npj | computational materials

Article

Published in partnership with the Shanghai Institute of Ceramics of the Chinese Academy of Sciences

https://doi.org/10.1038/s41524-025-01651-0

Scalable training of neural network potentials for complex interfaces through data augmentation

In Won Yeu[1,2], Annika Stuke[1,2], Jon López-Zorrilla[3,4], James M. Stevenson[5], David R. Reichman[2,6], Richard A. Friesner[2,6], Alexander Urban[1,2,7] ✉ & Nongnuch Artrith[2,4] ✉

Artificial neural network (ANN) potentials enable accurate atomistic simulations of complex materials at unprecedented scales, but training them for potential energy surfaces (PES) of diverse chemical environments remains computationally intensive, especially when the PES gradients are trained on atomic force data. Here, we present an efficient methodology incorporating forces into ANN training by translating them to synthetic energy data using Gaussian process regression (GPR), leading to accurate PES models with fewer additional first-principles calculations and a reduced computational effort for training. We evaluated the method on hybrid density-functional theory data for ethylene carbonate (EC) molecules and their interfaces with Li metal, which are relevant for Li-metal batteries. The GPR-ANN potentials achieved an accuracy comparable to fully force-trained ANN potentials with a significantly reduced computational and memory overhead, establishing the method as a powerful and scalable framework for constructing high-fidelity ANN potentials for complex materials systems.

Interactions at materials interfaces are essential to technologically relevant phenomena, such as crystal growth[1,2], catalytic activity[3], and interphase formation[4,5]. A concrete example is lithium (Li) metal batteries, which are a promising alternative to conventional Li-ion batteries due to their potential for higher energy density and lower cost per kWh[6–11]. However, their commercialization has been hindered by a lack of understanding regarding the reaction between Li metal and liquid electrolytes[12,13]. An atomistic understanding of interface structures and reaction dynamics would provide an opportunity to control interfaces in devices such as Li-metal batteries. Unfortunately, experimental characterization of interfaces *in operando* remains challenging, and current simulation approaches either face prohibitive computational costs or lack sufficient accuracy.

Interactions at interfaces typically involve different types of bonding, e.g., metallic bonding within Li metal and covalent and ionic bonding in the electrolyte, which are not well captured by conventional interatomic potentials. Additionally, interface simulations typically require structure models with several hundred to thousands of atoms, i.e., system sizes that are challenging for accurate first-principles electronic structure methods. Density-functional theory (DFT) has become a standard tool for materials discovery and has proven accurate for predicting a wide range of materials properties from first principles[14,15]. DFT expresses the electronic structure problem in terms of the electron density, thereby avoiding the need for an accurate approximation of the all-electron wavefunction. Semilocal DFT based on the generalized-gradient approximation (GGA) or meta-GGA, is comparatively computationally efficient and the most widely used electronic structure method for materials simulations, and it is generally considered reliable for metals and main-group compounds. Still, additional corrections are often required for transition-metal compounds[16] and for molecular systems with dispersive interactions[17], which can lead to significant errors for interfacial properties if the two materials in contact exhibit different types of bonding. Computationally significantly more demanding DFT hybrid functionals are often the only choice to reliably describe the electronic structure and potential energy in interface regions[18].

Machine-learning (ML) potentials trained on first-principles and quantum chemistry methods have emerged as a new family of reactive interatomic force fields[19–24]. Early methods, including Gaussian Process regression[25] and feed-forward artificial neural networks (ANNs)[26], laid the groundwork for ML potentials, while recent innovations[27–34] have achieved improvements in terms of computational efficiency and accuracy on public benchmarks, bringing ML potentials closer to replacing DFT in some applications. Especially, ML potentials based on ANNs with atomic descriptors[35] have been applied to a wide range of materials and phenomena,

---

[1]Department of Chemical Engineering, Columbia University, New York, NY, USA. [2]Columbia Center for Computational Electrochemistry, Columbia University, New York, NY, USA. [3]Physics Department, University of the Basque Country (UPV/EHU), Leioa, Basque Country, Leioa, Spain. [4]Debye Institute for Nanomaterials Science, Utrecht University, Utrecht, The Netherlands. [5]Schrödinger, Inc., New York, NY, USA. [6]Department of Chemistry, Columbia University, New York, NY, USA. [7]Columbia Electrochemical Energy Center, Columbia University, New York, NY, USA. ✉e-mail: a.urban@columbia.edu; n.artrith@uu.nl

including metals[36–41], oxides[41–44], alloys[45,46], molecular systems[47–50] and amorphous phases[51–54] due to their computational efficiency and easy accessibility. Carefully trained ML potentials can represent the potential energy surface (PES) of materials with thousands to millions of atoms with an accuracy close to that of ab initio methods at a significantly reduced computational cost and scaling. Accordingly, extensive research has been devoted to constructing ANN potentials for interfaces, for instance, between copper clusters and zinc oxide[41], water and copper[55], water and zinc oxide[56], and heterogeneous catalysts[57,58].

Constructing reliable ML potentials for surfaces and interfaces is especially challenging because abrupt changes in atomic environments and different bonding types are involved. As a consequence, a huge amount of reference data points can be needed to capture the drastically changing PES with sufficient precision[59]. These challenges have limited the construction of ANNs for interfaces, and atomistic understanding of interfacial reactions remains limited despite their crucial impacts in various technology areas[60–62].

Effective learning strategies are desirable to avoid any unnecessary first-principles calculations with expensive hybrid functionals or higher-level theory. In this regard, including atomic force information in the ANN potential training was found to greatly reduce data requirements, improve PES accuracy, and increase transferability[35,43,63–67]. Additionally, active learning can be employed, where training data for ML potential construction is generated incrementally based on the current state of the potential, adding new data to training sets in a systematic and non-redundant fashion[68–73]. A typical active learning strategy is to perform additional first-principles calculations for atomic structures for which the ML potential reports an uncertainty that exceeds a user-defined threshold. Such an approach avoids redundant first-principles calculations and increases the transferability of ML potentials by adapting the model to a new structure domain.

However, training ANNs not only on function values (energies) but also on derivatives (atomic forces, stress tensors, etc.) comes at a significant computational and memory overhead because such direct force training needs to evaluate and store the second (or higher order) derivative of the ANN potential, which scales quadratically with the number of atoms within the cutoff range[74]. This unfavorable scaling can be prohibitively expensive for complex, dense systems, or at least calls for expensive specialized hardware. Furthermore, conventional ANNs do not directly provide an uncertainty estimate that could be used for active learning, so either the predictions from multiple independently-trained ANN potentials need to be combined (*query by committee*)[38,75] or the ANN architecture needs to be modified, for example, by introducing dropout layers[76].

In this article, we introduce a new data-augmentation approach where ANN training is seamlessly integrated with Gaussian process regression (GPR), a non-parametric regression model, to overcome these downsides of ANN training. The *GPR-ANN* approach indirectly learns the information from the PES gradients (i.e., the interatomic forces) by translating the gradients to additional energy data via local interpolation and extrapolation using separate GPR models, simultaneously fitting to data points and their derivatives of subsystems of overall heterogeneous reference data (Fig. 1). The general idea follows the same spirit as the first-order Taylor-expansion extrapolation method that some of the present authors proposed previously[74] and simple extrapolation based on the zeroth order[77]. The non-linear, Bayesian nature of GPR models leads to greatly improved performance compared to the above methods, as we will show in the following sections. GPR-based ML potentials have been used with great success, such as the Gaussian approximation potential (GAP) by Bartók et al.[25,78–80], which performs exceptionally well with limited data compared to other ML potential methods[81,82]. We will show how the GPR-ANN approach enables scalable force data utilization without relying on direct force training by combining the *best of both worlds*: leveraging the superior interpolation and extrapolation capabilities of GPR with small data sets and uncertainty estimation at negligible additional computational cost, while maintaining the efficiency of ANN training for large data sets.

In the following *Results* section, we first detail the working principle of the GPR-ANN method and then demonstrate its improved performance in comparison with conventional ANN potentials by applying the method to three benchmark cases with increasing complexity: (i) a Lennard-Jones (LJ) potential of the H-H bond in the H_2 molecule, (ii) a hybrid-functional DFT PES of two ethylene carbonate (EC) molecules, and (iii) an EC molecule on the surface of Li metal.

Results

Energy training

A popular ANN potential architecture is the high-dimensional PES proposed by Behler and Parrinello[26], which describes the total energy, $E(\sigma)$, of a structure, $\sigma = \{(\vec{R}_1, t_1), (\vec{R}_2, t_2), \ldots, (\vec{R}_N, t_N)\}$ where $\vec{R}_i$ are the coordinates of atom i and t_i is its chemical species, as a sum of atomic energy contributions

$$E(\sigma) \approx E^{\mathrm{ANN}}(\sigma; \{w\}) = \sum_{i \in \sigma} \mathrm{ANN}_{t_i}\left(\sigma_i^{R_c}; \left\{w_{t_i}\right\}\right) \tag{1}$$

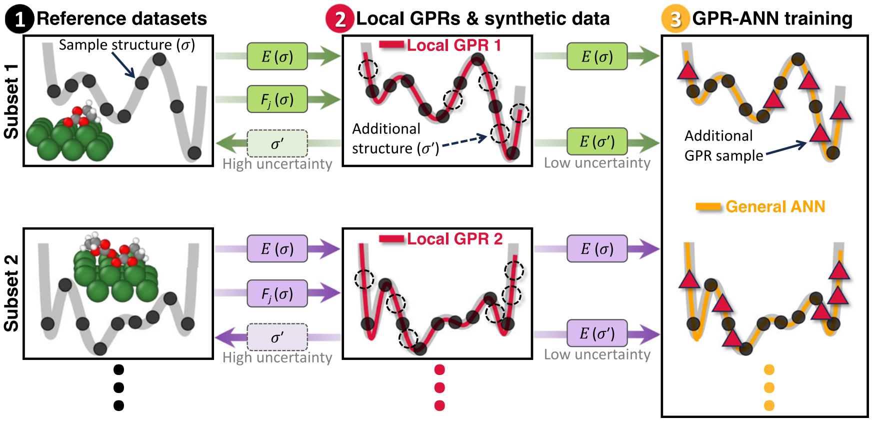


Fig. 1 | Indirect force training with the GPR-ANN approach. (**Step 1**) The reference data (black circles) consists of atomic structures (σ), their energies ($E(\sigma)$) and corresponding atomic forces ($F_j(\sigma)$) from electronic structure calculations for structures sampling target potential energy surfaces (PES, thick gray lines). Each subset contains related structures with the same number of atoms. (**Step 2**) For each subset, Gaussian process regression (GPR) models can efficiently interpolate the potential energy surface based on the energies and atomic forces (red lines). The GPR models can then be used to generate synthetic data by labeling additional related structures (empty circles) with energies. Structures for which the GPR model reports a high uncertainty are evaluated with the reference electronic structure method. (**Step 3**) Finally, the original structures and their energies can be combined with the additional structures and their GPR energies (red triangles) into a unified overall data set that can be used for efficient energy-only training of general ANN potentials (yellow lines).

where $\sigma_i^{R_c}$ in Eq. (1) is a descriptor (i.e., feature vector) representing the atomic environment of atom i within a cutoff radius R_c that serves as input to a multilayer perceptron feedforward neural network, ANN_{t_i}, specific to the chemical species of atom i. Each neural network ANN_{t_i} is defined by its weight parameters $\{w_{t_i}\}$, and we denote the set of all weight parameters for all chemical species $\{w\}$.

