Unified Graph Neural Network Force-field for the Periodic Table: Solid State Applications

Kamal Choudhary (0000-0001-9737-8074),[*,†,‡] Brian DeCost (0000-0002-3459-5888),[¶] Lily Major (0000-0002-5783-8432),[§,‖] Keith Butler (0000-0001-5432-5597),[‖] Jeyan Thiyagalingam (0000-0002-2167-1343),[‖] and Francesca Tavazza (0000-0002-5602-180X)[¶]

†*Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, 20899, MD, USA.*

‡*Theiss Research, La Jolla, 92037, CA, USA.*

¶*Materials Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, 20899, MD, USA.*

§*Department of Computer Science, Aberystwyth University, SY23 3DB, UK*

‖*Scientific Computing Department, Rutherford Appleton Laboratory, Science and Technology Facilities Council, Harwell Campus, Didcot, OX11 0QX, UK*

E-mail: kamal.choudhary@nist.gov

Abstract

Classical force fields (FF) based on machine learning (ML) methods show great potential for large scale simulations of solids. MLFFs have hitherto largely been designed and fitted for specific systems and are not usually transferable to chemistries beyond the specific training set. We develop a unified atomisitic line graph neural network-based FF (ALIGNN-FF) that can model both structurally and chemically diverse solids with any combination of 89 elements from the periodic table. To train the

ALIGNN-FF model, we use the JARVIS-DFT dataset which contains around 75000 materials and 4 million energy-force entries, out of which 307113 are used in the training. We demonstrate the applicability of this method for fast optimization of atomic structures in the crystallography open database and by predicting accurate crystal structures using genetic algorithm for alloys.

Introduction

Large scale atomistic simulation of multi-component systems is a difficult task but they are highly valuable for industrial applications such as designing alloys, designing electrical contacts, touch screens, transistors, batteries, composites and catalysts.[1–4] Quantum chemistry methods, such as density functional theory, are obvious approaches to simulate such systems, however, they are computationally very expensive for large systems.[5] Classical force-fields, or interatomic potentials, such as embedded-atom method (EAM),[6–8] modified embedded-atom method (MEAM), reactive bond-order (ReaxFF), charge-optimized many-body (COMB) etc.,[9–13] can be used for such simulations but they are usually parameterized for a very narrow chemical phase-space limiting their applicability and transferability. Moreover, it can be quite strenuous and time consuming to develop such traditional classical FFs.

Recently, machine learning based FFs[14,15] are being used to systematically improve accuracy of FFs and have successfully been used for multiple systems. One of the pioneer MLFFs were developed by Behler-Parinello in 2007 using neural network.[16] It was initially used for molecular systems and now has been extended to numerous other applications.[17] Although neural network is one of the most popular regressors, other methods such as Gaussian processes-based Gaussian approximation potentials (GAP),[18] as well as linear regression and basis functions-based spectral neighbor analysis potential (SNAP)[19] have also been thoroughly used. Such FFs use two and three body descriptors to describe local environment. Other popular MLFF formalism include smooth overlap of atomic positions (SOAP),[20] moment tensor potentials (MTP),[21,22] symbolic regression[23] and polynomial-

based approaches.[24] One of the critical issues in developing and maintaining classical force-fields is that they are hard to update with software and hardware changes. Luckily, MLFFs are more transparently developed and maintained than other classical FFs. A review article on this topic can be found elsewhere.[17] Nevertheless, early-generation MLFFS are also limited to a narrow chemical space and may require hand-crafted descriptors which may take time to be identified. Conventional MLFFs are usually trained on specific chemistry only as the number of parameters (cross-terms) exponentially increases with the number of elements in the system. This is where GNN based methods can be particularly useful for generalizability. We note that the number of model parameters do not scale explicitly with the number of elements, because there are no explicit cross-terms for interactions between different pairs of chemical species. Moreover, the model does not require retraining for new systems. So it could be shared and many researchers can make use of pre-trained models.

Graph neural network (GNN) based methods have shown remarkable improvements over descriptor based machine learning methods and can capture highly non-Euclidean chemical space.[15,25–33] GNN FFs have been also recently proposed and are still in the development phase.[34,35] We developed an atomistic line graph neural network (ALIGNN) in our previous work[36] which can capture many body interactions in graph and successfully models more than 70 properties of materials, either scalar or vector quantities, such as formation energy, bandgap, elastic modulus, superconducting properties, adsorption isotherm, electron and density of states etc.[36–41] The same automatic differentiation capability that allows training these complex models allows for physically consistent prediction of quantities such as forces and energies; this enables GNNs to be used in quickly identifying relaxed or equilibrium states of complex systems. However, there is a need for a large and diverse amount of data to train unified force-fields.

In this work, we present a dataset of energy and forces with 4 million entries for around 75000 materials in the JARVIS-DFT dataset which have been developed over past 5 years.[42] We extend the ALIGNN model to also predict derivatives that are necessary for FF for-

malism. There can be numerous applications of such a unified FF, however, in this work we limit ourselves to pre-optimization of structures, genetic algorithm based structure, and molecular dynamics applications. The developed model will be publicly available on the ALIGNN GitHub page (`https://github.com/usnistgov/alignn`) with several examples and a brief documentation.

Methodology

A flow-chart for training ALIGNN-FF is shown in Fig. 1a. To train the FF, we use a large DFT dataset, JARVIS-DFT, which contains about 75000 materials with a variety of atomic structures and chemistry and has been generated over the last 5 years. JARVIS-DFT is primarily based on Vienna Ab initio Simulation Package (VASP)[43,44] software and OptB88vdW[45] functional but also contains data obtained using other functionals and methods. In this work, only OptB88vdW-based data has been used. The OptB88vdW functional was shown to be very well applicable to solids in Ref.[45] and, ever since, it has been used to model rare-gas dimers and metallic, ionic, and covalent bonded solids, polymers, and small molecular systems.[46] Energies and forces are available for each structure optimization and elastic constant calculation runs. The total number of such entries is around 4 million. Although it would be justified to train on the entire dataset, we choose to use only a subset of it because of the computational budget and hardware requirements available to us. Instead of the 4 million datapoints, we use 307113 points i.e. more than an order magnitude less, by taking unique set of first, last, middle, maximum energy and minimum energy structures only. If some snapshots for a run are identical, say the last step and the minimum energy configuration are the same, then we only include one of them. The dataset consists of perfect structures only.

