Extended Software Development Workshop: Topics in Classical MD

Location: CECAM-FR-RA, ENS Lyon
Webpage: https://www.e-cam2020.eu/event/esdw-topics-in-classical-md/?instance_id=6
Dates: April 3, 2019 to April 12, 2019
Organizers: David Swenson and Ralf Everaers

1 State of the art

This workshop was intended to provide both a venue for scientific discussion, and also an opportunity to learn about the best practices in modern software development, so that participating researching could hone their tools.

From the scientific standpoint, classical molecular dynamics is a broad umbrella, and researchers within specific topics of the field often aren't familiar with the state of the art within other topics. The scientific goal of this workshop was to introduce participants to the core science of several topics and to software related to those topics, while also encouraging them to dive more deeply into a topic of their interest by implementing new functionality as an E-CAM "module."

Of course, there are more possible topics than could be covered during the workshop. The three we selected were based on topics that proved interesting at the E-CAM State of the Art Workshop held in Vienna, in October 2018. The topics and related software packages we selected were:

* Trajectory-based rare event methods, with software package OpenPathSampling
* CV-based rare event methods, with software package PLUMED
* Neural network potential energy surfaces, with software package N2P2.

This workshop was also designed to introduce participants to software development best practices. These best practices have evolved over decades of experience in the software industry. They result in more maintainable and robust software, and facilitate rapid development of new approaches. Many research scientists do not use these best practices, and indeed, many are even unaware of them. As a result, scientific innovation is slowed by bad software development practices. This workshop included training about best practices, and participants were encouraged to use these best practices while implementing their modules.

2 Training provided

The presentations during the workshop fell into three categories: (1) introductions to best practices in software development; (2) presentations on the theories and software behind the specific topics the workshop focused on; and (3) contributed presentations from participants regarding their own research.

The presentations on software best practices included introductions to: git and related tools (Alan O'Cais); software testing (David Swenson); object oriented programming (Jony Castagna); performance and benchmarking (Emmanuel Quémener); and git/GitHub/GitLab workflows (David Swenson).

The presentations on the theories and software for the workshop topics were: trajectory-based rare event methods and OpenPathSampling (David Swenson); neural network potential energy surfaces and N2P2 (Christoph Dellago; Andreas Singraber); and CV-based rare event methods and PLUMED (Gareth Tribello).

Finally, the contributed presentations were:

* "Augmented Harmonic Linear Discriminant analysis" (Faidon Brotzakis)

* "The analysis of chromatin configurations in the nucleus using MD simulation" (Ali Farnudi)

* "Gold nanoparticles as amyloid-β fibril inhibitors" (Francesco Tavanti)

* "Progress index-guided sampling: an unsupervised protocol to boost molecular simulations" (Cassiano Langini)

Lectures from this workshop were recorded and stored on E-CAM's training portal at https://training.e-cam2020.eu/spaces/5ca35151e4b0fed490540623.

Emmanuel Quémener's presentation on performance dealt directly with some topics in HPC. The approach of the three software packages presented is generally to gain as much performance as possible from the underlying molecular dynamics codes that they integrate with.

The workshop had 21 attendees, including the 8 speakers. In addition, Christoph Dellago's presentation was open to public, and had a total of around 50 attendees (including workshop participants). Ten of the workshop attendees opened merge requests for E-CAM modules they intend to contribute, at https://gitlab.e-cam2020.eu/e-cam/E-CAM-Library/merge_requests.

One recurring theme in the discussions was the question of when and how to contribute to community codes, instead of developing your own from scratch.

