ANL/ALCF/ESP-13/14

# NAMD - The Engine for Large-Scale Classical MD Simulations of Biomolecular Systems Based on a Polarizable Force Field

*ALCF-2 Early Science Program Technical Report*

**Argonne Leadership Computing Facility**

**About Argonne National Laboratory**
Argonne is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC under contract DE-AC02-06CH11357. The Laboratory's main facility is outside Chicago, at 9700 South Cass Avenue, Argonne, Illinois 60439. For information about Argonne and its pioneering science and technology programs, see www.anl.gov.

**Availability of This Report**
This report is available, at no cost, at http://www.osti.gov/bridge. It is also available on paper to the U.S. Department of Energy and its contractors, for a processing fee, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
phone (865) 576-8401
fax (865) 576-5728
reports@adonis.osti.gov

**Disclaimer**
This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor UChicago Argonne, LLC, nor any of their employees or officers, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of document authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, Argonne National Laboratory, or UChicago Argonne, LLC.

ANL/ALCF/ESP-13/14

# NAMD - The Engine for Large-Scale Classical MD Simulations of Biomolecular Systems Based on a Polarizable Force Field

*ALCF-2 Early Science Program Technical Report*

prepared by
Benoit Roux[1], Yun Luo[2], and Wei Jiang[2]
[1]University of Chicago
[2]Argonne Leadership Computing Facility, Argonne National Laboratory

May 7, 2013

# NAMD - The Engine for Large-Scale Classical MD Simulations of Biomolecular Systems Based on a Polarizable Force Field

PI: Benoit Roux
ESP Postdoc: Yun (Lyna) Luo
Catalyst: Wei Jiang

*ALCF Early Science Program Technical Report – April 2013*

Large-scale Molecular dynamics (MD) simulation based on atomic models provide a powerful tool to understand the structure-dynamics-function relationships of important biological systems. Recent advances in computing power, especially in supercomputer, make MD simulation more and more popular in modern scientific community. However, the power of classical MD simulation has been limited mainly by the accuracy of the potential energy function and the efficiency of the dynamic algorithm enabling the adequate configurational sampling. This Early Science Project is aimed to make MD simulation method go beyond current limit using the leadership supercomputer Blue Gene/Q Mira. To address the issue of potential energy accuracy, classical Drude oscillator is developed to take into account the electronic polarizability in molecular systems. In order to develop and test the Drude polarizable force field for biomolecules, extensive MD simulations of typical polypeptides, lipids are needed. This requires efficient dynamic sampling algorithm that allows fast sampling. The sampling issue can be addressed by using advanced strategies based on multiple copies in order to enhance the sampling efficiency of brute-force MD. One such method is called replica-exchange MD (REMD).

In the replica-exchange MD (REMD) approach, several copies of the molecular system are simulated concurrently under slightly different conditions, e.g., different temperatures or Hamiltonians. Attempts are periodically made to exchange parameters between different replicas using a Metropolis Monte Carlo acceptance criterion, thus insuring Boltzmann-weighted statistics. Such multiple copy algorithms (MCAs) have been shown that enhance the sampling and free energy convergence.

NAMD is currently one of the most optimal and efficient programs to carry out classical MD simulations of biomolecular systems. NAMD is built on top of charm++ that obtains adaptive overlap of communication and computation. However the previous implementation of REMD on NAMD is driven by external job script, which can only be used on small clusters. This limits the use of REMD algorithm on the leadership supercomputer platform. The real power of REMD relies on the use large number of replicas (up to thousands of replicas), and high frequency of exchange (>1 $ps^{-1}$). Such requirement is beyond the reach of small computer clusters. It has become essential to develop extremely scalable MPI level REMD scheme to make full use of leadership supercomputers evolving toward multi-millions of cores.