We convert the atomic structures to a graph representation using atomistic line graph neural network (ALIGNN). Details on ALIGNN can be found in the related paper.[36] In

brief, each node in the atomistic graph is assigned 9 input node features based on its atomic species: electronegativity, group number, covalent radius, valence electrons, first ionization energy, electron affinity, block and atomic volume. The inter-atomic bond distances are used as edge features with radial basis function up to 8 Å cut-off. We use a periodic 12-nearest-neighbor graph construction. This atomistic graph is then used for constructing the corresponding line graph using interatomic bond-distances as nodes and bond-angles as edge features. ALIGNN uses edge-gated graph convolution for updating nodes as well as edge features. One ALIGNN layer composes an edge-gated graph convolution on the bond graph with an edge-gated graph convolution on the line graph. The line graph convolution produces bond messages that are propagated to the atomistic graph, which further updates the bond features in combination with atom features.

In this work, we developed the functionality for atomwise and gradient predictions in the ALIGNN framework. Quantities related to gradients of the predicted energy, such as forces on each atom, are computed by applying the chain rule through the automatic differentiation system used to train the GNN. Usually MLFF datasets are not transferable. We share a large dataset of energy and forces (available on Figshare repository: `https://doi.org/10.6084/m9.figshare.21667874`), which can be used for other applications as well. Importantly, this dataset is continuously expanding making it very systematic and transferable across multiple elements and their combinations. In a closed system, the forces on each atom i depend on its position with respect to every other particle j through a force-field as:

$$m_i \frac{d^2 r_i(t)}{dt^2} = \sum_j F_{ij}(t) = -\sum_j \nabla_i U(r_{ij}(t)) \tag{1}$$

where r_{ij} is the distance between atom i and j. For stress predictions, we use the Virial stress[47] which is a measure of mechanical stress on an atomic scale for homogeneous systems:

$$\sigma_{\alpha\beta} = -\frac{1}{V} \sum_i \sum_{i \neq j} F_{ij}^{\alpha} r_{ij}^{\beta} + m_i v_i^{\alpha} v_i^{\beta} \tag{2}$$

The 307113 data points are split into 90:5:5 ratio for training, validation and testing. We train the model for 250 epochs using the same hyper-parameters as in the original ALIGNN model.[36] ALIGNN is based on deep graph library (DGL),[48] PyTorch[49] and JARVIS-Tools packages.[42] We optimize a composite loss function (l) with weighted mean absolute error terms for both forces and energies:

$$l = \left|E^{DFT} - E^{GNN}\right| + w \sum_{i}^{N_{atoms}} |F_i^{DFT} - F_i^{GNN}| \tag{3}$$

where, E^{DFT} and E^{GNN} are energies per atom using DFT and ALIGNN, F_i^{DFT} and F_i^{GNN} are forces acting on an atom using DFT and ALIGNN, w is a tunable weighting factor scaling the force contribution to the loss relative to the energies; in this work we set $w = 10$. Without such weighting, it is difficult to learn forces as they vary within a wide range. All the DFT data were obtained from the JARVIS-DFT dataset which uses VASP calculations with OptB88vdW functional.

The ALIGNN-FF model has been integrated with atomic simulation environment (ASE)[50] as an energy, force and stress calculator for structure optimization and MD simulation. This calculator can be used for optimizing atomic structures, using genetic algorithm,[51] and running molecular dynamics simulations, for example constant-temperature, constant-volume ensemble (NVT) simulations. The structural relaxations are carried out with the fast inertial relaxation engine (FIRE),[52] available in ASE. In order to predict equation of state/ energy-volume-curve (EV) simulation, we apply volumetric strains between ranges of -0.1 to 0.1 with an interval of 0.01. Although the current implementation is in ASE only, we plan to implement this FF in high-performance MD codes such as LAMMPS[53] in future which can provide significantly better performance.

Results and discussion

Performance on test set

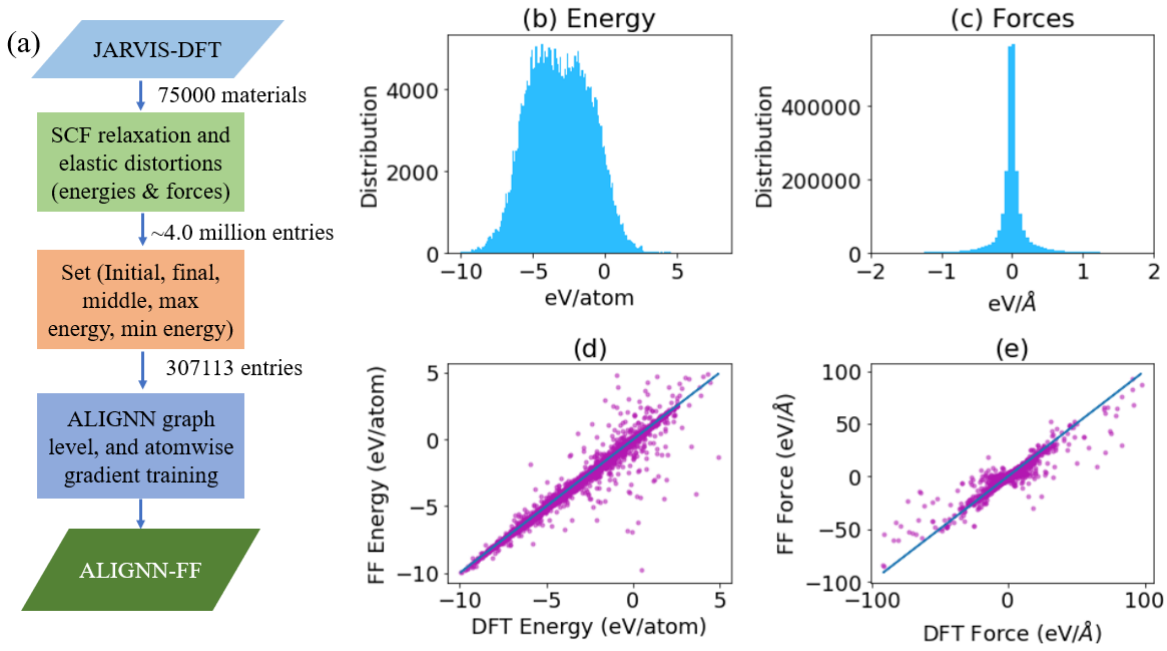

Figure 1: Schematic showing flow chart for developing the force-field, visualization of available data and model performance on test dataset. The left panel shows how dataset was obtained from JARVIS-DFT and model was trained. Fig. b and c shows energy and force-distribution in the dataset. While energy dataset ranges from -9.98 to 9.97 eV and is well-dispersed over this range, the force dataset varies from -591.12 to 591.40 eV/Å and is highly localized around zero. The force-dataset contains the x,y,z force on each atom. Fig. d and Fig. e shows energy and force predictions on the test set.

