BLUE WATERS

SUSTAINED PETASCALE COMPUTING

June 3, 2019

The Use of Petaflop Simulations in Optimization of Amber Force Field

Victor Anisimov, Valeri Poltev, Rodrigo Galindo-Murillo, Emil Briggs, Wenchang Lu, Thomas E. Cheatham III, Jerry Bernholc

2019 Blue Waters Symposium, June 3 – June 6, 2019, Sunriver, OR

NCSA NSF GREAT LAKES CONSORTIUM FOR PETASCALE COMPUTATION CRAY

I use Blue Waters to:

- **Initiate multidisciplinary cross team collaboration of 4 labs**
  - *Bernholc, Cheatham, NCSA, Poltev*
- **Run large-scale (1K-4K nodes) electronic structure simulations**
  - *CCSD(T) and DFT MD*
- **Create experimental and theoretical dataset for validation of approximate computational methods**
  - *Experimental Cluster Geometries (Dimer, Trimer, Tetramer)*
  - *Computed Interaction Energies*
- **Optimize electrostatic model of Amber force field**
  - *Fit Charges to Condensed Phase*
- **Run Lennard-Jones Parameter Sweep**
  - *Grid Search*

Project Motivation

- MD simulations based on classical force fields are the major consumers of supercomputing cycles
- The quality of underlying parameters determines the accuracy of bio-molecular force fields
- Well-optimized parameters can save tons of compute time and human resources
- Force Field optimization is challenging
  - Need to reproduce condensed-phase experimental data

Real-space Multi-Grid method (RMG)

- **Density functional equations solved directly on the grid**
- **Multigrid techniques remove instabilities by working on one length scale at a time**
- **Non-periodic boundary conditions are as easy as periodic**
- **Short-range discretizations**
- **Allows for efficient massively parallel implementation**

See E. L. Briggs, D. J. Sullivan and J. Bernholc Phys. Rev. B 54, 14362 (1996).

Ultrasoft pseudopotentials:

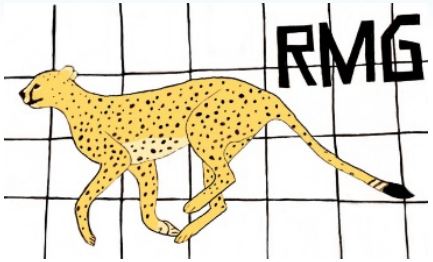
M. Hodak, S. Wang, W. Lu and J. Bernholc, PRB 76, 085108 (2007)

Kohn-Sham DFT/Orbital-free DFT for embedding biomolecules:

M. Hodak, W. Lu, J. Bernholc, JCP 128, 014101 (2008)

Performance on 3,872 Cray XK7 (K20x GPU)
Blue Water nodes: 1.14 PFLOPS

RMG Main Features



- Gamma and k-point calculations, spin polarization
- Runs on Linux/Unix, Windows and Mac
- Scalable to hundreds of thousands of CPU cores
- Supports thousands of GPU-nodes with multiple GPUs per node
- OpenBabel support for importing atomic structures
- Supports both ultrasoft and norm-conserving pseudopotentials
- Wide choice of DFT functionals including van der Waals via vdW-DF
- Hellman-Feynman forces, structure optimization and molecular dynamics
- To be released: optimally localized orbitals, quantum transport via self-consistent non-equilibrium Green's function (NEGF) formalism
- Upcoming capabilities: DFT+U, hybrid functionals, semi-local pseudopotentials

