


Learning Grain-Boundary Segregation: From First Principles to Polycrystals

Malik Wagih¹ and Christopher A. Schuh^{1*}*Department of Materials Science and Engineering, Massachusetts Institute of Technology,
77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA* (Received 9 March 2022; revised 6 May 2022; accepted 19 June 2022; published 21 July 2022)

The segregation of solute atoms at grain boundaries (GBs) can strongly impact the structural and functional properties of polycrystals. Yet, due to the limited availability of simulation tools to study polycrystals at the atomistic scale (i.e., interatomic potentials), there is a minimal understanding of the variation of solute segregation tendencies across the very complex space of GB microenvironments and the large range of alloys in which it can occur. Here, we develop an algorithmic framework that can directly learn the full spectrum of segregation energies for a metal solute atom in a metal polycrystal from *ab initio* methods, bypassing the need for alloy interatomic potentials. This framework offers a pathway to a comprehensive catalog of GB solute segregation with quantum accuracy, for the entire alloy space. As an initial demonstration in this pursuit, we build an extensive GB segregation database for aluminum-based alloys across the periodic table, including dozens of alloys for which there are substantially no prior data.

DOI: [10.1103/PhysRevLett.129.046102](https://doi.org/10.1103/PhysRevLett.129.046102)

A perennial fundamental problem in metal physics is solute segregation at grain boundaries (GBs) in polycrystalline alloys [1–3]. Such segregation can occur with only a trace presence of a solute, and yet can strongly impact mechanical, electrochemical, electrical, and magnetic properties, to name a few [4–6]. And, as most metals are used in a polycrystalline form, this phenomenon has wide-ranging engineering implications for the design and use of alloys [6]. The main driving force for solute segregation comes from the atomically disordered nature of GBs, which results in an array of local atomic environments (site types) that can be more accommodating to the solute atom than the intragrain (bulk) highly ordered lattice environment. The nature of the spectrum of site types at the GBs, and the degree to which each site type accommodates the solute atom, i.e., the segregation energy for a solute atom at that site type, determines the extent of equilibrium GB solute segregation [7–10]. A major challenge is that very little is known about this spectrum of GB site types (and the corresponding segregation energies for solutes) in polycrystals.

Recently, in Ref. [11], we took some first steps to tackle this challenge. Using molecular statics and dynamics simulations to exhaustively compute the segregation energies (as defined in Supplemental Material [12], Sec. 1) for an Mg solute at all GB site types in a randomly oriented aluminum polycrystal, we found the spectrum of GB site types to be captured by a skew-normal distribution. Subsequently, in Ref. [13], we developed a learning framework that can predict the segregation energy of a solute atom at a GB site, based solely on its pre-segregation (undecorated) local atomic environment, for systems that have extant interatomic potentials. Unfortunately, the use of

interatomic potentials is a severe limitation. There are only a limited number of available potentials, and few of those are accurate for the special conditions of GB segregation [14]. One obvious solution to these limitations is to develop more alloy interatomic potentials that are quantum accurate, which is becoming increasingly possible with the advent of machine-learning-based potentials [15,16]. However, this remains far from a trivial task, since state-of-the-art interatomic potentials are expensive to develop [17,18]. Their fitting typically requires a reference database of thousands of expensive *ab initio* calculations of atomic configurations, often chosen painstakingly through a manual procedure to ensure good coverage of the structural and chemical space. And though there are new interesting developments [19,20] that could potentially reduce the number of required training configurations and automate their selection procedure, this is still an active area of research that needs to be tested for its ability to capture (or extrapolate to) complex grain-boundary atomic environments.

Therefore, in this Letter, we take an alternative approach that bypasses the need for accurate interatomic potentials for the problem of GB solute segregation, by directly passing from *ab initio* calculations to a full macroscale GB segregation isotherm for polycrystals. The framework combines a learning model with hybrid quantum mechanics–molecular mechanics (QM-MM) calculations, and requires only an interatomic potential for the base (solvent) metal. Using aluminum [21] as the model base metal, we showcase the power of our learning framework by calculating—for the first time—GB solute segregation spectra in polycrystals for all group II, transition, and post-transition metals (excluding lanthanides, actinides, and synthetic

