

**ADVANCES IN ELECTRONIC STRUCTURE THEORY: GAMESS A DECADE  
LATER  
MARK S. GORDON AND MICHAEL W. SCHMIDT  
DEPARTMENT OF CHEMISTRY AND AMES LABORATORY, IOWA STATE  
UNIVERSITY, AMES, IA 50011**

**Abstract.** Recent developments in advanced quantum chemistry and quantum chemistry interfaced with model potentials are discussed, with the primary focus on new implementations in the GAMESS electronic structure suite of programs. Applications to solvent effects and surface science are discussed.

## **1. Introduction**

The past decade has seen an extraordinary growth in novel new electronic structure methods and creative implementations of these methods. Concurrently, there have been important advances in “middleware”, software that enables the implementation of efficient electronic structure algorithms. Combined with continuing improvements in computer and interconnect hardware, these advances have extended both the accuracy of computations and the sizes of molecular systems to which such methods may be applied.

The great majority of the new computational chemistry algorithms have found their way into one or more broadly distributed electronic structure packages. Since this work focuses on new features of the GAMESS (General Atomic and Molecular Electronic Structure System<sup>1</sup>) suite of codes, it is important at the outset to recognize the many other packages that offer both similar and complementary features. These include ACES<sup>2</sup>, CADPAC<sup>3</sup>, DALTON<sup>4</sup>, GAMESS-UK<sup>5</sup>, HYPERCHEM<sup>6</sup>, JAGUAR<sup>7</sup>, MOLCAS<sup>8</sup>, MOLPRO<sup>9</sup>, NWChem<sup>10</sup>, PQS<sup>11</sup>, PSI3<sup>12</sup>, Q-Chem<sup>13</sup>, SPARTAN<sup>14</sup>, TURBOMOLE<sup>15</sup>, and UT-Chem<sup>16</sup>. Some<sup>1,3,4,10</sup> of these packages are distributed at no cost to all, or to academic users, while others are commercial packages, but all are generally available to users without restriction or constraint. Indeed, the developers of these codes frequently collaborate to share features, a practice that clearly benefits all of their users.

The remainder of this paper is organized as follows. In Section 2, recently developed and implemented methods in quantum mechanics (QM) are discussed. New scalable methods for correlated wavefunctions are presented in Section 3, and several approaches for interfacing quantum mechanics with molecular mechanics (MM) are considered in Section 4. Conclusions and projections into the future are provided in Section 5.

## 2. QM Methods

*2.1 Variational Methods.* The simplest variational methods are Hartree-Fock (HF) and density functional theory (DFT), both of which are available in GAMESS. In principle, DFT is an exact, *ab initio* method, but in practice, since one does not know the exact density, it has become a very successful semi-empirical method, with several parameters that are fitted either to experiment or to simple models like the non-interacting homogeneous electron gas. DFT has been implemented in GAMESS using both the traditional grid-based approach and a grid-free method, due originally to Almlöf<sup>17a</sup> and extended by Glaesemann and Gordon<sup>17b</sup>, that employs the resolution of the identity (RI) to facilitate formulating DFT in algebraic form. Several of the most commonly used functionals are available in GAMESS, including B3LYP and BLYP.

In the 1950s Löwdin<sup>18</sup> showed that a wavefunction that includes all possible excitations from the reference wavefunction (usually the electronic ground state) is the exact wavefunction for the given atomic basis. Therefore, this level of theory, commonly called full configuration interaction (full CI), is the benchmark against which all advanced QM methods that include electron correlation may be measured. Indeed any level of CI, perturbation theory, or coupled cluster theory can be extracted from a full CI wavefunction and compared with the exact result. It is therefore very useful to develop and implement a full CI method that can be applied to as large an array of atomic and molecular species as possible. Such a full CI code based on a determinant, rather than a configuration, expansion has been developed by Ivanic and Ruedenberg<sup>19</sup> and implemented into GAMESS.

A special case of full CI is the CASSCF (complete active space self-consistent field) or FORS (fully optimized reaction space) approach in which one defines an active space of orbitals and corresponding electrons that are appropriate for a chemical process of interest<sup>20</sup>. The FORS wavefunction is then obtained as a linear combination of all possible electronic excitations (configurations) from the occupied to the unoccupied (virtual) orbitals in the active space, so a FORS wavefunction is a full CI within the specified active space. Since a full CI provides the exact wavefunction for a given atomic basis, there is no need to re-optimize the component molecular orbitals. On the other hand, a FORS wavefunction generally corresponds to an incomplete CI, in the sense that only a subset of configuration (or determinant) space is included. Therefore, one also optimizes the molecular orbital coefficients to self-consistency. The calculation of a full CI wavefunction is extremely computationally demanding, scaling exponentially with the number of atomic basis functions. Its use is therefore currently limited to either very small molecules with extended basis sets or molecules of modest size with modest basis sets. Therefore, even though new algorithms and computer hardware are expanding the range of applicability of full CI wavefunction, the FORS/CASSCF multi-configurational self-consistent-field (MCSCF) method is very powerful for a variety of applications. The method is especially important when one encounters near degeneracies. There are a great many phenomena in chemistry and related fields in which near degeneracies occur. Examples include:

- crossings or near-crossings of potential energy surfaces, in which non-adiabatic interactions become important;
- diradicals, such as those that occur during bond-breaking processes and near transition states in chemical reactions;
- atomic and molecular species that have multiple close-lying electronic states, such as unsaturated transition metal compounds and high energy metastable compounds.

Because of the associated resource (e.g., memory and disk) requirements for MCSCF calculations, the size of the active space is effectively limited to approximately 16 active electrons in 16 active orbitals. This seems somewhat limiting at first thought, since such a (16,16) active space is only slightly larger than a full valence active space for a

molecule like ethane. However, in many chemical processes that involve large molecules only a small number of electrons pairs are changing in any significant manner. So, MCSCF methods can frequently be applied to rather large problems, as long as the active space remains tractable. In recent years FORS calculations have been applied to such complex species as 7-azaindole (7AI)<sup>21</sup>, 7-AI interacting with water in ground and excited electronic states<sup>22</sup>, the 7-AI dimer<sup>23</sup>, large clusters that simulate the Si (100) surface, including up to five dimers (Si<sub>13</sub>H<sub>28</sub>)<sup>24</sup>, and the oxidation and etching reaction mechanisms of these silicon clusters<sup>25</sup>. Such large applications are made possible, in part, by a suite of MCSCF algorithms in GAMESS that render the iterative process more effective and efficient. These methods range from very simple first order methods that have small resource requirements and require more iterations to highly resource-demanding full Newton-Raphson methods that are rapidly convergent but considerably more resource-demanding. A compromise method employs an approximate second order iterative (SOSCF<sup>26</sup>) procedure that is the default convergence option in GAMESS.

Even with efficient algorithms, MCSCF calculations eventually run out of steam as the size of the required active space increases. In order to expand the sizes of active spaces that can be included in a molecular calculation, Ivanic has developed and implemented into GAMESS the ORMAS (occupation restricted multiple active spaces) method<sup>27</sup>. In the spirit of methods like RASSCF (restricted active space SCF<sup>28</sup>) and QCAS (quasi CAS)<sup>29</sup>, ORMAS divides the desired active space into multiple active subspaces that are chemically separable, thereby expanding the effective size of the orbitals and electrons in an active space. The advantage of the ORMAS approach is the complete flexibility in the number of active spaces and the manner in which they are defined. **Figure 1** illustrates the efficiency of the ORMAS method on an important biological compound, oxo(Mn)Salen<sup>27</sup>. It is clear from this figure that subdividing the full CASSCF (12,11) active space six subspaces results in errors on the order of only 1-3 mh in the total energy and 2 kcal/mol in the dissociation energy for different electronic spin states. This is extraordinary, given the fact that ORMAS requires two orders of magnitude fewer determinants than the full CASSCF!

Since, as noted above, a full CI provides the exact wavefunction for a given atomic basis, it is useful to expand the size of molecules for which a full CI wavefunction is accessible. Because of the large resource demands for full CI calculations, it is therefore important to devise ways in which one can approach the accuracy of a full CI wavefunction with a significantly reduced effort. The Ruedenberg group has devised two approaches that achieve this goal, both of which have been implemented in GAMESS. In the first, Ruedenberg, Ivanic and Bytautas have used the full CI code and a systematic analysis of single, double, triple, ... excitations to develop a general method for eliminating the “deadwood” from the full CI wavefunction<sup>30</sup>. Making use of localized MCSCF orbitals (LMOs), they have shown for several test cases that roughly 90% of the configurations in a full CI list can be eliminated while retaining millihartree accuracy. Figure 2 illustrates the effectiveness of this approach by comparing the model energies vs. the known CCSD(T) correlation energies for 38 small to moderate size molecules, with a mean absolute deviation of less than 3 mh. The error in *relative* energies, for example for chemical reactions, is likely to be much less.

