

# UPGRADE REPORT

Submitted by **ALEXANDER PAUL HOLMAN**, to the University of Exeter as  
evidence of suitability to upgrade from the degree program of Masters of  
Philosophy in Physics to the degree program of Doctor of Philosophy in Physics,  
**23 JUNE 2025.**

This report is available for Library use on the understanding that it is copyright  
material and that no quotation from this report may be published without proper  
acknowledgement.

I certify that all material in this report which is not my own work has been  
identified and that any material that has previously been submitted and approved  
for the award of a degree by this or any other University has been acknowledged.

(Signature) .....

**ALEXANDER PAUL HOLMAN**

## Abstract

Material interfaces underpin the performance of a wide range of functional devices, from microelectronic transistors to catalytic systems. Their properties often emerge from interfacial effects not captured by bulk characterisation, making the prediction of interface stability a critical and computationally demanding task. While Density Functional Theory (DFT) provides accurate energetics, its high cost limits its applicability to large-scale screening. This study investigates whether machine-learned interatomic potentials (MLPs), specifically the MACE framework, can replicate DFT-derived rankings of interface favourability across a diverse set of semiconductor heterostructures.

Interfaces were generated using the ARTEMIS toolkit across three compositional stages: initial Si|Ge binaries, expanded Si|Ge binaries, and a broader C-Si-Ge-Sn set. Each interface structure was independently relaxed using both DFT and MACE. Energies were compared using rank correlation metrics (Spearman  $\rho$ , Kendall  $\tau$ ), absolute error measures (RMSE, MAE), and Top- $N$  overlap to assess rank preservation.

Results indicate that MACE predictive performance is system-dependent. Rank correlation was weak for silicon-only interfaces but improved markedly for chemically ordered alloy systems, such as SiGe. Cross-method evaluations suggested that structural relaxation, rather than energy evaluation, is the primary source of deviation between DFT and MLP predictions.

Future work will focus on training delta-learning models to correct systematic MLP errors using DFT data and on developing a predictive machine-learning interface (MLI) model based on slab features. The role of surface orientation and structural features in determining interface stability will also be examined.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation for the Research . . . . .	3
1.2	Research Objectives . . . . .	4
1.3	Scientific and Practical Significance . . . . .	6
<b>2</b>	<b>Background and Literature Review</b>	<b>8</b>
2.1	Defining Material Interfaces . . . . .	8
2.1.1	What is a material interface in the context of this research?	8
2.1.2	What physical or electronic properties are commonly affected by interfaces? . . . . .	11
2.1.3	Why are interfaces critical in material engineering, electronics, or energy systems? . . . . .	14
2.1.4	How do crystallographic orientations, phase boundaries, or defects influence interfaces? . . . . .	15
2.1.5	Examples of important technological applications involving interfaces . . . . .	17
2.2	Conventional Prediction Methods . . . . .	18
2.2.1	What are traditional computational approaches to modelling interfaces? . . . . .	19

2.2.2	Density Functional Theory (DFT) . . . . .	21
2.2.3	Other Methods . . . . .	23
2.2.4	What are the advantages and limitations of DFT, MD, and KMC? . . . . .	25
2.2.5	How are properties like adhesion energy or electron transport predicted? . . . . .	26
2.3	Emergence of Machine Learning in Materials Science . . . . .	28
2.3.1	How has machine learning been introduced into materials prediction? . . . . .	29
2.3.2	What types of models and data are being used? . . . . .	31
2.3.3	What are the strengths and limitations of ML in this field? . . . . .	32
2.3.4	Machine-Learned Interatomic Potentials (MLPs) . . . . .	33
2.4	Limitations of Current Approaches . . . . .	35
2.4.1	What are the current bottlenecks in interface prediction? . . . . .	37
2.4.2	Where do both ML and physics-based methods fall short? . . . . .	39
2.4.3	Are there data or generalisation issues? . . . . .	40
2.5	State-of-the-Art in Interface Prediction . . . . .	42
2.5.1	What recent work has been done in this area? . . . . .	43
2.5.2	How does my proposed approach compare and build upon it? . . . . .	45
2.5.3	Emerging trends . . . . .	46
<b>3</b>	<b>Computational Investigation of Interface Favourability</b>	<b>48</b>
3.1	Methodology . . . . .	48

3.1.1	Iterative Scope and Dataset Composition . . . . .	49
3.1.2	Interface Generation Protocol . . . . .	50
3.1.3	Relaxation Workflows . . . . .	50
3.1.4	Energy Evaluation Scheme . . . . .	51
3.1.5	Automation and Data Infrastructure . . . . .	52
3.1.6	Ranking and Performance Metrics . . . . .	52
3.2	Results and Analysis . . . . .	53
3.2.1	Overview of Rank Agreement . . . . .	53
3.2.2	Cross-Method Evaluation . . . . .	54
3.2.3	System-Specific Observations . . . . .	54
3.2.4	DFT vs MLP: Computational Cost Analysis . . . . .	56
<b>4</b>	<b>Next Steps</b>	<b>57</b>
4.1	Further Steps Towards Rank Preservation . . . . .	57
4.1.1	Motivation: Why Rank Preservation Is Central . . . . .	57
4.1.2	Comparison of Training Strategies . . . . .	57
4.1.3	Planned Benchmarking Criteria . . . . .	58
4.2	Surface Effects on Interface Favourability . . . . .	59
4.2.1	Surface Energy as a Proxy for Interface Stability . . . . .	59
4.2.2	Systematic Correlation Studies . . . . .	59
4.2.3	Implications for Interface Screening . . . . .	60
4.3	Crystallographic Structure and Interface Behaviour . . . . .	60
4.3.1	Symmetry and Orientation Effects . . . . .	60

4.3.2	Structure-Resolved Performance Analysis . . . . .	60
4.3.3	Extension to Broader Crystal Families . . . . .	60
4.4	Integration of RAFFLE into Structure Generation . . . . .	61
4.4.1	Motivation for RAFFLE Use: Enhanced Interface Diversity .	61
4.4.2	Current ARTEMIS-Only Pipeline: Limitations . . . . .	61
4.4.3	RAFFLE Capabilities: Refill Methods, Semi-Coherent Boundaries . . . . .	61
4.4.4	Planned Integration Workflow . . . . .	61
4.4.5	Validation and Comparison with ARTEMIS-Only Structures	62
4.5	Towards a Machine-Learned Interface (MLI) Predictor . . . . .	62
4.5.1	Objective: Predict Likely Interface Configurations from Material Pairs . . . . .	62
4.5.2	Feature and Descriptor Design . . . . .	62
4.5.3	Model Development Roadmap . . . . .	63
4.5.4	Outlook: Integrating MLI into Structure Generation Pipelines	63
<b>5</b>	<b>Conclusions</b>	<b>64</b>
5.1	Summary of Findings . . . . .	64
5.1.1	Overview of the study objectives and what was undertaken	64
5.1.2	Recap of major results: . . . . .	64
5.2	Interface Rank Preservation . . . . .	66
5.2.1	Effectiveness of MLPs in reproducing DFT interface energetics . . . . .	66
5.2.2	Top-N performance and rank correlations . . . . .	67

5.2.3	Strengths vs weaknesses . . . . .	67
5.3	Insights from Work Undertaken . . . . .	68
5.3.1	What was learned from each system . . . . .	68
5.3.2	Contrast perfect alloys vs heterostructures . . . . .	68
5.4	Challenges in Modelling Silicon Interfaces . . . . .	69
5.4.1	Discussion of why MLPs fail for Si: . . . . .	69
5.4.2	Open questions . . . . .	69
5.5	Broader Lessons and Limitations . . . . .	70
5.5.1	When can MLPs be trusted for interface screening? . . . . .	70
5.5.2	Necessity of retraining, delta-correction, or hybrid workflows . . . . .	70
5.6	Concluding Remarks . . . . .	70
5.6.1	What this work contributes to the field of ML-accelerated interface modelling . . . . .	70
5.6.2	Final note on significance for device and materials design . . . . .	71

# Chapter 1

## Introduction

Material interfaces are critical to the function and design of advanced solid-state systems. In semiconductors, they govern charge transport, carrier confinement, defect formation, and phase stability. In energy devices, they regulate ion exchange, chemical compatibility, and thermal response. In emerging heterostructures, interfaces can even induce novel behaviours not present in the constituent materials; such as encapsulated metastable phases or engineered Dirac cones. As device dimensions shrink and surface-to-volume ratios rise, the influence of interfaces grows, placing renewed emphasis on their accurate prediction and control.

Yet despite their importance, predicting which interfaces are energetically favourable remains a significant challenge. Interface stability is not solely determined by bulk lattice matching, but also by local bonding, chemical inhomogeneity, and atomic-scale reconstruction. The structural space is vast: each pair of surfaces can be stacked with arbitrary lateral shifts, terminations, and alignments, resulting in a combinatorial explosion of configurations. Strain relaxation, interlayer interactions, and charge transfer further complicate the energy landscape. Accurately assessing favourability across this space demands a balance between resolution and scale that conventional methods struggle to achieve.

First-principles techniques such as Density Functional Theory (DFT) remain the standard for computing accurate interfacial energies and relaxations. However, their high computational cost, scaling cubically with system size, renders exhaus-



tive sampling intractable for realistic supercells, which often comprise hundreds of atoms. This bottleneck precludes broad screening efforts, especially when exploring multiple chemistries, orientations, and stacking arrangements. As such, DFT is best suited to final validation rather than initial discovery.

To address this, the present project explores the use of machine-learned interatomic potentials (MLPs) for scalable interface screening. MLPs are data-driven models trained to reproduce DFT-level energies and forces, but at orders of magnitude lower cost. In particular, this work focuses on the MACE framework; an equivariant message-passing architecture that preserves rotational symmetries and captures local atomic interactions with high fidelity. By using MACE to rapidly evaluate large numbers of candidate interfaces, it becomes possible to rank structures by predicted stability and prioritise high-value configurations for subsequent DFT refinement.

This approach supports a hierarchical modelling workflow in which DFT and MLPs are used in tandem: MLPs for breadth, and DFT for depth. It enables rigorous testing of whether MLPs can replicate DFT-derived favourability rankings across a diverse set of semiconductor systems, including homostructures (e.g. Si|Si) and heterostructures (e.g. Si|Ge, Ge|C). Moreover, it lays the foundation for further improvements through techniques such as delta-learning, structure-based prediction models, and surface-to-interface generalisation.

The remainder of this report is structured as follows:

- **Chapter 2** introduces the theoretical foundations and relevant literature, covering interface energetics, DFT formalism, and the development of MLPs.
- **Chapter 3** presents a benchmarking study comparing MLP and DFT predictions of interface favourability, including statistical correlation analysis and system-specific case studies.
- **Chapter 4** outlines planned extensions to the current methodology, including the development of delta-corrected potentials and predictive models linking surface properties to interface behaviour.
- **Chapter 5** concludes with a discussion of findings, limitations, and broader

implications for autonomous materials modelling and the design of functional heterostructures.

Through this integrated modelling framework, the project seeks to reduce the cost and uncertainty of interface prediction while preserving the fidelity needed for scientific insight and technological relevance.

## 1.1 Motivation for the Research

Interfaces rarely form under ideal conditions. Instead, they emerge through complex, often non-equilibrium processes such as high-temperature annealing, epitaxial growth, and chemically driven surface reconstructions. These dynamics frequently trap systems in metastable configurations or generate local disorder, resulting in atomic arrangements that deviate significantly from those predicted by ideal lattice-matching rules. Such complexity challenges efforts to model interface behaviour using conventional approaches alone.

While Density Functional Theory (DFT) provides a reliable route to total energy prediction and structural relaxation, its unfavourable scaling with system size ( $\mathcal{O}(N^3)$  in electron count) severely limits its use in high-throughput workflows. In particular, DFT becomes prohibitive when applied to interface supercells exceeding a few hundred atoms, as required to capture long-range strain, multi-layer reconstruction, or compositional variation. Moreover, standard DFT approximations, such as semi-local exchange-correlation functionals, may fail to capture interfacial band alignments, polarisation effects, or weak van der Waals interactions critical to layered systems.

Machine-learned interatomic potentials (MLPs) offer a promising route to overcoming these limitations. By learning to emulate DFT-level energies and forces, MLPs like MACE provide a means to conduct large-scale structure relaxations and energy evaluations at dramatically reduced cost. This enables exploration of broader configurational spaces, including low-symmetry systems and extended defects, that would otherwise remain intractable. However, this advantage is only realised if the models generalise well across diverse interfacial environments.

The core motivation of this work, therefore, is to assess whether MLPs can replicate the DFT-derived ordering of interface favourability, defined here in terms of relaxed formation energy, across a wide range of semiconductor heterostructures. The benchmarking in Chapter 3 demonstrates that MLPs can often preserve the relative rankings of DFT energies, especially in chemically similar systems. However, notable discrepancies arise in cases such as pure-Si interfaces, where model performance degrades, prompting the need for refinement or retraining. These results motivate the delta-learning strategies explored in Chapter 4, which aim to correct systematic MLP-DFT residuals through supervised learning on interface-specific error distributions.

