

## ARTICLE OPEN



## Machine learned interatomic potentials using random features

Gurjot Dhaliwal<sup>1</sup>, Prasanth B. Nair<sup>2,4</sup> and Chandra Veer Singh<sup>1,3,4</sup>

We present a method to model interatomic interactions such as energy and forces in a computationally efficient way. The proposed model approximates the energy/forces using a linear combination of random features, thereby enabling fast parameter estimation by solving a linear least-squares problem. We discuss how random features based on stationary and non-stationary kernels can be used for energy approximation and provide results for three classes of materials, namely two-dimensional materials, metals and semiconductors. Force and energy predictions made using the proposed method are in close agreement with density functional theory calculations, with training time that is 96% lower than standard kernel models. Molecular Dynamics calculations using random features based interatomic potentials are shown to agree well with experimental and density functional theory values. Phonon frequencies as computed by random features based interatomic potentials are within 0.1% of the density functional theory results. Furthermore, the proposed random features-based potential addresses scalability issues encountered in this class of machine learning problems.

npj Computational Materials (2022)8:7; <https://doi.org/10.1038/s41524-021-00685-4>

## INTRODUCTION

Classical molecular dynamics (MD) based on empirical force fields has emerged as a powerful technique for predicting material properties and behaviour at the atomistic scale. This is because the use of empirical models of interatomic interactions enable significantly faster and scalable computations compared to ab initio methods such as density functional theory (DFT)<sup>1</sup>. The central component of any MD simulation is the Interatomic Potential (IP) function that determines interatomic interactions<sup>1,2</sup>. The choice of IP can significantly alter the property predictions, thus affecting the derived insights from the corresponding MD simulations<sup>3,4</sup>. It has been shown that IPs based on physical and chemical theories can be sensitive to small changes in their parameters<sup>5,6</sup>. Furthermore, due to their rigid physically motivated functional forms, they cannot be generalized across different material systems. An alternative is to use IPs developed using machine learning (ML) tools<sup>7–17</sup>. These models make fewer assumptions on the functional form of the IP and can be easily generalized to different material systems<sup>10,18</sup>.

The most popular ML approaches for IP development include linear regression<sup>12,14</sup>, kernel regression models<sup>7,15</sup>, and neural networks<sup>8,11</sup>. All of these models use atomic descriptors, functions describing local atomic environment, as inputs and predict energy, and its higher-order derivatives as outputs<sup>19,20</sup>. More recently, Zhang et al.<sup>21</sup> and Jia et al.<sup>22</sup> proposed neural network-based IPs, that use the atomic coordinates as input and learn the underlying descriptors through an embedding neural network. Training of embedding as well as prediction network is performed end-to-end. A review of ML-based IPs can be found in refs. <sup>16,17,23–27</sup>. The present study focuses on kernel regression models such as Gaussian approximation potential (GAP)<sup>7</sup>. GAP is based on Gaussian processes and approximates the local atomic energy with a kernel defining the similarity between different atomic descriptors. This kernel-based IP has been used to predict the energy/forces for a wide range of

materials including metals, semiconductors, amorphous solids etc.<sup>9,28–32</sup>. Another popular IP based on kernel regression is Adaptive Generalizable Neighbourhood Informed (AGNI) potential, which has been mainly used for modelling interactions in metals<sup>15</sup>.

Parameter estimation in standard kernel methods costs typically  $\mathcal{O}(N^3)$ , where  $N$  is the number of training points. In addition, standard kernel methods require storage and factorization of a dense ( $N \times N$ ) Gram matrix, which poses a significant computational and memory bottleneck for large  $N$ . In IP learning applications,  $N$  can easily be in the range of  $10^6$ – $10^7$ , requiring significant computational resources for parameter estimation. The large Gram matrix also hinders the hyper-parameter tuning process where multiple model runs are required to identify good settings for the hyper-parameters such as kernel length scales, cut-off radius etc. Slow training becomes a bottleneck when more than one element is involved e.g., alloys, or when active learning methods are used for training<sup>13,33</sup>. Another major disadvantage of kernel methods is that prediction at a new point costs  $\mathcal{O}(ND)$ , where  $D$  is the number of input features. Depending on the descriptors used,  $D$  can easily take values of the order  $10^2$ , leading to slower MD run time. In the GAP approach, a sparse approximation is used to reduce the training complexity to  $\mathcal{O}(N^2N)$  and prediction complexity to  $\mathcal{O}(N'D)$ , where  $N'$  denotes the number of sparse points<sup>7</sup>.

In order to address the high computational cost associated with the application of ML to learn IPs, this study presents a method that approximates the local atomic energy as a linear combination of random features associated with a kernel. The basis function used for energy modelling is a low-dimensional representation of the infinite-dimensional feature space associated with a kernel. The models presented here are tested using random features based on both stationary and non-stationary kernels. Energy and forces are predicted for three material systems—graphene, diamond, and tungsten. For each material system, mean absolute

<sup>1</sup>Department of Mechanical and Industrial Engineering, University of Toronto, 5 King's College Rd, Toronto M5S 3G8 ON, Canada. <sup>2</sup>Institute for Aerospace Studies, University of Toronto, 4925 Dufferin St, Toronto M3H 5T6 ON, Canada. <sup>3</sup>Department of Materials Science and Engineering, University of Toronto, 184 College St, Toronto M5S 3E4 ON, Canada. <sup>4</sup>These authors jointly supervised this work: Prasanth B. Nair, Chandra Veer Singh. ✉email: pbn@utias.utoronto.ca; chandraveer.singh@utoronto.ca

error (MAE) for force predictions are of the order  $10^{-2}$  eVÅ<sup>-1</sup> as compared to the corresponding DFT results. Performance of IPs based on random features is compared against state-of-the-art ML-based IPs such as GAP<sup>7</sup> and empirical IP such as AIREBO<sup>34</sup>, EAM<sup>35</sup> etc. Apart from providing better energy/force fittings, IPs based on random features significantly reduce the training time by 96% as compared to GAP. The main reason for this is the smaller number of parameters needed to fit the energy function of random features based IPs. The low-dimensional parameter space has a direct impact on the computational cost associated with energy/force evaluations during MD run time, with fewer parameters leading to faster evaluations as detailed in Supplementary Note 3.

## RESULTS

### Random feature models

In this and the following sections, an atomic configuration will refer to a particular group of atoms in ordered or disordered state. The total number of atomic configurations in the training set will be denoted by  $L$ , with the  $l$ th configuration having  $N_l$  atoms, where  $l = 1, 2, \dots, L$ . Each atom is described by rotationally and translationally invariant descriptors. A simple descriptor can be the distance between two atoms or functions of distance as discussed by Behler<sup>11</sup>. More complex descriptors include functions based on spherical harmonics of charge density expansion such as smooth overlap of atomic positions (SOAP)<sup>20</sup>. The goal is to learn a model for energy and forces given the atomic descriptors.

The local atomic energy,  $E$ , for an atom described by the descriptor  $\mathbf{q}$ , can be approximated as follows

$$E(\mathbf{q}) = \sum_{l=1}^L \sum_{t=1}^{N_l} w_t K(\mathbf{q}, \mathbf{q}^t), \quad (1)$$

where  $K(\mathbf{q}, \mathbf{q}^t)$  is a kernel function denoting the similarity measure between atomic descriptors  $\mathbf{q}$  and  $\mathbf{q}^t$ . The contribution of each training atomic descriptor,  $\mathbf{q}^t$ , is weighted by a scalar weight,  $w_t$ , where the index  $t$  goes over all the atoms in the training set. The above model of energy approximation forms the basis of various ML-based interatomic potentials<sup>7</sup>. Two types of kernels: stationary such as squared exponential (used mostly for two-body and three-body descriptors) or non-stationary such as polynomial kernel (used mostly for SOAP descriptors) are used to approximate the atomic energy as per Eq. 1. However, as mentioned earlier the training cost and runtime associated with kernel methods can be exorbitant for large values of  $N$  and  $D$ . More specifically, the training data used to learn IP models can easily scale to 100,000 atomic configurations, which makes full kernel evaluation infeasible in practice.

