ARTICLE OPEN

Check for updates

Performance of two complementary machine-learned potentials in modelling chemically complex systems

Konstantin Gubaev [1,5 ✉], Viktor Zaverkin [2,4,5 ✉], Prashanth Srinivasan [1], Andrew Ian Duff [3], Johannes Kästner [2] and Blazej Grabowski [1]

Chemically complex multicomponent alloys possess exceptional properties derived from an inexhaustible compositional space. The complexity however makes interatomic potential development challenging. We explore two complementary machine-learned potentials—the moment tensor potential (MTP) and the Gaussian moment neural network (GM-NN)—in simultaneously describing configurational *and* vibrational degrees of freedom in the Ta-V-Cr-W alloy family. Both models are equally accurate with excellent performance evaluated against density-functional-theory. They achieve root-mean-square-errors (RMSEs) in energies of less than a few meV/atom across 0 K ordered and high-temperature disordered configurations included in the training. Even for compositions not in training, relative energy RMSEs at high temperatures are within a few meV/atom. High-temperature molecular dynamics forces have similarly small RMSEs of about 0.15 eV/Å for the disordered quaternary included in, and ternaries not part of training. MTPs achieve faster convergence with training size; GM-NNs are faster in execution. Active learning is partially beneficial and should be complemented with conventional human-based training set generation.

INTRODUCTION

Alloys consisting of multiple principal components have attracted considerable attention over the last decade[1–8]. Apart from the conventional subsystems like binary and ternary alloys, one can also access the interior regions of the highly complex compositional space of such alloys, revealing a multitude of untapped compositions. High entropy alloys (HEAs) are one such class of alloys consisting of at least four to five elements in near-equal proportions. They are known for remarkable properties such as high tensile and yield strength, high ductility, and fracture toughness[9,10].

A high mixing entropy stabilises the disordered phase of a HEA wherein elements randomly occupy sites of the simple lattices, i.e., bcc[11,12], fcc[7,13], or hcp[14]. However, this is the predominant effect only at high temperatures, accompanied by large vibrations in the system. At lower temperatures, there is a possibility of segregation of individual elements, short-range ordering, and phase decomposition[15–19]. Therefore, to fully analyse the thermodynamics of the alloy across the entire temperature range, both vibrational and configurational entropy contributions are critical, and one needs to study both the high-temperature disordered phase and the low-temperature segregated sub-phases. In addition, the properties of a HEA cannot be deduced by simply performing a linear combination of properties of individual elements and subsystems, as the interaction between the elements leads to unusual behaviour.

Computational techniques can assist in a rapid exploration of complex compositional spaces. However, the accuracy of results relies critically on the computational method. Running ab initio molecular dynamics (AIMD) simulations using density-functional theory (DFT) to obtain accurate results is computationally demanding. In particular, studying configurations with short-range order requires much larger length scales making DFT

prohibitively expensive. Atomistic simulations employing interatomic potentials are a possible alternative. Most potentials assume the locality of interatomic interactions, facilitating much faster computation and better scaling with system size than DFT.

The key challenge lies in developing accurate and robust interatomic potentials. In particular, simultaneously accounting for both the vibrational and configurational degrees of freedom is extremely difficult. Conventional potentials, e.g., based on the embedded-atom method (EAM)[20] or the modified embedded-atom method (MEAM)[21], are restricted to simple fixed analytical forms for the energy with a small number of parameters. The sparsity of the respective parameter space does not provide enough flexibility to model complex systems.

In this regard, state-of-the-art machine learning (ML)-based models have a more systematic approach toward the parameterization of interatomic interactions. The atomic environment around a central atom is encoded by a local representation, which depends on vibrational (coordinates) and configurational (atomic types) degrees of freedom. Typically, local representations are composed of a radial part (depending only on interatomic distances) and an angular part (depending on the angles). The representations are invariant with respect to atomic permutations, translations and rotations of physical space and form a set of functions that describe the local atomic arrangement. This set of functions provides suitable input to an ML model. The mapping from this input to the energy is determined by the values of model parameters, which are obtained by training on DFT energies, atomic forces, and stresses.

ML-based models can also robustly estimate the uncertainty in their predicted energies and forces[22–28]. Such estimations are used to develop active learning algorithms to select the data most informative to the model. The data are obtained from the

[1]Institute for Materials Science, University of Stuttgart, Pfaffenwaldring 55, 70569 Stuttgart, Germany. [2]Institute for Theoretical Chemistry, University of Stuttgart, Pfaffenwaldring 55, 70569 Stuttgart, Germany. [3]Scientific Computing Department, STFC Daresbury Laboratory, Warrington WA4 4AD, UK. [4]Present address: NEC Laboratories Europe GmbH, Kurfürsten-Anlage 36, 69115 Heidelberg, Germany. [5]These authors contributed equally: Konstantin Gubaev, Viktor Zaverkin. ✉email: yamir4eg@gmail.com; zaverkin@theochem.uni-stuttgart.de

configurational and vibrational spaces sampled during, e.g., a molecular dynamics (MD) simulation using the existing model, either beforehand or in an on-the-fly fashion. The model is then retrained with the updated training data set, which includes DFT values of these sampled configurations. Hence, in addition to their unique way of accurately describing a complex multidimensional phase space with a certain number of model parameters, the learning-on-the-fly feature allows ML-based models to be a more robust and reliable approximation of the DFT potential energy surface.

Several ML-based models have been proposed in the literature, each with its own representation of the atomic environment[29–32]. In the present work, we analyse two complementary ML-based approaches, the moment tensor potential (MTP)[23,33] and the Gaussian moment neural network (GM-NN)[34,35]. MTPs have been developed for thermodynamic property prediction up to the melting point of a single disordered phase of a TaVCrW HEA[36]. MTPs have also been built to study partial compositional spaces of a TiZrHfTa$_x$ HEA[37]. GM-NNs, on the other hand, have so far been developed for a different class of chemical structures. For instance, they were used to study the dynamics of N and H_2 on water and CO ice surfaces, respectively[38–40], and the dynamic behaviour of magnetic anisotropy tensors of $[Co(N_2S_2O_4C_8H_{10})_2]^{2-}$, $[Fe(tpa)^{Ph}]^-$, and $[Ni(HIM_2 - py)_2NO_3]^+$ molecular crystals[41]. However, they have not been tested for metallic, let alone, complex HEA systems.

Here, we develop and compare an MTP and a GM-NN potential for the Ta-V-Cr-W HEA family, and assess their performance in modelling chemically-complex multicomponent systems. We include the vibrational and configurational degrees of freedom of the system in the potential development. We do so by attempting to predict energies and forces in Ta-V-Cr-W sub-systems under two scenarios: (i) the 0 K energies and forces in binaries, ternaries and quaternaries; (ii) near-melting temperature energies and forces in 3- and 4-component disordered alloys. The subsystems are categorized into two groups—the in-distribution, and the out-of-distribution subsystems, where the former includes those subsystems used for training, and the latter includes non-equiatomic ternaries that are not a part of the training. In addition to the energy and force estimates, we assess the quality of the potentials by also computing thermodynamic properties.

The MTP and the GM-NN models are equally accurate in describing the Ta-V-Cr-W HEA family. The 0 K energies and forces of the in-distribution subsystems are predicted extremely accurately, while the RMS errors are slightly larger for the out-of-distribution subsystems. The high-temperature configurational space of the disordered quaternary (part of the in-distribution subsystems) is accurately captured to within 0.18 eV/Å RMS error in forces. Likewise, even non-equiatomic ternaries (part of the out-of-distribution subsystems) are accurately predicted, with RMS errors in forces of less than 0.16 eV/Å. While the absolute energies of the out-of-distribution subsystems are not predicted well, the relative energies—which are relevant for thermodynamics properties—are accurate to within 5.5 meV/atom. Classical EAM/MEAM potentials are unsuitable for such a complex phase space and have errors one to two orders higher in magnitude. Finally, we investigate the role of active learning strategies applied to HEAs and find that human-based inspection is crucial in this specific case.

RESULTS

Features of the two ML models

Figure 1 is a schematic illustration of the similarities and differences between the MTP and the GM-NN approaches. MTPs and GM-NNs assume locality of interatomic interactions. The total energy of an atomic system is split into the contributions of individual atoms. For every atom, only the interactions within a cutoff R_{cut} are explicitly considered, as shown in Fig. 1a. Since the total energies (written as a sum of site energies for all atoms) are trained on DFT systems which include long-range effects, the interactions beyond R_{cut} get implicitly included during the fitting.

The building blocks of the MTPs and GM-NNs are the atomic environment descriptors, shown in Fig. 1b. MTPs and GM-NNs encode the structural and chemical composition of a local atomic environment by utilizing symmetry preserving representations motivated by invariant polynomials and Cartesian Gaussian-type orbitals (GTOs)[42], respectively. Interestingly, both approaches lead to a common representation invariant to translations and permutations but equivariant to rotations.

Descriptors, invariant to rotations, are obtained by computing tensor contractions and eliminating possible linear dependencies; see Fig. 1c. For MTP, the selected contractions are restricted to the ones of a particular level lev$_{max}$[33], providing restriction on both radial and angular parts and thus defining the basis in which the atomic energy is expanded. Larger values of lev$_{max}$ account for high-order tensors participating in contractions, thereby producing more descriptors and increasing the fitting capabilities of the potential. In fact, the EAM and MEAM potentials can be viewed as reduced MTP potentials with lev$_{max} = 0$ and lev$_{max} = 2$, respectively. For GM-NN, linear dependencies are eliminated by identifying unique generating graphs corresponding to the tensor contractions[43] and taking into account the permutational symmetry of the respective tensors or, equivalently, nodes on the graph[34,35]. The intermediate equivariant representation relates MTP and GM-NN to the recently proposed equivariant message passing neural network (NN) architectures[44,45].

The invariant features of the atomic environments computed within both approaches are then used as an input to linear regression or a neural network-based model for MTP and GM-NN, respectively, as illustrated in Fig. 1d. The GM-NN has ~500 times more parameters as it utilizes artificial NNs to map the structure and the respective properties. Finally, the models are trained on DFT data (energies, atomic forces and stresses) by minimizing a specified loss function. The reader is referred to the 'Methods' section for more details about the construction of the ML-based models, and the relation of the MTP and GM-NN models to other ML potentials.

From configurations generated by the trained model, new samples are iteratively selected for active learning based on uncertainty quantification. In the MTP and GM-NN approach, uncertainty measures are based on gradient feature maps, i.e., $\nabla_{\boldsymbol{\theta}}E$ with $\boldsymbol{\theta}$ being the model's parameters. In the simplest case, $\nabla_{\boldsymbol{\theta}}E$ would correspond to input features of the linear regression in Fig. 1d. Specifically, for MTP, the uncertainty measure is based on the maximum volume (MaxVol) principle for the matrix of the basis function derivatives with respect to the fitting coefficients[46]. The actual measure of the uncertainty is the determinant of the corresponding MaxVol submatrix. For GM-NN, the uncertainty is computed by employing gradient features of a trained NN (with $\boldsymbol{\theta}$ being the weights and biases)[26–28], corresponding to the finite-width neural tangent kernel[47]. In this work, the largest cluster maximum distance (LCMD) method[28] for selecting new structures and last-layer gradient features (FEAT(LL)) to compute the similarity between structures and the model's uncertainty from it is used. Other active learning approaches are also possible[27,28].

