ACCOUNTS of chemical research

pubs.acs.org/accounts Article

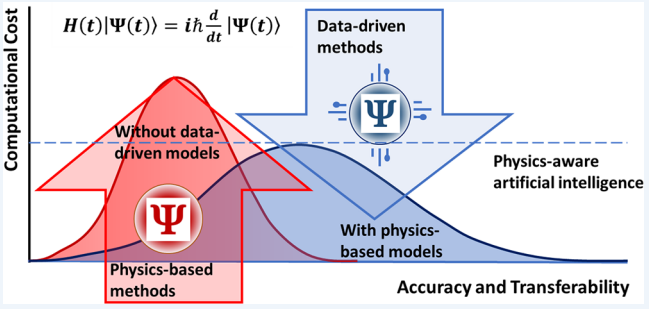
Development of Multimodal Machine Learning Potentials: Toward a Physics-Aware Artificial Intelligence

Published as part of the Accounts of Chemical Research special issue "Data Science Meets Chemistry".

Tetiana Zubatiuk and Olexandr Isayev*

Cite This: *Acc. Chem. Res.* 2021, 54, 1575–1585 Read Online

CONSPECTUS: Machine learning interatomic potentials (MLIPs) are widely used for describing molecular energy and continue bridging the speed and accuracy gap between quantum mechanical (QM) and classical approaches like force fields. In this Account, we focus on the out-of-the-box approaches to developing transferable MLIPs for diverse chemical tasks. First, we introduce the "Accurate Neural Network engine for Molecular Energies," ANAKIN-ME, method (or ANI for short). The ANI model utilizes Justin Smith Symmetry Functions (JSSFs) and realizes training for vast data sets. The training data set of several orders of magnitude larger than before has become the key factor of the knowledge transferability and flexibility of MLIPs. As the quantity, quality, and types of interactions included in the training data set will dictate the accuracy of MLIPs, the task of proper data selection and model training could be assisted with advanced methods like active learning (AL), transfer learning (TL), and multitask learning (MTL). Next, we describe the AIMNet "Atoms-in-Molecules Network" that was inspired by the quantum theory of atoms in molecules. The AIMNet architecture lifts multiple limitations in MLIPs. It encodes long-range interactions and learnable representations of chemical elements. We also discuss the AIMNet-ME model that expands the applicability domain of AIMNet from neutral molecules toward open-shell systems. The AIMNet-ME encompasses a dependence of the potential on molecular charge and spin. It brings ML and physical models one step closer, ensuring the correct molecular energy behavior over the total molecular charge.

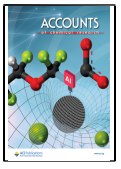


We finally describe perhaps the simplest possible physics-aware model, which combines ML and the extended Hückel method. In ML-EHM, "Hierarchically Interacting Particle Neural Network," HIP-NN generates the set of a molecule- and environment-dependent Hamiltonian elements $\alpha_{\mu\mu}$ and $K^{\ddagger}$. As a test example, we show how in contrast to traditional Hückel theory, ML-EHM correctly describes orbital crossing with bond rotations. Hence it learns the underlying physics, highlighting that the inclusion of proper physical constraints and symmetries could significantly improve ML model generalization.

KEY REFERENCES

- Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8*, 3192−3203.[1] *The first transferable NNP with accuracy comparable to DFT that is applicable to broad classes of organic molecules.*
- Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching Coupled Cluster Accuracy with a General-Purpose Neural Network Potential through Transfer Learning. *Nat. Commun.* **2019**, *10*, 2903.[2] *TL implementation to train NNP that approaches CCSD(T) accuracy on diverse benchmarks: thermochemistry, isomerization, molecular torsion.*
- Zubatyuk, R.; Smith, J. S.; Leszczynski, J.; Isayev, O. Accurate and Transferable Multitask Prediction of Chemical Properties with an Atoms-in-Molecules Neural Network. *Sci. Adv.* **2019**, *5*, eaav6490.[3] *Development of AIMNet modular deep NNP. The AIMNet shows a new dimension of transferability: the aptitude in learning new features from foregoing training through multimodal information.*
- Zubatyuk, R.; Smith, J.; Nebgen, B. T.; Tretiak, S.; Isayev, O. Teaching a Neural Network to Attach and Detach Electrons from Molecules. *ChemRxiv*, July 28, **2020**.

Received: December 22, 2020
Published: March 13, 2021



https://dx.doi.org/10.1021/acs.accounts.0c00868
Acc. Chem. Res. 2021, 54, 1575−1585

Table 1. Main Computational Distinctions among QM and ML-QM Methods

	method	scaling	time scale of simulations	training data req	transferability[a]	accuracy	algorithms for improvements
physics	QM	$\sim O(N^{3-7})$	ps−ns		high	high	
	semiempirical	$\sim O(N^{2})$	ns	low	medium	medium	refitting
	force fields	$\sim O(N^{1-2})$	μs−ms	medium	variable	low	refitting
ML	kernel methods	$\sim O(N^{3})$	ns	medium	low	variable	Δ,[b] AL
	HDNNP	$\sim O(N^{1-2})$	μs	medium	low	variable	Δ, AL
	ANI	$\sim O(N^{1-2})$	μs	high	medium	variable	Δ, AL, TL
	AIMNet	$\sim O(N^{1-2})$	μs	high	medium	variable	Δ, AL, TL, MTL
physics + ML	?	?	ps−ms	low	high	high	all of the above

[a]Transferability is defined as the ability to transfer the knowledge of a model trained from one data domain to the other data domain. [b]Δ-Learning (see section 2 for details).

10.26434/chemrxiv.12725276.v1.[4] *Extension of AIMNet framework toward open-shell molecular systems. AIMNet-Me examines a new dimension of transferability by including charge-spin space and feasibility to model conceptual DFT quantities without QM.*

1. INTRODUCTION

"*If, in some cataclysm, all of scientific knowledge were to be destroyed, and only one sentence passed on to the next generation of creatures, what statement would contain the most information in the fewest words? I believe it is theatomic hypothesisthat all things are made of atoms. In that one sentence, you will see, there is an enormous amount of information about the world, if just a little imagination and thinking are applied.*"[5] (Richard P. Feynman). The crucial property of atoms that determines that enormous amount of information is how they interact. Specifically, what is their *interatomic potential* (IP)? This question can be answered from the first-principle or *ab initio* by solving the Schrödinger equation (SE). While this is undoubtedly the most reliable approach, there is a practical reason for speed to consider simpler approximations that capture quantum mechanics' essential physics.

The computational costs of accurate quantum mechanical (QM) calculations scale exponentially with system size. Since the dimension of the wave function of methods such as configuration interaction[6] (CI) or coupled cluster[7] (CC) hits a performance wall with only tens of atoms, large systems simulation is impossible. Application of *ab initio* electronic structure calculations for large molecular systems (10^2−10^3 atoms) leads to the computations of such a formidable order of magnitude that it would seem hopeless to tackle them without the new approaches. In general, the simpler descriptions or alternative representations of IPs could give a systematic view of the complex and, at first glance, incomparable data. Today, the latter statement is even more relevant. Density functional theory (DFT)[8] and QM calculations of the electronic structure have vastly extended the body of accurate data available for training data sets generation. Keeping in mind only two mentioned reasons, the recent remarkable progress in numerical computations in physical and chemical science becomes apparent. The importance of skillful numerical analysis in chemistry was shrewdly pointed out by Löwdin[9] over half a century ago when quantum chemistry started to explore areas beyond the study of single molecules to systems of biological interest.

From the mathematical point of view, problems of complex chemical systems can be described by linear or nonlinear functions. One of the most popular statistical methods to analyze linear problems is partial least-squares[10] (PLS) analysis. In a single-component system, for example, the IP is linear in relation to the unknown variables and it is easy to fit them using PLS. However, the IP becomes nonlinear in a multicomponent system. In this case, a nonlinear function of several variables could be efficiently approximated with artificial neural networks (NNs).[11−13] NNs provide accurate results for complex nonlinear problems and reduce the computational work associated with a given calculation.

