ARTICLE   OPEN

A machine learning approach to model solute grain boundary segregation

Liam Huber[1], Raheleh Hadian[1], Blazej Grabowski[1] and Jörg Neugebauer[1]

Even minute amounts of one solute atom per one million bulk atoms may give rise to qualitative changes in the mechanical response and fracture resistance of modern structural materials. These changes are commonly related to enrichment by several orders of magnitude of the solutes at structural defects in the host lattice. The underlying concept—segregation—is thus fundamental in materials science. To include it in modern strategies of materials design, accurate and realistic computational modelling tools are necessary. However, the enormous number of defect configurations as well as sites solutes can occupy requires models which rely on severe approximations. In the present study we combine a high-throughput study containing more than 1 million data points with machine learning to derive a computationally highly efficient framework which opens the opportunity to model this important mechanism on a routine basis.

npj Computational Materials (2018)4:64 ; doi:10.1038/s41524-018-0122-7

INTODUCTION

A metallic alloy's strength is typically inversely correlated with the size of its crystalline grains.[1] Designing strong materials by ensuring a nanograined structure is therefore highly desirable. However, the excess energy stored in the grain boundaries (GBs) drives the system towards larger and fewer grains. Solute atoms can stabilize finely grained structures in two principal ways: kinetically, by providing a dragging force as the boundary migrates during grain growth; and thermodynamically, by lowering the interfacial energy.[2] The latter effect has received significant attention recently for its important role in producing nanograined materials by ball milling.[3–6]

Early efforts to model changes in GB interfacial energy due to solute segregation had to rely on severe approximations since neither the atomic structure at GBs nor the segregation energies of solutes to possible GB sites were known. One such approximation was to ignore GB details altogether and describe the interaction by a single per-solute value.[7–9] Under this single interaction value assumption, the solute enrichment at the boundary can be analytically described by the Langmuir–McLean (LM) segregation isotherm.[8] This model predicts a sigmoidal variation with temperature as the boundary transitions from undecorated to saturated with solute.

Since then, advances in experiment and modelling have clearly shown that the concept of describing all solute-GB interaction with a single effective energy is not justified.[10] For instance, transmission electron microscopy reveals a preferential segregation of specific solutes to particular sites, e.g., in ref. [11] The need to go beyond a single interaction parameter was also realized early on in theory: White and Coghlan (WC)[12] introduced a model in 1977 that treats segregation to each site individually. Unlike LM, this model allows the description of solute-GB interaction by a *distribution* of segregation energies rather than by a single value and does not require an explicit saturation limit.

A practical limitation in going beyond the LM model is that calculating segregation energy distributions is challenging. Even today only very few studies explicitly consider these distributions,[13–17] all of which use empirical potentials. Quantum mechanical calculations of GB segregation, being computationally much more expensive, are even more restricted and only allow the computation of high-symmetry GBs with few unique sites. Typically, these QM studies focus their analysis only on the most attractive site per boundary. Consequently, comparatively few per-solute-per-site investigations have been reported.[16,18–24]

To overcome these limitations we derive a computational framework that describes the full distribution of solute energies at GBs with modest computational effort. The framework is based on a two-step approach: First we perform a high-throughput study of six solute species segregating to thousands of sites at 38 low and high-symmetry boundaries. Using this large data set of more than a million segregation energies we show that the LM model is not even qualitatively able to reproduce the adsorption isotherm obtained by using the full distribution of energies. Secondly, we identify a set of machine learning descriptors which depend only on local properties of the solute-free GB. These descriptors allow an accurate prediction of the segregation energy distribution, and thus the segregation isotherm, using information available solely from the solute-free GB. This provides the opportunity to investigate solute segregation at arbitrary GBs on a routine basis.

RESULTS AND DISCUSSION

To study GB segregation we select aluminum as the host material. Empirical potentials are available for aluminum which allow for a variety of possible segregating elements. We construct a representative set of symmetry-inequivalent 53.1°[001] Σ5 GBs (i.e., with a *disorientation*[25] of 36.9°) following the concept of "fundamental zones" as introduced by Patala and Schuh[26,27] and Homer et al.[28] While a subset of these boundaries with higher
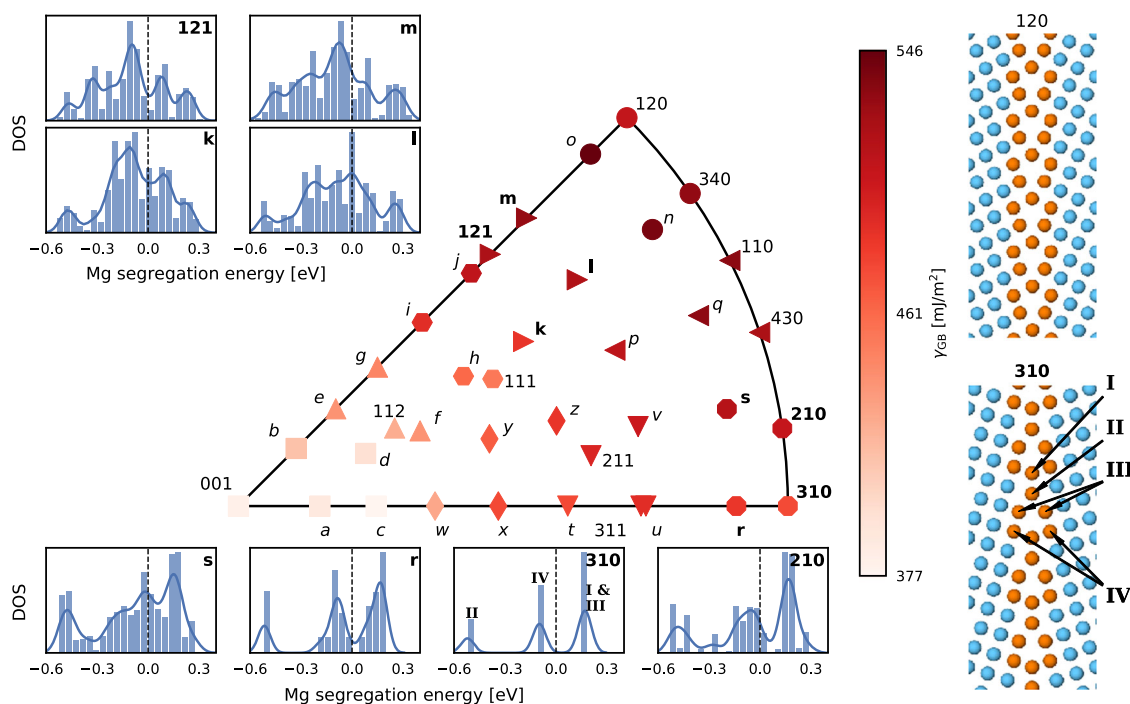
Fig. 1 Grain boundaries, their energies, and segregation energy distributions. The polar plot in the middle shows the GB interfacial energies for 38 different boundaries in the boundary-plane fundamental zone for pure Al 53.1°[001] boundaries with different normal vectors (see text). The energy distributions when substituting a Mg atom at the various GB sites are shown for selected boundaries below and above the polar plot. The blue bars are the data from a binning analysis. The blue curve shows the smoothed distribution. The black vertical dashed line in these plots marks the energy where a Mg atom has the identical energy as a Mg atom substituted in bulk. Thus, all energies left of this line represent sites attractive (exothermic) for segregation while the ones to the right represent sites which are not attractive. To the right of the polar plot, the atomic structure for two high-symmetry tilt boundaries projected on the [001] plane is shown. GB sites as identified by Common Neighbour Analysis (CNA) are shown in orange. For the (310) boundary the four inequivalent sites are labelled from I to IV