In addition to correcting prediction accuracy, this research investigates whether interface favourability can be inferred directly from structural or surface-level features. Chapter 4 proposes the use of learned models that map surface properties to interfacial energies, potentially bypassing the need for full supercell construction. This complements the structure-generation pipelines, based on ARTEMIS, that systematically enumerate viable interface candidates across various stacking shifts and Miller plane pairings.

Ultimately, this project is motivated by the need for more efficient and predictive tools to accelerate the design of functional heterostructures. By embedding MLPs within automated screening workflows and augmenting them with delta-learning corrections and structure-based heuristics, it becomes possible to bridge the gap between physical realism and computational tractability. In doing so, this research contributes to the broader aim of enabling scalable, data-driven exploration of complex interfacial systems; particularly in regimes where conventional symmetry-based rules or bulk-derived heuristics break down.

## 1.2 Research Objectives

This project aims to evaluate and enhance the use of machine-learned interatomic potentials (MLPs) for predicting the relative stability of material interfaces. By replacing expensive density functional theory (DFT) calculations in the early

stages of interface screening, MLPs offer a route to scalable evaluation across diverse chemical systems and stacking configurations.

Interfaces govern the functional behaviour of materials in microelectronics, energy storage, and optoelectronics, mediating phenomena such as band alignment, carrier confinement, and local defect formation. However, the configurational landscape of interface structures is combinatorially large. Even for lattice-matched systems, variations in Miller planes, interfacial shifts, and terminations can lead to significant changes in formation energy and electronic response. Traditional DFT methods, while accurate, are computationally prohibitive when applied to hundreds or thousands of such configurations.

This work builds on the integration of MLPs, particularly the MACE model, into high-throughput interface generation and evaluation pipelines. It focuses not only on energy prediction, but also on favourability ranking, robustness to system-specific errors, and hybrid workflows where MLPs are used as a pre-screening stage ahead of final DFT validation.

**The research addresses the following core objectives:**

- 1. Benchmark MLP-predicted rankings against DFT across diverse interfaces.** Using MACE-evaluated relaxations and energies, this study quantifies how well MLPs preserve the DFT ordering of interface favourability across multiple group-IV homostructures and heterostructures.
- 2. Diagnose system-specific discrepancies and explore delta-learning corrections.** Particular attention is given to cases, such as pure-Si systems, where MLP rank fidelity degrades. Chapter 4 proposes training delta models to correct these systematic energy errors.
- 3. Develop structure-based predictors of interface energy.** Surface and slab-level features (e.g. orientation, element type, symmetry, strain) will be used to train models capable of predicting favourability prior to full interface assembly, reducing reliance on supercell enumeration.
- 4. Integrate MLPs and predictors into automated workflows.** Tools such

as ARTEMIS and RAFFLE will be extended to incorporate MLP-based scoring or filtering, enabling efficient identification of favourable configurations before expensive DFT refinement.

In the long term, this approach supports scalable interface screening, facilitates the design of functional heterostructures, and lays groundwork for autonomous modelling pipelines. It targets regimes where standard coherence rules or bulk-derived heuristics fail, providing a data-driven alternative grounded in atomic-level energetics.

### 1.3 Scientific and Practical Significance

Material interfaces are central to the behaviour of functional solids, especially at the nanoscale where surface and interfacial effects dominate bulk properties. In semiconductors, they define charge transport pathways by setting potential barriers and band offsets. In batteries, they regulate ionic conductivity, chemical stability, and degradation kinetics. In photovoltaics and quantum devices, interfacial dipoles, tunnelling channels, and confinement effects underpin key device operations.

Subtle differences in stacking, registry, or atomic reconstruction can shift critical electronic properties by several electronvolts; altering conductivity, carrier lifetimes, or activation barriers. These shifts emerge from a complex interplay of interfacial strain, chemical bonding, and electronic redistribution, and are sensitive to both global structure and local environment. As such, predicting interface stability and functionality from atomic configuration alone remains a demanding task.

While Density Functional Theory (DFT) offers a robust *ab initio* framework for computing interfacial energies and electronic properties, its computational cost scales poorly with system size. Realistic supercells, often comprising hundreds of atoms, are therefore difficult to study in large numbers. DFT is well-suited to final validation, but ill-suited to exhaustive exploration across configurational space.

Machine learning (ML) models, particularly machine-learned interatomic potentials (MLPs), provide a way forward. When trained on DFT data, MLPs can reproduce formation energies and atomic forces at orders-of-magnitude lower cost, making them suitable for broad screening applications. Their integration into automated workflows, using tools such as ARTEMIS and RAFFLE, enables systematic generation and evaluation of interface candidates across crystallographic orientations, stacking shifts, and chemical combinations.

Importantly, the utility of ML does not stop at static energetics. Many applications require insight into how interfaces behave under external stimuli; electric fields, temperature gradients, or chemical exposure. Emerging interfacial phases, such as field-tunable dielectric layers or metastable heterostructures, challenge traditional heuristics and call for more adaptive and scalable modelling approaches.

This project contributes to that broader effort. It assesses the ability of MLPs, specifically MACE, to recover DFT-derived favourability rankings, identifies where discrepancies arise, and proposes delta-learning corrections to address them. It also lays groundwork for predictive models that estimate interface properties from top and bottom slabs, enabling pre-screening before full interface construction. By focusing on group-IV semiconductor systems and both homostructure and heterostructure cases, this work targets a class of technologically relevant interfaces where modelling uncertainty remains high and conventional rules often fail.

Ultimately, this research supports the advancement of data-driven, scalable methods for interface discovery; bridging the gap between computational feasibility and the physical fidelity required for next-generation materials design.

# Chapter 2

## Background and Literature Review

### 2.1 Defining Material Interfaces

This subsection establishes a foundational understanding of material interfaces, the boundary regions where two distinct phases, often different materials, meet. These interfaces are critical in determining the structural, electronic, and thermal behaviour of composite systems. In solid-state devices, subtle atomic arrangements and local electronic environments at interfaces can profoundly influence macroscopic properties. This is especially true in modern micro- and nanoscale technologies, where the continual miniaturisation of components has shifted the paradigm from bulk-dominated behaviour to interface-dominated performance. As devices shrink, interfaces no longer represent isolated regions but instead constitute a significant fraction of the material, making their precise definition and understanding essential for predictive modelling and materials design.

#### 2.1.1 What is a material interface in the context of this research?

A material interface is the region where two distinct solids meet and form a junction, commonly arising in heterostructures, grain boundaries, or thin film systems. These regions are not only geometrical boundaries but exhibit unique atomic,

electronic, and thermal properties that distinguish them from the bulk phases of the parent materials. The local structure at the interface can include strain fields, atomic reconstruction, and intermixing of constituent elements, all of which can significantly influence the behaviour of the composite system.

In this research, material interfaces are studied predominantly through *ab initio* methods, particularly Density Functional Theory (DFT), with a focus on predicting and analysing the structural and energetic properties of 2D|2D and 2D|3D junctions. Interfaces are modelled as systems with finite thickness that bridge two crystalline lattices. Depending on their geometric orientation, interfaces are categorised as either *stacked*, where a 2D layer rests atop a substrate, or *lateral*, where edge-on growth leads to in-plane bonding between dissimilar materials.

A critical motivation for interface studies lies in their role in determining key device properties such as band alignment, carrier transport, and thermal resistance. Band offsets at interfaces, often deviating from predictions by simple models like Anderson rule, are governed by local charge redistribution, dipole formation, and interface-specific states. The emergence of new electronic phases at interfaces, such as Ohmic contact formation or type-II band alignments, necessitates accurate modelling beyond bulk considerations.

Several methodologies have been developed to address the challenges of constructing realistic interface models. The ARTEMIS tool enables high-throughput generation of interface structures through lattice matching and surface termination selection, allowing systematic exploration of the complex interfacial phase space. Similarly, the RAFFLE method uses statistical learning and void-filling algorithms to construct interface phases that are energetically favourable and representative of plausible morphologies.

Interfaces also significantly influence phonon transport. They can reduce thermal conductivity compared to bulk values; an effect exploitable in thermoelectric applications. Interfacial scattering, acoustic impedance mismatch, and atomic-scale disorder all contribute to this behaviour, reinforcing the interface role as a functional component rather than a passive boundary.

Thus, in the context of this research, a material interface is not merely a geometric



boundary but an emergent, functionally distinct region. It governs critical physical properties, demands precise structural modelling, and serves as a fertile ground for material innovation and device engineering.

### Types of interfaces

Material interfaces can be broadly categorised by their crystallographic orientation, dimensionality, and degree of lattice registry. These classifications are instrumental in understanding interfacial phenomena and in designing simulation protocols that are physically realistic and computationally tractable.

**By geometry: stacked vs lateral.** Stacked interfaces occur when one material is deposited atop another, forming an out-of-plane junction. This is characteristic of many van der Waals heterostructures, where weak interlayer forces enable a range of stacking configurations and rotational alignments. In contrast, lateral interfaces arise when two materials are grown side-by-side, often leading to in-plane covalent bonding at the junction. These interfaces are especially relevant for in-plane heterostructures such as  $\text{MoS}_2|\text{WS}_2$  or graphene-based edge-junctions.

**By coherence: coherent, semi-coherent, incoherent.** Interfaces can also be distinguished by the degree of lattice match across the boundary:

- **Coherent interfaces** exhibit perfect registry between atomic planes, typically occurring when lattice mismatch is negligible. These interfaces preserve the translational symmetry across the junction and are often found in homoepitaxial or pseudomorphic systems.
- **Semi-coherent interfaces** accommodate lattice mismatch through periodic misfit dislocations. These dislocations relieve interfacial strain while maintaining partial registry, allowing larger domains of coherency interspersed with localised defects.

- **Incoherent interfaces** display no periodicity match across the interface, leading to structurally disordered junctions. These are often found in polycrystalline or amorphous/crystalline composites and can act as strong phonon or electron scattering centres.

**By chemical contrast: heterojunctions vs homojunctions.** Chemical composition also plays a defining role. *Homojunctions* involve structurally similar regions with subtle variations (e.g. doping or strain-modulated domains), while *heterojunctions* involve distinct materials with differing electronegativity, band gaps, and ionic/covalent character. Heterojunctions are central to semiconductor devices, solar cells, and catalysis, often giving rise to emergent interfacial dipoles, states, or reconstructed bonding motifs.

These classifications are not mutually exclusive; real interfaces may be simultaneously lateral and semi-coherent, or stacked but chemically graded. Understanding these distinctions is essential when generating or interpreting simulation results for both idealised and realistic interfacial systems.

### 2.1.2 What physical or electronic properties are commonly affected by interfaces?

Material interfaces influence a wide array of physical and electronic properties, often determining the functional behaviour of composite systems and devices. At the atomic level, interfaces introduce discontinuities in crystal symmetry, strain fields, and chemical potential that can lead to emergent phenomena not observed in bulk constituents. These effects are particularly pronounced in systems involving 2D materials, complex oxides, or strongly correlated electron systems.

#### **Band Alignment and Electronic Transport**

Perhaps the most critical electronic property influenced by interfaces is the band alignment between materials. Misalignment of conduction and valence band

edges creates energy barriers or wells that determine whether an interface behaves as a Schottky barrier, Ohmic contact, or heterojunction. These alignments are crucial for device operation in field-effect transistors, diodes, and photovoltaics. Interfacial dipoles, charge transfer, and electronic reconstruction can shift the effective band edges, often invalidating simple rules like Anderson model.

### **Charge Carrier Dynamics**

Interfaces often host localised states arising from atomic mismatch, bonding reconstruction, or impurity segregation. These can act as traps or recombination centres, significantly affecting carrier lifetimes and mobility. In 2D|3D systems, tunnelling, leakage, or enhanced recombination may occur depending on the barrier height and band offset at the junction.

### **Dielectric Response and Permittivity**

The dielectric properties of materials can be strongly altered at interfaces. In systems like  $\text{CaCu}_3\text{Ti}_4\text{O}_{12}$ , colossal permittivity arises not from the bulk crystal but from insulating grain boundaries and a semiconducting interior, exemplifying the Maxwell-Wagner effect. Interface-induced polarisation, charge accumulation, and ionic displacements all contribute to enhanced dielectric responses.

### **Thermal Conductivity**

Thermal transport across interfaces is typically suppressed due to phonon scattering, mass disorder, and acoustic impedance mismatch. This results in thermal boundary resistance (or Kapitza resistance), a key factor in thermoelectric design and heat management in microelectronics. In 2D heterostructures, the interface often acts as a dominant phonon scattering site, reducing the effective thermal conductivity below that of either constituent layer.

## Structural Stability and Phase Formation

Interfaces can stabilise new or metastable phases not accessible in bulk. For instance, the interface between  $\text{BaTiO}_3$  and  $\text{SiO}_2$  has been shown to promote the spontaneous formation of the fresnoite phase  $\text{Ba}_2\text{TiSi}_2\text{O}_8$ , altering both electronic and mechanical response. A more nuanced example is the graphene|MgO interface: while monolayer MgO confined between graphene layers often adopts a rocksalt-like structure, this is not simply a case of one phase being energetically favoured. Instead, the local environment, including charge transfer and encapsulation, stabilises the rocksalt phase relative to the expected hexagonal phase in unsupported monolayers. However, experimental and theoretical studies suggest that a mixed or hybrid phase may also emerge, with phase stability dependent on thickness and interface strain.

