

# Hartree–Fock perturbative corrections for total and reaction energies

Jia Deng, Andrew T. B. Gilbert, and Peter M. W. Gill<sup>a)</sup>  
*Research School of Chemistry, The Australian National University,  
 Australian Capital Territory 0200, Australia*

(Received 11 May 2010; accepted 23 June 2010; published online 30 July 2010)

We have performed an assessment of the Hartree–Fock perturbative correction (HFPC) on a large and diverse set of molecules and reactions. Errors in both absolute and reaction energies with respect to converged secondary basis Hartree–Fock results are reported for a wide spectrum of primary/secondary basis set combinations. These results show that using an adequate primary basis, HFPC can accurately reproduce secondary basis energies at a substantially reduced cost. Comparisons of HFPC with the related dual basis Hartree–Fock (DBHF) scheme are also made for several molecules and target secondary basis sets. Our results indicate that HFPC is faster and more accurate than DBHF for approaching triple- $\zeta$  basis sets. For quadruple- $\zeta$  secondary basis sets, HFPC is capable of yielding more accurate energies at a marginally increased cost over DBHF. © 2010 American Institute of Physics. [doi:10.1063/1.3463800]

## I. INTRODUCTION

*Ab initio* quantum chemistry provides a systematic pathway toward the exact solution of the Schrödinger equation. There is a hierarchy of methods with various levels of electron-correlation treatment that, when coupled with a large one-electron basis set, are capable of determining very accurate molecular properties.

Our interest here, however, is in the Hartree–Fock (HF) theory, the usual starting point for more sophisticated correlated methods. The HF method scales formally as  $O(n^4)$ , where  $n$  is the number of basis functions, although this can be reduced to  $O(n^2)$  using cutoffs. Despite this, and the fact that HF calculations converge more quickly than correlated calculations with respect to the size of the basis, obtaining accurate HF energies for larger molecules can be very expensive. Indeed, large-basis HF calculations form a significant part of the cost of the high-level G4 method.<sup>1</sup>

It is well-known that the energy and occupied molecular orbitals from a small-basis HF calculation are often good approximations to those of a large basis counterpart, and that the associated computational time is significantly less. This has encouraged the development of various dual-basis schemes by Hirao and Huzinaga,<sup>2,3</sup> King and co-workers,<sup>4,5</sup> Jurgens-Lutovsky and Almlöf,<sup>6</sup> Wolinski and Pulay,<sup>7</sup> and Head-Gordon and co-workers.<sup>8–10</sup>

Recently, we proposed three Hartree–Fock perturbative corrections<sup>11</sup> (HFPCs) that improve energies of a primary-basis HF calculation toward those of a larger secondary basis. Preliminary results for small systems showed that significant improvements can be achieved even if the primary basis is of only modest quality. Furthermore, our best method, which removes a previously overlooked self-interaction error, achieves a quadratic reduction in the energy error.

To validate a theoretical model, testing it on a large set containing diverse molecular systems is essential. With this goal in mind, we have implemented HFPC within the Q-CHEM package<sup>12</sup> and have calculated energies of the neutral molecules in the G3 set,<sup>13</sup> and a selection of organic reactions. Because our schemes are related to the dual-basis HF (DBHF) techniques,<sup>8–10</sup> a comparison of these two methods in terms of accuracy and efficiency is also presented.

## II. HARTREE–FOCK PERTURBATIVE CORRECTIONS

There are three HFPC variants, differing in the way that they obtain the improved orbitals. The simplest and cheapest approach, HFPC[1], is outlined below and henceforth denoted as HFPC. We refer the interested reader to our earlier communication<sup>11</sup> for technical details of the HFPC[2] and HFPC[3] variants.

A HFPC calculation consists of an iterative HF calculation in a small primary basis followed by a single Fock matrix formation, diagonalization, and energy evaluation in a larger secondary basis. We denote a conventional HF calculation by HF/basis, and a HFPC calculation by HFPC/primary/secondary. Using a primary basis of  $n$  functions, the restricted HF matrix elements for a  $2m$ -electron system are<sup>14</sup>

$$F_{\mu\nu} = h_{\mu\nu} + \sum_{\lambda\sigma}^n P_{\lambda\sigma} \left[ (\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\lambda|\nu\sigma) \right]. \quad (2.1)$$

Solving the Roothaan–Hall equation in the primary basis results in molecular orbitals and an associated density matrix,  $\mathbf{P}$ . In a HFPC calculation,  $\mathbf{P}$  is subsequently used to build a new Fock matrix,  $\mathbf{F}^{[1]}$ , in a larger secondary basis of  $N$  functions,

<sup>a)</sup>Electronic mail: peter.gill@anu.edu.au.

