

# JCTC

Journal of Chemical Theory and Computation

## An Error and Efficiency Analysis of Approximations to Møller–Plesset Perturbation Theory

Michael S. Marshall,<sup>†</sup> John S. Sears,<sup>\*,‡</sup> Lori A. Burns,<sup>†</sup> Jean-Luc Brédas,<sup>‡</sup> and C. David Sherrill<sup>\*,†</sup>

*Center for Computational Molecular Science and Technology, School of Chemistry and Biochemistry, and School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0400, United States, and Center for Computational Molecular Science and Technology, Center for Organic Photonics and Electronics, and School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332-0400, United States*

Received August 18, 2010

**Abstract:** We present a systematic study of the synergistic effects of popular approximations to Møller–Plesset perturbation theory through the second order (MP2). This work applies the density-fitting (DF) approximation for two-electron integrals, the dual-basis (DB) approximation for the Hartree–Fock reference, and the use of “heavy augmented” Dunning basis sets for basis set reduction, as well as combinations of these, to the S22 benchmark set of weakly bound dimers. For each approach, we report an error analysis as well as relative speedups for the 22 interaction energies in the set. Compared to the MP2/aug-cc-pVTZ level of theory, the DB-DF-MP2/heavy-aug-cc-pVTZ approach achieves an average speedup of 18 with a root-mean-square error of only 0.076 kcal mol<sup>−1</sup> (2%).

### 1. Introduction

In the past decade, the field of computational chemistry has demonstrated that high-level calculations on small molecules can in some cases achieve an accuracy comparable to that of experimental results.<sup>1–5</sup> A current challenge lies in the development of approximations to robust levels of theory to address larger systems of interest. Dispersion-dominated interactions, for which dynamic electron correlation has been shown to play an important role,<sup>6–11</sup> have attracted significant recent attention. When applying computational techniques to large-scale problems, long-range interactions can accumulate and must be accounted for properly. While an accurate description is provided by highly correlated methods such as coupled-cluster theory,<sup>12–14</sup> the steep computational cost of such methods constrains their applicability to systems

of but a few dozen atoms with modest basis sets. To overcome this problem, ongoing research efforts focus on two fronts: (1) the modification of established methods by adding adjustable parameters fit to experimental results or a higher level of theory or (2) the development of approximations to robust levels of theory that maintain their inherent accuracy while reducing the cost.

The strategy of incorporating *ad hoc* terms with fitted parameters has shown great success for methods such as DFT-D (which adds a scaled, damped dispersion correction to a DFT functional).<sup>15–17</sup> However, there remains no means of systematically improving the accuracy, and such methods sometimes require numerous parameters trained upon specific test sets to produce high-quality results. Correlated wave function methods have also been modified by fitted parameters in spin-component-scaled Møller–Plesset perturbation theory (SCS-MP2),<sup>18–22</sup> spin-opposite-scaled Møller–Plesset perturbation theory (SOS-MP2),<sup>23</sup> and spin-component-scaled coupled-cluster with singles and doubles theory (SCS-CCSD).<sup>24</sup> MP2 tends to give reasonably reliable results for certain types of noncovalent interactions (such as alkane–alkane interactions and H-bonded interactions). In cases where MP2

\* Corresponding authors: e-mail: sherrill@gatech.edu; john.sears@gatech.edu.

<sup>†</sup> School of Chemistry and Biochemistry and School of Computational Science and Engineering.

<sup>‡</sup> Center for Organic Photonics and Electronics and School of Chemistry and Biochemistry.

exhibits significant errors (e.g.,  $\pi$ -stacking interactions), the scaled MP2 methods such as SCS-MP2 tend to perform well.<sup>21</sup> Even in cases where very accurate binding energies are desired for noncovalent interactions, MP2 remains a critical ingredient in the theoretical procedure. When benchmark-quality results are needed, the current standard procedure is to evaluate the binding energies in the MP2 complete basis set limit and then to correct for higher-order correlation terms by adding a  $\Delta$ CCSD(T) correction [evaluated as the difference between CCSD(T) and MP2 binding energies in a smaller basis set].<sup>6,7</sup> Thus, whether one uses bare MP2, scaled MP2, or MP2 in conjunction with CCSD(T) corrections, MP2 computations remain important in studies of noncovalent interactions, and it is useful to explore approximations for speeding up these MP2 computations and to assess the associated errors.

