UNIVERSITÉ DU LUXEMBOURG

UMONS Université de Mons

PhD-FSTM-2025-120

The Faculty of Science, Technology and Medicine

The Faculty of Science

DISSERTATION

Defence held on 14/11/2025 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG EN PHYSIQUE

AND

DOCTEUR DE L'UNIVERSITÉ DE MONS EN SCIENCES

by

Anton CHARKIN-GORBULIN

Born on 18th July 1997 in Khrakiv (Ukraine)

Quantum Mechanics and Machine Learning for Organic Electronics: From Structure to Dynamics

Dissertation defence committee

Dr Igor POLTAVSKYI, dissertation supervisor
Research scientist, UNIVERSITÉ DU LUXEMBOURG

Prof. Dr David BELJONNE, dissertation co-supervisor
Professeur (Chargé de Cours), UNIVERSITÉ DE MONS

Prof. Dr Ludger WIRTZ, Chair
Full Professor in Physics, UNIVERSITÉ DU LUXEMBOURG

Prof. Dr Alex REDINGER
Full Professor in Physics, UNIVERSITÉ DU LUXEMBOURG

Prof. Dr Stéphane DUPONT
Professeur (Chargé de Cours), UNIVERSITÉ DE MONS

Prof. Dr Leeor KRONIK
Professor, WEIZMANN INSTITUTE OF SCIENCE

Contents

Abstract **7**

Acknowledgements **11**

1 Introduction **13**

2 Methology: From Quantum Structure to Data-Driven Dynamics **17**
2.1 *Ab initio* Methods . . . 17
2.1.1 Density-Functional Theory . . . 19
2.1.2 D3 Dispersion Correction . . . 24
2.1.3 Nonlocal Many-Body Dispersion . . . 25
2.1.4 DFT Implementation Details . . . 26
2.1.5 Summary of Methodological Choices . . . 27
2.2 Into Atomistic Simulations . . . 27
2.2.1 Molecular Dynamics . . . 28
2.2.2 Ensemble Control: NVT and NPT Algorithms . . . 30
2.2.3 Thermodynamic Integration . . . 33
2.2.4 Summary of Methodological Choice . . . 33
2.3 Machine-Learning Force Fields . . . 34
2.3.1 Rise of Learned Potentials . . . 34
2.3.2 Kernel-Based MLFF . . . 36
2.3.3 Neural Network Potentials . . . 47
2.3.4 Evaluation Metrics, Benchmarking, and Databases . . . 56
2.4 Halide Perovskite Materials . . . 60
2.4.1 Composition and Crystal Structure . . . 61
2.4.2 Challenges in Modeling Halide Perovskites . . . 66

3 Halide Perovskite Interfaces: from Dispersion Interaction to Electronic Structure at Finite Temperature **69**
3.1 Frozen Atomistic and Electronic Structure . . . 71
3.2 MLFF Accuracy and Dispersion Correcetions . . . 74
3.3 Finite-Temperature Dynamics and Electronic Structure . . . 76
3.3.1 Framework Dynamics and Phase Behavior . . . 76
3.3.2 Electronic Structure Evolution . . . 79
3.4 Summary . . . 81

4 Crash Testing Machine Learning Force Fields for Materials and Interfaces **85**
4.1 Assessing Models Accuracy via Test Set Analysis . . . 86
4.1.1 Interface Benchmark: 1,8-naphthyridine on Graphene . . . 88

4.1.2 Solid Benchmark: $MAPbI_3$ Perovskite . . . 90
4.2 Assessing Models Fidelity via Molecular Dynamic Simulations . . . 91
4.2.1 1,8-naphthyridine on Graphene . . . 94
4.2.2 Methylammonium Lead Iodide Perovskite . . . 98
4.3 Pitfalls and Best Practices for Reliable MLFFs . . . 101
4.4 Summary . . . 103

5 Orbits for Improved Machine Learning Force Fields 105
5.1 Graph-Based Symmetry Search Procedure . . . 106
5.1.1 Representative Graph Construction . . . 106
5.1.2 Permutational Symmetry Extraction . . . 109
5.1.3 Integration into Machine Learning Methods . . . 110
5.2 Organic molecules absorbed on graphene . . . 111
5.2.1 Free-Energy Profile of 1,8-Naphthyridine/Graphene . . . 113
5.3 Ac-Phe-Ala5-Lys . . . 115
5.4 Graphene with pyridinic-N defect . . . 118
5.5 Cesium Lead Iodine Perovskite Slab . . . 122
5.6 Summary . . . 126

6 Conclusions 129
6.1 Key Contributions and Findings . . . 129
6.2 Limitations and Future Directions . . . 130
6.3 Concluding Remarks . . . 131

References 133

A Dataset generation procedure 161
A.1 1,8-naphthyridine on Graphene . . . 161
A.2 $MAPbI_3$ Perovskite . . . 161
A.3 Graphene/molecule interfaces . . . 162
A.4 Graphene with pyridinic-N defect . . . 162
A.5 Perovskite slab . . . 163

B Supplementary Information for Chapter 4 165
B.1 Normalized errors on the test sets . . . 165
B.2 Radial Distribution Functions for methylammonium lead iodide perovskite . . . 166

C Supplementary Information for Chapter 5 171
C.1 Graphene with a pyridinic-N defect . . . 171
C.2 Phonon Spectra Consistency . . . 172
C.3 Epoch-Resolved Training and Validation Losses . . . 173

Acronyms

$\mathcal{GP}$ Gaussian Process. 37

ACE Atomic Cluster Expansion. 53

AIMD *Ab Initio* Molecular Dynamics. 35, 162, 163

APES Average Potential Energy Surface. 113, 114, 116, 126

BIGDML Bravais-Inspired Gradient-Domain Machine Learning. 44, 45

CBM Conduction Band Minimum. 72, 73, 81, 82

CG Coarse-Grained. 107, 108

COM Center of Mass. 81, 82, 113, 114, 127

CPU Central Processing Unit. 93, 102

D3 Grimme's third generation dispersion correction. 24, 26, 27, 71–84, 129, 163

DFT Density Functional Theory. 18–22, 24, 26, 35, 56, 57, 66, 67, 70, 74, 77, 83–85, 95, 96, 98–101, 111, 122, 129, 161–163, 165–169

FCHL Faber-Christensen-Huang-Lilienfeld. 34, 45–47, 86–88, 92, 93, 97, 103, 115, 129

FFs Force Fields. 28, 34, 35, 47, 85, 105, 110

GAP Gaussian approximation potential. 34, 39–41, 86, 92, 93, 96–101, 103, 115, 129

GDML Gradient-Domain Machine Learning. 41, 42

GGA Generalized Gradient Approximation. 19, 21–23, 26, 67

GPR Gaussian Process Regression. 36–39

GPU Graphics Processing Unit. 47, 49, 89, 93, 102

GPW Gaussian Plane Waves. 26, 27, 74, 163

GTH Goedecker-Teter-Hutter. 27, 163

GUC Graphene's Unit Cell. 113, 115

KRR Kernel Ridge Regression. 36–39, 41, 42, 47

LDA Local Density Approximation. 19, 21, 22

MAE Mean Absolute Error. 57, 59, 77, 86–91, 96, 97, 102, 103, 111, 112, 115, 117, 119, 121–124, 126, 165, 166, 171

MAXE Maximum Error. 57, 86, 87, 91, 103, 165, 166

MBD Many-Body Dispersion. 24–26

MBD-NL Nonlocal Many-Body Dispersion. 25–27, 71–84, 100, 129, 161–163

MD Molecular Dynamics. 28–30, 33, 44, 49, 56, 58, 59, 66, 67, 74–76, 78–86, 90–101, 103, 113, 127, 131, 161–164, 166–169

ML Machine Learning. 17, 49, 110

MLFF Machine Learning Force Field. 17, 23, 26–30, 33–36, 53, 56–59, 67, 69, 70, 74–79, 82–86, 88–98, 100–105, 108, 110, 115, 118, 122, 126–131, 161–163

MLP Multilayer Perceptron. 51, 54, 118

MPNN Message-Passing Neural Network. 17, 34, 50, 67, 86, 87, 92, 100, 104, 105, 110, 113, 117, 118, 124–126, 129, 130

MSE Mean Squared Error. 48

MST Minimum Spanning Tree. 43

MTK Martyna-Tuckerman-Klein. 32–34

NAO Numeric Atom-Centered Orbitals. 27, 74

NN Neural Network. 17, 45, 86, 87, 92, 93, 95–97, 99, 103

NNP Neural Network Potential. 48–50, 56, 57

PBC Periodic Boundary Conditions. 44, 62

PBE Perdew-Burke-Ernzerhof. 21, 23, 24, 26, 27, 71–74, 76–83, 100, 161–163, 166–169

PES Potential Energy Surface. 18, 28, 31, 34, 36, 41, 42, 44, 47–49, 59, 95, 100

PSMF Potential Surface of Mean Force. 113, 114, 116, 126

QC quantum-chemical. 34, 35, 44, 56, 109, 113

QHOs Quantum Harmonic Oscillators. 25, 26

RFF Random Fourier Features. 47

RMSE Root Mean Square Error. 57, 58, 60, 83, 86, 87, 91, 97, 102, 111, 121, 122, 125, 126, 165, 166, 171, 172

SCF Self-Consistent Field. 74

SD Standard Deviation. 81–83

sGDML Symmetric Gradient-Domain Machine Learning. 34, 42–45, 86, 91–97, 103, 105, 110–113, 115, 117, 126, 129, 130, 161, 163

SOAP Smooth Overlap of Atomic Positions. 34, 39–41, 86, 92, 93, 96–101, 103, 115, 129

SOC Spin-Orbit Coupling. 23, 27

TI Thermodynamic Integration. 33, 34, 113

VBM Valence Band Maximum. 72, 73, 81, 82

vdW van der Waals. 18, 23–25, 27, 35, 66, 67, 70, 71, 74–76, 83, 84, 106–108, 162, 163

XC Exchange-Correlation. 19, 21, 22, 24, 163

Abstract

Machine-learning force fields (MLFFs) bridge the gap between the high accuracy of quantum-chemical (QC) calculations and the efficiency required for large-scale atomistic simulations, enabling QC-level studies on spatial and timescales previously out of reach. This scalability not only expands the reach of fundamental simulations but also accelerates the discovery of functional materials, from advanced photovoltaic devices to next-generation pharmaceuticals. Realizing these gains, however, demands careful choices at every step – from the selection of the QC reference method and sampling of relevant configurational and chemical space to the design, training, improvement, and benchmarking of the MLFF models themselves.

In this thesis, we present a unified framework for systematically developing and validating MLFFs for complex semiconducting materials and interfaces. First, we conduct large-scale MLFF simulations of a perovskite slab system to provide detailed insights into its geometric and electronic structure. We perform a comprehensive analysis of state-of-the-art MLFFs to identify the most suitable architectures and models, pinpoint potential avenues for improvement, and establish a rigorous benchmarking protocol for atomistic simulation of materials and interfaces. Finally, we introduce a novel data-driven strategy that systematically incorporates molecular and crystalline symmetries, thereby enhancing the predictive accuracy and efficiency of MLFFs.

Practically, we have conducted large-scale simulations of cesium lead iodide ($CsPbI_3$) perovskite slab. We demonstrate that the nonlocal many-body dispersion (MBD-NL) method accurately reproduces the experimentally observed phase diagram, unlike another widely used D3 scheme, which incorrectly predicts the cubic phase at room temperature instead of the orthorhombic phase. Furthermore, by extending our study to finite temperatures for a slab structure, we demonstrate the influence of the van der Waals interaction on the electronic structure near the surface and within the slab. In combination with state-of-the-art MLFF models, these findings establish a foundation for large-scale and predictive modeling of the geometric and electronic properties of two-dimensional perovskite interfaces, thereby advancing their mechanistic understanding and facilitating the design of next-generation materials.

By conducting a comprehensive assessment of state-of-the-art MLFF architectures within the TEA Challenge 2023, we benchmarked five representative models—sGDML, SOAP/GAP, FCHL19, MACE, and SO3krates—across periodic materials and interfaces. While modern equivariant message-passing neural networks deliver the highest overall accuracy, all architectures display considerable

maximum force errors and heterogeneous per-atom performance. In molecular dynamics simulations, kernel-based methods often fail outside well-sampled regions, whereas E(3)-equivariant neural networks generally preserve stability but remain limited in capturing long-range interactions critical for adsorption and desorption processes. These findings provide practical guidelines for MLFF development, establish a streamlined evaluation workflow, and highlight key directions for improving model robustness and transferability in the study of materials and molecular systems.

Heterogeneity in force predictions for different atomic environments underscores the need for a data-driven symmetry search to automatically identify atomic "orbits"—chemically distinct local environments that conventional MLFFs often conflate under their global permutation-invariance assumptions. We demonstrate significant improvements in force prediction accuracy when integrating these orbits into both kernel-based (sGDML) and E(3)-equivariant neural message-passing (MACE) architectures for organic interfaces and slab perovskite systems.

For sGDML, trained on ethanol, 1,8-naphthyridine, D-alanine, and D-histidine molecules adsorbed on graphene, we establish a strong correlation between force prediction accuracy and chemical diversity, quantified by orbit count. Using an orbit-enhanced sGDML model, we constructed the average potential energy surface and potential surface of mean force for 1,8-naphthyridine on graphene, finding energy barriers exceeding 300 K between unit-cell minima, preferential diffusion along narrow channels that avoid atop-carbon sites, and substantial rotational barriers about the surface-normal axis – suggesting that, at room temperature, transport proceeds via coupled in-plane translation and rotation. Incorporating orbits into MACE enables us to reduce the model size by an order of magnitude while preserving predictive accuracy, as demonstrated for the $CsPbI_3$ perovskite slab and graphene with a pyridinic-N defect.

Overall, this work advances a robust pipeline for reliable and efficient atomistic simulations by integrating high-quality datasets, physically informed ML model architectures, and community-wide efforts toward rigorous validation protocols that extend beyond static error metrics to include dynamic stability and chemical transferability.

Publications

Part of the thesis is based on the following publications:

1. Anton Charkin-Gorbulin, Artem Kokorin, Huziel E Sauceda, Stefan Chmiela, Claudio Quarti, David Beljonne, Alexandre Tkatchenko and Igor Poltavsky. "Atomic orbits in molecules and materials for improving machine learning force fields". In: Mach. Learn.: Sci. Technol. 6.3 (Jul. 2025) p. 035005. DOI: 10.1088/2632-2153/adea0b

2. Igor Poltavsky, Anton Charkin-Gorbulin, Mirela Puleva, *et al.* "Crash testing machine learning force fields for molecules, materials, and interfaces: model analysis in the TEA Challenge 2023". In: Chem. Sci. 16.8 (2025), pp. 3720-3737. DOI: 10.1039/d4sc06529h.

3. Igor Poltavsky, Mirela Puleva, Anton Charkin-Gorbulin, *et al.* "Crash testing machine learning force fields for molecules, materials, and interfaces: molecular dynamics in the TEA Challenge 2023". In: Chem. Sci. 16.8 (2025), pp. 3738-3754. DOI: 10.1039/d4sc06530a.

Acknowledgements

This thesis would not have been possible without the unwavering support of my mum, the courage and strength of my dad, and the encouragement of my grandad Viktor Charkin.

Special thanks to my close friends and to those whose resilience and sacrifice made it possible to keep going during difficult times.

Chapter 1

Introduction

Semiconductors are materials with band gaps that enable intermediate and tunable conductivity. Their carrier populations and mobilities can be engineered by doping, electric fields, temperature, and illumination, which makes them the physical basis of transistors, integrated circuits, solar cells, and light emitters [1, 2]. As device dimensions shrink toward the nanoscale, in two-dimensional and quasi-2D semiconductors, interfaces determine band alignment and mediate charge separation, transport, and recombination, so surface chemistry and van der Waals (vdW) assembly become decisive for device function. [3, 4, 5, 6].

Among emerging semiconductor materials, two classes have demonstrated exceptional promise for next-generation electronics. Metal halide perovskites combine strong optical absorption with unusually long minority-carrier diffusion lengths, a combination that supports high-efficiency photovoltaic and optoelectronic devices [7, 8]. Meanwhile, graphene offers high carrier mobility, chemical robustness, and an atomically clean template for vdW heterostructures that enable designer band alignments at interfaces [4]. Recent experimental advances include certified record efficiencies in perovskite-silicon tandem photovoltaics and the observation of correlated insulating and superconducting phases in magic-angle twisted bilayer graphene, highlighting the transformative role of engineered 2D interfaces [9, 10]. These developments motivate a quantitative, atomistic understanding of structure-property relationships at semiconductor interfaces.

However, accurate atomistic-scale modeling of organic/inorganic complex materials remains a formidable challenge due to the intricate interplay of covalent bonding, weak noncovalent interactions, and dynamic disorder at interfaces. Traditional quantum-chemical methods deliver high fidelity but are prohibitively expensive, while classical force fields sacrifice transferability and fail to capture subtle electronic polarization and charge-transfer phenomena.

Machine learning (ML) force fields (FFs) have emerged as transformative tools that bridge this gap by learning potential-energy surfaces (PES) and interatomic forces from high-quality reference data. The field was pioneered by the Behler-Parrinello high-dimensional neural-network potential, which introduced atom-wise energy decomposition trained on electronic-structure data, and atom-centered symmetry functions later formalized the required rotational, translational, and permutational invariances [11, 12]. Kernel methods followed, notably the Gaussian Approximation Potential utilizing Gaussian process regression and, soon after, atomic environment descriptors such as SOAP and ACSF, for smooth local environment comparisons [13, 14]. Kernel-based global models such as GDML and its symmetric extension sGDML improved accuracy for small molecules by learning forces with a global descriptor and enforcing energy conservation, with sGDML additionally imposing permutational symmetries. The BIGDML variant generalizes this framework to periodic materials by encoding Bravais and space-group invariances [15, 16, 17, 18]. More recently, invariant and equivariant neural architectures — including SchNet [19], PaiNN [20], NequIP [21], Allegro [22], SO3krates [23], and MACE [24, 25]— have extended MLFF applicability to complex, chemically diverse systems by integrating local and sometimes long-range interactions within symmetry-aware message-passing frameworks.

Despite rapid progress in MLFF, accurately modeling halide perovskites remains challenging due to their complex phase behavior, defect dynamics, and interfacial phenomena. Particular interest presents $CsPbI_3$, it offers improved thermal robustness but its photoactive black perovskite polymorphs (α, β, and γ) are metastable at room temperature relative to the non-perovskite yellow δ phase, motivating stabilization by composition, nanostructuring, and interfacial control.

Modeling $CsPbI_3$ slabs shows that temperature-dependent surface reconstructions and mobile ions heighten the sensitivity of MLFF predictions to the dispersion treatment used in the density functional theory (DFT) reference [26, 27, 28, 29]. Across training sets that differ only by the chosen vdW method, surface forces shift relative to bulk-like layers and the predicted band gap changes, because vdW-induced geometry differences (lattice constants and octahedral tilts) feed directly into the electronic structure [30, 31, 32, 33, 34, 27]. This functional dependence at interfaces aligns with community guidance and recent assessments of vdW interactions. Consistent with our benchmarking, force-error heterogeneity is largest for surface atoms and differs from interior layers, which impacts molecular dynamic (MD) stability and the physical fidelity of trajectories and derived observables [35, 36, 37].

To address field-wide challenges rather than a single material, we report results from the 2023 “Crash Testing Machine Learning Force Fields” (TEA) Challenge, a multi-team bench-

mark spanning molecules, molecule-surface interfaces, and periodic solids. The evaluation of kernel-based SOAP/GAP, sGDML, FCHL19* models, and state-of-the-art MACE and SO3krates equivariant models were assessed on held-out tests and via identically configured MD workflows on a shared HPC system, enabling comparison of resulting structural, thermodynamic, and kinetic observables at identical conditions. By comparing structural, thermodynamic, and kinetic properties across architectures, the challenge reveals both strengths and pitfalls of current MLFF methodologies in realistic simulation workflows — including stability at elevated temperature, the importance of long-range interactions, and error heterogeneity across systems and atomic environments.

The TEA Challenge results and $CsPbI_3$ study highlighted a fundamental aspect requiring further investigation: the treatment of symmetry in MLFFs. Symmetry is a central design principle across modern MLFFs. Atom-centered descriptors such as ACSFs and SOAP encode translational, rotational, and permutational invariances directly into the representation, which improves data efficiency. Equivariant graph neural networks then enforce E(3) symmetries at the model level and propagate tensorial features through message passing. As datasets now span larger, interfacial, and multi-element systems, blanket assumptions of full permutation invariance over identical species can conflate chemically distinct sites and bonding motifs.

To complement current practice, we introduce a data-driven symmetry search method that constructs a representative atomic graph—based on van der Waals radii and interatomic geometry—and applies graph-automorphism algorithms to uncover permutational symmetry across all configurations in a dataset [38, 39]. This procedure yields (i) the total number of symmetries, (ii) a minimal generating set of symmetry operations, and (iii) automorphism orbits, i.e., partitions of atoms that are mapped onto one another by automorphisms of the constructed graph, defining equivalence classes of structurally equivalent neighborhoods.

We demonstrate the utility of this symmetry analysis by integrating it into both kernel-based (sGDML) and neural message-passing (MACE) MLFF frameworks. For periodic organic interfaces — exemplified by the adsorption of heterocyclic molecules on graphene — and perovskite slab, our approach refines chemical neighborhoods and improves force-prediction accuracy by treating orbit-distinct atoms as separate types. In the case of 1,8-naphthyridine on graphene, the sGDML model that incorporates the extracted permutational symmetries reconstructs the adsorption free energy landscape at room temperature in days, replacing over seven million *ab initio* single-point evaluations — an effort infeasible with direct quantum-chemical sampling.

The manuscript has the following structure. The Chapter 2 presents a survey of electronic-structure methods, the evolution of MLFF architectures (kernel and equivariant message-

passing neural networks), handling symmetry, and long-range interactions. The Chapter 3 investigates $CsPbI_3$ surfaces, revealing vdW-functional sensitivity, surface-bulk force-error heterogeneity, and implications for electronic structure. The Chapter 4 reports a one-to-one comparison benchmark across GAP/SOAP, sGDML, FCHL19*, MACE, and SO3krates MLFF models evaluating their accuracy, stability, and reliability. The Chapter 5 introduces a data-driven symmetry search via graph automorphisms, integrates it into sGDML and MACE, and demonstrates accuracy gains for periodic organic/graphene interfaces and perovskite slabs, including reconstruction of adsorption free-energy landscapes. The Chapter 6 summarizes the findings and outlines future directions.

Chapter 2

Methology: From Quantum Structure to Data-Driven Dynamics

This chapter establishes the computational methodology for probing atomistic and electronic structure on nanometer and nanosecond scales. The workflow integrates quantum-mechanical *ab initio* methods, molecular dynamics with ensemble-control algorithms, enhanced sampling, and free-energy integration. These components enable the creation of high-fidelity training data for Machine Learning Force Field (MLFF) and support quantitative analysis of electronic structure.

The landscape of MLFF architectures is then surveyed, covering kernel methods, Neural Networks (NNs), and equivariant Message-Passing Neural Network (MPNN) models, together with the evaluation procedures used to quantify accuracy, robustness, and efficiency. Within this methodological framework halide perovskites provide the central application and validation platform as photovoltaic materials whose structural softness, anharmonic lattice dynamics, and complex defect chemistry make them stringent tests for Machine Learning (ML)-accelerated simulation.

2.1 *Ab initio* Methods

Materials stability, reactivity, and overall behavior are fundamentally determined by how the energy of constituent atoms changes in response to the atoms' motion and environment. To define atomic dynamics and quantify interactions, one must specify both the location of

the nucleus and the distribution of the electrons in the system. Electrons are at least three orders of magnitude lighter than nuclei and therefore respond on much shorter time scales to changes in atomic environments. The Born-Oppenheimer approximation exploits this fact in the separation of time scales: nuclear motion is treated as slow relative to electronic motion, so the many-body problem decouples into an *electronic problem* at fixed nuclear positions and a *nuclear problem* on the resulting Potential Energy Surface (PES). For systems near equilibrium, the ground-state electronic structure at fixed nuclei defines a PES that is quantitatively reliable for nuclear dynamics and permits computationally efficient evaluation of energies and forces. This reliability holds when static correlation is weak and excited-state character is negligible. When these conditions are not met, methods beyond a ground-state treatment are required. In this thesis, we restrict attention to regimes where both the Born-Oppenheimer separation and a ground-state electronic description are valid, and these assumptions underpin the *ab initio* methods used here.

Ab initio electronic-structure methods aim to predict atomic behavior and materials properties from fundamental constants, avoiding system-specific fitting while relying on controlled numerical choices (basis sets or plane-wave cutoffs, k-point sampling, pseudopotentials, exchange-correlation approximations). For notational simplicity we adopt the time-independent many-electron Schrödinger equation:

$$\left(-\frac{1}{2} \sum_{i=1}^{N} \nabla_i^2 + \sum_{i=1}^{N} V(\mathbf{r}_i) + \sum_{i=1}^{N} \sum_{j<i} U\left(\mathbf{r}_i, \mathbf{r}_j\right) \right) \psi = E\psi, \tag{2.1}$$

where N is the number of electrons, ψ is the many-electron wavefunction, E is the total energy, $V(\mathbf{r}_i)$ is the potential energy of the i-th electron in the external field of nuclei, and $U(\mathbf{r}_i, \mathbf{r}_j)$ is the electron-electron interaction potential.

Direct solution of the many-electron wavefunction scales exponentially with N and is infeasible for extended systems. In practice, Density Functional Theory (DFT) delivers a practical compromise between accuracy and cost, reliably describing covalent bonding, charge transfer and basic lattice energetics, via the Hohenberg-Kohn theorems and Kohn-Sham formalism [40, 41]. To capture interactions missing in semilocal DFT, dispersion-augmented schemes — either pairwise or many-body treatments capturing collective van der Waals (vdW) effects — are indispensable for adsorption of organics on metal and 2D surfaces, interlayer binding and exfoliation in layered materials, stability and ranking of molecular-crystal polymorphs, and host-guest binding and framework cohesion in MOFs [42, 43, 44, 45].

2.1.1 Density-Functional Theory

DFT reformulates the many-body electronic-structure problem as a minimization over the ground-state electron density ($\rho = \sum_{i=1}^{N} |\psi_i|^2$, where ψ_i is the wavefunction of the i-th electron), thereby circumventing the steep computational scaling associated with solving the Schrödinger equation for the full many-electron wavefunction in systems with hundreds of atoms [40, 41, 46, 47]. In practice, the Kohn-Sham formulation introduces a set of single-particle orbitals whose self-consistent solution (the full cycle shown in Figure 2.1) yields both kinetic and Coulomb contributions exactly, leaving only the exchange-correlation energy to be approximated. The Kohn-Sham energy functional is expressed as:

$$E_{\mathrm{KS}}[\rho] = T[\rho] + V_{\mathrm{ext}}[\rho] + E_{\mathrm{H}}[\rho] + E_{\mathrm{xc}}[\rho], \tag{2.2}$$

where $T[\rho]$ is the kinetic energy functional, $V_{\mathrm{ext}}[\rho]$ is the external potential from nuclei, $E_{\mathrm{H}}[\rho]$ is the Hartree energy from electron-electron repulsion, and $E_{\mathrm{xc}}[\rho]$ is the Exchange-Correlation (XC) energy functional.

$E_{\mathrm{xc}}[\rho]$ is the only unknown in the total-energy functional and must be approximated [40, 41]. The simplest nonempirical choice is the Local Density Approximation (LDA)[48], which assumes that at each point in space the XC energy density is the same as that of a uniform electron gas of density $\rho(\mathbf{r})$. In other words, the LDA approximates the XC energy as a local function of the electron density:

$$E_{\mathrm{xc}}[\rho] = \int \rho(\mathbf{r})\, \epsilon_{\mathrm{xc}}(\rho(\mathbf{r}))\, \mathrm{d}\mathbf{r}, \tag{2.3}$$

where $\epsilon_{\mathrm{xc}}(\rho)$ is the XC energy per particle of the uniform electron gas at density ρ[49]. In practice, the LDA yields surprisingly good ground-state energies and bond lengths across a wide range of systems, despite its simplicity.

Early success with the LDA demonstrated that even a simple functional of the density can capture basic bonding trends, but systematic errors in lattice constants and reaction energies spurred the development of gradient-corrected forms [50, 51]. In particular, the Generalized Gradient Approximation (GGA) replaces the purely local XC energy in Eq. 2.3 with a semi-local form that explicitly depends on both the density and its gradient:

$$E_{\mathrm{xc}}[\rho] = \int \rho(\mathbf{r})\, \epsilon_{\mathrm{xc}}(\rho(\mathbf{r}), \nabla\rho(\mathbf{r}))\, \mathrm{d}\mathbf{r}. \tag{2.4}$$

By incorporating the local density gradient in both exchange and correlation terms, GGA functionals systematically correct LDA's tendency to overbind and to misestimate surface-

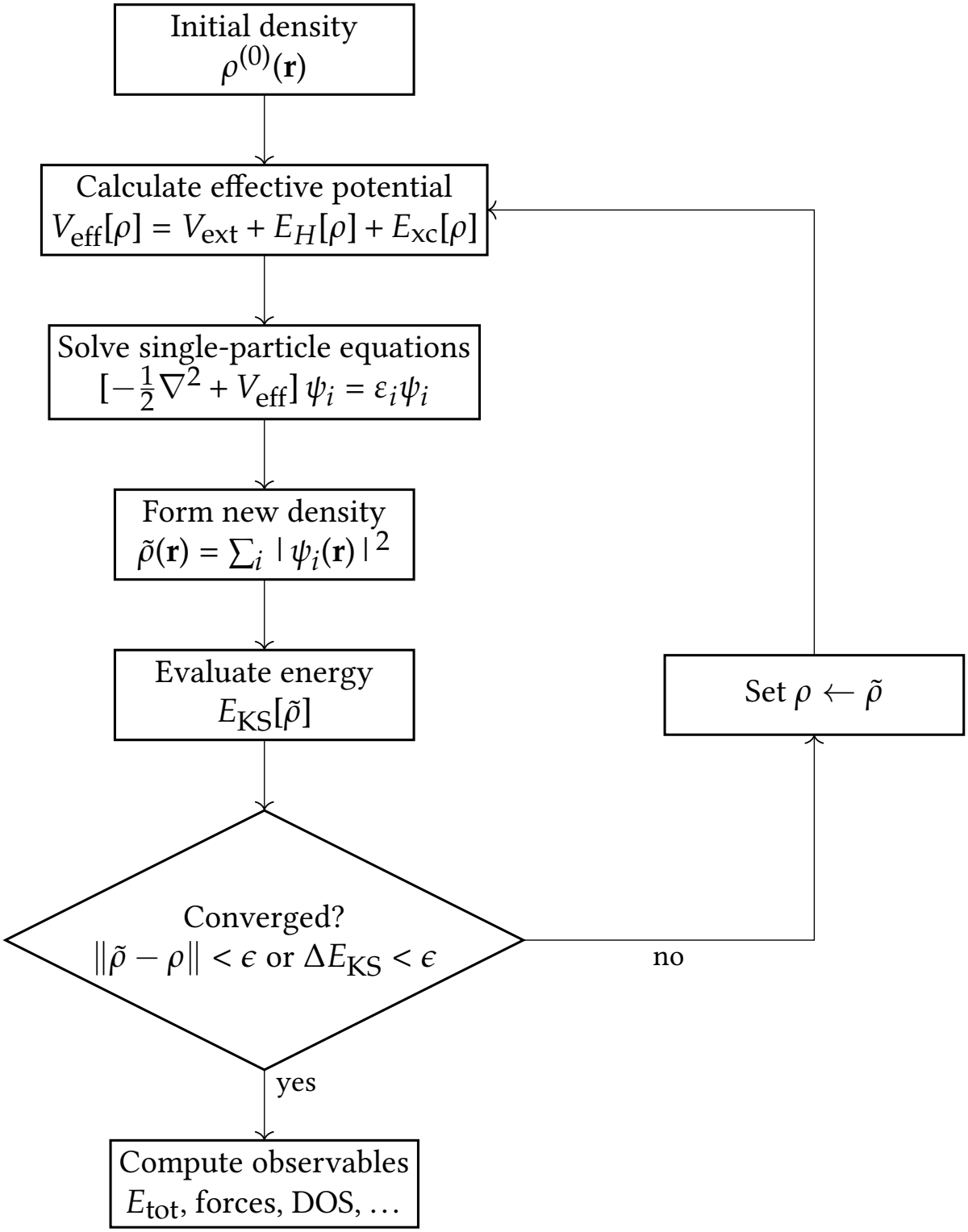


Figure 2.1 Workflow for self-consistent field calculations in DFT.

tension contributions [52, 53].

In Perdew's "Jacob's ladder", shown in Figure 2.2, a five-rung hierarchy of XC approximations in DFT that climb toward "chemical accuracy", LDA and GGA are the foundation. Beyond them, three more rungs are widely used: meta-GGAs, hybrids, and double-hybrids. Meta-GGAs augment semilocal forms by including kinetic-energy density $\tau(\mathbf{r})$ and/or the density Laplacian $\nabla^2\rho(\mathbf{r})$ to improve structures and thermochemistry at near-GGA cost (e.g., SCAN). Hybrids mix a fixed or range-separated fraction of exact Hartree-Fock exchange with semilocal exchange-correlation, which reduces self-interaction errors at increased cost from nonlocal exchange (e.g., PBE0, B3LYP). Double-hybrids further add a perturbative second-order (MP2-like) correlation term evaluated with Kohn-Sham orbitals (e.g., B2PLYP and related DHDFs). Each new rung aims toward improving accuracy at the price of computational cost [54, 55, 56, 57, 58].

PBE Functional

Among the most widely adopted members of the GGA family is the Perdew-Burke-Ernzerhof (PBE) functional [52]. Its common usage arises from its non-empirical parametrization: rather than fitting to experimental observations, it is constructed entirely from first-principles constraints. This makes PBE broadly transferable and robust across a vast range of chemical systems [59]. PBE is designed to satisfy several exact constraints of the true XC functional:

1. **Uniform electron gas limit:** reduces to the Local Density Approximation (LDA), reproducing the known exchange-correlation energies of the homogeneous electron gas.

2. **Second-order gradient expansion:** recovers the second-order expansion of the exchange and correlation energies for slowly varying densities.

3. **Lieb-Oxford bound:** enforces a rigorous lower bound on the indirect part of the Coulomb repulsion energy [60].

4. **Spin-scaling relation:** satisfies the exact relation between spin-polarized and spin-unpolarized contributions [61].

5. **Uniform density scaling:** exhibits the proper scaling of the exchange-correlation energy under uniform density scaling [62].

6. **Exchange-correlation hole sum rule:** ensures the hole integrates to -1 and remains nonpositive [63].

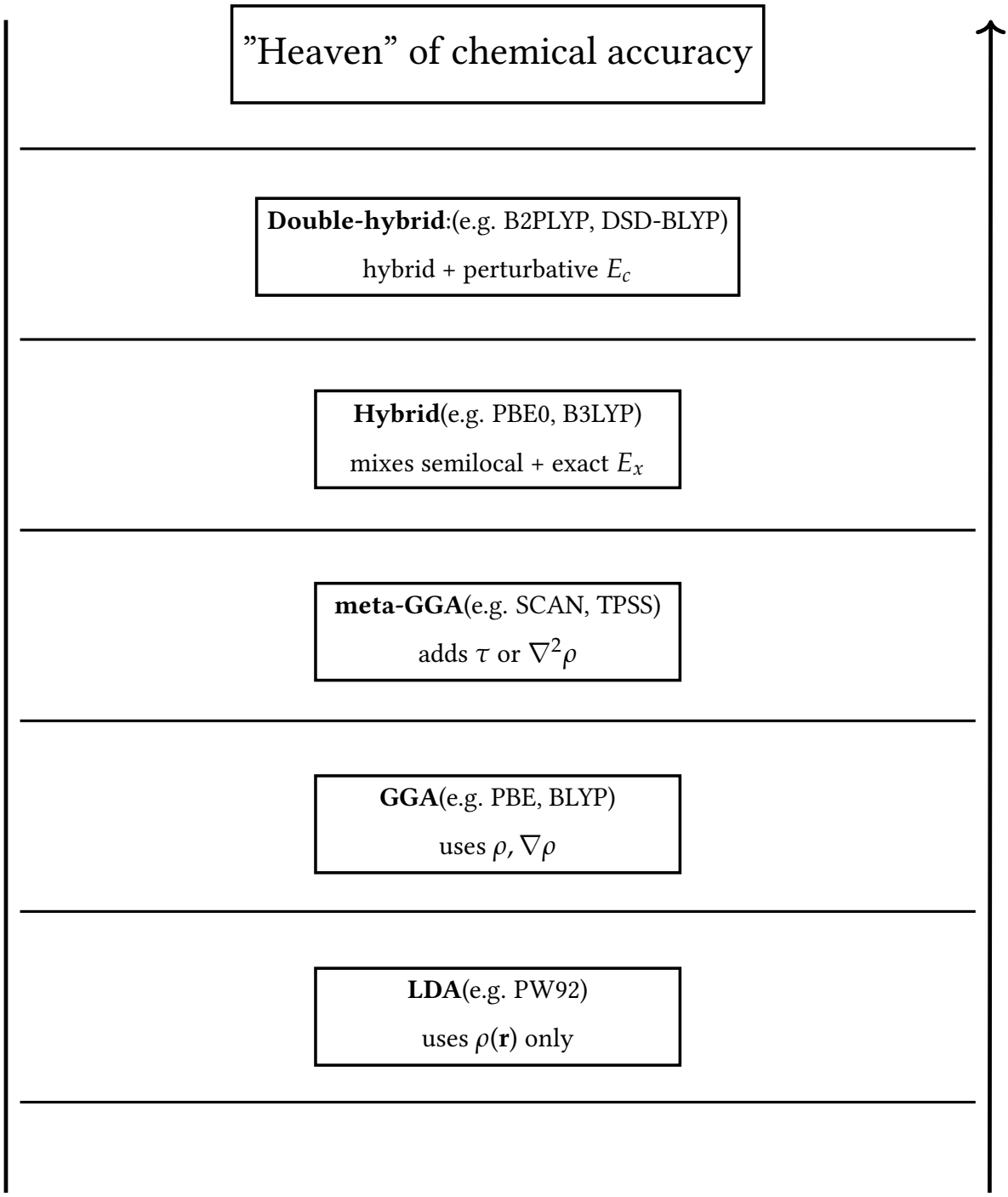


Figure 2.2 Jacob's ladder classification of XC approximations in DFT. The five rungs comprise LDA, GGA, meta-GGA, hybrid functionals, and fifth-rung orbital-dependent methods such as random-phase approximation and double-hybrids. Higher rungs introduce increasingly nonlocal ingredients and typically improve accuracy at increased computational cost.

Moreover, by adopting a semilocal form that depends only on the local electron density and its gradient, PBE enforces all of the above physical constraints without introducing any non-local operators. Consequently, it retains the conventional $O(N^3)$ computational scaling—where N denotes the dimension of the Kohn-Sham eigenproblem (i.e., the number of basis functions or orbitals to be diagonalized), which in practice grows approximately linearly with the atom count—making it a workhorse for solid-state physics and materials science [64, 47]. It generally provides good predictions for structural properties like bond lengths and lattice constants, as well as for the energies of chemical reactions. This combination of efficiency and reasonable accuracy has established PBE as a *de facto* standard for generating reference data for training MLFFs [65].

Despite its successes, the PBE functional is not without its limitations. Like most GGAs, it suffers from inherent self-interaction error, where an electron spuriously interacts with its own charge density. This deficiency leads to several well-known failures:

- **Incorrect Dissociation Limits:** The self-interaction error causes problems in describing the dissociation of molecules and charge localization, often leading to fractional charges on separated fragments.
- **Underestimation of Reaction Barriers:** PBE tends to underestimate the energy barriers of chemical reactions, which can result in inaccurate predictions of reaction rates and mechanisms.
- **Underestimation of Band Gaps:** PBE systematically underestimates the electronic band gaps of semiconductors and insulators, often by 30-50%, which can lead to a qualitatively incorrect description of their electronic properties.

A notable exception to the band gap underestimation occurs in materials containing heavy elements, such as the lead-halide perovskites central to this thesis. In lead-halide perovskite systems, a fortuitous error cancellation can occur when Spin-Orbit Coupling (SOC) effects are neglected. SOC, which is significant for heavy elements like lead and iodine, acts to reduce the band gap. Standard PBE calculations that omit SOC benefit from an accidental compensation: the inherent underestimation of the band gap by the PBE functional is offset by the failure to account for the gap-reducing SOC effect. This can lead to PBE band gaps that are surprisingly close to experimental values, albeit for the wrong reasons [66, 67]. This makes PBE a pragmatic, if imperfect, choice for generating reference data for such systems.

Furthermore, as a semi-local functional, PBE cannot capture long-range vdW (dispersion) interactions, which are non-local in nature. These forces are critical for accurately modeling molecular crystals, layered materials, adsorption phenomena, and biological systems.

Consequently, PBE is almost always used in conjunction with a dispersion method, such as the Grimme's third generation dispersion correction (D3) or Many-Body Dispersion (MBD) schemes, to account for these essential interactions.

2.1.2 D3 Dispersion Correction

Density-functional approximations based on local or semilocal XC functionals systematically underestimate long-range vdW interactions [68, 69]. D3 dispersion correction is constructed to address this challenge by adding an explicit atom-pairwise term to the DFT total energy and restoring missing attractive forces [70]. In its simplest form, the D3 correction is expressed as a sum over all pairs of atoms i and j:

$$E_{\text{disp}} = -\sum_{i<j}\left[s_6 \frac{C_6^{ij}}{R_{ij}^6} f_{\text{damp}}(R_{ij}) + s_8 \frac{C_8^{ij}}{R_{ij}^8} f_{\text{damp}}(R_{ij})\right], \tag{2.5}$$

where R_{ij} is the interatomic distance, C_n^{ij} are the dispersion coefficients for dipole-dipole ($n = 6$) and higher-order ($n = 8$) interactions, s_n are functional-dependent scaling factors (e.g. for PBE or B3LYP), and f_{damp} smoothly attenuates the correction at short range to avoid double-counting correlation effects [71]. The dispersion coefficients C_6^{ij} and C_8^{ij} are derived from the atomic polarizabilities of the constituent atoms, which are typically obtained from quantum mechanical calculations or empirical fits [72].

A key advance in D3 is the environment-dependent determination of dispersion coefficients. Instead of fixed atomic C_6 values, D3 computes coordination-number-corrected polarizabilities on the fly, so that, for example, a carbon atom in an sp^2 environment carries a different C_6 than one in an sp^3 environment. This coordination-based scheme markedly improves transferability across diverse chemical motifs without introducing a separate empirical parameter for each atomic pair.

Computationally, the D3 correction adds only an $O(N^2)$ pairwise summation to the underlying DFT calculation, incurring negligible overhead compared to the $O(N^3)$ cost of the Kohn-Sham solution. This efficiency makes D3 a good choice for systems where long-range dispersion is important–supramolecular assemblies, surface adsorption, layered solids, and biomolecular complexes–while preserving the generality and performance of standard DFT codes.

However, as an empirical, pairwise-additive scheme, D3 cannot capture true many-body dispersion effects, sometimes leading to errors in systems with strong many-body interactions or delocalized metallic character. Its functional- and element-specific parameterization may

also be less reliable for atypical chemical environments (e.g., heavy elements or highly charged ions). Finally, the short-range damping functions can induce overbinding or underbinding compared to nonlocal vdW treatments, so more advanced models (e.g., D4 [73], VV10 [74], or MBD(-NL) [75, 76]) may be needed when higher accuracy is required.

2.1.3 Nonlocal Many-Body Dispersion

The original MBD model represents long-range correlation via a system of dipole-coupled Quantum Harmonic Oscillators (QHOs), one per atom, whose parameters reproduce the atom-in-material polarizability [75]. The Hamiltonian is

$$\hat{H}^{\mathrm{MBD}} = -\frac{1}{2}\sum_{i=1}^{N}\nabla^2_{\boldsymbol{\xi}_i} + \frac{1}{2}\sum_{i=1}^{N}\omega_i^2\,\boldsymbol{\xi}_i^2 + \frac{1}{2}\sum_{i\neq j}^{N}\omega_i\omega_j\sqrt{\alpha_{0,i}\alpha_{0,j}}\,\boldsymbol{\xi}_i^{\top}\,\mathbf{T}_{ij}^{\mathrm{LR}}\,\boldsymbol{\xi}_j, \tag{2.6}$$

where $\boldsymbol{\xi}_i$ are the (mass-weighted) dipole displacements of oscillator i, ω_i and $\alpha_{0,i}$ are its characteristic frequency and static polarizability, and $\mathbf{T}_{ij}^{\mathrm{LR}}$ is the long-range dipole-dipole interaction tensor with damping function . The MBD correlation energy is the change in zero-point energy upon coupling the oscillators,

$$E_{\mathrm{disp}}^{\mathrm{MBD}} = \frac{1}{2}\sum_{k=1}^{3N}\left(\tilde{\omega}_k - \omega_k\right), \tag{2.7}$$

where $\{\tilde{\omega}_k\}$ are the eigenvalues of the Hamiltonian in Eq.2.6 and $\{\omega_k\}$ those of the uncoupled QHOs [75].

Nonlocal Many-Body Dispersion (MBD-NL) preserves the coupled-oscillator Hamiltonian of MBD and modifies it by deriving parameters in the long-range dipole-dipole interaction tensor from a nonlocal, density-based polarizability functional [76]. Starting from the Kohn-Sham density $n(\mathbf{r})$, a Vydrov-Van Voorhis (VV) polarizability functional $\alpha^{\mathrm{VV}}[n](\mathbf{r}, iu)$ provides a spatially and frequency-resolved estimate of the local dipolar response [77]. Coarse-graining this field with Hirshfeld weights yields atomic dynamic polarizabilities $\alpha_i(iu)$ and the corresponding $C_{6,ii}$, which are then renormalized to accurate free-atom reference values to enforce the correct atomic limit. To prevent double counting with semilocal correlation and to enforce a vanishing response in jellium-like regions, MBD-NL smoothly attenuates contributions identified by two local descriptors: the local ionization potential $I[n] = \tau_W[n]/n$ [78] and the iso-orbital indicator $\chi[n] = \left(\tau_{KS}[n] - \tau_W[n]\right)/\tau_{\mathrm{unif}}[n]$ [79, 80, 81], where $\tau_W[n]$ is the von Weizsäcker kinetic energy density and $\tau_{KS}[n]$ is the Kohn-Sham kinetic energy density, while $\tau_{\mathrm{unif}}[n] = 3(3\pi^2)^{2/3}n(\mathbf{r})^{5/3}(\mathbf{r})/10$ is the uniform electron gas kinetic energy density.

Regions with $I \approx 0$ and $\chi \approx 1$ (metallic or slowly varying densities) are down-weighted via a logistic filter, while molecular and bonding regions are retained. The resulting site polarizabilities parameterize the long-range dipole-dipole tensor $\mathbf{T}^{\mathrm{LR}}$ in Eq. (2.6), and the dispersion energy is obtained as the change in zero-point energy upon coupling the oscillators. Therefore, no separate short-range self-consistent screening step is required because short-/mid-range screening is already encoded in $\alpha^{\mathrm{VV}}[n]$.

In practice, MBD-NL can be integrated self-consistently into the Kohn-Sham cycle: atomic polarizabilities feed the nonlocal kernel, the resulting dispersion energy and potential are back-projected onto the electron density, and the process iterates to convergence. Despite MBD-NL requiring the evaluation of the kinetic energy density that is absent in GGA, it adds only a small overhead comparable to a fraction of one meta-GGA self-consistent cycle. Meanwhile, the Hamiltonian in Eq. (2.6) diagonalisation required for the energy evaluation scales as $\mathcal{O}(N^3)$ with a modest prefactor and benefits from efficient reciprocal-space or localized-basis implementations.

Benchmark studies show that MBD-NL markedly outperforms both pairwise and original MBD methods across diverse systems. Compared to PBE+D3 or SCS-MBD [82], it reduces binding-energy errors in molecular dimers by up to 30 %, predicts lattice constants of molecular crystals within 1 % of experiment, and accurately captures adsorption energies on surfaces. By uniting the physical rigor of many-body QHOs models with nonlocal density functionals, MBD-NL offers a parameter-free, variationally consistent description of dispersion—making it an ideal reference for electronic-structure studies and the generation of high-quality training data for MLFFs [83, 84, 85, 86, 87, 88, 89].

2.1.4 DFT Implementation Details

In practice, the one-electron orbitals are expanded in a basis set: a finite collection of functions that spans a connected subspace of the infinity-dimensional Hilbert space. The electron-nuclear interaction is represented either by the full Coulomb potential in an *all-electron* treatment or by an *effective core approximation* that replaces the singular nuclear potential plus tightly bound core electrons with a smooth, element-specific potential (pseudopotentials). The latter reproduces valence scattering while reducing the number of explicit electrons and the stiffness of the basis, thereby lowering cost. This is commonly achieved with norm-conserving pseudopotentials.

The CP2K DFT package, used in this thesis, employs the Gaussian Plane Waves (GPW) method: Kohn-Sham orbitals are expanded in localized atom-centered Gaussian functions, whereas the total electron density is mapped to an auxiliary plane-wave grid used to evaluate

Hartree and related electrostatic terms efficiently. In this work we used MOLOPT Gaussian basis sets together with norm-conserving Goedecker-Teter-Hutter (GTH) pseudopotentials, chosen consistently with the exchange-correlation functional. This combination delivers reliable valence accuracy for large periodic cells with modest memory requirements. GPW in CP2K provided favorable scaling and memory use for large systems and long trajectories, while GTH pseudopotentials with MOLOPT bases reduced cost without compromising the description of valence bonding and charge redistribution.

As a complementary route used in this work, FHI-aims treats all electrons explicitly and expands the Kohn-Sham orbitals in Numeric Atom-Centered Orbitals (NAO). Basis quality is tightened by ascending predefined "tier" levels that systematically enrich the NAO space. No pseudopotentials are used and no shape approximation is made for the Coulomb potential. This framework is well suited when core-region fidelity matters, for example in core properties, hyperfine parameters, and situations where transferability across diverse chemical environments is critical. In our work, FHI-aims provided all-electron reference results and removed concerns about pseudopotential transferability, at increased cost for heavy elements.

2.1.5 Summary of Methodological Choices

This thesis adopts a multi-level computational strategy to balance accuracy and efficiency in *ab initio* calculations. For broad exploration of configurational space, we use the PBE functional, which provides a cost-effective and robust baseline. To account for ubiquitous vdW forces, we augment PBE with the computationally efficient D3 correction for routine structural and energetic evaluations. In practice, CP2K with PBE+D3 is used for large-scale sampling and configuration generation. For high-precision benchmarks and for systems where many-body dispersion is critical—such as molecular crystals—we employ the more rigorous MBD-NL method in FHI-aims. This tiered approach enables the generation of large and representative datasets for training MLFF. For the lead-halide perovskites central to this work, PBE without SOC proves to be a pragmatic choice, as a fortuitous cancellation of errors often yields reasonable electronic structures, making it a cost-effective and robust choice for large-scale simulations.

2.2 Into Atomistic Simulations

Atomistic simulations connect microscopic interactions with macroscopic observables through ensemble and time-correlation sampling, providing quantitative thermodynamic, transport, and kinetic predictions that guide the elucidation of conformational dynamics, reaction path-

ways, adsorption and diffusion mechanisms, phase transitions, and the design of functional materials [90]. Trajectory-based atomistic simulations propagate nuclear coordinates by integrating Newton's equations with forces computed from *ab initio* electronic-structure methods, empirical Force Fields (FFs), or MLFFs[91, 92, 11]. Thermostat and barostat algorithms control intensive variables to enforce target equilibrium conditions, enabling sampling of canonical (NVT) or isothermal–isobaric (NPT) ensembles and thereby capturing temperature- and pressure-dependent regions of configurational space [93, 94, 95, 96, 97]. To quantify phase stability, defect-formation energetics, or reaction equilibria, one evaluates differences in the appropriate thermodynamic potentials (Helmholtz in NVT and Gibbs in NPT). Thermodynamic integration estimates these free-energy differences by introducing a continuous coupling parameter that transforms a reference state into a target state and by numerically integrating ensemble-averaged generalized forces along this reversible path [98, 99, 100] [98, 99, 100].

In MLFF workflows, atomistic simulations also serve as sampling schemes that systematically select a finite set of data points, configurations, or trajectories from a much larger, often continuous, phase space so as to ensure that subsequent analysis or model training adequately reflects the underlying phenomena of interest [65, 101, 102].

The following subsections describe the atomistic simulation strategies employed in this work for training MLFF, exploring free-energy surfaces, and investigating the evolution of the geometric and electronic structures of materials.

2.2.1 Molecular Dynamics

Molecular Dynamics (MD) provides a time-resolved, atomistic trajectory by numerically integrating Newton's equations of motion on a given PES [103, 104, 105]. Whether driven by empirical FFs, by *ab initio* forces, or by MLFFs, MD enables direct exploration of atomic vibrations, diffusion events and collective fluctuations that underlie finite-temperature material behavior [106, 107, 108, 109, 110].

Classical MD evolves in the $6N$-dimensional phase-space of atomic positions $\mathbf{R}$ and momenta $\mathbf{P}$ according to Hamiltonian $H(\mathbf{R}, \mathbf{P}) = \sum_i \frac{\mathbf{p}_i^2}{2m_i} + E(\mathbf{R})$, where $E(\mathbf{R})$ is the PES defined by the method used to calculate energies and forces. Hamilton's equations define the flow on this space and imply preservation of phase-space volume (Liouville's theorem), which underpins correct equilibrium sampling and stable long-time dynamics.

Time evolution can be written compactly using the Liouville operator $i\mathcal{L}$,

$$A(\mathbf{R}(t), \mathbf{p}(t)) = e^{i\mathcal{L}t} A(\mathbf{R}(0), \mathbf{p}(0)), \tag{2.8}$$

for any phase-space observable $A(\mathbf{R}, \mathbf{p})$, as $\frac{d}{dt}A = i\mathcal{L}A$, with

$$i\mathcal{L} = \sum_{i}^{3N} \left((\mathbf{p}_i/m_i) \cdot \nabla_{\mathbf{R}_i} - \nabla_{\mathbf{R}_i} U \cdot \nabla_{\mathbf{p}_i} \right). \tag{2.9}$$

This operator generates the Hamiltonian flow and enables operator-splitting integrators. Symplectic schemes such as velocity-Verlet preserve the phase-space measure and can be interpreted as exactly integrating a nearby "shadow" Hamiltonian $\tilde{H} = H + \mathcal{O}(\Delta t^2)$, ensuring bounded energy error and stable long-time dynamics [111, 112, 113, 114, 115]. With velocity-Verlet, positions and velocities advance as

$$\begin{aligned} \mathbf{R}(t + \Delta t) &= \mathbf{R}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\Delta t^2, \\ \mathbf{v}(t + \Delta t) &= \mathbf{v}(t) + \frac{1}{2}\left(\mathbf{a}(t) + \mathbf{a}(t + \Delta t)\right)\Delta t, \\ \mathbf{a}(t) &= -\frac{1}{m}\nabla_{\mathbf{R}} E(\mathbf{R}(t)), \end{aligned} \tag{2.10}$$

where $\mathbf{R}(t)$ is the atomic position vector, $\mathbf{v}(t) = \mathbf{p}(t)/m$ is the velocity vector, and $\mathbf{a}(t)$ is the acceleration vector obtained from the interatomic forces $-\nabla_{\mathbf{R}} E$ at time t divided by the atomic mass m.

The choice of time step Δt is dictated by the highest vibrational frequency in the system: too large a step leads to integration drift, while too small a step unnecessarily increases computational cost. In practice, Δt is chosen as a small fraction of the fastest period $T_{\min} = 1/(c \cdot \tilde{\nu}_{\max})$, typically $\Delta t \lesssim T_{\min}/10$. For an H-X stretch near $\tilde{\nu}_{\max} \approx 3500$ cm^{-1} (e.g., O-H), $T_{\min} \approx 9.5$ fs so $\Delta t \approx 0.5$-1 fs. For heavy-atom modes such as Cs-Pb-I lattices with highest optical phonons $\tilde{\nu}_{\max} \approx 80$-200 cm^{-1}, $T_{\min} \approx 160$-500 fs so the $T/10$ bound gives $\Delta t \approx 16$-50 fs, although in practice one still uses $\sim$1-5 fs to avoid resonances and maintain accuracy. [112, 116, 117]

Despite its theoretical elegance and practical utility, classical MD simulations have two fundamental limitations: the accessible time scales (typically nanoseconds to microseconds) restrict the observation of rare events or slow relaxation processes, and the classical-mechanical treatment neglects quantum effects such as zero-point energy and tunneling that can be significant for light atoms or low-temperature phenomena [106, 112]. Both limitations motivate the development of enhanced sampling techniques and interatomic forces generative methods (including MLFFs) to complement classical MD [118, 119, 120, 121, 122, 123].

Production runs then sample phase space under constant-energy (NVE), constant-tempera-

ture (NVT) or constant-temperature-pressure (NPT) conditions, as dictated by the properties of interest. Analysis of these trajectories yields not only structural information — radial distribution functions, mean-square displacements, angle distributions — but also dynamical observables such as diffusion coefficients, vibrational density of states and time-correlation functions [124].

Beyond structural and dynamical observables, MD trajectories furnish ensembles of configurations that sample thermally accessible regions of configurational space. When coupled to an *ab initio* method, each configuration yields reference energies and forces. MD-generated configurations are particularly valuable for MLFF training compared to alternative sampling approaches (such as random structure generation or grid-based sampling) for several key reasons: (1) they naturally concentrate sampling density in physically relevant regions that the system actually visits; (2) they include correlated atomic motions absent in independent sampling schemes; (3) they capture both near-equilibrium fluctuations and rare transition events at appropriate frequencies; (4) and they provide a continuous series of configurations that spans the configuration space along realistic pathways rather than disconnected snapshots. Consequently, short *ab initio* MD runs are a natural primary source of training data for MLFFs targeting finite-temperature behavior, as they efficiently represent the regions of configuration space most relevant to the system's actual thermodynamic and kinetic properties.

