

# Approaching the Basis Set Limit in Density Functional Theory Calculations Using Dual Basis Sets without Diagonalization<sup>†</sup>

WanZhen Liang<sup>‡</sup> and Martin Head-Gordon\*

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

Received: November 13, 2003; In Final Form: January 13, 2004

Dual basis sets are employed as an economical way to approximate self-consistent field (SCF) calculations, such as Kohn–Sham density functional theory (DFT), in large basis sets. First, an SCF calculation is performed in a small subset of the full set of basis functions. The density matrix in this small basis is used to construct the effective Hamiltonian operator in the large basis, from which a correction for basis set extension is obtained for the energy. This correction is equivalent to a single Roothaan step (diagonalization) in the large basis. We present second order nonlinear equations that permit this step to be obtained without explicit diagonalization. Numerical tests on part of the Gaussian-2 dataset, using the B3LYP density functional, show that large-basis results can be accurately approximated with this procedure, subject to some limitations on the smallness of the small basis. Computational savings are approximately an order of magnitude relative to a self-consistent DFT calculation in the large basis.

## Introduction

In ab initio self-consistent field (SCF) calculations, the molecular orbitals are expanded in an atom-centered basis set of Gaussian atomic orbitals.<sup>1</sup> Larger basis sets more closely approach the complete basis set limit and therefore yield higher accuracy. However, this is at the expense of steep increases in computational cost, since the number of significant elements in the effective Hamiltonian grows quadratically with the number of basis functions used per atom. Computational cost also increases between linearly and quartically with the number of atoms. Therefore, for calculations that are not on very small molecules, it is generally necessary to choose a basis set that is relatively small, so that the total computational cost is kept manageable. For example, most calculations on very large molecules reported using linear scaling methods<sup>2–4</sup> use only small basis sets, of double- $\zeta$  plus polarization size or smaller.

Different theoretical models have different requirements for basis set size in order for results to approach those in a complete basis set. Traditional wave function approaches to describe electron correlation are particularly notorious for slow convergence,<sup>5</sup> and extrapolations or semiempirical corrections for basis set incompleteness are essential to obtain reasonably well-converged thermochemical calculations. We shall not consider these methods further here. By contrast, density functional theory (DFT) and Hartree–Fock (HF) SCF calculations are considered to converge much faster with basis set. However, even these SCF methods require quite large basis sets to approach the basis set limit, which is where popular functionals such as Becke's B3LYP are parametrized.<sup>6,7</sup> For organic molecules, a basis of roughly 6-311++G(3df,3pd) size is satisfactory to approach the basis set limit with B3LYP, but this is about two to three times as large as a standard basis of 6-31G+(d) size. Indeed,

Adamson, Gill and Pople<sup>8</sup> showed that their EDF1 density functional, which is specially parametrized for the 6-31+G(d) basis, outperforms B3LYP in that small basis. However the errors are still about twice as large as B3LYP can achieve in the complete basis set limit.

It is therefore important to develop approaches that can yield the higher accuracy associated with DFT calculations in very large basis sets, at reduced computational cost. This paper develops a dual basis formalism for this purpose. The essential idea is to use a small basis set to perform a relatively economical reference calculation, and then correct this result perturbatively in a large basis set to mimic a full calculation in the large basis at much reduced cost. There have been several previous efforts in this direction, beginning with Hartree–Fock (HF) calculations for Rydberg states by King and co-workers,<sup>9,10</sup> and followed by wave function based approaches to correlation in which the HF reference in a small basis is followed by a correlation correction in a large basis.<sup>11,12</sup> A related idea is the use of an adaptive atom-centered minimal basis to span the occupied space,<sup>13</sup> which is obtained as an atom-blocked transformation from a conventional extended basis. The SCF energy obtained in the basis of these polarized atomic orbitals (PAO's) is necessarily higher than a full SCF calculation in the extended basis, as variational freedom is restricted in defining the PAO's. This energy can be corrected by perturbation theory in the same way as the dual basis approaches mentioned above, which substantially improves the quality of both absolute and relative energies.<sup>14</sup>

The second order perturbative corrections used in the dual basis set and PAO SCF calculations are quite simple. Since the reference SCF calculation is not performed in the large basis set, the Brillouin theorem is not satisfied in the large basis. Therefore Brillouin-condition violating matrix elements of the molecular orbital (MO) basis Fock matrix between occupied and virtual orbitals are used as the perturbation. The orbital basis comes from separate canonicalizations of the occupied and virtual spaces. While clearly successful in previous applications,

<sup>†</sup> Part of the special issue "Fritz Schaefer Festschrift".

