New Methods for Ab Initio Quantum Mechanical Calculations in Molecular and Crystalline Systems

Thesis by

Jean-Marc Langlois

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1994

(Submitted October 7, 1993)

To Tania Lesk

Acknowledgments

I would to thank my advisor Bill Goddard for his encouragement and his enthusiasm in science. His support was always appreciated and his vast knowledge in chemistry, physics and materials science has been a constant source of inspiration.

I would also like to thank Murco Ringnalda for being a good friend and collaborator. Our many discussions have given momentum to my work. Xiaojie Chen has been a very valuable collaborator on the latter part of this thesis and his contributions were beneficial. I would like to thank Terumasa Yamasaki and Rick Muller for discussions and contributions to the first part of this thesis. I also had fruitful interactions in the earlier days with Yuejin Guo, Guanhua Chen and Naoki Karasawa.

I also want to acknowledge the following members of the group: Siddharth Dasgupta, Gilles Ohanessian, Kevin Plaxco, Jason Perry, Charles Musgrave, Anne Miller, Terry Coley and especially Jamil Tahir-Kheli for his own special way of mixing science with life.

Abstract

This thesis deals with the development of new methods for doing *ab initio* quantum mechanical calculations of electronic wavefunctions of large molecules and crystalline systems with the emphasis on inclusion of electronic correlation or many body effects using generalized valence-bond (GVB) wavefunctions.

Chapters 1 and 2 describe two necessary steps for using the generalized valence-bond (GVB) formalism in large molecular systems. In Chapter 1 a fast method for generating GVB trial wavefunctions is described. The method is based on piecewise atomic and diatomic localization and makes possible calculations with large numbers of GVB pairs. The efficacy of the method is illustrated by application to several cases including GVB wavefunctions with up to 26 pairs. In Chapter 2 the pseudospectral (PS) method for self-consistent-field calculations is applied to the GVB formalism. In the GVB perfect pairing approximation, the PS method is shown to reduce the scaling cost of the calculation from N^4 to N^3, where N is the number of basis functions. This makes possible the calculation of GVB wavefunctions for large molecular systems.

Chapter 3 describes a density-functional method for calculations on crystalline systems using Gaussian type orbitals. Accurate and efficient strategies were developed for computing both the Hamiltonian matrix elements and the Coulomb field. The Hamiltonian matrix elements are computed by decomposing the multicenter numerical integrations into single-center integrations via a projection technique and the Coulomb field is evaluated analytically using a dual-space approach based on the Ewald method. The self-consistent field is obtained by a fast conjugate gradient method which uses both first and second derivative information and an efficient preconditioning strategy. Illustrative calculations are performed on two allotropes of carbon: diamond and C_{60} crystals.

Table of Contents

Introduction

The relentless increase in computer power is making ab initio computer simulations a more viable tool for predicting the properties of new materials. Along with these hardware advances, dramatic improvements in theoretical methods and associated computer software are essential for prediction of accurate and interesting properties. This thesis describes some progress in developing these new methods for calculating electronic wavefunctions of large molecules and crystals.

In molecular systems, the Hartree-Fock method (HF), based on a simple single Slater determinant of spin orbitals, is the most common approach for doing ab initio calculations. In HF theory the Euler-Lagrange equation for optimum orbitals leads to a set of equations :

$$H^{HF}\psi_\alpha = \epsilon_\alpha \; \psi_\alpha \tag{1}$$

for the optimum orbitals $\{\psi_\alpha, \alpha = 1, ...N_{orb} = N_{elec}/2\}$. Each Fock operator:

$$H^{HF} = h + \sum_{\alpha=1}^{N_{orb}} (2J_\alpha - K_\alpha) \tag{2}$$

involves the field due to the electrons in occupied orbitals where J_α and K_α are referred to as the Coulomb and exchange operators. The HF approximation is useful for many purposes, however, it leaves out the correlation energy so important to chemical reactions, excited states, transition metals, etc. Ab initio approaches for estimating this correlation energy involve including additional configurations to the HF determinental wavefunction (each configuration involves various number of electrons excited from the occupied orbitals of the HF wavefunction to various unoccupied orbitals). For large systems, cost consideration demands that the number of such configurations be kept small. Thus it is important to choose a method which gets the best description of the system with the fewest configurations.

The simplest generalization to the HF method achieving such a compact description is the Generalized Valence Bond method[1-3] (GVB), where each pair of electrons of interest is described by a pair of orbitals instead of a single orbital as in HF. In most cases each pair of GVB orbitals localizes near a bond or an atom, and each GVB pair contributes to two configurations which describe the local correlation in that bond or atom.

The orbitals, ψ_a, ψ_b, of a GVB pair resemble hybridized valence bond orbitals and provide useful interpretation for complicated molecules (e.g., transition metals). However for computational purposes it is convenient to combine the GVB orbitals (which overlap) into orthogonal orbitals called GVB natural orbitals (NO) :

$$\psi_g = \frac{\psi_a + \psi_b}{\sqrt{2(1+S)}} \qquad \text{and} \qquad \psi_{\mathrm{u}} = \frac{\psi_{\mathrm{a}} - \psi_{\mathrm{b}}}{\sqrt{2(1-\mathrm{S})}} \tag{3}$$

where ψ_g resembles a localized HF bond orbital and ψ_u resembles a localized antibonding orbital. The two GVB orbitals lead to two Fock operators and the self-consistent equations to solve consist of a system of $\mathrm{N}_{GVB} + 1$ coupled eigenvalue equations where N_{GVB} is the number of GVB orbitals.

Much success has been obtained with the application of the GVB method to small molecules. Examples of this can be found in Appendix 1 and 2 where the GVB method was applied to cluster of atoms describing the important sites of the high Tc materials. The GVB formalism was used to study the local spin properties of the Cu atoms and to determine physical parameters relevant to understanding the origin of the superconductivity.

The extension of the GVB method to larger systems has been slowed by several difficulties:

(i) the need for an automatic method for generating good trial wavefunctions,

(ii) the need for a fast and reliable iterative approach for achieving self-consistency, and

(iii) the need for much faster construction of the GVB Fock operators.

We have addressed each of these issues.

The solution to the first problem is discussed in Chapter 1. For small molecules the initial GVB orbitals have traditionally been constructed by hand using chemical intuition. This was an acceptable solution because of the simplicity of the wavefunctions and the small number of degrees of freedom involved. For large molecules, this simple approach becomes cumbersome and may lead to orbitals too different from the self-consistent GVB orbitals. Chapter 1 describes an automatic method for generating GVB initial orbitals in large molecules using piecewise localization. For a pair of GVB orbitals localized on a bond, the first GVB natural orbital (i.e., the bonding orbital) is obtained by finding the pair of atomic orbitals that leads to the largest overlap onto the bond space. The second GVB natural orbital is the orthogonal anti-bonding combination. For GVB lone pairs, the first GVB orbital consists of a lobe-like orbital and the second GVB orbital is obtained by applying a localized gradient operator to the first GVB orbital. This procedure is automatic and only requires the selection of the bond or lone pair of interest.

The second issue is that for larger systems it becomes critical to have an efficient iterative approach to converge the GVB wavefunctions. For HF wavefunctions, the method of choice is the Direct Inversion in the Iterative Subspace (DIIS) method.[4] This method produces rapid convergence because it makes use of the gradient information obtained for all the SCF iterations. In the DIIS method an effective Fock operator is constructed from a linear combination of the previous Fock operators where the coefficients are obtained by minimizing the corresponding linear combinations of the gradient vectors. Previously the method had been limited to wavefunctions with only one GVB pair because of the difficulty in finding an appropriate effective Fock operator. By incorporating the proper orbital mixing terms into a multi shell Fock operator, the DIIS method has now been generalized to wavefunctions with any number of closed shell orbitals, open shell orbitals, and GVB orbitals.[5]

The third issue to be resolved is the efficient construction of the Fock operators. The molecular orbitals or band orbitals, $\{\psi_\alpha\}$, are described in term of basis functions, $\{\phi_i\}$:

$$\psi_\alpha = \sum_i c_{i,\alpha}\phi_i \tag{4}$$

where the expansion coefficients, $\{c_{i,\alpha}\}$, determine the optimum orbitals. In molecular systems, the most common basis functions are atomic Gaussian-type orbitals (GTO). The use of a basis set converts the Fock *operator* into Fock *matrix*, H_{ij}. In standard HF theory, the Fock matrix is evaluated in term of one-electron and two-electron integrals as

$$H_{ij} = h_{ij} + 2J_{ij} - K_{ij} \tag{5}$$

where h_{ij} is a one-electron matrix element (kinetic energy plus nuclear attraction). The Coulomb J_{ij} and exchange K_{ij} matrices are given by :

$$J_{ij} = \sum_{kl} < ij||kl > D_{kl} \tag{6}$$

$$K_{ij} = \sum_{kl} < il||kj > D_{kl} \tag{7}$$

where $D_{kl} = \sum_{\alpha} C_{k\alpha} C_{l\alpha}$ is the density matrix and the two-electron integrals $< ij||kl >$ are defined as :

$$< ij||kl > = \int d^3r \int d^3r' \; \frac{\phi_i(r)\,\phi_j(r)\,\phi_k(r')\,\phi_l(r')}{|r - r'|}. \tag{8}$$

In this formulation of HF theory, the overall computational effort scales as N^4 where N is the number of GTOs used to represent the orbitals.

In contrast to ab initio methods, the application of density functional theory (DFT) to molecular systems[6] leads to a scaling of N^3 when using GTOs . This is achieved by fitting the density $\rho(r)$ and exchange-correlation potential $V_{xc}(r)$ with an auxiliary set of GTOs. This approach cannot be applied to HF theory because of the the non-local nature of the exchange operator, K.

A solution to this N^4 problem for ab initio theory is the pseudospectral (PS) method.[7–8] With PS both numerical grids and Gaussian basis functions are used to describe the orbitals.[8] The Coulomb and exchange matrices are given by :

$$J_{ij} = \sum_{g} Q_i(r_g)\,\phi_j(r_g)(\sum_{kl} 2D_{kl}\,A_{kl}(r_g)) \tag{9}$$

$$K_{ij} = \sum_{g} Q_i(r_g)\,(\sum_{k} A_{jk}(r_g)\,(\sum_{l} 2D_{kl}\,\phi_l(r_g))) \tag{10}$$

where the sum $\sum_g$ is over the physical space gridpoints and the one-electron field integrals are given by:

$$A_{jk}(r_g) = \int d^3r \; \frac{\phi_j(r)\,\phi_k(r)}{|r - r_g|}. \tag{11}$$

The matrix $Q_i(r_g)$ is used to transform a function represented on the grid onto a function represented by the basis set. In practice, the matrix $Q_i(r_g)$ is a least squares fitting operator, which is designed to fit any product of the form $A_{kl}(r_g)\,\phi_j(r_g)$.

With PS the construction of the Fock matrices (5) scales as N^3 rather than N^4. The number of grid points scales as the number of atoms and hence as N. This leads to N^3 work in evaluating Coulomb and exchange matrix elements in equations (9) and (10). Since the various integrals decrease with distance, the use of an integral cutoff reduces the scaling to N^2. The application of the PS method to the GVB formalism is similar to its application to HF and the details are given in Chapter 2.

In Chapter 2 the least squares fitting operators $Q_i(r_g)$ are constructed with the use of a fitting basis $X_q(r_g)$:

$$Q_i(r_g) = \sum_{pq} S_{ip}\,[S_{pq}^{num}]^{-1}\,X_q(r_g)\,w(r_g) \tag{12}$$

where $w(r_g)$ is the weight of the gridpoint, S_{ip} is the overlap integral between basis function ϕ_i and fitting function X_p, and S_{pq}^{num} is the *numerical* overlap integral between two fitting functions :

$$S_{pq}^{num} = \sum_{g} X_p(r_g)\,X_q(r_g)\,w(r_g). \tag{13}$$

When the basis function ϕ_i is short range (SR) it is advantageous to restrict the fitting to the region of space where the basis function is non-zero. A very complete basis set localized in the same region is used as the fitting basis set. The fit is constrained by multiplying the weight by an envelope function centered on the same atom as basis function ϕ_i.

When the basis function is long range (LR), then the fit needs to be done over the whole molecule. Hence the fitting basis set includes functions on the whole molecule as well. For small molecules this is not a problem but for large molecules this becomes prohibitive because of the matrix inverse operation required in forming $Q_i(r_g)$ which scales as the number of fitting functions to the third power, M^3. Since $M > N$ this step can easily dominate the overall calculation.

The solution to this problem results from the realization that the fitting operator is always more precise for SR functions because the fit can be of higher quality. This allows the Coulomb and exchange matrices, equations (9) and (10), to be rewritten in such a way that the basis function index of $Q(r_g)$ is a SR function. This can always be done except when all four basis functions in a two-electron integral are LR. In that case the fitting operator $Q(r_g)$ needs only to include LR fitting functions since all the SR components have been removed from the quantity to be fit. This makes the cost of constructing this LR fitting operator practical because the number of LR fitting functions is of the same order as the number of basis functions. This length scale algorithm has been implemented successfully. Coupled with the use of diatomic corrections introduced in Chapter 2, this leads to an efficient method for doing ab initio calculations with HF or GVB wavefunctions on large molecules.[9]

Unfortunately the added complexity required by these improvements have precluded the application of the PS method to crystalline systems. Because of the length scale algorithm and the diatomic corrections, the exchange and Coulomb fields are not constructed completely and this greatly complicates the handling of the lattice sums in crystalline systems.

For the Coulomb field the lattice sums do not converge in real space because of the long range $1/r$ tail. This is usually dealt with by screening the Coulomb field.

However one cannot do this in the present framework of the PS method because the Coulomb field cannot be separated into short range and long range pieces.

For the exchange field, the lattice sums are convergent but the exchange field becomes more problematic because it decays slowly. Because of its non-local nature, the exchange field cannot be screened and hence it decays more slowly than a screened Coulomb field where the long range $1/r$ component has been removed.

At this point it is useful to reexamine the features required in order to apply the GVB formalism to crystalline systems.

(i) use of localized basis functions (GTOs) to describe the GVB orbitals,

(ii) accurate evaluation of the Fock matrix elements without using density or field fitting in order to handle the non-local HF exchange operator, and,

(iii) effective screening of the Coulomb field.

The simplest method that encompasses the first two requirements is a scheme where the fields are evaluated on a grid and the matrix elements of the Fock matrix are evaluated by numerical integration over the grid points. Historically, numerical integrations in ab initio calculations have been avoided because of the difficulty in reaching sufficient accuracies. It is only recently[10−12] that good accuracy has been achieved for reasonable grid sizes in molecular and crystalline systems.

The solutions proposed for crystalline systems[11−12] have been based on the division of space into regions of different shapes over which independent integrations are performed. Three different regions are used: spherical regions around atomic nuclei, regular polyhedra in the interatomic region and truncated polyhedra to make the transition between the other two regions. The location and weight of the gridpoints are relatively easy to select in the case of the spherical regions and regular polyhedra but they are very difficult to choose for the truncated polyhedra.

A different approach, which avoids the truncated polyhedra regions, consists in transforming the integration over space into atomic integrations by using atomic projection functions.[10] The method has previously been employed for small molecules. We have generalized its use to crystalline systems. In crystalline systems, the larger number of neighbors makes the integration less accurate. By adding a new functional form to the projection function, the accuracy was significantly improved. The atomic integrations were also modified to better take advantage of the Gaussian nature of the basis set used. The description of grid generation along with accuracy results are given in Chapter 3.

In Chapter 3, a solution to the evaluation of the Coulomb field in crystalline systems is presented. Unlike density fitting approaches which tends to spherically average out the field around each atom, the current approach evaluates the Coulomb field by an exact method based on a perfect screening of the field.[13] For each pair of GTOs a smooth screening field is defined which exactly cancels out the long range part of the field. Like in the Ewald method, the screening field is added and subtracted and this leads to an efficient dual-space approach. The combined normal field minus screening field now rapidly converges in real space, whereas the field of the opposite screening density converges rapidly in reciprocal space. This dual-space method for evaluating the Coulomb field scales as $N^{1.5}$ just like the Ewald method.

The soundness of previous methods for evaluating the Fock matrix elements and the Coulomb field in crystalline systems were tested on crystalline calculations of diamond and C_{60}. At this point the HF exchange field was replaced by the exchange-correlation field of density functional theory in the local density approximation (LDA) and comparison with other LDA calculations using plane waves (PW) and GTOs were made.

VII. References

1. F. W. Bobrowicz and W. A. Goddard III, in *Modern Theoretical Chemistry: Methods of Electronic Structure Theory*, edited by H. F. Schaefer III (Plenum, New Yok, 1977), Vol. 3, p. 79.
2. E. A. Carter and W. A. Goddard III, J. Phys. Chem. **92**, 2109 (1988).
3. S. K. Shin and W. A. Goddard III, J. Chem. Phys. **93**, 4986 (1990).
4. T.P. Hamilton and P. Pulay, *J. Chem. Phys.* **84**, 5728 (1986).
5. R.P. Muller, J.-M. Langlois, M.N. Ringnalda, R.A. Friesner and W.A. Goddard III, *J. Chem. Phys.*, (in press).
6. J. Andzelm, in *Density Functional Methods in Chemistry*, J. Labanowski and J. Andzelm, eds., Springer-Verlag, Berlin, 1992.
7. S.A. Orszag, *Stud. Appl. Math.* **51**, 253 (1972); D. Gottlieb and S. Orszag, *Numerical Analysis of Spectral Methods; Theory and Application* (SIAM, Philadelphia, 1977).
8. R.A. Friesner, *J. Chem. Phys.* **86**, 3522 (1986); R.A. Friesner, *Ann. Rev. Phys. Chem* **42**, 341 (1991)
9. B.H. Greeley, T.V. Russo, D. Mainz, R.A. Friesner, J.-M. Langlois, W.A. Goddard III and M.N. Ringnalda, (in preparation).
10. A.D. Becke, *J. Chem. Phys.* **88**, 2547 (1988).
11. M.R. Pederson and K. A. Jackson, *Phys. Rev.* **B 41**, 7453 (1990).
12. G. te Velde and E.J. Baerends, *J. Comp. Phys.* **99**, 84 (1992)
13. C. Pisani, R. Dovesi and C. Roetti, "Hartree-Fock Ab Initio Treatment of Crystalline Systems," *Lecture Notes in Chemistry* Vol. 48 (1988).

Chapter 1

Generating Trial Functions for Generalized Valence Bond Wavefunctions

Abstract

A common difficulty with obtaining wavefunctions for Generalized Valence Bond (GVB) and other multiconfigurational self-consistent field (MC-SCF) wavefunctions is finding suitable initial guesses or trial functions. We present a general method (GVB-INIT) for obtaining such initial guesses. This method uses pseudo-HF molecular orbitals formed from HF *atomic* orbitals. The occupied HF orbitals are projected onto atomic basis functions to obtain GVB first natural orbitals and unoccupied HF orbitals are projected to obtain GVB second natural orbitals. We find for a variety of systems that these trial functions lead to excellent convergence and exhibit good overlap with the optimum GVB orbitals. Constructing the GVB-INIT wavefunction is fast because HF wavefunctions are not calculated and the localization is piecewise atomic.

In conjunction with the recently developed GVB-DIIS method for converging GVB wavefunctions and new pseudospectral programs (PS-GVB) for the Fock matrix elements, GVB-INIT makes calculation of highly correlated GVB wavefunctions now quite practical. The efficacy of GVB-INIT is illustrated by application to several cases including GVB wavefunctions with up to 26 pairs. It can also be used for other MC-SCF wavefunctions such as CAS-SCF.

I. Introduction

Essentially all methods for calculating the electronic wavefunctions (i.e., solving the Schrödinger equation) of atoms, molecules and solids use a first step involving optimization of orbitals (atomic orbitals, molecular orbitals, or band orbitals). The most common starting point for accurate wavefunctions is the Hartree-Fock (HF) method in which a product of orbitals and spins is antisymmetrized (using a Slater determinant) and the orbitals optimized self-consistently (SCF for self-consistent field). This might then be used as the starting point for obtaining more accurate wavefunctions [configuration interaction (CI) or perturbation theory] by including excitations where various numbers of electrons are excited from the occupied orbitals of the HF wavefunction to unoccupied or virtual orbitals.

An alternative approach is the Generalized Valence Bond method[1-3] (GVB) in which each *pair* of electrons is described with *two* orbitals (rather than one, as in HF), the wavefunction is generalized from a Slater determinant to a form that ensures that both the Pauli principle and spin symmetry are satisfied, and the orbitals are optimized self-consistently. In general, this GVB approach leads to twice the number of orbitals to be optimized as HF. Generally each pair of GVB orbitals localize uniquely near a bond or an atom and can be viewed as involving one orbital (the first NO for natural orbital) very similar to a localized occupied HF orbital and a second orbital (the second NO) that corresponds to a localized unoccupied (virtual) HF orbital. As with HF, many GVB calculations are followed by some kind of CI. However, since the dominant correlating orbitals have been solved self-consistently, the GVB wavefunction often allows[17] an accurate description of the wavefunction to be obtained with very compact CI's.

The two-electron interactions lead to nonlinearity in the optimization of wavefunctions. Critical to such SCF methods as HF or GVB is a prediction of good trial functions followed by an iterative approach that converges quickly and reliably to the desired state.

For HF wavefunctions the trial function is generally taken from an approximate or semi-empirical HF-based method such as extended Hückel, CNDO/INDO, MINDO, etc. This calculation is very fast and generally leads to adequate trial orbitals. However, for GVB there is not yet an analogous semi-empirical method. Previously GVB calculations have involved constructing localized HF orbitals (both occupied and unoccupied). However, these methods tend to be unreliable and not automatic.

In addition, the HF and localization calculations can be almost as expensive as the GVB calculations, making the total process unacceptably costly.

We present here an approach, GVB-INIT, that solves the problem of making reliable initial guesses, providing excellent trial functions for general GVB wavefunctions. Combined with the GVB-DIIS method for converging these wavefunctions,[4] this leads to an automatic procedure for calculating GVB wavefunctions.

II. The GVB-PP Wavefunction

The GVB[5] wavefunction generally involves a product of orbitals with a general spin function χ that is an eigenfunction of the total spin operator $[\hat{S}^2\chi = S(S+1)\chi]$

$$\Phi^{GVB} = \mathcal{A}[\phi_i \cdots \phi_N \chi] \tag{2.1}$$

where both the orbitals and spin functions are optimized. More commonly, the Perfect Pairing (PP) restriction[6] is made in which each pair of electrons is written as a valence bond (VB) pair

$$(\phi_a\phi_b + \phi_b\phi_a)\,(\alpha\beta - \beta\alpha)\,. \tag{2.2}$$

The GVB-PP formalism allows any number of electrons ($2n_{core}$) to be described with doubly-occupied orbitals (Φ_{core}) and any number of electrons, (n_{open}) to be high-spin coupled ($S = n_{open}/2$), leading to the composite wavefunction

$$\Psi = \mathcal{A}\left[\Psi_{core}\Psi_{open}\Psi_{pair}\right], \tag{2.3}$$

where $\mathcal{A}$ is the antisymmetrizer,

$$\Psi_{core} = \prod_{i=1}^{n_{core}} (\phi_i^c\alpha)(\phi_i^c\beta) \tag{2.4}$$

$$\Psi_{open} = \prod_{i=1}^{n_{open}} (\phi_i^o\alpha) \tag{2.5}$$

$$\Psi_{pair} = \prod_{i=1}^{n_{pair}} (\psi_{ai}\psi_{bi} + \psi_{bi}\psi_{ai})\,(\alpha\beta - \beta\alpha)\,, \tag{2.6}$$

where $\{\phi_i^c\}$, $\{\phi_i^o\}$ are orthogonal spatial orbitals and where α and β are spin functions. Here ψ_{ai} and ψ_{bi} denote the GVB orbitals in pair i. Within a pair, the GVB orbitals overlap, $S_i =< \psi_{ai}|\psi_{bi} >$. However, in the PP restriction we take the orbitals of pair i as orthogonal to all other pairs (and to $\{\phi_i^c\}$, and $\{\phi_j^o\}$). For computational purposes it is convenient to rewrite the GVB wavefunction (2.6) as

$$\Psi_{pair} = \prod_{i=1}^{n_{pair}} (C_{gi}\phi_{gi}\phi_{gi} + C_{ui}\phi_{ui}\phi_{ui})\,(\alpha\beta - \beta\alpha)\,, \tag{2.7a}$$

where

$$\phi_{gi} = \frac{(\psi_{ai} + \psi_{bi})}{\sqrt{2(1+S_i)}} \qquad \text{and} \qquad \phi_{\mathrm{ui}} = \frac{(\psi_{\mathrm{ai}} - \psi_{\mathrm{bi}})}{\sqrt{2(1-\mathrm{S_i})}} \tag{2.7b}$$

are orthogonal GVB natural orbitals. Here the GVB Configuration Interaction (CI) coefficients satisfy

$$\frac{C_{ui}}{C_{gi}} = -\frac{1 - S_i}{1 + S_i} \tag{2.8}$$

and

$$C_{gi}^2 + C_{ui}^2 = 1. \tag{2.9}$$

III. Initial Guesses for GVB Wavefunctions, GVB-INIT

A. Introduction

The generation of the trial wavefunction is composed of two steps.

(*i*) In the first step a HF trial wavefunction is generated based on *atomic* SCF orbitals.

(*ii*) In the second step, the GVB trial orbitals are obtained by piecewise localization of the HF trial wavefunction.

This is our standard GVB initial guess. Alternatively, one could use (*i*) as the starting point for a molecular HF-SCF calculation, which could in turn be used in place of the HF trial function for step two. However, use of the HF optimum orbitals in (*ii*) generally saves only about one or two iterations in the GVB-SCF calculation, and sometimes gives worse results for highly polar or stretched bonds. Hence the extra expense of the HF SCF calculation is not justified and is *not* a part of GVB-INIT.

We find that the HF-SCF orbitals of the individual atoms can be used to form an HF trial wavefunction for a molecule. This is plausible since the molecular density is not dramatically different from the sum of the atomic densities. The problem is how to go from a set of atomic SCF orbitals (SCF-AO) to a set of molecular orbitals (MO) without actually solving the HF equations for the molecular orbitals.

The procedure we use is:

(α): Atomic HF orbitals $\{\theta_n^A\}$: For each atomic basis set $\{\chi_\mu^A\}$, where A is the atom label, we define orthonormal HF atomic orbitals $\{\theta_n^A\}$ obtained from occupation-averaged atomic Hamiltonians (so that orbitals of the same symmetry are equivalent). Thus an oxygen atom would have 4/3 electrons in each p orbital. This atomic calculation is done once and stored with the basis set. The atomic occupation number $\{f_n^A\}$ is associated with each HF-AO. The orbitals are then sorted into core orbitals, $\{\theta_n^{core,A}\}$, valence orbitals, $\{\theta_n^{val,A}\}$, and unoccupied orbitals, $\{\theta_n^{unocc,A}\}$. The core orbitals are (1s) for first row elements and (1s,2s,2p) for second row elements. The valence orbitals are (1s) for hydrogen atoms, (2s,2p) for the first row elements and (3s,3p) for second row elements.

(β): VB Assignment: We next partition the electrons of the system in terms of bond pairs, lone pairs, core orbitals just as in a simple valence bond structure. This can be done in several ways: (*i*) automatically using the autotyping features of commercial molecular dynamics software, (e.g., BIOGRAF/POLYGRAF[7]), (*ii*) by drawing a molecule (with bonds) on a screen, or (*iii*) by inputting a list of atoms to be involved in various bonds, lone pairs, etc. GVB-PP is oriented toward molecules with one dominant valence bond structure, and we consider only this case herein. GVB does *not* require that every bond pair be correlated. In this case one designates which bonds are to be correlated and second natural orbitals (step δ4 below) are formed only for these.

(γ): Build occupied HF-MO's $\{\phi_n^{HF}\}$:

(γ1): Core HF MO's: each core HF-AO $\{\theta_n^{core}\}$ is considered as a doubly-occupied HF MO $\{\phi_n^{core}\}$

(γ2): Valence HF MO's: The remaining valence HF-AO's $\{\theta_n^{val}\}$ are used to construct a subset of HF-MO's $\{\phi_n^{val}\}$ as follows: We construct the overlap matrix over all valence HF-AO's $\{\theta_n^{val}\}$,

$$S_{ij} = \langle\theta_i^{val}|\theta_j^{val}\rangle; \tag{3.1}$$

this matrix is diagonalized and the eigenvectors with the largest eigenvalues are selected. The number to be selected, n_{occ}, is obtained by adding the atomic occupation numbers for the various atoms and correcting by the total charge of the molecule. The resulting $\{\phi_n^{val}\}$, which are not orthogonal, include the combinations of HF-AO having the largest overlap, a good first approximation to the best bonding orbitals. We refer to this as the HF-INIT space. It is an approximation to the HF wavefunction of the molecule (which we do *not* calculate).

(γ3): Orthonormalize the HF-MOs $\{\phi_n\}$. Since the core orbitals were not included in the diagonalization of the overlap matrix in $\gamma2$, we now order the $\{\phi_n\}$ with the core orbitals first (ordered by atomic orbital energy) and then the valence orbitals (ordered by eigenvalue from diagonalizing the valence *AO* overlap matrix). This set of occupied orbitals is then Schmidt orthogonalized.

(γ4): Create a new set of orthogonal orbitals $\{\eta_n^A\}$ on each atom. Here we want the atomic orbitals ordered according to their projection onto the initial HF-MO space (the occupied HF orbitals). To do this we construct the atomic density matrix over AO

$$P_{\mu\nu}^{AA} = \langle \chi_\mu^A | \rho_{HF} | \chi_\nu^A \rangle, \tag{3.2}$$

where ρ_{HF} is the electron density of the occupied HF-MO's

$$\rho_{HF} = \sum_{n=1}^{n_{occ}} | \phi_n^{HF} \rangle \langle \phi_n^{HF} | \, . \tag{3.3}$$

The associated eigenvalue problem to be solved is:

$$P^{AA} \; c \;=\; S^{AA} \; c \, \gamma^{AA} \tag{3.4a}$$

where

$$S_{\mu\nu}^{AA} = \langle \chi_\mu^A | \chi_\nu^A \rangle \tag{3.4b}$$

is the overlap matrix over atomic basis functions. The HF-projected orthogonal orbitals $\{\eta_n^A\}$ are associated with the eigenvectors $\{c_n\}$. Near unity eigenvalues indicate large projections onto the HF occupied orbital space. At this point the orbitals in $\{\eta_n^A\}$ are sorted by projections into core orbitals, $\{\eta_n^{core,A}\}$, valence orbitals, $\{\eta_n^{val,A}\}$, and unoccupied orbitals, $\{\eta_n^{unocc,A}\}$. The valence atomic orbitals will define the space used to create the GVB orbitals for both bond pairs (section III.B) and lone pairs (section II.C). For the 2nd GVB natural orbitals, the unoccupied atomic orbitals will be used (section III.C).

B. Bond Pairs

GVB-INIT is based on piecewise localization approach to obtain trial GVB orbitals. Each pair of GVB orbitals is generated independently. They are later combined by performing a symmetric orthogonalization.

The most common case is a GVB bond-pair localized between two nearest-neighbor atoms. In this case there are two GVB orbitals [ψ_{ai} and ψ_{bi} of (2.6)], one localized on each atom, which overlap. We then combine these to obtain the GVB natural orbitals [ϕ_{gi} and ϕ_{ui} of (2.7)], which are orthogonal.

The task here is to choose the localized GVB natural orbitals as suitable linear combinations of HF-MO's. The most important characteristic of the GVB orbitals forming a bond pair, (ψ_{ai} and ψ_{bi}), is the large overlap between them. The GVB orbitals can thus be found by searching for the linear combination of the valence atomic projected HF orbitals $\{\eta_n^{val,A}\}$ (which will form ψ_{bi}) and $\{\eta_n^{val,B}\}$ (which will form ψ_{bi}) so that ψ_{ai} and ψ_{bi} have maximum probability of sharing electrons.[18] The procedure is as follows:

(δ): Consider one by one each bond and its pair of atoms (A and B). We will obtain one pair of orbitals $\{\psi_{ai}\}$ and $\{\psi_{bi}\}$ for each bond between A and B.

(δ1): Evaluate the off-diagonal density matrix elements of the occupied HF-MO's in the basis of the orthogonal valence atomic orbitals $\{\eta_i^{val,A}\}$ and $\{\eta_j^{val,B}\}$. This defines the bond space from which we create the overlapping GVB orbitals:[11]

$$P_{ij}^{AB} = \langle \eta_i^{val,A} | \rho_{HF} | \eta_j^{val,B} \rangle. \tag{3.5}$$

(δ2): Find the pair of unitary transformations, $\mathbf{U}^A$ and $\mathbf{U}^B$, that makes each orbital on atom A overlap with only one orbital on atom B. This biorthogonalization process[12] transforms the density matrix $\mathbf{P}^{AB}$ into a diagonal matrix $\mathbf{\Lambda}^{AB}$

$$(U^A)^t P^{AB} U^B = \Lambda^{AB} \tag{3.6}$$

where the largest diagonal element of $\mathbf{\Lambda}$ indicates the largest overlap for the pair of localized orbitals (one each on atom A and B). Since the matrix $\mathbf{P^{AB}}$

is always non-symmetric, equation (3.6) is solved by singular-value decomposition.

($\delta 3$): Starting with the largest singular-value of Λ^{AB}, select GVB-like orbitals ($\bar{\psi}_{ai}$ and $\bar{\psi}_{bi}$) as biorthogonal pairs. [Here the bar indicates that the $\bar{\psi}_{ai}$ contains only basis functions on center A.] For multiple bonded cases, select the appropriate number of pairs. Generally the highest Λ^{AB} corresponds to a σ-bond, the next two largest eigenvalues correspond to π-bonds, the next two to δ bonds etc.

($\delta 4$): The two overlapping GVB orbitals of each pair ($\bar{\psi}_{ai}$ and $\bar{\psi}_{bi}$) from $\delta 3$ are combined to form the GVB natural orbitals [ϕ_{gi} and ϕ_{ui} of (2-7)]. The actual GVB orbitals (ψ_{ai} and ψ_{bi}) will be linear combinations of the atomic $\bar{\psi}_{ai}$ and $\bar{\psi}_{bi}$ ($\psi_{ai} = \bar{\psi}_{ai} + \lambda_a \bar{\psi}_{bi}$ and $\psi_{bi} = \psi_{bi} + \lambda_b \bar{\psi}_{ai}$ where $\lambda_i \simeq 0.1$) and thus the GVB-NO's, (ϕ_{gi} and ϕ_{ui}), will be combinations of $\bar{\psi}_{ai}$ and $\bar{\psi}_{bi}$ as in (2.7) but with coefficients that are not ± 1. The 1st GVB-NO is written as $\phi_{gi} = c_a \bar{\psi}_{ai} + c_b \bar{\psi}_{bi}$ where the coefficients c_a, c_b are chosen to produce the largest projection onto the occupied HF space. The 2nd GVB-NO, ϕ_{ui}, is just the orthogonal linear combination. This is done by solving the 2×2 corresponding eigenvalue problem:

$$P^{GVB} \; c \; = \; S^{GVB} \; c \; \lambda^{GVB} \tag{3.7a}$$

where

$$P^{GVB}_{\alpha\beta} = \langle \bar{\psi}_{\alpha i} | \rho_{HF} | \bar{\psi}_{\beta i} \rangle \tag{3.7b}$$

and

$$S^{GVB}_{\alpha\beta} = \langle \bar{\psi}_{\alpha i} | \bar{\psi}_{\beta j} \rangle . \tag{3.7c}$$

Generally the projection of the 1st GVB-NO is very large ($\lambda_{1st}^{GVB} > 0.95$) and we replace the GVB-NO by its projection to the occupied HF-MO space. On the other hand the projection of the 2nd GVB-NO is very small. Thus the 1st GVB-NO can be viewed as a linear combination of the initial occupied HF-MO's whereas the 2nd GVB-NO corresponds to some linear combination of the initially unoccupied MO's.

C. Lone Pairs

For the GVB wavefunction of a molecule well described in terms of a single valence bond structure, the above procedure leads to the starting guess for a full GVB wavefunction. In this case lone pair orbitals (and core orbitals) of the wavefunction remain doubly occupied (not correlated). However, for many purposes it is also desirable to correlate the lone pairs. In this case, the 1st GVB-NO consists of a localized lobe-like orbital (say, an sp^3-like hybrid corresponding to an orbital centered off the atom). The 2nd GVB-NO has the same overall direction and location, but with a nodal plane passing through the middle of the first NO, leading to an off-center p-like orbital. With GVB-INIT we choose the lobe 1st GVB-NO from a linear combination of the atomic valence orbitals $\{\eta_n^{val}\}$ not previously used in making GVB bond-pair orbitals. The lobe 2nd GVB-NO is chosen as the linear combination of unoccupied atomic orbitals $\{\eta_n^{unocc}\}$ having the largest gradient matrix element with the lobe 1st GVB-NO. The overall method is summarized as follows:

(ϵ1): Evaluate the overlap matrix between the valence orthogonal atomic orbitals $\{\eta_i^{val,A}\}$ in the subspace of the GVB bond-pair orbitals involved with this atom:

$$P_{ij}^{AA} = \delta_{ij} - \langle \eta_i^A \left| \sum_{n=1}^{n_{pair}} |\psi_{gn}^A\rangle\langle\psi_{gn}^A| \right| \eta_j^A \rangle \tag{3.8}$$

where $\{\psi_{gn}^A\}$ are the bond-pair 1st GVB-NO's involved with atom A. The matrix $\mathbf{P}^{AA}$ is diagonalized and the appropriate number of lobe 1st GVB-NO's, $\{\psi_{gi}^{lobe}\}$, are chosen as the eigenvectors with the largest eigenvalue. If there is more than one (e.g., O, F), they are Boys localized to obtain sp^3-like orbitals.

(ϵ2): The lobe 2nd GVB-NO's, $\{\psi_{ui}^{lobe}\}$, are formed by linear combinations of unoccupied orthogonal atomic orbitals $\{\eta_i^{unocc,A}\}$ having the maximum overlap with the gradient of the 1st GVB-NO's:

$$\langle \psi_{ui}^{lobe} \left| \frac{\vec{R}_{lobe} \cdot \vec{\nabla}}{|r - R_{lobe}|^3} \right| \psi_{gi}^{lobe} \rangle \tag{3.9}$$

where the $\vec{R}_{lobe}$ are the centroids of the lobe 1st GVB-NO's:

$$\vec{R}_{lobe} = \langle \psi_{gi}^{lobe} | \vec{r} | \psi_{gi}^{lobe} \rangle . \tag{3.10}$$

Empirically, the $1/|r - R_{lobe}|^3$ term was found useful in restricting the effect of the gradient to a small region of space near the centroid location. For computational simplicity the $1/r^3$ term was approximated by an expansion over eight Gaussian s functions.

D. Symmetric Orthogonalization

So far we have described how to obtain each pair of GVB natural orbitals independently. We must now merge these GVB orbitals with the existing HF orbitals. The procedure is as follows.

(λ1): Symmetrically orthogonalize all the 1st GVB-NO's $\{\psi_{gi}\}$; then symmetrically orthogonalize all the 2nd GVB-NO's $\{\psi_{ui}\}$.

(λ2): Form the remaining occupied orbitals by projecting out the bond and lone pair 1st GVB-NO's $\{\phi_{gn}^{GVB}\}$ from the initial set of HF-MO's, $\{\phi_i^{HF}\}$. This is done by selecting the near unity eigenvalues of the following matrix:

$$P_{ij}^{HF} = \delta_{ij} - \langle \phi_i^{HF} \left| \sum_{n=1}^{n_{pair}} |\psi_{gn}^{GVB}\rangle\langle\psi_{gn}^{GVB}| \right| \phi_j^{HF} \rangle. \tag{3.11}$$

(λ3): Combine all occupied GVB orbitals. The sequence is:

i. all doubly occupied orbitals (core, uncorrelated lobe and bonding orbitals),

ii. all the 1st GVB-NO's $\{\psi_{gi}^{GVB}\}$,

iii. any singly occupied (high-spin) molecular orbitals,

iv. all 2nd GVB-NO's $\{\psi_{ui}^{GVB}\}$.

(λ4): Schmidt orthogonalize the 2nd GVB-NO's to all doubly occupied and redo the symmetric orthogonalization of the 2nd GVB-NO's to remove any overlap between the 2nd GVB-NO's introduced by the Schmidt orthogonalization.

