



Accelerating high-throughput searches for new alloys with active learning of interatomic potentials

Konstantin Gubaev^a, Evgeny V. Podryabinkin^a, Gus L.W. Hart^b, Alexander V. Shapeev^{a,*}

^a Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, Nobel str. 3, Moscow 143026, Russia

^b Department of Physics and Astronomy, Brigham Young University, Provo, UT 84602, USA

ARTICLE INFO

Keywords:

Alloy phase prediction
Machine learning
Active learning
Interatomic potentials
Cluster expansion
Moment Tensor Potentials

ABSTRACT

We propose an approach to materials prediction that uses a machine-learning interatomic potential to approximate quantum-mechanical energies and an active learning algorithm for the automatic selection of an optimal training dataset. Our approach significantly reduces the amount of density functional theory (DFT) calculations needed, resorting to DFT only to produce the training data, while structural optimization is performed using the interatomic potentials. Our approach is not limited to one (or a small number of) lattice types (as is the case for cluster expansion, for example) and can predict structures with lattice types not present in the training dataset. We demonstrate the effectiveness of our algorithm by predicting the convex hull for the following three systems: Cu-Pd, Co-Nb-V, and Al-Ni-Ti. Our method is three to four orders of magnitude faster than conventional high-throughput DFT calculations and explores a wider range of materials space. In all three systems, we found unreported stable structures compared to the AFLOW database. Because our method is much cheaper and explores much more of materials space than high-throughput methods or cluster expansion, and because our interatomic potentials have a systematically improvable accuracy compared to empirical potentials such as embedded atom model, it will have a significant impact in the discovery of new alloy phases, particularly those with three or more components.

1. Introduction

Advances in computer power, improvements in first-principles methods, and the generation of large materials databases like AFLOWLIB [1], OQMD [2], CMR [3], NOMAD [4], and Materials Project [5] have enabled modern data analysis tools to be applied in the field of materials discovery [6–8]. There have been growing efforts in computational search for materials with superior properties, including metallic alloys [9–11], semiconductor materials [12], and magnetic materials [13]. In this work we consider the problem of predicting stable phases in multicomponent alloys. A typical prediction algorithm consists of sampling structures across the configurational space and evaluating their energies. The sampling is done by searching through structures that are either selected from some carefully assembled pool of possible structures, often called crystal prototypes [14], or are generated by some sampling algorithm, see, e.g., Refs. [15,16].

The evaluation of the energy of the structures in the pool is often done with density functional theory (DFT). Even despite its favorable accuracy/efficiency trade-off as compared to other quantum-mechanical algorithms, the DFT calculations remain the bottleneck in

materials prediction workflows, making an exhaustive search impractical. Machine learning (ML) for materials prediction has the potential to dramatically reduce the number of quantum-mechanical calculations performed and thus reduce the computational expense of predicting new materials via computation. The reduction of the computational time is achieved by constructing a surrogate model that “interpolates” the quantum-mechanical training data and makes subsequent energy evaluations much faster (by orders of magnitude). This is similar in spirit to the cluster expansion method which has been broadly used in different materials discovery applications [17–19,12]. Cluster expansion is quite successful when the stable structures are derivatives of a particular structure (fcc, bcc, etc.) but is not useful when this is not the case. Its accuracy also converges slowly when atomic size mismatch is not negligible [20]. Additionally, more classical machine-learning algorithms such as decision trees [21], support vector machines [22], and other ML algorithms [23,24] have been tried. Surrogate models such as the cluster expansion and standard machine learning approaches do not have the broad applicability and exceptional accuracy of the moment tensor potentials-based [25] approach we demonstrate here.

* Corresponding author.

E-mail address: alexander@shapeev.com (A.V. Shapeev).

The two important features of our approach are a completely general form for the interatomic potentials and an active learning algorithm for generating and refining the training set. In this work we extend the approach [26] for predicting the structure of a single-component material. In our approach, a ML model reproduces DFT for off-equilibrium structures that are not restricted to any lattice. Furthermore, the model learns the DFT interaction actively (on-the-fly) while equilibrating the candidate structures, completely automating the construction of the training set. Thus, structural optimization of the training structures can be performed via the interatomic potentials, rather than via DFT, further accelerating the construction of the training set.

Our method is based on moment tensor potentials (MTPs [25]) and the active learning algorithm [27]. Namely, we solve the following problem: given a set of elements, find the most stable structures (in the sense of lying on the convex hull of formation enthalpies) consisting of these elements, each characterized by composition, unit cell geometry and atomic positions within the unit cell. In this work we extend the interatomic potential [25] and active learning algorithm [27] to handle atomistic configurations with multiple types of atoms, similarly to the approach used in cheminformatics [28]. The differences between the algorithms from Ref. [28] and this work include that (1) we need derivatives of the energy, whereas in Ref. [28] we needed only the energy (or other predicted properties); and (2) that in Ref. [28] we were concerned with the selection from a finite set of predefined structures, whereas in this work we need to solve the problem predicting the energy with a fitted potential and assembling the training set used for the fitting at the same time (in other words, exploring the potential energy landscape and constructing the training set at the same time).

The idea of applying neural networks, as a broad class of machine-learning algorithms, to constructing interatomic potentials was pioneered in Ref. [29]. Application of Gaussian process regression, another class of machine-learning algorithms, was then proposed in Ref. [30]. The promising results obtained in these works have motivated many research groups to pursue this research direction [31–50,25,51]. However, the application of such algorithms to the problem of materials prediction has proven difficult since following such a methodology requires one to collect all the representative structures in the training set which is as hard as predicting materials structure itself. In our view, it is the active learning [27,26,33,52,53] that paves the way for machine-learning interatomic potentials to accelerate computational materials discovery.

This paper is organized as follows: in Section 2 we introduce the algorithms we use, including the moment tensor potentials (Section 2.1), active learning (Section 2.2), and the “relaxation while learning on-the-fly” algorithm (Section 2.3). In Section 3 we test the proposed algorithm on predicting the stable structures of the Cu-Pd, Co-Nb-V, and Al-Ni-Ti systems and discuss the performance of our algorithm. In particular, we compare our results to those obtained by high-throughput DFT calculations as reported in the AFLOW database [10,1]. In all three systems we have discovered new structures below the reported convex hull of ground-state structures. Finally, in Section 4 we make concluding remarks.

2. Methodology

2.1. Machine-learning potentials

We use the moment tensor potentials (MTPs) for approximating a quantum-mechanical energy. The potential is parametrized by a set of parameters θ that are found from minimizing the loss functional expressing that the predicted energy E is close to the reference quantum-mechanical energy E^{qm} :

$$L(\theta) = \sum_j (E(\theta, x^{(j)}) - E^{\text{qm}}(x^{(j)}))^2 \rightarrow \min, \quad (1)$$

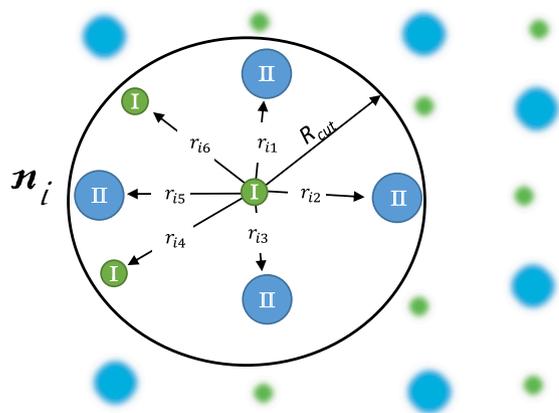


Fig. 1. Partitioning scheme: energy E is composed from contributions V_i of individual neighborhoods n_i . The neighborhood n_i of the i th atom is described by the relative position of neighboring atoms, r_{ij} , and the types of atoms z_j (I or II in this illustration).

where $x^{(j)}$ are the configurations in the training set and $E^{\text{qm}}(x^{(j)})$ are their reference energies.