A basic requirement of $\sigma_i^{R_c}$ is to obey the invariances of the total energy with respect to translation/rotation of the entire structure and permutation of equivalent atoms, and in this work, we used a Chebyshev descriptor method[83] that allows for an efficient representation of multi-element compounds. Details of the ANN architecture and the parameters for the Chebyshev descriptor are given in the *Methods* section.

Given reference data sets of structures σ and energies $E^{\text{ref}}(\sigma)$, energy-only training minimizes the energy loss function

$$\mathcal{L}^{\text{energy}} = \sum_{\sigma} \frac{1}{2} \left\{ E^{\text{ANN}}(\sigma; \{w\}) - E^{\text{ref}}(\sigma) \right\}^2 \tag{2}$$

by optimizing the weight parameters $\{w\}$

$$\{w^{\text{opt}}\} = \arg \min_{\{w\}} \{\mathcal{L}^{\text{energy}}\} \quad . \tag{3}$$

We refer to the process of minimizing the loss function $\mathcal{L}^{\text{energy}}$ as *energy training*. The minimization of $\mathcal{L}^{\text{energy}}$ with respect to $\{w\}$ requires the derivative

$$\begin{aligned} \frac{\partial \mathcal{L}^{\text{energy}}}{\partial w} &= \sum_{\sigma} \Delta E(\sigma) \frac{\partial E^{\text{ANN}}(\sigma;\{w\})}{\partial w} \\ &= \sum_{\sigma} \Delta E(\sigma) \sum_{i \in \sigma} \frac{\partial \text{ANN}_{t_i}(\sigma_i^{R_c};\{w_{t_i}\})}{\partial w} \qquad , \\ &\text{where } \Delta E(\sigma) = E^{\text{ANN}}(\sigma; \{w\}) - E^{\text{ref}}(\sigma) \end{aligned} \tag{4}$$

which can be efficiently calculated using backpropagation. The computational cost and memory requirement of energy training is, per data point, independent of the size of the data set but instead scales as $\mathcal{O}(N_w)$ where N_w is the number of weight parameters. Therefore, for training data sets containing a total of N_{atom} atoms, the total computational cost per epoch scales linearly with data points $\mathcal{O}(N_w N_{\text{atom}})$, which makes this approach feasible for large data sets up to millions of data points.

However, while energy training is computationally efficient, it does not fully utilize the reference data, since it discounts the interatomic forces, which provide valuable high-dimensional information about the PES gradient and can be obtained from many electronic structure methods using the Hellmann-Feynman theorem without significant computational overhead[84]. Consequently, energy training requires larger data sets to sample the PES more finely to accurately reproduce its gradient and curvature, leading to increased computational overhead for electronic structure reference calculations. Moreover, training exclusively on energies can result in large uncertainties and unreliable force reconstruction, as low energy errors do not necessarily correlate with accurate force predictions. This issue is further exacerbated by the presence of noise in the energy data, which amplifies force prediction errors as the model overfits to the noise, undermining the reliability of the reconstructed PES[85].

Direct force training

The chemical complexity of interface systems might require electronic-structure methods that are computationally demanding, such as hybrid functional DFT calculations, so that an excess of reference data for the ANN potential training must be avoided. Including atomic force information in ANN potential training significantly reduces data requirements by causing the training to enforce a physical constraint, the conservation of total energy $\vec{F}_i = -\vec{\nabla}_i E(\sigma)$[85]. Incorporating the physics into ANN training not only enables ANN potentials to accurately reproduce the gradient of PES with less training data but also helps prevent overfitting to noise in the energy data, ensuring more reliable energy and force predictions.

ANN potentials can be trained simultaneously on energies $E^{\text{ref}}(\sigma)$ and forces $\vec{F}_j^{\text{ref}}(\sigma)$, where j indexes the individual atoms within a structure σ, by including the force error $\mathcal{L}^{\text{force}}$ in the total loss function,

$$\mathcal{L}^{\text{total}} = (1-\alpha)\mathcal{L}^{\text{energy}} + \alpha \mathcal{L}^{\text{force}} \tag{5}$$

$$\mathcal{L}^{\text{force}} = \sum_{\sigma} \frac{1}{2} \left[\sum_{j \in \sigma} \left\{ -\vec{\nabla}_j E^{\text{ANN}}(\sigma; \{w\}) - \vec{F}_j^{\text{ref}}(\sigma) \right\} \right]^2 , \tag{6}$$

where α is a parameter determining the relative contribution of the force loss $\mathcal{L}^{\text{force}}$ to the overall loss function $\mathcal{L}^{\text{total}}$. We refer to training that minimizes the loss function $\mathcal{L}^{\text{total}}$, $\{w^{\text{opt}}\} = \arg\min\{\mathcal{L}^{\text{total}}\}$, as *direct force training*.

However, direct force training also has a critical drawback: overhead in training cost and memory, since it requires evaluating the derivative $\mathcal{L}^{\text{total}}$ with respect to $\{w\}$

$$\begin{aligned} \frac{\partial \mathcal{L}^{\text{total}}}{\partial w} &= (1-\alpha)\frac{\partial \mathcal{L}^{\text{energy}}}{\partial w} \\ &\quad -\alpha \sum_{\sigma} \sum_{j \in \sigma} \Delta \vec{F}_j(\sigma) \sum_{j \in \sigma} \frac{\partial}{\partial w} \vec{\nabla}_j E^{\text{ANN}}(\sigma; \{w\}) \qquad , \\ &\quad \text{where } \Delta \vec{F}_j(\sigma) = -\vec{\nabla}_j E^{\text{ANN}}(\sigma; \{w\}) - \vec{F}_j^{\text{ref}}(\sigma) \end{aligned} \tag{7}$$

and calculating the term

$$\begin{aligned} &\sum_{j \in \sigma} \frac{\partial}{\partial w} \vec{\nabla}_j E^{\text{ANN}}(\sigma; \{w\}) \\ &= \sum_{j \in \sigma} \sum_{i \in \sigma} \frac{\partial}{\partial w} \vec{\nabla}_j \text{ANN}_{t_i}\left(\sigma_i^{R_c}; \left\{w_{t_i}\right\}\right) \end{aligned} \tag{8}$$

requires evaluating the second derivative of the neural networks.

The derivative within the double sum of Eq. (8) is zero when the local atomic environments of atoms i and j do not overlap. As a consequence, the total computational cost of direct force training scales with $\mathcal{O}(N_w N_{\text{atom}} N_{\text{local}})$ where N_w is the number of weight parameters, N_{atom} is the total number of atoms in the training set, and N_{local} is the average number of atoms within $2R_c$, where R_c is the cutoff distance of the local atomic environment. In other words, direct force training scales approximately quadratically with N_{local}, which, in turn, increases cubically with R_c, so that the computational cost scales with the cutoff radius as $\mathcal{O}(R_c^6)$. This unfavorable scaling makes direct force training expensive for condensed phases and infeasible for large cutoff radii R_c, prompting us to develop a more efficient training method.

Representing potential energy surfaces with Gaussian process regression

Here, we propose using GPR models as surrogate models to efficiently incorporate atomic forces in ANN potential training in an indirect fashion. Unlike ANNs, which are sometimes referred to as *parametric* models since they are defined by their architecture and weight parameters $\{w\}$, GPRs are *non-parametric* kernel-based ML models, for which the model construction depends solely on the reference data. For small data sets, full GPR simultaneously fitted to function values and derivatives is the method of choice with respect to accuracy, remarkably reproducing target PESs, and it provides an uncertainty estimate without additional computational overhead, making it possible to further reduce data requirements with active learning[25,78–80].

The downside is that the computational cost and memory requirements for constructing a full dense GPR model scale as $\mathcal{O}(N^3)$ and $\mathcal{O}(N^2)$, respectively, with the training set size N[86]. In addition, the cost of inference or prediction for new data points also increases with the size of the reference training data set[80]. The unfavorable scaling can be improved with sparse GPR techniques, but the fundamental dependence on the data size remains.

