



Contents lists available at ScienceDirect

Journal of Materiomics

journal homepage: www.journals.elsevier.com/journal-of-materiomics/

Towards universal neural network interatomic potential

So Takamoto^a, Daisuke Okanohara^a, Qing-Jie Li^b, Ju Li^{b,*}^a Preferred Networks, Inc., 100-0004, 1-6-1 Otemachi, Chiyoda-ku, Tokyo, Japan^b Department of Nuclear Science and Engineering and Department of Materials Science and Engineering, MIT, Cambridge, MA, 02139, USA

ARTICLE INFO

Article history:

Received 23 October 2022

Received in revised form

27 December 2022

Accepted 31 December 2022

Available online 20 January 2023

The universal gravitational force law $f = -Gm_1m_2/r^2$ enabled Newton to predict the trajectory of celestial bodies in 1686, propelling human civilization into the modern age. Atoms and ions underlie chemistry, biology and materials science. Knowing their interaction potential $U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$, where $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$ are the positions and types of atoms/ions, would also allow us to predict their trajectories ($\mathbf{f}_i = -\nabla_i U$) and collective behavior. In 1871, Dmitri Mendeleev organized the periodic table. There are 92 natural elements from Hydrogen to Uranium. If we can have easy numerical access to $U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ for arbitrary sets of chemical elements, the so-called universal interatomic potential, it would greatly expedite the understanding of our natural world through simulations, and assist the design and synthesis of drug molecules, new materials for the transition to sustainable energy, etc.

Today such universal potentials exist, although they are computationally quite expensive and practically accessible for only a few hundred atoms. This is because we need to solve the quantum mechanical Schrödinger equation for the thousands of electrons for given nuclei positions to get $U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$. The 1998 Nobel Prize in Chemistry was awarded to John Pople “for his development of computational methods in quantum chemistry,” and to Walter Kohn “for his development of the density-functional theory” (DFT). Essentially, Pople and collaborators solved the many-body problem of quantum electrons directly, while Kohn and collaborators simplified this many-body problem to many effective single-body problems. Because one is solving many coupled integro-differential equations on the fly, the memory and computational complexities of these first-generation (full configuration

interaction, quantum Monte Carlo, coupled cluster, etc.) and second-generation (DFT) universal potentials scale as $O(N^3)$ or more. Currently neither “Pople’s universal potential” nor “Kohn’s universal potential” can be applied routinely to more than a few hundred atoms without undue loss of accuracy, even with the amazing computational resources we have, because of the exorbitant complexity scaling. If we run *ab initio* molecular dynamics (MD) simulations with these universal potentials, the simulated time duration is typically limited to 10^2 ps.

How about the accuracy of universal potentials? 1 kcal/mol, or 43 meV/atom, is termed “chemical accuracy” because this is the level of measurement error in experimental thermochemistry, *i.e.* measuring and inferring the heat of formation/binding energy of molecules and solids. “Pople’s universal potential”, when fully converged, can beat the chemical accuracy, but is practically only feasible for at most tens of atoms. “Kohn’s universal potential”, or DFT, broadly gives 2–3 kcal/mol (~100 meV/atom) error due to the approximate density functional used, but it can deal with hundreds of atoms. Despite these constraints, both have achieved resounding success in the last few decades, principally because of their abilities to *treat chemical complexities* robustly. That is, we can throw arbitrary types of atoms from Hydrogen to Uranium into the simulation box, and expect reasonable chemical behaviors from the simulations, including chemical reactions that break and form bonds. The most basic and important requirement on any universal interatomic potential is to reproduce *qualitative* features of “high-school chemistry” robustly. That is, when one throws Na metal into liquid water, one should get H₂ gas. When one throws NaCl crystal into liquid water, one should get hydrated and well-separated Na⁺(aq) and Cl⁻(aq) ions. When liquid water freezes, the ice should be less dense than the liquid water, etc..

Interatomic potentials bypassing on-the-fly electronic-structure calculations, the so-called empirical potentials, are developed to reduce the aforementioned computational complexity barrier. These calculation methods observe and approximate the electronic interactions with off-line fitting. For example, the Lennard-Jones potential models the van der Waals dispersion interaction with r^{-6} type fits. Embedded atom method (EAM) potential has local electron density as an atomic expression that describes interaction in metallic systems. Bond-order potentials such as the Tersoff-Brenner potentials model the tight-binding electronic bandwidth theory and can describe the chemical bonding effect including the

* Corresponding author.

E-mail address: liju@mit.edu (J. Li).

Peer review under responsibility of The Chinese Ceramic Society.

change of the number of bonds and bond angles. However, these empirical potentials typically fail the “general high-school chemistry test”. Because of the limited representation power of pre-conceived fitting forms and limited variety of *ab initio* or experimental data as fitting targets, previous empirical potentials are severely constrained by the *number of atomic types* allowed. Most empirical potentials used today are still monatomic (like pure Cu, Si) or binary (like Si-C, Ni-Al) potentials. One can for example browse the list of all multi-element empirical potentials collected by National Institute of Standards and Technology [1] in 2022, and see only ten empirical potentials with more than 4 elements, the largest set being Br-Cl-Cs-F-I-K-Li-Na-Rb. This is far from enough to cover high-school chemistry.