2.2.2 Ensemble Control: NVT and NPT Algorithms

MD simulations derive their predictive power from the faithful sampling of statistical-mechanical ensembles. In the canonical (NVT) ensemble, a thermostat continuously exchanges thermal energy between the simulated system and an external heat bath, ensuring that the temperature remains fixed on average throughout the trajectory. In the isothermal-isobaric (NPT) ensemble, in addition to the thermostat, a barostat couples the system to a pressure reservoir, allowing the volume to fluctuate so that both temperature and pressure are maintained at their target values. Together, these ensemble-specific controls guarantee that the simulation samples the correct canonical distribution for the desired thermodynamic conditions.

Temperature control in the NVT ensemble can be achieved via either deterministic "extended-system" thermostats or stochastic schemes [95, 96, 97].

The deterministic thermostat used in this work is Nosé–Hoover chain thermostat [125]. It introduces an extended Hamiltonian with an auxiliary variable (or a chain of variables) to mimic a heat bath. In its simplest (single-chain) form, the equations of motion read

$$\dot{\mathbf{r}}_i \;=\; \frac{\mathbf{p}_i}{m_i}, \qquad \dot{\mathbf{p}}_i \;=\; \mathbf{F}_i \;-\; \xi\,\mathbf{p}_i, \qquad \dot{\xi} \;=\; \frac{1}{Q}\Big(\sum_i \frac{\mathbf{p}_i^2}{m_i} - g\,k_B T_0\Big), \tag{2.11}$$

where $\mathbf{p}_i = m_i \mathbf{v}_i$ is the momentum of particle i, ξ is the friction (thermostat) variable, Q is its "mass", g is the number of degrees of freedom, and T_0 is the target temperature. In practice, Nosé-Hoover chains produce smooth, continuous trajectories that preserve the underlying Hamiltonian dynamics and, in principle, sample the NVT ensemble exactly. However, because they rely on deterministic coupling to a fictitious "heat-bath" variable, they can fail to explore phase space efficiently in systems with rough PES or widely separated timescales—i.e., their ergodicity may be limited, leading to "sticky" trajectories that do not fully sample all relevant configurations [126].

By contrast, the Langevin thermostat continuously modifies each particle's motion through a combination of frictional damping and stochastic forcing [127, 128]:

$$m_i \dot{\mathbf{v}}_i = \mathbf{F}_i - \gamma_i m_i \mathbf{v}_i + \boldsymbol{\eta}_i(t), \tag{2.12}$$

where γ_i is a friction coefficient (with dimensions of inverse time) and $\boldsymbol{\eta}_i(t)$ is a random force satisfying

$$\langle \eta_{i,\alpha}(t)\, \eta_{j,\beta}(t') \rangle \;=\; 2\,\gamma_i\, m_i\, k_B T_0\, \delta_{ij}\, \delta_{\alpha\beta}\, \delta(t - t'). \tag{2.13}$$

This ensures that (1) the mean kinetic energy per degree of freedom is $\frac{1}{2} k_B T_0$ and (2) the fluctuations around this mean obey the canonical distribution of kinetic energy statistics. The random "kicks" from $\boldsymbol{\eta}_i(t)$ help the system escape local minima and sample configurational space more thoroughly, improving ergodicity, albeit at the cost of slightly modified dynamical correlation functions compared to fully Newtonian motion [129, 130]. The $\boldsymbol{\eta}_i(t)$ is controlled by γ_i balancing how strongly the thermostat enforces the target temperature against how closely the system follows natural (Newtonian) dynamics: a small γ_i leads to weak coupling and long-time correlations, while a large γ_i overdamps the motion and can lead to unphysical behavior.

In practice, the choice of thermostat depends on the system and the properties of interest. Nosé-Hoover chains are often preferred for their exact sampling of the NVT ensemble and preservation of Hamiltonian structure. At the same time, while Langevin thermostats are favored for their robustness in exploring rough or high-dimensional PES.

However, when one wishes to sample the NPT ensemble, temperature control alone is not sufficient: pressure fluctuations must also be accommodated by treating the simulation

cell as a dynamical variable. Pressure control in the NPT ensemble is therefore achieved by embedding the simulation cell itself into the dynamics via an extended-system Lagrangian. In this work, the extended-system formulation of Melchionna et al [131] is used, which combines the Nosé-Hoover thermostat with a symmetric barostat. Denoting the scaled (or "internal") coordinates by $\mathbf{s}_i$, so that is given by

$$\begin{aligned}\mathcal{L} = &\sum_{i=1}^{N} \frac{1}{2} m_i \, \dot{\mathbf{s}}_i^T \, \mathbf{G} \, \dot{\mathbf{s}}_i - U\big(\{\mathbf{r}_i\}\big) + \frac{1}{2} Q \, \dot{\eta}^2 - g \, k_B \, T \, \eta \\ &+ \frac{1}{2} W \, \mathrm{Tr}\big(\dot{\mathbf{h}}^T \dot{\mathbf{h}}\big) - P_{\mathrm{ext}} \, \det(\mathbf{h}) \, ,\end{aligned} \tag{2.14}$$

here, $\mathbf{r}_i = \mathbf{h}\,\mathbf{s}_i$ defines the real-space positions $\mathbf{r}_i$ in terms of scaled coordinates $\mathbf{s}_i$ and the cell matrix $\mathbf{h}$, with $\mathbf{G} = \mathbf{h}^T\mathbf{h}$ the associated metric tensor. The thermostat is introduced via the Nosé-Hoover coordinate $s = e^{\eta}$, whose kinetic term $\frac{1}{2} Q \, \dot{\eta}^2$ carries fictitious mass Q and whose logarithmic "entropy" variable η appears in the potential $-\, g \, k_B T \, \eta$, with g the total number of degrees of freedom. Pressure fluctuations are controlled by the symmetric barostat: the term $\frac{1}{2} W \, \mathrm{Tr}(\dot{\mathbf{h}}^T \dot{\mathbf{h}})$ endows the cell with inertia W, setting a natural barostat frequency $\omega_b \sim 1/\sqrt{W}$ (typically chosen so that volume oscillations occur on a few hundred femtoseconds to picosecond time scale), while $-P_{\mathrm{ext}} \, \det(\mathbf{h})$ drives the simulation toward the external pressure P_{ext}.

Varying Eq. 2.14 yields equations of motion in which the usual interparticle force $-\nabla_i U$ acquires an additional coupling term $\mathbf{G}^{-1}\dot{\mathbf{G}}\,\dot{\mathbf{r}}_i$, and the cell matrix $\mathbf{h}$ evolves under the instantaneous internal virial pressure relative to P_{ext}. Physically, the barostat makes the box act like a massive piston: if the internal pressure exceeds the target, the box expands; if it falls below, the box contracts. Compared with the original Parrinello-Rahman formulation, the symmetric form used here avoids spurious rotational torques on the cell.

Martyna-Tuckerman-Klein (MTK) scheme refines this approach by introducing Nosé-Hoover chains for both thermal and barostatic degrees of freedom [125, 132]. In addition to the particle coordinates and cell matrix, one adds auxiliary variables $\{\eta_j\}$ and $\{\xi_j\}$ with associated masses Q_j^{th} and Q_j^{bar} [133]. These variables exchange energy with the system, enforcing a canonical distribution of kinetic energy and volume fluctuations. The extended Lagrangian acquires terms of the form

$$\frac{1}{2} Q_j^{\mathrm{th}} \, \dot{\eta}_j^2 - g \, k_B T \, \eta_j \quad \text{and} \quad \frac{1}{2} Q_j^{\mathrm{bar}} \, \dot{\xi}_j^2 - f \, k_B T \, \xi_j \tag{2.15}$$

for thermostat and barostat chains of lengths g and f, respectively. The resulting equations of motion preserve time-reversibility and admit a symplectic integration scheme, ensuring

long-term stability and exact sampling of the isothermal-isobaric ensemble. In practice, the thermostat-chain masses determine how tightly kinetic fluctuations are controlled, while the barostat-chain masses regulate the coupling between cell fluctuations and the pressure reservoir. By coupling the symmetric barostat to a Nosé-Hoover chain, MTK guarantees that both the shape and volume of the cell evolve without spurious drift, faithfully reproducing the canonical distribution at the desired T and P .

Careful selection of thermostat and barostat parameters is essential: thermostat relaxation times on the order of 0.1-1 ps balance energy exchange with the preservation of realistic dynamics, and barostat masses must be tuned to match the natural oscillation frequencies of the simulation cell. By tuning those parameters, one ensures that MD trajectories faithfully reproduce equilibrium fluctuations, enabling reliable computation of thermodynamic observables and providing physically meaningful ensembles for downstream analysis or MLFF training.

2.2.3 Thermodynamic Integration

Thermodynamic Integration (TI) provides a rigorous route to free-energy differences by integrating ensemble averages of a suitable derivative of the system Hamiltonian with respect to a coupling parameter [98, 134]. Consider a family of Hamiltonians $H(\lambda)$ that continuously interpolates between an initial state at $\lambda = 0$ and a final state at $\lambda = 1$ [135, 136]. The Helmholtz free-energy difference is then given by

$$\Delta F = F(1) - F(0) = \int_0^1 \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda, \tag{2.16}$$

where $\langle \cdot \rangle_\lambda$ denotes an ensemble average performed at fixed λ. In practice, one selects a discrete set of λ-windows, performs MD (or Monte Carlo) simulations under NVT or NPT conditions at each window, and computes the mean force $\langle \partial H / \partial \lambda \rangle_\lambda$. Numerical quadrature—typically the trapezoidal or Simpson's rule—then reconstructs ΔF with systematically controllable accuracy.

2.2.4 Summary of Methodological Choice

This chapter has outlined the atomistic simulation techniques employed throughout this work. Our methodological choices were guided by the specific requirements of developing accurate MLFF and characterizing molecular and material systems at finite temperatures.

For MD simulations, we employed velocity-Verlet–based integrators across the used ensembles, with time steps of 1 fs for hydrogen-containing systems and 4 fs for heavy-atom

systems to balance accuracy and efficiency. These choices ensure stable long-time dynamics while maintaining sufficient temporal resolution to capture the relevant vibrational modes in our systems.

Temperature control in NVT simulations was implemented through two complementary approaches: Nosé-Hoover chains for preserving Hamiltonian structure and sampling accuracy when generating training data, and Langevin dynamics for enhanced exploration of rough potential energy surfaces, particularly for interfaces. For NPT simulations requiring volume fluctuations, we utilized the MTK scheme with symmetric barostats.

TI was strategically applied to compute free-energy differences across complex landscapes, such as molecule-surface interactions.

2.3 Machine-Learning Force Fields

MLFF models provide computationally efficient alternatives to electronic-structure calculations, achieving this by learning the functional relationship between atomic coordinates and the associated energies and forces. Trained on reference data from *ab initio* calculations and experimental measurements, MLFFs reproduce reference energies and forces with quantum-chemical (QC) accuracy while significantly reducing computational costs, thereby enabling large-scale and long-time atomistic simulations that would otherwise be prohibitive. Over the past decade, these models have evolved from kernel-based regressors (Smooth Overlap of Atomic Positions (SOAP)/Gaussian approximation potential (GAP) [137, 14, 138], Symmetric Gradient-Domain Machine Learning (sGDML) [15, 17, 16], Faber-Christensen-Huang-Lilienfeld (FCHL)19* [139, 140]) and high-dimensional neural networks (Behler-Parrinello [11], SchNet [19]) to modern equivariant MPNN architectures (NequIP [21], Allegro [22], E^2GNN[141], MACE [24, 25], SO3krates [142, 23]) and emerging foundation-models (SO3LR [143], MACE-OFF [144], MACE-MP [145]). In this section, we examine the foundational principles behind each MLFF model class in this thesis and review the evaluation frameworks and benchmark datasets used to quantify their accuracy, robustness, and transferability.

2.3.1 Rise of Learned Potentials

Empirical FFs approximate the PES by means of analytic, physically motivated functions whose parameters are fitted to reproduce experimental measurements or reference *ab initio* energies and forces [146, 147]. A typical bonded term includes harmonic or Morse-type potentials for bond stretches and angular bends, while dihedral rotations are captured through periodic

torsional functions [148, 149, 147, 150]. Nonbonded interactions combine Lennard-Jones or Buckingham pair potentials for vdW forces with Coulomb's law for electrostatics, often augmented by fixed atomic charges or simple polarization schemes [151, 146, 152]. The fixed functional forms and limited parameter sets render empirical FFs extremely efficient—force and energy evaluations scale nominally as $\mathcal{O}(N)$ with system size—enabling simulations of millions of atoms over microsecond timescales [124, 150]. However, the reliance on fixed connectivity and pre-tabulated parameters limits their transferability and prevents the spontaneous formation or breaking of chemical bonds [153, 146]. Although reactive empirical models (e.g. bond-order or ReaxFF) can handle changing bonding topology [154, 155], they require extensive, system-specific parameterization and often degrade outside the fitted domain.

In contrast, *ab initio* potentials compute energies and forces by solving the electronic Schrödinger equation within an approximate framework (most commonly DFT). *Ab Initio* Molecular Dynamics (AIMD)s therefore capture bond breaking and formation, charge transfer, and polarization without recourse to predetermined connectivity. The principal limitation of this approach lies in its steep $\mathcal{O}(N^3)$ (or worse) scaling and significant prefactor, which restricts practical simulations to hundreds of atoms and tens of picoseconds. The existing linear-scaling DFT schemes reduce the formal cost when the electronic structure is localized, yet trade generality for computational efficiency.

From a modeling perspective, empirical FFs excel in large-scale sampling of equilibrium fluctuations—phase equilibria, conformational ensembles, and long-time diffusion—provided the chemistry remains within the scope of the parameter set. *Ab initio* potentials, by contrast, deliver a transferable, reactive description of chemical processes, crucial for exploring reaction pathways, defect formation and charge-transfer phenomena, albeit at a high computational cost. The choice between these paradigms thus reflects a trade-off between efficiency and fidelity: empirical FFs offer unparalleled speed for routine simulations, while *ab initio* methods supply the accuracy and generality necessary to capture true quantum effects.

The complementary strengths of empirical and *ab initio* approaches have motivated hybrid strategies—QM/MM and machine-learning surrogates—that seek to combine the accuracy of electronic-structure theory with the efficiency of analytic force fields. In this thesis, the limitations of both traditional FFs and direct DFT-based dynamics are addressed by MLFF, which learn flexible functional forms from reference data and recovers QC accuracy at a fraction of the cost.

2.3.2 Kernel-Based MLFF

Kernel methods, such as Kernel Ridge Regression (KRR) and Gaussian Process Regression (GPR), provide a non-parametric framework for learning PES by expressing the target function as a linear combination of similarity measures (kernels) centered on training configurations. One can define a kernel function:

$$k(x_i, x_j) = \langle D(x_i), D(x_j) \rangle, \tag{2.17}$$

that quantifies the similarity between two descriptor vectors $D(x_i)$ and $D(x_j)$. For atomistic systems, $D(x)$ encodes the spatial arrangement of atoms and their identities — either for the entire system or within a local atomic environment — by mapping coordinates into a fixed-length vector of features. For MLFF applications, a descriptor must rigorously uphold the system's underlying physical symmetries:

- **Translational invariance:** The descriptor must not change if the entire system is translated, $D(mathbf x_i + \mathbf{T}) = D(\mathbf{x}_i)$, for any global shift $\mathbf{T}$.
- **Rotational invariance:** The descriptor must not change if the system is rotated, $D(Rx_i) = D(x_i)$, for any rotation, $R \in SO(3)$. Here $SO(3)$ denotes the special orthogonal group in three dimensions, the group of orientation-preserving rotations of $\mathbb{R}^3$ under composition.

Additionally, the descriptor must be invariant under any permutation of the indices of atoms within the same chemical species: $D(\sigma x_i) = D(x_i)$for all such permutations σ. While one can sidestep this by enforcing a fixed atom ordering, doing so drastically limits extensibility: it ties the descriptor to a specific system size and composition, makes it cumbersome (or impossible) to add or remove atoms without redefining the entire ordering, and prevents seamless transfer of the same feature map to larger or differently indexed structures.

The similarity between descriptors is quantified by the kernel by implicitly mapping them into a (typically high- or infinite-dimensional) feature space, called a reproducing kernel Hilbert space, and evaluating the inner product $\langle D(x_i), D(x_j) \rangle$ in that space. When the predicted values are obtained at a new point x^{new} using KRR, it is expressed as a linear combination of kernels that quantify the similarity between new and M reference configurations:

$$\mathbf{y}(x^{\text{new}}) = \sum_{j=1}^{M} \alpha_j \, k(x^{\text{new}}, x_j). \tag{2.18}$$

where α_j are regression weights that are learned from reference data by minimizing a regu-

larized squared-error loss:

$$\min_{\boldsymbol{\alpha}} \|\mathbf{K}\,\boldsymbol{\alpha} - \mathbf{y}\|_2^2 + \lambda\,\boldsymbol{\alpha}^\top \mathbf{K}\,\boldsymbol{\alpha}, \tag{2.19}$$

where $\mathbf{K}_{ij} = k(x_i, x_j)$ is the kernel matrix that contains the pairwise similarity evaluations for each pair of configurations in training data, $\mathbf{y}$ contains the target values (e.g., energies or forces), and $\lambda > 0$ is a regularization parameter (typically chosen by cross-validation) that controls overfitting. The regression coefficients $\boldsymbol{\alpha}$ are then obtained in closed form by solving the linear system analytically:

$$\boldsymbol{\alpha} \;=\; \big(\mathbf{K} + \lambda\,\mathbf{I}\big)^{-1}\mathbf{y}. \tag{2.20}$$

The explicit analytical expression reveals how the system's behavior depends on its parameters, allowing one to understand and predict its response without the need for extensive simulations.

To guarantee a unique solution for the weight vector $\boldsymbol{\alpha}$, the kernel matrix $\mathbf{K}$ must be symmetric and positive semidefinite. In particular, for all inputs x_i, x_j and any real coefficients $\{c_i\}$, one requires

$$k(x_i, x_j) = k(x_j, x_i), \qquad \text{and} \qquad \sum_{i,j} c_i\,c_j\,k(x_i, x_j) \;\geq\; 0. \tag{2.21}$$

Although many kernel functions have been proposed in the literature, we restrict our discussion to those actually implemented in the architectures employed in this work.

GPR extends the KRR into a fully Bayesian framework by placing a Gaussian Process ($\mathcal{GP}$) prior directly over the latent function $f(x)$, parameterized by a mean function $m(x)$ (often set to zero) and a covariance (kernel) function $k(x, x')$:

$$f(x) \sim \mathcal{GP}\big(m(x),\; k(x, x')\big)\,, \tag{2.22}$$

Before to observing any data, this prior asserts that the function values $\{f(x_i)\}$ at any finite collection of inputs $\{x_i\}$ follow a joint multivariate normal distribution. The $\mathcal{GP}$ prior encodes our beliefs about the function's smoothness, amplitude, and other structural properties, thereby forming the basis for Bayesian inference, where training data are used to update these beliefs and yield a posterior $\mathcal{GP}$ that quantifies predictive uncertainty [156]. Upon observing noisy targets:

$$y_i = f(x_i) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2), \tag{2.23}$$

the joint distribution of training outputs $\mathbf{y}$ and the function values at a new input x^{new} is

Gaussian:

$$\begin{pmatrix} \mathbf{y} \\ f(x^{\text{new}}) \end{pmatrix} \sim \mathcal{N}\left(\mathbf{m}(x),\ \begin{bmatrix} \mathbf{K} + \sigma^2 \mathbf{I} & \mathbf{k}_* \\ \mathbf{k}_*^\top & k(x^{\text{new}}, x^{\text{new}}) \end{bmatrix} \right), \tag{2.24}$$

where $\mathbf{K}_{ij} = k(x_i, x_j)$ is the $M \times M$ kernel matrix and cross-covariance vector

$$\mathbf{k}_* = [\, k(x^{\text{new}}, x_j) \,]_{j=1}^{M}. \tag{2.25}$$

Conditioning on the observed outputs $\mathbf{y}$ updates the joint multivariate normal distribution at a new input to reflect the actual data, yielding a posterior distribution for $f(x^{\text{new}})$:

$$\mathbb{E}\left[f(x^{\text{new}}) \mid \mathbf{y} \right] = \mathbf{k}_*^\top \left(\mathbf{K} + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{y}, \tag{2.26}$$

which can be interpreted as a weighted sum of the training targets. Here, the kernel matrix $\mathbf{K}$ encodes all pairwise similarities among training inputs, while the vector $\mathbf{k}_*$ collects the similarities between the new input x^{new} and each training point. Each weight ($\boldsymbol{\alpha} = \left(\mathbf{K} + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{y}$) in the sum measures how strongly a past observation should influence the prediction, based on its similarity to x^{new}. Algebraically, the posterior mean is determined by the same regularized optimization that defines KRR, highlighting their formal equivalence. Therefore, both KRR and GPR exhibit the same quadratic memory and cubic training-time scaling in the number of training points M, i.e., $\mathcal{O}(M^2)$ and $\mathcal{O}(M^3)$, respectively for analytical solutions. Limiting the number of training points to a few thousand is necessary to keep the training time manageable, which is a significant limitation for large datasets. At evaluation time, prediction at a new point costs linear time in the number of training points $\mathcal{O}(M)$, as it requires computing the kernel between the new point and each training point.

Nevertheless, GPR goes further by also providing a posterior variance

$$\mathrm{Var}\left[f(x^{\text{new}}) \mid \mathbf{y} \right] = k(x^{\text{new}}, x^{\text{new}}) - \mathbf{k}_*^\top \left(\mathbf{K} + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{k}_*, \tag{2.27}$$

providing principled uncertainty estimates and enabling hyperparameter learning (e.g. kernel length-scales and noise variance) via maximization of the log-marginal likelihood

$$\log p(\mathbf{y} \mid X, \theta) = -\tfrac{1}{2}\, \mathbf{y}^\top (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{y} - \tfrac{1}{2}\, \log \det(\mathbf{K} + \sigma^2 \mathbf{I}) - \tfrac{M}{2} \log(2\pi). \tag{2.28}$$

This objective embodies an automatic Occam's-razor: the first term rewards explanatory power, while the second penalizes overly flexible models. Once the optimal hyperparameters are obtained, they fully determine the posterior predictive mean and variance at new inputs, enabling both accurate predictions and principled uncertainty quantification.

By recasting both the weight fitting in KRR and the hyperparameter tuning in GPR as gradient-based learning problems, one can entirely avoid expensive kernel-matrix inversions and brute-force grid searches. In KRR, the regression weights are found by minimizing a regularized loss with respect to the regression weights (**ff**) using iterative solvers that rely only on matrix—vector products, which dramatically lowers both memory and time requirements for large number of training points (M). In GPR, kernel hyperparameters (length-scales, amplitudes, noise variance, etc.) are optimized by maximizing the log-marginal likelihood via gradient-based methods—leveraging analytic derivatives to home in on optimal values and using multiple restarts to escape local optima. This unified, gradient-driven workflow naturally supports warm-starts, mini-batch or stochastic updates, and seamless integration with automatic-differentiation frameworks, thereby boosting the scalability, flexibility, and automation of kernel methods for complex materials-simulation tasks.

SOAP/GAP

The GAP framework employs GPR with the SOAP descriptor to capture many-body interactions in a rotationally and permutationally invariant form [137, 14, 138]. GAP decomposes the total energy of an N-atom system into a sum of atomic contributions:

$$E = \sum_{i=1}^{N} \varepsilon_i = \sum_{i=1}^{N} \sum_{j=1}^{M} \alpha_j \, K(x_i, x_j) \tag{2.29}$$

where M is the number of "inducing" environments — representative atomic neighbourhoods chosen from the full training set to serve as support points in the sparse GPR approximation — selected from the full training set — typically $M \ll N$.

GAP leverages two key simplifications to remain both accurate and efficient:

- **Sparse Gaussian-process regression:** Rather than using all atomic environments in the training set, GAP selects only $M_s \ll N$ representative "inducing" environments. These inducing points serve as the basis for the regression model, effectively compressing the training data while preserving the essential information. This sparse approximation reduces the computational complexity from $\mathcal{O}(N^3)$ to $\mathcal{O}(N\,M_s^2)$, making it feasible to train on large datasets. The selection of inducing environments is typically performed using clustering algorithms or other sparsification techniques that identify the most informative atomic configurations.

- **Composite many-body kernel:** The kernel function in GAP measures similarities between atomic environments and systematically incorporates many-body interactions

beyond simple pairwise terms. This approach enables GAP to capture the complex bonding patterns and interactions at different length scales that determine material properties. The mathematical implementation of this kernel relies on the SOAP descriptor, which will be detailed below.

Once α_j have been determined, the energy of any new configuration follows immediately from (2.29), and atomic forces are obtained by

$$F_{i\alpha} = -\frac{\partial E}{\partial r_{i\alpha}} = -\sum_{j=1}^{M_s} \alpha_j \frac{\partial K(x_i, x_j)}{\partial r_{i\alpha}}, \tag{2.30}$$

The required derivatives $\partial K/\partial r_{i\alpha}$ are computed analytically by differentiating the kernel arguments with respect to atomic coordinates, ensuring that forces are both consistent with the energy model and evaluated at the same cost-scaling as the energy itself.

SOAP descriptor encodes local atomic environments via a high-dimensional power spectrum whose computational complexity scales as $\mathcal{O}\left(n_{\max}^2 \, l_{\max} \, S_n^2\right)$, where $n_{\max}$ is the maximum number of neighbors, $l_{\max}$ is the maximum angular momentum, and S_n is the number of radial basis functions. This leads to a high-dimensional descriptor that captures both local and nonlocal interactions, but at the cost of significant computational and memory overhead.

In practice, each atomic environment x_i is represented by a continuous, Gaussian-smeared neighbor density:

$$\rho_i(\mathbf{r}) = \sum_{j \in \mathcal{N}(i)} \exp\Big(-\frac{\|\mathbf{r}-\mathbf{r}_{ij}\|^2}{2\sigma^2}\Big), \tag{2.31}$$

where $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ is the distance vector between atoms i and j, σ is the smearing width, and the sum runs over neighbors within a cutoff r_{cut}. This density is expanded in a product basis of radial functions $g_n(r)$ and spherical harmonics $Y_{lm}(\hat{\mathbf{r}})$:

$$c^i_{nlm} = \int \rho_i(\mathbf{r}) \, g_n(r) \, Y_{lm}(\hat{\mathbf{r}}) \, \mathrm{d}\mathbf{r}, \tag{2.32}$$

where $r = \|\mathbf{r}\|$ is the radial distance and $\hat{\mathbf{r}} = \mathbf{r}/r$ is the corresponding unit vector. A rotationally invariant power spectrum is then formed:

$$p^i_{nn'l} = \sum_{m=-l}^{l} \left(c^i_{nlm}\right)^* c^i_{n'lm}, \tag{2.33}$$

collecting all (n, n', l) channels into the descriptor vector $\mathbf{p}^i$. Finally, the SOAP kernel between two environments x_i and x_j is defined as the normalized dot product of their power spectra,

optionally raised to an exponent ζ:

$$K_{\mathrm{SOAP}}(x_i, x_j) = \left[\frac{\mathbf{p}^i \cdot \mathbf{p}^j}{\|\mathbf{p}^i\| \, \|\mathbf{p}^j\|} \right]^{\zeta}. \tag{2.34}$$

ζ is typically set to 4 and was found to give good discrimination between different environments by accentuating high-overlap (i.e. very similar) pairs. The SOAP descriptor is rotationally invariant and through symmetrization enforces permutational invariance, making it well-suited for multi-element systems. Leveraging these invariances and its rich, high-dimensional representation, the GAP framework — when employing SOAP as its kernel — inherits the descriptor's computational cost, scaling as $\mathcal{O}\left(n_{\mathrm{max}}^2 \, l_{\mathrm{max}} \, S_n^2\right)$, in addition to the $\mathcal{O}(M^3)$ compute and memory costs of $\mathcal{O}(M^2)$ for the kernel matrix $\mathbf{K}$ inverse during training with respect to the number of training configurations M. Therefore, SOAP/GAP is typically applicable to moderately sized datasets and making inference increasingly expensive as the reference database grows.

sGDML

Gradient-Domain Machine Learning (GDML) employs KRR to learn molecular force fields by fitting the vector of atomic forces directly, constructing them as the gradient of a single latent PES so that energy conservation is enforced by construction [15]. Given M reference configurations $\{x_i\}_{i=1}^{M}$ of an N-atom system and their associated Cartesian force vectors, the block-Hessian kernel:

$$\mathbf{K}_{\mathrm{Hess}}(k)_{ij} \;=\; \nabla_x \nabla_{x'}^T \, k(x_i, x_j) \quad (i, j = 1, \ldots, M), \tag{2.35}$$

where $k(x_i, x_j)$ is a scalar kernel on molecular descriptors. The regularized gradient-domain regression problem is

$$\left[\mathbf{K}_{\mathrm{Hess}}(k) + \lambda \, \mathbf{I}\right] \, \boldsymbol{\alpha} \;=\; \nabla E \;=\; -\,\mathbf{F}, \tag{2.36}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^{M \times 3N}$ collects the learned coefficients, λ is a regularization hyperparameter, and $\mathbf{F}$ concatenates all force components.

The predicted force on a new configuration x is then

$$\hat{\mathbf{F}}(x) \;=\; -\sum_{i=1}^{M} \sum_{k=1}^{3N} \alpha_{i,k} \, \frac{\partial}{\partial x_k} \, \nabla k\left(x, x_i\right), \tag{2.37}$$

and the corresponding PES $\hat{U}(x)$ is recovered (up to an additive constant U_0) by path-

independent integration of $\hat{\mathbf{F}}(x)$:

$$\hat{U}(x) = U_0 \; - \; \int_{\mathbf{x}_0}^{\mathbf{x}} \hat{\mathbf{F}}(\mathbf{x}') \cdot d\mathbf{x}' = U_0 \; + \; \sum_{i=1}^{M} \sum_{d=1}^{3N} \alpha_{i,d} \, \frac{\partial \, k(x, \, x_i)}{\partial x_d} \, , \tag{2.38}$$

where the additive constant U_0 is restored from training data by matching the average energy of the predicted PES to that of the reference data.

In the GDML framework, each molecular geometry is represented by a global inverse-distance descriptor, defined as the vector of $1/R_{ij}$ for all unique atom-atom pairs $\{i < j\}$. This descriptor is inherently translationally and rotationally invariant and, by construction, contains all pairwise distance information needed to capture long-range effects. However, tailored with KRR method, it treats all atoms as distinct, leading to a non-symmetric kernel that does not enforce permutational invariance or any kind of permutational symmetry. It leads to challenges for model extendability, transferability, data efficiency, and accuracy.

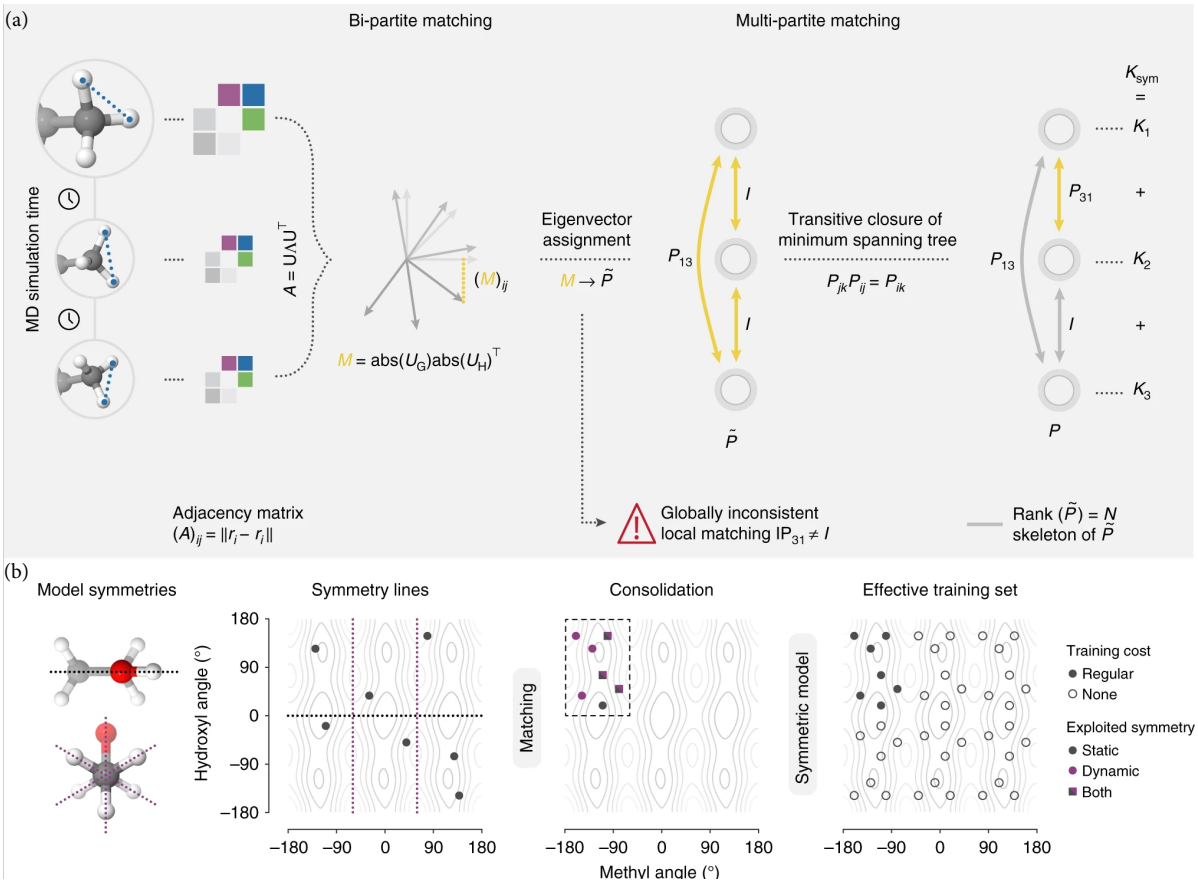


Figure 2.3 (a) The schematic illustration of symmetry search algorithm adopted in sGDML. (b) The recovered permutations effectively extend the training dataset by mapping each configuration to its symmetric counterparts. Adapted from Ref. [17].

To address data efficiency and accuracy limitations, the sGDML framework extends the original GDML model by incorporating both global rigid space-group symmetries (e.g.

point-group operations) and fluxional, non-rigid automorphisms (e.g. methyl-group rotations, torsional displacements) directly discovered from the training trajectories via a fully data-driven algorithm [17, 16], as illustrated in Figure 2.3. To extract these permutational symmetries, each molecular conformation x_i is first represented as a weighted, undirected graph G_i with adjacency matrix $A^{(i)}$, where nodes correspond to atoms and edge weights encode inverse interatomic distances:

$$A^{(i)}_{uv} = \frac{1}{\|r_u - r_v\|}. \tag{2.39}$$

Two configurations G_i and G_j are graph-isomorphic if there exists a permutation matrix $P \in \mathcal{P}_N$ such that

$$A^{(i)} = P\, A^{(j)}\, P^\top, \tag{2.40}$$

reflecting the relabeling of atom indices under the symmetry operation. To infer each P, sGDML applies Umeyama's eigendecomposition approach to weighted graph matching, which computes eigenvalue decompositions:

$$A^{(i)} = U^{(i)}\, \Lambda^{(i)}\, U^{(i)\top} \tag{2.41}$$

where $U^{(i)}$ are the eigenvectors and $\Lambda^{(i)}$ are the eigenvalues of the adjacency matrix $A^{(i)}$. It derives a continuous correspondence matrix from the leading eigenvectors before mapping to a discrete permutation via the Hungarian algorithm.

After estimating pairwise assignments τ_{ij} for each pair of configurations x_i and x_j, with permutation matrices $P_{ij} = P(\tau_{ij})$, sGDML must reconcile them into a single, self-consistent symmetry group $\{P_q\}$. Each τ_{ij} minimizes the bipartite matching cost:

$$c_{ij} = \min_{\tau} \| P(\tau) A^{(j)} P(\tau)^\top - A^{(i)} \|_F^2, \tag{2.42}$$

where $\| \cdot \|_F$ is the Frobenius norm. Although each permutation is individually optimal, the set may violate transitivity ($P_{jk} \circ P_{ij} \neq P_{ik}$) across triples of configurations.

To enforce a single, globally-consistent assignment, a graph in which each of the M configurations is a vertex and each edge (i, j) is weighted by matching cost c_{ij} corresponding to the permutation matrix P_{ij} is constructed. The computation of Minimum Spanning Tree (MST) of the cost graph allows to pick out the most reliable pairwise permutation assignments while minimizing the sum of their matching costs [157, 158]. Traversing the MST and taking the transitive closure of its edges synchronizes all pairwise matchings into a consistent multipartite matching $\mathcal{P}$. The resulting $\mathcal{P}$ has rank N and encodes a set of permutation

automorphisms $\{P_q\}$ that capture both rigid and fluxional symmetries discovered in the training trajectories.

Rigid symmetries arise from the inherent point-group operations of a molecule that leave its intra-molecular graph unchanged (e.g. rotations, reflections, inversions). In the context of sGDML, these are captured when different MD snapshots sample equivalent orientations of the same PES. For instance, benzene's D_{6h} point-group operations (six-fold rotations and mirror planes) manifest as permutation matrices P_q within $\{P_q\}$ that map each carbon (and its bonded hydrogen) to an equivalent position.

Fluxional symmetries correspond to non-rigid, internal rearrangements—such as methyl-group rotations, torsional flips, or ring inversions—that also preserve the molecular adjacency graph. These automorphisms are "discovered" in sGDML without prior chemical bias: whenever a trajectory traverses a torsional barrier and returns to an isomorphic geometry, the bipartite matching identifies a permutation matrix P_q encoding that fluxional operation [17, 16]. Incorporating these dynamic symmetries into the kernel ensures that the learned PES respects both static and time-dependent invariances inherent to the molecule.

Discovered permutation matrices, then, are incorporated into the block-Hessian kernel by symmetrizing the kernel function through an average over the symmetry group:

$$k_{\mathrm{sym}}(x_i, x_j) \;=\; \frac{1}{|\,\mathcal{G}\,|} \sum_{g \in \mathcal{G}} k\big(x_i,\, P_g\, x_j\big), \tag{2.43}$$

ensuring the learned force field is exactly invariant under both global point-group operations and local fluxional automorphisms.

sGDML was originally formulated for isolated molecules and therefore lacks native support for Periodic Boundary Conditions (PBC). To overcome this limitation, Bravais-Inspired Gradient-Domain Machine Learning (BIGDML) generalizes the sGDML framework to periodic systems by introducing descriptor that enforce the PBC using the minimal-image convention and by embedding the full Bravais-lattice symmetry including both lattice translations and point-group operations directly into its kernel construction in from of permutational symmetries [18]. By carrying over the data efficiency and rigorous enforcement of physical priors from sGDML , BIGDML achieves QC accuracy for periodic systems with only hundreds of training configurations. It enables the reconstruction of accurate PES for adsorbates on surfaces and bulk materials, and an account for nuclear quantum effects by coupling the model with path-integral molecular dynamics.

A key challenge arises, however, from the fact that the spectral decomposition of adjacency matrices in periodic systems is non-unique, rendering eigenvector-based correspondence

ill-posed. As a result, the original sGDML algorithm for discovering permutational symmetries cannot be directly applied to periodic systems. To overcome this, BIGDML adopts an alternative strategy: it uses a single optimized structure to identify the full Bravais lattice symmetry and permutation symmetry group. These symmetries are then employed to symmetrize the kernel function.

One of the principal contributions of this thesis — detailed in Chapter 5 — is the development of a data-driven symmetry discovery algorithm applicable for both periodic and isolated systems. By constructing a frequency-filtered, angle-screened representative graph over the entire dataset and resolving subsystem independence via coarse-graining, the method uncovers the automorphism group without explicitly enumerating permutations and returns the group size, a minimal generator set, and atomic orbits, i.e., equivalence classes of atoms that remain indistinguishable across configurations. These outputs provide three concrete benefits: (i) the group order quantifies whether enforcing the full symmetry during training and evaluation is computationally feasible; (ii) the generator set reconstructs the entire group without storing all permutations, improving reproducibility and storage efficiency; and (iii) orbit labels refine atomic identity beyond elemental types and can be used as species tags in models using local descriptors, simplifying symmetry handling during training and improving improves parameter efficiency.

Despite its exceptional data efficiency — often achieving sub-kcal mol^{-1} Å^{-1} force errors with only a few hundred reference points — sGDML exhibits several practical limitations. First, even though the global nature of its kernel allows to capture long-range interactions, it incurs an $\mathcal{O}((MN)^3)$ scaling in computational time and $\mathcal{O}((MN)^2)$ memory usage with respect to the number of atoms N and number of configurations M, which effectively restricts its application to small-to-medium size molecules. Second, while latter works have introduced the gradient-based optimization to avoid the direct inversion of the kernel matrix, the performance for large datasets and systems remains suboptimal in comparison to the equivariant NN [159, 160, 161, 21, 22, 24].

QML (FCHL19*)

The FCHL family provides an atom-centered, local representation designed to encode many-body chemistry while rigorously enforcing translational, rotational, and permutational invariances [139]. In FCHL19, each atom i is described by a feature vector $\mathbf{q}_i$ that aggregates two-body and three-body distributions, stratified by element types.

The two-body part represents, for each central atom and each neighbor element type, a spectrum of radial basis functions placed on an equidistant grid between $r_{cut}/n^{(2)}_{R_s}$ and r_{cut},

where r_{cut} is a cutoff radius and $n_{R_s}^{(2)}$ is the number of radial bins. Each neighbor at distance r_{ij} contributes to every grid point via a log-normal basis with parameters

$$\mu(r_{ij}) = \ln\Big(\frac{r_{ij}}{\sqrt{1 + w/r_{ij}^2}}\Big), \quad \sigma^2(r_{ij}) = \ln\Big(\frac{r_{ij}}{\sqrt{1 + w/r_{ij}^2}}\Big), \tag{2.44}$$

where w is a width parameter controlling the broadness of the basis functions. To bias learning toward physically more relevant near-field interactions, the contribution is rescaled by a distance-decay factor

$$\xi_2(r_{ij}) = \frac{1}{r_{ij}^{N_2}}, \tag{2.45}$$

where N_2 is a hyperparameter, and smoothly truncated by a cosine cutoff function:

$$f_{\text{cut}}(r_{ij}) = \begin{cases} 0.5\big[\cos(\pi r_{ij}/r_{\text{cut}}) + 1\big], & r_{ij} \le r_{\text{cut}} \\ 0, & r_{ij} > r_{\text{cut}} \end{cases}. \tag{2.46}$$

Therefore, the two-body contribution to the feature vector of atom i is

$$G^{(2)} = \xi_2(r_{ij})\, f_{\text{cut}}(r_{ij})\, \frac{1}{R_s\sigma\sqrt{2\pi}}\, \exp\Big(-\frac{(\ln R_s - \mu(r_{ij}))^2}{2\sigma^2}\Big), \tag{2.47}$$

where R_s is the location of the radial gridpoints, summed over all neighbors j of a given element type.

In FCHL19*, the three-body term encodes, for each central atom i, all unordered neighbor pairs (j,k) by multiplying four factors: (i) a radial basis:

$$G_{\text{rad}}^{(3)} = \sqrt{\frac{\eta_3}{\pi}}\exp\big(-\eta_3\left(\frac{1}{2}\left(r_{ij} + r_{ik}\right) - R_s\right)^2\big), \tag{2.48}$$

where η_3 is a parameter that controls the width of the radial distribution functions; (ii) a Fourier-type angular basis $G_{\text{Ang}}^{(3)}$ built from Gaussian-damped $\cos(n\theta_{ijk})$ and $\sin(n\theta_{ijk})$ terms (in practice tunicated to $n = 1$ for efficiency), where θ_{ijk} is the angle formed at the central atom i by neighbors j and k; (iii) three-body scaling:

$$\xi_3 = c_3\frac{1 + 3\cos(\theta_{ijk})\cos(\theta_{ijk})\cos(\theta_{ijk})}{(r_{ij}r_{ik}r_{jk})^{N_3}}, \tag{2.49}$$

where c_3 is weighting factor for the balance between two- and three-body terms, and N_3 exponent controls the distance decay; (iv) a product of three cutoff functions to smoothly

limit the range of three-body interactions. The full three-body contribution to the feature vector of atom i is then

$$G^{(3)} = \xi_3 \, f_{\mathrm{cut}}(r_{ij}) f_{\mathrm{cut}}(r_{ik}) f_{\mathrm{cut}}(r_{jk}) \, G^{(3)}_{\mathrm{rad}} \, G^{(3)}_{\mathrm{Ang}}, \tag{2.50}$$

summed over all unordered neighbor pairs (j, k) of given element types.

Collecting the element-resolved two- and three-body contributions over all unordered neighbor pairs defines the per-atom descriptor $\mathbf{q}_i$. These atomic feature vectors are then compared across structures in the regression stage using an element-screened Gaussian kernel and summed over atoms to ensure size-extensivity and permutation invariance (Eq. 2.51).

The global kernel is additive over atomic environments, with elemental screening:

$$K(A, B) \;=\; \sum_{i \in A} \sum_{j \in B} \delta_{Z_i Z_j} \; \exp\Big(- \frac{\|\mathbf{q}_i - \mathbf{q}'_j\|^2}{2\sigma^2} \Big), \tag{2.51}$$

where σ is a length scale and $\delta_{Z_i Z_j}$ restricts comparisons to like elements. This construction yields an extensive kernel and embeds permutation invariance via the sum over environments, in contrast to explicit symmetrization. A practical limitation of exact KRR is the $\mathcal{O}(M^3)$ training and $\mathcal{O}(M)$ prediction cost in the number of training configurations M. FCHL19* alleviates this by approximating the Gaussian environment kernel in Equation (2.51) with Random Fourier Features (RFF) [162], producing a finite map $\boldsymbol{\phi}(\cdot) \in \mathbb{R}^D$ such that

$$K(A, B) \;\approx\; \boldsymbol{\phi}(A)^\top \boldsymbol{\phi}(B), \qquad \boldsymbol{\phi}(A) \;=\; \sum_{i \in A} \boldsymbol{\phi}(\mathbf{q}_i). \tag{2.52}$$

Training then reduces to linear regression in D dimensions (typically $D \ll M$), with memory $\mathcal{O}(MD)$ and compute $\mathcal{O}(MD^2)$ (direct solve) or $\mathcal{O}(MD)$ (iterative), and prediction cost $\mathcal{O}(D)$ per structure [140]. Crucially, locality ensures representation cost scales linearly with the number of atoms (fixed cutoff), making FCHL19* suitable for large systems and datasets. Moreover, efficient Graphics Processing Unit (GPU) implementation in QML-lightning combine FCHL19* with structured RFF to reach orders-of-magnitude speedups while retaining accuracy [140].

2.3.3 Neural Network Potentials

Kernel-based FFs have demonstrated remarkable accuracy for small- to medium-sized datasets by leveraging fixed, handcrafted descriptors and nonparametric regression to interpolate PES and atomic forces. However, their cubic scaling with training-set size, growing memory re-

quirements, and limited extrapolative power beyond the convex hull of sampled configurations constrain their applicability to large datasets and complex chemistries [156].

Neural Network Potential (NNP) overcome these limitations by employing parametric, data-driven architectures that combine the expressive power, flexibility and numerical robustness of kernel methods with end-to-end representation learning [11, 163, 19, 164]. Critically, they achieve training-time complexity that grows only linearly with the number of reference points, while ensuring a fixed, constant cost at inference.

Formally, an NNP attempts to learn $V(\mathbf{R}_1, ..., \mathbf{R}_N)$ — the multi-dimensional function that maps atomic coordinates to the total potential energy. To respect fundamental symmetries, modern NNP architectures are constructed to be invariant to overall translations, rotations, and permutations of identical atoms. A common approach is to decompose the total energy as a sum of environment-dependent atomic contributions:

$$E_{\text{tot}} = \sum_{i=1}^{N} E_i, \tag{2.53}$$

with each

$$E_i = \text{NN}_i(\mathbf{x}_i) \tag{2.54}$$

given by a neural network that takes a descriptor of atom i's local environment $\mathbf{x}_i$ as input [12].

Constructing an accurate NNP requires a sufficiently broad and high-quality training dataset and an appropriate loss function to fit the PES. Typically, reference energies $E^{(\alpha)}$ and their corresponding forces obtained as analytic derivatives:

$$\mathbf{F}_i^{(\alpha)} = -\nabla_{\mathbf{R}_i} E_{\text{tot}}^{(\alpha)}, \tag{2.55}$$

are computed via *ab initio* methods for a representative ensemble of structures $\{\alpha\}$ sampling the relevant configurational space. This ensures strict energy—force consistency and thus stable integration of the equations of motion in molecular-dynamics simulations. The network is then trained by minimizing a loss functional $L[\theta]$ (with respect to network parameters θ) that measures the error in reproducing these reference values. A common choice is a weighted Mean Squared Error (MSE) on energies and forces, for example:

$$L = \frac{1}{N_s} \sum_{\alpha=1}^{N_s} \left(w_E \left| E_{\text{NN}}^{(\alpha)} - E^{(\alpha)} \right|^2 + w_F \sum_i \left\| \mathbf{F}_{\text{NN},i}^{(\alpha)} - \mathbf{F}_i^{(\alpha)} \right\|^2 \right), \tag{2.56}$$

where w_E and w_F tune the relative importance of energy vs. force matching [165]. Because forces provide dense information ($3N$ data points per structure for N atoms), including them

in training greatly improves PES fidelity and MD stability, essentially teaching the NNP the slope of the energy surface in addition to its absolute level. Care must be taken to balance energy and force contributions in the loss to prevent the model from neglecting subtle energy differences in favor of large-magnitude forces or vice versa. In practice, training is performed with stochastic gradient-based optimizers on minibatches of structures, typically with early stopping or cross-validation to prevent overfitting [166].

NNP are trained by minimizing a generally non-convex loss function, and are typically optimized with stochastic gradient-based methods. Consequently, the training procedure is not strictly deterministic. Commonly used choices such as random parameter initialization, shuffling/sampling of minibatches, and stochastic regularizers can lead to different optimization trajectories across runs. Retraining the same architecture on the same dataset can therefore yield slightly different fitted parameters and, in some cases, small differences in generalization performance. Such effects are often most visible in data-sparse or extrapolative regions of the PES. In long MDs simulations, small force differences can also accumulate over time, so comparisons between models should be interpreted relative to this run-to-run variation and, where relevant, supported by repeated trainings, cross-validation, or ensembles. Finally, even when pseudo-random seeds are fixed, exact bitwise reproducibility can be limited by non-deterministic GPU kernels and floating-point reductions. Most ML frameworks provide options to request deterministic algorithms when needed, typically at some performance cost.

It is crucial that the training set be representative of all important regions of phase space; otherwise, the NNP may exhibit poor generalization or instabilities when encountering unseen configurations [167, 168]. Strategies such as iterative refinement or active learning are often employed, wherein MD simulations with a preliminary NNP are used to identify poorly sampled configurations that are then added (with new *ab initio* calculations) to the training set [169, 170]. This iterative workflow helps ensure that the final model has "seen" a broad swath of relevant configurations, reducing extrapolation risk [171, 172, 169].

The entire training/validation workflow is analogous to fitting a classical force field, but with a far more flexible functional form and a data-driven approach: the neural network automatically discovers the complex functional relationships needed to reproduce quantum-mechanical observables, rather than requiring a human to postulate analytical formulas for each interaction.

Once trained, NNPs can be deployed in standard molecular dynamics engines, where they compute energies and forces on-the-fly. They typically scale linearly with system size (owing to the locality of interactions within a cutoff) and can be accelerated on GPUs, making simulations of thousands of atoms over nanoseconds feasible [24, 25, 173, 164, 166]. However,

the design of NNPs still requires careful choices of cutoff radius R_c, network depth and width, and training-data coverage to balance locality, expressivity, and transferability.

In recent years, MPNN architectures have become the dominant framework for NNPs [174]. In an MPNN, the molecular structure is represented as a graph $G = (V, E)$, where each atom corresponds to a node $v \in V$ with initial feature vector $\mathbf{h}_v^0$, and edges $(v, w) \in E$ connect atoms within a cutoff, optionally carrying edge attributes $\mathbf{h}_{e,vw}$ [19, 175, 35, 176]. Over K message-passing steps, node features are updated by

$$\mathbf{m}_v^{t+1} = \sum_{w \in \mathcal{N}(v)} M_t\big(\mathbf{h}_v^t, \mathbf{h}_w^t, \mathbf{h}_{e,vw}\big), \qquad \text{(message)} \tag{2.57}$$

$$\mathbf{h}_v^{t+1} = U_t\big(\mathbf{h}_v^t, \mathbf{m}_v^{t+1}\big), \qquad \text{(update)} \tag{2.58}$$

where M_t and U_t are learnable neural networks that respectively compute messages and update embeddings [177, 178]. The schematic illustration of the message passing process is shown in Figure 2.4. After K passes, a permutation-invariant readout aggregates the final node embeddings into the total energy:

$$E = R\big(\{\mathbf{h}_v^K \mid v \in V\}\big), \tag{2.59}$$

with R another learnable function (e.g. a sum or more general pooling network) [179, 20]. This iterative, graph-based scheme naturally captures many-body interactions through multi-hop neighbour aggregation—far beyond simple pairwise terms—and eliminates the need for fixed, handcrafted descriptors [180, 181].

While standard MPNNs guarantee invariance to translations and permutations, they do not explicitly encode directional (tensorial) information—an aspect addressed by equivariant architectures through the incorporation of vectorial features that transform covariantly under rotations. In an equivariant MPNN, internal features are organized into irreducible representations (irreps) of the rotation group $SO(3)$, typically expressed in a spherical tensor basis. Each atomic node carries features of type l, where $l = 0$ denotes scalars, $l = 1$ vectors, and $l = 2$ rank-2 tensors, up to some cutoff $l_{\max}$. A feature of type l is a $(2l + 1)$-dimensional vector that transforms under a rotation $R \in SO(3)$ according to the Wigner D-matrix of order l:

$$\mathbf{H}^{(l)} \xrightarrow{R} D^{(l)}(R)\mathbf{H}^{(l)}, \tag{2.60}$$

ensuring that scalar features remain invariant, vector features rotate as 3D vectors, and higher-order tensors rotate with the appropriate representation, as shown in Figure 2.5.

To combine features of different types, equivariant networks employ tensor products

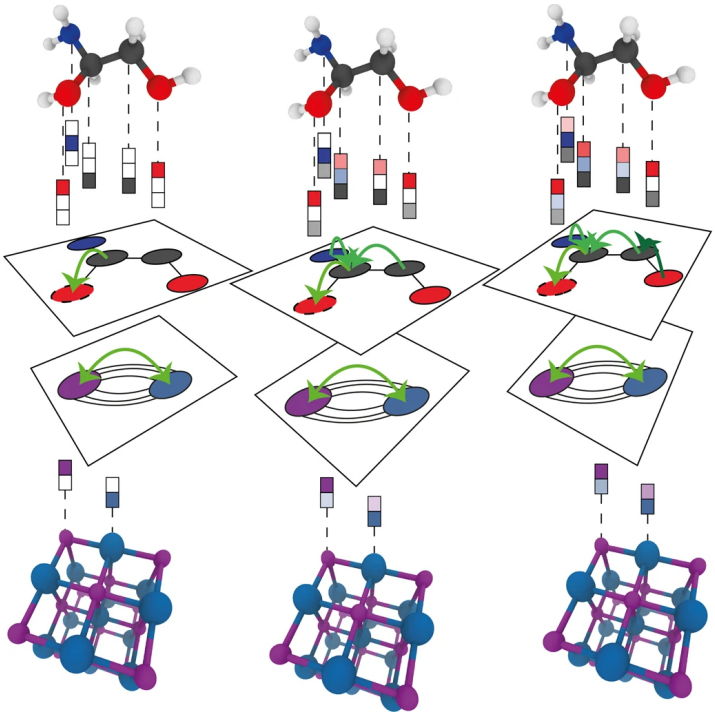


Figure 2.4 Schematic depiction of a message passing operation for molecules (top) and crystalline materials (bottom). Figure was taken from [182]

coupled by Clebsch-Gordan coefficients, which decompose products of irreps into valid output irreps:

$$(l_1 \otimes l_2) = \bigoplus_{l=|l_1-l_2|}^{l_1+l_2} l \,. \tag{2.61}$$

In component form, the projection onto an output type l_3 reads

$$\left(A^{(l_1)} \otimes B^{(l_2)}\right)^{(l_3)}_{m_3} = \sum_{m_1,m_2} C^{(l_1,l_2,l_3)}_{m_1 m_2 m_3} \cdot A^{(l_1)}_{m_1} \cdot B^{(l_2)}_{m_2}, \tag{2.62}$$

where $C^{(l_1,l_2,l_3)}_{m_1 m_2 m_3}$ are Clebsch-Gordan coefficients ensuring the result transforms as a proper type-l_3 tensor.

In message passing, node i aggregates information from its neighbors j by combining the neighbor features $\mathbf{H}^{(l_1)}_j$ with the interatomic direction $\hat{\mathbf{x}}_{ij} = (\mathbf{x}_j - \mathbf{x}_i)/\|\mathbf{x}_j - \mathbf{x}_i\|$ that is expanded in spherical harmonics $Y^{(l_2)}(\hat{\mathbf{x}}_{ij})$. The dependence on the interatomic distance $r_{ij} = \|\mathbf{x}_j - \mathbf{x}_i\|$ enters through a learnable scalar weight function $\phi_{l_1,l_2\to l_3}(r_{ij})$, often implemented as a radial Multilayer Perceptron (MLP).