symmetry has been studied by Olmsted et al.,[25] we uniformly sample the space of boundary-plane normal vectors, (lmn), by using an approach proposed by Lee and Choi.[29] This approach represents the GB by a large vacuum cluster rather than by periodic boundary conditions, which would enforce geometric constraints.

The position of these GBs in the boundary-plane fundamental zone is shown in Fig. 1 where they are labelled from a to z. We also calculate another twelve boundaries with integer (lmn) values, which are labelled by their indices. The shapes of the symbols group boundaries with similar (lmn) values. Colour indicates interfacial energy, γ_{GB}. For all these GBs structural optimization has been performed using empirical potentials[30–34] together with the lammps simulation package.[35] Further computational details are given in the Methods section.

The computed grain boundary energies for Al follow a similar trend to those reported for Ni:[25,28] low interfacial energies for boundaries which are mostly twist (near (001)), and higher energies as we move towards pure tilt boundaries (the $(lm0)$ line). The highly symmetric (120) and (310) boundaries have exceptionally low energy compared to other nearby boundaries. The atomic geometries of these two high-symmetry boundaries are shown to the right of the polar plot in Fig. 1.

To investigate the interaction of the GBs with various solutes we compute the segregation energy of six solutes (Mg, Fe, Ti, Co, Ni, and Pb) to every substitutional site at each of these 38 GBs in the dilute limit. While the exact number of GB sites where substitutional solutes can be placed varies depending on the microscopic GB structure, most have roughly six thousand sites. Examples of the resulting segregation energy distributions are

shown in Fig. 1 for Mg using two of the groups in the boundary-plane fundamental zone.

While each boundary has thousands of GB sites, these sites are not always all unique. For example, the high-symmetry (310) boundary has only four symmetry-inequivalent sites (see Fig. 1). The sites on the GB plane (I and II) are singly degenerate, while sites III and IV exist both to the left and right of the plane and are thus two-fold degenerate. The degeneracy and energy of these sites are directly reflected in the energy distribution: The lowest (most attractive) peak corresponds to site II. The less but still attractive peak corresponds to site IV and is a factor of two higher due to the two-fold degeneracy of this site. Sites I and III are very close in energy (within 0.04 eV) and thus appear as a single peak with a height three times larger than the peak for site II.

The three peaks observed for the (310) boundary, are still clearly evident as we rotate the normal plane about the shared axis towards the asymmetric (210) tilt boundary or introduce a twist component towards r (see Fig. 1 bottom). The observed peak broadening is a direct result of introducing disorder into the local structures when moving away from a high symmetry GB. Broadening can become large even in the proximity of a symmetric GB: Comparing boundary s to (310) the main peaks are still evident but broadening leads to an approximately continuous distribution of segregation energies. A full set of grain boundary energies and site counts as well as solute-GB segregation energy distributions can be found in the Supplementary Material Table 1 and Fig. 1 through 11. They show very similar trends as discussed here for Mg.

With these segregation energy distributions, we can use WC model for non-interacting solutes to predict the GB solute
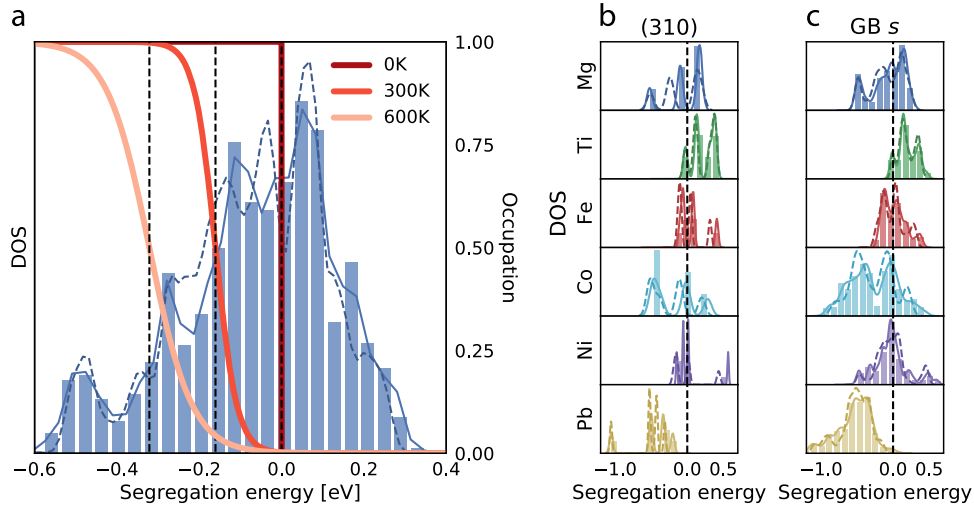
Fig. 2 Segregation DOS. **a** Segregation energy density of states shown for Mg (across all 38 boundaries with equal weight). Bars show a histogram of calculated data, while solid and dashed lines show smoothed distributions of calculated segregation energies and those predicted by machine learning, respectively. Using the same energy axis, red lines show the White-Coghlan site occupation value for three temperatures. Dashed black lines show the corresponding half-occupancy "Fermi level" at these temperatures with a bulk solute concentration of 0.2%at. **b, c** show the density of states for all solutes at the symmetric (310) and disordered *s* boundaries, respectively

concentration:

$$c_{GB} = \frac{1}{N} \sum_i \frac{1}{1 + \frac{1-c_{bulk}}{c_{bulk}} \exp\left(E_{seg}^{X@i}/k_B T\right)}, \tag{1}$$

where $E_{seg}^{X@i}$ is the segregation energy of the solute, *X*, to the *i*th of *N* sites at the GB, *T* is the temperature, and c_{bulk} is the concentration of solute in the bulk. The above equation has the same form as the traditional LM segregation isotherm except that it uses an energy distribution rather than a single energy, as indicated by the sum over sites, and it does not need an explicit saturation limit. Indeed, the LM model can be considered as a special case of the WC model where the true segregation energy distribution has been replaced by a single, "effective" value. Except for the $(1-c_{bulk})/c_{bulk}$ factor, this is also the same expression given by Fermi-Dirac statistics. This similarity is not surprising since both occupation curves are derived assuming an exclusion principle—in this case once a GB site is occupied it cannot be occupied by a second solute.

