NEW METHODS FOR ELECTRONIC STRUCTURE CALCULATIONS ON LARGE MOLECULES

Richard A. Friesner

Department of Chemistry, Columbia University, New York, NY 10027

INTRODUCTION

The equations of ab initio quantum chemistry—the clamped-nuclei Schroedinger or Dirac-Fock equations—have been known for nearly 60 years. Initially, accurate solutions could only be obtained for the simplest systems, e.g. the hydrogen atom, helium atom, and hydrogen molecule. The mathematical complexity of these ab initio equations for many-electron atoms and molecules led chemists to focus on construction of approximate, semiempirical molecular orbitals [e.g. Huckel theory (1)], an approach that has contributed (and continues to contribute) enormously to our understanding of chemical phenomena.

Forty years ago, the development of digital computers revived the hope that one might obtain molecular wavefunctions from first principles calculations. On the theoretical side, the work of Hartree (2, 3), Fock (4), and others led to a systematic hierarchy of approximation schemes based on a self-consistent field (SCF) or Hartree-Fock (HF) methodology as a starting point; the SCF solutions were then corrected for electron correlation by various techniques. Somewhat later, the development of density functional (DF) theory by Kohn, Hohenberg, and Sham (5, 6) provided a different approach to reducing the electronic Schroedinger equation to

341

tractable form. These two formalisms remain the alternatives if one pursues a first principles method.

The early digital computers had only a small fraction of the memory and central processing unit (CPU) capabilities of current machines. Thus, it was extremely difficult to make progress in developing numerical algorithms for the solution of these complicated partial differential equations. Over the first two decades, a dominant computational algorithm emerged for HF and related methods, which we refer to as conventional quantum chemistry (CQC). The critical component of CQC is expansion of the self-consistent molecular orbitals in a Gaussian basis set, which allows the resultant two electron repulsion integrals that appear in the Roothaan-Hall (RH) equations to be computed analytically. The refinement of this approach, by the use of contracted Gaussian basis functions (7, 8) remedied a major deficiency of the original method, namely the unsuitability of a single Gaussian function for representing atomic or molecular orbitals, particularly tight core orbitals. The announcement of the GAUSSIAN 70 program system (9), which contained the first efficient methods for computing integrals over the contracted Gaussians, initiated a long period of CQC research in which the basic theoretical and numerical framework was fixed and enormous efforts were invested in software and algorithm design, both for HF and correlated methods. There are now several large, complex codes of more than 100,000 Fortran statements (e.g. GAUSSIAN 90) that provide access to the latest CQC technology for nonspecialists in a relatively painless fashion.

In some ways, this research program has been extremely successful. Over the past 20 years the explosive growth in the power of computational hardware has allowed CQC treatments of small- and medium-sized molecules at increasing levels of accuracy. For small (2–6 atom) molecules, genuine predictive power for a wide variety of properties is generally within reach (albeit at a substantial cost for the extensive inclusion of electron correlation), whereas for medium-sized (7–20 atom) molecules, insightful studies can often be performed. Significant problems remain even for the small systems, e.g. accurate treatment of transition metals, excited states, or truly multiconfigurational transition states is far from trivial. However, these difficulties do not appear insuperable, and continued improvements in hardware, software, and CQC methodology are expected to bring most remaining obstacles under control.

The treatment of larger systems by CQC techniques is in a much less satisfactory state. The principal barrier has been the scaling of CQC methods with system size. Formally, the computational effort for solving the RH-SCF equations scales as N^4, where N is the size of the basis set (ordinarily proportional to the size of the molecule). Correlation methods

begin with a scaling of N^5, e.g. for second order perturbation theory, and become as large as N^7 for highly accurate methods; full configuration interaction, a brute force inclusion of all possible determinants within a given basis set, grows exponentially. This behavior makes it virtually impossible to imagine applications to systems in the 50–100 atom range in the near future, even with the doubling of hardware cost/performance every few years.

In the early days of computational quantum chemistry, scaling behavior was not critically important. Because of highly restrictive memory limitations, one could, in any case, only treat small molecules, where the additional powers of N required for the higher level correlation methods was not an insuperable burden. Furthermore, the emphasis of quantum chemists in this period was on accurate comparisons with gas phase measurements, where the systematic improvability of CQC methods proved to be enormously attractive to experimentalists. In the struggle to understand the elements of chemical bonding and potential energy surfaces for small molecules, CQC has become an invaluable tool.

Over the past 20 years, however, an alternative vision of the uses to which quantum chemical calculations can be put to has developed. In molecular simulation (or molecular modeling) methods (10, 11), interaction potentials were initially constructed by empirically fitting to both simple gas phase experiments (e.g. vibrational frequencies) and bulk thermodynamic properties. When coupled to formal statistical mechanical techniques, which now range from classical Newtonian dynamics to free energy perturbation theory, such simulations offer the exciting possibility of direct theoretical modeling of condensed phase systems. Concomitantly, a host of powerful new experimental techniques for directly studying condensed phases has been invented. This research program offers an alternative to the older picture of chemical physics in which knowledge of chemical structure and reactivity was to be obtained by detailed investigation of small molecules in the gas phase, transferring the resulting insights (and interaction potentials, if they could be extracted) to condensed systems, in which most (although not all) practical chemistry takes place.

Enhanced hardware capabilities have now made molecular modeling a serious possibility for virtually any system one cares to study. A crucial question, however, is the reliability of the interaction potentials, particularly if one is interested in obtaining actual numbers, rather than just an attractive (but perhaps quantitatively incorrect) graphical representation. This, in turn, poses a challenge to ab initio quantum chemists: Can one carry out calculations efficiently enough on large systems to contribute to molecular modeling approaches? An affirmative answer to this question is

absolutely essential for the development of truly predictive molecular modeling methods. There are simply too many adjustable parameters and too few experiments to derive accurate potential surfaces without recourse to quantum chemical calculations. These difficulties are particularly acute when considering reactive (bond breaking) processes, excited states, transition metals, or any other potential surface that is not, for whatever reason, well-characterized experimentally. Ab initio methods (principally at the HF level) are already utilized extensively in computing effective point charges and force fields for small- and medium-sized molecules or molecular fragments as inputs to molecular modeling programs (10, 11); the development of more effective correlation methods and better scaling of the calculations with system size can only accelerate this trend.

What are the prospects of performing sufficiently accurate quantum chemical calculations on large (e.g. 50–100 atom) systems in reasonable CPU times? One possibility is to employ semiempirical methods (1, 12–15), which are in general considerably less expensive than any of the first principles approaches. The problem here is to develop parametrized Hamiltonians of sufficient generality and reliability. At present, this is an active area of research; however, it lies outside the scope of this review. The relevant point for our purposes is that existing semiempirical methods, while impressive in selected applications, have not yet reached a level of accuracy where the use of ab initio techniques becomes redundant.

There remain three alternatives, if one intends to stick with some sort of first principles approach. First, one can modify existing CQC methods for application to large systems, by incorporating the rapid improvements in computational hardware technology. There has been remarkable progress along these lines over the past ten years, particularly in the solution of the HF equations. Second, substantial efforts have been made in this same period to improve the accuracy and computational efficiency of DF methods. In the 1970s, for various historical reasons, the practitioners of this approach did not put the same level of effort into software and algorithm development as did the CQC workers, particularly for molecular (as opposed to solid state) calculations. This state of affairs has now altered significantly, and DF methods must be considered as serious competitors for large molecule quantum chemical calculations. Significant progress has been made in the construction of effective gradient corrections (16–19), which represent real improvements in the energy functional beyond the local density approximation, and in the development of reliable numerical techniques to evaluate the total energy and its gradient (20, 21).

Finally, one can ask whether new algorithmic approaches to solving the molecular Schroedinger equation are possible. For example, it is easy enough to design numerical methods for SCF calculations, which have

much better scaling properties with system size than the RH approach, e.g. plane wave expansion methods that formally scale as N^2 (22). However, even for large systems, the prefactor multiplying the scaling term is crucial. A characteristic feature of electronic structure calculations is that one needs enormously high relative precision (as a function of geometry) in the energy, typically on the order of one part in 10^7. To achieve this precision, methods that have superior scaling properties may require huge prefactors, so that even for 50–100 atom systems, CQC methods, suitably modified for large scale calculations could be greatly superior. Similar considerations are paramount in the development of new correlation methods, e.g. quantum Monte Carlo techniques (23–28), which are very promising in theory because of their favorable scaling properties, but to date have had great difficulty in obtaining adequate relative precision in reasonable CPU times.

