


Machine-learning surrogate model for accelerating the search of stable ternary alloysM. Minotakis , H. Rossignol , M. Cobelli , and S. Sanvito 
School of Physics and CRANN Institute, Trinity College, Dublin 2, Ireland (Received 29 March 2023; revised 30 June 2023; accepted 22 August 2023; published 18 September 2023)

The prediction of phase diagrams in the search for new phases is a complex and computationally intensive task. Density functional theory provides, in many situations, the desired accuracy, but its throughput becomes prohibitively limited as the number of species involved grows, even when used with local and semilocal functionals. Here, we explore the possibility of integrating machine-learning models in the workflow for the construction of ternary convex hull diagrams. In particular, we train a set of spectral neighbor-analysis potentials (SNAPs) over readily available binary phases, and we establish whether this is good enough to predict the energies of novel ternaries. Such a strategy does not require any new calculations specific for the construction of the model, but just avails of data stored in binary-phase-diagram repositories. We find that a so-constructed SNAP is capable of accurate total-energy estimates for ternary phases close to the equilibrium geometry but, in general, is not able to perform atomic relaxation. This is because during a typical relaxation path, a given phase traverses regions in the parameter space poorly represented by the training set. Different metrics are then investigated to assess how well an unknown structure is described by a given SNAP model, and we find that the standard deviation of an ensemble of SNAPs provides a fast and non-specie-specific metric. Overall, we show that it is possible to train machine-learning interatomic potentials on readily available binary-compound data to effectively screen ternary compounds in a high-throughput search.

DOI: [10.1103/PhysRevMaterials.7.093802](https://doi.org/10.1103/PhysRevMaterials.7.093802)**I. INTRODUCTION**

The rapid advancements in existing technologies such as logic electronic devices, as well as the need to develop new energy solutions, make the search for novel materials of paramount importance. Following the improvement in hardware performance and the development of user-friendly *ab initio* algorithms for materials modeling [1–3], it has become possible to use such computational methods to accelerate the discovery process. Successful examples of theory-driven materials discovery include the development of Li-ion cathodes [4], high entropy alloys [5], and magnetic materials [6]. The state-of-the-art workflow takes the form of a high-throughput search [7]. This makes use of efficient density-functional-theory (DFT) calculations to predict the material properties of a large pool of compounds created virtually, in the hope that these contain some “hidden compounds” yet to be discovered [8]. Collections of such prototypes exist in material databases [9,10]. The first step in the workflow is a stability screening, which ensures that any compound selected is chemically reasonable and thermodynamically stable, namely that it might form. Such screening requires the calculation of the total energy of the system. In this work, we focus on demonstrating that machine learning (ML) can assist this first step.

Two key ingredients are needed to find new stable phases: a method for generating candidate structures and a method for evaluating their energies. The first aspect, which is not the focus of this work, is usually performed by either constructing a library of prototypes [6] or via dedicated methodologies for proposing novel structures [11–13]. Total energies are then typically evaluated by local/semilocal DFT, which ensures a general good accuracy at a moderate computational cost.

Unfortunately, even when combined with simple approximations, performing *ab initio* calculations remains the rate-limiting step of any materials search workflow, posing constraints on both the number of candidate structures that can be tested and the number of atoms in the unit cells of such structures. To speed up this process, ML has been deployed in a number of ways. For instance, it has been used to predict better starting charge densities for the self-consistent loop needed by Kohn-Sham DFT [14–16]. Furthermore, ML has also been used in active-learning frameworks to accelerate geometry optimization [17], as well as in *ab initio* molecular dynamics (AIMD) [18] through the direct prediction of energies and forces. In these cases, ML models are used to map local atomic configurations to energies and forces by tuning parameters based on training data provided from DFT. Once the models are trained, subsequent predictions are orders of magnitude faster than DFT calculations. We follow this approach in our work by developing a machine-learning model acting as a surrogate for the energy predictions made by DFT, which could act as a screening tool in high-throughput searches.

Several approaches have been employed to construct energies surrogate to the DFT ones in the computational materials discovery process [19,20]. A commonly used strategy is the cluster expansion, in which the energy of a system is written as a sum of energy contributions from different clusters made up by the constituent atoms. Here, the strengths of the effective cluster interactions are computed by fitting the model to a set of DFT energies. Thus, this surrogate model uses DFT data to perform interpolation on new structures [21], a task that is also performed when using machine-learning techniques. A swarm of studies have utilized this class of methods in the

context of materials discovery. These include, for instance, component predictions [22], where the most likely compositions for a given set of atomic species are computed, or structure predictions [23], where the most likely crystal structure is forecast, as well as for the direct prediction of the distance of a material's energy from the convex hull [24,25]. In general, these schemes make use of databases with a variety of materials made of different species, in combination with various machine-learning algorithms such as Bayesian optimizers, random forests, and support vector machines. The feature vectors typically include general information about the material structure as well as species characteristics of the constituent atoms. However, they are not typically trained to distinguish small differences between compounds with similar structures. For this task, one has to rely on machine-learning interatomic potentials (MLIAPs).

MLIAPs combine fingerprints of the atomic configurations, most frequently locally defined, with a ML algorithm to predict energies, forces, and stress tensors. Several atomic fingerprints, acting as the feature vectors, have been successfully deployed. Their locality enforces invariance under translation and atomic permutation, while they are usually constructed to be locally rotational invariant. The most successful MLIAPs notably include Behler-Parinello symmetry functions, combined with neural networks in the neural network potential (NNP) [26], bispectrum components, combined with ridge regression in spectral neighbor analysis potential (SNAP) [27] and quadratic SNAP (qSNAP) [28], the smooth overlap of atomic position (SOAP) descriptors with Gaussian process regression in Gaussian approximation potential (GAP) [29,52], invariant polynomials with linear regression in moment tensor potential (MTP) [30], and N -bond basis functions with linear regression in atomic cluster expansion (ACE) [31]. As for the cluster expansion method, MLIAPs make use of a DFT data set to fit the model parameters, and they are capable of predicting energies and forces at *ab initio* accuracy, provided these are made for structures for which the model interpolates [32]. This makes such potentials ideal for molecular-dynamic simulations at high accuracy, for large systems and over long timescales [33–36].

In the context of predicting materials stability, MLIAPs are used to map the potential energy surface of multiple phases and hence to reconstruct the $T = 0$ K phase diagram to determine the lowest energy structures, namely to construct the convex hull. SNAPs and NNPs have been previously used against this task for metallic alloys [37–40]. In the case of SNAP, however, the range of different structures and stoichiometries probed was limited and the potential was not used to find new stable alloys. Instead, in the case of NNPs, the training set used was very large ($\sim 10^3$ – 10^4 structures), so that the actual structures computed by DFT were as many as those needed to construct a fully *ab initio* convex hull. This is also the case of the GAP models trained as general potentials across the phase diagram of C [41] and Si [42]. One of the very few examples of using a MLIAP trained over a limited number of structures to predict materials stability at an accelerated pace has been recently provided by Gubaev *et al.* [43]. In their work, *ab initio* calculations were performed over between 383 and 2393 structures to train a MTP able to reproduce binary and ternary convex hulls. These structures were selected

through an efficient active-learning process [44], which probed on the order of 10^4 – 10^5 configurations for each phase diagram. This method proved that MLIAPs could accelerate the computational high-throughput search of new alloys.

In this work, we show how SNAPs can be used to drastically accelerate the search for new stable ternary intermetallic compounds, without the need to generate large training DFT data sets. Our approach for constructing a MLIAP for rapid screening is similar in philosophy to the *specialized* MLIAP training proposed by Artrith *et al.* to compute the binary convex hull of Li_xSi [45]. Our selection of the training database, however, is different. In fact, rather than investing resources on curating a training set and on running DFT calculations for the sole purpose of training the MLIAP, we make use of existing materials convex hull databases, namely the AFLOWlib [9] repository. In this way, data are already available and the computational efforts put into any *ab initio* calculations are also relevant for the convex hull construction. In AFLOWlib, the binary phase diagrams are typically extensively explored, meaning that there is a significant range of data available and that there is a lower probability of finding new stable binary alloys. The ternary hulls are usually not as rich, despite there being a combinatorial explosion of the number of possible derivative structures that can be created from a prototype structure. This leaves more room for exploration. For ternary systems, the enthalpic term in the Gibbs free energy is still significant with respect to the entropic one, unlike quaternary compounds [46], meaning that free-energy calculations are still relevant. The method proposed here then exploits SNAP to guide the screening of the ternary space.

