Computational Materials Science 200 (2021) 110752

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci

Exploring the necessary complexity of interatomic potentials

Joshua A. Vita, Dallas R. Trinkle *

Department of Materials Science and Engineering, University of Illinois Urbana-Champaign, Urbana, Illinois 61801, USA

ARTICLE INFO

Keywords:
Potential fitting
MEAM
Machine learning

ABSTRACT

The application of machine learning models and algorithms towards describing atomic interactions has been a major area of interest in materials simulations in recent years, as machine learning interatomic potentials (MLIPs) are seen as being more flexible and accurate than their classical potential counterparts. This increase in accuracy of MLIPs over classical potentials has come at the cost of significantly increased complexity, leading to higher computational costs and lower physical interpretability and spurring research into improving the speeds and interpretability of MLIPs. As an alternative, in this work we leverage "machine learning" fitting databases and advanced optimization algorithms to fit a class of spline-based classical potentials, showing that they can be systematically improved in order to achieve accuracies comparable to those of low-complexity MLIPs. These results demonstrate that high model complexities may not be strictly necessary in order to achieve near-DFT accuracy in interatomic potentials and suggest an alternative route towards sampling the high accuracy, low complexity region of model space by starting with forms that promote simpler and more interpretable interatomic potentials.

1. Introduction

For nearly a century of designing interatomic potentials for use in materials simulations, a strong emphasis was placed on constructing classical potentials with physically-motivated forms derived from quantum mechanical theories [1–9]. More recently, with the enormous success of machine learning in various fields, machine learning interatomic potentials (MLIPs) have come to dominate the attention of the computational materials science community (an incomplete list: [10–20]). MLIPs have been shown to be able to predict energies and forces on diverse ranges of atomic configurations with unprecedented accuracy, which in combination with active research into improving their speed and interpretability [21–23] makes them good candidates for being fast and accurate general-purpose models of atomic interactions.

Much of the increased performance of MLIPs over classical potentials comes from the flexibility in their functional forms that allows them to be systematically extended in order to be able to model increasingly diverse sets of atomic environments. However, pushing MLIPs to this limit of high accuracy and generalizability often results in highly complex models that are computationally expensive to use and conceptually difficult to interpret. Classical potentials, on the other hand, are by construction much simpler to interpret due to their basis in known physics and usually have much lower computational costs than even the simplest MLIPs. These two forms thus are generally used in opposing regions of "model space": classical potentials in the low complexity, low accuracy region; MLIPs in the high complexity, high accuracy region.

Sampling the space between these two regions (low complexity, high accuracy) is a fundamental goal of computational materials science, and in recent years has most often been approached by attempting to decrease the complexity of MLIPs. In this work we show that a family of spline-based classical potentials, when leveraging traditional "machine learning" databases and fitting algorithms, can be systematically improved in order to achieve accuracies on existing benchmark databases that push them into the low complexity, high accuracy region of model space alongside low-complexity MLIPs. Despite the lower interpretability of spline-based potentials relative to classical potentials with explicit analytical definitions, our results show that spline-based classical potentials offer a good balance between speed, interpretability, and accuracy. These results suggest that spline-based classical potentials may be good options alongside low-complexity MLIPs for designing practical and computationally tractable general-purpose potentials.

* Corresponding author.
E-mail addresses: jvita2@illinois.edu (J.A. Vita), dtrinkle@illinois.edu (D.R. Trinkle).

https://doi.org/10.1016/j.commatsci.2021.110752
Received 18 May 2021; Received in revised form 19 July 2021; Accepted 23 July 2021
Available online 19 August 2021

2. Background

2.1. Machine learning interatomic potentials

For this work, the most important distinctions between MLIPs and classical potentials are in their functional forms and in how they describe local atomic environments. The main advantage of MLIPs is that their functional forms are more easily extended to account for different environments by increasing their degrees of freedom, and that they are able to leverage more advanced descriptors of local atomic environments. The commonly-assumed implications of these differences are that (1) their extensible functional forms let MLIPs be more flexible and general than classical potentials, and (2) their use of better local descriptors allow MLIPs to predict atomic energies and forces more accurately.

While new MLIPs are constantly being developed, we use the results from [24], which focused on five of the more popular forms: Gaussian Approximation Potentials (GAP) [11], Moment Tensor Potentials (MTP) [16], Neural Network Potentials (NNP) [10], Spectral Neighbor Analysis Potentials (SNAP) [13], and a variant on SNAP that uses quadratic components (denoted qSNAP) [19]. Conceptually, each of these MLIPs operate by first encoding the local atomic environments of a structure into atomic "fingerprints" (descriptors), then passing those fingerprints through an embedding function. Each of the MLIPs differ in the types of fingerprints that they use—atom-centered symmetry functions [25] for NNP, "smooth overlap of atomic positions" kernels [26] and related descriptors for GAP and SNAP/qSNAP, and rotationally covariant tensors [16] for MTP—and in their embedding functions— neural networks for NNP, Gaussian process models for GAP, linear/quadratic models for SNAP/qSNAP, and summations of tensor contractions for MTP. The performances of the MLIPs can then be tuned by adjusting the number of fingerprints used or increasing the complexity of their embedding functions. In the interest of keeping this paper concise, a proper explanation of the details of the MLIPs [11,10,13,16,19] and their descriptors [25,26] is reserved for existing work in the literature.

2.2. s-MEAM

We compare the MLIP results from [24] with results obtained by fitting new spline-based modified embedded-atom method (s-MEAM) potentials, a variant of the analytical MEAM [4] that introduces some additional flexibility. The analytical MEAM is a popular potential form that has been widely applied to various metals and alloys, and to some covalently-bonded materials [4,27,28]. In the s-MEAM formalism, the energy of a system is written as:

$$E = \sum_{i<j} \phi_{ij}(r_{ij}) + \sum_i U_i(n_i) \tag{1}$$

$$n_i = \sum_{j\neq i} \rho_j(r_{ij}) + \sum_{\substack{j<k,\\ j,k\neq i}} f_j(r_{ij}) f_k(r_{ik}) g_{jk}\left(\cos(\theta_{jik})\right) \tag{2}$$

In Eq. (1) the total energy E is composed of a pair term (ϕ) and an embedding energy contribution (U) due to the electron density (n_i) around an atom. The electron density is further decomposed into 2-body (ρ) and 3-body (products of f and g) contributions. The total energy is computed by summing over the pair distances r_{ij} between each atom i and its neighbors j and the angles θ_{jik} and pair distances r_{ij} and r_{ik} defined by a triplet of atoms i, j, and k with i at the center. The subscripts on the functions indicate that different functions are used for evaluation depending on the bond types between atoms i, j, and k.

The s-MEAM formalism is very similar to the analytical MEAM, but differs in that s-MEAM removes the explicit analytical forms of the functions in a MEAM potential and replaces them with cubic splines (ϕ and U in Eq. (1); ρ, f, and g in Eq. (2)). s-MEAM was originally developed for elemental Si [29], but has since been applied to Nb, Mo, Ti, and Ti-O [30–34]. The use of splines theoretically gives s-MEAM better flexibility in reproducing interactions for complex atomic environments as well as better generalizability, but much like with MLIPs these added benefits come at the cost of decreased interpretability of s-MEAM relative to classical MEAM or other classical potentials. In particular, non-linearities in the spline shapes of the density terms (ρ, f, and g) and the embedding function (U) can theoretically lead to complex and uninterpretable behaviors. Additionally, the density values n_i lack any physically meaningful units and are perfectly capable (likely even) of sampling negative values.

In addition to the increased flexibility of s-MEAM (and despite its reduced interpretability) relative to analytical MEAM, s-MEAM has a couple of advantages that make it a particularly useful form to study in practice. First, s-MEAM encompasses an entire family of commonly-used classical potentials including the Lennard-Jones [1] (LJ), embedded-atom method [3] (EAM), modified embedded-atom method [4] (MEAM), Stillinger–Weber [6] (SW), and Tersoff [5] potentials, meaning it can be used to systematically study the impact of including/removing various terms in the functional form. And second, fitting software for spline-based potentials can take advantage of specialized data structures that make it extremely efficient to calculate energies and forces for large collections of different parameter sets simultaneously during fitting (for more details see Section 3.4).

