Modeling environment-dependent atomic-level properties in complex-concentrated alloys

Mackinzie S. Farnell
School of Materials Science & Engineering,
University of California Berkeley, Berkeley, California, USA

Zachary D. McClure, Shivam Tripathi, and Alejandro Strachan
School of Materials Engineering and Birck Nanotechnology Center,
Purdue University, West Lafayette, Indiana 47907, USA

(Dated: January 7, 2022)

Abstract

Complex-concentrated-alloys (CCAs) are of interest for a range of applications due to a host of desirable properties, including, high-temperature strength and tolerance to radiation damage. Their multi-principal component nature results in a vast number of possible atomic environments with the associated variability in chemistry and structure. This atomic-level variability is central to the unique properties of these alloys but makes their modeling challenging. We combine atomistic simulations using many body potentials with machine learning to develop predictive models of various atomic properties of CrFeCoNiCu-based CCAs: relaxed vacancy formation energy, atomic-level cohesive energy, pressure, and volume. A fingerprint of the local atomic environments is obtained combining invariants associated with the local atomic geometry and periodic-table information of the atoms involved. Importantly, all descriptors are based on the unrelaxed atomic structure, thus, they are computationally inexpensive to compute. This enables the incorporation of these models into macroscopic simulations. The models show good accuracy and we explore their ability to extrapolate to compositions and elements not used during training.

I. INTRODUCTION

Complex concentrated alloys (CCAs) and multiple principal component alloys are crystalline materials consisting of four or more elements combined in similar fractions. They have attracted significant attention since their introduction in 2004 [1, 2] due to a range of desirable properties including high strength, even at high temperatures, thermal stability, and resistance to fatigue[3]. In addition, the vast space of potential alloy compositions makes them tailorable to specific applications [2–4]. The inherent variability in the local atomic configurations is the driving factor behind many of their unique properties, but also poses significant challenges to modeling and experimental characterization [5]. For example, the distribution of vacancy formation energies determines vacancy concentration which, in turn, dominates creep. Another key example is single crystal strength, which is dominated by local changes in the core energy along dislocation lines [6–8]. The local atomic environments govern the energy landscape under which dislocations move and their variability hinders their mobility, resulting in strengthening. For other examples of the relationship between local variability and properties see Refs. [9–12]. These local properties can be assessed computationally via intensive atomistic simulations, but given the enormous number of local atomic configurations individual atoms can encounter in CCAs, computationally efficient models for local properties are highly desirable. For example, Chen et al. studied vacancy formation energy (VFE) in CrFeCoNi alloys using density functional theory (DFT) on special quasirandom structures (SQS) [13]. The authors found a wide distribution in VFE ranging from 1.5 to 2 eV with averages between 1.58 and 1.89 eV depending on the element. This work explored 24 of the most likely configurations given a 20 atom cell, a small subset of all the possibilities. For example, the number of local first nearest neighboring configurations in a five-element alloy is 5^Z, Z the coordination number, divided by the multiplicity due to symmetry operations; clearly brute force *ab initio* calculations and even lower-fidelity interatomic force field calculations are out of the question.

Efforts to efficiently explore and characterize this enormous space have turned to machine learning methods for phase prediction, material screening, and through that best practices have begun to emerge [14–16]. The foundations for these screening processes built on early work for formation energy determination using cluster expansion (CE) methods [17]. Extensions of this model beyond binary components have shown great success in ternary semicon-

ductors for predicting possible phase formations and separation [18], and multi-component CCA [19]. However, the method relies on unpacking 1^{st}, 2^{nd}, and higher order pairwise interactions in a symmetric, unrelaxed system. For systems that have been relaxed, and symmetry disrupted, the CE models begin to break down [20]. To overcome this limitation, rather than describing atomic interactions through the CE formalism, Shapeev used tensor descriptions to represent the energetics of multicomponent systems and showed better convergence rate with respect to training set size than CE for total energies [21, 22]. Each of these respective methods consider pair-wise interactions within a system, and sum their total contributions to determine total system energy. However, many of these methods focus on the macroscale properties and not on the local variability. To inform single crystal strength models, approximations to the local stresses have been developed from atomic radii and elastic constants [6, 8]. These model are easy to evaluate but involve several approximations and the associated uncertainties have not been quantified. In this paper we develop predictive models for various atomic-level properties of CCAs from molecular mechanics simulation data using invariant descriptors of local atomic environments and chemistry and neural networks. Recent work on high entropy diborides used atomistic simulations to develop models for VFE depending on the local environment. The authors showed the ability of pair approximation models with linear models and local structure up several neighboring shells to provide accurate descriptions [23].

In summary, the development of validated and computationally expedient models capable of predicting a variety of atomic-level properties of CCAs remains an active area of research and we are unaware of models capable of predicting a range of atomic-level properties needed to inform constitutive laws required for macroscopic predictions. To address this gap, we combine molecular static calculations using a many-body interatomic potential with machine learning to create predictive models for local atomic properties of face centered cubic CCAs containing Co, Cr, Fe, and Ni. We model several properties (relaxed vacancy formation energies, atomic pressures and volumes, and cohesive energies) and assess the ability of the models to generalize and predict properties for new compositions and new chemistries. Importantly, the descriptors of local chemistry and geometry used as inputs to the models are generated from unrelaxed atomic configurations; thus, evaluating the models does not require computationally intensive structural relaxations.

Our work builds on the significant recent progress in the use of machine learning for

atomistic simulations and a long history of modeling multicomponent systems [17]. Neural networks[24], gaussian processes [25], and even linear regression [26] have been shown to be powerful models to relate local atomic environment and atomic energies, resulting in a new class of interatomic potentials. In these models, local atomic structures are described with descriptors that capture the symmetries of the underlying physics (e.g. translational and rotational invariance). Moment tensor potentials have also shown great promise to describe multicomponent systems [21, 27] Approaches to descrisbe local atomic environments include smooth-atomic-overlaps (SOAP) [28], two- and three-body symmetry functions [24], tensor formalisms [21], and bispectrum coefficients [26]. In this paper, we use bispectrum coefficients to relate the local, first nearest neighbor, environment of the *unrelaxed* structure to various *relaxed* local properties. Thus, our models need to learn not just the mapping between structure and property but also the relaxation of the local structure. In addition to the geometry, we use standard description of chemical properties of each environment. We explore the ability of the models to predict environments not seen during training including those originating from unseen compositions as well as the inclusion of new elements.

