

Dual-basis second-order Møller-Plesset perturbation theory: A reduced-cost reference for correlation calculations

Ryan P. Steele^{a)} and Robert A. DiStasio, Jr.

*Department of Chemistry, University of California, Berkeley, California 94720
and Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720*

Yihan Shao and Jing Kong

Q-Chem, Inc., Pittsburgh, Pennsylvania 15213

Martin Head-Gordon^{b)}

*Department of Chemistry, University of California, Berkeley, California 94720
and Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720*

(Received 13 March 2006; accepted 26 June 2006; published online 18 August 2006)

The resolution-of-the-identity (RI) approximation has placed the onus of the cost of a second-order Møller-Plesset (MP2) calculation on the underlying self-consistent field (SCF) calculation for many moderately sized molecules. A dual-basis approach to the SCF calculation, based on previous methods demonstrated for density functional theory, is combined with RI-MP2 calculations, and small basis subsets for cc-pVTZ, cc-pVQZ, and 6-311++G(3df,3pd) are presented. These subsets provide time savings of greater than 90%, with negligible errors in absolute and relative energies, compared to the associated full-basis counterpart. The method is tested with a series of rotational barriers, relative conformational energies of alanine tetrapeptides, as well as the full G3/99 molecular set. RI-MP2 calculations on alanine octapeptides (40 heavy atoms, 3460 basis functions), using cc-pVQZ, are presented. Results improve upon previous methods that diagonalize the virtual space separately. © 2006 American Institute of Physics. [DOI: 10.1063/1.2234371]

I. INTRODUCTION

One of the most common and fundamental approximations in electronic structure theory is the expansion of molecular orbitals as linear combinations of atomic basis functions. Several recent case studies demonstrate that large basis sets are needed in order to accurately capture physical properties,^{1–10} and, in some cases, these large basis sets are required even for qualitatively correct predictions.^{11,12} While self-consistent field (SCF) calculations are believed to converge relatively quickly with respect to basis set size, electron correlation calculations are often strongly basis set dependent^{6,13} and may require enormous basis sets with many angular and radial nodes to reach completeness.

Second-order Møller-Plesset perturbation theory¹⁴ (MP2) calculations are often the first single-determinant electron correlation method of choice beyond the SCF level. The local MP2 algorithms of Pulay and co-workers^{15–19} and Werner and co-workers^{20,21} have very successfully reduced the scaling of MP2 correlation calculations and approach linear (system size) scaling for large systems. Ayala *et al.* have also implemented an atomic-orbital-based linear scaling MP2,^{22,23} based upon the Laplace transform methods of Häser²⁴ and Almlöf and co-workers.^{25–27} An alternative approach, based on auxiliary basis set fitting and often called the resolution-of-the-identity (RI) approximation,^{28–33} has reduced the overall computational prefactor while leaving the

formal scaling unchanged. RI methods,²¹ as well as further approximations (RI-DF-LMP2,²¹ RI-TRIM,¹² RI-MOS,³⁴ RI-SOS,³⁵ etc.), have drastically reduced the MP2 correction time, leaving the SCF step time dominant for a large class of molecules. For example, in the recent publication of DiStasio, Jr. *et al.*,¹² the RI-TRIM MP2 correction time for a sample tetrapeptide (at the cc-pVQZ level) was 2 hours, while the preceding SCF calculation required approximately 6 days.

The dual-basis SCF (DB-SCF) approach demonstrated by Liang and Head-Gordon³⁶ is well suited for addressing this timing imbalance. In this method, an SCF calculation is converged in a small basis set; the density matrix is subsequently projected into the large, or target, basis set, and an energy correction is applied after only a single Roothaan step is taken. For basis sets yielding accurate correlation energies, tractable calculations are still in the regime where construction of the Fock matrix is computationally dominant. Thus, if we assume that at least ten SCF cycles are required for convergence, computational savings at the SCF level should be approximately 90%. For the *small/target* combination of 6-311G*/6-311++G(3df,3pd), both absolute energy and atomization energy errors were shown to be less than 0.2 kcal/mol for the G2/91 (Ref. 37) set of molecules, relative to B3LYP^{38,39}/6-311++G(3df,3pd) results. In this work, we utilize the molecular orbitals generated by the DB-SCF procedure to obtain correlation corrections at the MP2 (or RI-MP2) level of theory. One additional advantage to the dual-basis scheme is its intrinsic amenability to other SCF algorithms. Linear scaling,^{40–45} RI-HF,^{28,46–48} and pseu-

^{a)}Electronic mail: ofer4@bastille.cchem.berkeley.edu

^{b)}Author to whom correspondence should be addressed. Electronic mail: mhg@bastille.cchem.berkeley.edu

dospectral methods^{49–54} all provide possible future augmentations to our method.

Extending a dual-basis scheme to MP2 calculations was originally proposed by Jurgens-Lutovsky and Almlöf⁵⁵ and, more recently, implemented by Wolinski and Pulay.⁵⁶ Lee and Head-Gordon⁵⁷ also implemented a similar perturbative method for polarized atomic orbital (PAO-SCF) calculations. Though these methods are discussed in more detail below, their distinguishing feature is the separate diagonalization of the occupied-occupied and virtual-virtual (*oo* and *vv*) subspaces of the large-basis Fock matrix, often termed “pseudocanonization.” Difficulty in accurately capturing the virtual space in correlation calculations logically leads to this expansion for the MP2 calculations; however, maintaining the strong orthogonality between small- and large-basis orbitals is not required and, as shown below, leads to unnecessary inflexibility in the orbitals. The method we present here utilizes a full diagonalization of the large-basis Fock matrix, in which extension basis functions are allowed to mix in the occupied and virtual orbital coefficients.

II. METHODS

In this section, the details of the DB-SCF/MP2 method and its implementation, as well as the calculations to test its robustness, are presented. Lowercase letters are used to represent matrices in the small basis; uppercase letters denote matrices in the target basis. When applicable, Greek letters $\{\mu, \nu\}$ signify atomic orbital (AO) basis indices, while Roman letters $\{i, j\}$ and $\{a, b\}$ signify occupied and virtual molecular orbital (MO) indices, respectively.

A. Dual-basis SCF/MP2

Upon choosing a suitable small basis (sub)set, an SCF calculation is performed to convergence, yielding a small-basis SCF energy,

$$E_{\text{small}} = \text{Tr}\left[p h + \frac{1}{2} p I p\right], \quad (1)$$

where h is the one-electron Hamiltonian and I is the fourth-rank tensor containing the full set of electron-repulsion integrals in the AO basis.

The one-particle density matrix p is subsequently projected onto the large basis set. From this small-basis density matrix (now defined in the larger space as P), a new Fock matrix F is built in the large-basis space via $F = H + I P$. A full diagonalization (or equivalent³⁶) provides the new MO coefficients C' and, subsequently, a new density matrix P' . The energy after this single Roothaan step is given as

$$E = \text{Tr}\left[(P + \Delta P)H + \frac{1}{2}(P + \Delta P)I(P + \Delta P)\right], \quad (2)$$

where ΔP is defined as $P' - P$. By ignoring terms that are quadratic in the change in density matrix (the terms that would require a second large-basis Fock build), the energy lowering due to the single Roothaan step is represented by the dual-basis energy correction:

$$E_{\text{dual}} = \text{Tr}[(\Delta P)F], \quad (3)$$

where F is the Fock matrix built from P . In other words, only a single⁵⁸ Roothaan step is taken in the large basis, and

the energy of this step is approximated by Eq. (3).

Using the new MO coefficients and orbital energies, the MP2 energy correction can be constructed in the usual fashion:

$$E_{\text{MP2}} = -\frac{1}{4} \sum_{ij} \sum_{ab} \frac{(ij \| ab)^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j}, \quad (4)$$

or can be implemented using more advanced local or density-fitting algorithms. The sum of contributions (1), (3), and (4) defines a dual-basis MP2 energy, where (3) provides the leading contribution due to orbital relaxation and (4) is the pair correlation energy.

