All-atom dynamics simulations are an indispensable quantitative tool in physics, chemistry, and materials science, but large systems and long simulation times remain challenging due to the trade-off between computational efficiency and predictive accuracy. To address this challenge, we combine effective two- and three-body potentials in a cubic B-spline basis with regularized linear regression to obtain machine-learning potentials that are physically interpretable, sufficiently accurate for applications, as fast as the fastest traditional empirical potentials, and two to four orders of magnitude faster than state-of-the-art machine-learning potentials. For data from empirical potentials, we demonstrate the exact retrieval of the potential. For data from density functional theory, the predicted energies, forces, and derived properties, including phonon spectra, elastic constants, and melting points, closely match those of the reference method. The introduced potentials might contribute towards accurate all-atom dynamics simulations of large atomistic systems over long-time scales.

npj Computational Materials (2023)9:162; https://doi.org/10.1038/s41524-023-01092-7

INTRODUCTION

All-atom dynamics simulations enable the quantitative study of atomistic systems and their interactions in physics, chemistry, materials science, pharmaceutical sciences, and related areas. The simulations’ capabilities and limits depend on the potential used to calculate the forces acting on the atoms, with an inherent correlation between the accuracy of the underlying physical model and the required computational effort: On the one hand, electronic-structure methods tend to be accurate, slow, applicable to many systems and require little human parameterization effort. On the other hand, traditional empirical potentials are fast but limited in accuracy and applicability, with often high parameterization effort.

Machine-learning potentials (MLPs) are flexible functions fitted to reference energy and force data, e.g., from electronic-structure methods. Their computational advantage does not primarily arise from simplified physical models but from avoiding redundant calculations through interpolation. Hence, they can stay close to the accuracy of the reference method while being orders of magnitude faster (see Fig. 1), with little human parameterization effort. However, current accurate MLPs are still orders of magnitude slower than traditional empirical potentials, limiting their use for dynamics simulations of large atomistic systems over long time scales.

In this work, we develop an interpolable MLP based on effective two- and three-body potentials using a flexible cubic B-spline basis. Figure 1 demonstrates how this ultra-fast (UF) potential is close in error to state-of-the-art MLPs while being as fast as the fastest traditional empirical potentials, such as the Morse and Lennard-Jones potentials.

The many-body expansion of an atomistic system’s potential energy

\[
E = E_0 + \frac{1}{2!} \sum_i V_i (R_i, \sigma_i) + \frac{1}{3!} \sum_{ij} V_{ij} (R_{ij}, R_i, \sigma_i, \sigma_j) + \frac{1}{4!} \sum_{ijk} V_{ijk} (R_{ijk}, R_{ij}, \sigma_{ij}, \sigma_k) + \ldots
\]  

(1)

is a sum of N-body potentials \( V_N \) that depend on atom positions \( R_i \) and element species \( \sigma_i \), where the prefactors account for double-counting and \( V_N = 0 \) if two indices are the same. Assuming the transferability of the \( V_N \) across configurations provides the basis for empirical and machine-learning potentials.

In this study, we use elemental tungsten as an example and hence omit the species dependency as well as the reference energies \( E_0 \) and \( \frac{1}{2} \sum_{ij} V_{ij} (R_{ij}, \sigma_i, \sigma_j) \), and where appropriate subsume the factorial prefactors into the potentials \( V \) for simplicity. The extension to multi-component systems is straightforward and implemented in the accompanying program code.

Traditional empirical potentials often have rigid functional forms with a small number of tunable parameters. These are optimized to reproduce experimental quantities, such as lattice parameters and elastic coefficients, as well as calculated quantities from first principles, such as energies of crystal structures, defects, and surfaces. Typically this requires global optimization, e.g., via simulated annealing, and substantial human effort.

Pair potentials truncate Equation (1) after two-body terms,

\[
E = \frac{1}{2} \sum_{ij} V_{ij} (r_{ij}),
\]  

(2)

where \( r_{ij} \) is the distance between atoms \( i \) and \( j \). These potentials are limited to systems where higher-order terms such as angular and dihedral interactions are negligible. The functional forms of the Lennard–Jones (LJ) potential

\[
V_{\text{LJ}}(r_{ij}) = 4\epsilon \left[ \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^{6} \right]
\]  

(3)

and the Morse potential

\[
V_{\text{Morse}}(r_{ij}) = D_0 \left( e^{-2a(r_{ij} - r_c)} - 2e^{-a(r_{ij} - r_c)} \right)
\]  

(4)

were originally developed for their numerical efficiency, where \( \epsilon, \sigma \), and \( D_0, a, r_c \) are model parameters.