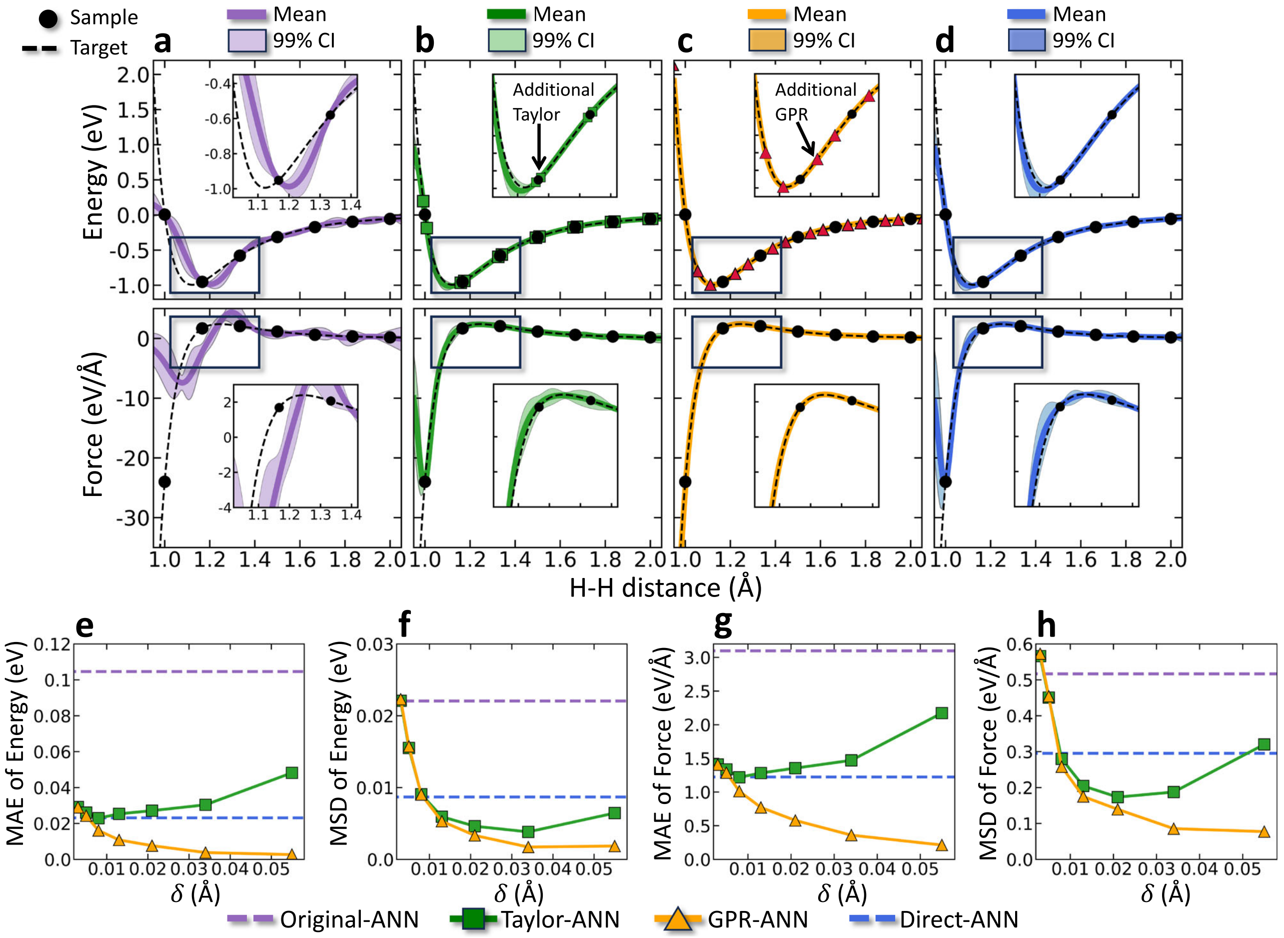


Fig. 2 | Comparison of the different ANN potential training strategies for an H_2 molecule. The same seven reference data points (black circles) sampled from the target potential energy surface of a H_2 dimer (dashed black line) were used to assess the accuracy and robustness of ANN potentials obtained by training with the four strategies detailed in the main text: **a**, energy-only training, indirect force training with **b**, the Taylor-expansion method and, **c**, the GPR-ANN method, and **d**, direct force training. The insets show zoomed-in views of the regions marked with rectangles. The mean predicted energies (top) and forces (bottom) of 10 ANN potentials are shown as solid lines, and the shaded regions indicate the 99% confidence interval (CI) as a measure of uncertainty. For the data-augmentation approaches, Taylor-ANN and GPR-ANN, the seven reference energies were supplemented with 14 predicted energies (green squares in **b** and red triangles in **c**), and the corresponding H_2 structures were generated with atomic displacements of $\delta = \pm 0.008$ Å and $\delta = \pm 0.055$ Å, respectively. The Taylor-ANN and GPR-ANN potentials corresponding to the optimal atomic displacements are shown, and results from other δ variables can be found in Figs. S4–7. The accuracy and robustness of the training strategies are quantified by the **e**, mean absolute error (MAE) and, **f**, mean standard deviation (MSD) of the energy and the **g**, MAE and **h**, MSD of the force, respectively. For the data-augmentation methods, these measures depend on the displacement length and are shown as a function of δ.

Indirect force training (GPR-ANN)

The intrinsic pros and cons of *parametric* ANN and *non-parametric* GPR prompted us to consider a way to integrate the advantages of both approaches. For interface systems, the overall heterogeneous reference data is naturally comprised of data for subsystems, e.g., bulk structures of the involved materials, cluster structures with different numbers of atoms and compositions, and periodic surface slab models (black boxes in Fig. 1). Each of these homogeneous subsystems can be individually fitted using separate local GPR models (red lines in Fig. 1), enabling more accurate and specialized representations of their respective PESs. Using the local GPR models, the overall PES can be finely sampled by perturbing the atomic structures in the subsystem data sets and augmenting synthetic data outside the observed regions with GPR-predicted structure-energy data. Limiting each GPR model to subsets of the total reference data mitigates scalability issues with large data sets, simplifies the fit, and facilitates highly efficient inference.

This process also makes it easy to perform active learning iterations: The GPR uncertainty of each additionally sampled structure (σ' in Fig. 1) is evaluated, and when it exceeds a user-defined threshold, the structure can be re-evaluated using the reference electronic-structure method and incorporated into the reference data set. This was not necessary in the benchmark examples of this study, as the GPR uncertainty remained low for perturbations applied to the relatively homogeneous subsets of data considered in the present study. In applications where larger perturbations are used to actively explore the structural domain or where subsets of data are sparsely distributed, the GPR uncertainty measure would enable refining the training data quality via Bayesian learning, motivating additional first-principles calculations during the GPR augmentation preprocessing stage (steps 1 and 2 in Fig. 1). Finally, ANN potentials (orange line in Fig. 1) are trained on GPR-augmented energies (red triangles) as well as the original electronic-structure reference energies (black circles) within the efficient energy-only training scheme described above. Since the synthetic data points generated with the local GPR models are based on the energies and atomic forces, the resulting ANN potentials are also implicitly trained on force information. We refer to this approach as *indirect force training*.

As will be shown in the *Results* section, a multiple of M between 10 and 40 in synthetic data points relative to the original data is sufficient to obtain optimal sampling of the GPR-ANN potentials, which leads to a total computational cost that is significantly lower than that of direct force training, especially for interface systems consisting of a large number of atoms.

In the following, we compare the training performance of the GPR-ANN approach in terms of accuracy, robustness, and computational efficiency against energy-only training, direct force training, and training using a first-order data augmentation method[74] here referred to as Taylor-ANN.

H_2 molecule

As a first example, we consider a Lennard-Jones (LJ) potential roughly approximating the H-H dimer for the purpose of an intuitive PES visualization. In Fig. 2, the target PES in the bond length range from 0.95 to 2.05 Å is displayed as a dashed black line, and seven reference samples are marked by black circles. The accuracy and robustness of the different ANN training methods are examined in terms of the mean and standard deviation (SD) over a committee of 10 ANN potentials, differing by random seed, to visualize how accurately the mean reproduces the target PES and how robust the training result is based on the SD between the 10 ANNs. Results for the predicted energies and forces are shown in Fig. 2a–d for each of the four ANN training methods with the mean and the 99% confidence interval (CI) represented by solid lines and shaded regions, respectively.

In the case of the data-augmentation approaches, Taylor-ANN and GPR-ANN, the seven reference forces were translated into additional energies for 14 synthetic structure-energy data points, and the combined total of $7 + 14 = 21$ energy data points were used as training data. The additional structures were generated by displacing the atoms in the seven original reference structures by small displacements δ, and the energies approximated by linear Taylor expansion and a GPR model are shown in the figure as well.

Note that the performance of the Taylor-ANN and GPR-ANN potentials depends on the choice of the displacement length δ, and the results shown in Fig. 2b, c are from the potentials with the optimal displacements δ. Figure S1 shows the approximate energies in comparison to the reference LJ PES for different displacements δ ranging from ± 0.003 to ± 0.055 Å. The mean absolute errors (MAE) relative to the LJ reference energies are also summarized in Fig. S1d as a function of the displacement length. As the displacement length increases, the Taylor-ANN energies deviate farther from the reference due to the limitations of the first-order approximation, while the GPR-ANN energies agree closely with the LJ PES throughout the displacements considered. The ANN potentials trained on augmented energy data highly depend on the perturbation size and correctness of corresponding energies that the potentials were trained on, and Figs. S2–7 summarize the mean, SD, and error of the mean with respect to the target LJ PES. The best Taylor-ANN potentials were obtained with $\delta = \pm 0.008$ Å and the best GPR-ANN potentials were obtained with the largest $\delta = \pm 0.055$ Å.

All of the four training methods exactly reproduce the energies given as training data with negligible uncertainty. However, the error of the mean over 10 ANN potentials obtained from energy-only training increases in between training structures (Fig. 2a). The errors come from an incorrect reproduction of the PES gradient, even at the reference points, as is evident from the force errors shown in the bottom panels of the figures. In addition, for the smooth PES region at long H-H separations, the SD is not negligible despite the mean of the 10 ANN potentials aligning well with the target PES, showing the interpolation instability of energy-only training with insufficient data.

Indirect force training with the Taylor-ANN approach (Fig. 2b) or direct force training (Fig. 2d) corrects the slope of the ANN potentials near the reference samples, leading to great improvements in interpolation as shown in a significant reduction in the error and SD for structures not included in the training data but located nearby. However, the SD still remains non-negligible, particularly for structures in between the training data as shown in the insets. Additionally, the mean demonstrates a significant underestimation of absolute forces in the repulsive region where the H-H distance is below 1 Å, highlighting the intrinsic extrapolation limitations of ANN training beyond the scope of local training data, small δ of ± 0.008 Å in the Taylor-ANN method, and local forces in direct force training. This issue persists even with the use of large perturbations in the Taylor-ANN approach, as the model learns from inaccurate additional energy values (Figs. S4g, S6g).

On the other hand, the GPR-ANN potentials trained on accurate augmented energies for the diverse structures generated with large δ achieve the most accurate reproduction of the PES and its derivatives, maintaining negligible SD both within and beyond the training data range as shown in Fig. 2c.

The four training methods are further compared in terms of their mean absolute error (MAE) and mean SD (MSD) over 200 test points in Fig. 2e–h. The results are shown as a function of the displacement length for the Taylor-ANN and GPR-ANN approaches. Taylor-ANNs (green squares) show the best accuracy for a small displacement of $\delta = 0.008$ Å, but their uncertainty is the lowest at a much larger displacement of $\delta = 0.034$ Å. Additional structures generated by small displacements are very similar to the reference structures, and thus, there are still large PES regions that are not well sampled. In general, if the synthetic data points are too similar to the original data, i.e., if the displacements δ are chosen too small, the data-augmentation methods Taylor-ANN and GPR-ANN do not show any notable improvement in robustness compared to energy-only training, as seen in Fig. 2f, h. As the displacement increases, the additional structures are more distinct from the original reference structures, and these well-distributed training points greatly reduce the SD between ANN potentials. At the same time, however, Taylor-ANN potential energies become less accurate as the displacement increases (Fig. S1), and including the inaccurate synthetic data in the training data degrades the ANN potential accuracy despite decreasing the uncertainty in the predicted energies and forces. Thus, the Taylor-ANN augmentation method suffers from a trade-off between data diversity and accuracy that needs to be accounted for when it is used.

In contrast, we can see that the GPR-ANN augmentation method is able to provide accurate energy labels for highly displaced unique structures. As a result, ANN potentials trained on the GPR-augmented energy data set show a gradual improvement in accuracy and uncertainty with increasing displacement length. The GPR-ANNs with the largest displacement of $\delta = 0.055$ Å, which results in the most uniform sampling of the PES regions, almost perfectly represent the LJ PES, exhibiting excellent accuracy and robustness in both interpolation and extrapolation regions. The MAEs of the GPR-ANN potentials for energies and forces are 3 meV and 0.21 eV/Å, lower than the MAE achieved by energy-only training (105 meV, 3.10 eV/Å) and direct force training (23 meV, 1.22 eV/Å). In addition, the MSDs of the GPR-ANN potentials for energies and forces are 2 meV and 0.08 eV/Å, i.e., also lower than the MSD for energy-only (22 meV, 0.52 eV/Å) and direct force training (9 meV, 0.30 eV/Å).