Since a new set of occupied-space MO coefficients is obtained (and hence, the density matrix has changed), the current Fock matrix is no longer the Fock matrix built from the new density, as would be the case for a converged SCF. For completeness, the contribution of the perturbative singles (rigorously nonzero since the Brillouin theorem is not satisfied in the large basis) should also be included:

$$E_{\text{singles}} = -\sum_{ia} \frac{F'_{ia}{}^2}{\epsilon_a - \epsilon_i}, \quad (5)$$

where F' is a new (second) Fock matrix built from P' . The magnitude of the singles contribution is discussed below.

This code has been implemented in a development version of Q-CHEM 3.0.⁵⁹ We have also implemented, for comparison, the dual-basis method of Wolinski and Pulay.⁵⁶ Since the separate *oo/vv* diagonalization leaves the density matrix unchanged, term (3) is, by definition, zero. Term (5) is the only correction to the SCF energy.

The computational savings of DB-SCF/MP2 are twofold. First, since only one large-basis Fock build is required, the cost of the SCF portion is reduced by a factor approaching the number of SCF steps typically required for convergence. The formal scaling of an SCF calculation is $O(N^4)$, where N is the size of the basis set. In practice, for a fixed basis set size per atom and varying molecular size, this scaling has been reduced to methods approaching $O(N)$. However, for fixed molecular size and varying basis set size, the aforementioned $O(N^4)$ scaling is still accurate. Savings due to DB-SCF, therefore, should be on the order of

$$\frac{t_{\text{full}}}{t_{\text{dual}}} \geq \left[\frac{(N/n)^4 \times C_N}{[(N/n)^4 \times 1] + C_n} \right], \quad (6)$$

where C_i is the number of SCF cycles in each basis, and n and N correspond to the size of the small and large basis sets, respectively. [Note that the $(N/n)^4$ savings alone neglects the often dominant cost of one SCF step in the large basis]. Second, strong integral screening can be applied to the Fock build and accounts for the inequality in the factor above. During the AO-integral portion of the Fock build ($I_{\mu\nu\lambda\sigma} P_{\lambda\sigma}$), integrals involving $\lambda\sigma$ pairs corresponding to $P_{\lambda\sigma} = 0$ can be excluded. Note that these integrals are not rigorously zero; however, they do not contribute to the dual-basis energy, since the projected density matrix has all new basis function terms equal to zero.⁶⁰ Given this screening, the scaling of the lone Fock build can be estimated. For the simultaneous con-

TABLE I. Subsets used for DB-SCF/MP2 calculations using Dunning-type basis sets. The most diffuse functions are listed at the bottom of each set.

| Atoms | Basis | Full | Subset | Basis | Full | Subset |
|-----------------|---------|-------------|----------------|----------|--------------|-------------------|
| H–He | cc-pVTZ | <i>spd</i> | <i>sp</i> | cc-pVQZ | <i>spdf</i> | <i>s</i> |
| | | <i>sp</i> | <i>s</i> | | <i>spd</i> | <i>sp</i> |
| | | <i>s</i> | <i>s</i> | | <i>sp</i> | <i>s</i> |
| | | | | | <i>s</i> | <i>s</i> |
| Basis functions | | 14 | 6 | | 30 | 7 |
| Li–Ne | | <i>spdf</i> | <i>spd</i> | | <i>spdfg</i> | <i>spd</i> |
| | | <i>spd</i> | <i>spd</i> | | <i>spdf</i> | <i>spd</i> |
| | | <i>sp</i> | <i>sp</i> | | <i>spd</i> | <i>spd</i> |
| | | <i>s</i> | <i>s</i> | | <i>sp</i> | <i>sp</i> |
| | | | | <i>s</i> | <i>s</i> | |
| Basis functions | | 30 | 23 | | 55 | 32 |
| Na–Ar | | (Varies) | (No <i>f</i>) | | (Varies) | (No <i>f, g</i>) |

struction of exchange and Coulomb contributions, this scaling is

$$\mathcal{O}\left(\frac{3N'^2n^2}{4} + \frac{N'n^3}{2} + \frac{n^4}{8}\right), \quad (7)$$

where N' is the number of strictly large-basis functions, i.e., $N' = N - n$. Contrast this cost with the standard scaling of the Fock matrix construction, $\mathcal{O}((N' + n)^4/8)$, where the factor of 8 stems from the permutational symmetry of the integrals.

In previous dual-basis MP2 implementations, integral savings have also been exploited during the first two quarter-transformations at the MP2 level. The motivation for this paper, however, is based upon the significant savings already inherent in the RI approximation. No additional screening has been implemented at the RI-MP2 level, though its addition to a more optimized code is possible.

B. Basis sets

Several preliminary tests were performed to obtain optimal basis set combinations that are (i) conservative enough to maintain useful accuracy and (ii) disparate enough to make use of the savings available in a large truncation. We choose to use only proper subsets of the target basis set (a property that is also exploited in the DB-SCF gradient in a forthcoming paper). Dunning-type sets^{61–66} are common and attractive sets for correlation calculations, particularly due to the availability of accurate extrapolation schemes⁶⁷ that permit the estimation of properties beyond those of the largest basis set used. However, the cc-pVXZ ($X = D, T, Q, 5, 6, \dots$) series is not comprised of sequential supersets, so we have constructed our own truncated versions of these basis sets. The scheme for Dunning-type subsets is summarized in Table I. Since the T \rightarrow Q extrapolation (see below) is our ultimate goal, the truncation of cc-pVTZ is more modest than that for cc-pVQZ. Only the set of *f* functions has been removed from second-row atoms in cc-pVTZ, whereas both the set of *f* functions and the set of *g* functions have been removed from cc-pVQZ. (Retaining only one set of *d* functions was also considered but yielded slightly higher errors than desired.) Interestingly, one *p*-type (polarization)

function was found to be necessary on hydrogen in both subsets, in addition to the retention of all *s* functions. The most compact *p* function is retained in the cc-pVTZ subset; the middle *p* function is retained in cc-pVQZ. Adding additional *p* functions in the small basis did not produce significant changes in results. In the end, the cc-pVQZ subset (hereafter called dual-cc-pVQZ) is of a size roughly intermediate between cc-pVDZ and cc-pVTZ for most molecules. Both subsets are available in the supplementary materials.

Previous density functional theory (DFT) tests³⁶ on a small test set also indicate that 6-311G* serves as a sufficient subset for 6-311+G(3*df*, 3*pd*),^{68–71} the largest Pople-style basis set currently available. Though the 6-311G*/6-311+G(3*df*, 3*pd*) pairing continues to perform well for B3LYP on our larger test sets, this truncation is too drastic for an MP2 calculation. This observation again demonstrates the faster convergence of SCF calculations with respect to basis set size, relative to correlation calculations. We choose to present results for both 6-311+G** and 6-311+G(3*d*, *p*) subsets. The latter pairing is the Pople-style analog of the conservative cc-pVTZ truncation described previously and, as will be shown below, is the subset of choice for near reproduction of 6-311+G(3*df*, 3*pd*) properties. However, the 6-311+G** subset still makes a marked improvement upon the gap in single-basis properties, though the reproduction of 6-311+G(3*df*, 3*pd*) properties is not exact. While our goal is to accurately approach the basis set limit at reduced computational cost, this smaller pairing is included to demonstrate the fact that drastic truncations still dramatically improve upon a single-basis calculation in the small basis. This smaller subset may be a suitable compromise, since RI auxiliary basis sets are not currently available for these large Pople-style basis sets. Typical SCF savings for the 6-311+G** subset are roughly a factor of 10, while the larger and more accurate 6-311+G(3*d*, *p*) subset produces SCF savings of roughly a factor of 6.