Our model is local, which we enforce by partitioning the energy, E , into the contributions, V , of individual atomic neighborhoods. To define a neighborhood of the i th atom, we let r_{ij} be the position of j th atom relative to the i th atom (thus, r_{ij} is a vectorial quantity) and z_j be the type of the j th atom. Then n_i is the collection of r_{ij} and z_j , and $E(x) = \sum_i V(n_i)$. The locality of the model is expressed by the requirement that V does not depend on atoms that are farther from i than some cutoff distance R_{cut} , which is usually around 5 Å. An illustration of an atomic neighborhood is sketched in Fig. 1. Mathematically, each atom in the neighborhood introduces four degrees of freedom, on which n_i depends: these are three coordinates in Euclidean space, and a discrete variable representing the chemical type. Typically, neighborhoods include a few dozen atoms, which means that the function $V(n_i)$ depends on the order of hundred scalar variables. To somewhat reduce the dimensionality, we embed all physical symmetries into $V(n)$ so they will not have to be learned by the model. These symmetries arise from the isotropy and translational symmetry of the physical space, and from the fact that the interaction between atoms does not depend on their ordering (see Fig. 2).

As in the work [25] devoted to the single-component moment tensor potentials, $V(n)$ is linearly expanded through a set of *basis functions* B_α :

$$V(n) = \sum_\alpha \xi_\alpha B_\alpha(n). \quad (2)$$

The basis functions, in turn, depend on the set of *moment tensor descriptors*

$$M_{\mu,\nu}(n_i) = \sum_j f_\mu(|r_{ij}|, z_i, z_j) \underbrace{r_{ij} \otimes \dots \otimes r_{ij}}_{\nu \text{ times}}, \quad (3)$$

where the index j enumerates all the atoms in the neighborhood n_i . The functions $f_\mu(|r_{ij}|, z_i, z_j)$ depend only on the interatomic distances and atomic types, therefore we call them *radial functions*. The terms $r_{ij} \otimes \dots \otimes r_{ij}$ contain the angular information about the neighborhood n_i and are tensors of rank ν . We next explain how to construct the basis functions from the moment tensor descriptors, following which we present a simple illustration of the structure of the descriptors and basis functions.

The functions $B_\alpha(n_i)$ enumerate all possible contractions of any number of $M_{\mu,\nu}(n_i)$ yielding a scalar. Note that $M_{\mu,\nu}(n_i)$ are invariant, by construction, with respect to translations of the system and permutations of equivalent atoms. Their scalar contractions are invariant with respect to rotations of the neighborhood. Thus the resulting function

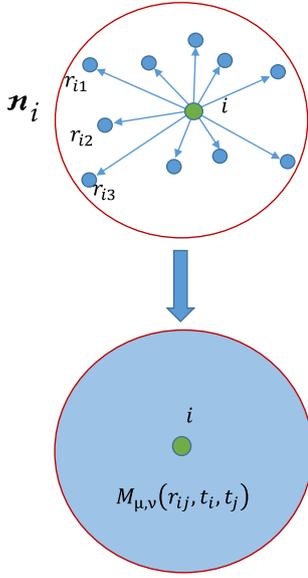


Fig. 2. For the purpose of fitting the interatomic interaction energy E , the neighborhood n_i is described by the moment tensors $M_{\mu,\nu}$ exhibiting all the physical symmetries that E has.

$V(n)$ also has these symmetries. Although all the descriptors $M_{\mu,\nu}(n_i)$ are composed of two-body terms depending only on r_{ij} , their contractions $B(n_i)$ can depend on many-body terms of higher order.

For the purpose of illustration, assume, for the moment, that the vectors r_{ij} are two-dimensional and that we can express them in polar coordinates (ρ, θ) centered at the i th atom. Let us look closer at the term $r_{ij} \otimes \dots \otimes r_{ij} = :r^{\otimes \nu}$. $r_{ij}^{\otimes 0}$ is a scalar with no angular information, while $r_{ij}^{\otimes 1} = r_{ij} = |r_{ij}|(\cos\theta_{ij}, \sin\theta_{ij})$ does contain angular information. A vectorial contraction is simply a dot product: $r_{ij} \cdot r_{ik} = |r_{ij}||r_{ik}|\cos(\theta_{ij}-\theta_{ik})$ —in this way we introduce angular terms into the potential. An arbitrary function of angle can be expanded into a sum of powers of cosine. Such higher-order terms are contributed to the potential by higher-rank tensors, e.g.,

$$r_{ij}^{\otimes 2} = r_{ij} r_{ij}^T = |r_{ij}|^2 \begin{pmatrix} \cos^2 \theta_{ij} & \sin \theta_{ij} \cos \theta_{ij} \\ \sin \theta_{ij} \cos \theta_{ij} & \sin^2 \theta_{ij} \end{pmatrix}.$$

The contractions of two matrices are given by the Frobenius product

$$r_{ij}^{\otimes 2} : r_{ik}^{\otimes 2} = |r_{ij}|^2 |r_{ik}|^2 \cos^2(\theta_{ij}-\theta_{ik}).$$

A more complicated expression can be constructed with a matrix and two vectors:

$$(r_{ij}^{\otimes 2} r_{ik}) \cdot r_{i\ell} = |r_{ij}|^2 |r_{ik}| |r_{i\ell}| \cos(\theta_{ij}-\theta_{ik}) \cos(\theta_{ij}-\theta_{i\ell}).$$

Terms of this form are rotationally invariant. Permutation invariance is achieved by summing those terms over all atoms in the neighborhood weighted by the radial functions.

As an illustration, assume that we have two radial functions,

$$f_{\mu}(\rho, z_i, z_j) = \exp\left(-\frac{|\rho-R_{\mu}|^2}{2\sigma^2}\right),$$

$\mu = 1, 2$, where ρ has the meaning of distance to the central, i th atom. In the sum (3) they “extract” two shells of atoms, around the distances R_1 and R_2 from the i th atom, smeared over the width of σ . We did not, but could assume the dependence of these functions on the types of atoms z_i and z_j —this would discriminate the importance of these atom types to these two shells. Thus, $M_{1,0}$ and $M_{2,0}$ are the atom count in these two shells and both could serve as basis functions. $M_{i,1}$ are vectorial quantities indicating eccentricity of these shells: if $M_{i,1} = 0$ then the i th shell is symmetric (to the first order) while $M_{i,1} \neq 0$ indicate that there

are “more atoms” in the direction $M_{i,1}$ than in the opposite direction.

As vectorial quantities, $M_{i,1}$ are not valid basis functions, however, the valid ones are $M_{i,1} \cdot M_{i,1}$ indicating the magnitude of eccentricity and $M_{i,1} \cdot M_{i,2}$ indicating how these two eccentricities are aligned with respect to each other. One can make many more basis functions from these quantities, e.g., $M_{i,0}(M_{i,1} \cdot M_{i,1})$, $(M_{i,1} \cdot M_{i,1})(M_{i,1} \cdot M_{i,2})$, etc. One can then continue by analogy: $M_{i,2}$ are the second moments of inertia of these shells indicating the degree to which these shells are “squeezed” in the respective directions, forming the basis functions $M_{i,2} : M_{j,2}$, $(M_{i,2} \cdot M_{j,1}) \cdot M_{k,1}$, $(M_{i,2} \cdot M_{j,2} \cdot M_{k,1}) \cdot M_{\ell,1}$, etc. We remark that this way of enforcing symmetries in the potential is related to the ideas from Refs. [54,55].