With the ascendance of machine learning (ML) approaches including neural network (NN) models, representation power of empirical potentials can be greatly enhanced [2–4]. Various high-dimensional kernels, corresponding to the descriptors of the local environment of atoms, were proposed. Gaussian approximation potential (GAP), moment tensor potentials (MTP), linear and quadratic Spectral Neighbor Analysis Potential (SNAP/qSNAP), *etc.*, achieves excellent balance between computational speed and accuracy (see Fig. 1). One technique with a large user base in ML is the architecture of graph neural networks (GNN). It models the complex interactions of elements as the stack of message passing and has the power to represent the interaction between atoms which are not directly bonded, *i.e.*, aromatic bonds or metals. With the rapid growth of deep neural networks (DNN) in recent years, GNN-based methods quickly surpassed the chemical accuracy of well-known energy estimation benchmarks, and it is not uncommon for them to be in the few meV/atom range nowadays (Fig. 1) [2–5] that is much better than the chemical accuracy. But so far this has mostly resulted in more accurate interatomic potentials [2–5], and less in the number of supported chemical elements. It has become widely accepted to develop tailor-made empirical potentials for each specific material system being simulated. However, the benchmarks they choose might be too simple in terms of chemical complexity and the number of elements supported. A universal empirical potential broadly applicable to *any* material in the world was simply not considered to be an achievable goal before. Two advances are needed for the breakthrough: the architecture and the dataset.

In 2019 Takamoto, Izumi and Li published *TeaNet* (Tensor Embedded Atom Network) which is a GNN potential satisfying special Euclidean group $E(3)$ equivariance (translation, rotation,

and reflection) implemented by both atom- and bond-centered tensorial attributes that include scalar, vector (3×1) and rank-2 tensor (3×3) quantities with full operations (dyad/contraction). GNN has rich expressive power and achieves translational equivariance, but previously was employed to transport scalars and vectors, not rank-2 tensors, in the context of atomistic simulations. As classic interatomic potentials were inspired by tight-binding electronic relaxation framework where the electronic Hamiltonian are rotationally equivariant tensors, we wanted to represent such iterative propagation of rank-2 tensor information by GNN. In *TeaNet*, the angular interaction is translated into graph convolution through the incorporation of Euclidean tensors, vectors and scalars. Previous empirical potentials, such as three-body angular dependent bond order potential, can be naturally mapped to shallow graph convolutions. *TeaNet* uses a much deeper (16 layers) GNN to simulate an iterative electronic relaxation, *i.e.* 16 steps, enabled by the residual network (ResNet) architecture and training with recurrent GNN weights initialization. This architecture design provides an effective inductive bias to learn the multi-element potential energy surface. The training dataset was generated by DFT single-point energy calculation of chemically and structurally randomized configurations. We demonstrate that arbitrary structures and reactions involving the *first 18 elements* on the periodic table (H to Ar), including hydrocarbon molecules, metals, amorphous SiO_2 and water, show surprisingly good performance and robustness, including chemical reaction pathways involving elements from H to Ar [6].

In early 2022, the collaborative work of Preferred Networks, Inc. (PFN) and ENEOS Corporation (ENEOS) extended the *TeaNet* architecture and constructed a large dataset with $\sim 10^7$ DFT configurations, that aims to handle arbitrary combinations of the entire Periodic Table. They have created the dataset by combining multiple sampling methods in order to cover the diverse chemical phase space as much as possible. For example, it contains crystal lattice structures with arbitrary combinations of elements and multi-element structures melted at high temperatures. They also provided a part of the dataset as the benchmark (HME21) and revealed that the *TeaNet* architecture performs very well for such a chemically diverse dataset, with an average error of only 19.6 meV/atom, which handily beats the chemical accuracy (43 meV/atom or 1 kcal/mol), namely the typical *experimental* error level involved in thermochemistry measurements. They released the first *empirical universal* interatomic potential covering 45 elements on the periodic table [7]. The series of trained models are named PFP

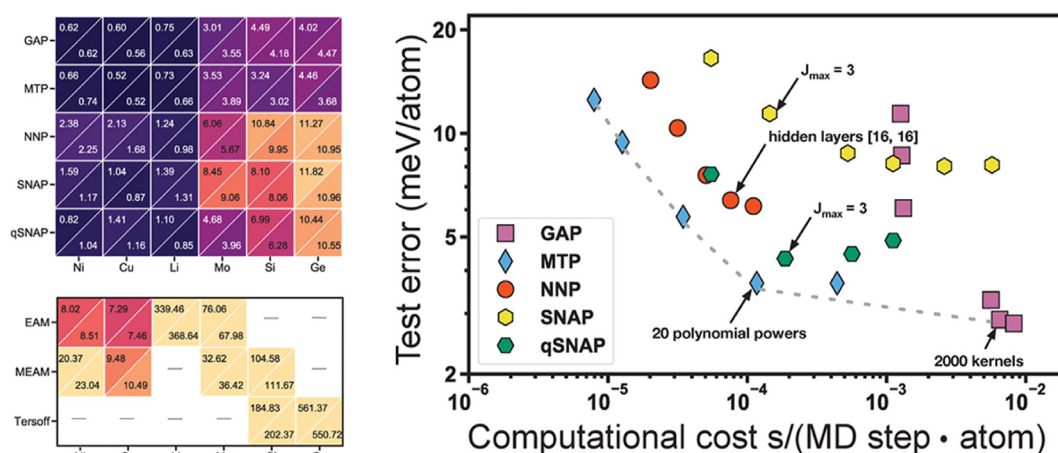


Fig. 1. Reprinted with permission from [5]. Copyright 2020 American Chemical Society. It shows the relationship of various types of interatomic potentials from the view of computational cost and accuracy. Note that this measures the performance in the quite limited task in terms of the number of elements and chemical complexity.

(PreFerred Potential). The number of elements was expanded to 72 as of the summer of 2022 (see Fig. 2). It was used as the core technology for the atomistic simulation software named Matlantis™ [8]. The number of elements supported is expected to rise from 72 as of now, to 94 (Pu), covering rare-earth elements and actinides, sometime in 2023.