Many-body potentials extend the pair formalism by including additional many-body interactions, either in the form of many-body functions as in Equation (1), or via environment-dependent
functionals, such as in the embedded atom method (EAM)\textsuperscript{11},

$$E_{\text{EAM}} = \frac{1}{2} \sum_{ij} V(r_{ij}) + \sum_{i} F \left( \sum_{j} \rho(r_{ij}) \right).$$  \hfill (5)

Here, the embedding energy \( F \) is a non-linear function of the electron density \( \rho \), which is approximated by a pairwise sum.

While traditional empirical potentials have seen success in applications across decades of research, their rigid functional forms limit their accuracy. More recently, MLPs with flexible, functional forms and built-in physics domain knowledge in the form of engineered features\textsuperscript{21,22} or deep neural network architectures have emerged as an alternative\textsuperscript{13-16}. State-of-the-art MLPs can simulate the dynamics of large ("high-dimensional") atomistic systems with an accuracy close to the underlying electronic-structure reference method but orders of magnitude faster\textsuperscript{17,18}. However, they are still orders of magnitude slower than fast models based on polynomial basis sets.

Recent efforts to improve the speed and accuracy of MLPs include using linear regression models, which can be faster to train and evaluate than non-linear models\textsuperscript{19,20}. The spectral neighbor analysis potential (SNAP)\textsuperscript{21} and its quadratic variant (qSNAP)\textsuperscript{22} are linear models based on the bispectrum representation\textsuperscript{23}. Moment tensor potentials (MTP)\textsuperscript{24}, perform atomic cluster expansion (PACE) potentials\textsuperscript{19,25}, atomic permutationally-invariant polynomials (aPIP) potentials\textsuperscript{26}, and Chebyshev interaction model for efficient simulation (ChIMES) potentials\textsuperscript{27} are linear models based on polynomial basis sets.

A complementary approach to improve speed is to use basis functions that are fast to evaluate. In the context of MLPs, non-linear kernel-based MLPs have been trained and subsequently projected onto a spline basis, yielding a linear model\textsuperscript{28}. Similar to this work, the general two- and three-body potential approach employs a quadratic spline basis set, exceeding MTPs in speed when trained on the same data\textsuperscript{29}. Polynomial symmetry functions improve over Behler–Parrinello symmetry functions\textsuperscript{30} in speed and accuracy by introducing compact support\textsuperscript{31}. Methods for fitting spline-based modified EAM potentials were benchmarked against MLPs, demonstrating comparable accuracy despite the lower complexity of their functional forms\textsuperscript{32}.

Motivated by these observations, we developed a UF MLP that combines the speed of the fastest traditional empirical potentials with an accuracy close to state-of-the-art MLPs by employing regularized linear regression with spline basis functions with compact support to learn effective two- and three-body interactions. Figure 1 showcases the relation between prediction errors and computational costs for three traditional empirical potentials (LJ, Morse, EAM) and several MLPs benchmarked on a dataset of elemental tungsten\textsuperscript{33}. While the traditional empirical potentials are fast but limited by accuracy, the MLPs are accurate but limited by speed. UF MLPs improve on the Pareto frontier of predictive accuracy and computational costs. They are available as an open-source Python implementation (UF\textsuperscript{3} Ultra-Fast Force Fields)\textsuperscript{34} with interfaces to the VASP\textsuperscript{35} electronic-structure code and the LAMMPS\textsuperscript{36,37} molecular dynamics code.

**RESULTS AND DISCUSSION**

The central idea of UF MLPs is to learn an effective low-order many-body expansion of the potential energy surface, using basis functions that are efficient to evaluate. For this, we truncate the many-body expansion of Eq. (1) at two- or three-body terms and express each term as a function of pairwise distances (one distance for the two-body and three distances for the three-body term). This approach is general and can be extended to higher-order terms. To minimize the computational cost of predictions, we represent \( N \)-body terms in a set of basis functions with compact support and sufficiently many derivatives to describe energies, forces, and vibrational modes, i.e., cubic splines.

**Expansion in B-splines**

Splines are piecewise polynomial functions with locally simple forms joined together at knot positions\textsuperscript{38}. They are globally flexible and smooth but do not suffer from some of the oscillatory problems of polynomial interpolators (e.g., Runge’s phenomenon\textsuperscript{39}).

Spline interpolation is well-established in empirical potential development\textsuperscript{40,41}. The LAMMPS package\textsuperscript{36}, a leading framework for molecular dynamics simulations, implements cubic spline interpolation for pair potentials and selected many-body potentials, including EAM. The primary motivations for splines are the desire for computational efficiency and the opportunity to improve accuracy systematically\textsuperscript{42}. Their compact support and simple form make spline-based potentials the fastest choice to evaluate, and adding more knots increases their resolution.