The standard Charm++ was enhanced to support parallel/parallel simulations with multiple copies within a single program execution. The implementation of multiple

copies is completely at the MPI level. In the MPI machine layer of Charm++, MPI_Comm_split function splits the default MPI_COMM_WORLD into multiple local sub-communicators, each of which runs an independent Charm++ and NAMD instance. A secondary set of MPI communicators, each spanning like ranks within the local (sub)communicators, allows exchanges between the independent NAMD instances. This exchange communication is implemented through new APIs in both Charm++ converse layer and the NAMD Tcl scripting interface to provide a user-friendly interface. The breakthrough of such implementation is shown in three aspects: Scalability, Generality and Flexibility, Easy Post Processing.

### I. Scalability on Blue Gene/Q Mira

MPI parallel/parallel multiple copy algorithm in NAMD is much more efficient than the old Tcl server and socket connections driving a separate NAMD process for every replica. Because NAMD program does not need to be restarted after each exchange, and the communication overhead is minimized by swapping parameters (temperature, lambda or biasing potential in ColVars) instead of coordinates (as in CHARMM), there is almost no performance lost even with large number of replicas and high exchange frequencies.

Pthread is implemented in charm++ so NAMD runs well in SMP mode on BG/Q Mira. The machine layer of charm++ interfaces with PAMI. QPX is implemented in the pairwise force/energy calculation of NAMD with xlc compiler intrinsic functions. Multithreaded mpi-smp version Charm++ 6.4.0 is built for running REMD

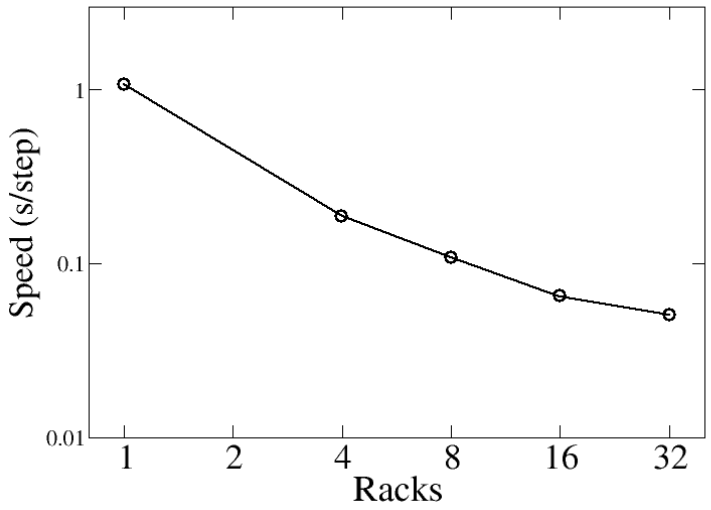


Figure 1 gives benchmarks for a multiple copies run with 1024 replicas. The calculation was carried out on BG/Q Mira using NAMD 2.9 with Charm++ 6.4.0 enhanced to support extremely scalable MCAs simulations. The REMD simulation scales up to 32 racks on BG/Q (i.e., 195 atoms/core) using16 MPI ranks per node and 2 working threads and 1 communication thread per rank.

## II. Generality and Flexibility

To apply REMD and its variant, one usually needs to change the source code of some molecular dynamics (MD) simulation packages, which may not be so easy for general users. Here, we tested that a variant of REMD algorithms (T-REMD, FEP/REMD, US/REMD and REXAMD) could be performed by a standard NAMD2.9 package without touching the source code. Through the Tcl script user can define which parameter (temperature, lambda or biasing potential in ColVars) to exchange and which acceptance criterion to apply. Both NVT and NPT ensemble can be applied to REMD. All of the topology is contained in the user-defined replica_neighbors Tcl proc. User can define a list of exchange neighbors for a given replica and define any exchange pattern on any topology. Thus it works for any-dimensional sampling in any-shaped regions.

## III. Easy Post Processing

Replica exchanges and energies are recorded in the .history files and the biasing potential applied in US/REMD is recorded in colvars.traj files. sortreplicas, found in the namd2 binary directory, is a program to un-shuffle replica trajectories to place the same temperature or order parameter frames in the same file.


**Argonne Leadership Computing Facility**
Argonne National Laboratory
9700 South Cass Avenue, Bldg. 240
Argonne, IL 60439

www.anl.gov


Argonne National Laboratory is a U.S. Department of Energy
laboratory managed by UChicago Argonne, LLC