To demonstrate the power of a universal potential (Fig. 3), four examples with very different application contexts were shown [7]. These examples were purposely chosen from out-of-domain applications, *i.e.*, all the chemical and structural configurations encountered were not part of the training dataset (and in fact very different from the training data), such that the results can faithfully indicate the extrapolation capability of PFP. The first example application is lithium diffusion in LiFeSO₄F cathode of lithium-ion batteries. To obtain the path and activation energy of the diffusion, not only the stable structure but also the unstable transition state should be reproduced. The nudged elastic band (NEB) method was applied for diffusion paths in three different crystal directions and the obtained activation energies reproduced the DFT predictions well, with an error on the order of 10⁻² eV or even less. A series of finite-temperature molecular dynamics simulations on Li diffusion were carried out and the Arrhenius plot showed similar activation energy. It indicates the energy surface of the entire phase space contained within a certain temperature range is smooth and there is no spuriously stable structure. Accurately predicting transition paths/states in chemically complex systems is an indispensable yet demanding task for practical applications, as kinetics information is often needed to identify/improve the rate-limiting steps of the underlying chemical/physical processes. TeaNet/PFP demonstrates great capability in such an application context in terms of both accuracy and efficiency (the above-mentioned NEB calculation can be completed within a few minutes).

The second example is the molecular adsorption in metal-organic frameworks (MOFs). They consist of metal centers bridged by organic linkers, thus containing both organic and inorganic parts as well as nanopores/surfaces. Their crystalline structures are diversified by the inclusion of various chemical groups and their ionic forms, and thus it was difficult to reproduce them using classical potentials because the interaction between many elements, including molecular complexes (ligands) and ions, should be reproduced in single potential. PFP was applied to a wide range of MOFs created from the Cambridge Structure Database and confirmed that both crystalline structure and the H₂O binding energy of MOFs are well reproduced, among different metal cation

centers (Cu, Zn, Al, Mg, Co, Ni). For example, the predicted cell volumes differ by only a few percent (in terms of mean absolute error) from experimental results and the H₂O binding energies only show a few percent error compared to reported values in the literature. It should be noted that, for some systems, adding DFT-D3 dispersion correction can improve the accuracy. This is because the standard DFT method itself can fall short for certain chemically complex systems; dispersion correction, Hubbard U correction, and even advanced exchange-correlation functionals are required for improvements. PFP provides different calculation modes to account for these corrections, which would benefit practical applications on chemically complex systems where certain correction is necessary.

The third example is the chemical order-disorder transition of Cu-Au alloys. This metallic system is a well-studied catalyst for the oxidation of CO and alcohols, and it is fully miscible over a wide composition range, but exhibiting a temperature-dependent chemical order-disorder transition. To reproduce this kind of alloy phase diagrams, the universal potential should have enough accuracy to resolve minute free-energy differences. The results show that the order-disorder transition temperatures of CuAu₃, CuAu, Cu₃Au are consistent with the experimental results, in the range of 300–400 K, 800–900 K, and 600–700 K, respectively. Many practical applications require the understanding of materials phase behaviors which are dictated by the underlying Gibbs free energy surface. Free energy evaluation generally involves both enthalpic and entropic (such as vibrational and configurational) terms that require a neural network interatomic potential (NNIP) to not only accurately predict the zero-K potential energy surface, but also the configurational evolution at finite temperatures. In addition, NNIP-based atomistic simulations can also provide useful input information for higher-level thermodynamic modeling and typical coarse-grained modeling. In such a context, TeaNet/PFP would benefit more practical applications with reliable phase behavior modeling, especially on chemically complex systems.

The fourth example is Fischer-Tropsch catalysis, which is the synthesis of hydrocarbons from H₂ and CO. It is selected as the typical example of heterogeneous system and catalytic reaction, which means the potential should reproduce not only the energy curvature of minimum energy path including transition state, but also its dependence on the surface environment. The results show that the mean absolute error of the activation energy is less than 0.1 eV among 20 elementary reactions involving hydrocarbon molecules on metal surfaces (Ag, Ce, K, Li, Mg, Mn, Na, Pt, Ru, V and Zn) as compared to the DFT calculations. In addition, the surface

1																	2															
H																	He															
3	Li	4												10																		
	Be												B	5	6	7	8	9	10													
11	Na	12												13	14	15	16	17	18													
	Mg												Al	Si	P	S	Cl	Ar														
19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36															
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr															
37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54															
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe															
55	56												61	62	63	64	65	66														
Cs	Ba												Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn					
87	88												104	105	106	107	108	109	110	111	112	113	114	115	116	117	118					
Fr	Ra												Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg	Cn	Nh	Fl	Mc	Lv	Ts	Og					
																		57	58	59	60	61	62	63	64	65	66	67	68	69	70	71
																		La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
																		89	90	91	92	93	94	95	96	97	98	99	100	101	102	103
																		Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr

Fig. 2. Chemical elements applicable in PFP universal potential as of the summer of 2022. Seventy-two elements highlighted as blue are fully supported in version 3.0.0. Tc and Tl are being experimentally supported.

