MASTER THESIS

Computational Study of Grain Boundary Segregation in Iron Using Universal Machine Learning Interatomic Potentials

Zhaolun Li (5740029)

Under the supervision of: Maria Santofimia Navarro, TU Delft; Poulumi Dey, TU Delft; Fei Shuang, TU Delft; Sebastián Echeverri Restrepo, SKF; Hanzheng Huang, SKF

Abstract

The detrimental effects of Cu contamination during steel recycling and production are primarily due to the segregation of Cu at Fe grain boundaries (GBs). A promising approach to mitigate these effects is the introduction of alloying elements that inhibit Cu segregation at Fe GBs, which can be investigated through atomistic simulations. Currently, popular simulation methods include density functional theory (DFT)-based simulations and empirical interatomic potentials (EIPs)-based simulations. DFT calculations can provide simulations with high accuracy, while the high computational cost limits their application to simpler GB structures. Moreover, simulations utilizing EIPs may lack reliable potentials, especially when predicting GB segregation tendencies across a broader range of alloy systems. To address these challenges, universal machine learning interatomic potentials (uMLIPs), which are trained on DFT data and applicable for most of elements, have emerged as a promising alternative. Although uMLIPs have shown potential in various materials simulation tasks, their reliability for out-of-distribution tasks, such as simulating GB segregation behavior, remains unproven.

This thesis project evaluates the performance of the best available uMLIPs, specifically MACE-MP-0, CHGNet, M3GNet, and SevenNet-0, in predicting single-solute GB segregation energies, GB energies for both body-centered cubic (BCC) Fe and face-centered cubic (FCC) Fe systems, and solution enthalpies for BCC Fe and cementite. The results were compared against existing studies conducted via DFT calculations to assess the accuracy and applicability of each uMLIP. Additionally, some EIPs were tested for comparison, serving as an extra reference. The findings reveal that MACE-MP-0 generally outperforms the other uMLIPs in both accuracy and stability of convergence. While all tested uMLIPs perform well in BCC Fe systems, CHGNet(v0.2.0) and SevenNet-0 show reduced accuracy in FCC Fe simulations. Although most of the simulations using uMLIPs converged well in BCC Fe GBs, many unconverged cases were reported in FCC Fe systems, particularly for uMLIPs other than MACE-MP-0 and CHGNet(v0.3.0). Furthermore, a consistent underprediction of segregation tendencies for highly segregating solute elements, such as Cu, is observed in the results of MACE-MP-0 and CHGNet. This suggests that while uMLIPs hold significant potential for atomistic simulations, fine-tuning pre-trained uMLIP models for out-of-distribution tasks, such as calculating GB segregation energy, is recommended to improve accuracy and convergence behavior. This work offers a valuable benchmark for using uMLIPs in future GB segregation studies.

List of Frequently Used Abbreviations

GB: Grain Boundary

DFT: Density Functional Theory

PES: Potential Energy Surface

EAM: Embedded Atom Method

MEAM: Modified Embedded Atom Method

EIP: Empirical Interatomic Potential

MLIP: Machine Learning Interatomic Potential

uMLIP: Universal Machine Learning Interatomic Potential

MAE: Mean Absolute Error

STGB: Symmetric Tilt Grain Boundary

Contents

1 Introduction

1.1 Cu Contamination during Steel Recycling

Steel plays an essential role in shaping our built environment and sustaining economic development, and at the same time, steel is the most recycled material in the world. According to Pauliuk et al. [1], by 2050 the global supply of end-of-life scrap will triple from current amounts, which indicates that there is an abundant supply for steel recycling. To meet the increasing need in steel recycling, a portion of steelmaking in Europe has shifted into the electric arc furnace (EAF) process, which relies on steel scrap as the main feedstock. Despite the evident benefits, the full potential of steel recycling remains underutilized. Currently, EAF is only responsible for no more than 50% of the EU steelmaking, with the rest of steelmaking still dominated by the conventional blast furnace route [2]. An important factor that limits the secondary steelmaking process like EAF is the contamination of tramp elements during recycling.

Tramp elements are undesirable impurities found within materials, particularly metals such as steel. These elements are often unintentionally introduced during the production or recycling process and can have negative effects on the properties and performance of the material. According to Nakajima et al. [3], the key tramp elements in steel recycling are identified as Cu and Ni based on thermodynamic analysis, with Cu contamination posing a significant challenge in the steel recycling industry. The presence of Cu in steel scrap introduces several detrimental effects during the manufacturing process, including surface defects, hot shortness, intergranular cracks during continuous casting, cracking during long time reheating in the furnace, surface cracking during hot forming, etc [4]. Daehn et al. indicated in their review [5] that the amount of Cu in steel scrap will surpass tolerable levels across all steel products by 2050 as shown in Fig. 1 based on Pauliuk's analysis, making it an urgent need to find an efficient solution to avoid Cu contamination during steel recycling.

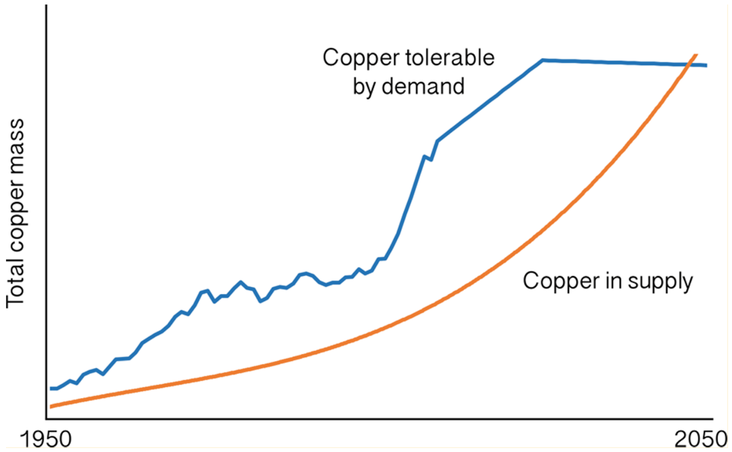


Figure 1: A predicted trend of mass of Cu in the end-of-life scrap supply (orange) and Cu tolerable by demanded products (blue) from 1950 to 2050 [5].

To mitigate Cu contamination effectively, the steel industry must adopt innovative methodologies. In recent years, methods for removing tramp element from steel scraps have been widely investigated by many researchers. However, so far, most of the methods more or less have some disadvantages. For example, Vacuum Arc Remelting is capable of effectively removing Cu and is able for commercial use. However, its application is restricted to high-quality, low-volume steels due to the necessity of re-melting in a vacuum at very slow rates, resulting in high energy expenses that are only justified for certain purposes [6].

1.2 Introduction of Grain Boundary Segregation

In the previous subsection, some detrimental effects due to high Cu content in steel were discussed. It is indicated in the literature that the detrimental effects are related to the segregation of Cu at Fe grain boundaries (GBs) [7–9]. In the past few decades, although research efforts have been dedicated to exploring methods for removing Cu during the recycling of steel, there are fewer studies focusing on understanding the segregation behavior of Cu at Fe GBs.

GBs are one of the most important defects in crystalline materials system. GBs are the regions where the crystal lattice abruptly changes orientation, marking the interface between adjacent grains in polycrystalline materials. The formation of GBs are due to variations in crystallographic orientation between neighboring grains, leading to higher energy levels compared to the crystalline interior. This increased energy prompts the system to minimize it, often through interactions with other defects within the crystal lattice [10]. One significant phenomenon observed at GBs is GB segregation, where foreign atoms, also known as solute atoms, preferentially accumulate at these interfaces. This segregation occurs due to the interaction between GBs and point defects, such as foreign atoms, leading to a reduction in the overall energy of the system. The distortion in atomic arrangement and the high free energy of the GB region facilitate the diffusion of solute atoms from the matrix to the GBs, effectively concentrating them there as thermodynamically preferred [10].

GB segregation has profound implications for the properties of materials. It affects various critical material properties, including cohesive energy, mobility, transport coefficients, and the nucleation of dislocations [11–14]. GB segregation can lead to the loss of cohesion, resulting in brittle intergranular fractures under loading, thereby limiting the practical use of the materials [15]. It can also adversely affect corrosive properties and significantly influence grain growth and recrystallization processes [15]. On the other hand, GB segregation also has advantageous effects. For example, it can stabilize grain size in nanocrystalline materials [16] and serve as a design tool for interfaces in metallic alloys [17]. For Fe, one of the most researched topic related to GB segregation is hydrogen embrittlement in steel. The hydrogen embrittlement in steel can be attributed to the cohesion loss due to hydrogen segregation at Fe GBs. In addition to that, the GB segregation of alloying elements at the Fe GBs also makes steel more prone to hydrogen embrittlement [18].

1.3 Methods for the Study of Grain Boundary Segregation

1.3.1 Experimental Methods

Experimental techniques are essential in offering insights into the GB segregation behavior. Traditionally, experimental investigations of GB segregation mainly rely on surface analysis techniques, such as Auger electron spectroscopy (AES), which have been widely utilized to study GB chemistry. However, these techniques often necessitate intergranular brittle fracture to expose the GBs, limiting the applicability to brittle materials. In recent years, with the rapid development of materials characterization techniques, the advent of techniques like

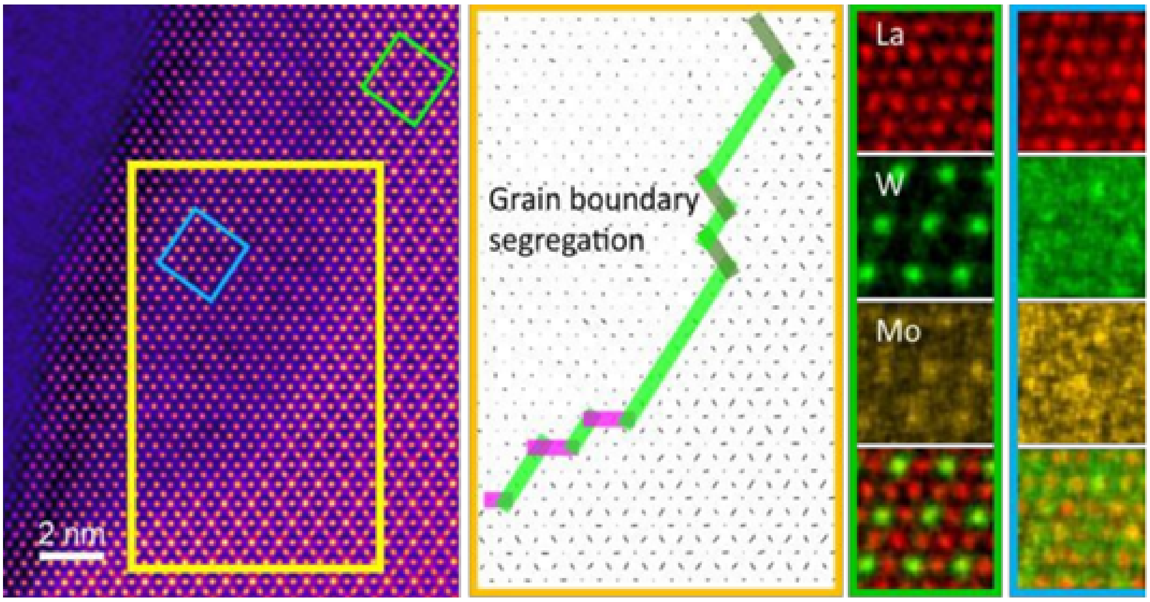


Figure 2: A high resolution high-angle annular dark-field (HAADF) image of membrane processing induced GB segregation in atomic-scale, showing how GB segregation is observed experimentally. EDX chemical mapping with atomic resolution was conducted by a ChemiSTEM microscope [19].

3-dimensional atom probe tomography (3D APT), high-resolution transmission electron microscopy (HRTEM) equipped with chemical analyzers, such as electron energy loss spectroscope (EELS), and aberration-corrected scanning transmission electron microscopy (AC-STEM) coupled with Energy-Dispersive X-ray Spectroscopy (EDX) offering the capability to investigate segregation in non-brittle systems, expanding the scope of GB segregation studies (Fig. 2 as an example). Meanwhile, these advanced techniques enable a two-dimensional or three-dimensional mapping across GBs, providing atomic-scale details of solute atoms' spatial distribution, which offer more insights into the study of GB segregation behavior [10, 15].

While experimental methods provide valuable insights into GB segregation by correlating it with the atomic structure of GBs, the comprehension of the mechanisms of various types of GB segregation and their effects on structure-property relationships remains restricted. For instance, due to the complex microstructures of materials, it is still difficult to isolate the effects caused by GB segregation experimentally even though using the most advancing techniques to date [10]. Hence, investigating GB segregation behavior using complementary approaches like computational modeling has become increasingly popular.

1.3.2 Computational Methods

The computational methods employed by researchers to study GB segregation can be classified into two categories. The first is first-principle (ab initio) quantum mechanical calculation methods, which allow for a quantum-level understanding of segregation phenomena by solving the Schrödinger equation to describe the electronic structure of materials accurately. The general formulation of the quantum mechanical equations for ab initio electronic structure calculations includes all known interactions among electrons and atomic nuclei in solids. Although the principle is relatively simple, solving these equations completely remains impossible. Thus, approximations are necessary for yielding understandable solutions. One significant simplification is density functional theory (DFT), which simplifies the complex many-body motion of all electrons into an equivalent problem of a single electron navigating within an effective potential [10].

A major challenge of DFT is that the exact energy functional, particularly the exchange-correlation component, remains unknown [15]. Therefore, DFT calculations rely on approximations, and the accuracy of these calculations depends on the quality of these approximations. Various approximations have been proposed for realistic calculations, with the local-density approximation (LDA), first proposed by Kohn and Sham [20], being the simplest among many other approximations. LDA has shown remarkable success in describing the ground-state properties of many physical systems. Nevertheless, LDA is still limited in systems with significant density gradients, Since then, there have been several attempts to improve upon the LDA. To date, a more accurate approximation that includes gradients of the electron density called generalized gradient approximation (GGA) is applied in many first-principle analyses [21, 22]. Traditional approximations like LDA might be limited in some cases (for example, as indicated by Lejcek et al. [15], when applying LDA for the exchange-correlation energy in Fe), but the improved approximation GGA has made DFT calculations able to provide a highly accurate segregation energy of certain solute element at GB.

The other approach for simulating GB segregation is through empirical interatomic potential (EIP)-based methods. EIPs are essential for atomistic simulations, which allow for modeling the interactions between atoms. These EIPs can be as simple as the pairwise potentials, such as the famous Lennard-Jones potential which can date back to a century ago, or the most widely used Embedded Atom Method (EAM) technique [23] and Modified Embedded Atom Method (MEAM) [24] developed decades ago. EIPs can be fitted through different methods, such as DFT calculations and experiments. EIP-based simulations have grown in popularity in recent years due to their computational efficiency. However, the accuracy of simulation results relies on the quality of how the EIPs are fitted. Although some EIPs with good performance can lead to simulation results closely matching experimental data, developing such potentials is a time-consuming and effort-intensive process. Additionally, the number of available EIPs is limited and a single potential model might be only available in certain conditions [25].

In recent years, rapid progress in machine learning techniques has enabled the development of machine learning interatomic potentials (MLIPs), which have transformed the way of modeling potential energy surfaces (PES). While the concept of MLIPs was introduced as early as the 1990s, it remained unrealized due to limitations in computational power and the availability of sufficient data [26]. Fig. 3 shows how MLIPs are constructed. MLIPs are trained on large reference datasets derived from quantum-mechanical calculations, making them fundamentally different from EIPs. Unlike traditional EIPs, where the PES is described as a function of atomic positions and is grounded in the physical understanding of chemical bonding, MLIPs rely on high-dimensional regression techniques to interpolate between reference energies, with the PES described as a function of local atomic environment descriptors. The major differences between EIPs and MLIPs are summarized in Table 1. While EIPs have a strong physical foundation and require only a few fitting parameters (around 10), MLIPs use thousands of fitting parameters, enabling greater accuracy, typically achieving an error of around 1 meV/atom. Although this increased accuracy comes at the cost of reduced computational speed, MLIPs are still significantly faster than DFT calculations. Despite these advantages, MLIPs exhibit poorer transferability in extrapolation compared to traditional EIPs, which makes them less reliable in scenarios outside of their training data (out-of-distribution scenarios). Nevertheless, the flexibility of MLIPs allows them to be systematically improved, which is impossible for traditional EIPs. A common limitation of both MLIPs and EIPs is that they are typically fitted to specific elements or particular combinations. Therefore, applying MLIPs to elements outside the training dataset remains infeasible.

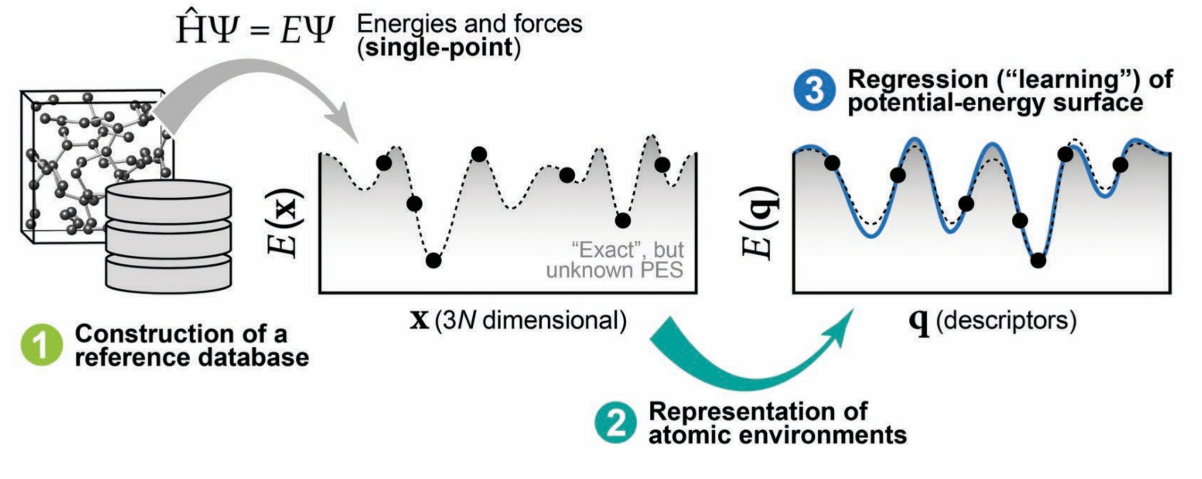


Figure 3: Schematic representation of how the MLIPs are constructured, adapted from Deringer et al [27]. The atomic structures, energies and forces computed via quantum-mechanical methods from the reference database should be interpreted as "machine-readable" descriptors, so that the PES can be constructed through mathematical regressions (learning). The descriptors might differ from different MLIPs.

Table 1: Comparison of two classes of potentials adapted from Mishin's paper [26].

Potential type	**EIPs**	**MLIPs**
Physical foundation	Strong	None
Number of fitting parameters	~ 10	$\gtrsim 10^3$
Computational speed	Very high	Slower[a]
Reference database	Small	Large
Accuracy	Limited	~ 1 meV/atom
Transferability	Reasonable	Poor
Specific to class of materials?	Yes	Depends on the training dataset
Systematically improvable?	No	Yes

[a] Slower than EIPs but orders of magnitude faster than straight DFT calculations.

In the past few years, several universal machine learning interatomic potentials (uMLIPs) have been developed. These MLIPs are termed "universal" because they are trained on data covering a broad range of elements from the periodic table, allowing them to be applied across various elements and combinations, rather than being restricted to specific ones. This technique offers a potentially powerful toolbox for materials engineers. However, their performance in relevant simulations still needs to be evaluated.

2 Literature Review

2.1 Universal Machine Learning Interatomic Potentials

2.1.1 Available Potentials

The first uMLIP is Three-Body Interaction Neural Network (M3GNet) proposed by Chen and Ong [28] in 2022. While MLIPs based on traditional material graph neural networks (GNN) lack physical constraints, the M3GNet GNN architecture takes many-body interactions into account. M3GNet utilizes a large dataset of intermediate structures generated during structural relaxation, along with their corresponding energies and forces, rather than focusing solely on the final equilibrium structures. The training data of the M3GNet (named "MPF.2021.2.8") includes energies a wide range of energies, forces and stresses of crystal structures obtained from the Materials Project database, covering 89 elements on the periodic table. The model includes 227.5k model parameters, and is trained by dataset contains 63k structures in total.

Following M3GNet, other uMLIPs trained on larger datasets, have been subsequently released. Another GNN-based uMLIP, Crystal Hamiltonian Graph Neural Network (CHGNet) [29], effectively accounts for the interaction between electrons in its simulations. Modeling electron interactions and their subtle effects in atomistic simulations has long been a challenge, specifically for classical interatomic potentials. Although some MLIPs are able to provide predictions at nearly quantum-mechanical accuracy efficiently, these MLIPs treat ions with different valence as they are different elements [29]. Such treatment fails to capture the chemical interactions such as site-preference of ions because the valence of the ions are not considered as a degree-of-freedom. However, in the model CHGNet, the atomic charges are defined in relation to the site-specific magnetic moments (magmoms), so that the effects due to the change in atomic charges can be incorporated. For example, Fig. 4 shows the results of a structural relaxation test performed by CHGNet, demonstrating the model's capability to differentiate V-ions without prior knowledge of their charge distribution. The training dataset (named "MPtrj_2022.9") of CHGNet also includes structures obtained from Materials Project dataset, which is much larger than the dataset used to train M3GNet, containing 146k structures. The dataset used to train M3GNet is a proper subset of the dataset "MPtrj_2022.9", with no materials deprecated in the later one [30].

Two additional GNN-based uMLIPs have been developed using the same training dataset as CHGNet. The first is MACE-MP-0, a uMLIP proposed by Batatia et al. based on their machine learning architecture MACE [31, 32]. MACE is a higher-order equivariant message-passing neural network that allows the efficient computation of high-body-order features. A more recent uMLIP is SevenNet-0 (Scalable EquiVariance Enabled Neural NETwork) developed based on the machine learning architecture NequIP [33], which enables parallel molecular dynamic simulations. The SevenNet-0 model replaces a large number of element-specific parameters to a linear message-passing layer. In this way, the model parameters are reduced to 842.4k without affecting the performance, thereby lowering the computational cost.

Beyond the above-mentioned uMLIPs, several new models with larger number of model parameters have also been developed, with MatterSim [34] and GNoME [35] represent notable advancements in the field. And at the time of writing, another state-of-the-art uMLIP, ORB MPtrj, trained on the "MPtrj_2022.9" dataset, has also been released. These models utilize a

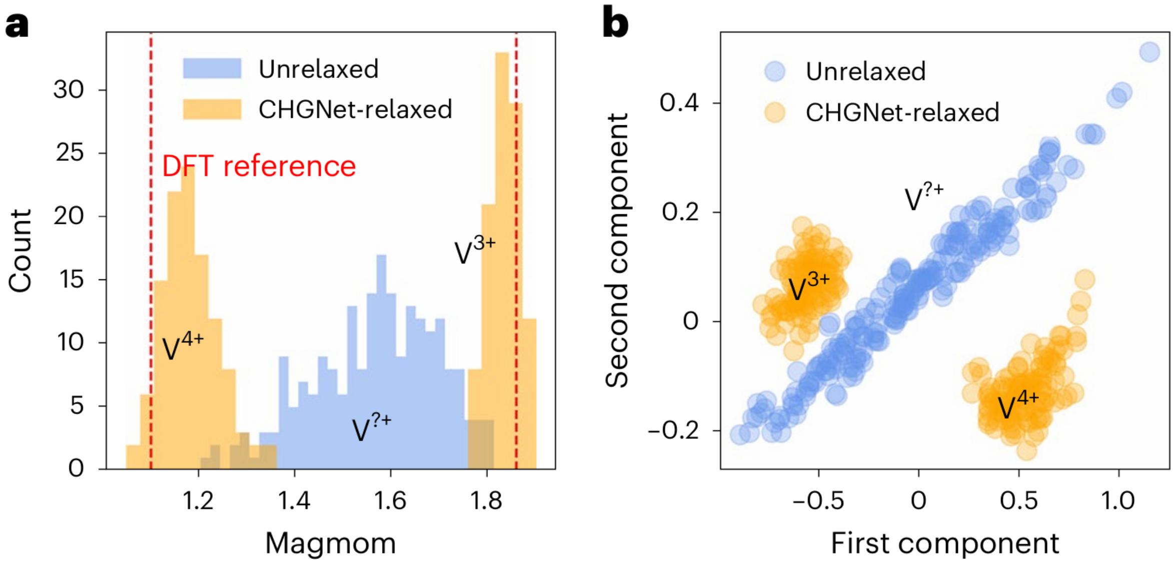


Figure 4: **a)** A graph of magmom distribution and **b)** a two dimensional visualization of the principal component analysis on V ions of unrelaxed structures and relaxed by CHGNet [29]. The creators of CHGNet generated a supercell structures to illustrate that CHGNet is able to differentiate V-ions without any prior knowledge about the charge distribution of V-ions.

greater number of model parameters compared to other uMLIPs (for instance, MatterSim has 182M model parameters), allowing them to achieve cutting-edge performance. The rapid development of uMLIPs highlights their vast potential and significant advancements in atomistic simulations with increasing accuracy and efficiency.

