Locality and Sparsity of Ab Initio One-Particle Density Matrices and Localized Orbitals

P. E. Maslen,* C. Ochsenfeld, C. A. White, M. S. Lee, and M. Head-Gordon*

Department of Chemistry, University of California—Berkeley, Berkeley, California 94720

Received: September 5, 1997; In Final Form: November 12, 1997

The cost of Hartree−Fock and local correlation methods is strongly dependent on the locality of the one-particle density matrix and localized orbitals. In this paper the locality and sparsity of the one-particle density matrix is investigated numerically and theoretically, primarily at the Hartree−Fock level, for linear alkanes containing up to 320 carbon atoms. A method for the calculation of localized, atom-centered, occupied orbitals is presented and compared with the Boys' localization procedure. The atom-centered orbitals are ideally suited for use in local-correlation calculations. The connection between the size of optimally localized orbitals, the locality of the density matrix, and the onset of linear scaling is investigated.

1. Motivation

Localized molecular orbitals have long been used to interpret the results of ab initio calculations.[1−6] More recently, localized orbitals have been employed to reduce the cost of SCF,[7−13] DFT,[14−19] tight-binding,[20−22] and correlated[23−37] calculations for large molecules. An alternative approach for SCF, DFT, and tight-binding calculations is to exploit the locality of the one-particle density matrix directly, without ever forming a set of molecular orbitals.[38−52] In order to determine the merits of the two approaches, we have conducted some theoretical and numerical investigations of the locality of ab initio orbitals and density matrices.

This paper presents several loosely connected results pertaining to the locality of orbitals and density matrices for large molecules. Section 2 contains a numerical study of the sparsity of the SCF one-particle density matrix for linear alkanes. Two algorithms for the production of nonorthogonal localized orbitals from a localized density matrix are described in section 3. Section 4 contains a numerical study of the locality of these orbitals for the linear alkanes and a comparison of these orbitals with the well-known Boys' localized orbitals.[1−3] In section 5 we present several theorems and conjectures relating the locality of the one-particle density matrix to the size of optimally localized orbitals.

2. Numerical Study of the Sparsity of the SCF One-Particle Density Matrix for Linear Alkanes

Theoretical models of periodic solids suggest that the locality of the one-particle density matrix is related to the size of the HOMO−LUMO (HOMO = higher occupied molecular orbital, LUMO = lowest unoccupied molecular orbital) gap G (also known as the band gap). The density matrix $\mathbf{P}(r,r')$ decays asymptotically as[53−62]

$$\mathbf{P}(r,r') \sim \mathrm{e}^{-\sqrt{G}|r-r'|}, \quad G > 0 \tag{1}$$

and in metals, for which $G = 0$,

$$\mathbf{P}(r,r') \sim 1/|x - x'||y - y'||z - z'|, \quad G = 0$$

The optimally localized occupied orbitals, known as Wannier orbitals, exhibit similar asymptotic decay.

While the relationship between the HOMO−LUMO gap and decay of the density matrix has not been proven for arbitrary nonperiodic systems,[92] it seems likely that sufficiently large linear alkanes (C_nH_{2n+2}) will exhibit behavior similar to that of the periodic solids. The linear alkanes are known to be insulators, and Hartree−Fock calculations using basis sets up to 6-31G* confirm that the HOMO−LUMO gap is large (>0.5 au), so the density matrix is expected to decay rapidly. If the density matrix is represented in a localized basis such as atomic orbitals, it is expected to be sparse for alkane chains appreciably longer than the decay length dictated by the band gap. The sparsity results in the remainder of this section refer to the atomic-orbital representation of the matrices. All calculations were performed using the software package QChem,[71] except where otherwise indicated.

Figure 1a shows the sparsity of the density matrix **P**, the overlap matrix **S**, the inverse overlap matrix $\mathbf{S}^{-1}$, and the Fock matrix **F** for Hartree−Fock STO-3G calculations on linear alkanes containing 10 to 320 carbon atoms. The carbon−carbon and carbon−hydrogen bond lengths have been fixed at 2.91 au and 2.08 au, respectively. Matrix elements 10^8 times smaller than the largest matrix element were set to zero.

The overlap matrix is considerably sparser than the density matrix. The overlap matrix is 40% sparse for $C_{10}H_{22}$, 64% sparse for $C_{20}H_{42}$ and reaches 90% sparsity for $C_{80}H_{162}$. In contrast the density matrix is dense for $C_{10}H_{22}$ and $C_{20}H_{42}$, with less than 20% sparsity, and the onset of sparsity is discouragingly slow. $C_{80}H_{162}$ is only 63% sparse and 90% sparsity is not attained until $C_{320}H_{642}$!

Hartree−Fock calculations performed by other groups have also found the onset of sparsity to be rather slow if a tight threshold is applied. The density matrix for the polyglycine $C_{40}N_{20}O_{21}H_{62}$ is found to be 50% sparse,[43] using a threshold of 10^{-8} and a 3-21G basis. The maximum density matrix element for the alkane $C_{15}H_{32}$ decays to $\sim 10^{-4}$ across the length of the molecule[51] if a split valence plus polarization basis[72] is used while the density matrix for the alkene $C_{20}H_{22}$ decays even more slowly, to $\sim 10^{-3}$ across the length of the molecule.[51] Large calculations exploiting sparsity have also been performed on water clusters and graphite sheets,[50] but the sparsity of the density matrix was not reported. Sparsity results are also