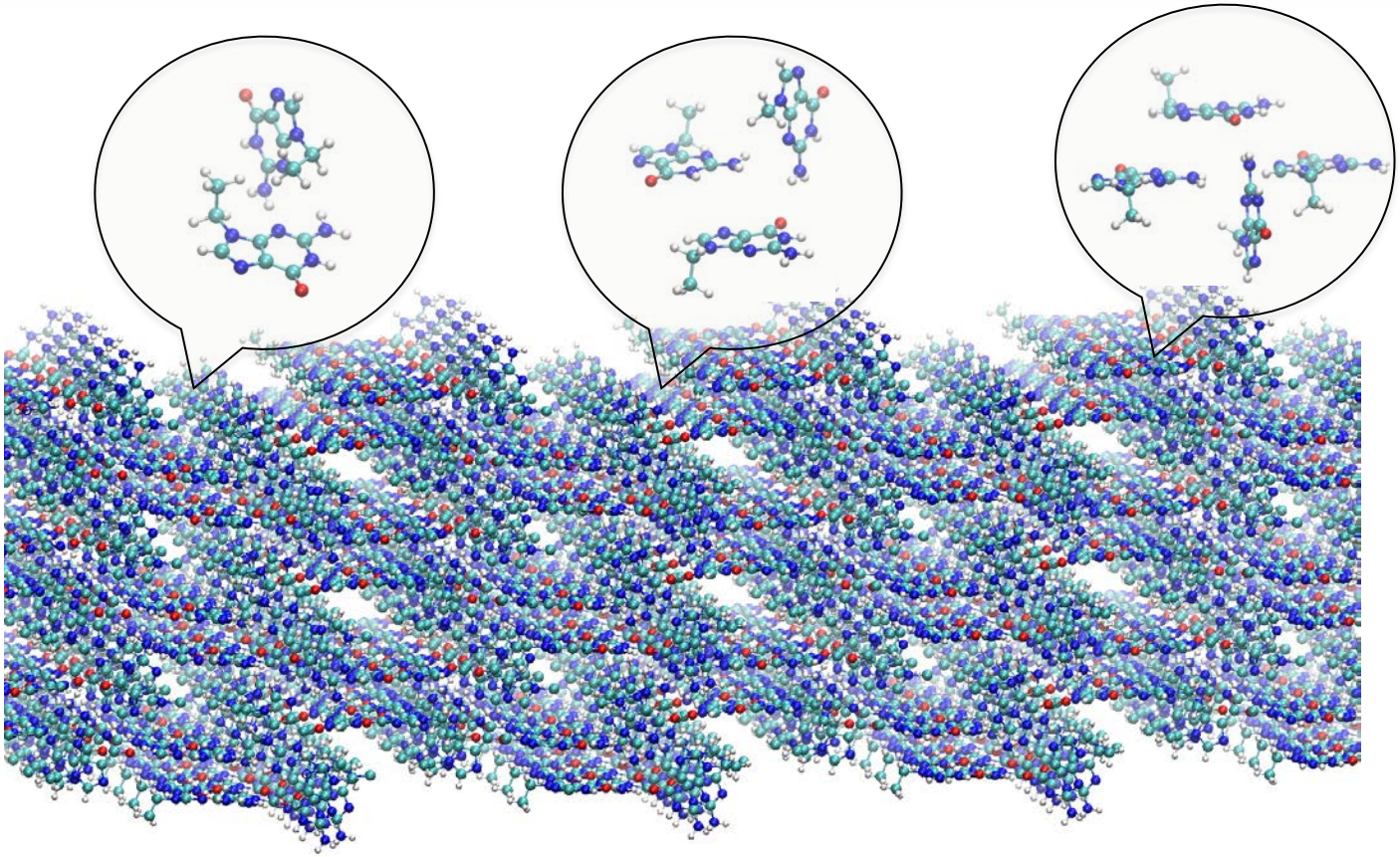
RMG Code Release

- Code & binaries (for Linux distr., Mac, Win)
- Blue Waters portal for Cray XE6/XK7

