ECAM State-of-the Art Workshop: Large scale activated event simulations

Location: CECAM-AT
Webpage: https://www.cecam.org/workshop-0-1642.html
Dates: October 1, 2018 to October 3, 2018

1 State of the art

Running on powerful computers, large-scale molecular dynamics (MD) simulations are used routinely to simulate systems of millions of atoms providing crucial insights on the atomistic level of a variety of processes of interest in physics, materials science, chemistry and biology. For instance, MD simulations are extensively used to study the dynamics and interactions of proteins, understand the properties of solutions or investigate transport in and on solids. From a technological point of view, molecular dynamics simulations play an important role in many fields such as drug development, the discovery of new materials, oil extraction or energy production. Indeed, enormous amounts of data are produced every day by molecular dynamics simulations running on high performance computers around the world and one of the big challenges related to such simulations is to make sense of the data and obtain mechanistic understanding in terms of low-dimensional models that capture the crucial features of the processes under study. Another central challenge is related to the time scale problem often affecting molecular dynamics simulations. More specifically, despite the exponential increase in computing power witnessed during the last decades and the development of efficient molecular dynamics algorithms, many processes are characterized by typical time scales that are still far beyond the reach of current computational capabilities. The central goal of this ECAM-State-of-the-Art Workshop was to discuss computational approaches capable of addressing such time scale problems in complex systems in materials science and biophysics. Another important goal of the workshop was to debate about how to facilitate the use of simulation and modelling in industrial settings.

2 Major outcomes

Scientific discussions at the workshop centered around three fundamental computational challenges closely related to the time scale problem of classical MD simulation:

1) The calculation of the populations of metastable states of an equilibrium system. Such populations can be expressed in terms of free energies and hence this problem boils down to the efficient calculation of free energies. Sampling methods for such free energy calculations were discussed in several talks at the workshop.

2) The sampling of transition pathways between long-lived (meta)stable states and the calculation of reaction rate constants. Here the problem consists in sampling dynamical trajectories which can be very long for complex systems. Several talks discussed this problem and showed how path-based approaches can be used to study, for instance, the nucleation of gas hydrates, the crystallization of metals or the unbinding of a ligand from a protein. Simulations of such processes are facilitated by new scientific software tools, such as the Open Path Sampling (OPS) package, which provide flexible frameworks that can be easily extended and provide the tools required to handle large and complex systems.

3) The extraction of useful mechanistic information from the simulation data and the construction of low-dimensional models that capture the essential features of the process under study. Such models serve as the basis for the definition of reaction coordinates that enable in-depth studies of the process at hand. It has become evident during the workshop that new machine learning approaches have a huge potential for making progress on this very important problem. A number of talks discussed how to use supervised and unsupervised learning methods to identify collective variables that can be used to characterize complex molecular rearrangement. What has become clear in the talks and discussions during the workshop is that the key point in the application of machine learning methods to molecular simulations lies in the appropriate definition of descriptors on which the learning process is based. For instance, artificial neural networks have been shown to be able to provide very accurate and efficient representations of potential energy surfaces. A condition for the application of this approach, however, is that a set of molecular fingerprints is defined that contains sufficient information on molecular arrangements to make the energy prediction possible. The smart selection of descriptors is crucial in other machine learning approaches as well and it can be viewed as the interface between physics and chemistry and the world of machine learning.

The various themes discussed in the talks were picked up in two open discussion sessions on the first and second day of the workshop. The first discussion revolved on efficient path sampling methods and the identification of reaction coordinates. In particular, it was discussed how machine learning approaches can be used to make progress in this area and how extreme scale computational resources can be used efficiently to address these questions. Discussions on machine learning continued also in the second discussion, in which interactions between academia and industry was the other important topic. The workshop participants with industrial experience emphasized the importance of detailed project management and, in particular, the need to have very clear agreements about intellectual property rights. Industrial participants to the workshop also pointed out that small companies that develop software solutions can help to bridge the gap between academic research and industry. The workshop also reinforced contacts with scientists from non-academic research centers (T. Trnka [Software for Chemistry & Materials], M.G. Mota [Simune Atomistic Simulations], A. Pan [D.E. Shaw Research], Tim Conrad [Modal AG, Berlin]).

3 Community needs

The workshop demonstrated clearly that in order to make progress the community needs openly available codes that can be easily adapted, extended and combined. Until a few years ago, the field of rare event simulation lacked such a code, but in the mean time this deficit has been closed by the Open Path Sampling (OPS) and the PyRETIS package. It is crucial that these packages, which implement a wide range of path based methods, are interfaced with programs to generate dynamical trajectories (such as Gromacs, Lammps or CP2k) and to

calculate collective variables (such as Plumed). Similarly, an open source software package (n2p2) is now available for the representation of atomistic potential energy surfaces. It is critical for the atomistic and molecular simulation community that development of these codes continues (for instance, within E-CAM Extended Software Development Workshops or similar CECAM activities) and that they are made ready for use on extreme scale computing infrastructure.