S1089-5639(97)02919-8 CCC: $15.00
Published on Web 02/26/1998

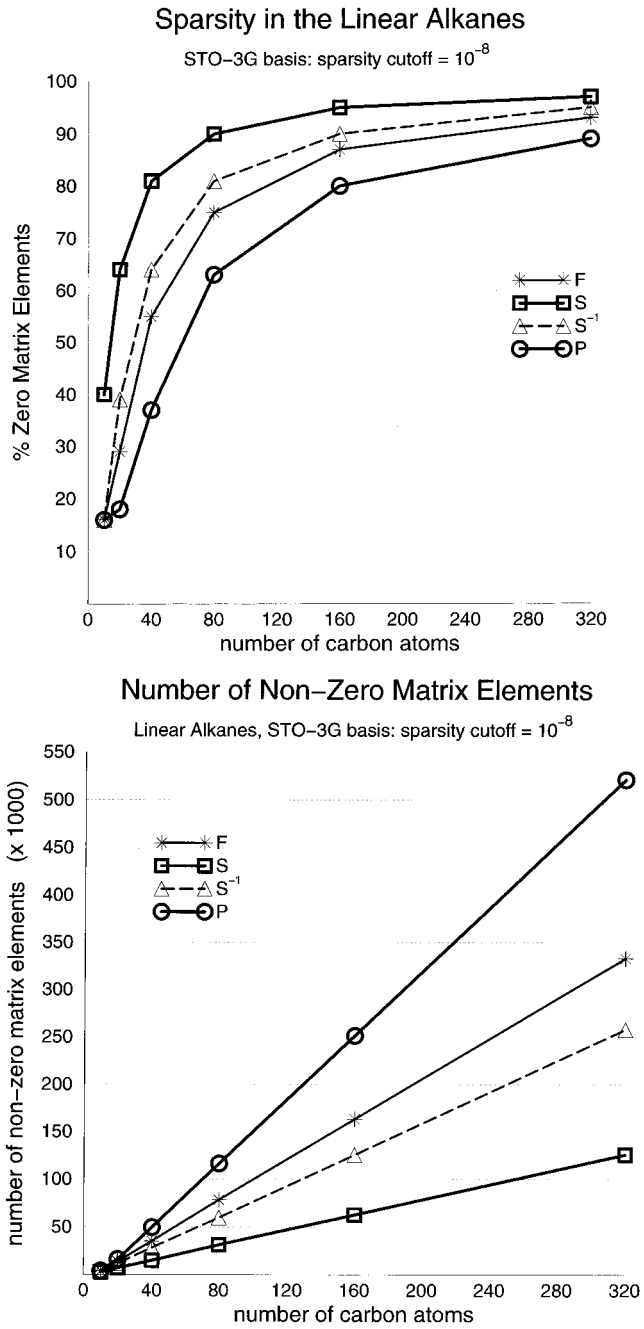

Figure 1. (a) Sparsity of the density matrix for the linear alkanes. (b) Number of nonzero elements in the density matrix for the linear alkanes. Hartree−Fock calculations with an STO-3G basis. Sparsity threshold = 10^{-8}.

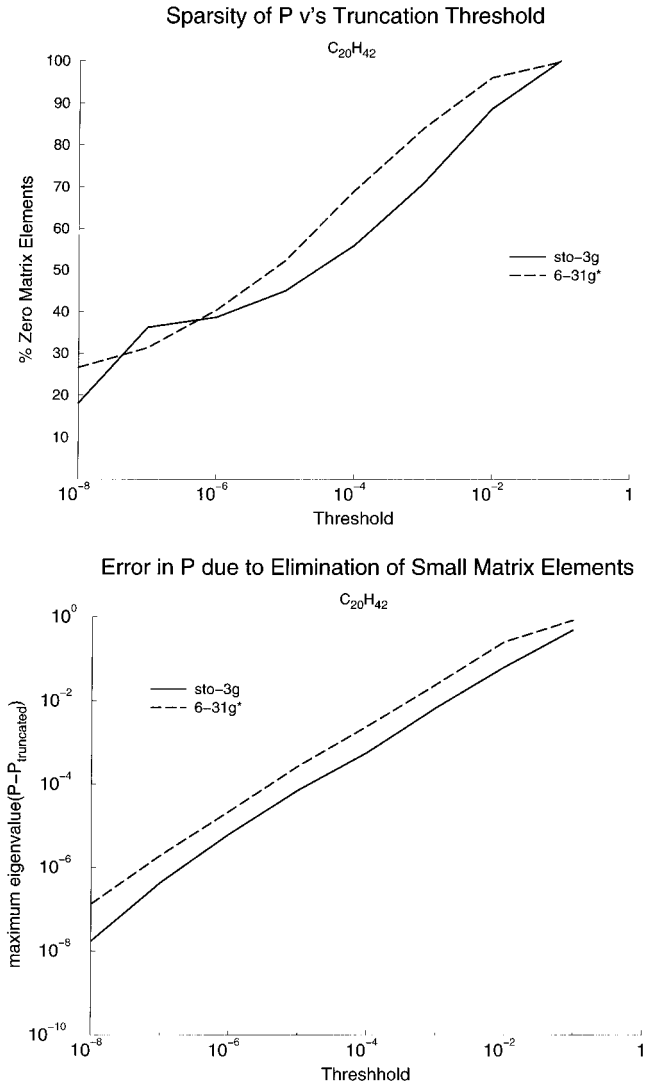

Figure 2. (a) Sparsity of the density matrix as a function of sparsity threshold for $C_{20}H_{42}$. (b) Error in the truncated density matrix as a function of sparsity threshold for $C_{20}H_{42}$. Hartree−Fock calculations.

available for tight-binding, DFT, and unconverged Hartree−Fock calculations, but they are not considered in this paper.

In Figure 1b the number of nonzero matrix elements of the matrices **P**, etc., is plotted as a function of alkane chain length, using the same sparsity threshold as in Figure 1a. All the matrices scale linearly for chain lengths beyond $C_{20}H_{42}$. While it is not obvious from the scale of Figure 1b, **S** scales linearly for the entire graph, while **P** commences linear scaling at $C_{20}H_{42}$. As discussed in sections 4 and 5, $C_{20}H_{42}$ is about the length of the largest localized occupied orbital, again for a cutoff of 10^{-8}. If the cutoff is relaxed to 10^{-6}, then linear scaling of **P** commences at $C_{15}H_{32}$. The variation of sparsity and locality with threshold is discussed further below and in section 4.

To investigate the effect of electronic correlation on the sparsity of the density matrix, we evaluated the Hartree−Fock one-particle density matrix and the MP2 relaxed[73] density matrix for STO-3G $C_{40}H_{82}$. The MP2 relaxed density matrix was calculated using the ab initio program TURBOMOLE.[74] The MP2 density matrix is noticeably less sparse than the Hartree−Fock density matrix; 28% vs 37% for a threshold of 10^{-8}. While the STO-3G basis is too small for an accurate treatment of electron correlation, the STO-3G results suggest that the one-particle density matrix will not become dramatically more sparse upon the inclusion of electron correlation.

These SCF and MP2 results indicate that algorithms which exploit sparsity of the density matrix will need to use a sparsity threshold considerably looser than 10^{-8} for molecules smaller than $C_{20}H_{42}$. For many purposes a looser threshold might be acceptable. Figures 2a,b examine the effect of varying the threshold on the sparsity and accuracy of the density matrix. The sparsity of the density matrix as a function of threshold is plotted in Figure 2a for $C_{20}H_{42}$. If a tight threshold is used, the density matrix has similar sparsity for STO-3G and 6-31G* basis sets. The sparsity increases comparatively slowly as the threshold is loosened for the STO-3G basis and considerably more rapidly for the 6-31G* basis. This might lead one to hope that sparsity could be exploited for larger basis sets, at least for fairly loose thresholds. However, Figure 2b, which plots the maximum eigenvalue of $(\hat{P} - \hat{P}_{\text{truncated}})$ as a function of

threshold, shows that a given threshold has a significantly larger effect on the accuracy of a 6-31G* density matrix than on a STO-3G density matrix. Apparently, the combined effect of the many small matrix elements within a given region of space in the 6-31G* density matrix is significant. Indeed, since the trace of the density matrix is equal to the number of electrons, it is obvious that the average size of the matrix elements must decrease as the basis set is expanded. Thus the size of a reasonable truncation threshold must decrease as the basis set is increased.

The error introduced into ab initio calculations by a loose sparsity threshold is strongly basis dependent, and this makes it difficult to judge the accuracy of the results. Tight (10^{-6}–10^{-8}) thresholds are probably safe, but the one-particle density matrix is effectively dense for alkane chains shorter than $C_{15}H_{32}$−$C_{20}H_{42}$, depending on the threshold.

Locality and sparsity of the density matrix is used in two places in the Hartree−Fock procedure, and we now comment briefly on these.

