

## ARTICLE OPEN



# Modeling antiphase boundary energies of Ni<sub>3</sub>Al-based alloys using automated density functional theory and machine learning

Enze Chen<sup>1,2</sup>, Artur Tamm<sup>3</sup>, Tao Wang<sup>4</sup>, Mario E. Epler<sup>4</sup>, Mark Asta<sup>1,2</sup> and Timofey Frolov<sup>3</sup>

Antiphase boundaries (APBs) are planar defects that play a critical role in strengthening Ni-based superalloys, and their sensitivity to alloy composition offers a flexible tuning parameter for alloy design. Here, we report a computational workflow to enable the development of sufficient data to train machine-learning (ML) models to automate the study of the effect of composition on the (111) APB energy in Ni<sub>3</sub>Al-based alloys. We employ ML to leverage this wealth of data and identify several physical properties that are used to build predictive models for the APB energy that achieve a cross-validation error of 0.033 J m<sup>-2</sup>. We demonstrate the transferability of these models by predicting APB energies in commercial superalloys. Moreover, our use of physically motivated features such as the ordering energy and stoichiometry-based features opens the way to using existing materials properties databases to guide superalloy design strategies to maximize the APB energy.

*npj Computational Materials* (2022)8:80; <https://doi.org/10.1038/s41524-022-00755-1>

## INTRODUCTION

Owing to their excellent high-temperature mechanical properties and corrosion resistance, Ni-based superalloys are the materials of choice for several important technologies, including airplane turbines and land-based power generators<sup>1–4</sup>. The microstructure of these alloys consists of a Ni-rich, disordered FCC matrix ( $\gamma$  phase) that envelops coherent Ni<sub>3</sub>Al precipitates ( $\gamma'$  phase) which give them their strength<sup>5</sup>. The Ni<sub>3</sub>Al precipitates are ordered in the L1<sub>2</sub> structure, where Al atoms are on the corners and Ni atoms are on the face centers of a simple cubic lattice. Upon the passage of a matrix  $\frac{a}{2}\langle 110 \rangle$ -type dislocation through the ordered precipitates, high-energy antiphase boundaries (APBs) are created on the {111} family of planes to contribute toward  $\gamma'$ -phase strengthening in superalloys<sup>6–8</sup>. One example of where this strengthening mechanism manifests is in the yield strength of the superalloy when dislocations shear the Ni<sub>3</sub>Al precipitates. Depending on the microstructure, dislocation pairs can be weakly-coupled or strongly-coupled, and Reppich found that the yield strength ( $\sigma_y$ ) reaches a peak at the transition point between the two regimes, where it can be approximated by:

$$\sigma_{y,\text{peak}} \propto \frac{\gamma_{\text{APB}} f^{1/2}}{2b} \quad (1)$$

where  $f$  is the volume fraction of the  $\gamma'$  phase,  $b$  is the magnitude of the Burgers vector, and  $\gamma_{\text{APB}}$  is the APB energy. Relationships such as Eq. (1) reinforce the importance of the APB energy in modulating superalloy properties and justify why it has garnered so much attention in superalloy design.

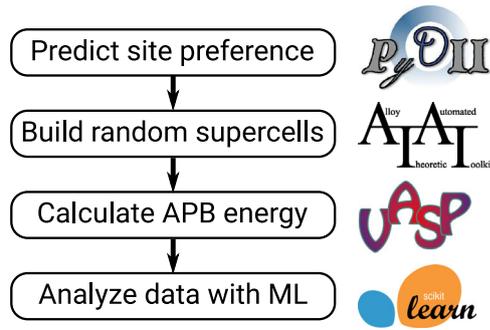
In addition, Ni-based superalloys have multicomponent chemistries often with ten or more elements<sup>1,2</sup>, and this alloying further improves the properties of the base Ni–Al alloy to enhance high-temperature performance<sup>9–11</sup>. The APB energy is sensitive to alloy composition, which provides a rich design space one can explore when optimizing the composition of alloys to provide

higher yield strength and creep resistance. With so many degrees of freedom, however, it would be difficult and prohibitively expensive to rely solely on experimental methods to probe the composition dependence of the APB energy<sup>12,13</sup>. As a result, the community has explored the use of computational methods to understand this correlation and exploit it in alloy design.

Numerous computational investigations into the APB energy in Ni<sub>3</sub>Al and Ni<sub>3</sub>Al-based alloys have been undertaken over the past two decades using a plethora of different methods. Ab initio approaches include density functional theory (DFT)<sup>14–19</sup> and DFT-based methods such as the cluster variation method<sup>20,21</sup>, coherent potential approximation (CPA)<sup>22,23</sup>, and cluster expansion (CE) with Monte Carlo (MC) sampling<sup>24,25</sup>. In addition, semi-empirical methods include molecular dynamics<sup>26</sup> and Monte Carlo simulations<sup>27</sup> based on classical potentials, and the CALculation of PHase Diagram (CALPHAD) approach<sup>18,28</sup>. These studies have contributed greatly to the fundamental understanding of the APB energy in Ni<sub>3</sub>Al-based alloys, including its dependence on temperature, spin polarization, composition, and ordering, all of which have implications for alloy design.

Focusing on the composition dependence, an early DFT study by Chandran and Sondhi found a large strengthening effect due to the addition of Ti, Ta, and Nb atoms, where the (111) APB energy (hereafter symbolized as  $\gamma_{111}$ ) increased from its intrinsic value of 0.181 J m<sup>-2</sup> in Ni<sub>3</sub>Al to over 0.6 J m<sup>-2</sup> in the ternary alloys<sup>15</sup>. Kumar et al. performed a more comprehensive DFT study with 16 different ternary solutes, considering a single atom near the APB in their 96-atom supercell for the ternary element<sup>19</sup>. That study also revealed the sensitivity of  $\gamma_{111}$  to sublattice occupancy of solute species in Ni<sub>3</sub>Al, which must be precisely determined in order to obtain accurate calculations for  $\gamma_{111}$ . Crudden et al. used a combination of DFT and CALPHAD to calculate  $\gamma_{111}$  in several ternary Ni<sub>3</sub>Al-based alloys with refractory metal solutes<sup>18</sup>, and they found compositional dependencies that reasonably agreed with

<sup>1</sup>Department of Materials Science and Engineering, University of California, Berkeley, CA 94720, USA. <sup>2</sup>Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. <sup>3</sup>Materials Science Division, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA. <sup>4</sup>Carpenter Technology Corporation, Philadelphia, PA 19103, USA. ✉email: mdasta@berkeley.edu; frolov2@llnl.gov



**Fig. 1 An outline of our computational workflow.** For each ternary species, we first compute its sublattice preference using PyDII, and then use this prediction to construct model supercells using ATAT. Next, we use VASP to perform DFT calculations to obtain  $\gamma_{111}$ . Finally, we use ML techniques implemented in scikit-learn to analyze correlations in the data and build predictive ML models for  $\gamma_{111}$ .

previous studies of the APB energy and yield strength in these alloys. The use of CALPHAD methods for calculating the APB energy is possible through its connection to the ordering energy ( $E_{\text{ord}}$ ), as detailed by Miodownik and Saunders<sup>28</sup> and leveraged in the work by Crudden et al.<sup>18</sup>. Intuitively, a correlation between these two quantities can be understood as the creation of an APB in the  $L_{12}$ -ordered  $\text{Ni}_3\text{Al}$  precipitates forms an interfacial defect that disrupts the preferred ordered atomic arrangement in the vicinity of the APB. The relationship was also studied by Gorbatov et al., who used the CPA method to discover a strong correlation between  $\gamma_{111}$  and  $E_{\text{ord}}$  in ternary  $\text{Ni}_3\text{Al}$ -based alloys with 3d transition metal solutes at low concentrations (2.5 at.%)<sup>23</sup>. The correlation between  $\gamma_{111}$  and  $E_{\text{ord}}$  has been extensively discussed in the community and is particularly attractive as it enables one to use materials thermodynamic databases, which are extensively developed for multicomponent Ni-based superalloy systems, to model the APB energy.

Building on the previous studies reviewed above, our goal is to further expand the predictive capabilities of APB energy models by leveraging machine-learning (ML) algorithms to automatically learn the sophisticated relationships between structural and energetic properties and the APB energy from data. ML allows us to combine the ordering energy with other physically meaningful features (derived from materials properties) as inputs into a flexible model which has the potential to predict the APB energy with greater accuracy and across a wider range of superalloy chemistries. The wealth of literature on the APB energy in  $\text{Ni}_3\text{Al}$ -based alloys suggests that having a transferable, fast, and predictive model can be impactful for alloy design in screening novel compositions. For this purpose, we require a training dataset that extends across a wider range of compositions than previously considered so that the ML model can learn the appropriate relationships for more exploratory designs with less data bias<sup>29</sup>.

In this work, we create an automated workflow to calculate APB energies at scale, leveraging recent advancements in high-throughput calculations for alloy design<sup>30</sup>. Using this workflow (see Fig. 1; detailed in “Methods”), we analyze the composition dependence of  $\gamma_{111}$  in 111 model ternary  $\text{Ni}_3\text{Al}$ -based alloys, generating significantly more DFT data than what are currently present in the literature. We present the results and compare them with previous studies to validate known behavior and uncover insights across a wide range of ternary elements. In particular, we employ ML to analyze feature correlations and build predictive models for the APB energy that can be used to rapidly screen alloy compositions at a much lower computational cost. Finally, we discuss how the infusion of scientific domain knowledge into data-driven methodologies can enhance superalloy design going forward.