Evaluation strategy

The objective of the ML-potentials is to capture both the vibrational and configurational degrees of freedom at near-DFT accuracy. The TaVCrW HEA stabilizes as a fully disordered structure at high temperature. As the temperature decreases, there is short-range ordering in the alloy, followed by segregation of individual elements or phases[18]. Hence, a robust ML model that accurately
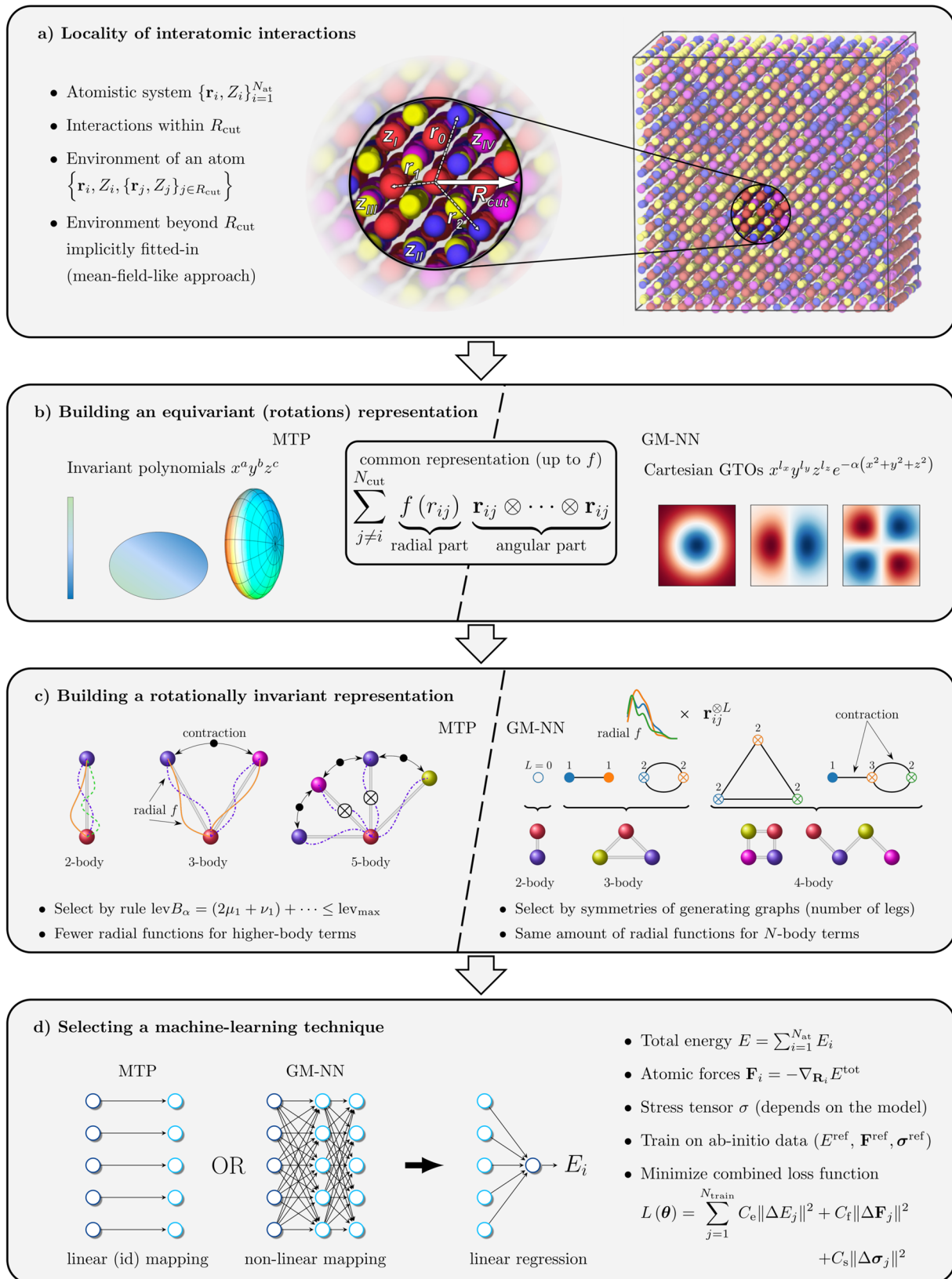
Fig. 1 Schematic representation of the moment tensor potential (MTP) and the Gaussian moment neural network (GM-NN) approaches.
a Both methods rely on the assumption of locality of interatomic interactions but implicitly include interactions beyond the cutoff radius R_{cut}.
b As an intermediate step in constructing invariant features, both methods obtain a common representation equivariant to rotations.
c Invariant features are obtained by computing tensor contractions and eliminating linear dependencies specific to the considered approach.
d Utilizing linear regression or neural networks, which take invariant features as an input, mapping from a structure to the respective energy is modelled.
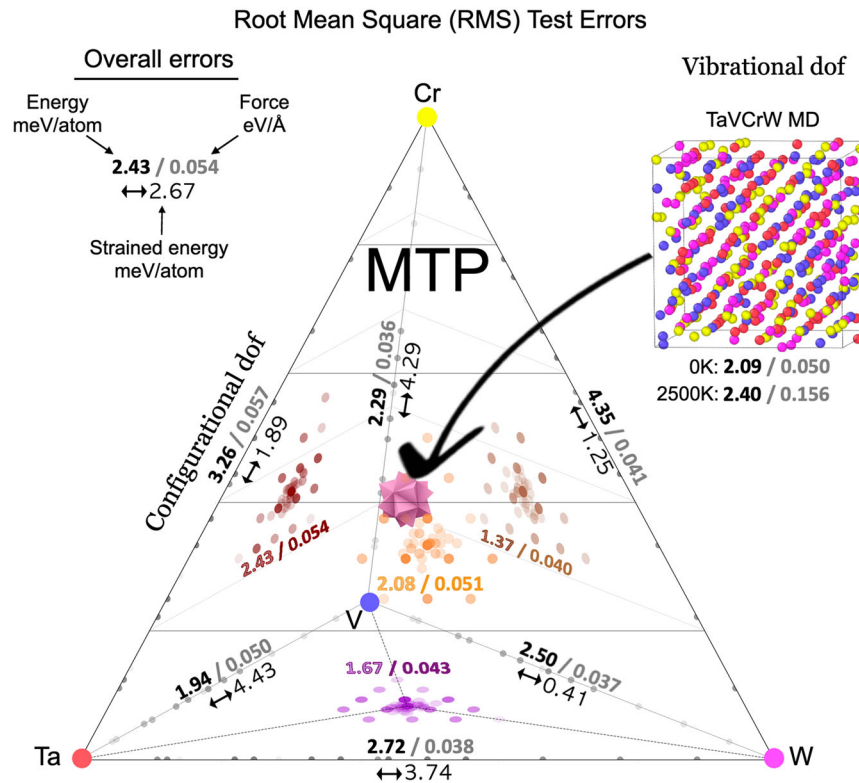
Fig. 2 RMS errors in predicted energies and forces for the MTP model evaluated on the test data for different in-distribution subsystems of Ta-V-Cr-W. Energy RMS errors in meV/atom are shown in bold, while the atomic force RMS errors in eV/Å are in grey. The numbers with arrows indicate the RMS errors for the energy per atom for strained binary structures.

predicts the complete HEA behaviour should be able to predict both the high-temperature disordered phase, and the low-temperature ordered phases.

Hence, we choose structures that represent both these temperature regimes for training and evaluation of the ML potentials. The training data includes 0 K energies and forces of ordered phases—binaries, ternaries and quaternaries. In addition, we also include configurations of strained binaries and of the disordered TaVCrW alloy at 2500 K. These subsystems that form the training set are classified as in-distribution subsystems.

To further investigate the generalization capabilities of the ML-potentials, we also evaluate their performance on subsystems from different parts of the compositional space of the Ta-V-Cr-W family that do not form a part of the training set. For this, we choose a set of non-equiatomic ternaries, and evaluate both the 0 K and high-temperature energy and force predictions of the ML-potentials. These subsystems are classified as out-of-distribution subsystems. The structures in the training and validation data set of the in-distribution and out-of-distribution subsystems are detailed in the 'Methods' section.

Accuracy for the in-distribution subsystems

Figure 2 shows the root-mean-square (RMS) errors in the predicted energies and forces for all in-distribution subsystems using the MTP. The RMS error-tetrahedra for the GM-NN potential and the EAM/MEAM potentials are compiled in the Supplementary Information. In addition to the 0 K subsystems (binaries, ternaries, quaternary), Fig. 2 also shows errors for the 0 K strained binaries and the 2500 K disordered HEA, all of which form the in-distribution subsystems. The 'Overall errors' correspond to the RMS errors on all configurations put together. The results are additionally compiled in Table 1. To guarantee reliable statistics, all

values refer to averages over ten separately fitted MTPs and GM-NNs.

From Fig. 2 and Table 1 (and Fig. S1 in the Supplementary Information), we see that both ML-based models MTP and GM-NN are almost equally accurate and reproduce DFT values extremely well for the in-distribution subsystems. For the 0 K subsystems, the RMS errors in energies and forces range from 1–4 meV/atom and 0.02–0.06 eV/Å, respectively, and for the 2500 K disordered structure, the RMS errors are around 2.5 meV/atom and 0.16 eV/Å, respectively. The strained binary structures are also predicted to within 4.5 meV/atom error in energy, with some binaries (CrW, TaCr, and VW) being predicted more accurately than the others, irrespective of the model chosen. The MTP is marginally better than the GM-NN in predicting the 2500 K structure (i.e., the vibrational degrees of freedom). The GM-NN predicts the 0 K structures marginally better (i.e., the configurational degrees of freedom).

Figure 3 shows the correlation between the predicted forces for the 0 K and the 2500 K TaVCrW structures. The 0 K data in the training and validation data set include a combination of relaxed and unrelaxed structures, and they span a broad range of force values. For example, as seen in Fig. 3, the forces range from −2 to 2 eV/Å for the TaVCrW 0 K structures. For the 2500 K MD structure, they range from −10.5 to 10.5 eV/Å. Irrespective of this, there is an excellent agreement in the forces predicted by the ML-based models for all structures of the in-distribution subsystems.

The overall accuracy achieved by the MTP and the GM-NN with respect to the DFT values is remarkable, given the complexity of the configurational phase space (which includes both 0 K subsystems and a 2500 K disordered quaternary). To put it into perspective, Table 1 also shows the RMS errors of an EAM that was fit to a combination of the 2500 K 4-component structure and 0 K subsystems. Both the energy and force errors are one to two orders of magnitude higher, deeming them almost unusable for

Table 1. RMS errors in the predicted energies/forces for the individual Ta-V-Cr-W in-distribution subsystems.

Subsystem	MTP[a]	GM-NN[b]	EAM[a]
TaV	1.94/0.050	**1.54/0.029**	32.0/0.404
TaCr	3.26/0.057	**2.98/0.038**	43.6/0.343
TaW	**2.72**/0.038	2.99/**0.025**	44.8/0.248
VCr	**2.29**/0.036	2.82/**0.025**	44.8/0.270
VW	2.50/0.037	**2.00/0.023**	21.3/0.292
CrW	4.35/0.041	**2.87/0.029**	23.4/0.248
TaVCr	2.43/0.054	**1.97/0.045**	34.1/0.313
TaVW	**1.67**/0.043	1.70/**0.034**	39.6/0.321
TaCrW	**2.08**/0.051	2.19/**0.039**	23.6/0.327
VCrW	**1.37**/0.040	1.94/**0.031**	19.4/0.314
TaVCrW (0 K)	**2.09**/0.049	2.16/**0.037**	50.8/0.488
TaVCrW (2500 K)	**2.40/0.156**	2.67/0.179	59.4/0.521
Overall	2.43/0.054	**2.32/0.043**	37.14/0.443
Deformed Structures[c]			
TaV	4.43	**3.63**	56.8
CrW	1.25	**1.04**	27.1
TaCr	1.89	**0.49**	13.3
VW	**0.41**	0.52	66.1
TaW	3.74	**3.11**	161.9
VCr	4.29	**2.70**	283.2
Overall	2.67	**1.91**	101.4

The results are provided for the MTP, GM-NN, and EAM models. Atomic energies are given in meV/atom, while forces are given in eV/Å. The best training result is written in boldface.
[a]The results for MTP and EAM are obtained by averaging over ten splits of the original data set, except for the deformed structures. The latter were obtained for all models trained on the whole data set (training + test).
[b]For GM-NN models, 500 structures were selected randomly from the respective training set to prevent overfitting using the early stopping technique[54]. Thus, GM-NN models were trained on 500 structures less than MTP and EAM. This procedure was repeated ten times for each of the original data splits and resulted in averaging over 100 splits in total. For the deformed structures, the results are averaged over ten splits.
[c]Binary structures are strained along the [100] direction and used as a part of the validation set. For different binaries, the strains are chosen based on certain accommodation rules (see 'Methods' section). The strains can vary up to 4.71%.

any accurate property prediction. The last column in Fig. 3 shows the force correlation between the EAM-predicted forces and the DFT forces on both 0 K and 2500 K TaVCrW structures. The 0 K forces predicted by the EAM are highly uncorrelated to the DFT values, and the 2500 K forces have a much broader distribution (larger RMS error) compared to the two ML-based models.