An s-MEAM potential can be easily extended to multi-component systems by adding additional splines for each new element and the cross-terms between pairs of elements. For example, in a binary system with A and B elements there would be a total of twelve splines: ϕ_{AA}, ϕ_{AB}, ϕ_{BB}, ρ_A, ρ_B, U_A, U_B, f_A, f_B, g_{AA}, g_{AB}, and g_{BB}. Though single- and multi-component s-MEAM potentials have the same physical assumptions as each other, it is likely that multi-component systems will contain more complex local atomic environments which may not be able to be as accurately described using spline representations. Because of this, multi-component s-MEAM potentials may be more difficult to fit and may not be able to achieve as high accuracies as compared to their single-component counterparts.

2.3. s-MEAM and low-complexity MLIPs

The nesting of potentially highly non-linear functions in the embedding term of s-MEAM has the result of making the s-MEAM functional form very similar in theory to some low-complexity MLIPs. In particular, s-MEAM is comparable in its formalism to a single-layer NNP or a low-order MTP since each of these forms involve only limited numbers of radial and angular terms as their "fingerprints" and relatively simple embedding functions. However, it is important to note that the s-MEAM splines in practice often optimize to relatively interpretable forms (e.g. LJ-like shapes for ϕ, ρ, and f, and nearly linear embedding functions U) as shown in Fig. 8, making it easier to attempt to understand how/why fitted s-MEAM potentials behave the ways that they do in practice. Furthermore, model developers can easily enforce different conditions to encourage interpretability, for example by manually setting boundary conditions or knot values to require repulsive forces for short bond lengths.

3. Methods

3.1. Fitting databases

In order to ensure a fair comparison, we train all models on the databases produced in [24], which were specifically designed to encompass a large variety of atomic environments. In total, we explored six different elements (three crystal systems): Ni and Cu (FCC), Li and Mo (BCC), and Si and Ge (diamond). Each database contains the ground state structure for the given element, strained supercells, slab structures, *ab initio* molecular dynamics (AIMD) sampling of supercells at different temperatures (300 K and 0.5 × , 0.9 × , 1.5 × , and 2.0× the melting

point), and AIMD sampling of supercells with single vacancies at 300 K and 2.0× the melting point. We construct training/testing sets using a 90:10 split, sampled from each sub-group of the databases (strained, slab, MD, and vacancy configurations). On average, each training database includes approximately 250 structures. A detailed summary of the contents of the databases can be found in [24]. The training and testing databases were made publicly-available on Github [35] by the authors of [24].

3.2. Model hyper-parameters

When fitting s-MEAM potentials, there are two hyper-parameters that are typically considered: the number of spline knots in each spline, and the x-positions of each of those knots. However, since s-MEAM has a strictly defined functional form, these hyper-parameters cannot be continuously adjusted to further increase the accuracy of the potential, and they do not significantly impact the computational cost of the potential. In [24] an extensive study of the hyper-parameters of the MLIPs was performed because increasing the degrees of freedom in the MLIPs can theoretically continuously increase their fitting capabilities (at higher computational costs). In order to provide a more appropriate comparison between the s-MEAM family of potentials and the MLIPs discussed in [24], instead of showing the results of changing the number of knots per spline we choose to demonstrate the impact of including only subsets of the terms in s-MEAM (e.g. potentials that only have a ϕ term, or don't include the 3-body embedding terms f and g). This shows the effects of systematically increasing the flexibility of the s-MEAM family of potentials, similar to how increasing the number of free parameters in an MLIP extends its functional form. The optimal s-MEAM potentials (starred points in Fig. 7) were further optimized by constructing potentials with 5, 7, 11, and 15 knots per spline and choosing the one that performed the best on the given tests of material property predictions. The U splines were often restricted to having only three knots since they are significantly easier to optimize and experiments with higher numbers of knots typically optimized to nearly linear functions anyways.

We choose the domains of the radial functions based on the minimum pair distance sampled in each database and the cutoff distances used by the MLIPs. The cutoff distances of the s-MEAM potentials for each element are taken to be 3.9 Å (Ni), 4.0 Å (Cu), 5.1 Å (Li), 5.2 Å (Mo), 5.0 Å (Si), and 5.3 Å (Ge), which are chosen to be similar to the cutoffs used for the MLIPs in [24]. As discussed in [24], these cutoff distances are consistent with previous studies using classical potentials and MLIPs which find second nearest neighbor interactions to be sufficient for FCC elements, while third nearest neighbor interactions are necessary for BCC and diamond systems. The domains of the U and g splines are $[-1, 1]$ (with the exception of the Ge s-MEAM potential, which uses $[-0.5, 0.5]$ for the U domain because it happened to result in a better potential during optimization) since the n_i values can be arbitrarily scaled during fitting to fall into the desired range [36] and g's inputs are $\cos(\theta)$ values.

3.3. Optimization algorithm: CMA-ES

We optimize the y-positions of the knots in the splines of the s-MEAM potentials using the Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES) [37,38]. CMA-ES is a derivative-free, population-based global optimization algorithm that is designed to work with continuous-domain, non-convex, non-linear functions, and is highly scalable and parallelizable. This algorithm has been shown previously to work well for optimization problems with dimensionalities similar to ours [39] (in our case, 45 fitting parameters: five splines, each with seven knots and two boundary conditions), has been successfully applied to fitting s-MEAM potentials in the past [34], and is readily available in an existing Python package [40].

We attempt to find a balance between exploration (exploring parameter space) and exploitation (moving toward good solutions) by tuning various hyper-parameters of the CMA-ES. The CMA-ES works by iteratively updating the mean and covariance matrix of a multi-variate normal distribution; at each CMA-ES step we sampled 100–2000 parameter sets and used the best 50% for the update (influencing the algorithm's exploitation). We adjust the "width" of the distribution, which influences the algorithm's exploratory capabilities, using a scaling parameter for the covariance matrix, with typical values in the range of 1–10.

3.4. Software implementation details

As mentioned in Section 2.2, a major benefit of using spline-based potentials when fitting a model is that the fitting software can take advantage of special data structures, which we call "structure vectors" here, in order to be able to compute energies and forces of atomic configurations for large batches of parameter sets simultaneously and efficiently. Unlike molecular dynamics, where a large number of atomic configurations need to be considered in serial, optimization considers a fixed set of atomic configurations. The processing of neighbor lists for each structure can be done once, and then used to construct an efficient algorithm for the evaluation of a large number of parameters. The use of cubic splines, which are linear functions of the knot values, permits sums over interatomic distances to be efficiently precalculated into "structure vectors."

The evaluation of a Hermitian cubic spline $f(x)$ with n knots, which is normally written as a piece-wise function with each piece being a linear combination of four basis functions, can instead be re-written as $f(x) = \vec{s}(x)\cdot\vec{\theta}$. Here, the evaluation of $f(x)$ is the dot product between two length-$(n+2)$ vectors: a vector of coefficients, $\vec{s}(x)$, that depends only on the input value x, and a vector of fitting parameters $\vec{\theta}$ composed of the two spline boundary conditions and the y-positions of the n knots in the spline. This form is particularly useful since it is linear in $\vec{\theta}$, which means that summations over multiple x can be efficiently computed as $f(x_1) + f(x_2) = [\vec{s}(x_1) + \vec{s}(x_2)]\cdot\vec{\theta}$. Thus, arbitrarily many spline evaluations can be performed with a single length-$(n+2)$ vector-vector dot product by defining $\vec{S} = \sum_i \vec{s}(x_i)$ (which is what we call a structure vector) to be the summation over the vectors of spline coefficients corresponding to all of the x_i values, then writing the summation as $\sum_i f(x_i) = \vec{S}\cdot\vec{\theta}$. A full derivation of the structure vector is found in Section SI in the Supplementary Material.