\* To whom correspondence should be addressed. E-mail: mhg@chem.berkeley.edu.

<sup>‡</sup> Present address: Laboratory of Bond Selective Chemistry, University of Science and Technology of China, Hefei, Anhui 230026, P. R. China.

this approach has a limitation that we seek to overcome in this work. In this paper, we reformulate the expressions for the correction in order to avoid diagonalizations. Our correction expression can be expressed in the atomic orbital basis set (or other localized representations). The projection operators in the AO basis are exponentially localized for insulators,<sup>15</sup> and therefore this formalism may be suitable for a linear scaling implementation in the future. This formalism is laid out in the second section. It leads to an energy correction that is identical to the energy lowering for a single Roothaan step. These equations when linearized give the second-order dual basis expressions used previously.

The third section contains detailed applications of the dual basis approach to the chemical problem for which basis set incompleteness effects are most severe, which is the calculation of atomization energies. The objective is to assess the quality of results that can be obtained with various sizes of small basis for a given large basis. The issue is that underlying the dual basis approach to DFT (or HF) calculations is the assumption that the Coulomb, exchange and correlation potentials due to the occupied orbitals can be adequately represented in the small basis. If so, then the dual basis correction will be relatively small, and a perturbative approach such as the dual basis method can be successful. On the other hand, if the small basis is too small, then important qualitative effects may be missing, and the dual basis results must become more erratic. For example, omission of diffuse functions in the treatment of anions or omission of polarization functions from the small basis may be overly drastic approximations. Of course, it is also important to examine the computational savings that can be obtained, which we do for a couple of examples. Our conclusions are given in the final section.

## Theory

We take the target energy lowering due to the dual basis as being the energy change associated with a single Roothaan step (i.e., diagonalization). We shall assume that the Kohn–Sham Fock matrix,  $\mathbf{F}$ , has been evaluated in the large basis using the Kohn–Sham density matrix,  $\mathbf{P}$ , which was obtained by a self-consistent calculation in the small basis. For simplicity, we assume that the small basis is a subset of the large basis. This assumption is convenient because it means that the small basis density operator can be exactly represented in the large basis. It is not essential, however, since the small basis density operator could, for example, be projected into the large basis. We shall not consider that possibility here. All matrices henceforth are in the large basis.

Operating in the molecular orbital basis for now, we can write the condition for a Roothaan step as the following eigenvalue problem, expressed in terms of the occupied space associated with  $\mathbf{P}$ , and the corresponding virtual space:

$$\begin{bmatrix} \mathbf{F}_{OO} & \mathbf{F}_{OV} \\ \mathbf{F}_{VO} & \mathbf{F}_{VV} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{OO} & \mathbf{U}_{OV} \\ \mathbf{U}_{VO} & \mathbf{U}_{VV} \end{bmatrix} = \begin{bmatrix} \mathbf{U}_{OO} & \mathbf{U}_{OV} \\ \mathbf{U}_{VO} & \mathbf{U}_{VV} \end{bmatrix} \begin{bmatrix} \mathbf{E}_{OO} & \mathbf{0}_{OV} \\ \mathbf{0}_{VO} & \mathbf{E}_{VV} \end{bmatrix} \quad (1)$$

The necessary and sufficient objective is to block-diagonalize the Kohn–Sham Fock matrix between occupied and virtual blocks by finding the unitary transformation  $\mathbf{U}$ . Explicitly writing out the equations for the OO and VO blocks gives us the following equations:

$$\mathbf{F}_{OO}\mathbf{U}_{OO} + \mathbf{F}_{OV}\mathbf{U}_{VO} = \mathbf{U}_{OO}\mathbf{E}_{OO} \quad (2)$$

$$\mathbf{F}_{VO}\mathbf{U}_{OO} + \mathbf{F}_{VV}\mathbf{U}_{VO} = \mathbf{U}_{VO}\mathbf{E}_{OO} \quad (3)$$

