

# A Deep Dive into Machine Learning Density Functional Theory for Materials Science and Chemistry

L. Fiedler,<sup>1,2,\*</sup> K. Shah,<sup>1,2,†</sup> M. Bussmann,<sup>1,2,‡</sup> and A. Cangi<sup>1,2,§</sup>

<sup>1</sup>*Center for Advanced Systems Understanding (CASUS), D-02826 Görlitz, Germany*

<sup>2</sup>*Helmholtz-Zentrum Dresden-Rossendorf, D-01328 Dresden, Germany*

(Dated: December 10, 2021)

With the growth of computational resources, the scope of electronic structure simulations has increased greatly. Artificial intelligence and robust data analysis hold the promise to accelerate large-scale simulations and their analysis to hitherto unattainable scales. Machine learning is a rapidly growing field for the processing of such complex datasets. It has recently gained traction in the domain of electronic structure simulations, where density functional theory takes the prominent role of the most widely used electronic structure method. Thus, DFT calculations represent one of the largest loads on academic high-performance computing systems across the world. Accelerating these with machine learning can reduce the resources required and enables simulations of larger systems. Hence, the combination of density functional theory and machine learning has the potential to rapidly advance electronic structure applications such as in-silico materials discovery and the search for new chemical reaction pathways. We provide the theoretical background of both density functional theory and machine learning on a generally accessible level. This serves as the basis of our comprehensive review including research articles up to December 2020 in chemistry and materials science that employ machine-learning techniques. In our analysis, we categorize the body of research into main threads and extract impactful results. We conclude our review with an outlook on exciting research directions in terms of a citation analysis.

## CONTENTS

I. Introduction	1	IV. Discussion	12
II. Theoretical Background	2	A. Citation Analysis	12
A. Electronic Structure Theory	2	B. Conclusions and Outlook	14
1. Exact framework	2	Acknowledgments	14
2. Practical methods	3	References	16
B. Machine learning	4		
1. Introduction	4		
2. Supervised Learning Models	4		
III. Machine-learning in Materials Science and Chemistry	5		
A. Property mappings	5		
1. Chemical Reactions	6		
2. Discovery of Novel, Stable Materials	7		
3. Electric Properties and Photovoltaics	7		
4. Elastic and Structural Properties	8		
5. Quantum Chemical Information	8		
6. Magnetic, Thermal, and Energetic Properties	9		
B. Interatomic Potentials	9		
1. Neural Network Potentials	9		
2. Gaussian Approximation Potentials	9		
3. Moment Tensor Potentials	10		
4. Spectral Neighborhood Analysis and Other Potentials	10		
5. Technical Aspects for the Constructing of Potentials	10		
C. Learning Electronic Structures	10		
D. Other Approaches	11		
E. Technical Aspects	12		

## I. INTRODUCTION

Electronic structure theory calculations enable the understanding of matter on the quantum level and complement experimental studies both in material science and chemistry. They play a central role in solving pressing scientific and technological problems. Large-scale electronic structure simulations have in turn been enabled by the advent of modern, high-performance computational resources. Yet, the ever increasing demand for accurate first-principles data renders even the most efficient simulation codes infeasible. On the other hand, the use of data-driven machine-learning methods has grown rapidly across a number of research fields. Such methods are increasingly gaining importance, as they are utilized to accelerate, replace, or improve traditional electronic structure theory methods.

The computational backbone of electronic structure workflows is often density functional theory (DFT). While DFT provides a convenient balance between computational cost and accuracy, enormous speed-ups can be achieved when combined with machine learning (ML). In the following, we provide an overview over recent research efforts that accelerate materials science with the aid of ML and tackle different aspects of the combined

ML-DFT workflow outlined schematically in Fig. 1. To that end, we compiled a database of over 300 research articles including the most recent publications up to December 2020. In Sec. II we introduce the necessary formalism and methodologies. Sec. III is the centerpiece of our manuscript, where we provide an in-depth review of the compiled publications according to the most relevant research lines and highlight the most prominent applications. Finally, in Sec. IV, we carry out a citation analysis and point to trends of future research.

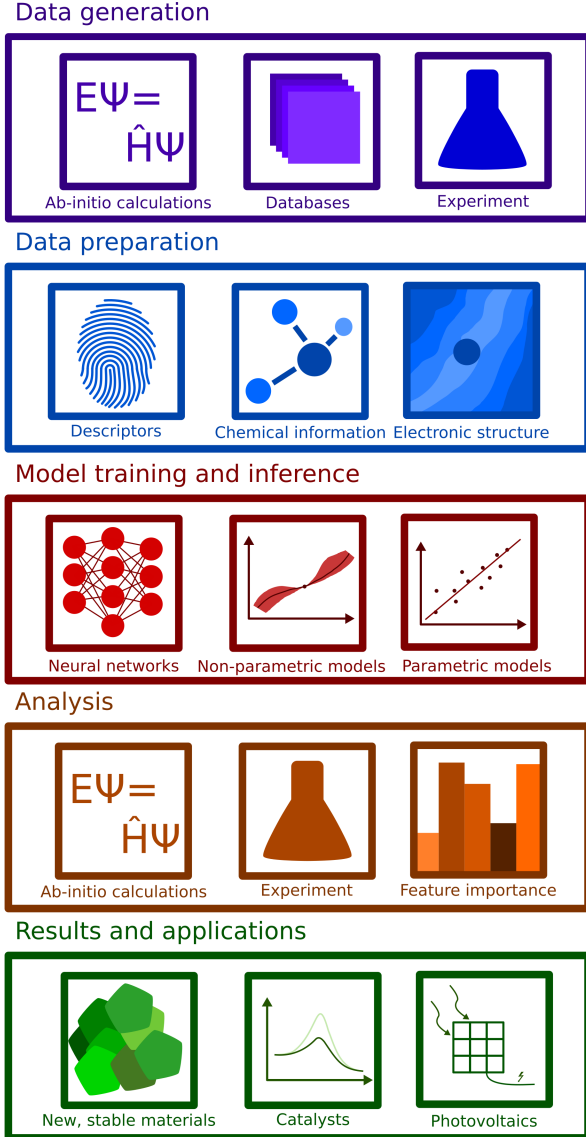


Figure 1. Schematic overview of the general workflow combining ML and DFT for material science or chemistry applications.

## II. THEORETICAL BACKGROUND

### A. Electronic Structure Theory

Developing practical methods for calculating the electronic structure of materials is a topic of active research. In the following, we provide a short overview of the basic formalism with a focus on DFT. We work within atomic units throughout, where  $\hbar = m_e = e^2 = 1$ , such that energies are expressed in Hartrees and length in Bohr radii.

#### 1. Exact framework

The most common theoretical framework for describing thermodynamic materials properties from first principles is within the scope of non-relativistic quantum mechanics. It provides a systematic representation of  $N_e$  electrons with collective coordinates  $\mathbf{r} = \{\mathbf{r}_1, \dots, \mathbf{r}_{N_e}\}$  that are coupled to  $N_i$  ions with collective coordinates  $\mathbf{R} = \{\mathbf{R}_1, \dots, \mathbf{R}_{N_i}\}$ , where  $\mathbf{r}_j \in \mathbb{R}^3$  refers to the position of the  $j$ th electron, while  $\mathbf{R}_\alpha \in \mathbb{R}^3$  denotes the position of the  $\alpha$ th ion. Note that for sake of brevity we do not take spins into account. The physics in this framework is governed by the Hamiltonian [1]

$$\hat{H}^{BO}(\mathbf{r}; \mathbf{R}) = \hat{T}^e(\mathbf{r}) + \hat{V}^{ee}(\mathbf{r}) + \hat{V}^{ei}(\mathbf{r}; \mathbf{R}) + E^{ii}(\mathbf{R}), \quad (1)$$

where  $\hat{T}^e$  refers to the kinetic energy of the electrons, while the Coulomb operators,  $\hat{V}^{ei}$  and  $\hat{V}^{ee}$  account for the electron-ion and electron-electron interactions. In this equation, the Born-Oppenheimer approximation [2] is employed to reduce the computational complexity of the underlying problem, i.e., the coupled electron-ion problem is separated and the ion-ion interaction amounts to a constant shift in energy  $E^{ii}$ . This approximation is valid as long as the motion of the ions happens on a much larger time scale than the motion of the electrons. Employing this Hamiltonian in the Schrödinger equation

$$\hat{H}^{BO}(\mathbf{r}; \mathbf{R})\Psi(\mathbf{r}; \mathbf{R}) = E\Psi(\mathbf{r}; \mathbf{R}), \quad (2)$$

with the many-electron wavefunction  $\Psi$ , which depends on the electronic coordinates and only parametrically on the ionic positions, provides the basis for the electronic structure of matter. It therefore enables quantitative predictions of physical phenomena in materials science. Its solution yields the electronic ground state based on which a vast amount of materials properties can be computed. Most importantly, the electronic ground-state energy can be identified as

$$E = \langle \Psi | \hat{H}^{BO} | \Psi \rangle. \quad (3)$$

Other important quantities include molecular or crystal structures [3], charge densities [4], cohesive energies [5], elastic properties [6], vibrational properties [7],

magnetic properties [8, 9], dielectric susceptibilities [10], magnetic susceptibilities [11], phase transitions [12], HOMO/LUMO energies [13–15], bond dissociations [16], enthalpies of formation [16], ionization potentials [17], electron affinities [17], band gaps [18, 19], and the equation of state [20].

## 2. Practical methods

The multitude of practical electronic structure methods employ different approximations to overcome the underlying complexity of the problem. In the following concise overview of these methods, we always work within the Born-Oppenheimer approximation and suppress denoting the parametric dependence on  $\mathbf{R}$ .

**Density Functional Theory** DFT is a very popular method for carrying out electronic structure calculations, as it balances acceptable accuracy with reasonable computational cost. The central quantity is the electronic density  $n(\mathbf{r})$ . The Hohenberg-Kohn theorems [21] guarantee a one-to-one correspondence between the electronic density and the external potential, e.g., the electron-ion potential  $\hat{V}^{ei}(\mathbf{r})$ . This means every property of interest can be determined as a functional of the density.

Practical DFT calculations usually rely on the Kohn-Sham ansatz [22]. Here, the system of interacting electrons is replaced by an auxiliary system of non-interacting electrons constrained to reproduce the electronic density of the interacting system. This is achieved in terms of the Kohn-Sham equations

$$\left(-\frac{1}{2}\nabla^2 + v_s(\mathbf{r})\right)\phi_j(\mathbf{r}) = \epsilon_j\phi_j(\mathbf{r}), \quad (4)$$

a set of  $N_e$  one-particle Schrödinger equations, where  $v_s^r(\mathbf{r})$  denotes the Kohn-Sham potential. The self-consistently calculated Kohn-Sham potential includes both the electron-ion interaction and the electron-electron interaction, the latter in terms of a mean-field description. The electronic density is obtained from the Kohn-Sham orbitals  $\phi_j(\mathbf{r})$  as

$$n(\mathbf{r}) = \sum_j |\phi_j(\mathbf{r})|^2. \quad (5)$$

The total energy functional is expressed within the Kohn-Sham formalism as

$$E_{\text{total}} = T_s[n] + E_H[n] + E_{xc}[n] + E^{ei}[n] + E^{ii}. \quad (6)$$

In Eq. (6),  $E_H[n]$  denotes the Hartree energy, i.e., the energy contribution from the electrostatic interaction of the density with itself,  $E^{ii}$  the energy contribution from the ion-ion interaction, and  $E^{ei}[n]$  the energy contribution from the electron-ion interaction.  $T_s[n]$  represents

the kinetic energy of the Kohn-Sham system. The final energetic contributions are the exchange and correlation energies  $E_{xc}[n]$ . Given that the form of  $E_{xc}[n]$  was known, it is ensured that DFT calculations yield *in principle* the exact result. In practice however,  $E_{xc}[n]$  has to be approximated. The fundamental approximation to  $E_{xc}[n]$  is the local density approximation (LDA) [22, 23], which by incorporation of the density gradient or further constraints can be expanded to (meta/hybrid-) generalized gradient approximations ((meta/hybrid-)GGA, e.g., PBE [24], SCAN [25], or B3LYP [26, 27]). Notable extensions of DFT are time-dependent DFT (TD-DFT) [28] (representing the time-dependent Schrödinger equation in terms of the single-particle picture) and finite-temperature DFT (FT-DFT), which extends DFT to temperatures  $\tau > 0$  [29–32].

**Beyond DFT** In contrast to DFT, wavefunction methods center on the many-particle wavefunction  $\Psi(\mathbf{r})$ . The simplest method is Hartree-Fock which constructs a single Slater determinant, but neglects electronic correlation by construction [33]. Thus, several post-Hartree-Fock methods have been developed that incorporate electronic correlation explicitly, such as the coupled-cluster method (CC / CCSD / CCST) [34–39], configuration interaction (CI) [40–50], and Møller-Plesset perturbation theory (MP) [51]. These methods generally outperform DFT calculations in terms of accuracy [52], but do so at a much higher computational cost.

