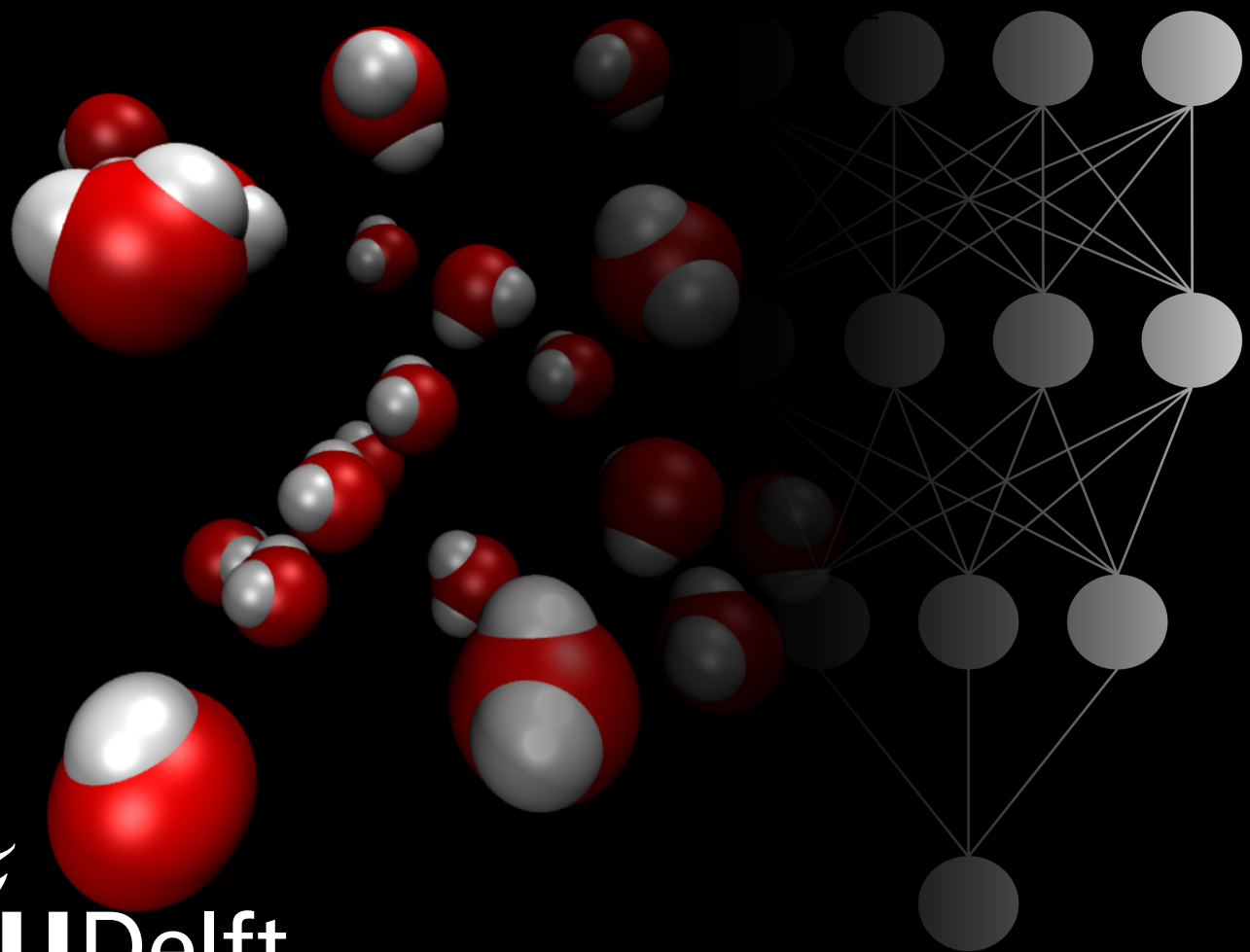
Evaluation of Machine-Learning Interatomic Potentials for Liquid Water Systems

ME55035 ME-EFPT Thesis

Wee Hong Xiang Declan (5691877)

Delft University of Technology



TU Delft

Evaluation of Machine-Learning Interatomic Potentials for Liquid Water Systems

by

Wee Hong Xiang Declan (5691877)

—

Supervisors: Dr. Othonas Moultos & Ir. Jelle Lagerweij
Faculty: Faculty of Mechanical Engineering, Delft

System of H2O modeled using AIMD, visualized with the VMD software package and Neural Network drawn using paint.

TU Delft

Contents

Nomenclature viii

1 Introduction 1

2 Methods 3
2.1 Density Functional Theory . . . 3
2.2 Machine Learning . . . 5
2.2.1 DeePMD model . . . 6
2.2.2 NequIP model . . . 8
2.2.3 Allegro model . . . 9
2.3 Current research landscape . . . 9

3 Results 11
3.1 Model Training . . . 11
3.1.1 DeePMD . . . 13
3.1.1.1 DPA-1 Model . . . 13
3.1.1.2 DPA-2 Models . . . 13
3.1.2 NequIP . . . 13
3.1.3 Allegro . . . 13
3.1.4 Training data efficiency . . . 14
3.2 MLMD with LAMMPS . . . 15
3.2.1 Simulation Speed . . . 15
3.2.2 Thermodynamic properties . . . 16
3.2.2.1 Densities . . . 16
3.2.2.2 Radial distribution functions and coordination numbers . . . 17
3.2.2.3 Molar Heat Capacity at Constant Pressure . . . 19
3.2.2.4 Isothermal Compressibility . . . 20
3.2.2.5 Diffusivity and Viscosity . . . 21

4 Conclusions and future outlook 25

References 27

A Summary of results 32

B Loss ratios 33

C MLIP Training Scripts 36

D MLIP Software Packages 66

Acknowledgements

This thesis marks the culmination of ten months of effort, made possible by the support and contributions of many. I would like to express my sincere gratitude to all those who helped bring this work to completion.

First and foremost, I thank my thesis supervisors, Ir. Jelle Lagerweij and Dr. Othonas Moultos, for their invaluable guidance, patience, and encouragement. Their deep knowledge and passion for the field have shaped both the direction of this work and my understanding of it. I would also like to thank the members of the Thermodynamics department for their willingness to assist whenever I stepped into the office seeking help.

Next, I would like to thank my friends for their support, as well as for the board game nights (with the notable exception of Tarun, who made it his mission to sabotage me every game), bouldering sessions, and meals we shared. These were the small mercies that kept me from losing my mind.

I also wish to thank the coffee from Harvest Café, Dune, and Lot, especially the Geisha pour-overs, which sustained me through long hours of reading and writing. A wise quote printed on the P&E lab's coffee machine reads, "Good things happen over good coffee." Thankfully, I found mine elsewhere.

To my fellow users of the Delftsblue A100 GPU cluster, I offer my sincere apologies for monopolizing the A100s over the past few months, at times occupying up to 32 of the 36 available GPUs. Thank you for your patience.

Finally, and most importantly, I thank my mother, my late father, and my late grandfather for their unwavering support—both emotional and financial. This past year has not been the easiest, but I made it through. I hope I have made you proud somehow.

Declan Wee
June 2025

Abstract

This report benchmarks the performance of several machine learning interatomic potentials (MLIPs) in simulating thermodynamic and transport properties of water. Classical force fields are efficient but often fail to capture complex interactions such as many-body effects and nuclear quantum effects. Quantum-scale methods, such as ab initio molecular dynamics, can account for these phenomena with high accuracy, but their computational cost limits their use to small systems and short timescales. MLIPs offer a practical compromise by learning interatomic potentials from quantum mechanical data, enabling simulations that are both accurate and scalable.

We evaluate two MLIPs, DeePMD and Allegro, trained on SCAN-metaGGA multiphase data, and assess their ability to reproduce key properties such as self-diffusivity and isothermal compressibility. The models yielded self-diffusivity values of $0.82 \times 10^{-9}\ \mathrm{m}^2/\mathrm{s}$ and $0.62 \times 10^{-9}\ \mathrm{m}^2/\mathrm{s}$, both of which are significantly lower than the experimental value of $2.30 \times 10^{-9}\ \mathrm{m}^2/\mathrm{s}$. This discrepancy is attributed to overly strong hydrogen bonding from the SCAN meta-GGA functional, consistent with prior studies. We also observed that training on multi-phase datasets caused the models to produce structural features characteristic of a mixture of phases, as revealed by the radial distribution functions. While DeePMD produced values closer to experiment, it exhibited unphysical trends, including decreasing diffusivity with increasing system size and decreasing isothermal compressibility with increasing temperature.

These results illustrate both the promise and limitations of MLIPs in capturing the complex behavior of water and suggest that care must be taken in dataset construction and model validation.

Nomenclature

Abbreviations

Abbreviation	Definition
TMD	Temperature of maximum density
ACSF	Atom-Centered Symmetry Function
BPSF	Behler-Parinello Symmetry Functions
DFT	Density Functional Theory
XC	Exchange Correlation
GGA	Meta-generalized gradient approximation
mGGA	Meta-generalized gradient approximation
SCAN	Strongly Constrained and Normed
ML	Machine-Learning
MC	Monte Carlo
MD	Molecular Dynamics
AIMD	Ab initio Molecular Dynamics
MLIP	Machine-Learning Interatomic Potentials
NN	Neural Network
NNP	Neural Network Potential
PES	Potential Energy Surface
NVT	Canonical Ensemble
NPT	Isothermal-Isobaric Ensemble

1

Introduction

Water is a fundamental requirement for life. The applications of water cover a broad spectrum that ranges from something as simple as daily showers to industrial applications electrocatalysis [1], desalinated pure water [2] solvents for chemical and biological processes [3], or as a cleaning fluid for semiconductor wafers [4]. A solid understanding of water's behavior under various thermodynamic conditions is essential to optimize these processes.

Despite the relatively simple molecular structure and small size of H_2O, the thermodynamic behavior of liquid water is atypical compared to most other liquids and exhibits several well-known anomalies [5]. These anomalies include the decrease in density upon freezing into ice Ih [6], a high specific heat capacity [7], and the presence of a minimum in the isothermal compressibility at $45°$C [8]. To predict and understand these anomalies, researchers have developed molecular models of water. A fundamental step in this process is understanding the molecular structure of water. The molecule is composed of two hydrogen atoms and one oxygen atom. Because the oxygen nucleus contains more protons than hydrogen, it attracts electrons more strongly, leading to an uneven charge distribution. Although the molecule is electrically neutral overall, the atoms carry partial charges. Experiments have shown that the oxygen atom carries a partial charge of approximately $0.7\,\mathrm{eV}$, while a total positive charge of equal magnitude is shared between the two hydrogen atoms [9].

Experimental data also indicate that the O–H bond length is approximately 0.9578 Å, with an angle of 104.48^o between the hydrogen atoms in liquid water [10]. This geometry arises from electron pair repulsion, which leads to a stable tetrahedral arrangement around the oxygen atom. The implications of having an electronegative oxygen atom and two electropositive hydrogen atoms are twofold. First, the molecule has a permanent dipole moment. Second, these partial charges lead to the formation of a tetrahedral hydrogen-bonding network between water molecules. While this network is fully realized in ice, liquid water contains both bonded and nonbonded molecules. As temperature decreases, the network becomes more structured and complete, reducing local density and resulting in the temperature of maximum density (TMD) at $277K$ [11].

The study of water at the molecular level has traditionally relied on experimental techniques such as Raman spectroscopy [12], X-ray and neutron scattering [13], X-ray absorption spectroscopy (XAS) [14], and infrared spectroscopy [15]. However, these methods are often limited by the ultrafast timescales over which hydrogen bonds form and break—on the order of picoseconds to nanoseconds [16], and by the rarity of certain events, such as self-ionization (approximately one in every two billion molecules) [17]. Additionally, experimental studies tend to be costly and time-consuming due to the requirements of equipment procurement, calibration, and execution.

Molecular simulations, typically in the form of Monte Carlo (MC) or Molecular Dynamics (MD) approaches [18], offer an alternative that addresses many of these limitations. The core of MD simulations is the potential energy surface (PES), from which the force field is derived by taking the local energy gradient with respect to the atomic positions. The accuracy of the PES critically influences simulation outcomes. Available approaches span from simple rigid non-polarizable models [19], to more accurate

but computationally expensive ab initio molecular dynamics (AIMD) where the potential energy of the system is derived through density functional theory (DFT) calculations [20]. To date, most large-scale simulations rely on classical force fields for computational efficiency, but this comes at the expense of accuracy, because the derived properties are sensitive to the PES. Notably, classical models lack the ability to describe reactive events such as hydrogen bonding and proton transfer in water [21].

The development of complex force fields that can handle bond breaking and formation, such as ReaxFF [22] and MB-pol [23–25] has advanced the accuracy of molecular simulations. However, such force fields often lack flexibility and require re-parameterization for different systems. For example, MB-pol was trained on data obtained using the computationally expensive CCSD(T) method [26–28] for water clusters, and its performance in aqueous mixtures or other solvents remains largely untested. This specificity implies that re-parameterization may be necessary, or the model may fail to deliver sufficiently accurate results in different chemical environments.

Ab initio methods provide a highly accurate PES directly from quantum mechanical principles. However, they come with a substantial computational cost, making them unfeasible for simulating large systems or accessing relevant timescales, even with state-of-the-art high-performance computing clusters [29].

To address the computational expense associated with AIMD, researchers have increasingly turned to machine learning (ML) techniques. In this approach, ML algorithms are trained on datasets generated from DFT calculations, with the aim of learning correlations between molecular configurations and their associated energies, forces and sometimes, stresses. Once trained, these models can approximate the PES and corresponding forces, thereby enabling efficient and accurate MD simulations.

The resulting interatomic potentials are commonly referred to as neural network potentials (NNPs), machine learning interatomic potentials (MLIPs), or machine learning force fields (MLFFs). The term “machine learning potentials” (MLPs) is occasionally used but is less preferred due to potential confusion with multi-layer perceptrons. In this report, we will adopt the term MLIPs. Notable MLIPs such as ANI-2X [30] and the implementation in VASP [31] have existed for several years, each with distinct architectural choices. More recently, advanced MLIPs incorporating deep learning techniques—such as self-attention [32] and graph-based convolutions [33]—have emerged. Examples include DeePMD v2 [34], NequIP [35], and Allegro [36].

While the accuracy of MLIPs has received considerable attention in recent literature [37–40], their computational cost has often been underexplored. This aspect becomes particularly critical when simulations are scaled to large systems or extended timescales. The total computational expense associated with MLIPs can generally be divided into three stages: the generation of the training dataset, typically using density functional theory (DFT); the time and computational resources required to train the ML model; and the cost of using the trained model to evaluate energies, forces, and stresses during molecular dynamics (MD) simulations.

The present study evaluates the performance of three widely used MLIPs (DeePMD, Allegro, and NequIP) by considering both predictive accuracy and computational efficiency. Specifically, we examine the extent to which each model can reproduce DFT-level accuracy in molecular dynamics trajectories, compare their computational performance during training and inference, and assess the sensitivity of key thermodynamic and transport properties.

It is important to note that the aim of ML-based potentials is to replicate the accuracy of their reference data, whether that data originates from DFT calculations, high-level quantum chemical methods, empirical force fields, or experimental measurements. While benchmarking MLIPs against their original training targets is the most direct validation strategy, it remains unclear how errors in the MLIP, and approximations in the reference data propagate into thermodynamic properties such as diffusion coefficients, densities, or structural distributions. Vega and Abascal [41] systematically benchmarked a range of rigid, non-polarizable water models against an extensive set of experimental data. In a similar spirit, this study undertakes a performance assessment for DeePMD, Allegro, and NequIP, using a representative subset of thermodynamic and transport properties adapted from their work.

2

Methods

2.1. Density Functional Theory

Based on Quantum mechanics, the PES of any medium at a given instance can be solved for by time-independent Schrodinger's equation [42]:

$$\hat{\mathcal{H}}\psi(\{r_i\},\{R_i\}) = E\psi(\{r_i\},\{R_i\}) \tag{2.1}$$

where $\{r_i\}$ and $\{R_i\}$ denotes the set of electron and nuclear coordinates respectively. Further simplifications are made by applying the Born-Oppenheimer approximation [43] which assumes that the nuclear terms contained in the Hamiltonian can be treated separately on the basis that nuclear movement relative to the electronic movement is insignificant, this is justified by the relatively large mass of the nucleus when compared to electrons, this gives:

$$\left[\frac{h^2}{2m}\sum_{i=1}^{N}\nabla_i^2 + \sum_{i=1}^{N}V(\boldsymbol{r_i}) + \sum_{i=1}^{N}\sum_{j<i}U(\boldsymbol{r_i},\boldsymbol{r_j})\right]\psi = E\psi(\{\boldsymbol{r_i}\}) \tag{2.2}$$

The first term on the left is the kinetic energy in each electron, the second is the energy from the interactions between electrons and nuclei and the third being the energy from Coulombic interactions between electrons. Analogously, the simplified Schrodinger's equation shown above as a function for a single-electron i, otherwise known as the Kohn-Sham equations [44], is given by:

$$\left[\frac{h^2}{2m}\nabla_i^2 + V(\boldsymbol{r_i}) + V_H(\boldsymbol{r_i}) + V_{XC}(\boldsymbol{r_i})\right]\psi_i(\boldsymbol{r_i}) = \epsilon_i\psi_i(\boldsymbol{r_i}) \tag{2.3}$$

By comparing Equation (2.2) and Equation (2.3), it is easy see that the $V(\boldsymbol{r_i})$ term is the same term that defines the electron and nuclei, and likewise, $V_H(\boldsymbol{r_i})$ term is the electron-electron Coulombic force term that is also known as the Hartree potential. $V_{XC}(\boldsymbol{r_i})$ is an unfamiliar term, however, it makes perfect sense when considering that in a single-electron configuration, the $V_H(\boldsymbol{r_i})$ term is an un-physical one that should not exist, hence, a correction term, $V_{XC}(\boldsymbol{r_i})$, was added to the equation. The reason for not simply removing the $V_H(\boldsymbol{r_i})$ term becomes clearer in the next few paragraphs. Despite the simplifications, Equation (2.2) remains a highly complex many-body equation that cannot be solved easily which brings us to the use of the electron density. The electron density for a given coordinate is expressed as follows:

$$n(\boldsymbol{r}) = 2\sum_i \psi_i^*(\boldsymbol{r})\psi_i(\boldsymbol{r}) \tag{2.4}$$

Where ψ_i^* represents the conjugate of the i-th wave equation and the pre-factor 2 is due the possibility of two separate electrons possessing the same wave function if they have different spins. This equation forms the basis of the Hohenburg and Kohn theorems [45], that are:

- The ground-state energy from Schrodinger's equation is a unique functional of the electron density.
- The electron density that minimizes the energy of the overall functional is the true electron density corresponding to the full solution of the Schrodinger equation.

Although the first theorem states the existence of a functional, an exact form is not provided, in fact, a general form of the function remains elusive till date. However, it is possible derive a form that is easy to re-parameterize for various systems. Since the first theorem states that Schrodinger's equation can be expressed in terms of its electron density, the existence of an equation that is equivalent to that of equation 2.2 should exist. The Hamiltonian functional for a single electron i is of the following form:

$$H[\{\psi_i\}] = H_{known}[\{\psi_i\}] + H_{XC}[\{\psi_i\}] \tag{2.5}$$

Where the known term is given by:

$$H_{known}[\{\psi_i\}] = \frac{h^2}{m}\sum_i \int \psi_i^* \nabla^2 \psi_i d^3r + \int V(\boldsymbol{r_i}) n(\boldsymbol{r_i}) d^3r + \frac{e^2}{2}\iint \frac{n(\boldsymbol{r_i}) n(\boldsymbol{r_i'})}{|r - r'|} d^3r d^3r' \tag{2.6}$$

in which, the electron density is used instead, hence, the integrals across all the terms. Performing the same operation for the Hartree term:

$$H_{XC}[\{\psi_i\}] = \int V_{XC}(\boldsymbol{r_i}) n(\boldsymbol{r_i}) d^3r \tag{2.7}$$

Or:

$$V_{XC}(\boldsymbol{r_i}) = \frac{\delta H_{XC}(\boldsymbol{r_i})}{\delta n(\boldsymbol{r_i})} \tag{2.8}$$

The H_{XC} and V_{XC} terms are known as the exchange–correlation functional and the "functional derivative" of the exchange–correlation energy respectively. The co-dependency of the Hartree potential and the electron density in the Kohn-Sham equations indicates that an iterative scheme is required to obtain a solution, such an iterative scheme would look like this:

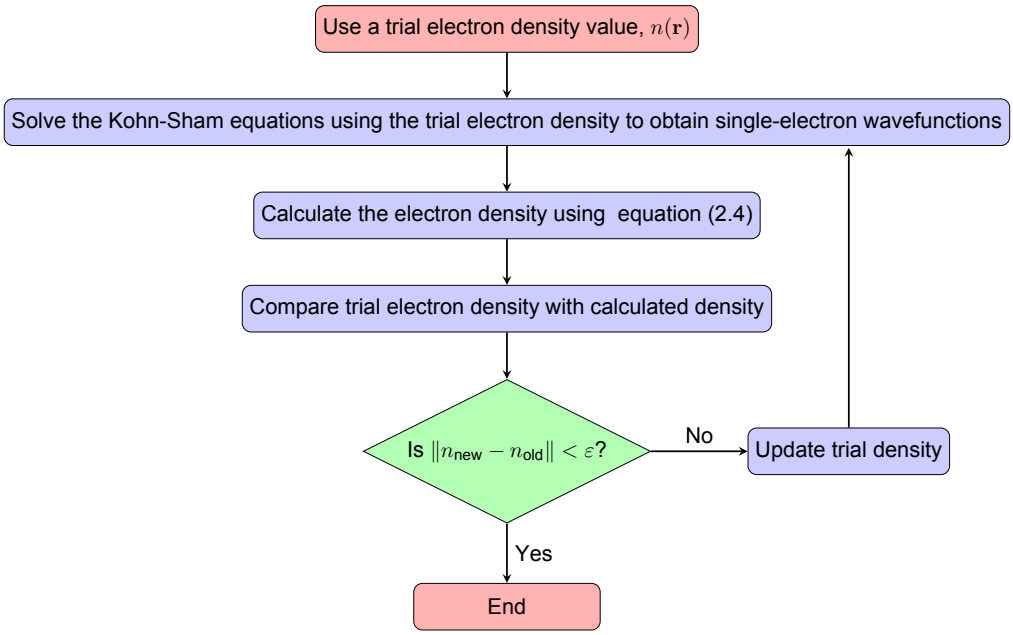


Figure 2.1: Flowchart of the self-consistent field (SCF) iteration scheme used in Kohn–Sham DFT. The procedure begins with an initial guess for the electron density, $n(\mathbf{r})$, which is used to construct the Kohn–Sham potential and solve the Kohn–Sham equations for the single-particle wavefunctions. These wavefunctions are then used to compute a new electron density via equation (2.4). The newly computed density is compared to the input trial density. If convergence criteria, defined by the difference in total energy or density between iterations, are satisfied, the process terminates. Otherwise, the trial density is updated, and the cycle repeats until self-consistency is achieved.

Much of the accuracy of DFT methods lies in how well the exchange–correlation (XC) functional is defined. The simplest case is the local density approximation (LDA), which assumes the functional derivative of the XC term behaves like that of a uniform electron gas. In this approximation, the electron density $n(\mathbf{r})$ is treated as constant throughout the system. While LDA performs reasonably well for systems with ordered crystalline structures, it often fails in disordered or inhomogeneous media, most fluids fall into this category.

Other classes of XC functionals have been proposed and systematically ranked by Perdew based on the number of physical constraints they satisfy. This ranking is known as Jacob's Ladder [46]. The SCAN functional [47] belongs to the class of meta-GGA methods, which occupy the third rung of Jacob's Ladder. Multiple studies have found that among non-hybrid density functionals, SCAN offers some of the most accurate performance for water systems [48]. While hybrid functionals and wavefunction-based methods such as CCSD(T) can provide higher accuracy, their computational cost often limits their applicability in large-scale simulations. This, combined with the availability of a high-quality SCAN dataset for water, motivated its use in this study. For a comprehensive review of exchange–correlation functionals and their relative performance, readers are referred to Kaplan et al. [49].

Ultimately, while DFT methods are widely used, they are not entirely accurate due to the lack of a general form for the exchange–correlation (XC) functional. The computational cost of solving the Kohn–Sham equations typically scales as $\mathcal{O}(N^3)$, where N is the number of electrons or basis functions, for relatively simple functionals such as the local density approximation (LDA) [50]. More sophisticated XC functionals impose additional constraints and incur significantly greater computational expense. For instance, the SCAN meta-GGA functional, which sits one rung above the generalized gradient approximation (GGA) on Jacob's Ladder, has been reported to require approximately five times the CPU time of the PBE GGA functional [51], despite having the same $\mathcal{O}(N^3)$ scaling. These demands underscore the increasing interest in machine learning-based approaches that aim to retain DFT-level accuracy while substantially reducing computational cost.

2.2. Machine Learning

As discussed in Section 2.1, DFT calculations offer high accuracy but are prohibitively expensive for large-scale simulations. ML methods provide a promising alternative by approximating the PES with significantly lower computational cost. In this context, the primary role of ML is to learn the mapping between atomic configurations and the corresponding DFT-calculated energies and forces.