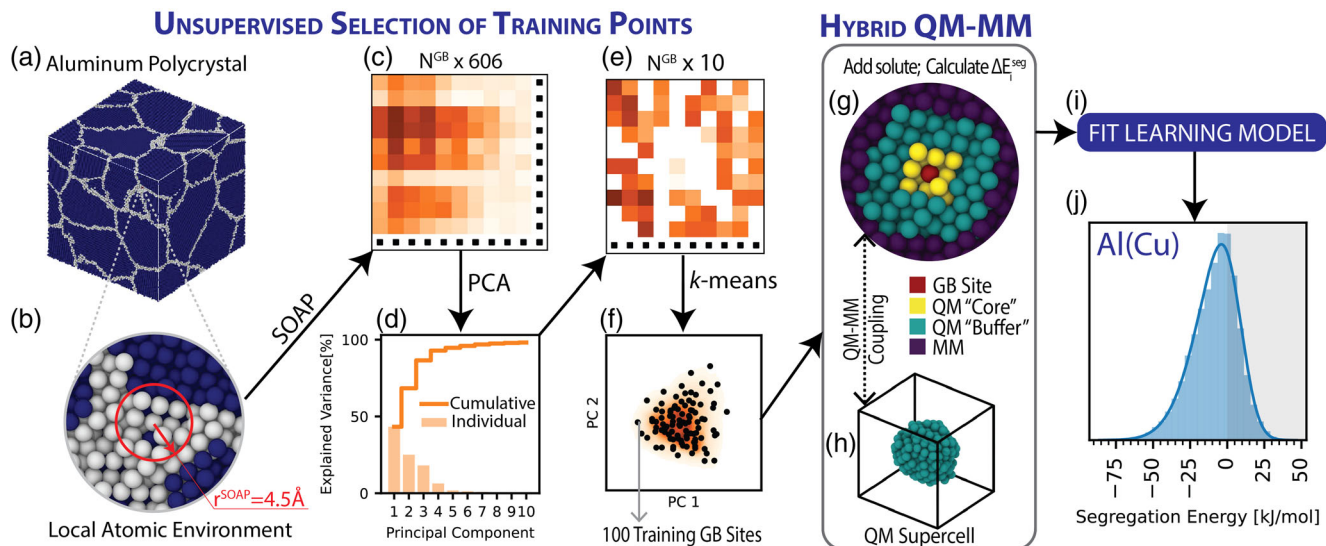


FIG. 1. An illustration of the *ab initio* based learning framework for GB segregation energies. For (a) the annealed $(20 \text{ nm})^3$ aluminum polycrystal [23–27], we convert the (b) local atomic environment of every GB site within a cutoff radius of 4.5 \AA into (c) a feature vector of 606 dimensions using the SOAP method [28,29]. (d) PCA [30,31] is used to transform the feature matrix into a low-dimensional space, and obtain the first 10 principal components (which capture $>99\%$ of the variance in feature space). Using (e) the reduced feature matrix, we apply (f) *k*-means clustering [32,33] to divide GB sites into 100 clusters, and use the centroids of the clusters as the training data points for the model, for which segregation energies need to be calculated using (g),(h) QM-MM method as detailed [34–42] in Supplemental Material, Sec. 4 [12]. (i) Linear regression is used to fit a learning algorithm to the training dataset, and to predict (j) the segregation energies of all GB sites in the polycrystal (i.e., the full spectrum) as shown, e.g., for Cu solute segregation in Al polycrystals.

elements)—a total of 39 solute metals. Our approach favors simplicity and flexibility, and is designed to be easily adapted to treat other base metals.

Our *ab initio* based framework has two key components, as illustrated in Fig. 1. First is the learning model, which enables the unsupervised selection of an optimal 100 training data points in polycrystalline GB environments, i.e., without any knowledge of the target properties (segregation energies), and the fitting of a regression model that can be used to accurately predict the segregation energies for all other GB sites in the polycrystal. In Ref. [13], we have shown that this learning model can accurately predict the segregation energies for large numbers of GB environments that are sampled in random microstructures (where the sampling probability of GB disorientation angles follows the Mackenzie distribution), and across multiple alloys. This model can be used to treat any subset of the GB disorientation space (e.g., low-angle boundaries, or sub-populations of the polycrystalline GB microenvironments, e.g., triple junctions and quadruple nodes [22]) by selecting a base polycrystal that samples the subset of interest). However, here we are interested in the full polycrystalline environment, and therefore use a randomly oriented aluminum polycrystal, as shown in Fig. 1(a). And thus, the selection of training data points is optimized to reduce overall errors for a random microstructure.

We refer the reader to Ref. [13] for a more detailed description of the learning model; we briefly reiterate its

main pieces here through the summary provided in Fig. 1. The model consists of four main steps: (i) feature extraction, (ii) reduction of the dimensions of features, (iii) selection of training data points (for which to calculate segregation energies), and (iv) the learning algorithm. For feature extraction, each GB site and its local atomic environment within a radius of 4.5 \AA , as illustrated in Fig. 1(b), is transformed into a feature vector or a “fingerprint” using the smooth overlap of atomic positions (SOAP) method [28,29]. The feature vector size is determined by the SOAP hyperparameters; here, we use 10 radial basis functions, 10 degrees of spherical harmonics, and 0.5 \AA smearing width for the Gaussian functions in SOAP, which results in a vector length of 606 features. Thus, the initial input feature matrix is $[N^{\text{GB}} \times 606]$ in size, as illustrated in Fig. 1(c). Using this matrix as input to the learning algorithm would require the fitting of hundreds of coefficients (even if simple linear regression were used), which would require thousands of training data points. This is far too expensive for QM calculations, especially for the purpose of screening. Thus, to simplify the learning problem, we use principal component analysis (PCA) [30,31] to reduce the dimensionality of the initial feature matrix by projecting it into a low-dimensional representation, i.e., principal components that maximize the variance of the data. As the first 10 principal components capture more than 99% of the variance in the data [Fig. 1(d)], we use them as input features $[N^{\text{GB}} \times 10]$ in size as shown in