For potential fitting, this means that the collection of $\vec{S}$ vectors for each spline involving atomic positions—ϕ, ρ, f, and g in s-MEAM—can be computed for each structure in the fitting database, saved, and reused at each optimization step rather than having to loop over full neighbor lists every single time. Due to the form of s-MEAM, the embedding energy $U(n_i)$ for each atom is evaluated using the standard algorithm for cubic splines. We find that the most time-consuming evaluation is the three-body product from f and g splines. Equally as important is the fact that $\vec{S}$ can be used to evaluate multiple parameter sets at the same time by converting $\vec{\theta}$ to instead be a matrix of parameter sets, with each column corresponding to a different parameterization of a potential; this use case is extremely common among global optimization algorithms, where objective function values need to be computed for batches of parameterizations at every step. For reference, on a single XE compute node of the Blue Waters supercomputer (two 8-core 2.3 GHz AMD 6276 Interlagos processors and 64 GB of memory), we are able to evaluate the energies and forces of a 108-atom cell for ten thousand different model parameterizations in approximately 3.4 seconds. In addition, the use of matrix–matrix products allows for efficient use of GPU acceleration, providing for additional hardware speedup. Readers interested in implementing structure vectors for different potential forms should see Section SI in the

Supplementary Material for further details, or refer to the Github repository associated with this paper [41].

4. Results and discussion

4.1. Fitting procedure

The following results were obtained by fitting s-MEAM potentials to a set of six benchmark databases published by [24] for Ni, Cu, Li, Mo, Si, and Ge. The databases were designed to cover a broad collection of atomic environments for each element, including ground state structures, strained configurations, surfaces, liquids, vacancies, and molecular dynamics snapshots. The potentials were fitted with software developed by the authors of this paper using the Covariance Matrix Adaptation Evolutionary Strategy [37,38], where the objective function was the mean absolute error computed by comparing the energies and forces predicted by the s-MEAM potentials to the values predicted by density functional theory (DFT). Optimal potentials were selected based on their ability to correctly reproduce the various material properties discussed in this section. Further details regarding the fitting databases and optimization procedures can be found in Section 3.

4.2. Training/testing errors

As shown in Fig. 1, for all elements the fitted s-MEAM potentials achieve train/test root-mean-square errors (RMSE) for both energies and forces that are similar in magnitude to the MLIPs despite the relatively large diversity of the training and testing sets. In most cases the s-MEAM potentials produce errors that are most similar to those of NNP and SNAP, typically making them only 2–3× the magnitudes of the errors for GAP and MTP. In the case of Ge, which has abnormally high energy errors relative to the GAP and MTP—but similar to other MLIPs—approximately 40% of the training and testing error comes from surface structures, indicating that s-MEAM is performing well for the majority of the Ge database, but failing at reproducing surface energies as accurately as other structures. This behavior is reflected in Fig. 2, which shows that the Ge s-MEAM potential has the highest surface energy errors of all the models.

Not only are the s-MEAM errors computed in this work comparable to the accuracies of the MLIPs, but they are also drastically lower than what is usually thought of as the errors for classical potentials. Zuo et al. [24] also reported the energy/force errors on these databases for other classical potentials available in literature [29,31,42–46], which were typically 10 to 100 times larger than the errors for the s-MEAM potentials and MLIPs shown here. However, in our work it is clear that the differences in errors between classical potentials and MLIPs trained on the same data can be as low as just a few meV·atom^{-1}, which is already nearing the accuracy of DFT [47]. The most likely reason that many of the potentials from the literature had such high energy/force errors on the databases from [24] was *not* because their functional forms are inherently less accurate than the MLIPs, but rather simply because the models weren't trained on databases that had as many different types of structures (in particular, liquids and surfaces may not have been included). Furthermore, a contributing factor to the perceived low accuracy of classical potentials is that many of the older literature potentials (LJ, EAM, and analytical MEAM in particular) are often constructed by fitting directly to material properties, rather than by using the "force-matching" method of Ercolessi-Adams [48] which is typically used for fitting MLIPs and the s-MEAM potentials in the literature.

4.3. Material properties

In Fig. 3 it is shown that in addition to having good energy/force errors, the s-MEAM potentials also perform comparably to the MLIPs in predicting material properties. For Ni, Cu, and Mo in particular the s-MEAM potentials show excellent agreement with the DFT-predicted values. Note that the vacancy migration energy E_m has the largest variation between models of any of the properties, and is the only property that was not explicitly included in the fitting databases. The large variation in the Li property predictions are partially due to the relatively small magnitudes of its cubic elastic constants, bulk modulus, and vacancy migration energy. Existing s-MEAM potentials from the literature for Si [29] and Mo [31] have been included for reference and show similar performance.

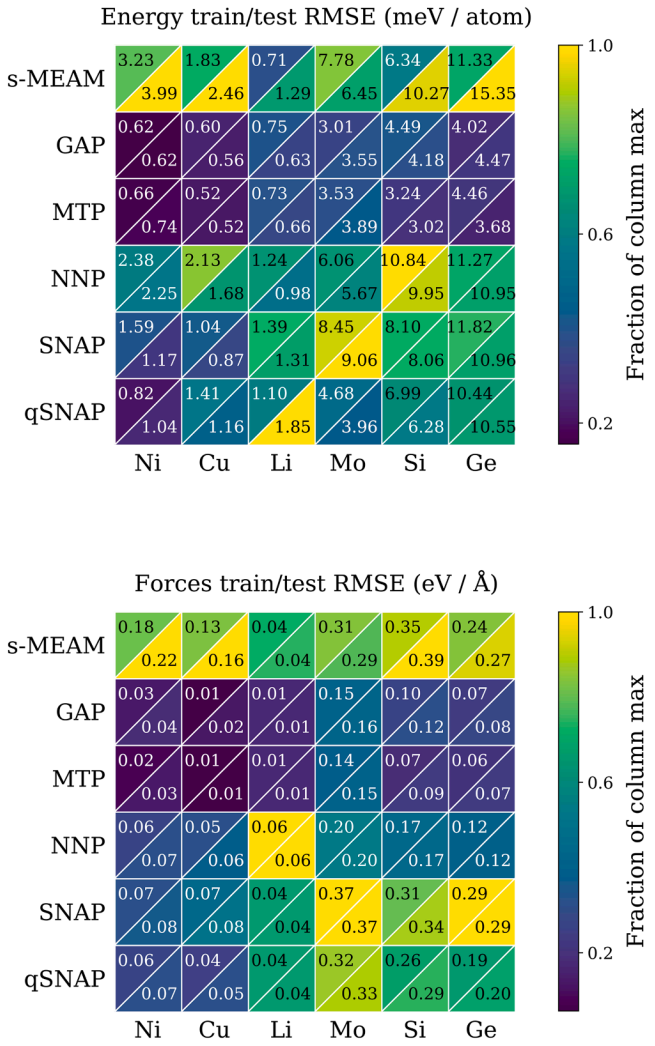

Fig. 1. Root-mean-square errors in predicted energies and forces for the s-MEAM potentials developed in this work and the MLIPs (GAP, MTP, NNP, SNAP, qSNAP) developed in [24]. Colors were scaled with respect to the maximum value in each column so that model performances can be more easily compared relative to each other for each element. Training/testing errors are shown in the upper/lower triangles respectively, where the testing databases represent a 10% split of the total data sampled from each group in the databases (MD samples of bulk solids and liquids, strained structures, vacancies, and surfaces). For all elements the s-MEAM potentials are able to reproduce the energies and forces in the databases with accuracies comparable to the optimal MLIPs.

In Fig. 4 we show the stacking fault energy (SFE) curves for four of the elements (Mo, Si, Ni, and Cu) that have been explored previously in the literature. MLIP curves were taken from [24], DFT curves were taken directly from the literature [31,49,50], and s-MEAM curves were computed in this work. Errors in unstable stacking fault energy predictions (peak heights) for the s-MEAM potentials are similar in magnitude to the MLIPs, but the s-MEAM potentials commonly have a

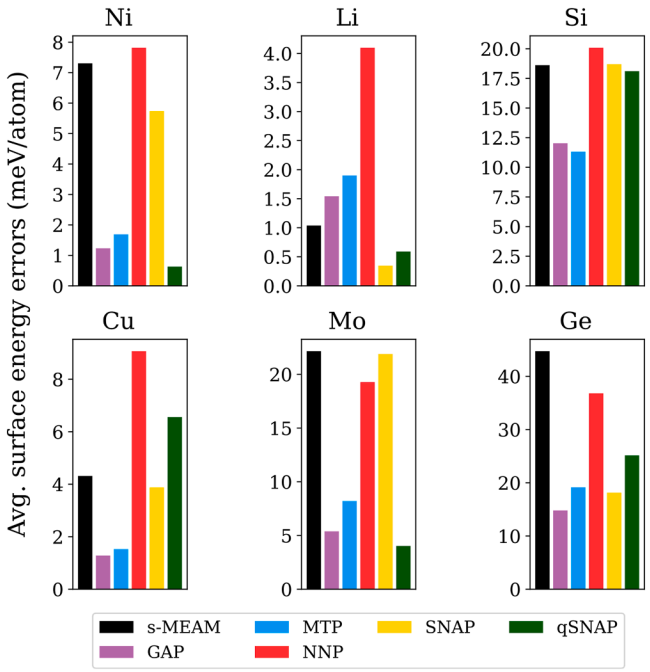

Fig. 2. Root-mean-square errors in predicted surface energies in the training set. For most elements the s-MEAM potentials maintain errors that are similar in magnitude to the errors of the MLIPs, but in the case of Ge the s-MEAM potential has particularly high errors that may indicate a lack of generalizability of the s-MEAM form. In general, the MTP and GAP have the lowest surface energy errors while s-MEAM and NNP have the highest. SNAP and qSNAP show the largest variation across different elements.

problem where the ground state structure is actually higher energy than the faulted configuration. In the case of elements with FCC ground state structures, the faulted configuration has a local structure that is similar to HCP, so inclusion of the HCP structure into the training database (which the databases in this work don't have) can help to address this issue.