Heat capacity prediction for the HEA

Having shown the remarkable accuracy in energies and forces achievable with the ML-based models for the in-distribution subsystems, we now test their reliability and robustness in predicting thermodynamic properties of the disordered TaVCrW HEA. In particular, we analyse the isobaric heat capacity, $C_p(T)$, up to 2500 K (close to the melting point). The heat capacity is calculated numerically using thermodynamic integration, the details of which are in the 'Methods' section.

Figure 4 shows the heat capacity curves obtained using the MTP and the GM-NN. The curves are compared to the DFT results from literature[36]. We have removed the electronic part from the full DFT

curve to be able to directly compare to the results predicted by the current potentials. Both the MTP and the GM-NN predict almost the same behaviour of the heat capacity with temperature. The curves are also very close to the full DFT curve up to around 2200 K after which the ML-based curves diverge. The heat capacity, calculated as the second derivative of the Gibbs energy with temperature, is extremely sensitive to small changes in the free energy. Hence, given that only 25% of the training set comprised of disordered quaternary structures (0 K and 2500 K), the accuracy achieved by both the MTP and GM-NN for the heat capacity is outstanding. The curves are also compared to the heat capacity calculated using the EAM in Fig. 4. The EAM that was fitted to the entire training set whose RMS errors were compared in Table 1 was not stable enough to run TI across the entire temperature and volume range. Hence, we have fitted an EAM only to the 2500 K disordered TaVCrW structures to calculate the heat capacity. Even then, the heat capacity calculated by the EAM diverges from the DFT curve after 750 K and is much less accurate.

Accuracy for the out-of-distribution subsystems

Figure 5 shows the errors in the predicted energies and forces (finite-temperature results at 2000 K and 0 K results) for the out-of-distribution subsystems (non-equiatomic ternaries on the x-axis) using both the MTP and the GM-NN potential. The first row of plots contains the errors in absolute energies of the various subsystems. It is observed that the prediction of the absolute energies by the ML-based potentials is off by as much as 20 meV/atom in some cases. The out-of-distribution subsystems are not a part of the training set. The ML-potentials—although trained on absolute energies of configurations of the in-distribution subsystems—cannot accurately predict the large, absolute chemical potentials. Hence, we observe these much higher errors in the absolute energies of the out-of-distribution subsystems compared to those from the in-distribution subsystems.

However, the temperature dependence of thermodynamic properties such as the heat capacity discussed in the previous section, are numerical functions of energy differences instead of absolute energies. Thus the absolute energy predictions do not serve as a rigorous assessment of the accuracy of the ML-based potentials in predicting thermodynamic properties. In the second row in Fig. 5, we show the respective RMS errors in the relative energy (with respect to the corresponding relaxed structure) of the out-of-distribution subsystems, and find that the errors are less than 5.5 meV/atom.

In the third row of Fig. 5, we also show the errors in atomic forces at both 2000 K (left) and 0 K (right). The errors observed in the 2000 K forces are similar to those for the in-distribution systems with values not exceeding 0.16 eV/Å. Accurate prediction of MD trajectories at elevated temperatures can be thus expected for various alloy compositions in the Ta-V-Cr-W family. The errors in the 0 K forces are even smaller, and comparable to the 0 K subsystems in the in-distribution set. An exception is the $Cr_{17}W_{17}V_{66}$ alloy, where the higher error comes from the distinctly higher absolute value of the force in the corresponding configuration (marked on the top of the plot). Hence, relative force errors are comparable among the different out-of-distribution subsystems.

In clear contrast to the ML models, the performance of the EAM on the out-of-distribution subsystems is much worse. The absolute energies are off by 120–200 meV/atom, while even the relative energies are off by 60–80 meV/atom. The RMS error in the forces in the ternaries is similar to the error observed for the TaVCrW quaternary with values ranging from 0.5 to 0.7 eV/Å. Such errors make the EAM unsuitable, even more so, for thermodynamic property predictions of out-of-distribution subsystems.

Overall, the performance of the ML-based potentials is equally and remarkably good even for out-of-distribution subsystems. The observed larger errors in the absolute energies could be, for
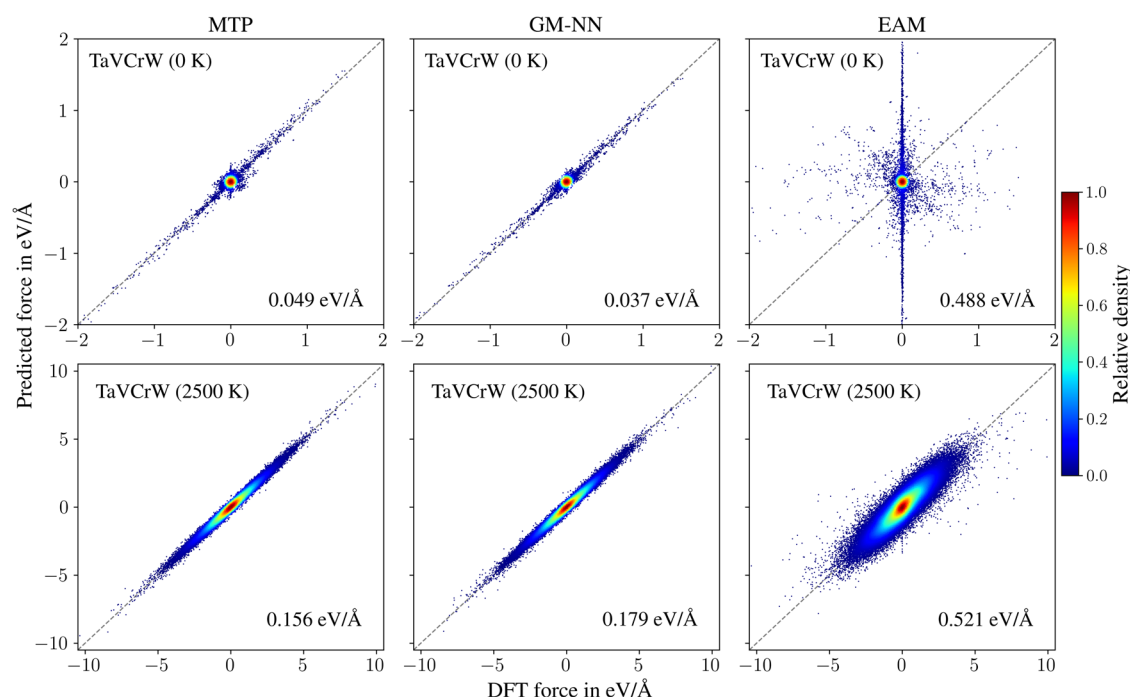
Fig. 3 Correlation between the predicted forces (by the MTP, the GM-NN and the EAM potential) and the DFT forces in eV/Å for the 0 K and 2500 K structures of the TaVCrW alloy. The colours indicate the relative density as estimated by a kernel density estimator.
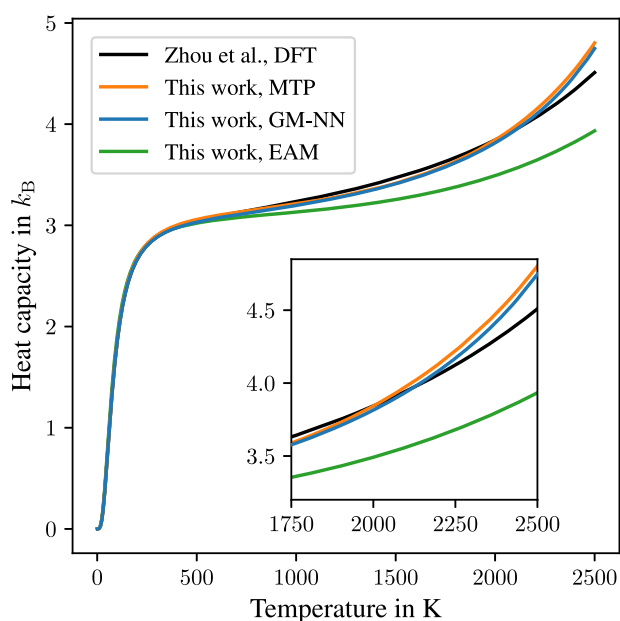

Fig. 4 Isobaric heat capacity as a function of temperature computed by employing the MTP, the GM-NN, and the EAM interatomic potentials. The reference DFT values (without the electronic contribution) are taken from literature[36].

example, corrected in an additional step by single static DFT calculations at the respective compositions.

Active learning applied to the TaVCrW HEA

Figure 6 compares the learning curves obtained by training the MTP and the GM-NN models on randomly and actively selected data. Figure 6a, b shows the RMS and MAX errors in energies. It can be seen that the MTP (orange lines) needs significantly fewer data to reach its final accuracy in the energies. The GM-NN model (blue

lines), on the other hand, shows a continuous (nearly linear) decrease in the error on the energies as the size of the training set increases. The final accuracy of the two models is similar. The reason for the faster convergence of the MTP can be attributed to the difference in the inherent nature of the two ML-based models—the relatively simpler formalism of the MTP with ~500 times fewer trainable parameters favoring quicker convergence *versus* the enhanced non-linear NN formalism of the GM-NN model making it continuously more accurate with increasing training data.

Indeed, the MTP reaches convergence in the energy already with a seemingly very small training set size of 128 structures. However, one should note that each structure contains several tens to hundreds of atoms with many distinct local chemical environments. The total number of atoms in each training set is tabulated in Table 2. For example, the MTP trained on 128-atom structures constitutes around 11,000 atoms. Hence, even with 128 structures, the MTP is trained on a sufficiently large number of atomic neighbourhoods, allowing it to make accurate predictions on every part of the configurational space.

Comparison of the dashed and solid lines in Fig. 6a, b reveals that there is no improvement in the RMS errors in energies using the active learning approach for neither of the two ML models. Since the random data selection was made across the entire diverse configurational space (with data chosen from all 0 K subsystems and MD), it is already sufficient to reach good energy accuracy for the Ta-V-Cr-W alloy family. An advanced active learning approach is superfluous for obtaining good energies.