Most MLIPs are built upon a common set of physical principles, with symmetry preservation being fundamental. Specifically, the model should be invariant to transformations that do not alter the physics of the system — such as translation, rotation, and the permutation of identical atoms. For example, rotating a single water molecule by 90° or exchanging the two hydrogen atoms should not change the total energy. These requirements correspond to invariance under the Euclidean group E(3).

Recent models such as NequIP, Allegro, MACE [52] and DeePMD's DPA-2 belong to a newer class of E(3)-equivariant MLIPs, these MLIPs can also be seen as a continuation of the invariant GNNs like SchNET [53] and DimeNET [54]. While equivariant models are the same as invariant models, in that, they produce a Hamiltonian that is invariant to E(3) actions, equivariant models go further by ensuring that the equivariant features like interatomic forces or directional features transform consistently with the system. In other words, when an E(3) transformation is applied to the input, the output changes in a predictable and physically meaningful way. This symmetry-aware design improves data efficiency and extends the generalizability of the model.

Another important concept to highlight is the locality approximation, introduced by Behler and Parrinello [55]. This approximation allows for the implementation of a local cut-off radius around each atom, based on the assumption that interatomic interactions decay rapidly with distance. A simple example is the Lennard-Jones potential, which sharply decreases beyond a certain separation, its attractive term scales as r^{-6}, and its repulsive term as r^{-12}. Beyond a cutoff (e.g., 8–10 Å), the interaction energy is negligible and can be safely ignored. Similarly, MLIPs adopt a cutoff radius to limit the region each atomic environment depends on, significantly reducing computational cost and enabling efficient parallelization.

Under this approximation, the total energy of a system can be decomposed into a sum of local atomic contributions:

$$U_{\text{total}} = \sum_{i}^{\text{all atoms}} U_i \tag{2.9}$$

To enforce locality for each central atom, MLIPs make use of descriptors which are mathematical representations of the surrounding atomic environment. These descriptors also serve to encode important physical symmetries, as discussed previously. A neural network trained directly on raw atomic coordinates would not respect such symmetries, since its predictions depend explicitly on the input order and position. For instance, consider a simple system with two atoms: atom 1 at position (0, 0) and atom 2 at (1, 1). If both atoms were translated uniformly along the y-axis to (0, 1) and (1, 2), respectively, the physics of the system should remain unchanged, and so should the Hamiltonian. However, a model trained directly on coordinates would not necessarily preserve this invariance because of how the NN weights are trained on the inputs.

The most commonly used symmetry-preserving descriptor is the interatomic distance. Using such descriptors not only respects physical symmetries, but also enables MLIPs to be trained on forces, which are defined as the gradient of the local energy with respect to distance:

$$\mathbf{F} = -\frac{dU}{dr} \tag{2.10}$$

Similarly, stress tensors can be included via the derivative of energy with respect to strain:

$$\sigma_{\alpha\beta} = \frac{1}{V}\frac{\partial U}{\partial \epsilon_{\alpha\beta}} \tag{2.11}$$

By incorporating forces and stresses into the training process, MLIPs can learn a more accurate and transferable representation of the potential energy surface, closely matching that of the reference *ab initio* data.

The input to an MLIP is typically a set of descriptors that encode the local atomic environment within a finite cutoff radius. These descriptors represent the positions and identities of neighboring atoms in a symmetry-preserving way and are designed to serve as effective features for predicting local atomic energies.

Descriptor design remains an active area of research, particularly for models like DeePMD DPA-1 that do not use learned features via message passing. In equivariant GNNs such as NequIP and Allegro, the representation learning is more tightly coupled with physical symmetries, which reduces the reliance on handcrafted input features and improves generalizability to new atomic environments. The functional form of these neural networks is the radial basis function as it facilitates the use of the backpropagation [56] which is needed to obtain the weights associated with the nodes and layers. Descriptors are available in various forms such as ACSF [57], SOAP [58] and Coulomb matrices [59], for extensive overview and evaluation of each descriptor, the reader is referred to available reviews [37, 58, 60–65].

Across all these approaches, the expressivity of the neural network which is dependent on its depth, width, and feature structure, plays a key role in capturing complex interatomic interactions. While deeper and more complex models can, in principle, achieve higher accuracy, they also require greater computational resources for both training and inference.

2.2.1. DeePMD model

The DeePMD machine learning potential was developed through a collaboration between Peking University, Princeton University, and Rutgers University. In DeePMD v1, atomic environments are encoded using descriptors based on local two-body and three-body interactions within a predefined cutoff radius. These are defined as:

$$(\mathcal{G}^i)_j = \mathcal{N}_{e,2}(s(r_{ij})) \qquad (\mathcal{G}^i)_j \in \mathbb{R}^{N_c \times M} \tag{2.12}$$

$$(\mathcal{G}^i)_{jk} = \mathcal{N}_{e,3}((\mathbf{R}^i_j) \cdot (\mathbf{R}^i_k)) = \mathcal{N}_{e,3}(\cos(\theta_{ijk})) \qquad (\mathcal{G}^i)_{jk} \in \mathbb{R}^{N_c \times N_c \times M} \tag{2.13}$$

Here, r_{ij} is the interatomic distance, and θ_{ijk} is the angle formed by atoms i, j, and k. These descriptors are passed through neural networks $\mathcal{N}_{e,2}$ and $\mathcal{N}_{e,3}$ to extract relevant local information.

DeePMD v2 introduced the DPA-1 framework [34], which brought two key innovations: a self-attention mechanism for improved many-body representations and the option for transfer learning through pretraining. In this study, only the attention-based descriptor is considered, as the available pretrained datasets do not include aqueous systems and thus offer limited benefit. The core motivation for introducing self-attention is to better capture many-body effects by dynamically weighting the contributions from neighboring atoms. In DPA-1, the cutoff function used to screen interactions is retained from DeePMD v1 and is given by:

$$s(r) = \begin{cases} \frac{1}{r}, & r \le r_s \\ \frac{1}{r}\left[x^3(-6x^2+15x-10)+1\right], & r_s < r < r_c, \quad x = \frac{r-r_s}{r_c-r_s} \\ 0, & r \ge r_c \end{cases} \tag{2.14}$$

This smooth switching function prevents unphysical divergences at short distances ($r < r_s$) and ensures continuity up to the cutoff r_c. In DPA-1, instead of using separate networks for 2-body and 3-body terms, the model uses a self-attention mechanism to learn many-body interactions directly from the atomic environment. The output of the self-attention network is then passed to a fitting network. Additionally, DeePMD incorporates a type embedding to handle multiple atomic species, represented as a one-hot encoded vector and mapped through a learnable transformation:

$$\mathcal{A}^i = \mathcal{N}_t(\text{one_hot}(\alpha_i)) \tag{2.15}$$

where α_i is the atomic species of atom i, and $\mathcal{N}_t$ is a neural network embedding layer. The output of the attention-based descriptor, concatenated with type embeddings, forms the atomic environment vector (AEV) for each atom. This is passed through a final neural network (the fitting network) to yield the local atomic energy E_i. The total system energy is then obtained by summing over all atomic contributions:

$$E_{\text{total}} = \sum_i E_i \tag{2.16}$$

This approach allows DeePMD to be applied to arbitrary molecular systems while maintaining permutation, translation, and rotation invariance. In summary, DPA-1 offers a significant advancement in descriptor flexibility and expressiveness compared to DeePMD v1, particularly for systems with complex many-body interactions like liquid water.

The DPA-2 framework was introduced with DeePMD v3. One of the main innovations of DPA-2 is the adoption of a unified descriptor, designed to generalize across multiple datasets containing systems with different atomic species and varying levels of DFT accuracy. The goal is to capture transferable features that are common across chemically and physically distinct environments. However, this generalization introduces challenges. The unified descriptor may encode unnecessary or system-specific features that are not relevant to the target system, potentially degrading model accuracy. In such cases, adding more training data from unrelated systems can increase the root mean squared error (RMSE) on a specific task, contrary to typical expectations in machine learning.

Architecturally, DPA-2 consists of a descriptor network and a fitting network, designed to predict the atomic energy contribution $E_i = F(D_i(R, Z))$, where R and Z denote atomic positions and atomic numbers, respectively. The descriptor D_i is a smooth, symmetry-preserving mapping that encodes

the local chemical environment of atom i and is constructed by concatenating two types of single-atom features: $D_i = \text{concat}(f_i^0, f_i^2)$. The first component, f_i^0, is an invariant feature derived from the atomic number Z_i. It is generated by passing a one-hot vector of Z_i through a NN. This vector is then passed through the ***RepInit*** module, which concatenates to f_i^0 the local geometric information extracted from pairwise descriptors g_{ij}^0 (invariant) and h_{ij}^0 (equivariant). These descriptors are based on the switched inverse distance function and relative position vectors of atom pairs within a cutoff radius r_c^0. Mathematically, these are defined as:

$$g_{ij}^0 = s(r_{ij}), \quad h_{ij}^0 = s(r_{ij}) \cdot \left(\frac{x_{ij}}{|r_{ij}|}, \frac{y_{ij}}{|r_{ij}|}, \frac{z_{ij}}{|r_{ij}|} \right),$$

where $s(r_{ij}) = w(|r_{ij}|)/|r_{ij}|$ is a smoothly decaying function to ensure differentiability at the cutoff. The RepInit module outputs an updated single-atom feature f_i^1. The second part of the descriptor, f_i^2, is generated by passing f_i^1, g_{ij}^1, and h_{ij}^1 through 12 stacked RepFormer layers. These transformer-style layers perform local updates using convolutional operations, MLPs, dot-product interactions, and localized self-attention mechanisms. The use of both invariant and equivariant channels in DPA-2 is similar to the architecture of Allegro and allows the model to account for the equivariant features within local atomic environments. As a result, the model captures both the equivariant features, as well as the many-body interactions.

Finally, the descriptor D_i is passed to the fitting network F. The model is constructed to be invariant with respect to translations, rotations, and permutations of atoms of the same species. Similar to DPA-1, it computes forces and stresses as derivatives of the energy, ensuring smoothness, energy conservation, and a physically consistent relationship among energy, forces, and stresses.

2.2.2. NequIP model

Both the NequIP and Allegro MLIP mentioned in the subsequent section was developed by the Kozinsky group at Harvard University as a message-passing neural network for atomistic simulations incorporating the concept of equivariance. In NequIP, every atom is associated with what is known as a feature vector, denoted by the variable V. Each feature vector contains tensors of various orders, starting from a scalar to vector and then higher-order tensors. These features are constantly updated via multiple layers of Message-passing networks. In the initialization phase, a self-interaction layer operates on a one-hot vector of the atomic species to fill the $0th$ degree (scalar) value of the feature vector. Then in the first layer update, the feature tensor is updated by what is know as the convolutional layer:

$$\mathcal{L}_{acm_o}^{l_o,p_o,l_f,p_f,l_i,p_i}\left(\boldsymbol{r}_a, V_{acm_i}^{l_i,p_i}\right) = \sum_{m_f,m_i} C_{l_i,m_i,l_f,m_f}^{l_o,m_o} \sum_{b\in S} (R(r_{ab})_{c,l_o,p_o,l_f,p_f,l_i,p_i}) \boldsymbol{Y}_{m_f}^{l_f}(\hat{\boldsymbol{r}}_{ab}) V_{bcm_i}^{l_i,p_i} \tag{2.17}$$

where S represents the convolutional filter between central atom a and neighbor b, for a detailed writeup of this, the reader should refer to the original NequIP article [35]. C represents the Clebsch-Gordon coefficients [66] which has been included to facilitate the tensor product of the different order tensors. The key ingredient here is the spherical harmonic projection of the interatomic vectors, from which an insight into how each of the pairwise vectors could possibly shift is provided, and this is easily visualized by the degrees (provides more angular resolution), l and order $m \in [-l, l]$ (azimuthal shift). The summation of these spherical projections for all pairwise vectors around a central atom for a given channel can then be concatenated to provide the NN with a description of how the rotated configuration would appear to be. The multiplication by the feature vector here is then just a means of updating the initial feature vector of the central atom by the feature vector of the neighboring atoms. The scalar output of the feature vector then goes through two self-interaction layers which then outputs the central atom's energy contribution. For a more detailed read, the reader is once again referred to the original NequIP article [35].

Based on this feature vector updating scheme, with each subsequent convolution layer implemented, the feature vectors of the neighbor atoms $b \in S$ gets populated by information from their own neighborhood. It is supposed that this flow of information allows for the MLIP to implicitly learn the long-range interactions and is a feature that has been previously reviewed before [67, 68]. One should also note that, this mechanism increases the computational expense and the dependency on atoms beyond the cut-off radius complicates parallelization techniques.

2.2.3. Allegro model

Allegro operates within the same E(3)-equivariant formalism introduced by NequIP; however, it employs a strictly localized message-passing network that computes interactions directly along edges within a fixed cutoff, rather than relying on iterative, node-based message passing across multiple layers as in NequIP. As a result, the update of equivariant features does not involve information from the extended environment but instead relies on weights derived from local invariant features.

In Allegro and similar equivariant models, neural network operations such as nonlinearities are applied only to scalar features. This saves the computational expense, but more importantly, the application of nonlinear functions directly to vectors or tensors would break the symmetry requirements of the model. Instead, equivariant features are updated using symmetry-preserving linear operations, while the learning happens through transformations of the scalars. This architectural separation also facilitates the parallel computation of invariant and equivariant components, improving efficiency and scalability. The invariant feature vector is first populated by the pairwise atomic information, that is:

$$\boldsymbol{x}^{ij,L=0} = MLP_{two-body}\left(Z_i||Z_j||B(r_{ij})\right) \tag{2.18}$$

And the equivariant feature vector is initialized by a weight w that is obtained by running the invariant feature vector through a NN. The update process for the equivariant feature vector is as follows:

$$\boldsymbol{V}^{ij,L=0}_{n,l,p} = w^{ij,L=0}_{n,l,p}\boldsymbol{Y}^{ij}_{l,p} \tag{2.19}$$

In a process that is analogous to the convolution layer seen in Equation 2.17, the equivariant feature vectors are updated by a summation of the feature vectors of the neighboring atoms. The main difference here is that the feature vectors are not repeatedly updating here but are just based on the weights derived for the local invariant features. The update is given as:

$$\boldsymbol{V}^{ij,L}_{n,(l_1,p_1,l_2,p_2} = \boldsymbol{V}^{ij,L-1}_{n,l_1,p_1} \otimes \left(\sum_{k\in\mathcal{N}_1} w^{ik,L}_{n,l_2,p_2}\boldsymbol{Y}^{ik}_{l_2,p_2}\right) \tag{2.20}$$

the summative terms are called the ”embedded environment” of the central atom and the above equation was derived by exploiting the bilinearity of the initial tensor product [36]. And the weights $w^{ik,L}_{n,l_2,p_2}$ are obtained via a NN operation on the pairwise invariant feature vector from the previous layer. With each layer, the feature vectors are populated with informations from the edges of other neighboring pairs, which therefore allows for many-body interactions to be included. From this point onwards, the remaining steps are the same as that of NequIP beyond the convolution layers.

The removal of feature vector transmission ensures the strict locality and there are pros and cons in doing so. The absence of interactions beyond the cutoff radius enables efficient parallelization in models like DeePMD, as atomic environments can be processed independently. However, this locality also limits the model’s ability to capture long-range interactions, such as electrostatics and dispersion, which may be important in systems like water or ionic materials.

2.3. Current research landscape

At present, various MD simulations using DeePMD, NequIP and Allegro have been conducted. In a paper by Maxson and Szilvási [69], an investigation into the ability of DeePMD v1 and Allegro in producing vapor-liquid equilibrium (VLE) properties was performed. Key findings of the paper were that Allegro generally outperformed DeePMD v1 in producing results that matched up to the input MB-Pol data. It was also noted that during the simulation of ice phases Ih, Ic, II, and III, the energy and force errors were approximately 3–4 times that of liquid water. It was postulated that this could be attributed to the missing long-range interactions in Allegro. It was also shown via a comparison of an invariant Allegro model against the equivariant Allegro model that a marked increase in the errors was observed when equivariant features were not accounted for.

Another relevant piece of literature is the same article introducing NequIP by Batzner et al. In that paper, a comparison between DeePMD and NequIP was made by once again comparing the energies and forces output by the MLIP. However, the main purpose of the comparison was to demonstrate NequIP's data efficiency: it was trained on a dataset that was three orders of magnitude smaller than DeePMD's and yet produced lower RMSE values for both energies and forces. Moreover, the DeePMD model used in the article was an older version [70] that did not use the newer smooth cutoff function, which affects the calculation of forces, i.e., the second derivative of the energy.

While many studies have examined these MLIPs based on force, energy, and stress RMSEs, few compare them in terms of their ability to reproduce macroscopic properties such as diffusivity or shear viscosity. It is possible that the differences in accuracy at the atomic scale are not significant enough to impact macroscopic properties meaningfully.

Another point of interest is the incorporation of the self-attention mechanism into the descriptor network introduced in DPA-1. Originally developed for natural language processing (NLP), self-attention in this context serves as a flexible and learned many-body representation of atomic environments, where each atom dynamically weighs the influence of its neighbors within the cutoff radius.

In contrast, models such as NequIP and Allegro capture many-body interactions through E(3)equivariant tensor products and messagepassing frameworks. While these approaches are fundamentally different in their architectural design, no direct benchmark comparisons between DPA-1's attention-based descriptor and the tensor based representations of NequIP or Allegro have yet been published.

Lastly, MLIPs are often compared on the basis of accuracy alone, but one of their key promises is to offer DFT-level accuracy at a significantly reduced computational cost. Therefore, MLIPs should be benchmarked not just against ab initio data but also against each other, with a consistent cost-performance framework.

While architectural differences across DeePMD, NequIP, and Allegro are well-documented, direct comparisons remain limited, particularly in terms of macroscopic properties and cost-performance tradeoffs. These gaps motivate the present study's focus on evaluating MLIP performance across both atomic and bulk thermodynamic scales. Table 2.1 presents part of the entire table that was compiled in Vega's paper.

Table 2.1: Compilation of thermodynamic properties containing experimental data in Vega et al's paper on rigid non-polarizable models [41]. Experimental results were sourced from Wagner et al. [71–73]. DeePMD values in this table used results from SCAN-metaGGA DeePMD models [48, 74, 75]. The equivalent benchmarks from NequIP and Allegro were not available, hence, were omitted from this table.

	Property	Experimental	DeePMD
Isothermal compressibility	$K_{T(1bar,298K)}[\frac{\mu}{bar}]$	45.3	53.1
	$K_{T(1bar,360K)}[\frac{\mu}{bar}]$	47	
Gas properties	$P_{v(350K)}[bar]$	0.417	
	$P_{v(450K)}[bar]$	9.32	
	$B_{2(450k)}[\frac{cm^3}{mol}]$	-238	
Self-diffusion coefficient	$\ln D_{278K}[\frac{cm^2}{s}]$	-11.24	-11.53
	$\ln D_{298K}[\frac{cm^2}{s}]$	-10.68	-10.98
	$\ln D_{318K}[\frac{cm^2}{s}]$	-10.24	-10.42
	$E_a[\frac{kJ}{mol}]$	18.4	
Shear viscosity	$\eta(1bar, 298K)[mPas]$	0.896	4.72
	$\eta(1bar, 373K)[mPas]$	0.284	0.473

3

Results

3.1. Model Training

For training to begin, there are a few hyperparameters that must be chosen for every MLIP, the main ones being interaction cutoff and the sizes of the NN components. Although DeePMD, NequIP, and Allegro are all trained on total energies, atomic forces, and global stresses, each employs a distinct architecture. DeePMD DPA-1 combines a self-attention descriptor network with a three-layer fitting network, whereas DeePMD DPA-2 inserts a four-head self-attention block into the descriptor stage before the same fitting network. NequIP uses multiple $E(3)$-equivariant interaction layers with 128 feature channels, and Allegro contains a scalar NN to output values after processing scalar quantities from the vector message-passing network and scalar values in that layer. Next, the cut-off radius as discussed for Equation (2.9), this cut-off radius determines the surrounding environment of the central atom that contributes to the interactions, a 6 Å cutoff adopted in all models, since that encapsulates the 1st solvation shell of each water molecule. These network sizes determine the overall capacity and inference speed. Loss coefficients weight the residuals of energy, force, and stress in the total loss. A ratio of 1 : 100 : 1 is used here to emphasise force accuracy, because forces are the gradients of the potential-energy surface as shown in Equation (2.10). A complete list of cutoffs, layer widths, optimiser settings, and loss weights is provided in the Appendix C.

All three MLIPs were trained on the DeePMD phase-diagram dataset [76], which contains 34 577 snapshots spanning liquid water from 0–2400 K and every experimentally known ice polymorph up to 50 GPa (Ih–XV). After optimisation, model generalisability was tested on a held-out set of configurations; parity plots of predicted versus reference energies, forces, and stresses are shown in Fig. 3.1.

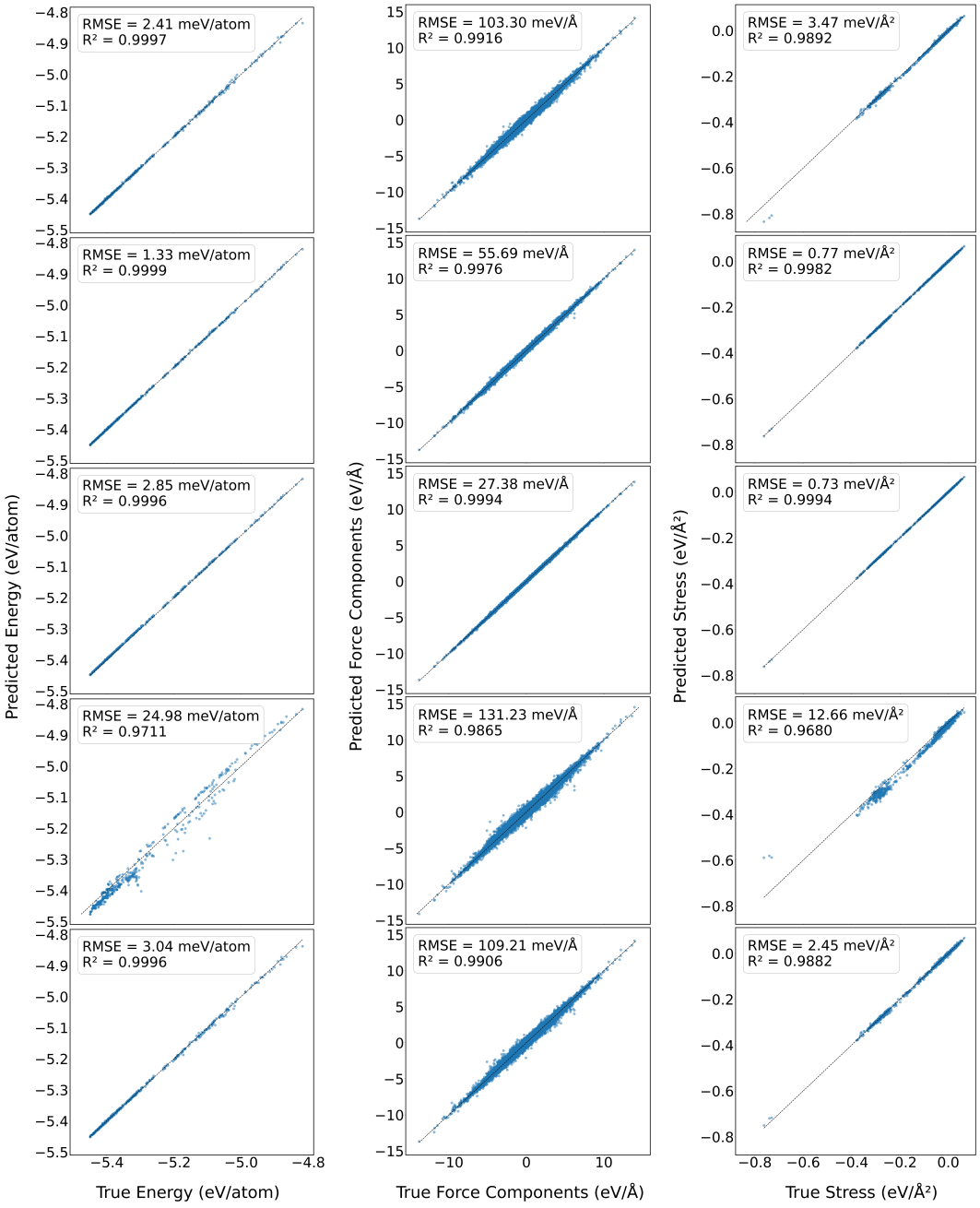


Figure 3.1: Row (1) Parity plots for the DeePMD DPA-1 model trained on a single water dataset using a self-attention descriptor. Results show accurate force and stress predictions, though energy predictions exhibit slight deviations. Row (2) Parity plots for the DeePMD DPA-2 model trained on a chemically diverse dataset using the unified descriptor and RepFormer architecture. Force predictions improve relative to DPA-1, but energy and stress predictions display greater scatter due to broader chemical coverage. Row (3) DPA-2 model trained on water-only datasets. This configuration achieves the lowest force RMSE among all DeePMD variants, but energy RMSEs increase. Row (4) Allegro model after pretraining with a high force loss weight ($\lambda_f = 100$) and low energy loss weight ($\lambda_e = 1$). The model performs well on forces but poorly on energy and stress. Row (5) Allegro model after fine-tuning with rebalanced loss weights, resulting in notable improvements across all properties, particularly for stress predictions. These parity plots collectively illustrate how architectural choices, dataset scope, and multi-phase training strategies affect the predictive performance of MLIPs.