Electronic structure calculations can also be carried out using Monte-Carlo methods, i.e., methods that use repeated, randomized sampling, allowing for generally trivial parallelizability and a formally exact quantum mechanical treatment of the electronic structure problem. Examples of such methods are variational Monte Carlo [53], fixed-node diffusion Monte Carlo [54], and path-integral Monte Carlo [55, 56], useful for  $\tau > 0$ .

**Coupling DFT to Molecular Dynamics** Investigating the electronic structure of a material is often only a part of a bigger systematic study. Simulations on larger time and length scales are often realized by combining classical mechanical molecular dynamics (MD) of the ions with electronic structure calculations. To this end, one calculates an interatomic potential (IAP), which determines how the ions in the system interact. Such IAPs can also be replaced with reasonably fast quantum mechanical calculations (DFT-MD) or be constructed using electronic structure data for later use. Traditional IAPs are the Lennard-Jones potential [57], the embedded atom model (EAM) [58] or other many-body potentials [58–66]. Recently, ML has given rise to a new class of IAPs that are constructed using data-driven algorithms (see Sec. II B). Notable examples include ANI-1 [67], GAP [68], BLAST [69], HIP-NN [70], SchNet [71], SNAP [72], DPMD [73], and the AGNI ML force fields [74]. The resulting IAPs come well within chemical accuracy while retaining an often negligible computational cost.

## B. Machine learning

ML is a form of computational pattern recognition which has had profound impact on multiple fields over the past decade. Advances in GPU hardware coupled with the development of high-level frameworks has led to many different applications in various fields. These applications include autonomous driving [75], high-energy particle-physics data analysis [76], and materials discovery [77]. In this section, a brief background on ML is given, as well as some specifics relevant to most common models in materials science applications.

### 1. Introduction

While there is no standard definition of ML, it is commonly defined as the study of computer algorithms that improve automatically through experience [78]. One usual way to divide ML models is into supervised models (models that make a prediction with respect to a ground truth), unsupervised models (models that operate on data without an underlying ground truth, and perform, e.g., dimensionality reduction) and reinforcement learning (models that explore complex environments). ML models can be further divided into parametric and non-parametric models. When the model involves a finite set of learning parameters  $\mathbf{W}$ , the model is said to be parametric. These parameters are not set explicitly and learned iteratively through the learning process. Models without a finite set of parameters are non-parametric. In non-parametric models, the set of weights  $\mathbf{W}$  is generally infinite.

Some examples of ML models include *supervised parametric models* (linear regression, neural networks), *unsupervised parametric models* (Gaussian mixture models, generative adversarial networks [79]), *supervised non-parametric models* (Gaussian Process Regression (GPR) [80], Decision Trees (DT) [81], Support Vector Machines (SVM) [82]), and *unsupervised non-parametric models* (k-means algorithm) [83].

Learning models can also be combined into ensembles. Using multiple learning models reduces overfitting and improves generalisability. Examples of ensemble learning techniques are Random Forests (Regression) (RF(R)) [84] and Gradient Boosting Regression (GB(R)) [85].

Supervised parametric learning problems generally follow a common structure. Given a dataset  $D$  consisting of  $N$  samples  $\mathbf{X}$  and labels  $\mathbf{y}$ , a model  $M : \mathbf{X} \mapsto \mathbf{y}$  with parameters  $\mathbf{W}$  is to be fitted such that the cost or loss function  $J$  is minimised. For this fitting only a subset of the dataset is used (training data), and once the model is sufficiently trained, its performance is analyzed on a separate subset of data (testing data). If the model performance is satisfactory, it can be used to predict  $\mathbf{y}$  on input data for which it is unknown (inference). Care has to be taken during training so as to avoid over- or underfitting a model, i.e., introducing a high error on unseen

data or not capturing the complexity of the underlying function, respectively. The optimization of the model is often done by gradient descent algorithms [86] in an iterative fashion. Lastly, one has to consider so called hyperparameters, i.e., parameters that describe the model design and are set before model optimization (e.g., characterization of the training process, kernel or activation functions). Their choice is crucial for reasonable model performance.

### 2. Supervised Learning Models

A brief description of commonly used supervised learning models in materials science is given in the following.

**Linear regression** Linear regression [87] is one of the most used algorithms for regression tasks. Given a dataset  $(\mathbf{X}, \mathbf{y})$  consisting of labeled data where  $(\mathbf{x}^{(i)}, y^{*(i)})$ ,  $i = 1, \dots, N$  with  $\mathbf{x}^{(i)} \in \mathbf{X}$  and  $y^{*(i)} \in \mathbf{y}$ , regression is used to determine the relationship between dependent variables  $y^{*(i)}$  and independent variables  $\mathbf{x}^{(i)} = \{x_1^{(i)}, x_2^{(i)}, x_3^{(i)}, \dots, x_d^{(i)}\}$ , where  $d$  is the dimension of the data. Given an input  $\mathbf{x}^{(i)}$ , the prediction  $y^{(i)}$  is defined as

$$y^{(i)} = \mathbf{W}^T \mathbf{x}^{(i)}, \quad (7)$$

where  $\mathbf{W}$  are the learnt parameters of the model, determined by minimizing a least squares cost function. While there exists an analytical solution for the optimal value of parameters  $\mathbf{W}^*$ , the computational complexity for calculating  $\mathbf{W}^*$  is  $\mathcal{O}(N^3)$  which becomes intractable for large data sets with high dimensions. If the number and range of the parameters is ill-suited for the dataset, the model might reach a sub-optimal state due to overfitting or underfitting. To overcome this, a regularisation term is added to the cost function; depending on the choice of this cost function one arrives at methods such as the Least Absolute Shrinkage and Selection Operator (LASSO) [88] or ridge regression [89]. Sure Independence Screening and Sparsifying Operators (SISSO) [90] is an extension to LASSO with better performance for feature selection. Kernel Ridge Regression (KRR) [91] is a non-parametric realisation of ridge regression. Instead of learned parameters  $\mathbf{W}$ , a kernel function measuring the distance between different data points is used. Linear regression models are not well suited for classification problems. When classification is needed, logistic regression [92] can be applied. Here, the input data is mapped to discrete labels by applying a sigmoid function on the linear regression model.

**Neural networks** Neural networks (NNs) are a promising approach to regression due to the *universal approximation theorem* [93] that shows that any function can be modeled by a sufficiently sophisticated NN. NNs have been widely applied to problems in different



domains including image classification [94] or biomedical engineering [95]. However, neural networks also have some challenges in determining their hyperparameters and training routines that avoid overfitting.

*a. Introduction:* A neural network is essentially a set of nested linear regression functions with non-linear activation functions. The base unit of a neural network is a perceptron or artificial neuron [96]. For data point  $\mathbf{x}^{(i)}$ , the perceptron is defined as

$$f^{(i)} = \sigma(\mathbf{W}\mathbf{x}^{(i)} + b), \quad (8)$$

where  $\sigma$  is a non-linear activation function,  $\mathbf{W}$  consists of the weights, and  $b$  denotes the bias. When nested together in layers, the model is known as multilayer perceptron or NN. The output of each layer is input to the subsequent layer. The first layer, that accepts the input samples, is called the input layer, and the final layer is called the output layer. The layers in between are referred to as hidden layers. A layer is defined by the number of its neurons which is often denoted as width. A NN is defined by the number and structure of hidden layers. A network with a large number of hidden layers is called a deep neural network (DNN). This gives rise to the term deep learning [97]. The design of a NN depends on the learning task. To successfully employ NNs, one has to make choices as to their activation functions [98] (e.g., Tanh and ReLU [99]), loss function and optimization scheme (standard gradient descent can be replaced by stochastic gradient descent [100] or the ADAM algorithm [101]).

*b. Architectures:* The operations performed by the neurons can be configured for each layer. The following are some of the most common architectures:

*Fully Connected Neural Network (FCN) / Feed-Forward Neural networks:* Each neuron in a layer of FCNs is connected to every neuron in the next layer [102].

*Convolutional Neural Network (CNN):* CNNs are useful for tasks with 2D or 3D data such as image classification [103].

*Graph Neural Network (GNN):* Instead of a vector input, GNNs [104] can be used with data points in the form of arbitrary graph structures, useful for tasks like many-body problems in cosmology [105].

*Recurrent Neural Network (RNN):* RNNs are used for tasks that involve sequential or temporal evolution such as language translation [106].

*Physics Informed Neural Network (PINN):* PINNs are used to solve partial differential equations. This is done by incorporating initial and boundary conditions in the loss term [107].

NNs are effective at modelling complex non-linear functions and can be generalised well to new data [108].

Shortcomings of NNs are that large training datasets are required for practically useful results, and that the black-box models lack explainability and interpretability [109].

**Gaussian Process Regression** Gaussian Processes (GPs) or Gaussian Process Regression (GPR) is a powerful non-parametric supervised learning technique with robust uncertainty quantification. It can be thought of as a distribution over functions. By constraining the probability distribution at training points, a learned function can be drawn over the test points. Instead of learning any learning parameters  $\mathbf{W}$ , inference is done by constructing the appropriate covariance matrix and drawing functions from it. Such a covariance matrix is defined by using a kernel function  $k$ , the choice of which impacts the usefulness of the model. With appropriate kernels, GPs are equivalent to other ML methods such as neural networks [110]. GPs provide in-built uncertainty quantification and are effective for small datasets. However, the inference step involves matrix inversion which scales as  $\mathcal{O}(N^3)$ . This makes GPs computationally expensive for larger datasets. Approximate inference to increase the scalability of GPs is an active area of research [111].

### III. MACHINE-LEARNING IN MATERIALS SCIENCE AND CHEMISTRY

Electronic structure methods (Sec. II A) suffer from their unfavorable computational scaling, as even the most efficient DFT implementations enable calculations with only few thousand atoms [112]. Increasing computational power can only alleviate this problem to a certain degree. Data-driven workflows (Sec. II B) offer speedups by augmenting or replacing standard DFT calculations.

In the following, we provide an overview over the most recent research developments combining ML and DFT aimed at materials science and chemistry applications. We have grouped them into five main categories based on the type of approach they employ and we highlight the most representative articles. With this grouping, we identify the overall ML-DFT workflow visualized in Fig. 2. By means of this diagram we link the individual categories presented in the following.

#### A. Property mappings

Perhaps the most intuitive approach to employing ML techniques in DFT is by learning distinct physical or chemical properties from data sets. While accessing these through DFT calculations is computationally cheap when compared to other methods of comparable accuracy, the computational cost can often be restrictive when investigating large systems. Likewise, such investigations are often focused on a specific property relevant to a given application rather than on a fundamental understanding of the electronic structure. Hence a feasible approach is

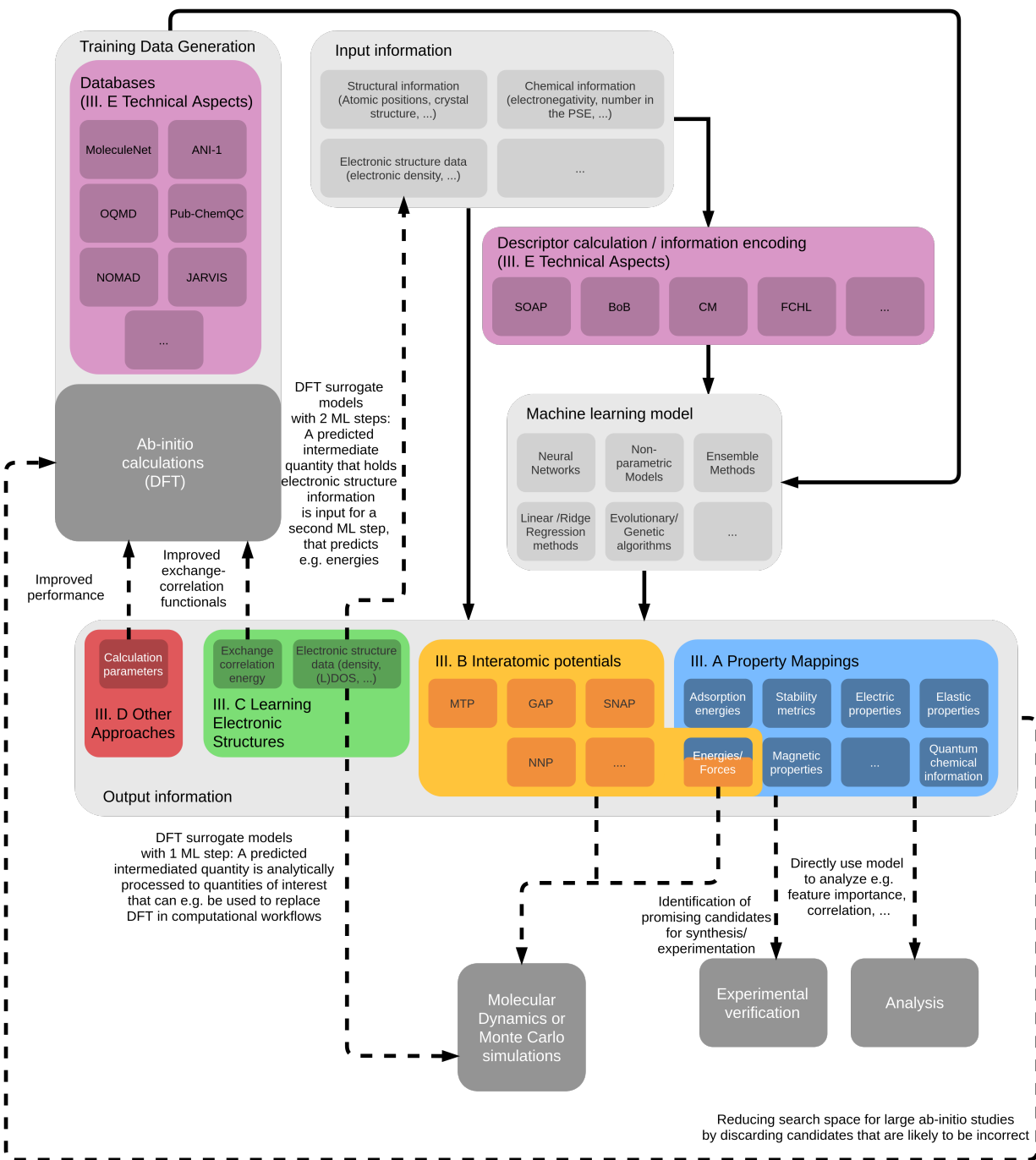


Figure 2. Comprehensive map of workflows combining ML and DFT in current practice. The map is color-coded: the colors highlight the five main categories we have identified in our comprehensive literature survey; the colors also correspond to those used in Fig. 3 which illustrate the results of our citation analysis. More specifically, blue corresponds to Sec. III A, orange to Sec. III B, green to Sec. III C, red to Sec. III D, and purple to Sec. III E.

