# Linear scaling computation of the Fock matrix. II. Rigorous bounds on exchange integrals and incremental Fock builds[1]

Eric Schwegler[2,3]
*Department of Chemistry, University of Minnesota*
*207 Pleasant Street S.E., Minneapolis, Minnesota 55455-0431*

Matt Challacombe[3,4] and Martin Head-Gordon[2]
*Department of Chemistry, University of California, Berkeley*
*410 Latimer Hall, Berkeley, California 94720-1460*
(March 6, 1997)

A new linear scaling method for computation of the Cartesian Gaussian-based Hartree-Fock exchange matrix is described, which employs a method numerically equivalent to standard direct SCF, and which does not enforce locality of the density matrix. With a previously described method for computing the Coulomb matrix (J. Chem. Phys. **105**, 1997), linear scaling incremental Fock builds are demonstrated for the first time. Microhartree accuracy and linear scaling are achieved for restricted Hartree-Fock calculations on sequences of water clusters and polyglycine $\alpha$-helices with the 3-21G and 6-31G basis sets. Eight fold speedups are found relative to our previous method. For systems with a small ionization potential, such as graphitic sheets, the method naturally reverts to the expected quadratic behavior. Also, benchmark 3-21G calculations attaining microhartree accuracy are reported for the P53 tetramerization monomer involving 698 atoms and 3836 basis functions.

## I. INTRODUCTION

The applicability of modern quantum chemistry is rapidly expanding due to the success and continued evolution of fast electronic structure methods. These fast methods may be loosely classified as those that involve hierarchical methods and those that do not. Hierarchical methods are able to reduce the computational complexity of building the Coulomb matrix $\mathbf{J}$ from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$, and include continuous versions of the Fast Multipole Method [1–6], the quantum chemical treecode [7], and FFT based KWIK technologies [8]. In the latter class are methods that exploit or enforce the locality of short-range quantum interactions to achieve a reduced computational complexity. In particular, a wide variety of methods have recently been introduced that enforce locality of the density matrix $\mathbf{D}$ [9–18] or the molecular orbital matrix $\mathbf{C}$ [19–27] to achieve linear scaling in various aspects of SCF theory, as for example computation of the Hartree-Fock (HF) exchange matrix $\mathbf{K}$ [17,18]. However, methods that enforce locality have not been able to simultaneously achieve the speed and accuracy associated with hierarchical methods. This is relevant to application of the increasingly useful HF and hybrid HF/DFT theories [28–33], which are currently limited by the computation of $\mathbf{J}$ and $\mathbf{K}$ (and in the latter case, the exchange-correlation matrix).

Recently, we introduced a linear scaling method for computation of the HF Fock matrix $\mathbf{F} = \mathbf{h} + \mathbf{J} - \frac{1}{2}\mathbf{K}$, where $\mathbf{h}$ is the core Hamiltonian,

$$J_{ab} = \sum_{cd} D_{cd} \left(\phi_a \phi_b | \phi_c \phi_d\right), \tag{1}$$

$$K_{ab} = \sum_{cd} D_{cd} \left(\phi_a \phi_c | \phi_b \phi_d\right), \tag{2}$$

and the four-index quantities

$$\left(\phi_a \phi_b | \phi_c \phi_d\right) = \int d\mathbf{r} \int d\mathbf{r}' \; \phi_a^*(\mathbf{r})\phi_b(\mathbf{r}) \left|\mathbf{r}-\mathbf{r}'\right|^{-1} \phi_c^*(\mathbf{r}')\phi_d(\mathbf{r}') \tag{3}$$

are electron repulsion integrals (ERIs) involving the Cartesian Gaussian functions

$$\phi_a(\mathbf{r}) = (x-A_x)^l (y-A_y)^m (z-A_z)^n \exp\left[-\zeta_a (\mathbf{r}-\mathbf{A})^2\right]. \tag{4}$$

To achieve linear scaling, we employed the quantum chemical treecode QCTC [7] to compute $\mathbf{J}$, and enforced the locality of exchange interactions with ONX [17], hearafter referred to as ONX_A. With this method, we found ONX_A to be 3-7 times slower than QCTC. Perhaps a more significant disadvantage of ONX_A is the reliance on system dependent parameters to balance performance with accuracy. Nevertheless, this approach allowed us to carry out HF calculations of unprecedented size, and has provided much of the motivation and insight for the present study.

---

[1] Accepted for publication in the Journal of Chemical Physics
[2] Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720
[3] Minnesota Supercomputer Institute, 1200 Washington Ave. South, Minneapolis, MN 55415
[4] mf10111@msi.umn.edu

Here, a new method for computation of the Hartree-Fock exchange matrix is described, which may be the first algorithm to achieve linear scaling without enforced locality. Also, linear scaling incremental Fock builds are demonstrated for the first time, and useful benchmarks are obtained.

This paper is organized as follows: In the next Section, locality of the density matrix and implications for fast electronic structure methods are discussed. Existing methods for computing the HF exchange matrix are reviewed in Section III and various issues relevant to computation of the exchange matrix are considered in Section IV. In Section V, a new implementation of subroutine ONX is outlined. Then, in Section VI the use of ONX and QCTC in incremental Fock builds is described. Benchmark RHF/3-21G and RHF/6-31G calculations on sequences of water clusters, polyglycine $\alpha$-helices, and single layer graphic sheets are presented in Section VII, along with calculations on the P53 tetramerization monomer. These results are discussed in Section VIII, and in Section IX we present our conclusions.

## II. LOCALITY OF THE DENSITY MATRIX

The long-range behavior of the density matrix,

$$\rho\,(\mathbf{r},\mathbf{r}') = \sum_{cd} D_{cd}\phi_c(\mathbf{r})\phi_d(\mathbf{r}') \tag{5}$$

is a subject of fundamental and practical importance. In Hartree-Fock theory, the density matrix describes the entire physical system [34]. Knowledge of its long-range behavior is essential to achieve linear scaling in a variety of electronic structure theories [7,9–16]. Also, locality of the density matrix is intimately related to orbital localization [19,35,36], and is therefore relevant to local-MO implementations of various SCF theories [19–27].

While the density matrix is dependent on both model chemistry and the physical system, generalizations of its asymptotic decay have been made for infinite systems [37–41]. For insulators that can be represented by localized Wannier functions, the density matrix decays as

$$\rho\,(\mathbf{r},\mathbf{r}') \sim e^{-\sqrt{G}|\mathbf{r}-\mathbf{r}'|} \tag{6}$$

where $G$ is proportional to the band gap of the system [42,43]. For conductors, which do not have a band gap, the decay rate is slower and a power law dependence

$$\rho\,(\mathbf{r},\mathbf{r}') \sim |\mathbf{r}-\mathbf{r}'|^{-d} \tag{7}$$

is expected [39,40,44], where $d$ is related to the dimensionality of the system.

For finite systems, especially disordered ones, the situation is less clear cut [45]. Kohn has conjectured that the density matrix behaves as

$$\rho\,(\mathbf{r},\mathbf{r}') \sim e^{-|\mathbf{r}-\mathbf{r}'|/\ell} \quad (|\mathbf{r}-\mathbf{r}'| \to \infty)\ , \tag{8}$$

where $\ell$ is a system-specific length-scale parameter [44].