As seen from the above, the idea to somehow construct the IPs, not in terms of quantum theory but utilizing regression, was inspired by the technical innovations and development of statistical analysis. With the pioneering work of Doren et al.,[14] the field of NN potentials (NNPs) for the representation of potential energy surfaces (PESs) has become an independent area of research. Since then, many variations on NNPs have been introduced. The first NNPs[14,15] accurately reproduced only the regions near the minima on the PES ("low- dimensional" NNPs) of the small molecular systems. The subsequent introduction of "high-dimensional" NNPs by Behler and Parrinello[16] (HDNNP) has enabled studies for molecular systems of thousands of atoms. The first HDNNP was designed for bulk silicon.[16] It introduced Behler and Parrinello symmetry functions (BPSFs) as a structural fingerprint. The main idea of BPSF was to describe the relevant local environment of each atom as an *atomic environment vector* (AEV) and subsequently transform AEVs as input for the HDNN. However, applications of HDNNPs were limited by the homogeneity of the atomic environment of the studied chemical systems. This caused a lack of transferability and limited HDNNPs to simulations of a single chemical system at a time. In other words, HDNNPs need to be retrained for every individual application.

It should be noted that, besides NNPs, IPs have also been developed with many other ML methods. For example, Gaussian approximation potentials,[17] kernel ridge regression,[18] and support vector machines[19] are widely used to construct PES. A comprehensive review of the MLIPs recent advances was given elsewhere.[20] A presentation of the theory behind NNs and modern deep learning[21] is beyond the scope of this Account too. Therefore, we will focus this Account on the pursuit of universal HDNNPs undertaken by our group, identifying advanced construction methods. On this destination, we discuss automatic *active learning* (AL),[22] *transfer learning* (TL),[2] and *multitask learning* (MTL)[3] techniques for sampling vast chemical space or taking advantage of different approximations to SE.

2. MACHINE LEARNED INTERATOMIC POTENTIAL, MLIP

The IP $U(\boldsymbol{r}_1,\boldsymbol{r}_2,...,\boldsymbol{r}_N)$ describes the dependence of the potential energy of a system of N atoms on their coordinates $\boldsymbol{r}_1,\boldsymbol{r}_2,...,\boldsymbol{r}_N$. The well-established hierarchy of IPs concerning the structure of

https://dx.doi.org/10.1021/acs.accounts.0c00868
Acc. Chem. Res. 2021, 54, 1575−1585

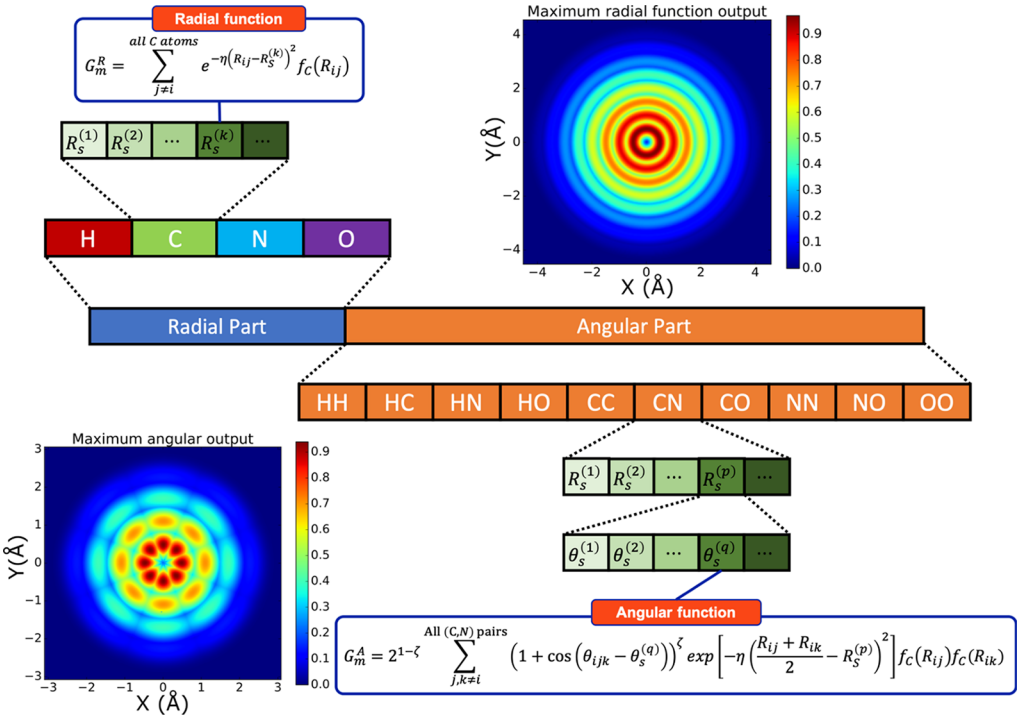


Figure 1. Structure of the ANI AEVs. The sum of j and k is on all neighbor atoms of selected atoms/pairs. f_C, cutoff cosine function; R_C, cutoff radius. Reprinted with permission from ref 33. Copyright 2020 American Chemical Society.

matter is represented by the increase of sophistication in the description of electron correlation, as follows: *empirical*, *semiempirical*, and *ab initio* IPs. The *ab initio* IPs provide a unifying approach for describing the behavior of molecular systems. Unfortunately, the accessible time scales for QM simulations are extremely limited (Table 1). Many important chemical and biological phenomena cannot be modeled with QM. Meanwhile, the interpolation of *ab initio* calculations with NN makes it possible to extend the time scale in molecular dynamics simulations while maintaining the accuracy of the reference QM method.[15−18] The accessible simulation length of ML methods, as of today, is approaching a microsecond. Table 1 summarizes the basic features of IPs of different approaches and shows that MLIPs provide good accuracy and transferability with comparable computational cost to traditional force fields in many applications.

The development of accurate and transferable MLIPs is a challenging task. The optimal procedure for the selection of either training data or the NN model is not well understood. Most ML models trained with kernel methods like Kernel ridge regression or BP-type HDNNPs are parametrized to specific chemical systems. Despite the fact, the MLIPs achieve high accuracy with small training data ($\sim 10^3-10^5$ samples). Their downside that one must retrain new IPs for every new chemical application and generate new QM training data.

The ANI and AIMNet models show our attempts to design universal MLIPs for neutral organic molecules. Our key innovations are increasing the angular-radial resolution of AEV descriptors and radical increasing training data set sizes up to 10^6-10^7. These two features force the NN to learn low-level interatomic interaction and transferability between different classes of molecules.

2.1. ANAKIN-ME Neural Network Potential (ANI)

Justin Smith Symmetry Functions or JSSFs. Similar to many-body potential, standard HDNNP models the total energy of a molecule as a sum over local contributions at each atom. Meanwhile, in contrast to many-body expansion, HDNNPs is not a rigid functional expression but a flexible nonlinear ML model. The key point of such expression is the introduction of special type coordinates, which are appeared to be invariant to the exchange of equivalent atoms along with translation and rotation of the system. Those special type coordinates are called *symmetry functions*, and they are many-body functions of all atomic positions inside the cutoff spheres. While many different types of transformed coordinates are known,[23−27] ANI approach implements the idea of BPSFs[16] and encodes the arrangement of atoms in a molecule in terms of fixed-sized AEV representation. AEV introduces the i^{th} atom in the molecule in terms of its (i) *local structure* described by a set of coordinates of all atoms within a cutoff distance and the (ii) *local chemical composition* given by the types of chemical elements. With the ANI model, we have introduced Justin Smith Symmetry Functions or JSSFs.[1]

Similarly to BPSFs, JSSFs encode the atomic configurations using interatomic distances with i^{th} atom neighbors within a cutoff radius R_c into invariant fixed-length AEV, $\vec{G}_i^Z = \{G_1, G_2, G_3, ..., G_M\}$. Elements G_M probe specific regions of i^{th} atom *radial* and *angular* chemical environment. Each ith atom with atomic number Z is encoded with vector $\vec{G}_i^Z$. Vector $\vec{G}_i^Z$ is then used as input of a single NN. With invariant AEV $\vec{G}_i^Z$, the total HDNNP energy of a molecule m is expressed as

$$E_{total}(m) = \sum_{i}^{all\ atoms} NNP_{Z_i}(\vec{G}_i^Z) \quad (1)$$

The form of $\vec{G}_i^Z$ vectors defines the flexibility and transferability of the HDNNPs. The BP-type AVEs either do not account for the type of neighboring atoms or account for them with separate

https://dx.doi.org/10.1021/acs.accounts.0c00868
Acc. Chem. Res. 2021, 54, 1575−1585

sub-AEVs. The latter is done with the "one-hot encoding," which assigns each atom type its orthogonal bit vector. The size of the radial AEV grows as $O(N^2)$ with the number of atom types and causes a corresponding growth of the *complexity* of the HDNNPs. This scaling has so far limited applications of HDNNPs to molecular systems with up to four types of atoms.[14,16,28−30]

JSSFs mitigate this problem with decreased complexity of AEVs and a substantially bigger data set. The angular and radial parts of JSSFs vectors $\bar{G}_i^Z$ are modified with *angular shift*, θ_s, and *radial shift*, R_s hyperparameters. The structure of the JSSFs AEV is shown in Figure 1. The radial AEV is divided into sub-AEVs according to atom types (H, C, etc.). Similarly, the angular AEV is divided into sub-AEVs according to types of atom pairs (HH, HC, etc.). New parameters θ_s and R_s allow an arbitrary number of shifts in the angular and radial environments correspondently.

