

## ARTICLE OPEN



# Mining of lattice distortion, strength, and intrinsic ductility of refractory high entropy alloys

Christopher Tandoc<sup>1</sup>, Yong-Jie Hu<sup>1✉</sup>, Liang Qi<sup>2</sup> and Peter K. Liaw<sup>3</sup>

Severe lattice distortion is a prominent feature of high-entropy alloys (HEAs) considered a reason for many of those alloys' properties. Nevertheless, accurate characterizations of lattice distortion are still scarce to only cover a tiny fraction of HEA's giant composition space due to the expensive experimental or computational costs. Here we present a physics-informed statistical model to efficiently produce high-throughput lattice distortion predictions for refractory non-dilute/high-entropy alloys (RHEAs) in a 10-element composition space. The model offers improved accuracy over conventional methods for fast estimates of lattice distortion by making predictions based on physical properties of interatomic bonding rather than atomic size mismatch of pure elements. The modeling of lattice distortion also implements a predictive model for yield strengths of RHEAs validated by various sets of experimental data. Combining our previous model on intrinsic ductility, a data mining design framework is demonstrated for efficient exploration of strong and ductile single-phase RHEAs.

*npj Computational Materials* (2023)9:53; <https://doi.org/10.1038/s41524-023-00993-x>

## INTRODUCTION

High-entropy alloys (HEA), or namely the compositionally complex alloys for a broader definition, have attracted a great deal of attention as a promising material solution for addressing the urgent societal goals of reduced carbon emissions and increased energy efficiency. Many HEAs have already been demonstrated to have remarkable yield strengths<sup>1–8</sup>, great corrosion<sup>9–14</sup> and fatigue resistance<sup>15–17</sup>, reasonable ductility<sup>3–6,18–22</sup>, and/or thermal stability<sup>3,23,24</sup>. More excitingly, the multicomponent chemistry grants HEAs a nearly unlimited design space, making this group of alloys full of potential as revolutionary structural materials for both light weighting and high-temperature applications. The outstanding performances of HEAs are recognized to largely originate from their unique non-dilute, multicomponent chemistry and atomic structures at the fundamental level<sup>25–28</sup>.

A signature characteristic of the atomic structures of HEAs is severe lattice distortion<sup>29,30</sup>. In HEAs, multiple types of atoms with different atomic sizes and electronegativities are randomly mixed in a single-crystal lattice. The size mismatch and charge transfer between individual atoms result in significant variations in interatomic distances and angles. As a result, the atoms cannot reside perfectly on the ideal lattice site, giving rise to a considerable lattice distortion throughout the entire crystal. With respect to the perfect lattice, the distorted lattice of HEAs generates atomic strains that fluctuate from one atom to another, which produces intense interactions with the strain field of dislocations. As a result, every atom in HEAs can actually be considered as a point-pinning obstacle to retard dislocation movement, leading to an outstanding solid-solution type strengthening effect<sup>31–35</sup>. For instance, extremely high yield strengths have been achieved in HEAs by elaborate manipulation of lattice distortion<sup>4,36–39</sup>. Additionally, lattice distortion is also recognized to play essential roles in the phase stability<sup>40,41</sup>, sluggish diffusion<sup>42,43</sup>, electrical<sup>44,45</sup>, and thermal conductivities<sup>45</sup> of HEAs. Therefore, a deep understanding of lattice distortion

could assist the design of high-performance HEAs without wandering in the endless compositional space.

Lattice distortion in HEAs can be accurately characterized for individual alloy compositions through various experimental methods, including neutron diffraction, synchrotron X-ray diffraction, and high-resolution transmission electron microscopy<sup>4,37,38,46</sup>. Additionally, first-principles calculations based on density functional theory (DFT) have been extensively applied to study lattice distortion owing to the accurate description of interatomic bonding. The supercell method based on the special quasi-random structure (SQS)<sup>47</sup> is widely employed to mimic the multicomponent random mixing in HEAs<sup>33,47–49</sup>. The alloy's atomic structure at the ground state can be straightforwardly obtained by performing a relaxation calculation. Based on the relaxed structure, lattice distortion can be accurately quantified via a variety of analysis approaches, such as radial or pair distribution function<sup>50,51</sup>, bond length distribution and fluctuation<sup>36,49</sup>, local atomic volume<sup>52</sup>, least-square atomic strain<sup>53</sup>, and mean square average atomic displacement (MSAD)<sup>33</sup>. Nevertheless, like the experimental approaches, most of the time the DFT calculations were only able to selectively study a small group of discrete alloy compositions due to large computational costs. Apart from the accurate experimental characterization and DFT-based computation, several mathematical models have been proposed for predicting lattice distortion simply according to the alloy composition. However, those models usually oversimplified lattice distortion as the difference in the pure-element atomic size of the constituent elements by assuming that each atom retains its original size even in a complex solid-solution environment. This assumption fails to account for the changes in atomic radii due to charge transfer among constituent atoms, which thus becomes problematic for predicting lattice distortion in the systems containing elements with similar atomic sizes but different electronegativities<sup>37,54</sup>. The lack of a generally applicable model for rapid and accurate predictions of lattice distortion for arbitrary alloy compositions significantly limits our ability to explore the large design spaces of HEA efficiently.

<sup>1</sup>Department of Materials Science and Engineering, Drexel University, Philadelphia, PA 19104, USA. <sup>2</sup>Department of Materials Science and Engineering, University of Michigan, Ann Arbor, MI 48109, USA. <sup>3</sup>Department of Materials Science and Engineering, University of Tennessee, Knoxville, TN 37996, USA. ✉email: yh593@drexel.edu

In the present work, a surrogate model based on physics-informed statistical learning and training data from first-principles calculations is developed to efficiently predict the lattice distortion in the body-centered-cubic (bcc) refractory HEAs (RHEAs) with compositions arbitrarily varying in chemical space composed of ten elements. The study is centered on RHEAs because they are not studied as comprehensively as the face-centered-cubic (fcc) 3d-transition-metal HEAs though lattice distortion is indeed found to be significant in individual cases and critically contribute to the mechanical strength<sup>4,35,36</sup>. The lattice distortion is quantified using the root mean squared atomic displacement (RMSAD), which measures the average displacement of relaxed atoms from their ideal positions in the undistorted crystal lattice<sup>33,55</sup>. On the basis of physical descriptors that characterize atomic bonds of pure metals and their binary ordered compounds, the developed surrogate model delivers high-throughput predictions of lattice distortion quantified in a way that previously can be only done for individual compositions by costly experiments or first-principles calculations. The high-throughput prediction enables a systematic overview of lattice distortion in the alloys with different levels of chemical complexity, from which the physical mechanism of HEA's unique cocktail effect on lattice distortion is revealed.

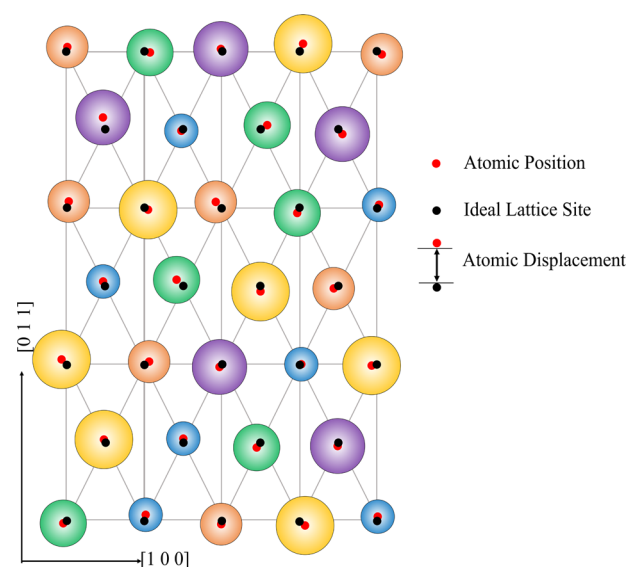
Based on the predictions of the lattice distortion model, it is further found that the RMSAD parameter quantitatively correlates with the room-temperature yield strengths of the bcc refractory alloys in a general linear form. Accordingly, a strengthening model is established to effectively predict yield strengths of RHEAs according to their lattice distortion. Together with our previous surrogate models on intrinsic ductility<sup>56</sup> and CALculation of PHase Diagram (CALPHAD)-based phase stability prediction, a high-throughput alloy design framework is demonstrated to perform comprehensive screening of potential alloy compositions with a synergy of high strength and good ductility in a large compositional space far from fully explored.

## RESULTS AND DISCUSSION

### Initial data exploration

As schematically illustrated in Fig. 1, due to lattice distortion, atoms at equilibrium positions in a solid-solution alloy are locally displaced away from their ideal positions in a perfect, undistorted lattice. A mean squared average of those atomic displacements (i.e., the MSAD) has been demonstrated an accurate quantitative measure of lattice distortion in 3d-transition-metal HEAs both experimentally and computationally<sup>33,57,58</sup>. In the present work, the square root of the mean squared atomic displacement (i.e., the RMSAD parameter) is used to quantify lattice distortion because of its strong linear correlation with yield strength<sup>33,57</sup>. Through a DFT relaxation calculation, the RMSAD parameter for a given alloy composition can be straightforwardly derived as described in the method section in detail. To generate the training and testing data for the development of a surrogate predictive model for lattice distortion, we computed the RMSAD parameters of 215 individual alloy compositions as listed in Supplementary Table 1. Prior to using those DFT data for direct statistical regressions, we first utilized them to perform an initial data exploration, aiming to gain enough physical insights of lattice distortion to better develop the surrogate model. We limited the exploration process to the training dataset, which only contains binary and ternary alloys. As a result, the subsequent development of the surrogate model based on the data exploration is completely blind to the validation and testing datasets (i.e., quaternary and quinary alloys), and those datasets were thus valid to use for testing the model's ability of extrapolative prediction.

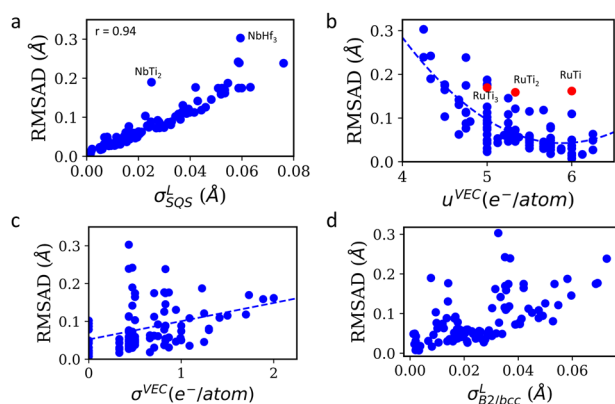
The standard deviation of local atomic-bond lengths was naturally considered as a starting point for exploring RMSAD, because an ideal, undistorted bcc lattice would have uniform local bond lengths for every atom, and any displacement of an atom from its ideal



**Fig. 1** A schematic shows the local atomic displacements due to lattice distortion. The red dots stand for the equilibrium atomic positions in a bcc solid-solution alloy with lattice distortion, while the black dots stand for the ideal positions in an undistorted lattice. The schematic corresponds to a projection view along the  $[0\bar{1}1]$  direction. The big solid circles in different colors represent individual types of atoms in the alloy.

lattice site would introduce the variance to the interatomic-bond lengths between the displaced atom and its neighbors. In the present work, we calculated the length of every first-nearest neighbor (FNN) atomic bond in the relaxed SQS's using Voronoi-tessellation analysis, as implemented in the Pymatgen Python package<sup>59</sup>, and then derived the standard deviation of the bond lengths ( $\sigma_{505}^L$ ) in comparison with the RMSAD parameter obtained for the same structure. A scatter plot shown in Fig. 2a depicts a strong positive correlation between the  $\sigma_{505}^L$  and RMSAD for the alloy samples in the training dataset, with a correlation coefficient,  $r = 0.94$ , and significance level,  $p < 0.001$ . The correlation indicates that the lattice distortion in solid-solution alloys tightly connects to the length divergence of individual local atomic bonds. In addition, there are two apparent outliers to be off from the correlation displayed in Fig. 2a, which are NbTi<sub>2</sub> and NbHf<sub>3</sub>. A common feature of the two alloys is their high content of the group IV element, giving valence electron concentrations (VECs) that are at the lower end of those represented in the training dataset. This trend implies that the *d*-band filling effect may also play a key role to affect the lattice distortion in bcc refractory solid-solution alloys.