3.1.1. DeePMD

In all five models, the stress datapoints were noticeably concentrated at the higher end, whereas the lower-stress region exhibited significant outliers. This suggests that, for the phases of interest, low-stress configurations may have been underrepresented or deemed less critical during data generation. In future work, this imbalance could be mitigated by augmenting the dataset with additional training samples in the low-stress regime.

3.1.1.1. DPA-1 Model

For a model using the attention descriptor and a fitting net of 3 layers and 250 neural nodes each, the model took 11 hours and 36 minutes to train.

3.1.1.2. DPA-2 Models

The DPA-2 framework was developed to enable large-scale atomic modeling by introducing a unified descriptor architecture in combination with E(3) equivariance via the RepFormer component [77]. One of the key features of DPA-2 is its ability to incorporate DFT training data of varying quality. The underlying rationale is that common physical patterns across datasets, even those of differing quality, can be learned and embedded within a shared descriptor. When evaluated using the publicly available multi-dataset pretrained model provided by the DeePMD team, DPA-2 achieves lower force and energy RMSEs than both the earlier self-attention-based DeePMD DPA-1 model and the Allegro model.

However, training on a chemically diverse set of systems introduces the possibility that irrelevant configurations could potentially push the model away from an optimal solution for specific systems. To test this hypothesis, we reproduced the multi-task pretraining procedure but limited the training data exclusively to DFT datasets of water. The resulting water-specific model achieved a force RMSE approximately 50% lower than that of the original multi-dataset pretrained model. However, this improvement came at the cost of a 114% increase in energy RMSE, rising from 1.33 to 2.85.

3.1.2. NequIP

Training a NequIP model on the same dataset as Allegro or DeePMD was not feasible, as NequIP does not support data parallelization across the training dataset, and the generated neighbor lists for this dataset exceeded the available VRAM on a single A100 GPU. This limitation could potentially be mitigated by using a smaller dataset, particularly given that NequIP was shown to be data-efficient in its original publication [35]. However, this approach was not pursued, as it would have undermined a direct comparison of the MLIPs presented in this study.

3.1.3. Allegro

The training of the Allegro model was conducted in two sequential stages: pre-training and fine-tuning. This approach was chosen because a one-shot training strategy, such as that employed with DeePMD, led to poor RMSE values and early overfitting in the Allegro model. During the pre-training phase, the model was optimized with an emphasis on forces, using a loss coefficient ratio of $\lambda_f = 100$ and $\lambda_e = 1$, similar to the strategy used for DeePMD. Pre-training required only 1 to 4 epochs, beyond which overfitting became apparent. Indicators of overfitting included a continued decline in training RMSEs while validation RMSEs plateaued, as well as unstable molecular dynamics runs characterized by unphysical volume fluctuations. Pre-training was therefore terminated after 3 epochs to ensure that the model had learned an adequate representation of forces while retaining capacity for improvement during fine-tuning. At this stage, the model reproduced forces and stress components reasonably well on test data, although energy and stress errors remained within approximately $\pm 5\%$. In the fine-tuning phase, the force loss coefficient was reduced in order to allow the model to shift focus to energy and stress refinement. A lower learning rate of 0.0005 was employed during this stage to preserve the pretrained weight parameters while enabling continued learning and improved accuracy.

We also examined how different stress-energy loss ratios affected the accuracy of the final model. We tested four different models, and the differences were largely negligible, as shown in table 3.1.

Table 3.1: Evaluation of model performance under varying stress-to-energy loss weight ratios during training. Each model was trained with a fixed energy loss weight of 20, while the stress loss weight was set to 10, 20, 50, or 100, respectively. The resulting RMSEs for forces, energies, and stresses are reported for each configuration. Overall, the variation in stress weighting had a negligible effect on model accuracy, with only minor fluctuations observed across all metrics.

Model	Stress Loss	Energy Loss	Force RMSE [meV/Å]	Energy RMSE [meV/atom]	Stress RMSE [meV/Å^2]
Model 1	10	20	95.50	2.61	2.29
Model 2	20	20	95.90	2.67	2.65
Model 3	50	20	96.64	2.62	2.19
Model 4	100	20	98.52	2.64	2.08

For the Allegro model, the model pretraining time was 37 minutes and 46 seconds, while the fine-tuning time was 6 hours, 25 minutes and 32 seconds. Hence, the total time was 7 hours, 3 minutes and 18 seconds.

3.1.4. Training data efficiency

To examine how efficiently the MLIPs use the training data, we trained multiple Allegro and DeePMD DPA-1 models on 25%, 50% and 100% of the total available dataset and compared the RMSEs from the respective model inference.

Table 3.2: RMSE values for energy, forces, and stresses as a function of the amount of training data used for the Allegro and DeePMD DPA-1 models. Each model was trained using 25%, 50%, and 100% of the available dataset to evaluate its data efficiency. DeePMD DPA-1 shows consistent force RMSEs across varying data sizes and marginal differences in energy and stress RMSEs, suggesting better generalization from limited data. In contrast, Allegro exhibits a clear deterioration in performance with reduced training data, indicating a greater sensitivity to dataset size. These results highlight differences in data efficiency between the two architectures, with DeePMD DPA-1 demonstrating more robust performance when trained on smaller datasets.

	Training data used [%]	Energy RMSE [meV/atom]	Force RMSE [meV/Å]	Stress RMSE [meV/Å^2]
Allegro	100	3.04	109.21	2.45
	50	4.30	136.41	3.22
	25	6.31	194.57	4.53
DeePMD DPA-1	100	2.41	103.3	3.47
	50	2.50	105.06	5.51
	25	2.41	104.49	4.59

The values observed suggest that the DeePMD DPA-1 model requires less training data to achieve consistently low RMSEs.The slight drop in DeePMD's RMSE for the smallest dataset may be attributed to the specific subset of data selected for training; ideally, multiple random samplings at each data fraction would be used to assess statistical variance. In contrast, the Allegro model exhibited a substantial decline in performance with reduced training data, a trend unlikely to arise from sampling fluctuations alone. This finding appears to contradict claims in the literature [78], where Allegro was reported to be more data-efficient than DeePMD. However, it is important to note that the comparisons in those studies used the earlier DeepPot-SE descriptors (see Equation (2.12), Equation (2.13)), whereas the present study employs the more recent DPA-1 descriptor architecture.

3.2. MLMD with LAMMPS

3.2.1. Simulation Speed

Simulation speeds for the DeePMD DPA-1 and Allegro models are generally comparable, particularly for simulations run on one or two GPUs. However, as both the simulation domain size and the number of GPUs increase, Allegro exhibits a noticeable decline in performance. This drop becomes more apparent when examining parallel efficiency, defined as the ratio of the actual simulation speed achieved on multiple GPUs to the ideal speed assuming perfect linear scaling from a single GPU. Mathematically, the parallel efficiency η is given by:

$$\eta = \frac{\text{katom-step/s on } x \text{ GPUs}}{x \times \text{katom-step/s on 1 GPU}} \tag{3.1}$$

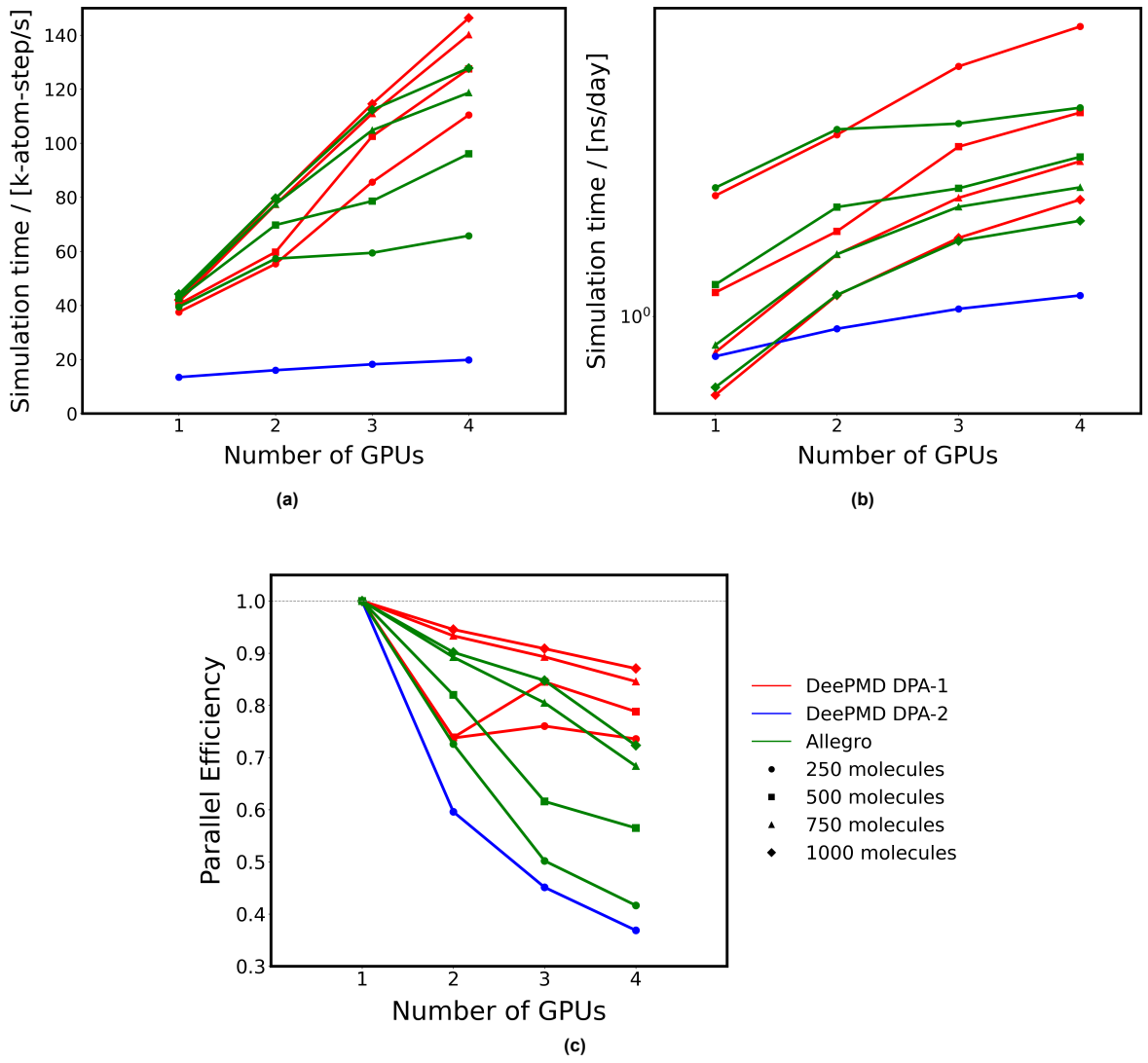


Figure 3.2: (a) Simulation speed in k-atom·steps/s as a function of the number of A100 GPUs used, where k-atom·steps denotes 10^3 atom time-steps per second (i.e. 1 k-atom·step = 1 000 atom·steps). (b) Simulation speed in nanoseconds per day (ns/day) as a function of the number of A100 GPUs used. (c) Parallel efficiency relative to single-GPU performance, illustrating how well each model scales across multiple GPUs. All simulations were performed in the *NPT* ensemble using LAMMPS with a timestep of 0.5 fs. The comparison includes the Allegro, DeePMD DPA-1, and DeePMD DPA-2 models. DeePMD DPA-1 and Allegro exhibited comparable simulation speeds, although Allegro showed poorer scalability. The DeePMD DPA-2 model was the slowest overall and demonstrated poor scalability.

Allegro's poorer scaling is expected, as its architectural design requires explicit computation of pairwise

energies and message passing of interatomic vectors across a large number of neighbors. These operations are inherently more computationally expensive and memory-intensive, particularly at large batch sizes.

Similarly, DPA-2, also an E(3)-equivariant model, exhibits performance bottlenecks due to its more expressive but heavier descriptor networks. While equivariant architectures improve accuracy by respecting the symmetries of physical space, they often sacrifice computational efficiency, especially during inference in large-scale molecular dynamics simulations. Furthermore, DPA-2's transform style architecture which incorporates attention mechanisms, further increase the inference cost.

Although the use of a trained NequIP model was not possible in this study, its computational cost is expected to exceed that of Allegro. This is because NequIP, while also an E(3)-equivariant model, lacks Allegro's strict locality constraints, resulting in the need for dense message passing across all neighbors within the cutoff, as well as the neighbor atoms of these neighboring atoms. As a consequence, NequIP's computational graph becomes increasingly complex, limiting its ability to leverage spatial domain decomposition or GPU parallelization effectively. Nonetheless, NequIP remains a compelling choice for applications where long-range interactions are significant or when predicting non-size-extensive properties, such as reaction barriers, defect formation energies, adsorption energies, or redox potentials, where accurate modeling of broader spatial correlations is essential.

The inavailability of a NequIP model and the poor compute times and scaling of the DeePMD DPA-2 model made it impossible or difficult to proceed with the computation of some following properties. In the case of the DeePMD DPA-2 model , we were able to compute the density for a system domain of 250 molecules, and the isothermal compressibility, however, it was not computationally feasible to compute the transport properties such as the diffusion coefficient given the time constraints of the project, and the long simulation runtimes necessary for the computation of these properties.

3.2.2. Thermodynamic properties

3.2.2.1. Densities

The densities of the system with respect to the number of molecules were obtained by performing *NPT* simulations at 300 K and 1 bar. Each system was equilibrated over 0.375 ns, after which the density was sampled and averaged using the `fix ave/time` command in LAMMPS. To quantify uncertainty, the time series of sampled density values was divided into five equal-sized blocks. The average density was computed for each block, and the overall mean and standard deviation were reported based on these five block averages. As shown in Table 3.3, both DeePMD DPA-1 and Allegro predicted densities that are higher than the experimental value of 997 kg/m^3. This overestimation is likely due to the overly strong hydrogen bonding network predicted by SCAN-based DFT calculations [79], which has been shown to produce an overstructured liquid and a water density of approximately 1050 kg/m^3.

Table 3.3: Comparison of water densities at 300 K and 1 bar obtained from *NPT* simulations using DeePMD DPA-1 and Allegro, across systems of increasing size. Densities are reported as mean $\pm$ standard deviation based on block averaging. All predicted densities are slightly higher than the experimental value (997 kg/m^3), likely due to SCAN's tendency to overbind via strong hydrogen bonding interactions [79].

	Experimental [kg/m^3]	**TIP4P/2005 [kg/m^3]**	**DeePMD DPA-1 [kg/m^3]**	**Allegro [kg/m^3]**
250 Molecules	997	997	1025.26 $\pm$ 4.97	1013.8 $\pm$ 2.25
500 Molecules	997	997	1026.61 $\pm$ 2.79	1013.48 $\pm$ 1.48
750 Molecules	997	997	1024.28 $\pm$ 1.83	1013.31 $\pm$ 1.18
1000 Molecules	997	997	1025.07 $\pm$ 1.27	1012.01 $\pm$ 0.75

3.2.2.2. Radial distribution functions and coordination numbers

The structure of water molecules from the *NVT* simulations can be assessed by analyzing the radial distribution functions (RDFs) for O–H and H–H atomic pairs. These RDFs provide insight into the model's ability to accurately capture the molecular geometry of water. The first peak of the O–H RDF corresponds to the bond length between the oxygen and hydrogen atoms within a water molecule. Similarly, the H–H RDF reveals the distance between the two hydrogen atoms. Using simple trigonometric relations, this distance can be used to calculate the H–O–H bond angle, θ_{HOH}, providing a quantitative measure of the intramolecular structure.

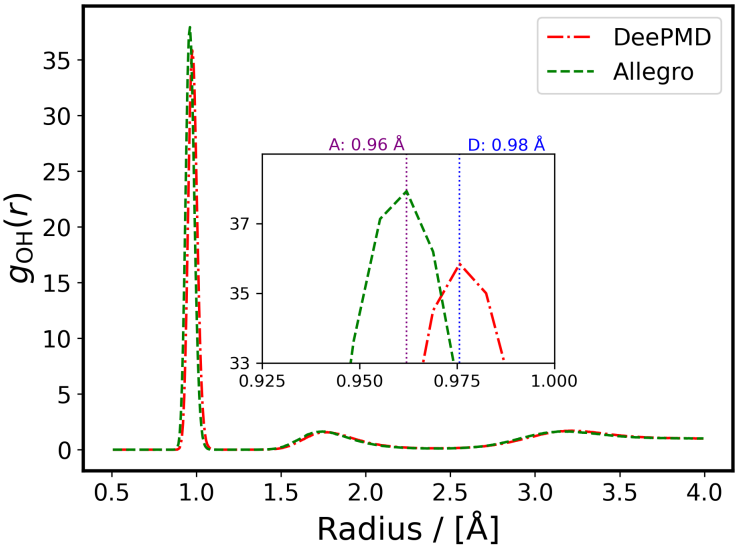


Figure 3.3: RDF for O–H atomic pairs from *NVT* molecular dynamics simulations. The first peak corresponds to the intramolecular O–H covalent bond length in a water molecule. Allegro predicts a slightly shorter bond length of 0.960 Å, whereas DeePMD yields a longer value of 0.980 Å. This feature serves as an indicator of how well each model reproduces intramolecular bond distances.

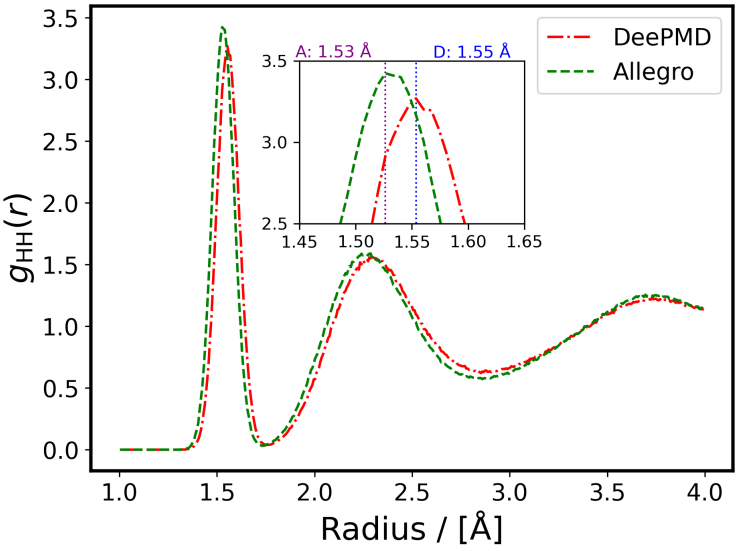


Figure 3.4: Radial distribution function (RDF) for H–H atomic pairs from *NVT* molecular dynamics simulations. The first peak corresponds to the intramolecular H–H distance in a water molecule, with Allegro predicting a shorter distance of 1.53 Å and DeePMD a slightly longer distance of 1.55 Å. These values reflect the underlying molecular geometry predicted by each model.

The O-H bond lengths and HOH bond angles are presented in Table 3.4. Both Allegro and DeePMD reproduce values in reasonable agreement with the experimental values of water, though minor deviations are observed. Allegro predicts a slightly shortened O–H bond length of 0.960 Å and a larger bond angle of $105.7°$, suggesting wider molecule geometry. DeePMD, on the other hand, slightly overestimates the bond length at 0.980 Å but yields a bond angle $104.5°$ that closely matches the experimental value. Given that both models were trained on datasets containing both liquid water and ice, These discrepancies likely reflect a compromise between the structural signatures of each phase. Overall, both ML models provide a structure that is within the range of TIP4P/2005 and SPC/E, and thus, is acceptable.

Table 3.4: Comparison of intramolecular structural parameters for water molecules, specifically the O–H bond length and the H–O–H bond angle, as predicted by Allegro and DeePMD DPA-1, alongside values from experimental measurements and classical water models. Experimental data for liquid water are taken from [10], and for ice Ih from [80]. Reference values for TIP4P/2005 and SPC/E models are obtained from [81] and [82], respectively. The machine-learned models show good overall agreement with experimental data, with small discrepancies that likely reflect influences of both liquid and ice training datasets.

Model	**O–H bond length (Å)**	**HOH bond angle θ_{HOH} (°)**
Experimental Water	0.9578	104.48
Experimental Ice Ih	1.01	109.5
TIP4P/2005	0.9572	104.52
SPC/E	1.000	109.47
Allegro	0.960	105.7
DeePMD DPA-1	0.980	104.5

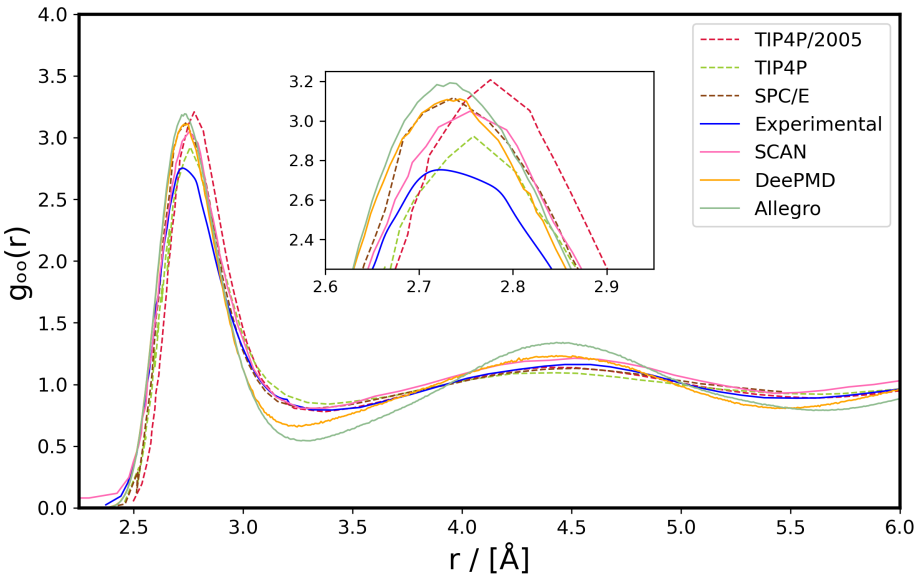


Figure 3.5: Radial distribution function (RDF) for O–O atomic pairs obtained from *NVT* molecular dynamics simulations. Both the DeePMD and Allegro models produce a first solvation shell with a higher g_{OO} peak compared to experiment, consistent with trends observed in other classical models. However, unlike classical models, the peak is shifted to shorter O–O distances, suggesting a more compressed molecular structure. This shift is in agreement with the higher densities observed in the corresponding *NPT* simulations and may indicate stronger hydrogen bonding interactions.

The O–O radial distribution function (RDF) shows that the first solvation shell has a higher g_{OO} peak in the DeePMD and Allegro models compared to experiment, a trend also seen in other classical models. However, unlike the classical models, DeePMD and Allegro predict this peak at a shorter distance from the central atom. This shift is consistent with the higher densities observed in *NPT* simulations and may also reflect stronger hydrogen bonding, as discussed earlier. In the coordinate number plot,

both DeePMD and Allegro yield lower n_{OO} values than the experiment and the classical models. This reduction coincides with a smaller cutoff radius defining the first solvation shell, further suggesting a more compressed local structure around each water molecule.

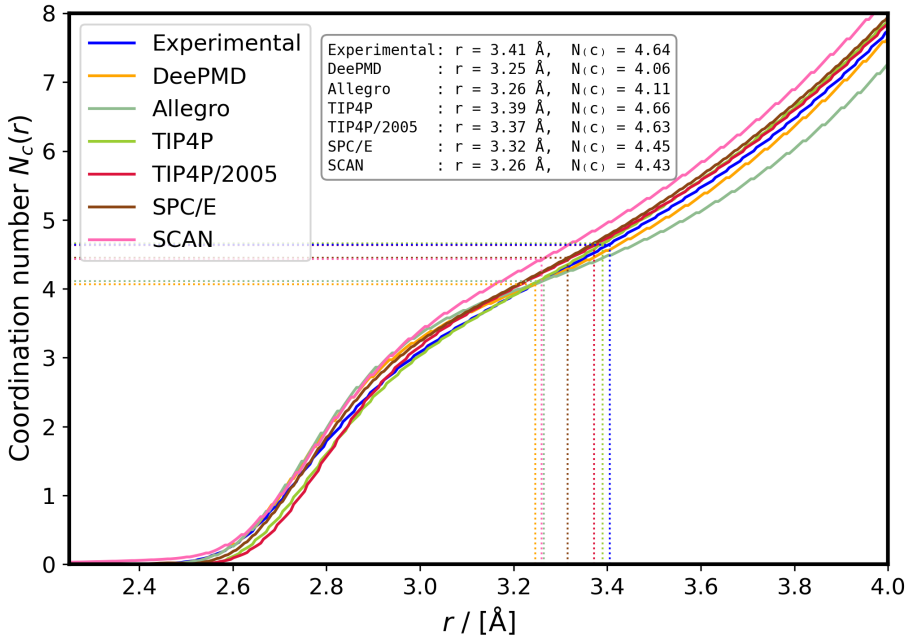


Figure 3.6: Coordination number $n_{OO}(r)$ as a function of radial distance, computed by integrating the O–O radial distribution function up to its first minimum. The position of this minimum, determined from the RDF, defines the radius of the first solvation shell. Lower coordination numbers are predicted by the DeePMD DPA-1 and Allegro due to the shorter solvation shell radius, suggesting a more compact local structure than in experiment or classical models.

