XSCI

eXtreme Scale Computing Initiative

Initiative Leadership

Jarek Nieplocha, Initiative Lead
(509) 372-4469
Jarek.nieplocha@pnl.gov

T.P.Straatsma, Initiative Lead
(509) 375-2802
tps@pnl.gov

Operational Team

Myra Wyant, Finance

Jill Blaylock, Administrator

Christine Novak, Communications

Website

http://www.pnl.gov/xsci/

Pacific Northwest NATIONAL LABORATORY

Challenges in Extreme Scale Computing

Effective use of high performance systems at the current peta-scale and potential exa-scale in the near future requires radical changes to the operating software as well as the science applications. These changes are as disruptive as the shift from vector processors to distributed memory massively parallel supercomputers over a decade ago. Scalability to millions of processor cores will be very difficult to achieve, and will not only require novel software development and performance analysis tools, but also a re-evaluation of mathematical and algorithmic approaches. Fault resiliency becomes a necessity on machines with millions of components.

Many scientific disciplines have the potential to benefit tremendously from computing at the extreme scale, but not without addressing the many algorithmic and implementation challenges for existing application codes. Novel, innovative approaches are needed both in fundamental and applied mathematics, the computer science approaches, and in the computational sciences. The eXtreme Scale Computing Initiative (XSCI) was established to build capabilities that will enable PNNL to address these challenges for computational science domains, in particular focusing on application areas supporting the Laboratory Missions.

The Initiative is organized around three focus areas. The main objective is to provide scalable and fault resilient algorithmic solutions to domain applications. The initial target scientific areas are molecular modeling and subsurface simulations. To accomplish this, two other main focus areas are applied computational mathematics and high performance computer science.

XSCI Research Accomplishments

Each of the XSCI bi-monthly newsletters will highlight the research approach and accomplishments of a different research projects supported by the. In the current issue, the “Extreme Scale Molecular Modeling” project is described in more detail.

Extreme Scale Molecular Modeling

K. Kowalski, E.J. Bylaska, M. Valiev

The main goal of the Extreme Scale Computing Initiative launched by PNNL is to develop a new generation of codes and tools which can efficiently utilize massively parallel machines characterized by hundreds of thousands of CPUs. In our project “Development of extreme-scaling algorithms for molecular modeling” we are striving to achieve this goal by developing highly diversified methodologies to describe systems at different time and spatial scales.

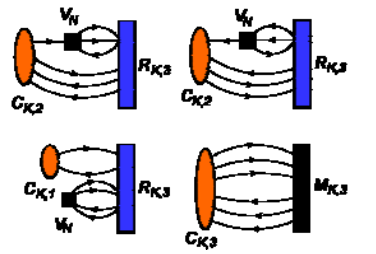

Figure 1. Diagrammatic representation of the Em-EOMCCSD(T) excited-state

An important component of this project is to develop highly-scalable ab-initio equation-of-motion coupled cluster (EOMCC) methodologies for excited-state calculations Below we briefly summarize recent progress in developing the iterative and non-iterative EOMCC methodologies based on the reduced iterative core: (1) iterative models with all singles (S) and selected subset of doubly excited amplitudes (d) defined within the so-called active space (CCSd/EOMCCSd; small case letter "d" designates limited set of doubles) were implemented with full inclusion of active-space logic imposed by the doubly excited cluster/excitation amplitudes. (2) Non-iterative approaches accounting for triply excited configuration were implemented for CCSd and EOMCCSd approaches. The resulting completely renormalized EOMCCSd(t) method uses the $\left|\Phi_{iJK}^{ABc}\right\rangle$ determinants only (capital indices correspond to active orbitals) in constructing the excited-state correction. In all cases the active space can be made arbitrarily large. (3) In order to minimize the inter-processor communication the most important low-rank tensors were replicated. Preliminary tests show that this procedure results in 10% speed-up of the iterative part. (4) The fully in-core algorithm for 4-index transformation was implemented. (5) A novel non-iterative approach, the so-called embedded EOMCCSD(T) method (Em-EOMCCSD(T), see Figure 1), was developed for full CCSD/EOMCCSD formalisms as well as for active-space methodologies. It is worth noticing that this approach solves the long-standing problem of balancing ground- vs. excited-state corrections.

In a planned series of publications we are going to discuss several aspects of developed and implemented formalism. This will include the discussion of accuracies of the Em-EOMCCSD(T) approach in describing the excited-state potential energy surfaces, the parallel performance of the Em-EOMCCSd(t) formalism based on the reduced iterative core, and the application of the Em-EOMCCSD(T) approach to large molecular systems. We hope that the developed suite of codes will provide a comprehensive understanding of photo-induced effects in large molecules, which due to their size pose a great challenge for accurate yet very expensive excited-state method based on the wavefunction approach

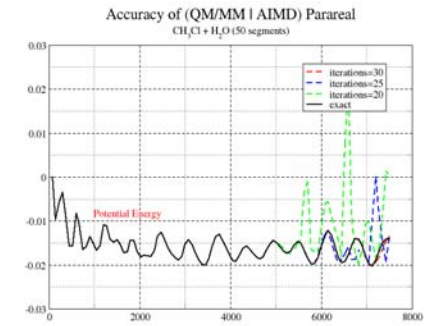

Figure 2. Illustration of the accuracy of the parareal multiscale (QM/MM | AIMD) model for the CH_3Cl+H_2O system.

In this project, we are also developing new solver strategies for the integration in time to be done in parallel. More aggressive methods of propagation in time, based on local in time multilevel solvers, such as the "parareal" algorithm of Lions *et. al.*, are being investigated. The original parareal algorithm is based upon a multiple shooting integration algorithm in which a coarse time step ($G_{\Delta T}$) is used to define initial conditions for fine time-step ($F_{\delta t}$) integration within each coarse grid solution, and then a coarse grid correction is effected. This process is repeated to define an iterative method. In principle the parareal algorithm is also applicable to multi-scale models involving different models, since there is no reason that in the algorithm, the model used for fine scale temporal integration has to be the same as that used at the coarse scale. We have recently implemented a two-level parareal multi-scale model composed of a coarse grain AIMD/MM model (*ab initio* molecular dynamics/Molecular mechanics) and a find grain AIMD model:

- $G_{\Delta T}$ – AIMD/MM simulation
- $F_{\delta t}$ – Full AIMD simulation

This multi-scale model was implemented using in the ab initio molecular dynamics module of the NWChem program package. Not surprisingly, initial studies have shown that the number of parareal solver iterations is sensitive to the choice of QM/MM model as well as the size of the coarse grain time step. We are currently testing and optimizing this model on a variety of systems, including CH_3Cl+nH_2O and $Zn^{2+}+nH_2O$ (see Figure 2).

K. Kowalski, "*Embedded variant of the Method of Moments of Coupled Cluster Equations for vertical excitation energies and excited-state potential energy surfaces*," (in preparation).