## Mechanical Properties and Elastic Moduli

Mechanical properties such as bulk modulus, shear strength, and adhesion energy are modified at interfaces due to altered bonding environments and lattice mismatch. These variations are important in the design of composite materials, flexible electronics, and nanoindentation studies.

## Optical Properties and Absorption

Interfaces affect the optical absorption spectrum by introducing new electronic states and modifying the joint density of states. This can lead to bandgap narrowing or broadening, shifts in refractive index, and enhancement of excitonic effects; features exploited in photodetectors and photovoltaic devices.

**In Summary** material interfaces act as multifunctional domains that modulate a broad spectrum of material properties. Their control and prediction are vital for the design of next-generation electronic, optoelectronic, and thermal devices.

### **2.1.3 Why are interfaces critical in material engineering, electronics, or energy systems?**

Interfaces are not merely passive boundaries between materials; they are active, functional regions that frequently control and even dominate the physical properties of the system. As materials and devices are engineered at ever smaller scales, the role of interfaces becomes more pronounced. In conventional bulk systems, interfaces may be limited to grain boundaries or inter-phase regions. However, in nanomaterials, layered heterostructures, and device-scale systems, interfacial regions can constitute a non-negligible fraction of the total volume or surface area. Their influence is therefore not ancillary but foundational to system performance.

In materials engineering, interfaces govern a wide array of properties including adhesion, corrosion resistance, mechanical strength, and diffusion kinetics. The strength and ductility of composite systems or coatings often depend on interfacial bonding and lattice compatibility. Tailoring interfacial chemistry or morphology, through methods such as surface functionalisation, alloying, or atomic layer deposition, is often essential for achieving desired macroscopic performance metrics.

In electronic devices, interfaces critically influence charge transport, band alignment, and contact resistance. Semiconductor heterojunctions, metal-semiconductor contacts, and dielectric boundaries are fundamental to the operation of transistors, diodes, and capacitors. Localised changes in atomic registry, roughness, or defect density at the interface can significantly alter tunnelling behaviour, barrier heights, and mobility. This is especially true in 2D/3D systems, where interfacial dipoles, charge redistribution, and electrostatic reconstruction result in complex electronic responses not captured by bulk models. As devices shrink to the nanometre scale, quantum confinement and interfacial electrostatics become dominant factors, necessitating atomically precise interface control.

Energy systems are similarly interface-dominated. In lithium-ion batteries, for example, the solid-electrolyte interphase (SEI) acts as a metastable interfacial layer that dictates both stability and charge transport. The performance and longevity

of solid-state batteries, fuel cells, and thermoelectric devices depend on the engineered properties of interfaces, which affect ionic conductivity, catalytic activity, thermal transport, and electronic insulation. In photovoltaics, interfacial band offsets between donor and acceptor layers, or between absorber materials and charge transport layers, govern charge separation efficiency and open-circuit voltage.

Overall, interfaces are not incidental features but active design parameters. They serve as sites of emergent phenomena, such as interfacial polarisation, bandgap renormalisation, and phonon scattering, that cannot be predicted from bulk properties alone. As such, mastering interfacial science is a gateway to developing high-performance materials and devices.

#### **2.1.4 How do crystallographic orientations, phase boundaries, or defects influence interfaces?**

Material interfaces are regions where two or more distinct phases meet, and their atomic-level structure plays a critical role in determining the physical, electronic, and thermal properties of the overall system. These interfaces are not merely passive boundaries; rather, they are active regions that can be engineered for specific functionality, particularly in devices involving electronic, optoelectronic, and energy materials. The structural features that influence interface behaviour include crystallographic orientation, the nature of the phase boundary, and the presence of defects.

##### **Crystallographic Orientation**

The relative orientation of crystals on either side of an interface can significantly affect the atomic registry, bond continuity, and symmetry breaking across the interface. Epitaxial alignment, where lattice planes across the interface maintain a well-defined relationship, typically results in low-strain, coherent interfaces. In contrast, a mismatch in orientation can lead to misfit dislocations or interfacial strain, influencing charge carrier transport, interfacial states, and mechanical sta-

bility. Orientation also affects the formation energy and electronic reconstruction at the interface, which can lead to phenomena like interface-induced states or altered band alignments.

### **Phase Boundaries**

When the interface forms between distinct structural phases, such as cubic and hexagonal polymorphs, unique bonding geometries and strain fields arise. For instance, in layered 2D|3D systems like  $\text{HfS}_2|\text{HfO}_2$ , the band alignment changes depending on whether the materials are stacked or laterally connected, with lateral interfaces showing increased reconstruction due to bonding mismatch between the different crystalline motifs. These reconstructions are not merely geometric; they result in different electrostatic and electronic behaviours, influencing band offsets and charge transfer dynamics.

### **Defects at Interfaces**

Defects, including vacancies, interstitials, dislocations, and grain boundaries, are often intrinsic to interface formation and can profoundly modify interfacial properties. They may serve as scattering centres, recombination sites, or sources of local doping. In grain boundaries of polycrystalline materials, for example, variations in local stoichiometry and strain can create regions of localised metallicity or enhanced dielectric response, as shown in systems exhibiting colossal permittivity. In some systems, such as CCTO ( $\text{CaCu}_3\text{Ti}_4\text{O}_{12}$ ), interfacial defects and microstructural features have been shown to dominate the macroscopic dielectric properties.

In computational studies, predicting the impact of such defects requires careful treatment of their spatial distribution, charge state, and interaction with surrounding lattice atoms. Modern approaches like the RAFFLE algorithm allow for machine learning-driven structure prediction that accounts for void-filling and local bonding environments, enabling more realistic models of defect-laden interfaces.

### 2.1.5 Examples of important technological applications involving interfaces

Material interfaces, where two distinct phases or materials meet, play a pivotal role in determining the performance and function of a wide array of modern technologies. These regions often exhibit structural, electronic, or thermal properties that differ markedly from those of the bulk constituents. Depending on their nature, interfaces may enhance or impede material performance. This subsection surveys several application domains in which understanding and engineering interfacial properties is essential.

In semiconductor devices, the interface between the gate dielectric and the semi-conducting channel is a critical design parameter. As traditional  $\text{SiO}_2$  dielectrics are replaced by high- $\kappa$  materials such as  $\text{HfO}_2$  to meet miniaturisation demands, understanding charge transfer and band alignment across interfaces like  $\text{HfS}_2|\text{HfO}_2$  becomes essential for ensuring device reliability and performance. Notably, in such 2D|3D heterostructures, the type and strength of bonding (e.g. van der Waals vs. chemical) directly influences charge redistribution and can cause significant deviation from band alignments predicted by Andersonrule.

In layered 2D electronics, especially those using transition metal dichalcogenides (TMDCs), the stacking geometry, layer orientation, and interfacial disorder all strongly influence the band alignment. These factors affect carrier dynamics, optical transitions, and switching behaviour. Traditional models such as Andersonrule have proven insufficient; corrections using terms such as  $\Delta E_T$  and  $\Delta E_{IF}$  have been developed to capture hybridisation and interface dipole contributions more accurately. These corrections allow for predictive modelling of heterostructure behaviour and support targeted band offset engineering.

Interfaces also govern performance in energy storage systems. In all-solid-state batteries, the interface between electrode and solid electrolyte dictates ion mobility, chemical stability, and space charge layer formation. Such interfacial processes can lead to resistive interphases or dendrite formation, reducing battery life and safety. Density Functional Theory (DFT) modelling has shown that interfacial reconstructions and electrostatic gradients can dominate device behaviour.



In thermoelectric materials, interfaces are employed to decouple electrical and thermal transport. Interface-induced phonon scattering suppresses thermal conductivity while preserving electronic conductivity; enhancing the thermoelectric figure of merit,  $ZT$ . Layered TMDCs and heterostructures with nanostructured interfaces exploit this mechanism to improve thermal management.

Finally, in optoelectronics and photovoltaics, interfaces play a central role in determining band bending, exciton dissociation, and charge carrier collection. In oxide-based devices, such as proposed all-oxide solar cells (e.g.  $\text{CaO}|\text{(Sn:Ca)}_{7:1}\text{O}|\text{TiO}_2$ ), interface energetics must be carefully tuned to support efficient charge separation and transport. The formation of intermediate phases or interfacial dipoles can critically affect performance, necessitating first-principles modelling for reliable device design.

Across these examples, it is evident that interfaces are not passive boundaries, but active, tunable regions that critically determine macroscopic functionality. Their predictive modelling, through DFT, machine learning potentials, or interface-specific structure search tools like ARTEMIS, remains a central challenge in modern materials design.

## 2.2 Conventional Prediction Methods

Understanding and predicting the behaviour of material interfaces requires a firm foundation in computational methods that are rooted in quantum and statistical mechanics. This section introduces the core approaches historically used to model interfaces at the atomic and mesoscale level, focusing primarily on first-principles methods and classical simulation techniques. Although each of these methods offers powerful insights, their accuracy, computational cost, and suitability for interface modelling vary substantially.

A primary tool in this field is *Density Functional Theory* (DFT), a quantum-mechanical framework capable of describing electronic structure with high precision. DFT is well-suited to calculating key interfacial properties such as formation ene.g. charge redistribution, and band alignment, making it an essential component

of most modern interface studies. Within this domain, software packages such as VASP have become widely adopted for their efficient implementations of the Kohn-Sham formalism.

However, DFTcubic scaling with system size makes it computationally prohibitive for large or disordered interface models. As a result, researchers also turn to *classical methods*, such as *Molecular Dynamics* (MD) and *Kinetic Monte Carlo* (KMC). MD enables the simulation of dynamic processes and thermal fluctuations over nanosecond timescales, whereas KMC offers access to much longer time evolution by modelling stochastic transitions between states. Each method has its own niche—MD is adept at capturing thermal effects and atomic rearrangements, while KMC is particularly valuable for studying diffusion and growth kinetics over experimental timescales.

These conventional approaches provide the groundwork for predicting critical interfacial properties—including adhesion ene.g. charge transfer behaviour, and electron or phonon transport. They remain indispensable despite their limitations, particularly when used in tandem. This section surveys these methods in turn, addressing their theoretical underpinnings, practical implementations, and their relative advantages and constraints in the context of interface science.

### **2.2.1 What are traditional computational approaches to modelling interfaces?**

Traditional computational approaches to modelling interfaces have historically focused on atomistic methods grounded in quantum mechanics or statistical physics. The most widely adopted of these are *first-principles* or *ab initio* methods, particularly Density Functional Theory (DFT), as well as Molecular Dynamics (MD) and Kinetic Monte Carlo (KMC). Each offers a different perspective on interfacial structure, energetics, and evolution.

## **Lattice Matching and Geometry-Driven Construction**

A foundational step in many interface modelling workflows is lattice matching between two crystalline materials. This involves aligning compatible Miller planes to minimise strain and identify viable supercell geometries. The method introduced by Zur et al. (1984) laid the groundwork for many subsequent algorithms used in tools such as METADISE and ARTEMIS, which systematically enumerate low-strain combinations, slips, terminations, and atomic registries. These approaches assume idealised surfaces and often neglect more complex processes such as surface reconstructions, chemical interdiffusion, or local disorder.

## **Density Functional Theory (DFT)**

DFT remains the most accurate and widely used quantum mechanical method for predicting interfacial properties. It solves the Kohn–Sham equations to determine the electronic ground state, providing access to total energies, charge densities, and band edge positions. DFT can reliably predict key quantities such as adhesion ene.g. charge transfer, and local band alignment. However, it scales cubically with system size ( $\mathcal{O}(N^3)$ ), limiting its application to relatively small supercells and coherent interfaces. For disordered systems or long-range relaxations, DFT becomes prohibitively expensive.

## **Molecular Dynamics (MD)**

MD simulates the atomic trajectories under classical Newtonian mechanics, governed by interatomic potentials. Classical MD is well-suited to studying thermal processes such as diffusion, defect migration, and strain relaxation on picosecond to nanosecond timescales. It can handle much larger system sizes than DFT but sacrifices electronic accuracy, particularly in capturing bond rearrangements or charge redistribution. Reactive MD and ab initio MD (e.g. Car-Parrinello) offer improved fidelity at increased computational cost.

### Kinetic Monte Carlo (KMC)

KMC provides a probabilistic approach to simulate rare events and long-term evolution, such as grain growth, interdiffusion, or defect clustering. It models a system as a sequence of stochastic transitions between predefined states, with rates typically derived from DFT or MD simulations. KMC can simulate processes on experimentally relevant timescales (up to seconds), but lacks atomic-resolution trajectories and requires a comprehensive catalogue of rate-limiting events.

### Summary

Together, these methods form the traditional computational toolkit for interface modelling. DFT provides high-accuracy predictions of interfacial energies and electronic structure; MD captures time-resolved atomistic behaviour; and KMC enables long-timescale simulations of kinetic evolution. While powerful individually, these methods are increasingly applied in tandem to span multiple length and time scales.

## 2.2.2 Density Functional Theory (DFT)

Density Functional Theory (DFT) is a first-principles quantum mechanical method used extensively to model the electronic structure of condensed matter systems, including molecules, solids, and, critically, material interfaces. It provides a variational framework in which the ground-state properties of a many-electron system can be obtained from a functional of the electron density  $\rho(\mathbf{r})$ , rather than from the many-body wavefunction  $\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$ . This simplification, formalised through the Hohenberg-Kohn theorems, underpins the modern implementation of DFT via the Kohn-Sham approach, where the interacting system is replaced by an auxiliary system of non-interacting electrons subject to an effective potential that captures Coulomb, exchange, and correlation interactions.

