SI2-CHE: Development and Deployment of Chemical Software for Advanced Potential Energy Surfaces

Teresa Head-Gordon[1], Lorna Smith[2]

Omar Demerdash[1], Richard Bradshaw[4], Henry Boateng[7], Weronika Fillinger[2], Alex Albaugh[1], Yuezhi Zhou[1], Jacek Dziedzic[4], Arno Proeme[2], and Mario Antonioletti[2]

David Case[3], Jonathan Essex[4], Neil Chue Hong[2], Martin Head-Gordon[1], Paul Nerenberg[5], Jay Ponder[6], Chris Skylaris[4], Illian Todorov[7], Mark Tuckerman[8]

[1]UC Berkeley, [2]U Edinburgh, [3]Rutgers U, [4]U Southampton, [5]Caltech, [6]Washington U, St. Louis, [7]Daresbury Lab, [8]New York U

Introduction

Molecular simulation and quantum chemistry software is an integral part of chemistry and chemical biology, and has been broadly adopted by academic researchers and industry scientists. Next generation scientific breakthroughs that utilize chemical software will be enabled by the deployment of state of the art theoretical models and algorithms that are translated into a sustainable software framework rapidly implemented on emergent high performance computing platforms.

Advanced potential energy surfaces, defined as classical or quantum mechanical treatments beyond widely available classical fixed charge pairwise-additive force fields, are encountering software-related obstacles that inhibit their application to grand challenge chemistry problems: computational cost of the theoretical models, lack of innovation in approximate polarizable models and algorithms that can mitigate the cost, limited dissemination to a broad range of community codes, under-developed software interfaces between theoretical models, and lagging quality software implementations on HPC architectures and newer GPU and multicore hardware.

We have organized a UK and US consortium that represents a broad cross section of the computational chemistry software community involved in chemical and biochemical applications, force field development, electronic structure methods, molecular dynamics algorithms, and software engineering and computer science experts, to directly tackle these obstacles. Our US and UK collaboration between universities and HPC centers will ensure that chemical software investments made in advanced potential energy surface models, successfully deployed on current and emergent hardware, will also realize a long term payoff in community sustainability and the training of the next generation of scientists.

Intellectual Merit

Classical potential energy surfaces have reached a generational transition, moving away from well-established but intrinsically limited fixed point charge models, and towards more intricate and expensive potential energy surface models that can allow accurate reproduction of a much broader range of reference data and better predictions. However the greater accuracy introduced by improvements in short range forces, polarizability, charge transfer, or explicit quantum mechanical treatments open up several challenges in their manifestation as algorithms and software on current or emergent hardware platforms, that in turn limits their wide adoption by the computational chemistry community. We will overcome these obstacles via multiple but integrated directions:

- optimally implement advanced potential energy surfaces across multi-core, GPUs, desktops to large scale NSF and EPSRC HPC resources
- develop a hierarchy of advanced polarizable models (iAMOEBA, 3-AMOEBA, Drude-AMOEBA, and full AMOEBA) that alter the trade off between accuracy and computational speed;
- new multiple time stepping methods;
- a QM/MM API that fully supports mutual polarization and new advanced linear scaling algorithms to grow the QM region
- adopt software best practices to ensure growth of a self-sustaining community, providing longevity beyond the funded phase of project
- to provide exemplar calculations with the new software in the areas of biological complexity.

Broader Impact

- The UK and US PIs represent a broad cross section of the computational chemistry and computer science community
- It will directly interface experts in the two main approaches of quantum and classical models to develop new software interfaces, and to develop new state-of-the-art classical models and algorithms to enable grand challenge applications at length and time scaless beyond the reach of current software
- A long-term payoff in community availability and sustainability through software development in 3 US and 2 UK chemistry software packages that serve different communities, offering a core of common functionality while providing additional targeted functionality and platform availability
- expertise in a range of grand challenge chemistry problems ranging from reactive chemistry to biomolecular simulation.
- Outreach and training workshops that organize around the emergence of the advanced potential energy software.
- Molecular simulation software boot camp for undergraduate/graduate students at multiple institutions will prepare them to directly participate in grand challenge prototyping efforts, thus enabling these students to gain research experience early on in their careers.

