Machine learning reveals transferable rules for grain boundary segregation energies

Jingbei Bai [1 2] **Che Fan** [3] **Beilin Ye** [1] **Bo Hu** [1] **Siyu Liu** [1] **Tongqi Wen** [1]
[1]*Center for Structural Materials, Department of Mechanical Engineering, The University of Hong Kong, Hong Kong SAR, China* [2]*Xingjian College, Tsinghua University, Beijing, China* [3]*Department of Materials Science and Engineering, City University of Hong Kong, Hong Kong SAR, China*. Correspondence to: Tongqi Wen tongqwen@hku.hk.

1. Introduction

Grain boundaries (GBs) fundamentally influence material properties [1, 2, 3, 4]. Solute segregation at these interfaces alters local chemistry and atomic structure, effectively modulating macroscopic performance [5, 6, 7, 8, 9, 10, 11, 12]. Consequently, understanding segregation behavior is essential for tailoring structural and functional reliability [13, 14]. While the segregation energy spectrum is critical for controlling this behavior, conventional first-principles density functional theory (DFT) and molecular dynamics (MD) methods are often limited by computational cost or the accuracy of interatomic potentials [15, 16]. Machine learning (ML) offers a scalable solution, provided robust descriptors and training protocols are established [17, 18].

In this work, we demonstrate accurate and transferable ML prediction of GB segregation energies and the associated segregation trends across metallic binary alloys by extracting transferable rules from a consistent dataset and modeling method. Using polycrystalline models as a common foundation, we systematically quantify how polycrystal size, descriptor choice, and algorithm selection influence predictive accuracy and computational cost. This analysis identifies practical modeling choices that enable reliable training data generation and robust prediction, including an efficient polycrystal size (~170 Å with 10 grains), a high-fidelity descriptor (Smooth Overlap of Atomic Positions (SOAP), retaining strong performance under dimensionality reduction), and an accurate, stable learning strategy (ensemble methods, particularly Extra Trees). Collectively, these results provide a transferable approach for high-throughput prediction of GB segregation energies and behavior, supporting accelerated screening and microstructure-guided design of alloys with targeted interfacial chemistry and performance.

2. Substantial section

This work establishes a systematic computational workflow integrating large-scale polycrystal modeling, high-throughput MD calculations, and ML to predict GB segregation behavior. As shown in Figure 1, the unified framework (1) generates Voronoi tessellated polycrystals relaxed via LAMMPS; (2) computes segregation energies using the site-by-site substitution method:

$$\Delta E_i^{seg} = E_{gb,i}^{solute} - E_{bulk}^{solute} \tag{1}$$

and (3) optimizes ML models mapping local atomic environments to energies.

As summarized in Figure 2, our systematic evaluation identifies the specific modeling parameters required for accurate and efficient GB segregation prediction. First, regarding structural representation, a polycrystal size of ~170 Å containing 10 grains is found to offer the optimal balance, sufficiently sampling diverse GB motifs while maintaining computational tractability. Second, in terms of feature determination, SOAP descriptor demonstrates the highest fidelity in encoding local environments, with an optimized configuration of $n_{max} = 10$, $l_{max} = 5$, and a 5 Å cutoff. Structurally, we show that SOAP is robust to dimensionality reduction; approximating the descriptor with ~20 principal components via PCA preserves near-full accuracy while significantly reducing computational cost. Finally, comparative algorithm analysis reveals that the Extra Trees model yields the strongest performance ($R^2 > 0.95$ for most binary alloys).

2.1 Related work

Traditional atomistic simulations, such as MD and DFT, are standard for probing GB segregation but are computationally limited when sampling the immense structural diversity of polycrystalline networks [16, 19, 20]. Consequently, data-driven approaches have evolved from utilizing simple geometric descriptors—such as coordination analysis and Voronoi volume [17], to employing high-fidelity, atom-centered descriptors like the SOAP. To address the high dimensionality of such descriptors, techniques like PCA are increasingly integrated. Wagih et al. have demonstrated that coupling PCA-reduced SOAP features with regression models enables efficient and accurate predictions across diverse binary and complex alloy systems [19].

Parallel to feature engineering, the selection of regression algorithms has shifted to capture the nonlinear nature of interfacial energetics. While linear models provide interpretability for general spectral trends [19], our studies indicate that tree-based ensemble methods, such as Random Forests and Extra Trees, significantly outperform linear regression in reproducing complex site-specific segregation behaviors. These findings highlight the necessity of a systematic workflow that concurrently optimizes polycrystal sampling, descriptor fidelity, and algorithmic complexity to achieve robust transferability.