To study how temperature affects GB occupation we fix the bulk concentration at 0.2 at% and consider the segregation energy density of states (DOS) that is an equally weighted superposition of results from all 38 individual GBs. Figure 2a shows the resulting DOS and WC occupancy curves for Mg. At zero Kelvin all of the attractive sites are occupied. Unlike an electronic DOS, which contains all possible states, the GB DOS shown does not include the DOS from the bulk state. At elevated temperatures, the high configurational entropy available for the solute in the bulk makes this state increasingly appealing. This has the effect of shifting the solute chemical potential (Fermi level) down to more attractive energies, and is the source of the $(1-c_{bulk})/c_{bulk}$ deviation of Eq. (1) from Fermi-Dirac statistics. The individual DOS for the high-symmetry (310) and low-symmetry *s* boundaries are shown for all solutes in Figs. 2b, c, respectively.

By integrating over all occupied states we can directly compute the total amount of GB segregation predicted by the WC model. The resulting GB concentration as function of temperature is shown for Co in Fig. 3 (solid blue line) with the corresponding DOS shown in the inset. The GB concentrations are rescaled by the number of sites per unit area to have physical units.

Using the analogy of the segregation DOS to an electronic DOS allows us to exploit ideas originally developed by the electronic
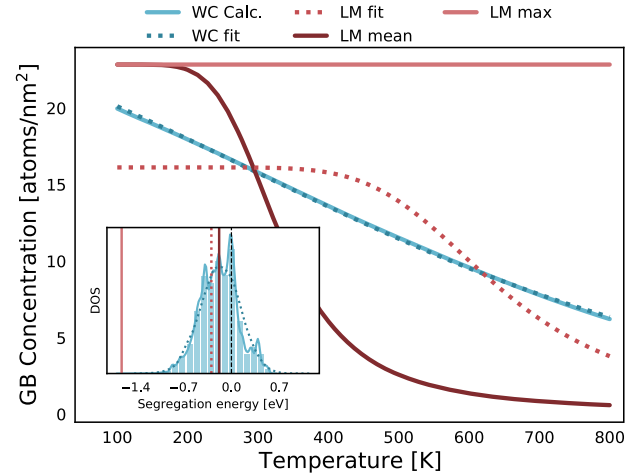

Fig. 3 GB solute concentration by several models. The WC prediction (Eq. (1)) for the Co concentration (solid blue) for the aggregated DOS (see text) using a bulk concentration of 0.2 at%. The LM prediction (DOS replaced by *δ*-function) using the calculated mean (solid dark red) or most attractive (solid light red) segregation energy. The dotted red line shows the LM model with the effective interaction energy and saturation limit used as fitting parameters. The dotted blue line shows a fit of the WC model based on a Gaussian DOS (see text). The inset shows the corresponding segregation energy DOS and uses the same colour and line scheme

structure community. One powerful concept to analyze and approximate the electronic DOS is moment expansion, with the *n*th order moment defined as:

$$M_n = \int f_{DOS}(E) E^n dE. \tag{2}$$

The first order moment, $n = 1$, replaces the full DOS by a single Dirac delta-function at the energy average over the entire DOS. Expanding up to this moment is equivalent to the LM model. As shown in Fig. 3, using this model gives qualitatively incorrect temperature dependence: this is true whether the mean or maximum segregation energy from our calculations is used (dark and light red lines, respectively) along with the true saturation limit, i.e., the fraction of attractive sites. Even using the LM model's
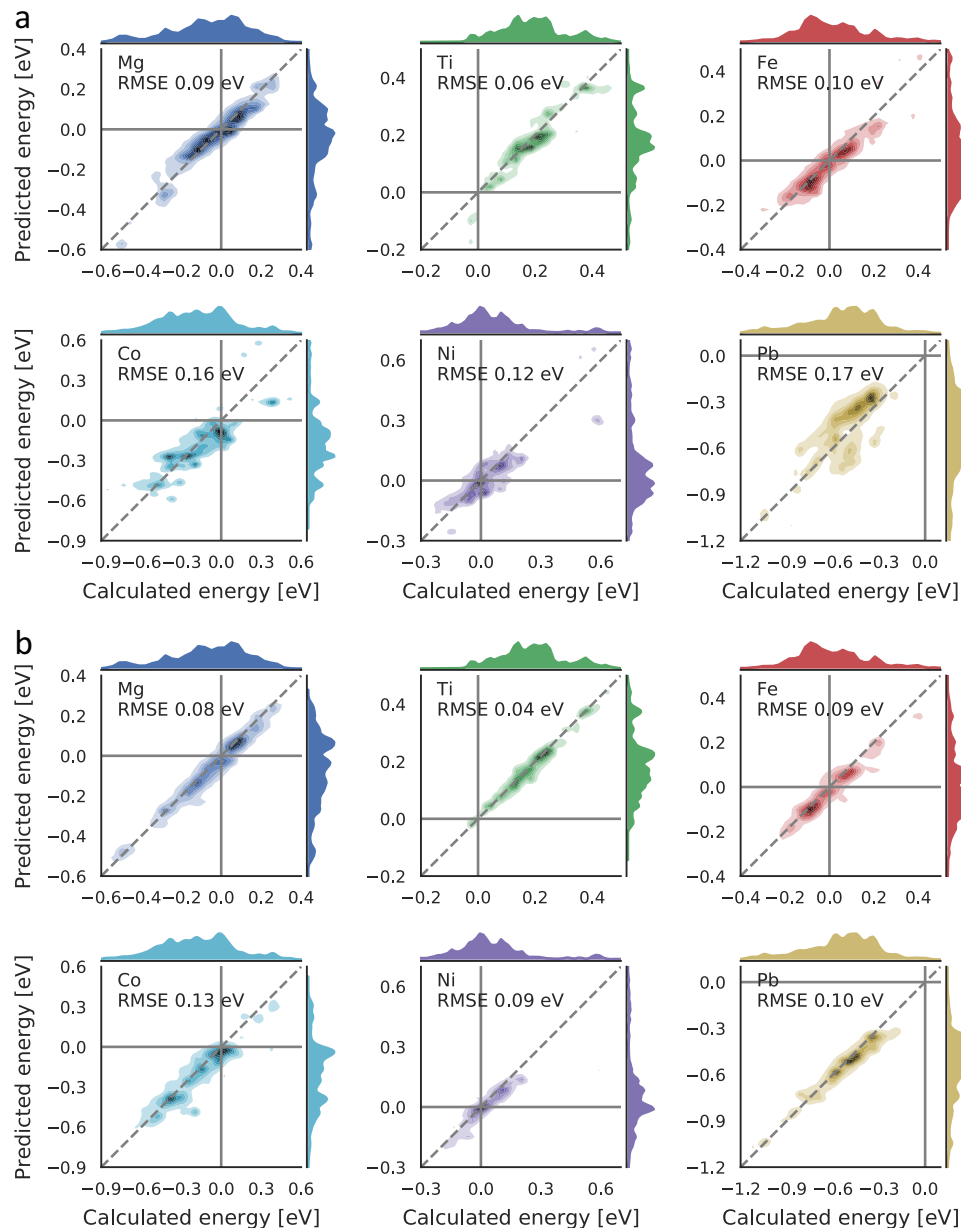
Fig. 4 Model performance. Predicted segregation energies using **a** a linear fit and **b** gradient-boosted decision trees plotted against calculated values for all six solutes. Results are taken from all 38 boundaries and data distribution plots are shown joint to their corresponding axes. Darkness of colour indicates data frequency, and results are smoothed. The black dashed line indicates perfect agreement and is a guide for the eye