Fig. 1(e)] to the learning algorithm instead of the initial SOAP matrix ($[N^{\text{GB}} \times 606]$ in size). The reduced PCA features reduce the learning problem to a great extent, and based on our previous work in Ref. [13], are expected to perform well (in a learning model) with errors on the order of ~ 5 kJ/mol for a spectrum that spans ~ 100 kJ/mol.

The next step in the model is the selection of training data points, whose number is determined by the complexity of the learning algorithm. In our framework, we use linear regression as the learning algorithm, and thus as a rule of thumb, approximately 100 data points are sufficient [43] to fit the 10 coefficients and the intercept. To choose the training data points, we use k -means clustering [32,33] to divide the GB space into 100 clusters based on the proximity of local atomic environments in PCA feature space. Then, for each cluster, we choose the GB site that is closest (i.e., shortest Euclidean distance) to its centroid, and use this site as the representative training data point, as shown in Fig. 1(f). (The use of k -means clustering, as opposed to random sampling, ensures better coverage of the full GB feature space.) The end result of this unsupervised learning component of the framework [Figs. 1(a)–1(f)] is the choice of 100 optimal GB sites for which we need to calculate the solute segregation energy to train the regression algorithm [Fig. 1(I)].

Second, the hybrid QM-MM scheme, which we use to compute the segregation energy for the 100 training data points. This type of hybrid scheme has been used in the literature, though in a limited capacity, to model complex processes for large metallic systems [44,45] inaccessible to QM methods that are typically limited to $O(100)$ atoms (see Ref. [46] for a detailed review); for example, dislocation motion [47–49], and solute interactions with dislocations [50], and grain boundaries [51,52].

A typical setup for the QM-MM method [46] is shown in Figs. 1(g) and 1(h). We use QM to treat the local atomic environment of the GB segregation site, which we call the “core” region [the yellow region in Fig. 1(g)], and use MM to treat the long-range elastic interactions in the rest of the polycrystal. To do the QM calculation, we carve out a cluster that contains the core region, and a “buffer” surrounding region [the green region in Fig. 1(g)], and feed it to QM, after adding a vacuum layer of 10 Å in all directions to remove periodic image effects in the supercell [Fig. 1(h)]. The buffer region facilitates the seamless back-and-forth coupling between QM and MM. It should be large enough to minimize any changes in the electronic structure of the core region atoms induced by the introduced free surface [outer area of the buffer region Fig. 1(h)] in the QM supercell calculation. The implementation details for the hybrid QM-MM scheme used here are described in Supplemental Material, Sec. 4 [12].

The main goal for the QM-MM procedure is to compute similar GB solute segregation energies as if the whole system were treated by QM. To test against this goal, we

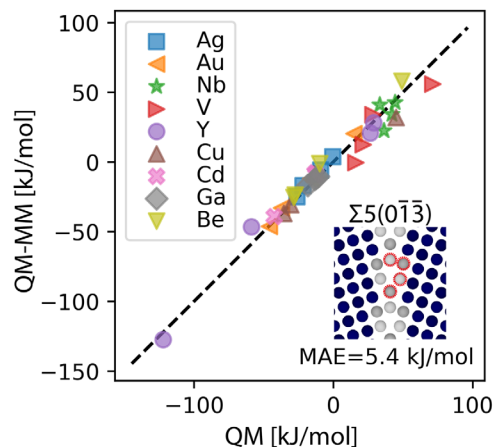


FIG. 2. Comparison between QM-MM and full QM computed segregation energies for Ag, Au, Nb, V, Y, Cu, Cd, Ga, and Be in an aluminum $\Sigma 5(0\bar{1}\bar{3})$ GB, which has four unique sites highlighted by dashed red circles.

compare the segregation energies obtained by QM-MM [Supplemental Material, Eq. (S.3)] against full QM calculations in Fig. 2. Because full QM is not possible for most of the polycrystalline environments of interest, for this validation we are limited to special, symmetric sites. For nine different solute elements that sample the alloy space ($>20\%$ of all solutes studied), we explore the $\Sigma 5(0\bar{1}\bar{3})$ GB, which has four unique GB site types (highlighted with red-dashed circles) as shown in Fig. 2 (see Supplemental Material [12], Sec. 5 for more details [53] on the computational setup). Overall, as shown in Fig. 2, the QM-MM approach results in low errors in segregation energies, with the mean absolute error being less than 6 kJ/mol on a problem with a scale spanning ~ 200 kJ/mol. Additionally, the method correctly captures the variation in segregation energies across both the alloy and GB spaces for this narrowly defined test problem. We therefore expect a similar performance level (error on the order of ~ 5 kJ/mol) for the QM-MM scheme for the polycrystal.