For the purpose of choosing which (out of the infinite number of) basis functions to include in the interatomic potential, we define the degree-like measure, *level*, of $M_{\mu,\nu}$ by $\text{lev}M_{\mu,\nu} = 2\mu + \nu$ and the level of B_{α} obtained by contracting $M_{\mu_1,\nu_1}, M_{\mu_2,\nu_2}, \dots$ as $\text{lev}B_{\alpha} = (2\mu_1 + \nu_1) + (2\mu_2 + \nu_2) + \dots$. Thus, to define an MTP we choose some lev_{\max} and include in (2) each B_{α} with $\text{lev}B_{\alpha} \leq \text{lev}_{\max}$. Thus, by increasing lev_{\max} we increase the number of parameters in the potential, including the contributions of three-body, four-body, etc., terms. In this sense, $V(n)$ has a systematically improvable functional form.

The radial functions $f_{\mu}(|r_{ij}|, z_i, z_j)$ (cf. (3)) have the form:

$$f_{\mu}(\rho, z_i, z_j) = \sum_k c_{\mu,z_i,z_j}^{(k)} Q^{(k)}(\rho), \quad \text{where} \quad (4)$$

$$Q^{(k)}(\rho) := T_k(\rho)(R_{\text{cut}}-\rho)^2.$$

Here $T_k(\rho)$ are the Chebyshev polynomials on the interval $[R_{\min}, R_{\text{cut}}]$. The term $(R_{\text{cut}}-\rho)^2$ was introduced to ensure smoothness with respect to the atoms leaving and entering the cut-off sphere. Taking into account that in real systems atoms never stay too close to each other, we can always choose some reasonable value for R_{\min} .

The difference from the single-component MTPs [25] is that now the functions $f_{\mu}(\rho, z_i, z_j)$ depend on the types of the central and the neighboring atoms. As follows from (4) a number of parameters $c_{\mu,z_i,z_j}^{(k)}$ exist for each pair of species and each μ . Note that the number of these parameters is proportional to n^2 , where n is the number of species, while number of parameters ξ_i from (2) does not depend on the number of species. Thus, the total number of model parameters $\theta = \left\{ \{\xi_i\}, \left\{ c_{\mu,z_i,z_j}^{(k)} \right\} \right\}$ to be found in the minimization procedure (1) grows less than quadratically with the number of species, despite accounting for many-body interactions in $V(n)$. It was proven [25] that the descriptors of the form (3) provide a complete description of an atomic neighborhood, in the sense that any function of atomic neighborhood with the same symmetries as $V(n)$ can be approximated as a polynomial of these descriptors with an arbitrary accuracy. The proof [25] holds for a single-component case, but can be easily extended to a multicomponent case.

2.2. Active learning

The accuracy of a machine-learning potential depends as much on a good functional form (an efficient representation, in ML parlance), as on the quality of the training set. Roughly speaking, a good training set should include all the representative structures, so that the potential does not have to “extrapolate” while searching for the stable phases. In cluster expansion-like approaches in which the energies of relaxed structures are predicted based on a representation that uses unrelaxed structure geometries, extrapolation results in higher prediction errors [20]. However, in our approach structural relaxation is treated explicitly, and we accelerate the relaxation by using a machine-learning potential instead of DFT. Because of the added flexibility of the MTP (as compared to cluster expansion), avoiding extrapolation is even more important—it is crucial to the reliability of the algorithm—as highly

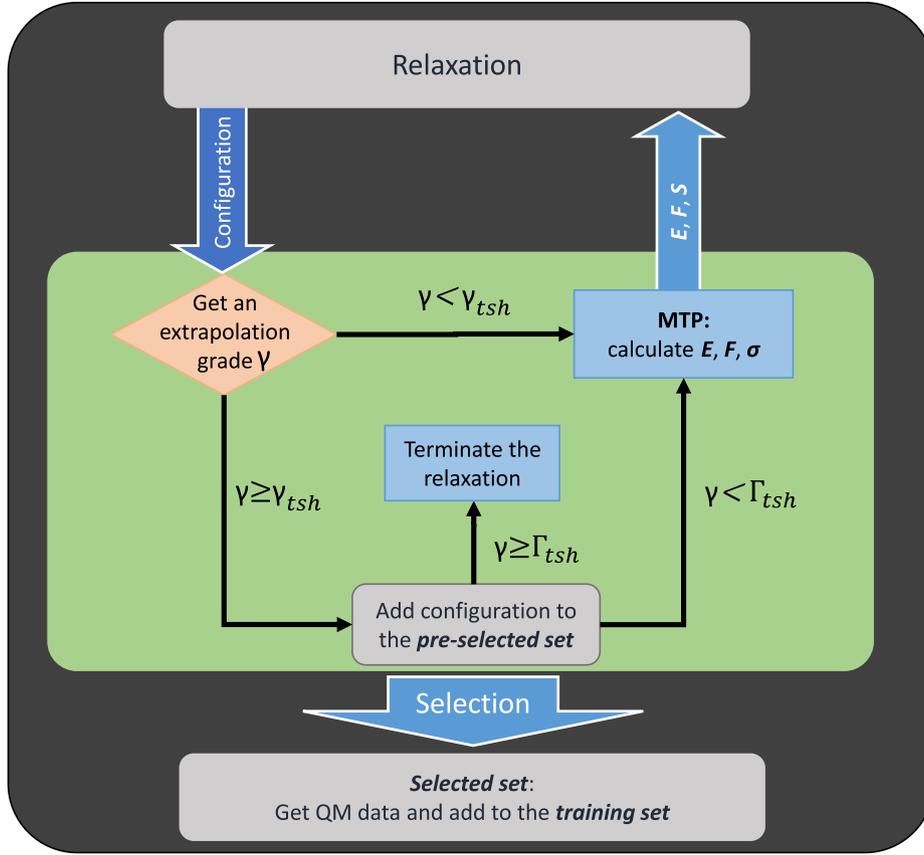


Fig. 3. Relaxation with active learning. If MTP encounters an extrapolative configuration ($\gamma \geq \gamma_{tsh}$), the configuration is added to the preselected set for further selection. In the case of significant extrapolation ($\gamma \geq \Gamma_{tsh}$) the relaxation is terminated. For configurations with $\gamma < \Gamma_{tsh}$, the MTP provides energies, forces and stresses. If no configuration with $\gamma \geq \Gamma_{tsh}$ is encountered, the relaxation stops at some equilibrium configuration.

unphysical structures can arise during the relaxation if the extrapolation is severe.

We avoid extrapolation by using active learning. The active learning algorithm developed in this work is a generalization of the algorithm proposed for linearly parametrized models [27]. It is based on a D-optimality criterion as we explain below. As the model in this paper has a nonlinear dependence on its parameters, we apply a generalization of the D-optimality criterion to the nonlinear case.

To that end, we interpret fitting as solving the following over-determined system of equations with respect to θ :

$$E(\theta, x^{(i)}) = E^{qm}(x^{(i)}).$$

We assume that we have some initial guess, $\bar{\theta}$, for the optimal values of parameters. Then if we linearize these equations around $\bar{\theta}$ then the left-hand side will be the following tall (Jacobi) matrix

$$B = \begin{pmatrix} \frac{\partial E}{\partial \theta_1}(\bar{\theta}, x^{(1)}) & \dots & \frac{\partial E}{\partial \theta_m}(\bar{\theta}, x^{(1)}) \\ \vdots & \ddots & \vdots \\ \frac{\partial E}{\partial \theta_1}(\bar{\theta}, x^{(n)}) & \dots & \frac{\partial E}{\partial \theta_m}(\bar{\theta}, x^{(n)}) \end{pmatrix},$$

where each row corresponds to a particular structure from the training set.