Accordingly, a scatter plot of VEC against RMSAD is displayed in Fig. 2b for the alloys in the training dataset. It was found that RMSAD showed a typical dependence on the average filling fraction of the *d*-band. Specifically, the RMSAD parameter becomes lower around VEC of 5.5  $e^-$ /atom, which corresponds to an averagely half-filled *d*-band, and becomes higher when the VEC is outside the range of 5.0–6.0  $e^-$ /atom, at which the *d*-band deviates from the half-filled state. In addition to this general *d*-band filling dependence, Fig. 2b also highlights an interesting phenomenon that RMSAD can still vary greatly for the alloys with the same VEC. A typical example is the Ti-Ru binary system. As presented in Fig. 2b, although all the three studied Ti-Ru alloys have VECs within the range of 5.0–6.0  $e^-$ /atom, they tend to have more severe lattice distortion (i.e., a larger RMSAD), compared to other alloys with the same VECs. It should be noted that the bcc solid-solution phase is metastable for studied Ti-Ru compositions<sup>60</sup>. The observed severe lattice distortion actually reflects the strong phase-transformation tendency from the disordered solid solution to ordered intermetallic



**Fig. 2** Correlations between the root-mean-squared atomic displacement (RMSAD) parameter and fundamental properties of atomic bonds in refractory solid-solution alloys. **a** RMSAD vs. standard deviation of lengths of the first-nearest neighbor (FNN) atomic bonds in the relaxed SQS's ( $\sigma_{SQS}^L$ ); **b** RMSAD vs. average valence electron concentration ( $u^{VEC}$ ); **c** RMSAD vs. standard deviation of the number of valence electrons of the constituent elements relative to VEC ( $\sigma^{VEC}$ ); and **d** RMSAD vs. standard deviation of atomic bond lengths estimated from single-element bcc and binary B2 crystals ( $\sigma_{B2/bcc}^L$ ) based on Eq. (6).

compounds. Another obvious distinction of the studied Ti-Ru alloys from others with similar VECs is that its constituent elements, Ti and Ru, have very different numbers of valence electrons. Considering the strong localization tendency of valence  $d$  electrons, one can expect that the local distribution of the  $d$  electrons at the Ti-Ti, Ti-Ru, and Ru-Ru atomic bonds in the alloy are inhomogeneous, giving local fluctuations in the VEC. Given this insight, the standard deviation of the number of valence electrons of the constituent elements relative to VEC ( $\sigma^{VEC}$ ) is also calculated and plotted to against RMSAD in Fig. 2c. As expected,  $\sigma^{VEC}$  shows a general positive correlation with RMSAD, suggesting that in addition to the bond-length divergence, the divergence in the  $d$ -band filling fraction at local atomic bonds, which is induced by the variance of constituent element valences, could also be an important contributor to lattice distortion.

### A physically interpretable surrogate model for RMSAD prediction

Although the DFT-SQS calculations provide a general approach to assess the lattice distortion of each individual alloy composition by predicting the RMSAD parameter, the relatively large size of SQS supercells make the calculation computationally expensive and consequently, ill-suited for high-throughput screening. Therefore, to fully explore the complex compositional space of RHEAs, it is necessary to develop surrogate models for more efficient prediction of RMSAD.

The data exploration discussed above provides a theoretical basis to raise a set of physical descriptors and an interpretable linear model to quantitatively describe RMSAD. As shown in Fig. 2a, the RMSAD parameter displays a strong linear correlation with the standard deviation of the FNN bond lengths in relaxed SQS's ( $\sigma_{SQS}^L$ ). However,  $\sigma_{SQS}^L$  cannot be directly used as a descriptor for the prediction of RMSAD because its obtainment requires expensive DFT calculations to relax the SQS for each alloy composition of interest. A surrogate approach is thus needed to efficiently estimate the variance in lengths of FNN atomic bonds in a solid-solution alloy.

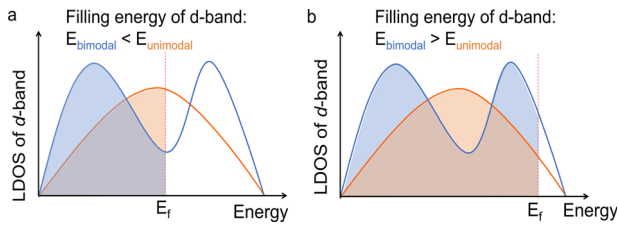
As discussed later in the methods section, the FNN atomic bonds in the single-element bcc and binary B2 crystals can serve as an approximation to those in the solid-solution lattice if only the atomic-pair interactions in the FNN shell are considered. We expect that this assumption is particularly reasonable for the

transition metal alloy systems as the valence  $d$  electrons are strongly localized. As a validation, the lengths of the FNN atomic bonds derived from the DFT-relaxed SQS's are compared with their respective lengths in the single-element bcc and binary B2 crystals by showing the histogram of relative deviations in Supplementary Fig. 1. By comparing across a considerable amount of FNN bonds in the SQS's of 68 binary and ternary alloys, the mean error and root mean square error (RMSE) are only 0.0009 Å and 0.0312 Å (−0.0567% mean relative error and 2.258% root mean square relative error), respectively, which validates our assumption. Therefore, the standard deviation of atomic bond lengths in a solid-solution alloy can be effectively estimated from the lengths of the FNN bonds in the single-element bcc and binary B2 structures using the bond-counting approach described in the method section. Correspondingly, a descriptor for the bond length standard deviation,  $\sigma_{bcc/B2}^L$  is derived. Based on Eq. (6),  $\sigma_{bcc/B2}^L$  can be easily calculated for any alloy composition of interest without the need to run expensive DFT-SQS calculations.

As shown in Fig. 2d, the correlation between the RMSAD and the bond-length standard deviation ( $\sigma_{bcc/B2}^L$ ) calculated using Eq. (6) is also significant ( $p < 0.001$ ). On the other hand, significant variability in RMSAD remains unaccounted for by  $\sigma_{bcc/B2}^L$  alone as evidenced by an increase in the number of outliers compared to  $\sigma_{SQS}^L$ , suggesting that it is insufficient on its own. The results of data exploration in Fig. 2b, c suggest that the  $d$ -band filling effect could be another important contributor to lattice distortion in addition to the bond-length standard deviation. Therefore, it is natural to consider VEC as a possible descriptor to explain the discrepancies in the RMSAD vs.  $\sigma_{bcc/B2}^L$  correlation, especially given that all the outliers share VECs much lower than those of other samples. To test this hypothesis, the residuals of a linear regression of RMSAD on  $\sigma_{bcc/B2}^L$  is plotted to against the VEC of each sample in the training set in Supplementary Fig. 2a, where the residuals show a clear parabolic dependence on VEC with a vertex between VECs of 5 and 6. Based on this observation, we modified VEC into a new descriptor,  $u_{5.7}^{VEC}$ , with the form  $(VEC-X)^2$ , where  $X$  is a constant tuned to be  $5.7 e^-/atom$  by selecting the value that maximizes the correlation coefficient (Pearson's  $r$ ) between the transformed feature and RMSAD (Supplementary Fig. 2b). Similarly, another descriptor,  $\sigma_{5.7}^{VEC}$ , which describes the fluctuation of VEC at different local atomic bonds in the solid-solution lattice. Similar to the feature standard deviation described in Eq. (6),  $\sigma_{5.7}^{VEC}$  is calculated as a modified variance of a discrete probability distribution as follows,  $\sigma_{5.7}^{VEC} = \sum x_i (VEC_i - 5.7)^2$ . Where  $x_i$  is the chemical composition of the constituent element,  $i$ , and  $VEC_i$  is its number of valence electrons. As expected,  $\sigma_{5.7}^{VEC}$  also shows a strong correlation with RMSAD (Supplementary Fig. 2c).

In addition to VEC, the bimodality of the electronic density of state (DOS) of the  $d$ -orbitals is also a key factor to determine the  $d$ -band filling effects<sup>61</sup>. The localized characteristics of  $d$  electrons cause their DOS to display considerable shape features that strongly depend on the lattice structure. For example, a perfect bcc-type lattice generally results in a bimodal DOS of the  $d$ -orbitals. Once the bcc lattice is distorted, such as due to the presence of crystal defects, the shape of the DOS becomes less bimodal<sup>61</sup>. An extreme situation is that the bcc lattice changes to the fcc/hcp (hexagonal-close-packed) lattice through phase transformation. Correspondingly, the shape of the DOS also changes from bimodal to unimodal. More importantly, the change of bimodality can strongly influence the  $d$ -band-filling energy as well as the stability of the crystal lattice when the filling fraction varies. As shown in Fig. 3a, when the  $d$ -band is about half-filled, which corresponds to a VEC between 5 and 6, a bimodal DOS would have more occupied states far below Fermi Level ( $E_f$ ) and fewer occupied states close  $E_f$ , compared to that of the unimodal DOS. This trend leads to a more negative band-filling energy, and





**Fig. 3** Illustration of the difference in the filling energy of the  $d$ -bands with a unimodal and bimodal electronic DOS. **a** The filling fraction is close to the band center. **b** The filling fraction is close to the band edge. The position of the Fermi level is represented by the red dashed line.

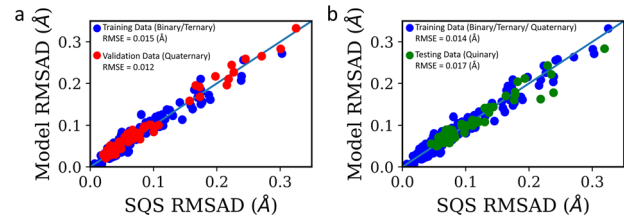
correspondingly make the bcc lattice stable. Inversely, when the filling fraction of the  $d$ -band towards to the edges, a lower band filling energy is expected for the less bimodal DOS, correspondingly stabilizing the non-/distorted-bcc lattice (Fig. 3b). This principle well explains the effect of  $d$ -band bimodality on lattice distortion. For instance, if two elements with fewer  $d$  electrons (e.g., Ti, Zr, Hf) are forced to form an atomic bond in an undistorted bcc lattice, the DOS of the local  $d$ -band between them will then have a bimodal shape but a filling fraction near the left-band edge. As a result, the filling energy of the local  $d$ -band will be high, resulting in a negative effect on the lattice stability. To lower the band-filling energy and stabilize the bcc lattice, a local lattice distortion is thus necessary to reduce the bimodality of the local  $d$ -band DOS. In other words, a stronger local lattice distortion shall be expected between two atoms if their atomic bond in the undistorted lattice has a local  $d$ -band that has a more bimodal DOS and a filling fraction more away from the half-filled state.

