A simplified density matrix minimization for linear scaling SCF theory[1]

Matt Challacombe
Los Alamos National Laboratory,
Theoretical Division
Group T-12, MS B268
Los Alamos, New Mexico 87545
EMail: MChalla@T12.LANL.Gov
(October 11, 1998)

A simplified version of the Li, Nunes and Vanderbilt [Phys. Rev. B **47** 10891 (1993)] and Daw [Phys. Rev. B **47** 10895 (1993)] density matrix minimization is introduced that requires 4 fewer matrix multiplies per minimization step relative to previous formulations. The simplified method also exhibits superior convergence properties, such that the bulk of the work may be shifted to the quadratically convergent McWeeny purification, which brings the density matrix to idempotency. Both orthogonal and non-orthogonal versions are derived. The AINV algorithm of Benzi, Meyer and Tůma [SIAM J. Sci. Comp. **17** (1996)] is introduced to linear scaling electronic structure theory, and found to be essential in transformations between orthogonal and non-orthogonal representations. These methods have been developed with an atom-blocked sparse matrix algebra that achieves sustained MFLOPS rates as high as 50% of theoretical, and implemented in the MondoSCF suite of linear scaling SCF programs. For the first time, linear scaling Hartree-Fock theory is demonstrated with three-dimensional systems, including water clusters and estane polymers. The non-orthogonal minimization is shown to be uncompetitive with minimization in an orthonormal representation. An early onset of linear scaling is found for both minimal and double zeta basis sets, and crossovers with a highly optimized eigensolver are achieved. Calculations with up to 6000 basis functions are reported. The scaling of errors with system size is investigated for various levels of approximation.

I. INTRODUCTION

Computation of the Fock matrix has historically been the limiting step in quantum chemical applications of the Hartree-Fock (HF) and Kohn-Sham (KS) theories to large systems. This is due to the expensive $\mathcal{O}(N_{\rm bas}^2)$ cost of two-electron integral computation and manipulation [1] in the direct method [2,3], where $N_{\rm bas}$ is the number of basis functions and is proportional to system size. Recently, methods with a computational complexity of $\mathcal{O}(N_{\rm bas})$ have been introduced for computing the Fock matrix [4–16], allowing access to systems large enough that the $\mathcal{O}(N_{\rm bas}^3)$ eigensolution of the self-consistent-field (SCF) equations is presently the bottleneck in the large scale application of HF and KS theories.

The computationally demanding aspects of quantum chemical SCF theory are now identical to those encountered in condensed matter formulations of the local density approximation (LDA) and tight binding theories, for which a variety of linear scaling algorithms exist. These include density matrix minimization (DMM) [17–24], orbital minimization (OM) [25–29], and the Fermi operator expansion (FOE) [30–35]. See Refs. [36–38] for excellent reviews and comparisons of these methods, as well as a more complete bibliography.

These methods achieve linear scaling for insulating systems and finite temperature metals by exploiting the short range nature of quantum interactions, the locality of which manifests itself in exponential decay of the density matrix [38–40]

$$\rho(\mathbf{r},\mathbf{r}') \sim \exp\left(-\sqrt{E_{\rm gap}}\,|\mathbf{r}-\mathbf{r}'|\right) , \tag{1}$$

where $E_{\rm gap}$ is the HOMO-LUMO gap, which is the difference between eigenvalues of the highest occupied molecular orbital and the lowest unoccupied molecular orbital. If a basis of local functions ϕ is used in calculation of the density matrix,

$$\rho(\mathbf{r},\mathbf{r}') = \sum_{ij} P_{ij}\phi_i(\mathbf{r})\phi_j(\mathbf{r}'), \tag{2}$$

then it follows that the discrete representation $\mathbf{P}$ has similar localization properties. This has been observed for both HF [14,40] and KS [38] models. Thus, for a sufficiently large system only $\mathcal{O}(N_{\rm bas})$ matrix elements are expected to remain numerically significant, and sparse matrix methods may be used to obtain the SCF in linear scaling CPU time.

The DMM, OM, and FOE methods employ a single variable, such as a localization radius or a matrix element threshold, which defines an approximate sparse matrix algebra, and which allows deviation from the exact result to be controlled systematically. However, the extension of these methods to the quantum chemical domain has faced a number of challenges related to the use of large non-orthogonal basis sets, an exacting demand for error control, and the slow (approximately exponential but slow) decay in off-diagonal elements of $\mathbf{P}$ with increasing system size [14,40], which can be severe in the case of three dimensions.

Clearly, the off-diagonal behavior of $\mathbf{P}$ depends on the physical system, the model chemistry, and the basis set. For example, minimal basis set calculations may lead to large HOMO-LUMO gaps. Likewise, HF theory tends to overestimate the HOMO-LUMO gap while LDA tends to underestimate it. One approach to limiting the bandwidth of $\mathbf{P}$ is to construct basis sets with only local support [41,42]. Another approach is the use of orthogonal bases such as wavelets

[1] To appear in the Journal of Chemical Physics.
Cite this preprint as LA-UR-98-3682

[43–45] or grids [46–48] which tend to yield very sparse matrices. A less exotic approach, pursued here, is to use standard quantum chemical basis sets together with efficient numerical algorithms to access system sizes that yield a sparse representation.

To this end, a simplified version of the Li, Nunes, Vanderbilt [17] and Daw [18] (LNVD) density matrix minimization (SDMM) is introduced that requires 4 fewer matrix multiplications than standard versions of the density matrix minimization, and which permits the time dominant work to be carried out with the quadratically convergent McWeeny purification [49]. The simplified density matrix minimization is implemented in both an orthogonal and non-orthogonal representation using an atom-blocked sparse matrix algebra that exploits both data locality and available sparsity. The AINV [50,51] method for computation of the incomplete inverse Cholesky factor is introduced to electronic structure theory as an efficient and accurate method for enabling the congruence transformation to and from an orthogonal representation, and for applying the inverse overlap matrix in non-orthogonal methods.

The paper is organized as follows. In Section II, the conventional and LNVD approach to solving the SCF equations is reviewed. In Section III, details of the LNVD method are presented, and a diagonal guess is shown to greatly simplify gradient and line searches. In Section IV, the sparse atom-blocked matrix algebra is introduced, and details of its implementation in the MondoSCF suite of linear scaling SCF programs [52] are given. Then in Section V, timings and errors are presented and discussed for water clusters and estane polymers. Finally, conclusions are drawn in Section VI.

II. OVERVIEW

A. SCF theory

In SCF theory, the density matrix $\mathbf{P}$ defines the entire physical system [53]. In particular, the electronic energy [54] is

$$E_{\rm el} = {\rm Tr}\left\{ \left(\mathbf{h} + \mathbf{F}[\mathbf{P}]\right) \mathbf{P} \right\} , \tag{3}$$

where $\mathbf{F}$ is the Fock matrix, which depends on $\mathbf{P}$ in a complicated way, and $\mathbf{h}$ is the core Hamiltonian, which does not depend on $\mathbf{P}$. For mathematical convenience, the closed shell density matrix may be factored in terms of the occupied molecular orbital coefficients $\mathbf{C}_{\rm occ}$ as

$$\mathbf{P} = 2\,\mathbf{C}_{\rm occ}\mathbf{C}_{\rm occ}^{\dagger} . \tag{4}$$

The molecular orbital matrix $\mathbf{C}$, including both occupied and virtual orbitals, is orthogonal;

$$\mathbf{C}^{\dagger}\mathbf{C} = \mathbf{I} . \tag{5}$$

The minimization of $E_{\rm el}$ with respect to $\mathbf{C}$, and with the constraint of orthogonality, leads to Roothaan's form of the SCF equations [55,54]

$$\mathbf{F}[\mathbf{P}^{n-1}]\mathbf{C}^{n} = \mathbf{C}^{n}\,\epsilon , \tag{6}$$

which is an eigenproblem with $\mathcal{O}(N_{\rm bas}^3)$ computational complexity. Iteration of Eq. (6) leads to the minimum energy solution at self-consistency, when $\mathbf{P}^{n} = \mathbf{P}^{n-1}$. For well behaved problems, typically $n \sim 10$ SCF cycles are required to reach convergence.

In the preceding matrix equations, an orthogonal basis has been tacitly assumed. In practice however, matrix elements are evaluated in the non-orthogonal basis of atomic orbitals (AOs). Introducing the metric or overlap matrix with elements

$$S_{ab} = (\phi_a, \phi_b) \tag{7}$$

between AO basis functions ϕ_A and ϕ_B, the Roothaan equations may be posed as the generalized eigenvalue problem [55]

$$\mathbf{F}_{\rm ao}\mathbf{C}_{\rm ao} = \mathbf{S}\,\mathbf{C}_{\rm ao}\,\epsilon , \tag{8}$$

where ao implies a non-orthogonal representation.