The remainder of the paper is organized as follows. Section II describes the data, descriptors, and models used. Section III focuses on results of training the models and using them to predict properties for new compositions and chemistries and Section V provides access to the code used to produce these results. Finally, conclusions are drawn in Section IV.

II. METHODS

A. LAMMPS Simulations

The atomic properties of interest (relaxed vacancy formation energy, cohesive energy, stress, and volume) were obtained using the LAMMPS simulation package [29] with an embedded atom model interatomic potential developed by Farkas et al. [30]. Initial structures of the CCA alloys of interest, equiatomic Cr, Fe, Co, Ni, Cu, were obtained using an FCC lattice with lattice parameter a_0=3.56 Å with atoms assigned following the SQS method. [31] All descriptors used as inputs for the neural network models are calculated from these initial structures, as described in sub-section II B.

After the descriptors are extracted, we relax the structure using molecular statics. We

minimize the total energy with respect to both lattice parameters and atomic coordinates under ambient pressure with thresholds of 10^{-12} and 10^{-12} eV/Å for scaled energy and force, respectively.

After relaxation, we compute the atomic energy (defined as the potential energy contribution of each atom), local atomic stress from the virial theorem [32], and local volume from a Voronoi tessellation [33]. Finally, the vacancy formation energy of each atomic site is computed by sequentially removing each atom and re-relaxing the structure (maintaining the simulation cell parameters constant). We define the relaxed vacancy formation energy (E_v^i) for site i from the energy difference between the perfect crystal E_0 and the system after the removal of corresponding atom E_i.

$$E_v^i = (E_i + \mu_i) - E_0, \tag{1}$$

where μ_i is the chemical potential of atoms of element corresponding to atom i. This chemical potential is obtained as the cohesive energy of a pure element system.

The distributions of the resulting properties for each atom type obtained from a 5,000-atom SQS structure are shown in Fig. 1. These distributions compare well with prior ab initio calculations [13]. Our average relaxed vacancy formation energies for Cr, Fe, Co, and Ni are 1.52, 1.58, 1.44 and 1.63 eV, respectively. These points compare well with *ab initio* results reporting average values of 1.61, 1.58, 1.70, and 1.89 eV for Cr, Fe, Co, and Ni obtained in 4-element CCAs.

We note that we use an interatomic potential since our goal is to establish the validity and accuracy of our proposed model of relaxed atomic-level properties. For more accurate models the interatomic potential would be replaced by DFT calculations that provide a good balance between accuracy and computational cost and can capture properties associated with the electronic structure of the systems, such as magnetism.

B. Model Features

We use a combination of chemical and geometrical descriptors to describe individual atoms. As described above, all descriptors are obtained from the initial, unrelaxed, structures. To describe the local geometrical environment we use bispectrum coefficients [26] that start from the local atomic density around an atom and create a list of translationally

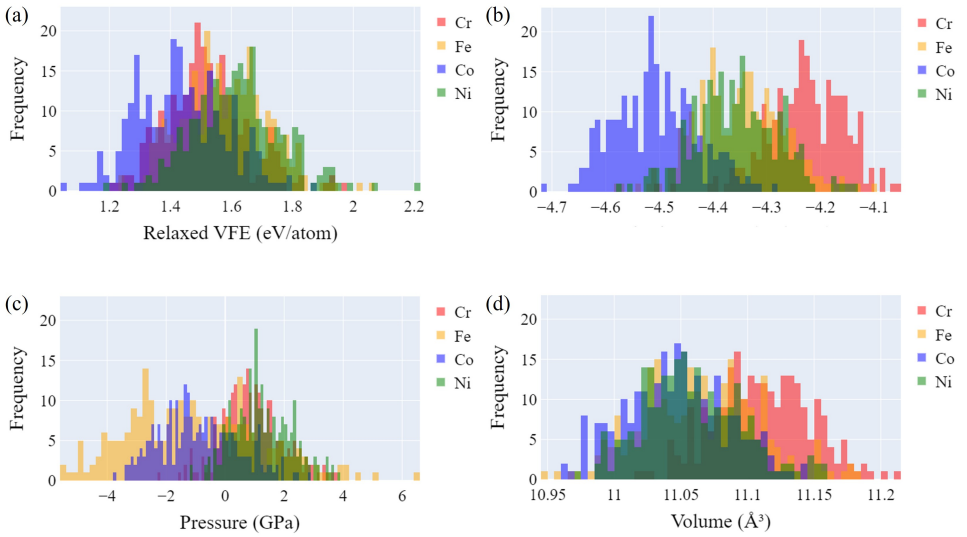

FIG. 1: Distribution of values for Relaxed VFE (a), Cohesive Energy (b), Pressure (c), and Volume (d).

and rotationally invariant descriptors. To distinguish between atom types in the bispectrum calculation, we use atomic numbers as prefactors in the local density during the coefficient calculations. Bispectrum coefficients are obtained using a radial cutoff 10% beyond the nearest neighbor distance (1.1 a_0 $\sqrt{2}/2$) and a band limit of eight for the resulting in a total of 55 coefficients. We note that the bispectrum coefficients capture up to four-body correlations and do not provide a complete description of atomic environments [34] and multiple local environments can lead to identical coefficients. This issue is less of a concern for multi-component systems and, from a practical point of view, near-DFT accuracy has been obtained for simple metals [35]. Thus, we believe the bispectrum coefficients provide an appropriate description for the problem at hand. In addition to the geometric descriptors, we use the atomic number of the central atom and the following chemical descriptors for the central atom queried from Pymatgen:[36] atomic radius, atomic mass, Poisson's ratio, electrical resistivity, thermal conductivity, and Brinell hardness. These properties were chosen to describe the size, bonding, and electronic structure of the central atom. We also studied the effects of using descriptors capturing the central atom and the 12 nearest neighboring atoms using a rule of mixtures, but found that these did not improve model performance; these results are discussed in the Supplemental Information in the section "Train Neural

Network on Equiatomic CrFeCoNi". These descriptors were added as additional physics informed descriptors, and have good overlap with previously investigated descriptors used in material classifications[37].