The bimodality of a DOS can be quantitatively measured through the Hartigan's dip test<sup>61,62</sup>. A completely unimodal DOS corresponds to a test statistic (i.e., the dip value) of 0, while a more bimodal DOS has a larger dip value. In the present work, the average bimodality of the  $d$ -band DOSs of the FNN atomic bonds ( $u_{B2/bcc}^{\text{dip}}$ ) in a bcc solid-solution alloy at the undistorted state is estimated using Eq. (5) based on the bond-counting approach described in the method section. Considering the positive correlation of lattice distortion with both the filling fraction and bimodality of the local  $d$ -band DOS, we further weigh the two VEC-associated descriptors,  $u_{5,7}^{\text{VEC}}$  and  $\sigma_{5,7}^{\text{VEC}}$  with the bimodality parameter  $u_{B2/bcc}^{\text{dip}}$  for a complete description of the  $d$ -band effect on lattice distortion. By testing the Pearson's correlation with RMSAD, the square of  $u_{B2/bcc}^{\text{dip}}$  (i.e.,  $(u_{B2/bcc}^{\text{dip}})^2$ ) turns to be the optimal weighing factor to result in the largest correlation coefficients.

With the descriptors discussed above, a physically interpretable model is developed to efficiently predict the RMSAD parameter of arbitrary bcc solid-solution alloys within the 10-element composition space studied in the present work. The model expresses RMSAD in a linear form with three terms,

$$\text{RMSAD} = a_1 \sigma_{B2/bcc}^1 + a_2 \left(u_{B2/bcc}^{\text{dip}}\right)^2 u_{5,7}^{\text{VEC}} + a_3 \left(u_{B2/bcc}^{\text{dip}}\right)^2 \sigma_{5,7}^{\text{VEC}} \quad (1)$$

where  $a_1$ ,  $a_2$ , and  $a_3$  are the fitting coefficients obtained by performing ordinary least squares (OLS) regression with the available DFT data. As shown in Fig. 4a, by only regression with the binary and ternary data, the resulting model can already accurately predict the RMSAD of the quaternary alloys in the validation dataset with an RMSE of 0.012 Å, which is comparable to its training RMSE, 0.015 Å. The regressed values of  $a_1$ ,  $a_2$ , and  $a_3$  are summarized in Table 1. Additionally, to confirm the stability and generalization of the model, we re-performed the regression on the data of binary, ternary, and quaternary alloys and tested the correspondingly obtained model with the DFT-calculated RMSAD of quinary alloys. As expected, the values of the fitting coefficients (i.e.,  $a_1$ ,  $a_2$ , and  $a_3$ ) are negligibly varied after



**Fig. 4** Prediction performance of the statistical surrogate model for RMSAD. **a** Trained on binary/ternary data and validated on quaternary data. **b** Trained on binary/ternary/quaternary data and tested on quinary data.

additionally including the quaternary data in the regression, indicating the good stability of the proposed linear model. Additionally, as shown in Fig. 4b, the testing performance of the model on the quinary data is also promising, yielding an RMSE of 0.017 Å over 52 testing compositions. The results of Fig. 4 strongly suggest that the proposed linear model well captures the underlying physical mechanism of RMSAD, which grants the model outstanding capability to make accurate extrapolative predictions for multicomponent systems by only training with a finite amount of binary and ternary data.

It is further interesting to point out that  $a_1$ ,  $a_2$ , and  $a_3$  all have positive values, which is consistent with the physical motivation of the model. First, a positive  $a_1$  indicates that a more severe lattice distortion should be expected in an alloy if it has larger deviations in the lengths of local atomic bonds, which is consistent with the observations in Fig. 2a, d. In the previous classical models of lattice distortion, the standard deviation of the atomic radii of pure elements are commonly adopted to describe the variance of atomic bond lengths by assuming atoms are all rigid particles. This assumption fails to consider the bond-length changes due to the possible electron density overlap between heteroelements, which leads to an overestimation on the lattice distortion in some bcc refractory alloys<sup>37,54</sup>. In the present work, we tackled this problem by using a bond-counting approach (Eqs. (5) and (6) in the methods section). Particularly, the bond length between two heteroelements can be approximated from their binary B2 structure by taking into account the possible charge transfer when forming a bond. Second, a positive value of  $a_2$  is consistent with our above analysis of the  $d$ -band effects on lattice distortion; the lattice distortion in an alloy becomes stronger if the average filling fraction of the  $d$  orbitals of the alloy is more away from half-filled (i.e., a larger value of  $u_{5,7}^{\text{VEC}}$ ) or the shape of the local  $d$ -orbital DOSs is more bimodal (i.e., a larger value of  $u_{B2/bcc}^{\text{dip}}$ ). Third, the positive contribution of VEC standard deviation to lattice distortion is also confirmed by a positive  $a_3$  coefficient. Overall, the proposed model suggests that the lattice distortion of bcc refractory solid-solution alloys is mainly attributed to two key factors: 1) the significance of the variance in the lengths of local atomic bonds; and 2) the  $d$ -band effect described by the filling fraction and bimodality of  $d$ -orbital DOS.

### Prediction of yield strength based on the RMSAD parameter

One of the primary motivations behind developing in-depth understanding of lattice distortion is its physical correlation with solid-solution strengthening, which consequently provides a possible route to design high-strength HEAs. As discussed in the introduction section, ubiquitous lattice distortion over the crystal lattice could introduce significant atomic strains with random spatial fluctuations, which can elastically interact with the strain field of dislocations and provide a pinning force to retard the dislocation movement. Outstanding yield strengths have been generally achieved in the HEAs with severe lattice distortion<sup>3,4,63</sup>. The correlation between the lattice distortion and yield strength has also been discussed previously in a few individual alloy systems,

**Table 1.** Comparison of values of the fitting coefficients when the RMSAD surrogate model (Eq. 1) is trained with two different datasets.

Descriptors	Coefficient Label	Model coefficient values	
		Trained on Binary/Ternary Compositions	Trained on Binary/Ternary/Quaternary Compositions
$\alpha_{B2/bcc}^L$	$a_1$	1.24	1.29
$(u_{B2/bcc}^{dip})^2 \times u_{5,7}^{VEC}$	$a_2$	18.66	19.56
$(u_{B2/bcc}^{dip})^2 \times \sigma_{5,7}^{VEC}$	$a_3$	10.85	10.56

including fcc HEAs composed of 3d transition metal elements<sup>33</sup>, and bcc Nb-based<sup>35,64</sup> and fcc Ni-based solid-solution alloys<sup>34</sup>.

In spite of the great success of these previous works, the correlation has so-far been only quantitatively confirmed for a few equimolar alloy compositions in rather limited compositional spaces. Our surrogate model of lattice distortion provides an opportunity for performing a more comprehensive assessment in a much broader space, because the model enables a rapid prediction of RMSAD for any given alloy compositions with the studied ten refractory elements. Towards this goal, we have tried to collect an exhaustive set of the room-temperature experimental yield strength and hardness data from a recent HEA database developed by Citrine Informatics and other literatures<sup>2,19,20,65–95</sup>. In total 83 yield strength and 28 hardness data were collected, which covered 66 individual compositions (30 equiatomic, and 33 non-equiatomic cases), ranging from quaternary to senary systems (Supplementary Table 2). The data collection is limited to as-cast and annealed alloys, so the strengthening effects due to processing such as grain refinement and strain hardening can be minimized. The detail of our data-collection process is described in the methods section. Then, regarding each of the experimental data, the surrogate model is employed to predict RMSAD according to the alloy composition in order to investigate the potential correlation between the lattice distortion and yield strength.

A modified theory of solid-solution strengthening for random alloys has been recently developed by Maresca and Curtin<sup>28</sup>. The theory expresses the athermal yield strength ( $\tau_Y$ ) of a bcc solid-solution alloy as,

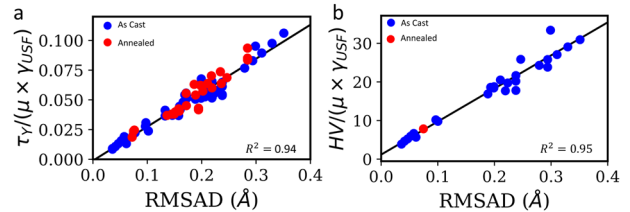
$$\tau_Y = 0.051a^{-\frac{1}{3}}\mu \left(\frac{1+\nu}{1-\nu}\right)^{\frac{4}{3}} f^{\frac{2}{3}} \left[\frac{\sum_n c_n \Delta V_n^2}{b^6}\right]^{\frac{2}{3}} \quad (2)$$

In the equation,  $a$  is the dislocation line-tension parameter, which is a constant for a given type of dislocation, and  $\left(\frac{1+\nu}{1-\nu}\right)^{\frac{4}{3}}$  is a term of elastic anisotropy calculated by the Poisson's ratio ( $\nu$ ), which is usually insensitive to the variation of alloy compositions<sup>96</sup>. From the equation, one can consider the variation of the athermal yield strength with alloy composition is dictated by three parts of contributions, the isotropic shear modulus of the alloy,  $\mu$ , an effective parameter describing the gradient of dislocation stress field,  $f^{\frac{2}{3}}$ , and the volume misfit quantity,  $\left[\frac{\sum_n c_n \Delta V_n^2}{b^6}\right]^{\frac{2}{3}}$ , which is actually closely related to the  $\delta$  parameter, a commonly-used parameter to describe lattice distortion based on the mismatch of atomic radii. Recently, the volume-misfit quantity has also been found to be strongly correlated with the RMSAD parameter in several HEA systems<sup>55</sup>, correspondingly evidencing the correlation of the yield strength with lattice distortion.

Therefore, inspired by Eq. (2) above, here we propose a simplified linear expression based on RMSAD to effectively model and predict the room-temperature yield strength of bcc refractory HEAs, which can be written as,

$$\tau_Y \approx a\mu\gamma_{USF} \times \text{RMSAD} \quad (3)$$

where  $\mu$  is still the isotropic shear modulus,  $\gamma_{USF}$  is the average unstable stacking fault (USF) energy of the {110} plane of the alloy, and  $a$  is a constant coefficient, which should be universal to



**Fig. 5 Linear dependence of mechanical properties of RHEAs on the RMSAD parameter.** **a** Normalized yield strength ( $\tau_Y$ ) vs. RMSAD, and **b** Normalized Vicker's hardness (HV) vs. RMSAD. The normalization factor is a product of the alloy's isotropic shear modulus ( $\mu$ ) and USF energy ( $\gamma_{USF}$ ). The yield strength and hardness data are collected from the previous experimental measurements in the literature. The unit of the y axis is  $\frac{\text{GPa}}{(\text{GPa} \times \frac{1}{\text{m}^2})} = \text{m}^2/\text{J}$  in **(a)** and  $(\text{HV} * \text{m}^2)/(\text{GPa} * \text{J})$  in **(b)**.

different alloy compositions and systems. The derivation of Eq. (3) is based on the three main contributions to yield strength in Eq. (2), where we use RMSAD to parameterize the local atomic misfit strain generated by lattice distortion and use the USF energy ( $\gamma_{USF}$ ) to substitute  $f^{\frac{2}{3}}$  to reflect the gradient of the dislocation-stress field. This is because the derivation of  $f^{\frac{2}{3}}$  is usually a complex process involving extensive data fitting, and in the meanwhile, the gradient of the dislocation-stress field should be generally proportional to the width of the dislocation core, which essentially relates to the USF energy according to the Peierls-Nabarro model<sup>96</sup>. Moreover, it is noteworthy that the three parameters,  $\mu$ ,  $\gamma_{USF}$ , and RMSAD, in Eq. (3) at a given alloy composition can be respectively predicted using Vegard's law<sup>97,98</sup>, a statistical learning model developed by us previously<sup>56</sup>, and the present lattice distortion model, without requiring performing any additional time-consuming calculations. Therefore, Eq. (3) naturally offers a high-throughput prediction approach of yield strength for efficient screening over large compositional spaces.