The manuscript is organised as follows. In the next section, we will present the main computational ingredients needed for our workflow, namely the AFLOWlib data set, the DFT numerical implementation used, and the SNAP model trained. Then we will proceed with presenting the results for three prototypical ternary systems, namely one composed by noble metals Cu-Ag-Au, and two composed by early, mid, and late transition metals, namely Ti-Mo-Pt and Cd-Hf-Rh. In that section, we will discuss the SNAP training and its performance against known ternary compounds and newly constructed prototypes. Several error metrics are then presented, which are aimed at identifying compounds that lie outside of the model's training space. Finally, we will conclude.

II. METHODS

This work explores the ability of SNAP to be used as an efficient energy predictor of novel ternary compounds. Our general philosophy, however, is to achieve such a goal without generating DFT data serving the sole purpose of training the ML model, but instead we aim at reusing the same DFT data computed to construct the binary phase diagrams. As such, the constructed ML model will have a negligible computational payload. With this in mind, the structures used to train SNAP are all taken from the AFLOWlib database [47]. Then the SNAP is trained on the total energies of binary crystalline compounds and tested on ternary materials, either in their equilibrium geometry or as suggested from a prototypes' generator. Different error metrics that can be used to identify structures far from the training set are then assessed. Here, all

DFT energy values are obtained with the VASP package [1]. Computational details are given in the upcoming sections.

A. AFLOWlib data

The training data for all the models are directly taken from the AFLOWlib repository [9] through the AFLOW-CHULL tool set [47]. The access and manipulation of the data is performed with the AFLOW application programming interface (API) [48]. In particular, we consider three ternary systems—Cu-Ag-Au, Ti-Mo-Pt, and Cd-Hf-Rh—for which we have extracted data for all the binary and ternary phases contained in AFLOWlib.

B. DFT

Ab initio total-energy calculations are performed using the Vienna *Ab initio* Software Package [1] (VASP), which makes use of periodic boundary conditions, a plane-wave basis set, and the projected-augmented-wave (PAW) method with pseudopotentials. All calculations are performed with a plane-wave cutoff of 600 eV and an energy convergence criterion of 10^{-4} eV. The standard generalized gradient approximation as parametrized by Perdew, Burke, and Ernzerhof [49,50] is used throughout, together with the corresponding VASP pseudopotential library.

For all DFT calculations, we use the convergence criteria set of the AFLOW standard [51]. The k -mesh is constructed with the Monkhorst-Pack scheme and ensuring that the mesh is Gamma-centered for the hexagonal (hP) and rhombohedral (hR) Bravais lattices. The number of sampling points, N_i , is proportional to the norm of each corresponding reciprocal Bravais lattice vector, \vec{b}_i , and they are minimized ensuring the following condition:

$$N_{\text{KPPRA}} \leq \min \left[\prod_{i=1}^3 N_i \right] \times N. \quad (1)$$

Here, N_{KPPRA} is the number of k -points per reciprocal atom, and N is the number of atoms in the cell. In particular, N_{KPPRA} is chosen at 10000 for all static calculations and at 6000 for all geometry relaxations. The geometry relaxations are considered converged when the atomic forces are smaller than 10^{-3} eV/Å.

C. Spectral neighbour analysis potential

Our machine-learning model of choice for total energy predictions is SNAP [27], which is briefly recalled here. Like many MLIAPs, SNAP assumes that the total energy of an N -atom system (molecule or crystal) can be broken down into individual contributions, E_i , associated with each atom i and element Z_i ,

$$E_{\text{tot}} = \sum_i E_i^{Z_i}. \quad (2)$$

Each energy contribution $E_i^{Z_i}$ is assumed to be linearly dependent on a feature vector, $\vec{B}_i^{Z_i}$, describing the local atomic environment, where the coefficient of expansion, $\vec{\alpha}^{Z_i}$, repre-

sents the training parameters of the model,

$$E_i^{Z_i} = \vec{\alpha}^{Z_i} \cdot \vec{B}_i^{Z_i}. \quad (3)$$

In SNAP the descriptors, $\vec{B}_i^{Z_i}$, are the bispectrum components [29], quantities that are invariant upon local rotations.

In a nutshell, the atomic-neighbor density function within a sphere of radius R_c and centered at the i th atom at \vec{R}_i can be written in terms of a sum over the neighbors j within the sphere,

$$\rho_i(\vec{r}) = \delta(\vec{r} - \vec{R}_i) + \sum_j w_{Z_j} \delta(\vec{r} - \vec{R}_j) f_c(R_{ij}), \quad (4)$$

where $R_{ij} = |\vec{R}_i - \vec{R}_j|$, w_{Z_j} are weights associated with each atomic species (treated as hyperparameters), and f_c is a cutoff function, as defined by Behler and Parinello [26]. Such density is then projected onto the four-dimensional sphere of radius r_0 and expanded over hyperspherical harmonics $U_{m',m}^J(\theta, \phi, \theta_0)$ as

$$\rho_i(\vec{r}) = \sum_{J=0}^{\infty} \sum_{m,m'=-J}^J c_{m',m}^J U_{m',m}^J(\theta, \phi, \theta_0). \quad (5)$$

Here, J , m , and m' are parameters distinguishing the individual hyperspherical harmonics, and $c_{m',m}^J$ are the appropriate expansion coefficients. Details on the conversion from \vec{r} to the four-dimensional polar angles can be found in Ref. [52]. In practice, the expansion is truncated at $J = J_{\text{max}}$. The bispectrum components are then built as an appropriate triproduct of the expansion coefficients,

$$B_i^{J_1, J_2} = \sum_{m'_1, m_1 = -J_1}^{J_1} c_{m'_1, m_1}^{J_1} \sum_{m'_2, m_2 = -J_2}^{J_2} c_{m'_2, m_2}^{J_2} \times \sum_{m', m = -J}^J C_{mm_1 m_2}^{J, J_1, J_2} C_{m' m'_1 m'_2}^{J, J_1, J_2} (c_{m', m}^J)^*,$$

where $C_{mm_1 m_2}^{J, J_1, J_2}$ and $C_{m' m'_1 m'_2}^{J, J_1, J_2}$ are the Clebsch-Gordan coefficients, which determine the coupling between the different values of J . The individual components with J values between 0 and J_{max} are then collected to form the vector \vec{B}_i .

The calculation of the bispectrum coefficients is performed with the LAMMPS software [53], while the energy fitting is performed with an in-house Python library that makes use of the SCIKIT-LEARN package [54]. The hyperparameters of the model are J_{max} , R_c , and the set of weights w_{Z_j} , which are optimized with TPE [55–57] as implemented in OPTUNA [58].

III. RESULTS AND DISCUSSION

The first ternary system selected comprises three noble metals: Cu, Ag, and Au. These display less complexity in their chemical behavior than other transition metals with partially filled d -shells, so that they represent a good playground to present our concept. The process chosen for identifying novel ternary prototypes involves training a SNAP model on data *already* available and only related to binary phases. The so-constructed ML model is then used to predict the energy, and hence the stability, of a range of ternary compounds that, by definition, do not appear in the training set.

A. Fitting energies for binary phases

As a first step, a series of SNAP models are trained individually over each of the three binary systems, namely Ag-Au, Cu-Ag, and Cu-Au, for which AFLOWlib contains 261, 190, and 260 structures, respectively. The associated unary systems are included as well. Although we could have proceeded by using the AFLOWlib energies directly, for consistency we have preferred here to rerun static DFT calculations, according to the standards presented in Sec. II B, for all the unary and binary structures. Within each binary system, the data set is split into a training and a cross-validation set, while the ternary phases form the test set. The split is 80% training and 20% cross-validation, as suggested by the learning curves shown in the Appendix. Monte Carlo cross-validation is the strategy of choice for splitting the data sets. Since there is significant structural diversity in the data sets, and in order to avoid the effect that a small subset of structures may have on the training, an ensemble of 10 SNAP models is trained for each binary system. The average of the predictions made by each SNAP model of the ensemble is used as the final output. We call this ensemble of SNAP models *SNAPs*. Details on the optimal hyperparameter selection for J_{\max} , R_c , and w_{Z_i} are given in the Appendix.