The first place where locality is used is in the construction of the exchange matrix **K** from the density matrix[47−50,52]

$$\mathbf{K}_\sigma(r,r') = -\frac{1}{|r-r'|}\mathbf{P}_\sigma(r,r')$$

where σ refers to α or β spin. **K** clearly has similar locality properties to **P**, so if **P** scales linearly with system size then **K** will too and the cost of constructing **K** also scales linearly. There is almost no overhead associated with exploiting locality in the construction of **K** and the linear-scaling algorithm provides some computational savings even when **P** is only slightly sparse.[52] For the linear alkanes we expect the cost of constructing **K** to scale linearly beyond $C_{15}H_{32}$−$C_{20}H_{42}$, depending on the threshold.

The second place where sparsity is used is in matrix multiplication. This is particularly important in density-matrix-based schemes[38−46,51] which replace a cubic scaling matrix diagonalization of the Fock matrix **F** with several evaluations of matrix products such as **FPSPS**. The multiplication based scheme is more expensive than matrix diagonalization if the matrices are dense and only becomes economical when considerable sparsity is present. The method is nevertheless useful, at least for the linear alkanes, because the cost of conventional matrix diagonalization is only significant for very large systems, of the order of 3000 basis functions or $C_{160}H_{322}$ with a 6-31G* basis, and the density matrices for such systems exhibit considerable sparsity.

3. Algorithms for Production of Nonorthogonal Localized Orbitals from a Localized Density Matrix

Given an atomic orbital $|\alpha\rangle$, one can project out its occupied component $|\hat{P}\alpha\rangle$ using the density matrix operator $\hat{P}$,

$$|\hat{P}\alpha(r)\rangle \equiv \int \hat{P}(r,r')|\alpha(r')\rangle\, dr' \qquad (2)$$

$|\hat{P}\alpha(r)\rangle$ will be about as localized as $\hat{P}(r,r')$ assuming that the atomic orbital $|\alpha(r')\rangle$ is much more localized than $\hat{P}$. Application of $\hat{P}$ to the complete set of N atomic orbitals $\{|\alpha\rangle\}_\alpha$ leads to a complete set of localized occupied orbitals $\{|\hat{P}\alpha\rangle|_\alpha$. This set of occupied orbitals is highly redundant, since the number of atomic orbitals is far greater than the dimension of the occupied space. In this section we describe two ways to remove some or all of the redundant occupied orbitals. The first algorithm selects a number of occupied orbitals equal to the number of functions in a minimal basis (one for hydrogen, five for carbon, etc.). This reduced set still contains some redundancies but is well suited to local correlation algorithms. Indeed, we have been able to greatly increase the locality of our local-correlation method[37] by using a redundant set of occupied orbitals. The second algorithm removes all redundancies from the set of occupied functions.

Expressing eq 2 in the atomic orbital basis one obtains

$$|\hat{P}\alpha\rangle = \sum_{\alpha\beta}(\mathbf{PS})_{\beta\alpha}|\beta\rangle \qquad (3)$$

where **S** is the overlap matrix and **P** the contravariant density matrix in the atomic orbital basis. We find it helpful to use atomic orbitals $|\alpha'\rangle$ which are orthonormal to other orbitals on the same nucleus

$$|\alpha'\rangle = \sum_{\alpha\in A}(\mathbf{S}_{AA})^{-1/2}_{\alpha\alpha'}|\alpha\rangle$$

where $\mathbf{S}_{AA}$ is the portion of the overlap matrix describing orbitals $|\alpha\rangle$ centered on each Ath nucleus. The overlap matrix in the $|\hat{P}\alpha'\rangle$ basis is then

$$\langle\hat{P}\alpha'|\hat{P}\beta'\rangle = \langle\alpha'|\hat{P}|\beta'\rangle = (\mathbf{SPS})_{\alpha'\beta'} = \sum_{\alpha\beta}(\mathbf{S}_{AA})^{-1/2}_{\alpha'\alpha}(\mathbf{SPS})_{\alpha\beta}(\mathbf{S}_{BB})^{-1/2}_{\beta\beta'}$$

where we have taken advantage of the idempotency condition for the density matrix, $\hat{P}\hat{P} = \hat{P}$. We are free to perform a unitary rotation of the atomic orbitals $\{|\alpha'\rangle\}$ on each Ath nucleus such that the set of projected orbitals $\{|\hat{P}\alpha''\rangle\}$ on that nucleus are orthogonal,

$$|\alpha''\rangle = \sum_{\alpha'\in A}U_{\alpha'\alpha''}|\alpha'\rangle$$

$$\langle\hat{P}\alpha''|\hat{P}\beta''\rangle = \delta_{\alpha''\beta''}\langle\hat{P}\alpha''|\hat{P}\beta''\rangle$$

α'' and β'' on the same Ath nucleus

To determine whether each atomic orbital $|\alpha''\rangle$ is weakly or strongly occupied, we calculate the Mulliken population,[87] $(\mathbf{PS})_{\alpha''\alpha''}$. Strongly occupied atomic orbitals have approximately unit occupancy, while weakly occupied orbitals have approximately zero occupancy. We note in passing that it is possible to define atomic orbitals which diagonalize each atomic block $(\mathbf{PS})_{AA}$, though the corresponding projected atomic orbitals on atom A are not orthogonal. We now consider the elimination of linear dependencies from $\{|\hat{P}\alpha''\rangle\}$ involving orbitals $|\hat{P}\alpha''\rangle$, $|\hat{P}\beta''\rangle$, ... centered on two or more different nuclei *A*, *B*, ...

Algorithm One. This algorithm does not seek to eliminate all linear dependencies from the basis of projected atomic orbitals. Rather, it reduces the size of the occupied basis to the size of a minimal basis.[93] To do this, consider the number of basis functions m_A in a minimal basis for each Ath atom in turn (one for hydrogen, five for carbon, etc.) and choose the m_A projected atomic orbitals $|\hat{P}\alpha''\rangle$ centered on the Ath nucleus with the largest orbital occupancies.

For molecules with simple bonding such as the alkanes there are usually m_A strongly occupied orbitals on each atom, with the remainder being weakly occupied. Under these conditions we have always found that the minimal basis of projected orbitals spans the occupied space. Some molecules with complicated bonding such as SF_6 exhibit a fairly even distribution of orbital occupancies on the atoms responsible for the complicated behavior (sulfur in this example). A minimal basis

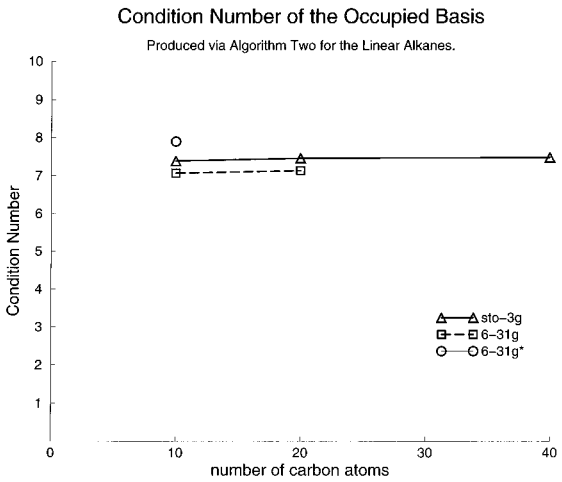

Figure 3. Condition number of the occupied basis produced via algorithm two for the linear alkanes. Hartree−Fock calculations.

may not be appropriate for atoms exhibiting a continuous range of orbital occupancies and we suggest including all projected atomic orbitals with nonnegligible occupancies for such atoms.