We find that the energy per atom dataset (with 307113 entries) varies from -9.98 eV/atom to 9.97 eV/atom. The force dataset contains 9593385 entries (resulting from x,y,z forces on each atom) is mostly centered around zero, but ranges from -591.12 eV/Å to 591.40 eV/Å. The mean absolute deviation (MAD) for energies and forces are 1.80 eV and 0.10 eV/Å respectively. The distributions of the energy and force used in the training are shown in Fig. 1b and 1c, respectively. The total dataset was split in 90:5:5 train-validation-test sets. We show the performance on test set for energies and forces in Fig. 1d and 1e respectively. We note that these points represent a variety of chemistry and structures unlike usual MLFFs

which are usually focused on a specific chemistry. The mean absolute error (MAE) of the predicted energy per atom and force component per atom are 0.086 eV and 0.047 eV/Å respectively. In comparison to the previous ALIGNN model for energy only, which has an MAE of 0.03 eV/atom, 0.086 eV/atom might seem high, however, we note that the previous energy model was trained on relaxed structures only while the current model also captures several un-relaxed structures which can be much higher in energy scale. The MAD:MAE ratios for energies and forces are 20.93 and 2.12 which are high.

While the above results are for force tunable weighting factor of 10, we also train the models with other weigthing factors as shown in Table 1. We find that as we increase the weighting factors, the MAEs increase for energies increase but decrease for forces. As the MAD for forces is 0.1 eV/Å, we choose to work with the model obtained for the lowest MAE for forces and to analyze its applications in the rest of the paper. Nevertheless, we share model parameters for other weighting factors for those interested in analyzing its effect on property predictions.

Table 1: Effect of different weighting factors for energy and force predictions.

Weight	MAE-Energies (eV/atom)	MAE-Forces (eV/$\AA$)
0.1	0.034	0.092
0.5	0.044	0.089
1.0	0.051	0.088
5.0	0.082	0.054
10.0	0.086	0.047

Energy-volume curves

The energy-volume (E-V) curves are crucial to understand the behavior of an FF. We obtain EV-curves for a few test case materials by applying volumetric strains between ranges of -0.1 to 0.1 with an interval of 0.01. Specifically, we compare the energy volume curves for Ni_3Al (JVASP-14971), Al_2CoNi (JVASP-108163), CrFeCoNi (4 atoms cell with spacegroup number 216 and lattice parameter of 4.007 Å), NaCl (JVASP-23862), MgO (JVASP-116), and

$BaTiO_3$ (JVASP-8029) using EAM potential,[7,8] GPAW DFT,[54] and ALIGNN-FF in Fig. 2. The energy scales for these methods differ so we align them with respect to the corresponding minimum energies, for comparison. We used less EV-curve points for a structure in GPAW to save computational cost. We notice all the EV-curves are parabolic in nature and smooth, indicating a smooth potential energy surface.

We find that the EV curves from these methods coincide near the minimum for all systems but for Al_2CoNi and CrFeCoNi, the GPAW equilibrium volume is slightly smaller than for EAM and ALIGNN-FF, suggesting that the lattice constants for ALIGNN-FF and EAM might be overestimated compared to GPAW. Nevertheless, EAM and ALIGNN-FF data agree well. Comparing EAM and ALIGN-FF, it's important to remember that GNNs have no fundamental limitation to number of species they can model (i.e. high chemical diversity), and can in principle even extrapolate to species not contained in the training set, which is extremely powerful compared to conventional FFs like EAM.

Note that there are many other conventional FF repositories available (such as Interatomic Potential Repository[55] and JARVIS-FF[56]) which contain data for a variety of systems. It is beyond the scope of the current work to compare all of them with ALIGNN-FF, however, it would be an interesting effort for the future work.

While the above examples are for individual crystals, it is important to distinguish different polymorphs of a composition system for materials simulation (i.e. structural diversity). In Fig. 3, we analyze the energy-volume (EV) curve of four systems and their polymorphs using ALIGNN-FF. We choose such four example systems because representative of different stable structures. In general, however, the EV-curve can be computed for any arbitrary system and structure. In Fig. 3a, we show the EV-curve for 4 silicon materials (JARVIS-IDs: JVASP-1002, JVASP-91933, JVASP-25369, JVASP-25368) with diamond cubic correctly being the lowest in energy. Similarly, the EV-curve for naturally prevalent SiO_2 systems (JARVIS-IDs:JVASP-58349, JVASP-34674, JVASP-34656, JVASP-58394), binary alloy Ni_3Al (JARVIS-IDs:JVASP-14971, JVASP-99749, JVASP-11979) and vdW bonded material

