Machine Learning for the Hydrogen Potential

Wladyslaw Zawadzki
Supervisor: Prof Andreas Hermann

June - August 2025

Lay Summary
In this project, we attempt to train a machine learned inter-atomic potential for hydrogen that will avoid predicting non-physical scenarios, such as overlapping atoms, which are currently prevalent in many trained neural network potentials due to a lack of low-separation data. To achieve this, we build a short-range interaction-simulator based on computational hydrogen energy data, which we use to modify the training data — by removing interactions at low separations — before feeding it into the machine learning algorithm. Ideally, the resulting neural network potential would ignore all short-range interactions and only reconstruct the atomic behaviour at longer ranges, such that the simulator script could be run in reverse at deployment to recombine the neural network data with the subtracted energy and forces.

Abstract
In this project, we attempt to overcome the issues experienced by traditional neural network potentials in predicting the behaviour of atoms at small separations by subtracting the short-range interactions from the dataset before the training procedure. The short-range interactions were simulated through an 'extended Lennard-Jones-style" pair-potential fitted to density-functional theorem data, used to calculate the energy and forces due to every atom within a specified cut-off radius of another. After modifying the training data, n2p2 was used to train the neural network potential, but the desired results were not observed. The underlying issue has not yet been identified, however, there are inconsistencies present in the validation output. We hope that further work will uncover the root of the problem and produce a simulator that effectively describes both short-range and long-range interactions in systems extending beyond molecular hydrogen.

1 Introduction

Despite advancements in atomic-scale modelling and simulation of materials, further progress is limited by the computational cost of explicit electronic-structure methods, such as density-functional theory (DFT)[1]. To address this problem, machine learning has developed as an exciting solution. Rather than relying on expensive quantum-mechanical calculations, machine learning extrapolates information on unfamiliar systems by seeking patterns in available data. A major application of machine learning in materials science has been in producing Neural Network Potentials (NNPs), which are used to learn how the atomic structure maps out to the potential energy surface (PES)[2].

However, this task isn't always straightforward. With insufficient data, the NNP may start to hallucinate during extrapolation. Such is the case for molecular hydrogen, which naturally does not form structures with very small (< 0.5 Å) atomic separations. Ultimately, the Neural Network does not learn that placing every hydrogen atom in the same location is incorrect.

In this project, we attempt to overcome this issue by isolating the short-range from the long-range interactions. Given DFT data about a many-atom hydrogen structure, forces within a certain cut-off radius of each atom must be subtracted from the system to obtain a modified training data set, such that the NNP learns only about forces on the intermolecular scale. The idea is that a given atoms nearest neighbours will be seen as having no effect on the total force, thus allowing the short-range forces to be added back into the input data when deploying the NNP on (e.g.) a molecular dynamics simulation.

2 Methods and Outcomes

2.1 Overview

The short-range interactions were all treated as coming from a single pair-potential, calculated for each atom found within a specified cut-off radius of another atom. A simulation would then be run to obtain the total energy of the system and the force vector on each atom. These simulated "short-range" values were then subtracted from the original DFT data input to the

program, to generate a set of modified training data. Ideally, this modified data would be used to train an NNP with the n2p2 software package[3], where it would learn to disregard the short-range while still accurately predicting the long-range interactions. Finally, this NNP would be deployed with the LAMMPS simulation tool[4] to observe how the structure evolves through time. At each time interval, the forces and total energy of the system would be manually adjusted by running the original data modification script in reverse, fully reintegrating the short-range interactions to the system.

2.2 Fitting the Pair-Potential

The first task was to fit a pair-potential given data obtained from the CASTEP DFT calculator[5]. The program was run on a single hydrogen molecule with varying atomic separation in a $10 \times 10 \times 10$ Å lattice cell obeying periodic boundary conditions (PBC, see fig. 1 for details on how this value was chosen), which gave an associated total energy value. This was done for about 300 data points, mainly focused around the potential minimum, which was found to be near 0.753 Å. Data was taken for separations in the range of $0.4 - 2.0$ Å, both to avoid unnaturally small separations and to extend far enough out to cover the next nearest-neighbour.

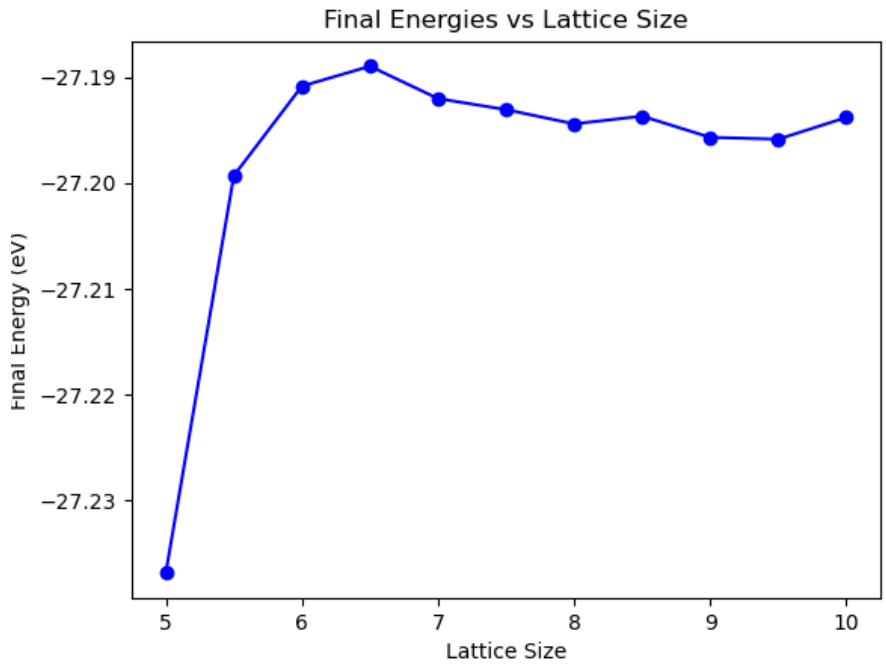


Figure 1: CASTEP was used to find the energy of a single hydrogen molecule with constant bond length inside a cubic lattice cell of varying side length (in Angstrom). 10 Å was chosen for fitting the pair-potential as it was well beyond the energy plateau at $\approx$ 8 Å, signifying that the molecule was outside the influence of "image" atoms in neighbouring cells; a natural consideration when working with PBC.