**B-splines** constitute a basis for splines of arbitrary order. They are recursively defined as\textsuperscript{38}

$$B_{n,1}(r) = \begin{cases} 1, & t_n \leq r < t_{n+1} \\ 0, & \text{otherwise} \end{cases}$$

$$B_{n,d+1}(r) = \frac{r-t_n}{t_{n+d}-t_n} B_{n,d}(r) + \frac{t_{n+d+1}-r}{t_{n+d+1}-t_{n+1}} B_{n+1,d}(r),$$  \hfill (6)

where \( t_n \) is the \( n \)-th knot position, and \( d \) is the degree of the polynomial. The position and number of knots, a non-decreasing sequence of support points that uniquely determine the basis set,
may be fixed or treated as free parameters. B-Splines are well suited for interpolation due to their intrinsic smoothness and differentiability. Their derivatives are also defined recursively:

$$B_{n,d+1}(r) = d \left( B_{n,d}(r) - \frac{B_{n+1,d}(r)}{t_{n+1} - t_n} \right) \quad \text{(7)}$$

The UF potential describes the energy $E$ of an atomistic system via two- and three-body interactions:

$$E = \sum_{ij} V_2(r_{ij}) + \sum_{ijk} V_3(r_{ij}, r_{ik}, r_{jk}). \quad \text{(8)}$$

For finite systems such as molecules or clusters, indices $i, j, k$ run over all atoms. For infinite systems modeled via periodic boundary conditions, $i$ runs over the atoms in the simulation cell, and $j, k$ run over all neighboring atoms, including those in adjacent copies of the simulation cell. While these are infinitely many, the sums are truncated in practice by assuming locality, that is, finite support of $V_2$ and $V_3$.

Modeling $N$-body interactions requires $(N^2)$-dimensional tensor product splines. The UF potential therefore expresses $V_2$ and $V_3$ as linear combinations of cubic B-splines, $B_n = B_{n,3,1}$, and tensor product splines:

$$V_2(r_{ij}) = \sum_{n=0}^{K} c_n B_n(r_{ij}) \quad \text{(9)}$$

$$V_3(r_{ij}, r_{ik}, r_{jk}) = \sum_{n=0}^{K} \sum_{m=0}^{K} \sum_{l=0}^{K} c_{mn} B_n(r_{ij}) B_m(r_{ik}) B_l(r_{jk}),$$

where $K, K_i, K_m$, and $K_n$ denote the number of basis functions per spline or tensor spline dimension, and $c_n$ and $c_{mn}$ are corresponding coefficients.

The B-spline basis set spans a finite domain and is bounded by the end knots $[t_0, t_f]$. Figure 2A illustrates the compact support of the cubic B-spline basis functions. We constrain the coefficients such that at the upper limit $t_f$, the potential smoothly goes to zero, and near the lower limit $t_0$ it monotonically increases with shorter distances to prevent atoms from getting unphysically close. By construction, each basis function is non-zero across four adjacent intervals. Therefore, evaluating the two- and three-body potentials involves evaluating at most 4 and 43 intervals. Therefore, evaluating the two- and three-body potentials, two- and three-body potentials, each basis function is non-zero across four adjacent distances to prevent atoms from getting unphysically close. By construction, each basis function is non-zero across four adjacent intervals.

The force $F_{rij}$ acting on atom $a$ is given as the negative gradient $-\nabla V_{E} E$ of the energy $E$ with respect to the atom’s Cartesian coordinate $R_{a}$, and obtained analytically from the derivatives of the two- and three-body potentials,

$$\frac{\partial V_2(r_{ij})}{\partial R_{a}} = \sum_{n=0}^{K} c_n B'_n(r_{ij}) \frac{\partial R_{a}}{\partial R_{a}} \quad \text{(10)}$$

$$\frac{\partial V_3(r_{ij}, r_{ik}, r_{jk})}{\partial R_{a}} = \sum_{n=0}^{K} \sum_{m=0}^{K} \sum_{l=0}^{K} c_{mn} B'_n(r_{ij}) B'_m(r_{ik}) B'_l(r_{jk}) \frac{\partial R_{a}}{\partial R_{a}}$$

with

$$\frac{\partial R_{a}}{\partial R_{a}} = \frac{(\delta_{aj} - \delta_{ai})(R_{aj} - R_{ai})}{r_{ij}} \quad \text{(11)}$$

where $l$ is the Cartesian coordinate.

Figure 2B illustrates, for the two-body potential, that the choice of the cubic B-spline basis results in a smooth, continuous first derivative comprised of quadratic B-splines and a continuous second derivative of linear B-splines.