Given identical reference energy and force data, the GPR-ANN data-augmentation strategy makes optimal use of the available information and leads to the most accurate and robust (least uncertain) potentials among the four considered ANN training methods. In practice, this means the GPR-ANN method requires the least number of reference electronic-structure calculations to reach a desirable level of accuracy and uncertainty.

While the dihydrogen molecule is a test system that is easy to conceptualize, it does not reflect the complexity of real-world applications. Therefore, we next compare the ANN training approaches for a higher-dimensional system comprised of two ethylene carbonate (EC) molecules.

Ethylene carbonate molecule dimers

To assess the GPR-ANN data-augmentation method for a relevant application, we first turned to the electrolyte side of the electrolyte-electrode interface that we seek to model. The energies and atomic forces of 1000 ethylene carbonate (EC) dimer structures were evaluated with hybrid-functional DFT calculations, and the resulting data set was divided into 250 training and 750 test data points. See the *Methods* section for details of the DFT calculations and structure generation.

As for the H_2 example before, we compared the accuracy and robustness of the four different ANN training methods by evaluating the MAE and MSD for the energy and force predictions for the 750 test

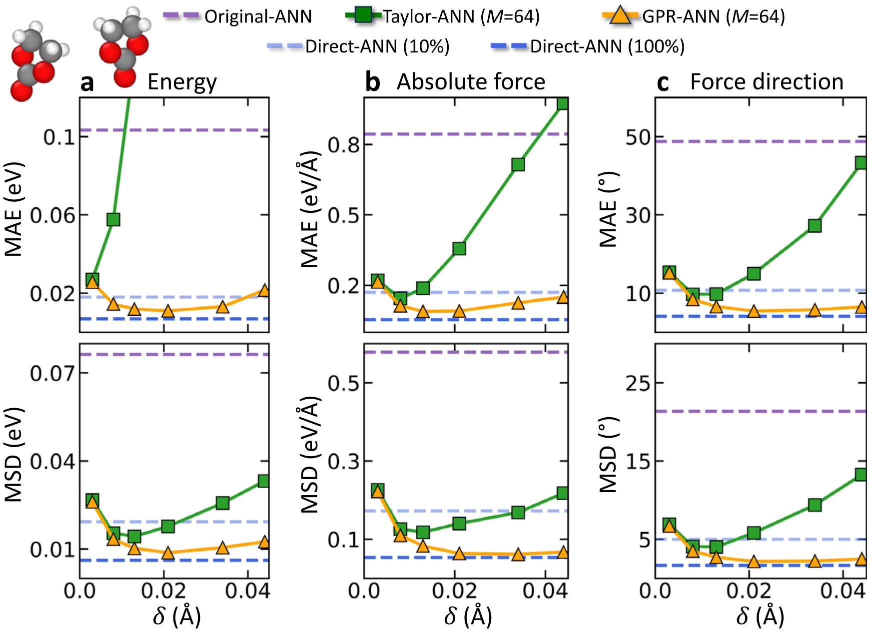


Fig. 3 | Comparison of the accuracy and robustness of the four ANN training methods for ethylene carbonate dimer structures. The mean absolute error (MAE) and mean standard deviation (MSD) over a committee of 10 ANN potentials are shown for **a** the energy, **b** the absolute magnitude of the forces, and **c** the force direction. These metrics are shown for ANN potentials obtained from energy-only training (dashed purple line), indirect force training with the Taylor-ANN (green squares), and the GPR-ANN (orange triangles) approach, and direct force training with 10% forces (dashed light blue line) and 100% force information (dashed dark blue line).

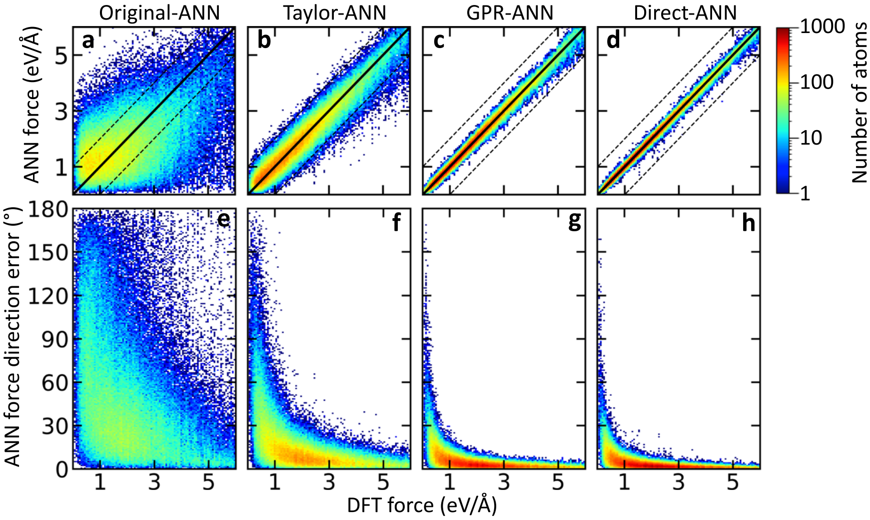


Fig. 4 | Detailed analysis of the atomic forces in ethylene carbonate dimers predicted with the different training approaches. a–d Correlation of the magnitude of the forces predicted by ANN potentials with the DFT reference. **e–h** Error in force direction with respect to DFT reference. The predictions were made by a committee of 10 potentials obtained from energy-only training (**a**, **e**), implicit force training with the Taylor-ANN (**b**, **f**) and GPR-ANN (**c**, **g**) methods, and direct force training (**d**, **h**). The color indicates the frequency of occurrence using a logarithmic scale. The solid black line in the top panels **a–d** corresponds to perfect correlation with the DFT reference, and the dashed black lines indicate differences of 1 eV/Å. Optimal parameters were used for all force training methods: Taylor-ANN ($\delta = 0.003$ Å, multiple = 64), GPR-ANN ($\delta = 0.021$ Å, multiple = 64), and direct force training (100% forces, alpha = 0.3).

structures using a committee of 10 ANN potentials. Figure 3a–c shows the MAE and MSD of the energy, absolute force, and force direction, respectively, as a function of the displacement amplitude δ.

Each EC, $(CH_2O)_2CO$, consists of 10 atoms, so for two molecules the total number of force components is 60 for each EC dimer structure. With direct force training, the percentage of force components to consider during training is a user parameter, and prior research suggests that ~10% of the force components can already be sufficient and provide an optimal balance of computational efficiency and training outcome[87]. Therefore, Fig. 3a–c shows the performance metrics of direct force training with 10 and 100% force information, respectively.

As in the H_2 example, the indirect force training with the Taylor-ANN approach achieves the lowest MAE for $\delta = 0.003$ Å, which is again a smaller displacement than the one that minimizes the MSD ($\delta = 0.013$ Å). Atomic forces predicted by the Taylor-ANN potentials are comparable to those predicted by GPR-ANN potentials for $\delta < 0.01$ Å, but the performance of the GPR-ANN potentials remains robust even for larger displacements due to an increase in data diversity with more accurate energy augmentation achieved through GPR models. In all considered metrics, accuracy and robustness for energies and forces, the GPR-ANN potentials with $\delta > 0.003$ Å are better than ANN potentials from direct force training with 10% of the force information. With an optimal δ of 0.021 Å, GPR-ANN potentials are comparable to direct force training with 100% forces, exhibiting MAE and MSD values that improve over energy-only training by about one order of magnitude.

Note that the data shown in Fig. 3a–c is for Taylor-ANN and GPR-ANN potentials with a multiple of $M = 64$. We used an unnecessarily large multiple for this comparison to ensure this parameter does not bias the results. As seen in supplementary Fig. S8, GPR-ANN potentials improve with increasing multiple and converge quickly, outperforming direct force training with 10% of the force data already at a multiple of $M = 16$ and for $M > 36$ becoming comparable to direct force training with 100% forces.

Figure 4a–d shows a more detailed analysis of the atomic forces predicted by ANN potentials trained with the different strategies. The figure shows the correlation of the predicted absolute force magnitude with the DFT reference forces. ANN potentials trained on the energy only (Original-ANN in Fig. 4a), clearly do not provide accurate force predictions without expanding the EC dimer database. For a large fraction of the atoms, the force

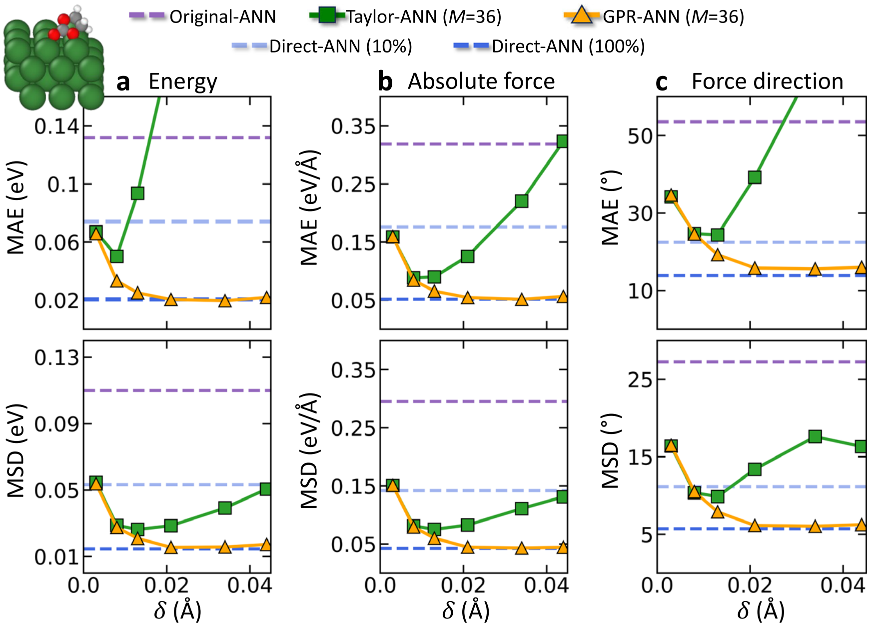


Fig. 5 | Comparison of the accuracy and robustness of the four ANN training methods for an ethylene carbonate molecule adsorbed on the lithium metal (100) surface. The mean absolute error (MAE) and mean standard deviation (MSD) based on a committee of 10 ANN potentials are shown for the **a**, energy, **b**, absolute force magnitude, and **c** force direction. Results are shown for energy-only training (dashed purple lines), indirect force training with the Taylor-ANN (green squares) and GPR-ANN (orange triangles) data-augmentation methods, and direct force training with 10% (dashed light blue lines) and 100% (dashed dark blue lines) force information.

error is greater than 1 eV/Å, which is the range indicated with thin dashed lines in the figure. The three other training approaches greatly improve the absolute force distribution, and the figure shows results for optimal parameters: Taylor-ANN potentials ($\delta = 0.003$ Å, multiple = 64), GPR-ANN potentials ($\delta = 0.021$ Å, multiple = 64), and direct force training (100% forces, $\alpha = 0.3$). Potentials trained with the GPR-ANN approach and direct force training show similar performance in force prediction, such that the absolute forces of most of the atoms are predicted to be close to the DFT reference, i.e., close to the $x = y$ diagonal of the plots.