Algorithm Two. This algorithm attempts to remove all linear dependencies from the projected atomic orbitals. We begin by selecting a projected atomic orbital $|\hat{P}\alpha''\rangle$ with a large orbital occupancy. Next, we select another projected atomic orbital $|\hat{P}\beta''\rangle$ with a large orbital occupancy and construct the 2×2 overlap matrix in the projected basis. A Cholesky decomposition[85] of the overlap matrix is performed to produce a decomposed matrix $\mathbf{S}^C$. If $|\hat{P}\beta''\rangle$ is orthogonal to $|\hat{P}\alpha''\rangle$, the diagonal matrix element of $\mathbf{S}^C$ corresponding to $|\hat{P}\beta''\rangle$ equals 1, and if $|\hat{P}\beta''\rangle$ is nearly linearly dependent with $|\hat{P}\alpha''\rangle$, then the diagonal matrix element is $\simeq 0$. If a near linear dependence is found, $|\hat{P}\beta''\rangle$ is discarded and the corresponding row and column of $\mathbf{S}$ and $\mathbf{S}^C$ are deleted. The procedure continues with the selection of additional basis functions $|\hat{P}\gamma''\rangle$, adding a row and a column to the overlap matrix and its Cholesky decomposition at each stage, until a complete set of nonredundant occupied functions is obtained.

The cost of a full Cholesky decomposition scales cubically with the number of atomic orbitals, though for sparse overlap matrices linear scaling could probably be achieved using a sparse matrix version of the incomplete[86] Cholesky decomposition. The incomplete decomposition has the virtue that the sparsity of the overlap matrix $\mathbf{S}$ is preserved in $\mathbf{S}^C$. While the cubic cost of a full decomposition is negligible compared with the cost of a correlated ab initio calculation, it would be significant for very large SCF calculations ($\geq$3 000 basis functions) using linear-scaling algorithms.

Numerical studies of the projected orbitals $|\hat{P}\alpha''\rangle$ produced by algorithm two suggest that they contain no linear dependencies or near linear dependencies. To test for linear dependencies, the overlap matrix was constructed in the $|\hat{P}\alpha''\rangle$ basis and diagonalized. Figure 3 shows the condition number of the overlap matrix (i.e., the ratio of the largest to the smallest eigenvalue) for linear alkanes with chain lengths from 10 to 40 carbon atoms. An orthonormal basis has a condition number of one, while a very-large condition number indicates a near linear dependency. The condition number of the projected atomic orbitals is small, around seven, and is insensitive to chain length and basis-set size. The weak basis-set dependence indicates that the occupied orbitals undergo small, subtle changes as the basis increases. This is not too surprising, given that even a minimal basis of atomic orbitals is sufficient for a reasonable description of the occupied space. Encouragingly, the condition number does not increase with increasing chain length, indicating that the projected orbitals $|\hat{P}\alpha''\rangle$ form a suitable basis for the occupied space irrespective of molecule size. For the cases considered in Figure 3, the orbitals produced by algorithm two were always a subset of the minimal basis produced by algorithm one.

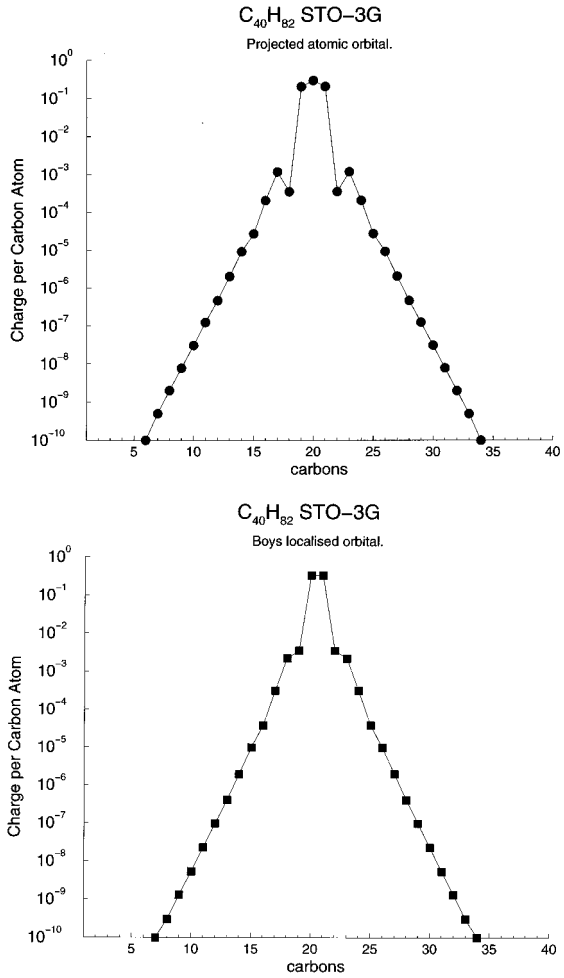

Figure 4. Localized orbitals for $C_{40}H_{82}$: (a) Projected atomic orbital, (b) Boys' localized orbital Hartree−Fock calculations, STO-3G basis.

4. Numerical Study of the Size of Localized Orbitals for Linear Alkanes

In this section we look at the size of the projected atomic orbitals described in section 3. As discussed in the introduction, highly compact orbitals increase the efficiency of local SCF and local correlation calculations, and in this context it is interesting to compare the size of projected atomic orbitals and Boys' localized orbitals.

For each localized orbital, we have employed Mulliken's prescription[87] for partitioning the corresponding electronic charge distribution into atomic contributions belonging to a single atom and bonding contributions belonging to a pair of atoms. To simplify Figures 4−6, we have plotted only the atomic charges on the carbon atoms, omitting both the bonding charges and also the atomic charges on the hydrogen atoms. Omitting the bonding contributions causes some minor humps in the wings of the orbitals in Figures 4−6, but doing so provides the clearest picture of the exponential decay of the orbital tails.

Only the largest localized orbital from each set is displayed. For each displayed orbital, a large number of nearly identical

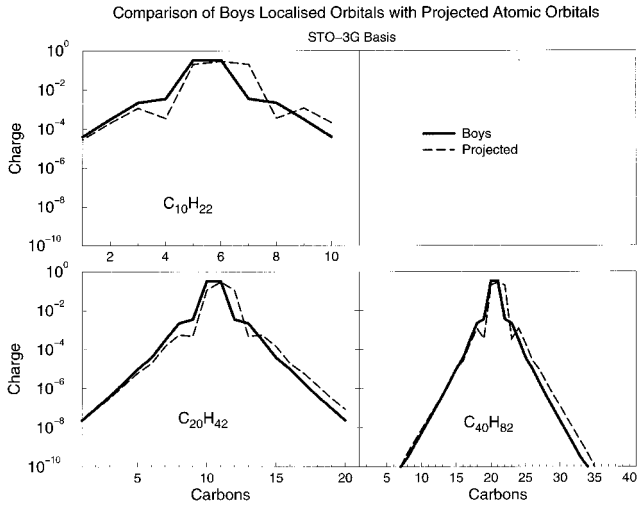

Figure 5. Comparison of Boys' localized orbitals with projected atomic orbitals for the linear alkanes. Hartree−Fock calculations, STO-3G basis.