TABLE I. MAEs and maximum absolute errors (MAX) of HF and HFPC for the neutral molecules in the G3 set (kcal/mol).

Basis		Nonhypervalent				Hypervalent			
		$\Delta E^{\text{HF}}$		$\Delta E^{\text{HFPC}}$		$\Delta E^{\text{HF}}$		$\Delta E^{\text{HFPC}}$	
Primary	Secondary	MAE	MAX	MAE	MAX	MAE	MAX	MAE	MAX
6-31G	6-31G( <i>d</i> )	50	211	2.1	19	185	280	17	27
6-31G( <i>d</i> )	6-311+G(3 <i>df</i> ,2 <i>p</i> )	47	161	0.5	1.5	126	195	1.5	2.6
6-311G( <i>d</i> )	6-311+G(3 <i>df</i> ,2 <i>p</i> )	18	57	0.4	1.9	59	99	2.3	3.6
pc-0	cc-pVDZ	368	1415	4.7	31	886	1534	22	37
cc-pVDZ	cc-pVTZ	35	145	0.2	1.4	100	182	1.2	2.1
cc-pVTZ	cc-pVTZ	4.6	15	0.05	0.2	16	30	0.2	0.4
cc-pVDZ	cc-pVQZ	43	178	0.4	2.2	123	224	1.9	3.1
cc-pVTZ	cc-pVQZ	8.4	33	0.02	0.1	23	42	0.07	0.1
cc-pVQZ	cc-pVQZ	4.9	15	0.04	0.2	16	25	0.2	0.3
pc-0	pc-1	330	1269	4.6	27	819	1397	22	35
pc-1	pc-2	72	301	0.3	1.3	167	337	1.3	2.4
pc-1	pc-3	81	331	0.5	2.0	192	375	2.3	3.5
pc-2	pc-3	9.0	30	0.05	0.3	24	38	0.2	0.3

$$F_{ab}^{[1]} = h_{ab} + \sum_{\lambda\sigma}^n P_{\lambda\sigma} \left[ (ab|\lambda\sigma) - \frac{1}{2}(a\lambda|b\sigma) \right], \quad (2.2)$$

where  $\lambda$ ,  $\sigma$  indicate primary basis functions and  $a$ ,  $b$  represent secondary basis functions. Diagonalization of  $\mathbf{F}^{[1]}$  yields improved molecular orbitals and an associated density matrix  $\mathbf{P}^{[1]}$ . The HFPC energy is given by

$$E^{\text{HFPC}} = \sum_{ab}^N P_{ab}^{[1]} h_{ab} + \frac{1}{2} \sum_{abcd}^N P_{ab}^{[1]} P_{cd}^{[1]} [2(ab|cd) - (ac|bd)], \quad (2.3)$$

where  $a$ ,  $b$ ,  $c$ , and  $d$  represent secondary basis functions. This differs from the DBHF energy evaluation where  $\mathbf{P}\mathbf{P}^{[1]}$ , rather than  $\mathbf{P}^{[1]}\mathbf{P}^{[1]}$ , is used. The inclusion of contributions that are quadratic in  $\mathbf{P}^{[1]}$  is the key reason for the fact that HFPC is more accurate than DBHF.

A weakness of this approach is that because they are not self-consistent in the secondary basis, the primary molecular orbitals have a potential that is not self-interaction free. This results in orbitals and an energy which are also corrupted by self-interaction errors. Our more sophisticated schemes, HFPC[2] and HFPC[3], address this shortcoming by removing the offending terms in Eq. (2.2) and forming a different Fock operator for each orbital. These latter methods are more accurate but incur the additional cost of multiple Fock formations and will not be considered further in this paper.

Previous tests of HFPC on small systems demonstrated that near-complete-basis HF energies can be achieved, if the primary calculation is reasonably accurate and the secondary basis is sufficiently large. However, for routine calculations on large systems, it is important to choose basis set pairings that exploit the method's full potential while still remaining computationally attractive. Our target, therefore, is to identify a small established primary basis that allows satisfactory reproduction ( $\leq 0.1$  kcal/mol per atom deviation) of the HF energy in a chosen secondary basis.