In order to circumvent the issues associated with full kernel evaluation, we use the idea of random features. Kernel approximations based on random features have been used to solve a range of challenging problems in machine learning<sup>36–38</sup>. However, the use of these models to develop IPs has not been explored yet. In the following sections, we discuss how random features corresponding to stationary and non-stationary kernels can be used to efficiently learn IPs from DFT simulation datasets.

To approximate stationary kernels, we use the notion of random Fourier features (RFF). The main idea behind RFF is that stationary kernels can be approximated by a dot product of random features as follows

$$K(\mathbf{q}, \mathbf{q}^t) \approx \mathbf{z}(\mathbf{q})^T \mathbf{z}(\mathbf{q}^t), \quad (2)$$

where  $\mathbf{z}(\mathbf{q})$  is a vector of random features associated with the atom descriptor  $\mathbf{q} \in \mathbb{R}^D$ . For stationary kernels such as the squared exponential, these random features are obtained by

invoking Bochner's theorem which states that a kernel is positive definite if and only if it is a Fourier transform of a non-negative measure. Writing the Fourier transform of the kernel as  $K(\mathbf{q} - \mathbf{q}^t) = \int p(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}^T(\mathbf{q} - \mathbf{q}^t)} d\boldsymbol{\omega}$ , and using the ideas presented by Rahimi and Recht<sup>39</sup>, the kernel can be approximated in the form of Eq. 2, with

$$\mathbf{z}(\mathbf{q}) = \sqrt{\frac{2}{M}} [\cos(\boldsymbol{\omega}_1^T \mathbf{q}), \sin(\boldsymbol{\omega}_1^T \mathbf{q}), \dots, \cos(\boldsymbol{\omega}_M^T \mathbf{q}), \sin(\boldsymbol{\omega}_M^T \mathbf{q})] \in \mathbb{R}^{2M}, \quad (3)$$

and  $\boldsymbol{\omega}_m \sim p(\boldsymbol{\omega})$ ,  $m = 1, 2, \dots, M$ , are random vectors of length  $D$  (for details see algorithm 1). For different kernels, the distribution  $p(\boldsymbol{\omega})$  takes the specific forms as detailed by Rahimi and Recht<sup>39</sup>. Detailed derivations for the above basis functions and generalization error bounds can be found in Rahimi and Recht<sup>39–41</sup>. The local atomic energy can be approximated using the random features as follows

$$E(\mathbf{q}) = \sum_{m=1}^{2M} \alpha_m z_m(\mathbf{q}), \quad (4)$$

where  $\alpha_m$  are the model weights and  $z_m(\mathbf{q})$  denotes the  $m$ th component of the vector  $\mathbf{z}(\mathbf{q})$  defined in (3). A smooth function that zeroes out the energy contributions of atoms falling outside a cutoff radius can be used with the above energy model. The mathematical form of the cutoff function used in this work is described in the Methods section.

For many body features such as SOAP, a dot product or polynomial kernel provides a better approximation of energy than a stationary kernel. A dot product kernel is of the form  $K(\mathbf{q}, \mathbf{q}^t) = (\mathbf{q}^T \mathbf{q}^t + r)^n$ , where  $n \in \mathbb{N}$  and  $r \in \mathbb{R}^+$ . This kernel can be approximated using a random features map (RFM) based on results from harmonic analysis<sup>42</sup>. According to Shor's theorem, a kernel is positive definite if it is an analytic function admitting Maclaurin expansion<sup>42</sup>. The  $m$ th component of the random feature map that provides an unbiased estimate of the dot-product kernel can be written as

$$z_m(\mathbf{q}) = \sqrt{\gamma} \prod_{j=1}^n \boldsymbol{\omega}_j^T \mathbf{q}, \quad (5)$$

where  $\gamma$  is a randomly chosen scaled Maclaurin coefficient as defined in algorithm 2 and  $\boldsymbol{\omega}_j \in \{-1, 1\}^D$  is a random vector. To reduce the variance due to random sampling,  $n$  copies of  $\boldsymbol{\omega}$  are generated and multiplied together in the preceding equation. Kar and Karnick<sup>42</sup> emphasized that the coefficients of the Maclaurin expansion needs to be non-negative, otherwise it will lead to indefinite kernels. Similar to stationary kernel approximation, the energy of an atomic configuration, considering a non-stationary kernel, is modelled by a linear combination of these random features as follows

$$E(\mathbf{q}) = \sum_{m=1}^M \beta_m z_m(\mathbf{q}), \quad (6)$$

where  $\beta_m$  are model weights to be estimated from the training data. Algorithms to generate random features for both stationary and non-stationary kernels are detailed in the Methods section.

### Parameter estimation

The energy of an atom can be modelled using random feature models as per Eqs. 4 and 6. The weight vectors,  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$ , are estimated by minimizing the regularized least-squares loss

function

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{l=1}^L \left( \left( \sum_{i=1}^{N_l} E_i^l - E^{l,*} \right)^2 + \sum_{i=1}^{N_l} \sum_{j=1}^{N_l} \sum_{k=1}^3 \left( F_{ijk}^l - F_{ijk}^{l,*} \right)^2 \right) + \lambda \left( \|\boldsymbol{\alpha}\|_2^2 + \|\boldsymbol{\beta}\|_2^2 \right), \quad (7)$$

where  $E_i^l$  denotes the predicted values of local energy for the  $i$ th atom belonging to the  $l$ th atomic configuration and  $F_{ijk}^l$  denotes the predicted forces in direction  $k$  for the  $(i, j)$  atom pair in atomic configuration  $l$ . The local energy is modelled using (3) and/or (6) while the predicted forces are obtained by differentiating the energy models as outlined in the Methods section. Quantities of interest calculated using DFT (available in the training dataset) are marked by an asterisk, where  $E^{l,*}$  denotes the DFT calculated energy for configuration  $l$  and  $F_{ijk}^{l,*}$  denotes DFT calculated force in direction  $k$  for atom pair  $(i, j)$ .  $\lambda$  denotes the regularization parameter which can be estimated by cross-validation. Since this loss function is quadratic in  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$ , the weights can be estimated by solving a system of linear algebraic equations. The complexity of weight estimation through this method is  $\mathcal{O}(M^2N)$  which is significantly lower than traditional full kernel methods ( $\mathcal{O}(N^3)$ ). Similarly, the computational cost of prediction at a test point is  $\mathcal{O}(M)$  as compared to  $\mathcal{O}(ND)$  for classical kernel methods. Here,  $N$  denotes the training set size,  $D$  denotes the number of descriptors and  $M$  denotes the total number of random features.

In the present study, we approximate the squared exponential kernel using RFFs. By writing the Fourier transform of this kernel, the corresponding probability distribution for  $\boldsymbol{\omega}$  can be written as  $\boldsymbol{\omega} \sim \mathcal{N}(\mathbf{0}, \sigma^{-1}\mathbf{I})$ , where  $\sigma$  denotes the shape parameter of the kernel<sup>39</sup> and  $\mathbf{I}$  denotes the  $D \times D$  identity matrix. From the above probability distribution, random samples for  $\boldsymbol{\omega}$  are generated, which are used to obtain  $\mathbf{z}(\mathbf{q})$  as defined in Eq. 3. To make this approach deterministic,  $\boldsymbol{\omega}$  can be generated from the above distribution using quasi-random sequences<sup>43</sup>. Random features are obtained for dot product kernels of order 4 using RFM as described in the Results section.

## Data

We tested the IP based on random features on three classes of materials namely metals, semiconductors, and 2D materials. Datasets for tungsten, diamond and graphene were obtained from libatoms.org, as accessed on March 3, 2019. The data consist of energy and forces for a range of configurations computed through first principle methods, details of which can be found in references<sup>7,28,31</sup>. For graphene, we used the same training and validation dataset as detailed in Rowe et al.<sup>28</sup>. For the tungsten dataset, we used GAP<sub>5</sub> as training and GAP<sub>6</sub> as test set<sup>31</sup>. The atomic configurations included in the sets GAP<sub>5</sub> and GAP<sub>6</sub> are described in ref. <sup>31</sup>. Random partition of diamond data was performed to obtain training and test set. For diamond and tungsten, we used SOAP descriptors, and thus only the RFM model was applied to these datasets. To isolate the performance of the RFF model, we employed two-body and three-body descriptors only for the graphene datasets. The parameters for the descriptors are kept the same as presented in refs. <sup>7,28,31</sup>.