3.2.2.3. Molar Heat Capacity at Constant Pressure

One approach to obtaining the constant pressure molar heat capacity, c_p, in molecular dynamics simulations is via the fluctuation formula [83]:

$$c_p = \frac{\langle H^2 \rangle - \langle H \rangle^2}{k_B T^2} \tag{3.2}$$

Table 3.5: Molar heat capacity c_p of liquid water at $300\,\mathrm{K}$, computed via the enthalpy fluctuation formula from 1 bar *NPT* simulations. Experimental reference (18 cal mol^{-1}K^{-1}) is taken from Ref. [84]. Results for MLIPs (DeePMD DPA-1, DPA-2, Allegro) and classical force fields (TIP4P/2005, TIP5P) are shown. The MLIPs substantially underpredict c_p, likely because the strong hydrogen bond network in water induces long enthalpy correlation times, so even extensive trajectories yield relatively few statistically independent enthalpy samples and thus an underestimated fluctuation variance.

Model	**Molar heat capacity,** c_p $(\mathrm{cal\,mol^{-1}\,K^{-1}})$
Experimental	18
DeePMD DPA-1	4.12
DeePMD DPA-2	4.13
Allegro	4.21
TIP4P/2005	21.1
TIP5P	29

However, values obtained using this method were significantly lower than the experimental value. It is possible that this was caused by the overly strong hydrogen bonds discussed earlier, as the enthalpies

sampled might not be sufficiently accurate. Alternatively, the finite difference method [85] may be used, and it is given by:

$$c_p \approx \frac{\langle H_2 \rangle - \langle H_1 \rangle}{T_2 - T_1} \tag{3.3}$$

where $\langle H_i \rangle$ is the average enthalpy of the system at temperature T_i. The simulations were performed in the isothermal-isobaric (*NPT*) ensemble to ensure the correct sampling of enthalpy fluctuations at constant pressure.

Table 3.6: Molar heat capacity (c_p) of liquid water at $300\,\mathrm{K}$ computed via the finite difference method from enthalpy–temperature gradients obtained in *NPT* simulations. DeePMD DPA-1, DPA-2, and Allegro yield similar predictions, all significantly overestimating the experimental value by approximately 67–70%. This overestimation is consistent with trends observed in classical models such as TIP4P and TIP5P, and is likely due to the omission of nuclear quantum effects, which require more advanced techniques like path-integral molecular dynamics for accurate treatment [86].

Model	**Molar heat capacity,** c_p $(\mathrm{cal\,mol^{-1}\,K^{-1}})$
Experimental	18
DeePMD DPA-1	30
DeePMD DPA-2	32.07
Allegro	30
TIP4P/2005	21.1
TIP5P	29

Table 3.7: Molar heat capacity c_p of liquid water at $300\,\mathrm{K}$ computed via two methods in 1 bar *NPT* simulations. The experimental reference (18 $\mathrm{cal\,mol^{-1}\,K^{-1}}$) is from Ref. [84]. Fluctuation results [(a)] show that ML-based potentials substantially underpredict c_p, likely due to long enthalpy correlation times, whereas finite-difference results [(b)] overestimate c_p by 67–70%, consistent with omission of nuclear quantum effects (see Ref. [86]).

Model	**(a) Fluctuation** c_p $(\mathrm{cal\,mol^{-1}\,K^{-1}})$	**(b) Finite-difference** c_p $(\mathrm{cal\,mol^{-1}\,K^{-1}})$
Experimental	18	18
DeePMD DPA-1	4.12	30
DeePMD DPA-2	4.13	32.07
Allegro	4.21	30
TIP4P/2005	21.1	21.1
TIP5P	29	29

The DeePMD DPA-1, DPA-2 and Allegro models obtained values that were of neglible difference, suggesting that the enthalpy–temperature gradients in all three simulations were comparable. Although the predicted values were approximately 67% to 70% higher than the experimental reference, the overestimation of heat capacity is more aligned with what is observed in classical models like TIP4P, TIP4P/2005 and TIP5P. As noted by Shiga et al. [86], this overestimation is likely attributable to the neglect of nuclear quantum effects (NQEs), which are not captured in classical molecular dynamics. Accurate treatment of NQEs requires path-integral methods, such as ring-polynomials MD.

3.2.2.4. Isothermal Compressibility

The isothermal compressibility κ_T was computed from volume fluctuations in a 1 ns *NPT* simulation using the fluctuation–dissipation relation:

$$\kappa_T = \frac{\langle V^2 \rangle - \langle V \rangle^2}{k_B T} \tag{3.4}$$

where $\langle V \rangle$ is the time-averaged volume, $\langle V^2 \rangle$ is the time-averaged squared volume, k_B is the Boltzmann constant, and T is the temperature.

Table 3.8: Isothermal compressibility (k_T) of water at 1 bar computed from volume fluctuations using molecular dynamics simulations. Results are reported for DeePMD DPA-1, DeePMD DPA-2, and Allegro models at 300 K and 360 K, and compared with experimental values. At 300 K, the DeePMD models produce values close to the experimental reference, whereas Allegro slightly underestimates k_T. However, with increasing temperature, both DeePMD models exhibit a counterintuitive decrease in compressibility, deviating from the expected trend. In contrast, Allegro captures the increasing trend more accurately, yielding values closer to experiment at 360 K.

	Experimental ($1 \times 10^{-6}\,\mathrm{bar}^{-1}$)	**TIP4P/2005** ($1 \times 10^{-6}\,\mathrm{bar}^{-1}$)	**DeePMD DPA-1** ($1 \times 10^{-6}\,\mathrm{bar}^{-1}$)	**DeePMD DPA-2** ($1 \times 10^{-6}\,\mathrm{bar}^{-1}$)	**Allegro** ($1 \times 10^{-6}\,\mathrm{bar}^{-1}$)
Isothermal compressibility k_T [1 bar, 300K]	45.3	46	44.2	40.82	32.75
Isothermal compressibility k_T [1 bar, 360K]	47	50.9	38.36	37.3	40.54

The DeePMD DPA-1 model was in good agreement with the experimental value for 300K, however, as it progressed to 360K, the isothermal compressibility decreased instead of increasing, this could be reasoned by a temperature shift necessary for the model to accurately produce properties at the respective temperatures, this temperature shift is discussed in Section 3.2.2.5. The Allegro model on the other hand, consistently underestimated the isothermal compressibilities at both temperatures but can be seen to get closer to the experimental value at 360K.

3.2.2.5. Diffusivity and Viscosity

Figure 3.7 and Figure 3.8 shows the diffusivities and viscosities computed using Allegro and DeePMD respectively, with the OCTP plugin [87]. When compared to the experimental self-diffusivity value of $2.3 \times 10^{-9}\,\mathrm{m}^2/\mathrm{s}$, both models underestimated the diffusivity. Moreover, the DeePMD model exhibited an unphysical trend of decreasing diffusivity with an increasing system size. The Yeh–Hummer correction [88] was applied to account for finite-size effects; however, the corrected values remained inconsistent and retained the same system-size-dependent trend as the uncorrected results, rather than converging towards a single consistent value. It is possible that for simulations using MLIPs, the hydrodynamic effects properly captured as these models are trained only on local environments within a cutoff radius.. Also, there are no explicit long-range forces like electrostatics in these models, which may affect how particles interact across the periodic boundaries. The viscosity remains relatively constant across all system sizes which aligns with our expectations.

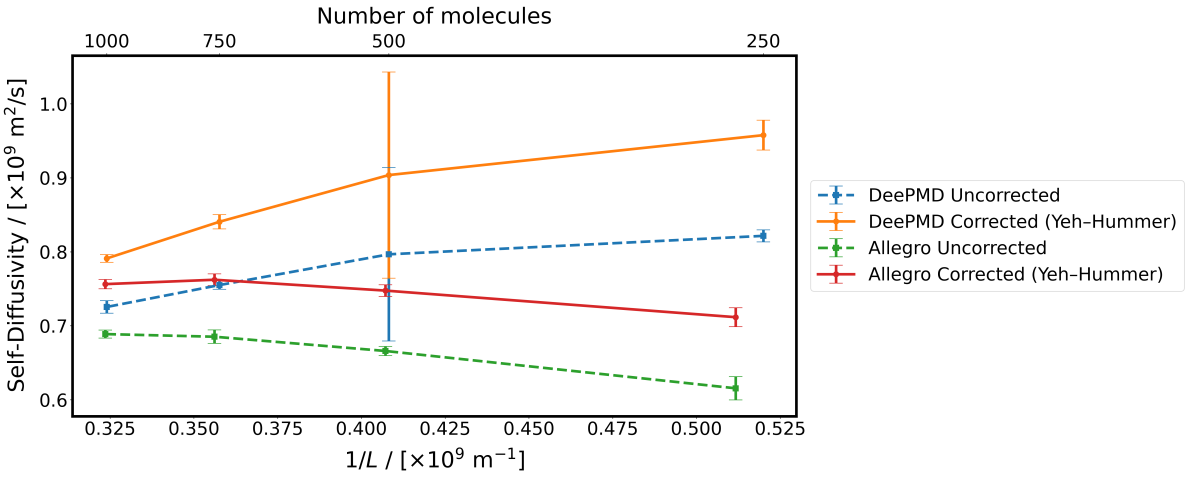


Figure 3.7: Self-diffusivity of water as a function of inverse simulation box length $(1/L)$, calculated using the OCTP plugin [87] for both the DeePMD DPA-1 and Allegro models. The results are shown before and after applying the Yeh–Hummer finite-size correction [88]. While Allegro exhibits a modest system-size dependence, the DeePMD model shows an unphysical decrease in diffusivity with increasing system size, even after correction. Experimental diffusivity at 300 K is approximately $2.3 \times 10^{-9}\,\mathrm{m}^2/\mathrm{s}$. Both models systematically underestimate this value.

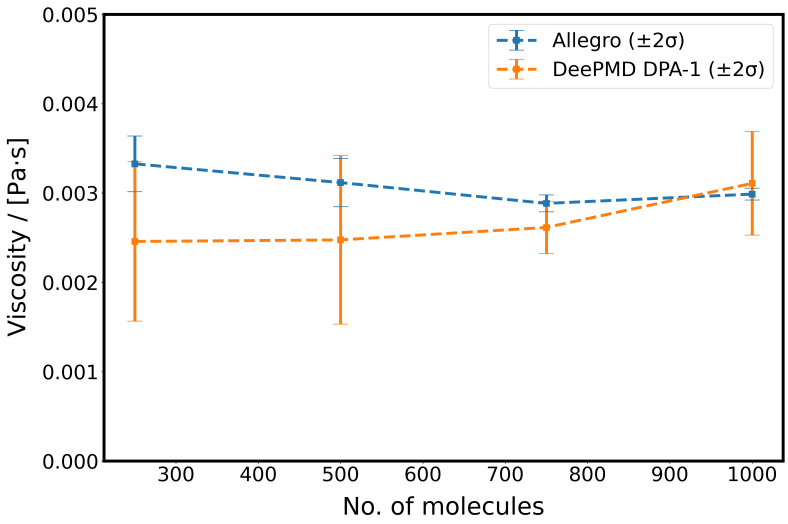


Figure 3.8: Viscosity as a function of system size, measured by the number of water molecules in the simulation. Across all system sizes, the viscosity remains relatively constant for both DeePMD and Allegro, indicating that viscosity is largely insensitive to finite-size effects in the range considered. This behavior is consistent with expectations. The flat trend contrasts with the system-size dependence observed in the diffusivity results.

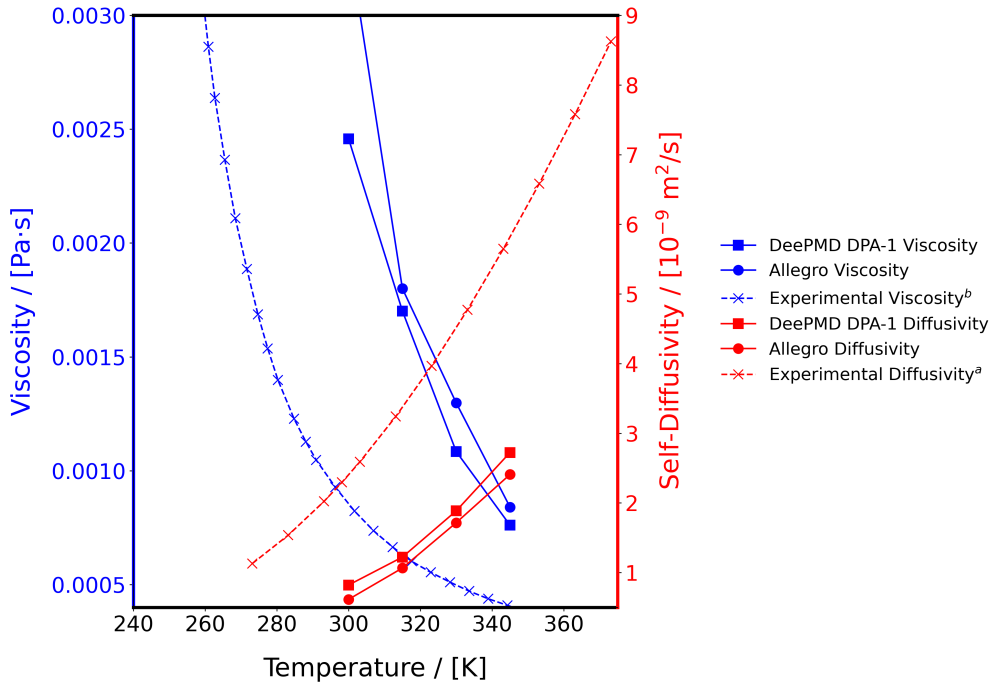


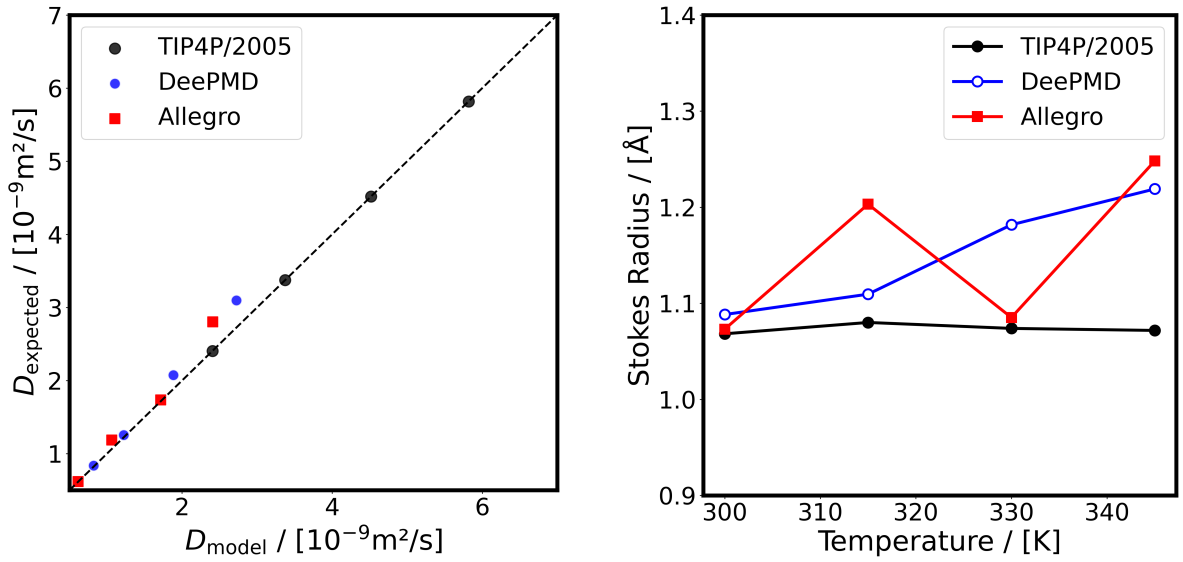
Figure 3.9: Viscosities and self-diffusivition coefficients as functions of temperature. An inverse relationship between the two properties is evident: as temperature increases, diffusivity decreases while viscosity increases, this is in line with the Stokes-Einstein relation. Diffusivity predictions from the Allegro and DeePMD DPA-1 models are also shown, both exhibiting the same qualitative trends as the experimental data, although with a noticeable temperature offset. Experimental diffusivity data are taken from Ref. [84], and viscosity data from Ref. [89].

The diffusivity and viscosity values for the system containing 250 molecules were evaluated across 300K to 345K, Figure 3.9. A key observation made is the temperature offset that has been reported in several studies done by others such as Tien et al [90, 91]. By shifting the diffusivity and viscosity values to correspond to temperatures 30 K higher, the results more closely align with experimental data. A plausible explanation for this offset is the presence of stronger hydrogen bonding, as discussed in Section 3.2.2.1. Stronger hydrogen bonds cause water molecules to be held together more tightly, resulting in increased resistance to shear and, consequently, higher viscosity.

More importantly, we observed the expected inverse relationship between diffusivity and viscosity, consistent with the Stokes–Einstein relation, which expresses this connection as:

$$D = \frac{k_B T}{6\pi\eta R_h} \tag{3.5}$$

where k_B is the Boltzmann constant, T is the absolute temperature, η is the dynamic viscosity, and R_h is the hydrodynamic radius of the molecules [92]. Using experimentally measured values of diffusivity and viscosity, we obtained an estimate of the water's hydrodynamic radius. This radius was then combined with the viscosities obtained from our *NVT* simulations to compute expected diffusivities via the Stokes–Einstein equation. As shown in Figure 3.10, the resulting values were in reasonably good agreement with those obtained directly from simulations.



(a) (b)

Figure 3.10: (a) Comparing the predicted self-diffusivities from molecular dynamics simulations (x-axis) with the values computed using the Stokes–Einstein (SE) equation (y-axis), based on model viscosities and an experimentally derived hydrodynamic radius. The plot demonstrates good agreement between the two methods for the TIP4P/2005 [81] DeePMD DPA-1 and Allegro models. (b) Temperature dependence of the Stokes radius calculated from model viscosities and predicted self-diffusivities. TIP4P/2005 TIP4P/2005 [81] DeePMD DPA-1 and Allegro models show consistent temperature trends.

However, one observation that runs counter to expectations is that the Allegro model, despite producing lower densities, yielded water that was less diffusive and more viscous than in the DeePMD simulations. A possible explanation is that, because the training dataset includes configurations of ice, the Allegro model may have learned to reproduce more characteristics of the ice phase. This is consistent with the fact that ice is less dense than liquid water and is what we observed for the Allegro model in Section 3.3.

4

Conclusions and future outlook

The primary objective of this project was to compare MLIPs, namely, Allegro, NequIP, and DeePMD, based on model accuracy and computational efficiency across the three key stages of an MLIP workflow: data generation via DFT, training, and inference through MD.

Although Allegro achieved marginally faster training times, this phase accounted for a relatively small fraction of the total computational cost, rendering the difference in training speed less significant in practical terms. To further assess training efficiency, we compared Allegro and DeePMD DPA-1 using reduced datasets. Based on the resulting RMSEs, DeePMD demonstrated the ability to achieve lower errors than Allegro despite having access to less training data. Depending on the complexity of the DFT reference data, this improved data efficiency could translate into substantial savings in computational cost.

In AIMD simulations, the DeePMD DPA-1 model required a compute budget comparable to that of Allegro, but exhibited better scalability to larger simulation domains and more effective GPU utilization during MD runs.

In terms of accuracy, DeePMD DPA-1 produced results that more closely matched experimental values across several thermodynamic properties. However, the cause of the unphysical decrease in diffusivity coefficients with increasing system sizes is something that remains unclear and further investigation might be needed. The decrease in isothermal compressibility was also another abnormality observed, it is possible that this was caused by the use of a multi-phase training dataset. In multi-phase systems like water, combining configurations from distinct states (e.g., liquid and ice) without accounting for their different structural and thermodynamic characteristics can lead to a model that compromises between phases. This can result in degraded phase-specific accuracy, even if overall RMSEs remain low. This issue is particularly relevant for water, where properties like density exhibit non-monotonic behavior near the liquid–ice transition.

However, in cases where one aims to model multi-phase properties, such as phase transitions, coexistence, or interfacial phenomena, multi-phase training can still be advantageous and possibly necessary even. In such contexts, it is important to be aware of the potential for compromised phase-specific accuracy and to adopt checks and strategies to mitigate this effect. One possible approach is to evaluate the model's performance using phase-resolved parity plots, examining RMSEs and correlation coefficients for each phase separately. If the model is found to be biased toward a particular phase, additional training data from the underrepresented phase may help to rebalance the fit. Or, one could also check for the O-H bond length and H-O-H angle to ensure that the molecules modelled accurately represent the phase they are supposed to. Lastly, one could explore increasing the complexity of the model architecture as a means of improving phase-specific accuracy. A more expressive functional form could enable better differentiation between distinct thermodynamic regimes.

The relatively minor differences in the macroscopic property predictions of the Allegro and DeePMD DPA-1 models suggest that the primary differentiator lies in the training data rather than the model

architecture. Nonetheless, the accuracy of all MLIP predictions remains highly sensitive to the quality and nature of the underlying DFT reference data. This dependence is particularly pronounced for water, where the choice of exchange–correlation (XC) functional significantly influences the strength of hydrogen bonding, which is a critical factor in accurately capturing its thermodynamic and structural behavior.

Towards the conclusion of this project, significant updates were released for both Allegro and NequIP. Firstly, parallelized training set handling was introduced for both NequIP and Allegro. More importantly, software optimizations related to PyTorch allowed for a reported 5× speedup in MD runs for systems of approximately 1000 water molecules [93]. In Section 3.1 and the Supporting Information (SI), we presented the RMSEs and parity plots from multiple Allegro models trained with different stress-to-energy loss ratios. Although more complex, these models achieved lower RMSEs than DeePMD DPA-1. With the new performance improvements, it is possible that more complex Allegro models could surpass DeePMD DPA-1 in both accuracy and MD performance.

Finally, research in this field is trending towards unified MLIP architectures (UMLIPs) such as DeePMD's DPA-2. While these are primarily designed for materials discovery, we observed that limiting the training dataset to phase-relevant configurations significantly improved force RMSEs. For the purposes of optimizing property predictions, the use of a more specific training dataset should be considered.

Future work could involve extending this study by evaluating additional thermodynamic properties. In light of the observations made regarding models trained on multi-phase datasets, it would be particularly interesting to investigate whether a multi-phase MLIP can be successfully trained using the recommended strategies in the previous paragraphs. Notably, classical force fields often struggle to accurately capture multi-phase behavior, making DFT the only reliable reference method in such contexts.

Alternatively, an evaluation of DeePMD DPA-2's zero-shot performance on systems it has not been explicitly trained on would be a valuable direction for future work, particularly given its intended application in materials discovery. Such a study would, however, need to focus on systems composed of elements present in the training set, but in previously unseen combinations.

Lastly, research in the direction of aqueous electrolytes could be pursued. Classical models have been shown to perform reasonably well for modelling diffusivity of pure water systems [94]. However, this agreement does not extend to aqueous electrolyte solutions. Classical non-polarizable models tend to predict a monotonic decrease in water diffusivity with increasing salt concentration, which contradicts experimental observations [95]. This discrepancy highlights the limitations of classical force fields in accurately capturing ion–water interactions and the collective dynamical behavior of solvated systems. Previously, Ding and Parrinello published on the need for the explicit treatment of electronic degrees of freedom in order to accurately model the influence that ions have on water structure and dynamics [96].

References

(1) Yuan, Y.; Li, J.; Zhu, Y.; Qiao, Y.; Kang, Z.; Wang, Z.; Tian, X.; Huang, H.; Lai, W. “Water in Electrocatalysis”. *Angewandte Chemie International Edition* **2025**, *64*, e202425590.

(2) Hou, Y.; Wang, M.; Chen, X.; Hou, X. “Continuous water-water hydrogen bonding network across the rim of carbon nanotubes facilitating water transport for desalination”. *Nano Research* **2021**, *14*, 2171–2178.

(3) Ball, P. “Water as an Active Constituent in Cell Biology”. *Chemical Reviews* **2007**, *108*, 74–108.

(4) Libman, S.; Wilcox, D.; Zerfas, B. “(Invited) Ultrapure Water for Advance Semiconductor Manufacturing: Challenges and Opportunities”. *ECS Meeting Abstracts* **2015**, *MA2015-02*, 1012–1012.

(5) Shi, R.; Tanaka, H. “The anomalies and criticality of liquid water”. *Proceedings of the National Academy of Sciences* **2020**, *117*, 26591–26599.

(6) Stanley, H. E., *Introduction to Phase Transitions and Critical Phenomena*; International Series of Monographs on Physics; Oxford University Press: New York, 1987.