C. Neural network architecture

Machine learning models were implemented in the Jupyter notebook environment [38] on nanoHUB [39] using Tensorflow [40] and Keras [41] libraries. The models use shallow neural networks with a first hidden layer containing 512 neurons connected to the 63 input features. This hidden layer used exponential linear unit (elu) activation functions and was followed by a dropout layer with dropout ratio of 0.2. During training, the loss function was mean squared error and the *Adagrad* optimizer was used[42]. Also, the learning rate was 0.002 and the models were trained for 5000 epochs. This model architecture and hyperparameters were chosen after testing several models, as detailed in the supplemental material.

To train the model, the data was split into testing and training sets, with 80% of data used for training and 20% used for testing. The inputs and outputs were normalized using the standard approach of subtracting the mean and dividing by the standard deviation of the training data. During training, 10% of the training data was used for validation. The validation data differs from the testing data in that it is used during the training of the model to assess convergence, while the testing data is hidden during training and only used after training to evaluate the model. Initially, an early stopping criterion based on validation data was used to determine number of epochs for training. However, models had similar errors when trained with early stopping and with 5000 epochs, so 5000 epochs were used to train all models. Independent models were developed for each property of interest to describe all elements in the system. The initial model architecture was developed using equiatomic CrFeCoNi structures with a data set containing 5000 atoms. However, we found that training with 2000 atoms was sufficient. Thus, models were then trained and tested on equiatomic four-element alloys CrFeNiCu, FeCoNiCu, CrCoNiCu, and CrFeCoCu with data sets containing 2000 entries (atoms) each. The predictive ability of these models was tested on the five element alloy CrFeCoNiCu and on non-equiatomic alloys.

III. MODELS FOR ATOMISTIC PROPERTIES OF CCAS

As described above, we trained neural network models to predict relaxed vacancy formation energy, atomic cohesive energy, atomic pressure, and local Voronoi volume. Figure 2 shows parity plots of the four properties for CrFeCoNi alloy. Only testing data points are shown, these have not been used in training. The results highlight the large atomic variability of all the properties studied, the range for each element is larger than the difference in mean values between elements. The dash lines bound errors corresponding to 10% of the range of each property. In absolute terms, the the mean absolute errors are 0.042 eV for cohesive energy, 0.059 eV for VFE, 0.809 GPa for pressure, and 0.020 $Å^3$ for atomic volume. Figure 3 compares the accuracy of the models for the five four-element alloys used for training. We show the mean absolute error of all predictions normalized by the range over the testing data points. Our models have comparable performance across the different chemistries. Importantly, models can predict properties with an accuracy of approximately 10% of the range for each of the properties studied. This level of accuracy is comparable to that achieved in high-entropy borides using first nearest descriptors [23].

A. Predicting properties for new compositions

The model trained on equiatomic CrFeCoNi was used to predict properties of alloys with different compositions with the same four elements. Neural network predictions are compared to molecular statics predictions in Figures 4 and 5. Figure 4 assesses the model accuracy for $Cr_{20}Fe_{40}Co_{20}Ni_{20}$. We find the model to be able to make accurate predictions across all properties. The normalized MAE values are slightly larger than those for the composition used for training, with models predicting with an accuracy of roughly 20% of the range of each property. The slight underestimation of the Voronoi volumes is due to the larger overall volume of this Fe-rich alloy. Figure 5 assesses the ability of the model trained on equiatomic CoCrFeNi to predict on $Cr_{15}Fe_{55}Co_{15}Ni_{15}$. For this composition, with environments more rich in Fe that deviate further from the training data, the model accuracy degrades further. The model is still able to capture overall trends in properties but the trend observed above of underestimating atomic volumes accentuates with increasing Fe. Going from the equiatomic systems to the $Cr_{15}Fe_{55}Co_{15}Ni_{15}$, the average volume computed

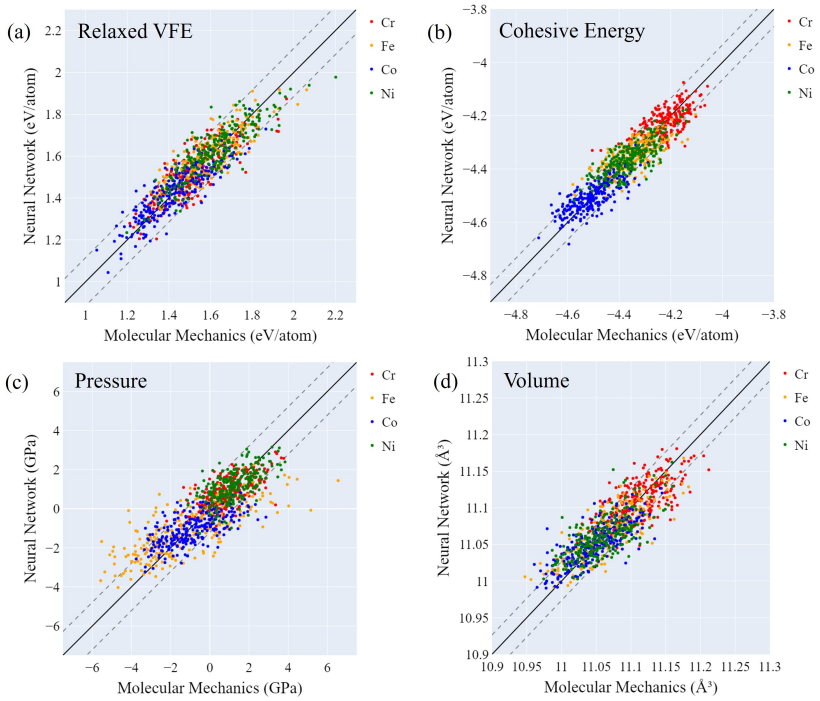

FIG. 2: Machine learning model predictions compared to molecular statics results for relaxed VFE (a), cohesive energy (b), atomic pressure (c), and atomic volume (d) for equiatomic CoCrFeNi configurations belonging to the testing set. The grey, dashed lines indicate errors of

$\pm$

10% of the range for each property, in absolute terms these represent

$\pm$

0.115 eV for relaxed VFE,

$\pm$

0.065 eV for cohesive energy,

$\pm$

1.213 GPa for atomic pressure, and

$\pm$

0.026 Å^3 for atomic volume.