In general, our *ab initio* based learning approach is easy to adapt and fine-tune. Errors from the different components can be systemically reduced. For the ML component, lower errors can be obtained by using: more data points for training; more features and/or a larger radius cutoff to describe the local atomic environment; or a more sophisticated learning algorithm. And similarly, errors for the QM-MM component can be reduced by using: a larger buffer region; a more accurate functional and/or basis set for the density functional theory calculations; or a higher-level quantum chemical simulation method. Such modifications, however, will come at a higher computational cost, and thus should be reserved for postscreening efforts, e.g., adding more training data points to alloys of high interest. However, for screening, we find the parameters defined

herein and in Supplemental Material [12] for the learning and QM-MM approach to be a good balance between accuracy and speed (cost). We also note that our framework provides a pathway to tackle, with QM accuracy, the more complex problems of nondilute [54–56] and higher-order alloys [57–59] (i.e., ternary and higher), which, in addition to understanding the spectrum of segregation energies, requires a comprehensive understanding of the magnitude of solute-solute interactions in the system. We expect this to be a major direction for future studies.

In Fig. 3, we plot the learned GB segregation spectra for all 39 solutes in an aluminum polycrystal. These spectra capture the complete range of possible segregation states with near quantum-level accuracy, and are widely variable for different alloying elements. As an illustration of this, each panel in Fig. 3 is colored by the magnitude of the specific solute excess, $\beta^{\text{GB}} = [\bar{X}^{\text{GB}}/(1 - \bar{X}^{\text{GB}})] [(1 - X^c)/X^c]$, that is expected at a total solute concentration of 5% and a temperature of 700 K [where \bar{X}^{GB} , X^c are the equilibrium solute concentration at the GB and bulk (intragrain) regions, respectively; see Supplemental Material, Sec. 6 [12] for more details]. For other conditions (concentration, temperature), the spectra in Fig. 3 can be fitted with the following skew-normal distribution function:

$$F_i^{\text{GB}}(\Delta E_i^{\text{seg}}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\Delta E_i^{\text{seg}} - \mu)^2}{2\sigma^2}\right] \times \text{erfc}\left[-\frac{\alpha(\Delta E_i^{\text{seg}} - \mu)}{\sqrt{2}\sigma}\right], \quad (1)$$

which requires only three fitted parameters—the location μ , the scale σ , and the shape (skewness) α . The fitted

parameters for each solute in aluminum are shown in Fig. 3, which should permit accurate thermodynamic calculations of GB segregation for future works on aluminum alloys, using the continuous form of the spectral segregation isotherm [11]:

$$X^{\text{tot}} = (1 - f^{\text{GB}})X^c + f^{\text{GB}} \int_{-\infty}^{\infty} F_i^{\text{GB}}(\Delta E_i^{\text{seg}}) \times \left[1 + \frac{1 - X^c}{X^c} \exp\left(\frac{\Delta E_i^{\text{seg}}}{k_B T}\right)\right]^{-1} d(\Delta E_i^{\text{seg}}), \quad (2)$$

where X^{tot} is the total solute concentration in the polycrystal, f^{GB} is the GB atomic site fraction, k_B is Boltzmann’s constant, T is the temperature.

It is scientifically important that all solute segregation spectra are well approximated by the skew-normal distribution function (as shown in Fig. 3 by the solid black lines). This result matches earlier observations based on calculations with interatomic potentials [13] but extends the utility of Eq. (1) to DFT-computed segregation energies. In fact, among the 39 solutes represented in Fig. 3, only 16 have available potentials in the NIST Interatomic Potentials Repository [60,61]. The present result thus offers a pathway not only to higher physical accuracy in the calibration of Eq. (1), but also a means of covering the vast gaps in the alloy space not addressed by existing potentials. In fact, the present approach may be preferable to using existing interatomic potentials to study GB solute segregation. Consider the case of Al(Cu), which is a well-studied alloy with multiple available interatomic potentials. In Fig. 4(a), we plot the segregation spectra for Al(Cu) for five different interatomic potentials [62–66], alongside the QM-MM

