Present and Future Computing Requirements for

Sampling diffusive dynamics on long timescales, and simulating the coupled dynamics of electrons and nuclei

Thomas Miller

Caltech

NERSC BES Requirements for 2017
October 8-9, 2013
Gaithersburg, MD

1. Project Description

Thomas F. Miller / Caltech

- Summarize your project(s) and its scientific objectives through 2017

- Our present focus is:
  - To understand reactive tunneling in enzyme-catalyzed hydrogen-transfer, electron-transfer, and proton-coupled electron transfer reactions.
  - To understanding and controlling protein targeting and protein expression in cells, at the amino acid level.

- By 2017 we expect to develop and utilize new methods to achieve these focuses.

2. Computational Strategies

- We approach this problem computationally at a high level by utilizing quantized molecular dynamics (MD), classical MD, and coarse-grained simulation methods.

- The codes we use are either developed in-house or modified versions of existing programs, such as NAMD, GROMACS, DL_POLY, or AMBER.

- These codes are characterized by these algorithms: molecular dynamics (ODE) integration, FFT

- Our biggest computational challenges are: FLOP availability, queue times.

2. Computational Strategies

- Our parallel scaling is limited by: potential energy surface evaluation, number of DOFs, number of weakly coupled trajectories

- We expect our computational approach and/or codes to change (or not) by 2017 in this way:
  - Increased use of GPUs
  - Increased computational cost (and parallel scalability) of potential energy surface evaluations.
  - Increased system sizes and desired timescales of simulation

3. Current HPC Usage (see slide notes)

- Machines currently using: Edison, Hopper, Carver

- Hours used in 2012-2013 (list different facilities):
  - NERSC: 13 M in 2013 cycle (22M in 2012 cycle)
  - OLCF (Jaguar): 9 M
  - ALCF (Intrepid, Vesta): 8M

- Typical parallel concurrency and run time, number of runs per year:
  - 500-2000 processors per set of jobs (tight parallelization w/i trajectories, weak among)
  - 6-12 hours per trajectory
  - Hundreds of runs per year

- Data read/written per run:
  - Modest: 1-5 GB

3. Current HPC Usage (see slide notes)

- Memory used per (node | core | globally)
  - 50-100 MB per core
- Necessary software, services or infrastructure
  - MD packages, FFT packages, electronic structure packages (Molpro, Gaussian)
- Data resources used (/scratch,HPSS, NERSC Global File System, etc.) and amount of data stored
  - Low storage requirements (<1 TB), only 167K SRUs used in 2013 for HPSS

4. HPC Requirements for 2017

(Key point is to directly link NERSC requirements to science goals)

- Compute hours needed (in units of Hopper hours)
  - 50-100 M (increasing demands of potential energy surface (PES) evaluations)
- Changes to parallel concurrency, run time, number of runs per year
  - Greater PES parallelizability (10-20x increase), no other major changes
- Changes to data read/written
  - No major changes expected
- Changes to memory needed per (core | node | globally)
  - Greater memory requirements (1 GB per core)
- Changes to necessary software, services or infrastructure
  - None anticipated

5. Strategies for New Architectures

- Our strategy for running on new many-core technologies (GPUs or MIC) is:
  - Expanded utilization of GPUs in many/most simulation studies
- To date we have prepared for many core by:
  - Utilization of local (Caltech-based) GPU machines
  - Utilization of GPU implementations of classical MD packages
  - Working with CS groups at Caltech and Pomona Colleges to develop efficient GPU versions of the coarse-grained simulations
- We are already planning to do:
  - Develop and implement GPU versions of existing codes
- To be successful on many-core systems we will need help with:
  - Efficient implementation, scaling tests

5. Summary

- What new science results might be afforded by improvements in NERSC computing hardware, software and services?
  - All aspects of our work utilize NERSC resources – a critical resource for our progress
  - Many applications are limited by available computational resources: More hardware access, more progress. Also, NERSC support is important for efforts.
  - Public software is a less critical bottlenecks for us at this time, but the currently available codes are essential to our work.

- Recommendations on NERSC architecture, system configuration and the associated service requirements needed for your science :
  - Don't make queues too short.
  - Recognize the value of ensemble calculations, while encouraging efficient parallelization.

5. Summary

- NERSC generally refreshes systems to provide on average a 2X performance increase every year. What significant scientific progress could you achieve over the next 5 years with access to 32X your current NERSC allocation?
  - Quantitate prediction of biological and catalytic processes, design of new catalysts
  - Theory-guided enhancement of integral membrane protein expression.
  - Progress towards the computational design of electrolyte and electrode materials.

- What "expanded HPC resources" are important for your project?
  - Increased numbers of processers.
  - Increased processor speed.

- General discussion