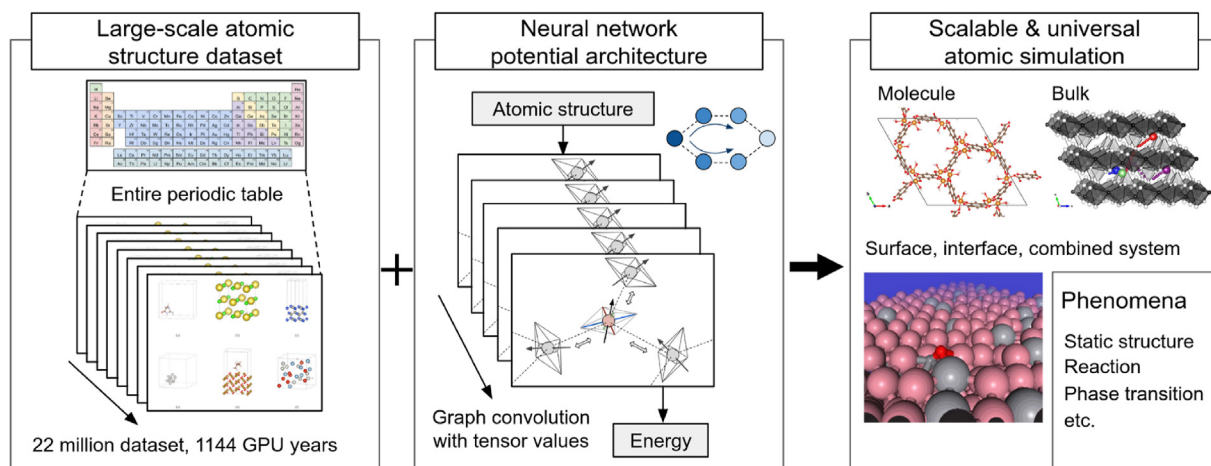


Fig. 3. Schematic illustration of the key technical ingredients to achieve PFP universal interatomic potential. For the dataset side, a large-scale dataset was constructed which does not limit its target domain and covers the entire phase space as much as possible. For the architecture side, a graph convolution neural network which treats higher-order tensors was developed. The single trained model can reproduce various material properties including molecule, bulk, and combined system. Figures are taken from reference [7], CC-BY 4.0 (<http://creativecommons.org/licenses/by/4.0/>).

promoter element exploration for the critical reaction was carried out. The result indicates that Vanadium reduces the activation barrier the most, and later it was confirmed that it is consistent with the experimental results. This is an example of the use case of PFP for materials discovery. Accurate yet efficient materials properties evaluation has been long pursued for materials discovery and optimization. Integrating together TeaNet/PFP's capability of predicting chemical reactions, phase behaviors, and kinetics with proper autonomous search algorithms such as Bayesian Optimization would open more opportunities for rational materials design.

The four problems above, each quite complex, would traditionally require a separate empirical potential and individualized fitting. But with PFP, all four problems can be simulated with the same potential, illustrating the strong transferability. The error versus the DFT is generally significantly lower than the error between DFT and experiments. More importantly, its computational cost is much smaller than conventional DFTs, with order- N complexity scaling. For example, PFP is more than a million times faster than DFT in a 3000-Pt-atoms calculation, so a calculation that takes PFP an hour to finish would take more than a century with DFT. Thus, one can perform DFT-accuracy simulations, but with a much larger number of atoms, for much longer physical times. The detailed information of the simulations and the related data are available in Ref. [7]. More application examples are forthcoming, for example in heterogeneous catalysis with strong metal-support interactions [9] or ionic transport under an electric field bias [10].

What makes TeaNet/PFP robust in describing universal interatomic interactions? The fundamental reason for such robustness is the effectiveness in transforming raw input data (atomic numbers $\{Z_n\}$ which are scalars, and positions $\{\mathbf{r}_n\}$ which are vectors) into *better* representations of the local electronic environments centered around “atoms” and “bonds” (Fig. 4). These atom-centered and bond-centered quantities are represented by tensors. Scalars and vectors are rank-0 and rank-1 tensors, respectively, but there are other higher-order tensors, for example the electric quadrupole moment (belonging to say, a pocket of electronic charge, well defined on a covalent bond center) would need to be represented by a rank-2 tensor T_{ij} , where subscript indices i, j takes integer values from 1 to 3 corresponding to the input space coordinate (xyz). Recall that the definition of a rank- M tensor $T_{i_1 i_2 \dots i_M}$ is not just this numerical object has M subscript indices, but also that it must satisfy tensor transformation law upon any observational-frame

rotation:

$$\tilde{T}_{\tilde{i}_1 \tilde{i}_2 \dots \tilde{i}_M} = T_{i_1 i_2 \dots i_M} R_{i_1 \tilde{i}_1} R_{i_2 \tilde{i}_2} \dots R_{i_M \tilde{i}_M} \quad (1)$$

where $R_{i\tilde{i}}$ is the frame-rotation matrix, and $\tilde{T}_{\tilde{i}_1 \tilde{i}_2 \dots \tilde{i}_M}$ is the read-out of the *same physical object* in the rotated observation frame. Physically meaningful quantities must be represented by tensors, since one can choose to measure a physically meaningful quantity in any arbitrary observation frame. In (1) we have already used the Einstein index-contraction convention, where repeated indices (indices not surviving on the left-hand side) are automatically summed over from 1 to 3. There are also well-defined tensor outer products and index contraction operations which give tensorial outcomes. Thus, the generalized dyad operations

$$T_{ijkl} \equiv U_{ij} V_{kl} \quad (2)$$

or arbitrary index contractions

$$v_i \equiv U_{ijk} V_{jk} \quad (3)$$

are automatically guaranteed to give tensor outputs. Just following the standard tensor inner products and outer products rules automatically guarantee the so-called Euclidean space equivariance in the final output, scalar, vector or otherwise. Translational and permutational invariances are automatically built in by using only the difference vector $\mathbf{r}_m - \mathbf{r}_n$ in Cartesian frame and $\{Z_n\}$ scalars. So, the $E(3)$ equivariant network is really just as simple/trivial as following the Einstein index-contraction rule between all tensorial quantities in the graph convolutional neural network.

The goal of TeaNet/PFP is to emulate iterative electronic relaxation (and possibly other physically motivated phenomena), where each layer can be thought of as one “timestep” in the relaxation step. In the construction of the electronic Hamiltonian with local orbital basis, or the vibrational dynamical matrix, or in dealing with the phenomena of elastic stress-strain and piezoelectricity, one must deal with $M > 1$ tensors. Hence when constructing a graph neural network and passing “information” from atoms to bonds, and bonds to atoms, across multiple repeated layers, an essential question is how to best represent the physical state with the combination of scalar ($M = 0$), vector ($M = 1$), matrix ($M = 2$), $M = 3, \dots$ tensors with a fixed computational resource budget. For