In practice, to obtain ϕ, most architectures first embed r_{ij} using a fixed radial basis expansion and then apply a standard MLP nonlinearity (e.g. SiLU) [184]. Two widespread choices for the radial basis are: (i) Gaussian(-like) functions, which provide a smooth and localized distance resolution but may become redundant or poorly conditioned if too many

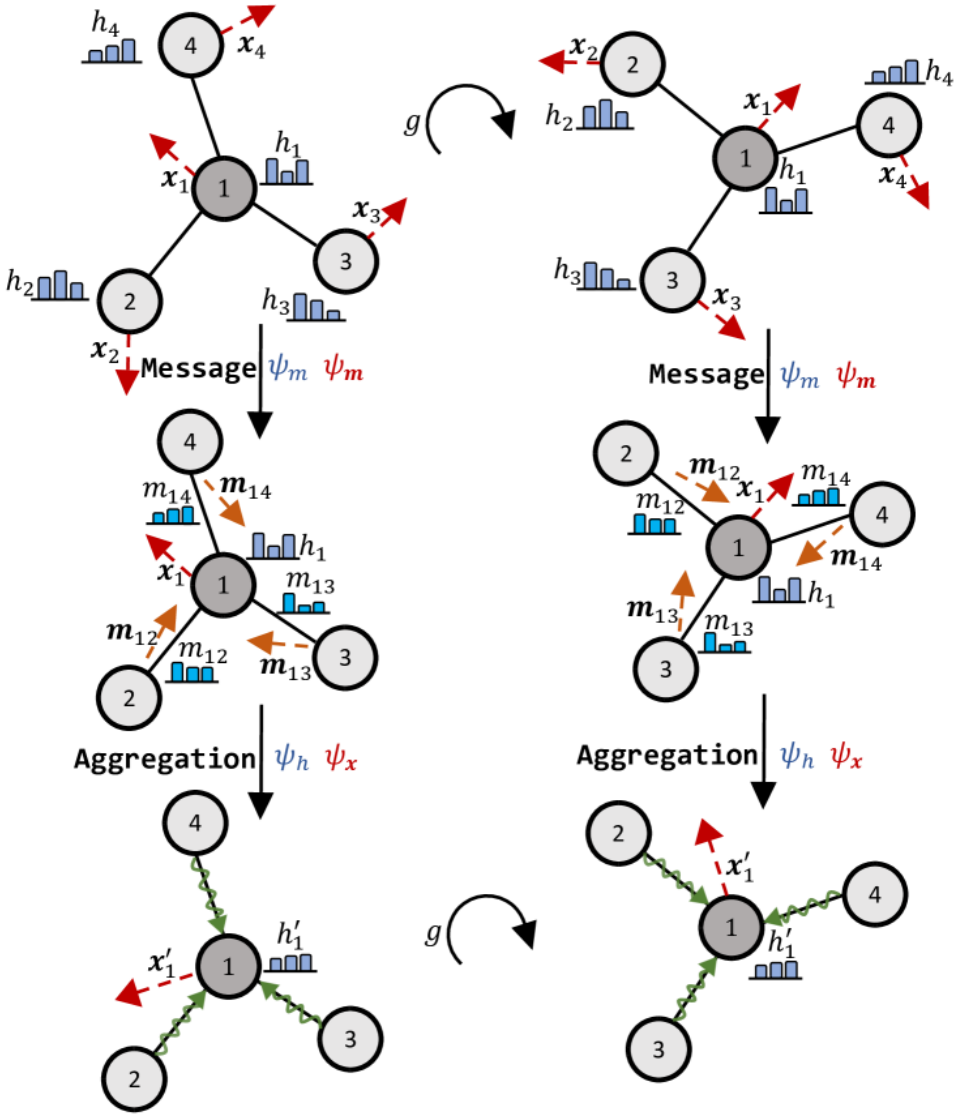


Figure 2.5 Schematic depiction of equivariant message passing in the case of rotation. Both scalar and vector (in bold) features are produced and the aggregated by combining neighbor features with geometric information (directions). Figure adapted from [183].

functions are used or if centers/widths are poorly chosen; and (ii) orthogonal bases such as Bessel functions on $[0, r_{\text{cut}}]$. In both cases, the basis is typically multiplied by a smooth cutoff so that interactions vanish smoothly at r_{cut}.

The resulting equivariant message ($SO(3)$ convolution) for output type l_3 is

$$[M_{ij}]_{m_3}^{(l_3)} = \sum_{l_1,l_2} \sum_{m_1,m_2} C_{m_1 m_2 m_3}^{(l_1,l_2,l_3)} \cdot [H_j^{(l_1)}]_{m_1} \cdot Y_{m_2}^{(l_2)}(\hat{\mathbf{x}}_{ij}) \cdot \phi_{l_1,l_2 \to l_3}(r_{ij}), \tag{2.63}$$

and the node update is obtained by summing over all neighbors,

$$H_{i;m_3}^{(t)(l_3)} = \sum_{j \in \mathcal{N}(i)} [M_{ij}]_{m_3}^{(l_3)}. \tag{2.64}$$

This design guarantees that rotating the input geometry by R produces a corresponding rotation of all equivariant hidden features and covariant outputs (e.g., force vectors), while scalar predictions such as energies remain invariant. Models like NequIP [21] and PaiNN [20] first demonstrated the advantage of this equivariant message passing for MLFFs, significantly reducing errors by embedding $E(3)$ symmetry. More recently, MACE [24, 25] and SO3krates [142, 23] have reached state-of-the-art accuracy across diverse systems by systematically enforcing the $SO(3)$ symmetry rules in their message-passing layers through the operations outlined in Eqs. (2.60)-(2.62).

MACE

Most earlier equivariant MPNNs passed only two-body messages (interactions depending on pairs of atoms), forcing them to stack many message-passing layers to capture higher-order effects. The MACE architecture combining equivariant message passing with Atomic Cluster Expansion (ACE) framework. In ACE, many-body interactions are expanded in a basis of symmetrized spherical harmonics and radial functions, allowing a k-body term to be built from lower-order basis products while keeping the cost linear in the number of neighbors.

MACE leverages this insight by using a "*density trick*" to efficiently include higher-body effects: instead of enumerating all neighbor k-tuples, it first aggregates neighbor contributions into per-atom features (Equations 2.63-2.64) and then forms tensor powers of this aggregate. Higher-body features are constructed as:

$$B^{(t)(L)}_{i;\eta_\nu,M} = \sum_{l_\xi} \sum_{m_\xi} C^{(l_1,\ldots,l_\nu)\to L}_{m_1,\ldots,m_\nu M;\eta_\mu} \prod_{\xi=1}^{\nu} H^{(t)(l_\xi)}_{i;m_\xi}, \tag{2.65}$$

where $C^{(l_1,\ldots,l_\nu)\to L}_{m_1,\ldots,m_\nu M;\eta_\mu}$ are generalized Clebsch-Gordan coefficients that couple irreps $(l_1, \ldots, l_\nu)$ into a resultant irrep L, with η_ν indexing different coupling schemes. The ν-fold product of neighbor-summed features expands into a sum over all ν-tuples of neighbors, so it captures all interactions up to $k = \nu + 1$ bodies. In this way, all k-neighbor interaction terms are implicitly included in the expansion of Equation 2.65, without needing to iterate over every neighbor tuple explicitly. By exchanging the order of summation and multiplication, MACE avoids the steep combinatorial cost that a naive k-tuple loop would incur - the cost remains linear in the number of neighbors even as k increases.

Rotational symmetry is enforced by the Clebsch-Gordan couplings, so internal features are organized by $SO(3)$ irreps (scalars, vectors, etc.) and transform correctly under rotations. In practice MACE truncates angular channels at a manageable level (typically $l_{\max} = 1$ or

2), yet still capturing high body-order geometry.Targeting $k = 4$ (i.e., a three-fold tensor power, $\nu = 3$) was found to be sufficient for MACE to reach state-of-the-art accuracy across a wide range of benchmarks, requiring only one message-passing iteration in the network, in contrast to conventional equivariant models that often need 4-6 layers of pairwise message passing to achieve comparable expressivity [24, 25].

The MACE architecture introduces only a few key hyperparameters, which control the trade-off between model complexity and accuracy:

- The maximum order of irreps l_{max} that determine the complexity of the internal feature space;
- The number of channels, i.e., features belonging to each hidden irreps, indicates the dimensionality of the internal representation at a given symmetry level.
- The design of the MLP that defines the radial resolution of the model.

These hyperparameters ultimately govern the model size and accuracy trade-off. For example, a default MACE configuration might include $l = 0$ and $l = 1$ features (scalars and vectors) with on the order of 128 channels each, and a small 2-3 layer MLP that acts on a distance expansion into 8 Bessel basis functions. Increasing the number of channels or extending l_{max} can improve accuracy at the cost of more parameters, whereas adjusting the radial MLP (e.g. adding more neurons or layers) improves how finely the model can fit distance-dependent interactions.

SO3krates

SO3krates is an $SO(3)$-equivariant graph neural network that replaces conventional spherical convolutions with an equivariant self-attention mechanism, thereby avoiding explicit tensor-product operations while preserving rotational symmetry. In this architecture, atomic features and geometric information are handled in separate channels: each atom i carries an invariant feature vector f_i (transforming as an $SO(3)$ scalar) alongside equivariant spherical harmonic coordinates $x_i^{(l)}$ (irreducible feature vectors of order l, transforming in the $2l$+1 dimensional irreps of $SO(3)$). As it was for general equivariant architecture, the directional basis of real spherical harmonics $Y_l^m(\hat{r}_{ij})$ is used to expand relative positions $\hat{r}_{ij}$, providing a fixed equivariant basis for geometric content. SO3krates's message-passing step is then formulated as a dual-branch self-attention: one branch updates the invariant features and the other updates the equivariant (spherical) features, both using attention weights that depend only on rotation-invariant quantities (like interatomic distances and scalar features). Concretely, for

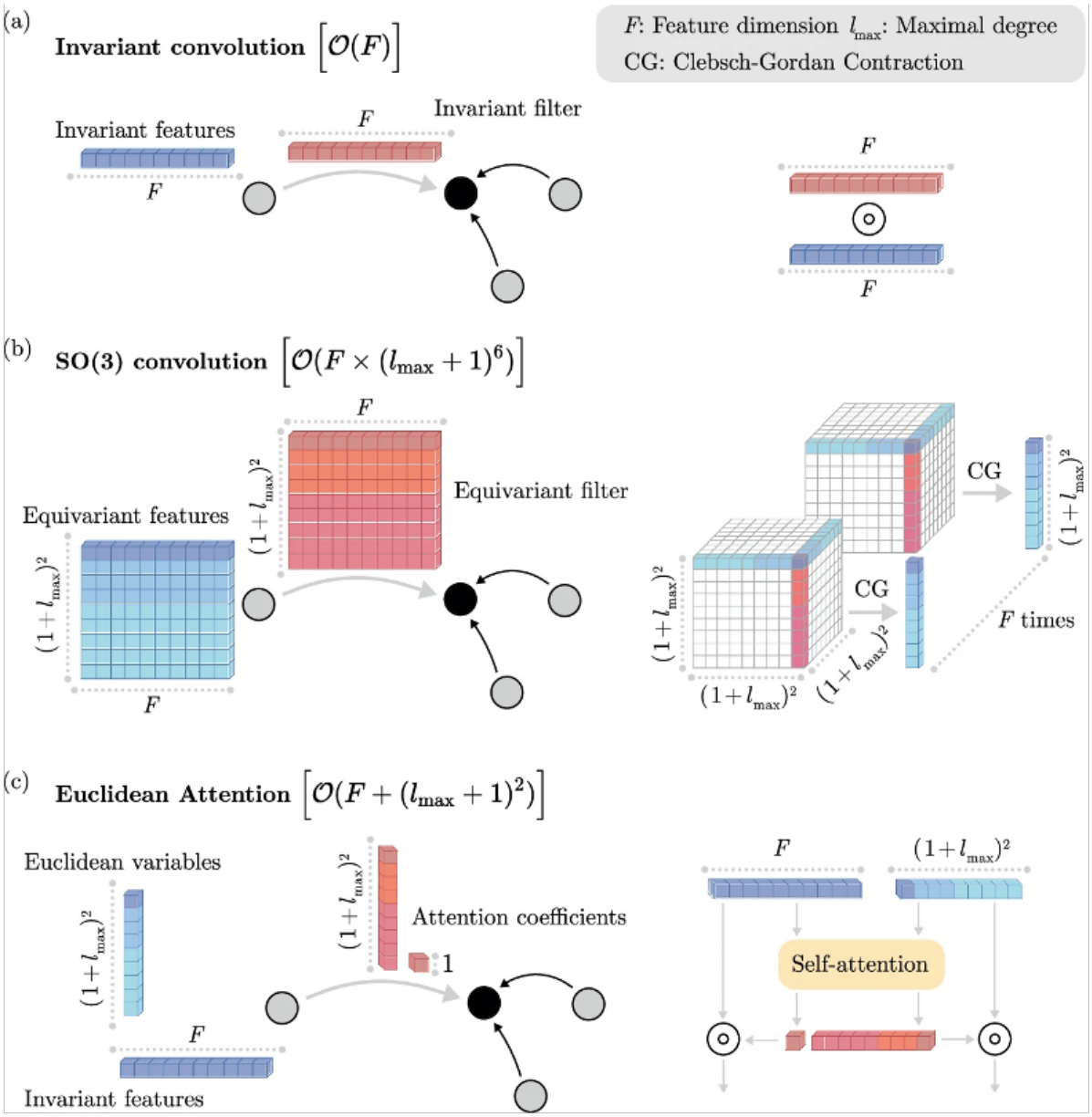


Figure 2.6 Schematic ilustraton of the difference between (a) invariant convolution, (b) $SO(3)$ convolution, and (c) Euclidean attention mechanism used in SO3krates architecture. The figure was taken from [23].

each neighbor j of i, an attention coefficient α_{ij} is computed via a learned filter w_{ij} acting on the distance r_{ij} and orientation (spherical coordinate) differences, and this scalar weights the neighbor's contribution. The invariant feature update at layer t can be written as a weighted sum:

$$f_i^{(t+1)} = f_i^{(t)} + \sum_{j \in N(i)} \alpha_{ij} \cdot \phi(r_{ij}) \cdot f_j^{(t)}, \tag{2.66}$$

while the equivariant update adds a projected spherical harmonic from each neighbor:

$$x_i^{(l,t+1)} = x_i^{(l,t)} + \sum_{j \in N(i)} \alpha_{ij}^{(l)} \cdot \phi(r_{ij}) \cdot Y^{(l)}(\hat{r}_{ij}). \tag{2.67}$$

Here $\phi(r_{ij})$ is a radial cutoff/filter and $\alpha_{ij}^{(l)}$ denotes a degree-specific attention weight (ensuring, for example, that p-orbital (l=1) contributions can be weighted differently than d-orbital (l=2) ones). Because α_{ij} and $\alpha_{ij}^{(l)}$ are functions of invariants (distances and feature scalars), they are themselves invariant under rotations. Multiplying these weights with the fixed basis $Y^{(l)}(\hat{r}_{ij})$ (which rotates covariantly) yields updates to $x_i^{(l)}$ that transform properly with respect to the lth irrep - thus the overall message-passing is equivariant by construction.

SO3krates departs from standard tensor-product-based message passing by decoupling the geometric degrees of freedom: rather than repeatedly coupling neighbor irreps via Clebsch-Gordan coefficients at every interaction (the costly "triplet" tensor products of conventional $SO(3)$ convolutions, shown in Equation 2.63), the model uses the attention filters on invariants to achieve a similar effect with far less computation. Only after each self-attention update does the model perform a local interaction (mixing) step that fuses geometric and feature information: the updated f_i and $x_i^{(l)}$ are combined through learnable functions that incorporate the magnitudes $|x_i^{(l)}|$ and even cross-term contractions of different spherical channels (using Clebsch-Gordan-based mixing) to produce new irreps. This design allows efficient mixing of geometric and feature information - higher-order angular features influence the scalar embedding and vice versa - without breaking equivariance. In summary, SO3krates's architecture leverages explicit irreps and a spherical harmonic basis to handle orientations, but achieves rotational equivariance through a content-driven attention mechanism in lieu of computationally heavy $SO(3)$ convolutions, resulting in a more efficient yet expressive equivariant message-passing paradigm.

In summary, neural network potentials represent a major advancement in the modeling of interatomic interactions. They marry the accuracy of first-principles calculations with the efficiency of classical force fields, and modern architectures (especially equivariant message-passing networks like MACE and SO3krates) have demonstrated the ability to simulate complex molecules, interfaces, and materials with remarkable fidelity. Their introduction has broadened the scope of molecular simulations, enabling routine nanosecond-length MD for systems with hundreds of atoms at near-DFT quality. At the same time, ensuring that NNPs are employed safely and effectively requires careful attention to training data completeness, adherence to physical symmetries, and thorough validation of predicted observables.

2.3.4 Evaluation Metrics, Benchmarking, and Databases

Rigorous evaluation of MLFFs is essential to establish their reliability and scope of applicability. MLFFs have rapidly progressed to achieve near QC accuracy on a wide range of chemical systems while maintaining computational efficiency. This progress enables simulations of large

molecular and materials systems with close-to *ab initio* accuracy. However, to trust MLFFs in critical applications, one must subject them to stringent benchmarks and validation protocols, analogous to the verification efforts in electronic structure theory. In the DFT community, the Δ-Project showed that, for fundamental solid-state properties, different DFT codes and computational setups agree to within experimental uncertainty [185]. It also introduced the Δ metric as a rigorous benchmarking framework that users and developers can apply to validate and refine new implementations. This systematic benchmarking of DFT implementations greatly improved confidence in DFT results. By establishing an accepted "error bar" for DFT calculations, such efforts set a precedent for reproducibility and precision benchmarking in computational chemistry. Similarly, MLFFs require community-wide evaluation to quantify their accuracy relative to trusted references and to uncover failure modes.

Benchmarking MLFFs against reference data is essential because modern MLFF errors now match the intrinsic variability of those references. Equivariant NNPs can reach force errors of just a few meV/atom—smaller than some differences between DFT functionals or among DFT codes. As a result, an MLFF may reproduce its chosen reference within uncertainty yet still inherit its biases. Therefore, benchmarking must place MLFF errors in the wider context of reference accuracy, using higher-level methods or experiments when available, or cross-comparing multiple *ab initio* references to set chemically meaningful error limits.

Static error metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Maximum Error (MAXE) are the foundation of most MLFF evaluations. For a set of N reference configurations with properties $y_i^{(ref)}$ (energies or forces) and MLFF predictions $y_i^{(pred)}$, the definitions are straightforward:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i^{(pred)} - y_i^{(ref)}\right)^2}, \tag{2.68}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i^{(pred)} - y_i^{(ref)}\right|, \tag{2.69}$$

$$\text{MAXE} = \max_{i=1,\ldots,N}\left|y_i^{(pred)} - y_i^{(ref)}\right| . \tag{2.70}$$

In practice, separate error values are often reported for energies (e.g. in kcal/mol or meV per atom) and forces (e.g. in kcal/(mol·Å) or eV/Å). MAE measures average absolute deviation and it is robust against outliers and providing directly interpretable error magnitudes. Because it squares residuals, RMSE amplifies large errors and conforms to Gaussian-noise assumptions, making it more sensitive to outliers and useful for flagging potential catastrophic failures in MLFF extrapolations.

However, static metrics have important limitations. A single aggregated error value cannot capture the full distribution of model deviations. An MLFF may achieve a low overall RMSE while still making large errors on a small subset of configurations – errors that might be "averaged out" in the RMSE but could be catastrophic if those configurations occur during an MD simulation. The maximum error, while highlighting worst-case performance, does not convey how rare or relevant that worst case is.

Beyond static test-set metrics, it is essential to monitor how the loss evolves during optimisation. For a given model f_τ after τ optimisation epochs and a loss function ℓ combining errors in energies and forces, the training and validation losses are defined as

$$L_{\text{train}}(\tau) = \frac{1}{|D_{\text{train}}|} \sum_{(\mathbf{x},\mathbf{y})\in D_{\text{train}}} \ell\big(f_\tau(\mathbf{x}), \mathbf{y}\big), \tag{2.71}$$

$$L_{\text{val}}(\tau) = \frac{1}{|D_{\text{val}}|} \sum_{(\mathbf{x},\mathbf{y})\in D_{\text{val}}} \ell\big(f_\tau(\mathbf{x}), \mathbf{y}\big). \tag{2.72}$$

Here D_{train} and D_{val} denote the training and validation datasets, respectively. The difference

$$\Delta L(\tau) = L_{\text{val}}(\tau) - L_{\text{train}}(\tau) \tag{2.73}$$

provides a simple measure of the *generalization gap*. A small positive ΔL reflects the expected degradation in performance on unseen configurations. A rapidly growing positive gap signals overfitting, whereas a nearly vanishing or mildly negative gap typically indicates either a capacity-limited model (underfitting) or small statistical differences between the training and validation subsets.

In practice, the models employed in this thesis are trained with early stopping based on $L_{\text{val}}(\tau)$: the final parameters are chosen at the epoch $\tau^\star$ that minimises the validation loss, or after a fixed patience interval without improvement. This procedure helps to avoid overfitting by stopping training once the validation loss ceases to decrease, even if the training loss continues to decline.

For K-fold cross-validation experiments, training and validation losses are recorded separately for each fold $f = 1, \ldots, K$, yielding curves $L^{(f)}_{\text{train}}(\tau)$ and $L^{(f)}_{\text{val}}(\tau)$. To summarise the behaviour across folds in a compact way, one can aggregate these into mean learning curves

with an uncertainty band:

$$\bar{L}_{\text{train}}(\tau) = \frac{1}{K}\sum_{f=1}^{K} L_{\text{train}}^{(f)}(\tau), \quad \sigma_{\text{train}}(\tau) = \sqrt{\frac{1}{K-1}\sum_{f=1}^{K}\left(L_{\text{train}}^{(f)}(\tau) - \bar{L}_{\text{train}}(\tau)\right)^2}, \tag{2.74}$$

$$\bar{L}_{\text{val}}(\tau) = \frac{1}{K}\sum_{f=1}^{K} L_{\text{val}}^{(f)}(\tau), \quad \sigma_{\text{val}}(\tau) = \sqrt{\frac{1}{K-1}\sum_{f=1}^{K}\left(L_{\text{val}}^{(f)}(\tau) - \bar{L}_{\text{val}}(\tau)\right)^2}. \tag{2.75}$$

Plots of $\bar{L}_{\text{train}}(\tau)$ and $\bar{L}_{\text{val}}(\tau)$ with shaded bands corresponding to $\pm\sigma(\tau)$ provide an aggregate view of convergence and fold-to-fold variability.

While static metrics assess overall system performance on test configurations, many applications require pinpointing exactly where a model excels or struggles. An MLFF's error might concentrate in specific local environments — e.g. it might systematically overestimate forces for certain under-coordinated atoms, or struggle with particular dihedral angle orientations. Recent analysis tools (such as the FFAST software) enable per-atom and per-configuration error inspection [186]. One can plot the distribution of force errors on each atom type, identifying if, say, all lead atoms have higher force MAE than iodines, which might point to a poorly learned part of chemistry, as shown in Figure 2.7.

An additional necessity for thorough validation arises from the fact that MLFFs can deliver smooth, physically plausible—but ultimately incorrect—predictions when asked to extrapolate beyond their training data. Unlike traditional force fields with rigid functional forms that typically crash on extrapolation, MLFFs may quietly generate wrong forces and energies, seeding subtle yet systematic errors into simulations. Spotting these failures in complex or large-scale systems is especially difficult when direct high-level reference calculations are out of reach. Therefore, robust evaluation protocols must move beyond simple training-and-test error metrics to detect and guard against hidden extrapolation errors.

Beyond stability, one can evaluate derived observables from MD. These include thermodynamic and kinetic properties that result from long-time sampling, which a good MLFF should reproduce if it truly captures the PES. For example, radial distribution functions in liquids, diffusion coefficients, phase transition temperatures, vibrational spectra, or conformational free energy differences can be calculated from MLFF MD and compared to either experimental data or reference simulations.

Ultimately, no single metric is sufficient: reliable MLFF evaluation combines traditional error metrics with stress tests (long MD simulations, high-temperature stability), consistency checks (comparing different MLFFs and reference data), and observable-based comparisons (does the MLFF reproduce key physical observables like structures, transition probabilities,

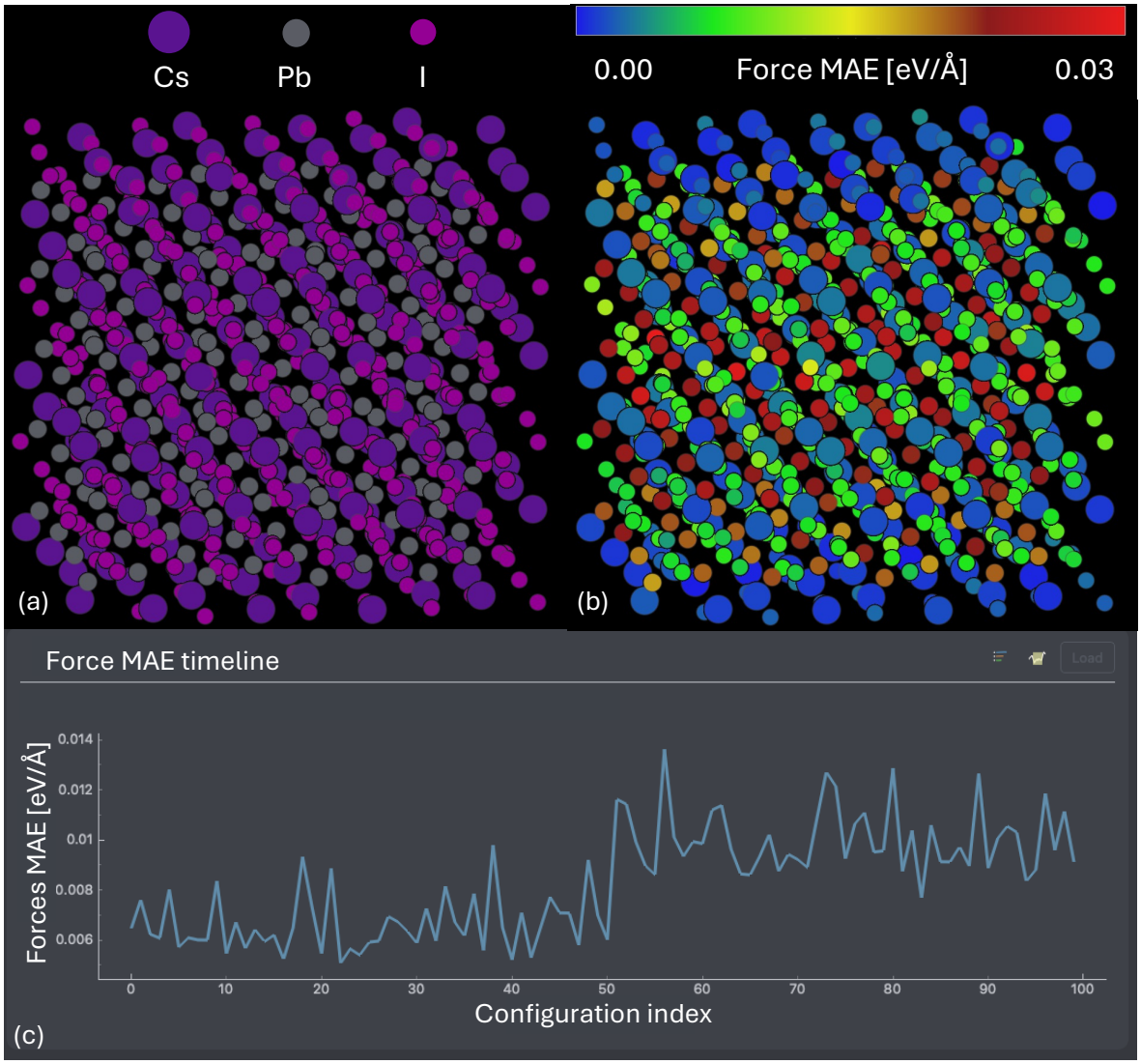


Figure 2.7 Per-atom and per-configuration force-error analysis for the $CsPbI_3$ $4 \times 4 \times 5$ slab. (a) Slab geometry. (b) Atoms colored by per-atom force MAE; a single color scale is used for all species. (c) Per-configuration force MAE for each of the 100 configurations.

and spectra?). Developing standard protocols to quantify these aspects is an active area of research, and is necessary for establishing community benchmarks beyond just "low test RMSE"

2.4 Halide Perovskite Materials

Conventional semiconductors face distinct limits. Crystalline silicon has an indirect band gap, which makes light emission inefficient, and weak near-edge absorption requires comparatively thick photovoltaic absorbers [187]; III-V compound semiconductors achieve record efficiencies but rely on costly lattice-matched epitaxy and expensive single-crystal substrates. Meanwhile, thin-film chalcogenides face tellurium-availability concerns, and organic semiconductors

exhibit large exciton binding energies and low mobilities [188, 189, 190, 191]. These limitations motivate the search for semiconductors that combine strong absorption in thin films, benign defect physics, tunable band gaps, and low-temperature, large-area fabrication — without sacrificing stability or supply-chain robustness [192, 193]. Adopting the general formula ABX_3, where A is an organic (e.g., $CH_3NH_3^+$, $HC(NH_2)_2^+$) or inorganic (Cs^+) cation, B is a divalent metal (Pb^{2+}, Sn^{2+}), and X is a halide (I^-, Br^-, Cl^-), halide perovskites have revolutionized optoelectronics by combining low-cost processability via solution techniques, compositional tunability, and exceptional optoelectronic performance. Through strategic alloying at the A, B, and X sites, their direct band gaps can be continuously tuned across the visible spectrum, enabling high-efficiency photovoltaics and light emission [188]. Despite rapid efficiency gains, recent outdoor studies and reliability reviews show that field stability and bankability remain limiting, sharpening the need to resolve light-/bias-driven ion motion, phase segregation, and interfacial chemistry at the atomic scale [194, 195, 196]. With absorption coefficients exceeding 10^5 cm^{-1}, ultrathin perovskite layers harvest light more effectively than most conventional semiconductors [197, 198]. At the same time, micron-scale carrier diffusion lengths support efficient charge extraction in solar cells and photodetectors [7, 199, 200]. This context motivates the application of atomistic simulation to investigate the mechanisms that govern halide perovskite performance and stability.

2.4.1 Composition and Crystal Structure

The perovskite structure consists of a three-dimensional framework of corner-sharing BX_6 octahedra, shown in Figure 2.8. This geometry imposes steric constraints on the ionic radii of A, B, and X, commonly summarized by the Goldschmidt tolerance factor,

$$t = \frac{r_A + r_X}{\sqrt{2}\,(r_B + r_X)}, \tag{2.76}$$

where r_A, r_B, and r_X are the ionic radii of the A, B, and X ions, respectively. Values of t near 1 indicate a stable cubic structure, while deviations suggest octahedral tilting or even the disruption of the corner-sharing network. For example, formamidinium lead iodide ($FAPbI_3$) with $t \approx 0.98$ is at the edge of the stable range for the perovskite structure because the FA cation is almost too large for the A-site [201, 202], methylammonium lead iodide ($MAPbI_3$) has $t \approx 0.91$, favoring a slightly distorted perovskite structure [203], and cesium lead iodide ($CsPbI_3$) with $t \approx 0.81$ tends to form non-perovskite phases unless stabilized by temperature or strain [27, 204]. In $CsPbI_3$, on cooling from the high-temperature perovskite, the structure undergoes $Pm\bar{3}m \rightarrow P4/mbm \rightarrow Pbnm$, and near room temperature it spontaneously converts

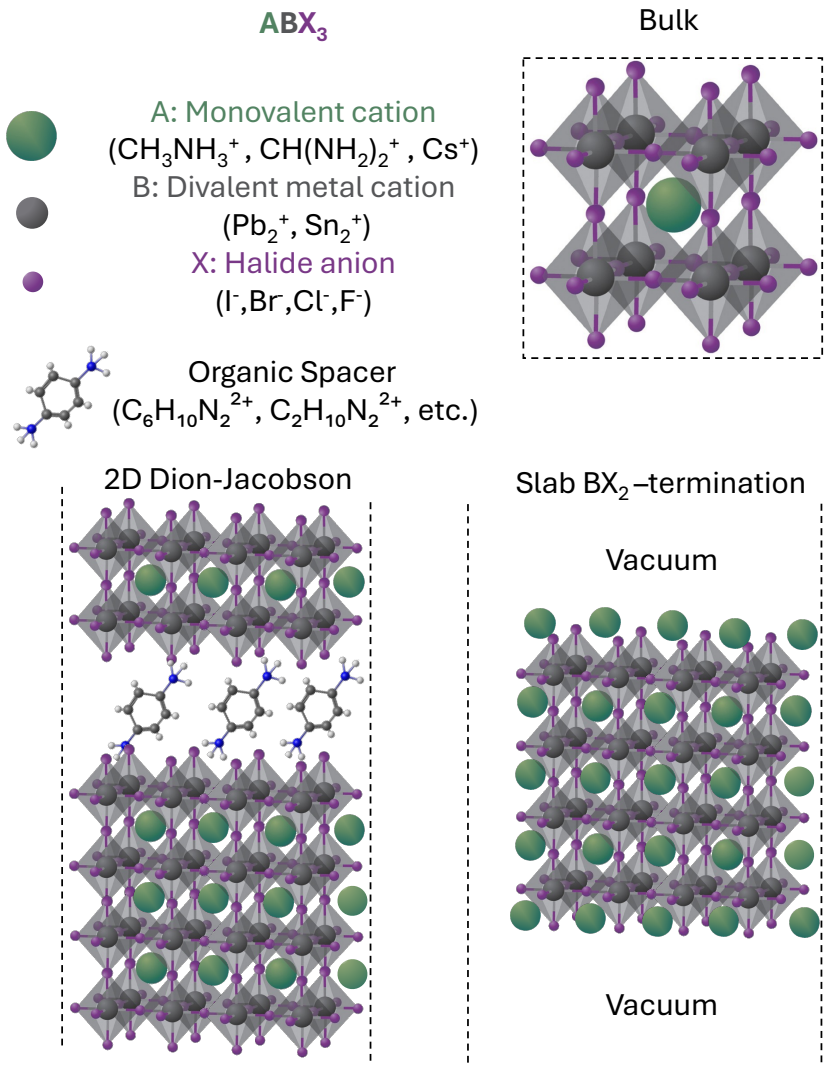


Figure 2.8 Schematics of (a) ideal cubic perovskite ABX_3 ($Pm\bar{3}m$); (b) a Dion–Jacobson layered perovskite with $n = 2$, where two-octahedra-thick $[BX_6]$ slabs are bridged by the divalent spacer benzene-1,4-bis(aminium); and (c) the slab model. A-site cations in green, B-site cations in gray, X-site anions in purple; PBCs indicated by dashed black lines.

to the non-perovskite δ phase (Pnma) if no stabilization techniques are used. The reported $\alpha \rightarrow \beta$ and $\beta \rightarrow \gamma$ transitions occur near 510 K and 325 K, respectively, while α-$CsPbI_3$ is stable only above ≈590-630 K at ambient pressure, as shown in Figure 2.9 [204, 27, 205].

Reduced-Dimensionality Perovskites

Beyond the three-dimensional perovskite framework, the Goldschmidt tolerance factor is not applicable as a design rule, as using bulky organic molecules (e.g., short alkyl or aryl-based compounds) disrupts the corner-sharing connectivity. As a result, two-dimensional (2D) halide perovskites can be isolated in the form of Ruddlesden-Popper and Dion-Jacobson

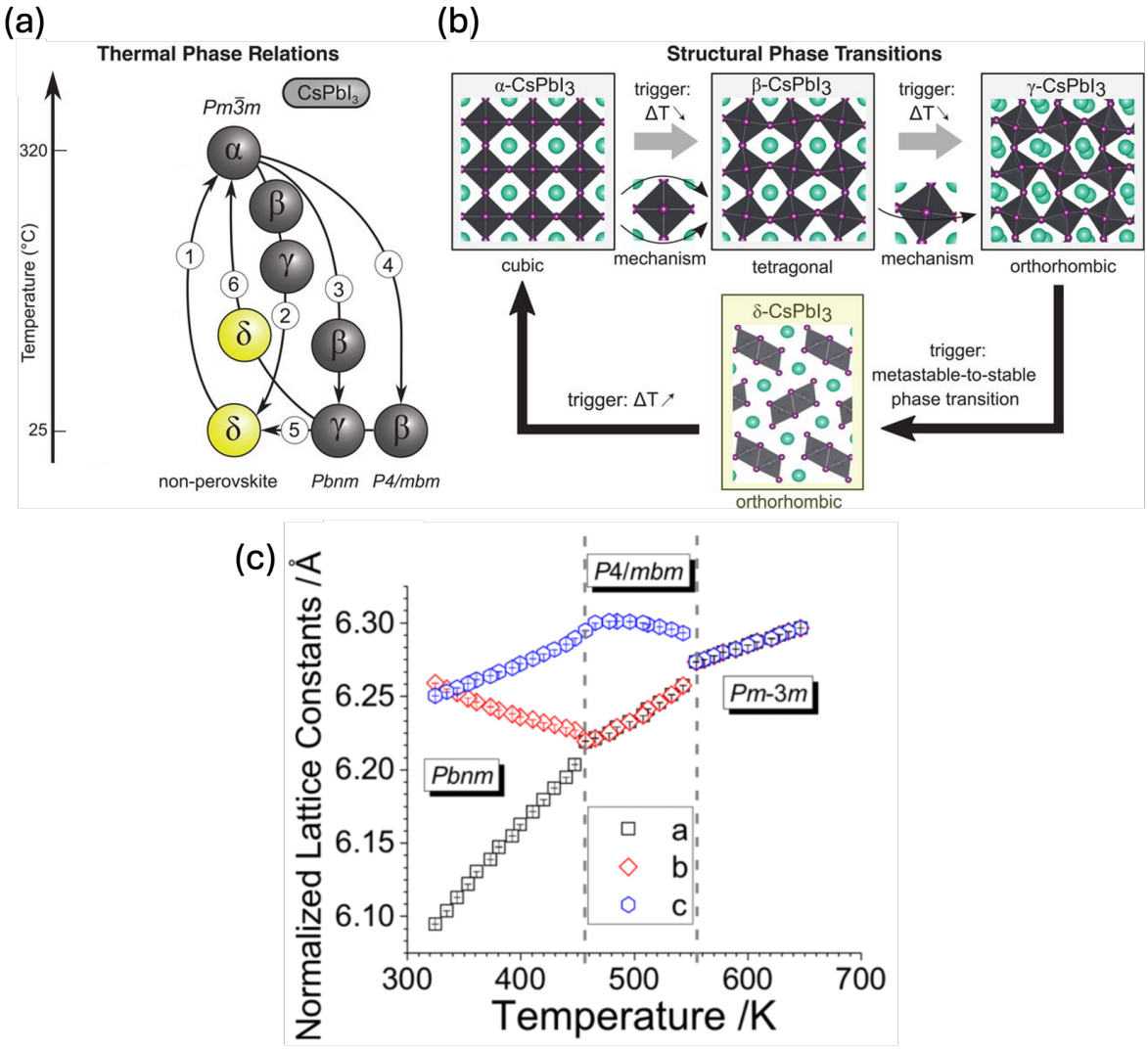


Figure 2.9 (a) Temperature-phase diagram for $CsPbI_3$ with representative thermal pathways between the cubic, tetragonal, and orthorhombic polymorphs. (b) Crystal structures of these phases and the associated octahedral tilt modes that mediate the transitions (PbI_6 octahedra in black, I in purple, Cs in cyan). Reproduced from Ref. [205]. (c) Temperature evolution of the lattice parameters of $CsPbI_3$. Reproduced from Ref. [27].

phases. In the Dion-Jacobson/Ruddlessden-Popper series, the general formula is respectively:

$$(A')_m(A)_{n-1}B_nX_{3n+1}, \tag{2.77}$$

where A' is the bulky organic molecule, and $m = 1$ or 2 for the Dion-Jacobson or Ruddlessden-Popper, respectively, describes a structure in which organic spacer cations separate inorganic BX_6 octahedral layers of thickness n. Figure 2.8 illustrates the Dion-Jacobson 2D structure. By varying n, one systematically tunes the materials' quantum-well excitonic properties and enhances moisture stability. These 2D structures retain the corner-sharing BX_6 connectivity within each layer while introducing strong dielectric and structural anisotropy, making them ideal platforms for probing the influence of reduced dimensionality on charge-carrier

dynamics, exciton binding energies, and overall optoelectronic performance at the nanoscale.

Nonlayered halide-perovskite slabs are finite-thickness cuts of the 3D ABX_3 lattice that expose AX- or BX_2-terminated surfaces and lack organic spacer layers, providing an intermediate platform between bulk 3D perovskites and strictly layered Ruddlesden-Popper or Dion-Jacobson phases. Figure 2.8 illustrates a slab structure with BX_2-terminated surfaces. Experimentally, these correspond to thin films, epitaxial layers, or colloidal nanosheets that isolate surface and interface effects [206, 207]. Their increased surface-to-volume ratio magnifies surface states and defect chemistry, which can shift band alignment and alter carrier recombination and transport compared with bulk [208, 209]. As slab thickness approaches the exciton Bohr radius, quantum confinement strengthens exciton binding and weakens dielectric screening [210, 211, 212]. The free-carrier population and mobility fall [213, 214]. Transport shifts toward exciton diffusion and becomes strongly anisotropic, with cross-plane conduction suppressed [215, 216]. Increased surface and edge scattering reduces the diffusion coefficient, and faster surface or defect recombination shortens carrier lifetimes, together shortening diffusion lengths [215, 217]. Slabs, therefore, provide a controlled platform to probe how reduced dimensionality and surface chemistry govern defect formation, ion migration, and stability under operating conditions, while avoiding the dielectric and transport effects introduced by bulky organic spacers in Ruddlesden-Popper and Dion-Jacobson perovskites.

Structural Descriptor of Dynamic Disorder

An effective way to capture phase transitions in halide perovskites systems from atomistic simulation is to quantify octahedral tilts by assigning each BX_6 octahedron three Euler angles θ, ϕ, and ψ defined with respect to the pseudocubic axes as rotation about the x-, y-, and z-axes, respectively, (Figure 2.10) [218, 219]. It is done by first identifying B-X bonds, matching each octahedron to an ideal cubic octahedron via polyhedral template matching, then extracting the best-fit rotation (quaternion) and converting it to Euler angles, and selecting the internal-angle branch that orders the angles by magnitude.

Knowledge of the instantaneous rotation angle at each site is insufficient to fully characterize tilt order. While angle magnitudes quantify local distortions, they do not encode the relative phase between neighboring octahedra, nor do they capture domain morphology or the persistence of local order. Consequently, two complementary characterizations are required: spatial organization, which determines in-phase versus anti-phase tilts and sets correlation lengths; and temporal organization, which governs the persistence of local order, interwell-hopping kinetics, and relaxation times near structural transitions. For these reasons, spatial and temporal two-point correlation functions are employed to quantify the underlying

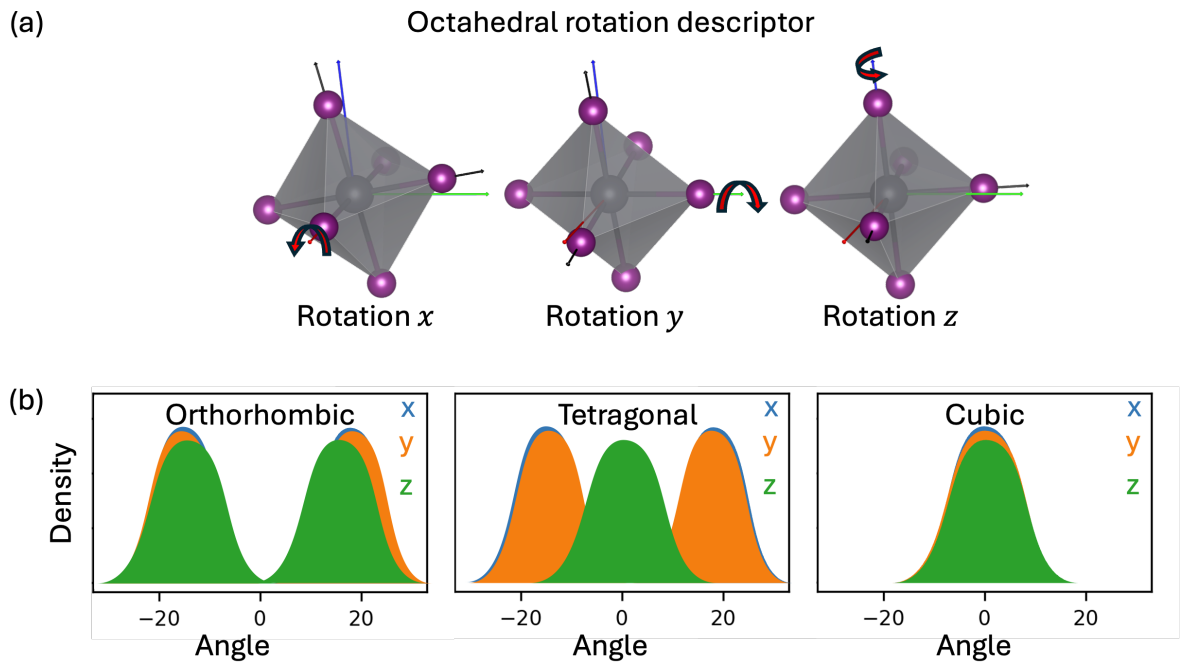


Figure 2.10 (a) Definition of octahedral tilt angles θ, ϕ, and ψ as rotations about the pseudocubic x-, y-, and z-axes, respectively. (b) Schematic of octahedral tilt patterns and corresponding phases: orthorhombic, tetragonal, and cubic.

order.

Let $\theta_i(\mathbf{n}, t)$ be the rotation of the BX_6 octahedron at lattice site $\mathbf{n}$ about pseudocubic axis $i \in \{x, y, z\}$ at time t. Let $\hat{\mathbf{e}}_j$ be the unit vector along axis $j \in \{x, y, z\}$. Averages $\langle \cdot \rangle_{t,\mathbf{n}}$ run over time and sites. Spatial correlation along axis j at separation d:

$$C_{ij}^{spat}(d) = \frac{\langle \theta_i(\mathbf{n}, t)\, \theta_i(\mathbf{n} + d\hat{\mathbf{e}}_j, t) \rangle_{t,\mathbf{n}}}{\langle \theta_i(\mathbf{n}, t)^2 \rangle_{t,\mathbf{n}}} . \tag{2.78}$$

Temporal fluctuations are characterized by the normalized autocorrelation function:

$$C_{ii}^{temp}(\tau) = \frac{\langle \theta_i(\mathbf{n}, t)\, \theta_i(\mathbf{n}, t + \tau) \rangle_{t,\mathbf{n}}}{\langle \theta_i(\mathbf{n}, t)^2 \rangle_{t,\mathbf{n}}} , \tag{2.79}$$

where τ is the lag time, i.e., the time separation between two evaluations of the same tilt variable along a trajectory.

These correlation functions quantify the spatial extent and temporal persistence of local tilt order, thereby characterizing the underlying phase behavior. However, capturing these polymorphs and their transition pathways requires extensive simulations on large time scales ($\sim$ 1 ns), preferably on large supercells ($\sim 10^3$ atoms) to sample collective tilts and phase coexistence.

2.4.2 Challenges in Modeling Halide Perovskites

Accurate computational modeling of halide perovskites presents unique challenges due to their complex structural dynamics, multiscale phenomena, and unusual defect physics — all of which demand specialized simulation approaches beyond conventional methods.

Halide perovskites are structurally flexible and dynamically disordered, with complex ionic and electronic responses that span multiple time and length scales. Additionally, organic A-site cations (MA^+, FA^+) add fast, coupled degrees of freedom. These exhibit rapid reorientations within their inorganic cages, and the BX_6 octahedral network displays pronounced anharmonic phonon modes, this further impacts the electronic properties of these materials, as shown by early DFT-based molecular dynamics simulations [220]. These dynamics invalidate the harmonic approximation. Standard harmonic phonon calculations fail to reproduce experimentally observed vibrational features (e.g. a mode near 200 cm^{-1} in $MAPbI_3$), whereas velocity-autocorrelation-based analyses in MD recover these anharmonic signatures [221].

These modeling challenges become even more pronounced when considering lower-dimensional perovskite structures. In 2D perovskites, surface and interface effects become dominant, requiring accurate treatment of electrostatics and dispersion interactions. Asymmetric repeated-slab models generate artificial fields unless a dipole correction is applied. Slabs require generalized schemes to suppress spurious potential ramps and recover reliable adsorption energies, reaction barriers, and interfacial band offsets. Dispersion interactions must also be included, since nonlocal vdW contributions substantially affect lattice parameters, binding distances, and molecular orientations at perovskite surfaces [218]. It requires accuracy beyond classical force fields, since neglecting either a dipole correction or a dispersion interaction yields qualitatively incorrect adsorption profiles and interfacial dynamics [222].

Beyond structural dynamics, the interplay between electronic and ionic transport presents additional modeling challenges. Halide perovskites combine relative defect tolerance in recombination with pronounced ionic mobility [223, 224, 225]. Point defects and mobile ions govern local fields in halide perovskites [226, 227, 228]. Vacancy-mediated halide transport has low activation energies ($\sim$0.1-0.6 eV), yielding mixed ionic-electronic response under bias and illumination [229, 230, 228, 231, 232]. At the atomistic scale, realistic modeling must resolve defect formation and annihilation, sample charge-state equilibria, and evaluate correlated hopping under controlled local electrostatic or photochemical perturbations [233, 234, 232].

The structural dynamics and ionic transport discussed above have a fundamental impact on electronic properties, creating additional computational challenges. The structural flexibility of halide perovskites strongly affects their electronic structure, producing a sizable thermal

dependence of the band gap [235] and altering the effective carrier mass, dielectric screening, and exciton binding [236]. This flexibility also creates shallow trap states, as revealed by temperature-resolved Urbach-tail analysis [237]. Accurate electronic structure simulations must simultaneously account for relativistic effects (particularly with heavy Sn, Bi, or Pb atoms) and electronic correlation. Notably, early studies on $MAPbI_3$ using standard DFT within the GGA approximation yielded band gaps in surprisingly good agreement with experiment [238, 239]. Subsequent research revealed this agreement was fortuitous, resulting from error cancellation between spin-orbit coupling and electronic correlation effects [240, 241].

The multifaceted challenges described above can be effectively addressed through MLFFs trained on high-fidelity data. By enabling long, large-cell MD, MLFFs capture dynamic disorder and anharmonic phonons, reproducing phase behavior and octahedral-tilt statistics in hybrid and inorganic perovskites [242, 170, 34, 218]. In 2D perovskites, models trained on vdW-corrected data retain the nonlocal physics needed for reliable adsorption energies, reaction barriers, and interfacial band offsets. For ionic transport, MLFF MD resolves vacancy and interstitial diffusion pathways and activation barriers in $CsPbX_3$, yielding diffusion constants and mechanistic insight at device-relevant scales [243]. Coupling MLFF-generated structural ensembles to electronic-structure evaluation then quantifies how thermal disorder modulates band edges and trap statistics. Surface simulations indicate that dynamics bias toward shallow rather than deep states [222]. In terms of scale and throughput, equivariant MPNN (e.g., MACE) and kernel models (e.g., GAP) deliver near-first-principles accuracy at far lower cost, enabling nanosecond trajectories in 10^3-10^4-atom supercells to map free-energy landscapes, compute diffusion, and sample interfacial configurations for band-edge analysis [24, 25, 137]. Thus, MLFF-based simulations of halide perovskites provide insights that traditional approaches struggle to deliver, placing them at the forefront of current research.

Chapter 3

Halide Perovskite Interfaces: from Dispersion Interaction to Electronic Structure at Finite Temperature

Halide perovskites have emerged as revolutionary optoelectronic materials, renowned for their exceptional combination of tunable band gaps, high charge-carrier mobilities, and defect tolerance that rivals traditional semiconductors [244, 245]. These ABX_3 compounds (with A = monovalent cation, B = divalent metal, X = halide) have enabled solar cells with power conversion efficiencies exceeding 25% [188] and light-emitting diodes with external quantum efficiencies above 20%. However, despite rapid progress, the long-term stability of halide perovskite devices remains a critical challenge: these materials are structurally soft and chemically volatile, prone to phase instabilities and ionic migration under operating conditions [246, 231]. Understanding how the intricate structural dynamics and ionic transport at the atomistic level impact electronic properties is crucial for improving the robustness and performance of perovskite-based devices. Finite-temperature phenomena such as dynamic disorder, ion vacancy migration, and band gap fluctuations must be captured by accurate atomic-scale simulations to inform strategies for enhanced stability and efficiency.

Despite the success of MLFF approaches in bulk perovskites, extending them to surfaces and heterogeneous interfaces remains challenging. Broken translational symmetry at terminations lowers coordination numbers and produces distinct chemical environments relative to the bulk, promoting relaxation, reconstruction, and the emergence of under-coordinated sites

and electronic surface states. Polar terminations and interfacial dipoles generate long-range electrostatic fields that decay slowly into the material or vacuum, while dispersion interactions can remain significant across gaps between facing slabs. Accurate modeling therefore requires careful control of finite-size effects, including sufficiently thick slabs, large lateral supercells, and appropriate vacuum spacing and dipole corrections, as well as systematic exploration of multiple metastable reconstructions.

Traditional *ab initio* treatments of such surfaces are computationally demanding because realistic models often contain hundreds of atoms and many competing configurations. This is precisely where MLFFs can be transformative: trained on dispersion-corrected and field-sensitive quantum data, they deliver near-first-principles accuracy at much lower cost, enabling large-scale slab calculations, exhaustive sampling of reconstruction pathways, and statistically meaningful thermodynamic averages. For interfacial systems, an effective MLFF must capture long-range electrostatics and vdW contributions, either by incorporating them explicitly (e.g., long-range baselines with short-range machine-learned residuals) or by training on datasets that span varying slab thicknesses, terminations, and applied fields. With these elements in place, MLFFs make it feasible to probe surface relaxations, defect formation at terminations, reconstruction energetics, and the coupling between structural motifs and electronic surface states at scales inaccessible to routine *ab initio* simulations.

In this chapter, we address the above challenges by developing and applying MLFF-driven simulations for halide perovskite interfaces, with a focus on surfaces and their influence on electronic properties. We employ slab geometry models to explicitly represent the perovskite surfaces and heterostructures of interest. This approach enables us to capture the broken-symmetry environment of a surface (including any unique surface reconstruction or distortion) and to address the challenges presenting in taking into account interactions across the interface between 2D and 3D perovskite regions. By leveraging a MLFF trained on DFT data, we bridge the quantum and atomistic scales: the MLFF provides near-DFT accuracy for forces and energies, while being efficient enough to sample extensive atomic configurations and thermal fluctuations at the interface. The following sections will detail how this MLFF is constructed and validated for the target halide perovskite system, and how it is used to investigate structural dynamics, finite-temperature effects, and electronic structure at the surface. Through this multiscale modeling strategy, the present work takes a step toward connecting atomistic insights with device-level behavior, helping to unravel how interfacial atomic structure and dynamics influence the macroscopic performance and stability of halide perovskite materials.

3.1 Frozen Atomistic and Electronic Structure

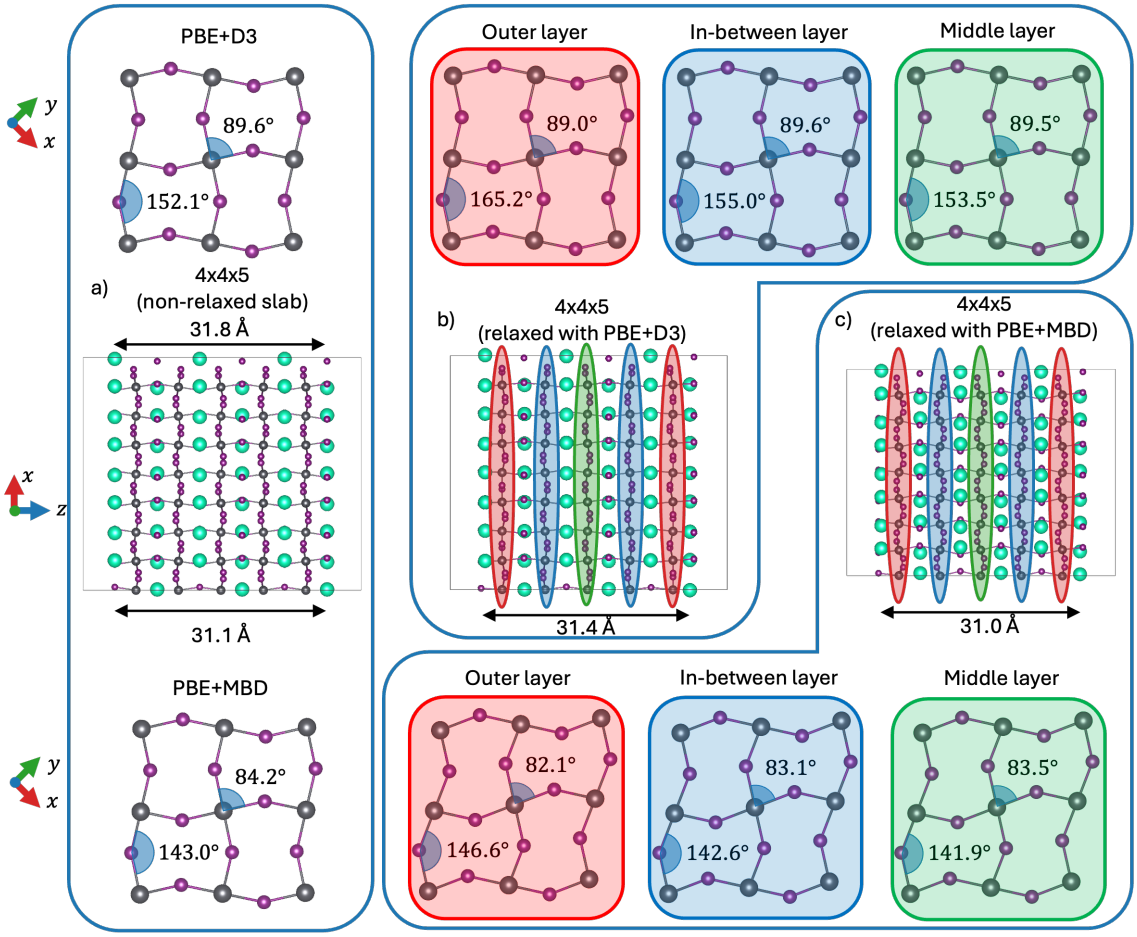


Figure 3.1 The visualization $CsPbI_3$ slab structures geometry (a) without relaxation of the optimized bulk structure after introducing surface (b) after relaxation using the PBE+D3 level of theory, and (c) after relaxation using the PBE+MBD-NL level of theory.

Orthorhombic (γ) $CsPbI_3$ exhibits cooperative PbI_6 octahedral tilting at room temperature that fundamentally determines its electronic structure. Surface formation disrupts these tilt patterns through coordination changes and modified dielectric screening, affecting both structural and electronic properties. Because these perturbations are mediated by dispersion interactions, comparing two widely used dispersion-corrected density-functional treatments—PBE+D3 and PBE+MBD-NL—reveals how the underlying vdW physics controls near-surface properties and bulk-like behavior recovery [70, 76].