Advanced Potential Energy Surfaces

Our consortium is organized in part around the classical force field AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications) which belongs to the class of molecular mechanics force fields that aims for high fidelity to ab initio calculations. The computational cost makes it suited for primarily small molecule condensed phase studies, but it is more limited in its application to large biomolecule studies and/or when extensive statistical mechanical sampling is necessary. AMOEBA has the following functional form for the interactions among atoms:

$$U = U_{bond} + U_{angle} + U_{b\theta} + U_{oop} + U_{torsion} + U_{vdW} + U_{ele}^{perm} + U_{ele}^{ind} \quad (1)$$

Advanced Potential Energy Surfaces

where the first six terms describe the short-range valence interactions and the last three terms are the nonbonded van der Waals (vdW) and electrostatic contributions from atomic multipoles through the quadrupoles, as well as polarizable dipoles. The inclusion of explicit dipole polarization allows the AMOEBA model to respond to changing or heterogeneous molecular environments, and allows direct parameterization against gas phase experimental data and high-level quantum mechanical results. More specifically, the induced dipole vector on any polarizable site i can be expressed as

$$\mu_{i,\beta} = \alpha_i \left(\sum_j T_\beta^{ij} M_j^{perm} + \sum_{j'} T_{\beta\delta}^{ij'} \mu_{j',\delta} \right) \quad (2)$$

Where $T_{ik}^{11} = \nabla_{ik}^2$ is a 3×3 sub matrix, consists of elements in corresponding to the inducible dipole moments. Since the atomic polarizability, α_i, is isotropic, the off-diagonal elements of the tensor are all zero and the three diagonal elements take the same scalar value. Initially the induced dipoles in Eq. (2) were solved for iteratively to produce the final induced dipoles, the self-consistent field (SCF) step, with convergence accelerated via a conservative successive over-relaxation (SOR) procedure. The iterative induced dipole costs for a convergence of 10^{-6} to 10^{-8} D is up to 15-20 times the cost of a fixed charge model time step. This steep rise in computational cost sorely limits the computational chemistry community in its ability to make their science more predictive using advanced theoretical models such as AMOEBA. The iAMOEBA, 3-AMOEBA, Drude-AMOEBA models provide for well-defined approximations to Eq. (2) that could aid in its computational speed-up [1].

In preliminary work in 2012/2013, we ported and tested the TINKER and Amber implementation of the AMOEBA force field on the XSEDE platforms Kraken and Lonestar and the UK machine HECToR. We utilized the JAC benchmark of ~25,000 (protein and water) atoms to show that the TINKER OpenMP software implementation scales only to 4-8 cores and the Amber MPI implementation to 20-30 cores. *It perfectly illustrates the need for software management and engineering to improve the computational performance of advanced potential energy models on real and emergent hardware platforms where the science is actually performed, and to ensure that those improvements are translated into a broad range of software packages to reach a wider user and developer community.*

A hybrid QM/MM description is often used when full QM modeling is computationally prohibitive. However available QM/MM approaches do not fully support polarizable force fields, which would necessitate including not only the long-range polarization of the QM from the MM, but also the polarization of the MM due to the QM, both happening concurrently and iterated to self-consistency. Furthermore, inaccuracies that develop when embedding a QM region in a MM environment are often due to the severe limitations on the number of atoms that can be treated via QM. While linear scaling electronic structure methods can partly address this challenge, new algorithms are required to achieve simultaneously maximum precision and performance gains. Algorithmic developments will be needed in order to achieve improved scaling of the ONETEP and Q-Chem codes on new and emerging non-homogeneous computer architectures, requiring implementation of mixed-mode MPI-OpenMP parallelism.

Software

- **TINKER:** serves as the reference code for the AMOEBA model, and enables modest parallel speedups of periodic MD simulations on commodity SMP hardware via OpenMP, and is free of charge with source code (http://dasher.wustl.edu/tinker/). Versions are available for all modern computer systems, and prebuilt executables are supplied for Linux, Windows and Apple OSX.
- **AMBER:** is a molecular dynamics software suite, which is used in more than a thousand sites. Support for the AMOEBA force field is split off into a separate executable, pmemd.amoeba, to allow for AMOEBA-specific optimizations, and uses MPI for message passing, PME for long-range electrostatics (including multipoles) and dynamic load-balancing to distribute work.
- **DL_POLY:** DL_POLY_4 is the UK flagship for a general purpose classical MD simulation software written in fully self-contained, free-formatted Fortran90 and provides scalable performance from a single processor workstation or on a high performance parallel computer provided an MPI2 instrumentation is available on the parallel machine. DL_POLY_4 has extended support to multipoles for AMOEBA (a goal of this work) and treating long range electrostatics with its own memory distributed implementations of Smoothed PME (SPME) and 3D FFTs (DaFT).
- **ONETEP:** is a general purpose linear-scaling Density Functional Theory (DFT) code capable of handling up to tens of thousands of atoms. The code has been developed with modern software engineering principles and in modular fashion which will facilitate the proposed developments and interfacing with other MM and QM packages.
- **Q-Chem:** is a comprehensive ab initio quantum chemistry package for accurate predictions of molecular structures, reactivities, and vibrational, electronic and NMR spectra using state-of-the-art of methodology such as DFT/HF calculations (linear scaling algorithms) to post-HF correlation methods. Q-Chem currently performs QM/MM for large systems via both internal capability and an API with CHARMM, using fixed charge force fields, and is reorganizing QM/MM API's to address mutual polarization with AMOEBA.