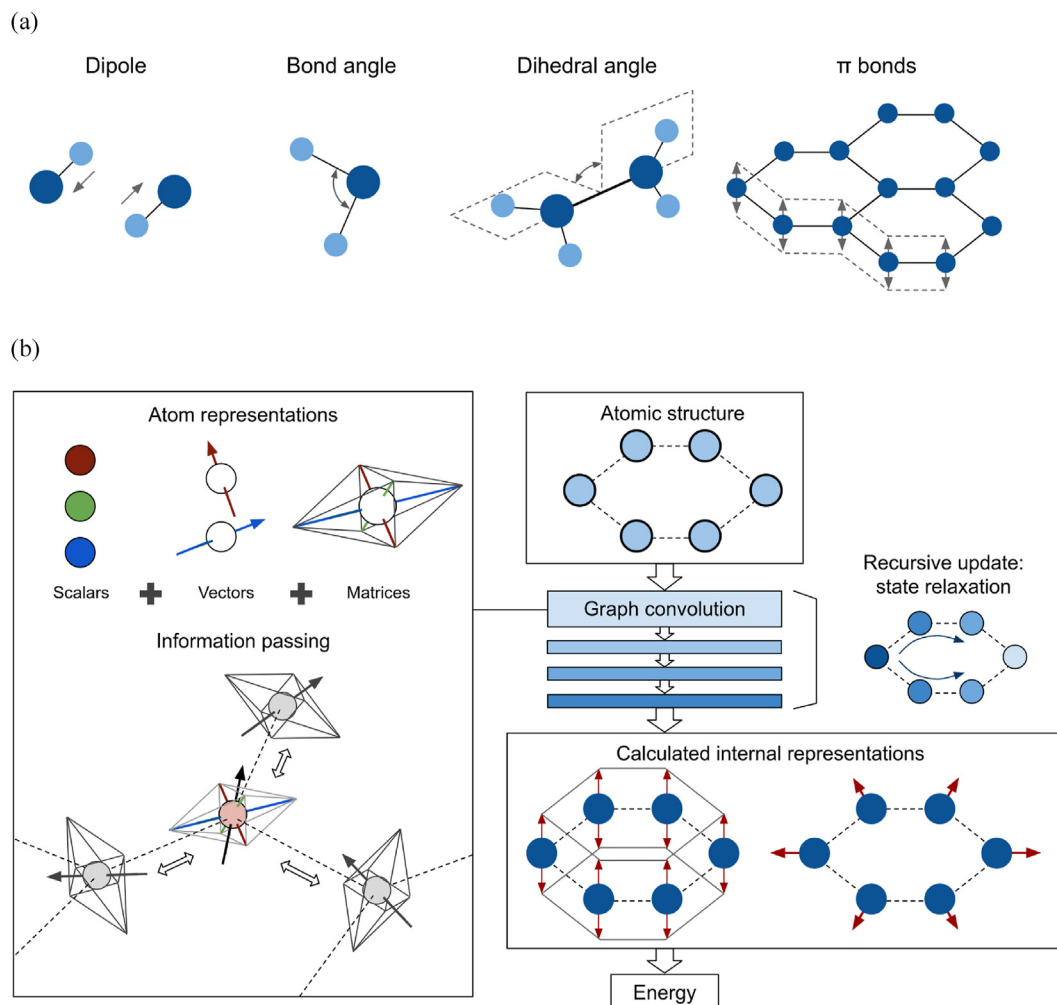


Fig. 4. (a) Example of atomic interactions depending on the geometry. These angle-dependent interactions are not always localized to a particular atom but can transfer among networks. (b) Schematic illustration of graph convolution operation with higher-order tensors. Each atom contains scalar ($M = 0$), vector ($M = 1$), and matrix ($M = 2$) tensors (shown in left top). In the graph convolution operation, those tensor values interact with each other, which represent higher-order geometric interactions. In addition, unlike the angle-based representation, those geometric features expand through the network by multiple graph convolution layers.

example, given one can afford to allocate 256 “double-precision” (8-byte) numbers for each atom which defines the “fatness” of the GNN, should one allocate

$$166 \text{ scalars} + 30 \text{ vectors} \quad (166 \times 1 + 30 \times 3) \quad (4)$$

or

$$115 \text{ scalars} + 20 \text{ vectors} + 6 \text{ matrices} + 1 \text{ T3} \quad (115 \times 1 + 20 \times 3 + 6 \times 3^2 + 1 \times 3^3) \quad (5)$$

to best represent the physical state of the iterative electronic relaxation? In terms of computational complexity, the cost of treating matrices or T3 or T4 (e.g., inner and tensor products) are not exorbitant compared to treating vectors and scalars – just think what one might do with 115 scalars as compared to what one might do with 6 three-by-three matrices (represented by 54 values). However, the 6 matrices can represent geometric features that 115 unstructured scalars simply cannot carry, as shown by (1)–(3). Indeed, an “asset allocation scheme” like (5) may be thought of as a Taylor series approximation of the target quantity in rotational complexity. Even though the current implementation of TeaNet/

PFM has a truncation at $M = 2$, there is really no reason why it has to be such, and we believe the best allocation would depend on the physical problem. Currently the “tensorial asset allocation” was done manually [6], but in the future we can optimize this allocation automatically as part of the hyperparameter tuning.