The accuracy of our models is assessed by computing both the mean absolute error (MAE) and the root-mean squared error (RMSE), averaged over the 10 different SNAP models constructed for each binary system. The results are presented in Table I for both the training and cross-validation sets, while a visual appreciation of the performance is shown in the parity plot of Fig. 1 that displays the results for one SNAP model randomly chosen from the ensemble. We find that all three binary systems are fitted to a similar standard. The cross-validation RMSEs are on the order of 10^{-2} eV/atom due to the presence of high enthalpy structures, present for all systems. These have proved more difficult to fit, even when present in the training data, as demonstrated by their departure from the parity line in Fig. 1. When looking at the hyperparameter optimization, we find the R_c value of Ag-Au to be lower than that of the other two binary systems. It is optimal to have a larger weight for Au, while having a lower weight on Cu compared to the other two elements helps to reduce the error.

B. Testing over ternary compounds

The SNAP models presented in the previous section have been individually trained over each binary system, i.e., they contain information for only two species at a time. These cannot be utilized to predict a ternary structure. As such,

TABLE I. Summary of the average errors over 10 SNAP models for the three binary systems. T = training, CV = cross-validation.

Error (meV/atom)	Ag-Au	Cu-Ag	Cu-Au
MAE (T)	2.63	2.82	4.30
RMSE (T)	4.37	4.01	6.25
MAE (CV)	6.85	7.27	8.57
RMSE (CV)	13.83	15.15	15.73

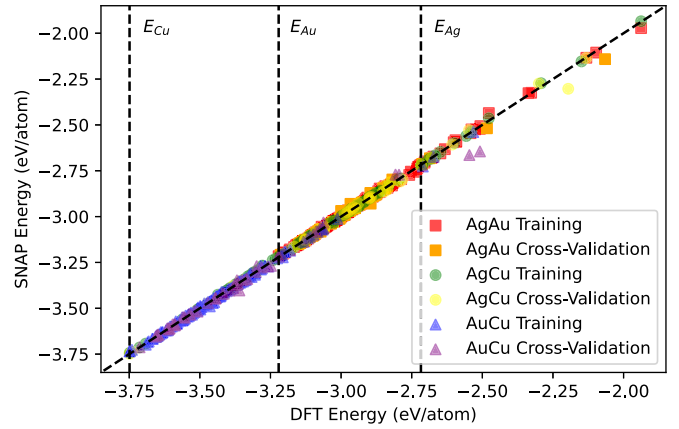


FIG. 1. Parity plot, showing the SNAP-predicted energies against the DFT ones, for the Ag-Au, Cu-Ag, and Cu-Au binary systems. The DFT energies of the lowest-energy single-element structures are marked by vertical dashed lines.

we now train another ensemble SNAP, this time over the entire library of unique unary and binary compounds available (677 after removing unaries that appear twice), by using the same strategy described before (e.g., an 80/20 training/cross-validation split; see the hyperparameters in the Appendix). The SNAPs ensemble model is then tested over the 78 ternary structures contained in AFLOWlib for Cu-Ag-Au. The parity plot for this new model is presented in Fig. 2, where again data are presented for the SNAP predictions of a model randomly chosen from the 10.

The figure clearly shows that our ensemble SNAP performs almost identically over the binary training set and the ternary test one, demonstrating that models having knowledge of enough binary structures are capable of offering good energy predictions for ternaries. More quantitatively, the RMSE of the model is 5.87 meV/atom for the training set and it actually decreases marginally to 4.84 meV/atom for the test one. Interestingly, this is even lower than the cross-validation error found for the individual binary SNAPs (see Table I), a result that we attribute to the more extended diversity of the chemical environments that the model has to fit now.

It is therefore established that by training on data extracted from the binary phase diagrams SNAP models are able to predict the ground-state energy of fully relaxed ternary structures.

C. Ternary prototypes

To put our SNAP models against a more severe task, we now investigate whether these can be used to predict the energy of novel prototypes not present in the AFLOWlib database. The dictionary method, as implemented in the AFLOW encyclopedia [59–61], is initially employed to create 42 new structures. These, in general, span a wide energy range and most of them are far away from the convex hull. Note, in fact, that the lattice parameters of these prototypes are not optimized, but just estimated through a Vegard-like law. Then, DFT relaxation is performed for all the newly created ternaries until the forces are below 10^{-3} eV/Å. We finally assess the extrapolation ability of the SNAP models to predict

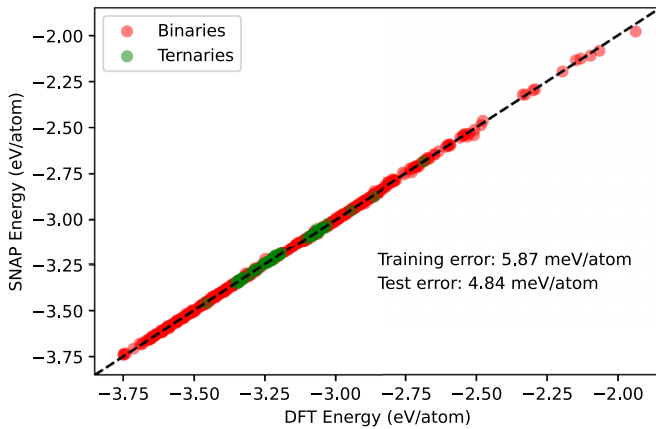


FIG. 2. Parity plot, showing the SNAP-predicted energies against the DFT ones, for both the binary (training and cross-validation set) and ternary (test set) compounds contained in AFLOWlib for the Cu-Ag-Au system. Here the SNAP models forming the ensemble have been trained on the entire pool of binary phases. The RMSE is reported in the legend.

the ground-state energy of both the relaxed (R) and the initial nonrelaxed (NR) ternary prototypes. The ensemble SNAP used here is the same as the one introduced in Sec. III B. Our results are presented in Fig. 3.

The figure shows that once the structures are fully relaxed by DFT, the error remains satisfactory, confirming that the good energy estimate for ternary compounds does not remain limited to the AFLOWlib data. In this case, the RMSE grows from 4.84 meV/atom for the original ternaries included in AFLOWlib to about 27.74 meV/atom for the new relaxed prototypes, reflecting the more diverse set of structures generated for this test. In contrast, the error is significantly higher

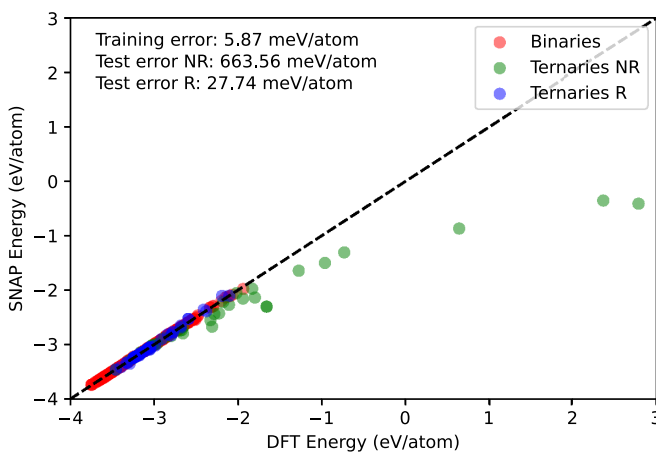


FIG. 3. Parity plot, showing the SNAP-predicted energies against the DFT ones, for both the binary compounds contained in AFLOWlib (training set) and for a new set of ternary phases either in their prototypes, nonrelaxed (NR), geometry or after full DFT relaxation (R). Here the SNAP models forming the ensemble have been trained on the entire pool of binary phases as described in Sec. III B. All data are for the Cu-Ag-Au system.

when one tries to estimate the energy of the prototypes as constructed, namely before DFT relaxation. We now find a RMSE of 663.56 meV/atom, with SNAP systematically underestimating the DFT energy. Interestingly, there is still a fraction of the created structures on which SNAP performs well, and these appear to have lattice parameters close to their relaxed ones. Surprisingly, SNAP is able to correctly identify the DFT energy trend, suggesting that such a trained model can be employed as an effective screening tool in materials prediction workflows. In general, this analysis proves that an ensemble SNAP model, constructed on binaries, can extrapolate to associated generic ternary structures when these are close to their equilibrium geometry and can accurately estimate the energy ordering for the nonequilibrium ones.

