

<https://doi.org/10.1038/s41524-024-01251-4>

Accurate machine learning force fields via experimental and simulation data fusion

Check for updates

Sebastien Röcken¹ & Julija Zavadlav^{1,2}✉

Machine Learning (ML)-based force fields are attracting ever-increasing interest due to their capacity to span spatiotemporal scales of classical interatomic potentials at quantum-level accuracy. They can be trained based on high-fidelity simulations or experiments, the former being the common case. However, both approaches are impaired by scarce and erroneous data resulting in models that either do not agree with well-known experimental observations or are under-constrained and only reproduce some properties. Here we leverage both Density Functional Theory (DFT) calculations and experimentally measured mechanical properties and lattice parameters to train an ML potential of titanium. We demonstrate that the fused data learning strategy can concurrently satisfy all target objectives, thus resulting in a molecular model of higher accuracy compared to the models trained with a single data source. The inaccuracies of DFT functionals at target experimental properties were corrected, while the investigated off-target properties were affected only mildly and mostly positively. Our approach is applicable to any material and can serve as a general strategy to obtain highly accurate ML potentials.

With their ability to accelerate the discovery of new materials and decipher the properties of existing materials, Molecular Dynamics (MD) simulations have become a cornerstone of material science¹. Nevertheless, the true capability is often hindered by the accuracy vs. efficiency trade-off of traditional approaches. Ab initio MD provides high-accuracy predictions at low computational efficiency, while the contrary holds for the MD simulations based on classical force fields. In theory, Machine Learning (ML) approaches^{2,3} and, in particular, ML potentials^{4–8} can overcome this compromise due to the multi-body construction of the potential energy with unspecified functional form. In practice, the success of ML potentials hinges primarily on the training data, the source of which can be either simulations or experiments, or both.

Typically the former source is used with ab initio calculations providing energy, forces, and potentially virial stress (target labels) for different atomic configurations (inputs)^{9–16}. Such a setup, also known as bottom-up learning, has the benefit of straightforward training and should result in ML potentials that reproduce all properties of the underlying model. However, generating ab initio training data that is sufficiently accurate, large, and broad (without distribution shift) is challenging.

CCSD(T) (coupled cluster with single, double, and perturbative triple excitations) method, regarded as the gold standard of electronic structure theory, is generally computationally infeasible for large dataset generation. Thus, most ML potentials are trained on the more affordable but less accurate

Density Functional Theory (DFT) calculations. These are not always in quantitative agreement with experimental predictions, and consequently, neither are ML potentials trained on DFT data. For example, a recent ML-based model of titanium¹⁷ does not quantitatively reproduce the experimental temperature-dependent lattice parameters and elastic constants. For these properties, it achieved a similar level of agreement with experiments as the classical MEAM (modified embedded atom method) potential¹⁸. Deviations in the phase diagram predictions are also frequent^{19–22}. In all cases, these deviations were attributed to DFT inaccuracies. To approach the CCSD(T) level accuracy, transfer learning²³ or Δ -learning²⁴ techniques, exploiting a large DFT and a small CCSD(T) dataset, can be used.

Nevertheless, DFT training data is, albeit cheaper, still computationally expensive, and an optimal selection of atomic configurations is needed for diverse and non-redundant training data. Typically, training datasets are carefully prepared and contain specialized sub-datasets based on the target application, such as surfaces, defects, lattice distortions, thermal displacements, configurations along the phase transformation pathways, etc.^{25–28} Alternatively, an active learning approach^{29–33} is used, where the dataset is increased on the fly during training. These methods require a robust uncertainty quantification scheme, which remains problematic for Neural Network (NN)-based potentials^{34–39}.

Apart from the dataset size, the system size (number of atoms per configuration) can also play a significant role in the optimal model

¹Department of Engineering Physics and Computation, Multiscale Modeling of Fluid Materials, TUM School of Engineering and Design, Technical University of Munich, Munich, Germany. ²Munich Data Science Institute, Technical University of Munich, Munich, Germany. ✉e-mail: julija.zavadlav@tum.de

components and hyperparameters and, consequently, the resulting trained model⁴⁰. Due to the cubic scaling of DFT implementations, the average number of atoms is typically below one hundred for dense systems under periodic boundary conditions. It is questionable whether long-range interactions⁴¹ can be learned from such databases, considering the recent finding that features related to interatomic distances as large as 15 Å can play an essential role in describing non-local interactions⁴².

The difficulties of ab initio data generation can be circumvented if ML potentials are instead trained top-down, i.e., on experimental data^{43–46}. While experimental data is also scarce, potentially laborious to obtain, and contains measurement errors, the obtained information per data sample is much larger compared to bottom-up learning. Experimentally observable properties of a system are in simulations computed as an ensemble average, i.e., averaged over a very large number of atomic configurations. This fact also complicates training since it requires running forward simulations to calculate the properties and, in principle, subsequent gradient backpropagation through the simulation. Automatic differentiation⁴⁷ and recent end-to-end differentiable software^{48–50} have made such endeavors technically possible. In practice, backpropagation through the simulation is unfeasible for properties that require long simulations due to issues such as memory overflow, exploding gradients, and high computational costs^{43,51,52}. However, for time-independent properties, these issues can be avoided with the Differentiable Trajectory Reweighting (DiffTRe) method⁴³ that, rather than backpropagating through the trajectory, employs a reweighting technique. For a test case diamond system, the method yielded an ML potential that reproduced the target experimental mechanical properties at ambient conditions. Yet, for out-of-target phonon density of states, substantially different results were obtained for different random initializations, showcasing that the high-capacity ML potentials are under-constrained when trained on a handful of experimental observations⁴³. Combining both simulation and experimental data sources, an idea used for decades to construct classical force fields⁵³, should, therefore, yield the best approach also for ML potentials. This idea was recently used also in ref.⁵⁴ where a two-body correction trained on

structural experimental data was added to a fixed ML potential trained on DFT data. However, such Δ -learning approach is limited as two-body potentials cannot reproduce many experimental observables simultaneously. On the other hand, replacing a two-body potential with another deep ML potential would double the computational cost.

In this work, we demonstrate the benefits of training a single deep ML potential to simultaneously reproduce simulation and experimental data. In particular, we train a Graph Neural Network (GNN) potential for titanium on DFT calculated energies, forces, and virial stress for various atomic configurations and experimental mechanical properties and lattice parameters of hcp titanium in the temperature range of 4 to 973 K. We then test the resulting model that faithfully reproduces all target properties on several out-of-target properties, i.e., phonon spectra, bcc titanium mechanical properties, and liquid phase structural and dynamical properties. We find that the out-of-target properties are only mildly and mostly positively affected by the combined training approach, revealing a remarkably large capacity of the state-of-the-art ML potentials.

Results

Fused data training approach

A concurrent training on the DFT and experimental data can be achieved by iteratively employing both a DFT trainer and an EXP trainer (Fig. 1). The former involves a standard regression problem. The ML potential takes as an input atomic configuration S and predicts the potential energy U from which the forces on all atoms F and virial stress tensor V are computed by differentiating with respect to atoms' positions. The parameters θ are modified using batch optimization for one epoch to match the ML potential's predictions and the target values in the DFT database. We reuse the previously published DFT calculations for titanium^{17,55}. The DFT database consists of 5704 samples. It includes equilibrated, strained, and randomly perturbed hcp, bcc, and fcc titanium structures, as well as configurations obtained via high-temperature MD simulations and an active learning approach. Further details are in the Supplementary Information.