Figure 1 shows the predicted forces versus the DFT computed forces for graphene, diamond, and tungsten. As can be seen from Fig. 1, the forces predicted closely agree with the DFT computed values. R-squared value for these plots is more than 0.99 exhibiting the excellent fit between predicted and true forces. MAE in energy and force prediction in these three classes of materials is detailed in the inset of Fig. 1. Energy and force fittings exhibit low MAE with energy fittings approaching the chemical accuracy of 1 kcal/mol (0.043 eV).

## Graphene potential

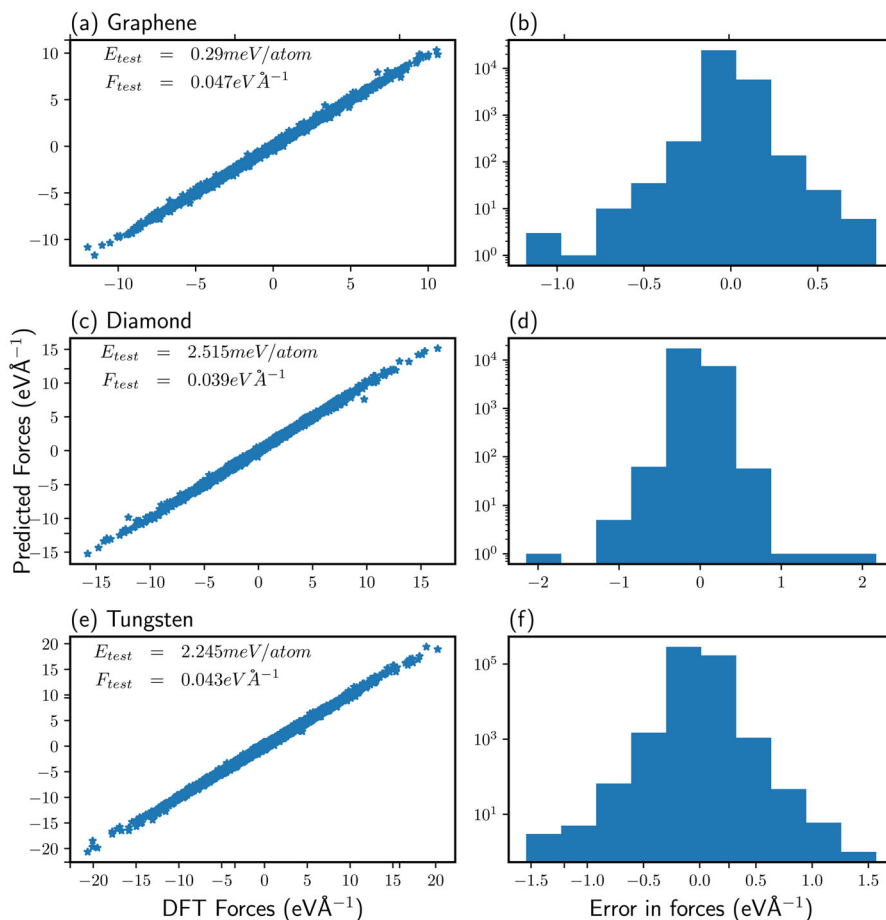
Performance of IPs based on random features was also compared against existing classical IPs. On the graphene dataset, Rowe et al.<sup>28</sup> compared GAP IP to classical IPs as well as other DFT methods. We computed the same error metric considering RFF, i.e., root mean squared error of in-plane and out-plane forces, and found these numbers to be 0.070 and 0.044 eVÅ<sup>-1</sup>, respectively. Though these error metrics are slightly higher than GAP, they are significantly lower than the other IPs<sup>28</sup>.

Apart from excellent energy/force fittings, another computational advantage considering RFFs is discussed here. Table 1 compares the MAE in energy and forces as predicted by GAP potential and RFF potential for two-body (GAP(2B), RFF(2B)) and three-body (GAP(3B), RFF(3B)) descriptors, respectively. It can be easily seen from Table 1 that the RFF model provides a much better energy fitting than GAP while also maintaining comparable force-fitting accuracies. Particularly, MAE for energy fittings considering three body descriptors is lower than GAP (3B) by two orders of magnitude while both IPs exhibit similar performance for force fittings. The last column in Table 1 shows the time taken to train each potential on the same dataset. It is clearly evident that RFF based IP not only provides better energy and force fittings, they also outperform GAP in terms of training time. The main drawback of RFF potential is that the number of parameters required for energy/force fittings is higher than the corresponding GAP potentials. This will result in slower prediction of energy/forces during MD run-time.

In order to reduce the number of parameters, we constrained each sample of  $\boldsymbol{\omega}$  to be orthogonal to each other. The importance of orthogonal random numbers has been discussed previously in other works and corresponding performance bounds can be found in refs. <sup>44–46</sup>. We will refer to RFF model with orthogonal random numbers as O-RFF and corresponding performance metrics are detailed in Table 1. By introducing orthogonal random numbers, the number of parameters reduced by one order of magnitude, while maintaining accuracy in energy/force predictions. Furthermore, training time is reduced by approximately 96% as compared to RFF, leading to significant computational savings. Supplementary Note 2 details the energy and force error convergence as number of random features is increased in O-RFF based IPs. It can be easily seen that, for both the two-body and three-body descriptors, error drops at much faster rate and plateaus earlier when orthogonal random numbers are used.

To test the ability of RFF based IP, an extensive set of structural, thermal, and elastic MD simulations were performed on monolayer graphene. MD simulations were performed using the Atomic Simulation Environment (ASE) library<sup>47</sup>. Predictions made through RFF based IP were compared against those obtained from classical and GAP potentials. Calculations involving GAP and classical potentials were performed using the QUIP framework within the ASE library<sup>47</sup>. Reproducibility of DFT and experimental values is also assessed and comparison to RFF based predictions is presented in Table 3.

The first set of MD simulations that we analyzed correspond to structural properties such as cohesive energy and lattice constant. In-plane lattice parameter was obtained by fitting Birch-Murnaghan equation of state (EOS)<sup>48</sup> on energy and volume data. Corresponding minimum energy configuration provided the cohesive energy of the graphene sheet. From Table 3, it can be seen that RFF prediction of cohesive energy is close to GAP. The predicted value of cohesive energy also lies within 1% of the DFT predictions. From Table 3, it is evident that prediction using AIREBO potential is off by 4% from DFT computed values. It is worth noting that both GAP and RFF



**Fig. 1 Prediction accuracy of potential based on random features.** Plot of forces predicted through random features versus the DFT computed values. Data was obtained from libatoms.org. Mean absolute error in energy ( $E_{test}$ ) and force ( $F_{test}$ ) prediction is shown in inset for the corresponding test sets (a, c, e). Histogram of the force prediction errors is provided for graphene (b), diamond (d), and tungsten (f).

Potential	MAE $E_{test}$ (meV/atom)	MAE $F_{test}$ ( $eV\text{\AA}^{-1}$ )	Parameters	Training time (seconds)
GAP (2B)	52.705	0.234	50	144
RFF (2B)	28.9459	0.325	200	103
O-RFF (2B)	27.18	0.353	20	51.4
GAP (3B)	11.185	0.058	250	6060
RFF (3B)	0.415	0.049	4000	4229
O-RFF (3B)	0.29	0.047	200	256

The metric used is mean absolute error. GAP model was trained considering only 2 body interactions, denoted as GAP (2B) and considering 3 body interactions only, denoted as GAP (3B). Training time is measured in seconds.