Progress Y2 (Since May 2014)

PI Synergy/Interactions:

- We had two face-to-face meetings involving US-UK PIs in Y2. One meeting was held September 2014 in Berkeley, and a multi-site visit of US-based postdoc Dr. Omar Demerdash to Daresbury (with THG), Edinburgh, and Southampton.
- Small sub-groups and full consortium are meeting regularly via ReadyTalk, and code is shared via a GitHub account (https://github.com/apes-soft), as well as an internal wiki to disseminate project information.
- Since Oct. 2014, Edinburgh and Dr. Demerdash have had weekly Skype meetings discussing parallel profiling and load balancing of the hybrid OpenMP/MPI implementation of 3-AMOEBA in Tinker.

Parallelization Efforts:

- Edinburgh and Dr. Omar Demerdash hand-instrumented the PME code in Tinker to verify that placement of the multipoles on the grid (subroutine grid_mpole), and the extraction of the potential from the grid (subroutine fphi_mpole) are the primary bottlenecks in the TINKER implementation for the permanent electrostatics. They are now focused on a replicated data hybrid MPI/OpenMP approach that benefits the hierarchy of models iAMOEBA, 3-AMOEBA, Drude-AMOEBA, AMOEBA.
- DL_POLY 4 is incorporating full AMOEBA functionality in the domain decomposition parallelization environment, including multipole electrostatics [2].
- For pmemd.amoeba in AMBER there is ongoing work on making the recently implemented DIIS SCF procedure scale better to large numbers of threads.

TINKER/AMOEBA Improvements:

- TINKER has implemented a pre-conditioned conjugate gradient approach to the SCF problem to improve polarization speeds by a factor of 3
- Cleanup, documentation and tutorial creation for the "TINKER_to_AMBER" script that allows users to create the input files needed for pmemd.amoeba.
- We have worked with Stanford on the ForceBalance reparameterization of the AMOEBA water model [3]

Hierarchy of Polarization Models

- **3-AMOEBA:** A hybrid OpenMP/MPI implementation for polarization and real-space permanent electrostatics has been integrated into the molecular dynamics routine in Tinker. 3-AMOEBA has a speedup of 2-4 fold over Tinker's N-AMOEBA implementation and ~2 fold speedup over N-AMOEBA in the pmemd.amoeba routine of AMBER and speeds comparable and at best near ~2 fold speedup over N-AMOEBA in OpenMM run with double precision. The polarization calculation exhibits excellent strong scaling. In the limit of 3,000-6,000 cores, polarization is no longer the computational bottleneck, but rather the permanent electrostatics. Current effort is focusing on an efficient replicated-data, MPI/OpenMP for the reciprocal-space permanent electrostatics (currently just MPI) and improving the MPI implementation of real-space permanent electrostatics.
- **Drude-AMOEBA:** The typical Drude model relates polarizability, α; Drude charge, q_D; and Drude spring force constant, k_D, by ssuming that the electric field is the same at the atomic center and Drude particle position

$$\mu = \alpha E = q_D \vec{d}; \qquad F_{D,net} = q_D E - k_d d = 0; \qquad k_d = q_D^2/\alpha \quad (3)$$

Eq. (3b) represents the Born-Oppenheimer approximation where the electronic degrees of freedom (Drude particles) must be instantaneously at equilibrium with the atomic positions. However, the assumption that the electric field at the Drude particle and the atomic center is the same is incorrect, and we have developed a Drude formulation that takes into account the disparate electric fields at the two positions.

$$\begin{bmatrix} k_{D,xx} & 0 & 0 \\ 0 & k_{D,yy} & 0 \\ 0 & 0 & k_{D,xz} \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} = k_0 \begin{bmatrix} E_{D,x} \\ E_{D,y} \\ E_{D,z} \end{bmatrix} \quad (4)$$
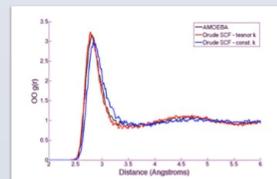

Figure 1 now shows that the AMOEBA Drude model using Eq. (4) is equivalent to the induced dipole model compared to the isotropic Drude using Eq. (3). We are currently exploring tractable and stable extended Lagrangian schemes for propogation of dynamics.

- iAMOEBA: We are extending the direct polarization model to new parameterization using Force Balance for a large class of ions.

