Freie Universität Berlin

A fast implementation of deep neural-network potentials for molecular dynamics simulations of alloys

by

Sandip Kumar Sah

Master Thesis in Computational Science

Submission: 09 January 2026 Supervisor: Prof. Dr. Felix Höfling

Freie Universität Berlin | Department of Mathematics and Computer Science
Institute of Mathematics

Statutory Declaration

Family Name, Given/First Name	Sah, Sandip Kumar
Matriculation number	5589263
Kind of thesis submitted	Master Thesis

English: Declaration of Authorship

I hereby declare that the thesis submitted was created and written solely by myself without any external support. Any sources, direct or indirect, are marked as such. I am aware of the fact that the contents of the thesis in digital form may be revised with regard to usage of unauthorized aid as well as whether the whole or parts of it may be identified as plagiarism. I do agree my work to be entered into a database for it to be compared with existing sources, where it will remain in order to enable further comparisons with future theses. This does not grant any rights of reproduction and usage, however.

This document was neither presented to any other examination board nor has it been published.

German: Erklärung der Autorenschaft (Urheberschaft)

Ich erkläre hiermit, dass die vorliegende Arbeit ohne fremde Hilfe ausschließlich von mir erstellt und geschrieben worden ist. Jedwede verwendeten Quellen, direkter oder indirekter Art, sind als solche kenntlich gemacht worden. Mir ist die Tatsache bewusst, dass der Inhalt der Thesis in digitaler Form geprüft werden kann im Hinblick darauf, ob es sich ganz oder in Teilen um ein Plagiat handelt. Ich bin damit einverstanden, dass meine Arbeit in einer Datenbank eingegeben werden kann, um mit bereits bestehenden Quellen verglichen zu werden und dort auch verbleibt, um mit zukünftigen Arbeiten verglichen werden zu können. Dies berechtigt jedoch nicht zur Verwendung oder Vervielfältigung.

Diese Arbeit wurde noch keiner anderen Prüfungsbehörde vorgelegt noch wurde sie bisher veröffentlicht.

.....

Date, Signature

Abstract

A concise 200–300 word summary of:

- Motivation (accelerating atomistic simulations using neural network potentials)
- Goal (integrating DeepMD potential calculation into HALMD)
- Method (extracting weights, replicating inference in C++/CUDA)
- Key results and performance (accuracy vs. speed trade-off)
- Conclusions

Contents

1 Introduction

1.1 Motivation

Molecular dynamics (MD) simulations play a central role in materials science by providing atomistic insight into the behavior of complex systems. Their predictive power depends on the interatomic potential used to approximate the underlying potential energy surface (PES). Classical empirical potentials are efficient but often too rigid to describe complex bonding environments, whereas *ab initio* methods offer high accuracy at prohibitive computational cost. Machine-learned interatomic potentials—particularly the Deep Potential Molecular Dynamics (DeepMD) framework—bridge this gap by achieving near–quantum mechanical accuracy with classical-MD performance.

The work by Andrés Cruz, titled *"Deep Neural Networks Potentials for Scalable Molecular Dynamics Simulations on Accelerator Hardware"* [1], represents an important step toward integrating Deep Potential models into the high-performance GPU-accelerated simulation engine HALMD. His work reconstructs the DeePMD-kit inference pipeline inside HALMD, extracts network weights from a trained TensorFlow model, and validates energy and force predictions for a **single-species Copper system**. In particular, his implementation focuses on the **two-body embedding DeepPot-SE descriptor** and reproduces the filter and fitting networks **only for a monoatomic system**, making it suitable for single metal.

However, the implementation in [1] remains limited to the simplest case of DeepMD-v2 architectures. It does not include multi-species support or the full range of descriptor and network features available in DeepMD-kit v2.

Additionally, several intermediate steps present in the full DeepMD-v2 computational graph—such as normalization layers, ghost body extension for periodicity in boundary condition, species-wise descriptor partitioning, and certain derivative pathways—were simplified or omitted in the previous implementation. Cruz's work successfully established a working single-species DeepMD integration within HALMD.