to train ML models to replicate a mapping of the form  $\mathbf{R} \rightarrow p$  for a relevant property  $p$  that one would initially obtain from a DFT calculation or from post-processing. In the following, an overview of such approaches is given, grouped by their respective application domain. Noteworthy high-impact results are discussed in detail.

### 1. Chemical Reactions

The demand to both reduce and capture carbon emissions motivates improving the efficiency of chemical reactions employed in industrial settings and searching for novel reactive pathways. In the context of computational

materials science and chemistry this usually translates to identifying catalysts for reactions of interest (for instance  $\text{CO}_2$  reduction) and their respective activity and selectivity. The computational demand of such studies can be drastically decreased by employing ML, and a wide range of such investigations has been performed [113–160].

A typical target quantity is the adsorption energy of a molecule on a catalyst. In Ref. [128], the  $\text{CO}_2$  reduction is analyzed where its efficiency is highly dependent on the adsorption properties of CO and H on the chosen catalyst. The catalyst was a high entropy alloy (HEA) consisting of at least five elements [161]. The configuration space of possible potential surface adsorption sites is vast. A GPR model built on a subset of those enables the prediction of adsorption energies on the entire configuration space. In this manner, a candidate catalyst for this reaction was identified, which was then also studied experimentally [162]. Similarly, Ref. [122] also uses GPR to predict activation barriers of dihydrogen on metal complexes. There are various approaches of using ML models to facilitate the design or discovery of catalysts by performing such high-throughput searches. Other notable examples include the  $\text{N}_2$  electroreduction on transition metal alloys using GNNs [130] and the combination of linear models based on experimental data and RFR/GBR models on ab-initio data to directly predict activities and selectivities for ethanol reforming on bi-atom catalysts [117].

Ref. [114], which is concerned with the  $\text{O}_2$  reduction on bi-atom catalysts, follows a slightly different approach. Here, ML is used to unveil correlations in the underlying DFT data. By training a RF regressor to correctly predict adsorption energies from a number of chemical descriptors, the correlation between the adsorption energy and these descriptors is evaluated. Using ML to identify the relative importance of individual descriptors, i.e., catalyst design features, is a technique used for a variety of reaction/catalyst combinations. These include  $\text{CO}_2$  capture on MOFs [139],  $\text{O}_2$  reduction on single atom catalysts [134], olefin oligomerization on Cr based catalysts [154], and the hydrogen evolution reaction (HER) [141].

## 2. Discovery of Novel, Stable Materials

The discovery of novel materials is at the very core of computational materials science, as computational methods reduce the cost of exploring large configuration spaces drastically compared to experimental methods. The ever increasing complexity and size of these configuration spaces makes their exploration an ideal application for ML methods that can be employed to accelerate such investigations [163–196].

Such studies usually sample a certain space of possible candidate structures using some measure of stability, and possibly, specific values for properties of interest. Ref. [177] uses DNNs and a minimal set of chemically motivated descriptors (electronegativity and ionic

radii) to analyze a range of mixed-series crystals and perovskites. The formation energy is then predicted for these materials within chemical accuracy but at negligible computational cost. This enables rapid identification of stable crystal structures which can then be synthesized in the lab. A similar approach is employed in Ref. [188], where a class of multicomponent Ti-alloys are investigated. It extends the previous study by predicting elastic properties, which are important for aerospace applications. The combination of stability metrics with application-centered properties is also demonstrated in Ref. [189], where a SISSO model is trained to predict formation enthalpies of 2D materials to assess their thermodynamic stability. The proposed application domain of this method is the identification of materials suitable for photoelectrocatalytic water splitting. Here, the band gap is the primary metric. While the presented model is not capable of predicting the band gap, it significantly speeds up the search for structures by reducing the search space by more than a half. A similar example of screening a chemical for thermodynamic stability is Ref. [170], where hybrid organic-inorganic perovskites (HOIP) for photovoltaics are at the center of attention.

A somewhat unconventional approach is followed in Ref. [176] where a CNN is used. The CNN is directly applied to the periodic system of elements. To this end, a given structural blueprint is represented as a matrix that takes the form of the periodic table (periodic table representation, PTR). The entries of that matrix reflect the proportion of the elements present in the material. The CNN then predicts formation energies and lattice parameters from these input matrices. It is important to note that these input matrices not only carry information on the proportionalities of the involved elements as an arbitrary list would, but also information on chemical relationships in the form of positions of elements in the periodic table.

## 3. Electric Properties and Photovoltaics

Driven by the effort towards a sustainable future, another important application domain is the search for materials useful in electronic applications such as batteries for electric cars, efficient photovoltaics, and semiconductors for microelectronics. While the underlying problem is similar to before, i.e., finding suitable materials from large configuration spaces, here the constraints on suitability are based on electrical properties such as the band gap or the photoelectric conversion coefficient (PCE). ML has been employed in a number of studies [193, 197–224].

A lot of studies are directed towards photovoltaic materials, for which the PCE, i.e., the percentage of light transferred to electrical energy by a photovoltaic material, is one of the most important metrics. In Ref. [198], Bayesian-Regularized NNs were trained on DFT data and used to predict the PCE for organic photovoltaic materials (OPVs). A prediction accuracy of  $\pm 0.5\%$  was

achieved, allowing for efficient screening of promising materials. The basis for these models are simple, chemical descriptors. A similar strategy has been employed in Ref. [208], where simple structural and electronic similarity features served as inputs for KRR models, again for OPVs. The investigation of photovoltaic materials is not limited to PCE models.

Another important target quantity for ML approaches is the band gap, as demonstrated in Ref. [197], where the vast chemical space of HOIP is sampled, similar to the aforementioned Ref. [170], however with an emphasis on the electrical properties of candidate materials. Using a small set chemical descriptors band gaps computed using DFT were learned and predicted by GBR. The results of Ref. [170] were incorporated in this investigation by means of progressive ML [225], in order to enrich the underlying database of the ML model. Yet, such models suffer from the underestimation of band gaps by usual DFT functionals, an issue addressed in Ref. [212]. The established way of overcoming these inaccuracies inherent to DFT are by using many-body theories, such as the GW approximation, which are computationally very costly. By using GPR models, the corrections applied by the GW approximation [226–228] to band gaps obtained from DFT can be evaluated in a fraction of the time and independently from DFT, since the GPR models are only directly dependent on chemical descriptors.

Other electrical properties of interest addressed in ML-based studies are piezoelectric and dielectric responses of inorganic materials [223], Li ion conductivity for Li ion batteries [209] or impurity levels of semiconductors [202].

#### 4. Elastic and Structural Properties

The design and discovery of materials with specified structural properties constitutes another important research thrust where DFT data and property mappings in terms of ML play a central role. These investigations [229–240] are similar in nature to the ones discussed in Sec. III A 2, however, with a stronger emphasis on property selection.

Typical quantities of interest in this context are elastic constants or Young’s modulus. NNs and RFRs were used in Ref. [229] to predict bulk, shear, and Young’s modulus for ternary Ti-Nb-Zr alloys in order to accelerate an otherwise costly high-throughput DFT search for low-modulus alloys, similar to Ref. [188]. Identification of such alloys is in demand from the biomedical industry, and due to the ML assisted study, a promising candidate could be found.

Another important property often considered in high-throughput calculations is the elastic constant. This is addressed by Ref. [236] and [238] in terms of NN models to provide both direct predictions of elastic constants.

ML models have been also used to unveil relationships between structure and properties of materials, such as in Ref. [234] by means of a GPR based analysis of molecu-

lar crystals. Here, similarities between crystal structures were classified along with the prediction of lattice energies where the resulting structural landscapes shed light on molecular crystallization.

#### 5. Quantum Chemical Information

While there is no precise distinction between quantum chemistry and materials science, these closely related fields can be distinguished in terms of either the systems considered or the properties under investigation. While the articles and groupings of property mappings discussed above are drawn from both quantum chemical and materials science applications, ML approaches are also heavily utilized for the rapid prediction of quantum chemical properties, such as polarizabilities, partial charges, HOMO-LUMO levels or bond dissociation energies [241–273].

In Ref. [257], polarizabilities are calculated with both DFT and CCSD. The latter provides better overall accuracy at a drastically increased computational cost. GPR models trained on either of the resulting data sets are capable of predicting polarizabilities accurately. Yet the best performance with respect to the baseline provided by CCSD is achieved when learning and predicting the difference  $\Delta$  between these two levels of theory and then applying the output of such a model to DFT results. Such an approach has been followed in combined ML/DFT applications (see Sec. III C and the approaches for exchange-correlation functionals discussed therein) and is often referred to as  $\Delta$ -learning [274]. It was originally introduced for the prediction of energies. Another example in the context of these target quantities is Ref. [246], where NNs predict corrections on top of DFT-level energies for non-covalent interactions.

Other approaches follow an outline more consistent with the workflows described above, such as Ref. [245], wherein a GNN is utilized to predict bond dissociation energies for a large database of organic molecules. The resulting model can be applied for instance in drug design. Similarly, in Ref. [265] NNs are used to predict spin-state splittings of transition metal complexes, employing a test data set that includes experimental data in addition to DFT data.

Related to these works are applications of TDDFT, such as Ref. [251], wherein corrections to DFT calculated electronic spectra are learned from CCSD calculations within the  $\Delta$ -learning methodology.

Finally, another interesting approach is presented in Refs. [242] and [271]. Models based on KRR and NN were employed to predict whether a specific configuration should be treated with DFT or more accurate (and costly) multireference methods.



## 6. Magnetic, Thermal, and Energetic Properties

Finally, there are ML property mappings that do not directly fall into the application-based groupings above. Properties that can be derived from DFT and predicted or analyzed with ML methods but have not yet been discussed at length are, for instance, transport and diffusion properties [275–277], surface energies [278], and magnetic properties [279].

Some studies that combine ML and DFT fall into scientific disciplines not mentioned yet, such as studying RNA conformations [280] or the DFT prediction of nuclear magnetic resonance shifts (NMR) [281–283]. DFT can also be used as a pre-processing or pre-screening tool for ML workflows based on experimental data, as demonstrated in Ref. [284] and Ref. [285], where Curie and transition temperatures are learned, respectively. A purist’s approach to combining DFT and ML is taken in Ref. [286], which is concerned with melting temperatures.

Lastly, an important direction of research is concerned with energetic materials, i.e., materials that hold large amounts of chemical energy, such as fuels or explosives. A comprehensive study of ML techniques for energetic materials can be found in Ref. [287].

### B. Interatomic Potentials

A specialized type of property mappings in terms of ML are interatomic potentials (IAPs) (Sec. II A 2). Here, a ML mapping of atomic positions to the potential energy and atomic forces is performed, also referred to as force field or potential energy surface (PES). Most commonly, these ML-IAPs parametrize the Born-Oppenheimer PES defined in Eq. (3). The trained models can be used to perform MD simulations at an accuracy close to the first-principles methods used for the calculation of the training data (e.g., DFT) with no significant computational overhead. In the following, an overview of developments in the field of ML-IAPs is given. The reviewed publications have been grouped by the type of IAP used.

#### 1. Neural Network Potentials

A large number of IAP workflows rely on neural networks in order to approximate energies and forces from atomic positions [273, 288–325].

