

Recent Advances in QM and QM/MM Methods

Mark S. Gordon and Michael W. Schmidt

Department of Chemistry and Ames Laboratory, Iowa State University, Ames, IA 50011
mark@si.fi.ameslab.gov

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

1 Introduction

In view of the limited space and the impressive array of recent advances in electronic structure theory, this summary will focus on new methods that have recently been implemented into GAMESS (General Atomic and Molecular Electronic Structure System)[1]. The following discussion is divided into three general topics: Recently developed and implemented methods in quantum mechanics (QM), New scalable methods for correlated wavefunctions, and Approaches for interfacing quantum mechanics with molecular mechanics (MM) in order to treat solvation and surface science.

2 QM Methods

In the 1950s Löwdin 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[2] 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[3]. 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. Since the FORS wavefunction generally corresponds to an incomplete CI, one also optimizes the molecular orbital coefficients to self-consistency. The resulting FORS/CASSCF multi-configurational

self-consistent-field (MCSCF) method is very powerful for a variety of application but the size of the active space for actual calculations is generally limited to roughly 14 electrons in 14 orbitals, or a full valence active space for a molecule the size of ethane.

Of course, even with the most efficient code, a full CI calculation or a full valence FORS calculation is limited to very small atoms and molecules. It is therefore very important to think about ways in which one can approach the accuracy of a full CI wavefunction with a much smaller effort than that required for full CI. In principle, one would expect that a great many of the configurations in a full CI or a FORS wavefunction have little effect on the total molecular energy, but it is not obvious how one would identify the important contributors to the total wavefunction. Ruedenberg, Ivanic and Bytautas have used the full CI code and a systematic analysis of single, double, triple, ... excitations in order to develop a general method for eliminating the "deadwood" from the full CI wavefunction[4]. 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. Fig. 1 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,

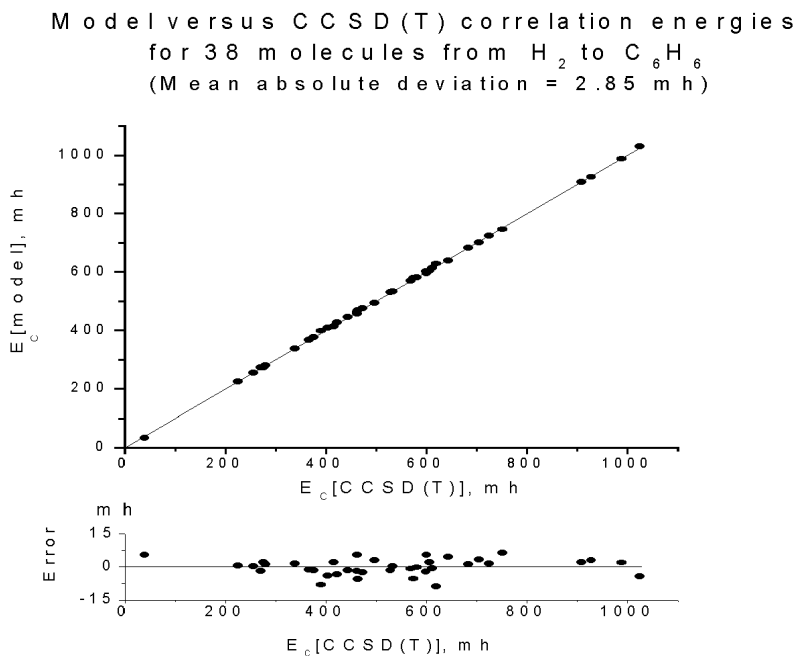


Fig. 1

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.

Once one has implemented a general CI such as that described above, the next logical step is to use this same approach for a MCSCF *ansatz*. The extension of a general CI method to a general MCSCF method is non-trivial, but it has been accomplished and implemented into GAMESS in collaboration with the Ruedenberg group[5]. 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.

Since a prime motivation for the general MCSCF approach is to expand the size of a chemical system that one can study at this level of theory, an especially exciting development is the very new ORMAS (Occupation Restricted Multiple Active Space) method developed by Ivanic[6]. This direct CI method, recently implemented into GAMESS, permits one to subdivide a FORS/CASSCF active space into multiple subspaces in a completely general way, and then imposes limits on the electron occupancies in each subspace. Since each subspace is treated individually as a complete active space, the ORMAS method enjoys all of the advantages of any FORS/CASSCF method (e.g., size-extensivity, good SCF convergence, straightforward formulation of energy gradients), while having the potential to greatly expand the accessible size of chemical systems. The method has already been applied with success to the very difficult N_2O_4 system, as well as a complicated transition metal complex[6]. Since MCSCF provides the correct zeroth order wavefunction, but does not include dynamic correlation, Ivanic has also developed an analog to the second order CI method; that is, single and double excitations out of all determinants in the ORMAS wavefunction. It is anticipated that this suite of methods will be heavily used in the future.