The effect of JSSFs modification is that AEV elements become smaller and provide a distinctive image of various molecular features. JSSFs assist NNs in learning molecular features of specific bonding patterns, rings, functional groups, or other molecular moieties. The changes allowed to push the boundary and develop HDNNPs with up to eight chemical elements.[31,32] Interested readers are referred to[1,22] for in-depth technical discussions.

More generally, this limitation has been solved by introducing *learnable representation*, *i.e.*, embedding atomic type information into fixed-size AEVs. Several alternative approaches were also proposed in DTNN[34] SchNet,[35] HIP-NN,[36] and AIMNet[3] models. In AIMNet[3] (see section 3), we introduced atomic feature vectors (AFVs). The AIMNet model explicitly maintains the distinction between local structure and local chemical composition. It uses two sets of invariant vectors that separately encode the atomic environments and atomic species. The latter concept provides the constant complexity of AEVs, which does not scale with the number of chemical elements, and thus enables high transferability and flexibility of HDNNPs.

2.2. Training Data

Occam's Razor states: "The simplest explanation compatible with the observations should be preferred".[37] This principle argues that simplicity is preferred over complexity. Let us discuss how the principle of simplicity manifests with ML models of high complexity. On the one hand, ML uses the so-called *Minimum Description Length* (MDL) principle[38] for comparing different ML models and model selection. MDL is based on the following insight: learning is viewed as data compression through pattern recognition. Therefore, the more we compressed the data, the more we have learned about it with the ability to transfer the knowledge to unseen samples.

On the other hand, HDNNPs have an extremely large number of parameters compared to other ML models. If we use MDL to measure the complexity of an HDNNP, it will look really bad. Therefore, such NNs models with high complexity would be considered overfitted, yet they often obtain high accuracy on test data. Indeed, one of the central concepts of statistical learning, the bias-variance trade-off, appears to be at odds with the observed behavior of HDNNPs. The trade-off between a model's ability to minimize prediction errors implies that a model should balance bias and variance, which means to be rich enough to reveal real data structure and simple enough to avoid fitting the noise.[37] The apparent contradiction has stimulated to study this phenomenon. Belkin et al. suggested that NNs do not work in this regime but follow the "double-descent risk curve."[39]

The traditional ML uses the U-shape risk curve to measure the trade-off between bias and variance. It quantifies how generalizable a model is and is directly linked to model complexity. Modern deep learning models are heavily overparameterized and could be easily "overfitted" to reach zero train error. In the traditional view of bias-variance trade-off, this would result in catastrophic model performance and the total inability of generalization. Belkin hypothesized that increasing function class capacity of a ML model (*e.g.*, number of parameters or size of the NN architecture) allows the model to find "simpler" solutions and thus improve the performance on out-of-sample examples in the overfitting regime.[39] Considering the larger function classes that enable finding of "simpler" interpolating functions is a form of Occam's Razor for NN models. This hypothesis might rationalize the unreasonable effectiveness of ANI and AIMNet models vs standard HDNNPs. The overall ANI architecture is at least 10−50 times more complex ($\sim 10^6$ weights). The HDNNPs are typically heavily regularized and must be trained in a traditional bias-variance trade-off regime using a small data set.

The accuracy of MLIPs also critically depends on the amount, quality, and types of interactions included in the training data set. With developing transferable HDNNPs, our goal is to predict energies of the very distinct organic molecules. Ideally, the training data set should cover a full sampling of configurations in 3D space and include all conformers of every possible combination of all atom types in a molecule. Since the size of the conformational space scales exponentially with the molecule size, the first generation of ANI or ANI-1 was limited to the data set of organic molecules with only four atom types: C, O, N, and H. For practical applicability, we also restricted the data set to nonequilibrium conformations within ∼200 kcal/mol window. One important aspect should be noted that available public databases like PubChem, ChEMBL, or GDB cover only *configurational space* but not *conformational space* of the molecules, while the latter is a critical requirement for the NN training procedure. The conformation generation process was carried using two sampling techniques: normal vibrational modes sampling and molecular dynamics.[40] Final ANI-1 training data set includes ∼22 million molecular energies of randomly selected molecular conformations. The conformational space has been generated from GDB[41] database for ∼58K molecules with up to 8 types of heavy atoms.

ANI-1x Training Data Set. The second generation of training data set ANI-1x[22] is produced through *active learning*, AL (see section 2). The size of the ANI-1x data set was considerably reduced to ∼5 million conformations for CHNO molecules and 9.5 M for seven elements. Despite the smaller training data set, the ANI-1x potential vastly outperforms ANI-1. The ANI-1x data set uses less than 100 conformations per molecule, compared to ∼400 in the ANI-1 data set. The accuracy of the ANI-1x potential is comparable to that of the best HDNNPs, while most of the HDNNPs require vast conformational space for the parametrization. This observation further validates the "big data" sampling philosophy introduced in the original ANI-1 work.[1] Moreover, in the ANI-1x training data set, the mean molecule size is only 15 atoms and it can therefore be continuously expanded with more accurate post-Hartree−Fock data.

https://dx.doi.org/10.1021/acs.accounts.0c00868
Acc. Chem. Res. 2021, 54, 1575−1585

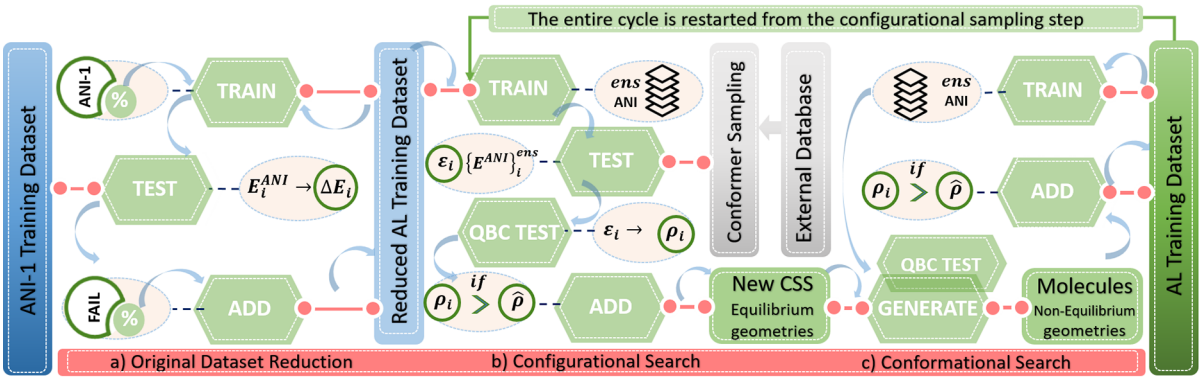


Figure 2. Active learning pipeline for generating ANI-1x data set. Here, $\hat{\rho}$ is an inclusion criterion (from QBC), ρ_i defines the standard deviation of predictions from an ensemble of ANI MLIPs, and ε_i is a prediction error per atom (from ensemble of ANI MLIPs).