Figure 6c, d show the errors in forces with increasing training set size. Once again, the MTP converges faster to its minimum RMS error in the forces in comparison to the GM-NN which has a more gradual decrease. The reason for this is the same as above, i.e., the simpler formalism of the MTP. However, unlike in the energies, here we notice a difference in convergence in the RMS error between the random and actively learnt MTP (dashed and solid orange lines respectively). The MTP trained on actively learnt training sets converges faster than the MTP trained on randomly selected training data. There is no such clear difference seen in the GM-NN curves (dashed and solid blue lines). To explain the improvement of
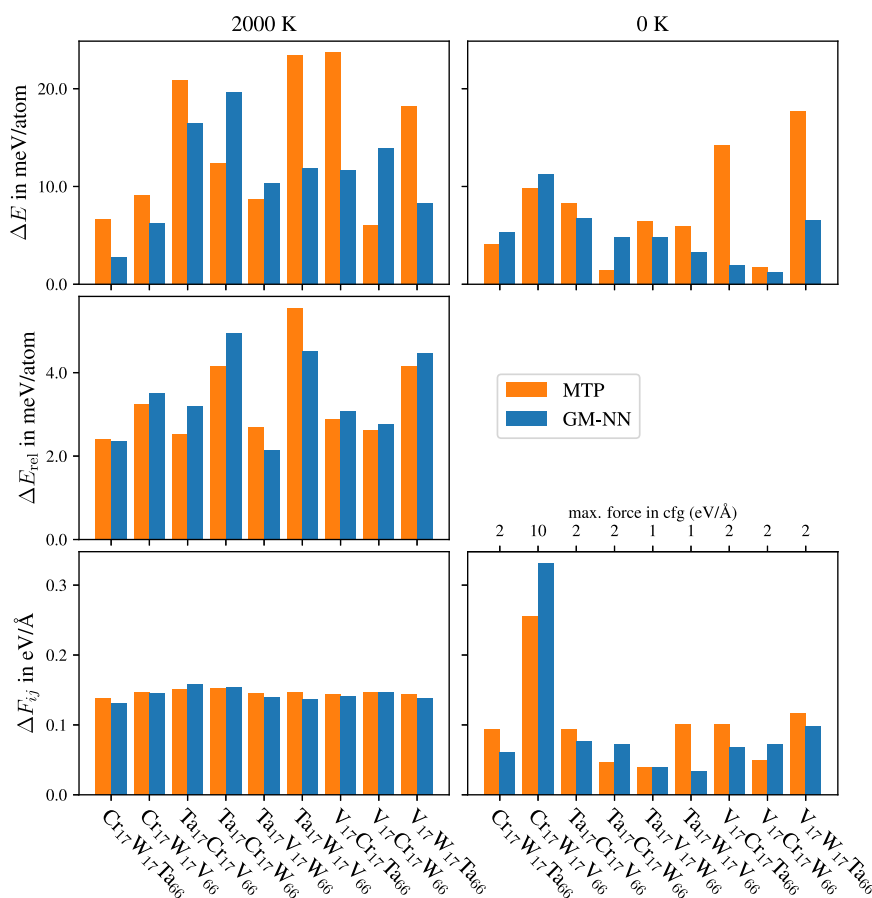
Fig. 5 Errors in predicted energies and atomic forces for finite-temperature (2000 K) and 0 K out-of-distribution non-equiatomic ternaries. For the 2000 K data, twelve structures have been sampled for each system. Thus, ΔE and ΔF_{ij} denote the respective RMS errors. For the relative energy, we used the first structure in each subsystem as a reference. Therefore, RMS errors ΔE_{rel} are evaluated for the remaining eleven structures. For the 0 K test, we employed a single configuration such that ΔE denotes the mean-absolute error, while ΔF_{ij} denotes the RMS error.

active learning over a random selection for the MTP and the absence of such an improvement for the GM-NN, we track the number of 2500 K configurations in the actively learnt training set.

Table 2 shows the fraction of 2500 K configurations, given in brackets, at different stages of the MTP and GM-NN active learning process. The active learning scheme of the MTP selects a higher fraction of 2500 K configurations in comparison to the GM-NN model, especially as the training set size increases. For example, for the MTP trained on 256 structures, 34% of the training set comprises of 2500 K configurations. This is reflected in the considerably better performance of the actively learnt MTP in predicting atomic forces at small training set sizes. In contrast, the active learning approach employed by GM-NN selects roughly the same amount of 2500 K data across all active learning steps, but a much smaller fraction than the corresponding MTP scheme. The GM-NN method constrains representativity of the selected batch, i.e. takes data distribution into account, and almost 80% of the total training set comprises of 0 K data. Thus, there is only a very small decrease in the RMS error in forces using the GM-NN model coming from the actively learnt training set. In conclusion, active learning might be beneficial as long as multiple structures are provided before labelling, i.e., computing DFT energies and forces. One can only take advantage of active learning in combination with a conventional human-based training set generation.

Training and inference times
The time taken to train an ML-based model and to perform MD calculations with it depends on the number of model parameters.

In the case of the GM-NN potential, this number is dictated by the neural network representation, whereas for an MTP the contractions are decisive (check 'Methods' section for details). To elucidate the training and inference times, we compare five GM-NNs and five MTPs built with different numbers of model parameters. The GM-NNs are designated as: 593-512-512-1, 593-128-128-1, 593-32-32-1, 360-512-512-1, and 910-512-512-1. Here the first number represents the number of input features, the following two numbers represent the number of hidden neurons, while the 1 in the end stands for the single output neuron (which corresponds to the energy of the system). For the MTP, we compare potentials with levels 16, 20, 24, 26, and 28 (see Equation (6)) resulting in 92, 288, 864, 1464, and 2445 descriptors, respectively.

Table 3 lists the training times for the different MTP and GM-NN models fitted to the in-distribution subsystems. Naturally, the time taken to train the MTP increases with the level as the number of parameters increase. Similarly, the time taken to fit a GM-NN potential increases with the number of input features, and with the number of neurons. The lev24 MTP—which is used eventually for the thermodynamic properties—takes eight hours to reach convergence in 400 iterations of the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm[48]. The corresponding GM-NN model takes four hours for 1000 epochs. In addition, in every epoch, the GM-NN evaluates 500 other configurations to avoid overfitting.

Figure 7 shows the inference time of the ML-models in comparison to the classical potentials and in relation to the achievable accuracy. In general, the MTPs (orange circles) are faster than the GM-NN potentials (blue triangles). However, the
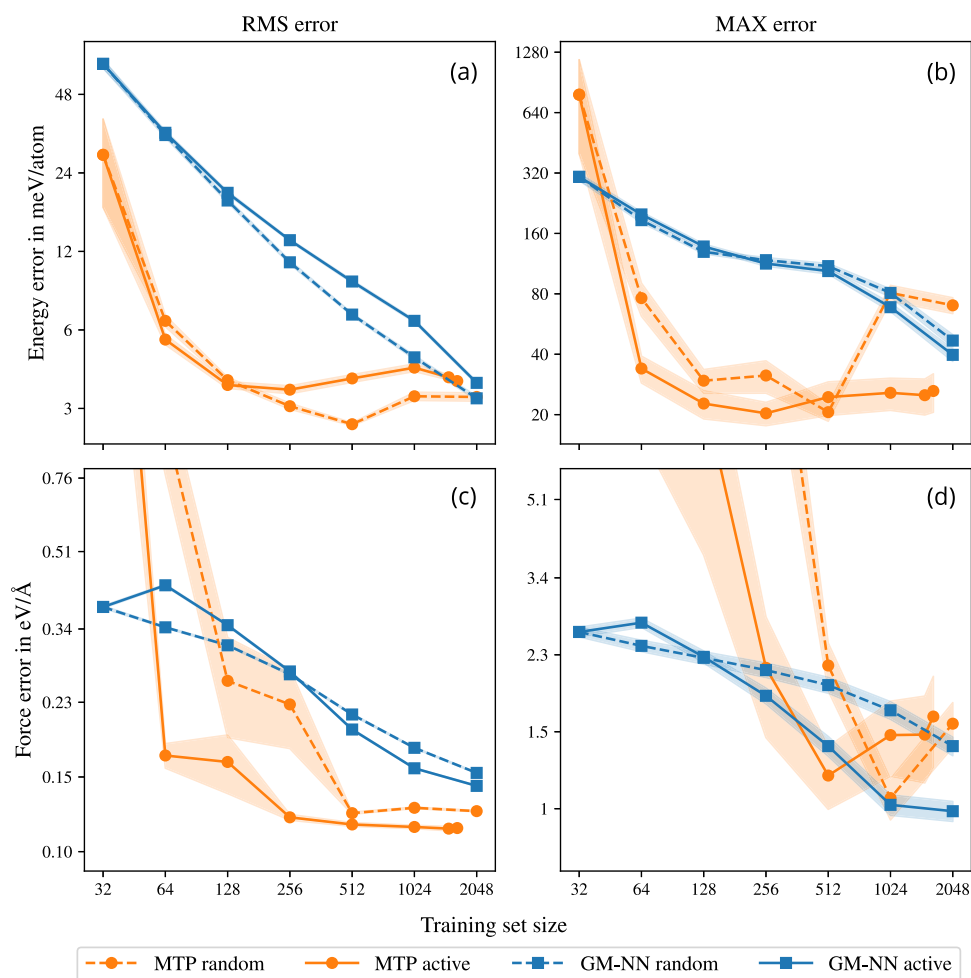
Fig. 6 Learning curves for MTP and GM-NN models. Shaded areas denote the standard error on the mean evaluated over 10 or 100 independent runs for MTP and GM-NN, respectively. An additional factor of ten appears for GM-NN as for each of the ten data splits, another ten splits for early stopping have been performed. The root-mean-squared (RMS) and maximal (MAX) errors for energies are shown in (**a**) and (**b**) respectively in meV/atom; The RMS and MAX errors for atomic force are shown in (**c**) and (**d**) respectively in eV/Å.

Table 2. The average number of atoms in the differently sized training sets during active learning (N_{train} = number of training structures).

N_{train}	MTP	GM-NN
64	5845 (0.21)	3101 (0.17)
128	10784 (0.18)	6427 (0.16)
256	20322 (0.34)	12222 (0.17)
512	45371 (0.53)	23601 (0.19)
1024	99549 (0.66)	46249 (0.19)

The fraction of 2500 K MD structures contained in the resulting training set are shown in brackets.

GM-NNs have been designed to run on graphical processing units (GPUs), which makes them faster (blue stars) than the higher-level MTPs. More specifically, the MTP used in this work for the HEA modelling—the lev24 MTP—is faster by a factor of 2.5 than the corresponding GM-NN—(593-512-512-1)—on the same CPU. However, on an NVIDIA GeForce RTX 3090 Ti 12GB GPU, the GM-NN is 2.5 times faster than the lev24 MTP.

It can also be observed that the inference time of an MTP gradually increases as the level of the MTP increases, although the accuracy (normalized error) saturates to a minimum after lev24. On the other hand, all GM-NN models are comparably

computationally efficient on a GPU, and only minor differences in inference times have been observed on CPU nodes. The accuracy of GM-NN deteriorates when decreasing the number of hidden neurons or input features. For example, decreasing the number of neurons by a factor of 16 leads to similar results as decreasing the number of features by a factor of 1.6. Thus, the employed representation affects the final result more than the network size.

Obviously, the ML-based models are slower than conventional potentials such as the EAM and the MEAM model. For example, the lev24 MTP is slower by a factor of 35 in comparison to the EAM. However, the speed of the classical potentials is accompanied by an extreme inaccuracy compared to the ML potentials, at least an order of magnitude in the normalized error.