In electronic structure theory, the evaluation and storage of four-index integrals is a common bottleneck. Various approaches to this problem have been explored, such as resolution of the identity<sup>25–33</sup> [now commonly referred to as density fitting (DF)], Cholesky decompositions<sup>34–43</sup> (CD), and pseudospectral techniques.<sup>44–46</sup> In the DF treatment, four-index integrals  $(\mu\nu|\rho\sigma)$  are approximated by summations over three-index quantities:

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

where  $[J^{-1}]_{PQ}$  is the inverse of the Coulomb metric evaluated in an auxiliary basis set:

$$[J]_{PQ} = \int P(\mathbf{r}_1) \frac{1}{r_{12}} Q(\mathbf{r}_2) d^3\mathbf{r}_1 d^3\mathbf{r}_2 \quad (2)$$

The three-index quantity  $(\mu\nu|P)$  serves to cast the product  $(\mu\nu|)$  onto the auxiliary basis via the Coulomb metric

$$(\mu\nu|P) = \int \mu(\mathbf{r}_1) \nu(\mathbf{r}_1) \frac{1}{r_{12}} P(\mathbf{r}_2) d^3\mathbf{r}_1 d^3\mathbf{r}_2 \quad (3)$$

While density-fitting does not lower the asymptotic scaling of MP2, it does reduce the prefactor significantly, with speedups in the range of 2 to 5.5 reported.<sup>32,47</sup> There exist many more methods for speeding up the evaluation of the correlation energy (e.g., local molecular orbital approaches such as local-MP2<sup>32,33</sup>), yet the application of DF alone is often sufficient to reduce the cost of the correlation energy computation to the point that the time needed for the underlying self-consistent field (SCF) becomes the rate-determining step.

Numerous algorithmic advances have been achieved over the past three decades to improve SCF efficiency. These range from Pulay's direct inversion of iterative subspace<sup>48,49</sup> (DIIS), which minimizes the number of SCF iterations, to modern linear scaling methods.<sup>50–54</sup> Two recent, similar advances in SCF theory are dual-basis techniques<sup>55–60</sup> to project the SCF energy from a smaller basis set and perturbative corrections to estimate the SCF complete basis set limit.<sup>61</sup>

The dual-basis (DB) approximation proposed in the work of Steele et al.<sup>60</sup> involves performing an iterative SCF in a

small basis and then taking a single Roothaan diagonalization step in a larger target basis set. In practice, the small basis is typically a specially designed subset of the target basis set, although this restriction is not imposed by the theory. Once the SCF is converged with the small basis set, the occupied molecular orbital (MO) coefficients are projected onto the larger basis via

$$\mathbf{C}_{\bar{\mu}i} = \sum_{\bar{\mu}\nu} \sum_{\lambda} \mathbf{S}_{\bar{\mu}\nu}^{-1} \mathbf{S}_{\nu\lambda} \mathbf{C}_{\lambda i} \quad (4)$$

where  $\mathbf{S}$  is the atomic orbital (AO) overlap matrix,  $i$  represents a MO index, Greek letters represent AO indices, and barred indices signify large-basis quantities. Using the newly constructed coefficient matrix, the new density matrix  $\mathbf{P}$  is formed, and a single Fock matrix is built and diagonalized. After including some first-order corrections, the DB-SCF energy is shown to be

$$E_{\text{dualbasis}} = E_{\text{smallbasis}} + \sum_{\bar{\mu}\nu} \Delta\mathbf{P}_{\bar{\mu}\nu} \mathbf{F}_{\bar{\mu}\nu} \quad (5)$$

where  $\Delta\mathbf{P} = \mathbf{P}' - \mathbf{P}$  is the difference between the postdiagonalization density matrix  $\mathbf{P}'$  and the small basis density matrix  $\mathbf{P}$  (projected into the large basis). The small truncated basis sets used in the dual-basis methods have already been implemented<sup>62</sup> in the Q-Chem 3.2 program suite for several Pople and Dunning basis sets.

Another broadly employed approximation is the truncation of the aug-cc-pVXZ ( $X = \text{D, T, Q}$ ) basis sets by eliminating diffuse functions from hydrogen atoms. These truncated basis sets are commonly referred to as heavy-aug-cc-pVXZ and are often abbreviated as haXZ ( $X = \text{D, T, Q}$ ). For biological applications and polymer studies, where a large number of hydrogens are present, haXZ can introduce a significant savings. Dropping augmented functions on hydrogen has been shown to have a small effect on properties such as interaction energies for nonbonded complexes.<sup>47</sup> The DF, DB, and haXZ approximations have all been developed independently. In this work, we systematically examine the practicability of combining these approximations and evaluate the magnitude of accumulated errors and attainable speedups. The S22 benchmark set<sup>63</sup> has been adopted because of its focus on noncovalent complexes, which are theoretically challenging.

Recent work by Steele et al. has shown that by combining DB and DF approximations within MP2, one can expect root-mean-square errors (RMSEs) of 0.043 and 0.019 kcal mol<sup>−1</sup> for the MP2/aug-cc-pVDZ and MP2/aug-cc-pVTZ levels of theory, respectively, for the S22 set.<sup>60</sup> Their timings focus mainly on evaluating DB-DF-MP2 analytical gradients. In this work, we present an error and efficiency analysis for each approximation independently, then repeat with the S22 set for the combination of approximations, thereby permitting dissection of any errors incurred in the energy, as well as elucidating the origins of the speedup. We also consider a series of linear alkanes to examine how these approximations behave as a function of increasing system size. Timings are compared to those from some other methods such as density functional theory.