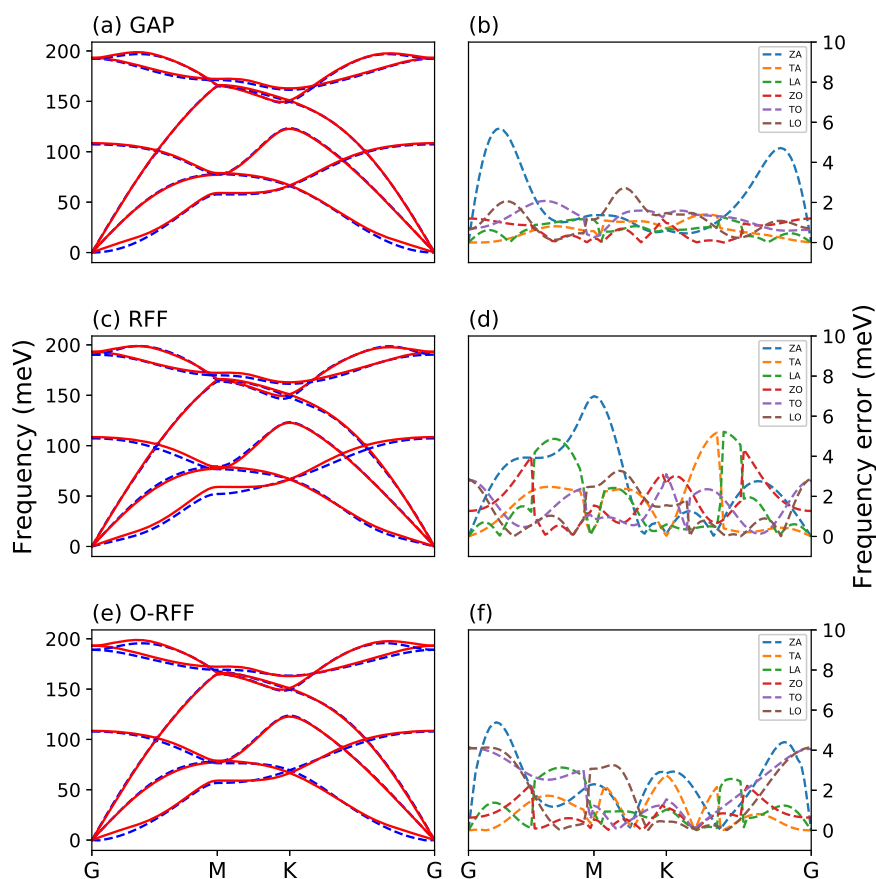
over-predicts cohesive energy as compared to experimental value of  $-7.73$  eV.

We also consider the elastic properties of monolayer graphene sheet. The elastic constants,  $C_{11}$  and  $C_{12}$  were computed by fitting elastic tensor to stress–strain data. Independent normal and shear strained configurations were obtained and corresponding stress tensors were computed after performing force relaxation. For stress calculations, we

considered the thickness of graphene sheet to be  $3.34 \text{ \AA}$ . MD simulations reveal that  $C_{11}$  as predicted by RFF was within 0.2% of the GAP predictions. Value of  $C_{12}$  as obtained from RFF is  $221.93 \text{ GPa}$ , which is showing a deviation of 12% from the GAP prediction. Compared to the value obtained from experiments, RFF computed value for  $C_{11}$  is underestimated by 8%. This deviation can be addressed by including more strained configurations during the training phase.

Table 3 also shows the properties predicted using the O-RFF potential. O-RFF predicted values of lattice constant and cohesive energy are in close agreement with the full RFF model. The values of  $C_{11}$  are off by 2% while  $C_{12}$  are off by 3.2% as compared to RFF.

Another important quantity of interest is the phonon dispersion curve for graphene. To model thermal transport properties such as thermal expansion coefficient and specific heat, it is highly critical to accurately determine phonon frequencies. It has been shown in previous studies that classical IPs fail to reproduce experimentally or DFT obtained phonon dispersion curves. Considering ML-based IP, GAP has been shown to accurately reproduce the experimentally obtained phonon frequencies<sup>28</sup>. In the present study, we compared the phonon frequency as obtained from DFT to those obtained using GAP, RFF, and O-RFF potentials. Phonon dispersion curve and frequencies were obtained using small displacement method as implemented in the ASE library<sup>47</sup>.



**Fig. 2 Comparison of phonon properties.** Phonon dispersion curves as obtained by DFT (shown by solid lines) and by GAP (a), RFF (c), and O-RFF (e). Absolute error in frequency prediction between DFT and respective potentials is shown in subfigure (b), d, and (f). Maximum absolute error in frequency prediction through RFF is 7 meV while from O-RFF its 6 meV.

The DFT study was conducted using ASE with VASP as underlying calculator.

From Fig. 2c, it is evident that the RFF predictions closely match DFT obtained phonon dispersion curves. We also compared the phonon frequencies as obtained from RFF and GAP. Absolute error in frequencies as predicted by both the potentials is less than 7 meV, thus confirming that RFF based IP can accurately reproduce phonon dispersion curve within the experimental tolerance. Figure 2e compares the phonon dispersion curves as obtained from O-RFF and DFT. The absolute error in frequencies is shown in Fig. 2f. It is clear that the O-RFF predictions also closely match DFT obtained phonon dispersion curve with an accuracy of 6 meV.

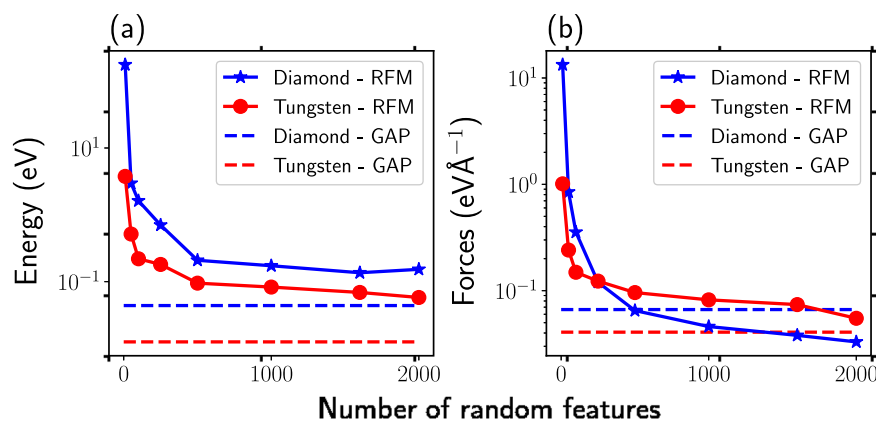
### Diamond and Tungsten potentials

Considering diamond and tungsten data, Fig. 3 shows the dependence of (a) MAE in energy and (b) force on the number of random features. For both the material systems, a clear trend of reduction of MAE in energy and force fittings is evident. Performance is compared against the GAP IP, where dashed lines show corresponding MAE in energy and forces. As the number of random features is increased, errors in force fittings, Fig. 3b, for both diamond and tungsten falls below the GAP benchmarks. MAE in energy fittings for both the materials reduces as we increase the number of random features but at different rate than the corresponding error in force fittings. Supplementary Note 1 details the variation of energy and force errors as the training sample size is increased.

Both energy and force errors converge with increase in training sample size.

Accuracy of force fittings for RFM potentials is also compared to those obtained from classical IPs. We considered the diamond data and obtained forces using AIREBO<sup>34</sup>, AIREBO-M<sup>49</sup>, and Tersoff<sup>50</sup> potentials. Force prediction for each potential is compared against DFT predictions and MAE is presented in Table 2. It can be noted that RFM has the lowest MAE for both training and test set.

To validate the proposed RFM potential, we performed a series of MD simulations and compared the results to DFT and experimental values. In order to perform MD simulations, we implemented RFM potentials using the open-source ASE library<sup>47</sup>. We start the performance analysis of the RFM potential by first assessing the structural properties of both diamond and tungsten. From the energy–volume curves, ground state energy and corresponding lattice constant is obtained for the respective material systems. Figure 4 shows energy–volume curves for diamond and BCC tungsten using both RFM and GAP<sup>7,31</sup> potentials. It is clearly evident that RFM matches GAP predicted energies with high accuracy for both the material systems. Lattice constant values are listed in Table 3 and are compared to DFT and experimental values. Comparing RFM and DFT values, the error in lattice constant prediction is within 0.01% for both the material systems. We also obtained bulk modulus, by fitting the Birch–Murnaghan equation to the energy volume data. For both the material systems the values lie closely to DFT predicted values.