In steps λ1 and λ4 it is important to use symmetrical orthogonalization in order to preserve any symmetry between the GVB orbitals. The 1st GVB-NO's are symmetrically orthogonalized independently of the 2nd GVB-NO since the former have much larger occupation.

In summary, the notation used in the above steps is as follows:

i. The atomic basis functions are $\{\chi_\mu\}$.

ii. The atomic HF-SCF orbitals are $\{\theta_n^A\}$ where f_n^A are the occupation numbers. Step α.

iii. The molecular orbitals of the HF-INIT wavefunction are $\{\phi_n\}$. This contains only the occupied MO's for the HF wavefunctions. Steps $\gamma 1, \gamma 2, \gamma 3$.

iv. Atomic projections of the HF-MO's $\{\eta_n^A\}$, sorted into core, valence and unoccupied orbitals according to occupations. Step $\gamma 4$.

v. Bond Pairs (if correlated); 1st and 2nd GVB-NO $\{\psi_{gi}^{GVB}\}$ and $\{\psi_{ui}^{GVB}\}$ formed from valence orbitals $\{\eta_n^A\}$ and $\{\eta_n^B\}$ with maximum overlap. Steps $\delta 1, \delta 2, \delta 3, \delta 4$.

vi. Lone Pairs (if correlated); 1st GVB-NO $\{\psi_{gi}^{lobe,A}\}$ selected from valence orbitals $\{\eta_n^{val,A}\}$ after eliminating $\{\psi_{ui}^{GVB}\}$ on center A; 2nd GVB-NO $\{\psi_{ui}^{lobe,A}\}$ obtained from unoccupied orbitals $\{\eta_n^{unocc,A}\}$ by maximizing overlap with gradient operator applied to $\{\psi_{gi}^{lobe,A}\}$. Steps $\epsilon 1$ and $\epsilon 2$.

vii. Symmetrically orthogonalize with highest occupations first (steps $\lambda 1, \lambda 2, \lambda 3, \lambda 4$).

Computationally, the method described above is very efficient. The most time consuming steps are the construction of the projection operator over the HF orbitals, the construction and diagonalization of the matrix $\mathbf{P}^{HF}$ of (3.11), and the final Schmidt orthogonalization. All three steps scale as M^3. Note that this method does not require a global localization of all the orbitals of the molecule[13] (an iterative process) but only piecewise noniterative localization based on simple chemical bonding ideas.

IV. Results and Discussion

The GVB-INIT method as described above has been used as the starting point for a number of studies of various wavefunctions. In particular, a recent paper[4] on GVB-DIIS utilizes it for a number of systems, including glycine correlated with up to 10 GVB pairs. In this paper we have concentrated on making comparisons between the GVB-INIT wavefunctions and the converged wavefunctions in order to assess the soundness of the approach proposed.

A. Hartree-Fock SCF Calculations

In Table 1 we compare the efficacy of the HF initial guess method defined in section III.A (called HF-INIT hereafter) with that of the standard initial guess from GAUSSIAN90 (G90-INIT). All calculations used the 6-31G** basis sets with the geometries specified in Table 1. As a measure of the efficacy, we evaluate the overlap population between the initial guess wavefunction and the SCF wavefunction. The overlap population, n_{guess}, is defined as follows:

$$n_{guess} = \sum_{i,j}^{n_{occ}} \langle \phi_i^{guess} | \phi_j^{HF-SCF} \rangle^2 \qquad (4.1)$$

where the $\{\phi_i^{guess}\}$ are occupied orbitals of the initial guess wavefunction and the $\{\phi_i^{HF-SCF}\}$ are occupied orbitals of the final HF-SCF wavefunction. Thus $0 \leq n_{guess} \leq n_{occ}$.

The test cases shown in Table 1 are for systems up to intermediate size such as porphine (with 81 occupied orbitals and 430 basis functions). Both methods do quite well with overlaps between the initial guess orbitals and the final SCF orbitals that are always greater than 98%. The average error per orbital is 1.2% for GAUSSIAN90, which is slightly better than the value of 1.7% for HF-INIT.

Given the simplicity of the HF-INIT method it is encouraging that the accuracy is almost as good as the semi-empirical method used in GAUSSIAN90, particularly since HF-INIT involves no adjustable parameters.

The success of HF-INIT validates the basic hypotheses:

i. the molecular density is not significantly different from the sum of the atomic densities in covalent systems, and

ii. the HF-INIT orbitals can be selected out of the space spanned by the valence atomic orbitals by choosing the linear combination of valence atomic orbitals with maximal overlap.

It should be noted that the HF-INIT orbitals are different from the orbitals obtained by diagonalizing the sum of atomic densities. In that case the sum of atomic densities is given by

$$\rho_{average} = \sum_{A=1}^{n_{atoms}} \sum_{n=1}^{n_{occ}(A)} \left|\theta_n^A\right\rangle f_n^A \left\langle\theta_n^A\right| \tag{4.2}$$

where $n_{occ}(A)$ is the number of occupied HF-AO on atom A. Here the associated initial guess orbitals are the eigenvectors of the corresponding eigenvalue problem:

$$P\,c = S\,c\,\gamma \tag{4.3a}$$

with

$$P_{\mu\nu} = \langle\chi_\mu \left|\rho_{average}\right| \chi_\nu\rangle, \tag{4.3b}$$

and

$$S_{\mu\nu} = \langle\chi_\mu|\chi_\nu\rangle, \tag{4.3c}$$

where the $\{\chi_\mu\}$ are the molecular basis functions. The occupied orbitals are chosen according to their eigenvalue which correspond to projection onto the space formed by $\rho_{average}$. Those density selected orbitals are different from the the HF-SCF orbitals by the same amount as the HF-INIT orbitals, i.e., an average orbital error of 1.7% when compared with the converged HF-SCF orbitals, but the average difference per orbital between the HF-INIT wavefunction and the wavefunction coming from the density selected orbtials is 1.3 %.

B. GVB-SCF Calculations

The GVB-INIT approach has been applied to a variety of molecules and wavefunctions, as summarized in Table 2. We use two criteria to assess the soundness of the approach.

1. **Orbital Space**:
    a) how well does GVB-INIT separate the space of core orbitals from the space of 1st GVB-NO's,
    b) how well does it separate the space of 2nd GVB-NO's from the space of unoccupied orbitals.

    As a measure we compare the initial guess orbitals to the final SCF orbitals. In this *orbital space* test we consider only the space of 1st GVB-NO's (or 2nd GVB-NO's) and do *not* distinguish whether or not the 1st GVB-NO's (or the 2nd GVB-NO's) are mixed together by some unitary transformation.

2. **Orbital Pairs**
    a) how well each 1st GVB-NO and each 2nd GVB-NO from GVB-INIT compares with the self-consistent GVB orbitals.

The *orbital space* efficacy (separating core orbitals from the 1st GVB-NO's and the 2nd GVB-NO's from the unoccupied orbitals) is given in Table 3 for all the systems studied. Here we consider both the GVB-INIT approach and also for a Boys localization method. The efficacy for each set of orbitals (core orbitals , 1st GVB-NO's, 2nd GVB-NO's) was defined as n_{guess}/n_{orbs} where n_{orbs} is the number of orbitals in that set and n_{guess} is defined in equation (4.1).

In order to have as close as possible a comparison between GVB-INIT and the Boys method, the Boys orbitals were obtained as follows:

1. a Boys localization was performed on the HF-INIT guess orbitals with the unoccupied orbitals localized independently from the occupied orbitals and the inner core orbitals (i.e., 1s orbital for C,N,O atoms and 1s,2s,2p for S atoms) kept frozen from HF-INIT,
2. the separation of the core orbitals from the 1st GVB-NO's and the 2nd GVB-NO's from the unoccupied orbitals was done by selecting the Boys orbitals having the maximum overlap with the SCF orbitals.

For Boys localization one needs some form of analysis of the location and composition of each Boys orbitals in order to properly assign which are core orbitals and which are lone pairs or 1st GVB-NO's. In contrast GVB-INIT does this assignment automatically. For these tests we did this analysis for Boys by hand. For all systems studied here the assignment for the Boys localized orbitals is simple since the electron distributions are consistent with simple valence bond structures. Assigning the 2nd GVB-NO's with a Boys localized orbital becomes very difficult since the 2nd GVB-NO's must be localized in the same region as the corresponding 1st GVB-NO.

Table 3 shows that the Boys method is slightly better than GVB-INIT for orbital space separation of the core orbitals from the 1st GVB-NO's. The Boys efficacies vary from 81% to 98% for the 1st GVB-NO's, whereas the efficacies vary from 77% to 98% for GVB-INIT. The advantage of the Boys method is largest when only some valence orbitals are correlated. If all possible bond and lone pair orbitals are correlated, then the core space is reduced to the inner core orbitals and the Boys and GVB-INIT methods become equivalent for the core orbitals and the 1st GVB-NO's. The reason for the slight disadvantage for GVB-INIT is that the

1st GVB-NO's are formed from projections without any mechanism to deliberately make the orbital localized. Of course GVB-INIT does automatic assignment of the orbitals whereas we had to do this by hand for Boys.

For the 2nd GVB-NO orbital space, the GVB-INIT method is markedly superior to the Boys method. Here the Boys efficacies range from 46% to 75%, whereas the GVB-INIT efficacies range from 82% to 96%. Clearly, this indicates that the 2nd GVB-NO's should be obtained directly from the 1st GVB-NO's as in the GVB-INIT method. In the Boys method, the localization principle does not necessarily lead to 2nd GVB-NO's that correspond well with the 1st GVB-NO's.

More relevant for convergence is the *orbital pair* criteria, that is how well each GVB-NO is predicted. In order to assess this, we evaluate the sum of the populations over the 1st GVB-NO's or the 2nd GVB-NO's.

$$n_{SumPop.} = \sum_{i}^{n_{pair}} \langle \phi_i^{guess} | \phi_i^{GVB-SCF} \rangle^2 \tag{4.4}$$

where the $\{\phi_i^{guess}\}$ are GVB-NO's of the initial guess wavefunction and the $\{\phi_i^{GVB-SCF}\}$ are GVB-NO's of the final HF-SCF wavefunction. The results are shown in Table 4.

Clearly, for both the 1st GVB-NO's and 2nd GVB-NO's, the GVB-INIT method gives much better orbital pair results than the Boys method for all cases except for the 1st GVB NO's of water. One difference with Boys localization is that double bonds and triple bonds look like "banana" bonds with two or three equivalent bonds, whereas the optimum GVB-NO's generally have the character of sigma and pi orbitals. A similar situation occurs for the lone pair orbitals on oxygen and sulphur atoms. GVB-INIT always gives lobe orbitals as in the final

GVB wavefunction, whereas Boys localization sometimes leads to σ and π lone pair orbitals.

The fact that Boys localization is better for the orbital space of 1st GVB-NO's for water indicates that choosing the most localized orbitals within the space of 1st GVB-NO's might lead to better trial GVB orbitals. To test whether this is true we performed a series of calculations where the Boys localization is followed by simple linear combination of the 1st GVB-NO's in order to reproduce the GVB-INIT orbitals as closely as possible. Those results are tabulated in Table 4 under the GVB-Boys heading. They consistently give the best results.

The most difficult GVB orbitals to generate are the 2nd GVB-NO's for lone pairs. For example in the case of glycine with 15 GVB pairs, the average error per orbital is about 8.1 % for bond pair 1st GVB-NO's, 13.0 % for lone pair 1st GVB-NO's, 11.7 % for bond pair 2nd GVB-NO's and 48.2 % for lone pair 2nd GVB-NO's. Unlike the bond pair 2nd GVB-NO's, the lone pair 2nd GVB-NO's are not constructed directly from the 1st GVB-NO's. In the bond pair case, the 1st GVB-NO and 2nd GVB-NO are obtained from the set of two localized atomic orbitals centered on the nearest neighbor atoms. The 1st GVB-NO corresponds to the bonding orbital and the 2nd GVB-NO to the anti-bonding orbital. For lone pair, we have only one atomic-centered orbital involved from which one can construct only the 1st GVB-NO. The 2nd GVB-NO must be obtained by some extra criteria. The guiding principle is that the 1st GVB-NO is typically s-like and the 2nd GVB-NO is p-like and pointing away from the nucleus in the direction of the 1st GVB-NO centroid. Two simple operators, the dipole moment and the gradient, can be used to transform an s-like orbital to a p-like orbital. If the s-like orbital is composed of a single s Gaussian function, then the two operators generate the same p-like orbital. If the s-like function contains many different Gaussian functions, then the p-like orbital generated by the gradient operator emphasizes tighter Gaussian functions than the dipole moment operator. The effect of the operators can also be limited to a small region of space around the 1st GVB-NO centroid by multiplying the operator by $1/|r - R_{lobe}|^n$. Different values of the $1/r$ exponent were tried (see Table 5) and the best results were obtained with $n = 3$ for the gradient operator.

So far we have compared directly the initial guess wavefunction to the final SCF wavefunction. Another important criteria is to compare the SCF convergence obtained with the initial guess wavefunctions. In Table 6 the GVB-INIT guess is compared to the Boys guess for three different convergence methods (for formaldehyde with four GVB pairs). The convergence criteria requires that the sums of squares of changes in the orbital coefficients be less than:10^{-9}

$$\sum_{\mu=1}^{n_{basis}} \sum_{i=1}^{n_{occ}} (\Delta C_{\mu i})^2 \leq 10^{-9}. \tag{4.5}$$

In all cases the GVB-INIT initial guess leads to much better convergence than Boys localization. The fastest convergence was obtained with the GVB-DIIS method.

V. Conclusions

The GVB-INIT method of obtaining initial guesses for GVB wavefunctions is very inexpensive, using only piecewise localization, and yet provides trial guesses which are very close to the final orbitals and have the correct localization properties for GVB wavefunctions.

Together with GVB-DIIS, these two methods make it now possible to use physically accurate wavefunctions to calculate force-fields, to describe bond distortion and dissociation processes, and to obtain highly converged wavefunctions for the purpose of calculating more exact molecular properties such as charges, dipole moments and solvation energies.

The GVB-INIT automatic scheme for GVB trial wavefunctions and the GVB-DIIS convergence scheme have already been implemented in the PS-GVB electronic structure program. We expect that the pseudospectral approach, when combined with the methods detailed here, should allow GVB calculations on much larger systems than have been possible before.

References

1. F. W. Bobrowicz and W. A. Goddard III, in Modern Theoretical Chemistry: Methods of Electronic Structure Theory, edited by H. F. Schaefer III (Plenum, New Yok, 1977), Vol. 3, p. 79.
2. L. G. Yaffe and W. A. Goddard III, J. Chem. Phys. **67**, 1777 (1977).
3. W. J. Hunt, W. A. Goddard III, and T. H. Dunning, Jr., Chem. Phys. Lett. **6**, 147 (1970). See also Chem. Phys. Lett. **4** 231 (1969) and Chem. Phys. Lett **3**,606 (1969).
4. R. P. Muller, J-M. Langlois, M. N. Ringnalda, R. A. Friesner, and W. A. Goddard III, "A Generalized Direct Inversion of the Iterative Subspace Approach for Converging Generalized Valence Bond Wavefunction (GVB-DIIS)," to be published.
5. R. C. Ladner and W. A. Goddard III, J. Chem. Phys. **51**, 1073 (1969). W. A. Goddard III, T. H. Dunning, Jr., W. J. Hunt, and P. J. Hay, Accts. Chem. Res. **6**, 368 (1973). W. A. Goddard III and L. B. Harding, Ann. Rev. Phys. Chem. **29**, 363 (1978). W. A. Goddard III, Science **227**, 917 (1985).
6. W. J. Hunt, P. J. Hay, and W. A. Goddard III, J. Chem. Phys. **57**, 738 (1972). P. J. Hay, W. J. Hunt, and W. A. Goddard III, J. Am. Chem. Soc. **94**, 8293 (1972).
7. BIOGRAF/POLYGRAF from Molecular Simulation Inc., Burlington, Mass.
8. M. J. Frisch, *et al.*, *GAUSSIAN 90*, Gaussian, Inc., Pittsburgh, Pennsylvania, 1990.
9. R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, J. Chem. Phys. **72**, 650 (1980).

10. GAMESS program suite, M. W. Schmidt, K. K. Baldridge, J. A. Boatz, J. H. Jensen, S. Doseki, M. S. Gordon, K. A. Nguyen, T. L. Windus, S. T. Elbert.

11. R. Polack, Int. J. Quant. Chem. **4**, 271 (1970).

12. A. T. Amos and G. G. Hall, Proc. R. Soc. **A263**, 483 (1961).

13. J. M. Foster and S. F. Boys, Rev. Mod. Phys. **32**, 300 (1960).

14 C. Edmiston and K. Ruedenberg, Rev. Mod. Phys. **35**, 457 (1963).

15 T. F. Koetzle, M. N. Frey, M.S. Lehmann and W. C. Hamilton, Acta Cryst. **B29**, 2571 (1973).

16 B. M. L. Chen and A. Tulinsky, J. Am. Chem. Soc. **94**, 4144 (1972).

17. E. A. Carter and W. A. Goddard III, J. Phys. Chem. **92**, 2109 (1988). S. K. Shin and W. A. Goddard III, J. Chem. Phys. **93**, 4986 (1990).

18. H. Fujimoto, N. Koga, K. Fukui, J. Am. Chem. Soc., **103**, 7452 (1981).

Table 1. Population difference $n_{occ} - n_{guess}$ [see (4.1)] between initial and converged HF wavefunctions.

Molecule	No. of orbitals	HF-INIT	GAUSSIAN90	Geometry
water	5	0.149	0.066	STO-3G optimized
formaldehyde	8	0.167	0.084	STO-3G optimized
ethylene	8	0.141	0.077	Reference 9
Si_2H_6 [a]	17	0.297	0.249	STO-3G
glycine	20	0.455	0.224	STO-3G optimized
glutamine	39	0.808	0.454	Reference 15
TTF[b]	52	0.819	0.707	6-31G** optimized
porphine	81	1.213	0.832	Reference 16

[a] Staggered geometry.

[b] Tetra-thio-fulvelene.

Table 2. Description of GVB-PP wavefunctions used for tests.

Molecule	No. of GVB pairs	Pair distribution σ	 π	 lone	Pair correlated
water	2	2	0	0	all bond pairs
water	4	2	0	2	all bond and lone pairs
formaldehyde	2	1	1	0	CO double bond pairs
formaldehyde	4	3	1	0	all bond pairs
formaldehyde	6	3	1	2	all bond and lone pairs
ethylene	2	1	1	0	CC double bond pairs
ethylene	6	5	1	0	all bond pairs
glycine	5	4	1	0	all NC,CC,CC,CO bond pairs
glycine	10	9	1	0	all bond pairs
glycine	15	9	1	5	all bond and lone pairs
TTF	18	15	3	0	all bond pairs
TTF	26	15	3	8	all bond and lone pairs
porphine	22	11	11	0	all double bond pairs

Table 3. Orbital space efficacies.

Molecule	GVB	Core Orb.		1st GVB-NO		2nd GVB-NO	
	pairs	Boys	GVB-INIT	Boys	GVB-INIT	BOYS	GVB-INIT
water	2	0.9201	0.8632	0.8813	0.7862	0.4607	0.8432
water	4	0.9998	0.9998	0.9629	0.9629	0.4904	0.8935
formaldehyde	2	0.9570	0.9272	0.8992	0.8085	0.7039	0.8978
formaldehyde	4	0.9551	0.9008	0.9379	0.8831	0.6562	0.9084
formaldehyde	6	0.9998	0.9998	0.9737	0.9737	0.5983	0.9180
ethylene	2	0.9732	0.9522	0.9630	0.8980	0.7508	0.9654
ethylene	6	0.9999	0.9999	0.9766	0.9766	0.6004	0.9375
glycine	5	0.9244	0.9118	0.8154	0.7715	0.6622	0.8906
glycine	10	0.8866	0.8827	0.8750	0.8696	0.6014	0.9005
glycine	15	0.9998	0.9998	0.9693	0.9693	0.5549	0.9124
TTF	18	0.9594	0.9485	0.9162	0.8954	0.7077	0.9249
TTF	26	0.9996	0.9996	0.9670	0.9670	0.6399	0.8208
porphine	22	0.9647	0.9435	0.9242	0.8667	0.6946	0.9445

Table 4. Orbital pair efficacies. Sum of populations between initial and converged GVB Natural Orbitals.

Molecule	GVB pairs	Boys	1ts GVB-NO GVB-INIT	GVB-Boys	2nd GVB-NO Boys	GVB-INIT
water	2	1.756	1.547	1.756	0.919	1.681
water	4	3.724	3.435	3.724	1.419	3.349
formaldehyde	2	0.899	1.617	1.798	0.704	1.795
formaldehyde	4	2.836	3.530	3.744	1.918	3.633
formaldehyde	6	4.807	5.523	5.761	2.501	5.269
ethylene	2	0.963	1.796	1.926	0.782	1.930
ethylene	6	4.867	5.858	5.851	2.883	5.624
glycine	5	3.168	3.853	4.470	3.252	4.447
glycine	10	7.816	8.647	9.112	5.944	8.975
glycine	15	11.662	13.542	14.053	7.139	13.105
TTF	18	13.501	16.061	16.434	10.447	16.623
TTF	26	17.749	23.891	24.095	13.897	20.921
porphine	22	10.839	19.021	19.626	9.040	20.737

Table 5. Lone pair 2nd natural orbital efficacies. Sum of populations between initial and converged 2nd GVB Natural Orbitals.

Molecule	No. of GVB pairs	Operator	Power of $1/r^n$			
			$n = 0$	$n = 1$	$n = 2$	$n = 3$
water	4	Dipole	1.857	1.762	1.627	1.760
water	4	Gradient	1.809	2.007	2.826	3.334
formaldehyde	6	Dipole	3.856	3.840	3.730	3.746
formaldehyde	6	Gradient	3.766	3.882	3.949	4.147
glycine	15	Dipole	9.383	9.292	9.010	9.401
glycine	15	Gradient	9.221	9.630	10.722	11.477
TTF	26	Dipole	17.132	16.609	17.492	16.549
TTF	26	Gradient	16.972	18.244	20.452	20.724

Table 6. Number of iterations for converging the GVB wavefunctions of formaldehyde.

	No. of	Convergence Method[a]		
Trial Function	GVB pairs	GVB-DIIS	GVB2P5	GAUSSIAN90
GVB-INIT	2	15	28	34
Boys	2	21[b]	83	95
GVB-INIT	4	15	58	69
Boys	4	21[b]	>300[c]	>300[c]

[a] Convergence criteria of 10^-9.

[b] Converged to nearest root.

[c] Unconverged after 300 iterations.

Chapter 2

Pseudospectral Generalized Valence-Bond Calculations: Application to Methylene, Ethylene and Silylene

Pseudospectral generalized valence-bond calculations: Application to methylene, ethylene, and silylene

Jean-Marc Langlois, Richard P. Muller, Terry R. Coley, and William A. Goddard III
Contribution No. 8101 from the Arthur Amos Noyes Laboratory of Chemical Physics, California Institute of Technology, Pasadena, California 91125

Murco N. Ringnalda, Youngdo Won, and Richard A. Friesner
Department of Chemistry, The University of Texas at Austin, Austin, Texas 78712

(Received 2 February 1990; accepted 28 February 1990)

The pseudospectral (PS) method for self-consistent-field calculations is extended for use in generalized valence-bond calculations and is used to calculate singlet–triplet excitation energies in methylene, silylene, and ethylene molecules and bond dissociation and twisting energies in ethylene. We find that the PS calculations lead to an accuracy in total energies of <0.1 kcal/mol and excitation energies to <0.01 kcal/mol for all systems. With effective core potentials on Si, we find greatly improved accuracy for PS.

I. INTRODUCTION

Recently, we have shown that the pseudospectral (PS) numerical method provides an attractive alternative to the usual all-integral methods of calculating Hartree–Fock (HF) wave functions for molecules.[1–6] Thus, for a valence double-zeta plus polarization description of glycine (100 basis functions), PS calculations[5] leading to an accuracy in the total energy of 0.04 kcal/mol now take about 75 CPU seconds on a Cray X-MP (recalculating integrals at every self-consistent iteration) vs 200 CPU seconds using GRADSCF, a fast program for the traditional all-integral approach (which calculates integrals only once). For HF wave functions the PS approach scales as M^3 (where M is the number of basis functions), whereas the all-integral approaches scale as M^4, so that PS should become even more advantageous for very large systems. Using integral cutoffs, we have recently achieved M^2 scaling for PS for systems as small as benzene and uracil.[6] In this paper we extend the PS approach to generalized valence-bond (GVB) wave functions (which include electron correlation) and to second-row compounds (Si) treated either with all electrons or with core effective potentials.

Earlier attempts to use purely numerical methods for electronic wave functions of polyatomic molecules have failed to provide uniform accuracy, primarily because the accurate description of the cusps at atomic centers requires an impractical number of grid points. The PS method uses both a basis set and a numerical grid and preserves the advantages of both approaches. Using a basis set for the calculation of one-electron and selected two-electron integrals accurately reproduces the description of the wave function near atomic centers. Use of a grid for the evaluation of the majority of the two-electron terms both (a) preserves the M^3 scaling of the numerical calculation with basis-set size and (b) increases the speed of calculating the integrals that must be evaluated.

In previous papers the general algorithm has been proposed and applied to atomic,[1] diatomic,[2] and polyatomic molecules.[3] A suitable automatic grid generation scheme has been developed,[4] and the method has been applied to glycine.[5] It has been demonstrated that the method promises significant reduction of computation time (even when compared with the GRADSCF program that was specifically designed for Cray computers). In all cases the PS method has shown accurate reproduction of total and relative energies to within 0.1 kcal/mol, as compared with all-integral approaches.

There remain, however, many issues to be addressed. Only a few basis sets have been used, and it has not yet been determined whether the agreement with all-integral method energies will degrade with different basis sets. Since calculations have only been performed on first-row elements, the ability to replace the core electrons with an effective core potential (ECP)—a virtually essential condition to effectively model systems with heavy atoms—has not yet been established. The calculations have not yet attempted to solve for excited electronic states, or for the broken bonds of stretched or twisted molecular conformations. Finally, no amount of electron correlation has entered into any of the calculations. In each of these areas it is necessary to test the efficacy of the PS method.

In this paper we examine the above issues for methylene, ethylene, and silylene molecules. In the methylene and silylene molecules we investigate (a) the effect of basis sets where *s*, *p*, and *d* orbitals have different exponents, (b) the use of diffuse functions, and (c) multiple sets of polarization functions. In silylene we also examine the use of an effective potential to replace the core electrons in the wave function. For the ethylene system we determine how well the PS method reproduces all-integral results for broken bonds and twisted geometries. In all of these systems we have examined both restricted HF wave functions and generalized valence-bond (GVB) multiple configuration wave functions for both the singlet and triplet electronic states. We find excellent agreement with all-integral results for these more complex systems, suggesting that the PS method can be applied to a variety of chemically relevant systems.

In Secs. II and III we review the relevant aspects of GVB wave functions and the PS method and indicate how the PS approach is used with GVB. In Sec. IV we report the results on CH_2, SiH_2, and C_2H_4 while Sec. V discusses some key points.

II. OVERVIEW OF THE GVB FORMALISM

A. The GVB-PP wave function

The GVB wave function for an N electron system has the general form[7,8]

$$\Psi^{\mathrm{GVB}}_{S,M_s}=\mathscr{A}[\phi_a^{(1)}\phi_b^{(2)}\cdots\phi_w^{(N)}\Theta_{S,M_s}(1\cdots N)], \tag{2.1}$$

where the orbitals are singly occupied and allowed to overlap, $\Theta^{(1\cdots N)}_{S,M_s}$ is a general spin eigenfunction

$$S^2\Theta_{S,M_s}=S(S+1)\Theta_{S,M_s}, \quad S_z\Theta_{S,M_s}=M_S\Theta_{S,M_s}, \tag{2.2}$$

for an N electron system, and $\mathscr{A}$ is the antisymmetrizer. There are f_{NS} ways of combining the spin to satisfy Eqs. (2.2) so that Θ_{S,M_s} is written as

$$\Theta=\sum_{i=1}^{f}\gamma_i\Theta_i, \tag{2.3}$$

where each Θ_i satisfies Eqs. (2.2). (Thus $f=2$ for $N=4$ and $S=0$, and $f=5$ for $N=6$ and $S=0$.) The GVB orbitals are obtained by simultaneously optimizing the orbitals ϕ_i and spin function Θ [i.e., the γ_i for Eq. (2.3)].

A restricted form of GVB that is much more tractable for computation is the GVB-PP (for perfect-pairing) wave function, in which the spin function in Eq. (2.3) is restricted to the simple valence-bond form

$$\Theta_{\mathrm{VB}}=\Theta_1=(\alpha\beta-\beta\alpha)(\alpha\beta-\beta\alpha)\cdots(\alpha\beta-\beta\alpha)\alpha\alpha\cdots\alpha \tag{2.4}$$

with $(1/2)N-S$ singlet pairs and $2S$ electrons coupled high spin.

In this case the orbital functions in Eq. (2.1) corresponding to the jth singlet pair in Eq. (2.4) have the valence-bond form

$$(\phi_{1j}\phi_{2j}+\phi_{2j}\phi_{1j}). \tag{2.5}$$

For computational purposes it is convenient to rewrite Eq. (2.5) as

$$C_{gj}\phi_{gj}\phi_{gj}-C_{uj}\phi_{uj}\phi_{uj} \tag{2.6}$$

using the GVB natural orbitals

$$\phi_{gj}=(\phi_{1j}+\phi_{2j})/\sqrt{2(1+S_j)},$$
$$\phi_{uj}=(\phi_{1j}-\phi_{2j})/\sqrt{2(1-S_j)}, \tag{2.7}$$

where $S_j=\langle\phi_{1j}|\phi_{2j}\rangle$ and $C^2_{gj}+C^2_{uj}=1$. Taking natural orbitals of different pairs to be orthogonal leads to the general energy expression for GVB-PP wave functions:

$$E=2\sum_i^{\mathrm{occ}}f_i h_{ii}+\sum_{i,j}^{\mathrm{occ}}(a_{ij}J_{ij}+b_{ij}K_{ij}), \tag{2.8}$$

where

$$h_{ii}=\langle\phi_i|h|\phi_i\rangle, \tag{2.9}$$

$$J_{ij}=(ii|jj)=\int d^3r_1\,\phi_i(1)\hat{\mathscr{J}}_j(1)\phi_i(1), \tag{2.10}$$

$$\hat{\mathscr{J}}_j(1)=\int d^3r_2\frac{\phi_j(2)\phi_j(2)}{r_{12}}, \tag{2.11}$$

$$K_{ij}=(ij|ij)=\int d^3r_1\,\phi_i(1)\hat{K}_j(1)\phi_i(1), \tag{2.12}$$

$$\hat{K}_j(1)=\int d^3r_2\frac{\phi_j(2)\tau_{ij}\phi_j(2)}{r_{12}}, \tag{2.13}$$

τ_{ij} = transposition operator

are standard expressions for the one-electron, Coulomb, and exchange energies. (Here h may include core effective potentials in addition to kinetic energy and nuclear potential terms.)

The coefficients f_i, a_{ij}, and b_{ij} in Eq. (2.8) depend on the wave function, and the general prescription is given in Ref. 9. Some special cases are that doubly occupied orbitals lead to $f_i=1$, while a GVB natural orbital leads to $f_i=C_i^2$. For two orbitals of the same GVB pair, $a_{ii}=f_i$, $b_{ii}=0$, $a_{1j,2j}=0$, $b_{1j,2j}=-C_{1j}C_{2j}$. For orbitals i,j $(i\neq j)$ that are doubly occupied or of different GVB pairs, $a_{ij}=2f_if_j$ and $b_{ij}=-f_if_j$.

With Eq. (2.8), the general condition for the optimum orbitals is that the first-order change in the energy due to changes in the various orbitals is zero, leading to

$$\sum_i\langle\delta\phi_i|F_i|\phi_i\rangle=0, \tag{2.14}$$

where F_i is the Fock operator

$$F_i=f_i\,\hat{h}+\sum_j^{\mathrm{occ}}(a_{ij}\hat{\mathscr{J}}_j+b_{ij}\hat{K}_j). \tag{2.15}$$

Because F_i depends on the orbitals, solution of the variational equation (2.14) is nonlinear, requiring an iterative procedure to solve. The general conditions for the orbitals to be optimal is that the matrix

$$A_{ji}=\langle j|(F_j-F_i)|i\rangle \tag{2.16}$$

be zero,

$$A_{ji}=0, \tag{2.17}$$

for all pairs of orbitals: If j is unoccupied, then $F_j=0$ is used in Eq. (2.16). The iterative GVB equations are based on expanding (2.8) through second order in the orbital changes and lead to the general equation

$$\sum_{\nu j}B_{\mu i,\nu j}\Delta_{\nu j}=-A_{\mu i}, \tag{2.18}$$

where $\nu>j$ for index νj and where j is occupied and ν is occupied or virtual. The antisymmetric matrix Δ is the generator of the orthogonal transformation matrix

$$T=e^{\Delta} \tag{2.19}$$

that transforms an initial set of orthogonal orbitals $\{\phi_i^0\}$ into a final set of orthogonal functions $\{\phi_i^1\}$ which minimizes the energy (2.8) expanded to second order in the changes. The general formulas for $A_{\mu i}$ and $B_{\mu i,\nu j}$ can be found in Refs. 9–11.

The diagonal elements of B have the form

$$B_{\mu i,\mu i}=\langle\mu|(F_i-F_\mu)|\mu\rangle+\langle i|(F_\mu-F_i)|i\rangle + \alpha_{\mu i}J_{\mu i}+\beta_{\mu i}K_{\mu i}, \qquad (2.20)$$

where the $\alpha_{\mu i}$ and $\beta_{\mu i}$ depend on the a_{ij} and b_{ij}. Since only Coulomb and exchange operators are involved, the quantities required for the diagonal terms can be evaluated from a list of integrals (an M^4 procedure) without carrying out integral transformations (an M^5 procedure). On the other hand, the off-diagonal terms generally involve more complicated integrals requiring integral transformations. In the GVB2P5 program we ignore any terms that cannot be evaluated with Coulomb and exchange operators, and in the present calculations we used only the diagonal elements of B. As a result, Eq. (2.18) reduces to

$$\Delta_{\mu i}=-A_{\mu i}/B_{\mu i,\mu i}. \qquad (2.21)$$

In addition to Δ, the configuration-interaction (CI) coefficients in Eq. (2.6) must be optimized each iteration, leading to the two-by-two diagonalizations[9]

$$\sum_{l=g}^{u} Y_{kj,lj}C_{lj}=E_{kj}C_{kj}, \qquad (2.22)$$

where $l=g$ or u and

$$Y_{kj,kj}=h_{kj,kj}+\tfrac{1}{2}J_{kj,kj}+\sum_{i\neq gj,uj}^{\text{occ}} f_i(2J_{i,kj}-K_{i,kj}), \qquad (2.23)$$

$$Y_{gj,uj}=\tfrac{1}{2}K_{gj,uj}.$$

An important feature of the GVB-PP wave function is that the calculations require *only* Coulomb and exchange integrals [see Eqs. (2.8), (2.15), (2.20), and (2.23)].

B. Basis-set expansions

The standard approach to solving for the GVB-PP orbitals is to expand each orbital ϕ_i as

$$\phi_i=\sum_{\mu=1}^{M}\chi_\mu C_{\mu i},\quad \delta\phi_i=\sum_{\mu=1}^{M}\chi_\mu(\delta C_{\mu i}) \qquad (2.24)$$

in a finite basis set $\{\chi_\mu;\mu=1\cdots M\}$ and solve for the coefficients $\{C_{\mu\nu}\}$ leading to orbitals satisfying Eqs. (2.17) and (2.22). With a basis set, all quantities needed for optimizing the orbitals can be calculated in terms of the one-electron integrals,

$$h_{\mu\nu}=\langle\mu|h|\nu\rangle, \qquad (2.25)$$

the matrix elements of the Coulomb operators,

$$J^j_{\mu\nu}=\langle\mu|\hat{J}_j|\nu\rangle=(\mu\nu|jj), \qquad (2.26)$$

and the exchange operators,

$$K^j_{\mu\nu}=\langle\mu|\hat{K}_j|\nu\rangle=(\mu j|\nu j). \qquad (2.27)$$

These operators can be calculated from the fundamental list of two-electron integrals by a simple two-index transformation:

$$J^j_{\mu\nu}=\sum_{\sigma,\eta}(\mu\nu|\sigma\eta)C_{\sigma j}C_{\eta j},\quad K^j_{\mu\nu}=\sum_{\sigma,\eta}(\mu\sigma|\nu\eta)C_{\sigma j}C_{\eta j}, \qquad (2.28)$$

so that it is never necessary to transform the two-electron set of integrals. This contrasts with more general multiconfiguration–self-consistent-field (MC-SCF) procedures where a full transformation of the two-electron integrals (a M^5 procedure) is necessary every iteration.

III. PS GVB

A. Overview

In the usual all-integral GVB calculations, a set of $M^4/8$ integrals $(\mu\nu|\sigma\eta)$ is evaluated and used to calculate the matrix elements for various Coulomb and exchange operators as in Eq. (2.28). In contrast, for PS GVB calculations, the Coulomb and exchange operators are evaluated directly using a numerical grid. This involves evaluating the $M^2/2$ quantities

$$A_{\sigma\eta}(r_g)=\int d^3r_2\frac{\chi_\sigma(2)\chi_\eta(2)}{|r_g-r_2|} \qquad (3.1)$$

directly for a set of grid points $\{r_g\}$. With Gaussian-type basis functions, $\chi_\sigma=r^le^{-\alpha r^2}Y_{lm}(\Theta,\phi)$, on various centers, the potentials $A_{\sigma\eta}(r_g)$ can be evaluated analytically for each grid point.

Given the potentials (3.1), the fundamental Coulomb and exchange matrices (2.26) and (2.27) could be calculated numerically using

$$J^j_{\mu\nu}=\langle\mu|\hat{J}_j|\nu\rangle=\sum_g\chi_\mu(g)\chi_\nu(g)\sum_{\sigma\eta}C_{\sigma j}C_{\eta j}A_{\sigma\eta}(g), \qquad (3.2)$$

$$K^j_{\mu\nu}=\langle\mu|\hat{K}_j|\nu\rangle=\sum_g\chi_\mu(g)\sum_\eta\chi_\eta(g)\sum_\sigma C_{\sigma j}C_{\eta j}A_{\sigma\nu}(g).$$

However, Eq. (3.2) leads to large numerical errors that can be overcome only with huge grids. The reason is that the functions

$$\hat{J}_j\chi_\nu \quad\text{and}\quad \hat{K}_j\chi_\nu \qquad (3.3)$$

lead to components *outside* the basis space $\{\chi_\mu\}$ even though ϕ_i is described within this space. Use of Eq. (3.2) is equivalent to expanding $\hat{J}^j\phi_i$ and $\hat{K}^j\phi_i$ defined over the grid in terms of the basis $\{\chi_\mu\}$. If the grid were infinite, this would be acceptable, but with finite grids it leads to large errors in (3.2), referred to as the *alias*.[2] This manifests itself as noise in the wave function.[3] The solution to this difficulty is to expand Eq. (3.3) in terms of a larger *dealiasing basis* $\{\chi^*_\sigma;\sigma=1\cdots M^*\}$:

$$J^j(g)\chi_\nu(g) = \sum_\sigma^{M^*} \chi_\sigma^*(g) W_{\sigma,\nu}^{jJ}$$

$$K^j(g)\chi_\nu(g) = \sum_\sigma^{M^*} \chi_\sigma^*(g) W_{\sigma,\nu}^{jK}, \tag{3.4}$$

used only for evaluating Eq. (3.2)

$$J_{\mu\nu}^j = \langle\mu|J^j|\nu\rangle = \sum_\sigma^{M^*} S_{\mu\sigma}^* W_{\sigma,\nu}^{jJ},$$

$$K_{\mu\nu}^j = \langle\mu|K^j|\nu\rangle = \sum_\sigma^{M^*} S_{\mu\sigma}^* W_{\sigma,\nu}^{jK}, \tag{3.5}$$

where the $S_{\mu\sigma}^* = \langle\chi_\mu|\chi_\sigma^*\rangle$ is the overlap matrix element between the normal basis and the augmented dealiasing set (normal basis functions plus dealiasing functions). This set of dealiasing functions then filters out the noise due to incompleteness of $\langle\chi_\mu\rangle$. The coefficients $W_{\sigma,\nu}^{jJ}$ and $W_{\sigma,\nu}^{jK}$ in Eq. (3.4) are obtained by a least-squares fit over the grid points.[2]

The normal basis $\langle\chi_\mu\rangle$ is augmented to form the dealiasing basis $\langle\chi_\mu^*\rangle$ by adding new radial wave functions so that the augmented basis has functions whose exponents differ by a factor of at least 1.5 and at most 2.5. When two exponents are any closer than this, linear dependencies start to appear, making the final energy unreliable. When two consecutive exponents are spaced by more than this, then the least-squares fit is less accurate, again making the final energy less reliable. In addition, the angular functions of the basis are augmented so that the d basis contains exponents corresponding to the p functions of the normal basis, and so there are f functions corresponding to the d functions of the normal basis.