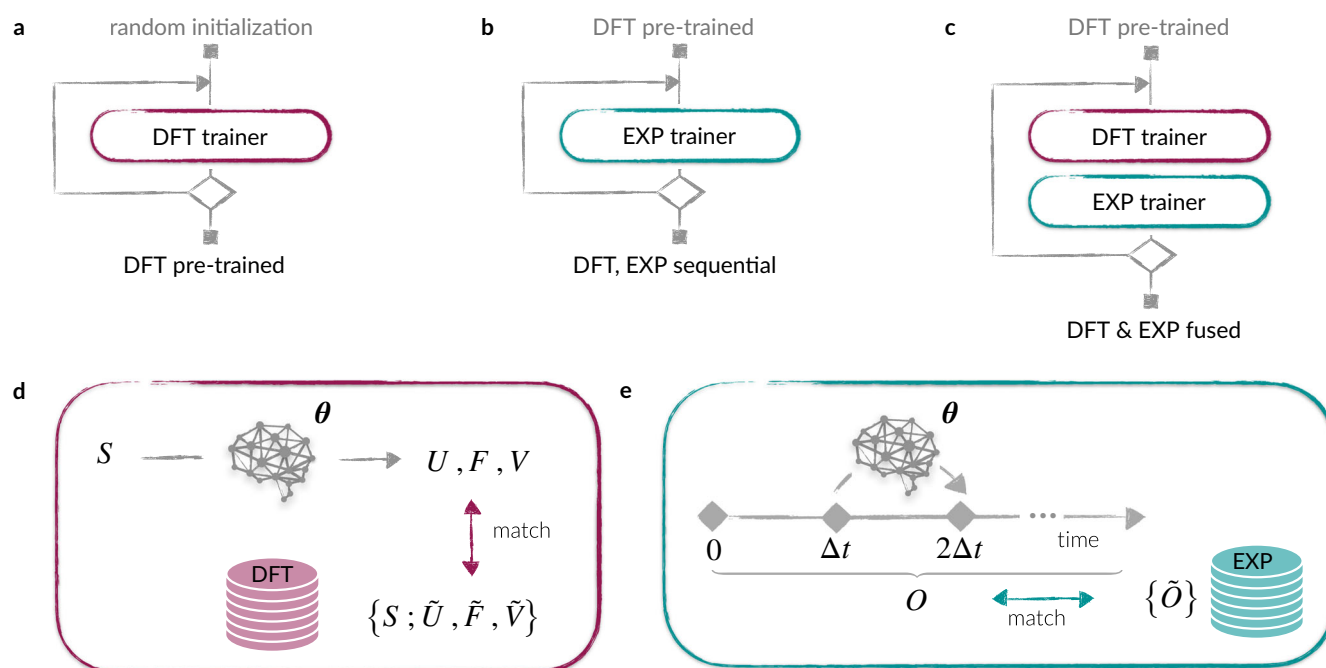


Fig. 1 | Investigated models. The DFT pre-trained model (a) is trained only with DFT trainer (d), which optimizes the parameters of the ML potential to match the reference DFT potential energy \tilde{U} , forces \tilde{F} , and virial \tilde{V} for different atomic environments S . For the DFT, EXP sequential model (b), the ML potential is initialized with the parameters of the DFT pre-trained model and trained with EXP trainer (e), where the ML potential is trained to reproduce experimental observables

\tilde{O} . EXP trainer requires simulations since the observables are not a direct output of the ML model but computed as a time average over the simulated trajectory. The DFT & EXP fused model (c) is obtained by alternating between the DFT and EXP trainers, starting from the DFT pre-trained model. In all cases, the DFT and/or EXP trainers are repeatedly applied for one epoch until the total number of epochs is reached.

The EXP trainer, on the other hand, performs optimization of parameters θ for one epoch such that the properties of titanium (observables) computed from the ML-driven simulation's trajectory match experimental values where the gradients are computed with the DiffTRe method⁴³. We consider temperature-dependent, solid-state elastic constants of hcp titanium as target experimental properties. Elastic constants of titanium were measured experimentally at 22 different temperatures in the range of 4–973 K⁵⁶. Nevertheless, we select only the following four temperatures: 23, 323, 623, and 923 K for the experimental training database. With this choice, we reduce the computational cost per epoch and include our expectation that the models will be, to some degree, temperature transferable. The elastic constants are evaluated in the NVT ensemble, where the box size is set according to the experimentally determined lattice constants⁵⁷ (see Supplementary Information). Thus, by adding the additional target of zero pressure, we indirectly match also the experimental lattice constants.

To investigate the impact of DFT and EXP trainers, we compare three different approaches: (i) the DFT pre-trained model, employing only the DFT trainer (ii) the DFT, EXP sequential model, employing only the EXP trainer, and (iii) the DFT & EXP fused model, obtained with the alternating use of the DFT and EXP trainers. The switching between the trainers is performed after processing all respective training data, i.e., after one epoch. Alternatively, a batch-wise switching could be employed. For the last two approaches, the parameters of the ML potential are not initialized randomly but with the values of the DFT pre-trained model. This allows us to circumvent the use of prior potentials, typical for top-down learning^{43,52}. The prior potentials are simple classical potentials added to the ML potential to avoid unphysical trajectories and, therefore, slow learning in the initial learning stage. The models are trained for a fixed number of epochs, and the final model is selected with early stopping. For further information, see Supplementary Information.

Table 1 | Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) of energy, force, and virial predictions computed on the test DFT dataset

Approach	RMSE/MAE		
	Energy [meV atom ⁻¹]	Force [meV Å ⁻¹]	Virial [meV atom ⁻¹]
DFT pre-trained	6.0/4.4	92.5/62.1	406.6/261.2
DFT, EXP sequential	385.1/384.9	123.6/83.8	401.5/267.1
DFT & EXP fused	7.9/6.2	111.2/76.6	405.6/263.2

Simultaneously learning DFT and experimental target properties

We compute the energy, force, and virial errors on the DFT test dataset (Table 1) for all three investigated models. For the DFT pre-trained model, the obtained energy error is below 43 meV, generally accepted within the chemistry community as the chemical accuracy⁵⁸. In Supplementary Table 3, we additionally show the errors for a portion of the test dataset containing only strained and perturbed hcp or bcc samples. The force and virial errors are an order of magnitude lower when high-temperature configurations are excluded. This difference is due to larger force magnitudes in high-temperature configurations. Indeed, the force relative errors are similar for all test datasets (Supplementary Table 4). We compare favorably with the previously published ML-based potential model¹⁷ for the force errors, while the energy errors are somewhat higher. However, precedence can be given to energy, force, or virial error by changing the weights of the loss function (Eq. (1)). We give a higher emphasis on the forces as these are relevant for carrying out MD simulations.

When training on both DFT and experimental data (DFT & EXP fused model), the errors are only slightly increased compared to training only on DFT data (DFT pre-trained model). An increase is expected as the model has to satisfy both DFT and experimental objectives, which are partially conflicting due to the DFT inaccuracies as well as experimental errors. The fact that the errors do not change drastically indicates that the DFT errors in energy, force, and virial predictions are minor. Nevertheless, a small difference in force prediction can amount to large differences in MD simulations and subsequent evaluation of properties, as we later show for the mechanical properties.

For the DFT, EXP sequential model, the force and virial errors are still comparable to the DFT pre-trained model, but the energy error is drastically increased. The Supplementary Fig. 2 shows the energy RMSE during training. This is not surprising considering that MD simulations and our target experimental properties do not depend on energy but only on its derivatives. The EXP trainer, therefore, leaves the energy undetermined up to a constant, as confirmed by the predicted vs. DFT energy plot (Supplementary Fig. 1). Consequently, any energy-related quantity will also be predicted incorrectly. For example, the energy versus volume equation of state curves for hcp, fcc, and bcc structures are all shifted by a constant and equal value (Fig. 2). Nevertheless, this shift can be evaluated in post analysis. In particular, we compute the mean energy shift in the training dataset and apply it to the test dataset. With this correction, the energy RMSE and MAE are 14.0 and 9.5 meV atom⁻¹, respectively. The errors are slightly higher but comparable to the errors of the DFT & EXP fused model. The DFT, EXP sequential model demonstrates the importance of including DFT data in training, especially when the experimental dataset does not include properties directly related to energies.