The generalized D-optimality criterion hence states that the best training set of m configurations corresponds to a square $m \times m$ submatrix A of the matrix B of maximal volume (i.e., with maximal value of $|\det(A)|$). In practice, it is sufficient, for a given configuration x^* to compute its *extrapolation grade* $\gamma(x^*)$ defined as the maximal factor by which $|\det(A)|$ can grow if x^* is added to the training set. We do it by using the so-called maxvol algorithm [56], according to which

$$\gamma(x^*) = \max_{1 \leq j \leq n} \left(|c_j| \right), \quad \text{where}$$

$$c = \left(\frac{\partial F}{\partial \theta_1}(\bar{\theta}, x^*) \dots \frac{\partial F}{\partial \theta_n}(\bar{\theta}, x^*) \right) A^{-1} = : b^* A^{-1}.$$

Thus, we add x^* to the training set if $\gamma(x^*) \geq \gamma_{tsh}$, where $\gamma_{tsh} \geq 1$ is a tunable threshold parameter that can control how much extrapolation is allowed.

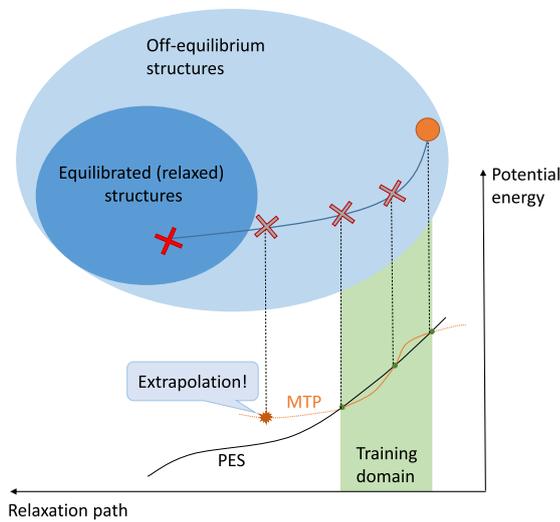
The (generalized) D-optimality criterion serves to detect structures on which the potential extrapolates. Hence, training on such structures prevents extrapolation and thus ensures that all the structures occurring during relaxation are interpolative with respect to the structures in the training set. In this work we use the active learning algorithm to construct the training set by selecting some of the configurations arising during relaxation. It also can be used to compose an optimal training set from configurations belonging to some pre-defined set [28].

2.3. Algorithm

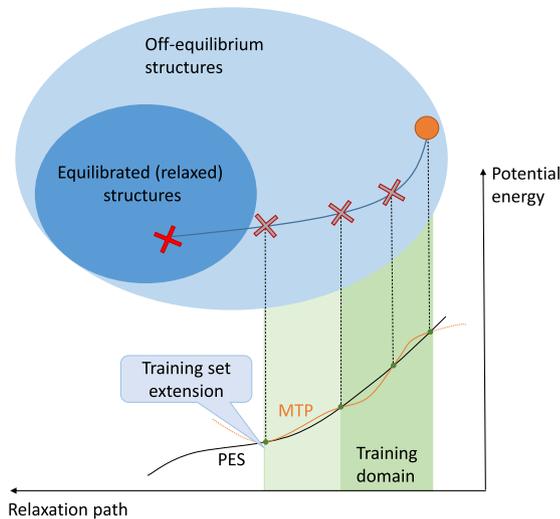
Next we describe the algorithm for constructing the convex hull.

Input The input to the algorithm is:

1. A set of candidate structures among which we expect to find the groundstate structures. (We can afford to select a much broader and more diverse set of structures as compared to the approaches based solely on DFT.)
2. A functional form of MTP, $E = E(\theta, x)$. We initialize θ randomly and let the training set be empty.
3. A quantum-mechanical model $E^{qm}(x)$. In this work we used DFT as implemented in VASP 5.4.1.



(a)



(b)

Fig. 4. If an MTP encounters some extrapolative configuration during relaxation, as shown in (a), the relaxation is terminated and restarted after retraining the MTP on this configuration, as shown in (b).

4. Two thresholds γ_{tsh} and Γ_{tsh} , such that $\Gamma_{\text{tsh}} > \gamma_{\text{tsh}} > 1$. If the extrapolation grade $\gamma(x^*)$ is greater than 1, the algorithm makes two decisions: to add x^* to the training set if $\gamma(x^*) > \gamma_{\text{tsh}}$ and to terminate the relaxation if $\gamma(x^*) > \Gamma_{\text{tsh}}$ (assuming in the latter case that we cannot make reliable predictions of energy, forces, and stresses for x^*), as explained below.

- Step 1** For each candidate structure we perform the structure relaxation with the current MTP (defined by the current values of θ). There can be two outcomes of the relaxation: (1) the relaxation completed successfully and we get an equilibrium structure as a result, (2) the relaxation was not successful because we encountered a structure on which the MTP attempted to extrapolate. More precisely, the following scenarios can emerge:
- a. The relaxation successfully converges to an equilibrium configuration and on each configuration from the relaxation trajectory (see Fig. 5) the MTP does not significantly extrapolate, i.e., the extrapolation grade of each intermediate configuration is less than Γ_{tsh} . During the relaxation there

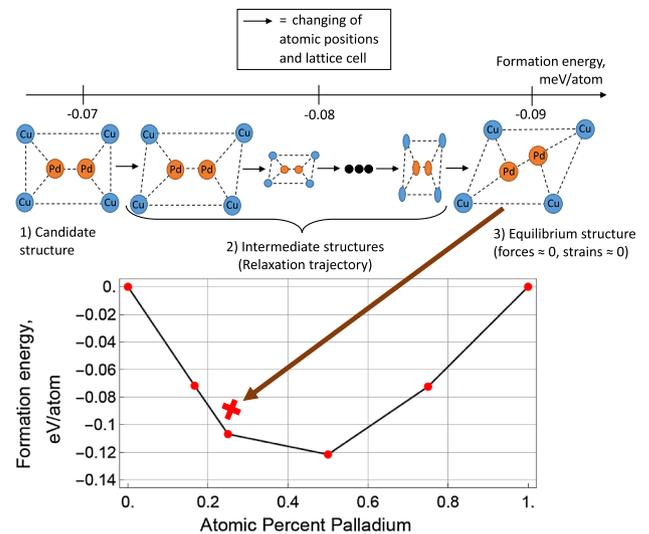


Fig. 5. Graphical illustration of the relaxation process. By “relaxation trajectory” we mean a sequential list of structures that occur during the relaxation which have similar but distinct atomic displacements and lattice parameters, and which typically have decreasing energy.

could be, however, configuration with extrapolation grade exceeding γ_{tsh} —in this case we add such a configuration to the *preselected set* (see Fig. 3 and Section 2).

b. At some step of the relaxation we obtain a configuration with the extrapolation grade exceeding Γ_{tsh} . This means that MTP cannot provide a reasonable prediction as it extrapolates significantly on this configuration and needs to be retrained with more *ab initio* data (see Fig. 4). We then terminate the relaxation. The last and all the previous configurations with the grade exceeding γ_{tsh} are added to the *preselected set*.

- Step 2** Out of the *preselected set* from the step 1b, we select a smaller number of configurations that will be added to the training set. The *preselected set* can be very large and contain hundreds of thousands configurations (note that during the first iteration of the algorithm all the relaxations will be terminated according to the scenario (b), as the training set is empty and the MTP extrapolates on every configuration.) Therefore we use our active learning algorithm to select up to few hundred most representative configurations, according to the D-optimality criterion from Section 2.2. Thus, we extend the training domain of the MTP as much as possible while keeping the amount of *ab initio* calculations relatively small. After the calculation of *ab initio* energies, forces and stresses of the selected configurations are added to the training set.
- Step 3** Fit the MTP on the updated training set. As the size of the training set grows on each iteration of the algorithm, this step will take more and more time during each subsequent iteration, but still this time is a small fraction of the time spent on *ab initio* calculations.
- Step 4** Repeat the steps 1–3, unless all the relaxations have successfully converged to the respective equilibrium configurations.