The type, architecture, and data representation of the employed NNs vary drastically between different approaches. Arguably the most common type of NNs within these approaches are deep feed-forward networks as used in Ref. [291]. Using such *deep potentials*, the authors were able to investigate a proposed liquid-liquid state transition in water, a study unattainable with regular DFT calculations. Similarly, in Ref. [304] an investigation of the phase-change material GeTe, which is

highly relevant for the development of non-volatile memory devices, is presented. These simulations would have been challenging with classical IAPs. Instead, a NN potential is used, drawing on Ref. [320] and Ref. [321]. The transferability of models highlights another advantage of ML-IAPs in comparison to highly specialized workflows, as they aim to replace DFT or classical potentials with often no application-induced constraints.

Traditional NNs can be improved by incorporating physical considerations, such as in Ref. [309]. The thought behind developing such physically enhanced networks is to enable extrapolation capabilities outside of the range of training data by encoding information on the chemical bonding of the atomic structures. Extrapolation is traditionally a problematic task for NNs, yet often necessary when performing simulations on new materials. They demonstrate the improved capabilities of such IAPs based on physically enhanced NNs for the case of aluminum.

The combination of DFT and ML is also leveraged to construct IAPs with an accuracy better than DFT itself, as shown in Ref. [307]. There, transfer learning is used to construct an IAP with CCSD accuracy. First an IAP is calculated using DFT data which is retrained via transfer learning on CCSD data. This enables rapid predictions of total energies with CCSD accuracy. A similar approach is taken in Ref. [326]. Here, a neural network is trained to correct a classical IAP to within DFT accuracy by training it on DFT data. Both of these approaches follow the principal idea of  $\Delta$ -learning mentioned in Sec. III A 5.

#### 2. Gaussian Approximation Potentials

Another important class of IAPs are Gaussian Approximation Potentials (GAPs), primarily based on Ref. [68]. Based upon GPs, the total energy is decomposed into atomic contributions, which are then learned by GPR from DFT data. Information on the atomic neighborhoods is encoded by a bispectrum representation of the atomic density. A range of publications have employed GAPs to model the PES [327–343].

For example, in Ref. [342], a GAP is trained for amorphous carbon. This potential is then used by the same authors in Ref. [339] to investigate electrode materials based on amorphous carbon through a combination of GAP and MD, a task usually unattainable with quantum mechanical accuracy. The same GAP is also employed in Ref. [343] and [341], further highlighting the universal re-usability of these IAPs.

A GAP has also been used to model phase-change materials [338], similar to Ref. [304] which was based on NNs. As one would expect, there exists no singular optimal choice for IAPs, and several different ML workflows can be used to tackle similar problems. Furthermore, GAPs have been shown to be able to model complicated PES of magnetic materials in Ref. [340].

Finally, GAPs can be incorporated in larger ML

workflows, such as the CALYPSO structure prediction method [337]. They also facilitate automatic workflows for optimal IAP construction [344].

### 3. Moment Tensor Potentials

Similar to GAPs, moment tensor potentials (MTPs) [345] are based on linear regression and polynomial expansion. They can, in principle, approximate any PES. MTPs pose another class of powerful ML-IAPs that have been used in many publications [346–365], often in conjunction with active learning approaches.

For example, Ref. [349] investigates solid-state batteries, which are of high interest due to the increasing demand for powerful and reliable electrical power sources. They employed a MTP that is trained on the fly that improves the accuracy of the IAP with ab-initio data where needed.

Using MTPs for active learning was first proposed by the authors of Ref. [345]. A linearly parameterized IAP is constructed on the fly based on the D-optimality criterion [350]. This method has subsequently been used to study diffusion processes [355] or lattice dynamics [354]. As the authors point out, this technique is not limited to MTPs and can be employed for other linear IAPs as well.

Another prominent application of MTPs is the investigation of both phononic properties and thermal conductivities [346, 347, 362, 365] and the search for new alloys [356].

### 4. Spectral Neighborhood Analysis and Other Potentials

Given the amount of possibilities ML-DFT techniques provide, it is no surprise that there are a number of IAP construction methods [72, 160, 322–325, 344, 366–376] that do not fall into the three categories outlined above.

The most prominent of these is the spectral neighborhood analysis potential (SNAP) [72, 376–378]. It is based on a linear model, similar in construction to GAP and MTPs. SNAP has been used to investigate lithium nitride [379], carbon under extreme conditions [380], and molybdenum alloys [367, 373].

Other IAP approaches draw on a range of different ML techniques, such as GPR or KRR [366, 368–370, 372, 375, 381]. Of note is further the (s)GDML approach [382–386], which combines a kernel based ML approach (GDML) with physical symmetries (sGDML) to construct efficient and accurate IAPs.

Finally, another procedure for constructing ML-IAPs is based on optimizing classical IAPs. This is done in Ref. [374], where a genetic algorithm is utilized to fit a bond order potential to investigate metal-organic structures.

### 5. Technical Aspects for the Constructing of Potentials

Finally, there are technical aspects when dealing with ML-IAPs. The construction of these IAPs is often a recurring task, since one is often not only interested in a specific system at specific conditions, but in rather large-scale systematic studies. The methodologies outlined above are not limited to specific elements or conditions apart from limitations introduced by DFT itself. Therefore, frameworks exist to simplify the repeated construction of such potentials. These include PANNA [387], ANI-1 [388], TensorMol [389] and Amp [390], which provide frameworks to build IAPs based on NNs. For MTPs, MLIP [391] provides a software framework incorporating active learning. GAPs can be build using the QUIP code [392] and SNAPs can be constructed with the LAMMPS [393] code.

Two other closely related aspects worth mentioning in this context are active learning and uncertainty quantification. Ref. [394] introduces a framework to realize such active learning procedures for IAPs, while Ref. [395] introduces potentials based on dropout-uncertainty neural networks (DUNN) which are capable of assessing uncertainties in their predictions. Such estimations are important, because NNs usually do not perform well during extrapolation tasks, yet such situations might occur during automated materials calculations.

## C. Learning Electronic Structures

While the preceding sections discussed ML approaches which center on specific properties extracted from DFT data, be it a wide variety of materials properties (Sec. III A) or specifically the energy-force-landscape (Sec. III B), the emphasis in the following is on using ML techniques to tackle the electronic structure problem directly. In these investigations, DFT is not solely a data acquisition tool, but rather the centerpiece of a ML-powered workflow aimed at a better understanding of electronic structures.

The most common mapping deals with the electronic exchange-correlation energy [396–417]. As discussed in Sec. II A 2, the accuracy of the chosen exchange-correlation approximation primarily determines the accuracy of the overall DFT simulation. The exchange-correlation functionals discussed in Sec. II A 2 mainly employ exact conditions for their constructions, but fitting to either experimental or high-level calculations data is also a common practice for the construction of exchange-correlation functionals. These have been in use for a long time (see Ref. [418]), especially in quantum chemistry. Using ML methods to perform similar fits is a natural consequence. An example of this is Ref. [401]. Here, the exact exchange-correlation energy is calculated for a system of limited size and then approximated using a NN, where the electronic density serves as the input quantity. By using automatic differentia-

tion, the exchange-correlation potential is obtained as well. The resulting model incorporates non-local effects, the lack thereof being a usual shortcoming of traditional exchange-correlation functionals at a negligible computational overhead compared to exact calculations. A similar approach is taken in Ref. [407], which introduces NeuralXC, a framework for constructing such NN exchange-correlation functionals, or Ref. [397], where CNNs are used to estimate the exchange-correlation energy from the density by drawing on the convolutions of the CNN architecture to featurize the electronic density. They recover the exact exchange of the B3LYP functional. This is especially important given that the B3LYP functional achieves exact exchange by evaluating the Kohn-Sham orbitals rather than the density, thus demonstrating how NNs can extract meaningful information from the electronic density. Improving exchange-correlation energies with ML is not necessarily limited to the construction of ML functionals. In Ref. [396], range-separation parameters for the established long-range corrected Becke-Lee-Yang-Parr (LC-BLYP) functional [27, 418–422] are learned from molecular features. The application of exact conditions in combination with ML functionals is also a subject of current research. As shown by Ref. [415], the kinetic energy functional rather than the exchange-correlation functional is investigated.

Instead of directly learning the exchange-correlation functional itself, corrections to these functionals can be learned. This  $\Delta$ -DFT method was already briefly introduced in Sec. III A 5 and has been applied in Ref. [404]. Here, differences between DFT and CC predicted total energies were learned using a KRR model. This model is in turn used to enhance DFT calculations towards chemical accuracy at almost no computational overhead, thereby, enabling highly accurate MD simulations. The advantage of methods is that constructing an ML correction rather than a ML functional reduces the training samples drastically. This is also shown in Ref. [410], where non-covalent interaction corrections to standard DFT functionals are learned at a much smaller cost.

Going beyond the electronic exchange-correlation energy, other studies focus on intermediate electronic quantities. In Ref. [423], pattern-learning techniques are used to predict the density of states (DOS) of an alloy system based on chemical features with high accuracy. The DOS is an important quantity in solid-state physics that can be used to discover or assess the properties of new materials in terms of their band energy. Ref. [424] utilizes a similar workflow to predict the DOS, but further includes uncertainty estimates. This workflow is mainly applied to silicon structures. The authors further compare the results of deriving the band energy from a predicted DOS with ML models that directly predict the band energy. They find that the former approach outperforms the latter. This makes intuitive sense given that the ML model learns from a larger amount of data with direct access to information about the electronic structure.

Another logical quantity of interest in this context is

the electronic density itself, for which various ML approaches exist [425–429]. Most notably, Ref. [425] uses KRR to learn and predict electronic densities, although these densities are in turn only intermediate quantities. The principle quantity of interest is the total energy. Similar to the results of Ref. [424], it is found that learning the electronic density from atomic positions and then the total energy from the electronic density, leads to more accurate results than can be achieved by a direct mapping of atomic positions to total energies. It should be noted that the mapping from the electronic density to the total energy has to be done via ML in the context of KS-DFT, since the KS energy functional includes terms that are only implicit functionals of the density, namely the kinetic energy of the non-interacting system, which depend on the KS wave functions explicitly.

In order to gain direct access to the total energy, another target quantity is needed. One possible candidate is the local density of states (LDOS), which gives the DOS at any point in space. In Ref. [430], NNs are used to predict both the electronic density and the LDOS individually at each grid point. By integrating the LDOS over the real space afterwards, the DOS is obtained. We have recently proposed a similar ML workflow [431] that follows along similar lines. It uses predicted LDOS values to obtain the total energy from the inferred DOS and density. Predicted densities can also be used to construct a  $\Delta$ -DFT [432] approach. On the other hand, Ref. [433] and Ref. [434] are concerned with the ML-aided construction of density functionals that enable an efficient prediction of the total energy directly from electronic densities, irrespective of how these densities are actually obtained.

However, electronic structure theory is not limited to an energy-based analysis. Refs. [435] and [436] apply ML methods to Hamiltonians in order to calculate different quantities of interest. Similarly, Ref. [437] uses GPR to fit potentials on DFT data, that can be used to improve calculations at lower levels of theory (e.g., density functional tight-binding). Finally, ML can also be used to accelerate the numerical treatment of DFT by reducing the numerical overhead of the solution of the KS equations [438] or identifying efficient, adaptive basis sets [439].

## D. Other Approaches

There are approaches that do not fit in the categories outlined above, either dealing with more technical aspects on combining ML and DFT (see Sec. III E) or simply mappings and calculations not discussed yet [371, 440–447] [448–451].

An example of such an approach is Ref. [441], which is concerned with the already discussed prediction of NMR shifts, but follows a different scheme to directly machine-learn chemical shifts from DFT by using DFT calculated properties as descriptors for a NN. Similar approaches are developed in Ref. [447] and Ref. [445].



Contrarily, Ref. [444] uses DFT as part of a loss function by calculating NMR spectra via DFT for molecules generated by a ML algorithm. Differences in DFT calculated and experimentally measured spectra are used to train this molecular generator. A similar approach is employed in Ref. [449] for finding organic molecules. Generally, models trained on data are used as key components of such search algorithms in the form of an "evaluator" of candidate structures proposed by a genetic algorithm [448, 450], an evolutionary algorithm [440], or a Monte Carlo simulation [442].

ML techniques are also used to predict "hyperparameters" of ab-initio simulations. These are usually adjusted manually through experience or by converging them with subsequently more costly and accurate calculations. An example is Ref. [446], where a DT was used to predict the number of k-points and the cutoff energy in plane-wave DFT calculations. Similarly, in Ref. [443] the choice of the exchange-correlation functional is predicted from chemical descriptors. Finally, DFT-based models are also integrated into elaborate statistical mechanics frameworks as in Ref. [451].