Next, we evaluate the elastic constants of hcp titanium (Supplementary Fig. 3), which are the target properties of the EXP trainer. Additionally, we

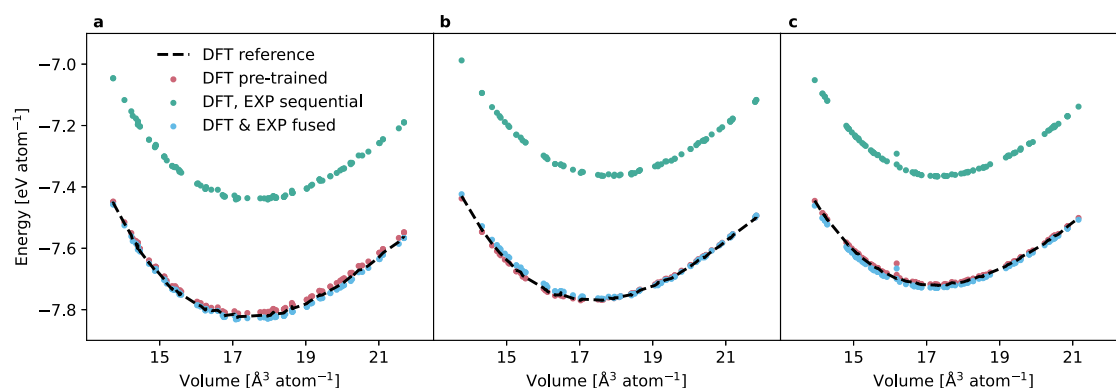


Fig. 2 | Agreement with DFT data. Energy vs. volume for the hcp (a), fcc (b), and bcc (c) titanium crystal structures computed for samples in the test DFT dataset. The predictions of the DFT pre-trained, DFT, EXP sequential, and DFT & EXP fused

models are denoted with red, green, and blue points, respectively. DFT calculations are denoted with a black dashed line.

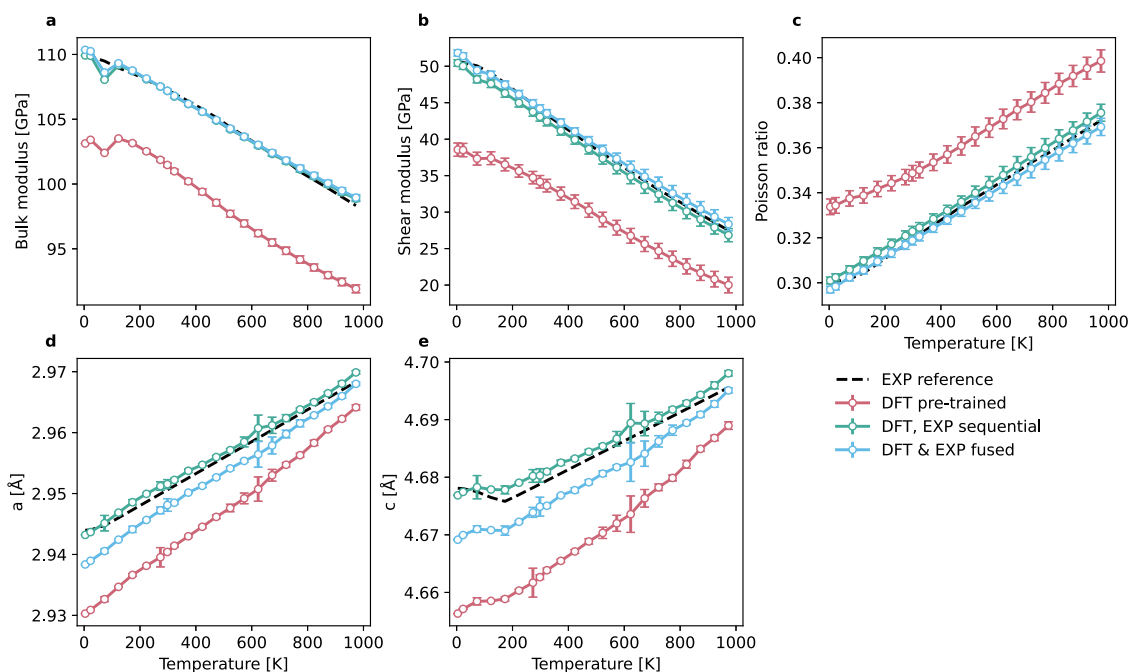


Fig. 3 | Agreement with EXP data. Bulk modulus (a), shear modulus (b), Poisson’s ratio (c), and lattice constants *a* (d) and *c* (e) as a function of temperature for hcp titanium. The DFT pre-trained, DFT, EXP sequential, and DFT & EXP fused models are denoted with red, green, and blue line points, respectively. The experimental results are denoted with a black dashed line. Error bars denote the standard deviation computed via block-averaging with ten blocks.

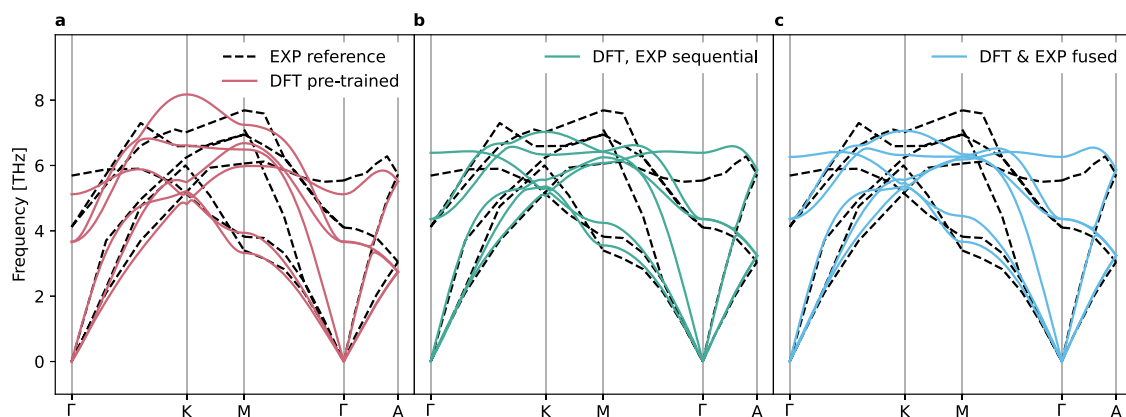


Fig. 4 | Off-target solid state property. Phonon dispersion curves of hcp titanium for DFT pre-trained (a), DFT, EXP sequential (b), and DFT & EXP fused (c) models. The ML potential models’ predictions match well the black dashed lines denoting the experimental prediction measured at 295 K²⁹.

report in Fig. 3a–c the bulk modulus, shear modulus, and Poisson’s ratio, which are all directly related to elastic constants. These properties are computed for all 22 temperatures in the range of 4–973 K where experimental data is available. Training only on DFT data (DFT pre-trained model) fails to reproduce the mechanical properties. On average, the model deviates from the experimental data by 6, 24, and 9% in bulk modulus, shear modulus, and Poisson’s ratio, respectively (Supplementary Table 5). In terms of elastic constants, the predictions are for some components off by more than 20 GPa. Similar deviations in mechanical properties were reported for other ML potentials^{17,19,22}. Per contra, for the two models that include the EXP trainer, the elastic constants are within a few GPa of the experimental values, while the relative errors for the bulk modulus, shear modulus, and Poisson’s ratio are below 3%. We obtain a good agreement with experimental observations on the entire investigated temperature range, even though we fit the elastic constants only at four temperatures. Naturally, the agreement is better for the DFT, EXP sequential model because the DFT and experimental datasets are erroneous and, thus, somewhat incompatible.