constant segregation energy and saturation limit as free fitting parameters (dotted red line) results in unsatisfactory agreement. This indicates that the failure of the LM model is due to its unrealistic functional form and cannot be "cured" by choosing optimally fitted parameters. Thus, for an accurate description of the segregation isotherm it is crucial to go beyond the LM model.

Between the two extrema of the 1st-order LM expansion and the full WC model, there is a series of models using increasingly sophisticated DOS expansions. Including the second order moment, M_2, represents the DOS by a Gaussian. Despite the fact that the aggregated DOS is not strictly Gaussian, as can be seen in Figs. 2a, 3 (inset), and 4 (joint axes), this level of approximation works surprisingly well. Using the Gaussian normalization factor, mean, and standard deviation as fitting parameters to optimize the GB solute concentration, the resulting fit-Gaussian WC-model prediction (dotted dark blue line) in Fig. 3 agrees almost perfectly,

and the fit mean and variance match the first and second moments of the calculated DOS (c.f. Supplementary Material Table 4 for numeric values).

This fit-Gaussian procedure suggests a promising way to directly connect experimental segregation measurements to atomistic calculations. When concentration values are known for a variety of temperatures or bulk concentration (or both), an implicit minimization scheme can be applied to find the Gaussian approximation to the DOS which minimizes the WC model prediction error. When performed consistently for the same boundary, or over a broad enough sampling of boundaries to get a representative DOS—similar to the aggregate DOS in Fig. 2a—a corresponding atomistic calculation can be performed. Therefore, to connect atomistic simulations with experimental segregation isotherms it will be sufficient in most cases to go one order

beyond the LM model, i.e., to increase the number of fitting parameters from two to three.

Figure 3 clearly shows that a realistic description of the segregation isotherm can only be achieved if the distribution of the segregation energies is known. A practical challenge is that with presently available tools such calculations are difficult to perform on a routine basis. In the following we therefore explore alternative descriptions of the DOS which do not require calculations for each solute on every possible site in the GB. Specifically, we will analyze whether information that can be obtained from the *chemically pure* GB is sufficient. This would reduce the number of atomistic calculations from $\mathcal{O}$(number of GB sites) to $\mathcal{O}(1)$ per GB. To achieve this goal, we have to identify suitable descriptors from the GB that capture pertinent information for each possible segregation site and a relation that connects this data with the segregation energy. Our large set of more than a million segregation energies will allow us to address both aspects.

We start by applying a simple linear model[16] that connects two quantities which can be easily obtained for each site i on the grain boundary—excess volume,[22] ΔV_i, and the change in coordination number,[36] ΔN_i—with the segregation energy of a solute X:

$$E_{seg}^{X@i} = P_X \Delta V_i + E_X^{bond} \Delta N_i. \tag{3}$$

The solute-specific parameters P_X and E_X^{bond} both have simple physical interpretations as the solute pressure and the energy-per-bond, respectively. Here we use these as fitting parameters, and their signs indicate whether the solute prefers compressive/expansive sites, and whether it prefers to be under/over-bonded, respectively.

Conventional models for GB segregation, such as McLean's model based on the total relief of strain energy upon segregation, or Seah's model based on changes in bonding state, typically omit details of the GB altogether and function only on a per-solute basis (as in the models just mentioned) or provide analysis on a per-solute-per-*GB*, as in a handful of more recent works.[37–40] In contrast, we present here a fully atomistic model that has predictive capability since it relies on GB properties solely derived from the *undecorated* GB.

In order to estimate the predictive error of such a fit model, i.e., how well the model will perform when applied to a GB for which it was not fit, we employ cross validation. The data is structured into nine groups (indicated by different symbols in Fig. 1, see the "cross validation" section of the Supplementary Material for details) and for nine iterations one group is held out from the fitting procedure and predicted on. This grouping is an effort to best represent generalization to an arbitrary new boundary, since boundaries nearby in this space are likely to share more microscopic atomic configurations. Nevertheless, the resulting nine sets of P_X and E_X^{bond} are highly similar even though they were fitted to different data. The results are robust to using fewer but larger groups. Fit values of P_X and E_X^{bond} are in the Supplementary Material Table 3.

For each solute, an aggregation of these predictions across all 38 boundaries is plotted in Fig. 4a against the calculated segregation energies. The energy distributions for both the model and calculations are shown joint to their respective axes. For all six solutes, the simple linear model captures qualitative trends in segregation. The plot also shows root mean square errors (RMSE) which range from 0.06 (Ti) to 0.17 eV (Pb). To set this error in perspective we refer to a previous study.[16] There, QM/MM multiscale calculations show that the potentials for Mg and Pb reproduce quantum mechanical values for GB segregation energies with RMSE values of 0.09 and 0.16 eV, respectively.