To validate Eq. (3), experimental yield strength data is normalized by the isotropic shear modulus and USF energy to evaluate its linear dependence on RMSAD. As shown in Fig. 5a, a linear relationship is observed between the normalized yield strength, denoting as  $\tau_Y/(\mu\gamma_{USF})$ , and RMSAD, giving an  $R^2$  value as high as 0.94. Additionally, the value of the constant,  $a$ , in Eq. (3) can be easily obtained via an OLS regression of  $\tau_Y/(\mu\gamma_{USF})$  on RMSAD, which is  $0.29 \text{ \AA}/\text{eV}$ . An RMSE of 135 MPa was obtained by comparing the predicted yield strength predicted with the experimental truth. Additionally, it is well known that the hardness of most alloys shows a linear dependence on yield strength in general. Therefore, Eq. (3) should also be capable to model the hardness of bcc refractory HEAs, if it truly sketches the underlying physics of solid-solution strengthening. As expected, a clear linear relationship is also observed between the normalized hardness,  $\text{HV}/(\mu\gamma_{USF})$ , and RMSAD, returning a RMSE of prediction as small as 29.5 HV (Fig. 5b).

### Data mining and interpretation

Based on the developed surrogate models, a systematic data mining over massive alloy compositions is performed to uncover

the physical relations of lattice distortion to phase stability, strength, and ductility of RHEAs. First, the distribution of lattice distortion in bcc refractory alloys is mapped in a compositional space composed of ten major RHEA elements. Second, with the mapping results of lattice distortion, the room-temperature yield strengths of those mapped compositions are also predicted, using Eq. (3). Third, the predicted yield strength is leveraged with a previously developed ductility model<sup>56</sup> and the CALPHAD approach<sup>99</sup> to validate an integrated computational framework for the data-driven design of RHEAs with high strength and good ductility.

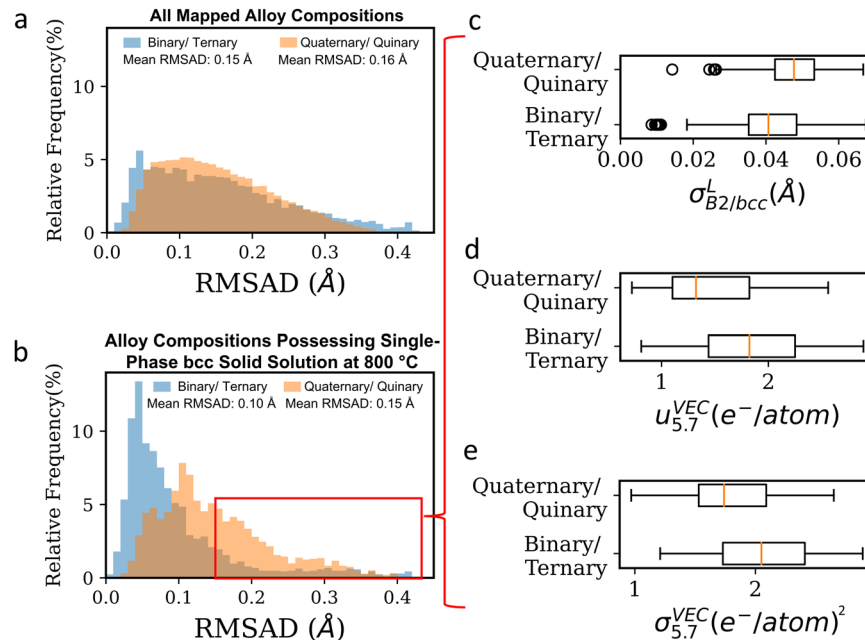
### Systematic mapping of lattice distortion

Mapping of lattice distortion is performed for various binary, ternary, quaternary, and quinary alloys in a vast compositional space consisting of ten different elements, namely Ti, Zr, Hf, V, Nb, Ta, Mo, W, Re, and Ru. Compositions considered are limited such that every individual element composes between 10 and 60 atomic percent (at%) to maintain focus on the concentrated, high-entropy region of the compositional space. For a homogenous sampling, the binary, ternary, and quaternary alloys are assessed at 5 at% compositional intervals with each element while a 10 at% interval is applied to the quinary alloys. As such, a total of 76,816 compositions are investigated. Of those, 140 compositions are binary alloys; 5208 are ternary; 36,400 are quaternary; and 35,068 are quinary. At each mapped composition, we also utilize the CALPHAD approach<sup>99</sup> to evaluate the phase stability of the single-phase bcc solid solution at 800 °C, which is a common temperature used for the homogenization processing of RHEAs. After the phase-stability evaluation, 5916 compositions are found to result in single-phase bcc solid solutions at 800 °C. Among them, 71 are binary; 1989 are ternary; 2012 are quaternary; and 1844 are quinary.

The mapping results of lattice distortion are shown in Fig. 6a, b where relative frequency histograms showing the distribution of RMSAD for all the screened alloy compositions and those possessing a single bcc phase at 800 °C, respectively. The compositions are grouped into binary/ternary (low/medium

entropy) and quaternary/quinary (high entropy) categories to highlight the effect from the number of component elements, i.e., the cocktail effect of HEAs. In Fig. 6a, it can be seen that the RMSAD distributions are nearly identical between the binary/ternary and the quaternary/quinary groups. The mean values of the distributions are also similar, which are 0.1544 Å for the low/medium-entropy alloys and 0.1573 Å for the high-entropy alloys. However, when the solid-solution phase stability is considered, obvious difference is observed in the RMSAD distributions between the low/medium-entropy and high-entropy alloys. As presented in Fig. 6b, with the single-phase constraint, the RMSAD distribution of the low/medium-entropy group is biased towards the compositions with less lattice distortion. This feature is also evidenced by the increased skewness, which more than doubled (0.70 vs. 1.92) and reduced distribution mean (0.1544 vs. 0.0974 Å). In contrast, the distribution of RMSAD for the high-entropy group (quaternary/quinary alloys) does not show significant bias after applying the phase-stability restriction, which still holds a mean RMSAD as high as 0.1502 Å and a weak skewness. Comparing the results of Fig. 6a, b suggests that while highly distorted binary and ternary bcc solid-solution alloys are predicted to be possible, a majority of them cannot remain thermodynamically stable at typical processing temperatures. On the other hand, the quaternary and quinary alloys with severe lattice distortions are likely to be stabilized by the increased configurational entropy from the inclusion of multiple principal elements. There has been an enduring debate whether the severe lattice distortion is a feature unique to HEAs or it can be generally existing in any kind of solid-solution alloys regardless compositional complexity<sup>100</sup>. The present observation clearly evidences the importance of compositional complexity in preserving solid-solution alloys with severe lattice distortion from phase decomposition.

To uncover the mechanism of the HEA's cocktail effect on lattice distortion, the single-phase-alloy compositions with a RMSAD larger than 0.15 Å are collected for a further investigation, as marked by the red box in Fig. 6b. Here, 0.15 Å is chosen as an indicator for severe lattice distortion, as it represents the 84th



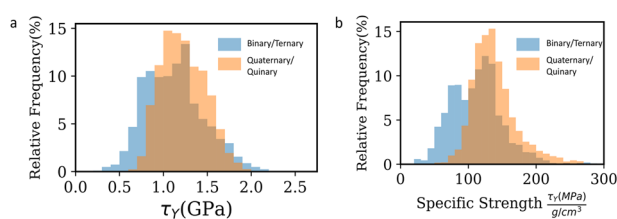
**Fig. 6** Distribution of lattice distortion in bcc refractory solid-solution alloys. Relative frequency histograms of lattice distortion in the alloys (a) before and (b) after screening for single-phase stability at 800 °C. The histogram colored in blue corresponds to the distribution of the RMSAD parameter of the binary and ternary alloys, while the histogram associated with the quaternary and quinary alloys is colored in orange. c–e Boxplots describing the distribution of physical descriptors that contribute to RMSAD in the alloys with severe lattice distortion (i.e., alloys with RMSAD > 0.15 Å).



percentile of the binary/ternary distribution and the 58th percentile of the quaternary/quinary distribution. In this collective dataset, a comparison between the binary/ternary and quaternary/quinary alloys could clearly elucidate the effects of compositional complexity on severe lattice distortion. According to Eq. (1), a higher degree of lattice distortion can be achieved in an alloy that has a larger deviation in local atomic bond lengths ( $\sigma_{B2/bcc}^L$  in Eq. (1)), a VEC more away from  $5.7 e^-/\text{atom}$  ( $u_{5.7}^{\text{VEC}}$  in Eq. (1)), and/or a large variability of the number of valence electrons ( $\sigma_{5.7}^{\text{VEC}}$  in Eq. (1)) among the constituent elements. The distributions of the  $\sigma_{B2/bcc}^L$ ,  $u_{5.7}^{\text{VEC}}$ , and  $\sigma_{5.7}^{\text{VEC}}$  parameters in the binary/ternary and quaternary/quinary alloys with RMSAD > 0.15 Å are visualized via boxplots, as exhibited in Fig. 6c, d, and e, respectively. It is interesting to note that the quaternary/quinary alloys generally have a larger  $\sigma_{B2/bcc}^L$  than the binary/ternary alloys, but vice versa for the  $u_{5.7}^{\text{VEC}}$  parameter. Also, the aforementioned differences are demonstrated to be statistically significant via Welch's t-tests<sup>101</sup>. This observation suggests that the severe lattice distortion in the quaternary/quinary alloys is more attributed to the deviations in local atomic bond lengths of the constituent elements. However, significant differences in the valences and atomic volumes of the constituent element in a solid-solution phase also leads to a strong driving force for phase decomposition, according to the classic Hume-Rothery rule. Thanks to the multicomponent chemistry, such a driving force in the quaternary and quinary alloys could be balanced by the stabilization effect induced by the increased configurational entropy. As a result, those alloys can still remain as a single-phase solid solution while bearing a large lattice distortion. On the other hand, it is understandable that the severe lattice distortion in the binary and ternary alloys is more associated with a larger  $u_{5.7}^{\text{VEC}}$  parameter (Fig. 6d), since too large  $\sigma_{B2/bcc}^L$  would introduce phase decomposition, which cannot be sufficiently balanced by the entropy stabilization effect. In fact, the  $u_{5.7}^{\text{VEC}}$  parameter cannot be too large either because the bcc lattice structure becomes unstable for a transition-metal alloy if its  $d$  orbital band is away from the half-filled state<sup>102</sup>. Therefore, a majority of screened binary and ternary alloys with severe lattice distortion are thermodynamically unstable.

### Lattice distortion induced solid solution strengthening

With the results of the RMSAD screening above, we could further evaluate how the yield strengths of the bcc refractory alloys distribute across a large compositional space using the developed strengthening model (Eq. (3)). Here we restrict the yield-strength prediction to the same alloy compositions in Fig. 6b, which are predicted to be stable as a single bcc solid-solution phase at 800 °C. The compositions with RMSAD less than 0.01 Å, however, are omitted. It is important to note that this exclusion would only disproportionately affect a select few dilute compositions in the Mo-W binary series, which tend to exhibit low lattice distortion and yet could potentially still be strong due to a high lattice friction stress, as reflected by their shear moduli (Supplementary Table 3). The predicted yield strengths are visualized as distribution histograms to reveal the difference between the binary/ternary and quaternary/quinary alloys. As shown in Fig. 7a, the quaternary/quinary alloys generally possess a higher yield strength and a narrower distribution than those of the binary/ternary alloys. The higher average strength of the quaternary/quinary alloys is likely to be a benefit of higher levels of lattice distortions in those alloys (Fig. 6b), which contributes positively to yield strength according to Eq. (3). Additionally, in our strengthening model, the yield strength of a bcc refractory alloy is also proportional to its shear modulus and USF energy, which are two material features showing strong dependences on VEC. As exhibited in Supplementary Fig. 3, an alloy with a lower VEC that is



**Fig. 7** Distribution of yield strength in the bcc refractory alloys screened to be stable as a single-phase solid solution at 800 °C. Relative frequency histograms of (a) room-temperature yield strength and (b) room-temperature specific yield strength.

