2006

Scalable Correlated Electronic Structure Theory

Mark S. Gordon
Iowa State University, mgordon@iastate.edu

Klaus Ruedenberg
Iowa State University, klausr@iastate.edu

Michael Schmidt
Iowa State University, mike@si.fi.ameslab.gov

Laimis Bytautas
Iowa State University

Timothy J. Dudley
Iowa State University

See next page for additional authors

Follow this and additional works at: http://lib.dr.iastate.edu/ameslab_pubs

Part of the Chemistry Commons

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/ameslab_pubs/334. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.
Scalable Correlated Electronic Structure Theory

Abstract
The approach taken in Ames to advance high-level electronic structure theory has been a combination of the development and implementation of new and novel methods with the continuing development of strategies to optimize scalable computing. This work summarizes advances on both fronts. Several new methods have been implemented under the Distributed Data Interface (DDI), most recently including analytic Hessians for both Hartree-Fock and CASSCF (complete active space self-consistent field) wavefunctions, gradients for restricted open shell second order perturbation theory, and the fragment molecular orbital method (FMO). Exciting new method developments include the FMO method and the CEEIS (Correlation Energy Extrapolation by Intrinsic Scaling) method for efficiently approaching the exact energy for atomic and molecular systems.

Disciplines
Chemistry

Comments

Rights
This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Authors
Mark S. Gordon, Klaus Ruedenberg, Michael Schmidt, Laimis Bytautas, Timothy J. Dudley, Takeshi Nagata, Ryan Olson, and Sergey Aleksandrovich Varganov

This article is available at Iowa State University Digital Repository: http://lib.dr.iastate.edu/ameslab_pubs/334
Scalable correlated electronic structure theory

This content has been downloaded from IOPscience. Please scroll down to see the full text.


(http://iopscience.iop.org/1742-6596/46/1/032)

View the table of contents for this issue, or go to the journal homepage for more

Download details:

IP Address: 129.186.176.217
This content was downloaded on 03/12/2015 at 19:44

Please note that terms and conditions apply.
Scalable correlated electronic structure theory

Mark S. Gordon, Klaus Ruedenberg, Michael W. Schmidt, Laimis Bytautas, Timothy J. Dudley, Takeshi Nagata, Ryan Olson, Sergey Varganov
Ames Laboratory, Iowa State University, Ames, IA 50011

Abstract. The approach taken in Ames to advance high-level electronic structure theory has been a combination of the development and implementation of new and novel methods with the continuing development of strategies to optimize scalable computing. This work summarizes advances on both fronts. Several new methods have been implemented under the Distributed Data Interface (DDI), most recently including analytic Hessians for both Hartree-Fock and CASSCF (complete active space self-consistent field) wavefunctions, gradients for restricted open shell second order perturbation theory, and the fragment molecular orbital method (FMO). Exciting new method developments include the FMO method and the CEEIS (Correlation Energy Extrapolation by Intrinsic Scaling) method for efficiently approaching the exact energy for atomic and molecular systems.

1. Introduction.
Advances in both the accuracy and efficiency of electronic structure theory and methods are critical in order to enable the prediction of properties that directly connect with experiment. Accuracy can, in principle, be attained by exactly solving the Schrödinger equation by, for example, solving the full configuration interaction (Full CI) problem; however, such direct solutions are limited to atoms and molecules with only small numbers of electrons, even with the most powerful computers. Consequently, novel new methods, that approach the exact answer in a much more efficient manner, are needed. More approximate methods that include some electronic correlation, but do not approach the exact solution to the Schrödinger equation, can also be extremely useful for chemists; however, even approximate correlated methods have very demanding scaling requirements as a function of N, the number of atomic basis functions. For example, the most popular coupled cluster method, CCSD(T), scales as N^7, and the CASSCF (complete active space self-consistent field) and related methods scale exponentially. In order to make these methods more applicable to problems of interest to researchers in chemistry, biology, chemical engineering, physics and materials science, more efficient methods and algorithms are needed. The following paragraphs summarize recent advances in both the development of new methods and the development and implementation of new scalable algorithms.