The remaining question is then whether or not the so-constructed SNAP can be used to drive the atomic relaxation. This is investigated by employing the LAMMPS package [53] for geometry minimization, where the energy and force convergence criteria are set at 10^{-4} eV and 10^{-3} eV/Å, respectively. The relaxation is performed in two steps, where first the atomic positions are optimized, and then we relax both the cell parameters and the atomic positions. This procedure is repeated five times for each structure to ensure convergence. Again the mean SNAP model, averaged over all those of the ensemble, is chosen to perform the relaxation. We find that, although the resulting SNAP-optimized structures generally have a lower DFT-computed total energy than the unrelaxed ones, they are still far from the optimal DFT-computed geometries. This means that, although the ensemble SNAP is capable of some relaxation, in general it is not able to find the equilibrium structure.

To obtain some insight into this aspect, principal component analysis (PCA) is performed on the feature vectors forming the training set and on those of the structures encountered along the DFT relaxation path of the ternary prototypes. This analysis is performed for each one of the three species considered here, namely Cu, Ag, and Au. An illustrative plot of the PCA of the first two components is shown in Fig. 4 for Ag, and similar graphs have been obtained for the other two species. In the plot, blue circles represent the PCA components associated with structures included in the training set (binary phases), while the colored ones are for a DFT relaxation trajectory starting from a new ternary prototype. Clearly, the binary feature space is quite heterogeneously distributed, with relatively large portions poorly known by SNAP. The DFT relaxation is found to travel sparsely populated regions, in particular for the starting structure. As a result, energy and forces during the SNAP-driven relaxation may be poorly predicted, since the relaxation trajectory has to travel regions in the parameter space of which the SNAP has little knowledge. This then results in structures that differ from the *ab initio* relaxed ones.

There are a number of possible strategies to mitigate this problem. On the one hand, one can simply generate additional DFT-computed prototypes, cleverly covering a more uniform distribution of local atomic environments, and enlarge the SNAP training set. This avenue is usually pursued in conjunction with some active-learning strategy [18,29,44,62] to generate force fields for stable molecular-dynamics simulations, but it often requires a rather large number of new DFT

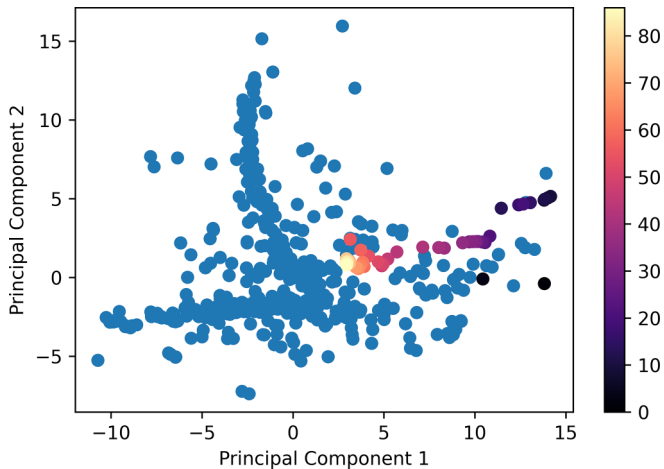


FIG. 4. Plot of the first two principal components of a PCA performed over the Ag bispectrum of the training set (blue circles). The colored circles are for the ternary structures encountered during the *ab initio* relaxation path of a prototype ternary compound. The color code encodes the relaxation-step index. Note that the DFT relaxation begins in a region poorly covered by the training set, but ends in a rather high-density region.

calculations. When the problem is that of constructing phase diagrams, this does not seem like an ideal solution, since the DFT effort is mostly spent in structures of no particular interest other than training the ML potential, and the total DFT-calculation count may be similar to that of computing the entire phase diagram from DFT alone. On the other hand, one can notice from Fig. 4 that the typical DFT relaxation moves towards regions of the feature space well covered by the binary training set. This means that constructing the initial prototypes with local atomic environments closer to the training set may represent a better initial choice for novel ternary phases that are closer to the final optimized structures and can be naturally relaxed with SNAP.

We close this section by discussing how our work differs from a more “conventional” construction of MLIAPs. First, recall that the aim of the SNAPs model presented here is not to act as a robust force field suitable for accurate molecular-dynamics simulations across the phases of a ternary system, but rather to be used as a screening tool in the context of the high-throughput search for stable materials. As such, we seek a model error close to or lower than 50 meV/atom, consistent with other high-throughput studies that make use of a machine learning energy screening tool instead of or in combination with DFT [24,63–65]. This comes as a consequence of the choice of training database. In fact, the training structures are not generated using scientific intuition or an automatic scheme, but are taken directly from a readily available online repository, AFLOWlib in our case. The motivation for this choice is twofold. On the one hand, only preexisting data are used to avoid having to curate a tailor-made database by running DFT calculations specific for it. On the other hand, all the DFT calculations performed by AFLOWlib to create the database are directly relevant for the construction of the convex hull, since they concern structures with a variety of compositions and space groups. Such diversity in the struc-

tures is necessary for the robustness in the determination of the phase diagram.

To provide a more quantitative understanding of this issue, we now give a brief comparison between the training databases used in this study and those employed in the construction of several different MLIAPs. The work of Chen *et al.* [66] offers a good example of how a typical SNAP model is deployed as a general force field, in this case for Mo. This has been designed to accurately predict energies, forces, and stress tensors, as well as more complex properties, such as the lattice constants, vacancy formation, surface, and grain boundary energies, the phonon spectrum of bcc Mo, and even the melting temperature. SNAP must then be able to run accurate MD simulations over large lengthscales and timescales (~ 2000 atoms for 500 ps). The training database was physically informed with $\sim 10^3$ training points (note that in this case only one element is present). These included ground-state phases from the Materials Project repository [10] (eight structures), surface slabs, vacancy, and grain boundary structures (respectively, 11, 24, and 13), along with large structures (54 atoms) that are elastically deformed (67 structures) and used for different sets of AIMD simulations (169 structures). All of these data classes are necessary to expose SNAP to a variety of atomic configurations, notably with missing atoms (surfaces and vacancies) and unusual angular and radial distributions (grain boundaries). Such data diversity enables the force field to be robust during the MD simulations. The role of our model is not to achieve that type of accuracy, but rather to work as a rough screening tool for the energy of different structures across a phase diagram, meaning that our training database would only include the Materials Project part of the database [10], namely about 1% of the data.