The auxiliary basis sets used throughout this work for the RI approximation to cc-pVTZ and cc-pVQZ calculations are those of Weigend, Köhn, and Hättig.^{72,73}

TABLE II. Dual-basis MP2 rotational barriers. Results are errors relative to single-basis MP2 barriers (in kcal/mol). For comparison, single-basis MP2 basis set dependence is also shown. The RI approximation is used for Dunning-type basis sets. %=average absolute percent error.

| Method | Basis | msd | mad | rms | % |
|---------------------|------------------------------------|---------|--------|--------|-------|
| DB-SCF/MP2 | 6-311+G**/6-311++G(3df,3pd) | 0.0158 | 0.0370 | 0.0454 | 2.28 |
| DB-SCF/MP2+singlets | | 0.0181 | 0.0397 | 0.0463 | 2.58 |
| DB-SCF/MP2 | 6-311+G(3d,p)/6-311++G(3df,3pd) | 0.0036 | 0.0111 | 0.0174 | 0.50 |
| DB-SCF/MP2+singlets | | 0.0032 | 0.0097 | 0.0146 | 0.41 |
| DB-SCF/MP2 | dual-cc-pVTZ/cc-pVTZ | 0.0040 | 0.0159 | 0.0215 | 0.59 |
| DB-SCF/MP2+singlets | | 0.0032 | 0.0142 | 0.0190 | 0.52 |
| DB-SCF/MP2 | dual-cc-pVQZ/cc-pVQZ | -0.0028 | 0.0108 | 0.0148 | 0.34 |
| DB-SCF/MP2+singlets | | -0.0027 | 0.0102 | 0.0139 | 0.29 |
| DB-SCF/MP2 | T→Q | -0.0042 | 0.0123 | 0.0162 | 0.45 |
| DB-SCF/MP2+singlets | extrapolation | -0.0042 | 0.0116 | 0.0155 | 0.41 |
| MP2 | 6-311+G** vs 6-311++G(3df,3pd) | -0.2856 | 0.4286 | 0.5090 | 19.34 |
| MP2 | 6-311+G(3d,p) vs 6-311++G(3df,3pd) | -0.0709 | 0.1627 | 0.2065 | 7.23 |
| MP2 | cc-pVDZ vs cc-pVTZ | -0.1072 | 0.5828 | 1.0327 | 18.52 |
| MP2 | cc-pVTZ vs cc-pVQZ | 0.1463 | 0.1735 | 0.2101 | 5.94 |

C. Test sets

The methods and basis sets described above were tested on three sets of molecules. Our intent is to demonstrate the broad applicability of DB-SCF/MP2 across a range of molecular systems. In all cases, we have explored cc-pVTZ, cc-pVQZ, and the T→Q extrapolation; and the two Pople-style combinations are included in two of the applications for comparison. Given the immense amount of data tabulated for these test sets, only set-wide statistics are reported. All SCF calculations were converged to at least a rms DIIS⁷⁴ error of 10^{-8} a.u. between successive density matrices, using integral cutoffs of 10^{-12} a.u.

- A series of nine conjugated hydrocarbons⁷⁵⁻⁷⁷ was used to examine 16 relative conformational energies and rotational barriers. They include six vinyl-*G* (*G*=NO₂, CHO, CFO, Vi, OH, NH₂) and three phenyl-*G'* (*G'*=CHO, CFO, C₂H₃) compounds, for which geometries were optimized at MP2/6-31G*. Only relative energies are examined, which range in magnitude from 0.15 to 8.5 kcal/mol. The RI approximation is used for the Dunning-type basis sets.
- A set of 27 alanine tetrapeptide conformations^{78,12} was used as a quantitative and qualitative application of DB-SCF/RI-MP2. The conformations range from linear to globular geometries and have relative energy differences (normalized to the lowest-energy conformation) from 0.5 to 8.0 kcal/mol. The quantitative agreement with full-basis relative energies, as well as the relative ordering of the conformations, was examined for cc-pVTZ, cc-pVQZ, and the T→Q extrapolation. The conformational geometries were optimized at HF/6-31G**,⁷⁸ and coordinates are available in the supplementary materials of Ref. 12. The frozen core approximation is used in these calculations.

- The G3/99 (Refs. 37, 79, and 80) set of 223 open- and closed-shell (neutral) molecules was used to examine absolute energies and atomization energies across a wide range of molecules. The latter comparison provides a realistic upper bound to the errors inherent to the method.

Throughout the tables, msd=mean signed deviation, mad=mean absolute deviation, Max.=maximum absolute deviation, and rms=root mean squared deviation, and a reasonably chosen “acceptance” cutoff is presented in the final column.

III. RESULTS

A. Energetics

1. Rotational barriers

Table II includes results for the set of rotational barriers and conformational energies. Dual-basis results agree to within 0.019 and 0.014 kcal/mol (rms) of their full-basis RI counterparts, respectively, for cc-pVTZ and cc-pVQZ. Despite the use of a more drastic truncation in the small basis, the cc-pVQZ results are actually closer to their full-basis counterpart, most likely due to the fact that the dual-cc-pVQZ is still more complete than the cc-pVTZ subset.

The error in the large Pople-style set (0.015 kcal/mol) is intermediate to the error in cc-pVTZ and cc-pVQZ. The smaller subset (6-311+G**) produces errors that are, in general, two to three times larger, though these errors are still quite tolerable and consist of only hundredths of a kcal/mol. The average absolute percentage error is also dominated by the result for CH₂CH-CFO, which has a *cis-trans* conformational energy difference of only 0.2 kcal/mol. The dual-basis approximation introduces an error of 0.03 kcal/mol in this case.

TABLE III. Dual-basis RI-MP2 tetrapeptide conformational energies. Results are errors relative to single-basis RI-MP2 conformational energies (in kcal/mol).

| Method | Basis | rms | Max. | No. of errors |
|--------------------|-------------|-------|-------|---------------|
| DB-SCF/MP2 | cc-pVTZ | 0.023 | 0.060 | 4 |
| | cc-pVQZ | 0.015 | 0.039 | 3 |
| | T→Q extrap. | 0.014 | 0.039 | 0 |
| DB-SCF/MP2+singles | cc-pVTZ | 0.020 | 0.051 | 4 |
| | cc-pVQZ | 0.014 | 0.033 | 3 |
| | T→Q extrap. | 0.013 | 0.033 | 0 |

In all cases, dual-basis errors are significantly less than the relative errors associated with single-basis calculations in the small basis sets. For example, single-basis calculations in cc-pVTZ produce a rms error of 0.210 kcal/mol, relative to cc-pVQZ single-basis calculations. Dual-basis calculations with dual-cc-pVQZ have an associated error that is at least an order of magnitude less. The Pople-style combinations reduce the rms error from 0.509 to 0.046 kcal/mol (6-311+G^{**}) and from 0.207 to 0.015 kcal/mol [6-311+G(3*d*,*p*)].

The T→Q extrapolation was also applied to these results, using the extrapolation scheme of Halkier *et al.*⁶⁷ for cc-pV(*XY*)Z basis sets, with $X < Y$:

$$E_{XY} = E_Y^{\text{HF}} + \frac{X^3 E_X^{\text{corr}} - Y^3 E_Y^{\text{corr}}}{X^3 - Y^3}. \quad (8)$$

The singles energy, when applicable, is included in E_{HF} . Extrapolated barriers agree with the full-basis barriers to within 0.02 kcal/mol, an average of less than one-half of 1%. This error is comparable in magnitude to the error introduced by the RI approximation, and we can reliably conclude that it is sufficiently small to replace full MP2 calculations with DB-SCF/RI-MP2 calculations. It should be emphasized that, using the DB-SCF/RI-MP2 scheme and these extrapolations,

we are able to obtain results of approximately cc-pV5Z quality, using basis sets between cc-pVDZ and cc-pVTZ sizes.

The data for rotational barriers also illustrate a common trend among our calculations. Since the SCF energy has already been corrected using our dual-basis formalism, the contribution from the perturbation theory singles is quite small (only differences in rms errors of 0.0025 and 0.0009 kcal/mol for cc-pVTZ and cc-pVQZ). This negligible contribution to the energy does not justify the construction of a second Fock matrix, a Fock matrix for which integral screening can no longer be exploited, effectively doubling the cost of the total SCF calculation. Thus, the singles can safely be neglected without significant loss in quality of the results.