### III. RESULTS AND DISCUSSION

Primary and secondary basis set pairings were formed from popular Pople, Dunning, and Jensen basis sets, and the HFPC energies for the 257 neutral molecules in the G3 data set,<sup>13</sup> and 44 selected reactions<sup>15,16</sup> were calculated. The geometries for the structures in the reaction set were taken from literature,<sup>15,16</sup> and those for the G3 neutral molecules were optimized at the B3LYP/6-31G(2*df*,*p*) level using the SG-1 quadrature grid.<sup>17</sup> Incremental Fock formation<sup>18</sup> was switched off and all self-consistent field (SCF) calculations were converged to a DIIS (direct inversion in the iterative subspace) error of  $10^{-7}$ . Our target was the HF/secondary energy and we use the errors

$$\Delta E^{\text{HF}} = E(\text{HF/primary}) - E(\text{HF/secondary}), \quad (3.1)$$

$$\Delta E^{\text{HFPC}} = E(\text{HFPC/primary/secondary}) - E(\text{HF/secondary}), \quad (3.2)$$

$$\Delta E^{\text{DBHF}} = E(\text{DBHF/primary/secondary}) - E(\text{HF/secondary}), \quad (3.3)$$

to gauge the numerical accuracy of the HFPC and DBHF schemes. We note that HFPC energies, unlike DBHF ones, are variational and therefore all  $\Delta E^{\text{HFPC}}$  values are positive. Pure *d*, *f*, and *g* functions were used throughout.

#### A. HFPC total energies

Total energies of the 257 neutral molecules in the G3 data set were computed using HF/primary and HFPC/primary/secondary levels of theory and the mean absolute errors (MAEs) and maximum errors (MAX) with respect to the HF/secondary level are listed in Table I. The errors for hypervalent molecules are consistently larger and for this reason they are listed separately from the nonhypervalent molecules.

For the basis set pairings considered, the MAE of the  $\Delta E^{\text{HF}}$  values for nonhypervalent molecules range from a few

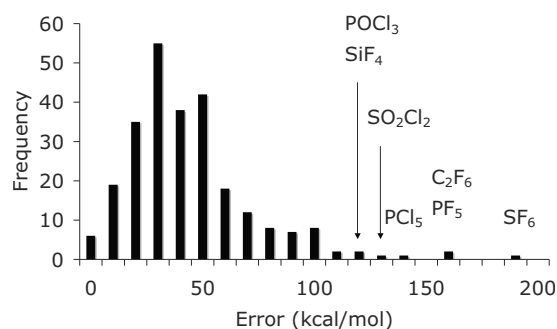


FIG. 1. Distribution of  $\Delta E^{\text{HF}}$  values for the neutral molecules in the G3 set. Primary is 6-31G(*d*) and secondary is 6-311+G(3*df*,2*p*).

kcal/mol to several hundreds of kcal/mol, and the corresponding values for hypervalent molecules are several times larger. These values are a measure of the differences between the primary and secondary bases, and the larger they are, the more difficult it is for HFPC to recover the secondary energies. Overall, for nonhypervalent molecules, HFPC is able to obtain energies that are approximately two orders of magnitude more accurate than the primary HF calculations on which they are based, whereas the corresponding improvements for hypervalent molecules are not as large.

The relative improvements from HFPC are poorest for the subset/superset combinations 6-31G/6-31G(*d*), 6-311G(*d*)/6-311+G(3*df*,2*p*), and *rcc*-pVTZ/*cc*-pVTZ and in these cases the MAEs are reduced by roughly 1.5 orders of magnitude. This demonstrates the importance of having a balanced primary-secondary basis pairing to extract maximum benefit from the HFPC method. The greatest relative improvements are seen for HFPC/*cc*-pVTZ/*cc*-pVQZ where the errors drop by over 2.5 orders of magnitude.

The 6-31G and *pc*-0 basis sets lack polarization functions and HF calculations with these primary bases yield poor potentials. As a result, the HFPC/6-31G/6-31G(*d*), HFPC/*pc*-0/*cc*-pVDZ, and HFPC/*pc*-0/*pc*-1 calculations do not satisfactorily reproduce the secondary basis energies and the  $\Delta E^{\text{HFPC}}$  values are an order of magnitude larger than any of the other basis set combinations. Polarization functions are particularly important for hypervalent molecules such as SF<sub>6</sub>, PF<sub>5</sub>, and PCI<sub>5</sub>, and we therefore recommend the use of primary basis sets with at least one set of *d* functions for such systems.