Multiple time stepping

- The resonance-avoiding multiple time stepping methods XI-RESPA and XO-RESPA are working in TINKER for non-polarizable models, and two students are working on extending the approach to the polarizable AMOEBA model.
- AMBER has implemented several versions of the isokinetic integrators, including a stochastic version recommended for polarizable force fields. Outer time steps can now be in the range of 30 to 100 fs, as found by other groups, and ongoing tests will tune the parameters in these models in an attempt to find "standard" parameters suitable for most biomolecular simulations.

Progress Y2 (Since May 2014)

QM-MM

- The Skylaris group has finished the implementation of a distributed multipole analysis (DMA) using spherical waves, to treat the QM/polarizable-MM electrostatic interaction efficiently in any code, regardless of underlying representation. Two papers are in preparation.
- Work is underway on the Q-Chem-MM API with MM=AMOEBA, based on development of the open source LibEFP library. Previous results with a nono-polarizable force field illustrate the QM-MM approach on zeolites [4] that will be extended to polarizable MM in the future
- A new approach to create adaptive basis functions to reveal sparsity in one-particle density matrices that is absent in many standard representations such as a large gaussian atomic orbital basis, to enable large scale DFT calculations, as part of the QM/MM effort. First paper is currently being written.

Polarization Applications

- pK_a calculations on Staphylococcus nuclease (Southampton and UCB). BAR free energy calculations have been carried out for 3 residues in Staph. nuclease using AMOEBA and compared with BAR and constant pH simulations using the fixed-charge force field ff99SB in AMBER. Staph. nuclease is a challenging system with multiple residues exhibiting coupled protonation and large pK_a shifts. Preliminarily, pK_a shifts calculated with BAR using AMOEBA are closer to experiment than those calculated with BAR using ff99SB. Exhaustive exploration of other residues, particularly those with coupled protonation microstates, is ongoing.
- Solvation free energies (Southampton) AMOEBA and the GAFF fixed-charge force field gave results of comparable accuracy in calculating a set of 47 small molecule hydration free energies. Investigation of AMOEBA small molecule parameterization identified clear stepwise improvements to an existing automated protocol, which will be transferable to other applications. Solvation free energy calculations in non-aqueous solvents have shown small improvements using AMOEBA. Work is ongoing to determine whether functional group trends for these improvements exist.
- Crystal simulations (Rutgers) of the "fav8" peptide in its crystal environment have been carried out at a 100 ns time scale with AMOEBA. Structures and fluctuations are in modestly but systematically better agreement with experiment than with competing non-polarizable force fields.

Outreach and Sustainability:

- We continue the second year of our molecular simulation bootcamp that will take place June 14-20 at Caltech. The first 5-day workshop will introduce general scientific computing + basic MD techniques directed at undergraduates and first year grad students who are simulation novices. (2) parallelization.
- We will also host a 1-2-day advanced topics workshop in August 2015 in Berkeley that would be intended primarily for students with prior simulation experience to focus on the use of QM/MM with AMOEBA.
- We have organized a Telluride meeting for Advanced Potential Energy Surfaces for June 14-18, 2015 that is open to scientists working on polarizable models.
- We have also submitted a proposal to SC15 for a full day session "Producing high performance and sustainable software for molecular simulation".
- We have defined a benchmark suite for all the code and different developments will be fully integrated into the main code releases and made available to others beyond the project.
- We plan to run a set of courses for the UK's national service (ARCHER) users in the final year of the project for DL_POLY and ONETEP and possibly TINKER and AMBER.

Publications (NSF-SI2 Supported)

1. O. N. Demerdash, E.-H. Yap and T. Head-Gordon (2014). Advanced potential energy surfaces for condensed phase simulation. Ann. Rev. Phys. Chem. 65, 149-174.
2. H.A. Boateng and I. T. Todorov, "Arbitrary order permanent Cartesian multipolar electrostatic interactions," J. Chem. Phys. 142, 034117 (2015); doi: 10.1063/1.4905952
3. M. Laury, L.-P. Wang, V. S. Pande, T. Head-Gordon, J. Ponder (2015). Revised parameters for the AMOEBA polarizable atomic multipole water model. J. Phys. Chem. B (Branka Ladanyi Festschrift) in press
4. Y.-P. Li, J. Gomes, S.M. Sharada, A.T. Bell, and M. Head-Gordon (2015). Improved Force-Field Parameters for QM/MM Simulations of the Energies of Adsorption for Molecules in Zeolites and a Free Rotor Correction to the Rigid Rotor Harmonic Oscillator Model for Adsorption Enthalpies, J. Phys. Chem. C. 119, 1840-1850.

Acknowledgments