## RESULTS

### Solute site preference

Previous works have demonstrated the sensitivity of the APB energy<sup>19</sup> and mechanical properties<sup>31</sup> of  $\text{Ni}_3\text{Al}$ -based alloys on the solute site preference, i.e., the preference for the solute to reside on the Al or Ni sublattices of the  $L_{12}$  structure. Therefore, we first determine the site preference of ternary species in  $\text{Ni}_3\text{Al}$  using the DFT-based dilute-solution thermodynamic framework implemented in PyDII<sup>32,33</sup> (see “Methods” for details), which has been previously used to predict the solute site preference in intermetallics such as  $\text{NiAl}$ <sup>32</sup> and  $\text{NiTi}$ <sup>34</sup>. From our DFT calculations and analysis with PyDII, we compute the fractional site occupancy on the Al sublattice ( $f_i(\text{Al}) \equiv f$ ) of 46 ternary elements in  $\text{Ni}_3\text{Al}$  as defined by the following equation:

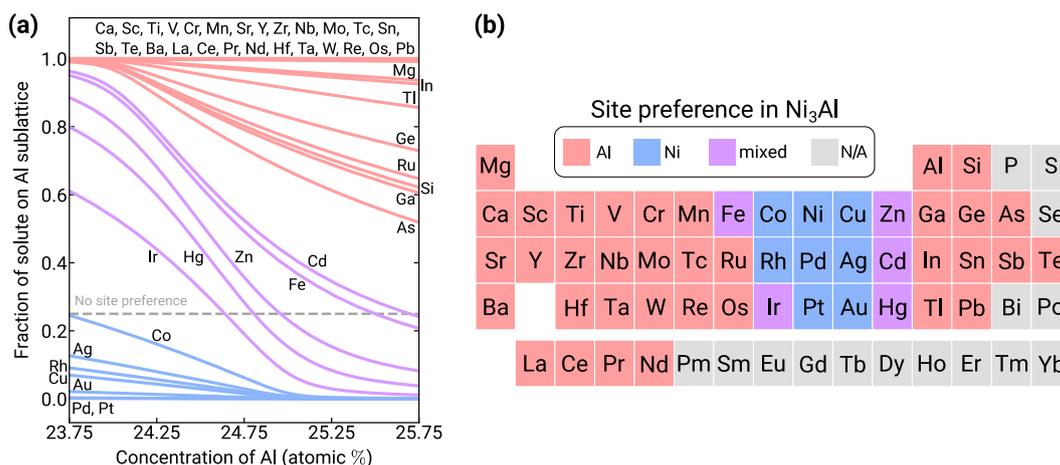
$$f_i(\text{Al}) = \frac{\langle c_i(\text{Al}) \rangle}{\langle c_i(\text{Al}) \rangle + 3\langle c_i(\text{Ni}) \rangle} \quad (2)$$

where  $\langle c_i(X) \rangle$  is the average concentration of species  $i$  on each site of the  $X$  sublattice. We perform the calculations at  $T = 1000$  K and obtain the results shown in Fig. 2, where  $f$  is plotted as a function of the overall mole fraction (concentration) of Al in the  $\text{Ni}_3\text{Al}$  compound. The horizontal dashed line indicates no site preference and corresponds to  $f^* = 0.25$ , when atomic species  $Z$  randomly occupies the two sublattices independent of the Al concentration. By comparing the behavior of  $f$  with the threshold  $f^*$  in the composition range shown in Fig. 2a, we classify the sublattice preference of the ternary species into the following three categories:

- I. If  $f < f^*$  across the entire composition range (blue curves in Fig. 2a), then that solute prefers the Ni sublattice. For the model ternary supercells used to compute the APB energies (see below), these solutes are distributed entirely on the Ni sublattice at all concentrations. The resulting alloy has the composition  $\text{Ni}_{0.75-x}\text{Al}_{0.25}Z_x$ .
- II. If  $f > f^*$  across the entire composition range (red curves), then that solute prefers the Al sublattice. For the APB supercell calculations, these solutes are distributed entirely on the Al sublattice at all concentrations, resulting in a model ternary alloy with composition  $\text{Ni}_{0.75}\text{Al}_{0.25-x}Z_x$ .
- III. Otherwise, we classify the solute as having a mixed sublattice preference (purple curves). These species are distributed such that 3/4 of the solute atoms are on the Ni sublattice and 1/4 of the atoms are on the Al sublattice. The resulting alloy has a composition  $\text{Ni}_{0.75(1-x)}\text{Al}_{0.25(1-x)}Z_x$ .

By automating these calculations and using the above classification system, we identify patterns across the entire periodic table (Fig. 2b). We observe that most elements in groups IX, X, and XI strongly prefer the Ni sublattice (type I) when alloyed in  $\text{Ni}_3\text{Al}$  in dilute quantities. The exception is Ir, which—along with Fe and three group XII elements (Zn, Cd, Hg)—is a type III solute whose occupancy on the Al sublattice is strongly composition-dependent, and these elements are adjacent to the type I elements in the periodic table. The vast majority of the elements prefer the Al sublattice (type II), although that preference is weaker among the  $p$ -block post-transition metals. These results may be rationalized by the larger size of Al atoms compared to Ni atoms (metallic radii of 143 pm and 124 pm, respectively)<sup>1,35</sup>, causing the vast majority of the elements to prefer the Al sublattice. The exception for elements near group X may be due to favorable  $d$ -orbital hybridization with the  $p$  orbitals in Al, which is a prominent interaction in  $\text{Ni}_3\text{Al}$ <sup>36</sup>. Additionally, we note that PyDII predicts most of the common alloying elements in Ni-based superalloys<sup>1,2</sup> to strongly prefer the Al sublattice in  $\text{Ni}_3\text{Al}$ , at least in the dilute regime in Fig. 2.

Our approach using PyDII to calculate site preference of ternary solutes in  $\text{Ni}_3\text{Al}$  can be validated by comparing these results to the



**Fig. 2** Site preference of ternary solute Z in  $\text{Ni}_{0.99-x}\text{Al}_x\text{Z}_{0.01}$ . **a** The calculations are performed at  $T = 1000\text{ K}$ , and the fractional site occupancy  $f$  (Eq. (2)) is plotted as a function of Al concentration. The curves are labeled with the ternary element and the group of elements at the top (bottom) all have overlapping profiles at  $f = 1.0$  ( $f = 0.0$ ), indicating a strong sublattice preference. Based on the composition dependence in the range shown here, we group the elements into three categories for sublattice preference: Ni preference (blue), Al preference (red), and mixed preference (purple). The horizontal dashed line at  $f^* = 0.25$  indicates no site preference. **b** An alternate representation of the solute site preference as a heatmap of the periodic table. A majority of the elements prefer the Al sublattice (red), except for a few near group X that prefer Ni (blue) or exhibit mixed preference (purple). Elements in gray are not tested.

previous studies<sup>37–42</sup>. In a similar study where first-principles calculations were used to obtain defect formation enthalpies to parametrize a Wagner-Schottky model, Jiang and Gleeson<sup>39</sup> predicted the fractional sublattice occupancy in  $\text{Ni}_3\text{Al}$  at 1273 K and produced concentration profiles that are consistent with those in Fig. 2a. They also found a majority of their transition metal solutes to strongly prefer the Al sublattice ( $f \approx 1$  across the whole composition range) while elements like Ir and Fe had a strong composition-dependent site preference in their model. They calculated a stronger composition-dependent site preference for Co and Mn (PyDII predicts Ni and Al sublattice preference, respectively), but other solute profiles were consistent with ours in the composition range we considered. Liu et al. used DFT calculations to parametrize the formation energies in a dilute-solution model and also found a strong Al site preference for most transition metal elements<sup>41</sup>, which they attributed to an atomic size effect when the solute radius exceeds that of Ni by more than 15%. When the size mismatch is below this threshold, they found through analysis of the differential charge density that the Al site preference for a solute Z may be correlated to the stronger Ni–Z interactions from  $d$ -band electrons compared to Al–Z interactions.

### APB energy

Table 1 shows the DFT-calculated values for  $\gamma_{111}$  in model ternary  $\text{Ni}_3\text{Al}$ -based alloys for 37 ternary species at 1.39 at.%, 5.56 at.%, and 9.72 at.% (111 compositions in total). These elements are chosen for their known contributions to high-temperature strength, oxidation resistance, and creep resistance<sup>1,2</sup>, including a few rare-earth metals<sup>43</sup>, or otherwise to span a diverse range of compositions for our data-driven approach. The locations (i.e., site preference) of the ternary species in the supercells are determined using the predictions from PyDII and then randomized over that sublattice using special quasirandom structures (SQSs)<sup>44</sup> generated using the software developed by van de Walle et al.<sup>45,46</sup>. We find that an alloy with 9.72 at.% Ta has the highest APB energy of  $0.464\text{ J m}^{-2}$ , which exceeds the intrinsic (111) APB energy for  $\text{Ni}_3\text{Al}$  by nearly  $0.3\text{ J m}^{-2}$ . The strong influence of Ta on increasing  $\gamma_{111}$  is corroborated by several prior studies<sup>15,17,18</sup>. Notably, we find the highest value for  $\gamma_{111}$  at the medium solute concentration (5.56 at.%) is not for Ta ( $0.353\text{ J m}^{-2}$ ), but rather for W ( $0.369\text{ J m}^{-2}$ ), which

appears to give rise to a strong increase in  $\gamma_{111}$  at lower concentrations. Our calculations show that alloying with Re produces the highest mean APB energy ( $0.227\text{ J m}^{-2}$ ) at low concentration (1.39 at.%), although the variance in  $\gamma_{111}$  means that several other elements, including Ce, Ta, and W, are equally potent within the statistical scatter in our calculations. The contributions from Ta, W, and Re in maximizing  $\gamma_{111}$  at low solute concentrations have also been discussed in other DFT-based studies in the literature<sup>16,19</sup>.