2. Alanine tetrapeptides

Another test of relative conformational energies reiterates the numerical robustness of DB-SCF/MP2, as well as the qualitative consistency of the method (see Table III). For the 27 tetrapeptide conformations, described above, the dual-basis approximation (without singles) introduces relative rms errors of 0.023 and 0.015 kcal/mol, respectively, for cc-pVTZ and cc-pVQZ. Addition of the singles provides very little improvement. Worth noting is the fact that these errors are slightly *less* than the errors inherent to some MP2 localization schemes.¹² Relative to a full RI-MP2 calculation, these two basis sets make four and three qualitative errors in the relative ordering of the conformations. It should be emphasized, however, that differences in relative conformational energies can be as small as 0.005 kcal/mol in this test set. These qualitative errors are eliminated at the T→Q extrapolated level.

TABLE IV. Dual-basis MP2 absolute energies on G3/99 set of 223 neutral molecules. The RI approximation is used for Dunning-type basis sets. Results are errors relative to single-basis MP2 energies (in kcal/mol).

| Method | Basis | msd | mad | Max. | rms | No. above 1 mHartree |
|------------------------|---|--------|-------|-------|-------|-------------------------|
| Pseudocanonicalization | 6-311+G ^{**} /6-311++G(3 <i>df</i> ,3 <i>pd</i>) | -0.501 | 0.760 | 4.071 | 0.981 | 114 |
| DB-SCF/MP2 | | 0.401 | 0.427 | 1.837 | 0.543 | 39 |
| DB-SCF/MP2+singles | | 0.147 | 0.306 | 2.239 | 0.419 | 19 |
| Pseudocanonicalization | 6-311+G(3 <i>d</i> , <i>p</i>)/6-311++G(3 <i>df</i> ,3 <i>pd</i>) | -0.248 | 0.272 | 2.194 | 0.415 | 22 |
| DB-SCF/MP2 | | 0.194 | 0.209 | 0.932 | 0.252 | 3 |
| DB-SCF/MP2+singles | | 0.147 | 0.176 | 0.939 | 0.218 | 2 |
| Pseudocanonicalization | dual-cc-pVTZ/cc-pVTZ | -0.152 | 0.249 | 1.906 | 0.366 | 16 |
| DB-SCF/MP2 | | 0.199 | 0.213 | 1.293 | 0.263 | 2 |
| DB-SCF/MP2+singles | | 0.148 | 0.179 | 1.234 | 0.226 | 1 |
| Pseudocanonicalization | dual-cc-pVQZ/cc-pVQZ | -0.104 | 0.227 | 1.405 | 0.319 | 13 |
| DB-SCF/MP2 | | 0.217 | 0.225 | 0.972 | 0.265 | 2 |
| DB-SCF/MP2+singles | | 0.168 | 0.188 | 0.926 | 0.226 | 1 |
| Pseudocanonicalization | T→Q | -0.073 | 0.239 | 2.000 | 0.354 | 15 |
| DB-SCF/MP2 | extrap. | 0.226 | 0.236 | 0.891 | 0.279 | 3 |
| DB-SCF/MP2+singles | | 0.177 | 0.201 | 0.650 | 0.239 | 2 |

TABLE V. Dual-basis MP2 atomization energies on the G3/99 set of 223 neutral molecules. Results are errors relative to single-basis MP2 atomization energies (in kcal/mol). For comparison, single-basis standard MP2 basis set dependence is also shown. The RI approximation is used for Dunning-type basis sets.

| Method | Basis | msd | mad | Max. | rms | No. above 1 kcal/mol |
|------------------------|------------------------------------|---------|--------|---------|--------|----------------------|
| Pseudocanonicalization | 6-311+G**/6-311++G(3df,3pd) | 0.069 | 0.569 | 4.071 | 0.811 | 43 |
| DB-SCF/MP2 | | -0.292 | 0.326 | 1.806 | 0.460 | 15 |
| DB-SCF/MP2+singles | | -0.141 | 0.251 | 1.145 | 0.334 | 6 |
| Pseudocanonicalization | 6-311+G(3d,p)/6-311++G(3df,3pd) | 0.286 | 0.307 | 2.373 | 0.461 | 7 |
| DB-SCF/MP2 | | -0.172 | 0.192 | 0.936 | 0.240 | 0 |
| DB-SCF/MP2+singles | | -0.145 | 0.174 | 0.889 | 0.216 | 0 |
| Pseudocanonicalization | dual-cc-pVTZ/cc-pVTZ | 0.192 | 0.282 | 2.113 | 0.410 | 4 |
| DB-SCF/MP2 | | -0.176 | 0.207 | 1.305 | 0.259 | 1 |
| DB-SCF/MP2+singles | | -0.145 | 0.185 | 1.246 | 0.231 | 1 |
| Pseudocanonicalization | dual-cc-pVQZ/cc-pVQZ | 0.103 | 0.220 | 1.507 | 0.316 | 3 |
| DB-SCF/MP2 | | -0.185 | 0.204 | 0.990 | 0.251 | 0 |
| DB-SCF/MP2+singles | | -0.165 | 0.189 | 0.945 | 0.229 | 0 |
| Pseudocanonicalization | T→Q | 0.010 | 0.235 | 2.031 | 0.350 | 5 |
| DB-SCF/MP2 | extrap. | -0.220 | 0.233 | 0.764 | 0.277 | 0 |
| DB-SCF/MP2+singles | | -0.200 | 0.217 | 0.598 | 0.254 | 0 |
| Standard MP2 | 6-311+G** vs 6-311++G(3df,3pd) | -38.127 | 38.127 | 110.346 | 43.969 | 223 |
| Standard MP2 | 6-311+G(3d,p) vs 6-311++G(3df,3pd) | -25.396 | 25.396 | 71.873 | 29.763 | 221 |
| Standard MP2 | cc-pVTZ vs cc-pVQZ | -7.040 | 7.040 | 24.266 | 7.881 | 221 |

3. G3 set

Absolute energy errors for the G3/99 set are presented in Table IV. Results for dual-cc-pVTZ agree to within 0.263 kcal/mol, and results for dual-cc-pVQZ agree to within 0.265 kcal/mol. Again, the contribution of the singles is quite small, less than 0.040 kcal/mol across the set, confirming our previous assertion that the singles can safely be ignored.

Particularly of note is the fact that our full diagonalization algorithm, with or without singles, shows improvement over previous algorithms that only expand the virtual space (denoted by “Pseudocanonicalization” in the data tables). By retaining the flexibility of the orbitals (and their associated energies) to mix small- and large-basis functions, we obtain better orbitals and energies, at nearly identical cost.

The results are consistently reliable: only two (P_4 and C_6H_5) of the 233 test molecules exhibited errors in absolute energy above 1 mHartree. Additionally, our results are mostly one sided, as shown by the similarity in signed and absolute errors. Though the DB-SCF energy is not strictly variational, we have found very few cases where DB-SCF/MP2 overcorrects (only 17 and 8 of the 233 calculations were overcorrected, for cc-pVTZ and cc-pVQZ, respectively), leaving open the possibility for further systematic improvement in DB-SCF/MP2 energies.

A more strenuous and chemically relevant test of DB-SCF/MP2 is shown in the atomization energies of Table V. Here, errors are larger than the errors in Table II; however, the relative energies also range from 7 to 2600 kcal/mol. In general, DB-SCF/MP2 errors are roughly 0.25 kcal/mol. The dual-cc-pVQZ error of 0.251 kcal/mol is a vast improvement over the error in

single-basis cc-pVTZ calculations (7.881 kcal/mol), despite the fact that the small basis is often smaller than cc-pVTZ.