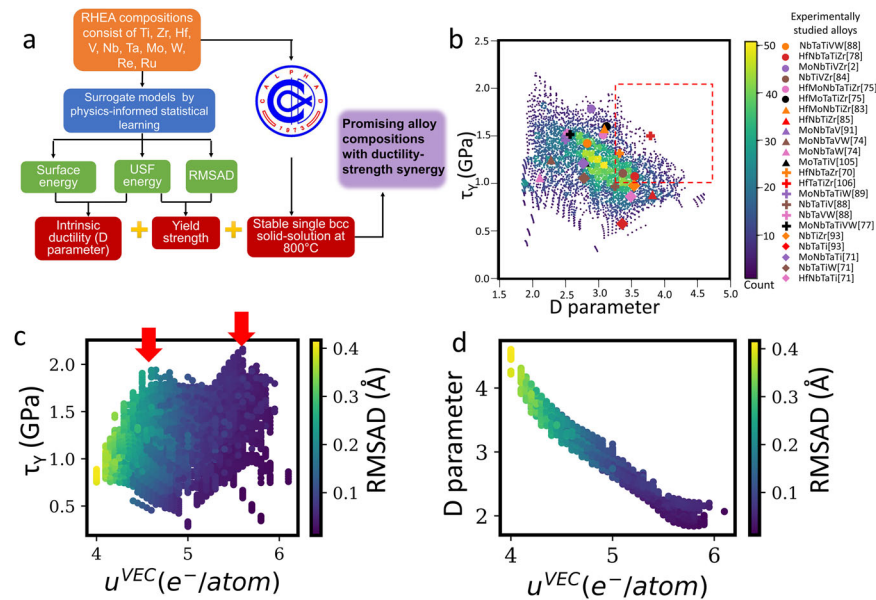
away from a half-filled  $d$  band state generally has a smaller  $\mu$  and  $\gamma_{\text{USF}}$ . Therefore, to achieve a high yield strength in an alloy would require it to have a large RMSAD and meanwhile not a too small VEC. This combination of features is more possible to achieve in the quaternary/quinary alloys due to the entropy-stabilization effect.

The potentials of RHEAs for light weighting applications are also evaluated through a prediction of specific strength, which is achieved by combining the yield strength data obtained above with a lattice-parameter model developed in our recent work for the prediction of theoretical density<sup>103</sup>. The predicted specific strengths are also overviewed, using the distribution histograms, as shown in Fig. 7b. It is exciting to note that the HEAs also averagely yield higher specific strengths than the alloys with low/medium configurational entropies. The results of Fig. 7 here suggest that more opportunities of light-weight, high-strength alloy compositions shall be expected by increasing the number of components through the “high-entropy” design strategy.

### Data-driven design of high-performance RHEAs

The current screening model for alloy's yield strength can be seamlessly integrated with a ductility assessment model developed in our previous work for a data-driven exploration of strong and ductile bcc refractory alloys. Taking alloy composition as input, the ductility model yields rapid prediction for a so-called  $D$  parameter, which is defined as the ratio between the surface energy of a crack fracture plane and unstable stacking fault energy of the common dislocation-slip plane. The magnitude of the  $D$  parameter physically reflects the likelihood of an alloy to be intrinsically ductile, based on the Rice fracture mechanics<sup>104</sup>. More importantly, the  $D$  parameter of RHEAs has been demonstrated to positively correlate with the compression fracture strains measured experimentally, which makes it an effective indicator to assess alloy's ductility<sup>56</sup>. A combination of the strength and ductility models with the CALPHAD-based assessment of phase stability forms a data-driven design framework that is powerful for performing massive searches in a vast design space for promising alloy compositions with the enhanced ductility and strength. Figure 8a shows the flow chart of the framework. With an input alloy composition, the CALPHAD-based thermodynamic assessment is first carried out to check whether the bcc solid solution can be stable as a single phase at a given processing temperature. If yes, the strength and ductility models are sequentially applied to assess the mechanical properties.

As a demonstration, we applied this framework to the same composition dataset used for our lattice-distortion mapping, which contains 76,816 compositions homogeneously sampling all the binary, ternary, quaternary, and quinary systems in the 10-component compositional space. The results are displayed by an Ashby plot of the strength-ductility combination in Fig. 8b. Positions of some experimentally-studied RHEAs<sup>2,19,65,67,68,70–75,77,78,80,82–95,105</sup> are also highlighted to manifest the potential area to explore. As shown in Fig. 8b, despite a weak trend of the typical strength-ductility trade-off, the quite dispersed distribution between yield strength and the  $D$  parameter illuminates a range of RHEAs that can possess a good combination of strength and ductility. If we consider



**Fig. 8** **A high-throughput framework for screening of RHEAs with a good combination of ductility and strength.** **a** A flowchart detailing the screening process of the framework. The RMSAD and yield strength are predicted, using the surrogate models developed in the present work, while intrinsic ductility is predicted from the models in our previous work<sup>56</sup>. The phase stability assessment is performed using the TCHEA4 database implemented in the Thermo-Calc Software<sup>99</sup>. **b** A scatter histogram of the yield strength ( $\tau_y$ ) and intrinsic ductility (D parameter) of the screened alloys that are single-phase solid solution at 800 °C. The data points that overlap with more other points within a bin area are coded by a warmer color. Experimentally studied RHEAs are also identified to show where the current experimental literature on equimolar compositions exists and where there is potential to optimize properties by exploring non-equimolar compositions. The red-dashed box illuminates a range of previously unexplored chemical compositions promising for achieving a single-phase RHEAs with a combination of high strength ( $\tau_y > 1$  GPa) and reasonable ductility (D parameter  $> 3.3$ ). Scatterplots showing effect of VEC on **(c)** yield strength and **(d)** the D parameter of single-phase RHEAs. In **(c)** and **(d)**, RMSAD is color mapped onto points to show how VEC and RMSAD can be optimized in tandem for high-performance alloy design.

the alloys with a D parameter larger than 3.3 being potentially ductile, as it generally corresponds to a compression failure strain greater than 40%, there are still a substantial proportion of alloy compositions (marked by a dashed box) could have yield strength greater than 1 GPa but are not fully experimentally explored yet. A recently discovered RHEA, HfTaTiZr, falls in this window, which was indeed measured to have a yield strength about 1.5 GPa and be ductile in tension deformation with limited plasticity, in a good agreement with the present prediction<sup>67</sup>. Additionally, the alloy's ductility was found to be significantly improved via proper heat-treatment processing<sup>106</sup>. Furthermore, the scattered distribution of the experimentally reported equimolar alloys in Fig. 8b indicates that maximizing configurational entropy may not always be an optimal design strategy for HEAs and there could be considerable spaces to tune alloy compositions for the improved strength and ductility combination relative to the currently known RHEAs.

Several practical insights are also derived from the current screening, which may enlighten the future design of RHEAs. VEC is unsurprisingly found to be quite influential for both yield strength and ductility (i.e., D parameter). The yield strength and D parameter of the screened alloys is plotted against their VEC in Fig. 8c, d, respectively, with data points color coded corresponding with RMSAD. As shown in Fig. 8c, the yield strength seems to peak at two VEC values, 4.7 and 5.5 e<sup>-</sup>/atom, respectively. As discussed in Eq. (3), the yield strength is a function of RMSAD,  $\mu$ , and  $\gamma_{USF}$ . The relation between RMSAD and VEC is discussed in Fig. 2b with lower VECs corresponding with higher RMSAD. When plotted as a function of VEC (Supplementary Fig. 3),  $\mu$ , and  $\gamma_{USF}$  show a positive correlation with VEC. Based on these observations, The peak at 4.7 e<sup>-</sup>/atom is attributed to increased lattice distortion and the peak at 5.5 e<sup>-</sup>/atom is more likely to correspond to higher  $\mu$ , and  $\gamma_{USF}$ . On the other hand, the D parameter shows a negative dependence on VEC in general, which means that the alloys with larger VEC are likely have poor

ductility due to a smaller D parameter. Particularly, regarding the alloys near the yield strength peak with VEC around 5.5 e<sup>-</sup>/atom, their D parameter is only about 2, which is a value exhibited by the alloys showing brittle deformation behaviors<sup>56</sup>. Therefore, a good balance between the strength and ductility may be achieved by tuning the alloy's VEC to be around 4.7 e<sup>-</sup>/atom. Furthermore, the lattice distortion also positively affects both the strength and ductility. For example, as shown in Fig. 8d, alloys with a larger RMSAD generally have a larger D parameter. Also, at a certain level of VEC, increasing the alloy's RMSAD leads to a further enhancement of the yield strength, as shown by the colormap in Fig. 8c. These observations suggest that a good blend of strength and ductility can be expected if the alloy is designed to have a VEC close to 4.7 e<sup>-</sup>/atom and constitute elements that having large variances in their number of valence electrons and bcc/B2 bond lengths to maximize the lattice distortion (i.e., RMSAD).

In summary, a statistical surrogate model was developed to efficiently predict lattice distortion in bcc refractory non-dilute, high-entropy solid solution alloys with ten constituent elements: Ti, Zr, Hf, V, Nb, Ta, Mo, W, Re, and Ru. The lattice distortion quantified by the RMSAD parameter was previously only obtained for a few individual alloys by costly experiments or calculations but can now be rapidly predicted by the developed model in a high-throughput manner simply according to alloy composition. Unlike the conventional models based on atomic size mismatch of pure elements, the present model uses descriptors that characterize the length and electronic structure of interatomic bonds in solid-solution alloys for a more accurate description of lattice distortion. With physics-informed statistical learning, the model exhibits a promising capability to make reliable extrapolative predictions for the multicomponent alloy compositions outside the training data range. Moreover, a strong correlation between the lattice distortion and the room-temperature yield strength is



revealed in bcc RHEAs. Accordingly, a surrogate model enlightened by the recent solid-solution strengthening theory<sup>28</sup> is proposed for rapid prediction of yield strength of RHEAs simply according to their compositions.

The developed surrogate models are applied to screen over a massive amount of alloy compositions to systematically investigate the compositional dependence of lattice distortion and yield strength in RHEAs. The multicomponent nature (i.e., cocktail effect) of RHEAs is revealed to play a critical role in preventing the highly distorted solid-solution lattice from phase transformation. As a result, severe lattice distortion is found to generally be more achievable in the alloys with higher chemical complexity due to increased configurational entropy. The increased level of lattice distortion also benefits RHEAs to generally have a higher strength than their binary and ternary counterparts. Integration of the developed models with a model of intrinsic ductility<sup>56</sup> and phase stability assessment based on the CAPHAD approach enables an alloy design platform for high-throughput search of high-performance RHEAs in a large compositional space that is far from being fully explored. The high-throughput search illuminates a range of previously unexplored chemical compositions promising for achieving a single-phase RHEAs with a combination of high strength (>1 GPa) and reasonable ductility. The present work demonstrates a promising route to combine physical principles with statistical approaches for developing efficient surrogate models to accelerate the pace of HEAs' design and development, especially for those with non-equimolar compositions.