With the Coulomb and exchange operators $J_{\mu\nu}^j$ and $K_{\mu\nu}^j$ in Eq. (3.5), we proceed to calculate the wave functions as in normal GVB-PP calculations.

B. The grid

The general grid generation approach is outlined in Ref. 4. For each atom in the molecule, there are angular and radial grids using Legendre polynomials and hyperbolic tangent functions, respectively. These atomic grids are then merged to form a molecular grid using Lagrange interpolation to ensure that the grid produces functions that dissociate smoothly as atoms are separated to infinity.

We use a multigrid strategy as outlined in Ref. 5. This strategy involves generating three different grids denoted as medium, fine, and ultrafine with about 200, 400, and 1100 points per carbon atom. The medium grid is used for the vast majority of the SCF iterations. Once the wave function begins to converge, the ultrafine grid is used for one iteration, followed by the fine grid for several iterations (usually three). After this reconsolidation, all subsequent iterations can be carried out using the medium grid. Fock matrix updating is used subsequent to the ultrafine grid iteration, so that the remaining iterations require increasingly less accuracy in the integrals.

The great advantage in using such a multigrid strategy is that one can obtain wave functions whose energies are more accurate than those obtained from using the medium grid alone, while using the medium grid for all but about four of the iterations. The integrals for the fine and ultrafine grids need never be stored, and consequently the storage requirement for the PS algorithm is proportional to the medium grid size. More recently, we have switched entirely to a direct SCF scheme in which no disk space is required for integral storage. Regeneration of the medium grid integrals is very inexpensive, so that with an optimized integral package and efficient cutoffs there is only a ~15% increase in CPU time for recalculation.

C. Monoatomic and diatomic corrections

A major challenge in using numerical methods for the wave functions of molecules is the accurate description of the core orbitals. These orbitals vary rapidly near the nucleus, requiring a large number of grid points and dealiasing functions for reasonable accuracy. An efficient way to deal with this problem is to calculate selected two-electron integrals analytically for these core orbitals. This method has been applied on two levels, which we denote as monoatomic (1C) and diatomic (2C) corrections.

1. 1C corrections

In the 1C approach the two-electron integrals having all four of the atomic orbitals centered on one atom are evaluated analytically. Such an approach is computationally inexpensive and is easy to implement. The number of two-electron integrals required is $M\eta^3/8$, where η is the number of basis functions per atom and M the total number of basis functions. For large molecules this scales as M, with a prefactor of η^3. Since η is always small (10–20) and a large fraction of the 1C two-electron terms are zero by symmetry, the 1C calculations are always inexpensive.

The procedure is as follows. The Coulomb and exchange operators containing all one-center contributions are calculated as in Eqs. (2.26) and (2.27), but with the sum restricted to one-center terms only, to obtain

$$J_{\mu,\nu}^{j,1c} \quad \text{and} \quad K_{\mu\nu}^{j,1c}.$$

The time required to evaluate analytically the necessary 1C two-electron integrals and to assemble these operators is negligible.

Next we must correct for these 1C terms in evaluating $J_j\chi_2$ and $K_j\chi_2$ over the grid. This is done by calculating the 1C operators on the grid using

$$J_a^{j,1c}(g) = \sum_{\sigma,\eta\in a} A_{\sigma\eta}(g) C_{\sigma j} C_{\eta j}, \tag{3.6a}$$

$$K_a^{j,1c}(g)\chi_{\nu_a}(g) = \sum_{\sigma,\eta\in a} A_{\sigma\nu_a}(g) C_{\sigma j} C_{\eta j} \chi_\eta(g), \tag{3.6b}$$

where the sums in Eq. (3.6) are over all basis functions of atoms a. In Eq. (3.6b) the ν_a is only for basis functions on a.

These functions Eq. (3.6) are then described with the antialiasing functions:

$$J_a^{j,1c}(g)\chi_{\nu_a}(g)=\sum_{\sigma}^{M^*} W_{\sigma,\nu_a}^{j,J,1c}\chi_\sigma^*(g),$$

$$K_a^{j,1c}(g)\chi_{\nu_a}(g)=\sum_{\sigma}^{M^*} W_{\sigma,\nu_a}^{j,K,1c}\chi_\sigma^*(g). \qquad (3.7)$$

Using these quantities the total Coulomb and exchange operators are calculated as

$$J_{\mu\nu}=\sum_{\sigma}^{M^*} S_{\mu\sigma}^*(W_{\sigma,\nu}^{j,J}-\delta_{a_\mu a_\nu}W_{\sigma,\nu}^{j,J,1c})+J_{\mu\nu}^{1c},$$

$$K_{\mu\nu}=\sum_{\sigma}^{M^*} S_{\mu\sigma}^*(W_{\sigma,\nu}^{j,K}-\delta_{a_\mu a_\nu}W_{\sigma,\nu}^{j,K,1c})+K_{\mu\nu}^{1c}, \qquad (3.8)$$

where $\delta_{a_\mu a_\nu}$ ensures that basis functions μ and ν be on the same center a ($J_{\mu\nu}^{1c}$ and $K_{\mu\nu}^{1c}$ are zero for μ and ν on different centers).

2. 2C corrections

In the diatomic corrections approach, the two-electron integrals having all four of the atomic orbitals centered on one or two atoms are evaluated analytically and used to calculate

$$J_{\mu\nu}^{2c} \quad \text{and} \quad K_{\mu\nu}^{2c} \qquad (3.9)$$

The number of such terms is $M^2\eta^2$, and for large molecules this scales as M^2 with a prefactor of η^2. The implementation of the diatomic correction is much more complicated than for the monatomic correction. New correction operators need to be defined for every pair of atoms,

$$J_{ab}^{j,2c}(g)=\sum_{\sigma,\eta\in a,b} A_{\sigma\eta}(g)C_{\sigma j}\,C_{\eta j},$$

$$K_{ab}^{j,2c}(g)\chi_{\nu_a}(g)=\sum_{\sigma,\eta\in a,b} A_{\sigma\nu}(g)C_{\sigma j}\,C_{\eta j}\chi_\eta(g) \qquad (3.10)$$

(where ν_{ab} indicates ν on either a or b). These are in turn described with antialiasing functions as in Eq. (3.7),

$$J_{ab}^{j,2c}(g)\chi_{\nu_a}(g)=\sum_{\sigma}^{M^*} W_{\sigma,\nu_{ab}}^{j,J,2c}\chi_\sigma^*(g),$$

$$K_{ab}^{j,2c}(g)\chi_{\nu_a}(g)=\sum_{\sigma}^{M^*} W_{\sigma,\nu_{ab}}^{j,K,2c}\chi_\sigma^*(g). \qquad (3.11)$$

The Coulomb and exchange operators are then defined as

$$J_{\mu\nu}^{j}=\sum_{\sigma} S_{\mu\sigma}^*\left[W_{\sigma,\nu}^{j,J}-\left(\sum_{b}\right)W_{\sigma,\nu,b}^{j,J,2c}\right]+J_{\mu\nu}^{j,2c},$$

$$K_{\mu\nu}^{j}=\sum_{\sigma} S_{\mu\sigma}^*\left[W_{\sigma,\nu}^{j,K}-\left(\sum_{b}\right)W_{\sigma,\nu,b}^{j,K,2c}\right]+K_{\mu\nu}^{j,2c}, \qquad (3.12)$$

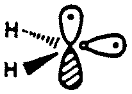
where the sum over all the atoms, Σ_b, needs to be done if $a_\nu = a_\mu$. The number of correction operators that need to be evaluated scales as the square of the number of atoms in the molecule. The time required to include these corrections is not as trivial as in the monoatomic corrections, but the diatomic corrections give results that are significantly more accurate than the monoatomic corrections.

IV. RESULTS

We picked three systems designed to test how effective the PS method would be for GVB calculations. In each case the quantity of interest is an energy difference, either between two excited states or for breaking a bond. For CH_2 and SiH_2 the quantities of interest are the excitation energy to the excited triplet state. In addition, a core effective potential was tested for Si where an issue is the accuracy of PS approaches for such cases. For C_2H_4 the quantities of interest include the π-to-π^* excitation energy, the rotational barrier for twisting the molecule by 90° about the CC bond, and the CC bond energy (energy to double the CC bond length).

A. Methylene

For CH_2 the ground state has $^3B_1$ symmetry with σ and π nonbonding orbitals as in Eq. (4.1a), whereas the singlet has the two nonbonding electrons spin paired as in Eq. (4.1b):



(4.1a)

(4.1b)

Because of the difference in the nonbonding orbitals, the geometries are quite different. Thus, with GVB configuration-interaction (GVB-POL-CI), Harding and Goddard[12] calculated $R_{CH} = 1.084$ Å, $\Theta_{HCH} = 133.2°$ for the triplet state, and $R_{CH} = 1.113$ Å, $\Theta_{HCH} = 101.8°$ for the singlet. The important quantity here, the triplet–singlet excitation energy, is $\Delta_{ST} = 9.09$ kcal/mol $= 0.0145$ hartrees $= 14.5$ mhartrees.

In studying methylene, our goals are (a) to test the accuracy of the PS method for good quality basis sets [valence double-zeta plus polarization (VDZ*p*)] and for basis sets where additional polarization or diffuse functions are added, and (b) to test the change in accuracy when GVB wave functions are used rather than HF wave functions.

All-integral HF calculations lead to excitation energies ranging from 26.1 to 25.1 kcal/mol for various basis sets, while GVB-PP leads to 9.1 to 7.8 kcal/mol. As described below, the PS-1C leads to a maximum error of 0.09 kcal/mol for both HF and GVB calculations, while PS-2C leads to a maximum error of 0.01 kcal/mol. We conclude that the PS method is quite adequate for these systems.

TABLE I. Dealiasing set for methylene.

Atom	Basis	L block	Exponent[a]				
C	VDZ_p	SP	10.24	20.48			
		SPD	0.32	0.64	1.28	2.56	5.12
H	VDZ_p	SP	4.0	8.0			
		SPD	0.1775	0.32	0.64	1.0	2.0
C	$VDZ2_p$	SP	8.0	16.0			
	$VDZ2_{pn}$	SPD	0.06	0.422	0.971	2.0	4.0
H	$TZ2_p$	SP	4.0	8.0			
		SPD	0.0822	0.2246	0.6	1.38	2.0

[a] Exponents that already exist in the basis set are not included in the dealizing set.

TABLE II. Total energy for CH_2 using GVB-PP wave functions.

Method	State[a]	Basis sets (carbon/hydrogen)		
		VDZ_p/VDZ_p	$VDZ2_p/TZ2_p$	$VDZ2_{pn}/TZ2_p$
		Total energy (hartrees)		
All-ints[b]	$^1A_1$	$-38.938\,766$	$-38.942\,557$	$-38.943\,274$
All-ints	$^3B_1$	$-38.953\,285$	$-38.955\,554$	$-38.955\,770$
		Energy difference (kcal/mol)[c]		
PS-1C[d]	$^1A_1$	0.015	0.032	0.024
PS-2C[e]	$^1A_1$	-0.005	0.005	0.004
PS-1C	$^3B_1$	-0.054	0.117	0.112
PS-2C	$^3B_1$	<0.001	0.013	0.013

[a] GVB-PP (3/6) wave function for the $^1A_1$ state and GVB-PP (2/4) for the $^3B_1$ state.

[b] All-integrals result from the GVB2P5 program.

[c] Energy difference, $E_{PS} - E_{AI}$.

[d] PS results with analytic two-electron integrals for monoatomic terms.

[e] PS results with analytic two-electron integrals for diatomic terms.

1. Geometry and basis sets

For CH_2, we used the geometries optimized by Harding and Goddard[12] with GVB-POL-CI calculations (see below).

Three different basis sets of increasing size were used:

(a) VDZp/DZp: The C uses the Dunning–Huzinaga valence double-zeta basis set[13,14] [based on $(9s/5p)$] with one set of d functions on C ($\alpha = 0.64$). The H uses the Dunning–Huzinaga basis (based on four s functions contracted to two, with the exponents scaled by[15] $\zeta = 1.2$) and one set of p functions on H ($\alpha = 1.0$).

(b) VDZ2p/TZ2p: Here we replaced the 4s H basis with the 6s (unscaled) basis[14] contracted to 3s triple zeta. The d and p polarization functions from VDZp/DZp were replaced by a pair of functions scaled by the factors $\sqrt{2.3}$ and $1/\sqrt{2.3}$ from the above values.

(c) VDZ2pn/TZ2p: To the basis in b was added a diffuse set of s and p functions on the C atom ($\alpha_s = 0.45$, $\alpha_p = 0.34$) optimized[13] for the negative ion of C.

2. Grids and dealiasing basis sets

Table I shows the dealiasing basis sets used for methylene. As in previous work,[5] the dealiasing basis sets are formed of SP and SPD blocks of atomic orbitals with the same exponents. This makes the assembly of the least-squares-fitting matrix more efficient. The molecular grid for CH_2 is generated using atomic grids similar to the ones used in studying glycine.[5] Thus, for methylene the sizes of the grids are 181, 409, and 1114 grid points.

3. Discussion

For the triplet state, the two C–H bonds are correlated, leading to the GVB-PP (2/4) wave function. For the singlet state, the two C–H bonds and the singlet lone pair are all correlated, leading to the GVB-PP (3/6) wave function. These wave functions have the same total number of orbitals, leading to a consistent level of description for the two states.

The results for methylene are summarized in Tables II and III. In Table II we compare the total energies for the various calculations. The all-integral results are expressed in hartrees, and the PS results are expressed as the difference from the all-integral results. These are quoted in kcal/mol to simplify comparison with bond energies and other chemical data.

For the singlet state using the PS-1C correction, the errors in total energy (GVB) range from 0.015 to 0.032 kcal/mol, which drops to a maximum error of 0.005 kcal/mol for PS-2C. In the triplet state the errors are a factor of 3 to 5 larger, ranging from 0.054 to 0.117 kcal/mol for PS-1C and up to 0.013 kcal/mol for PS-2C. The larger error for the triplet arises because the grid is less complete in the directions perpendicular to the molecule. The grid is more complete in directions parallel to the plane since each atom contributes its own grid. Such difficulties could be ameliorated by a more sophisticated grid design procedure. Similar energy differences are obtained for the HF wave functions, again a maximum error of about 0.120 kcal/mol was found for the triplet-state geometry with monoatomic corrections.

Table III shows the singlet–triplet gap obtained for the HF and GVB-PP wave functions. For both wave functions, the error is smaller than 0.09 kcal/mol using PS-1C corrections and smaller than 0.02 kcal/mol for the PS-2C corrections.

TABLE III. The singlet–triplet gap (ΔE_{ST}) for CH_2.[a]

Method	Wave function	Basis sets (carbon/hydrogen)		
		VDZp/VDZ_p	VDZ2p/$TZ2_p$	VDZ2pn/$TZ2_p$
All-ints[b]	HF	26.11	25.48	25.13
PS-1C[c]	HF	26.18	25.41	25.06
PS-2C[d]	HF	26.11	25.48	25.12
All-ints	GVB-PP	9.11	8.16	7.84
PS-1C	GVB-PP	9.18	8.07	7.75
PS-2C	GVB-PP	9.11	8.15	7.83

[a] All energies in kcal/mol. Experimental value 9.09 kcal/mol.

[b] See footnote b of Table II.

[c] See footnote c of Table II.

[d] See footnote d of Table II.

An important result is that the errors do not increase as the size of the basis set is increased. This shows that the PS method does not have any intrinsic problems in manipulating the diffuse functions present in the largest basis set. More significant is the fact that the same grid was used for all three basis sets, so that the number of one-electron potentials (3.1) only increases as M^2. This scaling is much better than the traditional all-integral method, where the number of two-electron integrals grows as M^4. The PS method is thus expected to be particularly advantageous for larger basis sets.

B. Silylene

For SiH_2 the two low-lying states have the character of the two states in CH_2 as in Eq. (4.1); however, the singlet state is the ground state (see Ref. 12 for a discussion). The studies of silylene emphasize two points: (a) to test the PS method for a second-row element, and (b) to test the use of effective core potentials for PS calculations. Two sets of calculations were carried out, one with all the core electrons present and the other using the shape and Hamiltonian consistent (SHC) effective core potential to replace them.

As discussed below, we find that use of the ECP leads to an enormous improvement of the accuracy for PS without atomic corrections (i.e., so that 1C and 2C corrections are *not* necessary). For the all-electron calculations, use of 1C and 2C corrections leads to results for SiH_2 comparable with those of CH_2, suggesting that PS will be effective for the later rows of the Periodic Table.

1. Geometry and basis sets

For the $^1A_1$ state, we used $R_{\mathrm{SiH}} = 1.508$ Å and $\Theta_{\mathrm{HSiH}} = 92.4°$, while for the $^3B_1$ state, we used $R_{\mathrm{SiH}} = 1.471$ Å and $\Theta_{\mathrm{HSiH}} = 118.2°$. [These were based on geometry optimizations using second-order Møller–Plesset perturbation theory (MP2) and unrestricted MP2 with the 6-31G^{**} basis (valence double zeta with polarization functions on Si and H)].

When no effective core potential is used, a Huzinaga valence double-zeta basis set (11*s*8*p*/4*s*3*p*) was used for silicon and a Huzinaga unscaled double-zeta basis set (4*s*/2*s*) was used for hydrogen. A set of *d* polarization functions ($\alpha = 0.42$) was used with the silicon basis, and a set of *p* polarization functions ($\alpha = 0.6$) was used with the hydrogen basis.

The SHC effective core potential (Rappé *et al.*[16]) was used to eliminate the [17] core electrons of Si. The corresponding VDZ basis set (3*s*3*p*/2*s*2*p*) was augmented with the same 3*d* polarization functions as before. The same VDZ*p* basis was used for hydrogen.

2. Grids and dealiasing basis sets

The dealiasing basis sets are chosen by the same method described for methylene, but slightly different atomic grids are used. For the silicon atom, the size of the bonding region is changed from 4.5 to 6.5 bohrs for the ultrafine grid. This increases the size of the grids from 1114 of carbon to 1389 for silicon. This was found to be necessary to achieve good accuracy.

TABLE IV. Energies comparison for SiH_2.

Calculation[a]	State	All-ints[b] Energy[f]	PS[c] error (kcal/mol)	PS-1C[d] error (kcal/mol)	PS-2C[e] error (kcal/mol)
		(a) All-electron			
HF	$^3B_1$	− 289.920 507	22.7	0.065	0.007
HF	$^1A_1$	− 289.929 006	22.6	0.080	0.025
ΔE_{ST}[g]		0.008 499	0.152	− 0.015	− 0.019
GVB[h]	$^3B_1$	− 289.940 694	22.7	0.062	0.008
GVB[i]	$^1A^1$	− 289.972 813	22.9	0.075	0.025
ΔE_{ST}		0.032 119	0.150	− 0.013	− 0.017
		(b) Si core electrons replaced by the effective potential			
HF	$^3B_1$	− 289.994 255	0.003	0.002	0.010
HF	$^1A_1$	− 290.003 034	0.066	0.063	0.024
ΔE_{ST}		0.008 779	− 0.064	− 0.061	− 0.014
GVB	$^3B_1$	− 290.014 773	0.003	0.001	0.004
GVB	$^1A_1$	− 290.048 020	0.068	0.063	0.024
ΔE_{ST}		0.033 247	− 0.066	− 0.061	− 0.020

[a] Experimental geometry with the VDZ*p* basis set for Si and H atoms.

[b] All-integral calculations.

[c] PS results with no atomic corrections.

[d] PS results with analytic bielectronic integrals for monoatomic terms.

[e] PS results with analytic bielectronic integrals for diatomic terms.

[f] All total energies in hartrees.

[g] Singlet–triplet gap $= E_{\mathrm{Triplet}} - E_{\mathrm{Singlet}}$.

[h] GVB-PP (2/4) wave function for the $^3B_1$ state.

[i] GVB-PP (3/6) wave function for the $^1A_1$ state.

3. Discussion

The errors obtained for SiH_2 are of the same order as those for CH_2. The maximum error is 0.079 kcal/mol when PS-1C corrections are included and 0.014 kcal/mol with PS-2C corrections. As can be seen in Table IV, the errors are always similar for the HF and GVB wave functions at a given geometry.

The presence of an effective core potential improves the accuracy of the PS method. This is because the largest errors in the PS method are caused by the tight core orbitals. Thus, when no 1C or 2C corrections are made, the error in the PS result is 22 kcal/mol for all electrons, which drops to less than 0.07 kcal/mol when the core electrons are replaced by the effective potential. On the other hand, with 1C or 2C corrections, the accuracy is about the same for all electrons as for ECP; this shows that the 1C and 2C corrections are very effective in removing large errors associated with core orbitals.

Note that for the triplet state with an effective core potential, the error using the PS method is unusually small, 0.001–0.002 kcal/mol. We believe this results from a fortuitous cancellation of errors arising from the particular geometries and has nothing to do with either the effective core potential or the nature of the wave function.

C. Ethylene

In the ethylene system we shift our focus away from electronic criteria (basis sets and effective core potentials) and concentrate on the effect of larger geometric changes on the accuracy of the PS method. We considered four geometries:

TABLE V. Summary of energy differences for C_2H_4.

			PS energy error (kcal/mol)[a]			
	All-integral total energy (hartrees)[b]		PS-1C[c]		PS-2C[d]	
State	HF	GVB[e]	HF	GVB	HF	GVB
		Experimental geometry ($R_{CC} = 1.338$ Å, $\phi = 0°$)				
$^3A_2$	− 77.931 907	− 77.942 874	− 0.137	− 0.119	− 0.036	− 0.038
$^1A_1$	− 78.060 723	− 78.099 240	− 0.243	− 0.209	− 0.045	− 0.045
ΔE_{ST}[f]	0.128 816	0.156 366	0.106	0.089	0.008	0.007
		Twisted geometry ($R_{CC} = 1.338$ Å, $\phi = 90°$)				
$^3A_2$	− 77.973 417	− 77.984 305	− 0.183	− 0.022	0.003	0.004
$^1A_1$	− 77.888 273	− 77.981 006	0.016	− 0.025	0.007	0.004
ΔE_{ST}	− 0.085 144	− 0.003 299	− 0.199	0.003	− 0.004	0.000
Barrier[g] $^3A_2$	− 0.041 510	− 0.041 431	0.046	− 0.097	− 0.039	− 0.042
Barrier $^1A_1$	0.172 450	0.118 234	− 0.259	− 0.184	− 0.053	− 0.049
		Stretched geometry ($R_{CC} = 2.676$ Å, $\phi = 0°$)				
$^3A_2$	− 77.772 934	− 77.884 587	− 0.199	− 0.196	− 0.013	− 0.013
^1A_I	− 77.726 946	− 77.847 384	− 0.128	− 0.209	− 0.038	− 0.016
ΔE_{ST}	− 0.045 988	0.002 797	− 0.070	0.013	− 0.009	0.003
Bond energy $^3A_2$	0.158 973	0.098 287	0.062	0.077	0.023	0.025
Bond energy $^2A_1$	0.333 777	0.251 856	0.115	0.000	0.007	0.029
		Stretched/twisted geometry ($R_{CC} = 2.676$ Å, $\phi = 90°$)				
$^3A_2$	− 77.773 688	− 77.845 388	− 0.128	− 0.104	0.003	− 0.013
$^1A_1$	− 77.727 351	− 77.845 284	− 0.139	− 0.104	0.012	0.013
ΔE_{ST}	− 0.046 337	− 0.000 104	0.011	0.000	− 0.006	− 0.000

[a] Energy difference, $E_{PS} - E_{RH}$.

[b] All total energies in hartrees.

[c] PS results with analytic bielectronic integrals for diatomic terms.

[d] PS results with analytic bielectronic integrals for monoatomic terms.

[e] GVB-PP (3/6) wave function for $^1A_1$ state and GVB-PP (2/4) for $^3A_2$ state.

[f] Single–triplet gap $E_{\text{triplet}} - E_{\text{singlet}}$.

[g] $E(\phi = 90°) - E(\phi = 0)$.

(a) Planar: The ground-state geometry with a torsion angle of $\phi = 0°$, and $R_{CC} = 1.338$ Å. The other geometric parameters are $R_{CH} = 1.085$ Å, $\Theta_{HCH} = 117.8°$.

(b) Twisted: The torsion angle is $\phi = 90°$ but no other parameters are changed. (The equilibrium twisted geometry would have a larger CC bond and HCH angle.)

(c) Stretched: With $\phi = 0°$ the CC bond distance is doubled to $R_{CC} = 2.676$ Å (all other quantities fixed). At this distance the bond is broken,

(d) Stretched/twisted: With $R_{CC} = 2.676$ Å, we twist to $\phi = 90°$. Since there is essentially no bond, the energy changes should be small.

We find that with PS-1C the errors in the GVB wave functions have a maximum error of 0.09 kcal/mol, while PS-2C leads to a maximum error of 0.007 kcal/mol. These are quite acceptable results and indicate that PS-GVB has sufficient accuracy for chemical applications.

1. Basis sets, grids, and wave functions

For all of the calculations on the ethylene system we use the VDZ_p/VDZ_p basis described above for CH_2 and the same grid and dealiasing set.

In these calculations we correlate the two pairs of electrons associated with the CC bond, leading to a GVB-PP (2/4) wave function for the singlet and a GVB-PP (1/2) wave function for the triplet state.

2. Discussion

Table V summarizes the calculations performed on ethylene. The first two columns tabulate the standard all-integral results for both HF and GVB-PP wave functions for each of the four geometries. Included are the absolute singlet and triplet energies, the singlet–triplet gap, and the rotational barriers.

The next four columns detail the corresponding results for PS calculations. Columns three and four include results with 1C corrections, and columns five and six include results with 2C corrections. The PS results are expressed as a difference from the all-integral results and are quoted in kcal/mol to indicate the significance for bond energies and other chemical data.

We note that with 1C corrections the absolute energies for HF wave functions agree with the analytic ones to within 0.25 kcal/mol, and the relative energies agree to within 0.2 kcal/mol. For GVB wave functions the absolute energies agree to within 0.2 kcal/mol, and the relative energies to within 0.1 kcal/mol.

For calculations using the 2C corrections, the absolute energies for HF wave functions agree with the analytic ones to within 0.05 kcal/mol, and the relative energies to within 0.008 kcal/mol. For GVB wave functions the similar results are 0.05 and 0.007 kcal/mol.

For singlet–triplet excitation energies (same geometry), the PS-1C calculations lead to errors of 0.01 to 0.20 kcal/mol for HF and 0.00 to 0.09 kcal/mol for GVB. These errors are quite acceptable. With PS-2C the error drops to a maximum of 0.009 kcal/mol, which is negligible.

In general, for ethylene, the PS method does a better job at reproducing the results for GVB wave functions than it does for HF wave functions. We presume this to be an artifact to the particular grids and dealiasing sets used, since we see no physical reason for this effect.

V. DISCUSSION

We find that the PS method with 1C corrections leads to an accurate description of the GVB-PP wave functions and that the energy differences for singlet–triplet excitations and bond dissociation are well described. In addition, we find that with use of the effective potential (ECP) on Si, the uncorrected PS results are quite accurate. This suggests that PS approaches will be useful in calculations of chemically interesting phenomena, such as excitation energies and bond dissociation processes.

The code used to carry out the calculations reported above was obtained by interfacing the standard GVB2P5 program[18] with the original PS HF program. The combined code has not yet been optimized for efficiency, so we do not present timing results in this paper. An optimized implementation is now nearing completion so that timing tests should be available soon. Preliminary results indicate that substantial reductions in CPU time and disk storage similar to those obtained for HF studies of large systems can be achieved for GVB calculations.

With regard to higher levels of electron correlation, the PS method can be applied to general CI calculations if the CI matrix can be expressed in terms of Coulomb and exchange operators (the integrals are evaluated each iteration, a direct method). In these cases, a factor of M (as in the HF and GVB algorithms) can be gained in formulating the action of the CI matrix on a trial vector. Of course, the actual performance of the method for such purposes is best tested with calculations.

A particularly attractive application would be to some of the less computationally demanding forms of CI that could be implemented for very large systems. Among these are restricted CI, which properly spin couples the GVB-PP wave function, and correlation-consistent CI, in which only a few electron pairs are allowed to enter the virtual space.[17,19] We plan to begin implementation of these methods in the near future.

It is straightforward to incorporate analytic first and second derivatives into PS calculations and such work is in progress.

These studies have given us some additional insight into the role of the grids in determining accuracy, and we believe that it will be possible to develop similar sized grids with even higher accuracy.

The 2C corrections give results that are more accurate than the 1C corrections by a factor of about three or more. However, 2C corrections require greater computation time than is required for 1C corrections, and we have not determined whether adding diatomic corrections is faster than simply increasing the grid size.

VI. CONCLUSIONS

We find that the PS algorithm for the direct evaluation of Coulomb and exchange operators leads to accurate results for GVB wave functions, suggesting a wide range of applications in quantum chemistry. The PS method provides uniform accuracy for a range of basis sets (including diffuse functions of *s*, *p*, and *d* with various exponents), for systems with effective core potentials, and for a wide range of geometric conformations (including bond dissociation). Additional studies are underway using the PS algorithm for transition-metals complexes. In summary, we believe that the PS algorithm has a future in *ab initio* quality studies of molecular systems.

ACKNOWLEDGMENTS

This work was supported in part by grants to R.A.F. from the National Institutes of Health and the National Science Foundation, and in part by grants to W.A.G. from the National Science Foundation (Grant No. CHE-8318041) and the Air Force Office of Scientific Research (Grant No. AFOSR-88-0051). R.A.F. is a Camille and Henry Dreyfus Teacher-Scholar and the recipient of a Research Career Development Award from the National Institutes of Health, Institute of General Medical Sciences. W.A.G. acknowledges the support of the Endowment Fund for the Charles and Mary Ferkel Chair.

[1] R. Friesner, Chem. Phys. Lett. **116**, 39 (1985).

[2] R. Friesner, J. Chem. Phys. **85**, 1462 (1986).

[3] R. Friesner, J. Chem. Phys. **86**, 3522 (1987).

[4] R. Friesner, J. Phys. Chem. **92**, 3091 (1988).

[5] M. Ringnalda, Y. Won, and R. Friesner, J. Chem. Phys. **92**, 1163 (1990).

[6] M. Ringnalda, M. Belhadj, and R. A. Friesner (unpublished).

[7] R. C. Ladner and W. A. Goddard III, J. Chem. Phys. **51**, 1073 (1969); W. A. Goddard III, J. Am. Chem. Soc. **94**, 793 (1972).

[8] W. A. Goddard III and L. B. Harding, Annu. Rev. Phys. Chem. **29**, 363 (1978).

[9] F. W. Bobrowicz and W. A. Goddard III, in *Modern Theoretical Chemistry: Methods of Electronic Structure Theory*, edited by H. F. Schaefer III (Plenum, New York, 1977), Vol. 3, p. 79.

[10] L. G. Yaffe and W. A. Goddard III, Phys. Rev. A **13**, 1682 (1976).

[11] W. J. Hunt, W. A. Goddard III, and T. H. Dunning, Jr., Chem. Phys. Lett. **6**, 147 (1970).

[12] L. B. Harding and W. A. Goddard III, J. Chem. Phys. **67**, 1777 (1977).

[13] T. H. Dunning, Jr. and P. J. Hay, in *Modern Theoretical Chemistry: Methods of Electronic Structure Theory*, edited by H. F. Schaefer III (Plenum, New York, 1977), Vol. 3, Chap. 1.

[14] S. Huzinaga, J. Chem. Phys. **42**, 1293 (1965).

[15] When discussing scale factors for basis functions, the use of ζ denotes an effective Slater exponent, whereas α denotes a Gaussian exponent. Thus, for H, $\zeta = 1.2$ implies that the standard exponents for the Gauss-

ian expansion have all exponents scaled by 1.44.

[16] A. K. Rappé, T. A. Smedley, and W. A. Goddard III, J. Phys. Chem. 85, 1662 (1981); P. J. Hay and W. R. Wadt, J. Chem. Phys. 82, 270 (1985).

[17] S. K. Shin, W. A. Goddard III, and J. L. Beauchamp, J. Chem. Phys. (to be submitted).

[18] W. A. Goddard III, R. A. Bair, F. W. Bobrowicz, W. R. Wadt, L. G. Yaffe, and A. K. Rappé, Ref. 9; R. A. Bair, Ph.D. thesis, California Institute of Technology, 1975 (unpublished).

[19] E. A. Carter and W. A. Goddard III, J. Chem. Phys. 86, 862 (1987).

Chapter 3

Local Density Approximation Electronic Structure Calculations using Gaussian Type Orbitals

Abstract

We present a method for large-scale electronic calculations in the framework of local density approximation. The Bloch orbitals are expanded in terms of real-space atom-centered Gaussian type orbitals. The Coulomb potential is evaluated over a set of gridpoints using an analytical dual-space approach. The approach is based on an extension of the Ewald method to electron density where every pair of basis functions in the density is screened independently. The Fock matrix elements are evaluated by replacing the three-dimensional analytical integrations over the whole space by a set of atom-centered numerical integrations via a projection technique. The self-consistent field is obtained by a modified conjugate-gradient method with preconditioning which also avoids expensive minimization.

This method was used to calculate the band structures of the diamond and C_{60} allotropes of carbon. Standard double zeta basis sets were used and the results are compared to other recent theoretical calculations and experimental measurements.

I. Introduction

The standard approach to calculating electronic wavefunctions of solids uses plane waves as basis functions. To make such calculations practical it is necessary to use pseudopotentials in order to avoid the large number of plane waves which would otherwise be required to describe the core electrons. This allows the focus to be on valence electrons which are usually well described as band orbitals. For larger unit cells the bottleneck of such calculations becomes Fock matrix diagonalization, which scales as N^3 where N is the number of plane waves at each point in the Brillouin zone. For example, about 27000 plane waves were required for calculations on C_{60} crystal.[1] In global minimization approaches like the Car-Parrinello method[2] or the conjugate-gradient method,[3] the diagonalization step is avoided and the time limiting factor becomes the calculation of forces, which scales as N^2n where n is the number of occupied orbitals.

An alternative approach is to use localized atom-centered basis functions. This is essential if the pseudopotential approximation is to be avoided. Such localized basis sets can be much smaller, typically a factor of 5 to 10 for a close packed structure. For C_{60}, the atoms are far from being compactly packed and the atom-centered basis may be 20 to 40 times smaller than the plane wave basis.

Gaussian type orbitals (GTO's) have been used extensively and successfully in molecular systems. Their popularity is based on the ease with which multicenter integrals can be computed analytically (as compared to Slater type orbitals) and on the accuracy for treating the bonding in molecules. Many approaches have been explored to speed up the evaluation of multicenter integrals over GTO's.[4–6] The most powerful methods involve recursion relations to improve the evaluation of integrals involving high angular momentum orbitals.[7]

In our approach the Fock matrix is constructed by first calculating the Coulomb and exchange-correlation field on a set of gridpoints and then integrating the field with a pair of basis functions. The integration is done numerically over the gridpoints. It is only recently that numerical integrations with reliable accuracy have been developed.[8–10] One variation consists of dividing the space into regions of different shapes over which independent integrations are performed. The most important regions are the regions around the atomic nuclei, where atomic fields are very large, requiring high accuracy. This leads naturally to defining a spherical region around each atom inside which the integration is performed in spherical coordinates. The remaining space is divided into regular polyhedra and truncated polyhedra having a spherical boundary. The location and weight of the gridpoints are relatively easy to select in the case of the spherical regions and regular polyhedra. The difficulty is in choosing gridpoints for the truncated polyhedra. Proposed solutions have been to map the truncated polyhedra back onto a regular polyhedra[9] and to use transformed Gaussian-quadrature mesh with a continuous parameter which is adjusted variationaly to minimize an objective function.[10]

A different approach, which avoids the truncated polyhedra regions, is based on transforming the integration over all space into atom centered integrations via a projection technique.[8] The atomic projection functions have the properties that they have unit value on the atom on which they are centered, and that they go to zero on all other atoms. This is done by defining the projection function as a product of diatomic functions. The challenge in this case is to find the best diatomic functions. In Becke's original work, the diatomic functions were iterative in nature which means that only a discrete set of functions were obtained. We have modified the Becke approach to include a continuous degree of freedom which can be adjusted to improve the accuracy. The atom centered integrations use spherical grids decomposable in radial and angular components. The angular integrations use Gaussian-type formulas for selecting the location and weight of the gridpoints.[11] For the radial integrations, we use a geometrically scaled mesh with the weights chosen to exactly integrate Gaussian functions. This takes advantage of the Gaussian nature of the basis set.

In GTO based methods, the computationally most intensive step is not the Fock diagonalization (as in plane wave methods), but rather the evaluation of the Coulomb field over the grid. We evaluate the Coulomb field efficiently by using a dual-space approach similar in spirit to the Ewald method.[12] In the dual-space approach, the density is written as a sum of pairs of GTOs where each pair is screened independently. The screening density for each pair of functions is defined in such a way that the potential generated by the combined density (normal plus screening density) goes rapidly to zero away from the product center of the GTO. This *perfect screening* is possible because we are using GTOs.[13] The Coulomb field of the combined density now converges rapidly in real space whereas the field of the opposite screening density converges rapidly in reciprocal space. Like in the Ewald method, the screening density is more slowly varying than the normal density. By choosing the screening exponent appropriately, the evaluation of the Coulomb field scales as $N^{1.5}$ with proper cutoffs. This approach contrasts with the popular density fitting procedure[14] which has the disadvantage of requiring more expensive three-center integrals of GTO's instead of the two-center integrals of the current approach.

The density fitting procedure also leads to extra components in the atomic forces when doing geometry optimization[15] which are subject to instabilities due to aliasing problems.

A critical component of any self-consistent field (SCF) theory is the iterative procedure used to achieve convergence. A simple technique is the linearization technique of te Velde and Baerends,[16] denoted VB. In this method a new Fock matrix is expressed as a linear combination of the Fock matrices from previous cycles. If the fit is accurate, then the new density is obtained by the same linear combination of the corresponding density matrices. This could be improved by implementing the Direct Inversion in the Iterative Subspace (DIIS) method[17] used in molecular Hartree-Fock (HF) calculations. In DIIS the new Fock matrix is a linear combination of the current and previous Fock matrices where the coefficients of the matrices are obtained by minimizing the corresponding linear combinations of the error matrices (deviation from convergence). The major drawback of both VB and DIIS is the large amount of disk space required to store the information from the previous iterations. A simpler approach is the conjugate-gradient technique[18] but this often exhibits unacceptably slow convergence. By preconditioning the conjugate-gradient vectors the number of iterations required for self-consistency can be greatly reduced.[3] In our approach the preconditioning is obtained by a Green function-like operator. Essentially, the expensive-to-evaluate second-order term, required for the fast convergence of Newton-Raphson type methods, is replaced by an inexpensive but effective approximation.

II. Density Functional Theory (DFT)

Hohenberg and Kohn[19] showed that the electron density can be used as the basic variable in minimizing the energy of an electronic system. However,the antisymmetry implied by the Pauli principle leads to a kinetic energy functional whose character precludes a simple and accurate variational method based only on the electron density. The simplest solution is the Kohn-Sham (KS) method[20] where an orbital representation of the density is used to obtain a simple expression for the kinetic energy evaluation. The KS method consists of mapping the original system of interacting electrons onto a system of independent electrons moving in an effective external potential:

$$[-\frac{1}{2}\nabla^2 + V_{eff}(r)]\,\phi_{nk}(r) = \epsilon_{nk}\,\phi_{nk}(r) \tag{1}$$

where $\phi_{nk}(r)$ is the one electron state with wavevector k, band index n and energy ϵ_{nk}; the kinetic energy operator is given by $-\frac{1}{2}\nabla^2$ in atomic units.