Pb, for which the model performs worst, also has the largest size mismatch with Al among the solutes studied. The poor performance of the linear model may have been expected, since we are attempting to predict the Pb segregation energy only from

site properties of the *initial, undecorated* GB. In reality, a solute with such a large atomic radius shows non-linear elastic interactions, and also induces significant relaxations of the local environment.[16]

An advantage of the linear model is that its parameters can be interpreted physically. However, its functional simplicity limits predictive power. To improve accuracy, we test a second more sophisticated approach based on machine learning using gradient boosted decision trees.[41] In addition to per-site values of ΔV and ΔN, we also include as descriptors their second and third powers, which effectively permit the construction of a truncated Taylor series with respect to these variables. Going beyond a linear relation allows the model to infer that a solute may have preferred values of ΔV and ΔN, with energetic penalties to either side of the preferred value. Further, we include various properties of the Voronoi analysis[42] such as total surface area, the ratio of surface area to volume, total edge length, the number of faces, and the number of edges. We also use the Steinhardt bond-orientational order parameters[43] Q_1 through Q_8 with a cut-off half way between first and second-nearest neighbour distances. These descriptors all have in common that they are available from the undecorated GB structure.

Segregation energies predicted by this machine-learned model show a clear improvement over the linear model (see Fig. 4b). For each solute the error is decreased, although only slightly for Mg and Fe which were already well described. The largest improvement is for Pb whose error drops from 0.17 to 0.10 eV.

While decision trees do not allow the same level of direct interpretation that a linear model does, they still allow identification of the most relevant descriptors. For Pb, which has the largest atomic radius mismatch to Al, it is not surprising to see that powers of the excess volume, ΔV and ΔV^3, are the two most significant descriptors with the square also playing an important role. For all other investigated solutes the most important descriptor is the ratio between the undecorated site's surface area and volume. Not only does this descriptor encode information about the excess volume at a specific site, but, in conjunction with ΔV and/or the area, it also encodes information about the site's structural anisotropy. This piece of physics which was completely neglected in the simple linear model. Despite their usefulness in describing Re segregation to W boundaries,[23] here the Steinhardt parameters are always secondary to at least one other descriptor. More detailed information on the relative importance of the various features can be found in the Supplementary Material Fig. 12.

Figure 2a shows the smoothed aggregate DOS predictions (dashed darker lines) for Mg, while Figs. 2b, c show the DOS for the specific boundaries (310) and s, respectively. Overall, the DOS for these machine-learned segregation energy predictions agree reasonably well with the calculated DOS. This shows that these machine learned models capture the overall trends in segregation and also demonstrates the degree to which they reproduce the detailed behaviour of each boundary for each solute individually.

While the DOS is important, the experimentally measurable quantity is the GB solute concentration, c_{GB}. To evaluate the importance of the (small) differences between the calculated DOS and the DOS generated from our decision tree segregation energy predictions, Fig. 5 shows GB concentration curves for all six solutes. Across a wide temperature range, the predicted energies give GB concentrations that are very close to the reference with the explicitly computed segregation energies (dashed and solid lines, respectively). We also show fits of the LM (light dotted lines) and Gaussian-WC-model (dark dotted lines) concentrations. The first moment LM model shows systematic failure—except for the "null case" of Ti where no segregation is predicted—while the machine learned energies and Gaussian fits both perform well.

For some of these solutes the predicted GB concentration at room temperature is significant and the assumption of non-
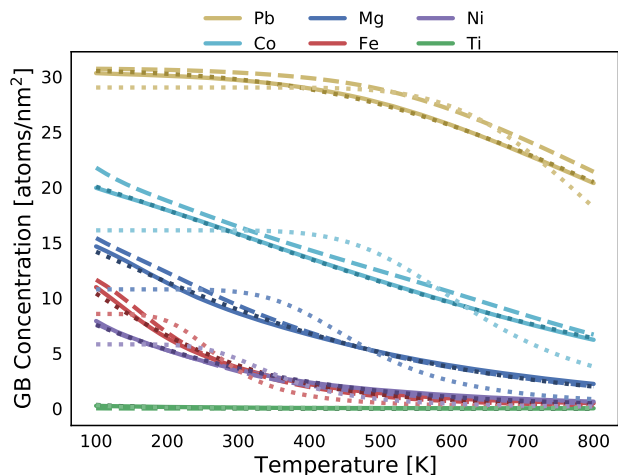
Fig. 5 GB enrichment. WC segregation isotherms using the calculated aggregate DOS (solid lines) and the aggregate DOS predicted by machine learning (dashed lines). Best fits of the LM model (lighter dotted lines) and WC model assuming a Gaussian DOS (darker dotted lines) are also shown. For all solutes the bulk solute concentration is 0.2 at%
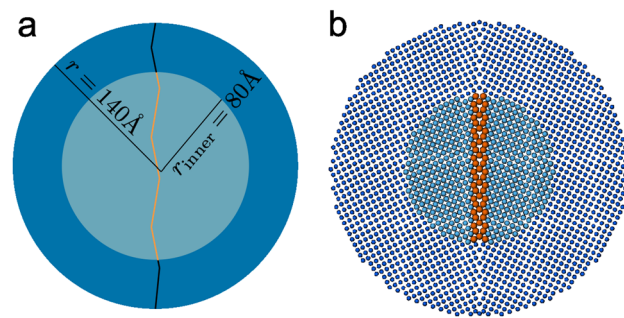

Fig. 6 Geometric setup. **a** A schematic representation of the cluster setup with the inner domain used for calculating GB energies and searching for GB sites shown in light blue, while the remainder of the cluster which serves as a buffer from the surface are dark blue. The portion of the GB which is eligible for segregation energy calculations is bright orange. **b** A slice from a small toy cluster of the 53.13°[001](120) boundary showing atomistic resolution of this scheme with the same colour scheme

interaction may break down. For example, ignoring displacement from the GB plane, a planar concentration of 15 atoms/nm^2 implies that the average inter-solute distance is approximately 2.8 Å—i.e., the solutes would be first-nearest-neighbours. Solute interaction may also occur at much lower concentrations when the attractive sites are not uniformly distributed across the GB, e.g., our calculations on Mg segregation to the symmetric (310) boundary indicate a preference for columnar segregation. Solute-solute interaction will thus be a natural extension of this study and another area where machine learning may prove useful.

