Contribution ID: **67** Type: **Professional Micro-Talk**

Scalable In-Situ Analytics for High-Throughput Ensemble Molecular Dynamics

Next-generation computing systems are expected to have dramatically higher performance than current systems. However, I/O bandwidth and parallel file system capacity will not grow at the same rate. This I/O bottleneck constitutes a major challenge for the analysis of data-intensive high-performance computing (HPC) simulations.

This work targets one of the most common applications on petascale: molecular dynamics (MD) simulations studying the classical time evolution of a molecular system at atomic resolution. We enable the development of new comprehensive workflows for MD simulations in which HPC meets data analytics. In situ processing, combined with in-memory computing, has emerged as a solution to overcome I/O bottlenecks in HPC systems. How to do this efficiently from the infrastructure and human resources perspectives is still a challenging open problem. We address this challenge by designing and developing A4MD: a workflow engine that enables better scientific knowledge discovery in the domain of MD. A4MD features plug-and-play workflow composition of reusable building blocks; transparent in situ annotation capability for user-defined methods; simulator-agnostic runtime trajectory analysis; support for stopping, starting and restarting simulations; and ensemble workflow patterns optimized for HPC environments. A4MD empowers scientists to integrate simulation and analytics into complex workflows for runtime detection of changes in structural and temporal molecular properties. This knowledge of molecular structures' transformations at runtime can be used to steer simulations to more promising areas of the simulation space, identify the data that should be written to congested parallel file systems, and index generated data for retrieval and post-simulation analysis.

We demonstrate the capabilities of A4MD by studying a case of enhanced adaptive sampling for the exploration of the conformational space of the FS peptide protein. We model the execution of an ensemble of trajectories and analyze the overall throughput obtained by the workflow. Using A4MD, we integrate runtime trajectory analysis to automatically detect inefficiencies in the behavior of the simulations.

Student?

No

Supervisor name

Supervisor email

Primary authors: Dr CAINO-LORES, Silvina (Department of Electrical Engineering and Computer Science, The University of Tennessee); Dr CUENDET, Michel (Department of Physiology and Biophysics, Weill Cornell

Medical College); Mr KETRON, Ross (Department of Electrical Engineering and Computer Science, The University of Tennessee); Dr WEINSTEIN, Harel (Department of Physiology and Biophysics, Weill Cornell Medical College); Dr TAUFER, Michela (Department of Electrical Engineering and Computer Science, The University of Tennessee)

Presenter: Dr CAINO-LORES, Silvina (Department of Electrical Engineering and Computer Science, The University of Tennessee)

Session Classification: Micro-talks

Track Classification: Materials Science