The use of localized orbitals leads to the recognition that interactions that are further from each other than vicinal (three bonds) can be safely ignored in a correlation calculation. The extension of a general CI method such as that developed by the Ruedenberg group to a general MCSCF method in which the molecular orbital coefficients, as well as the CI coefficients, is non-trivial, but it has been accomplished and implemented into GAMESS in collaboration with the Ruedenberg group<sup>31</sup>. It is important to recognize that there are clear advantages and some disadvantages to the general MCSCF approach. The obvious advantage is the dramatic reduction in computation time that one attains by eliminating most of the configurations. What one gives up in this approach is the built-in size-consistency that is guaranteed by the complete active space approach. Until the method has been extensively tested, it is not clear how serious a matter this is. Indeed, it is possible that eliminating essentially non-contributing configurations has only a small effect on size-consistency. Similarly, it is not clear how the MCSCF convergence will be affected when a complete active space is not used, but one suspects that the use of less than a FORS active space will lead to deterioration of the MCSCF convergence.

An alternative approach to achieving full CI at a reduced computational cost has been developed by Bytautas and Ruedenberg<sup>32</sup>. This method is built upon a careful analysis of single, double, triple, ... excitations, and the observation of nearly linear relationships between lower and higher excitations. These linear relationships allow one to use extrapolation techniques to very accurately predict the energies due to higher excitations without performing the actual calculations. The impact of this method is that one can predict the full CI energy for a given atomic basis to sub millihartree accuracy. Combined with an extrapolation to the complete basis set limit, this method is capable of predicting bond energies to within chemical (1 kcal/mol) accuracy. The method has already been applied to the first row homonuclear diatomics<sup>32</sup>.

In general, an MCSCF calculation is based on a compact active space and provides the correct zeroth order wavefunction for a chemical process, such as bond dissociation or a more complex chemical reaction. In this sense, the MCSCF wavefunction may be thought of as a replacement for Hartree-Fock (HF) when a single electron configuration cannot provide a qualitatively correct description of a process of interest. That is, an MCSCF calculation accounts only for “non-dynamic” correlation. To obtain accurate energies, and sometimes accurate molecular structures, one must add dynamic correlation, just as one must add dynamic correlation to a HF wavefunction when accurate structures and energies are desired. Several methods are available in GAMESS for recovering dynamic correlation on top of MCSCF wavefunctions. These include multi-reference (MR) configuration interaction (MRCI) and multi-reference second order perturbation theory (MRPT2). The MRCI wavefunctions can include either single or single and double excitations out of the active space, referred to as first and second order CI, respectively. These methods may be applied to both FORS and ORMAS wavefunctions. The version of MRPT2 in GAMESS is that developed by Hirao, Nakano and co-workers. When only one electronic state is included in the calculation, the method is referred to as MRMP2<sup>33</sup>. For multiple state calculations, the appropriate method is second order quasi-degenerate perturbation theory (MCQDPT2<sup>34</sup>). A multi-reference perturbation theory that is built upon ORMAS wavefunctions is in progress.

Whereas both FORS and full CI wavefunctions are size consistent and size extensive, none of these multi-reference methods for recovering dynamic correlation have this property. Interestingly, it appears that both the choice of the unperturbed Hamiltonian  $H_0$  and the manner in which the excitations are included (configuration state functions (CSF) vs. internally contracted configurations (ICC)) impacts how close to size consistency a particular MRPT method is. The ICC approach appears to provide a much more nearly size consistent result than does the choice of CSF<sup>35</sup>. The MRMP2 and MCQDPT2 methods in GAMESS are based on CSFs.

*2.2 Many-Body Methods.* All of the methods discussed above are based on a multi-reference (MR) approach to obtaining wavefunctions and properties. Such MR approaches are often necessary, because many chemical problems involve species with considerable configurational mixing due the existence of near degeneracies (diradical character). However, the amount of diradical character in a chemical system can span a very broad range, from essentially zero (e.g., HOMO occupancy  $\sim 2$ , LUMO occupancy  $\sim 0$ ) to fully diradical (HOMO occupancy  $\sim 1$ , LUMO occupancy  $\sim 1$ ). As one approaches fully diradical character, all single reference methods break down, but they do not break down at the same rate as one approaches this limit. In particular, there is considerable evidence that coupled cluster (CC) methods, particularly those like CCSD(T) that incorporate a triples correction, can overcome the deficiency of a single reference wavefunction for problems with non-trivial diradical character. This has been demonstrated, for example, by examining the  $N_2$  dissociation curves for MP2 and CCSD(T) vs. various MR methods<sup>36</sup>. The breakdown in the CCSD(T) calculation appears much later in the dissociation process than does the MP2 breakdown. Recent developments by Piecuch and co-workers<sup>37</sup> are particularly exciting, since they extend this breakdown even further out in the dissociation curve.. Termed re-normalized and completely re-normalized methods (e.g., R-CCSD(T) and CR-CCSD(T)), these methods are designed to account for an increasing amount of diradical character. Although they do eventually break down at large distances for multiple bonds, they are clearly more robust for intermediate cases. The full suite of closed shell CC, R-CC and CR-CC methods are now available in GAMESS. Since the coupled cluster methods are not fully

variational, one must use the relaxed density<sup>38</sup> that requires the gradient of the wavefunction (non-Hellman-Feynman term) to evaluate molecular properties such as the dipole moment and electrostatic potential. The formulation of the relaxed density is very similar to that of the energy gradient, so the evaluation of properties is considerably more complex for non-variational methods than is it for variational methods. The Piecuch group has completed the derivation and coding of the relaxed density and associated properties for the coupled cluster and renormalized coupled cluster methods<sup>39</sup>. These new features will be in a new GAMESS release shortly.

Even though single reference second order perturbation theory (MP2) cannot describe bond breaking very well, this level of theory is still the most efficient *ab initio* approach that includes electron correlation<sup>40</sup>. For closed shells, the restricted MP2 energy is the standard choice, while for open shells there are two possible choices, unrestricted and restricted MP2. Unrestricted methods (UMP2) have the advantage that the orbitals and orbital energies are well defined, but these methods are generally not spin correct. It is common to encounter modest “spin contamination”, in which the spin expectation value differs from the correct value by 0.05 – 0.1. However, during bond breaking processes, this deviation can become very large, so that the identity of the spin state is effectively lost. The orbitals and orbital energies in restricted open shell Hartree-Fock are not uniquely defined. Consequently, there are several ways in which restricted open shell second order perturbation theory can be formulated. The most popular of these methods, referred to as RMP2 was independently developed by Bartlett and co-workers<sup>41</sup> and Pople, Handy and co-workers<sup>42</sup>. This method is spin correct only through second order in the energy. A more complex and more rigorously spin correct method, called Z-averaged perturbation theory (ZAPT2), was developed by Lee and Jayatilaka<sup>43</sup>. Both methods are available in GAMESS, while the default is ZAPT2.

### 2.3 Excited States, Non-Adiabatic and Relativistic Methods.

The simplest approach to the calculation of excited electronic states is to perform a singles configuration interaction (CIS) calculation. While the accuracy of CIS is limited, especially if an excited state of interest has significant contributions from doubly excited



configurations, such as charge transfer states, it is a computationally efficient method that can frequently provide at least useful qualitative trends correctly. The new CIS code in GAMESS (developed by Webb<sup>44</sup>) provides energies and analytic gradients, so one can predict excited state geometries, determine transition states, and follow reaction paths at this level of theory with approximately the computational cost of Hartree-Fock. At the other end of the computational complexity spectrum, one can use the full CI option to obtain energies and wavefunctions for both ground and excited states. While this level of theory provides the exact wavefunction for the atomic basis used, it is generally limited to atoms and small molecules. For this reason, the Bytautas-Ruedenberg method for extrapolating to the full CI solution is very exciting, since the computational complexity is reduced by several orders of magnitude. As noted in Section 2.1, this method is capable of predicting ground state dissociation energies of diatomic molecules to within chemical accuracy. The ability of this method to predict full potential energy curves and surfaces is currently under investigation, and the ability of this method to treat excited states with equivalent accuracy is promising.

There are several choices for excited state calculations that are intermediate in accuracy and computational complexity between singles CI and full CI. These include the MCSCF, multi-reference CI (called first order or second order CI in GAMESS) and second order multi-reference perturbation theory methods discussed above. Excited states may be analyzed for each of these methods simply by requesting multiple roots. The most recent excited state method to be implemented in GAMESS is the suite of equations-of-motion (EOM) coupled cluster methods that have been developed by Piecuch and co-workers<sup>37a,45</sup>. EOM-CCSD and EOM-CCSD(T) and their completely renormalized analogs (using the method of moments to obtain triples corrections) are available. The EOM methods are usually initiated by performing a CI singles (CIS) calculation to obtain the starting wavefunction. This can be somewhat limiting if the state(s) of interest have significant contributions from two-electron excitations. Using a full singles + doubles CI (CISD) would be prohibitively expensive in most cases. The EOM-CC implementation in GAMESS permits an alternative approach called CISd, in which a subset of double excitations is added to the CIS wavefunction. The choice of

which double excitations to include is flexible, but the most sensible approach is to treat the choice of doubles in a manner that is similar to choosing an MCSCF active space. That is, one would choose that subset of orbitals that are most likely to be involved in the excited state(s) of interest.