In conclusion, based on an extensive set of over a million segregation energies for six solute species and 38 GBs we show that the standard model to describe grain boundary segregation —the Langmuir-McLean model—qualitatively fails in describing segregation isotherms. Using the mathematical similarity of the segregation and the electronic DOS we adopt concepts originally developed in electronic structure theory. Specifically, expanding the DOS in moments we show that the LM model corresponds to an expansion up to first order moments whereas for a qualitatively correct description of the segregation isotherms an expansion of at least second order is needed. Since the computational effort to routinely compute the full segregation DOS is prohibitive, we used a machine learning approach to identify a set of efficient descriptors. These descriptors are solely based on information that can be extracted from the solute-free grain boundary, i.e., rather than having to explicitly compute thousands of segregation energies for each GB, a single calculation is sufficient. This reduces the effort to compute segregation isotherms by up to four orders of magnitude, presenting the opportunity to routinely determine segregation isotherms for arbitrary GBs. Since the amount of solute concentration can be well controlled in metallurgy, having a tool to accurately determine the relation between bulk solute concentration, GB segregation, and GB energy provides the materials engineer a new set of design strategies. Since GBs are a key element in structural materials, many interesting extensions of this concept, e.g., to control and prevent GB related embrittlement, can be envisioned.

METHODS

To construct a simulation cell containing a GB with the character $\theta[ijk](lmn)$ —giving misorientation, the shared axis of misorientation, and one grain's normal plane, respectively—we follow the work of Lee and Choi.[29] Our

execution of this approach can be briefly summarized as follows: first, construct two spheres of radius r with the same crystal structure and orientation; second, rotate each by $\theta/2$ about $[ijk]$ in opposite directions; third, rotate both spheres in tandem until one of them has its internal $[lmn]$ direction pointing in the x-direction of our supercell coordinates; next, take the left half of one sphere and the right half of the other, and combine them to a single sphere surrounded by vacuum with a planar GB in the yz-plane of the supercell. Finally, minimize the microscopic degrees of freedom (DOF). The format $\theta[ijk](lmn)$ completely defines the GB character once one grain has been chosen to assign (lmn) to, but it is not unique.

From here, our approach expands upon ref. [29]. We introduce a second geometric parameter, a smaller radius r_{inner}, which defines a sub-sphere with the same centre. Since empirical potentials provide per-atom energies, we can easily calculate the energy of this sub-sphere separately from the whole system.[44] The GB interfacial energy is then

$$\gamma_{GB} = \left(\tilde{E}_{GB} - \frac{\tilde{N}_{GB}}{\tilde{N}_{bulk}} \tilde{E}_{bulk} \right) / \tilde{A}_{GB}, \quad (4)$$

where $\tilde{E}$ is the energy of these *inner spheres*. These spheres have $\tilde{N}$ atoms within r_{inner} for clusters with a GB or bulk-like structure as indicated by the subscript, and $\tilde{A}_{GB} = \pi r_{inner}^2$ is the sub-sphere's cross-sectional area.

When $r - r_{inner}$ is sufficiently large, the inner sphere is isolated from surface effects. For this case the sub-sphere represents a spherical sample out of an effectively infinite structure. For non-coincident site lattice (non-CSL) boundaries or CSL boundaries with low symmetry the GB structure repeat distance might be larger than $2r_{inner}$, but as long as this sub-sphere is large enough, we are still sampling a *representative* area of the GB.

For segregation energies, we use this inner sphere again to restrict our search for GB sites (of the substitutional variety, identified by their non-FCC common neighbour analysis[45,46] (CNA) signature) to ensure that they are also well isolated from surface effects. The segregation energy of a solute X to the ith site of a GB is

$$E_{seg}^{X@i} = \left(E_{GB}^{X@i} + E_{bulk} \right) - \left(E_{GB} + E_{bulk}^X \right), \quad (5)$$

where we have returned to using the total energy of the clusters, E, because of the good cancellation of surface energies in this formulation, and superscripts indicate the presence of a solute at the ith site of the GB-containing cluster, or in the centre of the bulk cluster. This formulation represents the infinitely dilute limit. Note that we use a convention with favourable segregation energies defined negative. The set-up is shown schematically in Fig. 6a. A cross-section of a small cluster with a symmetric (120)Σ5 boundary is shown in Fig. 6b visualized using the Ovito software.[47]

We found that $r = 140$ Å and $r_{inner} = 80$ Å were sufficient to give a GB energy converged to within 5 mJ/m^2 and worst-case Mg-GB segregation energy converged to within 0.01 eV when cluster calculations were compared to calculations using regular periodic boundary conditions for a boundary of the (120) character shown in Fig. 6b. The resulting buffer distance of 60 Å from the vacuum surface is considerably larger than the cut-off distance of the empirical potentials. This is important because all of the atoms in the cluster are completely free to relax, and elastic information from the relaxation of the cluster surface can propagate

significantly farther than the cut-off distance, which is designed to capture the limit of chemical interaction between two atoms.

While the macroscopic state of the GB is fully defined by the relative orientation of the two grains and the cut of the GB plane, we still need to optimize the microscopic DOF. To this end, we begin by minimizing the GB energy with respect to translations of the two grains relative to each other along with the distance at which two nearby atoms from opposite grains are merged into a single atom. This minimization is performed using simultaneous perturbation stochastic approximation (SPSA).[48] We use an initial step size of 0.0405 Å, i.e., 1% of the lattice parameter. All other parameters, including the decay of the step size and learning rate follow the suggested values in ref. [49] with one addition: we have included a momentum term as found in other gradient-based optimizers which ramps from 0 to 0.8 as the learning rate decays with the form $0.8\left(1 - \frac{a}{a_0}\right)$ where a and a_0 are the current and initial learning rates, respectively. SPSA steps are repeated until the best GB energy is within 0.1 mJ/m^2 of the last best GB energy *and* these two best states are within 0.1 Å of each other in the 4D space of translations and merge distance, *or* the step count reaches 200. Because this process is stochastic, we repeat it ten times with different initial conditions and keep the lowest GB energy structure found.

This stochastic optimization is followed by an annealing period of up to 5 ps at 622 K, which is forked every picosecond to perform a 1 ps quench with a friction constant of 0.01 eVps/Å^2 and then a force-minimization of atomic positions. As with the SPSA minimization, the annealed state with the lowest GB energy is retained.[50,51] This annealing time is not long enough to facilitate faceting of the boundary, but gives the opportunity for local metastable structures to relax into lower energy states.[52] This two-step minimization process is evaluated in the Supplementary Material (Fig. 14) by comparison to the data of Olmsted et al.[25] and is found to perform well. While we have restricted ourselves to 53.1°[001] boundaries in this work, the methodology can be applied to any GB character. Although the boundaries studied differ only in their normal vector, we nonetheless sample a wide range of boundaries from highly ordered to very disordered, including tilt, twist, and tilt-twist combination GBs. This gives us confidence in the applicability of our results to boundaries beyond 53.1° [001].