As we keep refitting the MTP during the relaxation on a dynamically updated training set, we call this algorithm as “relaxation while learning on-the-fly”.

3. Results and discussion

3.1. CuPd system

To test the applicability of our algorithm (Section 2.3) to the prediction of stable alloy structures we first used it to construct the CuPd

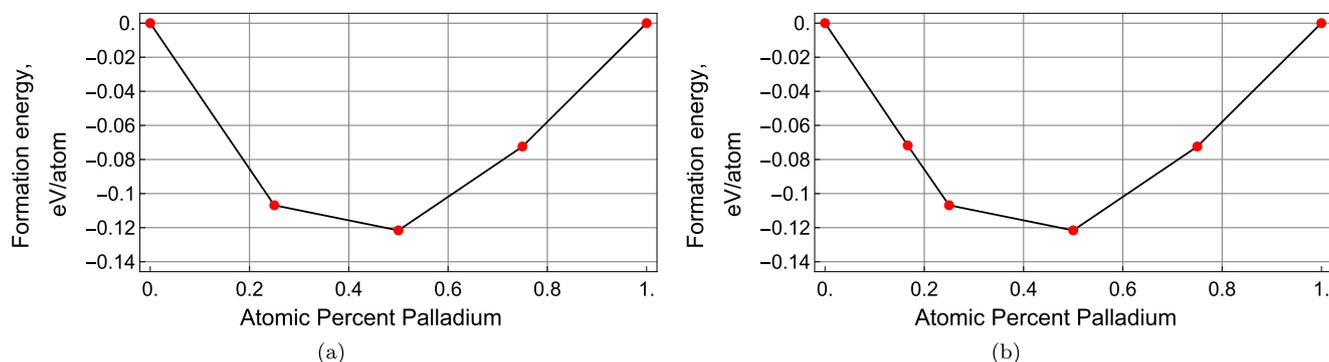


Fig. 6. Comparison of the convex hulls (a) as obtained from AFLOW and re-calculated with DFT, and (b) as found by MTP and re-calculated with DFT. We have discovered a structure at 16.6% Pd which is 0.5 meV lower than AFLOW's convex hull.

convex hull. We chose the Cu-Pd system because the structure of both pure Cu and Pd is fcc, while the stable equimolar CuPd structure is a bcc derivative structure. This system is a good test of whether or not our MTP-based model is able to simultaneously handle multiple lattice types, a challenging case for cluster expansion.

Using the algorithm of Ref. [15], we prepared 40,000 unrelaxed configurations with the bcc lattice and close-packed lattices (fcc and hcp), each configuration with 12 or less atoms in the unit cell and different concentrations of Cu and Pd. These were the candidate structures served as the input to our relaxation while learning-on-the-fly algorithm. We then equilibrated them and constructed a convex hull based on their relaxed energies. As follows from the scheme from Section 2.3, the training set increases on each iteration. The final training set was formed by 523 configurations. We call the training set “final” since an MTP trained on this set is able to relax all the candidate structures without exceeding threshold for the extrapolation grade. The root-mean-square error (RMSE) of energy per atom (σ) measured on this training set was equal to 2.3 meV/atom. We used $\text{lev}_{\max} = 16$ (refer to Section 2.1) to construct the MTP with about 200 parameters θ .

Fig. 6 shows the convex hulls constructed by the MTP and by high-throughput DFT calculations as reported in AFLOW. To make a direct comparison possible, both convex hulls were post-relaxed with DFT using the same settings (such as pseudopotentials, k-point mesh, etc.). As a result, we have found a structure with 16.6% concentration of Pd that is not presented in the AFLOW library and has energy per atom 0.5 meV below the AFLOW convex hull level. Though such a shallow ground state is typically not significant beyond academic interest, Cu-rich ground states are believed to have an effect on the experimental Cu-Pd phase diagram and have been discussed in Refs. [57,58] as a way of explaining the peculiar “off-stoichiometry” behavior on the Cu-rich side of the phase diagram.

It is illustrative to show the convex hull predicted by MTP and not post-relaxed with DFT. In Fig. 7, only structures within the 4σ (10 meV/atom) interval from the MTP convex hull are shown. Visually, the MTP convex hull looks slightly different due to the approximation errors of

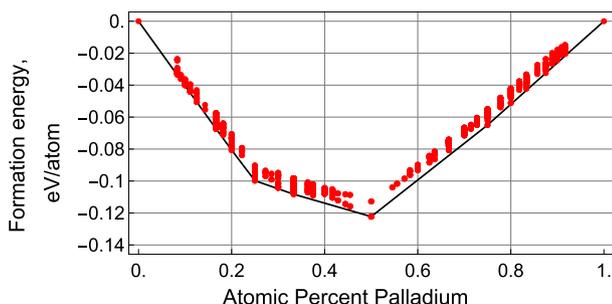


Fig. 7. Convex hull constructed by MTP and structures with formation energy within 10 meV/atom above the convex hull.

MTP leading to different relative levels of the structures on the “energy per atom” axis. Still, MTP reproduced the stable phases present in AFLOW library.

During the entire procedure, most of the computational expense (about 90%) was DFT calculations. In total, we did 523 single-point DFT calculations. If we relaxed all the 40,000 configurations using DFT, it would have taken about 10,000 times more computing time.

3.2. Co-Nb-V

We next test our algorithm on constructing a convex hull for the ternary Co-Nb-V system in the region where the concentration of Co is 50% or more. The number of initial candidates was about 27,000 and they were bcc-like and close-packed (fcc, hcp, etc.) configurations with 8 or less atoms in the unit cell and different concentrations of Co, Nb and V.

The MTP was trained on-the-fly and the final training set consisted of 383 configurations. The resulting convex hull is shown in Fig. 8. Remarkably, we have discovered a new structure with composition $\text{Co}_3\text{Nb}_2\text{V}$. It has a formation energy of 50 meV/atom below the AFLOW convex hull. Its unit cell and a layer-by-layer plot are shown in Fig. 9. We remark that geometrically this structure is different from any of those in the initial pool—e.g., the Nb atoms have 16 nearest neighbors with distances between 2.76 and 2.98 Å. It would hence be impossible to accurately treat such a configuration for both an on-lattice model,

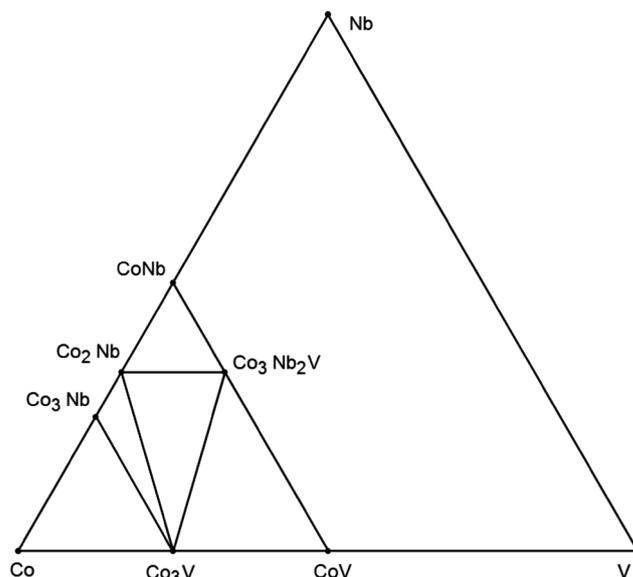


Fig. 8. Convex hull of the Co-Nb-V system constructed by MTP in the Co-rich region.