DISCUSSION

We have developed two ML-based interatomic potentials—an MTP and a GM-NN—that can accurately capture the atomic interactions of the Ta-V-Cr-W alloy family. This is the first instance of exploring the limits of applicability of ML-based models to an extent where both the configurational and the vibrational degrees of freedom are included in the training and testing data sets for a chemically complex system. Specifically, the MTP and the GM-NN are able to extremely accurately reproduce energies and forces of both the 0 K subsystems, and the high-temperature (2500 K)

Table 3. Training times for both ML-based models evaluated for different numbers of model parameters on two Intel Xeon E6252 Gold (Cascade Lake) CPUs.
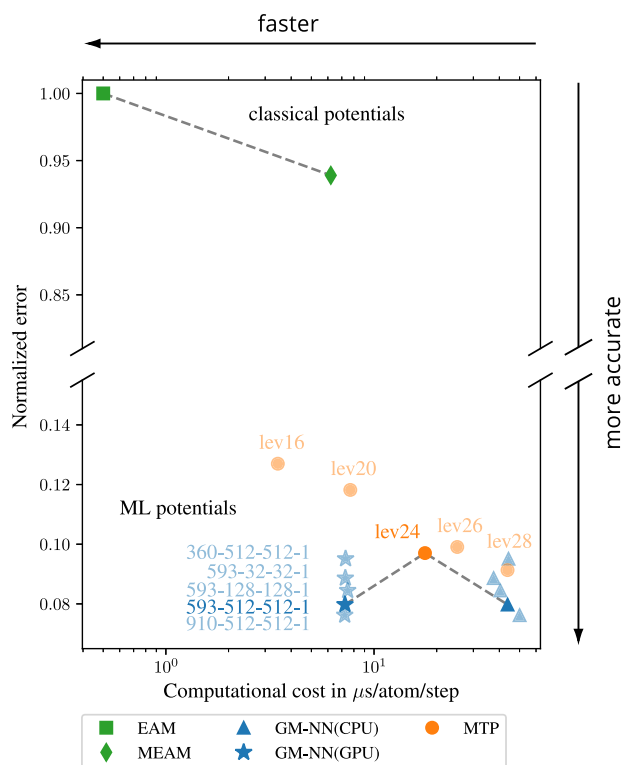
Model	No. parameters	No. cfgs	No. epochs/iterations	Time (h)
MTP Models				
lev16	608	5373	400 iterations	0.9
lev20	932	5373	400 iterations	2.1
lev24	**1636**	**5373**	**400 iterations**	**7.1**
lev26	2236	5373	400 iterations	9.2
lev28	3217	5373	400 iterations	10.8
GM-NN Models				
593-32-32-1	20032	4873	1000 epochs	2.8
593-128-128-1	92416	4873	1000 epochs	3.0
593-512-512-1	**566272**	**4873**	**1000 epochs**	**3.4**
360-512-512-1	446976	4873	1000 epochs	3.0
593-512-512-1	**566272**	**4873**	**1000 epochs**	**3.4**
910-512-512-1	728576	4873	1000 epochs	3.9

The bold font emphasizes the default potentials used throughout this work.

Fig. 7 The trade-off between computational cost and normalized error (Equation (13)) evaluated for MTP, GM-NN, EAM, and MEAM. For the normalized error, the overall errors from Table 1 have been chosen. The darkened colour represents the default MTP and GM-NN models used throughout this work. Blurred colours show models with varying hidden neurons and/or input features. For MTP, the respective level defines the number of employed input features; see Equation (6). For GM-NN, we use a convention where the first number indicates the number of features, while all following numbers define the number of neurons in the respective layer (two hidden and one output layer).

disordered equiatomic TaVCrW HEA, which make up the in-distribution sets. They are also able to accurately predict the energies of the 0 K strained binaries. The accuracy achieved in the present work is very similar to the MTP errors obtained in ref. [36] for the same disordered alloy, although their MTP was trained only on the equiatomic 2500 K structures. Adding the configurational degrees of freedom by introducing many 0 K structures in the training data set does not affect the predictive capabilities of the MTP (nor the GM-NN) on the HEA behaviour, suggesting that there is scope for further expanding the training set.

The models have also been tested on out-of-distribution subsystems (not a part of the training set) comprising of non-equiatomic ternaries. Both models successfully provide a similar performance for the out-of-distribution subsystems as compared to the in-distribution subsystems in the prediction of high-temperature MD forces. The observed RMS errors in absolute energies are larger than for the in-distribution cases, arising from the inability to predict accurate absolute chemical potentials. Nevertheless, the relative energies relevant for calculating temperature-dependent thermodynamic properties are still predicted with a low RMS error. These results indicate a good generalization of the MTP and GM-NN models even to out-of-distribution data.

In an earlier work [18], lattice-based Monte Carlo and cluster expansion techniques were used to predict chemical SRO and phase separation below 1300 K in the TaVCrW alloy. With the newly developed ML-based potentials, we can now model off-lattice behaviour that enables us to capture strong lattice relaxations. In phase-separated structures, lattice mismatch at the interface can lead to strained subsystems, which are likewise well-predicted by the MTP and GM-NN. Consequently, the low-temperature phase-separated structures in combination with strained interfaces, the intermediate SRO structures, as well as the high-temperature fully disordered structure can now be simultaneously and more accurately modelled with a single potential. The here-developed ML models also offer scope to combine MD with Monte Carlo methods to further enhance the sampling of phase spaces of chemically complex alloys. The vibrational degrees of freedom that the ML-based models entail enable, for example, a more accurate prediction of the order-disorder transition temperature. A well-determined transition temperature is crucial in deciding the

operational temperature of the HEA since ordering and phase segregation can lead to changes in mechanical behaviour [15].

The robustness of the ML-based models is also noticeable in the heat capacity curves shown in Fig. 4. Even small differences in the free energies are known to cause large deviations in the computed heat capacity [49]. Nevertheless, the MTPs and GM-NNs are reasonably accurate (< few meV/atom) in the free energy prediction that they are able to predict the heat capacities to DFT accuracy (without electronic contributions) up to around 2200 K.

Even though the ML-based models are remarkable in their free energy and the eventual thermodynamic property prediction, one needs to keep in mind they have been fitted to DFT data that only contain the vibrational and configurational degrees of freedom. To reach 'full DFT' accuracy, one also needs to include the electronic degrees of freedom (electronic free energy and the impact of vibrations on it) and the magnetic degrees of freedom. The state-of-the-art method to include the electronic contribution uses the direct upsampling technique [49], wherein we can perform DFT calculations (including electronic contributions) on snapshots generated by the ML-based models. The number of snapshots needed to reach a certain accuracy in the free energy can be roughly estimated based on the RMS errors in our models. In the current case, for an RMS error of 3 meV/atom, we can reach to within ± 1 meV/atom accuracy using approximately 35 snapshots [49] for a given volume and temperature. By doing so, we can include the temperature-dependent (and vibration-coupled) electronic degree of freedom into the thermodynamics of the alloy. To

account for the magnetic free energy, there have been propositions put forth in the literature, such as the magnetic MTP[50], that can be used for magnetic systems. In consequence, on top of the predictions of the currently fitted ML-based models, further expanding the configurational space of the training set (e.g., also including temperature-dependent behaviour of binaries and ternaries) and adding the electronic and magnetic degrees of freedom would lead to a more complete and accurate picture of the behaviour of chemically complex multicomponent alloys.

For a complex system such as the Ta-V-Cr-W family, active learning is only partially beneficial. The RMS errors in predicted energies converge similarly irrespective of an active or random selection of configurations. This is true for both ML models. The random selection done uniformly across the diverse set of subsystems is already sufficient to reach good accuracy in the energies. The advantage of active learning is more evident in the convergence of the atomic forces, arising likely from the data set size and complexity of the atomic-force landscape compared to the energies. The active learning algorithm employed by MTP preferably selects more high-temperature configurations which have large force values. In contrast, the algorithm employed by GM-NN has been designed to select new data considering the data distribution. Thus, the GM-NN scheme selects almost equal portions of MD data in every iteration.

The convergence of the current models with respect to the training set size is also the result of the number of trainable parameters. The MTP has ~ 500 times fewer fitting parameters than the GM-NN model. Hence, it reaches its final accuracy faster while showing no improvement when increasing the training set size above a certain amount. On the other hand, the GM-NN model shows gradual yet continuous improvement in accuracy with the training set size, owing to a larger number of trainable parameters. This fact allows us to assume that the GM-NN would outperform the MTP on larger data set sizes. However, the accuracy of the MTP on larger data sets can be improved by increasing the level of the selected MTP model, i.e., by increasing the effective number of trainable parameters.

Ultimately, in complex systems such as the TaVCrW HEA, which require large simulation cells to allow for events such as SRO to occur, the choice of the interatomic model is an interplay between the accuracy and the inference (simulation) time. Although the ML-based models are one to two orders slower than classical potentials (yet negligibly small in comparison to ab initio simulation times), they are at least an order of magnitude more accurate. Classical empirical potentials such as the EAM/MEAM are restricted in their functional form. Although such potentials need a smaller training set leading to less severe phase space sampling problems, their inaccuracy far outweighs their partial benefits. The inaccuracy of classical models might not only affect their robustness and stability while running MD across the volume and temperature range but would also highly alter, for example, the order-disorder transition temperature, making them unsuitable for such applications. Hence, for predicting challenging multicomponent system behaviour, one needs to resort to ML-based models.

In summary, despite certain prominent differences between MTP and GM-NN, they provide similar results, evaluated on a test data set or during a real-time simulation. Considering that the two models are cutting-edge representatives of the ML-potential family, these results may reflect the constantly growing boundaries of the machine-learning potentials applied to current material systems of interest.

METHODS

Local interatomic potentials

An atomic structure is defined by $S = \{\mathbf{r}_i, Z_i\}_{i=1}^{N_{at}}$, where N_{at} is the number of atoms, $\mathbf{r}_i \in \mathbb{R}^3$ denotes the Cartesian coordinates of

atom i, and $Z_i \in \mathbb{N}$ is the respective atomic number. We consider the mapping f from the atomic structure to a scalar electronic energy E, i.e., $f : S \mapsto E \in \mathbb{R}$. All models employed here assume locality of interatomic interactions (Fig. 1a), thus the total energy of the system is split into contributions from individual atoms

$$E = \sum_{i=1}^{N_{at}} E_i. \tag{1}$$

The energy contributed by atom i is defined by a parameterized function, i.e., $E_i = E_i(S_i, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ are trainable parameters and S_i is the local atomic environment of atom i. The parameters $\boldsymbol{\theta}$ are learned from the training data by minimizing a predefined loss function $L(\boldsymbol{\theta})$. The data set typically contains total energies, atomic forces, and stresses obtained by an ab initio method. The local environment, $S_i = \{\mathbf{r}_i, Z_i, \{\mathbf{r}_j, Z_j\}_{j \in R_{cut}}\}_{i=1}^{N_{at}}$, is defined by the chosen cutoff radius R_{cut}. In this work, we choose a cutoff radius of 5.0 Å.

Moment tensor potential—MTP

The MTP site energy for each atom i is linearly expanded through a set of *basis functions* B_α:

$$E_i = \sum_{\alpha=1}^{\alpha_{max}} \xi_\alpha B_\alpha^{(i)}, \tag{2}$$

which are contractions (e.g., dot products of vectors and matrices, see below) of the *moment tensor descriptors* $M_{\mu,\nu}$[33]. The ξ_α are expansion coefficients and the parameter α_{max} is determined from Equation (6) below. Each descriptor $M_{\mu,\nu}$ is composed of a radial part (scalar function depending only on the interatomic distances) and an angular part (a tensor of a certain rank) (Fig. 1b),

$$M_{\mu,\nu}(\mathfrak{n}_i) = \sum_j \overbrace{f_\mu(|\mathbf{r}_{ij}|, Z_i, Z_j)}^{\text{radial}} \underbrace{\overbrace{\mathbf{r}_{ij} \otimes \dots \otimes \mathbf{r}_{ij}}^{\text{angular}}}_{\nu \text{ times}}, \tag{3}$$

where $\mathfrak{n}_i$ stands for the neighbourhood of the atom i (all atoms within the radius R_{cut} around atom i), μ stands for the number of the radial functions, ν is the rank of the tensor in the angular part, and $\mathbf{r}_{ij}$ the vector between atom i and j. The radial functions $f_\mu(|\mathbf{r}_{ij}|, Z_i, Z_j)$ are expanded in a basis of Chebyshev polynomials, usually up to rank seven,

$$f_\mu(|\mathbf{r}_{ij}|, Z_i, Z_j) = \sum_k c_{\mu,Z_i,Z_j}^k Q^k(|\mathbf{r}_{ij}|), \tag{4}$$

where c_{μ,Z_i,Z_j}^k are expansion coefficients and

$$Q^k(\mathbf{r}) = T_k(\mathbf{r})(R_{cut} - \mathbf{r})^2. \tag{5}$$

Here $T_k(\mathbf{r})$ are the Chebyshev polynomials of order k and the $(R_{cut} - \mathbf{r})^2$ term is introduced to ensure a smooth decreasing to 0 at R_{cut}.