This work builds directly on the work of Cruz [1] and advances it substantially. The primary contributions of the present work are:

1. Generalizing the HALMD Deep Potential integration to **multi-species, multi-body DeepMD-v2 models**, enabling simulations of binary and multicomponent alloy systems.
2. Reconstructing descriptor and fitting networks for the **full multi-species DeepPot-SE architecture**, including multi-body descriptor terms.
3. Improving the **accuracy** of the HALMD implementation by adding several computational steps missing in the previous work, such as proper cutoff normalization, multi-species descriptor algebra, and the full force backpropagation pipeline.

Through these extensions, the present thesis transforms HALMD from supporting a prototype single-component DeepMD potential into a **fully general, high-accuracy, multi-species DeepMD-v2 engine**. This significantly broadens HALMD's applicability, enabling large-scale molecular dynamics simulations of technologically relevant multicomponent materials at near–quantum mechanical accuracy.

1.2 Objectives and Scope

The objective of this thesis is to extend and generalize the Deep Potential (DP) implementation in HALMD beyond the single-species, two-body descriptor developed by Cruz [1]. While Cruz's work successfully demonstrated that HALMD can evaluate a DeePMD-v2 model for a monoatomic copper system, several components required for full multi-species DeepMD-v2 inference were missing. Most notably, the previous implementation did not include (i) the periodic coordinate extension and neighbor-list construction used in DeepMD [2], (ii) normalization and scaling layers defined in the DP-v2 framework [3], (iii) species-dependent descriptors, or (iv) descriptor and filter weights that depend simultaneously on the central and neighbor species, as introduced in the multi-species DeepPot-SE descriptor [3].

This thesis addresses these limitations by implementing the complete multi-species DeepMD-v2 inference pipeline, including accurate periodic handling of atomic environments. Specifically, the thesis pursues the following objectives:

1. **Implement coordinate normalization, ghost-cell extension, and multi-type neighbor list construction.** The previous implementation did not include the canonical DeePMD preprocessing steps for periodic systems—wrapping coordinates into the primary simulation cell, generating ghost atoms to cover the cutoff radius, and constructing species-grouped neighbor lists—as described in the DeePMD methodology [2, 3]. Implementing these steps ensures that HALMD reproduces the correct local environments required by the DeepPot-SE descriptors under periodic boundary conditions.

2. **Generalize the descriptor pipeline to multi-species systems.** This includes implementing species-dependent neighbor counts $N_c(a)$, species-aware embedding matrices G_i, and the full multi-species DeepPot-SE (se_e2_a) descriptor introduced in DP-v2 [3].

3. **Implement species-dependent filter networks.** Extend the single-species embedding and filter networks used in [1] to support arbitrary numbers of atomic types, following the species-indexed filter network formulation defined in the DP-v2 architecture [3].

4. **Reproduce the full DeepMD-v2 inference procedure with higher accuracy.** Incorporate several computational steps omitted in previous work, including normalization layers, descriptor scaling operations, smooth cutoff functions, and full force backpropagation through descriptor derivatives, consistent with the DeepMD-v2 formulation [3, 2].

5. **Validate energy and force predictions for multi-component systems.** Compare results from HALMD against the DeePMD-kit reference implementation for multi-species models and quantify numerical deviations, following standard validation procedures established in DeepMD literature [2, 3].

Scope. This thesis focuses exclusively on the inference stage of DeepMD-v2, i.e., the computation of energies and forces from pre-trained models. Model training is outside the scope of this work. The implementation targets the multi-species DeepPot-SE descriptor and does not cover other descriptor families such as end-to-end or message-passing potentials. The work provides support for multi-species atomic systems relevant to alloys and chemically complex materials, while the underlying molecular dynamics algorithms

(integrators, thermostats, barostats) rely on HALMD's existing infrastructure.

2 Background

2.1 HALMD Software

The High-Accuracy Large-scale Molecular Dynamics (HALMD) package is a high-performance, open-source molecular dynamics (MD) simulation framework designed to study the microscopic dynamics of liquids, glasses, and other condensed matter systems. HALMD is built around a modular C++ architecture with a strong emphasis on numerical precision, extensibility, and efficient parallel computation on graphics processing units (GPUs) [4]. It provides a versatile platform for conducting large-scale classical molecular dynamics simulations involving millions of particles.

