Accurate and efficient QM/MM molecular dynamics on 86,016 cores of SuperMUC Phase 2

[Extended Abstract]

Magnus Schwörer
Ludwig-Maximilians-Universität München, Lehrstuhl für BioMolekulare Optik
Oettingenstrasse 67
80538 München, Germany
magnus.schwoerer@physik.uni-muenchen.de

Momme Allalen
Leibniz Supercomputing Centre (LRZ)
Boltzmannstrasse 1 85748
Garching, Germany
allalen@lrz.de

Ferdinand Jamitzky
Leibniz Supercomputing Centre (LRZ)
Boltzmannstrasse 1 85748
Garching, Germany
ferdinand.jamitzky@lrz.de

Helmut Satzger
Leibniz Supercomputing Centre (LRZ)
Boltzmannstrasse 1 85748
Garching, Germany
helmut.satzger@lrz.de

Gerald Mathias
Ludwig-Maximilians-Universität München, Lehrstuhl für BioMolekulare Optik
Oettingenstrasse 67
80538 München, Germany
gerald.mathias@physik.uni-muenchen.de

1. INTRODUCTION

The combination of a quantum mechanical (QM) electronic structure methods for a biologically relevant solute molecule with a computationally much cheaper force field treatment of the environment [1] is now a standard tool in biophysics and medicine [2]. We have recently [3] presented an approach to such hybrid molecular dynamics (MD) simulations in which the atomic forces are calculated by grid-based density functional theory (DFT) for a solute molecule, while a polarizable molecular mechanics (PMM) force field is applied to the solvent environment. In contrast to many other QM/MM approaches, this DFT/PMM model dynamically accounts for electronic polarization effects in the solvent subsystem. For the solute, DFT provides the accuracy required to calculate biomolecular properties, e.g. infrared spectra, by MD [3].

The combination of these highly accurate physical descriptions pose a computational challenges to algorithms, implementation and parallelization. Our implementation, which combines the two MPI/OpenMP-parallel codes IPHIGENIE [5] and CPMD [6] has already proven to scale very well on HPC systems [7].

However, when the conformational space of a biomolecular solute, e.g. a protein, is probed, even the most efficient plain MD algorithms yield a poor sampling performance, because at ambient temperature the time to visit all conformations is much larger than the simulation time. This sampling problem is overcome by so-called generalized ensemble methods like simulated solute tempering (SST) [8]. To enable the study of biomolecular conformational dynamics by DFT/PMM-MD, we have therefore extended our approach by SST [10]. Concurrently, another parallelization layer is introduced now enabling the use tens of thousands of cores.

2. ALGORITHMS

The key task of calculating the DFT/PMM interaction forces from the corresponding Hamitonian is efficiently performed. Long-range interactions are treated using hierarchically nested fast multipole expansions [4]. Extending the original algorithm, we now use equally sized small subvolumes (voxels) of the DFT grid as smallest clusters of the fast multipole hierarchy. This measure largely enhanced the algorithmic accuracy and the efficiency while the scalability was preserved [9].

The sampling problem is resolved by a parallel generalization of SST [8] to fully polarizable systems (pSST) [10]. It exploits the fact that at higher temperatures, the transitions between molecular conformational states occur much more often. Therefore, pSST simulates multiple replicas of

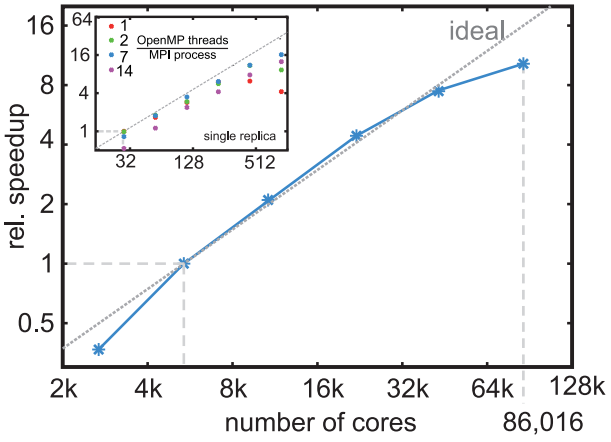

Figure 1: Scaling of IPHIGENIE/CPMD for 22 DFT and about 26,000 PMM atoms in a pSST ensemble spanning 128 replicas. Inset: scaling of a single replica for different MPI/OpenMP parallel setups.

the simulation system in parallel at different temperatures $T \in [T_{\text{low}}, T_{\text{high}}]$. Periodically, the replicas may jump to another temperature [8]. All replicas contribute to the weight parameters that ensure uniform replica distribution in temperature space.