An additional target property of the EXP trainer is zero pressure (Supplementary Fig. 4) at fixed, experimentally determined simulation box sizes. In Fig. 3d, e, we show an equivalent result, i.e., the temperature-dependant lattice constants evaluated in the isothermal-isobaric ensemble. Similarly, as for the mechanical properties, the addition of the EXP trainer improves the results for both target and non-target temperatures, with the DFT, EXP sequential model being the closest to experimental reference values. Note that the DFT & EXP fused model’s relative deviations from experimental values are below 0.1%, i.e., smaller than deviations in mechanical properties (Supplementary Table 5).

Generalization to off-target properties and thermodynamic states

As a first test of the generalization capabilities to off-target properties, we compute the phonon spectra of hcp titanium (Fig. 4). All models agree well with experimental prediction, with the DFT pre-trained model in closest agreement based on the phonon density of states (Supplementary Fig. 5). Good agreement is expected for the DFT pre-trained and DFT & EXP fused

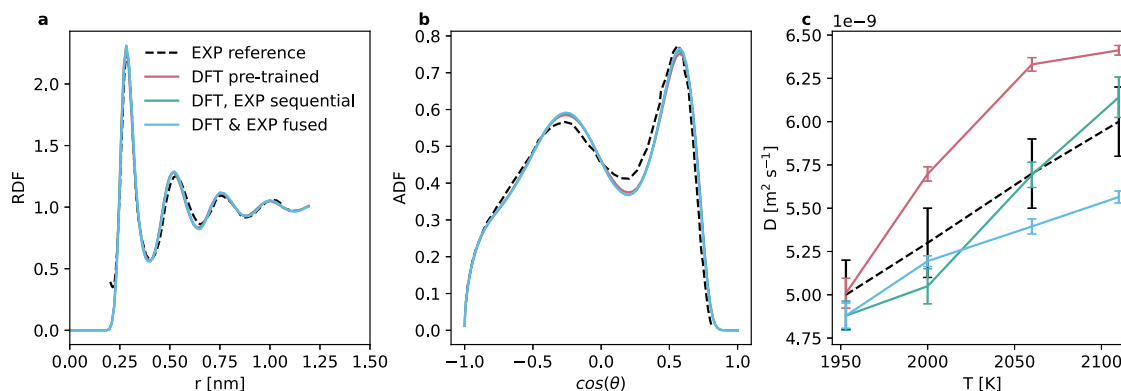


Fig. 5 | Off-target liquid state properties. Radial distribution function (RDF, **a**), angular distribution function (ADF, **b**), and self-diffusion coefficients (D , **c**) for the DFT pre-trained (red), DFT, EXP sequential (green), and DFT & EXP fused (blue) models. The RDFs and ADFs are computed at 1965 K and compared with

experimentally determined RDF⁸⁰ at 1965 K and ADF⁸¹ at 1973 K (black, dashed). The self-diffusion is evaluated at 1953, 2000, 2060, and 2110 K for comparison to experiments^{82–84}. The experimental error bars are estimated based on the experimental error bar at 2000 K⁸⁴.

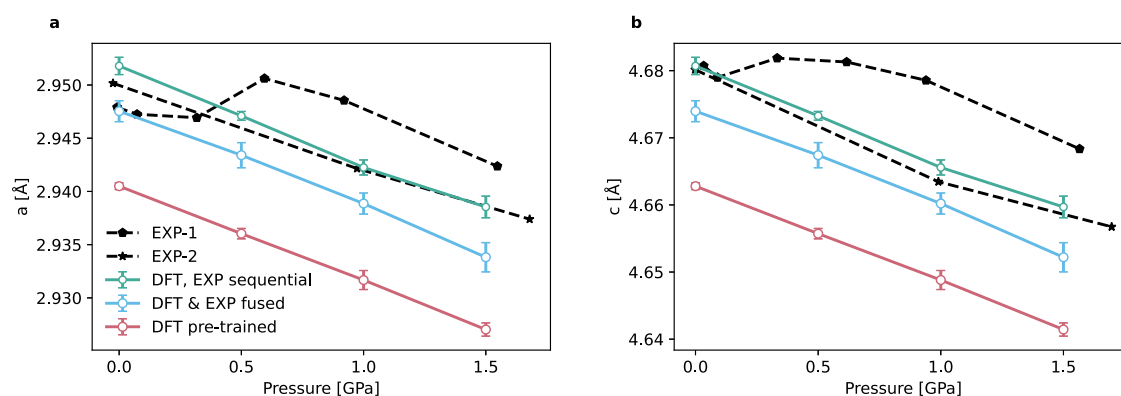


Fig. 6 | Off-target thermodynamic states. Lattice constants a (**a**) and c (**b**) of hcp titanium for varying pressures at temperature 300 K. The DFT pre-trained (red), DFT, EXP sequential (green), and DFT & EXP fused (blue) models are compared to experimental values EXP-1 (SPRING-8 data) and EXP-2 (NLSLS data)⁸⁵.

models since ML potentials trained on DFT data typically reproduce the phonon dispersion curves well and much better than the classical potentials^{17,20,27,59}. Interestingly, we obtain a good agreement also for the DFT, EXP sequential model. Our previous study⁴³ showed that training randomly initialized ML potentials on mechanical properties leads to models with drastically different phonon densities of states, i.e., high-capacity models are underconstrained when trained on a small set of target properties. Additional properties could be included to converge toward a unique potential energy solution. However, the required experimental database size is unknown a priori. Our results in Fig. 4b indicate an alternative route. Pretraining on DFT data seems to constrain the solution to a particular region in parameter space which is only locally modified by the subsequent training on the experimental data. This hypothesis is in accordance with the observed similar force errors for DFT pre-trained and DFT, EXP sequential models (Table 1).

To further validate our rationale, we examine the liquid-state titanium's structural and dynamical properties. Two-body and three-body local structural order is measured with radial distribution function (RDF) and angular distribution function (ADF). For all investigated models, the results are indistinguishable within the line thickness (Fig. 5a, b). Moreover, the obtained RDFs are very close to the experimental measurement. For ADFs, the position of the minima and maxima agrees very well with the experiments, while the absolute values slightly differ. The largest deviation for the three models is observed in Fig. 5c, which presents the self-diffusion coefficients calculated via the velocity autocorrelation function (Supplementary Fig. 6). Both models trained on experimental data yield better results on

average than the DFT-pretrained model. The DFT, EXP sequential model performs best.

Next, we consider generalization to different pressures. To this end, we compute the lattice constants of hcp titanium at the temperature 300 K and elevated pressures (Fig. 6). Similarly, as in the case of diffusion, we find the closest agreement with experimental values for the DFT, EXP sequential model. However, such an outcome is not always guaranteed, as we show next.

We evaluate all three models on the bcc elastic constants at 1273 K (Table 2). Concrete conclusions are difficult given significant deviations between the three experimental references at equal or similar temperature. Nevertheless, assuming that the latest experimental results by Ledbetter et al.⁶⁰ are the most accurate, the DFT & EXP fused model is best overall. In

Table 2 | Elastic constants in GPa of bcc titanium at 1273 K

Approach	C_{11}	C_{12}	C_{44}
DFT pre-trained	98.4	79.4	27.4
DFT, EXP sequential	119.9	87.4	34.9
DFT & EXP fused	112.0	85.3	32.5
EXP Ledbetter et al. ⁶⁰	97.7	82.7	37.5
EXP Petry et al. ⁷⁷	134.0	110.0	36.0
EXP Fischer et al. ⁷⁸	99.0	85.0	33.6

The experimental refs. ^{60,78} are at 1273 K, while ref. ⁷⁷ is at 1293 K.