Comparison of the minima-aligned energy versus volume equation of state curves shows that the s-MEAM potentials agree with DFT at non-equilibrium volumes to within the threshold that is used for comparing equivalent DFT codes. As mentioned in Fig. 5, the Δ_{EOS} gauge is the RMSE of the difference between two EOS curves within a range of $\pm 6\%$ of $V_{0,\text{DFT}}$, where values below 2 meV·atom^{-1} are considered to define "indistinguishable EOS curves" [51]. In all cases, the s-MEAM models agree with DFT to within the desired threshold. These results aren't surprising, since the Δ_{EOS} gauge depends largely on a model's ability to reproduce the bulk elastic modulus of the given systems, which most models do to within about 5% error relative to the DFT-predicted bulk moduli.

Energy differences between the ground state structure and the DFT-predicted low-energy polymorph structures are used as a test of the ability of the models to extrapolate to data that they weren't explicitly trained on (since only ground state crystals were included in the training databases). As seen in Fig. 6, for Ni, Li, and Ge the s-MEAM potentials agree with the DFT predictions to within ≈ 20 meV·atom^{-1}. The Cu s-MEAM errors are about 10 meV·atom^{-1} larger than the errors of the worst MLIP, but the Si s-MEAM severely under-predicts the polymorph energy. All of the Mo potentials have larger absolute errors due to the larger magnitude of the DFT energy differences, but the error of the s-MEAM potential is still within the error range of the MLIPs.

While classical potentials have already been shown to be able to predict properties of these material systems with good accuracy [29,31,42–46], the results here emphasize the fact that they can perform well even when trained on the larger and more diverse databases that

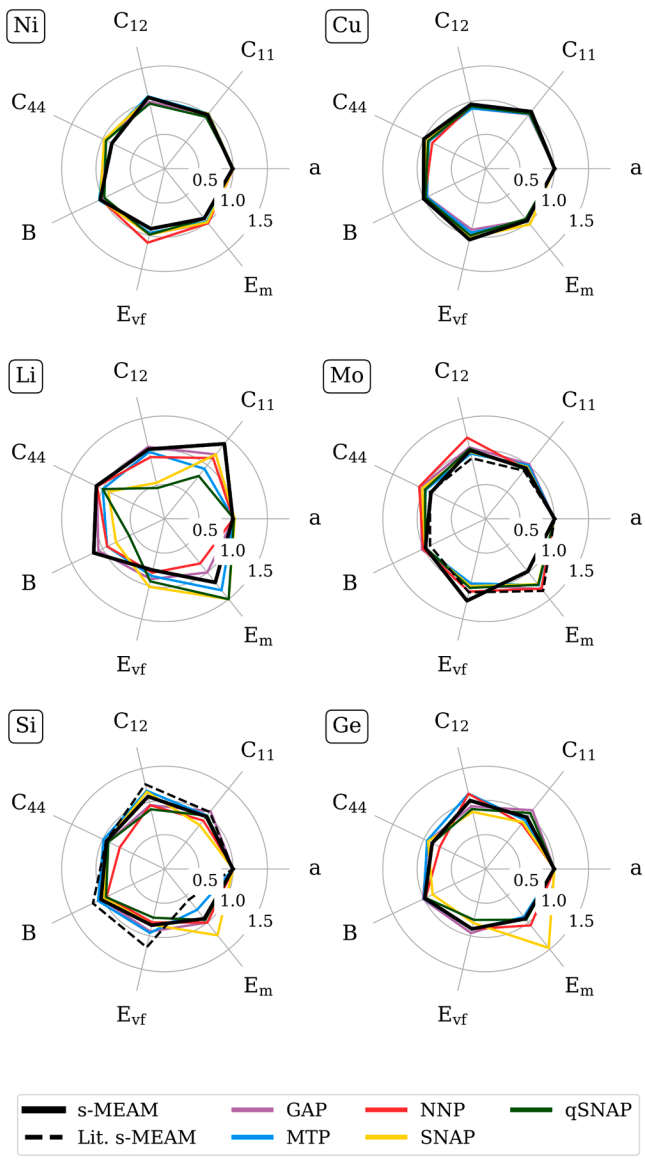

Fig. 3. Predictions of the lattice constant (a), cubic elastic constants (C_{11}, C_{12}, C_{44}), bulk modulus (B), and vacancy formation and migration energies (E_{vf} and E_m) for each model for each element. All properties have been normalized with respect to the DFT values, so an ideal potential would produce a regular heptagon whose vertices fall on the unit circle. The dashed black lines correspond to the predictions for previously-published s-MEAM potentials for Si [29] and Mo [31].

are usually reserved for MLIPs. However, one issue that became apparent when constructing the optimal potentials was that lower total RMSE values did not necessarily correspond to better property predictions. Potentials with training errors that were even just 1–2 meV·atom^{-1} lower than the optimal potentials shown here often began to see a trade-off where the percent errors in their property predictions began to be as large as 30%–40% for all properties except the lattice constant. A possible cause of this behavior could be that some of the databases contain certain types of structures with a frequency that is disproportionate to the relative importance of their related material properties. As an example, in this work all of the predictions were of bulk or defected solid properties, but the training databases also included large amounts of liquid and slab (surface) structures. When errors for each structure are equally weighted, an "imbalanced" database such as this could cause an insufficiently generalizable model to be fitted to structures that don't significantly contribute to the properties of interest.

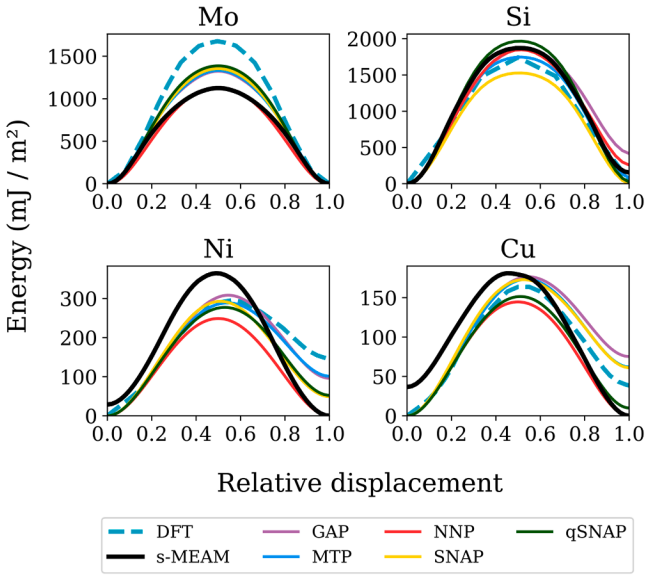

Fig. 4. Cross-sections of the relaxed γ surfaces for each model for a selection of elements that have been explored previously in literature. DFT curves were taken from previous works [31,49,50], while MLIP curves were provided by the authors of [24]. Displacements are in the (011) plane in the [111] direction for Mo, and in the (111) plane along the [112] direction for Si, Ni, and Cu.

There were many cases, especially when fitting the Li potentials, where the property predictions could be improved by allowing larger errors in reproducing the energies of the surface structures. This issue begins to highlight a limitation of the s-MEAM form, where although the average errors are nearing MLIP accuracy, it is still seeing a trade-off in errors that implies that it is not yet as generalizable as the MLIPs.