The effective potential in equation (1) is written as:

$$V_{eff}(r) = V_{nuc}(r) + J(r) + V_{xc}(r) \tag{2}$$

where $V_{nuc}(r)$ is the Coulomb potential generated by the nuclear charges, $J(r)$ is the Coulomb potential generated by the electron density and everything else is lumped into the term $V_{xc}(r)$ referred to as the exchange-correlation potential. The nuclear charge potential is given by:

$$V_{nuc}(r) = \sum_{aR} \frac{Z_a}{|r - \xi_a - R|}, \tag{3}$$

where R labels the unit cells in the crystal and a labels the atoms in the primitive unit cell. The atom locations are given by ξ_a and the atomic charges by Z_a. The electronic Coulomb potential is given

$$J(r) = \int d^3r'\ \frac{\rho(r')}{|r - r'|}, \tag{4}$$

where $\rho(r)$ is the electron density. Both of these electrostatic potentials involve infinite sums which are dealt with by a dual-space approach in the spirit of the Ewald method.

The exchange-correlation potential $V_{xc}(r)$ written as the functional

$$V_{xc}(r) = \delta E_{xc}[\rho]/\delta\rho(r), \tag{5}$$

where E_{xc} is referred to as the exchange-correlation energy. The exact functional form of the exchange-correlation energy is unknown, but local approximations,

$$E_{xc}[\rho] = \frac{1}{N_{cell}} \int d^3r \;\; \rho(r)\epsilon_{xc}[\rho(r)], \tag{6}$$

have been developed over the years. In the above equation, $\epsilon_{xc}[\rho]$ is the exchange-correlation energy per electron, which can be obtained numerically from studies of interacting electron system of constant density. We have used the exchange-correlation potential as parametrized by Perdew and Zunger[21] from the Quantum Monte Carlo simulations of Ceperley and Alder.[22] We also included the relativistic correction of Bachelet.[23] The exact form of the exchange-correlation potential is given in Appendix A. We also used other exchange-correlation functionals for comparison purposes. The overall theory described above is referred to as Density Functional Theory (DFT).

In DFT, the total energy per unit cell can be evaluated as follows:

$$E_{total} = E_{kin} + E_{xc} + E_{e-e} + E_{e-n} + E_{n-n} \tag{7}$$

where E_{kin} is the kinetic energy, E_{xc} is the exchange-correlation energy defined above and the last three terms are the various electrostatic energies. The kinetic energy is given by

$$E_{kin} = \sum_{nk} 2\,\Theta(\epsilon_{nk} - \epsilon_{Fermi}) < \phi_{nk}| - \frac{1}{2}\nabla^2|\phi_{nk} >, \tag{8}$$

where $\Theta(\epsilon_{nk} - \epsilon_{Fermi})$ is the Heavyside step function with ϵ_{Fermi} the energy of the Fermi level. Here k denotes the wavevector and n is the band index. The expressions for the electrostatic energies are:

$$E_{e-e} = \frac{1}{2}\frac{1}{N_{cell}} \int d^3r \int d^3r' \; \frac{J(r)\rho(r')}{|r - r'|} \tag{9}$$

for the electron-electron repulsion,

$$E_{e-n} = \frac{1}{N_{cell}} \int d^3r \int d^3r' \; \frac{V_{nuc}(r)\rho(r')}{|r - r'|} \tag{10}$$

for the electron-nuclear attraction, and

$$E_{n-n} = \sum_{a} \sum_{b,R'} \frac{Z_a Z_b}{|\xi_a + R - \xi_b - R'|}, \tag{11}$$

for the nuclear-nuclear repulsion.

The orthogonal one electron band orbitals $\phi_{nk}(r)$, can be expanded in linear combinations of Bloch basis functions $X_{uk}(r)$:

$$\phi_{nk}(r) = \sum_{u} C_{un}(k) X_{uk}(r). \tag{12}$$

The expansion coefficients $C_{un}(k)$ obey the orthogonality constraint:

$$\delta_{nm} = \sum_{uv} C^*_{un}(k) S_{uv}(k) C_{vm}(k), \tag{13}$$

where the overlap matrix elements $S_{uv}(k)$ between Bloch basis functions are given by:

$$S_{uv}(k) =< X_{uk}|X_{vk} >= \int d^3r \;\; X^*_{uk}(r) X_{vk}(r). \tag{14}$$

A. Basis Set Expansion

The expansion of the Bloch basis functions in terms of atomic orbitals(AO) is:

$$X_{uk}(r) = \frac{1}{\sqrt{N}} e^{ik\xi_u} \sum_R e^{ikR} X_{uR}(r), \tag{15}$$

where X_{uR} denotes an AO composed of one or many Gaussian functions centered on atom ξ_u in unit cell R:

$$\begin{aligned} X_{uR}(r) = \sum_i^{primitives} & c_{ui} N_{ui} (x - \xi_{ux} - R_x)^{lx} (y - \xi_{uy} - R_y)^{ly} \\ & \times (z - \xi_{uz} - R_z)^{lz} e^{-\alpha_i (r - \xi_u - R)^2}. \end{aligned} \tag{16}$$

Here i labels the primitive Gaussian functions, lx, ly, lz label the exponents of the x, y, z component of the primitive function. The coefficients of the contraction are denoted by c_{ui} and the normalization constant for each Gaussian function is given by N_{ui}. The number of primitives is typically large ($\sim 5-10$) for AO representing core orbitals and is small ($\sim 1-2$) for AO representing valence orbitals or unoccupied orbitals.

The Bloch basis functions have translational symmetry:

$$X_{uk}(r + R) = e^{ikR} X_{uk}(r) \tag{17}$$

and

$$X_{uk+G}(r) = e^{iG\xi_u} X_{uk}(r), \tag{18}$$

where equation (17) is the Bloch theorem and equation (18) comes from the phase factor $e^{ik\xi_u}$ in equation (15). The Bloch basis functions are orthogonal for different wavevectors, $k \neq k'$:

$$< X_{uk} | X_{vk'} >= 0. \tag{19}$$

However, Bloch basis functions with the same wavevector are not generally orthogonal,

$$S_{uv}(k) \neq \delta_{uv}. \tag{20}$$

The spatial electronic density is given by:

$$\rho(r) = 2 \sum_{nk} W_k \; |\phi_{nk}(r)|^2 \; \Theta(\epsilon_{nk} - \epsilon_{Fermi}) \tag{21.a}$$

$$\rho(r) = \sum_{uvR} D_{uvR} \sum_{R'} X_u(r-R') \; X_{vR}(r-R'), \tag{21.b}$$

where the factor of 2 is for the spin degeneracy and W_k is the wavevector-dependent weight used in the integration over the Brillouin zone. In equation (21.b) the density matrix in real space, D_{uvR}, is defined as:

$$D_{uvR} = \sum_k e^{ik(\xi_v - \xi_u)} \; e^{ikR} \; D_{uv}(k), \tag{22}$$

where the k space density matrix is given by:

$$D_{uv}(k) = 2 \sum_n W_k \; C^*_{un}(k) \; C_{vn}(k) \; \Theta(\epsilon_{nk} - \epsilon_{Fermi}). \tag{23}$$

For the evaluation of the spatial density over gridpoints, $\rho(r_g)$, equation (21.b) is less expensive computationally than equation (21.a) because efficient cutoffs can be applied to the construction of the atomic orbitals $X_u(r_g)$ and $X_{vR}(r_g)$ over the gridpoints.

B. Evaluation of Effective Hamiltonian

In the Kohn-Sham method the matrix elements of the effective Hamiltonian between a pair of Bloch basis functions, $X_{uk}(r)$ and $X_{vk}(r)$, are given by:

$$H_{uv}(k) = T_{uv}(k) + V_{uv}(k). \tag{24}$$

The kinetic energy matrix elements T_{uvk} are given by

$$T_{uv}(k) = < X_{uk}| - \frac{1}{2}\nabla^2 |X_{vk} > \tag{25}$$

which can be evaluated analytically. The effective potential matrix elements $V_{uv}(k)$ include the exchange-correlation and electrostatic potentials. They are evaluated numerically over a set of gridpoints within the first primitive cell:

$$V_{uv}(k) = \sum_{g}^{cell} W_g \; X_{uk}(r_g) \; [V_{xc}(r_g) + J(r_g) + V_{nuc}(r_g)] \; X_{vk}(r_g) \tag{26}$$

with W_g being the weight of gridpoint r_g.

In equation (26) the matrix elements of the nuclear attraction potential $V_{nuc}(r_g)$ are evaluated numerically so that the $1/r$ tails of $V_{nuc}(r_g)$ and $J(r_g)$ cancel exactly. In molecular application of DFT, we found that this cancellation is crucial in minimizing numerical integration errors.

For a purely covalent system, such as C_{60}, each atom is neutral and the $1/r$ tail is effectively eliminated. This is not the case for ionic systems. For those systems we add and subtract the field created by the net atomic charge:

$$Q^{(a)}_{net} = Z_a + Q^{(a)}_{elec} \tag{27}$$

where $Q^{(a)}_{elec}$ is the effective atomic electronic charge. This effective electronic charge is evaluated from the Mulliken population:

$$Q^{(a)}_{elec} = \sum_{u}^{on(a)} \sum_{v}^{all} \sum_{k} \Theta(\epsilon_{nk} - \epsilon_{Fermi}) \; S_{uv}(k) \; D_{uv}(k), \tag{28}$$

where $D_{uv}(k)$ is the density matrix for each wavevector k. We should emphasize that this definition of the effective electronic charge is not unique. However, Mulliken populations are simple and well behaved.

In crystalline systems the Coulomb potential and the nuclear attraction potential are both evaluated using an Ewald-like method . This guarantees that the two potentials have zero average value:[24]

$$0 = \int d^3r \ \ J(r) \tag{29}$$

$$0 = \int d^3r \ \ V_{nuc}(r). \tag{30}$$

This means that the potential $V_{nuc}(r) + J(r)$ does not tend to go to zero in the interatomic regions but instead go to to some constant value V_{shift} introduced by the Ewald reference point of the potential. Since the numerical integration is least accurate precisely in the interatomic regions, we want to subtract the constant potential shift, V_{shift}, from the numerical integration and add its contribution analytically.

Taking into account the two previous points, the final expression for constructing the Hamiltonian is given by:

$$\begin{aligned} H_{uv}(k) &= T_{uv}(k) + V_{uv}^{net} + S_{uv}(k)V_{shift} \\ &+ \sum_{g}^{cell} W_g \ X_{uk}(r_g) \ \ (V_{xc}(r_g) + J(r_g) + V_{nuc}(r_g) - V_{net}(r_g) - V_{shift}) \ \ X_{vk}(r_g), \end{aligned} \tag{31}$$

where $V_{net}(r_g)$ is the field of the net atomic charges $Q_{net}^{(a)}$ and V_{uv}^{net} is the analytical matrix of the same potential evaluated for a pair of basis functions.

C. Evaluation of Total Energy

The total energy is given by:

$$E_{total} = E_{kin} + E_{xc} + E_{e-e} + E_{e-n} + E_{n-n} \tag{7}$$

where the exchange-correlation energy E_{xc} and the Coulomb energy E_{e-e} are evaluated by numerical integration over the gridpoints:

$$E_{xc} = \sum_{g}^{cell} W_g \, \epsilon_{xc}(r_g) \, \rho(r_g), \tag{32}$$

$$E_{e-e} = \frac{1}{2} \sum_{g}^{cell} W_g \, J(r_g) \, \rho(r_g). \tag{33}$$

The same accuracy enhancing features applied to the construction of $H_{uv}(k)$ can also be applied to the Coulomb energy evaluation. This is done most simply by evaluating the Coulomb energy as part of the sum of the one electron orbital energies:

$$E_{one} = \sum_{nk} 2 \, \Theta(\epsilon_{nk} - \epsilon_{Fermi}) \epsilon_{nk}. \tag{34}$$

The total energy is then recast in term of E_{one}:

$$E_{total} = E_{xc} + E_{n-n} + \frac{1}{2}(E_{one} + E_{kin} + E_{e-n} - V_{sum,xc}). \tag{35}$$

The subtraction of the exchange-correlation potential from E_{one} is done by:

$$V_{sum,xc} = \sum_{g}^{cell} W_g \, V_{xc}(r_g) \, \rho(r_g). \tag{36}$$

The electron-nuclear attraction energy is evaluated analytically as:

$$E_{e-n} = \sum_{uvR} V_{uvR} \, D_{uvR} \tag{37}$$

and the matrix elements of the electron-nuclear attraction are given by:

$$V_{uvR} = \int d^3r \;\; X_u(r) V_{nuc}(r) X_{vR}(r), \tag{38}$$

which can be evaluated analytically using the dual-space approach. The kinetic energy is also evaluated via the real space density matrix:

$$E_{kin} = \sum_{uvR} T_{uvR}\, D_{uvR}. \tag{39}$$

Finally the nuclear-nuclear repulsion is evaluated using the standard Ewald method.

The overlap and kinetic energy matrix elements in k space are calculated by Fourier transform of the matrix elements in R space:

$$S_{uv}(k) = e^{ik(\xi v-\xi u)} \sum_{R} e^{ikR} S_{uvR} \tag{40}$$

and

$$T_{uv}(k) = e^{ik(\xi v-\xi u)} \sum_{R} e^{ikR} T_{uvR}, \tag{41}$$

where

$$S_{uvR} = \int d^3r \;\; X_u(r) X_{vR}(r) \tag{42}$$

and

$$T_{uvR} = -\frac{1}{2} \int d^3r X \;\; _u(r) \nabla^2 X_{vR}(r). \tag{43}$$

The evaluations of S_{uvR} and T_{uvR} are performed analytically using recursion relations described in Appendix B.

III. Dual-Space Approach

A. Ewald Method

In calculating the electrostatic forces we need to sum over the infinite set of lattice vectors. In order to make this sum convergent, we have used the Ewald method and its extension to a continuous distribution of charges.

Following the derivation of Tosi,[25] a non-converging lattice sum like the nuclear attraction potential:

$$V_{nuc}(r) = \sum_{a}^{atoms} \sum_{R}^{lattice} \frac{Z_a}{|r - \xi_a - R|}, \tag{3}$$

is made convergent by subtracting and adding a screening field $V_{Gauss}(r)$ created by an infinite set of atom centered functions:

$$V_{Gauss}(r) = \sum_{aR} Z_a \int d^3r' \ \frac{g_\alpha(r' - \xi_a - R)}{|r - r'|}. \tag{44}$$

If the screening functions are chosen as normalized Gaussian densities, $g_\alpha(r)$ defined by:

$$g_\alpha(r) = \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} e^{-\alpha r^2}, \tag{45}$$

then the field can be rewritten as:

$$V_{Gauss}(r) = \sum_{aR} Z_a \frac{erf(\sqrt{\alpha}|r - \xi_a - R|)}{|r - \xi_a - R|} \tag{46}$$

in term of the error function, $erf(x)$,

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x dt \ e^{-t^2}. \tag{47}$$

The combined field of the nuclear charges minus the screening density is given by:

$$V_{Rsum}(r) = \sum_{aR} Z_a \left(\frac{1 - erf(\sqrt{\alpha}|r - \xi_a - R|)}{|r - \xi_a - R|} \right). \tag{48}$$

This is now convergent since for large value of R the field created by the Gaussian density $g_\alpha(r' - \xi_a - R)$ goes as $1/|r - \xi_a - R|$ in the limit that $|r - \xi_a - R| >> 1/\sqrt{\alpha}$.

The compensating screening field contribution is transformed to reciprocal space by using the relationship:

$$\sum_R \frac{1}{|r-R|} = \frac{4\pi}{V_{cell}} \sum_{G\neq 0} \frac{e^{iGr}}{G^2} \tag{49}$$

where the G vectors are reciprocal lattice vectors and V_{cell} is the volume of the unit cell. The result is

$$V_{Gsum}(r) = \sum_a Z_a \left(\frac{4\pi}{V_{cell}} \sum_{G\neq 0} \frac{e^{-G^2/4\alpha} e^{iG(r-\xi_a)}}{G^2} \right), \tag{50}$$

which converges rapidly for large G vectors because of the exponential factor.

The screening exponent α is chosen such that both the real space sum and reciprocal space sum converge equally rapidly. The relevant scale for choosing α is $1/V_{cell}^{2/3}$

By omitting the $G=0$ term in the sum over reciprocal vectors we have neglected two contributions. The first contribution is the $1/G^2$ singularity which is easily removed by adding a constant background density of value $\sum_a Z_a/V_{cell}$ that compensates the original nuclear charges. The second contribution is a constant whose value depends on the exact nature of the charge distribution at the surface of the crystal. Since in real crystals the surface charge distribution (nuclear plus electronic) is generally unknown, the constant is set to zero. Thus the reciprocal potential is only defined up to a constant value. The constant value is conventionally chosen such that the average value of the potential $V_{nuc}(r)$ is equal to zero. From equations (48) and (50), the average value of $V_{Gsum}(r)$ is zero because the $G=0$ has been omitted. The average value of V_{Rsum} is evaluated by using equation (49) into equation (48) and keeping only the $G=0$ term which gives the average value when divided by V_{cell}:[26]

$$
\begin{aligned}
< V_{Rsum}(r) > &= \frac{1}{V_{cell}} \int d^3 r V_{Rsum}(r) \\
&= (\sum_a Z_a) \frac{4\pi}{V_{cell}} \lim_{G^2 \to 0} \frac{1 - e^{G^2/4\alpha} e^{iG(r-\xi_a)}}{G^2} \\
< V_{Rsum}(r) > &= (\sum_a Z_a) \frac{\pi}{\alpha V_{cell}}. \qquad (51)
\end{aligned}
$$

The final expression for evaluating the potential generated by the nuclear charges is:

$$
\begin{aligned}
V_{nuc}(r) = &\sum_a Z_a \left(\sum_R \frac{1 - erf(\sqrt{\alpha}|r - \xi_a - R|)}{|r - \xi_a - R|} - \frac{\pi}{\alpha V_{cell}} \right) \\
&+ \sum_a Z_a \left(\frac{4\pi}{V_{cell}} \sum_{G \neq 0} \frac{e^{-G^2/4\alpha} e^{iG(r-\xi_a)}}{G^2} \right). \qquad (52)
\end{aligned}
$$

In a similar fashion, the potential $V_{net}(r)$ from the net atomic charges $Q_{net}^{(a)}$ is obtained from equation (52) by replacing the nuclear charges Z_a by the charges $Q_{net}^{(a)}$.

The nuclear-nuclear repulsion energy is most simply written as $E_{n-n} = \sum_a Z_a V_{nuc}(\xi_a)$. However, the singular self-repulsion terms for $R = 0$ need to be removed in the sum over real space vectors. The final expression is:

$$
\begin{aligned}
E_{n-n} = &\sum_{ab} Z_a Z_b \left(\sum_{R \neq 0} \frac{1 - erf(\sqrt{\alpha}|\xi_a - \xi_b - R|)}{|\xi_a - \xi_b - R|} - 2\frac{\sqrt{\alpha}}{\sqrt{\pi}} - \frac{\pi}{\alpha V_{cell}} \right) \\
&+ \sum_{ab} Z_a Z_b \left(\frac{4\pi}{V_{cell}} \sum_{G \neq 0} \frac{e^{-G^2/4\alpha} e^{iG(\xi_a - \xi_b)}}{G^2} \right), \qquad (53)
\end{aligned}
$$

which is the standard form using Ewald method.[25]

B. Application to Electron Density

So far we have applied the Ewald method to the nuclear point charges. We now describe how the lattice sums are evaluated for the continuous electron distribution. As defined previously the electron-nuclear attraction energy per cell is given by:

$$E_{e-n} = \sum_{uvR} V_{uvR} D_{uvR} \tag{37}$$

with

$$V_{uvR} = \int d^3r \ \ X_u(r) X_{vR}(r) V_{nuc}(r). \tag{38}$$

In this form the matrix element V_{uvR} can be interpreted as the energy of the basis function pair density, $\rho_{uvR}(r) = X_u(r) X_{vR}(r)$, in the field $V_{nuc}(r)$ of the periodic nuclear charges. By using equation (52), V_{uvR} can be rewritten as:

$$\begin{aligned} V_{uvR} = & \sum_{aR'} Z_a \big((g_{\infty(r'-\xi_a-R')} || \rho_{uvR(r)}) - (g_{\alpha(r'-\xi_a-R')} || \rho_{uvR(r)}) \big) \\ & - \frac{\pi}{\alpha V_{cell}} (\sum_a Z_a) \int d^3r \ \ \rho_{uvR}(r) \\ & \frac{4\pi}{V_{cell}} \sum_{G \neq 0} \frac{e^{-G^2/4\alpha}}{G^2} (\sum_a Z_a e^{-iG\xi_a}) \int d^3r \ \ e^{iGr} \rho_{uvR}(r) \end{aligned} \tag{54}$$

where $g_\infty(r'-\xi_a-R)$ is a delta function density centered at ξ_a+R and the Coulomb interaction between the two densities, $g_a(r')$ and $\rho_b(r)$, is given by

$$(g_{a(r')} || \rho_{b(r)}) = \int d^3r \int d^3r' \ \ \frac{g_a(r')\rho_b(r)}{|r - r'|}. \tag{55}$$

Because the density $\rho_{uvR}(r)$ does not contain a delta function-like singularity, the sum over reciprocal vectors in equation (54) is always convergent, independent of the value of the screening exponent α. In practice there is no need to introduce the screening densities if $\rho_{uvR}(r)$ is slowly varying on the length of the unit cell. In that case equation (54) reduces to

$$V_{uvR} = \frac{4\pi}{V_{cell}} \sum_{G \neq 0} \frac{\sum_a Z_a e^{-iG\xi_a}}{G^2} \int d^3r \; e^{iGr} \rho_{uvR}(r). \tag{56}$$

The criteria used in choosing between equation (54) and equation (56) is the exponent $\gamma_{uv} = \alpha_u + \alpha_v$ of the pair density ρ_{uvR} where α_u and α_v are the Gaussian exponents of the two atomic orbitals. (For simplicity of discussion, we will ignore the fact that the atomic orbitals are contracted basis functions involving many primitive Gaussian functions.) So for the long range (LR) densities when $\gamma_{uv} < \alpha$ equation (56) will be used, whereas for short range densities (SR) when $\gamma_{uv} > \alpha$ equation (54) will be used instead.

C. Alternate Expression for V_{uvR}

In practice we found it better to rewrite equations (54) by using an alternative interpretation for the matrix elements V_{uvR}. The matrix elements can be interpreted as the energy of the nuclear charges Z_a sitting in the potential $V[\rho_{uvR}](r)$ created by the periodic density $\sum_{R'} \rho_{uvR}(r - R')$:

$$V_{uvR} = \sum_a Z_a V[\rho_{uvR}](\xi_a) \tag{57}$$

where

$$V[\rho_{uvR}](r) = \int d^3r' \; \sum_{R'} \frac{\rho_{uvR}(r' - R')}{|r - r'|}. \tag{58}$$

When the density $\rho_{uvR}(r)$ is slowly varying then equation (56) is obtained again. When the density $\rho_{uvR}(r)$ is rapidly varying then we need to introduce some screening. But instead of screening the nuclear point charges, we now need to screen the continuous electron density $\rho_{uvR}(r)$. Because the density $\rho_{uvR}(r)$ can be expressed in terms of a single Gaussian function

$$\rho_{uvR}(r) = \sum_{ijk} C_{ijk} \frac{\partial^i}{\partial P_x^i} \frac{\partial^j}{\partial P_y^j} \frac{\partial^k}{\partial P_z^k} e^{-\gamma_{uv}(r-P)^2} \tag{59}$$

where the vector P denotes the product center of two basis functions (see Appendix B), then there exists an efficient screening density given by:[13]

$$\rho^{\Omega}_{uvR}(r) = \sum_{ijk} C_{ijk} \frac{\partial^i}{\partial P_x^i} \frac{\partial^j}{\partial P_y^j} \frac{\partial^k}{\partial P_z^k} e^{-\Omega(r-P)^2} (\frac{\Omega}{\gamma_{uv}})^{\frac{3}{2}}, \tag{60}$$

where Ω is the screening exponent. Both densities have the same integrated value

$$C_{000} = \int d^3r \;\; \rho_{uvR}(r) = \int d^3r \;\; \rho^{\Omega}_{uvR}(r). \tag{61}$$

Also the combined density, $\rho_{uvR}(r) - \rho^{\Omega}_{uvR}(r)$, generates a non-zero field only inside the sphere where the combined density is non-zero. Outside the sphere the field generated is exactly zero. This can be seen by evaluating the field of the combined density written in the following form

$$\begin{aligned} V[\rho - \rho^{\Omega}](r) &= \sum_{ijk} C_{ijk} \frac{\partial^i}{\partial P_x^i} \frac{\partial^j}{\partial P_y^j} \frac{\partial^k}{\partial P_z^k} \\ &\times \frac{2\pi}{|r-P|\,(\gamma_{uv})^{\frac{3}{2}}} \int_{\Omega|r-P|}^{\gamma_{uv}|r-P|} du \;\; e^{-u^2}, \end{aligned} \tag{62}$$

where we used the fact that $\Omega < \gamma_{uv}$. In the limit where the density is small, i.e., $e^{-\Omega(r-P)^2} << 1$, then the main contribution left is of order $e^{-\Omega(r-P)^2}$.

Having defined the screening density ρ^{Ω}_{uvR}, we can add and subtract its contribution from $V[\rho uvR](r)$ to give the alternate expression for $V_{uvR}(r)$:

$$\begin{aligned} V_{uvR} &= \sum_a Z_a \left(\sum_{R'} (A_{uvR}(\xi_a + R') - A^{\Omega}_{uvR}(\xi_a + R')) + (\frac{1}{\gamma_{uv}} - \frac{1}{\Omega})\frac{\pi}{v} \right) \\ &+ \sum_a Z_a \left(\frac{4\pi}{v} \sum_{G \neq 0} \frac{e^{-iG\xi_a}}{G^2} \int d^3r \;\; e^{iGr} \rho^{\Omega}_{uvR}(r) \right) \end{aligned} \tag{63}$$

where

$$A_{uvR}(r) = \int d^3r' \frac{\rho_{uvR}}{|r - r'|} \tag{64}$$

and

$$A^{\Omega}_{uvR}(r) = \int d^3r' \frac{\rho^{\Omega}_{uvR}}{|r - r'|}. \tag{65}$$

Effective recursion formulas have been derived[7] to evaluate the potential $A_{uvR}(r)$ of the density $\rho_{uvR}(r)$. For the potential $A^{\Omega}_{uvR}(r)$ generated by the screening density, the recursion formulas need to be modified as shown in Appendix B. Also in Appendix B, the recursive formulas for the Fourier transform of the density and the screening density are obtained by decomposing the integrals into one-dimensional integrals.

Comparing equation (54) with equation (63), it is obvious that the two equations should be identical. Clearly by definition we have that

$$(g_{\infty(r-C)}||\rho_{uvR(r')}) = A_{uvR}(C), \tag{66}$$

and by inspection we must also have that

$$(g_{\alpha(r-C)}||\rho_{uvR(r')}) = A^{\Omega}_{uvR}(C), \tag{67}$$

$$\int d^3r \;\; e^{iGr} e^{-G^2/4\alpha} \rho_{uvR}(r) = \int d^3r \;\; e^{iGr} \rho^{\Omega}_{uvR}(r) \tag{68}$$

and

$$\frac{1}{\Omega} = \frac{1}{\gamma_{uv}} + \frac{1}{\alpha}. \tag{69}$$

Equation (68) is easily derived from the knowledge that the only dependence on Ω of $\int d^3r \, e^{iGr} \rho^{\Omega}_{uvR}(r)$ is of the form $e^{-G^2/4\Omega}$. Equation (67) effectively says that the potential $A^{\Omega}_{uvR}(r)$, created by the screening density $\rho^{\Omega}_{uvR}(r)$ is equal to the Coulomb repulsion between the original density $\rho_{uvR}(r)$ and a normalized screening Gaussian function $g_\alpha(r)$, whose exponent is given by equation (69). Note that those two identities, which are proven in Appendix B, hold only because the densities are expanded in term of Gaussian functions.

In actual calculations we find it more desirable to fix Ω as the screening exponent instead of α. This makes equation (63) more desirable than equation (54). As shown previously, the combined density $\rho_{uvR}(r) - \rho^{\Omega}_{uvR}(r)$ has the property that the generated field falls off as $e^{-\Omega(r-P)^2}$ from the center P of the density. This makes possible the implementation of an efficient cutoff scheme in constructing V_{uvR} as in equation (63).

The Coulomb field $J(r)$ created by the total electron density $\rho(r)$ can be obtained from:

$$J(r) = \sum_{uvR}^{SR} V^{SR}_{uvR} D_{uvR} + \sum_{uvR}^{LR} V^{LR}_{uvR} D_{uvR} \tag{70}$$

where V^{SR}_{uvR} is equation (63) used for the SR contributions of the density and V^{LR}_{uvR} is equation (56) used for the LR contributions of the density. Replacing V^{SR}_{uvR} and V^{LR}_{uvR} by their respective formulas, the final expression for the Coulomb field is:

$$\begin{aligned} J(r) = & \sum_{uvR}^{SR} \sum_{R'} (A_{uvR}(r-R') - A^{\Omega}_{uvR}(r-R')) D_{uvR} \\ & + \sum_{uvR}^{SR} S_{uvR} D_{uvR} \left(\frac{1}{\gamma_{uv}} - \frac{1}{\Omega}\right) \frac{\pi}{V_{cell}} \\ & + \frac{4\pi}{V_{cell}} \sum_{G \neq 0} e^{-iGr} \frac{\rho^{SR}(G) + \rho^{LR}(G)}{G^2}, \end{aligned} \tag{71}$$

where the FT of the SR component of the density is:

$$\rho^{SR}(G) = \sum_{uvR}^{SR} D_{uvR} \int d^3r' \; e^{iGr'} \rho^{\Omega}_{uvR}(r') \tag{72}$$

and the FT of the LR component of the density is:

$$\rho_e^{LR}(G) = \sum_{uvR}^{LR} D_{uvR} \int d^3r' \; e^{iGr'} \rho_{uvR}(r'). \tag{73}$$

IV. Grid Generation

The basic principle used in designing grids for numerical integration has been the divide and conquer strategy. The particular strategy we used was first introduced by Becke. The idea is to replace three-dimensional analytical integrations over the whole space by a set of single center numerical integrations. The single center integrations are performed over atom centered grids using spherical coordinates. For example, the Coulomb matrix element involving basis functions, $X_u(r)$ and $X_{vR}(r)$,

$$J_{uvR} = \int d^3r \;\; X_u(r) J(r) X_{vR}(r), \tag{74}$$

is replaced by a numerical integration:

$$J_{uvR} = \sum_{R'ag} W_a(r_{ag} - R') X_u(r_{ag} - R') J(r_{ag} - R') X_{vR}(r_{ag} - R'), \tag{75}$$

where R' labels the unit cells, a the atoms in the unit cell and g the gridpoints associated with that atom. The gridpoints r_{ag} are given by:

$$r_{ag} = \xi_a + r_g, \tag{76}$$

where ξ_a is the atom position and r_g the gridpoint location relative to the atom position. The weights W_a can be separated into two components:

(1) the projection weight, $W^a_{proj}(r_{ag} - R')$, which depends on all the other atoms and which controls how the multicenter integration is reduced into atomic integration by projection,

(2) the atomic weight $W^a_{atom}(r_g)$, which only depends on the location of the gridpoint with respect to the atom center.

A. Atomic Projection Function

In order to ensure that the volume is conserved, the atomic projection functions w^a_{proj} are normalized at each of the gridpoints:

$$\sum_{aR} w^{aR}_{proj}(r) = 1. \tag{77}$$

This only ensures volume conservation in the limit of infinite number of gridpoints, but it remains a very close approximation for finite grids. Thus the Becke weight of point r with respect to atom a is given by:

$$w_a(r) = \frac{P_a(r)}{\sum_{aR} P_a(r-R)}, \tag{78}$$

where the P_a's are the unnormalized atomic projection function. We want the projection function $P_a(r)$ to have the following properties:

(1) near unit value close to the atom a,

(2) near zero value close to every other atom.

In Becke's approach this is done by choosing P_a as a product of pairwise functions:[8]

$$P_a(r) = \prod_{b \neq a}^{atoms} s_3(\mu_{ab}(r)). \tag{79}$$

The pair projection functions are given by:

$$s_3(\mu) = \frac{1}{2}(1 - p_3(\mu)) \tag{80}$$

where

$$p_3(\mu) = p(p(p(\mu))) \ , \quad p(\mu) = \frac{3}{2}\mu - \frac{1}{2}\mu^3 \tag{81}$$

and the μ_{ab} are the hyperbolic coordinates:

$$\mu_{ab}(r) = \frac{|r_a| - |r_b|}{|\xi_a - \xi_b|}, \tag{82}$$

where $r_a = r - \xi_a$ and $r_b = r - \xi_b$. This approach is motivated as follows: the μ_{ab} function goes linearly from -1 on atom a to $+1$ on atom b along the $a - b$ axis. With the use of the $p(\mu)$ function this is transformed into a function which has the same value at atoms a and b but with zero derivatives. The function $p(\mu)$ is the simplest polynomial function with those properties but it varies too slowly. The recursive application of function p is used to obtain the desired sharpness in the intermediate region (i.e., $\mu \sim 0$). Finally equation (106) is used to obtain a function varying between $[0, 1]$ in going from atom b to a.

We have improved on the above scheme in two ways:

(1) As defined, equation (82) doesn't decouple atoms a and b from each other as the distance $|\xi_a - \xi_b|$ becomes large. Physically we expect that the two atomic grids should decouple totally when they are far away from each other. Empirically we have found that limiting the denominator of equation (82) to a maximum value of 5.6 Bohr has the desired result. Thus we used the following expressions

$$\begin{aligned} \mu_{ab}(r) &= \frac{|r_a| - |r_b|}{Rcut} \quad if \quad \big||r_a| - |r_b|\big| < Rcut, \\ \mu_{ab}(r) &= -1 \quad if \quad |r_a| - |r_b| < -Rcut, \\ \mu_{ab}(r) &= 1 \quad if \quad |r_a| - |r_b| > Rcut, \end{aligned} \tag{83}$$

whenever $|\xi_a - \xi_b| > Rcut = 5.6$ Bohr.

(2) In equation (81) the recursive function can only be applied an integer number of times. Becke has shown that only the function $p_3(\mu)$ has the right sharpness.

We have found it desirable to add to $s_3(\mu)$ a function that is non-zero only near $\mu = 0$ but that also contains an adjustable parameter. So we use the following expression for $s_3(\mu)$ instead of equation (80):

$$s_3(\mu) = \frac{1}{2}(1 - p_3(\mu)) + \frac{\beta}{2} sin(p(\mu)\pi), \tag{84}$$

where reasonable values for β are in the range 0.03 – 0.07.

By adjusting *Rcut* and β for a series of molecules, we have improved the accuracy by more than an order of magnitude with our choice of atomic grids.

B. Atomic Spherical Grid

The atomic grids used for the single center integration have spherical symmetry. The grid is formed as the product of a radial grid times an angular grid. The angular distribution of gridpoints is based on quadratures developed by Lebedev.[11] The Lebedev grids have O_h symmetry and are very efficient with respect to integrating spherical harmonics on a sphere. The radial gridpoints are obtained from an exact quadrature for Gaussian function in one dimension. By choosing the radial positions, r_i, and the radial weights, w_i, as follows:

$$r_{i+1} = \gamma r_i \quad , \qquad w_i = r_i^3 log(\gamma), \tag{85}$$

then the following numerical integration:

$$I_{\alpha,l} = \sum_i w_i r_i^{l+1} e^{-\alpha r_i^2}, \tag{86}$$

can be made very accurate. Accuracy better than one in part in 10^{-8} is obtained for a wide range of α and l (α between 0.08 and 1000.0, and l from 0 to 5) by choosing $r_1 = 0.00001$, $\gamma = 1.18$ and using 72 radial points. The final atomic grid have on the order of 3000 to 4000 gridpoints.

In order to have an efficient cutoff scheme for skipping negligible contributions the grid must be cut into small blocks of about 200 gridpoints each. This is done by cutting the atomic grids into 6 wedges (one along each six directions). Each wedge is then cut into a group of shells with thickness less than some maximum value (4.0 Bohr). This scheme is well adapted to incorporate the effect of symmetry (see Appendix C).

V. Self-Consistent Procedure

A. Initial Kohn-Sham Orbitals

The self-consistent procedure consists of the following two steps:

(1): form a new effective potential from the current density and generate the associated Hamiltonian,

(2): obtain the new Kohn-Sham orbitals from diagonalization of the Hamiltonian and construct a new density.

The cycle can be started by having a trial guess for either the potential or the density. Typically the sum of the atomic densities is used to construct the initial density. We have used a different approach where the initial Kohn-Sham orbitals are chosen as the best bonding orbitals. This was found to be a better starting point for the SCF procedure. The initial Kohn-Sham orbitals are constructed as follows:

(1): On each atom in the first unit cell a set of atomic orbitals, $\{\theta_u\}$, is created from an atomic SCF calculation using the occupation-averaged Hamiltonian.

The $\{\theta_u\}$ are partitioned into core and valence orbitals $\{\theta^{core}\}$, $\{\theta^{val}\}$.

(2): The Bloch coefficients for the core orbitals, $C^{core}_{un}(k)$, are obtained directly from the coefficients, θ_{un}, of the core atomic SCF orbitals

$$C^{guess}_{un}(k) = \theta^{core}_{un}. \tag{87}$$

The necessary normalization is done in step (4) below together with the valence Bloch orbitals.

(3): In general the dimension of the valence space formed by the atomic SCF orbitals is larger than the valence space for the actual system because of fractional occupation (e.g., in diamond we have 8 atomic SCF valence orbitals but only 4 occupied valence orbitals in the band structure). Thus we need to reduce the number of valence orbitals. As a criteria we choose the linear combinations of atomic valence SCF orbitals having the largest overlap. This is good first approximation at forming the best bonding orbitals.

The overlap matrix between the valence atomic SCF orbitals, θ^{val}_u, is evaluated as follows:

$$S_{nm} = \sum_{uv} \theta^*_{un} S_{uv}(k) \theta_{vm}. \tag{88}$$

The matrix is then diagonalized

$$S = U \; \lambda \; U^t. \tag{89}$$

The eigenvectors with the largest eigenvalue, λ_n, are selected and saved

$$C^{guess}_{un}(k) = \sum_m C^{va}_{um}(k) U_{mn}. \tag{90}$$

(4): The guess orbitals, $\{C^{guess}_{un}\}$, are orthonormalized by Schimdt orthogonalization. The orbitals are ordered with the core orbitals first (ordered by atomic orbital energy) and then the valence orbitals (ordered by eigenvalue from diagonalizing the valence AO overlap matrix) and then Schmidt orthogonalized.

B. Preconditioned Conjugate-Gradient (PCG)

The preconditioned conjugate-gradient (PCG) method is used to directly minimize the total energy. Much progress has been made in the application of the CG methods to self-consistent field calculations using plane wave basis sets.[3] The issues encountered were: (1) orbitals normalization constraints, (2) the expense of line minimization, (3) slow convergence due to the large number of degrees of freedom and the large bandwidth of the energy spectra. For GTO basis sets, the same issues are present but some solutions are different.

Recall the standard CG algorithm to minimize a function $f(x)$ in the space spanned by the multidimensional vector x:

1) start with initial position vector $x^{(1)}$,

2) evaluate the gradient vector, $g^{(m)} = -\nabla f(x^{(m)})$ for the current iteration (m),

3) evaluate the CG vector, $h^{(m)} = g^{(m)} + \gamma^{(m-1)} h^{(m-1)}$, where

$$\gamma^{(m-1)} = \frac{g^{(m)} \times g^{(m)}}{g^{(m-1)} \times g^{(m-1)}},$$

with the initial condition that $\gamma^{(0)} = 0$,

4) do a line minimization along the CG direction $h^{(m)}$ and obtain the new position vector $x^{(m+1)}$.

Steps (2) through (4) are repeated until the magnitude of the gradient is small enough.