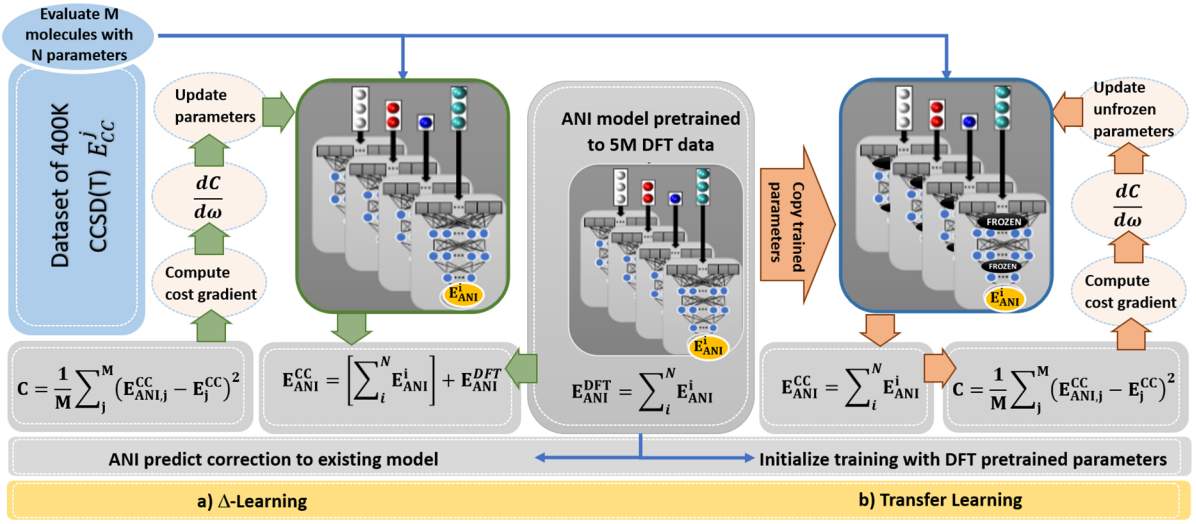


Figure 3. TL and Δ-learning techniques: (a) Δ-learning uses two separate neural networks. The first model is trained using low-fidelity (DFT) energy. The second model is then trained to correct the difference between DFT and the high-fidelity (CCSD(T)/CBS) method. (b) TL starts from a pretrained low-fidelity (DFT) model and then retrained to higher accuracy CCSD(T)/CBS-extrapolated data.

3. ADVANCED METHODS FOR ACCURACY AND TRANSFERABILITY IMPROVEMENT

The MLIP accuracy could be addressed from two orthogonal but data-related avenues: *quantity* and *quality* of training data. *Active learning*, AL, or on-the-fly learning helps one to obtain good ML models without having to rely on unnecessary large data sets. In this scenario, the MLIPs transferability problem is solved by rigorous *uncertainty quantification* (UQ) and the selection of poorly predicted samples for a model improvement. Most studies concur that adaptive AL training is more accurate and data-efficient than passive learning with fixed training data set. The *quality* of training data is dictated by the accuracy of reference QM calculations. The solution to this issue was found through *transfer learning* TL,[42,43] which could take advantage of combining the multiple levels of theory. TL has been shown to reduce the amount of high-quality data required for training.

Active Learning

In the context of MLIP training, AL aims to identify which new data points should be added during training data set generations.[22] This is done by predicting whether the potential correctly describes a new molecule. The model outputs the "yes"/"no" answer to the question of whether to run QM calculation for this structure or "*labeling*" the data. Here, the unlabeled data are abundant, but labeling (getting new QM energy) is expensive. Different labeling approaches are known from the literature: Deep Potential Generator (DP-GEN)[44] and AL-NN[45] with a Nested Ensemble Monte Carlo scheme. To develop ANI-1x MLIP, we have generated the AL training data set with an updated on-the-fly via Query by Committee (QBC)[46] algorithm (Figure 2). QBC iteratively queries a large chemical space of molecular conformations and samples high error regions and updates the training pool minimizing the need for redundant QM calculations. The fully automated AL workflow for data set generation has three steps: initial data set reduction, configurational search, and conformational search. Details can be found in refs 22 and 47.

(a) Initial *data set reduction*. ANI HDNNP is trained with a randomly subsampled ANI-1 data set. The remaining ANI-1 data are tested, and the small percentage of high-error structures are added to the training data set.

(b) *Configurational search* is performed by screening an external database with an ensemble of previously trained ANI models. All molecules that fail the QBC test form a new conformer sampling set (CSS).

https://dx.doi.org/10.1021/acs.accounts.0c00868
Acc. Chem. Res. 2021, 54, 1575−1585

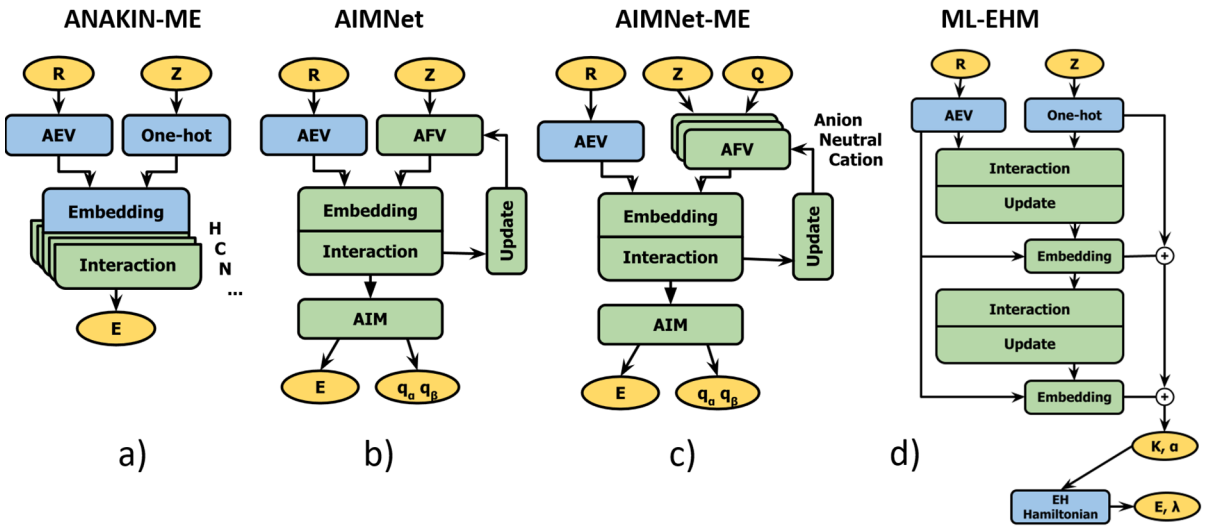


Figure 4. Basic NN architectures considered in this Account: ANI, AIMNet, multiembedding AIMNet-ME, and ML-EHM. The yellow blocks show input data (coordinates R, atomic numbers Z, total molecular charge Q) and output (energies E, spin-polarized charges q, EHM diagonal matrix elements α, and empirical fitting coefficient K). The green blocks indicate NN modules for training and the blue blocks show mathematical transformations.

(c) *Conformational search* cycle generates a set of nonequilibrium conformers for the molecules from new CSS via one of the three sampling techniques: diverse normal mode, K random trajectory, and MD. The conformers which fail subsequent QBC test are labeled and added to the AL training data set. New ensemble of five ANI NNs is trained. The conformational search cycles are repeated until the model stops improving. Then the entire cycle is restarted.

Transfer Learning

Smith and co-workers[2] have demonstrated how TL could be efficiently implemented for the development of HDNNPs with accuracy comparable to coupled-cluster CCSD(T).[48] The resulting ANI-1ccx model afforded a training set consisting of 90% of DFT calculations, while only a fraction of data (∼10%) is required for expensive CCSD(T) calculations. The generalized workflow is schematically depicted in Figure 3b. First, the ANI model is trained to the DFT data and provides HDNNP which is equivalent to ANI-1x. Then, ANI-1ccx HDNNP is retrained to the CCSD(T)*/CBS data with some optimized NN parameters remaining constant. NN parameters are organized in a series of hidden layers. The ANI models trained in this work have four hidden layers. The two hidden layers are left fixed to avoid overfitting to the smaller CCSD(T)*/CBS data set. The other two hidden layers are left to be optimized during the TL procedure.