4.4. Accuracy/cost trade-off

Fig. 7 shows that for all elements, the optimal s-MEAM potentials consistently fall on or near the empirical accuracy/cost Pareto front. The optimal s-MEAM potentials are usually 1–2 orders of magnitude faster than the respective MLIPs while still having comparable testing errors. We plot testing set error in Fig. 7 since it is commonly used for evaluating "general-purpose" potentials (potentials that are designed for general use with a given chemical system, rather than for a single application of that system), and because it can detect possible overfitting of a model. Although we use cost per molecular dynamics step for this comparison, it is important to note that many other factors must be taken into consideration such as the software implementation of the model, the hardware on which it is being run, and the memory requirements of the model.

For the MLIPs, the authors of [24] modified the model costs and accuracy by increasing the degrees of freedom in each type of model. In the case of s-MEAM, we choose to explore the extensibility of the potential form by altering which of the terms (ϕ, ρ, U, f, and g) in Eqs. (1) and (2) to include in the potential. Note that although s-MEAM does have other hyper-parameters that can be adjusted (number of knots per spline, and x-positions of knots), they don't significantly impact the computational cost of the potential, therefore adding/removing terms from the s-MEAM form is the most appropriate method for analyzing the accuracy/complexity trade-off.

The consistent presence of the s-MEAM potentials and their variants on or near the Pareto front alongside the low-complexity MLIPs suggests that the s-MEAM family is an ideal candidate for beginning to sample the low complexity, high accuracy region of model space. However, the fact that the s-MEAM potentials occasionally appear to be out-performed by some of it's simpler variants (e.g. the Cu s-MEAM variant that doesn't include the 3-body f and g term) also helps to demonstrate a weakness of s-MEAM. In all cases where the s-MEAM potential doesn't fall on the Pareto front (Ni, Cu, and Ge), it had higher testing set errors than one of its variants, but was still chosen as the "optimal" potential because it had better overall material property predictions on the tests shown in Section 4.3. The ability of s-MEAM potentials and it's variants to lower their errors further than what is shown in Fig. 1 and Fig. 7 in exchange for worse property predictions was experienced ubiquitously throughout this work. The most likely explanation for this behavior is that achieving lower average errors on the database is forcing s-MEAM to allow larger errors on some of the structures that influence the material property predictions more heavily. This explanation suggests that the s-MEAM form is not yet as generalizable as some of the MLIPs (MTP and GAP in particular), and would most likely require the introduction of additional

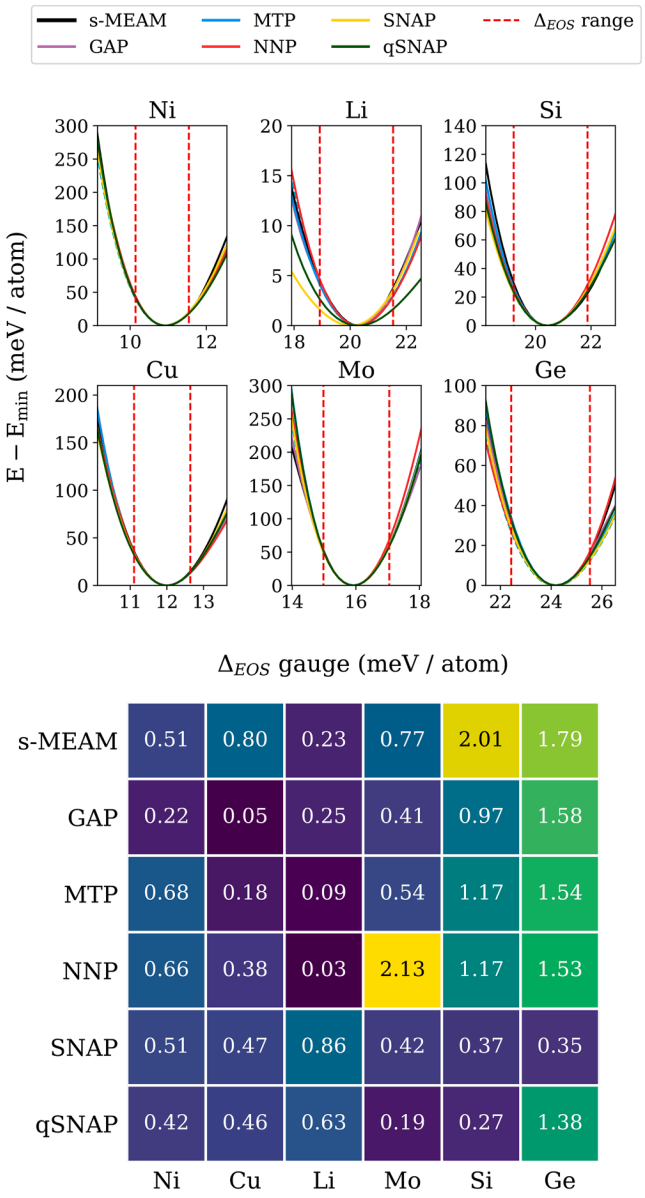

Fig. 5. Computed EOS curves (top) and Δ_{EOS} gauge values (bottom) calculated using 3rd-order Birch-Murnaghan fits. The Δ_{EOS} gauge is computed by integrating the root-mean-square difference between a model's EOS curve and the EOS curve predicted by DFT within a range of ±6% of $V_{0,DFT}$ (dashed red) after aligning the two curves with respect to their minima. $\Delta_{EOS} \lesssim 2$ meV·atom^{-1} has been used in the past as a threshold to define "indistinguishable EOS curves" when comparing DFT codes [51].

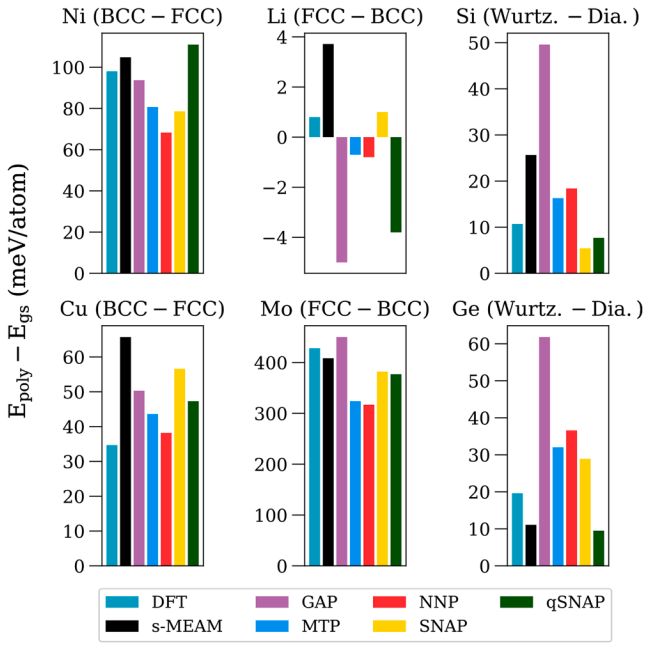

Fig. 6. Structural energy comparisons between the ground state structures and the DFT-predicted low-energy polymorph structures. Polymorph structures are labelled for each element. Although MLIPs are often considered to be less transferable (i.e. more prone to over-fitting) than classical potentials, they perform slightly better for low energy structure predictions than s-MEAM for most of the elements that were explored in this work.

terms in order to no longer see this trade-off. In many of the plots in Fig. 7 the empirical estimation of the Pareto front (dashed lines) becomes concave at the intersection of the s-MEAM and the MTP curves (see for example the Ge plot in Fig. 7), suggesting the possibility of a theoretical potential form that would have accuracies similar to some of the s-MEAM variants, but with lower complexity. A hypothetical potential form that could fill in these "holes" in the estimated Pareto front could be constructed by varying the number of 2- (ϕ, ρ) and 3-body (f and g) terms and the embedding number of embedding terms (U) beyond what is done in this work.