In ab initio calculations,[3] the function to minimize is the total energy E_{total} and the multidimensional position vectors are the orbitals $\phi_{ik}(r)$. For each orbital

$|\phi_{ik}>$, a gradient vector is formed

$$|g_{ik}>= -(H - \epsilon_{ik})|\phi_{ik}> \tag{91}$$

where H is the Kohn-Sham Hamiltonian and where the presence of the orbital energy $\epsilon_{ik} =< \phi_{ik}|H|\phi_{ik}>$ enforces the orbital normalization. The CG direction is given by

$$|h_{ik}>= |g_{ik}> +\gamma'_{ik}|h'_{ik}> \tag{92}$$

where

$$\gamma'_{ik} = \frac{< g_{ik}|g_{ik}>}{< g'_{ik}|g'_{ik}>} \tag{93}$$

and the primed quantities, $|h'_{ik}>$, $|g'_{ik}>$, γ'_{ik}, refer to the previous iteration. At this point the CG vector is made orthogonal to all the occupied orbitals to preserve the orthogonality constraint between the different orbitals:

$$|h_{ik}>= |h_{ik}> - \sum_j |\phi_{jk}>< \phi_{jk}|h_{ik}> . \tag{94}$$

The new orbital, $|\phi^{new}_{ik}>$, is obtained by doing a line minimization to find the best mixing angle θ between the old orbital $|\phi_{ik}>$ and the CG vector $|h_{ik}>$:

$$|\phi^{new}_{jk}>= |\phi_{jk}> cos(\theta) + |h_{ik}> sin(\theta). \tag{95}$$

The expensive line minimization can be avoided in ab initio calculations by evaluating the first and second derivative of the energy with respect to the mixing angle:

$$\frac{\partial E}{\partial \theta}\Big|_{\theta=0} = 2Real\{ < h_{ik}|H|\phi_{ik}> \} \tag{96}$$

and

$$\frac{\partial^2 E}{\partial \theta^2}\Big|_{\theta=0} =< h_{ik}|H|h_{ik}> - < \phi_{ik}|H|\phi_{ik}> +\delta\epsilon_{ik} \tag{97}$$

where $\delta\epsilon_{ik}$ is small and related to derivative of the field:

$$\delta\epsilon_{ik} = 2Real\{ < h_{ik}|\left(\frac{\partial J(r)}{\partial\theta} + \frac{\partial Vxc(r)}{\partial\theta}\right)|\phi_{ik} > \}. \tag{98}$$

The angle that minimizes the energy is then given by:

$$\theta_{min} = tan^{-1}\left(- \frac{\partial E}{\partial\theta}\Big|_{\theta=0} \Big/ \frac{1}{2}\frac{\partial^2 E}{\partial\theta^2}\Big|_{\theta=0}\right). \tag{99}$$

The first derivative is obtained for free since the Hamiltonian is available. The same is true for the first two terms of the second derivative. Only the small term $\delta\epsilon_{ik}$ requires extra computation. For calculations with plane waves, the extra cost is small and the second derivative (including the $\delta\epsilon_{ik}$ term) is evaluated analytically.[3] For calculations with GTO's, the cost of building a new field and evaluating $\delta\epsilon_{ik}$ is too expensive. Instead we use the following empirical expression:

$$\delta\epsilon_{ik} = f \sum_{j}^{unocc.} \frac{| < h_{ik}|H|\psi_{jk} > |^2}{< h_{ik}|H|h_{ik} > -\epsilon_{ik}} \tag{100}$$

where the sum is over the unoccupied orbitals $\psi_{jk}(r)$ of the current Kohn-Sham Hamiltonian and the constant f is introduced to account for the approximate nature of $\delta\epsilon_{ik}$.

Beside avoiding the line minimization, we also used a preconditioning technique.[27] Here the idea is to precondition the gradient in such a way that it becomes parallel to the direction obtained by a second-order perturbative solution of the Hamiltonian. For calculations with plane waves this is difficult to achieve and only the diagonal kinetic energy contribution is corrected. For calculations with GTO's, this can be easily obtained by using a Green function like operator:

$$\sum_{j}^{occ.} \frac{|\psi_{jk} >< \psi_{jk}|}{\sqrt{\left(\epsilon_{ik} - < \psi_{jk}|H|\psi_{jk} > \right)^2 + \omega}} \tag{101}$$

where the sum is over the occupied orbitals $\psi_{jk}(r)$ of the current Kohn-Sham Hamiltonian and where ω describes the energy scale over which the mixing between orbitals occurs. One would expect the scale ω to be of the order of the gap ϵ_{gap} at the beginning of the self-consistent loop and it should become smaller close to convergence. Empirically, we have found the following expression for ω to be satisfactory:

$$\omega = \epsilon_{gap}\sqrt{| < g_{ik}|h_{ik} > |^2}, \tag{102}$$

which has the property that close to convergence

$$\omega = \epsilon_{gap}(\epsilon_{ik} - < \psi_{jk}|H|\psi_{jk} >)^2. \tag{103}$$

VI. Results and Discussion

A. Grid Optimization

Molecular Hartree-Fock calculations were performed to optimize the grid parameters. The grid parameters were only optimized for first row and hydrogen atoms. In these calculations the exchange and Coulomb operators were evaluated over the grid and integrated numerically. The resulting exchange and Coulomb matrices over atomic orbitals can then be compared to the exact matrices obtained with standard HF theory. The parameters controlling the grid generation are:

(1) γ which controls the spacing of the radial atomic shells (equation 85),

(2) $N_{ang}(r_i)$ which controls the number of angular points for each radial shell,

(3) *Rcut* which controls the point of separation between two atoms in a pair projection function (equation 83),

(4) β which fine tunes the shape of the pair projection function (equation 84).

The parameter γ was first optimized for a single atom. With γ chosen to be 1.20, all Gaussian functions with exponents in range 0.08 to 10000.0 were integrated to better than 10^{-7}. The first radial shell was chosen to be equal to 10.0 Bohr. All the other shells were then obtained by dividing the previous radial shell position by γ and the series was terminated at 0.0001 Bohr.

For small molecules up to benzene (see Table 1) the angular parameters $N_{ang}(r_i)$ were adjusted. This was performed using past experience in constructing grid for the pseudospectral method.[28] Only a few angular points were required near the core. The number of angular points was then increased up to a maximum value of 194 in the bonding region, where large nonsymmetric fluctuations can occur. For larger distances the number of angular points is reduced again to smaller values. The radial space was separated into five regions with boundaries at 0.0, 0.18, 0.4, 0.6, 3.3 and 12.0 Bohr and the optimized number of angular points in each region was found to be 12, 26, 50, 194 and 116.

The pair separation parameter, *Rcut* was then optimized for larger molecules (porphine and C_{60}). Up to this point the Coulomb field was not screened, i.e., the nuclear attraction potential was evaluated analytically. The largest errors occurred for Coulomb matrix elements involving long range basis functions. Adding the nuclear attraction field to the Coulomb field effectively made the resulting field short range (for the covalent molecules used here) and the error was reduced by a factor proportional to the number of atoms in the molecule. In order to integrate the diverging part of the nuclear attraction potential to sufficient accuracy the value of the smallest radial shell was reduced to 0.00001 Bohr.

At this point the total energy error for C_{60} was about 0.6 eV per atom. The

β parameter was then introduced and was adjusted at the same time as *Rcut* and γ. The final total energy error was brought down to about 0.1 eV per atom with $\gamma = 1.18$, $rcut = 5.6$ Bohr and $\beta = 0.07$ for first row atoms and 0.03 for hydrogen atoms. The final atomic grid has 3740 gridpoints per atom. The effect of the parameter β on the pair projection functions can be seen in Figure 1 where pair projection functions, $s_3(\mu)$, $s_2(\mu)$ and $s_1(\mu)$ (defined similarly to equation(85)), are plotted along the line in between the pair of atoms *ab*. The effect of $\beta = 0.07$ is seen on the line labeled $s_3 + s_{add}$ which is markedly tighter than the original s_3 functions.

As can be seen in Table 1 the final accuracy of the total energy is very good, and we expect the same or better accuracy for calculations with the local density approximation since the exchange-correlation field is shorter range than the HF exchange operator. Beside the total energy, we are also interested in the energy of the HF orbitals. For glycine and benzene the orbital energy error was less than 0.3 meV for all states within 30.0 eV of the vacuum level; for the C_{60} molecule, the error was less than 3.5 meV.

B. Calculations on Diamond Crystal

The diamond allotrope of carbon has a close packed fcc structure with two atoms per unit cell. In diamond each carbon atom is bonded tetragonally to its four nearest neighbors. This leads to a covalently bonded insulator with an indirect band gap of 5.47 eV.[29]

The theoretical electronic properties of diamond have been extensively studied. We compare our results to three studies, each based on a different methodology and believed to be well converged in terms of basis set:

(1) LCAO, Linear Combination of Atomic Orbitals,

(2) LAPW, Linear Augmented Plane Waves and

(3) PS-PW, Pseudopotential Plane Waves.

These three studies have used the same exchange-correlation potential (Hedin-Lundqvist[30]), the same experimental lattice constant of 3.567 Å, and the same number of special k points[31] for the integration over the Brillouin zone.

The LCAO calculation[32] uses a real space method with Gaussian type orbitals. The Coulomb field is evaluated on a physical space grid by using both a spline cutoff algorithm for the core part of the density and a fast Fourier transform algorithm followed by linear interpolation for the non-core part of the density. The total field (Coulomb plus exchange-correlation plus nuclear attraction) is then fitted to S Gaussian functions of the form $r^n e^{-\alpha r^2}$ where $n = 0, -1$. Most fitting functions are located on atoms, with a few located on bond centers. The matrix elements of the Hamiltonian are then evaluated analytically using the fitting functions. Their basis set contained 27 Gaussian functions per atom (6s5p1d), but they did not report the exponents of the Gaussian functions and the coefficients used for contracting them.

The LAPW calculation[33] used 175 PW's per unit cell, an atomic radius of 1.42 Bohr and an expansion of up to $l = 8$ inside the sphere. The PS-PW calculation[34] used 300 PW's using a norm-conserving pseudopotential with angular component up to $l = 2$.[35]

Table 3 contains the energy levels for the valence and conduction bands at the Γ (000), X (100) and L $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ points and the band structure along high symmetry directions is shown in Figure 2. In both cases, the zero of energy is taken as the top of the valence band, $\Gamma_{25'}$. The first two columns of Table 3 give the results of our calculations with both the Dunning/Huzinaga (DH) double zeta basis and the Pople 6-31G basis set. These two basis sets are standard bases used for molecular studies. The core $1s$ orbital is represented by a single contracted basis function expanded in term of 6 or 7 Gaussian functions. The $2s$ orbital is described in term of two basis functions, with the inner one contracted with 2 or 3 GTOs and the other one uncontracted; and similarly for the $2p$ orbitals. The exact exponents and contraction coefficients are given in Table 2. Both basis sets were complemented by a single set of Gaussian d function with exponent of 0.75 and 0.8 respectively.

The agreement between all five calculations in Table 3 is good. The average energy difference between any pair of calculations is about 0.04 to 0.06 eV except for the LAPW calculation which has consistently a larger energy difference, 0.15 to 0.18 eV. The average difference between with DH and the 6-31G basis set calculations is only 0.036 eV.

The largest differences occurs at the X points, $\frac{2\pi}{a}(1,0,0)$, of the conduction band. The lowest point, X_1, is very close to the energy minimum of the conduction band which is along the Γ-X direction. At this point the LCAO and PS-PW disagree by 0.15 eV.

We calculate a band gap of 4.04 eV in good agreement with PS-PW value of 4.05 eV. This is about 30% smaller than the experimental value of 5.47 eV. It is well known that LDA calculations consistently underestimate the band gap by 20%-40%. To improve on this may require ab initio methods.

When the d polarization functions are removed from the DH and the 6-31G basis sets, the total energy increases (less strongly bonded) by 25 meV and 23 meV respectively and the energy levels change on average by 0.18 and 0.16 eV respectively. We should note that we used Cartesian Gaussian functions. This means that the d polarization set contains six functions, five spherical harmonics for $l = 2$ angular momentum and one spherical harmonic for $l = 0$ (corresponding to the $x^2 + y^2 + z^2$ linear combination). Looking at the Mulliken population for the DH basis set, the population for the x^2, y^2, z^2 orbitals is 0.0058 and the population for the xy, xz, yz orbitals is 0.0133. Because of the local T_d symmetry around a carbon atom the x^2, y^2, z^2 orbitals only contribute through the A_1 irreducible representation, i.e., $x^2 + y^2 + z^2$ linear combination, whereas the xy, xz, yz orbitals all contribute in the triply degenerate T_1 irreducible representation. The $x^2+y^2+z^2$ orbital should have a smaller effect on the energy levels than the xy, xz, yz orbitals because most of the 0.0058 population is due to overlap with the already occupied $l = 0$ space whereas the 0.0133 population reflects the addition of new degrees of freedom for $l = 2$. Thus we conclude that the presence of d functions is necessary in the calculations. This is in contrast to the conclusion from the LCAO calculations[32] where the d functions ($l = 2$ only) were found unnecessary.

We also have performed an optimization of the lattice constant (see Figure 3) for the 6-31G basis set. The resulting lattice constant is 0.6% smaller than the experimental value of 3.567 Å. This is similar to other LDA calculations where errors of up to a few percent are reported.[36]

C. Calculations on C_{60} Crystal

The second allotrope of carbon we studied is the recently discovered crystalline form of C_{60}.[37] This has sparked much interest because of the superconducting behavior of the related compounds $A_{3-x}B_xC_{60}$ with A,B= K,C_s,R_b.[38-41] The molecular form of C_{60} also has attracted much interest because of its possible chemical and biological properties.

Due to the high symmetry of C_{60} (I_h symmetry) it is not surprising that the crystalline form of C_{60} is the fcc close packed structure.[42] At high temperature the C_{60} molecules rotate freely around their center and the structure is strictly fcc. Below 249K, the molecules become rotationally ordered[42] and the crystal is then a simple-cubic lattice with four C_{60} molecules at (000), $(0, \frac{1}{2}, \frac{1}{2})$, $(\frac{1}{2}, 0, \frac{1}{2})$ and $(\frac{1}{2}, \frac{1}{2}, 0)$ each with a different orientation.

Ab initio calculations of the simple-cubic lattice structure, which has 240 atoms per unit cell, are too expensive and all ab initio studies have been performed using an hypothetical fcc structure having a single C_{60} molecule per cell[43] which are oriented in such a way that the resulting space group is of cubic symmetry. The space group is $T_h^3(Fm3)$ and the point group symmetry around a C_{60} center is reduced from I_h symmetry (with 120 operations) to T_h symmetry (with 24 operations). Because of this reduction in symmetry, the C_{60} molecule is free to distort within the T_h symmetry restriction. In practice the distortion is very small[44] and can be neglected. Thus both the molecular and crystalline structure of C_{60} can be described by two parameters: R_1 is the short carbon-carbon bond distance, $R_1 \sim 1.41$ Å , and R_2 is the long C-C bond distance, $R_2 \sim 1.45$ Å.

Loosely speaking the shorter bond corresponds to a double C-C bond (σ and π bonds) and the longer bond to a single C-C bond (σ bond). Because of the curvature of C_{60}, neither the pure sp^3 hybridization of diamond or the pure sp^2 hybridization of graphite are possible. Hence there are no pure σ or π bonds in this allotrope of carbon. The structure of C_{60} can be described as having 12 pentagons of singly bonded carbon atoms where each pentagon is connected to its five nearest neighbors with a double C-C bond.

For crystalline C_{60}, the experimental values for R_1 and R_2 are 1.40 Å and 1.45 Å from NMR data[45] and 1.391 Å and 1.455 Å from x-ray data.[46] Theoretical values are 1.38 to 1.39 Å and 1.44 to 1.45 Å (with the hypothetical fcc (Fm3) structure with one C_{60} per unit cell) using plane waves LDA.[44,47]

For our calculations on C_{60} we have used the values of $R_1 = 1.414$ Å and $R_2 = 1.455$ Å from molecular simulations using a graphite force field.[48] The calculations were performed using both the Dunning/Huzinaga and 6-31G basis sets which have 9 functions per atoms (see Table 2) for a total of 540 basis functions per unit cell. For the 6-31G basis set, the integration over the Brillouin zone was done with one or two special k points.[31] The resulting orbital energy differences were less than 0.01 eV for energy levels within 20 eV of the Fermi energy. Thus calculations with two k points are necessary only for evaluating sensitive properties like the density of states or structural properties.

The band structure of C_{60} using the 6-31G basis set and two k points is given in Figure 4 where the zero of energy is the top of the valence band. The band structure reveals a direct band gap insulator with a energy gap of $\epsilon_{gapX} = 0.97$ eV at the X point, $\frac{2\pi}{a}(1,0,0)$. This about 50% smaller than the gap of $\sim 1.85eV$ found by microwave conductivity experiments[49] and much smaller than the values of 2.3 to 2.6 eV inferred from direct and inverse photoemission experiments.[50-52] In Figure 4, the largest band dispersion is always at the X point for the three bands shown. The band width of the valence band, labeled H_u from its I_h molecular symmetry, is $W_{Hu} = 0.82$ eV and the band width of the two conduction bands are $W_{T1u} = 0.61$ eV and $W_{T1g} = 0.73$ eV.

Also of interest is the energy gap at the Γ point, $\frac{2\pi}{a}(0,0,0)$, which has a value of $\epsilon_{gap\Gamma} = 1.69$ eV. At the Γ point the top three levels of the valence band are degenerate forming a T_u irreducible representation (from the local T_h symmetry) with an angular character of $l = 5$ (with respect to the center of C_{60}).[44] The first conduction band also has the same irreducible representation and angular character and one can expect those two bands to be affected in the same way by the crystal field. Indeed comparing molecular C_{60} with crystalline C_{60}, the gap changes by 0.01 eV which is less than 1%. In contrast, the splitting T_u-T_g at the Γ point between the two conduction bands decreases by 0.33 eV in going from the molecular form to the crystalline form. The same observation also applies to K_3C_{60} where the band gap $\epsilon_{gap\Gamma}$ changes by 0.02 eV and the T_u-T_g splitting decreases by 0.29 eV when a pressure of 3 G Pa is applied.[53]

The C_{60} calculations were repeated with one k point for both the 6-31G and the Dunning/Huzinaga basis sets. The resulting band gaps and band widths were very similar to the 6-31G calculation with two k points. The changes were less than 0.01 eV and 0.02 eV respectively. Thus the two basis sets give almost identical results, just as observed for diamond.

The band parameters from the C_{60} calculations are summarized in Table 4, along with theoretical results obtained by four other groups [all using the hypothetical fcc (Fm3) structure]. The two plane wave calculations, PS-PW(1)[54] and PS-PW(2)[44], use similar numbers of plane waves, 27000 and 28000, and both use norm conserving pseudopotentials with s and p components. Note that the size of the plane wave basis set is about 45 times larger than the GTO basis sets we use. The difference between the two pseudopotentials is in the size of the core radii over which the norm conversation is applied. For the PS-PW(1) calculation, the core radii, $r_s \sim r_p \sim 0.4\mathring{A}$, are about half the size of the core radii used in the PS-PW(2) calculation, $r_s \sim r_p \sim 0.8\mathring{A}$. Larger core radii lead to a smoother, more well-behaved pseudopotential,[36] but also to a larger approximation. The pseudopotentials were generated for the ground state configuration of carbon, $2s^2 2p^2$, and the resulting pseudopotential only has s and p components. These two calculations agree well with each other although the geometry and the nature of the exchange-correlation used in PS-PW(1) was not stated. The Ceperley-Alder exchange-correlation was used in PS-PW(2).

The PS-LCAO calculation[43] also uses a norm conserving pseudopotential to represent the core $1s$ electrons. Two uncontracted Gaussian functions were used to represent the valence $2s$ and $2p$ orbitals. The exponents are 0.299 and 2.908 for the $2s$ orbital and 0.362 and 2.372 for the $2p$ orbitals. The smaller exponents are substantially larger than the exponents in the 6-31G basis set (0.16 for s and p) and the Dunning/Huzinaga basis set (0.15 for s and 0.11 for p). Small exponents are known to be essential for calculations on finite molecules in order to describe the outer part of the valence orbitals. Eliminating these small exponents should lead to an artificially large band gap like (as observed in the PS-LCAO calculation). In contrast, we obtained the same band gap for two basis sets where the smallest p function exponents vary by 40%.

The LCAO calculation[55] is an all electron calculation with the carbon $1s$,$2s$,$2p$ orbitals represented by 16 GTOs with exponents in the range 0.15 to 50000. The GTOs are contracted to 5 or 9 basis functions per carbon atom. The matrix elements of the Kohn-Sham Hamiltonian were evaluated by a fitting procedure where both the density and field are fitted. However, this calculation uses the Wigner exchange-correlation potential rather than the Ceperley-Alder exchange-correlation potential. Just as for the PS-LCAO calculation, their gap $\epsilon_{gap\Gamma}$ is about 0.2 eV larger than our value of 1.6 eV (which is close to the two PS-PW values). On the other hand, our values for the different band widths are consistently larger by 20-30% as compared to the other four calculations.

The band width is directly related to the interaction between nearest neighbor C_{60} molecules. Taking into account the values of R_1, R_2 and the lattice parameter, the smallest distance between two carbon atoms located on different C_{60} clusters is 2.94 $\AA$ for the PS-PW(2) and our calculations, and 3.09 $\AA$ for the LCAO and PS-LCAO calculations. Because of the exponential decay of the orbital the 5% difference in distance should translate into a much larger effect on the band width. Assuming a *pp* hopping integral between nearest neighbor C_{60} the band width can be parametrized as follows:[56]

$$W = W_o(d/d_o)e^{-(d-d_o)/L}$$

where reasonable values of L are in the range 0.54 $\AA$ to 0.63 $\AA$. Thus the 5% increase in distance translates into a decrease of 20-25% in band width. This may explain the difference between our band widths and the LCAO band widths.

We have not included the *d* polarization functions in the C_{60} calculations. Because of the large value of their exponents, 0.75-0.8, the *d* functions should have negligible effect on the band width. However, the *d* functions *can* affect the gap for intramolecular interactions which would result in rigid shift of the bands. This incompleteness of the basis set can be estimated by comparing calculations on molecular C_{60}. Table 5 compares the energy levels obtained with using the 6-31G basis with the energy levels obtained with a well converged (triple zeta) basis set.[57] The comparison is very good. The energies of the highest occupied molecular orbital (HOMO) and of the lowest unoccupied molecular orbitals (LUMO) are both shifted upward by 0.02 eV in the calculation with the 6-31G basis set. A HOMO-LUMO gap value of 1.68 eV is obtained in both calculations. For higher energy unoccupied orbitals, the energy difference between the two calculations is as large as 0.21 eV for the first 4 eV. For the lower occupied orbitals, the energy difference is at most 0.11 eV for the levels within 10 eV of the HOMO orbital energy. The absolute value of the HOMO energy we obtained, 5.92 eV, can be used to estimate the first ionization potential (IP) of C_{60}. The experimental first IP of C_{60} is 7.6 eV.[58–59] The comparison can be greatly improved by including orbital relaxation effects.[57]

Turning our attention back to the crystal calculation, Table 6 compares energy levels at the Γ point with the energy levels for the PS-PW(2) calculations (setting the top of the valence band at 0.0 eV). Over the whole valence band energy spectrum, the worst energy difference is 0.3 eV. Near the Fermi energy the comparison is even better, with the band gaps within 0.02 eV, and the splitting of the molecular H_u HOMO into E_u-T_u within 0.03 eV. For the conduction band the difference are larger, as expected. Not included in Table 6 are five energy levels, at A_g4.2, A_g6.50, A_u6.97, A_g7.21, T_u7.43 and T_g8.30 eV, which are observed in the plane wave calculations. Those levels where associated with nonmolecular states, i.e., states where the density is located at interstitial positions: the center of C_{60} clusters, the interstitial fcc tetrahedral and octahedral positions. For the 6-31G basis set, no such level is identified in the first 10 eV. For the Dunning/Huzinaga, two such levels are observed, A_u at 7.82 eV and A_g at 8.24 eV. Clearly the smaller exponents in the

Dunning/Huzinaga basis are much better at describing such delocalized states.

VII. Conclusions

We have presented a method for doing ab initio electronic structure calculations on large systems using localized Gaussian type orbitals in the local density approximation. This method avoids the density fitting procedure by evaluating analytically the Coulomb field using a dual-space approach that scales as $N^{1.5}$. The dual-space approach is made possible by screening perfectly each component of the density matrix and by the efficient computations of different Gaussian integrals using a recursion relation scheme. Avoiding the density fitting procedure makes the algorithm more compatible with current parallel computer architecture.

The matrix elements of the Kohn-Sham Hamiltonian are evaluated by breaking up the integration in real space over an atom centered grid using a projection technique. Test calculations on molecular systems as large as C_{60} demonstrate the accuracy of the grid generated. The total energy error was less than 0.2 meV per atom using 3700 gridpoints per atom.

Band structure calculations on diamond were shown to be very accurate using Dunning/Huzinaga or 6-31G double zeta basis sets with *d* polarization functions. For both basis sets, the energy levels were converged to within 0.02 eV over a large spectrum of energy. The optimized lattice parameter was found to be very close to the experimental value (0.6% smaller) using only 30 basis functions per cell.

For crystalline C_{60}, we showed that the energy levels near the Fermi level were well described (to within 0.1 eV) using a double zeta basis sets (540 basis functions per cell). The Dunning/Huzinaga and the 6-31G basis sets were found equivalent to within 0.03 eV for all valence bands. For the conduction bands, the differences were an order of magnitude larger. The smaller exponent in the Dunning/Huzinaga basis set lead to an extra band at 7.8 eV which was associated with a nonmolecular state. Adding less than 15 Gaussian functions at the center of the C_{60} cluster and at the fcc octahedral and tetrahedral interstitial positions should make possible the correct description of the nonmolecular states which were better described in the plane wave calculations at the price of a very large basis sets (28000 basis functions per cell).

Appendix A. Exchange-Correlation Potential

We use the exchange-correlation potential based on the Quantum Monte Carlo simulation of the interacting electron gas done by Ceperley and Alder[22] which was parametrized by Perdew and Zunger.[21] The parametrization of Perdew and Zunger gives the correct low-density and high-density limits. The final formula for the exchange correlation energy is:

$$\epsilon_{xc}(\rho) = f(\rho)\,\epsilon_x(\rho) \,+\, \epsilon_c(\rho) \tag{A.1}$$

where $\epsilon_x(\rho)$ is exchange energy, $f(\rho)$ is the Bachelet[23] relativistic correction, and $\epsilon_c(\rho)$ is the correlation energy. The exchange energy $\epsilon_x(\rho)$ is approximated by the average exchange energy in the HF free electron gas

$$\epsilon_x(\rho) = -\frac{0.4582}{r_s} \tag{A.2}$$

where r_s is related to the density by

$$r_s(\rho) = (\frac{3}{4\pi\rho})^{\frac{1}{3}}. \tag{A.3}$$

The relativistic factor is:

$$f(\rho) = 1 - \frac{3}{2}\left(\frac{(1+\beta)^{\frac{1}{2}}}{\beta} - \frac{ln(\beta + (1+\beta^2)^{\frac{1}{2}})}{\beta^2}\right)^2 \tag{A.4}$$

with $\beta = 0.0140/r_s$. Finally, the parametrization of the correlation energy $\epsilon_c(\rho)$ is given by:

$$\epsilon_c(\rho) = \frac{-0.1432}{1 + 1.0529\sqrt{r_s} + 0.3334\,r_s} \qquad r_s \geq 1, \tag{A.5}$$

$$\begin{aligned}\epsilon_c(\rho) &= -0.048 + 0.0311 ln(r_s)\\ &\quad - 0.0116\,r_s + 0.002\,r_s\,ln(r_s) \qquad r_s < 1.\end{aligned}$$

Appendix B. One-Electron Integrals

With Cartesian Gaussian functions, efficient recursion formulas can be used to evaluate the different one-electron integrals: overlap integral S_{uv}, kinetic energy integral T_{uv} , Fourier transform integral $\rho_{uv}(G)$, and two-center potential integral $A_{uv}(r)$. Recursion relations for $S_{uv}, T_{uv}, A_{uv}(r)$ integrals have been derived using three-center overlap integrals.[7] Traditionally, the most expensive integrals are the nuclear attraction integrals for which the evaluation has been shown to be very efficient using recursion relations.

The purpose of this Appendix is to show how to obtain the recursion relations for the Fourier transform integrals, $\rho_{uv}(G)$ and $\rho^{\Omega}_{uv}(G)$, of the density $\rho_{uv}(r)$, and of the screened density $\rho^{\Omega}_{uv}(r)$, and the two-center potential integrals $A^{\Omega}_{uv}(r)$ for the screened density $\rho^{\Omega}_{uv}(r)$. In the dual-space approach, the screening exponent is chosen to balance the cost of the Fourier transform integrals with the cost of two-center potential integrals. Thus we need to have efficient evaluation of the Fourier transform integrals.

For Fourier transform integrals involving cubic lattices, it is very advantageous to separate the x,y,z components of the integrals because for each pair of basis functions, we want to evaluate $\rho_{uv}(G)$ for many G vectors of the form $G = (n_x, n_y, n_z)\,\Delta G$. In this way, the one-dimensional integrals over the x,y,z components can be reused to evaluate all the possible $\rho_{uv}(G)$ integrals. For non-cubic-lattice, the one-dimensional integrals are more numerous but the final product of the three components, the expensive part, doesn't change.

Surprisingly the recursion formulas for the $\rho_{uv}(G)$ and $A_{uv}(r)$ integrals only need to be slightly modified for the integrals involving the screened density ρ^{Ω}_{uv}.

Section B.1 gives the basic definitions for Gaussian functions. Sections B.2 and B.3 respectively gives the recursion relations for the S_{uv} and T_{uv} integrals. Those relations are only given here for the sake of completeness. Sections B.4 and B.5 respectively give the recursion relations for the $\rho_{uv}(G)$, $A_{uv}(r)$ integrals, and their modifications for the screened density.

B.1 Basic Definitions

The normalized Cartesian Gaussian function is given by

$$X_{ul}(r) = N_{ul}\,(x-x_u)^{lx}(y-y_u)^{ly}(z-z_u)^{lz}e^{-\alpha_u(r-\xi_u)^2}, \tag{B.1}$$

where $\xi_u = (x_u, y_u, z_u)$ is the origin of the Gaussian function, $l = (l_x, l_y, l_z)$ defines its angular momentum, and α_u its exponent. The normalization constant, N_{ul}, which makes the integration of $|X_{ul}(r)|^2$ over space equal to one, is given by

$$N_{ul} = \left(\frac{(2\alpha_u/\pi)^{3/2}(4\alpha_u)^{lx+ly+lz}}{(2l_x-1)!!(2l_y-1)!!(2l_z-1)!!}\right)^{\frac{1}{2}}. \tag{B.2}$$

The product $\rho_{ul,vm}(r)$ of a pair of Gaussian functions, $X_{ul}(r)$ and $X_{vm}(r)$, is also a Gaussian function centered on the line joining the two centers ξ_u and ξ_v:

$$\begin{aligned}\rho_{ul,vm}(r) &= N_{ul}N_{vm}e^{-\alpha_u\alpha_v(\xi_u-\xi_v)^2/\gamma_{uv}}(x-x_u)^{lx}(y-y_u)^{ly}(z-z_u)^{lz}\\ &\times(x-x_v)^{mx}(y-y_v)^{my}(z-z_v)^{mz}e^{-\gamma_{uv}(r-P)^2},\end{aligned} \tag{B.3}$$

where $\gamma_{uv} = \alpha_u + \alpha_v$ is the sum of the two basis function exponents and P is the product center which is given by

$$P = \frac{\alpha_u\xi_u + \alpha_v\xi_v}{\gamma_{uv}}. \tag{B.4}$$

The density product can also be written in terms of derivatives as in equation (59):

$$\rho_{ul,vm}(r) = N_{ul}N_{vm}\sum_{ijk} c_i^x c_j^y c_k^z \frac{\partial^i}{\partial P_x^i}\frac{\partial^j}{\partial P_y^j}\frac{\partial^k}{\partial P_z^k} e^{-\gamma_{uv}(r-P)^2} \qquad (B.5)$$

where the explicit form of the coefficients c_i^x,c_j^y,c_k^z will never be needed. The associated screening density is given by:

$$\rho_{ul,vm}^{\Omega}(r) = N_{ul}N_{vm}\sum_{ijk} c_i^x c_j^y c_k^z \frac{\partial^i}{\partial P_x^i}\frac{\partial^j}{\partial P_y^j}\frac{\partial^k}{\partial P_z^k} i\, e^{-\Omega(r-P)^2}(\frac{\Omega}{\gamma_{uv}})^{\frac{3}{2}}, \qquad (B.6)$$

where Ω is the screening exponent. Equations (B.5) and (B.6) will only be used to derive the recursion relations for the integrals involving the screening density.

B.2 Overlap Integrals

The overlap integral between a pair of normalized Gaussian functions is given by:

$$S_{ul,vm} = \int d^3r\ X_{ul}(r)X_{vm}(r). \qquad (B.7)$$

The integral can be separated into its x,y,z components:

$$S_{ul,vm} = N_{ul}N_{vm}[l_x,m_x][l_y,m_y][l_z,m_z], \qquad (B.8)$$

where the one-dimensional integral over the x component is given by

$$[l_x,m_x] = \int_{-\infty}^{+\infty} dx\ (x-x_u)^{lx}e^{-\alpha_u(x-x_u)^2}(x-x_v)^{mx}e^{-\alpha_v(x-x_v)^2}, \qquad (B.9)$$

and similarly for the y and z components. Two special identities can be derived from the form of equation (B.9):

1) The first identity comes from the invariance of the overlap under translation of both Gaussian function

$$2\alpha_u[l_x+1,m_x] + 2\alpha_v[l_x,m_x+1] = l_x[l_x-1,m_x] + m_x[l_x,m_x-1]. \quad (B.10)$$

To prove the above equation start with the following identity

$$0 = \int_{-\infty}^{+\infty} dx \ \frac{\partial}{\partial_x}\left((x - x_u)^{lx} e^{-\alpha_u (x-x_u)^2}(x - x_v)^{mx} e^{-\alpha_v (x-x_v)^2}\right), \qquad (B.11)$$

which depends on the Gaussian functions being 0 at $\pm\infty$. Applying the partial derivative $\frac{\partial}{\partial_x}$ to each term in (B.11) leads directly to (B.10).

2) The second identity,

$$(x_v - x_u)[l_x, m_x] = [l_x + 1, m_x] - [l_x, m_x + 1], \qquad (B.12)$$

is related to the fact that the integrand in equation (B.9) is only a function of $(x - x_u)$ and $(x - x_v)$ which is easily obtained by placing $(x_v - x_u)$ under the integration and by writing it as $(x - x_u) - (x - x_v)$.

When equations (B.10) and (B.12) are combined to zero out the $[l_x, m_x + 1]$ contribution, the recursion relation that describes how to increment the value of l_x is obtained:

$$[l_x + 1, m_x] = (P_x - x_u)[l_x, m_x] + \frac{l_x}{2\gamma_{uv}}[l_x - 1, m_x] + \frac{m_x}{2\gamma_{uv}}[l_x, m_x - 1]. \qquad (B.13)$$

Equations (B.10) and (B.12) can also be combined to zero out the $[l_x + 1, m_x]$ contribution. This results in the recursion relation that describes how to increment the value of m_x:

$$[l_x, m_x + 1] = (P_x - x_v)[l_x, m_x] + \frac{l_x}{2\gamma_{uv}}[l_x - 1, m_x] + \frac{m_x}{2\gamma_{uv}}[l_x, m_x - 1]. \qquad (B.14)$$

In practice the above recursion relations are used as follows:

1) Evaluate $[0, 0] = (\pi/\gamma_{uv})^{1/2} Exp(-\alpha_u \alpha_v (x_u - x_v)^2/\gamma_{uv})$

2) Evaluate all $[l_x, 0]$ using equation (B.13) (where the last term can be skipped since it is zero).

3) Evaluate all $[l_x, m_x]$ using equation (B.14):

   Loop over all l_x.

   Loop over all $m_x > 0$.

   Use equation (B.14).

4) Repeat steps 1-3 for the y,z components.

5) Get $S_{ul,vm}$ from equation (B.8).

B.3 Kinetic Energy Integrals

The kinetic energy matrix integral between two normalized Gaussian functions is given by

$$T_{ul,vm} = -\frac{1}{2}\int d^3r\ \phi_{ul}\nabla^2\phi_{vm}, \tag{B.15}$$

in atomic units. After integration by part and separation of the x,y,z components, equation (B.15) can be rewritten as

$$T_{ul,vm} = \frac{N_{ul}N_{vm}}{2}(J_xI_yI_z + I_xJ_yI_z + I_xI_yJ_z), \tag{B.16}$$

where

$$I_x = [l_x, m_x] = \int_{-\infty}^{+\infty} dx\ (x - x_u)^{lx} e^{-\alpha_u(x-x_u)^2}(x - x_v)^{mx} e^{-\alpha_v(x-x_v)^2}, \tag{B.17}$$

which is the one-dimensional overlap integral obtained previously and

$$\begin{aligned} J_x &= [\![l_x, m_x]\!] \\ &= \int_{-\infty}^{+\infty} dx\ \frac{\partial}{\partial x}\left((x - x_u)^{lx} e^{-\alpha_u(x-x_u)^2}\right)\frac{\partial}{\partial x}\left((x - x_v)^{mx} e^{-\alpha_v(x-x_v)^2}\right), \end{aligned} \tag{B.18}$$

which is the one-dimensional kinetic energy integral. Similar expressions are defined for the y,z components.

Two special identities can be derived from the form of equation (B.18):

1) The first identity comes from the invariance of the kinetic energy integral under translation of both Gaussian function

$$2\alpha_u \llbracket l_x + 1, m_x \rrbracket + 2\alpha_v \llbracket l_x, m_x + 1 \rrbracket = l_x \llbracket l_x - 1, m_x \rrbracket + m_x \llbracket l_x, m_x - 1 \rrbracket, \quad (B.19)$$

which is derived in the same way as in the overlap integral.

2) The second identity is related to the fact that the integrand in equation (B.18) is only a function of $(x - x_u)$ and $(x - x_v)$

$$(x_v - x_u)\llbracket l_x, m_x \rrbracket = \llbracket l_x + 1, m_x \rrbracket - \llbracket l_x, m_x + 1 \rrbracket + \frac{\partial [l_x, m_x]}{\partial x_v} - \frac{\partial [l_x, m_x]}{\partial x_u}, \quad (B.20)$$

where the derivatives are given by:

$$\frac{\partial [l_x, m_x]}{\partial x_u} = 2\alpha_u [l_x + 1, m_x] - l_x [l_x - 1, m_x] \qquad (B.21)$$

and

$$\frac{\partial [l_x, m_x]}{\partial x_v} = 2\alpha_v [l_x, m_x + 1] - m_x [l_x, m_x - 1]. \qquad (B.22)$$

Equation (B.20) is obtained by placing $(x_v - x_u)$ under the integration and by writing it as $(x - x_u) - (x - x_v)$. The first term, $x - x_u$, leads to

$$\int_{-\infty}^{+\infty} dx \; (x - x_u) \frac{\partial}{\partial x}\left((x - x_u)^{lx} e^{-\alpha_u (x - x_u)^2} \right) \times \frac{\partial}{\partial x}\left((x - x_v)^{mx} e^{-\alpha_v (x - x_v)^2} \right), \qquad (B.23)$$

where an integration by part of the first derivative gives directly equation (B.20).

Equations (B.19) and (B.20) can be combined to zero out the $[l_x, m_x + 1]$ and $\llbracket l_x, m_x + 1 \rrbracket$ contributions. This gives the recursion relation that describes how to increment the value of l_x:

$$\begin{aligned} \llbracket l_x + 1, m_x \rrbracket = {} & (P_x - x_u)\llbracket l_x, m_x \rrbracket + \frac{l_x}{2\gamma_{uv}}\llbracket l_x - 1, m_x \rrbracket + \frac{m_x}{2\gamma_{uv}}\llbracket l_x, m_x - 1 \rrbracket \\ & + \frac{4\alpha_u\alpha_v}{\gamma_{uv}}\left([l_x + 1, m_x] - \frac{l_x}{2\alpha_u}[l_x - 1, m_x] \right). \end{aligned} \tag{B.24}$$

Equations (B.19) and (B.20) can also be combined to give the recursion relation that describes how to increment the value of m_x:

$$\begin{aligned} \llbracket l_x, m_x + 1 \rrbracket = {} & (P_x - x_v)\llbracket l_x, m_x \rrbracket + \frac{l_x}{2\gamma_{uv}}\llbracket l_x - 1, m_x \rrbracket + \frac{m_x}{2\gamma_{uv}}\llbracket l_x, m_x - 1 \rrbracket \\ & + \frac{4\alpha_u\alpha_v}{\gamma_{uv}}\left([l_x, m_x + 1] - \frac{m_x}{2\alpha_v}[l_x, m_x - 1] \right). \end{aligned} \tag{B.25}$$

In practice the above recursion relations are used as follows:

1) Evaluate $\llbracket 0, 0 \rrbracket$ using

$$\llbracket 0, 0 \rrbracket = [0, 0]\frac{2\alpha_u\alpha_v}{\gamma_{uv}}\left(1 - \frac{2\alpha_u\alpha_v}{\gamma_{uv}}(x_u - x_v)^2\right)$$

2) Evaluate all $\llbracket l_x, 0 \rrbracket$ using equation (B.24).