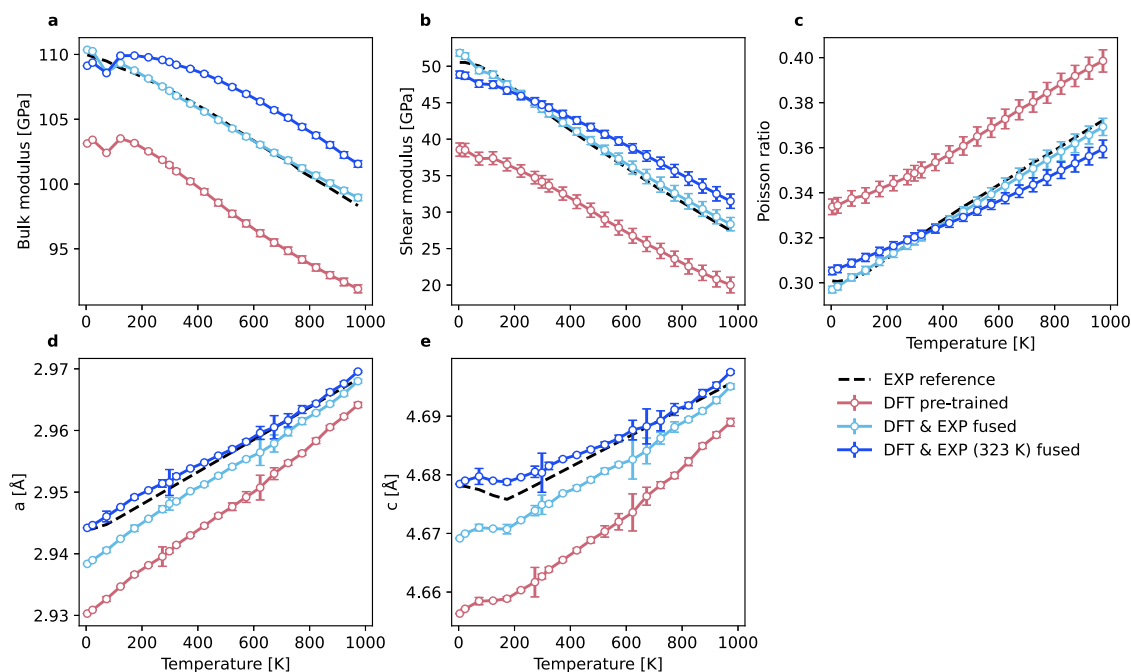


Fig. 7 | Data ablation study. Bulk modulus (a), shear modulus (b), Poisson's ratio (c), and lattice constants a (d) and c (e) as a function of temperature for hcp titanium. The DFT pre-trained, DFT & EXP fused, and DFT & EXP (323 K) fused models are denoted with red, blue, and dark blue line points, respectively. The last two models

differ only in experimental training data, i.e., the DFT & EXP (323 K) fused model is trained on data at a single temperature of 323 K. The experimental reference values are marked with a black dashed line. Error bars denote the standard deviation computed via block-averaging with ten blocks.

particular, it performs best on C_{12} and second best on C_{11} and C_{44} . Note that when training on the EXP database, the bcc lattice is never seen, while the DFT dataset also contains the equilibrated, strained, and perturbed bcc structures.

Experimental data ablation

Lastly, we consider a data ablation study. An additional model, labeled DFT & EXP (323 K) fused, is trained with the same approach as the DFT & EXP fused model, but with experimental training data containing elastic constants and pressure only at a single temperature of 323 K. The aim is to reveal the effect of experimental data size as well as the model's temperature transferability. As shown in Fig. 7, the DFT & EXP (323 K) fused model yields improved mechanical properties and lattice parameters on the entire temperature range compared to training only on DFT data, i.e., DFT pre-trained model. The predicted elastic constants are shown in Supplementary Fig. 7. However, as expected, the mechanical properties are not as accurate as training on experimental data at four different temperatures (DFT & EXP fused model trained at 23, 323, 623, 923 K). In general, due to the temperature transferability of the models, it seems more beneficial to enlarge the experimental dataset with diverse properties rather than with a single property at densely sampled temperatures.

Discussion

Using titanium as a test case system, we have demonstrated the advantages of using both experimental and simulation data to train ML potentials. We tested two strategies of employing the DFT and experimental data, i.e., sequential and fused, and referenced them against using only DFT data. Note that training only on experimental data is difficult without a prior potential and was therefore not attempted.

The addition of experimental data resulted in ML potentials that reproduced target experimental properties, thus correcting for the inaccuracies of the DFT calculations and limited DFT training dataset. Moreover, some of the off-target properties (e.g., diffusion) improved even though the relevant (e.g., liquid) configurations were never seen by the EXP trainer.

On the other hand, pretraining on the DFT data has the effect of regularizing the solution, evidenced by very similar or only mildly different out-of-target properties. This is especially important when the experimental dataset is scarce. As we have shown previously⁴³, ML potentials fitted only on a handful of observations can substantially differ on out-of-target properties due to the large capacity of these models. In general, top-down training lacks theoretical guarantees of bottom-up approaches and can result in deteriorated out-of-target properties. For this reason, we advocate for the DFT & EXP fused approach rather than the DFT, EXP sequential approach, even though the latter performed better on some out-of-target properties. With minimal computational overhead, the fused training ensures that the solution remains close to the DFT solution, which might deviate from experiments somewhat but not drastically. Furthermore, experimental measurements also contain errors, and conflicting results might be reported in the literature, e.g., mechanical properties of bcc titanium [67]. The DFT & EXP fused approach can, therefore, to some extent overcome the deficiencies of pure bottom-up or top-down training.

In this paper, the experimental properties were elastic and lattice constants. However, the DiffTRe approach is general, and, in principle, any other static structural or thermodynamic property could be used⁴³. In practice, training on properties requires running simulations, the spatio-temporal scales of which should be sufficiently large to reasonably estimate the target properties and, consequently, obtain informative gradients. Thus, observables involving rare events might be out of reach for conventional computational resources.

The number of required simulation runs can be reduced with reweighting techniques. DiffTRe method employs the simplest Zwanzig approach⁶¹ that reweights observables from a single reference state. In this work, simulations were initialized at every parameter update to avoid an additional layer of complexity. Nevertheless, the reweighting ansatz is still used to provide a relation between the observables and the parameters of the ML potential, enabling a direct route to the gradient computation. Other reweighting approaches could also be employed. For example, the multistate Bennett acceptance ratio (MBAR)^{62–64}, where information from multiple states is used to probe the configuration space of the unsampled state⁶⁵. Note

that the computational overhead of evaluating the potential energy for multiple states is minor compared to forward simulations. Multistate reweighting techniques are typically more accurate in estimating ensemble averages and could provide more accurate gradients. On the other hand, deep ML methods sometimes benefit from noisy gradients⁶⁶. Additionally, all reweighting methods require sufficient configuration overlap, and choosing appropriate reference states is a non-trivial task. Therefore, the best reweighting technique is an open question that we leave for future work.

Methods

ML potential architecture

We employ a message passing GNN DimeNet++¹¹ using our implementation in JaxMD⁴³, which takes advantage of neighbor lists for efficient computation of the sparse atomic graph. We select the same neural network hyperparameters (Supplementary Table 1) as in the original publication¹¹ except for the embedding sizes, which we reduced by factor 4 for computational speed-up. The cut-off is set to 0.5 nm.