## METHOD

### Details of DFT calculations

In the present work, DFT calculations were performed, using the projector augmented wave method<sup>107</sup> with the exchange-correlation functional depicted by the general gradient approximation from Perdew, Burke, and Ernzerhof (GGA-PBE)<sup>108</sup>, as implemented in the Vienna ab-initio simulation package (VASP)<sup>109</sup>. The calculation parameters used were the same as those employed in our previous work<sup>56</sup> with a plane-wave basis energy cutoff of 400 eV, first-order Methfessel Paxton smearing of 0.2 eV, and using k-point grids generated by the VASP's automatic meshing scheme with an  $R_k$  length of 30 Å. An Energy convergence criterion of  $10^{-6}$  eV was used for electronic self-consistency cycles. The atomic structural relaxations were performed by minimizing the atomic forces to be smaller than 0.001 eV/Å. The aforementioned settings of the k-point density and smearing width were validated by convergence test calculations in our previous work<sup>56</sup>. The supercells used for DFT calculations were constructed, using the SQS approach to approximate the chemical disorder in solid-solution alloys<sup>47</sup>. For binary, ternary, and quaternary alloy compositions, the generated SQSs contains 72 atoms, and 120-atom SQSs were employed for quinary compositions. Overall, as listed in Supplementary Table 1, the relaxed atomic structures of 69 binary, 36 ternary, 64 quaternary, and 52 quinary bcc refractory alloys were calculated from a compositional space comprised of 10 group IV–VIII elements, including Ti, Zr, Hf, V, Nb, Ta, Mo, W, Re, and Ru. Among the DFT results, 106 of the compositions were adopted from our previous work<sup>56</sup>.

### Scrutiny of bcc-to- $\omega$ phase transformation in training data

A bcc-based SQS for refractory alloys with low VECs could be dynamically unstable and transform into a lattice closer to the crystal structure of the  $\omega$  phase after relaxation<sup>40,41,110</sup>. To ensure that no  $\omega$ -like structures were included in the training data, an approach developed by Ikeda et al.<sup>41</sup> was used. The bcc-to- $\omega$  transformation has the martensitic nature and can be caused by atomic shuffling of two-third atoms in the bcc lattice along a  $111_{\text{bcc}}$  direction. Therefore, a SQS of a multicomponent alloy can be considered as being " $\omega$ -like" if the atomic displacements

caused by relaxation display a preference for a particular  $111_{\text{bcc}}$  direction whereas there should be no significant preference if the bcc symmetry is preserved<sup>41</sup>. The preference can be quantified through a four-step analysis. First, the atomic displacement of each atom in the supercell is projected along four distinct  $[111]_{\text{bcc}}$ ,  $[\bar{1}\bar{1}\bar{1}]_{\text{bcc}}$ ,  $[\bar{1}\bar{1}1]_{\text{bcc}}$  and  $[1\bar{1}\bar{1}]_{\text{bcc}}$  directions, respectively. Second, the projection magnitude for each  $111_{\text{bcc}}$  direction is averaged over all the atoms in the supercell. Third, the average projection magnitudes along different  $111_{\text{bcc}}$  directions are sorted based on their values. Fourth, a structural descriptor is defined by taking the difference between the largest projection magnitude ( $p_1$ ) and second largest projection magnitude ( $p_2$ ), i.e.,  $p_1 - p_2$ . A relaxed SQS is considered as being " $\omega$ -like" if  $p_1 - p_2$  is larger than a threshold value,  $0.04c_\omega$ , where  $c_\omega$  is the lattice parameter of the c axis of a perfect  $\omega$  crystal which is calculated as a function of the bcc lattice parameter ( $a_{\text{bcc}}$ ) with the relationship:  $c_\omega = \frac{\sqrt{3}}{2}a_{\text{bcc}}$ . If  $p_1 - p_2$  is smaller than  $0.04c_\omega$ , the bcc symmetry is preserved. By implementing Ikeda's method, we found that concerning the SQSs used in the present work no structures exceeded this threshold. The calculated  $p_1 - p_2$  of each SQS is shown in Supplementary Fig. 4. In addition, NbTi<sub>3</sub> was found to be an outlier relative to the other compositions and was thus removed from subsequent training process as a precautionary measure.

### RMSAD parameter

In the present work, the RMSAD parameter was derived from DFT relaxation calculations. Specifically, the RMSAD parameter was calculated by taking the root-mean square average of the Euclidian distances between atomic positions in the relaxed SQS supercell and their corresponding ideal positions in the perfect bcc lattice, which can be expressed as,

$$\text{RMSAD} = \sqrt{\frac{\sum_i^n (x_i^{\text{real}} - x_i^{\text{ideal}})^2}{n}} \quad (4)$$

where  $n$  is the number of atoms in the supercell,  $x_i^{\text{ideal}}$  is the ideal lattice site for atom,  $i$ , and  $x_i^{\text{real}}$  is the DFT relaxed position of atom,  $i$ . Since the volume of the supercell can also change after relaxation, the ideal position of each atom was attained by scaling the initially unrelaxed supercell to have the same atomic volume as the relaxed one. The Euclidian distance was then calculated in Cartesian coordinates. As such, the calculated Euclidian distance should only reflect the atomic-position deviation due to the local lattice distortion rather than that caused by the change in the volume of the supercell after relaxation.

### Physics-informed descriptors

The selection of the physical feature to construct descriptors was driven by a data exploration described in the result section. It was found that lattice distortion was largely determined by interatomic-bonding characteristics like the first nearest neighbor (FNN) bond length and valence electron concentration (VEC) and their variations due to the local chemistry fluctuation caused by random mixing. Therefore, other than VEC, which has a simple derivation, the rest of descriptors were constructed using a bond-counting-based method developed in our previous work, by which the physical characteristics of interatomic bonds in a random alloy can be effectively integrated with its composition<sup>56</sup>.

According to the bond-counting method<sup>111,112</sup>, a bulk property of an alloy, such as the cohesive energy, can be approximated as a summation of the individual contributions from each atomic bond in the alloy. In random alloys, the probability of finding an interatomic bond between two constituent elements,  $i$  and  $j$ , should be statistically equal to the product of the concentration of  $i$  and  $j$ . Therefore, the summation of individual bond contribution can be considered as a weighted average over different  $i$ - $j$  bonds in the alloy, and the weighting factor is the presence probability of

each type of  $i$ - $j$  bonds. Accordingly, the descriptor can be written as,

$$u_{\text{B2/bcc}}^p = \sum_i x_i \sum_j x_j p_{ij} \quad (5)$$

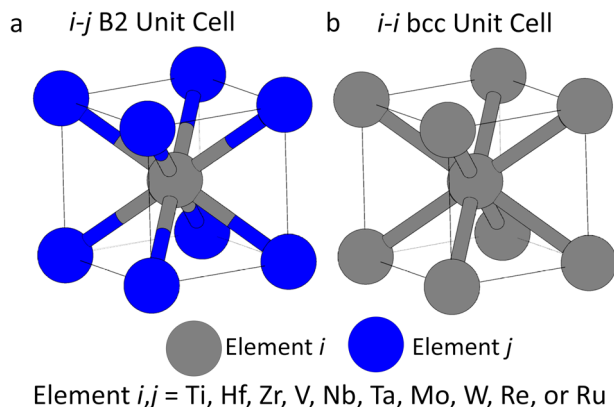
where  $x_i$  and  $x_j$  are the chemical compositions of elements,  $i$  and  $j$ , respectively, which are among the 10 refractory elements studied in the present work (i.e., Ti, Zr, Hf, V, Nb, Ta, Mo, W, Re, and Ru).  $u_{\text{B2/bcc}}^p$  is a descriptor based on an interatomic-bonding feature,  $p$ , which has a value of  $p_{ij}$  for the  $i$ - $j$  type of atomic bond. Clearly, all values of  $p_{ij}$  can be summarized as a  $10 \times 10$  matrix in which each component stands for a pair permutation between the ten alloying elements. Additionally, to reflect the fluctuations in local atomic environment, the standard deviation of  $u_{\text{B2/bcc}}^p$  is also taken as a descriptor with an expression as,

$$\sigma_{\text{B2/bcc}}^p = \sqrt{\sum_i x_i \sum_j x_j (p_{ij} - u_{\text{B2/bcc}}^p)^2} \quad (6)$$

To derive the values of  $p_{ij}$  is not trivial. Assuming that the atomic interactions between the atoms as FNN are much stronger than others, we could apply the physical properties of a group of single-element bcc and ordered binary B2 structures as the bond feature parameter. For example, if we consider the FNN bond length as a bond-feature parameter, its  $p_{ij}$  component thus to be the length of the FNN  $i$ - $i$  bond in the single-element bcc structure composed of element,  $i$ , while the  $p_{ij}$  and  $p_{ji}$  components are the FNN bond lengths of the binary B2 structure composed of elements,  $i$  and  $j$ . This assumption is valid because the only type of the FNN atomic bond in a bcc crystal purely comprised of element,  $i$ , is the  $i$ - $i$  bond along the  $\langle 111 \rangle$  direction, and the only type of the FNN bond in the binary B2 structure composed of elements,  $i$  and  $j$ , is the  $i$ - $j$  bond, as illustrated in Fig. 9. Since there are only two atoms in the unit cell of the single-element bcc and ordered binary B2 structures, DFT calculations can be easily applied to these structures to predict the values of  $p_{ij}$  without demanding extensive calculations. In the present work, we derived two types of bond-feature parameters from the DFT calculations, which are the FNN bond length ( $L$ ) and an electronic structure parameter (dip) measuring the bimodality of the local DOS of the  $d$  orbitals in the bcc/B2 structures. The dip parameter was designed to account for the nuance of the  $d$ -band filling effects in bcc transition metals beyond VEC, and demonstrated to great success capturing the effects of interatomic bonding on defect energetics of refractory alloys in our previous work<sup>56,61</sup>.

### Data splitting and statistical regression

Due to the high computational cost, the DFT-SQS calculations were only able to generate a finite amount of RMSAD data for 215



**Fig. 9** Unit cells of the crystal structures used to model the physical features of the first nearest neighbor (FNN) atomic bonds in bcc solid-solution alloys. **a** binary B2 unit cell and **b** single element bcc unit cell.

individual alloy compositions, which is only a tiny fraction of the full composition space studied in the present work. Therefore, to wisely leverage this small set of DFT data for reliable extrapolative predictions, we did not randomly split them into training and testing datasets. Instead, the obtained DFT data were split into three datasets based on the chemical complexity for training, validation, and testing. Specifically, only the data of binary and ternary alloys were used as training data, quaternary ones as the validation data to determine the best descriptors, and quinary ones were reserved exclusively for a final testing to examine the regression stability of the finalized model and its prediction ability for unseen compositions. This data-splitting strategy allows us to maximally test the prediction capability of the built surrogate models for extrapolative exploration of multicomponent alloy compositions not covered by the training set. Based on the developed descriptors, a modified OLS regression process was employed to develop a predictive model for the lattice distortion (i.e., the RMSAD parameter) in refractory solid-solution alloys. Specifically, the model was constructed piecewise, starting with the descriptors with the strongest correlation, and additional descriptors were sequentially added until no significant improvement in the prediction performance can be achieved.

### Collection of the experimental strength data of RHEAs

To explore the deep connection between the lattice distortion and solid-solution strengthening in refractory HEAs, a set of experimental yield-strength and hardness data were collected from a HEA database recently developed by Citrine Informatics and other literatures<sup>2,19,20,65–95</sup>. Specifically, data points were filtered to collect the measurements made at room temperature on bcc single-phase HEAs with constituent elements among the ones studied in the present work. The collection is limited to as-cast and annealed alloys to minimize the extrinsic strengthening effects due to processing such as strain hardening and grain refinement and to compression yield strengths because of a relative lack of tensile tested data. Overall, 83 yield-strength and 28 hardness data were collected, which covered 66 individual alloy compositions, ranging from quaternary to senary systems, as summarized in Supplementary Table 2. For each of the assessed experimental data, its corresponding RMSAD parameter is predicted by the surrogate model developed in the present work for exploring the correlation between lattice distortion and yield strength.

### DATA AVAILABILITY

All the data associated with the findings of this work is available to the public through an open-access repository, GitHub (<https://github.com/tandocca/Lattice-Distortion>) as well as in the supplementary information file associated with this manuscript.