**Fig. 3 Learning capacity of random features model.** Dependence of MAE in energy and force prediction on the number of random features for RFM model. Mean absolute error in energy (a) and forces (b) are shown for diamond and tungsten RFM IP. Dashed horizontal lines show the performance of GAP IP.

**Table 2.** Comparison of MAE in force predictions as predicted through RFM and other existing IPs for the diamond dataset.

	MAE $F_{\text{train}}$ ( $\text{eV}\text{\AA}^{-1}$ )	MAE $F_{\text{test}}$ ( $\text{eV}\text{\AA}^{-1}$ )
RFM	0.022	0.024
GAP <sup>7</sup>	0.067	-
AIREBO <sup>34</sup>	3.281	3.341
AIREBO-M <sup>49</sup>	3.156	3.229
Tersoff <sup>50</sup>	2.407	2.376

MAE in forces is lower by two orders of magnitude as compared to classical IPs and about three times lower than GAP.

Using linear elasticity theory, elastic constants  $C_{11}$ ,  $C_{12}$ , and  $C_{44}$  were computed for both diamond and tungsten. Pristine atomic configurations were strained within 1% and corresponding stresses were computed. The results for RFM and GAP are given in Table 3 and compared to the corresponding DFT and experimental results. Considering tungsten, the values of  $C_{11}$  and  $C_{44}$  are within 4% of the DFT values. There is a slight overprediction of  $C_{12}$  with the RFM predicted value to be 225.07 GPa while the corresponding DFT value is 203 GPa. The elastic constants obtained for the diamond dataset were within 3% of the DFT values except  $C_{44}$ . We also computed mechanical properties such as Young's modulus and Poisson's ratio and found them to be in excellent agreement with the DFT results (see Table 3). Another validation of RFM based potentials is assessed by performing simulations considering defects such as vacancies for tungsten. From Table 3, it is clear that RFM based potentials can model defects such as vacancy formation and provide accuracy levels comparable to DFT.

## DISCUSSION

The present study provides a computationally efficient approach to model interatomic interactions. The generalized linear models presented in this study provide energy and force predictions with DFT level accuracy (MAE less than  $0.03 \text{ eV}\text{\AA}^{-1}$  when compared to DFT results). By providing potentials for three different classes of materials, we have shown that the same functional form, as presented in Eqs. 4 and 6, can be used to model a wide range of material classes.

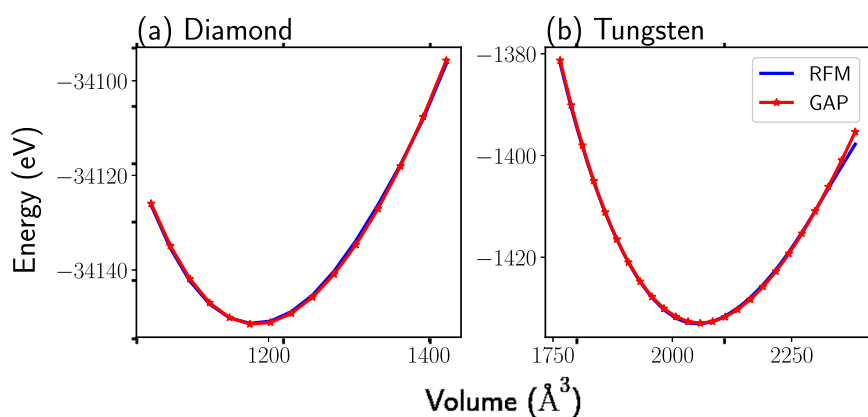
It was shown that MD simulations performed using the above approach provided results within 7% of the experimental values or DFT predictions.

In certain applications, material behaviour is accurately represented by a combination of two body, three body, and many body descriptors. In those cases, a linear combination of RFF and RFM can be easily applied. We also like to highlight that there is no restriction on using the descriptors discussed in this study. In fact, the algorithms used to generate random features (RFF/RFM) can be easily applied to Behler's symmetry functions<sup>19</sup>, bispectrum coefficients<sup>12</sup>, or the descriptors used in AGNI framework<sup>15</sup> etc.

The major advantage of the RFF/RFM based IP is the low memory requirements as compared to standard kernel methods such as GAP and AGNI. Prediction through RFF/RFM potentials does not need storage of training set and requires computational effort of  $\mathcal{O}(M)$ , where  $M$  is the number of random features. This low memory advantage of RFF/RFM based IP makes them scalable compared to standard kernel methods. This computational advantage is achieved without sacrificing prediction accuracy as is evident through MD simulation results.

The potential's functional form can be easily extended to model complex material interactions that involve different material classes. These interactions are hard to capture by classical IPs, as they are tuned to very specific chemical compositions, bonding types, and surface geometry. On the other hand, ML-based potentials can have challenges with descriptor and parameter identification for the large dataset involving different elements. In the light of these issues, RFF/RFM based potentials come up as a worthy candidate for accurate modelling of such complex interactions in a computationally efficient way. Not only have these potentials been shown to provide forces with DFT level accuracy, the simple generalized linear form can provide faster training and prediction of energy/forces in comparison to standard kernel methods. To test our above hypothesis, we developed O-RFF potentials for two different material systems (a) Li-Si-P-S<sup>51</sup> and (b) Li-P-S-Sn<sup>51</sup>. The data for both the material systems was obtained from DeepMD library (<http://dplibrary.deepmd.net/>). For both the material systems, we obtained forces agreeing well with DFT calculations, with MAE between DFT and O-RFF predicted forces around  $0.08 \text{ eV}\text{\AA}^{-1}$ . Detailed results are presented in Supplementary Note 4.

One major issue that we encountered was slow run time of RFF based IP as compared to the classical IP. The main reason



**Fig. 4** Fitting equation of states for diamond and tungsten data. Energy volume relation for **a** diamond and **b** BCC-tungsten. The relations are obtained using RFM (blue) and GAP (red)<sup>7,31</sup> IP, by fitting Birch–Murnaghan equation of state<sup>48</sup> to energy–volume data.

		Experiment	DFT	Classical IP	GAP	RFF/RFM	O-RFF
Graphene	Lattice constant (Å°)	2.459 <sup>59</sup>	2.408–2.6 <sup>60</sup>	2.419	2.467	2.453	2.461
	Cohesive energy (eV)	–7.374 <sup>61</sup>	–7.73 <sup>62</sup> , –7.906 <sup>63</sup>	–7.427	–7.964	–7.963	–7.976
	C <sub>11</sub> (GPa)	1017.96 <sup>64</sup>	1072.16 <sup>65</sup>	992.43 <sup>5</sup>	934.49	936.27	917.55
	C <sub>12</sub> (GPa)	–	194.61 <sup>66</sup>	335.58 <sup>5</sup>	198.33	221.93	214.74
Tungsten	Lattice constant (Å°)	3.165 <sup>67</sup>	3.172 <sup>68</sup>	3.1652	3.18	3.178	–
	C <sub>11</sub> (GPa)	501–521 <sup>69</sup>	517 <sup>31</sup> , 523 <sup>70</sup>	553.09	518.04	519.77	–
	C <sub>12</sub> (GPa)	199–207 <sup>69</sup>	198 <sup>31</sup> , 203 <sup>70</sup>	183.59	198.47	225.07	–
	C <sub>44</sub> (GPa)	151–160 <sup>69</sup>	142 <sup>31</sup> , 160 <sup>70</sup>	148.44	142.95	138.71	–
	Vacancy formation energy (eV)	3.1–4.0 <sup>69</sup>	3.49 <sup>70</sup>	3.69	3.29	3.39	–
	Bulk modulus (GPa)	308–314 <sup>69</sup>	320 <sup>69</sup>	283.68	303.81	323.3	–
	Poisson ratio	0.28 <sup>71</sup>	–	0.25	0.28	0.3	–
Diamond	Young’s modulus (GPa)	405 <sup>71</sup>	–	461.58	408.09	383.75	–
	Lattice constant (Å°)	3.567 <sup>72</sup>	3.568 <sup>73</sup>	3.561	3.527	3.524	–
	C <sub>11</sub> (GPa)	1074–1084 <sup>74</sup>	1060 <sup>73</sup> 1118 <sup>7</sup>	1051.96	1059.17	1043.05	–
	C <sub>12</sub> (GPa)	119–129 <sup>74</sup>	125 <sup>73</sup> 151 <sup>7</sup>	100.78	98.86	114.61	–
	C <sub>44</sub> (GPa)	576–580 <sup>74</sup>	562 <sup>73</sup> 603 <sup>7</sup>	629.68	575.11	507.02	–
	Bulk modulus (GPa)	441–445 <sup>75</sup>	463.1 <sup>76</sup>	428.45	476.56	478.48	–
	Poisson ratio	0.079 <sup>77</sup>	–	0.087	0.085	0.09	–
Young’s modulus (GPa)	1164 <sup>77</sup>	–	1034.34	1042.29	1020.38	–	