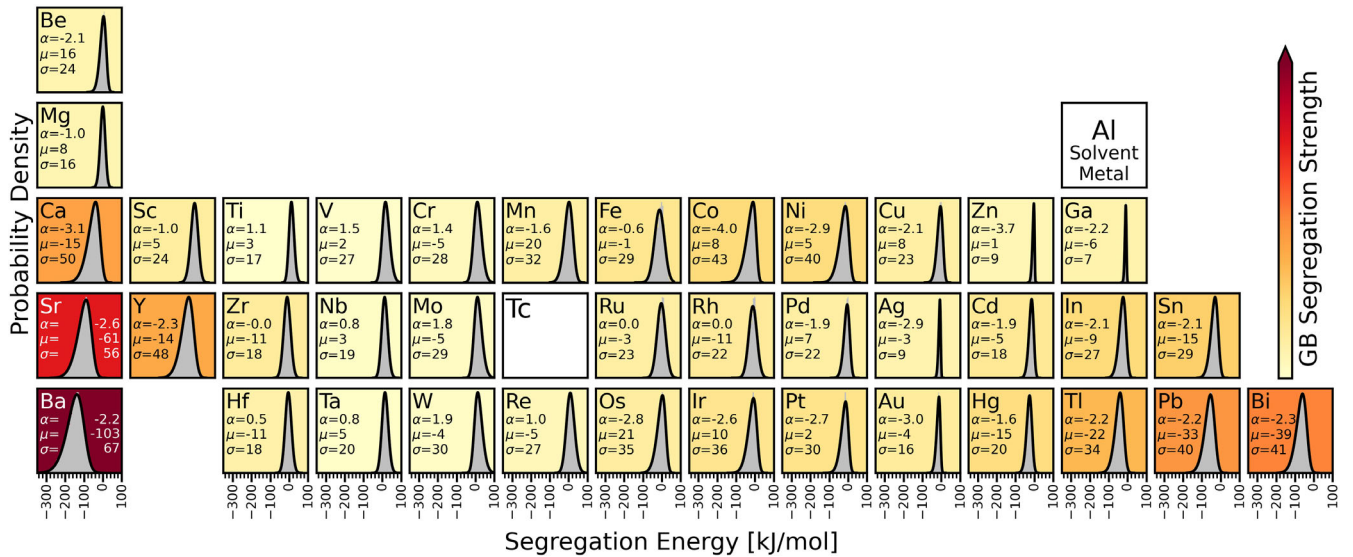


FIG. 3. The QM-MM learned GB solute segregation spectra for 39 solutes in the $(20 \text{ nm})^3$ aluminum polycrystal shown in Fig. 1(a). The segregation strength, $\ln(\beta^{\text{GB}})$, is quantified by solving for the equilibrium segregation state for $X^{\text{tot}} = 5\%$ and $T = 700 \text{ K}$. We fit the spectra to the skew-normal function (solid black line), Eq. (1), and list the values for its three parameters, μ (kJ/mol), σ (kJ/mol), and α .

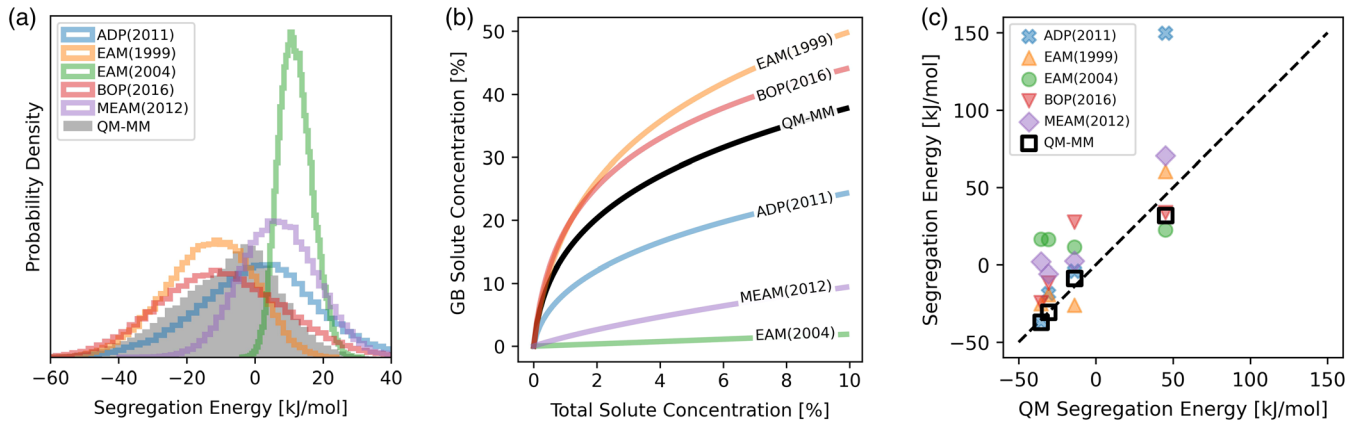


FIG. 4. (a) Segregation spectra for Cu solute in an aluminum polycrystal obtained using five different interatomic potentials [62–66] from Ref. [13] against the QM-MM spectrum computed here. (b) Predictions for \bar{X}^{GB} of Cu segregation in an aluminum polycrystal of 100 nm grain size at $T = 700$ K for the six different spectra in (a). (c) A comparison of computed segregation energies of Cu in an aluminum $\Sigma 5(0\bar{1}\bar{3})$ GB for the five interatomic potentials and QM-MM against full QM simulations.