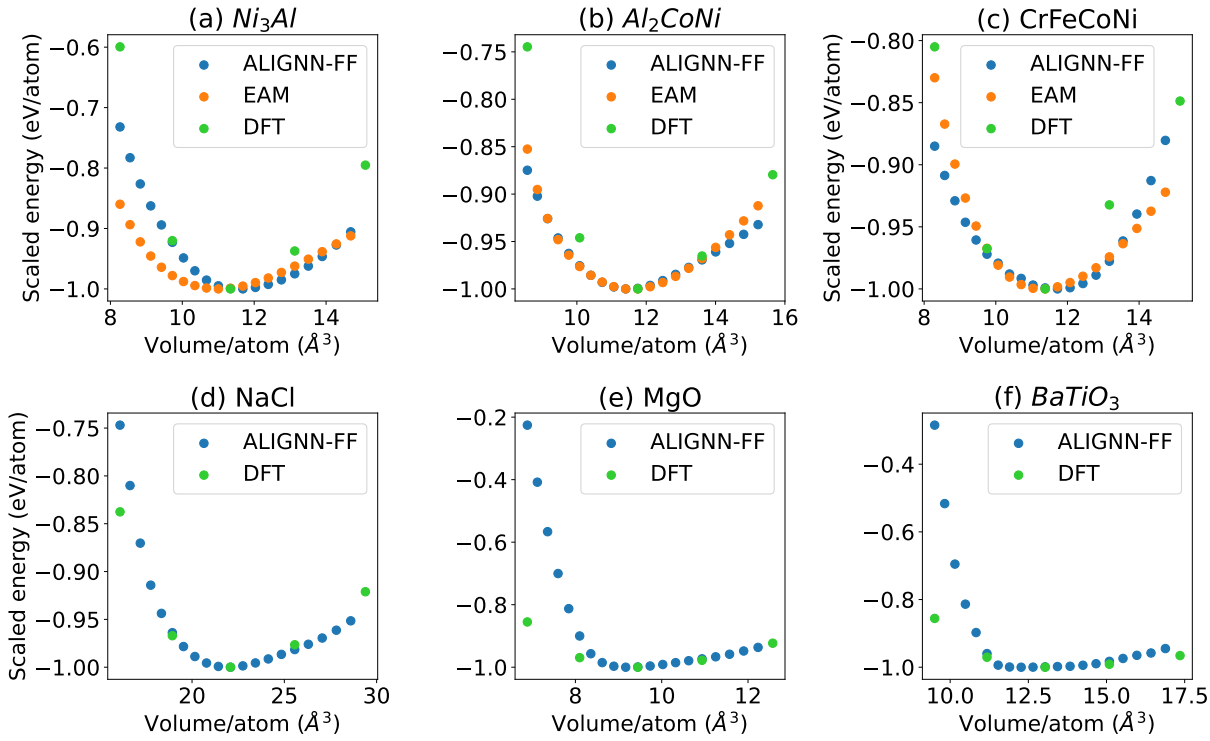

Figure 2: Energy-volume/expansion-contraction curves for a few example systems: a) Ni_3Al, b) Al_2CoNi, c) CrFeCoNi, d) NaCl, e) MgO, f) $BaTiO_3$ with ALIGNN-FF, EAM force-fields and VASP DFT methods. We apply volumetric strains between ranges of -0.1 to 0.1 with an interval of 0.01. For DFT (VASP) calculations, we use an interval of 0.05. Note that energy-volume curves were not explicitly included during the model training.

MoS_2 (JARVIS-IDs:JVASP-28733, JVASP-28413, JVASP-58505) all have the correct structure corresponding to the minimum energy. Therefore, while the MAE for our overall energy model is high, such model is able to distinguish polymorphs of compounds with meV level accuracy which is critical for atomistic applications.

Lattice constants and formation energies

In this section, we compare the lattice constants and formation energies for elemental and multi-component solids from the JARVIS-DFT after optimizing them using ALIGNN-FF as shown in Fig. 4a-4d. For ALIGNN-FF we choose solids with less than 10 atoms in a cell from the JARVIS-DFT database. We optimize 23495 such materials, and present the results in Fig

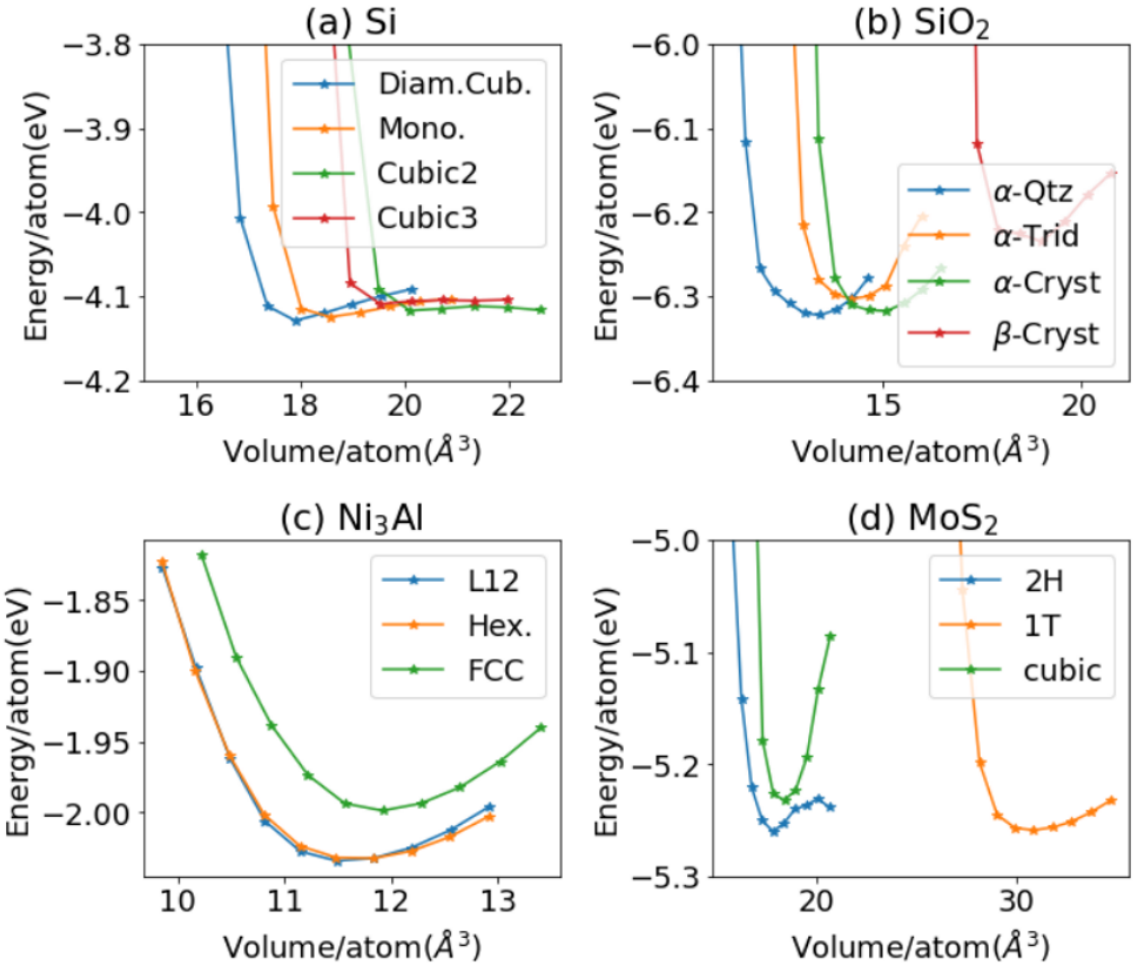

Figure 3: Energy-volume/expansion-contraction curves for a few example systems: a) silicon, b) SiO_2, c) Ni_3Al, d) MoS_2 polymorphs. We optimize the structures and then apply volumetric strains between ranges of -0.05 to 0.05 with an interval of 0.01.