Over the past five years, we have developed the first successful alternative to RH methods for solving nonlocal ab initio SCF equations for poly-atomic molecules (29–35). The method uses conventional quantum chemi-cal basis sets in conjunction with the pseudospectral method, a numerical technique borrowed from hydrodynamic simulations of turbulence (36–39). The methods bear some resemblance to approaches used to solve the DF self-consistent field equations; indeed, an important feature of our results is the demonstration that the efficacy of such numerical methods is not dependent upon the use of local exchange operators, as had been widely thought in the past. We have shown that significant reductions in computational effort, as compared with modern CQC programs, can be achieved with the method without sacrifice in accuracy for both HF and multiconfigurational self-consistent field (MCSCF) calculations.

The objective of this article is to describe the developments that have occurred over the past five years for each of these three methods and to assess prospects for the future. We give a more detailed presentation of our new methodology, as this is likely to be less familiar to readers than the two alternatives. The conclusions arising from these examinations is an exciting one; over the next five to ten years, reasonably accurate first principles methods will become applicable to large molecular structures on a routine basis. Furthermore, the development of user-friendly, com-mercialized software insures that this technology will be accessible to chemists in all fields and institutions.

The orientation of this review is towards presenting concepts, as opposed to conducting an exhaustive survey of the recent literature. A particular manifestation of this is the focus on three (among many) first principles electronic structure codes: GAUSSIAN 90, a CQC-based program; DGAUSS, a DF program utilizing Gaussian basis functions; and our own

pseudospectral codes, PSHF and PSGVB. Although the last of these are at present unique, there are a significant number of alternatives to the first two packages. Both GAUSSIAN 90 and DGAUSS are representative of the state-of-the-art in their respective methodologies; however, we intend no particular implications concerning the relative performance of these codes as compared with the above-mentioned alternatives.

CONVENTIONAL AB INITIO QUANTUM CHEMISTRY

Algorithms for Hartree-Fock Computations

The HF level of ab initio theory is the method most commonly used in actual applications and is conceptually the most straightforward first principles approach. The method is variational; it assumes a single determinant trial wavefunction and optimizes the constituent molecular orbits to yield the lowest values of the ground state electronic energy. For pedagogical purposes, we consider only the closed-shell version of the method, assuming an even number of electrons.

The HF equations for an occupied molecular orbital ϕ_α and electronic eigenvalue ε_α of the ground state of a closed-shell molecule are

$$(h_0 + 2J - K)\phi_\alpha = \varepsilon_\alpha \phi_\alpha \qquad 1.$$

where, defining r_a as the coordinate of electron a, R_b as the position of nucleus b, and Z_j as the charge on nucleus j,

$$h_0 = -1/2\nabla_a^2 - \sum_b \frac{Z_b}{|r_a - R_b|} \qquad 2.$$

is the one electron operator, and the nonlinear Coulomb and exchange operators J and K are defined in physical space as:

$$J(r_a) = \int \sum_\alpha \frac{|\phi_\alpha(r)|^2}{|r_a - r|} d^3r = \int \frac{\rho(r)}{|r_a - r|} d^3r \qquad 3.$$

$$K(r_a)^*\phi_\beta = \sum_\alpha \phi_\alpha(r_a) \int \frac{\phi_\alpha(r)\phi_\beta(r)}{|r_a - r|} d^3r. \qquad 4.$$

In Equation 3, we define the electron density $\rho(r)$ as the sum over the squares of the occupied molecule orbitals. The exchange operator K is a nonlocal integral operator that acts on a function ϕ_β by placing it inside the integral sign.

In the RH approach, one expands the molecular orbital ϕ_α in a localized

atomic basis set χ, multiplies both sides of Equation 1 by a basis function χ_j, and integrates over coordinate space, thus yielding the matrix equations

$$F\mathbf{c}_\alpha = \varepsilon_\alpha S\mathbf{c}_\alpha \qquad \qquad 5.$$

where the Fock operator F is just $h_0 - 2J + K$, S is the overlap matrix of the nonorthogonal basis functions, and $\mathbf{c}_\alpha$ is a vector of basis function coefficients that represent the molecular orbital α. The Coulomb and exchange operators are now assembled from two electron integrals over atomic basis functions $(ij|kl)$ and density matrix elements ρ_{kl} via the formulas

$$J_{ij} = \sum_{kl} (ij|kl)\rho_{kl} \qquad \qquad 6.$$

$$K_{ij} = \sum_{kl} (ik|jl)\rho_{kl}. \qquad \qquad 7.$$

The basic CQC strategy for solving the RH equations is straightforward. The one electron matrix elements of h_0 and S are easily computed analytically and stored. The two electron repulsion integrals $(ij|kl)$ can be computed analytically in a variety of ways for contracted Gaussian basis functions. One starts with an initial guess for the density matrix (e.g. from a semiempirical calculation) and uses this and the two electron integrals to assemble the Coulomb and exchange operators J and K from Equations 6 and 7. From these operators, a new guess for the density matrix can be obtained, either via diagonalization of the Fock matrix or (when close to the solution) Newton-Raphson methods (40) or extrapolation procedures, the most efficient of which has been developed by Pulay (41). This process is repeated until the change in the density matrix and/or the total energy is less than a preset tolerance, whereupon the calculation is declared to be converged to the desired level of accuracy.

A brief discussion of the structure of CQC contracted Gaussian basis sets will prove useful in what follows; in the interest of simplicity, we confine our remarks to first row atoms (although it should be noted that the development of basis sets for heavier atoms, e.g. transition metals, is still an ongoing process). For HF calculations of equilibrium geometries and vibrational frequencies, one can obtain rather good results from a double zeta (DZ) basis. For any given first row atom, except hydrogen, this consists of three s-type functions (two of which are contracted) and two p-type functions (one of which is contracted) for a total of nine functions; the analogous set for hydrogen contains just two s functions. The next level of accuracy is double zeta plus polarization (DZP), in which the DZ basis is augmented by one set of six Cartesian d (polarization) functions (or, for hydrogen, a set of p functions). The d functions are

necessary for properly describing some angular bonding structures and are essential in carrying out any sort of correlated calculation. Additional s, p, and d functions [e.g. triple zeta, double polarization (TZDP) would contain four s functions, three p functions, and two d functions] lead to improved results for specific properties (e.g. NMR chemical shifts, intermolecular forces) and allow recovery of a larger part of the correlation energy. The necessity for utilizing these extended basis sets, instead of DZP (which is currently more or less standard for second order perturbation theory calculations, for example) is not yet established. At present, this is an active research area, with considerable effort being expended on the construction of basis sets specifically designed to produce optimal results with various correlation strategies.

For sufficiently small molecules, the two electron integrals can be generated once and stored in fast memory. However, there are $N^4/8$ two electron integrals for a molecule with N basis functions and no symmetry, so this approach begins to be impractical when $N \sim 100$ for current hardware. Assuming a DZP and first row atoms, this restricts such "in core" methods to about six heavy atoms, a rather small calculation by today's standards.

An alternative approach is to store the two electron integrals on disk and read them into fast memory on each iteration. The tradeoffs here are different, as they depend upon the capacity and speed of the I/O subsystem of a particular machine. With typical (gigabyte) levels of disk storage, one is limited to systems with about 200 basis functions; for this size, I/O costs begin to compete significantly with CPU time as a bottleneck in job throughput on a multiple-user machine.