Figure 4e–h shows the corresponding distribution of ANN prediction errors in the direction of the atomic force vectors as a function of the absolute value of the DFT force. Again, the failure of the ANN potentials trained on only 250 energies is obvious: the forces acting on a significant fraction of the atoms are predicted in opposite direction (180°) relative to the reference. The errors in atomic force direction are significantly reduced when force information is included, especially with the GPR-ANN approach or direct force training. For these two approaches, errors in force direction larger than 30° only occur in a small fraction of the atoms and for force vectors with small magnitudes, below 0.5 eV/Å.

While the performance of the GPR-ANN approach looks promising for the EC molecule example, interface systems are yet more challenging to model. Therefore, as a final test, we will compare the different training methods for EC adsorbed on and interacting with the surface of Li metal.

EC on the surface of lithium metal

We generated 800 reference structures of an EC molecule on the Li(100) surface, 46 atoms in total, by applying random displacements to all atoms in the ground-state configuration. All structures were labeled with energies and atomic forces from hybrid-functional DFT calculations, the details of which are provided in the *Methods* section. The data set was split into 200 training and 600 test data points (see *Methods* section).

As for the previous systems, committees of 10 ANN potentials were used to predict the energy and force of the 600 test structures, and the MAE and MSD are summarized in Fig. 5. Overall, the trends are similar as for the EC dimer structures from the previous section, yet more pronounced: Implicit force training with the Taylor-ANN approach achieves the lowest MAE and MSD for a δ of 0.008 Å, and the accuracy and robustness of the method for energies and the magnitude of the forces lie between those of direct force training with 10% and 100% force information and is comparable to direct force training with 10% for the force direction. However, the errors and variance rapidly increase with increasing δ as the first-order Taylor expansion becomes unreliable. In practice, it can be expected to be challenging to find an optimal δ that both yields data diversity and provides sufficient accuracy, since the perturbation parameter depends on the unknown PES.

In contrast, the GPR-ANN data-augmentation approach is much more robust with respect to the choice of the displacement amplitude. As seen in Fig. 5, all GPR-ANN potentials with $\delta > 0.013$ Å show accuracy and robustness comparable to direct force training with 100% force information across all of the metrics.

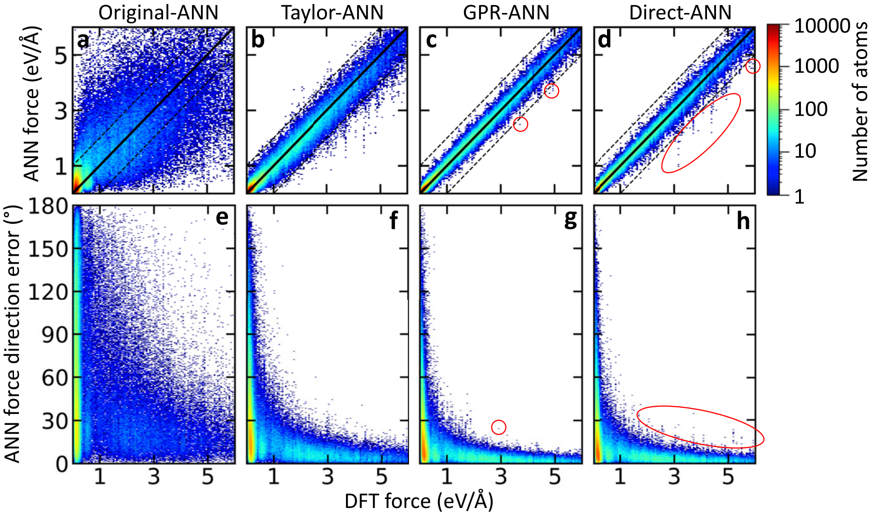
The MAE and MSD as a function of the augmentation multiple is plotted in Fig. S9 for fixed optimal δ values of 0.008 Å for the Taylor-ANN and 0.034 Å for the GPR-ANN approach, respectively. The GPR-ANN approach with a multiple of 16 already reaches the accuracy and robustness of direct force training with 100% forces for this complex Li-EC system.

Figure 6 shows a detailed analysis of the atomic forces with correlation plots and the directional errors, in the same fashion as above in Fig. 4 for the EC dimer case. The trends are similar to those seen for the EC dimers, and energy-only training on 200 energy data points proves certainly unreliable regarding the predictions for both magnitude and direction of forces showing severe errors for a large fraction of atoms. The best agreement with the DFT reference was achieved by the GPR-ANN potentials and direct force training, consistent with the average analysis of Fig. 5. For this more challenging system, direct force training shows significantly more outliers with errors greater than 1 eV/Å (highlighted in red circles and ellipses in Fig. 6) than training with the GPR-ANN approach, implying that direct force training is more vulnerable to critically large errors despite the comparable MAE and MSD of the two methods.

Computational efficiency of the GPR-ANN method

As discussed above, the computational cost of direct force training scales with $\mathcal{O}(N_w N_{\mathrm{atom}} N_{\mathrm{local}})$ and that of synthetic data GPR-ANN training with $\mathcal{O}(N_w N_{\mathrm{atom}} M)$. As the examples of the previous sections demonstrated, a fixed multiple of $M = 10$–40 additional structures generated via random atomic displacement is sufficient to obtain GPR-ANN potentials comparable to direct force training with 100% force information with regard to the accuracy and robustness for all considered metrics. For condensed phases and typical cutoff radii, the number of atoms within twice the local atomic environment, N_{local}, is at least one order of magnitude greater than M and can be even larger for materials with high density. For example, for water

Fig. 6 | Detailed analysis of the atomic forces in ethylene carbonate adsorbed on the lithium (100) surface predicted by different ANN potentials. **a–d** Correlation between the predicted absolute force and the DFT reference. **e–h** Error in the force direction with respect to DFT reference. The predictions are based on a committee of 10 ANN potentials obtained from energy-only training (**a**, **e**), indirect force training with the Taylor-ANN (**b**, **f**) and GPR-ANN (**c**, **g**) methods, and direct force training with 100% force information (**d**, **h**). The color encodes the frequency of occurrence with a logarithmic scale. The solid black line in the top panels **a–d** corresponds to a perfect agreement with the DFT reference, and the dashed black lines indicate a difference of 1 eV/Å, and the red circles and ellipses highlight outliers with force errors exceeding 1 eV/Å and critical force direction errors in high-magnitude forces. For all force training methods, optimal parameters were used: Taylor-ANN ($\delta = 0.008$ Å, multiple = 36), GPR-ANN ($\delta = 0.034$ Å, multiple = 36), and direct force training (100% forces, alpha = 0.3).



and copper, a recent review reported that typical local atomic environments contain between 50 and 150 atoms within cutoff range R_c[88], so that 400 to 1200 atoms are within the range $2R_c$ required for force evaluation. On the other hand, the fitting of the GPR models also requires computation, which gives rise to a pre-factor. Therefore, in the following, we benchmark the efficiency of the indirect force training approach by comparing the memory and computer time required by GPR-ANN training and direct force training.

For this benchmark, we used the entire reference data set of 5168 Li-EC DFT interface calculations comprised of 17 heterogeneous subsets. This includes the above example of a single EC molecule adsorbed on the Li(100) surface (subset 17), and the other subsets consist of different numbers of atoms and compositions generated with different protocols to sample diverse structural configurations of the interface (see the "Methods" section for details). Each of the subsets was divided into training and test data, with 2100 and 3068 training and test points, respectively (see the "Methods" section for details).

Figure 7a–d compares the efficiency of direct force training and GPR-ANN training in terms of memory usage and training time per epoch across various choices of batch sizes and cutoff radii (R_c) for the atomic environment descriptors ($\sigma_i^{R_c}$ in Eq. (1)). The ANN potentials were trained on a single CPU of our local computer cluster (Intel Xeon Gold 6226 2.9 GHz). Both methods were trained using the optimal parameters identified in the previous Li-EC example, i.e., $\delta = 0.034$ Å and $M = 36$ for the GPR-ANN approach and 100% forces and $\alpha = 0.3$ for direct force training.

The GPR-ANN approach benefits from the inherent parallelism of the local GPR models, which can be fitted separately for each data subset (Fig. 1). Furthermore, fitting GPR models on small homogeneous data sets containing 50–150 structures and generating additional structures within the local structural spaces contributes negligibly to the overall computational cost, especially in terms of memory use, since it only needs to be done once before the first ANN training epoch. For example, for the combined Li/EC interface data system, the total time of GPR preprocessing including the construction of the 17 GPR models and data augmentation took around 5–20 min, compared to the total ANN training time of 5–10 h for 5000 epochs. Therefore, for relevant cutoff radii and batch sizes, the GPR-ANN training consistently requires less memory than direct force training (Fig. 7a, b), and the computer time is lower or comparable (Fig. 7c, d). In addition, memory usage and training time with the GPR-ANN approach are essentially independent of the cutoff radius, whereas the computational cost of direct force training grows as $\mathcal{O}(R_c^6)$ with the cutoff radius for condensed phases. Despite its lower memory usage compared to direct force training, the GPR-ANN method achieves comparable accuracy in both energy and force predictions, with over an order of magnitude improvement in energy and nearly two orders of magnitude in force predictions compared to energy-only training, as demonstrated in Figs. S10, S11.

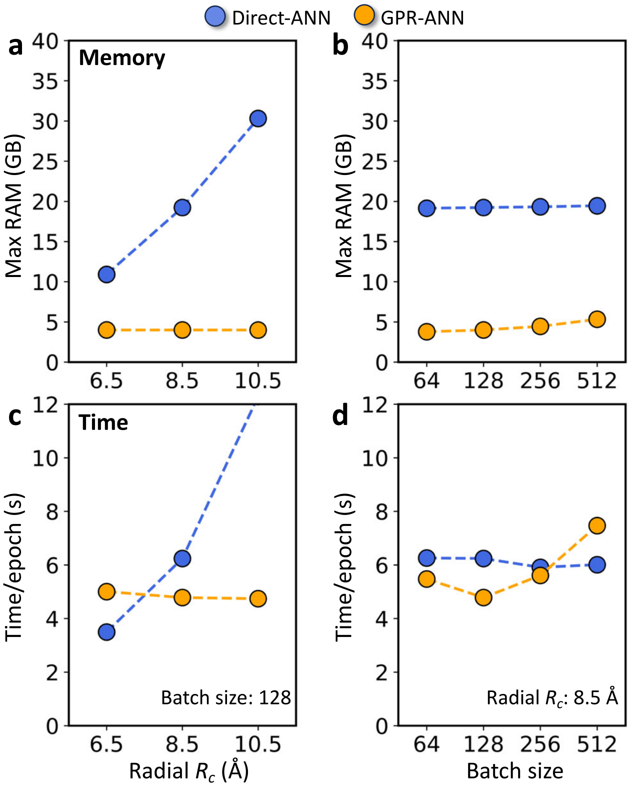


Fig. 7 | Comparison of the memory and computer time required by GPR-ANN and direct force training for Li-EC interface structures. Maximum random-access memory (RAM) required for direct force training (blue circles) and GPR-ANN training (orange circles) as a function of **a** the radial cutoff radius and **b** the batch size. The training times per epoch for both training methods are shown as a function of **c** the radial cutoff radius and **d** the batch size.