4a-4d. The formation energies require chemical potential of elemental systems. The energy per atom of elemental solid systems with minimum energy are used as chemical potentials. We optimize the lattices with the FIRE algorithm as implemented in ASE and find reasonable agreement between the DFT and ALIGNN-FF lattice constants. The MAE for a,b,c lattice constants are 0.11, 0.11 and 0.13 Å respectively.The mean absolute error (MAE) for formation energies between ALIGNN-FF and JARVIS-DFT is 0.08 eV/atom which is reasonable for pre-screening applications. Note, the above validation is different from the performance measurement in Fig. 1 for the 5 % test dataset because we optimize the crystals rather than directly predicting the formation energies on unrelaxed structures. Similarly, we

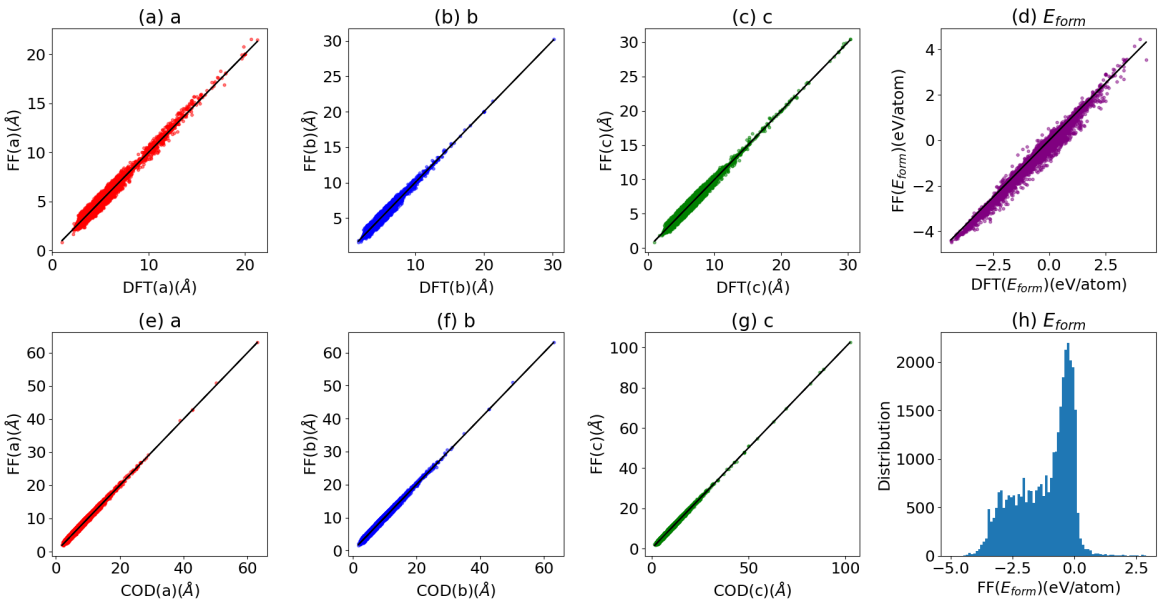

Figure 4: Comparison of DFT and FF data for a) lattice constant in x-direction, b) lattice-constant in y direction, c) lattice constant in z-direction, d) formation energies for stable binary solids in JARVIS-DFT. Comparison of crystallography open database (COD) and FF lattice constants in e) x-direction, f) y-direction, g) z-direction and h) formation energy distribution.

apply the ALIGNN-FF on the crystallography open database (COD). The COD database contains more than 431778 atomic structures with different types of chemical bondings and environments. We optimize 34615 structures in the COD database with number of atoms in a cell less than 50 and show the results in Fig. 4e-4h. Here, we find the MAE for a, b, c lattice parameter as 0.20, 0.20 and 0.23 Å respectively. Most of the systems in COD have been derived from experiments, and hence we see many of them have negative formation energies as shown in Fig. 4h.

Genetic algorithm based structure search

Computational prediction of the ground-state structures of a chemical system is a challenging task. Some of the common methods for this task include genetic algorithm (GA), simulated annealing and basin or minima hopping.[57] In the following examples, we show using GA together with ALIGNN-FF to search for crystal structures of Ni-Al and Cu-Al example

systems. Genetic algorithms mimic the biological evolution process to solve optimization problems. A GA optimization consists of 1) inheritance, 2) mutation, 3) selection, and 4) crossover operations to produce new structures and search for better survivors from generation to generation based on "survival of the fittest" idea which in our case would be energetics criteria. While all GAs follow similar strategy, the details of the individual operations can vary a lot from problem to problem and can be critical to search efficiency. We start with face-centered cubic (FCC) structures of the individual components (FCC Al (JVASP-816) and FCC Ni (JVASP-943) for Ni-Al search and FCC Al and FCC Cu (JVASP-867) for Cu-Al search) as the initial population. We perform relaxation of these systems with ALIGNN-FF. We choose a population size of 10 individuals and create the initial population by randomly selecting the elements. We use 40 generations to evolve the system and store the entries which are also relaxed with ALIGNN-FF.

After this example GA search for structures, we plot the convex hull diagram of these systems in Fig. 5. We find that the GA predicts AB and A_3B compounds which are in fact observed experimentally in such binary alloys.[51,58] Additionally, Ni_3Al (spacegroup: Pm-3m) is known to be one of the best performing super-alloys[51] which is reproduced in the above example. We also found that the formation energy of this structure (-0.47 eV/atom) is similar to Johannesson's[51] findings of -0.49 eV/atom.Although the above example is carried out for binary systems,in principle the same methodology can be applied for any other systems as well.

Timing study

Now, we compare the time for a single step potential energy calculation for FCC Aluminum (JVASP-816) with varied supercell sizes with EAM, ALIGNN-FF and GPAW DFT[54] methods. All of these calculations are performed on CPU on a personal laptop; GPU performance scaling may differ, particularly for large cell sizes. We start with a unit cell and start making supercells with [2,2,2], [3,3,3]. [4,4,4] and [5,5,5] dimensions. For GPAW case we use 8x8x8