2.1.2 Performance Evaluations

Although benchmarks reported by the developers of each uMLIP showed that most uMLIPs exhibit small training and test errors compared to DFT calculations, these models lack evaluations for out-of-distribution simulations. Consequently, comprehensive evaluation of these uMLIPs remains necessary.

Matbench Discovery [30] provides a standardized benchmark specifically designed to evaluate the performance of uMLIP models. It provides comprehensive testing based on diverse datasets, including training dataset (dataset from Materials Project) and unseen (using WBM test dataset [36]) scenarios, which assess models' ability to generalize to unseen data. Table 2 highlights the best available uMLIPs (models) based on Matbench Discovery evaluation results. A clear trend from the table is that models trained on larger datasets generally exhibit lower mean absolute error (MAE) and root mean square error (RMSE). For instance, MatterSim, trained on 17M structures, has the lowest MAE (0.026 eV/atom) and RMSE (0.080 eV/atom) values, reflecting the advantage of extensive training data. Conversely, models trained on smaller datasets, such as M3GNet (trained on 63k structures), show higher errors (MAE of 0.075 eV/atom and RMSE of 0.118 eV/atom), indicating the limitations in generalization from limited data. Models with a higher number of parameters, such as MatterSim (182M) and ORB (25.2M), tend to perform better, likely due to their enhanced capacity to capture complex patterns in the data. This size-performance relationship of uMLIPs has also been reported in the literature [37]. However, increased model size does not always guarantee the better performance, as seen with MACE, which, despite having more model parameters, does not outperform SevenNet according to this evaluation. This suggests that while larger models typically perform better, smaller, well-optimized models or machine learning architectures can still achieve reasonable accuracy.

Table 2: uMLIPs performance comparison adapted from Matbench Discovery [30]. MAE: mean absolute error (eV/atom); RMSE: root mean square error (eV/atom); R^2: coefficient of determination. Training set: Number of materials (number of structures). The training datasets of ORB, MatterSim and GNoME are different from the dataset mentioned in the main text.

Model	**MAE**	**RMSE**	$\mathbf{R^2}$	**Training Set**	**Model Params**	**Date Added**
ORB	0.031	0.079	0.816	3M (32.1M)	25.2M	2024-09-02
MatterSim	0.026	0.080	0.812	17M	182.0M	2024-06-16
GNoME	0.035	0.085	0.785	6M (89M)	16.2M	2024-02-03
ORB MPtrj	0.046	0.094	0.740	146k (1.6M)	25.2M	2024-09-02
SevenNet	0.048	0.092	0.750	146k (1.6M)	842.4k	2024-07-13
MACE	0.057	0.101	0.697	146k (1.6M)	4.7M	2023-07-14
CHGNet	0.063	0.103	0.689	146k (1.6M)	412.5k	2023-03-03
M3GNet	0.075	0.118	0.585	63k (188.3k)	227.5k	2022-09-20

By providing a more robust assessment environment, Matbench Discovery facilitates a more reliable comparison of uMLIPs, highlighting their strengths and weaknesses in predicting material properties. However, the evaluation conducted by Matbench Discovery put more emphasis on assessing the performance of uMLIPs in the context of materials discovery. Currently, except for the evaluation work done by Matbench Discovery, only few other studies have systematically evaluated the performance of the mentioned uMLIPs. Focassio et al. [38] assessed the performance of three uMLIPs (MACE, CHGNet, and M3GNet) in predicting surface energies and Yu et al. [39] evaluated the performance of four uMLIPs (M3GNet, CHGNet, MACE and ALIGNN) in calculating formation energy for geometry optimization. Both assessments were carried out by comparing calculation results with DFT results obtained from Materials Project database, which included both data used for model training and unseen data for validation.

The evaluation done by Focassio et al. focused on surface energy predictions, an out-of-distribution task that presents challenges since most uMLIP training datasets include only bulk materials. According to their assessment, CHGNet exhibits the highest precision in calculating surface energy. The performance of MACE is also good while M3GNet exhibits the poorest performance among the three tested uMLIPs. However, discrepancies are observed for some different surface orientation and different chemical system. MACE, for example, shows excellent accuracy in predicting surface energies for the Cu system, where its error is the smallest, even outperforming several traditional MLIPs such as Moment Tensor Potential (MTP) and Neural Equivariant Interatomic Potential (NequiIP) as shown in Fig. 5. While MACE produced large errors for the W system (not shown in the figure), with a RMSE reaching nearly 3 eV/atom. This inconsistency could be attributed to the model's difficulty in generalizing to systems dissimilar to its training data, particularly surfaces, which are structurally different from bulk materials. Although in this assessment the calculated feature is mainly surface energy, the performance of uMLIPs following the sequence MACE > (better than) CHGNet > M3GNet is in good agreement with the assessment given by Matbench Discovery.

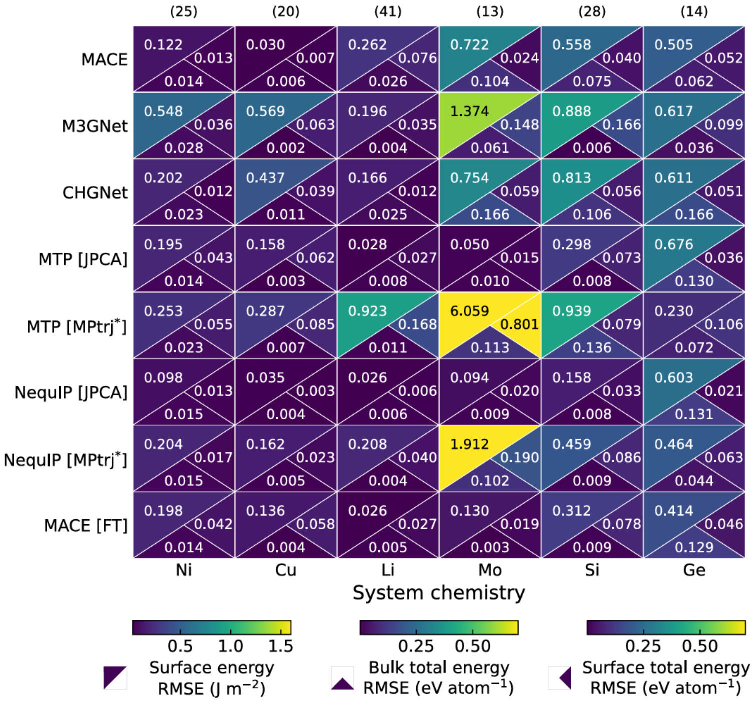


Figure 5: **(a)** Performance comparison between the three uMLIPs, other traditional MLIPs and fine-tuned MACE model. The RMSE of the universal interatomic potentials vs relation to the surface chemistry, for selected elements [38].

Similar conclusions were also made by Yu et al. [39] in their assessments. Generally, MACE and CHGNet outperform the other uMLIPs in calculating formation energies for structural relaxation, while M3GNet and ALIGNN exhibit significantly larger errors. This can be clearly seen from the heat maps in Fig. 6, the colors of elements of MACE and CHGNet are mostly purple (0-100 meV/atom), with only a few cases in blue region (100-200 meV/atom). In contrast, in

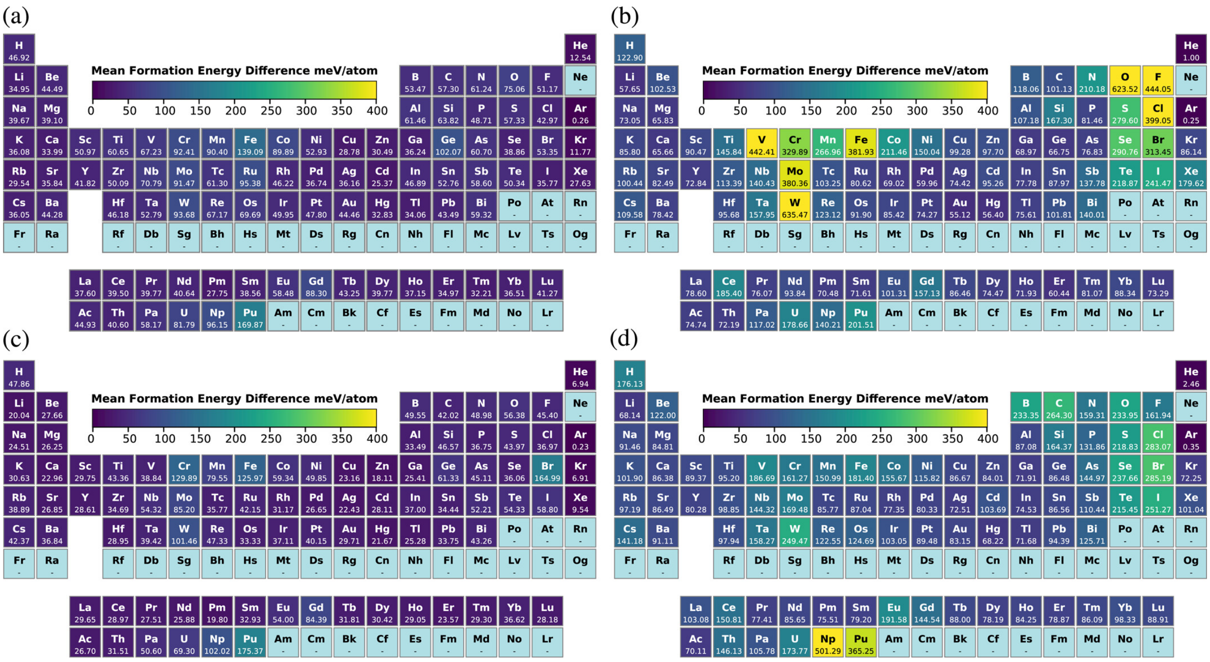


Figure 6: Heat maps of mean formation energy difference of (a) CHGNet, (b) M3GNet, (c) MACE and (d) ALIGNN [39].

the heat maps of M3GNet and ALIGNN, most of the elements are in blue region, the formation energies predicted for some elements exhibit significant errors in green or even yellow region (more than 200 meV/atom). For the prediction of transition metals, M3GNet and ALIGNN predicted formation energies with significantly larger errors compared to MACE and CHGNet, especially for elements like V, Cr, Mo and W.

2.1.3 Limitations

Yu et al. [39] reported the convergence issues of uMLIPs during structural relaxation. From their analysis (Fig. 7) it can be seen that all tested uMLIPs have limited unconverged cases when relaxing the atomic positions only, while the unconverged cases increases significantly if relaxing both the atomic positions and the cell parameters. Based on Yu's assessment, it can be concluded that among the four tested uMLIPs, CHGNet and MACE can perform the structural relaxation more robustly. In another assessment, Gonzales et al. [40] investigated the convergence behavior of MACE, CHGNet, and M3GNet by performing structural relaxation on 10,773 structures taken from the Materials Project. Their results showed a different convergence behavior compared to those of Yu et al. Specifically, the convergence rate of CHGNet is only 40.9%, while that of M3GNet is 62.9%. Their study confirmed the robustness (of convergence behavior) of MACE, with a convergence rate of 95.1%. The two studies indicated that the convergence issues of uMLIPs are a significant limitation when using uMLIPs for molecular statics simulations. In contrast, traditional EIPs can provide more robust simulations with better convergence.

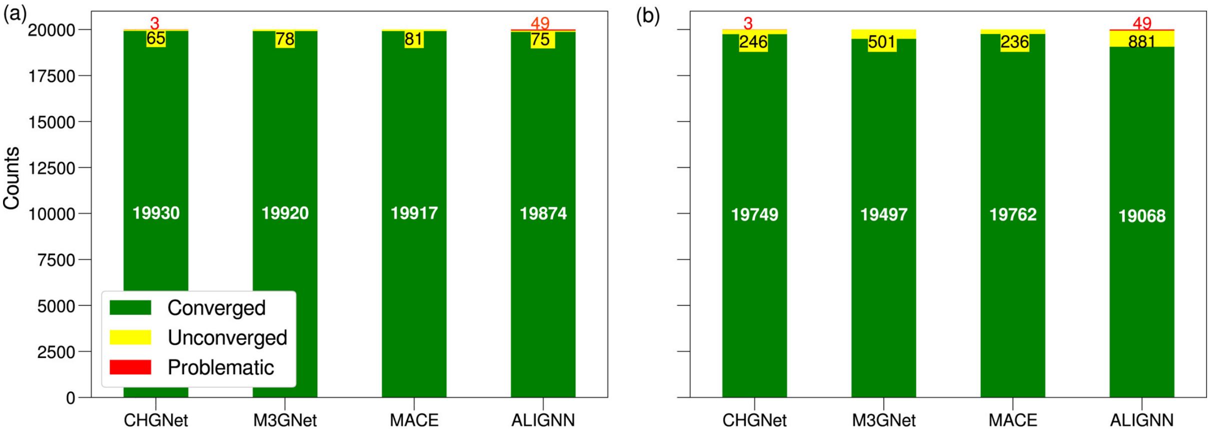


Figure 7: Convergence assessment of uMLIPs. [39] (a) Only the atomic position was relaxed. (b) Both the atomic position and the cell size were relaxed.

The training dataset of uMLIPs are obtained from Materials Project database, where only data of perfect lattice structures are included [41]. Since GB segregation is a type of defect of the lattice, the prediction of GB segregation is out-of-distribution for uMLIPs. Deng et al. [42] investigated the performance of M3GNet, CHGNet and MACE-MP-0 on several out-of-distribution material modeling tasks such as calculating surface energies. A PES softening behavior which was characterized by a systematic underprediction of energies and forces is observed from their benchmark tests. This phenomenon is illustrated in Fig. 8, as we can see that the high-energy regions (indicated in red) of the PES of uMLIPs are lower than the PES described by DFT. This softening behavior is attributed to the uMLIP training datasets, which include only structures at near equilibrium states. It can be corrected by fine-tuning with high-energy out-of-distribution data points, such as those from intermediate states during structural relaxation.

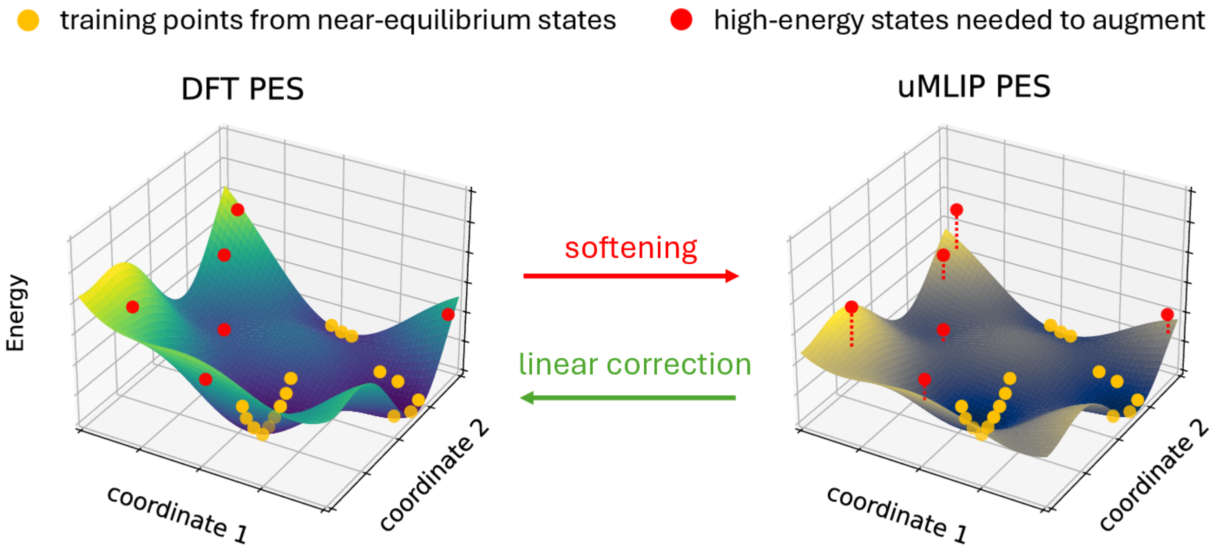


Figure 8: Schematic illustration of PES softening behavior, PES described by DFT (left) and uMLIPs (right) [42].

The PES softening behavior of pre-trained uMLIP models highlights the importance of fine-tuning the models. The accuracy of out-of-distribution simulation tasks such as calculating GB segregation energies might be reduced when using pre-trained model of uMLIPs due to this systematic underprediction phenomenon. Deng et al. also indicated that the PES softening might be mitigated by increasing the number of model parameters of the uMLIPs. Therefore, uMLIPs such as MACE-MP-0 and SevenNet-0 might be less influenced by the PES softening compared to CHGNet and M3GNet due to a larger model size.

2.2 Grain Boundary Segregation Simulations Based on DFT Methods

In this subsection, literature that simulated GB segregation based on DFT calculations will be reviewed. The review aims at learning the methodology of simulating GB segregation, and more importantly, how to present the data and find the correlation between GB segregation behavior and the investigated material properties. Since simulation based on DFT calculations is the most accurate methods, they can serve as a valuable reference to evaluate the performance of uMLIPs.

Ito et al. [43] studied the segregation behavior of nine transition metals (Ti, V, Cr, Mn, Co, Ni, Cu, Nb and Mo) at nine face-centered cubic (FCC) Fe GBs as shown in Fig. 9. Their research employed the first-principles calculations based on DFT using the projector augmented wave (PAW) method within the Vienna Ab-initio Simulation Package (VASP). For exchange–correlation effects the generalised gradient approximation with the Perdew-Burke-Ernzerhof (PBE) functional is used.

The segregation energies of all inequivalent sites within 2.5 Å for each GBs were calculated and the results with respect to the Voronoi volume of the segregation site were presented as shown in Fig. 10(a) and in a histogram as shown in Fig. 10(b). The segregation energy can be understood

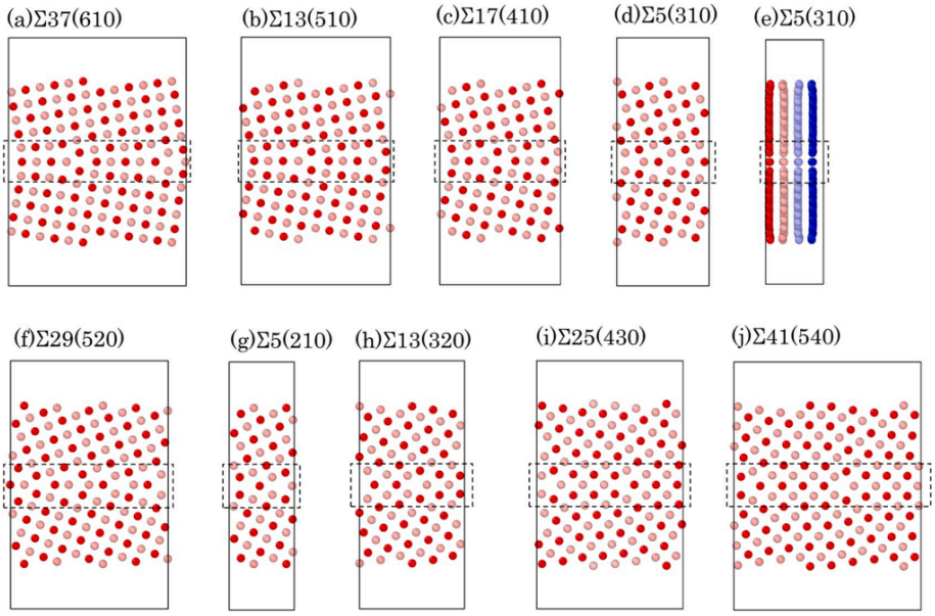


Figure 9: The nine simulated GB strucutures in the study of Ito et al. [43]. The atoms are presented in the axis direction [001]. The sites within region 2.5Å from the GB plane are indicated by the dashed lines. Fig. (e) is the same GB as (d) which is viewed from a different axis.

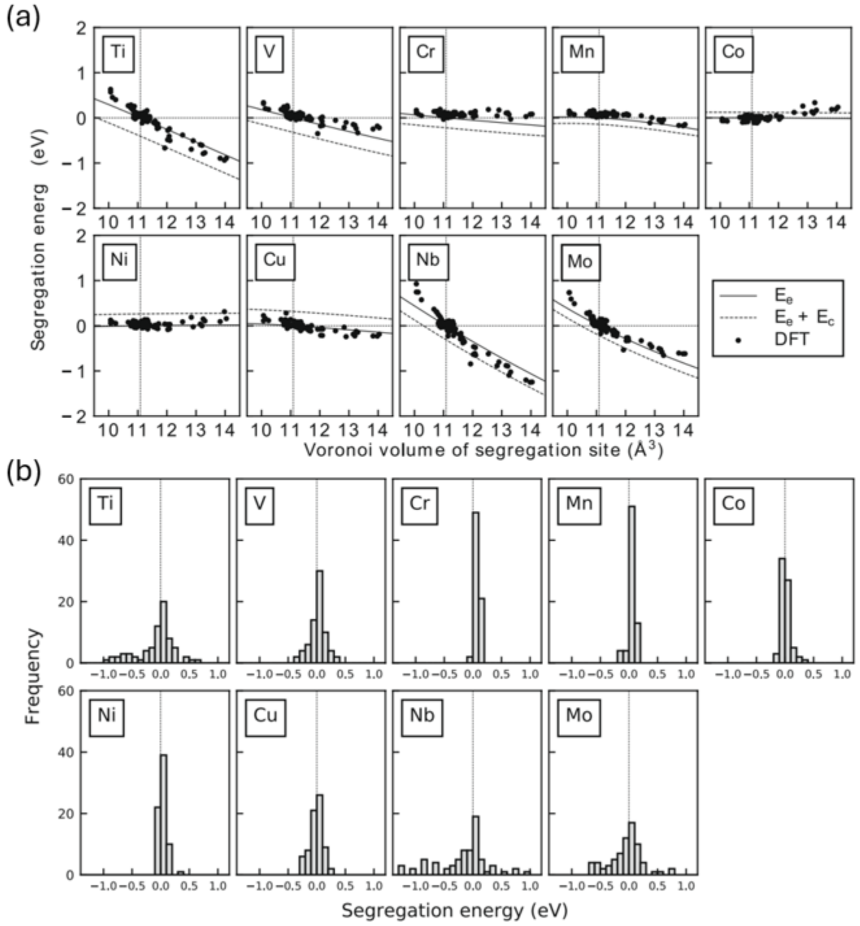


Figure 10: Results of segregation energies calculated by Ito et al. via DFT calculations [43]. (a) Segregation energy vs. Voronoi volume of the segregation sites, E_{seg} is the elastic contribution of the segregation energy. (b) Histogram of the segregation energy calculated at different sites of the nine GBs.

as the difference in energy between having the solute at the GB compared to being in the bulk material. If E_{seg} is negative, it suggests that the solute prefers to segregate to the GB, while a positive value indicates a preference for the solute to remain in the bulk. According to their results, the segregation energy can be well correlated with the Voronoi volume of the segregation site. For example, for Cu and other alloying elements including Ti, Mo, V, and Nb, the segregation energy will shift to more negative at segregation sites with a larger Voronoi volume than the bulk. When the Voronoi volume at segregation sites is smaller compared to the bulk, the segregation energy will be positive, which means segregation is unfavorable. Different trends are observed for Ni, Co, and Cr. However, the dependence of segregation energy on Voronoi volume is different. Among the mentioned alloying elements, Cr is almost independent of Voronoi volume, while Cu exhibits the least dependence among the remaining elements. The dependence can also be well reflected from the histogram, the alloying element having a larger dependence will show a larger variance in segregation energy in the histogram. For instance, Cr, as mentioned with the least dependence, exhibits a histogram with segregation energy localized around zero.

To further explain the physical origin of the segregation energy and the dependence on Voronoi volume, Ito el al. quantified the elastic contribution to the segregation energy. They concluded that the grain boundary segregation energy of alloying elements can be attributed to the elastic energy arising from the atomic radius difference between the Fe atom and the solute atom. Solute atoms with larger radii prefer to segregate at sites with larger Voronoi volumes, as this reduces the elastic energy, leading to more negative segregation energy. Additionally, the segregation tendency of solutes to each investigated GBs were quantified by calculating the effective segregation energy for each GB. Effective segregation energy can be calculated based on temperature, the bulk composition of the solute element, and the solute composition at GB. This allows for the calculation of effective segregation energy under conditions consistent with existing experimental studies, enabling direct comparison with experimental results. From Fig. 11 it can be seen that, the segregation of Cu at the investigated GBs of FCC Fe are not intense, and it is of weak dependence on GB character. According to Ito et al., the calculated effective segregation energy can be validated by previous experimental results, confirming the accuracy of their simulation.