B. Orthogonal and non-orthogonal bases

It is well known that the generalized eigenproblem, Eq. (8), can be transformed to the standard eigenproblem, Eq. (6), using a congruence transformation involving factorization of the metric (overlap) matrix [56,57]. A key matrix in this transformation is $\mathbf{Z}$, the inverse factor, which relates non-orthogonal (AO) and orthogonal representations of the Fock matrix

$$\mathbf{F} = \mathbf{Z}\,\mathbf{F}_{\rm ao}\,\mathbf{Z}^{\mathbf{T}} \tag{9}$$

and the density matrix

$$\mathbf{P}_{\rm ao} = \mathbf{Z}^{T}\,\mathbf{P}\,\mathbf{Z} . \tag{10}$$

The inverse factor has the property

$$\mathbf{Z}^{T}\mathbf{S}\mathbf{Z} = \mathbf{I}, \tag{11}$$

which is satisfied by $\mathbf{Z} = \mathbf{S}^{-1/2}$ or $\mathbf{Z} = \mathbf{L}^{-1}$, where $\mathbf{L}$ is the Cholesky factor for which $\mathbf{S} = \mathbf{L}\mathbf{L}^{T}$. The choice $\mathbf{Z} = \mathbf{S}^{-1/2}$ corresponds to Löwdin's symmetric orthogonalization [58,59], which is commonly used in modern quantum chemistry. The choice $\mathbf{Z} = \mathbf{L}^{-1}$ is widely used in solution of the generalized eigenproblem [56,57], and has recently been introduced to linear scaling SCF theory by Millam and Scuseria (MS) [23].

Symmetric orthogonalization requires an eigensolve, which is $\mathcal{O}(N^3)$. As discussed by Millam and Scuseria (MS) [23], the incomplete Cholesky factorization is potentially $\mathcal{O}(N)$ when sparse matrix methods are employed in conjunction with an elimination tree. In the MS approach, the approximate inverse factor $\mathbf{Z}$ is obtained through an incomplete linear solve. This tends to be inaccurate unless tight thresholds are employed, in which case the solve may become expensive. Moreover, inverting the Cholesky factor introduces two levels of incompleteness (one in the factorization and one in the inversion), which may introduce errors that are difficult to control.

A different approach introduced here, is to solve for $\mathbf{Z} = \mathbf{L}^{-1}$ directly using the AINV algorithm [50,51]. AINV is state of the art in the theory and application of inverse preconditioners, and may be used to obtain the inverse factors $\mathbf{Z}$ to within arbitrary accuracy. The matrix $\mathbf{Z}$ is obtained by

S-orthogonalization of the standard basis vector, where sparsity is preserved with the use of a drop tolerance [50,51]. We have recently implemented a block version of AINV [60] and found it to be very fast under the demands of high accuracy.

When working in a non-orthogonal representation, the inverse overlap matrix $\mathbf{S}^{-1}$ enters the OM [25,26], DMM [61], and FOE [34] formulations of SCF theory. While $\mathbf{S}^{-1}$ is typically dense, its Cholesky factors $\mathbf{Z}$ may be quite sparse, depending on the ordering of $\mathbf{S}$ [62,63]. Thus, if the product $\mathbf{S}^{-1}\mathbf{X}$ is sparse, it is possible to efficiently apply the inverse as

$$\mathbf{Z}\left(\mathbf{Z}^T\mathbf{X}\right) = \mathbf{S}^{-1}\mathbf{X}\,, \tag{12}$$

without ever referencing a potentially dense $\mathbf{S}^{-1}$.

C. The LNVD objective

Beginning with the seminal work of Löwdin [53] and McWeeny [49], it has long been recognized that eigensolution of the SCF equations can be avoided by the direct minimization of $E_{\rm el}$ subject to the constraint of normalization

$$N_{\rm el} = \mathrm{Tr}\left\{\mathbf{P}\right\}\,, \tag{13}$$

and idempotency,

$$\mathbf{P} = \mathbf{PP}\,. \tag{14}$$

This latter condition is equivalent to requiring orthogonality of $\mathbf{C}_{\rm occ}$.

Variation of the electronic energy leads to the stationary condition [49,64]

$$\delta E_{\rm el} = 2\mathrm{Tr}\left\{\,\mathbf{F}\,\delta\mathbf{P}\,\right\} = 0\,, \tag{15}$$

allowing $E_{\rm el}$ to be replaced with the equivalent objective

$$\mathcal{E} = \mathrm{Tr}\left\{\mathbf{PF}\right\} \tag{16}$$

in which $\mathbf{F}$ no longer depends on $\mathbf{P}$.

A number of workers have formulated different versions of density matrix minimization [65–71], but these were not competitive with eigensolution due to the expense associated with enforcing Eq. (14), and the high efficiency of direct eigensolvers [72]. A breakthrough came with the density matrix minimization of Li, Nunes, Vanderbilt [17] and Daw [18] (LNVD), which imposes the constraint on idempotency implicitly through substitution of the McWeeny purification [49],

$$\mathbf{P} = 3\mathbf{PP} - 2\mathbf{PPP}\,, \tag{17}$$

into Eq. (16). The purification transform brings an approximately idempotent matrix closer to idempotency, a process that is quadratically convergent upon iteration [49]. With this substitution, Eq. (14) is imposed approximately during optimization, and may be used after convergence to restore idempotency to within the accuracy allowed by the underlying matrix algebra. In the LNVD density matrix minimization, normalization is imposed through a Lagrange multiplier μ, yielding the LNVD objective

$$\Omega = \mathrm{Tr}\left\{\left(3\mathbf{PP} - 2\mathbf{PPP}\right)\mathbf{F}\right\} + \mu\left(N_{\rm el} - \mathrm{Tr}\mathbf{P}\right)\,. \tag{18}$$

III. DENSITY MATRIX MINIMIZATION

In the LNVD method, a density matrix minimization is carried out after forming each new Fock matrix. To achieve an $\mathcal{O}(N)$ complexity, it is necessary to employ an approximate sparse matrix algebra and to forego methods that require storing a Hessian. One possibility is steepest descent (SD) in which the gradient,

$$\mathbf{G}_j = -\nabla_{\mathbf{P}_j}\Omega \tag{19}$$

is simply followed. A better alternative is the non-linear conjugate gradient (NLCG) method, which employs properties of the Hessian implicitly through a sequence of A-orthogonal search directions $\mathbf{H}_j$ [73–75]. In the non-linear conjugate gradient method, if an exact line search is performed, that is if a steplength λ_j is chosen such that the updated density matrix

$$\mathbf{P}_{j+1} = \mathbf{P}_j + \lambda\mathbf{H}_j \tag{20}$$

exactly minimizes the objective, then each gradient in the sequence will be orthogonal to the others [76];

$$(\mathbf{G}_k, \mathbf{G}_j) = 0 \quad (k \neq j)\,. \tag{21}$$

If conjugacy is preserved, the non-linear conjugate gradient method is in general much more efficient than steepest descent. In practice, incomplete matrix algebra, deviation from quadratic behavior, and inexact line searches can spoil this property [75–77], leading to an algorithm that may be less efficient than steepest descent. Note that modifying the gradient or objective during optimization, through a purification step or extrapolation, will tend to destroy conjugacy.

A. Gradient and line search

Using the trace algebra [78], the gradient of Ω is

$$\begin{aligned}\nabla\Omega = {} & 3\left(\mathbf{PF} + \mathbf{FP}\right) \\ & -2\left(\mathbf{PPF} + \mathbf{PFP} + \mathbf{FPP}\right) - \mu\mathbf{I}\,,\end{aligned} \tag{22}$$

where it has been assumed that μ does not depend on $\mathbf{P}$. The Lagrange multiplier μ is chosen as

$$\begin{aligned}\mu = {} & \mathrm{Tr}\left\{3\left(\mathbf{PF} + \mathbf{FP}\right)\right. \\ & \left.-2\left(\mathbf{PPF} + \mathbf{PFP} + \mathbf{FPP}\right)\right\}/N_{\rm bas}\,,\end{aligned} \tag{23}$$

which renders the gradient traceless at each step [20,23]. With each step an analytic line search is carried out as in Refs. [20,23] by solving

$$\frac{d\Omega[\mathbf{P} + \lambda\mathbf{H}]}{d\lambda} = b + 2c\lambda + 3d\lambda^2 = 0 \tag{24}$$

for the minimum root, where

$$b = \mathrm{Tr}\left\{\mathbf{H}\nabla\Omega\right\} \tag{25}$$

$$c = \mathrm{Tr}\left\{3\mathbf{HHF} - 2(\mathbf{HHPF} + \mathbf{HPHF} + \mathbf{PHHF})\right\} \tag{26}$$

$$d = -2\mathrm{Tr}\{\mathbf{HHHF}\}\,. \tag{27}$$

B. A simplifying guess

At the start of each minimization, it is necessary to choose a guess density matrix. A natural choice is the density matrix from the previous SCF cycle. However, this guess leads to very slow convergence, typically requiring on the order of 50-100 non-linear conjugate gradient cycles.