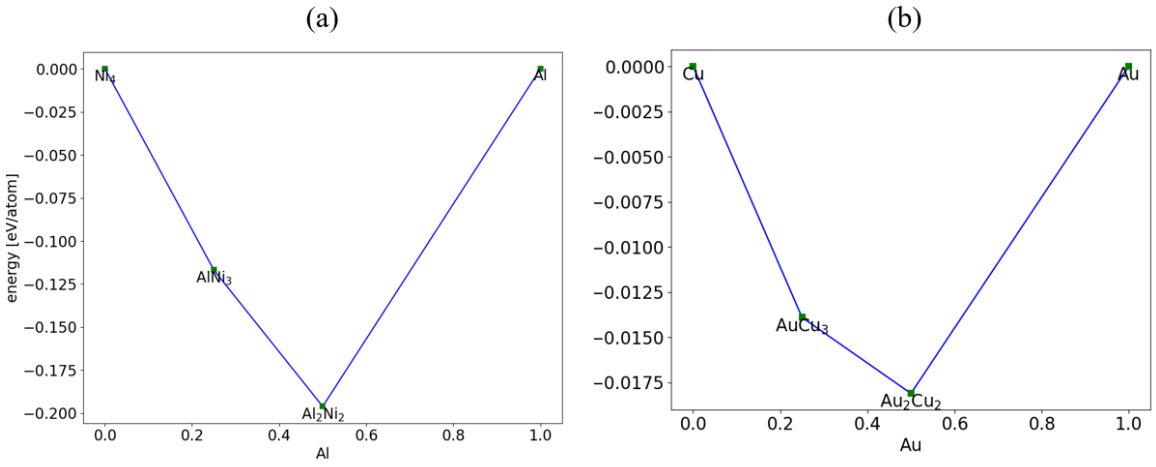

Figure 5: Convex hulls obtained from a genetic algorithm search for alloys starting with elemental solids only. The x-axis values represent mole fraction of corresponding element. We show the stable structures only for a) Ni-Al system, b) Cu-Au system.

k-points and as we make supercells, we reduce the k-points inversely proportional to the supercell size. In Fig. 7, we find that out of these three, EAM is the fastest method. The ALIGNN-FF is an order of magnitude slower than EAM, as expected. While EAM potentials are considerably faster, we note that they are difficult to train for multi-component systems. The GPAW single step calculations are almost 100 times slower than ALIGNN-FF. As we add multiple electronic and ionic relaxation steps in DFT, the computational cost drastically increases. Note that we chose a very generalized DFT set up of 330 eV plane wave cutoff and PBE[59] exchange correlation functional for GPAW calculations. As we add more k-points and plane wave cut-off, the computational cost will increase. Moreover, this comparison is only for an elemental system and as the system complexity (such as number of elements, number of electrons, vacuum padding etc.) increases, we believe the ALIGNN-FF will get a significantly higher boost in speed. ALIGNN-FF does not explicitly suffer from DFT parameters such as plane-wave cutoff, K-point settings which can add much higher computational cost.

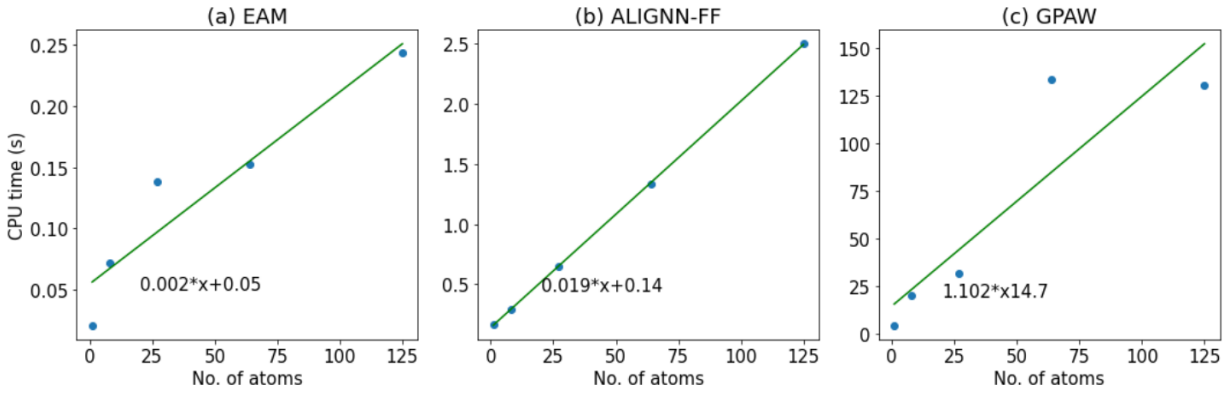

Figure 6: Timing comparison for FCC Al system for EAM, ALIGNN-FF and GPAW DFT methods.

Conclusion

In summary, we have developed a unified atomistic line graph neural network based force-field (ALIGNN-FF) that can model diverse set of materials with any combination of 89 elements from the periodic table. Using several test cases, we demonstrate the application of this method to determine the several properties such as lattice constants, formation energies, EV-curve for different materials including metallic, ionic and van der Waal bonded materials.

Although, the above examples are based on a few test cases, ALIGNN-FF can in principle be used for several applications such as investigating defective systems, high-entropy alloy, metal-organic framework, catalyst, battery designs etc. and its validity needs to be tested for other applications which is beyond the scope of the present work. Also, as the methods for including larger datasets improve (such as training on millions of data), and integration of active learning and transfer learning strategies are achieved, we believe we can train more accurate models. Moreover, such universal FF model can be integrated with universal tight-binding model[60] so that for classical and quantum properties can be predicted for large systems. In the current state, such FFs can be very useful for structure optimization, however, there is lot of room for improvement in terms of other physical characteristics such as defects, magnetism, charges, electronic levels etc. Under-parameterized potentials get drastically better energy and force accuracy, but are much narrower in scope and have

much higher training data density in their regions of applicability, based on highly tailored (and expensive) fit-for-purpose dataset generation. Overparameterized GNNs are currently trained on extremely sparse datasets that are repurposed from high throughput materials discovery efforts (e.g. JARVIS-DFT). Their potential to generalize to much greater chemical diversity is good, but a lot of research is needed to close the gap in accuracy with more narrow MLFF methods. A happy path forward for future research is probably intermediate in terms of the breadth of chemical and structural space and dataset density. Additionally, availability of standard benchmark datasets of multi-component systems with important properties such as diffusion coefficients for a diverse set of solute species, stacking fault energies, defect formation energies (solute substitution, vacancy, schottky/frenkel, etc), thermal conductivity, mechanical properties, and interface properties would play a pivotal role in the development of universal force-fields.

Integration with external coding interface such as Alloy Theoretic Automated Toolkit (ATAT),[61] CALculation of PHAse Diagrams (CALPHAD),[62] Universal Structure Predictor(USPEX),[63] Ab initio Random Structure Searching (AIRSS),[64] Genetic Algorithm for Structure and Phase Prediction (GASP),[65,66] RASPA[67] etc. would further extend the applications for alloy design, structure predictions and designing nanoporous materials in future. We note that the current force-field has been mainly evaluated for solids only, but in principle can be extended to polymers, molecular and hybrid systems as well in future. While there are several areas of improvements for such a unified force-field, we believe this work would spark the interest among materials scientists and engineers to enable a very wide range of atomistic applications.