The work of Gubaev *et al.* [43] provides another suitable example of a MLIAP, in this case trained to construct phase diagrams. The moment tensor potential (MTP) framework is chosen to accelerate the search for novel stable alloys as exemplified for the Cu-Pd, Co-Nb-V, and Al-Ni-Ti systems. Their force field is an efficient screening tool, and it possesses the ability to perform relaxation. The training database is constructed on the fly using an active learning scheme, centered around the extrapolation grade (presented here in Sec. III E). A set of structures (10^4 – 10^5) are created from hcp, bcc, and fcc supercells and are decorated for different compositions using the methodology presented in Ref. [67]. The maximum number of atoms varies from 8 to 20 depending on the system. The MTP is then trained on the fly as the structures are being relaxed. This leads to having 523 and 383 training structures for Cu-Pd and Co-Nb-V, respectively. For Al-Ni-Ti, two different MTPs are necessary, one with 2393 and the other with 976 training structures. It is interesting to note that for this system, AFLOWlib prototypes are also used in the training set, highlighting the additional diversity brought by such structures. The key difference with our work is that many of the structures used for the training of the MTPs are off equilibrium as they appear in the relaxation path. These are judged by the active learning scheme as being the most relevant for training a robust force field. In our case, all structures included are at equilibrium, they do not need to be generated, and they are directly relevant for the convex hull construction. Note that in both cases, forces and stress

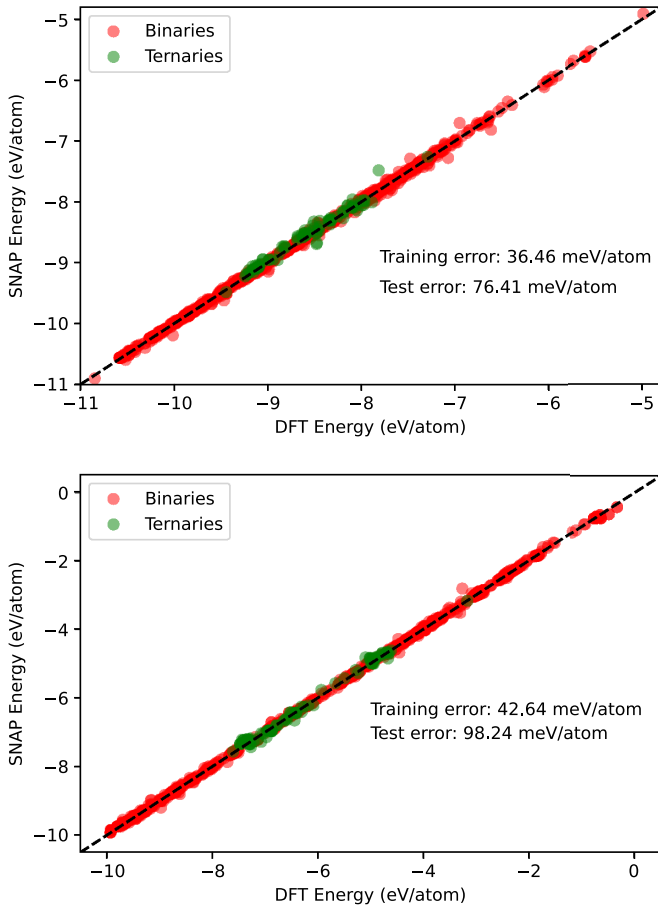


FIG. 5. Parity plot, showing the SNAP-predicted energies against the DFT ones, for both the binary (training and cross-validation set) and ternary (test set) compounds contained in AFLOWlib. Data are presented for the Ti-Mo-Pt (upper graph) and for the Cd-Hf-Rh (lower graph) ternary systems. Here the SNAP models forming the ensemble have been trained on the entire pool of binary phases. The RMSE is reported in the legend.

tensors are used for training, as opposed to our study, since such quantities are not provided by AFLOWlib.

D. Ternary phase diagrams for Ti-Mo-Pt and Cd-Hf-Rh

To validate our results and to demonstrate that the method works well beyond noble metals, we have performed the same workflow on two different new ternary systems, namely Ti-Mo-Pt and Cd-Hf-Rh. These have been selected based on two criteria: (i) they contain phases made of early (Ti and Hf), mid (Mo and Rh), and late (Pt and Cd) transition metals, thus presenting a chemical variety larger than that found in noble-metal compounds; (ii) for these ternary systems, AFLOWlib contains a high enough number of prototypes, enabling the training of a reliable model. In fact, these two systems contain more than 1500 compounds, with around 90 ternaries for each set. It should be noted, however, that for each system, one binary combination is usually over-represented when compared to the others. These are Ti-Mo for Ti-Mo-Pt and Cd-Hf for Cd-Hf-Rh. Importantly, the full database of DFT energies was taken from AFLOWlib, rather than having been recalculated.

The ensemble SNAP parity plots are shown in Fig. 5, along with the RMSE of the fit. Also in this case the training (cross-validation) set comprises 80% (20%) of all available unary and binary compounds, the test set is only made of DFT-relaxed ternary phases, and careful optimization of the hyperparameters R_c , J_{\max} , and w_{Z_i} is performed. From the figures, it is clear that for these relaxed structures, taken from AFLOWlib's convex hull, the predictions are satisfactory and the parity line is followed closely. The absolute errors on all sets are higher than those found for Cu-Ag-Au, but the total-energy percentage error is similar. In fact, for Cu-Ag-Au, the percentage error relative to the average energy per atom was 0.19% for the training set and 0.15% for the test set. These must be compared to 0.40% and 0.86% for Ti-Mo-Pt and to 0.74% and 1.57% for Cd-Hf-Rh. It should be noted that the ternary convex hull diagrams for Ti-Mo-Pt and Cd-Hf-Rh are much deeper than that of Cu-Ag-Au as the lowest points are at an enthalpy of formation of 951 and 921 meV/atom, respectively, compared to 61 meV/atom for Cu-Ag-Au. This makes the SNAP absolute error less significant.

E. Error metrics

The aim of this work is to develop a ML model surrogate to DFT to predict total energies and hence to identify stable ternary compounds. As shown in the previous sections, a SNAP potential constructed over binary phases is able to provide accurate energy predictions for ternary compounds only close to their equilibrium structure, but it is not reliable to perform relaxation. This is because during the relaxation process, the system crosses regions in the parameter space poorly covered by the training set. Therefore, it would be of interest to have a metric capable of distinguishing structures that are far from the feature space spanned by the training set from those that are within it. We expect SNAP to perform well for the second set of structures but not for the first. In this section, three such metrics are presented, and their effectiveness is assessed.

The first error metric used is the Euclidean distance, d_{\min} , which can be defined here in different ways. The easiest approach is to take the distance between the feature vector of the test system and the average feature vector of the training set. Unfortunately, the total feature vector is made up of a sum of bispectrum components, some of which are associated with different atomic species. Instead, the so-defined distance is evaluated independently of each species. For a given species, two sets are created, namely the set of all bispectrum components in the training set and all those in the test set. Then the Euclidean distance between the vectors of each set is evaluated, and the metric used is the minimum of all these distances, namely

$$d_{\min} = \min\{\|\vec{B}_i - \vec{B}_j\|\}_{i \in \Omega_{\text{Training}}, j \in \Omega_{\text{Test}}}, \quad (6)$$

where Ω_{Training} (Ω_{Test}) is the ensemble of structures contained in the training (test) set. Since a given compound could have several atoms of the same species, the maximum of all d_{\min} is assigned to that compound. Note that for each test ternary system, three values of this distance metric are obtained, one for each species.

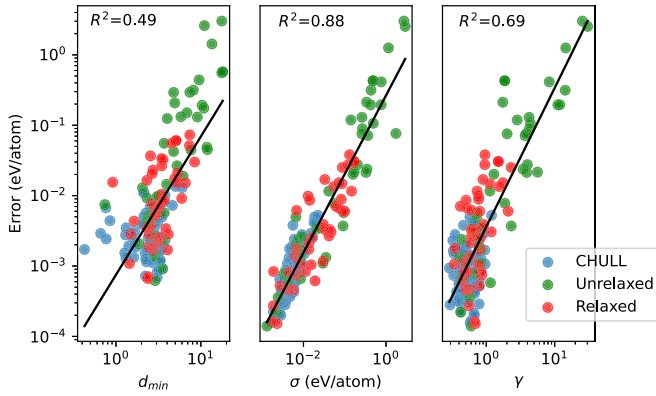


FIG. 6. Plot of the SNAP error against our defined three different metrics for the three ternary data sets in the Cu-Ag-Au system: the compounds residing on the convex hull diagram (CHULL), the unrelaxed new prototypes (Unrelaxed), and the new prototypes after relaxation (Relaxed). The metric on the x -axis from left to right is the distance from the training, d_{\min} ; the standard deviation, σ ; and the extrapolation grade, γ . These are shown for Au for d_{\min} and for Ag for γ (note that σ is species-independent). The black lines show the linear fits on the data points, and the associated R^2 is given.