The total Kohn-Sham energy is given by:

$$E_{\text{KS}}[\rho] = T_s[\rho] + \int V_{\text{ext}}(\mathbf{r})\rho(\mathbf{r}) \, d\mathbf{r} + \frac{1}{2} \int \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, d\mathbf{r} \, d\mathbf{r}' + E_{\text{xc}}[\rho], \quad (2.1)$$

where  $T_s[\rho]$  is the kinetic energy of the non-interacting electrons,  $V_{\text{ext}}$  is the external (ionic) potential, and  $E_{\text{xc}}[\rho]$  is the exchange-correlation functional, the exact form of which remains unknown.

Common approximations include the Local Density Approximation (LDA), Generalised Gradient Approximation (GGA), and hybrid functionals. While LDA assumes a locally uniform electron gas and often leads to overbinding, GGA incorporates density gradients and improves accuracy for molecular and inhomogeneous systems. Hybrid functionals, such as HSE06, incorporate a portion of exact exchange and offer improved bandgap and localisation predictions at significantly greater computational cost.

DFT has become foundational in materials modelling due to its ability to predict not only total energies and optimised geometries, but also derived properties including electronic band structures, charge densities, dielectric responses, phonon spectra, and interfacial energetics. Its applications to interface modelling, however, are computationally constrained. Systems containing hundreds of atoms, often required to model misfit dislocations, strain relief, and interfacial disorder, can become prohibitively expensive, with computational scaling on the order of  $\mathcal{O}(N^3)$  with system size.

### **VASP and Practical Implementations**

All DFT simulations in this project are conducted using the Vienna Ab initio Simulation Package (VASP), a widely used plane-wave-based implementation of Kohn-Sham DFT. VASP employs the projector-augmented wave (PAW) method to model core-valence interactions, and supports a variety of exchange-correlation functionals including GGA-PBE and hybrid HSE06. Structural relaxations are typically performed with conjugate gradient algorithms until residual forces fall below a chosen threshold (e.g. 0.01 eV/Å).

Periodic boundary conditions are employed in all three spatial directions, necessitating the introduction of vacuum regions for slab models and sufficient supercell dimensions to minimise image-image interactions in interface simulations. For phonon and dielectric property calculations, VASP supports both finite displace-

ment and Density Functional Perturbation Theory (DFPT) approaches, though care must be taken when treating low-symmetry systems or low-frequency modes.

The simulation input is controlled via the INCAR, POSCAR, KPOINTS, and POTCAR files, with convergence typically assessed by k-point density and energy cutoff tests. Pulay and Broyden mixing schemes are used for charge density convergence, and dipole corrections may be employed for asymmetric slab geometries.

While VASP provides accurate and robust results for interfacial modelling, it remains computationally intensive. For extended or complex interfacial systems, this motivates the introduction of machine-learned potentials (MLPs), which aim to reproduce DFT accuracy at substantially lower cost; enabling exploration of larger supercells, disordered motifs, and thermodynamic ensembles that are otherwise intractable.

### 2.2.3 Other Methods

In addition to Density Functional Theory (DFT), two other classes of atomistic modelling, Molecular Dynamics (MD) and Kinetic Monte Carlo (KMC), have proven valuable for exploring material interfaces, particularly when temporal evolution or thermal effects are of interest.

#### Molecular Dynamics

Molecular Dynamics is a time-resolved simulation method in which the trajectories of atoms are computed by numerically integrating Newton equations of motion. Forces on each atom are derived either from empirical interatomic potentials (classical MD) or quantum-mechanical calculations (ab initio MD, e.g. Car-Parrinello or Born-Oppenheimer approaches). In interface science, MD is used to study thermal stability, strain relaxation, diffusion, and defect evolution at finite temperatures.

One advantage of MD is its capacity to resolve dynamic processes at the atomic

level, such as dislocation glide, grain boundary motion, or interfacial melting. However, timescales are severely limited: typical simulations span nanoseconds, far shorter than the seconds to minutes often relevant in experiments. Classical MD can scale to millions of atoms but sacrifices accuracy; *ab initio* MD is more accurate but typically limited to fewer than 1,000 atoms and short timescales due to computational cost.

For interface modelling, MD is particularly useful when studying thermally activated restructuring, interdiffusion, and local vibrational behaviour, especially under conditions such as annealing or irradiation. It also serves as a preliminary stage in workflows involving machine-learned potentials, providing dynamical data for training.

### **Kinetic Monte Carlo (KMC)**

Kinetic Monte Carlo methods simulate the time evolution of materials by selecting thermally activated events (such as diffusion, adsorption, or defect migration) based on their rates, derived from either experimental data or calculated energy barriers. Rather than resolving individual atomic vibrations, KMC focuses on probabilistic transitions between states, making it well-suited to model slow processes over long timescales, up to seconds or beyond.

KMC has found particular utility in modelling interface evolution under conditions of growth, ion exchange, oxidation, and defect diffusion. In grain boundary studies, for example, KMC can capture the gradual restructuring of interfaces under chemical potential gradients or during irradiation recovery.

Its main limitation is the requirement for a comprehensive event catalogue with well-characterised energy barriers and transition rates. These are often obtained from DFT or empirical fitting, and inaccuracies in this input can compromise the reliability of the predictions. Furthermore, KMC lacks atomic-resolution trajectories and cannot directly model phonon-mediated transport or structural vibrations.

## Summary

Both MD and KMC complement DFT in the modelling of interfaces. While DFT offers high-fidelity energetics and electronic structure, MD enables atomistic dynamics over nanoseconds, and KMC extends the accessible timescale to macroscopic phenomena. As such, hybrid approaches are increasingly common, with DFT used to parameterise MD or KMC models, or machine learning techniques introduced to accelerate sampling across scales.

### 2.2.4 What are the advantages and limitations of DFT, MD, and KMC?

Each computational method traditionally employed in the modelling of interfaces, Density Functional Theory (DFT), Molecular Dynamics (MD), and Kinetic Monte Carlo (KMC), offers a distinct balance between accuracy, computational efficiency, and physical scale. While often applied in isolation, their complementary regimes suggest the need for hybrid or hierarchical approaches in the study of real interfaces.

**Density Functional Theory (DFT)** is the most widely used *ab initio* method for predicting atomic-scale properties at interfaces. It provides access to ground-state energies, relaxed geometries, charge densities, and band alignments with a strong quantum mechanical foundation. Crucially, DFT is capable of modelling materials across the periodic table and is the backbone of modern electronic structure prediction. Its limitations, however, are twofold. Firstly, the method scales poorly with system size, typically as  $\mathcal{O}(N^3)$  with respect to the number of electrons or basis functions, making simulations involving more than a few hundred atoms prohibitively expensive. Secondly, conventional exchange-correlation functionals (e.g. GGA) systematically underestimate band gaps and cannot capture excited-state dynamics without costly extensions such as GW or time-dependent DFT.

**Molecular Dynamics (MD)** offers time-resolved access to thermally activated processes and interfacial dynamics. Classical MD, using empirical or machine-



learned interatomic potentials, enables simulation of thousands to millions of atoms over nanosecond to microsecond timescales. It is well-suited to studying lattice relaxation, phonon scattering, or diffusion across interfaces. However, MD inherits limitations from its force field or potential model: empirical potentials may not transfer well across bonding environments, while high-fidelity machine-learned potentials require extensive and chemically diverse training data. Furthermore, standard MD remains constrained to accessible timescales, making it ill-suited for capturing rare events such as defect migration, reconstruction, or long-term phase evolution.

**Kinetic Monte Carlo (KMC)** bridges the gap between atomistic fidelity and long-timescale evolution by explicitly simulating stochastic processes with predefined transition rates. It is particularly useful for modelling diffusion, growth, or reaction networks at interfaces. The method can span experimental timescales (seconds to hours), offering a unique perspective on kinetic stability and metastability. The key limitation lies in its dependency on an accurate catalogue of possible events and associated rates. These must be either calculated from first principles or parametrised experimentally, introducing significant uncertainty and limiting transferability to novel systems.

**In summary** DFT is unrivalled in electronic accuracy but severely limited in scale. MD extends spatial and temporal reach at the cost of potential fidelity, while KMC pushes further still into kinetic regimes but demands well-characterised reaction landscapes. The combined use of these methods, e.g. DFT-informed MD or KMC parametrised by *ab initio* barriers, has emerged as a powerful paradigm for interface modelling across scales.

### 2.2.5 How are properties like adhesion energy or electron transport predicted?

The accurate prediction of interfacial properties, such as adhesion energy and electron transport, is central to understanding and designing functional materials. These properties emerge from complex interplays of bonding, charge distri-

bution, and band alignment at the interface, and are typically predicted through first-principles simulations based on density functional theory (DFT). Increasingly, these predictions are augmented or accelerated using machine-learned inter-atomic potentials (MLPs).

**Adhesion energy** is a thermodynamic quantity describing the work required to separate two materials at their interface. It is commonly computed from the total energies of the isolated constituents and the relaxed interface configuration. Within DFT, the adhesion energy  $E_{\text{adh}}$  is given by:

$$E_{\text{adh}} = \frac{1}{A} (E_{\text{slab1}} + E_{\text{slab2}} - E_{\text{interface}}),$$

where  $A$  is the interface area, and the energies refer to fully relaxed geometries of the individual slabs and the interface supercell. For 2D|2D lateral heterostructures, where the interface is line-like rather than planar, a line energy formulation in eV/Å is typically used to account for the dimensionality of bonding.

**Electron transport** across interfaces is more nuanced and often requires evaluation of both band alignment and carrier mobility. Band alignment can be predicted via several approaches. The simplest, Anderson rule, aligns vacuum levels to estimate offsets from electron affinities or ionisation potentials. However, this often fails to capture interface-specific effects such as charge transfer, built-in fields, or hybridisation. More accurate predictions therefore rely on self-consistent DFT calculations that extract the local electrostatic potential across the heterostructure to determine band offsets directly.

For example, the offset between conduction or valence band edges across an interface can be obtained by aligning the macroscopic average electrostatic potentials of the constituent slabs with that of the full interface. This method accounts for relaxation, dipole formation, and interfacial reconstruction. In cases where electronic coupling is strong, or the valence states are spatially delocalised across layers (as in type-II or type-III band alignment), transport characteristics such as charge separation, carrier confinement, or tunnelling probability can change significantly.

Additionally, charge transport properties such as electrical conductivity, Seebeck coefficient, or mobility may be approximated from the band dispersion near the Fermi level using Boltzmann transport theory, or more rigorously via non-equilibrium Greenfunction (NEGF) methods, especially for low-dimensional or nanoscale interfaces.

Overall, both adhesion and electronic transport predictions depend on the level of theory and the nature of the interface. While DFT remains the most widely used method for accuracy, its computational cost limits system size. For larger systems or rapid screening, surrogate models such as machine-learned potentials or analytical approximations may be employed, with the understanding that they must be carefully validated against reference calculations.

## **2.3 Emergence of Machine Learning in Materials Science**

Over the past decade, machine learning (ML) has emerged as a transformative tool in materials modelling, offering a route to accelerate predictions and expand accessible system sizes beyond those tractable by first-principles methods. Traditional computational techniques, such as Density Functional Theory (DFT), Molecular Dynamics (MD), and Kinetic Monte Carlo (KMC), have established themselves as accurate and rigorous, but they often scale poorly with system size or configurational complexity. This limitation becomes particularly acute in the modelling of interfaces, where subtle geometric, chemical, and electronic variations can dramatically affect material behaviour. ML offers a complementary paradigm, enabling the construction of surrogate models trained on high-fidelity data, which can interpolate or even extrapolate to predict key material properties across vast configuration spaces.

This section introduces the growing role of ML in materials prediction, highlighting its application in interface modelling and the broader landscape of data-driven approaches. Subsequent subsections will examine the types of models in use, including classical regressors and deep neural architectures, alongside the na-

ture of training data, typical input descriptors, and validation strategies. Special attention will be given to machine-learned interatomic potentials (MLPs), particularly those with demonstrated success in interfacial contexts such as MACE and ChgNet. Finally, we reflect on the benefits and ongoing limitations of ML methods in materials science, including issues of transferability, uncertainty quantification, and integration into automated modelling pipelines.

### **2.3.1 How has machine learning been introduced into materials prediction?**

Machine learning (ML) has been introduced into materials prediction as a complementary tool to traditional first-principles methods, offering improved scalability, predictive flexibility, and the potential for workflow automation. Its adoption has been driven by limitations in conventional approaches such as Density Functional Theory (DFT), particularly in capturing the structural and chemical complexity of real-world systems such as interfaces.

#### **Computational Motivation**

First-principles methods scale poorly with system size, typically as  $\mathcal{O}(N^3)$ , where  $N$  is the number of electrons or basis functions. This makes simulations of large or disordered systems, especially those involving interfaces with long-range strain relaxation, misfit dislocations, or extended defects, computationally prohibitive. Machine-learned interatomic potentials (MLPs) such as MACE or ChgNet offer near-DFT accuracy at a fraction of the cost, enabling simulations with thousands of atoms.