To better illustrate the APB energy data in Table 1, we plot in Fig. 3 a few representative trends for the dependence of  $\gamma_{111}$  on solute concentration. In a majority of ternary systems,  $\gamma_{111}$  varies approximately linearly with concentration up to 10 at.%, which matches previous studies on  $\gamma_{111}$  in  $\text{Ni}_3\text{Al}$ -based alloys<sup>17,23–25</sup>. A more comprehensive comparison of the composition dependence between this work and previously published studies may be found in Supplementary Figs. 1 and 2. Elements in groups IV and V of the periodic table, like Ti, help stabilize the  $\gamma'$  phase<sup>1</sup> and greatly increase  $\gamma_{111}$  as a linear function of their concentration. A few of the elements behave similarly to Co, which has little effect on the APB energy<sup>23</sup>, while others like Rh cause a roughly linear reduction in the APB energy with increasing solute concentration. We observe these non-increasing dependencies (exemplified by Co and Rh) most often in alloys where the ternary species is the type I or type III, corresponding to Ni and mixed sublattice preference, respectively.

Perhaps the most striking dependence on concentration is displayed by the elements in groups VI and VII, with a representative example given by Mo in Fig. 3. Adding these solutes increases the value of  $\gamma_{111}$  up to medium concentrations (5.56 at.%), but any further addition induces a drop in the APB energy. Gorbатов et al. reported a similar non-monotonic dependence in their CPA calculations for  $\gamma_{111}$  for Cr additions<sup>23</sup>, while CALPHAD calculations by Crudden et al. revealed a similar trend in  $\gamma_{111}$  for Mo and W<sup>18</sup>. When we look at the variation in  $\gamma_{111}$  with respect to composition across the first three periods of the  $d$ -block (Fig. 4), we see that this non-monotonic behavior is exhibited by all of the group VI and VII transition metals. Follow-up DFT calculations for  $\gamma_{111}$  without allowing for structural relaxations result in the same qualitative trend for Mo and W,

**Table 1.**  $\gamma_{111}$  (in  $\text{J m}^{-2}$ ) of ternary  $\text{Ni}_3\text{Al}$ -based alloys.

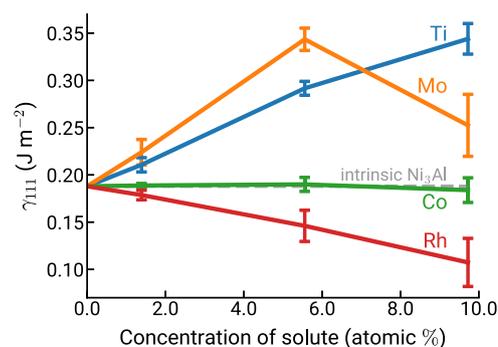
Solute	Site pref.	Low (1.39 at.%)	Medium (5.56 at.%)	High (9.72 at.%)
Ag	Ni	0.187 (01)	0.193 (15)	0.210 (13)
Au	Ni	0.191 (02)	0.205 (06)	0.233 (22)
Cd	Mixed	0.192 (07)	0.183 (21)	0.218 (27)
Ce	Al	0.227 (14)	0.335 (10)	0.374 (31)
Co	Ni	0.189 (02)	0.190 (07)	0.184 (13)
Cr	Al	0.217 (09)	0.281 (04)	0.282 (05)
Cu	Ni	0.186 (02)	0.197 (13)	0.229 (19)
Fe	Mixed	0.203 (07)	0.185 (16)	0.198 (33)
Ga	Al	0.185 (02)	0.173 (02)	0.163 (03)
Ge	Al	0.189 (03)	0.204 (03)	0.223 (05)
Hf	Al	0.205 (29)	0.320 (11)	0.390 (28)
Hg	Mixed	0.193 (07)	0.190 (20)	0.228 (22)
Ir	Mixed	0.177 (05)	0.164 (06)	0.158 (24)
La	Al	0.223 (10)	0.291 (22)	0.262 (80)
Mg	Al	0.186 (04)	0.169 (02)	0.148 (07)
Mn	Al	0.187 (07)	0.187 (07)	0.202 (21)
Mo	Al	0.224 (14)	0.344 (12)	0.252 (33)
Nb	Al	0.221 (13)	0.340 (13)	0.428 (29)
Os	Al	0.216 (08)	0.206 (15)	0.096 (28)
Pb	Al	0.188 (03)	0.199 (05)	0.216 (13)
Pd	Ni	0.185 (04)	0.179 (12)	0.166 (10)
Pt	Ni	0.187 (03)	0.189 (13)	0.180 (10)
Re	Al	0.227 (15)	0.304 (14)	0.122 (38)
Rh	Ni	0.179 (05)	0.146 (17)	0.107 (25)
Ru	Al	0.202 (08)	0.184 (09)	0.120 (20)
Sc	Al	0.208 (05)	0.269 (06)	0.293 (20)
Si	Al	0.189 (04)	0.201 (02)	0.219 (07)
Sn	Al	0.189 (04)	0.207 (05)	0.232 (12)
Ta	Al	0.225 (23)	0.353 (12)	0.464 (30)
Tc	Al	0.221 (10)	0.256 (13)	0.142 (27)
Ti	Al	0.211 (08)	0.292 (07)	0.344 (16)
Tl	Al	0.181 (04)	0.151 (06)	0.069 (08)
V	Al	0.215 (09)	0.308 (08)	0.382 (14)
W	Al	0.225 (16)	0.369 (11)	0.295 (41)
Y	Al	0.215 (08)	0.247 (07)	0.245 (38)
Zn	Mixed	0.197 (07)	0.194 (15)	0.247 (33)
Zr	Al	0.219 (11)	0.318 (11)	0.374 (27)

The APB energy is calculated for three different concentrations of the ternary species. The uncertainty in the final two decimal places is the standard deviation of five different DFT calculations of  $\gamma_{111}$  for each composition. Each row also contains that solute's PyDII-calculated sublattice preference.

leading us to hypothesize that the origin of such behavior is predominantly electronic. Vamsi and Karthikeyan found that W and Mo increase the contributions to the APB energy from Al–Al violations across the APB<sup>47</sup>, which matches the dramatic strengthening effect we observe at the lower solute concentrations.

### Machine-learning modeling

To better understand the composition dependence of  $\gamma_{111}$ , we engineer several physically motivated features and use ML methods to analyze their predictive capabilities in modeling the APB energy. We first use DFT to calculate the ordering energy ( $E_{\text{ord}}$ ) according to Eq. (6) (see “Methods”) and obtain a value of  $-0.118 \text{ eV atom}^{-1}$  for  $\text{Ni}_3\text{Al}$ , which is consistent with the value of

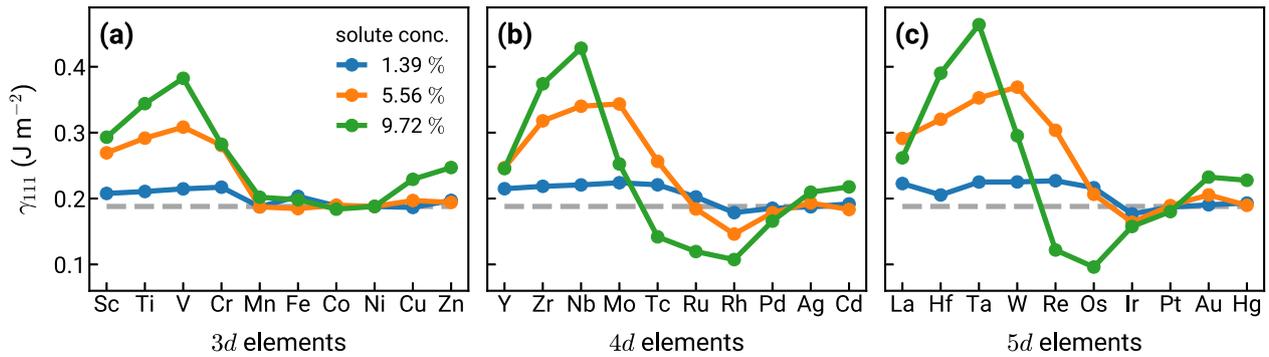


**Fig. 3** Representative profiles of the composition dependence of the APB energy. We plot  $\gamma_{111}$  against solute concentration for select species in model ternary  $\text{Ni}_3\text{Al}$ -based alloys, where error bars represent one standard deviation from the mean value. The intrinsic  $\gamma_{111}$  for  $\text{Ni}_3\text{Al}$  ( $0.188 \text{ J m}^{-2}$ ) is given by the gray dashed line. For a majority of solutes,  $\gamma_{111}$  exhibits an approximately linear dependence on solute concentration, whether that trend is increasing (like Ti), decreasing (like Rh), or relatively constant (like Co). Solutes in groups VI (like Mo) and VII behave non-monotonically where  $\gamma_{111}$  increases up to the medium concentration and then decreases as the solute concentration further increases.

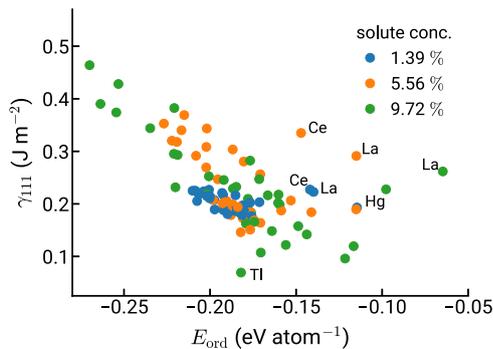
$-0.12 \text{ eV atom}^{-1}$  obtained by Gorbatov et al. using DFT<sup>23</sup>. The trend for the ordering energy ratio ( $E_{\text{ord}}/E_{\text{ord}}^{\text{Ni}_3\text{Al}}$ ) among the 3d elements also agrees qualitatively with the CPA results obtained by Gorbatov et al., although our energy ratios are higher by  $\sim 40\%$ . We plot  $\gamma_{111}$  against  $E_{\text{ord}}$  in Fig. 5 and further differentiate the data points based on the ternary solute concentration. There exists a clear negative correlation between the two variables, with a Pearson correlation coefficient of  $r = -0.541$ , indicating a moderately strong linear correlation. This correlation appears to be present at all solute concentrations and agrees qualitatively with previous studies on the 3d transition metals<sup>23</sup>.