TL provides a pragmatic way to lower the computational cost of obtaining high-level QM data, although yet not used widely in training HDNNPs. Our case studies with ANI-1ccx potential show a significant reduction in errors and even outperform DFT methods in different applications such as isomerization energies, reaction energies, molecular torsion profiles in nonequilibrium geometries.[2]

An alternative to TL is Δ-learning.[49] With the Δ-learning technique (Figure 3a), a new ML model is trained to correct the difference between accurate CCSD(T)/CBS and the existing ML model taught previously on DFT data. Although Δ-learning and TL allow HDNNPs to reach the same accuracy, Δ-learning must evaluate two separate NNPs in order to make a prediction.

4. AIMNET NEURAL NETWORK POTENTIAL—INCLUDING MORE PHYSICS

So far, we focused our discussion on data-related aspects of HDNNPs construction. We have discussed how to encode molecules for training and how to get an optimal data set. Meanwhile, the success of MLIPs is also attributed to a flexible functional form of NNs for data fitting. Deep learning enables a model to learn features automatically and recognize patterns, therefore finding a representation of the data. Obtained patterns and statistical correlations are entirely data-driven. They do not necessarily capture the underlying physics and chemistry of atoms in molecules. This problem is closely associated with the transferability of MLIPs. Current cutting-edge research in HDNNPs is focused on capturing the correct physical behavior by combining physical models with ML or so-called physics-aware artificial intelligence (PhAI).[50−52] The PhAI models promise to improve generalization by constraining NN to obey physical laws and reducing the possibility of overfitting and data memorization.

In the AIMNet[3], which takes inspiration from Bader's quantum theory of atoms in molecules (QTAIM),[53] we made an attempt to improve the HDNNP performance for long-range interactions. Atoms and bonds in the QTAIM model are expressed as the function of the electron density, which describes the average distribution of electron charges in the nuclei's attractive field. In AIMNet, atoms and bonds are expressed by the *atomic feature vectors* (AFV). The interactions with the neighboring atoms are transmitted through messages within the neural network. The main distinction of AFV from AEV is that instead of assigning fixed AEV to each atom type, in AIMNet atoms are characterized by *learnable* AFVs, which encode similarities between chemical elements. The latter eliminates the dependence of the AEV size on the number of element types.

Figure 4 lists all architectures considered in this Account. The diagram in Figure 4b shows the architecture of the AIMNet

https://dx.doi.org/10.1021/acs.accounts.0c00868
Acc. Chem. Res. 2021, 54, 1575−1585

which consists of several blocks: input data, embedding block, iterative block, and learned atomic interaction layer (AIM). First, the Cartesian coordinates $\vec{R}$ are submitted to the embedding block which encodes information about the AFVs and AEVs. Due to the implementation of JSSFs in the embedding block, the molecules being encoded in a rotation-translation invariant way.

To address the presence of different atom types in the chemical environment, the AIMNet operates with a tensor product of two coordinate vectors $\vec{G}_i^Z \otimes A_j$, where $\vec{G}_i^Z$ is denoted by the AEV of the i^{th} atom and A_j is denoted by the AFV of the i^{th} atom. The subsequent summation of the outer products of AEV with AFV restricts the AEV dimension and prevents its growth with a number of neighboring atom types in AIMNet. For comparison, the ANI-type HDNNPs use only a concatenation of the sums of JSSFs $[\Sigma_j G_{ij}^{(r)}\ \Sigma_{jk} G_{ijk}^{(a)}]$, where $G_{ij}^{(r)}$ and $G_{ijk}^{(a)}$ encode radial contribution from neighboring j^{th} atom and angular contribution from neighboring jk atom pair to the AEV of the i^{th} atom. AIMNet embedded the same radial and angular JSSFs complementing them with corresponding AFVs of the atom A_j and the atom pair $A_{jk}^{(p)}$. Overall, the AIMNet AEV vector $\vec{G}_i^Z$ is defined as follows:

$$\vec{G}_i^Z(\text{AIMNet}) = G_{ij}^{(r)} \cdot A_j,\ \ G_{ijk}^{(a)} \cdot A_{jk}^{(p)} \tag{3}$$

The AEV vector $\vec{G}_i^Z$ further undergoes a nonlinear transformation through the NN block, which produces the *environmental field layer*, F_i. The influence of the neighboring atoms on the central i^{th} atom through the F_i layer is expressed, as follows:

$$F_i = f_{\text{MLP}}([G_i^{(r)},\ \ G_i^{(a)}]) \tag{4}$$

where the f_{MLP} perceptron function extracts information about the environment of the central i^{th} atom. Next, in the *interaction* block (Figure 4b), DNN transforms F_i layer to the new *Atom-In-Molecule layer (AIM), Mi*. Due to the network architecture, the AIM layer encodes the information used to predict multiple atomic and molecular properties as well as to update AFV to get better predictions on the next "SCF-like" iteration. This information is further decoded with separate DNN blocks. Finally, we showed that the information encoded in the AFVs could be used to differentiate chemical elements and a chemical environment of an atom in a molecule at the same time. We refer interested readers to the original paper[3] for a detailed discussion.

Multitask and Multimodal Learning

The type of training data usually limits the applicability domain of NNPs. At the same time, transferable ML models are intended to be extensible in the description of the chemical and conformational diversity of organic molecules. However, due to the architectural limitations, most HDNNPs, including ANI and AIMNet, were parametrized only for the neutral molecules or closed-shell ions. In our subsequent work, we made an attempt to extend the AIMNet framework toward open-shell molecular systems. We have included charge-spin dependence to enable the MLIP to explore a new dimension of transferability. Figure 4c represents the resulting *multiembedding* AIMNet-ME[4] model, which provides a discrete and physically correct behavior of MLIP with respect to the spin and charge. This architecture brings ML models one more step closer toward PhAI, where the ML model correctly describes the energetics of the systems with long-range dependency in the electronic structure.

NNP with Charge Equilibration and Spin-Charge

For most NNPs, including ANI-type models which are based on local short-range descriptors, the accurate prediction of the system energetics with long-range interactions or ionization states cannot be performed efficiently. That is because the total charge and spin multiplicity are highly nonlocal. The local atomic environment representation does not correctly describe this phenomenon. Following first-principles QM to incorporate nonlocal information, the model should adapt according to the electronic structure changes. Hence, with our AIMNet-ME potential, we have attempted to include the long-range effects into the atomic environment representation of the molecule. That could be achieved with a "*message-passing*" through the neural network that redistribute spin-charges. It should be noted that the AIMNet *charge redistribution* technique[4] is an alternative to the *charge equilibration NN* technique (CENT).[54] AIMNet builds MLIP through the iterative "SCF-like" updates made to the AFVs and atomic embeddings. This procedure achieves the optimal electronic charge distribution over the *charge-spin space* of the molecular system. As it has been shown in the original paper,[4] the AIMNet potential excellently reproduces three global CDFT indexes: electronegativity (χ), chemical hardness (η), and electrophilicity (ω) with R^2 ranging from 0.93 to 0.97 and overall achieves "chemical accuracy" for spin-polarized atomic charges of organic molecules with up to nine atom types.

Figure 4c shows that AIMNet-ME shares the same interaction and update blocks for neutral and charged molecules, keeping only their AFVs as different inputs. NN training for neutral and charged molecules is carried out jointly, and errors for each charged state are averaged in the loss function. That enables AIMNet-ME to learn different atomic representations inside one model according to their iterative charge redistribution through the NN blocks. The AIMNet family of methods could thus serve as a neural charge redistribution scheme. Flexible integration of QM data and adaptable electronic description within AIMNet is a step toward developing a general-purpose single NN architecture capable of quantitatively predicting nonlocal molecular properties.