One method to accelerate convergence is to precondition with a diagonal Hessian [74,75,79], which has been used by Millam and Scuseria [23]. Another approach employed by Millam and Scuseria is to use direct inversion in the iterative subspace (DIIS) [80] to minimize the commutator $[\mathbf{F}^i, \mathbf{P}_j^{i+1}]$ between non-linear conjugate gradient cycles (DIIS-CG), where superscripts have been used to emphasize the SCF iteration, and j is the CG cycle. Note that $[\mathbf{F}^i, \mathbf{P}^i]$ will be zero only when $\mathbf{P}^i = \mathbf{P}^{i+1}$.

An alternative guess for the start of a density matrix minimization is

$$\mathbf{P}_0 = \left(\frac{N_{\rm el}}{N_{\rm bas}}\right)\mathbf{I}\,, \tag{28}$$

which preserves symmetry of the converged result, namely

$$[\mathbf{P}, \mathbf{F}] = [\mathbf{H}, \mathbf{F}] = 0\,, \tag{29}$$

throughout a non-linear conjugate gradient or steepest descent sequence.

Starting with Eq. (28) it is found that after some characteristic number of conjugate gradient cycles, $N_{\rm cg} \sim 3$, purification will establish the exact answer, even though the gradient may still be large. If too few steps are taken, purification will lead to the loss of electrons. As $N_{\rm cg}$ increases, fewer purification steps $N_{\rm pur}$ are needed to obtain the correct result. In every case, it is found possible to cast the dominant portion of the work into the purification step, which is quadratically convergent.

In addition to enhanced convergence properties, Eq. (29) leads to a significant reduction in complexity of the gradient evaluation and the line search. In particular, Eqs. (22) and (23) reduce to

$$\nabla\Omega = 6\,(\mathbf{I}-\mathbf{P})\,\mathbf{P}\mathbf{F} - \mu\mathbf{I} \tag{30}$$

and

$$\mu = 6{\rm Tr}\left\{(\mathbf{I}-\mathbf{P})\,\mathbf{P}\mathbf{F}\right\}/N_{\rm bas}\,, \tag{31}$$

while Eq. (26) simplifies to

$$c = 3{\rm Tr}\left\{(\mathbf{I}-2\mathbf{P})\,\mathbf{H}\mathbf{H}\mathbf{F}\right\}\,. \tag{32}$$

Inspection of Eq. (30) shows that during minimization, $\mathbf{P}$ will start with the sparsity pattern of $\mathbf{F}$, and will broaden with additional iterations. In practice, this results in initial cycles that are quite fast.

Note that preconditioning will destroy the commutation property; that is, if preconditioning is used then Eqs. (30), (31), and (32) will become invalid.

C. Non-orthogonal density matrix minimization

Just as with eigensolution, it is possible to cast the LNVD method in a non-orthogonal basis. There are at least two ways to perform the density matrix minimization in a non-orthogonal basis: One way is to simply follow the gradient of the non-orthogonal objective [19,21,22]

$$\Omega_{\rm ao} = {\rm Tr}\left\{(3\mathbf{P}_{\rm ao}\mathbf{S}\mathbf{P}_{\rm ao} - 2\mathbf{P}_{\rm ao}\mathbf{S}\mathbf{P}_{\rm ao}\mathbf{S}\mathbf{P}_{\rm ao})\,\mathbf{F}_{\rm ao}\right\} - \mu N_{\rm el}\,. \tag{33}$$

Another approach is to follow the gradient of the orthogonal objective, Eq. (18), but in a non-orthogonal basis as suggested by White *et al.* [61];

$$(\nabla\Omega)_{\rm ao} = \mathbf{S}^{-1/2}\,[\nabla\Omega]\,\mathbf{S}^{-1/2}\,. \tag{34}$$

This latter approach is independent of basis set transformations, and may have improved convergence properties over the former approach. One disadvantage is that the inverse overlap matrix $\mathbf{S}^{-1}$ must be applied at every iteration, and it is no longer possible to use the non-linear conjugate gradient without a back transformation as the gradient corresponds to the orthogonal objective, Eq. (18), rather than the non-orthogonal one, Eq. (33). On the other hand, $\mathbf{F}_{\rm ao}$ tends to be much sparser than $\mathbf{F}$, and it may be that working in the AO results in faster gradient evaluations and line minimizations.

Starting with the guess,

$$\mathbf{P}_{\rm ao,0} = \left(\frac{N_{\rm el}}{N_{\rm bas}}\right)\mathbf{S}^{-1} \tag{35}$$

it is possible to achieve simplification and improved convergence properties in the AO basis that are equivalent to those of the orthogonal simplified density matrix minimization developed in Section III B. With this guess, Eq. (34) reduces to

$$(\nabla\Omega)_{\rm ao} = 6\,\mathbf{S}^{-1}\left[(\mathbf{I}-\mathbf{S}\mathbf{P}_{\rm ao})\mathbf{F}_{\rm ao}\mathbf{P}_{\rm ao} - \mu\mathbf{I}\right]\,. \tag{36}$$

with

$$\mu = {\rm Tr}\left\{(\mathbf{I}-\mathbf{S}\mathbf{P}_{\rm ao})\mathbf{F}_{\rm ao}\mathbf{P}_{\rm ao}\right\}/N_{\mathbf{bas}}\,. \tag{37}$$

Rather than compute and manipulate $\mathbf{S}^{-1}$, it is much more efficient to apply the AINV factors as in Eq. (12). As before, an analytic line search is carried out. However, this time the non-orthogonal objective is used, yielding the coefficients

$$b = 6\,{\rm Tr}\{(\mathbf{I}-\mathbf{P}_{\rm ao}\mathbf{S})\mathbf{H}\mathbf{F}_{\rm ao}\} \tag{38}$$

$$c = 3\,{\rm Tr}\{(\mathbf{I}-2\mathbf{P}_{\rm ao}\mathbf{S})\mathbf{H}\mathbf{S}\mathbf{H}\mathbf{F}_{\rm ao}\} \tag{39}$$

$$d = -2{\rm Tr}\{\mathbf{H}\mathbf{S}\mathbf{H}\mathbf{S}\mathbf{H}\mathbf{F}_{\rm ao}\}\,. \tag{40}$$

IV. IMPLEMENTATION

The orthogonal and non-orthogonal simplified density matrix minimization have been implemented in the MondoSCF [52] suite of programs for linear scaling SCF theory, with details as follows.

Atom-blocked sparse matrix algebra A library of routines for carrying out variable block size, sparse matrix algebra have been developed and employed in the implementation of MondoSCF. These routines exploit both existing sparsity and data locality. Related methods have been shown to yield significant speedups for matrix-vector multiplies [81–83]. Note that in Ref. [82], Goedecker *et al.* employ an atom-blocked algebra for matrix-vector multiplies in the context of the tight binding FOE method.

Matrix multiplies In the sparse atom-blocked matrix multiply, the matrices are stored in a modified compressed sparse row (CSR) format [84–86], in which all pointers index blocks rather than individual matrix elements. An additional pointer is required to address the start of each block since the blocks are not grouped by dimension. A modified version of Gustavson's sparse matrix multiply [87] is employed, with DGEMM-like routines to carry out the block-block multiplies. DGEMM is a level three BLAS [72] routine for performing double precision matrix multiplies. In the limit that the block size is reduced to 1×1, an algorithm equivalent to the standard [87] sparse matrix multiply is obtained.

Truncation Small matrix elements are dropped after each matrix operation such as multiplication or addition. Blocks are dropped when they meet the criterion

$$\| \mathbf{C}_{IK} \|_F < \texttt{TrixtNeglect}\,, \tag{41}$$

where `TrixtNeglect` is a drop tolerance and F demarks the Frobenius norm [88]

$$\| \mathbf{C}_{IK} \|_F = \sqrt{\mathrm{Tr}\,\{\mathbf{C}_{IK}^T \mathbf{C}_{IK}\}} \tag{42}$$

This approach should be compared with truncation of individual elements [23,89], and separation-based truncation [17,22]. With truncation at the level of elements, a change of basis may lead to very different behavior of errors. This is because the average magnitude of the matrix elements must decrease with increasing basis set size [40]. Truncation at the level of atoms is expected to yield errors that are more transferable, as the amount of charge in a block should be less sensitive to a change of basis.