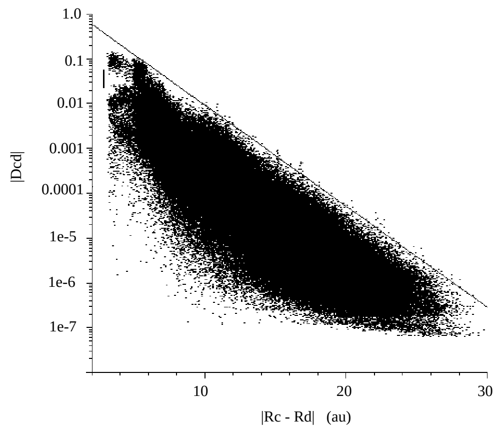


FIG. 1. Magnitudes of RHF/3-21G density matrix elements with separation of basis function centers for the $(H_2O)_{110}$ water cluster. The density matrix has been converged to within $10^{-7}$ au. The line is $f(x) = ae^{-bx}$ with $a$=1.7 and $b$=0.52 au$^{-1}$.

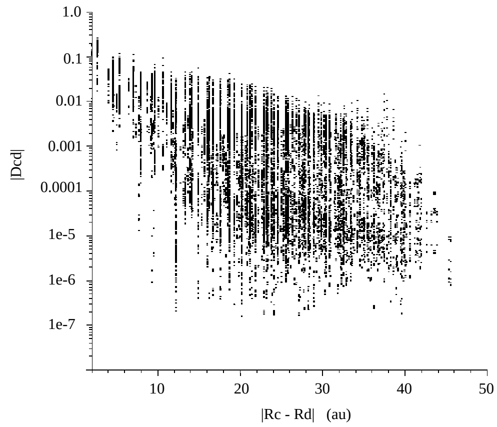


FIG. 2. Magnitudes of RHF/3-21G density matrix elements with separation of basis function centers for the graphitic sheet $C_{150}H_{30}$. The density matrix has been converged to within $10^{-7}$ au. The stratification of density matrix elementsis due to molecular symmetry.

In practice, we find that for systems such as proteins and water clusters, elements of the density matrix may be bounded (approximately) by an exponential, which decays with basis function separation. For example, see

Figure 1. On the other hand, elements of the density matrix corresponding to molecular systems with a small HOMO-LUMO gap exhibit a decay that is much slower, consistent with the algebraic decay of conductors. For example, see Figure 2.

## III. METHODS FOR K BUILDS

The first specialized methods for computing $\mathbf{K}$ were developed to efficiently truncate the potentially infinite summations that arise in the HF theory of periodic systems [37,39–41]. These methods, implemented in the periodic HF program CRYSTAL, are based on truncating density and exchange matrices with distance dependent cutoffs.

Locality of the density matrix for insulating systems has been used to argue the existence of reduced scaling methods for computing exchange interactions in HF theory [3,17,46–50]. The most straightforward approach to exploiting this locality is to use the exchange component of the density weighted integral estimates commonly employed in direct SCF methods [51,52];

$$\begin{aligned}\mathrm{Max}\left[\left|D_{cd}\right|,\left|D_{ab}\right|,\left|D_{cb}\right|,\left|D_{ad}\right|\right]\times \\ \left|\left(\rho_{ac}\middle|\rho_{ac}\right)\right|^{1/2}\left|\left(\rho_{bd}\middle|\rho_{bd}\right)\right|^{1/2}\leq \text{ thresh} .\end{aligned} \tag{9}$$

However, this approach results in an algorithm with $N^2$ complexity, due at least to logical testing.

The first linear scaling method for computing $\mathbf{K}$ was developed in Ref. [17]. This method, which we refer to as ONX_A, is based on the assumed exponential decay of the density matrix, and the corresponding estimate

$$\widetilde{K}_{ab}=\int d\mathbf{r}\int d\mathbf{r}'\;\phi_a\left(\mathbf{r}\right)\phi_b\left(\mathbf{r}'\right)\left|\mathbf{r}-\mathbf{r}'\right|^{-1}e^{-\left|\mathbf{r}-\mathbf{r}'\right|/\ell} \tag{10}$$

for thresholding insignificant elements of $\mathbf{K}$.

More recently, Burant *et al.* have presented an alternative method, near-field exchange (NFX), for reducing the cost of $\mathbf{K}$ builds [18]. NFX limits exchange interactions to domains that are analogous to the near-field regions of the Gaussian very Fast Multipole Method (GvFMM) [6]. In the lowest order approximation $R+0$, the region of exchange interaction is identical to the GvFMM near-field, and the cost of constructing the near-field $\mathbf{J}$ can be amortized in the $\mathbf{K}$ build.

## IV. EXCHANGE RECONSIDERED

The aforementioned methods for computing $\mathbf{K}$ [17,18] have demonstrated reduced computational scalings while maintaining errors at a reasonable level. However, the success of these methods relies on an empirically determined knowledge of density matrix asymptotics to achieve a balance between speed and accuracy. Implicit is the assumption that exchange interactions become negligible when basis function separations exceed certain distance dependent cutoff criteria. A subtle difference between ONX_A and NFX is that when the centers $A$ and $B$ are close, ONX_A may employ the entire density matrix, while NFX will not compute elements of $\mathbf{K}$ when $C$ and $D$ are well separated, which is equivalent to employing a sparse density matrix. In any case, as can be clearly seen in Fig. 1, simply bounding these interactions with a cutoff is necessarily inefficient as elements of $\mathbf{D}$ corresponding to a given separation can vary dramatically in size, in this example by four orders of magnitude.

If an effective zero is considered $10^{-8}$ au, numerical studies show that well converged density matrices remain dense, even for very extended systems with large HOMO-LUMO gaps [36,53]. See for example Fig. 1. On careful examination, we find that thresholding itself easily dominates the computation of ERIs. Moreover, retaining a dense density matrix does not necessarily lead to an $N^2$-scaling computation of ERIs, as for insulating systems small but numerically significant elements of the density matrix generally incur negligible work (see Section VIII for an illustration of this principle).

## V. ORDER N EXCHANGE

A new implementation of ONX is described that employs the density weighted integral estimate

$$\left|D_{cd}\right|\left|\left(\rho_{ac}\middle|\rho_{ac}\right)\right|^{1/2}\left|\left(\rho_{bd}\middle|\rho_{bd}\right)\right|^{1/2}\leq \texttt{thresh}\,, \tag{11}$$

which is equivalent to standard direct methods. As demonstrated in Section VII A, this approach achieves linear scaling for systems with a large HOMO-LUMO gap, and reduces naturally to a quadratic dependence in systems where the HOMO-LUMO gap is small.

```
do c =1,N
   do a =1,N
      compute |(ρac|ρac)|^1/2
      if(|(ρac|ρac)|^1/2 > thresh /[Dcd^max|(ρbd^max|ρbd^max)|^1/2])
        then add ρac to {ρac}
      endif
   enddo
   for current c, sort {ρac} by decreasing |(ρac|ρac)|^1/2
   store pointers first(c) and last(c)
enddo
```

FIG. 3. Ordering the list of significant distributions.

The $N^2$ complexity of thresholding can be avoided by foregoing permutational symmetry and employing a structured list of distributions $\{\rho_{ac}\}$, which for each basis function $c$ are ordered by decreasing value of the integral estimate $|(\rho_{ac}|\rho_{ac})|^{1/2}$. To avoid an $N^2$ list, only those distributions that satisfy

$$\left|(\rho_{ac}|\rho_{ac})\right|^{1/2} \le \frac{\texttt{thresh}}{D_{cd}^{\max}\left|(\rho_{bd}^{\max}|\rho_{bd}^{\max})\right|^{1/2}} \tag{12}$$

are considered, where $D_{cd}^{\max}$ is the largest element of the density matrix and $\left|(\rho_{bd}^{\max}|\rho_{bd}^{\max})\right|$ is the largest integral estimate. The procedure for structuring this list of distributions is outlined in Fig. 3.

