Computational Physics, Machine Learning, Web Design

Group leader David van der Spoel, Dept. of Cell- and Molecular Biology, UU.
david.vanderspoel@icm.uu.se

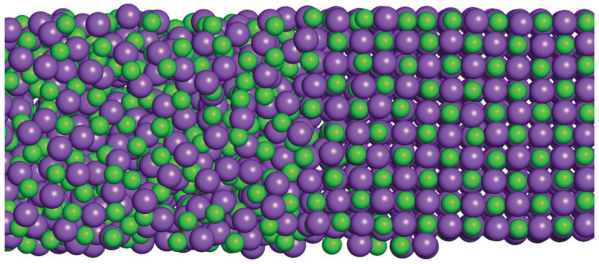

Figure 1 Melting simulation with liquid on the left side and solid on the right side. As can be seen, the solid state is more ordered and therefore has a lower entropy than the liquid state. We can detect melting in simulations by storing the files containing the atomic coordinates as a function of time, and compressing each frame. The higher entropy in the liquid state files makes them harder to compress and therefore we can detect a phase change from checking the file size.

In my research group we design physical models for the simulation of molecular processes. The models are based on heterogeneous data, from both experiments and from high-level quantum chemistry. The aim is to build a model that can be used to study complex physical processes in both the gas- and the condensed phases. As a proof-of-principle, we have developed a model for alkali-halide salts such as NaCl[1] and used it to predict the melting temperature of these susbstances[2] (Figure 1) among other things. In order to arrive at these models a Bayesian Monte-Carlo algorithm was used to perform simulations / optimizations in parameter space with the objective to reproduce known data. The algorithms are conceptually simple, however due to the large parameter space, extensive searching is needed. All of the projects deal with performing simulations and analyzing the results using computational science and mathematics in order to improve models for physics and chemistry.

1. Sensitivity analysis in high-dimensional space (30 hp)

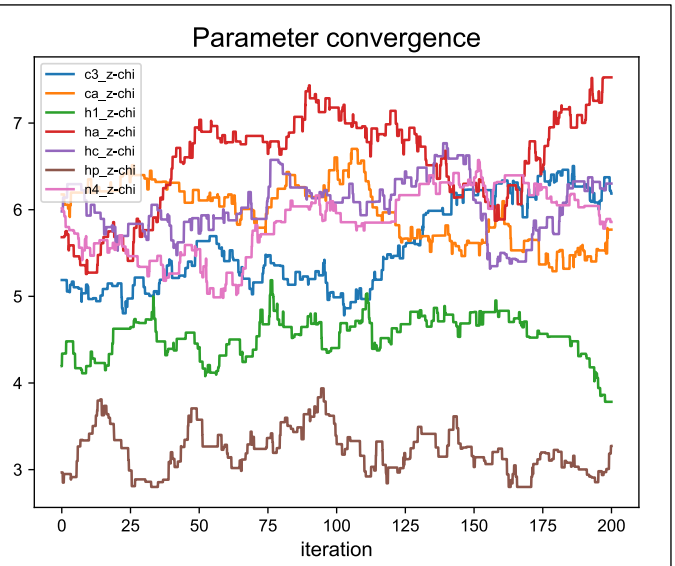

Figure 2 Results from Monte-Carlo simulation in parameter space, each line corresponding to the electronegativity of a specific chemical atom-type.

In this project you will perform machine learning using Monte-Carlo parametrization simulations and analyze the results of the simulations in a high-dimensional parameter space. The simulations are performed many times and typically we aim to converge to reach a certain maximum "allowable" deviation χ^2 from reference data. For each such simulation a trajectory in parameter space is obtained containing many iterations (Figure 2). Each iteration corresponds to a χ^2 value (not plotted). The idea here is to find clusters of parameters that yield low χ^2 values such that we can estimate uncertainties on the parameters. Since the parameters are highly correlated, individual correlations are not sufficient to describe

the total variance in χ^2 as a function of the parameters[3]. In addition to cluster analysis, you will perform sensitivity analyses by systematically altering parameters and checking the χ^2 values, and in this way establish an estimate of the error in the observable of interest. The final aim of this project is to get robust error estimates in the parameters.

Requirements: fluent in Python or R, Linux, English, mathematics and statistics. Good to have as additional knowledge: C++, cmake, basic chemistry and physics, Swedish.

2. Re-design webserver for chemists (15 hp)

The http://virtualchemistry.org website is a popular site for computational chemists where my group provides models and input data for molecular simulations[4]. The website data is supplied by an SQLite database of about 70 Mb which is processed using a combination of Ajax and PHP. Javascript is used for interactive plots and, using a third-party plugin, to visualize molecules. In addition, we have designed another website to provide further inputs for molecular simulations, this one is called "gentop" which stands for "generate molecular topology". It is written in Angular and embedded in a Docker container. The goal of this project is to redesign the Virtual Chemistry server and incorporate the Angular functionality into it. The whole new website should be stored in the docker environment for portability.

Requirements: some experience with Javascript, Angular, PHP, Linux, English, Docker. Good to have as additional knowledge: basic chemistry and physics, Swedish.

3. Implement Langevin Integrator with Polarizable Particles (30 hp)

Molecular simulations are typically performed in a way to mimic experimental conditions. That means we need to control the temperature and pressure in a simulation, in a similar way as we would in a laboratory experiment. Algorithms for this have been developed and implemented in many codes. Here the idea is to implement a combination of polarizable models[1], and the Langevin method for integrating the atomic motions into the OpenMM software (http://openmm.org). This software is very flexible and supports both Python code for proofs of principle, as well as optimized code for CPU and GPU.

Requirements: experience with Python, C++, Cuda, Linux, English. Good to have as additional knowledge: basic chemistry and physics, Swedish, OpenCL.

References

(1) Walz, M.-M.; Ghahremanpour, M. M.; van Maaren, P. J.; van der Spoel, D. Phase-Transferable Force Field for Alkali Halides. *J. Chem. Theory Comput.* **2018**, *14* (11), 5933–5948. https://doi.org/10.1021/acs.jctc.8b00507.

(2) Walz, M.-M.; van der Spoel, D. Systematically Improved Melting Point Prediction: A Detailed Physical Simulation Model Is Required. *Chem. Commun.* **2019**, *55* (80), 12044–12047. https://doi.org/10.1039/C9CC06177K.

(3) Yildirim, A.; Ghahremanpour, M. M.; van der Spoel, D. Propagation of Uncertainty in Physicochemical Data to Force Field Predictions. *Phys Rev Res* **2020**, *2*, 033277.

(4) van der Spoel, D.; van Maaren, P. J.; Caleman, C. GROMACS Molecule & Liquid Database. *Bioinformatics* **2012**, *28* (5), 752–753. https://doi.org/10.1093/bioinformatics/bts020.