In separation-based truncation, elements of a matrix product are computed only if the corresponding basis function separation is less than a cut-off radius R_c. This has the advantage that the structure of all matrices is known *a priori*. On the other hand, this approach requires a careful system, basis set, and model chemistry dependent parameterization. Also, when examining the fall off of density matrix elements with basis function separation [14], it is seen that many small elements correspond to small separations. This suggests that, for a given level of error, truncation schemes based on magnitude may yield sparser matrices.

SPAMM A method related to thresholding schemes used in the order N exchange method ONX [8,14] has been extended to matrix multiplication, and may be well suited for matrices with elements that decay away from the diagonal. In this sparse approximate matrix multiply (SPAMM), small matrix elements are treated in a more approximate manner than are the large ones. For the matrix multiply $\mathbf{C} = \mathbf{AB}$, each block $\mathbf{C}_{IK}$ is formed as the sum over blocks

$$\mathbf{C}_{IK} = \sum_{J=1}^{M} \mathbf{A}_{IJ} \mathbf{B}_{JK}\,. \tag{43}$$

The error

$$\| \,\delta\, \mathbf{C}_{IK} \|_F < \texttt{MultNeglect} \tag{44}$$

may be rigorously bounded while avoiding all contractions $\mathbf{A}_{IJ}\mathbf{B}_{JK}$ for which

$$\| \mathbf{B}_{JK} \|_F < \frac{\texttt{MultNeglect}}{M \, \| \mathbf{A}_{IJ} \|_F}\,, \tag{45}$$

where M is the number of blocks in the contraction, and `MultNeglect` is a threshold. Just as in the case of ONX, repeatedly executing a conditional like Eq. (45) in an innermost loop can create a significant overhead. This is avoided in SPAMM by sorting each row in decreasing order on the basis of $\|.\|_F$, and using a binary search to find the limit ($\leq M$) of the innermost loop.

Trace A substantial savings may be achieved in the execution of $\mathrm{Tr}\{.\}$ by performing both the contraction and the diagonal accumulation at the same time as

$$\mathrm{Tr}\,\{\mathbf{AB}\} = \sum_{ik} A_{ik} B_{ki}\,, \tag{46}$$

rather than first forming the product $\mathbf{AB}$.

AINV MondoSCF employs a blocked version of AINV, which is described in detail elsewhere [60].

A. SCF

Initial Guess For a minimal basis such as STO-3G, a sparse linear scaling guess is built from the superposition of spherically averaged atomic density matrices in an orthogonal (diagonal) representation. The AINV factor $\mathbf{Z}$ is then used to transform to an AO representation [90]. For more complicated basis sets, a partially converged density matrix from a minimal basis set calculation is used to construct a mixed basis Fock matrix [90]. These guesses are linear scaling.

Eigensolver The Roothaan equations, Eq. (6), are solved with the LAPACK divide and conquer eigensolver DSYEVD [72] as implemented in the SGI scientific library SCSL. This a competitive algorithm that has been highly optimized for the SGI platform.

Fock builds The multipole accelerated symmetrized order N exchange algorithm (MASONX) [16] is used to compute the exchange matrix, and the quantum chemical tree code (QCTC) [13] is used to compute the Coulomb matrix. Tight integral ($\texttt{TwoENeglect} = 10^{-9}$) and distribution ($\texttt{DistNeglect} = 10^{-11}$) thresholds are used throughout.

DIIS A sparse implementation of DIIS for extrapolation of the Fock matrix [91] is used in all calculations. In this implementation, the error vector

$$\mathbf{e}_{\mathrm{ao}} = \mathbf{F}_{\mathrm{ao}} \mathbf{P}_{\mathrm{ao}} \mathbf{S} - \mathbf{S} \mathbf{P}_{\mathrm{ao}} \mathbf{F}_{\mathrm{ao}} \tag{47}$$

is transformed into an orthonormal basis with the transformation

$$\mathbf{e} = \mathbf{Z}^T \mathbf{e}_{\mathrm{ao}} \mathbf{Z}\,. \tag{48}$$

The use of sparse matrix algebra will eventually cause the DIIS method to stagnate. Tightening up the matrix thresholds with SCF convergence leads to premature stagnation that is incommensurate with the thresholds. One solution is to simply restart the DIIS after each decrement of a threshold, and this should lead to significantly reduced average CPU times per SCF cycle. For clarity and simplicity, this approach has not been followed here; rather, constant thresholds are used throughout.

Non-linear conjugate gradient The Polak-Ribière version of non-linear conjugate gradient [92] has been implemented for the orthogonal minimization. As discussed in Section III, the non-quadratic behavior of the objective, approximate arithmetic, and deviation of the gradient from the objective due to the normalization constraint may destroy conjugacy as measured by

$$\alpha = \frac{|(\mathbf{G}_j, \mathbf{G}_{j+1})|}{(\mathbf{G}_{j+1}, \mathbf{G}_{j+1})}, \tag{49}$$

which is ideally 0. While the Polak-Ribière non-linear conjugate gradient tends to restart along $\mathbf{G}_{j+1}$ when conjugacy is lost, Powell [93] suggests a complete steepest descent restart when α exceeds 0.2. In the calculations performed here, this condition typically occurs only for large values of $N_{\rm cg}$, and no advantage has been found in performing a steepest descent restart. However, for systems with a small HOMO-LUMO gap this may yield an advantage.

Overview of the orthogonal method After each Fock build, DIIS extrapolation of the Fock matrix is used, and the extrapolated Fock matrix is transformed to an orthogonal basis. A diagonal guess is used to start, and $N_{\rm cg}$ non-linear conjugate gradient steps are taken using the simplified gradient and line minimization particular to the SDMM. Then, the resulting density matrix is brought to idempotency through purification. The degree of idempotency achieved depends on the matrix thresholds `TrixNeglect` and `MultNeglect`. Stagnation manifests itself in an inability to further reduce the maximum block of the difference density $P_{I_{\rm pur}} - P_{I_{\rm pur}+1}$, and is said to occur after $N_{\rm pur}$ iterations. Thus, $N_{\rm pur}$ is a function of $N_{\rm cg}$ and the matrix thresholds. The resulting density matrix is then transformed back to a non-orthogonal atomic orbital representation and used to construct a new Fock matrix. Stagnation of the DIIS procedure also occurs with more approximate matrix thresholds, and the SCF procedure is halted after the DIIS error fails to improve.

V. RESULTS AND DISCUSSION

All computations were carried out on ASCI Bluemountain, which is a large collection of 195 MHz SGI Origin 2000s. The calculations were all performed in serial and in a shared memory (non-dedicated) environment. Atomic units are used throughout.

A. Benchmark systems

Two three-dimensional benchmark suites are used in this study: a sequence of water clusters and a sequence of estane polymers. Linear systems of polyglycine chains, alkanes, and nucleic acids yield overly optimistic results with respect to both computational complexity and error control, and are not considered here.

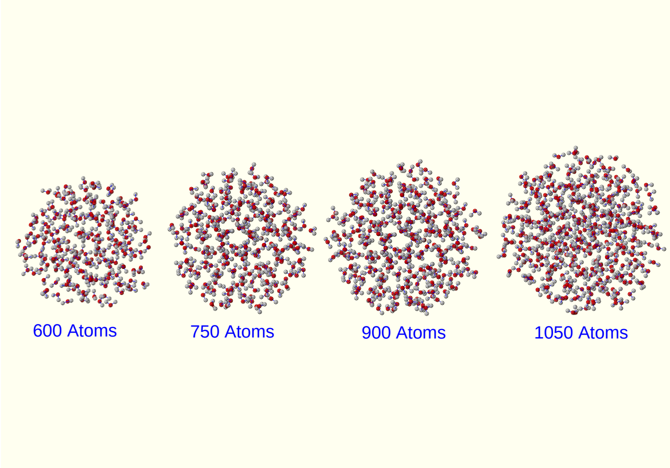

FIG. 1. Configuration of the clusters with 200, 250, 300 and 350 water molecules.