### E. Technical Aspects

Building complicated workflows involving DFT and ML poses several technical problems that have to be addressed. First and foremost, any data-driven approach is fundamentally reliant on data. In the context of ab-initio simulations, data is both costly to acquire and digitally exchangeable by nature. To that end, it makes intuitive sense to recycle already pre-calculated data and models in terms of databases, a number of which are given in Tab. I, alongside a short overview over useful software frameworks for ML-DFT approaches. For example, JARVIS includes properties not only on around 40,000 materials, but also ready-to-use ML models trained on DFT data. The database is constantly expanded and, thereby, facilitates the construction of new workflows. Similarly, acquiring new training data is facilitated with integrated workflows [171].

Another important aspect concerns the choices with respect to the ML methodologies. A wealth of ML techniques can be applied to DFT calculations and data. Likewise, there is a sizable amount of procedures to encode information in terms of descriptors. Investigations [452–454] as to which of these techniques hold advantages over others are crucial. In Ref. [454] different types of neural networks (CNN, DNN, single-layer NN) have been assessed in terms of predicting molecular properties, with a shallow DNN performing best. Similarly, in Ref. [452], different ways to encode compositional information for a GB model were compared. None of the investigated methods was able to produce models capable of assessing the stability of new materials reliably.

Indeed, the development of new descriptors and techniques to represent chemical information for specialized

use cases is subject of ongoing research [455–464]. For example, Ref. [464] introduces persistent image descriptors (PI), based on concepts from applied mathematics, to represent chemical structures in the search for functional groups capable of capturing CO<sub>2</sub>. Different ML models were trained on data encoded by these descriptors and compared to other commonly used descriptors. Common descriptors are SOAP [465], the Coulomb matrix [466], BoB [467], FCHL [468], and ACE [469, 470]. This list is not comprehensive, and a number of ML workflows use customized descriptors or simply easily accessible chemical information to represent materials. The utility of these descriptors is demonstrated in the aforementioned publications. Ref. [466–468] use their respective descriptors and KRR to learn and predict atomization energies. On the other hand, Ref. [465] uses SOAP descriptors to construct and compare GAPs, with Ref. [469] building ACE based IAPs as well.

Lastly, transferability and uncertainty quantification have to be taken into account in order to apply combined ML-DFT workflows in future applications. Correctly treating such aspects reduces the need to retrain models. As Ref. [471] suggests, the transferability of ML models can be utilized with certain limitations, as demonstrated for moderate volume changes in the liquid phase. Uncertainty quantification, as investigated in [472], deals with assessing whether a particular prediction can be trusted or additional data is needed to improve the model.

## IV. DISCUSSION

### A. Citation Analysis

We conclude this review by assessing general trends in the field in terms of a citation analysis based on the 365 research articles collected. The first part of this collection was gathered at the end of 2020 using Web of Science [519]. Grouping of these articles was performed manually. Citation numbers were retrieved via Google Scholar [520] by an automated python script. Both the script and the database of research articles are provided alongside this publication [521].

The results of this citation analysis are shown in Fig. 3. We note that we use the average number of citations as a measure in Fig. 3. It denotes how often a research article published in a particular year was *cited per year* on average in subsequent years. In this manner, articles published in later years can be judged on the same basis.

It is apparent from Fig. 3 that the number of publications increases rapidly, highlighting the growing importance of this research area. The number of publications has more than tripled in the years between 2017 and 2020. Interestingly, this growth in number of publications is not necessarily reflected in the number of citations. The average number of citations is consistent for works published in 2017–2020, despite the total number of articles grow-



Table I. Non-comprehensive selection of useful software frameworks for constructing ML-DFT workflows. The list of DFT software packages is ordered by the number of citations [473].

Task	Software
DFT calculations (periodic systems)	VASP [474–476], QuantumESPRESSO [477–479], WIEN2k [480, 481], CASTEP [482]
DFT calculations (isolated systems)	Gaussian [483], GAMESS [484, 485], ORCA [486, 487], MOLPRO [488–499]
Neural Network construction	PyTorch [500, 501], TensorFlow [502], Keras [503], Flux.jl [504, 505]
General ML library	Shogun [506], Scikit-Learn [507], Weka [508]
Gaussian Process Regression	GPytorch [509]
Constructing NN-IAPs	PANNA [387], ANI-1 [388], TensorMol [389], Amp [390]
Constructing other IAPs	MLIP [391], QUIP code [392], LAMMPS [393]
ML-exchange-correlation	NeuralXC [407]
Toolchaining	ASE [510], AiiDA [511, 512], DeepChem [513]
Databases	MoleculeNet [514], OQMD [515], ANI-1 [67], PubChemQC [516], NOMAD [517], JARVIS [518]

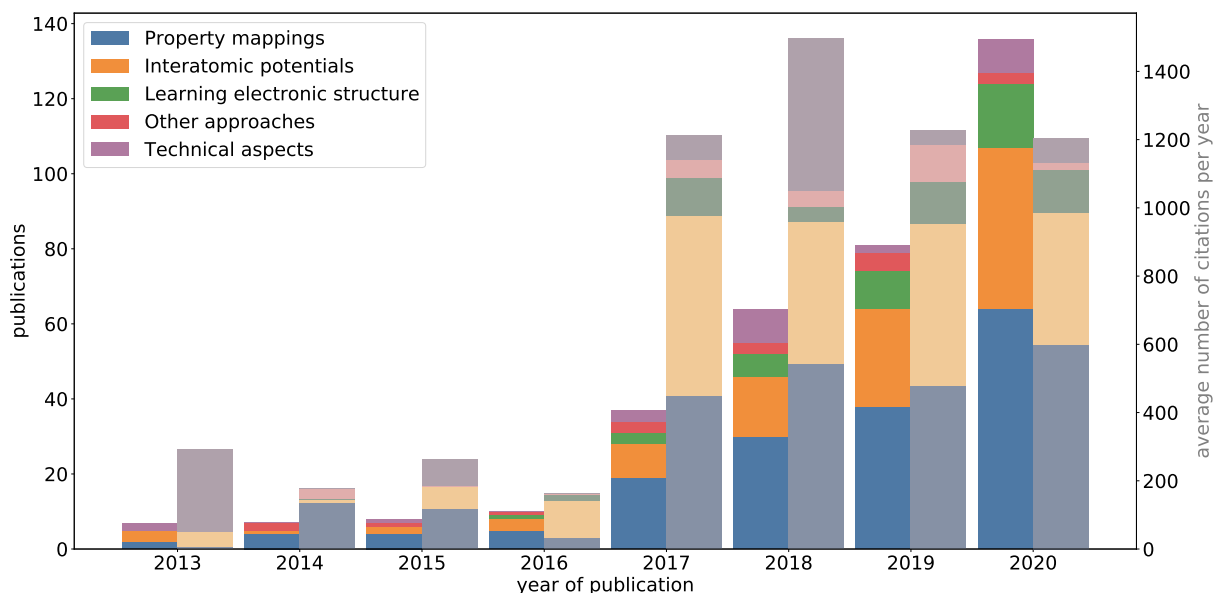


Figure 3. Number of publications and average number of citations per year by year of publication and category. Average number of citations marks how often a publication was cited per year after publication (including year of publication). The total number of publications for which data is shown here was 350, with data for 15 articles published prior to 2013 not shown for better visibility.

ing rapidly in these years. A possible explanation lies in the relatively young age of this research area. Recently published research articles are consequently very relevant for more recent publications. A clear illustration of this fact is apparent in the set of publications in 2013, where only two articles, Ref. [465] and [515], have accumulated a large number of citations.

Fig. 3 further illustrates how most research is directed towards property mappings and interatomic potentials. This makes intuitive sense, given that DFT calculations often underpin large-scale simulations that are based on data-driven methodologies. Yet, also a growing number of publications is concerned with applying ML to the electronic structure problem itself.

We perform the same analysis on the two largest cat-

egories in Fig. 3 separately. The results are illustrated in Fig. 4. Except for an increasing interest in applications concerning electrical properties and photovoltaics, another dominating trend cannot be identified for property mapping applications. Likewise, it can be seen that while in 2020 nearly half of all publications using IAPs employed NNs, the relative number of publications for the respective types of IAPs are relatively evenly distributed throughout the years. The large number of NN related publications is further illustrated in Fig. 5, which analyzes the number of publications mentioning certain ML methods. A prevalence for NNs is apparent, with Ridge Regression and GPR being the second and third-most used methods, respectively.

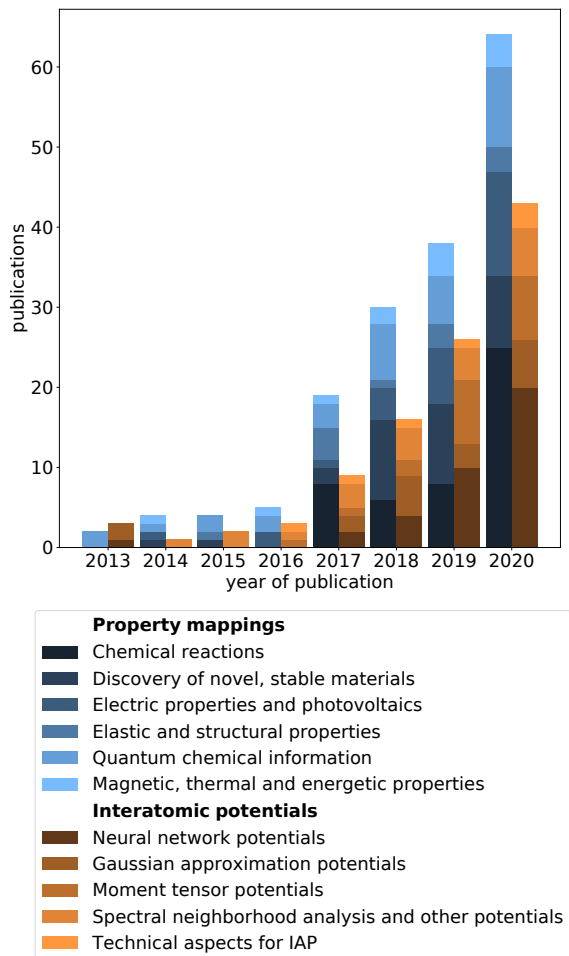


Figure 4. Number of publications by year for property mapping and interatomic potential related publications.

## B. Conclusions and Outlook

An extensive review over research combining ML and DFT was performed. By compiling and analyzing over 300 peer-reviewed research articles, we were able to identify five principal categories of research, presented in Sec. III. Most current efforts are focused on using data-driven methodologies to assist conventional DFT studies of materials, either by directly working with application-specific quantities (Sec. III A) or by providing ML-IAPs (Sec. III B) to facilitate dynamical calculations. The addressed range of topics is vast and follows pressing scientific problems, such as novel materials discovery, reducing the human carbon footprint, or enabling novel energy solutions. A number of publications (see Secs. III B 5 and III E) is concerned with technical aspects of enabling ML-DFT workflows by providing fundamental software solutions. Of particular importance are databases, which are the backbone for any data-driven approach. Another, still underrated research area focuses directly on addressing the shortcomings of DFT as an electronic structure

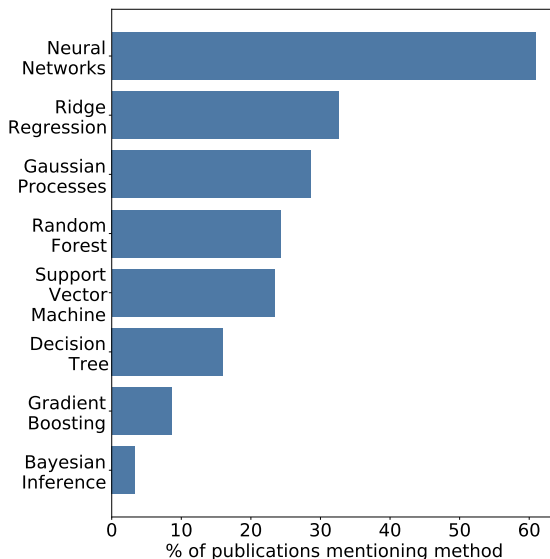


Figure 5. Mentions of ML methods throughout the reviewed articles. Note that only 325 publications could be included in this particular study, as no full text was available for the remaining 40.

method. These efforts have been discussed in Sec. III C. Drawing on this analysis, a general ML-DFT workflow can be identified, as is shown in Fig. 2, representing the vast majority of ML-DFT approaches, with only a small number of very specialized methods not being included. Both the variety as well as similarities in different methods can be identified in Fig. 2. In general, the research interest in ML-DFT methods is growing fast, as shown in Sec. IV A, and methods addressing the electronic structure problem directly are gaining traction. Considering the rising number of publication, this research field is expected to grow further. Future investigations may address topics such as uncertainty quantification and active learning approaches. Thereby, they may broaden the utility of ML in computational material science and computational chemistry.

With ever improving ML-DFT models, new applications that were previously unattainable will become possible, such as large-scale automated materials discovery, multi-scale modelling of materials, and digital twins of complex systems.

## ACKNOWLEDGMENTS

This work was funded by the Center for Advanced Systems Understanding (CASUS) which is financed by the German Federal Ministry of Education and Research (BMBF) and by the Saxon Ministry for Science, Art, and Tourism (SMWK) with tax funds on the basis of the budget approved by the Saxon State Parliament. We thank Aidan Thompson and Sebastian Schwalbe for useful com-

munications.