#### **Exploration of Complex Configuration Spaces**

ML enables efficient exploration of vast configurational landscapes, where brute-force DFT sampling would be infeasible. This is particularly important in interface modelling, where multiple stacking orders, terminations, and slip configurations

can exist. ML-driven screening allows researchers to rapidly identify promising candidates for further evaluation.

### **Initial Integration Pathways**

Early ML applications focused on surrogate models trained to replicate DFT-derived energies, forces, or properties. Descriptor-based regression models, using quantities such as coordination numbers, Bader charges, or bond lengths, were employed to predict interfacial stability or band alignments. These approaches have evolved toward end-to-end learning frameworks using physically-informed descriptors or graph-based representations.

### **Workflow Integration**

ML is now embedded into iterative prediction–generation–evaluation cycles. For example, predicted interface favourability informs structure generation (e.g. stacking vectors or surface terminations), which are then evaluated by DFT or ML surrogates. The results feed back into model training, supporting a self-refining loop. Tools such as ARTEMIS benefit from such integration.

### **Modern Model Types**

More recent work employs graph neural networks (GNNs), such as MACE or NequIP, which are equivariant to Euclidean transformations and thus well-suited for atomistic systems. These models learn directly from atomic positions and species, avoiding the need for handcrafted features while preserving symmetry constraints essential for interface modelling.

### **Toward Autonomy**

Ultimately, ML enables a shift from static modelling pipelines to more autonomous workflows. When coupled with structure-generation protocols and physical con-

straints, ML can be used not only to accelerate predictions but to guide the discovery of novel interfacial phenomena in mixed-dimensional and disordered materials.

### 2.3.2 What types of models and data are being used?

Recent advances in machine learning (ML) have introduced a diverse range of models for predicting materials properties, including those relevant to interface stability, e.g. and structure. These models can be broadly categorised into classical learning algorithms and deep learning approaches. Classical methods include random forests, gradient boosting, support vector machines (SVMs), and Gaussian process regression (GPR), the latter being particularly valuable for small datasets due to its uncertainty estimation capability. Deep learning architectures include feedforward neural networks and, more notably, graph neural networks (GNNs), which are well-suited for materials modelling due to their ability to represent variable-sized atomic graphs with complex bonding topologies.

The input to these models typically consists of numerical representations of atomic configurations known as descriptors. Geometric descriptors include bond lengths, coordination numbers, angular distributions,  $n$ -body distribution functions, and radial symmetry functions; many of which feature prominently in Behler-Parrinello-type neural networks. Chemical descriptors such as electronegativity differences and Bader charges capture bonding character and charge transfer. Electronic descriptors like partial density of states (PDOS) and local work functions are relevant for predicting band alignment and interfacial transport properties. Interface-specific features, such as interfacial roughness or vacuum level offsets, are particularly crucial in 2D/3D and mixed-dimensional heterostructures.

To support learning across such heterogeneous and high-dimensional features, symmetry-invariant and scale-independent representations have become increasingly important. These ensure that models generalise across materials with different atom counts or lattice sizes. Dimensionality reduction methods and learned embeddings (e.g. via autoencoders or message passing) are frequently used to compress structural information into a lower-dimensional latent space optimised

for prediction.

The datasets used to train such models vary in size and fidelity. They often derive from high-throughput DFT calculations available in public repositories such as the Materials Project or AFLOW. In cases targeting interfaces, the training data may originate from internal simulations using structure generation pipelines like ARTEMIS or RAFFLE, where candidate structures are pre-screened by DFT or empirical methods for energetic viability.

The combination of curated descriptors, structured data from known interfaces, and model architectures sensitive to local and long-range atomic interactions underpins the growing success of ML in materials interface prediction.

### **2.3.3 What are the strengths and limitations of ML in this field?**

Machine learning (ML) has emerged as a promising strategy for accelerating materials modelling, particularly in the context of complex interfacial systems. Its strengths lie chiefly in efficiency, scalability, and adaptability. ML models, once trained, can rapidly approximate formation energies, surface reactivities, and stability metrics for thousands of configurations, many of which would be prohibitively expensive to assess using first-principles methods alone. This computational efficiency enables exploration of broader configurational and chemical spaces, including systems with disorder, stacking faults, or metastable structures that pose challenges for traditional simulations.

Another key advantage is the capacity of ML to capture subtle, non-linear correlations in high-dimensional datasets. When used in conjunction with tools such as ARTEMIS or RAFFLE, ML can assist not only in predicting energetics, but also in prioritising candidate geometries during structure generation. Moreover, deep learning approaches such as graph neural networks (GNNs) and equivariant message-passing architectures have demonstrated good performance on systems with varied local coordination environments, as frequently found at 2D|3D or lateral heterointerfaces.

However, ML approaches also exhibit critical limitations. A foremost concern is

their dependence on the training dataset: models typically interpolate well within the domain of known examples but generalise poorly to unseen chemistries or extreme geometries. This limits their reliability when extrapolating to novel material classes or interface types, unless training sets are curated with care to ensure coverage of relevant chemical and structural motifs.

Another limitation is interpretability. While classical models (e.g. those based on density functional theory) yield physically grounded outputs, such as charge densities or band structures, many ML models operate as statistical black boxes. Though efforts are underway to integrate physically meaningful descriptors (e.g. *n*-body distribution functions, local work function, or Bader charges), ensuring transparency and robustness remains a work in progress.

Finally, ML alone cannot fully replace physics-based methods. For systems exhibiting emergent interfacial phases, electronic reconstructions, or coupled phonon-electron effects, high-fidelity methods like DFT or many-body perturbation theory remain necessary for validation and refinement. As such, a hybrid strategy is often advocated; using ML to pre-screen candidate structures and flag configurations for further high-accuracy assessment.

**In summary** ML constitutes a valuable addition to the materials modelling toolkit, particularly for screening and guiding interface prediction. Yet its performance remains contingent on data quality and physical context, and it is best deployed in tandem with more rigorous electronic structure methods.

### 2.3.4 Machine-Learned Interatomic Potentials (MLPs)

Machine-learned interatomic potentials (MLPs) have emerged as a powerful alternative to classical force fields and first-principles methods for atomistic simulations. By learning from high-fidelity data, typically generated from Density Functional Theory (DFT), these models aim to reproduce the accuracy of *ab initio* methods while offering orders-of-magnitude reductions in computational cost. This is particularly valuable for interface modelling, where large supercells and



long-range relaxation effects often render direct DFT approaches intractable.

## MACE

The MACE (Many-body Attention and Correlation Equivariant) model represents a recent advancement in equivariant message passing neural networks (MPNNs), tailored for high-accuracy and scalable force field construction. Unlike earlier MPNNs which rely primarily on two-body invariant messages, MACE employs many-body interactions and equivariant tensor contractions to capture angular and chemical complexity with improved expressivity and data efficiency.

One notable strength of MACE is its capacity to reach state-of-the-art accuracy on diverse benchmarks, including rMD17, 3BPA, and AcAc, while requiring only two message passing iterations, due to its use of high-body-order messages. For instance, in extrapolation tasks involving out-of-distribution molecular geometries or high-temperature trajectories, MACE (particularly with  $L = 2$ ) has demonstrated superior performance and robustness relative to competing models such as NequIP and BOTNet.

Critically for materials modelling, MACE supports parallel evaluation across many GPUs by limiting its receptive field through local message construction and efficient tensor operations. This makes it attractive for large-scale interface studies involving thousands of atoms; especially where relaxation and strain distribution across the interface play a pivotal role.

## ChgNet

ChgNet is a neural network potential that incorporates both atomic positions and charge distributions into its prediction framework, thereby explicitly learning charge transfer phenomena. This is particularly relevant for interfaces where electrostatics, dipole formation, and local charge accumulation significantly affect electronic properties such as band alignment.

While detailed benchmark results for ChgNet are less mature compared to MACE,

its underlying design philosophy, embedding local charge environments directly into the model, offers a complementary advantage. For instance, ChgNet could be better suited to capturing interface-induced dipoles, charge spillover, or mixed-valence behaviour, which are challenging to encode via purely structural descriptors.

### **Use in Interface Modelling**

In the context of this project, MLPs like MACE and ChgNet offer distinct but synergistic advantages. MACE, with its speed, high accuracy, and symmetry-respecting architecture, is well-suited for structure relaxation and force prediction across multi-material junctions. ChgNet, on the other hand, may provide more faithful predictions of electronic interface effects arising from local charge redistribution.

Nonetheless, these models are not without limitations. MLPs depend heavily on the representativeness of their training data and can struggle to generalise to unseen chemistries or dissimilar interface environments. Therefore, hybrid workflows that use MLPs for broad structural exploration, followed by targeted DFT recalculations for ambiguous configurations, are increasingly being adopted.

The eventual goal is to integrate MLPs into an active learning pipeline alongside tools like RAFFLE or ARTEMIS, facilitating the generation, screening, and optimisation of candidate interface structures at scale.

## **2.4 Limitations of Current Approaches**

Despite notable advances in both physics-based and machine learning (ML) approaches to materials modelling, several persistent limitations constrain the predictive fidelity, scalability, and generalisability of current methods; particularly when applied to realistic interface systems. This section outlines three critical domains in which these limitations manifest, motivating the hybrid methodologies explored later in this report:

- **What are the current bottlenecks in interface prediction?**

Interface prediction remains an inherently multiscale problem, combining combinatorial complexity at the atomic level with long-range elastic and electrostatic effects. First-principles approaches such as density functional theory (DFT) offer high accuracy but scale cubically with system size, making them infeasible for large or disordered systems typical of interfaces. Even with machine-learned interatomic potentials (MLPs), exploring configurational space, e.g. over terminations, interlayer slips, and misfit reconstructions, requires extensive sampling and often yields flat energy landscapes with many metastable minima.

- **Where do both ML and physics-based methods fall short?**

DFT underestimates band gaps and struggles to capture emergent interfacial phenomena such as dipole formation, phase reconstruction, or charge transfer without recourse to hybrid functionals or many-body perturbation theory. Conversely, ML models, especially those trained on bulk or idealised configurations, exhibit limited transferability to chemically complex, strained, or disordered interfaces. Without embedded physical constraints or domain-specific descriptors, ML models often fail to resolve interfacial states, misfit dislocations, or partial intermixing.

- **Are there data or generalisation issues?**

Interface-specific datasets are significantly less developed than those for bulk crystals, due in part to the vast configurational diversity of junctions and the lack of standardised repositories. This data sparsity hampers the ability of ML models to generalise across terminations, chemistries, or stacking motifs. Additionally, many early ML workflows lack symmetry-invariant and scale-independent representations, increasing the risk of overfitting to specific atom counts or lattice types.

These limitations underscore the need for hybrid workflows that integrate ML potentials (e.g. MACE, ChgNet) for large-scale screening, DFT for high-fidelity validation, and structure-generation tools such as ARTEMIS or RAFFLE for efficient sampling of the configurational landscape. Such frameworks aim to balance accuracy, efficiency, and transferability in modelling the behaviour of realistic inter-

facial systems.

### 2.4.1 What are the current bottlenecks in interface prediction?

Despite considerable advances in computational materials science, reliable prediction of interface structures remains a fundamental bottleneck. This difficulty stems from the intrinsic complexity of real interfaces, which are often chemically heterogeneous, strain-mediated, and structurally disordered. The configurational degrees of freedom, ranging from slip vectors, surface terminations, and intermixing patterns to stacking sequences and misfit accommodation, create a vast structural landscape that resists exhaustive exploration via conventional methods.

**Configurational Complexity.** At the atomic scale, even between lattice-matched materials, variations in terminations, registry shifts, and interlayer relaxations produce an exponential number of possible configurations. The search space rapidly becomes unmanageable, especially in systems requiring large supercells to capture long-range strain or disorder. This renders brute-force density functional theory (DFT) calculations infeasible for comprehensive screening.

**Computational Cost.** DFT, while accurate and chemically grounded, exhibits cubic scaling with respect to the number of electrons or basis functions, making simulations of interfaces with hundreds to thousands of atoms prohibitively expensive. Moreover, structural relaxations in such systems often converge poorly due to shallow energy gradients and multiple metastable minima. These numerical instabilities pose a challenge even for single configuration evaluations.

**Lack of Generalisable Descriptors.** Unlike bulk systems, for which symmetry and stoichiometry often impose constraints, interfaces exhibit broken symmetries, compositional gradients, and non-uniform electrostatics. There is no broadly applicable analogue to Anderson's rule for predicting interfacial band alignment, necessitating ad hoc corrections such as  $\Delta E_{\Gamma}$  and  $\Delta E_{\text{IF}}$  that work only for narrow material classes.

**Limitations of Current Structure Generation Tools.** Automated interface builders like ARTEMIS and RAFFLE enable systematic generation of interface candidates but are typically restricted to heuristic approaches such as rigid lattice matching or void-filling. These tools often fail to account for elastic mismatch, surface reconstructions, or stacking instabilities. Furthermore, the lack of integration with energetic feedback loops results in large pools of candidates requiring expensive post hoc evaluation.