The energy lowering associated with the Roothaan step is

$$\delta E = \text{Tr}(\mathbf{E}_{OO} - \mathbf{F}_{OO}) \quad (4)$$

We therefore premultiply eq 2 by  $\mathbf{U}_{OO}^{-1}$  and rearrange inside the matrix trace to obtain

$$\begin{aligned} \text{Tr}(\mathbf{E}_{OO}) &= \text{Tr}(\mathbf{U}_{OO}^{-1}\mathbf{F}_{OO}\mathbf{U}_{OO} + \mathbf{U}_{OO}^{-1}\mathbf{F}_{OV}\mathbf{U}_{VO}) \\ &= \text{Tr}(\mathbf{F}_{OO} + \mathbf{F}_{OV}\mathbf{U}_{VO}\mathbf{U}_{OO}^{-1}) \end{aligned} \quad (5)$$

Defining a new matrix of variables as

$$\mathbf{X}_{VO} = \mathbf{U}_{VO}\mathbf{U}_{OO}^{-1} \quad (6)$$

we then have the following expression for the energy lowering

$$\delta E = \text{Tr}(\mathbf{F}_{OV}\mathbf{X}_{VO}) \quad (7)$$

We now seek equations that determine the new variables,  $\mathbf{X}$ . A suitable set may be found by first post-multiplying eq 3 by  $\mathbf{U}_{OO}^{-1}$

$$\mathbf{F}_{VO} + \mathbf{F}_{VV}\mathbf{X}_{VO} = \mathbf{U}_{VO}\mathbf{E}_{OO}\mathbf{U}_{OO}^{-1} \quad (8)$$

We then take eq 2 and premultiply by  $\mathbf{X}_{VO}$  and postmultiply by  $\mathbf{U}_{OO}^{-1}$  to obtain

$$\mathbf{X}_{VO}\mathbf{F}_{OO} + \mathbf{X}_{VO}\mathbf{F}_{OV}\mathbf{X}_{VO} = \mathbf{U}_{VO}\mathbf{E}_{OO}\mathbf{U}_{OO}^{-1} \quad (9)$$

We can then subtract eq 9 from eq 8 to eliminate the right-hand sides, leaving a set of equations that determine the unknowns,  $\mathbf{X}_{VO}$ , in terms of the Fock matrix elements partitioned between the starting occupied and virtual spaces:

$$\mathbf{F}_{VO} + \mathbf{F}_{VV}\mathbf{X}_{VO} - \mathbf{X}_{VO}\mathbf{F}_{OO} - \mathbf{X}_{VO}\mathbf{F}_{OV}\mathbf{X}_{VO} = \mathbf{0}_{VO} \quad (10)$$

Iterative solution of eq 10 followed by evaluation of the energy lowering from eq 7 is a complete prescription for a dual basis calculation. Although this work is only concerned with the dual basis energy, it is feasible and desirable to also formulate the dual basis analytical gradient with respect to nuclear coordinates (or other perturbations). This is quite feasible, although, because the dual basis energy is not variational, it involves the solution of linear response equations<sup>16</sup> in the small basis. We intend to report on this problem separately, in due course.

Although we have presented eq 10 starting from a molecular orbital description, the matrix of unknowns,  $\mathbf{X}$ , does not have to be expressed in an MO basis. For example, these equations can be equivalently expressed in the atomic orbital basis, by first representing  $\mathbf{F}_{VO}, \mathbf{F}_{VV}, \mathbf{F}_{OO}, \mathbf{F}_{OV}$  in the AO basis by projection. In this representation, the various matrices become sparse for large systems,<sup>15</sup> which offers potential computational advantages. We can write  $\mathbf{F}_{VO}^{\text{AO}} = (\mathbf{1} - \mathbf{S}\mathbf{P})\mathbf{F}\mathbf{P}$ ,  $\mathbf{F}_{OO}^{\text{AO}} = \mathbf{S}\mathbf{P}\mathbf{F}\mathbf{P}$ ,  $\mathbf{F}_{VV}^{\text{AO}} = (\mathbf{1} - \mathbf{S}\mathbf{P})\mathbf{F}(\mathbf{1} - \mathbf{P}\mathbf{S})$ , etc, where  $\mathbf{F}$  and  $\mathbf{S}$  are the AO basis Fock and overlap matrices and  $\mathbf{P}$  is the initial (small basis) density matrix. With these definitions, the equations for  $\mathbf{X}$  in the atomic orbital basis become

$$\mathbf{F}_{VO}^{\text{AO}} + \mathbf{F}_{VV}^{\text{AO}}\mathbf{X}\mathbf{S} - \mathbf{S}\mathbf{X}\mathbf{F}_{OO}^{\text{AO}} - \mathbf{S}\mathbf{X}\mathbf{F}_{OV}^{\text{AO}}\mathbf{X}\mathbf{S} = \mathbf{0} \quad (11)$$