While the electronic ground state is usually, although not always, well separated from excited electronic states, there are frequently multiple excited states within a small energy gap. When electronic states come close to, or cross, each other, the adiabatic assumption can break down. Non-adiabatic effects can be manifested in either spin-orbit coupling (a fundamentally relativistic effect) or derivative (sometimes called vibronic) coupling. The former phenomenon arises from the interactions of orbital angular momentum and the magnetic moment due to spin, while the latter derives from a breakdown in the Born-Oppenheimer approximation according to which one ignores the changes in the electronic wavefunction that are induced by nuclear motion. Several methods for dealing with spin-orbit coupling have been implemented in GAMESS, while derivative coupling codes are planned for the near future.

The array of methods in GAMESS for treating spin-orbit coupling effects has recently been the subject of two reviews<sup>46,47</sup>. These methods include the full Breit-Pauli spin-orbit operator and approximations to it, primarily developed by Koseki and Fedorov. All of the methods require a multi-reference wavefunction as a starting point. This can be MCSCF, first or second order CI or MRPT2. The simplest method is a semi-empirical Breit-Pauli implementation, in which the most time-consuming two-electron part of the spin-orbit operator is neglected. This approximation is made viable by replacing the nuclear charge  $Z$  in the one-electron operator by an effective nuclear charge  $Z_{\text{eff}}$  that is fitted to atomic spectra. This method has been parametrized for virtually the entire periodic table, for both all electron and effective core potential basis sets, and then applied to a number of interesting problems<sup>48</sup>. Of course, the accuracy of this semi-empirical approach is inherently limited. For greater accuracy, one can use the full Breit-Pauli approach in which both one- and two-electron operators are included. An interesting and effective compromise is the partial two-electron (P2E) method<sup>49</sup>, in which

only the core-valence two-electron term is included. This method is accurate because the core-valence term is the most important one, and it is computationally efficient because this term can be reformulated to resemble an effective one-electron interaction.

Scalar relativistic effects (e.g., mass-velocity and Darwin-type effects) can be incorporated into a calculation in two ways. One of these is simply to employ effective core potentials (ECPs), since the core potentials are obtained from calculations that include scalar relativistic terms<sup>50</sup>. This may not be adequate for the heavier elements. Scalar relativity can be variationally treated by the Douglas-Kroll (DK)<sup>51</sup> method, in which the full four-component relativistic *ansatz* is reduced to a single component equation. In GAMESS, the DK method is available through third order and may be used with any available type of wavefunction.

*2.4 Properties related to nuclear energy derivatives.* Analytic energy derivatives are as important as the energies themselves. One needs first derivatives for geometry optimizations, reaction path following, dynamics simulations, and (if analytic second derivatives are not available) second derivatives via finite differencing. Second derivatives are necessary for the computation of vibrational frequencies and, subsequently, thermodynamic properties via the appropriate partition functions.

Analytic first derivatives (gradients) are available in GAMESS for open and closed shell Hartree-Fock, open and closed shell density functional theory, GVB (generalized valence bond), MCSCF and ORMAS, CIS, and MP2 for closed and (unrestricted) open shells. Analytic second derivatives (Hessians) are available for open and closed shell HF, GVB and MCSCF wavefunctions. One can also calculate Hessians using numerical finite differencing of analytic gradients using any method for which analytic gradients are available. Fully numerical first and second derivative codes have very recently been implemented, so one can optimize molecular geometries with any level of theory. Fully numerical derivatives are inherently less efficient than analytic derivatives, however, the numerical derivatives are more amenable to very efficient coarse-grained parallel computing.

The most common method for determining vibrational frequencies is the normal mode analysis, based on the harmonic force constant matrix of energy second derivatives (Hessians). Of course, vibrations are not truly harmonic, and the anharmonicity generally increases as the frequency of the vibration (steepness of the potential) decreases. That is, the more anharmonic a motion is, the less applicable is the traditional approach to determining vibrational frequencies. In such cases, one can use the vibrational self-consistent field (VSCF) method to obtain anharmonic frequencies<sup>52,53,54</sup>. This method uses energies and gradients along a given normal mode direction to obtain the anharmonic cubic and quartic terms. In analogy with electronic structure theory, the VSCF method can be augmented by a CI or perturbation theory, in order to obtain improved vibrational frequencies. The VSCF and related methods in GAMESS have been developed by Chaban, Matsunaga, and Taketsugu.

Energy derivatives are essential for the computation of dynamics properties. There are several dynamics-related methods available in GAMESS. The intrinsic reaction coordinate (IRC) or minimum energy path (MEP) follows the infinitely damped path from a first order saddle point (transition state) to the minima connected to that transition state. In addition to providing an analysis of the process by which a chemical reaction occurs (e.g., evolution of geometric structure and wavefunction), the IRC is a common starting point for the study of dynamics. Example are variational transition state theory (VTST<sup>55</sup>) and the modified Shepard interpolation method developed by Collins and co-workers<sup>56</sup>. Indeed, interfaces are maintained between GAMESS and the VTST codes (GAMESS-PLUS)<sup>57</sup> and between GAMESS and the Collins Grow code (GAMESS-Grow)<sup>58</sup>. The GAMESS-Grow interface, in particular, benefits greatly from the availability of analytic energy derivatives. Since Grow builds global potential energy surfaces and GAMESS has multi-state capability via its multi-reference methods, this interface can build multiple potential energy surfaces, a precursor for the dynamical analysis of such phenomena as conical intersections.

In addition to the IRC functionality, GAMESS also has a direct dynamics capability, the dynamic reaction path, DRC<sup>59</sup>. The DRC allows one to perform dynamics “on-the-fly”, by performing classical trajectories at any level of theory for which analytic gradients are available. One can, for example, put an amount of energy equal to  $n$  quanta into any vibrational mode(s), in order to model mode specific chemistry.

The gradient extremal is another trajectory, available in GAMESS, defined as that curve for which the nuclear gradient vector is an eigenvector of the nuclear Hessian<sup>60</sup>. Gradient extremals may lead from minima to any of the following: dissociation products, minima, transition states, or higher order saddle points (stationary points); or united atom collisions. For small numbers of atoms, tracing all gradient extremals may lead to a global understanding of all relevant parts of the potential energy surface<sup>61</sup>.

## *2.5 Other Properties*

Although the generation of wavefunctions and corresponding potential energy surfaces is the primary use for quantum chemistry, other molecular properties are also of interest. In addition to the set of electrostatic properties and localized orbitals described earlier, GAMESS now contains a number of other useful properties or analysis procedures.

Analytic computation of IR intensities<sup>62</sup> and numerical computation of Raman intensities<sup>63</sup> for the harmonic vibrational spectra have been programmed. The analytic computation<sup>64</sup> of closed shell static or frequency dependent polarizabilities is possible, and numerical computation<sup>65</sup> of the static polarizabilities of most wavefunctions can be performed. A program to explore solvent effects on NMR shifts has been developed<sup>66</sup>.

Analysis of molecular wavefunctions is less directly tied to experiment but is of considerable interest to chemists looking for trends or explanations. Atomic charges are of considerable interest to force field developers, and a procedure for fitting them to the computed electrostatic potential and total molecular dipole and quadrupole is available<sup>67</sup>. The localized orbital programs have been extended to include a detailed orbital analysis

of energy contributions, called the Localized Charge Distribution, available for closed shell SCF and MP2 wavefunctions<sup>68</sup>. As part of the EFP solvent project (see below), the Stone distributed multipole analysis<sup>69</sup> and the Kitaura-Morokuma or Stevens-Fink dimer energy analysis<sup>70</sup> have been added, and generalized to any number of weakly interacting monomers.

### 3. Scalable Electronic Structure Theory

One approach to growing the size of a chemical system that can be realistically treated by the most sophisticated electronic structure methods is to devise new methods that are inherently more efficient, as discussed in the previous section. Another, complementary approach is to devise algorithms in such a manner that the calculations are scalable; that is, the computationally most demanding tasks may be distributed among whatever processors are available. Often referred to as parallel programming, this approach is *relatively* straightforward for low-level methods like Hartree-Fock and density functional theory energies and gradients using a replicated data (RD) algorithm, in which the necessary data is either replicated on each available processor (e.g., density and Fock matrices) or recomputed on the fly each iteration (e.g., “direct” calculation of the two-electron integrals). However, the design of scalable algorithms becomes increasingly complicated for the more sophisticated, correlated methods. The disadvantage of the RD approach is that although a calculation proceeds more rapidly than it would on a single processor, the feasible size of a chemical system is limited by the amount of memory and disk on the smallest node. Therefore, the RD approach is sensible when only two-dimensional matrices are involved, but becomes much less viable for correlated methods for which the four-dimensional electron repulsion integrals must be manipulated (i.e., transformed between the AO and MO basis).