## 2. Theoretical Approach

**2.1. Efficiency Study of Approximate MP2 on Linear Alkanes.** To evaluate gains by approximate MP2 methods, we examine a series of linear alkanes ( $C_nH_{2n+2}$ ). Recent work has considered the effect of RI,<sup>64,65</sup> Cholesky,<sup>66,67</sup> and atomic-orbital-based MP2<sup>68</sup> approximations on linear alkanes. Single-point energy computations were performed using B3LYP, DB-B3LYP, MP2, DB-MP2, DF-MP2, and DB-DF-MP2 with the aug-cc-pVDZ basis set. For all dual-basis approximations, we employed the optimized basis set of Steele et al.<sup>60</sup> referred to as racc-pVDZ, which has been shown to reproduce the target basis set (aug-cc-pVDZ) with minimal error in total energy. For the density-fitting auxiliary basis sets, we employed the basis set from Hättig and co-workers<sup>69,70</sup> referred to as rimp2-aug-cc-pVDZ. In this work, we will only be density-fitting the MP2 contribution, not the underlying SCF, as that capability is not currently implemented in Q-Chem 3.2. The frozen-core approximation was employed for all of the perturbative methods. Alkane geometries were constructed from the following parameters:  $r_{CC} = 1.53 \text{ \AA}$ ,  $r_{CH} = 1.09 \text{ \AA}$ , and  $\theta_{CCC} = 109.5^\circ$ . For each level of theory, we report the overall user time as well as a decomposition of SCF and MP2 user times. For the tests performed, I/O time was typically minor; hence user times were very similar to wall times.

All computations were performed without taking advantage of spatial symmetry. All alkane computations used the Q-Chem 3.2<sup>62</sup> program suite on an Altus 1702 server featuring dual AMD Opteron 2378 processors (2.4 GHz, Quad Core), 32 GB of DDR2 RAM, and  $2 \times 1 \text{ TB}$  7200 rpm RAID-0 local disks. The SCF was converged to  $10^{-8}$  hartree, and the integral threshold was  $10^{-13}$ .

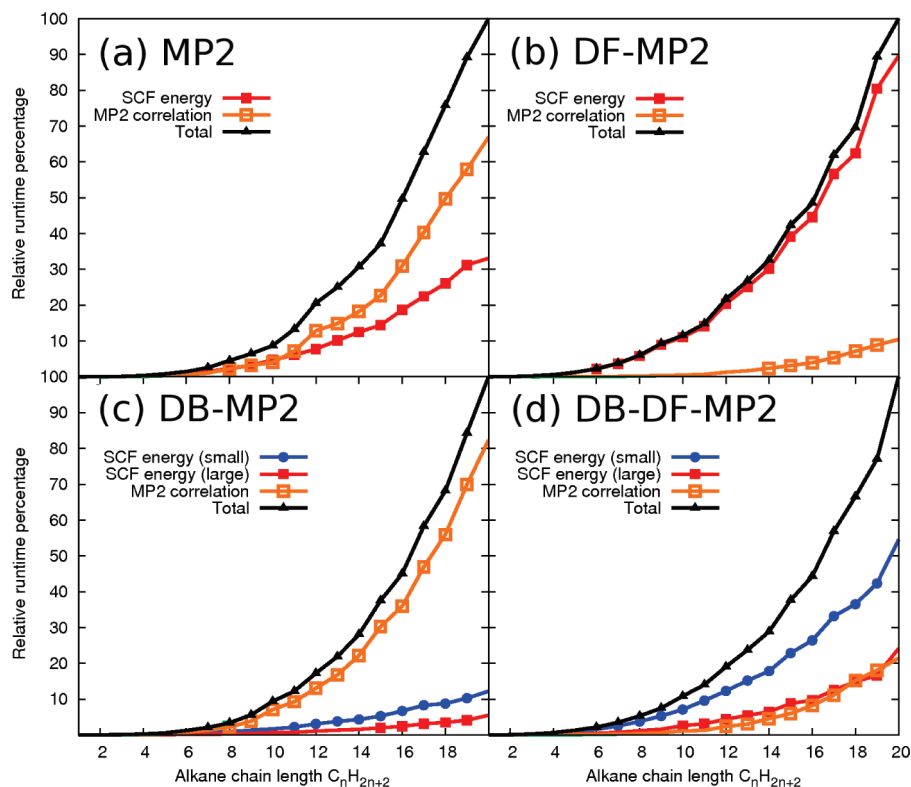
**2.2. Performance Analysis of Approximate MP2 on the S22 Set.** For a detailed analysis of the performance (both speedups and errors) by various approximate MP2 methods, we chose the S22 benchmark set,<sup>63</sup> which features diverse types of nonbonded interactions over a wide range of system sizes, from a water dimer (six atoms) to a hydrogen-bonded adenine-thymine complex (30 atoms). Benchmark-quality CCSD(T)/CBS reference binding energies are available for this test set.<sup>63,71</sup> For each of the 22 complexes, we report the interaction energy and the total user time for MP2, DB-MP2, DF-MP2, and DB-DF-MP2 with the aug-cc-pVDZ, aug-cc-pVTZ, heavy-aug-cc-pVDZ, and heavy-aug-cc-pVTZ Dunning basis sets [a heavy-aug-cc-pVXZ ( $X = D, T$ ) basis set consists of cc-pVXZ on hydrogen atoms and aug-cc-pVXZ on all other atoms]. The choice of the DB basis set and DF auxiliary basis is as described above. All interaction energies were corrected for basis-set superposition error (BSSE) using the counterpoise correction scheme outlined by Boys and Bernardi,<sup>72</sup> and individual calculations employed the frozen-core approximation. The benchmark machine for the S22 test set is an Intel Xeon (3.2 GHz, single core), with 4 GB of DDR2 RAM and a 150 GB local disk. The SCF was converged to  $10^{-8}$  hartree, and the integral threshold was  $10^{-13}$ . This work focuses on approximating three computations: MP2/aug-cc-pVDZ, MP2/aug-cc-pVTZ, and MP2/CBS(aDZ,aTZ), where CBS(aDZ,aTZ) refers to a