Although both schemes are detailed in 2.1.2 and 2.1.3, their underlying physics diverges in ways that matter at surfaces. PBE+D3 adds pairwise C_6/R^6 terms with empirical damping, so the dispersion energy is a sum over atom pairs with no explicit collective screening. PBE+MBD-NL treats atoms as coupled quantum harmonic oscillators, yielding screened

many-body dispersion that is non-additive, anisotropic, and environment dependent. At surfaces, coordination decreases and dielectric screening is inhomogeneous. These conditions amplify many-body and screening effects that MBD-NL can represent but D3 cannot by construction. The choice of scheme therefore affects adsorption energies, interlayer binding, and reconstruction energetics.

Starting from the experimentally determined orthorhombic bulk structure of $CsPbI_3$, we optimized a $4 \times 4 \times 4$ supercell at both the PBE+D3 and PBE+MBD-NL levels. From each relaxed bulk, we cleaved a stoichiometric slab with symmetric CsI terminations, consistent with reports that CsI-terminated facets are among the lowest-energy surfaces of γ-$CsPbI_3$ [27, 247, 248]. For each dispersion treatment we then performed variable-cell slab relaxations, allowing both ionic positions and lattice vectors to relax. We quantified the slab thickness and layer-resolved octahedral metrics—the in-plane I-Pb-I angle and the Pb-I-Pb bridging angle within the PbI_2 planes.

Figure 3.1 contrasts the PBE+D3 and PBE+MBD-NL for the relaxed bulks and slabs structures. Already in the bulk, PBE+MBD-NL yields a 7.7% smaller unit cell volume than PBE+D3, consistent with prior benchmarks showing that MBD-NL typically produces more compact structures for soft materials [76]. Introducing the surfaces and relaxing the structure, a net thickness contraction of the slab, with the PBE+D3 method yielding a 1.6% (0.4 Å) contraction. In the outermost PbI_2 layer the in-plane I-Pb-I angle decreases from its bulk-like value $\sim$89.6° to $\sim$89.1° , while the Pb-I-Pb angle experiences an increase from $\sim$152.1° to $\sim$165.7°. The distortion decays rapidly: the in-plane I-Pb-I angle returns to $\sim$89.5° and the Pb-I-Pb angle to $\sim$153.5° after just two layers, making it possible to use the inner layer as a good approximation of the bulk structure. Relative to PBE+D3, PBE+MBD-NL method yields a smaller overall contraction of 0.1 Å and the layer-resolved angle distortions across the slab. The Pb-I-Pb angle shows the largest sensitivity in all PbI planes, reaching $\sim$146.6° in the outermost layer and to $\sim$141.9° in the middle layer. Overall, inclusion of MBD-NL yields a denser equilibrium lattice and slightly stronger near-surface angular distortions than D3, noting that these 0 K relaxations neglect thermal expansion and anharmonic effects presented in the experimental data.

These structural differences translate directly into electronic properties. We analyzed band-decomposed Valence Band Maximum (VBM) and Conduction Band Minimum (CBM) charge densities for all slab models. The as-cleaved slabs cut from PBE+D3- and PBE+MBD-NL-relaxed bulks produced indistinguishable depth profiles within numerical uncertainty, so to avoid redundancy Figure 3.2 reports three representative cases: the as-cleaved slab cut from the PBE+D3-relaxed bulk, and the slabs fully relaxed with PBE+D3 and with PBE+MBD-NL.

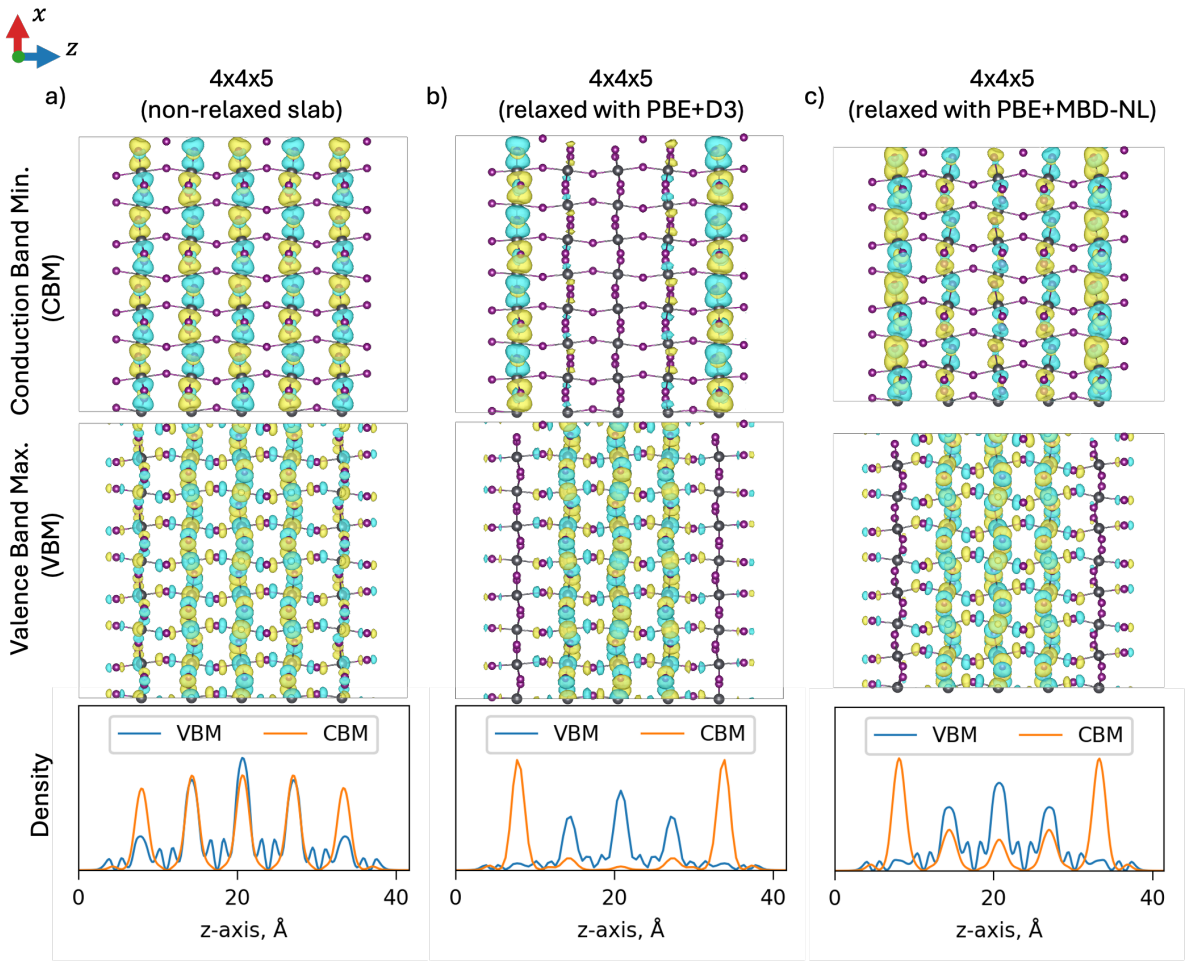


Figure 3.2 Spatial visualization of the Kohn-Sham orbitals for $CsPbI_3$ slab structures, illustrating the conduction band minimum (the upper row), the valence band maximum (the middle row), the projection of these orbitals along the z-axis (the bottom row). Results are presented for three scenarios: (a) without relaxation after creation from the optimized bulk structure, (b) after relaxation using the PBE+D3 level of theory, and (c) after relaxation using the PBE+MBD-NL level of theory. Cesium atoms are not presented in figures for convenience.

For each case we rendered real-space isosurfaces and planar-averaged the band densities along the surface normal to obtain depth-resolved localization profiles. Column (a) shows that introducing a surface while keeping atomic positions unrelaxed drives a localization of the VBM. In contrast, the CBM is only weakly perturbed and remains largely delocalized across the slab. The band gap increases by 0.20 eV from the bulk (1.59 eV) to the unrelaxed five-layer slab (1.79 eV). This opening is consistent with residual quantum confinement at this thickness, with shifts reported to persist up to ≈10 layers [249]. After relaxation the dispersion model governs both gap and localization. With PBE+D3 (column (b)), straightening of Pb-I-Pb linkages reduces the gap to 1.65 eV and enhances VBM confinement in the central layer while shifting the CBM toward the surfaces. With PBE+MBD-NL (column (c)), stronger residual octahedral tilts increase the gap to 1.86 eV and yield weaker changes in localization, consistent with more uniform layer-by-layer distortions that attenuate the sensitivity of

orbital localization to relaxation.

These pronounced method-dependent differences at 0 K prompt the question of how they propagate to finite-temperature dynamics. This motivates the development of accurate MLFFs that encode dispersion and enable long, large-scale molecular dynamics to capture thermal distortions and dynamical surface responses.

3.2 MLFF Accuracy and Dispersion Correcetions

Having analysed the structural and electronic properties of halide perovskite slabs and having assessed distinct behaviours from D3 and MBD-NL, we now move to develop MLFF for this system. To achieve this, we employ MACE-based MLFFs, discussed in 2.3.3 to achieve near-DFT accuracy while capturing the complex interactions in $CsPbI_3$ slabs. Due to code- and element-specific costs rather than the dispersion model itself, we adopted a two-stage protocol. Long MD trajectories were generated at the PBE+D3 level with CP2K/Quickstep, whose mixed GPW formalism and pseudopotentials are optimized for scalable MD. We then re-evaluated selected frames with PBE+MBD-NL in FHI-aims. FHI-aims is an all-electron NAO code, so Pb-containing systems are markedly more expensive per Self-Consistent Field (SCF) step than in pseudopotential-based GPW calculations. The dominant cost difference in our workflow therefore stems from the electronic-structure engines and electron treatment; while MBD-NL has higher asymptotic cost than pairwise D3, this overhead is secondary here. This procedure yields paired datasets—identical geometries with distinct energy and force labels—for direct comparison of dispersion models via subsequent MLFFs, with simulation details in Section A.5.

The structural differences observed between $CsPbI_3$ slab structures optimised using PBE+MBD-NL and PBE+D3 are results of differences with the underlying driving forces. Figure 3.3 compares the surface-normal (z) component of the vdW forces from both methods across all configurations in created datasets. The scatter plot for all atoms, depicted in Figure 3.3(a), shows three distinct clusters: a dominant one centred near zero MBD-NL force, and two smaller clusters centred at ± 3 kcal/mol·Å. These smaller clusters arise primarily from the z-components of surface Cs atoms, while the large one mainly consist of atoms in layers inside the slab. Restricting the analysis to surface Cs and I atoms, shown in Figure 3.3(b), confirms that the largest method-dependent force differences occur at the surface. This surface-selective force disparity is consistent with the greater thickness contraction and enhanced octahedral-angle distortions observed after MBD-NL relaxation (Figure 3.1), which in turn produce a larger band-gap increase relative to D3. These different surface forces drive

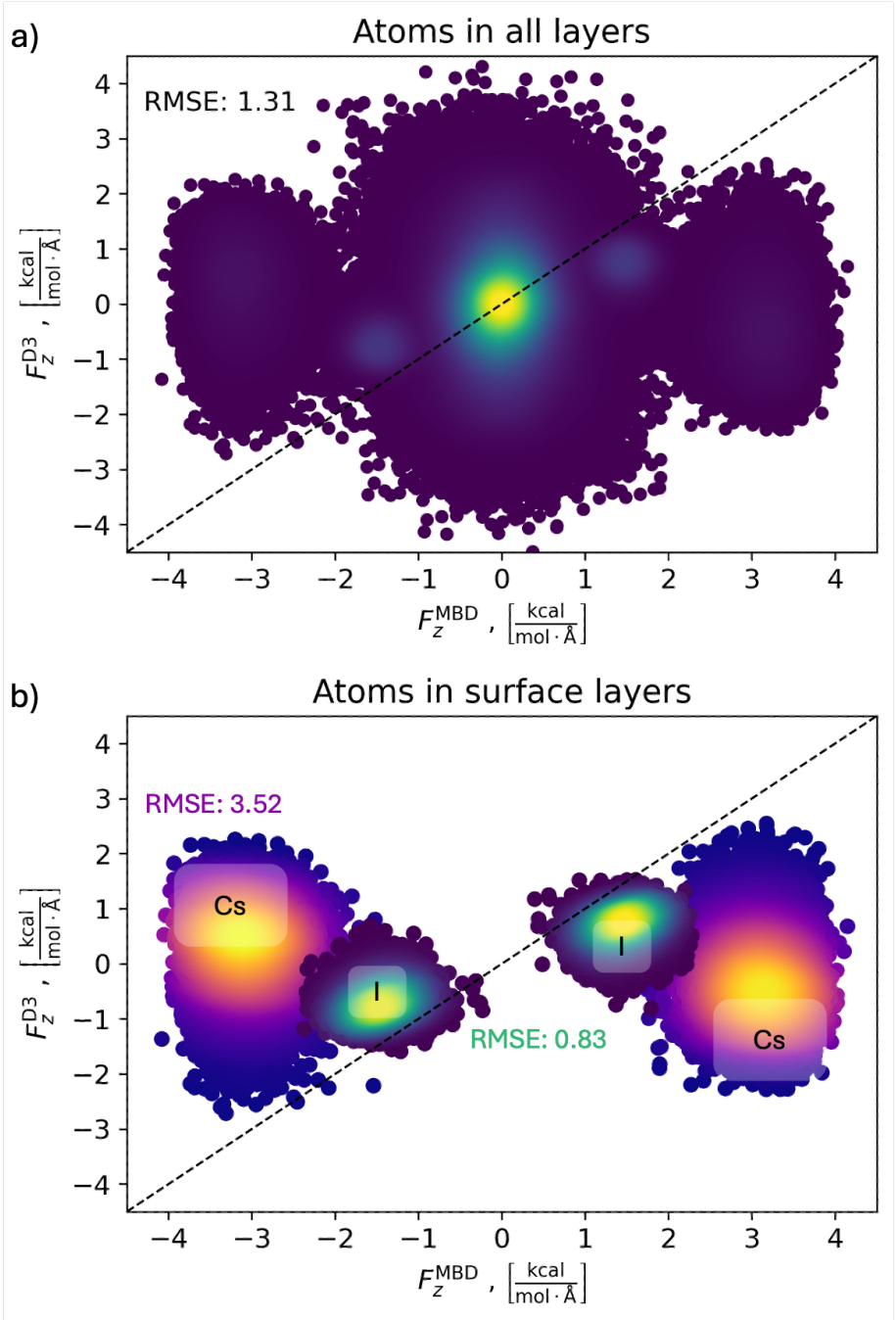


Figure 3.3 Density scatter plot of the z-component of vdW forces computed with the MBD-NL and D3 schemes for sampled frames from the 300 K MD trajectory of a $CsPbI_3$ slab used to train the MLFF models. Each point represents a pair of z-component forces acting on a single atom. The local point density encoded by a continuous heat map, where darker colors indicate lower density and lighter colors indicate higher density. Local point density is encoded by a continuous heat map, with lighter colors denoting higher density and darker colors denoting lower density. The black dashed line indicates perfect parity between the two methods. (a) Compared the vdW forces for all atoms in the slab structure. (b) Compared the vdW forces for the Cesium and Iodine atoms in the surface layer.

distinct dynamic behaviors that ultimately affect electronic properties, as we explore in the following section.

To ensure our MLFFs accurately reproduce their respective reference theories, Figure 3.4 presents a comprehensive error analysis using held-out validation configurations. Panel (a) shows that both models achieve similar overall accuracy, with median MAEs of 0.19 kcal/(mol·Å) for PBE+MBD-NL (red, dashed) and 0.21 kcal/(mol·Å) for PBE+D3 (green, solid). The configuration-resolved errors in panel (b) reveal expected variation across different geometries, with higher errors for structures sampled at 600 K due to greater thermal distortions. Panel (c) decomposes errors by atomic species and slab position, showing that both models exhibit similar error patterns: larger errors for lead atoms (middle) compared to cesium (top) and iodine (bottom), and higher errors in the middle layer (green) versus outer layers (blue). These consistent error patterns across both models ensure that any observed differences in subsequent dynamics arise from the underlying dispersion physics rather than model accuracy disparities.

3.3 Finite-Temperature Dynamics and Electronic Structure

Having established that (1) PBE+D3 and PBE+MBD-NL predict significantly different surface forces and (2) both MLFFs accurately reproduce their respective reference forces with comparable error distributions, we now investigate how these distinct dispersion treatments control finite-temperature properties. By applying these validated models to large-scale molecular dynamics simulations, we can directly observe how the differences in vdW physics manifest in thermal framework dynamics and electronic structure evolution.

While static calculations provide valuable insights into ground-state properties, understanding finite-temperature behavior is essential for predicting device performance under operating conditions. The choice of dispersion treatment becomes especially critical at elevated temperatures, where soft vibrational modes govern framework flexibility and electronic coupling.

3.3.1 Framework Dynamics and Phase Behavior

Using the two trained MLFFs, we performed NPT MD on an $8 \times 8 \times 5$ $CsPbI_3$ slab ($N = 3456$ atoms) to probe how the dispersion model governs finite-temperature structure and electronic properties. Figure 3.5 demonstrates octahedral-tilt statistics and correlations in an $8 \times 8 \times 5$

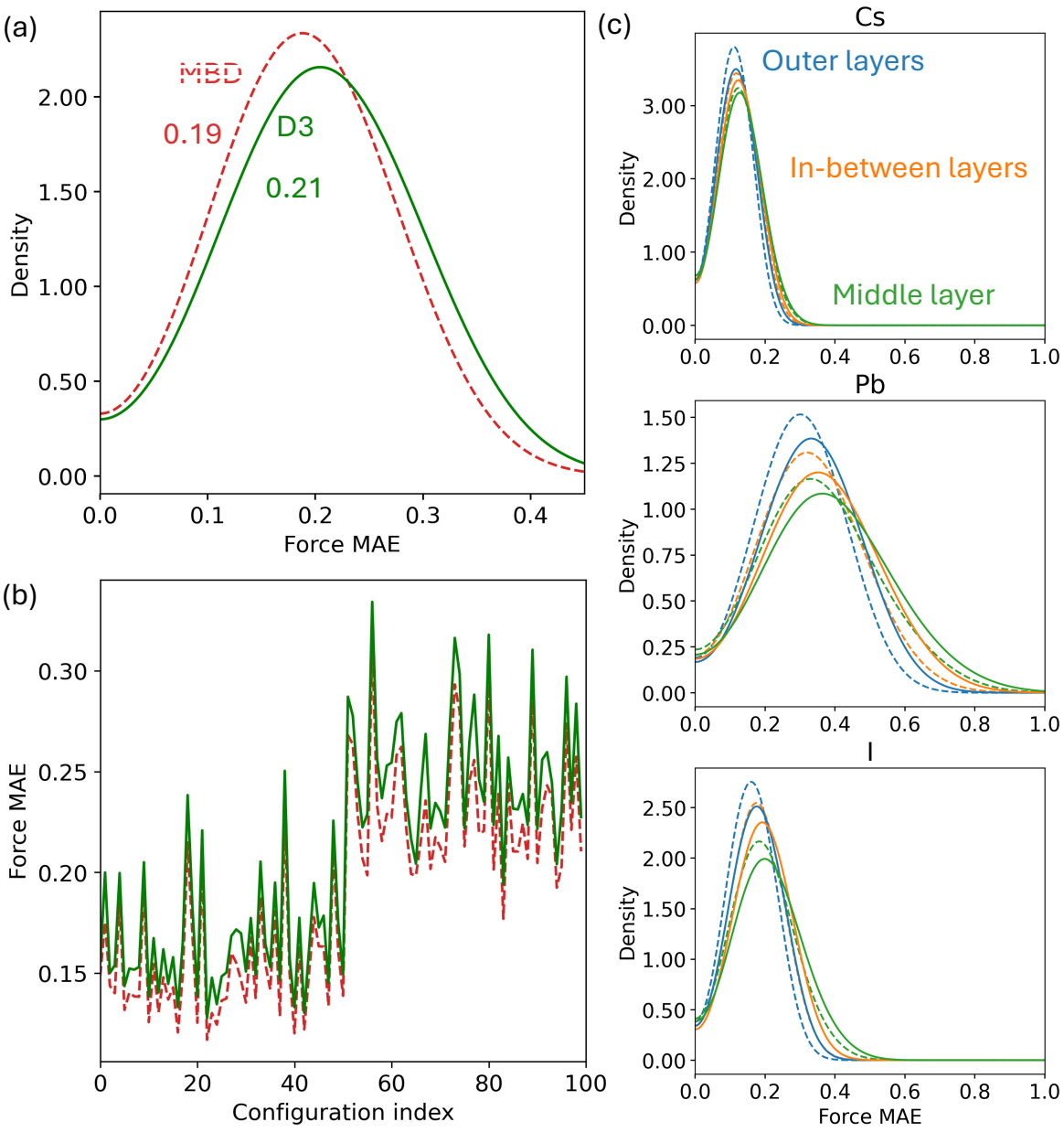


Figure 3.4 Comparative analysis of force-error statistics on the validation set for MLFFs trained on dispersion-corrected DFT references for $CsPbI_3$ slabs. (a) Probability density of force MAE, comparing models trained on PBE+D3 (green, solid line) and PBE+MBD-NL (red, dashed line). (b) Configuration-resolved force MAE over 100 validation configurations for the same two models, revealing variability across distinct slab geometries extracted from 300 K and 600 K dynamics. (c) Species- and layer-resolved force MAE densities for Cs (top), Pb (middle), and I (bottom), partitioned by atomic position within the slab: outer layers (blue), in-between layers (orange), and middle layer (green); solid lines PBE+D3 and dashed lines show PBE+MBD-NL. All reported force errors are in kcal/(mol·Å);

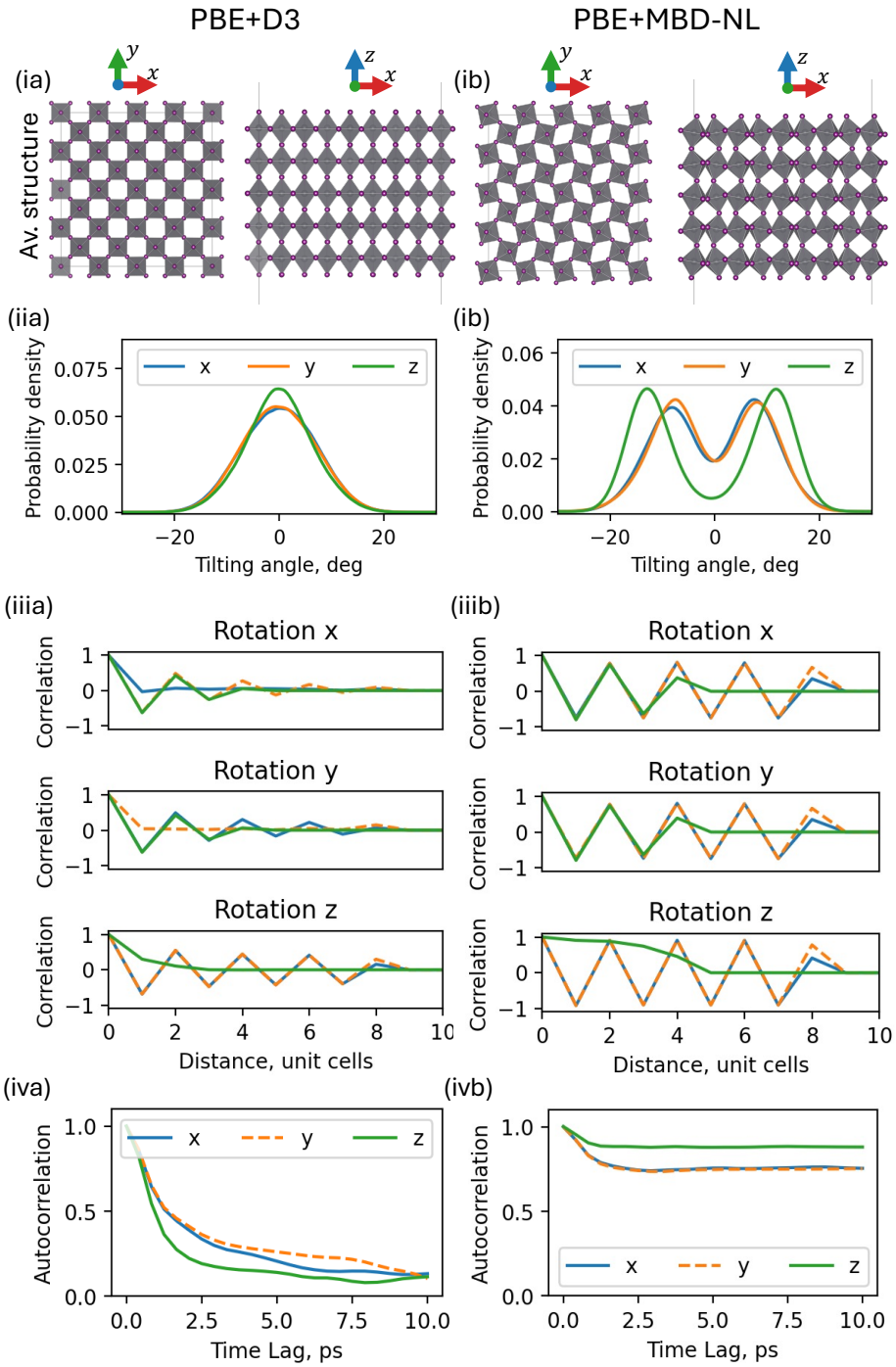


Figure 3.5 Result of octahedral tilting analysis from NPT MD at 300 K of $8 \times 8 \times 5$ $CsPbI_3$ slab structure using MLFF models trained on forces, energies, and stress tensors obtained from (column a) PBE+D3 and (column b) PBE+MBD-NL levels of theory. Rows: (i) Time-averaged structures of an $CsPbI_3$ slab from MD simulations, Cs omitted to highlight PbI_6 octahedra, (ii) rotational angles distributions, (iii) spatial correlation of octahedra rotational angles along all tilting directions (iv) temporal auto-correlations of all octahedra rotational angles extracted from NPT MD at 300 K of $8 \times 8 \times 5$ $CsPbI_3$ slab structure using MLFF models trained on forces, energies, and stress tensors obtained from (a) PBE+D3 and (b) PBE+MBD-NL levels of theory.

$CsPbI_3$ slab at 300 K from 5 ns of NPT MD with each model. For the MLFF model trained on PBE+D3 data (column (a)), the measured distributions in (iii) are narrow and centered at 0°, while the temporal autocorrelations in (iv) decay rapidly, and spatial correlations in (v) are short-ranged with weak oscillations. All of it being evidence that the PBE+D3 model reproduce cubic (α) phase behavior. In contrast, the PBE+MBD-NL model (column (b)) yields bimodal angle distributions with finite tilts along all axes; the temporal autocorrelations remain high over picosecond lags, and the spatial correlations display longer-range, alternating order across unit cells, matching the orthorhombic phase signature. Thus, inclusion of many-body dispersion stabilizes orthorhombic tilt pattern at 300 K, whereas pairwise D3 biases the slab toward an artificially cubic response. The former agrees with experiments that observe γ-$CsPbI_3$ near room temperature in stabilized thin films and in single crystals [250, 251].

Seeing that the two dispersion treatments yield qualitatively different structural dynamics at 300 K, one might expect them to also predict distinct phase behaviors upon temperature changes. Starting from the respective relaxed PBE+D3 and PBE+MBD-NL slabs, we heated from 50 K to 500 K over 4.5 ns, i.e. at a rate of 0.1 $\mathrm{K\,ps^{-1}}$. Figure 3.6 reports the temperature evolution of the in-plane lattice parameters a and b, and the slab thickness c_{slab} (defined as the instantaneous surface-to-surface separation along the normal, excluding vacuum), alongside probability densities of PbI_6 rotation angles about the x, y, and z axes. The PBE+MBD-NL model captures the experimentally observed thermal response of $CsPbI_3$, shown in Figure 2.9 (c) more faithfully: with increasing temperature, the in-plane lattice parameters exhibit anisotropic expansion and the octahedral tilt distributions progressively shift toward 0° merging into a single peak near 450 K, indicating a gradual transition from orthorhombic to cubic-like behavior. In contrast, the PBE+D3 model collapses the tilts prematurely and drives $|a - b| \to 0$ substantially lower temperatures (near 175 K), yielding an early onset of near-cubic behavior. The slab thickness c also expands more rapidly with D3, reflecting the weaker interlayer binding that we observed in static calculations. Experimentally for the bulk $CsPbI_3$, the stabilization of the cubic phase occurs at much higher temperatures, around 600 K [27, 205]. However, surface-, size-, and strain-related effects, known to stabilize the black perovskite phase at reduced temperatures, are amplified in slab geometries and rationalize part of the downward shift in the stabilization temperature relative to bulk measurements [252, 253, 254, 205, 255].

3.3.2 Electronic Structure Evolution

To assess how these structural differences influence electronic properties, we sampled configurations from each trajectory and evaluated band-structure properties at the PBE level. As was

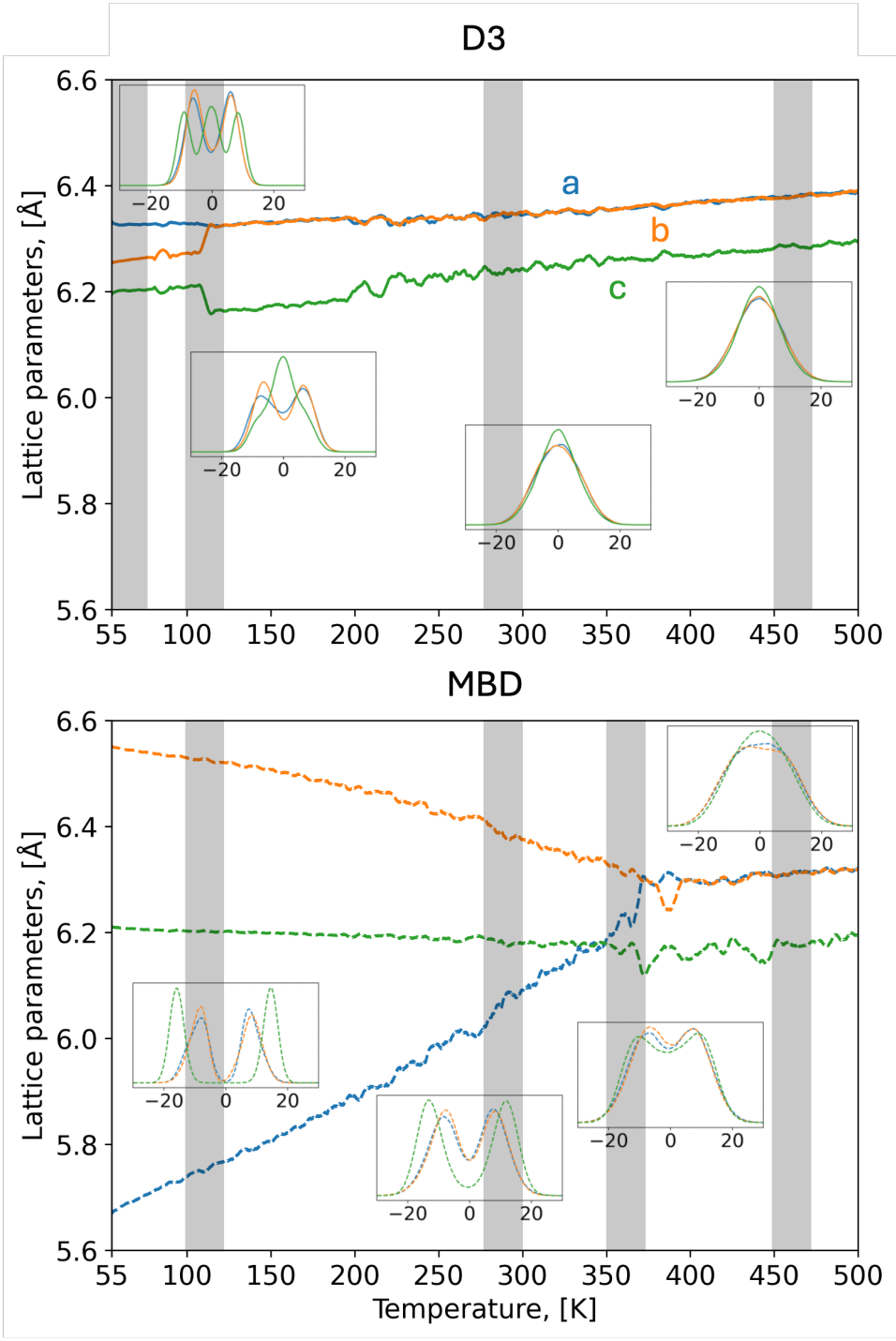


Figure 3.6 Temperature dependence of lattice parameters and octahedral tilts in a $CsPbI_3$ 8x8x5 slab from heating MD with MACE force fields. Top: model trained on PBE+D3. Bottom: model trained on PBE+MBD-NL. Curves show the three lattice parameters a, b, and c (left) as functions of temperature. Grey shaded bands mark temperature windows for which the insets report the probability densities of PbI_6 octahedral rotation angles (in degrees) around the x, y, and z axes. The panels inside plots illustrate the rotational angle distributions of PbI_6 octahedra at different temperature ranges, shown as a grey area under the panels on the plots.

discussed in 2.1.1, PBE often yields band gaps in reasonable agreement with experiment due to fortuitous error cancellation rather than a systematic treatment of exchange-correlation. Nevertheless, the size of our models (>3,000 atoms) renders higher-level approaches computationally prohibitive.

In the absence of thermally induced ionic motion, i.e., 0 K, introducing vacuum-exposed surfaces and relaxing the slab yields preferential localization of the VBM in the interior and the CBM at the surface, as shown in Figure 3.2 (b-c). Figure 3.7 extends this analysis to finite temperature by tracking the time evolution of the VBM and CBM in a $8 \times 8 \times 5$ $CsPbI_3$ slab at 300 K. Panel (a) displays representative isosurfaces from a randomly selected MD frame and outlines the workflow used to track the time evolution of the center of mass (COM) and SD of the VBM and CBM. Depth profiles in panels (b) and (c) (top) show strong CBM weight at the outermost PbI_2 planes and a broader, more interior VBM, consistent with Pb-6p contributions dominating the CBM and I-5p character dominating the VBM. Thermal motion drives continuous evolution of frontier-orbital localization: although the CBM and VBM preferentially reside at the surface and in the bulk, respectively, their positions oscillate on picosecond timescales, and their spatial extent varies, as captured by the SD. The time-depth maps (bottom) resolve band-edge center switching between equivalent surfaces and show no persistent mid-gap surface states, as expected for clean CsI-terminated slabs [248]. Relative to MBD-NL, the D3 trajectory exhibits longer residence times and tighter CBM confinement at the outer PbI_2 planes, while the VBM remains more interior; MBD-NL yields broader band-edge spreads and more frequent surface switching.

It is not evident that the observed VBM/CBM localization alone sets the D3–MBD-NL band-gap difference. Figure 3.8 shows the 3 ps evolution of the total dipole-moment magnitude, $|\boldsymbol{\mu}(t)|$, and the instantaneous band gap, $E_g(t)$, for each trajectory. Both models are centrosymmetric on average ($\langle\boldsymbol{\mu}\rangle \approx 0$), yet their dynamics differ: the standard deviation of $|\boldsymbol{\mu}|$ is 3.86 $e \cdot$ Å for D3 and 2.58 $e \cdot$ Å for MBD-NL. The gap distributions are likewise separated, with a larger time-averaged E_g under MBD-NL (1.55 eV) than under D3 (1.32 eV). The coincidence of suppressed dipole fluctuations and an enlarged gap in MBD-NL is consistent with weaker instantaneous internal fields and reduced transient disorder. Thus, the gap trend is governed by mechanisms beyond band-edge localization.

3.4 Summary

Finite-temperature fluctuations in $CsPbI_3$ halide perovskite slabs profoundly affect its structural and electronic properties, as demonstrated in this chapter, whereas these effects were

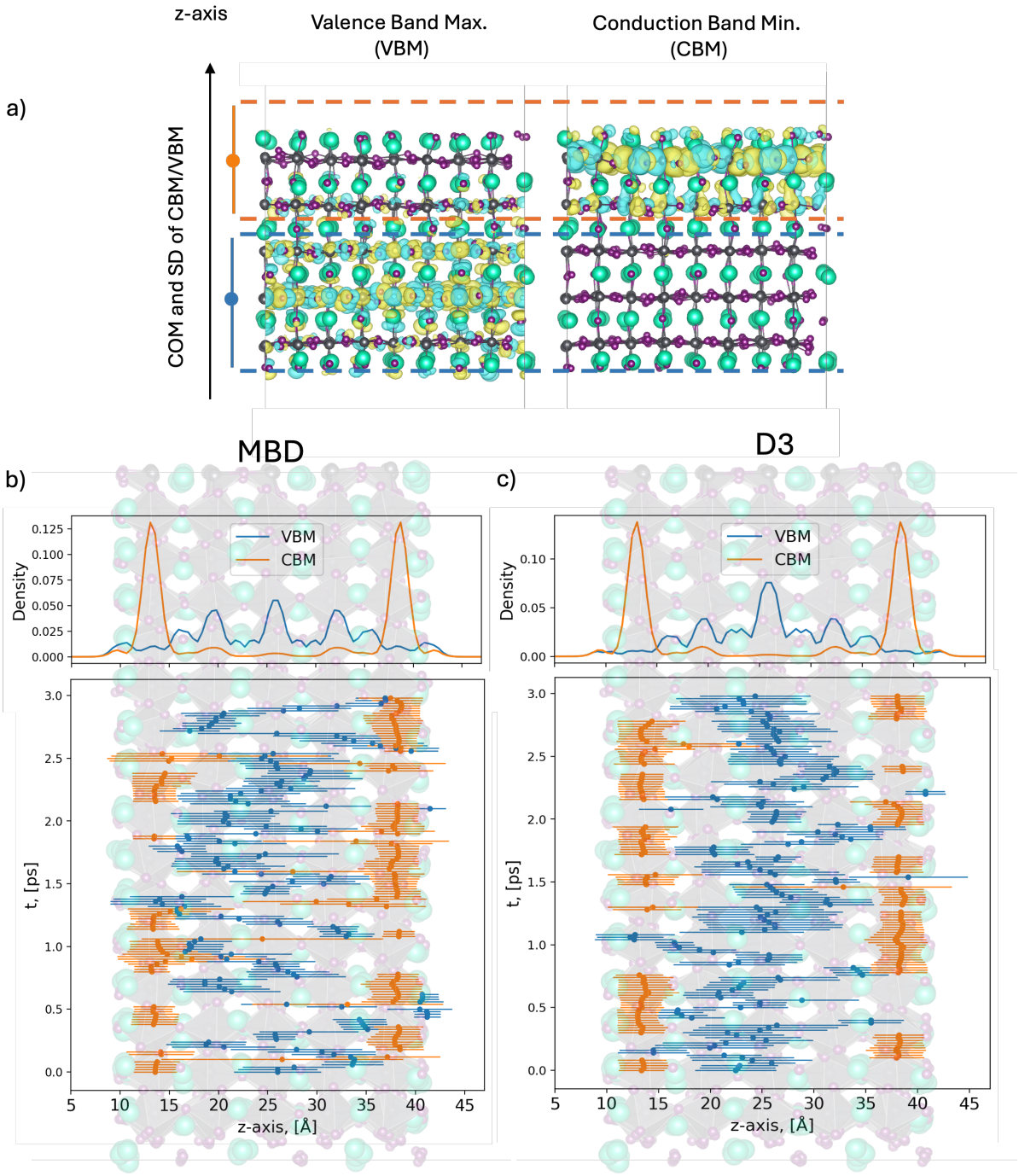


Figure 3.7 Frontier-orbital analysis of a $CsPbI_3$ slab structure at 300 K. (a) Isosurfaces of the VBM (left) and CBM (right) for a randomly selected frame from a MD trajectory of a $4 \times 4 \times 5$ slab. (b,c) For a $8 \times 8 \times 5$ slab, , time series of the VBM and CBM Center of Mass (COM) along the slab normal with Standard Deviation (SD) envelopes, together with time-averaged orbital density projected onto the slab's principal z axis. Results are from MACE MLFFs trained on PBE+D3 (b) and PBE+MBD-NL (c).

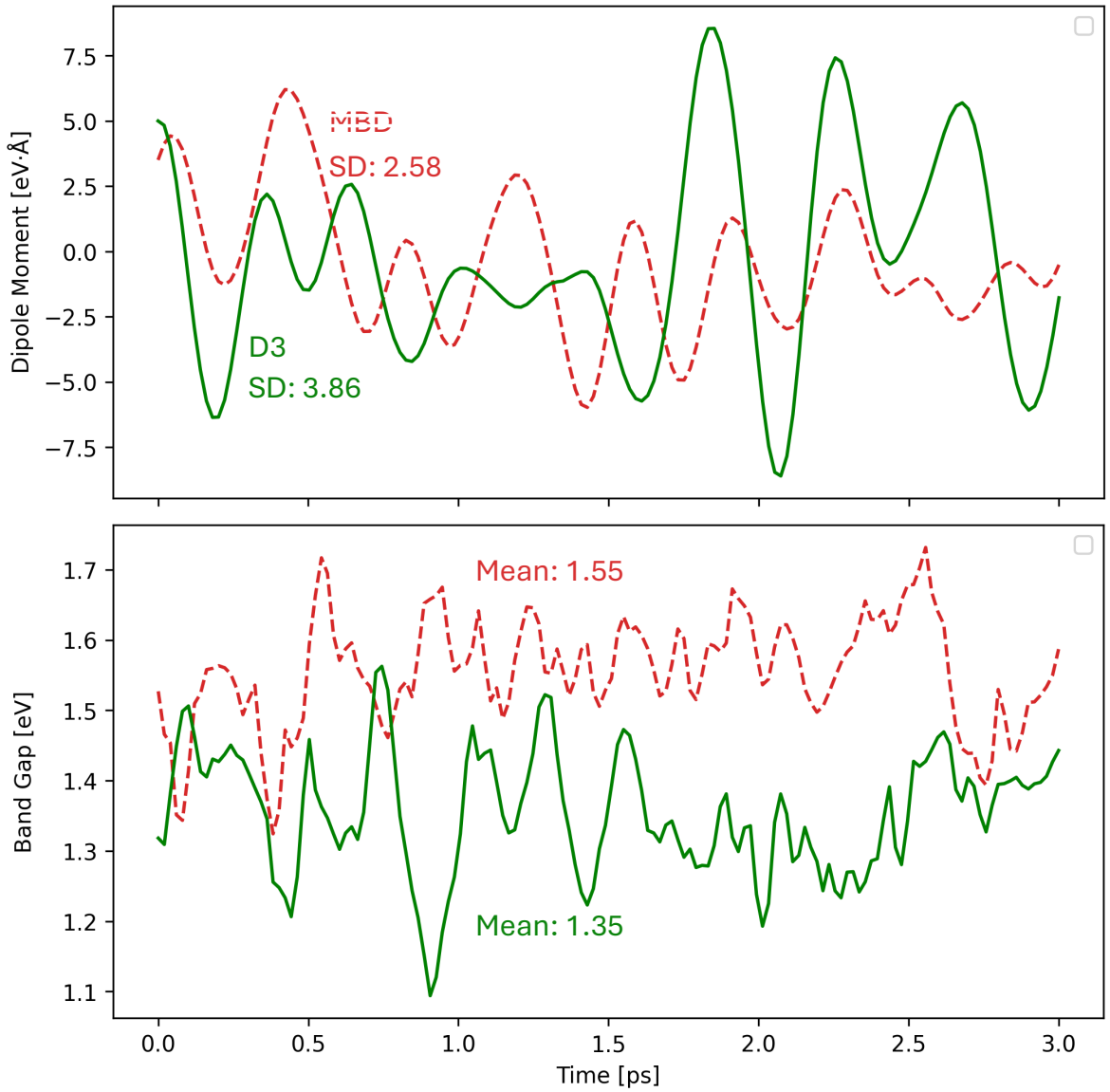


Figure 3.8 Time evolution of (a) the total dipole moment magnitude and (b) electronic band gap over a 3 ps window sampled from a 5 ns MD trajectory. Dynamics were generated with MACE models trained on PBE+D3 (green, solid) and PBE+MBD-NL (red, solid). Summary statistics: SD for dipole moment shown in eV·Åand mean band gap value shown in eV are computed over the 3 ps segment.

not fully captured by static, 0 K DFT calculations. To sample the thermal fluctuations, we trained a MLFF based on the MACE architecture and used it to perform accurate and efficient MDs. In addition to reproducing reference DFT forces with high accuracy, we found that the treatment of vdW interactions in the underlying DFT significantly affected both the structural dynamics and the electronic structure of the slab.

We compared MLFF trained on DFT data with two treatments of vdW interactions, namely MBD-NL and D3. Both models showed similar RMSE around 0.2 kcal/mol·Å and error heterogeneity across atomic species and slab layers, which allows to perform subsequent MD simulations to compare dispersion treatments.

The two models produced markedly different structural dynamics at ambient conditions. MBD-NL stabilized larger and longer-lived octahedral tilts, whereas D3 favored near-cubic structures with suppressed tilts and short-range correlations. In temperature ramps, MBD-NL preserved orthorhombic-like tilt order to higher temperatures and reproduced a gradual orthorhombic-to-cubic evolution consistent with reported $CsPbI_3$ phase diagrams, while D3 predicted an earlier loss of tilt order than expected from experiment. These dynamical differences carried over to the electronic properties: MBD-NL yielded wider thermally averaged band gaps and smaller dipole fluctuations, while both models remained centrosymmetric on average

The trained MLFF enabled large-scale finite-temperature MD of $CsPbI_3$ slabs beyond the reach of direct DFT. However, the cost of subsequent electronic-structure evaluations constrained the supercell sizes and the number of thermally sampled configurations analyzed, limiting the statistical precision of electronic-property trends. Future work should expand configurational and temporal sampling and include finite-size analysis, using more efficient electronic-structure methods or surrogate models for band gap, dipole, and other electronic structure observables to quantify these effects with tighter uncertainties.

Our results showed that the choice of treatment vdW controlled the finite-temperature slab dynamics and the associated electronic response. This established an MLFF-driven sampling protocol that scaled to realistic surface models and yielded statistically grounded trends. The same protocol can be applied to 2D perovskites and defected slabs to assess band edges, dipole responses, and defect energetics. Future studies should be applicable to systems with tens of thousands of atoms by coupling MLFF sampling with low-cost perovskite-specific electronic-structure workflows on representative snapshots, to quantify dispersion-controlled trends in band edges, interfacial dipoles, and defect energetics across temperature, and to relate these microscopic descriptors to transport metrics such as carrier mobilities that influence device performance.

Chapter 4

Crash Testing Machine Learning Force Fields for Materials and Interfaces

Chapter 3 demonstrated the strengths of MLFF for modeling complex halide-perovskite interfaces at length and time scales inaccessible to traditional methods — with empirical FFs limited by accuracy and DFT by prohibitive computational cost. It also identified key issues that require attention in such simulations, particularly the reliability of the reference data used for training. This chapter presents an extended independent, systematic evaluation of several state-of-the-art MLFF models, outlined in Section 2.3, for materials and interfaces. Our analysis is anchored in the "Crash Testing Machine Learning Force Fields" (TEA) Challenge 2023, which the author co-organized and co-authored.

The participating models in the TEA Challenge 2023 were trained by their developers using the provided reference datasets, with limited knowledge on data-generation protocols to prevent unilateral extension of the given datasets. Developers optimized their own trade-offs between model size, accuracy, and computational cost, emulating real-world deployment decisions. Once training concluded, the models could no longer be altered, nor could developers influence the outcomes of the analysis. This simulates realistic application conditions for MLFFs, where the ground truth is often unknown, highlighting potential issues practitioners might encounter. All final MLFF models were subjected to identical MD simulations on the same High-Performance Computing platform, under fixed integrator, thermostat, and timestep settings.

The systems selected for these benchmarks — designed to be both tractable and challenging

Method	Metric	E	F	F(H)	F(C)	F(N)	F(Mol)
MACE	MAE	**0.24**	**0.05**	**0.06**	**0.04**	**0.14**	**0.12**
	RMSE	0.31	0.08	0.09	0.07	0.20	0.16
	MAXE	1	4	2	4	3	1
SO3	MAE	**0.41**	**0.56**	**0.28**	**0.58**	**0.53**	**0.58**
	RMSE	0.52	0.93	0.40	0.95	0.75	0.77
	MAXE	3	44	9	44	13	5
sGDML	MAE	**0.46**	**0.23**	**0.18**	**0.22**	**0.45**	**0.16**
	RMSE	0.58	0.37	0.26	0.37	0.70	0.24
	MAXE	3	12	3	12	12	3
SOAP/GAP	MAE	1.31	**0.56**	**0.44**	**0.57**	**0.93**	1.57
	RMSE	1.59	0.76	0.63	0.75	1.23	2.63
	MAXE	5	69	25	69	21	20
FCHL19*	MAE	366	1.44	1.29	1.43	2.21	4.03
	RMSE	366	1.91	1.70	1.89	2.92	6.12
	MAXE	380	190	134	190	90	41

Table 4.1 MAE, RMSE, and MAXE, in that order, for energy and forces are reported in units of kcal/mol and kcal/(mol·Å), respectively, of the models for 1,8-naphthyridine on graphene system. Values in bold represent MAE below 1 kcal/mol and 1 kcal/(mol·Å) for energies and forces, respectively.

— span molecule-surface complexes and extended solids, ensuring broad applicability and minimizing bias toward any specific architecture. The primary objective was to assess both the theoretical fidelity of modern MLFF frameworks (from kernel methods such as sGDML and SOAP/GAP to equivariant MPNNs like MACE and SO3krates) on rigorous test sets, and their dynamical robustness when deployed in classical MD under ambient conditions.

4.1 Assessing Models Accuracy via Test Set Analysis

Tables 4.1 and 4.2 present a detailed error analysis for the 1,8-naphthyridine on graphene and $MAPbI_3$ systems, respectively, listing MAE, RMSE, and MAXE values for both energies and forces, with MAEs below 1 kcal/mol or 1 kcal/(mol·Å) highlighted in **bold**.

For both systems, the MACE and SO3krates models demonstrate superior performance in terms of MAE and RMSE for both energies and forces compared to the kernel-based models: sGDML, SOAP/GAP, and FCHL19*. Traditionally, the benchmark for MLFF accuracy in quantum chemistry has been 1 kcal/mol MAE for energy and 1 kcal/(mol·Å) MAE for forces. *Equivariant architectures reduce typical MAEs by a factor of 2 to 5*, depending on the number of parameters within the NN. This reduction represents a substantial leap in predictive accuracy, establishing *new standards for force predictions in the range of 0.2 – 0.5 kcal/(mol·Å).*

Despite the advancements in MLFF architectures, *MAXE remain substantial even for the most recent MLFF models.* While the MACE and SO3krates models achieve low MAE and RMSE values, their MAXE errors are still significant, particularly in the case of SO3krates for the

Method	Metric	E	F	F(H)	F(C)	F(N)	F(Pb)	F(I)	F(MA)
MACE	MAE	**0.50**	**0.17**	**0.12**	**0.18**	**0.21**	**0.28**	**0.20**	**0.2**
	RMSE	0.65	0.23	0.16	0.24	0.27	0.36	0.26	0.3
	MAXE	2	5	5	4	3	3	4	4
SO3	MAE	**0.75**	**0.20**	**0.13**	**0.19**	**0.23**	**0.41**	**0.26**	**0.2**
	RMSE	0.94	0.28	0.17	0.25	0.29	0.52	0.34	0.3
	MAXE	3	4	2	4	2	3	3	2
sGDML	MAE	–	–	–	–	–	–	–	–
	RMSE	–	–	–	–	–	–	–	–
	MAXE	–	–	–	–	–	–	–	–
SOAP/GAP	MAE	1.66	**0.43**	**0.29**	**0.50**	**0.54**	**0.83**	**0.50**	**0.6**
	RMSE	2.09	0.59	0.38	0.64	0.68	1.05	0.65	0.7
	MAXE	7	7	7	7	5	7	5	5
FCHL19*	MAE	134	2.81	2.75	2.60	2.95	3.40	2.78	6.2
	RMSE	138	3.62	3.55	3.42	3.80	4.30	3.53	7.9
	MAXE	240	68	68	68	62	34	34	53

Table 4.2 MAE, RMSE, and MAXE for energy and forces are reported in units of kcal/mol and kcal/(mol·Å), respectively, of the models for the $MAPbI_3$ system. Values in bold represent MAE below 1 kcal/mol and 1 kcal/(mol·Å) for energies and forces, respectively.

1,8-naphthyridine/graphene system, where the maximum force error reaches 44 kcal/(mol·Å). This discrepancy highlights that even models with high average accuracy can produce outliers that may lead to unphysical behavior in simulations.

Particular attention should be paid to the total forces acting on the 1,8-Naphthyridine and the Methylammonium molecules in both systems. These forces arise from non-covalent interactions between the molecular and periodic components of the system, which are naturally weaker in magnitude compared to atom-atom interactions dominated by covalent bonding. Despite this, the MAE and RMSE for molecular forces are similar in magnitude to those for atomistic forces across all considered ML models. This is likely due to the fact that all loss functions contain total forces rather than non-covalent forces, and results in *a higher relative error for non-covalent interactions*, highlighting a critical area for enhancing ML architectures.

Advantages of equivariant MPNNs are more evident for the 1,8-naphthyridine molecule on graphene substrate system then for the $MAPbI_3$ system. In the former case, the MACE model achieves MAE and RMSE values roughly half those of SO3krates and five to ten times smaller than those of kernel-based methods. By contrast, for $MAPbI_3$, MACE's advantage over SO3krates shrinks to about 30 % in energy prediction and 15 % in force prediction. Likewise, the SOAP/GAP model exhibits MAE and RMSE only twice as large as those of the NN. Although the gap in energy-prediction accuracy remains larger than that for forces, it is markedly reduced for all models — except FCHL19* — when moving from the 1,8-naphthyridine/graphene system to $MAPbI_3$.

4.1.1 Interface Benchmark: 1,8-naphthyridine on Graphene

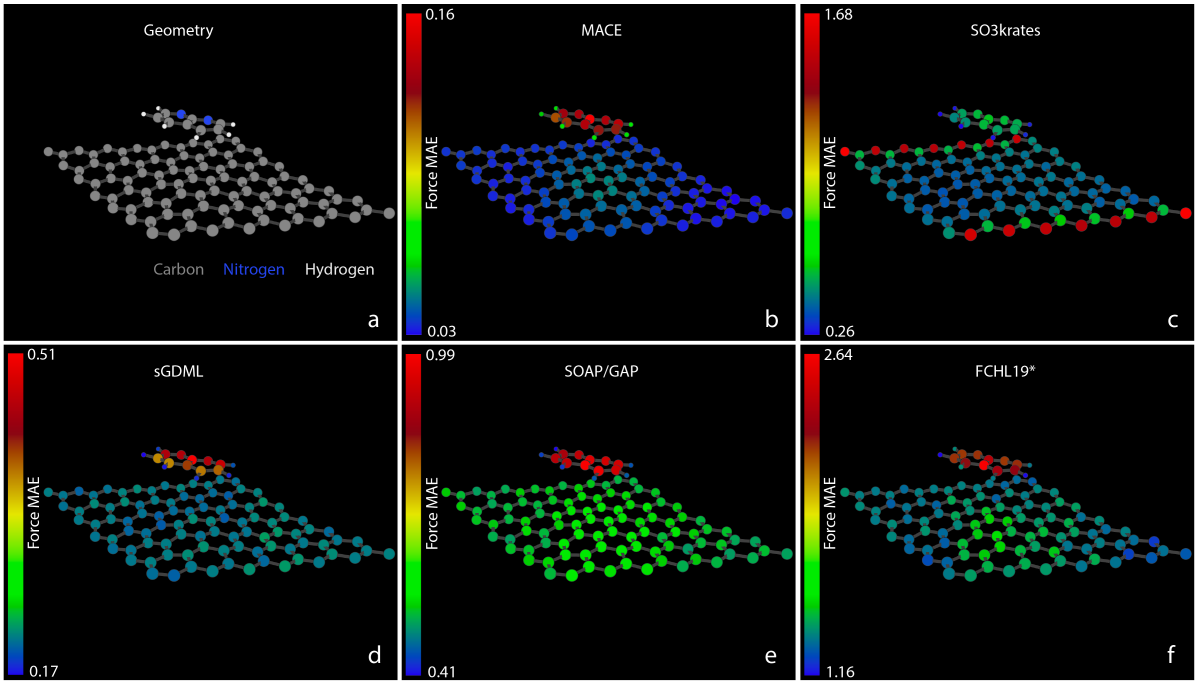


Figure 4.1 Atomic Force MAEs for $C_8H_6N_2/C_{98}$. Figure (a) shows a snapshot of the system geometry and atom types: carbons – grey, nitrogens – blue, and hydrogens – white. Figs. (b) through (f) display the MAEs for forces, measured in kcal/(mol·Å), acting on individual atoms within the $C_8H_6N_2/C_{98}$ system. The MAEs correspond to the MLFFs predictions on the test set and are represented in different colors according to the colour bars shown with the corresponding scaling numbers, which are different for different MLFFs, with scales ranging from 0.16 for MACE to 2.64 kcal/(mol·Å) for FCHL19*.

The analyses presented above follow the typical approach based on aggregated accuracy metrics. While that is a well-established method reliable for small and relatively simple systems in terms of chemical composition, it becomes insufficient as system size and composition diversity increase. In order to address this gap, the FFAST software[186] is used to provide detailed performance quantifiers on the atomistic level.

Figure 4.1 shows the MAEs for atomistic forces for the 1,8-naphthyridine molecule adsorbed on a graphene on the test set for all five MLFFs. In Figure 4.1a, the system geometry is shown, with the 1,8-naphthyridine molecule in the center and the graphene sheet surrounding it. The MAEs for forces acting on individual atoms are presented in Figs. 4.1b-f, with each figure corresponding to a different MLFF. The color bars indicate the range of MAE values for each model, with MACE showing the lowest MAE of 0.16 kcal/(mol·Å) and FCHL19* exhibiting the highest at 2.64 kcal/(mol·Å). While there is a vast difference between the absolute accuracy of the best (MACE) and the worst (FCHL19*) models, all MLFFs, except SO3krates, exhibit similar patterns of relative force accuracy between different atoms. The H atoms and C atoms in graphene are treated with relative ease, while C and N atoms in the 1,8-naphthyridine molecule show significantly larger force errors. The variation of atomic

force error is particularly pronounced for the MACE model, where the MAEs for C atoms in graphene are five times smaller than those for the C atoms in the 1,8-naphthyridine molecule. Interestingly, the exceptionally high accuracy of force reconstruction within the graphene layer, with an MAE of approximately 0.03 kcal/(mol·Å), results in a noticeable 'shadow' of the molecule on the graphene surface in Figure 4.1b. This 'shadow' represents the region most affected by molecule-surface interactions where the MACE model exhibits lower accuracy compared to regions away from the molecule. This observation aligns with the aforementioned challenging nature of non-covalent interactions. This effect is not seen for other models whose errors on the carbon atoms in the graphene sheet are much higher thus masking the above effect.