DFT trainer

We use a weighted mean squared error loss function

$$L_{\text{DFT}} = \frac{1}{N_{\text{data}}} \sum_{i=1}^{N_{\text{data}}} \left[\omega_U (U_i - \tilde{U}_i)^2 + \frac{\omega_F}{3N_{\text{atoms}}} \sum_{j=1}^{N_{\text{atoms}}} \sum_{k=1}^3 (F_{ijk} - \tilde{F}_{ijk})^2 + \frac{\omega_V}{9} \sum_{k=1}^3 \sum_{l=1}^3 (V_{ikl} - \tilde{V}_{ikl})^2 \right] \quad (1)$$

where U_i is the energy of the i -th atomic environment in a batch, F_{ijk} is the force in the k -direction of the j -th atom, and V_{ikl} is the virial in the k,l -direction. The reference DFT values are denoted with $\tilde{\cdot}$. The weights for the energy and force are set to $\omega_U = 1e^{-6}$ and $\omega_F = 1e^{-2}$, while for the virial contribution, only the uniformly deformed supercells contribute with $\omega_V = 4e^{-6}$. The numerical optimization hyperparameters are reported in Supplementary Table 2.

EXP trainer

We define the loss function as

$$L_{\text{EXP}} = \frac{1}{N_{\text{temp}}} \sum_{n=1}^{N_{\text{temp}}} \sum_{m=1}^{N_{\text{obs}}} w_m (O_{m,n} - \tilde{O}_{m,n})^2 = \frac{1}{N_{\text{temp}}} \sum_{n=1}^{N_{\text{temp}}} \left[\omega_P (P_n)^2 + \frac{\omega_C}{5} \left\{ (C_{11,n} - \tilde{C}_{11,n})^2 + (C_{12,n} - \tilde{C}_{12,n})^2 + (C_{13,n} - \tilde{C}_{13,n})^2 + (C_{33,n} - \tilde{C}_{33,n})^2 + (C_{44,n} - \tilde{C}_{44,n})^2 \right\} \right], \quad (2)$$

where $O_{m,n}$ is a m -th observable at n -th temperature in a batch and $\tilde{\cdot}$ denotes the experimental value. The observables are scalar pressure P_n and elastic constants in Voigt notation C_{**} . The weights are $\omega_P = 1e^{-9}$ and $\omega_C = 1e^{-10}$. The gradient of the loss with respect to the parameters of the ML potential is obtained with the DiffTRe method⁴³, where the ensemble average of an observable $O_{m,n}$ is computed with the reweighting ansatz for the canonical ensemble^{61,67,68}

$$\langle O_{m,n}(U_\theta) \rangle \simeq \sum_{i=1}^{N_{\text{traj}}} w_i O_{m,n}(S_i, U_\theta) \quad \text{with} \quad w_i = \frac{e^{-\beta(U_\theta(S_i) - U_\theta(S_i))}}{\sum_{j=1}^N e^{-\beta(U_\theta(S_j) - U_\theta(S_j))}}. \quad (3)$$

The summation runs over the trajectory states/atomic environments S , $\beta = 1/(k_B T)$, k_B is the Boltzmann constant, and T is the temperature. U_θ and U_θ denote the reference and perturbed ML potentials. We initialize the

forward trajectory generation for every parameter update. Thus, $U_\theta = U_\theta$ and $w = 1$ for every sample. Nevertheless, $\nabla_\theta L_{\text{EXP}}$ is generally non-zero. Further details can be found in Ref. ⁴³. The numerical optimization hyperparameters are reported in Supplementary Table 2.

ML potential-driven MD simulations

All MD simulations are performed in JaxMD⁴⁸ using a velocity Verlet integrator with a time step of 0.5 fs. The simulated system contains 256 atoms unless otherwise stated. The mass of titanium atoms is set to 47.867 a.u. For NVT simulations during training and to compute the elastic constants and pressure in postprocessing, we use the Langevin thermostat with a friction constant of 4 ps^{-1} . For the remaining post-processing, we run NVT simulations using a Nose-Hoover thermostat and NPT simulations with a Nose-Hoover thermostat and barostat. For the Nose-Hoover chains, we use a chain length of 5, 2 chain steps, and 3 Suzuki-Yoshida steps, and set the thermostat damping parameter to $\tau = 50 \text{ fs}$ and the barostat damping parameter to $\tau = 500 \text{ fs}$. The pressure is set to 0.

In the EXP trainer, the elastic constants and pressure are computed from 80 ps NVT simulation, where the first 10 ps are disregarded as equilibration, and the state is saved every 0.1 ps. The isothermal elasticity tensor is computed with the stress-fluctuation method^{43,69}.

To analyze the properties of trained models, we perform the following simulations. For hcp elastic constants and pressure, we perform a 100 ps NVT equilibration run followed by a 1 ns NVT production run. As in training, the box size is set according to the experimental lattice parameters at a given temperature. The elastic constants are saved every 0.1 ps. The bulk modulus and shear modulus are computed from elastic constants (in Voigt notation)^{56,70,71} as $K = 2/9(C_{11} + C_{12} + 2C_{13} + 1/2C_{33})$ and $G = 1/30(12C_{44} + 7C_{11} - 5C_{12} + 2C_{33} - 4C_{13})$. The Poisson ratio is computed with $\sigma = (3K - 2G)/(2G + 6K)$. For hcp lattice constants, we perform a 100 ps NPT equilibration followed by a 100 ps NPT production run, where a state is saved every 0.25 ps. For phonon frequency analysis, we generate a $5 \times 5 \times 3$ hcp super cell in Avogadro⁷² and employ Phonopy^{73,74} to compute the phonon densities via finite displacements of 0.01 Å. To compute the RDF and the ADF, we perform a 100 ps NPT equilibration at 2400 K, 100 ps NPT equilibration at 1965 K, and 80 ps NVT production run at 1965 K, which we sample every 0.1 ps. For these simulations, we double the box size in each dimension, yielding a total of 2048 atoms. The ADF is computed for all triplets within 0.4 nm, corresponding to the first minimum of the experimental RDF. For VACF, we perform 100 ps NPT equilibration at 2400 K, 100 ps NPT equilibration at 2000 K, 100 ps NVT equilibration at 2000 K, and 80 ps NVT production run from which we sample every 0.01 ps. The VACF is computed by averaging over 160 different starting points that are 0.5 ps apart. We use the Green-Kurbo relation to compute the self-diffusion^{75,76}. The errors are estimated with block-averaging using 10 blocks. The bcc elastic constants were obtained by creating a bcc titanium structure with 128 atoms as input for a 100 ps NPT followed by 100 ps NVT equilibration at 1273 K, and a 1 ns NVT production run at 1273 K. We confirmed the adequateness of equilibration protocols by repeating the analysis for RDF, ADF, VACF, and high temperature hcp lattice constants with doubled NPT equilibration lengths (i.e., using 200 ps) for the DFT & EXP fused model (Supplementary Fig. 8).

Data availability

The dataset is publicly available at <https://github.com/tummfm/Fused-EXP-DFT-MLP/tree/main/Dataset>.

Code availability

The code is publicly available at <https://github.com/tummfm/Fused-EXP-DFT-MLP.git>.