Although the moment tensor descriptors themselves are two-body functions (depending only on $\mathbf{r}_{ij}$), many-body interactions get included into the basis functions B_α by building various contractions of the $M_{\mu,\nu}$, under the condition that each contraction has to yield a scalar value. Note that the coefficients c_{μ,Z_i,Z_j}^k enter Equation (2) non-linearly, as the basis functions B_α contain multiplications of different radial functions, see examples below. This way, the MTP fitting parameters $\boldsymbol{\theta} = \{\xi_\alpha, c_{\mu,Z_i,Z_j}^k\}$ are composed of the coefficients ξ_α and c_{μ,Z_i,Z_j}^k, which enter Equation (2) linearly and non-linearly, respectively.

For the purpose of choosing which (out of the infinite number of) basis functions to include in the interatomic potential, we define a degree-like measure, '*level*', of $M_{\mu,\nu}$ defined by $\text{lev} M_{\mu,\nu} = 2\mu + \nu$ and the level of B_α obtained by contracting M_{μ_1,ν_1}, $M_{\mu_2,\nu_2}, \dots$, as

$$\text{lev} B_\alpha = (2\mu_1 + \nu_1) + (2\mu_2 + \nu_2) + \dots. \tag{6}$$

Thus, to define an MTP we choose a particular $\text{lev}_{\max}$ and expand the site energy function linearly through such B_α, which satisfy $\text{lev}B_\alpha \leq \text{lev}_{\max}$. The MTP part of Fig. 1c (left-hand side) has examples of how some of the basis functions are constructed:

- For the 2-body case, the scalar moments $M_{0,0}$, $M_{1,0}$, and $M_{2,0}$ already serve as basis functions, which differ by the radial functions involved.
- For the 3-body case, the moments $M_{0,1}$ and $M_{1,1}$ produce basis functions $M_{0,1} \cdot M_{0,1}$ and $M_{1,1} \cdot M_{1,1}$, respectively, which differ by the radial functions involved.
- For the 5-body case (the most complex case for MTP in this work), the basis function has the form $M_{0,1} \cdot M_{0,2} \cdot M_{0,2} \cdot M_{0,1}$.

Note how the condition specified in Equation (6) reduces the number of different radial functions involved in the many-body moments (Fig. 1c left). By increasing $\text{lev}_{\max}$, MTPs with more parameters emerge, which are capable of more accurately fitting or handling a larger training set, but which require more computational time for fitting and inference (simulation time). In this work, we use an MTP with $\text{lev}_{\max} = 24$.

The parameters of an MTP are obtained as those minimizing the mean squared loss on the training data (Fig. 1d). We use the second-order quasi-Newton BFGS[48] optimization method, which requires exact first derivatives of the loss function with respect to the model parameters and approximates the second derivatives by constructing the so-called *hessian matrix*. To balance the relative contributions of energies, forces, and stresses for a configuration j with $N_{\text{at}}^{(j)}$ atoms, weight factors of $C_\text{e} = 1/N_{\text{at}}^{(j)}$, $C_\text{f} = 0.001$ Å^2, and $C_\text{s} = 0.0001/N_{\text{at}}^{(j)}$ are used, respectively.

Gaussian moment neural network—GM-NN

GM-NN uses an artificial NN to map an atomic structure S to the total energy $E(S; \boldsymbol{\theta})$. Specifically, GM-NN employs a fully-connected feed-forward NN consisting of two hidden layers[34,35]

$$
\begin{aligned}
y_i = {} & 0.1 \cdot \mathbf{b}^{(3)} + \frac{1}{\sqrt{d_2}}\mathbf{W}^{(3)}\phi\Big(0.1 \cdot \mathbf{b}^{(2)} \\
& + \frac{1}{\sqrt{d_1}}\mathbf{W}^{(2)}\phi\Big(0.1 \cdot \mathbf{b}^{(1)} + \frac{1}{\sqrt{d_0}}\mathbf{W}^{(1)}\mathbf{G}_i\Big)\Big),
\end{aligned}
\tag{7}
$$

with $\mathbf{W}^{(l+1)} \in \mathbb{R}^{d_{l+1} \times d_l}$ and $\mathbf{b}^{(l+1)} \in \mathbb{R}^{d_{l+1}}$ being the weights and biases of the layer $l+1$, respectively. The default network in this work consists of $d_0 = 593$ input neurons (dimension of the feature vector $\mathbf{G}_i = \mathbf{G}_i(S_i)$), $d_1 = d_2 = 512$ hidden neurons, and a single output neuron $d_3 = 1$. The weights of the network $\mathbf{W}^{(l+1)}$ are initialized by picking the respective entries from a normal distribution with zero mean and unit variance. In contrast, the trainable bias vectors $\mathbf{b}^{(l+1)}$ are initialized to zero. To improve the accuracy and convergence of the GM-NN model, we employ a neural tangent parameterization (factors of 0.1 and $1/\sqrt{d_l}$)[47]. For the activation function ϕ, we use the Swish/SiLU function[51,52].

The network's output is scaled and shifted to aid the training process even further,

$$
E_i(\mathbf{G}_i, \boldsymbol{\theta}) = c \cdot (\sigma_{Z_i} y_i + \mu_{Z_i}),
\tag{8}
$$

where the trainable shift parameters μ_{Z_i} are initialized by solving a linear regression problem, and the trainable scale parameters σ_{Z_i} are initialized to 1. The constant c is given by the RMS error per atom of the regression solution[35].

To encode the invariance of the total energy to translations, rotations, and permutations of the same species, we employ the Gaussian moment (GM) representation[34]. The construction of GM features is based mainly on defining pairwise distance vectors $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ and splitting them into their radial and angular components: $r_{ij} = \|\mathbf{r}_{ij}\|_2$ and $\hat{\mathbf{r}}_{ij} = \mathbf{r}_{ij}/r_{ij}$. Thus, the equation from Fig. 1b takes a slightly different form as the L-fold outer product is

computed for the normalized distances,

$$
\Psi_{i,L,s} = \sum_{j \neq i} R_{Z_i, Z_j, s}(r_{ij}, \boldsymbol{\beta}) \hat{\mathbf{r}}_{ij}^{\otimes L},
\tag{9}
$$

where $\hat{\mathbf{r}}_{ij}^{\otimes L} = \hat{\mathbf{r}}_{ij} \otimes \cdots \otimes \hat{\mathbf{r}}_{ij}$. The non-linear radial functions $R_{Z_i, Z_j, s}(r_{ij}, \boldsymbol{\beta})$ are defined as a finite sum of Gaussians $\Phi_{s'}(r_{ij})$ ($N_{\text{Gauss}} = 8$ for this work)[35]

$$
R_{Z_i, Z_j, s}(r_{ij}, \boldsymbol{\beta}) = \frac{1}{\sqrt{N_{\text{Gauss}}}} \sum_{s'=1}^{N_{\text{Gauss}}} \beta_{Z_i, Z_j, s, s'} \Phi_{s'}(r_{ij}).
\tag{10}
$$

The factor $1/\sqrt{N_{\text{Gauss}}}$ influences the effective learning rate and is inspired by the neural tangent parameterization[47]. The individual radial functions are centred at an equidistant grid of points between $R_{\min} = 1.5$ Å (specific to this work) and R_{cut}, and are re-scaled by a cosine cutoff function[29]. The chemical information is encoded in the GM representation by trainable parameters $\beta_{Z_i, Z_j, s, s'}$. The index s runs over the number of independent radial basis functions, set to $N_{\text{basis}} = 6$ for this work.

The features invariant to rotations $\mathbf{G}_i$ (Fig. 1c right) are obtained by computing contractions of the respective tensors, as any full contraction of Cartesian tensors $\hat{\mathbf{r}}_{ij}^{\otimes L}$ yields a rotationally invariant scalar[34]. For the employed GM-NN model, contractions of up to three tensors of the maximum rank of three are used. Thus, the many-body expansion is restricted to the maximum of four-body terms, e.g.,

$$
G_{i,s_1,s_2,s_3} = (\Psi_{i,1,s_1})_a (\Psi_{i,1,s_2})_b (\Psi_{i,2,s_3})_{a,b},
\tag{11}
$$

where we use Einstein notation, i.e., the right-hand sides are summed over $a, b \in \{1, 2, 3\}$. Particular contractions are defined by identifying unique generating graphs[43]. In a practical implementation all GM features are computed at once providing tensors $\mathbf{G}^{(j)} \in \{\mathbb{R}^{N_{\text{basis}}}, \mathbb{R}^{N_{\text{basis}} \times N_{\text{basis}}}, \dots\}$. However, the number of invariant features used as input to NNs is reduced by considering the permutational symmetries of the respective graphs (see Fig. 1c right)[34], i.e., only tensors of the same rank can be exchanged.

The parameters $\boldsymbol{\theta} = \{\mathbf{W}, \mathbf{b}, \boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\mu}\}$ of the NN, i.e., weights $\mathbf{W}$ and biases $\mathbf{b}$, as well as the trainable parameters $\boldsymbol{\beta}$ of the local representation and the parameters that scale, $\boldsymbol{\sigma}$, and shift, $\boldsymbol{\mu}$, the output of the NN, are optimized by minimizing the mean squared loss on training data (Fig. 1d). We employ the Adam optimizer[53] to minimize the loss function. The respective parameters of the optimizer are $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-7}$. We use a mini-batch of 32 molecules. The layer-wise learning rates are set to 0.0075 for the parameters of the fully connected layers, 0.005 for the trainable representation, as well as 0.0125 and 0.00025 for the shift and scale parameters of atomic energies, respectively. The training is performed for 1000 training epochs. To balance the relative contributions of energies, forces, and stresses, weight factors of $C_\text{e} = 1/N_{\text{at}}^{(j)}$, $C_\text{f} = 0.005$ Å^2, and $C_\text{s} = 0.0001/N_{\text{at}}^{(j)}$ are used, respectively. To prevent overfitting during training, we employ the early stopping technique[54]. All models are trained within the TensorFlow framework[55].