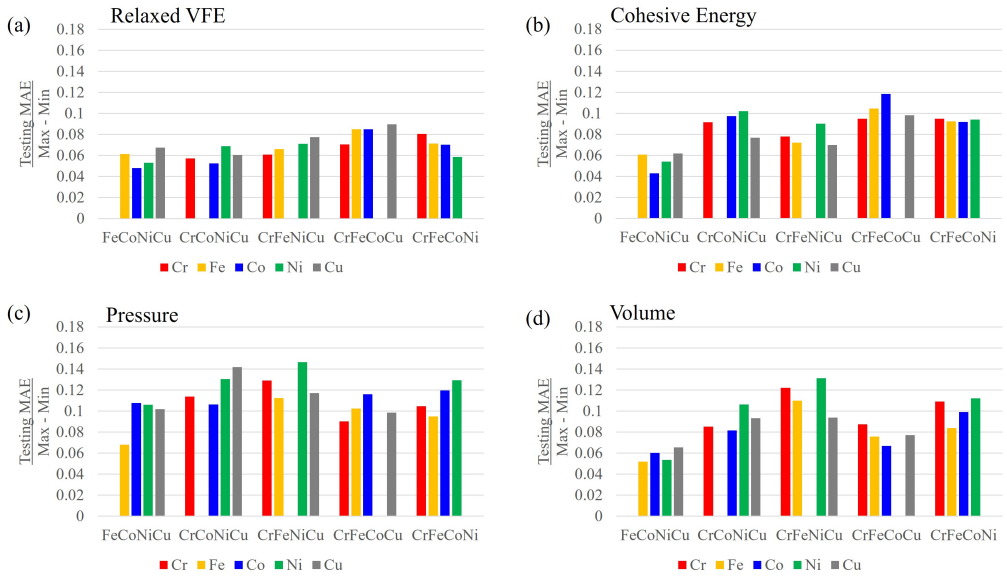

FIG. 3: MAE normalized by range for the testing data for each of the four-atom systems for Relaxed VFE (a), Cohesive Energy (b), Pressure (c), and Volume (d).

using molecular mechanics increases from 11.070 Å^3 to 11.146 Å^3. In contrast, the model average volume predictions are essentially unchanged. This indicates that the model cannot capture the overall expansion observed with increasing Fe content, this is not surprising as this information was not provided to the model during training.

The model trained on equiatomic CrFeCoNi was also used to make predictions on several other alloys with different compositions. The error in these predictions, for the four properties of interest, is shown in Figure 6. The first composition in each panel of Figure 6 represents the one used for training. These results indicate that the model has some predictive power on unseen compositions, giving better predictions on compositions closer to training set. For compositions with 40% of a particular atom and 20% of each of the other atoms, the model accuracy is roughly 20% of the property range. For compositions with 55% of a specific atom and 15% of each of the other atoms, the model accuracy is roughly 30% of the property range for relaxed vacancy formation energy and cohesive energy and 50% of the range for atomic volume.

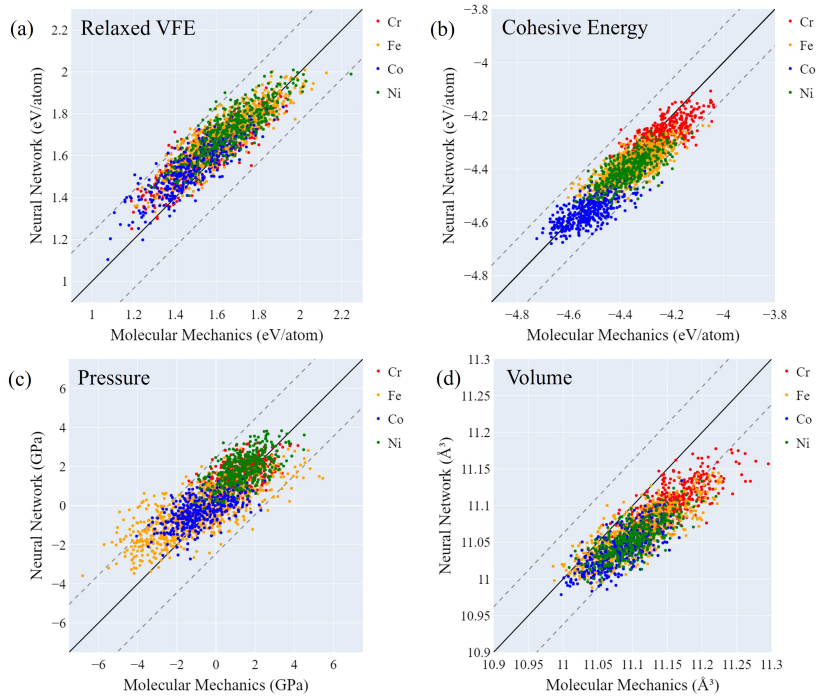

FIG. 4: Parity plots for $Cr_{20}Fe_{40}Co_{20}Ni_{20}$ for Relaxed VFE (a), Cohesive Energy (b), Pressure (c), and Volume (d). Predictions were made using model trained on equiatomic CrFeCoNi. The grey, dashed lines bound

$$\pm$$

20% of the range for each property.

B. Predicting properties for new chemistries: CrFeCoNiCu

Finally, we tested the model's ability to predict properties of systems with unseen elements. We used five models trained on single four-element alloys (CrFeCoNi, CrFeNiCu, FeCoNiCu, CrCoNiCu, and CrFeCoCu) to make predictions on CrFeCoNiCu. Results for vacancy formation energies are shown in Figure 7, with the other properties included in the supplemental information. Figure 7 indicates that the relaxed vacancy formation predictions of all elements on the CoCrCuFeNi are accurately described by the models trained on CrFeCoCu (missing Ni), CrFeNiCu (missing Co), and CrCoNiCu (missing Fe) but rather poorly by the models trained on FeCoNiCu (missing Cr) and CrFeCoNi (missing Cu); note that Cr and Cu are the end elements within our group in terms of atomic number.