- 
- \* [l.felder@hzdr.de](mailto:l.felder@hzdr.de)  
† [k.shah@hzdr.de](mailto:k.shah@hzdr.de)  
‡ [m.bussmann@hzdr.de](mailto:m.bussmann@hzdr.de)  
§ [a.cangi@hzdr.de](mailto:a.cangi@hzdr.de)
- [1] A. Abedi, N. T. Maitra, and E. K. U. Gross, *Journal of Chemical Physics* **137**, 22A530 (2012).
  - [2] M. Born and R. Oppenheimer, *Annales de Physique* **389**, 457 (1927).
  - [3] S. K. Seth, S. Banerjee, and T. Kar, *Journal of Molecular Structure* **965**, 45 (2010).
  - [4] Z. Fang, Y. Zhao, H. Wang, J. Wang, S. Zhu, Y. Jia, J.-H. Cho, and S. Guan, *Applied Surface Science* **470**, 893 (2019).
  - [5] B. Civalleri, K. Doll, and C. M. Zicovich-Wilson, *Journal of Physical Chemistry B* **111**, 26 (2006).
  - [6] Y.-H. Duan, Z.-Y. Wu, B. Huang, and S. Chen, *Systematic Organisation of Information in Fuzzy Systems* **110**, 10 (2015).
  - [7] M. Ben Yahia, F. Lemoigno, T. Beuvier, J.-S. Filhol, M. Richard-Plouet, L. Brohan, and M.-L. Doublet, *Journal of Chemical Physics* **130**, 204501 (2009).
  - [8] D. Eustace, D. McComb, and A. Craven, *Micron (Oxford, England : 1993)* **41**, 547 (2010).
  - [9] D. A. Pantazis, M. Orio, T. Petrenko, S. Zein, E. Bill, W. Lubitz, J. Messinger, and F. Neese, *Chemistry A European Journal* **15**, 5108 (2009).
  - [10] B. Andriyevsky, T. Pilz, J. Yeon, P. Halasyamani, K. Doll, and M. Jansen, *Journal of Physics and Chemistry of Solids* **74**, 616 (2013).
  - [11] F. Mauri and S. G. Louie, *Physical Review Letters* **76**, 4246 (1996).
  - [12] F. A. La Porta, L. Gracia, J. Andrés, J. R. Sambrano, J. A. Varela, and E. Longo, *Journal of The American Ceramic Society* **97**, 4011 (2014).
  - [13] G. Zhang and C. B. Musgrave, *Journal of Physical Chemistry A* **111**, 1554 (2007).
  - [14] S. Schwalbe, T. Hahn, S. Liebing, K. Trepte, and J. Kortus, *Journal of Computational Chemistry* **39**, 2463 (2018).
  - [15] A. Dittmer, R. Izsák, F. Neese, and D. Maganas, *Inorganic Chemistry* **58**, 9303 (2019).
  - [16] G. M. Khrapkovskii, R. V. Tsyshevsky, D. V. Chachkov, D. L. Egorov, and A. G. Shamov, *Journal of Molecular Structure: THEOCHEM* **958**, 1 (2010).
  - [17] C.-G. Zhan, J. A. Nichols, and D. A. Dixon, *Journal of Physical Chemistry A* **107**, 4184 (2003).
  - [18] N. Hernández-Haro, J. Ortega-Castro, Y. B. Martynov, R. G. Nazmitdinov, and A. Frontera, *Chemical Physics* **516**, 225 (2019).
  - [19] M. Barhoumi, N. Sfina, and M. Said, *Solid State Communications* **329**, 114261 (2021).
  - [20] S. M. Daghash, P. Servio, and A. D. Rey, *Molecular Simulation* **45**, 1524 (2019).
  - [21] P. Hohenberg and W. Kohn, *Physical Review* **136**, B864 (1964).
  - [22] W. Kohn and L. J. Sham, *Physical Review* **140**, A1133 (1965).
  - [23] D. M. Ceperley and B. J. Alder, *Physical Review Letters* **45**, 566 (1980).
  - [24] J. P. Perdew, K. Burke, and M. Ernzerhof, *Physical Review Letters* **77**, 3865 (1996).
  - [25] J. Sun, A. Ruzsinszky, and J. P. Perdew, *Physical Review Letters* **115**, 036402 (2015).
  - [26] A. D. Becke, *Journal of Chemical Physics* **98**, 1372 (1993).
  - [27] C. Lee, W. Yang, and R. G. Parr, *Physical Review B* **37**, 785 (1988).
  - [28] E. Runge and E. K. U. Gross, *Physical Review Letters* **52**, 997 (1984).
  - [29] N. D. Mermin, *Physical Review* **137**, A1441 (1965).
  - [30] A. Pribram-Jones, S. Pittalis, E. K. U. Gross, and K. Burke, in *Lecture Notes in Computational Science and Engineering*, Vol. 96, edited by F. Graziani, M. P. Desjarlais, R. Redmer, and S. B. Trickey (Springer International Publishing, Cham, 2014) pp. 25–60.
  - [31] S. Pittalis, C. R. Proetto, A. Floris, A. Sanna, C. Bersier, K. Burke, and E. K. U. Gross, *Physical Review Letters* **107**, 163001 (2011).
  - [32] V. V. Karasiev, D. Chakraborty, O. A. Shukruto, and S. B. Trickey, *Physical Review B* **88**, 161108 (2013).
  - [33] N. Oliphant and R. J. Bartlett, *Journal of Chemical Physics* **100**, 6550 (1994).
  - [34] T. H. Dunning Jr, *The Journal of chemical physics* **90**, 1007 (1989).
  - [35] D. E. Woon and T. H. Dunning, *Journal of Chemical Physics* **100**, 2975 (1994).
  - [36] D. E. Woon and T. H. Dunning, *Journal of Chemical Physics* **98**, 1358 (1993).
  - [37] J. Koput and K. A. Peterson, *Journal of Physical Chemistry A* **106**, 9595 (2002).
  - [38] N. B. Balabanov and K. A. Peterson, *Journal of Chemical Physics* **123**, 064107 (2005).
  - [39] A. K. Wilson, D. E. Woon, K. A. Peterson, and T. H. Dunning, *Journal of Chemical Physics* **110**, 7667 (1999).
  - [40] P. E. M. Siegbahn, *Journal of Chemical Physics* **72**, 1647 (1980).
  - [41] H. Lischka, R. Shepard, F. B. Brown, and I. Shavitt, *International Journal of Quantum Chemistry* **20**, 91 (2009).
  - [42] B. Liu and M. Yoshimine, *Journal of Chemical Physics* **74**, 612 (1981).
  - [43] P. Saxe, D. J. Fox, H. F. Schaefer, and N. C. Handy, *Journal of Chemical Physics* **77**, 5584 (1982).
  - [44] V. Saunders and J. van Lenthe, *Molecular Physics* **48**, 923 (1983).
  - [45] H.-J. Werner and E.-A. Reinsch, *Journal of Chemical Physics* **76**, 3144 (1982).
  - [46] H.-J. Werner, Special Volume in Memory of Ilya Prigogine **69**, 1 (1987).
  - [47] H.-J. Werner and P. J. Knowles, *Theoretica chimica acta* **78**, 175 (1991).
  - [48] P. E. Siegbahn, *Chemical Physics Letters* **109**, 417 (1984).
  - [49] P. Knowles and N. Handy, *Handy NC (1989) Comput Phys Commun* **54**, 75 (1984).
  - [50] T. H. Dunning and P. J. Hay, in *Methods of electronic structure theory* (Springer, 1977) pp. 1–27.
  - [51] C. Møller and M. S. Plesset, *Physical Review* **46**, 618 (1934).
  - [52] C. Suellen, R. G. Freitas, P.-F. Loos, and D. Jacquemin, *Journal of Chemical Theory and Computation* **15**, 4581 (2019).



- [53] L. K. Wagner, M. Bajdich, and L. Mitas, *Journal of Computational Physics* **228**, 3390 (2009).
- [54] J. Kim, A. D. Baczewski, T. D. Beaudet, A. Benali, M. C. Bennett, M. A. Berrill, N. S. Blunt, E. J. L. Borda, *et al.*, *Journal of Physics: Condensed Matter* **30**, 195901 (2018).
- [55] J. A. Barker, *Journal of Chemical Physics* **70**, 2914 (1979).
- [56] B. Militzer, F. González-Cataldo, S. Zhang, K. P. Driver, and F. Soubiran, *Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* **103**, 013203 (2021).
- [57] J. E. Jones, *Proceedings of the Royal Society of London A* **106**, 463 (1924).
- [58] M. S. Daw and M. I. Baskes, *Physical Review B* **29**, 6443 (1984).
- [59] M. S. Daw and M. I. Baskes, *Physical Review Letters* **50**, 1285 (1983).
- [60] M. I. Baskes, *Physical Review Letters* **59**, 2666 (1987).
- [61] J. Tersoff, *Physical Review B* **37**, 6991 (1988).
- [62] D. W. Brenner, *Physical Review B* **42**, 9458 (1990).
- [63] D. G. Pettifor and I. I. Oleinik, *Physical Review B* **59**, 8487 (1999).
- [64] S. J. Stuart, A. B. Tutein, and J. A. Harrison, *Journal of Chemical Physics* **112**, 6472 (2000).
- [65] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, *Journal of Physical Chemistry A* **105**, 9396 (2001).
- [66] J. Yu, S. B. Sinnott, and S. R. Phillpot, *Physical Review B* **75**, 085311 (2007).
- [67] J. S. Smith, O. Isayev, and A. E. Roitberg, *Scientific data* **4**, 170193 (2017).
- [68] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, *Physical Review Letters* **104**, 136403 (2010).
- [69] H. Chan, B. Narayanan, M. Cherukara, T. D. Loeffler, M. G. Sternberg, A. Avarca, and S. K. Sankaranarayanan, *MRS Advances* **6**, 21 (2021).
- [70] N. Lubbers, J. S. Smith, and K. Barros, *Journal of Chemical Physics* **148**, 241715 (2018).
- [71] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, *Journal of Chemical Physics* **148**, 241722 (2018).
- [72] A. Thompson, L. Swiler, C. Trott, S. Foiles, and G. Tucker, *Journal of Computational Physics* **285**, 316 (2015).
- [73] L. Zhang, J. Han, H. Wang, R. Car, and W. E, *Physical Review Letters* **120**, 143001 (2018).
- [74] T. D. Huan, R. Batra, J. Chapman, S. Krishnan, L. Chen, and R. Ramprasad, *npj computational materials* **3** (2017), 10.1038/s41524-017-0042-y.
- [75] A. Prakash, K. Chitta, and A. Geiger, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2021) pp. 7077–7087.
- [76] M. D. Schwartz, *Harvard Data Science Review* (2021), 10.1162/99608f92.beeb1183.
- [77] G. Pilania, *Systematic Organisation of Information in Fuzzy Systems* **193**, 110360 (2021).
- [78] T. M. Mitchell, *Machine Learning*, 1st ed. (McGraw-Hill Education, New York, 1997).
- [79] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, in *Advances in Neural Information Processing Systems*, Vol. 27 (Curran Associates, Inc., 2014).
- [80] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, Adaptive Computation and Machine Learning (The MIT Press, Cambridge, Mass, 2005).
- [81] L. Rokach and O. Z. Maimon, *Data Mining with Decision Trees: Theory and Applications* (World Scientific, 2008).
- [82] C. Cortes and V. Vapnik, *Machine learning* **20**, 273 (1995).
- [83] J. MacQueen *et al.*, in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. 1 (Oakland, CA, USA, 1967) pp. 281–297.
- [84] L. Breiman, *Machine Learning* **45**, 5 (2001).
- [85] J. H. Friedman, *The Annals of Statistics* **29**, 1189 (2001).
- [86] S. Ruder, arXiv:1609.04747 [cs] (2017), arXiv:1609.04747 [cs].
- [87] X. Yan and X. G. Su, *Linear Regression Analysis. Theory and Computing* (WORLD SCIENTIFIC, 2009).
- [88] R. Tibshirani, *Journal of the Royal Statistical Society. Series B (Methodological)* **58**, 267 (1996).
- [89] A. E. Hoerl and R. W. Kennard, *Technometrics : a journal of statistics for the physical, chemical, and engineering sciences* **12**, 55 (1970).
- [90] R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler, and L. M. Ghiringhelli, *Physical Review Materials* **2**, 083802 (2018).
- [91] V. Vovk, in *Empirical Inference*, edited by B. Schölkopf, Z. Luo, and V. Vovk (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013) pp. 105–116.
- [92] J. M. Hilbe, *Logistic Regression Models* (Taylor & Francis, 2009).
- [93] K. Hornik, *Neural Networks* **4**, 251 (1991).
- [94] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, *et al.* (2020).
- [95] E. G. Julie, Y. H. Robinson, and S. M. Jaisakthi, *Handbook of Deep Learning in Biomedical Engineering and Health Informatics. Biomedical Engineering: Techniques and Applications* (Apple Academic Press, 2021).
- [96] F. Rosenblatt, *The Perceptron: A Perceiving and Recognizing Automaton (Project PARA)*, Tech. Rep. 85-460-1 (Cornell Aeronautical Laboratory, 1957).
- [97] Y. LeCun, Y. Bengio, and G. Hinton, *Nature* **521**, 436 (2015).
- [98] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, arXiv:1811.03378 [cs] (2018), arXiv:1811.03378 [cs].
- [99] V. Nair and G. E. Hinton, in *ICML* (2010) pp. 807–814.
- [100] L. Bottou, in *Proceedings of COMPSTAT'2010*, edited by Y. Lechevallier and G. Saporta (Physica-Verlag HD, Heidelberg, 2010) pp. 177–186.
- [101] D. P. Kingma and J. Ba, arXiv:1412.6980 (2017), arXiv:1412.6980 [cs.LG].
- [102] M. Minsky and S. A. Papert, *Perceptrons. An Introduction to Computational Geometry*, expanded, subsequent edition ed. (The MIT Press, Cambridge, Mass, 2017).
- [103] Z. Dai, H. Liu, Q. V. Le, and M. Tan, arXiv:2106.04803 [cs] (2021), arXiv:2106.04803 [cs].
- [104] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, *IEEE Transactions on Neural Networks* **20**, 61 (2009).
- [105] M. Cranmer, A. Sanchez Gonzalez, P. Battaglia, R. Xu, K. Cranmer, D. Spergel, and S. Ho, *Advances in Neural*