HALMD employs the *classical molecular dynamics* approach, in which atomic motion is governed by Newton's equations of motion,

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = -\nabla_i U(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N),$$

where U is a predefined analytical potential describing interatomic interactions. The accuracy of a classical MD simulation depends on the fidelity of this potential. HALMD supports several built-in empirical pair potentials, such as Lennard–Jones, Gaussian Core, and Yukawa models, and additionally allows users to supply tabulated pair potentials. These empirical models do not involve quantum-mechanical calculations and thus differ from *ab initio* molecular dynamics (AIMD), where forces are computed directly from electronic structure methods.

One of HALMD's defining strengths is its high-performance GPU backend. The software employs NVIDIA's CUDA to accelerate computationally intensive tasks such as force evaluation, neighbor-list construction, and time integration [4], and more recent versions extend support to heterogeneous computing platforms through SYCL [5]. Spatial domain decomposition and cell binning enable efficient scaling on multi-GPU systems, achieving orders-of-magnitude speedups compared to CPU-only implementations. HALMD also supports mixed-precision arithmetic, selectively applying double precision to critical operations to ensure long-term numerical stability without compromising performance.

HALMD follows a modular design philosophy. Core components—integrators, interaction potentials, neighbor-list builders, and observables—are implemented as interchangeable modules configurable through a Lua scripting interface [6]. This modularity makes it straightforward to extend HALMD with new functionality, including custom potentials, analysis routines, and data exporters. Simulation data are stored in the H5MD format [7], an HDF5-based standard widely supported by analysis tools in computational physics and chemistry.

These characteristics make HALMD an ideal platform for integrating machine-learned interatomic potentials. Its GPU-accelerated, modular architecture provides the necessary infrastructure to replace conventional analytical potentials with neural-network-based force fields. In this thesis, HALMD serves as the host MD engine into which the Deep Potential Molecular Dynamics (DeepMD) model is embedded. The goal is to achieve *ab initio*-level accuracy within HALMD's efficient large-scale simulation framework by implementing a full multi-species DeepMD-v2 inference pipeline.

2.2 Neural Network Potentials

In recent years, machine learning techniques—particularly deep neural networks (DNNs)—have revolutionized the construction of interatomic potentials for molecular dynamics simulations. Traditional analytical potentials, while computationally efficient, are often limited in their ability to accurately capture many-body interactions and chemical complexities across diverse atomic environments. Neural Network Potentials (NNPs) overcome these limitations by learning the potential energy surface (PES) directly from *ab initio* reference data [8, 9].

In the NNP framework, a neural network is trained to approximate the mapping between an atomic configuration and its corresponding total energy and atomic forces. The total potential energy of a system is commonly expressed as a sum of atomic contributions [8]:

$$E = \sum_i E_i(\mathcal{R}_i),$$

where E_i denotes the atomic energy associated with atom i, and $\mathcal{R}_i$ represents its local environment within a cutoff radius. This decomposition ensures extensivity and locality and enables efficient scaling with system size.

Training a neural network potential involves minimizing a loss function that combines the errors in energies, forces, and optionally virials. Following the formulation used in DeePMD-kit [2, 3], the loss is written as

$$\mathcal{L} = p_E L_E + p_F L_F + p_V L_V, \tag{1}$$

where the terms are defined as

$$L_E = \frac{1}{N_E} \sum_{n=1}^{N_E} \left(E_n^{\mathrm{NN}} - E_n^{\mathrm{DFT}}\right)^2, \tag{2}$$

$$L_F = \frac{1}{3N_F} \sum_{n=1}^{N_F} \sum_{i=1}^{N_{\mathrm{atoms}}} \left\|\mathbf{F}_{i,n}^{\mathrm{NN}} - \mathbf{F}_{i,n}^{\mathrm{DFT}}\right\|^2, \tag{3}$$

$$L_V = \frac{1}{9N_V} \sum_{n=1}^{N_V} \left\|\mathbf{V}_n^{\mathrm{NN}} - \mathbf{V}_n^{\mathrm{DFT}}\right\|^2. \tag{4}$$

The factors $1/3$ and $1/9$ arise from averaging the force and virial errors over their three and nine Cartesian components, respectively, ensuring consistent normalization across all contributions to the loss.