The second error metric tested is the extrapolation grade, γ , as introduced by Podryabinkin *et al.* [44]. The approach followed here is closest to their Query Strategy 3, as it is defined specieswise. From the bispectrum components of all the k configurations in the training set, a rectangular $k \times m$ matrix, \mathbf{B} , is formed, where m is the length of the feature vector. The MAXVOL [68] algorithm is then applied to find the $m \times m$ submatrix \mathbf{A} with maximal determinant. This corresponds to the selection of the active set. Then, for a new structure outside of the training set, the extrapolation coefficient γ is calculated for each bispectrum component of that system, and the maximum value is selected. As for the distance metric, this is done specieswise.

The final error metric chosen is the standard deviation, σ , of the SNAPs prediction. As described above, an ensemble of SNAP models is trained over different batches of the training set. The mean value of the predictions of each model is taken as the prediction of the ensemble. This approach also allows one to obtain the standard deviation of the predictions, which is used here as an error metric.

These metrics are assessed and compared by evaluating their correlation with the SNAP absolute error. A variety of three test sets in the Cu-Ag-Au ternary system is used for this exercise, namely the ternaries presented in Sec. III B (compounds on the convex hull), as well as the unrelaxed and relaxed prototype ternary compounds introduced in Sec. III C. In Fig. 6 we present the SNAP mean absolute error against the error metric for each phase contained within the sets. These are plotted on a log scale, since both the error metrics and errors span several orders of magnitude. A linear fitting is performed on the log of all data points for each plot, and the associated values of R^2 are also given. For the d_{\min} and the γ metrics, the results are shown for the bispectrum of a specific specie, Au and Ag, respectively. As discussed previously, the structures in the AFLOWlib CHULL and in the relaxed data sets, that are close to equilibrium, have lower errors in gen-

eral, with error ranges of 13 and 72 meV/atom, respectively, whereas it is ~ 3 eV/atom for the structures of the unrelaxed data set. This implies that as a general feature of all plots, there is a group of points at lower errors that are within a small range of the error metrics. Therefore, there is modest correlation between metric and error. This is to be expected, since this range is of the same order of magnitude as the cross-validation error of the SNAP model. The metrics then become more relevant at larger errors, where the structures are expected to be outside of the training range, as for those in the unrelaxed data set.

Going into more detail, we find that the d_{\min} minimum distance metric is not much larger for the structures with large errors, resulting in a low R^2 . This indicates that the Euclidean distance between test and training vectors is not adequate for measuring the extent of extrapolation in our work. In contrast, the extrapolation grade, γ , and especially the standard deviation of the model prediction, σ , correlate far better with the error, with R^2 values of 0.69 and 0.88, respectively. Both of these metrics are significantly larger for the structures of the unrelaxed data sets compared to those in the AFLOWlib CHULL and relaxed data sets. It is interesting to note that the best performing metric is the standard deviation metric, which is not a specie-defined metric, so it can only be calculated for an entire structure. The minimum distance metric and extrapolation grade can be made global for one structure by taking the maximum metric value of the three species present. This increases the R^2 values to 0.67 and 0.79, respectively. The standard deviation is also the only metric that makes use of the performance of the trained model, since it relies on the SNAP coefficients. The other two metrics only make use of the feature vectors of the training set.

IV. CONCLUSIONS

In conclusion, we have examined the extrapolation ability of SNAP models trained on data extracted from binary phase diagrams to predict the ground-state energy and hence the thermodynamic stability of unseen ternary prototypes. Such analysis has been performed over a noble-metal ternary system, Cu-Ag-Au, and over two mixed systems, namely Ti-Mo-Pt and Cd-Hf-Rh. Furthermore, we have explored different error metrics in the search for a criterion to detect structures that are far from the training set. In this way, one could easily identify prototypes with which SNAP would struggle.

We have been able to establish that an ensemble of SNAP models, trained on binary-phase data, is able to predict the total energy of ternary compounds if those are close to their energy minimum, namely if their geometries are relaxed. In contrast, full relaxation, starting from an out-of-equilibrium prototype, is typically not possible, since along the relaxation path one is likely to encounter structures falling outside the range of applicability of the SNAP. To identify those structures for which SNAP will perform well, we have examined different error metrics. We have found that the best performing error metric is the ensemble SNAP standard deviation, σ . This is a global quantity, which is independent of the particular specie, and it is simple and efficient to calculate. With these two results at hand, we have a method that is able to evaluate

total energies of unknown ternary compounds and establish whether these energies are reliable. Note that this comes at a minimal computational cost, since no DFT calculations have to be performed to train the model.

We can then envision the use of our tool in a workflow for the construction of convex hull diagrams across the ternary space. This does not replace DFT, but it simply provides a means to establish which DFT calculations to perform and with which priority. Thus, if for a given prototype SNAP returns a high energy with a high confidence, further DFT calculations will not be necessary. Indeed, the fact that full relaxation is not possible precludes the use of our SNAP in global structure optimization processes, such as those implemented with a genetic algorithm [12] or *ab initio* random structure searching [11]. In this case, the construction of suitable machine-learning potentials has to follow a completely different approach and necessitates completely different, much larger and more diverse, training data sets [69].

ACKNOWLEDGMENTS

This work has been supported by the Irish Research Council Advanced Laureate Award (IRCLA/2019/127), and by the Irish Research Council postgraduate program (M.C.). We acknowledge the DJEI/DES/SFI/HEA Irish Centre for High-End Computing (ICHEC) and Trinity Centre for High Performance Computing (TCHPC) for the provision of computational resources.

This section is written according to the CRediT system. M.M. and H.R. contributed equally to this work. M.M. and H.R. contributed to conceptualization, methodology, software, data curation, formal analysis, investigation, validation, visualization, writing the original draft, as well as reviewing and editing the manuscript. M.C. contributed to investigation, formal analysis, software, as well as reviewing and editing the manuscript. S.S. contributed to conceptualization, funding acquisition, project administration, resources, supervision, as well as reviewing and editing the manuscript.

APPENDIX

1. Learning curves

In this subsection, the learning curves for the models trained on the three binary systems of Cu-Ag-Au are presented. These are used to establish the amount of data needed to train the SNAP model. As can be seen from Fig. 7, a 80% (training)/20% (cross-validation) split appears to be ideal.

TABLE II. Summary of the optimal hyperparameters for the SNAP models individually trained over the binary systems. Here w_X and w_Y refer to the weights of the X and Y species of the X - Y binary system.

Hyperparameter	Ag-Au	Cu-Ag	Cu-Au
J_{\max}	3	3	3
R_c	4.144006	4.358127	4.452703
w_X	0.526867	0.174029	0.193797
w_Y	0.701251	0.225576	0.341620

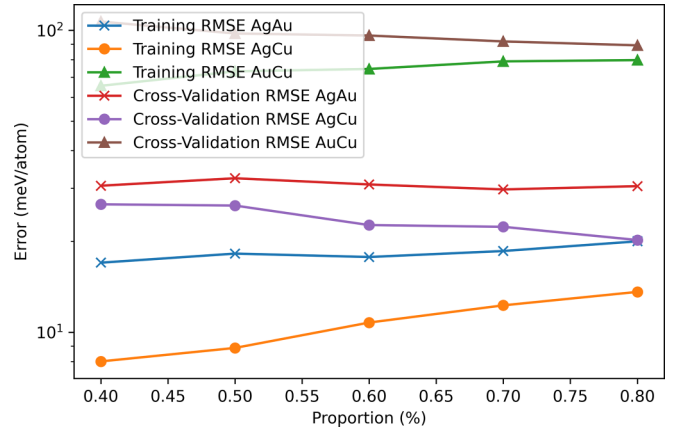


FIG. 7. Learning curves of the models trained on the binary data. As one can notice, the optimal value for the data splitting is at 80% for the training sets and 20% for the test sets.

Note that similar curves are obtained for the other two ternary systems investigated (not presented here).