(7) Angell, C. A.; Shuppert, J.; Tucker, J. C. “Anomalous properties of supercooled water. Heat capacity, expansivity, and proton magnetic resonance chemical shift from 0 to -38%”. *Journal of Physical Chemistry* **1973**, *77*, 3092–3099.

(8) Speedy, R. J.; Angell, C. A. “Isothermal compressibility of supercooled water and evidence for a thermodynamic singularity at −45°C”. *Journal of Chemical Physics* **1976**, *65*, 851–858.

(9) Martin, F.; Zipse, H. “Charge distribution in the water molecule—A comparison of methods”. *Journal of Computational Chemistry* **2005**, *26*, 97–105.

(10) Császár, A. G.; Czakó, G.; Furtenbacher, T.; Tennyson, J.; Szalay, V.; Shirin, S. V.; Zobov, N. F.; Polyansky, O. L. “On equilibrium structures of the water molecule”. *Journal of Chemical Physics* **2005**, *122*, 214305.

(11) Debenedetti, P. G.; Stanley, H. E. “Supercooled and Glassy Water”. *Physics Today* **2003**, *56*, 40–46.

(12) Rull, F. “Structural investigation of water and aqueous solutions by Raman spectroscopy”. *Pure and Applied Chemistry* **2002**, *74*, 1859–1870.

(13) Head-Gordon, T.; Hura, G. “Water Structure from Scattering Experiments and Simulation”. *Chemical Reviews* **2002**, *102*, 2651–2670.

(14) Näslund, L.-Å.; Lüning, J.; Ufuktepe, Y.; Ogasawara, H.; Wernet, P.; Bergmann, U.; Pettersson, L. G. M.; Nilsson, A. “X-ray Absorption Spectroscopy Measurements of Liquid Water”. *Journal of Physical Chemistry B* **2005**, *109*, 13835–13839.

(15) Fecko, C.; Eaves, J.; Loparo, J.; Tokmakoff, A.; Geissler, P. “Ultrafast hydrogen-bond dynamics in the infrared spectroscopy of water”. *Science* **2003**, *301*, 1698–1702.

(16) Lawrence, C. P.; Skinner, J. L. “Vibrational spectroscopy of HOD in liquid D2O. III. Spectral diffusion, and hydrogen-bonding and rotational dynamics”. *Journal of Chemical Physics* **2003**, *118*, 264–272.

(17) Calegari Andrade, M.; Car, R.; Selloni, A. “Probing the self-ionization of liquid water with ab initio deep potential molecular dynamics”. *Proceedings of the National Academy of Sciences* **2023**, *120*.

(18) Frenkel, D.; Smit, B. In *Understanding Molecular Simulation (Third Edition)*, Frenkel, D., Smit, B., Eds., 3rd; Academic Press: San Diego, CA, 2023, pp 1–8.

(19) Jorgensen, W. L. “Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water”. *Journal of the American Chemical Society* **1981**, *103*, 335–340.

(20) Kohn, W.; Sham, L. J. “Self-Consistent Equations Including Exchange and Correlation Effects”. *Physical Review* **1965**, *140*, A1133–A1138.

(21) Yue, S.; Panagiotopoulos, A. Z. “Dynamic properties of aqueous electrolyte solutions from non-polarisable, polarisable, and scaled-charge models”. *Molecular Physics* **2019**, *117*, 3538–3549.

(22) Senftle, T. P.; Hong, S.; Islam, M. M.; Kylasa, S. B.; Zheng, Y.; Shin, Y. K.; Junkermeier, C.; Engel-Herbert, R.; Janik, M. J.; Aktulga, H. M., et al. “The ReaxFF reactive force-field: development, applications and future directions”. *npj Computational Materials* **2016**, *2*, 1–14.

(23) Babin, V.; Leforestier, C.; Paesani, F. “Development of a “First Principles” Water Potential with Flexible Monomers: Dimer Potential Energy Surface, VRT Spectrum, and Second Virial Coefficient”. *Journal of Chemical Theory and Computation* **2013**, *9*, 5395–5403.

(24) Babin, V.; Medders, G. R.; Paesani, F. “Development of a “First Principles” Water Potential with Flexible Monomers. II: Trimer Potential Energy Surface, Third Virial Coefficient, and Small Clusters”. *Journal of Chemical Theory and Computation* **2014**, *10*, 1599–1607.

(25) Medders, G. R.; Babin, V.; Paesani, F. “Development of a “First-Principles” Water Potential with Flexible Monomers. III. Liquid Phase Properties”. *Journal of Chemical Theory and Computation* **2014**, *10*, 2906–2910.

(26) Noga, J.; Bartlett, R. J. “The full CCSDT model for molecular electronic structure”. *Journal of Chemical Physics* **1987**, *86*, 7041–7050.

(27) Scuseria, G. E.; Schaefer, H. F. “A new implementation of the full CCSDT model for molecular electronic structure”. *Chemical Physics Letters* **1988**, *152*, 382–386.

(28) Watts, J. D.; Bartlett, R. J. “The coupled-cluster single, double, and triple excitation model for open□shell single reference functions”. *Journal of Chemical Physics* **1990**, *93*, 6104–6105.

(29) Fiore, S.; Bakhouya, M.; Smari, W. W. “On the road to exascale: Advances in High Performance Computing and Simulations—An overview and editorial”. *Future Generation Computer Systems* **2018**, *82*, 450–458.

(30) Devereux, C.; Smith, J. S.; Huddleston, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E. “Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens”. *Journal of Chemical Theory and Computation* **2020**, *16*, 4192–4202.

(31) Jinnouchi, R.; Karsai, F.; Kresse, G. “On-the-fly machine learning force field generation: Application to melting points”. *Phys. Rev. B* **2019**, *100*, 014105.

(32) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. “Attention is all you need”. *Advances in neural information processing systems* **2017**, *30*.

(33) Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; Sun, M. “Graph neural networks: A review of methods and applications”. *AI Open* **2020**, *1*, 57–81.

(34) Zeng, J. et al. “DeePMD-kit v2: A software package for deep potential models”. *Journal of Chemical Physics* **2023**, *159*, 054801.

(35) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. “E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials”. *Nature Communications* **2022**, *13*.

(36) Musaelian, A.; Batzner, S.; Johansson, A.; Sun, L.; Owen, C. J.; Kornbluth, M.; Kozinsky, B. “Learning local equivariant representations for large-scale atomistic dynamics”. *Nature Communications* **2023**, *14*, 579.

(37) Zuo, Y.; Chen, C.; Li, X.; Deng, Z.; Chen, Y.; Behler, J.; Csányi, G.; Shapeev, A. V.; Thompson, A. P.; Wood, M. A.; Ong, S. P. “Performance and Cost Assessment of Machine Learning Interatomic Potentials”. *Journal of Physical Chemistry A* **2020**, *124*, 731–745.

(38) Li, H.; Wang, Z.; Zou, N.; Ye, M.; Xu, R.; Gong, X.; Duan, W.; Xu, Y. “Deep-learning density functional theory Hamiltonian for efficient ab initio electronic-structure calculation”. *Nature Computational Science* **2022**, *2*, 367–377.

(39) Pinheiro, M.; Ge, F.; Ferré, N.; Dral, P. O.; Barbatti, M. “Choosing the right molecular machine learning potential”. *Chem. Sci.* **2021**, *12*, 14396–14413.

(40) Bowman, J. M.; Qu, C.; Conte, R.; Nandi, A.; Houston, P. L.; Yu, Q. “The MD17 datasets from the perspective of datasets for gas-phase “small” molecule potentials”. *Journal of Chemical Physics* **2022**, *156*, 240901.

(41) Vega, C.; Abascal, J. “Simulating water with rigid non-polarizable models: A general perspective”. *Physical chemistry chemical physics : PCCP* **2011**, *13*, 19663–88.

(42) Schrödinger, E. “Quantisierung als Eigenwertproblem. I”. *Annalen der Physik* **1926**, *79*, 361–376.

(43) Born, M.; Oppenheimer, R. “Zur Quantentheorie der Molekeln”. *Annalen der Physik* **1927**, *389*, 457–484.

(44) Kohn, W.; Sham, L. J. “Self-Consistent Equations Including Exchange and Correlation Effects”. *Phys. Rev.* **1965**, *140*, A1133–A1138.

(45) Hohenberg, P.; Kohn, W. “Inhomogeneous Electron Gas”. *Phys. Rev.* **1964**, *136*, B864–B871.

(46) Perdew, J. P.; Schmidt, K. “Jacob’s ladder of density functional approximations for the exchange-correlation energy”. *AIP Conference Proceedings* **2001**, *577*, 1–20.

(47) Sun, J.; Ruzsinszky, A.; Perdew, J. P. “Strongly Constrained and Appropriately Normed Semilocal Density Functional”. *Physical Review Letters* **2015**, *115*, 036402.

(48) Gartner, T. E.; Zhang, L.; Piaggi, P. M.; Car, R.; Panagiotopoulos, A. Z.; Debenedetti, P. G. “Signatures of a liquid–liquid transition in an ab initio deep neural network model for water”. *Proceedings of the National Academy of Sciences* **2020**, *117*, 26040–26046.

(49) Kaplan, A. D.; Levy, M.; Perdew, J. P. “The Predictive Power of Exact Constraints and Appropriate Norms in Density Functional Theory”. *Annual Review of Physical Chemistry* **2023**, *74*, 193–218.

(50) Mohr, S.; Ratcliff, L. E.; Genovese, L.; Caliste, D.; Boulanger, P.; Goedecker, S.; Deutsch, T. “Accurate and efficient linear scaling DFT calculations with universal applicability”. *Phys. Chem. Chem. Phys.* **2015**, *17*, 31360–31370.

(51) Kingsbury, R.; Gupta, A. S.; Bartel, C. J.; Munro, J. M.; Dwaraknath, S.; Horton, M.; Persson, K. A. “Performance comparison of $r^2\mathrm{SCAN}$ and SCAN metaGGA density functionals for solid materials via an automated, high-throughput computational workflow”. *Phys. Rev. Mater.* **2022**, *6*, 013801.

(52) Batatia, I.; Kovacs, D. P.; Simm, G.; Ortner, C.; Csányi, G. “MACE: Higher order equivariant message passing neural networks for fast and accurate force fields”. *Advances in neural information processing systems* **2022**, *35*, 11423–11436.

(53) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. “SchNet – A deep learning architecture for molecules and materials”. *Journal of Chemical Physics* **2018**, *148*.

(54) Gasteiger, J.; Groß, J.; Günnemann, S. “Directional Message Passing for Molecular Graphs”. *arXiv preprint arXiv:2003.03123* **2022**.

(55) Behler, J.; Parrinello, M. “Generalized Neural-Network Representation of High-Dimensional Potential Energy Surfaces”. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(56) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. “Learning representations by back-propagating errors”. *Nature* **1986**, *323*, 533–536.

(57) Behler, J. “Atom-centered symmetry functions for constructing high-dimensional neural network potentials”. *Journal of Chemical Physics* **2011**, *134*, 074106.

(58) Bartók, A. P.; Kondor, R.; Csányi, G. “On representing chemical environments”. *Phys. Rev. B* **2013**, *87*, 184115.

(59) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. “Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning”. *Phys. Rev. Lett.* **2012**, *108*, 058301.

(60) Parsaeifard, B.; Sankar De, D.; Christensen, A. S.; Faber, F. A.; Kocer, E.; De, S.; Behler, J.; Anatole von Lilienfeld, O.; Goedecker, S. “An assessment of the structural resolution of various fingerprints commonly used in machine learning”. *Machine Learning: Science and Technology* **2021**, *2*, 015018.

(61) Himanen, L.; Jäger, M. O.; Morooka, E. V.; Federici Canova, F.; Ranawat, Y. S.; Gao, D. Z.; Rinke, P.; Foster, A. S. “DScribe: Library of descriptors for machine learning in materials science”. *Computer Physics Communications* **2020**, *247*, 106949.

(62) Pozdnyakov, S. N.; Willatt, M. J.; Bartók, A. P.; Ortner, C.; Csányi, G.; Ceriotti, M. “Incompleteness of Atomic Structure Representations”. *Phys. Rev. Lett.* **2020**, *125*, 166001.

(63) Nguyen, T. T.; Székely, E.; Imbalzano, G.; Behler, J.; Csányi, G.; Ceriotti, M.; Götz, A. W.; Paesani, F. “Comparison of permutationally invariant polynomials, neural networks, and Gaussian approximation potentials in representing water interactions through many-body expansions”. *Journal of Chemical Physics* **2018**, *148*, 241725.

(64) Langer, M. F.; Goeßmann, A.; Rupp, M. “Representations of molecules and materials for interpolation of quantum-mechanical simulations via machine learning”. *npj Computational Materials* **2022**, *8*.

(65) Jindal, S.; Chiriki, S.; Bulusu, S. S. “Spherical harmonics based descriptor for neural network potentials: Structure and dynamics of Au147 nanocluster”. *Journal of Chemical Physics* **2017**, *146*, 204301.

(66) Griffiths, D. J., *Introduction to Quantum Mechanics*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2017.

(67) Kang, S. “How Graph Neural Network Interatomic Potentials Extrapolate: Role of the Message-Passing Algorithm”. *arXiv preprint arXiv:2405.00290* **2024**.

(68) Zhang, Y.; Xia, J.; Jiang, B. “Physically Motivated Recursively Embedded Atom Neural Networks: Incorporating Local Completeness and Nonlocality”. *Phys. Rev. Lett.* **2021**, *127*, 156002.

(69) Maxson, T.; Szilvási, T. “Transferable Water Potentials Using Equivariant Neural Networks”. *Journal of Physical Chemistry Letters* **2024**, *15*, PMID: 38547514, 3740–3747.

(70) Wang, H.; Zhang, L.; Han, J.; E, W. “DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics”. *Computer Physics Communications* **2018**, *228*, 178–184.

(71) Saul, A.; Wagner, W. “A Fundamental Equation for Water Covering the Range from the Melting Line to 1273 K at Pressures up to 25□000 MPa”. *Journal of Physical and Chemical Reference Data* **1989**, *18*, 1537–1564.

(72) Wagner, W.; Saul, A.; Pruss, A. “International Equations for the Pressure Along the Melting and Along the Sublimation Curve of Ordinary Water Substance”. *Journal of Physical and Chemical Reference Data* **1994**, *23*, 515–527.

(73) Feistel, R.; Wagner, W. “A New Equation of State for H2O Ice Ih”. *Journal of Physical and Chemical Reference Data* **2006**, *35*, 1021–1047.

(74) Liu, J.; Lan, J.; He, X. “Toward High-level Machine Learning Potential for Water Based on Quantum Fragmentation and Neural Networks”. *Journal of Physical Chemistry A* **2022**, *126*, Published: June 23, 2022, 3926–3936.

(75) Malosso, C.; Zhang, L.; Car, R.; Baroni, S.; Tisi, D. “Viscosity in water from first-principles and deep-neural-network simulations”. *npj Computational Materials* **2022**, *8*, 139.

(76) Zhang, L.; Wang, H.; Car, R.; E, W. “Phase Diagram of a Deep Potential Water Model”. *Phys. Rev. Lett.* **2021**, *126*, 236001.

(77) Zhang, D.; Liu, X.; Zhang, X.; et al. “DPA-2: A Large Atomic Model as a Multi-Task Learner”. *arXiv* **2024**.

(78) Loose, T. D.; Sahrmann, P. G.; Qu, T. S.; Voth, G. A. “Coarse-Graining with Equivariant Neural Networks: A Path Toward Accurate and Data-Efficient Models”. *Journal of Physical Chemistry B* **2023**, *127*, 10564–10572.

(79) Duignan, T. “Liquid water simulation using hydrogen bond corrected SCAN and neural network potentials.” **2021**.

(80) Kuhs, W. F.; Lehmann, M. S. “The structure of the ice Ih by neutron diffraction”. *Journal of Physical Chemistry* **1983**, *87*, 4312–4313.

(81) Valle, J. V. L.; Mendonça, B. H. S.; Barbosa, M. C.; Chacham, H.; de Moraes, E. E. “Accuracy of TIP4P/2005 and SPC/Fw Water Models”. *Journal of Physical Chemistry B* **2024**, *128*, 1091–1097.

(82) Mark, P.; Nilsson, L. “Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K”. *Journal of Physical Chemistry A* **2001**, *105*, 9954–9960.

(83) Allen, M. P.; Tildesley, D. J., *Computer Simulation of Liquids*, 2nd ed.; Oxford University Press: Oxford, UK, 2017.

(84) Easteal, A. J.; Price, W. E.; Woolf, L. A. “Diaphragm cell for high-temperature diffusion measurements. Tracer Diffusion coefficients for water to 363 K”. *J. Chem. Soc., Faraday Trans. 1* **1989**, *85*, 1091–1097.

(85) Wang, Q.; Keffer, D. J.; Petrovan, S.; Thomas, J. B. “Molecular Dynamics Simulation of Poly(ethylene terephthalate) Oligomers”. *Journal of Physical Chemistry B* **2010**, *114*, 786–795.

(86) Shiga, M.; Shinoda, W. “Calculation of heat capacities of light and heavy water by path-integral molecular dynamics”. *Journal of chemical physics* **2005**, *123*, 134502.

(87) Jamali, S. H.; Wolff, L.; Becker, T. M.; de Groen, M.; Ramdin, M.; Hartkamp, R.; Bardow, A.; Vlugt, T. J. H.; Moultos, O. A. “OCTP: A Tool for On-the-Fly Calculation of Transport Properties of Fluids with the Order-n Algorithm in LAMMPS”. *Journal of Chemical Information and Modeling* **2019**, *59*, 1290–1294.

(88) Yeh, I.-C.; Hummer, G. “System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions”. *Journal of Physical Chemistry B* **2004**, *108*, 15873–15879.

(89) Sengers, J. V.; Watson, J. T. R. “Improved International Formulations for the Viscosity and Thermal Conductivity of Water Substance”. *Journal of Physical and Chemical Reference Data* **1986**, *15*, 1291–1314.

(90) Tian, W.; Wang, C.; Zhou, K. “The Dynamic Diversity and Invariance of Ab Initio Water”. *Journal of Chemical Theory and Computation* **2024**, *20*, 10667–10675.

(91) Malosso, C.; Zhang, L.; Car, R.; Baroni, S.; Tisi, D. “Viscosity in water from first-principles and deep-neural-network simulations”. *npj Computational Materials* **2022**, *8*, 139.

(92) McQuarrie, D. A., *Statistical Mechanics*; University Science Books: Sausalito, CA, 1976.

(93) Tan, C. W.; Descoteaux, M. L.; Kotak, M.; de Miranda Nascimento, G.; Kavanagh, S. R.; Zichi, L.; Wang, M.; Saluja, A.; Hu, Y. R.; Smidt, T.; Johansson, A.; Witt, W. C.; Kozinsky, B.; Musaelian, A. “High Performance Training and Inference for Deep Equivariant Interatomic Potentials”. version 1 *arXiv* **2025**.

(94) Tsimpanogiannis, I. N.; Moultos, O. A.; Franco, L. F. M.; de M. Spera, M. B.; Erdős, M.; and, I. G. E. “Self-diffusion coefficient of bulk and confined water: a critical review of classical molecular simulation studies”. *Molecular Simulation* **2019**, *45*, 425–453.

(95) Kim, J. S.; Wu, Z.; Morrow, A. R.; Yethiraj, A.; Yethiraj, A. “Self-Diffusion and Viscosity in Electrolyte Solutions”. *Journal of Physical Chemistry B* **2012**, *116*, : 22967241, 12007–12013.

(96) Ding, Y.; Hassanali, A. A.; Parrinello, M. “Anomalous water diffusion in salt solutions”. *Proceedings of the National Academy of Sciences* **2014**, *111*, 3310–3315.

(97) Morozova, T. I.; García, N. A.; Barrat, J.-L. “Temperature dependence of thermodynamic, dynamical, and dielectric properties of water models”. *Journal of Chemical Physics* **2022**, *156*, 126101.

A

Summary of results

We recompiled the same thermodynamic properties seen in [41] with the results from Section 3.2.2.1 to Section 3.2.2.5. The diffusivity values of both the DeePMD DPA-1 and Allegro models have not been finite-size corrected, as Figure 3.7 shows that the correction does not apply.

Table A.1: Comparison of selected thermodynamic and transport properties for water at various temperatures, predicted by different MLIPs, against experimental reference values. Models include DeePMD DPA-1, DPA-2, and Allegro. Experimental values are taken from Ref. [84] and Ref. [89]. Note that self-diffusion coefficients for DPA-1 and Allegro are reported without finite-size correction, as detailed in Figure 3.7, due to the correction method's failure to yield a consistent corrected value. DPA-2 results are omitted where simulations were not performed. Properties shown include isothermal compressibility (K_T), self-diffusion coefficient (D), constant-pressure heat capacity (c_p), and shear viscosity (η). TIP4P/2005 values were taken from [41] and [97].

	Property	**Exp.**	**TIP4P**	**DPA-1**	**DPA-2**	**Allegro**
Isothermal compressibility	K_T (298 K) [μ/bar]	45.3	46	44.2	40.82	30.75
	K_T (360 K) [μ/bar]	47.0	50.9	38.36	37.3	40.54
Self-diffusion coefficient	$D(298\,\mathrm{K})$ [$10^{-9}\,\mathrm{m^2\,s^{-1}}$]	2.30	2.27	0.82	—	0.62
	$D(315\,\mathrm{K})$ [$10^{-9}\,\mathrm{m^2\,s^{-1}}$]	3.37	3.16	1.22	—	1.07
	$D(330\,\mathrm{K})$ [$10^{-9}\,\mathrm{m^2\,s^{-1}}$]	4.37	4.10	1.89	—	1.72
	$D(345\,\mathrm{K})$ [$10^{-9}\,\mathrm{m^2\,s^{-1}}$]	5.79	5.19	2.72	—	2.41
Heat capacity	c_p (298 K) [cal/mol·K]	18.0	21.1	30.00	32.07	30.00
Shear viscosity	η (298 K) [mPa·s]	0.89	0.89	2.46	—	3.33
	η (315 K) [mPa·s]	0.63	0.61	1.70	—	1.80
	η (330 K) [mPa·s]	0.49	0.48	1.08	—	1.30
	η (345 K) [mPa·s]	0.40	0.39	0.761	—	0.84

B

Loss ratios

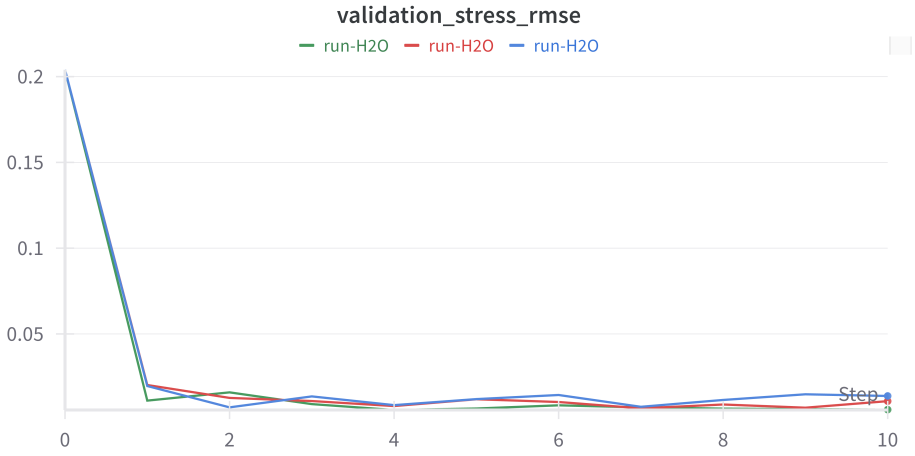


Figure B.1: Validation loss vs. epoch for the Allegro model trained with a learning rate of 1×10^{-3}.



Figure B.2: Validation loss vs. epoch for the Allegro model trained with a learning rate of 5×10^{-4}.



Figure B.3: Validation loss vs. epoch for the Allegro model trained with a learning rate of 1×10^{-4}.

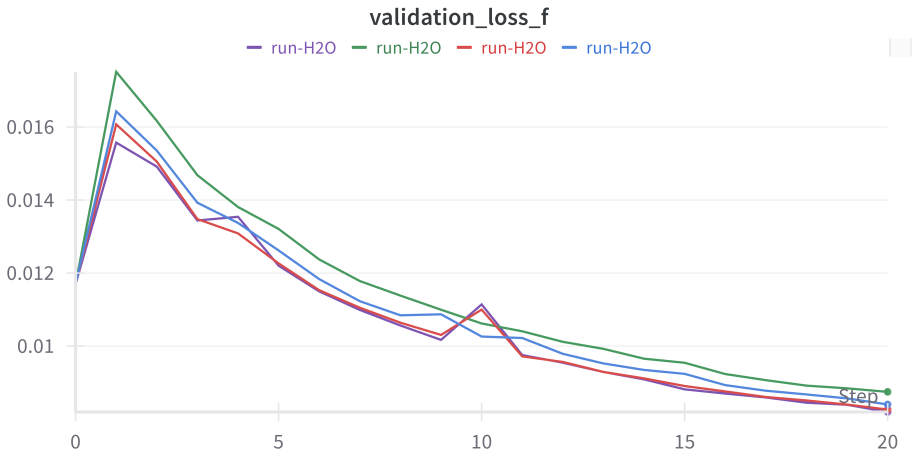


Figure B.4: Validation force RMSE vs. epoch for Allegro models trained with different stress-to-energy loss ratios. Each curve represents a distinct model configuration.