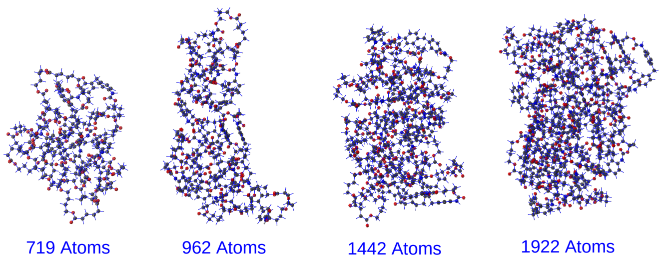

FIG. 2. Local minimum energy configurations of the estane polymers with 3, 4, 6, and 8 segments, corresponding to 719, 962, 1442 and 1922 atoms respectively.

The sequence of clusters includes 50, 70, 90, 110, 150, 200, 250, 300, and 350 water molecules, with a constant density corresponding to standard temperature and pressure. The four largest clusters are shown in Fig. 1. These water clusters have been used in a number of other studies [7,6,13,10,14,23,89,15,16], and have proven a challenge to achieving linear scaling in computation of the Coulomb matrix [7,6,10], the exchange matrix [10], and the density matrix [23,89].

Estane 5703 polyester urethane is a segmented copolymer which is a constituent in the plastic-bonded high explosive PBX-9501 [94]. A fundamental repeat unit, $N_{\rm seg}$ =1, is a length of chain comprised of two 4,4'-diphenylmethane diisocyanate and 1,4-butanediol segments bonded to five poly tetramethylene adipate segments. Configurations for degrees of polymerization corresponding to 1, 2, 3, 4, 6, and 8 (242, 482, 719, 962, 1442, and 1922 atoms, respectively) have been generated by molecular dynamics equilibration followed by energy minimization. The four largest configurations are shown in Fig. 2.

B. Scaling

Orthogonal and non-orthogonal simplified density matrix minimization In Fig. 3, the scaling of the non-orthogonal

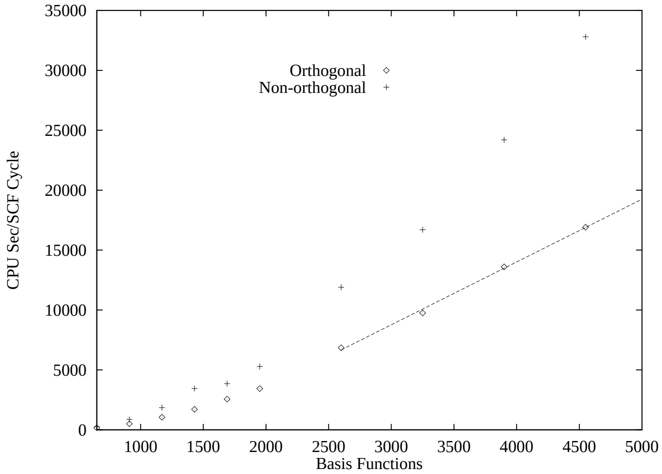

FIG. 3. Scaling of CPU time per SCF for the restricted Hartree-Fock (RHF)' 3-21G sequence of waters clusters using the non-orthogonal and orthogonal simplified density matrix minimization.

and orthogonal simplified density matrix minimization are shown for the sequence of water clusters at the RHF/3-21G level of theory. In both cases, the thresholds `TrixNeglect` $= 10^{-6}$ and `MultNeglect` $= 10^{-8}$ were used. Seven minimization steps were taken in both the steepest descent non-orthogonal and the non-linear conjugate gradient orthogonal case. The non-orthogonal simplified density matrix minimization is much slower than the orthogonal one, and does not appear to achieve linear scaling.

While $\mathbf{F}$ is denser than $\mathbf{F}_{\rm ao}$, the interspersal of $\mathbf{S}$ in the non-orthogonal expressions leads to intermediate matrices, *e.g.* $\mathbf{SF}_{\rm ao}$, with an increased density. In addition, the non-orthogonal simplified density matrix minimization requires many more matrix multiplies than does the orthogonal version.

TABLE I. HOMO and LUMO energies and the HOMO-LUMO gap for a cluster of 30 waters. All values were obtained with GAUSSIAN 94 [95].

Theory	Basis Set	HOMO	LUMO	$E_{\rm gap}$
HF	STO-3G	-.318	0.497	0.815
HF	3-21G	-.393	0.164	0.557
HF	6-31G**	-.423	0.126	0.549
LDA	STO-3G	0.005	0.200	0.195
LDA	3-21G	-.133	-.039	0.094
LDA	6-31G**	-.190	-.061	0.129
B3LYP	3-21G	-.178	0.008	0.186
B3LYP	6-31G**	-.226	-.013	0.213

Water The scaling of the orthogonal simplified density

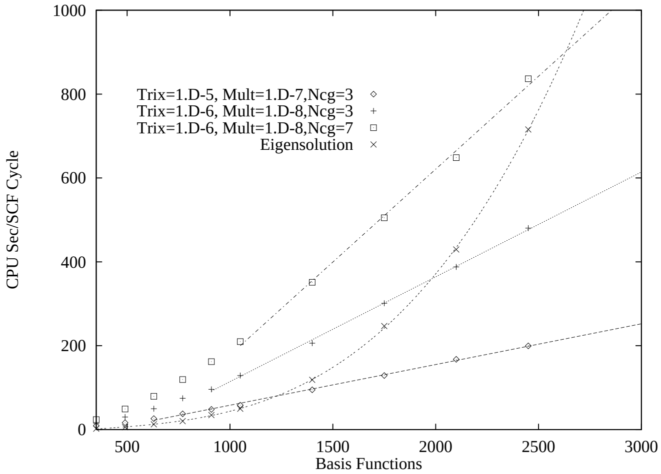

FIG. 4. Scaling of CPU time per SCF for the RHF/STO-3G sequence of water clusters.

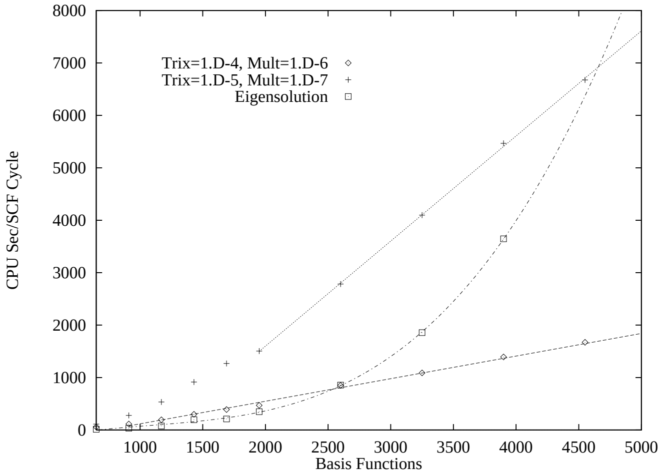

FIG. 5. Scaling of CPU time per SCF for the RHF/3-21G sequence of water clusters.

matrix minimization is shown in Fig. 4 for the water clusters at the RHF/STO-3G level of theory. A factor of two reduction in CPU time is observed on going from $N_{\rm cg} = 7$ to $N_{\rm cg} = 3$ for the `TrixNeglect` $= 10^{-6}$ calculations. A similar gain is found (but not shown) for the `TrixNeglect` $= 10^{-5}$ calculation.

These results should be compared with Fig. 5 from Ref. [23], which shows the inability of the CG-DMS algorithm to achieve linear scaling for LDA/STO-3G calculations on the same set of water clusters. In that work, the onset of linear scaling is predicted to occur at about 400 water molecules. Here, the onset of linear scaling is observed at 90 water molecules for the `TrixNeglect` $= 10^{-5}$ calculation, and at 130 water molecules for the `TrixNeglect` $= 10^{-6}$ calculations. In addition to differences in the algorithms and implementation, it is known that the LDA tends to underestimate the HOMO-LUMO gap, while HF tends to overestimate it. This effect is shown in Table I. As the asymptotic behavior of the density matrix is governed by $\sqrt{E_{\rm gap}}$ as in Eq. (1), this phenomenon may explain some of the differences between this work and Ref. [23].

The scaling of the orthogonal simplified density matrix

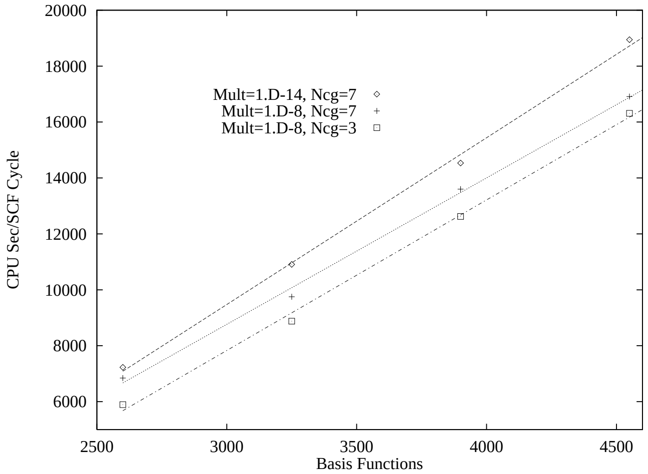

FIG. 6. Scaling of CPU time per SCF for the RHF/3-21G sequence of water clusters with TrixNeglect $= 10^{-6}$.