Regularized least-squares optimization with energies and forces

During the fitting procedure, we optimize all spline coefficients $c$ simultaneously with the regularized linear least-squares method. Given atomic configurations $S$, energies $E$, and forces $F$, we fit the potential energy function $E$ of Equation (8) by minimizing the loss function

$$L = \frac{1}{2} \sum_{s \in S} (E(s) - \bar{E}_s)^2 + \frac{1}{2} \sum_{s \in S} (\nabla E(s) - \bar{F}_s)^2 + \lambda_1 \sum_n c_n^2 + \lambda_2 \sum_{n=0}^{K} c_{mn}^2 \quad \text{(12)}$$

where the second sum is taken over force components. Here, $\lambda \in [0, \infty]$ is a weighting parameter that controls the relative contributions between energy and force-component residuals, $|E|$ and $|F|$ are the number of energy and force component observations in the training set, and $\sigma_E$ and $\sigma_F$ are the sample standard deviations of energies and force components across the training set. This normalization yields dimensionless residuals, allowing $\lambda$ to balance the relative contributions from energies and forces independently of the size and variance of the training energies and forces.
The minimization of $L$ with respect to the spline coefficients $c$ is a linear least-squares problem with Tikhonov regularization and solution

$$c = (X^TX + \lambda_1 I + \lambda_2 D_2^T D_2)^{-1} X^T y,$$  \hspace{1cm} (13)

where $I$ is the identity matrix, $y$ contains energies and forces, and each element of $X$ is the sum of B-spline values taken over all relevant pair distances in each configuration (rows) for each basis function (columns). Subsets of columns correspond to different body orders. Similarly, for multi-component systems, subsets of columns correspond to different chemical interactions. When forces are included in $y$, the corresponding rows of $X$ are generated according to Eqs. (10) and (11). This optimization problem is strongly convex, allowing for an efficient and deterministic solution, e.g., with LU decomposition.

The used Tikhonov regularization controls the magnitude of spline coefficients $c$ through the parameter $\lambda_1$ via the ridge penalty and the curvature and local smoothness across adjacent spline coefficients through $\lambda_2$ via the difference penalty\(^43,44\). For the two-body case, $D_2$ is given by

$$D_2 = \begin{pmatrix} 1 & -2 & 1 & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 1 & -2 & 1 \end{pmatrix}.$$ \hspace{1cm} (14)

For higher-order potential terms, the difference penalty affects the spline coefficients that are adjacent in each dimension of a tensor product spline. This difference penalty is related to a penalty on the integral of the squared second derivative of the potential\(^26,45,46\). However, the difference penalty is less complex because the dimensionality of the corresponding regularization problem is simply the number of basis functions $K$\(^44\).

Figure 2C compares optimized UF$_2$, LJ, and Morse potentials for tungsten. Although the three curves have similar minima and behavior for greater pair distances $r$, the UF$_2$ potential exhibits additional inflection points. We attribute the ability of the UF$_2$ potential to reproduce the properties of a bcc metal, a traditionally difficult task for pair potentials, to its flexible, functional form.

The fitting of a UF potential maps energy and force data onto effective two- and three-body terms, as shown in Fig. 3 for the tungsten dataset. Both terms can be visualized directly, providing interpretability by disentangling contributions to the interatomic interactions. The inspection of minima, repulsive and attractive contributions and inflection points enables insight into the chemical bonding characteristics of the material. This straightforward and visual analysis makes the UF potentials more directly interpretable than most MLPs.

**Convergence of error in energy and force predictions**

UF potentials should exactly reproduce any two- and three-body reference potential by construction, given sufficiently many basis functions and training data. To establish baseline functionality, we fit the UF$_2$ potential to energies and forces from the LJ potential for elemental tungsten and the UF$_{2,3}$ potential to energies and
forces from the Stillinger-Weber potential on elemental silicon, which they both reproduce with negligible error (see Supplementary Figs. 2 and 3 for learning curves and details).

To assess the accuracy of UF potentials, we measure their ability to predict density-functional theory (DFT) energies and forces in tungsten as a function of the amount of training data. Figure 4 compares learning curves for the two-body UF$_2$, two- and three-body UF$_{2,3}$, SNAP, and qSNAP potentials. To quantify prediction performance, we use the root-mean-squared error (RMSE) on randomly sampled hold-out test sets. In this, we ensured that each test set contained all available configuration types (see Section “Data”). Each learning curve is fit with a soft plus function $\ln(1 + e^n)$, where $n$ is the number of training data. This function captures both the initial linear slope in log-log space and the observed saturation due to the models’ finite complexity.