An additional, more transferable analysis on a per-bond basis is provided in Table VI as well. In this analysis, only atom connectivity is taken into account (e.g., a triple bond is considered one connection). The range of molecules in the G3 set is thus compared on a more size-independent footing and provides a rough measure of the error involved in treating chemical reactions. The 6-311+G**/6-311++G(3df,3pd) combination introduces 0.19 kcal/mol error per bond, whereas the large Pople subset, dual-cc-pVTZ, and dual-cc-pVQZ combinations all introduce only 0.08 kcal/mol error. These small errors remain roughly the same at the T→Q extrapolated level. The DB-SCF/MP2 method is thus well suited for chemical reaction energies, particularly in the context of the error associated with single-basis calculations in the small basis.

Methods that diagonalize the virtual space separately still fare well in this analysis, as rms errors are still on the order of tenths of a kcal/mol. However, this analysis highlights the advantages of a full diagonalization. rms errors per bond are roughly half those of their virtual-only counterparts, at no additional cost.

While the Dunning-style set results shown in Tables IV–VI are for DB-SCF/RI-MP2, the RI approximation introduces a negligible error, which is not dependent on the dual-basis approximation used. In calculating atomization energies, for example, the RI approximation introduced rms errors of 0.061 kcal/mol across the same set for both standard SCF/MP2 and DB-SCF/MP2. The dual-basis error is roughly four times larger than the error intrinsic to RI. This

TABLE VI. Dual-basis MP2 atomization energies *per bond* on the G3/99 set of 223 neutral molecular. Results are errors relative to single-basis MP2 atomization energies per bond (in kcal/mol). RI approximation is used for Dunning-type basis sets.

| Method | Basis | msd | mad | Max. | rms | No. above 0.25 kcal/mol |
|--------------------|------------------------------------|--------|-------|--------|--------|-------------------------|
| Pseudocanonization | 6-311+G**/6-311++G(3df,3pd) | 0.041 | 0.218 | 4.071 | 0.448 | 53 |
| DB-SCF/MP2 | | -0.083 | 0.099 | 1.304 | 0.186 | 16 |
| DB-SCF/MP2+singles | | -0.040 | 0.091 | 1.145 | 0.170 | 18 |
| Pseudocanonization | 6-311+G(3d,p)/6-311++G(3df,3pd) | 0.081 | 0.090 | 0.908 | 0.167 | 22 |
| DB-SCF/MP2 | | -0.040 | 0.047 | 0.445 | 0.070 | 2 |
| DB-SCF/MP2+singles | | -0.035 | 0.044 | 0.433 | 0.065 | 1 |
| Pseudocanonization | dual-cc-pVTZ/cc-pVTZ | 0.055 | 0.092 | 0.895 | 0.171 | 22 |
| DB-SCF/MP2 | | -0.044 | 0.054 | 0.520 | 0.083 | 4 |
| DB-SCF/MP2+singles | | -0.037 | 0.049 | 0.507 | 0.077 | 3 |
| Pseudocanonization | dual-cc-pVQZ/cc-pVQZ | 0.034 | 0.076 | 0.639 | 0.141 | 21 |
| DB-SCF/MP2 | | -0.042 | 0.051 | 0.518 | 0.076 | 2 |
| DB-SCF/MP2+singles | | -0.038 | 0.049 | 0.507 | 0.073 | 2 |
| Pseudocanonization | T→Q | 0.007 | 0.080 | 0.694 | 0.143 | 20 |
| DB-SCF/MP2 | extrap. | -0.051 | 0.059 | 0.526 | 0.085 | 3 |
| DB-SCF/MP2+singles | | -0.048 | 0.057 | 0.514 | 0.083 | 3 |
| Standard MP2 | 6-311+G** vs 6-311++G(3df,3pd) | -8.821 | 8.821 | 47.360 | 11.099 | 223 |
| Standard MP2 | 6-311+G(3d,p) vs 6-311++G(3df,3pd) | -5.696 | 5.969 | 32.566 | 7.452 | 223 |
| Standard MP2 | cc-pVTZ vs cc-pVQZ | -1.810 | 1.810 | 8.089 | 2.227 | 223 |

error remains around 0.26 kcal/mol for atomization energies, most likely well below the error inherent to a single-reference MP2 calculation.

One additional aspect of our DB-SCF/MP2 method (as well as those expanding only the virtual subspace⁵⁶) merits discussion. In general, our approximation underestimates the SCF energy, since we are only taking one (approximate) Roothaan step. Since the SCF is not fully converged, however, orbital energy gaps are presumably smaller, leading to an overestimate of the correlation energy. A fortuitous cancellation of error between the SCF and MP2 energies is the ultimate result and is partially responsible for the small overall errors. For example, the G3 set absolute energies contained an average signed SCF error of 0.44 kcal/mol

for both dual-cc-pVTZ and dual-cc-pVQZ; the corresponding signed correlation energy errors were -0.24 kcal/mol (0.25 kcal/mol unsigned) for dual-cc-pVTZ and -0.22 kcal/mol (0.23 kcal/mol unsigned) for dual-cc-pVQZ. (Note that this does not imply that the SCF energy is “less converged” with respect to basis set, as the SCF energy is two to three orders of magnitude larger than the correlation energy.) Standalone DB-HF energies would, unfortunately, not benefit from this error cancellation. However, this dual-basis error (0.44 kcal/mol) is still a marked improvement over the 9.2 kcal/mol average basis set error for absolute energies in single-basis cc-pVTZ calculations, relative to cc-pVQZ.

TABLE VII. Errors in dual-basis DFT (B3LYP) absolute and atomization energies on the G3/99 set of 223 neutral molecules. Results are errors relative to single-basis DFT energies (in kcal/mol). The 6-311+G** subset is used for 6-311++G(3df,3pd); the dual-cc-pVTZ and dual-cc-pVQZ subsets are used for the Dunning-type sets.

| Basis | msd | mad | Max. | rms | No. above 1 mHartree |
|------------------------------------|--------|-------|-------|-------|----------------------|
| Absolute energy | | | | | |
| 6-311++G(3df,3pd) | 0.058 | 0.183 | 1.868 | 0.285 | 10 |
| cc-pVTZ | 0.049 | 0.067 | 0.446 | 0.089 | 0 |
| cc-pVQZ | 0.026 | 0.032 | 0.114 | 0.041 | 0 |
| Atomization energy | | | | | |
| 6-311++G(3df,3pd) | -0.019 | 0.140 | 2.027 | 0.285 | 12 |
| cc-pVTZ | -0.030 | 0.066 | 0.373 | 0.085 | 0 |
| cc-pVQZ | -0.047 | 0.080 | 0.264 | 0.099 | 0 |
| Atomization energy <i>per bond</i> | | | | | |
| 6-311++G(3df,3pd) | -0.007 | 0.037 | 0.412 | 0.071 | 3 |
| cc-pVTZ | -0.006 | 0.017 | 0.143 | 0.025 | 0 |
| cc-pVQZ | -0.008 | 0.020 | 0.169 | 0.029 | 0 |

TABLE VIII. CPU timings (in minutes) for cc-pVQZ calculations on styrene.

| Method | SCF | DB correction | Singles | MP2 | Total job | % savings |
|---------------|-------|---------------|---------|-------|--------------------|-------------------|
| DFT/B3LYP | 488.6 | ... | ... | ... | 488.6 | |
| DB-DFT/B3LYP | 23.5 | 32.3 | ... | ... | 55.8 | 88.6 |
| Standard MP2 | 507.1 | ... | ... | 314.5 | 821.6 | |
| RI-MP2 | 507.1 | ... | ... | 5.2 | 512.3 | 37.6 |
| DB-SCF/MP2 | 22.0 | 32.1 | 45.6 | 334.4 | 434.1 | 47.2 |
| | | | | | 388.5 ^a | 52.7 ^a |
| DB-SCF/RI-MP2 | 22.0 | 32.1 | 45.6 | 5.2 | 104.9 | 87.2 |
| | | | | | 59.3 ^a | 92.8 ^a |

^aTime excluding the singles.