A major advance in the manner in which QM (especially correlated QM) calculations may be performed on parallel computers was provided by the development at PNNL of the global array (GA) tools<sup>71</sup>, a one-sided message passing library that facilitates the distribution of large sets of data across all available nodes. The development of the

distributed data interface (DDI)<sup>72</sup> in GAMESS, initially led by Fletcher and more recently by Olson, benefited considerably from the prior development of GA technology. DDI performs best when it can take advantage of the SHMEM library, especially on Cray or SGI systems, but it has also been very successful on IBM computers running under LAPI and clusters of UNIX and Linux computers. The point-to-point messages required for the implementation of DDI on such hardware are carried by TCP/IP socket messages or, sometimes, an MPI-1 library. In a related development, DDI now makes optimal use of the processors in an SMP box by avoiding message passing between CPUs in a given node<sup>73</sup>, by using System V memory regions accessible by all processes within a multiprocessor node.

The initial implementation of DDI was for closed shell MP2 energies and gradients<sup>74</sup>. This has been extremely successful. As long as the size of the system of interest is increased as the number of CPUs is increased, the method scales almost linearly up through 512 T3E processors<sup>75</sup>. For species with unpaired electrons, the implementation of restricted open shell energies is equally efficient, and the UMP2 energy and gradient code<sup>76</sup> scales as well as the closed shell analog. Restricted open shell gradients using the ZAPT *ansatz* have been derived<sup>43d</sup>, and the coding of both sequential and parallel codes is in progress. DDI has also been used to develop a distributed parallel MR second order perturbation method in collaboration with the Koseki group<sup>77</sup>. It appears that the parallel MRMP2 method currently scales well up to about 32 processors. Of course, the most demanding many-body method in GAMESS is the set of coupled cluster methods. The coupled cluster codes in GAMESS are currently sequential, but the development of parallel coupled cluster methods is in progress.

Since MCSCF is an important starting point for so many chemical problems, it is very important to develop parallel MCSCF methods as well. The initial attempt at this was a RD approach which scaled well only to ~4-8 processors<sup>78</sup>. Very recently, a DD parallel MCSCF algorithm has been developed using the full Newton-Raphson convergence algorithm<sup>79</sup>. This DD MCSCF method addresses the integral transformation and orbital rotation steps, but not the CI coefficient optimization, which is discussed below. Initial

tests suggest that this algorithm will scale well up to ~32-64 processors, a major advance over the RD algorithm. Particularly exciting is the very recent implementation of a parallel MCSCF analytic Hessian code<sup>80</sup>. Analytic derivatives are generally preferable to numerical finite differencing schemes, and this is especially so for MCSCF wavefunctions. The reason for this is that numerical schemes usually decrease the molecular symmetry upon some subset of the coordinate displacements. Even for high symmetry species, some displacements reduce the symmetry all the way down to  $C_1$ . So, for example, if one is interested in an excited state that belongs to an irreducible representation that is different from that of the ground state (say,  ${}^1B_2$  vs.  ${}^1A_1$  in  $C_{2v}$  symmetry), both states will become  ${}^1A$  in  $C_1$  symmetry. Because of this, one will frequently encounter root flipping during a numerical Hessian calculation, such that the  ${}^1B_2$  state flips to the  ${}^1A_1$  state when the symmetry is reduced to  $C_1$ . This is avoided if the Hessian calculation is done analytically. The implementation of an analytic Hessian code is also an important step in the development of a derivative (i.e., vibronic) coupling code that is in progress. Derivative coupling plays an important role in the chemistry and physics of excited states, since the crossing of potential energy surfaces and the associated conical intersections are common phenomena in photochemistry, photophysics and photobiology. Since excited state calculations can be very computationally intensive, especially for molecules that are important in biochemistry, the availability of an MCSCF Hessian code that scales well with the number of processors is also important. The scalability of the MCSCF analytic Hessian code in GAMESS, developed by Dudley, is illustrated in **Figure 3** for  $H_2C_4O$ , using a 10 electrons in 9 orbitals active space.

As noted in the Introduction, the ultimate wavefunction for a given basis is the full CI wavefunction, so it is important to extend the sizes of chemical species that can be realistically approached using full CI. Equally important is the recognition that a full CI within a specified set of orbitals and corresponding electrons is just a FORS/CASSCF wavefunction. So, the development of a scalable Full CI method serves a dual purpose. Both RD and DD full CI codes have been developed and implemented into GAMESS by Gan<sup>81</sup>. The algorithm uses a CI driven approach, in which communication is controlled by a string-driven method. The success of the DD/FCI method is especially encouraging,



as is illustrated in **Figure 4**. This figure demonstrates a test on a cluster of 64-bit IBM Power 3II dual processor computers running AIX. The illustrated problems are CH<sub>3</sub>OH (14,14) and H<sub>2</sub>O<sub>2</sub> (14,15), where the numbers in parentheses signify the number of electrons and orbitals, respectively. These problems include ~11,800,000 and 40,400,000 determinants, respectively, and the scalability through 32 processors is excellent. Similar performance is observed on Linux clusters up through the 16 processors that were available for testing.

One can think of the parallel methods discussed above as fine-grained parallelism, in that each subtask in a single energy or energy + gradient evaluation is individually distributed among available processors. There are also problems for which a very coarse-grained approach is appropriate. Examples are the computation of numerical derivatives (e.g., gradients and Hessians) for which each displaced geometry is separate from the others, and all displacements may be identified at the beginning of the calculation. Other examples are multiple Monte Carlo simulations and multiple classical trajectories, since the associated energy and gradient evaluations are independent of one another. A development underway in GAMESS is the GDDI (generalized DDI) method which makes use of the concept of groups and subgroups (in a computational science sense) to make use of both fine-grained and coarse-grained parallelism<sup>82</sup>. For example, if one wishes to perform a CCSD(T) fully numerical Hessian, one can distribute the large number of CCSD(T) energy evaluations among all available nodes. At the same time, if each node is a multi-processor (e.g., SMP) computer, each CCSD(T) energy calculation can itself, in principle be distributed among the processors on its node.

Of course parallel algorithms are only one way in which one can improve the scalability of a particular method. Two other approaches that have recently been implemented into GAMESS are linear scaling methods for HF and DFT<sup>83</sup> and the fragment molecular orbital (FMO) method<sup>84</sup>. Linear scaling methods are based on the premise that the further away atoms are from each other, the more their mutual interactions become similar to classical electrostatic interactions that can be accounted for by multipolar expansions. The key is to be able to devise a procedure in which the linear scaling interaction is turned on when it is appropriate. Among the leaders in this field have been

Head-Gordon<sup>85</sup>, Scuseria<sup>86</sup>, and others<sup>87</sup>. The algorithm in GAMESS, developed by Choi, follows the lead of Wilhite and Head-Gordon<sup>85</sup>, but implements an optimization scheme that determines the optimal parameters in a self-consistent manner, depending on the desired accuracy. The latter is an input parameter, so that one can design the accuracy to fit the application. The linear scaling method in GAMESS is also scalable in the parallel sense, with nearly linear scaling with the number of processors up to at least 16 nodes<sup>83b</sup>.

The basic idea of the FMO method is to take advantage of the observation that exchange interactions and self-consistency in molecular calculations are generally local phenomena. This is even more true for extended systems, such as clusters, crystals and polymers. This locality permits one to treat interactions between remote parts of a system by purely Coulombic operators, ignoring exchange interactions. One can therefore perform the usual quantum calculations on separate units (fragments) in the near field regions, while fragment-fragment interactions are reduced to Coulombic interactions. In practice, one divides the system of interest into fragments and performs *n*-mer calculations of these fragments in the Coulomb field of the other fragments. There are no empirical parameters, and the only approximation is the fragmentation itself. So, in this sense, there are common features with the linear scaling method discussed in the previous paragraph. The fragmentation scheme introduces little error as long as the fragments are chosen in a physically sensible manner. The most important impact on accuracy is the size of the fragments. The FMO method has been enabled by Fedorov for SCF<sup>84b</sup>, DFT<sup>84c</sup>, and MP2<sup>84d</sup> methods at present. Parallelization of the FMO method is implemented by assigning each monomer, dimer, or trimer computation to a subgroup of the full processor set<sup>84a</sup>.

#### **4. QM/MM Methods**

Even with the most clever and efficient methods and scalable algorithms, as the size of the system of interest grows, sooner or later the available compute power is not up to the task if one uses fully QM methods, especially correlated ones. Two important areas of research that fall into this category are solvent effects (more generally liquid behavior)

and surface science. An effective alternative to fully QM methods is the combination of QM with molecular mechanics (MM) methodology. MM is a term that generally suggests that one is using classical techniques with no wavefunction; such methods vary broadly in sophistication. Two types of MM methods that are very different in their level of sophistication are discussed here.