learned spectra from this work. It is clear that the spectra are highly variable, and as a result, the expected solute segregation behavior is essentially not accurately predictable for any of the potentials; compare the solute content at 100 nm grain size and 700 K for the potentials (colored lines) with the quantum-based result from the present model (black line) in Fig. 4(b). Two of the potentials [65,66] give predictions for \bar{X}^{GB} that are an order of magnitude different from QM-MM. And it is indeed the QM-MM approach that most closely matches fully electronic results, as shown in Fig. 4(c) for four unique sites of the $\Sigma 5(0\bar{1}\bar{3})$ GB. This result highlights that even when interatomic potentials exist, they may not produce accurate results without extensive validation and testing against QM (assuming that QM is the ground truth); in light of this, the present combined learning and QM-MM approach may be an ideal balance of speed and quantum accuracy for any system, even if potentials already exist.

In conclusion, we have developed a flexible and easy-to-adapt *ab initio*-based framework that can learn the spectrum of solute segregation energies in a polycrystal, without the need for alloy interatomic potentials. To illustrate its power, we used the framework to compute the GB segregation spectra for a large swath of the aluminum-based alloy space (a total of 39 solutes)—a task currently infeasible using interatomic potentials. The proposed framework provides researchers with a simple but powerful tool to fully understand the variation of solute segregation across both the polycrystalline GB site space, as well as the binary alloy space. The combination of rapid computability with quantum-level accuracy for all realistic GB segregation sites in a polycrystalline environment represents a significant shift in the scope of GB segregation modeling toward genuine physical correctness.

This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences under Grant

No. DE-SC0020180. Work on entropy and grain size effects is supported by the U.S. National Science Foundation under Grant No. DMR-2002860.

*Corresponding author.
schuh@mit.edu

- [1] D. McLean, *Grain Boundaries in Metals* (Clarendon Press, Oxford, 1957).
- [2] M. P. Seah, Grain boundary segregation, *J. Phys. F* **10**, 1043 (1980).
- [3] P. Lejček and S. Hofmann, Thermodynamics and structural aspects of grain boundary segregation, *Crit. Rev. Solid State Mater. Sci.* **20**, 1 (1995).
- [4] P. Lejček, *Grain Boundary Segregation in Metals* (Springer, Berlin, Heidelberg, 2010), Vol. 136.
- [5] A. P. Sutton and R. W. Balluffi, *Interfaces in Crystalline Materials* (Clarendon Press, Oxford, 1995).
- [6] D. Raabe, M. Herbig, S. Sandlöbes, Y. Li, D. Tytco, M. Kuzmina, D. Ponge, and P.-P. Choi, Grain boundary segregation engineering in metallic alloys: A pathway to the design of interfaces, *Curr. Opin. Solid State Mater. Sci.* **18**, 253 (2014).
- [7] D. N. Seidman, Subnanoscale studies of segregation at grain boundaries: Simulations and experiments, *Annu. Rev. Mater. Res.* **32**, 235 (2002).
- [8] C. L. White and W. A. Coghlan, The spectrum of binding energies approach to grain boundary segregation, *Metall. Trans. A* **8**, 1403 (1977).
- [9] C. L. White and D. F. Stein, Sulfur segregation to grain boundaries in Ni_3Al and $\text{Ni}_3(\text{Al}, \text{Ti})$ alloys, *Metall. Trans. A* **9**, 13 (1978).
- [10] R. Kirchheim, Hydrogen solubility and diffusivity in defective and amorphous metals, *Prog. Mater. Sci.* **32**, 261 (1988).
- [11] M. Wagih and C. A. Schuh, Spectrum of grain boundary segregation energies in a polycrystal, *Acta Mater.* **181**, 228 (2019).
- [12] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.129.046102> for further