During inference, atomic coordinates are transformed into symmetry-preserving descriptors that are invariant under translation, rotation, and permutation of atoms of the same type [10, 11]. These descriptors serve as input to the neural network, which predicts atomic energies and corresponding forces in real time, achieving *ab initio*-level accuracy at computational cost comparable to classical MD.

A variety of neural-network-based interatomic potentials have been proposed, including Behler–Parrinello networks [8], Gaussian Approximation Potentials (GAP) [12], SchNet [13], and the E(3)-equivariant NequIP model [14]. Among these, the Deep Potential Molecular Dynamics (DeepMD) framework has emerged as one of the most widely adopted and computationally efficient implementations due to its smooth descriptor formulation and native GPU acceleration. The following subsection focuses specifically on **DeepMD-kit version 2**, which forms the theoretical and computational foundation for the present work.

2.3 Deep Potential Molecular Dynamics (DeepMD)

This work focuses on **Deep Potential Molecular Dynamics version 2 (DeepMD-kit v2)**, the second-generation implementation of the Deep Potential framework. DeepMD-kit v2 represents a major advancement over the original DeepMD formulation [2], introducing improved descriptor smoothness, enhanced multi-species handling, and a refined software architecture that enables highly efficient GPU execution [3]. These features make DeepMD-kit v2 particularly suitable for modeling chemically complex systems such as binary alloys within HALMD.

Deep Potential Molecular Dynamics is a machine-learning approach that approximates interatomic interactions with near *ab initio* accuracy while retaining computational efficiency comparable to classical MD. A neural network is trained on quantum-mechanical reference data (typically from DFT or AIMD) to map atomic configurations to energies and forces. The total potential energy of the system is decomposed into atomic energy contributions [2, 3]:

$$E = \sum_i E_i(\mathcal{R}_i),$$

where E_i is the atomic energy of atom i, and $\mathcal{R}_i$ denotes its local environment within a cutoff radius r_c. This locality assumption yields linear scaling in the number of atoms and enables efficient parallelization on GPUs.

A defining element of DeepMD is the use of symmetry-preserving descriptors that are invariant under translations, rotations, and permutations of atoms of the same type. DeepMD-kit v2 employs the *Smooth Edition* (SE) descriptor family, provided in angular (`se_e2_a`) variants [3]. It is constructed from two matrices:

- the R**-matrix**, containing relative position vectors $R_{ij} = \mathbf{r}_j - \mathbf{r}_i$ for neighbors j within the cutoff radius;
- the G**-matrix**, a normalized representation incorporating angular information and many-body geometric correlations.

Distances $r_{ij} = |R_{ij}|$ are modulated by a smooth cutoff function to ensure continuity of energies and forces at the cutoff boundary [2].

The Deep Potential model in both DP-v1 and DP-v2 consists of two neural networks: an **embedding (filter) network** and a **fitting network** [2, 3].

1. **Embedding (Filter) Network.** This smaller feed-forward network transforms raw neighbor information into high-dimensional filter features. In DP-v2, a distinct embedding network is assigned to each atomic species, allowing the model to capture chemically specific interactions (e.g., Cu–Ag, Ni–Al) while preserving permutation invariance within each species [3]. The embedding network processes each neighbor independently.
2. **Fitting Network.** The outputs of the embedding networks are aggregated into a descriptor vector $\mathbf{D}_i$, which is passed to a larger fully connected fitting network that predicts the atomic energy E_i. Each species has its own fitting network parameters, enabling accurate modeling of multi-component materials.

Forces are obtained as the negative gradients of the total energy:

$$\mathbf{F}_i = -\frac{\partial E}{\partial \mathbf{r}_i}.$$

During training, these derivatives are computed automatically through TensorFlow's back-propagation engine. During inference—which is the focus of this thesis—the gradients are obtained by applying the chain rule explicitly through the descriptor and neural-network layers, following the DP-v2 computational graph [3].