*4.1 Discrete Solvent Approaches.* The approach taken in GAMESS to study solvation is a multi-layer one in which the innermost layer consists of the solute plus some number of solvent molecules that one feels must be treated with explicit QM. Examples of the latter are water molecules that act as conduits in H-transfer reactions. The second layer consists of a sophisticated potential, the effective fragment potential (EFP) that is derived from rigorous QM<sup>88</sup>. The outermost layer is represented by a reliable continuum method to represent the bulk liquid. In its original EFP1/HF version, this method described solvent molecules (i.e., water) by three terms that are added to the QM (i.e., HF) Hamiltonian. The first term represents the Coulomb interactions by a distributed multipole analysis (DMA) expanded through octopoles. The entire Coulomb term is multiplied by a distance-dependent cutoff to account for overlapping charge distributions. The second, induction, term accounts for the polarization of charge densities in a self-consistent manner using localized molecular orbitals (LMOs). The third term is fitted to the remainder of the HF water dimer potential, determined from 192 points on the dimer surface, after subtracting the Coulomb and induction contributions. This remainder represents exchange repulsion and (implicitly) charge transfer. This EFP1/HF method has been very successful for problems that are well described by the HF method, but it is limited in two respects. First, HF includes no electron correlation which invades all of the terms mentioned above and introduces entirely new interactions, most notably dispersion. Second, the process of fitting to obtain the exchange repulsion/charge transfer term is not something one wants to do for every solvent or liquid of interest.

The first of these limitations has been partially addressed by reformulating the EFP1 approach with density functional theory (DFT), using the popular B3LYP functional<sup>89</sup>. Denoted EFP1/DFT, this method, developed by Adamovic, includes some correlation, although not long-range dispersion, and therefore produces much better binding energies,

for example, in water clusters. So, this approach only partially accounts for the correlation problem. Likewise, an MP2 EFP1 analog has been developed by Song in a similar manner, except that the dispersion contribution is fitted in a separate term using the usual  $C_n/R^n$  expansion<sup>90</sup>. So far, only the fragment-fragment interaction has been coded, while the fragment-QM interaction is in progress. So far, the EFP1/MP2 method looks very promising: A molecular dynamics simulation using this method maps almost exactly onto the experimental  $g(OO)$ ,  $g(OH)$  and  $g(HH)$  spectra, whereas the EFP1/HF method does very poorly and EFP1/DFT not nearly as well as EFP1/MP2.<sup>91</sup> A parallel code has also been developed by Netzloff for all of the EFP1 methods. The scaling is very good as long as a sufficient number of fragments is included in the calculation.

All of the EFP1 methods involve fitting the repulsive and dispersion terms, so it is difficult to extend the method to other species, such as solvents that may be of interest to organic or inorganic chemists. In this sense, it is desirable to derive the exchange repulsion, dispersion, and charge transfer from “first principles” instead of employing fitting procedures. This has been accomplished for the exchange repulsion by expanding this interaction as a power series in the intermolecular overlap. This is not a new idea, but combining this approach with highly transferable LMOs to calculate these integrals and the related intermolecular kinetic energy integrals has been very successful for a wide variety of solvents<sup>88</sup>. The exchange repulsion calculated by this method maps the exact HF intermolecular exchange typically to within 0.5 kcal/mol. More recently, the dispersion interaction has been derived by Adamovic, again from first principles, based on the frequency-dependent polarizabilities in the imaginary frequency range obtained from the time-dependent Hartree-Fock method.<sup>92</sup> In keeping with the other terms in the interaction energy, the dispersion is obtained in a distributed manner, by summing over individual interactions between localized molecular orbitals located on the two interacting fragments to obtain the final  $C_6$  terms. Combined with an estimate of the  $C_8$  term, this provides an accurate evaluation of the dispersion energy. An important feature of this new EFP2 implementation is that one can generate an effective fragment for any species, simply by performing a single “MAKEFP” run to generate all of the pieces for a given fragment. Indeed, performing calculations on mixed species (e.g., methanol-water)

is just as simple, because no parameter fits are required. At this time, EFP2 is primarily an EFP only method, since not all of the QM-EFP interactions have been developed, but these terms are in progress. In addition, the EFP2 method has been extended by Jensen and co-workers<sup>93</sup> to the treatment of intramolecular covalent interactions, in addition to intermolecular interactions. This is accomplished by interfacing a “buffer region” constructed from frozen localized orbitals that connects the QM part of the molecule with the EFP part. This allows one to study very large molecules (e.g., proteins and polymers) by treating the “action region” with quantum mechanics and the remainder of the system with EFP. This is similar to traditional QM/MM methods, except that the EFP provides a much more sophisticated classical potential that captures most of the physics.

Although the cost of an EFP calculation is several orders of magnitude smaller than that of a corresponding QM (e.g., HF, DFT, MP2) calculation, the cost can rise considerably if one incorporates a large number of solvent molecules. This cost reflects not only the inherent cost of a single energy + gradient calculation, but also the fact that the number of arrangements of solvent or liquid molecules expands rapidly with the number of molecules. To put this in context, a molecular dynamics run to predict bulk properties of some species should include at least 64 molecules, and 256-512 molecules is considered to be more realistic. As another example, the process of finding a global minimum for some reacting system that consists of a QM “solute” and some number of solvent molecules requires a Monte Carlo or similar calculation that frequently requires tens of thousands of energy evaluations, plus geometry optimizations at a subset of these points. Both Monte Carlo/simulated annealing<sup>94</sup> and molecular dynamics<sup>95</sup> codes have been implemented in GAMESS, combined with the EFP methods. To make such calculations more feasible for several hundred fragments, each term in the EFP method has been made scalable by Netzloff<sup>96</sup>. As for any other application, the scalability relies on the size of the problem: the larger the system, the more effective in the parallel code. There are two scalability considerations with the EFP code. One is the inherent scaling of the EFP code itself. The other is the composite scalability of a QM + EFP calculation. For small numbers of water molecules, the pure EFP calculation does not scale very well, since the time requirements are very small. However, the scalability on up to 16 processors

improves dramatically as the number of EFPs increases<sup>96</sup>. When a QM molecule is present, the inherent QM scalability<sup>1</sup> plus the greater CPU demands of the QM part cause the scalability to improve for even relatively small numbers of solvent molecules.

*4.2 Surface Chemistry.* For surface chemistry, a more traditional QM/MM approach, SIMOMM<sup>97</sup> (surface integrated molecular orbital molecular mechanics), has been developed and implemented in GAMESS. SIMOMM is an embedded cluster approach in which the QM part of the system is embedded into a much larger MM cluster to represent the bulk. Any level of QM theory in GAMESS can be used for the QM part, while the TINKER<sup>98</sup> code is used for the MM part. The interface between the QM and MM parts is represented by link atoms that appear in the QM part (typically as hydrogens, although this is not a requirement) and in the MM part as the actual surface atoms of interest. Gradients for both the QM and MM methods are generally available, so full geometry optimizations are both feasible and recommended. The method has been most extensively applied to problems that involve the Si(100) surface, including addition of organic molecules to the surface<sup>99</sup>, etching<sup>25</sup>, and diffusion of metal atoms along the surface<sup>100</sup>. More recently, it has been applied to the growth of diamond<sup>101</sup> and silicon carbide<sup>102</sup> surfaces.

*4.3 Continuum Solvent Methods.* As noted above, solvent effects may be modeled in GAMESS as discrete solvent molecules using the EFP model. Continuum solution models are also useful, in that these remove the solvent structural problem completely (no need to sample many solvent minima structures). The most sophisticated continuum solvent model in GAMESS is the Polarizable Continuum Model (PCM) of Tomasi and coworkers, which models the molecule as a space-filling surface surrounded by continuum<sup>103</sup>. PCM permits the computation of the effect of solvation on the energy<sup>104</sup>, nuclear gradients<sup>105</sup>, and even frequency dependent polarizabilities<sup>106</sup>. The present PCM implementation has been greatly extended by the Jensen group, to use the conductor-like PCM model<sup>107</sup>, and to treat all SCF wavefunctions from RHF to MCSCF and their DFT counterparts. The numerical solution of the Poisson equation to obtain apparent surface charges has been made possible for large systems by using an iterative linear equation solver<sup>108</sup>. Numerically stable nuclear gradients have been obtained by using an “area

scaling” tessellation<sup>109</sup> that deals with surface elements that are too close to each other by rescaling their sizes, permitting the study of solute geometry optimization. The PCM computations just described are enabled for parallel computation<sup>108</sup>, and they have been interfaced with the effective fragment potential<sup>110</sup>.

Two other approaches are used less often. The original SCRF spherical cavity model<sup>111</sup> is less preferable than the space filling cavity used in the PCM code. The COSMO model has been implemented in GAMESS to describe the electrostatic interactions between the solute and the continuum, for RHF and its corresponding MP2<sup>112</sup>.