ONX **K** builds begin with an outer loop over density matrix elements, as shown in Fig. 4. An intermediate list of distributions significant to basis function $c$ is then formed using the criterion

$$\left|D_{cd}\right|\left|(\rho_{ac}|\rho_{ac})\right|^{1/2}\left|(\rho_{b_f d}|\rho_{b_f d})\right|^{1/2} \le \texttt{thresh}\,, \tag{13}$$

where $\rho_{b_f d}$ corresponds to the largest (first) integral estimate involving basis function $d$. An intermediate list of distributions significant to $d$ are likewise formed.

```
     do c = 1,N
        do d = 1,N
           a_f = first(c); a_l = last(c)
           b_f = first(d); b_l = last(d)
           do  a = a_f,a_l
              if(|D_cd| * |(rho_ac|rho_ac)|^1/2* |(rho_bfd|rho_bfd)|^1/2 <
                thresh)goto 201
              buffer intermediate values for rho_ac
           enddo
201        continue
           do b = b_f,b_l
              if(|D_cd| * |(rho_afc|rho_afc)|^1/2* |(rho_bd|rho_bd)|^1/2<
                thresh)goto 301
              buffer intermediate values for rho_bd
           enddo
301        continue
           do  a = a_f,a_l
              if(|D_cd| * |(rho_ac|rho_ac)|^1/2*|(rho_bfd|rho_bfd)|^1/2<
                thresh)goto 501
              do  b = b_f,b_l
                 if(|D_cd| * |(rho_ac|rho_ac)|^1/2*|(rho_bd|rho_bd)|^1/2<
                   thresh)goto 601
                 compute (rho_ac|rho_bd)
                 K_ab = K_ab + D_cd * (rho_ac|rho_bd)
              enddo
601           continue
           enddo
501        continue
        endddo
     enddo
```

FIG. 4. The ONX loops. Skipping entirely out of loops (`goto 201`, `301`, `501`, and `601`) is possible due to the ordering of distribution and integral estimate lists.

Next, a two-fold loop over the intermediate lists of $\rho_{ac}$ and $\rho_{bd}$ is begun in which ERIs are computed and contracted with the density matrix to form elements of **K**. For each $\rho_{ac}$, satisfaction of Eq. (13) validates `goto 501`, while the inner-most loop over $\rho_{bd}$ can be skipped out of when Eq. (11) is met.

## VI. RECURSIVE F BUILDS

A unique feature of both ONX and QCTC is that they employ density based error estimates, and can therefore employ incremental **F** builds [51,52,54]. In the incremental approach, **F** is computed by recursion as

$$\mathbf{F}^m = \mathbf{F}^{m-1} + \mathbf{\Delta J}^{m-1} - \frac{1}{2}\mathbf{\Delta K}^{m-1}\,, \tag{14}$$

where $m$ is the SCF cycle, and $\mathbf{\Delta J}^m$ and $\mathbf{\Delta K}^m$ are computed using the difference density matrix $\mathbf{\Delta D}^m = \mathbf{D}^m - \mathbf{D}^{m-1}$. For $m = 1$, the core Hamiltonian matrix and a guess density matrix are used.

With recursion, the computation of $\mathbf{\Delta J}$ and $\mathbf{\Delta K}$ is greatly accelerated as convergence is neared. However, the most recent full **F** build determines the maximum precision achievable, and errors that grow like $\sqrt{m}$ propagate during recursion [52]. To minimize this effect, a full Fock build can be carried out periodically, and/or the thresholding parameter can be continuously tightened.

We have implemented a simple algorithm for recursive Fock builds. Minimized density differences [52] and other acceleration techniques are applicable to our approach and are likely to substantially improve performance, but have not been implemented at this time. Subroutines ONX and QCTC [7] have been interfaced with a modified version of the *ab initio* electronic structure program HONDO 95.3 [55] (MONDO). A single thresholding parameter `thresh`, described above for ONX and in Ref. [7] for QCTC, is used to control errors in **F**. With this arrangement, the errors in **J** are smaller than those in **K**, as `thresh` has a somewhat different function in QCTC than it does in `onx`. As **J** builds are still more competitive than **K** builds, balancing the errors from **J** and **K** builds is presently unwarranted.

In the following, two levels of precision are considered; those achieved with `thresh` $= 10^{-7}$ and those with `thresh` $= 10^{-9}$. All calculations are started from an extended Hückel guess. The first three SCF cycles are then performed without recursion using `thresh` $= 10^{-5}$. A full Fock build with `thresh` $= 10^{-7}$ is then performed, followed by incremental updates with decreasing values of `thresh` to avoid error propagation. Convergence is achieved when the criteria $\Delta E < 10^{-5}$ and $\max[\mathbf{\Delta D}] < 10^{-4}$ are satisfied. This protocol is illustrated in Table I. To reach an error commensurate with `thresh` $= 10^{-9}$, an identical calculation is carried out, with a full `thresh` $= 10^{-9}$ Fock build after the aforementioned criteria are met. This is followed with recursive **F** builds involving decreasing values of `thresh`, reaching convergence when $\Delta E < 10^{-7}$ and $\max[\mathbf{\Delta D}] < 10^{-5}$, typically within three to four additional cycles.

TABLE I. Typical output from a RHF/3-21G calculation of a 150 water molecule cluster with `thresh` $= 10^{-7}$. Note that after SCF cycle 4, `thresh` is tightened to avoid the propagation of errors. For the more accurate calculations, SCF cycle 13 would be a full Fock build with `thresh` $= 10^{-9}$, followed by recursion.

| | | | | | cpu time (seconds) | |
|---|---|---|---|---|---|---|
| SCF cycle | Energy (hartree) | $\Delta$E | max[$\mathbf{\Delta D}$] | `thresh` | ONX | QCTC |
| 1[a] | -11239.101 259 | 1.1e04 | 5.5e-01 | 1.0e-05 | 377.2 | 305.3 |
| 2[a] | -11335.567 964 | 9.6e01 | 1.3e-01 | 1.0e-05 | 271.6 | 299.4 |
| 3[a] | -11340.406 789 | 4.8e00 | 4.1e-02 | 1.0e-05 | 257.7 | 296.2 |
| 4[a] | -11341.005 010 | 6.0e-01 | 2.3e-02 | 1.0e-07 | 632.8 | 621.7 |
| 5 | -11341.104 223 | 9.9e-02 | 7.4e-03 | 8.1e-08 | 280.9 | 320.3 |
| 6 | -11341.122 070 | 1.8e-02 | 3.1e-03 | 3.6e-08 | 223.6 | 295.8 |
| 7 | -11341.126 944 | 4.9e-03 | 1.6e-03 | 1.7e-08 | 270.6 | 293.3 |
| 8 | -11341.127 859 | 9.1e-04 | 7.1e-04 | 9.4e-09 | 246.3 | 271.4 |
| 9 | -11341.128 046 | 1.9e-04 | 3.5e-04 | 4.5e-09 | 246.0 | 248.2 |
| 10 | -11341.128 086 | 4.0e-05 | 1.6e-04 | 2.2e-09 | 231.9 | 235.7 |
| 11 | -11341.128 097 | 1.1e-05 | 7.5e-05 | 1.1e-09 | 226.9 | 216.2 |
| 12 | -11341.128 101 | 3.5e-06 | 3.8e-05 | 5.9e-10 | 204.1 | 202.5 |

[a]Full Fock matrix computed in this SCF cycle.