Taking energy as a function of the separation vector, several pair-potential fits were tried before finding one that balanced accuracy with the number of parameters being used. All curve-fitting work was done via SciPy in Python. Initially, these were attempted with models taken from literature. The first such fit used the extended Morse potential[6], but the obtained curve varied far too abruptly from the data and was not pursued.

After the shortcomings of the Morse-potential fit, we decided that some form of Extended Lennard-Jones potential must be used, as more parameters were clearly required to capture the unique shape of the data. The next fit was inspired by Hajigeorgiou et al.'s paper[7], but the exact expression could not be replicated due to an implicit discontinuity in their method (instead, the general form of the expression was taken). This expression was used to find a 7-parameter fit which closely resembled the data, but was not used for two reasons; the residuals were too large and too correlated, and the gradient of the expression would be too computationally expensive to evaluate for the eventual force expression. The final parametrised expression for the "ELJ Hajigeorgiou" pair-potential fit is given by eqn. 1, where x is the separation and the parameters are represented by capitalised letters in alphabetical order, and visualised in fig. 2.

$$f_{ELJ}(x) = A(1 - r^{n_r})^2 + B, \tag{1}$$

where

$$\begin{aligned} n_r(x) &= C + DJ + EJ^2 \\ J(x) &= (r - G)/(rz^F + G) \\ F &= \text{int}(2F) \\ z(x) &= (r - G)/(r + G) \\ r(x) &= G/x. \end{aligned}$$

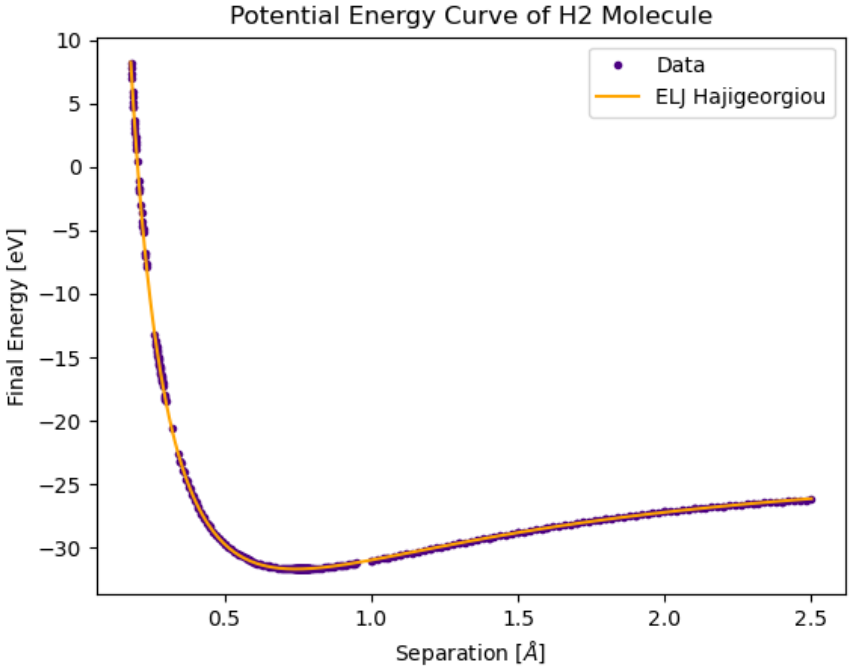


Figure 2: The 7-parameter pair-potential fit (orange) is compared to the CASTEP data (indigo). Although the data was matched quite well, there were significant trends in the residuals. Note that this graph contains separation data below 0.4 Å and above 2.0 Å, but this was disregarded by the time a useable pair-potential was discovered.

An expression derived from Wang et al.'s paper[8], the "Wang potential", was considered due to its elegant form, which would lead to simplified differentiation when solving for force, and due to its built in cut-off behaviour, which would be required at a later stage. This pair-potential was extended by considering a linear combination of terms as an inverse power series, where the parameters μ and ν were held constant at varying fractions $1/n$, σ was set to ≈ 0.38 Å, and only the coefficients of each term were left as free parameters. At 14 total parameters (including an extra energy offset term — added manually), the resultant curve can be seen in fig. 3. Although a poor fit, significant improvements were observed as the number of terms was increased.

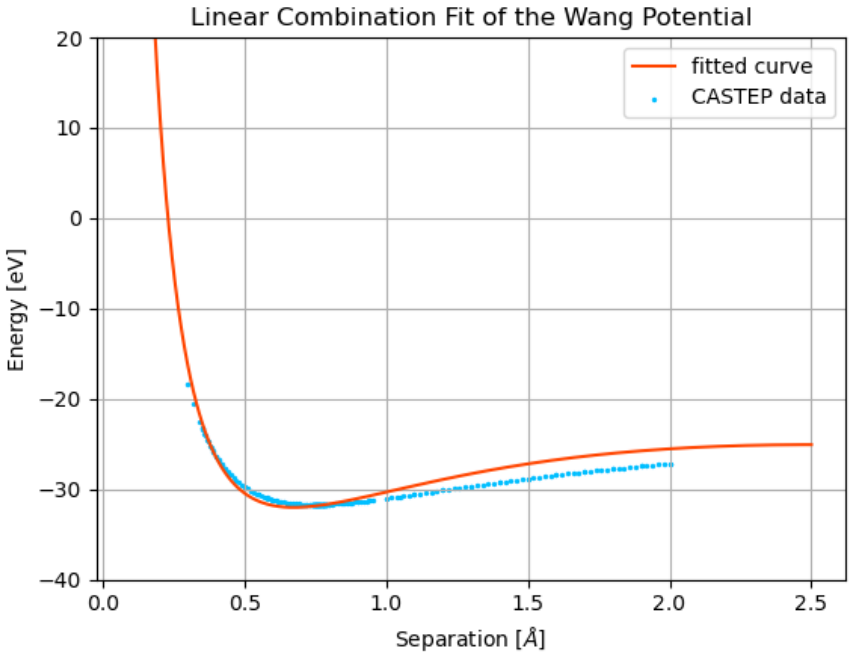


Figure 3: The 14-parameter Wang potential (red) is compared to the CASTEP data (light-blue). With each new term added, the fitted curve grew wider, gradually improving the fit. A manual y-axis offset of $-25.068eV$ was added to the expression, corresponding to the energy of two isolated hydrogen atoms, which levelled the curve with the data.