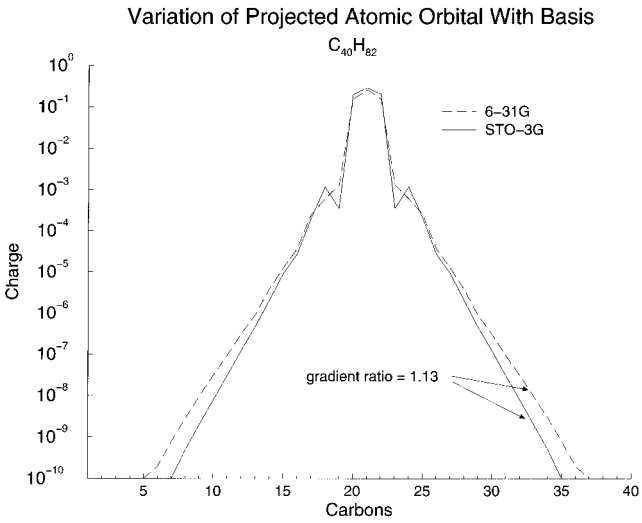

Figure 6. Variation of projected atomic orbital with basis set for $C_{40}H_{82}$. Hartree−Fock calculations.

orbitals exist, differing chiefly in a translation by one or more atoms along the carbon backbone. The sequence of translated orbitals is disrupted near the ends of the carbon chain. Orbitals near the end of the chain appear to be squashed and somewhat more localized than orbitals in the middle of the chain.

Figures 4a,b depict the charge distribution of the largest projected atomic orbital and Boys' localized orbital for $C_{40}H_{82}$, using an STO-3G basis. Most of the charge in the Boys' orbital is concentrated around a bonding pair of two carbon atoms. In contrast, most of the charge in the projected orbital is distributed evenly between three consecutive carbon atoms. This is a necessary consequence of translational near symmetry in the midsections of large linear alkanes; when the projection operator is applied to an atomic orbital on a given carbon atom, the resulting projected orbital will be approximately symmetric about that atom. The most important feature of Figure 4a,b, however, is not the distribution of the bulk of the charge about two or three central atoms but rather the charge distribution in the tails of the orbitals. Both the projected orbital and the Boys' orbital have exponentially decaying tails, as evidenced by a straight line on the logarithmic plots. This is exactly the kind of decay found in the Wannier orbitals[53,54] and density matrices of periodic systems with a band gap (eq 1). Intriguingly, the Boys' orbital and the projected orbital exhibit similar decay rates and orbital widths, even though they are produced via very different algorithms. This leads us to speculate that *the size of optimally localized orbitals is determined by the decay rate of the density matrix*, a point we return to in section 5.

To aid in the comparison of Boys' orbitals and projected orbitals, they have been superposed in Figure 5 for alkane chain lengths ranging from 10 to 40 carbon atoms. Because the projected orbital is centred on three atoms whereas the Boys' orbital is centered on two, it is not possible to superpose the two orbitals in a symmetrical manner. We have displayed the projected orbital whose left-hand tail aligns with the tail of the corresponding Boys' orbital. As noted above a second projected orbital exists, differing from the first by a translation of one carbon atom to the left, and the right-hand tail of the second projected orbital aligns with the right-hand tail of the Boys' orbital. Figure 5 clearly shows that for large enough molecules the tails of Boys' orbitals and projected orbitals exhibit identical exponential decay, while the orbital widths are similar, differing by only one carbon atom.

Very small basis sets such as STO-3G usually have an artificially large HOMO−LUMO gap, with the gap rapidly decreasing to an asymptotic limit as the basis size is increased.[94] For $C_{20}H_{42}$ the STO-3G HOMO−LUMO gap is 0.89 au, the 6-31G HOMO−LUMO gap is 0.60 au, and the 6-31G* HOMO−LUMO gap is 0.60 au, while for $C_{40}H_{82}$ the STO-3G HOMO−LUMO gap is 0.88 au and the 6-31G HOMO−LUMO gap is 0.59 au. Equation 1 leads us to anticipate that the size of the localized orbitals will increase as the gap decreases, and hence that the size of the orbitals will initially increase as the basis set is enlarged.

Figure 6 shows the largest projected orbital for $C_{40}H_{82}$, for the STO-3G and 6-31G basis sets. On the basis of eq 1, one would expect the tail of the STO-3G orbital to decay 1.22 times more rapidly than the tail of the 6-31G orbital. Figure 6 agrees qualitatively with this prediction, though we find that the STO-3G orbital decays 1.13 times more rapidly than the 6-31G orbital. Since eq 1 is derived from a tight-binding Hamiltonian for an infinite periodic system, whereas Figure 6 is based on a Hartree−Fock calculation for a large molecule, it is not surprising that the numbers are in only qualitative agreement.

We have previously indicated that, in order for local Hartree−Fock and local correlation methods to afford appreciable time savings, the molecule under consideration must be larger than the localized orbitals. The results presented in this section suggest that the size of the orbitals is determined by the decay rate of the density matrix and that this is only weakly dependent (13%) on the basis set, with most of the change occurring on going from a minimal (STO-3G) basis to a 6-31G basis.

To increase the efficiency of local algorithms, it is common practice to truncate the exponentially decaying tails of the orbitals below some threshold. Figure 7 shows the size of the STO-3G projected orbital for $C_{40}H_{82}$ as a function of truncation threshold. For a very loose threshold of $10^{-3}-10^{-4}$ the orbital is highly localized, with a width of just 7−9 carbon atoms. The central carbon atom in the orbital is connected to only 3−4 carbon atoms on either side—roughly the number of neighbors commonly included in tight-binding calculations.[38−40,42] To achieve the accuracy required in most ab initio calculations, a tighter threshold is required. For a threshold of 10^{-6} the projected orbital covers 15 carbon atoms, increasing to 21 atoms for a threshold of 10^{-8} and 29 carbon atoms—about three quarters of the entire $C_{40}H_{82}$ carbon chain—for a very tight threshold of 10^{-10}. As noted in section 2, the number of nonzero

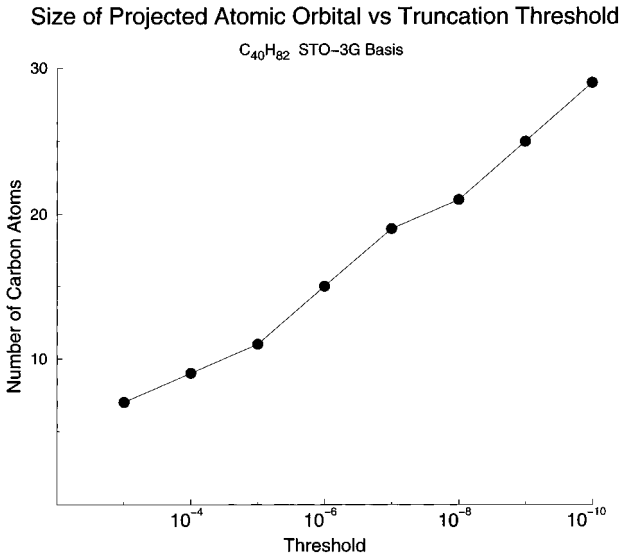

Figure 7. Size of projected atomic orbital as a function of truncation threshold for $C_{40}H_{82}$. Hartree−Fock calculations, STO-3G basis.

elements in the density matrix scales linearly with molecule size once the number of carbon atoms exceeds the length of a projected orbital.