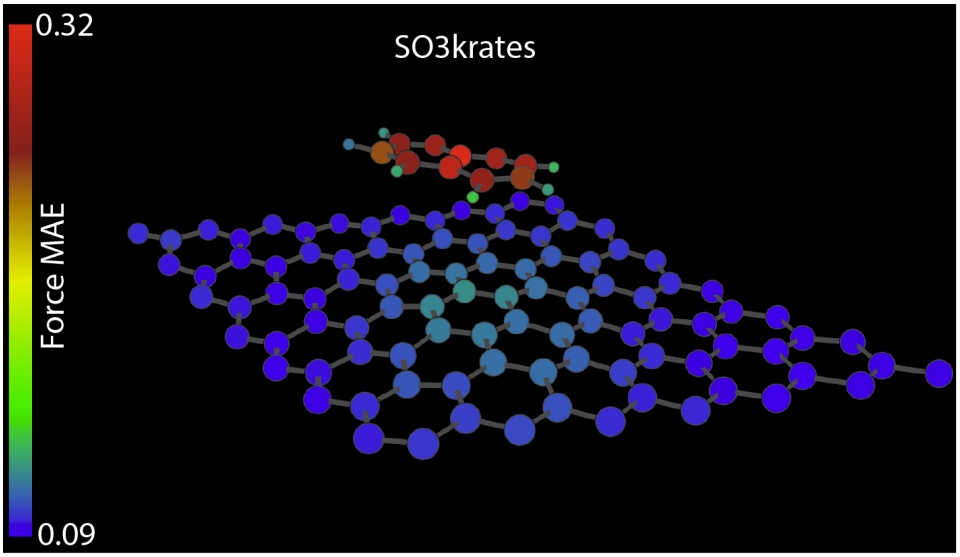


Figure 4.2 Atomic Force MAEs for $C_8H_6N_2/C_{98}$. The MAEs in kcal/(mol·Å) correspond to the retrained SO3krates model predictions on the test set and are represented in different colors according to the colour bar shown with the corresponding scaling numbers.

The significant errors for C atoms on the two borders of the graphene cell produced by the SO3krates model are unexpected and due to precision design choices employed in the JAX ecosystems for certain GPU hardware (see explanation below). It further does not affect the faithfulness of simulation and derived observables. It is important to note that only the *visualisation of atomistic errors allows for identifying subtle but essential prediction problems.* This underscores the necessity of employing software packages such as FFAST, which are designed for comprehensive MLFF performance analysis, to ensure the reliability of the predicted forces and the results of consequent simulations. The cause of the errors for SO3krates is due to the JAX default of running in tensorflow32 precision on A100 and H100 GPU. This problem can be alleviated by explicitly disabling tensorflow32 precision (which performs some operations in float16) and instead running training and evaluation in standard float32 precision. This not only alleviates the boundary effects but also gives significantly

lower errors, ranging from 0.09 to 0.32 kcal/(mol·Å) as shown in Figure 4.2

4.1.2 Solid Benchmark: $MAPbI_3$ Perovskite

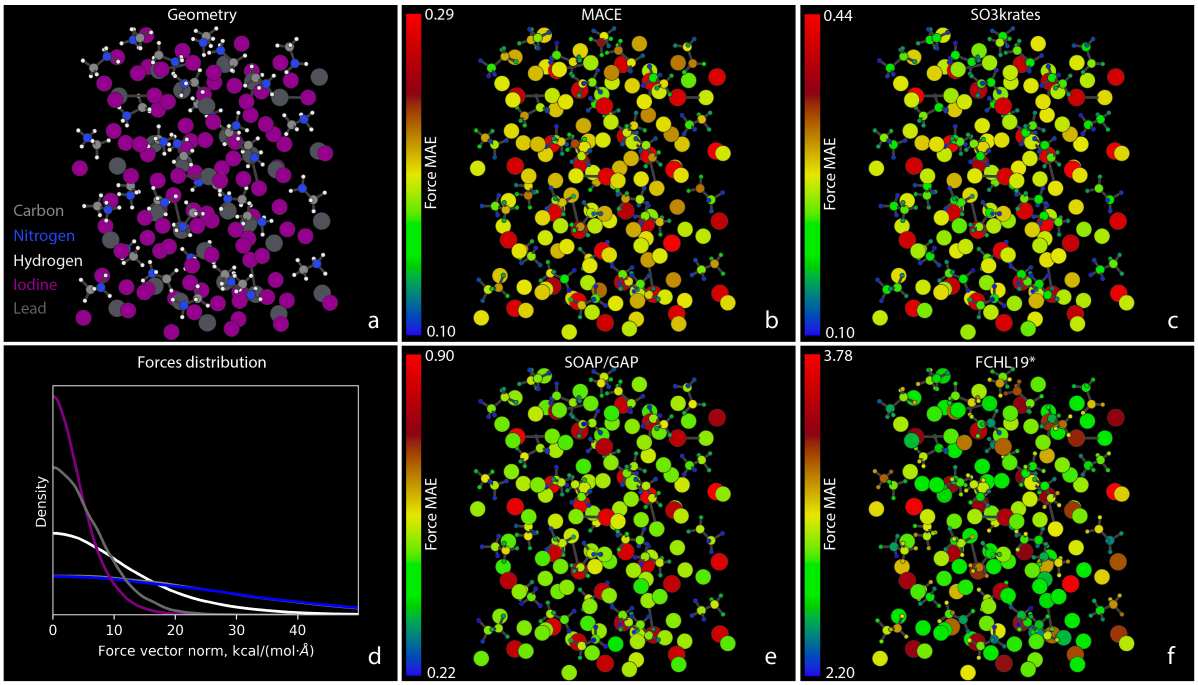


Figure 4.3 Atomic Force MAEs for $MAPbI_3$. Panel **(a)** shows a snapshot of the system geometry and atom types: Carbons – grey, Nitrogens – blue, Hydrogens – white, Iodines – purple, and Leads – dim grey. Panels **(b)**, **(c)**, **(e)**, and **(f)** display the MAEs for forces, measured in kcal/(mol·Å), acting on individual atoms within the $MAPbI_3$ system. The MAEs correspond to the MLFFs predictions on the test set and are represented in different colors according to the color bars shown with the corresponding scaling numbers, which are different for different MLFFs. Figure **(d)** illustrates the distribution of the norms of atomic force vectors for different chemical elements. The carbon curve lies under the nitrogen one, as expected due to the 3rd Newton's law.

The observations made thus far are also consistent with the results for $MAPbI_3$, as presented in Figure 4.3 for the atomic force MAEs for $MAPbI_3$. Similar force prediction heterogeneity patterns are observed across all MLFFs. Figure 4.3 b, c, and e appear nearly identical when disregarding the color bar values that indicate the vast difference between absolute errors. Surprisingly, the most problematic atoms for all MLFFs are the Pb atoms. The *largest MAEs* for these atoms *anti-correlate with the strength of the forces* they experience. This can be seen in Figure 4.3d, which illustrates the distribution of the magnitudes of atomic forces for different chemical elements in the system. The forces on Pb atoms, shown in dim grey, are the second smallest, only slightly larger than those on I atoms. Meanwhile, the MAEs for force reconstruction on I, C, N, and H atoms throughout the system are more than twice as small as those for Pb atoms. This clearly indicates the need for improvements in either MLFF architectures, loss functions, or training datasets. As will be showen in MD analysis of

this systems, the large MAEs for Pb atoms directly impact the stability of MD simulations.

Finally, for $MAPbI_3$, only a minor difference in accuracy for the forces between MACE and SO3krates were observed, although the energy error of MACE is 30% lower. This suggests that, beyond a certain level of system complexity, integrating even more principles of chemistry and physics is necessary for more accurate and reliable modeling. Therefore, despite the tremendous progress in MLFF architectures over the last decade, there remains room for improvement and further developments.

In summary, novel equivariant MLFF architectures like MACE and SO3krates significantly enhance the faithful reconstruction of potential energy surfaces for molecular interfaces and periodic systems. However, a major challenge remains in eliminating rare but substantial errors in force predictions. The observed maximum force errors are comparable to those of previous MLFF generations, such as sGDML. Another area for improvement is the description of non-covalent interactions, which are often obscured by strong covalent bonding. Additionally, force reconstruction demonstrates significant heterogeneity between different atoms of a given system that is consistent across all tested models, including both kernel-based and neural network approaches. The consistent error patterns are primarily due to the L2 nature of the loss function applied to the forces vector across all models. However, the observed heterogeneity, which does not always align with the magnitude of the force acting on an atom, suggests a direction for further development of MLFFs. The analysis emphasizes the need to move beyond aggregate accuracy measurements, such as system MAE or RMSE, to uncover potential pitfalls of MLFFs.

4.2 Assessing Models Fidelity via Molecular Dynamic Simulations

One of the critical requirements for any MLFF is the ability to run efficient and reliable simulations. Beyond achieving low MAE, RMSE, or MAXE on test sets, an MLFF must also ensure stable and computationally efficient MD trajectories. To assess this, 12 independent MD simulations were conducted, each for one million steps, using all MLFF architectures participating in the TEA Challenge 2023.

Table 4.3 provides a summary of the stability test results. Stability was assessed using a broken bond criterion, terminating the simulation if any covalent chemical bond in the system exceeded a length of 2 Å. Under the given simulation conditions, the test systems were not expected to undergo proton transport or other processes that would alter their bonding patterns.

Simulation		Systems	
Model	Temperature, K	1,8-naphthyridine/Graphene	$MAPbI_3$
MACE	300	12/0	12/0
	500	12/0	0/12
SO3krates	300	12/0	12/0
	500	6/6	10/2
sGDML	300	12/0	–
	500	12/0	–
SOAP/GAP	300	2/10	0/12
FCHL19*	300	1/11	0/12

Table 4.3 Molecular dynamics stability in A/B ratio. Here, A is the number of completed trajectories (1 ns), and B is the number of failed trajectories (broken chemical bonds).

The MACE architecture demonstrated the highest stability in these tests, successfully completing most MD trajectories. Failure occurred only in the $MAPbI_3$ system at 500 K, where a single 1 ns long MD trajectory could not be provided. The second most stable architecture, SO3krates, performed well in this scenario, successfully completing 10 out of 12 MD runs. However, SO3krates faced stability issues for the 1,8-naphthyridine/graphene system when the temperature increased to 500 K and above. The causes of the difference in stability between the equivariant MPNN architectures remain unclear and require further studies.

Among the kernel-based potentials, sGDML demonstrated the greatest dynamic stability for the 1,8-naphthyridine/graphene system, retaining robustness up to 500 K. In contrast, the SOAP/GAP and FCHL19* potentials exhibited pronounced instability, with several trajectories terminating after only a few thousand timesteps. Such limited stability is expected for these kernels in the absence of active-learning training protocols[256, 242, 257, 258]. Nonetheless, when confined to well-sampled, near-equilibrium regions of the potential-energy surface, SOAP/GAP and FCHL19* simulations can readily achieve durations on the order of tens to hundreds of picoseconds.

Overall, the results highlight the varying degrees of stability among different MLFF architectures and emphasize the need for continued improvements to ensure reliable MD simulations, cf. also [259, 260]. Although low benchmark errors do not invariably guarantee robust simulations – as exemplified by SO3krates's failure under conditions in which MACE remains stable, and conversely by MACE's breakdown in regimes where SO3krates remains robust – models with the best aggregate accuracy (MACE and SO3krates) nonetheless occupy the highest-stability regime.

Building on this, incorporating uncertainty estimations into the workflow will represent a crucial step in enhancing MLFF modeling reliability. For kernel-based models, Bayesian methods provide a relatively straightforward way to achieve this. However, NN models

such as MACE and SO3krates currently lack intrinsic mechanisms for estimating prediction uncertainty. Meanwhile, uncertainty estimation in NNs has been a widely researched topic in other fields [261]. Adapting and integrating the developed methodologies into MLFF codes should become a key research focus, especially as the complexity and scale of target systems grow, making reference calculations computationally prohibitive.

Table 4.4 summarizes the average computational time required for each ML model to produce 1 ns of MD dynamics (1 million steps) on a single NVIDIA A100-40 GPU. The reported times are averaged across all MD trajectories within each challenge. All run conditions and scripts were identical for all models except for the ASE calculator and the specific ML architecture provided by the participant [262]. It is important to note that the SOAP/GAP models cannot be executed on a GPU. Instead, these models were run on Central Processing Unit (CPU) nodes consisting of 2 AMD Rome CPUs, each with 64 cores at 2.6 GHz, for a total of 128 cores and 512 GB of RAM. This hardware configuration was used to ensure that SOAP/GAP models could complete the MD simulations within a reasonable timeframe.

Simulation	Systems	
Model	1,8-naphthyridine/Graphene	$MAPbI_3$
MACE	45.4	36.1
SO3krates	4.2	6.7
sGDML	224.8	–
SOAP/GAP	64.9	606.2
FCHL19*	9.9	29.2

Table 4.4 Average simulation time in hours per 1 ns of molecular dynamics. The red color indicates models that failed to generate a single 1 ns MD trajectory.

SO3krates delivers the highest production throughput, outperforming its nearest neural-network rival, MACE, by factors of 6–25 depending on the system size. FCHL19* ranks second in speed but incurs the greatest instability in MD trajectories. Under its current hyperparameter settings, MACE occupies third place. By contrast, the explicit enforcement of permutational symmetry renders sGDML the least efficient of all TEA 2023 entrants. For 1,8-naphthyridine/graphene system, MACE achieves only a 1.5–2.5 times speedup over SOAP/GAP—even though SOAP/GAP runs solely on CPU—highlighting that raw accelerator performance does not directly translate into end-to-end throughput. Although the TEA 2023 systems are modest in size compared with those used in large-scale MD studies, the non-linear scaling of simulation time with system size underscores the need for further algorithmic and code-level optimization. As system size grows, workloads may shift from memory-bound to compute-bound regimes, introducing additional performance complexities.

4.2.1 1,8-naphthyridine on Graphene

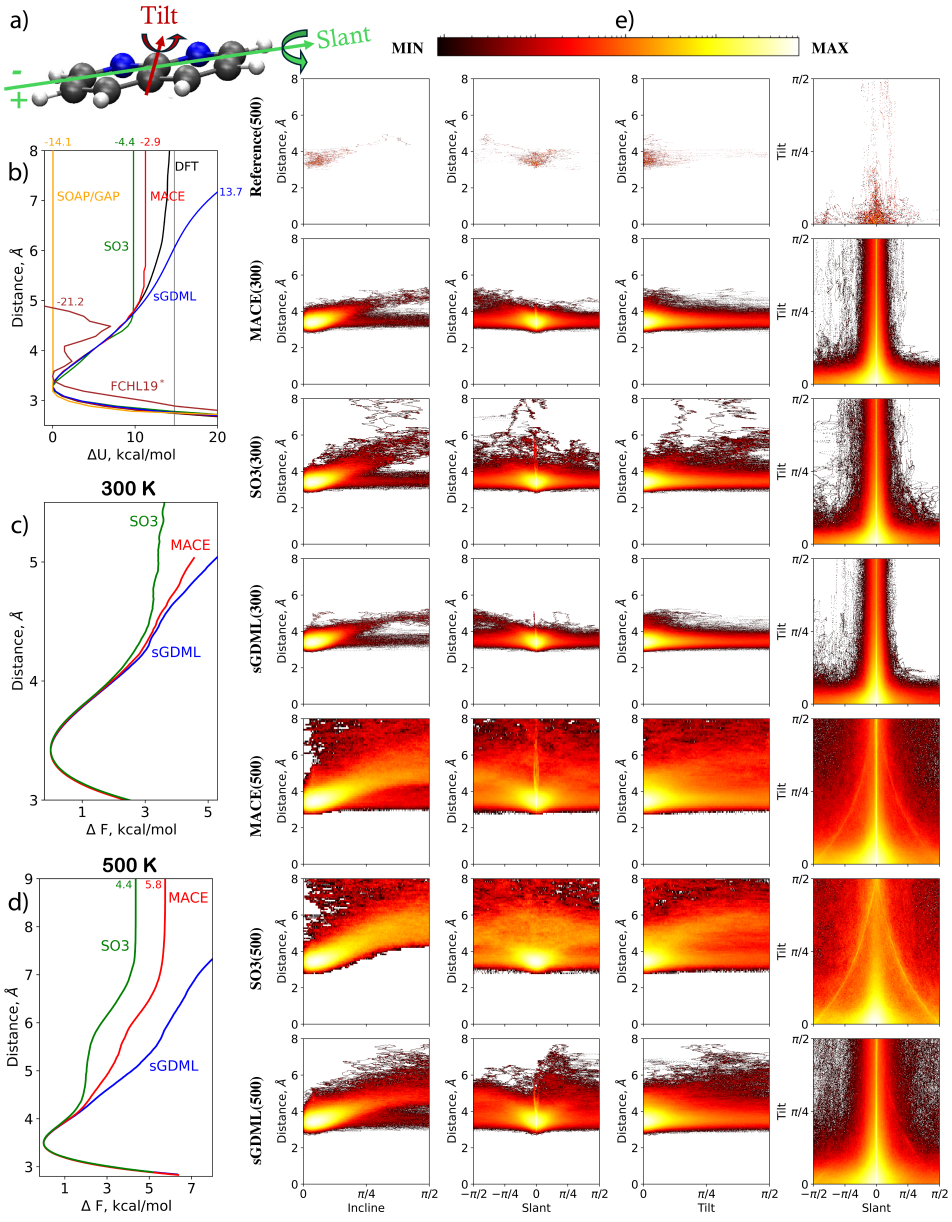


Figure 4.4 Analysis of MDs for 1,8-naphthyridine molecule adsorbed on graphene. **a)** Schematic representation of the molecule and definitions of tilt and slant angles. **b)** Interaction energy profile (ΔU) as a function of the distance between the centers of mass of the molecule and graphene. **c)** and **d)** Free energy change profile from MD simulations at 300 and 500 K, respectively. **e)** Distributions of incline, tilt, and slant angles at different molecule-to-surface distances, and the tilt-slant angle distribution in MD simulations at 300 and 500 K using MACE, SO3krates, and sGDML models compared to the reference dataset.

While modern MLFFs can accurately describe an isolated 1,8-naphthyridine molecule or a pristine graphene sheet, their combination gives rise to the subtleties of molecule–substrate interactions that govern many key properties. Capturing these interactions requires the MLFF to simultaneously solve the comparatively weak adsorption forces and the strong covalent bonds intrinsic to each subsystem. Interfacial behavior is assessed using four complementary

metrics. First, the distribution of the separation distance, D, is examined, where D denotes the distance between the averaged graphene plane (AGP) and the molecule's center of mass. Second, the incline angle, θ, is analyzed, with θ defined as the angle between the averaged molecular plane (excluding H atoms) and the AGP. The third and the fourth metrics are the *tilt* and *slant* angles, respectively. Plots of all aspects of the interaction are shown in Figure 4.4. Note that the molecule is symmetrical for the tilt rotation (the tilt angle is defined in the range $[0, \pi/2]$), while it is asymmetric for the slant rotation (the slant angle is defined in the range $[-\pi/2, \pi/2]$). Figure 4.4 illustrates the results for MLFF MD simulations conducted at 300 K. Notably, all ML models encounter angles and distances that extend significantly beyond those observed in shorter reference MD simulations at 500 K, indicating that the MLFF models operate in an extrapolation regime.

At 300 K, sGDML, SO3krates, and MACE produced similar patterns for all four examined dependencies. This includes subtle features found in agreement between models observed in the plot of incline angle probability as a function of surface-to-molecule distance. A notable split for angles close to $\pi/2$ (molecular orientation perpendicular to the graphene surface) is visible for sGDML, SO3krates, and MACE. The split arises from the asymmetry in the molecule's slant rotation, where the side containing N atoms can approach the surface closer than the sides with H atoms attached to C atoms. The molecular asymmetry also explains the trend to negative slant values in the distance versus slant angle plots. Furthermore, it is the reason for the asymmetric large fluctuations of the slant angle in the tilt versus slant angle plots - a negative slant angle indicates that the side of the molecule with N atoms is closer to the surface. The main disagreement between the models is a tendency for the molecule to desorb from the surface when SO3krates models is employed. This tendency is absent in the MD simulations produced with the MACE and sGDML MLFFs.

At 500 K, the predictions between the MLFFs begin to diverge. The MD simulations obtained with MACE and SO3krates models differ drastically from those produced by sGDML. Both the MACE and SO3krates models predict the molecule's desorption from the surface, while the sGDML model keeps the molecule within an 8 Å distance. This suggests that the PES profile for the molecule in the direction perpendicular to the graphene surface predicted by sGDML differs from that predicted by the NNs. To verify this conjecture, the molecule-surface interaction energy is computed as a function of the distance, see Figure 4.4b. The computation starts from the relaxed structure (obtained using the reference DFT setup), and confirmations are produced, for which the z-coordinates of the atoms belonging to the molecule are moved further from or closer to the surface in 0.1 Å steps. This provides us with the $\Delta U(D)$ dependence computed at the reference level of accuracy. The same calculations

are repeated using the MACE, SO3krates, sGDML, SOAP/GAP, and FCHL19* models. Notably, some MLFFs failed to provide reasonable energy predictions at large molecule-to-surface distances. Therefore, the minimum energy within the distance range of 3 to 4 Å is chosen for each method as the zero energy level. The thin black line in Figure 4.4b corresponds to the DFT energy at an infinite molecule-to-surface separation.

In complete agreement with the MD simulations, the *SO3krates and MACE models significantly underestimate the potential energy minima compared to the DFT reference.* Figure 4.4b shows the difference between the DFT and MLFFs energy predictions at an 8 Å distance as small numbers on the top. Notably, this distance is beyond the cutoff for both NNs, leading them to predict zero interaction between the molecule and the surface. The MACE model underestimates the molecule-surface interaction energy at 8 Å by 2.9 kcal/mol and SO3krates by 4.4 kcal/mol. For the adsorption minimum, when aligning the DFT and MLFFs energies at infinite distance, the MACE and SO3krates models underestimate the value by 3.5 and 4.9 kcal/mol, respectively.

Interestingly, *this underestimation* of the molecule-surface interaction minimum *aligns with the high accuracy in predicting the force acting on the molecule as a whole in the test set.* The reference molecule-to-surface distances are limited to an interval of 3 to 5 Å. Within this range, the curvature of the $\Delta U(D)$ function obtained from DFT calculations and those predicted by MACE and SO3krates models agree. Moreover, these distances fall within the cutoff radii for MACE and SO3krates, which are 6 Å and 5 Å, respectively. This results in minor mean absolute errors (MAEs) in forces, 0.12 kcal/(mol·Å) for MACE and 0.58 kcal/(mol·Å) for SO3krates, despite a significant underestimation of the barrier.

Figs. 4.4c and 4.4d illustrate the differences of free energies of adsorption calculated employing different MLFFs. The free energies derived from MD trajectories via the thermodynamic integration method [124], are significantly smaller than the adsorption energy estimations calculated as the difference between the energy of the DFT-optimized structure and the state, for which the molecule and the surface are at infinite separation. This discrepancy arises due to the substantial rotational freedom of the molecule's plane relative to the surface plane. Specifically, the free energy minima predicted by the MACE and SO3krates models at 500 K are only 5.8 and 4.4 kcal/mol, respectively, whereas the adsorption energies predicted by these models are 11.3 and 9.8 kcal/mol.

The kernel-based models also face significant challenges in reproducing the molecule-surface interaction. The sGDML model over-stabilizes the system due to an artificial barrier at intermediate distances. For instance, at 8 Å, the model overestimates the DFT energy by 13.6 kcal/mol. The artificial barrier only disappears at much larger distances, above 30 Å.

The SOAP/GAP model fails to reproduce any molecule-to-surface attraction, providing a flat interaction profile immediately after the expected physisorption minima around 3.2 Å. The behavior of $\Delta U(D)$ within the FCHL19* MLFF lies somewhere between SOAP/GAP and the NNs. Qualitatively, the model reproduces the repulsion and a part of the attraction regions but with additional peculiar minima and the largest error at 8 Å distance of 21.2 kcal/mol among all MLFFs participating in the TEA Challenge 2023.

A significant point is that adding more training points with a better sampling of molecule-to-surface distances can only help the global sGDML MLFF to improve the description of the molecule-surface interaction. The finite cutoffs intrinsic to all other MLFFs participating in TEA 2023 would prevent them from correctly describing the molecule adsorption process or long-term dynamics unless additional elements targeting long-range interactions are incorporated into these MLFFs. It is emphasized that the above statement pertains specifically to the versions of the MLFF models that participated in the TEA Challenge 2023. The updated version of SO3krates (the pre-trained SO3LR model [143] for bio-molecular simulations includes long-range electrostatic and universal dispersion interactions) includes the option to incorporate long-range interactions. With the appropriate reference data, SO3LR could accurately describe the adsorption process.

Detailed analysis on the broken chemical bonds responsible for the instability of MD simulations across all MLFF architectures is provided in Table 4.5.

Temperature	Bond	MACE	SO3	sGDML	SOAP/GAP	FCHL19*
300K	C-H	0	0	0	10	10
	C-C	0	0	0	0	16
500K	C-H	0	5	0	12	10
	C-C	0	1	0	0	16

Table 4.5 Chemical bonds broken by stretching or atoms fusing in the system with 1,8-naphthyridine molecule adsorbed on graphene. The same MD test of 1M steps is run 12 separate times with each model. The possible bond types that can be broken are: 'C-H', 'C-C', and 'C-N'. If a bond type is not listed, then all models have performed well for that bond, which was not broken in the termination of the MD run. The SOAP/GAP model's results are for the best-performing 'heavy' model for this task.

In summary, molecule–surface interfaces need careful attention when using MLFFs for simulations. Modern MLFFs achieve high accuracy, capturing detailed system behavior with MAEs and RMSEs in the fraction of one kcal/mol and kcal/(mol·Å). However, they require embedding long-range interaction models into MLFF architectures to handle system elements beyond the intrinsic cutoff distances, enhancing simulation accuracy and reliability.

4.2.2 Methylammonium Lead Iodide Perovskite

The periodic structure of methylammonium lead iodide ($MAPbI_3$) features an organic methylammonium (MA) cation at the center of the inorganic eight corner-sharing $[PbI_6]_4$-octahedral. The stability of the simulations was assessed by applying two different thresholds for the organic and inorganic parts of the structure: 2 and 4.3 Å, respectively. The organic threshold was defined using the covalent radius. In contrast, the inorganic threshold was based on experimental results from various scattering techniques for the radial distribution function of the Pb-I ionic bond [263].

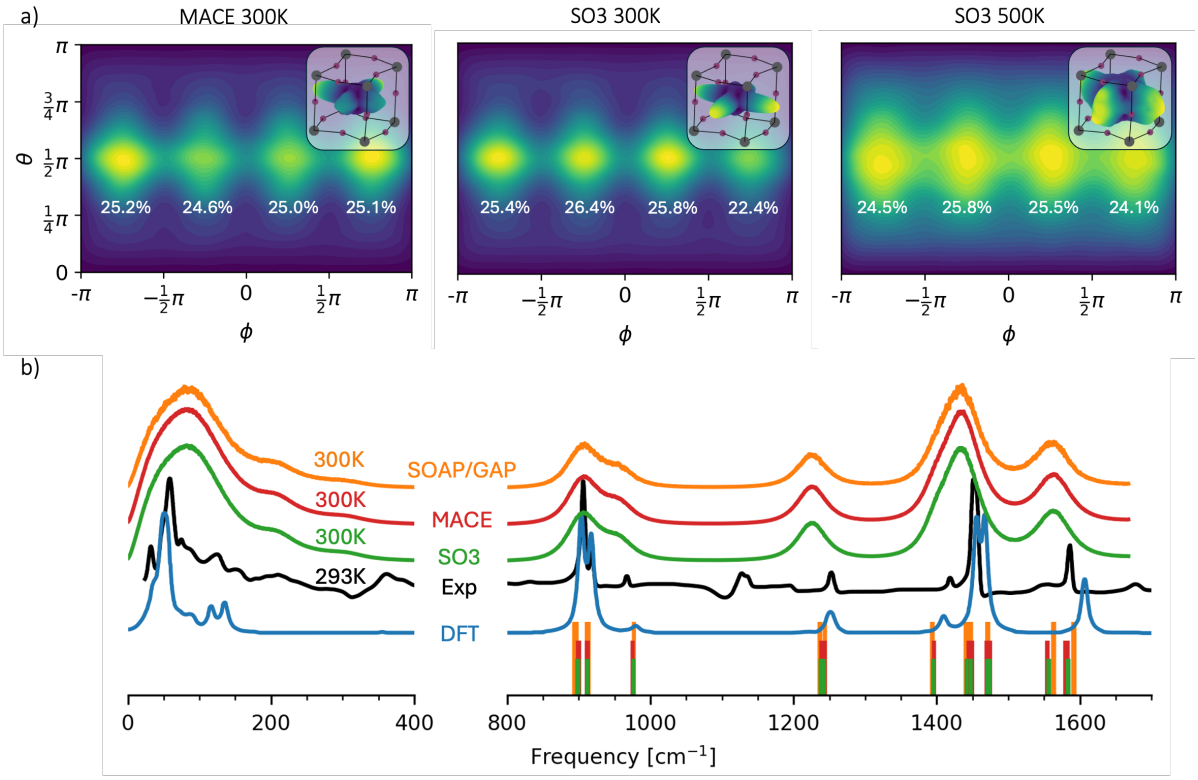


Figure 4.5 Observables extracted from the MD simulations of the orthorhombic phase of $MAPbI_3$. **a)** Distribution of C-N bond orientation in spherical coordinates. The numbers near the maxima indicate the participation ratios of the bond orientation for four different regions of ϕ. 3D spatial representations of the distribution within the PbI framework on a logarithmic scale are shown in the upper right corners. **b)** Spectral analysis of the velocity autocorrelation function (solid lines) obtained from MD simulations using SO3krates (green), MACE (red), and SOAP/GAP (orange) models at 300 K, compared to the experimental infrared spectrum[264] (black) and DFT within the harmonic approximation [264] (blue). The vertical bars indicate phonon frequencies calculated within the harmonic approximation for each MLFF model, spanning a frequency range of 800 cm^{-1} to 1800 cm^{-1}.

MA cations within the PbI framework are characterized by ionic interactions with the rest of the lattice. The distribution of MA-cation C–N bond orientations was computed to assess the ability of the ML models to reproduce these weak interactions. The results in spherical coordinates are shown in Figure 4.5a for the MACE and SO3krates models in cases where MD simulations produced sufficiently long (over 10 ns) trajectories. The polar

angle θ is the angle between the z-axis and the C-N bond, while the azimuthal angle ϕ is the angle between the orthogonal projection of the bond on the x-y plane and the x-axis. The upper right corner of the distribution plot for each set of MDs presents a 3-dimensional representation of the C-N orientation on a log scale for peak visibility. The numbers near the maxima of the distributions indicate the probability of the bond orientation for four regions: $-\pi < \phi < -\frac{1}{2}\pi$, $-\frac{1}{2}\pi < \phi < 0$, $0 < \phi < \frac{1}{2}\pi$, and $1/2\pi < \phi < \pi$. The bond tends to orient into cavities in the PbI framework along the x and y axes for all presented MDs. As expected, the distribution for SO3krates at 500 K is more smeared than that for both NNs at 300 K. The unlikely orientation of the bond along the z-axis relates to the alignment of this axis with the c-axis of the orthorhombic structure of $MAPbI_3$.

The MACE model permitted significant fluctuations in the positions of Pb atoms, which ultimately compromised the structural integrity of the system when these atoms approached the MA molecule. This instability can be attributed to the inadequate sampling of large atomic fluctuations in the reference dataset, which was derived from relatively short *ab initio* MD simulations [265]. Additionally, the training dataset lacked supplementary information necessary for the ML models to capture these large-scale atomic movements accurately. Notably, the escape events involving Pb atoms in our simulations occurred after significantly longer times compared to the reference dynamics. Despite this, the SO3krates model demonstrated a markedly improved stability. In our tests, SO3krates resulted in only 2 out of 12 failed 1 ns long trajectories, compared to 12 out of 12 failures with MACE. However, it is important to note that SO3krates also encountered similar issues twice, underscoring the current limitations in accurately describing interatomic interactions at distances beyond the typical covalent bond range in a sparse training data regime. A comparative analysis of atomic radial distribution functions from MACE and SO3krates MD simulations at 500 K, alongside 1 ps *ab initio* MD results at the same temperature, is provided in SI.

Figure 4.5b presents the spectral analysis of the $MAPbI_3$ system for MACE, SO3krates, and SOAP/GAP models. This analysis yields the vibrational frequencies of the system, derived through the Fourier transform of the velocity autocorrelation function (VAF), which provides insights into vibrational dynamics. The experimental infrared (IR) spectrum and a DFT-simulated IR spectrum (within the harmonic approximation) are also shown for reference. Since IR intensities are governed by changes in dipole moments, while VAF spectra are based on velocity correlations, the resulting peaks from the VAF method are generally broader, with different intensity ratios. Nevertheless, the VAF-based frequency positions can still be compared meaningfully to the IR spectra. Unlike the harmonic approximation assumed in DFT, the VAF approach can recover anharmonic vibrational modes. For instance, a notable

feature at 200 cm^{-1} is observed in both the MD-derived and experimental IR spectra but is absent from the DFT predictions. Overall, *all MLFF models produce nearly identical spectra, showing good agreement with experimental data and DFT* in terms of peak positions.

At the same time, *the experimental peaks* observed at 360 and 1140 cm^{-1} *were not detected by either MLFF*. This discrepancy is attributed to the limitations of the reference DFT calculations. Notably, these peaks are also absent in the DFT spectra presented in Figure 4.5. However, it is important to note that the DFT functionals used in this study differ from those in Ref [264]. Unfortunately, computing the IR spectra for $MAPbI_3$ with the unit cell size used for this TEA Challenge at the level of accuracy PBE+MBD-NL is not feasible.

Additionally, phonon frequencies, depicted as vertical bars on Figure 4.5b, were calculated within the harmonic approximation for each model, covering the frequency range from 800 cm^{-1} to 1800 cm^{-1}. To obtain these frequencies, the $2 \times 2 \times 2$ $MAPbI_3$ system was first optimized separately by each MLFF model using the BFGS algorithm implemented in ASE, with a convergence criterion set to a maximum force component of 0.005 eV/Å [262]. The phonon frequencies were then derived from these optimized structures using the finite difference method for each respective model, employing a 0.01 Å displacement. These frequencies are consistent across the models and align well with the IR spectra, demonstrating the MLFFs' ability to accurately capture the region of the PES near the system's equilibrium state. However, differences between the MACE, SO3krates, and SOAP/GAP models are slightly more pronounced in the harmonic spectra compared to the VAF spectra. This highlights the greater sensitivity of physical properties that depend on specific regions of the PES to the accuracy of the MLFFs, in contrast to more "global" statistical properties like VAFs, which reflect a broader portion of the PES.

As it shown in Table 4.6, the stability of MD simulations of $MAPbI_3$ was primarily compromised by broken covalent bonds within MA cations, although C-N bond breaks were less frequent. Interestingly, the instabilities observed in the MD trajectories generated using the MACE model were triggered by significant fluctuations of the Pb atoms rather than covalent bond breakage. The displacement of Pb atoms closer to the MA cation led to the MA breaking and further destabilization of the system. This behavior was not observed in the SO3krates model, which maintained stability throughout the simulations. In recent publication on MACE training process this problem was addressed by introducing a new training protocol that includes adding to the training dataset carefully constructed dimer interactions.

In summary, kernel-based MLFFs failed to provide stable MD trajectories, with only the SOAP/GAP model successfully generating 2 out of 12 1 ns-long trajectories without loss of structural integrity. In contrast, equivariant MPNNs demonstrated reliable stability and

Temperature	Bond	MACE	SO3	sGDML	SOAP/GAP	FCHL19*
300K	C-H	0	0	-	2	6
	N-H	0	0	-	2	4
	C-N	0	0	-	1	5
	Pb-I	0	0	-	8	0
500K	C-H	0	0	-	8	8
	N-H	0	0	-	3	8
	C-N	0	2	-	0	1
	Pb-I	13	0	-	1	0

Table 4.6 Chemical bonds broken by stretching or atoms fusing in $MAPbI_3$ system. The same MD test of 1M steps is run 12 separate times with each model. The possible bond types that can be broken are: 'C-H', 'N-H', 'C-N', and 'Pb-I'; '-' indicates that no MD was run for the specific model at this temperature. If a bond type is not listed, then all models have performed well for that bond, which was not broken in the termination of the MD run. The SOAP/GAP model's results are for the best-performing 'heavy' model for this task.

efficiency. Spectral analysis showed that MACE, SO3krates, and SOAP/GAP models aligned well with experimental and DFT-derived spectra, despite missing some peaks, likely due to limitations in the reference DFT calculations. Therefore, the main bottleneck for atomistic simulations in this class of systems is obtaining high-quality and representative reference data rather than the MLFF architectures themselves.

4.3 Pitfalls and Best Practices for Reliable MLFFs

This analysis identifies common pitfalls encountered when integrating modern MLFFs into atomistic simulation workflows for bulk materials and surfaces and distills best practices to identify and mitigate them, enabling robust, reproducible simulations.

Among these pitfalls, *version drift* is particularly pervasive: models trained with one release of an MLFF library or its transitive dependencies may no longer deserialize or evaluate correctly after an update. Such changes can manifest as outright crashes, but more insidiously as silent deviations in predicted energies and forces due to altered kernels, activation functions, cutoffs, or default hyperparameters. Tracking these changes is complicated by *incomplete or non-granular changelogs*, which obscure whether a given release introduces a new feature, fixes a numerical issue, or breaks a previously valid interface. Friction also arises from the absence of *unified conventions* across implementations—units (eV vs. kcal/mol; Åvs. nm), file formats for trained weights and datasets, and input/output API semantics often differ, making model exchange and comparative evaluation unnecessarily error-prone.

Practical deployment further suffers from *under-documented build, runtime, and hardware requirements*. Missing details about compiler flags, CUDA/BLAS variants, or supported

CPU/GPU architectures lead to installations that appear successful but deliver suboptimal throughput or nondeterministic results. Because many packages ship without portable, built-in benchmarks, it is difficult to verify that a fresh installation reproduces published accuracy and speed on a new system. Finally, several MLFF interfaces do not surface *model-specific energy or force offsets* introduced during training (e.g., baseline shifts, per-species reference energies, or regularization terms). When these offsets are applied implicitly—or are omitted in downstream codes—otherwise accurate models produce systematic discrepancies that accumulate in thermodynamic estimates or destabilize long trajectories.

Robust use of modern MLFFs should start with *cross-validation* rather than reliance on a single architecture. Comparing predictions from independently trained models provides a sanity check when reference data are sparse or absent.

Reproducibility hinges on rigorous record-keeping and openness. Persist complete *training metadata*—hyperparameters, software versions, and the exact composition of training/validation sets—and, where possible, embed this information in the model artifact to enable exact retraining. Developers and users alike should prioritize *transparency* by publishing clear changelogs, documenting recommended preprocessing beyond intrinsic MLFF routines, and exposing any energy or force offsets. Clear *documentation and examples*, including notes on software/hardware compatibility and common failure modes, further reduce integration risk.

Model reliability is tightly coupled to *training dataset quality*. Training sets must be balanced across relevant states and conditions; over-representing a narrow subset can lower global error but harm transferability. Integrating *active learning* or similar iterative curation into the workflow helps identify blind spots efficiently. It is also beneficial to include configurations at very short and very long interatomic separations to enforce correct *asymptotic behavior*, even if such extremes are rare in the target application.

Equally important is a *detailed performance analysis*: aggregate metrics such as overall MAE or RMSE can conceal failure modes in realistic systems. Performance should be resolved per atom, per chemical element, and per local environment, and the error distributions should be inspected directly. Reducing *error heterogeneity*—even at fixed global accuracy—typically yields more stable dynamics and more trustworthy thermodynamic estimates.

For hierarchical systems—atoms forming molecules, molecules assembling into clusters or surfaces—training objectives should reflect the operative scales. Incorporating *multiscale forces* (and, where appropriate, energies) into the loss with sensible weights can stabilize collective behavior; minimizing only atom-wise errors may be insufficient for mesoscale fidelity. At the same time, accuracy targets should be matched to the intended use. Choosing *appropriate accuracy levels* avoids unnecessary computational cost: models that differ in

per-atom MAE (e.g., 0.5 vs. 0.1 kcal/(mol·Å)) can yield indistinguishable observables in many molecular-dynamics settings, while the higher-accuracy model is costlier to train and deploy.

Finally, initiatives such as the TEA Challenge 2023, where developers can have independent evaluations, open discussions, and structured feedback, provide valuable insights into the practical challenges and limitations of current MLFF approaches for developers and users alike.

4.4 Summary

This chapter presented a systematic evaluation of state-of-the-art MLFF through the TEA Challenge 2023, revealing significant advances in MLFF field alongside persistent challenges requiring further attention.

Equivariant NN architectures, presented in the TEA Challenge by MACE and SO3krates, demonstrated remarkable improvements over traditional kernel-based methods, presented by SOAP/GAP, sGDML, and FCHL19*, establishing new benchmarks for force field accuracy with MAEs in the range of 0.2-0.5 kcal/mol·Å. However, even these advanced models exhibited significant MAXE errors and struggled with long-range interactions. The heterogeneous error distribution across different atom types and chemical environments was particularly evident in the force predictions for Pb atoms in $MAPbI_3$ and for C atoms in 1,8-naphthyridine/graphene system — emphasizing that aggregate error metrics alone are insufficient to characterize MLFF reliability, and highlighting potential areas for architectural improvements.

The MD analyses show a clear trade-off across models. MACE was the most stable across systems, while SO3krates delivered the highest computational throughput. Both equivariant models reproduced vibrational spectra and molecular orientation distributions for $MAPbI_3$, but underbound 1,8-naphthyridine on graphene due to missing long-range interactions beyond their cutoff radii. SOAP/GAP achieved reasonable force test errors yet produced unstable MD, with recurrent bond breaking in $MAPbI_3$, failure to sustain 1 ns at elevated temperature, and an adsorption scan that missed the expected physisorption minimum near $\sim$3.2 Å. sGDML matched observables at 300 K for 1,8-naphthyridine/graphene system being slow in production MD. Being a global model sGDML was still not able to reproduce long-range behavior, introducing a spurious barrier and overestimated energies at large separations. Moreover, the global nature of the sGDML model prohibited it from being trained for the large $MAPbI_3$ system, which contains 384 atoms. FCHL19* reached high raw throughput but was the least stable, with extensive organic-bond breaking in 1,8-naphthyridine/graphene and $MAPbI_3$ systems. Its adsorption curve captured short-range repulsion and part of the

attraction, yet still underbound overall. All in all, equivariant MPNN based models were the only ones to combine stability, accuracy, and transferability at scale.

Several common pitfalls were identified in MLFF deployment: version drift between software releases, incomplete documentation of system requirements, inconsistent conventions across implementations, and hidden model-specific energy or force offsets. These challenges can be mitigated through best practices, including thorough cross-validation, comprehensive training metadata preservation, active learning protocols, and careful selection of appropriate accuracy targets for the intended application.

Future MLFF development should focus on addressing rare but substantial force prediction errors, improving the representation of long-range interactions, reducing error heterogeneity across chemical environments, and incorporating robust uncertainty quantification. Additionally, integrating multiscale force components into training objectives appears crucial for accurately modeling hierarchical systems like molecule-surface interfaces.

Initiatives like the TEA Challenge provide essential feedback for both MLFF developers and practitioners by independently evaluating models under realistic conditions. Such systematic benchmarking reveals practical limitations that might otherwise remain hidden in research settings, thereby accelerating progress toward more robust, reliable, and widely applicable machine learning force fields for materials science and chemistry.

Chapter 5

Orbits for Improved Machine Learning Force Fields

The architecture-independent heterogeneity of atomic errors, seen in every MLFF studied in Chapter 4, indicates an area where further refinement is needed to extend their applicability. In complex systems, atoms of the same element often participate in various interaction patterns, for example, surface versus bulk atoms in a slab, or defect versus pristine sites in a material. Distinguishing the atoms of the same element during FFs construction is the strategy used for decades in empirical force fields, where, for example, the bond order of carbon atoms determines their interaction parameters [266, 267]. However, to achieve broad applicability, it is crucial to go beyond identifying symmetries tied to specific chemical compositions and geometries. Instead, symmetries should be defined across the entire relevant chemical and configurational space.

This chapter presents a novel data-driven symmetry search method that can reveal all relevant permutations in molecules and materials present in a given dataset [268]. These permutations are used to differentiate atoms based on their surrounding chemical environments, allowing us to improve the accuracy of MLFFs by adjusting the level of symmetry exercised by the model.

We demonstrate the applicability of the developed approach on examples of a kernel-based model (sGDML) and a equivariant MPNN architecture (MACE) [18, 24, 25].

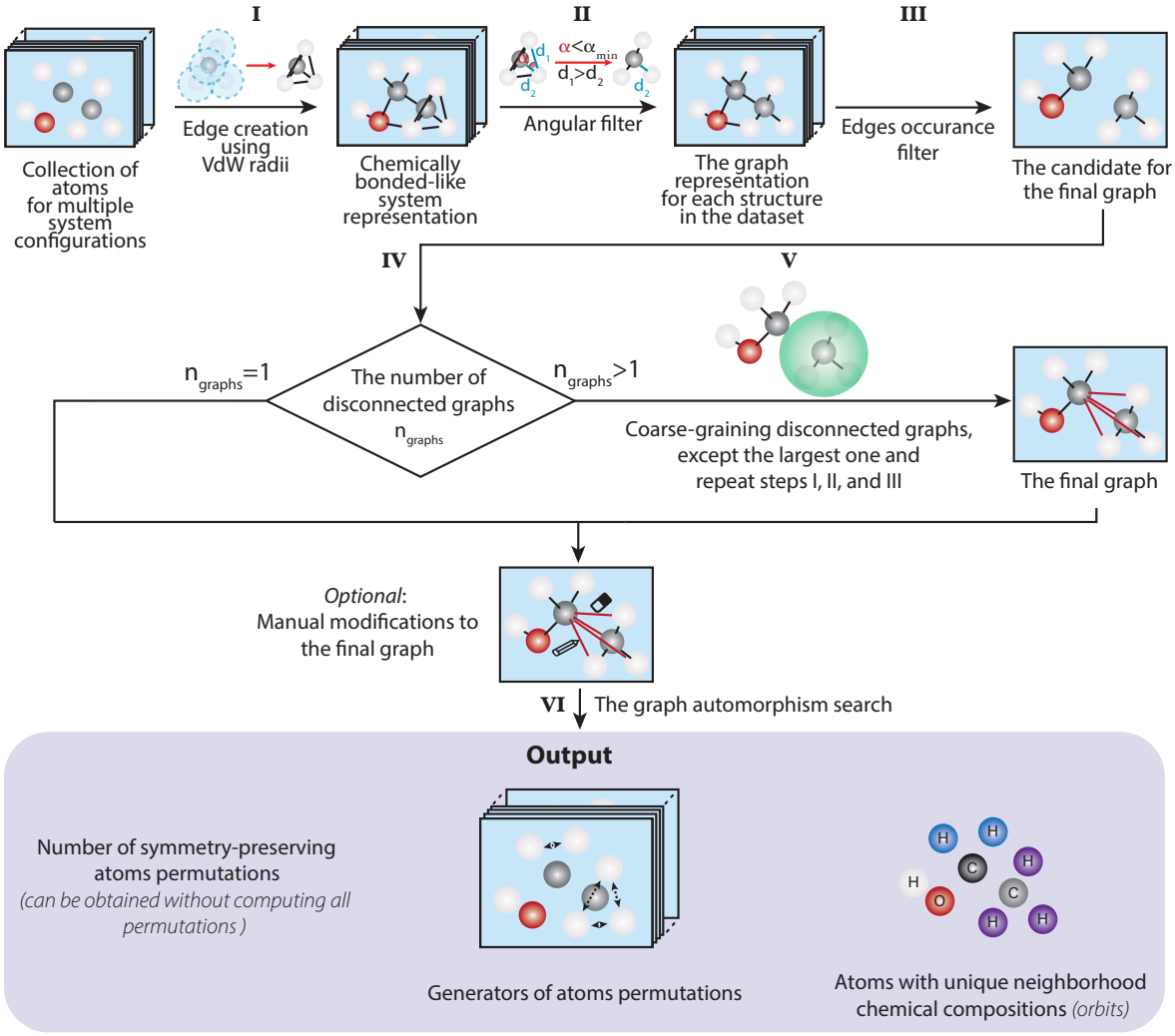


Figure 5.1 An illustrative description of the main components of the graph-based symmetry search procedure. The procedure consists of six steps: (I) creating of edges between atoms using vdW radii as cutoff; (II) filtering out the longest of two created edges if the formed angle between them is greater than the threshold angle; (III) filtering edges by occurrence frequency in the dataset; (IV) estimating the number of disconnected components; (V) coarse-graining the disconnected components; (VI) performing the graph automorphism search.

5.1 Graph-Based Symmetry Search Procedure

5.1.1 Representative Graph Construction

The representation of a system as a graph, where the vertices represent atoms and the edges represent the links between them, is widely used in chemistry and crystallography [269, 270]. The permutational symmetries of the system form the graph automorphism group, i.e., they preserve the edge-vertex connectivity. Consequently, the graph representation reduces the problem of finding symmetries in a system to finding elements of an automorphism group, a well-studied problem in the field of graph theory [271, 272]. However, building a representative graph encompassing the symmetries in a given relatively complex system

along its finite temperature evolution is not straightforward. This work proposes a procedure for constructing such graph representation that describes an entire dataset of configurations rather than one particular structure.

The first step establishes edges between individual atoms at distances smaller or equal to their respective elements' vdW radii to account for all physically and chemically relevant neighbors for each individual configuration, as shown in Figure 5.1 (I). This provides a system representation resembling chemical bonds in which neighboring atoms become directly connected.

Some edges created during the first step form *sharp angles* that are not chemically or physically meaningful. Therefore, the second step is to remove the longer of the two edges that form such angles, as depicted in Figure 5.1 (II). The threshold defining sharp angles differs from system to system and should be set manually. We found that taking this threshold as half of the smallest angle formed by translation vectors of a super-cell in the case of a periodic system is an excellent choice to obtain a structural-informative graph. After performing the first two steps, a graph representation is formed for each system configuration at our disposal. In practice, we can use only a limited number of representative structures, making the method easily applicable for large datasets.

During the third step, we sum the adjacency matrices of the constructed graphs, obtaining the frequencies of edge formations. Only the edges whose frequencies are above some predefined threshold become a part of the candidate for the *final graph*, as illustrated in Figure 5.1 (III). This eliminates the edges that are not characteristic of most configurations. The eliminating threshold value varies from system to system and depends on both the softness of the system and the presence of transient states and should be set manually. A good choice is between 70 – 90%. After the threshold is applied, the candidate for the final graph is constructed.

As shown in Figure 5.1 (IV), the fourth step is determining the number of non-connected components in the candidate's final graph. It is common for these components to arise from subsystems either distant from one another or moving quickly enough to prevent them from forming stable edges. If the candidate for the final graph has no separate components, it is considered the final graph. At this point, one can proceed directly to the search for its automorphisms.

However, if the candidate for the final graph contains disconnected subgraphs, a fifth step, shown in Figure 5.1 (V), is necessary to analyze the relationships between them. This step begins by replacing all fragments—represented by disconnected subgraphs, except the largest—with coarse-grained (Coarse-Grained (CG)) particles positioned at each fragment's

center of mass. The initial four steps are then repeated on this semi-CG system dataset to create a new graph with CG vertices corresponding to CG particles. The radius for edge creation around each CG particle is defined as the distance from its center of mass to its most distant atom plus the vdW radius of that atom. Once these steps are complete, each CG vertex is replaced by its original disconnected subgraph, with new connections added to each subgraph's vertices based on the CG vertex links. By the end of the fifth step, any previously missed interatomic connections are re-established, and the distinct fragments are identified.

Whether the final graph was obtained after the fourth or the fifth step, it can be modified manually before searching for symmetries. Manual modifications can offer advantages in systems like the chemisorbed Pd/MgO-surface, see Figure 5.2. In this scenario, when a palladium atom becomes trapped in one of the numerous local minima, it can introduce bias into the final graph. Accordingly, the biased final graph does not have information associated with other local minimum configurations missing in the dataset. However, manual modifications applied to the final graph can help to restore the absent information in the trajectory to train a robust MLFF model. Therefore, being an optional step, the manual modifications can both compensate for insufficient configurational space sampling and include physical and chemical intuitions.

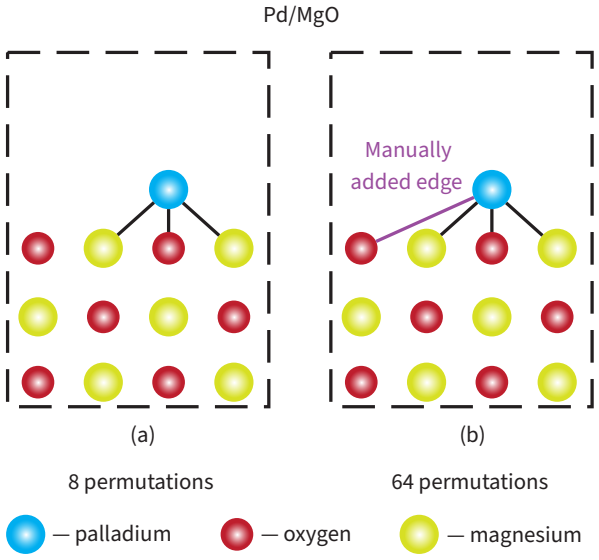


Figure 5.2 The example of manual modifications of the final graphs for the chemisorbed Pd/MgO-surface. All edges of the final graph that are not connected to the palladium atom are hidden in both (a) and (b). (a) Illustration of the final graph before the manual modifications, where permutational symmetries are biased toward single local minima. (b) Illustration of the final graph after the manual modifications, where permutational symmetries take into account other possible locations of a palladium atom.

5.1.2 Permutational Symmetry Extraction

The group theory-based graph automorphism search algorithm implemented in the `nauty` package efficiently determines the total size of the permutation group and its generators without requiring the explicit generation of all possible permutations [273]. Moreover, the information on orbits – the equivalence classes of the graph's vertices under the action of the automorphisms – can be obtained without additional computations. Different orbits represent atoms with distinct *global* neighborhood chemical compositions. Compared to the number that can be expected for a given system, a large number of different orbits reveals that the dataset is insufficiently representative or that a transition state is present, requiring careful construction of the dataset to capture its behavior accurately.

For the final multicomponent graph, the graph automorphism search is applied to each component separately to examine subsystem symmetries. As the number of components increases, the total number of symmetries grows exponentially and equals the product of the number of symmetries of each non-connected part. On the other hand, the total number of orbits shows a linear growth and is equal to the sum of the number of atoms with distinct neighborhood chemical compositions in each component. For instance, the system consisting of a 7×7 graphene sheet and 1,8-naphthyridine molecule has a total of 1176 permutational symmetries. Specifically, 588 of these symmetries come from graphene, while two come from the 1,8-naphthyridine molecule. However, this system has ten distinct neighborhood chemical compositions, consisting of one orbit from graphene and nine orbits from the 1,8-naphthyridine molecule.

The presence of local symmetry perturbations in the final graph representation reduces the number of permutational symmetries while inflating the number of orbits. Focusing on local neighborhood chemical compositions—rather than relying exclusively on global orbits—can be more advantageous. The extent of these environments is determined by the number of nearest neighbors considered, a choice guided by the range of relevant interactions in the system. In the subsequent section, we will showcase a graphene system with a pyridinic-N defect to illustrate this scenario.

Although the graph-based symmetry search procedure was discussed in the context of data produced by QC methods, it can be applied to any data represented by atomic species and their coordinates, including those sourced from an experiment.

5.1.3 Integration into Machine Learning Methods

Orbit labels and permutational symmetries, obtained via graph-based symmetry search, are used to quantify and mitigate atomic force reconstruction heterogeneity across MLFF architectures. Using results from the TEA Challenge 2023 for the Ac-Phe-Ala_5-Lys system, we first assess heterogeneity by linking error patterns to orbit labels [88]. This analysis, which yields consistent results across various deep learning architectures and kernel-based ML models, including MACE, SO3krates, and sGDML models, highlights the limitations of classifying atoms solely by their elemental identity [24, 25, 23, 16]. We next incorporate these insights in sGDML.

The sGDML model employs a global representation that treats the system as a whole rather than partitioning it into localized atomic contributions, as described in 2.3.2. The model ensures permutational invariance by explicitly symmetrizing its predictions through integration over the permutation group of a system. Consequently, incorporating permutational symmetries into the model requires obtaining a comprehensive list of these symmetries. The current implementation uses a spectral graph-matching algorithm limited to non-periodic systems. Our method extends this approach to periodic systems such as those featuring organic molecules interacting with a graphene surface.

From the same automorphism computation we also extract the atomic orbit — groups of atoms identified by permutational symmetry analysis as sharing equivalent chemical environments—as distinct atomic types. To improve the accuracy and efficiency of MACE, we extend its architecture by explicitly incorporating atomic orbits. Incorporating orbit information into any MPNN including MACE requires no changes to the model architecture. The modification is applied at the data level: each atomic species label is replaced with its corresponding orbit identifier, effectively redefining species based on neighborhood chemical compositions. As a demonstration, we select two systems: the cesium lead iodide perovskite slab and graphene with a pyridinic-N defect. The proposed symmetry search method notably enhances FFs reconstruction for complex systems when combined with the MPNN architecture. This benefits systems where atoms of the same species can be classified into multiple groups with varying neighborhood chemical compositions. For example, for graphene with a pyridinic-N defect, the presence of the defect produces different carbon environments based on proximity to the defect site, while in the $CsPbI_3$ perovskite slab system, the presence of surfaces leads to different atomic environments for the same atomic species in different layers.