Acknowledgement

K.C. thanks National Institute of Standards and Technology for funding and computational support. In addition, K.C thanks Ruth Pachter and Kiet Ngyuen from the Air Force Research

Laboratory for helpful discussions. This work was also partially supported by Wave 1 of the UKRI Strategic Priorities Fund under the EPSRC grant EP/T001569/1, particularly the 'AI for Science' theme within that grant, by the Alan Turing Institute and UKRI AIMLAC CDT, grant no. EP/S023992/1.

References

(1) Ogale, S. B. *Thin films and heterostructures for oxide electronics*; Springer Science & Business Media, 2006.

(2) Andersson, M. P.; Bligaard, T.; Kustov, A.; Larsen, K. E.; Greeley, J.; Johannessen, T.; Christensen, C. H.; Nørskov, J. K. Toward computational screening in heterogeneous catalysis: Pareto-optimal methanation catalysts. *Journal of Catalysis* **2006**, *239*, 501–506.

(3) Liang, T.; Shan, T.-R.; Cheng, Y.-T.; Devine, B. D.; Noordhoek, M.; Li, Y.; Lu, Z.; Phillpot, S. R.; Sinnott, S. B. Classical atomistic simulations of surfaces and heterogeneous interfaces with the charge-optimized many body (COMB) potentials. *Materials Science and Engineering: R: Reports* **2013**, *74*, 255–279.

(4) Li, X.; Zhu, Y.; Cai, W.; Borysiak, M.; Han, B.; Chen, D.; Piner, R. D.; Colombo, L.; Ruoff, R. S. Transfer of large-area graphene films for high-performance transparent conductive electrodes. *Nano letters* **2009**, *9*, 4359–4363.

(5) Srolovitz, D. J.; Vitek, V. *Atomistic Simulation of Materials: Beyond Pair Potentials*; Springer Science & Business Media, 2012.

(6) Daw, M. S.; Baskes, M. I. Semiempirical, quantum mechanical calculation of hydrogen embrittlement in metals. *Physical review letters* **1983**, *50*, 1285.

(7) Pun, G. P.; Yamakov, V.; Mishin, Y. Interatomic potential for the ternary Ni–Al–Co

system and application to atomistic modeling of the B2–L10 martensitic transformation. *Modelling and Simulation in Materials Science and Engineering* **2015**, *23*, 065006.

(8) Farkas, D.; Caro, A. Model interatomic potentials for Fe–Ni–Cr–Co–Al high-entropy alloys. *Journal of Materials Research* **2020**, *35*, 3031–3040.

(9) Daw, M. S.; Baskes, M. I. Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals. *Physical Review B* **1984**, *29*, 6443.

(10) Liang, T.; Devine, B.; Phillpot, S. R.; Sinnott, S. B. Variable charge reactive potential for hydrocarbons to simulate organic-copper interactions. *The Journal of Physical Chemistry A* **2012**, *116*, 7976–7991.

(11) Brenner, D. W. Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films. *Physical review B* **1990**, *42*, 9458.

(12) Van Duin, A. C.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: a reactive force field for hydrocarbons. *The Journal of Physical Chemistry A* **2001**, *105*, 9396–9409.

(13) Brenner, D. W. The art and science of an analytic potential. *physica status solidi (b)* **2000**, *217*, 23–40.

(14) Poltavsky, I.; Tkatchenko, A. Machine learning force fields: Recent advances and remaining challenges. *The Journal of Physical Chemistry Letters* **2021**, *12*, 6551–6564.

(15) Choudhary, K.; DeCost, B.; Chen, C.; Jain, A.; Tavazza, F.; Cohn, R.; Park, C. W.; Choudhary, A.; Agrawal, A.; Billinge, S. J., et al. Recent advances and applications of deep learning methods in materials science. *npj Computational Materials* **2022**, *8*, 1–26.

(16) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical review letters* **2007**, *98*, 146401.

(17) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müllerr, K.-R. Machine learning force fields. *Chemical Reviews* **2021**, *121*, 10142–10186.

(18) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Physical review letters* **2010**, *104*, 136403.

(19) Wood, M. A.; Thompson, A. P. Extending the accuracy of the SNAP interatomic potential form. *The Journal of chemical physics* **2018**, *148*, 241721.

(20) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Physical Review B* **2013**, *87*, 184115.

(21) Shapeev, A. V. Moment tensor potentials: A class of systematically improvable interatomic potentials. *Multiscale Modeling & Simulation* **2016**, *14*, 1153–1173.

(22) Novikov, I. S.; Gubaev, K.; Podryabinkin, E. V.; Shapeev, A. V. The MLIP package: moment tensor potentials with MPI and active learning. *Machine Learning: Science and Technology* **2020**, *2*, 025002.

(23) Hernandez, A.; Balasubramanian, A.; Yuan, F.; Mason, S. A.; Mueller, T. Fast, accurate, and transferable many-body interatomic potentials by symbolic regression. *npj Computational Materials* **2019**, *5*, 1–11.

(24) Drautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. *Physical Review B* **2019**, *99*, 014104.

(25) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. Schnet–a deep learning architecture for molecules and materials. *The Journal of Chemical Physics* **2018**, *148*, 241722.

(26) Xie, T.; Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters* **2018**, *120*, 145301.

(27) Chen, C.; Ye, W.; Zuo, Y.; Zheng, C.; Ong, S. P. Graph networks as a universal machine learning framework for molecules and crystals. *Chemistry of Materials* **2019**, *31*, 3564–3572.

(28) Chen, C.; Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *arXiv preprint arXiv:2202.02450* **2022**,

(29) Kearnes, S.; McCloskey, K.; Berndl, M.; Pande, V.; Riley, P. Molecular graph convolutions: moving beyond fingerprints. *Journal of computer-aided molecular design* **2016**, *30*, 595–608.

(30) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural message passing for quantum chemistry. International conference on machine learning. 2017; pp 1263–1272.

(31) Klicpera, J.; Giri, S.; Margraf, J. T.; Günnemann, S. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. *arXiv preprint arXiv:2011.14115* **2020**,

(32) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications* **2022**, *13*, 1–11.

(33) Musaelian, A.; Batzner, S.; Johansson, A.; Sun, L.; Owen, C. J.; Kornbluth, M.; Kozinsky, B. Learning Local Equivariant Representations for Large-Scale Atomistic Dynamics. *arXiv preprint arXiv:2204.05249* **2022**,

(34) Park, C. W.; Kornbluth, M.; Vandermause, J.; Wolverton, C.; Kozinsky, B.; Mailoa, J. P. Accurate and scalable graph neural network force field and molecular dynamics with direct force architecture. *npj Computational Materials* **2021**, *7*, 1–9.