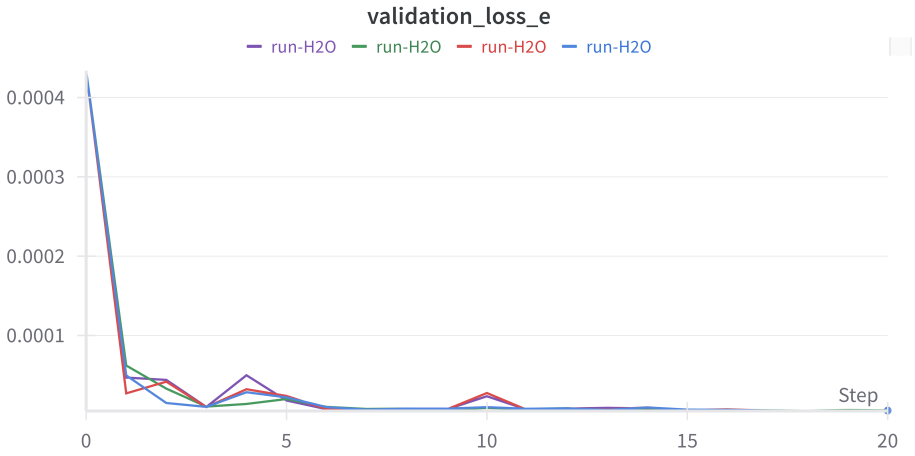


Figure B.5: Validation energy RMSE vs. epoch for Allegro models trained with varying stress-to-energy loss ratios.

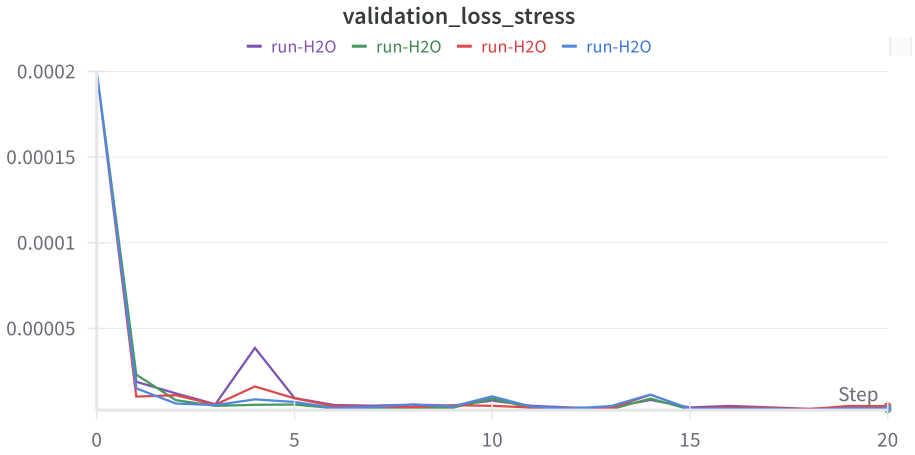


Figure B.6: Validation stress RMSE vs. epoch for Allegro models trained with different stress-to-energy loss ratios.

C

MLIP Training Scripts

```
# Two folders will be used during the training: 'root'/process and 'root'/'run_name'
# run_name contains logfiles and saved models
# process contains processed data sets
# if 'root'/'run_name' exists, 'root'/'run_name'_'year'-'month'-'day'-'hour'-'min'-'s' will be used instead.
root: results/H2O
run_name: run-H2O
seed: 123               # model seed
dataset_seed: 456       # data set seed
append: true            # set true if a restarted run should append to the previous log file

# see https://arxiv.org/abs/2304.10061 for discussion of numerical precision
default_dtype: float64
# model_dtype: float32
# allow_tf32: true    # consider setting to false if you plan to mix training/inference over any devices that are not NVIDIA Ampere or later

# -- network --
# tell nequip which modules to build
model_builders:
 - allegro.model.Allegro
 # the typical model builders from `nequip` can still be used:
 - PerSpeciesRescale
 - StressForceOutput
 - RescaleEnergyEtc

r_max: 6 # cutoff radius in length units, here Angstrom, this is an important hyperparamter to scan

# average number of neighbors in an environment is used to normalize the sum, auto precomputed it automitcally
avg_num_neighbors: auto

# radial basis
# set true to train the bessel roots
BesselBasis_trainable: true

# p-parameter in envelope function, as proposed in Klicpera, J. et al., arXiv:2003.03123
# sets it BOTH for the RadialBasisProjection AND the Allegro_Module
PolynomialCutoff_p: 12

l_max: 2                                                                    # the maximum irrep order (rotation order) for the network's features, l=1 is a good default, l=2 is more accurate but slower

# allowed: o3_full, o3_restricted, so3
parity: so3

# number of tensor product layers, 1-3 usually best, more is more accurate but slower
num_layers: 1

# number of features, more is more accurate but slower, 1, 4, 8, 16, 64, 128 are good options to try depending on data set
env_embed_multiplicity: 32

# whether or not to embed the initial edge, true often works best
embed_initial_edge: true

# hidden layer dimensions of the 2-body embedding MLP
```

```yaml
# two_body_latent_mlp_latent_dimensions: [128, 256, 512, 1024]
two_body_latent_mlp_latent_dimensions: [32]
# nonlinearity used in the 2-body embedding MLP
two_body_latent_mlp_nonlinearity: silu
# weight initialization of the 2-body embedding MLP
two_body_latent_mlp_initialization: uniform

# hidden layer dimensions of the latent MLP
# these MLPs are cheap if you have have large l/env_embed_multiplicity, so a good place to put model capacity if you can afford it
# only if you are in the ultra-fast/scalable regime, make these smaller
# latent_mlp_latent_dimensions: [1024, 1024, 1024]
latent_mlp_latent_dimensions: [32, 32, 32]

# nonlinearity used in the latent MLP
latent_mlp_nonlinearity: silu

# weight initialization of the latent MLP
latent_mlp_initialization: uniform

# whether to use a resnet update in the scalar latent latent space, true works best usually
latent_resnet: true

# hidden layer dimensions of the environment embedding mlp, none work best (will build a single linear layer)
env_embed_mlp_latent_dimensions: []

# nonlinearity used in the environment embedding mlp
env_embed_mlp_nonlinearity: null

# weight initialzation of the environment embedding mlp
env_embed_mlp_initialization: uniform

# - end allegro layers -

# Final MLP to go from Allegro latent space to edge energies:

# hidden layer dimensions of the per-edge energy final MLP
edge_eng_mlp_latent_dimensions: [32]

# nonlinearity used in the per-edge energy final MLP
edge_eng_mlp_nonlinearity: null

# weight initialzation in the per-edge energy final MLP
edge_eng_mlp_initialization: uniform

# As an alternative option to npz, you can also pass data ase ASE Atoms-objects
# This can often be easier to work with, simply make sure the ASE Atoms object
# has a calculator for which atoms.get_potential_energy() and atoms.get_forces() are defined
dataset: ase
dataset_file_name: ./input_files/mlip_input_data/H2O_PD_with_stress.extxyz                                                    # need to be a format accepted by ase.io.read
ase_args:
  format: extxyz
key_mapping:
  energy: total_energy
dataset_include_keys:
  - forces
  - energy
  - stress

# Alternatively, you may explicitly specify which chemical species in the input will map to NequIP atom type 0, which to atom type 1, and so on.
# Other than providing an explicit order for the NequIP atom types, this option behaves the same as `chemical_symbols`
chemical_symbol_to_type:
   O: 0
   H: 1

# logging
wandb: true                                                                                                                   # we recommend using wandb for logging
wandb_project: H2O-PD-Simple-Model                                                                                            # project name used in wandb
# wandb_watch: true

# see https://docs.wandb.ai/ref/python/watch
# wandb_watch_kwargs:
#   log: all
#   log_freq: 1
#   log_graph: true

verbose: info                                                                                                                 # the same as python logging, e.g. warning, info, debug, error. case insensitive
```

```yaml
# log_batch_freq: 100                                                    # batch frequency, how often to print training errors withinin the same epoch
# log_epoch_freq: 1                                                      # epoch frequency, how often to print
# save_checkpoint_freq: -1                                               # frequency to save the intermediate checkpoint. no saving of intermediate checkpoints when the value is not positive.
# save_ema_checkpoint_freq: -1                                           # frequency to save the intermediate ema checkpoint. no saving of intermediate checkpoints when the value is not positive.

# alternatively, n_train and n_val can be set as percentages of the dataset size:
n_train: 85%  # 70% of dataset
n_val: 14%    # 30% of dataset (if validation_dataset not set), or 30% of validation_dataset (if set)
learning_rate: 0.001                                                     # learning rate, we found values between 0.01 and 0.005 to work best - this is often one of the most important hyperparameters to tune
batch_size: 1                                                            # batch size, we found it important to keep this small for most applications including forces (1-5); for energy-only training, higher batch sizes work better
max_epochs: 2                                                            # stop training after _ number of epochs, we set a very large number here, it won't take this long in practice and we will use early stopping instead
train_val_split: random                                                  # can be random or sequential. if sequential, first n_train elements are training, next n_val are val, else random, usually random is the right choice
shuffle: true                                                            # If true, the data loader will shuffle the data, usually a good idea
metrics_key: validation_loss # validation_loss                                                                  # metrics used for scheduling and saving best model. Options: `set`_`quantity`, set can be either "train" or "validation, "quantity" can be loss or anything that appears in the validation batch step header, such as f_mae, f_rmse, e_mae, e_rmse
use_ema: true                                                            # if true, use exponential moving average on weights for val/test, usually helps a lot with training, in particular for energy errors
ema_decay: 0.99                                                          # ema weight, typically set to 0.99 or 0.999
ema_use_num_updates: true                                                # whether to use number of updates when computing averages
# report_init_validation: true                                           # if True, report the validation error for just initialized model

# loss function
loss_coeffs:                                                             # different weights to use in a weighted loss functions
  forces: # if using PerAtomMSELoss, a default weight of 1:1 on each should work well
  - 100.0
  - PerSpeciesMSELoss
    # stress: 50.0
  total_energy:
    - 1.0
    - PerAtomMSELoss
# note that the ratio between force and energy loss matters for the training process. One may consider using 1:1 with the PerAtomMSELoss. If the energy loss still significantly dominate the loss function at the initial epochs, tune the energy loss weight lower helps the training a lot.

# output metrics
metrics_components:
  - - forces                                 # key
    - rmse                                    # "rmse" or "mae"
  - - forces
    - rmse
    - PerSpecies: True                       # if true, per species contribution is counted separately
      report_per_component: False            # if true, statistics on each component (i.e. fx, fy, fz) will be counted separately
  - - total_energy
    - rmse
  - - total_energy
    - rmse
    - PerAtom: True                          # if true, energy is normalized by the number of atoms
  - - stress
    - rmse

# optimizer, may be any optimizer defined in torch.optim
# the name `optimizer_name`is case sensitive
# IMPORTANT: for NequIP (not for Allegro), we find that in most cases AMSGrad strongly improves
# out-of-distribution generalization over Adam. We highly recommed trying both AMSGrad (by setting
# optimizer_amsgrad: true) and Adam (by setting optimizer_amsgrad: false)
optimizer_name: Adam
optimizer_amsgrad: false
optimizer_betas: !!python/tuple
  - 0.9
  - 0.999
optimizer_eps: 1.0e-08
optimizer_weight_decay: 0

```

```yaml
# gradient clipping using torch.nn.utils.clip_grad_norm_
# see https://pytorch.org/docs/stable/generated/torch.nn.utils.clip_grad_norm_.html#torch.nn.utils.clip_grad_norm_
# setting to inf or null disables it
max_gradient_norm: null

# lr scheduler
# first: on-plateau, reduce lr by factory of lr_scheduler_factor if metrics_key hasn't improved for lr_scheduler_patience epoch
# you can also set other options of the underlying PyTorch scheduler, for example lr_scheduler_threshold
lr_scheduler_name: ReduceLROnPlateau
lr_scheduler_patience: 5
lr_scheduler_factor: 0.5

# early stopping based on metrics values.
# LR, wall and any keys printed in the log file can be used.
# The key can start with Training or validation. If not defined, the validation value will be used.
early_stopping_patiences:                                                    # stop early if a metric value stopped decreasing for n epochs
  validation_loss: 50
  # validation_f_rmse: 50
early_stopping_delta:                                                        # If delta is defined, a decrease smaller than delta will not be considered as a decrease
  validation_loss: 0.005
  # validation_f_rmse: 0.001

early_stopping_cumulative_delta: false                                       # If True, the minimum value recorded will not be updated when the decrease is smaller than delta

early_stopping_lower_bounds:                                                 # stop early if a metric value is lower than the bound
  LR: 1.0e-5

early_stopping_upper_bounds:                                                 # stop early if a metric value is higher than the bound
  cumulative_wall: 1.0e+100
```

Listing C.1: Full YAML training configuration used for Allegro pretraining

```yaml
root: results/H2O
run_name: run-H2O
seed: 123                                                                    # model seed
dataset_seed: 456                                                            # data set seed
append: true                                                                 # set true if a restarted run should append to the previous log file

# see https://arxiv.org/abs/2304.10061 for discussion of numerical precision
default_dtype: float64
# model_dtype: float32
# allow_tf32: true    # consider setting to false if you plan to mix training/inference over any devices that are not NVIDIA Ampere or later

# -- network --
# tell nequip which modules to build
model_builders:
 - allegro.model.Allegro
 # the typical model builders from `nequip` can still be used:
 - PerSpeciesRescale
 - StressForceOutput
 - RescaleEnergyEtc
 - initialize_from_state
initial_model_state: ./pretraining_output_files/results/H2O/run-H2O/best_model.pth
r_max: 6 # cutoff radius in length units, here Angstrom, this is an important hyperparamter to scan

# average number of neighbors in an environment is used to normalize the sum, auto precomputed it automitcally
avg_num_neighbors: auto

# radial basis
# set true to train the bessel roots
BesselBasis_trainable: true

# p-parameter in envelope function, as proposed in Klicpera, J. et al., arXiv:2003.03123
# sets it BOTH for the RadialBasisProjection AND the Allegro_Module
PolynomialCutoff_p: 12

l_max: 2                                                                     # the maximum irrep order (rotation order) for the network's features, l=1 is a good default, l=2 is more accurate but slower

# allowed: o3_full, o3_restricted, so3
parity: so3

# number of tensor product layers, 1-3 usually best, more is more accurate but slower
num_layers: 1

# number of features, more is more accurate but slower, 1, 4, 8, 16, 64, 128 are good options to try depending on data set
env_embed_multiplicity: 32

# whether or not to embed the initial edge, true often works best
embed_initial_edge: true

# hidden layer dimensions of the 2-body embedding MLP
# two_body_latent_mlp_latent_dimensions: [128, 256, 512, 1024]
two_body_latent_mlp_latent_dimensions: [32]
# nonlinearity used in the 2-body embedding MLP
two_body_latent_mlp_nonlinearity: silu
# weight initialization of the 2-body embedding MLP
two_body_latent_mlp_initialization: uniform

# hidden layer dimensions of the latent MLP
# these MLPs are cheap if you have have large l/env_embed_multiplicity, so a good place to put model capacity if you can afford it
# only if you are in the ultra-fast/scalable regime, make these smaller
# latent_mlp_latent_dimensions: [1024, 1024, 1024]latent_mlp_latent_dimensions: [64, 64]
latent_mlp_latent_dimensions: [32, 32, 32]
# nonlinearity used in the latent MLP
latent_mlp_nonlinearity: silu

# weight initialization of the latent MLP
latent_mlp_initialization: uniform

# whether to use a resnet update in the scalar latent latent space, true works best usually
latent_resnet: true

# hidden layer dimensions of the environment embedding mlp, none work best (will build a single linear layer)
env_embed_mlp_latent_dimensions: []

# nonlinearity used in the environment embedding mlp
env_embed_mlp_nonlinearity: null
```

```yaml

# weight initialzation of the environment embedding mlp
env_embed_mlp_initialization: uniform

# - end allegro layers -

# Final MLP to go from Allegro latent space to edge energies:

# hidden layer dimensions of the per-edge energy final MLP
edge_eng_mlp_latent_dimensions: [32]

# nonlinearity used in the per-edge energy final MLP
edge_eng_mlp_nonlinearity: null

# weight initialzation in the per-edge energy final MLP
edge_eng_mlp_initialization: uniform

# As an alternative option to npz, you can also pass data ase ASE Atoms-objects
# This can often be easier to work with, simply make sure the ASE Atoms object
# has a calculator for which atoms.get_potential_energy() and atoms.get_forces() are defined
dataset: ase
dataset_file_name: ./input_files/mlip_input_data/H2O_PD_with_stress.extxyz
                                                    # need to be a format accepted by ase.io.read
ase_args:
  format: extxyz
key_mapping:
  energy: total_energy
dataset_include_keys:
  - forces
  - energy
  - stress

# Alternatively, you may explicitly specify which chemical species in the input will map to NequIP atom type 0, which to atom type 1, and so on.
# Other than providing an explicit order for the NequIP atom types, this option behaves the same as `chemical_symbols`
chemical_symbol_to_type:
   O: 0
   H: 1

# logging
wandb: true                                                                     # we recommend using wandb for logging
wandb_project: H2O-PD-Simple-Model                                              # project name used in wandb
# wandb_watch: true

# see https://docs.wandb.ai/ref/python/watch
# wandb_watch_kwargs:
#   log: all
#   log_freq: 1
#   log_graph: true

verbose: info                                                                   # the same as python logging, e.g. warning, info, debug, error. case insensitive
# log_batch_freq: 100                                                             # batch frequency, how often to print training errors withinin the same epoch
# log_epoch_freq: 1                                                               # epoch frequency, how often to print
# save_checkpoint_freq: -1                                                        # frequency to save the intermediate checkpoint. no saving of intermediate checkpoints when the value is not positive.
# save_ema_checkpoint_freq: -1                                                    # frequency to save the intermediate ema checkpoint. no saving of intermediate checkpoints when the value is not positive.

# alternatively, n_train and n_val can be set as percentages of the dataset size:
n_train: 85%  # 70% of dataset
n_val: 14%    # 30% of dataset (if validation_dataset not set), or 30% of validation_dataset (if set)
learning_rate: 0.0005                                                           # learning rate, we found values between 0.01 and 0.005 to work best - this is often one of the most important hyperparameters to tune
batch_size: 1                                                                   # batch size, we found it important to keep this small for most applications including forces (1-5); for energy-only training, higher batch sizes work bettemax_epochs: 75
                                                                                # stop training after _ number of epochs, we set a very large number here, it won't take this long in practice and we will use early stopping instead
max_epochs: 20                                                                  # stop training after _ number of epochs, we set a very large number here, it won't take this long in practice and we will use early stopping instead
shuffle: true                                                                   # If true, the data loader will shuffle the data, usually a good idea
metrics_key: validation_loss # validation_loss                                  # metrics used for scheduling and saving best model. Options: `set`_`quantity`, set can be either "
```

```yaml
    train" or "validation, "quantity" can be loss or anything that appears in the validation batch step
    header, such as f_mae, f_rmse, e_mae, e_rmse
use_ema: true                                                                   # if true, use
    exponential moving average on weights for val/test, usually helps a lot with training, in particular
    for energy errors
ema_decay: 0.99                                                                 # ema weight, typically
     set to 0.99 or 0.999
ema_use_num_updates: true                                                       # whether to use number
     of updates when computing averages
# report_init_validation: true                                                    # if True, report the
     validation error for just initialized model

# loss function
loss_coeffs:                                                                    # different weights to
    use in a weighted loss functions
  forces: # if using PerAtomMSELoss, a default weight of 1:1 on each should work well
  - 1.0
  - PerSpeciesMSELoss
  stress: 50.0
  total_energy:
    - 20.0 # Change to 25 later for fine-tuning
    - PerAtomMSELoss
# note that the ratio between force and energy loss matters for the training process. One may consider
    using 1:1 with the PerAtomMSELoss. If the energy loss still significantly dominate the loss function
    at the initial epochs, tune the energy loss weight lower helps the training a lot.

# output metrics
metrics_components:
  - - forces                                # key
    - rmse                                   # "rmse" or "mae"
  - - forces
    - rmse
    - PerSpecies: True                      # if true, per species contribution is counted separately
      report_per_component: False           # if true, statistics on each component (i.e. fx, fy, fz) will
          be counted separately
  - - total_energy
    - rmse
  - - total_energy
    - rmse
    - PerAtom: True                         # if true, energy is normalized by the number of atoms
  - - stress
    - rmse

# optimizer, may be any optimizer defined in torch.optim
# the name `optimizer_name`is case sensitive
# IMPORTANT: for NequIP (not for Allegro), we find that in most cases AMSGrad strongly improves
# out-of-distribution generalization over Adam. We highly recommed trying both AMSGrad (by setting
# optimizer_amsgrad: true) and Adam (by setting optimizer_amsgrad: false)
optimizer_name: Adam
optimizer_amsgrad: false
optimizer_betas: !!python/tuple
  - 0.9
  - 0.999
optimizer_eps: 1.0e-08
optimizer_weight_decay: 0

# gradient clipping using torch.nn.utils.clip_grad_norm_
# see https://pytorch.org/docs/stable/generated/torch.nn.utils.clip_grad_norm_.html#torch.nn.utils.
    clip_grad_norm_
# setting to inf or null disables it
max_gradient_norm: null

# lr scheduler
# first: on-plateau, reduce lr by factory of lr_scheduler_factor if metrics_key hasn't improved for
    lr_scheduler_patience epoch
# you can also set other options of the underlying PyTorch scheduler, for example lr_scheduler_threshold
lr_scheduler_name: ReduceLROnPlateau
lr_scheduler_patience: 10
lr_scheduler_factor: 0.5

# early stopping based on metrics values.
# LR, wall and any keys printed in the log file can be used.
# The key can start with Training or validation. If not defined, the validation value will be used.
early_stopping_patiences:                                                       # stop early if a
    metric value stopped decreasing for n epochs
  validation_loss: 50
  # validation_f_rmse: 50
early_stopping_delta:                                                           # If delta is defined,
    a decrease smaller than delta will not be considered as a decrease
  validation_loss: 0.005
  # validation_f_rmse: 0.001

early_stopping_cumulative_delta: false                                          # If True, the minimum
    value recorded will not be updated when the decrease is smaller than delta
```

```
206
207 early_stopping_lower_bounds:                                    # stop early if a
        metric value is lower than the bound
208   LR: 1.0e-5
209
210 early_stopping_upper_bounds:                                    # stop early if a
        metric value is higher than the bound
211   cumulative_wall: 1.0e+100
```