3) Evaluate all $\llbracket l_x, m_x \rrbracket$ using equation (B.25):

    Loop over all l_x.

    Loop over all $m_x > 0$.

    Use equation (B.25).

4) Repeat steps 1-3 for the y,z components.

5) Get $T_{ul,vm}$ from equation (B.16).

B.4 Fourier Transform Integrals

The plane wave integral between a pair of normalized Gaussian functions is defined as

$$\begin{aligned}\rho_{ul,vm}(G) &= \int d^3r\ X_{ul}(r)X_{vm}(r)e^{iGr} \\ &= N_{ul}N_{vm}[l_x|e^{iG_x x}|m_x][l_y|e^{iG_y y}|m_y][l_z|e^{iG_z z}|m_z], \qquad (B.26)\end{aligned}$$

where

$$[l_x|e^{iG_x x}|m_x] = \int_{-\infty}^{+\infty} dx\ e^{iG_x x}(x-x_u)^{lx}e^{-\alpha_u(x-x_u)^2}(x-x_v)^{mx}e^{-\alpha_v(x-x_v)^2}. \quad (B.27)$$

Two special identities can be derived from the form of equation (B.27):

1) The first identity comes from the invariance of the integral under translation of both Gaussian functions (except for the $e^{iG_x x}$ term)

$$\begin{aligned}&2\alpha_u[l_x+1|e^{iG_x x}|m_x] + 2\alpha_v[l_x|e^{iG_x x}|m_x+1] = \qquad (B.28) \\ &\quad iG_x[l_x|e^{iG_x x}|m_x] + l_x[l_x-1|e^{iG_x x}|m_x] + m_x[l_x|e^{iG_x x}|m_x-1].\end{aligned}$$

To prove the above equation, start with the following identity

$$0 = \int_{-\infty}^{+\infty} dx\ \frac{\partial}{\partial_x}\left(e^{iG_x x}(x-x_u)^{lx}e^{-\alpha_u(x-x_u)^2}(x-x_v)^{mx}e^{-\alpha_v(x-x_v)^2}\right), \quad (B.29)$$

which depends upon the Gaussian functions being 0 at ±∞. Applying the partial derivative $\frac{\partial}{\partial_x}$ to each term in (B.29) leads directly to (B.28).

2) The second identity,

$$(x_v - x_u)[l_x|e^{iG_x x}|m_x] = [l_x+1|e^{iG_x x}|m_x] - [l_x|e^{iG_x x}|m_x+1], \qquad (B.30)$$

is related to the fact that the integrand in equation (B.27) is only a function of $(x - x_u)$ and $(x - x_v)$ which is easily obtained by placing $(x_v - x_u)$ under the integration and by writing it as $(x - x_u) - (x - x_v)$.

When equations (B.28) and (B.30) are combined to zero out the $[l_x|e^{iG_x x}|m_x+1]$ contribution, the recursion relation that describes how to increment the value of l_x is obtained:

$$\begin{aligned}[l_x+1|e^{iG_x x}|m_x] &= (P_x - x_u + \frac{iG_x}{2\gamma_{uv}})[l_x|e^{iG_x x}|m_x] \\ &+ \frac{l_x}{2\gamma_{uv}}[l_x-1|e^{iG_x x}|m_x] + \frac{m_x}{2\gamma_{uv}}[l_x|e^{iG_x x}|m_x-1].\end{aligned} \tag{B.31}$$

Equations (B.28) and (B.30) can also be combined to zero out the $[l_x+1|e^{iG_x x}|m_x]$ contribution. This results in the recursion relation that describes how to increment the value of m_x:

$$\begin{aligned}[l_x|e^{iG_x x}|m_x+1] &= (P_x - x_v + \frac{iG_x}{2\gamma_{uv}})[l_x|e^{iG_x x}|m_x] \\ &+ \frac{l_x}{2\gamma_{uv}}[l_x-1|e^{iG_x x}|m_x] + \frac{m_x}{2\gamma_{uv}}[l_x|e^{iG_x x}|m_x-1].\end{aligned} \tag{B.32}$$

In practice the above recursion relations are used as follows:

1) Evaluate $[0|e^{iG_x x}|0] = [0,0]e^{iG_x P_x}e^{-G_x^2/4\gamma_{uv}}$.

2) Evaluate all $[l_x,0]$ using equation (B.31) .

3) Evaluate all $[l_x,m_x]$ using equation (B.32):

    Loop over all l_x.

    Loop over all $m_x > 0$.

    Use equation (B.32).

4) Repeat steps 1-3 for the y,z components.

5) Get $\rho_{ul,vm}(G)$ from equation (B.26).

B.4.1 Fourier Transform Integrals for Cubic Lattices

In dealing with crystalline systems, the plane wave integrals are used to evaluate the electron density in reciprocal space. The plane wave integrals are required for a large number of G vectors. But for cubic lattices, the G vectors can be written in the following form:

$$G_n = (n_x, n_y, n_z)\Delta G \qquad (B.33)$$

where the ΔG's are the vector steps in each direction. In that case, the 1D plane wave integrals

$$[l_x|e^{iG_x x}|m_x], \qquad [l_y|e^{iG_y y}|m_y], \qquad \text{and} \qquad [\mathrm{l_z}|\mathrm{e^{iG_z z}}|\mathrm{m_z}] \qquad (B.34)$$

can be evaluated in advance, i.e., before entering the loop over the 3D G vectors. That keeps the operations count within the loop over the 3D vectors to a minimum. The algorithm is as follows:

```
1) Evaluate range of $n_x, n_y$ and $n_z$ needed for 1D integrals.
1) Evaluate all 1D integrals using the recursion relations.
2) Evaluate all 3D integrals:
      Loop over all 3D $G$ vectors.
         Find $(n_x, n_y, n_z)$.
         Multiply the three 1D integrals together.
```

B.4.2 Screened Fourier Transform Integrals

The recursion relation for the Fourier transform integrals involving the screening density are the same as for the un-screened density except that the recursion relations are started with:

$$[0|e^{iG_x x}|0]^{\Omega} = [0,0]e^{iG_x P_x}e^{-G_x^2/4\Omega} \qquad (B.35)$$

instead of

$$[0|e^{iG_x x}|0] = [0,0]e^{iG_x P_x}e^{-G_x^2/4\gamma_{uv}}. \qquad (B.36)$$

To prove the above assertion, start with the screening density, $\rho_{ul,vm}(r)$, written as in equation (B.6) and look only at the x component of the integration:

$$\begin{aligned} I_x^{\Omega} &= (\frac{\Omega}{\gamma_{uv}})^{\frac{1}{2}} \sum_i c_i^x \frac{\partial^i}{\partial P_x^i} \int dx \, e^{iG_x x} e^{-\Omega\,(x-P_x)^2} \\ &= (\frac{\Omega}{\gamma_{uv}})^{\frac{1}{2}} \left(\int dx \; e^{iG_x x} e^{-\Omega\, x^2} \right) \sum_i c_i^x \frac{\partial^i}{\partial P_x^i} e^{iG_x P_x} \\ &= (\frac{\pi}{\gamma_{uv}})^{\frac{1}{2}} e^{-G_x^2/4\Omega} \sum_i c_i^x \frac{\partial^i}{\partial P_x^i} e^{iG_x P_x} . \end{aligned} \tag{B.37}$$

For the normal density, $\rho_{ul,vm}(r)$, just replace Ω everywhere by γ_{uv} to obtain:

$$I_x = (\frac{\pi}{\gamma_{uv}})^{\frac{1}{2}} e^{-G_x^2/4\gamma_{uv}} \sum_i c_i^x \frac{\partial^i}{\partial P_x^i} e^{iG_x P_x} . \tag{B.38}$$

The proof is completed by noting that the only difference between equations (B.37) and (B.38) is in the prefactor term, $e^{-G_x^2/4\Omega}$ versus $e^{-G_x^2/4\gamma_{uv}}$.

B.5 Two-Center Potential Integrals

The two-center potential integral between a pair of normalized Gaussian functions is given by

$$A_{ul,vm}(r) = \int d^3r' \; \frac{X_{ul}(r')X_{vm}(r')}{|r - r'|} . \tag{B.39}$$

The recursion relation is obtained by inserting the following identity:

$$\frac{1}{|r - r'|} = \frac{2}{\sqrt{\pi}} \int_0^{\infty} du \; e^{u^2\,(r-r')^2} \tag{B.40}$$

in the above equation and by using the three-center overlap recursion relation similar to equation (B.13). The end result is:[7]

$$A_{ul,vm}^{(n)}(C) = < X_{u,l+\delta_i}(r) | \frac{1}{|r - C|} | X_{v,m}(r) >^{(n)}$$

$$\begin{aligned}
&= (P_i - \xi_{u,i}) < X_{u,l+\delta_i}|\frac{1}{|r-C|}|X_{v,m} >^{(n)} \qquad (B.41)\\
&+ (P_i - C_i) < X_{u,l+\delta_i}|\frac{1}{|r-C|}|X_{v,m} >^{(n+1)}\\
&+ \frac{l_i}{2\gamma_{uv}}\left(< X_{u,l-\delta_i}|\frac{1}{|r-C|}|X_{v,m} >^{(n)} - < X_{u,l-\delta_i}|\frac{1}{|r-C|}|X_{v,m} >^{(n+1)}\right)\\
&+ \frac{m_i}{2\gamma_{uv}}\left(< X_{u,l}|\frac{1}{|r-C|}|X_{v,m-\delta_i} >^{(n)} - < X_{u,l}|\frac{1}{|r-C|}|X_{v,m-\delta_i} >^{(n+1)}\right),
\end{aligned}$$

where $A^{(0)}_{ul,vm}(C)$ is the integral we are interested in, the C is the point where the potential is evaluated and the starting point of the recursion are the following:

$$< X_{u,l=0}|\frac{1}{|r-C|}|X_{v,m=0} >^{(n)}= 2(\frac{\gamma_{uv}}{\pi})^{\frac{1}{2}} < X_{u,l=0}|X_{v,m=0} > F_n(T) \qquad (B.42)$$

with $T = \gamma_{uv}(P-C)^2$ and

$$F_n(T) = \int_0^1 dt\; t^{2n}\, e^{-Tt^2}. \qquad (B.43)$$

Many methods have been used to evaluate the $F_n(T)$'s based on either two formulas seven-term Taylor expansion[4] or a seven formulas scheme.[6] With the advent of better computers, it becomes much cheaper to use a two formulas method based on a simple interpolation scheme.[60] This method involves fewer operations for each value of $F_n(T)$ desired, but requires storing 10000 values of $F_n(T)$ at preselected values of T. In order to improve the accuracy, the interpolation is done using the variable $x = 1/(1+T)$ which covers the interval $[0,1]$ and over which the 10000 values of x are spaced equally. In order to reproduce the asymptotic behavior of $F_n(T)$,

$$\lim_{T\to\infty} F_n(T) = \frac{(2n-1)!!}{2(2T)^n}(\frac{\pi}{T})^{\frac{1}{2}}, \qquad (B.44)$$

the function $F_n(T(x)) \times T^{n+1/2}$ is stored for $T > 1$ instead of $F_n(T(x))$. The final absolute accuracy is better than 10^{-8} for any value of T.

In addition the careful evaluation of $F_n(T)$, the evaluation of the two-center potential integrals can be sped up by computing the integrals in a rotated frame where the atoms ξ_u and ξ_v lie along one of the axes.[61] In that frame, 2/3 of the prefactors of the form $(P_i - \xi_{u,i})$ and $(P_i - C_i)$ are now zero.

B.5.1 Screened Two-Center Potential Integral

The recursion relations for the two-center potential integrals involving the screening density are the same as for the un-screened density except that the function $F_n(\gamma_{uv}(P-C)^2)$ is replaced by $(\Omega/\gamma_{uv})^{1/2} F_n(\Omega(P-C)^2)$ in equation (B.42).

To prove the above assertation, start with the two-center potential integral

$$A^{\Omega}_{ul,vm}(C) = \int d^3r' \, \frac{\rho^{\Omega}_{ul,vm}(r')}{|r-r'|} \tag{B.45}$$

and replace the screening density by equation (B.6):

$$\begin{aligned} A^{\Omega}_{ul,vm}(C) = \sum_{ijk} C_{ijk} \, \frac{\partial^i}{\partial P^i_x} \, \frac{\partial^j}{\partial P^j_y} \, \frac{\partial^k}{\partial P^k_z} \\ \times (\frac{\Omega}{\gamma_{uv}})^{\frac{3}{2}} \int d^3r \, \frac{e^{-\Omega\,(r-P)^2}}{|r-C|} \end{aligned} \tag{B.46}$$

where

$$C_{ijk} = N_{ul} N_{vm} c^x_i c^y_j c^z_k. \tag{B.47}$$

Using the following relation:

$$\int d^3r \, \frac{e^{-\Omega\,(r-P)^2}}{|r-C|} = \frac{2\pi}{\Omega} F_0(\Omega\,(P-C)^2) \tag{B.48}$$

equation (B.46) becomes:

$$A^{\Omega}_{ul,vm}(C) = \sum_{ijk} C_{ijk} \, \frac{\partial^i}{\partial P^i_x} \, \frac{\partial^j}{\partial P^j_y} \, \frac{\partial^k}{\partial P^k_z} \, (\frac{\Omega}{\gamma_{uv}})^{\frac{1}{2}} F_0(\Omega\,(P-C)^2). \tag{B.49}$$

Applying one of the partial derivatives leads to:

$$\begin{aligned}\frac{\partial^i}{\partial P_x^i}F_0(\Omega\,(P-C)^2) &= \int_0^1 dt\;\frac{\partial^i}{\partial P_x^i}\,e^{-Tt^2} \\ &= i!\,(-)^i\sum_r^{[i/2]}\frac{2^{i-2r}}{r!\,(i-2r)!}\Omega^{i-2r}F_{i-2r}(\Omega\,(P-C)^2).\end{aligned} \tag{B.50}$$

In the above equation the function $F_n(\Omega\,(P-C)^2)$ is always multiplied by a factor of the form Ω^n. For the un-screened density $\rho_{ul,vm}(r)$ the function $F_n(\gamma_{uv}\,(p-C)^2)$ is always multiplied by a factor of the form γ_{uv}^n. The proof is completed by noting that there is an extra factor of $(\frac{\Omega}{\gamma_{uv}})^{1/2}$ in equation (B.49).

As stated in equation (67), the screened potential $A^{\Omega}_{ul,vm}(C)$ is equal to the integral $(g_{\alpha(r-C)}||\rho_{ul,vm}(r'))$ [defined in equation (55)] where the exponent of the screening Gaussian density $g_\alpha(r-C)$ [defined in equation (45)] is related to Ω by equation (69). For s basis functions, i.e., $l=0$ and $m=0$, the identity can be reduced to:

$$(\frac{\alpha}{\pi})^{\frac{3}{2}}\int d^3r\int d^3r'\;\frac{e^{-\alpha\,(r-C)^2}e^{-\gamma_{uv}\,(r'-P)^2}}{|r-r'|}=(\frac{\Omega}{\gamma_{uv}})^{\frac{3}{2}}\int d^3r'\;\frac{e^{-\Omega\,(r'-P)^2}}{|C-r'|} \tag{B.51}$$

which can be proven by using identity (B.40) on both sides and performing the integrations over r and r'. The proof for higher angular momentums is done by applying the derivatives $\partial^i/\partial P_x^i$, $\partial^j/\partial P_y^j$ and $\partial^k/\partial P_z^k$ to equation (B.51) and using equations (B.5) and (B.6) that defines the screened and un-screened density.

Appendix C. Symmetry

C.1 Convention for Symmetry Operations

The application of symmetry operation l on a 3D vector r is given by:

$$r_l = U_p(l)\ r + t_l, \tag{C.1}$$

where $U_p(l)$ is the 3×3 rotation matrix for symmetry operation l.

The application of a symmetry operation on atom a position ξ_a is given by

$$\xi_{al} + R_{al} = U_p(l)\ \xi_a + t_l, \tag{C.2}$$

where ξ_{al} labels the rotated atom position translated into the first unit cell and R_{al} is the lattice vector used for doing the translation.

C.2 Rotation of Cartesian Basis Functions

In dealing with basis functions, we will make use of the above equation with a basis function index instead of an atom index

$$\xi_{ul} + R_{ul} = U_p(l)\ \xi_u + t_l, \tag{C.3}$$

where ξ_u and ξ_{ul} are atom positions of the original basis function u and of the rotated basis function u_l. In rotating Cartesian basis functions, the effect of the rotation on the angular part of the basis function must also be taken care of. For S basis functions, the rotation matrix $U^{(S)}$ is just a unit scalar. For P basis functions, the rotation matrix $U^{(P)}_{(i)(j)}$ is identical to the 3×3 matrix $U_p(l)$ used to rotate 3-dimensional vectors. For D basis functions, if the six Cartesian functions, x^2, y^2, z^2, xy, xz, yz, are written as (ij) then the 6×6 rotation matrix is given by:

$$U^{(D)}_{(ij)(kl)} = \sqrt{\frac{N_{(kl)}}{N_{(ij)}}} \sum_{(kl)} U^{(P)}_{(i)(k)} U^{(P)}_{(j)(l)} \tag{C.4}$$

where the sum is over all the permutations of the pair (kl) and $N_{(ij)}$ is the normalization constant of the Cartesian function represented by (ij).

In a similar manner, the ten F basis functions x^3, y^3, z^3, x^2y, x^2z, y^2x, y^2z, z^2x, z^2y, xyz, are written as (ijk) and the 10×10 rotation matrix is given by:

$$U^{(F)}_{(ijk)(lmn)} = \sqrt{\frac{N_{(lmn)}}{N_{(ijk)}}} \sum_{(lmn)} U^{(P)}_{(i)(l)} U^{(P)}_{(j)(m)} U^{(P)}_{(k)(n)}. \qquad (C.5)$$

After all the angular rotation matrices $U^{(S)}$, $U^{(P)}$, $U^{(D)}$ and $U^{(F)}$ are built, the total rotation matrix, T, that transforms all the basis functions at once can be generated.

C.3 Rotation of Matrix Elements

The effect of symmetry on the Hamiltonian matrix elements $H_{uvR} =< u|H|vR >$ is given by

$$H_{u'v'R'} = T_l \; H_{uvR} \; T_l^t, \qquad (C.6)$$

where T_l is the rotation matrix for the basis functions associated with symmetry operation l and u', v' and R' are the rotated indices of the original product (uvR). The full expression for R' is

$$R' = U_p(l) \; R + R_{vl} - R_{ul}. \qquad (C.7)$$

The above equation is valid not only for the Fock operator but also for any other operator including the identity (which leads to evaluating overlap matrix).

For density matrix elements, D_{uvR}, the effect of symmetry is given by

$$D_{u'v'R'} = T_{l'}^t \; D_{uvR} \; T_{l'}, \qquad (C.8)$$

where $T_{l'}$ is the rotation matrix associated to the inverse of symmetry operation l. The above can be easily derived from the following identity

$$T_{race}(D_{uvR}\ S_{uvR}) = T_{race}(D_{u'v'R'}\ S_{u'v'R'}). \tag{C.9}$$

Beside using the rotations described in equations (C.6) and (C.7), we will also make use of their opposites

$$H_{uvR} = T_{l'}\ H_{u'v'R'}\ T_{l'}^{t} \tag{C.10}$$

and

$$D_{uvR} = T_{l}^{t}\ D_{u'v'R'}\ T_{l}. \tag{C.11}$$

C.4 Application of Symmetry to Grid Generation

For each atom in the first unit cell an atomic grid is generated. By using symmetry only the grid for the non-equivalent atoms is saved. Those atomic grids are further reduced by using the symmetry operations that leave the atom unchanged. The weights of the remaining gridpoints are changed to include the contribution of all the other gridpoints that can be obtained by symmetry.

References

1. J.H. Weaver, J.L. Martins, T. Komeda, Y. Chen, T.R. Ohno, G.H. Kroll, N. Troullier, R.E. Haufler and R.E. Smalley, *Phys. Rev. Lett.* **66**, 1741 (1991).
2. R. Car and M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 (1985).
3. M.C. Payne, M.P. Teter, D.C. Allan, T.A. Arias and J.D. Joannopoulos, *Rev. Mod. Phys.* **64**, 1045 (1992).
4. I. Shavitt in *Methods in Computational Physics*, edited by B. Alder, S. Fernbach and M. Rotenberg (Academic Press, New York, 1963), Vol2.
5. H. Taketa, S. Huzinaga and K. O-ohata, *J. Phys. Soc. Jap.* **21**, 2313 (1966).
6. L.E. McMurchie and E.R. Davidson, *J. Comp. Phys.* **26**, 218 (1978).
7. S. Obara and A. Saika, *J. Chem. Phys.* **84**, 3963 (1986).
8. A.D. Becke, *J. Chem. Phys.* **88**, 2547 (1988).
9. G. te Velde and E.J. Baerends, *J. Comp. Phys.* **99**, 84 (1992)
10. M.R. Pederson and K. A. Jackson, *Phys. Rev.* **B 41**, 7453 (1990).
11. V.I. Lebedev, *Sib. Mat. Zh.* **18**, 132 (1977); *Zh. Vychisl. Mat. Mat. Fiz.* **16**, 293 (1976); *Zh. Vychisl. Mat. Mat. Fiz.* **15**, 48 (1975).
12. P.P. Ewald, *Ann. Phys. (Leipzig)* **64**, 253 (1921).
13. C. Pisani, R. Dovesi and C. Roetti, "Hartree-Fock Ab Initio Treatment of Crystalline Systems," *Lecture Notes in Chemistry* Vol. 48 (1988).
14. H. Sambe and R.H. Felton, *J. Chem. Phys.* **61**, 3862 (1974).
15. B.I. Dunlap, J. Andzelm and J.W. Mintmire, *Phys. Rev.* **A 42**, 6354 (1990).
16. G. te Velde and E.J. Baerends, *Phys. Rev.* **B 44**, 7888 (1991)
17. T.P. Hamilton and P. Pulay, *J. Chem. Phys.* **84**, 5728 (1986)
18. M.P. Teter, M.C. Payne and D.C. Allan, *Phys. Rev.* **B 40**, 12255 (1989).

19. P. Hohenberg and W. Kohn, *Phys. Rev.* **B 136**, 864 (1964).

20. W. Kohn and L.J. Sham, *Phys. Rev.* **A 140**, 1133 (1965).

21. J.P. Perdew and A. Zunger, *Phys. Rev.* **B 23**, 5048 (1981).

22. D.M. Ceperley and B.I. Alder, *Phys. Rev. Lett.* **45**, 566 (1980).

23. G.B. Bachelet, D.R. Hamann and M. Schluter, *Phys. Rev.* **B 26**, 4199 (1982).

24. F.E. Harris in *Theoretical Chemistry Advances and Perspective* edited by H. Eyring and D. Henderson (Academic Press, New York, 1975).

25. M.P. Tosi, *Solid State Physics* **16**, 107 (1964).

26. R.N. Euwena and G.T. Surratt, *J. Phys. Chem.* **36**, 67 (1975).

27. P.E. Gill, W. Murray and M.H. Wright, *Practical Optimization*, 1981.

28. R.A. Friesner, *J. Phys. Chem.* **92**, 3091 (1988).

29. C.D. Clark, P.J. Dean and P.V. Harris, *Proc. R. Soc. London*, Ser. **A 277**, 312 (1964).

30. L. Hedin and B.I. Lundqvist, *J. Phys.* **C 4**, 2064 (1971).

31. D.J. Chadi and M.L. Cohen, *Phys. Rev.* **B 8**, 5747 (1973)

32. S.C. Erwin, M.R. Pederson and W.E. Pickett, *Phys. Rev.* **B 41**, 10347 (1990).

33. S.-H. Wei, H. Krakauer and M. Weinert, *Phys. Rev.* **B 32**, 7792 (1985).

34. W.E. Pickett and C.S. Wang, *Phys. Rev.* **B 30**, 4719 (1984).

35. G.B. Bachelet, H.S. Grenside, G.A. Baraff and M. Schluter, *Phys. Rev.* **B 24**, 4736 (1981); **B 24**, 4736 (1981).

36. N. Troullier and J.L. Martins, *Phys. Rev.* **B 43**, 1993 (1991).

37. K. Kratschmer, L.D. Lamb and K. Fostipopoules, *Science* **242**, 1139 (1988).

38. A.F. Hebard et al., *Nature (London)* **350**, 600 (1991).

39. M.J. Rosseinsky et al., *Phys. Rev. Lett.* **66**, 2230 (1991).

40. S.P. Kelty, C.-C. Chen and C.M. Lieber, *Nature (London)* **352**, 233 (1991).

41. K. Tanigaki et al., *Nature (London)* **352**, 220 (1991).

42. P.A. Heiney et al., *Phys. Rev. Lett.* **66**, 2911 (1991).

43. S. Saito and A. Oshiyama, *Phys. Rev. Lett.* **66**, 2637 (1991).

44. N. Troullier and J.L. Martins, *Phys. Rev.* **B 46**, 1754 (1992).

45. C.S. Yannoni et al., *J. Am. Chem. Soc.* **113**, 3190 (1991).

46. W.I.F. David et al., *Nature (London)* **353**, 147 (1991).

47. Q.-M. Zhang, J.-Y. Yi and J. Bernholc, *Phys. Rev. Lett.* **66**, 2633 (1991).

48. Y. Guo, N. Karasawa and W.A. Goddard, *Nature (London)* **351**, 464 (1991).

49. T. Rabenau, A. Simon, R.K. Kremmer and E. Sohmen, *Z. Phys.* **B 90**, 69 (1993)

50. T. Takahashi et al., *Phys. Rev. Lett.* **68**, 1232 (1992)

51. J.H. Weaver, P.J. Benning, F. Stepniak and D.M. Poirier, *J. Phys. Chem. Phys.* **53**, 1707 (1992)

52. R.W. Lof, M.A. van Veenendaal, B. Koopmans, H.T. Jonkman and G.A. Sawatzky, *Phys. Rev. Lett.* **68**, 3924 (1992)

53. J.L. Martins and N. Troullier, *Phys. Rev.* **B 46**, 1766 (1992)

54. E.L. Shirley and S.G. Louie, *Phys. Rev. Lett.* **71**, 133 (1993).

55. W.Y. Ching, M.-Z. Huang, Y.-N. Xu, W.G. Harter and F.T. Chan, *Phys. Rev. Lett.* **67**, 2045 (1991).

56. S. Satpathy, V.P. Antropov, O.K. Andersen, O. Jepsen, O. Gunnarsson and A.I. Liechtenstein, *Phys. Rev.* **B 46**, 1773 (1992).

57. B.I. Dunlap, D.W. Brenner, J.W. Mintmire, R.C. Mowrey and C.T. White, *J. Phys. Chem.* **95**, 8737 (1991).

58. D.L. Lichtenberger, K.W. Nebesny, C.D. Ray, D.R. Huffman and L.D. Lamb, *Chem.*

Phys. Lett. **176**, 203 (1991).

59. J.A. Zimmerman, J.R. Eyler, S.B. Bach and S.W. McElvany, *J. Chem. Phys.* **94**, 3556 (1991).

60. R.A. Friesner *J. Chem. Phys.* **86**, 3522 (1987).

61. M.N. Ringnalda M. Belhadj and R.A. Friesner, *J. Chem. Phys.* **93**, 3397 (1990).

Table 1. Accuracy of HF total energy molecular calculations using numerical integration.

Molecule	Number of atoms	Number of bas. func.	Total Energy error (mH)	Error/atom (mH)
Oxygen	2	30	0.004	0.002
Water	3	25	-0.007	-0.002
Carbon Dioxide	5	45	-0.018	-0.006
Methane	12	35	-0.016	-0.003
Benzene	12	120	0.022	0.002
Glycine	10	100	0.024	0.002
Glutamine	20	200	0.072	0.004
Porphine	38	430	0.160	0.004
C_{60} (I_h)	60	540	0.206	0.003
C_{60} (T_h)	60	540	0.255	0.004

Table 2. Exponents and contraction coefficients for carbon atom in Dunning/Huzinaga and Pople 6-31G double zeta basis sets.

	Dunning/Huzinaga[a]		Pople/6-31G[b]	
	Exponent	Coefficient	Exponent	Coefficient
		first s orbital		
S	4233.	0.001220	3047.5249	0.0018347371
S	634.9	0.009342	457.36952	0.014037323
S	146.1	0.045452	103.94868	0.068842622
S	42.5	0.154657	29.210155	0.23218444
S	14.19	0.358866	9.286663	0.46794135
S	5.148	0.438632	3.163927	0.36231199
S	1.967	0.145918		
		second s orbital		
S	5.148	-0.168367	7.8682723	-0.11933242
S	0.4962	1.060091	1.8812885	-0.16085415
S			0.54424926	1.1434564
		first p orbital		
P	18.16	0.018539	7.8682723	0.068999067
P	3.986	0.115436	1.8812885	0.31642396
P	1.143	0.386188	0.54424926	0.74430829
P	0.3594	0.640114		
		third s orbital		
S	0.1533	1.0000000	0.16871448	1.000000000
		second p orbital		
P	0.1146	1.0000000	0.16871448	1.000000000

(a) S. Huzinaga, Ed., *Gaussian Basis Sets for Molecular Calculations*, Elsevier: Amsterdam, 1984.

(b) W.J. Hehre, R. Ditchfield and J.A. Pople, *J. Chem. Phys.* **56**, 2257 (1972); P.C. Hariharan and J.A. Pople *Theor. Chim. Acta* **28**, 213 (1973); M.S. Gordon, *Chem. Phys. Lett.* **76**, 163 (1980); M.J. Frisch, J.A. Pople and J.S. Binkley, *J. Chem. Phys.* **80**, 3265 (1984).

Table 3. Diamond energy levels in the valence and conduction bands for several different methods.

	This Work				
Root[(a)]	DH[(b)]	6-31G[(c)]	LCAO[(d)]	LAPW[(e)]	PS-PW[(f)]
Γ_1	-21.35	-21.42	-21.34	-21.06	-21.38
$\Gamma_{25'}$	0.00	0.00	0.00	0.00	0.00
$\Gamma_{15'}$	5.53	5.50	5.57	5.51	5.51
$\Gamma_{2'}$	13.55	13.52	13.56	13.13	13.56
X_1	-12.64	-12.70	-12.66	-12.48	-12.67
X_4	-6.30	-6.33	-6.34	-6.18	-6.31
X_1	4.70	4.66	4.79	4.68	4.64
X_4	16.60	16.62	16.71	16.41	16.81
$L_{2'}$	-15.51	-15.57	-15.52	-15.33	-15.53
L_1	-13.38	-13.44	-13.41	-13.14	-13.47
$L_{3'}$	-2.82	-2.83	-2.81	-2.73	-2.81
L_3	8.36	8.38	8.45	8.33	8.37
L_1	9.01	9.01	9.07	8.75	8.97

(a) All energies in eV with Γ'_{25} shifted to 0.0 eV. All calculations done with $a = 3.567$ Å, Hedin-Lundqvist exchange-correlation potential and 10 special k points.

(b) Dunning/Huzinaga basis set ($9s5p/3s2p$) plus 0.75 d function.

(c) Pople 6-31G basis set ($10s4p/3s2p$) plus 0.8 d function.

(d) Linear combination of atomic orbitals, Ref. 32.

(e) Linear augmented-plane-wave, Ref. 33.

(f) Pseudopotential plane wave, Ref. 34.

Table 4. Comparison of C_{60} fcc (Fm3) crystalline calculations for several different methods.

	This Work$^{(a)}$	PS-PW(1)$^{(b)}$	PS-PW(2)$^{(c)}$	LCAO$^{(d)}$	PS-LCAO$^{(e)}$
Geometry ($\AA$)					
R_1 $^{(f)}$	1.414	-	1.382	1.40	1.40
R_2 $^{(f)}$	1.455	-	1.444	1.46	1.46
Lattice $^{(g)}$	14.00	-	13.88	14.20	14.20
Energies (eV)					
Gap X $^{(h)}$	0.96/0.96	1.04	1.18	1.34	1.5
Gap Γ $^{(h)}$	1.69/1.69	1.6-1.7	1.71	1.87	1.9
W_{Hu} $^{(i)}$	0.84/0.83	0.6	0.58	0.55	0.42
W_{T1u} $^{(i)}$	0.62/0.62	0.5	0.5	0.5	0.4
W_{T1g} $^{(i)}$	0.74/0.73	0.6	0.5	-	-
$E_{exch.-corr.}$ $^{(j)}$	CP	-	CP	W	CP
Pseudopotential	No	Yes	Yes	No	Yes

(a) With 6-31G basis set and Dunning/Huzinaga basis set.

(b) Pseudopotential plane waves, small core radii ~ 0.4$\AA$, Ref. 54.

(c) Pseudopotential plane waves, large core radii ~ 0.8$\AA$, Ref. 44.

(d) Linear combination of atomic orbitals, Ref. 55.

(e) Pseudopotential linear combination of atomic orbitals, Ref. 43.

(f) Carbon-carbon distances in short bond and long bond for C_{60}.

(g) Lattice constant.

(h) Direct gaps at X and Γ points. Lowest gap is at the X point.

(i) Widths of the valence band, W_{Hu}, and first two conduction bands, W_{T1u} and W_{T1g}.

(j) Exchange-correlation potentials: CP for Ceperley-Alder and W for Wigner.

Table 5. Energy levels (eV) in molecular C_{60} using Ceperley-Alder exchange correlation potential.

Symmetry	6-31G	LCGTO$^{(a)}$	$\Delta\epsilon^{(b)}$
Gg	-0.12	-0.11	-0.01
Hu	-0.61	-0.56	-0.05
Hg	-1.94	-2.07	0.13
T2u	-2.04	-2.25	0.21
T1g	-3.25	-3.15	-0.10
T1u $^{(c)}$	-4.24	-4.26	0.02
Hu $^{(d)}$	-5.92	-5.94	0.02
Gg	-7.05	-7.12	0.07
Hg	-7.17	-7.24	0.07
Hu	-8.70	-8.78	0.08
Gu	-8.87	-8.83	-0.04
Hg	-9.17	-9.09	-0.08
T2u	-9.31	-9.38	0.07
Gu	-10.18	-10.10	0.08
T2g	-10.57	-10.58	0.01
Hg	-10.73	-10.70	-0.03
Gg	-10.88	-10.85	-0.03
Hg	-11.32	-11.26	-0.06
Gu	-11.67	-11.75	0.08
T1u	-11.84	-11.84	0.00
Ag	-12.24	-12.44	0.20
T2u	-12.42	-12.47	0.05
T1u	-13.12	-13.05	0.07
Hu	-13.36	-13.25	0.11

(a) Linear combination of Gaussian Type Orbitals basis set (11s7p/5s3p) with 0.6 *d* functions, $R_1 = 1.39\AA$ and $R_2 = 1.45\AA$, Ref. 57.

(b) Orbital energy difference $\epsilon_{6-31G} - \epsilon_{LCGTO}$

(c) Lowest unoccupied molecular orbital.

(d) Highest occupied molecular orbital.

Table 6. Energy levels (eV) at the Γ point in crystalline C_{60} for Ceperley-Alder exchange correlation potential.

	This...Work				This...Work		
Label	DH(a)	6-31G(b)	PS-PW(c)	Label	DH	6-31G	PS-PW
Ag	-18.74	-18.73	-18.77	Tg	-5.05	-5.08	-4.93
Tu	-18.15	-18.13	-18.15	Eg	-4.90	-4.92	-5.00
Tg	-17.30	-17.29	-17.26	Tg	-4.86	-4.88	-4.59
Eg	-17.27	-17.25	-17.22	Tu	-4.30	-4.32	-4.11
Tu	-16.24	-16.22	-16.10	Au	-4.30	-4.31	-4.09
Au	-15.34	-15.36	-15.38	Au	-3.48	-3.47	-3.60
Tu	-15.30	-15.29	-15.33	Eg	-3.39	-3.41	-3.19
Eg	-14.21	-14.21	-14.14	Tu	-3.37	-3.38	-3.45
Tg	-14.06	-14.07	-13.98	Tg	-3.33	-3.36	-3.12
Ag	-13.60	-13.61	-13.58	Eu	-3.01	-3.03	-2.77
Tg	-13.56	-13.56	-13.54	Tu	-3.01	-3.02	-2.76
Eu	-12.10	-12.12	-12.02	Tu	-2.54	-2.53	-2.63
Tu	-12.08	-12.09	-11.99	Eg	-1.83	-1.83	-1.85
Tu	-12.00	-12.01	-11.87	Ag	-1.56	-1.52	-1.62
Tu	-10.92	-10.94	-10.92	Tg	-1.04	-1.03	-1.08
Tg	-9.92	-9.94	-9.83	Tg	-0.94	-0.95	-0.97
Eg	-9.80	-9.82	-9.75	Eu	-0.25	-0.26	-0.23
Tg	-9.75	-9.78	-9.56	Tu (d)	0.00	0.00	0.00
Ag	-9.44	-9.45	-9.42	Tu	1.68	1.69	1.71
Ag	-8.36	-8.37	-8.25	Tg	2.25	2.26	2.37
Tg	-8.24	-8.27	-8.13	Tu	3.63	3.64	3.47
Eu	-7.57	-7.59	-7.37	Eg	4.11	4.08	4.05
Tu	-7.47	-7.49	-7.27	Tg	4.14	4.14	4.10
Tu	-7.25	-7.28	-7.11	Eu	4.83	4.81	4.97
Ag	-6.94	-6.95	-7.11	Ag	5.21	5.24	5.15
Tu	-6.55	-6.58	-6.53	Tu	5.37	5.39	5.52
Au	-6.00	-6.03	-5.83	Au	5.80	5.75	5.88
Tu	-5.97	-6.01	-5.80	Tg	6.15	6.15	6.24
Eg	-5.51	-5.55	-5.37	Tu	6.33	6.22	6.42
Tu	-5.46	-5.50	-5.56	Tg	8.23	7.17	7.45
Tg	-5.40	-5.39	-5.39	Tu	7.90	7.90	7.82
Tg	-5.23	-5.25	-5.26	Eg	8.84	8.65	8.32
Ag	-5.06	-5.09	-4.95	Tu	8.58	8.98	8.75

(a) Dunning/Huzinaga double zeta basis set.

(b) Pople 6-31G double zeta basis set.

(c) Pseudopotential plane waves, Ref. 44.
(d) Highest valence band (shifted to 0.0 eV).

Figure Captions

Figure 1. The diatomic projection functions along the line between the two atoms for different recursion order: s_1, s_2 and s_3. The projection function labeled $s_3 + s_{add}$ includes the adjustable sine-like function.

Figure 2. The band structure of fcc diamond near the Fermi energy. The top of the valence is shifted to 0.0 eV.

Figure 3. The total energy of diamond lattice as function of lattice parameter. The experimental lattice parameter is 3.567 $\AA$ which compares well to the calculated value of 3.546 $\AA$.

Figure 4. The band structure of C_{60} for the fcc (Fm3) structure near the Fermi energy. The top of the valence is shifted to 0.0 eV.