Of the four models, the UF$_2$ potential, using 28 basis functions, converges earliest. The SNAP potential, using 56 basis functions based on hyperspherical harmonics, converges slightly later with lower errors. The qSNAP potential, using 496 basis functions including quadratic terms, converges last, with modest improvements in error over SNAP. Finally, the UF$_{2,3}$ potential, using 915 basis functions, is comparable to SNAP in convergence speed with lower errors in energies and similar errors in forces. Separate energy and force learning curves are included in Supplementary Fig. 5.

The convergence speed is related to both the number of basis functions and the complexity of the many-body interactions. This indicates that the simpler UF$_2$ potential may be more suitable than other MLPs when data is scarce.

**Validation with derived quantities**

To benchmark performance for applications, we computed several derived quantities that were not included in the fit, such as the phonon spectrum and melting temperature, using each potential. Figure 5 shows the relative errors in 12 quantities: energy, forces,
potential outperforms empirical potentials (Morse, EAM4) in Phonon dispersion curves for bcc tungsten. A
errors to the UF2 and EAM4 potentials. Interactions allows the UF2,3 potential to approach the performance of the MTP and GAP potentials.

![Figure 6](https://example.com/figure6)

**Figure 6** Phonon dispersion curves for bcc tungsten. A The UF2 potential outperforms empirical potentials (Morse, EAM4) in reproducing the reference phonon frequencies (pink squares). The LJ curve, omitted, exhibits large oscillations and negative phonon frequencies. B SNAP and qSNAP have similar phonon frequency errors to the UF2 and EAM4 potentials. C The addition of three-body interactions allows the UF2,3 potential to approach the performance of the MTP and GAP potentials.

| Table 1. Melting temperature predictions alongside energy, force, and phonon frequency benchmarks for bcc tungsten. |
|---|
| Relative Comp. Melting Temp. | Energy (eV/atom) | Force (eV/Å) | Phonon RMSE (THz) |
| DFT | 3465 ± 105 | 0.110 | 1.400 | 3.914 |
| LJ | 5695 ± 90 | 0.040 | 1.140 |
| Morse | 2681 ± 45 | 0.027 | 0.387 | 0.230 |
| UF2 | 2985 ± 68 | 0.005 | 0.152 | 0.263 |
| EAM4 | 2973 ± 78 | 0.017 | 0.146 | 0.376 |
| UF2,3 | 3651 ± 31 | 0.010 | 0.167 | 0.256 |
| MTP | 3961 ± 82 | 0.014 | 0.189 | 0.270 |
| qSNAP | 3136 ± 63 | 0.006 | 0.169 | 0.291 |

Compared to LJ, Morse, and EAM4, the UF2 potential achieves a low error in the melting temperature for a similar cost. The UF2,3 potential prediction is even closer to the DFT reference at the cost of one order of magnitude in speed. MTP, SNAP, qSNAP, and GAP require one to three orders of magnitude more computational resources for the same large-scale simulation. See Sections IID and IIID for details.

Figure 6 compares the calculated phonon spectra of the various MLPs with the DFT reference values. Perhaps surprisingly, the UF2 pair potential is comparable in error to the potentials with many-body terms. In contrast, the Morse pair potential is rather inaccurate, and the LJ pair potential, not shown, yields a phonon spectrum with imaginary frequency and large frequency oscillations. Table 1 summarizes the RMSE of the phonon frequencies computed across the 26 DFT reference values. As in other calculated properties, the addition of three-body interactions in the UF2,3 potential significantly improves the agreement with the DFT reference compared to the UF2 pair potential and even surpasses the other computationally more expensive MLPs.
As an example of a practical, large-scale calculation, we predict the melting temperature of tungsten (see Section Calculations for computational details). Since the training set of the MLPs does not include liquid configurations, the melting point predictions measure the models’ extrapolative capacity. Table 1 compares the predicted melting temperatures of the MLPs to the ab-initio reference value of 3465 K. We observe that the other pair potentials are limited in their predictive capabilities while the UF2 potential is comparable in accuracy to the more complex potentials, which yield reasonable predictions. The qSNAP melting calculation failed due to numerical instability at higher temperatures, which is known to occur in high-dimensional potentials. Bonds break at higher temperatures, leading to many local atomic configurations that are underrepresented in the training set. We expect that training the qSNAP potential on a suitable, more extensive dataset would remove this instability. The extrapolative capacity of the UF2 and UF2,3 potentials in melting-temperature simulations illustrates that the UF potentials can provide an accurate description with only a moderately sized training dataset, indicating their usefulness for materials simulations with limited reference data.