Relation of MTP and GM-NN to other ML potentials

Interestingly, the MTP and GM representations can be related to many other approaches of modelling interactions proposed in the literature by considering a general function, e.g., ref. [56]

$$
\sum_{j,k,l} f(r_{ij}) f(r_{ik}) f(r_{il}) \langle \hat{\mathbf{r}}_{ij} \hat{\mathbf{r}}_{ik} \rangle \langle \hat{\mathbf{r}}_{ik} \hat{\mathbf{r}}_{il} \rangle.
\tag{12}
$$

This expression describes a four-body interaction but scales cubically with the number of atoms within R_{cut}. It is reminiscent of, e.g., atom-centred symmetry functions[57]. However, by writing the definition of scalar products $\langle \hat{\mathbf{r}}_{ij} \hat{\mathbf{r}}_{ik} \rangle$ explicitly and pulling out the sum with respect to spatial coordinates we arrive at expressions proposed in ref. [33] and ref. [34], which scale linearly with the number of atoms within R_{cut}. Specifically, for GM-NN the expression is

obtained in Equation ([11]), while for MTP the analogue is contracting the tensor moments from Equation ([3]), each of which scales linearly with the number of atoms. Since a Cartesian tensor of rank L can be written as a linear combination of spherical harmonics up to angular momentum L[58], both MTP and GM-NN representations (which use Cartesian tensors) are closely related to linear ACE[59] and message-passing NequIP and MACE models[45,60], which use spherical harmonics. Note that for message-passing models, the resulting many-body order and the effective cutoff radius increase with the increasing number of interaction blocks. While it often leads to a better performance in terms of error metrics such as RMS errors, it may lead to larger inference times and, thus, less efficient potentials.

The recently proposed extension of HIP-NN models[61] with tensor sensitivity information is closely related to MTP and GM-NN potentials, as it employs irreducible Cartesian tensors to pass tensor information between atomic sites[62]. Another example of local representations based on spherical harmonics is SOAP or the respective GAP model[63]. Unlike the above-mentioned models, it employs Gaussian process regression to map the structure to the respective energy. A more detailed overview of existing interatomic ML potentials is out of the scope of this work, and the reader is referred elsewhere[32,64,65].

EAM and MEAM potentials

Classical interatomic potentials based on the embedded-atom method (EAM)[20] and the modified embedded-atom method (MEAM)[21] are well understood in the literature and are not discussed here. We just draw attention to their relation to the ML-based models. EAM can be viewed as an MTP potential with one radial function per species pair and without the angular part, i.e., $lev_{max} = 0$, see Equation ([6]). In turn, MEAM would be close to an MTP with $lev_{max} = 2$ (see Equation ([6])), allowing for different radial functions, and a vectorial angular part (rank 1 tensor).

In the main text, we compare the performance of the ML-based models to an EAM that is fitted to the same training data. In addition, in the Supplementary Information, we also compare to a reference-free MEAM in order to test whether there is any improvement coming from the angular three-body terms. The EAM and reference-free MEAM fitting is performed with MEAM-fit[66,67]. Fitting weights for the energies, atomic forces and stresses are $1/N_{at}^{(j)}$, 0.1, and 0.001, respectively.

Data set generation - DFT calculations

All DFT calculations are performed using VASP[68,69], with projector augmented wave (PAW) potentials[70], within the generalized gradient approximation (GGA)[71–73], and in the spin-unpolarized state. We use Methfessel-Paxton smearing of order 1 with $\sigma = 0.1$, a 350 eV energy cutoff and an automatic k-mesh generation with KSPACING = 0.12. The number of valence electrons is 11 for Ta and V, and 12 for Cr and W.

Description of the data set

The training and test data set includes energies, atomic forces and stresses from the DFT calculations. For the 0 K configurations with small supercells (item 1 below), we perform structural relaxation with DFT. For the 0 K configurations with large supercells (item 2 below), we have a combination of relaxed and unrelaxed structures. All configurations sampled from MD (item 3 below) are unrelaxed. We generate 2-, 3-, and 4-component BCC configurations in the following categories:

1. 0 K configurations in small supercells (2–8 atoms):
    These configurations are sampled with the enumlib library[74,75], which enumerates all possible symmetries to obtain non-repetitive structures. We first pick 2500 binary

structures from enumlib, representing different atomic combinations of each of the TaV, TaCr, TaW, VCr, VW, and CrW subsystems on the BCC lattice. To obtain the 4-component structures, we take 8-atom binaries from above and substitute the elements giving 248 equiatomic quarternary structures.

2. 0 K configurations in medium to large supercells (16, 128, 432, 8192 atoms):
    These configurations are designed to mimic the possible low-temperature ordering in TaVCrW. For each of these sizes, we create certain types of binaries and quaternaries. Binary structures for each of TaV, TaCr, TaW, VCr, VW, CrW subsystems include:

    - B2 ordering
    - B32 ordering
    - random binary solid solution
    - BCC interface: half of the cubic supercell is one unary, and the other half is the other unary

    Quaternary structures represent all possible phase separations on the TaVCrW BCC lattice, i.e., half of the cubic supercell is one binary system (e.g., TaV) and the other half is another binary system (e.g., CrW), and they include:

    - B2/B2 ordering
    - B2/B32 ordering
    - B32/B32 ordering
    - B2/random binary ordering
    - B32/random binary ordering
    - random binary/random binary ordering

3. High-temperature (2500 K) disordered structures (128 and 432 atom supercells):
    We use the MTP from ref. [36] to run NPT MD at 2500 K. From the obtained MD trajectories, we sample 983 and 48 structures with 128 and 432 atoms in the supercell, respectively. Sampling is performed by employing active learning designed for MTP models[76,77], with MTP potentials initialized by training on 0 K data. Thus, the selected data may introduce some bias towards MTP performance. However, a more thorough investigation may be required to prove this hypothesis and is postponed to future works. The energies, forces and stresses of the sampled structures are calculated using DFT without relaxation.

4. Deformed structures (432 atoms):
    Equiatomic TaVCrW can phase segregate at low temperatures into: Ta+Cr/V+W (predicted by MTP in this work), Ta+W/Cr+V (predicted in ref. [18]), and Ta+V/Cr+W. Each binary part undergoes certain strain in the [100] direction to accommodate the same cross-section between the two parts. Hence we also create deformed binaries as a part of the test data set (whose RMS errors are in Table [1]) to accurately replicate such a phase-segregated B2/B2 quaternary structure. The strains applied along the [100] direction corresponding to the different binaries are: TaV 3.7%, CrW −1.28%, TaCr 0.27%, VW −0.14%, TaW 4.06% and VCr −4.71%.

In total, 6711 configurations are computed with DFT for usage as training and validation sets. More precisely, there are 5680 0 K structures: 4491 binary, 595 ternary, and 594 quaternary structures, along with 1031 structures sampled from MD at 2500 K. The number of binary and ternary structures is more or less equally distributed between the subsystems, and the difference comes from the non-convergence of some DFT calculations for some subsystems.

Testing the models

To assess the accuracy of the models, we compute the RMS error in energies and atomic forces on every Ta-V-Cr-W in-distribution

subsystem, including the TaVCrW alloy at 0 K and the equiatomic TaVCrW disordered alloy sampled at 2500 K. The RMS errors are obtained by the 10-fold cross-validation method. From each subsystem and disordered alloy, 80% of the respective data set is selected for training and 20% for testing the resulting models. The selected training data is gathered into a final training set consisting of 5373 configurations. Then the potentials are tested on the selected test data set including 1336 configurations. We perform ten such instances with different combinations of the (80%, 20%) splitting to obtain a reasonable statistical averaging. The RMS errors presented in Table 1 and Fig. 2 are averaged over the 10 independent splits. We also train a single potential for each MTP/GM-NN/EAM/MEAM model to the entire data set (100%) to calculate the heat capacity of the disordered alloy.

While the above strategy applies to the in-distribution data, for the out-of-distribution subsystems (Fig. 5), we generate nine random ternary $A_{17}B_{17}C_{66}$ structures with 54 atoms, for each possible non-equiatomic ternary except for the Cr-rich ones to avoid magnetism-related issues. For the 0 K case we take one sample per composition, while for the MD case we sample 12 structures per composition using NPT MD at 2000 K.

It should be mentioned here that the NN-based models are over-parametrized given a larger number of parameters than MTPs. The possibility of overfitting is accounted for by, e.g., employing a separate holdout or a validation data set and the early stopping technique[54]. For the GM-NN, we take 500 structures from the resulting training data sets as validation data sets, i.e., the GM-NNs are trained on fewer structures compared to the other potentials. We make ten independent splits for each training set to avoid any bias in the final models. Thus, the RMS errors for the GM-NNs are averaged over 100 runs.

The normalized error (NE) in Fig. 7 is calculated as

$$\text{NE} = \frac{1}{2}\left(\frac{F - \text{RMSE}}{\max F - \text{RMSE}} + \frac{E - \text{RMSE}}{\max E - \text{RMSE}}\right), \tag{13}$$

where E-RMSE and F-RMSE are overall energy and force errors, respectively, taken from Table 1.

Heat capacity simulations

The heat capacity of the TaVCrW HEA up to 2500 K is numerically calculated as the second derivative of the Gibbs free energy with respect to temperature. The Gibbs free energy is calculated by a Legendre transformation of the Helmholtz energy. The Helmholtz energy up to the level of the interatomic model is calculated using thermodynamic integration using Langevin dynamics (TILD). For a set of volume and temperature (V, T) points, we run TILD on a 128-atom special quasi-random structure (SQS) of the TaVCrW alloy. TILD is performed from a reference effective quasiharmonic model[78,79] (fitted to 2500 K MD forces) to the interatomic potential to obtain the free energy difference. By summing the effective quasiharmonic reference energy with the free energy difference, we obtain the full Helmholtz energy for that (V, T) point. After doing the same for the different (V, T) points, we parametrize the Helmholtz energy surface using a fourth-order polynomial in V and T, on which the transformation and numerical derivatives are performed. Further details on the TILD methodology can be found in ref. [36].

DATA AVAILABILITY

All data sets used in this study can be accessed via the link: https://doi.org/10.18419/darus-3516.

CODE AVAILABILITY

MTPs can be used as a part of the MLIP package accessible through mlip.skoltech.ru/download/. The GM-NN software employed in this work is implemented within the

TensorFlow framework[55] and is published at gitlab.com/zaverkin_v/gmnn, including examples of usage.

