High-Throughput Multiobjective Genetic-Algorithm Workflow for In Situ Training of Reactive Molecular-Dynamics Force Fields

Ho Ching Justin Cheng, Pankaj Rajak, Chunyang Sheng, Rajiv K. Kalia, Aiichiro Nakano, Priya Vashishta, Evan Brown

Collaboratory for Advanced Computing & Simulations
Dept. of Computer Science, Dept. of Physics & Astronomy,
Dept. of Chemical Engineering & Materials Science,
Dept. of Biological Sciences
University of Southern California

Email: {hochingc, rajak}@usc.edu

Spring Simulation Multi-conference
Chair: Dr. Lukas Polok
Westin Pasadena, U.S.A.
April 6, 2016

H. C. Cheng et al.,
in Proc SpringSim HPC2016 (2016)
Reactive Force Field (ReaxFF)

The computational bottleneck is iterative determination of atomic charges to minimize $E_{\text{Coulomb}}$ using conjugate gradient (CG) method at every MD step.

$E_{\text{Coulomb}}(\mathbf{r}_N, q^N) = \sum_i \chi_i q_i + \frac{1}{2} \sum_i \sum_j q_i H(r_{ij}) q_j$

Electronegativity Coulombic interaction

---


For a small (150 atoms) system, both QMD & RMD simulations show
• Formation of much more Si-O bonds compared with C-O
• Condensation of clusters
Genetic Algorithm Parameter Fitting

Multiobjective genetic algorithm (MOGA)

To minimize the error for Si-O, Si-O and C-C between QMD and ReaxFF simulation NSGA-II is used


Since formation of Si-C, Si-O & C-C is conflicting in nature, the final solution is a Pareto front
File-based Workflow Diagram (MOGA)

- Separate directory for each gene
- Four files per directory for every generation
Evolution of Pareto Front

- Population converges after 180 generations

Red: final Pareto solutions
Blue: current population
Parameter Training with Time Series

- With published values of parameters, there are discrepancies in number of C-C bonds as a time series
- After training with QMD results, the time series of bond counts fit much better
Client-server Based Workflow (iMOGA)

- The file-based workflow is not scalable for large population size (\(4N\) files for each generation; \(N = \text{GA population}\))
- Remove bottleneck of file I/O
- Replace with piping within each node, TCP/IP socket communication across nodes
Reduced File I/O Achieves Better Scalability

- **Weak scaling test:** number of processors scales linearly with population size.

- In the file-based workflow (MOGA), runtime is linear with population size since communication is expensive.

- With the TCP/IP & piping methods, iMOGA achieves weak-scaling parallel efficiency of 0.848 with 120 nodes.
Oxidation of SiC Nanoparticle

- Reactive molecular dynamics (RMD) simulations: Diameter = 10 nm (100K atoms), 46 nm (10M atoms) & 100 nm (112M atoms) on 786,432-processor IBM Blue Gene/Q

- Formation of nanocarbon, embedded within SiO$_2$ shell

K. Nomura et al., Sci. Rep. 6, 24109 ('16); J. Insley et al., IEEE/ACM SC16
Conclusion

1. MOGA is used to train ReaxFF to give more accurate simulations
2. Scalable workflow is developed with piping & sockets to remove the file I/O bottleneck
3. Optimized force field was used in 112 million-atom simulation

Research supported by DOE-CMS/BES/INCITE, NSF-CDI, DTRA, AFOSR