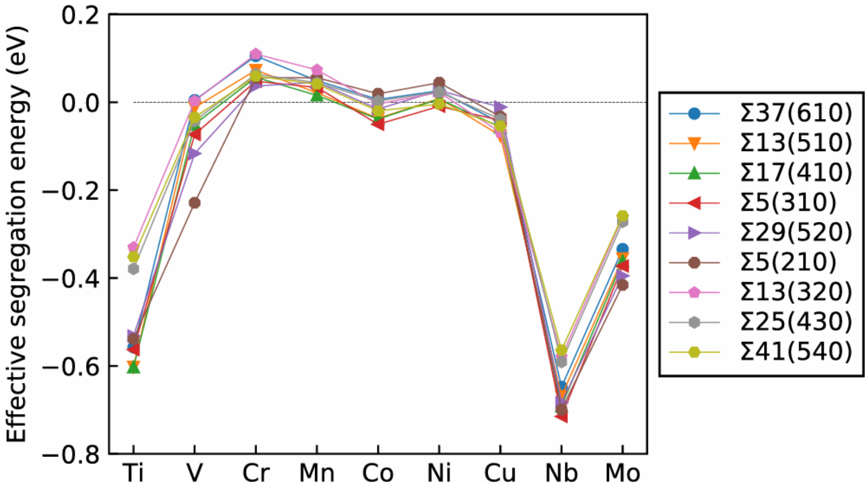


Figure 11: Effective segregation energies (eV) per alloying element per GB calculated by Ito [43]. Effective segregation energies were calculated to validate their simulations by comparing them with existing experimental results.

Similar to Ito's work, Mai et al. [12] investigated the segregation and co-segregation behavior of transition metals (including phosphorus (P) in their later paper [44]) at body-centered cubic (BCC) Fe GBs via first-principle calculations. They selected four symmetric tilt grain boundaries (STGBs) as shown in Fig. 12 for simulations: $\Sigma3(1\bar{1}1)[110]$, $\Sigma3(1\bar{1}2)[110]$, $\Sigma9(2\bar{2}1)[110]$ and $\Sigma11(3\bar{3}2)[110]$. The co-segregation was assessed by calculating the increment segregation energy $E_{\text{seg}}^{\text{inc}}$, which is defined as the energy changed when a second solute atom Y segregating to a GB with solute X already presented at the most energetically favourable site at the pure GB. While most GB co-segregation studies have focused binary systems, Mai's work expanded this to include ten transition metals, providing valuable insights for research in the related field.

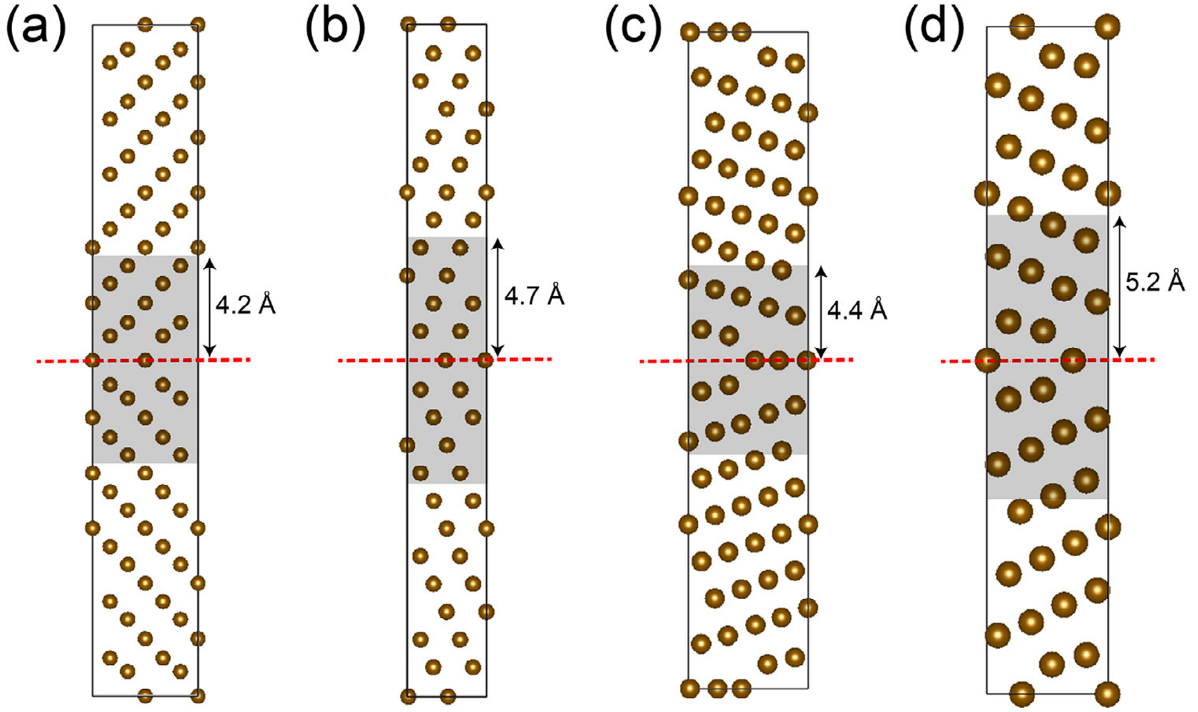


Figure 12: The four simulated GB strucutures in the study of Mai et al. [12]: **(a)** $\Sigma3(1\bar{1}1)[110]$, **(b)** $\Sigma3(1\bar{1}2)[110]$, **(c)** $\Sigma9(2\bar{2}1)[110]$ and **(d)** $\Sigma11(3\bar{3}2)[110]$. The GB plane is indicated with red dashed lines, and the calculated sites are within the shaded region.

From the results of Mai et al. as shown in Figs. 13 & 14, it can be seen that the Cu has a high tendency of both single solute segregation and co-segregation with other elements. There's a high co-segregation tendency for Cu with Ti, Ni and Nb. The co-segregation of V as the second solute seems unfavorable for Cu and all other alloying elements. In addition to the alloying elements, the GB character also influece the segregation tendency. The segregation at GB $\Sigma3(1\bar{1}2)$ is generally less favorable compared to other GBs. The lower segregation tendency of stacking-fault/twin-type $\Sigma3(1\bar{1}2)$ GB can be attributed to the lower relative excess volume in the GB, which is at the same time a distinguishing characteristic of this type of GBs. When analyzing the results of co-segregation, Mai et al. argued that the bulk solute-solute interaction energies cannot adequately represent the GB co-segregation interaction. As an example, the Cu-Cu interactions in the bulk exhibit an attractive force ranging from -0.23 eV to -0.08 eV within the first to third coordination shells. While the Cu-Cu interaction in the minimum energy configurations across the four investigated GBs varies from 0.60 eV to -0.13 eV.

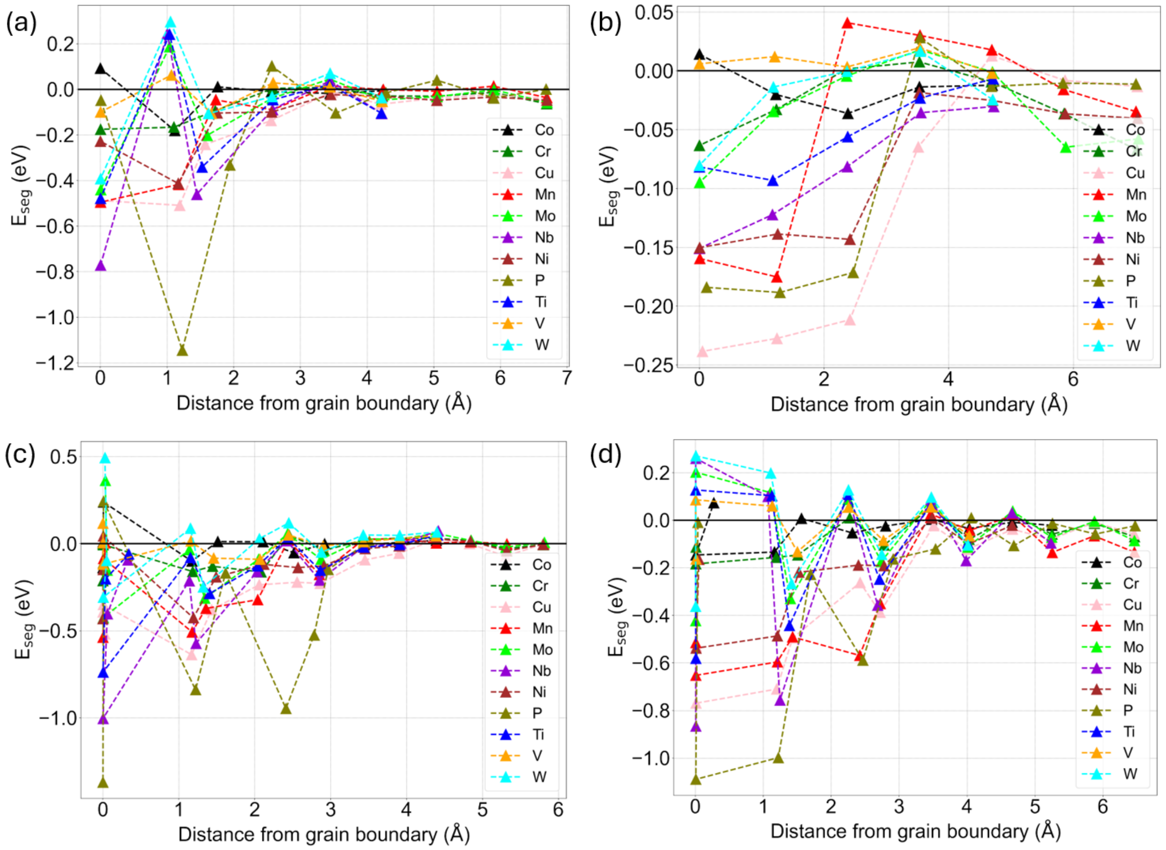
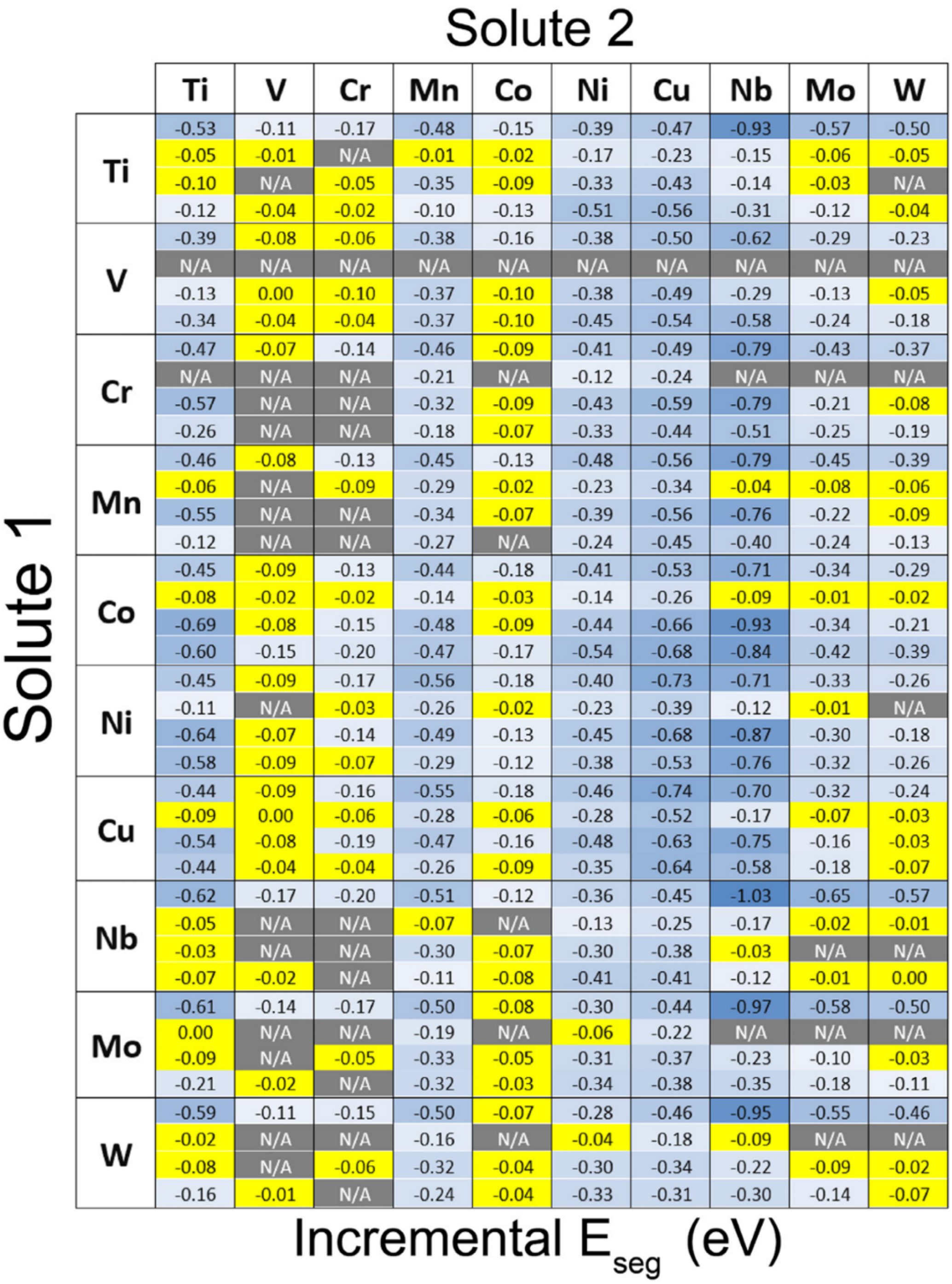


Figure 13: Segregation energy profiles for a particular element to segregate (1 atom) to a GB plotted against the distance from the GB plane. (a) $\Sigma3(1\bar{1}1)[110]$, (b) $\Sigma3(1\bar{1}2)[110]$, (c) $\Sigma9(2\bar{2}1)[110]$ and (d) $\Sigma11(3\bar{3}2)[110]$ [44].

Both studies by Ito's group and Mai's group indicated that the segregation behaviors of transition metals cannot be solely studied on an elemental basis, the GB character should also be considered. The dependence of segregation preference on Voronoi volume was also discussed in both studies. In Ito's work, the physical origin of this dependence was examined through analysis of the force contributions to segregation energy. Mai's work, meanwhile, examined the co-segregation of transition metals in detail and investigated how segregation influenced cohesion. The two studies mentioned above covers the segregation behaviors of transition metals at FCC Fe and BCC Fe GBs, providing valuable insights to the thesis project, especially on presenting the calculated data and performing a quantitative analysis. There are other relevant literature based on first-principle calculations: Xu et al. [45] studied the electrical origin of 3d-transition metals segregation at Σ3(111) and Σ11(332) BCC Fe GBs. Yuasa and Mabuchi [9] investigated the effects of Cu segreagtion at Σ3(111)[110] BCC Fe GB using fully relaxed tensile and shear test via DFT simulations. Numerous other studies have examined the segregation of various elements, including hydrogen segregation in Fe GBs, which is a primary cause of hydrogen embrittlement. As the methodologies in these studies are similar, they will not be discussed here to avoid repetition.



Solute 2

Solute 1

	Ti	V	Cr	Mn	Co	Ni	Cu	Nb	Mo	W
Ti	-0.53	-0.11	-0.17	-0.48	-0.15	-0.39	-0.47	-0.93	-0.57	-0.50
	-0.05	-0.01	N/A	-0.01	-0.02	-0.17	-0.23	-0.15	-0.06	-0.05
	-0.10	N/A	-0.05	-0.35	-0.09	-0.33	-0.43	-0.14	-0.03	N/A
	-0.12	-0.04	-0.02	-0.10	-0.13	-0.51	-0.56	-0.31	-0.12	-0.04
V	-0.39	-0.08	-0.06	-0.38	-0.16	-0.38	-0.50	-0.62	-0.29	-0.23
	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	-0.13	0.00	-0.10	-0.37	-0.10	-0.38	-0.49	-0.29	-0.13	-0.05
	-0.34	-0.04	-0.04	-0.37	-0.10	-0.45	-0.54	-0.58	-0.24	-0.18
Cr	-0.47	-0.07	-0.14	-0.46	-0.09	-0.41	-0.49	-0.79	-0.43	-0.37
	N/A	N/A	N/A	-0.21	N/A	-0.12	-0.24	N/A	N/A	N/A
	-0.57	N/A	N/A	-0.32	-0.09	-0.43	-0.59	-0.79	-0.21	-0.08
	-0.26	N/A	N/A	-0.18	-0.07	-0.33	-0.44	-0.51	-0.25	-0.19
Mn	-0.46	-0.08	-0.13	-0.45	-0.13	-0.48	-0.56	-0.79	-0.45	-0.39
	-0.06	N/A	-0.09	-0.29	-0.02	-0.23	-0.34	-0.04	-0.08	-0.06
	-0.55	N/A	N/A	-0.34	-0.07	-0.39	-0.56	-0.76	-0.22	-0.09
	-0.12	N/A	N/A	-0.27	N/A	-0.24	-0.45	-0.40	-0.24	-0.13
Co	-0.45	-0.09	-0.13	-0.44	-0.18	-0.41	-0.53	-0.71	-0.34	-0.29
	-0.08	-0.02	-0.02	-0.14	-0.03	-0.14	-0.26	-0.09	-0.01	-0.02
	-0.69	-0.08	-0.15	-0.48	-0.09	-0.44	-0.66	-0.93	-0.34	-0.21
	-0.60	-0.15	-0.20	-0.47	-0.17	-0.54	-0.68	-0.84	-0.42	-0.39
Ni	-0.45	-0.09	-0.17	-0.56	-0.18	-0.40	-0.73	-0.71	-0.33	-0.26
	-0.11	N/A	-0.03	-0.26	-0.02	-0.23	-0.39	-0.12	-0.01	N/A
	-0.64	-0.07	-0.14	-0.49	-0.13	-0.45	-0.68	-0.87	-0.30	-0.18
	-0.58	-0.09	-0.07	-0.29	-0.12	-0.38	-0.53	-0.76	-0.32	-0.26
Cu	-0.44	-0.09	-0.16	-0.55	-0.18	-0.46	-0.74	-0.70	-0.32	-0.24
	-0.09	0.00	-0.06	-0.28	-0.06	-0.28	-0.52	-0.17	-0.07	-0.03
	-0.54	-0.08	-0.19	-0.47	-0.16	-0.48	-0.63	-0.75	-0.16	-0.03
	-0.44	-0.04	-0.04	-0.26	-0.09	-0.35	-0.64	-0.58	-0.18	-0.07
Nb	-0.62	-0.17	-0.20	-0.51	-0.12	-0.36	-0.45	-1.03	-0.65	-0.57
	-0.05	N/A	N/A	-0.07	N/A	-0.13	-0.25	-0.17	-0.02	-0.01
	-0.03	N/A	N/A	-0.30	-0.07	-0.30	-0.38	-0.03	N/A	N/A
	-0.07	-0.02	N/A	-0.11	-0.08	-0.41	-0.41	-0.12	-0.01	0.00
Mo	-0.61	-0.14	-0.17	-0.50	-0.08	-0.30	-0.44	-0.97	-0.58	-0.50
	0.00	N/A	N/A	-0.19	N/A	-0.06	-0.22	N/A	N/A	N/A
	-0.09	N/A	-0.05	-0.33	-0.05	-0.31	-0.37	-0.23	-0.10	-0.03
	-0.21	-0.02	N/A	-0.32	-0.03	-0.34	-0.38	-0.35	-0.18	-0.11
W	-0.59	-0.11	-0.15	-0.50	-0.07	-0.28	-0.46	-0.95	-0.55	-0.46
	-0.02	N/A	N/A	-0.16	N/A	-0.04	-0.18	-0.09	N/A	N/A
	-0.08	N/A	-0.06	-0.32	-0.04	-0.30	-0.34	-0.22	-0.09	-0.02
	-0.16	-0.01	N/A	-0.24	-0.04	-0.33	-0.31	-0.30	-0.14	-0.07

Incremental E_{seg} (eV)

Figure 14: The incremental energy of segregation of solute 2 when solute 1 is present. The 4 values in each cell from top to bottom correspond to $\Sigma3(1\bar{1}1)[110]$, $\Sigma3(1\bar{1}2)[110]$, $\Sigma9(2\bar{2}1)[110]$ and $\Sigma11(3\bar{3}2)[110]$ GBs, respectively. More negative values (in deeper color) indicate that segregation is more likely to occur. "N/A" values indicate that segregation of the second solute is unfavourable. Yellow color indicate that the co-segregation is negligible [12].

2.3 Grain Boundary Segregation Simulations Based on EIPs

Simulations based on DFT calculations provide high accuracy in predicting atomic-scale properties. These methods are based on solving Schrödinger equations without using any empirical parameters, ensuring the accuracy of calculation output. However, the expensive computational cost of the calculations limits the computational scale to hundreds of independent atoms. Such GB models might not be realistic in some simulation tasks compared to a real GB structure with enormous amount of atoms. On the other hand, simulations using EIPs are computationally much faster, allowing the modeling of systems with millions of atoms.

Currently, most of the EIPs for simulating pure metal systems or metallic alloy systems are developed based on EAM or MEAM. The EAM potential is an approach combining both pairwise interactions and contributions from multiple atoms, making it particularly suitable for modeling metallic systems [10]. In these systems, the behavior of atoms is significantly affected by the surrounding electron density provided by nearby atoms. MEAM potentials are more versatile by incorporating directional bonding, allowing them to handle a broader range of materials such as ceramics. MEAM potentials are more straightforward for simulating alloy systems compared to EAM potentials [24, 46]. Although MEAM potentials include an additional angular term over EAM potentials and some literatures have shown that the performance of MEAM potential are better than EAM potentials in some applications [47, 48], it remains unclear which method outperforms the other in simulating GB segregation.

Many studies have examined the GB segregation behavior via EIP-based molecular dynamic (MD) simulation. For example, Gao et al. [49] investigated segregation behavior of Cu at STGB $\Sigma3(112)$ via MD/MC (Monte Carlo) simulations with a Fe–Cu EAM potential. Similarly, Zhou et al. [50] studied the segregation preference of Cr at several BCC Fe GBs for different GB structures and different Cr composition using MD/MC simulation with a Fe-Cr EAM potential. The methodology of simulating GB segregation based on EIPs is similar to the simulation using DFT methods which has been reviewed in previous subsection. Although EIP-based simulations are less accurate than DFT, they offer the advantage of simulating more complex structural models due to their lower computational cost. Unlike the bi-crystal GB models that are usually used for DFT-based simulations, polycrystalline GB models can be used for EIP-based simulation. These polycrystalline GB models can be randomly generated and contain tens of grains, allowing the simulation of more realistic GB features such as triple-junctions and more GB structures simultaneously. Additionally, It is indicated in literature that STGBs with low Σ value cannot adequately represent the diverse and intricate structures found in polycrystals [51]. The generation of polycrystalline GB models can incorporate GBs with higher Σ values in simulations, which can enhance the sampling of local atomic environment space, and thereby improving the efficiency of segregation energy predictions.

One illustrative example is the segregation study by Ito et al. [18], where they investigated the hydrogen segregation at BCC Fe GBs using a polycrystalline model. Their study comprehensively analyzed the segregation energies of hydrogen at approximately 17 million interstitial sites, a scale made possible by the use of EIPs. The calculated effective segregation energies ranged from -0.48 to -0.42 eV, closely aligning with experimental data, demonstrating the model's reliability. Their study identified that hydrogen predominantly segregates at octahedral sites (as shown in Fig. 15) with larger Voronoi volumes due to uniaxially distorted crystal structures near GBs. This ability to simulate diverse GB structures simultaneously demonstrates the strength of using polycrystalline models with EIP-based simulations.