In addition to the DFT-calculated ordering energy, density, and change in lattice parameters, we use several composition-based features generated using the matminer Python® package (Python® is a registered trademark of Python Software Foundation Corporation)<sup>48</sup> (see “Methods”). When we compute the Pearson correlation coefficient between each feature and  $\gamma_{111}$ , we find that four of the features have a stronger linear correlation to the APB energy than the ordering energy (Table 2). The feature with the strongest linear correlation ( $r = 0.692$ ) is the weighted standard deviation of the number of unfilled *d* orbitals in each constituent element. The second strongest correlation ( $r = 0.570$ ) is for the weighted standard deviation of the column number of each constituent element while the third ( $r = 0.547$ ) is the weighted mean of the covalent radius of each constituent element. We find that the weighted mean Pettifor number<sup>49</sup> also has a marginally stronger linear correlation ( $r = -0.543$ ) than the ordering energy. These features are all automatically generated from the alloy’s composition using matminer<sup>48</sup> and their correlations to  $\gamma_{111}$  reflect the trends shown in Fig. 4. For example, adding transition metals from the left side of the d-block is shown to dramatically boost the APB energy and it increases the value of the first two features (those elements have more unfilled *d* orbitals than Ni (2) and Al (0) and they have a smaller column number than Ni (10) and Al (13)). We also generally observe  $\gamma_{111}$  to decrease across a row and increase down a group, which agrees qualitatively with the trends in elemental covalent radii.

After filtering out features with undefined values and features that are highly correlated to another feature, we apply recursive feature elimination (RFE) to further downselect the features and use grouped cross-validation (CV)<sup>50</sup> to obtain the optimal number of nine features. We note that  $E_{\text{ord}}$  is the only physical parameter



**Fig. 4**  $\gamma_{111}$  of model ternary  $\text{Ni}_3\text{Al}$ -based alloys for solutes from the first three periods of the  $d$ -block.  $\gamma_{111}$  at low, medium, and high solute concentrations are given by the blue, orange, and green profiles, respectively, for the **a**  $3d$ , **b**  $4d$ , and **c**  $5d$  transition metals.  $\gamma_{111}$  for Ni at all concentrations is set to the  $\gamma_{111}$  value for  $\text{Ni}_3\text{Al}$  ( $0.188 \text{ J m}^{-2}$ ), which is also represented by the gray dashed lines.



**Fig. 5** Visualizing the correlation between APB energy ( $\gamma_{111}$ ) and ordering energy ( $E_{\text{ord}}$ ) in  $\text{Ni}_3\text{Al}$ -based alloys. The two properties exhibit a negative, moderately strong linear correlation ( $r = -0.541$ ). Points are color-coded based on solute concentration and select outliers are labeled with the ternary element.

**Table 2.** Linear correlations between the features and  $\gamma_{111}$ .

Feature	$r$ value
Mean Pettifor number	-0.543
Ordering energy	-0.541
Mean # of unfilled $p$ orbitals	-0.527
Std. dev. of the # of unfilled $d$ orbitals	0.692
Std. dev. of the column number	0.570
Mean covalent radius	0.547

The features with the three most negative and three most positive  $r$  values are shown, with four features having a stronger linear correlation to  $\gamma_{111}$  than  $E_{\text{ord}}$ . Except for the ordering energy, the feature name explains the statistic that is computed for the elemental values of that property.

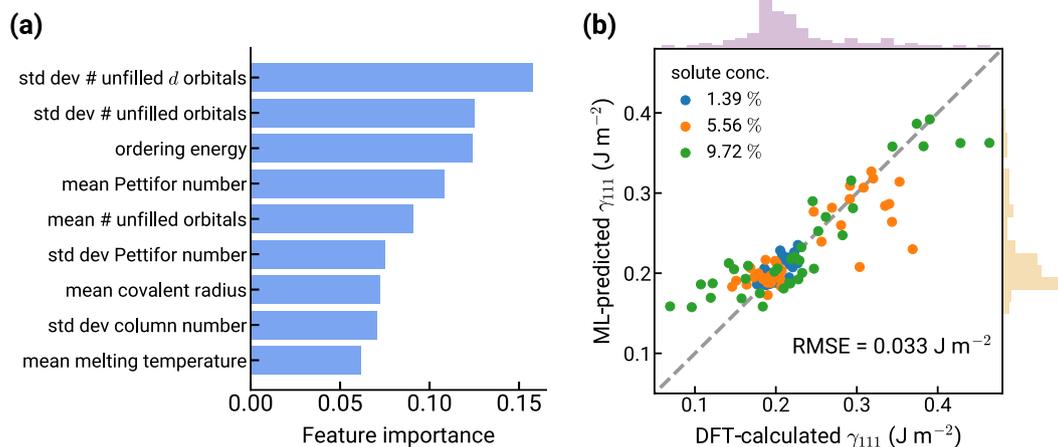
of the  $L1_2$  structure that remains. We use these nine features to build a final random forest (RF) model and perform grouped fivefold CV to predict  $\gamma_{111}$ . Figure 6a shows the normalized feature importances<sup>51</sup> computed by the RF algorithm for the nine features selected from RFE. The standard deviation of the elemental number of unfilled  $d$  orbitals surfaces as the most important feature, which follows from its large positive correlation to  $\gamma_{111}$ . The ordering energy is only the third most important feature in our model, preceded by the standard deviation of the elemental number of unfilled orbitals. The parity plot in Fig. 6b compares the cross-validated predictions of  $\gamma_{111}$  using the RF model against the DFT-calculated values of  $\gamma_{111}$ . We again color the data points by

solute concentration and observe relatively good agreement between the ML predictions and the DFT calculations. The overall RMSE from grouped fivefold CV is  $0.033 \text{ J m}^{-2}$  ( $R^2 = 0.753$ ), which is comparable in magnitude with the average experimental error of  $0.027 \text{ J m}^{-2}$  for multinary ( $>2$  elements) compositions in Table 1 from Crudden et al.<sup>18</sup> As there may be interest in an ML model without  $E_{\text{ord}}$  as a feature, particularly for multicomponent compositions where this information may not be readily available, we report the fivefold CV results for such a model in the Supplementary Methods.

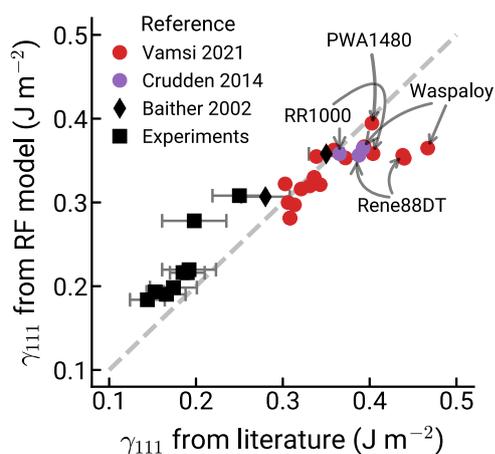
#### Extrapolation to multicomponent chemistries

To demonstrate the generalizability of our ML model, we use our methodology to train an RF model to predict  $\gamma_{111}$  for real superalloy chemistries reported in the literature. Specifically, we use the compositions for the  $\gamma'$  phase reported by Baither et al.<sup>13</sup> and Vamsi and Karthikeyan<sup>47</sup>, which were obtained through experiments and thermodynamic modeling, respectively. The  $\gamma'$  phase in these commercial superalloys contain multicomponent chemistries ranging from 5 to 11 constituent elements. As we do not have DFT-calculated data nor access to a thermodynamic database for these alloys, we first retrain the RF model with only composition-based features (does not include  $E_{\text{ord}}$ ) for the full training set of 111 ternary  $\text{Ni}_3\text{Al}$ -based alloys (training error of  $0.011 \text{ J m}^{-2}$ ;  $R^2 = 0.974$ ). We then use this trained model to predict  $\gamma_{111}$  for the unseen test data (i.e., 23 commercial, multicomponent superalloys) and compare the ML predictions with the  $\gamma_{111}$  values from the literature (Fig. 7). While direct application of the model for ternary systems (which includes  $E_{\text{ord}}$ ) is not possible, our analysis shows that models with comparable performance can be developed without  $E_{\text{ord}}$  (see Supplementary Methods). We also use several other experimental measurements from the literature<sup>12,52–56</sup> to test our model and we include those results in Fig. 7 (black squares). The overall test RMSE is  $0.038 \text{ J m}^{-2}$ , which compares favorably to the CV error, even though the test set contains compositions that are more complex than what exist in our training set. Conversely, when we build a model that simply uses a linear combination of ternary APB energies<sup>18,25</sup>, i.e.,  $\Delta\gamma^{\text{total}} = \sum_{i=1}^n k_i \Delta\gamma_i^{\text{ternary}}$ , we obtain a higher extrapolation error of  $0.048 \text{ J m}^{-2}$  on the test set.