## VII. BENCHMARKS

### A. CPU Timings

To demonstrate the scaling properties of our new approach, MONDO RHF calculations with the 3-21G and 6-31G basis sets and `thresh` $= 10^{-7}$ and `thresh` $= 10^{-9}$ have been performed on a sequence of three-dimensional water clusters, polyglycine $\alpha$-helices, and graphitic sheets. All calculations have been performed on a single processor IBM RS/6000 model 590. These results are shown in Figures 5-14, where ONX(avg) and QCTC(avg) refer to the cpu time per formation of the exchange and Coulomb matrix averaged over the entire calculation. MONDO(avg) is simply the sum of ONX(avg) and QCTC(avg) times. MONDO(max) is the most expensive Fock matrix construction time incurred in the calculation. This corresponds to the fourth SCF cycle when `thresh` is set to $10^{-7}$, or the full Fock build with `thresh` $= 10^{-9}$.

The coordinates for the water clusters are the same as in Refs. [7,17,18]. Note that the coordinates for the polyglycine calculations are not the same as those employed by Burant *et. al.* [18]; their glycine coordinates correspond to an extended linear chain, while ours is in the $\alpha$-helix conformation [17]. A carbon-carbon bond length of 1.42 au has been used to produce the coordinates of graphitic sheets with stoichiometry $C_{6n^2}H_{6n}$, for $n$=2 to 7 [56].

To illustrate the potential $N^2$ scaling of thresholding, we have also included cpu timings labeled No Skipout in Figure 5. These timings correspond to full Fock builds in which the ONX loop skipouts have been disabled. In particular, `601 continue` and `501 continue` have been moved within their respective loops (see Figure 4). The cpu times for this method have been fit to a quadratic and are also shown in Figure 5 as No Skipout(fit).

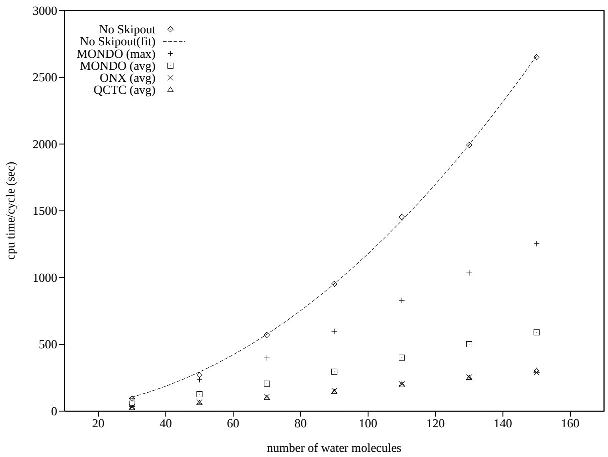


FIG. 5. RHF/3-21G water cluster timings with `thresh` $= 10^{-7}$.

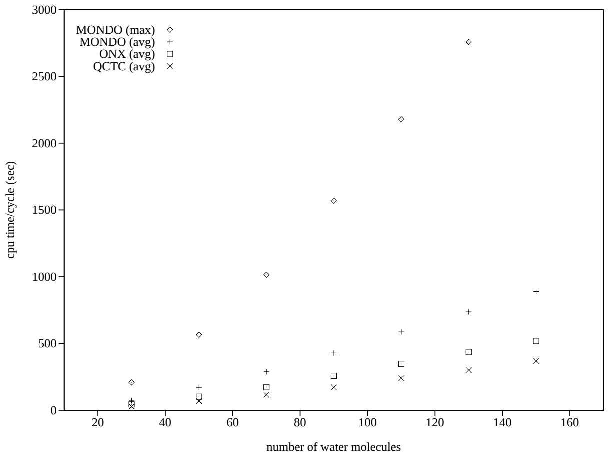


FIG. 6. RHF/3-21G water cluster timings with `thresh` $= 10^{-9}$.

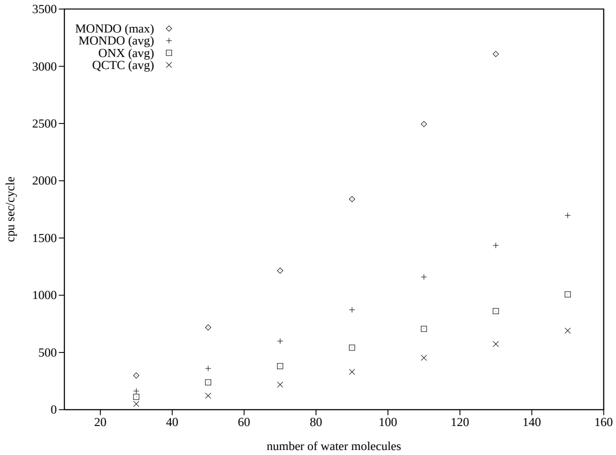


FIG. 7. RHF/6-31G water cluster timings with `thresh` $= 10^{-7}$.

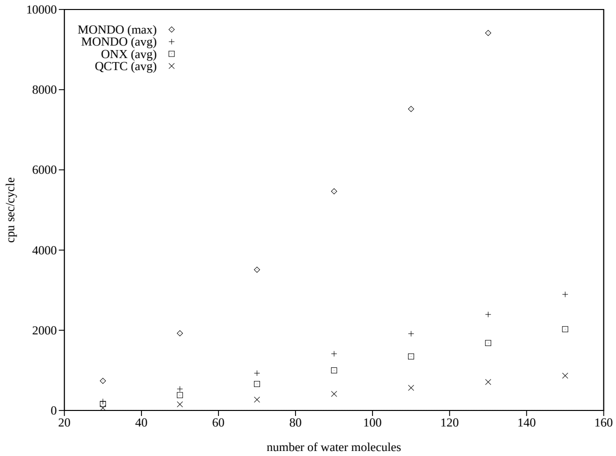


FIG. 8. RHF/6-31G water cluster timings with `thresh` $= 10^{-9}$.

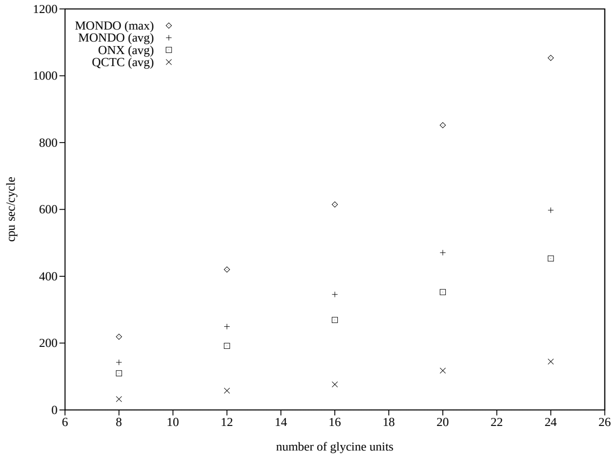


FIG. 9. RHF/3-21G polyglycine $\alpha$-helix timings with $\texttt{thresh} = 10^{-7}$.

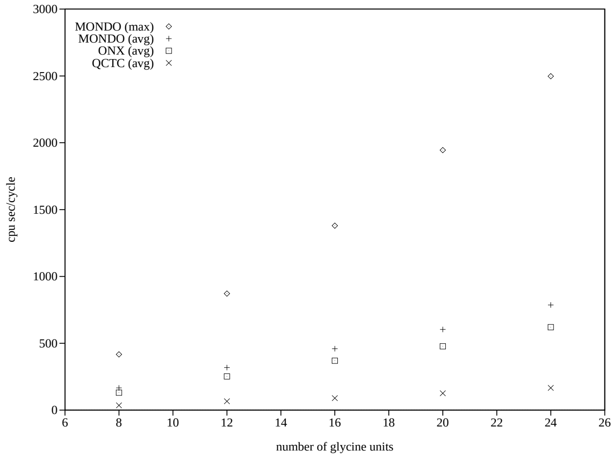


FIG. 10. RHF/3-21G polyglycine $\alpha$-helix timings with $\texttt{thresh} = 10^{-9}$.