Following on from the Wang potential, a different approach was found that reduced the number of parameters while greatly improving the fit. The final, accepted expression was a 12-parameter inverse power series expansion (ELJ12), given by eqn. 2. The fitted parameters and their calculated numerical values are displayed in table 1. In this fit, the 11^{th} power was omitted for efficiency, as it was observed during testing to make the least notable improvements to the overall fit.

$$A + \frac{B}{x} + \frac{C}{x^2} + \frac{D}{x^3} + \frac{E}{x^4} + \frac{F}{x^5} + \frac{G}{x^6} + \frac{H}{x^7} + \frac{I}{x^8} + \frac{J}{x^9} + \frac{K}{x^{10}} + \frac{L}{x^{12}} \tag{2}$$

Initially, ELJ12 was fitted with a polynomial cut-off function (eqn. 3) taken from the n2p2 documentation[3], which decayed smoothly between 2.0 Å and the cut-off radius, $r_c = 2.5$ Å, beyond which the energy was set to a constant value of $E_c = -25.068eV$ — the total energy of two entirely isolated hydrogen atoms. This was done to avoid (later) calculations on atoms

Parameter	Value
A	-5.65429887e+01
B	3.17871633e+02
C	-1.36366521e+03
D	3.28161439e+03
E	-5.10301803e+03
F	5.38631149e+03
G	-3.91966493e+03
H	1.95568837e+03
I	-6.48949185e+02
J	1.32615401e+02
K	-1.33895098e+01
L	8.60247231e-02

Table 1: Numerical values of parameters used for ELJ12.

separated by more than 2.5 Å, which would provide negligible contributions to the total energy/force but would come at a significant computational cost. Once the full (cut-off included) energy expression was obtained, the negative gradient was computed analytically to find an expression for the force on each atom. The force expression cut-off was achieved by using the product rule on eqns. 2 and 3 in the cut-off region, which massively simplified the derivative. The energy/force curves for the pair-potential with $r_c = 2.5$ Å are shown in fig. 4.

$$f_{cut-off}(x) = ((15 - 6x)x - 10)x^3 + 1 \tag{3}$$

Before the final form of ELJ12 was settled on, an alternative cut-off procedure was attempted, in an effort to overcome the unnatural "dip" in the force curve. The same cut-off function (eqn. 3) was applied, this time to the force expression rather than the energy, and the energy cut-off was found through integration. The resultant pair-potential, shown in fig. 5, was not used due to the discontinuity in energy, which was deemed less acceptable than the dip in force (fig. 4). The cut-off function applied to the energy expression, then the derivative taken to find the force expression was chosen as the preferred method for the ELJ12 pair-potential.

Ultimately, as will be discussed further in section 2.3, the cut-off had to be changed from $2.0 - 2.5$ Å to $1.0 - 1.5$ Å. The final form of the ELJ12 pair-potential and its residuals are displayed in figs. 6 and 7, respectively. Although the differences between the final form of ELJ12 and the data are drastic, the intended outcome was for the NNP to pick up on this pattern, so long as the data modification procedure stays consistent.

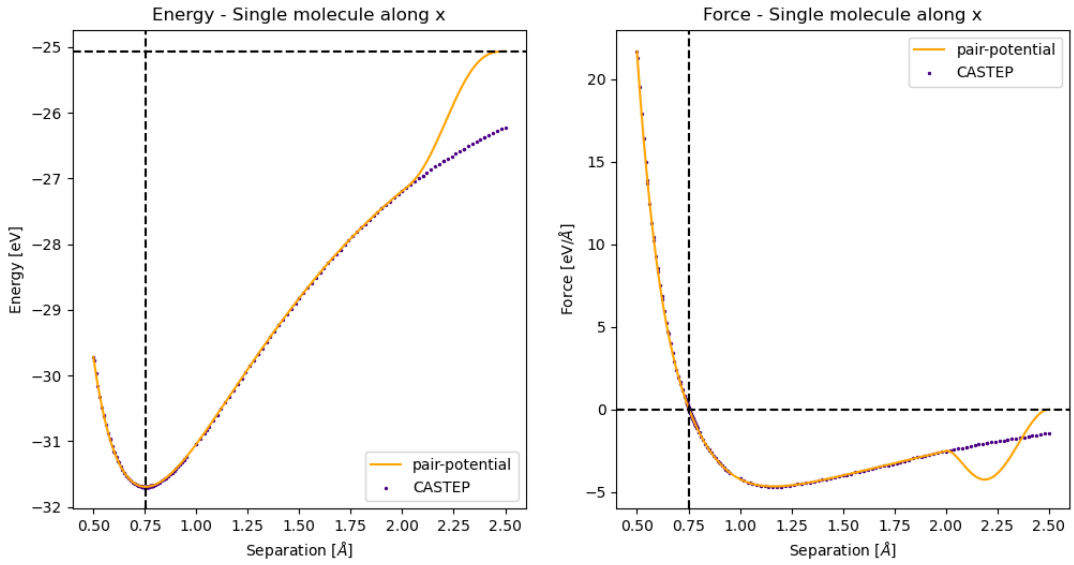


Figure 4: The ELJ12 pair-potential (orange) is compared to the CASTEP data (indigo), for both energy (left) and force (right). The horizontal black dotted line corresponds to E_c for the energy and 0 for the force, whilst the vertical black dotted line corresponds to the equilibrium separation of two hydrogen atoms, ≈ 0.753 Å. The cutoff behaviour can be observed in both graphs in the $2.0-2.5$ Å range. There is a significant dip for force in the cut-off region resulting from the strict physical requirement that the force must correspond to the negative gradient of the potential. Note that the force graph gives the component of force along the x-axis.