Compared with the experimental measurements by Baither et al. (black diamonds in Fig. 7), our model predicts values for  $\gamma_{111}$  that are consistent with those reported in that study<sup>13</sup>. We find the error from our RF model reasonable given the comparison to experimental data and our model still outperforms other predictive models in the literature<sup>25,47</sup> (Supplementary Fig. 3). The red points in Fig. 7 represent the APB energies from the study by Vamsi and Karthikeyan, who used an environment-dependent nearest-neighbor bond (EDNNB) model that was also fit to DFT



**Fig. 6 Results from training the random forest (RF) model using grouped fivefold CV. a** Ranked feature importances (normalized to sum to 1.0) for the nine features used in the final RF model. **b** ML-predicted  $\gamma_{111}$  vs. DFT-calculated  $\gamma_{111}$  for model ternary Ni<sub>3</sub>Al-based alloys. Perfect agreement is given by the gray dashed line and points are color-coded based on solute concentration. The RF model achieves a RMSE of 0.033 J m<sup>-2</sup> in a fivefold CV.



**Fig. 7 Testing the ML model on superalloy chemistries given in the literature.** The ML model trained on only the ternary alloys with composition-based features achieves a test error of 0.038 J m<sup>-2</sup> when used to predict  $\gamma_{111}$  in these systems. The computational data from Crudden et al.<sup>18</sup> and Vamsi and Karthikeyan<sup>47</sup> for commercial superalloys are shown in purple and red, respectively. Experimental measurements<sup>12,52–56</sup> with error bars are shown in black, with data for commercial superalloys from Baither et al.<sup>13</sup> shown as black diamonds. Select superalloys are labeled for clarity.

data<sup>47</sup>. They used a thermodynamic database to predict the composition of the  $\gamma'$ -phase for commercial superalloys and then used those compositions as inputs to their EDNNB model. Our ML model predicts values for  $\gamma_{111}$  that are consistent with their EDNNB model predictions for a majority of the superalloys, but under-predicts  $\gamma_{111}$  at higher values. We note, however, that our ML predictions for Waspalloy, Rene88DT, and RR1000 are in closer agreement with the predicted values from Crudden et al. (purple points)<sup>18</sup>, which exemplifies the variability between different models for the APB energy.

## DISCUSSION

The APB energy is an important parameter for alloy design as it influences several mechanical properties of precipitation-strengthened superalloys and can be tuned by adjusting the

chemical composition. By creating a computational workflow to automate DFT calculations, we generate a wealth of data including solute site preference, (111) APB energies, and physically meaningful features such as the ordering energy to enable a data-driven assessment of the chemical contributions to the APB energy. The synergy between high-throughput calculations and machine learning presents an opportunity to explore vast swaths of alloy composition space and accelerate the process of alloy development.

It is interesting to note PyDII's predictions for Co and Cr, which are the two species with the greatest variability in site preference among previously published studies. For Co, we classify its behavior as preferring the Ni sublattice, but its concentration profile varies the most among type I elements, suggesting it has the possibility of occupying both sublattices<sup>35,39,42</sup> when analyzed with different compositions or at different temperatures. For Cr, PyDII predicts a very strong preference for the Al sublattice, which reasonably agrees with previous computational studies, but differs from the MC study by Saito and Harada<sup>57</sup> and a few atom probe tomography experiments<sup>58,59</sup>. These three studies notably used multicomponent superalloy chemistries and Booth-Morrison et al. found that Ta has a propensity to kick Cr out of Al sites<sup>59</sup>, which could help explain the differences with our results for Cr. Competition between multiple solute species, which more closely mimics real superalloy chemistries, deserves further scrutiny in future investigations and can be incorporated in the dilute-solution thermodynamic formalism employed by PyDII. In general, validation against the literature confirms the viability of using PyDII as an automated tool to predict site preferences of solute species in ordered intermetallics, which can potentially be extended to other precipitation-strengthened alloys<sup>60,61</sup>.

Our DFT results also reveal that several elements exhibit a non-monotonic concentration dependence for  $\gamma_{111}$ , which can have important consequences for alloy design. As we scan across the *d*-block elements (Fig. 4), Ta may be the element that maximizes  $\gamma_{111}$  at high concentrations, but the same cannot be said at lower concentrations, where other solutes like Mo and W perform comparably or better. Furthermore, several predictive models for the APB energy in the literature use a linear regression model based on each solute's concentration<sup>18,25</sup>. Crudden et al.<sup>18</sup> calculated  $\gamma_{111}$  for 12.5 at.% Mo and W, and they assumed that the APB energy varied linearly between 0 at.% and 12.5 at.%. Our results show that for several elements, particularly those in groups VI and VII which are currently used in superalloys, the choice of

those endpoints can affect the estimate of its chemical contribution to the APB energy in the  $\gamma'$  phase, and we must be more cautious of making assumptions about monotonicity. These data can also be used to guide the selection of alloying elements to maximize APB energy when facing additional constraints (e.g., total solute quantity, weight, cost, processing techniques).

Building on previous analytical models for the APB energy, our nonlinear ML model uses a derived set of physically motivated features and is not constrained to a particular analytical form. Seeing ordering energy surface as one of the most important features for predicting  $\gamma_{111}$  agrees favorably with our physical intuition, given the known relationships between the APB energy and the ordering energy<sup>18,23,28</sup>. Furthermore, the ability to extract the ordering energy of an alloy system from CALPHAD<sup>28</sup> makes it a viable input feature for modeling the APB energy by integrating ML workflows with existing tools and databases from alloy design<sup>62,63</sup>. Using stoichiometry alone, we can calculate several additional features that improve the generalizability and interpretability of our ML model. Evidenced by the grouped CV performance, these features are already effective at making predictions for alloy chemistries that are not present in the training dataset, albeit with room for improvement (W at medium concentration is the severely under-predicted orange data point in Fig. 6b). This predictive ability is important for exploratory alloy development, where such a model is sufficiently accurate to be able to screen for promising compositions (i.e., rank those with high APB energies as shown in Fig. 7) in new regions of chemical space and can couple with existing tools for increased predictability in the broader context of alloy design. The continued expansion of databases for multicomponent alloys presents exciting opportunities to enrich our ML models and extrapolate the methods presented here to other extended defects (e.g., stacking faults). As materials data become increasingly digitized, abundant, and accessible, we foresee the coupling of materials thermodynamic databases and data-driven methodologies to have an outside impact in this field.

## METHODS

We outlined our computational workflow in Fig. 1 and elaborate on the details in the following subsections.

### Solute site preference

PyDII<sup>32</sup> employs the grand canonical, dilute-solution model (DSM) thermodynamic formalism based on a statistical-mechanical low-temperature expansion following the framework described by Woodward et al.<sup>64</sup> to predict equilibrium defect concentrations in intermetallics as a function of composition and temperature. As a brief review of the PyDII framework, we can define concentration variables  $c_i(p)$  to denote the atomic fraction of species  $i$  at site  $p$  in an ordered intermetallic. Here,  $p$  is either the Ni or Al sublattice and  $i$  can be Ni, Al, or the solute species. When we write out the expression for the grand potential  $\Omega$ , we can take the first-order low-temperature expansion of  $\Omega$  to obtain (assuming zero stress):

$$\Omega = E^0 - \sum_i \mu_i \sum_p \lambda(p) c_i^0(p) - k_B T \sum_p \lambda(p) \sum_c \exp\left(-\left[\delta E^c(p) - \sum_i \mu_i \delta c_i^c(p)\right]/k_B T\right) \quad (3)$$

In Eq. (3),  $E^0$  and  $c_i^0(p)$  are the energy and concentrations, respectively, in the ground-state configuration, which corresponds to binary, stoichiometric Ni<sub>3</sub>Al in this study.  $\mu_i$  is the chemical potential of species  $i$ ,  $\lambda(p)$  denotes the multiplicity of site  $p$ ,  $k_B$  is the Boltzmann constant and  $T$  is temperature.  $\delta E^c(p)$  and  $\delta c_i^c(p)$  are the defect “excitation energy” and change in site concentration, respectively, associated with introducing a substitutional or vacancy point defect at site  $p$ .

Equation (3) can be differentiated to derive an equation for the average concentration:

$$\langle c_i(p) \rangle = c_i^0(p) + \sum_c \delta c_i^c(p) \times \exp\left(-\left[\delta E^c(p) - \sum_j \mu_j \delta c_j^c(p)\right]/k_B T\right) \quad (4)$$

where  $\langle c_i(p) \rangle$  denotes the ensemble-averaged concentration of species  $i$  at site  $p$ . Combining Eqs. (3) and (4) and the condition of thermodynamic equilibrium with respect to vacancies<sup>64</sup> produces a nonlinear system of equations that can be solved at a specified  $T$  to obtain a set of chemical potentials,  $\{\mu_i\}$ . The  $\{\mu_i\}$  are then used to determine sublattice concentrations for each species as a function of the Al concentration.

For Ni<sub>3</sub>Al, the fraction of a species  $i$  on the Al sublattice can then be computed using Eq. (2) in the main text. The PyDII code<sup>33</sup> is used to generate the default structures and settings for the DFT calculations of the excitation energies using the Vienna Ab initio Simulation Package (VASP<sup>®</sup>) (VASP<sup>®</sup> is a registered trademark of VASP Software GmbH Limited.)<sup>65–68</sup> with the projector augmented-wave (PAW) method<sup>69,70</sup> and the generalized gradient approximation of Perdew, Burke, and Ernzerhof (GGA-PBE)<sup>71</sup>. The default VASP settings include a fixed plane wave energy cutoff of 520 eV, 5000  $k$ -points per reciprocal atom, spin polarization, ionic relaxation, and convergence criteria for electronic self-consistency and ionic relaxation of  $1 \times 10^{-6}$  eV and  $1 \times 10^{-3}$  eV, respectively.