The minimum molecule size at which local methods become efficient is strongly dependent on the required accuracy. *If appreciable time savings can be achieved for a molecule size approximately double the length of a lo*calized *orbital, then the minimum alkane chain length will be* 14−58 *carbon atoms, depending on the truncation threshold*. While these results are encouraging for one-dimensional chains, they do not augur well for dense two- and three-dimensional solids such as diamond, where it seems quite likely that the minimum molecule sizes will be 14 × 14 and 14 × 14 × 14 carbon atoms, respectively. Liquids and other weakly bonded systems are more amenable to linear scaling algorithms because they are less dense, and the density matrix tends to be more localized.[43,50] While all local methods to date are based on truncation of the nonlocal regions of the density matrix or the tails of the orbitals, it seems that for dense three-dimensional solids and highly delocalized systems such as metals a more sophisticated treatment of nonlocality will be required to produce time savings without sacrificing accuracy.[95]

5. Theorems and Conjectures Connecting the Locality of the Hartree−Fock One-Particle Density Matrix and the Size of Optimally Localized Occupied Orbitals

This section contains several theorems and conjectures relating the locality of the Hartree−Fock one-particle density matrix to the size of optimally localized orbitals spanning the occupied space. More generally, the results apply to any projection operator defining a subspace. While the theorems are not as strong as we would like, they serve to deepen our understanding of the numerical results.

Definitions. A *well-conditioned* overlap matrix is one containing no linear dependencies or near linear dependencies. Its *condition number*, the ratio of its largest and smallest eigenvalues, is small. A *well-conditioned set of orbitals* is a set of orbitals whose corresponding overlap matrix is well conditioned.

Theorem 1. *If a well-conditioned set of localized occupied orbitals exists, then the density matrix describing the occupied space is similarly localized.*

Proof. Denote the localized occupied orbitals by ϕ_i, ϕ_j, and the corresponding overlap matrix by $\mathbf{S}_{ij}$. Because the orbitals are localized, the overlap matrix is localized. The density matrix is related to the occupied orbitals by,

$$\mathbf{P}(r,r') = \sum_{ij} \phi_i(r) \mathbf{S}_{ij}^{-1} \phi_j(r') \tag{4}$$

Because $\mathbf{S}$ is localized and well conditioned, it can be shown that $\mathbf{S}^{-1}$ is similarly localized, as noted by Nunes and Vanderbilt.[40] Nunes and Vanderbilt derived this result from a general theorem concerning the locality of the Green's function $(\mathbf{S} - \epsilon\mathbf{I})^{-1}$. A direct proof is presented in the Appendix. Since all quantities on the right-hand side of eq 4 are localized, with a localization width dependent on the size of the occupied orbitals, $\mathbf{P}$ must also be localized.

Corollary 1.1. *For the special case where the localized orbitals are orthogonal, the localization width of* $\boldsymbol{P}$ *is less than or equal to the size of the most delocalized orbital.*

Proof. This follows directly from eq 4.

Corollary 1.2. *If the density matrix is delocalized, then the set of optimally localized occupied orbitals is either ill conditioned or delocalized.*

Proof. This is just a restatement of the theorem.

Conjecture. *If the density matrix is localized, then a well-conditioned set of localized occupied orbitals exists.*

Supporting Evidence. Algorithm Two in section 3 produces a localized set of orbitals directly from a localized density matrix. They are not guaranteed to be well conditioned, but numerical evidence supports this conjecture. For the special case of a periodic solid the Wannier orbitals are known to be localized.

Discussion of the above Results. The above results pertain only to linearly independent sets of occupied orbitals. It is of considerable practical interest to know whether a set of localized occupied orbitals can always be found, even when the density matrix is delocalized, by admitting larger sets of occupied orbitals containing several exact linear dependencies. A simple counter example shows that localized orbitals do not always exist. Consider the ion formed from a single electron orbiting a long chain of evenly spaced identical nuclei. The single occupied orbital for the system encompasses all the nuclei and hence is highly delocalized. Since the occupied space is one dimensional, the occupied orbital is uniquely defined and there is no possibility of forming localized occupied orbitals, though even in this extreme case the occupied orbital can be represented as the projection of a single atomic orbital onto the occupied space.

6. Conclusion

Numerical and theoretical investigations of the locality of Hartree−Fock one-particle density matrices and localized orbitals have uncovered several results with significant practical ramifications for large scale ab initio calculations.

Ramification 1. Two algorithms have been presented for the production of atom-centered occupied orbitals. The orbitals are localized and nonorthogonal. Each occupied orbital is obtained by projecting a single atomic orbital onto the occupied space. One algorithm produces a linearly independent set of occupied orbitals, while the other algorithm produces a minimal set of occupied orbitals, one for hydrogen, five for carbon, etc. The minimal set of localized orbitals is ideally suited for use in local-correlation calculations.[37]

Ramification 2. The size of optimally localized occupied orbitals is comparable with the nonlocal extent of the one-particle density matrix. More specifically, the exponential decay rate of the tails of the largest orbital(s) matches the exponential decay of the density matrix. For molecules which are larger than the largest localized orbital, the number of nonzero elements of the density matrix grows linearly with molecule size.

Ramification 3. Even for highly insulating systems such as the linear alkanes, the nonlocal extent of the density matrix and the localized orbitals is considerable—7 carbon atoms using a very loose truncation threshold of 10^{-3} or 29 carbon atoms using a tight threshold of 10^{-10}. If appreciable time savings can be achieved for a molecule size approximately double the length of a localized orbital, then the minimum chain length amenable to sparse matrix calculations will be 14−58 carbon atoms, depending on the truncation threshold. For a fairly strict sparsity threshold of 10^{-8}, the orbital width is 21 carbon atoms and the density matrix does not reach 90% sparsity until $C_{320}H_{642}$. In contrast, recent results indicate that the derivative of the density matrix with respect to the displacement of a nucleus is very sparse.[51]

Ramification 4. Loose sparsity thresholds have an unpredictable, basis-set dependent effect on the accuracy of an ab initio calculation. Because the trace of the density matrix is equal to the number of electrons, the average size of the matrix elements must decrease as the basis set is expanded. Thus the size of a reasonable truncation threshold must decrease as the basis set is increased. Tight (10^{-6}−10^{-8}) sparsity thresholds appear to be safe but are only useful for chain lengths of more than 15−20 carbon atoms.

Ramification 5. These results suggest that schemes which exploit the locality of the density matrix in SCF calculations will enable considerable savings for large linearly bonded insulators such as long alkane chains and proteins. Liquids and other weakly bonded two- and three-dimensional systems are also amenable to linear scaling algorithms because of their relatively low-density and localized density matrices.[43,50] Locality will be less useful for dense systems with three-dimensional bonding such as diamond and for highly delocalized systems such as metals, except in very low-accuracy minimum basis calculations using loose thresholds.