A third alternative is to recompute the two electron integrals at each iteration; this is referred to as "direct" SCF (42–45). This has the obvious advantage of eliminating the memory and disk storage issues, even for very large systems. The disadvantage, equally obvious, is the increase in CPU time associated with integral recomputation, which is the most expensive step in an HF cycle. During the past five years, substantial progress has been made in speeding up two-electron integral evaluation (46–48), in part by exploiting vector architectures (a painful process because of the nature of the integral generation algorithm) and in part by reducing operation counts, particularly those scaling as high powers of the number of contracted primitives in the Gaussian basis functions. As an example, consider the performance on a Cray X-MP of GAUSSIAN 86 (which stores integrals on disk, but uses old integral generation routines) as compared with GAUSSIAN 88, which employs a recursive, vectorizable scheme devised by Head-Gordon & Pople (47) [based partially on the work of Obara & Saika (46) along the same lines]. For the largest molecules

accessible to GAUSSIAN 86 (e.g. glutamine at the DZP level, which requires 200 basis functions), GAUSSIAN 88 provides significant reductions in CPU time despite its need to recompute integrals for 7–12 SCF cycles. These results are partly explained by the fact that one does not require the same accuracy for every SCF cycle; early cycles, being distant from the minimum, can utilize approximate integrals, whereas later ones can also be done cheaply by using the Fock matrix updating scheme of Almlof and coworkers (42). Nevertheless, the new integral generation algorithms clearly are already five to ten times faster on vector machines than the old ones, and further improvements are currently in progress (48).

Assuming a direct SCF scheme, we can next examine the actual scaling of computational effort with system size for large molecules. A substantial reduction from the formal N^4 is readily obtained by using integral cutoffs; the integral $(ij|kl)$ is not computed at all if the product of the overlap of functions i and j with that of k and l are less than a certain threshold. These cutoffs lead to an asymptotic scaling of N^2. Further reduction of the asymptotic scaling to order $N \log N$ is then possible, in principle, by employing multipole expansions of the product ij and kl charge densities, assuming that the density matrix falls off with distance reasonably quickly. Thus, we have reason to be optimistic about applying CQC to large molecular systems. In practice, for typical systems in the range of 200–400 basis functions, GAUSSIAN 90 appears to achieve a scaling of around $N^{2.7}$ (49).

Nevertheless, the computation times required for GAUSSIAN 90 are substantial; there is still a significant barrier to carrying out HF calculations on larger systems. For example, a DZP treatment of porphine, which utilizes 430 basis functions, consumes 2759 CPU seconds on a Cray Y-MP. Improvements in scaling and overall timing are, therefore, still essential if we are to utilize ab initio methods in large scale molecular mechanics simulations. One possibility is that these improvements can be achieved by employing new generations of massively parallel supercomputers. The second possibility, that the CQC algorithm can be qualitatively improved, are examined in subsequent sections.

Electron Correlation Algorithms

In contrast to HF methodology, CQC electron correlation methods are widely diverse in formalism and implementation. Many of these approaches have been extensively tested on small molecules, and a great deal is known about the sort of correlated wavefunction that is required to achieve a specified degree of accuracy for particular molecular properties. It is not the objective of the present article to examine this issue in detail. Rather, we seek to assess the promise of various approaches for

large molecule calculations. We therefore focus on the computational effort required as a function of system size, with only a brief characterization of the accuracy available from each method.

There are three basic approaches to electron correlation. One approach is to utilize a perturbation series, typically the Moller-Plesset (MP) expansion (50) (although other partitionings are possible). In the simplest methodology, the reference system is the HF Hamiltonian, and the MP series is carried out to the desired order, most frequently the second or fourth. Second order MP theory (MP2) is one of the least expensive correlation methods. It requires the generation of two electron integrals over two occupied molecular orbitals and two virtual molecular orbitals. The production of these matrix elements from two electron integrals over atomic basis functions, referred to as a partial four index transform, scales as nN^4, where n is the number of occupied orbitals and N is the basis set size. Furthermore, the transform must only be carried out once, in contrast to other correlation methods that require a moderate number (10–15) of iterations. For small- or medium-sized systems, the CPU time for MP2 is often less than for the initial HF calculation.

Usually, MP2 recovers a substantial fraction of the correlation energy (defined as the difference between the HF energy and the exact non-relativistic eigenvalue of the molecular Schroedinger equation), often in the 80–90% range if a very large basis set is used. For the first row, errors in vibrational frequencies and equilibrium geometries are reduced by as much as a factor of two or three from corresponding HF values; calculation of isodesmic (equivalent numbers of electron pairs) reaction energies is also significantly improved (51). The closed-shell MP series does not properly separate to open-shell fragments, so quantitatively reliable bond dissociation curves cannot be generated throughout the entire geometry of separation. Nevertheless, significant improvements over HF are usually obtained for bond dissociation energies and reaction barriers (51). On the other hand, chemical (2 kcal/mole) accuracy is rarely available for these quantities.

The MP series appears to converge rather slowly beyond second order; for cases where MP2 has substantial errors, MP4 rarely provides a large fraction of the corrections, and the results tend to oscillate as function of increasing the number of perturbation terms. As the computational effort grows rapidly with increasing order (MP4 scales as N^6 or N^7, depending upon whether triple substitutions are included), it is not clear whether this pathway is suitable for large molecule applications.

MP2 is one of the few correlation strategies for which a serious attempt has been made to reformulate the algorithm for large scale calculations. The most significant barrier to large scale MP2 calculations has been

storage of the two electron integrals on disk during the integrals trans-
formation steps. By employing "semidirect" methods (52–54) that involve
recalculation of two electron integrals over basis functions when necessary
(as in direct SCF approaches), it is possible to limit the disk space require-
ments for the integral transformation at the cost of some CPU time, thus
making possible MP2 calculations on systems with a few hundred basis
functions. However, larger systems still present difficulties, because of the
increase in overall computational effort as the fifth power of the system
size and the inability to employ cutoffs as a result of the delocalized nature
of the molecular orbitals.

A second class of methods (55) involves the optimization of multi-
configurational wavefunctions (i.e. sums of determinants of molecular
orbitals) via a variational procedure; typically, both the SCF orbitals and
the coefficients of the determinants are optimized. Such MCSCF methods
include complete active space SCF (CASSCF) [in which M electrons are
distributed in all possible ways among K molecular orbitals (56)] and
generalized valence bond (GVB) approaches [in which classes of valence
bond wavefunctions are optimized self-consistently (57–60)], as well as
configuration lists tailored to specific problems. These methods properly
dissociate to open-shell fragments, and almost always provide a quali-
tatively reasonable picture of molecular structure. Furthermore, quantities
like vibrational frequencies and equilibrium geometries are often sub-
stantially improved as compared with HF results. However, the inclusion
of only a limited number of excited determinants leads to quantitative
errors, particularly in dissociation energies.

The computational effort involved in MCSCF calculations depends
heavily upon how many determinantal configurations are included in the
MCSCF expansion and how many individual parameters are associated
with this expansion. In most CQC-based MCSCF algorithms, one must
carry out a full four-index transform of the atomic integrals to integrals
over molecular orbitals; this procedure scales as N^5 and must be performed
at every SCF iteration. For small systems, it is, nevertheless, relatively
inexpensive because the transforms can be written as matrix multiplies,
which are well suited to modern computational hardware. Calculations
with large configuration lists are dominated by the configuration inter-
action (CI) part of the calculation (see below), with scaling of N^6 or N^7
for various ways of systematically including configurations as a function
of system size. Not much work has been done to extend these methods to
very large systems, so we do not yet have much information concerning
practical scaling behavior. However, the use of delocalized molecular orbi-
tals clearly is inimical to the employment of physical space cutoffs. If such
an extension is to be accomplished, the localization of the SCF orbitals

to be utilized will clearly become a key problem, which has been examined in this context by only a few investigators to date (61, 62). Generalized valence bond methods appear to have a significant advantage in this regard, because the GVB orbitals are localized automatically.

A third methodology, CI (63), employs diagonalization of the Hamiltonian in a basis of excited determinants, typically constructed from HF orbitals. A conceptually simple approach is to use all single and double excitations into HF virtual orbitals from the HF ground state (SDCI); this approach scales as N^6. Over the past decade, it has become clear that for a relatively modest computational increase, one can, by employing an exponential ansatz, carry out a modified SDCI-type computation in which higher excitations (triples, quadruples, etc) are implicitly included in the wavefunction via factorization assumptions. These coupled-cluster methods (64) have proven to be capable of high accuracy (in many cases, chemical accuracy) for most properties of interest to chemists. However, the N^6 scaling remains a formidable barrier to large molecule applications, and little work has been done to investigate the use of localized orbitals or the implementation of cutoffs in coupled-cluster-based approaches to reduce the scaling.