**Summary**

We developed and implemented an MLP that is fast to train and evaluate, provides an interpretable form, is extendable to higher-order interactions, and accurately describes materials even for comparatively sparse training sets. The approach is based on an effective many-body expansion and utilizes a flexible B-spline basis. The resulting regularized linear least-squares optimization problem is strongly convex, enabling efficient training.

For the example of elemental tungsten, the UF2 pair potential produces energy, force, and property predictions rivaling those of SNAP and qSNAP while matching the cost of the Morse potential, corresponding to a reduction in computational cost by two orders of magnitude. Despite the intrinsic limitations of the pair potential in capturing physics, we find that the UF2 pair potential yields reasonable predictions in property benchmarks, such as for elastic constants, phonons, surface energies, and melting temperature. The UF2,3 potential, which accounts for three-body interactions, approaches the accuracy of MTP and GAP while reducing computational cost by one to three orders of magnitude.

The rapidly increasing number of MLPs, from ultra-fast linear models to graph neural network potentials, highlights the trade-offs between computational efficiency, robustness, and model capacity. Complex, high-dimensional MLPs are expected to yield higher accuracy for complex systems at the price of greatly increased computational cost and data requirements. The UF approach yields potentials that are fast and robust at the price of reduced flexibility and possibly greater errors for complex systems. Future work on UF potentials will explore the use of active learning for increased robustness and data efficiency, as well as the addition of the four-body term, which is necessary for modeling dihedral angles that are critical to describing organic molecules and protein structures. The software for fitting UF potentials and running them in the LAMMPS molecular dynamics code is freely available in our GitHub repository.

**METHODS**

**Data**

To compare the UF potential against existing potentials, we use a dataset by Szlachta et al., which has been used before to benchmark the GAP, SNAP, and aPIP potentials. This dataset of 693 tungsten configurations includes body-centered cubic (bcc) primitive cells, bulk snapshots from molecular dynamics, surfaces, vacancies, gamma surfaces, gamma surface vacancies, and dislocation quadrupoles. Energies, forces, and stresses in the dataset were computed using DFT with the Perdew–Burke–Ernzerhof functional.

**B-spline basis**

The UF potential uses natural cubic B-splines: Their first derivative is continuous and smooth, while their second derivative is continuous. They are natural rather than clamped in that their second derivative is zero at the boundary conditions. These properties, which are critical for accurately reproducing forces and stresses, motivated our choice of basis set. Other B-spline schemes have been explored for interpolation in empirical potential development. Wen et al. discuss clamped and Hermite splines as well as quartic and quintic splines. The natural cubic spline is sufficient except when computing properties that rely on the third and fourth derivative of the potential, such as thermal expansion and finite-temperature elastic constants.

Uniform spacing of knots is a reasonable choice in many cases. However, one may adjust the density of knots to control resolution in regions of interest. Due to compact support in this basis, each pair-distance energy requires the evaluation of exactly four B-splines. Hence, the potential’s speed scales with neither the number of knots nor the number of basis functions.

On the other hand, the minimum distance between knots limits the maximum curvature of the function. The optimum density of knots thus depends on the training set. Underfitting or overfitting may arise from insufficient or excessive knot density, respectively. Based on convergence tests (see Supplementary Fig. 1), we fit UF potentials in this work using 25 uniformly spaced knots.

By construction, each spline coefficient influences the overall function across five adjacent knots. As a result, one can tune the shape of the potential further according to additional constraints. For instance, soft-core and smooth-cutoff requirements may be satisfied by adjusting spline coefficients at the ends. In this work, we ensure that the potential and its first derivative smoothly go to zero when approaching the cutoff at \( r_0 = 1 \) by setting the last three coefficients to 0.

**Models**

In this work, we partitioned the data using a random split of 20% training and 80% testing data. The testing set was used to evaluate the RMSE in energies and forces, as reported in Table 1 and Fig. 1. The LJ and Morse potentials were optimized using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm as implemented in the SciPy library. The SNAP and qSNAP potentials were retrained using the MAterials Machine Learning (MAML) package. The GAP potential was retrained using the QUantum mechanics and Interatomic Potentials (QUIP) package. We used the EAM4 potential, obtained from the NIST Interatomic Potential Repository, without modification. The MTP potential was fit using the MLIP package.

The size of the training set was selected to represent common, data-scarce scenarios. With a larger training set, the MTP, SNAP, qSNAP, and GAP models would likely produce better predictions. We refer the reader to the original works and existing benchmarks for details regarding convergence with training examples and model complexity.

**Calculations**

All potentials in Fig. 1 were benchmarked using one thread on an AMD EPYC 7702 Rome (2.0 GHz) CPU. Computational cost measurements for each potential are reported as the average over ten simulations.