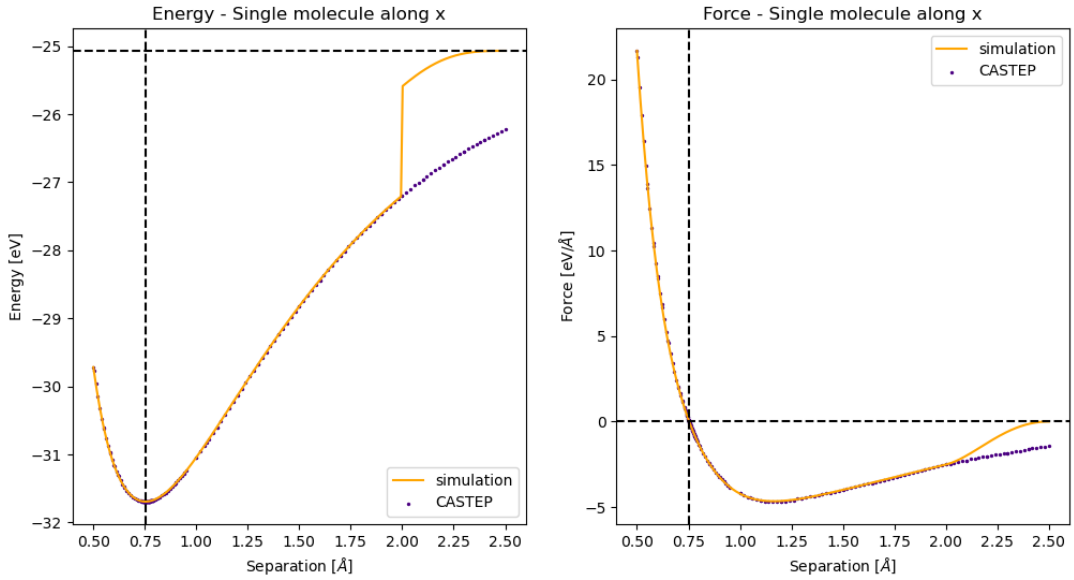


Figure 5: The ELJ12 pair-potential (orange) is compared to the CASTEP data (indigo), for both energy (left) and force (right). In this fit, the cut-off function has been applied to the force first, then integrated to obtain the full expression for the energy. Although the force curve seems better, the discontinuous energy curve could have led to more issues when training the NNP. The black dotted lines represent the same values of interest as in fig. 4.

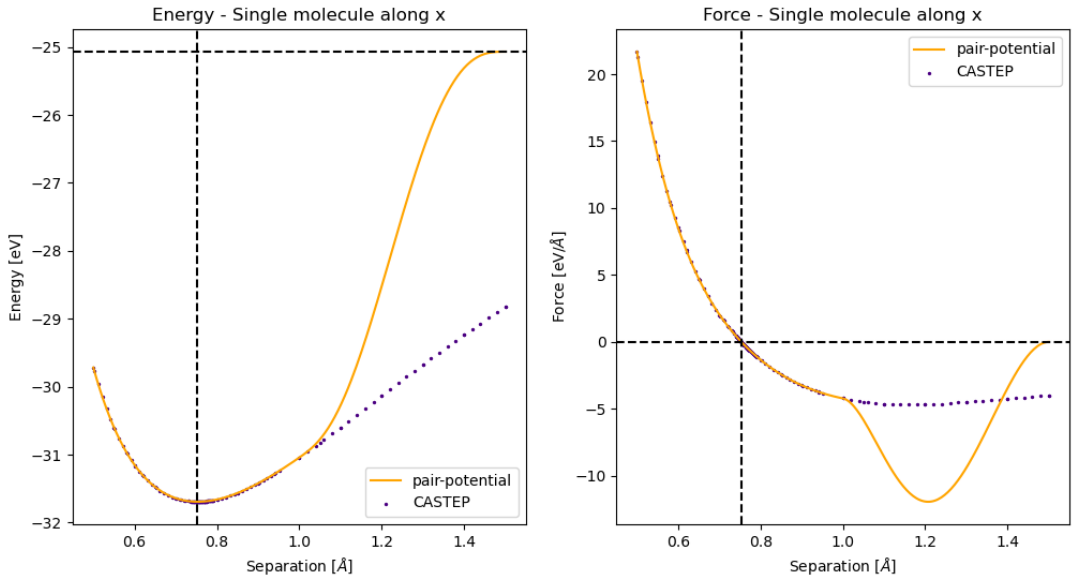


Figure 6: The final ELJ12 pair-potential (orange) is compared to the CASTEP data (indigo) in the $0.4 - 1.5$ Å range, for both energy (left) and force (right). We can observe that, in this reduced range, the force dip has become far more prominent. The black dotted lines represent the same values of interest as in fig. 4.