It is possible to combine two of these approaches into a more complex scheme. The most commonly used method is to carry out CI starting with a multiconfigurational reference state, as opposed to a HF reference. These multireference configuration interactions (MR-CI) approaches have also proven capable of yielding high accuracy for a wide variety of small molecule applications (26, 55, 57–60, 63, 65). Both MR-CI and coupled-cluster have occasional failures, which often receive a great deal of publicity [e.g. the chromium dimer calculations of Goddard and coworkers (66)]. This should not obscure the fact that, overall, either approach is highly satisfactory for small molecule computations. The difficulties lie in extension to larger systems.

One promising approach is to perform a CI expansion for only a small subset of localized electron pairs. An example of this methodology is the correlation-consistent CI (CCCI) approach of Goddard and coworkers (59, 60), in which bond dissociation energies are computed by correlating only the electron pairs associated with the bond in question; the localized nature of the GVB pair orbitals is exploited here. Chemically accurate results have been obtained for several small molecule test cases, although the number of results is not yet sufficient to establish the generality of the methodology, and applications to large systems have not yet been presented. An overall scaling of N^5 is obtained for a large molecule if only a few electron pairs are correlated at the SDCI level.

Recently, there has been some interest in combining MCSCF and per-

turbation theory approaches in an attempt to devise an accurate method with more favorable scaling properties: If the MCSCF expansion does not have an extremely large configuration list, a formal scaling of N^5 can be obtained (67, 68). This idea appears to be very promising, although there are difficulties in the definition of the appropriate zeroth order Hamiltonian corresponding to an MCSCF wavefunction. Further experiments on small molecules are required to determine the accuracy of this approach for various properties of interest.

To summarize, correlation methods that are chemically accurate for many (but not all) systems have been developed that scale in the N^6 regime. Other methods, which are less reliable but still represent a significant improvement over HF theory, scale in the N^5 regime, and there are several directions in which these methods can be developed without increasing the scaling beyond N^5. To reduce the scaling below N^5, it is necessary to employ localized orbitals and physical space cutoffs in matrix elements and integral transformations. Work in this area is, at present, in its infancy.

DENSITY FUNCTIONAL METHODS

Density functional theory is based upon the formal theorem, derived by Hohenberg and Kohn, which states that the total ground state electronic energy of a molecule can be expressed as a functional of the ground state electron density $\rho(r)$. If one then guesses the functional dependence of the energy on the density, variational methods can be employed to minimize the total energy within the limitations of the functional form, thus yielding SCF equations that are very similar mathematically to the HF equations. The principal advantage of this approach is the possibility of including electron correlation in the SCF equations from the start. The disadvantage is that the quality of the answer depends upon how well one has guessed the density functional, an enterprise in which it is difficult to generate systematically improved results.

Our focus in this article is on two issues: the functional form of the local density functional (LDF) equations, as compared with the HF equations, and numerical methods of solution of each; and the prospects for obtaining reasonably accurate large molecule structures and energies by utilizing DF approaches. Additionally, the discussion is confined to the performance of the LDF equations for isolated molecules, as opposed to the crystalline and amorphous condensed phase applications in which this methodology has traditionally been dominant. New issues arise in the solid state (e.g. the qualitatively incorrect behavior of the HF approximation near the Fermi surface), which would favor the use of one method over another. Although these issues are interesting in and of themselves, the comparison

354   FRIESNER

of methods, even in the small molecule case, is quite involved, and it seems best in this short space to avoid additional complexities.

Local density functional theory (6) assumes that the exchange and correlation energies of a molecule are a local functional of $\rho(r)$. The functional itself is typically obtained by a parametrization that ensures recovering the exact exchange and correlation energy for the uniform electron gas; several such parametrizations are available in the literature (69). The density is expanded in a set of "orbitals," and the SCF equations for these orbitals reads:

$$\{h_0 + 2J - V_{xc}[\rho(r)]\}\phi_\alpha = \varepsilon_\alpha\phi_\alpha. \qquad 8.$$

The one electron and Coulomb terms h_0 and J are exactly the same here as in the HF equations above, and $\rho(r)$ is formally identical to its definition in Equation 3, i.e.

$$\rho(r) = \sum_\beta |\phi_\beta(r)|^2 \qquad 9.$$

although the meaning of the self-consistent orbitals ϕ (we refer to them as "density orbitals") is not equivalent in HF and DF methodology.

Equations 8 and 9 together form a self-consistent set that can be solved by approaches analogous to those employed for the HF equations. The orbitals ϕ_β are expanded in a Gaussian basis set, the coefficients of which are determined by solution of a matrix version of Equation 8. The SCF iteration process is initiated via an initial guess for $\rho(r)$, which is used to construct the operators in Equation 8; a new set of orbitals then yields a new $\rho(r)$ from Equation 9, and the process is repeated until the change in energy and/or electron density is sufficiently small.

The new term in Equation 8, compared with Equation 1, is the exchange-correlation potential V_{xc}, which is a local functional of the electron density $\rho(r)$; in compensation, the nonlocal integral exchange operator K is not present. These LDF equations are thus three-dimensional, nonlinear partial differential equations in which the potential operator is local in coordinate space, in contrast to the HF equations where the exchange operator is nonlocal. On the other hand, V_{xc} is typically specified numerically, so that one cannot do all of the integrals appearing in a RH-like version of the theory analytically. Thus, LDF practitioners have been forced to rely on "numerical" methods of solution from the start, which involves the use of grids in physical space on which to evaluate V_{xc} and, hence, to carry out the evaluation of the exchange-correlation matrix elements. This has led, quite naturally, to the use of similar methods to evaluate the Coulomb term, as well.

These historical developments have led to a misconception concerning the computational effort involved in solving the HF and LDF equations,

which we hope to dispel in this article. The fact is that the locality of the LDF equations is of minor (although nonnegligible) importance in ease of solution, and one can develop analogous grid-based methods for solving the nonlocal HF equations that have a very similar scaling with system size and are similar in overall operation counts, as is explained below. The crucial question is the computational efficiency of methods that use numerical grids, as compared with RH procedures, which do not use such grids. This issue has to be examined carefully, because one must compare the methods when they yield the same average level (over many molecular examples) of relative precision. Comparisons of CPU time for a grid-based method, which routinely exhibits spurious energy errors of 10 kcal/mole for local energy differences with an RH-based procedure reliable to 0.1 kcal/mole, are not very informative. Note that the relevant errors here are energy differences as a function of geometry for a particular equation (HF or DF); we wish our computational scheme to reproduce as closely as possible the differences that would be obtained via exact solution of the actual partial differential equations.

Early methods for solving the LDF equations for molecules (e.g. muffin-tin approximations) did not, in general, yield highly accurate total energies (70, 71). To some extent, this was a consequence of the objectives of those calculations. Much of the work, for example, was aimed at qualitative large molecule applications, and the DF approach was presented as one intermediate between semiempirical and ab initio methods in both accuracy and computational cost. However, the development of numerically accurate LDF algorithms for small molecules demonstrated that, in many cases, LDF solutions yielded chemical properties at a level often superior to HF (but not, generally, MP2) calculations. This has spurred the development of efficient and reliable LDF codes with complex, careful numerical schemes (20–22, 70, 72, 73).

There is still a great deal of evaluative work to be done to assess the convergence properties of the existing DF numerical methods and the performance of the LDF approximation when numerical precision is removed as an issue. For example, studies of diatomics (74) indicate that the LDF approximation often makes substantial (~ 1 eV) errors in dissociation energies; however, it usually provides a reasonable qualitative picture, typically better than HF and in some cases better than MR-CI methods [e.g. for Cr_2 (75)]. Similarly, equilibrium geometries and vibrational force constants are normally quite reasonable, again in many cases superior to HF. At present, however, there is not a large data base concerning accuracies for larger molecules or for a wide variety of properties. For example, how well does the LDF theory describe hydrogen bonded structures or rotational barriers around torsional angles? With the

development of numerically reliable LDF programs, it should be possible to assemble substantive data bases analogous to those that already exist for HF computations of molecules that contain first row atoms. Only in this fashion can we obtain a truly unbiased performance evaluation.