4.5. Fitted models

As can be seen in Fig. 8, the fitted splines for each element often optimize to recognizable and simple forms (e.g. LJ-like radial functions ϕ, ρ, and f, and nearly linear embedding functions U) that lend themselves to much easier interpretation than most MLIPs in practice. While these splines are more difficult to interpret than most analytical MEAM potentials (e.g. due to the roughness of some of the splines, or unexpected trends in some of the radial functions), they provide model developers with a way of identifying key trends in model behavior that would be difficult to recognize in an MLIP. In particular, because of the nearly linear shape typical of the U splines, it is possible to identify how different features in the splines will raise/lower the predicted energy at different distances or angles. Additionally, the fact that the U splines consistently optimized to nearly linear functions for all elements regardless of the number of knot points shows that high accuracies can be achieved while still maintaining simple forms for the embedding functions (which is not usually the case with MLIPs).

5. Conclusion

We demonstrate that a family of spline-based classical potentials offer a viable route towards sampling the low complexity, high accuracy region of model space while promoting physical interpretability, making

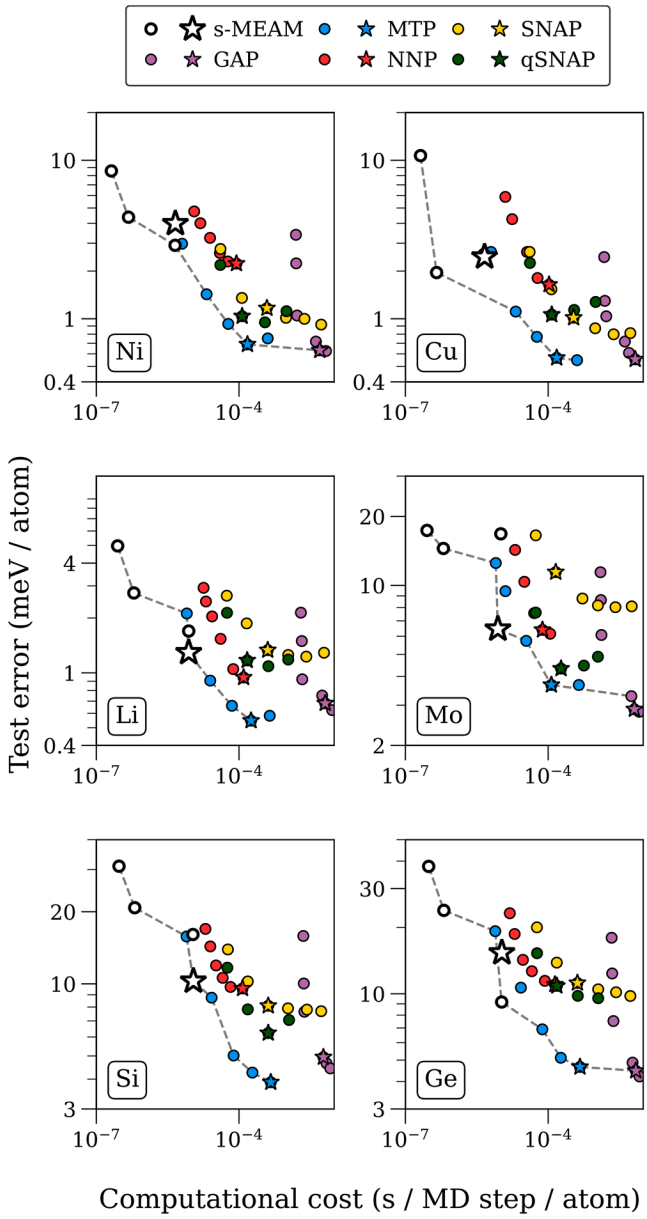

Fig. 7. Accuracy/cost trade-off for all elements and models. Stars are "optimal" potentials for each model that were used to compute all errors and material properties, while circles are different hyper-parameter choices of the MLIPs (e.g., the number of nodes in a NNP or the number of kernels in a GAP). For s-MEAM the stars are s-MEAM potentials (with all 5 terms included), but the circles moving from left to right are variants of s-MEAM that have (1) only the ϕ term, (2) only ϕ, U, and ρ, and (3) only ϕ, U, f, and g. In the case of Cu, the third variant which doesn't include ρ is obscured by the starred point. For all elements, the optimal s-MEAM potential falls on or near the empirical estimation of the Pareto front (dashed lines) while maintaining errors comparable to the MLIPs, indicating that it has an optimal balance between speed and accuracy.

them good candidates to be general-purpose potentials alongside existing low complexity MLIPs. In related recent work [52], an example of an MLIP modeled after the embedded-atom method also achieved high accuracy on the databases used in this paper, highlighting some of the efforts of approaching the problem by increasing MLIP interpretability. Given the already small error magnitudes for the s-MEAM potentials developed here and the MLIPs developed in [24], additional research that more exhaustively tests the generalizability of each of the model forms by constructing databases for more chemical systems and applications would be extremely useful. However, given that s-MEAM shows

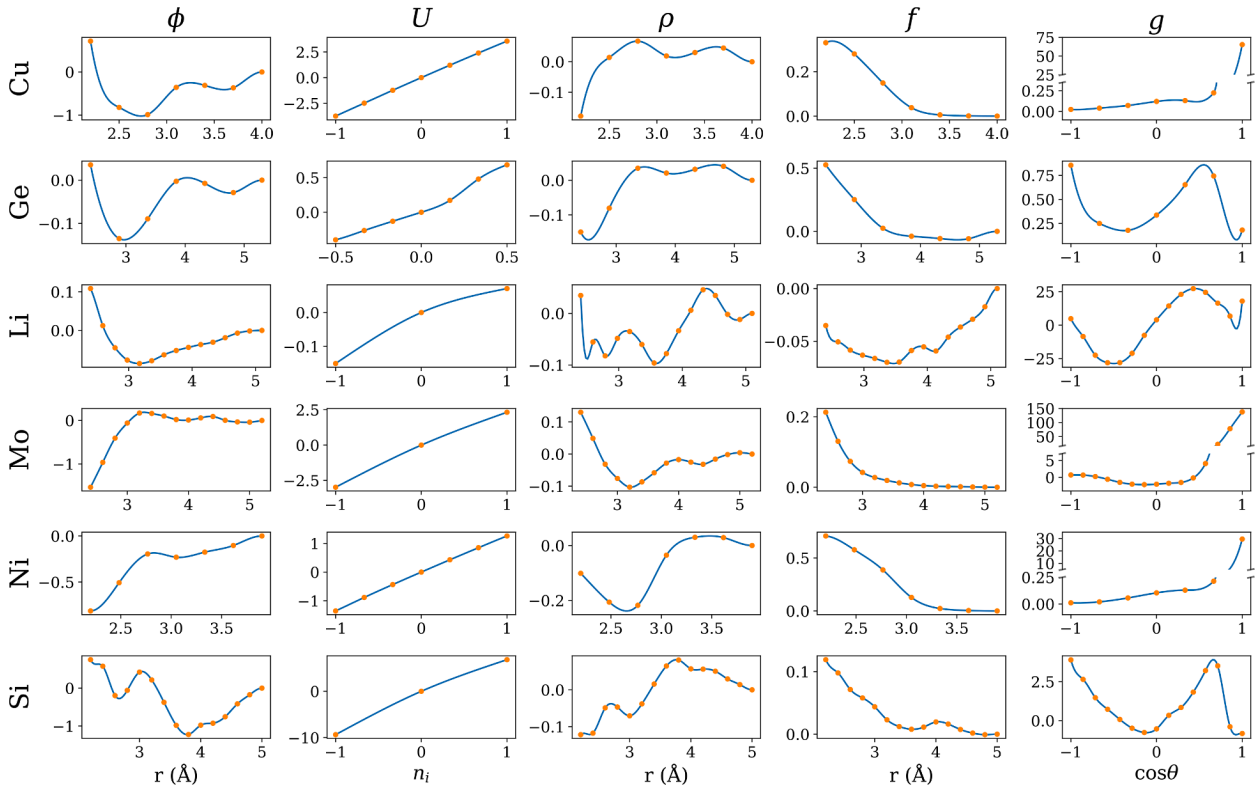

Fig. 8. Plots of fitted splines for each term in the s-MEAM equation for each element. In order to facilitate interpretation of the functions, some of the density terms have been scaled by -1 (by scaling U, ρ, and g by -1) so that all of the U splines have similar slopes, and some of the f splines have also been multiplied by -1. These changes are purely aesthetic, and don't impact the performance of the potentials, as discussed in [36]. The g splines for Cu, Mo, and Ni have been plotted with broken y axes to account for the large magnitudes of some of the knots at the ends of the splines, which is caused by poor sampling of $\cos\theta = 1$ in the databases leading to poorly-constrained knot values. The U domain for the Ge potential was chosen to be $[-0.5, 0.5]$ since it yielded a better potential duri.ng optimization.

signs of trade-offs in accuracies for the systems studied in this work, it's clear that further research is necessary to explore how its functional form might be modified or extended in order to make it more generalizable. While for some cases this might be possible by simply adding additional 2- or 3-body terms (or higher order, similar to what is done in [53,56]) or different embedding functions, there are other situations—i.e., organic systems, 2D materials, chemical reactions—where it would be helpful to design entirely new spline functions (similar to the COMB [9] approach of adding more terms). In those situations, other classical potentials already exist which would provide useful guidelines for how to introduce new terms into a fully general spline-based potential. Ideally, a more exhaustive study that covers a variety of classical and machine learning potentials on even more data sets than the six shown here would be performed to perform a more complete comparison of any inherent limitations between different potential forms.