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 (frequently referred to as 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 ~ 2 , LUMO occupancy ~ 0) to fully diradical (HOMO occupancy ~ 1 , LUMO occupancy ~ 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 was demonstrated, for example, by examining the N_2 dissociation curves for MP2 and CCSD(T) vs. various MR methods[7]. 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[8] 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, and their equations-of-motion (EOM) analogs (especially important for investigating electronically excited states) will become available within the next six months.

3 Scalable Methods

One approach to growing the size of a chemical system that can be realistically treated by the most sophisticated 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, but become increasingly complicated for the more sophisticated correlated methods. Early approaches to parallel methods relied on replicated data (RD) algorithms, in which the data sets, such as Fock and density matrices, are replicated on each compute node, while the two-electron integrals are computed in a “direct” manner, on-the-fly. The disadvantage to 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[9], 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)[10] in GAMESS benefited considerably from the prior development of GA technology. DDI performs best when it can take advantage of the SHMEM library, especially on Cray systems, but it has also been very successful on 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.

The initial implementation of DDI was for closed shell MP2 energies and gradients[11]. 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[12]. For species with unpaired electrons, the implementation of restricted open shell energies is equally efficient, and it is anticipated that UMP2 energies and gradients (currently in progress) will scale as well as the closed shell analog. Restricted open shell gradients using the ZAPT *ansatz*

have been derived[13], and the coding of both sequential and parallel codes will begin shortly. DDI has also been used to develop a distributed parallel MR perturbation method in collaboration with the Koseki group[14]. It appears that the parallel MRMP2 method currently scales well up to about 32 processors.

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[15]. Very recently, a DD parallel MCSCF algorithm has been developed using the full Newton-Raphson convergence algorithm[16]. This DD MCSCF method addresses the integral transformation and orbital rotation steps, but not the CI coefficient optimization, which is treated in the next paragraph. Initial tests suggest that this algorithm will scale well up to ~32-64 processors, a major advance over the RD algorithm.

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[17]. 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 Fig. 2.

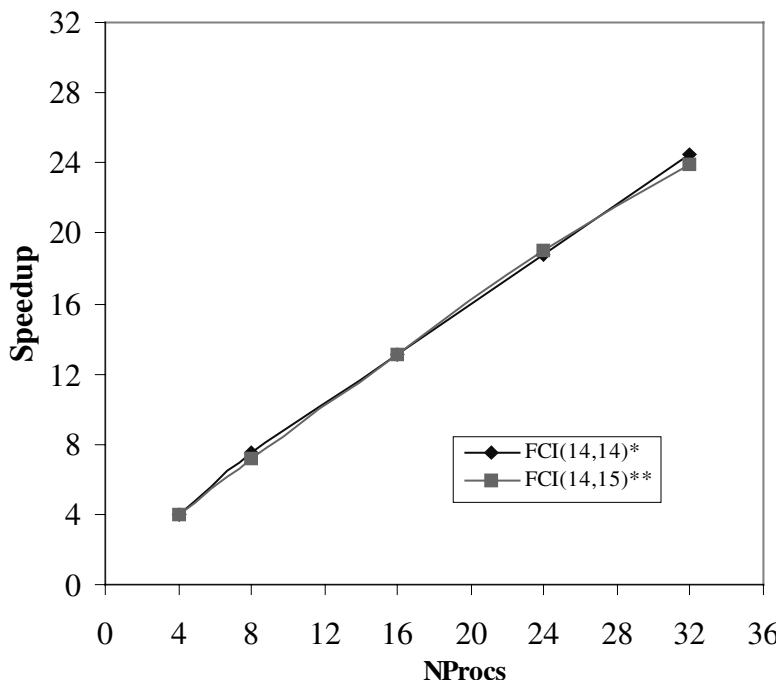


Fig. 2

This figure demonstrates a test on a cluster of 64-bit IBM Power 3II dual processor computers running AIX. The illustrated problems are CH_3OH (14,14) and H_2O_2 (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. Another example is a Monte Carlo calculation, since again, the energy 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[18]. For example, if one wishes to perform a MP2 Monte Carlo study, one can distribute the large number of MP2 energy evaluations among all available nodes. At the same time, if each node is a multi-processor (e.g., SMP) computer, each MP2 energy calculation can itself be distributed among the processors on its node.

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 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.