- details on the implementation of the combined learning and hybrid QM-MM framework.
- [13] M. Wagih, P. M. Larsen, and C. A. Schuh, Learning grain boundary segregation energy spectra in polycrystals, *Nat. Commun.* **11**, 6376 (2020).
- [14] Z. Pan, V. Borovikov, M. I. Mendeleev, and F. Sansoz, Development of a semi-empirical potential for simulation of Ni solute segregation into grain boundaries in Ag, *Model. Simul. Mater. Sci. Eng.* **26**, 075004 (2018).
- [15] V. L. Deringer, M. A. Caro, and G. Csányi, Machine learning interatomic potentials as emerging tools for materials science, *Adv. Mater.* **31**, 1902765 (2019).
- [16] Y. Mishin, Machine-learning interatomic potentials for materials science, *Acta Mater.* **214**, 116980 (2021).
- [17] M. A. Wood, M. A. Cusentino, B. D. Wirth, and A. P. Thompson, Data-driven material models for atomistic simulation, *Phys. Rev. B* **99**, 184305 (2019).
- [18] X. G. Li, C. Hu, C. Chen, Z. Deng, J. Luo, and S. P. Ong, Quantum-accurate spectral neighbor analysis potential models for Ni-Mo binary alloys and fcc metals, *Phys. Rev. B* **98**, 094104 (2018).
- [19] J. Vandermause, S. B. Torrisi, S. Batzner, Y. Xie, L. Sun, A. M. Kolpak, and B. Kozinsky, On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events, *npj Comput. Mater.* **6**, 20 (2020).
- [20] C. W. Rosenbrock, K. Gubaev, A. V. Shapeev, L. B. Pártay, N. Bernstein, G. Csányi, and G. L. W. Hart, Machine-learned interatomic potentials for alloys and alloy phase diagrams, *npj Comput. Mater.* **7**, 24 (2021).
- [21] M. I. Mendeleev, M. J. Kramer, C. A. Becker, and M. Asta, Analysis of semi-empirical interatomic potentials appropriate for simulation of crystalline and liquid Al and Cu, *Philos. Mag.* **88**, 1723 (2008).
- [22] N. Turchinda and C. A. Schuh, Grain size dependencies of intergranular solute segregation in nanocrystalline materials, *Acta Mater.* **226**, 117614 (2022).
- [23] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, *J. Comput. Phys.* **117**, 1 (1995).
- [24] W. M. Brown, A. Kohlmeyer, S. J. Plimpton, and A. N. Tharrington, Implementing molecular dynamics on hybrid high performance computers—Particle-particle particle-mesh, *Comput. Phys. Commun.* **183**, 449 (2012).
- [25] P. Hirel, ATOMSK: A tool for manipulating and converting atomic data files, *Comput. Phys. Commun.* **197**, 212 (2015).
- [26] A. Stukowski, Visualization and analysis of atomistic simulation data with OVITO—The open visualization tool, *Model. Simul. Mater. Sci. Eng.* **18**, 015012 (2010).
- [27] A. Stukowski, Structure identification methods for atomistic simulations of crystalline materials, *Model. Simul. Mater. Sci. Eng.* **20**, 045021 (2012).
- [28] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons, *Phys. Rev. Lett.* **104**, 136403 (2010).
- [29] A. P. Bartók, R. Kondor, and G. Csányi, On representing chemical environments, *Phys. Rev. B* **87**, 184115 (2013).
- [30] M. E. Tipping and C. M. Bishop, Probabilistic principal component analysis, *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61**, 611 (1999).
- [31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, SCIKIT-LEARN: Machine learning in PYTHON, *J. Mach. Learn. Res.* **12**, 2825 (2011), <https://www.jmlr.org/papers/v12/pedregosa11a.html>.
- [32] S. P. Lloyd, Least squares quantization in PCM, *IEEE Trans. Inf. Theory* **28**, 129 (1982).
- [33] C. Elkan, Using the triangle inequality to accelerate K-means, in *Proceedings of the Twentieth International Conference on Machine Learning, Washington, DC* (AAAI, Palo Alto, California, 2003), Vol. 1, pp. 147–153.
- [34] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, Real-space grid implementation of the projector augmented wave method, *Phys. Rev. B* **71**, 035109 (2005).
- [35] J. Enkovaara *et al.*, Electronic structure calculations with GPAW: A real-space implementation of the projector-wave method, *J. Phys. Condens. Matter* **22**, 253202 (2010).
- [36] A. H. Larsen, M. Vanin, J. J. Mortensen, K. S. Thygesen, and K. W. Jacobsen, Localized atomic basis set in the projector augmented wave method, *Phys. Rev. B* **80**, 195112 (2009).
- [37] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [38] N. Marzari, D. Vanderbilt, A. De Vita, and M. C. Payne, Thermal Contraction and Disordering of the Al(110) Surface, *Phys. Rev. Lett.* **82**, 3296 (1999).
- [39] S. R. Bahn and K. W. Jacobsen, An object-oriented scripting interface to a legacy electronic structure code, *Comput. Sci. Eng.* **4**, 56 (2002).
- [40] A. Hjorth Larsen *et al.*, The atomic simulation environment—A PYTHON library for working with atoms, *J. Phys. Condens. Matter* **29**, 273002 (2017).
- [41] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch, Structural Relaxation Made Simple, *Phys. Rev. Lett.* **97**, 170201 (2006).
- [42] N. Govind, Y. A. Wang, A. J. R. Da Silva, and E. A. Carter, Accurate ab initio energetics of extended systems via explicit correlation embedded in a density functional environment, *Chem. Phys. Lett.* **295**, 129 (1998).
- [43] F. E. Harrell, Jr., *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis* (Springer, New York, 2015).
- [44] N. Choly, G. Lu, E. Weinan, and E. Kaxiras, Multiscale simulations in simple metals: A density-functional-based methodology, *Phys. Rev. B* **71**, 094101 (2005).
- [45] J. Q. Broughton, F. F. Abraham, N. Bernstein, and E. Kaxiras, Concurrent coupling of length scales: Methodology and application, *Phys. Rev. B* **60**, 2391 (1999).
- [46] N. Bernstein, J. R. Kermode, and G. Csányi, Hybrid atomistic simulation methods for materials systems, *Rep. Prog. Phys.* **72**, 026501 (2009).
- [47] F. Bianchini, A. Glielmo, J. R. Kermode, and A. De Vita, Enabling QM-accurate simulation of dislocation motion in γ -Ni and α -Fe using a hybrid multiscale approach, *Phys. Rev. Mater.* **3**, 043605 (2019).
- [48] T. D. Swinburne and J. R. Kermode, Computing energy barriers for rare events from hybrid quantum/classical