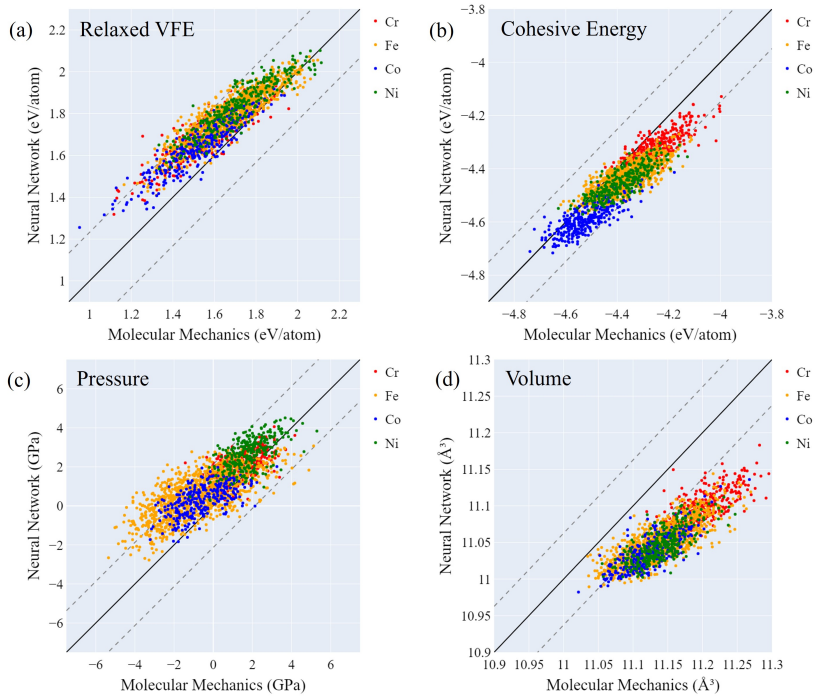

FIG. 5: Parity plots for $Cr_{15}Fe_{55}Co_{15}Ni_{15}$ for Relaxed VFE (a), Cohesive Energy (b), Pressure (c), and Volume (d). Predictions were made using model trained on equiatomic CrFeCoNi. The grey, dashed lines bound

$\pm$

20% of the range for each property.

To understand the underlying reason for these differences, we compared the inputs between the various alloys, specifically the unrelaxed bispectrum coefficients for CrFeCoNiCu with those for the four-element alloys. Figure 8 shows the distributions of the first coefficient. We find that the systems trained without Fe, Co, and Ni have relatively similar local descriptors (bispectrum coefficients) to the CrFeCoNiCu system. However, the descriptors for the alloys lacking Cu or Cr show significantly different distributions of descriptors as compared to the 5-element CCA. For FeCoNiCu (without Cr), the differences in the local environments are more pronounced than for CrFeCoNi (without Cu), explaining why the model shows very poor performance. We observe the same trends for the other bispectrum coefficients. This is due to the use of atomic number as prefactors in the construction of bispectrum coefficients. Ni, Fe, and Co lie between the elements trained on while Cr has

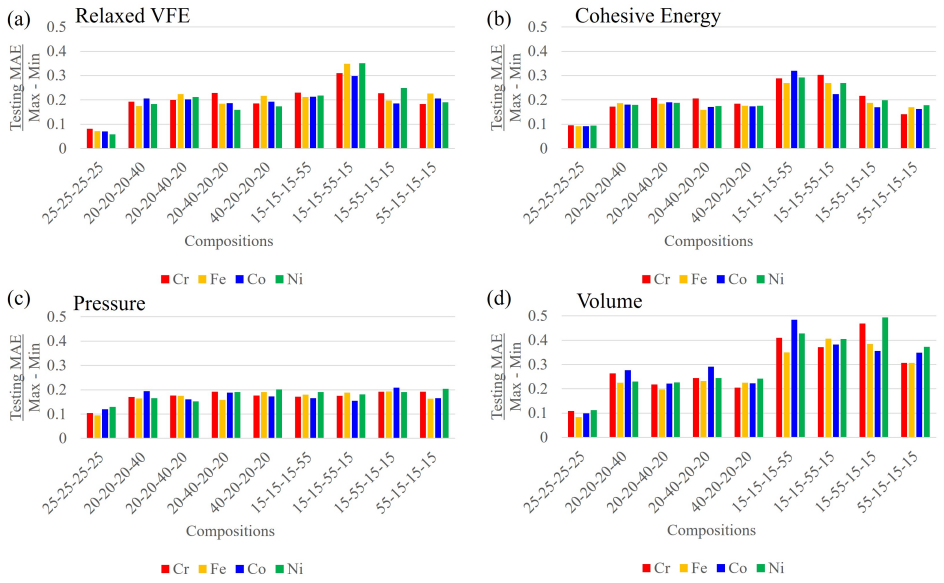

FIG. 6: MAE for predictions on untrained compositions for: (a) Relaxed vfe, (b) Cohesive Energy, (c) Pressure, and (d) Volume. The model was trained on equiatomic CrFeCoNi.

the lowest atomic number of the group and Cu has the highest atomic number.

IV. DISCUSSION AND CONCLUSIONS

We combined molecular statics, atomic level featurization, and data science to develop models for atomic properties in high entropy alloys from local atomic environment and elemental information. Our approach relates descriptors that are easy to obtain from unrelaxed atomic structures to properties that require atomic relaxations and, thus, are computationally more intensive to obtain. Evaluation of the models requires simply generating an atomic structure, performing a local structure calculation, computing atomic-based descriptors, and evaluating a neural network. For testing data, the model predictions were within 10% of the range for each of the properties studied. This level of accuracy is comparable with that of the pair approximation models of Daigle et al. when only the first neighboring cell is used. [23] The authors demonstrate improvements in accuracy as additional shells are included. We assessed the ability of our models to predict concentrations and chemistries not used during training, and we find that the model has can predict properties for several unseen concentrations and chemistries.

The local atomic properties modeled are important in determining several macroscopic