To put these MP2 results in perspective, a final comparison to dual-basis DFT is provided in Table VII. Absolute and atomization energies for DB-DFT show, in general, a more consistent correspondence with single-basis DFT calculations than do DB-HF/MP2 energies. However, this result is to be expected, as we are no longer trying to capture a strictly correlation-based space with the dual-basis formalism. Additionally, for the now well-known cases where standard DFT methods fail (dispersion interactions, van der Waals complexes, etc.), DB-SCF/RI-MP2 calculations are now approaching the same feasibility regime as DFT calculations for a wide range of molecules (see Table VIII). Table VII also shows that the Dunning-type subset constructions continue to work extremely well for DFT calculations. For the same time savings discussed previously, DB-DFT/dual-cc-pVQZ atomization energies are within 0.1 kcal/mol of single-basis DFT atomization energies and within 0.03 kcal/mol per bond. In general, DB-DFT errors are also significantly reduced relative to the standalone DB-HF absolute energy errors discussed previously.

B. Timings

All calculations discussed in this work were performed on a 2 GHz, single-processor Opteron node with 8 Gbytes of random access memory (RAM) and a 7200 rpm hard drive with up to 1 Tbytes of available disk space (though it should be noted that even the largest calculations described in this section did not require these large amounts of memory or disk space).

To demonstrate the drastic savings introduced, timings for the most stable styrene (phenyl-C₂H₃) structure are shown in Table VIII. The full cc-pVQZ calculation utilizes 680 basis functions, while the dual-cc-pVQZ subset consists of 312 basis functions. Since the SCF calculations took 14

and 15 cycles to converge for dual- and single-basis calculations, respectively, the SCF speedup predicted in Sec. II becomes

$$\frac{t_{\text{full}}}{t_{\text{dual}}} \geq \frac{\left(\frac{680}{312}\right)^4 \times 15}{\left(\frac{680}{312}\right)^4 + 14} = 9.3,$$

or 89.2% savings. (These calculations used the superposition of atomic densities (SAD) guess for the initial density matrix. We have generated these atomic densities for the Dunning basis subsets; these densities are available in the supplementary materials.) Indeed, we do see a tremendous speedup due to the dual-basis approximation, as shown in Table VIII. Total SCF savings is 89.3%, and total job savings using the RI approximation is 92.8%, with only 0.3 kcal/mol error in total energy and 0.003 kcal/mol in the rotational barrier. Note that use of the RI approximation alone—while impressively reducing the correlation time (by 98%)—still only reduces the total job time by 38% due to the imbalance in SCF and correlation computation times. The calculations shown in Table VIII represent a full MP2 time of almost 14 h and a DB-SCF/RI-MP2 time of only 1 h.

The energies in Table IX also exhibit the same error cancellation trend mentioned previously. In the case of styrene, the SCF energy is underestimated by 0.68 kcal/mol, and the MP2 correction is overestimated by 0.31 kcal/mol (the RI approximation adds an additional 0.04 kcal/mol overestimate).

Despite the savings demonstrated in Table VIII, the RI approximation still introduces enough correlation speedup to render the SCF portion of the calculation dominant. Figures 1 and 2 explore the regime in which the DB-HF/RI-MP2 crossover occurs. We should note that these calculations were performed with fairly conservative SCF convergence

TABLE IX. Energies for cc-pVQZ calculations on styrene.

| Method | SCF | DB correction | Singles | MP2 | Total |
|---------------|--------------|---------------|------------|------------|--------------|
| Standard MP2 | -307.704 734 | ... | ... | -1.571 650 | -309.276 384 |
| RI-MP2 | -307.704 734 | ... | ... | -1.571 712 | -309.276 446 |
| DB-SCF/MP2 | -307.690 399 | -0.013 246 | -0.000 099 | -1.572 141 | -309.275 786 |
| DB-SCF/RI-MP2 | -307.690 399 | -0.013 246 | -0.000 099 | -1.572 204 | -309.275 849 |

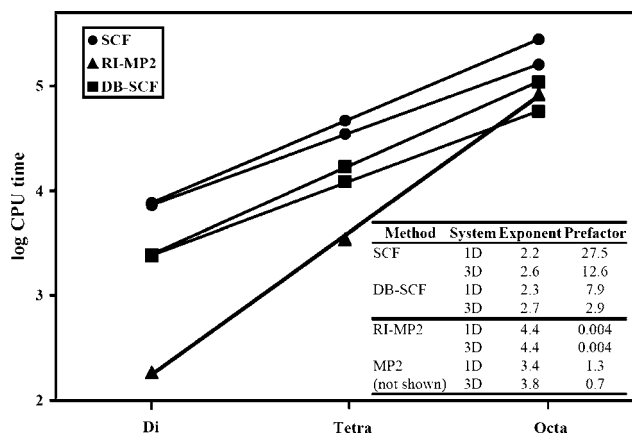


FIG. 1. System size dependence for cc-pVTZ calculations. Shown are both linear (1D) and globular (3D) conformations for di-, tetra-, and octa-alanine peptides. No singles are included in the DB-SCF times, and the frozen core approximation is used for the correlation calculations.

and integral thresholding criteria (10^{-8} and 10^{-12} , respectively) so that any observed errors are solely due to the approximations described herein. They also allow for direct comparison to the results of Ref. 12. Shown are timings for two, four, and eight alanine residues in the polypeptide. Using linear least-squares fits for these data (summarized in the figures), the system size crossovers are established. For standard SCF and RI-MP2, the cc-pVTZ crossover occurs at 56 and 80 heavy atoms for linear [one-dimensional (1D)] and globular [three-dimensional (3D)] systems, respectively. The (albeit conservative) truncation using dual-cc-pVTZ reduces this crossover to 35 and 48 heavy atoms. For cc-pVQZ, the SCF/RI-MP2 crossover occurs at 77 and 106 heavy atoms, well beyond the practical reach of current computing technology for this basis set. The DB-SCF/RI-MP2 crossover now occurs at 28 and 37 heavy atoms. Both standard SCF and DB-SCF methods are noticeably dependent upon system dimensionality; the RI-MP2 correction shows essentially no dimensionality dependence. For reference, reduction of the convergence and threshold parameters to $10^{-6}/10^{-9}$ reduces the DB-SCF/RI-MP2 crossover to 22 (1D) and 24 (3D) heavy atoms, with only an additional 0.006 kcal/mol error in the relative 1D-3D conformational gap.

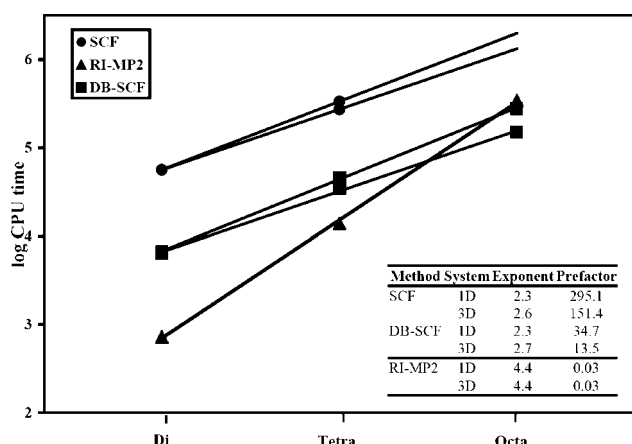


FIG. 2. System size dependence for cc-pVQZ calculations. Shown are both linear (1D) and globular (3D) conformations for di-, tetra-, and octa-alanine peptides.