During this process, it will be important for researchers to consider what level of model and descriptor complexity is tolerable for increasing the accuracies of interatomic potentials, especially if the differences between models are only on the order of a few meV·atom^{-1} as seen here. Related work exploring the necessary complexity of atomic descriptors [54,55] will be useful as researchers continue to evaluate model complexity. Finally, with the already excellent performances of both classical potentials and MLIPs, this suggests expanding efforts towards building, documenting, and curating suitable fitting databases, which would enable exhaustive model performance comparisons and significantly lower the barrier for developing interatomic potentials for new computational studies.

Data and Code Availability

The databases of atomic configurations were made available on Github [35] by the authors of [24], who originally created them. Fitted s-MEAM models from our work, and software used for fitting the s-MEAM potentials, are available on Github [41].

CRediT authorship contribution statement

Joshua A. Vita: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Dallas R. Trinkle:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing - review & editing.

Declaration of competing interest

The authors declare no competing financial or non-financial interests.

Acknowledgements

The authors thank Michael R. Fellinger for useful comments and discussions, and Yunxing Zuo for providing the data from [24] and helping to verify their results. This research was supported by the National Science Foundation through awards NSF/BD-SPOKE-1636929, NSF/NRT-1922758, and NSF/HDR-1940303. This research is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, athttps://doi.org/10.1016/j.commatsci.2021.110752.

References

[1] J.E. Jones. On the determination of molecular fields. —II. from the equation of state of a gas. Proc. R. Soc. Lond. Ser. A 106 (738): 463–477, October 1924. doi: 10.1098/rspa.1924.0082.

[2] R.A. Buckingham, The classical equation of state of gaseous helium, neon and argon, Proc. R. Soc. Lond. Ser. A 168 (933) (1938) 264–283, https://doi.org/10.1098/rspa.1938.0173.

[3] Murray S. Daw and M.I. Baskes. Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals. Phys. Rev. B, 29 (12): 6443–6453, June 1984. doi: 10.1103/physrevb.29.6443.

[4] M.I. Baskes, Modified embedded-atom potentials for cubic materials and impurities, Phys. Rev. B 46 (5) (1992) 2727–2742, https://doi.org/10.1103/physrevb.46.2727.

[5] J. Tersoff, New empirical model for the structural properties of silicon, Phys. Rev. Lett. 56 (6) (1986) 632–635, https://doi.org/10.1103/physrevlett.56.632.

[6] Frank H. Stillinger, Thomas A. Weber, Computer simulation of local order in condensed phases of silicon, Phys. Rev. B 31 (8) (1985) 5262–5271, https://doi.org/10.1103/physrevb.31.5262.

[7] Adri C.T. van Duin, Siddharth Dasgupta, Francois Lorant, William A. Goddard, ReaxFF: a reactive force field for hydrocarbons, J. Phys. Chem. A 105 (41) (2001) 9396–9409, https://doi.org/10.1021/jp004368u.

[8] Donald W Brenner, Olga A Shenderova, Judith A Harrison, Steven J Stuart, Boris Ni, and Susan B Sinnott. A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons. J. Phys.: Condens. Matter, 14 (4): 783–802, January 2002. doi: 10.1088/0953-8984/14/4/312.

[9] Tzu-Ray Shan, Bryce D. Devine, Travis W. Kemper, Susan B. Sinnott, and Simon R. Phillpot. Charge-optimized many-body potential for the hafnium/hafnium oxide system. Phys. Rev. B, 81 (12), March 2010. doi: 10.1103/physrevb.81.125328.

[10] Jörg Behler, Michele Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, Phys. Rev. Lett. 98 (14) (2007), https://doi.org/10.1103/physrevlett.98.146401.

[11] Albert P. Bartók, Mike C. Payne, Risi Kondor, Gábor Csányi, Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons, Phys. Rev. Lett. 104 (2010), 136403, https://doi.org/10.1103/PhysRevLett.104.136403.

[12] V. Botu, R. Ramprasad, Learning scheme to predict atomic forces and accelerate materials simulations, Phys. Rev. B 92 (9) (2015), https://doi.org/10.1103/physrevb.92.094306.

[13] A.P. Thompson, L.P. Swiler, C.R. Trott, S.M. Foiles, G.J. Tucker, Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials, J. Comput. Phys. 285 (2015) 316–330, https://doi.org/10.1016/j.jcp.2014.12.018.

[14] Nongnuch Artrith, Alexander Urban, An implementation of artificial neural-network potentials for atomistic materials simulations: Performance for TiO2, Comput. Mater. Sci. 114 (2016) 135–150, https://doi.org/10.1016/j.commatsci.2015.11.047.

[15] Alireza Khorshidi, Andrew A. Peterson, Amp: A modular approach to machine learning in atomistic simulations, Comput. Phys. Commun. 207 (2016) 310–324, https://doi.org/10.1016/j.cpc.2016.05.010.

[16] Alexander V. Shapeev, Moment tensor potentials: A class of systematically improvable interatomic potentials, Multiscale Model. Simul. 14 (3) (2016) 1153–1173, https://doi.org/10.1137/15m1054183.

[17] K.T. Schütt, H.E. Sauceda, P.-J. Kindermans, A. Tkatchenko, K.-R. Müller, SchNet – a deep learning architecture for molecules and materials, J. Chem. Phys. 148 (24) (2018), 241722, https://doi.org/10.1063/1.5019779.

[18] Han Wang, Linfeng Zhang, Jiequn Han, Weinan E. DeePMD-kit, A deep learning package for many-body potential energy representation and molecular dynamics, Comput. Phys. Commun. 228 (2018) 178–184, https://doi.org/10.1016/j.cpc.2018.03.016.

[19] Mitchell A. Wood, Aidan P. Thompson, Extending the accuracy of the SNAP interatomic potential form, J. Chem. Phys. 148 (24) (2018), 241721, https://doi.org/10.1063/1.5017641.

[20] Yaolong Zhang, Hu. Ce, Bin Jiang, Embedded atom neural network potentials: Efficient and accurate machine learning with a physically inspired representation, J. Phys. Chem. Lett. 10 (17) (2019) 4962–4967, https://doi.org/10.1021/acs.jpclett.9b02037.

[21] G.P. Purja Pun, R. Batra, R. Ramprasad, Y. Mishin, Physically informed artificial neural networks for atomistic modeling of materials, Nat. Commun. 10 (1) (2019), https://doi.org/10.1038/s41467-019-10343-5.

[22] Rahulkumar Gayatri, Stan Moore, Evan Weinberg, Nicholas Lubbers, Sarah Anderson, Jack Deslippe, Danny Perez, and Aidan P. Thompson. Rapid Exploration of Optimization Strategies on Advanced Architectures using TestSNAP and LAMMPSl. 2020.

[23] Yu Xie, Jonathan Vandermause, Lixin Sun, Andrea Cepellotti, and Boris Kozinsky. Bayesian force fields from active learning for simulation of inter-dimensional transformation of stanene. npj Comput. Mater., 7 (1), March 2021. doi: 10.1038/s41524-021-00510-y.

[24] Yunxing Zuo, Chi Chen, Xiangguo Li, Zhi Deng, Yiming Chen, Jörg Behler, Gábor Csányi, Alexander V. Shapeev, Aidan P. Thompson, Mitchell A. Wood, and Shyue Ping Ong. Performance and cost assessment of machine learning interatomic potentials. J. Phys. Chem. A, 124 (4): 731–745, January 2020. doi: 10.1021/acs.jpca.9b08723.