More generally, we can use the tensor properties of these operator matrices<sup>17–19</sup> to conclude that the equations are invariant to any change of basis spanning the occupied and virtual spaces. In terms of covariant and contravariant functions spanning the occupied (indices  $i, j, \dots$ ) and virtual spaces (indices

$a, b, \dots$ ) associated with  $\mathbf{P}$ , we then have the following expression which exhibits all invariances to changes of basis:

$$F_{\bullet i}^a + F_{\bullet b}^a X_{\bullet i}^b - X_{\bullet j}^a F_{\bullet i}^j - X_{\bullet j}^a F_{\bullet b}^j X_{\bullet i}^b = 0 \quad (12)$$

(more details on this notation are available in the literature<sup>19</sup>).

It is also interesting to compare eq 10 with other expressions that have been used for dual basis corrections. Equation 10 is, as already emphasized, equivalent to the energy lowering associated with a single diagonalization, if it is evaluated using the given Fock matrix. Since the dual basis correction is presumed to be relatively small, one approximation that is likely to be effective is to neglect the quadratic terms in eq 10. The resulting linear equation is equivalent to dual basis corrections

$$\mathbf{F}_{VO} + \mathbf{F}_{VV}\mathbf{X}_{VO} - \mathbf{X}_{VO}\mathbf{F}_{OO} = \mathbf{0}_{VO} \quad (13)$$

that have been used before (for single excitations<sup>9,11,14</sup>). Evaluating the energy with the solution to eq 13 gives the energy lowering that would be obtained via second-order perturbation theory. In wave function language, that is the energy lowering associated with single excitations from the reference determinant. These single excitations give nonvanishing energy-lowerings because the Brillouin theorem is not satisfied using the small basis density matrix with the large-basis Fock matrix.

Solution of eq 13 gives the same energy lowering as an approximate Newton step, based on the gradient,<sup>20</sup>  $\mathbf{g}$ , of the energy with respect to orbital rotations coupling occupied levels,  $i$ , with virtual levels,  $a$ , with elements  $(\mathbf{g})_{ai} = \partial E / \partial \theta_{ai} = 2F_{ai}$ . Furthermore, the diagonal orbital second derivatives can be approximated<sup>21</sup> as differences between canonical virtual and occupied Fock matrix elements,  $\partial^2 E / \partial \theta_{ai}^2 \cong 2(\epsilon_a - \epsilon_i)$ , so that the step vector  $\mathbf{d}$  will have elements  $d_{ai} = -\partial E / \partial \theta_{ai} / \partial^2 E / \partial \theta_{ai}^2 = -F_{ai} / (\epsilon_a - \epsilon_i)$ . From the gradient  $\mathbf{g}$  and the chosen step vector  $\mathbf{d}$ , the energy lowering in the canonical basis is finally estimated in the quadratic model as

$$\Delta E = \frac{1}{2} \mathbf{g} \mathbf{d} = \frac{1}{2} \sum_{ai} (2F_{ai}) \left( \frac{-F_{ai}}{\epsilon_a - \epsilon_i} \right) = \sum_{ai} \frac{-F_{ai}^2}{\epsilon_a - \epsilon_i} \quad (14)$$

When the quadratic correction in eq 10 is small, this will be close to what we obtain iteratively. Either approach (either Eqs. (10) or the linearized version, eq 13) can be solved without diagonalization, which may be useful for future large-scale applications of dual basis corrections.

## Results and Discussion

In this section, we examine several aspects of the computational performance of the dual basis approach described here. All calculations reported below are based on the formulation described above, which is the energy lowering predicted for a Roothaan step in the large basis, after an SCF calculation has been completed in the small basis. These calculations use a preliminary implementation of an algorithm to solve the second-order nonlinear equations for  $\mathbf{X}$  in eq 10. A conjugate gradient search is used to solve the equations in the Cholesky orthogonalized basis. The equations are first converged without the nonlinear term. This procedure was implemented into a development version of the Q-Chem program.<sup>22</sup> The SCF convergence criterion for all calculations reported below was  $10^{-8}$  for the RMS difference between successive density matrices. The convergence criterion in the equations for  $\mathbf{X}$  was  $10^{-8}$ .