We use two methods to estimate the computational cost of the UF2,3 potential and present both values in Fig. 1 and Table 1. The lower estimate, 0.76 ms/step, is computed by multiplying the computational cost of UF2 by the ratio of floating point
operations used by UF$_{2}$ and UF$_{2}$. The higher estimate, 2.03 ms/step, is computed using the ratio of speeds in the Python implementation multiplied by the reported cost of UF$_{2}$. We show performance bounds instead of the current UF$_{2,3}$ implementation’s computational cost because it has not been fully optimized yet.

Elastic constants were evaluated using the Elastic package. Phonon spectra were evaluated using the Phonopy package. Melting temperatures were calculated in LAMMPS using the two-phase method and a time step of 1 fs. The initial system, a $16 \times 8 \times 8$ bcc supercell (2048 atoms), was equilibrated at a selected temperature for 40,000-time steps. Next, the solid-phase atoms were fixed while the liquid-phase atoms were heated to 5000 K and cooled back to the initial temperature over 80,000 time steps. Finally, the system was equilibrated over 200,000 steps using the isenthalpic-isobaric (NPH) ensemble. If the final configuration did not contain both phases, the procedure was repeated with a different initial temperature. Reported melting temperatures were computed by taking the average over the final 100,000 steps.

DATA AVAILABILITY
Example notebooks, LAMMPS input files, and parameters for all potentials fit in this work are available in our open-source “Ultra-Fast Force Fields” GitHub repository. The UF$_{2}$ potential is natively supported by LAMMPS, GROMACS, and other molecular dynamics suites that can construct potentials from interpolation tables. In LAMMPS, the pair-style “table” is available for execution on both CPU and GPUs, enabling the UF$_{2}$ potential to benefit from various computer architectures. A LAMMPS package for using UF$_{2,3}$ potentials is also available in the repository.

Received: 1 October 2021; Accepted: 23 July 2023; Published online: 02 September 2023