GB triple junctions or more complex GB junctions are considered to be more "realistic" compared to bi-crystal GB models [52]. Understanding segregation behavior at GB triple junctions is critical when simulating polycrystalline structures. For example, Adlakha and Solanki [53] studied the solute segregation behavior at triple junctions of BCC Fe GBs using EIP-based MD

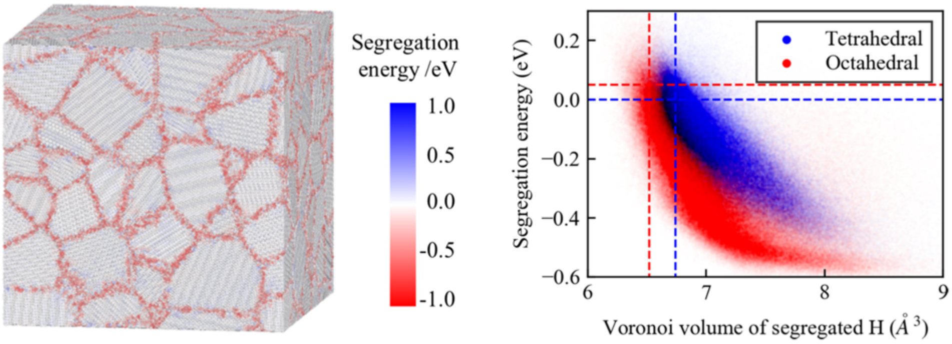


Figure 15: Nano-polycrystalline GB model with segregation energy computed. The right image is the plot of segregation energy against Voronoi volume, highlighting the prefered segregation of H at octahedral sites [18].

simulation. Their simulation results indicated large solute binding energies at sites within the core regions of triple junctions would lead to an intense solute segregation, which cannot be observed if using bi-crystal GB models.

In recent years, EIP-based MD and MC simulations have become popular in the study of high-entropy alloys, the segregation behavior at GBs of alloying elements can be well simulated using a polycrystalline GB model. He et al. [54] studied the influence of elemental segregation at GBs on the mechanical properties of FeNiCrCoCu high-entropy alloys using combined MD and MC simulations with a five elements (FeNiCrCoCu) EAM potential. The simulations were applied on a polycrystalline model containing 40 randomly orientated Voronoi cells. Their results, as shown in Fig. 16, revealed that Cu (indicated in blue in the polycrystalline model) dominated the segregation at GBs, while the other alloying elements preferred to segregate into the bulk. This segregation behavior was confirmed by EDX elemental mapping. It's indicated that the high concentration of Cu at GBs can weaken the mechanical properties by initiating GB cracking based on their results.

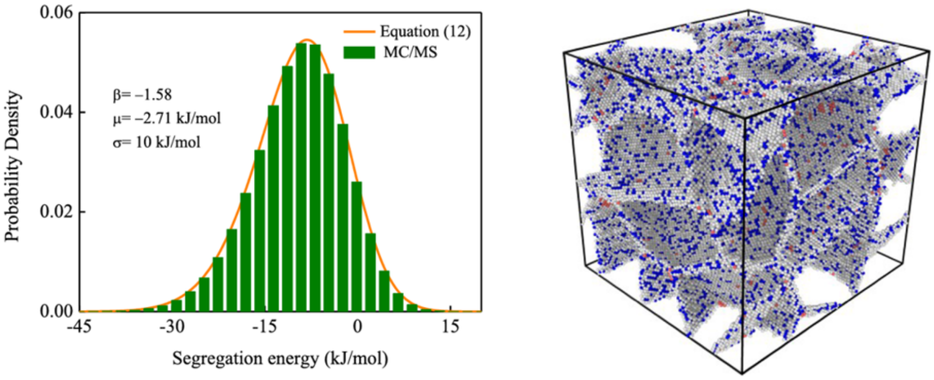


Figure 16: Equillibrium state of polycrystalline GB model of FeNiCrCoCu high-entropy alloys simulated by MD and MC, 4 at.% Cu and T=300K [54]. Segregation energy distribution is plotted, the orange line is the theoretical values calculated. In the polycrystalline GB model, Cu atoms are highlighted in blue.

In computational simulation studies, presenting the results in an organized and structured manner is crucial for effective comprehension and interpretation. In many simulation studies related to GB segregation, the segregation behaviors are usually described by a single-value segregation energy (for example, the DFT studies reviewed in previous subsection). However, such simplification ignores the variation of substitutional sites within individual GB and the fact that a polycrystal contains different types of GBs [55]. To address these limitations, researchers have developed strategies that utilize a GB energy spectrum to represent the segregation behavior. The GB energy spectrum is more effective because it captures the full range of segregation energies across different GB sites, accounting for the complex distribution of elements and their specific interactions with various GBs. For example, in the work of He et al. as shown in Fig. 16, the GB segregation energy distribution of Cu (4 at.% Cu and T=300K) highlights that Cu dominates the segregation behavior, while other elements like Ni and Fe are depleted from the GB region. From this spectrum, correlation functions can be constructed, offering insights into how segregation behavior may influence alloy properties when combined with results on strengthening effects.

2.4 Grain Boundary Segregation Studies Based on Machine Learning Techniques

The construction of GB segregation energy spectra offers a more comprehensive approach to analyzing segregation behavior, as previously mentioned. The computations needed for the construction of a segregation energy spectrum remain challenging to this day. With the rapid development of machine learning techniques, these methods are not only used in developing uMLIPs but also in reducing the computational cost of building segregation energy spectra.

Huber et al. [56] designed a new computational framework to calculate the full distribution of GB segregation energies with modest computational effort. This framework enables the computation using quantum mechanics DFT calculations with higher accuracy but more computational cost. For the calculation of the segregation spectrum, firstly, investigation involving

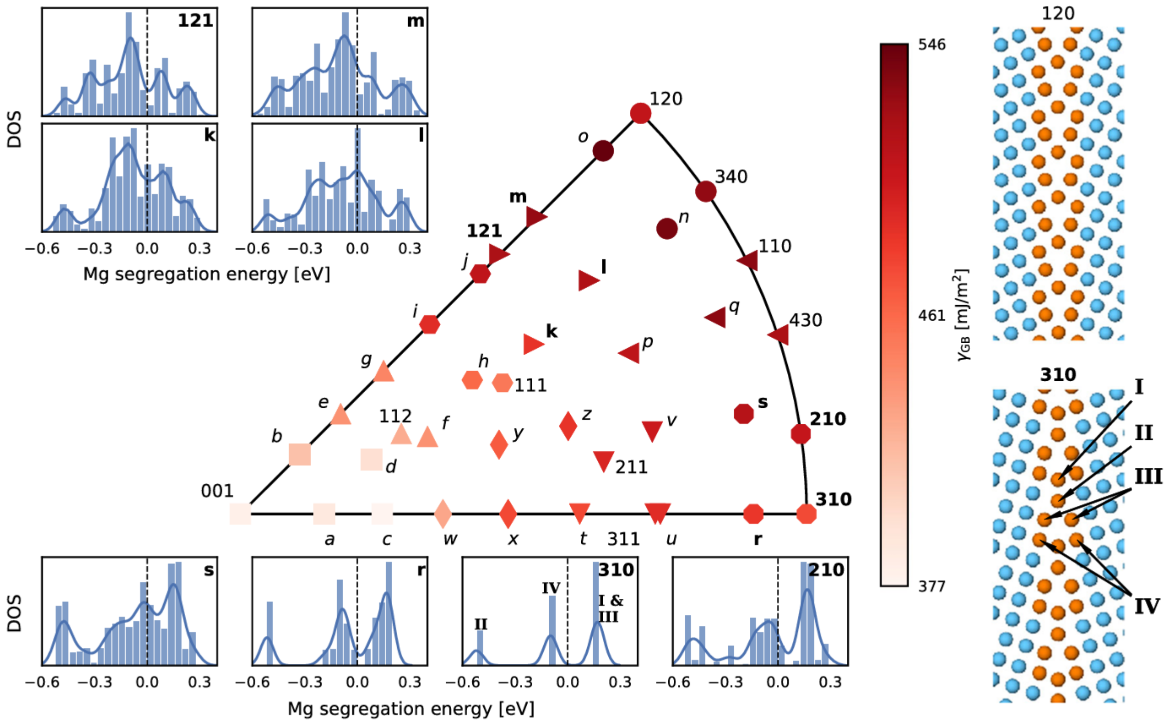


Figure 17: Segregation energy distributions predicted based on machine learning. The polar plot includes GB interfacial energies for 38 different boundaries in the boundary-plane fundamental zone for pure Al 53.1°[001] GBs with different normal vectors [56].

six solute species segregating across numerous sites at 38 low and high-symmetry boundaries were conducted, yielding a dataset consisting of over a million segregation energies. Then, Huber et al. identified a set of machine learning descriptors based solely on the local atomic environment around an atom within a cutoff radius of the unsegregated GB. These machine learning descriptors enable to forecast the segregation energy distribution and the segregation isotherm accurately. They applied this computational framework to study the segregation behavior of Mg at Al GBs as shown in Fig. 17, making it possible to predict GB segregation energy at arbitrary GBs for Al–Mg system.

In more recent works done by Schuh and Wagih, they provided two new machine learning-based frameworks for developing the segregation energy spectrum. The first framework (the upper framework shown in Fig. 18), proposed in 2020 [57], utilized segregation energies calculated from structural relaxation using EIPs to train machine learning models. This framework significantly improves computational efficiency by replacing resource-intensive atomistic calculations with machine-learned predictions based on atomic descriptors. The model can predict the segregation

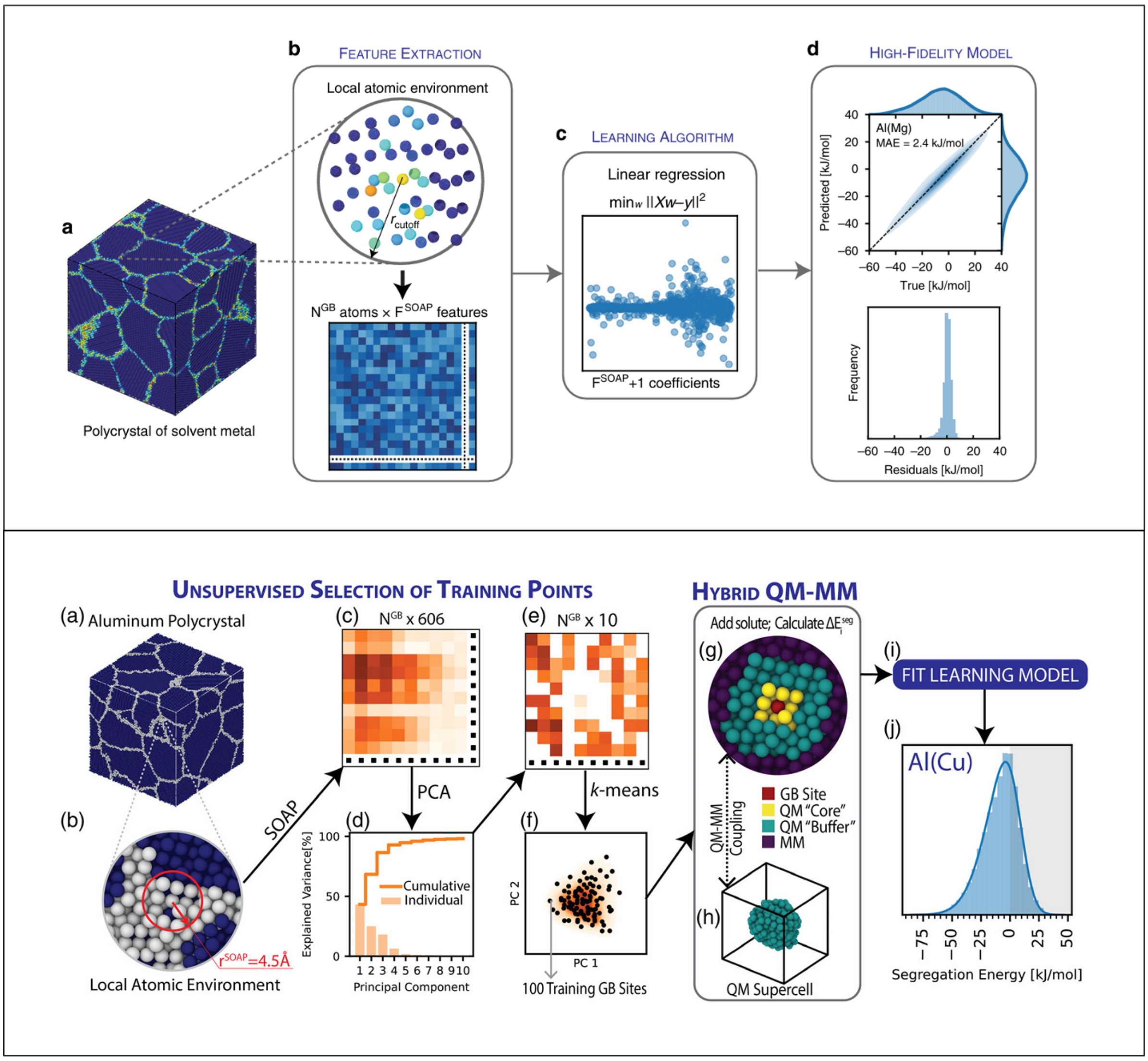


Figure 18: Workflow of the frameworks developed by Schuh and Wagih. Top: Segregation energy spectra for polycrystals developed based on machine learning that trained by results structural relaxation using EIPs [57]. Bottom: Workflow of DFT based learning framework for GB segregation energies [58].

energies for solute atoms at different GB sites in a polycrystal based solely on the undecorated local atomic environment of the grain boundary sites. This approach enabled the creation of an extensive database of GB segregation spectra for over 250 metal-based binary alloys as shown in Fig. 19, offering valuable insights into how different solutes interact with GBs across a wide variety of alloy systems. However, this model has a limitation that the segregation energies are calculated via strutural relaxation at zero K, making predictions at elevated temperature less accurate.

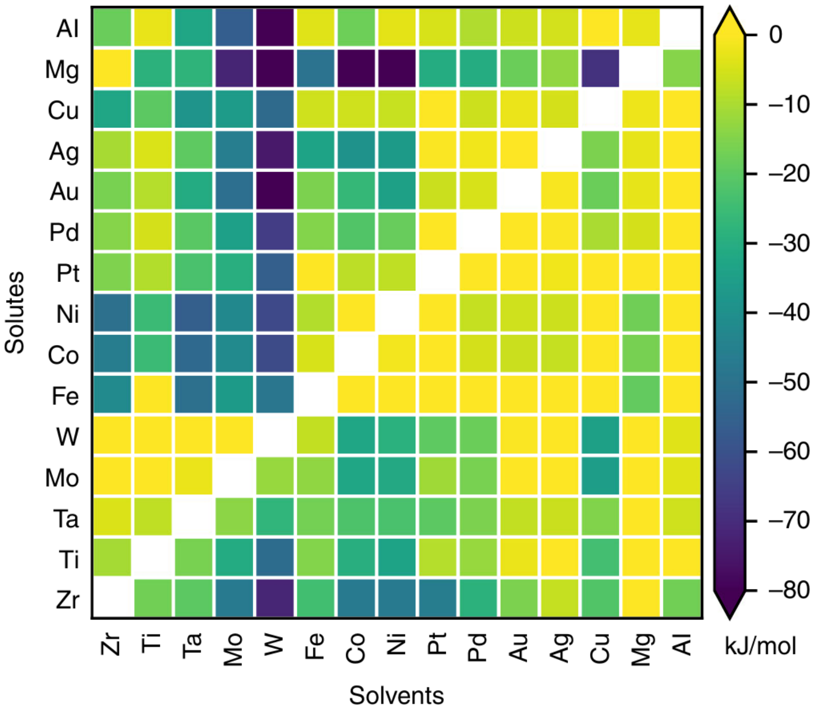


Figure 19: A summarization (Pettifor map) of segregation energy specra covering 15 alloying elements which can be used to predict the segregation tendency [57].

The accuracy of the first framework proposed by Schuh and Wagih is limited by the training dataset calculated using EIPs. While the second framework (the bottom framework shown in Fig. 18), developed in 2022 [58], improves the accuracy of GB segregation predictions by directly integrating DFT calculations into the machine learning model. By combining quantum mechanics with machine learning, this model achieves near quantum-level accuracy in predicting segregation tendencies across a broader range of alloy systems, particularly those where reliable EIPs are not available. The framework provides a tool that not only covers a larger chemical space but also produces higher fidelity segregation spectra for complex alloys. This method has been successfully applied by Schuh and Wagih to a comprehensive database of Al-based alloys, which is able to fill a crucial gap in the understanding of GB segregation behavior in polycrystals.

These above-mentioned works represent the most state-of-the-art methods for the study of GB segregation. Both frameworks proposed by Schuh and Wagih represent a significant improvement over earlier models, such as those developed by Huber et al., which were limited to bi-crystal systems and specific GB configurations. By focusing on polycrystals, Schuh and Wagih's models better capture the complexities of real-world materials, where a wide variety of GB types and orientations coexist within the same structure. These models offer the most accurate predictions to date for polycrystals, enabling more effective alloy design. Their database of GB segregation energies for over 250 binary alloys is a key resource for researchers seeking to design materials with tailored properties, including improved strength, corrosion resistance, or embrittlement resistance [51]. More importantly, their frameworks are extensible, meaning they can be adapted to other base metals and alloy systems, which makes them particularly useful for future materials discovery and optimization efforts.

For the framework proposed by Schuh and Wagih, the machine learning techniques are applied to the local atomic environment descriptors, which are restricted to predicting segregation energies exclusively. While most recently, Scheiber et al. [59] proposed another method to investigate GB segregation through machine learning, which also allows a broader range of prediction tasks. The machine learning descriptors of their model include physical quantities, such as partial cohesive energy, the volume mismatch between the solute and matrix, and the Voronoi volume of the GB site, providing their model with a stronger physical foundation compared to uMLIPs or the previously reviewed machine learning frameworks. The model can predict GB segregation energies with good accuracy (as shown in Fig. 20) based on the input computational data from available literature.

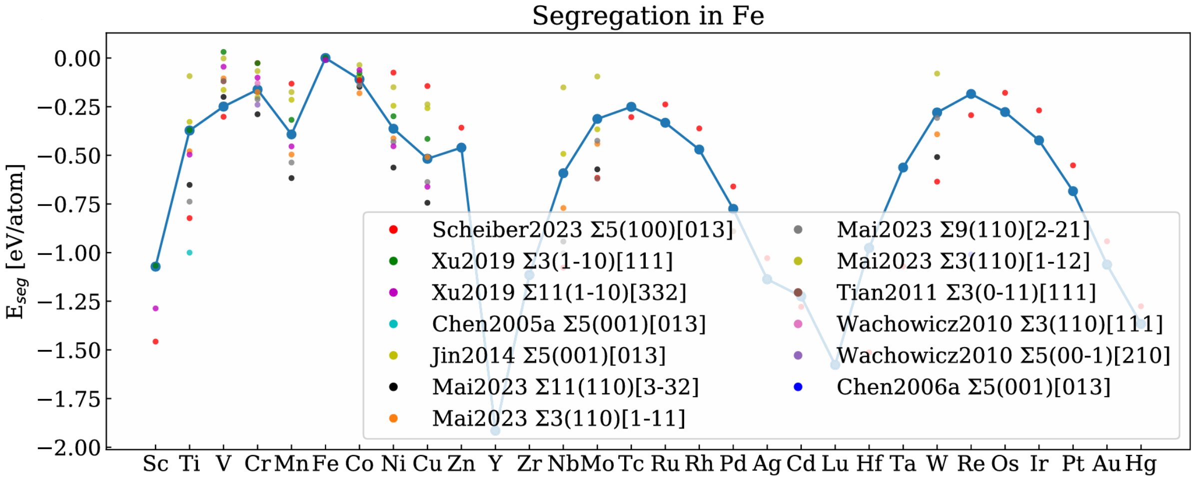


Figure 20: Comparison of predicted lowest segregation energy between the model of Scheiber et al. (blue) with data from the other literature [59].

A significant challenge in assessing the performance of uMLIPs on predicting GB segregation arises from the scattered and inconsistent available data. The model of Scheiber et al., however, offers an effective approach for predicting segregation behavior in a consistent manner (consistency in computational parameters). Furthermore, due to the strong physical foundation of the machine learning model, their model can also predict other related physical properties, such as strengthening energies in addition to GB segregation energy. However, as shown in Fig. 20, their model seems to overlook the influence of different GB structures on segregation tendency. If this is the case, the model may only be suitable for predicting the general segregation tendencies of solute elements within a given matrix, rather than providing detailed, structure-specific predictions. This also highlights the difficulty of proposing an evaluation framework for assessing the performance of uMLIPs, due to the limited and scattered available GB segregation data from various literature.

3 Research Question and Objective

This section provides a brief summary of the research question and objectives of this Master's thesis project. Since experimental methods face challenges in precisely analyzing the effects of GB segregation, computational simulation is a more promising approach to study the GB segregation behavior of Cu and to explore solutions for Cu contamination. Although DFT calculations offer high accuracy, their high computational cost limits them to simulating systems with up to 1000 atoms. Simulations using EIPs, on the other hand, face the challenge of a lack of reliable potentials, especially when modeling the co-segregation of Cu with various other alloying elements at Fe GBs to investigate Cu segregation behavior. Recently developed uMLIPs offer a potential solution to this issue, as they can be applied to simulations involving nearly all elements in the periodic table. However, evaluations of uMLIPs remain limited, particularly regarding their performance in simulating GB segregation. Therefore, an evaluation of these potentials is necessary before they can be fully applied in the study of GB segregation. The evaluation will be applied on the best available uMLIP models according to the Matbench Discovery (Table 2), with an emphasis on their performance in simulating GB segregation across various GBs and as many solute elements as possible, to validate their accuracy.

4 Methodology

In the Literature Review (section 2.3), we highlighted that segregation energy spectra can provide a more comprehensive representation of segregation behavior. However, for this project, in order to evaluate the performance of uMLIPs by computing segregation energy, the evaluation is still limited to comparing single-value segregation energies. The simulation of polycrystalline GB models using uMLIPs remains unavailable due to the high computational cost and the absence of reference data.

The best reference for evaluating the performance of uMLIPs comes from existing GB segregation studies based on DFT methods, as quantum mechanical methods are considered the most accurate and the reviewed evaluations typically use DFT data for comparison [30, 38, 39]. Systematically evaluating uMLIPs for simulating GB segregation is challenging. Firstly, the available DFT results are limited, and most of these DFT-based studies focus either on merely one or two GBs or the segregation of a single solute at Fe GBs. Consequently, the most relevant comparative studies currently available are the DFT-based GB segregation studies by Mai et al. [12], which examined four BCC Fe GBs and ten transition metals, and Ito et al. [43], which investigated nine FCC Fe GBs and nine transition metals. Additionally, since segregation energy calculations can also provide GB energy data, and both of the aforementioned studies include such data, GB energies will also serve as a reference for evaluation. Finally, we also followed the work of Ande and Sluiter [60] to calculate the solution enthalpy of BCC Fe and cementite with a single solute element substituted, to see if this can offer further insights for our study. The evaluation details of this project are summarized in the following table:

Table 3: Details of the evaluation.

Calculation Targets	**Substituted Elements (Solutes)**	**Relaxed Structures**
Segregation Energy	Ti, V, Cr, Mn, Co, Ni, Cu, Nb, Mo, W	4 BCC Fe GBs, 4×4×4 BCC Fe supercell
GB Energy	\	4 BCC Fe GBs
Segregation Energy	Ti, V, Cr, Mn, Co, Ni, Cu, Nb, Mo	9 FCC Fe GBs, 3×3×3 FCC Fe supercell
GB Energy	\	9 FCC Fe GBs
Solution Enthalpy	Al, Si, P, S, Ti, V, Cr, Mn, Co, Ni, Cu, Nb, Mo, W	3×3×3 BCC Fe supercell, cementite

Five pretrained uMLIP models were employed for the simulations: MACE-MP-0 (using the "large" model that includes more model parameters for higher accuracy), two versions of CHGNet, v0.2.0 and v0.3.0, M3GNet (MP-2021.2.8-PES), and SevenNet-0. CHGNet (v0.3.0) enhances the previous v0.2.0 version by increasing the AtomGraph cutoff, resolving discontinuities when no BondGraph is present, adding normalization layers, and improving accuracy in energy, force, and stress predictions. In addition to the uMLIPs used, several EIPs featuring Cu, the solute of primary interest, were selected from the Interatomic Potential Repository of the National Institute of Standards and Technology (NIST) database [61, 62] to serve as a secondary reference for the simulation (segregation energy and GB energy only). These EIPs include three EAM potentials and one MEAM potential. For the BCC Fe systems, the EAM potentials from Zhou et al. [63] and Bonny et al. [64] were employed, along with the MEAM potential from Lee et al. [65]. An EAM potential specifically fitted for studying high-entropy alloys, proposed by Deluigi et al. [66] and also the MEAM developed by Lee et al., were used for the FCC Fe system. The EAM potentials from Zhou and Bonny were used exclusively for BCC Fe systems because the pre-testing of this work showed that their predictions for FCC Fe systems were unreliable. Similarly, Deluigi's EAM potential was employed only for FCC Fe, as it was fitted for high-entropy alloys and exhibited significant errors when applied to BCC Fe systems.