Discussion

Data augmentation has previously been proposed as an approach for implicit force training, and we compared the GPR-ANN method with the Taylor-ANN method by Cooper et al.[74] that is based on a first-order Taylor expansion. As seen in the benchmark results, the GPR-ANN approach is significantly more robust with respect to the choice of the additional structures that are labeled with synthetic energies. Specifically, when additional structures are derived from reference structures via the random displacement of atoms, the Taylor-ANN approach works best for small displacement amplitudes where the potential energy varies approximately linearly with respect to the reference energy. In contrast, the non-linear GPR models are also able to fit the PES in regions further away from the reference structures. This is important since the optimal displacement amplitude for the Taylor-ANN method depends on the curvature of the PES and is, therefore, system-dependent. For example, the optimal displacement amplitude for the EC dimers and the adsorbed EC molecule was 0.003 Å and 0.008 Å, respectively. Soft bonds (e.g., Li-Li bonds in Li metal) can tolerate greater displacements than stiff bonds (e.g., C-C bonds in EC molecules), and in complex systems such as interfaces, it can become challenging to select displacement amplitudes in practice. The GPR-ANN approach mostly avoids this parameter dependence and works well in our test systems for a wide range of displacement amplitudes.

GPR models excel at reproducing unknown PESs based on small reference datasets. Additionally, GPR models provide an intrinsic measure of model uncertainty that can be used to confirm whether the predicted energy is robust, ensuring the diversity and reliability of augmented energy data. Hence, the GPR-ANN approach can provide benefits not only for efficiently sampling reference training data but also for preventing the inclusion of inaccurate synthetic energy data (Fig. 1) as compared to methods such as Taylor-ANN.

While the primary purpose of the GPR-ANN method is indirect force training, its model uncertainty also enables efficient active learning, reducing overall data requirements. Estimating uncertainties with ANNs alone requires computationally expensive committees of multiple ANNs, and thus, the computational effort of training uncertainty models is a multiple of the effort of direct force training. In contrast, in the GPR-ANN training process, GPR surrogate models provide uncertainty estimates at no additional cost, allowing for non-redundant reference data sampling with Bayesian learning strategies prior to training multiple ANNs. Note that the DFT reference data for the present work were obtained using traditional sampling methods, i.e., molecular dynamics simulations, random atomic displacements, and conformal sampling (Figs. S12, S14, S16). In particular, for the heterogeneous EC/Li interface data sets, the structures in each subset were independently generated by perturbation of an initial structure via sampling so that only the atomic coordinates changed, maintaining compositional homogeneity. As a result, these data sets were inherently partitioned into homogeneous subsets. In cases where training begins with an existing database that requires division into homogeneous subsets, structures can be grouped based on their composition, number of atoms, and (for example) the fingerprint-based similarity of their chemical environments[89,90].

Finally, we conclude that indirect force training is not always the best option. For low-density materials or molecular data sets, the memory requirements for direct force training can be moderate so that the additional pre-factor of ~40 due to additional synthetic data is less favorable than the cost of direct force training. Hence, it depends on the target system and the potential cutoff whether the GPR-ANN approach is effective, and its utility is greatest for condensed-matter systems.

In conclusion, we have introduced a GPR-based data-augmentation approach that indirectly incorporates atomic forces into the training of ANN potentials via synthetic energy data. The approach bypasses directly training ANN potentials on interatomic forces, which is computationally demanding and can become infeasible due to the $\mathcal{O}(R_c^6)$ scaling with the range of the potential R_c. For four test systems with increasing complexity—the dihydrogen molecule, ethylene carbonate dimers, an ethylene carbonate molecule adsorbed on the surface of lithium metal, and heterogeneous data for diverse Li-EC interfaces—we showed that the GPR-ANN approach yields ANN potentials with accuracy and robustness on a par with direct force training across various metrics. We showed that scaling challenges of the GPR models can be avoided using separate local GPR models, each trained on small subsets of the overall data. For training on hybrid-functional DFT data of the Li-EC interface system, indirect force training with the GPR-ANN approach significantly lowers the memory requirement compared to traditional direct force training without compromising the training time, ANN potential accuracy, robustness, or transferability. The GPR-ANN approach, furthermore, provides an estimate of a model uncertainty without a need for ensemble models that can be used for Bayesian active learning strategies. As system complexity grows, the GPR-ANN method provides a scalable alternative to traditional direct force training in developing accurate potentials and reduces the need for costly additional reference calculations. This paves the way for constructing ANN potentials for complex condensed matter systems, such as the interfaces in lithium-ion and lithium-metal batteries.

Methods

In addition to the description below, details of the reference data, the GPR models, and the ANN potentials are summarized in supplementary Tables S1-S3.

Reference data

A Lennard–Jones potential, as implemented in the atomic simulation environment (ASE)[91] library, was used to generate an approximate PES for the H_2 molecule. The energy and force for 7 equally-spaced H-H bond lengths between 1 and 2 Å were generated as training reference data, and 200 equally-spaced points between 0.95 and 2.05 Å were generated for testing.

The 1000 EC dimer reference structures were generated by random displacement of the ground-state structure. The energies and interatomic forces were evaluated with hybrid-functional DFT calculations using the all-electron electronic structure program FHI-aims in which the Kohn-Sham states are expanded as linear combinations of numerical atomic orbitals[92,93]. The HSE06 functional[94,95] and FHI-aims' default *tight* basis set were employed for the non-periodic EC dimer structures. Relativistic effects were taken into account on the level of the zeroth Order Regular Approximation (ZORA)[96].

The 1000 reference data points were divided into 250 training and 750 test data points, and their relative energy distribution with respect to the minimum ground-state energy is shown in Fig. S12. Additional structures for implicit force training with the Taylor-ANN and GPR-ANN methods were generated by randomly displacing all the constituent atoms of the 250 training structures. Random displacements were obtained from a Gaussian distribution, and the amount of displacement was controlled via the standard deviation parameter δ. Figure S13a shows the relative energy distribution of Taylor-augmented (green bar) and GPR-augmented (red bar) energies of the same additional structures with different δ parameters ranging from 0.003 Å to 0.044 Å. For the GPR-ANN approach, full GPR models were constructed using all the energy and force information of the 250 training structures to predict the energies of the additional structures.

The representation plot shows the predicted energies of 1000 additional (synthetic) structures as an example, which is 4 times the number of the reference training structures, i.e., the augmentation multiple is $M = 4$. For $\delta \geq 0.013$ Å, some of the energies predicted with linear Taylor expansion are lower than the ground-state energy, i.e., have values below zero. This shows the limitation of the linear Taylor-ANN method since the energy of no structure can be lower than the ground-state energy.

Figure S13b shows the distribution of energy difference between the Taylor- and GPR-augmented energies. When δ is small, the Taylor and GPR energies are almost identical, as expected. However, as δ increases, the Taylor-expansion energies become increasingly lower than the GPR-augmented energies. This is the case when the curvature of relevant PES is positive, as schematically described in Fig. S13c.

The 800 reference structures of EC molecules on Li metal were sampled by random displacement of the ground-state structure, and they were evaluated using HSE06 DFT calculations using FHI-aims. All details of the calculation were the same as in the previous section, except that the Li-EC structures were represented as periodic slab models, and the DFT calculations were performed with 5 × 5 × 1 k-point meshes. It should be noted that atomic forces obtained from DFT calculations are very sensitive with respect to the density of the k-point meshes, and 5 × 5 × 1 meshes gave converged results in our tests. Figure S14 shows the relative energy distribution of the reference structures, which were divided into 200 training and 600 test data points.

Additional structures for the data-augmentation approaches were generated by random displacements, as described above for EC dimers, using the same standard deviation parameter δ as in the previous section. Figure S15a shows the relative distribution of Taylor-augmented (green bar) and GPR-augmented (red bar) energies of the same synthetic structures corresponding to different δ values ranging from 0.003 Å to 0.044 Å. As for the EC dimer, the first-order Taylor expansion underestimates the energy of perturbed structures for large δ values compared to the GPR energies (Fig. S15b). However, there are some structures $0.008\ \text{Å} \le \delta \le 0.034\ \text{Å}$ where the Taylor-expansion energies are higher than the GPR energies, implying that the PES of the Li-EC system is more complex than that of the EC dimers, and reference samples around regions with negative curvature are included as well.

All data was combined into a heterogeneous dataset for diverse Li-EC interface structures. For each initial structure in subsets with a different number of EC molecules and Li atoms, ab initio molecular dynamics (AIMD) simulations, molecular dynamics simulations using preliminary ANN potentials, random displacement, and internal (i.e., conformer) sampling were performed to generate subset data. All the generated structures were evaluated with HSE06 DFT calculations using FHI-aims with 5 × 5 × 1 k-point meshes for periodic cells and a single Γ k-point for non-periodic cells. For each data subset, a representative atomic structure is shown in Fig. S16 along with each of the sampling methods and the number of training and test data points.

GPR surrogate model

All GPR construction and GPR-based data augmentation were performed using the ænet-GPR package developed for the present work and available at https://github.com/atomisticnet/aenet-gpr. In all examples presented here, full GPR models were utilized, accounting for the covariance between two function values, between a function value and a derivative, and between two derivatives[80,85,86] with the squared exponential as the kernel function. System-specific parameters are detailed below.

A GPR model for H_2 molecule was constructed based on the energies and forces of seven equally-spaced training points. We used flattened Cartesian coordinates as the global fingerprint. Using PyTorch's *autograd* functionality, the weight and scale parameters of the kernel function were optimized by iteratively minimizing the energy loss function for the 200 test points. After 100 iterations, the default parameters of the weight and scale converged to 8.5 and 0.2, respectively (Fig. S17). Figure 17a, b shows the GPR energy and force predictions with the default kernel parameters while Fig. S17c, d shows the predictions after the parameter optimization. This GPR model with optimized hyperparameters was used to augment energy data for the GPR-ANN training.

For the EC molecule dimers, as for the H_2 molecule, flattened Cartesian coordinates were used as structural fingerprints. We performed a grid search to optimize the GPR kernel parameters for this high-dimensional system, and the optimized parameters, which minimize the energy loss function for the 750 test data points, are 1.0 and 1.5 for the weight and scale, respectively. Figure S18a, b shows the correlation between the GPR-predicted absolute force and the DFT reference before and after the kernel parameter optimization. The GPR model fitted to the energy and forces of 250 reference training data points with the optimized kernel parameters was used to evaluate GPR-augmented energies.