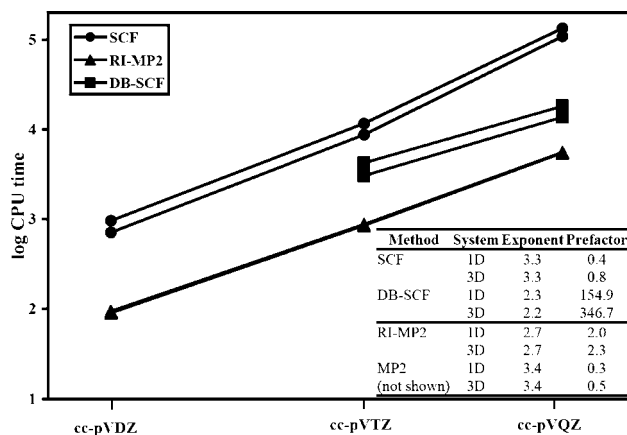


FIG. 3. Basis set dependence for tetrapeptide calculations. Shown are both linear (1D) and globular (3D) conformations for cc-pVDZ, cc-pVTZ, and cc-pVQZ. The DB-SCF calculations are plotted with respect to the target basis set size.

The prefactors and exponents extrapolated from these plots demonstrate two noteworthy aspects of the dual-basis formalism. With respect to system size, DB-SCF calculations still scale identically to their single-basis counterparts. However, since only one step is taken in the large basis, the overall prefactor is reduced by a factor of 3.5–4.4 for cc-pVTZ and by a factor of 8.5–11.2 for cc-pVQZ. Note that *greater* speedup is obtained for the 3D systems. With respect to basis set size, the scaling itself is reduced, due to the $(N/n)^4$ basis set dependence. Standard SCF calculations are shown to scale as $N^{3.3}$, while DB-SCF calculations scale as $N^{2.2}$ with respect to the target basis set size. This reduction is particularly significant since most attempts to reduce the scaling of SCF calculations focus upon system size while leaving the basis size dependence unchanged. The noticeable deviation from linearity in the SCF results of Fig. 3 is most likely due to angular momentum dependence of the evaluation of the Fock matrix integrals.

These reduced scalings and prefactors have subsequently made possible the DB-SCF/RI-MP2 cc-pVQZ calculation of the alanine octapeptides, $C_{24}N_8O_8H_{42}$ (1574 small-basis functions, 3460 large-basis functions, and 7590 auxiliary-basis functions). To our knowledge, this is the largest system to date treated with cc-pVQZ, and the entire calculation took less than 6 days on a single processor. Using the linear fits (as well as standard MP2 fits not shown in the plots, for clarity), we can estimate the cost of using previous methodologies. The timing progression is thus as follows: A standard SCF/MP2 calculation would require 3–4 months, an SCF/RI-MP2 calculation would require 3–4 weeks, and our DB-SCF/RI-MP2 calculations required 6–7 days.

In the large-molecule limit, the RI-MP2 portion of the calculation will continue to dominate; however, since the dual-basis approximation to the SCF calculation is essentially standalone, further modifications to the MP2 portion (local algorithms, etc.) could easily be implemented to extend the range in which savings are significant and in which the HF and post-HF portions of the calculations are competitive.

IV. CONCLUSIONS

The RI approximation to MP2 integrals and the availability of high-quality RI auxiliary basis sets have shifted the time cost of an MP2 calculation to the underlying SCF calculation for many molecules of interest. The combination of RI with our dual-basis method has demonstrated the following characteristics.

- Subsets of 6-311++G(3df,3pd), cc-pVTZ, and cc-pVQZ have been established that produce errors sufficiently small to safely replace MP2 calculations with DB-SCF/RI-MP2 calculations. Standard MP2 energies are accurately reproduced by DB-SCF/RI-MP2 calculations, with absolute energy errors typically on the order of 0.3 kcal/mol and correspondingly small errors in relative conformational energies (0.02 kcal/mol) and atomization energies (0.3 kcal/mol). Errors in chemical reaction energies are roughly 0.08 kcal/mol.
- Total time savings of a DB-SCF/RI-MP2 calculation are 90%–95%, relative to a full (nonlocal) MP2 calculation for the quadruple-zeta basis sets presented. The regime in which the SCF/RI-MP2 crossover occurs has been reduced from 106 to 37 heavy atoms for cc-pVQZ.
- The inclusion of singles introduces a small enough improvement to DB-SCF/MP2 energies that its contribution can be neglected. The cost of adding this contribution roughly doubles the cost of the SCF calculation, which further justifies its exclusion.
- A full diagonalization of the large-basis Fock matrix, followed by the dual-basis correction scheme, produces more accurate and more consistent results than methods which diagonalize the virtual space separately.
- A DB-SCF calculation reduces the computational prefactor with respect to system size, as well as the scaling with respect to basis set size. Molecules previously untouched with large basis sets are now tractable on single processors.
- DB-SCF/RI-MP2 calculations are significantly faster than standard DFT calculations for most currently attainable system sizes. DB-DFT calculations attenuate this timing imbalance; however, DB-SCF/RI-MP2 calculations provide a useful and competitive alternative in cases where DFT might not be applicable.

More aggressive basis set truncation schemes are certainly possible and would lead to even greater computational savings; however, we recommend the above pairings for energy calculations as a conservative compromise between speed and accuracy. The analytic DB-HF gradient has already been implemented in Q-CHEM and has shown promising results for derivative timings and properties. Upcoming papers detailing the interfacing of the DB-HF and RI-MP2 gradients, as well as DB-DFT gradients, will explore the possibility of more aggressive basis set truncation schemes for derivative properties and geometry optimizations.

ACKNOWLEDGMENT

Funding has been provided by the National Science Foundation under Grant No. CHE-0535710. One of the authors (M.H.-G.) is a part owner of Q-CHEM.

APPENDIX: SUPPORTING MATERIAL

The dual-cc-pVTZ and dual-cc-pVQZ basis sets, as well as their associated SAD guesses, are available for download at <http://www.cchem.berkeley.edu/~mhggrp>.