**Transferability of Machine-Learned Potentials.** Machine-learned interatomic potentials (MLPs), such as MACE and ChgNet, offer pathways to accelerate structural screening at near-DFT accuracy. However, their performance is contingent upon the representativeness of the training data. Interfacial environments frequently exhibit local chemistries and coordination motifs absent from bulk-derived datasets. As such, model extrapolation can lead to large errors, particularly in strained, defected, or reactive regimes.

### **Summary of Bottlenecks.**

- The combinatorial complexity of plausible interface configurations.
- High computational cost and poor scaling of DFT-based relaxations.
- Absence of generalised physical descriptors for predicting interfacial properties.
- Limited adaptability of current structure generation frameworks.
- Transferability limitations of MLPs in interfacial chemical environments.

Overcoming these bottlenecks likely requires hybrid workflows that combine MLP-based structural exploration with targeted DFT refinements, supported by uncertainty quantification and descriptor-led screening. Such frameworks are increasingly positioned as a pathway to scalable, realistic, and physics-informed interface prediction.

## 2.4.2 Where do both ML and physics-based methods fall short?

Despite significant progress in both first-principles and machine learning (ML) approaches, key limitations persist in the predictive modelling of material interfaces. These limitations arise not only from computational constraints but also from intrinsic mismatches between real interfacial complexity and the assumptions underpinning current models.

Physics-based methods, most notably Density Functional Theory (DFT), provide high accuracy for local bonding environments and band structure prediction. However, the computational scaling of standard Kohn-Sham DFT—typically  $\mathcal{O}(N^3)$  with respect to system size—renders large or disordered interface simulations intractable without approximation. Even when feasible, DFT struggles to capture extended features such as long-range strain relaxation, misfit dislocations, and spatially diffuse defect states that dominate realistic interface behaviour.

Machine learning potentials (MLPs), such as those based on graph neural networks or equivariant message passing architectures, have emerged as promising tools to extend predictive power to larger systems at near-DFT accuracy. Nonetheless, they too face structural limitations. MLPs require extensive, representative training datasets, typically derived from DFT calculations, and their performance deteriorates when applied to configurations involving unseen chemistries, surface reconstructions, or high strain regimes. Furthermore, many ML models lack rigorous uncertainty quantification, making it difficult to detect failure modes in out-of-distribution predictions.

Both DFT and ML approaches also fall short in capturing emergent interfacial phenomena such as charge transfer, interface-induced dipoles, or the formation of new interfacial phases. These effects are often strongly dependent on the local atomic registry and chemical heterogeneity, and cannot be inferred from the properties of the isolated constituents alone. Notably, simple models like Anderson rule have been shown to systematically fail in 2D and 2D|3D heterostructures, necessitating corrections such as  $\Delta E_{\Gamma}$  and  $\Delta E_{\text{IF}}$  that explicitly depend on the atomic-scale interface structure and electrostatics.

**In summary** while ML and physics-based models have complementary strengths, neither is sufficient in isolation for the reliable prediction of interfacial behaviour. Their limitations motivate the development of hybrid workflows, such as embedding physical constraints into ML architectures, or coupling MLP-based screening with selective DFT refinement. These approaches aim to achieve a balance between scalability, accuracy, and generalisability—a balance that remains at the forefront of interface modelling challenges.

### 2.4.3 Are there data or generalisation issues?

The predictive performance of machine learning (ML) models for interfacial systems is fundamentally constrained by data availability, quality, and representational fidelity. While ML models have achieved notable success in bulk materials prediction, their application to interface science introduces distinct challenges; chief among them data scarcity, limited transferability, and inadequate descriptor design.

#### Data scarcity and imbalance

In contrast to bulk crystals, where structured repositories like the Materials Project or AFLOW provide extensive coverage, interface-specific datasets remain sparse. This is due to the combinatorial explosion of viable terminations, slip vectors, and orientations across material pairs, especially in mixed-dimensional systems. Existing databases tend to overrepresent a narrow set of lattice-matched, technologically prominent interfaces, introducing bias and hindering the model's capacity to generalise across broader chemical or structural classes.

todo

#### Transferability and chemical diversity

Machine-learned potentials (MLPs) such as MACE and ChgNet often fail to extrapolate to configurations that deviate from the training distribution; particularly

at strained, chemically heterogeneous, or electronically reconstructed interfaces. The emergence of site-specific hybridisation, localised charge transfer, and dipole formation presents physics not typically encountered in bulk datasets. Without explicit exposure to such environments during training, even state-of-the-art equivariant models exhibit elevated uncertainty and reduced fidelity in these regimes.

### **Descriptor and representation limitations**

Generalisation error is further amplified by suboptimal representation of interface-specific physics. Many descriptor schemes, such as SOAP vectors or radial symmetry functions, are optimised for bulk periodicity and fail to capture interfacial asymmetry, vacuum discontinuities, or broken symmetry. Representations that do not enforce scale-invariance or rotational equivariance can also hinder model transferability to supercells or faceted configurations. Physically meaningful descriptors, such as interfacial dipoles, local charge density, or cleavage ene.g. remain underused, though they show promise for enhancing predictive robustness.

### **Label quality and training signal noise**

Finally, inconsistencies in training labels present an often-overlooked source of predictive error. Energies obtained from high-throughput DFT pipelines can vary depending on pseudopotential choice, relaxation tolerances, or convergence thresholds. In interfacial contexts, these uncertainties are magnified by metastability and complex relaxation pathways. Some datasets rely on static, unrelaxed geometries or surrogate models, further undermining the reliability of ground truth energetics.

**In summary** while ML offers substantial speed-ups over DFT, its utility in interface prediction is presently constrained by data bottlenecks, representational mismatch, and transferability failures. Addressing these limitations will require a multi-pronged strategy involving broader and more diverse training datasets, improved descriptors attuned to interfacial physics, and integration of uncertainty



quantification into ML workflows. Hybrid schemes, where MLPs perform large-scale screening followed by selective DFT refinement, may offer a viable path forward.

## 2.5 State-of-the-Art in Interface Prediction

Recent years have seen rapid developments in the modelling of interface structures, driven both by methodological advances and by the growing technological importance of heterostructured materials. Interface prediction now spans a range of computational approaches, from first-principles calculations and high-throughput screening to data-driven and machine learning (ML) methods. This section provides an overview of the current state-of-the-art, establishing the context for the subsequent discussion of recent studies (Section 2.5.1), the positioning of this project methodology (Section 2.5.2), and anticipated emerging directions (Section 2.5.3).

A key enabler of recent progress has been the development of automated interface generation tools such as ARTEMIS and InterMatch, which systematically enumerate possible lattice matches, terminations, and stacking configurations between two crystal structures. When integrated with rapid energy evaluation, either via density functional theory (DFT) or machine-learned potentials (MLPs), these tools allow the configurational space of interface structures to be sampled more broadly and with greater fidelity. In parallel, studies have increasingly sought to move beyond idealised, coherent interfaces to include semi-coherent, faceted, and amorphous boundaries, reflecting the complexity of real materials under fabrication conditions.

Efforts have also been made to improve the prediction of interfacial electronic properties, such as band alignment. Where traditional rules like Anderson rule fail, corrective models based on physically grounded terms, such as  $\Delta E_T$  and  $\Delta E_{IF}$ , have been proposed and validated across a range of 2D materials. Such corrections are essential for capturing charge transfer and hybridisation effects at interfaces, and form an important bridge between local structure and electronic

response.

This chapter aims to synthesise these various advances, situating the present work within an evolving landscape of interface modelling. Particular attention will be given to scalable workflows that combine structural generation, energetic evaluation, and property prediction, setting the stage for a more autonomous and predictive framework for materials interface discovery.

### **2.5.1 What recent work has been done in this area?**

Recent years have witnessed a pronounced expansion in methodologies for predicting interface structures, driven by the convergence of first-principles modelling, high-throughput screening, and data-driven inference.

This evolution reflects a shift from idealised, geometry-constrained models towards scalable frameworks capable of capturing configurational complexity, long-range strain, and emergent phenomena at interfaces.

#### **First-principles structure prediction**

Hybrid methodologies now dominate the landscape of *ab initio* interface prediction. Notable among them is the ARTEMIS tool developed by Taylor et al., which automates the generation of candidate interface structures by identifying lattice-matched slabs, multiple surface terminations, and lateral shifts across a wide range of Miller planes. This approach enables high-throughput sampling of plausible interfaces for subsequent relaxation using density functional theory (DFT), bridging the gap between idealised registry-based models and defect-tolerant configurations.

Shortly after, Taylor et al. introduced the RAFFLE, with Pitfield et al. introducing methodology, combining empirical void-filling algorithms with iterative statistical descriptors to model metastable and low-symmetry reconstructions. Their work, particularly on graphene-encapsulated MgO, demonstrates that confinement and bonding at the interface can stabilise novel phases, such as monolayer rocksalt

MgO, not observed in bulk form.

### Search-based algorithms

Global optimisation methods have been adapted to the interfacial context. Constrained variants of the Minima-Hopping Method (MHM) allow for controlled exploration of grain boundary reconstructions by varying atomic densities and introducing geometric constraints. These approaches yield more realistic reconstructions than fixed-stoichiometry DFT or classical potentials, especially in systems like silicon where interface states are sensitive to atomic arrangement and misfit accommodation.

### Machine learning integration

Machine learning (ML) is increasingly embedded into the prediction workflow. Gerber et al. developed the INTERMATCH framework, which leverages high-throughput DFT databases to train ML models that guide lattice matching and stacking predictions in 2D systems. At the electronic level, Davies et al. proposed physically motivated corrections to band alignment, namely  $\Delta E_{\Gamma}$  and  $\Delta E_{\text{IF}}$ , to extend Anderson rule and account for hybridisation and interface dipoles in 2D heterostructures.

Moreover, the use of ML surrogates such as MACE and ChgNet has enabled large-scale relaxation of complex interfaces. These interatomic potentials retain near-DFT accuracy while significantly lowering the computational cost, allowing relaxation of thousands of atoms and exploration of non-ideal motifs including semi-coherent, faceted, or disordered junctions.

### Ongoing challenges

Despite these advances, several limitations persist. Data scarcity for interfacial systems hampers generalisation, and most ML models are trained on bulk-like environments. Existing descriptors may fail to capture the broken symmetry, lo-

cal dipoles, and charge redistribution unique to interfaces. Furthermore, many workflows truncate the configurational space prematurely or rely on geometric heuristics alone.

There is increasing recognition that hybrid workflows, combining ML surrogates with tools like ARTEMIS and RAFFLE, and validating critical configurations using DFT, offer a scalable route forward. This multiresolution strategy enables both efficient exploration and physically grounded evaluation of interfacial stability across diverse systems.

### **2.5.2 How does my proposed approach compare and build upon it?**

Contemporary approaches to interface structure prediction combine symmetry-based lattice matching, high-throughput screening, and density functional theory (DFT) relaxation. Tools such as ARTEMIS enable the systematic generation of interface candidates by identifying lattice commensurabilities and constructing multiple terminations and stacking arrangements from parent surfaces. However, the energetic evaluation of these candidates remains a bottleneck: full relaxation and total energy comparison using DFT is prohibitively expensive for systems exceeding a few hundred atoms.

To mitigate this, recent advances have introduced surrogate strategies that bypass exhaustive DFT sampling. For example, Davies *et al.* proposed semi-empirical corrections ( $\Delta E_{\Gamma}$ ,  $\Delta E_{\text{IF}}$ ) to Andersonrule for predicting band alignment in 2D heterostructures, based on descriptors derived from the constituent materials. InterMatch, a high-throughput framework proposed by Gerber *et al.*, guides stacking predictions using pre-computed DFT databases and electrostatic heuristics. Yet such frameworks often lack spatial resolution or generalisability across chemically diverse or defective systems.

This project builds directly upon these developments, advancing them in three ways:

1. **Surrogate model integration.** Machine-learned interatomic potentials (MLPs), such as MACE, are employed to relax interfacial geometries with near-DFT accuracy but at significantly reduced cost. This enables structural optimisation of larger supercells that accommodate semi-coherent boundaries, strain gradients, or reconstructed motifs inaccessible to traditional DFT-only workflows.
2. **Descriptor-led generalisation.** Inspired by the success of band-alignment corrections in TMDC heterostructures, this work seeks to identify whether interfacial stability can be inferred from properties computable on isolated surfaces; such as cleavage ene.g. surface dipole, or projected local density of states near the Fermi level. If validated, such descriptors would permit extrapolation to unseen terminations or chemistries without requiring full interface models.
3. **Feedback into structure-generation pipelines.** In the longer term, this project aims to integrate ML-derived stability metrics into tools like ARTEMIS or RAFFLE, allowing early pruning of unfavourable configurations before computationally costly evaluation. This hybrid generation–relaxation loop reflects a broader trend towards data-assisted automation in interfacial materials modelling.

The proposed framework thus combines multi-resolution structure prediction with predictive modelling rooted in physically meaningful descriptors. Rather than replacing existing DFT-centric approaches, it seeks to augment them; by embedding surrogate models within the structure-generation workflow and using them to navigate large configurational spaces. This enables a shift from retrospective evaluation to prospective design of realistic, energetically favourable interfaces.