**CPU Time.** As an indication of the computational cost advantage of using the dual basis approach, we report some

**TABLE 1: Computer Timings for Dual Basis Calculations Using the 6-311G\* Basis as the Primary Basis and 6-311++G(3df,3pd) as the Secondary Basis, against Full Self-Consistent Calculations in the Large Secondary Basis<sup>a</sup>**

molecule	CPU time (min.)		SCF cycles		basis set size	
	full	dual	full	dual	full	6-311G*
CH <sub>3</sub> COCl <sup>b</sup>	18.1	2.2	15	15	218	89
CH <sub>3</sub> COCH <sub>3</sub> <sup>c</sup>	25.6	2.4	12	12	264	90

<sup>a</sup> These timings were recorded on a 375 MHz IBM RS/6000 Power-3 system, using a single processor, and the convergence parameters and tolerances described in the text. <sup>b</sup> The absolute energies at the B3LYP/6-31G(d,p) optimized geometry are  $-613.515502$  au for the small basis,  $-613.540635$  au for the dual basis calculation, and  $-613.541243$  au for the large basis. <sup>c</sup> The absolute energies at the B3LYP/6-31G(d,p) optimized geometry are  $-193.204914$  au for the small basis,  $-193.231674$  au for the dual basis calculation, and  $-193.232607$  au for the large basis.

CPU timings in Table 1 for calculations in which the small basis is 6-311G\* and the large basis is 6-311++G(3df, 3pd). These calculations are on two quite small molecules so that the computational cost is dominated by the formation of the Fock matrix. In the dual basis approach, this is only done once in the large basis, which leads to a theoretical speedup that is the number of SCF cycles (i.e., typically between 10 and 20). The results in Table 1 show speedups in practice of around a factor of 10, reflecting contributions of the small basis SCF and the matrix operations needed to solve the equations for  $\mathbf{X}$ . These speedups clearly show the value of the dual basis approach from the standpoint of efficiency. We now turn to an assessment of accuracy.

**Accuracy of Dual Basis Set Calculations for Atomization Energies.** Among the most challenging properties to calculate accurately by electronic structure methods are atomization energies. This is because there is no possibility of cancellation of error, unlike reactions in which the number of paired electrons is conserved such as isogyric and isodesmic reactions. Systematic testing of electronic structure methods for atomization energies has been greatly aided by the development of the Gaussian-2<sup>23,24</sup> and Gaussian-3<sup>25,26</sup> databases of reliable experimental data for small molecules. In the calculations reported in this section, we compute atomization energies for 56 small molecules to assess the performance of the dual basis approach. The large basis set is taken as 6-311++G(3df,3pd), which is expected to yield DFT results close to the basis set limit. Several smaller subsets of this large basis are tested as the small basis set. All calculations used the B3LYP density functional,<sup>6,7</sup> which remains among the most accurate available for DFT calculations of atomization energies.

The atomization energy results are summarized in Table 2, and the full set of calculations are available as Tables in Supporting Information. Table 2 summarizes the mean absolute derivation (MAD), the root square (RMS) deviation and maximum deviation (MaxD) of atomization energies (AE) versus the database values for calculations using a single basis set. Zero-point energies were included at the B3LYP/6-31G\* level. Table 2 additionally includes a comparison of the performance of single basis calculations versus dual basis calculations for atomization energies. Table 3 then summarizes MAD, RMS, and MaxD for absolute molecular energies (E) using both single and dual basis sets, against the full SCF values in the large 6-311++G(3df,3pd) basis.

First, let us compare the performance of SCF calculations of atomization energies using four different basis sets with the B3LYP functional, as summarized in the single basis columns



**TABLE 2: Comparison of Single Basis B3LYP Calculations and Dual Basis B3LYP Calculations against Experiment for 56 Atomization Energies, as a Function of Basis Set<sup>a</sup>**

small basis	MAD (kcal/mol)		RMS (kcal/mol)		MaxD (kcal/mol)	
	single	dual	single	dual	single	dual
6-311G	24.3	4.0	32.3	5.2	-149.7	14.1
6-311G*	7.0	2.2	9.2	3.0	-36.8	-8.6
6-311G**	4.9	2.2	7.4	3.0	-36.8	-8.6
6-311++G(3df,3pd)	2.2		3.1		-8.6	

<sup>a</sup> In all dual basis calculations, the large basis set is taken as 6-311++G(3df,3pd). Mean absolute deviations (MAD), root-mean-square (RMS), and maximum deviations (MaxD) for the calculations against experiment are reported.