Efforts have been underway for many years to improve the LDF approximation and produce more accurate density functionals, a task that certainly seems feasible in principle. However, only recently have useful methods for including density gradient corrections (16–19) (the logical systematic way to improve a functional based on an electron gas of uniform density) been developed. In preliminary tests (21), these methods appear to lead to systematically better results for a wide variety of molecular properties, although the quality of the results are not yet at the level of chemical accuracy, and the degree of improvement obtained is far from uniform. Although it is too early to know exactly what level of accuracy can be achieved by this approach, this is certainly an exciting research area with potentially great promise.

There are other issues that must be considered when applying LDF methods. The treatment of open-shell systems via a local spin density approximation (76), for example, is difficult to improve systematically. Excited state calculations are possible (77), but again the quality of the approximations is unclear and difficult to upgrade. There are numerous other properties, e.g. magnetic resonance coupling constants, electric field gradients, and optical activity parameters, with which there is little or no experience. We also do not know how well the LDF method works in determination of transition states for chemical reactions, as there have been few quantitative applications of this type.

Nevertheless, the LDF method has had enough successes in molecular calculations over the past five years to become a serious alternative to wavefunction-based methodology. One important question, which we have not yet examined, is that of computational effort. The formal scaling of the bulk of the numerical methods currently used to obtain accurate solutions to the molecular LDF equations is N^3; with the use of cutoffs, this can be reduced to on the order of N^2, and a scaling of $N^{2.2}$ is reported for medium to large molecules (21). Recent work of Andzelm and co-workers (21) has begun to address issues, such as convergence of the energy and equilibrium geometry, with respect to the basis set, number of grid points, and other numerical parameters. Reasonably well-converged (~ 0.1 kcal/mole) relative energies can apparently be obtained with moderate grid (~ 1000 pts/atom) sizes. Andzelm and coworkers (21) report a CPU time of 20 minutes on a Cray Y-MP for an energy plus gradient calculation involving 62 atoms.

By comparison, the estimated time for GAUSSIAN 90 to complete a

comparable calculation, based on a scaling of $N^{2.7}$ and a 45 minute CPU time for porphine as described above, is on the order of 3 hours (180 minutes). This comparison is not completely straightforward, however, because one must estimate what sort of basis set will render the GAUSSIAN and LDF calculations equivalent; in general, RH-based algorithms can utilize smaller basis sets, because they display superior cancellation of error across geometry. The above estimation assumes a DZP and then an average of ten basis functions per atom, plus 50% additional time for the gradient calculation. On the basis of these numbers, there is a clear computational advantage in utilizing the DF methodology, an advantage that increases as the molecules become larger. And, as we have seen, the accuracy of even the LDF approach is typically as good as or better than HF, and promises to become greater still with improved functionals, which do not substantially increase the computational effort of DF methods.

Traditional ab initio methods are in no danger of losing their primacy in chemically accurate studies of small molecules, where they can achieve a precision far greater than DF methods will be able to in the foreseeable future (e.g. on the order of 1–2 kcal/mole accuracy for dissociation energies), and the unpleasant computational scaling of the accurate correlation methods required to obtain such agreement does not lead to prohibitive CPU times. For large systems, however, DF approaches may become the methods of choice unless new strategies can be implemented to reduce the computational scaling of both HF and correlated methods for large systems. In the next few sections we describe one effort along these lines, which has been pursued in our research group.

NUMERICAL METHODS FOR AB INITIO QUANTUM CHEMISTRY

Historical Background and Analysis

The first HF calculations for atoms were performed by Hartree and Hartree (3), who used standard numerical methods for solving differential equations. The spherical symmetry of the atomic problem can be used to reduce the numerical computations to one (radial) dimension, which is readily dealt with via finite difference algorithms. Such calculations are still carried out and generally yield results with lower total energies than even quite large Gaussian basis sets.

The extension of these numerical methods to higher dimensions poses formidable problems of numerical precision. One would like to know relative conformational energies, for example, to an accuracy of 1 kcal/mole at a minimum and preferably to 0.1 kcal/mole. Leaving aside for the

moment the question of the reliability of the correlation approximation being made (HF rotational barriers are, in fact, quite reasonable), this requires a precision of one part in 10^6 for a small molecule and one part in 10^8 for a large molecule. As the energy difference cannot be directly evaluated, each total energy must be computed to this accuracy unless one can rely upon a systematic cancellation of error across geometry. As we see below, such a cancellation is much more difficult to achieve with numerical methods than with analytical ones.

Over the past 40 years the key to the dominance of the Gaussian-based RH methodology is the remarkable convergence of relative conformational energies with basis set size. The basis sets used in molecular calculations—even those like DZP that are considered high quality—are astonishingly small considering the precision demanded in the energy. Consider, as a simple example, the water molecule. The energy near equilibrium is on the order of 76.023 au for a standard DZP basis set, 6–31 G** (78). A near HF limit calculation yields a total energy of 76.065 au (78). However, the O–H bond lengths obtained from the two basis sets differs by a few thousandths of an angstrom, and the bond angle by less than one degree. In other words, the potential surfaces obtained from these two computations are very similar, despite the difference in total energy of on the order of 25 kcal/mole. This error cancellation across geometry is not as consistent when the energies of different molecules (e.g. separated fragments) are computed, and one is advised to correct for this basis set superposition error via now-standard techniques (79). Nevertheless, the local reliability of energy differences is crucial to many applications of CQC methodology.

Now consider the same sort of conformational energy difference calculation carried out with a numerical method involving grid points. Any reasonable numerical method has the grid points move with the atomic nuclei; a uniformly dense grid, in which the position of the nuclei are ignored, would require an inordinate number of points per atom and would be prohibitively expensive in accurate polyatomic calculations. Thus, the integration grid used is altered as the molecule changes conformation, and there is a spurious energy difference due to inequivalent numerical integration schemes. It is the absence of such differences in the RH approach (because all integrations are done analytically) that give the method its power, despite the formally worse scaling properties.

One way to overcome the above dilemma is to utilize analytical gradient technology. In the context of a numerical method, what one is doing is computing changes in the energy as a function of geometry over a small region without changing the integration scheme. The LDF numerical approaches now exploit this idea extensively. However, for determining

conformational energy differences between local minima and barrier tops, this approach is not adequate, and one must directly confront the precision problem.

Over the past 20 years, there has been a lot of work on numerical HF methods for diatomic molecules, which has principally employed finite difference techniques. Here, one can utilize a special (elliptical) set of coordinates and exploit symmetry so that the numerical grid need only be in the plane (rather than fully three-dimensional) (80, 81). Highly accurate results have been obtained for diatomics, albeit at a rather large CPU cost. However, methods based on specialized diatomic coordinate systems are not very useful for a general polyatomic molecule. To our knowledge, there has not been a single accurate HF grid-based calculation for a polyatomic molecule that contains more than two electrons until our work in this area began several years ago.

Pseudospectral Hartree-Fock Methodology

To understand more clearly the difficulties involved in utilizing grid-based methods to solve the HF equations, we first examine one method currently used in solving the LDF equations. As indicated above, it is now possible to obtain reasonably reliable total energies by using this approach (which itself is a fairly recent development). Many of the ideas used in our HF methodology are quite similar to the LDF technique that we describe here. The key difference, treatment of nonlocal exchange and the modification of the algorithms it requires, is examined in some detail below.

The one electron kinetic energy and electron-nuclear attraction operators (i.e. h_0) in both theories can be treated trivially by computing analytical matrix elements. In LDF methods based on expansion of the density orbitals in a Gaussian basis set, one must evaluate integrals of the form $J_{ij} = (ij|J(r)) = (ij|\rho(r)/(r-r'))$ and $V_{ij} = (ij|V_{xc}(r))$ where i and j are Gaussian basis functions, $J(r)$ is the Coulomb operator given by Equation 3, and $V_{xc}(r)$ is the exchange-correlation potential. The important point here is that there are only two functions, $J(r)$ and $V_{xc}(r)$, for which one needs matrix elements with pairs of basis functions.