2. Hyperparameter optimization

Hyperparameter optimization is performed with the use of the OPTUNA [58] package. Tree-structured Parzen Estimator (TPE) [55–57], a Bayesian algorithm, is selected to explore the hyperparameters space. The hyperparameters R_c and w_{z_i} are optimized simultaneously. For the radius cutoff, an interval of 3–10 Å is searched, corresponding to a distance greater than that of the first neighbors for the systems of interest. For the elemental weights, positive values between 0 and 5 are screened. An ensemble strategy is followed during the hyperparameter optimization in order to reduce any bias resulting from the small number of data points available. For each optimization step, 10 different SNAP models are trained, each obtained from a different 80% (training)/20% (cross-validation) split of the full data set. The RMSE of the ensemble is defined as the average RMSE on the cross-validation set for each model trained. The hyperparameter tuning is carried out 50 times for 400 steps for each system explored, and the hyperparameters that minimize the RMSE are selected. The values given in Table II are found to minimize the RMSE for the single binary SNAP models. Note that from the construction of learning curves, the optimal value for J_{\max} is found to be 3 for all single binary systems.

The same procedure is followed to minimize the RMSE for the SNAP models trained on all three binaries. In this case, the optimal value for J_{\max} is 4. The results are given in Table III.

TABLE III. Summary of the hyperparameters used for the Cu-Ag-Au SNAP models.

Hyperparameter	Cu-Ag-Au
J_{\max}	4
R_c	4.647073
w_{Ag}	0.305086
w_{Au}	0.418890
w_{Cu}	0.245647

- [1] G. Kresse and J. Furthmüller, Efficiency of *ab-initio* total energy calculations for metals and semiconductors using a plane-wave basis set, *Comput. Mater. Sci.* **6**, 15 (1996).
- [2] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos *et al.*, Quantum espresso: a modular and open-source software project for quantum simulations of materials, *J. Phys.: Condens. Matter* **21**, 395502 (2009).
- [3] X. Gonze, B. Amadon, G. Antonius, F. Arnardi, L. Baguet, J.-M. Beuken, J. Bieder, F. Bottin, J. Bouchet, E. Bousquet, N. Brouwer, F. Bruneval, G. Brunin, T. Cavignac, J.-B. Charraud, W. Chen, M. Côté, S. Cottenier, J. Denier, G. Geneste *et al.*, The ABINIT project: Impact, environment and recent developments, *Comput. Phys. Commun.* **248**, 107042 (2020).
- [4] G. Hautier, A. Jain, H. Chen, C. Moore, S. P. Ong, and G. Ceder, Novel mixed polyanions lithium-ion battery cathode materials predicted by high-throughput *ab initio* computations, *J. Mater. Chem.* **21**, 17147 (2011).
- [5] Y. Lederer, C. Toher, K. S. Vecchio, and S. Curtarolo, The search for high entropy alloys: A high-throughput *ab-initio* approach, *Acta Mater.* **159**, 364 (2018).
- [6] S. Sanvito, C. Oses, J. Xue, A. Tiwari, M. Zic, T. Archer, P. Tozman, M. Venkatesan, M. Coey, and S. Curtarolo, Accelerated discovery of new magnets in the heusler alloy family, *Sci. Adv.* **3**, e1602241 (2017).
- [7] S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, The high-throughput highway to computational materials design, *Nat. Mater.* **12**, 191 (2013).
- [8] A. R. Oganov, G. Saleh, and A. G. Kvashnin, *Computational Materials Discovery* (Royal Society of Chemistry, Croydon, UK, 2018).
- [9] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, and O. Levy, AFLOWLIB.ORG: A distributed materials properties repository from high-throughput *ab initio* calculations, *Comput. Mater. Sci.* **58**, 227 (2012).
- [10] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, Commentary: The materials project: A materials genome approach to accelerating materials innovation, *APL Mater.* **1**, 011002 (2013).
- [11] C. J. Pickard and R. Needs, *Ab initio* random structure searching, *J. Phys.: Condens. Matter* **23**, 053201 (2011).
- [12] A. O. Lyakhov, A. R. Oganov, H. T. Stokes, and Q. Zhu, New developments in evolutionary structure prediction algorithm USPEX, *Comput. Phys. Commun.* **184**, 1172 (2013).
- [13] M. K. Bisbo and B. Hammer, Global optimization of atomistic structure enhanced by machine learning, *Phys. Rev. B* **105**, 245404 (2022).
- [14] F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, Bypassing the Kohn-Sham equations with machine learning, *Nat. Commun.* **8**, 872 (2017).
- [15] A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf, and M. Ceriotti, Transferable machine-learning model of the electron density, *ACS Cent. Sci.* **5**, 57 (2019).
- [16] B. Focassio, M. Domina, U. Patil, A. Fazzio, and S. Sanvito, Linear Jacobi-Legendre expansion of the charge density for machine learning-accelerated electronic structure calculations, *npj Comput. Mater.* **9**, 87 (2023).
- [17] Y. Yang, O. A. Jiménez-Negrón, and J. R. Kitchin, Machine-learning accelerated geometry optimization in molecular simulation, *J. Chem. Phys.* **154**, 234704 (2021).
- [18] R. Jinnouchi, K. Miwa, F. Karsai, G. Kresse, and R. Asahi, On-the-fly active learning of interatomic potentials for large-scale atomistic simulations, *J. Phys. Chem. Lett.* **11**, 6946 (2020).
- [19] Y. Hinuma, T. Hatakeyama, Y. Kumagai, L. A. Burton, H. Sato, Y. Muraba, S. Iimura, H. Hiramatsu, I. Tanaka, H. Hosono, and F. Oba, Discovery of earth-abundant nitride semiconductors by computational screening and high-pressure synthesis, *Nat. Commun.* **7**, 11962 (2016).
- [20] X. Zhang and M. H. Sluiter, Cluster expansions for thermodynamics and kinetics of multicomponent alloys, *J. Phase Eq. Diffus.* **37**, 44 (2016).
- [21] Q. Wu, B. He, T. Song, J. Gao, and S. Shi, Cluster expansion method and its application in computational materials science, *Comput. Mater. Sci.* **125**, 243 (2016).
- [22] G. Hautier, C. C. Fischer, A. Jain, T. Mueller, and G. Ceder, Finding nature's missing ternary oxide compounds using machine learning and density functional theory, *Chem. Mater.* **22**, 3762 (2010).
- [23] A. O. Oliynyk, L. A. Adutwum, J. J. Harynyk, and A. Mar, Classifying crystal structures of binary compounds *AB* through cluster resolution feature selection and support vector machine analysis, *Chem. Mater.* **28**, 6672 (2016).
- [24] K. Kim, L. Ward, J. He, A. Krishna, A. Agrawal, and C. Wolverton, Machine-learning-accelerated high-throughput materials screening: Discovery of novel quaternary heusler compounds, *Phys. Rev. Mater.* **2**, 123801 (2018).
- [25] J. Schmidt, L. Chen, S. Botti, and M. A. Marques, Predicting the stability of ternary intermetallics with density functional theory and machine learning, *J. Chem. Phys.* **148**, 241728 (2018).
- [26] J. Behler and M. Parrinello, Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces, *Phys. Rev. Lett.* **98**, 146401 (2007).
- [27] A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials, *J. Comput. Phys.* **285**, 316 (2015).
- [28] M. A. Wood and A. P. Thompson, Extending the accuracy of the SNAP interatomic potential form, *J. Chem. Phys.* **148**, 241721 (2018).
- [29] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons, *Phys. Rev. Lett.* **104**, 136403 (2010).
- [30] A. V. Shapeev, Moment tensor potentials: A class of systematically improvable interatomic potentials, *Multiscale Model. Simul.* **14**, 1153 (2016).
- [31] R. Drautz, Atomic cluster expansion for accurate and transferable interatomic potentials, *Phys. Rev. B* **99**, 014104 (2019).
- [32] Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A. V. Shapeev, A. P. Thompson, M. A. Wood, and S. P. Ong, Performance and cost assessment of machine learning interatomic potentials, *J. Phys. Chem. A* **124**, 731 (2020).
- [33] X.-G. Li, C. Chen, H. Zheng, Y. Zuo, and S. P. Ong, Complex strengthening mechanisms in the NbMoTaW multi-principal element alloy, *npj Comput. Mater.* **6**, 1 (2020).