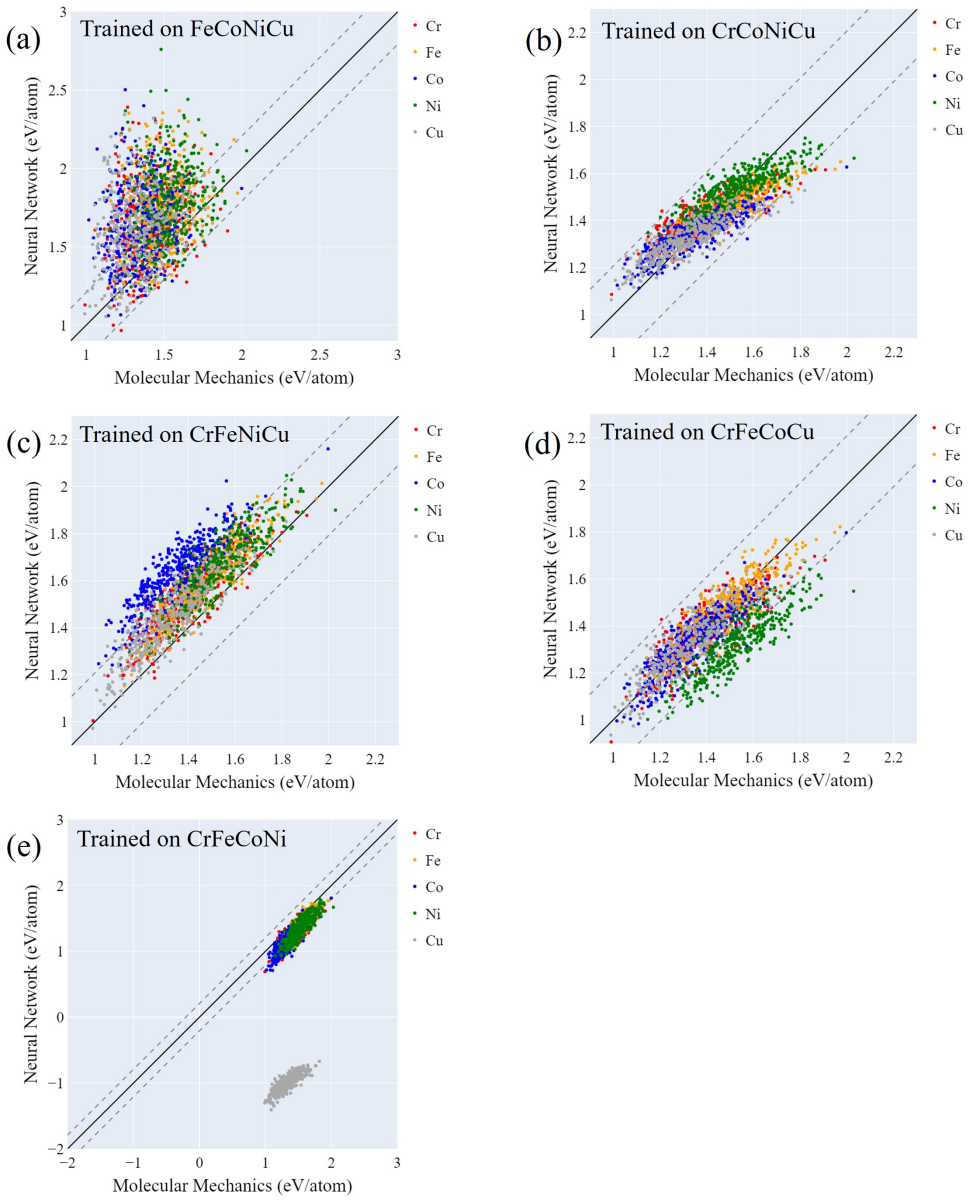

FIG. 7: Parity plots for relaxed vacancy formation energy for predictions on CrFeCoNiCu. Model was trained on FeCoNiCu (a), CrCoNiCu (b), CrFeNiCu (c), CrFeCoCu (d), and CrFeCoNi (e). The grey, dashed lines bound

$\pm$

20% of the range for each property.

properties of CCAs. As mentioned above, models for local volumes and stresses can inform single crystal strength models [8]. In addition, the distribution of VFEs affect vacancy con-

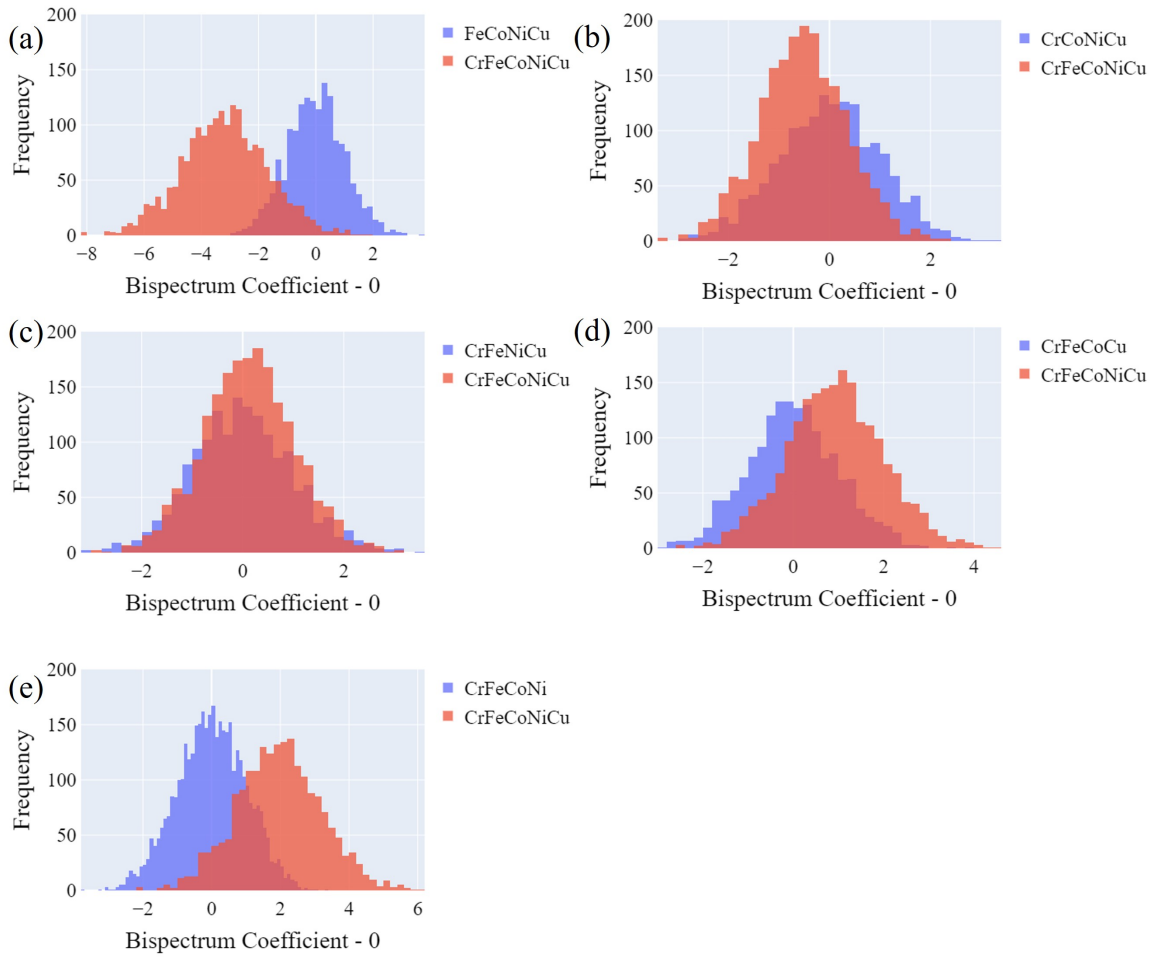

FIG. 8: Distribution for zeroth bispectrum coefficient for FeCoNiCu (a), CrCoNiCu (b), CrFeNiCu (c), CrFeCoCu (d), and CrFeCoNi (e) compared with CrFeCoNiCu. Bispectrum coefficient was normalized using the mean and standard deviation for FeCoNiCu (a), CrCoNiCu (b), CrFeNiCu (c), CrFeCoCu (d), and CrFeCoNi (e).