5. INTERFACING PHYSICS AND DEEP NEURAL NETWORKS

With training HDNNPs to approximate results of QM calculations, one can predict properties and energies within milliseconds. Although advanced HDNNPs are nonparametric IPs and can be systematically improved with complex AEV descriptors, they still are not tied to the physical form and symmetries in the Schrödinger equation. This HDNNP deficiency causes uncertainties and issues with extrapolation. Here, we discuss our efforts to combine HDNNP with an effective Hamiltonian model. With the latter, we expect to maintain the accuracy of HDNNP along with the interpretability of the Hamiltonian model and low cost of the semiempirical QM approach. In addition, mapping the original QM problem to simple Hamiltonian mechanics provides a practical method to address the locality challenge of molecular representation for HDNNPs. As it was repeatedly highlighted above, the energy of the system in ML models is constructed as a sum of atomic energy contributions, which in turn depend on the local chemical environment up to an empirically determined cutoff radius. Meanwhile, the total energy of the molecular system could be simply expressed in terms of Hamiltonian mechanics with physics-based interpretation.

https://dx.doi.org/10.1021/acs.accounts.0c00868
Acc. Chem. Res. 2021, 54, 1575−1585

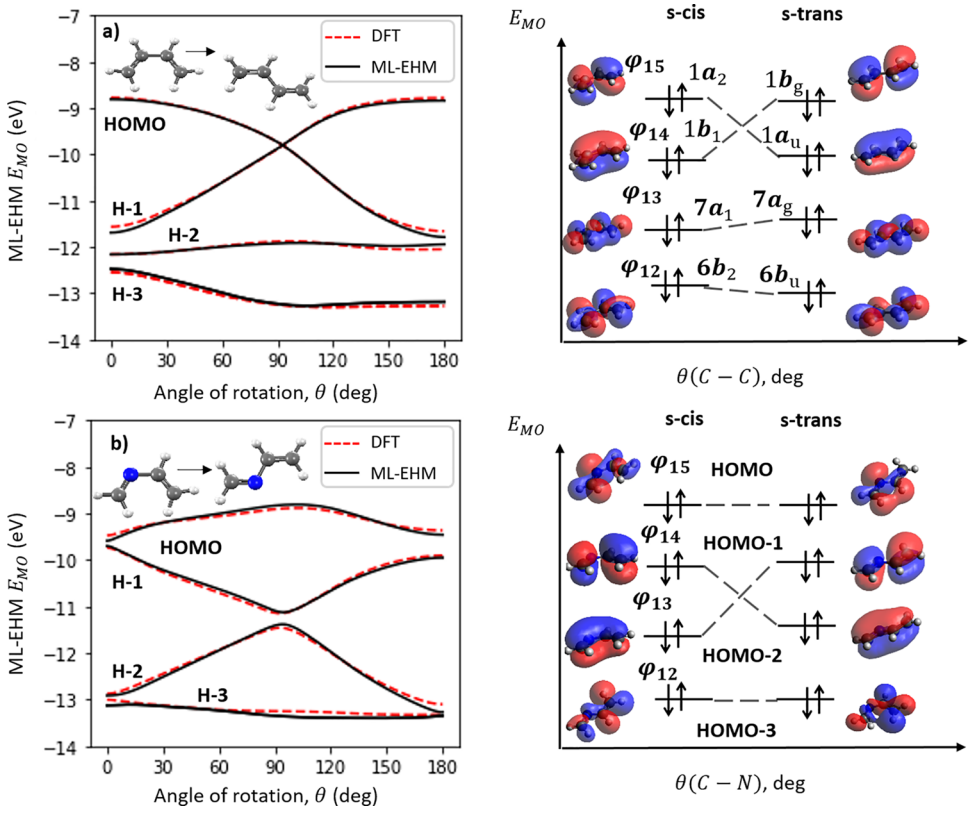


Figure 5. Right: MO energy for the rotation around a central bond in (a) butadiene and (b) aza-butadiene. Left: frontier MO diagrams.

Machine Learned Hückel Hamiltonian

Today, Hückel-like methods cannot be described as mainstream, but they are used as part of the tight-binding algorithms. The Hückel model is still appreciated for its simplicity and interpretability. Our recent work[55] has introduced ML semi-empirical time-binding scheme based on the extended Hückel method (ML-EHM, Figure 4(d)). In the Hückel theory, the definition of Hamiltonian $\boldsymbol{H}$ for the molecular system in a basis set of N atomic functions is given as follows:

$$\boldsymbol{H} = \begin{bmatrix} \alpha_{11} & \frac{1}{2}K^{\ddagger}(\alpha_{22}+\alpha_{11})S_{21} & \cdots & \frac{1}{2}K^{\ddagger}(\alpha_{NN}+\alpha_{11})S_{N1} \\ \frac{1}{2}K^{\ddagger}(\alpha_{11}+\alpha_{22})S_{12} & \alpha_{22} & \cdots & \frac{1}{2}K^{\ddagger}(\alpha_{NN}+\alpha_{22})S_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}K^{\ddagger}(\alpha_{11}+\alpha_{NN})S_{1N} & \frac{1}{2}K^{\ddagger}(\alpha_{22}+\alpha_{NN})S_{2N} & \cdots & \alpha_{NN} \end{bmatrix} \quad (5)$$

The atomic functions are chosen to include *1s* orbitals of the hydrogen atoms and 2s, $2p_x$, $2p_y$, and $2p_z$ orbitals of the carbons and heteroatoms in the molecule. The matrix equation for all the molecular orbitals in the extended Hückel method is given as eigenvalue problem:

$$\boldsymbol{H}C = E\boldsymbol{S}C \quad (6)$$

The square matrix $\boldsymbol{H}$ is defined in (eq 5), and C is the matrix of coefficients for the atomic orbitals. S is the matrix of overlap integrals, and E is the diagonal matrix of orbital energies. In ML-EHM, the elements of the $\boldsymbol{H}$ matrix (eq 5) are assigned using *learnable parameters* α_{NN} and $K^{\ddagger}$ which makes the method a *ML molecular orbital method*. The eq 6 is iteratively solved until the eigenvalues match the orbital energies of a high-level QM method.

The choice of parameters is deliberate since the Coulomb integral α_{NN} is a function of the nuclear charge and the type of orbital, thus it encodes the information about the *local chemical composition* given by the types of chemical species. The Hoffmann vector $K^{\ddagger}$ scales the contributions of the energy of the atomic orbitals and their overlapping to the molecular energy. Thus, vector $K^{\ddagger}$ is bond-dependent[56] and therefore introduces the i^{th} atom in the molecule in terms of its *local structure*. As the atomic chemical composition and the atomic structure depend upon the entire chemical environment, *e.g.*, bond order, conjugation, and hybridization, it is reasonable to approximate those unknown complex functions with NN.

Figure 4d shows that ML-EHM generates the set of parameters $\{K, \alpha\}$, accounting for the AEV of each atom and atom pairs, to create an effective extended Hückel Hamiltonian. ML-EHM is based on the HIP-NN[36] which was trained to predict the $\boldsymbol{H}$ matrix elements based on DFT. It is important to emphasize that, in contrast to the original EHM formulation, ML parameters $\{K, \alpha\}$ are continuously varying with molecular geometry allowing to capture the orbital physics observed in molecule. By learning the orbital behavior with MLIP we could fastly estimate the molecular structure, the energy barriers to rotation around a bond or predict the transition states for reactions.

ML-EHM Case Study for Internal Rotation

As a practical application of ML-EHM model, we demonstrate a behavior of the frontier molecular orbitals upon rotation around a central bond in butadiene and aza-butadiene. These two molecules have the same number of electrons, while aza-butadiene has one s-orbital less. As it can be seen from Figure 5 HOMO and HOMO − 1 in butadiene shift down and become HOMO − 1 and HOMO − 2 in aza-butadiene. Moreover, the orbital crossing is perfectly reflected in Figure 5 as butadiene undergoes a torsional rotation. According to the orbital symmetries:

https://dx.doi.org/10.1021/acs.accounts.0c00868
Acc. Chem. Res. 2021, 54, 1575−1585

cis-butadiene X^1A_1: $(6b_2)(7a_1)(\mathbf{1b_1})(\mathbf{1a_2})$

trans-butadiene X^1A_g: $(6b_u)(7a_g)(\mathbf{1a_u})(\mathbf{1b_g})$

we expect the crossing between HOMO and HOMO − 1. The aza-butadiene molecule lacks inversion symmetry due to the presence of nitrogen. In aza-butadiene, due to the valence orbital shift, HOMO − 1 and HOMO − 2 are predicted to cross. This is in clear contrast with the original EHM, which uses fixed empirical parameters and is thus unable to predict smooth orbital energies change with bond rotations and bond stretching. This example illustrates how ML-EHM harmonizes to capture orbital physics across substantial geometry changes. With the matrix diagonalization, the EHM captures rapid changes in orbital physics, and the learnable diagonal elements provide smooth modulation of the underlying Hamiltonian. In this way, the ML-EHM model learns the underlying physics to reproduce the quantum effects caused by the electronic interactions in the molecular system.