Ramification 6. It is highly desirable to develop a fast treatment of the nonlocal regions of the density matrix which yields a rigorous error bound for the ab initio energy, or which can be converged to the correct result in a controlled manner.

Acknowledgment. Early development of this work was supported by a grant from the Hellman Family Fund. We gratefully acknowledge funding from Sandia National Laboratory (subcontract LS-5503), the Packard Foundation, and the National Science Foundation (Grant CHE9357129). C.O. acknowledges financial support by a postdoctoral fellowship from the DFG (Deutsche Forschungsgemeinschaft).

Appendix: Comment on the Locality of the Inverse of a Hermitian Positive Definite Localized Matrix

Given a Hermitian positive definite localized matrix **S**, we wish to determine an upper bound to the nonlocality of $\mathbf{S}^{-1}$. This has been considered previously for banded matrices[90] and for general Hermitian positive definite matrices.[40] The previous discussion of the locality of $\mathbf{S}^{-1}$ was couched in terms of the the Green's function for $(\mathbf{S} - \epsilon\mathbf{I})^{-1}$. In this appendix we adopt a more direct approach, expanding $\mathbf{S}^{-1}$ in a binomial series.

Without loss of generality we assume that **S** is the overlap matrix for a set of normalized basis functions. We also assume that **S** has been scaled so that its eigenvalues lie on the interval $[1/C,1]$, where C is the condition number of the matrix, $C \geq 1$. **S** can be written as

$$\mathbf{S} = \mathbf{I} - \epsilon$$

where ϵ is a local positive semidefinite matrix with maximum eigenvalue $1-1/C$, so

$$|\epsilon_{ij}| \leq 1 - 1/C, \quad \forall i,j$$

$\mathbf{S}^{-1}$ can then be expanded in a binomial series in ϵ,

$$\mathbf{S}^{-1} = I + \epsilon + \epsilon^2 + ...$$

The $(n+1)^{\text{th}}$ term in the series is ϵ^n, which can be up to n times as delocalized as **S**. ϵ^n decays as,

$$|\epsilon^n_{ij}| \leq (1 - 1/C)^n, \quad \forall i,j$$
$$\approx \exp(-n/C), \quad C \gg 1 \qquad \text{and} \qquad n \gg 1 \quad \text{(A1)}$$

ϵ^n decays exponentially with n, with the rate of decay decreasing as the condition number C increases.

References and Notes

(1) Boys, S. F. *Rev. Mod. Phys.* **1960**, *32*, 296.
(2) Foster, J. M.; Boys, S. F. *Rev. Mod. Phys.* **1960**, *32*, 300.
(3) Boys, S. F. In Löwdin (ref 91), 253.
(4) Edmiston, C.; Ruedenberg, K. *Rev. Mod. Phys.* **1963**, *35*, 457.
(5) Edmiston, C.; Ruedenberg, K. In Löwdin (ref 91), 263.
(6) Pipek, J.; Mezey, P. G. *J. Chem. Phys.* **1989**, *90*, 4916.
(7) Lefebvre, R. *C. R. Acad. Sci. (Paris)* **1955**, *240*, 1094.
(8) Lefebvre, R. *J. Chim. Phys.* **1957**, *54*, 168.
(9) Lefebvre, R.; Moser, C. M. *J. Chim. Phys.* **1956**, *53*, 393.
(10) Adams, W. H. *J. Chem. Phys.* **1961**, *34*, 89.
(11) Gilbert, T. L. In *Molecular Orbitals in Chemistry, Physics and Biology*; Löwdin, P.-O., Pullman, B., Eds.; Academic: New York, 1964; p 405.
(12) Daudey, J. P. *Chem. Phys. Lett.* **1974**, *24*, 574.
(13) Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1977**, *46*, 11.
(14) Yang, W. *Phys. Rev. Lett.* **1991**, *66*, 1438.
(15) Galli, G.; Parrinello, M. *Phys. Rev. Lett.* **1992**, *69*, 3547.
(16) Mauri, F.; Galli, G.; Car, R. *Phys. Rev. B* **1993**, *47*, 9973.
(17) Stechel, E. B.; Williams, A. R.; Feibelman, P. J. *Phys. Rev. B* **1994**, *49*, 10088.
(18) Lewis, J. P.; Ordejón, P.; Sankey, O. F. *Phys. Rev. B* **1997**, *55*, 6880.
(19) Fernandez, P.; Dal Corso, A.; Baldereschi, A.; Mauri, F. *Phys. Rev. B* **1997**, *55*, 1909.
(20) Ordejón, P.; Drabold, D. A.; Grumbach, M. P.; Martin, R. M. *Phys. Rev. B* **1993**, *48*, 14646.
(21) Goedecker, S.; Colombo, L. *Phys. Rev. Lett.* **1994**, *73*, 122.
(22) Goedecker, S.; Teter, M. *Phys. Rev. B* **1995**, *51*, 9455.
(23) Sinanoglu, O. *Adv. Chem. Phys.* **1964**, *6*, 315.
(24) Diner, S.; Malrieu, J. P.; Calverie, P. *Theor. Chim. Acta* **1969**, *13*, 8.
(25) Meyer, W. *Int. J. Quantum Chem., Quantum Chem. Symp.* **1971**, *5*, 341.
(26) Robb, M. A. In *Computational Techniques in Quantum Chemistry and Molecular Physics*; Diercksen, G. H. F., Sutcliffe, B. T., Veillard, A., Eds.; *NATO Advanced Study Institutes Series. Series C, Mathematical and Physical Sciences* 15; Reidel: Boston, 1975; p 435.
(27) Ahlrichs, R.; Lischka, H.; Staemmler, V.; Kutzelnigg, W. *J. Chem. Phys.* **1975**, *62*, 1225.
(28) Meyer, W. *J. Chem. Phys.* **1976**, *64*, 2901.
(29) Stollhoff, G.; Fulde, P. *J. Chem. Phys.* **1980**, *73*, 4548.
(30) Dykstra, C. E.; Chiles, R. A.; Garrett, M. D. *J. Comput. Chem.* **1981**, *2*, 266.
(31) Pulay, P. *Chem. Phys. Lett.* **1983**, *100*, 151.
(32) Saebø, S.; Pulay, P. *Chem. Phys. Lett.* **1985**, *113*, 13.
(33) Kirtman, B.; Dykstra, C. E. *J. Chem. Phys.* **1986**, *85*, 2791.
(34) Murphy, R.; Beachy, B.; Friesner, F.; Ringnalda, M. N. *J. Chem. Phys.* **1995**, *103*, 1481.
(35) Kirtman, B. *Int. J. Quantum Chem.* **1995**, *55*, 103.