### CODE AVAILABILITY

All the code associated with the findings of this work is available to the public through an open-access repository, GitHub (<https://github.com/tandocca/Lattice-Distortion>).

Received: 31 October 2022; Accepted: 6 March 2023;

Published online: 05 April 2023

### REFERENCES

- Lee, C. et al. Strength can be controlled by edge dislocations in refractory high-entropy alloys. *Nat. Commun.* **12**, 1–8 (2021).
- Zhang, Y., Yang, X. & Liaw, P. K. Alloy design and properties optimization of high-entropy alloys. *Jom* **64**, 830–838 (2012).
- Zou, Y., Ma, H. & Spolenak, R. Ultrastrong ductile and stable high-entropy alloys at small scales. *Nat. Commun.* **6**, 1–8 (2015).

4. Lee, C. et al. Lattice-distortion-enhanced yield strength in a refractory high-entropy alloy. *Adv. Mater.* **32**, 2004029 (2020).
5. Pan, Q. et al. Gradient cell-structured high-entropy alloy with exceptional strength and ductility. *Science* **374**, 984–989 (2021).
6. Yang, T. et al. Multicomponent intermetallic nanoparticles and super mechanical behaviors of complex alloys. *Science* **362**, 933–937 (2018).
7. Fu, Z. et al. A high-entropy alloy with hierarchical nanoprecipitates and ultrahigh strength. *Sci. Adv.* **4**, eaat8712 (2018).
8. Feng, R. et al. Superior high-temperature strength in a supersaturated refractory high-entropy alloy. *Adv. Mater.* **33**, 2102401 (2021).
9. Qiu, X.-W., Zhang, Y.-P., He, L. & Liu, C.-G. Microstructure and corrosion resistance of AlCrFeCuCo high entropy alloy. *J. Alloy. Compd.* **549**, 195–199 (2013).
10. Qiu, Y., Thomas, S., Gibson, M. A., Fraser, H. L. & Birbilis, N. Corrosion of high entropy alloys. *NPJ Mater. Degrad.* **1**, 1–18 (2017).
11. Shi, Y., Yang, B. & Liaw, P. K. Corrosion-resistant high-entropy alloys: a review. *Metals* **7**, 43 (2017).
12. Chou, Y. L., Yeh, J. W. & Shih, H. C. The effect of molybdenum on the corrosion behaviour of the high-entropy alloys Co<sub>1</sub>. 5CrFeNi<sub>1</sub>. 5Ti<sub>0</sub>. 5Mox in aqueous environments. *Corros. Sci.* **52**, 2571–2581 (2010).
13. Mao, X. et al. Microstructure and corrosion performance of Al<sub>0</sub>. 5FeCoNiCrMn coating prepared by plasma spraying mechanically alloyed powders. *Surf. Eng.* **38**, 383–392 (2022).
14. Wang, C. et al. Chloride corrosion property evolution of CrMnFeCoNi high-entropy alloy coating. *Mater. Sci. Technol.* **38**, 1418–1424 (2022).
15. Chen, P. et al. Fatigue behavior of high-entropy alloys: a review. *Sci. China Technol. Sci.* **61**, 168–178 (2018).
16. Hemphill, M. A. et al. Fatigue behavior of Al<sub>0</sub>. 5CoCrCuFeNi high entropy alloys. *Acta Mater.* **60**, 5723–5734 (2012).
17. Li, W., Chen, S. & Liaw, P. K. Discovery and design of fatigue-resistant high-entropy alloys. *Scr. Mater.* **187**, 68–75 (2020).
18. Li, P., Wang, A. & Liu, C. T. A ductile high entropy alloy with attractive magnetic properties. *J. Alloy. Compd.* **694**, 55–60 (2017).
19. Juan, C.-C. et al. Solution strengthening of ductile refractory HfMoxNbTaTiZr high-entropy alloys. *Mater. Lett.* **175**, 284–287 (2016).
20. Sheikh, S. et al. Alloy design for intrinsically ductile refractory high-entropy alloys. *J. Appl. Phys.* **120**, 164902 (2016).
21. Chen, R. et al. Composition design of high entropy alloys using the valence electron concentration to balance strength and ductility. *Acta Mater.* **144**, 129–137 (2018).
22. Wei, S. et al. Natural-mixing guided design of refractory high-entropy alloys with as-cast tensile ductility. *Nat. Mater.* **19**, 1175–1181 (2020).
23. Laurent-Brocq, M. et al. Insights into the phase diagram of the CrMnFeCoNi high entropy alloy. *Acta Mater.* **88**, 355–365 (2015).
24. Otto, F., Yang, Y., Bei, H. & George, E. P. Relative effects of enthalpy and entropy on the phase stability of equiatomic high-entropy alloys. *Acta Mater.* **61**, 2628–2638 (2013).
25. Wang, F. et al. Multiplicity of dislocation pathways in a refractory multiprincipal element alloy. *Science* **370**, 95–101 (2020).
26. George, E. P., Raabe, D. & Ritchie, R. O. High-entropy alloys. *Nat. Rev. Mater.* **4**, 515–534 (2019).
27. Miracle, D. B. & Senkov, O. N. A critical review of high entropy alloys and related concepts. *Acta Mater.* **122**, 448–511 (2017).
28. Maresca, F. & Curtin, W. A. Mechanistic origin of high strength in refractory BCC high entropy alloys up to 1900K. *Acta Mater.* **182**, 235–249 (2020).
29. Yeh, J.-W. et al. Formation of simple crystal structures in Cu-Co-Ni-Cr-Al-Fe-Ti-V alloys with multiprincipal metallic elements. *Met. Mater. Trans. A* **35**, 2533–2536 (2004).
30. Jien-Wei, Y. E. H. Recent progress in high entropy alloys. *Ann. Chim. Sci. Mat.* **31**, 633–648 (2006).
31. Rao, S. I., Woodward, C., Akdim, B., Senkov, O. N. & Miracle, D. Theory of solid solution strengthening of BCC chemically complex alloys. *Acta Mater.* **209**, 116758 (2021).
32. Wen, C. et al. Modeling solid solution strengthening in high entropy alloys using machine learning. *Acta Mater.* **212**, 116917 (2021).
33. Okamoto, N. L., Yuge, K., Tanaka, K., Inui, H. & George, E. P. Atomic displacement in the CrMnFeCoNi high-entropy alloy—a scaling factor to predict solid solution strengthening. *AIP Adv.* **6**, 125008 (2016).
34. Zhao, Y. Y. & Nieh, T. G. Correlation between lattice distortion and friction stress in Ni-based equiatomic alloys. *Intermetallics* **86**, 45–50 (2017).
35. Zhao, Y. Y., Lei, Z. F., Lu, Z. P., Huang, J. C. & Nieh, T. G. A simplified model connecting lattice distortion with friction stress of Nb-based equiatomic high-entropy alloys. *Mater. Res. Lett.* **7**, 340–346 (2019).
36. Chen, H. et al. Contribution of lattice distortion to solid solution strengthening in a series of refractory high entropy alloys. *Met. Mater. Trans. A* **49**, 772–781 (2018). *Phys Metall Mater.*
37. Tong, Y. et al. Severe local lattice distortion in Zr-and/or Hf-containing refractory multi-principal element alloys. *Acta Mater.* **183**, 172–181 (2020).
38. Lee, C. et al. Lattice distortion in a strong and ductile refractory high-entropy alloy. *Acta Mater.* **160**, 158–172 (2018).
39. Li, J. et al. Heterogeneous lattice strain strengthening in severely distorted crystalline solids. *Proc. Natl Acad. Sci.* **119**, e2200607119 (2022).
40. Samolyuk, G. D., Osetsky, Y. N., Stocks, G. M. & Morris, J. R. Role of static displacements in stabilizing body centered cubic high entropy alloys. *Phys. Rev. Lett.* **126**, 025501 (2021).
41. Ikeda, Y., Gubaev, K., Neugebauer, J., Grabowski, B. & Körmann, F. Chemically induced local lattice distortions versus structural phase transformations in compositionally complex alloys. *NPJ Comput. Mater.* **7**, 1–8 (2021).
42. Tsai, K. Y., Tsai, M. H. & Yeh, J. W. Sluggish diffusion in co–cr–fe–mn–ni high-entropy alloys. *Acta Mater.* **61**, 4887–4897 (2013).
43. Wang, R. et al. Effect of lattice distortion on the diffusion behavior of high-entropy alloys. *J. Alloy. Compd.* **825**, 154099 (2020).
44. Mu, S., Wimmer, S., Mankovsky, S., Ebert, H. & Stocks, G. M. Influence of local lattice distortions on electrical transport of refractory high entropy alloys. *Scr. Mater.* **170**, 189–194 (2019).
45. Zhang, Y. et al. Influence of chemical disorder on energy dissipation and defect evolution in concentrated solid solution alloys. *Nat. Commun.* **6**, 1–9 (2015).
46. Zou, Y., Maiti, S., Steurer, W. & Spolenak, R. Size-dependent plasticity in an Nb<sub>25</sub>Mo<sub>25</sub>Ta<sub>25</sub>W<sub>25</sub> refractory high-entropy alloy. *Acta Mater.* **65**, 85–97 (2014).
47. Zunger, A., Wei, S. H., Ferreira, L. G. & Bernard, J. E. Special quasirandom structures. *Phys. Rev. Lett.* **65**, 353 (1990).
48. Song, H. et al. Local lattice distortion in high-entropy alloys. *Phys. Rev. Mater.* **1**, 023404 (2017).
49. Oh, H. S. et al. Lattice distortions in the FeCoNiCrMn high entropy alloy studied by theory and experiment. *Entropy* **18**, 321 (2016).
50. Ishibashi, S., Ikeda, Y., Körmann, F., Grabowski, B. & Neugebauer, J. Correlation analysis of strongly fluctuating atomic volumes, charges, and stresses in body-centered cubic refractory high-entropy alloys. *Phys. Rev. Mater.* **4**, 023608 (2020).
51. Feng, B. & Widom, M. Elastic stability and lattice distortion of refractory high entropy alloys. *Mater. Chem. Phys.* **210**, 309–314 (2018).
52. Guo, W. et al. Local atomic structure of a high-entropy alloy: an X-ray and neutron scattering study. *Met. Mater. Trans. A Phys. Met. Mater. Sci.* **44**, 1994–1997 (2013).
53. Ye, Y. F. et al. Atomic-scale distorted lattice in chemically disordered equimolar complex alloys. *Acta Mater.* **150**, 182–194 (2018).
54. Meng, F. et al. Charge transfer effect on local lattice distortion in a HfNbTiZr high entropy alloy. *Scr. Mater.* **203**, 114104 (2021).
55. Noehring, W. G. & Curtin, W. A. Correlation of microdistortions with misfit volumes in high entropy alloys. *Scr. Mater.* **168**, 119–123 (2019).
56. Hu, Y.-J., Sundar, A., Ogata, S. & Qi, L. Screening of generalized stacking fault energies, surface energies and intrinsic ductile potency of refractory multi-component alloys. *Acta Mater.* **210**, 116800 (2021).
57. Dunitz, J. D., Schomaker, V. & Trueblood, K. N. Interpretation of atomic displacement parameters from diffraction studies of crystals. *J. Phys. Chem.* **92**, 856–867 (1988).
58. Zhang, F. et al. Chemical complexity induced local structural distortion in NiCoFeMnCr high-entropy alloy. *Mater. Res. Lett.* **6**, 450–455 (2018).
59. Ong, S. P. et al. Python Materials Genomics (pymatgen): a robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).
60. Liu, J., Zhou, C. & Wang, H. Thermodynamic re-optimizations of the Ru–X (X = Al, Hf, Mo, Ti) binary systems. *Calphad* **72**, 102239 (2021).
61. Hu, Y.-J. et al. Local electronic descriptors for solute-defect interactions in bcc refractory metals. *Nat. Commun.* **10**, 1–11 (2019).
62. Hartigan, J. A. & Hartigan, P. M. The dip test of unimodality. *Ann. Stat.* **13**, 70–84 (1985).
63. Li, Z. & Raabe, D. Strong and ductile non-equiatom high-entropy alloys: design, processing, microstructure, and mechanical properties. *JOM* **69**, 2099–2106 (2017).
64. Ye, Y. X. et al. Evaluating elastic properties of a body-centered cubic NbHfZrTi high-entropy alloy—A direct comparison between experiments and ab initio calculations. *Intermetallics* **109**, 167–173 (2019).
65. Borg, C. K. H. et al. Expanded dataset of mechanical properties and observed phases of multi-principal element alloys. *Sci. Data* **7**, 1–6 (2020).
66. Feuerbacher, M., Heidelmann, M. & Thomas, C. Plastic deformation properties of Zr–Nb–Ti–Ta–Hf high-entropy alloys. *Philos. Mag.* **95**, 1221–1232 (2015).
67. Huang, H. et al. Phase-transformation ductilization of brittle high-entropy alloys via metastability engineering. *Adv. Mater.* **29**, 1701678 (2017).
68. Hu, Y.-M. et al. Microstructure and mechanical properties of NbZrTi and NbHfZrTi alloys. *Rare Met.* **38**, 840–847 (2019).
69. Qiao, D.-X. et al. A novel series of refractory high-entropy alloys Ti<sub>2</sub>ZrHf<sub>0</sub>. 5VNbx with high specific yield strength and good ductility. *Acta Met. Sin.-ENGL* **32**, 925–931 (2019).
70. Maiti, S. & Steurer, W. Structural-disorder and its effect on mechanical properties in single-phase TaNbHfZr high-entropy alloy. *Acta Mater.* **106**, 87–97 (2016).