As fingerprint for the GPR construction for the EC/Li interface system, we tested both flattened Cartesian coordinates and smooth overlap of atomic positions (SOAP) descriptors[97] as implemented in the DScribe library[98]. A SOAP descriptor was generated with a cutoff radius of 5.0 Å, 6 radial basis functions, and a maximum degree of spherical harmonics of 4. As summarized in Fig. S19, the GPR model based on the SOAP descriptor (kernel parameters: weight = 5.0, scale = 6.0) optimized by the grid search shows an optimal correlation with the DFT reference, and this model was adopted as a surrogate model to augment energy data.

For the heterogeneous interface data set, the descriptor and kernel parameters of the GPR model were not further optimized, and we used the parameters identified as optimal in the previous section. Using the same kernel parameters, 17 local GPR models were individually fitted to homogeneous training data in each data subset, and the separate local GPR models representing respective PESs were used to generate local synthetic energies for each subsystem.

ANN potential training

All of the ANN training and prediction was carried out using the atomic energy network (ænet)[42] and ænet-PyTorch[87] packages. The *AdamW* optimization algorithm[99,100] with a learning rate of 0.0001 and a regularization parameter of 0.001 was used for all of the training runs. Atomic environments were represented using a Chebyshev descriptor[83]. The system-specific parameters of the Chebyshev descriptors and ANN architectures are described in the following.

For the H-H dimer, a Chebyshev descriptor was constructed with a radial cutoff radius of 8.0 Å and a radial expansion order of 10. No angular expansion was used for this linear molecule. The ANN architecture was N-5-5-1, where N is the descriptor dimension, the ANN gives a single output value (the atomic energy), and the two hidden layers each had five nodes. Hyperbolic tangent activation functions were used.

For the EC-EC dimers, the Chebyshev descriptors for the elements C, H, and O were constructed as follows: the radial and angular expansion orders were 12 and 4, respectively, and the radial and angular cutoff radii were 6.5 and 4.0 Å, respectively. The ANN architecture for each of the elements was 36-10-10-1 with hyperbolic tangent activation functions. The batch size was 32 for energy-only and direct force training, while a batch size of 256 was used for Taylor-ANN and GPR-ANN training.

For the Li-EC structures, the Chebyshev descriptors used the same parameters as those of the EC molecule above for all elements, and the ANN architecture was also identical.

For the heterogeneous Li-EC database, the same radial and angular expansion orders (12 and 4) were used for the Chebyshev descriptors. In order to compare the memory and cost overhead with respect to the cutoff radius for atomic descriptors, several different radial and angular cutoff radii were tested: 6.5 and 4.0 Å, 8.5 and 6.0 Å, and 10.5 and 8.0 Å. The ANN architecture for each of the elements was 36-10-10-1 with hyperbolic tangent activation functions as before, and different batch sizes were tested as described in the main text.

Data availability

The reference Li/C/H/O dataset can be obtained from the Materials Cloud repository (https://doi.org/10.24435/materialscloud:w6-9a). The data set contains atomic structures and interatomic forces in the XCrySDen structure format (XSF), and total energies are included as additional meta information.

Code availability

This work made use of the free and open-source atomic energy network (ænet), ænet-PyTorch package. The source code can be obtained either from the ænet Web site (http://ann.atomistic.net) or from GitHub (https://github.com/atomisticnet/aenet-PyTorch). The GPR code with example input and output files and a tutorial can also be obtained from the ænet-GPR GitHub (https://github.com/atomisticnet/aenet-gpr).

Received: 19 December 2024; Accepted: 8 May 2025;

References

1. Yeu, I. W., Han, G., Hwang, C. S. & Choi, J.-H. An ab initio approach on the asymmetric stacking of GaAs nanowires grown by a vapor-solid method. *Nanoscale* **12**, 17703–17714 (2020).
2. Cheula, R. et al. Local ordering of molten salts at NiO crystal interfaces promotes high-index faceting. *Angew. Chem.* **133**, 25595–25600 (2021).
3. Artrith, N. & Kolpak, A. M. Understanding the composition and activity of electrocatalytic nanoalloys in aqueous solvents: A combination of DFT and accurate neural network potentials. *Nano Lett.* **14**, 2670–2676 (2014).
4. Winter, M. The solid electrolyte interphase–the most important and the least understood solid electrolyte in rechargeable li batteries. *Z. Phys. Chem.* **223**, 1395–1406 (2009).
5. Heiskanen, S. K., Kim, J. & Lucht, B. L. Generation and evolution of the solid electrolyte interphase of lithium-ion batteries. *Joule* **3**, 2322–2333 (2019).
6. Xu, W. et al. Lithium metal anodes for rechargeable batteries. *Energ. Environ. Sci.* **7**, 513–537 (2014).
7. Eroglu, D., Ha, S. & Gallagher, K. G. Fraction of the theoretical specific energy achieved on pack level for hypothetical battery chemistries. *J. Power Sources* **267**, 14–19 (2014).
8. Pang, Y., Pan, J., Yang, J., Zheng, S. & Wang, C. Electrolyte/electrode interfaces in all-solid-state lithium batteries: A review. *Electrochem. Energy Rev.* **4**, 169–193 (2021).
9. Famprikis, T., Canepa, P., Dawson, J. A., Islam, M. S. & Masquelier, C. Fundamentals of inorganic solid-state electrolytes for batteries. *Nat. Mater.* **18**, 1278–1291 (2019).
10. Banerjee, A., Wang, X., Fang, C., Wu, E. A. & Meng, Y. S. Interfaces and interphases in all-solid-state batteries with inorganic solid electrolytes. *Chem. Rev.* **120**, 6878–6933 (2020).
11. Xiao, Y. et al. Understanding interface stability in solid-state batteries. *Nat. Rev. Mater.* **5**, 105–126 (2020).
12. Li, Y. et al. Revealing nanoscale passivation and corrosion mechanisms of reactive battery materials in gas environments. *Nano Lett.* **17**, 5171–5178 (2017).
13. Yuan, X., Liu, B., Mecklenburg, M. & Li, Y. Ultrafast deposition of faceted lithium polyhedra by outpacing SEI formation. *Nature* **620**, 86–91 (2023).
14. Jain, A., Shin, Y. & Persson, K. A. Computational predictions of energy materials using density functional theory. *Nat. Rev. Mater.* **1**, 15004 (2016).
15. Schleder, G. R., Padilha, A. C., Acosta, C. M., Costa, M. & Fazzio, A. From dft to machine learning: recent approaches to materials science–a review. *J. Phys. Mater.* **2**, 032001 (2019).
16. Artrith, N., Garrido Torres, J. A., Urban, A. & Hybertsen, M. S. Data-driven approach to parameterize scan+ u for an accurate description of 3 d transition metal oxide thermochemistry. *Phys. Rev. Mater.* **6**, 035003 (2022).
17. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *J. Chem. Phys.* **132**, 154104 (2010).
18. Debnath, S. et al. Accurate quantum chemical reaction energies for lithium-mediated electrolyte decomposition and evaluation of density functional approximations. *J. Phys. Chem. A* **127**, 9178–9184 (2023).
19. Behler, J. First principles neural network potentials for reactive simulations of large molecular and condensed systems. *Angew. Chem. - Int. Ed.* **56**, 12828–12840 (2017).
20. Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **145**, 170901 (2016).
21. Himanen, L., Geurts, A., Foster, A. S. & Rinke, P. Data-driven materials science: Status, challenges, and perspectives. *Adv. Sci.* **6**, 1900808 (2019).
22. Deringer, V. L., Caro, M. A. & Csänyi, G. Machine learning interatomic potentials as emerging tools for materials science. *Adv. Mater.* **31**, 1902765 (2019).
23. Schmidt, J., Marques, M. R. G., Botti, S. & Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Mater.* **5**, 1–36 (2019).
24. Handley, C. M. & Popelier, P. L. A. Potential energy surfaces fitted by artificial neural networks. *J. Phys. Chem. A* **114**, 3371–3383 (2010).
25. 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).
26. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
27. Shapeev, A. V. Moment Tensor Potentials: A Class of Systematically Improvable Interatomic Potentials. *Multiscale Model. Sim.* **14**, 1153–1173 (2016).
28. Schütt, K. et al. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. *Adv. Neural Inf. Process. Syst.* **30**, 992–1002 (2017).
29. Batzner, S. et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).
30. Chen, C., Ye, W., Zuo, Y., Zheng, C. & Ong, S. P. Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals. *Chem. Mater.* **31**, 3564–3572 (2019).
31. Chen, C. & Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* **2**, 718–728 (2022).
32. Deng, B. et al. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat. Mach. Intell.* **5**, 1031–1041 (2023).
33. Drautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys. Rev. B* **99**, 014104 (2019).
34. Batatia, I., Kovács, D. P., Simm, G. N. C., Ortner, C. & Csányi, G. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. *Adv. Neur.* **35**, 11423–11436 (2022).
35. Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **134**, 074106 (2011).
36. Khaliullin, R. Z., Eshet, H., Kühne, T. D., Behler, J. & Parrinello, M. Nucleation mechanism for the direct graphite-to-diamond phase transition. *Nat. Mater.* **10**, 693–697 (2011).
37. Eshet, H., Khaliullin, R. Z., Kühne, T. D., Behler, J. & Parrinello, M. Ab initio quality neural-network potential for sodium. *Phys. Rev. B* **81**, 184107 (2010).
38. Artrith, N. & Behler, J. High-dimensional neural network potentials for metal surfaces: A prototype study for copper. *Phys. Rev. B* **85**, 045439 (2012).
39. Boes, J. R., Groenenboom, M. C., Keith, J. A. & Kitchin, J. R. Neural network and ReaxFF comparison for Au properties. *Int. J. Quantum Chem.* **116**, 979–987 (2016).
40. Sun, G. & Sautet, P. Metastable structures in cluster catalysis from first-principles: Structural ensemble in reaction conditions and metastability triggered reactivity. *J. Am. Chem. Soc.* **140**, 2812–2820 (2018).
41. Artrith, N., Hiller, B. & Behler, J. Neural network potentials for metals and oxides - First applications to copper clusters at zinc oxide. *Phys. Status Solidi B* **250**, 1191–1203 (2013).
42. Artrith, N. & Urban, A. An implementation of artificial neural-network potentials for atomistic materials simulations: Performance for TiO2. *Comput. Mater. Sci.* **114**, 135–150 (2016).