		Ethanol	1,8-naph-thyridine	D-alanine	D-histidine
# Orbits		7	10	11	19
# Permutational Symmetries	Graphene	588	588	588	588
	Molecule	6	2	6	4
Force $\left[\frac{\text{kcal}}{\text{mol}\cdot\text{Å}}\right]$	MAE	0.31	0.45	0.55	0.59
	RMSE	0.54	0.72	1.07	1.20
Energy $\left[\frac{\text{kcal}}{\text{mol}}\right]$	MAE	0.53	0.61	1.07	1.35
	RMSE	0.67	0.78	1.39	1.92

Table 5.1 Accuracy of the sGDML model, along with the number of permutational symmetries and atoms with distinct neighborhood chemical compositions (orbits), for molecule/graphene systems. The models were trained on 400 configurations and tested on 12,000 configurations for each system.

5.2 Organic molecules absorbed on graphene

sGDML is a global kernel ridge regression model designed for efficient force field reconstruction of extended systems [18]. It leverages permutational symmetries to enhance data efficiency and employs a periodic Coulomb descriptor to capture system periodicity accurately. To demonstrate the capabilities of the developed graph-based symmetry search algorithm in combination with global kernel model, we selected four periodic systems in which a 7×7 graphene surface interacts with various organic molecules: ethanol, 1,8-naphthyridine, D-alanine, and D-histidine (details provided in A.3). For each system, we extracted 400 configurations from DFT-generated molecular dynamics trajectories to identify permutational symmetries and integrate them into the sGDML model in closed form training process.

Table 5.1 presents the MAE and RMSE values for the sGDML models applied to molecular systems on graphene, along with the number of orbits and permutational symmetries. The results show that the force MAE remains below 1 kcal/(mol·Å) across all systems. While both permutational symmetries and orbits influence the accuracy of sGDML models, the impact of orbits appears to be more significant. Notably, the MAE strongly correlates with the number of orbits, representing atoms in distinct neighborhood chemical compositions. Furthermore, the increasing disparity between RMSE and MAE for forces with a growing number of orbits suggests that atomic force prediction errors are highly heterogeneous, likely due to variations in interaction patterns for different orbits.

Figure 5.3 presents the molecules with atoms color-coded according to two criteria: their classification into orbits, representing distinct neighborhood chemical compositions, and their individual force MAE, averaged over the entire test set [274, 186]. As illustrated by the example of hydrogen atoms in ethanol shown in Figure 5.3 (a), three distinct types of hydrogen atoms can be identified based on their neighborhood chemical compositions. Hydrogen atoms in the R-CH_3 group share the same neighborhood chemical composition

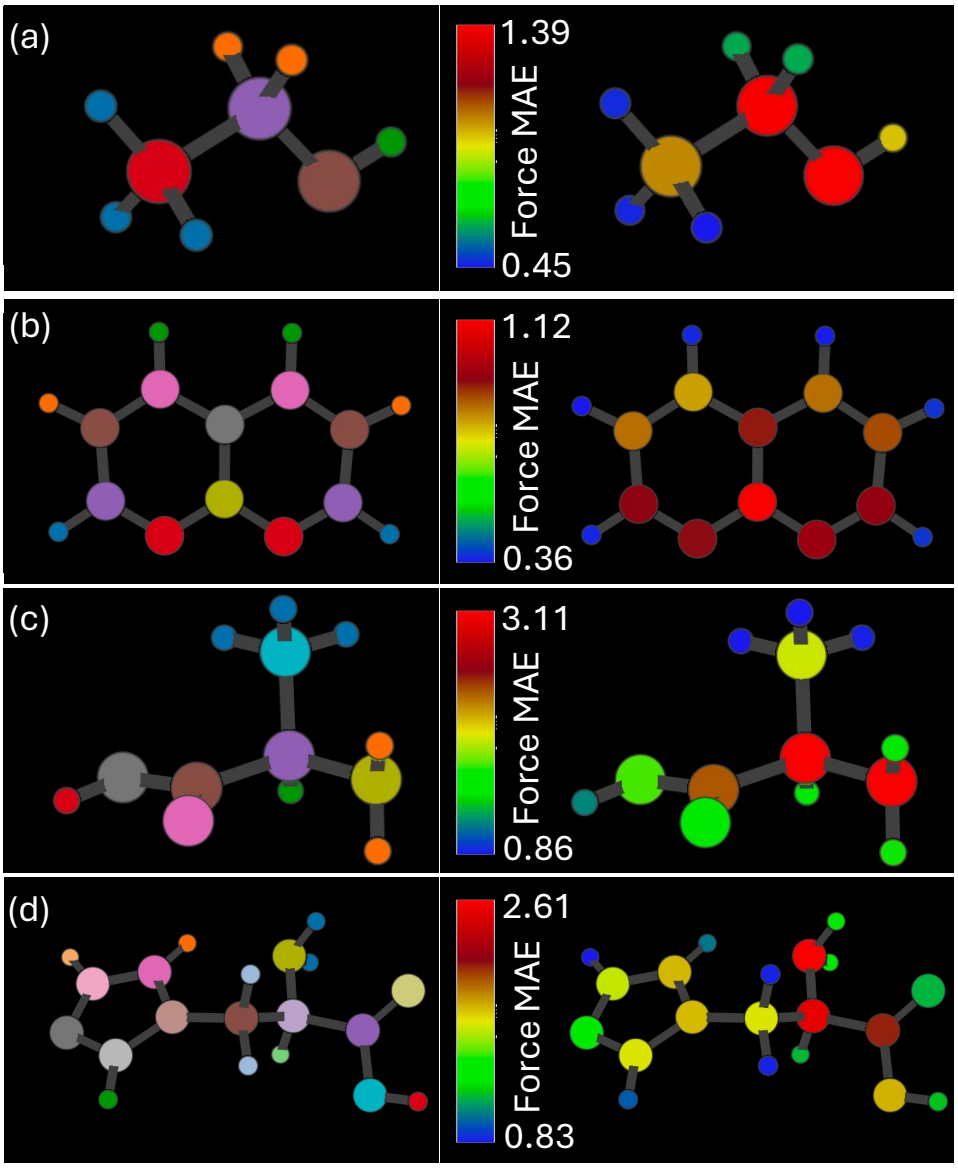


Figure 5.3 3D visualizations of molecule/graphene systems, where the first column shows atoms colored according to orbit classifications, reflecting distinct neighborhood chemical compositions, and the second column shows atoms colored by their mean average force error in kcal/(mol·Å) as predicted by the sGDML models. Figures correspond to: (**a**) ethanol, (**b**) 1,8-naphthyridine, (**c**) D-alanine, and (**d**) D-histidine molecules in the respective molecule/graphene systems.

and thus belong to the same orbit (blue). In contrast, hydrogen atoms bonded to the adjacent carbon (orange) or to the oxygen atom (green) form distinct orbits, reflecting differences in their local chemical surroundings. Although all three groups are hydrogen, their force MAEs differ across orbits because local neighborhood composition shifts the error distribution. The impact of neighborhood chemical compositions on the distribution of errors is particularly evident in 1,8-naphthyridine, shown Figure 5.3 (b). Reflection symmetry inherent to the molecule ensures that symmetrically positioned atoms share identical neighborhood chemical compositions, resulting in comparable MAEs. Consistent conclusions are observed when applying similar analyses to each atom in the other systems examined.

Overall, the analysis suggests that while the sGDML model accounts for symmetries by considering the complete set of all possible permutations without explicitly distinguishing orbits, integrating orbit-based approaches could enhance performance in architectures capable of leveraging this distinction, such as MPNN.

5.2.1 Free-Energy Profile of 1,8-Naphthyridine/Graphene

The reliable force field can be used to construct the over time Average Potential Energy Surface (APES) and the Potential Surface of Mean Force (PSMF) for the molecule on the graphene surfaces. APES and PSMF are sources of valuable insights into the free-energy landscape, the reaction pathways, and the energetics of adsorption processes. Their construction requires proper phase space sampling, accurately capturing the entire energy landscape, making *ab initio* methods hardly usable due to computation expenses. To demonstrate the performance of the BIGDML models trained using symmetries extracted by graph-based symmetry search in practical applications, the system with 1,8-naphthyridine was chosen.

A reliable depiction of the interaction between 1,8-naphthyridine and graphene necessitates thorough configurational space sampling, encompassing diverse molecule positions relative to the hexagonal graphene surface. The graphene surface has a periodic structure with repetitive hexagonal blocks characterized by the point group D6h. Consequently, the graphene surface consists of rhombic unit cells with corners located at the centers of neighboring hexagons, and each unit cell contains two carbon atoms. Thus, molecule positions on the graphene surface were bounded by a selected Graphene's Unit Cell (GUC) instead of the whole surface without losing generality. Thirty-six uniformly distributed points were selected within the GUC to properly sample the graphene surface. For each node, an MD using the obtained BIGDML model was run. The movement of the COM of the molecule was constrained in the $x - y$ plane at those points to ensure equal sampling across the surface of the GUC. This approach enabled the exploration of phase space regions that are typically difficult to access using unconstrained dynamics due to the presence of energetic barriers. In addition to the constraints on the molecule's COM, undesirable translational motions of the graphene surface were prevented by fixing positions of the carbon atom furthest from the selected GUC. All 36 MD simulations were run for 0.2 ns at 300 K using an NVT ensemble. A Langevin thermostat with a 50 fs time constant was employed to maintain the desired temperature. As a result, trajectories of 7.2 ns in total length were generated, providing accurate results comparable to QC methods. This constrained MD approach effectively implements TI, discussed in 2.2.3, with the molecule's position serving as the coupling parameter λ, allowing us to compute free-energy differences across the graphene surface by integrating the mean

forces at each sampled position.

For each grid point, the average potential energy and mean force acting on the molecule were computed after the system reached thermal equilibrium. Using radial basis function interpolation, these values were then used to calculate the smooth APES and PSMF, shown in Figure 5.4. In Figure 5.4 (a), the resulting APES reveals that the local minima within the elementary cells are separated by energy barriers with a height of around 300 K. Such energetic barriers indicate that at room temperature, the movement of molecules along the surface becomes significantly more complex than simple translational motion. Additionally, in the presence of 6-fold rotational symmetry around the z-axis of the graphene, interaction with a molecule with only mirror symmetry leaves the APES only with mirror and translational symmetry along the lattice vectors. The absence of rotational symmetry of the APES implies that the molecules exhibit a preferred orientation in the $x - y$ plane.

In Figure 5.4 (b), the PSMF is presented as a vector field, with average forces acting on the molecule indicated by red arrows. The absolute magnitude of these forces is visualized through a colored map, offering a more comprehensive insight into the intricate interactions occurring between the molecule and the substrate. For instance, on the PSMF one can see that saddle and stable critical points are positioned between the graphene atoms, while the unstable ones are on top. All three types of critical points within the system are interconnected by channels characterized by minimal changes in the acting forces on the molecule from the graphene surface. These channels represent the most favorable paths for molecular motion due to the minimal alteration in free-energy along their trajectories. Thus, the most favorable motion for the molecule's center of mass is along channels rarely crossing the carbon atoms, given the repulsive nature of the force around them.

To further delve deeper into the absence of rotational symmetry in the APES, the distribution of the angle θ was plotted for each of the 36 COM positions in Figure 5.4 (a). Here, θ represents the angle between the x-axis and a molecule orientation vector in the $x - y$ plane. The molecule orientation vector passes through the axis of symmetry of the 1,8-naphthyridine, with the direction pointing toward pyridine rings. All generated trajectories were initialized with $\theta = 266^\circ$. As shown from the angular distributions, the 1,8-naphthyridine has one, rarely two, preferred orientations within the energetic barrier zone. These preferred orientations deviate from the initial orientation by $35^\circ \pm 5^\circ$. The existence of preferable orientation and the reversal of the molecule from the initial position after thermalization within highly energetic regions imply the existence of energy barriers governing the rotational degree of freedom. At the same time, the angular distribution in the region of the global minimum is wider and tends to align more closely with the initial orientation. However, more of these distributions

have secondary peaks to indicate the presence of energy barriers to the rotation motion that the molecule can overcome. Consequently, even at room temperature, the molecule must combine translational motion within the surface plane with a rotational motion to traverse from one GUC to another.

5.3 Ac-Phe-Ala5-Lys

Orbit	MACE			SO3krates			sGDML		
	com	incom	unkn	com	incom	unkn	com	incom	unkn
$C^{(1)}$	0.85	0.74	1.08	2.22	2.10	2.79	8.40	7.44	13.42
	0.73	0.63	0.99	1.98	1.80	2.59	8.11	7.07	13.35
	0.72	0.63	0.94	1.87	1.72	2.41	8.24	7.21	13.78
	0.68	0.61	0.82	1.74	1.64	2.05	7.92	7.14	12.03
	0.71	0.64	0.90	1.90	1.73	2.37	8.04	7.10	13.37
	0.83	0.72	1.11	2.28	2.13	2.89	8.30	7.33	13.38
$C^{(2)}$	0.50	0.45	0.60	1.25	1.22	1.46	5.55	4.91	8.70
	0.51	0.45	0.59	1.24	1.21	1.45	5.53	4.89	8.62
$C^{(3)}$	0.40	0.35	0.44	0.98	0.97	1.09	5.06	4.46	7.70
	0.39	0.34	0.44	0.96	0.93	1.10	5.00	4.37	7.84
	0.38	0.33	0.41	0.88	0.85	0.93	4.95	4.41	7.48
$C^{(4)}$	0.65	0.57	0.79	1.73	1.62	2.13	8.62	7.72	14.05
	0.63	0.55	0.76	1.66	1.56	2.00	8.45	7.47	14.40
	0.58	0.52	0.68	1.51	1.43	1.73	8.16	7.27	13.08
	0.61	0.55	0.74	1.62	1.53	1.87	8.43	7.39	14.05
	0.69	0.61	0.84	1.86	1.74	2.27	8.51	7.47	13.90

Table 5.2 Normalized atomistic force MAE (in %) for Ac-Phe-Ala5-Lys molecule. Orbits correspond to carbon atoms highlighted in Figure 5.5. For details on the MLFF architectures and different test sets (complete (com), incomplete (incom), and unknown (unkn)), refer to Ref [88].

To further explore the influence of orbits on atomic force reconstruction across various chemical systems and MLFF architectures, we analyze the results from the TEA Challenge 2023 [88]. The TEA Challenge 2023 is a benchmarking initiative that evaluates MLFF performance across diverse molecular systems, systematically assessing models with various architectures such as MACE, SO3krates, sGDML, SOAP/GAP, and FCHL19* [24, 25, 142, 23, 16, 17, 15, 137, 139]. The test systems included two biomolecular structures (alanine tetrapeptide and N-acetylphenylalanyl-pentaalanyl-lysine), a 1,8-naphthyridine/graphene interface, and a methylammonium lead iodide perovskite. The challenge focused on key tasks, including the reconstruction of potential energy surfaces, the handling of incomplete reference data, the modeling of multi-component systems, and the simulation of complex periodic structures. The results revealed pronounced heterogeneity in atomic force prediction errors, with worst-to-best atomistic MAE ratios ranging from 5:1 to 6:1, consistently observed

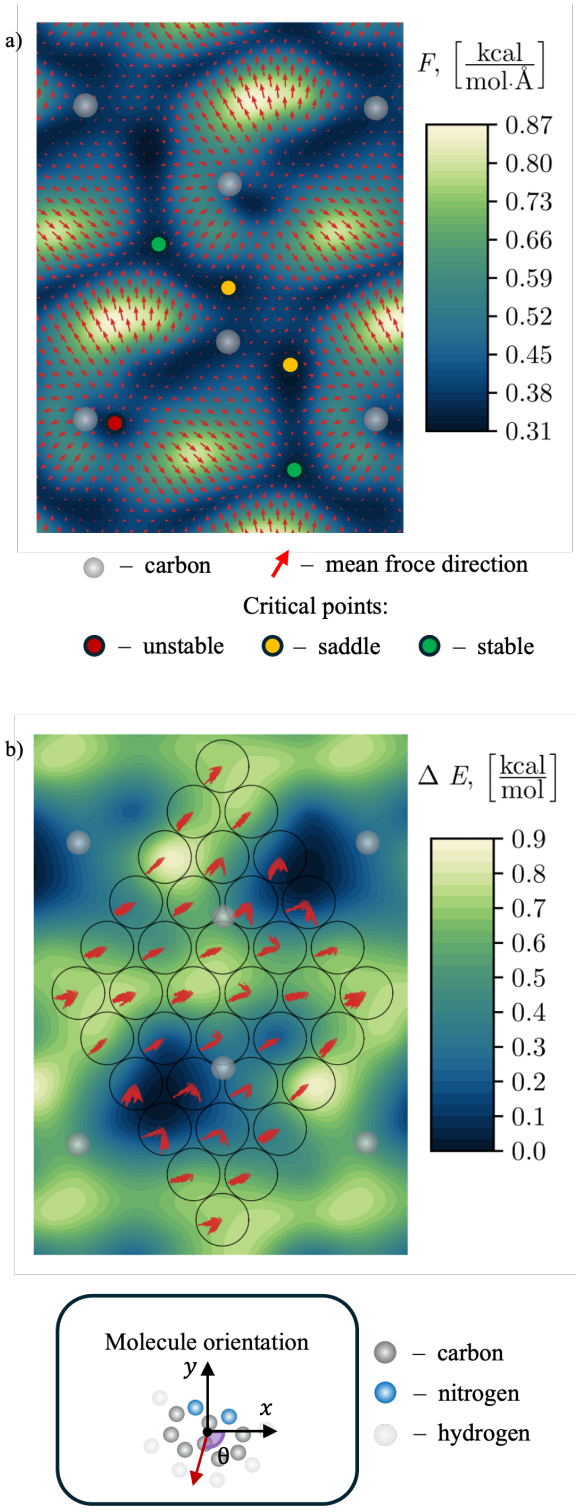


Figure 5.4 The 1,8-naphthyridine molecule on a pristine graphene layer at ambient conditions. (**a**) Potential Surface of Mean Force (PSMF), with key points marked as red (unstable), yellow (saddle), and green (stable) dots, indicating regions of instability, saddle points, and stability, respectively. (**b**) Angular orientation distribution of the molecule at different center of mass positions within graphene's unit cell, overlaid on the Average Potential Energy Surface (APES).

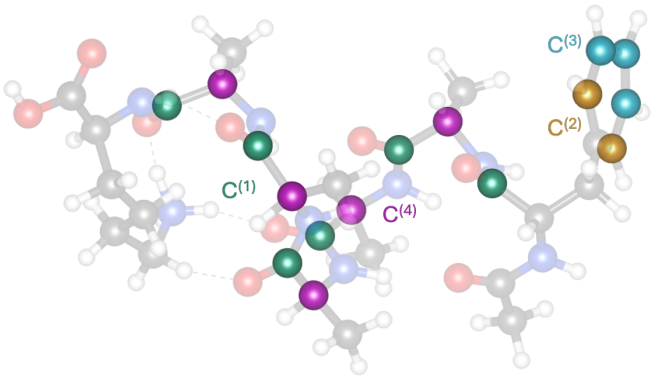


Figure 5.5 A snapshot of the Ac-Phe-Ala5-Lys molecule geometry. $C^{(1)}$ – $C^{(4)}$ represent sets of carbon atoms belonging to four different orbits.

across different architectures. Here, we selected the Ac-Phe-Ala5-Lys molecule (100 atoms) to investigate the origin of this heterogeneity using orbits identified through the developed permutational symmetry search method.

Since orbits are defined as sets of vertices that remain indistinguishable under all graph automorphisms, applying graph automorphism analysis directly to complex molecules can lead to an excessive number of chemically distinct environments, even for atoms that should be treated identically based on chemical intuition. Distinguishing atomic species based on their local neighborhood chemical compositions is more effective than relying solely on global permutational symmetry in such systems. Therefore, while orbits traditionally refer to global permutations, hereafter, we extend the term to include atoms characterized by distinct local environments.

Figure 5.5 presents a molecular snapshot highlighting four sets of carbon atoms that form orbits with identical neighborhood chemical compositions up to the second nearest neighbors, considering only orbits containing multiple atoms. Table 5.2 reports the mean absolute errors (MAEs) for atomic force reconstruction using two MPNN architectures, MACE and SO3krates, along with the kernel-based sGDML model. The errors are evaluated on the same test set and model prediction data used to generate Table 1 in Ref [88]. To account for variations in force magnitudes across the molecule, MAEs are expressed as percentages of the averaged norm of the true force acting on an atom. Carbon atoms within the same orbit are grouped together, revealing a clear pattern: relative atomic force MAEs remain highly consistent within each orbit but differ significantly across different orbits. Similar results can be observed for absolute atomic force MAE values, as well as for other chemical elements and systems benchmarked in the TEA Challenge 2023.

This trend aligns with our findings for smaller molecules on graphene in the previous subsection, further emphasizing the limitations of treating all atoms of a given chemical

element identically in machine learning models, despite their diverse bonding environments.

5.4 Graphene with pyridinic-N defect

Beyond providing insight into the origin of force reconstruction heterogeneity observed across all ML models and chemical systems analyzed, orbits can also be directly leveraged to improve the performance of state-of-the-art MLFF architectures. As a demonstration, we considered a 15×15 graphene lattice containing a pyridinic-N defect, shown in Figure 5.6 (a), which disrupts the symmetry among carbon atoms. For this study, we choose the MACE model, an equivariant MPNN, and employ orbits (atoms with persistent neighborhood chemical compositions) extracted using the developed graph-based symmetry search algorithm to enhance training efficiency and predictive performance [24, 25]. The standard MACE architecture has three hyperparameters that are critical for accuracy and ultimately define the model's size:

- The maximum order of irreducible representations (irreps) that determine the complexity of the internal feature space;
- The number of channels, i.e., features belonging to each hidden irreps, indicates the dimensionality of the internal representation at a given symmetry level.
- The design of the MLP that defines the radial resolution of the model.

The default configuration for the MACE model incorporates hidden representations up to the second order, with each order containing 128 features. Additionally, the radial multilayer perceptron comprises three fully connected layers, each with 64 neurons. However, for experiments with models' sizes that were changed with by varying channel counts, the radial MLP consists of three fully connected layers, each with only 8 neurons, since that was enough to obtain error below 1 kcal/(mol·Å). All MACE models were trained under identical conditions, employing early stopping on the validation loss with a patience of 2048 epochs to ensure convergence.

As an equivariant model, MACE captures symmetries by translating discrete permutation symmetry into a continuous representation governed by rotational equivariance. This approach enables efficient parameterization through the use of irreducible representations (irreps) of the rotation group $SO(3)$. However, this conversion also introduces a limitation: the model is constrained to operate within the full rotation group and cannot explicitly emphasize specific symmetry subgroups, such as particular atom permutations. By preserving a direct

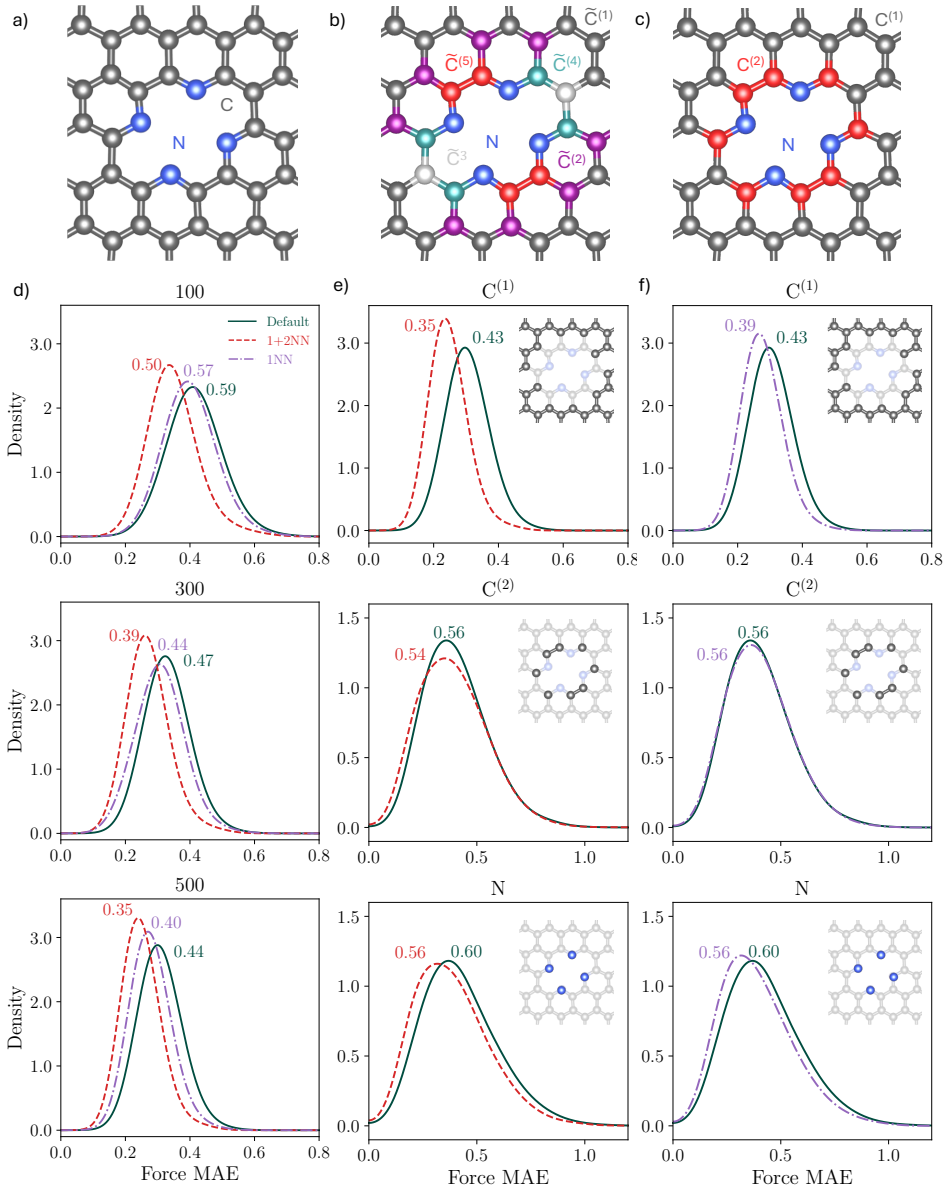


Figure 5.6 Graphene with a pyridinic-N defect system is represented using three approaches: (**a**) default, (**b**) first- and second-neighbor (1+2NN), and (**c**) first-neighbor structural encoding (1NN). (**d-f**) Distributions of Mean Absolute Error (MAE) in force predictions, in kcal/(mol·Å), for MACE models applied to the graphene system with pyridine-type defect. Solid green lines represent the default MACE architecture, red dashed lines indicate the first and second-neighbor-based model, and purple dashed-dotted line denote the first-neighbor-based model. (**d**) Force MAE distributions calculated for all atoms in the system, for the MACE models trained on 100, 300, and 500 reference geometries. (**e, f**) Force MAE distributions specifically for nitrogen, carbons bonded to nitrogen, and other carbon atoms in the system using MACE models trained on 500 configurations, where default MACE architecture compared with (**e**) the first and second-neighbor-based MACE model, and with (**f**) the first-neighbor-based MACE model.

encoding of permutation symmetry, it becomes possible to treat both discrete permutation and continuous rotational symmetries explicitly and independently, offering greater flexibility and potentially improved inductive bias in the model. By maintaining an explicit connection to permutation symmetry—specifically by partitioning atoms into equivalence classes based on their local chemical environments—the model can separately and directly encode both discrete permutation and continuous rotational symmetries. In this framework, permutation invariance within each equivalence class is enforced a priori, allowing the network to concentrate its capacity on learning rotational equivariance within each class. This approach offers finer control over symmetry treatment and improves both the accuracy and efficiency of MACE. By organizing atoms into classes with stable neighborhood chemical environments, the model reduces hypothesis space complexity, accelerates convergence, and lowers generalization error. Crucially, it focuses the network on learning meaningful geometric correlations rather than rediscovering redundant permutation symmetries.

To further explore how these equivalence classes influence model performance, we assessed the impact of different orbit-based structural encodings on MACE's performance by constructing three datasets containing 200, 600, and 1,000 configurations each (for details, see Appendix A.4). For every dataset, we trained three types of models using standard hyperparameter settings: one with default atomic species (formally classified as a 0NN model) and two incorporating 1NN and 1+2NN structural encodings. Training process followed a 50/50 split between training and validation datasets.

The default structural encoding of the system shown in Figure 5.6 (a) distinguishes only carbon and nitrogen atoms. It is then refined by adjusting the number of nearest neighbors that define each atom's neighborhood chemical composition, capturing a more detailed view of local symmetries. For example, including first and second neighbors into consideration (1+2NN) yields four additional inequivalent orbits for carbon. Specifically, $\tilde{C}^{(1)}$ is unaffected by the defect, $\tilde{C}^{(2)}$ has one nitrogen atom among its second neighbors, $\tilde{C}^{(3)}$ has two nitrogen atoms among its second neighbors, $\tilde{C}^{(4)}$ has one nitrogen atom as a first neighbor, and $\tilde{C}^{(5)}$ has one nitrogen atom as a first neighbor plus another as a second neighbor. In contrast, Figure 5.6 (c), which considers only first-neighbor structural encoding (1NN), introduces just two inequivalent orbits for carbon: $C^{(1)} = \tilde{C}^{(1)} \cup \tilde{C}^{(2)} \cup \tilde{C}^{(3)}$, unaffected by the defect, and $C^{(2)} = \tilde{C}^{(4)} \cup \tilde{C}^{(5)}$, which has one nitrogen atom as a first neighbor.

The MACE models that were trained on default (0NN), 1NN, and 1+2NN structural encodings use a 5 Å cutoff and two interaction blocks, effectively increasing their effective interaction range to 10 Å. In the considered defective graphene system, every atom's third-nearest neighbors lie within 4 Å. Thus, orbit partitioning does not extend the model's spatial

reach under any encoding, but simply reorganizes the same local information into symmetry-informed classes, letting the network focus on geometric features rather than relearning permutation invariance.

Figure 5.6 (d) illustrates the force MAE distributions predicted for 0NN, 1NN, and 1+2NN models, each trained on 100, 300, and 500 structures, respectively. The force MAE distributions for graphene with a pyridinic-N defect display a single Gaussian-like peak across all dataset sizes. For 100 training points, the default model yields a force MAE distribution peaking around 0.40 kcal/(mol·Å) with RMSE of 0.59 kcal/(mol·Å). In contrast, the 1NN and 1+2NN models produce noticeably narrower distributions that peak near 0.38 and 0.30 kcal/(mol·Å), with corresponding RMSE values of 0.57 and 0.50 kcal/(mol·Å), respectively. As the number of training points increases, the relative performance ranking of the models remains consistent, while the error distributions shift toward smaller values. With 500 training points, the RMSEs decrease to 0.35, 0.40, and 0.44 kcal/(mol·Å) for the 1+2NN, 1NN, and default MACE models, respectively.

To go deeper into the effects of different structural encodings on specific subsets of atoms, we separately analyze the precision of the predicted forces for nitrogen atoms (N), carbons bonded to nitrogen ($C^{(2)}$), and all other carbons ($C^{(1)}$). Figures 5.6 (e) and (f) show the force MAE distributions for those atoms, illustrating the influence of the local pyridinic-N defect on the performance of MACE models when trained with 500 configurations. By examining these subsets individually, one finds that orbits-based models outperform the default one most noticeably for $C^{(1)}$ – the largest subset of atoms. For $C^{(1)}$, the MAE distributions of the models narrow and shift toward smaller errors, reducing the RMSE from 0.43 (0NN) to 0.39 (1NN) and 0.35 (1+2NN) kcal/(mol·Å).

However, the MAE distributions and RMSE for the $C^{(2)}$ subset remain largely unchanged in both orbit-based models, likely due to its role as a "force outlier" for carbon atoms. By employing orbits, the larger $C^{(1)}$ subset is effectively decoupled from smaller and chemically distinct $C^{(2)}$, allowing orbit-based models to significantly enhance force reconstruction for $C^{(1)}$. Similarly, nitrogen atoms also exhibit improved force predictions when orbit-based models are used. This can be attributed to the fact that, in the default MACE model, the nitrogen subset initially has the highest error, 0.60 kcal/(mol·Å), making it more responsive to orbit-based encoding. As a result, its MAE distributions improve, ultimately aligning the nitrogen atoms and $C^{(2)}$ RMSEs at 0.56 kcal/(mol·Å).

These findings indicate that orbits with the greatest impact on the loss function—typically those containing the largest subset of atoms or exhibiting the highest errors—experience the most substantial improvements.

To validate that the 1+2NN orbit-based model accurately captures key physical properties, we compare the Γ-point phonon spectrum of the system with that obtained from the default MACE model. Although the default model has approximately 1.6 times more parameters, both models achieve nearly identical force RMSE and MAE. As shown in Figure C.2, the phonon frequencies at the Γ-point closely align across all vibrational modes, demonstrating that the orbit-based model reproduces vibrational behavior with accuracy comparable to the default, despite its significantly reduced complexity.

5.5 Cesium Lead Iodine Perovskite Slab

To validate our conclusions regarding the advantages of incorporating orbits in MLFFs across different system sizes, we extend our analysis to a larger system—a $CsPbI_3$ perovskite slab with 216 atoms per unit cell. As shown in Figure 5.7 (a), the $CsPbI_3$ perovskite slab exhibits only reflective symmetry along the z-axis, resulting in three distinct, non-equivalent layers of PbI_2 and CsI: the outer layer, the in-between layer, and the middle layer. Consequently, despite containing only three chemical elements, the system features 12 distinct orbits defined by global permutational symmetry—six for iodine, three for cesium, and three for lead—providing a more granular structural representation for MACE models. By enforcing these orbits, the orbit-based variant effectively extends its receptive field beyond the nominal 13 Å with 6.5 Å cutoff, allowing it to capture longer-range structural information that the default model cannot access.

To assess the impact of orbits on data efficiency and model accuracy, we assembled three DFT-based datasets, comprising 600, 1,000, and 2,000 configurations each (for details, see Appendix A.5). Subsequently, two MACE models with standard hyperparameter settings were trained for each dataset – one utilizing default atomic species and the other incorporating orbits. The training process used a 50/50 split for training and validation subsets. For experiments on model size, four distinct models were trained for each model type and channel count using a cross-validation procedure on 2,000 configurations, with a 50/50 split between training and validation datasets. Additionally, early stopping on the validation loss with a 2048-epoch patience was used to guarantee full convergence for all models. All obtained ML models were tested on the test sets consisting of 8,000 configurations, ensuring the stability and reliability of our results.

Figure 5.7 (b) illustrates the distribution of the MAE of forces predicted on the same test set by the MACE models trained on 300, 500, and 1,000 structures, respectively. For all training set sizes except 1,000, two distinct peaks are shown in the distribution. The left and right

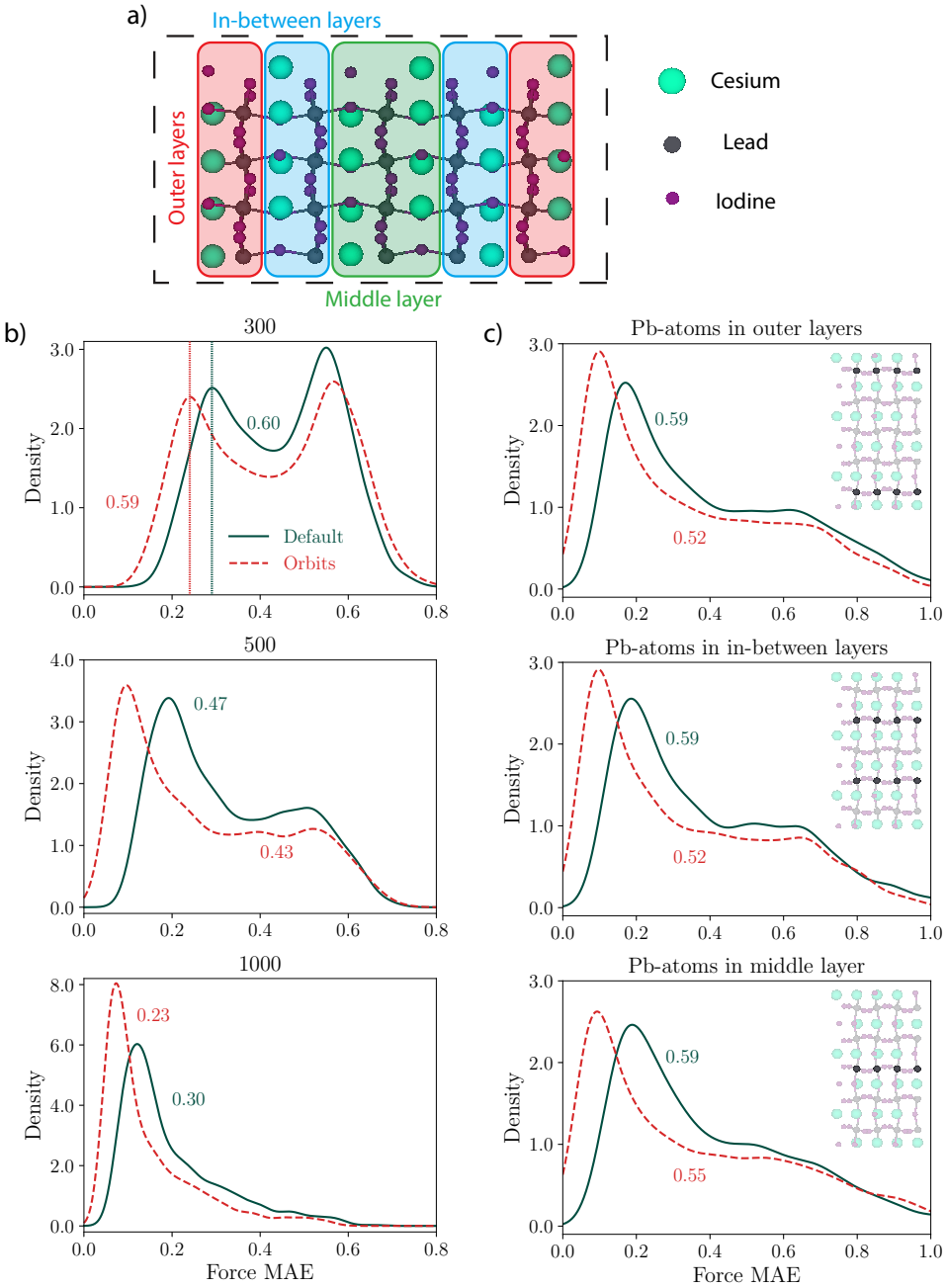


Figure 5.7 (**a**) The perovskite $CsPbI_3$ slab system in the zx projection. (**b-c**) Distributions of Mean Absolute Error (MAE) in force predictions, in kcal/(mol·Å), for MACE models applied to the $CsPbI_3$ perovskite slab system. Solid green lines represent the default MACE architecture, while red dashed lines indicate the orbit-based MACE models. (**b**) Force MAE distributions calculated for all atoms in the system, for the MACE models trained on 300, 500, and 1,000 reference slab geometries. (**c**) Force MAE distributions specifically for lead atoms, analyzed across different slab layers, using MACE models trained on 500 configurations.

peaks in the low and high MAE regimes highlight the existence of properly sampled and under-sampled regions of the configuration phase space, respectively. For 300 training points, the right-side peak is centered at 0.54 kcal/(mol·Å) for the model with default species and thus has slightly better accuracy for those under-sampled regions than the model with orbits, which shows a broader right-side peak centered at 0.58 kcal/(mol·Å). The difference between the two models is highlighted in well-sampled regions of configurational space. For the left-side peak, the model with orbits has higher accuracy than the default model with a peak centered around 0.23 kcal/(mol·Å) as opposed to 0.30 kcal/(mol·Å). In total, the two models present similar overall accuracies, with the inclusion of orbits favouring the well-sampled regions.

With the increasing number of training points, the right-side peak vanishes faster for the models with orbits than for those with default species, pushing the overall accuracy in the orbits' favor. At 1,000 training points, both models demonstrate a single peak distribution. The MACE model with orbits is narrower, with a peak of around 0.08 kcal/(mol·Å) compared to the MACE model with default species presenting a broader distribution and a peak at around 0.18 kcal/(mol·Å).

Figure 5.7 (c) depicts the distributions of the MAE across configurations for three lead orbits separately. This analysis pertains to both types of models trained on 500 configurations, thereby showcasing the individual contribution of each lead's orbit to the overall force MAE distribution. Including orbits in the model yields more significant improvements in accuracy for both the outer and in-between layers than for the middle layer. This difference is attributed to the middle layer having two times fewer lead instances than other layers. Other chemical elements within the system show similar trends and distributions.

In MPNN models, the increase in atom types correlates with a rise in trainable parameters. However, our findings demonstrate that the observed improvement of MACE models within the proposed approach is more than just a result of an increased number of parameters. Figure 5.8 (a) illustrates the accuracy of these models as a function of the number of channels, which represent the dimensionality of the internal representation at a specific symmetry level. Each model type and channel count underwent training and testing on four distinct datasets, with the resulting variability in accuracy depicted as error bars. Graphs for all model types exhibit linear trends in the logarithmic scale, where the models treating all atoms as the same species have the lowest accuracy. In contrast, orbit models display the highest accuracy for the same number of channels. Notably, the difference in accuracy among these models remained nearly constant when plotted on a logarithmic scale. This suggests that increasing the number of parameters for a given tensor representation can provide sufficient capacity

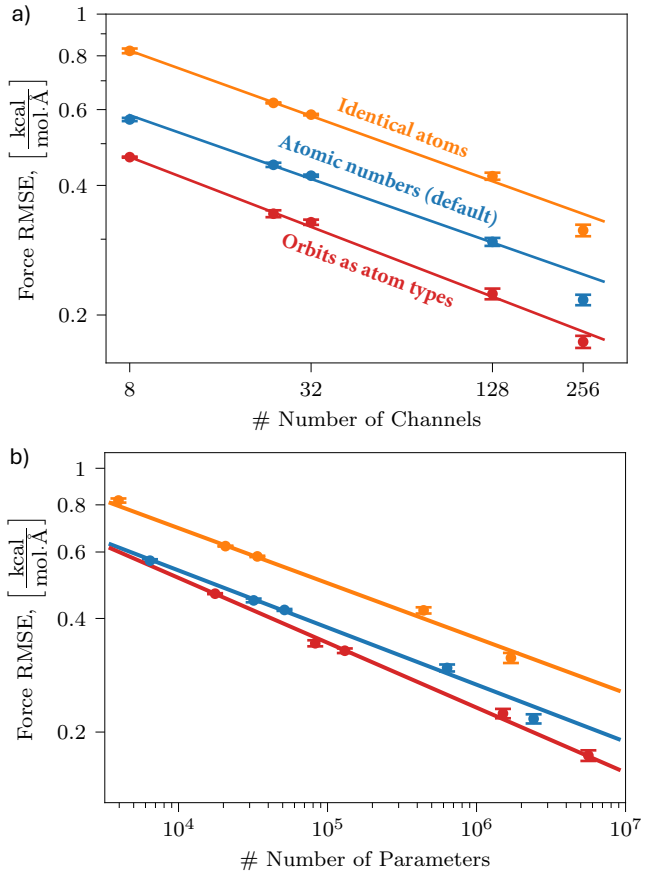


Figure 5.8 Force prediction accuracy of MACE models in the $CsPbI_3$ slab system as a function of the models' size. The Root Mean Square Error (RMSE) is evaluated across three model architectures: the default RMSE architecture (blue), an orbit-based MACE architecture (red), and a MACE architecture treating all atoms as identical (orange). Uncertainty was quantified using cross-validation testing. (**a**) The RMSE as a function of the number of channels in hidden irreducible representations, with other model parameters held constant. (**b**) The RMSE as a function of the total number of trainable parameters.

to capture the neighborhood chemical composition, even without considering specific atom types. This is particularly true in systems where the neighborhood chemical composition remains consistent during the system's evolution.

Any MPNN, including MACE, is invariant under permutation of the same chemical species, using the same function for force predictions. Employing orbits expands the number of different chemical species in the models, increasing the number of distinct functions used. To benefit from such an increase, the training set needs to be large enough for every orbit to be well sampled; otherwise, the parameters of those functions cannot be accurately trained. As a demonstration, Figure 5.8 (b) shows a relationship between the accuracy of different types of the MACE models and the total number of trainable parameters they have, where the variability in accuracy is depicted as error bars. The figure corresponds precisely to the same models as Figure 5.8 (a). Notably, orbit-based MACE models achieve comparable accuracy with significantly fewer parameters than the default MACE models. For example, to reach

a target accuracy of 0.4 kcal/(mol·Å), an orbit-based MACE model requires approximately 9×10^4 parameters. In contrast, the default MACE model, which relies on atomic species, requires around 7×10^5 parameters—an order of magnitude more. In C.3 of Appendix, we provide epoch-resolved training and validation losses for all models discussed in this section, illustrating their convergence behaviour during optimization.

Our results show that while increasing the number of trainable parameters in MPNN models generally improves prediction accuracy, the proposed orbit-based approach achieves comparable or even superior accuracy with significantly fewer parameters. This suggests that explicitly incorporating chemically distinct atomic environments through orbits enables a more efficient parameterization, reducing model complexity without sacrificing accuracy.

A similar analysis for graphene with a pyridinic-N defect system is presented in Appendix C.1, confirming the generality of our conclusions across different systems.

5.6 Summary

We have developed a graph-based method for identifying permutational symmetries, applicable to both periodic and non-periodic chemical systems. Our approach systematically uncovers symmetries across the entire configurational space by analyzing atomic positions from a dataset of configurations. It outputs the size and generators of atomic permutation sets, along with a classification of atoms into chemically distinct environments (orbits). Notably, this information is determined prior to explicitly computing permutations. Our method has been successfully integrated with the kernel-based sGDML framework and the equivariant MPNN MACE model.

The practical utility of our approach was demonstrated by training accurate MLFFs for four organic molecules: ethanol, 1,8-naphthyridine, D-alanine, and D-histidine adsorbed on a pristine graphene substrate using the sGDML architecture. We established that the accuracy of atomic force predictions is strongly influenced by the number of orbits in a system. Greater chemical diversity (resulting in more orbits) leads to increased force reconstruction heterogeneity and larger MAE and RMSE values. This trend was further validated across multiple MLFF architectures, including MACE, SO3krates, and sGDML, for the N-acetylphenylalanyl-pentaalanyl-lysine using the results of the TEA Challenge 2023 benchmark.

To demonstrate the capability and efficiency of trained sGDML models, the APES and the PSMF were constructed for the system with 1,8-naphthyridine. The built APES revealed energy barriers exceeding 300 K between the local minima of the unit cell of graphene. Moreover, it was found from the PSMF that 1,8-naphthyridine has preferable trajectories

along narrow channels with minimal alterations in the free energy of the system, avoiding the COM of the molecule being on top of the graphene carbon atoms. Furthermore, the presence of preferable angular orientations of 1,8-naphthyridine at various positions on the graphene surface showed the existence of energy barriers for the rotational degree of freedom along the axis perpendicular to the surface. Accordingly, the molecule's movement at room temperature involves an interplay between translational motion within the surface plane and a rotational motion in the molecule plane.

By incorporating orbit information into the MACE training process, we observed significant improvements over the default MACE architecture, which relies solely on atomic types. The tests were run for the graphene with pyridinic-N defect and $CsPbI_3$ perovskite slab. For example, to achieve a target accuracy of 0.4 kcal/(mol·Å) for the $CsPbI_3$ slab (sufficient for multiple practical applications), an orbit-based MACE model required approximately 9×10^4 parameters—an order of magnitude fewer than the 7×10^5 parameters needed for the default MACE model. The advantages of using orbits become more pronounced as system complexity and dataset size increase. Our results consistently demonstrate that orbit-based models outperform standard approaches in force prediction accuracy, owing to their more efficient parameterization.

In this work, we have focused on a structurally well-defined benchmark system to demonstrate the accuracy improvements enabled by explicit permutation-symmetry encoding. Extending orbit-based symmetry methods to fully dynamic settings, where atoms may alter their local chemical environments through processes such as defect migration, bond breaking and formation, or other structural transformations, will require the development of dynamic orbit assignments. In particular, future MLFF architectures should smoothly embed atomic neighborhoods to allow orbit labels to evolve consistently throughout MD trajectories. We consider this a critical and promising direction for future research.

Additionally, leveraging orbits offers several key benefits for building better universal MLFFs, which aim to cover essentially all chemistries and structures. Notable examples include SO3LR (based on SO3krates architecture) [142, 23, 143], MACE-MP-0 for materials and MACE-OFF23 for organic molecules (based on MACE architecture) [24, 25, 145, 144], SevenNet-MF-0 (based on SevenNet) [275, 276], AIMNet2 [277], and Orb [278] and many others. First, orbit analysis may provide a principled way to balance and augment datasets: it highlights underrepresented atomic environments, guiding the inclusion of additional configurations or the weighting of data such that all relevant orbits are adequately sampled. Second, the set of orbits present in a new system may serve as a fingerprint to assess model applicability. If a queried structure contains orbits that were never seen during training, the

model's predictions lie in an out-of-distribution regime and should be treated with caution. Third, incorporating orbit information directly into the universal MLFFs architecture can improve its accuracy, efficiency, and reliability.

In summary, the proposed graph-based symmetry search method represents a significant step forward in modeling large, complex, and chemically diverse systems using state-of-the-art MLFFs. By leveraging orbits, our approach enhances predictive accuracy and computational efficiency, making it a valuable tool for future applications in computational chemistry and materials science.

Chapter 6

Conclusions

This thesis investigates MLFF models for atomistic simulations of semiconducting materials and their interfaces. We assess how the choice of reference electronic-structure method used to generate training data affects predicted structural dynamics and electronic properties. We further benchmark multiple MLFF architectures across different systems and then explore the crucial role of symmetry for both accuracy and computational efficiency of the models.

6.1 Key Contributions and Findings

A comprehensive investigation of $CsPbI_3$ halide perovskite slabs demonstrates that the choice of dispersion method in the underlying DFT reference significantly influences both the structural and electronic properties obtained from subsequent MLFF simulations. The comparison between D3 and MBD-NL methods demonstrated that MBD-NL stabilizes orthorhombic tilt patterns consistent with experimental observations, while the pairwise D3 artificially favors cubic-like behavior at room temperature. This fundamental difference in structural dynamics directly impacts electronic properties, with MBD-NL producing wider band gaps (mean 1.55 eV versus 1.32 eV for D3) and smaller dipole fluctuations. These findings underscore how the physical assumptions embedded in the reference calculations propagate through the MLFF, ultimately shaping finite-temperature properties, with a pronounced 0.23 eV variation in band-gap predictions arising solely from the chosen dispersion treatment.

Differences in observables may also stem from the choice of MLFF architecture. Benchmarking in the TEA Challenge 2023 showed that equivariant MPNN architectures (MACE and SO3krates) significantly outperform kernel-based methods (SOAP/GAP, sGDML, FCHL19*), achieving force prediction accuracies with MAEs in the range of 0.2-0.5 kcal/mol·Å. However, even these advanced models exhibited persistent challenges: maximum force errors often

exceeded 10 kcal/mol·Å, heterogeneous error distributions across atomic environments, and limitations in capturing long-range interactions. This systematic evaluation across multiple systems demonstrated that equivariant models achieve the best balance of accuracy, stability, and computational efficiency, particularly for complex interfaces and solids.

A graph-based symmetry search method was developed to address errors heterogeneity seen first in $CsPbI_3$ slab studies for MACE model and then in other systems and models in the TEA Challenge 2023. By identifying permutational symmetries and classifying atoms into orbits based on their local chemical environments, this approach enables more efficient parameterization and decreasing error heterogeneity of both kernel-based sGDML framework and the equivariant MPNN MACE model. We demonstrated practical utility by training accurate MLFFs for ethanol, 1,8-naphthyridine, D-alanine, and D-histidine molecules adsorbed on pristine graphene using the sGDML architecture. Force-prediction accuracy depends strongly on the number of orbits: greater chemical diversity yields more orbits, higher force-reconstruction heterogeneity, and larger MAE and RMSE. This trend was confirmed across MACE, SO3krates, and sGDML on N-acetylphenylalanyl-pentaalanyl-lysine molecule using TEA Challenge 2023 results. Injecting orbit information into MACE delivered clear gains over the atom-type-only default, with tests on graphene with a pyridinic-N defect and on a $CsPbI_3$ slab. To reach 0.4 kcal/(mol·Å) on the $CsPbI_3$ slab, the orbit-based MACE used $\sim 9 \times 10^4$ parameters versus 7×10^5 for the default architecture. As system size and dataset complexity increase, orbital-based models consistently achieve superior performance owing to their more efficient parameterization.

6.2 Limitations and Future Directions

Despite significant advances achieved in this thesis, several limitations and opportunities remain for future research. The TEA Challenge 2023 exposed gaps in the scope, physics, uncertainty, and tooling. The next TEA-like benchmark should span a broader range of chemistries and conditions, including organics, metals, ionic and magnetic systems, stress states, and other relevant environments. It should add explicit long-range evaluations using interaction curves, lattice energies, and adsorption on realistic catalyst surfaces. An active-learning track with a fixed *ab initio* budget would measure data efficiency and reinforce safe exploration. Transfer tests should quantify zero-shot and few-shot performance of universal or pre-trained MLFFs on unseen materials and conditions. The scorecard should expand beyond energies and forces to stresses, phonons, elastic constants, diffusion barriers, and other application-relevant observables. This will allow more comprehensive assessment of

strengths and weaknesses of newly developed model architectures.

The orbit-based symmetry approach developed here could be extended to fully dynamic settings, where atoms may alter their local chemical environments through processes such as defect migration, bond breaking and formation, or other structural transformations. This will require the development of dynamic orbit assignments; in particular, future MLFF architectures should smoothly embed atomic neighborhoods to allow orbit labels to evolve consistently throughout MD trajectories, which we consider a critical and promising direction for future research. Additionally, leveraging orbits offers several key benefits for building better universal MLFFs that aim to cover essentially all chemistries and structures. First, orbit analysis may provide a principled way to balance and augment datasets by highlighting underrepresented atomic environments and guiding the inclusion or weighting of data so that all relevant orbits are adequately sampled. Second, the set of orbits present in a new system may serve as a fingerprint to assess model applicability, since predictions for structures containing unseen orbits lie in an out-of-distribution regime and should be treated with caution.

6.3 Concluding Remarks

The intersection of machine learning, quantum mechanics, and materials science is rapidly transforming our ability to model and understand complex semiconductor interfaces. This thesis contributes methodological innovations and systematic benchmarks that address key challenges in this evolving field. By improving the accuracy, efficiency, and reliability of MLFFs through symmetry-aware design and rigorous validation, particularly for interfacial systems, this work helps bridge the gap between atomistic modeling and functional understanding of semiconductor materials.

The quantitative demonstration that dispersion physics controls both structural dynamics and electronic properties of halide perovskite surfaces suggests that many previous computational studies may have systematically underestimated band gaps and framework flexibility at interfaces. Similarly, the identification of persistent error heterogeneity across chemical environments provides a roadmap for targeted improvements in next-generation MLFFs. The orbit-based symmetry method established here represents a practical step toward universal MLFFs capable of accurately modeling the diverse chemical environments of real devices.

These advances bring us closer to the goal of computationally guided design of next-generation electronic and optoelectronic devices, where accurate modeling of interfaces, defects, and finite-temperature effects is essential for predicting and optimizing performance. As MLFFs continue to evolve toward greater accuracy, efficiency, and reliability, their integra-

tion into multiscale modeling workflows will enable increasingly predictive simulations of complex semiconductor systems under realistic operating conditions.

References

[1] S. M. Sze and K. K. Ng. *Physics of Semiconductor Devices*. 3rd ed. Wiley, 2006.

[2] B. G. Streetman and S. Banerjee. *Solid State Electronic Devices*. 7th ed. Pearson, 2015.

[3] J. Xue, R. Wang, and Y. Yang. "The Surface of Halide Perovskites from Nano to Bulk". In: *Nat. Rev. Mater.* 5 (2020), pp. 809–827. DOI: 10.1038/s41578-020-0221-1.

[4] A. K. Geim and I. V. Grigorieva. "Van der Waals Heterostructures". In: *Nature* 499.7459 (2013), pp. 419–425. DOI: 10.1038/nature12385.

[5] A. Castellanos-Gomez, X. Duan, Z. Fei, H. R. Gutierrez, Y. Huang, X. Huang, et al. "Van der Waals Heterostructures". In: *Nat. Rev. Methods Primers* 2.1 (2022), p. 58. DOI: 10.1038/s43586-022-00139-1.

[6] L. K. Ono and Y. Qi. "Surface and Interface Aspects of Organometal Halide Perovskite Materials and Solar Cells". In: *J. Phys. Chem. Lett.* 7.22 (2016), pp. 4764–4794. DOI: 10.1021/acs.jpclett.6b01951.

[7] S. D. Stranks, G. E. Eperon, G. Grancini, C. Menelaou, M. J. P. Alcocer, T. Leijtens, L. M. Herz, A. Petrozza, and H. J. Snaith. "Electron-Hole Diffusion Lengths Exceeding 1 Micrometer in an Organometal Trihalide Perovskite Absorber". In: *Science* 342.6156 (2013), pp. 341–344. DOI: 10.1126/science.1243982.

[8] Q. Dong, Y. Fang, Y. Shao, P. Mulligan, J. Qiu, L. Cao, Y. Wang, and J. Huang. "Electron-hole Diffusion Lengths >175 μm in Solution-grown $CH_3NH_3PbI_3$ Single Crystals". In: *Science* 347.6225 (2015), pp. 967–970. DOI: 10.1126/science.aaa5760.

[9] M. A. Green, E. D. Dunlop, J. Hohl-Ebinger, M. Yoshita, N. Kopidakis, and X. Hao. "Solar Cell Efficiency Tables (Version 64)". In: *Prog. Photovolt. Res. Appl.* 32.1 (2024), pp. 3–22. DOI: 10.1002/pip.3775.

[10] Y. Cao, V. Fatemi, S. Fang, and et al. "Unconventional Superconductivity in Magic-Angle Graphene Superlattices". In: *Nature* 556.7699 (2018), pp. 43–50. DOI: 10.1038/nature26160.

[11] J. Behler and M. Parrinello. “Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces”. In: *Phys. Rev. Lett.* 98 (14 Apr. 2007), p. 146401. DOI: 10.1103/PhysRevLett.98.146401.

[12] J. Behler. “Atom-Centered Symmetry Functions for Constructing High-Dimensional Neural Network Potential-Energy Surfaces”. In: *The Journal of Chemical Physics* 134.7 (2011), p. 074106. DOI: 10.1063/1.3553717.