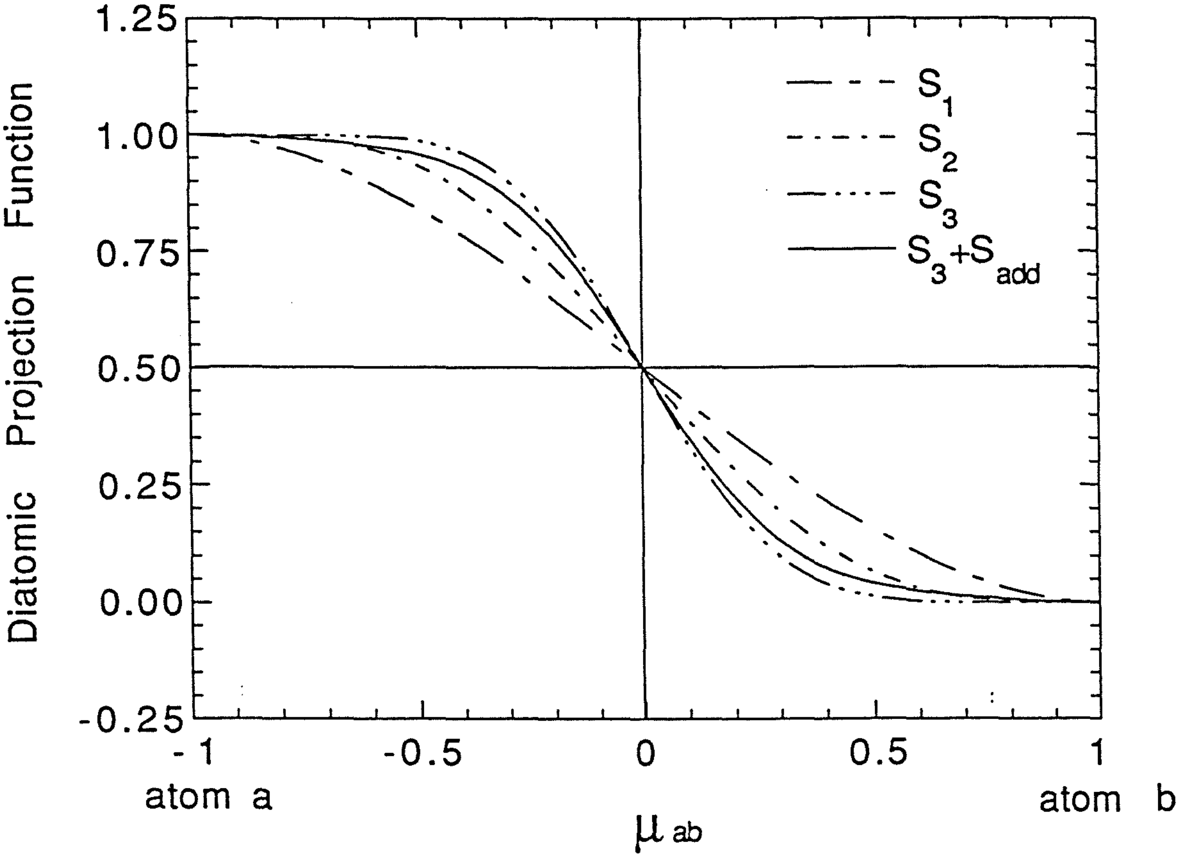

1.25
1.00
0.75
0.50
0.25
0.00
-0.25
Diatomic Projection Function
- 1
-0.5
0
0.5
1
atom a
atom b
μ_{ab}
S_1
S_2
S_3
S_3+S_{add}

Figure 1.

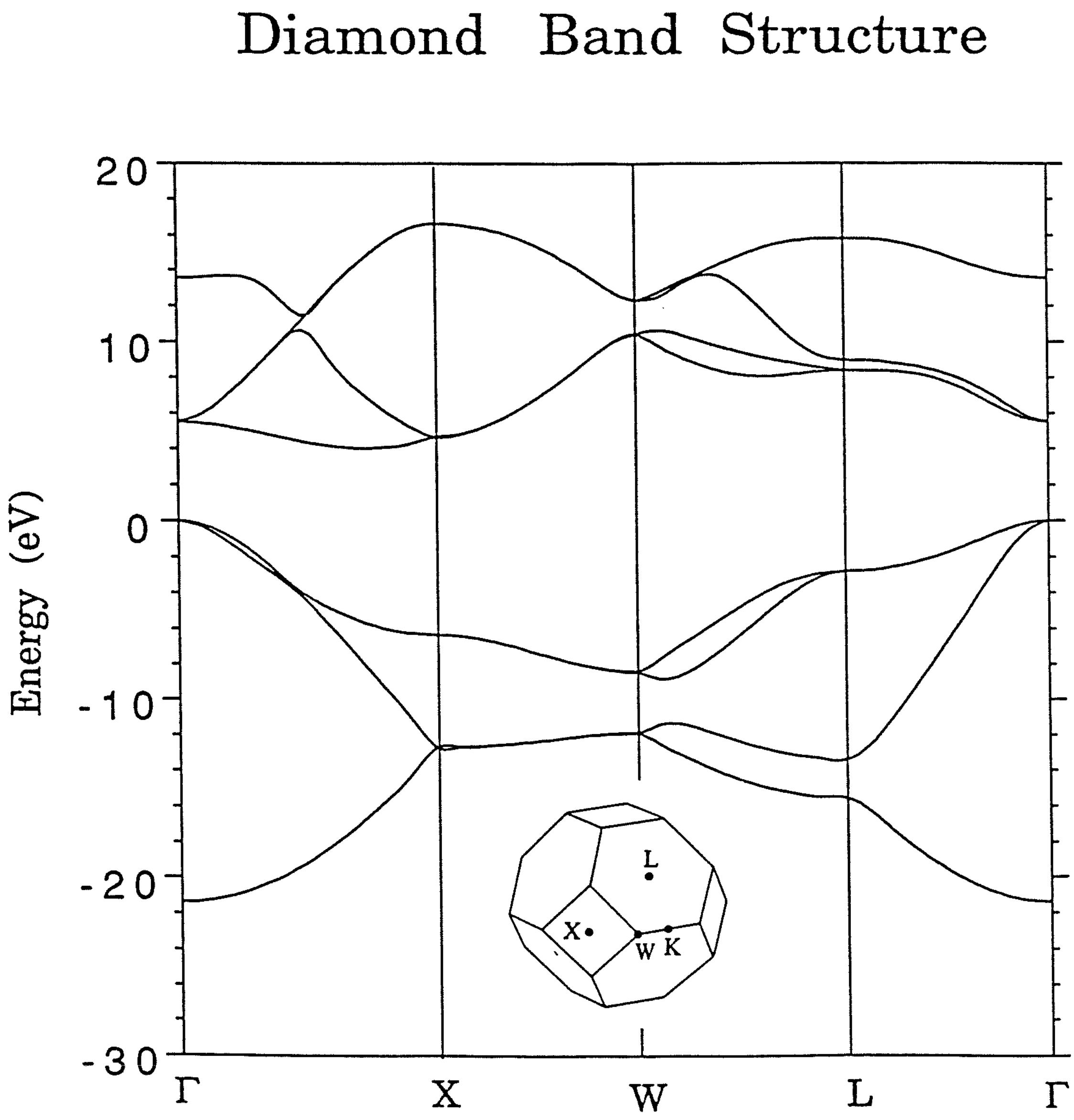


Figure 2.

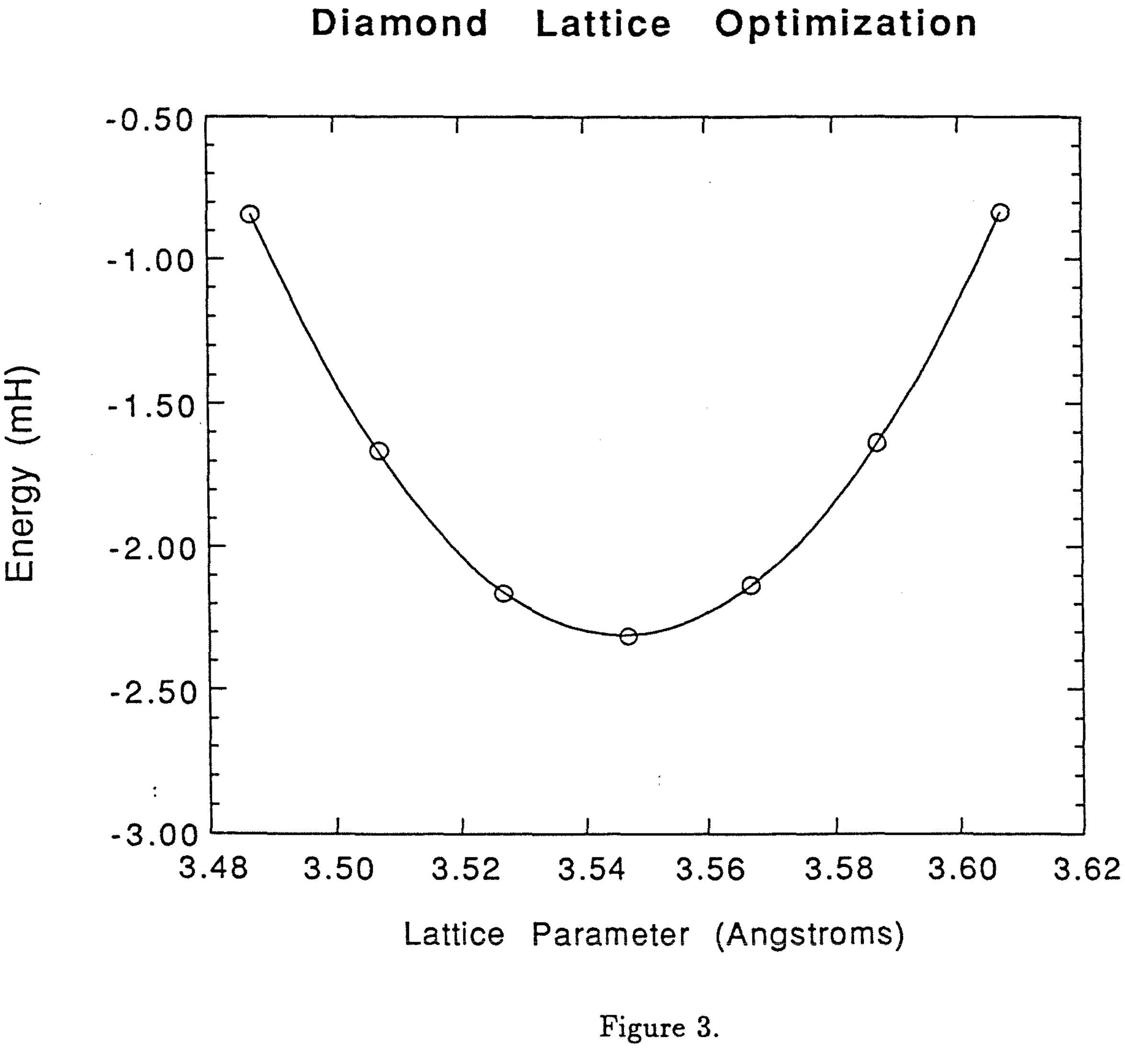

Diamond Lattice Optimization
-0.50
-1.00
-1.50
-2.00
-2.50
-3.00
Energy (mH)
3.48
3.50
3.52
3.54
3.56
3.58
3.60
3.62
Lattice Parameter (Angstroms)

Figure 3.

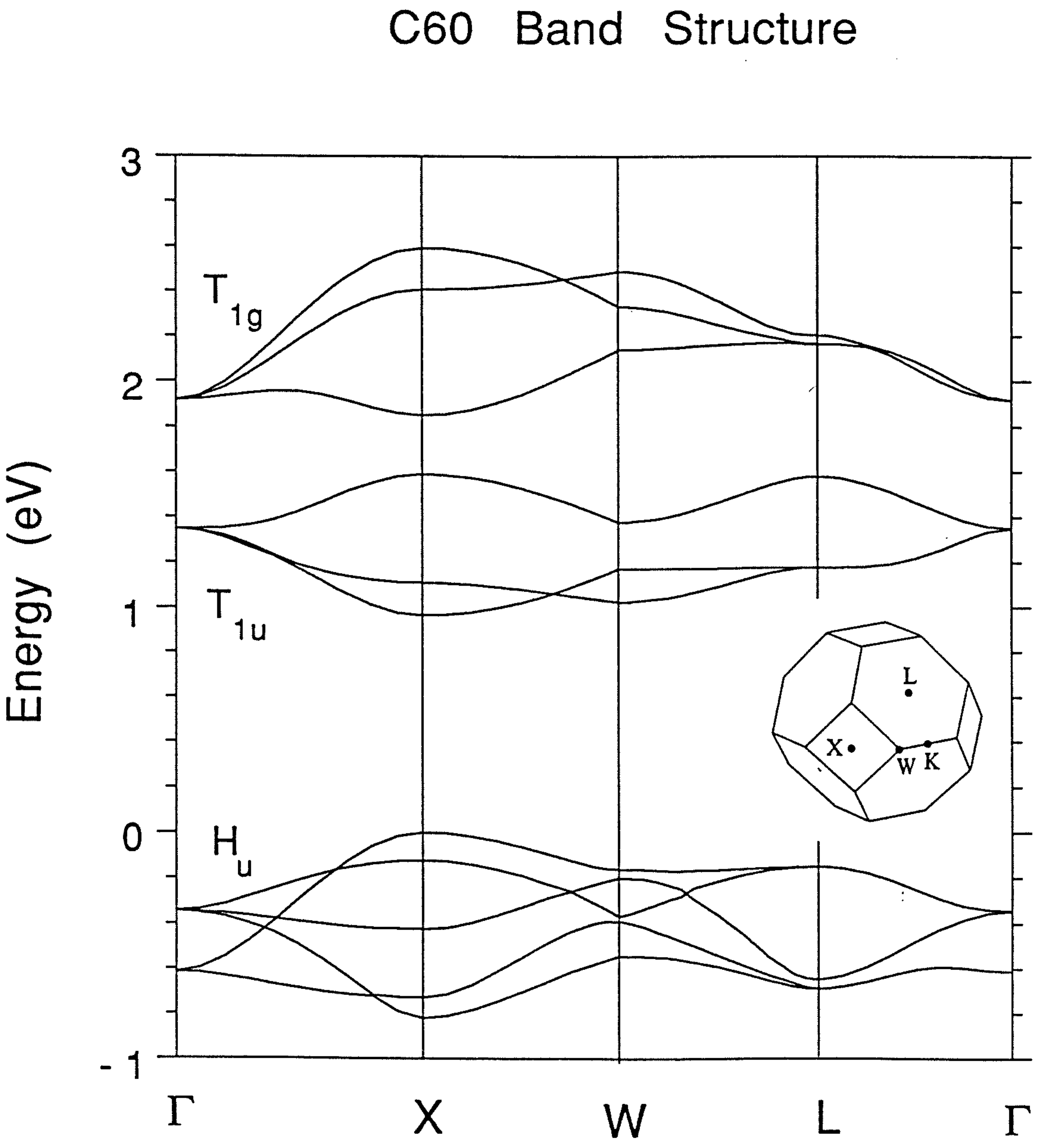


Figure 4.

Appendix 1

Electronic Structure and Valence-Bond Band Structure of Cuprate Superconducting Materials

Reprint Series
19 February 1988, Volume 239, pp. 896–899

SCIENCE

Electronic Structure and Valence-Bond Band Structure of Cuprate Superconducting Materials

YUEJIN GUO, JEAN-MARC LANGLOIS, AND WILLIAM A. GODDARD III

Electronic Structure and Valence-Bond Band Structure of Cuprate Superconducting Materials

YUEJIN GUO, JEAN-MARC LANGLOIS, WILLIAM A. GODDARD III

From ab initio calculations on various clusters representing the $La_{2-x}Sr_xCu_1O_4$ and $Y_1Ba_2Cu_3O_7$ classes of high-temperature superconductors, it is shown that (i) all copper sites have a Cu^{II} (d^9) oxidation state with one unpaired spin that is coupled antiferromagnetically to the spins of adjacent Cu^{II} sites; (ii) oxidation beyond the cupric (Cu^{II}) state leads not to Cu^{III} but rather to oxidized oxygen atoms, with an oxygen $p\pi$ hole bridging two Cu^{II} sites; (iii) the oxygen $p\pi$ hole at these oxidized sites is ferromagnetically coupled to the adjacent Cu^{II} d electrons despite the fact that this is opposed by the direct dd exchange; and (iv) the hopping of these oxygen $p\pi$ holes (in CuO sheets or chains) from site to site is responsible for the conductivity in these systems (N-electron band structures are reported for the migration of these localized charges).

TO ELUCIDATE THE MECHANISM RESPONSIBLE for high-temperature superconductivity in various cuprates, we carried out first principles quantum chemical calculations on models of the $La_{2-x}Ba_xCu_1O_4$ (denoted 2-1-4) and the $Y_1Ba_2Cu_3O_7$ (denoted 1-2-3) classes of systems (*1*, *2*). The resulting wavefunctions indicate that electrical conduction in these systems is dominated by hopping of oxygen $p\pi$ holes from site to site in the CuO sheets and chains, and we report the band structures based on these valence-bond localized states. In addition, there are important magnetic couplings between spins on adjacent copper atoms and between the conduction electrons (oxygen $p\pi$ holes) and the copper spins that are critical in the superconductivity.

Electronic structure of the reduced (Cu^{II}) system. The electronic wavefunctions were calculated with the generalized valence bond (GVB), Hartree-Fock (HF), or configuration interaction (CI) methods (*3*). Calculations (*4*–*7*) were carried out on finite clusters as indicated in Fig. 1. In each case we included explicitly all electrons on the atoms shown plus the point charge approximation to all other ions within about 8 Å. The sphere size was adjusted slightly (up to 0.2 Å) so that balanced sets of ions were included, with the outer boundary always being oxygen. The charge on the outer layer was scaled so that the whole cluster is neutral. The number of explicit electrons corresponds to having O^{2-} and Cu^{2+} or Cu^{3+} at each atom shown. The wavefunctions were calculated self-consistently for each state.

We find in all cases (four-, five-, and six-coordinate) that the optimum wavefunctions have nine electrons in d orbitals on each copper, with the singly occupied orbital pointing at the four short bonds. For example, the singly occupied d-like orbitals for the 1-2-3 chain are shown in Fig. 2, a and b (cluster Cu_3O_{10}). The total population in d orbitals on the three copper sites is $d^{9.15}$, $d^{9.17}$, and $d^{9.15}$. The electrons formally considered on O^{2-} are partially shared with the copper so that the total charge on the copper is +0.41, +0.23, and +0.41 for this cluster (rather than +2, +2, and +2). Similar results are obtained for the 1-2-3 stub (Cu_3O_{12}), where the d populations are $d^{9.05}$, $d^{9.17}$, and $d^{9.05}$ (total charges 0.44, 0.45, and 0.44). For the 2-1-4 system (Cu_2O_{11}), the d populations are $d^{9.11}$ and $d^{9.11}$ (total charges −0.23 and −0.23).

With one singly occupied orbital per copper, the Cu_2 systems lead to two low-lying spin states ($S = 0$ and $S = 1$), and the Cu_3 systems lead to three low-lying states ($S = 1/2$, $S = 1/2$, and $S = 3/2$). We solved self-consistently for each of these states using the GVB wavefunction and fitted the resulting energies to a Heisenberg spin Hamiltonian,

$$H = -\sum_{i<j} 2J_{ij}\, S_i \cdot S_j \qquad (1)$$

For the Cu_2 cases, the singlet is lower because the orbitals overlap slightly ($\sigma = 0.05$), leading to an "antiferromagnetic" exchange integral J, where $J = -205$ K for the Cu_2O_{11} sheet and $J = -244$ K for the Cu_2O_7 chain. From the Cu_3 systems we find that $|J_{dd}| < 0.1$ K for second nearest neighbors ($\sigma = 0.0006$). In one and two dimensions, a system described by Eq. 1 with negative J leads to an ordering (Néel) temperature of $T_N = 0$ K (*8*). Hence, long-range order is not expected at finite T. Experimental evidence on La_2CuO_{4-y} suggests that there are strong short-range couplings; however, evidence for long-range order is conflicting, with one report of $T_N = 220$ K (*9*). For the Cu_3 stub (Cu_3O_{12}), the $d_{z^2-y^2}$ orbitals of the center (chain) copper are orthogonal to the $d_{x^2-y^2}$ orbitals of the top and bottom (sheet) copper sites, leading to weak ferromagnetic coupling ($J_{dd} = +11$ K).

Although each system has an array of singly occupied copper d orbitals, it does not lead to electrical conduction. The reason is that the charge transfer state ($Cu^{II} \cdots Cu^{II} \rightarrow Cu^{I} \cdots Cu^{III}$) is very high in energy (about 10 eV for the Cu_2O_{11} cluster). In band language, the system has a large Hubbard U parameter (one center electron-electron repulsion), leading to an exceedingly narrow band.

The oxidized ("Cu^{III}") system. The cuprates exhibiting superconductivity all are oxidized further than Cu^{2+}. Assuming no oxygen vacancies, the copper of $La_{1.85}Sr_{0.15}Cu_1O_4$ is 15 percent oxidized, and the copper of $Y_1Ba_2Cu_3O_7$ is 33 percent oxidized. However, other systems (for example, $La_4BaCu_5O_{13}$ and $La_5SrCu_6O_{15}$) have similar levels of oxidation but do not exhibit superconductivity (down to 5 K) (*10*). For all systems discussed here we find that oxidation of the Cu^{II} (d^9) leads not to Cu^{III} (d^8)

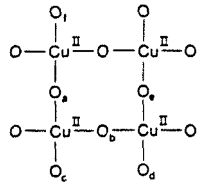
$$\text{O–Cu}^{II}\text{–O–Cu}^{II}\text{–O} \rightarrow \text{O–Cu}^{III}\text{–O–Cu}^{II}\text{–O} \qquad (2)$$

but rather to oxidation out of an oxygen $p\pi$ singly occupied orbital (see Fig. 2c) located between two Cu^{II} sites,

$$\text{O–Cu}^{II}\text{–O–Cu}^{II}\text{–O} \rightarrow \text{O–Cu}^{II}\text{–O}^{+}\text{–Cu}^{II}\text{–O} \qquad (3)$$

Thus, in the oxidized systems, all coppers have the Cu^{II} (d^9) oxidation state, but each oxidation leads to a singly occupied oxygen $p\pi$ orbital that is spin-coupled to the various singly occupied copper d orbitals. We will show below that hopping of such oxygen $p\pi$ holes from site to site leads to electrical conduction if the hole is on the proper oxygen.

The $La_{2-x}Sr_xCu_1O_4$ system has two-dimensional sheets of copper and oxygen as in



I

where each Cu-O bond is 1.89 Å. In addition, there are apex oxygens 2.40 Å above and below each copper. For $x = 0$, we find that oxidation of the in-plane (sheet) oxygen

Arthur Amos Noyes Laboratory of Chemical Physics, California Institute of Technology, Pasadena, CA 91125.

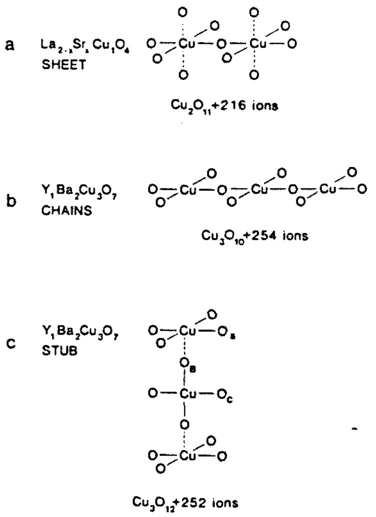


Fig. 1. Clusters used in GVB calculations. The array of point charges extends out to ~8 Å from each copper (adjusted so that the outer shells are oxygens). The charges on the point charges are the nominal values (O^{2-}, Y^{3+}, Ba^{2+}) except for Cu where $Cu^{+2.15}$ was used for $La_{1.85}Sr_{0.15}CuO_4$ and $Cu^{+2.33}$ was used for $Y_1Ba_2Cu_3O_7$. In addition, for $La_{2-x}Sr_xCuO_4$ an averaged charge $Z = 3 - \frac{1}{2}x$ was used for the LaSr system. The number of explicit electrons was based on the nominal charges [for example, $(Cu_3O_{10})^{14-}$ with three Cu^{II}]. In all calculations, all occupied orbitals are strongly bound (lowest IP = 5.4 eV).

is favored by 0.39 eV. Each of these oxygens has two $p\pi$ orbitals perpendicular to the Cu-O-Cu axis, but the $p\pi$ orbital in the plane of structure I is the preferred one for oxidation (by 0.4 eV). This leads to good electrical conduction as discussed below. On the other hand, such a subtle change as replacing part of the La^{3+} with Sr^{2+} can shift the relative ionization potentials (IP) of the apex and sheet oxygens significantly. Thus, with $x = 0.3$, we find that the apex oxygen is preferentially oxidized (by 0.45 eV). We believe that it is subtle shifts in relative IP that lead to a loss in superconductivity above $x = 0.3$ and that this is responsible for the semiconductor character of systems such as $La_4BaCu_5O_{13}$ and $La_5SrCu_6O_{15}$ (which have sheets or chains similar to the 2-1-4 and 1-2-3 systems). Thus, subtle changes in the cations (charges or placement) or in the structure (for example, owing to higher pressure) might change these semiconductors into superconductors.

For the $Y_1Ba_2Cu_3O_{7-y}$ system, there are three types of oxygens, O_C, O_S, and O_B for chain, sheet, and bridge, as indicated in Fig. 1c, but we find a very strong preference for oxidizing the chain oxygens. For $y = 0$, there must be one oxygen hole per central copper and hence an idealized structure would have every bridging O_C oxidized. However, with sufficient positive charge along the chain, it becomes favorable to oxidize sheet oxygens. Holes in O_C can hop easily to adjacent O_C, leading to high mobility if the chain is perfect. Holes in O_S would lead to conduction in the sheets much as in $La_{2-x}Sr_xCu_1O_4$. These two-dimensional sheets should be less sensitive to defects than the chains. Holes in O_B probably serve to communicate between the O_C and O_S sites. Thus, we expect an equilibrium population of holes among O_B, O_C, and O_S, with O_C and O_S important in electrical conduction.

Magnetic coupling in oxidized states. With three singly occupied orbitals, the spins in a Cu^{II}–O^+–Cu^{II} triad can be coupled in three ways ($S = 3/2$ and two $S = 1/2$). Since the singly occupied oxygen $p\pi$ orbital is orthogonal to the Cu^{II} orbitals, there is ferromagnetic coupling between the oxygen $p\pi$ and adjacent copper d orbitals. The spin coupling between the Cu^{II} spins still prefers singlet; however, the Cu-O exchange is much stronger, leading to a ferromagnetic (quartet) ground state. Thus the optimal magnetism of a chain is as in

$$\mathrm{O{-}Cu^{II}\uparrow{-}O^{+}\uparrow{-}Cu^{II}\uparrow{-}O{-}Cu^{II}\downarrow{-}O} \qquad \text{(II)}$$

We solved self-consistently for all possible spin states of the various clusters and fitted the results with Eq. 1 to obtain coupling terms of $J_{OCu} = 383$ K, $J_{dd} = -205$ K, and $J^{\dagger}_{dd} = -139$ K for sheets of $La_{2-x}Sr_xCu_1O_4$, and $J_{OCu} = 405$ K, $J_{dd} = -244$ K, and $J^{\dagger}_{dd} = -164$ K for chains of $Y_1Ba_2Cu_3O_7$. Here J_{OCu} is for adjacent oxygen $2p$ and copper $3d$ orbitals, $J^{\dagger}_{dd}$ is for two copper atoms with the intervening oxygen oxidized (the center and left Cu of structure II), and J_{dd} is for a copper pair where the intervening oxygen is not oxidized (the central and right Cu of structure II). Note that the J_{dd} with an intervening O^+ is smaller than that for a normal oxygen (164 K versus 244 K); this is because superexchange decreases when the central atom (oxygen) is less negative (*11*).

In the accompanying report (*12*) it is shown that the interplay between the O-Cu and Cu-Cu magnetic couplings is responsible for the superconductivity in these systems.

Electron correlation. It is important to emphasize that the many-body electron correlations implicit in the GVB wavefunction are essential to a proper description of these clusters (*3*). For example, an ordinary HF calculation on the Cu_2O_7 cluster yields a closed-shell singlet state 6.89 eV above the triplet, while the GVB wavefunction puts the singlet 0.04 eV below the triplet. The problem with the HF description for these systems is that separate singly occupied d orbitals as in Fig. 2, a and b, are not allowed. Thus, for Cu_2O_7, the GVB wavefunction has the form

$$\Psi^{\mathrm{GVB}}_{\mathrm{singlet}} = \mathcal{A}[\Phi(\phi_L\phi_R + \phi_R\phi_L)(\alpha\beta - \beta\alpha)] \qquad (4)$$

where ϕ_L and ϕ_R denote $d_{y^2-z^2}$-like orbitals centered mainly on the left and right copper (as in Fig. 2 for Cu_3), Φ contains all other orbitals and spins, and $\mathcal{A}$ is the antisymmetrizer (determinant operator). All orbitals are calculated self-consistently with no restriction on shape, overlap, or character of the orbitals. Similarly, the triplet state is described as

$$\Psi^{\mathrm{GVB}}_{\mathrm{triplet}} = \mathcal{A}[\Phi(\phi_L\phi_R - \phi_R\phi_L)(\alpha\beta + \beta\alpha)] \qquad (5)$$

where all orbitals are solved for self-consis-

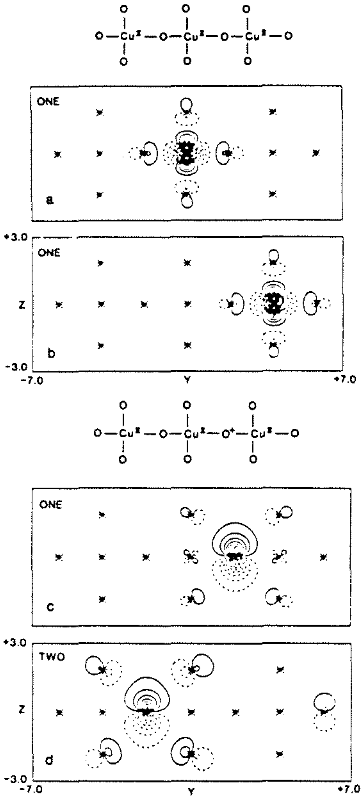


Fig. 2. The GVB orbitals (amplitudes) of Cu_3O_{10} (**a**) and (**b**) show two of the three singly occupied (d-like) orbitals located on the copper centers. Atoms in the plane are indicated with asterisks. Positive contours are solid, whereas negative contours are dotted; the increments are 0.10 atomic units. (**c**) The new singly occupied (oxygen $2p\pi$-like) orbital obtained by ionizing the Cu_3O_{10} cluster. No noticeable change occurs in the other orbitals shown in (a) and (b). (**d**) The corresponding doubly occupied orbital at an adjacent oxygen.

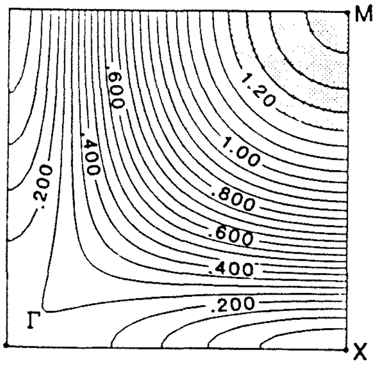


Fig. 3. The upper oxygen $2p\pi$ band for the Cu-O sheets of $La_2Cu_1O_4$ (based on GVB energy band calculations). The contour lines are labeled in electron volts. The Fermi energy for $La_{1.85}Sr_{0.15}Cu_1O_4$ is at the boundary of the dotted and undotted regions.

tently; however, the final orbitals are very similar (indistinguishable in a plot such as Fig. 2) from the optimum orbitals of the singlet. As a result, the Heisenberg-type description (Eq. 1) is suitable. In contrast, the HF functions have the form

$$\Phi^{HF}_{singlet} = \mathcal{A}[\Phi(\phi_g\phi_g)(\alpha\beta - \beta\alpha)] \qquad (6)$$

$$\Phi^{HF}_{triplet} = \mathcal{A}[\Phi(\phi_g\phi_u - \phi_u\phi_g)(\alpha\beta + \beta\alpha)] \qquad (7)$$

where all orbitals are calculated self-consistently. For the optimum triplet state, the final orbitals have the form

$$\phi_g = (\phi_L + \phi_R) \qquad (8)$$

$$\phi_u = (\phi_L - \phi_R) \qquad (9)$$

(ignoring normalization), leading to a wavefunction identical to $\Psi^{GVB}_{triplet}$. However, for the singlet state, the HF wavefunction has the form

$$\phi_g\phi_g = (\phi_L\phi_R + \phi_R\phi_L) + (\phi_L\phi_L + \phi_R\phi_R) \qquad (10)$$

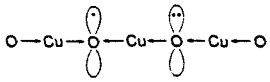
which includes equal amounts of covalent and ionic character. These ionic terms correspond to equal mixtures of Cu^{III}-Cu^{I} character into the Cu^{II}-Cu^{II} wavefunction, leading to an artificially high energy. This HF wavefunction also leads to strong mixtures between the copper $d\sigma$ and oxygen $p\sigma$ orbitals ($d_{y^2-z^2}$ and p_y in Fig. 2). To remove this difficulty for HF wavefunctions, one can relax the spin symmetry restriction and use a wavefunction of the form

$$\Phi^{UHF} = \mathcal{A}[\Phi(\phi_L\alpha)(\phi_R\beta)] \qquad (11)$$

(the unrestricted HF or UHF wavefunction). This leads to much lower energy, but the wavefunction is a mixture of singlet and triplet character and hence one cannot directly obtain parameters for Eq. 1. (The UHF wavefunction leads to singly occupied orbitals that are essentially pure $d_{y^2-z^2}$ in character.) The GVB wavefunction corresponds to converting Eq. 11 into a singlet state (leading to Eq. 4) and then reoptimizing the orbitals self-consistently. In terms of band concepts, the HF description has a half-filled band (ϕ_g occupied and ϕ_u empty), whereas the GVB description has every band orbital half-occupied (ϕ_L and ϕ_R). For the infinite system, localized, singly occupied orbitals such as in Fig. 2 lead to very narrow energy bands. It is well known that such systems are not well described with normal band theory, and an empirical modification (the Hubbard Hamiltonian) is used to obtain UHF-like wavefunctions. For the GVB method, an N-body approach to band calculations is required.

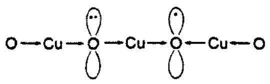
Valence bond band structure—localized versus delocalized states. Consider the Cu_3O_{10} cluster modeling the chains of $Y_1Ba_2Cu_3O_7$. Ionizing this cluster leads to two equivalent localized wavefunctions, Ψ_L and Ψ_R,

Ψ_L



III

Ψ_R



IV

where structure IV is shown in Fig. 2, c and d. The wavefunction is localized because the shape of a doubly occupied orbital needs to be more expanded (electron repulsion) than that of a singly occupied orbital and because in Ψ_L the other bonds polarize to stabilize a charge on the left, whereas in Ψ_R they polarize to stabilize a charge on the right. These two wavefunctions overlap with a sizeable matrix element between them,

$$\sigma_{LR} = \langle\Psi_L|\Psi_R\rangle = -0.1412 \qquad (12)$$

$$H_{LR} = \langle\Psi_L|H|\Psi_R\rangle \qquad (13)$$

In the resonating-GVB description of these states, the total wavefunction is written

$$\Psi_u = (\Psi_L + \Psi_R)/\sqrt{1+S} \qquad (14)$$

$$\Psi_g = (\Psi_L + \Psi_R)/\sqrt{1-S} \qquad (15)$$

We calculate these states using the R-GVB program (*13*), which allows every orbital of Ψ_L to overlap every orbital of Ψ_R. This many-electron resonating generalized valence bond (R-GVB) hole description of the ion states is equivalent to a polaron-hopping description in which the hole is completely dressed with polarization for each site. The R-GVB states in structures III and IV lead to a resonance stabilization of 0.33 eV for Ψ_g and destabilization of 0.43 eV for Ψ_u. This leads to a one-dimensional band with bandwidth = 1.62 eV. As a result, holes in the oxygen $2p\pi$ bonds of the Cu-O chains or Cu-O sheets lead to high mobility charge carriers for electrical conductivity.

N-Electron band theory. We report here a band calculations based on such localized N-electron valence bond wavefunctions rather than the traditional one-electron molecular orbitals (MO) or band orbitals. The N-electron band states Ψ_k are described as a linear combination of localized N-electron valence bond wavefunctions,

$$\Psi_k = C_{1k}\Phi_1 + C_{2k}\Phi_2 + \cdots = \sum_i C_{ik}\Phi_i \qquad (16)$$

where each term Φ_i is an N-electron wavefunction describing a fully polarized description of the hole on a particular site i. Here the coefficients are calculated by solving

$$\sum_j H_{ij}C_{jk} = \sum_j S_{ij}C_{jk}\lambda_k \qquad (17)$$

where the H_{ij} are the N-electron matrix elements (*13*) between oxidized sites i and j and S_{ij} is the N-electron overlap for sites i and j, as in Eqs. 12 and 13. This process is analogous to a tight-binding band calculation except that the matrix elements arise from N-electron valence bond calculations, $\langle\Psi_L|H|\Psi_R\rangle$, where H is the total N-electron Hamiltonian rather than from one-electron matrix elements, $\langle\Phi_L|H^{HF}|\Phi_R\rangle$, where H^{HF} is an effective one-particle Hamiltonian. A wavepacket constructed from the band states, Ψ_k in Eq. 16, describes the motion of a fully dressed polaron.

This contrasts with the tight-binding MO description where the coefficients describe a particular one-electron band orbital

$$\psi_k = C_{1k}\phi_1 + C_{2k}\phi_2 + \cdots = \sum_i C_{ik}\phi_i \qquad (18)$$

(in terms of localized atomic-like orbitals ϕ_i), which is occupied along with the other band orbitals to construct a many-electron wavefunction,

$$\Psi^{MO} = \mathcal{A}[(\psi_1)^2(\psi_2)^2\cdots] \qquad (19)$$

For the CuO sheets of $La_2Cu_1O_4$, we calculated the valence band shown in Fig. 3. This represents the Bloch N-electron states for a hole moving in the CuO sheets. The most stable ion state is at the M point (highest state in a standard band calculation). The least stable state is the X point. The total bandwidth is 1.38 eV.

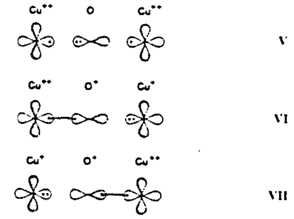
For $La_{2-x}Sr_xCu_1O_4$, the maximum transition temperature T_c occurs for $x = 0.15$. Assuming all these holes go into the oxygen $2p\pi$ band leads to the Fermi energy at 0.16 eV below the top of the band, as indicated in Fig. 3. The density of states at the Fermi energy is $N(0) = 1.14$ eV^{-1} per copper atom. For $Y_1Ba_2Cu_3O_7$ the band arising from the chains would be half full if all holes were in this band. This would lead to $N(0) = 0.21$ eV^{-1} per chain copper. These results are used in the accompanying report (*12*) to derive the T_c for superconductivity in cuprates.

The Hubbard model. As pointed out above, standard HF methods lead to a very bad description of systems such as these Cu^{II} systems having weakly overlapping orbitals. The result is a strong mixture of the singly occupied $d\sigma$ orbitals ($d_{x^2-y^2}$ pointing from Cu to O) with oxygen $p\sigma$ orbitals, leading to a partially filled band of mixed copper $d\sigma$ and oxygen $p\sigma$ character. With GVB, the electron correlation effects lead to singly occupied $d_{x^2-y^2}$ orbitals on each copper. In terms of band concepts this GVB description corresponds to half occupation of every orbital of the band constructed from $d_{x^2-y^2}$ on each center, whereas HF would start with half occupation of the band. The Hubbard approximation (*14*) to UHF builds in a similar improvement upon standard HF band theory by introducing a repulsive one-center term to split the HF band into up-spin and down-spin bands on separate sublattices (leading to spin waves). The GVB approach treats the electron correlation problem rigorously, leading to pure spin states. However, this leads to the complication that the band states must be calculated in terms of N-body wavefunctions (as presented above) rather than the usual one-particle orbitals.

Several standard one-electron band calculations (based on local density functionals) have been reported (*15*) on $La_2Cu_1O_4$. These band calculations suggest an overall population of d^9 on each copper, in agreement with the GVB results but, as expected, they all involve strong mixing of copper $d\sigma$ and oxygen $p\sigma$ character, leading to a partially occupied band of σ character. A properly parameterized Hubbard Hamiltonian might mimic the GVB results (with the copper $d\sigma$ orbitals forming a narrow band and a highest band that is oxygen $p\pi$–like).