71. Coury, F. G., Kaufman, M. & Clarke, A. J. Solid-solution strengthening in refractory high entropy alloys. *Acta Mater.* **175**, 66–81 (2019).
72. Senkov, O. N., Gorse, S. & Miracle, D. B. High temperature strength of refractory complex concentrated alloys. *Acta Mater.* **175**, 394–405 (2019).
73. Fazakas, E. et al. Experimental and theoretical study of Ti<sub>20</sub>Zr<sub>20</sub>Hf<sub>20</sub>Nb<sub>20</sub>X<sub>20</sub> (X = V or Cr) refractory high-entropy alloys. *Int. J. Refract. Hard Met.* **47**, 131–138 (2014).
74. Senkov, O. N., Wilks, G. B., Scott, J. M. & Miracle, D. B. Mechanical properties of Nb sub (25 Mo) sub (2) 5 Ta sub (25 W) sub (2) 5 and V sub (20 Nb) sub (2) 0 Mo sub (20 Ta) sub (2) 0 W sub (20 refractory high entropy alloys). *Intermetallics* **19**, 698–706 (2011).
75. Juan, C.-C. et al. Enhanced mechanical properties of HfMoTaTiZr and HfMoNb-TaTiZr refractory high-entropy alloys. *Intermetallics* **62**, 76–83 (2015).
76. Guo, N. N. et al. Microstructure and mechanical properties of in-situ MC-carbide particulates-reinforced refractory high-entropy Mo<sub>0.5</sub>NbHf<sub>0.5</sub>ZrTi matrix alloy composite. *Intermetallics* **69**, 74–77 (2016).
77. Han, Z. D. et al. Effect of Ti additions on mechanical properties of NbMoTaW and VNbMoTaW refractory high entropy alloys. *Intermetallics* **84**, 153–157 (2017).
78. Lin, C.-M., Juan, C.-C., Chang, C.-H., Tsai, C.-W. & Yeh, J.-W. Effect of Al addition on mechanical properties and microstructure of refractory AlxHfNbTaTiZr alloys. *J. Alloy. Compd.* **624**, 100–107 (2015).
79. Liu, Y. et al. Microstructure and mechanical properties of refractory HfMo<sub>0.5</sub>NbTiV<sub>0.5</sub>Six high-entropy composites. *J. Alloy. Compd.* **694**, 869–876 (2017).
80. Yao, H. W. et al. Mechanical properties of refractory high-entropy alloys: experiments and modeling. *J. Alloy. Compd.* **696**, 1139–1150 (2017).
81. Wei, Q. et al. Microstructure evolution, mechanical properties and strengthening mechanism of refractory high-entropy alloy matrix composites with addition of TaC. *J. Alloy. Compd.* **777**, 1168–1175 (2019).
82. Dirras, G. et al. Mechanical behavior and microstructure of Ti<sub>20</sub>Hf<sub>20</sub>Zr<sub>20</sub>Ta<sub>20</sub>Nb<sub>20</sub> high-entropy alloy loaded under quasi-static and dynamic compression conditions. *Mater. Charact.* **111**, 106–113 (2016).
83. Guo, N. N. et al. Microstructure and mechanical properties of refractory MoNbHfZrTi high-entropy alloy. *Mater. Des.* **81**, 87–94 (2015).
84. Wu, Y. D. et al. Phase composition and solid solution strengthening effect in TiZrNbMoV high-entropy alloys. *Mater. Des.* **83**, 651–660 (2015).
85. Wu, Y. D. et al. A refractory Hf<sub>25</sub>Nb<sub>25</sub>Ti<sub>25</sub>Zr<sub>25</sub> high-entropy alloy with excellent structural stability and tensile properties. *Mater. Lett.* **130**, 277–280 (2014).
86. Couzinié, J. P. et al. On the room temperature deformation mechanisms of a TiZrHfNbTa refractory high-entropy alloy. *Mater. Sci. Eng.: A* **645**, 255–263 (2015).
87. Dirras, G. et al. Elastic and plastic properties of as-cast equimolar TiHfZrTaNb high-entropy alloy. *Mater. Sci. Eng. A* **654**, 30–38 (2016).
88. Yao, H. W. et al. NbTaV-(Ti, W) refractory high-entropy alloys: experiments and modeling. *Mater. Sci. Eng. A* **674**, 203–211 (2016).
89. Han, Z. D. et al. Microstructures and mechanical properties of Ti<sub>x</sub>NbMoTaW refractory high-entropy alloys. *Mater. Sci. Eng. A* **712**, 380–385 (2018).
90. Chen, S. Y., Yang, X., Dahmen, K. A., Liaw, P. K. & Zhang, Y. Microstructures and crackling noise of AlxNbTiMoV high entropy alloys. *Entropy* **16**, 870–884 (2014).
91. Yao, H. et al. MoNbTaV medium-entropy alloy. *Entropy* **18**, 189 (2016).
92. Tseng, K.-K. et al. Effects of Mo, Nb, Ta, Ti, and Zr on mechanical properties of equiatomic Hf-Mo-Nb-Ta-Ti-Zr alloys. *Entropy* **21**, 15 (2018).
93. Zyka, J. et al. Microstructure and room temperature mechanical properties of different 3 and 4 element medium entropy alloys from HfNbTaTiZr system. *Entropy* **21**, 114 (2019).
94. Qiao, D. X., Jiang, H., Chang, X. X., Lu, Y. P. & Li, T. J. Microstructure and mechanical properties of VTaTiMoAl<sub>x</sub> refractory high entropy alloys. *Mater. Sci. Forum* **898**, 638–642 (2017).
95. Guo, N. N. et al. Effect of composing element on microstructure and mechanical properties in Mo-Nb-Hf-Zr-Ti multi-principle component alloys. *Intermetallics* **69**, 13–20 (2016).
96. Varvenne, C., Leyson, G. P. M., Ghazisaeidi, M. & Curtin, W. A. Solute strengthening in random alloys. *Acta Mater.* **124**, 660–683 (2017).
97. Vegard, L. Die konstitution der mischkristalle und die raumfüllung der atome. *Z. für. Phys.* **5**, 17–26 (1921).
98. Noble, B., Harris, S. J. & Dinsdale, K. The elastic modulus of aluminium-lithium alloys. *J. Mater. Sci.* **17**, 461–468 (1982).
99. Andersson, J.-O., Helander, T., Höglund, L., Shi, P. & Sundman, B. Thermo-Calc & DICTRA, computational tools for materials science. *Calphad* **26**, 273–312 (2002).
100. Owen, L. R. & Jones, N. G. Lattice distortions in high-entropy alloys. *J. Mater. Res.* **33**, 2954–2969 (2018).
101. Welch, B. L. The generalization of ‘STUDENT’S’ problem when several different population variances are involved. *Biometrika* **34**, 28–35 (1947).
102. Sutton, A. P., Finnis, M. W., Pettifor, D. G. & Ohta, Y. The tight-binding bond model. *J. Phys. C: Solid State Phys.* **21**, 35 (1988).
103. Tandoc, C. & Hu, Y.-J. A bond counting model Using B2 intermetallic calculations for efficient estimation of BCC MPEA lattice parameter with superior accuracy. *arXiv preprint at* <https://doi.org/10.48550/arXiv.2203.07137> (2022).
104. Rice, J. R. Dislocation nucleation from a crack tip: an analysis based on the Peierls concept. *J. Mech. Phys. Solids* **40**, 239–271 (1992).
105. Yang, X., Zhang, Y. & Liaw, B. Microstructure and compression properties of high-entropy alloys NbTiVTaAl x. *Procedia Eng.* **36**, 292–298 (2012).
106. Huang, H. et al. On cooling rates dependence of microstructure and mechanical properties of refractory high-entropy alloys HfTaTiZr and HfNbTiZr. *Scr. Mater.* **211**, 114506 (2022).
107. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953 (1994).
108. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).
109. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169 (1996).
110. Körmann, F., Ikeda, Y., Grabowski, B. & Sluiter, M. H. Phonon broadening in high entropy alloys. *NPJ Comput. Mater.* **3**, 1–9 (2017).
111. Clouet, E., Nastar, M. & Sigli, C. Nucleation of Al 3 Zr and Al 3 Sc in aluminum alloys: From kinetic Monte Carlo simulations to classical theory. *Phys. Rev. B* **69**, 064109 (2004).
112. Shu, S., Wells, P. B., Almirall, N., Odette, G. R. & Morgan, D. D. Thermodynamics and kinetics of core-shell versus appendage co-precipitation morphologies: an example in the Fe-Cu-Mn-Ni-Si system. *Acta Mater.* **157**, 298–306 (2018).

## ACKNOWLEDGEMENTS

C.T. and Y.J.H. thank the financial support from the startup fund from Drexel University. L.Q. acknowledge the financial support from the National Science Foundation (NSF) Award DMR-1847837. P.K.L. very much appreciates the supports from (1) the National Science Foundation (DMR-1611180, 1809640, and 2226508) and (2) the US Army Research Office (W911NF-13-1-0438 and W911NF-19-2-0049). The computational support from the Drexel’s University Research Computing Facility is greatly acknowledged. A part of the calculations was also carried out using the Extreme Science and Engineering Discovery Environment (XSEDE) Stampede2 at the TACC through allocation TG-DMR190035, and allocation MAT220033 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation Grants #2138259, #2138286, #2138307, #2137603, and #2138296.

## AUTHOR CONTRIBUTIONS

Y.J.H. conceptualized and supervised the research project. C.T. performed the first-principles calculations, development of descriptors and statistical learning models, and data mining of alloy properties. C.T., Y.J.H., L.Q., and P.K.L. prepared the manuscript. All authors discussed the results and contributed to the manuscript.

## 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/s41524-023-00993-x>.

**Correspondence** and requests for materials should be addressed to Yong-Jie Hu.

**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) 2023