two-point extrapolation as defined by Halkier et al.<sup>73</sup> using aug-cc-pVDZ and aug-cc-pVTZ correlation energies. All computations in this part of the study use one of these three canonical MP2 results as a reference point.

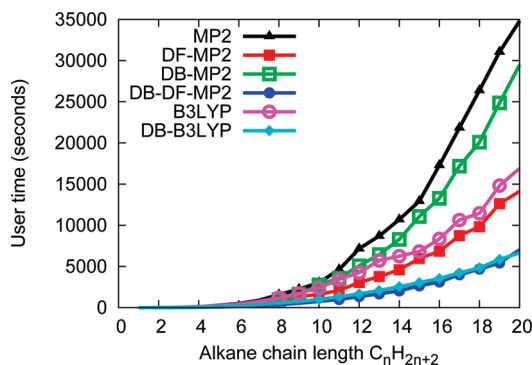
## 3. Results and Discussion

**3.1. Efficiency of Approximations to MP2 for Linear Alkanes.** In the following analysis, MP2 will be discussed in terms of two components: (a) the underlying SCF computation and (b) the evaluation of the MP2 correlation energy (including the transformation of the atomic orbital integrals to the molecular orbital basis). Using the alkane test cases, we investigate how the DF, DB, and haXZ approximations affect the speed and accuracy of the computation. For a medium-sized molecule such as  $C_{20}H_{42}$ , the underlying SCF takes 33% of the total user time as shown in Figure 1a. The MP2 contribution, which formally scales as  $O(N^5)$ , dominates over the SCF, which formally scales as  $O(N^4)$ , because of the AO to MO transformation required. (Of course, actual computational scalings with respect to system size will be lower than these formal scalings because of techniques such as integral prescreening; moreover, integral prescreening will work best in one-dimensional systems such as these.) As system size or the basis set increases, an increasing fraction of time will be spent in the MP2 portion of the computation. It is this costly step that DF abates by changing the complexity of the AO to MO transformation from  $O(N^5)$  to  $O(N^4)$  while increasing the correlation energy evaluation from  $O(N^4)$  to  $O(N^5)$ , but with a much lower prefactor than before. Figure 1b shows how DF-MP2 reduces the time to compute the MP2 correlation energy, shifting the majority of the computation time to the underlying SCF. For  $C_{20}H_{42}$ , 90% of the time to compute the DF-MP2 energy is spent in the SCF, to obtain an overall speedup of 2.46 relative to traditional MP2.

Now that the majority of the work has been shifted to the underlying SCF, we investigate dual-basis techniques that can drastically reduce the time to compute this stage. Figure 2 shows the DB-MP2 speedup to be only 1.18 relative to MP2 for  $C_{20}H_{42}$ , but this is only because of the large amount of time spent computing the correlation energy (and accordingly, the small cost of SCF) in the absence of the DF approximation. Figure 2 shows that the combination of the DF and DB approximations can yield speedups of 4.94 for  $C_{20}H_{42}$ . Within DB-DF-MP2, the bottleneck is the SCF in the small basis, which consumes 54% of the total compute time for  $C_{20}H_{42}$ . To put these improvements into context, we compared each method to B3LYP and dual-basis B3LYP (DB-B3LYP) in Figure 2, demonstrating that DF-MP2 is competitive with B3LYP and that DB-DF-MP2 is competitive with DB-B3LYP for system sizes upward of  $C_{20}H_{42}$ . This implies that the overhead in computing the correlation contribution within DFT is comparable to the time to compute the density fitted correlation energy in MP2. Note, however, that DF-MP2 and DB-DF-MP2 are still not competitive with any pure DFT method that lacks Hartree–Fock exchange.



**Figure 1.** The relative user time and decompositions of (a) MP2, (b) DF-MP2, (c) DB-MP2, and (d) DB-DF-MP2 are shown for the alkane series methane ( $CH_4$ ) through dodecane ( $C_{20}H_{42}$ ). The percentages are computed by taking the  $C_{20}H_{42}$  as a reference, with the total broken into SCF and MP2 correlation components. For the DB approximations, SCF (small) refers to the percentage of time to solve the iterative part, and SCF (large) refers to the percentage of time to perform the single Fock build in the target basis.