REFERENCES AND NOTES

1. J. G. Bednorz and K. A. Muller, *Z. Phys. B* **64**, 189 (1986); J. M. Tarascon, L. H. Greene, W. R. McKinnon, G. W. Hull, T. H. Geballe, *Science* **235**, 1373 (1987).
2. M. K. Wu, *et al.*, *Phys. Rev. Lett.* **58**, 908 (1987).
3. W. J. Hunt, P. J. Hay, W. A. Goddard III, *J. Chem. Phys.* **57**, 738 (1972); W. A. Goddard III and T. C. McGill, *J. Vac. Sci. Technol.* **16**, 1308 (1979).
4. The geometries of these clusters are based on neutron diffraction data (*5*). The Dunning ($9s5p/3s2p$) double zeta contraction (*6*) of the Huzinaga Gaussian basis set is used for the oxygen atoms. The Los Alamos core effective potential (*7*) is used to replace the 18 core electrons of the copper atom. This leads to a ($3s,2p,2d$) Gaussian basis for the 11 valence electrons of the copper atom.
5. P. Day *et al.*, *J. Phys. C* **20**, L429 (1987); W. I. F. David *et al.*, *Nature (London)* **327**, 310 (1987).
6. T. H. Dunning, Jr., and P. J. Hay, in *Modern Theoretical Chemistry* (Plenum, New York, 1977), vol. 3, p. 1.
7. P. J. Hay and W. R. Wadt, *J. Chem. Phys.* **82**, 270 (1985).
8. N. D. Mermin and H. Wagner, *Phys. Rev. Lett.* **17**, 1133 (1966).
9. D. Vaknin *et al.*, *ibid.* **58**, 2802 (1987).
10. J. B. Torrance, Y. Tokura, A. Nazzal, S. S. P. Parkin, in preparation; Y. Tokura, J. B. Torrance, A. I. Nazzal, T. C. Huang, C. Ortiz, in preparation.
11. Superexchange involves mixing of charge transfer states (structures VI and VII) into the dominant wavefunction (top structure V diagram). This charge transfer is much less favorable if the oxygen is made more positive.



12. G. Chen and W. A. Goddard III, *Science* **239**, 899 (1988).
13. A. F. Voter and W. A. Goddard III, *J. Am. Chem. Soc.* **108**, 2830 (1986). The matrix elements as defined in Eq. 13 would depend on cluster size. We use instead

$$\overline{H}_{ij} = E^{N-1}S_{ij}^{N-1} - H_{ij}^{N-1} \tag{20}$$

where E^{N-1} is the energy of the cluster with one hole, and H_{ij}^{N-1} and S_{ij}^{N-1} are for systems with the hole localized on sites i and j, respectively. The sign chosen in $\overline{H}_{ij}$ is so that $E_k = -IP_k$, where IP_k is the ionization potential out of Bloch state k, leading to band diagrams analogous to one-electron band diagrams.
14. J. Hubbard, *Proc. R. Soc. London Ser. A.* **281**, 401 (1964).
15. L. F. Mattheiss, *Phys. Rev. Lett.* **58**, 1028 (1987); K. Takegahara, H. Harima, A. Yanase, *Jpn. J. Appl. Phys.* **26**, L352 (1987); T. Oguchi, *ibid.*, p. L417.
16. We thank the Office of Naval Research and the donors of the Petroleum Research Fund (administered by the American Chemical Society) for partial support of this research. We thank C. M. Kao and G. Chen for assistance and useful discussions. The GVB and R-GVB calculations were carried out on the Alliant FX8/8 and DEC VAX 8650 computers in the Caltech Materials Simulation Facility [funded by the National Science Foundation–Materials Research Groups (grant DMR-84-21119); the Office of Naval Research/Defense Advanced Research Projects Agency (contract N00014-86-K-0735); the Department of Energy–Energy Conversion and Utilization Technology (JPL code 49-242-E0403-0-3550), the National Science Foundation–Chemistry (grant CHE-8318041), and the Office of Naval Research (contract N00014-84-K-0637)].

7 December 1987; accepted 20 January 1988

Appendix 2

Superconducting Properties of Copper Oxide High-Temperature Superconductors

Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 3447–3451, May 1989
Chemistry

Superconducting properties of copper oxide high-temperature superconductors

(YBaCuO/LaSrCuO/magnon pairing/generalized valence bond calculations/Heisenberg coupling term)

GUANHUA CHEN, JEAN-MARC LANGLOIS, YUEJIN GUO, AND WILLIAM A. GODDARD III

Arthur Amos Noyes Laboratory of Chemical Physics, California Institute of Technology, Pasadena, CA 91125

Contributed by William A. Goddard III, February 9, 1989

ABSTRACT The equations for the magnon pairing theory of high-temperature copper-oxide-based superconductors are solved and used to calculate several properties, leading to results for specific heat and critical magnetic fields consistent with experimental results. In addition, the theory suggests an explanation of why there are two sets of transition temperatures ($T_c \approx 90$ K and $T_c \approx 55$ K) for the $Y_1Ba_2Cu_3O_{6+x}$ class of superconductors. It also provides an explanation of why $La_{2-x}Sr_xCuO_4$ is a superconductor for only a small range of x (and suggests an experiment to independently test the theory). These results provide support for the magnon pairing theory of high-temperature superconductors. On the basis of the theory, some suggestions are made for improving these materials.

We recently proposed the magnon pairing mechanism (1) to explain the high-temperature superconductivity in ceramic copper oxide superconductors (2–4). This model was derived from the results of generalized valence bond (GVB) calculations (5) and was used to predict some qualitative features of these systems. With the magnon pairing mechanism (1) we have now calculated several properties of the superconducting phase. The results on specific heat, critical magnetic fields, Hall effect, penetration depth, coherence length, and dependence upon doping generally agree with experiment and, in some cases, explain rather puzzling results. Several predictions are made that could be tested with further experiments.

REVIEW OF THE MODEL

The GVB calculations (5) indicate that the $La_{2-x}Sr_xCu_1O_4$, $Y_1Ba_2Cu_3O_{6+x}$, and $Tl_2Ba_2Ca_{n-1}Cu_nO_{2n+4}$ classes of superconductor oxides (2–4) have the following essential features:

(*i*) $Cu^{II}(d^9)$ sites lie in an essentially square Cu—O sheet having linear Cu—O—Cu bonds with the singly occupied Cu d orbital localized in the plane of four short Cu—O bonds ($R_{CuO} \approx 1.9$ Å).

(*ii*) The spins of adjacent d orbitals are coupled antiferromagnetically (by means of the intervening oxygen) with a Heisenberg coupling term, J_{dd}, ranging from −100 K to −250 K (depending on the system).

(*iii*) Oxidation of the system beyond cupric (Cu^{II}) leads not to Cu^{III} but rather to holes localized in the $p\pi$ (nonbonding) orbitals of oxygens ($O^{2-} \rightarrow O^-$). These $p\pi$ orbitals are localized in the plane containing the short Cu—O bonds to the adjacent copper.

(*iv*) The migration of an oxygen hole from one site to another leads to energy bands with a reasonably high density of states ($N_0 \approx 1.2$ states per eV per sheet of Cu) and to high electrical conductivity.

(*v*) The magnetic coupling of the singly occupied oxygen orbital (the hole) with the two adjacent copper atoms is ferromagnetic, with a Heisenberg coupling term of $J_{OCu} = +330$ to $+400$ K (depending on the system). This leads to ferromagnetic coupling of the spins of the two Cu atoms, despite the antiferromagnetic J_{dd}. Because of the more positive oxygen, the value of J^{+}_{dd} for the Cu atoms adjacent to the hole is about 30% smaller than J_{dd}.

A qualitative view of the magnon pairing model of superconductivity is as follows. Adjacent Cu spins have a tendency to be opposite when separated by a normal O^{2-} but tend to repolarize parallel (ferromagnetic) when separated by an oxygen hole (O^-). As the conduction electron (O^- hole) moves from site to site, it tends to leave behind a wake where adjacent copper spins are ferromagnetically paired. As a second conduction electron is scattered, it is favorable to be scattered into the wake of the first electron, since there is already ferromagnetic polarization of the copper spins, leading to a favorable interaction. The net result is an attractive interaction between conduction electrons which leads to superconductivity. In the next section we outline the approach used to calculate the quantitative aspects of superconductivity.

THE SUPERCONDUCTING STATE

Interaction Potential. For a two-dimensional CuO sheet with Cu–Cu distances of a and b in the x and y directions, the lowest-order interaction between two O$p\pi$ holes is given by

$$V_{\bar{\mathbf{k}}\downarrow-\bar{\mathbf{k}}\uparrow,\mathbf{k}\uparrow-\mathbf{k}\downarrow} = -P(\mathrm{Q})W, \qquad [1]$$

where $W = J^2_{pd}/\Delta E$,

$$P(\mathrm{Q}) = 1 + \frac{1}{2}\cos \mathrm{Q}\cdot\mathbf{a} + \frac{1}{2}\cos \mathrm{Q}\cdot\mathbf{b} + 2\cos\left(\frac{1}{2}\mathrm{Q}\cdot\mathbf{a}\right)\cos\left(\frac{1}{2}\mathrm{Q}\cdot\mathbf{b}\right),$$

and $\mathrm{Q} = \mathbf{k} - \bar{\mathbf{k}}$. In deriving Eq. **1**, we use the random phase approximation for the Cu spins and write the average excitation energy for a Cu spin-flip as $\Delta E = 8\tau\beta|J_{dd}|$. Here τ is related to the average spin correlation between adjacent sites (1), and $\beta = (J^{+}_{dd} + J_{dd})/2J_{dd} \approx 0.85$ accounts for the decrease in $|J_{dd}|$ when the neighboring oxygen is O^- rather than O^{2-}.

Coupling the ($\mathbf{k}$, $-\mathbf{k}$) pair into a triplet state leads to an attractive net interaction potential of

$$V_{\mathbf{kk}} = -[P(\mathrm{Q}) - P(\mathrm{Q}')]W, \qquad [2]$$

where $\mathrm{Q}' = \mathbf{k} + \bar{\mathbf{k}}$.

Factorization of the Interaction Potential. To solve the gap equation $\Delta_{\mathbf{k}} = -\sum_{\bar{\mathbf{k}}} V_{\mathbf{k}\bar{\mathbf{k}}}\Delta_{\bar{\mathbf{k}}}(1 - 2f_{\bar{\mathbf{k}}})/2E_{\bar{\mathbf{k}}}$, where $E_{\mathbf{k}} = (\xi^2_{\mathbf{k}} +$

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: GVB, generalized valence bond; BCS, Bardeen–Cooper–Schrieffer.

$|\Delta_k|^2)^{1/2}$ and $f_k = 1/(e^{\beta E_k} + 1)$, we decompose the Q-dependent factor in Eq. 2 to obtain

$$V_{k\bar{k}} = -W \sum_{i=1}^{4} \lambda_i \Psi_i(\mathbf{k}) \Psi_i(\bar{\mathbf{k}}). \qquad [3]$$

This leads to the new gap equation

$$g_i = \sum_{j=1}^{4} B_{ij} g_j, \qquad [4a]$$

where

$$B_{ij} = W\lambda_i \sum_{\mathbf{k}} \Psi_i(\mathbf{k})\Psi_j(\mathbf{k}) \frac{(1 - 2f_k)}{2E_k} \qquad [4b]$$

and

$$\Delta_k = \sum_{i=1}^{4} g_i \Psi_i(\mathbf{k}). \qquad [4c]$$

For each temperature we solve Eq. 4 iteratively until self-consistency is achieved.

We find two sets of solutions to Eq. 4. Assuming the gap to be real, we find a self-consistent solution of the form $\Delta_k^R = g(\Psi_1 \pm \Psi_2)$, leading to a gap that goes through zero (Ψ_3 and Ψ_4 are small compared with Ψ_1 and Ψ_2). On the other hand, allowing a complex gap leads to a self-consistent solution of the form $\Delta_k^C = g(\Psi_1 \pm i\Psi_2)$, where $|\Delta_k^C|$ depends on k but never goes to zero. In each system, the complex gap solution leads to an energy stabilization about 20% greater than for the real gap, and we report properties based only on Δ_k^C. It is possible that near T_c and near surfaces or grain boundaries the solutions would be more complicated.

Transition Temperature. As T approaches T_c, the gap and hence the g_i must go to zero. This leads to $B_{ij} = \delta_{ij}$ and hence to

$$T_c = 1.13\, \alpha |J_{dd}| \exp\left[-\frac{8\tau\beta|J_{dd}|}{\lambda N_0 (J_{pd})^2}\right], \qquad [5]$$

where $\lambda = \lambda_1 = \lambda_2 >> \lambda_3 = \lambda_4$, N_0 is the density of electron states, and αJ_{dd} is the effective range of the $K_\perp$ integration (perpendicular to the Fermi surface). The upper bound on α is 4, and we estimate $\alpha \approx 2$.

RESULTS

Using Eq. 4, we solved numerically to obtain the $\Delta(\mathbf{k})$ at various temperatures for three classes of systems,

$La_{2-x}Sr_xCu_1O_4$ (denoted 2-1-4),
$Y_1Ba_2Cu_3O_{6+x}$ (denoted 1-2-3),

and

$(TlO)_2Ba_2Ca_{n-1}Cu_nO_{2n+2}$ (with n layers of Cu—O sheets, denoted Tl-n layer),

and used these results to calculate specific heats, critical fields, and other properties at various temperatures.

The 2-1-4 system involves six-coordinate Cu sites (with apex oxygens above and below the Cu—O sheets), while the 1-2-3 and Tl-n systems (with $n \geq 2$) involve five-coordinate sites with apex oxygen on only one side of the Cu—O sheet. In all cases we used the experimental crystal structure and carried out the GVB calculations as in ref. 5, with Cu_2O_p clusters (p = 11, 9, and 9 for 2-1-4, 1-2-3, and Tl-2, respectively) using an array of additional ions (216, 200, and 299, respectively) to represent the electrostatic field of the crystal. These calculations were carried out for two charge states to yield J_{dd} for the neutral and J_{OCu} and J_{dd}^{+} for the positive ion, as in Table 1. The electron transfer matrix elements were calculated as in ref. 5 to obtain similar band structures for all three systems, leading to the density of states N_0. We used $\alpha = 2$ (see above) for all systems. The λ comes from Eq. 3 but depends sensitively on the concentration of holes in the Cu—O sheets (x_s), as discussed in the next section.

The average spin correlation of adjacent Cu spins, τ, is difficult to calculate because of the dynamic nature of the coupling between the oxygen and copper spins. Given all other parameters, the value of τ needed for Eq. 5 to yield T_c = 93 K for $Y_1Ba_2Cu_3O_7$ is 0.0167. Using $\tau = 0.0167$ for $La_{1.85}Sr_{0.15}CuO_4$ would lead to T_c = 32.6 K, in reasonable agreement with the observed T_c = 37 K. This close agreement provides support for the overall magnon pairing mechanism; however, in the balance of this paper we will use the value $\tau = 0.0159$ for $La_{2-x}Sr_xCuO_4$ (but will assume τ to be independent of x), adjusted to yield T_c = 37 K at $x = 0.15$ ($x_s = 0.10$, see below).

In general, for Bi-, Tl-, and Pb-containing materials (6), $(AO)_mM_2Ca_{n-1}Cu_nO_{2n+2}$, the $n = 1$ case leads to six-coordinate Cu—O sheets (as in 2-1-4) and the four known such systems lead to T_c = 0, 12, 50, and 90 K, whereas the $n \geq 2$ systems have two five-coordinate Cu—O sheets (as in 1-2-3, with additional four-coordinate Cu—O sheets for $n > 2$) and the five known $n = 2$ systems lead to T_c = 90, 90, 90, 90, and 110 K, while the four known $n = 3$ systems lead to T_c = 110, 110, 122, and 122 K, and the $n = 4$ system leads to T_c = 122 K. We calculated only the $n = 2$ case but presume the results to be relevant for $n \geq 2$.

Using the calculated parameters (J_{dd}, J_{OCu}, N_0, β) for the

Table 1. Quantities used in calculating superconducting properties

Parameter	$La_{1.85}Sr_{0.15}Cu_1O_4$	$Y_1Ba_2Cu_3O_7$	$Tl_2Ba_2Ca_1Cu_2O_8$
J_{dd}, K	−204.7	−94.0	−175.4
J_{OCu}, K	383.2	330.1	396.0
β*	0.841	0.860	0.872
x_s	0.10	0.25	(0.25)†
λ	0.558	1.17	1.17
N_0‡	1.23	1.19	1.42
α	(2.0)§	(2.0)§	(2.0)§
T_c (τ = 0.0159), K	(≈37.0)	≈97	≈167
T_c (τ = 0.0167), K	≈33	(≈93)	≈160¶

*$\beta = (J_{dd} + J_{dd}^{+})/2J_{dd}$.

†Assumed, based on $Y_1Ba_2Cu_3O_7$.

‡Units are eV^{-1} per sheet Cu atom.

§Assumed.

¶Using $\lambda = 0.91$ (for $x_s = 0.18$) and all other quantities the same leads to T_c = 123 K.

Tl-2 layer system and assuming other parameters estimated as for $Y_1Ba_2Cu_3O_7$ (x_s = 0.25, τ = 0.0167) leads to T_c = 160 K for Tl-2. This is in reasonable agreement with experiment, providing additional support for the magnon pairing model. We believe that the fluctuations in T_c for these Tl-*n* systems are due to variations in x_s and suggest that x_s = 0.25 (the same value as for 1-2-3) would yield $T_c \approx 160$ K for Tl-*n* (with $n \geq 2$). The highest observed T_c of 122 K suggests x_s = 0.18 for the Tl-*n* systems ($n \geq 2$).

Dependence on Doping. The number of holes in the Cu—O sheets x_s is critical to the superconductivity. The valence band maximum is at the M point, $(\pi/a, \pi/b)$, and thus as $x_s \rightarrow 0$, the Fermi surface collapses around the M point. From Eq. 1 this leads to $V_{\text{trip}} \rightarrow 0$ and hence to $\Delta \rightarrow 0$ as $x_s \rightarrow 0$.

1-2-3 system. For $Y_1Ba_2Cu_3O_{6+x}$, Hall measurements (7) have been reported for various *x* and used to estimate x_H, the concentration (per cell) of carriers contributing to the conductivity. We assume that hole in the chains (which are disordered and incomplete) cannot contribute significantly to x_H, so that the Hall coefficient measures the concentration of holes per sheet (x_s), $x_H = 2x_s$. For *x* = 1, the experiments lead to $x_s \approx 0.25$ and T_c = 92 K, while for smaller *x*, the experiments lead to smaller x_s and decreased T_c. In Fig. 1 we show the predicted T_c versus x_s (using the same values for *J*, N_0, τ, α, and β but allowing λ to change with x_s) and compare with the experimental data. The close correspondence provides support for the magnon pairing model. It is important to note that such a rapid change of T_c with x_s would not be expected for a simple Bardeen–Cooper–Schrieffer (BCS) system (singlet pairs rather than triplet pairs).

The observation that the $Y_1Ba_2Cu_3O_{6+x}$ system tends to have either a high T_c around 93 K or a lower T_c around 60 K we associate with the special stabilization of x_s = 0.25 for *x* near 1 (because holes in the oxygen chain cannot get closer than alternate sites) and x_s = 0.125 for *x* near 0.5 (again related to the capacity for chains to carry holes). Assuming all calculated parameters except x_s (and hence λ) are unchanged, the theory predicts $T_c \approx 45$ K for x_s = 0.125, while T_c = 93 K for x_s = 0.25.

2-1-4 systems. In Fig. 2 (solid line) we show how the predicted T_c varies with x_s for the 2-1-4 system. These results are in agreement with experiment for $x < 0.1$ but not for higher values. However, as predicted earlier (5), the GVB calculations lead to a relative ionization potential (IP) from the sheet O versus the apex O that varies linearly with doping, $IP_a - IP_s = 0.38 - (0.38/0.13)x$, so that for $x > 0.13$, the most stable location of the hole is predicted to be at the apex oxygen, not at the sheet oxygen. (This occurs because $La^{3+} \rightarrow Sr^{2+}$ leads to a less attractive potential near the apex oxygen.) This suggests that for $x > 0.1$, an increasing fraction of the holes are on the apex oxygens rather than the sheet oxygens. At atmospheric pressure, oxygen vacancies are observed for $x > 0.15$. We suggest that this is because a high concentration of holes on the apex oxygens favors oxygen vacancy formation.

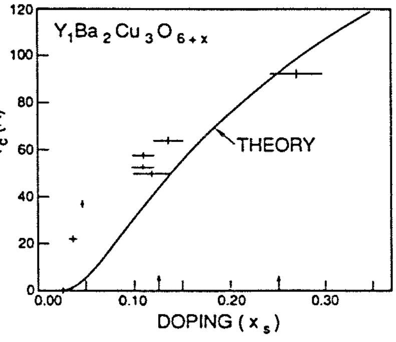


FIG. 1. Predicted dependence (solid line) of T_c on x_s for 1-2-3. Experimental results from Hall effect measurements (7) are shown as crosses with magnitude representing experimental errors.

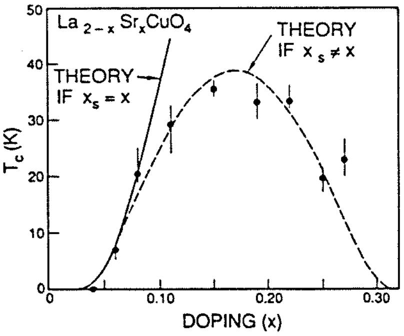


FIG. 2. The solid line shows the predicted dependence of T_c on x_s for 2-1-4, while the broken line shows the predicted T_c versus *x* using the relation between *x* and x_s from the text. Experimental T_c values are given versus *x* (x_s has not been measured).

In Fig. 2 (broken line) we show the T_c predicted from magnon pairing theory, assuming that $x_s = x$ for $x \leq 0.06$, $x_s = 0$ for $x \geq 0.34$, with x_s/x varying linearly between these limits (assuming no vacancy formation). This approximates the variation of x_s with *x* expected if $IP_a = IP_s$ at $x = 0.17$ (calculated crossing point at $x = 0.13$). As indicated in Fig. 2, the predictions are in good agreement with experiment (8), explaining the experimental results that $T_c \approx 0$ K for $x < 0.06$ and for $x > 0.3$. These predictions could be tested independently by measuring the number of holes (x_s) contributing to the Hall current as a function of *x*. This is a difficult experiment because apex holes also contribute to the normal conductivity for the 2-1-4 system.

Specific Heat. We calculated the electronic specific heat for the 2-1-4 system and found a change at T_c of $(\Delta C_s/T_c)$ = 8.4 mJ/mol·K^2 and an electronic specific heat for the normal state of $\gamma_e = 5.79$. Combining these results leads to $\Delta C_s/\gamma_e T_c$ = 1.45, which is close to the value for BCS theory, $\Delta C_s/\gamma_e T_c$ = 1.43.

A direct measurement of $\Delta C/T_c$ was made by Batlogg *et al.* (9), using quasi-adiabatic methods. They found a specific heat gap of $\Delta C/C$ = 1% at T_c and $(\Delta C/T_c)_{\exp} = 7.6 \pm 1.8$ mJ/mol·K^2, in excellent agreement with theory. However, we should point out that the experimental situation is not settled. Similar experiments by Maple *et al.* (10) revealed no transition in ΔC, while similar experiments by Nieva *et al.* (11) led to $\Delta C/C$ = 8%. This latter experiment leads to a best estimate of $\gamma_e = 18 \pm 2$ mJ/mol·K^2.

Critical Magnetic Fields. Above the critical magnetic field, H_c, the free energy of the normal state is lower than that of the superconducting state (because of the Meissner effect). Assuming a magnetic field in the *c* direction (perpendicular to the Cu—O sheet), we calculated H_c as a function of *T*. For the 1-2-3 system these calculations lead to $H_c(0)$ = 0.74 T and $(dH_{c2\perp/d})_{T_c} = -0.80$ T/K (where T denotes tesla).

To compare with experiment, it is necessary to estimate the magnetic penetration length λ_L and the coherence length ξ. Using the London formula $\lambda_L = (m^*c^2/4\pi n_s e^2)^{1/2}$, the calculated effective mass (m^*/m_e = 1.9 for 2-1-4) and assuming n_s to be the number of O holes in the sheets, we obtain λ_L = 1400 Å from 1-2-3 (x_s = 0.25) and λ_L = 2200 Å for 2-1-4 (x_s = 0.10). To obtain the coherence length ξ, we used the calculated $H_c(0)$, the above value for λ_L, and the Ginzburg–Landau relation for $H_c(0) = \phi_0/(2\pi\sqrt{2}\,\xi\lambda)$. The result is ξ = 23 Å for

1-2-3 and $\xi = 38$ Å for 2-1-4. The conversion to the experimentally measured critical field H_{c2} is $H_{c2} = \sqrt{2}\, H_c\lambda/\xi$.

The only measurement (on single crystals) with the field perpendicular to the CuO plane is by Worthington *et al.* (12) for 1-2-3. For a sample with $T_c = 87$ K, they found

$$\left(\frac{dH_{c2\perp}}{dT}\right) = -0.71 \text{ T/K} \quad \text{for} \quad T < 78 \text{ K}$$

and

$$\left(\frac{dH_{c2\perp}}{dT}\right) = -0.46 \text{ T/K} \quad \text{for} \quad 78 < T < 87.$$

These values are in reasonable agreement with our predicted value of −0.80, providing support for the current model. The high-temperature knee in H_c versus T at $T \approx 78$ K is not understood. It might be an experimental artifact (the measured $T_c = 87$ K is below the usual value of $T_c = 93$ K). It might result from a sensitivity to surface regions where the properties may be different. For example, the calculated slope depends on x_s, which might be different near the surface. [For the real gap solution, we calculate $(dH_{c2\perp}/dT) = -0.65$ T/K.]

Magnitude of J_{dd}. There are no directly determined values of J_{dd} for the CuO systems. However, accurate experiments have been reported on K_2NiF_4, which has the same layered structure as $La_2Cu_1O_4$. The Neél temperature (13†) for K_2NiF_4 is $T_N = 97$ K, and neutron scattering leads to an (in-plane) Heisenberg coupling term of $J_{dd} = -52$ K (13) and −56 K (14†) [the out-of-plane value is less than 0.2 K (14)].

To determine the accuracy of our calculations for J_{dd}, we carried out the same GVB calculations on Ni_2F_{11} clusters, leading to $J_{dd} = -51$ K. This is in good agreement with experimental results, suggesting that the exchange integrals calculated with GVB are accurate (to about 10%).

Estimates have previously been made for J_{dd} of the CuO systems. Lyons *et al.* (15†, 16†), using Raman light scattering, found an inelastic peak at 0.37 eV for $La_2Cu_1O_4$ and 0.32 eV for $Y_1Ba_2Cu_3O_6$ (both semiconductors, not superconductors). They found that this peak rapidly disappears as the systems are doped ($x > 0$) and interpreted this inelastic transition as a double Cu spin-flip. Linear magnon theory leads to $\Delta E = 5.4\, J_{dd}$ for this process, suggesting $J_{dd} \approx -790$ K for 2-1-4 and $J_{dd} \approx -680$ K for 1-2-3. There is no direct evidence showing that the observed process involves the Cu spins, and we believe that the large discrepancy with the calculated J_{dd} values argues against this interpretation. We suggest that the undoped system may have a small number of oxygen vacancies, leading to extra electrons in the system, which would lead to local Cu^{I} (d^{10}) sites. From GVB calculations on the clusters used for 2-1-4 but with one O vacancy, we calculate the electron transfer process Cu^{II}—Cu^{I} to Cu^{I}—Cu^{II} to have an excitation energy of 0.4 eV and suggest that the transition observed in Raman light scattering is associated with such electron transfer excitations. Small amounts of doping would remove the extra electron from these Cu^{I} sites, leading to disappearance of the peak with small x, as observed. For the 2-1-4 system this might be testable directly by experiments at high O_2 pressure, which might decrease the number of oxygen vacancy sites and by our suggestion lead to the disappearance of the 0.4 eV peak near $x = 0$.

Maximum T_c. In Eq. 5, the maximum T_c would occur if the term in the exponent were zero, leading to $T_c^{max} = 1.13\, \alpha J_{dd}$. However, the derivation of this equation includes only the lowest-order interactions and is not valid for the limit of very large coupling. To estimate the form for T_c at stronger coupling, one can use (17) $T_c = [\langle\omega\rangle/1.20]\exp[-1.04(1 + \eta)/\eta]$, valid for $\eta \approx 1$, where in our case $\eta = \lambda N_0(J_{pd})^2/8\tau\beta|J_{dd}|$. The upper limits are $\langle\omega\rangle < 4J_{dd}$ (for $\alpha = 4$) and the exponential term $< e^{-1}$. This leads to $T_c < 1.23\, J_{dd}$ (Eq. 5 would give the same result if $\alpha \rightarrow 1.09$ as $\eta \rightarrow \infty$). For $\eta >> 1$, the correct formula (17) is $T_c = 0.18(\langle\omega^2\rangle\eta)^{1/2}$, where $\langle\omega^2\rangle = 2/\eta \int \alpha^2F(\omega)\, \omega\, d\omega$; estimating the integral in $\langle\omega^2\rangle$ using various forms for $F(\omega)$ leads to $T_c < J_{dd}$. Using $T_c^{max} \approx 1.23\, J_{dd}$ with our calculated values of $J_{dd} \leq 205$ K leads to $T_c^{max} < 250$ K. (We consider this to be an upper bound; a more conservative estimate of $T_c^{max} < J_{dd} \approx 200$ K given in ref. 1 is probably closer to the real limit.) This provides hope that the recent enormous advances in increasing T_c may continue (currently the highest confirmed T_c is ≈125 K). On the other hand, it suggests that the current class of systems based on CuO sheets will not achieve room temperature T_c.

Non-Cu—O High-Temperature Systems. The ceramic $Ba_{1-x}K_xBiO_3$ shows superconductivity with $T_c = 30$ K for $x = 0.4$ (18, 19). The magnon pairing mechanism has nothing to say about this system since there are no localized magnetic atoms. Systems such as $BaBiO_3$ with a formal Bi^{IV} oxidation state tend to disproportionate to a mixed valence state $Ba_2Bi^{III}Bi^{V}O_6$, with clearly defined (20) Bi^{III} sites [larger cavity because of the $(6s)^2$ pair of valence electrons] and Bi^{V} sites (small cavity). Upon doping with K, such an ordered arrangement is not possible (the Bi atoms are equivalent). However, it is plausible that conduction involves hopping of the electrons between Bi^{III} and Bi^{IV} sites and/or between Bi^{IV} to Bi^{V}, either of which should couple strongly to lattice vibrations. Thus we believe that this system involves lattice coupling much as in BCS theory.

NEW HIGH T_c MATERIALS

Cu—O Type. Among compounds with CuO layers, some are superconducting and some are not (21). We suggested (1, 5) that the major issue here is arranging the cations so that the oxygen holes are in sheets rather than in apex or other locations. Thus, systems that are not superconducting might be converted into superconductors by rearranging the cations (varying the ionic radius and charge). Indeed, we believe that the critical issue in all current materials is arranging the environment of cations so as to maximize x_s. For example, the magnon pairing theory suggests that if the $La_{2-x}Sr_xCuO_4$ system could be modified to have $x_s = 0.25$ (leading to $\lambda = 1.17$, but leaving all other parameters the same), then T_c would increase to 139 K!

A second approach would be to eliminate all oxygens not in the sheets in such a way that the holes must remain in the sheets. For example, if the apex O^{2-} were all replaced by F^-, the holes would be expected to lead to O^- in the sheets (rather than F radicals in the apex position), allowing $x_s = x$. (However, if some F ions end up in the sheet, the superconducting properties might be much worse.) Such a system might be designed by finding a structure in which cations near the apex positions would prefer F to O.

Cu—X Type. To replace the O with other anions X, we believe that it is important to retain the CuX_2 sheets with linear Cu—X—Cu bonds. To obtain the highest possible T_c, the equation for T_c^{max} suggests maximizing $|J_{dd}|$. Based on the Neél temperatures for the MnO, MnS, and MnSe systems, we expect $|J_{dd}|$ to be about 40% larger for S or Se in place of O. Thus the limiting temperature might be $T_c^{max} \approx 350$ K for Cu—S sheets rather than $T_c^{max} \approx 250$ K for Cu—O sheets (bear in mind that these are upper limits). There are, however, many other factors to consider. Most important, the Cu—S sheets must be two-dimensional (with weak coupling between adjacent layers) and must yield S $p\pi$ holes upon doping. A three-dimensional system with strong coupling between Cu spins in different layers would lead to long-

---

†The J in this paper is defined as twice our value.

range order in the Cu spins (τ large), making it unfavorable for the conduction electrons (S $p\pi$ holes) to flip the Cu spins. Thus one would need layered systems analogous to the 2-1-4, 1-2-3, or Tl-n systems. We know of no such examples *involving S or Se. In addition, the structure must be stable* for large concentrations of holes (x_s). The ferromagnetic coupling J_{SCu} is likely to be smaller than J_{OCu} (because of the longer bond length), so that it may be more difficult to achieve the limiting temperature for S systems.

Other Metals. To find replacements for Cu, there are several factors to consider. First, any such replacement will probably lead to a longer M—O bond and hence a smaller J_{OM}. Other things being equal, a large J_{dd} is favored by both small bond distance and more electropositive metals, suggesting Sc–Cr. However, for metals with n unpaired d electrons (e.g., Cu^{2+} has $n = 1$, Ni^{2+} has $n = 2$, Mn^{2+} has $n = 5$), the J_{dd} decreases as n increases. Thus, these criteria favor d^1 and d^9 systems, making Sc^{2+}, Ti^{3+}, V^{4+}, or Cr^{5+} reasonable possibilities. However, for a d^1 system in a pseudooctahedral environment, the three t_{2g}-like orbitals are of similar energy, making such systems less favorable (since d_{xy} must be stabilized). In addition to Cu^{2+}, potential d^9 systems would include Ag^{2+} and Au^{2+}; however, considerations of size, electronegativity, and spin-orbit coupling all seem to favor Cu. Thus these considerations suggest that Cu^{2+} is the optimal choice for the metal.

SUMMARY

The agreement with experiment for various properties predicted by using the magnon pairing model of superconductivity provides strong support for the validity of this model for the Cu—O systems. All quantities are related to the fundamental parameters of the system (J_{dd}, J_{OCu}, band structure). Some approximations have been made in the solutions to these equations. Nevertheless, the fundamental parameters are well defined, and hence improved calculational approximations will eventually lead to precise predictions of all properties. In this theory, the superconducting properties are related to fundamental structural, chemical, and physical properties, allowing one to use qualitative reasoning in contemplating how to improve the properties.

Note Added in Proof. After submission of this paper, a news report (22) appeared stating that Z. Kąkol, J. Spałek, and J. Honig (Purdue University) have evidence suggesting that $\mathrm{La}_{2-x}\mathrm{Sr}_x\mathrm{NiO}_4$ might be superconducting, with a T_c between 4 and 70 K. We have not calculated the properties for this system, but assuming that holes lead to $\mathrm{O}^{2-} \rightarrow \mathrm{O}^-$ (as in Cu) rather than to $\mathrm{Ni}^{2+} \rightarrow \mathrm{Ni}^{3+}$ and assuming $J_{dd\mathrm{Ni}} = \frac{1}{2} J_{dd\mathrm{Cu}} = -100$ K (see above) with all other parameters (including x_s and τ) as for $\mathrm{La}_{1.85}\mathrm{Sr}_{0.15}\mathrm{Cu}_1\mathrm{O}_4$, we predict $T_c = 66$ K for $\mathrm{La}_{1.85}\mathrm{Sr}_{0.15}\mathrm{NiO}_4$. Thus it is plausible that the Ni system leads to superconductivity *if* the holes go on the NiO_2 sheets. Because of the lower J_{dd}, the maximum T_c for a NiO_2 system should be about half of that for a CuO_2 system (i.e., $\approx$100 K instead of $\approx$200 K).

In addition, a second report (23) shows that $\mathrm{Nd}_{1.85}\mathrm{Ce}_{0.15}\mathrm{Cu}_1\mathrm{O}_4$ is a superconductor ($T_c = 24$ K) but that conduction is dominated by electrons rather than holes. GVB calculations on this system lead to $J_{dd} = -137$ K for the undoped system. For the doped system (extra electron), the GVB calculations lead to a resonating state involving Cu^{I} (d^{10}) and $\mathrm{Cu}^{\mathrm{II}}$ (d^9), with a nearest-neighbor resonance stabilization of 0.32 eV. This leads to a description of conduction involving a heavy magnon (one down-spin, several up-spin) whose motion is impeded by the antiferromagnetic coupling. This leads to magnon-mediated attractive coupling between the heavy magnons, much as for oxygen holes. However, we have not yet succeeded in estimating the T_c.

This research was supported by the Office of Naval Research with assistance from the Donors of the Petroleum Research Fund, administered by the American Chemical Society. The calculations were carried out on the Alliant FX8/8 computer and also on a DEC VAX 8650 computer. These computer facilities were provided by the Defense Advanced Research Projects Agency/Office of Naval Research, National Science Foundation (Division of Materials Research, Materials Research Groups), Department of Energy/Energy Conversion and Utilization Technologies, and the National Science Foundation (Division of Chemistry). This is contribution no. 7881 from the Arthur Amos Noyes Laboratory of Chemical Physics.

1. Chen, G. & Goddard, W. A., III (1988) *Science* 239, 899–902.
2. Bednorz, J. G. & Müller, K. A. (1986) *Z. Phys. B* 64, 189–193.
3. Wu, M. K., Ashburn, J. R., Torng, C. J., Hor, P. H., Meng, R. L., Gao, L., Huang, Z. J., Wang, Y. Q. & Chu, C. S. (1987) *Phys. Rev. Lett.* 58, 908–910.
4. Sheng, Z. Z. & Hermann, A. M. (1988) *Nature (London)* 332, 55–58.
5. Guo, Y., Langlois, J.-M. & Goddard, W. A., III (1988) *Science* 239, 896–899.
6. Sleight, A. W. (1988) *Science* 242, 1519–1527.
7. Wang, Z. Z., Clayhold, J., Ong, N. P., Tarascon, J. M., Greene, L. H., McKinnon, W. R. & Hull, G. S. (1987) *Phys. Rev. B* 36, 7222–7225.
8. Torrance, J. B., Tokura, Y., Nazzal, A. I., Bezinge, A., Huang, T. C. & Parkin, S. S. P. (1988) *Phys. Rev. Lett.* 61, 1127–1130.
9. Batlogg, B., Ramirez, A. P., Cava, R. J., van Dover, R. B. & Rietman, E. A. (1987) *Phys. Rev. B* 35, 5340–5342.
10. Maple, M. B., Yang, K. N., Torikachvili, M. S., Ferreira, J. M., Neumeier, J. J., Zhou, H., Dalichaouch, Y. & Lee, B. W. (1987) *Solid State Commun.* 63, 635–639.
11. Nieva, G., Martinez, E. N., dela Cruz, F., Esparza, D. A. & D'Ovidio, C. A. (1987) *Phys. Rev. B* 36, 8780–8782.
12. Worthington, T. K., Gallagher, W. J. & Dinger, T. R. (1987) *Phys. Rev. Lett.* 59, 1160–1163.
13. Birgeneau, R. J., Als-Nielsen, J. & Shirane, G. (1977) *Phys. Rev. B* 16, 280–292.
14. Birgeneau, R. J., Skalyo, J. & Shirane, G. (1970) *J. Appl. Phys.* 41, 1303–1310.
15. Lyons, K. B., Fleury, P. A., Remeika, J. P., Cooper, A. S. & Negran, T. J. (1988) *Phys. Rev. B* 37, 2353–2356.
16. Lyons, K. B., Fleury, P. A., Schneemeyer, I. F. & Waszczak, J. V. (1988) *Phys. Rev. Lett.* 60, 732–735.
17. Allen, P. B. & Dynes, R. C. (1975) *Phys. Rev. B* 12, 905–922.
18. Cava, R. J., Batlogg, B., Krajewski, J. J., Farrow, R., Rupp, L. W., Jr., White, A. E., Short, K., Peck, W. R. & Kometani, T. (1988) *Nature (London)* 332, 814–816.
19. Jin, S., Tiefel, T. H., Sherwood, R. C., Ramirez, A. P., Gyorgy, E. M., Kammlott, G. W. & Fastnacht, R. A. (1988) *Appl. Phys. Lett.* 53, 1116–1118.
20. Cox, D. E. & Sleight, A. W. (1976) *Solid State Commun.* 19, 969–973.
21. Torrance, J. B., Tokura, Y., Nazzal, A. & Parkin, S. S. P. (1988) *Phys. Rev. Lett.* 60, 542–545.
22. Pool, R. (1989) *Science* 243, 741.
23. Tokura, Y., Takagi, H. & Uchida, S. (1989) *Nature (London)* 337, 345–347.