[25] Jörg Behler, Atom-centered symmetry functions for constructing high-dimensional neural network potentials, J. Chem. Phys. 134 (7) (2011), 074106, https://doi.org/10.1063/1.3553717.

[26] Albert P. Bartók, Risi Kondor, Gábor Csányi, On representing chemical environments, Phys. Rev. B 87 (18) (2013), https://doi.org/10.1103/physrevb.87.184115.

[27] M.I. Byeong-Joo Lee, Hanchul Kim Baskes, Yang Koo Cho, Second nearest-neighbor modified embedded atom method potentials for bcc transition metals, Phys. Rev. B 64 (18) (2001), https://doi.org/10.1103/physrevb.64.184102.

[28] S. Nouranian, M.A. Tschopp, S.R. Gwaltney, M.I. Baskes, M.F. Horstemeyer, An interatomic potential for saturated hydrocarbons based on the modified embedded-atom method, Phys. Chem. Chem. Phys. 16 (13) (2014) 6233–6249, https://doi.org/10.1039/c4cp00027g.

[29] T.J. Lenosky, B. Sadigh, E. Alonso, V.V. Bulatov, T.D. de la Rubia, J. Kim, A. F. Voter, J.D. Kress, Highly optimized empirical potential model of silicon, Modell. Simul. Mater. Sci. Eng. 8 (6) (2000) 825–841, https://doi.org/10.1088/0965-0393/8/6/305.

[30] Michael R. Fellinger, Hyoungki Park, John W. Wilkins, Force-matched embedded-atom method potential for niobium, Phys. Rev. B 81 (14) (2010), https://doi.org/10.1103/physrevb.81.144119.

[31] Hyoungki Park, Michael R. Fellinger, Thomas J. Lenosky, William W. Tipton, Dallas R. Trinkle, Sven P. Rudin, Christopher Woodward, John W. Wilkins, Richard G. Hennig, Ab initio based empirical potential used to study the mechanical properties of molybdenum, Phys. Rev. B 85 (21) (2012), https://doi.org/10.1103/physrevb.85.214121.

[32] Pinchao Zhang, Dallas R. Trinkle, A modified embedded atom method potential for interstitial oxygen in titanium, Comput. Mater. Sci. 124 (2016) 204–210, https://doi.org/10.1016/j.commatsci.2016.07.039.

[33] R.G. Hennig, T.J. Lenosky, D.R. Trinkle, S.P. Rudin, J.W. Wilkins, Classical potential describes martensitic phase transformations between the α, β, and ω titanium phases, Phys. Rev. B 78 (5) (2008), https://doi.org/10.1103/physrevb.78.054121.

[34] Chaoming Yang, Liang Qi, Modified embedded-atom method potential of niobium for studies on mechanical properties, Comput. Mater. Sci. 161 (2019) 351–363, https://doi.org/10.1016/j.commatsci.2019.01.047.

[35] materialsvirtuallab/mlearn repository. URL:https://github.com/materialsvirtuallab/mlearn. Accessed: 2010-09-30.

[36] Peter Brommer, Franz Gähler, Potfit: effective potentials from ab initio data, Modell. Simul. Mater. Sci. Eng. 15 (3) (2007) 295–304, https://doi.org/10.1088/0965-0393/15/3/008.

[37] N. Hansen, A. Ostermeier, Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation, in: Proceedings of IEEE International Conference on Evolutionary Computation IEEE, 1996, https://doi.org/10.1109/icec.1996.542381.

[38] Nikolaus Hansen, The CMA Evolution Strategy, A Tutorial (2016).

[39] Nikolaus Hansen, Stefan Kern, Evaluating the CMA evolution strategy on multimodal test functions, in: Lect. Notes Comput. Sci., Springer, Berlin Heidelberg, 2004, pp. 282–291, https://doi.org/10.1007/978-3-540-30217-9_29.

[40] Nikolaus Hansen, Youhei Akimoto, and Petr Baudis. CMA-ES/pycma on Github. Zenodo, DOI:10.5281/zenodo.2559634, February 2019.

[41] Trinklegroup/s-meam repository. https://github.com/TrinkleGroup/s-meam. Accessed: 2010-09-30.

[42] Sayyed Jalil Mahdizadeh and Golnoosh Akhlamadi. Optimized tersoff empirical potential for germanene. J. Mol. Graph. Model., 72: 1–5, March 2017. doi: 10.1016/j.jmgm.2016.11.009.

[43] T. Kumagai, S. Izumi, S. Hara, S. Sakai, Development of bond-order potentials that can reproduce the elastic constants and melting point of silicon for classical molecular dynamics simulation, Comput. Mater. Sci. 39 (2) (2007) 457–464, https://doi.org/10.1016/j.commatsci.2006.07.013.

[44] Ebrahim Asadi, Mohsen Asle Zaeem, Sasan Nouranian, and Michael I. Baskes. Two-phase solid–liquid coexistence of ni, cu, and al by molecular dynamics simulations using the modified embedded-atom method. Acta Materialia, 86: 169–181, March 2015. doi: 10.1016/j.actamat.2014.12.010.

[45] A. Nichol, G.J. Ackland, Property trends in simple metals: An empirical potential approach, Phys. Rev. B 93 (18) (2016), https://doi.org/10.1103/physrevb.93.184101.

[46] X.W. Zhou, R.A. Johnson, H.N.G. Wadley, Misfit-energy-increasing dislocations in vapor-deposited CoFe/NiFe multilayers, Phys. Rev. B 69 (14) (2004), https://doi.org/10.1103/physrevb.69.144113.

[47] K. Lejaeghere, V. Van Speybroeck, G. Van Oost, S. Cottenier, Error estimates for solid-state density-functional theory predictions: An overview by means of the ground-state elemental crystals, Crit. Rev. Solid State Mater. Sci. 39 (1) (2013) 1–24, https://doi.org/10.1080/10408436.2013.772503.

[48] F. Ercolessi and J.B Adams. Interatomic potentials from first-principles calculations: The force-matching method. Europhys. Lett. (EPL), 26 (8): 583–588, June 1994. doi: 10.1209/0295-5075/26/8/005. doi: 10.1209/0295-5075/26/8/005.

[49] Albert P. Bartók, James Kermode, Noam Bernstein, Gábor Csányi, Machine learning a general-purpose interatomic potential for silicon, Phys. Rev. X 8 (4) (2018), https://doi.org/10.1103/physrevx.8.041048.

[50] Jonathan A Zimmerman, Huajian Gao, and Farid F Abraham. Generalized stacking fault energies for embedded atom FCC metals. Modell. Simul. Mater. Sci. Eng., 8 (2): 103–115, February 2000. doi: 10.1088/0965-0393/8/2/302.
[51] K. Lejaeghere et al. Reproducibility in density functional theory calculations of solids. Science, 351 (6280): aad3000–aad3000, March 2016. doi: 10.1126/science.aad3000.
[52] Yaolong Zhang, Hu. Ce, Bin Jiang, Accelerating atomistic simulations with piecewise machine-learned ab initio potentials at a classical force field-like cost, PCCP 23 (3) (2021) 1815–1821, https://doi.org/10.1039/d0cp05089j.
[53] Ralf Drautz. Atomic cluster expansion for accurate and transferable interatomic potentials. Phys. Rev. B, 99 (1), January 2019. doi: 10.1103/physrevb.99.014104.
[54] Sergey N. Pozdnyakov, Michael J. Willatt, Albert P. Bartók, Christoph Ortner, Gábor Csányi, and Michele Ceriotti. Incompleteness of atomic structure representations, Phys. Rev. Lett. 125 (16) (2020), https://doi.org/10.1103/physrevlett.125.166001.
[55] Claudio Zeni, Kevin Rossi, Aldo Glielmo, Stefano de Gironcoli, Compact atomic descriptors enable accurate predictions via linear models, J. Chem. Phys. 154 (22) (2021), 224112, https://doi.org/10.1063/5.0052961.
[56] Y. Lysogorskiy, C.v.d. Oord, A. Bochkarev, et al., Performant implementation of the atomic cluster expansion (PACE) and application to copper and silicon, npj Comput. Mater. 7 (2021) 97, https://doi.org/10.1038/s41524-021-00559-9.