(35) Chmiela, S.; Sauceda, H. E.; Müller, K.-R.; Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nature communications* **2018**, *9*, 1–10.

(36) Choudhary, K.; DeCost, B. Atomistic Line Graph Neural Network for improved materials property predictions. *npj Computational Materials* **2021**, *7*, 1–8.

(37) Choudhary, K.; Yildirim, T.; Siderius, D. W.; Kusne, A. G.; McDannald, A.; Ortiz-Montalvo, D. L. Graph neural network predictions of metal organic framework CO2 adsorption properties. *Computational Materials Science* **2022**, *210*, 111388.

(38) Choudhary, K.; Garrity, K. Designing High-Tc Superconductors with BCS-inspired Screening, Density Functional Theory and Deep-learning. *arXiv preprint arXiv:2205.00060* **2022**,

(39) Choudhary, K.; Sumpter, B. G. A Deep-learning Model for Fast Prediction of Vacancy Formation in Diverse Materials. *arXiv preprint arXiv:2205.08366* **2022**,

(40) Kaundinya, P. R.; Choudhary, K.; Kalidindi, S. R. Prediction of the electron density of states for crystalline compounds with Atomistic Line Graph Neural Networks (ALIGNN). *JOM* **2022**, *74*, 1395–1405.

(41) Gurunathan, R.; Choudhary, K.; Tavazza, F. Rapid Prediction of Phonon Structure and Properties using the Atomistic Line Graph Neural Network (ALIGNN). *arXiv preprint arXiv:2207.12510* **2022**,

(42) Choudhary, K.; Garrity, K. F.; Reid, A. C.; DeCost, B.; Biacchi, A. J.; Hight Walker, A. R.; Trautt, Z.; Hattrick-Simpers, J.; Kusne, A. G.; Centrone, A.,

et al. The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design. *npj Computational Materials* **2020**, *6*, 1–13.

(43) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical review B* **1996**, *54*, 11169.

(44) Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational materials science* **1996**, *6*, 15–50.

(45) Klimeš, J.; Bowler, D. R.; Michaelides, A. Chemical accuracy for the van der Waals density functional. *Journal of Physics: Condensed Matter* **2009**, *22*, 022201.

(46) Choudhary, K.; Cheon, G.; Reed, E.; Tavazza, F. Elastic properties of bulk and low-dimensional materials using van der Waals density functional. *Physical Review B* **2018**, *98*, 014107.

(47) Subramaniyan, A. K.; Sun, C. Continuum interpretation of virial stress in molecular simulations. *International Journal of Solids and Structures* **2008**, *45*, 4340–4346.

(48) Wang, M.; Zheng, D.; Ye, Z.; Gan, Q.; Li, M.; Song, X.; Zhou, J.; Ma, C.; Yu, L.; Gai, Y., et al. Deep graph library: A graph-centric, highly-performant package for graph neural networks. *arXiv preprint arXiv:1909.01315* **2019**,

(49) Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. Automatic differentiation in pytorch. **2017**,

(50) Larsen, A. H.; Mortensen, J. J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C., et al. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **2017**, *29*, 273002.

(51) Johannesson, G. H.; Bligaard, T.; Ruban, A. V.; Skriver, H. L.; Jacobsen, K. W.; Nørskov, J. K. Combined electronic structure and evolutionary search approach to materials design. *Physical Review Letters* **2002**, *88*, 255506.

(52) Bitzek, E.; Koskinen, P.; Gähler, F.; Moseler, M.; Gumbsch, P. Structural relaxation made simple. *Physical review letters* **2006**, *97*, 170201.

(53) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D., et al. LAMMPS-a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications* **2022**, *271*, 108171.

(54) Enkovaara, J.; Rostgaard, C.; Mortensen, J. J.; Chen, J.; Dułak, M.; Ferrighi, L.; Gavnholt, J.; Glinsvad, C.; Haikola, V.; Hansen, H., et al. Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *Journal of physics: Condensed matter* **2010**, *22*, 253202.

(55) 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. *Current Opinion in Solid State and Materials Science* **2013**, *17*, 277–283.

(56) Choudhary, K.; Congo, F. Y. P.; Liang, T.; Becker, C.; Hennig, R. G.; Tavazza, F. Evaluation and comparison of classical interatomic potentials through a user-friendly interactive web-interface. *Scientific data* **2017**, *4*, 1–12.

(57) Ji, M.; Wang, C.-Z.; Ho, K.-M. Comparing efficiencies of genetic and minima hopping algorithms for crystal structure prediction. *Physical Chemistry Chemical Physics* **2010**, *12*, 11617–11623.

(58) van de Walle, A.; Ceder, G. Automating first-principles phase diagram calculations. *Journal of Phase Equilibria* **2002**, *23*, 348–359.

(59) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Physical review letters* **1996**, *77*, 3865.

(60) Garrity, K. F.; Choudhary, K. Fast and Accurate Prediction of Material Properties with Three-Body Tight-Binding Model for the Periodic Table. *arXiv preprint arXiv:2112.11585* **2021**,

(61) Van De Walle, A.; Asta, M.; Ceder, G. The alloy theoretic automated toolkit: A user guide. *Calphad* **2002**, *26*, 539–553.

(62) Liu, Z.-K. First-principles calculations and CALPHAD modeling of thermodynamics. *Journal of phase equilibria and diffusion* **2009**, *30*, 517–534.

(63) Glass, C. W.; Oganov, A. R.; Hansen, N. USPEX—Evolutionary crystal structure prediction. *Computer physics communications* **2006**, *175*, 713–720.

(64) Pickard, C. J.; Needs, R. High-pressure phases of silane. *Physical review letters* **2006**, *97*, 045504.

(65) Revard, B. C.; Tipton, W. W.; Hennig, R. G. Structure and stability prediction of compounds with evolutionary algorithms. *Prediction and Calculation of Crystal Structures* **2014**, 181–222.

(66) Choudhary, K.; Liang, T.; Mathew, K.; Revard, B.; Chernatynskiy, A.; Phillpot, S. R.; Hennig, R. G.; Sinnott, S. B. Dynamical properties of AlN nanostructures and heterogeneous interfaces predicted using COMB potentials. *Computational Materials Science* **2016**, *113*, 80–87.

(67) Dubbeldam, D.; Calero, S.; Ellis, D. E.; Snurr, R. Q. RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Molecular Simulation* **2016**, *42*, 81–101.