Deep learning - Project synopsis
Atomic structure generation with recurrent neural networks

Bjarke Arnskjær Hastrup and Michael Wamberg
{s151602, s146791}@student.dtu.dk

October 2019

Motivation and background **:**
Deep generative models have a great potential in molecular science, as they can reduce the usually very high computational costs associated with traditional quantum mechanical calculations. In this project we will focus on the chemical element silicon. What makes silicon particularly interesting to study is its wide use as a semiconductor and the fact that silicon anodes are generally considered to be the next development within lithium-ion battery technologies. The goal of the project is to take around 500.000 silicon structures and process these structures into training data for a Long short-term memory (LSTM) network. The ambition is for the network to learn the patterns in physically viable atomic structures and make the probabilistic generation of amorphous silicon structures.

Milestones **:**

- Project week 1: *Install and learn about The Atomic Simulation Environment (ASE) and setup training data for our network. Study relevant research articles.*
- Project week 2: *Consider and study the possible network structures.*
- Project week 3: *Build the variational autoencoder and generate sequences of atom coordinates.*
- Project week 4: *Extend the network with the LSTM-based encoder.*
- Project week 5: *Training, testing, debugging, fintuning and optimizing of the network.*

References **:**

Data-Driven Approach to Encoding and Decoding 3-D Crystal Structures, Hoffmann, Jordan and Maestrati, Louis and Sawada, Yoshihide and Tang, Jian and Sellier, Jean and Bengio, Yoshua, 2019.

Symmetry-adapted generation of 3d point sets for the targeted discovery of molecules, Niklas W. A. Gebauer and Michael Gastegger and Kristof T. Schütt, 2019.

Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning, Frank Noé and Simon Olsson and Jonas Köhler and Hao Wu, 2018.

Deep Latent Variable Models for Sequential Data, Marco Fraccaro, 2018.