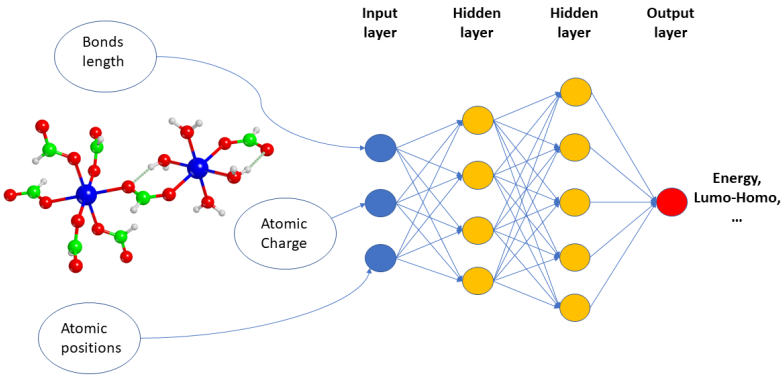
Neural Networks in Quantum Chemistry

Joseph Sleiman

We aim to implement a neural network in Python using the *TensorFlow* machine learning library to predict the properties of given chemicals based on their molecular structure. This structure is encoded using molecular ”descriptors” describing features such as atomic charge and bond length, via the library *DScribe*. Finally, we hope to benchmark the capabilities of both GPUs and CPUs in running parallel neural network applications.



Machine Learning (ML) and Artificial Intelligence (AI) have provided comprehensive solutions to various natural and scientific phenomena, ranging from galaxy detection[1] to genome sequencing,[2] and paired with the boom in data and growing computational power, means the applications for ML and AI will continue to stretch further and wider into even the most established and niche sectors.

Likewise, quantum chemistry - a branch of chemistry focused on predicting chemical and physical properties of molecules and materials[3] - has seen an increase in ML applications, particularly through the use of neural networks. This deep learning architecture consists of groups of "neurons" (a term inspired inspired by the architecture and functionality of the brain - see image) all densely connected to form a network where the input is typically a vector consisting of your data representation and the output is a value representing your prediction. In between these layers, there are hidden layers consisting of an arbitrary number of neurons which are optimised via weights to produce the best predictions possible on well-understood datasets, while also generalising to the new data we wish to make predictions about.

Accordingly, in this project we use neural networks to predict quantum mechanical properties of molecules as opposed to the traditional method of solving the Schrödinger equation. This

is a very promising avenue considering that it gets increasingly difficult and computationally expensive to solve the Schrödinger equation for molecules more complex than Helium (2 atoms). Hence, the numerical optimisation approach inherent in ML could result in accurate predictions without the need for large computing power and a lot of time!

More specifically, we will represent molecules using descriptors via the Python library *DScribe*[4] to take their atomic positions as input and output property predictions such as energies, the charges on atoms, and evidence of hydrogen bonds.

Final Product

An efficient implementation of a neural network using *TensorFlow* that can predict chemical properties of given molecules to a high degree of accuracy, without the need for exponential computing time at greater molecular complexity (traditional drawback of the Schrödinger equation). Parallel code to be bench-marked and compared to GPU implementation.

Final Product Presentation

We will create a slide presentation outlining our project and have a voice-over commentary explaining the finer details.

Project Timeline

Week 3 (5 July – 25 July)

Introduction to neural networks, *TensorFlow*, and the remote server we will be running our code on. Aim to run a basic ML experiment remotely.

Week 4: (26 July – 1 August)

Introduction to molecular descriptors, *DScribe* and how to encode the relevant information necessary to represent a molecule. Complete some basic *DScribe* tutorials.

Week 5 (2 August – 08 August)

Continue to learn more advanced techniques using *DScribe* and start looking at ways to efficiently implement neural network algorithms.

Week 6 (08 August – 15 August)

Implementation and optimisation of neural network for our quantum chemistry problem as well as a deep look into the dataset specific to our project.

Week 7 (16 August – 22 August)

Extensive testing and bench-marking of our code. Aim to finalise project and compare relative speed-ups using GPUs and CPUs.

Week 8 (23 August – 31 August)

Writing up the project in our LaTeX report, as well as completing the presentation/video summarising the last 8 weeks.

PRACE SoHPC Project Title

Neural Networks in Quantum Chemistry

PRACE SoHPC Site

Slovak Academy of Sciences, Slovakia

PRACE SoHPC Authors

Joseph Sleiman , University of Edinburgh, UK

Joseph Sleiman

PRACE SoHPC Mentor

Marián Gall, Slovakia

PRACE SoHPC Contact

Joseph Sleiman, University of Edinburgh
E-mail: josephsleiman16@gmail.com

PRACE SoHPC Software applied

TensorFlow, Python, DScribe, MPI

PRACE SoHPC Acknowledgement

Special thanks to Marián Gall for his supervision during this project and the team at ICHEC for their insightful training week programme.

PRACE SoHPC Project ID

2107

PRACE SoHPC References

[1] R. González, R. Muñoz, and C. Hernández, "Galaxy detection and identification using deep learning and data augmentation," *Astronomy and Computing*, vol. 25, pp. 103–109, 2018.

[2] R. Poplin, P.-C. Chang, D. Alexander, S. Schwartz, T. Colthurst, A. Ku, D. Newburger, J. Dijamco, N. Nguyen, P. T. Afshar, S. S. Gross, L. Dorfman, C. Y. McLean, and M. A. DePristo, "Creating a universal snp and small indel variant caller with deep neural networks," *bioRxiv*, 2018.

[3] T. A. Profitt and J. K. Pearson, "A shared-weight neural network architecture for predicting molecular properties," *Phys. Chem. Chem. Phys.*, vol. 21, pp. 26175–26183, 2019.

[4] L. Himanen, M. O. Jäger, E. V. Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke, and A. S. Foster, "Dscribe: Library of descriptors for machine learning in materials science," *Computer Physics Communications*, vol. 247, p. 106949, 2020.