Listing C.2: Full YAML training configuration used for Allegro finetuning

```
{
  "_comment1": " model parameters",
  "model": {
    "type_map": [
      "O",
      "H"
    ],
    "descriptor": {
      "type": "se_atten_v2",
      "sel": 120,
      "rcut_smth": 0.50,
      "rcut": 6.00,
      "neuron": [
        25,
        50,
        100
      ],
      "resnet_dt": false,
      "axis_neuron": 16,
      "seed": 1,
      "attn": 128,
      "attn_layer": 1,
      "attn_dotr": true,
      "attn_mask": false,
      "precision": "float64",
      "_comment2": " that's all"
    },
    "fitting_net": {
      "neuron": [
        240,
        240,
        240
      ],
      "resnet_dt": true,
      "precision": "float64",
      "seed": 1,
      "_comment3": " that's all"
    },
    "_comment4": " that's all"
  },

  "learning_rate": {
    "type": "exp",
    "decay_steps": 5000,
    "start_lr": 0.001,
    "stop_lr": 3.51e-8,
    "_comment5": "that's all"
  },

  "loss": {
    "type": "ener",
    "start_pref_e": 0.02,
    "limit_pref_e": 1,
    "start_pref_f": 1000,
    "limit_pref_f": 1,
    "start_pref_v": 0.02,
    "limit_pref_v": 1,
    "_comment6": " that's all"
  },

  "training": {
    "training_data": {
      "systems": [
        "H2O-Phase-Diagram/data/data.iters/iter.000025/02.fp/data.065/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000025/02.fp/data.069/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000025/02.fp/data.041/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.006/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.026/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.062/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.010/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.058/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.022/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.038/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.002/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.054/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.018/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.050/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.014/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.066/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.046/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000007/02.fp/data.035/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000007/02.fp/data.044/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000007/02.fp/data.031/train",
```

```
        "H2O-Phase-Diagram/data/data.iters/iter.000021/02.fp/data.065/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000021/02.fp/data.069/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000021/02.fp/data.041/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000015/02.fp/data.029/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000015/02.fp/data.033/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000015/02.fp/data.041/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000029/02.fp/data.065/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000029/02.fp/data.069/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000029/02.fp/data.041/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000010/02.fp/data.030/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000010/02.fp/data.034/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000010/02.fp/data.042/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000019/02.fp/data.035/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000019/02.fp/data.043/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000019/02.fp/data.031/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.036/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.020/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.056/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.004/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.000/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.052/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.016/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.048/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.012/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.064/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.044/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.008/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.060/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.024/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000028/02.fp/data.040/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000028/02.fp/data.064/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000028/02.fp/data.068/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000008/02.fp/data.040/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000008/02.fp/data.028/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000008/02.fp/data.032/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000003/02.fp/data.029/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000003/02.fp/data.033/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000003/02.fp/data.041/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.011/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.047/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.067/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.051/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.063/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.027/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.007/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.059/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.023/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.017/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.039/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.003/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.055/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.013/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000035/02.fp/data.071/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000035/02.fp/data.043/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000035/02.fp/data.067/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.006/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.026/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.062/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.010/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.058/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.022/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.038/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.002/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.054/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.018/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.050/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.066/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.046/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000027/02.fp/data.071/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000027/02.fp/data.043/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000027/02.fp/data.067/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000024/02.fp/data.040/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000024/02.fp/data.064/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000024/02.fp/data.068/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000017/02.fp/data.030/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000017/02.fp/data.034/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000017/02.fp/data.042/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000030/02.fp/data.066/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000030/02.fp/data.070/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000030/02.fp/data.042/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000011/02.fp/data.035/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000011/02.fp/data.043/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000011/02.fp/data.031/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000013/02.fp/data.040/train",
```

```
        "H2O-Phase-Diagram/data/data.iters/iter.000013/02.fp/data.028/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000013/02.fp/data.032/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.036/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.020/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.056/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.004/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.000/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.052/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.016/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.048/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.012/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.064/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.044/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.008/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.060/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.024/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000001/02.fp/data.028/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000001/02.fp/data.032/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000020/02.fp/data.064/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000020/02.fp/data.068/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000020/02.fp/data.040/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000032/02.fp/data.040/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000032/02.fp/data.064/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000032/02.fp/data.068/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.045/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.009/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.061/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.025/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.005/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.057/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.021/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.037/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.001/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.053/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.017/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.049/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.013/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.065/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000034/02.fp/data.066/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000034/02.fp/data.070/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000034/02.fp/data.042/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.025/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.061/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.009/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.045/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.005/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.057/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.021/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.037/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.001/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.053/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.017/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.049/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.013/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.065/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000009/02.fp/data.029/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000009/02.fp/data.033/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000009/02.fp/data.041/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000005/02.fp/data.030/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000005/02.fp/data.034/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000005/02.fp/data.042/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000031/02.fp/data.071/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000031/02.fp/data.043/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000031/02.fp/data.067/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000023/02.fp/data.071/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000023/02.fp/data.043/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000023/02.fp/data.067/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000033/02.fp/data.065/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000033/02.fp/data.069/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000033/02.fp/data.041/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.063/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.047/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.067/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.051/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.027/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.007/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.059/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.023/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.017/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.003/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.055/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.013/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000026/02.fp/data.066/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000026/02.fp/data.070/train",
```

```
        "H2O-Phase-Diagram/data/data.iters/iter.000026/02.fp/data.042/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000022/02.fp/data.066/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000022/02.fp/data.070/train",
        "H2O-Phase-Diagram/data/data.iters/iter.000022/02.fp/data.042/train",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0400.1.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0800.3.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0800.5.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0400.3.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1200.3.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0800.1.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.2000.1.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1600.5.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1200.1.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.2000.5.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1600.1.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.2000.3.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0400.5.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1600.3.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1200.5.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0200.3.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0400.1.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0200.5.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0400.3.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0200.1.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0400.5.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0400.1.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0800.3.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0800.5.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0400.3.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1200.3.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0800.1.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.2000.1.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1600.5.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1200.1.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.2000.5.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1600.1.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.2000.3.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0400.5.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1600.3.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1200.5.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data.init/init_system.012/train",
        "H2O-Phase-Diagram/data/data.init/init_system.002/train",
        "H2O-Phase-Diagram/data/data.init/init_system.011/train",
        "H2O-Phase-Diagram/data/data.init/init_system.008/train",
        "H2O-Phase-Diagram/data/data.init/init_system.013/train",
        "H2O-Phase-Diagram/data/data.init/init_system.007/train",
        "H2O-Phase-Diagram/data/data.init/init_system.016/train",
        "H2O-Phase-Diagram/data/data.init/init_system.005/train",
        "H2O-Phase-Diagram/data/data.init/init_system.009/train",
        "H2O-Phase-Diagram/data/data.init/init_system.006/train",
        "H2O-Phase-Diagram/data/data.init/init_system.000/train",
        "H2O-Phase-Diagram/data/data.init/init_system.014/train",
        "H2O-Phase-Diagram/data/data.init/init_system.003/train",
        "H2O-Phase-Diagram/data/data.init/init_system.010/train",
        "H2O-Phase-Diagram/data/data.init/init_system.001/train",
        "H2O-Phase-Diagram/data/data.init/init_system.004/train",
        "H2O-Phase-Diagram/data/data.init/init_system.017/train",
        "H2O-Phase-Diagram/data/data.init/init_system.015/train",
        "H2O-Phase-Diagram/data/data_051920/ice13/state.0050.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice13/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice05/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice05/state.0250.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice08/state.0150.5.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice08/state.0250.5.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice15/state.0050.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice15/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/liq/state.0400.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/liq/state.0300.1.00e00.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice12/state.0050.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice12/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice06/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice06/state.0250.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice09/state.0050.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice09/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice07/state.0150.5.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice07/state.0250.5.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice03/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice03/state.0250.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice14/state.0050.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice14/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice01h/state.0150.1.00e00.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice01h/state.0250.1.00e00.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice11/state.0150.1.00e00.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice11/state.0050.1.00e00.00/vasp/data/train",
```

```
        "H2O-Phase-Diagram/data/data_051920/ice04/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice02/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_051920/ice02/state.0250.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0100.1.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0200.3.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0200.5.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0100.5.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0300.3.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0300.5.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0200.1.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0300.1.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0100.3.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0100.1.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0200.3.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0200.5.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0100.5.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0300.3.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0300.5.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0200.1.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0300.1.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0100.3.00e05.00/deepmd/train",
        "H2O-Phase-Diagram/data/data_050820/ice13/state.0050.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice13/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice05/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice05/state.0250.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice08/state.0150.5.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice08/state.0250.5.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice15/state.0050.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice15/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/liq/state.0400.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/liq/state.0300.1.00e00.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice12/state.0050.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice12/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice06/state.0250.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice06/state.0150.5.00e03.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice06/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice09/state.0050.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice09/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice07/state.0150.5.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice07/state.0250.5.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice03/state.0250.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice03/state.0150.5.00e03.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice03/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice14/state.0050.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice14/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice01h/state.0150.1.00e00.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice01h/state.0250.1.00e00.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice11/state.0050.1.00e00.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice11/state.0150.1.00e00.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice04/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice02/state.0150.1.00e04.00/vasp/data/train",
        "H2O-Phase-Diagram/data/data_050820/ice02/state.0250.1.00e04.00/vasp/data/train"
      ],
      "batch_size": "auto",
      "_comment7": "that's all"
    },
    "validation_data": {
      "systems": [
        "H2O-Phase-Diagram/data/data.iters/iter.000025/02.fp/data.065/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000025/02.fp/data.069/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000025/02.fp/data.041/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.006/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.022/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.002/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.054/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.018/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.066/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000007/02.fp/data.035/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000007/02.fp/data.031/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000021/02.fp/data.065/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000021/02.fp/data.069/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000021/02.fp/data.041/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000015/02.fp/data.029/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000015/02.fp/data.033/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000015/02.fp/data.041/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000029/02.fp/data.065/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000029/02.fp/data.069/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000029/02.fp/data.041/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000010/02.fp/data.042/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000019/02.fp/data.035/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000019/02.fp/data.043/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000019/02.fp/data.031/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.036/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.020/valid",
```

```
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.056/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.004/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.000/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.052/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.016/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.048/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.012/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.064/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.044/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.008/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.060/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.024/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000028/02.fp/data.040/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000028/02.fp/data.064/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000028/02.fp/data.068/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000008/02.fp/data.040/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000008/02.fp/data.028/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000003/02.fp/data.029/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000003/02.fp/data.033/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000003/02.fp/data.041/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.047/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.067/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.051/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.063/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.027/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.007/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.023/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.017/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.039/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.003/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.055/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.013/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000035/02.fp/data.071/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000035/02.fp/data.043/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000035/02.fp/data.067/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.006/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.002/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.018/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.066/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.046/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000027/02.fp/data.071/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000027/02.fp/data.067/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000024/02.fp/data.064/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000017/02.fp/data.030/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000017/02.fp/data.042/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000030/02.fp/data.066/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000030/02.fp/data.070/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000030/02.fp/data.042/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000011/02.fp/data.043/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000013/02.fp/data.040/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000013/02.fp/data.028/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000013/02.fp/data.032/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.036/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.020/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.056/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.004/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.000/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.052/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.016/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.048/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.012/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.064/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.044/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.008/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.060/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.024/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000001/02.fp/data.028/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000001/02.fp/data.032/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000020/02.fp/data.064/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000020/02.fp/data.068/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000020/02.fp/data.040/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000032/02.fp/data.040/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000032/02.fp/data.064/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000032/02.fp/data.068/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.045/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.009/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.061/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.025/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.005/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.057/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.021/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.037/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.001/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.053/valid",
```

```
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.017/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.049/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.013/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.065/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000034/02.fp/data.066/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000034/02.fp/data.070/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000034/02.fp/data.042/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.025/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.061/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.009/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.045/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.005/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.057/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.021/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.037/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.001/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.053/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.017/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.049/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.013/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.065/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000009/02.fp/data.041/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000005/02.fp/data.034/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000005/02.fp/data.042/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000031/02.fp/data.071/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000031/02.fp/data.043/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000031/02.fp/data.067/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000023/02.fp/data.071/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000023/02.fp/data.043/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000023/02.fp/data.067/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000033/02.fp/data.065/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000033/02.fp/data.069/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000033/02.fp/data.041/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.067/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.003/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000026/02.fp/data.066/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000026/02.fp/data.070/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000026/02.fp/data.042/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000022/02.fp/data.066/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000022/02.fp/data.070/valid",
        "H2O-Phase-Diagram/data/data.iters/iter.000022/02.fp/data.042/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0400.1.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0800.3.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0800.5.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0400.3.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1200.3.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0800.1.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.2000.1.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1600.5.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1200.1.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.2000.5.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1600.1.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.2000.3.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0400.5.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1600.3.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1200.5.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0200.3.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0400.1.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0200.5.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0400.3.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0200.1.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0400.5.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0400.1.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0800.3.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0800.5.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0400.3.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1200.3.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0800.1.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.2000.1.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1600.5.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1200.1.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.2000.5.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1600.1.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.2000.3.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0400.5.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1600.3.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1200.5.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data.init/init_system.012/valid",
        "H2O-Phase-Diagram/data/data.init/init_system.002/valid",
        "H2O-Phase-Diagram/data/data.init/init_system.011/valid",
        "H2O-Phase-Diagram/data/data.init/init_system.008/valid",
        "H2O-Phase-Diagram/data/data.init/init_system.013/valid",
        "H2O-Phase-Diagram/data/data.init/init_system.007/valid",
        "H2O-Phase-Diagram/data/data.init/init_system.016/valid",
```

```
        "H2O-Phase-Diagram/data/data.init/init_system.005/valid",
        "H2O-Phase-Diagram/data/data.init/init_system.009/valid",
        "H2O-Phase-Diagram/data/data.init/init_system.006/valid",
        "H2O-Phase-Diagram/data/data.init/init_system.000/valid",
        "H2O-Phase-Diagram/data/data.init/init_system.014/valid",
        "H2O-Phase-Diagram/data/data.init/init_system.003/valid",
        "H2O-Phase-Diagram/data/data.init/init_system.010/valid",
        "H2O-Phase-Diagram/data/data.init/init_system.001/valid",
        "H2O-Phase-Diagram/data/data.init/init_system.004/valid",
        "H2O-Phase-Diagram/data/data.init/init_system.017/valid",
        "H2O-Phase-Diagram/data/data.init/init_system.015/valid",
        "H2O-Phase-Diagram/data/data_051920/ice13/state.0050.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice13/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice05/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice05/state.0250.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice08/state.0150.5.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice08/state.0250.5.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice15/state.0050.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice15/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/liq/state.0400.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/liq/state.0300.1.00e00.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice12/state.0050.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice12/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice06/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice06/state.0250.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice09/state.0050.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice09/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice07/state.0150.5.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice07/state.0250.5.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice03/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice03/state.0250.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice14/state.0050.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice14/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice01h/state.0150.1.00e00.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice01h/state.0250.1.00e00.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice11/state.0150.1.00e00.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice11/state.0050.1.00e00.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice04/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice02/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_051920/ice02/state.0250.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0100.1.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0200.3.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0200.5.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0300.3.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0200.1.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0300.1.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0100.1.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0200.3.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0200.5.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0300.3.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0300.5.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0200.1.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0100.3.00e05.00/deepmd/valid",
        "H2O-Phase-Diagram/data/data_050820/ice13/state.0050.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice13/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice05/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice05/state.0250.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice08/state.0150.5.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice08/state.0250.5.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice15/state.0050.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice15/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/liq/state.0400.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/liq/state.0300.1.00e00.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice12/state.0050.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice12/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice06/state.0250.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice06/state.0150.5.00e03.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice06/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice09/state.0050.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice09/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice07/state.0150.5.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice07/state.0250.5.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice03/state.0250.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice03/state.0150.5.00e03.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice03/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice14/state.0050.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice14/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice01h/state.0150.1.00e00.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice01h/state.0250.1.00e00.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice11/state.0050.1.00e00.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice11/state.0150.1.00e00.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice04/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice02/state.0150.1.00e04.00/vasp/data/valid",
        "H2O-Phase-Diagram/data/data_050820/ice02/state.0250.1.00e04.00/vasp/data/valid"
```

```
      ],
      "batch_size": "auto",
      "numb_btch": 1,
      "_comment8": "that's all"
    },
    "numb_steps": 1000000,
    "seed": 10,
    "disp_file": "lcurve.out",
    "disp_freq": 100,
    "save_freq": 1000,
    "_comment9": "that's all"
  },

  "_comment10": "that's all"
}
```