Received: 17 August 2023; Accepted: 22 March 2024; Published online: 05 April 2024

References

- Pilania, G., Goldsmith, B., Yoon, M. & Dongare, A. M. Recent advances in computational materials design: methods, applications, algorithms, and informatics. *J. Mater. Sci.* **57**, 10471–10474 (2022).
- Hart, G. L., Mueller, T., Toher, C. & Curtarolo, S. Machine learning for alloys. *Nat. Rev. Mater.* **6**, 730–755 (2021).
- Vlachas, P. R., Zavadlav, J., Praprotnik, M. & Koumoutsakos, P. Accelerated simulations of molecular systems through learning of effective dynamics. *J. Chem. Theory Comput.* **18**, 538–549 (2022).
- Friederich, P., Häse, F., Proppe, J. & Aspuru-Guzik, A. Machine-learned potentials for next-generation matter simulations. *Nat. Mater.* **20**, 750–761 (2021).
- Mueller, T., Hernandez, A. & Wang, C. Machine learning for interatomic potential models. *J. Chem. Phys.* **152**, 050902 (2020).
- Mishin, Y. Machine-learning interatomic potentials for materials science. *Acta Mater.* **214**, 116980 (2021).
- Zuo, Y. et al. Performance and cost assessment of machine learning interatomic potentials. *J. Phys. Chem. A* **124**, 731–745 (2020).
- McSloy, A. et al. Tbmalt, a flexible toolkit for combining tight-binding and machine learning. *J. Chem. Phys.* **158**, 034801 (2023).
- Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
- Erhard, L. C., Rohrer, J., Albe, K. & Deringer, V. L. A machine-learned interatomic potential for silica and its relation to empirical models. *NPJ Comput. Mater.* **8**, 1–12 (2022).
- Gasteiger, J., Giri, S., Margraf, J. T. & Günnemann, S. Fast and Uncertainty-Aware Directional Message Passing for Non-Equilibrium Molecules. *Machine Learning for Molecules Workshop, NeurIPS* (2020).
- Schütt, K. et al. Schnet: a continuous-filter convolutional neural network for modeling quantum interactions. *Adv. Neural Inf. Process. Syst.* **30**, 991–1001 (2017).
- Unke, O. T. & Meuwly, M. Physnet: a neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory Comput.* **15**, 3678–3693 (2019).
- Batzner, S. et al. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).
- Musaelian, A. et al. Learning local equivariant representations for large-scale atomistic dynamics. *Nat. Commun.* **14**, 579 (2023).
- Sivaraman, G. et al. Experimentally driven automated machine-learned interatomic potential for a refractory oxide. *Phys. Rev. Lett.* **126**, 156002 (2021).
- Wen, T. et al. Specialising neural network potentials for accurate properties and application to the mechanical response of titanium. *Npj Comput. Mater.* **7**, 206 (2021).
- Lee, B.-J., Baskes, M. I., Kim, H. & Cho, Y. K. Second nearest-neighbor modified embedded atom method potentials for bcc transition metals. *Phys. Rev. B* **64**, 184102 (2001).
- Dickel, D., Francis, D. & Barrett, C. Neural network aided development of a semi-empirical interatomic potential for titanium. *Comput. Mater. Sci.* **171**, 109157 (2020).
- Bartók, A. P., Kermode, J., Bernstein, N. & Csányi, G. Machine learning a general-purpose interatomic potential for silicon. *Phys. Rev. X* **8**, 041048 (2018).
- Rosenbrock, C. W. et al. Machine-learned interatomic potentials for alloys and alloy phase diagrams. *Npj Comput. Mater.* **7**, 1–9 (2021).
- Li, X.-G., Chen, C., Zheng, H., Zuo, Y. & Ong, S. P. Complex strengthening mechanisms in the nbmotaw multi-principal element alloy. *Npj Comput. Mater.* **6**, 1–10 (2020).
- Smith, J. S. et al. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nat. Commun.* **10**, 2903 (2019).
- Ramakrishnan, R., Dral, P. O., Rupp, M. & von Lilienfeld, O. A. Big data meets quantum chemistry approximations: the δ -machine learning approach. *J. Chem. Theory Comput.* **11**, 2087–2096 (2015).
- Botu, V., Batra, R., Chapman, J. & Ramprasad, R. Machine learning force fields: construction, validation, and outlook. *J. Phys. Chem. C* **121**, 511–522 (2017).
- Huan, T. D. et al. A universal strategy for the creation of machine learning-based atomistic force fields. *Npj Comput. Mater.* **3**, 1–8 (2017).
- Takahashi, A., Seko, A. & Tanaka, I. Conceptual and practical bases for the high accuracy of machine learning interatomic potentials: application to elemental titanium. *Phys. Rev. Mater.* **1**, 063801 (2017).
- Zong, H. et al. Developing an interatomic potential for martensitic phase transformations in zirconium by machine learning. *Npj Comput. Mater.* **4**, 1–8 (2018).
- Smith, J. S., Nebgen, B., Lubbers, N., Isayev, O. & Roitberg, A. E. Less is more: sampling chemical space with active learning. *J. Chem. Phys.* **148**, 241733 (2018).
- Kostiuchenko, T. et al. Impact of lattice relaxations on phase transitions in a high-entropy alloy studied by machine-learning potentials. *Npj Comput. Mater.* **5**, 1–7 (2019).
- Podryabinkin, E. V. & Shapeev, A. V. Active learning of linearly parametrized interatomic potentials. *Comput. Mater. Sci.* **140**, 171–180 (2017).
- Zhang, L., Lin, D.-Y., Wang, H., Car, R. & Weinan, E. Active learning of uniformly accurate interatomic potentials for materials simulation. *Phys. Rev. Mater.* **3**, 023804 (2019).
- Sivaraman, G. et al. Machine-learned interatomic potentials by active learning: amorphous and liquid hafnium dioxide. *npj Comput. Mater.* **6**, 104 (2020).
- Tan, A. R., Urata, S., Goldman, S., Dietschreit, J. C. B. & Gómez-Bombarelli, R. Single-model uncertainty quantification in neural network potentials does not consistently outperform model ensembles. *npj Comput. Mater.* **9**, 225 (2023).
- Thaler, S., Doehner, G. & Zavadlav, J. Scalable bayesian uncertainty quantification for neural network potentials: promise and pitfalls. *J. Chem. Theory Comput.* **19**, 4520–4532 (2023).
- Kahle, L. & Zipoli, F. Quality of uncertainty estimates from neural network potential ensembles. *Phys. Rev. E* **105**, 015311 (2022).
- Zhu, A., Batzner, S., Musaelian, A. & Kozinsky, B. Fast uncertainty estimates in deep learning interatomic potentials. *J. Chem. Phys.* **158**, 164111 (2023).
- Musil, F., Willatt, M. J., Langovoy, M. A. & Ceriotti, M. Fast and accurate uncertainty estimation in chemical machine learning. *J. Chem. Theory Comput.* **15**, 906–915 (2019).
- Imbalzano, G. et al. Uncertainty estimation for molecular dynamics and sampling. *J. Chem. Phys.* **154**, 074102 (2021).
- Gasteiger, J. et al. Gemnet-oc: developing graph neural networks for large and diverse molecular simulation datasets. *Transactions on Machine Learning Research* (2022).
- Anstine, D. M. & Isayev, O. Machine learning interatomic potentials and long-range physics. *J. Phys. Chem. A* **127**, 2417–2431 (2023).
- Kabylda, A., Vassilev-Galindo, V., Chmiela, S., Poltavsky, I. & Tkatchenko, A. Efficient interatomic descriptors for accurate machine learning force fields of extended molecules. *Nat. Commun.* **14**, 3562 (2023).
- Thaler, S. & Zavadlav, J. Learning neural network potentials from experimental data via differentiable trajectory reweighting. *Nat. Commun.* **12**, 6884 (2021).
- Wang, W., Wu, Z., Dietschreit, J. C. & Gómez-Bombarelli, R. Learning pair potentials using differentiable simulations. *J. Chem. Phys.* **158**, 044113 (2023).
- Navarro, C., Majewski, M., & Fabritius, G. D. Top-down machine learning of coarse-grained protein force-fields. *J. Chem. Theory Comput.* **19**, 7518–7526 (2023).