4.1 Calculation

4.1.1 Grain Boundary Energy

In order to evaluate the performance of the uMLIPs in simulating GB segregation, GB energies and the segregation energies were calculated for comparison. The GB energy γ_{GB} can be represented as the excess energy associated with the presence of a GB compared to the energy of the bulk material. It quantifies the disruption of atomic structure and bonding that occurs at the interface between two grains. A higher value of γ_{GB} indicates a greater energy penalty for having the GB, which can influence the material's stability and mechanical properties. The GB energy was calculated as follows:

$$\gamma_{\mathrm{GB}} = \frac{E_{\mathrm{GB}} - E_{\mathrm{bulk}} \times \frac{N_{\mathrm{GB}}}{N_{\mathrm{bulk}}}}{2A} \tag{1}$$

E_{bulk} is the energy of the bulk cell in its pure form, containing N_{bulk} Fe atoms. Similarly, E_{GB} is the energy of the GB cell in its pure form, containing N_{GB} Fe atoms. Area A represents the area of the grain boundary, and the factor of two is due to the periodic boundary condition of the GB cells.

4.1.2 Segregation Energy

The segregation energy E_{seg} can be understood as the difference in energy between having the solute at the grain boundary compared to being in the bulk material. If E_{seg} is negative, it suggests that the solute prefers to segregate to the grain boundary, while a positive value indicates a preference for the solute to remain in the bulk. The segregation energy was calculated as following:

$$E_{\mathrm{seg}} = E_{\mathrm{GB}}(n-1)\mathrm{Fe,X} - E_{\mathrm{GB}} - E_{\mathrm{bulk}}(m-1)\mathrm{Fe,X} - E_{\mathrm{bulk}} \tag{2}$$

$E_{\mathrm{GB}}(n-1)\mathrm{Fe,X}$ represents the total energy of a GB structure with n atoms, one of which is the substitutional solute X. $E_{\mathrm{bulk}}(m-1)\mathrm{Fe,X}$ represents the total energy of a bulk cell with m atoms, one of which is the substitutional solute X. E_{bulk} and E_{GB} are the energy of the bulk cell and GB cell in its pure form, containing m Fe atoms in the bulk cell and n Fe atoms in the GB cell.

4.1.3 Solution Enthalpy

The calculation of solution enthalpy in BCC Fe system is done by finding the formation enthalpy difference of the solute-free bulk BCC Fe system and the system substituted by one solute element. The calculation of solution enthalpy for cementite is also the same.

Using cementite as example, the solution enthalpy of cementite where a Fe atom is substituted by an atom of element M can be calculated as:

$$\Delta H_{\mathrm{f}} = H_{\mathrm{f}}[\mathrm{Fe_{3q\text{-}1}MC_q}] - H_{\mathrm{f}}[\mathrm{Fe_{3q}C_q}] \tag{3}$$

where $H_{\mathrm{f}}[\mathrm{Fe_{3q\text{-}1}MC_q}]$ represents the formation enthalpy of the alloying-element-substituted-cementite system and $H_{\mathrm{f}}[\mathrm{Fe_{3q}C_q}]$ is the formation enthalpy of the solute-free cementite system.

4.2 Structures

All the bulk models were constructed using Python module Atomic Simulation Environment (ASE) [67]. A large supercell was built for the bulk to prevent solute-solute interaction. The bulk model for BCC Fe consists of a supercell containing 128 atoms, arranged in a 4×4×4 configuration of unit cells. While for the FCC Fe system, bulk model is a supercell containing 108 atoms in a 3×3×3 configuration of unit cells.

In the work of Mai et al., four STGB structures were used for GB segregation simulations as shown in Fig. 12. The initial GB models can simply be found in the appendix of Mai's paper [12]. And these GB structures were rescaled according to the lattice parameters of BCC Fe obtained using tested potentials as listed in Table A.1. While for Ito's work, the GB structures were obtained in the same way. The structures of the nine FCC Fe GBs (as shown in Fig. 9) were created according to the structures of stable [001] STGBs of Cu provided by Tschopp et al. in their dataset [43, 68]. These GB structures of Cu were rescaled according to the lattice parameters of FCC Fe of the potentials used in this work.

The 54 atoms BCC Fe supercell used for Ande and Sluiter's work was built via the ASE module in unit cell configuration of 3×3×3. While the ferromagnetic cementite structure was obtained from Materials Project database. The cementite structure is orthorhombic with space group *pnma*, and the lattice constants are $a = 4.49085$ Å, $b = 5.03018$ Å, and $c = 6.73931$ Å.

Table 4: Size of GB cells (number of atoms n_{GB}) and the cell parameters. Note that the cell parameters correspond to lattice parameters 2.832 Å (BCC Fe) and 3.540 Å (FCC Fe) used for DFT simulations. The cell parameters were rescaled according to the lattice parameter obtained for each potential as listed in Table A.1. The cell parameter c includes the thickness of the vacuum layer.

System	GB	n_{GB}	a (Å)	b (Å)	c (Å)
BCC Fe [110]	$\Sigma3(1\bar{1}1)$	72	4.005	6.937	44.144
	$\Sigma3(1\bar{1}2)$	48	4.005	4.905	41.620
	$\Sigma9(2\bar{2}1)$	68	4.005	6.332	50.973
	$\Sigma11(3\bar{3}2)$	42	4.005	4.696	50.973
FCC Fe [001]	$\Sigma5(210)$	100	7.92	7.08	30.00
	$\Sigma5(310)$	144	11.19	7.08	30.00
	$\Sigma13(320)$	160	12.76	7.08	30.00
	$\Sigma13(510)$	224	18.05	7.08	30.00
	$\Sigma17(410)$	180	14.60	7.08	30.00
	$\Sigma25(430)$	228	17.70	7.08	30.00
	$\Sigma29(520)$	240	19.06	7.08	30.00
	$\Sigma37(610)$	276	21.53	7.08	30.00
	$\Sigma41(540)$	288	22.67	7.08	30.00

4.3 Computational Details

For all the simulations, structural relaxation was conducted using either the Fast Inertial Relaxation Engine (FIRE) [69] or the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm [70], implemented in ASE. The relaxation convergence criterion was $f_{max} = 0.0001$ eV/Å, which means the relaxation will stop when the force on all individual atoms is less than f_{max}. BFGS was primarily utilized for minimization, while FIRE was employed in cases where minimization using BFGS did not converge. The maximum step for structural relaxation was set to 2000, any relaxation not converging within 2000 steps will be reported as unconverged in this work.

When calculating the segregation energies and solution enthalpies for the bulk systems, relaxations were performed on both the atomic positions and the cell parameters. While for the

GB system, only the atomic positions were relaxed. This approach is consistent with the relaxation methodologies employed by Mai et al. and Ito et al. in their first-principles calculations. Furthermore, the segregated solutes simulated for the FCC Fe GBs consist of nine transition metals, including Ti, V, Cr, Mn, Co, Ni, Cu, Nb, and Mo. The same solutes were also investigated by Mai et al., incorporating an additional transition metal W. It is noteworthy that the EIPs utilized as references included only the covered solutes. For example, for the EAM potential proposed by Bonny et al. [64] (will be referred as "FeMnNiCu_Bonny" in this work), the Fe atom was substituted by an atom of Mn, Ni or Cu when performing simulation using this potential.

The simulated substitutional sites for all GB structures are identical to those simulated by Mai et al. [12] and Ito et al. [43]. For the four BCC Fe GBs, the simulated substitutional sites are within the shaded region as shown in Fig. 12. And for the nine FCC Fe GBs, the simulated substitutional sites are indicated in Fig. 9, where all inequivalent substitutional sites within region 2.5 Å from the GB plane were studied. Note that for each GB, the segregation energies discussed in this work will be the lowest segregation energy among the results of all the simulated sites. For the cementite structure, the substituting sites of alloying elements are the 4c and 8d site of Fe atom as shown in Fig. 21.

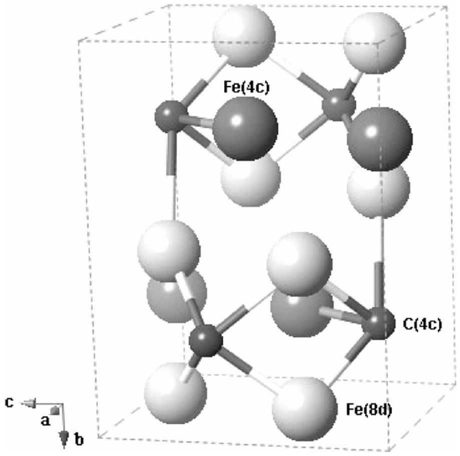


Figure 21: Schematic of a cementite cell with substituting sites labeled. The large sphere represents Fe atom and the small ones are C. The white Fe shpere is the 8d site and the dark sphere is the 4c site [71].

The GB models used for segregation energy calculations included a vacuum layer, whereas the models used to calculate GB energies had the vacuum layers removed. This treatment was adopted to maintain consistency with the work of Mai et al. [12]. Additionally, the relaxation performed on the GB cells without vacuum was conducted in the same way as in Mai's work, with both the atomic positions and the cell direction along the GB relaxed. The simulation of FCC Fe systems for calculating GB energies also follows this procedure, as Ito et al. did not specify how their GB energies were obtained.

5 Results

5.1 Grain Boundary Energy

Table 5: BCC Fe GB energies (J/m^2) for different potentials compared with DFT results from Mai et al.[12]. The GB misorientation axis is [110]. Missing data: results unconverged.

GB Energy [J/m^2]	**$\Sigma 3(1\bar{1}1)$**	**$\Sigma 3(1\bar{1}2)$**	**$\Sigma 9(2\bar{2}1)$**	**$\Sigma 11(3\bar{3}2)$**	**MAE [J/m^2]**
MACE-MP-0	1.16	0.22	1.87	0.75	0.36
CHGNet(v0.3.0)	0.72	0.10	1.23	0.53	0.67
CHGNet(v0.2.0)	0.90	0.27	1.35	0.82	0.48
M3GNet	0.01	0.20			0.91
SevenNet-0	0.04	0.22	0.48	0.59	0.98
FeMnNiCu_Bonny	1.31	0.27	2.01	1.04	0.28
FeTiCoNiCuMoW_Zhou	1.41	0.26	2.13	1.04	0.28
FeCu_Lee	1.16	0.24	2.12	0.86	0.39
DFT (Mai)	1.58	0.45	1.77	1.45	

The GB energies for all simulated GB structures are presented in Table 5 for BCC Fe and Table 6 for FCC Fe. There are convergence issues for the simulation of M3GNet and SevenNet-0. The relaxation of M3GNet on $\Sigma 9(2\bar{2}1)$ and $\Sigma 11(3\bar{3}2)$ BCC Fe GBs didn't converge, and none of the relaxation performed on FCC Fe GBs converged within 2000 steps. For SevenNet-0, one unconverged case was noted when relaxing the $\Sigma 37(610)$ FCC Fe GB. Although the relaxation of BCC Fe GB $\Sigma 3(1\bar{1}1)$ using M3GNet and SevenNet-0 converged, the calculated GB energies are only 0.01 and 0.04 J/m^2, respectively. Compared to the GB energy 1.58 J/m^2 obtained by Mai et al. via DFT calculation, the GB energies obtained by M3GNet and SevenNet-0 for this GB are unreasonably small.

For BCC Fe systems, the two EAM potentials achieve the smallest error in predicting GB energies, with an MAE of only 0.28 J/m^2. Among the uMLIPs, the model with the smallest error is MACE-MP-0, which has an MAE of 0.36 J/m^2, slightly smaller than the MAE of the MEAM potential at 0.39 J/m^2. The MAEs of the two CHGNet models are 0.67 J/m^2 for version 0.3.0 and 0.48 J/m^2 for version 0.2.0. Although the GB energy errors obtained by uMLIPs are larger compared to the EIPs, MACE-MP-0 and both CHGNet versions still demonstrate reasonable accuracy in predicting the GB energies of BCC Fe. For the GB energies of the nine FCC Fe GBs, all models show significant errors compared to the DFT results, with the EAM potential proposed by Deluigi having the smallest MAE at 0.51 J/m^2. In general, the simulated GB energies tend to be lower than the DFT results.

Table 6: FCC Fe GB energies (J/m^2) for different potentials compared with DFT results from Ito et al.[43], the GB misorientation axis is [001]. Missing data: result unconverged.

GB Energy [J/m^2]	**Σ37(610)**	**Σ13(510)**	**Σ17(410)**	**Σ5(310)**	**Σ29(520)**
MACE-MP-0	0.36	0.43	0.41	0.37	0.42
CHGNet(v0.3.0)	0.23	0.24	0.27	0.27	0.28
CHGNet(v0.2.0)	0.19	0.29	0.30	0.32	0.21
SevenNet-0		0.40	0.42	0.39	0.37
FeCrCoNiCu_Deluigi	0.49	0.53	0.70	0.48	0.33
FeCu_Lee	0.36	0.44	0.36	0.35	0.36
DFT (Ito)	1.07	1.13	1.20	1.24	1.29
	Σ5(210)	**Σ13(320)**	**Σ25(430)**	**Σ41(540)**	**MAE [J/m^2]**
MACE-MP-0	0.19	0.36	0.31	0.11	0.79
CHGNet(v0.3.0)	0.12	0.20	0.20	0.06	0.91
CHGNet(v0.2.0)	0.06	0.19	0.17	0.16	0.91
SevenNet-0	0.39	0.31	0.29	0.37	0.76
FeCrCoNiCu_Deluigi	0.52	0.86	0.82	0.75	0.51
FeCu_Lee	0.12	0.33	0.34	0.22	0.80
DFT (Ito)	1.33	1.08	0.91	0.81	

5.2 Segregation Energy

As this evaluation serves as a preparation for future simulation studies of Cu GB segregation, particular emphasis will be placed on the results related to Cu.

5.2.1 Overview

In the relaxation of the bulk system for both solute-free and solute-segregated, no instances of unconvergence were reported. However, when relaxing the solute-segregated BCC Fe GB system using M3GNet, unconverged cases were noted at several GB sites. While the relaxation of solute-segregated BCC Fe GB systems with other uMLIPs showed good convergence. For FCC Fe GB systems, there were significantly more cases of unconvergence. MACE-MP-0 and CHGNet(v0.3.0) showed no unconverged results, while CHGNet(v0.2.0) and SevenNet-0 exhibited several cases of unconvergence, which have been removed from the results. The convergence issues with M3GNet are particularly problematic, as almost all segregation at sites near the GB plane fails to converge, making result analysis nearly impossible. Therefore, the results from M3GNet cannot be presented due to these convergence issues.

Fig. 22 provides an overview of the lowest segregation energies predicted by uMLIPs compared to DFT results of all tested GBs and elements. For the BCC Fe system, MACE-MP-0 and both versions of CHGNet tend to underpredict the segregation tendency at BCC Fe GBs, as most data points lie above the dashed line. For the FCC Fe system, the accuracy of CHGNet(v0.3.0) is very high when the segregation tendency is small (from 0 to -0.3 eV). However, for data points with higher segregation tendencies at all FCC Fe GBs, CHGNet(v0.3.0) tends to underpredict

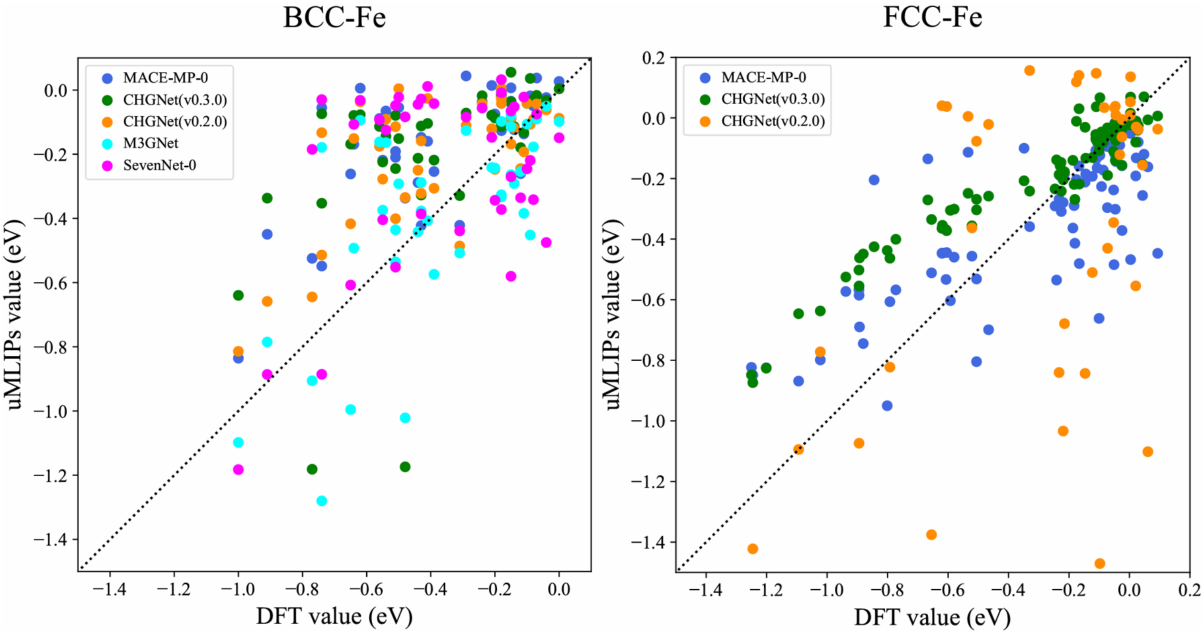


Figure 22: An Overview of the segregation energies results of all tested GBs and elements calculated by uMLIPs, plotted against DFT results from Mai et al.[12] and Ito et al.[43]. Note that for the FCC Fe include all solute elements and GBs. The results of SevenNet-0 are not presented on this graph due to large deviations.

the segregation tendency. The figure also shows that CHGNet(v0.2.0) performs poorly on the FCC Fe system.

All MAEs for both uMLIPs and EIPs are listed in Table 7. In general, the accuracy of uMLIPs is better than or at least as good as EIPs, including EAM and MEAM potentials. This is consistent with expectations, as uMLIPs are trained using datasets obtained from quantum mechanical methods. In the simulation of the BCC Fe system, the EAM potential proposed by Bonny et al. shows the largest error of 0.392 eV, which is almost twice that of other potentials. The uMLIP with the largest error is SevenNet-0, with an MAE of 0.257 eV. The accuracy of the uMLIPs follows the sequence: M3GNet > CHGNet(v0.2.0) > MACE-MP-0 > CHGNet(v0.3.0) > SevenNet-0. It's worth mentioning that although CHGNet(v0.3.0) was improved from version 0.2.0, version 0.2.0 outperforms version 0.3.0 in BCC Fe systems. The accuracy of the EAM potential from Zhou et al. and the MEAM potential from Lee et al. is high, comparable to that of uMLIPs.

The accuracy of different potentials in predicting GB segregation in FCC Fe GBs varies compared to BCC Fe GBs. The uMLIPs, particularly CHGNet(v0.3.0) and MACE-MP-0, demonstrate the highest accuracy, with MAE around 0.15 eV, surpassing selected traditional potentials, such as the EAM potential from Deluigi et al. (0.258 eV) and the MEAM potential from Lee et al. (0.307 eV). While M3GNet demonstrates the highest accuracy for BCC Fe structures, convergence issues limit its applicability in simulating the FCC Fe system. Similar convergence problems were noted with SevenNet-0, where seven data points were excluded due to unconvergence. CHGNet(v0.2.0) and SevenNet-0 exhibit significantly lower accuracy in FCC Fe systems, with the latter showing large deviations from the DFT results, reaching an MAE of 2.781 eV. Although CHGNet(v0.2.0) maintains accuracy for some solutes and GBs, many results show significant errors, which can also be seen in Fig. 22. Overall, the accuracy of these potentials in predicting segregation in FCC Fe GBs follows the sequence: CHGNet(v0.3.0) > MACE-MP-

Table 7: Overall MAE of segregation energy results compared to DFT results. Note that for EIPs, the MAE were calculated with available chemicals only. For example, the MAE of FeCu_Lee is the MAE of Cu only.

BCC Fe		**FCC Fe**	
Potential	**MAE (eV)**	**Potential**	**MAE (eV)**
MACE-MP-0	0.221	MACE-MP-0	0.169
CHGNet(v0.3.0)	0.245	CHGNet(v0.3.0)	0.150
CHGNet(v0.2.0)	0.193	CHGNet(v0.2.0)	0.786
M3GNet	0.176	SevenNet-0	2.781
SevenNet-0	0.257	FeCrCoNiCu_Deluigi	0.258
FeMnNiCu_Bonny	0.392	FeCu_Lee	0.307
FeTiCoNiCuMoW_Zhou	0.207		
FeCu_Lee	0.256		

0 > CHGNet(v0.2.0) > SevenNet-0. Among the tested uMLIPs, only CHGNet (v0.3.0) and MACE-MP-0 provide reliable predictions for FCC Fe systems.

5.2.2 Segregation Energy of BCC Fe GBs

Fig. 23 presents the lowest segregation energies of four BCC Fe GBs for different solutes. The variation in segregation energies across different solutes shows a similar pattern among the uMLIPs, which means that uMLIPs can capture the general segregation tendency of tested solute elements. It is evident that these uMLIPs exhibit relatively higher accuracy in predicting the segregation tendencies of V, Cr, and Co. However, the predictions for Mn, Ni, and Cu exhibit lower accuracy, as DFT results from Mai et al. show a much stronger segregation tendency for these solutes. This discrepancy is a significant contributor to the underprediction noted in the Overview subsection 5.2.1. While the segregation energies of Mn, Ni, and Cu predicted by uMLIPs are similar, including those from the three EIPs, the data points for these potentials cluster closely together in Fig. 23, except for M3GNet in the two high-angle GBs, $\Sigma3(1\bar{1}1)$ and $\Sigma9(2\bar{2}1)$, and SevenNet-0 in the twin GB, $\Sigma3(1\bar{1}2)$. In addition to Mn, Ni, and Cu, another major contribution of error is the predictions for Ti and Nb. Unlike Mn, Ni, and Cu, the predicted segregation energies for Ti and Nb show much greater variability across different potentials.

For the prediction of Cu segregation, the results show that M3GNet has the best performance in Cu segregation simulation with an MAE of 0.242 eV (MAE of the segregation energy of Cu only for the four studied GBs compared to DFT results), while the MEAM potential developed by Lee et al. specialzed in Cu has an MAE of 0.255 eV. The MAEs of Cu segregation of other uMLIPs and the two EAM potentials are around 0.4 eV or even larger. In addition to Cu, M3GNet also exhibits the best performance in simulating Ni segregation with an MAE (Ni only) down to 0.153 eV, while the other uMLIPs and the two EAM potentials have an MAE nearly twice as large. Despite the outperformance of M3GNet in simulating Ni and Cu segregation at BCC Fe GBs compared to other uMLIPs, M3GNet shows the worst performance in simulating Ti segregation, which can be seen clearly on Fig. 23. M3GNet always predicts a much larger segregation tendency of Ti which is nearly twice of the DFT results. Similar to M3GNet on

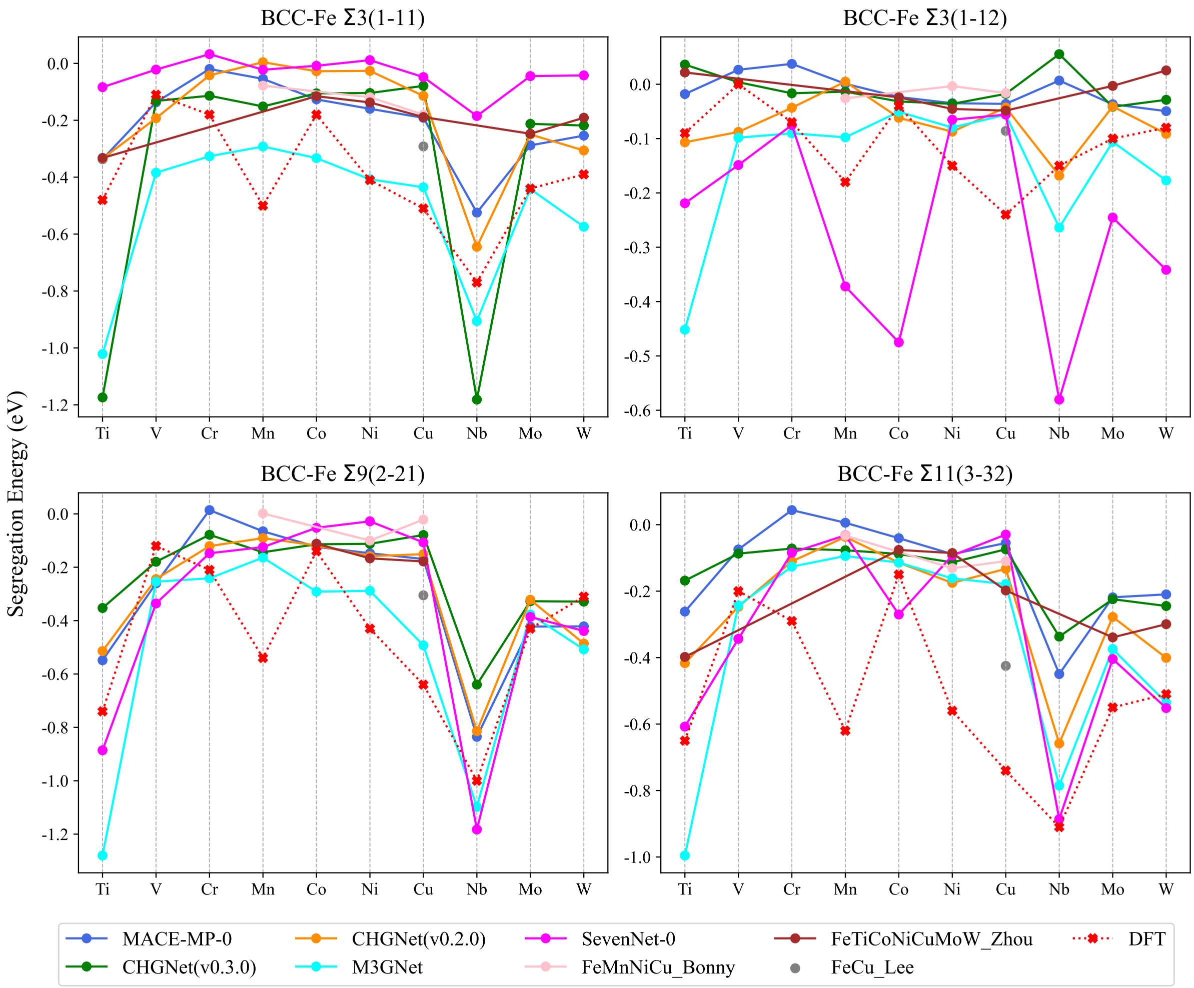


Figure 23: Lowest segregation energies (eV) of BCC Fe systems plotted against solute elements. GB line direction is [110].