6. OUTLOOK

In this Account, we have presented a perspective on the recent development of machine-learned atomistic potentials and their combination with physics-based approaches. With the rapid improvement of MLIPs, they are promising to change the way molecular simulations are conducted. Modern MLIPs can carry out high-throughput calculations for molecules and materials in millisecond time scales with accuracy approaching high-level QM. They are actively used to find reliable conformational energies for molecules, reparametrize force fields, and simulate protein−ligand interfaces. Recent results also show that the coupling of ML and MM models drastically improves the accuracy of protein−ligand free energy simulations.[57]

Unfortunately, the field of MLIPs is heavily dominated by the benchmark culture. New models are developed with the sole purpose of beating the current state of the art on standard benchmark data sets and improve accuracy for yet another 0.1%. The community should move toward solving real-life applications, addressing the challenges of scale and time, and aiming to reproduce experimental observables.

Given the prospects of artificial intelligence (AI), robotics, and intelligent software, we are currently witnessing a transformation of chemical sciences with data-driven automated discovery methods. As we show in this Account, the current MLIPs research is focused on *physics-aware artificial intelligence* (PhAI), where ML is combined with physical models. The envisioned PhAI of the future will imitate human decision-making by machine intelligence. It will incorporate data-driven and physics-based methods into one ultimate model. If successful, PhAI will revolutionize the way computational methods are developed. In PhAI, multiple simple components are combined into a single system. Consequently, their individual limitations diminish, but new challenges might arise.

AUTHOR INFORMATION

Corresponding Author

Olexandr Isayev − *Department of Chemistry, Mellon College of Science, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States;* orcid.org/0000-0001-7581-8497; Email: olexandr@olexandrisayev.com

Author

Tetiana Zubatiuk − *Department of Chemistry, Mellon College of Science, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.accounts.0c00868

Notes

The authors declare no competing financial interest.

Biographies

Tetiana Zubatiuk obtained her master's degree in chemistry in 2007 and her Ph.D. in 2012 from V.N. Karazin Kharkiv National University, Ukraine. After a postdoctoral stay at Ukraine-American Laboratory of Computational Chemistry, at the Institute for Single Crystals, National Academy of Sciences of Ukraine (2011−2015), she moved to the United States and joined the group of Jerzy Leszczynski at the Interdisciplinary Center for Nanotoxicity at Jackson State University as a research associate (2016−2019). Since 2020, she is a postdoctoral fellow at the Department of Chemistry at Carnegie Mellon University in Olexandr Isayev's research group. Her current research focuses on applying machine learning and quantum chemical simulations to model important electronic properties of molecular materials.

Olexandr Isayev is an assistant professor at the Department of Chemistry, Carnegie Mellon University. He obtained his master's degree in chemistry in 2002 from Oles Honchar Dnipro National University, Ukraine. He received his Ph.D. in computational chemistry under the guidance of Prof. Jerzy Leszczynski at Jackson State University in 2009. He was a postdoctoral research fellow at the Case Western Reserve University and a scientist at the government research laboratory. During 2016−2019, he was a research faculty at UNC Eshelman School of Pharmacy, the University of North Carolina at Chapel Hill. His main research interest is the acceleration of molecular discovery by combining AI, informatics, and molecular modeling.

ACKNOWLEDGMENTS

The authors appreciate all colleagues and coauthors, especially Dr. Justin S. Smith, Dr. Roman Zubatyuk, Dr. Adrian Roitberg, Dr. Kipton Barros, and Dr. Segeri Tretiak, for their invaluable discussions and contributions to the development of MLIPs. The authors acknowledge support from NSF CHE-1802789 and CHE-2041108. This work was performed, in part, at the Center for Integrated Nanotechnologies, an Office of Science User Facility operated for the US Department of Energy (DOE) Office of Science. We also acknowledge the Extreme Science and Engineering Discovery Environment (XSEDE) award CHE200122, which is supported by NSF grant number ACI-1053575. This research is part of the Frontera computing project at the Texas Advanced Computing Center. Frontera is made possible by the National Science Foundation award OAC-1818253. We gratefully acknowledge the support and hardware donation from NVIDIA Corporation and express our special gratitude to Jonathan Lefman. We want to thank the reviewers for their thoughtful comments and efforts towards improving this Account.

ABBREVIATIONS

AEV, atomic environment vector; AFV, atomic feature vector; AIM, atoms-in-molecules layer; AIMNet, atoms-in-molecules neural network; AL, active learning; ANAKIN-ME or ANI, accurate neural network engine for molecular energies; BPSFs,

Behler and Parrinello symmetry functions; HDNNP, high-dimensional neural network potential; HIP-NN, hierarchically interacting particle neural network; IP, interatomic potential; ML, machine learning; ML-EHM, machine learned extended Hückel method; MLIP, machine learned interatomic potential; MTL, multitask learning; NNP, neural network potential; NN, neural network; PES, potential energy surfaces; PhAI, physics-aware artificial intelligence; PLS, partial least-squares; QBC, query by committee; SE, Schrödinger equation; JSSFs, J. Smith symmetry functions; TL, transfer learning

REFERENCES

(1) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8*, 3192−3203.

(2) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching Coupled Cluster Accuracy with a General-Purpose Neural Network Potential through Transfer Learning. *Nat. Commun.* **2019**, *10*, 2903.

(3) Zubatyuk, R.; Smith, J. S.; Leszczynski, J.; Isayev, O. Accurate and Transferable Multitask Prediction of Chemical Properties with an Atoms-in-Molecules Neural Network. *Sci. Adv.* **2019**, *5*, eaav6490.

(4) Zubatyuk, R.; Smith, J.; Nebgen, B. T.; Tretiak, S.; Isayev, O. Teaching a Neural Network to Attach and Detach Electrons from Molecules. *ChemRxiv*, July 28, **2020**. DOI: 10.26434/chemrxiv.12725276.v1.

(5) Feynman, R. P.; Leighton, R. B.; Sands, M. *Six Easy Pieces: Essentials of Physics Explained by Its Most Brilliant Teacher*; Basic Books: New York, 2011.

(6) David Sherrill, C.; Schaefer, H. F. The Configuration Interaction Method: Advances in Highly Correlated Approaches. *Adv. Quantum Chem.* **1999**, *34*, 143−269.

(7) Bishop, R. F. An Overview of Coupled Cluster Theory and Its Applications in Physics. *Theor. Chim. Acta* **1991**, *80*, 95−148.

(8) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864−B871.

(9) Löwdin, P.-O. Nature of Quantum Chemistry. *Int. J. Quantum Chem.* **1967**, *1*, 7−12.

(10) Wold, S.; Sjöström, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109−130.

(11) Hornik, K.; Stinchcombe, M.; White, H. Multilayer Feedforward Networks Are Universal Approximators. *Neural Networks* **1989**, *2*, 359−366.

(12) Tianping Chen; Hong Chen. Universal Approximation to Nonlinear Operators by Neural Networks with Arbitrary Activation Functions and Its Application to Dynamical Systems. *IEEE Trans. Neural Networks* **1995**, *6*, 911−917.

(13) Bishop, C. M. *Neural Networks for Pattern Recognition*; Oxford University Press, Inc.: New York, 1995.

(14) Blank, T. B.; Brown, S. D.; Calhoun, A. W.; Doren, D. J. Neural Network Models of Potential Energy Surfaces. *J. Chem. Phys.* **1995**, *103*, 4129−4137.

(15) Lorenz, S.; Groß, A.; Scheffler, M. Representing High-Dimensional Potential-Energy Surfaces for Reactions at Surfaces by Neural Networks. *Chem. Phys. Lett.* **2004**, *395*, 210−215.

(16) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(17) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.

(18) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.