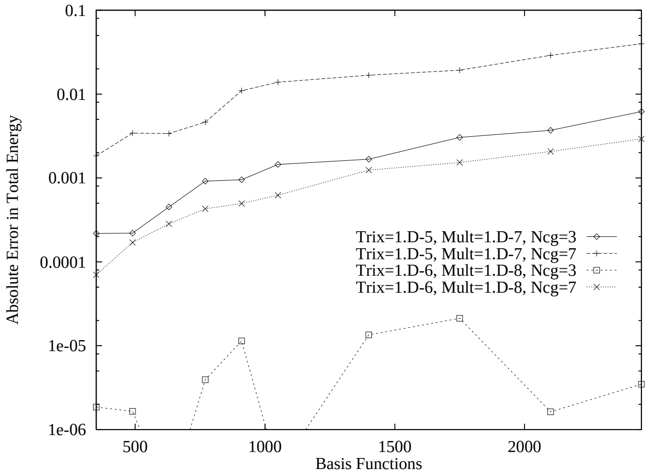

FIG. 8. Absolute errors in converged total energies for the RHF/STO-3G sequence of water clusters.

minimization with $N_{\rm cg} = 3$ is shown in Figs. 5 and 6 for the water series at the RHF/3-21G level of theory. For the 3-21G calculations, the differences in CPU time between $N_{\rm cg} = 3$ and $N_{\rm cg} = 7$ are much less pronounced than in the case of the STO-3G calculations. This is shown in Fig. 6 for TrixNeglect $= 10^{-6}$. A similar behavior is also found for looser values.

The onset of linear scaling is slower for the 3-21G calculations than for the STO-3G calculations, occurring at 90, 150, and 200 water molecules for TrixNeglect $= 10^{-4}, 10^{-5}$ and 10^{-6} respectively. Also, the STO-3G density matrix is less dense than the 3-21G density matrix. These differences are most likely because the STO-3G band gap is larger than the 3-21G band gap, as shown in Table I.

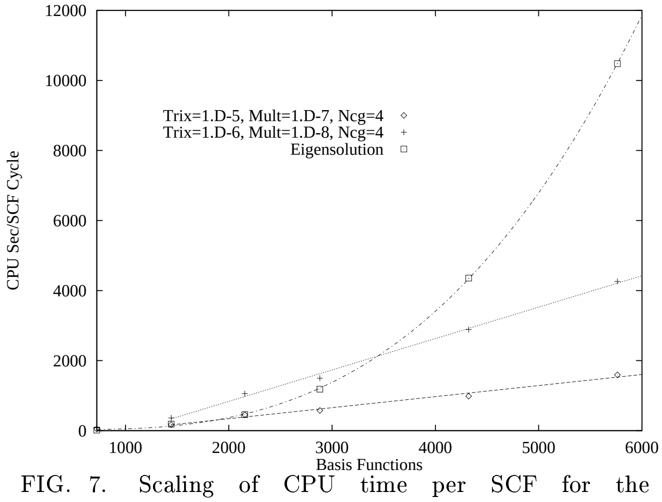

FIG. 7. Scaling of CPU time per SCF for the RHF/STO-3G sequence of estane polymers.

Estane The scaling of the orthogonal simplified density matrix minimization with $N_{\rm cg} = 4$ is shown in Fig. 7 for the estane polymers at the RHF/STO-3G level of theory. As in the water calculations, the $N_{\rm cg} = 4$ results are about a factor of two faster than those obtained with $N_{\rm cg} = 8$ (not shown). In all cases, the onset of linear scaling is seen to occur at the $N_{\rm seg} = 2$ polymer, which corresponds to 482 atoms, and 722

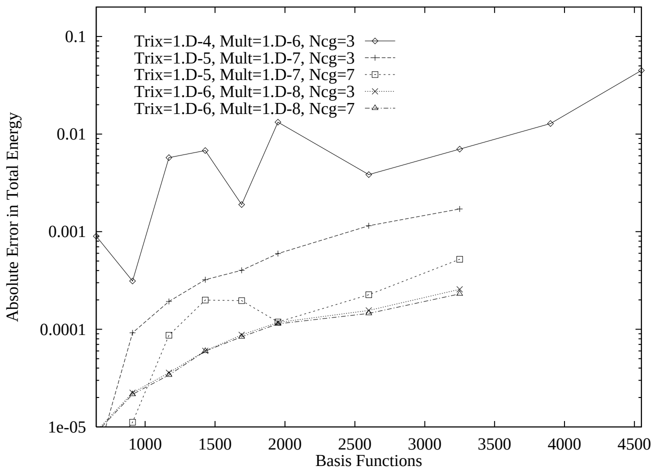

FIG. 9. Absolute errors in converged total energies for the RHF/3-21G sequence of water clusters.

basis functions.

C. Errors

Errors in converged total energies are shown in Fig. 8 for the RHF/STO-3G water series, in Fig. 9 for the RHF/3-21G water series, and in Fig. 10 for the RHF/STO-3G Estane sequence. A striking feature of the water errors is that the calculations with $N_{\rm cg} = 3$ yield more accurate results than the calculations with $N_{\rm cg} = 7$.

One explanation for this behavior is that, with approximate algebra, $[\mathbf{F}, \mathbf{P}]$ and $[\mathbf{F}, \mathbf{H}]$ are not exactly zero. With each additional step these commutators grow, leading to errors in the gradient and line search. As the matrix thresholds are tightened, results for different values of N_{cg} will become identical; this can be seen for the TrixNeglect $= 10^{-6}$ calculations in Fig. 9.

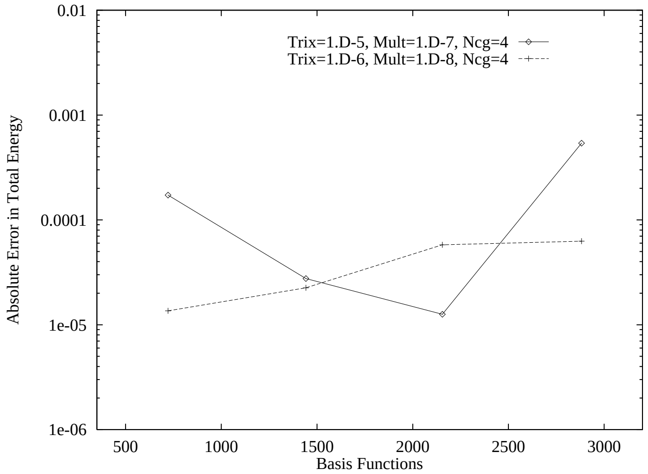

FIG. 10. Absolute errors in converged total energies for the RHF/STO-3G sequence of estane polymers.

TABLE II. Sustained MFLOPS rates as a function of different systems and basis sets. N_H/N_X is the ratio of hydrogen atoms to heavy atoms.

			MFLOPS		
System	N_H/N_X	Basis Set	GradE	LineM	Purify
Water	2	STO-3G	71	57	81
Estane	1	STO-3G	100	83	104
Fullerene	0	STO-3G	162	136	158
Water	2	3-21G	137	125	140
Estane	1	3-21G	154	135	150
Fullerene	0	3-21G	163	145	157

D. Performance of the matrix multiply

MFLOPS Mega floating point operations per second (MFLOPS) for gradient evaluation (GradE), line minimization (LineM), and purification (Purify) are given in Table II for water, estane and fullerene for the STO-3G and 3-21G basis sets. These are sustained values, and while dominated by multiplication, they include additions, trace, diagonal adds, matrix-scalar multiplication and IO. For reference, Warner's [96] optimized DGEMM using 50×50 sub-blocks achieves a peak rate of approximately 270 MFLOPS on the 195 MHz SGI Origin 2000.

The MFLOPS rate for line minimization is less than for purification or gradient evaluation. This is because cache and integer overheads are higher per FLOP for the evaluation of Tr{.} with Eq. (46) than for the matrix multiply and addition.

As the ratio of hydrogens to heavy atoms N_H/N_X decreases, the MFLOPS rate increases. This is because the number of basis functions per atom is smaller (1 for STO-NG and 2 for 3-21G) for hydrogen than for the heavy atoms (5 for STO-NG and 9 for 3-21G), and computations involving small dimensions are inefficient. This effect is less pronounced for the 3-21G basis set because the CPU time in this case is dominated by the $9 \times 9 \times 9$ DGEMM corresponding to the all heavy atom multiplies. As larger basis sets are used, the influence of N_H/N_X will become less significant, and the peak rate will be determined by DGEMM efficiency, which is limited by cache effects and is block size dependent [96,97].