https://bluewaters.ncsa.illinois.edu/rmg

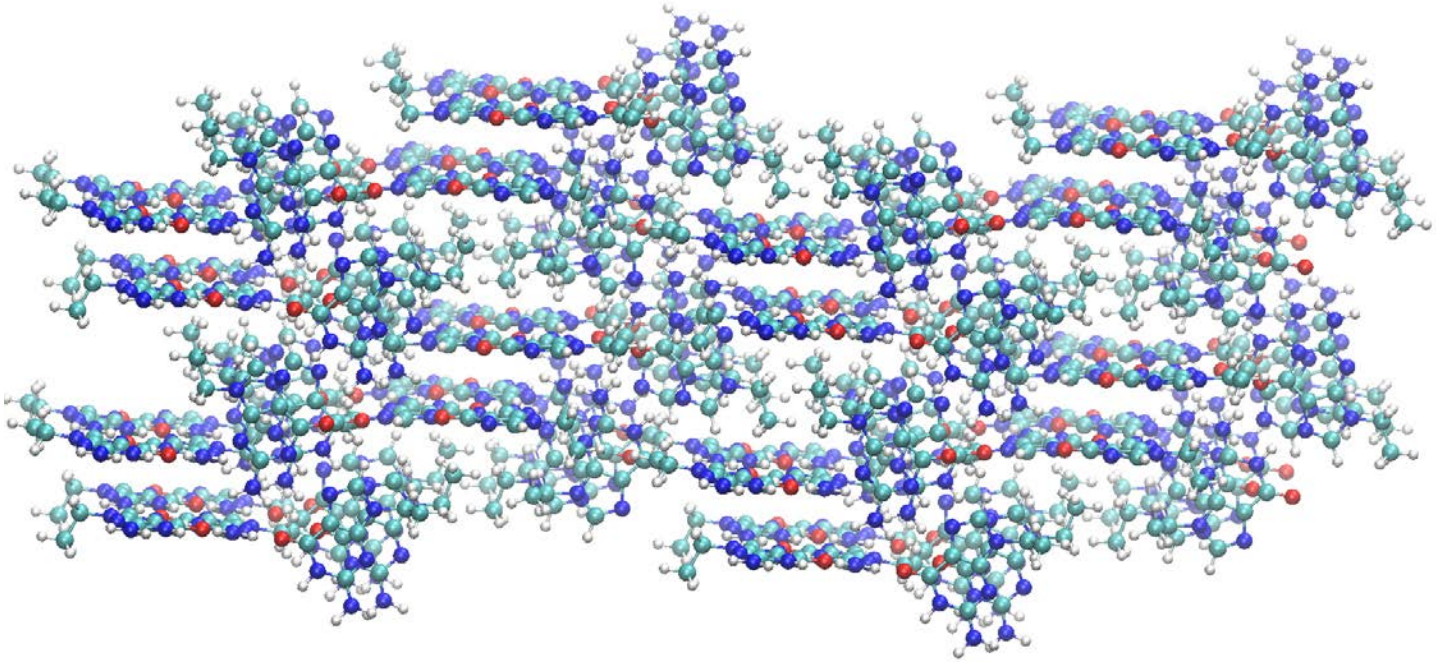
- Part of NSF's Sustained Petascale Performance benchmarks
  https://bluewaters.ncsa.illinois.edu/benchmarks
- Funded by Exascale Computing Project
- Graphical User Interface (GUI) for input & output
- Examples, documentation & discussion forum on sourceforge

Cluster Extraction from Crystals



- Processed 12 crystals of 5 nucleic acid bases
- Extracted 147 X-ray dimers, 1618 trimers, 11355 tetramers

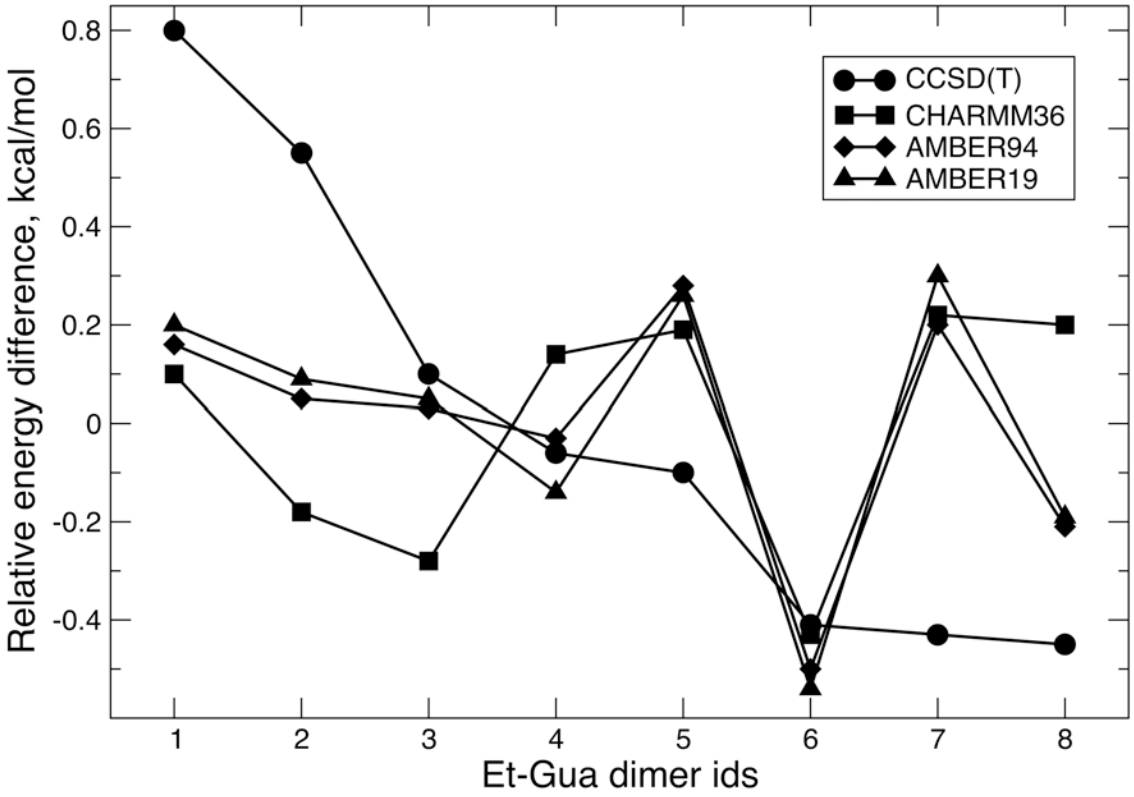
QM MD: Unit Cell of Et-Gua Crystal, 950 fs