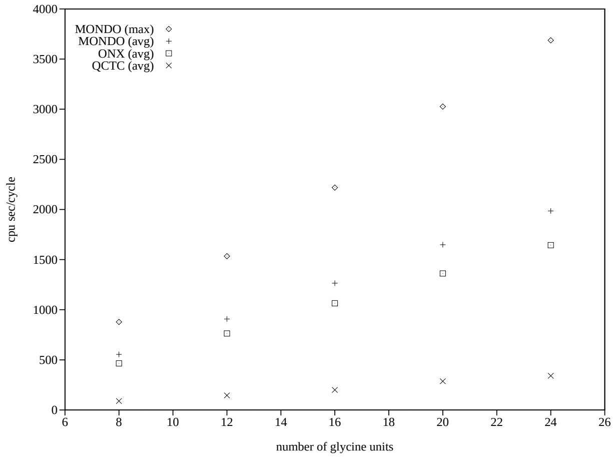


FIG. 11. RHF/6-31G polyglycine $\alpha$-helix timings with $\texttt{thresh} = 10^{-7}$.

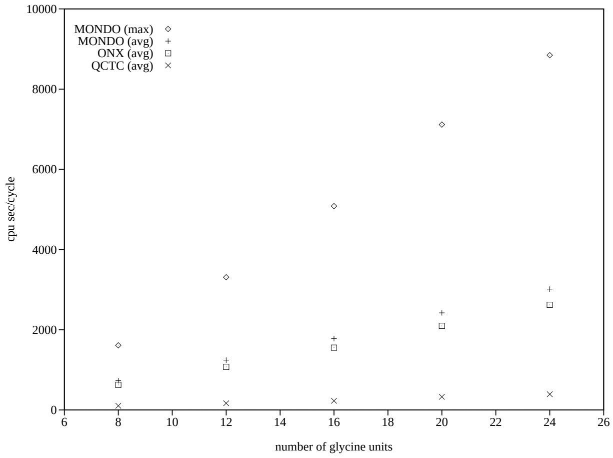


FIG. 12. RHF/6-31G polyglycine $\alpha$-helix timings with $\texttt{thresh} = 10^{-9}$.

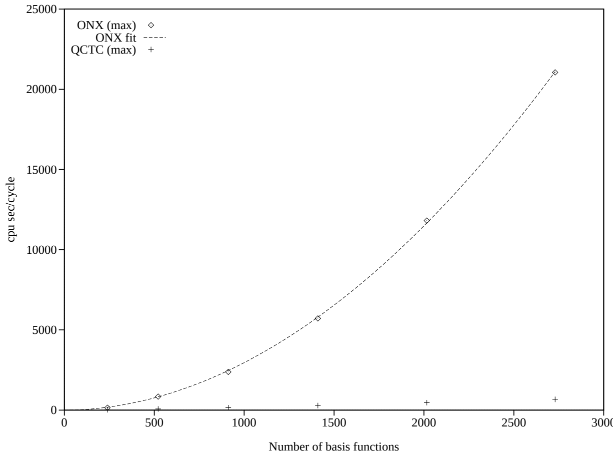


FIG. 13. RHF/3-21G graphitic sheet timings with $\texttt{thresh} = 10^{-7}$.

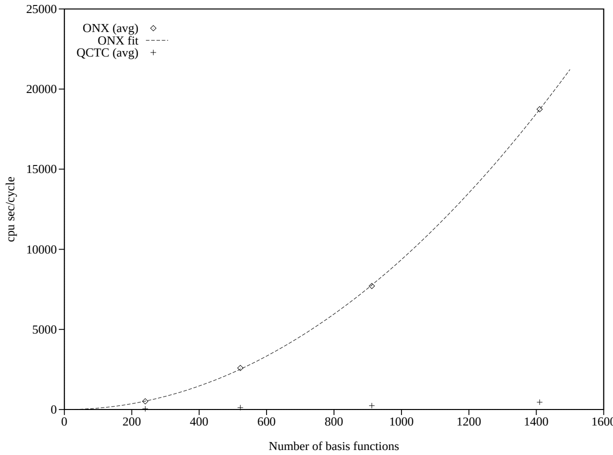


FIG. 14. RHF/6-31G graphitic sheet timings with $\texttt{thresh} = 10^{-7}$.

### B. Errors and convergence

Absolute errors in the converged total energies for each of the molecular sequences and basis sets are shown in Figures 15 to 19. In each case, the errors have been computed by comparison to MONDO with $\mathtt{thresh} = 10^{-11}$. In Table II, we have also listed three representative systems from each sequence along with corresponding converged total energies for decreasing values of `thresh`. The column labeled max$[\mathbf{\Delta D}]$ in Table II is the convergence criterion used. The error per atom for the the same systems is listed in Table III.

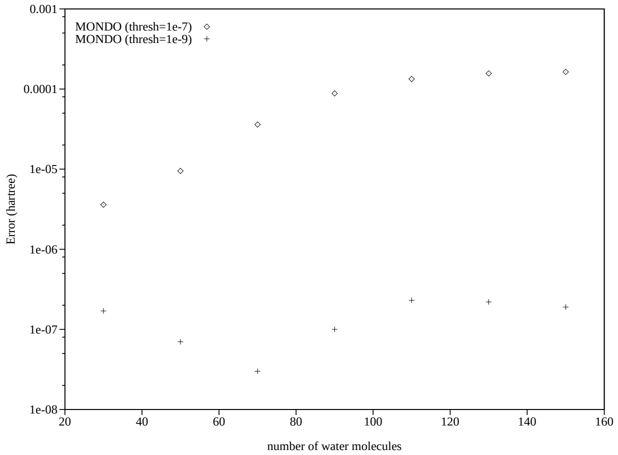


FIG. 15. RHF/3-21G water cluster errors in converged total energy.

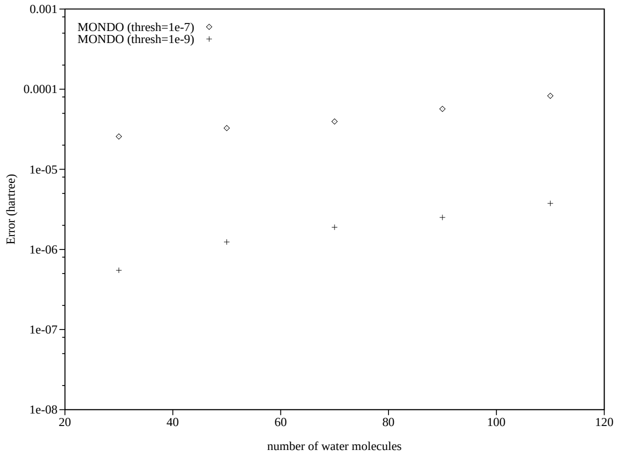


FIG. 16. RHF/6-31G water cluster errors in converged total energy.

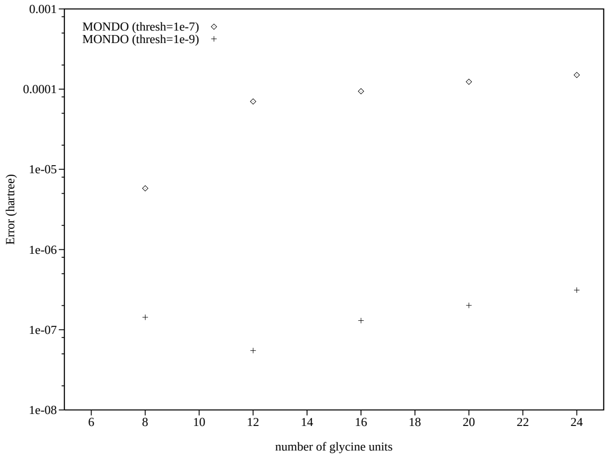


FIG. 17. RHF/3-21G glycine $\alpha$-helix errors in converged total energy.