- J. Sancho-García and J. Cornil, *J. Chem. Theory Comput.* **1**, 581 (2005).
- G. S. Fanourgakis, E. Aprà, and S. S. Xantheas, *J. Chem. Phys.* **121**, 2655 (2004).
- S. S. Xantheas, C. J. Burnham, and R. J. Harrison, *J. Chem. Phys.* **116**, 1493 (2002).
- I. Dąbkowska, P. Jurečka, and P. Hobza, *J. Chem. Phys.* **122**, 204322 (2005).
- N. X. Wang and A. K. Wilson, *J. Chem. Phys.* **121**, 7632 (2004).
- T. Helgaker, W. Klopper, H. Koch, and J. Noga, *J. Chem. Phys.* **106**, 9639 (1997).
- T. Helgaker, J. Gauss, P. Jørgensen, and J. Olsen, *J. Chem. Phys.* **106**, 6430 (1997).
- F. Pawłowski, A. Halkier, P. Jørgensen, K. Bak, T. Helgaker, and W. Klopper, *J. Chem. Phys.* **118**, 2539 (2003).
- K. Bak, P. Jørgensen, T. Helgaker, and W. Klopper, *J. Chem. Phys.* **112**, 9229 (2000).
- A. Halkier, W. Klopper, T. Helgaker, and P. Jørgensen, *J. Chem. Phys.* **111**, 4424 (1999).
- T. van Mourik, P. Karamertzanis, and S. Price, *J. Phys. Chem. A* **110**, 8 (2006).
- R. A. DiStasio, Jr., Y. Jung, and M. Head-Gordon, *J. Chem. Theory Comput.* **1**, 862 (2005).
- M. R. Nyden and G. A. Petersson, *J. Chem. Phys.* **75**, 1843 (1981).
- C. Møller and M. Plesset, *Phys. Rev.* **46**, 618 (1934).
- P. Pulay, *Chem. Phys. Lett.* **100**, 151 (1983).
- S. Saebo and P. Pulay, *Chem. Phys. Lett.* **113**, 13 (1985).
- P. Pulay and S. Saebo, *Theor. Chim. Acta* **69**, 357 (1986).
- S. Saebo and P. Pulay, *J. Chem. Phys.* **86**, 914 (1987).
- S. Saebo and P. Pulay, *Annu. Rev. Phys. Chem.* **44**, 213 (1993).
- M. Schütz, G. Hetzer, and H. Werner, *J. Chem. Phys.* **111**, 5691 (1999).
- H. Werner, F. Manby, and P. Knowles, *J. Chem. Phys.* **118**, 8149 (2003).
- P. Ayala and G. E. Scuseria, *J. Chem. Phys.* **110**, 3660 (1999).
- P. Ayala, K. Kudin, and G. E. Scuseria, *J. Chem. Phys.* **115**, 9698 (2001).
- M. Häser, *Theor. Chim. Acta* **87**, 147 (1993).
- J. Almlöf, *Chem. Phys. Lett.* **181**, 319 (1991).
- M. Häser and J. Almlöf, *J. Chem. Phys.* **96**, 489 (1992).
- A. Wilson and J. Almlöf, *Theor. Chim. Acta* **95**, 49 (1997).
- K. Eichkorn, O. Treutler, H. Ohm, M. Heiser, and R. Ahlrichs, *Chem. Phys. Lett.* **240**, 283 (1995).
- F. Weigend, M. Häser, H. Patzelt, and R. Ahlrichs, *Chem. Phys. Lett.* **294**, 143 (1998).
- M. Feyereisen, G. Fitzgerald, and A. Komornicki, *Chem. Phys. Lett.* **208**, 359 (1993).
- D. E. Bernholdt and R. J. Harrison, *Chem. Phys. Lett.* **250**, 477 (1996).
- F. Weigend and M. Häser, *Theor. Chem. Acc.* **97**, 331 (1997).
- Y. Jung, A. Sodt, P. M. W. Gill, and M. Head-Gordon, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6692 (2005).
- R. Lochan, Y. Jung, and M. Head-Gordon, *J. Phys. Chem. A* **109**, 7598 (2005).
- Y. Jung, R. C. Lochan, A. D. Dutoi, and M. Head-Gordon, *J. Chem. Phys.* **121**, 9793 (2004).
- W. Liang and M. Head-Gordon, *J. Phys. Chem. A* **108**, 3206 (2004).
- L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, *J. Chem. Phys.* **94**, 7221 (1991).
- A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993).
- R. H. Hertwig and W. Koch, *Chem. Phys. Lett.* **268**, 345 (1997).
- S. Goedecker, *Rev. Mod. Phys.* **71**, 1085 (1999).
- G. E. Scuseria, *J. Phys. Chem. A* **103**, 4782 (1999).
- W. Liang, C. Saravanan, Y. Shao, R. Baer, A. T. Bell, and M. Head-Gordon, *J. Chem. Phys.* **119**, 4117 (2003).
- M. Challacombe and E. Schwegler, *J. Chem. Phys.* **106**, 5526 (1997).

- ⁴⁴E. Schwegler, M. Challacombe, and M. Head-Gordon, *J. Chem. Phys.* **106**, 9708 (1997).
- ⁴⁵M. Challacombe, *J. Chem. Phys.* **110**, 2332 (1999).
- ⁴⁶O. Vahtras, J. Almlöf, and M. W. Feyereisen, *Chem. Phys. Lett.* **213**, 514 (1993).
- ⁴⁷H. A. Fruchtl, R. A. Kendall, R. J. Harrison, and K. G. Dyall, *Int. J. Quantum Chem.* **64**, 63 (1997).
- ⁴⁸F. Weigend, *Phys. Chem. Chem. Phys.* **4**, 4285 (2002).
- ⁴⁹R. A. Friesner, *Chem. Phys. Lett.* **116**, 39 (1985).
- ⁵⁰R. A. Friesner, *J. Chem. Phys.* **85**, 1462 (1986).
- ⁵¹R. A. Friesner, *J. Chem. Phys.* **86**, 3522 (1987).
- ⁵²M. N. Ringnalda, Y. Won, and R. A. Friesner, *J. Chem. Phys.* **92**, 1163 (1990).
- ⁵³M. N. Ringnalda, M. Belhadj, and R. A. Friesner, *J. Chem. Phys.* **93**, 3397 (1990).
- ⁵⁴R. A. Friesner, *J. Phys. Chem.* **92**, 3091 (1988).
- ⁵⁵R. Jurgens-Lutovsky and J. Almlöf, *Chem. Phys. Lett.* **178**, 451 (1991).
- ⁵⁶K. Wolinsky and P. Pulay, *J. Chem. Phys.* **118**, 9497 (2003).
- ⁵⁷M. S. Lee and M. Head-Gordon, *Comput. Chem. (Oxford)* **24**, 295 (2000).
- ⁵⁸Truncated SCF schemes could, of course, be crafted at any step in the SCF procedure. We have chosen a single step because it (a) provides significant savings in the single Fock matrix build, (b) eliminates the self-consistency problem altogether in the large basis, and (c) captures the vast majority of the remaining basis-set-dependent energy. The SCF could also be carried out to completion; this procedure essentially uses the small-basis density as a new SCF guess and was originally implemented in Q-CHEM by J. Kong.
- ⁵⁹Y. Shao, L. Fusti-Molnar, Y. Jung *et al.*, *Phys. Chem. Chem. Phys.* **8**, 3172 (2006).
- ⁶⁰This fact only holds for small basis sets that are proper subsets of the target basis set, which is why only subsets have been explored in this work.
- ⁶¹T. H. Dunning, Jr., *J. Chem. Phys.* **90**, 1007 (1989).
- ⁶²R. A. Kendall and T. H. Dunning, Jr., *Chem. Phys. Lett.* **96**, 6796 (1992).
- ⁶³D. E. Woon and T. H. Dunning, Jr., *J. Chem. Phys.* **98**, 1358 (1993).
- ⁶⁴D. E. Woon and T. H. Dunning, Jr., *J. Chem. Phys.* **103**, 4572 (1995).
- ⁶⁵D. E. Woon and T. H. Dunning, Jr., *J. Chem. Phys.* **100**, 2975 (1994).
- ⁶⁶A. K. Wilson, T. van Mourik, and T. H. Dunning, Jr., *J. Mol. Struct.: THEOCHEM* **388**, 339 (1996).
- ⁶⁷A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, *Chem. Phys. Lett.* **286**, 243 (1998).
- ⁶⁸R. Krishnan, J. Binkley, R. Seeger, and J. A. Pople, *J. Chem. Phys.* **72**, 650 (1980).
- ⁶⁹T. Clark, J. Chandrasekhar, G. W. Spitznagel, and P. von R. Schleyer, *J. Comput. Chem.* **4**, 294 (1983).
- ⁷⁰P. M. W. Gill, B. G. Johnson, J. A. Pople, and M. J. Frisch, *Chem. Phys. Lett.* **197**, 499 (1992).
- ⁷¹M. J. Frisch, J. A. Pople, and J. S. Binkley, *J. Chem. Phys.* **80**, 3265 (1984).
- ⁷²F. Weigend, A. Köhn, and C. Hättig, *J. Chem. Phys.* **116**, 3175 (2002).
- ⁷³C. Hättig, available for download at <ftp://ftp.chemie.uni-karlsruhe.de/pub/cbasen>.
- ⁷⁴P. Pulay, *Chem. Phys. Lett.* **73**, 393 (1980).
- ⁷⁵M. Head-Gordon and J. A. Pople, *J. Phys. Chem.* **97**, 1147 (1993).
- ⁷⁶R. J. Meier, *J. Phys. Chem.* **97**, 10248 (1993).
- ⁷⁷M. Head-Gordon, *J. Phys. Chem.* **97**, 10250 (1993).
- ⁷⁸M. D. Beachy, D. Chasman, R. B. Murphy, T. A. Halgren, and R. A. Friesner, *J. Am. Chem. Soc.* **119**, 5908 (1997).
- ⁷⁹L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, and J. A. Pople, *J. Chem. Phys.* **109**, 7764 (1998).
- ⁸⁰L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, *J. Chem. Phys.* **112**, 7374 (2000).