**Figure 2.** Total user times of MP2, DF-MP2, DB-MP2, DB-DF-MP2, B3LYP, and DB-B3LYP all with aug-cc-pVDZ basis set for the alkanes methane ( $CH_4$ ) through dodecane ( $C_{20}H_{42}$ ).

For the systems investigated, the errors of all of these approximations scale linearly with system size. DB incurs an average error of  $0.027 \text{ kcal mol}^{-1}$  per atom, DF incurs an error of  $0.006 \text{ kcal mol}^{-1}$  per atom, and DB-DF incurs an error of  $0.033 \text{ kcal mol}^{-1}$  per atom.

**3.2. Performance Analysis of Approximate MP2 on the S22 Set.** To analyze the error introduced by the DB, DF, and haXZ approximations, we compute interaction energies for each of the 22 complexes in the S22 benchmark test set.<sup>63</sup> Table 1 presents the root-mean-square error (RMSE) for the test set at each level of theory. In trying to reproduce MP2/aug-cc-pVDZ results, we see that the RMSE from DF is  $0.003 \text{ kcal mol}^{-1}$  while achieving an average

speedup of 1.3. The magnitude of this error is very reasonable when compared to other remaining errors such as basis set incompleteness error (BSIE). On the other hand, the DB approximation incurs a RMSE of  $0.043 \text{ kcal mol}^{-1}$ . While an order of magnitude larger, the DB error is still rather small, especially considering it has a speedup of 1.78. Applying both of the approximations simultaneously shows that they do indeed compound well, achieving a speedup of 3.1, but the errors are additive also, for a RMSE of  $0.045 \text{ kcal mol}^{-1}$ . The use of heavy-aug-cc-pVDZ in place of aug-cc-pVDZ yields a speedup of 1.74 but at the cost of  $0.120 \text{ kcal mol}^{-1}$  average error. This average error is still small considering that the S22 MP2/aug-cc-pVDZ interaction energies range from  $-0.39$  to  $-18.41 \text{ kcal mol}^{-1}$ , but it may not be acceptable in some high-accuracy applications. The largest error introduced by neglecting diffuse functions on H atoms is  $0.22 \text{ kcal mol}^{-1}$  for the ethylene dimer test case, which has four closely packed hydrogens. Heavy-augmented basis sets should be avoided for systems with multiple hydrogen–hydrogen contacts, such as methane and ethene dimers.

To determine how well these approximations perform for larger basis sets, they were also tested against the MP2/aug-cc-pVTZ level of theory. For this larger basis set, RMSEs are reduced for all three approximations: DB has a RMSE of  $0.017 \text{ kcal mol}^{-1}$  (down from  $0.043 \text{ kcal mol}^{-1}$ ), DF has a RMSE of  $0.001 \text{ kcal mol}^{-1}$  (down from  $0.003 \text{ kcal mol}^{-1}$ ), and the use of heavy-augmented basis sets has a RMSE of  $0.070 \text{ kcal mol}^{-1}$  (down from  $0.120 \text{ kcal mol}^{-1}$ ). When all



**Table 1.** Mean Unsigned Error (MUE), Root Mean Square Error (RMSE), Average Percent Error, and Average Speedup Analysis of Approximating MP2/aug-cc-pVXZ and MP2/CBS(aDZ,aTZ) for the S22 Test Set of Complexes<sup>a</sup>

reference	level of theory	speedup	MUE	RMSE	% error
MP2/aug-cc-pVDZ	MP2/haDZ	1.74	0.113	0.120	4.26
	DB-MP2/haDZ	2.61	0.106	0.120	4.05
	DF-MP2/haDZ	2.29	0.107	0.116	4.20
	DB-DF-MP2/haDZ	4.12	0.104	0.119	3.99
	DB-MP2/aDZ	1.78	0.034	0.043	0.82
	DF-MP2/aDZ	1.31	0.002	0.003	0.04
	DB-DF-MP2/aDZ	3.09	0.036	0.045 <sup>b</sup>	0.84
	MP2/haTZ	1.87	0.068	0.070	1.95
MP2/aug-cc-pVTZ	DB-MP2/haTZ	5.76	0.072	0.077	2.01
	DF-MP2/haTZ	2.43	0.066	0.069	1.91
	DB-DF-MP2/haTZ	18.04	0.071	0.076	1.97
	DB-MP2/aTZ	3.25	0.012	0.017	0.20
	DF-MP2/aTZ	1.30	0.001	0.001	0.02
	DB-DF-MP2/aTZ	10.73	0.012	0.017 <sup>c</sup>	0.19
	MP2/CBS(haDZ,haTZ)	1.86	0.039	0.044	0.94
	DB-MP2/CBS(haDZ,haTZ)	5.48	0.050	0.057	1.08
MP2/CBS(aDZ,aTZ)	DF-MP2/CBS(haDZ,haTZ)	2.42	0.038	0.043	0.92
	DB-DF-MP2/CBS(haDZ,haTZ)	15.73	0.049	0.056	1.05
	DB-MP2/CBS(aDZ,aTZ)	3.14	0.017	0.023	0.26
	DF-MP2/CBS(aDZ,aTZ)	1.30	0.001	0.001	0.01
	DB-DF-MP2/CBS(aDZ,aTZ)	9.65	0.017	0.022	0.26

<sup>a</sup> All errors in kilocalories per mole. <sup>b</sup> Reference 60 reports a RMSE of 0.043 kcal mol<sup>-1</sup>. <sup>c</sup> Reference 60 reports a RMSE of 0.019 kcal mol<sup>-1</sup>.