2. The CEEIS Method.
In general, Full CI solutions have been severely limited to atoms and small molecules and to modest basis sets, because of the exponential growth in computational demand as one increases the size of the basis set. For example, a Full CI calculation on F_2 using the cc-pVQZ basis set (a good, but not extensive basis set) would require ~10^17 determinants, a number that is still well beyond current computational capabilities. In order to extend the size and complexity of molecular systems that are amenable to Full CI calculations, or the equivalent, one must devise new and clever methods. The approach taken in Ames is to recognize that the vast majority of determinants in a Full CI wavefunction contribute very little to the total energy (are “deadwood”) and devise a clever way to
take advantage of this fact. The CEEIS (Correlation Energy Extrapolation by Intrinsic Scaling) method [1] cleverly estimates the contribution of the deadwood to sub-millihartree accuracy. This approach is combined with a basis set extrapolation to the complete basis set limit to provide a very accurate prediction of entire molecular potential energy curves. A CEEIS calculation is initiated by a preliminary SCF or MCSCF calculation to generate a set of natural orbitals that are subsequently used to generate all determinants in the calculation. The correlation energy is then calculated as

$$\Delta E_{\text{corr}} = \sum \Delta E(x).$$

Here, $\Delta E(x) = \text{limit} \Delta E(x|m)$, for $m \to N$, where $N$ is the number of basis functions. In $\Delta E(x|m)$, only the first $m$ correlating orbitals are used in the correlation calculation. The contributions of the remaining orbitals are estimated using a linear extrapolation. For example,

$$\Delta E(x|m) = a_x \Delta E(x-2|m) + c_x.$$

The coefficients are obtained by least square fitting for low values of $m$ with a moderate number of determinants. Then, the desired values of $\Delta E(x) = \Delta E(x|M)$ for $x \geq 4$ are obtained from the values for $x=2,3$ and $m=M$. The efficacy of the CEEIS method is illustrated in Figures 1 and 2. Figure 1 summarizes the binding energies of four first row diatomic molecules. After correcting for core-valence correlation, scalar relativistic effects, zero point vibrational energy, and spin-orbit coupling, the agreement with experiment is within 2 mh (1.2 kcal/mol) for all four molecules. A comparison between the number of determinants required for CEEIS vs. a Full CI calculation is revealing. For the smallest molecule, $C_2$, is $3.6 \times 10^7$ for CEEIS as compared with $3.6 \times 10^{12}$ for a Full CI calculation. The corresponding comparison for $F_2$ is $1.1 \times 10^8$ vs. $3.7 \times 10^{17}$. Clearly the CEEIS method can provide chemical accuracy at a small fraction of the cost of a Full CI calculation.

Figure 1. Binding Energies of $C_2$, $N_2$, $O_2$, $F_2$ (in mh)

<table>
<thead>
<tr>
<th>Energy</th>
<th>$2C \rightarrow C_2$</th>
<th>$2N \rightarrow N_2$</th>
<th>$2O \rightarrow O_2$</th>
<th>$2F \rightarrow F_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonrelativistic Valence</td>
<td>-232.3</td>
<td>-362.7</td>
<td>-193.2</td>
<td>-61.6</td>
</tr>
<tr>
<td>Core-Valence Correlations</td>
<td>-2.4</td>
<td>-1.4</td>
<td>-0.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Scalar Relativistic</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>0.0</td>
</tr>
<tr>
<td>Spin Orbit Coupling</td>
<td>0.3</td>
<td>0.0</td>
<td>0.7</td>
<td>1.2</td>
</tr>
<tr>
<td>CEEIS, Theory</td>
<td>-234.1</td>
<td>-363.9</td>
<td>-192.9</td>
<td>-60.4</td>
</tr>
<tr>
<td>Exp. Dissociation</td>
<td>-231.7 ± 0.8</td>
<td>-358.5 ± 0.04</td>
<td>-188.0 ± 0.002</td>
<td>-58.9 ± 0.2</td>
</tr>
<tr>
<td>Exp. Zero-Point Vibration</td>
<td>4.2</td>
<td>5.4</td>
<td>3.6</td>
<td>2.1</td>
</tr>
<tr>
<td>Experiment</td>
<td>-235.9 ± 0.8</td>
<td>-363.9 ± 0.04</td>
<td>-191.6 ± 0.002</td>
<td>-61.0 ± 0.2</td>
</tr>
</tbody>
</table>

Enthalpy of Formation of Water (hartree)