The workflow described above is coupled to the Large-scale Atomic/ Molecular Massively Parallel Simulator (LAMMPS) MD code[35] [http://lammps. sandia.gov] by a custom Python package [https://github.com/liamhuber/ clustergb] to facilitate high-throughput calculations of GB and segregation energies with empirical potentials. We have demonstrated previously[16] that the potentials of Mendelev et al.[30] and Landa et al.[34] reproduce QM segregation energies with reasonable accuracy for Mg and Pb at Al GBs, respectively. We have not performed this computationally expensive comparison for the remaining solute species and their potentials (Fe by Mendelev et al.,[31] Ti by Zope and Mishin,[32] and Ni and Co by Pun et al.[33]). These potentials were rescaled (if necessary) so that they all had a bulk Al lattice constant of 4.05 Å.

The LM and WC models for solute GB concentration differ in how they treat the interaction energy (a single value and the full, physical distribution of energies, respectively), but they otherwise rest on a shared set of assumptions. Both models assume that the solutes are non-interacting, which is indicated by the fact that the segregation energy is not itself a function of the GB solute concentration. Similarly, the solute-GB interaction is also assumed to be temperature independent, which indicates both that the phonon contributions are ignored, and that the GB structure is assumed to not change at higher temperatures. Making the same assumptions, our segregation energy calculations are made in the dilute limit with 0 K energy minimizations. Finally, both models treat the bulk solute concentration as fixed, i.e., the grain interiors represent infinite solute reservoirs with constant chemical potential of the solute and are thus not suitable for nanograined materials. Developing models which push beyond these assumptions is of great interest, but lies beyond the scope of this particular work.

In the present study, we have chosen to focus only on substitutional segregation sites. There is evidence that some solutes, e.g., Cu,[53] can act as either a substitutional *or* an interstitial solute in Al. While we do not expect this to be the case for over-sized solutes like Pb and Mg, it is possible that the smaller solutes studied here may exhibit similar behaviour. Further, when searching for substitutional sites we consider only those with a non-FCC CNA signature. In our high-angle Σ5 boundaries this accounts for almost the entirety of sites near the GB plane, since having a locally FCC structure by chance is exceedingly small. This technique also ignores all potential sites more than a few angstroms from the GB plane. This is justifiable as solute segregation energies have been found to go to zero very quickly as a function of distance from the GB plane, [23,37]

DATA AVAILABILITY

The relaxed, undecorated structure files and calculated solute segregation energies necessary to reproduce the results of this study are hosted on the Max Planck Society's Edmond repository [https://edmond.mpdl.mpg.de/imeji/collection/ TdKyGarXQf7VhkTB?q=]. GB site Voronoi volumes and coordination numbers are also included for the reader's convenience, although these and all other properties used in our models can be derived directly from the structure files. The code used to perform these high-throughput calculations has been be made public on GitHub [https://github.com/liamhuber/clustergb].

AUTHOR CONTRIBUTIONS

L.H. conceived the study, wrote the necessary code, and carried out the calculations. L.H. and R.H. determined the computational procedure. L.H., B.G., and J.N. analyzed the data. All four authors contributed to writing the manuscript.

ADDITIONAL INFORMATION

Supplementary information accompanies the paper on the *npj Computational Materials* website (https://doi.org/10.1038/s41524-018-0122-7).

Competing interests: The authors declare no competing interests.

REFERENCES

1. Hall, E. The deformation and ageing of mild steel: Iii discussion of results. *Proc. Phys. Soc. Sec. B* **64**, 747 (1951).
2. Kirchheim, R. Reducing grain boundary, dislocation line and vacancy formation energies by solute segregation. i. theoretical background. *Acta Mater.* **55**, 5129–5138 (2007).
3. Koch, C., Scattergood, R., Darling, K. & Semones, J. Stabilization of nanocrystalline grain boundary sizes by solute additions. *J. Mater. Sci.* **43**, 7264–7272 (2008).
4. Chookajorn, T., Murdoch, H. A. & Schuh, C. A. Design of stable nanocrystalline alloys. *Science* **337**, 951–954 (2012).
5. Kalidindi, A. R. & Schuh, C. A. Stability criteria for nanocrystalline alloys. *Acta Mater.* **132**, 128–137 (2017).
6. Amram, D. & Schuh, C. A. Interplay between thermodynamic and kinetic stabilization mechanisms in nanocrystalline Fe-Mg alloys. *Acta Mater.* **144**, 447–458 (2018).
7. Friedel, J. Electronic structure of primary solid solutions in metals. *Adv. Phys.* **3**, 446–507 (1954).
8. McLean, D. *Grain Boundary Segregation in Metals*. Clarendon Press, Oxford (1957).
9. Seah, M. Adsorption-induced interface decohesion. *Acta Metall.* **28**, 955–962 (1980).
10. Udler, D. & Seidman, D. Solute segregation at [001] tilt boundaries in dilute FCC alloys. *Acta Mater.* **46**, 1221–1233 (1998).
11. Nie, J., Zhu, Y., Liu, J. & Fang, X. Periodic segregation of solute atoms in fully coherent twin boundaries. *Science* **340**, 957–960 (2013).
12. White, C. & Coghlan, W. The spectrum of binding energies approach to grain boundary segregation. *Metall. Mater. Trans. A* **8**, 1403–1412 (1977).
13. Rittner, J. & Seidman, D. Solute-atom segregation to ⟨110⟩ symmetric tilt grain boundaries. *Acta Mater.* **45**, 3191–3202 (1997).
14. O'Brien, C. J. & Foiles, S. M. Hydrogen segregation to inclined twin grain boundaries in nickel. *Philos. Mag.* **96**, 2808–2828 (2016).
15. O'Brien, C. J. & Foiles, S. M. Misoriented grain boundaries vicinal to the twin in nickel part ii: thermodynamics of hydrogen segregation. *Philos. Mag.* **96**, 1463–1484 (2016).

16. Huber, L., Grabowski, B., Militzer, M., Neugebauer, J. & Rottler, J. Ab initio modelling of solute segregation energies to a general grain boundary. *Acta Mater.* **132**, 138–148 (2017).

17. Spearot, D. E., Dingreville, R. & O'Brien, C. J. Atomistic simulation techniques to model hydrogen segregation and hydrogen embrittlement in metallic materials. in *Handbook of Mechanics of Materials*, Springer, Singapore. 1–34 (2018).

18. Berthier, F., Legrand, B. & Tréglia, G. How to compare superficial and intergranular segregation? a new analysis within the mixed SMA-TBIM approach. *Acta Mater.* **47**, 2705–2715 (1999).