REFERENCES

1. Ikeda, Y., Grabowski, B. & Körmann, F. Ab initio phase stabilities and mechanical properties of multicomponent alloys: a comprehensive review for high entropy alloys and compositionally complex alloys. *Mater. Charact.* **147**, 464–511 (2019).
2. George, E. P., Curtin, W. & Tasan, C. C. High entropy alloys: a focused review of mechanical properties and deformation mechanisms. *Acta Mater.* **188**, 435–474 (2020).
3. George, E. P., Raabe, D. & Ritchie, R. O. High-entropy alloys. *Nat. Rev. Mater.* **4**, 515–534 (2019).
4. Li, Z., Körmann, F., Grabowski, B., Neugebauer, J. & Raabe, D. Ab initio assisted design of quinary dual-phase high-entropy alloys with transformation-induced plasticity. *Acta Mater.* **136**, 262–270 (2017).
5. Feng, R. et al. Phase stability and transformation in a light-weight high-entropy alloy. *Acta Mater.* **146**, 280–293 (2018).
6. Yang, Y.-C., Liu, C., Lin, C.-Y. & Xia, Z. Core effect of local atomic configuration and design principles in Al_xCoCrFeNi high-entropy alloys. *Scr. Mater.* **178**, 181–186 (2020).
7. Yin, B. & Curtin, W. Origin of high strength in the CoCrFeNiPd high-entropy alloy. *Mater. Res. Lett.* **8**, 209–215 (2020).
8. Zherebtsov, S., Yurchenko, N., Panina, E., Tikhonovsky, M. & Stepanov, N. Gum-like mechanical behavior of a partially ordered $Al_5Nb_{24}Ti_{40}V_5Zr_{26}$ high entropy alloy. *Intermetallics* **116**, 106652 (2020).
9. Oh, H. S. et al. Lattice distortions in the FeCoNiCrMn high entropy alloy studied by theory and experiment. *Entropy* **18**, 321 (2016).
10. Zaddach, A., Niu, C., Koch, C. & Irving, D. Mechanical properties and stacking fault energies of NiFeCrCoMn high-entropy alloy. *JOM* **65**, 1780–1789 (2013).
11. Laube, S. et al. Controlling crystallographic ordering in Mo–Cr–Ti–Al high entropy alloys to enhance ductility. *J. Alloys Compd.* **823**, 153805 (2020).
12. Bhandari, U., Zhang, C., Zeng, C., Guo, S. & Yang, S. Computational and experimental investigation of refractory high entropy alloy $Mo_{15}Nb_{20}Re_{15}Ta_{30}W_{20}$. *J. Mater. Res. Technol.* **9**, 8929–8936 (2020).
13. Chen, S. et al. Simultaneously enhancing the ultimate strength and ductility of high-entropy alloys via short-range ordering. *Nat. Commun.* **12**, 1–11 (2021).
14. Rogal, L. et al. Computationally-driven engineering of sublattice ordering in a hexagonal AlHfScTiZr high entropy alloy. *Sci. Rep.* **7**, 1–14 (2017).
15. Wu, Y. et al. Short-range ordering and its effects on mechanical properties of high-entropy alloys. *J. Mater. Sci. Technol.* **62**, 214–220 (2021).
16. Ma, Y. et al. Chemical short-range orders and the induced structural transition in high-entropy alloys. *Scr. Mater.* **144**, 64–68 (2018).
17. Huang, S. The chemical ordering and elasticity in FeCoNiAl$_{1-x}$Ti$_x$ high-entropy alloys. *Scr. Mater.* **168**, 5–9 (2019).
18. Sobieraj, D. et al. Chemical short-range order in derivative Cr–Ta–Ti–V–W high entropy alloys from the first-principles thermodynamic study. *Phys. Chem. Chem. Phys.* **22**, 23929–23951 (2020).
19. El-Atwani, O. et al. Outstanding radiation resistance of tungsten-based high-entropy alloys. *Sci. Adv.* **5**, eaav2002 (2019).
20. Daw, M. S. & Baskes, M. I. Embedded-atom method: derivation and application to impurities, surfaces, and other defects in metals. *Phys. Rev. B* **29**, 6443–6453 (1984).
21. Kim, Y.-M., Lee, B.-J. & Baskes, M. I. Modified embedded-atom method interatomic potentials for Ti and Zr. *Phys. Rev. B* **74**, 014101–014112 (2006).
22. Gastegger, M., Behler, J. & Marquetand, P. Machine learning molecular dynamics for the simulation of infrared spectra. *Chem. Sci.* **8**, 6924–6935 (2017).
23. Gubaev, K., Podryabinkin, E. V. & Shapeev, A. V. Machine learning of molecular properties: locality and active learning. *J. Chem. Phys.* **148**, 241727 (2018).
24. Janet, J. P., Duan, C., Yang, T., Nandy, A. & Kulik, H. J. A quantitative uncertainty metric controls error in neural network-driven chemical discovery. *Chem. Sci.* **10**, 7913–7922 (2019).
25. Vandermause, J. et al. On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events. *NPJ Comput. Mater.* **6**, 20 (2020).
26. Zaverkin, V. & Kästner, J. Exploration of transferable and uniformly accurate neural network interatomic potentials using optimal experimental design. *Mach. Learn.: Sci. Technol.* **2**, 035009 (2021).
27. Zaverkin, V., Holzmüller, D., Steinwart, I. & Kästner, J. Exploring chemical and conformational spaces by batch mode deep active learning. *Digit. Discov.* https://doi.org/10.1039/D2DD00034B (2022).

28. Holzmüller, D., Zaverkin, V., Kästner, J. & Steinwart, I. A framework and benchmark for deep batch active learning for regression. *J. Mach. Learn. Res.* **24**, 1–81 (2023).

29. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).

30. Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).

31. Thompson, A., Swiler, L., Trott, C., Foiles, S. & Tucker, G. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *J. Comput. Phys.* **285**, 316–330 (2015).

32. Musil, F. et al. Physics-inspired structural representations for molecules and materials. *Chem. Rev.* **121**, 9759–9815 (2021).

33. Shapeev, A. Moment tensor potentials: a class of systematically improvable interatomic potentials. *Multiscale Model. Simul.* **14**, 1153–1173 (2016).

34. Zaverkin, V. & Kästner, J. Gaussian moments as physically inspired molecular descriptors for accurate and scalable machine learning potentials. *J. Chem. Theory Comput.* **16**, 5410–5421 (2020).

35. Zaverkin, V., Holzmüller, D., Steinwart, I. & Kästner, J. Fast and sample-efficient interatomic neural network potentials for molecules and materials based on Gaussian moments. *J. Chem. Theory Comput.* **17**, 6658–6670 (2021).

36. Zhou, Y. et al. Thermodynamics up to the melting point in a TaVCrW high entropy alloy: systematic ab initio study aided by machine learning potentials. *Phys. Rev. B* **105**, 214302 (2022).

37. Gubaev, K. et al. Finite-temperature interplay of structural stability, chemical complexity, and elastic properties of bcc multicomponent alloys from ab initio trained machine-learning potentials. *Phys. Rev. Mater.* **5**, 073801 (2021).

38. Molpeceres, G., Zaverkin, V. & Kästner, J. Neural-network assisted study of nitrogen atom dynamics on amorphous solid water - I. adsorption and desorption. *Mon. Not. R. Astron. Soc.* **499**, 1373–1384 (2020).

39. Molpeceres, G., Zaverkin, V., Watanabe, N. & Kästner, J. Binding energies and sticking coefficients of H_2 on crystalline and amorphous CO ice. *Astron. Astrophys.* **648**, A84 (2021).

40. Zaverkin, V., Molpeceres, G. & Kästner, J. Neural-network assisted study of nitrogen atom dynamics on amorphous solid water - II. Diffusion. *Mon. Not. R. Astron. Soc.* **510**, 3063–3070 (2022).

41. Zaverkin, V., Netz, J., Zills, F., Köhn, A. & Kästner, J. Thermally averaged magnetic anisotropy tensors via machine learning based on Gaussian moments. *J. Chem. Theory Comput.* **18**, 1–12 (2022).

42. Boys, S. F. & Egerton, A. C. Electronic wave functions - I. A general method of calculation for the stationary states of any molecular system. *Proc. R. Soc. Lond. A* **200**, 542–554 (1950).

43. Suk, T. & Flusser, J. in *Computer Analysis of Images and Patterns* (eds. Real, P., Diaz-Pernil, D., Molina-Abril, H., Berciano, A. & Kropatsch, W.) Vol. 6855, 212–219 (Springer, 2011).

44. Schütt, K. T., Unke, O. T. & Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. *ICML* 1–13, 9377–9388 (2021).

45. Batzner, S. et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).

46. Shapeev, A., Gubaev, K., Tsymbalov, E. & Podryabinkin, E. *Active Learning and Uncertainty Estimation* 309–329 (Springer International Publishing, 2020).

47. Jacot, A., Gabriel, F. & Hongler, C. in *NeurIPS* (eds. Bengio, S. et al.) Vol. 31, 8580–8589 (Curran Associates, Inc., 2018).

48. Liu, D. C. & Nocedal, J. On the limited memory BFGS method for large scale optimization. *Math. Program.* **45**, 503–528 (1989).

49. Hyun Jung, J., Srinivasan, P., Forslund, A. & Grabowski, B. High-accuracy thermodynamic properties to the melting point from *ab initio* calculations aided by machine-learning potentials. *Npj Comput. Mater.* (2022, submitted).

50. Novikov, I., Grabowski, B., Körmann, F. & Shapeev, A. Magnetic Moment Tensor Potentials for collinear spin-polarized materials reproduce different magnetic states of bcc Fe. *Npj Comput. Mater.* **8**, 13 (2022).

51. Elfwing, S., Uchibe, E. & Doya, K. Sigmoid-Weighted linear units for neural network function approximation in reinforcement learning. *Neural Netw.* **107**, 3–11 (2018).

52. Ramachandran, P., Zoph, B. & Le, Q. V. Searching for activation functions. Preprint at http://arXiv:1710.05941 (2017).

53. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at http://arxiv.org/abs/1412.6980 (2015).

54. Prechelt, L. *Early Stopping—But When?* 53–67 (Springer, 2012).

55. Abadi, M. et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. https://www.tensorflow.org/ (2015).

56. Zaverkin, V. *Investigation of Chemical Reactivity by Machine-learning Techniques*. Ph.D. thesis, University of Stuttgart. https://doi.org/10.18419/opus-12182 (2022).

57. Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **134**, 074106 (2011).

58. Drautz, R. Atomic cluster expansion of scalar, vectorial, and tensorial properties including magnetism and charge transfer. *Phys. Rev. B* **102**, 024104 (2020).

59. Drautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys. Rev. B* **99**, 014104 (2019).

60. Batatia, I., Kovacs, D. P., Simm, G. N. C., Ortner, C. & Csanyi, G. in *Advances in Neural Information Processing Systems* (eds. Oh, A. H., Agarwal, A., Belgrave, D. & Cho, K.) https://openreview.net/forum?id=YPpSngE-ZU (2022).

61. Lubbers, N., Smith, J. S. & Barros, K. Hierarchical modeling of molecular energies using a deep neural network. *J. Chem. Phys.* **148**, 241715 (2018).

62. Chigaev, M. et al. Lightweight and effective tensor sensitivity for atomistic neural networks. *J. Chem. Phys.* **158**, 18 (2023).

63. Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).

64. Unke, O. T. et al. Machine learning force fields. *Chem. Rev.* **121**, 10142–10186 (2021). PMID: 33705118.

65. Langer, M. F., Goeßmann, A. & Rupp, M. Representations of molecules and materials for interpolation of quantum-mechanical simulations via machine learning. *Npj Comput. Mater.* **8**, 41 (2022).

66. Duff, A. I., Finnis, M., Maugis, P., Thijsse, B. J. & Sluiter, M. H. MEAMfit: a reference-free modified embedded atom method (RF-MEAM) energy and force-fitting code. *Comput. Phys. Commun.* **196**, 439–445 (2015).

67. Srinivasan, P. et al. The effectiveness of reference-free modified embedded atom method potentials demonstrated for NiTi and NbMoTaW. *Model. Simul. Mater. Sci. Eng.* **27**, 065013 (2019).

68. Kresse, G. & Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Phys. Rev. B* **49**, 14251–14269 (1994).

69. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169 (1996).

70. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).

71. Ceperley, D. M. & Alder, B. J. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.* **45**, 566–569 (1980).

72. Perdew, J. P. & Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **23**, 5048–5079 (1981).

73. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).

74. Hart, G. L. & Forcade, R. W. Algorithm for generating derivative structures. *Phys. Rev. B* **77**, 224115 (2008).

75. Morgan, W. S., Hart, G. L. & Forcade, R. W. Generating derivative superstructures for systems with high configurational freedom. *Comput. Mater. Sci.* **136**, 144–149 (2017).

76. Podryabinkin, E. V. & Shapeev, A. V. Active learning of linearly parametrized interatomic potentials. *Comput. Mater. Sci.* **140**, 171–180 (2017).

77. Novikov, I. S., Gubaev, K., Podryabinkin, E. V. & Shapeev, A. V. The MLIP package: moment tensor potentials with MPI and active learning. *Mach. Learn.: Sci. Technol.* **2**, 025002 (2020).

78. Hellman, O., Abrikosov, I. A. & Simak, S. I. Lattice dynamics of anharmonic solids from first principles. *Phys. Rev. B* **84**, 180301 (2011).

79. Hellman, O., Steneteg, P., Abrikosov, I. A. & Simak, S. I. Temperature dependent effective potential method for accurate free energy calculations of solids. *Phys. Rev. B* **87**, 104111 (2013).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

FUNDING

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-023-01073-w.

Correspondence and requests for materials should be addressed to Konstantin Gubaev or Viktor Zaverkin.

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

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