<table>
<thead>
<tr>
<th>Energy</th>
<th>$H_2$</th>
<th>$1/2$ $O_2$</th>
<th>$H_2O$</th>
<th>$\Delta H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonrelativistic Valence</td>
<td>-1.17468</td>
<td>-75.10240</td>
<td>-76.37740</td>
<td>-0.10032</td>
</tr>
<tr>
<td>Core-Valence Correlations</td>
<td>0.00000</td>
<td>-0.0624</td>
<td>-0.06280</td>
<td>-0.00035</td>
</tr>
<tr>
<td>Scalar Relativistic</td>
<td>0.0</td>
<td>0.00015</td>
<td>0.00044</td>
<td>0.00029</td>
</tr>
<tr>
<td>Spin Orbit Coupling</td>
<td>0.0</td>
<td>0.00035</td>
<td>0.00035</td>
<td>0.00000</td>
</tr>
<tr>
<td>CEEIS, Theory</td>
<td>-1.17468</td>
<td>-75.16435</td>
<td>-76.43941</td>
<td>-0.10038</td>
</tr>
<tr>
<td>Experiment</td>
<td>0.00999</td>
<td>0.0018</td>
<td>0.0211</td>
<td>-0.0094</td>
</tr>
<tr>
<td>Exp. Zero-Point Vibration</td>
<td>0.00999</td>
<td>0.0018</td>
<td>0.0211</td>
<td>-0.0094</td>
</tr>
<tr>
<td>Experiment</td>
<td>-0.1004</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


As shown in Figure 2, the accuracy of CEEIS is not limited to simple energy differences. To the contrary, CEEIS predicts the entire F$_2$ potential energy curve agree with experiment to within sub-millihartree accuracy. Although it is not shown in the figure, this excellent agreement leads to astonishing accuracy in the prediction of the full set of the F$_2$ vibrational spectrum [2].

3. DDI.
The distributed data interface (DDI) [3] is the code by which GAMESS (General Atomic and Molecular Electronic Structure System) [4] accomplishes distributed data parallelism. DDI is implemented on most computer systems using socket code, or MPI when sockets are not available, with a paired CPU/data server arrangement. The exception is that SHMEM is used where it is available, as on Cray computers. Several GAMESS functionalities have been implemented with DDI. These include RHF Hessians, energies and analytic gradients for both closed and open shell second order perturbation theory, CASSCF energies, gradients and Hessians, multi-reference second order perturbation theory energies, CI singles energies and gradients, Full CI energies, fragment molecular orbital (FMO) energies, and effective fragment potential (EFP) energies and gradients. So, a large majority of GAMESS functionalities can be performed with good scalability. Coming to DDI within the next few months are derivative coupling (critical for the study of conical intersections and therefore photochemistry and photobiology), CCSD(T) energies, generalized Van Vleck second order perturbation theory energies and gradients, and ORMAS (Occupation Restricted Multiple Active Spaces) energies and gradients.

4. CASSCF Hessians.
It is always preferable to compute energy derivatives analytically, rather than numerically, simply for the advantage ones has in the accuracy one has for predicting vibrational frequencies. For MCSCF, the availability of analytic Hessians is additionally important, because one is often interested in predicting frequencies for electronic excited states. If a numerical algorithm is used, “root flipping” to a lower electronic state is very likely, and if this happens, the calculation fails. Furthermore, analytic MCSCF Hessians are essential for the subsequent computation of vibronic (derivative) coupling matrix elements that are so important for understanding the electronic structure and dynamics of systems near conical intersections and curve crossings. Because analytic MCSCF Hessians are so computationally demanding, it is essential to implement them using a scalable algorithm. The scalability of the parallel MCSCF Hessian code in GAMESS [5] is illustrated in Figure 3, for the
biologically important molecule 7-azaindole. This calculation employs a relatively modest basis set and a modest MCSCF active space. The scalability is nearly perfect through 32 processors and very good through 64 processors. Of course, a more demanding problem is expected to scale well to more processors. Such a demanding problem is illustrated in Figure 4, where a comparison between the CASSCF and experimental IR spectrum is presented to Ti8C12, a metallocarbohedrene, the parent compound in a family of important new materials. Both the absolute frequencies and the shift induced by ionization to the cation are well reproduced by the calculations, in which a basis set of more than 350 basis functions was used.

Once analytic MCSCF Hessians are available, they can be employed to calculate derivative couplings. This is illustrated in Figure 5 for the two lowest lying sattes of ozone (O₃), in which these non-adiabatic coupling matrix elements (NACME) are plotted as a function of the O-O bond distance for a variety of O-O-O angles. One can see that there are many regions of the potential energy hypersurface in which this coupling becomes very large, and this can have significant impact on the chemistry of this important molecule.
5. Conclusions.
The electronic structure theory group in Ames have made great strides in the development of new models that are designed to approach the exact solution of the Schrödinger equation, as well as novel algorithms that enable excellent parallel scalability for highly correlated wavefunctions.

6. Acknowledgements.
This work was supported by a Department of Energy SciDAC (Scientific Discovery Through Advanced Computing) grant to Ames Laboratory and by several IBM hardware grants via the SUR (Shared University Research) program.

References.