19. Lezzar, B., Khalfallah, O., Larere, A., Paidar, V. & Duparc, O. H. Detailed analysis of the segregation driving forces for Ni (Ag) and Ag (Ni) in the Σ = 11 {113} and Σ = 11 {332} grain boundaries. *Acta Mater.* **52**, 2809–2818 (2004).

20. Duparc, O. H., Larere, A., Lezzar, B., Khalfallah, O. & Paidar, V. Comparison of the intergranular segregation for eight dilute binary metallic systems in the Σ 11{332} tilt grain boundary. *J. Mater. Sci.* **40**, 3169–3176 (2005).

21. Ghazisaeidi, M., Hector, L. & Curtin, W. Solute strengthening of twinning dislocations in Mg alloys. *Acta Mater.* **80**, 278–287 (2014).

22. Huber, L., Rottler, J. & Militzer, M. Atomistic simulations of the interaction of alloying elements with grain boundaries in Mg. *Acta Mater.* **80**, 194–204 (2014).

23. Scheiber, D., Razumovskiy, V. I., Puschnig, P., Pippan, R. & Romaner, L. Ab initio description of segregation and cohesion of grain boundaries in W-25at.% Re alloys. *Acta Mater.* **88**, 180–189 (2015).

24. Miyazawa, N., Suzuki, S., Mabuchi, M. & Chino, Y. Atomic simulations of the effect of Y and Al segregation on the boundary characteristics of a double twin in Mg. *J. Appl. Phys.* **122**, 165103 (2017).

25. Olmsted, D. L., Foiles, S. M. & Holm, E. A. Survey of computed grain boundary properties in face-centered cubic metals: I. grain boundary energy. *Acta Mater.* **57**, 3694–3703 (2009).

26. Patala, S. & Schuh, C. The topology of homophase misorientation spaces. *Philos. Mag.* **91**, 1489–1508 (2011).

27. Patala, S. & Schuh, C. A. Symmetries in the representation of grain boundary-plane distributions. *Philos. Mag.* **93**, 524–573 (2013).

28. Homer, E. R., Patala, S. & Priedeman, J. L. Grain boundary plane orientation fundamental zones and structure-property relationships. *Sci. Rep.* **5**, 15476 (2015).

29. Lee, B.-J. & Choi, S.-H. Computation of grain boundary energies. *Model. Simul. Mater. Sci. Eng.* **12**, 621 (2004).

30. Mendelev, M., Asta, M., Rahman, M. & Hoyt, J. Development of interatomic potentials appropriate for simulation of solid-liquid interface properties in Al-Mg alloys. *Philos. Mag.* **89**, 3269–3285 (2009).

31. Mendelev, M., Srolovitz, D., Ackland, G. & Han, S. Effect of Fe segregation on the migration of a non-symmetric Σ5 tilt grain boundary in Al. *J. Mater. Res.* **20**, 208–218 (2005).

32. Zope, R. R. & Mishin, Y. Interatomic potentials for atomistic simulations of the Ti-Al system. *Phys. Rev. B* **68**, 024102 (2003).

33. Pun, G. P., Yamakov, V. & Mishin, Y. Interatomic potential for the ternary Ni-Al-Co system and application to atomistic modeling of the B2-L10 martensitic transformation. *Model. Simul. Mater. Sci. Eng.* **23**, 065006 (2015).

34. Landa, A. et al. Development of glue-type potentials for the Al-Pb system: phase diagram calculation. *Acta Mater.* **48**, 1753–1761 (2000).

35. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).

36. Huang, L.-F., Grabowski, B., McEniry, E., Trinkle, D. R. & Neugebauer, J. Importance of coordination number and bond length in titanium revealed by electronic structure investigations. *Status Solidi B* **252**, 1907–1924 (2015).

37. Jin, H., Elfimov, I. & Militzer, M. Study of the interaction of solutes with Σ5 (013) tilt grain boundaries in iron using density-functional theory. *J. Appl. Phys.* **115**, 093506 (2014).

38. Käshammer, P. & Sinno, T. A mechanistic study of impurity segregation at silicon grain boundaries. *J. Appl. Phys.* **118**, 095301 (2015).

39. Karkina, L. et al. Solute-grain boundary interaction and segregation formation in Al: First principles calculations and molecular dynamics modeling. *Comp. Mater. Sci.* **112**, 18–26 (2016).

40. Cao, F., Jiang, Y., Hu, T. & Yin, D. Correlation of grain boundary extra free volume with vacancy and solute segregation at grain boundaries: a case study for Al. *Philos. Mag.* **6**, 1–20 (2017).

41. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

42. Rycroft, C. H. Voro++: A three-dimensional Voronoi cell library in C++. *Chaos* **19**, 041111 (2009).

43. Steinhardt, P. J., Nelson, D. R. & Ronchetti, M. Bond-orientational order in liquids and glasses. *Phys. Rev. B* **28**, 784 (1983).

44. Brown, J. & Mishin, Y. Dissociation and faceting of asymmetrical tilt grain boundaries: Molecular dynamics simulations of copper. *Phys. Rev. B* **76**, 134118 (2007).

45. Faken, D. & Jónsson, H. Systematic analysis of local atomic structure combined with 3D computer graphics. *Comput. Mater. Sci.* **2**, 279–286 (1994).

46. Tsuzuki, H., Branicio, P. S. & Rino, J. P. Structural characterization of deformed crystals by analysis of common atomic neighborhood. *Comput. Phys. Commun.* **177**, 518–523 (2007).

47. Stukowski, A. Visualization and analysis of atomistic simulation data with Ovito-the open visualization tool. *Model. Simul. Mater. Sci. Eng.* **18**, 015012 (2009).

48. Spall, J. C. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE T. Autom. Contr.* **37**, 332–341 (1992).

49. Spall, J. C. Implementation of the simultaneous perturbation algorithm for stochastic optimization. *IEEE T. Aero. Elec. Sys.* **34**, 817–823 (1998).

50. Hadian, R., Grabowski, B., Race, C. P. & Neugebauer, J. Atomistic migration mechanisms of atomically flat, stepped, and kinked grain boundaries. *Phys. Rev. B* **94**, 165413 (2016).

51. Hadian, R., Grabowski, B., Finnis, M. & Neugebauer, J. Migration mechanisms of a faceted grain boundary. *Phys. Rev. Mater.* **2**, 043601 (2018).

52. Han, J., Vitek, V. & Srolovitz, D. J. Grain-boundary metastability and its statistical properties. *Acta Mater.* **104**, 259–273 (2016).

53. Campbell, G. H. et al. Copper segregation to the Σ5 (310)/[001] symmetric tilt grain boundary in aluminum. *Interface Sci.* **12**, 165–174 (2004).