In the program DGAUSS of Andzlem and coworkers, $\rho(r)$ and $V_{xc}(r)$ are themselves expanded in Gaussian functions. Once this expansion is conducted, the resulting integrals over three Gaussian basis functions (in the Coulomb case, these integrals additionally involve the kernel $1/|r-r'|$) can be done analytically by using standard Gaussian integral technology; in fact, the recursive technology used in DGAUSS is very similar to that used in GAUSSIAN 90. The smaller number of such integrals, and the fact that they are, at most, three- (rather than four

centered, leads to significant reductions in computational effort. Of course, the fitting procedure has to be checked carefully to make sure it has not induced spurious numerical errors into relative energies. Apparently, enough effort has now been put into the DGAUSS program so that this problem is reasonably well understood.

The Coulomb terms in the HF equations could be treated analogously. The problem lies in the nonlocal exchange operators. Rather than fitting a single charge density, a corresponding algorithm for nonlocal exchange would require fitting all product densities consisting of one occupied molecular orbital and one basis function, to evaluate integrals of the form $K_{i\alpha j\beta} = (i\alpha|j\beta)$, where i and j are basis functions and α and β are occupied molecular orbitals. The saving of computational effort is clearly much smaller here than for the local operators, as the number of charge density fits (of the product functions $j\beta$) now rises from two to nN, where n is the number of occupied molecular orbitals and N is the basis set size.

There is, however, another alternative, although it is not immediately obvious from the previous discussion. Rather than fitting the charge densities, it is possible to fit the product of the integrated charge density multiplied by an orbital—or, more generally, sums of integrated charge densities multiplied by orbitals. For the Coulomb term, for example, this involves fitting the product $J(r)\chi_i(r)$ to a Gaussian expansion for each basis function χ_i and projecting the results back onto basis function χ_j to obtain the Fock matrix element F_{ij}. The algorithm to do this can be shown to scale formally like N^3. Perhaps somewhat surprisingly, an analogous algorithm for assembling the exchange operator K_{ij} can be devised that also scales like N^3 and is no more expensive in operation counts than the Coulomb algorithm. This observation led to the development of our new grid-based HF program (29–35).

The original idea for this new approach did not come from electronic structure theory, but rather from numerical methods used in fluid dynamics. In 1970, Orszag proposed that the nonlinear terms in the Navier-Stokes equations be evaluated by computing derivatives analytically, multiplying the results together on a physical space grid, and transforming the resulting product back to a basis set (in his case, sine and cosine functions) (36–39). This pseudospectral method (so-called because it utilized a basis set, i.e. a spectral expansion, but in addition a physical space grid) is now routinely employed in hydrodynamic simulations of turbulence; furthermore, after its introduction into chemical physics by Kosloff & Kosloff (82), it has been increasingly utilized for solving quantum mechanical vibrational and scattering problems (82–85). The use of the method in conjunction with Gaussian basis functions has been implemented in this area quite successfully by Light et al (84).

We do not discuss many of the numerical details here; the reader is referred to Refs. 29–35 for an extensive explication. However, a summary of the steps required to assemble a Fock matrix provides a useful overview. First, the action of the Coulomb and exchange operators on a basis function, which yields a new function expressed on the numerical grid, are computed exactly, principally by manipulating potential integrals involving two basis function indices and one grid index. Because these integrals contain only two (rather than four) basis function indices, they are significantly less expensive to generate per integral than the four-center integrals of CQC; furthermore, one can easily vectorize calculations over the grid, thus leading to a more efficient use (by roughly a factor of 3) of the vector processors present in most currently used supercomputers. The number of integrals required, ignoring the effects of cutoffs, is N^2M, where N is the basis set size and M is the number of grid points. As M is proportional to N, the formal scaling is N^3. The actual computational efficiency clearly is critically dependent upon the size of M; if the number of grid points required is so large that M is much greater than N^2 for practical values of N, the "advantage" of the method is nonexistent.

The second step in the method consists of fitting the new function, produced by the action of the Fock operator on a basis function that uses a least squares procedure, and projecting the resulting expansion onto a second basis function by utilizing overlap integrals, which are trivial to obtain computationally. An expanded, even-tempered Gaussian basis set, with exponents spaced by a factor of 2 and angular momentum varying from s to g functions depending upon the exponent, is used in the least squares expansion. The employment of analytical projection as the last step essentially filters components, which are outside of the basis set, from the least squares expansion. This filtering procedure, known as dealiasing in the Fourier transform literature, renders the procedure equivalent to the RH method in the limit that the least squares basis is complete.

Dealiasing allows one to approach the RH results for a given basis set without explicitly increasing the basis set size. An increase in the size of the dealiasing basis effects only the formation of the least squares matrices, and not assembly of the Fock operator. Thus, as long as the computational effort in this part of the algorithm does not become excessive, one can exploit the cancellation of error that has already been demonstrated for CQC Gaussian basis sets in RH calculations. This provides a practical solution to the problem of spurious numerical errors in the energy as a function of geometry; the filtering procedure essentially removes these errors and mimics the results obtained from an unbiased integration scheme.

Improvement of the least squares technology, both mathematically and

in the computer implementation, is ongoing and can be expected to reduce the required grid size and computational effort significantly in the future. For example, different, localized least squares schemes are now used depending upon which basis functions are to be projected onto (33). The design of the grid itself is crucial to efficient functioning. At present, a molecular grid is generated by joining atomic grids composed of spherical shells of points with weights that exactly integrate spherical harmonics up to a certain order (35). The integration of spherical harmonics by the radial shells (86–89) is essential and also employed in the DF numerical methods described above.

To make this technology function accurately and efficiently, it is necessary to optimize the parameters associated with the grid and least squares expansion functions. We performed this optimization by hand for a restricted set of test cases, then checked the quality of results for numerous additional molecules. Once a satisfactory set of parameters has been developed for a given atomic basis set, it can then be employed in an arbitrary molecular calculation. We additionally employ many, now standard techniques from direct SCF methodology, e.g. Fock matrix updating and extrapolation procedures. The Fock matrix updating is well complemented by a multigrid scheme, in which grids of different sizes (and hence accuracies) are utilized in differing iterations; both the early iterations and those subsequent to the initiation of the updating procedure can be carried out on grids that are roughly five times smaller than a grid yielding high accuracy.

A final important point is that as many terms as possible are evaluated analytically. The one electron operator can be treated in this fashion, as in LDF theory. However, we also currently utilize analytical two electron integrals when all four basis functions are on the same atom. This has negligible computational cost, and has been shown to reduce the grid size required to achieve accurate results significantly.

Our criterion for accuracy has been to reproduce the RH total energy for a given basis set to within 0.1 kcal/mole, regardless of molecular geometry. If the dealiasing procedure can be made sufficiently robust so that this can be done reliably, the pseudospectral algorithm becomes an isomorphic replacement for conventional RH programs and can take over the large literature of conclusions that rely on the error cancellation inherent in the RH method. As reported in Ref. 33, this goal has now been achieved for first row atoms; optimizations for second row atoms and transition metals are in progress. The major difficulty in the procedure is that, as one changes the algorithm significantly, new optimizations need to be performed. We have, therefore, set up a variety of software tools to facilitate optimization; these, combined with the increasing power of

computational hardware and the experience gained in previous optimizations, have led to a tolerable (but still significant) effort to complete this task. Further automation of the parameter optimization procedure is clearly possible, and work along these lines is in progress.

The numerical methods described above are not yet close to their final form, and we expect that factors of 5–10 reduction in computational effort are available from relatively straightforward modifications of the algorithm that are currently in progress. It is, nevertheless, useful to compare the CPU times of the present pseudospectral Hartree-Fock code, PSHF, with those of GAUSSIAN 90. The comparisons are for a single point energy calculation, starting from a semiempirical initial guess, utilizing the same basis set, and converging to 5.0×10^{-5} atomic units, the current GAUSSIAN 90 default value (a tighter convergence criterion would substantially increase the CPU time required by GAUSSIAN 90, while making virtually no difference in the PSHF timings). The programs are both run in direct mode, i.e. integrals are recalculated at every iteration.