[13] J. Behler. “Atom-centered symmetry functions for constructing high-dimensional neural network potentials”. In: *J. Chem. Phys.* 134.7 (2011). DOI: 10.1063/1.3553717.

[14] A. P. Bartók, R. Kondor, and G. Csányi. “On representing chemical environments”. In: *Phys. Rev. B* 87 (18 May 2013), p. 184115. DOI: 10.1103/PhysRevB.87.184115.

[15] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller. “Machine learning of accurate energy-conserving molecular force fields”. In: *Sci. Adv.* 3.5 (2017), e1603015. DOI: 10.1126/sciadv.1603015.

[16] S. Chmiela, H. E Sauceda, I. Poltavsky, and A. Müller K.-R.and Tkatchenko. “sGDML: Constructing accurate and data efficient molecular force fields using machine learning”. In: *Comput. Phys. Commun.* 240 (2019), pp. 38–45. DOI: 10.1016/j.cpc.2019.02.007.

[17] S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko. “Towards exact molecular dynamics simulations with machine-learned force fields”. In: *Nat. Commun.* 9.1 (2018). DOI: 10.1038/s41467-018-06169-2.

[18] H. E. Sauceda, L. E. Gálvez-González, S. Chmiela, L. O. Paz-Borbón, K.-R. Müller, and A. Tkatchenko. “BIGDML—Towards accurate quantum machine learning force fields for materials”. In: *Nat. Commun.* 13.1 (June 2022), p. 3733. DOI: 10.1038/s41467-022-31093-x.

[19] K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, and K.-R. Müller. “SchNet: a continuous-filter convolutional neural network for modeling quantum interactions”. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS’17. Long Beach, California, USA: Curran Associates Inc., 2017, pp. 992–1002.

[20] K. T. Schütt, O. T. Unke, and M. Gastegger. “Equivariant message passing for the prediction of tensorial properties and molecular spectra”. In: *International Conference on Machine Learning*. PMLR. 2021, pp. 9377–9388. DOI: 10.48550/arXiv.2002.08264.

[21] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky. "E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials". In: *Nat. Commun.* 13.1 (2022), p. 2453. DOI: 10.1038/s41467-022-29939-5.

[22] A. Musaelian, S. Batzner, A. Johansson, L. Sun, C. J. Owen, M. Kornbluth, and B. Kozinsky. "Learning local equivariant representations for large-scale atomistic dynamics". In: *Nat. Commun.* 14.1 (2023), p. 579. DOI: 10.1038/s41467-023-36329-y.

[23] J. T. Frank, O. T. Unke, K.-R. Müller, and S. Chmiela. "A Euclidean transformer for fast and stable machine learned force fields". In: *Nat. Commun.* 15.1 (2024), p. 6539. DOI: 10.1038/s41467-024-50620-6.

[24] I. Batatia, S. Batzner, D. P. Kovács, A. Musaelian, G. N. C. Simm, R. Drautz, C. Ortner, B. Kozinsky, and G. Csányi. "The design space of E(3)-equivariant atom-centred interatomic potentials". In: *Nat. Mach. Intell.* 7.1 (Jan. 2025), pp. 56–67. DOI: 10.1038/s42256-024-00956-x.

[25] I. Batatia, D. P. Kovacs, G. N. C. Simm, C. Ortner, and G. Csanyi. "MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields". In: *Adv. Neural Inf. Process. Syst.* Ed. by A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho. Vol. 35. 2022, pp. 11423–11436. DOI: 10.48550/arXiv.2206.07697.

[26] M. Pols, T. Hilpert, I. A. W. Filot, A. C. T. van Duin, S. Calero, and S. Tao. "What Happens at Surfaces and Grain Boundaries of Halide Perovskites: Insights from Reactive Molecular Dynamics Simulations of $CsPbI_3$". In: *ACS Appl. Mater. Interfaces* 14.37 (2022), pp. 42471–42482. DOI: 10.1021/acsami.2c09239.

[27] A. Marronnier, G. Roma, S. Boyer-Richard, L. Pedesseau, J.-M. Jancu, Y. Bonnassieux, C. Katan, C. C. Stoumpos, M. G. Kanatzidis, and J. Even. "Anharmonicity and Disorder in the Black Phases of Cesium Lead Iodide Used for Stable Inorganic Perovskite Solar Cells". In: *ACS Nano* 12.4 (2018), pp. 3477–3486. DOI: 10.1021/acsnano.8b00267.

[28] N. Yantara and N. Mathews. "Toolsets for Assessing Ionic Migration in Halide Perovskites". In: *Joule* 8.5 (2024), pp. 1239–1273. DOI: 10.1016/j.joule.2024.02.010.

[29] A. N. Arber, Vikram, F. C. Mocanu, and M. S. Islam. "Ion Migration and Dopant Effects in the Gamma-$CsPbI_3$ Perovskite Photovoltaic Material: Atomistic Insights through *Ab Initio* and Machine Learning Methods". In: *Chem. Mater.* 37.12 (2025), pp. 4416–4424. DOI: 10.1021/acs.chemmater.5c00503.

[30] D. A. Egger and L. Kronik. “Role of Dispersive Interactions in Determining Structural Properties of Organic-Inorganic Halide Perovskites: Insights from First-Principles Calculations”. In: *J. Phys. Chem. Lett.* 5.15 (2014), pp. 2728–2733. DOI: 10.1021/jz5012934.

[31] H. Beck, C. Gehrmann, and D. A. Egger. “Structure and Binding in Halide Perovskites: Analysis of Static and Dynamic Effects from Dispersion-Corrected Density Functional Theory”. In: *APL Mater.* 7.2 (2019), p. 021108. DOI: 10.1063/1.5086541.

[32] J. Hermann, R. A. DiStasio Jr., and A. Tkatchenko. “First-Principles Models for van der Waals Interactions in Molecules and Materials: Concepts, Theory, and Applications”. In: *Chem. Rev.* 117.6 (2017), pp. 4714–4758. DOI: 10.1021/acs.chemrev.6b00446.

[33] M. Stöhr, T. Van Voorhis, and A. Tkatchenko. “Theory and Practice of Modeling van der Waals Interactions in Electronic-Structure Calculations”. In: *Chem. Soc. Rev.* 48.13 (2019), pp. 4118–4154. DOI: 10.1039/C8CS00963D.

[34] E. Fransson, J. Wiktor, and P. Erhart. “Phase Transitions in Inorganic Halide Perovskites from Machine-Learned Potentials”. In: *J. Phys. Chem. C.* 127 (2023), pp. 13773–13781. DOI: 10.1021/acs.jpcc.3c01542.

[35] O. T. Unke and M. Meuwly. “PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges”. In: *J. Chem. Theory Comput.* 15.6 (2019), pp. 3678–3693. DOI: 10.1021/acs.jctc.9b00181.

[36] O. T. Unke, S. Chmiela, M. Gastegger, K. T. Schütt, H. E. Sauceda, A. Tkatchenko, and K.-R. Müller. “SpookyNet: Learning Force Fields with Electronic Degrees of Freedom and Nonlocal Effects”. In: *Nat. Commun.* 12.7273 (2021), pp. 1–14. DOI: 10.1038/s41467-021-27504-0.

[37] J. Westermayr, D. Achenbach, P. Marquetand, and R. J. Maurer. “Long-Range Dispersion-Inclusive Machine Learning Potentials for Structure Search and Optimization at Metal-Organic Interfaces”. In: *Digit. Discov.* 1.3 (2022), pp. 243–259. DOI: 10.1039/D2DD00016D.

[38] E. M. Luks. “Isomorphism of graphs of bounded valence can be tested in polynomial time”. In: *J. Comput. Syst. Sci.* 25.1 (1982), pp. 42–65. DOI: 10.1016/0022-0000(82)90009-5.

[39] A. Merkys, A. Vaitkus, A. Grybauskas, A. Konovalovas, M. Quirós, and S. Gražulis. “Graph isomorphism-based algorithm for cross-checking chemical and crystallographic descriptions”. In: *J. Cheminform.* 15.1 (Feb. 2023), p. 25. DOI: 10.1186/s13321-023-00692-1.

[40] P. Hohenberg and W. Kohn. “Inhomogeneous Electron Gas”. In: *Phys. Rev.* 136 (1964), B864–B871. DOI: 10.1103/PhysRev.136.B864.

[41] W. Kohn and L. J. Sham. “Self-Consistent Equations Including Exchange and Correlation Effects”. In: *Phys. Rev.* 140.4A (1965), A1133–A1138. DOI: 10.1103/PhysRev.140.A1133.

[42] J. Hoja, H.-Y. Ko, M. A. Neumann, R. Car, J. DiStasio Robert A., and A. Tkatchenko. “Reliable and practical computational description of molecular crystal polymorphs”. In: *Sci Adv.* 5.1 (2019), eaau3338. DOI: 10.1126/sciadv.aau3338.

[43] S. Zuluaga, P. Canepa, K. Tan, Y. J. Chabal, and T. Thonhauser. “Study of van der Waals bonding and interactions in metal organic framework materials”. In: *J. Phys.: Condens. Matter* 26.13 (2014), p. 133002. DOI: 10.1088/0953-8984/26/13/133002.

[44] R. J. Maurer, V. G. Ruiz, and A. Tkatchenko. “Many-body dispersion effects in the binding of adsorbates on metal surfaces”. In: *J. Chem. Phys.* 143.10 (2015), p. 102808. DOI: 10.1063/1.4922688.

[45] T. Björkman, A. Gulans, A. V. Krasheninnikov, and R. M. Nieminen. “van der Waals Bonding in Layered Compounds from Advanced Density-Functional First-Principles Calculations”. In: *Phys. Rev. Lett.* 108.23 (2012), p. 235502. DOI: 10.1103/PhysRevLett.108.235502.

[46] R. G. Parr and W. Yang. Vol. 16. International Series of Monographs on Chemistry. New York, NY: Oxford University Press, 1995, p. 352. DOI: 10.1093/oso/9780195092769.002.0001.

[47] R. O. Jones. “Density functional theory: Its origins, rise to prominence, and future”. In: *Rev. Mod. Phys.* 87 (2015), pp. 897–923. DOI: 10.1103/RevModPhys.87.897.

[48] D. M. Ceperley and B. J. Alder. “Ground State of the Electron Gas by a Stochastic Method”. In: *Phys. Rev. Lett.* 45 (1980), pp. 566–569. DOI: 10.1103/PhysRevLett.45.566.

[49] J. P. Perdew and A. Zunger. “Self-interaction Correction to Density-Functional Approximations for Many-Electron Systems”. In: *Phys. Rev. B* 23 (1981), pp. 5048–5079. DOI: 10.1103/PhysRevB.23.5048.

[50] J. P. Perdew and W. Yue. “Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation”. In: *Phys. Rev. B* 33.12 (1986), pp. 8800–8802. DOI: 10.1103/PhysRevB.33.8800.

[51] A. D. Becke. “Density-functional exchange-energy approximation with correct asymptotic behavior”. In: *Phys. Rev. A* 38.6 (1988), pp. 3098–3100. DOI: 10.1103/PhysRevA.38.3098.

[52] J. P. Perdew, K. Burke, and M. Ernzerhof. “Generalized Gradient Approximation Made Simple”. In: *Phys. Rev. Lett.* 77.18 (1996), pp. 3865–3868. DOI: 10.1103/PhysRevLett.77.3865.

[53] C. Lee, W. Yang, and R. G. Parr. “Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density”. In: *Phys. Rev. B* 37.2 (1988), pp. 785–789. DOI: 10.1103/PhysRevB.37.785.

[54] J. P. Perdew and K. Schmidt. “Jacob’s ladder of density functional approximations for the exchange-correlation energy”. In: *AIP Conference Proceedings* 577.1 (July 2001), pp. 1–20. DOI: 10.1063/1.1390175.

[55] J. Sun, A. Ruzsinszky, and J. P. Perdew. “Strongly Constrained and Appropriately Normed Semilocal Density Functional”. In: *Phys. Rev. Lett.* 115 (2015), p. 036402. DOI: 10.1103/PhysRevLett.115.036402.

[56] C. Adamo and V. Barone. “Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model”. In: *J. Chem. Phys.* 110 (1999), pp. 6158–6170. DOI: 10.1063/1.478522.

[57] S. Grimme. “Semiempirical Hybrid Density Functional with Perturbative Second-Order Correlation”. In: *J. Chem. Phys.* 124 (2006), p. 034108. DOI: 10.1063/1.2148954.

[58] L. Goerigk and S. Grimme. “Double-Hybrid Density Functionals”. In: *WIREs Comput. Mol. Sci.* 4 (2014), pp. 576–600. DOI: 10.1002/wcms.1193.

[59] J. Toulouse. “Review of Approximations for the Exchange-Correlation Energy in Density-Functional Theory”. In: *Density Functional Theory: Modeling, Mathematical Analysis, Computational Methods, and Applications*. Ed. by E. Cancès and G. Friesecke. Cham: Springer International Publishing, 2023, pp. 1–90. DOI: 10.1007/978-3-031-22340-2_1.

[60] E. H. Lieb and S. Oxford. “Improved Lower Bound on the Indirect Coulomb Energy”. In: *Int. J. Quantum Chem.* 19 (1981), pp. 427–439. DOI: 10.1002/qua.560190814.

[61] Y. Wang and J. P. Perdew. “Spin scaling of the electron-gas correlation energy in the high-density limit”. In: *Phys. Rev. B* 43 (1991), pp. 8911–8914. DOI: 10.1103/PhysRevB.43.8911.

[62] M. Levy. “Density-functional exchange correlation through coordinate scaling in adiabatic connection and correlation hole”. In: *Phys. Rev. A* 43 (1991), pp. 4637–4643. DOI: 10.1103/PhysRevA.43.4637.

[63] D. C. Langreth and J. P. Perdew. “The exchange-correlation energy of a metallic surface”. In: *Physical Review B* 21.10 (1980), pp. 5469–5493. DOI: 10.1103/PhysRevB.21.5469.

[64] D. S. Sholl and J. A. Steckel. *Density Functional Theory: A Practical Introduction*. John Wiley & Sons, 2009. DOI: 10.1002/9780470742819.

[65] J. Behler. “Perspective: Machine learning potentials for atomistic simulations”. In: *J. Chem. Phys.* 145 (2016), p. 170901. DOI: 10.1063/1.4966192.

[66] J. Even, L. Pedesseau, J.-M. Jancu, and C. Katan. “Importance of Spin-Orbit Coupling in Hybrid Organic/Inorganic Perovskites for Photovoltaic Applications”. In: *J. Phys. Chem. Lett.* 4.7 (2013), pp. 2999–3005. DOI: 10.1021/jz401532q.

[67] F. Brivio, K. T. Butler, A. Walsh, and M. van Schilfgaarde. “Relativistic Quasiparticle Self-Consistent Electronic Structure of Hybrid Halide Perovskite Photovoltaic Absorbers”. In: *Phys. Rev. B* 89.15 (2014), p. 155204. DOI: 10.1103/PhysRevB.89.155204.

[68] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist. “Van der Waals density functional for general geometries”. In: *Phys. Rev. Lett.* 92.24 (2004), p. 246401. DOI: 10.1103/PhysRevLett.92.246401.

[69] J. Klimeš, D. R. Bowler, and A. Michaelides. “Chemical accuracy for the van der Waals density functional”. In: *J. Phys.: Condens. Matter* 22.2 (2010), p. 022201. DOI: 10.1088/0953-8984/22/2/022201.

[70] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. “A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu”. In: *J. Chem. Phys.* 132.15 (2010), p. 154104. DOI: 10.1063/1.3382344.

[71] S. Grimme, S. Ehrlich, and L. Goerigk. “Effect of the damping function in dispersion corrected density functional theory”. In: *J. Comput. Chem.* 32.7 (2011), pp. 1456–1465. DOI: 10.1002/jcc.21759.

[72] S. Ehrlich, J. Möllmann, W. Reckien, and T. Bredow. “System-Dependent Dispersion Coefficients for the DFT-D3 Treatment of Adsorption Processes on Ionic Surfaces”. In: *ChemPhysChem* 12.17 (2011), pp. 3414–3420. DOI: 10.1002/cphc.201100521.

[73] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, and S. Grimme. "A generally applicable atomic-charge dependent London dispersion correction: The DFT-D4 model". In: *J. Chem. Phys.* 150.15 (2019), p. 154122. DOI: 10.1063/1.5090222.

[74] O. A. Vydrov and T. Van Voorhis. "Nonlocal van der Waals density functional: The simpler the better". In: *J. Chem. Phys.* 133.24 (2010), p. 244103. DOI: 10.1063/1.3521275.

[75] A. Tkatchenko, R. A. J. DiStasio, R. Car, and M. Scheffler. "Accurate and efficient method for many-body van der Waals interactions". In: *Phys. Rev. Lett.* 108 (2012), p. 236402. DOI: 10.1103/PhysRevLett.108.236402.

[76] J. Hermann and A. Tkatchenko. "Density functional model for van der Waals interactions: Unifying many-body atomic approaches with nonlocal functionals". In: *Phys. Rev. Lett.* 124 (2020), p. 146401. DOI: 10.1103/PhysRevLett.124.146401.

[77] O. A. Vydrov and T. Van Voorhis. "Nonlocal van der Waals Density Functional Made Simple". In: *Phys. Rev. Lett.* 103 (6 Aug. 2009), p. 063004. DOI: 10.1103/PhysRevLett.103.063004.

[78] C. Gutle, A. Savin, J. B. Krieger, and J. Chen. "Correlation energy contributions from low-lying states to density functionals based on an electron gas with a gap". In: *Int. J. Quantum Chem.* 75.4-5 (1999), pp. 885–888. DOI: https://doi.org/10.1002/(SICI)1097-461X(1999)75:4/5<885::AID-QUA53>3.0.CO;2-F.

[79] A. D. Becke and E. R. Johnson. "Exchange-hole dipole moment and the dispersion interaction revisited". In: *J. Chem. Phys.* 127.15 (Oct. 2007), p. 154108. DOI: 10.1063/1.2795701.

[80] S. Künmmel and J. P. Perdew. "Two avenues to self-interaction correction within Kohn—Sham theory: unitary invariance is the shortcut". In: *Mol. Phys.* 101.9 (2003), pp. 1363–1368. DOI: 10.1080/0026897031000094506.

[81] J. Sun, B. Xiao, Y. Fang, R. Haunschild, P. Hao, A. Ruzsinszky, G. I. Csonka, G. E. Scuseria, and J. P. Perdew. "Density Functionals that Recognize Covalent, Metallic, and Weak Bonds". In: *Phys. Rev. Lett.* 111 (10 Sept. 2013), p. 106401.

[82] T. Gould, S. Lebègue, J. G. Ángyán, and T. Bučko. "A fractionally ionic approach to polarizability and van der Waals many-body dispersion calculations". In: *J. Chem. Theory Comput.* 12 (2016), pp. 5920–5930. DOI: 10.1021/acs.jctc.6b00925.

[83] L. Hörmann, A. Jeindl, and O. T. Hofmann. "Reproducibility of Potential Energy Surfaces of Organic/Metal Interfaces on the Example of PTCDA on Ag(111)". In: *J. Chem. Phys.* 156.20 (2022), p. 206101. DOI: 10.1063/5.0094163.

[84] B. Emrem, R. Kempt, K. Finzel, and T. Heine. "London Dispersion-Corrected Density Functionals Applied to van der Waals Stacked Layered Materials: Validation of Structure, Energy, and Electronic Properties". In: *Adv. Theory Simul.* 5.7 (2022), p. 2200055. DOI: 10.1002/adts.202200055.

[85] J. Hermann, M. Stöhr, S. Göger, S. Chaudhuri, B. Aradi, R. J. Maurer, and A. Tkatchenko. "libMBD: A General-Purpose Package for Scalable Quantum Many-Body Dispersion Calculations". In: *J. Chem. Phys.* 159 (2023), p. 174802. DOI: 10.1063/5.0170972.

[86] B. Sohail, P. J. Blowey, L. A. Rochford, P. T. P. Ryan, D. A. Duncan, T.-L. Lee, P. Starrs, G. Costantini, D. P. Woodruff, and R. J. Maurer. "Donor-Acceptor Co-Adsorption Ratio Controls the Structure and Electronic Properties of Two-Dimensional Alkali-Organic Networks on Ag(100)". In: *J. Phys. Chem. C.* 127.5 (2023), pp. 2716–2727. DOI: 10.1021/acs.jpcc.2c08688.

[87] S. Banerjee and A. Tkatchenko. "Non-local Interactions Determine Local Structure and Lithium Diffusion in Solid Electrolytes". In: *Nat. Commun.* 16 (2025), p. 1672. DOI: 10.1038/s41467-025-56662-8.

[88] I. Poltavsky et al. "Crash testing machine learning force fields for molecules, materials, and interfaces: model analysis in the TEA Challenge 2023". In: *Chem. Sci.* 16 (8 2025), pp. 3720–3737. DOI: 10.1039/D4SC06529H.

[89] I. Poltavsky et al. "Crash testing machine learning force fields for molecules, materials, and interfaces: molecular dynamics in the TEA challenge 2023". In: *Chem. Sci.* 16 (8 2025), pp. 3738–3754. DOI: 10.1039/D4SC06530A.

[90] T. Maximova, R. Moffatt, B. Ma, R. Nussinov, and A. Shehu. "Principles and Overview of Sampling Methods for Modeling Macromolecular Structure and Dynamics". In: *PLoS Comput. Biol.* 12.4 (2016), e1004619. DOI: 10.1371/journal.pcbi.1004619.

[91] R. Car and M. Parrinello. "Unified Approach for Molecular Dynamics and Density-Functional Theory". In: *Phys. Rev. Lett.* 55.22 (1985), pp. 2471–2474. DOI: 10.1103/PhysRevLett.55.2471.

[92] J. Tersoff. "New empirical approach for the structure and energy of covalent systems". In: *Phys. Rev. B* 37.12 (1988), pp. 6991–7000. DOI: 10.1103/PhysRevB.37.6991.

[93] H. C. Andersen. “Molecular dynamics simulations at constant pressure and/or temperature”. In: *J. Chem. Phys.* 72.4 (1980), pp. 2384–2393. DOI: 10.1063/1.440845.

[94] M. Parrinello and A. Rahman. “Polymorphic transitions in single crystals: A new molecular dynamics method”. In: *J. Appl. Phys.* 52.12 (1981), pp. 7182–7190. DOI: 10.1063/1.328693.

[95] S. Nosé. “A unified formulation of the constant temperature molecular dynamics methods”. In: *J. Chem. Phys.* 81.1 (1984), pp. 511–519. DOI: 10.1063/1.447334.

[96] W. G. Hoover. “Canonical dynamics: Equilibrium phase-space distributions”. In: *Phys. Rev. A* 31.3 (1985), pp. 1695–1697. DOI: 10.1103/PhysRevA.31.1695.

[97] W. G. Hoover. “Constant-pressure equations of motion”. In: *Phys. Rev. A* 34 (3 Sept. 1986), pp. 2499–2500. DOI: 10.1103/PhysRevA.34.2499.

[98] J. G. Kirkwood. “Statistical Mechanics of Fluid Mixtures”. In: *J. Chem. Phys.* 3.5 (1935), pp. 300–313. DOI: 10.1063/1.1749657.

[99] T. P. Straatsma and H. J. C. Berendsen. “Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations”. In: *J. Chem. Phys.* 89.9 (1988), pp. 5876–5886. DOI: 10.1021/j100322a007.

[100] E. Darve. “Thermodynamic Integration Using Constrained and Unconstrained Dynamics”. In: *Free Energy Calculations*. Ed. by C. Chipot and A. Pohorille. Vol. 86. Springer Series in Chemical Physics. Springer, Berlin, Heidelberg, 2007, pp. 195–226. DOI: 10.1007/978-3-540-38448-9_4.

[101] L. B. Pártay, A. P. Bartók, and G. Csányi. “Efficient Sampling of Atomic Configurational Spaces”. In: *J. Phys. Chem. B.* 114.32 (2010). PMID: 20701382, pp. 10502–10512. DOI: 10.1021/jp1012973.

[102] H. Bhatia et al. “Machine-learning-based dynamic-importance sampling for adaptive multiscale simulations”. In: *Nat. Mach. Intell.* 3.5 (May 2021), pp. 401–409. DOI: 10.1038/s42256-021-00327-w.

[103] B. J. Alder and T. E. Wainwright. “Studies in Molecular Dynamics. I. General Method”. In: *J. Chem. Phys.* 31.2 (1959), pp. 459–466. DOI: 10.1063/1.1730376.

[104] L. Verlet. “Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules”. In: *Phys. Rev.* 159.1 (1967), pp. 98–103. DOI: 10.1103/PhysRev.159.98.

[105] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, 2017. DOI: 10.1093/oso/9780198803195.001.0001.

[106] D. Marx and J. Hutter. "Ab Initio Molecular Dynamics: Theory and Implementation". In: *Angew. Chem. Int. Ed.* 48.41 (2009), pp. 6986–6998. DOI: 10.1002/anie.200904748.

[107] A. Rahman. "Correlations in the Motion of Atoms in Liquid Argon". In: *Phys. Rev.* 136 (1964), A405–A411. DOI: 10.1103/PhysRev.136.A405.

[108] J. H. Irving and J. G. Kirkwood. "The Statistical Mechanical Theory of Transport Processes. IV. The Equations of Hydrodynamics". In: *J. Chem. Phys.* 18 (1950), pp. 817–829. DOI: 10.1063/1.1747782.

[109] J. A. McCammon, B. R. Gelin, and M. Karplus. "Dynamics of Folded Proteins". In: *Nature* 267 (1977), pp. 585–590. DOI: 10.1038/267585a0.

[110] M. Karplus and J. A. McCammon. "Molecular Dynamics Simulations of Biomolecules". In: *Nat. Struct. Biol.* 9 (2002), pp. 646–652. DOI: 10.1038/nsb0902-646.

[111] M. E. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. Oxford University Press, 2023. DOI: 10.1093/oso/9780198825562.001.0001.

[112] M. E. Tuckerman and G. J. Martyna. "Understanding Modern Molecular Dynamics: Techniques and Applications". In: *J. Phys. Chem. B.* 104.2 (2000), pp. 159–178. DOI: 10.1021/jp992433y.

[113] B. Leimkuhler and S. Reich. *Simulating Hamiltonian Dynamics*. Vol. 14. Cambridge Monographs on Applied and Computational Mathematics. Cambridge: Cambridge University Press, 2004. DOI: 10.1017/CBO9780511614118.

[114] J. M. Sanz-Serna. "Symplectic integrators for Hamiltonian problems: an overview". In: *Acta Numer.* 1 (1992), pp. 243–286. DOI: 10.1017/S0962492900002282.

[115] P. Minary, M. E. Tuckerman, and G. J. Martyna. "Dynamical Spatial Warping: A Novel Method for the Conformational Sampling of Biophysical Structure". In: *SIAM J. Sci. Comput.* 30.4 (2008), pp. 2055–2083. DOI: 10.1137/070686706.

[116] E. Braun, J. Gilmer, H. B. Mayes, D. L. Mobley, J. I. Monroe, S. Prasad, and D. M. Zuckerman. "Best Practices for Foundations in Molecular Simulations [Article v1.0]". In: *Living Journal of Computational Molecular Science* 1.1 (2019). Epub 2018-11-29, p. 5957. DOI: 10.33011/livecoms.1.1.5957.

[117] W. Chu, Q. Zheng, O. V. Prezhdo, J. Zhao, and W. A. Saidi. "Low-frequency lattice phonons in halide perovskites explain high defect tolerance toward electron–hole recombination". In: *Sci. Adv.* 6.7 (2020), eaaw7453. DOI: 10.1126/sciadv.aaw7453.

[118] O. Valsson, P. Tiwary, and M. Parrinello. "Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint". In: *Annu. Rev. Phys. Chem.* 67 (2016), pp. 159–184. DOI: 10.1146/annurev-physchem-040215-112229.

[119] G. M. Torrie and J. P. Valleau. "Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling". In: *J. Comput. Phys.* 23.2 (1977), pp. 187–199. DOI: 10.1016/0021-9991(77)90121-8.

[120] Y. Sugita and Y. Okamoto. "Replica-Exchange Molecular Dynamics Method for Protein Folding". In: *Chem. Phys. Lett.* 314.1-2 (1999), pp. 141–151. DOI: 10.1016/S0009-2614(99)01123-9.

[121] A. Laio and M. Parrinello. "Escaping Free-Energy Minima". In: *PNAS USA* 99.20 (2002), pp. 12562–12566. DOI: 10.1073/pnas.202427399.

[122] A. F. Voter. "Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events". In: *Phys. Rev. Lett.* 78 (1997), pp. 3908–3911. DOI: 10.1103/PhysRevLett.78.3908.

[123] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller. "Machine Learning Force Fields". In: *Chem. Rev.* 121.16 (2021), pp. 10142–10186. DOI: 10.1021/acs.chemrev.0c01111.

[124] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, 2007. DOI: 10.1063/1.881812.

[125] G. J. Martyna, M. L. Klein, and M. Tuckerman. "Nosé-Hoover Chains: The Canonical Ensemble via Continuous Dynamics". In: *J. Chem. Phys.* 97.4 (1992), pp. 2635–2643. DOI: 10.1063/1.463940.

[126] P. K. Patra and B. Bhattacharya. "Non-Ergodicity of Nosé-Hoover Chain Thermostat in Computationally Achievable Time". In: *Phys. Rev. E.* 90.4 (2014), p. 043304. DOI: 10.1103/PhysRevE.90.043304.

[127] J. A. Izaguirre, D. P. Catarello, J. M. Wozniak, and R. D. Skeel. "Langevin Stabilization of Molecular Dynamics". In: *J. Chem. Phys.* 114.5 (2001), pp. 2090–2098. DOI: 10.1063/1.1332569.

[128] B. Leimkuhler and C. Matthews. "Robust and Efficient Configurational Molecular Sampling via Langevin Dynamics". In: *J. Chem. Phys.* 138.17 (2013), p. 174102. DOI: 10.1063/1.4804167.

[129] M. Ceriotti, G. Bussi, and M. Parrinello. "Langevin Equation with Colored Noise for Constant-Temperature Molecular Dynamics Simulations". In: *Phys. Rev. Lett.* 102.2 (2009), p. 020601. DOI: 10.1103/PhysRevLett.102.020601.

[130] M. Ceriotti, G. Bussi, and M. Parrinello. “Colored-Noise Thermostats à la Carte”. In: *J. Chem. Theory Comput.* 6.4 (2010), pp. 1170–1180. DOI: 10.1021/ct900563s.

[131] S. Melchionna, G. Ciccotti, and B. L. Holian. “Hoover NPT dynamics for systems varying in shape and size”. In: *Molecular Physics* 78.3 (1993), pp. 533–544. DOI: 10.1080/00268979300100371.

[132] G. J. Martyna, D. J. Tobias, and M. L. Klein. “Constant pressure molecular dynamics algorithms”. In: *J. Chem. Phys.* 101.5 (1994), pp. 4177–4189. DOI: 10.1063/1.467468.

[133] S. Nosé. “A molecular dynamics method for simulations in the canonical ensemble”. In: *Molecular Physics* 52.2 (1984), pp. 255–268. DOI: 10.1080/00268978400101201.

[134] C. Chipot and A. Pohorille, eds. *Free Energy Calculations: Theory and Applications in Chemistry and Biology*. Vol. 86. Springer Series in Chemical Physics. Berlin, Heidelberg: Springer, 2007. DOI: 10.1007/978-3-540-38448-9.

[135] H. Resat and M. Mezei. “Studies on free energy calculations. I. Thermodynamic integration using a polynomial path”. In: *J. Chem. Phys.* 99.8 (1993), pp. 6052–6061. DOI: 10.1063/1.465111.

[136] M. Jorge, N. Garrido, A. Queimada, I. Economou, and E. Macedo. “Effect of the integration method on the accuracy and computational efficiency of free energy calculations using thermodynamic integration”. In: *J. Chem. Theory Comput.* 6.4 (2010), pp. 1018–1027. DOI: 10.1021/ct900661y.

[137] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi. “Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons”. In: *Phys. Rev. Lett.* 104 (13 Apr. 2010), p. 136403. DOI: 10.1103/PhysRevLett.104.136403.

[138] V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, and G. Csányi. “Gaussian Process Regression for Materials and Molecules”. In: *Chem. Rev.* 121.16 (2021), pp. 10073–10141. DOI: 10.1021/acs.chemrev.1c00022.

[139] A. S. Christensen, L. A. Bratholm, F. A. Faber, and O. A. von Lilienfeld. “FCHL revisited: Faster and more accurate quantum machine learning”. In: *J. Chem. Phys.* 152.4 (2020). DOI: 10.1063/1.5126701.

[140] N. J. Browning, F. A. Faber, and O. Anatole von Lilienfeld. “GPU-accelerated approximate kernel method for quantum machine learning”. In: *J. Chem. Phys.* 157.21 (Dec. 2022), p. 214801. DOI: 10.1063/5.0108967.

[141] Z. Yang, X. Wang, Y. Li, Q. Lv, C. Y.-C. Chen, and L. Shen. "Efficient equivariant model for machine learning interatomic potentials". In: *npj Comput Mater.* 11.1 (Feb. 2025), p. 49. DOI: 10.1038/s41524-025-01535-3.

[142] J. T. Frank, O. T. Unke, and K.-R. Müller. "So3krates: Equivariant attention for interactions on arbitrary length-scales in molecular systems". In: *Adv. Neural Inf. Process. Syst.* Ed. by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh. Vol. 35. 2022, pp. 29400–29413. DOI: 10.48550/arXiv.2205.14276.

[143] A. Kabylda, J. T. Frank, S. Suárez-Dou, A. Khabibrakhmanov, L. Medrano Sandonas, O. T. Unke, S. Chmiela, K.-R. Müller, and A. Tkatchenko. "Molecular Simulations with a Pretrained Neural Network and Universal Pairwise Force Fields". In: *J. Am. Chem. Soc.* 147.37 (Sept. 2025), pp. 33723–33734. DOI: 10.1021/jacs.5c09558.

[144] D. P. Kovács, J. H. Moore, N. J. Browning, I. Batatia, J. T. Horton, V. Kapil, I.-B. Magdău, D. J. Cole, and G. Csányi. "MACE-OFF23: Transferable machine learning force fields for organic molecules". In: *arXiv* (2023 [cited 2024 Aug 24]). DOI: 10.48550/arXiv.2312.15211.

[145] I. Batatia et al. "A foundation model for atomistic materials chemistry". In: *arXiv* (2023 [cited 2024 Aug 24]). DOI: 10.48550/arXiv.2401.00096.

[146] A. R. Leach. *Molecular Modelling: Principles and Applications*. 2nd. Prentice Hall, 2001.

[147] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules". In: *J. Am. Chem. Soc.* 117.19 (1995), pp. 5179–5197. DOI: 10.1021/ja955032e.

[148] P. M. Morse. "Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels". In: *Phys. Rev.* 34 (1 July 1929), pp. 57–64. DOI: 10.1103/PhysRev.34.57.

[149] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen. "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes". In: *J. Comput. Phys.* 23.3 (1977), pp. 327–341. DOI: 10.1016/0021-9991(77)90098-5.

[150] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, June 2017. DOI: 10.1093/oso/9780198803195.001.0001.

[151] J. E. Jones and S. Chapman. "On the determination of molecular fields. —II. From the equation of state of a gas". In: *Proc. R. Soc. Lond. Ser. A-Contain. Pap. Math. Phys. Character* 106.738 (1924), pp. 463–477. DOI: 10.1098/rspa.1924.0082.

[152] G. Lamoureux and B. Roux. “Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm”. In: *J. Chem. Phys.* 119.6 (Aug. 2003), pp. 3025–3039. DOI: 10.1063/1.1589749.

[153] A. D. J. MacKerell et al. “All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins”. In: *J. Phys. Chem. B.* 102.18 (Apr. 1998), pp. 3586–3616. DOI: 10.1021/jp973084f.

[154] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. G. III. “ReaxFF: A Reactive Force Field for Hydrocarbons”. In: *J. Phys. Chem. A* 105.41 (2001), pp. 9396–9409. DOI: 10.1021/jp004368u.

[155] K. Chenoweth, A. C. T. van Duin, and W. A. G. III. “ReaxFF Reactive Force Field for Molecular Dynamics Simulations of Hydrocarbon Oxidation”. In: *J. Phys. Chem. A* 112.5 (2008), pp. 1040–1053. DOI: 10.1021/jp709896w.

[156] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Nov. 2005. DOI: 10.7551/mitpress/3206.001.0001.

[157] J. B. Kruskal. “On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem”. In: *Proc. Amer. Math. Soc.* 7.1 (1956), pp. 48–50. DOI: 10.1090/S0002-9939-1956-0078686-7.

[158] R. C. Prim. “Shortest Connection Networks and Some Generalizations”. In: *Bell System Technical Journal* 36.6 (1957), pp. 1389–1401. DOI: 10.1002/j.1538-7305.1957.tb01515.x.

[159] J. R. Shewchuk. *An Introduction to the Conjugate Gradient Method Without the Agonizing Pain*. Tech. rep. USA, 1994.

[160] K. Wang, G. Pleiss, J. Gardner, S. Tyree, K. Q. Weinberger, and A. G. Wilson. “Exact Gaussian Processes on a Million Data Points”. In: *Adv. Neural Inf. Process. Syst.* Ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett. Vol. 32. Curran Associates, Inc., 2019. DOI: 10.48550/arXiv.1903.08114.

[161] Y. Saad. *Iterative methods for sparse linear systems*. SIAM, 2003.

[162] A. Rahimi and B. Recht. “Random Features for Large-Scale Kernel Machines”. In: *Adv. Neural Inf. Process. Syst.* Ed. by J. Platt, D. Koller, Y. Singer, and S. Roweis. Vol. 20. Curran Associates, Inc., 2007.

[163] J. S. Smith, O. Isayev, and A. E. Roitberg. “ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost”. In: *Chem. Sci.* 8.4 (2017), pp. 3192–3203. DOI: 10.1039/C6SC05720A.

[164] L. Zhang, J. Han, H. Wang, R. Car, and W. E. "Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics". In: *Phys. Rev. Lett.* 120.14 (2018), p. 143001. DOI: 10.1103/PhysRevLett.120.143001.

[165] D. Ocampo, D. Posso, R. Namakian, and W. Gao. "Adaptive loss weighting for machine learning interatomic potentials". In: *Comput. Mater. Sci.* 244 (2024), p. 113155. DOI: 10.1016/j.commatsci.2024.113155.

[166] J. Zeng et al. "DeePMD-kit v3: A Multiple-Backend Framework for Machine Learning Potentials". In: *J. Chem. Theory Comput.* 21.9 (May 2025), pp. 4375–4385. DOI: 10.1021/acs.jctc.5c00340.

[167] V. L. Deringer, M. A. Caro, and G. Csányi. "Machine Learning Interatomic Potentials as Emerging Tools for Materials Science". In: *Advanced Materials* 31.46 (2019), e1902765. DOI: 10.1002/adma.201902765.

[168] D. Milardovich, C. Wilhelmer, D. Waldhör, L. Cvitkovich, G. Sivaraman, and T. Grasser. "Machine Learning Interatomic Potential for Silicon-Nitride (Si_3N_4) by Active Learning". In: *The Journal of Chemical Physics* 158.19 (2023), p. 194802. DOI: 10.1063/5.0146753.

[169] L. Zhang, D.-Y. Lin, H. Wang, R. Car, and W. E. "Active learning of uniformly accurate interatomic potentials for materials simulation". In: *Phys. Rev. Mater.* 3 (2 Feb. 2019), p. 023804. DOI: 10.1103/PhysRevMaterials.3.023804.

[170] R. Jinnouchi, K. Miwa, F. Karsai, G. Kresse, and R. Asahi. "On-the-Fly Active Learning of Interatomic Potentials for Large-Scale Atomistic Simulations". In: *Journal of Physical Chemistry Letters* 11.17 (2020), pp. 6946–6955. DOI: 10.1021/acs.jpclett.0c01061.

[171] M Hodapp and A Shapeev. "In operando active learning of interatomic interaction during large-scale simulations". In: *Machine Learning: Science and Technology* 1.4 (Sept. 2020), p. 045005. DOI: 10.1088/2632-2153/aba373.

[172] Y. Xie, J. Vandermause, S. Ramakers, N. H. Protik, A. Johansson, and B. Kozinsky. "Uncertainty-aware molecular dynamics from Bayesian active learning for phase transformations and thermal transport in SiC". In: *npj Comput. Mater.* 9.1 (Mar. 2023), p. 36. DOI: 10.1038/s41524-023-00988-8.

[173] C. Chen and S. P. Ong. "A Universal Graph Deep Learning Interatomic Potential for the Periodic Table". In: *Nature Computational Science* 2.11 (2022), pp. 718–728. DOI: 10.1038/s43588-022-00349-3.

[174] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. "Neural Message Passing for Quantum Chemistry". In: *Proceedings of the 34th International Conference on Machine Learning*. JMLR, 2017, pp. 1263–1272. DOI: 10.48550/arXiv.1704.01212.

[175] K. T. Schütt, F. Arbabzadah, S. Chmiela, K.-R. Müller, and A. Tkatchenko. "Quantum-chemical insights from deep tensor neural networks". In: *Nat. Commun.* 8 (2017), p. 13890. DOI: 10.1038/ncomms13890.

[176] C. Chen, W. Ye, Y. Zuo, C. Zheng, and S. P. Ong. "Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals". In: *Chem. Mater.* 31.9 (2019), pp. 3564–3572. DOI: 10.1021/acs.chemmater.9b01294.

[177] J. Klicpera, J. Groß, and S. Günnemann. "Directional Message Passing for Molecular Graphs". In: *International Conference on Learning Representations*. 2020. DOI: https://doi.org/10.48550/arXiv.2003.03123.

[178] J. Gasteiger, C. Yeshwanth, and S. Günnemann. "Directional Message Passing on Molecular Graphs via Synthetic Coordinates". In: *Adv. Neural Inf. Process. Syst.* Vol. 34. 2021, pp. 24603–24617. DOI: 10.48550/arXiv.2111.04718.

[179] V. G. Satorras, E. Hoogeboom, and M. Welling. "E(n) Equivariant Graph Neural Networks". In: *Proceedings of the 38th International Conference on Machine Learning*. Vol. 139. 2021, pp. 9323–9332. DOI: 10.48550/arXiv.2102.09844.

[180] J. Behler. "Constructing high-dimensional neural network potentials: A tutorial review". In: *Int. J. Quantum Chem.* 115.16 (2015), pp. 1032–1050. DOI: 10.1002/qua.24890.

[181] M. Weiler, M. Geiger, M. Welling, W. Boomsma, and T. S. Cohen. "3D Steerable CNNs: Learning Rotationally Equivariant Features in Volumetric Data". In: *Adv. Neural Inf. Process. Syst.* Vol. 31. 2018, pp. 10402–10413. DOI: 10.48550/arXiv.1807.02547.

[182] P. Reiser et al. "Graph neural networks for materials science and chemistry". In: *Communications Materials* 3.1 (Nov. 2022), p. 93. DOI: 10.1038/s43246-022-00315-6.

[183] J. Han, Y. Rong, T. Xu, and W. Huang. "Geometrically Equivariant Graph Neural Networks: A Survey". In: *arXiv* (2022 [cited 2024 Aug 24]). DOI: 10.48550/arXiv.2202.07230.

[184] S. Elfwing, E. Uchibe, and K. Doya. "Sigmoid-Weighted Linear Units for Neural Network Function Approximation in Reinforcement Learning". In: (2017). DOI: 10.48550/arXiv.1702.03118.

[185] K. Lejaeghere et al. “Reproducibility in density functional theory calculations of solids”. In: *Science* 351.6280 (2016), aad3000. DOI: 10.1126/science.aad3000.

[186] G. Fonseca, I. Poltavsky, and A. Tkatchenko. “Force Field Analysis Software and Tools (FFAST): Assessing Machine Learning Force Fields under the Microscope”. In: *J. Chem. Theory Comput.* 19.23 (Dec. 2023), pp. 8706–8717. DOI: 10.1021/acs.jctc.3c00985.

[187] C. Battaglia, A. Cuevas, and S. De Wolf. “High-efficiency crystalline silicon solar cells: status and perspectives”. In: *Energy Environ. Sci.* 9 (2016), pp. 1552–1576. DOI: 10.1039/C5EE03380B.

[188] National Renewable Energy Laboratory. “Interactive Best Research-Cell Efficiency Chart”. In: (June 2025). Last updated June 16, 2025.

[189] U.S. Department of Energy, Solar Energy Technologies Office. *Cadmium Telluride Photovoltaics (CdTe) — Perspective Paper.* Tech. rep. Published 2025-01-16. DOE, 2025.

[190] A. Sugie, Y. Shibata, Y. Noguchi, K. Itami, T. Morifuji, M. Otani, and T. Kato. “Dependence of Exciton Binding Energy on Bandgap of Organic Semiconductors”. In: *J. Phys. Chem. Lett.* 14.48 (2023), pp. 10851–10858. DOI: 10.1021/acs.jpclett.3c02863.

[191] H. Sirringhaus. “25th Anniversary Article: Organic Field-Effect Transistors: The Path Beyond Amorphous Silicon”. In: *Adv. Mater.* 26 (2014), pp. 1319–1335. DOI: 10.1002/adma.201304346.

[192] A. Walsh and A. Zunger. “Instilling defect tolerance in new compounds”. In: *Nat. Mater.* 16.10 (2017), pp. 964–967. DOI: 10.1038/nmat4973.

[193] A. Zakutayev. “Brief Review of Emerging Photovoltaic Absorbers”. In: *Curr. Opin. Green Sustain. Chem.* 4 (2017), pp. 8–15. DOI: 10.1016/j.cogsc.2017.01.002.

[194] M. U. Ali, H. Mo, A. U. Rehman, T. L. Leung, and A. B. Djurišić. “Metal halide perovskites: stability under illumination and bias”. In: *Trends Chem.* 6.5 (2024), pp. 248–259. DOI: 10.1016/j.trechm.2024.03.009.

[195] S. Baumann et al. “Stability and reliability of perovskite containing solar cells and modules: degradation mechanisms and mitigation strategies”. In: *Energy Environ. Sci.* 17 (2024), pp. 7566–7599. DOI: 10.1039/D4EE01898B.

[196] M. Remec, M. Khenkin, U. Erdil, Q. Emery, G. Paramasivam, E. Unger, R. Schlatmann, S. Albrecht, M. Topič, and C. Ulbrich. “Seasonality in Perovskite Solar Cells: Insights from 4 Years of Outdoor Data”. In: *Adv. Energy Mater.* 15.35 (2025), p. 2501906. DOI: 10.1002/aenm.202501906.

[197] S. De Wolf, J. Holovsky, S.-J. Moon, P. Löper, B. Niesen, M. Ledinsky, F.-J. Haug, J.-H. Yum, and C. Ballif. "Organometallic Halide Perovskites: Sharp Optical Absorption Edge and Its Relation to Photovoltaic Performance". In: *J. Phys. Chem. Lett.* 5.6 (2014), pp. 1035–1039. DOI: 10.1021/jz500279b.

[198] Y. Wang, Y. Zhang, P. Zhang, and W. Zhang. "High Intrinsic Carrier Mobility and Photon Absorption in the Perovskite $CH_3NH_3PbI_3$". In: *Phys. Chem. Chem. Phys.* 17 (2015), pp. 11516–11520. DOI: 10.1039/C5CP00448A.

[199] G. Xing, N. Mathews, S. Sun, S. S. Lim, Y. M. Lam, M. Grätzel, S. Mhaisalkar, and T. C. Sum. "Long-Range Balanced Electron- and Hole-Transport Lengths in Organic-Inorganic $CH_3NH_3PbI_3$". In: *Science* 342.6156 (2013), pp. 344–347. DOI: 10.1126/science.1243167.

[200] C. Wehrenfennig, G. E. Eperon, M. B. Johnston, H. J. Snaith, and L. M. Herz. "High Charge Carrier Mobilities and Lifetimes in Organolead Trihalide Perovskites". In: *Adv. Mater.* 26.10 (2014), pp. 1584–1589. DOI: 10.1002/adma.201305172.

[201] G. Kieslich, S. Sun, and A. K. Cheetham. "An extended Tolerance Factor approach for organic-inorganic perovskites". In: *Chem. Sci.* 6 (6 2015), pp. 3430–3433. DOI: 10.1039/C5SC00961H.

[202] S. Sidhik et al. "Two-dimensional perovskite templates for durable, efficient formamidinium perovskite solar cells". In: *Science* 384.6701 (2024), pp. 1227–1235. DOI: 10.1126/science.abq6993.

[203] G. P. Nagabhushana, R. Shivaramaiah, and A. Navrotsky. "Direct calorimetric verification of thermodynamic instability of lead halide hybrid perovskites". In: *PNAS* 113.28 (2016), pp. 7717–7721. DOI: 10.1073/pnas.1607850113.

[204] G. E. Eperon, G. M. Paternò, R. J. Sutton, A. Zampetti, A. A. Haghighirad, F. Cacialli, and H. J. Snaith. "Inorganic caesium lead iodide perovskite solar cells". In: *J. Mater. Chem. A* 3 (39 2015), pp. 19688–19695. DOI: 10.1039/C5TA06398A.

[205] J. A. Steele et al. "Thermal unequilibrium of strained black $CsPbI_3$ thin films". In: *Science* 365.6454 (2019), pp. 679–684. DOI: 10.1126/science.aax3878.

[206] Q. A. Akkerman et al. "Solution Synthesis Approach to Colloidal Cesium Lead Halide Perovskite Nanoplatelets with Monolayer-Level Thickness Control". In: *J. Am. Chem. Soc.* 138.3 (Jan. 2016), pp. 1010–1016. DOI: 10.1021/jacs.5b12124.

[207] J. Shamsi, Z. Dang, P. Bianchini, C. Canale, F. Di Stasio, R. Brescia, M. Prato, and L. Manna. "Colloidal Synthesis of Quantum Confined Single Crystal CsPbBr3 Nanosheets with Lateral Size Control up to the Micrometer Range". In: *J. Am. Chem. Soc.* 138.23 (June 2016), pp. 7240–7243. DOI: 10.1021/jacs.6b03166.

[208] J. Xue, R. Wang, and Y. Yang. "The surface of halide perovskites from nano to bulk". In: *Nat. Rev. Mater.* 5.11 (Nov. 2020), pp. 809–827. DOI: 10.1038/s41578-020-0221-1.

[209] P. Schulz, D. Cahen, and A. Kahn. "Halide Perovskites: Is It All about the Interfaces?" In: *Chem. Rev.* 119.5 (Mar. 2019), pp. 3349–3417. DOI: 10.1021/acs.chemrev.8b00558.

[210] L. E. Brus. "Electron-electron and electron-hole interactions in small semiconductor crystallites: The size dependence of the lowest excited electronic state". In: *J. Chem. Phys.* 80.9 (May 1984), pp. 4403–4409. DOI: 10.1063/1.447218.

[211] P. Cudazzo, I. V. Tokatly, and A. Rubio. "Dielectric screening in two-dimensional insulators: Implications for excitonic and impurity states in graphane". In: *Phys. Rev. B* 84 (8 Aug. 2011), p. 085406. DOI: 10.1103/PhysRevB.84.085406.

[212] H. M. Jang, K. Y. Jang, S. H. Lee, J. Park, and T.-W. Lee. "Exciton Bohr radius of lead halide perovskites for photovoltaic and light-emitting applications". In: *arXiv* (2025 [cited 2025 Jul 24]). DOI: 10.48550/arXiv.2505.15565.

[213] M. C. Gélvez-Rueda, E. M. Hutter, D. H. Cao, N. Renaud, C. C. Stoumpos, J. T. Hupp, T. J. Savenije, M. G. Kanatzidis, and F. C. Grozema. "Interconversion between Free Charges and Bound Excitons in 2D Hybrid Lead Halide Perovskites". In: *The Journal of Physical Chemistry C* 121.47 (2017), pp. 26566–26574. DOI: 10.1021/acs.jpcc.7b10705.

[214] N. Ma and D. Jena. "Charge Scattering and Mobility in Atomically Thin Semiconductors". In: *Phys. Rev. X* 4 (1 Mar. 2014), p. 011043. DOI: 10.1103/PhysRevX.4.011043.

[215] S. Deng, E. Shi, L. Yuan, L. Jin, L. Dou, and L. Huang. "Long-range exciton transport and slow annihilation in two-dimensional hybrid perovskites". In: *Nat. Commun.* 11.1 (Jan. 2020), p. 664. DOI: 10.1038/s41467-020-14403-z.

[216] L. Mao, C. C. Stoumpos, and M. G. Kanatzidis. "Two-dimensional hybrid Halide perovskites: Principles and promises". en. In: *J. Am. Chem. Soc.* 141.3 (Jan. 2019), pp. 1171–1190. DOI: 10.1021/jacs.8b11501.

[217] T. P. Weiss, B. Bissig, T. Feurer, R. Carron, S. Buecheler, and A. N. Tiwari. "Bulk and surface recombination properties in thin film semiconductors with different surface treatments from time-resolved photoluminescence measurements". In: *Sci. Rep.* 9.1 (Mar. 2019), p. 5385. DOI: 10.1038/s41598-019-41716-x.

[218] J. Wiktor, E. Fransson, D. Kubicki, and P. Erhart. "Quantifying Dynamic Tilting in Halide Perovskites: Chemical Trends and Local Correlations". In: *Chem. Mater.* 35.17 (Sept. 2023), pp. 6737–6744. DOI: 10.1021/acs.chemmater.3c00933.

[219] W. J. Baldwin et al. "Dynamic Local Structure in Caesium Lead Iodide: Spatial Correlation and Transient Domains". In: *Small* 20.3 (2024), p. 2303565. DOI: 10.1002/smll.202303565.

[220] C. Quarti, E. Mosconi, and F. De Angelis. "Structural and electronic properties of organo-halide hybrid perovskites from ab initio molecular dynamics". In: *Phys. Chem. Chem. Phys.* 17 (14 2015), pp. 9394–9409. DOI: 10.1039/C5CP00599J.

[221] E. Mosconi, C. Quarti, T. Ivanovska, G. Ruani, and F. De Angelis. "Structural and electronic properties of organo-halide lead perovskites: a combined IR-spectroscopy and ab initio molecular dynamics investigation". In: *Phys. Chem. Chem. Phys.* 16 (30 2014), pp. 16137–16144. DOI: 10.1039/C4CP00569D.

[222] F. P. Delgado, F. Simões, L. Kronik, W. Kaiser, and D. A. Egger. "Machine-Learning Force Fields Reveal Shallow Electronic States on Dynamic Halide Perovskite Surfaces". In: *ACS Energy Lett.* 10.7 (July 2025), pp. 3367–3374. DOI: 10.1021/acsenergylett.5c01519.

[223] W.-J. Yin, T. Shi, and Y. Yan. "Unusual Defect Physics in $CH_3NH_3PbI_3$ Perovskite Solar Cell Absorber". In: *Appl. Phys. Lett.* 104.6 (2014), p. 063903. DOI: 10.1063/1.4864778.

[224] R. E. Brandt, V. Stevanović, D. S. Ginley, and T. Buonassisi. "Identifying Defect-Tolerant Semiconductors with High Minority-Carrier Lifetimes: Beyond Hybrid Lead Halide Perovskites". In: *MRS Commun.* 5.2 (2015), pp. 265–275. DOI: 10.1557/mrc.2015.26.

[225] G.-W. Kim and A. Petrozza. "Defect Tolerance and Intolerance in Metal–Halide Perovskites". In: *Adv. Energy Mater.* 10.37 (2020), p. 2001959. DOI: 10.1002/aenm.202001959.

[226] L. Bertoluzzi, C. C. Boyd, N. Rolston, J. Xu, R. Prasanna, B. C. O'Regan, and M. D. McGehee. "Mobile Ion Concentration Measurement and Open-Access Band Diagram Simulation Platform for Halide Perovskite Solar Cells". In: *Joule* 4.1 (2020), pp. 109–127. DOI: 10.1016/j.joule.2019.10.003.

[227] G. Richardson, S. E. J. O'Kane, R. G. Niemann, T. A. Peltola, J. M. Foster, P. J. Cameron, and A. B. Walker. "Can Slow-Moving Ions Explain Hysteresis in the Current–Voltage Curves of Perovskite Solar Cells?" In: *Energy Environ. Sci.* 9 (2016), pp. 1476–1485. DOI: 10.1039/C5EE02740C.

[228] W. Tress. "Metal Halide Perovskites as Mixed Electronic–Ionic Conductors: Challenges and Opportunities–From Hysteresis to Memristivity". In: *J. Phys. Chem. Lett.* 8.13 (2017), pp. 3106–3114. DOI: 10.1021/acs.jpclett.7b00975.

[229] C. Eames, J. M. Frost, P. R. F. Barnes, B. C. O'Regan, A. Walsh, and M. S. Islam. "Ionic Transport in Hybrid Lead Iodide Perovskite Solar Cells". In: *Nat. Commun.* 6 (2015), p. 7497. DOI: 10.1038/ncomms8497.

[230] S. Meloni et al. "Ionic Polarization-Induced Current–Voltage Hysteresis in $CH_3NH_3PbX_3$ Perovskite Solar Cells". In: *Nat. Commun.* 7 (2016), p. 10334. DOI: 10.1038/ncomms10334.

[231] P. Calado, A. M. Telford, D. Bryant, X. Li, J. Nelson, B. C. O'Regan, and P. R. F. Barnes. "Evidence for Ion Migration in Hybrid Perovskite Solar Cells with Minimal Hysteresis". In: *Nat. Commun.* 7 (2016), p. 13831. DOI: 10.1038/ncomms13831.

[232] D. W. deQuilettes, W. Zhang, V. M. Burlakov, D. J. Graham, T. Leijtens, A. Osherov, V. Bulović, H. J. Snaith, D. S. Ginger, and S. D. Stranks. "Photo-Induced Halide Redistribution in Organic–Inorganic Perovskite Films". In: *Nat. Commun.* 7 (2016), p. 11683. DOI: 10.1038/ncomms11683.

[233] D. Meggiolaro and F. De Angelis. "First-Principles Modeling of Defects in Lead Halide Perovskites: Best Practices and Open Issues". In: *ACS Energy Lett.* 3.9 (2018), pp. 2206–2222. DOI: 10.1021/acsenergylett.8b01212.

[234] C. Li, A. Guerrero, S. Huettner, and J. Bisquert. "Unravelling the Role of Vacancies in Lead Halide Perovskite Through Electrical Switching of Photoluminescence". In: *Nat. Commun.* 9 (2018), p. 5113. DOI: 10.1038/s41467-018-07571-6.