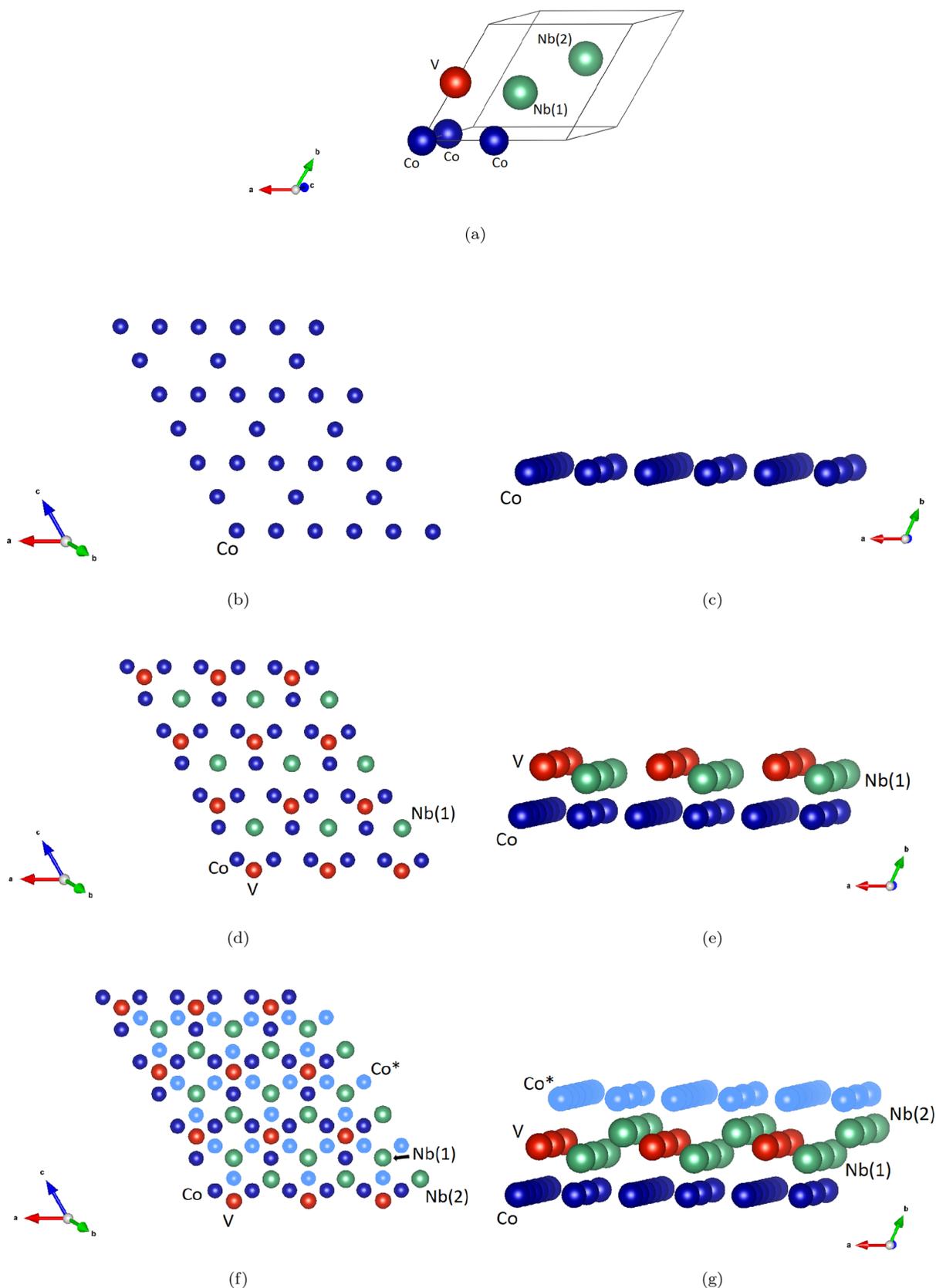


Fig. 9. The $\text{Co}_3\text{Nb}_2\text{V}$ discovered by MTP. The unit cell is shown in (a), while layer-by-layer plots in vertical and side projections are shown in (b)–(g). Co^* show were the next (periodically extended) layer of Co atoms are positioned. The structure was found, although no similar crystal prototypes were used.

such as cluster expansion, and an off-lattice model unless such a crystal prototype was known and explicitly added to the training set. This demonstrates the capabilities of our approach, combining an accurate off-lattice model and active learning.

3.3. Al-Ni-Ti

Finally, we applied our algorithm to the Al-Ni-Ti system. This system is well-studied and has many known ternary structures, some of which have over 20 atoms in the unit cell. We hence chose a set of candidate structures consisting of two parts: the first part has 1463 structures which were used in AFLOW as crystal prototypes.

The second part was generated by the algorithm from Ref. [15] and contains 375,000 binary and ternary structures enumerating all possible unit cells with different symmetries (bcc, fcc and hcp) and different number of atoms; we chose unit cells containing up to 12 atoms.

Including crystal structure prototypes adds extra difficulties: the structures may contain short interatomic distances (if, e.g., the original structure from which the prototype was derived had carbon-metal bonds which are shorter than typical metal-metal distances) and also smaller volume than that of the typical Al-Ni-Ti structures. Both of these features of the prototypes might result in unphysical structures with large stresses and forces which, in turn, lead to large MTP prediction errors. To make the unit cells of the candidate structures less deformed, we adjusted their volumes enforcing the relation:

$$v(n_{\text{Al}}, n_{\text{Ni}}, n_{\text{Ti}}) = n_{\text{Al}}v_{\text{Al}} + n_{\text{Ni}}v_{\text{Ni}} + n_{\text{Ti}}v_{\text{Ti}}, \quad (5)$$

where $v(n_{\text{Al}}, n_{\text{Ni}}, n_{\text{Ti}})$ is the volume per atom assigned to the unit cell with concentrations of Al, Ni, Ti equal to $n_{\text{Al}}, n_{\text{Ni}}, n_{\text{Ti}}$ respectively and $v_{\text{Al}}, v_{\text{Ni}}, v_{\text{Ti}}$ are the volumes per atom for equilibrium fcc-Al, fcc-Ni, hcp-Ti structures respectively. Resizing the unit cells in this way provides an initial guess for their volumes (a kind of “Vegard’s law” for different lattice types.)

To circumvent the large prediction errors that might occur for prototype structures with bond lengths and neighborhoods atypical of alloys, we performed a two-step relaxation as explained below. We used $\text{lev}_{\text{max}} = 20$ (see Section 2.1) to construct the MTP with about 650 parameters. This makes the potential more accurate, but requires more data for training, than with $\text{lev}_{\text{max}} = 16$. First, we did the same procedure as for the Cu-Pd and Co-Nb-V systems, which provided us with the training set of 2393 configurations with *ab initio* energies, forces and stresses. The MTP trained on this set has mean absolute error (MAE) and RMSE of energy per atom of 18 meV/atom and 27 meV/atom, respectively.

We next relaxed the 377,000 configurations and constructed a convex hull. Next, we picked all the configurations whose formation energy per atom is lower than 4σ (≈ 100 meV) from the convex hull level. This left us with 62,000 configurations.

Second, we repeated the procedure of relaxing the 62,000 configurations on-the-fly from scratch, starting from an empty training set. During this process a new training set with 976 structures was formed by the active learning algorithm. The MAE and RMSE on this training set was 7 meV/atom and 9 meV/atom, respectively. This way we constructed a convex hull based on more accurate formation energies, than would be possible after the first step.

