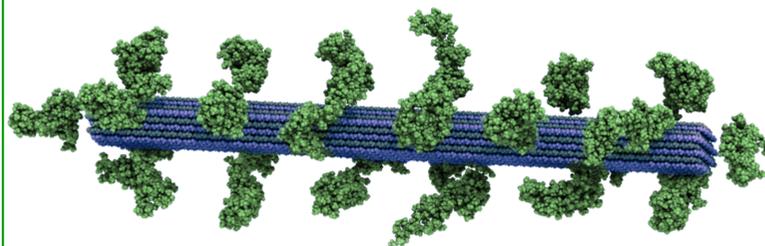


Roland Schulz, Benjamin Lindner, Loukas Petridis and Jeremy C. Smith
Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, TN-37831

Abstract: This poster outlines a strategy for all-atom molecular dynamics (MD) simulation of multimillion-atom biological systems, using lignocellulosic biomass as test system. In the era of petaflop supercomputers, such simulation is limited by the parallel efficiency of the MD algorithms. The bottleneck for highly-parallel all-atom simulations is the computation of the electrostatic interactions. To overcome the limitations of the commonly-used Particle Mesh Ewald we present the performance and accuracy of the reaction field (RF) method for electrostatics. Comparison of RF and PME simulations of cellulose and water show very good agreement. The RF method enables simulation to be performed of a 3.3 million atom system yielding 27.5ns/day.

Approach

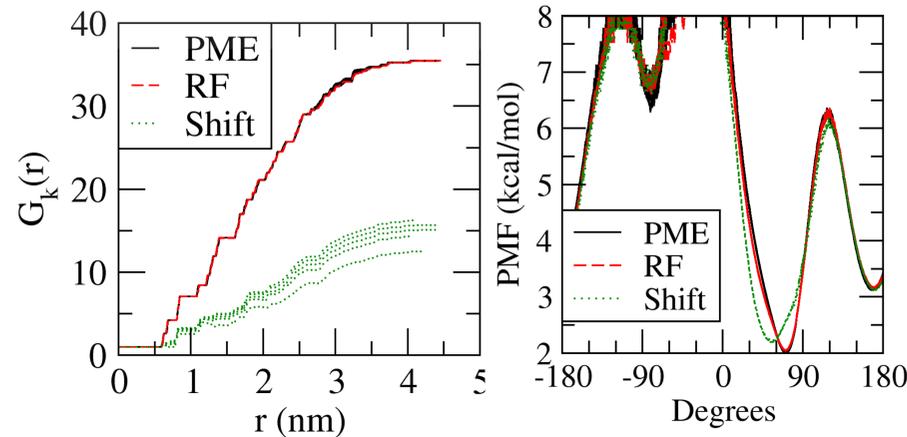
- Realistic atomic-detailed model of cell walls
- Reaction field for electrostatic
- GROMACS 4



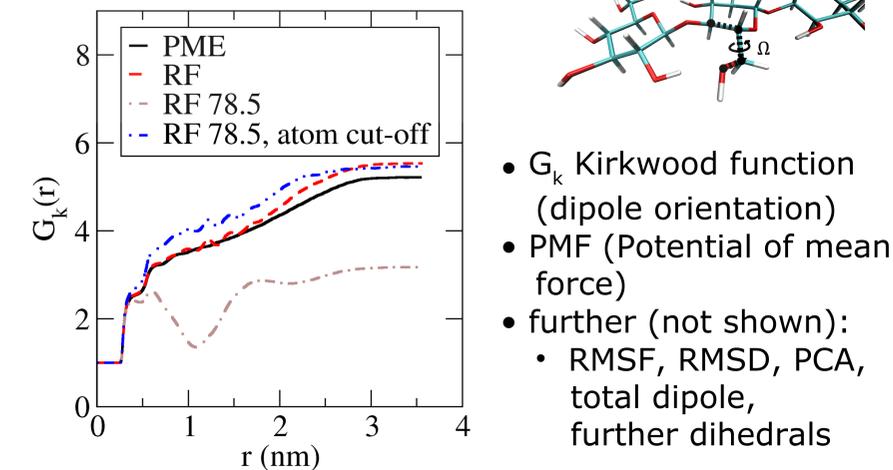
- 3.3 million atoms
- 52 lignin polymer (61 monomers)
- 36 cellulose chains (80 monomers)
- Explicit water