- Information Processing Systems 9 **33**, 17429 (2020).
- [106] I. Sutskever, O. Vinyals, and Q. V. Le, in *Advances in Neural Information Processing Systems*, Vol. 27 (Curran Associates, Inc., 2014).
- [107] M. Raissi, P. Perdikaris, and G. Karniadakis, *Journal of Computational Physics* **378**, 686 (2019).
- [108] S. Guan and M. Loew, in *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)* (IEEE, 2020) pp. 101–106.
- [109] K. Cheng, N. Wang, and M. Li, in *Advances in Natural Computation, Fuzzy Systems and Knowledge Discovery*, Advances in Intelligent Systems and Computing, edited by H. Meng, T. Lei, M. Li, K. Li, N. Xiong, and L. Wang (Springer International Publishing, Cham, 2021) pp. 475–486.
- [110] J. Lee, Y. Bahri, R. Novak, S. S. Schoenholz, J. Pennington, and J. Sohl-Dickstein, arXiv:1711.00165 (2018), arXiv:1711.00165 [stat.ML].
- [111] M. Kuss, T. Pfingsten, L. Csato, and C. E. Rasmussen, *Approximate Inference for Robust Gaussian Process Regression*, Tech. Rep. 136 (Max Planck Institute for Biological Cybernetics, 2005).
- [112] J. Dziedzic, A. Bhandari, L. Anton, C. Peng, J. C. Womack, M. Famili, D. Kramer, and C.-K. Skylaris, *Journal of Physical Chemistry C* **124**, 7860 (2020).
- [113] H. Li, L. Shi, M. Zhang, Z. Su, X. Wang, L. Hu, and G. Chen, *Journal of Chemical Physics* **126**, 144101 (2007).
- [114] C. Deng, Y. Su, F. Li, W. Shen, Z. Chen, and Q. Tang, *Journal of Materials Chemistry A: Materials for Energy and Sustainability* **8**, 24563 (2020).
- [115] L. Wu, T. Guo, and T. Li, *Journal of Materials Chemistry A: Materials for Energy and Sustainability* **8**, 19290 (2020).
- [116] V. Fung, G. Hu, Z. Wu, and D.-e. Jiang, *Journal of Physical Chemistry C* **124**, 19571 (2020).
- [117] N. Artrith, Z. Lin, and J. G. Chen, *Acs Catalysis* **10**, 9438 (2020).
- [118] F. Liu, S. Yang, and A. J. Medford, *ChemCatChem* **12**, 4317 (2020).
- [119] C. S. Praveen and A. Comas-Vives, *ChemCatChem* **12**, 4611 (2020).
- [120] J. Zheng, X. Sun, C. Qiu, Y. Yan, Z. Yao, S. Deng, X. Zhong, G. Zhuang, Z. Wei, and J. Wang, *Journal of Physical Chemistry C* **124**, 13695 (2020).
- [121] C. T. Ser, P. Žuvela, and M. W. Wong, *Applied Surface Science* **512**, 145612 (2020).
- [122] P. Friederich, G. dos Passos Gomes, R. De Bin, A. Aspuru-Guzik, and D. Balcels, *Chemical Science* **11**, 4584 (2020).
- [123] D. Ologunagba and S. Kattel, *Energies* **13**, 2182 (2020).
- [124] K. K. Rao, Q. K. Do, K. Pham, D. Maiti, and L. C. Grabow, *Topics in catalysis* **63**, 728 (2020).
- [125] S. Lin, H. Xu, Y. Wang, X. C. Zeng, and Z. Chen, *Journal of Materials Chemistry A: Materials for Energy and Sustainability* **8**, 5663 (2020).
- [126] J. Zhang, J. Chen, P. Hu, and H. Wang, *Chinese Chemical Letters* **31**, 890 (2020).
- [127] M. Sun, A. W. Dougherty, B. Huang, Y. Li, and C.-H. Yan, *Advanced Energy Materials* **10**, 1903949 (2020).
- [128] J. K. Pedersen, T. A. A. Batchelor, A. Bagger, and J. Rossmeisl, *Acs Catalysis* **10**, 2169 (2020).
- [129] A. J. Chowdhury, W. Yang, K. E. Abdelfatah, M. Zare, A. Heyden, and G. A. Terejanu, *Journal of Chemical Theory and Computation* **16**, 1105 (2020).
- [130] M. Kim, B. C. Yeo, Y. Park, H. M. Lee, S. S. Han, and D. Kim, *Chemistry of Materials* **32**, 709 (2019).
- [131] Z. Lu, S. Yadav, and C. V. Singh, *Catalysis Science & Technology* **10**, 86 (2020).
- [132] S. Saxena, T. S. Khan, F. Jalid, M. Ramteke, and M. A. Haider, *Journal of Materials Chemistry A: Materials for Energy and Sustainability* **8**, 107 (2020).
- [133] X. Li, R. Chiong, Z. Hu, D. Cornforth, and A. J. Page, *Journal of Chemical Theory and Computation* **15**, 6882 (2019).
- [134] X. Guo, S. Lin, J. Gu, S. Zhang, Z. Chen, and S. Huang, *Acs Catalysis* **9**, 11042 (2019).
- [135] S. Back, J. Yoon, N. Tian, W. Zhong, K. Tran, and Z. W. Ulissi, *Journal of Physical Chemistry Letters* **10**, 4401 (2019).
- [136] R. A. Hoyt, M. M. Montemore, I. Fampiou, W. Chen, G. Tritsarlis, and E. Kaxiras, *Journal of Chemical Information and Modeling* **59**, 1357 (2019).
- [137] T. A. Batchelor, J. K. Pedersen, S. H. Winther, I. E. Castelli, K. W. Jacobsen, and J. Rossmeisl, *Joule* **3**, 834 (2019).
- [138] H. A. Tahini, X. Tan, and S. C. Smith, *Adv. Theory Simul.* **2**, 1800202 (2019).
- [139] R. Anderson, J. Rodgers, E. Argueta, A. Biong, and D. A. Gómez-Gualdrón, *Chemistry of Materials* **30**, 6325 (2018).
- [140] T. Toyao, K. Suzuki, S. Kikuchi, S. Takakusagi, K.-i. Shimizu, and I. Takigawa, *Journal of Physical Chemistry C* **122**, 8315 (2018).
- [141] R. B. Wexler, J. M. P. Martinez, and A. M. Rappe, *Journal of The American Chemical Society* **140**, 4678 (2018).
- [142] A. Yada, K. Nagata, Y. Ando, T. Matsumura, S. Ichinoseki, and K. Sato, *Chemistry Letters* **47**, 284 (2018).
- [143] R. Jinnouchi, H. Hirata, and R. Asahi, *Journal of Physical Chemistry C* **121**, 26397 (2017).
- [144] Z. W. Ulissi, M. T. Tang, J. Xiao, X. Liu, D. A. Torelli, M. Karamad, K. Cummins, C. Hahn, N. S. Lewis, T. F. Jaramillo, *et al.*, *Acs Catalysis* **7**, 6600 (2017).
- [145] R. Gasper, H. Shi, and A. Ramasubramaniam, *Journal of Physical Chemistry C* **121**, 5612 (2017).
- [146] J. R. Boes and J. R. Kitchin, *Molecular Simulation* **43**, 346 (2017).
- [147] B. Sawatlon, M. D. Wodrich, B. Meyer, A. Fabrizio, and C. Corminboeuf, *ChemCatChem* **11**, 4096 (2019).
- [148] Z. Li, X. Ma, and H. Xin, *Catalysis Today* **280**, 232 (2017).
- [149] B. Meyer, B. Sawatlon, S. Heinen, O. A. von Lilienfeld, and C. Corminboeuf, *Chemical Science* **9**, 7069 (2018).
- [150] X. Wang, C. Wang, S. Ci, Y. Ma, T. Liu, L. Gao, P. Qian, C. Ji, and Y. Su, *Journal of Materials Chemistry A: Materials for Energy and Sustainability* **8**, 23488 (2020).
- [151] M. Shetty, M. A. Ardagh, Y. Pang, O. A. Abdelrahman, and P. J. Dauenhauer, *Acs Catalysis* **10**, 12867 (2020).
- [152] X. Sun, J. Zheng, Y. Gao, C. Qiu, Y. Yan, Z. Yao, S. Deng, and J. Wang, *Applied Surface Science* **526**, 146522 (2020).
- [153] H. Li, S. Xu, M. Wang, Z. Chen, F. Ji, K. Cheng, Z. Gao, Z. Ding, and W. Yang, *Journal of Materials Chemistry A: Materials for Energy and Sustainability* **8**, 17987 (2020).