TABLE III. The increase in CPU time per orthogonal non-linear conjugate gradient cycle for the 3-21G 350 water calculation with `TrixNeglect` $= 10^{-7}$ and `MultNeglect` $= 10^{-5}$.

$I_{\rm CG}$	CPU sec
0	407
1	505
2	714
3	780
4	749
5	905
6	828

SPAMM The choice

$$\texttt{MultNeglect} = 10^{-2}\,\texttt{TrixNeglect} \tag{50}$$

has been found to yield results that are very close to those obtained with `MultNeglect` $= 0$, and this parameterization has been used throughout. In Fig. 6, results are shown using SPAMM with this default parameterization and with `MultNeglect` $= 10^{-14}$. The effects of SPAMM here and in the other systems studied are small ($< 10\%$). Nevertheless, it may be more effective in systems which are not as well localized as those studied here.

TABLE IV. Decrease in the number of purification cycles $N_{\rm pur}$ on going from $N_{\rm cg} = 3$ to $N_{\rm cg} = 7$ for the 3-21G 350 water calculation with `TrixNeglect` $= 10^{-5}$ and `MultNeglect` $= 10^{-7}$.

	Max $\|\|\mathbf{P}_{IJ}\|\|_F$	
$I_{\rm pur}$	$N_{\rm cg} = 3$	$N_{\rm cg} = 7$
1	0.215D+00	0.147D+0
2	0.154D+00	0.760D-01
3	0.979D-01	0.434D-01
4	0.662D-01	0.227D-01
5	0.579D-01	0.702D-02
6	0.411D-01	0.793D-03
7	0.282D-01	0.327D-04
8	0.175D-01	0.328D-04
9	0.813D-02	
10	0.182D-02	
11	0.954D-04	
12	0.341D-04	
13	0.330D-04	

Non-linear conjugate gradient vs purification cycles As mentioned in Section III B, the CPU time per non-linear conjugate gradient step grows with the number of steps taken. This is shown in Table III for the 3-21G 350 water calculation. Note that the first cycles cost is half as much as the last cycle. The reduction in the number of purification steps required to reach stagnation is shown in Table IV for $N_{\rm cg} = 3$ and $N_{\rm cg} = 7$. Stagnation occurs when Max $||\mathbf{P}_{IJ}||_F$ fails to decrease with continued iteration due to the approximate matrix algebra.

VI. CONCLUSIONS

A simplified density matrix minimization (SDMM) has been introduced and shown to yield linear scaling CPU times that are highly competitive with a state-of-the-art dense eigensolver. This work is the first demonstration of linear scaling HF theory for three-dimensional systems, and one of only a few that consider double zeta basis sets or the behavior of errors with increasing system size.

The orthogonal simplified density matrix minimization uses 4 fewer matrix multiplies than conventional implementations of the LNVD density matrix minimization, and requires no transpositions. In addition, it is found that only a few minimization steps are required, and that the majority of the work can be shifted to the quadratically convergent McWeeny purification. For more approximate matrix thresholds, taking fewer non-linear conjugate gradient steps is found to yield both faster CPU times and more accurate results.

The non-orthogonal simplified density matrix minimization turns out to be non-competitive with the orthogonal version,and is unable to achieve linear scaling.

The AINV algorithm for computation of the inverse Cholesky factor $\mathbf{Z}$ has been introduced to linear scaling electronic structure theory, and found to be essential for working with non-orthogonal basis sets. In particular, $\mathbf{Z}$ is used in the transformation to and from an orthogonal representation, formation of a linear scaling initial guess, application of the inverse overlap, and in DIIS extrapolation.

Advantages of the AINV algorithm relative to iterative, incomplete inversion of the Cholesky factor include accuracy, memory, and speed. While in the context of a single point SCF calculation, it may be argued that the cost of inversion is small relative to the expense of ~ 10 SCF iterations, it should be pointed out that in a dynamics calculation or in a orbital minimization (OM) [25–29] scheme, the relative cost of inversion may be much larger. In addition, to achieve high accuracy in iterative inversion potentially requires $\mathcal{O}(N^2)$ memory as the sparsity pattern cannot be determined *a priori*. Also, because the computation of $\mathbf{Z}$ reduces to performing matrix multiplies rather than triangular solves (as in the case of computing $\mathbf{L}$), the AINV algorithm is ideally suited for parallel implementations [63].

A sparse atom-blocked matrix algebra was introduced and found to yield sustained performances of up to 160 MFLOPS, which is more than half of the 270 MFLOPS possible with an optimized DGEMM. Moreover, these methods should lead to very efficient parallel implementations, as sparse matrix multiplication is efficient in parallel, and blocking will improve this efficiency by reducing the integer overhead associated with decomposition and reordering.

Linear scaling Hartree-Fock theory for insulating three-dimensional systems was established for the first time in Refs [13,14] for the Fock build, and herein for solution of the SCF equations. While achieving linear scaling may be more difficult for model chemistries that yield a less inflated HOMO-LUMO gap, with parallel implementation and code maturation the outlook for large scale application of the simplified density matrix minimization is bright.

VII. ACKNOWLEDGEMENTS

The ASCI project at LANL is gratefully acknowledged for support of this work. The Advanced Computing Laboratory of Los Alamos National Laboratory, Los Alamos, NM 87545 is acknowledged. This work was performed on computing resources located at this facility.

Thanks to Michele Benzi for help with AINV, and for many enlightening conversations. Thomas D. Sewell is thanked for furnishing the estane coordinates. Thanks also to Yousef Saad for suggesting the use of a sparse approximate inverse in non-orthogonal minimization.

---

[1] D. L. Strout and G. E. Scuseria, J. Chem. Phys. **102**, 8448 (1995).
[2] J. Almlöf, K. Faegri, and K. Korsell, J. Comp. Chem. **3**, 385 (1982).
[3] M. Häser and R. Ahlrichs, J. Comp. Chem. **10**, 104 (1989).
[4] C. A. White, B. Johnson, P. Gill, and M. Head-Gordon, Chem. Phys. Letters **230**, 8 (1994).
[5] C. A. White, B. G. Johnson, P. M. W. Gill, and M. Head-Gordon, Chem. Phys. Lett. **253**, 268 (1996).
[6] M. Challacombe, E. Schwegler, and J. Almlöf, J. Chem. Phys. **104**, 4685 (1996).
[7] M. Challacombe, E. Schwegler, and J. Almlöf, in *Computational Chemistry: Review of Current Trends*, edited by J. Leczszynski (World Scientific, Singapore, 1996), pp. 53–107.
[8] E. Schwegler and M. Challacombe, J. Chem. Phys. **105**, 2726 (1996).
[9] J. P. Dombroski, S. W. Taylor, and P. M. W. Gill, J. Phys. Chem. **100**, 6272 (1996).
[10] J. C. Burant, R. E. Stratmann, and M. J. Frisch, J. Chem. Phys. **105**, 8969 (1996).
[11] R. E. Stratmann, G. E. Scuseria, and M. J. Frisch, Chem. Phys. Let. **257**, 213 (1996).
[12] M. C. Strain, G. E. Scuseria, and M. J. Frisch, Science **271**, 51 (1996).
[13] M. Challacombe and E. Schwegler, J. Chem. Phys. **106**, 5526 (1997).
[14] E. Schwelger, M. Challacombe, and M. Head-Gordon, J. Chem. Phys. **106**, 9708 (1997).
[15] C. Ochsenfeld, C. A. White, and M. Head-Gordon, J. Chem. Phys. **109**, 1663 (1998).
[16] E. Schwegler and M. Challacombe, manuscript in preparation.
[17] X. P. Li, R. W. Nunes, and D. Vanderbilt, Phys. Rev. B **47**, 10891 (1993).
[18] M. S. Daw, Phys. Rev. B **47**, 10895 (1993).
[19] R. W. Nunes and D. Vanderbilt, Phys. Rev. B **50**, 17611 (1994).
[20] S. Y. Qiu, C. Z. Wang, K. M. Ho, and C. T. Chan, J. Phys. Cond. Matt. **6**, 9153 (1994).