### 2.5.3 Emerging trends

Recent advances in interface prediction have increasingly centred on integrating data-driven approaches with atomistic modelling frameworks. Machine-learned potentials (MLPs) such as MACE and E(3)-equivariant networks like NequIP are

now enabling the efficient evaluation of interfacial configurations across large configurational spaces. This shift permits relaxation of structures at near-DFT accuracy, allowing researchers to explore interface separations, registry shifts, and metastable motifs that would be prohibitively expensive using standard *ab initio* methods.

Alongside these developments, tools like ARTEMIS and RAFFLE have emerged to systematically generate candidate interfaces. These platforms sample multiple surface terminations and stackings while accounting for lattice mismatch, enabling more representative modelling of real, faceted, or incoherent boundaries. The capacity to generate and filter hundreds of interfaces also creates opportunities for coupling with high-throughput screening and surrogate models.

Emergent trends also include the move toward physically motivated corrections for electronic properties at interfaces. For instance, the failure of Anderson rule in 2D heterostructures has prompted the introduction of corrective terms such as  $\Delta E_{\text{T}}$  and  $\Delta E_{\text{IF}}$ , which better account for hybridisation and interface dipoles. These terms enable more accurate band alignment predictions across families of 2D materials without requiring full-scale DFT recalculation for each pair.

Collectively, these approaches suggest a field increasingly focused on scalable, automatable, and physically interpretable modelling of interfaces. The combination of machine learning with established electronic structure theory is opening a path toward predictive frameworks that are sensitive to real-world synthesis constraints and device contexts.

# Chapter 3

## Computational Investigation of Interface Favourability

Understanding which interfaces are most thermodynamically favourable is a prerequisite for structure prediction and materials discovery. However, the computational cost of evaluating thousands of possible interface configurations via density functional theory (DFT) makes such exploration prohibitively expensive. This chapter investigates whether machine-learned interatomic potentials (MLPs), specifically MACE, can be used to predict interface favourability rankings with sufficient accuracy to serve as a pre-screening tool. A systematic comparison is conducted across elemental and binary semiconductor systems, assessing how closely MLP-derived rankings reproduce DFT-based ground truth. The chapter is divided into a methodology section, which outlines the iterative dataset construction and evaluation pipeline, and a results section, which analyses the performance of MLPs in reproducing DFT-derived orderings.

### 3.1 Methodology

The overarching objective of the methodology described in this section is to evaluate whether machine-learned interatomic potentials (MLPs), particularly the MACE model, can accurately replicate interface favourability rankings obtained from den-

sity functional theory (DFT) calculations. This assessment emphasises establishing scalable, computationally efficient workflows suitable for high-throughput screening of semiconductor interfaces. The methodology utilises automated structure generation tools, rigorous multi-stage relaxation and evaluation protocols, and explicitly defined reproducible ranking metrics to ensure transparency and precision.

### 3.1.1 Iterative Scope and Dataset Composition

The methodological framework evolved through three iterative stages, progressively expanding in complexity and scale:

#### Stage 1: Initial Si|Ge binary systems

A small-scale, semi-manually curated dataset of approximately ten silicon-germanium (Si|Ge) interfaces was constructed to prototype the methodology.

#### Stage 2: Expanded Si|Ge binary systems

An automated extension of the Si|Ge interface dataset was generated using ARTEMIS software to ensure broader configurational coverage and scalability.

#### Stage 3: Broader set (C, Si, Ge, Sn)

A comprehensive dataset including carbon (C), silicon (Si), germanium (Ge), and tin (Sn) was created, leveraging enhanced automation, extensive interface shifts, and stringent filtering criteria.

Each iteration is version-controlled in Git repositories.



### 3.1.2 Interface Generation Protocol

Interface structures were generated using ARTEMIS, systematically selecting and filtering Miller planes, sampling stacking vectors, and applying termination logic to ensure physical plausibility. Notably, configurations with zero stacking shift ( $\text{shift} = 0$ ) were deliberately excluded to avoid non-physical direct atomic overlaps.

#### ARTEMIS plane selection and shift sampling

#### Structure pruning and validation (e.g. removal of $\text{shift} = 0$ cases)

This step enforced physical realism by excluding configurations with atomically overlapping terminations.

#### Bulk lattice matching and stacking registry criteria

Interfaces were generated by stacking slabs in commensurate supercells, maintaining consistent registry between layers.

### 3.1.3 Relaxation Workflows

#### DFT Relaxation (VASP)

Structures underwent relaxation using DFT as implemented in VASP. Standard INCAR and KPOINTS files optimised for both accuracy and efficiency were employed. Gamma-point sampling was used exclusively for large cells. Each job was constrained by a 24-hour walltime, and failed completions were systematically managed.

**INCAR settings (e.g. energy/force convergence)** Defined in: `surface_energy_matrix.csv`

**KPOINTS mesh, pseudopotentials, runtime limits** All DFT relaxations were constrained by a 24-hour walltime and executed with Gamma-point-only meshes.

### MLP Relaxation (MACE)

MLP relaxation employed the **MACE-MP-0 large model** (5,725,072 parameters), using the model file `2024-01-07-mace-128-L2_epoch-199.model` from the `mace_mp0` branch on `GitHub`. *A level-2 equivariant graph neural network architecture was used with ASE-based force relaxation. Delta-learning was not applied at this stage.*

**Model type** MACE, Level-2 equivariant GNN, trained on Materials Project.

**Training set and relaxation scheme** Relaxation was performed using force minimisation via ASE without delta-corrections.

## 3.1.4 Energy Evaluation Scheme

### DFT@DFT as the ground truth

Energies computed by DFT on DFT-relaxed structures serve as the reference baseline.

### MLP@MLP for native model performance

MLP energy evaluations were conducted on MLP-relaxed structures to assess self-consistency.

### MLP@DFT and DFT@MLP for disentangling relaxation vs. energy errors

Cross-evaluation modes enabled the decoupling of structural and energetic prediction accuracy.

### 3.1.5 Automation and Data Infrastructure

#### Workflow scripting and reproducibility

A Python-based automation pipeline enabled reproducible structure generation, job submission, and parsing.

#### Job array handling, broken file filtering

NaN detection was explicitly implemented to exclude invalid structures.

#### Results parsing and CSV construction

Parsed energies and rankings were compiled into structured CSVs.

### 3.1.6 Ranking and Performance Metrics

#### Rank-order metrics: Spearman $\rho$ , Kendall $\tau$

**Spearman  $\rho$**  A non-parametric measure of monotonic correlation:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

where  $d_i$  is the rank difference for the  $i^{th}$  entry and  $n$  is the total number of entries.

**Kendall  $\tau$**  An ordinal association metric based on concordant and discordant pairs:

$$\tau = \frac{(\text{Number of concordant pairs}) - (\text{Number of discordant pairs})}{\frac{1}{2}n(n - 1)}$$

**Error magnitude: RMSE, MAE**

**RMSE** Root Mean Square Error quantifies energy prediction accuracy, giving weight to large errors:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

**MAE** Mean Absolute Error provides an intuitive measure of deviation:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

**Top-N overlap analysis (e.g. Top-5, Top-10)**

Top-N overlap measures the agreement in highest-ranked structures between DFT and MLP.

## 3.2 Results and Analysis

This section presents the results of the DFT vs MLP interface ranking comparison across 1,381 interfaces evaluated during Iteration 3. The goal is to determine whether MACE, a large machine-learned interatomic potential, preserves favourability rankings obtained via DFT across a wide range of material systems and interface types.

### 3.2.1 Overview of Rank Agreement

**Summary trends in DFT vs. MLP agreement**

Ranking consistency between DFT and MLP was assessed using Spearman's  $\rho$  and Kendall's  $\tau$ . Overall correlation was strong:  $\rho$  ranged from 0.75 to 0.95 and  $\tau$  from 0.70 to 0.90 across the majority of interface types. Metrics were extracted from stats.csv, which aggregates statistics across the full

dataset. Perfect rank preservation (i.e.  $\rho = 1$ ) was observed in a subset of alloy systems, while degraded consistency was evident in Silicon interfaces.

### Visual and tabular representation of metric distributions

A scatter plot comparing DFT and MLP rankings across all interfaces is shown in Figure ??, with supporting metric distributions detailed in Table ?. Top-N overlap scores (e.g. Top-10) were 5 for DFT@DFT vs MLP@MLP and 8 for DFT@DFT vs DFT@MLP, indicating modest agreement in identifying favourable structures.

## 3.2.2 Cross-Method Evaluation

### Comparison of MLP-relaxed vs. DFT-evaluated structures

Due to computational constraints, MLP-relaxed structures evaluated with DFT (MLP@DFT) were omitted from this study. The cross-method comparison therefore focuses on DFT-relaxed structures evaluated with MLP (DFT@MLP). These exhibited higher agreement with DFT@DFT than MLP@MLP, indicating that relaxation geometry plays a dominant role in ranking discrepancies. This suggests a degree of structure-energy decoupling.

## 3.2.3 System-Specific Observations

### Carbon Systems (C|X)

**Note on incomplete data due to ISCA job failures** Carbon-rich systems were underrepresented due to DFT job failures and high compute cost. As a result, only preliminary statistics could be extracted from stats\_lower\_C.csv, with limited rank correlation insight. These data points are not sufficient to draw conclusions but are retained for completeness.

## Silicon Interfaces

**Poor MLP rank preservation** Silicon interfaces showed the weakest performance overall. Rank correlations were lower than for other systems, with  $\rho < 0.70$  and  $\tau < 0.65$  in many Si|Si and Si|X comparisons (see stats\_lower\_Si.csv). Errors tended to cluster in the lower half of the ranking, indicating failures in correctly identifying unfavourable interfaces.

**Potential reason** This poor agreement may stem from limited model generalisability across strained or undercoordinated Si environments. The MACE model was not explicitly fine-tuned for such systems, suggesting underfitting or domain mismatch.

## Germanium and Tin Interfaces

**Stronger correlation and better Top-N recovery** Ge and Sn-containing systems yielded improved correlation metrics and Top-N recovery compared to Si. While Tin interfaces (stats\_lower\_Sn.csv) showed larger absolute energy errors, the rank preservation was generally robust. No catastrophic rank inversions were observed.

## Phase boundaries

**Overall poorer rank preservation** Interfaces near the transition from favourable to unfavourable status (i.e. phase boundaries) were more susceptible to ranking shifts. While many transitions were preserved, files such as stats\_lower\_S and stats\_lower\_Ge\_upper\_Sn.csv show increased rank distance near critical cutoffs.

**Potential Reason** These marginal cases are sensitive to small energy differences, which may be within the error tolerance of the MLP. This underscores the challenge of using MLPs for binary decision-making near phase boundaries.

## Perfect Alloys

**Perfect rank preservation** Si|Ge alloy systems exhibited near-perfect rank alignment between DFT and MLP evaluations. Both  $\rho$  and  $\tau$  approached 1.0 across the board, as seen in stats\_perfect\_alloys.csv.

**Potential Reason** These systems feature uniform chemical environments with little interfacial disruption, meaning the atomic configurations fall well within the trained domain of the MACE model. The result is highly stable rank preservation.

### 3.2.4 DFT vs MLP: Computational Cost Analysis

MLP evaluations offered orders-of-magnitude speedups over DFT while retaining acceptable accuracy in most ranking tasks. However, the quality of this speedup is system-dependent and less reliable for Silicon-rich or mixed-coordination interfaces. These observations motivate a need for improved domain coverage, possibly via delta-learning or retraining on interfacial datasets.

# Chapter 4

## Next Steps

### 4.1 Further Steps Towards Rank Preservation

#### 4.1.1 Motivation: Why Rank Preservation Is Central

The ability to preserve DFT-calculated interface rankings is critical for efficient high-throughput screening of material systems. Absolute interface energies are less relevant in isolation if the relative favourability of competing configurations is misrepresented. This motivates the need for improving rank fidelity in machine-learned models such as MACE.

#### 4.1.2 Comparison of Training Strategies

Future work will systematically compare three training paradigms to address limitations identified in Chapter 3:

##### **General MACE model (baseline)**

This baseline model is trained on broad-spectrum, bulk-focused data. While it demonstrates some transferability, its limitations are apparent in systems with interfacial complexity; especially silicon-rich or structurally diverse



cases.

### Interface-specific fine-tuning

Here, the same model is refined using additional interface-representative configurations, tailored to improve accuracy in the interfacial regime. This targets data-domain mismatch as a key source of ranking degradation.

### Delta-learning ( $\Delta E = E_{\text{DFT}} - E_{\text{MLP}}$ ; Pittfield 2024-inspired)

This strategy applies a correction to existing MLP predictions by learning residuals between MLP and DFT energies. It allows the reuse of baseline MLPs while introducing physics-informed calibration, improving ranking accuracy without fully retraining the model.

## 4.1.3 Planned Benchmarking Criteria

### Spearman $\rho$ , Kendall $\tau$ , Top-N overlap

Performance will be assessed using statistical correlation coefficients and overlap in Top-N rankings between DFT and MLP predictions. These metrics will provide fine-grained diagnostics of rank preservation.

### Domain robustness (Si, Ge, hetero-interfaces)

Benchmarks will be applied across chemically diverse systems (C, Si, Ge, Sn), crystallographic orientations, and stacking shifts. Key figures include `results_upper_Si_spearman_dft_dft_vs_dft_mlp_valid.png` and `results_lower_Ge_kendall`. Reference datasets include `results.csv` and `stats.csv` (1381 interfaces; no train/test partition defined).