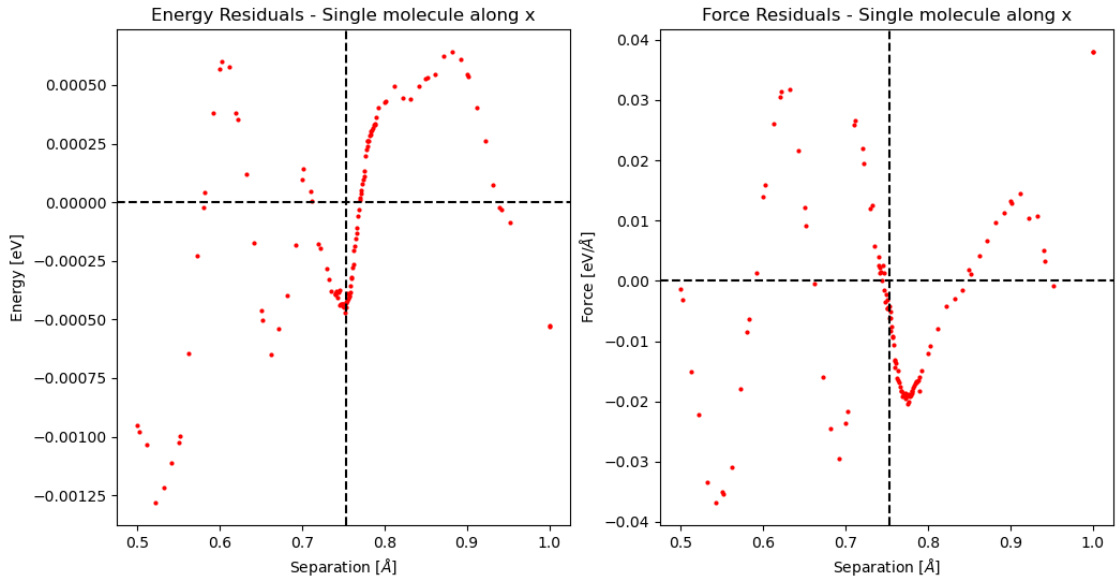


Figure 7: Graphs showing the energy residuals (left) and the force residuals (right) for the ELJ12 pair-potential. Only data up to the cut-off region is included. Although there are some emergent oscillations in the residuals, their magnitude is relatively small and there is no clear pattern. Note that the force residuals are larger than the energy residuals, most likely because the pair-potential was fitted to the energy data rather than the force data.

2.3 Modifying the Data

The data modification procedure relied on using a simulation which employed the newly fitted ELJ12 pair-potential to obtain the energy and forces due to all "short-range interactions" in a hydrogen structure, which would later be subtracted from the input data. During implementation, short-range interactions were considered as the resultant energy and forces from extending the pair-potential to every neighbouring atom within the cut-off radius of another. In reality, the full short-range behaviour cannot be captured by a pair-potential alone — quantum effects play a significant role once many atoms come into close proximity. However, all input data was taken in temperature/pressure conditions where such configurations were improbable, and hence the pair-potential could largely be used to reproduce the present energies and forces, and the NNP would learn that atoms within the cut-off radius do not impact the overall PES. All code used to build the simulation is original and available on GitHub*.

Before continuing, we note that the energy expression was modified to work in terms of relative energy differences, rather than the absolute value of energy (which included the isolated atomic energy), such that the energy expression returns $\approx -6.5eV$ at the potential minimum and $0eV$ beyond the cut-off radius. This was done so that the system is left with at least N times the isolated atomic energy after the short-range energy contributions have been subtracted, where N is the number of atoms in the structure, and was achieved by subtracting E_c from the energy expression. The force calculation was left unchanged.

```
def simulation(home, new_home):
    """Run code on a directory containing unmodified .data files and store the new files in a separate folder"""
    if not Path(home).exists():
        print("ERROR: Provided home path does not exist.")
        exit(1)
    for data_file in Path(home).rglob("*.data"):
            mols = M_Trajectory(data_file)
            mols.compute()
            mols.modify_data(new_home)
```

Figure 8: The simulation function taken from the original project's `M_Trajectory.py` code, used to run the pair-potential calculations on a set of hydrogen structures and modify the training data. The function operates as follows: `home` is a directory containing many unmodified .data files, each containing information pertaining to a single hydrogen structure (the individual structures may be split into separate files by using `separate data.py`). `new_home` is the target location of the program's output, i.e., individual modified data files. The function looks for every ".data" file in the `home` input directory, computes the short-range interactions (through `.compute()`), and modifies the data (through `.modify_data(new_home)`), which automatically places the output in the specified target directory.

A single class was used to run the simulation, which took a single .data file (made from processed CASTEP output data) as input. From this file, the program reads in the position data of every atom and the cell lattice vectors, which would be used in a script to generate the 26 adjacent "image cells", each containing another set of N "image atoms". These image cells were a natural consequence of using periodic boundary conditions (PBCs), which CASTEP already does, and must be considered to correctly quantify interactions between pairs of atoms

*https://github.com/Wladek-Z/Machine-Learning-for-Inter-atomic-Potentials

straddling a cell boundary. To compute the "short-range" energy of the system and the forces on the individual atoms, the simulation would iterate over every atom in the central cell and use the pair-potential to calculate the energy/force contribution due to each atom within the cut-off radius (r_c = 1.5 Å). Once these values were obtained, a modified copy of the input .data file could be made automatically, where the energy and force values would be altered by subtracting the results of the simulation. A snippet of code containing the simulation/data modification function, and how it works is displayed in fig. 8.

During the testing stages of the simulation, we made the decision to shift the pair-potential's cut-off region from 2.0 − 2.5 Å to 1.0 − 1.5 Å, as mentioned in section 2.2. This change came after visualising a comparison between the pair-potential and CASTEP (DFT) forces, displayed in fig. 9. First, data was collected from a modified version of the simulation, where, instead of calculating the energy/force contributions due to each neighbouring atom in a 2.5 Å radius, energy/force calculations were calculated due to the nearest-neighbour only. The expected result would be that the pair-potential forces closely resemble the DFT forces — which they did. The issue only arose when viewing the comparison of the full simulation, i.e. including all interactions within 2.5 Å. In fig. 9, the pair-potential forces can be seen as not only too large (likely owing to the big dip in the force cut-off region), but also seemingly entirely random. Any NNP trained using this data would learn nothing helpful.