Innovative deep learning architecture alone cannot lead to successful applications; it must be exposed to sufficiently diverse and large training data to fully develop the desired model capability. Current deep learning works extremely well for interpolation tasks but not for extrapolation tasks. Therefore, the training dataset should cover a much broader region that encompasses the entire “high-school” (or even PhD-level) chemistry. The construction of the training dataset for TeaNet/PFM follows an iterative and adaptive way. In the following, we first explain why once-for-all model training rarely works and then highlight the active-learning strategy [15,19] employed in the current TeaNet/PFM training. Traditional machine learning interatomic potential development often starts with constructing a ‘comprehensive’ dataset based on domain knowledge, in a one-shot fashion. Typical examples of such a dataset include MD17 [11], QM9 [12], and Materials Project [13]. These datasets have been widely used to prototype and benchmark various NNIP architectures, significantly contributing to the development of chemoinformatics and materials informatics. However, the construction of these datasets is highly biased toward

more stable structures that are considered reasonable. Such predefined datasets bring in two issues. First, it requires sufficient (if not exhaustive) consideration on structures that are possibly encountered in applications. This can be very challenging even for a domain-expert and hardly be sufficient; human knowledge of the atomic structures resulting from the simulation and quantum chemical calculations are actually quite limited. Second, although models can achieve high accuracy on test data from the same predefined dataset, they are likely to perform much worse in practical applications. This is because the empirically-crafted dataset generally has ‘blind-spots’ that show up frequently in atomistic simulations as unexpected configurations. For example, the Open Catalyst 2020 (OC20) dataset, which focuses on the energy and structure of small molecules adsorbed on various material surfaces, mostly contains stable structures that are pre-defined [14]. The authors found that the pre-trained DimeNet++ model on this OC20 dataset can lead to unexpected results in actual molecular dynamics simulations. While the model successfully performed structure optimization calculations with a molecule adsorbed on a crystal surface, it resulted in an unnaturally deformed structure if the calculation was done only for the surface structure. The structure was further severely broken when the molecule was placed there later. This broken structure was predicted to be even more stable than the reasonable molecular adsorption structure. This example demonstrates a situation in which an AI agent moves into an unexpected configuration space during practical applications, and starts nonsensical operations due to an illusionary low-energy basin. This indicates that even in-domain simulations can easily reach “unexpected extrapolation regions” due to the limitations of one-shot hand-crafted training datasets, which can have catastrophic consequences for the simulation results.

The discrepancy between in-domain benchmarks and real-world performance should always be checked, as typical structural optimization and molecular dynamics calculations are iterative calculations that use the results of current inference as the next input. It is not uncommon for the number of computation loops to exceed 1 million before obtaining the desired result. In such iterative loops, the input data used for inference in machine learning models is not explicitly given by a human or other external entity, but is entirely determined by the model's behavior and the simulator. This is similar to the situation of reinforcement learning or sequential prediction/control tasks where the input distribution is non i.i.d. and strongly depends on the previous prediction results. In such cases, it is quite possible that the performance of the machine learning model in real-world applications will be significantly degraded compared to the in-domain inference accuracy of the dataset.

The discussions above clearly show that empirically-crafted datasets are usually not broad enough and can significantly deviate from what molecular dynamics simulations reach. To reduce human bias in input data collection, TeaNet/PFP proposes a framework to generate a training dataset by molecular dynamics simulation using the NNIP itself (Fig. 5). Specifically, atomic structures are generated by mixing many different elements at once, then the system is heated up to a very high temperature by molecular dynamics simulation, and finally the system is cooled down to relax the energy to some extent. Quantum chemistry calculations are performed on the structures obtained during this process, which are augmented to the training dataset to retrain the model. The retrained NNIPs can be used to recursively run this process to automatically expand the distribution of training data and increase the robustness of NNIP. It is noted that there are preceding studies which suggest active learning-based sampling, with small numbers of the chemical types of elements [15,19].

In this way, the ML model itself, rather than humans, generates the training data, thereby minimizing the influence of assumed knowledge on the training dataset. In addition, because the data is generated by molecular dynamics simulation, the existence of “unexpected extrapolation regions” described earlier can be efficiently reduced. Specifically, if there is a structure that should not be stable but is incorrectly inferred to be stable (“illusionary low-energy basins”) due to an insufficient dataset, the corresponding structures are expected to be generated by the above Boltzmann sampling process, and a large error will be discovered, and subsequently incorporated into the training data to be relearned. In the game AI analogy in reinforcement learning, we can find similarities in the process of self-learning by the agent's own trial-and-error. Note that the MD at high temperature mentioned above is an example, and the learning of PFP also includes datasets created with other “self-play” strategies, such as molecular or crystal structures in which some elements are artificially substituted by other elements. All of these techniques have been used to be as diverse as possible for the purpose of universality.

The resulting dataset to train our first universal potential is extremely diverse, covering 72 elements from Hydrogen to Bismuth. A portion of this dataset called high-temperature multi-element 2021 (HME21) dataset is now made available [16]. We have so far used 4×10^5 GPU days, or 1,144 GPU years, to generate the active-learning-driven DFT calculations. The number of atomic configurations in the entire dataset has reached 22 million [17]. Diverse test simulations have shown this empirical potential has outstanding performance, with energy error significantly less than the chemical accuracy (43 meV/atom) for even chemically very complex systems, including reaction saddle-point energies [7]. Considering the energetic error of DFT itself is often a few kcal/mol (~ 100 meV/atom) against experimental thermochemical measurements, this means TeaNet/PFP can represent DFT quite a bit better than how much DFT itself can represent chemical reality. The universal potential can however run millions of times faster than DFT when dealing with several thousand atoms, and the latest Matlantis™ release allows for more than 10 000 atoms of arbitrary combinations of 72 elements (soon to be extended to 94) to be simulated together. TeaNet/PFP also provides analytical interatomic force for molecular dynamics simulations and gradient driven annealers by chain-rule differentiation of the DNN, with excellent energy conservation in MD simulations. The mean - average error for force prediction is 0.126 meV/Å for these very diverse configurations. In the future, with more accurate *ab initio* energetics generated by using better density functionals, quantum Monte Carlo or coupled-cluster quantum chemical calculations, we expect the true accuracy of the PFP universal empirical potential to keep improving.