- simulations through the virtual work principle, *Phys. Rev. B* **96**, 144102 (2017).
- [49] Y. Liu, G. Lu, Z. Chen, and N. Kioussis, An improved QM/MM approach for metals, *Model. Simul. Mater. Sci. Eng.* **15**, 275 (2007).
- [50] P. Grigorev, T. D. Swinburne, and J. R. Kermode, Hybrid quantum/classical study of hydrogen-decorated screw dislocations in tungsten: Ultrafast pipe diffusion, core reconstruction, and effects on glide mechanism, *Phys. Rev. Mater.* **4**, 023601 (2020).
- [51] L. Huber, B. Grabowski, M. Militzer, J. Neugebauer, and J. Rottler, Ab initio modelling of solute segregation energies to a general grain boundary, *Acta Mater.* **132**, 138 (2017).
- [52] L. Huber, B. Grabowski, M. Militzer, J. Neugebauer, and J. Rottler, A QM/MM approach for low-symmetry defects in metals, *Comput. Mater. Sci.* **118**, 259 (2016).
- [53] H. Zheng, X. G. Li, R. Tran, C. Chen, M. Horton, D. Winston, K. A. Persson, and S. P. Ong, Grain boundary properties of elemental metals, *Acta Mater.* **186**, 40 (2020).
- [54] M. Wagih and C. A. Schuh, Grain boundary segregation beyond the dilute limit: Separating the two contributions of site spectrality and solute interactions, *Acta Mater.* **199**, 63 (2020).
- [55] T. P. Matson and C. A. Schuh, Atomistic assessment of solute-solute interactions during grain boundary segregation, *Nanomaterials* **11**, 2360 (2021).
- [56] D. Scheiber, L. Romaner, R. Pippan, and P. Puschnig, Impact of solute-solute interactions on grain boundary segregation and cohesion in molybdenum, *Phys. Rev. Mater.* **2**, 093609 (2018).
- [57] H. L. Mai, X.-Y. Cui, D. Scheiber, L. Romaner, and S. P. Ringer, The segregation of transition metals to iron grain boundaries and their effects on cohesion, *Acta Mater.* **231**, 117902 (2022).
- [58] W. Xing, A. R. Kalidindi, D. Amram, and C. A. Schuh, Solute interaction effects on grain boundary segregation in ternary alloys, *Acta Mater.* **161**, 285 (2018).
- [59] M. Guttman, Equilibrium segregation in a ternary solution: A model for temper embrittlement, *Surf. Sci.* **53**, 213 (1975).
- [60] C. A. Becker, F. Tavazza, Z. T. Trautt, and R. A. Buarque De Macedo, Considerations for choosing and using force fields and interatomic potentials in materials science and engineering, *Curr. Opin. Solid State Mater. Sci.* **17**, 277 (2013).
- [61] L. M. Hale, Z. T. Trautt, and C. A. Becker, Evaluating variability with atomistic simulations: The effect of potential and calculation methodology on the modeling of lattice and elastic constants, *Model. Simul. Mater. Sci. Eng.* **26**, 055003 (2018).
- [62] X. W. Zhou, D. K. Ward, and M. E. Foster, An analytical bond-order potential for the aluminum copper binary system, *J. Alloys Compd.* **680**, 752 (2016).
- [63] F. Apostol and Y. Mishin, Interatomic potential for the Al-Cu system, *Phys. Rev. B* **83**, 054116 (2011).
- [64] X. Y. Liu, C. L. Liu, and L. J. Borucki, New investigation of copper's role in enhancing Al-Cu interconnect electromigration resistance from an atomistic view, *Acta Mater.* **47**, 3227 (1999).
- [65] B. Jelinek, S. Groh, M. F. Horstemeyer, J. Houze, S. G. Kim, G. J. Wagner, A. Moitra, and M. I. Baskes, Modified embedded atom method potential for Al, Si, Mg, Cu, and Fe alloys, *Phys. Rev. B* **85**, 245102 (2012).
- [66] X. W. Zhou, R. A. Johnson, and H. N. G. Wadley, Misfit-energy-increasing dislocations in vapor-deposited CoFe/NiFe multilayers, *Phys. Rev. B* **69**, 144113 (2004).