**TABLE 3: Comparison of Absolute Energies from Single and Dual Basis B3LYP Calculations for 56 Small Molecules against the Absolute Energies Calculated at the B3LYP/6-311++G(3df,3pd) Level<sup>a</sup>**

small basis	MAD (kcal/mol)		RMS (kcal/mol)		MaxD (kcal/mol)	
	single	dual	single	dual	single	dual
6-311G	28.0	3.2	35.6	4.8	151.8	-17.5
6-311G*	9.6	0.2	11.4	0.3	36.6	0.7
6-311G**	7.2	0.2	9.3	0.2	36.6	0.7

<sup>a</sup> Mean absolute deviations (MAD), root-mean-square (RMS), and maximum deviations (MaxD) are reported.

of Table 2. It is clear that the largest basis, 6-311++G(3df,3pd) is qualitatively superior to the other three by all measures. For example, the RMS deviation vs experiment in the largest basis is 3.1 kcal/mol, whereas it increases to 7.4, 9.2, and 32.3 kcal/mol in the successively smaller 6-311G\*\*, 6-311G\*, and 6-311G basis sets. This is consistent with the fact that density functionals are developed at the complete basis limit and require large basis sets to perform to their full capability for atomization energies. Although the three smaller basis sets are clearly inadequate for direct calculation of atomization energies, they may still be useful as the small basis reference for a dual basis calculation. This is the question that we address next.

From Table 2, the three dual basis set calculations can be assessed by comparing the MAD, RMS, and MaxD values against those for the explicit SCF calculations in the large basis. It is evident that dual basis calculations starting from either 6-311G\*\* or 6-311G\* reproduce the large basis SCF calculations remarkably well, whereas the 6-311G dual basis results are noticeably poorer. For example, the RMS deviation vs experiment is 3.1 kcal/mol in the large basis and is 3.0, 3.0, and 5.2 kcal/mol in the 6-311G\*\*, 6-311G\*, and 6-311G dual basis calculations. We conclude that 6-311G\* is the smallest basis capable of giving dual basis results that are virtually as good as those from an SCF calculation in the larger basis. It is a very economical approximation that scarcely degrades the quality of the results.

The origin of the poor results for atomization energies obtained with the small basis sets as well as the good results obtained using those same basis sets in a dual basis calculation can be seen from the absolute energy deviations against the large 6-311++G(3df,3pd) basis, as summarized in Table 3. Quantum chemistry calculations in small basis sets yield large RMS errors in absolute energies; those for the 6-311G\*\*, 6-311G\*, and 6-311G basis sets are 9.3, 11.4, and 35.6 kcal/mol versus the large basis. Virtually all of this deviation is then reflected directly in errors in the atomization energies. By contrast, the dual basis approximations in the same three basis sets reduce the RMS absolute energy deviation to 0.2, 0.3, and 4.8 kcal/

mol. In either 6-311G\* or 6-311G\*\* dual basis calculations, this RMS deviation is almost certainly smaller than the intrinsic accuracy of the B3LYP functional, and accounts for the success of these dual basis calculations.

## Conclusions

In this short paper, we have explored the use of a dual basis correction for high accuracy Kohn–Sham density functional theory (DFT) calculations, using the popular B3LYP functional. First, a self-consistent DFT calculation is performed in a small basis, such as 6-311G\*. Second, an energy correction is calculated in a much larger basis set, such as 6-311++G(3df,3pd), based on a single Roothaan step (or diagonalization) using the Fock matrix in the large basis built from the density matrix in the small basis. We have three principal conclusions:

(1) The energy lowering associated with the single Roothaan step in the large basis gives relative energies that are remarkably close to those obtained via fully self-consistent DFT calculations in the large basis. For atomization energies, this yields reductions in the mean absolute error, root-mean-square error, and maximum absolute error that are roughly a factor of 3 for the basis set combination mentioned above.

(2) A reduction of about a factor of 10 in computational cost is achieved even for the small molecules tested here. This basically reflects the fact that only a single Fock matrix must be constructed in the large basis in the dual basis approach, instead of roughly 10 or more in a self-consistent calculation.