(19) Balabin, R. M.; Lomakina, E. I. Support Vector Machine Regression (LS-SVM)—an Alternative to Artificial Neural Networks (ANNs) for the Analysis of Quantum Chemistry Data? *Phys. Chem. Chem. Phys.* **2011**, *13*, 11710.

(20) Dral, P. O. Quantum Chemistry in the Age of Machine Learning. *J. Phys. Chem. Lett.* **2020**, *11*, 2336−2347.

(21) Schmidhuber, J. Deep Learning in Neural Networks: An Overview. *Neural Networks* **2015**, *61*, 85−117.

(22) Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. Less Is More: Sampling Chemical Space with Active Learning. *J. Chem. Phys.* **2018**, *148*, 241733.

(23) Bartók, A. P.; Csányi, G. Gaussian Approximation Potentials: A Brief Tutorial Introduction. *Int. J. Quantum Chem.* **2015**, *115*, 1051−1057.

(24) Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M. Machine Learning Unifies the Modeling of Materials and Molecules. *Sci. Adv.* **2017**, *3*, No. e1701816.

(25) Shapeev, A. V. Moment Tensor Potentials: A Class of Systematically Improvable Interatomic Potentials. *Multiscale Model. Simul.* **2016**, *14*, 1153−1173.

(26) Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett.* **2015**, *6*, 2326−2331.

(27) Christensen, A. S.; Bratholm, L. A.; Faber, F. A.; Anatole von Lilienfeld, O. FCHL Revisited: Faster and More Accurate Quantum Machine Learning. *J. Chem. Phys.* **2020**, *152*, 044107.

(28) Behler, J. Perspective: Machine Learning Potentials for Atomistic Simulations. *J. Chem. Phys.* **2016**, *145*, 170901.

(29) Artrith, N.; Morawietz, T.; Behler, J. High-Dimensional Neural-Network Potentials for Multicomponent Systems: Applications to Zinc Oxide. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2011**, *83*, 153101.

(30) Hellström, M.; Behler, J. Neural Network Potentials in Materials Modeling. In *Handbook of Materials Modeling*; Springer International Publishing: Cham, 2020; pp 661−680.

(31) Devereux, C.; Smith, J. S.; Davis, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E. Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. *J. Chem. Theory Comput.* **2020**, *16*, 4192−4202.

(32) Stevenson, J.; Jacobson, L. D.; Zhao, Y.; Wu, C.; Maple, J.; Leswing, K.; Harder, E.; Abel, R. Schrodinger-ANI: An Eight-Element Neural Network Interaction Potential with Greatly Expanded Coverage of Druglike Chemical Space. *ChemRxiv*, December 12, 2019. DOI: 10.26434/chemrxiv.11319860.v1.

(33) Gao, X.; Ramezanghorbani, F.; Isayev, O.; Smith, J. S.; Roitberg, A. E. TorchANI: A Free and Open Source PyTorch-Based Deep Learning Implementation of the ANI Neural Network Potentials. *J. Chem. Inf. Model.* **2020**, *60*, 3408−3415.

(34) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-Chemical Insights from Deep Tensor Neural Networks. *Nat. Commun.* **2017**, *8*, 13890.

(35) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet - A Deep Learning Architecture for Molecules and Materials. *J. Chem. Phys.* **2018**, *148*, 241722.

(36) Lubbers, N.; Smith, J. S.; Barros, K. Hierarchical Modeling of Molecular Energies Using a Deep Neural Network. *J. Chem. Phys.* **2018**, *148*, 241715.

(37) Vapnik, V. N. *The Nature of Statistical Learning Theory*; Springer: New York, 1995.

(38) Grunwald, P. A Tutorial Introduction to the Minimum Description Length Principle. 2004.

(39) Belkin, M.; Hsu, D.; Ma, S.; Mandal, S. Reconciling Modern Machine-Learning Practice and the Classical Bias-Variance Trade-Off. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 15849−15854.

(40) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1, A Data Set of 20 Million Calculated off-Equilibrium Conformations for Organic Molecules. *Sci. Data* **2017**, *4*, 170193.

(41) Blum, L. C.; Reymond, J.-L. 970 Million Druglike Small Molecules for Virtual Screening in the Chemical Universe Database GDB-13. *J. Am. Chem. Soc.* **2009**, *131*, 8732−8733.

(42) Taylor, M. E.; Stone, P. Transfer Learning for Reinforcement Learning Domains: A Survey. *J. Mach. Learn. Res.* **2009**, *10*, 1633−1685.

https://dx.doi.org/10.1021/acs.accounts.0c00868
Acc. Chem. Res. 2021, 54, 1575−1585

(43) Pan, S. J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345−1359.
(44) Zhang, L.; Lin, D.-Y.; Wang, H.; Car, R.; E, W. Active Learning of Uniformly Accurate Interatomic Potentials for Materials Simulation. *Phys. Rev. Mater.* **2019**, *3*, 023804.
(45) Loeffler, T. D.; Manna, S.; Patra, T. K.; Chan, H.; Narayanan, B.; Sankaranarayanan, S. Active Learning A Neural Network Model For Gold Clusters & Bulk From Sparse First Principles Training Data. *ChemCatChem* **2020**, *12*, 4796−4806.
(46) Freund, Y.; Seung, H. S.; Shamir, E.; Tishby, N. Selective Sampling Using the Query by Committee Algorithm. *Mach. Learn.* **1997**, *28*, 133−168.
(47) Smith, J. S.; Zubatyuk, R.; Nebgen, B.; Lubbers, N.; Barros, K.; Roitberg, A. E.; Isayev, O.; Tretiak, S. The ANI-1ccx and ANI-1x Data Sets, Coupled-Cluster and Density Functional Theory Properties for Molecules. *Sci. Data* **2020**, *7*, 134.
(48) Bartlett, R. J.; Musiał, M. Coupled-Cluster Theory in Quantum Chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291−352.
(49) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087−2096.
(50) Li, H.; Collins, C.; Tanha, M.; Gordon, G. J.; Yaron, D. J. A Density Functional Tight Binding Layer for Deep Learning of Chemical Hamiltonians. *J. Chem. Theory Comput.* **2018**, *14*, 5764−5776.
(51) Glick, Z. L.; Metcalf, D. P.; Koutsoukas, A.; Spronk, S. A.; Cheney, D. L.; Sherrill, C. D. AP-Net: An Atomic-Pairwise Neural Network for Smooth and Transferable Interaction Potentials. *J. Chem. Phys.* **2020**, *153*, 044112.
(52) Qiao, Z.; Welborn, M.; Anandkumar, A.; Manby, F. R.; Miller, T. F. OrbNet: Deep Learning for Quantum Chemistry Using Symmetry-Adapted Atomic-Orbital Features. *J. Chem. Phys.* **2020**, *153*, 124111.
(53) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Halpen, J., Green, M. L. H., Eds.; Clarendon Press: Oxford, 1990.
(54) Ghasemi, S. A.; Hofstetter, A.; Saha, S.; Goedecker, S. Interatomic Potentials for Ionic Systems with Density Functional Accuracy Based on Charge Densities Obtained by a Neural Network. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2015**, *92*, 045131.
(55) Zubatyuk, T.; Nebgen, B.; Lubbers, N.; Smith, J. S.; Zubatyuk, R.; Zhou, G.; Koh, C.; Barros, K.; Isayev, O.; Tretiak, S. Machine Learned Hückel Theory: Interfacing Physics and Deep Neural Networks. *arXiv (Disordered Systems and Neural Networks)*, September 27, **2019**.
(56) Rincón, L.; Hasmy, A.; Gonzalez, C. A.; Almeida, R. Extended Hückel Tight-Binding Approach to Electronic Excitations. *J. Chem. Phys.* **2008**, *129*, 044107.
(57) Rufa, D. A.; Macdonald, H. E. B.; Fass, J.; Wieder, M.; Grinaway, P. B.; Roitberg, A. E.; Isayev, O.; Chodera, J. D. Towards Chemical Accuracy for Alchemical Free Energy Calculations with Hybrid Physics-Based Machine Learning/Molecular Mechanics Potentials. *BioRxiv*, July 30, **2020**. DOI: 10.1101/2020.07.29.227959.

https://dx.doi.org/10.1021/acs.accounts.0c00868
Acc. Chem. Res. 2021, 54, 1575−1585