Given this universal empirical potential can reach much larger space-time volumes of simulation for chemically complex systems, what are the new opportunities for materials modeling? While DFT allows the study of extended defects such as dislocations, grain boundaries and phase boundaries, the microstructures are typically too simple (i.e., straight dislocations or boundaries, with occasional kinks). Going from few hundred atoms in DFT to 10^4 atoms in Matlantis™, one can study realistic extended defects (e.g., with curvatures) and their interactions, such as dislocation-dislocation junction strength, dislocation-interface interactions [18], realistic phase transformations with heterogeneous nucleation near extended defects [2], plastic deformation and damage evolution, i.e., stress-corrosion cracking in molten salts [3,4], electrochemical interfaces [20], etc. The much longer time scales of simulation would enable accurate thermodynamic and kinetic evaluations beyond the quasi-harmonic phonon theory, for example superionic diffusion, potential-of-mean-force free-energy difference

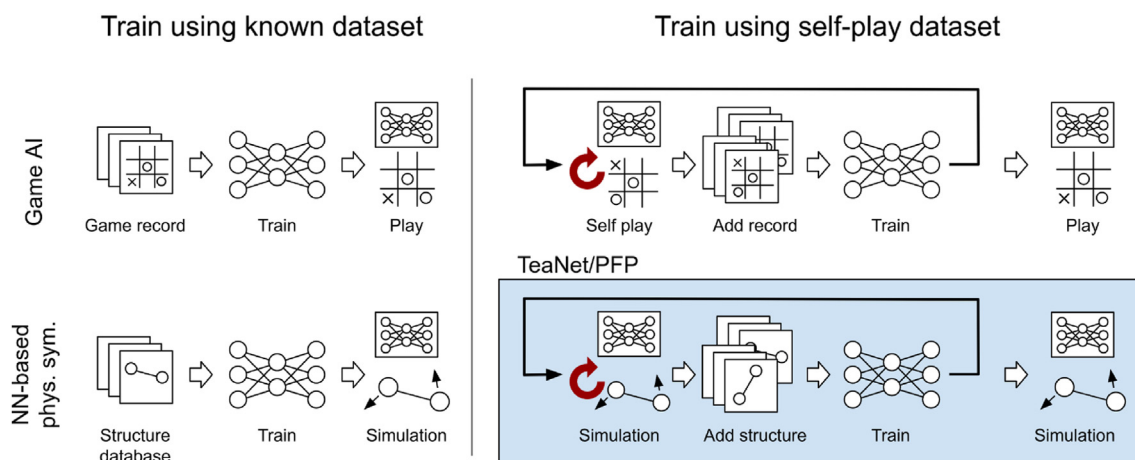


Fig. 5. Schematic illustration of dataset collection method in TeaNet/PFP. The left column shows the training method using the existing dataset and the right column shows the training method by self-play scheme.

sampling, high-temperature thermal conductivity calculation, *etc.*, that would not have been possible with direct DFT calculations, or at least without the strong limitations associated with the phonon theory. (Incidentally, we have also shown that TeaNet/PFP provides highly accurate interatomic Hessian as well as the nonlinear coefficients that would allow phonon scattering calculations of thermal conduction.)

Furthermore, the total energy and forces are just a subset of the physical tensorial measurable for an atomistic configuration. It would be no surprise if TeaNet/PFP architecture is extended to electronic-structure properties such as the electronic charge density distribution $\rho_e(\mathbf{r})$. It may also be applicable to the processing of physical phenomena on other scales, such as point cloud observation data. We have developed a user-friendly Application Programming Interface for these very diverse applications. Matlantis™ provides web-browser driven python Jupyter Notebook interface with many use cases, visualization tools and documentation. This user friendliness will help grow the application ecosystem of TeaNet/PFP. In turn, user feedback would drive further improvement of the empirical universal potential.

In summary, a universal interatomic potential applicable to an arbitrary set of chemical elements with high computational efficiency, which had not been achieved historically for a long time, has finally been realized. Its transferability and wide applicability have been shown in various material examples and phenomena, and opens the door for exploration into a myriad of materials, chemical and even biological systems.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: So Takamoto is working at Preferred Networks, Inc. Daisuke Okanohara is a representative director, and chief executive researcher of Preferred Network Inc. and is also the CEO of Preferred Computational Chemistry Inc, the commercial developer of PFP and Matlantis™ (<https://matlantis.com/>). Ju Li is a technical advisor and consultant to Preferred Networks, Inc., the commercial developer of Preferred Potential (PFP) and Matlantis™. See <https://www.preferred.jp/en/news/pr20210518/>