Ti, CHGNet(v0.2.0) also exhibits large errors when simulating Ti and Nb. For STGB $\Sigma3(1\bar{1}1)$, CHGNet(v0.3.0) predicted the most negative segregation energies when solutes are Ti or Nb. While for the other GBs, the segregation energies of Ti and Nb predicted by CHGNet(v0.3.0) become the most positive.

5.2.3 Segregation Energy of FCC Fe GBs

For FCC Fe results, MACE-MP-0 and CHGNet(v0.3.0) exhibit higher accuracy compared to the BCC Fe results, with a MAE of 0.17 eV and 0.15 eV, respectively. While the results of Ni and Cu segregation in BCC Fe GBs show large errors, the predictions for Ni and Cu by MACE-MP-0 and CHGNet(v0.3.0) in FCC Fe systems are the most accurate. The MAE for Ni and Cu segregation with MACE-MP-0 are 0.08 eV and 0.09 eV, respectively, while for CHGNet(v0.3.0), the values are as low as 0.04 eV for Ni and 0.05 eV for Cu. Based on the DFT results from Ito et al. in Table A.11, it appears that the smaller MAEs of MACE-MP-0 and CHGNet(v0.3.0) may be attributed to the smaller variation in segregation energies across different GBs. However, the higher MAEs for Deluigi's EAM potential (0.20 eV for Ni and 0.31 eV for Cu) and Lee's MEAM potential (0.31 eV) confirm that MACE-MP-0 and CHGNet(v0.3.0) outperform EIPs in FCC Fe systems. CHGNet(v0.3.0) also exhibits small errors in the prediction of V, Cr, Mn,

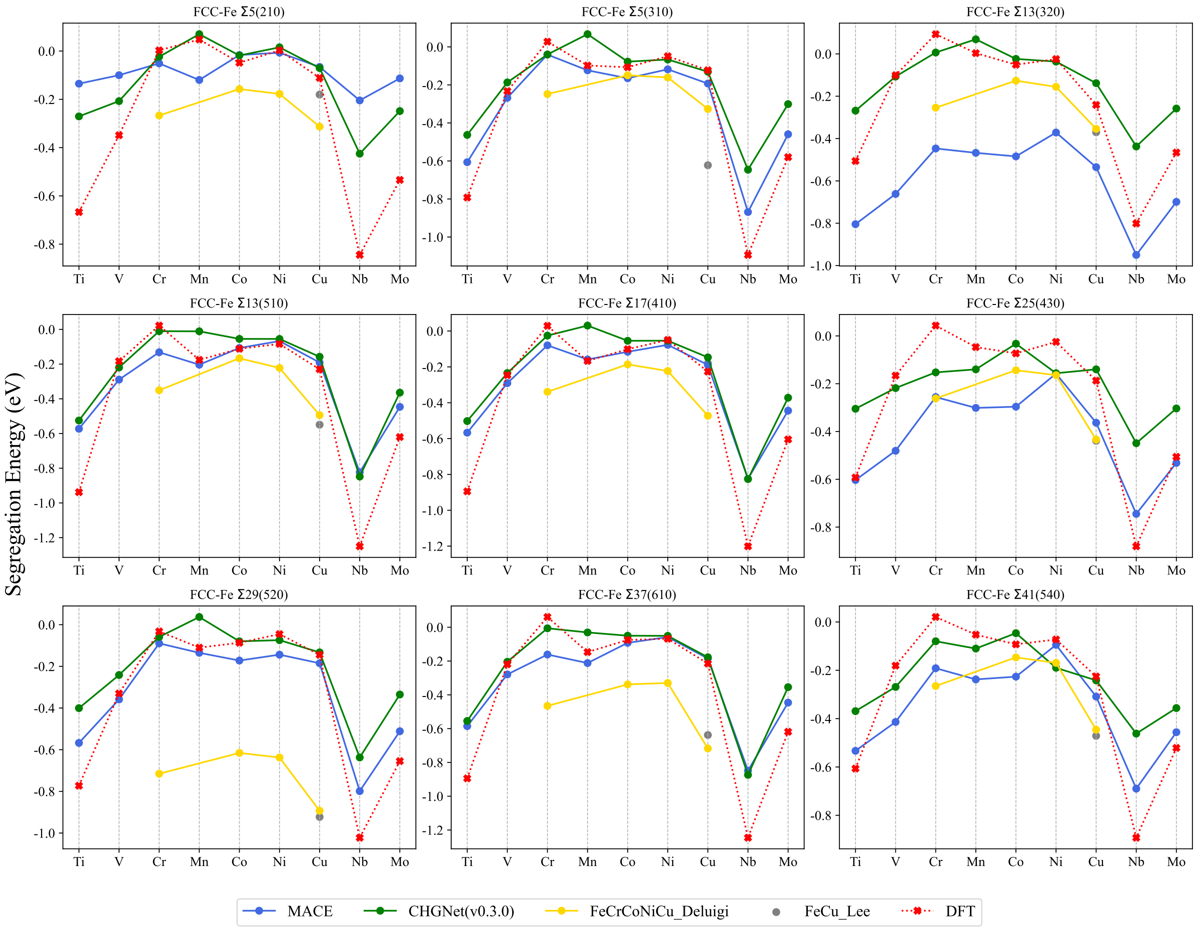


Figure 24: Lowest segregation energies (eV) of FCC Fe systems plotted against solute elements. Results of SevenNet-0 are not shown in order to avoid the stretch of the plots due to large errors.

and Co, with MAEs around 0.1 eV or smaller. However, CHGNet(v0.3.0) shows larger errors in the prediction of Ti, Nb, and Mo, with Ti and Nb errors similar to those observed in BCC Fe, where the variation is much larger. The primary contributors to the errors for MACE-MP-0 are also Ti and Nb, while predictions for other elements generally yield MAEs between 0.1 and 0.2 eV. It can be observed that the segregation tendencies of Ti and Nb predicted by DFT are generally much higher than those predicted by MACE-MP-0 and CHGNet(v0.3.0). This underprediction can also be observed in the BCC Fe ones.

While the results from MACE-MP-0 and CHGNet(v0.3.0) demonstrate smaller errors in FCC Fe systems, the accuracy of other uMLIPs remains inadequate. Despite the convergence issues associated with M3GNet, the converged results from CHGNet(v0.2.0) and SevenNet exhibit significant errors. As shown in Appendix Table A.15, the predictions of Ti, V, Cr, Co, Ni, and Cu for GB Σ13(320) are relatively reasonable, whereas the majority of the other predictions exhibit absolute errors exceeding 2 eV. As for M3GNet(v0.2.0), there are a few predictions with smaller errors, which are illustrated in Fig. 24. These include the segregation energies for Cr, Mn, Co, and Ni in FCC Fe GBs such as Σ5(210), Σ13(320), Σ13(510), Σ17(410), Σ25(430), and Σ29(520). However, the remaining predictions from CHGNet(v0.2.0) generally exhibit absolute errors around 1 eV.

From the results of FCC Fe GB segregation, no distinct dependence on the GB structures was observed. The two EIPs showed the largest errors when simulating the Σ29(520) GB.

CHGNet(v0.3.0) demonstrated consistent performance across different GBs, with only small variations in error, ranging from 0.14 eV to 0.16 eV (the MAE for each GB). In contrast, MACE-MP-0 exhibited errors ranging from 0.1 eV to 0.17 eV for most GBs, except for the $\Sigma5(210)$ GB, which had an MAE of approximately 0.24 eV, and the $\Sigma13(320)$ GB, which showed the largest MAE at around 0.37 eV. The discrepancies of MACE-MP-0 on these two GBs are clearly visible in Fig. 24: while the segregation tendency predicted by DFT decreases, MACE-MP-0 does not capture this reduction.

5.3 Solution Enthalpies of Alloying-Element-Substituted BCC Fe and Cementite

Table 8: Solution enthalpies (unit: eV) of the alloying-element-substituted BCC Fe supercell containing 54 atoms. DFT data from Ande and Sluiter [60].

Potential	Al	Si	P	S	Ti	V	Cr
MACE-MP-0	3.667	1.559	2.159	4.835	-0.595	-1.513	-1.303
CHGNet(v0.3.0)	3.725	1.833	1.562	3.657	-0.714	-1.331	-1.167
CHGNet(v0.2.0)	3.728	2.017	1.408	4.402	-0.577	-1.328	-1.113
M3GNet	4.385	2.175	1.189	4.438	0.798	-1.090	-0.922
SevenNet-0	3.944	1.782	1.825	4.333	-0.360	-1.264	-1.443
DFT	3.751	1.661	1.842	4.314	-0.394	-1.453	-1.412
Potential	Mn	Co	Ni	Cu	Nb	Mo	W
MACE-MP-0	-0.572	1.120	2.599	4.556	-2.044	-2.317	-4.559
CHGNet(v0.3.0)	-0.687	1.319	2.594	4.625	-2.118	-2.414	-4.595
CHGNet(v0.2.0)	-0.648	1.355	1.769	4.661	-2.019	-2.440	-4.793
M3GNet	-0.515	1.217	2.699	4.798	-1.303	-1.933	-4.245
SevenNet-0	-0.741	0.950	2.341	4.222	-1.744	-2.222	-4.034
DFT	-0.522	1.069	2.844	5.217	-1.992	-2.575	-4.796

The calculated solution enthalpies of the alloying-element-substituted BCC Fe supercell are shown in Table 8 and the MAEs are in Table 9. The general performance of the uMLIPs in predicting solution enthalpy follows the sequence that MACE-MP-0 > CHGNet(v0.2.0) > CHGNet(v0.3.0) > SevenNet-0 > M3GNet. According to the MAE of CHGNet(v0.3.0), M3GNet and MACE-MP-0 in calculating formation energy reported by Yu et al., the performance of these three uMLIPs follow the sequence that MACE-MP-0 > CHGNet(v0.3.0) > M3GNet, which aligns well with the results in Table 9. Furthermore, both results also show the small difference in MAE between MACE-MP-0 and CHGNet(v0.3.0) and the much larger error of M3GNet. It's worth mentioning that although M3GNet exhibits the best performance in calculating the segregation energy at BCC Fe GBs, the solution enthalpy calculated are with the largest errors among the uMLIPs instead.

Similar to the results for BCC Fe GB segregation energy, the solution enthalpy of Cu shows the largest error among all the alloying elements. For other alloying elements, uMLIPs consistently predict solution enthalpies with errors around or below 0.1 eV. However, for the solution enthalpy of Cu, the result with the smallest error is predicted by M3GNet, with an error of approximately 0.42 eV. The solution enthalpy of Cu predicted by DFT is significantly larger than those predicted by uMLIPs. M3GNet also exhibits a large error in calculating the solution enthalpy of Ti. The substitution process predicted by M3GNet is endothermic (positive solution enthalpy), whereas all other predictions and DFT results show negative solution enthalpies. This is the only case where a difference in thermal behavior compared to DFT results was observed.

Table 9: MAEs (eV) of solution enthalpies of alloying-element-substituted BCC Fe, uMLIPs vs. DFT. The MAE of specific solute, e.g., Al, represents the MAE of the Al results simulated using all potentials compared to DFT results. The MAE of specific potential, e.g., MACE-MP-0, represents the MAE of all solutes using MACE-MP-0 compared to DFT results.

Element	**MAE**	**Element**	**MAE**	**Element**	**MAE**	**Model**	**MAE**
Al	0.192	**Ti**	0.386	**Ni**	0.244	**MACE-MP-0**	0.218
Si	0.253	**V**	0.172	**Cu**	0.645	**CHGNet (v0.3.0)**	0.255
P	0.279	**Cr**	0.235	**Nb**	0.229	**CHGNet (v0.2.0)**	0.224
S	0.282	**Mn**	0.113	**Mo**	0.333	**M3GNet**	0.448
		Co	0.171	**W**	0.432	**SevenNet-0**	0.272

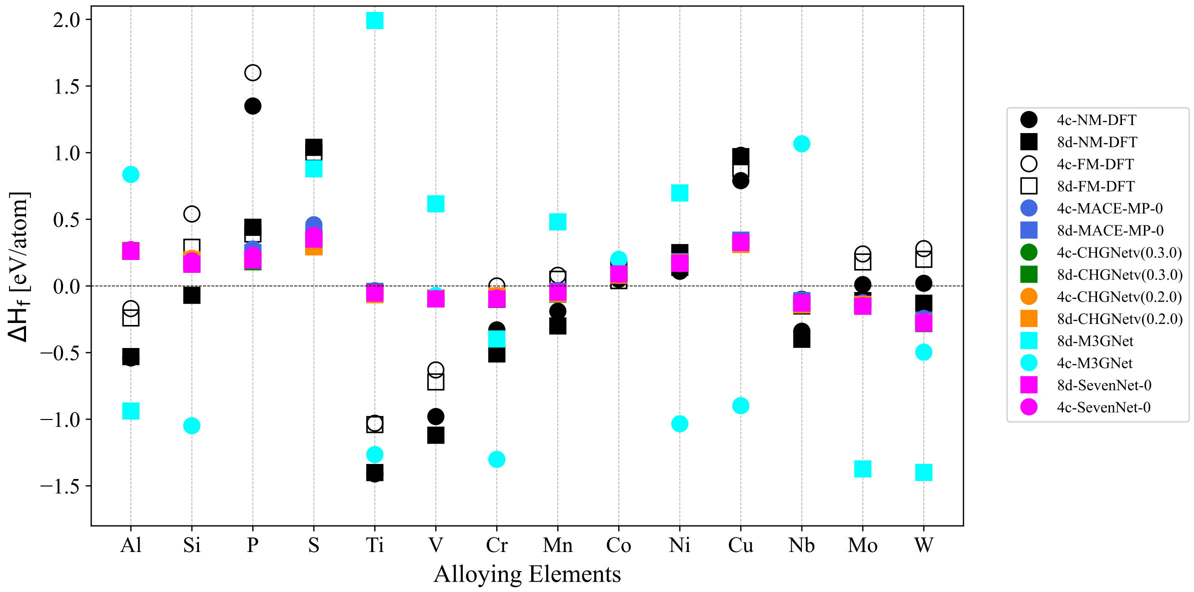


Figure 25: Solution enthalpies of alloying-element-substituted cementite compared with DFT data from Ande and Sluiter [60].

The solution enthalpies of alloying-element-substituted cementite are presented in Fig. 25. Since the solution enthalpies provided by DFT are distinguished as NM (non-ferromagnetic) and FM (ferromagnetic), the solution enthalpies calculated using uMLIPs were compared with both NM and FM DFT data. It can be observed that, with the exception of M3GNet, the other four uMLIPs predict almost the same solution enthalpies for all tested alloying elements. In contrast, the solution enthalpies predicted by M3GNet exhibit significant errors. This can be further confirmed by calculating the MAEs as listed in Table 10. The MAEs for MACE-MP-0, both versions of CHGNet, and SevenNet-0 are quite similar, while the MAEs for M3GNet are significantly higher—three times those of the other uMLIPs at the 4c site and almost ten times higher at the 8d site.

Table 10: MAEs of calculated solution enthalpies for each uMLIP compared to both the FM and NM DFT data.

MAE (eV)	**MACE-MP-0**		**CHGNetv3**		**CHGNetv2**		**M3GNet**		**SevenNet-0**	
Site	**4c**	**8d**	**4c**	**8d**	**4c**	**8d**	**4c**	**8d**	**4c**	**8d**
FM	0.55	0.32	0.58	0.34	0.57	0.33	1.77	3.10	0.57	0.33
NM	0.58	0.44	0.59	0.45	0.59	0.45	1.80	3.12	0.59	0.44

6 Discussion

6.1 Convergence Analysis

In the Results section, the convergence issues were briefly discussed. Structural relaxations performed on the bulk systems, including cementite, successfully converged across all uMLIPs. However, convergence issues were encountered during the relaxation of GB systems, particularly when using M3GNet, as detailed in Table 11. Notably, seven out of nine GBs failed to converge during the relaxation of pure FCC Fe GB systems using M3GNet. This is why results from M3GNet for FCC Fe are not included in this work. Among the other uMLIPs, MACE-MP-0 and CHGNet(v0.3.0) demonstrated the most robust relaxation performance, with no reported convergence failures. CHGNet(v0.2.0) had only one unconverged case when relaxing the FCC Fe $\Sigma37(610)$ GB system segregated by Co. This unconverged case might be resolved by increasing the maximum relaxation steps. For SevenNet-0, all unconverged cases were associated with the $\Sigma41(540)$ FCC Fe GB and Cu-segregated GBs (see Table A.15). It is also possible that both Cu and the $\Sigma41(540)$ GB (with 216 atoms) require more relaxation steps to achieve convergence.

Table 11: Convergence analysis on structrual relaxation of the GB structures substituted by tested solutes. For BCC Fe, there are 4 GBs and 10 solute elements, 40 cases for each uMLIP. For FCC Fe, there are 9 GBs and 9 solute elements, 81 cases for each uMLIP.

# of Convergence	**BCC Fe**	**Rate**	**FCC Fe**	**Rate**
MACE-MP-0	40	100%	81	100%
CHGNet(v0.3.0)	40	100%	81	100%
CHGNet(v0.2.0)	40	100%	80	99%
M3GNet	40	100%	<18	<22%
SevenNet-0	40	100%	74	91%

Since the structural relaxation of GB systems in calculating segregation energies was performed by relaxing the atomic position only, the significant increase in unconverged cases when relaxing both the atomic position and the cell cannot be assessed quantitatively in this work. However, the increase of possibility of unconvergence can be observed in this work when calculating the GB energies (relaxation was performed on atomic position and the direction along the GB of the cell parameter). For instance, when calculating the segregation energies at BCC Fe GBs, all relaxation using M3GNet converged, while unconvergence cases were found for the relaxation performed on the two BCC Fe GB systems ($\Sigma9(2\bar{2}1)$ and $\Sigma11(3\bar{3}2)$) when calculating GB energies.

The convergence problems reported in this work align with the analysis performed by Yu et al. [39], as mentioned in the Literature Review section 2.1.3, which concluded that MACE-MP-0 and CHGNet (v0.3.0) can perform relaxations more robustly. However, according to the assessment by Gonzales et al. [40], the robustness of CHGNet (v0.3.0) remains questionable, and the origin of the convergence issues requires further investigation. One assumption based on this work is that the convergence issues of uMLIPs might be attributed to a lack of training data for structures at intermediate states. Although the training datasets of the tested uMLIPs include data of intermediate structures [28, 30], fine-tuning the models with additional data of such structures might improve their convergence behavior.

6.2 Accuracy Analysis

Although the primary focus of this thesis is on assessing the performance of uMLIPs in simulating GB segregation, GB energies and solution enthalpies were also calculated as additional reference points. For BCC Fe systems, the GB energies calculated using selected EIPs are close to those obtained through DFT methods. In contrast, most of the GB energy results calculated by uMLIPs show relatively larger errors, with only the BCC Fe GB energies calculated by MACE-MP-0 with reasonable errors compared to EIPs. Since both uMLIPs and EIPs exhibit significant errors in predicting GB energies for FCC Fe systems. This suggests that the discrepancies may not be due to the inherent performance limitations of uMLIPs in FCC Fe GBs but may instead be attributed to issues with the simulated structures, as Ito et al. did not provide sufficient details about the GB energy calculation. Consequently, the GB energies calculated in this work may not provide a reliable basis for evaluating the performance of uMLIPs.

It has been reviewed (section 2.1.2) that the performance of uMLIPs highly depends on the quality of the training dataset, and increasing the size of the training dataset will improve the performance of uMLIPs. Additionally, the out-of-distribution performance of the tested pre-trained uMLIP models can be reduced due to a systematic underprediction as reviewed in section 2.1.3 [42]. Furthermore, the training datasets for MACE-MP-0, CHGNet, and SevenNet-0 are identical, with the training dataset of M3GNet being a subset of the former. Since the cementite structure was obtained from Materials Project database, this structures is included in the training dataset of the uMLIPs except for M3GNet. This is clearly shown in the results (Fig. 25) of calculating the solution enthalpies of alloying-element-substituted cementite. While the other four uMLIPs give almost the same solution enthalpy for each substituted alloying elements, the solution enthalpies predicted by M3GNet show larger deviations and errors.

In terms of performance on GB segregation simulation, M3GNet exhibited the smallest error in predicting the segregation energies for BCC Fe GBs among the uMLIPs evaluated. However, due to the relatively smaller size of M3GNet's training dataset and convergence issues, its performance requires further evaluation. While CHGNet(v0.2.0) outperforms CHGNet(v0.3.0) in BCC Fe systems, large errors and several unconverged cases were observed in FCC Fe systems. Overall, the performance of CHGNet(v0.3.0) is still superior to CHGNet(v0.2.0) across both systems, as it represents an improved version. According to the assessment of Matbench Discovery [30], SevenNet-0 is expected to perform best among the tested uMLIPs. However, the segregation energies predicted by SevenNet-0 for FCC Fe GB systems are found to be unreasonable. Further evaluation of SevenNet-0 is necessary, since currently no additional studies have assessed its performance. Since only four BCC Fe GBs were simulated in this study, the current evaluation is insufficient to definitively determine the performance of M3GNet and SevenNet-0 on predicting the segregation behavior, particularly given their performance in FCC Fe GB systems.

It is difficult to summarize the performance of each uMLIP in predicting the GB segregation of specific solute elements, as discrepancies are frequently observed across different uMLIPs. In the Results section 5.2.1, it was noted that MACE-MP-0 and both CHGNet models tend to underpredict the segregation tendency at BCC Fe GBs, particularly when the segregation tendency predicted by DFT is high. The elements with high segregation tendencies at BCC Fe GBs—Ti, Mn, Ni, Cu, and Nb, as shown in Table 12, generally exhibit underpredictions of over 0.3 eV (with few exceptions). For the GB $\Sigma 11(3\bar{3}2)$, the error due to underprediction increases significantly. A similar underprediction of segregation tendencies was observed in the results of all tested EIPs, as well as for certain solutes predicted by M3GNet and SevenNet-0, particularly Cu. However, this does not imply that the uMLIPs' predictions for these solute elements are inherently inaccurate, as some of their accuracy can be high when predicting solute segregation at FCC Fe GBs. For FCC Fe GBs, this underprediction is observed when predicting the segregation behavior of Ti, Nb, and Mo using MACE-MP-0 and CHGNet(v0.3.0). A common characteristic among these elements is their high segregation tendency in this particular structure. It is hypothesized that this underprediction phenomenon, when the segregation tendency predicted by DFT is high, may be the same phenomenon reported by Deng et al. [42]. However, since this phenomenon is also observed with the results of EIPs, further evaluation of its underlying cause is required. Based on these observations, it is suggested that fine-tuning the model, especially when simulating solutes with high segregation tendencies at specific GBs is necessary to achieve greater accuracy.