[235] C. Quarti, E. Mosconi, J. M. Ball, V. D'Innocenzo, C. Tao, S. Pathak, H. J. Snaith, A. Petrozza, and F. De Angelis. "Structural and optical properties of methylammonium lead iodide across the tetragonal to cubic phase transition: implications for perovskite solar cells". In: *Energy Environ. Sci.* 9 (1 2016), pp. 155–163. DOI: 10.1039/C5EE02925B.

[236] J. Even, L. Pedesseau, and C. Katan. "Analysis of Multivalley and Multibandgap Absorption and Enhancement of Free Carriers Related to Exciton Screening in Hybrid Perovskites". In: *The Journal of Physical Chemistry C* 118.22 (June 2014), pp. 11566–11572. DOI: 10.1021/jp503337a.

[237] M. Ledinsky, T. Schönfeldová, J. Holovský, E. Aydin, Z. Hájková, L. Landová, N. Neyková, A. Fejfar, and S. De Wolf. "Temperature Dependence of the Urbach Energy in Lead Iodide Perovskites". In: *J. Phys. Chem. Lett.* 10.6 (Mar. 2019), pp. 1368–1373. DOI: 10.1021/acs.jpclett.9b00138.

[238] T. Umebayashi, K. Asai, T. Kondo, and A. Nakao. “Electronic structures of lead iodide based low-dimensional crystals”. In: *Phys. Rev. B* 67 (15 Apr. 2003), p. 155405. DOI: 10.1103/PhysRevB.67.155405.

[239] E. Mosconi, A. Amat, M. K. Nazeeruddin, M. Grätzel, and F. De Angelis. “First-Principles Modeling of Mixed Halide Organometal Perovskites for Photovoltaic Applications”. In: *J. Phys. Chem. C* 117.27 (July 2013), pp. 13902–13913. DOI: 10.1021/jp4048659.

[240] P. Umari, E. Mosconi, and F. De Angelis. “Relativistic GW calculations on CH3NH3PbI3 and CH3NH3SnI3 Perovskites for Solar Cell Applications”. In: *Sci. Rep.* 4.1 (Mar. 2014), p. 4467. DOI: 10.1038/srep04467.

[241] J. Even, L. Pedesseau, J.-M. Jancu, and C. Katan. “Importance of Spin–Orbit Coupling in Hybrid Organic/Inorganic Perovskites for Photovoltaic Applications”. In: *J. Phys. Chem. Lett.* 4.17 (Sept. 2013), pp. 2999–3005. DOI: 10.1021/jz401532q.

[242] R. Jinnouchi, J. Lahnsteiner, F. Karsai, G. Kresse, and M. Bokdam. “Phase Transitions of Hybrid Perovskites Simulated by Machine-Learning Force Fields Trained on the Fly with Bayesian Inference”. In: *Phys. Rev. Lett.* 122 (22 June 2019), p. 225701. DOI: 10.1103/PhysRevLett.122.225701.

[243] V. Tyagi, M. Pols, G. Brocks, and S. Tao. “Tracing Ion Migration in Halide Perovskites with Machine Learned Force Fields”. In: *J. Phys. Chem. Lett.* 16.20 (2025), pp. 5153–5159. DOI: 10.1021/acs.jpclett.5c01139.

[244] A. Kojima, K. Teshima, Y. Shirai, and T. Miyasaka. “Organometal Halide Perovskites as Visible-Light Sensitizers for Photovoltaic Cells”. In: *J. Am. Chem. Soc.* 131.17 (2009), pp. 6050–6051. DOI: 10.1021/ja809598r.

[245] M. A. Green, A. Ho-Baillie, and H. J. Snaith. “The Emergence of Perovskite Solar Cells”. In: *Nat. Photonics* 8.7 (2014), pp. 506–514. DOI: 10.1038/nphoton.2014.134.

[246] W. Nie et al. “Light-Activated Photocurrent Degradation and Self-Healing in Perovskite Solar Cells”. In: *Nat. Commun.* 7 (2016), p. 11574. DOI: 10.1038/ncomms11574.

[247] A. Seidu, M. Dvorak, P. Rinke, and J. Li. “Atomic and electronic structure of cesium lead triiodide surfaces”. In: *J. Chem. Phys.* 154.7 (Feb. 2021), p. 074712. DOI: 10.1063/5.0035448.

[248] K. Li and M. Wang. “Density Functional Theory Study of Surface Stability and Phase Diagram of Orthorhombic CsPbI3”. In: *J. Phys. Chem. C.* 129.7 (2025), pp. 3809–3816. DOI: 10.1021/acs.jpcc.4c07358.

[249] L. Lodeiro, F. Barría-Cáceres, K. Jiménez, R. Contreras, A. L. Montero-Alejo, and E. Menéndez-Proupin. "Methodological Issues in First-Principle Calculations of $CH_3NH_3PbI_3$ Perovskite Surfaces: Quantum Confinement and Thermal Motion". In: *ACS Omega* 5.45 (2020), pp. 29477–29491. DOI: 10.1021/acsomega.0c04420.

[250] B. Zhao et al. "Thermodynamically Stable Orthorhombic γ-$CsPbI_3$ Thin Films for High-Performance Photovoltaics". In: *J. Am. Chem. Soc.* 140.37 (2018), pp. 11716–11725. DOI: 10.1021/jacs.8b06050.

[251] D. B. Straus, S. Guo, and R. J. Cava. "Kinetically Stable Single Crystals of Perovskite-Phase $CsPbI_3$". In: *J. Am. Chem. Soc.* 141.29 (2019), pp. 11435–11439. DOI: 10.1021/jacs.9b06055.

[252] F. Ke, C. Wang, C. Jia, N. R. Wolf, J. Yan, S. Niu, T. P. Devereaux, et al. "Preserving a robust CsPbI3 perovskite phase via pressure-directed octahedral tilt". In: *Nat. Commun.* 12.1 (2021), p. 158. DOI: 10.1038/s41467-020-20745-5.

[253] T. Braeckevelt, R. Goeminne, S. Vandenhaute, S. Borgmans, T. Verstraelen, J. A. Steele, M. B. J. Roeffaers, J. Hofkens, S. M. J. Rogge, and V. Van Speybroeck. "Accurately Determining the Phase Transition Temperature of $CsPbI_3$ via Random-Phase Approximation Calculations and Phase-Transferable Machine Learning Potentials". In: *Chem. Mater.* 34.19 (2022), pp. 8561–8576. DOI: 10.1021/acs.chemmater.2c01508.

[254] E. Fransson, J. M. Rahm, J. Wiktor, and P. Erhart. "Revealing the Free Energy Landscape of Halide Perovskites: Metastability and Transition Characters in CsPbBr3 and MAPbI3". In: *Chem. Mater.* 35.19 (Oct. 2023), pp. 8229–8238. DOI: 10.1021/acs.chemmater.3c01740.

[255] R. X. Yang and L. Z. Tan. "Understanding size dependence of phase stability and band gap in $CsPbI_3$ perovskite nanocrystals". In: *J. Chem. Phys.* 152.3 (2020), p. 034702. DOI: 10.1063/1.5128016.

[256] V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, and G. Csányi. "Gaussian process regression for materials and molecules". In: *Chem. Rev.* 121.16 (2021), pp. 10073–10141. DOI: 10.1021/acs.chemrev.1c00022.

[257] S. Stocker, H. Jung, G. Csányi, C. F. Goldsmith, K. Reuter, and J. T. Margraf. "Estimating Free Energy Barriers for Heterogeneous Catalytic Reactions with Machine Learning Potentials and Umbrella Integration". In: *J. Chem. Theory Comput.* 19 (19 2023), pp. 6796–6804. DOI: 10.1021/acs.jctc.3c00541.

[258] H. Jung, L. Sauerland, S. Stocker, K. Reuter, and J. T. Margraf. "Machine-learning driven global optimization of surface adsorbate geometries". In: *npj Comput. Mater.* 9.1 (2023), p. 114. DOI: 10.1038/s41524-023-01065-w.

[259] X. Fu, Z. Wu, W. Wang, T. Xie, S. Keten, R. Gomez-Bombarelli, and T. Jaakkola. "Forces are not enough: Benchmark and critical evaluation for machine learning force fields with molecular simulations". In: *arXiv* (2022 [cited 2024 Aug 24]). DOI: 10.1063/5.0147023.

[260] S. Stocker, J. Gasteiger, F. Becker, S. Günnemann, and J. T. Margraf. "How robust are modern graph neural network potentials in long and hot molecular dynamics simulations?" In: *Mach. Learn.: Sci. Technol.* 3 (4 2022), p. 045010. DOI: 10.1088/2632-2153/ac9955.

[261] J. Gawlikowski et al. "A survey of uncertainty in deep neural networks". In: *Artif. Intell. Rev.* 56.1 (2023), pp. 1513–1589. DOI: 10.1007/s10462-023-10562-9.

[262] A. H. Larsen et al. "The atomic simulation environment—a Python library for working with atoms". In: *J. Phys.: Condens. Matter* 29.27 (June 2017), p. 273002. DOI: 10.1088/1361-648X/aa680e.

[263] J. J. Choi, X. Yang, Z. M. Norman, S. J. L. Billinge, and J. S. Owen. "Structure of Methylammonium Lead Iodide Within Mesoporous Titanium Dioxide: Active Material in High-Performance Perovskite Solar Cells". In: *Nano Lett.* 14.1 (2014). PMID: 24266720, pp. 127–133. DOI: 10.1021/nl403514x.

[264] T. Ivanovska, C. Quarti, G. Grancini, A. Petrozza, F. De Angelis, A. Milani, and G. Ruani. "Vibrational Response of Methylammonium Lead Iodide: From Cation Dynamics to Phonon-Phonon Interactions". In: *ChemSusChem* 9.20 (2016), pp. 2994–3004. DOI: 10.1002/cssc.201600932.

[265] V. DiezCabanes, S. Giannini, D. Beljonne, and C. Quarti. "On the Origin of Energetic Disorder in Mixed Halides Lead Perovskites". In: *Adv. Opt. Mater.* 12.8 (2024), p. 2301105. DOI: 10.1002/adom.202301105.

[266] C. S. Ewig et al. "Derivation of class II force fields. VIII. Derivation of a general quantum mechanical force field for organic compounds". In: *J. Comput. Chem.* 22.15 (2001), pp. 1782–1800. DOI: 10.1002/jcc.1131.

[267] M. J. Hwang, T. P. Stockfisch, and A. T. Hagler. "Derivation of Class II Force Fields. 2. Derivation and Characterization of a Class II Force Field, CFF93, for the Alkyl Functional Group and Alkane Molecules". In: *J. Am. Chem. Soc.* 116.6 (Mar. 1994), pp. 2515–2525. DOI: 10.1021/ja00085a036.

[268] A. Charkin-Gorbulin, A. Kokorin, H. E. Sauceda, S. Chmiela, C. Quarti, D. Beljonne, A. Tkatchenko, and I. Poltavsky. “Atomic orbits in molecules and materials for improving machine learning force fields”. In: *Mach. Learn.: Sci. Technol.* 6.3 (July 2025), p. 035005. DOI: 10.1088/2632-2153/adea0b.

[269] D. Bonchev. *Chemical Graph Theory: Introduction and Fundamentals*. Chemical Graph Theory. Taylor & Francis, 1991.

[270] R. García-Domenech, J. Gálvez, J. V. de Julián-Ortiz, and L. Pogliani. “Some New Trends in Chemical Graph Theory”. In: *Chem. Rev.* 108.3 (2008), pp. 1127–1169. DOI: 10.1021/cr0780006.

[271] B. L. Douglas and J. B. Wang. “A classical approach to the graph isomorphism problem using quantum walks”. In: *J. Phys. A: Math. Theor.* 41.7 (Feb. 2008), p. 075303. DOI: 10.1088/1751-8113/41/7/075303.

[272] J. Lauri and R. Scapellato. *Topics in Graph Automorphisms and Reconstruction*. 2nd ed. London Mathematical Society Lecture Note Series. Cambridge University Press, 2016. DOI: 10.1017/CBO9781316669846.

[273] B. D. McKay and A. Piperno. “Practical graph isomorphism, II”. In: *J. Symb. Comput.* 60 (2014), pp. 94–112. DOI: 10.1016/j.jsc.2013.09.003.

[274] G. Fonseca, I. Poltavsky, V. Vassilev-Galindo, and A. Tkatchenko. “Improving molecular force fields across configurational space by combining supervised and unsupervised machine learning”. In: *J. Chem. Phys.* 154.12 (Mar. 2021), p. 124102. DOI: 10.1063/5.0035530.

[275] Y. Park, J. Kim, S. Hwang, and S. Han. “Scalable Parallel Algorithm for Graph Neural Network Interatomic Potentials in Molecular Dynamics Simulations”. In: *J. Chem. Theory Comput.* 20.11 (2024), pp. 4857–4868. DOI: 10.1021/acs.jctc.4c00190.

[276] J. Kim, J. Kim, J. Kim, J. Lee, Y. Park, Y. Kang, and S. Han. “Data-Efficient Multifidelity Training for High-Fidelity Machine Learning Interatomic Potentials”. In: *J. Am. Chem. Soc.* 147.1 (Jan. 2025), pp. 1042–1054. DOI: 10.1021/jacs.4c14455.

[277] D. M. Anstine, R. Zubatyuk, and O. Isayev. “AIMNet2: A Neural Network Potential to Meet Your Neutral, Charged, Organic, and Elemental-Organic Needs”. In: *Chem. Sci.* (2025). Advance Article. DOI: 10.1039/D4SC08572H.

[278] M. Neumann, J. Gin, B. Rhodes, S. Bennett, Z. Li, H. Choubisa, A. Hussey, and J. Godwin. “Orb: A Fast, Scalable Neural Network Potential”. In: *arXiv* (2024 [cited 2024 Dec 12]). DOI: 10.48550/arXiv.2410.22570.

[279] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler. "*Ab initio* molecular simulations with numeric atom-centered orbitals". In: *Comput. Phys. Commun.* 180.11 (2009), pp. 2175–2196. DOI: https://doi.org/10.1016/j.cpc.2009.06.022.

[280] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter. "Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach". In: *Comput. Phys. Commun.* 167.2 (2005), pp. 103–128. DOI: 10.1016/j.cpc.2004.12.014.

[281] J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele. "cp2k: atomistic simulations of condensed matter systems". In: *WIREs Comput. Mol. Sci.* 4.1 (2014), pp. 15–25. DOI: 10.1002/wcms.1159.

[282] J. VandeVondele and J. Hutter. "Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases". In: *J. Chem. Phys.* 127.11 (Sept. 2007), p. 114105. DOI: 10.1063/1.2770708.

[283] S. Goedecker, M. Teter, and J. Hutter. "Separable dual-space Gaussian pseudopotentials". In: *Phys. Rev. B* 54 (3 July 1996), pp. 1703–1710. DOI: 10.1103/PhysRevB.54.1703.

[284] M. Krack. "Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals". In: *Theor. Chem. Acc.* 114 (Jan. 2005), pp. 145–152. DOI: 10.1007/s00214-005-0655-y.

[285] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. "Molecular dynamics with coupling to an external bath". In: *J. Chem. Phys.* 81.8 (Oct. 1984), pp. 3684–3690. DOI: 10.1063/1.448118.

[286] P. Giannozzi et al. "QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials". In: *J. Phys.: Condens. Matter* 21.39 (2009), 395502 (19pp). DOI: 10.1088/0953-8984/21/39/395502.

[287] P Giannozzi et al. "Advanced capabilities for materials modelling with QUANTUM ESPRESSO". In: *J. Phys.: Condens. Matter* 29.46 (2017), p. 465901. DOI: 10.1088/1361-648x/aa8f79.

[288] P. Giannozzi et al. "Quantum ESPRESSO toward the exascale". In: *J. Chem. Phys.* 152.15 (2020), p. 154105. DOI: 10.1021/acs.jctc.3c00249.

Appendix A

Dataset generation procedure

A.1 1,8-naphthyridine on Graphene

The *1,8-Naphthyridine molecule adsorbed on graphene dataset* ($C_8H_6N_2/C_{98}$, 114 atoms) was generated using six independent NVT MD runs performed at a temperature of 500 K with a time step of 1 fs. The resulting dataset contains 15,000 reference structures with forces and energies computed at the PBE+MBD-NL level of theory.

This dataset was created to assess the ability of MLFFs to deal with multi-component interfaces. While the 1,8-Naphthyridine molecule and the graphene sheet are, by themselves, trivial challenges for modern MLFFs, their combination introduces a new element: molecule-surface interaction. This interaction is non-covalent and small in strength compared to interatomic forces within the subsystems. At the same time, molecule-surface interaction determines many effects of practical interest, such as surface friction or adsorption. Train and validation datasets of sizes 200, 400, 600, 800, and 1000 were provided to participants. The test dataset in all cases consisted of all configurations left in the 1,8-Naphthyridine molecule adsorbed on a pristine graphene sheet dataset after extracting the train and validation subsets. MD trajectories were generated using the MLFFs trained on the largest training set (1000 configurations) with periodic boundary conditions.

A.2 $MAPbI_3$ Perovskite

The tetragonal methylammonium lead iodide ($MAPbI_3$) dataset contains 698 structures selected from a preexisting DFT-based MD trajectory. Structures were clustered based on their energy distribution via the sGDML clustering procedure, then recomputed at the PBE+MBD-NL level under periodic boundary conditions to faithfully represent both the organic cation

dynamics and inorganic lattice interactions.

This system extends the complexity of multi-component modeling by combining heavy and light atoms within one relatively large system. Three types of interactions can be distinguished: interatomic forces within MA molecules, interatomic forces between heavy I and Pb atoms, and the interaction between the MA and PbI_3 subsystems. The additional complexity of the challenge arises from the sluggish motion of Pb and I atoms compared to that of H, C, and N atoms and the rotational freedom of the MA molecule within the PbI_3 cage. All these factors lead to different requirements for sufficient sampling of different system parts, further complicating the MLFF training process. The training and validation datasets are the same for all MLFFs; their sizes are 100, 200, 300, 400, and 500. Only the MLFF models trained on 500 reference structures were used for MD trajectory generation.

A.3 Graphene/molecule interfaces

To demonstrate the capabilities of the developed graph-based symmetry search algorithm, we chose four periodic systems where the graphene surface interacts with various organic molecules: ethanol, 1,8-naphthyridine, D-alanine, and D-histidine. All-electron DFT with the PBE exchange-correlation functional and nonlocal many-body dispersion method to include vdW interactions implemented in the FHI-AIMS package was used to generate datasets for chosen systems [279]. For all studied systems, the 7×7 graphene layer was used. Eight AIMD simulations with different initial velocities were performed at 500 K using a time step of 1 fs and 1700 cm^{-1} damping in the NVT Nose-Hoover thermostat for a total for 2 ps.

A.4 Graphene with pyridinic-N defect

The computational investigation of the graphene layer featuring a pyridinic-N defect was carried out using a 15×15 supercell. The electronic structure and dynamical properties of the system were computed using all-electron DFT as implemented in the FHI-AIMS software package [279]. PBE exchange-correlation functional was employed alongside the nonlocal MBD-NL method to accurately capture vdW interactions. Calculations were performed utilizing the standard "light" computational settings provided by FHI-AIMS. To sample the configurational space thoroughly, a total of 20 independent AIMD simulations, each initialized with distinct atomic velocities, were performed. The simulations were conducted under an NVT ensemble maintained at a temperature of 500 K, employing the Nose-Hoover thermostat with a damping parameter set at 1700 cm^{-1}. Each AIMD trajectory had a duration of 7 ps

with an integration time step of 1 fs.

A.5 Perovskite slab

The crystal structure $2 \times 2 \times 5$ $CsPbI_3$ perovskite slab corresponds to the cubic phase with supercell dimensions $a = b = 17.81$ Å and $c = 50.01$ Å. The AIMD simulation was carried out using the CP2K package [280, 281]. The mixed GPW approach was defined by a relative cutoff of 50 Ry, a plane wave cutoff of 300 Ry, and 16 grid points in real space. The electronic structure properties were calculated using the PBE as XC functional, DZVP-MOLOPT as basis sets, and norm-conservative GTH as pseudopotentials [282, 283, 284]. MD was performed for 44 ps using an integration step of 4 fs, under the NVT Nose-Hoover thermostat with a 20 fs time-constant and the target temperature of 600 K.

To obtain a dataset for further training of MLFF potential, the $4 \times 4 \times 5$ $CsPbI_3$ perovskite slab structure with a CsI layer surface in the $xy-$plane was constructed. The initial crystal structure corresponds to the orthorhombic phase with supercell dimensions $a = 35.07$ Å , $b = 35.29$ Å , and $c = 50.01$ Å after the cell was relaxed using the DFT with the PBE exchange-functional, D3 dispersion correction, DZVP-MOLOPT as basis sets, and norm-conservative GTH as pseudopotentials implemented in the CP2K package [280, 281, 282, 283, 284]. The mixed Gaussian and plane waves approach was defined by a relative cutoff of 50 Ry, a plane wave cutoff of 300 Ry, and 16 grid points in real space. Two AIMD simulations were conducted at target temperatures of 300 K and 600 K. Each simulation spanned 44 ps, utilizing an integration time step of 4 fs. The simulations were performed under the NPT ensemble with a flexible cell in the x-y plane. The system was coupled to a Nosé-Hoover thermostat and barostat, with time constants set to 20 fs and 1000 fs, respectively, to maintain the target temperature and a pressure of 1 bar.

Two datasets were constructed by selecting 300 configurations from each trajectory based on energy distribution, using the procedure implemented in the sGDML software package [15, 17, 16]. The first training dataset employed forces, energies, and stress tensors obtained from AIMD simulations performed at the PBE+D3 level of theory. In contrast, the second dataset involved recomputing the forces, energies, and stress tensors using all-electron DFT with PBE, incorporating vdW interactions via the MBD-NL as implemented in the FHI-aims package [279].

Two datasets were used to train separate MACE models to examine the impact of dispersion interactions on the large-scale dynamics of Perovskite slab materials. MACE, an equivariant message-passing neural network, encodes information on many body atomic geometry,

capturing the intricate relationships between atoms in the system [24, 25]. The trained models incorporate hidden representations up to the second order, with each order containing 64 features ('64x0e + 64x1o'). Additionally, the radial multilayer perceptron comprises three fully connected layers, each with 64 neurons.

Both MACE models were used to perform separate 5 ns long MD dynamics to investigate the impact of dispersion interactions on electronic structure. Each simulation utilizes an integration time step of 4 fs under NPT ensemble with a flexible cell in the x-y plane. The temperature and pressure were maintained at 300 K and 1 bar, respectively, using Parrinello-Rahman barostat with a 'pfactor' of $3.7 \cdot 10^5$ GPa·fs^2 and a Nose-Hoover thermostat with a time constant of 20 fs, as implemented in the ASE package [262].

Appendix B

Supplementary Information for Chapter 4

B.1 Normalized errors on the test sets

Quantity	Metric	Method				
		MACE	**SO3**	**sGDML**	**SOAP/GAP**	**FCHL19***
E	MAE	0.14	0.24	0.26	0.75	211
	RMSE	0.18	0.30	0.33	0.92	211
	MAXE	0.73	1.81	1.54	2.83	219
F	MAE	0.21	2.48	1.00	2.51	6.39
	RMSE	0.34	4.12	1.66	3.36	8.46
	MAXE	18.8	194	54.2	308	842
F(H)	MAE	0.42	1.87	1.19	2.92	8.59
	RMSE	0.58	2.66	1.72	4.18	11.4
	MAXE	14.9	57.4	18.4	165	892
F(C)	MAE	0.19	2.51	0.98	2.47	6.26
	RMSE	0.31	4.16	1.62	3.29	8.26
	MAXE	18.5	191	50.0	303	828
F(N)	MAE	0.52	2.05	1.74	3.57	8.52
	RMSE	0.76	2.90	2.71	4.73	11.2
	MAXE	10.7	51.6	47.1	80.9	348
F(mol)	MAE	5.90	28.3	8.03	76.6	197
	RMSE	7.90	37.8	12.0	128	299
	MAXE	67.0	243	130	965	2011

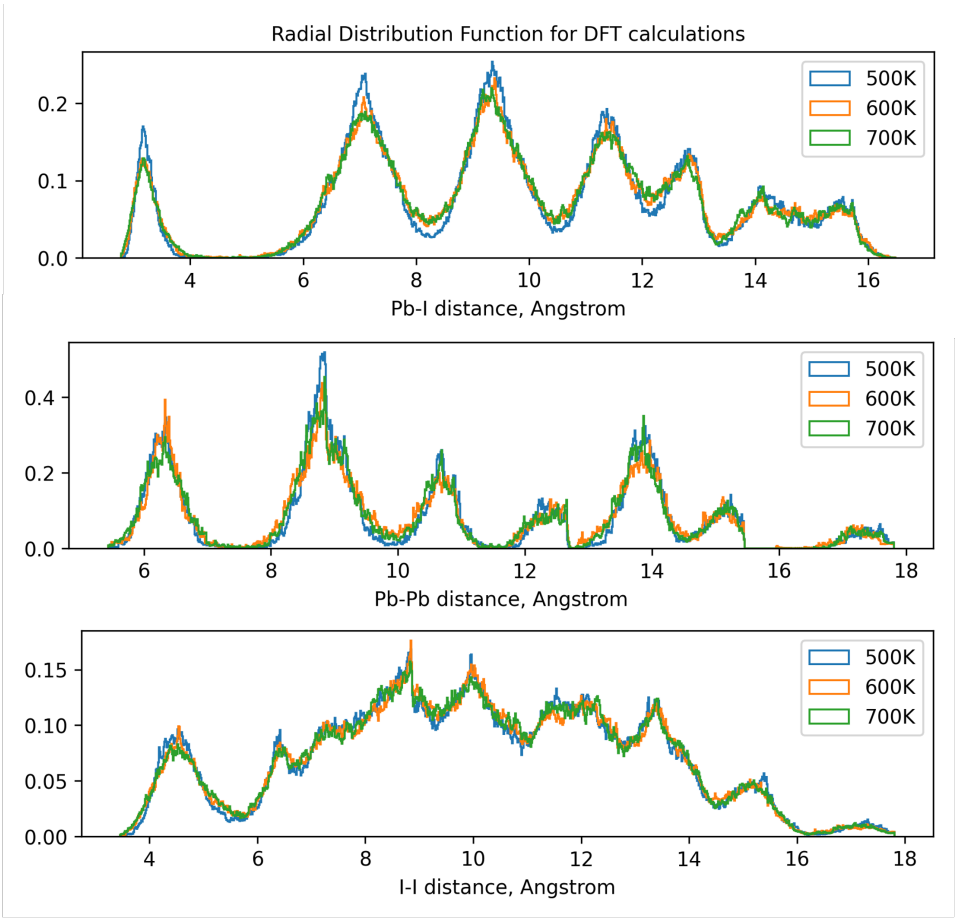
Table B.1 MAE, RMSE, and MAXE as rows. Values are % w.r.t. mean absolute energies and forces in the reference DFT datasets for the 1,8-naphthyridine on graphene system.

Quantity	Metric	Method				
		MACE	**SO3**	**sGDML**	**SOAP/GAP**	**FCHL19***
E	MAE	0.45	0.68	-	1.49	102
	RMSE	0.58	0.85	-	1.88	106
	MAXE	1.93	2.75	-	6.28	192
F	MAE	1.56	1.87	-	4.00	26.4
	RMSE	2.14	2.63	-	5.50	34.0
	MAXE	45.3	36.9	-	66.9	642
F(H)	MAE	1.10	1.19	-	2.63	24.8
	RMSE	1.47	1.57	-	3.46	32.1
	MAXE	43.6	22.3	-	64.4	618
F(C)	MAE	0.85	0.91	-	2.39	12.3
	RMSE	1.12	1.17	-	3.04	16.2
	MAXE	19.2	18.6	-	32.7	322
F(N)	MAE	0.96	1.03	-	2.44	13.4
	RMSE	1.24	1.31	-	3.09	17.3
	MAXE	11.5	8.10	-	22.9	283
F(Pb)	MAE	4.80	7.02	-	14.3	58.4
	RMSE	6.17	8.90	-	18.0	73.9
	MAXE	55.1	44.1	-	114	589
F(I)	MAE	4.72	6.18	-	11.9	66.2
	RMSE	6.22	8.16	-	15.5	84.1
	MAXE	86.2	60.6	-	118	799
F(mol)	MAE	6.20	6.12	-	15.0	171
	RMSE	8.03	7.77	-	19.1	217
	MAXE	102	63.0	-	141	1449

Table B.2 MAE, RMSE, and MAXE as rows. Values are % w.r.t. mean absolute energies and forces in the reference DFT datasets for the $MAPbI_3$ system.

B.2 Radial Distribution Functions for methylammonium lead iodide perovskite

One-picosecond molecular dynamics simulations were conducted at 500 K, 600 K, and 700 K with Berendsen thermostat using the Quantum ESPRESSO (QE) code to construct radial distribution functions (RDF) for various atom pairs in methylammonium lead iodide perovskite ($MAPbI_3$) system [285, 286, 287, 288]. Density functional theory (DFT) calculations were performed with the PBE functional, employing a kinetic energy cutoff of 200.0 Ry and ultrasoft pseudopotentials. The distances for each atom pair shown in Figs. B.1, B.2, and B.3 were extracted from every step in the MD trajectories in the DFT calculations and from every 100th step for each of the 12 MD trajectories generated for the MACE and SO3 potentials at 500 K.



xx

Figure B.1 Radial distribution function (RDF) for lead and iodine atoms in the orthorhombic phase of 2x2x2 $MAPbI_3$, derived from MD simulations at 500 K, 600 K, and 700 K using DFT with the PBE functional.

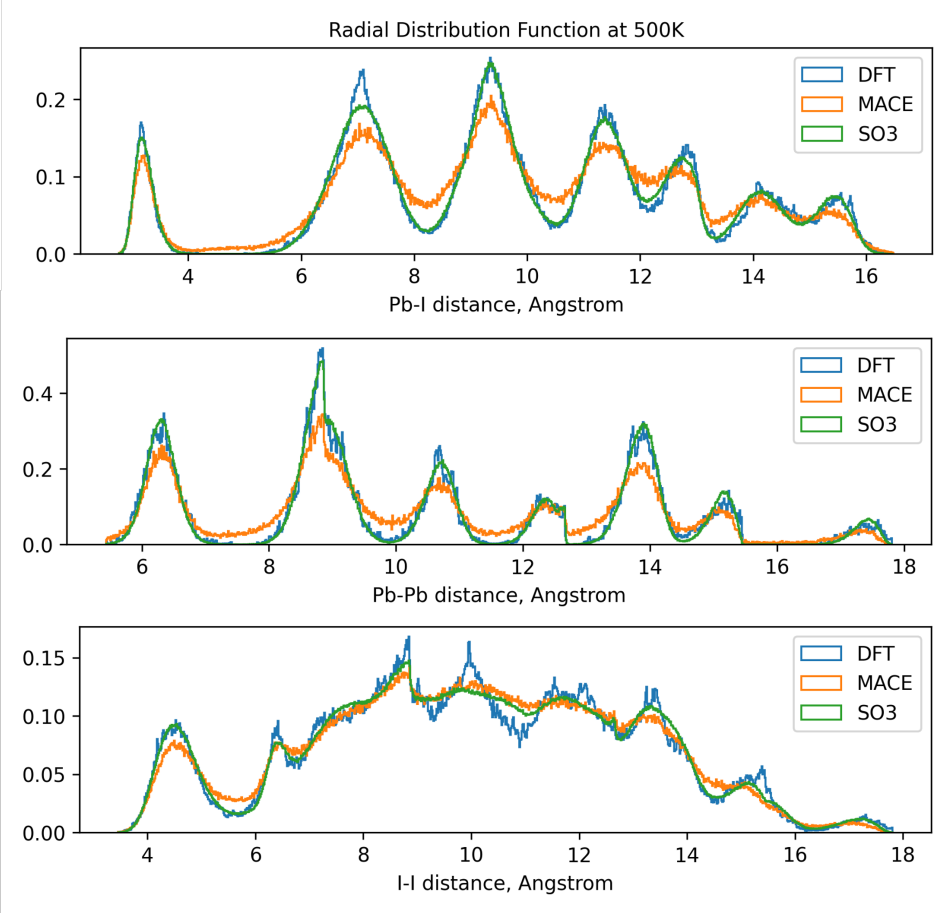


Figure B.2 Radial distribution function (RDF) for lead and iodine atoms in the orthorhombic phase of 2x2x2 $MAPbI_3$, derived from MD simulations at 500 K. The RDFs are compared for trajectories generated using DFT with the PBE functional, MACE, and SO3.

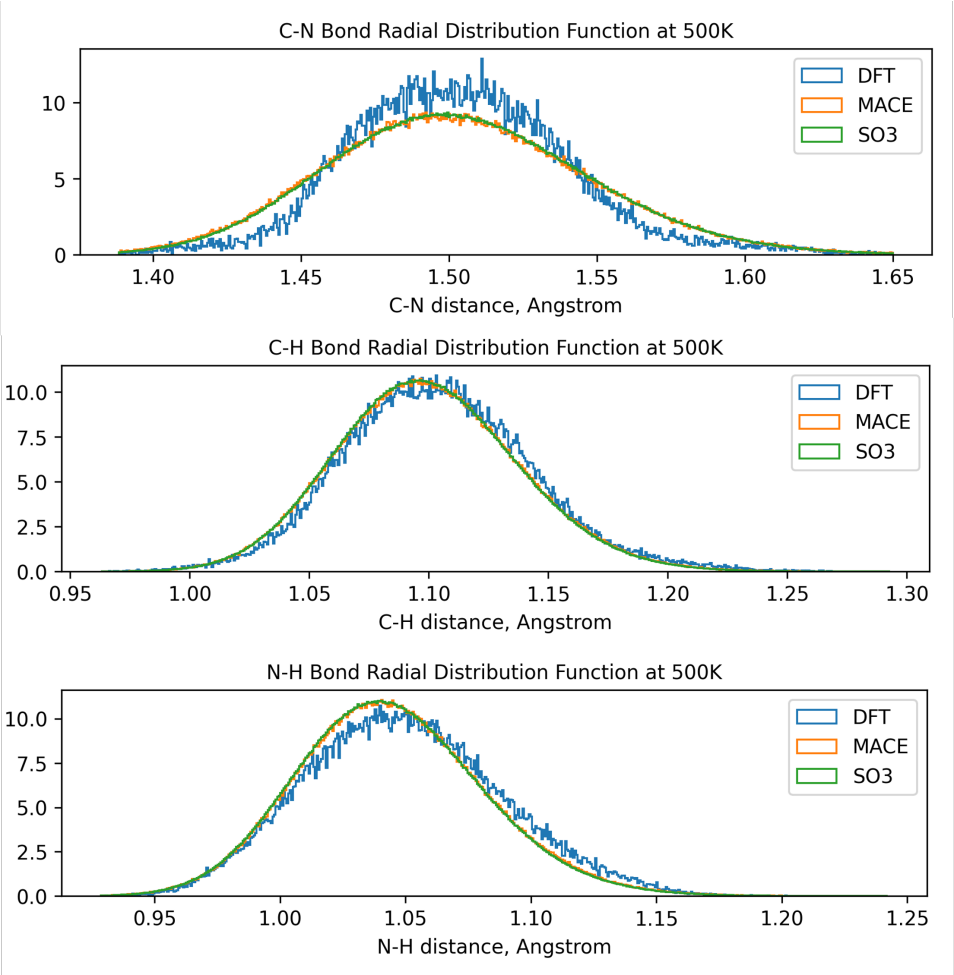


Figure B.3 Radial distribution function (RDF) for methylammonium extracted from the MD simulations of the orthorhombic phase of 2x2x2 $MAPbI_3$ generated using DFT(PBE), MACE, and SO3 at 500K. The radial distribution function of bonds in methylammonium in the orthorhombic phase of 2x2x2 $MAPbI_3$, derived from MD simulations at 500 K. The radial distribution functions of bonds are compared for trajectories generated using DFT with the PBE functional, MACE, and SO3.

Appendix C

Supplementary Information for Chapter 5

C.1 Graphene with a pyridinic-N defect

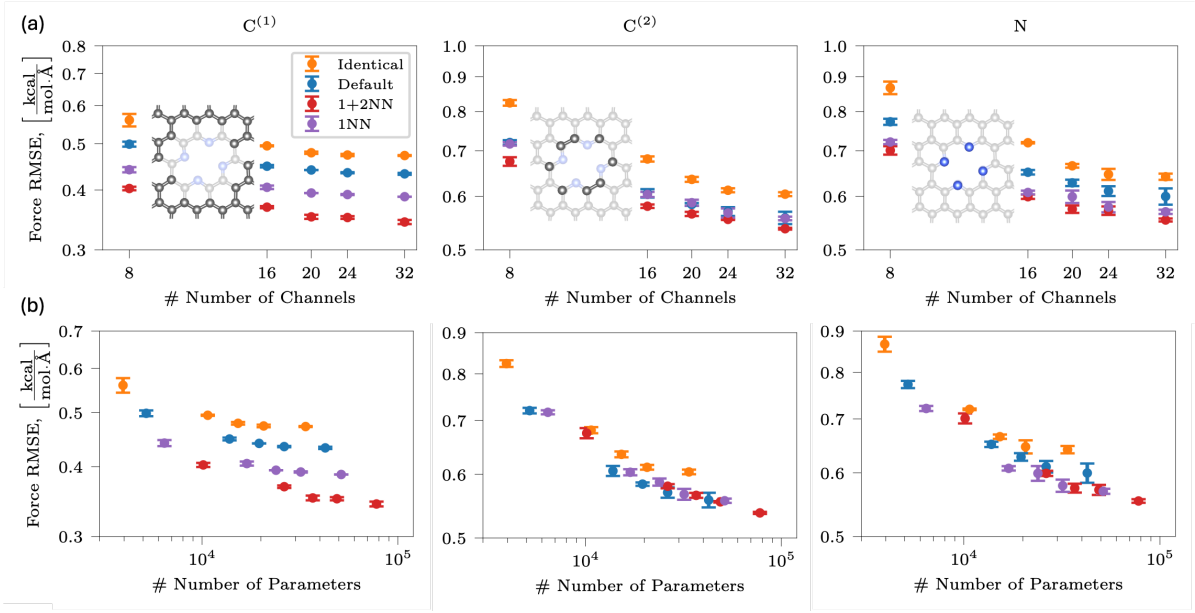


Figure C.1 Force prediction accuracy of MACE models in the graphene with a pyridinic-N defect system as a function of (a) the number of channels in hidden irreducible representations and (b) the number of parameters for three atoms group: nitrogen atoms, carbons bonded to nitrogen and all other carbons. RMSE is evaluated across four model architectures: the default RMSE architecture (blue), a first- and second-neighbor-based MACE architecture (red), a first-neighbor-based MACE architecture (purple), and a MACE architecture treating all atoms as identical (orange). Uncertainty was quantified using cross-validation testing.

The comparison of force MAE distributions across different MACE architectures for the graphene system with a pyridinic-N defect, in Figure 5.6 reveals a variation in the extent

of improvement across atomic subsets, with more pronounced gains observed for atoms farther from the defect. This variation is further analyzed in Figure C.1 and illustrates the relationship between the force RMSE of different model types and the number of channels in hidden irreducible representations, which determine the dimensionality of the internal feature space at a given symmetry level, and the number of parameters in the model. All model types exhibit a downward trend across all atomic subsets with the increasing number of channels and the number of parameters in the model. Among the different architectures, the identical atoms model, a MACE architecture that treats all atoms as indistinguishable, consistently yields the highest RMSE across all subsets. Although the identical model partially catches up with the 1NN and 1+2NN models for nitrogen and its directly bonded carbon neighbors after a substantial increase in the number of channels (from 8 to 20), it fails to achieve comparable accuracy for the remaining carbon atoms, where its RMSE remains significantly higher. Meanwhile, the default model exhibits systematically higher RMSE values than both the 1NN and 1+2NN models for the remaining carbon atoms subset. This outcome suggests the insufficiency of a purely elemental-based distinction in capturing the complexity of local neighborhood chemical compositions introduced by the pyridinic-N defect. However, the RMSE differences between the default model and the orbits-based models gradually decrease for nitrogen and its directly bonded carbon neighbors as the number of channels increases, indicating that broader network capacity compensates for the lack of local environment differentiation in these cases. Still, the orbit-based models consistently achieve higher accuracy than the default MACE model with fewer channels and trainable parameters.

C.2 Phonon Spectra Consistency

To demonstrate that the orbit-based model preserves key vibrational physics, we computed the Γ-point phonon frequencies in the harmonic approximation via the finite-difference method for graphene containing a pyridinic-N defect. We compared two MACE variants that achieve very close accuracy for the prediction force on the test set–despite different model sizes:

- Default MACE model with 42 608 learnable parameters that have force RMSE on the test set 0.43 kcal/(mol·Å)
- 1+2NN MACE model with 26 304 learnable parameters that have force RMSE on the test set 0.38 kcal/(mol·Å)

To obtain the phonon frequencies, the structures were first relaxed independently by each model, until the maximum force on any atom fell below $F_{max} = 0.001$ eV/Å. Phonon frequen-

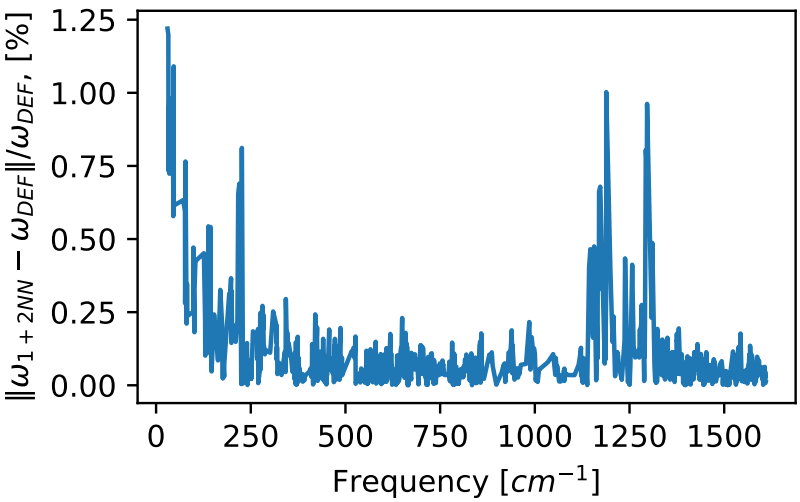


Figure C.2 Relative difference in Γ-point phonon frequencies between the Default MACE and 1+2NN MACE models for graphene with a pyridinic-N defect. Structures were relaxed independently by each model to a maximum force threshold of 0.001 eV/Å, then phonon frequencies were computed via 0.01 Åfinite displacements.

cies were then determined by displacing each atom by 0.01 Å and evaluating the resulting Hessian matrices.

As shown in Figure C.2, the maximum relative deviation in Γ-point phonon frequencies between Default and orbit-based models remains below 1.25 % for all vibrational modes, even though the 1+2NN model uses 1.6 times fewer parameters. This confirms that explicitly encoding permutation symmetry via atomic orbits reproduces vibrational properties with close accuracy–even at substantially reduced model size.

C.3 Epoch-Resolved Training and Validation Losses

The epoch-resolved learning curves in Figs. C.3–C.5 provide a compact diagnostic of optimization stability and generalization for the MACE models studied in this work. Plotting the training and validation losses as a function of epoch serves three purposes: (i) to verify that the optimization reaches a well-defined plateau (convergence), (ii) to quantify the generalization gap between training and validation performance as an indicator of overfitting, and (iii) to enable a controlled comparison of structural encodings and model capacity under otherwise comparable training conditions. In all cases, the validation loss is evaluated on held-out configurations that are not used for parameter updates, so separation between the training and validation curves directly reflects how well the learned representation transfers beyond the training set.

For the $CsPbI_3$ slab, Figs. C.3 and C.4 compare learning dynamics across alternative input

encodings. Figure C.3 focuses on the default species-based representation versus an orbit-based encoding in which distinct local environments are mapped to separate effective types, and repeats the comparison across different train/validation set sizes. This organization makes it straightforward to assess how data availability influences both the rate of convergence (how rapidly the loss decreases with epoch) and the extent of any generalization gap (the vertical separation between training and validation losses at late epochs). Figure C.4 extends this analysis by including an "all atoms identical" baseline and by reporting cross-validation statistics (mean and $\pm 1\sigma$ variability across folds), which is particularly informative for estimating sensitivity to data partitioning and for identifying regimes where performance is dominated by sampling variability rather than model capacity. In addition, the row-wise sweep over channel numbers in the hidden irreducible representations links learning-curve behaviour to model capacity: increased capacity can accelerate fitting of the training set, but may also increase the risk of a widening train–validation separation when the available data are limited.

For graphene with a pyridinic-N defect (Fig. C.5), the same diagnostic is used to contrast species-aware versus progressively more local-environment-aware encodings (default species, all atoms identical, 1NN, and 1+2NN). Presenting these curves across multiple train/validation set sizes highlights how representation choice interacts with data regime: more informative encodings can, in principle, reduce representational ambiguity by distinguishing chemically or topologically distinct environments, whereas less informative encodings force the network to explain heterogeneous environments with a smaller effective input alphabet. Consequently, these learning curves complement aggregate error metrics by showing whether improved validation performance (when present) arises from genuinely better generalization (lower validation loss with a modest gap) or from increased fitting capacity that primarily reduces training loss.

Overall, the epoch-resolved curves in this section are intended as qualitative evidence that the reported models are trained to a converged regime and that comparisons between encodings and hyperparameters are not confounded by premature stopping or unstable optimization. They also provide a visual basis for selecting training schedules (e.g., a sufficient number of epochs and/or early-stopping criteria) that balance convergence against overfitting, especially in the low-data settings where fold-to-fold variability can be substantial.

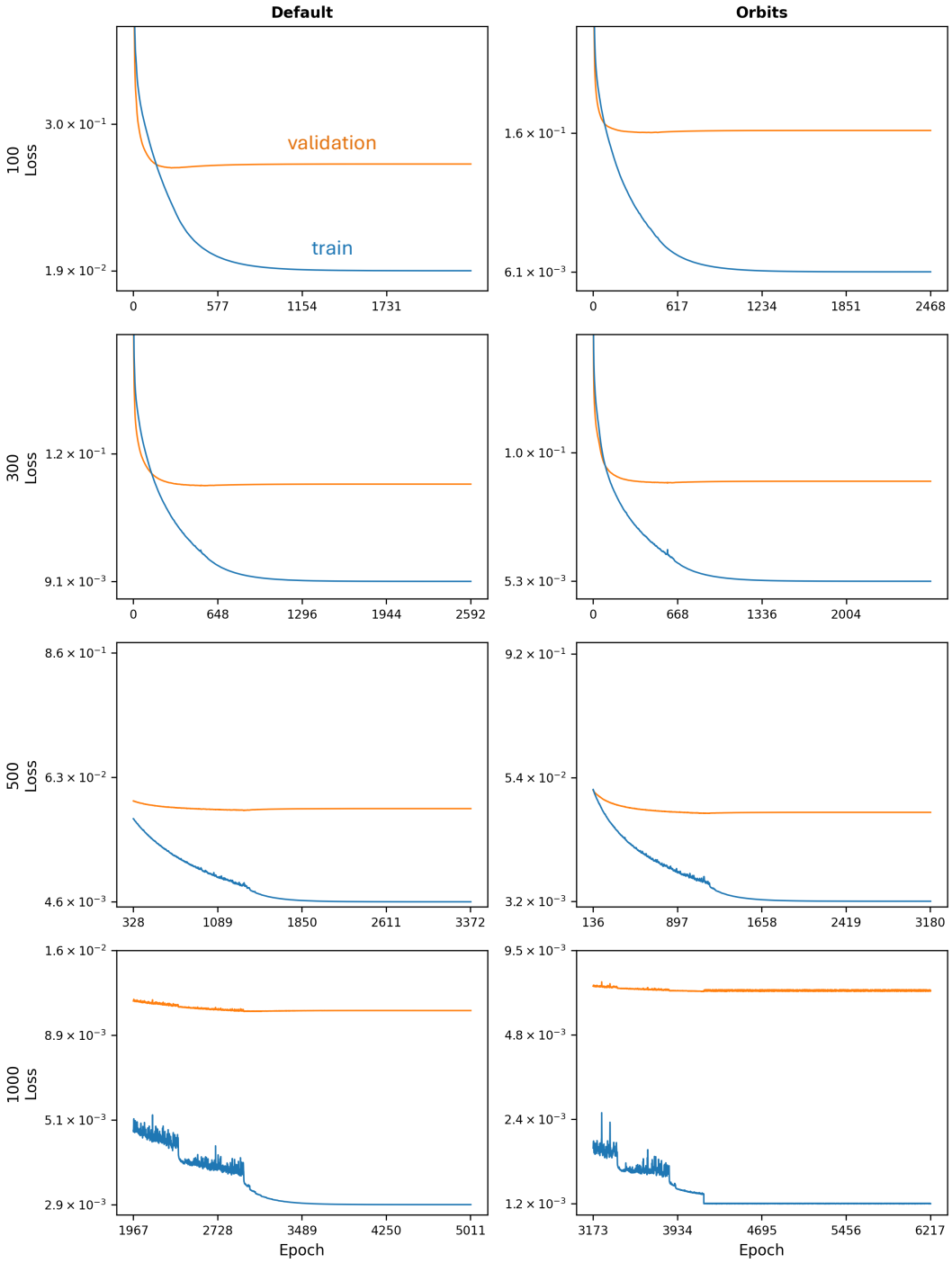


Figure C.3 Training and validation loss as a function of epoch for MACE models of the $CsPbI_3$ slab using two structural encodings: (first column) the default species representation and (second column) an orbit-based representation in which distinct atomic environments are treated as separate atom types. For each encoding, solid lines indicate the training (blue) and validation (orange) losses. The learning curves illustrate the convergence behaviour during optimization and the magnitude of the generalization gap between training and validation performance. The rows correspond to different train/validation set sizes.

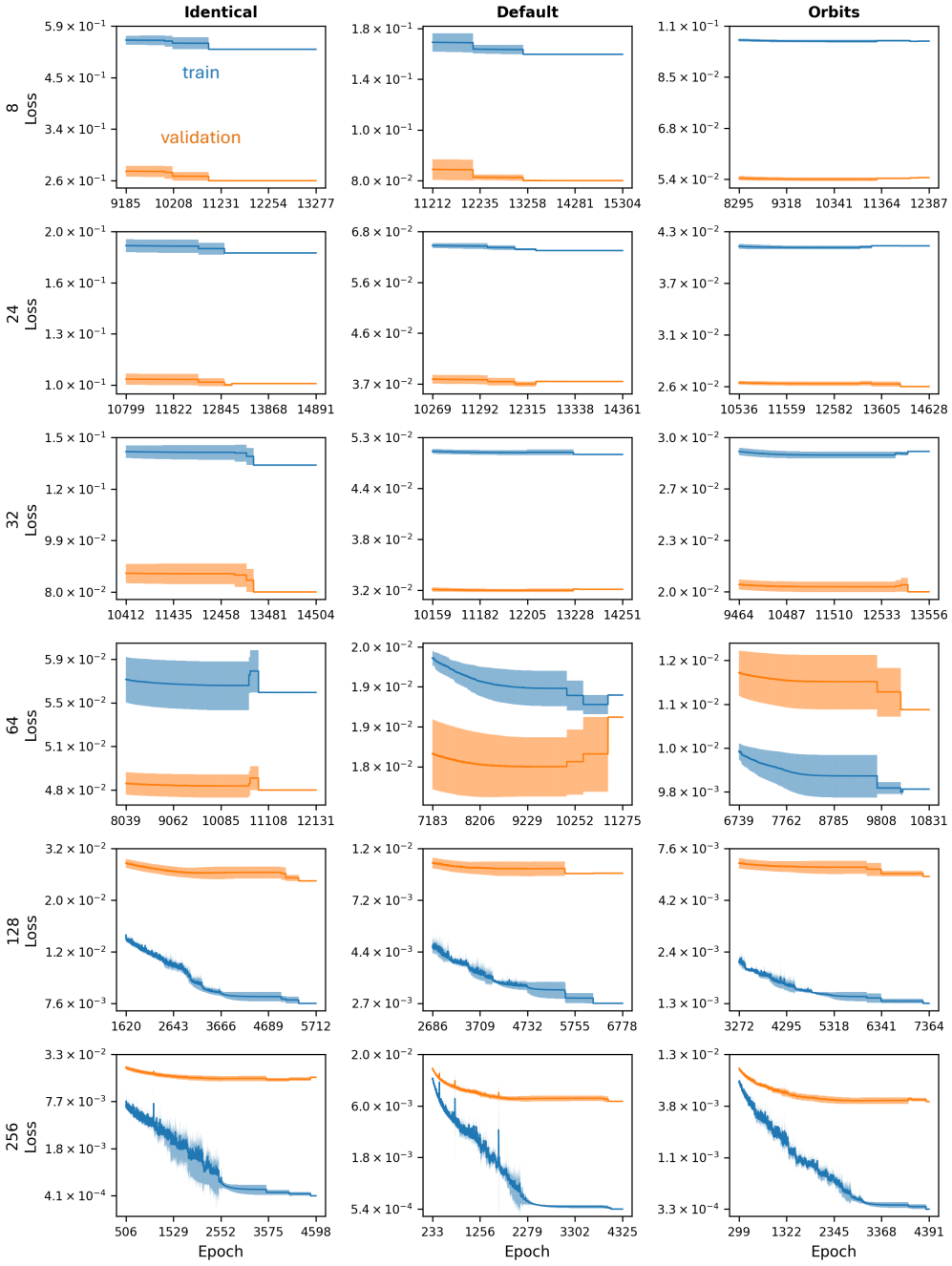


Figure C.4 Training and validation loss as a function of epoch for MACE models of the $CsPbI_3$ slab using three structural encodings: (first column) all atoms treated as identical, (second column) the default species representation, and (third column) an orbit-based representation in which distinct atomic environments are treated as separate atom types. For each encoding, solid lines show the mean training (blue) and validation (orange) loss across cross-validation folds, and shaded regions indicate the corresponding ± 1 standard deviation. The learning curves illustrate the convergence behaviour during optimization and the magnitude of the generalization gap between training and validation performance. The rows correspondence to different channels numbers in hidden irreducible representations.

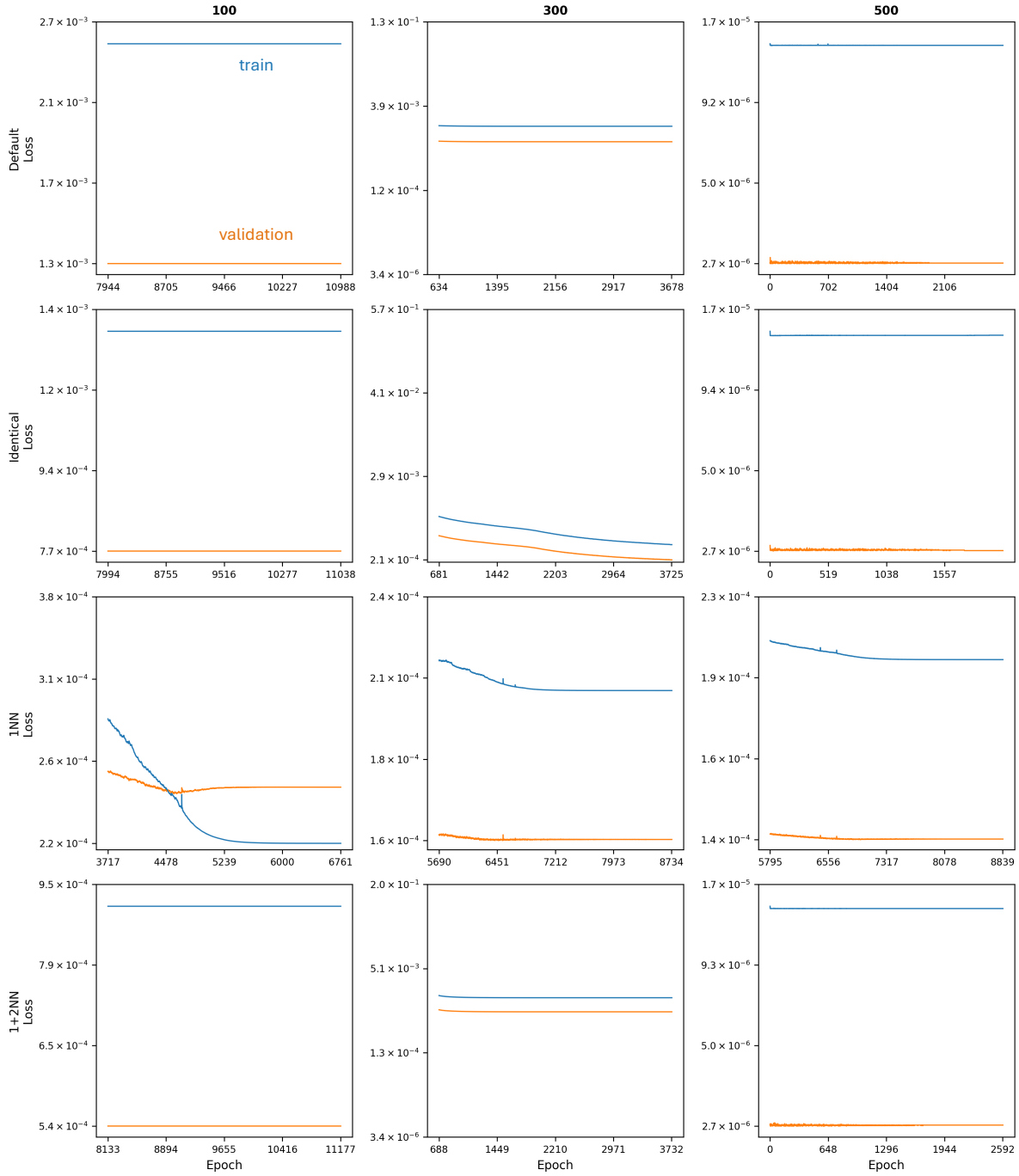


Figure C.5 Training and validation loss as a function of epoch for MACE models of the graphene with a pyridinic-N defect using three structural encodings: (first row) the default species representation, (second row) all atoms treated as identical, (third row) 1NN representation, and (forth row) 1+2NN representation. For each encoding, solid lines indicate the training (blue) and validation (orange) losses. The learning curves illustrate the convergence behaviour during optimization and the magnitude of the generalization gap between training and validation performance. The rows correspond to different train/validation set sizes.