3. IMPLEMENTATION

DFT/PMM-pSST is implemented in the MPI/OpenMP- parallelized PMM-MD C-code IPHIGENIE, which features an interface to the most recent version 4.1 of the DFT program CPMD [6] (FORTRAN90). For this purpose, CPMD is compiled as a library and statically linked to IPHIGENIE. This compilation strategy makes the hybrid program IPHIGENIE/CPMD easy to run on HPC systems, such as the SuperMUC petaflop system operated by Leibniz Supercomputing Centre (LRZ), Munich.

4. SCALING PROPERTIES

The strong scaling on SuperMUC Phase 2 was tested treating an alanine dipeptide molecule (22 atoms) as DFT fragment (MT/BLYP, cutoff E_{cut}=100 Ry) and describing its aqueous environment by 4487 polarizable six-point water molecules. Figure 1 demonstrates that the DFT/ PMM-pSST-MD hybrid approach with 128 replicas scales linearly up to 43,008 cores, with still good performance gain at all 86,016 cores of SuperMUC Phase 2.

The optimal parallel setup for the 2x14 core topology of the Intel Haswell processor employed by SuperMUC Phase 2 was determined by a reduced setup featuring only one replica (inset of Figure 1) and featured 4 MPI processes with 7 OpenMP threads each per node. The well-known excellent scaling performance of CPMD is obviously not hampered by the interface to IPHIGENIE in a single replica, and the additional pSST layer of parallelization largely increases the scaling.

5. SUMMARY

Our new DFT/PMM-pSST approach extends our combination of PMM force fields with grid-based DFT by the pSST generalized ensemble idea.

IPHIGENIE/CPMD is an unique tool which
- combines an accurate DFT solute description with a PMM description of the solvent environment
- scales to 86,016 Intel Haswell cores of SuperMUC Phase 2;
- makes studies of conformational and spectroscopic properties of large DFT molecules solvated in accurately modeled condensed phase feasible.

IPHIGENIE/CPMD now enables large-scale applications targeting e.g. the mechanisms of protein folding.

6. REFERENCES

[1] A. Warshel and M. Levitt. *J. Mol. Biol.*, 103:227, 1976.
[2] H. M. Senn and W. Thiel. *Angew. Chem. Int. Ed.*, 48:1198, 2009.
[3] M. Schwörer, B. Breitenfeld, P. Tröster, S. Bauer, K. Lorenzen, P. Tavan, and G. Mathias. *J. Chem. Phys.*, 138:244103, 2013.
[4] K. Lorenzen, C. Wichmann, and P. Tavan J. Chem. Theory Comput., 10:3244, 2014
[5] IPHIGENIE is developed under GPL and can be downloaded at https://sourceforge.net/projects/iphigenie
[6] *CPMD: Car–Parinello Molecular Dynamics, version 4.1.* © IBM Corp 1990–2015 and MPI für Festkörperforschung Stuttgart 1997–2001, `www.cpmd.org`.
[7] M. Schwörer, P. Tavan, F. Jamitzky, H. Satzger, and G. Mathias. Poster @ SC13, 2013
[8] R. Denschlag, M. Lingenheil, P. Tavan, and G. Mathias. *J. Chem. Theory Comput.*, 5:2847, 2009
[9] M. Schwörer, K. Lorenzen, G. Mathias, and P. Tavan. *J. Chem. Phys.*, 142:104108, 2015.
[10] M. Schwörer, C. Wichmann, E. Gawehn, and G. Mathias. *(submitted to JCTC)*