We then use the DFT-calculated defect excitation energies to calculate the equilibrium sublattice concentrations at  $T = 1000$  K and a fixed concentration of 1 at.% for the ternary species. The sublattice preference calculated from PyDII is then used to guide supercell construction for model ternary Ni<sub>3</sub>Al-based alloys according to the following distributions for the ternary species:

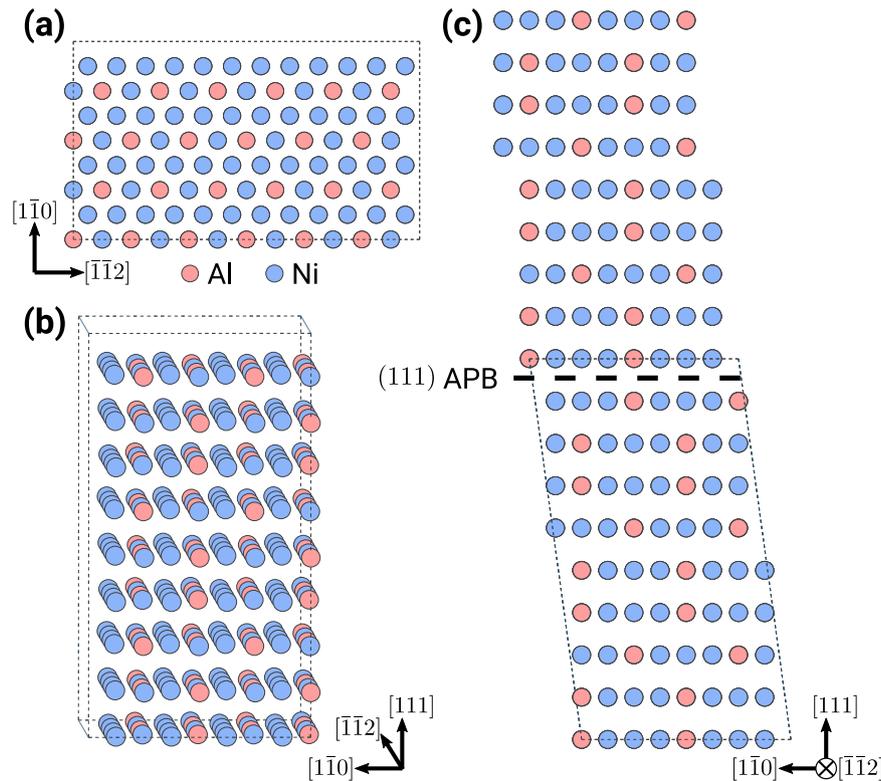
- I. Solutes with strong Ni sublattice preference are distributed entirely on the Ni sublattice at all concentrations. The resulting alloy has the composition Ni<sub>0.75–x</sub>Al<sub>0.25</sub>Z<sub>x</sub>.
- II. Solutes with strong Al sublattice preference are distributed entirely on the Al sublattice at all concentrations. The resulting alloy has the composition Ni<sub>0.75</sub>Al<sub>0.25–x</sub>Z<sub>x</sub>.
- III. The remaining solutes with composition-dependent site preference are distributed such that 3/4 of the solute atoms are on the Ni sublattice and 1/4 of the atoms are on the Al sublattice. The resulting alloy has a composition Ni<sub>0.75(1–x)}</sub>Al<sub>0.25(1–x)}</sub>Z<sub>x</sub>.

For each composition, we randomize the placement of atoms on the appropriate sublattices using special quasirandom structures (SQSs)<sup>44</sup> generated using the Alloy Theoretic Automated Toolkit (ATAT)<sup>45,46</sup>. We optimize for pair and triplet correlations out to 0.55 nm, which captures the first few NN correlations for each configuration. The maximum absolute value of the correlation difference between the SQS and ideally random is 0.017. We emphasize that these solute distributions are simply chosen to create model supercells for our high-throughput workflow. The species distribution in the  $\gamma'$  phase in a real superalloy will differ depending on the temperature, (relative) chemical potentials of the species, processing techniques, etc. Despite this simplification, we believe our method will still capture the dominant behaviors and trends in site preference—and, correspondingly, the APB energy—which we validate against other studies in Results and Supplementary Note 1. We find that the sensitivity of the APB energy to local ordering (e.g., using a different SQS) results in differences by as much as  $0.01 \text{ J m}^{-2}$  in  $\gamma_{111}$ , but should not affect our conclusions (see Supplementary Methods). This is an intrinsic limitation of the SQS/DFT approach and we have ongoing work to better quantify this sensitivity by comparing across random, SQS, and DFT-MC supercells.

### DFT calculations for $\gamma_{111}$

We perform DFT calculations for  $\gamma_{111}$  using VASP with the PAW method and the GGA-PBE functional. For these calculations, we set the plane wave cutoff energy for each calculation to be at least 30% greater than the highest cutoff energy (ENMAX in VASP) specified in the pseudopotential files (ENCUT is at least 350 eV). Spin-polarized calculations are performed to account for the effects of local magnetic moments on the APB energy<sup>14</sup>. Calculations with spin polarization appear to give closer agreement to literature values (both computational and experimental, see Supplementary Figs.) as non-spin-polarized calculations result in values for  $\gamma_{111}$  that are ~30% lower.

For all APB energy calculations, we use a 288-atom computational cell with the  $x$ -,  $y$ -, and  $z$ -axes oriented along the  $[\bar{1}\bar{1}2]$ ,  $[1\bar{1}0]$ , and  $[111]$  directions, respectively (Fig. 8). This corresponds to a  $2 \times 2 \times 3$  L<sub>12</sub> supercell and nine (111) planes separating the APBs, which we deem sufficient based on convergence tests. We use  $\Gamma$ -centered Monkhorst-Pack  $k$ -point grids<sup>72</sup> of  $2 \times 3 \times 2$ , corresponding to 5000  $k$ -points per reciprocal atom, based on a test of energy convergence for L<sub>12</sub> (Supplementary Fig. 4). We apply a first-order Methfessel-Paxton smearing method<sup>73</sup> with a smearing width of  $\sigma = 0.1$  eV. Structural relaxations are performed using a conjugate



**Fig. 8 Schematics of our 288-atom DFT supercell.** The Al and Ni sites are red and blue, respectively, and they are populated with the ternary species based on the predictions from PyDII followed by randomization with SQS. **a** Top view of the (111) plane. **b** Bulk cell without an APB. **c** A single (111) APB is created near the top of the supercell by tilting the cell by  $\mathbf{b} = a/2[1\bar{1}0]$ . The cell is projected along the  $[\bar{1}\bar{1}2]$  direction and has been repeated once to show the APB (note the new Al–Al and Ni–Ni bonds).

gradient algorithm in two steps: First, only the volume is allowed to relax for the bulk structure to capture the lattice expansion due to the solute. Then, the bulk structure is copied and tilted such that the top face translates by  $\mathbf{b} = a/2[1\bar{1}0]$  (Fig. 8c) to create a structure with a single (111) APB. Independent ionic relaxations are then performed on the bulk structure (without an APB) and the second structure with an APB. The convergence criteria for the electronic self-consistency and ionic relaxation loops are  $1 \times 10^{-5}$  eV and  $1 \times 10^{-3}$  eV, respectively. Other VASP settings and DFT convergence tests may be found in Supplementary Methods.

For binary  $\text{Ni}_3\text{Al}$ , our DFT calculations give an equilibrium lattice constant of approximately 0.3565 nm, which agrees well with reported computational (e.g., refs. <sup>19,26</sup>) and experimental values (0.357 nm)<sup>1</sup>. We find that  $\gamma_{111}$  for stoichiometric  $\text{Ni}_3\text{Al}$  is  $0.188 \text{ J m}^{-2}$ , and we observe a roughly linear increase in  $\gamma_{111}$  with increasing Al concentration (Supplementary Table 1), which are consistent with previous experimental<sup>52,74</sup> and computational (e.g., refs. <sup>14,15,23</sup>) studies. We also observe a variability in  $\gamma_{111}$  depending on the species arrangements around the APB for a fixed composition. To account for these differences, we shift the (111) planes in the tilted supercell to create five different APBs to help quantify the statistical scatter. The energy of a single (111) APB, denoted by  $\gamma_{111}^{(i)}$ , is calculated according to Eq. (5):

$$\gamma_{111}^{(i)} = \frac{E_{\text{apb}}^{(i)} - E_{\text{bulk}}}{A_{111}} \quad (5)$$

where  $E_{\text{apb}}^{(i)}$  is the total energy of the  $i$ th structure with an APB,  $E_{\text{bulk}}$  is the total energy of the non-tilted supercell, and  $A_{111}$  is the area of the APB (Fig. 8a). The mean and standard deviation are calculated for five different APB energies to obtain  $\gamma_{111}$  for each composition. We do this for each ternary alloy system at three different solute concentrations: low (1.39 at. %), medium (5.56 at. %), and high (9.72 at. %). We note that while these DFT calculations at 0 K do not account for the configurational entropy contributions to the free energy<sup>20</sup> nor the decrease in APB energy at high temperatures ( $>1000 \text{ K}$ )<sup>24</sup>, they can still provide quantitatively accurate data for room-temperature properties<sup>23</sup> that can be used to build predictive models for screening new compositions.

To calculate the ordering energy  $E_{\text{ord}}$ , we perform additional DFT calculations using smaller, 80-atom supercells of similar compositions and a  $\mathbf{k}$ -point grid of the same density as before.  $E_{\text{ord}}$  is the difference in energy between two structures: one with atoms on their preferred sublattice (which we determine using PyDII) and another with the same composition with atoms randomly distributed on all sites. In this case, all degrees of freedom are allowed to relax, but all other DFT settings remain the same as those for calculating  $\gamma_{111}$ .  $E_{\text{ord}}$  in units of  $\text{eV atom}^{-1}$  is then given by Eq. (6):

$$E_{\text{ord}} = \frac{E_{\text{L12}} - E_{\text{random}}}{N} \quad (6)$$

where  $E_{\text{L12}}$  and  $E_{\text{random}}$  are the total energies of the structures with correct and random sublattice distributions, respectively, and  $N$  is the total number of atoms. Eq. (6) is used to obtain the ordering energy for all ternary  $\text{Ni}_3\text{Al}$ -based alloys except for 9.72 at.% Ce whose calculations did not converge.