2.2 *Figures*

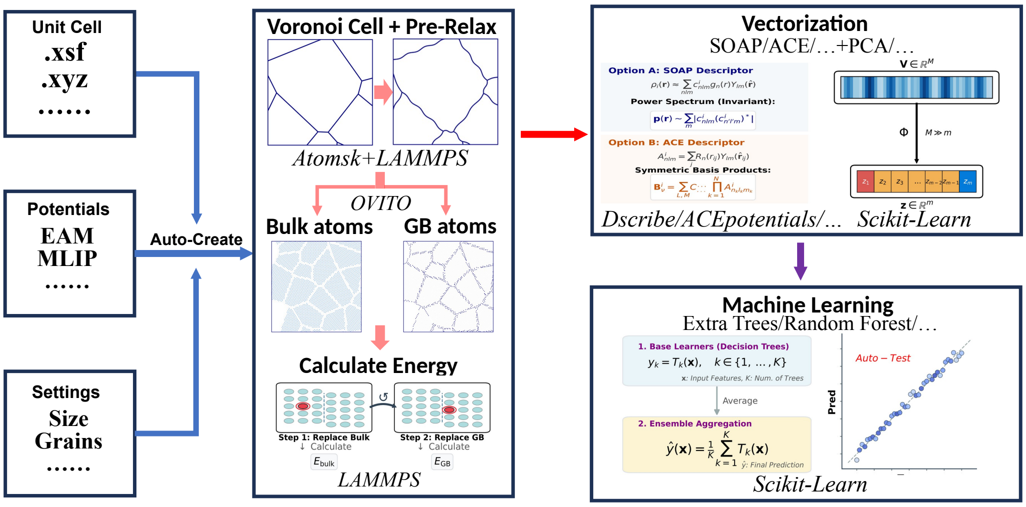

Fig. 1: Automated workflow for GB segregation energy prediction. (a) Structure generation and preprocessing: Polycrystalline supercells are constructed via Voronoi tessellation, followed by pre-relaxation in LAMMPS. Bulk and GB atoms are identified in OVITO, and site-specific energies are calculated (E_{bulk},E_{GB}). (b) Vectorization: Local atomic environments are encoded via SOAP, ACE, or related descriptors and reduced to low-dimensional feature vectors (**z**) using PCA. (c) Machine learning: Ensemble learning models such as Extra Trees and Random Forest regressors are trained to map descriptor vectors to atomic energies, enabling accurate and efficient prediction of GB segregation energies.

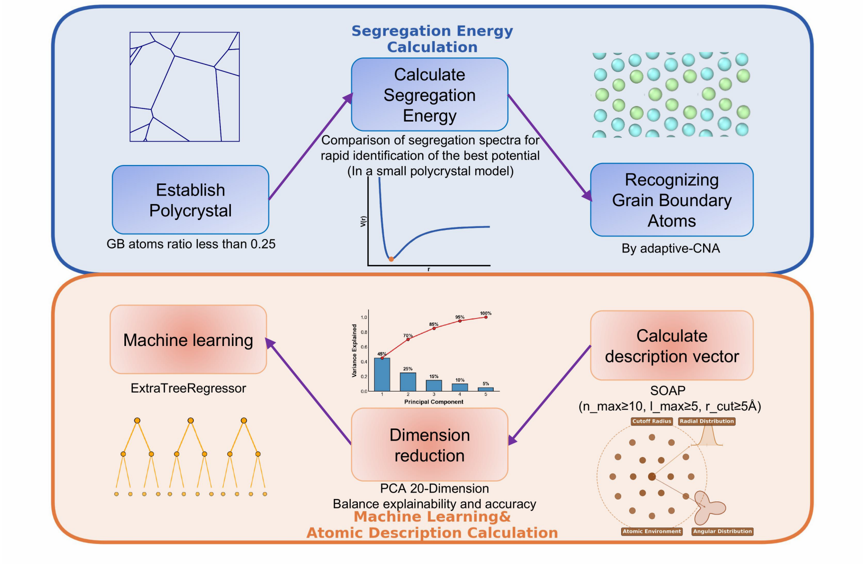

Fig. 2: Summary of the transferable rules and end-to-end workflow, along with the recommended parameter settings, for accurate and efficient machine-learning prediction of grain-boundary segregation energies.