The error distributions shown in Figs. 1 and 2 illustrate

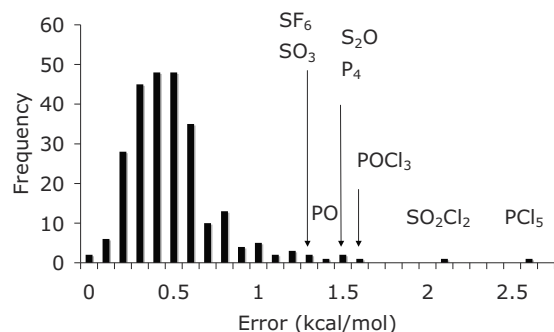


FIG. 2. Distribution of  $\Delta E^{\text{HFPC}}$  values for the neutral molecules in the G3 set. Primary is 6-31G(*d*) and secondary is 6-311+G(3*df*,2*p*).

the numerical robustness of the HFPC scheme. Taking 6-311+G(3*df*,2*p*) as the target secondary basis, the errors in the HF/6-31G(*d*) calculations span a broad range of values, mostly between 20 and 90 kcal/mol. In contrast, the error distribution in the HFPC calculations is significantly narrower, and concentrated between 0.3 and 0.9 kcal/mol. Most of the worst performing molecules are hypervalent.

## B. HFPC reaction energies

Accurate total energies are sufficient but not necessary to ensure the accuracy of HFPC for reaction energies because energy contributions from the core regions, which constitute a significant portion of the total energy, largely cancel when considering energy differences in chemical reactions. In this section, we test the accuracy of HFPC for reaction energies by applying it to the isomerization and Diels–Alder reactions shown in Figs. 3 and 4. The 6-311+G(3*df*,2*p*) basis was selected as the target and benchmark HF reaction energies were calculated using this basis. Six primary basis sets were considered and the MAE and MAX errors with respect to the benchmark data are listed in Table II.<sup>19</sup>

The changes in electronic structure in these reactions are relatively simple and comparisons of total and reaction energy errors demonstrate strong error cancellation. The STO-3G basis is clearly inadequate as a primary basis, with 39 of the 44 reactions having errors in excess of 1 kcal/mol. Although HFPC is able to improve the errors by an order of magnitude, the number of outliers is still unacceptably large. The 3-21G primary basis significantly improves upon STO-3G and leads to much better  $E^{\text{HFPC}}$  values. This is especially true for the reaction energies which have a MAE of only 0.4 kcal/mol. In terms of accuracy, there is little to choose between the 6-31G, 6-311G, 6-31G(*d*), and 6-311G(*d*) basis sets. The total  $\Delta E^{\text{HF}}$  values are better for the 6-311G(*d*) primary but, in all cases, HFPC is able to reduce the MAEs for the reactions to within 0.2 kcal/mol and remove all outliers. None of the reactions considered involve hypervalent species and so polarization functions in the primary basis are unnecessary.

## C. Comparison with the dual-basis method

The DBHF method is closely related to HFPC, and it is worth comparing the two methods in terms of both cost and accuracy. To this end, we calculated the reactions energies in Figs. 3 and 4 for the 6-311G(*d*)/6-311+G(3*df*,2*p*) pairing. The MAE and MAX for DBHF are 0.3 and 1.6 kcal/mol, whereas the corresponding HFPC errors (the last row of Table II) are three times smaller.

A detailed total energy comparison was performed for several molecules: PCI<sub>5</sub>,  $\beta$ -carotene (C<sub>40</sub>H<sub>56</sub>), porphine (C<sub>20</sub>H<sub>14</sub>N<sub>4</sub>), alanine octapeptide (C<sub>24</sub>H<sub>42</sub>N<sub>8</sub>O<sub>8</sub>), vitamin B2 (C<sub>17</sub>H<sub>20</sub>N<sub>4</sub>O<sub>6</sub>), and morphine (C<sub>17</sub>H<sub>19</sub>NO<sub>3</sub>). The structures of porphine, vitamin B2, and morphine were optimized using HF/6-31G(*d*). The structure of  $\beta$ -carotene was obtained from the Cambridge Structure Database<sup>20</sup> and the structure of alanine octapeptide was obtained from the literature.<sup>21</sup> We were unable to converge the HF/*cc*-pVQZ calculation for porphine in a reasonable (<50) number of SCF cycles and