To perform a comparison with the AFLOW convex hull, from the 62,000 relaxed configurations we eliminated all the configurations with formation energy per atom higher than 4σ from the convex hull level, where now $\sigma = 9$ meV/atom. This left us with about 7000 configurations, which were subsequently relaxed with DFT. After this, we constructed a final convex hull using the DFT formation energies. It has all the structures, present in AFLOW, and three new structures discovered by MTP (see Fig. 10). Their chemical formulas are given in Table 1 together with their position below the AFLOW convex hull level. Interestingly, all the structures are Ni-rich which makes their discovery relevant to the application of Ni-based alloys.

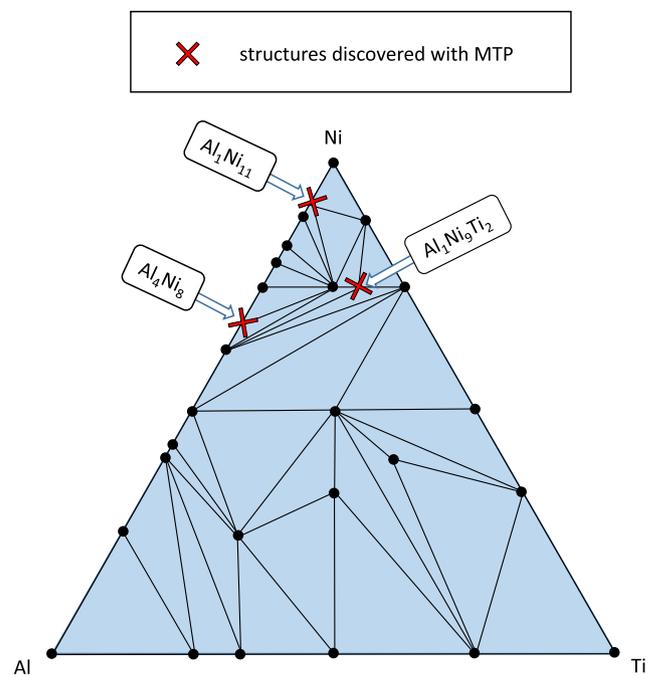


Fig. 10. Al-Ni-Ti convex hull constructed by MTP and compared to the one from AFLOW. The MTP convex hull contains all the structures from AFLOW plus three newly discovered ones.

Table 1

New Al-Ni-Ti structures found in this study. The “level below the convex hull” was computed using DFT.

Formula	Position below the convex hull (meV)
Al ₄ Ni ₈	−7.38
Al ₁ Ni ₁₁	−1.18
Al ₁ Ni ₉ Ti ₂	−0.34

Taking into account that after the first step we have obtained an MTP capable of relaxing all the 377,000 configurations, we call it the “robust” potential. After the second step we have obtained an MTP which is trained on (and thus able to relax) the low-energy near-equilibrium structures only. We refer to this MTP as the “accurate” potential. We attribute the difference in accuracies of the robust and accurate potentials to the fact that, at the second step, the trajectories of relaxations started from near-equilibrium structures (within the accuracy of the robust potential), see an illustration in Fig. 11. This reduces the region in the configurational space in which the MTP is fitted, thus improving the accuracy in comparison to the first step.

4. Conclusions

We have developed an algorithm for constructing a convex hull of stable alloy structures based on the moment tensor potentials (MTPs) to approximate *ab initio* energies, forces and stresses of atomistic configurations. This way the calculations for atomistic systems can be done much faster than with DFT, while the accuracy is comparable to that of DFT. The active-learning algorithm forms a training set automatically, removing the need for its manual design—the most tedious part of application of ML to atomistic modeling. We have verified the applicability of our algorithm by constructing the convex hulls for the Cu-Pd, Co-Nb-V and Al-Ni-Ti metallic alloy systems and comparing them to the convex hulls from AFLOW library. For all the systems we have discovered new stable structures, which are not listed in the AFLOW library. We attribute

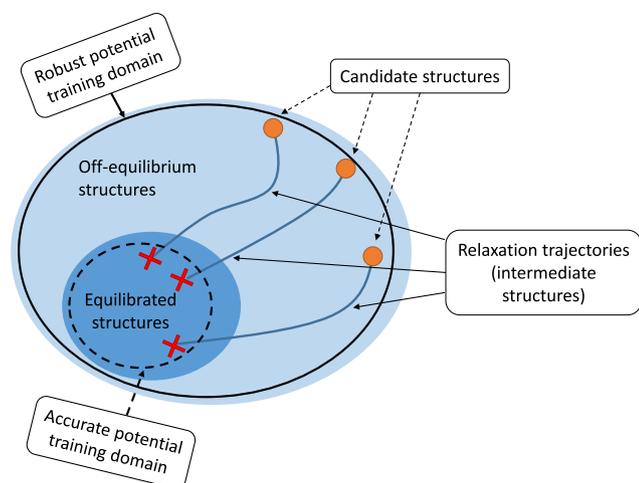


Fig. 11. The accurate potential is trained on a smaller domain of configurational space than the robust one. Thus, the accurate potential provides more accurate predictions at the interior of the “Relaxed structures” region.

this to the large amount of candidate structures (40,000 for Cu-Pd, 27,000 for Co-Nb-V, 377,000 for Al-Ni-Ti) we explored, which would be impossible to equilibrate using DFT. Instead, we performed relaxations using fast MTP calculations, referring to DFT only for the training data generation. In the cases covered by this paper, the amount of single-point DFT calculations was about 1% of the total amount of relaxed configurations. In comparison to the high-throughput DFT calculations, the speedup is three to four orders of magnitude.

5. Data availability

The data required to reproduce our results are available to download from http://gitlab.skoltech.ru/kgubaev/Data_for_MTP_with_active_learning.

CRedit authorship contribution statement

Konstantin Gubaev: Methodology, Software, Validation, Investigation, Writing – Original Draft, Visualization. **Evgeny V. Podryabinkin:** Methodology, Software. **Gus L.W. Hart:** Conceptualization, Validation, Writing – Review & Editing, Funding acquisition. **Alexander V. Shapeev:** Conceptualization, Methodology, Software, Writing – Original Draft, Supervision, Funding acquisition.

Acknowledgments

The work of K.G., E.V.P., and A.V.S. was supported by the Russian Science Foundation (Grant No. 18-13-00479). G.L.W.H. was supported by the Office of Naval Research (MURI N00014-13-1-0635). This work was performed, in part, by A.V.S. at the Center for Integrated Nanotechnologies, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science by Los Alamos National Laboratory (Contract DE-AC52-06NA25396) and Sandia National Laboratories (Contract DE-NA-0003525). This collaboration has started during the MPS2016 long program at the Institute of Pure and Applied Mathematics, UCLA.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.commatsci.2018.09.031>.