Acknowledgments

This work was supported by Guangdong Natural Science Fund (2025A1515012129) and The University of Hong Kong (HKU) via seed fund (2509100468).

References

[1] Helena Van Swygenhoven. Grain Boundaries and Dislocations. *Science*, 296(5565):66–67, April 2002. Publisher: American Association for the Advancement of Science.

[2] H. Gleiter. On the structure of grain boundaries in metals. *Materials Science and Engineering*, 52(2):91–131, February 1982.

[3] D. G Brandon. The structure of high-angle grain boundaries. *Acta Metallurgica*, 14(11):1479–1484, November 1966.

[4] Tadao Watanabe. Grain boundary engineering: historical perspective and future prospects. *Journal of Materials Science*, 46(12):4095–4115, June 2011.

[5] David N. Seidman. Subnanoscale Studies of Segregation at Grain Boundaries: Simulations and Experiments. *Annual Review of Materials Research*, 32(Volume 32, 2002):235–269, August 2002. Publisher: Annual Reviews.

[6] 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, August 2014.

[7] M. P. Seah. Grain boundary segregation. *Journal of Physics F: Metal Physics*, 10(6):1043, June 1980.

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

[9] Zhi Zhang, Jinshu Xie, Jinghuai Zhang, Xu-Sheng Yang, and Ruizhi Wu. Towards designing high mechanical performance low-alloyed wrought magnesium alloys via grain boundary segregation strategy: A review. *Journal of Magnesium and Alloys*, 12(5):1774–1791, May 2024.

[10] 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, June 2022.

[11] Chang Liu, Jing Rao, Zhongji Sun, Wenjun Lu, James P. Best, Xuehan Li, Wenzhen Xia, Yilun Gong, Ye Wei, Bozhao Zhang, Jun Ding, Ge Wu, and En Ma. Near-theoretical strength and deformation stabilization achieved via grain boundary segregation and nano-clustering of solutes. *Nature Communications*, 15(1):9283, October 2024.

[12] Michael A. Gibson and Christopher A. Schuh. Segregation-induced changes in grain boundary cohesion and embrittlement in binary alloys. *Acta Materialia*, 95:145–155, August 2015.

[13] Xiaoying Qian, Zhihua Dong, Bin Jiang, Bin Lei, Huabao Yang, Chao He, Lintao Liu, Cuihong Wang, Ming Yuan, Hong Yang, Baoqing Yang, Changyong Zheng, and Fusheng Pan. Influence of alloying element segregation at grain boundary on the microstructure and mechanical properties of Mg-Zn alloy. *Materials & Design*, 224:111322, December 2022.

[14] Haoyang Song, Chenyang Zhao, Haonan Bai, Xinke Ren, Hongfei Shao, Jinze Chi, Guojiang Dong, Jiang Bi, and Caiwang Tan. Investigation of grain boundary segregation evolution and corrosion behavior in 7050 aluminum alloy under oscillating laser melting. *Journal of Alloys and Compounds*, 1010:177524, January 2025.

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

[16] Malik Wagih, Yannick Naunheim, Tianjiao Lei, and Christopher A. Schuh. Grain boundary segregation predicted by quantum-accurate segregation spectra but not by classical models. *Acta Materialia*, 266:119674, March 2024.

[17] Joseph Messina, Renjie Luo, Ke Xu, Guanghong Lu, Huiqiu Deng, Mark A Tschopp, and Fei Gao. Machine learning to predict aluminum segregation to magnesium grain boundaries. *Scripta Materialia*, 204:114150, November 2021.

[18] Ke Xu, Shuo Jin, and Guang-Hong Lu. Predicting hydrogen segregation energy distributions in strained regions of tungsten using artificial neural network. *Nuclear Materials and Energy*, 39:101637, June 2024.

[19] Malik Wagih, Peter M. Larsen, and Christopher A. Schuh. Learning grain boundary segregation energy spectra in polycrystals. *Nature Communications,* 11(1):6376, December 2020.

[20] Doruk Aksoy, Jian Luo, Penghui Cao, and Timothy J Rupert. A machine learning framework for the prediction of grain boundary segregation in chemically complex environments. *Modelling and Simulation in Materials Science and Engineering,* 32(6):065011, June 2024. Publisher: IOP Publishing.