### Machine learning

We generate physically motivated features from the alloy composition using the matminer Python package<sup>48</sup>, as a similar approach by Ling et al. was shown to be effective for featuring superalloy chemistries<sup>75</sup>. Briefly, the matminer featurizer creates an equal-length feature vector for every composition by calculating statistical quantities (we use the composition-weighted mean and standard deviation) over a set of elemental properties<sup>76</sup> of the alloying elements. To this elemental feature vector from matminer we also include several more derived features that we believe to be influential in modeling the APB energy, including the DFT-calculated ordering energy ( $E_{\text{ord}}$ ), density, change in lattice parameter, and composition-weighted Pettifor number<sup>49</sup>, among others (see Supplementary Table S2). We generate a total of 70 features and remove all features with at least one undefined value and those that have a Spearman's rank correlation<sup>77</sup> greater than 0.9 with another feature.

We train a random forest (RF) regression model, as implemented in the scikit-learn Python package<sup>78</sup>, to predict  $\gamma_{111}$  using the root-mean-square

error (RMSE) as the error metric. The model is trained with 100 decision trees grown to full depth and with a third of the features considered at each split. To further downselect the features for predictive modeling, we use recursive feature elimination to recursively remove the least important feature from our model based on the feature importance metric in random forests<sup>51</sup>. The RF model's performance is evaluated on the full dataset using grouped fivefold cross-validation (CV)<sup>50</sup>, where the data are grouped by ternary species. This means that four of the folds are used as training data and the final fold is used for validation, while alloys with the same ternary species will always appear in the same fold. This training-validation procedure is repeated such that each of the fivefold is used once as validation data and the model's overall performance (RMSE) is computed as the average of all five RMSE scores from CV. The grouping process helps ensure that our ML model is extrapolating to new regions of chemical space and calculating more faithful validation metrics that are representative of novel alloy compositions encountered during the screening. We compare our choice of fivefold CV (a standard 80–20 training-validation split used in the community) with other techniques (e.g., tenfold, leave-one-solute-out) in Supplementary Fig. 5, but emphasize that a better demonstration of extrapolation performance is to new multicomponent chemistries as reported in Results.

## DATA AVAILABILITY

The data that support the findings of this study are available on MPContribs<sup>80</sup>. Other data are available from the corresponding authors upon reasonable request.

## CODE AVAILABILITY

The DFT calculations are performed using the Vienna Ab initio Simulation Package (VASP)<sup>65–68</sup>. All other software (PyDII<sup>33</sup>, ATAT<sup>45</sup>, and the Python packages<sup>48,78</sup>) is open source.

Received: 20 September 2021; Accepted: 15 March 2022;

Published online: 25 April 2022

## REFERENCES

- Reed, R. C. *The Superalloys: Fundamentals and Applications*, 1st edition. (Cambridge University Press, 2006).
- Pollock, T. M. & Tin, S. Nickel-based superalloys for advanced turbine engines: chemistry, microstructure and properties. *J. Propulsion Power* **22**, 361–374 (2006).
- Long, H., Mao, S., Liu, Y., Zhang, Z. & Han, X. Microstructural and compositional design of Ni-based single crystalline superalloys—a review. *J. Alloys Compounds* **743**, 203–220 (2018).
- Darolia, R. Development of strong, oxidation and corrosion resistant nickel-based superalloys: critical review of challenges, progress and prospects. *Int. Mater. Rev.* **64**, 355–380 (2019).
- Nembach, E. & Neite, G. Precipitation hardening of superalloys by ordered  $\gamma'$ -particles. *Progress Mater. Sci.* **29**, 177–319 (1985).
- Kear, B. H. & Wilsdorf, H. G. F. Dislocation configurations in plastically deformed polycrystalline  $\text{Cu}_3\text{Au}$  alloys. *Transact. Metallurg. Soc. AIME* **224**, 382–386 (1962).
- Reppich, B. Some new aspects concerning particle hardening mechanisms in  $\gamma'$  precipitating Ni-base alloys—I. Theoretical concept. *Acta Metallurgica* **30**, 87–94 (1982).
- Paidar, V., Pope, D. P. & Vitek, V. A theory of the anomalous yield behavior in  $\text{L}_{12}$  ordered alloys. *Acta Metallurgica* **32**, 435–448 (1984).
- Smith, T. M. et al. Phase transformation strengthening of high-temperature superalloys. *Nat. Commun.* **7**, 13434 (2016).
- Wu, X. et al. Unveiling the Re effect in Ni-based single crystal superalloys. *Nat. Commun.* **11**, 389 (2020).
- Yang, T. et al. Ultrahigh-strength and ductile superlattice alloys with nanoscale disordered interfaces. *Science* **369**, 427–432 (2020).
- Baluc, N., Schäublin, R. & Hemker, K. J. Methods for determining precise values of antiphase boundary energies in  $\text{Ni}_3\text{Al}$ . *Philosophical Magazine Lett.* **64**, 327–334 (1991).
- Baither, D., Rentenberger, C., Karnthaler, H. P. & Nembach, E. Three alternative experimental methods to determine the antiphase-boundary energies of the  $\gamma'$  precipitates in superalloys. *Philosophical Magazine A* **82**, 1795–1805 (2002).
- Manga, V. R., Saal, J. E., Wang, Y., Crespi, V. H. & Liu, Z.-K. Magnetic perturbation and associated energies of the antiphase boundaries in ordered  $\text{Ni}_3\text{Al}$ . *J. Appl. Phys.* **108**, 103509 (2010).
- Chandran, M. & Sondhi, S. K. First-principle calculation of APB energy in Ni-based binary and ternary alloys. *Model. Simul. Mater. Science Eng.* **19**, 025008 (2011).
- Yu, X.-X. & Wang, C.-Y. Effect of alloying element on dislocation cross-slip in  $\gamma'$ - $\text{Ni}_3\text{Al}$ : a first-principles study. *Philosoph. Magazine* **92**, 4028–4039 (2012).
- Vamsi, K. V. & Karthikeyan, S. Effect of off-stoichiometry and ternary additions on planar fault energies in  $\text{Ni}_3\text{Al}$ . in *Proceedings of the Twelfth International Symposium on Superalloys*, 521–530 (TMS, Champion, PA, 2012).
- Crudden, D. J., Mottura, A., Warnken, N., Raesinia, B. & Reed, R. C. Modelling of the influence of alloy composition on flow stress in high-strength nickel-based superalloys. *Acta Materialia* **75**, 356–370 (2014).
- Kumar, K., Sankarasubramanian, R. & Waghmare, U. V. Tuning planar fault energies of  $\text{Ni}_3\text{Al}$  with substitutional alloying: first-principles description for guiding rational alloy design. *Scripta Materialia* **142**, 74–78 (2018).
- Sluiter, M., Hashi, Y. & Kawazoe, Y. The effect of segregation and partial order on the thermodynamics of (111) antiphase boundaries in  $\text{Ni}_3\text{Al}$ . *Comput. Mater. Sci.* **14**, 283–290 (1999).
- Wang, H.-P., Sluiter, M. & Kawazoe, Y. Prediction of the effect of Ti on the (111) and (100) antiphase boundary energy in  $\text{Ni}_3\text{Al}$ . *Mater. Transactions JIM* **40**, 1301–1305 (1999).
- Liu, J. B. & Johnson, D. D. First principle predictions of anomalous yield strength in  $\text{L}_{12}$  materials. *Mater. Res. Innovations* **18**, S4–1021–S4–1025 (2014).
- Gorbatov, O. I. et al. Effect of composition on antiphase boundary energy in  $\text{Ni}_3\text{Al}$ -based alloys: Ab initio calculations. *Phys. Rev. B* **93**, 224106 (2016).
- Sun, R., Woodward, C. & van de Walle, A. First-principles study on  $\text{Ni}_3\text{Al}$  antiphase boundary with Ti and Hf impurities. *Phys. Rev. B* **95**, 214121 (2017).
- Dodaran, M. et al. Effect of alloying elements on the  $\gamma'$  antiphase boundary energy in Ni-base superalloys. *Intermetallics* **117**, 106670 (2020).
- Mishin, Y. Atomistic modeling of the  $\gamma$  and  $\gamma'$ -phases of the Ni-Al system. *Acta Materialia* **52**, 1451–1467 (2004).
- Becker, C. A., Mishin, Y. & Boettinger, W. J. The pre-wetting transition at antiphase boundaries: An atomistic modeling study of  $\text{Ni}_3\text{Al}$ . *J. Mater. Sci.* **43**, 3873–3880 (2008).
- Miodownik, A. P. & Saunders, N. The calculation of APB energies in  $\text{L}_{12}$  compounds using a thermodynamic database. in *Applications of Thermodynamics in the Synthesis and Processing of Materials* (eds Nash, P. & Sundman, B.) 91–104 (The Minerals, Metals & Materials Society, Warrendale, 1995).
- Jia, X. et al. Anthropogenic biases in chemical reaction data hinder exploratory inorganic synthesis. *Nature* **573**, 251–255 (2019).
- van de Walle, A. & Asta, M. High-throughput calculations in the context of alloy design. *MRS Bull.* **44**, 252–256 (2019).
- Rawlings, R. D. & Staton-Bevan, A. E. The alloying behaviour and mechanical properties of polycrystalline  $\text{Ni}_3\text{Al}$  ( $\gamma'$  phase) with ternary additions. *J. Mater. Sci.* **10**, 505–514 (1975).
- Ding, H. et al. PyDII: a Python framework for computing equilibrium intrinsic point defect concentrations and extrinsic solute site preferences in intermetallic compounds. *Comput. Phys. Commun.* **193**, 118–123 (2015).
- Medasani, B. Python framework for Defects In Intermetallics. *GitHub* <https://github.com/pydii/pydii> (2015).
- Singh, N. et al. Effect of ternary additions to structural properties of NiTi alloys. *Comput. Mater. Sci.* **112**, 347–355 (2016).
- Eriş, R., Akdeniz, M. V. & Mekhrabov, A. O. Atomic size effect of alloying elements on the formation, evolution and strengthening of  $\gamma'$ - $\text{Ni}_3\text{Al}$  precipitates in Ni-based superalloys. *Intermetallics* **109**, 37–47 (2019).
- Iotova, D., Kiousis, N. & Lim, S. P. Electronic structure and elastic properties of the  $\text{Ni}_3\text{X}$  (X=Mn, Al, Ga, Si, Ge) intermetallics. *Phys. Rev. B* **54**, 14413–14422 (1996).
- Sluiter, M. H. F. & Kawazoe, Y. Site preference of ternary additions in  $\text{Ni}_3\text{Al}$ . *Phys. Rev. B* **51**, 4062–4073 (1995).
- Ruban, A. V. & Skriver, H. L. Calculated site substitution in ternary  $\gamma'$ - $\text{Ni}_3\text{Al}$ : temperature and composition effects. *Phys. Rev. B* **55**, 856–874 (1997).
- Jiang, C. & Gleeson, B. Site preference of transition metal elements in  $\text{Ni}_3\text{Al}$ . *Scripta Materialia* **55**, 433–436 (2006).
- Wu, Q. & Li, S. Alloying element additions to  $\text{Ni}_3\text{Al}$ : site preferences and effects on elastic properties from first-principles calculations. *Comput. Mater. Sci.* **53**, 436–443 (2012).
- Liu, S. et al. Partitioning and diffusion of transition metal solutes in ternary model Ni-based single crystal superalloys. *Mater. Design* **130**, 157–165 (2017).
- Lu, B., Wang, C.-Y. & Du, Z. Site preferences of alloying transition metal elements in Ni-based superalloy: a first-principles study. *Chinese Phys. B* **27**, 097102 (2018).
- Kumar, A., Chernatynskiy, A., Hong, M., Phillpot, S. R. & Sinnott, S. B. An ab initio investigation of the effect of alloying elements on the elastic properties and magnetic behavior of  $\text{Ni}_3\text{Al}$ . *Comput. Mater. Sci.* **101**, 39–46 (2015).
- Zunger, A., Wei, S.-H., Ferreira, L. G. & Bernard, J. E. Special quasirandom structures. *Phys. Rev. Lett.* **65**, 353–356 (1990).
- van de Walle, A., Asta, M. & Ceder, G. The alloy theoretic automated toolkit: a user guide. *Calphad* **26**, 539–553 (2002).