43. Artrith, N., Morawietz, T. & Behler, J. High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide. *Phys. Rev. B* **83**, 153101 (2011).
44. Elias, J. S. et al. Elucidating the nature of the active phase in copper/ceria catalysts for CO oxidation. *ACS Catal.* **6**, 1675–1679 (2016).
45. Kobayashi, R., Giofré, D., Junge, T., Ceriotti, M. & Curtin, W. A. Neural network potential for Al-Mg-Si alloys. *Phys. Rev. Mater.* **1**, 053604 (2017).
46. Sosso, G. C., Miceli, G., Caravati, S., Behler, J. & Bernasconi, M. Neural network interatomic potential for the phase change material GeTe. *Phys. Rev. B* **85**, 174103 (2012).
47. Morawietz, T., Singraber, A., Dellago, C. & Behler, J. How van der Waals interactions determine the unique properties of water. *Proc. Natl Acad. Sci.* **113**, 8368–8373 (2016).
48. Smith, J. S., Isayev, O. & Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **8**, 3192–3203 (2017).
49. Cooper, A. M., Hallmen, P. P. & Kästner, J. Potential energy surface interpolation with neural networks for instanton rate calculations. *J. Chem. Phys.* **148**, 094106 (2018).
50. Morawietz, T. et al. Hiding in the crowd: Spectral signatures of overcoordinated hydrogen-bond environments. *J. Phys. Chem. Lett.* **10**, 6067–6073 (2019).
51. Li, W., Ando, Y., Minamitani, E. & Watanabe, S. Study of Li atom diffusion in amorphous Li3PO4 with neural network potential. *J. Chem. Phys.* **147**, 214106 (2017).
52. Artrith, N., Urban, A. & Ceder, G. Constructing first-principles phase diagrams of amorphous LixSi using machine-learning-assisted sampling with an evolutionary algorithm. *J. Chem. Phys.* **148**, 241711 (2018).
53. Deringer, V. L. et al. Realistic atomistic structure of amorphous silicon from machine-learning-driven molecular dynamics. *J. Phys. Chem. Lett.* **9**, 2879–2885 (2018).
54. Lacivita, V., Artrith, N. & Ceder, G. Structural and compositional factors that control the Li-Ion conductivity in LiPON electrolytes. *Chem. Mater.* **30**, 7077–7090 (2018).
55. Natarajan, S. K. & Behler, J. Neural network molecular dynamics simulations of solid-liquid interfaces: water at low-index copper surfaces. *Phys. Chem. Chem. Phys. PCCP* **18**, 28704–28725 (2016).
56. Quaranta, V. et al. Maximally resolved anharmonic OH vibrational spectrum of the water/ZnO(1010) interface from a high-dimensional neural network potential. *J. Chem. Phys.* **148**, 241720 (2018).
57. Omranpour, A., Elsner, J., Lausch, K. N. & Behler, J. Machine learning potentials for heterogeneous catalysis. *ACS Catal.* **15**, 1616–1634 (2025).
58. Vandermause, J., Xie, Y., Lim, J. S., Owen, C. J. & Kozinsky, B. Active learning of reactive Bayesian force fields applied to heterogeneous catalysis dynamics of H/Pt. *Nat. Commun.* **13**, 5183 (2022).
59. Artrith, N. Machine learning for the modeling of interfaces in energy storage and conversion materials. *J. Phys. Energy* **1**, 032002 (2019).
60. Zheng, J. et al. Reversible epitaxial electrodeposition of metals in battery anodes. *Science* **366**, 645–648 (2019).
61. Etxebarria, A. et al. Work function evolution in Li anode processing. *Adv. Energy Mater.* **10**, 2000520 (2020).
62. Li, X., Wang, Q., Guo, H., Artrith, N. & Urban, A. Understanding the onset of surface degradation in $LiNiO_2$ cathodes. *ACS Appl. Energy Mater.* **5**, 5730–5741 (2022).
63. Witkoskie, J. B. & Doren, D. J. Neural network models of potential energy surfaces: Prototypical examples. *J. Chem. Theory Comput.* **1**, 14–23 (2005).
64. Marques, M. R. G., Wolff, J., Steigemann, C. & Marques, M. A. L. Neural network force fields for simple metals and semiconductors: construction and application to the calculation of phonons and melting temperatures. *Phys. Chem. Chem. Phys. PCCP* **21**, 6506–6516 (2019).
65. Pukrittayakamee, A. et al. Simultaneous fitting of a potential-energy surface and its corresponding force fields using feedforward neural networks. *J. Chem. Phys.* **130**, 134101 (2009).
66. Gastegger, M. & Marquetand, P. High-dimensional neural network potentials for organic reactions and an improved training algorithm. *J. Chem. Theory Comput.* **11**, 2187–2198 (2015).
67. Zhang, L., Han, J., Wang, H., Car, R. & E, W. Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.* **120**, 143001 (2018).
68. Jinnouchi, R., Miwa, K., Karsai, F., Kresse, G. & Asahi, R. On-the-fly active learning of interatomic potentials for large-scale atomistic simulations. *J. Phys. Chem. Lett.* **11**, 6946–6955 (2020).
69. Yang, M., Bonati, L., Polino, D. & Parrinello, M. Using metadynamics to build neural network potentials for reactive events: The case of urea decomposition in water. *Catal. Today* **387**, 143–149 (2022).
70. 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).
71. Ang, S. J., Wang, W., Schwalbe-Koda, D., Axelrod, S. & Gómez-Bombarelli, R. Active learning accelerates ab initio molecular dynamics on reactive energy surfaces. *Chem* **7**, 738–751 (2021).
72. Zeng, J., Cao, L., Xu, M., Zhu, T. & Zhang, J. Z. H. Complex reaction processes in combustion unraveled by neural network-based molecular dynamics simulation. *Nat. Commun.* **11**, 5713 (2020).
73. Young, T. A., Johnston-Wood, T., Deringer, V. L. & Duarte, F. A transferable active-learning strategy for reactive molecular force fields. *Chem. Sci.* **12**, 10944–10955 (2021).
74. Cooper, A. M., Kästner, J., Urban, A. & Artrith, N. Efficient training of ANN potentials by including atomic forces via Taylor expansion and application to water and a transition-metal oxide. *npj Comput. Mater.* **6**, 54 (2020).
75. Schran, C., Brezina, K. & Marsalek, O. Committee neural network potentials control generalization errors and enable active learning. *J. Chem. Phys.* **153**, 104105 (2020).
76. Gal, Y. & Ghahramani, Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *Int. Conf. Mach. Learn.* **48**, 1050–1059 (2016).
77. Gibson, J., Hire, A. & Hennig, R. G. Data-augmentation for graph neural network learning of the relaxed energies of unrelaxed structures. *npj Comput. Mater.* **8**, 211 (2022).
78. Bartók, A. P. et al. Machine learning unifies the modeling of materials and molecules. *Sci. Adv.* **3**, e1701816 (2017).
79. Bartók, A. P., Kermode, J., Bernstein, N. & Csányi, G. Machine learning a general-purpose interatomic potential for silicon. *Phys. Rev. X* **8**, 041048 (2018).
80. Bartók, A. P. & Csányi, G. Gaussian approximation potentials: A brief tutorial introduction. *Int. J. Quantum Chem.* **115**, 1051–1057 (2015).
81. Kamath, A., Vargas-Hernández, R. A., Krems, R. V., Carrington, T. & Manzhos, S. Neural networks vs Gaussian process regression for representing potential energy surfaces: A comparative study of fit quality and vibrational spectrum accuracy. *J. Chem. Phys.* **148**, 241702 (2018).
82. Zuo, Y. et al. Performance and cost assessment of machine learning interatomic potentials. *J. Phys. Chem. A* **124**, 731–745 (2020).
83. Artrith, N., Urban, A. & Ceder, G. Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species. *Phys. Rev. B* **96**, 014112 (2017).
84. Politzer, P. & Murray, J. S. The Hellmann-Feynman theorem: A perspective. *J. Mol. Model.* **24**, 266 (2018).
85. Chmiela, S. et al. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **3**, e1603015 (2017).
86. Deringer, V. L. et al. Gaussian process regression for materials and molecules. *Chem. Rev.* **121**, 10073–10141 (2021).
87. López-Zorrilla, J. et al. net-PyTorch: A GPU-supported implementation for machine learning atomic potentials training. *J. Chem. Phys.* **158**, 164105 (2023).

88. Tokita, A. M. & Behler, J. How to train a neural network potential. *J. Chem. Phys.* **159**, 121501 (2023).
89. De, S., Bartók, A. P., Csányi, G. & Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **18**, 13754–13769 (2016).
90. Guo, H., Wang, Q., Urban, A. & Artrith, N. Artificial intelligence-aided mapping of the structure–composition–conductivity relationships of glass–ceramic lithium thiophosphate electrolytes. *Chem. Mater.* **34**, 6702–6712 (2022).
91. Hjorth Larsen, A. et al. The atomic simulation environment-"a Python library for working with atoms. *J. Phys. Condens. Matter* **29**, 273002 (2017).
92. Blum, V. et al. Ab initio molecular simulations with numeric atom-centered orbitals. *Comp. Phys. Commun.* **180**, 2175–2196 (2009).
93. Havu, V., Blum, V., Havu, P. & Scheffler, M. Efficient integration for all-electron electronic structure calculation using numeric basis functions. *J. Comput. Phys.* **228**, 8367–8379 (2009).
94. Heyd, J., Scuseria, G. E. & Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **118**, 8207–8215 (2003).
95. Heyd, J., Scuseria, G. E. & Ernzerhof, M. Erratum: -œHybrid functionals based on a screened Coulomb potential" [J. Chem. Phys. 118, 8207 (2003)]. *J. Chem. Phys.* **124**, 219906 (2006).
96. Van Lenthe, E., Baerends, E. J. & Snijders, J. G. Relativistic total energy using regular approximations. *J. Chem. Phys.* **101**, 9783–9792 (1994).
97. Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).
98. Himanen, L. et al. DScribe: Library of descriptors for machine learning in materials science. *Comput. Phys. Commun.* **247**, 106949 (2020).
99. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization (2017). ArXiv preprint arXiv:1412.6980.
100. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization (2019). ArXiv preprint arXiv:1711.05101.

Acknowledgements

We acknowledge support by the Columbia Center for Computational Electrochemistry (CCCE) and computing resources from Columbia University's Shared Research Computing Facility. J.L.Z. and N.A. thank the Project HPC-EUROPA3 (Grant No. INFRAIA-2016-1-730897) for its support, provided through the EC Research and Innovation Action under the H2020 Program. N.A. also acknowledges support by a start-up grant (Dutch Sector Plan) from Utrecht University. The authors gratefully acknowledge discussions with José A. Garrido Torres and Thomas Bligaard.

Author contributions

N.A. and A.U. conceived the project and the GPR-ANN workflow and supervised the research. I.W.Y. implemented the GPR-ANN methodology and workflow. I.W.Y., A.S., N.A., and A.U. implemented the structural sampling. I.W.Y. performed DFT calculations with A.S. and benchmarked calculation parameters. J.L.Z. implemented and tested direct force training in ænet-PyTorch. I.W.Y. and A.S. wrote the first draft of the manuscript. N.A., A.U., and I.W.Y. revised the manuscript. J.S., D.D.R., and R.A.F. contributed to discussions during project meetings. All authors reviewed and approved the final manuscript.

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/s41524-025-01651-0.

Correspondence and requests for materials should be addressed to Alexander Urban or Nongnuch Artrith.