## **5. SUMMARY AND PROGNOSIS**

The focus of this paper has been on the new developments in electronic structure theory during the past decade. These developments include new methods in quantum mechanics, including approaches for extrapolating to the full CI and complete basis set limits, novel methods for CASSCF calculations, new coupled cluster techniques, methods for evaluating non-adiabatic and relativistic interactions, new approaches for distributed parallel computing, and QM/MM methods for describing solvent effects and surface science. It is useful to note in this regard that GAMESS is a general purpose suite of electronic structure and QM/MM methods (including open- and closed-shell Hartree-Fock which has been essentially ignored here) that can be run on virtually any computer, cluster, massively parallel system, or for that matter a desktop Mac or PC. Indeed, GAMESS is used at many universities as an educational tool, making use of its graphical back end MacMolPlt<sup>113</sup>. GAMESS and MacMolPlt can be downloaded at no cost from [www.msg.ameslab.gov](http://www.msg.ameslab.gov), with only a simple license required. A great many people from essentially every region of the world have contributed to the development of GAMESS. This is due in no small measure to the philosophy that has stood us in good stead for more than two decades: Science is to be shared openly and freely, and the development of computer codes should be driven by the science that the community wishes to accomplish. This approach to disseminating GAMESS will continue into the foreseeable future.

It is anticipated that several new features will be implemented into GAMESS in the near future. These include:

- Restricted open shell second order perturbation theory (ZAPT2) analytic gradients
- Properties and gradients for the closed shell coupled cluster methods that are already in GAMESS
- Parallel coupled cluster methods
- Derivative coupling at the MCSCF level of theory
- Determinant-based multi-reference perturbation theory
- Model core potentials with the correct nodal behavior
- Interfaces between MP2, CIS, MRMP2, CCSD(T), and EOM-CCSD(T) codes with EFP, in order to facilitate the study of solvent effects on correlated systems and in excited electronic states
- Molecular dynamics capability for both EFP1 and EFP2
- An interface between EFP and MM codes

As always, additional code developments will be driven by the interests of our many colleagues who contribute to GAMESS and by the chemistry that we and our many users wish to study.

### **Acknowledgements**

The work described in this paper was supported by the Air Force Office of Scientific Research via a software development (CHSSI) grant, by the Department of Energy through a SciDAC grant, and by IBM via multiple SUR grants. The authors are most grateful to our many collaborators who have contributed their efforts to GAMESS. We mention in particular Professors Klaus Ruedenberg, Kim Baldridge, Jan Jensen, Shiro Koseki, and Piotr Piecuch, and Drs. Dmitri Fedorov, Graham Fletcher, Joe Ivanic, and Simon Webb, and of course all the present and past graduate students who have contributed to GAMESS.



## References

1. M.W.Schmidt, K.K. Baldrige, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S. Su, T.L. Windus, M. Dupuis, J.A. Montgomery J. Comput. Chem. 14 (1993) 1347.
2. ACES II is a program product of the Quantum Theory Project, University of Florida. Authors: J.F. Stanton, J. Gauss, J.D. Watts, M. Nooijen, N. Oliphant, S.A. Perera, P.G. Szalay, W.J. Lauderdale, S.R. Gwaltney, S. Beck, A. Balkova, D.E. Bernholdt, K.-K. Baeck, P. Rozyczko, H. Sekino, C. Huber, R.J. Bartlett.
3. The Cambridge Analytic Derivatives Package Issue 6, Cambridge, 1995. A suite of quantum chemistry programs developed by R.D. Amos with contributions from I.L. Alberts, J.S. Andrews, S.M. Colwell, N.C. Handy, D. Jayatilaka, P.J. Knowles, R. Kobayashi, K.E. Laidig, G. Laming, A.M. Lee, P.E. Maslen, C.W. Murray, J.E. Rice, E.D. Simandiras, A.J. Stone, M.-D. Su, D.J. Tozer.
4. Dalton, a molecular electronic structure program, Release 1.2 (2001), written by T. Helgaker, H.J.Aa. Jensen, P. Jørgensen, J. Olsen, K. Ruud, H. Agren, A.A. Auer, K.L. Bak, V. Bakken, O. Christiansen, S. Coriani, P. Dahle, E.K. Dalskov, T. Enevoldsen, B. Fernandez, C. Hättig, K. Hald, A. Halkier, H. Heiberg, H. Hettema, D. Jonsson, S. Kirpekar, R. Kobayashi, H. Koch, K.V. Mikkelsen, P. Norman, M.J. Packer, T.B. Pedersen, T.A. Ruden, A. Sanchez, T. Saue, S.P.A. Sauer, B. Schimmelpfennig, K. O. Sylvester-Hvid, P.R. Taylor, O. Vahtras.
5. M.F. Guest, J.M.H. Thomas, P. Sherwood, I.J. Bush, H.J.J. van Dam, J.H. van Lenthe, R.W.A. Havenith, J. Kendrick, Mol. Phys. in press.
6. HyperChem Molecular Modeling Software, Release X.Y, a product of Hypercube, Inc., 1115 NW 4th St. , Gainesville FL.
7. <http://www.scienceserve.com/Software/Schroedinger/Jaguar.htm>
8. MOLCAS. G. Karlstrom, R. Lindh, P.-E. Malmqvist, B.O. Roos, U. Ryde, V. Veryazov, P.-O. Widmark, M. Cossi, B. Schimmelpfennig, P. Neogrody, L. Seijo, Comput. Mater. Sci. 28 (2003) 222.
9. (a) H.-J. Werner and P.J. Knowles, J. Chem Phys. 89 (1988) 5803 (b) P.J. Knowles and H.-J. Werner, Chem. Phys. Lett. 145 (1988) 514. (c) MOLPRO is a package of ab-initio programs written by H.-J. Werner and P.J. Knowles, with contributions from R.D. Amos, A. Bernhardsson, A. Berning, P. Celani, D.L. Cooper, M.J.O. Deegan, A.J. Dobbyn, F. Eckert, C. Hampel, G. Hetzer, T. Korona, R. Lindh, A.W. Lloyd, S.J. McNicholas, F.R. Mandy, W. Meyer, M.E. Mura, A. Nicklass, P. Palmieri, R. Pitzer, G. Rauhut, M. Schutz, H. Stoll, A.J. Stone, R. Tarroni, T. Thorsteinsson.
10. T.P. Straatsma, E. Aprà, T.L. Windus, E.J. Bylaska, W. de Jong, S. Hirata, M. Valiev, M.T. Hackler, L. Pollack, R.J. Harrison, M. Dupuis, D.M.A. Smith, J.J. Nieplocha, V. Tipparaju, M. Krishnan, A.A. Auer, E. Brown, G. Cisneros, G.I. Fann, H. Fruchtl, J. Garza, K. Hirao, R. Kendall, J. Nichols, K. Tsemekhman, K. Wolinski, J. Anchell, D. Bernholdt, P. Borowski, T. Clark, D. Clerc, H. Dachsel, M. Deegan, K. Dyall, D. Elwood, E. Glendening, M. Gutowski, A. Hess, J. Jaffe, B. Johnson, J. Ju, R. Kobayashi, R. Kutteh, Z. Lin, R. Littlefield, X. Long, B. Meng, T. Nakajima, S. Niu, M. Rosing, G. Sandrone, M. Stave, H. Taylor, G. Thomas, J. van Lenthe, A. Wong, Z. Zhang