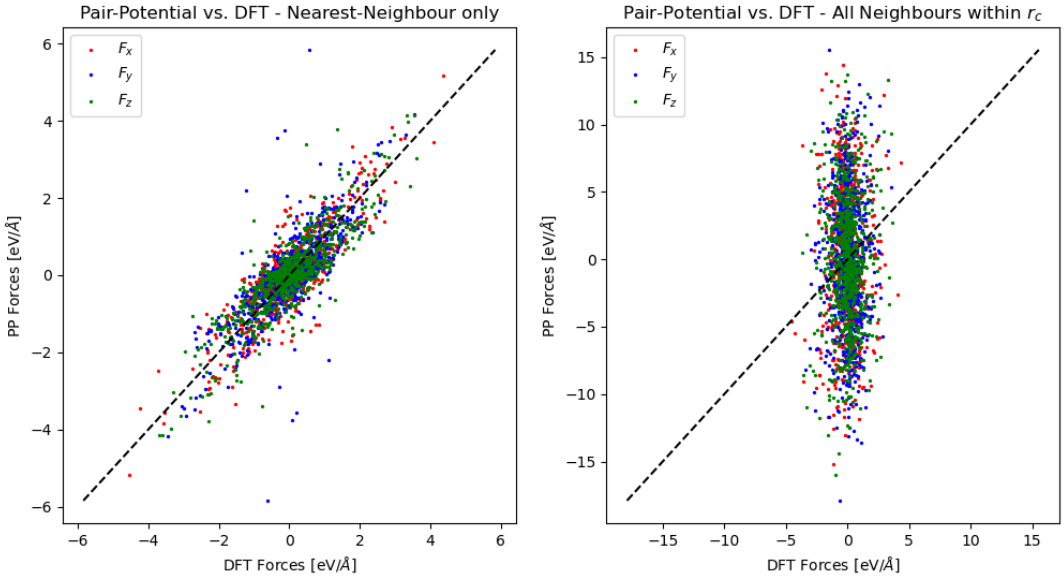


Figure 9: Pair-potential forces compared to the DFT output for a simulation involving nearest-neighbours only (left) and a simulation involving all interactions within the cut-off radius (2.5 Å. The x, y, and z force components have been visually segregated, but there is no reason why any component should show different behaviour than the rest. The graph on the right shows that the simulation seems to be predicting entirely random (and overestimated) forces compared to DFT data. All data was taken from a single structure involving 768 atoms at $T = 900K$ and $P = 20GPa$, with the DFT forces taken from a CASTEP output file.

When r_c = 2.5 Å, configurations of multiple atoms within the cut-off radius are commonplace and, as mentioned previously, the pair-potential alone does not accurately model the full short-range behaviour for groups of atoms in close proximity. Hence, using the 2.5 Å cut-off introduces

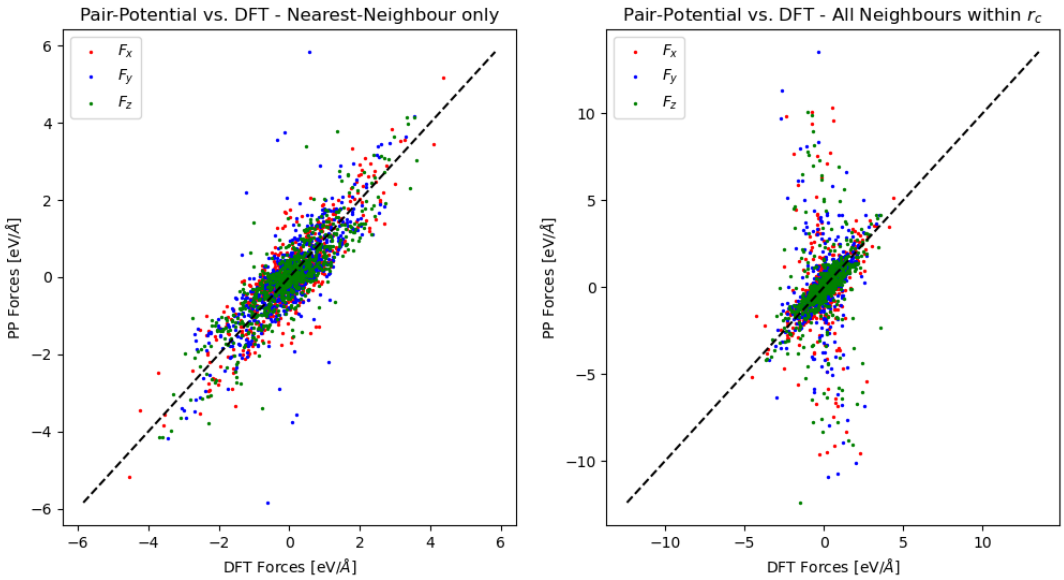


Figure 10: Pair-potential forces compared to the DFT output for a simulation involving nearest-neighbours only (left) and a simulation involving all interactions within the cut-off radius (1.5 Å. Using the new cut-off radius made the force estimations far more concentrated along the main diagonal, which implied that the short-range interactions were being captured more accurately.

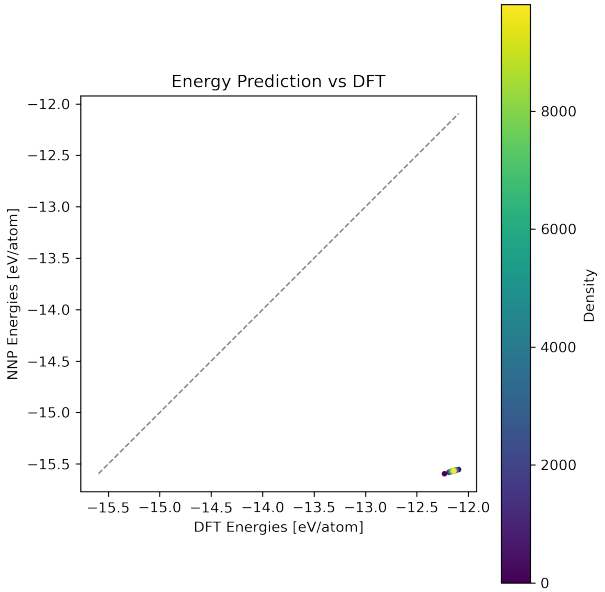
large errors in the results. A model was needed that would more accurately capture the short-range behaviour without limiting the simulation to nearest-neighbours only — the method should be extended eventually to cases where the nearest-neighbour is arbitrary (e.g. under high density). The cut-off range was reduced to $1.0 - 1.5$ Å, and the new pair-potential vs. DFT force comparison graphs can be seen in fig. 10. Although the forces dip still caused the simulation to overestimate the CASTEP results, multiple atoms were far less likely to appear within the new cut-off radius of each other, which led to significantly less overall randomness. This version of the simulation was finally used to modify the training data, in preparation for training the NNP.