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[19]. 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 dimer potential and represents exchange repulsion and (implicitly) charge transfer. This 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[20]. Denoted EFP1/DFT, this method 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 and does not address the issue of fitting the repulsion term. In this sense, it is desirable to derive the exchange repulsion 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 solvents¹⁹. The exchange repulsion calculated by this method maps the exact HF intermolecular exchange typically to within 0.5 kcal/mol. It remains to develop analogous expressions for charge transfer, but this EFP2/HF method is already a success, and it has been extended by Jensen and co-workers[21] to the treatment of intramolecular covalent, rather than intermolecular interactions. Neither EFP1 nor EFP2 include dispersion interactions at present, but the development of dispersion terms for both methods is in progress.

Although the cost of an EFP calculation is several orders of magnitude smaller than that of a corresponding (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. This means to find the global minimum, for example, one requires a Monte Carlo or similar calculation that requires a great many energy evaluations. Likewise, to predict bulk properties, one would employ a molecular dynamics scheme that generates a great many energy + gradient evaluations. Both Monte Carlo/simulated annealing[22] and molecular dynamics[23] 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. As for any other application, the scalability relies on the size of the problem, but the scalability looks promising up through the 16 Linux nodes that have been available for testing.

For surface chemistry, a more traditional QM/MM approach, SIMOMM[24] (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 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 as hydrogens 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, etching, and diffusion of metal atoms along the surface. More recently, it has been applied to the growth of diamond and silicon carbide surfaces.

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

References

- Schmidt, M.W., Baldrige, K.K., Boatz, J.A., Elbert, S.T., Gordon, M.S., Jensen, J., Koseki, S., Matsunaga, N., Nguyen, K. A., Su, S., Windus, T.L., Dupuis, M., Montgomery, J.A.: *J. Comput. Chem.* **14**, 1347-1363 (1993).
- Ivanic, J., Ruedenberg, K.: *Theoret. Chem. Acc.* **106**, 339-351 (2001).
- Schmidt, M.W., Gordon, M.S.: *Annu. Rev. Phys. Chem.* **49**, 233-266 (1998).
- a) Bytautas, L., Ruedenberg, K.: *Mol. Phys.*, **100**, 757-781 (2002).
b) Ivanic, J., Ruedenberg, K.: *Theoret. Chem. Accs.*, **107**, 220-228 (2002).
- Ivanic, J., Ruedenberg, K.: *J. Comp. Chem.*, in press.
- Ivanic, J.: *J. Chem. Phys.*, submitted.
- Gordon, M.S., Schmidt, M.W., Chaban, G.M., Glaesemann, K.R., Stevens, W.J., Gonzalez, C.: *J. Chem. Phys.* **110**, 4199-4207 (1999).
- Piecuch, P., Kucharski, S.A., Kowalski, K., Musial, M.: *Comput. Phys. Commun.*, **149**, 71-96 (2002).
- Nieplocha, J., Harrison, R.J., Littlefield, R.J.: *Proc. Supercomputing'94*, 340-349 (1994).
- Fletcher, G.D., Schmidt, M.W., Bode, B.M., Gordon, M.S.: *Comput. Phys. Commun.* **128**, 190-200 (2000).
- Fletcher, G.D., Schmidt, M.W., Gordon, M.S.: *Adv. Chem. Phys.*, **110**, 267-294 (1999).
- Kudo, T., Gordon, M.S.: *J. Phys. Chem. A*, **105**, 11276-11284 (2001).
- Fletcher, G.D., Gordon, M.S., Bell, R.S.: *Theoret. Chem. Accs.*, **107**, 57-70 (2002).
- Umeda, H., Koseki, S., Nagashima, U., Schmidt, M.W.: *J. Comput. Chem.*, **22**, 1243-1251 (2001).
- Windus, T.L., Schmidt, M.W., Gordon, M.S.: *Theoret. Chim. Acta* **89**, 77-88 (1994).
- Fletcher, G.D.: to be published.
- Gan, Z., Alexeev, Y., Kendall, R.A., Gordon, M.S.: *J. Chem. Phys.*, in press.
- Olson, R.M., Fedorov, D.G., Schmidt, M.W., Gordon, M.S.: to be published.
- Gordon, M.S., Freitag, M.A., Bandyopadhyay, P., Jensen, J.H., Kairys, V., Stevens, W.J.: *J. Phys. Chem. A*, **105**, 293-307 (2001).
- Adamovic, I., Freitag, M.A., Gordon, M.S.: *J. Chem. Phys.*, in press.

21. Kairys, V., Jensen, J.H.: *J. Phys. Chem. A*, **104**, 6656-6665 (2000).
22. Day, P.N., Pachter, R., Gordon, M.S., Merrill, G.N.: *J. Chem. Phys.*, **112**, 2063-2073 (2000).
23. Netzloff, H.M., Gordon, M.S.: to be published.
24. Shoemaker, J., Burggraf, L.W., Gordon, M.S.: *J. Chem. Phys.*, **112**, 2994-3005 (2000).