The first column exhibits experiment values. DFT predictions are stored in the second column and MD simulations using classical, GAP, and random features based IP are shown in the last four columns respectively. Classical IPs considered in this table are Airebo<sup>34</sup>, Finnis Sinclair<sup>78</sup>, and Tersoff<sup>50</sup> for graphene, tungsten and diamond respectively.

for this is the expensive evaluation of cos and sin functions at every timestep. We tackled this issue by introducing potentials based on orthogonal random numbers, which reduced the number of required random features by more than half. For different sheets of graphene, Supplementary Fig. 3 (ref. Supplementary Note 3) shows the average time taken by 10 steps of NVT ensemble considering O-RFF, RFF, and GAP<sup>28</sup> IPs. Run time of O-RFF is considerably lower than both GAP and RFF IP, while two orders of magnitude slower than the classical IP such as AIREBO<sup>34</sup>. Another way to reduce the runtime cost would be to include a sparsity-inducing regularizer in the objective function or by using Kronecker matrix algebra<sup>52</sup>. This is the focus of an ongoing study.

ML-based IPs can outperform classical empirical IP, by using a generic functional form, with DFT level accuracy for energy/force prediction. However, the main disadvantage of ML-based IP is the computational time needed for training the model and prediction of energy/forces during MD runs. Kernel-based methods are memory intensive as they have to store the training configurations for prediction purposes. The present study eliminated these limitations by providing a linear model-based IP using random features for stationary and non-stationary kernels. Energy and force fittings are comparable to DFT predictions. Low parametric space and linear functional form enabled these models to be computationally faster than DFT. Time to learn IP is significantly reduced by 96% as compared to GAP, thus allowing faster hyper parameter search and optimization.

## METHODS

### Algorithms to generate random features

The algorithms used to generate random features for both the models are presented below

---

#### Algorithm 1. RFF approximation

---

**Require:** A positive definite shift-invariant kernel i.e.  $k(\mathbf{x}, \mathbf{y}) = k(\mathbf{x} - \mathbf{y})$

**Ensure:** A randomized feature map  $\mathbf{z} : \mathbb{R}^D \rightarrow \mathbb{R}^M$  such that  $\langle \mathbf{z}(\mathbf{x}), \mathbf{z}(\mathbf{y}) \rangle \approx k(\mathbf{x}, \mathbf{y})$

- 1: Compute the Fourier transform  $p$  of the kernel  
 $k : p(\boldsymbol{\omega}) = \frac{1}{2\pi} \int e^{-j\boldsymbol{\omega}^T \boldsymbol{\delta}} k(\boldsymbol{\delta}) d\boldsymbol{\delta}$
  - 2: Draw  $M$  i.i.d samples  $\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_M \in \mathbb{R}^D$  from  $p$
  - 3:  $\mathbf{z}(\mathbf{x}) = \sqrt{\frac{2}{M}} [\cos(\boldsymbol{\omega}_1^T \mathbf{x}), \sin(\boldsymbol{\omega}_1^T \mathbf{x}), \dots, \cos(\boldsymbol{\omega}_M^T \mathbf{x}), \sin(\boldsymbol{\omega}_M^T \mathbf{x})]$
- 

---

#### Algorithm 2. RFM approximation

---

**Require:** A positive definite dot-product kernel i.e.  $k(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}^T \mathbf{y})$

**Ensure:** A randomized feature map  $\mathbf{z} : \mathbb{R}^D \rightarrow \mathbb{R}^M$  such that  $\langle \mathbf{z}(\mathbf{x}), \mathbf{z}(\mathbf{y}) \rangle \approx k(\mathbf{x}, \mathbf{y})$

- 1: Obtain the Maclaurin expansion of  $f(\xi) = \sum_{t=0}^{\infty} a_t \xi^t$  by setting  $a_t = \frac{f^{(t)}(0)}{t!}$
  - 2: Fix a value  $p_0 > 1$ .
  - 3: **for**  $i = 1, 2, \dots, M$  **do**
  - 4: Choose a nonnegative integer  $n \in \mathbb{N} \cup 0$  with  $\mathbb{P}[n = t] = \frac{1}{p_0^{t+1}}$ .
  - 5: Choose  $n$  vectors  $\boldsymbol{\omega}_1 \dots \boldsymbol{\omega}_n \in \{-1, 1\}^D$ , selecting each coordinate using fair coin tosses.
  - 6: Let feature map  $z_i : \mathbf{x} \rightarrow \sqrt{\frac{1}{M p_0^{n+1}}} \prod_{j=1}^n \boldsymbol{\omega}_j^T \mathbf{x}$ .
  - 7: **end for**
  - 8: Output  $\mathbf{z} : \mathbf{x} \rightarrow \frac{1}{\sqrt{M}} (z_1(\mathbf{x}), z_2(\mathbf{x}), \dots, z_M(\mathbf{x}))$
- 

### Cutoff radius

The cutoff function, denoted by  $f_{\text{cut}}(r_{ij})$ , is a smooth function that zeroes out the energy contributions for atoms falling outside a cutoff radius and is as described below

$$f_{\text{cut}}(r_{ij}) = 0.5 \times \left[ \cos \left( \pi \frac{r_{ij} - r_{\text{cut}} + w_{\text{cut}}}{w_{\text{cut}}} \right) + 1 \right], \quad (8)$$

where  $r_{ij}$  denotes the distance between two atoms and  $r_{\text{cut}}$  is the cutoff radius. Tuning can be achieved by setting  $w_{\text{cut}}$ . For three body descriptors such as  $[r_{ij} + r_{ik}, (r_{ij} - r_{ik})^2, r_{jk}]$ , the cutoff function takes the form  $f_{\text{cut}}(r_{ij})f_{\text{cut}}(r_{ik})$ .

### Expression for forces for RFF and RFM models

This section outlines the steps involved in computation of forces from potential energy expressions. The forces are computed as derivative of the energy and corresponding expressions were obtained through the chain rule. The derivative of energy contribution of RFF features (see Eq. 4) with respect to descriptor  $\mathbf{q}$  can be written as

$$\nabla E(\mathbf{q}) = \sum_{m=1}^{2M} a_m \nabla z_m(\mathbf{q}), \quad (9)$$

where

$$z_m(\mathbf{q}) = \begin{cases} \sqrt{\frac{2}{M}} \cos(\boldsymbol{\omega}_m^T \mathbf{q}) & \text{if } m \text{ is odd} \\ \sqrt{\frac{2}{M}} \sin(\boldsymbol{\omega}_m^T \mathbf{q}) & \text{if } m \text{ is even.} \end{cases} \quad (10)$$

