Instructions for completing the reproducible results

YC Wong

1 Files given in the GitHub repo

1. Pre-trained machine learning models (2x):
   (a) MACE-opt: **STO-MACE-opt.model**
   (b) MACE-ph: **STO-phonopy.model**
   (c) MACE-MP: See the installation section below
2. Dataset for evaluating the performance of models:
   (a) Test dataset used for verifying numerical result and performing uncertainty quantification in section 3: **final_test_set.xyz**
   (b) Molecular dynamics dataset used for evaluating the performance of the model between the MACE-opt and MACE-MP models in section 4: **STO_MD_dataset.xyz**

2 Setup and Software Installations

1. Login to the SCRTP theory bethe node for GPU usage:
   `ssh username@bethe.theory.warwick.ac.uk`
2. First change to your desired working directory, and do
   `git clone https://github.com/ycwong-chess/UQ_peer_ex`
3. Installing mace through the below commands (upgrade pip first if installation not working)
   `git clone https://github.com/ACEsuit/mace.git`
   `pip install ./mace`
4. Now open the jupyter notebook which will provide the basic instructions. Feel free to work directly on the jupyter notebook or make a copy of it.

3 Task 1: Reproducing the exact numerical result for the three models

In this part of the exercise, you should be able to calculate the force MAE and RMSE of the 3 models: **MACE-opt**, **MACE-ph** (phonopy model) **MACE-MP** model. The force MAE and RMSE are given as below (all in unit eV/Å):

	MACE-opt	Phonopy	MACE-MP
MAE	0.0136	0.284	0.0457
RMSE	0.0284	0.574	0.0953

4 Task 2: Compare the uncertainties of the model against DFT using molecular dynamics Data

In this part of the exercise, you are given a MD dataset that contains 50 configurations that have been evaluated using DFT calculations. you are expected to split the MD dataset into five subsets with roughly an equal number of configurations in each set (each set should contain $10 \times 625 \times 3 = 18750$ force data points), and try to evaluate the mean and standard deviation of the MAE and RMSE respectively using the MACE-opt and MACE-MP model. This should give a relatively good estimate of the performance for the dataset based on the mean and the spread, and you should be able to conclude which model is performing better.