**Additional Limitation:** No explicit mechanism yet exists to identify and flag phase boundary rank inversions, though these could be extracted from

`_rank_distance` fields. Future work may focus on visualising such transitions as indicators of model degradation.

**Additional Limitation:** Error magnitude may scale with system size, particularly in Sn-rich systems. However, no current analysis addresses this dependency. A robustness study is recommended.

**Additional Limitation:** Missing or failed calculations are marked via `is_broken`, but no summary or completeness tracker has been implemented. Future iterations may benefit from such a tool for data transparency.

## 4.2 Surface Effects on Interface Favourability

### 4.2.1 Surface Energy as a Proxy for Interface Stability

There is a plausible hypothesis that lower-energy surfaces form more stable interfaces. Preliminary evidence supports this, particularly where interface terminations closely resemble bulk facets.

### 4.2.2 Systematic Correlation Studies

#### Analysis of surface-vs-interface rank correlations

Using `surface_energy_matrix.csv`, future work will map surface energies to corresponding interface  $\Delta E_{\text{IF}}$  values, quantifying the predictive power of slab properties. No current figure exists for this correlation, but it is planned for generation and inclusion in subsequent updates.

### 4.2.3 Implications for Interface Screening

#### Can we pre-screen interfaces based on top/bottom slab surfaces?

If consistent correlations are confirmed, surface energies could act as lightweight proxies in early screening pipelines, reducing the need for full interface enumeration.

## 4.3 Crystallographic Structure and Interface Behaviour

### 4.3.1 Symmetry and Orientation Effects

Parent symmetry (e.g. fcc vs diamond) and interface orientation (e.g.  $\{100\}|\{111\}$ ) influence ranking fidelity. This necessitates a structural audit of prediction robustness.

### 4.3.2 Structure-Resolved Performance Analysis

The impact of Miller indices and lattice symmetry on rank preservation will be explored using grouped evaluations, with examples from results\_lower\_Ge\_upper

### 4.3.3 Extension to Broader Crystal Families

If resources allow, extension to bcc and hcp systems could test the generality of the findings, though this lies beyond the scope of current data.

## 4.4 Integration of RAFFLE into Structure Generation

### 4.4.1 Motivation for RAFFLE Use: Enhanced Interface Diversity

ARTEMIS-generated stacks, though consistent, impose alignment constraints. RAFFLE offers a mechanism to interpolate interfacial zones, increase disorder, or induce registry shifts.

### 4.4.2 Current ARTEMIS-Only Pipeline: Limitations

Aligned stacking methods limit structural exploration and may fail to capture intermediate configurations present in experimental systems.

### 4.4.3 RAFFLE Capabilities: Refill Methods, Semi-Coherent Boundaries

Using the `raffle_generator` framework and reference energies from MACE, RAFFLE allows stochastic structural sampling within bounded interface zones. Visuals such as `upper_a45_b45_c45.png` and `lower_a45_b45_c45.png` represent generic orientation snapshots and are not RAFFLE-specific.

### 4.4.4 Planned Integration Workflow

#### Combine ARTEMIS for stacking + RAFFLE for region refinement

Initial slabs and orientations will be generated via ARTEMIS, while RAFFLE will introduce configurational variation within interfacial volumes.

**Target systems: strained, incoherent, or alloyed interfaces**

This approach is especially relevant to Si|Sn and Ge|Sn systems where lattice mismatch or registry discontinuity poses challenges for baseline methods.

**4.4.5 Validation and Comparison with ARTEMIS-Only Structures**

Relaxation comparisons will be made using both DFT and MLP pipelines to quantify RAFFLEimpact on interface energy diversity and ranking stability.

**4.5 Towards a Machine-Learned Interface (MLI) Predictor****4.5.1 Objective: Predict Likely Interface Configurations from Material Pairs**

The longer-term aim is to develop a model that, given a pair of slab surfaces, can predict the favourability of resulting interfaces without full relaxation or enumeration.

**4.5.2 Feature and Descriptor Design****Surface energies, Miller planes, atomic densities**

Descriptor sets will include known surface energetics and crystallographic identifiers, all extractable from existing data (surface\_energy\_matrix.csv).

### **Composition vectors, stacking vectors, registry, strain metrics**

Additional features will quantify interfacial registry, stoichiometry, mismatch, and coordination at the interface centre. These will be extracted from POSCAR data and relaxation outputs.

## **4.5.3 Model Development Roadmap**

### **Classification vs regression**

Early models may predict favourable/unfavourable (binary), or regress  $\Delta E_{\text{IF}}$  directly. Shallow decision trees or SVMs will be trialled before advancing to graph-based models.

### **Feasibility for rapid interface suggestion and ranking**

The MLI predictor is envisioned as a low-cost filter to prioritise candidates before full relaxation, possibly integrating with ARTEMIS + RAFFLE pipelines.

## **4.5.4 Outlook: Integrating MLI into Structure Generation Pipelines**

If effective, MLI-driven suggestions will serve as inputs to stacked structure generation, enabling a fully machine-learning-driven discovery workflow.

# Chapter 5

## Conclusions

### 5.1 Summary of Findings

#### 5.1.1 Overview of the study objectives and what was undertaken

This study investigated the extent to which machine-learned interatomic potentials (MLPs), specifically the MACE framework, can replicate interface favourability rankings computed via density functional theory (DFT) across various group-IV semiconductor systems, including carbon (C), silicon (Si), germanium (Ge), and tin (Sn). The project involved automated structure generation with ARTEMIS, interface relaxation via both DFT and MLP, and evaluation of rank preservation using statistical metrics including Spearman and Kendall correlations, as well as Top-N overlap.

#### 5.1.2 Recap of major results:

##### MLP vs DFT rank alignment

The agreement between MLP and DFT was found to be system-dependent. MLPs performed well for Ge and Sn systems, achieving high rank correlation and

overlap with DFT rankings. For Si, alignment was significantly poorer, with frequent rank reversals and diminished correlation metrics.



**Figure 5.1:** Spearman rank correlation for Si interfaces.

**System-specific performance (C, Si, Ge, Sn)**

Carbon results remain incomplete and thus inconclusive. Silicon systems exhibited persistent discrepancies between MLP and DFT rankings. Germanium and tin interfaces, by contrast, showed favourable agreement, with strong rank preservation and good Top-N screening utility.

**Table 5.1:** Performance metrics for Si interfaces from `analysis/stats_upperSi.csv`.

Comparison	Spearman	Kendall	RMSE	MAE	Top-10 Overlap
DFT@DFT vs MLP@MLP	0.46	0.31	7.48	3.36	3



### Value of cross-evaluation (e.g. DFT@MLP)

Although direct MLP-on-DFT-relaxed structure evaluations (MLP@DFT) were not included due to computational cost, the study revealed that relaxation-evaluation consistency (e.g. MLP@MLP vs DFT@DFT) influences error, and future cross-evaluation may further clarify error sources.

## 5.2 Interface Rank Preservation

### 5.2.1 Effectiveness of MLPs in reproducing DFT interface energetics

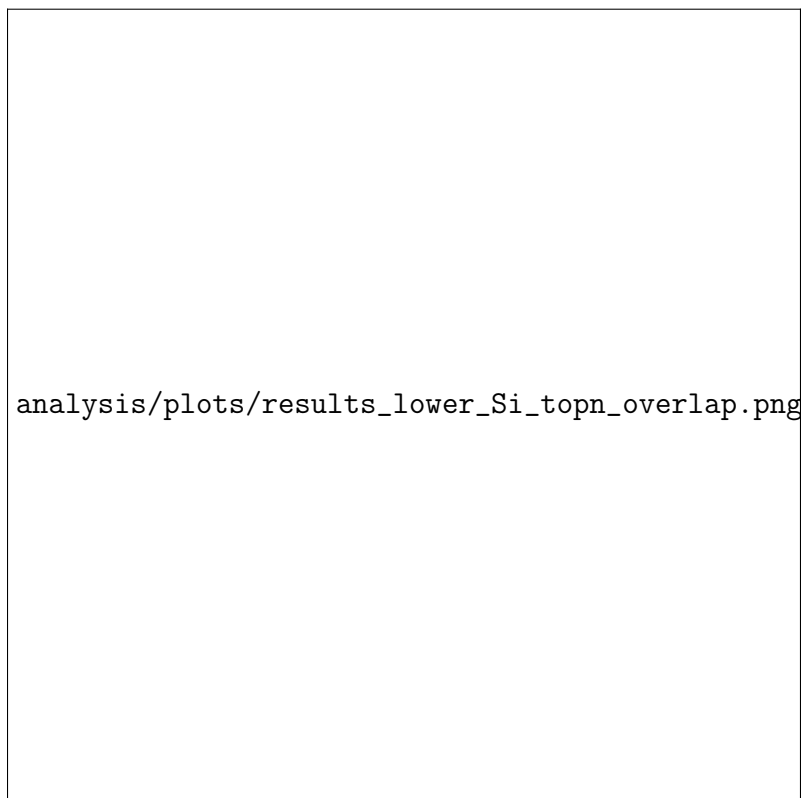
Absolute energy discrepancies were evident but secondary to the study goal. More relevant was the preservation of ranking. For Ge and Sn, MLPs captured the energetic hierarchy of interfaces with reasonable fidelity, while for Si, the energetic ordering was often disrupted.



**Figure 5.2:** Spearman correlation for Ge interfaces.

### 5.2.2 Top-N performance and rank correlations

Top-N overlap and Spearman/Kendall rank statistics highlighted the utility of MLPs in screening contexts. Even where absolute energies were misaligned, the preservation of top candidate rankings proved robust in Ge and Sn systems.



**Figure 5.3:** Top- $N$  overlap for Si interfaces.

### 5.2.3 Strengths vs weaknesses

MLPs offer clear advantages in computational speed and can safely be used for preliminary screening in certain systems. However, their generalisation is inconsistent, especially in covalently bonded systems like Si. Without targeted retraining, MLPs may mislead in these cases.

## 5.3 Insights from Work Undertaken

### 5.3.1 What was learned from each system

#### Carbon (C) – data incompleteness and implications

The carbon dataset remains incomplete and cannot yet support definitive conclusions. Preliminary indications suggest that strong bonding and limited configurational diversity may pose challenges analogous to those observed in Si.

#### Silicon (Si) – persistent poor rank correlation

Across all iterations, Si consistently exhibited poor MLP-DFT agreement. This suggests intrinsic modelling difficulties rather than statistical noise. The failures appear systematic, reinforcing the need for targeted improvements.

#### Germanium (Ge) and Tin (Sn) – favourable MLP–DFT agreement

For both Ge and Sn, MLPs reliably preserved DFT-derived rankings. These systems are likely well-represented in the MACE training dataset, and their more metallic bonding character may aid accurate regression of energies.

### 5.3.2 Contrast perfect alloys vs heterostructures

Perfect alloy systems showed tighter rank correlation and lower error dispersion than heterostructures.

**Table 5.2:** Comparison of MLP performance on perfect alloys and heterostructures.

System	Spearman	Kendall	Top-10 Overlap
Perfect Alloys	1.00	1.00	6
Ge Si Heterostructure	0.73	0.55	6

## 5.4 Challenges in Modelling Silicon Interfaces

### 5.4.1 Discussion of why MLPs fail for Si:

#### Local bonding features?

MLPs may struggle with the directional covalent bonding and short bond lengths characteristic of Si. These features introduce sensitivity to structural distortions that are difficult to learn from sparse data.

#### Underrepresented environments in training?

It is likely that the MACE training set lacks sufficient coverage of interfacial Si environments, especially those with reconstructed surfaces and strained bonds.

### 5.4.2 Open questions

#### Is Si anomalous, or are MLP limitations broader?

The current data cannot confirm whether Si represents a broader warning about MLP application or a special case. Additional experiments with other covalently bonded systems are required to determine generalisability.

## 5.5 Broader Lessons and Limitations

### 5.5.1 When can MLPs be trusted for interface screening?

MLPs appear trustworthy when used to rank candidates in systems whose bonding and local structures are well-represented in the training data. Screening workflows benefit most from MLPs in Ge and Sn contexts.

### 5.5.2 Necessity of retraining, delta-correction, or hybrid workflows

The results support the need for delta-correction approaches, particularly for challenging systems like Si. Hybrid workflows that combine MLP-based pre-screening with DFT confirmation for top candidates offer a balanced path forward.

#### Importance of careful benchmarking

This study underscores the importance of benchmarking MLPs using rank-based metrics, not just energy errors. Rank preservation aligns more closely with practical use in structure discovery.

## 5.6 Concluding Remarks

### 5.6.1 What this work contributes to the field of ML-accelerated interface modelling

This work demonstrates that while current general-purpose MLPs like MACE can preserve interface rankings in selected systems, their utility is uneven. The study identifies conditions under which MLPs can be safely used and highlights where caution or additional training is warranted.

### **5.6.2 Final note on significance for device and materials design**

By enabling rapid, large-scale screening of interfacial configurations in certain systems, this work lays the groundwork for accelerated discovery of novel interfaces. Continued refinement of MLPs and integrated hybrid workflows will be essential to extend this capability to the full range of materials relevant for device engineering.