3 List of software development projects

The following modules were in development during the ESDW:

* Add "nucleic" keyword to MDTraj atom selection language (https://gitlab.e-cam2020.eu/e-cam/E-CAM-Library/merge_requests/130)

* PLUMED dimensionality reductions (https://gitlab.e-cam2020.eu/e-cam/E-CAM-Library/merge_requests/131)

* Python bindings for "Open Dynamics Engine" (https://gitlab.e-cam2020.eu/e-cam/E-CAM-Library/merge_requests/132)

* Python bindings for PIGS/CAMPARI (https://gitlab.e-cam2020.eu/e-cam/E-CAM-Library/merge_requests/133)

* OpenMM implementation of d-AFED (https://gitlab.e-cam2020.eu/e-cam/E-CAM-Library/merge_requests/134)

* Update N2P2 weights via articifial MD (https://gitlab.e-cam2020.eu/e-cam/E-CAM-Library/merge_requests/135)

* Tools for analyzing training set size dependence in N2P2 (https://gitlab.e-cam2020.eu/e-cam/E-CAM-Library/merge_requests/136)

* Symmetry function parameter generator for N2P2 (https://gitlab.e-cam2020.eu/e-cam/E-CAM-Library/merge_requests/137)

* Gyration tensor for PLUMED (https://gitlab.e-cam2020.eu/e-cam/E-CAM-Library/merge_requests/141)

* Reweighted path ensemble for OPS (https://gitlab.e-cam2020.eu/e-cam/E-CAM-Library/merge_requests/142)

Both of the modules to create Python bindings will benefit from the HPC performance of the underlying codes. In addition, several modules are being designed specifically with performance considerations in mind. Implementing d-AFED for OpenMM is intended to benefit from OpenMM's GPU acceleration. One of the purposes of developing Python bindings for the PIGS routines in CAMPARI is to interface with MD engines in a way that avoids the performance costs that currently come from using disk as an intermediate.

We have not discussed the relevance of these modules with industrialists. However, it is reasonable to expect that some may be of interest. For example, the development of Python bindings for both the PIGS method in CAMPARI and for the ODE engine are intended to bring methods implemented in those software packages to a wider audience by lowering the barrier to entry. The modules related to N2P2 may be useful for the materials science industry.

Although these modules are not scheduled for any particular E-CAM deliverable, they would, once completed, be suitable to include in a future deliverable. These modules were selected based on participant interest; i.e., in response to the requests of users.

We expect that this workshop will be followed by a second face-to-face meeting. Pending confirmation of the availability of participants, we hope to schedule this for late October 2019.

Before the follow-up, we hope that the participants will have completely implemented all the functionality of their modules. The purpose of the follow-up will be to help them bring their modules up to the standards of the E-CAM Software Library, and to complete the documentation of their modules so they can be included in the E-CAM Software Library.

4 Participant list

Organizers

Everaers, Ralf
École Normale Supérieure de Lyon, France
Swenson, David
École Normale Supérieure de Lyon, France

Brotzakis, Zacharias Faidon - Department of Chemistry, University of Cambridge, Greece
Buchner, Florian - Vienna University of Technology, Austria
Carrivain, Pascal - Centre Blaise Pascal, ENS Lyon, France
Castagna, Jony - STFC Daresbury Laboratory, United Kingdom
Dellago, Christoph - University of Vienna, Austria
Farnudi, Ali - Department of Physics, Sharif University of Technology, Iran
Hammer, Valentin - University of Vienna, Austria
Kobsch, Anaïs - CNRS, France
Langini, Cassiano - University of Zurich, Switzerland
Li, Zhi - University of Lyon (ENS), France
Naleem, Nawavi - New York University Abu Dhabi, United Arab Emirates
O'Cais, Alan - Jülich Supercomputing Centre, Germany
Quemener, Emmanuel - Centre Blaise Pascal - Ecole Normale Supérieure de Lyon, France
Reiner, Madlen Maria - University of Vienna, Austria
Schaan, Renata - ENS Lyon, France
Singraber, Andreas - University of Vienna, Austria
Solomatova, Natalia - ENS de Lyon, France
Tavanti, Francesco - University of Modena and Reggio Emilia, Italy
Tribello, Gareth - Queen's University Belfast, United Kingdom