References

- [1] S. Curtarolo, W. Setyawan, G.L.W. Hart, M. Jahnatek, R.V. Chepulskii, R.H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, et al., *Comput. Mater. Sci.* 58 (2012) 218.
- [2] J.E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, *JOM* 65 (2013) 1501.
- [3] D.D. Landis, J.S. Hummelshøj, S. Nestorov, J. Greeley, M. Dułak, T. Bligaard, J.K. Nørskov, K.W. Jacobsen, *Comput. Sci. Eng.* 14 (2012) 51.
- [4] The novel materials discovery (NOMAD) laboratory, a European centre of excellence. <<http://nomad-repository.eu>>.
- [5] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, *APL Mater.* 1 (2013) 011002.
- [6] S.R. Kalidindi, M. De Graef, *Annu. Rev. Mater. Res.* 45 (2015) 171.
- [7] A. Agrawal, A. Choudhary, *Appl Mater.* 4 (2016) 053208.
- [8] K. Rajan, *Annu. Rev. Mater. Res.* 45 (2015) 153.
- [9] C. Nyshadham, C. Oses, J.E. Hansen, I. Takeuchi, S. Curtarolo, G.L.W. Hart, *Acta Mater.* 122 (2017) 438.
- [10] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R.H. Taylor, L.J. Nelson, G.L.W. Hart, S. Sanvito, M. Buongiorno-Nardelli, et al., *Comput. Mater. Sci.* 58 (2012) 227.
- [11] G.L.W. Hart, S. Curtarolo, T.B. Massalski, O. Levy, *Phys. Rev. X* 3 (2013) 041035.
- [12] Y. Hinuma, T. Hatakeyama, Y. Kumagai, L.A. Burton, H. Sato, Y. Muraba, S. Iimura, H. Hirata, I. Tanaka, H. Hosono, et al., *Nat. Commun.* 7 (2016) 11962.
- [13] S. Sanvito, C. Oses, J. Xue, A. Tiwari, M. Zic, T. Archer, P. Tozman, M. Venkatesan, M. Coey, S. Curtarolo, *Sci. Adv.* 3 (2017) e1602241.
- [14] M.J. Mehl, D. Hicks, C. Toher, O. Levy, R.M. Hanson, G.L.W. Hart, S. Curtarolo, *Comput. Mater. Sci.* 136 (2017) S1.
- [15] G.L.W. Hart, L.J. Nelson, R.W. Forcade, *Comput. Mater. Sci.* 59 (2012) 101.
- [16] A.O. Lyakhov, A.R. Oganov, H.T. Stokes, Q. Zhu, *Comput. Phys. Commun.* 184 (2013) 1172.
- [17] Q. Wu, B. He, T. Song, J. Gao, S. Shi, *Comput. Mater. Sci.* 125 (2016) 243.
- [18] C. Jiang, B.P. Uberuaga, *Phys. Rev. Lett.* 116 (2016) 105501.
- [19] M.C. Tropicovskiy, J.R. Morris, M. Daene, Y. Wang, A.R. Lupini, G.M. Stocks, *JOM* 67 (2015) 2350.
- [20] A.H. Nguyen, C.W. Rosenbrock, C.S. Reese, G.L.W. Hart, *Phys. Rev. B* 96 (2017) 014107.
- [21] B. Meredig, A. Agrawal, S. Kirklin, J.E. Saal, J. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, *Phys. Rev. B* 89 (2014) 094104.
- [22] S. Ubaru, A. Miedlar, Y. Saad, J.R. Chelikowsky, *Phys. Rev. B* 95 (2017) 214102.
- [23] A. Shapeev, *Comput. Mater. Sci.* 139 (2017) 1612.03359.
- [24] L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, *npj Comput. Mater.* 2 (2016) 16028.
- [25] A.V. Shapeev, *Multiscale Model. Simul.* 14 (2016) 1153.
- [26] E.V. Podryabinkin, E.V. Tikhonov, A.V. Shapeev, A.R. Oganov, Also Available at: arXiv preprint arXiv:1802.07605, 2018.
- [27] E.V. Podryabinkin, A.V. Shapeev, *Comput. Mater. Sci.* 140 (2017) 171.
- [28] K. Gubaev, E.V. Podryabinkin, A.V. Shapeev, *J. Chem. Phys.* 148 (2018) 241727.
- [29] J. Behler, M. Parrinello, *Phys. Rev. Lett.* 98 (2007) 146401.
- [30] A.P. Bartók, M.C. Payne, R. Kondor, G. Csányi, *Phys. Rev. Lett.* 104 (2010) 136403.
- [31] N. Artrith, A.M. Kolpak, *Comput. Mater. Sci.* 110 (2015) 20.
- [32] J. Behler, *Phys. Chem. Chem. Phys.* 13 (2011) 17930.
- [33] J. Behler, *J. Phys. Condens. Matter.* 26 (2014) 183001.
- [34] J.R. Boes, M.C. Groenenboom, J.A. Keith, J.R. Kitchin, *Int. J. Quant. Chem.* 116 (2016) 979.
- [35] P.E. Dolgirev, I.A. Kruglov, A.R. Oganov, *AIP Adv.* 6 (2016) 085318.
- [36] M. Gastegger, P. Marquetand, *J. Chem. Theory Comput.* 11 (2015) 2187.
- [37] S. Manzhos, R. Dawes, T. Carrington, *Int. J. Quant. Chem.* 115 (2015) 1012.
- [38] S.K. Natarajan, T. Morawietz, J. Behler, *Phys. Chem. Chem. Phys.* 17 (2015) 8356.
- [39] N. Lubbers, J.S. Smith, K. Barros, *J. Chem. Phys.* 148 (2018) 241715.
- [40] J.S. Smith, O. Isayev, A.E. Roitberg, *Chem. Sci.* 8 (2017) 3192.
- [41] B. Kolb, L.C. Licht, A.M. Kolpak, *Sci. Rep.* 7 (2017), <https://doi.org/10.1038/s41598-017-01251-z>.
- [42] W.J. Szlachta, A.P. Bartók, G. Csányi, *Phys. Rev. B* 90 (2014) 104108.
- [43] V.L. Deringer, G. Csányi, *Phys. Rev. B* 95 (2017) 094203.
- [44] V.L. Deringer, C.J. Pickard, G. Csányi, *Phys. Rev. Lett.* 120 (2018) 156001.
- [45] A. Grisafi, D.M. Wilkins, G. Csányi, M. Ceriotti, *Phys. Rev. Lett.* 120 (2018) 036002.
- [46] A. Thompson, L. Swiler, C. Trott, S. Foiles, G. Tucker, *J. Comput. Phys.* 285 (2015) 316.
- [47] V. Botu, R. Ramprasad, *Phys. Rev. B* 92 (2015) 094306.
- [48] Z. Li, J.R. Kermode, A. De Vita, *Phys. Rev. Lett.* 114 (2015) 096405.
- [49] I. Kruglov, O. Sergeev, A. Yanilkin, A.R. Oganov, *Sci. Rep.* 7 (2017) 8512.
- [50] S. Chmiela, H.E. Sauceda, K.-R. Müller, A. Tkatchenko, Also Available at: arXiv preprint arXiv:1802.09238, 2018.
- [51] K. Schütt, P.-J. Kindermans, H.E.S. Felix, S. Chmiela, A. Tkatchenko, K.-R. Müller, *Adv. Neural Inform. Process. Syst.* (2017) 992–1002.
- [52] V. Botu, R. Batra, J. Chapman, R. Ramprasad, Also Available at: arXiv preprint arXiv:1610.02098, 2016.
- [53] J.S. Smith, B. Nebgen, N. Lubbers, O. Isayev, A.E. Roitberg, *J. Chem. Phys.* 148 (2018) 241733.
- [54] R. Kondor, Also Available at: arXiv preprint arXiv:1803.01588, 2018.
- [55] M. Hirn, S. Mallat, N. Poilvert, *Multiscale Model. Simul.* 15 (2017) 827.
- [56] S. Goreinov, I.V. Oseledets, D.V. Savostyanov, E.E. Tyrtyshnikov, N. Zamarashkin, *Matrix Methods: Theory Algorithms And Applications: Dedicated to the Memory of Gene Golub*, World Scientific, 2010, pp. 247–256.
- [57] S. Bärthlein, G.L.W. Hart, A. Zunger, S. Müller, *J. Phys.: Condens. Matter* 19 (2007) 032201.
- [58] S. Bärthlein, E. Winning, G.L.W. Hart, S. Müller, *Acta Mater.* 57 (2009) 1660.