**Table 2.** Mean Unsigned Error (MUE), Root-Mean-Square Error (RMSE), and Average Percent Error for the Interaction Energies for Each Subgroup in the S22 Test Set in Kilocalories Per Mole<sup>a</sup>

level of theory	H bonding			dispersion			mixed		
	MUE	RMSE	% error	MUE	RMSE	% error	MUE	RMSE	% error
MP2/CBS(haDZ,haTZ)	0.059	0.061	0.64	0.029	0.033	1.26	0.031	0.032	0.88
DB-MP2/CBS(haDZ,haTZ)	0.057	0.065	0.56	0.051	0.057	1.60	0.043	0.047	1.00
DF-MP2/CBS(haDZ,haTZ)	0.058	0.060	0.63	0.027	0.032	1.21	0.030	0.031	0.86
DB-DF-MP2/CBS(haDZ,haTZ)	0.057	0.064	0.55	0.049	0.055	1.56	0.042	0.046	0.98
DB-MP2/CBS(aDZ,aTZ)	0.020	0.025	0.17	0.018	0.024	0.29	0.015	0.017	0.32
DF-MP2/CBS(aDZ,aTZ)	0.000	0.000	0.00	0.001	0.001	0.02	0.000	0.000	0.01
DB-DF-MP2/CBS(aDZ,aTZ)	0.019	0.025	0.17	0.018	0.023	0.28	0.015	0.017	0.32

<sup>a</sup> All values are relative to MP2/CBS(aDZ,aTZ).

three approximations are combined, speedups of 18.0 are achieved at the cost of 0.076 kcal mol<sup>-1</sup> RMSE. Considering the large gain in computational efficiency, these errors are tolerable, and DB-DF-MP2/haTZ is recommended for typical studies of nonbonded interactions.

We also examined how complete basis set (CBS) extrapolations affect the error for each approximation. CBS extrapolations consistently reduce the RMSE (shown at the bottom of Table 1) for the approximations considered. The extrapolations particularly abate the error caused by the use of heavy-augmented basis sets, reclaiming 0.026 kcal mol<sup>-1</sup> on average. The compounding of all three approximations and CBS extrapolations [DB-DF-MP2/CBS(haDZ,haTZ)] yields a RMSE of 0.056 kcal mol<sup>-1</sup> and a speedup of 15.7. This speedup is not quite as large as that observed for the DB-DF-MP2/haTZ (18.0), because the CBS extrapolations include haDZ computations which have a lesser efficiency gain. We note that MP2/CBS(aDZ,aTZ) has a 0.118 kcal mol<sup>-1</sup> RMSE compared to MP2/CBS(aTZ,aQZ)<sup>71</sup> for the S22 test set.

The DF-MP2 speedups in our study are not as large as might be expected. We were forced to use a core Hamiltonian guess to be consistent between the DF and DB tests, because in Q-Chem 3.2, one cannot use the DB technique in

conjunction with more advanced initial orbital guesses. The core Hamiltonian guess requires more SCF iterations to converge, thereby increasing the time spent in SCF. If superior SCF guesses were used, such as superposition of atomic densities (SAD) or a small basis projection, the overall computation would spend less time in the SCF and more time in the MP2 correlation. This would cause the DF methods to have better overall speedups and DB methods to have slightly smaller speedups.

To better understand the errors incurred through these approximations, a decomposition by binding type is shown in Table 2. Reference 63 defines the division of complexes between hydrogen-bonded, dispersion-bound, and mixed-influence subgroups. As shown in Table 2, dispersion-bound complexes experience a larger mean percent error than the hydrogen-bonded subset for every approximation examined, by a factor of 1.7–3.9, thereby suggesting the approximations examined in this work, particularly haXZ, may have difficulty with longer-range interactions. For the CBS limit, we report errors among dispersion-dominated systems of 0.02%, 0.29%, 0.28%, and 1.26% for DF, DB, DB-DF, and heavy-augmented basis sets, respectively, while the corresponding value for hydrogen-bonded complexes in the last case is only 0.64%.