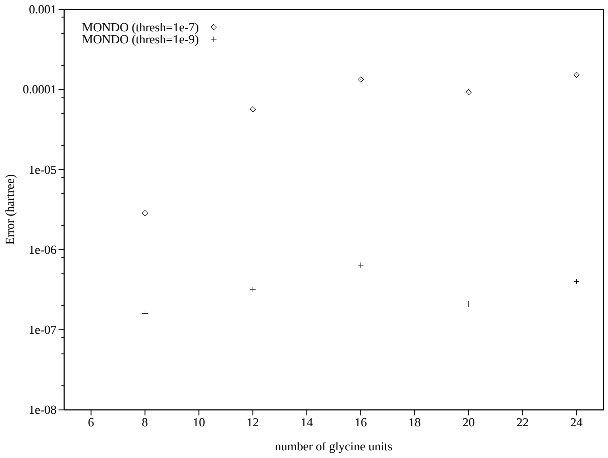


FIG. 18. RHF/6-31G glycine $\alpha$-helix errors in converged total energy.

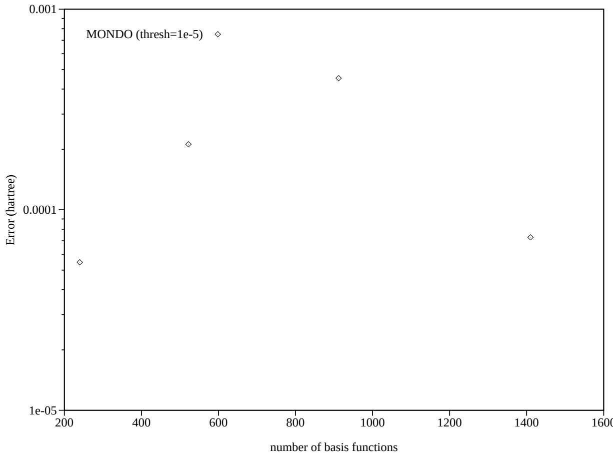


FIG. 19. RHF/3-21G graphitic sheet errors in converged total energy.

TABLE II. The convergence of the RHF/3-21G total energy for different molecular systems and values of `thresh`. For each calculation `thresh` was held constant and the incremental Fock build technique was not used.

| | | Total energy (hartree) | | |
|---|---|---|---|---|
| `thresh` | max[$\mathbf{\Delta D}$] | $(H_2O)_{110}$ | $(Gly)_{24}$ | $C_{54}H_{18}$ |
| 10e-6 | 1.0e-5 | -8316.844 140 | -4936.961 534 | -2044.145 492 |
| 10e-7 | 1.0e-6 | -8316.845 692 | -4936.961 942 | -2044.147 993 |
| 10e-8 | 1.0e-6 | -8316.845 757 | -4936.962 076 | -2044.147 809 |
| 10e-9 | 1.0e-6 | -8316.845 776 | -4936.962 087 | -2044.147 818 |
| 10e-11 | 1.0e-7 | -8316.845 776 | -4936.962 087 | -2044.147 819 |

TABLE III. The RHF/3-21G error per atom for different molecular systems and values of `thresh`. Within each calculation `thresh` was held constant and the incremental Fock build technique was not used.

| | Error per atom ($\mu$hartree) | | |
|---|---|---|---|
| `thresh` | $(H_2O)_{110}$ | $(Gly)_{24}$ | $C_{54}H_{18}$ |
| 10e-6 | 5.0 | 3.3 | 32 |
| 10e-7 | 0.25 | 0.86 | 2.4 |
| 10e-8 | 0.057 | 0.071 | 0.13 |
| 10e-9 | 0.00069 | 0.0080 | 0.020 |

### C. P53

In Ref. [7], we reported a RHF/3-21G calculation on the tetramerization monomer of P53, in what we believe to be the largest non-empirical HF calculation to date. In Table IV, the convergence of this calculation with decreasing values of `thresh` is reported, along with cpu times for ONX and QCTC.

TABLE IV. Convergence of the P53 RHF/3-21G total energy and corresponding cpu times per SCF cycle for decreasing values of `thresh`. Energies and timings correspond to full Fock builds and a constant value of `thresh`.

| | | | cpu time (hours) | |
|---|---|---|---|---|
| `thresh` | max[$\mathbf{\Delta D}$] | $E_{RHF}$ (hartree) | ONX | QCTC |
| 10e-6 | 1.0e-5 | -17115.306 141 | 0.67 | 0.36 |
| 10e-7 | 1.0e-6 | -17115.305 875 | 1.36 | 0.49 |
| 10e-8 | 1.0e-6 | -17115.305 227 | 2.31 | 0.66 |
| 10e-9 | 1.0e-6 | -17115.305 248 | 3.74 | 0.87 |
| 10e-11 | 1.0e-7 | -17115.305 249 | 8.35 | 1.46 |

## VIII. DISCUSSION

The new implementation of ONX is a marked improvement over previous linear scaling methods for computing the HF exchange matrix. For full Fock builds, ONX is approximately 3-4 times faster than ONX_A, and is numerically equivalent to the direct SCF methods now standard in modern quantum chemistry. Perhaps equally dramatic, an unsophisticated approach to incremental Fock builds has been found to increase this margin by another factor of 2, with similar gains for QCTC. It is likely that a more sophisticated approach to incremental Fock builds will improve this factor.

For the insulating systems studied, the computation of $\mathbf{K}$ with ONX scales linearly with system size. For graphitic sheets which have small HOMO-LUMO gaps tending to zero with $N$, a quadratic scaling is observed. It is satisfying that the transition from linear to quadratic is entirely controlled by the structure of $\mathbf{D}$.

Calculations with $\texttt{thresh} = 10^{-9}$ have demonstrated linear scaling and $\mu$hartree accuracy in converged total energies for the 3-21G and 6-31G basis sets. We reconcile this level of accuracy, the use of an effectively dense density matrix, and linear scaling by noting that even though $\mathbf{D}$ is an $N^2$ quantity, ONX is dominated by the evaluation of ERIs, which is $\mathcal{O}(N)$ for water clusters and $\alpha$-helices. This surprising result is highlighted by Figures 20-23, which show elements of $\mathbf{D}$ versus basis function separation, binned by the number of associated ERIs $M_{\rm int}$, for the RHF/3-21G $(H_2O)_{150}$ calculation of $\mathbf{K}$. These results are contrasted by Figures 24-27, which show the corresponding values for the $C_{150}H_{30}$ calculation of $\mathbf{K}$. For the water cluster, matrix elements involving large numbers of ERIs correspond to small separations, with a definite correlation between larger separations and fewer ERIs. For the graphitic sheet, density matrix elements corresponding to large $M_{\rm int}$ persist even for separations of up to 40 atomic units. Note that the percentages quoted in Figures 20-23 correspond to a much smaller ERI total than do those in Figures 24-27.