The availability of well documented and easily applicable software packages are also a condition for successful collaborations of academic researchers with industry.

4 Funding

The success of the E-CAM Center of Excellence in the development of software tools for rare event simulations in the field of classical molecular dynamics has demonstrated that a collective effort is needed in order to develop high quality scientific software for the community. Such projects cannot be sustained by individual research groups and joint initiatives such as the European Centers of Excellence are needed to provide the funding for software development. The EuroHPC Joint Undertaking, a Europe-wide initiative to create a European HPC infrastructure which is starting at the moment, might provide opportunities in this direction in the future.

5 Will these developments bring societal benefits?

It is unlikely that short term societal benefits follow from the workshop. But in the long term, activities such our E-CAM State-of-the-Art Workshop on "Large Scale Rare Event Simulation" can have an important societal impact in several distinct ways, for instance:

- Many processes of importance in the fields of materials science and drug design are determined by rare events. So a detailed molecular understanding of new materials or drugs relies on our ability to study rare events in large-scale computer simulations. The simulation algorithms discussed in this meeting are an important step in this direction. In fact, several talks and discussions at the meeting touched on topics of direct economic relevance for instance the binding and unbinding of ligands to and from proteins and the catalytic properties of disordered substrates.

- The meeting also had a strong emphasis on the development of scientific software, mainly for transition path sampling and machine learning. Such software projects can have a huge impact on the research field and, in particular, on possible collaborations with industry.

- Early stage researchers working on the scientific software developments discussed in the workshop will have interesting job opportunities and will facilitate the transfer of knowledge from academic research to industry.

6 Participant list

Organizers

Bolhuis, Peter
University of Amsterdam, The Netherlands
Dellago, Christoph
University of Vienna, Austria
Kahl, Gerhard
Vienna University of Technology, Austria

Arjun, A - University of Amsterdam, The Netherlands
Ceriotti, Michele - Swiss Federal Institutes of Technology Lausanne (EPFL), Switzerland
Chandrabose, Selvaraj - Masaryk University, Brno, Czech Republic
Chodera, John - Memorial Sloan Kettering Cancer Center, USA
Conrad, Tim - FU Berlin, Germany
Decherchi, Sergio - Italian Instiute of Technology, Genoa, Italy
Diaz Leines, Grisell - Ruhr-Universität Bochum, Germany
Everaers, Ralf - École Normale Supérieure de Lyon, France
García Mota, Mónica - Simune Atomistic Simulations, Spain
Hantal, Gyorgy - University of Vienna, Austria
Hummer, Gerhard - Max Planck Institute of Biophysics, Germany
Innerbichler, Max - University of Vienna, Austria
Janoš, Pavel - National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno, Czech Republic
Jurásek, Miroslav - Masaryk University, Faculty of Science, Brno, Czech Republic
Keller, Bettina - Free University of Berlin, Germany
Leimkuhler, Ben - University of Edinburgh , United Kingdom
Leitold, Christian - UC Santa Barbara, USA
Locatelli, Emanuele - University of Vienna, Austria
Lupi, Laura - Faculty of Physics, University of Vienna, Austria
Lyu, Wenping - RWTH Aachen, Germany
MacKernan, Donal - University College Dublin, Ireland
Minkowski, Marcin - University of Vienna, Austria
Moritz, Clemens - University of Vienna, Austria
Pan, Albert - D. E. Shaw Research, USA
Pérez de Alba Ortíz, Alberto - University of Amsterdam, The Netherlands
Peter, Christine - University of Konstanz, Germany
Peters, Baron - University of California, Santa Barbara, USA
Pietrucci, Fabio - Sorbonne Université, France
Qin, Lin - Technical University of Vienna, Austria
Rogal, Jutta - Ruhr University Bochum, Germany
Sega, Marcello - Helmholtz-Institut Erlangen-Nürnberg, Germany
Shabbir, Huzaifa - University of Vienna, Austria
Sikora, Mateusz - MPI for Biophysics, Germany
Singraber, Andreas - University of Vienna, Austria
Šmak, Pavel - Masaryk University, Czech Republic
Swenson, David - École Normale Supérieure de Lyon, France
Tiwary, Pratyush - University of Maryland, USA
Trnka, Tomáš - Software for Chemistry & Materials, The Netherlands
Troester, Andreas - University of Vienna, Austria
van Erp, Titus - NTNU:Norwegian University of Science and Technology, Trondheim, Norway
Wolf, Beatrix - ESI, Austria