It can be easily verified that for descriptor  $\mathbf{q}$  with dimension  $D$

$$\nabla z_m(\mathbf{q}) = \begin{cases} \sqrt{\frac{2}{M}} \nabla \cos(\boldsymbol{\omega}_m^T \mathbf{q}) & \text{if } m \text{ is odd} \\ \sqrt{\frac{2}{M}} \nabla \sin(\boldsymbol{\omega}_m^T \mathbf{q}) & \text{if } m \text{ is even.} \end{cases} \quad (11)$$

where

$$\nabla \cos(\boldsymbol{\omega}_m^T \mathbf{q}) = \begin{bmatrix} -\boldsymbol{\omega}_m(1) \sin(\boldsymbol{\omega}_m^T \mathbf{q}) \\ -\boldsymbol{\omega}_m(2) \sin(\boldsymbol{\omega}_m^T \mathbf{q}) \\ \vdots \\ -\boldsymbol{\omega}_m(D) \sin(\boldsymbol{\omega}_m^T \mathbf{q}) \end{bmatrix} \quad (12)$$

and

$$\nabla \sin(\boldsymbol{\omega}_m^T \mathbf{q}) = \begin{bmatrix} \boldsymbol{\omega}_m(1) \cos(\boldsymbol{\omega}_m^T \mathbf{q}) \\ \boldsymbol{\omega}_m(2) \cos(\boldsymbol{\omega}_m^T \mathbf{q}) \\ \vdots \\ \boldsymbol{\omega}_m(D) \cos(\boldsymbol{\omega}_m^T \mathbf{q}) \end{bmatrix}. \quad (13)$$

Note that  $\boldsymbol{\omega}_m(i)$  denotes the  $i$ th component of the vector  $\boldsymbol{\omega}_m$ . Following the same line of reasoning, the gradient for RFM features (Eq. 6 in the main text) can be written as

$$\nabla z_m(\mathbf{q}) = \begin{bmatrix} \sqrt{Y} \sum_{i=1}^n \boldsymbol{\omega}_{mi}(1) (\prod_{k \neq i} \boldsymbol{\omega}_{mk}^T \mathbf{q}) \\ \sqrt{Y} \sum_{i=1}^n \boldsymbol{\omega}_{mi}(2) (\prod_{k \neq i} \boldsymbol{\omega}_{mk}^T \mathbf{q}) \\ \vdots \\ \sqrt{Y} \sum_{i=1}^n \boldsymbol{\omega}_{mi}(D) (\prod_{k \neq i} \boldsymbol{\omega}_{mk}^T \mathbf{q}) \end{bmatrix}. \quad (14)$$

The final step in obtaining forces is to multiply  $\nabla z_m(\mathbf{q})$  with the gradient of the descriptor  $\mathbf{q}$  with respect to position vector  $\mathbf{r}$ . The last quantity is descriptor-dependent and should be handled accordingly.

### Phonon dispersion DFT calculation

Plane wave DFT calculations were conducted using Vienna Ab initio Simulation Package (VASP)<sup>53,54</sup> to get energy and forces. The interaction between the valence electrons and the ionic core was described by generalized gradient approximation (GGA) in the Perdew-Burke-Ernzerhof (PBE) formulation along with projector augmented wave (PAW) pseudopotentials<sup>55-57</sup>. A  $7 \times 7 \times 1$  supercell of graphene was constructed for the calculations and a  $4 \times 4 \times 1$  k-point mesh generated using Monkhorst–Pack scheme was used for Brillouin zone sampling<sup>58</sup>. An additional support grid was added for the evaluation of the augmentation charges and non-spherical contributions related to the gradient of the density in the PAW spheres were included.

### DATA AVAILABILITY

Data to train interatomic potentials was obtained from libatoms.org, as accessed on March 3, 2019.

### CODE AVAILABILITY

Software library to facilitate the reproduction of the results presented here can be found at <https://github.com/g7dhaliwal/RandomFeatures>.

Received: 4 May 2021; Accepted: 26 November 2021;

Published online: 17 January 2022

### REFERENCES

1. Allen, M. P. & Tildesley, D. J. *Computer Simulation of Liquids* (Oxford University Press, 2017).
2. Griebel, M., Knapek, S. & Zumbusch, G. *Numerical Simulation in Molecular Dynamics* Vol. 5 (Society for Industrial and Applied Mathematics, 2007).
3. Becker, C. A., Tavazza, F. & Levine, L. E. Implications of the choice of interatomic potential on calculated planar faults and surface properties in nickel. *Philos. Mag.* **91**, 3578–3597 (2011).
4. Becker, C. A., Tavazza, F., Trautt, Z. T. & de Macedo, R. A. B. Considerations for choosing and using force fields and interatomic potentials in materials science and engineering. *Curr. Opin. Solid State Mater. Sci.* **17**, 277–283 (2013).
5. Dhaliwal, G., Nair, P. B. & Singh, C. V. Uncertainty analysis and estimation of robust airebo parameters for graphene. *Carbon* **142**, 300–310 (2019).
6. Dhaliwal, G., Nair, P. B. & Singh, C. V. Uncertainty and sensitivity analysis of mechanical and thermal properties computed through embedded atom method potential. *Comput. Mater. Sci.* **166**, 30–41 (2019).