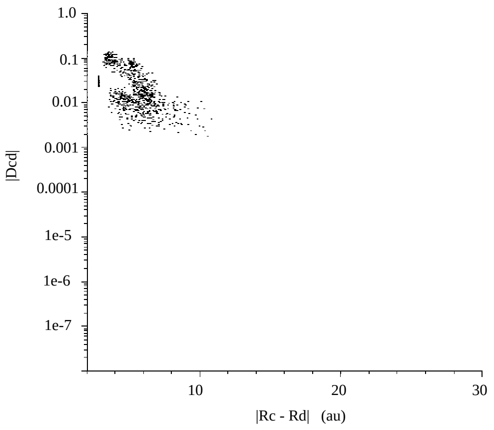


FIG. 20. $|D_{cd}|$ versus $|C - D|$ for the $(H_2O)_{110}$ cluster for which $M_{\rm int} > 1000$. This plot represents 6.5% of the integrals and 0.4% of **D**.

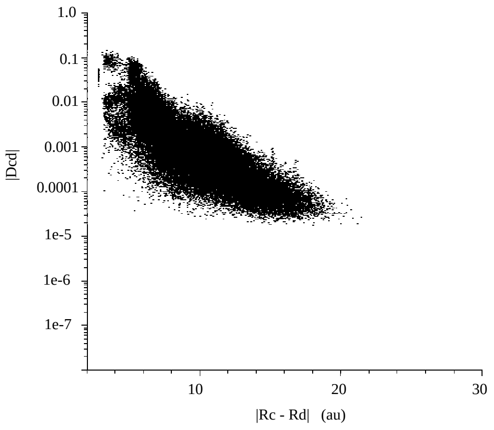


FIG. 21. $|D_{cd}|$ versus $|C - D|$ for the $(H_2O)_{110}$ cluster for which $1000 > M_{\rm int} > 100$. This plot represents 66.4% of the integrals and 20.8% of **D**.

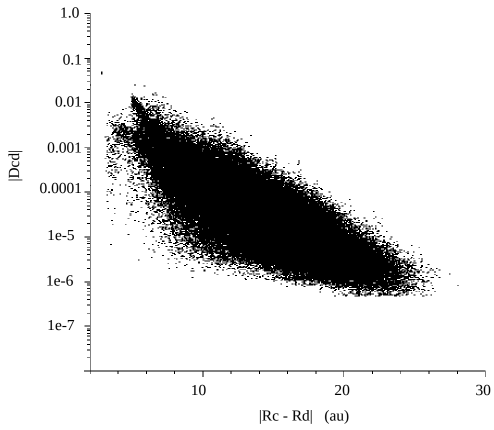


FIG. 22. $|D_{cd}|$ versus $|C - D|$ for the $(H_2O)_{110}$ cluster for which $100 > M_{\rm int} > 10$. This plot represents 25.6% of the integrals and 53.2% of **D**.

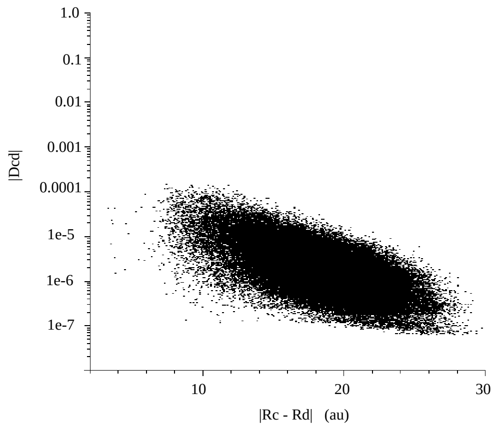


FIG. 23. $|D_{cd}|$ versus $|C - D|$ for the $(H_2O)_{110}$ cluster for which $10 > M_{\rm int}$. This plot represents 1.5% of the integrals and 25.5% of **D**.

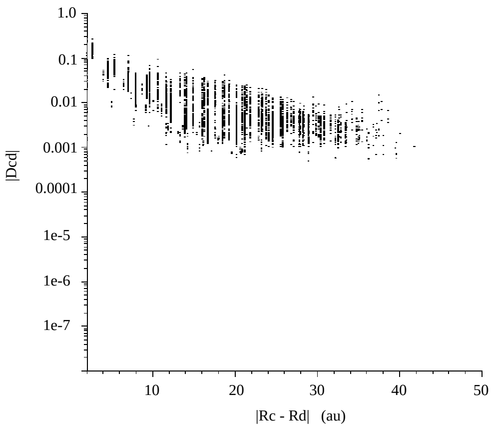


FIG. 24. $|D_{cd}|$ versus $|C - D|$ for $C_{150}H_{30}$ for which $M_{\rm int} > 1000$. This plot represents 58.6% of the integrals and 17.9% of $\mathbf{D}$.

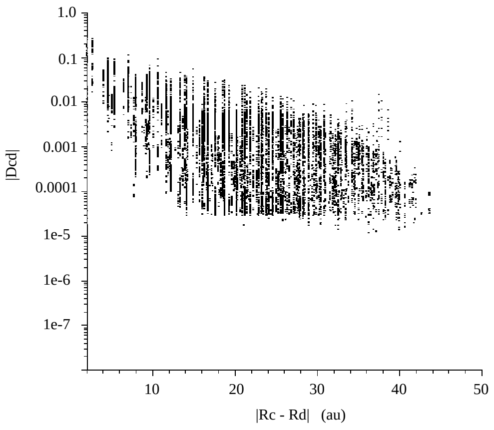


FIG. 25. $|D_{cd}|$ versus $|C - D|$ for $C_{150}H_{30}$ for which $1000 > M_{\rm int} > 100$. This plot represents 39.4% of the integrals and 52.0% of $\mathbf{D}$.

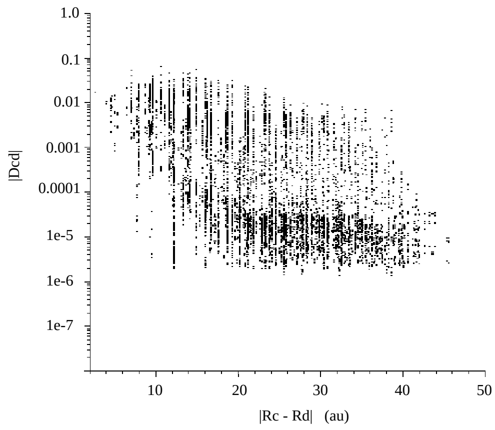


FIG. 26. $|D_{cd}|$ versus $|C - D|$ for $C_{150}H_{30}$ for which $100 > M_{\rm int} > 10$. This plot represents 1.9% of the integrals and 23.6% of $\mathbf{D}$.

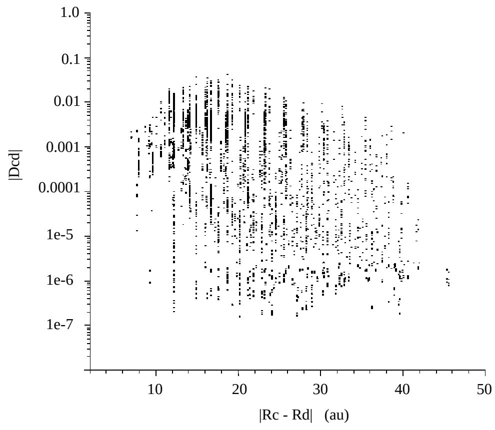


FIG. 27. $|D_{cd}|$ versus $|C - D|$ for $C_{150}H_{30}$ for which $10 > M_{\rm int}$. This plot represents 0.04% of the integrals and 6.4% of $\mathbf{D}$.

## IX. CONCLUSION

We have shown that locality of the Gaussian-based density matrix is directly exploitable without sacrificing accuracy, and without enforcing locality. However, this requires the careful minimization of costs associated with data handling and manipulation, such as the thresholding of integrals. One effective strategy for achieving this is to employ data structures that permit loop skip-outs. Note that this approach is also employed by QCTC in thresholding near-field Coulomb integrals [7]. Related strategies may also prove useful in reducing the complexity of MO-based theories of electron-correlation.