centrations. To exemplify the importance of capturing distributions, Figure 9 compares the equilibrium vacancy concentrations vs. inverse temperature for each element in a CrFeCoNi alloy considering the distribution of VFEs (solid circles) with the values assuming a constant value (set to the mean VFE for each element). The vacancy fraction calculated from neural network predictions of VFE compares well with the vacancy fraction calculated from molecular mechanics predictions of VFE. As also observed in shown borides, [23] a distribution of VFEs results in non-Arrhenius behavior as the relative contribution of different values is temperature dependent. All calculation details are includes as supplementary material in

online Jupyter notebooks [43].

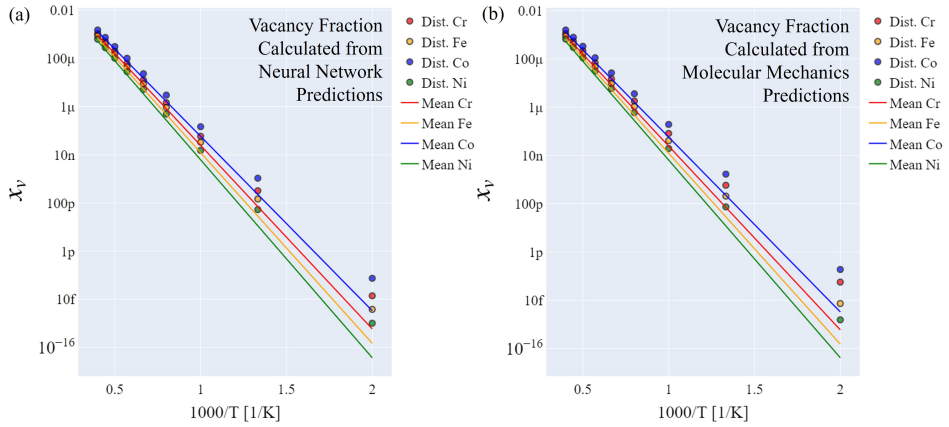

FIG. 9: Vacancy Fraction of HEA elements in an alloy given the mean VFE (solid lines), and calculating a population of vacancies based on the full distribution (circles) using neural network predictions (a) and molecular mechanics predictions (b)

In summary, atomic-level fluctuations in CCAs and other multi-principal component materials result in unique and often desirable properties. Our results indicate that atomic level simulations, appropriate descriptors, and machine learning tools can be used to capture such variability. In this paper we used properties computed from a many body force field for computational expediency, but the overall approach can be used with more accurate *ab initio* results.

V. DATA AND MODEL AVAILABILITY

The code developed and data used in this paper are available on the nanoHUB platform for online simulations [43]. The tool makes use of interactive Jupyter notebooks and includes the complete workflows in this work. They include: i) the generation of atomic structures, ii) determination of descriptors (bispectrum coefficients and chemistry based), iii) training of machine learning models, and iv) their application to predict various composition's properties within the CoCrCuFeNi alloy family. We also include code to plot the distributions of vacancy occupancy based on relaxed vacancy formation energy. While specific to this potential, the workflows and code provided are general enough to extend to other material systems.

VI. ACKNOWLEDGEMENTS

This effort was supported by the US National Science Foundation, DMREF program under contract number 1922316-DMR. The authors also acknowledge computational resources from nanoHUB and Purdue University.

---

[1] Jien-Wei Yeh, Swe-Kai Chen, Su-Jien Lin, Jon-Yiew Gan, Tsung-Shune Chin, Tao-Tsung Shun, Chun-Huei Tsau, and Shou-Yi Chang. Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes. *Advancd Engineering Materials*, 6(5):299–303, 2004.

[2] Brain Cantor, ITH Chang, P Knight, and AJB Vincent. Microstructural development in equiatomic multicomponent alloys. *Materials Science and Engineering: A*, 375:213–218, 2004.

[3] DB Miracle and ON Senkov. A critical review of high entropy alloys and related concepts. *Acta Materialia*, 122:448–511, 2017.

[4] Oleg N Senkov, Daniel B Miracle, Kevin J Chaput, and Jean-Philippe Couzinie. Development and exploration of refractory high entropy alloys—a review. *Journal of Materials Research*, 33(19):3092–3128, 2018.

[5] WG Nohring and WA Curtin. Correlation of microdistortions with misfit volumes in high entropy alloys. *Scripta Materialia*, 168:119–123, 2019.

[6] Francesco Maresca and William A Curtin. Mechanistic origin of high strength in refractory bcc high entropy alloys up to 1900k. *Acta Materialia*, 182:235–249, 2020.

[7] Francesco Maresca and William A Curtin. Theory of screw dislocation strengthening in random bcc alloys from dilute to "high-entropy" alloys. *Acta Materialia*, 182:144–162, 2020.

[8] Celine Varvenne, Aitor Luque, and William A Curtin. Theory of strengthening in fcc high entropy alloys. *Acta Materialia*, 118:164–176, 2016.

[9] George Kim, Haoyan Diao, Chanho Lee, AAT Samaei, Tu Phan, Maarten de Jong, Ke An, Dong Ma, Peter K Liaw, and Wei Chen. First-principles and machine learning predictions of elasticity in severely lattice-distorted high-entropy alloys with experimental validation. *Acta Materialia*, 181:124–138, 2019.

[10] Yan Zhang, Cheng Wen, Changxin Wang, Stoichko Antonov, Dezhen Xue, Yang Bai, and Yanjing Su. Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models. *Acta Materialia*, 185:528–539, 2020.

[11] Wenjiang Huang, Pedro Martin, and L Zhuang Zhuang, Houlong. Machine-learning phase prediction of high-entropy alloys. *Acta Materialia*, 169:225–236, 2019.