Table 12: Error ($E_{\text{seg}}^{\text{uMLIP}}$ - $E_{\text{seg}}^{\text{DFT}}$, unit: eV) of MACE-MP-0 and both CHGNet models for various elements, demonstrating the general underprediction of segregation tendency of listed elements. The results of twin-type GB $\Sigma 3(1\bar{1}2)$ were not compared since the general segregation tendencies of all solutes at this GB are small. The plotted data of this table can be found in the appendix Figure A.1.

Potential	**GB**	**Ti**	**Mn**	**Ni**	**Cu**	**Nb**
MACE-MP-0	$\Sigma 3(1\bar{1}1)$	0.14	0.45	0.25	0.32	0.25
	$\Sigma 9(2\bar{2}1)$	0.19	0.47	0.28	0.47	0.16
	$\Sigma 11(3\bar{3}2)$	0.39	0.63	0.47	0.68	0.46
CHGNet(v0.3.0)	$\Sigma 3(1\bar{1}1)$	-0.69	0.35	0.31	0.43	-0.41
	$\Sigma 9(2\bar{2}1)$	0.39	0.40	0.32	0.56	0.36
	$\Sigma 11(3\bar{3}2)$	0.48	0.54	0.45	0.67	0.57
CHGNet(v0.2.0)	$\Sigma 3(1\bar{1}1)$	0.15	0.50	0.38	0.40	0.13
	$\Sigma 9(2\bar{2}1)$	0.23	0.45	0.27	0.49	0.19
	$\Sigma 11(3\bar{3}2)$	0.23	0.58	0.38	0.61	0.25

Based on the results of BCC Fe GBs, it can be concluded that as the misorientation (Σ) of the GB increases, a general trend of increasing errors is observed for all potentials since greater structural complexity increases the difficulty of simulation. A notable example is the significant increase in errors at the $\Sigma11(3\bar{3}2)$ GB, as shown in Table 12. However, with the exception of SevenNet-0, the simulation results for SevenNet-0 at the two $\Sigma3$ GBs exhibit higher errors, which are the primary contributors to the overall MAE of SevenNet-0. These discrepancies are particularly evident for FCC Fe GBs, where the variation in segregation energies is smaller. The prediction of GB segregation is an out-of-distribution task for uMLIPs, and they are expected to have a weaker physical foundation. However, the results for the BCC Fe $\Sigma3(1\bar{1}2)$ GB demonstrate that uMLIPs are capable of capturing the characteristics of this twin-type GB, similar to EIPs and quantum mechanical methods, by predicting lower segregation energies compared to other GBs.

6.3 Summary of Performance Evaluation

Table 13: Ranked uMLIPs and EIPs according to the MAE of predicted segregation energy.

MAE (eV)	**Good (0.15-0.20)**	**Moderate (0.2-0.26)**	**Bad (0.39)**
E_{seg} **BCC Fe**	M3GNet CHGNet(v0.2.0)	MACE-MP-0 CHGNet(v0.3.0) SevenNet-0 FeTiCoNiCuMoW_Zhou FeCu_Lee	FeMnNiCu_Bonny
MAE (eV)	**Good (0.15-0.20)**	**Moderate (0.25-0.30)**	**Bad (0.79 & 2.78)**
E_{seg} **FCC Fe**	CHGNet(v0.3.0) MACE-MP-0	FeCrCoNiCu_Deluigi FeCu_Lee	CHGNet(v0.2.0) SevenNet-0

To better summarize the general performance of uMLIPs in predicting GB segregation, we categorized the uMLIPs, along with the EIPs, into three tiers based on their accuracy, as shown in Table 13. In BCC Fe systems, all uMLIPs exceed or at least meet the accuracy of EIPs, with models like M3GNet and CHGNet(v0.2.0) performing particularly well. In FCC Fe systems, MACE-MP-0 and CHGNet(v0.3.0) outperform the EIPs, while the other uMLIPs either exhibit inadequate accuracy or experience severe convergence issues.

Although M3GNet, CHGNet(v0.2.0) and SevenNet-0 exhibit good performance in the prediction of GB segregation at BCC Fe GBs. The reason why they are not recommended for future simulaions has been discussed (see 6.2). Therefore, the uMLIPs can finish the GB segregation simulation tasks robustly with relatively high accuracy are MACE-MP-0 and CHGNet(v0.3.0) in general. However, for predicting solute segregation with high segregation tendency at BCC Fe GBs, the accuracy of these two pre-trained models might be restricted. For example, considering the simulation of single segregation of Cu only at BCC Fe GBs, the MAE of MACE-MP-0 and CHGNet(v0.3.0) can achieve over 0.5 eV, while the MAE of using the Fe-Cu MEAM potential is half of that. Since the underprediction of the mentioned solutes with high segregation tendencies is the major contribution to the errors of MACE-MP-0 and CHGNet(v0.3.0), the prediction of other elements can reach near quantum-level accuracy.

Overall, CHGNet(v0.3.0) and MACE-MP-0 emerged as the two best-performing uMLIPs in this evaluation. Both models demonstrate strong performance in BCC Fe and FCC Fe systems without convergence issues. However, as CHGNet(v0.3.0) can occasionally be unstable (e.g., predicting negative elastic constants [72] and in the benchmarking of Gonzales et al. [40]), MACE-MP-0 is the preferred choice. While the performance of uMLIPs like MACE-MP-0 generally surpasses that of the tested EIPs, their accuracy can be limited by the underprediction phenomenon noted earlier. Another limitation of uMLIPs is their higher computational cost, which constrains their application to larger structures, such as polycrystalline GB models, compared to EIPs. It is noteworthy that EIPs can still outperform uMLIPs in certain cases.

7 Conclusion

This thesis study evaluates the performance of uMLIPs in simulating solute segregation at Fe GBs. The performance of the tested uMLIPs was compared with referenced DFT-based studies. It can be concluded that uMLIPs, specifically MACE-MP-0 and CHGNet(v0.3.0), are capable of replacing EIPs in GB segregation studies (with MACE-MP-0 more recommended). However, this study also highlighted the challenges in simulating the segregation of highly segregating solutes, where significant errors persisted for both MACE-MP-0 and CHGNet. This limitation could be problematic for future research aimed at using simulations to explore alloying elements that inhibit Cu segregation at Fe GBs, given that Cu is highly segregating. Another limitation of uMLIPs identified in this work is convergence issues. The issues identified in this work may be mitigated through further fine-tuning of uMLIP models, such as by expanding training datasets. Nevertheless, this thesis provides an important benchmark for using uMLIPs in simulating solute segregation at Fe GBs.

Acknowledgement

This work was supported through the computational resources and staff contributions provided for the TulipX high-performance computing facilities at SKF Research and Technology Development, Houten, The Netherlands.

References

[1] Stefan Pauliuk, Rachel L. Milford, Daniel B. Müller, and Julian M. Allwood. The steel scrap age. *Environmental Science Technology*, 47(7):3448–3454, 3 2013.

[2] European Steel Association (EUROFER). EUROFER Statistics, https://www.eurofer.eu/statistics.

[3] Kenichi Nakajima, Osamu Takeda, Takahiro Miki, Kazuyo Matsubae, and Tetsuya Nagasaka. Thermodynamic analysis for the controllability of elements in the recycling process of metals. *Environmental Science Technology*, 45(11):4929–4936, 5 2011.

[4] Olivier Rod, Christian Becker, and Margareta Nylén. Opportunities and dangers of using residual elements in steels: a literature survey. *Jernorets Forskning*, 1 2006.

[5] Katrin E. Daehn, André Cabrera Serrenho, and Julian M. Allwood. How will copper contamination constrain future global steel recycling? *Environmental Science Technology*, 51(11):6599–6606, 5 2017.

[6] Katrin E. Daehn, André Cabrera Serrenho, and Julian Allwood. Finding the Most Efficient Way to Remove Residual Copper from Steel Scrap. *Metallurgical and Materials Transactions B*, 50(3):1225–1240, 2 2019.

[7] Gadadhar Sahoo, Balbir Singh, and Atul Saxena. Effect of strain rate, soaking time and alloying elements on hot ductility and hot shortness of low alloy steels. *Materials Science and Engineering A*, 718:292–300, 3 2018.

[8] Seok-Jong Seo, Kentaro Asakura, and Koji Shibata. Effects of CU and other TrAmp elements on steel properties. Evaluation of susceptibility to surface hot shortness in CU-containing steels by tensile test. *ISIJ International*, 37(3):232–239, 1 1997.

[9] Motohiro Yuasa and Mamoru Mabuchi. Effects of segregated Cu on an Fe grain boundary by first-principles tensile tests. *Journal of Physics Condensed Matter*, 22(50):505705, 11 2010.

[10] Pavel Lejcek. *Grain boundary segregation in metals.* 1 2010.

[11] Liang Zhang, Zhihui Zhang, Xuan Zhang, and Xiaoxu Huang. Computational simulation of grain boundary segregation of solute atoms in nanocrystalline metals. *Journal of Materials Research and Technology*, 21:161–185, 11 2022.

[12] Han Lin Mai, Xiang-Yuan Cui, Daniel Scheiber, Lorenz Romaner, and Simon P. Ringer. The segregation of transition metals to iron grain boundaries and their effects on cohesion. *Acta Materialia*, 231:117902, 6 2022.

[13] Boris Bokstein and Alexey Rodin. Grain boundary diffusion and grain boundary segregation in metals and alloys. *Diffusion foundations*, 1:99–122, 4 2014.

[14] Valery Borovikov, Mikhail I. Mendelev, and Alexander H. King. Effects of solutes on dislocation nucleation from grain boundaries. *International Journal of Plasticity*, 90:146–155, 12 2016.

[15] Pavel Lejček, Mojmír Šob, and Václav Paidar. Interfacial segregation and grain boundary embrittlement: An overview and critical assessment of experimental data and calculated results. *Progress in Materials Science*, 87:83–139, 6 2017.

[16] Reiner Kirchheim. Grain coarsening inhibited by solute segregation. *Acta Materialia*, 50(2):413–419, 1 2002.

[17] D. Raabe, M. Herbig, S. Sandlöbes, Y. Li, D. Tytko, M. Kuzmina, D. Ponge, and P.-p. Choi. Grain boundary segregation engineering in metallic alloys: A pathway to the design of interfaces. *Current Opinion in Solid State and Materials Science*, 18(4):253–261, 8 2014.

[18] Kazuma Ito, Yuta Tanaka, Kazumasa Tsutsui, and Hideaki Sawada. Analysis of grain-boundary segregation of hydrogen in bcc-Fe polycrystals via a nano-polycrystalline grain-boundary model. *Computational Materials Science*, 225:112196, 6 2023.

[19] Ke Ran, Wendelin Deibert, Mariya E. Ivanova, Wilhelm A. Meulenberg, and Joachim Mayer. Direction observation of the grain boundary segregation in molybdenum substituted lanthanum tungstate membranes. *Nanoscale*, 12(34):17841–17848, 1 2020.

[20] W. Kohn and L. J. Sham. Self-Consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133–A1138, 11 1965.

[21] John P. Perdew, J. A. Chevary, S. H. Vosko, Koblar A. Jackson, Mark R. Pederson, D. J. Singh, and Carlos Fiolhais. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Physical review. B, Condensed matter*, 46(11):6671–6687, 9 1992.

[22] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865–3868, 10 1996.

[23] Murray S. Daw and M. I. Baskes. Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals. *Physical review. B, Condensed matter*, 29(12):6443–6453, 6 1984.

[24] M. I. Baskes. Modified embedded-atom potentials for cubic materials and impurities. *Physical review. B, Condensed matter*, 46(5):2727–2742, 8 1992.

[25] Zhiliang Pan, Valery Borovikov, Mikhail I Mendelev, and Frederic Sansoz. Development of a semi-empirical potential for simulation of Ni solute segregation into grain boundaries in Ag. *Modelling and Simulation in Materials Science and Engineering*, 26(7):075004, 9 2018.

[26] Y. Mishin. Machine-learning interatomic potentials for materials science. *Acta Materialia*, 214:116980, 8 2021.

[27] Volker L. Deringer, Miguel A. Caro, and Gábor Csányi. Machine Learning Interatomic Potentials as emerging Tools for Materials Science. *Advanced Materials*, 31(46), 9 2019.

[28] Chi Chen and Shyue Ping Ong. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science*, 2(11):718–728, 11 2022.

[29] Bowen Deng, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Kevin Han, Christopher J.

Bartel, and Gerbrand Ceder. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5(9):1031–1041, 9 2023.

[30] Janosh Riebesell, Rhys E. A. Goodall, Anubhav Jain, Philipp Benner, Kristin A. Persson, and Alpha A. Lee. Matbench Discovery – An evaluation framework for machine learning crystal stability prediction. *arXiv (Cornell University)*, 1 2023.

[31] Ilyes Batatia, Dávid Péter Kovács, Gregor N. C. Simm, Christoph Ortner, and Gábor Csányi. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. *arXiv (Cornell University)*, 1 2022.

[32] Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M. Elena, Dávid P. Kovács, Janosh Riebesell, Xavier R. Advincula, Mark Asta, William J. Baldwin, Noam Bernstein, Arghya Bhowmik, Samuel M. Blau, Vlad Cărare, James P. Darby, Sandip De, Della Pia Flaviano, Volker L. Deringer, Rokas Elijošius, Zakariya El-Machachi, Edvin Fako, Andrea C. Ferrari, Annalena Genreith-Schriever, Janine George, Rhys E. A. Goodall, Clare P. Grey, Shuang Han, Will Handley, Hendrik H. Heenen, Kersti Hermansson, Christian Holm, Jad Jaafar, Stephan Hofmann, Konstantin S. Jakob, Hyunwook Jung, Venkat Kapil, Aaron D. Kaplan, Nima Karimitari, Namu Kroupa, Jolla Kullgren, Matthew C. Kuner, Domantas Kuryla, Guoda Liepuoniute, Johannes T. Margraf, Ioan-Bogdan Magdău, Angelos Michaelides, J. Harry Moore, Aakash A. Naik, Samuel P. Niblett, Sam Walton Norwood, Niamh O'Neill, Christoph Ortner, Kristin A. Persson, Karsten Reuter, Andrew S. Rosen, Lars L. Schaaf, Christoph Schran, Eric Sivonxay, Tamás K. Stenczel, Viktor Svahn, Christopher Sutton, Van Der Oord Cas, Eszter Varga-Umbrich, Tejs Vegge, Martin Vondrák, Yangshuai Wang, William C. Witt, Fabian Zills, and Gábor Csányi. A foundation model for atomistic materials chemistry. *arXiv (Cornell University)*, 1 2024.

[33] Yutack Park, Jaesun Kim, Seungwoo Hwang, and Seungwu Han. Scalable parallel algorithm for graph neural network interatomic potentials in molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 20(11):4857–4868, 5 2024.

[34] Han Yang, Chenxi Hu, Yichi Zhou, Xixian Liu, Yu Shi, Jielan Li, Guanzhi Li, Zekun Chen, Shuizhou Chen, Claudio Zeni, Matthew Horton, Robert Pinsler, Andrew Fowler, Daniel Zügner, Tian Xie, Jake Smith, Lixin Sun, Qian Wang, Lingyu Kong, Chang Liu, Hongxia Hao, and Ziheng Lu. MatterSim: a deep learning atomistic model across elements, temperatures and pressures. *arXiv (Cornell University)*, 5 2024.

[35] Amil Merchant, Simon Batzner, Samuel S. Schoenholz, Muratahan Aykol, Gowoon Cheon, and Ekin Dogus Cubuk. Scaling deep learning for materials discovery. *Nature*, 624(7990):80–85, 11 2023.

[36] Hai-Chen Wang, Silvana Botti, and Miguel A. L. Marques. Predicting stable crystalline compounds using chemical similarity. *npj Computational Materials*, 7(1), 1 2021.

[37] Nathan C. Frey, Ryan Soklaski, Simon Axelrod, Siddharth Samsi, Rafael Gómez-Bombarelli, Connor W. Coley, and Vijay Gadepally. Neural scaling of deep chemical models. *Nature Machine Intelligence*, 5(11):1297–1305, 10 2023.

[38] Bruno Focassio, Luis Paulo M. Freitas, and Gabriel R. Schleder. Performance Assess-

ment of Universal Machine Learning Interatomic Potentials: Challenges and directions for materials' surfaces. *ACS Applied Materials Interfaces*, 7 2024.

[39] Haochen Yu, Matteo Giantomassi, Giuliana Materzanini, Junjie Wang, and Gian-Marco Rignanese. Systematic assessment of various universal machine-learning interatomic potentials. *Materials Genome Engineering Advances*, 7 2024.

[40] Carmelo Gonzales, Eric Fuemmeler, Ellad B. Tadmor, Stefano Martiniani, and Santiago Miret. Benchmarking of universal machine learning interatomic potentials for structural relaxation. In *AI for Accelerated Materials Design - NeurIPS 2024*, 2024.

[41] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1), 7 2013.

[42] Bowen Deng, Yunyeong Choi, Peichen Zhong, Janosh Riebesell, Shashwat Anand, Zhuohan Li, KyuJung Jun, Kristin A. Persson, and Gerbrand Ceder. Overcoming systematic softening in universal machine learning interatomic potentials by fine-tuning. *arXiv (Cornell University)*, 5 2024.

[43] Kazuma Ito and Hideaki Sawada. First-principles analysis of the grain boundary segregation of transition metal alloying elements in Fe. *Computational Materials Science*, 210:111050, 7 2022.

[44] Han Lin Mai, Xiang-Yuan Cui, Daniel Scheiber, Lorenz Romaner, and Simon P. Ringer. Phosphorus and transition metal co-segregation in ferritic iron grain boundaries and its effects on cohesion. *Acta Materialia*, 250:118850, 5 2023.

[45] Zhuo Xu, Shingo Tanaka, and Masanori Kohyama. Grain-boundary segregation of 3d-transition metal solutes in bcc Fe: ab initio local-energy and d-electron behavior analysis. *Journal of Physics Condensed Matter*, 31(11):115001, 1 2019.

[46] M. I. Baskes. Application of the Embedded-Atom Method to covalent materials: a semiempirical potential for silicon. *Physical Review Letters*, 59(23):2666–2669, 12 1987.

[47] F. J. Cherne, M. I. Baskes, and P. A. Deymier. Properties of liquid nickel: A critical comparison of EAM and MEAM calculations. *Physical review. B, Condensed matter*, 65(2), 12 2001.

[48] None Munaji, None Sudarno, D L Purwaningroom, and R Arifin. Performance of EAM and MEAM potential for NITI alloys: A comparative study. *IOP Conference Series Materials Science and Engineering*, 180:012252, 3 2017.

[49] N. Gao, A. Ghoniem, X. Gao, P. Luo, K.F. Wei, and Z.G. Wang. Molecular dynamics simulation of Cu atoms interaction with symmetrical grain boundaries of BCC Fe. *Journal of Nuclear Materials*, 444(1-3):200–205, 1 2014.

[50] Xuyang Zhou, Xiao-Xiang Yu, Tyler Kaub, Richard L. Martens, and Gregory B. Thompson. Grain boundary specific segregation in nanocrystalline FE(CR). *Scientific Reports*, 6(1), 10 2016.

[51] Malik Wagih and Christopher A. Schuh. Viewpoint: Can symmetric tilt grain boundaries represent polycrystals? *Scripta Materialia*, 237:115716, 8 2023.

[52] Chongze Hu, Rémi Dingreville, and Brad L. Boyce. Computational modeling of grain boundary segregation: A review. *Computational Materials Science*, 232:112596, 1 2024.

[53] I. Adlakha and K.N. Solanki. Atomic-scale investigation of triple junction role on defects binding energetics and structural stability in -Fe. *Acta Materialia*, 118:64–76, 10 2016.

[54] Tengwu He, Yuming Qi, Yanzhou Ji, and Miaolin Feng. Grain boundary segregation-induced strengthening-weakening transition and its ideal maximum strength in nanopolycrystalline FeNiCrCoCu high-entropy alloys. *International Journal of Mechanical Sciences*, 238:107828, 1 2023.

[55] Malik Wagih and Christopher A. Schuh. Spectrum of grain boundary segregation energies in a polycrystal. *Acta Materialia*, 181:228–237, 12 2019.

[56] Liam Huber, Raheleh Hadian, Blazej Grabowski, and Jörg Neugebauer. A machine learning approach to model solute grain boundary segregation. *npj Computational Materials*, 4(1), 11 2018.

[57] Malik Wagih, Peter M. Larsen, and Christopher A. Schuh. Learning grain boundary segregation energy spectra in polycrystals. *Nature Communications*, 11(1), 12 2020.

[58] Malik Wagih and Christopher A. Schuh. Learning Grain-Boundary Segregation: From first principles to Polycrystals. *Physical Review Letters*, 129(4), 7 2022.

[59] Daniel Scheiber, Vsevolod Razumovskiy, Oleg Peil, and Lorenz Romaner. High-throughput first-principles calculations and machine learning of grain boundary segregation in metals. *Advanced Engineering Materials*, 6 2024.

[60] Chaitanya Krishna Ande and Marcel H.F. Sluiter. First-principles prediction of partitioning of alloying elements between cementite and ferrite. *Acta Materialia*, 58(19):6276–6281, 11 2010.

[61] Chandler A. Becker, Francesca Tavazza, Zachary T. Trautt, and Robert A. Buarque De Macedo. Considerations for choosing and using force fields and interatomic potentials in materials science and engineering. *Current Opinion in Solid State and Materials Science*, 17(6):277–283, 12 2013.

[62] Lucas M Hale, Zachary T Trautt, and Chandler A Becker. Evaluating variability with atomistic simulations: the effect of potential and calculation methodology on the modeling of lattice and elastic constants. *Modelling and Simulation in Materials Science and Engineering*, 26(5):055003, 5 2018.

[63] X. W. Zhou, R. A. Johnson, and H. N. G. Wadley. Misfit-energy-increasing dislocations in vapor-deposited CoFe/NiFe multilayers. *Physical Review B*, 69(14), 4 2004.

[64] G. Bonny, D. Terentyev, A. Bakaev, E.E. Zhurkin, M. Hou, D. Van Neck, and L. Malerba. On the thermal stability of late blooming phases in reactor pressure vessel steels: An atomistic study. *Journal of Nuclear Materials*, 442(1-3):282–291, 11 2013.

[65] Byeong-Joo Lee, Brian D. Wirth, Jae-Hyeok Shim, Junhyun Kwon, Sang Chul Kwon, and Jun-Hwa Hong. Modified embedded-atom method interatomic potential for theFeCualloy system and cascade simulations on pure Fe andFeCu. . . . *Physical Review B*, 71(18), 5 2005.

[66] O.R. Deluigi, R.C. Pasianot, F.J. Valencia, A. Caro, D. Farkas, and E.M. Bringa. Simulations of primary damage in a High Entropy Alloy: Probing enhanced radiation resistance. *Acta Materialia*, 213:116951, 7 2021.

[67] Ask Hjorth Larsen, Jens Jørgen Mortensen, Jakob Blomqvist, Ivano E Castelli, Rune Christensen, Marcin Dułak, Jesper Friis, Michael N Groves, Bjørk Hammer, Cory Hargus, Eric D Hermes, Paul C Jennings, Peter Bjerre Jensen, James Kermode, John R Kitchin, Esben Leonhard Kolsbjerg, Joseph Kubal, Kristen Kaasbjerg, Steen Lysgaard, Jón Bergmann Maronsson, Tristan Maxson, Thomas Olsen, Lars Pastewka, Andrew Peterson, Carsten Rostgaard, Jakob Schiøtz, Ole Schütt, Mikkel Strange, Kristian S Thygesen, Tejs Vegge, Lasse Vilhelmsen, Michael Walter, Zhenhua Zeng, and Karsten W Jacobsen. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics Condensed Matter*, 29(27):273002, 6 2017.

[68] Mark A. Tschopp, Shawn P. Coleman, and David L. McDowell. Symmetric and asymmetric tilt grain boundary structure and energy in Cu and Al (and transferability to other fcc metals). *Integrating materials and manufacturing innovation*, 4(1):176–189, 10 2015.

[69] Erik Bitzek, Pekka Koskinen, Franz Gähler, Michael Moseler, and Peter Gumbsch. Structural relaxation made simple. *Physical Review Letters*, 97(17), 10 2006.

[70] R. Fletcher. *Practical methods of optimization.* 5 2000.