(3) We have presented simple equations for the Roothaan step that replace explicit diagonalization with the solution of a second order nonlinear equation. If the second order terms are neglected this expression is equivalent to the energy lowering due to an energy-weighted steepest descent step, or the “single excitation” contributions in second order perturbation theory. We shall shortly report on these equations in more detail elsewhere.

**Acknowledgment.** This work was supported by the National Science Foundation through Grant CHE-9981997. This work was also supported indirectly by funding from Q-Chem Inc via an SBIR subcontract from the National Institutes of Health (R43GM069255-01). M.H.-G. is a part owner of Q-Chem Inc.

**Supporting Information Available:** Tables containing the full set of calculations. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References and Notes

- (1) Davidson, E. R.; Feller, D. *Chem. Rev.* **1986**, *86*, 681–696.
- (2) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085–1123.
- (3) Scuseria, G. E. *J. Phys. Chem. A* **1999**, *103*, 4782–4790.
- (4) Liang, W. Z.; Saravanan, C.; Shao, Y. H.; Baer, R.; Bell, A. T.; Head-Gordon, M. *J. Chem. Phys.* **2003**, *119*, 4117–4125.
- (5) Bak, K. L.; Jorgensen, P.; Olsen, J.; Helgaker, T.; Klopper, W. *J. Chem. Phys.* **2000**, *112*, 9229–9242.
- (6) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (7) Hertwig, R. H.; Koch, W. *Chem. Phys. Lett.* **1997**, *268*, 345–351.
- (8) Adamson, R. D.; Gill, P. M. W.; Pople, J. A. *Chem. Phys. Lett.* **1998**, *284*, 6–11.
- (9) Havriliak, S.; King, H. F. *J. Am. Chem. Soc.* **1983**, *105*, 4–12.
- (10) Havriliak, S.; Furlani, T. R.; King, H. F. *Can. J. Phys.* **1984**, *62*, 1336–1346.
- (11) Jurgens-Lutovsky, R.; Almlöf, J. *Chem. Phys. Lett.* **1991**, *178*, 451–454.
- (12) Wolinski, K.; Pulay, P. *J. Chem. Phys.* **2003**, *118*, 9497–9503.
- (13) Lee, M. S.; Head-Gordon, M. *J. Chem. Phys.* **1997**, *107*, 9085–9095.
- (14) Lee, M. S.; Head-Gordon, M. *Comput. Chem.* **2000**, *24*, 295–301.
- (15) Maslen, P. E.; Ochsenfeld, C.; White, C. A.; Lee, M. S.; Head-Gordon, M. *J. Phys. Chem. A* **1998**, *102*, 2215–2222.

- (16) Handy, N. C.; Schaefer, H. F. *J. Chem. Phys.* **1984**, *81*, 5031–5033.
- (17) Artacho, E.; Delbosch, L. M. *Phys. Rev. A* **1991**, *43*, 5770–5777.
- (18) White, C. A.; Maslen, P.; Lee, M. S.; Head-Gordon, M. *Chem. Phys. Lett.* **1997**, *276*, 133–138.
- (19) Head-Gordon, M.; Maslen, P. E.; White, C. A. *J. Chem. Phys.* **1998**, *108*, 616–625.
- (20) Van Voorhis, T.; Head-Gordon, M. *Mol. Phys.* **2002**, *100*, 1713–1721.
- (21) Head-Gordon, M.; Pople, J. A. *J. Phys. Chem.* **1988**, *92*, 3063–3069.
- (22) Kong, J.; White, C. A.; Krylov, A. I.; Sherrill, D.; Adamson, R. D.; Furlani, T. R.; Lee, M. S.; Lee, A. M.; Gwaltney, S. R.; Adams, T. R.; Ochsenfeld, C.; Gilbert, A. T. B.; Kedziora, G. S.; Rassolov, V. A.; Maurice, D. R.; Nair, N.; Shao, Y. H.; Besley, N. A.; Maslen, P. E.; Dombroski, J. P.; Daschel, H.; Zhang, W. M.; Korambath, P. P.; Baker, J.; Byrd, E. F. C.; Van Voorhis, T.; Oumi, M.; Hirata, S.; Hsu, C. P.; Ishikawa, N.; Florian, J.; Warshel, A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M.; Pople, J. A. *J. Comput. Chem.* **2000**, *21*, 1532–1548.
- (23) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221–7230.
- (24) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063–1079.
- (25) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764–7776.
- (26) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, *112*, 7374–7383.