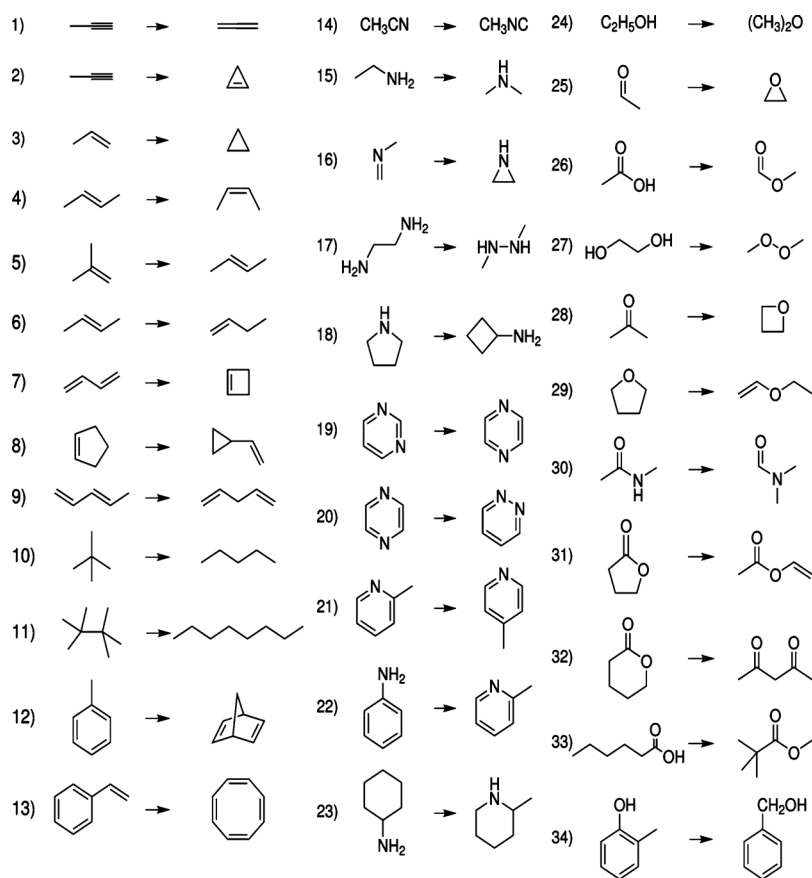


FIG. 3. Isomerization reactions included in the test set.

for this reason we removed the most diffuse *s* functions from each atomic basis. All timing calculations were carried out on a 3.0 GHz single processor with sufficient memory and disk space. The superposition of atomic densities initial guess,  $10^{-5}$  convergence threshold, and DIIS SCF algorithm were used except for the modified cc-pVQZ and rcc-pVQZ calculations, where the core Hamiltonian guess was used.

For the cc-pVTZ secondary basis, 6-31G(*d*) and rcc-pVTZ were used as the primary basis sets for HFPC and DBHF, respectively, as these were found to yield comparable errors. The HF/cc-pVTZ secondary calculations took between 7 and 11 SCF cycles. Table III shows that the HFPC times vary significantly with respect to the primary basis

size: the 6-31G(*d*) calculations take 26%–37% of the time of the full secondary calculation and introduce errors of 1–2 kcal/mol. The larger rcc-pVTZ primary basis causes the calculations to take much longer (73%–90% of the secondary time) but introduce errors of at most 0.4 kcal/mol. Comparing this performance to DBHF we see that except for the  $\text{PCl}_5$  molecule, HFPC is always more accurate and, if the 6-31G(*d*) primary basis is used, cheaper as well. As previously mentioned, hypervalent systems require more polarized primary basis sets and we see the  $\text{PCl}_5$  molecule benefits significantly from an additional set of *d* functions in the primary basis: The  $\Delta E^{\text{HFPC}}$  value is reduced from 2.3 kcal/mol for 6-31G(*d*) to only 0.3 kcal/mol for 6-31G(2*d*).

The HF/cc-pVQZ secondary calculations took between 7 and 14 SCF cycles. Table III shows that HFPC coupled with the cc-pVDZ primary basis usually surpasses the accuracy of DBHF using the rcc-pVQZ primary basis. Again  $\text{PCl}_5$  is problematic and requires an additional set of *d* functions. For the cc-pVQZ secondary basis, HFPC is not always as cost-effective as DBHF, and the reason for this is that the HFPC energy calculation is formally more expensive. Although both energies can be evaluated at  $O(N^2)$  cost after integral cutoffs, such tactics are less effective for dense basis sets and the formal  $O(N^4)$  cost of HFPC is greater than the formal  $O(N^3)$  cost of DBHF.