[71] Jae Hoon Jang, In Gee Kim, and H.K.D.H. Bhadeshia. Substitutional solution of silicon in cementite: A first-principles study. *Computational Materials Science*, 44(4):1319–1326, 10 2008.

[72] Marcel H.F. Sluiter Sebastián Echeverri Restrepo, Naveen K. Mohandas and Anthony T. Paxton. Applicability of universal machine learning interatomic potentials to the simulation of steels. 2024.

A Appendix: Simulation Results Data

Table A.1: Lattice Parameters for BCC Fe and FCC Fe obtained by all tested potentials. The lattice parameters of DFT for BCC Fe and FCC Fe are from Mai et al. [12] and Ito et al. [43], respectively.

Potential	BCC Fe (Å)	FCC Fe (Å)
MACE-MP-0	2.859	3.634
CHGNet(v0.3.0)	2.847	3.496
CHGNet(v0.2.0)	2.849	3.463
M3GNet	2.852	3.464
SevenNet-0	2.846	3.516
FeMnNiCu_Bonny	2.855	3.658
FeCrCoNiCu_Deluigi	2.860	3.560
FeTiCoNiCuMoW_Zhou	2.866	3.628
FeCu_Lee	2.864	3.611
DFT (Mai & Ito)	2.832	3.540

Table A.2: BCC Fe lowest segregation energies of DFT, data from Mai et al [12].

E_{seg} **(eV)**	**Ti**	**V**	**Cr**	**Mn**	**Co**	**Ni**	**Cu**	**Nb**	**Mo**	**W**
$\Sigma3(1\bar{1}1)$	-0.48	-0.11	-0.18	-0.50	-0.18	-0.41	-0.51	-0.77	-0.44	-0.39
$\Sigma3(1\bar{1}2)$	-0.09	0.00	-0.07	-0.18	-0.04	-0.15	-0.24	-0.15	-0.10	-0.08
$\Sigma9(2\bar{2}1)$	-0.74	-0.12	-0.21	-0.54	-0.14	-0.43	-0.64	-1.00	-0.43	-0.31
$\Sigma11(3\bar{3}2)$	-0.65	-0.20	-0.29	-0.62	-0.15	-0.56	-0.74	-0.91	-0.55	-0.51

Table A.3: BCC Fe lowest segregation energies of MACE-MP-0. The last row of MAE is the MAE of specific element, e.g. the MAE in second column is the MAE of Ti results. Similarly, the MAE in the last column is the MAE of specific GB, and the last cell is the overall MAE.

E_{seg} **(eV)**	**Ti**	**V**	**Cr**	**Mn**	**Co**	**Ni**	**Cu**	**Nb**	**Mo**	**W**	**MAE**
$\Sigma3(1\bar{1}1)$	-0.34	-0.14	-0.02	-0.05	-0.13	-0.16	-0.19	-0.52	-0.29	-0.25	0.19
$\Sigma3(1\bar{1}2)$	-0.02	0.03	0.04	0.00	-0.02	-0.04	-0.04	0.01	-0.04	-0.05	0.10
$\Sigma9(2\bar{2}1)$	-0.55	-0.26	0.01	-0.07	-0.12	-0.15	-0.17	-0.84	-0.42	-0.42	0.21
$\Sigma11(3\bar{3}2)$	-0.26	-0.08	0.04	0.01	-0.04	-0.09	-0.06	-0.45	-0.22	-0.21	0.38
MAE	0.20	0.08	0.21	0.43	0.05	0.28	0.42	0.26	0.14	0.14	**0.22**

Table A.4: BCC Fe lowest segregation energies of CHGNet(v0.3.0).

E_{seg} **(eV)**	**Ti**	**V**	**Cr**	**Mn**	**Co**	**Ni**	**Cu**	**Nb**	**Mo**	**W**	**MAE**
$\Sigma3(1\bar{1}1)$	-1.17	-0.13	-0.11	-0.15	-0.11	-0.10	-0.08	-1.18	-0.21	-0.22	0.28
$\Sigma3(1\bar{1}2)$	0.04	0.00	-0.02	-0.01	-0.03	-0.04	-0.02	0.06	-0.04	-0.03	0.10
$\Sigma9(2\bar{2}1)$	-0.35	-0.18	-0.08	-0.14	-0.11	-0.11	-0.08	-0.64	-0.33	-0.33	0.24
$\Sigma11(3\bar{3}2)$	-0.17	-0.09	-0.07	-0.08	-0.09	-0.11	-0.07	-0.34	-0.22	-0.24	0.37
MAE	0.42	0.05	0.12	0.36	0.04	0.30	0.47	0.39	0.18	0.13	**0.25**

Table A.5: BCC Fe lowest segregation energies of CHGNet(v0.2.0).

E_{seg} **(eV)**	**Ti**	**V**	**Cr**	**Mn**	**Co**	**Ni**	**Cu**	**Nb**	**Mo**	**W**	**MAE**
$\Sigma3(1\bar{1}1)$	-0.33	-0.19	-0.04	0.00	-0.03	-0.03	-0.11	-0.64	-0.25	-0.31	0.22
$\Sigma3(1\bar{1}2)$	-0.11	-0.09	-0.04	0.00	-0.06	-0.09	-0.04	-0.17	-0.04	-0.09	0.07
$\Sigma9(2\bar{2}1)$	-0.51	-0.24	-0.12	-0.09	-0.12	-0.16	-0.15	-0.81	-0.32	-0.49	0.21
$\Sigma11(3\bar{3}2)$	-0.42	-0.25	-0.11	-0.04	-0.11	-0.18	-0.13	-0.66	-0.28	-0.40	0.27
MAE	0.16	0.09	0.11	0.43	0.06	0.28	0.42	0.15	0.16	0.10	**0.19**

Table A.6: BCC Fe lowest segregation energies of M3GNet.

E_{seg} **(eV)**	**Ti**	**V**	**Cr**	**Mn**	**Co**	**Ni**	**Cu**	**Nb**	**Mo**	**W**	**MAE**
$\Sigma3(1\bar{1}1)$	-1.02	-0.38	-0.33	-0.29	-0.33	-0.41	-0.44	-0.91	-0.44	-0.57	0.17
$\Sigma3(1\bar{1}2)$	-0.45	-0.10	-0.09	-0.10	-0.05	-0.08	-0.06	-0.26	-0.11	-0.18	0.10
$\Sigma9(2\bar{2}1)$	-1.28	-0.25	-0.24	-0.16	-0.29	-0.29	-0.49	-1.10	-0.38	-0.51	0.19
$\Sigma11(3\bar{3}2)$	-1.00	-0.24	-0.13	-0.09	-0.11	-0.16	-0.18	-0.79	-0.37	-0.54	0.24
MAE	0.45	0.14	0.09	0.30	0.09	0.15	0.24	0.12	0.06	0.13	**0.18**

Table A.7: BCC Fe lowest segregation energies of SevenNet-0.

E_{seg} **(eV)**	**Ti**	**V**	**Cr**	**Mn**	**Co**	**Ni**	**Cu**	**Nb**	**Mo**	**W**	**MAE**
$\Sigma3(1\bar{1}1)$	-0.08	-0.02	0.03	-0.02	-0.01	0.01	-0.05	-0.18	-0.04	-0.04	0.36
$\Sigma3(1\bar{1}2)$	-0.22	-0.15	-0.08	-0.37	-0.47	-0.07	-0.06	-0.58	-0.25	-0.34	0.20
$\Sigma9(2\bar{2}1)$	-0.89	-0.34	-0.15	-0.12	-0.05	-0.03	-0.11	-1.18	-0.39	-0.44	0.22
$\Sigma11(3\bar{3}2)$	-0.61	-0.34	-0.08	-0.03	-0.27	-0.09	-0.03	-0.89	-0.40	-0.55	0.25
MAE	0.18	0.15	0.12	0.42	0.20	0.34	0.47	0.31	0.18	0.19	**0.26**

Table A.8: BCC Fe lowest segregation energies of FeMnNiCu_Bonny.

E_{seg} **(eV)**	**Mn**	**Ni**	**Cu**	**MAE**
$\Sigma3(1\bar{1}1)$	-0.08	-0.12	-0.18	0.35
$\Sigma3(1\bar{1}2)$	-0.03	-0.00	-0.02	0.17
$\Sigma9(2\bar{2}1)$	0.00	-0.10	-0.02	0.50
$\Sigma11(3\bar{3}2)$	-0.03	-0.13	-0.11	0.55
MAE	0.43	0.30	0.45	**0.39**

Table A.9: BCC Fe lowest segregation energies of FeTiCoNiCuMoW_Zhou. Unconverged results were not presented (empty cells in the table).

E_{seg} **(eV)**	**Ti**	**Co**	**Ni**	**Cu**	**Mo**	**W**	**MAE**
$\Sigma3(1\bar{1}1)$	-0.33	-0.12	-0.14	-0.19	-0.25	-0.19	0.20
$\Sigma3(1\bar{1}2)$	0.02	-0.02	-0.05	-0.05	-0.00	0.03	0.10
$\Sigma9(2\bar{2}1)$		-0.11	-0.17	-0.18			0.25
$\Sigma11(3\bar{3}2)$	-0.40	-0.08	-0.09	-0.20	-0.34	-0.30	0.29
MAE	0.17	0.05	0.28	0.38	0.17	0.17	**0.21**

Table A.10: BCC Fe lowest segregation energies of FeCu_Lee.

E_{seg} **(eV)**	**Cu**
$\Sigma3(1\bar{1}1)$	-0.29
$\Sigma3(1\bar{1}2)$	-0.09
$\Sigma9(2\bar{2}1)$	-0.31
$\Sigma11(3\bar{3}2)$	-0.42
MAE	**0.26**

Table A.11: FCC Fe lowest segregation energies of DFT, data from Ito et al [43].

E_{seg} **(eV)**	**Ti**	**V**	**Cr**	**Mn**	**Co**	**Ni**	**Cu**	**Nb**	**Mo**
$\Sigma37(610)$	-0.89	-0.22	0.06	-0.15	-0.08	-0.07	-0.22	-1.25	-0.62
$\Sigma13(510)$	-0.94	-0.18	0.02	-0.18	-0.11	-0.08	-0.23	-1.25	-0.62
$\Sigma17(410)$	-0.89	-0.25	0.03	-0.17	-0.10	-0.05	-0.23	-1.20	-0.60
$\Sigma5(310)$	-0.79	-0.23	0.03	-0.10	-0.11	-0.05	-0.12	-1.09	-0.58
$\Sigma29(520)$	-0.77	-0.33	-0.03	-0.11	-0.09	-0.05	-0.14	-1.02	-0.65
$\Sigma5(210)$	-0.67	-0.35	0.00	0.05	-0.05	0.00	-0.11	-0.84	-0.53
$\Sigma13(320)$	-0.51	-0.10	0.09	0.00	-0.05	-0.02	-0.24	-0.80	-0.47
$\Sigma25(430)$	-0.59	-0.17	0.04	-0.05	-0.07	-0.03	-0.19	-0.88	-0.51
$\Sigma41(540)$	-0.61	-0.18	0.02	-0.05	-0.09	-0.07	-0.23	-0.89	-0.52

Table A.12: FCC Fe lowest segregation energies of MACE-MP-0.

E_{seg} **(eV)**	**Ti**	**V**	**Cr**	**Mn**	**Co**	**Ni**	**Cu**	**Nb**	**Mo**	**MAE**
$\Sigma37(610)$	-0.59	-0.28	-0.16	-0.21	-0.09	-0.06	-0.19	-0.85	-0.45	0.14
$\Sigma13(510)$	-0.57	-0.29	-0.13	-0.20	-0.11	-0.07	-0.19	-0.82	-0.45	0.15
$\Sigma17(410)$	-0.57	-0.29	-0.08	-0.16	-0.12	-0.08	-0.19	-0.83	-0.44	0.12
$\Sigma5(310)$	-0.61	-0.27	-0.04	-0.12	-0.17	-0.12	-0.19	-0.87	-0.46	0.10
$\Sigma29(520)$	-0.57	-0.36	-0.09	-0.14	-0.17	-0.14	-0.19	-0.80	-0.51	0.10
$\Sigma5(210)$	-0.14	-0.10	-0.05	-0.12	-0.02	-0.01	-0.07	-0.21	-0.11	0.24
$\Sigma13(320)$	-0.80	-0.66	-0.45	-0.47	-0.48	-0.37	-0.54	-0.95	-0.70	0.37
$\Sigma25(430)$	-0.60	-0.48	-0.26	-0.30	-0.30	-0.16	-0.36	-0.75	-0.53	0.17
$\Sigma41(540)$	-0.53	-0.41	-0.19	-0.24	-0.23	-0.10	-0.31	-0.69	-0.46	0.13
MAE	0.26	0.18	0.19	0.14	0.11	0.08	0.09	0.31	0.17	**0.17**

Table A.13: FCC Fe lowest segregation energies of CHGNet(v0.3.0).

E_{seg} **(eV)**	**Ti**	**V**	**Cr**	**Mn**	**Co**	**Ni**	**Cu**	**Nb**	**Mo**	**MAE**
Σ37(610)	-0.55	-0.20	-0.01	-0.03	-0.05	-0.05	-0.18	-0.87	-0.35	0.15
Σ13(510)	-0.53	-0.22	-0.01	-0.01	-0.05	-0.05	-0.16	-0.85	-0.36	0.14
Σ17(410)	-0.50	-0.23	-0.03	0.03	-0.05	-0.05	-0.15	-0.83	-0.37	0.16
Σ5(310)	-0.46	-0.19	-0.04	0.07	-0.08	-0.07	-0.13	-0.65	-0.30	0.15
Σ29(520)	-0.40	-0.24	-0.06	0.04	-0.08	-0.07	-0.13	-0.64	-0.34	0.15
Σ5(210)	-0.27	-0.21	-0.02	0.07	-0.02	0.02	-0.07	-0.43	-0.25	0.15
Σ13(320)	-0.27	-0.11	0.01	0.07	-0.02	-0.04	-0.14	-0.44	-0.26	0.15
Σ25(430)	-0.30	-0.22	-0.15	-0.14	-0.03	-0.16	-0.14	-0.45	-0.30	0.12
Σ41(540)	-0.37	-0.27	-0.08	-0.11	-0.05	-0.19	-0.24	-0.46	-0.36	0.16
MAE	0.33	0.05	0.07	0.11	0.03	0.04	0.05	0.40	0.25	**0.14**

Table A.14: FCC Fe lowest segregation energies of CHGNet(v0.2.0). Unconverged results were not presented (empty cells in the table).

E_{seg} **(eV)**	**Ti**	**V**	**Cr**	**Mn**	**Ni**	**Cu**	**Nb**	**Mo**	**Co**	**MAE**
Σ37(610)	-1.07	-1.03	-1.10	-0.84	-1.99	-0.68	-1.42	-2.06		0.93
Σ13(510)	0.94	1.00	-0.03	0.12	0.03	1.23	0.62	0.04	0.00	0.85
Σ17(410)	1.01	0.95	-0.04	0.14	0.03	1.25	0.71	0.04	0.02	0.86
Σ5(310)	-0.82	-0.84	-1.82	-1.47	-1.77	-0.51	-1.09	-1.76	-1.83	0.98
Σ29(520)	0.97	0.16	-0.12	0.15	0.01	1.22	-0.77	-1.38	-0.01	0.62
Σ5(210)	1.04	1.09	0.05	0.26	0.01	1.23	0.80	0.01	0.01	0.78
Σ13(320)	1.03	1.05	-0.04	0.14	-0.00	1.23	0.85	-0.02	-0.02	0.73
Σ25(430)	1.01	1.07	-0.16	0.04	-0.06	1.19	0.84	-0.08	-0.06	0.74
Σ41(540)	0.60	0.74	-0.55	-0.35	-0.43	0.80	0.44	-0.36	-0.37	0.62
MAE	1.31	1.00	0.46	0.41	0.49	1.06	0.92	0.81	0.59	**0.79**

Table A.15: FCC Fe lowest segregation energies of SevenNet-0. Unconverged results were not presented (empty cells in the table).

E_{seg} **(eV)**	**Ti**	**V**	**Cr**	**Mn**	**Co**	**Ni**	**Cu**	**Nb**	**Mo**	**MAE**
Σ37(610)	1.64	3.29	2.94	4.49	3.79	2.38	2.25	0.50	2.62	3.03
Σ13(510)	1.58	2.82	3.27	4.10	3.52	2.37	2.25	0.77	3.22	3.05
Σ17(410)	1.39	2.62	2.92	3.50	2.60	1.77	1.92	0.01	3.04	2.58
Σ5(310)	2.36	3.08	3.31	4.43	3.54	2.69		1.39	3.75	3.43
Σ29(520)	2.12	2.84	3.30	4.00	3.73	2.32	2.61	0.55	2.87	3.06
Σ5(210)	2.06	2.94	3.14	4.27	3.39	2.25		1.16	3.56	3.14
Σ13(320)	-0.59	-0.29	0.31	1.51	0.54	-0.28	-0.44	-1.66	1.04	0.60
Σ25(430)	2.54	3.36	3.71	4.68	3.55	2.59	-0.71	1.54	3.86	3.18
Σ41(540)			3.20	4.32	3.54	2.35				3.40
MAE	2.41	2.86	2.87	4.01	3.22	2.15	1.76	1.79	3.57	**2.78**

Table A.16: FCC Fe lowest segregation energies of FeCrCoNiCu_Deluigi.

E_{seg} **(eV)**	**Cr**	**Co**	**Ni**	**Cu**	**MAE**
Σ37(610)	-0.47	-0.34	-0.33	-0.72	0.39
Σ13(510)	-0.35	-0.17	-0.22	-0.49	0.21
Σ17(410)	-0.34	-0.19	-0.22	-0.47	0.22
Σ5(310)	-0.25	-0.15	-0.16	-0.33	0.16
Σ29(520)	-0.72	-0.62	-0.64	-0.89	0.64
Σ5(210)	-0.27	-0.16	-0.18	-0.31	0.19
Σ13(320)	-0.25	-0.13	-0.16	-0.35	0.17
Σ25(430)	-0.26	-0.14	-0.16	-0.43	0.19
Σ41(540)	-0.26	-0.15	-0.17	-0.45	0.16
MAE	0.38	0.14	0.20	0.31	**0.26**

Table A.17: FCC Fe lowest segregation energies of FeCu_Lee.

E_{seg} **(eV)**	**Cu**
Σ37(610)	-0.64
Σ13(510)	-0.55
Σ17(410)	-0.18
Σ5(310)	-0.62
Σ29(520)	-0.92
Σ5(210)	-0.18
Σ13(320)	-0.37
Σ25(430)	-0.44
Σ41(540)	-0.47
MAE	**0.49**

Table A.18: DFT results (in eV) of solution enthalpies for different alloying elements, data from Ande and Sluiter [60]. FM: ferromagnetic; NM: non-ferromagnetic. 4c and 8d are different substitutional sites of Fe atoms in cementite.

Element	DFT_FM_4c	DFT_FM_8d	DFT_NM_4c	DFT_NM_8d
Al	-0.17	-0.24	-0.54	-0.53
Si	0.54	0.29	0.17	-0.07
P	1.6	0.39	1.35	0.44
S	2.77	1.0	2.77	1.04
Ti	-1.03	-1.04	-1.41	-1.40
V	-0.63	-0.72	-0.98	-1.12
Cr	0.0	-0.1	-0.33	-0.51
Mn	0.08	0.05	-0.19	-0.30
Co	0.09	0.04	0.05	0.13
Ni	0.22	0.14	0.11	0.25
Cu	0.98	0.88	0.79	0.97
Nb	-0.1	-0.15	-0.34	-0.40
Mo	0.24	0.18	0.01	-0.11
W	0.28	0.2	0.02	-0.13

Table A.19: Results (in eV) of solution enthalpies for different alloying elements calculated by MACE-MP-0.

Element	MACE_4c	MACE_8d
Al	0.26	0.26
Si	0.20	0.19
P	0.28	0.25
S	0.46	0.38
Ti	-0.04	-0.04
V	-0.08	-0.09
Cr	-0.07	-0.08
Mn	-0.03	-0.04
Co	0.09	0.09
Ni	0.18	0.18
Cu	0.34	0.34
Nb	-0.10	-0.11
Mo	-0.13	-0.14
W	-0.24	-0.26

Table A.20: Results (in eV) of solution enthalpies for different alloying elements calculated by CHGNet (v0.3.0).

Element	CHGNetv3_4c	CHGNetv3_8d
Al	0.27	0.27
Si	0.18	0.18
P	0.19	0.18
S	0.35	0.30
Ti	-0.05	-0.05
V	-0.09	-0.09
Cr	-0.09	-0.09
Mn	-0.06	-0.06
Co	0.08	0.09
Ni	0.17	0.17
Cu	0.32	0.32
Nb	-0.12	-0.13
Mo	-0.15	-0.15
W	-0.28	-0.28

Table A.21: Results (in eV) of solution enthalpies for different alloying elements calculated by CHGNet (v0.2.0).

Element	CHGNetv2_4c	CHGNetv2_8d
Al	0.27	0.26
Si	0.21	0.20
P	0.22	0.19
S	0.32	0.29
Ti	-0.06	-0.07
V	-0.09	-0.09
Cr	-0.07	-0.08
Mn	-0.05	-0.06
Co	0.08	0.08
Ni	0.17	0.17
Cu	0.31	0.31
Nb	-0.14	-0.14
Mo	-0.14	-0.14
W	-0.27	-0.27

Table A.22: Results for solution enthalpies (in eV) of different alloying elements calculated by M3GNet.

Element	M3GNet_4c	M3GNet_8d
Al	0.84	-0.94
Si	-1.05	5.36
P	8.17	5.54
S	3.09	0.88
Ti	-1.27	1.99
V	-0.07	0.62
Cr	-1.30	-0.40
Mn	5.40	0.48
Co	0.20	14.58
Ni	-1.03	0.70
Cu	-0.90	4.04
Nb	1.07	5.73
Mo	2.97	-1.37
W	-0.50	-1.40

Table A.23: Results for solution enthalpies (in eV) of different alloying elements calculated by SevenNet-0.

Element	SevenNet-0_4c	SevenNet-0_8d
Al	0.27	0.26
Si	0.19	0.16
P	0.23	0.19
S	0.38	0.35
Ti	-0.05	-0.05
V	-0.09	-0.10
Cr	-0.09	-0.10
Mn	-0.05	-0.05
Co	0.09	0.09
Ni	0.17	0.17
Cu	0.33	0.33
Nb	-0.13	-0.13
Mo	-0.15	-0.15
W	-0.28	-0.28

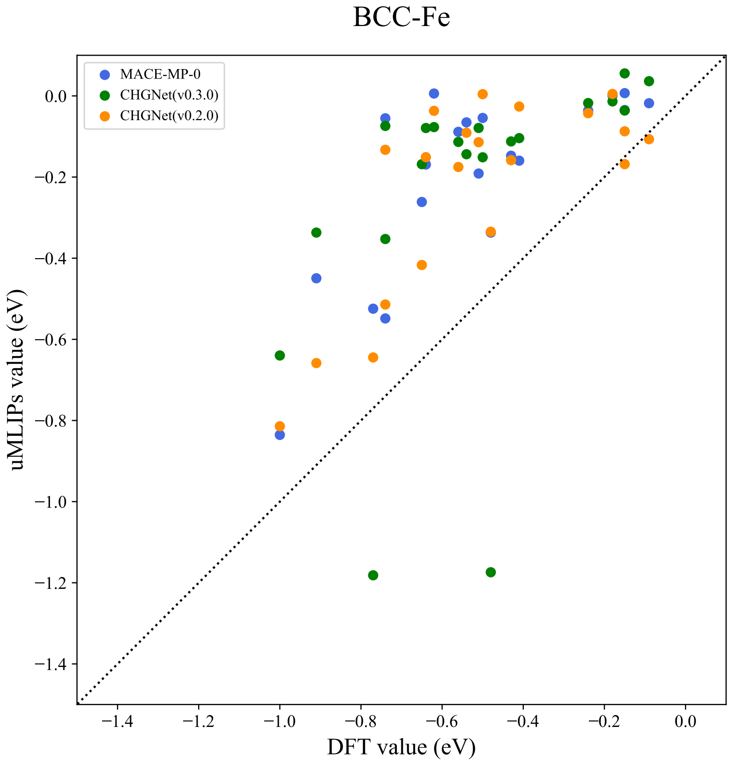


Figure A.1: GB segregation energies of MACE and CHGNet at three BCC GBs, $\Sigma3(1\bar{1}1)$, $\Sigma9(2\bar{2}1)$ and $\Sigma11(3\bar{3}2)$ for solutes Ti, Mn, Ni, Cu and Nb. The underprediction can be clearly observed with few exceptions.