For the porphine molecule, employing a 6–31 G** basis of 430 basis functions, PSHF requires 938 CPU seconds, as compared with the 2759 CPU seconds for GAUSSIAN 90. The scaling of PSHF with system size in the 200–400 basis function range is approximately $N^{2.2}$, as compared with the $N^{2.7}$ obtained for GAUSSIAN 90. From these results, we judge that the PSHF method is currently comparable in efficiency to the LDF program DGAUSS, with the actual CPU advantage dependent upon which assumptions one makes concerning what constitutes equivalent basis sets.

Pseudospectral Methods for Electron Correlation

Thus far, pseudospectral methods for electron correlation have been applied only in the context of the GVB perfect pairing (PP) methodology. As this approach can be formulated in terms of Coulomb and exchange operators for each GVB electron pair, it is straightforward to show that by using pseudospectral techniques the formal computational scaling is reduced from pN^4 to pN^3, where p is the number of GVB pairs; this is readily reduced to a pN^2 scaling via cutoffs. Reference 34 demonstrates that the numerical accuracy of pseudospectral correlated calculations is equivalent to that obtained for HF, i.e. on the order of 0.1 kcal/mole agreement with RH results.

An extremely important observation is that higher levels of GVB theory can be implemented using pseudospectral methods with essentially no increase in computational effort. In particular, the GVB-RCI (restricted configuration interaction) wavefunction (59, 60) can be determined with-

out the necessity of a four index transform (which scales as N^5); in fact, the formal scaling can be shown again to be pN^3, with only a small increase in the prefactor as compared with GVB perfect pairing. As GVB-RCI is a considerably better approximation than GVB perfect pairing (in many cases comparable to CASSCF), this offers the hope of obtaining a very good MCSCF wavefunction for quite large molecules in reasonable CPU times.

Application to other correlation methods also appears feasible. Preliminary investigation indicates that the use of pseudospectral techniques can be quite useful in addressing the problem of storage of configuration coefficients encountered in large scale configuration interaction calculations (E. A. Carter 1991, personal communication) and the integral disk storage problem in perturbation theory methods. The actual advantages gained in these approaches, if any, can be assessed only by actually developing a serious code, a lengthy undertaking.

A particularly attractive pseudospectral correlation method is obtained from the combination of GVB-RCI and second order perturbation theory. Because the perturbation theory can be carried out by using the highly localized GVB orbitals, the entire method can be expected to scale in the N^3 regime for large systems, even if every valence electron pair is treated at the GVB/perturbation theory level (and as N^2 if only a few electrons of critical importance are treated thusly, or in the extreme asymptotic limit in any case). Although there are problems associated with this theoretical formulation, there is every hope of solving them in a cost-effective fashion. This theory might then provide a competitive wavefunction-based alternative to the DF approach for large molecules, yielding higher accuracy and greater flexibility (e.g. reasonably accurate excited state wavefunctions) at only a moderately greater computational cost.

CONCLUSION

Whatever turns out to be the optimal first principles electronic structure method, we are clearly entering a new era in which high quality calculations will be available for increasingly larger molecular structures. The adaptation of the methods to massively parallel computers, while an extremely difficult task, will further extend the range of systems that can be profitably studied, as will continued improvements in the basic algorithms. Truly predictive chemistry for new materials is likely to occur, however, only when more powerful approaches to the correlation problem (either from the DF or wavefunction perspective) have been produced; but, there is reason

to be optimistic here, also, as there are many promising avenues to pursue.

Although DF numerical methods will surely continue to improve, the key to progress is the construction of better functionals. This research area deserves to attract a greatly enhanced effort, given the progress made over the past five years. The development of a large data base of results and experimentation with a wide variety of approaches to functional improvement, essential for further progress, will require a critical mass of investigators.

The additional complexity of the wavefunction-based techniques, especially those involving correlation, means that further development of numerical methods is both necessary and likely. One promising approach for very large systems is to utilize multipole approximations to the two electron integrals; it has already been demonstrated that this leads to an asymptotic scaling of $N \log N$ (or even N) for HF calculations. Multipole methods can also be profitably integrated into pseudospectral technology.

Further development of the pseudospectral approach logically involves a more intimate synthesis with CQC techniques, as these have been dramatically improved over the past five years. For example, one could calculate two- and selected three-center integrals analytically (by using the new, accelerated two electron integral technology) at a very low cost, and then use pseudospectral approximations only for the remaining three- and four-center integrals; this would allow the use of considerably sparser grids, because of the smaller size of the residual terms. We expect that a truly efficient method will utilize all of these techniques, integrated to yield the most efficient overall computational scheme. The end result should make HF and LDF calculations similar in their required computational effort. The situation is less clear for correlated wavefunctions, and estimates of attainable practical scalings and CPU times will have to await actual computational experiments.

Other approaches to the correlation problem are possible. For example, given the arguments we have made above, one might combine a correlation-only density functional with exact treatment of exchange. Quantum Monte Carlo methods may yet be radically improved in their computational efficiency and numerical precision. Finally, the addition of semiempirical elements to all of these "first principles" methods should be seriously considered; if one can achieve decent results with an incorrect (Hartree-Fock) functional form and an inflexible, single zeta basis set, the use of a more accurate ab initio formalism coupled to parametrization of some key elements of the theory might do much better at a relatively small computational cost. Encouraging results along these lines have been obtained by Goodgame & Goddard (66).

ACKNOWLEDGMENTS

Literature Cited

1. Hoffmann, R. J. 1963. *J. Chem. Phys.* 39: 1397–1412
2. Hartree, D. R. 1928. *Proc. Cambridge Philos. Soc.* 24: 328–36
3. Hartree, D. R. 1957. *The Calculation of Atomic Structures.* New York: Wiley. 310 pp.
4. Fock, V. A. 1930. *Z. Phys.* 15: 126–33
5. Hohenberg, P., Kohn, W. 1964. *Phys. Rev. B* 136: 864–71
6. Kohn, W., Sham, L. J. 1965. *Phys. Rev. A* 140: 1133–38
7. Huzinaga, S. 1965. *J. Chem. Phys.* 42: 1293–1302
8. Dunning, T. H. 1970. *J. Chem. Phys.* 53: 2823–33
9. Hehre, W. J., Lathan, W. A., Newton, M. D., Ditchfield, R., Pople, J. A. 1970. *GAUSSIAN 70 Quantum Chem. Program Exch. No. 236.* Indiana Univ., Bloomington, Ind.
10. Kollman, P. 1987. *Annu. Rev. Phys. Chem.* 38: 303–16
11. Jorgensen, W. L. 1989. *Acc. Chem. Res.* 22: 184–201
12. Pople, J. A., Beveridge, D. L. 1970. *Approximate Molecular Orbital Theory.* New York: McGraw-Hill. 254 pp.
13. Dewar, M. J. S., Thiel, W. 1977. *J. Am. Chem. Soc.* 99: 4899–4907
14. Warshel, A., Karplus, M. 1972. *J. Am. Chem. Soc.* 94: 5612–25
15. del Bene, J., Jaffe, H. 1968. *J. Chem. Phys.* 48: 1807–13
16. Becke, A. D. 1989. *Int. J. Quantum Chem. Symp.* 23: 599
17. Becke, A. D. 1989. *The Challenge of d and f Electrons: Theory and Computation, ACS Symp. Ser.,* ed. D. Salahub, M. Zerner, 394: 166–73
18. Perdew, J. P. 1986. *Phys. Rev. B* 33: 8822–24
19. Stoll, H., Pavlidou, C. M., Preuss, H. 1978. *Theor. Chim. Acta* 49: 143–49
20. Fournier, R., Andzelm, J., Salahub, D. R. 1989. *J. Chem. Phys.* 90: 6371–77
21. Andzelm, J. 1991. In *Density Functional Methods in Chemistry,* ed. J. Labanowski, J. Andzelm. New York: Springer-Verlag
22. Teter, M., Payne, M. C., Allan, D. C. 1989. *Phys. Rev. B* 40: 12255–63
23. Hammond, B. L., Reynolds, P. J., Lester, W. A. 1988. *Phys. Rev. Lett.* 61: 2312–15
24. Skinner, D. W., Moskowitz, J. W., Lee, M. A., Whitlock, P. A., Schmidt, K. E. 1985. *J. Chem. Phys.* 83: 4668–72
25. Umrigar, C. J., Wilson, K. G., Wilkins, J. W. 1988. *Phys. Rev. Lett.* 60: 1719–22
26. Werner, H.-J. 1987. *Adv. Chem. Phys.* 70: 1–62
27. Andersen, J. B. 1975. *J. Chem. Phys.* 63: 1499–1503
28. Reynolds, P. J., Ceperley, D. M., Alder, B. J., Lester, W. A. 1982. *J. Chem. Phys.* 77: 5593–5603
29. Friesner, R. 1985. *Chem. Phys. Lett.* 116: 39–43
30. Friesner, R. 1986. *J. Chem. Phys.* 85: 1462–68
31. Friesner, R. 1987. *J. Chem. Phys.* 86: 3522–31
32. Ringnalda, M. N., Won, Y., Friesner, R. 1990. *J. Chem. Phys.* 92: 1163–73
33. Ringnalda, M. N., Belhadj, M., Friesner, R. A. 1990. *J. Chem. Phys.* 93: 3397–3407
34. Langlois, J. M., Muller, R. P., Coley, T. R., Goddard, W. A. III, Ringnalda, M. N., et al. 1990. *J. Chem. Phys.* 92: 7488–97
35. Friesner, R. 1988. *J. Phys. Chem.* 92: 3091–96
36. Orszag, S. 1972. *Stud. Appl. Math.* 50: 293–327
37. Orszag, S. 1971. *Phys. Rev. Lett.* 26: 1100–3
38. Gottlieb, D., Orszag, S. 1977. *Numerical Analysis of Spectral Methods: Theory and Application.* Philadelphia: SIAM