2.4 Training the Neural Network Potential

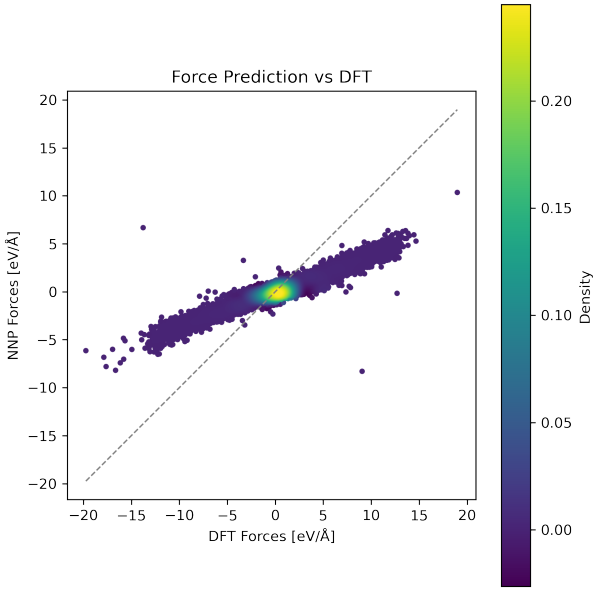
The NNP training mainly follows a predefined procedure utilising existing code, which will not be covered by this report, and some basic steps from n2p2[3]. Out of the available hydrogen trajectory data, we randomly selected 703 (out of a total 3389) structures at $T = 900K$ and $P = 20GPa$, each containing 768 hydrogen atoms. These conditions were chosen to maintain well-defined molecules while also allowing for configurations of multiple atoms in close proximity. Out of these 768, 100 were chosen as the training data set, and a different (entirely non-overlapping) 100 were chosen as the validation set. Out of the training data set, 1 structure would be used to filter the symmetry functions.

The symmetry functions were generated using code available on the GitHub, where the short-range/long-range interaction border was set to 1.5 Å. After running the initial filtering procedure, the 95% correlation filter was applied to retain 105 symmetry functions, which were transferred into a new input.nn file and used for the remainder of this task. The data was then

properly scaled (using 500 bins), and the training was completed. "weights.001.000018.out" was copied into "weights.001.data" as it resulted in the minimum energy RMSE (according to the learning curve). The results of validation, which visualised how the newly trained NNP performed against the unseen data set, can be seen in fig. 11.



(a) The energy scatter graph, showing how the predicted energy per atom compares against the energy per atom in the unseen data. The NNP seems to be underestimating all energies, which should be around $-12eV$, but instead come out around $-15.5eV$. The range of unseen energies is about 0.25 Å, but the range of predicted values appears to be smaller.



(b) The force scatter graph, showing how the predicted force components (decomposed) compare against the force components in the unseen data. Although the predicted values are clearly incorrect, there appears to be a very high degree of correlation with the data, with relatively low noise. Similarly to the energy, the NNP seems to be consistently underestimating all forces.

Figure 11: The scatter graphs output by the validation procedure, showing how energy/forces predicted by the NNP compare against the unseen data set. The unseen data was the set of 100 modified structures not used for training, prepared earlier.

3 Discussion

3.1 Pair-potential Simulation

Although the intended reasoning behind removing the short-range interactions from the system was for nearby atoms to appear as having no impact on each other, the introduction of the cut-off function resulted in a large dip in the force curve, as seen in fig. 6, which had inadvertently caused nearby atoms to have the greatest effect on the total force. Admittedly, these force contributions would be seen by the NNP as repulsive — subtracting large attractive forces would have this effect on the modified data — so the NNP may still prevent atomic positions from overlapping when deployed, as intended. However, it is unclear how the NNP interprets and deals with these forces — whether the final issues have resulted from elsewhere.

3.2 Neural Network Validation

It is clear from fig. 11 that the NNP was unable to predict the correct energy/forces from unseen data. Contrary to this is the output learning curve, which would imply that the training completed successfully (fig. 12). This graph shows how the root mean square error on values of energy/forces evolves with each iteration of the training loop. We can see that the NNP thinks it is within $\approx 1eV/$ Å of the training data forces — and it is at least $100\times$ closer to the training data energies. Moreover, the NNP had stabilised in only a few epochs, as evident by the convergence of the training curves, so there is no question of an incomplete training procedure.