DeepMD-kit v2 introduces several enhancements that are essential for accurate multi-species modeling. These include species-dependent embedding networks, species-specific filter weights, descriptor normalization layers, and refined smooth cutoff schemes [3]. Together, these improvements enable DP-v2 to handle chemically diverse systems with greater smoothness, transferability, and numerical stability compared to its predecessor.

A trained DeepMD model is stored as a TensorFlow `frozen_model.pb` file along with an accompanying `input.json` file specifying descriptor types, cutoff radii, hyperparameters, and precision settings. During inference, atomic coordinates are transformed into descriptors, processed by species-specific embedding networks, and fed into the fitting network to produce atomic energies and forces.

In this thesis, DeepMD-kit v2 serves as the machine-learned potential that is fully integrated into HALMD. All parameter extraction, descriptor reconstruction, and neural-network inference follow the DP-v2 specification precisely. By embedding this inference process into HALMD's GPU-accelerated architecture, the present work enables large-scale simulations of multi-species systems, such as binary alloys, with near *ab initio* accuracy.

3 Methodology

3.1 Overview of Implementation

1. Parameter Extraction (Python side).

2. Inference and Integration (C++/CUDA side).

3. Verification and Validation.

References

[1] Andres Cruz. “Deep Neural Networks Potentials for Scalable Molecular Dynamics Simulations on Accelerator Hardware”. Master’s Thesis. Freie Universität Berlin, Sept. 2025.

[2] Han Wang, Linfeng Zhang, Jiequn Han, et al. “DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics”. In: *Computer Physics Communications* 228 (2018), pp. 178–184.

[3] Jinzhe Zeng et al. “DeePMD-kit v2: A software package for deep potential models”. In: *The Journal of Chemical Physics* 159.5 (2023).

[4] Peter H Colberg and Felix Höfling. “Highly accelerated simulations of glassy dynamics using GPUs: Caveats on limited floating-point precision”. In: *Computer Physics Communications* 182.5 (2011), pp. 1120–1129.

[5] HALMD Developers. *HALMD: High-Accuracy Large-scale Molecular Dynamics*. https://github.com/halmd-org/halmd. Accessed: 2025-01-10.

[6] Peter H. Colberg and Felix Höfling. *HALMD Documentation and User Manual*. https://halmd.org/doc/. Accessed: 2025-01-10.

[7] Pierre De Buyl et al. “H5MD: A structured, efficient, and portable file format for molecular data”. In: *Computer Physics Communications* 185.6 (2014), pp. 1546–1553. DOI: 10.1016/j.cpc.2014.01.018.

[8] Jörg Behler and Michele Parrinello. “Generalized neural-network representation of high-dimensional potential-energy surfaces”. In: *Physical Review Letters* 98.14 (2007), p. 146401. DOI: 10.1103/PhysRevLett.98.146401.

[9] Frank Noé et al. “Machine learning for molecular simulation”. In: *Annual review of physical chemistry* 71.1 (2020), pp. 361–390.

[10] Albert P Bartók, Risi Kondor, and Gábor Csányi. “On representing chemical environments”. In: *Physical Review B* 87.18 (2013), p. 184115. DOI: 10.1103/PhysRevB.87.184115.

[11] Manzil Zaheer et al. “Deep Sets”. In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017. URL: https://proceedings.neurips.cc/paper/2017/file/f22e4747da1aa27e363d86d40ff442fe-Paper.pdf.

[12] Albert P Bartók et al. “Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons”. In: *Physical review letters* 104.13 (2010), p. 136403.

[13] Kristof T Schütt et al. “SchNet: A continuous-filter convolutional neural network for modeling quantum interactions”. In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017. URL: https://proceedings.neurips.cc/paper/2017/file/303ed4c69846ab36c2904d3ba8573050-Paper.pdf.

[14] Simon Batzner et al. “E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials”. In: *Nature Communications* 13.1 (2022), p. 2453. DOI: 10.1038/s41467-022-29939-5.