46. van de Walle, A. et al. Efficient stochastic generation of special quasirandom structures. *Calphad* **42**, 13–18 (2013).
47. Vamsi, K. V. & Karthikeyan, S. Modeling APB energies in multicomponent Ni-base superalloys. *Intermetallics* **132**, 107124 (2021).
48. Ward, L. et al. Matminer: An open source toolkit for materials data mining. *Comput. Mater. Sci.* **152**, 60–69 (2018).
49. Pettifor, D. G. A chemical scale for crystal-structure maps. *Solid State Commun.* **51**, 31–34 (1984).
50. Lu, H.-J. et al. Error assessment and optimal cross-validation approaches in machine learning applied to impurity diffusion. *Comput. Mater. Sci.* **169**, 109075 (2019).
51. Breiman, L. Random forests. *Machine Learn.* **45**, 5–32 (2001).
52. Dimiduk, D. M., Thompson, A. W. & Williams, J. C. The compositional dependence of antiphase-boundary energies and the mechanism of anomalous flow in Ni<sub>3</sub>Al alloys. *Philosophical Magazine A* **67**, 675–698 (1993).
53. Korner, A. Weak-beam study of superlattice dislocations moving on cube planes in Ni<sub>3</sub>(Al, Ti) deformed at room temperature. *Philosophical Magazine A* **58**, 507–522 (1988).
54. Hemker, K. J. & Mills, M. J. Measurements of antiphase boundary and complex stacking fault energies in binary and B-doped Ni<sub>3</sub>Al using TEM. *Philosophical Magazine A* **68**, 305–324 (1993).
55. Ezz, S. S. & Hirsch, P. B. The strain rate sensitivity of the flow stress and the mechanism of deformation of single crystals of Ni<sub>3</sub>(Al Hf)B. *Philosophical Magazine A* **69**, 105–127 (1994).
56. Sun, J., Lee, C. S., Lai, J. K. L. & Wu, J. S. Dislocation dissociations and fault energies in Ni<sub>3</sub>Al alloys doped with palladium. *Intermetallics* **7**, 1329–1335 (1999).
57. Saito, Y. & Harada, H. The Monte Carlo simulation of ordering kinetics in Ni-base superalloys. *Mater. Sci. Eng.: A* **223**, 1–9 (1997).
58. Bagot, P. A. J. et al. An atom probe tomography study of site preference and partitioning in a nickel-based superalloy. *Acta Materialia* **125**, 156–165 (2017).
59. Booth-Morrison, C., Mao, Z., Noebe, R. D. & Seidman, D. N. Chromium and tantalum site substitution patterns in Ni<sub>3</sub>Al(L1<sub>2</sub>)  $\gamma'$ -precipitates. *Appl. Phys. Lett.* **93**, 033103 (2008).
60. Suzuki, A., Inui, H. & Pollock, T. M. L1<sub>2</sub>-strengthened cobalt-base superalloys. *Annu. Rev. Mater. Res.* **45**, 345–368 (2015).
61. Li, Z. M. et al. Cuboidal  $\gamma'$  phase coherent precipitation-strengthened Cu–Ni–Al alloys with high softening temperature. *Acta Materialia* **203**, 116458 (2021).
62. Reed, R. C., Tao, T. & Warnken, N. Alloys-by-design: application to nickel-based single crystal superalloys. *Acta Materialia* **57**, 5898–5913 (2009).
63. Menou, E., Ramstein, G., Bertrand, E. & Tancret, F. Multi-objective constrained design of nickel-base superalloys using data mining- and thermodynamics-driven genetic algorithms. *Modelling Simulation Mater. Sci. Eng.* **24**, 055001 (2016).
64. Woodward, C., Asta, M., Kresse, G. & Hafner, J. Density of constitutional and thermal point defects in L1<sub>2</sub> Al<sub>3</sub>Sc. *Phys. Rev. B* **63**, 094103 (2001).
65. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558–561 (1993).
66. Kresse, G. & Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal-amorphous-semiconductor transition in germanium. *Phys. Rev. B* **49**, 14251–14269 (1994).
67. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
68. Kresse, G. & Furthmüller, J. Efficiency of ab initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
69. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
70. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
71. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
72. Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **13**, 5188–5192 (1976).
73. Methfessel, M. & Paxton, A. T. High-precision sampling for Brillouin-zone integration in metals. *Phys. Rev. B* **40**, 3616–3621 (1989).
74. Kruml, T., Conforto, E., Lo Piccolo, B., Caillard, D. & Martin, J. L. From dislocation cores to strength and work-hardening: a study of binary Ni<sub>3</sub>Al. *Acta Materialia* **50**, 5091–5101 (2002).
75. Ling, J. et al. Machine learning for alloy composition and process optimization. *ASME Turbo. Expo. 2018*, GT2018–75207 (2018).
76. Ward, L., Agrawal, A., Choudhary, A. & Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Comput. Mater.* **2**, 16028 (2016).
77. Spearman, C. The proof and measurement of association between two things. *Am. J. Psychol.* **15**, 72–101 (1904).
78. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Machine Learn. Res.* **12**, 2825–2830 (2011).
79. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
80. Chen, E. et al. Antiphase boundary energies of Ni<sub>3</sub>Al-based alloys. *MPContribs* [https://contribs.materialsproject.org/projects/apbe\\_Ni3Al](https://contribs.materialsproject.org/projects/apbe_Ni3Al) (2022).

## ACKNOWLEDGEMENTS

This work is partly performed under the auspices of the U.S. Department of Energy (DOE) by the Lawrence Livermore National Laboratory (LLNL) under Contract No. DE-AC52-07NA27344. The authors are grateful for project funding from the High-Performance Computing for Materials (HPC4Mtl) Program of the DOE Vehicle Technologies Office under Cooperative Research and Development Agreement (CRADA) No. TC02309. Computing support for this work comes from the LLNL Institutional Computing facilities, and the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility operated under Contract No. DE-AC02-05-CH11231. E.C. acknowledges a fellowship through the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1752814. M.A. acknowledges support for his contributions by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under Contract No. DE-AC02-05-CH11231 within the Materials Project program (KC23MP). All figures are produced using matplotlib<sup>79</sup>.

## AUTHOR CONTRIBUTIONS

E.C., M.A., A.T., and T.F. conceived the research. E.C. performed the simulations and modeling and analyzed the results with feedback from M.A. and T.F. T.W. performed the CALPHAD calculations. E.C. and M.A. drafted the manuscript. All authors discussed the results and commented on the manuscript. M.A. and T.F. supervised the project.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41524-022-00755-1>.

**Correspondence** and requests for materials should be addressed to Mark Asta or Timofey Frolov.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022