Accuracy

• Cellulose



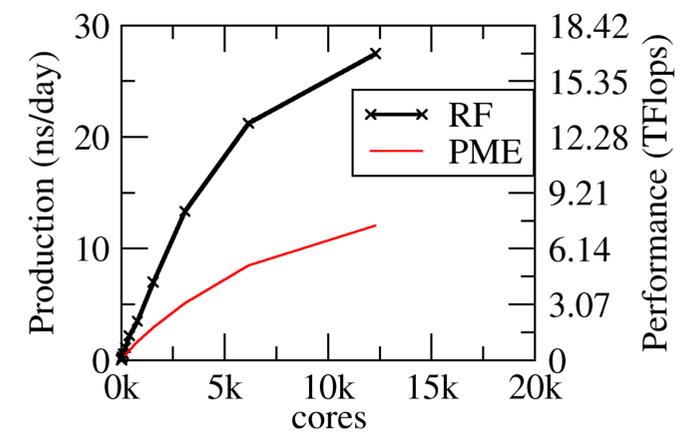
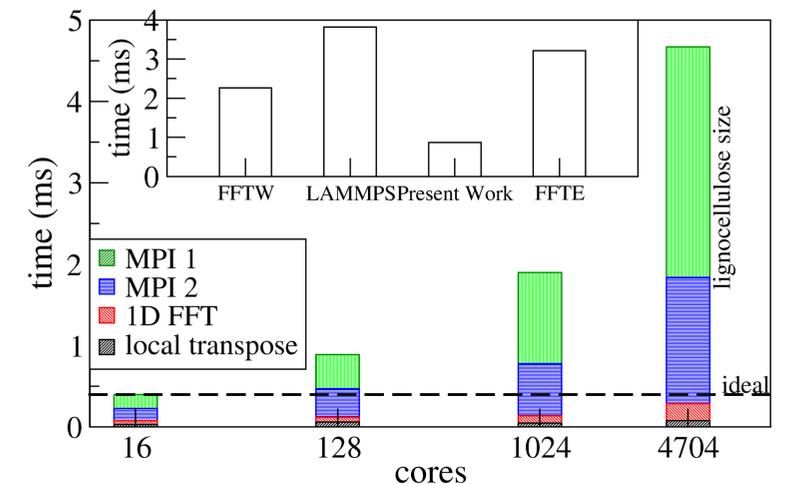
• Water



- G_k Kirkwood function (dipole orientation)
- PMF (Potential of mean force)
- further (not shown):
 - RMSF, RMSD, PCA, total dipole, further dihedrals
- Good agreement for all tests between PME and RF
- Agreement between NAMD and GROMACS

Performance

• Scaling of FFT (required for PME)



- PME with NAMD, GROMACS/PME currently slower
- 5.5 million atom model system
- 38ns/day with 5fs ts, scaling to 30k cores (2fs)

Discussion: The scaling of MD codes is restricted by global communications. MD simulations using the PME method faces weak-scaling problems. While for small systems simulations achieving over 100ns/day are possible, for larger systems the global communication for the FFT (MPI_Alltoall) takes longer than the time available for one time-step at that speed. As shown in the Scaling Figures, the use of the RF method greatly improves the strong scaling of million-atom systems, to the point where 28ns/day are obtained when a 5.4 million atom system is run on 30k cores with 2 fs time-step. The RF has already been used in numerous studies, but questions have remained about its accuracy. The present work demonstrates that the RF method does not appear to compromise the accuracy of MD simulation of the test system: lignocellulosic biomass. RF and Shift are similar in the sense that they do not consider explicitly electrostatic interactions between atoms separated by more than the cutoff distance. Therefore, one might have expected the RF and Shift methods to yield similar results. However, our findings suggest a different picture. All benchmarks show very good agreement between RF and PME, while the Shift method shows several significant artifacts. Also the RF and PME simulations show very good agreement for the long range electrostatics in water. Using RF for the electrostatic calculation, removes the biggest inherent limitation of the scaling of MD. The FFT part of PME requires two global MPI_Alltoall communications. These MPI_Alltoall communications do not exhibit a good weak scaling. Therefore the performance (in ns/day) for big systems is limited with PME and can only be improved by different electrostatic treatment, as shown here for RF. **Acknowledgments:** Thanks to GROMACS developers, P. Bjelkmar, NCCS at ORNL, Genomics:GTL Program, BESC, GST.