SCF convergence was slowest for the porphine molecule, which took 14 cycles for the HF/cc-pVQZ calculation. SCF calculations using large basis sets are often slow to converge and dual basis methods—where the SCF iterations

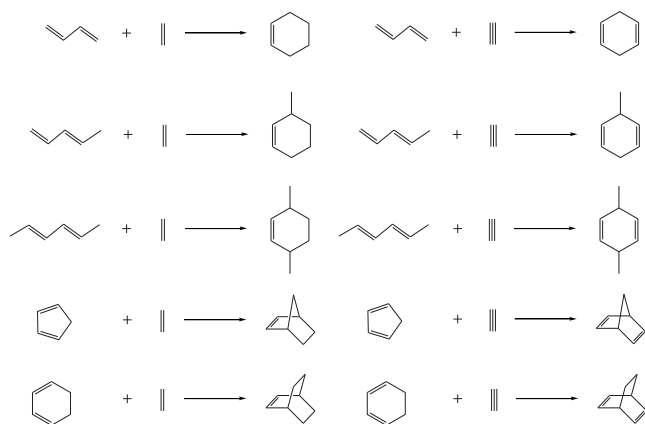


FIG. 4. Diels–Alder reactions included in the test set.



TABLE II. MAEs (kcal/mol), maximum absolute errors (MAX) (kcal/mol), and the number of outliers (NO) ( $>1$  kcal/mol) for HF and HFPC total and reaction energies (the former shown in parentheses). The reaction set contains 44 reactions that are derived from 80 molecules. The secondary basis was 6-311+G(3df,2p).

Primary basis	$\Delta E^{\text{HF}}$			$\Delta E^{\text{HFPC}}$		
	MAE	MAX	NO	MAE	MAX	NO
STO-3G	24 (1765)	71 (3113)	39 (80)	2.6 (12)	8.4 (24)	27 (80)
3-21G	5.2 (827)	23 (1427)	37 (80)	0.4 (4.7)	2.0 (9.0)	2 (80)
6-31G	2.1 (107)	9.4 (197)	24 (80)	0.2 (2.9)	0.8 (5.6)	0 (79)
6-311G	2.1 (79)	10 (142)	26 (80)	0.2 (2.6)	0.7 (5.3)	0 (78)
6-31G( <i>d</i> )	2.9 (49)	9.0 (89)	30 (80)	0.2 (0.6)	0.5 (1.0)	0 (0)
6-311G( <i>d</i> )	1.9 (21)	7.8 (36)	26 (80)	0.1 (0.4)	0.5 (0.9)	0 (0)

occur only in the small primary basis—can become very competitive. This advantage is reflected in the fact that the HFPC/cc-pVDZ/cc-pVQZ calculation took only 10% of the time of the HF/cc-pVQZ calculation.

While HFPC is both accurate and efficient for approximating large basis HF energies, its application to small basis HF energies is less promising. For the  $\beta$ -carotene molecule, HFPC/6-31G/6-31G(*d*) yields an error of 0.1 kcal/mol per atom, but achieves only a 25% reduction in time over the

HF/6-31G(*d*) calculation. This modest saving in time is due to the relatively small difference between the sizes of the primary and secondary basis sets. If smaller primary basis sets, such as STO-3G, are used, the cost savings become more attractive, but the errors become much larger.

#### IV. DEFAULT PRIMARY BASIS

Unlike DBHF, HFPC does not require proper subset/superset basis set combinations and is therefore able to jump

TABLE III. Errors (kcal/mol) and timing ratios for HF, DBHF, and HFPC calculations, error per atom values shown in parentheses. Bold numbers indicate values for our recommended basis set combinations.