39. Orszag, S., Patera, A. T. 1980. *Phys. Rev. Lett.* 45: 989–93
40. Douady, J., Ellinger, Y., Subra, R., Levy, B. 1980. *J. Chem. Phys.* 72: 1452–62
41. Pulay, P. 1980. *Chem. Phys. Lett.* 73: 393–98
42. Almlof, J., Faegri, K., Korsell, K. 1982. *J. Comput. Chem.* 3: 385–99
43. Cremer, D., Gauss, J. 1986. *J. Comput. Chem.* 7: 274–82
44. Haser, M., Ahlrichs, R. 1989. *J. Comput. Chem.* 10: 104–11
45. Almlof, J., Luthi, H. P. 1987. In *Supercomputer Research in Chemistry and Chemical Engineering*, ACS Symp. Ser., ed. K. F. Jensen, D. G. Truhlar, 353: 35. Washington, DC: ACS
46. Obara, S., Saika, A. 1986. *J. Chem. Phys.* 84: 3963–74
47. Head-Gordon, M., Pople, J. A. 1988. *J. Chem. Phys.* 89: 5777–86
48. Gill, P. M. W., Head-Gordon, M., Pople, J. A. 1990. *J. Phys. Chem.* 94: 5564–72
49. Frisch, M. J. 1990. *GAUSSIAN 90 Users Guide and Programming Reference*. Pittsburgh: Gaussian. 335 pp.
50. Moller, C., Plesset, M. 1934. *Phys. Rev.* 46: 618–22
51. Hehre, W., Radom, L., von Schleyer, P. v. R., Pople, J. A. 1986. *Ab Initio Molecular Orbital Theory*. New York: Wiley. 547 pp.
52. Head-Gordon, M., Pople, J. A., Frisch, M. J. 1988. *Chem. Phys. Lett.* 153: 503–6
53. Saebo, S., Almlof, J. 1989. *Chem. Phys. Lett.* 154: 83–89
54. Ahlrichs, R., Bar, M., Haser, M., Horn, H., Kolmel, C. 1989. *Chem. Phys. Lett.* 162: 165–69
55. Shepard, R. 1987. *Adv. Chem. Phys.* 69: 63–200
56. Roos, B. O., Taylor, P. R. 1980. *Chem. Phys.* 48: 157–73
57. Goddard, W. A., Harding, L. B. 1978. *Annu. Rev. Phys. Chem.* 29: 363–96
58. Bobrowicz, F. W., Goddard, W. A. III. 1977. In *Modern Theoretical Chemistry: Methods of Electronic Structure Theory*, ed. H. F. Schaefer, pp. 79–102. New York: Plenum. 265 pp.
59. Carter, E. A., Goddard, W. A. III. 1988. *J. Chem. Phys.* 88: 1752–63
60. Carter, E. A., Goddard, W. A. III. 1988. *J. Chem. Phys.* 88: 3132–40
61. Saebo, S., Pulay, P. 1988. *J. Chem. Phys.* 88: 1884–91
62. Kirtman, B., Dykstra, C. E. 1986. *J. Chem. Phys.* 85: 2791–98
63. Shavitt, I. 1977. In *Modern Theoretical Chemistry: Methods of Electronic Structure Theory*, ed. H. F. Schaefer, pp. 189–232. New York: Plenum. 265 pp.
64. Bartlett, R. J. 1981. *Annu. Rev. Phys. Chem.* 32: 359–401
65. Bauschlicher, C. W., Taylor, P. R. 1986. *J. Chem. Phys.* 85: 2779–83
66. Goodgame, M. M., Goddard, W. A. 1985. *Phys. Rev. Lett.* 54: 661–64
67. Andersson, K., Malmqvist, P. A., Roos, B. O., Sadlej, A. J., Wolinski, K. 1990. *J. Phys. Chem.* 94: 5483–88
68. Wolinski, K., Pulay, P. 1989. *J. Chem. Phys.* 90: 3647–59
69. Vosko, S. H., Wilk, L., Nusair, M. 1980. *Can. J. Phys.* 58: 1200–11
70. Dunlap, B. I., Connolly, J. W., Sabin, J. R. 1979. *J. Chem. Phys.* 71: 3396–3402
71. Case, D. A. 1982. *Annu. Rev. Phys. Chem.* 33: 151–71
72. Baerends, E. J., Ellis, D. E., Ros, P. 1973. *J. Chem. Phys.* 2: 41–51
73. Becke, A. D. 1988. *J. Chem. Phys.* 88: 2547–53
74. Jones, R. O. 1987. In *The Single Particle Density in Chemistry and Physics*, ed. N. H. March, pp. 73–89. London: Academic. 385 pp.
75. Baykara, N. A., McMaster, B. N., Salahub, D. R. 1984. *Mol. Phys.* 52: 891–905
76. Von Barth, U., Hedin, L. 1972. *J. Phys. C* 5: 1629–33
77. Theophilou, A. K. 1987. *The Single Particle Density in Chemistry and Physics*, ed. N. H. March, pp. 201–18. London: Academic. 385 pp.
78. Harihan, P. C., Pople, J. A. 1973. *Theor. Chim. Acta* 28: 213–22
79. Boys, B. F., Bernardi, F. 1970. *Mol. Phys.* 19: 553–66
80. McCullough, E. A. 1982. *J. Phys. Chem.* 86: 2178–82
81. McCullough, E. A. 1986. *Comput. Phys. Rep.* 4: 265–312
82. Kosloff, D., Kosloff, R. 1983. *J. Comput. Phys.* 52: 35
83. Peet, A. C. 1989. *J. Chem. Phys.* 90: 4363–69
84. Light, J. C., Hamilton, I. P., Lill, J. V. 1985. *J. Chem. Phys.* 82: 1400–9
85. Kosloff, R. 1988. *J. Phys. Chem.* 92: 2087–2100
86. Levedev, V. I. 1975. *Zh. Vychisl. Mat. Mat. Fiz.* 15: 48–51
87. Lebedev, V. I. 1976. *Zh. Vychisl. Mat. Mat. Fiz.* 16: 293–99
88. Lebedev, V. I. 1976. *Sib. Mat. Zh.* 18: 132–42
89. Stroud, A. H. 1971. *Approximate Calculation of Multiple Integrals*. Englewood Cliffs, NJ: Prentice-Hall. 356 pp.

CONTENTS

viii   CONTENTS (*continued*)