(36) Bernholdt, D.; Harrison, R. J. *J. Chem. Phys.* **1995**, *102*, 9582.
(37) Maslen, P. E.; Head-Gordon, M. *Chem. Phys. Lett.* **1997**. In press.
(38) Li, X. P.; Nunes, R. W.; Vanderbilt, D. *Phys. Rev. B* **1993**, *47*, 10891.
(39) Daw, M. S. *Phys. Rev. B* **1993**, *47*, 10895.
(40) Nunes, R. W.; Vanderbilt, D. *Phys. Rev. B* **1994**, *50*, 17611.
(41) Hernandez, E.; Gillan, M. J. *Phys. Rev. B* **1995**, *51*, 10157.
(42) Xu, C. H.; Scuseria, G. E. *Chem. Phys. Lett.* **1996**, *262*, 219.
(43) Millam, J. M.; Scuseria, G. E. *J. Chem. Phys.* **1997**, *106*, 5569.
(44) Daniels, A. D.; Millam, J. M.; Scuseria, G. E. *J. Chem. Phys.* **1997**, *107*, 425.
(45) White, C. A.; Head-Gordon, M. 1997, Manuscript in preparation.
(46) Lee, M. S.; Head-Gordon, M. *J. Chem. Phys.* **1997**, *107*, 9085.
(47) Schwegler, E.; Challacombe, M. *J. Chem. Phys.* **1996**, *105*, 2726.
(48) Burant, J. C.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **1996**, *105*, 8969.
(49) Challacombe, M.; Schwegler, E. M. *J. Chem. Phys.* **1997**, *106*, 5526.
(50) Schwegler, E.; Challacombe, M.; Head-Gordon, M. *J. Chem. Phys.* **1997**, *106*, 9708.
(51) Ochsenfeld, C.; Head-Gordon, M. *Chem. Phys. Lett.* **1997**, *270*, 399.
(52) Ochsenfeld, C.; White, C. A.; Head-Gordon, M. *J. Chem. Phys.* **1997**. Submitted for publication.
(53) Kohn, W. *Phys. Rev.* **1959**, *115*, 809.
(54) Kohn, W. *Phys. Rev. B* **1973**, *7*, 4388.
(55) Wannier, G. *Phys. Rev.* **1937**, *52*, 191.
(56) Des Cloizeaux, J. *Phys. Rev.* **1963**, *129*, 554.
(57) Des Cloizeaux, J. *Phys. Rev. A* **1964**, *135*, 685.
(58) Des Cloizeaux, J. *Phys. Rev. A* **1964**, *135*, 698.
(59) Blount, E. I. In *Solid State Physics*; Seitz, F., Turnbull, D., Eds.; Academic: NY, 1962; Vol. 13, p 305.
(60) Nenciu, G. *Commun. Math. Phys.* **1983**, *91*, 81.
(61) Marzari, N.; Vanderbilt, D. *Phys. Rev. B* **1997**. Submitted for publication.
(62) Baer, R.; Head-Gordon, M. *Phys. Rev. Lett.* **1997**, *79*, 3962. In this paper the density matrix is expanded in powers of the Hamiltonian matrix. One of the authors (P.E.M.) notes that this expansion[21] converges very slowly for typical Hamiltonian matrices and only provides a competitive alternative to matrix diagonalisation if the condition number of the matrix is small.[63−65] (A competitive alternative to diagonalization has recently been devised by combining the polynomial expansion with a renormalisation group approach.[88]) The locality of the density matrix is estimated from the rate of convergence of the expansion, but disordered systems[68,69] with a large condition number may be much more local than this estimate suggests.
(63) Denman, E. D.; Beavers, A. N., Jr. *Appl. Math. Comput.* **1976**, *2*, 63.
(64) Larin, V. B. *Dokl. Akad. Nauk SSSR* **1991**, *320*, 536 (in English, translated by Scripta. Technica, Inc.).
(65) Higham, N. J. *Linear Algebra Applications* **1994**, *212*, 3.
(66) Kohn, W.; Onffroy, J. *Phys. Rev. B* **1973**, *8*, 2485.
(67) Rehr, J. J.; Kohn, W. *Phys. Rev. B* **1974**, *10*, 448.
(68) Geller, M. R.; Kohn, W. *Phys. Rev. B* **1993**, *48*, 14085.
(69) Nenciu, A.; Nenciu, G. *Phys. Rev. B* **1993**, *47*, 10112.
(70) Ziman, J. M. *Models of Disorder: The Theoretical Physics of Homogeneously Disordered Systems;* Cambridge University Press: Cambridge, 1979.
(71) Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M.; White, C. A.; Baker, J.; Maurice, D. R.; Adamson, R. D.; Adams, T. R.; Kong, J.; Oumi, M.; Ishikawa, N. *Qchem Development version*; QChem, Inc.: Pittsburg, 1996.
(72) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571.
(73) Simandiras, E. D.; Amos, R. D.; Handy, N. C. *Chem. Phys.* **1987**, *114*, 9.
(74) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. H. *Chem. Phys. Lett.* **1989**, *162*, 165.
(75) Mulliken, R. S. *J. Chem. Phys.* **1962**, *36*, 3428.
(76) Davidson, E. R. *J. Chem. Phys.* **1967**, *46*, 3320.
(77) Roby, K. R. *Mol. Phys.* **1974**, *27*, 81.
(78) Heinzmann, R.; Ahlrichs, R. *Theor. Chim. Acta* **1976**, *42*, 33.
(79) Ehrhardt, C.; Ahlrichs, R. *Theor. Chim. Acta* **1985**, *68*, 231.
(80) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899.
(81) Mayer, I. *Chem. Phys. Lett.* **1995**, *242*, 499.
(82) Mayer, I. *J. Phys. Chem.* **1996**, *100*, 6249.
(83) Ruedenberg, K.; Schmidt, M. W.; Gilbert, M. M.; Elbert, S. T. *Chem. Phys.* **1982**, *71*, 41.
(84) Ruedenberg, K.; Schmidt, M. W.; Gilbert, M. M. *Chem. Phys.* **1982**, *71*, 51.
(85) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed.; Cambridge University Press: Cambridge, 1992.
(86) Golub, G. H.; Van Loan, C. F. *Matrix Computations;* Johns Hopkins University Press: Baltimore, MD, 1983.
(87) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.
(88) Baer, R.; Head-Gordon, M. *Phys. Rev. Lett.* **1997**. Submitted for publication.
(89) Ochsenfeld, C.; Head-Gordon, M. 1997. Manuscript in preparation.
(90) Demko, S.; Moss, W. F.; Smith, P. W. *Math. Comput.* **1984**, *43*, 491.
(91) Löwdin, P.-O., Ed. *Quantum Theory of Atoms, Molecules and the Solid State;* Academic: New York, 1966.

(92) Exponentially localized orbitals have been found for a one-dimensional periodic system with a point-defect[66] or surfaces.[67] Nearly periodic systems[68] and somewhat more general nonperiodic systems[69] have also been treated. It may well be impossible to find a simple relation between the HOMO−LUMO gap and the decay rate of the density matrix for an arbitrary system, since disordered systems do not possess a clear-cut band gap.[70]

(93) For examples of discussions of the role of minimal basis sets in performing and analyzing SCF calculations, see: refs 46, 75−84.

(94) The HOMO−LUMO gap may decrease further if diffuse Gaussians or low momentum plane waves are added to the basis.

(95) Stechel et al. have found that it is possible to exploit locality in some metals where only a small number of the occupied orbitals are delocalized.[17] Baer and Head-Gordon have recently developed a linear-scaling renormalization-group approach and applied it to a tight-binding Hamiltonian with a small band gap.[88]