[21] E. Hernández and M. J. Gillan, Phys. Rev. B **51**, 10157 (1995).
[22] E. Hernández, M. J. Gillan, and C. Goringe, Phys. Rev. B **53**, 7147 (1996).
[23] J. M. Millam and G. E. Scuseria, J. Chem. Phys. **106**, 5569 (1997).
[24] C. Ochsenfeld and M. Head-Gordon, Chem. Phys. Lett. **270**, 399 (1997).
[25] F. Mauri, G. Galli, and R. Car, Phys. Rev. B **47**, 9973 (1993).
[26] P. Ordejón, D. A. Drabold, M. P. Grumbach, and R. M. Martin, Phys. Rev. B **48**, 14646 (1993).
[27] J. Kim, F. Mauri, and G. Galli, Phys. Rev. B **51**, 1456 (1995).
[28] D. Sanchezportal, P. Ordejon, E. Artacho, and J. M. Soler, Int. J. Quant. Chem. **65**, 453 (1997).
[29] P. Ordejon, E. Artacho, and J. M. Soler, Phys. Rev. B **53**, 10441 (1996).
[30] S. Goedecker and L. Colombo, Phys. Rev. Let. **73**, 122 (1994).
[31] S. Goedecker, J. Comp. Phys. **118**, 261 (1995).
[32] Y. Huang, D. J. Kouri, and D. K. Hoffman, Chem. Phys. Lett. **243**, 367 (1995).
[33] A. F. Voter, J. D. Kress, and R. N. Silver, Phys. Rev. B **53**, 12733 (1996).
[34] H. Roder, R. N. Silver, D. A. Drablold, and J. J. Dong, Phys. Rev. B **55**, 15382 (1997).
[35] F. Gagel, J. Comp. Phys. **139**, 399 (1998).
[36] G. Galli, Cur. Op. Sol. State Mat. Sci. **1**, 864 (1996).
[37] D. R. Bowler *et al.*, Mod. Sim. Mat. Sci. Eng. **5**, 199 (1997).
[38] S. Goedecker, Rev. Mod. Phys. , submitted.
[39] W. Kohn, Int. J. Quantum Chem. **56**, 229 (1995).
[40] P. Maslen *et al.*, J. Phys. Chem. A **102**, 2215 (1998).
[41] E. Hernández, M. J. Gillan, and C. M. Goringe, Phys. Rev. B **55**, 13485 (1997).
[42] M. Lepetit, L. Lafon, and X. Lafage, Int. J. Quant. Chem. **64**, 411 (1997).
[43] S. Q. Wei and M. Y. Chou, Phys. Rev. Lett. **76**, 2650 (1996).
[44] C. J. Tymczak and X. Q. Wang, Phys. Rev. Lett. **78**, 3654 (1997).
[45] R. A. Lippert, T. A. Arias, and A. Edelman, J. Comp. Phys. **140**, 278 (1998).
[46] N. T. J. R. Chelikowsky and Y. Saad, Phys. Rev. Lett. **72**, 1240 (1994).
[47] K. W. J. R. Chelikowsky, N. Troullier and Y. Saad, Phys. Rev. B **50**, 11355 (1994).
[48] D. J. S. E. L. Briggs and J. Bernholc, Phys. Rev. B **52**, R5471 (1995).
[49] R. McWeeny, Rev. Mod. Phys. **126**, 1028 (1962).
[50] M. Benzi and C. D. Meyer, SIAM J. Sci. Comput. **16**, 1159 (1995).
[51] M. Benzi, C. D. Meyer, and M. Tůma, SIAM J. Sci. Comput. **17**, 1135 (1996).
[52] M. Challacombe and E. Schwegler, MondoSCF a suite of programs for linear scaling SCF theory, unpublished.
[53] P. O. Löwdin, Phys. Rev. **97**, 1490 (1955).
[54] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry*, first, revised ed. (Mc Graw–Hill Inc., New York, 1989).
[55] C. C. J. Roothaan, Rev. Mod. Phys. **23**, 69 (1951).
[56] J. H. Wilkinson, *The aglebraic Eigenvalue Problem* (Clarendon Press, Oxford, 1965).
[57] G. W. Stewart, *Introduction to Matrix Computations* (Acadmeic Press, London, 1973).
[58] P. O. Löwdin, J. Chem. Phys **18**, 365 (1950).
[59] P. O. Löwdin, Adv. Phys. **5**, 3 (1956).
[60] M. Challacombe, M. Benzi, and M. M. Tůma, to be submitted.
[61] C. A. White, P. Maslen, M. S. Lee, and M. Head-Gordon, Chem. Phys. Lett. **276**, 133 (1997).
[62] M. Benzi and M. Tůma, Technical Report No. LA-UR-98-2175, Los Alamos National Laborarory, Los Alamos, NM (unpublished).
[63] M. Benzi, J. Marin, and M. Tůma, in *Fourth IMACS International Symposium on Iterative Methods in Scientific Computation* (IMACS, Austin, TX, 1998).
[64] F. Liu, Theochem **230**, 47 (1991).
[65] L. Cohen and C. Frishberg, J. Chem. Phys. **65**, 4234 (1976).
[66] C. Frishberg, L. Cohen, and P. Blumenau, Int. J. Quant. Chem. Symp. **14**, 161 (1980).
[67] L. Massa and L. Cohen, Int. J. Quant. Chem. Symp. **14**, 167 (1980).
[68] A. Redondo, Phys. Rev. A **39**, 4366 (1989).
[69] A. Redondo and J. C. Marshall, J. Chem. Phys. **91**, 5492 (1989).
[70] C. Waggoner and L. L. Combs, J. Opt. Theor. App. **76**, 225 (1993).
[71] P. Fantucci and S. Polenzzo, Int. J. Quant. Chem. **52**, 817 (1994).
[72] E. Anderson *et al.*, *LAPACK Users' Guide*, release 2.0 ed., SIAM, 1994.
[73] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipies in FORTRAN* (Cambridge University Press, Port Chester, New York, 1992).
[74] L. C. W. Dixon, in *Nonlinear Optimization Theory and Algorithms*, edited by L. C. W. Dixon, E. Spedicato, and G. P. Szegö (Birkhäuser, Boston, 1980), pp. 124–135.
[75] J. R. Shewchuk, Technical Report No. CMU-CS-94-125, Carnegie Mellon University, Pittsburgh, PA (unpublished).
[76] L. C. W. Dixon, P. G. Ducksbury, and P. Singh, J. Opt. Theo. Appl. **47**, 285 (1985).
[77] J. F. Annett, Comp. Mat. Sci. **4**, 23 (1995).
[78] F. Liu, Theochem **226**, 197 (1991).
[79] A. G. Buckley, Math. Prog. **15**, 200 (1978).
[80] P. Pulay, Chem. Phys. Let. **73**, 393 (1980).
[81] R. T. McLay, S. Swift, and G. F. Carey, J. Par. Dist. Comp. **37**, 146 (1996).
[82] S. Goedecker *et al.*, Technical Report No. LA-UR-97-1504, Los Alamos National Laborarory, Los Alamos, NM (unpublished).
[83] M. R. Field, SIAM J. Sci. Comp. **19**, 27 (1998).
[84] S. Pissanetzky, *Sparse Matrix Technology* (Academic Press, London, 1984).
[85] Y. Saad, *Iterative methods for Sparse Linear Systems* (PWS Publishing, Boston, 1996).
[86] I. S. Duff, A. M. Erisman, and J. K. Reid, *Direct Methods*

for Sparse Matrices (Oxford University Press, London, 1986).
[87] F. G. Gustavson, ACM Trans. Math. Soft. **4**, 250 (1978).
[88] G. Golub and C. F. van Loan, *Matrix Computations* (Johns Hopkins University Press, Baltimore, 1996).
[89] A. D. Daniels, J. M. Millam, and G. E. Scuseria, J. Chem. Phys. **107**, 425 (1997).
[90] M. Challacombe and E. Schwegler, to be submitted.
[91] P. Pulay, J. Comp. Chem. **3**, 556 (1982).
[92] E. Polak, *Computational methods in optimization: a unified approach* (Academic Press, London, 1971).
[93] M. J. D. Powell, Math. Prog. **12**, 241 (1977).
[94] T. R. Gibbs and A. Popolato, *LASL Explosive Property Data* (University of California Press, Berkeley, 1980), p. 109.
[95] M. J. Frisch *et al.*, *Gaussian 94, Revision E.2*, Gaussian Inc., Pittsburgh PA, 1995.
[96] M. Smotherman, program mm.c, available by anonymous ftp from ftp.nosc.mil under /pub/aburto/mm. Compiled using `cc -DN=500 -DUNIX -O2 -64 -r10000` and executed as `mm -w 50`, 1997.
[97] M. J. Daydé and I. S. Duff, Lect. Not. Comp. Sci. **1215**, 108 (1997).