**Goal.** In this project we will use AI to go beyond machine learning in chemistry and create a new energy function for molecular simulation. In many present physical models, force fields, the energy for stretching a chemical bond is described by a harmonic potential and the same goes for angular distortions (for instance the H-O-H angle in a water molecule). These are very crude approximations and better models have been proposed long ago\(^1\) but not implemented at a large scale. *In reality we do not know what would be the best mathematical representation of these contributions to the energy function. It would be straightforward to apply supervised machine learning for instance through Bayesian Monte Carlo to fit a predefined function (and we have done similar things for other parts of the Alexandria force field\(^2,3\)). However, since we do not know in advance what mathematical model or combination of models will best fit the energy function, AI tools and indeed machine learning expertise are needed.*

**Background.** Higher accuracy for predictions and well-defined confidence intervals are needed to extend the applicability of force field simulations and this project will contribute significantly to that goal. The importance of computational methods for predicting physicochemical properties such as protein-ligand binding free energies or material properties can not be overstated since experimental measurements can be both complex, expensive or even dangerous. Studies of corrosive materials at very high temperatures, such as molten alkali halides are a good example and we have shown that very accurate force field modeling can be achieved\(^4-6\) through supervised machine learning.\(^2\) *AI can give the breakthrough that is needed to achieve the same level of accuracy for (halo)organic compounds as well.* With that in place the range of possible applications becomes virtually unlimited.

As an example, consider the compounds consisting of just C, H, O elements with general formula \(C_xH_yO_z\). The enthalpy of formation \(\Delta_fH^\ominus\) of such a compound is defined as the enthalpy of the following reaction:

\[
xC(s) + \frac{y}{2}H_2(g) + \frac{z}{2}O_2(g) \leftrightarrow C_xH_yO_z \quad (1)
\]

where the compound is built from its constituting elements in their standard state. This energy can be measured by, e.g., combustion experiments or computed using quantum chemistry methods. Quantum chemistry is needed since the chemical bonds are defined by the electronic structure of the compounds. Interestingly, compounds with the same formula but different atomic- and electronic structure have different \(\Delta_fH^\ominus\), for instance for ethanol (\(\text{C}_2\text{H}_6\text{O}\)) \(\Delta_fH^\ominus = -235\) kJ/mol and for methoxymethane with the same formula \(\Delta_fH^\ominus = -184\) kJ/mol. *The main goal in this project will be to derive an energy function to reproduce such energy differences.*

The research presented in this proposal will contribute to the development of the Alexandria force field. The first complete model, for alkali-halide salts in three different phases, was published in 2018.\(^2\) The model for alkali halides was developed using supervised machine learning and it has been applied in three papers so far\(^4-6\) using the GROMACS software\(^7\) of which DvdS is one of the core developers. In these studies we showed that the error with respect to experiment is a factor of three lower than previous models.

![Figure 1. Energy for a water molecule in many different conformations from quantum chemistry (X-axis) versus the Alexandria force field (Y-axis) after a machine learning optimization of the potential function.](image-url)
Implementation. Here we will use the Alexandria library\textsuperscript{8} of quantum chemistry calculations that we recently published as well as the ANI-1 library.\textsuperscript{9} The Alexandria library is currently being extended with off-equilibrium calculations using density functional theory at a higher level of theory than ANI-1 and featuring more elements. In this manner we will build up a very large database of reference values that will be used for machine learning and artificial intelligence.

More specifically, in this project the CIM graduate student will work on discovering mathematical relations between the atomic coordinates and the energy of the compound. Starting from known functions such as harmonic and Morse potentials\textsuperscript{1} we will explore many functions and combinations of functions beyond the current simplistic division of energy function into bond-, angle- and torsion-potential functions.\textsuperscript{10} Indeed, we will investigate using different mathematical representations of interactions in different parts of compounds, to allow for a more realistic energy function. A preliminary result for one compound is given in Fig. 1 and in this project this will be extended to, in principle, all (halo)organic compounds. Evaluation of results will be done by computing additional properties such as thermochemistry,\textsuperscript{11} through simulated vibrational infrared spectra (Figure 2) and, eventually, by evaluating condensed-phase properties.\textsuperscript{12,13}

![Figure 2. Experimental and theoretical infrared vibrational spectra for quinoline based on two quantum-chemistry methods and four force field calculations. There is quite some discrepancy between theory and experiment that this project will contribute to reducing.](image)

Status. A CIM graduate student will be recruited with a keen interest in both AI, mathematics and chemistry and biology with the CIM contribution in combination with faculty funding. Software is available to work with the Alexandria library\textsuperscript{8} and the ANI-1 database comes equipped with software tools as well,\textsuperscript{9} which implies the graduate student can be productive from day one. Crucially, the existing tools will be extended to allow discovery of physical models for chemicals or materials that can be described by classical potentials, thereby breaking new ground for AI in chemistry.

The combined expertise of main supervisor DvdS (physics, chemistry) and co-supervisor SE (computer science) will make sure the student will receive adequate supervision.

References

\begin{enumerate}
\item Morse, P. M. Diatomic molecules according to the wave mechanics. II. Vibrational levels. Phys. Rev. 1929, 34, 57–64.
\item Walz, M.-M.; van der Spoel, D. Systematically improved melting point prediction: a detailed physical simulation model is required. Chem. Comm. 2019, 55, 12044–12047.
\end{enumerate}