Achieving a linear scaling SCF method for all $N$ will ultimately require the use of sparse matrices to circumvent the $N^{2-3}$ cpu time and $N^2$ memory associated with dense MO and/or density matrices. However, this work demonstrates the independence of achieving linear scaling Fock builds and the use of sparse matrices.

In conclusion, a new method has been presented for computing the Hartree-Fock exchange matrix, and linear scaling incremental Fock builds have been demonstrated for the first time. This new approach retains the reliability associated with direct SCF, and is able to achieve $\mu$hartree accuracies and linear scaling for insulating systems with both the 3-21G and the 6-31G basis set. Significantly, a factor of 8 improvement over our previous method [7] has been achieved. For systems with small HOMO-LUMO gaps, such as graphitic sheets, the approach naturally reverts to an $\mathcal{O}(N^2)$ complexity.

MONDO RHF/3-21G calculations on the P53 tetramerization monomer, involving 698 atoms and 3836 basis functions, have been carried out to a precision of 1 $\mu$hartree on the commonly available IBM RS/6000 model 590, and may provide a useful benchmark for the development of new algorithms [56].

## ACKNOWLEDGMENTS

ES and MC acknowledge the Minnesota Supercomputer Institute for resources and facilities. Computational resources were provided by the University of Minnesota-IBM Shared Research Project.

[1] C. A. White and M. Head-Gordon, J. Chem. Phys. **101**, 6593 (1994).
[2] C. A. White, B. Johnson, P. Gill, and M. Head-Gordon, Chem. Phys. Letters **230**, 8 (1994).
[3] C. A. White and M. Head-Gordon, J. Chem. Phys. **104**, 2620 (1996).
[4] C. A. White and M. Head-Gordon, Chem. Phys. Lett. **257**, 647 (1996).
[5] C. A. White and M. Head-Gordon, J. Chem. Phys. **105**, 5061 (1996).
[6] M. C. Strain, G. E. Scuseria, and M. J. Frisch, Science **271**, 51 (1996).
[7] M. Challacombe and E. Schwegler, J. Chem. Phys. **105**, 0 (1997).
[8] J. P. Dombroski, S. W. Taylor, and P. M. W. Gill, J. Phys. Chem. **100**, 6272 (1996).
[9] X. P. Li, R. W. Nunes, and D. Vanderbilt, Phys. Rev. B **47**, 10891 (1993).
[10] S. Goedecker and L. Colombo, Phys. Rev. Lett. **73**, 122 (1994).
[11] E. Hernández, M. J. Gillan, and C. M. Goringe, Phys. Rev. B **53**, 7147 (1996).
[12] W. Yang and T. S. Lee, J. Chem. Phys. **103**, 5674 (1995).
[13] A. F. Voter, J. D. Kress, and R. N. Silver, Phys. Rev. B **53**, 12733 (1996).
[14] W. Kohn, Phys. Rev. Lett. **76**, 3168 (1996).
[15] A. P. Horsfield and A. M. Bratkovsky, Phys. Rev. B **53**, 15381 (1996).
[16] S. L. Dixon and K. M. Merz, J. Chem. Phys. **104**, 6643 (1996).
[17] E. Schwegler and M. Challacombe, J. Chem. Phys. **105**, 2726 (1996).
[18] J. C. Burant, G. E. Scuseria, and M. J. Frisch, J. Chem. Phys. **105**, 8969 (1996).
[19] P. Ordejón, D. A. Drabold, R. M. Martin, and M. P. Grumbach, Phys. Rev. B **51**, 1456 (1995).
[20] F. Mauri, G. Galli, and R. Car, Phys. Rev. B **47**, 9973 (1993).
[21] P. Ordejón and D. A. Drabold, Phys. Rev. B **48**, 14646 (1993).
[22] E. B. Stechel, A. R. Williams, and P. J. Feibleman, Phys. Rev. B **49**, 10088 (1994).
[23] X. Chen, J. M. Langlois, and W. A. Goddard III, Phys. Rev. B **54**, 2348 (1995).
[24] J. Kim, F. Mauri, and G. Galli, Phys. Rev. B **52**, 1640 (1995).
[25] P. Ordejón, E. Artacho, and J. M. Soler, Phys. Rev. B **53**, 10441 (1996).
[26] J. P. Stewart, Int. J. Quant. Chem. **58**, 133 (1996).
[27] A. Canning *et al.*, Comp. Phys. Comm. **94**, 89 (1996).
[28] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[29] A. D. Becke, J. Chem. Phys. **104**, 1040 (1996).
[30] B. Johnson, P. Gill, and J. Pople, J. Chem. Phys. **98**, 5612 (1993).
[31] C. W. Bauschlicher, Chem. Phys. Lett. **246**, 40 (1995).
[32] G. S. Tschumper, J. T. Fermann, and H. F. Schaefer III, J. Chem. Phys. **104**, 3676 (1996).
[33] V. Barone, C. Adamo, and F. Mele, Chem. Phys. Letters **249**, 290 (1996).
[34] P. O. Löwdin, Phys. Rev. **97**, 1490 (1955).
[35] P. R. Surján, M. Kertész, A. Karpfen, and J. Koller, Phys. Rev. B **27**, 7583 (1983).
[36] P. Maslen *et al.*, manuscript in preparation.
[37] R. Dovesi, C. Pisani, and C. Roetti, Int. J. Quant. Chem. **17**, 517 (1980).
[38] H. J. Monkhorst and M. Kertész, Phys. Rev. B **24**, 3015 (1981).
[39] M. Causà, R. Dovesi, C. Pisani, and V. R. Saunders, J. Phys. Chem. **92**, 909 (1988).
[40] C. Pisani, E. Aprà, and M. Causà, Int. J. Quant. Chem. **38**, 395 (1990).
[41] C. Pisani, M. Causà, and R. Orlando, Int. J. Quant. Chem. **38**, 419 (1990).
[42] W. Kohn, Phys. Rev. **115**, 809 (1959).
[43] J. des Cloizeaux, Phys. Rev. A **135**, 685 (1964).
[44] W. Kohn, Int. J. Quantum Chem. **56**, 229 (1995).

[45] J. M. Ziman, *Models of Disorder: The Theoretical Physics of Homogeneously Disordered Systems* (Cambridge University Press, Cambridge, England, 1979).
[46] H. Böhm and R. Ahlrichs, J. Chem. Phys. **77**, 2028 (1982).
[47] F. Maeder, J. Chem. Phys. **88**, 4934 (1988).
[48] I. Panas, J. Almlöf, and M. W. Feyereisen, Int. J. Quant. Chem. **40**, 797 (1991).
[49] V. Termath and N. C. Handy, Chem. Phys. Letters **230**, 17 (1994).
[50] G. R. Ahmadi and J. Almlöf, Chem. Phys. Letters **246**, 364 (1995).
[51] J. Almlöf, K. Faegri, and K. Korsell, J. Comp. Chem. **3**, 385 (1982).
[52] M. Häser and R. Ahlrichs, J. Comp. Chem. **10**, 104 (1989).
[53] E. Schwegler and M. Challacombe (unpublished).
[54] D. Cremer and J. Gauss, J. Comput. Chem. **7**, 274 (1986).
[55] M. Dupuis, A. Marquez, and E. R. Davidson, *HONDO 95.3 from CHEM-Station*, IBM Corporation, Neighborhood Road, Kingston, NY. 12401, 1995.
[56] Coordinate files available upon request.