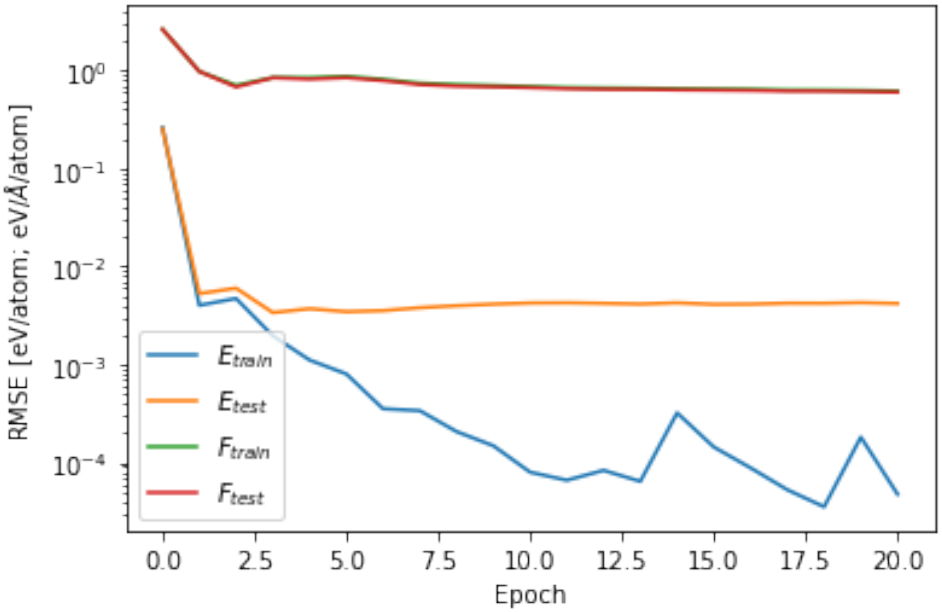


Figure 12: The RMSE (root mean square error) learning curve output from the training procedure. Both predicted forces and energy had stabilised fairly quickly, at reasonable values considering the spread of modified data. Judging by this graph, the predicted energy values should always match the data, and the forces should at least be centred on the data. We can see that the lowest energy RMSE occurs at epoch 18, hence why it was chosen for validation.

How does this reconcile with the results in fig. 11? Further clues may be found other files output by validation, which show that the energy is consistently predicted to be around $\approx -12 \times 10^{-4} eV$, rather than $\approx -9.3 \times 10^{-3}$ as provided by the modified data and corresponding to slightly higher (less negative) than the number of atoms times the isolated atomic energy. This is also why the energy per atom is comes out around $\approx -15.5eV$ instead of $\approx -12eV$. Interestingly enough, $\approx -12 \times 10^{-4} eV$ is actually the value of energy in the *unmodified* data — but the NNP does not have access to this! Furthermore, the predicted forces are consistently underestimated — not dissimilar to how the short-range simulation consistently *overestimated* the unmodified data energy! By all accounts, it would appear that the NNP is completely ignoring the modified data provided and trying to guess at what the true data should look like, but we provide no basis for this assumption. Further work on this project should investigate the inconsistency of the validation results with the learning curve — if there is a minor error causing the NNP to disregard provided data, or perhaps if there is some deeper issues with the theory behind our overall approach.

4 Conclusions

The project began with fitting an "extended Lennard-Jones-style" pair-potential to CASTEP output energies for configurations of a single hydrogen molecule with varying bond length. Although the "ELJ12" inverse power-expansion expression was satisfactory, the number of parameters needing to be fit was likely arbitrary, and residuals might have been improved by using some form of the Wang potential with $>> 14$ parameters. Nevertheless, it was not the pair-potential at fault for the issues with the final outcome.

Following on, short-range interactions were simulated by approximation through the pair-potential, where each nearby atom influenced other nearby atoms according to the pair-potential and nothing else. It is unclear whether this method resulted in the neural network potential failing to produce the desired outcomes due to the inconsistencies observed between learning and validation. Any further development should analyse the training output for evidence of deeper issues before reconsidering the method by which short-range interactions were removed. Should the NNP eventually produce satisfactory results, then it should be deployed in LAMMPS[4] together with a script to recombine the NNP output with the modified short-range interactions for analysis.

Acknowledgements

I would like to thank the School of Physics and Astronomy for granting me financial assistance through the 2025 Career Development Summer Scholarship scheme. I would also like to thank my supervisor, Professor Andreas Hermann, for his insights and guidance throughout the project. Finally, I extend my gratitude towards Yu Cai, who provided me with her n2p2 scripts and the technical assistance required to use them.

Personal statement

An important transferable skill I have developed during this project has been project-level communication, regularly scheduling meetings with my supervisor to discuss next steps, having the chance to present my progress at weekly social lunches to PhD students, and relaying key results via email. Also over the course of the project, I have learned to use Linux and Bash scripting, including the use of CASTEP for Density Functional Theory (DFT) calculations, deepened my understanding of inter-atomic potentials, improved my Programming Skills, and experienced training a Neural Network Potential. The experience with Linux and the command line has been especially useful, as these are skills that will carry over into any technical computer-based work I may do in the future. Despite not achieving the desired outcomes of this project, I am glad to have set the foundations for further development in this area.

References

[1] V. L. Deringer, M. A. Caro, and G. Csányi. "Machine Learning Interatomic Potentials as Emerging Tools for Materials Science". *Advanced Materials* 31.46 2019. DOI: 10.1002/adma.201902765.

[2] E. Kocer, T. W. Ko, and J. Behler. "Neural Network Potentials: A Concise Overview of Methods". 2021. DOI: 10.48550/ARXIV.2107.03727.

[3] A. Singraber et al. "Parallel Multistream Training of High-Dimensional Neural Network Potentials". *Journal of Chemical Theory and Computation* 15.5 2019. DOI: 10.1021/acs.jctc.8b01092.

[4] A. P. Thompson et al. "LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales". *Computer Physics Communications* 271 2022. DOI: https://doi.org/10.1016/j.cpc.2021.108171.

[5] S. J. Clark et al. "First principles methods using CASTEP". *Zeitschrift für Kristallographie - Crystalline Materials* 220.5–6 2005. DOI: 10.1524/zkri.220.5.567.65075.

[6] P. M. Morse. "Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels". *Phys. Rev.* 34 1929. DOI: 10.1103/PhysRev.34.57.

[7] P. G. Hajigeorgiou. "An extended Lennard-Jones potential energy function for diatomic molecules: Application to ground electronic states". *Journal of Molecular Spectroscopy* 263.1 2010. DOI: 10.1016/j.jms.2010.07.003.

[8] X. Wang et al. "The Lennard-Jones potential: when (not) to use it". *Physical Chemistry Chemical Physics* 22.19 2020. DOI: 10.1039/c9cp05445f.