## 4. Conclusions

This work demonstrates that with a careful choice of approximations, MP2-quality results can be computationally affordable for systems with a few dozen atoms or larger without introducing significant error. Density fitting reduces the time to compute the MP2 correlation energy. Dual-basis techniques abate the cost of the underlying SCF, and heavy-augmented functions speed up both parts of the computation relative to the fully augmented basis sets. Except for comparisons using the smaller heavy-aug-cc-pVDZ basis set, all of these approximations show a significant speedup while never incurring a RMSE greater than 0.045 kcal mol<sup>-1</sup> for the S22 test cases. We also demonstrate that all of these approximations do indeed combine very efficiently. In future tests, density fitting will be extended to the SCF stage (currently not implemented in Q-Chem). The use of DF within the DB-SCF framework should be a significant stride toward achieving a level of theory that is not only accurate but applicable to a wide range of systems. Q-Chem also will soon have the capability to perform perturbative SCF approaches as outlined in the work of Gill et al.<sup>61</sup> These new computational tools will open up larger systems of interest to *ab initio* techniques while introducing errors which are negligible in most applications.

**Acknowledgment.** This material is based upon work supported by the National Science Foundation (Grant No. CHE-1011360). The computer resources of the Center for Computational Molecular Science and Technology are funded through an NSF CRIF Award (CHE-0946869).

## References

- Lee, T. J.; Scuseria, G. E. Achieving Chemical Accuracy with Coupled-Cluster Theory. In *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy*; Langhoff, S. R., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1995; pp 47–108.
- Polyansky, O. L.; Zobov, N. F.; Viti, S.; Tennyson, J.; Bernath, P. F.; Wallace, L. *Science* **1997**, 277, 346.
- Polyansky, O. L.; Császár, A. G.; Shirin, S. V.; Zobov, N. F. *Science* **2003**, 299, 539–542.
- Martin, J. M. L. *Chem. Phys. Lett.* **1998**, 292, 411–420.
- Temelso, B.; Valeev, E. F.; Sherrill, C. D. *J. Phys. Chem. A* **2004**, 108, 3068–3075.
- Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2002**, 124, 104–112.
- Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, 124, 10887–10893.
- Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2004**, 108, 10200–10207.
- Lee, E. C.; Kim, D.; Jurečka, P.; Tarakeshwar, P.; Hobza, P.; Kim, K. S. *J. Phys. Chem. A* **2007**, 111, 3446–3457.
- Pitoňák, M.; Riley, K. E.; Neogrády, P.; Hobza, P. *ChemPhysChem* **2008**, 9, 1636–1644.
- Sherrill, C. D.; Sumpter, B. G.; Sinnokrot, M. O.; Marshall, M. S.; Hohenstein, E. G.; Walker, R. C.; Gould, I. R. *J. Comput. Chem.* **2009**, 30, 2187–2193.
- Purvis, G. D.; Bartlett, R. J. *J. Chem. Phys.* **1982**, 76, 1910–1918.
- Scuseria, G. E.; Scheiner, A. C.; Lee, T. J.; Rice, J. E.; Schaefer, H. F. *J. Chem. Phys.* **1987**, 86, 2881.
- Scuseria, G. E.; Janssen, C. L.; Schaefer, H. F. *J. Chem. Phys.* **1988**, 89, 7382.
- Wu, Q.; Yang, W. *J. Chem. Phys.* **2002**, 116, 515–524.
- Grimme, S. *J. Comput. Chem.* **2004**, 25, 1463–1473.
- Grimme, S. *J. Comput. Chem.* **2006**, 27, 1787–1799.
- Grimme, S. *J. Chem. Phys.* **2003**, 118, 9095–9102.
- Gerenkamp, M.; Grimme, S. *Chem. Phys. Lett.* **2004**, 392, 229–235.
- Hill, J. G.; Platts, J. A. *J. Chem. Theory Comput.* **2007**, 3, 80–85.
- Antony, J.; Grimme, S. *J. Phys. Chem. A* **2007**, 111, 4862–4868.
- Takatani, T.; Sherrill, C. D. *Phys. Chem. Chem. Phys.* **2007**, 9, 6106–6114.
- Jung, Y.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. *J. Chem. Phys.* **2004**, 121, 9793–9802.
- Takatani, T.; Hohenstein, E. G.; Sherrill, C. D. *J. Chem. Phys.* **2008**, 128, 124111.
- Whitten, J. L. *J. Chem. Phys.* **1973**, 58, 4496–4501.
- Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. *Int. J. Quantum Chem. Symp.* **1977**, 11, 81.
- Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. *J. Chem. Phys.* **1979**, 71, 3396–3402.
- Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, 208, 359–363.
- Bernholdt, D. E.; Harrison, R. J. *Chem. Phys. Lett.* **1996**, 250, 477–484.
- Kendall, R. A.; Fruchtl, H. A. *Theor. Chem. Acc.* **1997**, 97, 158–163.
- Weigend, F. *Phys. Chem. Chem. Phys.* **2002**, 4, 4285–4291.
- Werner, H.-J.; Manby, F. R.; Knowles, P. J. *J. Chem. Phys.* **2003**, 118, 8149–8160.
- Werner, H.; Manby, F. *J. Chem. Phys.* **2006**, 124, 054114.
- Beebe, N. H. F.; Linderberg, J. *Int. J. Quantum Chem.* **1977**, 12, 683–705.
- Roeggen, I.; Wisloff-Nilssen, E. *Chem. Phys. Lett.* **1986**, 132, 154–160.
- Koch, H.; de Meras, A. S.; Pedersen, T. B. *J. Chem. Phys.* **2003**, 118, 9481–9484.
- Aquilante, F.; Pedersen, T. B.; Lindh, R. *J. Chem. Phys.* **2007**, 126, 194106.
- Boström, J.; Aquilante, F.; Pedersen, T. B.; Lindh, R. *J. Chem. Theory Comput.* **2009**, 5, 1545–1553.
- Weigend, F.; Kattannek, M.; Ahlrichs, R. *J. Chem. Phys.* **2009**, 130, 164106.
- Aquilante, F.; Gagliardi, L.; Pedersen, T. B.; Lindh, R. *J. Chem. Phys.* **2009**, 130, 154107.
- Zienau, J.; Clin, L.; Doser, B.; Ochsenfeld, C. *J. Chem. Phys.* **2009**, 130, 204112.