References

- [1] Interatomic potentials repository. <https://www.ctcms.nist.gov/potentials/>; 2022.
- [2] Tang Hao, Zhang Yin, Li Qing-Jie, Xu Haowei, Wang Yuchi, Wang Yunzhi, Ju Li. High accuracy neural network interatomic potential for NiTi shape memory alloy. *Acta Mater* 2022;238:118217.
- [3] Li Qing-Jie, Küçükbenli Emine, Lam Stephen, Khaykovich Boris, Kaxiras Efthimios, Ju Li. Development of robust neural-network interatomic potential for molten salt. *Cell Reports Physical Science* 2021;2:100359.
- [4] Lam Stephen T, Li Qing-Jie, Ballinger Ronald, Forsberg Charles, Ju Li. Modeling LiF and LiBe molten salts with robust neural network interatomic potential. *ACS Appl Mater Interfaces* 2021;13:24582–92.
- [5] Zuo Yunxing, Chen Chi, Li Xiangguo, Deng Zhi, Chen Yiming, Behler Jorg, Csanyi Gabor, Shapeev Alexander V, Thompson Aidan P, Wood Mitchell A, et al. Performance and cost assessment of machine learning interatomic potentials. *J Phys Chem* 2020;124(4):731–45.
- [6] So Takamoto, Izumi Satoshi, Ju Li. Teanet: universal neural network interatomic potential inspired by iterative electronic relaxations. *Comput Mater Sci* 2022;207:111280.
- [7] So Takamoto, Shinagawa Chikashi, Motoki Daisuke, Nakago Kosuke, Li Wenwen, Kurata Iori, Watanabe Taku, Yayama Yoshihiro, Iriguchi Hiroki, Asano Yusuke, Onodera Tasuku, Ishii Takafumi, Kudo Takao, Ono Hideki, Sawada Ryohto, Ishitani Ryuichiro, Ong Marc, Yamaguchi Taiki, Kataoka Toshiki, Hayashi Akihide, Charoenphakdee Nontawat, Ibuka Takeshi. Towards universal neural network potential for material discovery applicable to arbitrary combination of 45 elements. *Nat Commun* May 2022;13(1):2991.
- [8] Matlantis. <https://matlantis.com/>; 2022.
- [9] Huerta Gerardo Valadez, Hisama Kaoru, Sato Katsutoshi, Nagaoka Katsutoshi, Koyama Michihisa. Cyber catalysis: N2 dissociation over ruthenium catalyst with strong metal-support interaction. 2022.
- [10] Hisama Kaoru, Huerta Gerardo Valadez, Koyama Michihisa. Molecular dynamics of electric-field driven ionic systems using a universal neural-network potential. *Comput Mater Sci* 2022;218:111955.
- [11] Chmiela Stefan, Alexandre Tkatchenko, Sauceda Huziel E, Poltavsky Igor, Kristof T Schütt, Müller Klaus-Robert. Machine learning of accurate energy-conserving molecular force fields. *Sci Adv* 2017;3(5).
- [12] Ramakrishnan Raghunathan, Dral Pavlo O, Rupp Matthias, Anatole von Lilienfeld O. Quantum chemistry structures and properties of 134 kilo molecules. *Sci Data* Aug 2014;1(1):140022.
- [13] Jain Anubhav, Ong Shyue Ping, Hautier Geoffroy, Chen Wei, Richards William Davidson, Dacek Stephen, Cholia Shreyas, Gunter Dan, Skinner David, Ceder Gerbrand, Persson Kristin A. The Materials Project: a materials genome approach to accelerating materials innovation. *Apl Mater* 2013;1(1):011002.
- [14] Chanussot Lowik, Das Abhishek, Goyal Siddharth, Lavril Thibaut, Shuaibi Muhammed, Riviere Morgane, Tran Kevin, Heras-Domingo Javier, Ho Caleb, Hu Weihua, Palizhati Aini, Sriram Anuroop, Wood Brandon, Yoon Junwoong, Parikh Devi, Lawrence Zitnick C, Ulissi Zachary. Open catalyst 2020 (oc20) dataset and community challenges. *ACS Catal* 2021;11(10):6059–72.
- [15] Zhang Linfeng, Lin De-Ye, Wang Han, Car Roberto, Weinan E. Active learning of uniformly accurate interatomic potentials for materials simulation. *Phys. Rev. Materials* Feb 2019;3:023804.
- [16] So Takamoto, Shinagawa Chikashi, Motoki Daisuke, Nakago Kosuke, Li Wenwen, Kurata Iori, Watanabe Taku, Yayama Yoshihiro, Iriguchi Hiroki, Asano Yusuke, Onodera Tasuku, Ishii Takafumi, Kudo Takao, Ono Hideki, Sawada Ryohto, Ishitani Ryuichiro, Ong Marc, Yamaguchi Taiki, Kataoka Toshiki, Hayashi Akihide, Charoenphakdee Nontawat, Ibuka Takeshi. High-temperature multi-element 2021 (HME21) dataset. 2022. p. 4.
- [17] PFP v3.0.0 news release. <https://tech.preferred.jp/ja/>; 2022.
- [18] Zhu Ting, Ju Li, Amit Samanta, Hyoung Gyu Kim, Suresh Subra. Interfacial plasticity governs strain rate sensitivity and ductility in nanostructured metals. *Proc Natl Acad Sci USA* 2007;104(9):3031–6.
- [19] Dane Morgan, Ghanshyam Paliana, Adrien Couet, Blas P. Uberuaga, Cheng Sun,

Ju Li. Machine learning in nuclear materials research. *Curr Opin Solid State Mater Sci* 2022;26:100975.

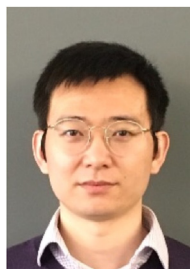
[20] Haoran Du, Yanhao Dong, Li Qing-Jie, Zhao Ruirui, Qi Xiaoqun, Wang Hay Kan. A New Zinc Salt Chemistry for Aqueous Zinc-Metal Batteries. *Adv Mater* 2023. <https://doi.org/10.1002/adma.202210055>.



So Takamoto is a researcher at Preferred Networks, Inc. He received his Ph.D. from the University of Tokyo in 2017 and awarded Dean's prize. After working as postdoctoral fellow at Massachusetts Institute of Technology and assistant professor at the University of Tokyo, he joined Preferred Networks, Inc. in 2019. His research includes machine learning-based atomistic simulation.



Daisuke Okanohara is a co-founder, Representative Director, and Chief Executive Researcher of Preferred Network Inc. Daisuke is also CEO of Preferred Computational Chemistry Inc. where they develop and provide a service for material research and development. He received his Ph.D. in Information Science and Engineering from the University of Tokyo in 2010. His research includes deep learning and its application.



Qing-jie Li is a postdoctoral associate at Massachusetts Institute of Technology. He received his Ph.D. from Johns Hopkins University in 2019 with the Donald S. Rodbell Fellowship. His research focuses on developing and applying atomistic modeling and machine learning techniques to understand materials mechanics and chemistry.



Ju Li has held faculty positions at the Ohio State University, the University of Pennsylvania, and is presently a chaired professor at MIT. His group (<http://Li.mit.edu>) investigates the mechanical, electrochemical and transport behaviors of materials as well as novel means of energy storage and conversion. Ju was elected Fellow of the *American Physical Society* in 2014, a Fellow of the *Materials Research Society* in 2017 and a Fellow of AAAS in 2020. He is the chief organizer of MIT A + B Applied Energy Symposia that aim to develop solutions to global climate change challenges with “A-Action before 2040” and “B-Beyond 2040” technologies.