Molecule	Primary	Secondary	$\Delta E^{\text{HF}}$		$T^{\text{P}}/T^{\text{S}}$	$\Delta E^{\text{DBHF}}$		$T^{\text{DBHF}}/T^{\text{S}}$	$\Delta E^{\text{HFPC}}$		$T^{\text{HFPC}}/T^{\text{S}}$
PCl <sub>5</sub>	6-31G( <i>d</i> )	cc-pVTZ	170	(28)	0.03	...	...	...	2.3	(0.4)	0.37
	6-31G(2 <i>d</i> )	cc-pVTZ	142	(24)	0.08	...	...	...	<b>0.3</b>	<b>(0.04)</b>	<b>0.44</b>
	rcc-pVTZ	cc-pVTZ	15	(2.5)	0.54	1.3	(0.2)	0.71	0.2	(0.04)	0.85
Morphine	6-31G( <i>d</i> )	cc-pVTZ	204	(5.1)	0.03	...	...	...	<b>1.2</b>	<b>(0.03)</b>	<b>0.35</b>
	rcc-pVTZ	cc-pVTZ	24	(0.6)	0.31	1.9	(0.05)	0.39	0.2	(0.005)	0.90
Porphine	6-31G( <i>d</i> )	cc-pVTZ	202	(5.3)	0.02	...	...	...	<b>1.1</b>	<b>(0.03)</b>	<b>0.26</b>
	rcc-pVTZ	cc-pVTZ	32	(0.9)	0.40	2.5	(0.07)	0.57	0.3	(0.007)	0.77
Vitamin B2	6-31G( <i>d</i> )	cc-pVTZ	307	(6.5)	0.03	...	...	...	<b>1.8</b>	<b>(0.04)</b>	<b>0.34</b>
	rcc-pVTZ	cc-pVTZ	36	(0.8)	0.33	2.8	(0.06)	0.39	0.3	(0.007)	0.81
Ala <sub>8</sub>	6-31G( <i>d</i> )	cc-pVTZ	445	(5.4)	0.03	...	...	...	<b>2.4</b>	<b>(0.03)</b>	<b>0.26</b>
	rcc-pVTZ	cc-pVTZ	49	(0.6)	0.29	3.8	(0.05)	0.34	0.4	(0.005)	0.73
PCl <sub>5</sub>	cc-pVDZ	cc-pVQZ	91	(15)	0.004	...	...	...	1.7	(0.3)	0.26
	6-31G(2 <i>d</i> )	cc-pVQZ	158	(26)	0.003	...	...	...	<b>0.4</b>	<b>(0.07)</b>	<b>0.27</b>
	rcc-pVQZ	cc-pVQZ	17	(2.8)	0.05	1.6	(0.3)	0.22	0.3	(0.05)	0.41
Morphine	cc-pVDZ	cc-pVQZ	191	(4.8)	0.005	...	...	...	<b>1.5</b>	<b>(0.04)</b>	<b>0.23</b>
	rcc-pVQZ	cc-pVQZ	26	(0.7)	0.05	1.9	(0.05)	0.13	0.2	(0.004)	0.39
Porphine	cc-pVDZ	cc-pVQZ <sup>a</sup>	186	(4.9)	0.003	...	...	...	<b>0.9</b>	<b>(0.02)</b>	<b>0.10</b>
	rcc-pVQZ <sup>a</sup>	cc-pVQZ <sup>a</sup>	32	(0.8)	0.08	2.1	(0.05)	0.15	0.004	(0.000 09)	0.21
Vitamin B2	cc-pVDZ	cc-pVQZ	284	(6.0)	0.006	...	...	...	<b>2.2</b>	<b>(0.05)</b>	<b>0.23</b>
	rcc-pVQZ	cc-pVQZ	39	(0.8)	0.06	2.7	(0.06)	0.12	0.2	(0.003)	0.35
$\beta$ -Carotene	STO-3G	6-31G( <i>d</i> )	11 714	(122)	0.05	...	...	...	53	(0.6)	0.34
	6-4G	6-31G( <i>d</i> )	750	(7.8)	0.16	145	(1.5)	0.29	37	(0.4)	0.51
	6-31G	6-31G( <i>d</i> )	358	(3.7)	0.26	53	(0.6)	0.36	9.9	(0.1)	0.75

<sup>a</sup>The most diffuse *s* function on each atom was removed.

between any two basis sets. This flexibility creates a bewildering array of combinations and makes it desirable to have a recommended default primary basis set for a target secondary basis. Such a default must, of course, be a compromise between cost and accuracy. For triple- $\zeta$  quality secondary bases, we recommend 6-31G(*d*) as the primary and, for quadruple- $\zeta$  quality secondary bases, we recommend cc-pVDZ as the primary. We do not advocate HFPC for calculations that target double- $\zeta$  secondary basis sets as the CPU savings are not significant. However, we note that HFPC/STO-3G/6-31G(*d*) is competitive with the recommended DBHF/6-4G/6-31G(*d*) level, and may be useful for rapid preliminary studies of large molecules. Our recommendations do not cover systems such as anions and excited states, whose diffuse nature places additional demands on the basis set.

## V. CONCLUDING REMARKS

We have performed a benchmark study of HFPC on a large and diverse data set of total and reaction energies. For a range of primary/secondary basis set combinations, the HFPC scheme can reduce the error of the primary calculation by around two orders of magnitude at a cost of about one third that of the full secondary calculation.