7. 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).
8. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
9. Nyshadham, C. Machine-learned multi-system surrogate models for materials prediction. *npj Comput. Mater.* **5**, 51 (2019).
10. Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A. & Kim, C. Machine learning in materials informatics: Recent applications and prospects. *npj Comput. Mater.* **3**, 54 (2017).
11. Behler, J. Representing potential energy surfaces by high-dimensional neural network potentials. *J. Phys.: Condens. Matter* **26**, 183001 (2014).
12. Thompson, A. P., Swiler, L. P., Trott, C. R., Foiles, S. M. & Tucker, G. J. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *J. Comput. Phys.* **285**, 316–330 (2015).
13. Podryabinkin, E. V. & Shapeev, A. V. Active learning of linearly parametrized interatomic potentials. *Comput. Mater. Sci.* **140**, 171–180 (2017).
14. Shapeev, A. V. Moment tensor potentials: A class of systematically improvable interatomic potentials. *Multiscale Modeling Simul.* **14**, 1153–1173 (2016).
15. Botu, V., Batra, R., Chapman, J. & Ramprasad, R. Machine learning force fields: Construction, validation, and outlook. *J. Phys. Chem. C* **121**, 511–522 (2016).
16. Hansen, K. Assessment and validation of machine learning methods for predicting molecular atomization energies. *J. Chem. Theory Comput.* **9**, 3404–3419 (2013).
17. Mueller, T., Kusne, A. G. & Ramprasad, R. Machine learning in materials science: Recent progress and emerging applications. *Rev. Comput. Chem.* **29**, 186–273 (2016).
18. Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **145**, 170901 (2016).
19. Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **134**, 074106 (2011).
20. Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).
21. Zhang, L. et al. End-to-end symmetry preserving inter-atomic potential energy model for finite and extended systems. Proceedings of the 32nd International Conference on Neural Information Processing Systems, 4441–4451 (2018).
22. Jia, W. et al. In *SC20: International Conference for High Performance Computing, Networking, Storage and Analysis* 1–14 (IEEE, 2020).
23. Rupp, M. Machine learning for quantum mechanics in a nutshell. *Int. J. Quantum Chem.* **115**, 1058–1073 (2015).
24. Liu, Y., Zhao, T., Ju, W. & Shi, S. Materials discovery and design using machine learning. *J. Mater. Chem.* **3**, 159–177 (2017).
25. Mishin, Y. Machine-learning interatomic potentials for materials science. *Acta Mater.* **214**, 116980 (2021).
26. Deringer, V. L., Caro, M. A. & Csányi, G. Machine learning interatomic potentials as emerging tools for materials science. *Adv. Mater.* **31**, 1902765 (2019).
27. Behler, J. Four generations of high-dimensional neural network potentials. *Chem. Rev.* **121**, 10037–10072 (2021).
28. Rowe, P., Csányi, G., Alfè, D. & Michaelides, A. Development of a machine learning potential for graphene. *Phys. Rev. B* **97**, 054303 (2018).
29. Deringer, V. L. & Csányi, G. Machine learning based interatomic potential for amorphous carbon. *Phys. Rev. B* **95**, 094203 (2017).
30. Bartók, A. P., Kermode, J., Bernstein, N. & Csányi, G. Machine learning a general-purpose interatomic potential for silicon. *Phys. Rev. X* **8**, 041048 (2018).
31. Szlachta, W. J., Bartók, A. P. & Csányi, G. Accuracy and transferability of Gaussian approximation potential models for tungsten. *Phys. Rev. B* **90**, 104108 (2014).
32. Rosenbrock, C. W. Machine-learned interatomic potentials for alloys and alloy phase diagrams. *npj Comput. Mater.* **7**, 1–9 (2021).
33. Li, Z., Kermode, J. R. & De Vita, A. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.* **114**, 096405 (2015).
34. Stuart, S. J., Tutein, A. B. & Harrison, J. A. A reactive potential for hydrocarbons with intermolecular interactions. *J. Chem. Phys.* **112**, 6472–6486 (2000).
35. Zhou, X. Atomic scale structure of sputtered metal multilayers. *Acta Mater.* **49**, 4005–4015 (2001).
36. Xiong, K. & Wang, S. The online random Fourier features conjugate gradient algorithm. *IEEE Signal Process. Lett.* **26**, 740–744 (2019).
37. Nelsen, N. H. & Stuart, A. M. The random feature model for input-output maps between Banach spaces. *SIAM J. Sci. Comput.* **43**, A3212–A3243 (2021).
38. Hung, T. H. & Chien, P. A random Fourier feature method for emulating computer models with gradient information. *Technometrics* **63**, 1–10 (2020).
39. Rahimi, A. & Recht, B. Random Features for Large-Scale Kernel Machines. *Adv. Neural Inf. Process. Syst.* 1177–1184 (2008).
40. Rahimi, A. & Recht, B. In *2008 46th Annual Allerton Conference on Communication, Control, and Computing* 555–561 (IEEE, 2008).
41. Avron, H., Kapralov, M., Musco, C., Musco, C., Velingker, A., & Zandieh, A. In *Proceedings of the 34th International Conference on Machine Learning* Vol. 70, 253–262 (JMLR. org, 2017).
42. Kar, P. & Karnick, H., Random feature maps for dot product kernels. *Artif. Intell. Stat.* 583–591 (2012).
43. Niederreiter, H. *Random Number Generation and Quasi-Monte Carlo Methods* (SIAM, 1992).
44. Yu, F. X. X., Suresh, A. T., Choromanski, K. M., Holtmann-Rice, D. N. & Kumar, S. Orthogonal random features. *Adv. Neural Inf. Process. Syst.* 1975–1983 (2016).
45. Choromanski, K. M., Rowland, M. & Weller, A. The Unreasonable Effectiveness of Structured Random Orthogonal Embeddings. *Adv. Neural Inf. Process. Syst.* 219–228 (2017).
46. Choromanski, K. et al. The geometry of random features. *Int. Conf. Artif. Intell. Stat.* 1–9 (2018).
47. Larsen, A. H. The atomic simulation environment—a python library for working with atoms. *J. Phys.: Condens. Matter* **29**, 273002 (2017).
48. Birch, F. Finite elastic strain of cubic crystals. *Phys. Rev.* **71**, 809 (1947).
49. O'Connor, T. C., Andzelm, J. & Robbins, M. O. Airebo-m: A reactive model for hydrocarbons at extreme pressures. *J. Chem. Phys.* **142**, 024903 (2015).
50. Tersoff, J. Modeling solid-state chemistry: Interatomic potentials for multi-component systems. *Phys. Rev. B* **39**, 5566 (1989).
51. Huang, J. et al. Deep potential generation scheme and simulation protocol for the li10gop2s12-type superionic conductors. *J. Chem. Phys.* **154**, 094703 (2021).
52. Evans, T. W. & Nair, P. B. Scalable Gaussian processes with grid-structured eigenfunctions (GP-GRIEF). *Int. Conf. Mach. Learn.* 1416–1425 (2018).
53. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558 (1993).
54. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169 (1996).
55. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).
56. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953 (1994).
57. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758 (1999).
58. Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **13**, 5188 (1976).
59. Li, L., Reich, S. & Robertson, J. Defect energies of graphite: Density-functional calculations. *Phys. Rev. B* **72**, 184109 (2005).
60. Behera, H. & Mukhopadhyay, G. In *AIP Conference Proceedings*, Vol. 1313, 152–155 (AIP, 2010).
61. Baskin, Y. & Meyer, L. Lattice constants of graphite at low temperatures. *Phys. Rev.* **100**, 544 (1955).
62. Shulenburger, L. & Mattsson, T. R. Quantum monte carlo applied to solids. *Phys. Rev. B* **88**, 245117 (2013).
63. Shin, H. et al. Cohesion energetics of carbon allotropes: Quantum Monte Carlo study. *J. Chem. Phys.* **140**, 114702 (2014).
64. Lee, C., Wei, X., Kysar, J. W. & Hone, J. Measurement of the elastic properties and intrinsic strength of monolayer graphene. *Science* **321**, 385–388 (2008).
65. Wei, X., Fragneaud, B., Marianetti, C. A. & Kysar, J. W. Nonlinear elastic behavior of graphene: Ab initio calculations to continuum description. *Phys. Rev. B* **80**, 205407 (2009).
66. Wang, R., Wang, S., Wu, X. & Liang, X. First-principles calculations on third-order elastic constants and internal relaxation for monolayer graphene. *Phys. B: Condens. Matter* **405**, 3501–3506 (2010).
67. Lide, D. R. *CRC Handbook of Chemistry and Physics* Vol. 85 (CRC Press, 2004).
68. Ochs, T., Beck, O., Elsässer, C. & Meyer, B. Symmetrical tilt grain boundaries in body-centred cubic transition metals: An ab initio local-density-functional study. *Philos. Mag. A* **80**, 351–372 (2000).
69. Wang, J., Zhou, Y., Li, M. & Hou, Q. A modified w-w interatomic potential based on ab initio calculations. *Model. Simul. Mater. Sci. Eng.* **22**, 015004 (2013).
70. Marinica, M. C. Interatomic potentials for modelling radiation defects and dislocations in tungsten. *J. Phys.: Condens. Matter* **25**, 395502 (2013).
71. Grünwald, E. et al. Young's modulus and Poisson's ratio characterization of tungsten thin films via laser ultrasound. *Mater. Today: Proc.* **2**, 4289–4294 (2015).
72. Donohue, J. *Structures of the Elements* (John Wiley and Sons, Inc., 1974).
73. Mounet, N. & Marzari, N. First-principles determination of the structural, vibrational and thermodynamic properties of diamond, graphite, and derivatives. *Phys. Rev. B* **71**, 205214 (2005).
74. McSkimin, H. & Andreatch Jr, P. Elastic moduli of diamond as a function of pressure and temperature. *J. Appl. Phys.* **43**, 2944–2948 (1972).
75. McSkimin, H., Andreatch Jr, P. & Glynn, P. The elastic stiffness moduli of diamond. *J. Appl. Phys.* **43**, 985–987 (1972).

76. Hebbache, M. First-principles calculations of the bulk modulus of diamond. *Solid State Commun.* **110**, 559–564 (1999).
77. Klein, C. A. & Cardinale, G. F. Young's modulus and Poisson's ratio of CVD diamond. *Diam. Relat. Mater.* **2**, 918–923 (1993).
78. Finnis, M. & Sinclair, J. A simple empirical n-body potential for transition metals. *Philos. Mag. A* **50**, 45–55 (1984).

## ACKNOWLEDGEMENTS

This research is supported by grants from Natural Sciences and Engineering Research Council of Canada, Hart Professorship, Canada Research Chairs programme, the University of Toronto, and Compute Canada. The authors would like to thank Abu Anand for providing DFT calculated phonon dispersion curves.

## AUTHOR CONTRIBUTIONS

Conceptualization, G.D., P.B.N., and C.V.S.; methodology, G.D. and P.B.N.; software, G.D.; formal analysis, G.D.; investigation, G.D. and P.B.N.; data curation, G.D.; writing—original draft, G.D.; writing—review and editing, G.D., P.B.N., and C.V.S.; visualization, G.D.; supervision, P.B.N. and C.V.S.; project administration, C.V.S.; funding acquisition: C.V.S.

## 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/s41524-021-00685-4>.

**Correspondence** and requests for materials should be addressed to Prasanth B. Nair or Chandra Veer Singh.

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

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



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

© The Author(s) 2022