Listing C.3: Full YAML training configuration used for DeePMD DPA-1

```
{
    "_comment": "that's all",
    "model": {
        "shared_dict": {
            "type_map_all": [
                "O",
                "H"
            ],
            "dpa2_descriptor": {
                "type": "dpa2",
                "repinit": {
                    "tebd_dim": 8,
                    "rcut": 6.0,
                    "rcut_smth": 0.5,
                    "nsel": 120,
                    "neuron": [
                        25,
                        50,
                        100
                    ],
                    "axis_neuron": 12,
                    "activation_function": "tanh",
                    "three_body_sel": 40,
                    "three_body_rcut": 4.0,
                    "three_body_rcut_smth": 3.5,
                    "use_three_body": true
                },
                "repformer": {
                    "rcut": 4.0,
                    "rcut_smth": 3.5,
                    "nsel": 40,
                    "nlayers": 6,
                    "g1_dim": 128,
                    "g2_dim": 32,
                    "attn2_hidden": 32,
                    "attn2_nhead": 4,
                    "attn1_hidden": 128,
                    "attn1_nhead": 4,
                    "axis_neuron": 4,
                    "update_h2": false,
                    "update_g1_has_conv": true,
                    "update_g1_has_grrg": true,
                    "update_g1_has_drrd": true,
                    "update_g1_has_attn": false,
                    "update_g2_has_g1g1": false,
                    "update_g2_has_attn": true,
                    "update_style": "res_residual",
                    "update_residual": 0.01,
                    "update_residual_init": "norm",
                    "attn2_has_gate": true,
                    "use_sqrt_nnei": true,
                    "g1_out_conv": true,
                    "g1_out_mlp": true
                },
                "add_tebd_to_repinit_out": false
            },
            "_comment": "that's all"
        },
        "model_dict": {
            "H2O_H2O-PD": {
                "type_map": "type_map_all",
                "descriptor": "dpa2_descriptor",
                "fitting_net": {
                    "neuron": [
                        240,
                        240,
                        240
                    ],
                    "activation_function": "tanh",
                    "resnet_dt": true,
                    "seed": 1,
                    "_comment": " that's all"
                }
            },
            "H2O_H2O-DPLR": {
                "type_map": "type_map_all",
                "descriptor": "dpa2_descriptor",
                "fitting_net": {
                    "neuron": [
                        240,
                        240,
                        240
                    ],
```

```
                    "activation_function": "tanh",
                    "resnet_dt": true,
                    "seed": 1,
                    "_comment": " that's all"
                }
            },
            "H2O_H2O-SCAN0": {
                "type_map": "type_map_all",
                "descriptor": "dpa2_descriptor",
                "fitting_net": {
                    "neuron": [
                        240,
                        240,
                        240
                    ],
                    "activation_function": "tanh",
                    "resnet_dt": true,
                    "seed": 1,
                    "_comment": " that's all"
                }
            }
        }
    },
    "learning_rate": {
        "type": "exp",
        "decay_steps": 5000,
        "start_lr": 0.001,
        "stop_lr": 3.51e-8,
        "decay_rate": 0.98,
        "_comment": "that's all"
    },
    "loss_dict": {
        "H2O_H2O-PD": {
            "type": "ener",
            "start_pref_e": 0.02,
            "limit_pref_e": 1,
            "start_pref_f": 1000,
            "limit_pref_f": 1,
            "start_pref_v": 0.02,
            "limit_pref_v": 1
        },
        "H2O_H2O-DPLR": {
            "type": "ener",
            "start_pref_e": 0.02,
            "limit_pref_e": 1,
            "start_pref_f": 1000,
            "limit_pref_f": 1,
            "start_pref_v": 0.02,
            "limit_pref_v": 1
        },
        "H2O_H2O-SCAN0": {
            "type": "ener",
            "start_pref_e": 0.02,
            "limit_pref_e": 1,
            "start_pref_f": 1000,
            "limit_pref_f": 1,
            "start_pref_v": 0.02,
            "limit_pref_v": 1
        }
    },
    "training": {
        "model_prob": {
            "H2O_H2O-PD": 1.0,
            "H2O_H2O-DPLR": 0.1,
            "H2O_H2O-SCAN0": 1.0
        },
        "data_dict": {
            "H2O_H2O-PD": {
                "stat_file": "H2O_H2O-PD.hdf5",
                "training_data": {
                    "systems": [
                        "H2O-Phase-Diagram/data/data.iters/iter.000025/02.fp/data.065/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000025/02.fp/data.069/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000025/02.fp/data.041/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.006/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.026/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.062/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.010/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.058/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.022/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.038/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.002/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.054/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.018/train",
```

```
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.050/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.014/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.066/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.046/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000007/02.fp/data.035/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000007/02.fp/data.044/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000007/02.fp/data.031/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000021/02.fp/data.065/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000021/02.fp/data.069/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000021/02.fp/data.041/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000015/02.fp/data.029/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000015/02.fp/data.033/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000015/02.fp/data.041/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000029/02.fp/data.065/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000029/02.fp/data.069/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000029/02.fp/data.041/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000010/02.fp/data.030/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000010/02.fp/data.034/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000010/02.fp/data.042/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000019/02.fp/data.035/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000019/02.fp/data.043/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000019/02.fp/data.031/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.036/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.020/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.056/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.004/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.000/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.052/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.016/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.048/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.012/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.064/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.044/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.008/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.060/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.024/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000028/02.fp/data.040/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000028/02.fp/data.064/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000028/02.fp/data.068/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000008/02.fp/data.040/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000008/02.fp/data.028/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000008/02.fp/data.032/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000003/02.fp/data.029/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000003/02.fp/data.033/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000003/02.fp/data.041/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.011/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.047/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.067/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.051/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.063/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.027/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.007/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.059/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.023/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.017/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.039/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.003/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.055/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.013/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000035/02.fp/data.071/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000035/02.fp/data.043/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000035/02.fp/data.067/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.006/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.026/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.062/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.010/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.058/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.022/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.038/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.002/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.054/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.018/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.050/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.066/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.046/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000027/02.fp/data.071/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000027/02.fp/data.043/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000027/02.fp/data.067/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000024/02.fp/data.040/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000024/02.fp/data.064/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000024/02.fp/data.068/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000017/02.fp/data.030/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000017/02.fp/data.034/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000017/02.fp/data.042/train",
```

```
                    "H2O-Phase-Diagram/data/data.iters/iter.000030/02.fp/data.066/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000030/02.fp/data.070/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000030/02.fp/data.042/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000011/02.fp/data.035/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000011/02.fp/data.043/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000011/02.fp/data.031/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000013/02.fp/data.040/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000013/02.fp/data.028/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000013/02.fp/data.032/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.036/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.020/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.056/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.004/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.000/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.052/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.016/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.048/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.012/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.064/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.044/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.008/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.060/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.024/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000001/02.fp/data.028/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000001/02.fp/data.032/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000020/02.fp/data.064/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000020/02.fp/data.068/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000020/02.fp/data.040/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000032/02.fp/data.040/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000032/02.fp/data.064/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000032/02.fp/data.068/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.045/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.009/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.061/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.025/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.005/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.057/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.021/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.037/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.001/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.053/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.017/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.049/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.013/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.065/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000034/02.fp/data.066/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000034/02.fp/data.070/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000034/02.fp/data.042/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.025/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.061/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.009/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.045/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.005/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.057/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.021/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.037/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.001/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.053/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.017/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.049/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.013/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.065/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000009/02.fp/data.029/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000009/02.fp/data.033/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000009/02.fp/data.041/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000005/02.fp/data.030/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000005/02.fp/data.034/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000005/02.fp/data.042/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000031/02.fp/data.071/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000031/02.fp/data.043/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000031/02.fp/data.067/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000023/02.fp/data.071/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000023/02.fp/data.043/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000023/02.fp/data.067/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000033/02.fp/data.065/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000033/02.fp/data.069/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000033/02.fp/data.041/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.063/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.047/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.067/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.051/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.027/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.007/train",
                    "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.059/train",
```

```
                        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.023/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.017/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.003/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.055/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.013/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000026/02.fp/data.066/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000026/02.fp/data.070/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000026/02.fp/data.042/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000022/02.fp/data.066/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000022/02.fp/data.070/train",
                        "H2O-Phase-Diagram/data/data.iters/iter.000022/02.fp/data.042/train",
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0400.1.00e05.00/deepmd/train"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0800.3.00e05.00/deepmd/train"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0800.5.00e05.00/deepmd/train"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0400.3.00e05.00/deepmd/train"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1200.3.00e05.00/deepmd/train"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0800.1.00e05.00/deepmd/train"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.2000.1.00e05.00/deepmd/train"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1600.5.00e05.00/deepmd/train"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1200.1.00e05.00/deepmd/train"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.2000.5.00e05.00/deepmd/train"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1600.1.00e05.00/deepmd/train"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.2000.3.00e05.00/deepmd/train"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0400.5.00e05.00/deepmd/train"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1600.3.00e05.00/deepmd/train"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1200.5.00e05.00/deepmd/train"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0200.3.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0400.1.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0200.5.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0400.3.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0200.1.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0400.5.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0400.1.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0800.3.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0800.5.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0400.3.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1200.3.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0800.1.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.2000.1.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1600.5.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1200.1.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.2000.5.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1600.1.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.2000.3.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0400.5.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1600.3.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1200.5.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data.init/init_system.012/train",
```

```
                        "H2O-Phase-Diagram/data/data.init/init_system.002/train",
                        "H2O-Phase-Diagram/data/data.init/init_system.011/train",
                        "H2O-Phase-Diagram/data/data.init/init_system.008/train",
                        "H2O-Phase-Diagram/data/data.init/init_system.013/train",
                        "H2O-Phase-Diagram/data/data.init/init_system.007/train",
                        "H2O-Phase-Diagram/data/data.init/init_system.016/train",
                        "H2O-Phase-Diagram/data/data.init/init_system.005/train",
                        "H2O-Phase-Diagram/data/data.init/init_system.009/train",
                        "H2O-Phase-Diagram/data/data.init/init_system.006/train",
                        "H2O-Phase-Diagram/data/data.init/init_system.000/train",
                        "H2O-Phase-Diagram/data/data.init/init_system.014/train",
                        "H2O-Phase-Diagram/data/data.init/init_system.003/train",
                        "H2O-Phase-Diagram/data/data.init/init_system.010/train",
                        "H2O-Phase-Diagram/data/data.init/init_system.001/train",
                        "H2O-Phase-Diagram/data/data.init/init_system.004/train",
                        "H2O-Phase-Diagram/data/data.init/init_system.017/train",
                        "H2O-Phase-Diagram/data/data.init/init_system.015/train",
                        "H2O-Phase-Diagram/data/data_051920/ice13/state.0050.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice13/state.0150.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice05/state.0150.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice05/state.0250.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice08/state.0150.5.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice08/state.0250.5.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice15/state.0050.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice15/state.0150.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/liq/state.0400.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/liq/state.0300.1.00e00.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice12/state.0050.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice12/state.0150.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice06/state.0150.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice06/state.0250.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice09/state.0050.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice09/state.0150.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice07/state.0150.5.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice07/state.0250.5.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice03/state.0150.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice03/state.0250.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice14/state.0050.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice14/state.0150.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice01h/state.0150.1.00e00.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice01h/state.0250.1.00e00.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice11/state.0150.1.00e00.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice11/state.0050.1.00e00.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice04/state.0150.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice02/state.0150.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_051920/ice02/state.0250.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0100.1.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0200.3.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0200.5.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0100.5.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0300.3.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0300.5.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0200.1.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0300.1.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0100.3.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0100.1.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0200.3.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0200.5.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0100.5.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0300.3.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0300.5.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0200.1.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0300.1.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0100.3.00e05.00/deepmd/
                            train",
                        "H2O-Phase-Diagram/data/data_050820/ice13/state.0050.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice13/state.0150.1.00e04.00/vasp/data/train",
```

```
                        "H2O-Phase-Diagram/data/data_050820/ice05/state.0150.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice05/state.0250.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice08/state.0150.5.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice08/state.0250.5.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice15/state.0050.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice15/state.0150.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/liq/state.0400.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/liq/state.0300.1.00e00.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice12/state.0050.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice12/state.0150.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice06/state.0250.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice06/state.0150.5.00e03.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice06/state.0150.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice09/state.0050.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice09/state.0150.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice07/state.0150.5.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice07/state.0250.5.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice03/state.0250.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice03/state.0150.5.00e03.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice03/state.0150.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice14/state.0050.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice14/state.0150.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice01h/state.0150.1.00e00.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice01h/state.0250.1.00e00.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice11/state.0050.1.00e00.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice11/state.0150.1.00e00.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice04/state.0150.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice02/state.0150.1.00e04.00/vasp/data/train",
                        "H2O-Phase-Diagram/data/data_050820/ice02/state.0250.1.00e04.00/vasp/data/train"
                    ],
                    "batch_size": "auto",
                    "_comment": "that's all"
                },
                "validation_data": {
                    "systems": [
                        "H2O-Phase-Diagram/data/data.iters/iter.000025/02.fp/data.065/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000025/02.fp/data.069/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000025/02.fp/data.041/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.006/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.022/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.002/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.054/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.018/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000016/02.fp/data.066/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000007/02.fp/data.035/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000007/02.fp/data.031/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000021/02.fp/data.065/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000021/02.fp/data.069/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000021/02.fp/data.041/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000015/02.fp/data.029/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000015/02.fp/data.033/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000015/02.fp/data.041/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000029/02.fp/data.065/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000029/02.fp/data.069/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000029/02.fp/data.041/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000010/02.fp/data.042/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000019/02.fp/data.035/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000019/02.fp/data.043/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000019/02.fp/data.031/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.036/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.020/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.056/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.004/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.000/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.052/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.016/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.048/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.012/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.064/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.044/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.008/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.060/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000000/02.fp/data.024/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000028/02.fp/data.040/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000028/02.fp/data.064/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000028/02.fp/data.068/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000008/02.fp/data.040/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000008/02.fp/data.028/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000003/02.fp/data.029/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000003/02.fp/data.033/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000003/02.fp/data.041/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.047/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.067/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.051/valid",
```

```
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.063/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.027/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.007/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.023/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.017/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.039/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.003/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.055/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000018/02.fp/data.013/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000035/02.fp/data.071/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000035/02.fp/data.043/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000035/02.fp/data.067/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.006/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.002/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.018/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.066/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000004/02.fp/data.046/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000027/02.fp/data.071/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000027/02.fp/data.067/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000024/02.fp/data.064/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000017/02.fp/data.030/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000017/02.fp/data.042/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000030/02.fp/data.066/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000030/02.fp/data.070/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000030/02.fp/data.042/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000011/02.fp/data.043/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000013/02.fp/data.040/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000013/02.fp/data.028/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000013/02.fp/data.032/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.036/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.020/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.056/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.004/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.000/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.052/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.016/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.048/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.012/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.064/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.044/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.008/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.060/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000012/02.fp/data.024/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000001/02.fp/data.028/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000001/02.fp/data.032/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000020/02.fp/data.064/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000020/02.fp/data.068/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000020/02.fp/data.040/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000032/02.fp/data.040/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000032/02.fp/data.064/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000032/02.fp/data.068/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.045/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.009/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.061/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.025/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.005/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.057/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.021/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.037/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.001/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.053/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.017/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.049/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.013/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000002/02.fp/data.065/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000034/02.fp/data.066/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000034/02.fp/data.070/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000034/02.fp/data.042/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.025/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.061/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.009/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.045/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.005/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.057/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.021/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.037/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.001/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.053/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.017/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.049/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.013/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000014/02.fp/data.065/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000009/02.fp/data.041/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000005/02.fp/data.034/valid",
```

```
                        "H2O-Phase-Diagram/data/data.iters/iter.000005/02.fp/data.042/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000031/02.fp/data.071/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000031/02.fp/data.043/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000031/02.fp/data.067/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000023/02.fp/data.071/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000023/02.fp/data.043/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000023/02.fp/data.067/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000033/02.fp/data.065/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000033/02.fp/data.069/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000033/02.fp/data.041/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.067/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000006/02.fp/data.003/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000026/02.fp/data.066/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000026/02.fp/data.070/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000026/02.fp/data.042/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000022/02.fp/data.066/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000022/02.fp/data.070/valid",
                        "H2O-Phase-Diagram/data/data.iters/iter.000022/02.fp/data.042/valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0400.1.00e05.00/deepmd/valid"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0800.3.00e05.00/deepmd/valid"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0800.5.00e05.00/deepmd/valid"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0400.3.00e05.00/deepmd/valid"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1200.3.00e05.00/deepmd/valid"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0800.1.00e05.00/deepmd/valid"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.2000.1.00e05.00/deepmd/valid"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1600.5.00e05.00/deepmd/valid"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1200.1.00e05.00/deepmd/valid"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.2000.5.00e05.00/deepmd/valid"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1600.1.00e05.00/deepmd/valid"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.2000.3.00e05.00/deepmd/valid"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.0400.5.00e05.00/deepmd/valid"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1600.3.00e05.00/deepmd/valid"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/liq/state.1200.5.00e05.00/deepmd/valid"
                            ,
                        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0200.3.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0400.1.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0200.5.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0400.3.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0200.1.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice08/state.0400.5.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0400.1.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0800.3.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0800.5.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0400.3.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1200.3.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0800.1.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.2000.1.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1600.5.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1200.1.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.2000.5.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1600.1.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.2000.3.00e05.00/deepmd/
                            valid",
```

```
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.0400.5.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1600.3.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure/ice07/state.1200.5.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data.init/init_system.012/valid",
                        "H2O-Phase-Diagram/data/data.init/init_system.002/valid",
                        "H2O-Phase-Diagram/data/data.init/init_system.011/valid",
                        "H2O-Phase-Diagram/data/data.init/init_system.008/valid",
                        "H2O-Phase-Diagram/data/data.init/init_system.013/valid",
                        "H2O-Phase-Diagram/data/data.init/init_system.007/valid",
                        "H2O-Phase-Diagram/data/data.init/init_system.016/valid",
                        "H2O-Phase-Diagram/data/data.init/init_system.005/valid",
                        "H2O-Phase-Diagram/data/data.init/init_system.009/valid",
                        "H2O-Phase-Diagram/data/data.init/init_system.006/valid",
                        "H2O-Phase-Diagram/data/data.init/init_system.000/valid",
                        "H2O-Phase-Diagram/data/data.init/init_system.014/valid",
                        "H2O-Phase-Diagram/data/data.init/init_system.003/valid",
                        "H2O-Phase-Diagram/data/data.init/init_system.010/valid",
                        "H2O-Phase-Diagram/data/data.init/init_system.001/valid",
                        "H2O-Phase-Diagram/data/data.init/init_system.004/valid",
                        "H2O-Phase-Diagram/data/data.init/init_system.017/valid",
                        "H2O-Phase-Diagram/data/data.init/init_system.015/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice13/state.0050.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice13/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice05/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice05/state.0250.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice08/state.0150.5.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice08/state.0250.5.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice15/state.0050.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice15/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/liq/state.0400.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/liq/state.0300.1.00e00.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice12/state.0050.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice12/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice06/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice06/state.0250.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice09/state.0050.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice09/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice07/state.0150.5.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice07/state.0250.5.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice03/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice03/state.0250.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice14/state.0050.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice14/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice01h/state.0150.1.00e00.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice01h/state.0250.1.00e00.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice11/state.0150.1.00e00.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice11/state.0050.1.00e00.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice04/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice02/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_051920/ice02/state.0250.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0100.1.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0200.3.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0200.5.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0300.3.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0200.1.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice07/state.0300.1.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0100.1.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0200.3.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0200.5.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0300.3.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0300.5.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0200.1.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_high_pressure2/ice08/state.0100.3.00e05.00/deepmd/
                            valid",
                        "H2O-Phase-Diagram/data/data_050820/ice13/state.0050.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice13/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice05/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice05/state.0250.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice08/state.0150.5.00e04.00/vasp/data/valid",
```

```
                        "H2O-Phase-Diagram/data/data_050820/ice08/state.0250.5.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice15/state.0050.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice15/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/liq/state.0400.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/liq/state.0300.1.00e00.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice12/state.0050.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice12/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice06/state.0250.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice06/state.0150.5.00e03.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice06/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice09/state.0050.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice09/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice07/state.0150.5.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice07/state.0250.5.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice03/state.0250.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice03/state.0150.5.00e03.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice03/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice14/state.0050.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice14/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice01h/state.0150.1.00e00.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice01h/state.0250.1.00e00.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice11/state.0050.1.00e00.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice11/state.0150.1.00e00.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice04/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice02/state.0150.1.00e04.00/vasp/data/valid",
                        "H2O-Phase-Diagram/data/data_050820/ice02/state.0250.1.00e04.00/vasp/data/valid"
                    ],
                    "batch_size": 1,
                    "_comment": "that's all"
                }
            },
            "H2O_H2O-DPLR": {
                "stat_file": "H2O_H2O-DPLR.hdf5",
                "training_data": {
                    "systems": [
                        "H2O-DPLR/data/water/dpgen.lowd.merged/iter.000016/02.fp/data.016/train",
                        "H2O-DPLR/data/water/dpgen.lowd.merged/iter.000016/02.fp/data.012/train",
                        "H2O-DPLR/data/water/dpgen.lowd.merged/iter.000016/02.fp/data.020/train",
                        "H2O-DPLR/data/water/dpgen.merged/iter.000000/02.fp/data.000/train",
                        "H2O-DPLR/data/water/dpgen.merged/iter.000010/02.fp/data.006/train",
                        "H2O-DPLR/data/water/dpgen.merged/iter.000009/02.fp/data.005/train",
                        "H2O-DPLR/data/water/dpgen.merged/iter.000001/02.fp/data.001/train",
                        "H2O-DPLR/data/water/dpgen.merged/iter.000008/02.fp/data.004/train",
                        "H2O-DPLR/data/water/dpgen.merged/init/01.liq/state.0400.1.00e04.00/vasp/data/
                            train",
                        "H2O-DPLR/data/water/dpgen.merged/init/01.liq/state.0300.1.00e00.00/vasp/data/
                            train",
                        "H2O-DPLR/data/water/dpgen.merged/init/00.liq/state.0400.1.00e04.00/vasp/data/
                            train",
                        "H2O-DPLR/data/water/dpgen.merged/init/00.liq/state.0300.1.00e00.00/vasp/data/
                            train",
                        "H2O-DPLR/data/water/dpgen.merged/iter.000005/02.fp/data.001/train",
                        "H2O-DPLR/data/water/dpgen.merged/iter.000002/02.fp/data.002/train"
                    ],
                    "batch_size": "auto",
                    "_comment": "that's all"
                },
                "validation_data": {
                    "systems": [
                        "H2O-DPLR/data/water/dpgen.merged/iter.000000/02.fp/data.000/test",
                        "H2O-DPLR/data/water/dpgen.merged/iter.000010/02.fp/data.006/test",
                        "H2O-DPLR/data/water/dpgen.merged/iter.000001/02.fp/data.001/test",
                        "H2O-DPLR/data/water/dpgen.merged/iter.000008/02.fp/data.004/test",
                        "H2O-DPLR/data/water/dpgen.merged/init/01.liq/state.0400.1.00e04.00/vasp/data/test
                            ",
                        "H2O-DPLR/data/water/dpgen.merged/init/01.liq/state.0300.1.00e00.00/vasp/data/test
                            ",
                        "H2O-DPLR/data/water/dpgen.merged/init/00.liq/state.0400.1.00e04.00/vasp/data/test
                            ",
                        "H2O-DPLR/data/water/dpgen.merged/init/00.liq/state.0300.1.00e00.00/vasp/data/test
                            ",
                        "H2O-DPLR/data/water/dpgen.merged/iter.000005/02.fp/data.001/test",
                        "H2O-DPLR/data/water/dpgen.merged/iter.000002/02.fp/data.002/test",
                        "H2O-DPLR/data/data.test/slab.md/md.slab.12/deepmd/test",
                        "H2O-DPLR/data/data.test/slab.md/md.slab.07/deepmd/test"
                    ],
                    "batch_size": 1,
                    "_comment": "that's all"
                }
            },
            "H2O_H2O-SCAN0": {
                "stat_file": "H2O_H2O-SCAN0.hdf5",
                "training_data": {
                    "systems": [
```

```
                        "H2O-scan0/data/data42/train",
                        "H2O-scan0/data/data35/train",
                        "H2O-scan0/data/data5/train",
                        "H2O-scan0/data/data_ex6/train",
                        "H2O-scan0/data/data26/train",
                        "H2O-scan0/data/data7/train",
                        "H2O-scan0/data/data2/train",
                        "H2O-scan0/data/data21/train",
                        "H2O-scan0/data/data33/train",
                        "H2O-scan0/data/data17/train",
                        "H2O-scan0/data/data12/train",
                        "H2O-scan0/data/data9/train",
                        "H2O-scan0/data/data_ex2/train",
                        "H2O-scan0/data/data4/train",
                        "H2O-scan0/data/data37/train",
                        "H2O-scan0/data/data32/train",
                        "H2O-scan0/data/data44/train",
                        "H2O-scan0/data/data28/train",
                        "H2O-scan0/data/data23/train",
                        "H2O-scan0/data/data_ex4/train",
                        "H2O-scan0/data/data19/train",
                        "H2O-scan0/data/data30/train",
                        "H2O-scan0/data/data14/train",
                        "H2O-scan0/data/data_ex3/train",
                        "H2O-scan0/data/data43/train",
                        "H2O-scan0/data/data27/train",
                        "H2O-scan0/data/data39/train",
                        "H2O-scan0/data/data34/train",
                        "H2O-scan0/data/data18/train",
                        "H2O-scan0/data/data41/train",
                        "H2O-scan0/data/data13/train",
                        "H2O-scan0/data/data25/train",
                        "H2O-scan0/data/data_ex5/train",
                        "H2O-scan0/data/data20/train",
                        "H2O-scan0/data/data38/train",
                        "H2O-scan0/data/data16/train",
                        "H2O-scan0/data/data6/train",
                        "H2O-scan0/data/data45/train",
                        "H2O-scan0/data/data1/train",
                        "H2O-scan0/data/data11/train",
                        "H2O-scan0/data/data29/train",
                        "H2O-scan0/data/data40/train",
                        "H2O-scan0/data/data24/train",
                        "H2O-scan0/data/data36/train",
                        "H2O-scan0/data/data8/train",
                        "H2O-scan0/data/data3/train",
                        "H2O-scan0/data/data31/train",
                        "H2O-scan0/data/data15/train",
                        "H2O-scan0/data/data_ex1/train",
                        "H2O-scan0/data/data10/train",
                        "H2O-scan0/data/data22/train"
                    ],
                    "batch_size": "auto",
                    "_comment": "that's all"
                },
                "validation_data": {
                    "systems": [
                        "H2O-scan0/data/data42/valid",
                        "H2O-scan0/data/data35/valid",
                        "H2O-scan0/data/data5/valid",
                        "H2O-scan0/data/data_ex6/valid",
                        "H2O-scan0/data/data26/valid",
                        "H2O-scan0/data/data7/valid",
                        "H2O-scan0/data/data2/valid",
                        "H2O-scan0/data/data21/valid",
                        "H2O-scan0/data/data33/valid",
                        "H2O-scan0/data/data17/valid",
                        "H2O-scan0/data/data12/valid",
                        "H2O-scan0/data/data9/valid",
                        "H2O-scan0/data/data_ex2/valid",
                        "H2O-scan0/data/data4/valid",
                        "H2O-scan0/data/data37/valid",
                        "H2O-scan0/data/data32/valid",
                        "H2O-scan0/data/data44/valid",
                        "H2O-scan0/data/data28/valid",
                        "H2O-scan0/data/data23/valid",
                        "H2O-scan0/data/data_ex4/valid",
                        "H2O-scan0/data/data19/valid",
                        "H2O-scan0/data/data30/valid",
                        "H2O-scan0/data/data14/valid",
                        "H2O-scan0/data/data_ex3/valid",
                        "H2O-scan0/data/data43/valid",
                        "H2O-scan0/data/data27/valid",
                        "H2O-scan0/data/data39/valid",
```

```
                        "H2O-scan0/data/data34/valid",
                        "H2O-scan0/data/data18/valid",
                        "H2O-scan0/data/data41/valid",
                        "H2O-scan0/data/data13/valid",
                        "H2O-scan0/data/data25/valid",
                        "H2O-scan0/data/data_ex5/valid",
                        "H2O-scan0/data/data20/valid",
                        "H2O-scan0/data/data38/valid",
                        "H2O-scan0/data/data16/valid",
                        "H2O-scan0/data/data6/valid",
                        "H2O-scan0/data/data45/valid",
                        "H2O-scan0/data/data1/valid",
                        "H2O-scan0/data/data11/valid",
                        "H2O-scan0/data/data29/valid",
                        "H2O-scan0/data/data40/valid",
                        "H2O-scan0/data/data24/valid",
                        "H2O-scan0/data/data36/valid",
                        "H2O-scan0/data/data8/valid",
                        "H2O-scan0/data/data3/valid",
                        "H2O-scan0/data/data31/valid",
                        "H2O-scan0/data/data15/valid",
                        "H2O-scan0/data/data_ex1/valid",
                        "H2O-scan0/data/data10/valid",
                        "H2O-scan0/data/data22/valid"
                    ],
                    "batch_size": 1,
                    "_comment": "that's all"
                }
            }
        },
        "numb_steps": 1000000,
        "warmup_steps": 50000,
        "gradient_max_norm": 5.0,
        "seed": 10,
        "disp_file": "lcurve.out",
        "disp_freq": 2000,
        "save_freq": 2000,
        "_comment": "that's all"
    }
}
```

Listing C.4: Full YAML training configuration used for DeePMD DPA-2 Water

D

MLIP Software Packages

Software Environment for DeepMD DPA-1

The training environment consisted of a Python pip environment and HPC modules provided via Spack. Below is a full list of installed packages and loaded modules.

```
Package            Version
------------------ ---------
annotated-types    0.7.0
appdirs            1.4.4
array_api_compat   1.11.1
ase                3.25.0
bracex             2.5.post1
certifi            2023.7.22
cffi               1.15.1
charset-normalizer 3.3.0
click              8.1.7
colorama           0.4.6
contourpy          1.0.7
cycler             0.11.0
dargs              0.4.10
deepmd-kit         3.0.2
Deprecated         1.2.18
docker-pycreds     0.4.0
fonttools          4.39.4
future             0.18.3
gitdb              4.0.9
GitPython          3.1.27
greenlet           3.1.1
h5py               3.13.0
idna               3.4
joblib             1.5.1
kiwisolver         1.4.5
matplotlib         3.7.1
mendeleev          0.20.1
ml_dtypes          0.5.1
mpmath             1.2.1
numpy              1.24.1
packaging          23.1
pathtools          0.1.2
Pillow             10.0.0
pip                23.0.1
protobuf           3.20.3
psutil             5.9.5
pybind11           2.11.0
pycparser          2.21
pydantic           2.10.6
pydantic_core      2.27.2
pyfiglet           0.8.post1
Pygments           2.19.1
pyparsing          3.0.9
python-dateutil    2.8.2
PyYAML             6.0
requests           2.31.0
scikit-learn       1.6.1
scipy              1.11.3
sentry-sdk         1.5.5
setproctitle       1.1.10
```

```
setuptools          63.4.3
six                 1.16.0
smmap               5.0.0
SQLAlchemy          2.0.39
sympy               1.11.1
threadpoolctl       3.6.0
torch               1.11.0
tqdm                4.66.1
typeguard           4.4.2
typing_extensions   4.8.0
urllib3             2.0.6
wandb               0.13.9
wcmatch             10.0
wrapt               1.17.2

Loaded HPC Modules:
2024r1
miniconda3
openmpi/4.1.6
py-torch/1.11.0
py-scipy/1.11.3
py-kiwisolver/1.4.5
py-sympy/1.11.1
py-mpmath/1.2.1
py-pillow/10.0.0
py-packaging/23.0
py-matplotlib/3.7.1
py-cycler/0.11.0
py-contourpy/1.0.7
py-fonttools/4.39.4
py-wandb/0.13.9
cmake/3.27.7
fftw/3.3.10_openmp_True
```

Listing D.1: Complete software environment including pip packages and loaded modules

Software Environment for DeepMD DPA-2

```
Package             Version
------------------- ---------
annotated-types     0.7.0
appdirs             1.4.4
array_api_compat    1.11.1
ase                 3.25.0
bracex              2.5.post1
certifi             2023.7.22
cffi                1.15.1
charset-normalizer  3.3.0
click               8.1.7
colorama            0.4.6
contourpy           1.0.7
cycler              0.11.0
dargs               0.4.10
deepmd-kit          3.0.0b4
Deprecated          1.2.18
docker-pycreds      0.4.0
fonttools           4.39.4
future              0.18.3
gitdb               4.0.9
GitPython           3.1.27
greenlet            3.1.1
h5py                3.13.0
idna                3.4
joblib              1.5.1
kiwisolver          1.4.5
matplotlib          3.7.1
mendeleev           0.20.1
ml_dtypes           0.5.1
mpmath              1.2.1
numpy               1.24.1
packaging           23.1
pathtools           0.1.2
Pillow              10.0.0
pip                 23.0.1
protobuf            3.20.3
psutil              5.9.5
pybind11            2.11.0
pycparser           2.21
pydantic            2.10.6
pydantic_core       2.27.2
pyfiglet            0.8.post1
Pygments            2.19.1
pyparsing           3.0.9
python-dateutil     2.8.2
PyYAML              6.0
requests            2.31.0
scikit-learn        1.7.0
scipy               1.11.3
sentry-sdk          1.5.5
setproctitle        1.1.10
setuptools          63.4.3
six                 1.16.0
smmap               5.0.0
SQLAlchemy          2.0.39
sympy               1.11.1
threadpoolctl       3.6.0
torch               1.11.0
tqdm                4.66.1
typeguard           4.4.2
typing_extensions   4.8.0
urllib3             2.0.6
wandb               0.13.9
wcmatch             10.0
wrapt               1.17.2

Loaded HPC Modules:
2024r1
miniconda3
openmpi/4.1.6
py-torch/1.11.0
py-scipy/1.11.3
py-kiwisolver/1.4.5
py-sympy/1.11.1
py-mpmath/1.2.1
py-pillow/10.0.0
py-packaging/23.0
py-matplotlib/3.7.1
py-cycler/0.11.0
py-contourpy/1.0.7
py-fonttools/4.39.4
```

```
py-wandb/0.13.9
cmake/3.27.7
fftw/3.3.10_openmp_True
```

Listing D.2: Software environment for DeepMD DPA-2

Allegro Software Environment

The Allegro model was trained using a custom Conda environment with the following key packages:

```
ase               3.24.0
e3nn              0.5.5
mir-allegro       0.2.0
mkl-include       2025.0.0
nequip            0.6.1
opt-einsum        3.4.0
opt-einsum-fx     0.1.4
pip               25.0
python            3.10.16
setuptools        75.8.0
torch-ema         0.3
torch-runstats    0.2.0
wheel             0.45.1
```

Listing D.3: Installed packages in the Conda environment `allegro_lammps`

The following HPC modules were loaded during training:

```
module use ~/software/spack/share/spack/lmod/linux-rhel8-x86_64/Core
module use ~/software/spack/share/spack/lmod/linux-rhel8-x86_64/openmpi/4.1.6-h2uag4k/Core

module load 2024r1
module load miniconda3
module load openmpi/4.1.6
module load py-torch/1.11.0
module load py-scipy/1.11.3
module load py-kiwisolver/1.4.5
module load py-sympy/1.11.1
module load py-mpmath/1.2.1
module load py-pillow/10.0.0
module load py-packaging/23.0
module load py-matplotlib/3.7.1
module load py-cycler/0.11.0
module load py-contourpy/1.0.7
module load py-fonttools/4.39.4
module load py-wandb/0.13.9
module load cmake/3.27.7
module load fftw/3.3.10_openmp_True
```

Listing D.4: Modules loaded in the HPC environment