- [34] N. Artrith and J. Behler, High-dimensional neural network potentials for metal surfaces: A prototype study for copper, *Phys. Rev. B* **85**, 045439 (2012).
- [35] V. L. Deringer and G. Csányi, Machine learning based interatomic potential for amorphous carbon, *Phys. Rev. B* **95**, 094203 (2017).
- [36] V. L. Deringer, N. Bernstein, G. Csányi, C. Ben Mahmoud, M. Ceriotti, M. Wilson, D. A. Drabold, and S. R. Elliott, Origins of structural and electronic transitions in disordered silicon, *Nature (London)* **589**, 59 (2021).
- [37] X.-G. Li, C. Hu, C. Chen, Z. Deng, J. Luo, and S. P. Ong, Quantum-accurate spectral neighbor analysis potential models for Ni-Mo binary alloys and fcc metals, *Phys. Rev. B* **98**, 094104 (2018).
- [38] K. Shimizu, E. F. Arguelles, W. Li, Y. Ando, E. Minamitani, and S. Watanabe, Phase stability of Au-Li binary systems studied using neural network potential, *Phys. Rev. B* **103**, 094112 (2021).
- [39] W. Ibarra-Hernández, S. Hajinazar, G. Avendaño-Franco, A. Bautista-Hernández, A. N. Kolmogorov, and A. H. Romero, Structural search for stable Mg-Ca alloys accelerated with a neural network interatomic model, *Phys. Chem. Chem. Phys.* **20**, 27545 (2018).
- [40] S. Hajinazar, J. Shao, and A. N. Kolmogorov, Stratified construction of neural network based interatomic models for multicomponent materials, *Phys. Rev. B* **95**, 014114 (2017).
- [41] P. Rowe, V. L. Deringer, P. Gasparotto, G. Csányi, and A. Michaelides, An accurate and transferable machine learning potential for carbon, *J. Chem. Phys.* **153**, 034702 (2020).
- [42] A. P. Bartók, J. Kermode, N. Bernstein, and G. Csányi, Machine Learning a General-Purpose Interatomic Potential for Silicon, *Phys. Rev. X* **8**, 041048 (2018).
- [43] K. Gubaev, E. V. Podryabinkin, G. L. Hart, and A. V. Shapeev, Accelerating high-throughput searches for new alloys with active learning of interatomic potentials, *Comput. Mater. Sci.* **156**, 148 (2019).
- [44] E. V. Podryabinkin and A. V. Shapeev, Active learning of linearly parametrized interatomic potentials, *Comput. Mater. Sci.* **140**, 171 (2017).
- [45] N. Artrith, A. Urban, and G. Ceder, Constructing first-principles phase diagrams of amorphous Li_xSi using machine-learning-assisted sampling with an evolutionary algorithm, *J. Chem. Phys.* **148**, 241711 (2018).
- [46] C. Toher, C. Oses, D. Hicks, and S. Curtarolo, Unavoidable disorder and entropy in multi-component systems, *npj Comput. Mater.* **5**, 69 (2019).
- [47] C. Oses, E. Gossett, D. Hicks, F. Rose, M. J. Mehl, E. Perim, I. Takeuchi, S. Sanvito, M. Scheffler, Y. Lederer, O. Levy, C. Toher, and S. Curtarolo, AFLOW-CHULL: Cloud-oriented platform for autonomous phase stability analysis, *J. Chem. Inf. Model.* **58**, 2477 (2018).
- [48] R. H. Taylor, F. Rose, C. Toher, O. Levy, K. Yang, M. Buongiorno Nardelli, and S. Curtarolo, A RESTful API for exchanging materials data in the AFLOWLIB.org consortium, *Comput. Mater. Sci.* **93**, 178 (2014).
- [49] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [50] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple [Phys. Rev. Lett. **77**, 3865 (1996)], *Phys. Rev. Lett.* **78**, 1396(E) (1997).
- [51] C. E. Calderon, J. J. Plata, C. Toher, C. Oses, O. Levy, M. Fornari, A. Natan, M. J. Mehl, G. Hart, M. Buongiorno Nardelli, and S. Curtarolo, The AFLOW standard for high-throughput materials science calculations, *Comput. Mater. Sci.* **108**, 233 (2015).
- [52] A. P. Bartók, R. Kondor, and G. Csányi, On representing chemical environments, *Phys. Rev. B* **87**, 184115 (2013).
- [53] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, *J. Comput. Phys.* **117**, 1 (1995).
- [54] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine learning in Python, *J. Machine Learning Res.* **12**, 2825 (2011).
- [55] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, Algorithms for hyper-parameter optimization, in *Advances in Neural Information Processing Systems*, edited by J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger (Curran Associates, Red Hook, NY, 2011), Vol. 24.
- [56] J. Bergstra, D. Yamins, and D. Cox, Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures, in *Proceedings of the 30th International Conference on Machine Learning*, Proceedings of Machine Learning Research, edited by S. Dasgupta and D. McAllester (PMLR, Atlanta, 2013), Vol. 28, pp. 115–123.
- [57] C. A. C. Coello, Y. Ozaki, Y. Tanigaki, S. Watanabe, and M. Onishi, Multiobjective tree-structured parzen estimator for computationally expensive optimization problems, in *Proceedings of the 2020 Genetic and Evolutionary Computation Conference (Association for Computing Machinery, New York, 2020)*, p. 533.
- [58] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, Optuna: A next-generation hyperparameter optimization framework, in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Association for Computing Machinery, New York, 2019)*.
- [59] D. Hicks, M. J. Mehl, M. Esters, C. Oses, O. Levy, G. L. Hart, C. Toher, and S. Curtarolo, The aflow library of crystallographic prototypes: Part 3, *Comput. Mater. Sci.* **199**, 110450 (2021).
- [60] M. J. Mehl, D. Hicks, C. Toher, O. Levy, R. M. Hanson, G. Hart, and S. Curtarolo, The aflow library of crystallographic prototypes: Part 1, *Comput. Mater. Sci.* **136**, S1 (2017).
- [61] D. Hicks, M. J. Mehl, E. Gossett, C. Toher, O. Levy, R. M. Hanson, G. Hart, and S. Curtarolo, The aflow library of crystallographic prototypes: Part 2, *Comput. Mater. Sci.* **161**, S1 (2019).
- [62] J. S. Smith, B. Nebgen, N. Mathew, J. Chen, N. Lubbers, L. Burakovsky, S. Tretiak, H. A. Nam, T. Germann, S. Fensin *et al.*, Automated discovery of a robust interatomic potential for aluminum, *Nat. Commun.* **12**, 1257 (2021).
- [63] C. W. Park and C. Wolverton, Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery, *Phys. Rev. Mater.* **4**, 063801 (2020).
- [64] R. Wang, W. Xia, T. J. Slade, X. Fan, H. Dong, K.-M. Ho, P. C. Canfield, and C.-Z. Wang, Machine learning guided discovery of ternary compounds involving La and immiscible Co and Pb elements, *npj Comput. Mater.* **8**, 258 (2022).
- [65] H. Sun, C. Zhang, W. Xia, L. Tang, R. Wang, G. Akopov, N. W. Hewage, K.-M. Ho, K. Kovnir, and C.-Z. Wang, Machine learning-guided discovery of ternary compounds

- containing La, P, and Group 14 elements, *Inorg. Chem.* **61**, 16699 (2022).
- [66] C. Chen, Z. Deng, R. Tran, H. Tang, I.-H. Chu, and S. P. Ong, Accurate force field for molybdenum by machine learning large materials data, *Phys. Rev. Mater.* **1**, 043603 (2017).
- [67] G. L. Hart, L. J. Nelson, and R. W. Forcade, Generating derivative structures at a fixed concentration, *Comput. Mater. Sci.* **59**, 101 (2012).
- [68] S. A. Goreinov, I. V. Oseledets, D. V. Savostyanov, E. E. Tyrtyshnikov, and N. L. Zamarashkin, How to find a good submatrix, in *Matrix Methods: Theory, Algorithms and Applications: Dedicated to the Memory of Gene Golub* (World Scientific, Singapore, 2010), pp. 247–256.
- [69] C. J. Pickard, Ephemeral data derived potentials for random structure search, *Phys. Rev. B* **106**, 014102 (2022).