- (42) Boström, J.; Delcey, M. G.; Aquilante, F.; Serrano-Andrés, L.; Pedersen, T. B.; Lindh, R. *J. Chem. Theory Comput.* **2010**, *6*, 747–754.
- (43) Chwee, T. S.; Carter, E. A. *J. Chem. Phys.* **2010**, *132*, 074104.
- (44) Martinez, T. J.; Mehta, A.; Carter, E. A. *J. Chem. Phys.* **1992**, *97*, 1876–1880.
- (45) Martinez, T. J.; Carter, E. A. Pseudospectral Methods Applied to the Electron Correlation Problem. In *Modern Electronic Structure Theory*; Yarkony, D. R., Ed.; World Scientific: Singapore, 1995; Vol. 2, pp 1132–1165.
- (46) Friesner, R. A.; Murphy, R. B.; Beachy, M. D.; Ringnalda, M. N.; Pollard, W. T.; Dunietz, B. D.; Cao, Y. *J. Phys. Chem. A* **1999**, *103*, 1913–1928.
- (47) Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. *J. Phys. Chem. A* **2009**, *113*, 10146–10159.
- (48) Pulay, P. *J. Phys. Lett.* **1980**, *73*, 323–398.
- (49) Pulay, P. *J. Comput. Chem.* **1982**, *3*, 556–560.
- (50) White, C.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett.* **1996**, *253*, 268–278.
- (51) Strain, M. C.; Scuseria, G. E.; Frisch, M. J. *Science* **1996**, *271*, 51–53.
- (52) Burant, J. C.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **1996**, *105*, 8969–8972.
- (53) Schwegler, E.; Challacombe, M. *J. Chem. Phys.* **1996**, *105*, 2726–2734.
- (54) Ochsenfeld, C.; White, C. A.; Head-Gordon, M. *J. Chem. Phys.* **1998**, *109*, 1663–1669.
- (55) Wolinski, K.; Pulay, P. *J. Chem. Phys.* **2003**, *118*, 9497–9503.
- (56) Steele, R. P.; DiStasio, R. A.; Shao, Y.; Kong, J.; Head-Gordon, M. *J. Chem. Phys.* **2006**, *125*, 074108.
- (57) Steele, R.; Shao, Y.; DiStasio, R.; Head-Gordon, M. *J. Chem. Phys. A* **2006**, *110*, 13915–13922.
- (58) Distasio, R.; Steele, R.; Head-Gordon, M. *Mol. Phys.* **2007**, *105*, 2731–2742.
- (59) Steele, R.; Head-Gordon, M. *Mol. Phys.* **2007**, *105*, 2455–2473.
- (60) Steele, R. P.; DiStasio, R. A.; Head-Gordon, M. *J. Chem. Theory Comput.* **2009**, *5*, 1560–1572.
- (61) Deng, J.; Gilbert, A. T. B.; Gill, P. M. W. *J. Chem. Phys.* **2009**, *130*, 231101.
- (62) Shao, Y.; et al. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.
- (63) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.
- (64) Sodt, A.; Beran, G. J. O.; Jung, Y. S.; Austin, B.; Head-Gordon, M. *J. Chem. Theory Comput.* **2006**, *2*, 300–305.
- (65) Sodt, A.; Subotnik, J. E.; Head-Gordon, M. *J. Chem. Phys.* **2006**, *125*, 194109.
- (66) Aquilante, F.; Pedersen, T. B. *Chem. Phys. Lett.* **2007**, *449*, 354.
- (67) Zienau, J.; Clin, L.; Doser, B.; Ochsenfeld, C. *J. Chem. Phys.* **2009**, *130*, 204112.
- (68) Doser, B.; Lambrecht, D. S.; Kussmann, J.; Ochsenfeld, C. *J. Chem. Phys.* **2009**, *130*, 064107.
- (69) Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, *116*, 3175–3183.
- (70) Hättig, C. *Phys. Chem. Chem. Phys.* **2005**, *7*, 59–66.
- (71) Takatani, T.; Hohenstein, E. G.; Malagoli, M.; Marshall, M. S.; Sherrill, C. D. *J. Chem. Phys.* **2010**, *132*, 144104.
- (72) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.
- (73) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243–252.

CT100468F