[12] Xianglin Liu, Jiaxin Zhang, Markus Eisenbach, and Yang Wang. Machine learning modeling of high entropy alloy: the role of short-range order, 2019.

[13] Weiliang Chen, Xueyong Ding, Yuchao Feng, Xiongjun Liu, Kui Liu, ZP Lu, Dianzhong Li, Yiyi Li, CT Liu, and Xing-Qiu Chen. Vacancy formation enthalpies of high-entropy fecocrni alloy via first-principles calculations and possible implications to its superior radiation tolerance. *Journal of Materials Science & Technology*, 34:355–364, 2018.

[14] Ronald Machaka. Machine learning-based prediction of phases in high-entropy alloys. *Computational Materials Science*, 188:110244, 2021. ISSN 0927-0256. doi: https://doi.org/10.1016/j.commatsci.2020.110244. URL `https://www.sciencedirect.com/science/article/pii/S0927025620307357`.

[15] Kevin Kaufmann and Kenneth S. Vecchio. Searching for high entropy alloys: A machine learning approach. *Acta Materialia*, 198:178–222, 2020. ISSN 1359-6454. doi: https://doi.org/10.1016/j.actamat.2020.07.065. URL `https://www.sciencedirect.com/science/article/pii/S1359645420305814`.

[16] J. M Rickman, G Balasubramanian, C. J Marvel, H. M Chan, and M.-T Burton. Machine learning strategies for high-entropy alloys. *Journal of applied physics*, 128(22), 2020. ISSN 0021-8979.

[17] Juan M Sanchez, Francois Ducastelle, and Denis Gratias. Generalized cluster description of multicomponent systems. *Physica A: Statistical Mechanics and its Applications*, 128(1-2): 334–350, 1984.

[18] Shiqiang Hao, Li-Dong Zhao, Chang-Qiang Chen, Vinayak P Dravid, Mercouri G Kanatzidis, and Christopher M Wolverton. Theoretical prediction and experimental confirmation of unusual ternary ordered semiconductor compounds in sr–pb–s system. *Journal of the American Chemical Society*, 136(4):1628–1635, 2014.

[19] I Toda-Caraballo, JS Wróbel, SL Dudarev, D Nguyen-Manh, and PEJ Rivera-Díaz-del Castillo. Interatomic spacing distribution in multicomponent alloys. *Acta Materialia*, 97:

156–169, 2015.

[20] JM Sanchez. Foundations and practical implementations of the cluster expansion. *Journal of Phase Equilibria and Diffusion*, 38(3):238–251, 2017.

[21] Alexander V Shapeev. Moment tensor potentials: A class of systematically improvable interatomic potentials. *Multiscale Modeling & Simulation*, 14(3):1153–1173, 2016.

[22] Alexander Shapeev. Accurate representation of formation energies of crystalline alloys with many components. *Computational Materials Science*, 139:26–30, 2017.

[23] SE Daigle and DW Brenner. Statistical approach to obtaining vacancy formation energies in high-entropy crystals from first principles calculations: Application to a high-entropy diboride. *Physical Review Materials*, 4(12):123602, 2020.

[24] Jörg Behler. Perspective: Machine learning potentials for atomistic simulations. *Journal of Chemical Physics*, 4:053208, 2016.

[25] Albert P Bartok, Mike C Payne, Risi Kondor, and Gabor Csanyi. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Physical review letters*, 104(13):136403, 2010.

[26] AP Thompson, LP Swiler, CR Trott, SM Foiles, and GJ Tucker. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *Journal of Computational Physics*, 285:316–330, 2015.

[27] Konstantin Gubaev, Evgeny V Podryabinkin, Gus LW Hart, and Alexander V Shapeev. Accelerating high-throughput searches for new alloys with active learning of interatomic potentials. *Computational Materials Science*, 156:148–156, 2019.

[28] Albert P Bartok, Risi Kondor, and Gabor Csanyi. On representing chemical environments. *Physical Review B*, 2013.

[29] S Plimpton. Fast parallel algorithms for short-range molecular dynamics, 1995. URL `http://lammps.sandia.gov`.

[30] Diana Farkas and Alfredo Caro. Model interatomic potentials and lattice strain in a high-entropy alloy. *Journal of Materials Research*, 33(19):3218–3225, 2018.

[31] Alex Zunger, S-H Wei, LG Ferreira, and James E Bernard. Special quasirandom structures. *Physical Review Letters*, 65(3):353, 1990.

[32] DH Tsai. The virial theorem and stress calculation in molecular dynamics. *The Journal of Chemical Physics*, 70(3):1375–1382, 1979.

[33] Chris Rycroft. Voro++: A three-dimensional voronoi cell library in c++. Technical report, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2009.

[34] Sergey N Pozdnyakov, Michael J Willatt, Albert P Bartók, Christoph Ortner, Gábor Csányi, and Michele Ceriotti. Incompleteness of atomic structure representations. *Physical Review Letters*, 125(16):166001, 2020.

[35] Yunxing Zuo, Chi Chen, Xiangguo Li, Zhi Deng, Yiming Chen, Jörg Behler, Gábor Csányi, Alexander V Shapeev, Aidan P Thompson, Mitchell A Wood, et al. Performance and cost assessment of machine learning interatomic potentials. *The Journal of Physical Chemistry A*, 124(4):731–745, 2020.

[36] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent Chevrier, Kristin A Persson, and Gerbrand Ceder. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis, 2013.

[37] Ankit Roy and Ganesh Balasubramanian. Predictive descriptors in machine learning and data-enabled explorations of high-entropy alloys. *Computational materials science*, 193, 2021. ISSN 0927-0256.

[38] nanoHUB. Jupyter notebook, Sep 2016. URL `https://nanohub.org/resources/jupyter`.

[39] Alejandro Strachan, Gerhard Klimeck, and Mark Lundstrom. Cyber-enabled simulations in nanoscale science and engineering. *Computing in Science & Engineering*, 12(2):12–17, 2010.

[40] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL `https://www.tensorflow.org/`. Software available from tensorflow.org.

[41] Francois Chollet et al. Keras, 2015. URL `https://github.com/fchollet/keras`.

[42] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12:2121–2159, 2011.

[43] Mackinzie Farnell, Zachary McClure, and Alejandro Strachan. Machine learning for high entropy atomic properties, 2021. URL https://nanohub.org/tools/mlatomprop.