46. Fröhling, T., Bernetti, M., Calonaci, N. & Bussi, G. Toward empirical force fields that match experimental observables. *J. Chem. Phys.* **152**, 230902 (2020).
47. Baydin, A. G., Pearlmutter, B. A., Radul, A. A. & Siskind, J. M. Automatic differentiation in machine learning: a survey. *J. Machine Learn. Res.* **18**, 1–43 (2018).
48. Schoenholz, S. & Cubuk, E. D. Jax md: a framework for differentiable physics. *Adv. Neural Inf. Process. Syst.* **33**, 11428–11441 (2020).
49. Doerr, S. et al. Torchmd: a deep learning framework for molecular simulations. *J. Chem. Theory Comput.* **17**, 2355–2363 (2021).
50. Wang, X. et al. Dmff: an open-source automatic differentiable platform for molecular force field development and molecular dynamics simulation. *J. Chem. Theory Comput.* **19**, 5897–5909 (2023).
51. Ingraham, J., Riesselman, A., Sander, C. & Marks, D. Learning protein structure with a differentiable simulator. In *International Conference on Learning Representations* (2019).
52. Wang, W., Axelrod, S. & Gómez-Bombarelli, R. Differentiable molecular simulations for control and learning. In *ICLR 2020 Workshop on Integration of Deep Neural Models and Differential Equations* (2020).
53. Purja Pun, G. & Mishin, Y. Development of an interatomic potential for the ni-al system. *Philos. Mag.* **89**, 3245–3267 (2009).
54. Matin, S. et al. Machine learning potentials with the iterative boltzmann inversion: Training to experiment. *J. Chem. Theory Comput.* **20**, 1274–1281 (2024).
55. <https://www.aissquare.com/datasets/detail?pageType=datasets&name=Ti>. accessed 21 Jul 2023.
56. Simmons, G. & Wang, H. *Single crystal elastic constants and calculated aggregate properties: a handbook*. (MIT Press, Cambridge, MA, 1971).
57. Souvatzis, P., Eriksson, O. & Katsnelson, M. Anomalous thermal expansion in α -titanium. *Phys. Rev. Lett.* **99**, 015901 (2007).
58. Faber, F. A. et al. Prediction errors of molecular machine learning models lower than hybrid dft error. *J. Chem. Theory Comput.* **13**, 5255–5264 (2017).
59. Seko, A. Machine learning potentials for multicomponent systems: the ti-al binary system. *Phys. Rev. B* **102**, 174104 (2020).
60. Ledbetter, H., Ogi, H., Kai, S., Kim, S. & Hirao, M. Elastic constants of body-centered-cubic titanium monocrystals. *J. Appl. Phys.* **95**, 4642–4644 (2004).
61. Zwanzig, R. W. High-temperature equation of state by a perturbation method. I. Nonpolar Gases. *J. Chem. Phys.* **22**, 1420–1426 (1954).
62. Shirts, M. & Chodera, J. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **129**, 124105 (2008).
63. Messerly, R., Razavi, S. & Shirts, M. Configuration-sampling-based surrogate models for rapid parameterization of non-bonded interactions. *J. Chem. Theory Comput.* **14**, 3144–3162 (2018).
64. Naden, L. & Shirts, M. Rapid computation of thermodynamic properties over multidimensional nonbonded parameter spaces using adaptive multistate reweighting. *J. Chem. Theory Comput.* **12**, 1806–1823 (2016).
65. Dybeck, E., König, G., Brooks, B. & Shirts, M. Comparison of methods to reweight from classical molecular simulations to qm/mm potentials. *J. Chem. Theory Comput.* **12**, 1466–1480 (2016).
66. Neelakantan, A. et al. Adding gradient noise improves learning for very deep networks. *arXiv preprint arXiv:1511.06807* (2015).
67. Norgaard, A. B., Ferkinghoff-Borg, J. & Lindorff-Larsen, K. Experimental parameterization of an energy function for the simulation of unfolded proteins. *Biophys. J.* **94**, 182–192 (2008).
68. Li, D. W. & Brüsweiler, R. Iterative optimization of molecular mechanics force fields from NMR data of full-length proteins. *J. Chem. Theory Comput.* **7**, 1773–1782 (2011).
69. Van Workum, K., Yoshimoto, K., de Pablo, J. J. & Douglas, J. F. Isothermal stress and elasticity tensors for ions and point dipoles using ewald summations. *Phys. Rev. E* **71**, 061102 (2005).
70. Li, Y., Vočadlo, L. & Brodholt, J. P. The elastic properties of hcp-fe alloys under the conditions of the earth's inner core. *Earth Planet. Sci. Lett.* **493**, 118–127 (2018).
71. Jafari, M., Zarifi, N., Nobakhti, M., Jahandoost, A. & Lame, M. Pseudopotential calculation of the bulk modulus and phonon dispersion of the bcc and hcp structures of titanium. *Phys. Scr.* **83**, 065603 (2011).
72. Hanwell, M. D. et al. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **4**, 1–17 (2012).
73. Togo, A. First-principles phonon calculations with phonopy and phono3py. *J. Phys. Soc. Jpn.* **92**, 012001 (2023).
74. Togo, A., Chaput, L., Tadano, T. & Tanaka, I. Implementation strategies in phonopy and phono3py. *J. Phys.: Condens. Matter.* **35**, 353001 (2023).
75. Green, M. S. Markoff random processes and the statistical mechanics of time-dependent phenomena. ii. irreversible processes in fluids. *J. Chem. Phys.* **22**, 398–413 (1954).
76. Kubo, R. Statistical-mechanical theory of irreversible processes. i. general theory and simple applications to magnetic and conduction problems. *J. Phys. Soc. Jpn.* **12**, 570–586 (1957).
77. Petry, W. et al. Phonon dispersion of the bcc phase of group-iv metals. i. bcc titanium. *Phys. Rev. B* **43**, 10933 (1991).
78. Fisher, E. & Dever, D. *Science, Technology, and Application of Titanium*. (Pergamon, New York, 1970).
79. Stassis, C., Arch, D., Harmon, B. & Wakabayashi, N. Lattice dynamics of hcp ti. *Phys. Rev. B* **19**, 181 (1979).
80. Holland-Moritz, D., Heinen, O., Bellissent, R. & Schenk, T. Short-range order of stable and undercooled liquid titanium. *Mater. Sci. Eng.: A* **449**, 42–45 (2007).
81. Kim, T. & Kelton, K. Structural study of supercooled liquid transition metals. *J. Chem. Phys.* **126**, 054513 (2007).
82. Horbach, J., Rozas, R., Unruh, T. & Meyer, A. Improvement of computer simulation models for metallic melts via quasielastic neutron scattering: a case study of liquid titanium. *Phys. Rev. B* **80**, 212203 (2009).
83. Meyer, A., Horbach, J., Heinen, O., Holland-Moritz, D. & Unruh, T. Self diffusion in liquid titanium: quasielastic neutron scattering and molecular dynamics simulation. In *Defect and Diffusion Forum*, vol. 289, 609–614 (Trans Tech Publ, 2009).
84. Meyer, A. The measurement of self-diffusion coefficients in liquid metals with quasielastic neutron scattering. In *EPJ Web of Conferences*, vol. 83, 01002 (EDP Sciences, 2015).
85. Zhang, J. et al. Thermal equations of state for titanium obtained by high pressure—temperature diffraction studies. *Phys. Rev. B* **78**, 054119 (2008).

Acknowledgements

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 534045056. The authors would like to thank Stephan Thaler for insightful discussions.

Author contributions

J.Z. conceptualized the study. S.R. implemented and applied the methods and conducted MD simulations as well as postprocessing. S.R. and J.Z. planned the study, analyzed and interpreted results, and wrote the paper.

Funding

Open Access funding enabled and organized by Projekt DEAL.

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/s41524-024-01251-4>.

Correspondence and requests for materials should be addressed to Julija Zavadlav.

Reprints and permissions 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024