- "NWChem, A Computational Chemistry Package for Parallel Computers, Version 4.6" (2004), Pacific Northwest National Laboratory, Richland, Washington 99352-0999, USA.
11. P. Pulay, J. Baker, K. Wolinski, <http://www.pqs-chem.com>.
  12. T.D. Crawford, C.D. Sherrill, E.F. Valeev, J.T. Fermann, R.A. King, M.L. Leininger, S.T. Brown, C.L. Janssen, E.T. Seidl, J.P. Kenny, W.D. Allen, *PSI* 3.2, 2003.
  13. QChem. J. Kong, C.A. White, A.I. Krylov, D. Sherrill, R.D. Adamson, T.R. Furlani, M.S. Lee, A.M. Lee, S.R. Gwaltney, T.R. Adams, C. Ochsenfeld, A.T.B. Gilbert, G. S. Kedziora, V.A. Rassolov, D.R. Maurice, N. Nair, Y. Shao, N.A. Besley, P.E. Maslen, J.P. Dombrowski, H. Daschel, W. Zhang, P.P. Korambath, J. Baker, E.F. Byrd, T. van Voorhis, M. Oumi, S. Hirata, C.-P. Hsu, N. Ishikawa, J. Florian, A. Warshel, B. G. Johnson, P.M.W. Gill, M. Head-Gordon, J.A. Pople, *J. Comput. Chem.* 21 (2000) 1532.
  14. See <http://www.wavefun.com>
  15. S. Brode, H. Horn, M. Ehrig, D. Moldrup, J.E. Rice, R. Ahlrichs, *J. Comput. Chem.* 14 (1993) 1142.
  16. T. Yanai, H. Nakano, T. Nakajima, T. Tsuneda, S. Hirata, Y. Kawashima, Y. Nakao, M. Kamiya, H. Sekino, K. Hirao *Computational Science - International Conference on Computational Science 2003, Lecture Notes in Computer Science (Springer)* pp 84-95 (2003).
  17. (a) Y.C. Zheng, J. Almlöf, *Chem. Phys. Lett.* 214 (1993) 397. (b) K.R. Glaesemann, M.S. Gordon, *J. Chem. Phys.* 108 (1998) 9959.
  18. P.-O. Löwdin, *Phys. Rev.* 101 (1956) 1730.
  19. J. Ivanic, K. Ruedenberg, *Theoret. Chem. Acc.* 106 (2001) 339.
  20. M.W. Schmidt, M.S. Gordon, *Ann. Rev. Phys. Chem.* 49 (1998) 233.
  21. M.S. Gordon, *J. Phys. Chem.* 100 (1996) 3974.
  22. G.M. Chaban, M.S. Gordon, *J. Phys. Chem. A* 103 (1999) 185.
  23. P. Bandyopadhyay, T.J. Dudley, M.S. Gordon, in preparation.
  24. Y. Jung, Y. Akinaga, K.D. Jordan, M.S. Gordon, *Theoret. Chem. Accts.* 109 (2003) 268.
  25. C.H. Choi, D.-J. Liu, J.W. Evans, M.S. Gordon, *J. Am. Chem. Soc.* 124 (2002) 8730.
  26. G. Chaban, M.W. Schmidt, M.S. Gordon, *Theor. Chem. Accts.* 97 (1997) 88.
  27. J. Ivanic, *J. Chem. Phys.* 119 (2003) 9364.
  28. J. Olsen, B.O. Roos, P. Jørgensen, H.J.A. Jensen, *J. Chem. Phys.* 89 (1998) 2185.
  29. H. Nakano, K. Hirao, *Chem. Phys. Lett.* 317 (2000) 90.
  30. a) L. Bytautas, K. Ruedenberg, *Mol. Phys.* 100 (2002) 757.  
b) J. Ivanic, K. Ruedenberg, *Theoret. Chem. Accts.* 107 (2002) 220.
  31. J. Ivanic, K. Ruedenberg, *J. Comput. Chem.* 24, 1250 (2003).
  32. L. Bytautas, K. Ruedenberg, *J. Chem. Phys.* 121, 10905 (2004); 121, 10919 (2004); 121, 10852 (2004).
  33. K. Hirao, *Chem. Phys. Lett.* 190 (1992) 374.
  34. H. Nakano, *J. Chem. Phys.* 99 (1993) 7983.
  35. J. Rintelman, I. Adamovic, S. Varganov, M.S. Gordon, *J. Chem. Phys.* 122, 044105 (2005).
  36. M.S. Gordon, M.W. Schmidt, G.M. Chaban, K.R. Glaesemann, W.J. Stevens, C. Gonzalez, C. *J. Chem. Phys.* 110 (1999) 4199.

37. (a) P. Piecuch, K. Kowalski, I.S.O. Pimienta, M. McGuire, *Int. Rev. Phys. Chem.* 21 (2002) 527. (b) P. Piecuch, S.A. Kucharski, K. Kowalski, M. Musial, *Comput. Phys. Commun.* 149 (2002) 71.
38. (a) K. Raghavachari, J.A. Pople, *Int. J. Quantum Chem.* 20 (1981) 1067. (b) G.H.F. Diercksen, B.O. Roos, A.J. Sadlej, *J. Chem. Phys.* 59 (1981) 29. (c) J.E. Rice, R.D. Amos, *Chem. Phys. Lett.* 122 (1985) 583. (d) K.B. Wiberg, C.M. Hadad, T.J. LePage, C.M. Breneman, M.J. Frisch, *J. Phys. Chem.* 96 (1992) 671. (e) A.P.L. Rendell, G.B. Backskay, N.S. Hush, N.C. Handy, *J. Chem. Phys.* 87 (1987) 5976.
39. K. Kowalski, P. Piecuch, in preparation.
40. C.M. Aikens, S.P. Webb, R.L. Bell, G.D. Fletcher, M.W. Schmidt, M.S. Gordon, *Theoret. Chem. Acc.* 110 (2003) 233.
41. W.J. Lauderdale, J.F. Stanton, J. Gauss, J.D. Watts, R.D. Bartlett, *Chem. Phys. Lett.* 187 (1994) 21.
42. P.J. Knowles, J.S. Andrews, R.D. Amos, N.C. Handy, J.A. Pople, *Chem. Phys. Lett.* 186 (1991) 130.
43. (a) D. Jayatilaka, T. J. Lee, *Chem. Phys. Lett.* 199 (1992) 211. (b) T.J. Lee, D. Jayatilaka, *Chem. Phys. Lett.* 201 (1993) 1. (c) D. Jayatilaka, T.J. Lee, *J. Chem. Phys.* 98 (1993) 9734. (d) G.D. Fletcher, M.S. Gordon, R.S. Bell, *Theoret. Chem. Accts.* 107 (2002) 57.
44. S. Webb, unpublished results.
45. K. Kowalski, P. Piecuch, *J. Chem. Phys.* 120 (2004) 1715.
46. D.G. Fedorov, S. Koseki, M.W. Schmidt, M.S. Gordon, *Int. Rev. Phys. Chem.* 22 (2003) 551.
47. D. Fedorov, M.W. Schmidt, S. Koseki, M.S. Gordon, in *Recent Advances in Relativistic Molecular Theory*, Vol. 5, K. Hirao and Y. Ishikawa, Eds., World Scientific, Singapore, pp. 107-136, 2004.
48. (a) S. Koseki, N. Matsunaga, M.W. Schmidt, M.S. Gordon, *J. Phys. Chem.* 99 (1995) 12764. (b) S. Koseki, M.W. Schmidt, M.S. Gordon, *J. Phys. Chem.* 102 (1998) 1043. (c) S. Koseki, D.G. Fedorov, M.W. Schmidt, M.S. Gordon, *J. Phys. Chem. A* 105 (2001) 8262. (d) S. Koseki, Y. Ishihara, H. Umeda, D.G. Fedorov, M.S. Gordon, *J. Phys. Chem.* 106 (2002) 785. (e) S. Koseki, Y. Ishihara, D.G. Fedorov, M.W. Schmidt, M.S. Gordon, *J. Phys. Chem. A* 104 (2004) 4707.
49. (a) D.G. Fedorov, M.S. Gordon, *J. Chem. Phys.* 112 (2000) 5611. (b) D.G. Fedorov, M.S. Gordon, *ACS Symposium Series* 828, *Low-Lying Potential Energy Surfaces*, M.R. Hoffmann and K.G. Dyall, Ed., Oxford Press, 2002.
50. M. Krauss, W.J. Stevens, *Ann. Rev. Phys. Chem.* 35 (1985) 357.
51. (a) M. Douglas, N.M. Kroll, *Ann. Phys.* 82 (1974) 89. (b) T. Nakajima, K. Hirao, *Chem. Phys. Lett.* 329 (2000) 511. (c) T. Nakajima, K. Hirao, *J. Chem. Phys.* 113 (2000) 7286. (d) W.A. DeJong, R.J. Harrison, D.A. Dixon, *J. Chem. Phys.* 114 (2001) 48. (e) T. Nakajima, K. Hirao, *J. Chem. Phys.* 119 (2003) 4105.
52. G.M. Chaban, J.O. Jung, R.B. Gerber, *J. Chem. Phys.* 111 (1999) 1823.
53. (a) K. Yagi, T. Taketsugu, K. Hirao, M.S. Gordon, *J. Chem. Phys.* 113 (2000) 1005. (b) K. Yagi, K. Hirao, T. Taketsugu, M.W. Schmidt, M.S. Gordon, *J. Chem. Phys.* 121 (2004) 1383.
54. N. Matsunaga, G.M. Chaban, R.B. Gerber, *J. Chem. Phys.* 117 (2002) 3541.
55. D.G. Truhlar, B.C. Garrett, *Acc. Chem. Res.* 13 (1980) 440.