We have compared the cost of HFPC with the closely related DBHF scheme for several molecules and secondary basis sets and found that for the cc-pVTZ secondary basis, HFPC in combination with 6-31G(*d*) is almost always faster and more accurate than DBHF using the rcc-pVTZ primary basis. For cc-pVQZ basis, HFPC/cc-pVDZ/cc-pVQZ is more accurate but slightly more costly than DBHF/rcc-pVQZ/cc-pVQZ. An advantage of HFPC over DBHF is that it can use any combination of primary/secondary basis sets, which allows errors to be systematically reduced when desired.

HFPC is a cost-effective alternative to large-basis HF calculations and should be useful in future high-level composite methods such as the HEAT (high accuracy extrapolated *ab initio* thermochemistry),<sup>22</sup>  $W_n$ ,<sup>23</sup> and  $G_n^1$  families. It should also be useful in large basis set calculations where the SCF often converges slowly, for HFPC calculations avoid SCF iterations in the secondary basis.

## ACKNOWLEDGMENTS

We thank the Australian Research Council for funding (Grant No. DP0771978) and APAC for a generous allocation

of supercomputer resources. J.D. thanks the ANU/RSC for a Ph.D. scholarship.

- <sup>1</sup>L. A. Curtiss, P. C. Redfern, and K. Raghavachari, *J. Chem. Phys.* **126**, 084108 (2007).
- <sup>2</sup>S. Huzinaga and K. Hirao, *J. Chem. Phys.* **66**, 2157 (1977).
- <sup>3</sup>K. Hirao and S. Huzinaga, *Chem. Phys. Lett.* **45**, 55 (1977).
- <sup>4</sup>S. Havriliak and H. F. King, *J. Am. Chem. Soc.* **105**, 4 (1983).
- <sup>5</sup>S. Havriliak, T. R. Furlani, and H. F. King, *Can. J. Phys.* **62**, 1336 (1984).
- <sup>6</sup>R. Jurgens-Lutovsky and J. Almlöf, *Chem. Phys. Lett.* **178**, 451 (1991).
- <sup>7</sup>K. Wolinski and P. Pulay, *J. Chem. Phys.* **118**, 9497 (2003).
- <sup>8</sup>W. Liang and M. Head-Gordon, *J. Phys. Chem. A* **108**, 3206 (2004).
- <sup>9</sup>R. P. Steele, R. A. DiStasio, Jr., Y. Shao, J. Kong, and M. Head-Gordon, *J. Chem. Phys.* **125**, 074108 (2006).
- <sup>10</sup>R. P. Steele, Y. Shao, R. A. DiStasio, Jr., and M. Head-Gordon, *J. Phys. Chem. A* **110**, 13915 (2006).
- <sup>11</sup>J. Deng, A. T. B. Gilbert, and P. M. W. Gill, *J. Chem. Phys.* **130**, 231101 (2009).
- <sup>12</sup>Y. Shao, L. Fusti-Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio, Jr., R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khallilulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock III, W. Zhang, A. T. Bell, and A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer III, J. Kong, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon *Phys. Chem. Chem. Phys.* **8**, 3172 (2006).
- <sup>13</sup>L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, *J. Chem. Phys.* **112**, 7374 (2000).
- <sup>14</sup>A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry* (McGraw-Hill, New York, 1989).
- <sup>15</sup>S. Grimme, M. Steinmetz, and M. Korth, *J. Org. Chem.* **72**, 2118 (2007).
- <sup>16</sup>S. N. Pieniazek, F. R. Clemente, and K. N. Houk, *Angew. Chem., Int. Ed.* **47**, 7746 (2008).
- <sup>17</sup>P. M. W. Gill, B. G. Johnson, and J. A. Pople, *Chem. Phys. Lett.* **209**, 506 (1993).
- <sup>18</sup>J. Almlöf, K. Faegri, Jr., and K. Korsell, *J. Comput. Chem.* **3**, 385 (1982).
- <sup>19</sup>See supplementary material at <http://dx.doi.org/10.1063/1.3463800> for raw HF and HFPC energies.
- <sup>20</sup>F. H. Allen, *Acta Crystallogr., Sect. B: Struct. Sci.* **58**, 380 (2002).
- <sup>21</sup>R. A. DiStasio, Jr., Y. Jung, and M. Head-Gordon, *J. Chem. Theory Comput.* **1**, 862 (2005).
- <sup>22</sup>A. Tajti, P. G. Szalay, A. G. Csaszar, M. Kallay, J. Gauss, and E. F. Valeev, *J. Chem. Phys.* **121**, 11599 (2004).
- <sup>23</sup>A. Karton, E. Rabinovich, and J. M. L. Martin, *J. Chem. Phys.* **125**, 144108 (2006).