REFERENCES
1. Deringer, V. L., Caro, M. A. & Csányi, G. Machine learning interatomic potentials as emerging tools for materials science. Adv. Mater. 31, 1902765 (2019).
2. 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).
3. Miksch, A. M., Morawietz, T., Kästner, J., Urban, A. & Arth, N. Strategies for the construction of machine-learning potentials for accurate and efficient atomic-scale simulations. Mach. Learn. Sci. Tech. 2, 031001 (2021).
4. Friederich, P., Häse, F., Poppe, J. & Aspuru-Guzik, A. Machine-learned potentials for next-generation matter simulations. Nat. Mater. 20, 750–761 (2021).
5. Dautz, R., Fähnle, M. & Sanchez, J. M. General relations between many-body potentials and cluster expansions in multicomponent systems. J. Phys. Condens. Matter 17, 263–270 (2013).
6. Martinez, J. A., Yilmaz, D. E., Liang, T., Sinnott, S. B. & Phillpot, S. R. Fitting empirical potentials: challenges and methodologies. Curr. Opin. Solid State Mater. Sci. 17, 263–270 (2013).
7. Ragasa, E. J., O’Brien, C. J., Hennig, R. G., Foiles, S. M. & Phillpot, S. R. Multi-objective optimization of interatomic potentials with application to MgO. Model. Simul. Mater. Sci. Eng. 27, 074007 (2019).
8. Jones, J. E. On the determination of molecular fields—I. from the variation of the viscosity of a gas with temperature. Proc. R. Soc. Lond. A 106, 441–462 (1924).
9. Morse, P. M. Diatomic molecules according to the wave mechanics. II. vibrational levels. Phys. Rev. 34, 57–64 (1929).
10. 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).
11. Musli, F. et al. Physics-inspired structural representations for molecules and materials. Chem. Rev. 121, 9759–9815 (2021).
12. Deringer, V. L. et al. Gaussian process regression for materials and molecules. Chem. Rev. 121, 10073–10141 (2021).
13. Huang, B. & von Lilienfeld, O. A. Ab initio machine learning in chemical compound space. Chem. Rev. 121, 10001–10036 (2021).
14. Behler, J. Four generations of high-dimensional neural network potentials. Chem. Rev. 121, 10037–10072 (2021).
15. Unke, O. T. et al. Machine learning force fields. Chem. Rev. 121, 10142–10186 (2021).
16. Pasaefard, B. et al. An assessment of the structural resolution of various fingerprinters commonly used in machine learning. Mach. Learn. Sci. Technol. 2, 015018 (2021).
17. Zuo, Y. et al. Performance and cost assessment of machine learning interatomic potentials. J. Phys. Chem. A 124, 731–745 (2020).
18. Lysogorskiy, Y. et al. Performant implementation of the atomic cluster expansion (PACE) and application to copper and silicon. npj Comput. Mater. 7, 97 (2021).
19. Kovács, D. P. et al. Linear atomic cluster expansion force fields for organic molecules: beyond RMSE. J. Chem. Theor. Comput. 17, 7696–7711 (2021).
20. 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).
21. Wood, M. A. & Thompson, A. P. Extending the accuracy of the SNAP interatomic potential formalism. J. Chem. Phys. 148, 241721 (2018).
22. 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).
23. Shapeev, A. V. Moment tensor potentials: a class of systematically improvable interatomic potentials. Multiscale Model. Simul. 14, 1153–1173 (2016).
24. Dautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. Phys. Rev. B 99, 141049 (2019).
25. van der Oord, C., Dusserg, G., Csányi, G. & Orter, C. Regularized atomic-body-ordered permutation-invariant polynomials for the construction of interatomic potentials. Mach. Learn. 1, 015004 (2020).
26. Lindsey, R. K., Fried, L. E. & Goldman, N. CHiMES: a force matched potential with explicit three-body interactions for molten carbon. J. Chem. Theory Comput. 13, 6222–6229 (2017).
27. Vandermousse, J. et al. On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events. npj Comput. Mater. 6, 20 (2020).
28. Pozdyakov, S., Oganov, A. R., Mazhnik, E., Mazitov, A. & Kruglov, I. Fast general two- and three-body interatomic potential. Phys. Rev. B 107, 125160 (2023).
29. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. Phys. Rev. Lett. 98, 146401 (2007).
30. Birchar, M. P., Singraber, A. & Delligo, C. Improved description of atomic environments using low-cost polynomial functions with compact support. Mach. Learn. 2, 035026 (2021).
31. S.R. Xie et al. LAMMPS—a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. Comput. Phys. Commun. 271, 108171 (2022).
32. de Boor, C. A Practical Guide to Splines (Springer, New York, 1978).
33. Runge, C. Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten. Z. Math. Phys. 46, 224–243 (1901).
34. Wolff, D. & Rudd, W. Tabulated potentials in molecular dynamics simulations. Comput. Phys. Commun. 120, 20–32 (1999).
35. Wen, M., Whalen, S. M., Elliott, R. S. & Tadmor, E. B. Interpolation effects in tabulated interatomic potentials. Model. Simul. Mater. Sci. Eng. 23, 074008 (2015).
36. Hennig, R., Lenosky, T., Trinkle, D., Rudin, S. & Wilkins, J. Classical potential describes martensitic phase transformations between the α, β, and ω titanium phases. Phys. Rev. B 78, 054121 (2008).
37. Whitaker, E. T. On a new method of graduation. Proc. Edinb. Math. Soc. 41, 63–75 (1922).
38. Eilers, P. H. & Marx, B. D. Flexible smoothing with B-splines and penalties. Statist. Sci. 11, 89–121 (1996).
39. Schoenberger, I. J. Spline functions and the problem of graduation. Proc. Natl. Acad. Sci. USA 52, 947–950 (1964).
40. Reinsch, C. H. Smoothing by spline functions. Numer. Math. 10, 177–183 (1967).
ACKNOWLEDGEMENTS

S.R. Xie et al.

S.R. Xie and R.G.H. were supported by the United States Department of Energy under contract number DE-SC0020385. R.G.H. was supported by the U.S. National Science Foundation under contract number DMR 2118718. M.R. acknowledges partial support by the European Center of Excellence in Exascale Computing TREX—Targeting Real Chemical Accuracy at the Exascale; this project has received funding from the European Union’s Horizon 2020 Research and Innovation program under Grant Agreement No. 952165. Part of the research was performed while the authors visited the Institute for Pure and Applied Mathematics, which is supported by the National Science Foundation (Grant No. DMS 1404015). Computational resources were provided by the University of Florida Research Computing Center. We thank Ajinkya Hire for the implementation of UF potentials in LAMMPS and Alexander Shapeev for fitting the MTP potentials. We thank Thomas Bischoff, Jason Gibson, Bastian Jackl, Hendrik Krall, Ming Li, Johannes Margraf, Paul-Rene Mayer, Pawam Prakash, Robert Schmid, and Benjamin Walls for testing of and contributing to the UF implementation.

AUTHOR CONTRIBUTIONS

All authors contributed extensively to the work presented in this paper. S.R.X, M.R., and R.G.H. jointly developed the methodology. S.R.X. implemented the algorithm and performed the model training and analysis. S.R.X, M.R., and R.G.H. wrote the paper.

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/s41524-023-01092-7.

Correspondence and requests for materials should be addressed to Richard G. Hennig.

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.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023