56. R.P.A. Bettens, M.A. Collins, *J. Chem. Phys.* 108 (1998) 2424, and references cited therein.
57. <http://comp.chem.umn.edu/gamesplus>
58. H.M. Netzloff, M.S. Gordon, M.A. Collins, *J. Chem. Phys.*, submitted.
59. (a) T. Taketsugu, M.S. Gordon, *J. Phys. Chem.* 99 (1995) 8467.  
(b) M.S. Gordon, G. Chaban, T. Taketsugu, *J. Phys. Chem.* 100 (1996) 11512.
60. (a) D.K. Hoffman, R.S. Nord, K. Ruedenberg, *Theoret. Chim. Acta* 69 (1986) 265.  
(b) J. Sun, K. Ruedenberg, *J. Chem. Phys.* 98 (1993) 9707.
61. K. Bondensgaard, F. Jensen, *J. Chem. Phys.* 104 (1996) 8025.
62. Y. Yamaguchi, M. Frisch, J. Gaw, H.F. Schaefer, J.S. Binkley, *J. Chem. Phys.* 84 (1986) 2262.
63. A. Komornicki, J.W. McIver *J. Chem. Phys.* 70 (1979) 2016.
64. P. Korambath, H.A. Kurtz, in "Nonlinear Optical Materials", ACS Symposium Series 628, S.P. Karna, A.T. Yeates, Eds., pp 133-144, Washington DC, 1996.
65. H.A. Kurtz, J.J.P. Stewart, K.M. Dieter, *J. Comput. Chem.* 11 (1990) 82.
66. M.A. Freitag, B. Hillman, A. Agrawal, M.S. Gordon, *J. Chem. Phys.* 120 (2004) 1197.
67. M.A. Spackman, *J. Comput. Chem.* 17 (1996) 1.
68. J.H. Jensen, M.S. Gordon, *J. Phys. Chem.* 99 (1995) 8091.
69. (a) A.J. Stone, *Chem. Phys. Lett.* 83 (1981) 233.  
(b) A.J. Stone, M. Alderton, *Mol. Phys.* 56 (1985) 1047.
70. W. Chen, M.S. Gordon, *J. Phys. Chem.* 100 (1996) 14316.
71. J. Nieplocha, R.J. Harrison, R.J. Littlefield, *Proc. Supercomputing '94* (1994) 340.
72. G.D. Fletcher, M.W. Schmidt, B.M. Bode, M.S. Gordon, *Comput. Phys. Commun.* 128 (2000) 190.
73. R.M. Olson, M.W. Schmidt, M.S. Gordon, A.P. Rendell, *Proc. Supercomputing 2003*, in press.
74. G.D. Fletcher, M.W. Schmidt, M.S. Gordon, *Adv. Chem. Phys.* 110 (1999) 267.
75. T. Kudo, M.S. Gordon, *J. Phys. Chem. A* 105 (2001) 11276.
76. C.M. Aikens, M.S. Gordon, *J. Phys. Chem. A* 108 (2004) 3103.
77. H. Umeda, S. Koseki, U. Nagashima, M.W. Schmidt, *J. Comput. Chem.* 22 (2001) 1243.
78. T.L. Windus, M.W. Schmidt, M.S. Gordon, *Theoret. Chim. Acta* 89 (1994) 77.
79. G.D. Fletcher, to be published.
80. T.J. Dudley, M.W. Schmidt, R.M. Olson, M.S. Gordon, *J. Comp. Chem.*, submitted.
81. Z. Gan, Y. Alexeev, R.A. Kendall, M.S. Gordon, *J. Chem. Phys.* 119 (2003) 47.
82. D.G. Fedorov, R.M. Olson, K. Kitaura, M.S. Gordon, S. Koseki, *J. Comp. Chem.*, 25, 872 (2004).
83. (a) C.H. Choi, J. Ivanic, M.S. Gordon, K. Ruedenberg, *J. Chem. Phys.* 111 (1999) 8825. (b) C.H. Choi, K. Ruedenberg, M.S. Gordon, *J. Comput. Chem.* 22 (2001) 1484.  
(c) C.H. Choi, *J. Chem. Phys.* 120 (2004) 3535.
84. (a) D.G. Fedorov, R.M. Olson, K. Kitaura, M.S. Gordon, S. Koseki, *J. Comput. Chem.* 25 (2004) 872. (b) D.G. Fedorov, K. Kitaura *J. Chem. Phys.* 120 (2004) 6832. (c) D.G. Fedorov, K. Kitaura *Chem. Phys. Lett.* 389 (2004) 129. (d) D.G. Fedorov, K. Kitaura, *J. Chem. Phys.* 121 (2004) 2483.

85. (a) C. White, M. Head-Gordon, *J. Chem. Phys.* 101 (1994) 6593. (b) C. White, M. Head-Gordon, *Chem. Phys. Lett.* 257 (1996) 647. (c) C. White, M. Head-Gordon, *J. Chem. Phys.* 105 (1996) 5061.
86. (a) M.C. Strain, G.E. Scuseria, M.J. Frisch, *Science* 271 (1996) 5245. (b) J.C. Burant, M.C. Strain, G.E. Scuseria, M.J. Frisch, *Chem. Phys. Lett.* 248 (1996) 43. (c) J.C. Burant, M. Strain, G.E. Scuseria, M.J. Frisch, *Chem. Phys. Lett.* 258 (1996) 45. (d) G.E. Scuseria, *J. Phys. Chem. A* 103 (1999) 4782.
87. (a) M. Challacombe, E. Schwegler, J. Almlöf, *J. Chem. Phys.* 104 (1996) 4685. (b) E. Schwegler and M. Challacombe, *J. Chem. Phys.* 105 (1996) 2726. (c) M. Challacombe, *J. Chem. Phys.* 113 (2000) 10037. (d) J.A. Board, J.W. Causey, J.F. Leathrum, A. Windemuth, K. Schulten, *Chem. Phys. Lett.* 198 (1992) 89. (e) J.M. Perez-Jorda, W. Wang, *J. Chem. Phys.* 104 (1996) 8003. (f) H.G. Petersen, D. Soelvason, J.W. Perram, E.R. Smith, *J. Chem. Phys.* 101 (1994) 8870.
88. M.S. Gordon, M.A. Freitag, P. Bandyopadhyay, J.H. Jensen, V. Kairys, W.J. Stevens, *J. Phys. Chem. A*, 105 (2001) 293.
89. I. Adamovic, M.A. Freitag, M.S. Gordon, *J. Chem. Phys.* 118 (2003) 6725.
90. J. Song, M.S. Gordon, in preparation.
91. H.M. Netzloff, M.S. Gordon, *J. Chem. Phys.* 121 (2004) 2711.
92. I. Adamovic, M.S. Gordon, *Mol. Phys.*, in press.
93. V. Kairys, J.H. Jensen, *J. Phys. Chem. A* 104 (2000) 6656.
94. P.N. Day, R. Pachter, M.S. Gordon, G.M. Merrill, *J. Chem. Phys.* 112 (2000) 2063.
95. H.M. Netzloff, M.S. Gordon, to be published.
96. H.M. Netzloff and M.S. Gordon, *J. Comput. Chem.* 25 (2004) 1926.
97. J. Shoemaker, L.W. Burggraf, M.S. Gordon, *J. Chem. Phys.* 112 (2000) 2994.
98. P. Ren, J. W. Ponder, *J. Phys. Chem. B* 107 (2003) 5933.; see also, <http://dasher.wustl.edu/tinker>
99. (a) C.H. Choi, M.S. Gordon, *J. Am. Chem. Soc.* 121 (1999) 11311. (b) Y. Jung, C.H. Choi, M.S. Gordon, *J. Phys. Chem.* 105 (2001) 4039. (c) C.H. Choi, M.S. Gordon, in *The Chemistry of Organic Silicon Compounds*, Vol. 3, Z. Rappaport and Y. Apeloig, Eds., p. 821 (2001). (d) C.H. Choi, M.S. Gordon, *J. Am. Chem. Soc.* 124 (2002) 8730 (2002). (e) C.H. Choi, M.S. Gordon, in *Computational Materials Chemistry: Methods and Applications*, L.A. Curtiss and M.S. Gordon, Eds., pp. 125-190 (2004). (f) H.S. Lee, C.H. Choi, M.S. Gordon, *J. Am. Chem. Soc.*, submitted.
100. D. Zorn, M.S. Gordon, in progress.
101. (a) P. Zapol, L.A. Curtiss, H. Tamura, M.S. Gordon, in *Computational Materials Chemistry: Methods and Applications*, L.A. Curtiss and M.S. Gordon, Eds., pp. 266-307 (2004); (b) H. Tamura, M.S. Gordon, in preparation.
102. H. Tamura, M.S. Gordon, *J. Chem. Phys.* 119 (2003) 10318.
103. J. Tomasi, M. Persico, *Chem. Rev.* 94 (1994) 2027.
104. E. Cancès, B. Mennucci, J. Tomasi, *J. Chem. Phys.* 107 (1997) 3032.
105. E. Cancès, B. Mennucci, *J. Chem. Phys.* 109 (1998) 249.
106. B. Mennucci, C. Amovilli, J. Tomasi, *Chem. Phys. Lett.* 286 (1998) 221.
107. (a) V. Barone, M. Cossi, *J. Phys. Chem. A* 102 (1998) 1995. (b) M. Cossi, N. Rega, G. Scalmani, V. Barone, *J. Comput. Chem.* 24 (2003) 669.
108. H. Li, C.S. Pomelli, J.H. Jensen, *Theoret. Chim. Acta* 109 (2003) 71.
109. H. Li, J.H. Jensen, *J. Comput. Chem.* 25 (2004) 1449.

110. (a) P. Bandyopadhyay, M.S. Gordon, *J. Chem. Phys.* 113 (2000) 1104. (b) P. Bandyopadhyay, B. Mennucci, J. Tomasi, M.S. Gordon, *J. Chem. Phys.* 116 (2002) 5023.
111. M.M.Karelson, A.R. Katritzky, M.C. Zerner, *Int. J. Quantum Chem.* S20 (1986) 521.
112. K.Baldrige, A. Klamt, *J. Chem. Phys.* 106 (1997) 6622.
113. B.M. Bode, M.S. Gordon, *J Molec. Graphics* 16 (1999) 133.