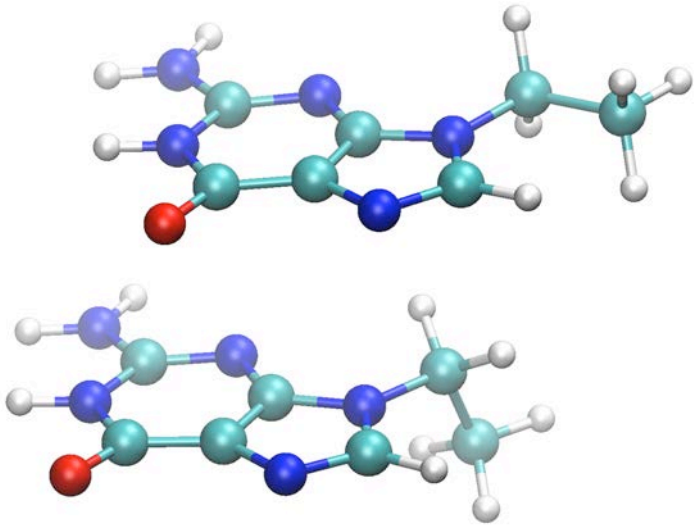


Molecular Dynamics simulation of Guanine crystal performed at the Density Functional Theory level to sample the conformational space of base-base interacting pairs in vicinity of the experimental thermodynamic energy minimum. Computed on 1024 BW nodes.

Extracted 216 dimer clusters from QM MD: 0.1 Å < RMSR < 0.2 Å

Relative Base-Base Interaction Energy





- Relative interaction energy difference vs. the average value
- Dimer id 3 corresponds to the crystal structure
- The remaining seven structures carry perturbed geometries obtained from QM MD

Resource Required for CCSD(T) Computations

- NWChem CCSD that includes previous NCSA contribution
- CCSD(T)/6-311++G** level of theory
- Singles and doubles for Et-Gua dimer (44 atoms including hydrogen) take 5651 seconds on 1024 XE6 nodes
- The portion of perturbative triples requires 5392 seconds on 4096 XE6 nodes
- Conducted 154 such computations

Intermolecular Interaction Energy in Clusters

Method	#records	AverErr	RMSD	MaxErr	Cluster ID
CH36	147	2.09 ± 1.95	2.85	8.34	018-dimer-guah-guah-01-06
AM94	147	2.09 ± 1.71	2.70	7.66	059-dimer-cytm-cytm-01-14
AM19	147	1.93 ± 1.57	2.48	7.14	018-dimer-guah-guah-01-06

Interaction energy is in kcal/mol

Amber19: Optimized force field

Amber19 has the lowest error

Conclusions:

- Optimization of bio-molecular force fields requires petascale resources
- Project produced the largest dataset of 147 dimer, 1618 trimer, and 11355 tetramer clusters of five nucleic acid bases extracted from 12 high-resolution crystals of nucleobases
- Cluster extraction from crystals produces the most important points on the potential energy profile of base-base interactions
- QM MD of crystals samples the energy profile near the experimental minima
- Fitting charges to X-ray clusters improves agreement with CCSD(T)
- New force field, Amber19 outperforms both Amber94 and CHARMM36

Acknowledgements

- This research is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (award number OCI 07-25070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications.

- This work is also part of the PRACs “Ensembles of molecular dynamics engines for assessing force fields, conformational change, and free energies of proteins and nucleic acids" and “Petaflops simulation and design of nanoscale materials and devices”. PRAC allocation support by the National Science Foundation (award numbers ACI-1515572 and ACI-1615114).