- [154] S. M. Maley, D.-H. Kwon, N. Rollins, J. C. Stanley, O. L. Sydora, S. M. Bischof, and D. H. Ess, *Chemical Science* **11**, 9665 (2020).
- [155] Z. Yang, W. Gao, and Q. Jiang, *Journal of Materials Chemistry A: Materials for Energy and Sustainability* **8**, 17507 (2020).
- [156] M. Hakala, R. Kronberg, and K. Laasonen, *Scientific reports* **7**, 15243 (2017).
- [157] R. Jinnouchi and R. Asahi, *Journal of Physical Chemistry Letters* **8**, 4279 (2017).
- [158] Z. W. Ulissi, A. J. Medford, T. Bligaard, and J. K. Nørskov, *Nature communications* **8**, 14621 (2017).
- [159] C. Lu, Q. Liu, C. Wang, Z. Huang, P. Lin, and L. He, *AAAI* **33**, 1052 (2019).
- [160] M. Scherbela, L. Hörmann, A. Jeindl, V. Obersteiner, and O. T. Hofmann, *Physical Review Materials* **2**, 043803 (2018).
- [161] J.-W. Yeh, S.-K. Chen, S.-J. Lin, J.-Y. Gan, T.-S. Chin, T.-T. Shun, C.-H. Tsau, and S.-Y. Chang, *Advanced Engineering Materials* **6**, 299 (2004).
- [162] S. Nellaiappan, N. Kumar, R. Kumar, A. Parui, K. D. Malviya, K. G. Pradeep, A. K. Singh, S. Sharma, C. S. Tiwary, *et al.*, *ChemRxiv* (2019), 10.26434/chemrxiv.9777218.v1.
- [163] J. Hu, C. Wang, Q. Li, R. Sa, and P. Gao, *APL Materials* **8**, 111109 (2020).
- [164] Z. Zhang, M. Li, K. Flores, and R. Mishra, *Journal of Applied Physics* **128**, 105103 (2020).
- [165] A. S. Yasin and T. D. Musho, *Systematic Organisation of Information in Fuzzy Systems* **181**, 109726 (2020).
- [166] E. M. D. Siriwardane, R. P. Joshi, N. Kumar, and D. Çakır, *ACS Applied Materials & Interfaces*, *acsami.0c03536* (2020).
- [167] S. Ulenberg, M. Belka, M. Król, F. Herold, W. Hewelt-Belka, A. Kot-Wasik, and T. Bączek, *PLoS ONE* **10**, e0122772 (2015).
- [168] B. Meredig, A. Agrawal, S. Kirklin, J. E. Saal, J. W. Doak, A. Thompson, K. Zhang, A. Choudhary, and C. Wolverton, *Physical Review B* **89**, 094104 (2014).
- [169] J. Noh, G. H. Gu, S. Kim, and Y. Jung, *Journal of Chemical Information and Modeling* **60**, 1996 (2020).
- [170] T. Wu and J. Wang, *Nano Energy* **66**, 104070 (2019).
- [171] A. Nandy, J. Zhu, J. P. Janet, C. Duan, R. B. Getman, and H. J. Kulik, *Acs Catalysis* **9**, 8243 (2019).
- [172] P. C. Jennings, S. Lysgaard, J. S. Hummelshøj, T. Vegge, and T. Bligaard, *npj computational materials* **5**, 46 (2019).
- [173] Y.-P. Li, K. Han, C. A. Grambow, and W. H. Green, *Journal of Physical Chemistry A* **123**, 2142 (2019).
- [174] Z. Li, Q. Xu, Q. Sun, Z. Hou, and W.-J. Yin, *Advanced Functional Materials* **29**, 1807280 (2019).
- [175] K. Kim, L. Ward, J. He, A. Krishna, A. Agrawal, and C. Wolverton, *Physical Review Materials* **2**, 123801 (2018).
- [176] X. Zheng, P. Zheng, and R.-Z. Zhang, *Chemical Science* **9**, 8426 (2018).
- [177] W. Ye, C. Chen, Z. Wang, I.-H. Chu, and S. P. Ong, *Nature communications* **9**, 3800 (2018).
- [178] Y. He, E. D. Cubuk, M. D. Allendorf, and E. J. Reed, *Journal of Physical Chemistry Letters* **9**, 4562 (2018).
- [179] W. Li, R. Jacobs, and D. Morgan, *Systematic Organisation of Information in Fuzzy Systems* **150**, 454 (2018).
- [180] J. P. Janet, L. Chan, and H. J. Kulik, *Journal of Physical Chemistry Letters* **9**, 1064 (2018).
- [181] M. Chandran, S. C. Lee, and J.-H. Shim, *Modelling and Simulation in Materials Science and Engineering* **26**, 025010 (2018).
- [182] A. Seko, H. Hayashi, H. Kashima, and I. Tanaka, *Physical Review Materials* **2**, 013805 (2018).
- [183] L. Ward, R. Liu, A. Krishna, V. I. Hegde, A. Agrawal, A. Choudhary, and C. Wolverton, *Physical Review B* **96**, 024104 (2017).
- [184] J. Schmidt, J. Shi, P. Borlido, L. Chen, S. Botti, and M. A. L. Marques, *Chemistry of Materials* **29**, 5090 (2017).
- [185] M. Kaneko, M. Fujii, T. Hisatomi, K. Yamashita, and K. Domen, *Journal of Energy Chemistry* **36**, 7 (2019).
- [186] T. Q. Hartnett, M. V. Ayyasamy, and P. V. Balachandran, *MRS Communications* **9**, 882 (2019).
- [187] C. W. Park and C. Wolverton, *Physical Review Materials* **4**, 063801 (2020).
- [188] S. Xiong, X. Li, X. Wu, J. Yu, O. I. Gorbatov, I. Di Marco, P. R. Kent, and W. Sun, *Systematic Organisation of Information in Fuzzy Systems* **184**, 109830 (2020).
- [189] G. R. Schleder, C. M. Acosta, and A. Fazzio, *ACS Applied Materials & Interfaces* **12**, 20149 (2019).
- [190] J. Lu, C. Wang, and Y. Zhang, *Journal of Chemical Theory and Computation* **15**, 4113 (2019).
- [191] Y. Li, B. Xiao, Y. Tang, F. Liu, X. Wang, F. Yan, and Y. Liu, *Journal of Physical Chemistry C* **124**, 28458 (2020).
- [192] M. Askerka, Z. Li, M. Lempen, Y. Liu, A. Johnston, M. I. Saidaminov, Z. Zajacz, and E. H. Sargent, *Journal of The American Chemical Society* **141**, 3682 (2019).
- [193] P. B. Jørgensen, M. Mesta, S. Shil, J. M. García Lastra, K. W. Jacobsen, K. S. Thygesen, and M. N. Schmidt, *Journal of Chemical Physics* **148**, 241735 (2018).
- [194] F. Curtis, X. Li, T. Rose, Á. Vázquez-Mayagoitia, S. Bhattacharya, L. M. Ghiringhelli, and N. Marom, *Journal of Chemical Theory and Computation* **14**, 2246 (2018).
- [195] B. Nebgen, N. Lubbers, J. S. Smith, A. E. Sifain, A. Lokhov, O. Isayev, A. E. Roitberg, K. Barros, and S. Tretiak, *Journal of Chemical Theory and Computation* **14**, 4687 (2018).
- [196] D. Choudhuri, *Systematic Organisation of Information in Fuzzy Systems* **187**, 110111 (2021).
- [197] T. Wu and J. Wang, *ACS Applied Materials & Interfaces* **12**, 57821 (2020).
- [198] N. Meftahi, M. Klymenko, A. J. Christofferson, U. Bach, D. A. Winkler, and S. P. Russo, *npj computational materials* **6**, 166 (2020).
- [199] O. Allam, R. Kuramshin, Z. Stoichev, B. Cho, S. Lee, and S. Jang, *Materials Today Energy* **17**, 100482 (2020).
- [200] R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, T. D. Hirzel, A. Aspuru-Guzik, D.-G. Ha, M. Einzinger, T. C. Wu, and M. A. Baldo, *SPIE* (2016).
- [201] M. W. Gaultois, A. O. Oliynyk, A. Mar, T. D. Sparks, G. J. Mulholland, and B. Meredig, *APL Materials* **4**, 053213 (2016).
- [202] A. Mannodi-Kanakkithodi, M. Y. Toriyama, F. G. Sen, M. J. Davis, R. F. Klie, and M. K. Y. Chan, *npj computational materials* **6**, 134 (2020).
- [203] J. P. Janet, S. Ramesh, C. Duan, and H. J. Kulik, *ACS Central Science* **6**, 513 (2020).
- [204] Z. Zhu, B. Dong, H. Guo, T. Yang, and Z. Zhang, *Chinese Physics B* **29**, 046101 (2020).



- [205] K. Choudhary, M. Bercx, J. Jiang, R. Pachter, D. Lamoen, and F. Tavazza, *Chemistry of Materials* **31**, 5900 (2019).
- [206] Y. Huang, C. Yu, W. Chen, Y. Liu, C. Li, C. Niu, F. Wang, and Y. Jia, *Journal of Materials Chemistry C: Materials for Optical and Electronic Devices* **7**, 3238 (2019).
- [207] H. Park, R. Mall, F. H. Alharbi, S. Sanvito, N. Tabet, H. Bensmail, and F. El-Mellouhi, *Physical Chemistry Chemical Physics* **21**, 1078 (2019).
- [208] D. Padula, J. D. Simpson, and A. Troisi, *Materials Horizons* **6**, 343 (2019).
- [209] A. D. Sendek, E. D. Cubuk, E. R. Antoniuk, G. Cheon, Y. Cui, and E. J. Reed, *Chemistry of Materials* **31**, 342 (2018).
- [210] A. Paul, D. Jha, R. Al-Bahrani, W.-k. Liao, A. Choudhary, and A. Agrawal, in *2019 International Joint Conference on Neural Networks (IJCNN)* (IEEE, Budapest, Hungary, 2019) pp. 1–8.
- [211] A. Mannodi-Kanakkithodi, M. Toriyama, F. G. Sen, M. J. Davis, R. F. Klie, and M. K. Chan, in *2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)* (IEEE, Chicago, IL, USA, 2019) pp. 0791–0794.
- [212] A. C. Rajan, A. Mishra, S. Satsangi, R. Vaish, H. Mizuseki, K.-R. Lee, and A. K. Singh, *Chemistry of Materials* **30**, 4031 (2018).
- [213] M. V. Tabib, O. M. Løvrvik, K. Johannessen, A. Rasheed, E. Sagvolden, and A. M. Rustad, in *Artificial Neural Networks and Machine Learning – ICANN 2018*, Vol. 11139, edited by V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis, and I. Maglogiannis (Springer International Publishing, Cham, 2018) pp. 392–401.
- [214] O. Allam, B. W. Cho, K. C. Kim, and S. S. Jang, *RSC Advances* **8**, 39414 (2018).
- [215] M. Fernandez, A. Bilić, and A. S. Barnard, *Zinc Oxide Materials and Devices Iv* **28**, 38LT03 (2017).
- [216] G. Zhang, J. Yuan, Y. Mao, and Y. Huang, *Systematic Organisation of Information in Fuzzy Systems* **186**, 109998 (2021).
- [217] L. Ju, M. Li, L. Tian, P. Xu, and W. Lu, *Materials Today Communications* **25**, 101604 (2020).
- [218] S. K. Kauwe, T. Welker, and T. D. Sparks, *Integrating materials and manufacturing innovation* **9**, 213 (2020).
- [219] K. Choudhary, K. F. Garrity, and F. Tavazza, *Journal of Physics: Condensed Matter* **32**, 475501 (2020).
- [220] N. C. Frey, D. Akinwande, D. Jariwala, and V. B. Shenoy, *ACS Nano* **14**, 13406 (2020).
- [221] E. Antono, N. N. Matsuzawa, J. Ling, J. E. Saal, H. Arai, M. Sasago, and E. Fujii, *Journal of Physical Chemistry A* **124**, 8330 (2020).
- [222] M. P. Lourenço, A. dos Santos Anastácio, A. L. Rosa, T. Fraunheim, and M. C. da Silva, *Journal of molecular modeling* **26**, 187 (2020).
- [223] K. Choudhary, K. F. Garrity, V. Sharma, A. J. Biacchi, A. R. Hight Walker, and F. Tavazza, *npj computational materials* **6**, 64 (2020).
- [224] K. T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K. R. Müller, and E. K. U. Gross, *Physical Review B* **89**, 205118 (2014).
- [225] H. M. Fayek, L. Cavedon, and H. R. Wu, *Neural Networks* **128**, 345 (2020).
- [226] L. Hedin, *Physical Review* **139**, A796 (1965).
- [227] W. G. Aulbur, L. Jönsson, and J. W. Wilkins, *Solid state physics* (New York. 1955) **54**, 1 (2000).
- [228] F. Aryasetiawan and O. Gunnarsson, *Reports On Progress in Physics* **61**, 237 (1998).
- [229] C. A. F. Salvador, B. F. Zornio, and C. R. Miranda, *ACS Applied Materials & Interfaces* **12**, 56850 (2020).
- [230] W. Tong, Q. Wei, H.-Y. Yan, M.-G. Zhang, and X.-M. Zhu, *Frontiers in Physics* **15**, 63501 (2020).
- [231] M. Rupp, *International Journal of Quantum Chemistry* **115**, 1058 (2015).
- [232] D. Jain, S. Chaube, P. Khullar, S. Goverapet Srinivasan, and B. Rai, *Physical Chemistry Chemical Physics* **21**, 19423 (2019).
- [233] S. A. Tawfik, O. Isayev, C. Stampfl, J. Shapter, D. A. Winkler, and M. J. Ford, *Adv. Theory Simul.* **2**, 1800128 (2018).
- [234] F. Musil, S. De, J. Yang, J. E. Campbell, G. M. Day, and M. Ceriotti, *Chemical Science* **9**, 1289 (2018).
- [235] P. V. Balachandran, T. Shearman, J. Theiler, and T. Lookman, *Acta Crystallogr Sect B* **73**, 962 (2017).
- [236] J. Wang, X. Yang, Z. Zeng, X. Zhang, X. Zhao, and Z. Wang, *Systematic Organisation of Information in Fuzzy Systems* **138**, 135 (2017).
- [237] T. Tamura, M. Karasuyama, R. Kobayashi, R. Arakawa, Y. Shiihara, and I. Takeuchi, *Modelling and Simulation in Materials Science and Engineering* **25**, 075003 (2017).
- [238] J. Wang, X. Yang, G. Wang, J. Ren, Z. Wang, X. Zhao, and Y. Pan, *Systematic Organisation of Information in Fuzzy Systems* **134**, 190 (2017).
- [239] H. Zhao, C. I. Ezech, W. Ren, W. Li, C. H. Pang, C. Zheng, X. Gao, and T. Wu, *Applied Energy* **254**, 113651 (2019).
- [240] M. V. Ayyasamy, J. A. Deijkers, H. N. Wadley, and P. V. Balachandran, *Journal of The American Ceramic Society* **103**, 4489 (2020).
- [241] R. M. Balabin and E. I. Lomakina, *Journal of Chemical Physics* **131**, 074104 (2009).
- [242] F. Liu, C. Duan, and H. J. Kulik, *Journal of Physical Chemistry Letters* **11**, 8067 (2020).
- [243] S. Bag, A. Aggarwal, and P. K. Maiti, *Journal of Physical Chemistry A* **124**, 7658 (2020).
- [244] M. Veit, D. M. Wilkins, Y. Yang, R. A. DiStasio, and M. Ceriotti, *Journal of Chemical Physics* **153**, 024113 (2020).
- [245] P. C. St. John, Y. Guan, Y. Kim, S. Kim, and R. S. Paton, *Nature communications* **11**, 2328 (2020).
- [246] T. Gao, H. Li, W. Li, L. Li, C. Fang, H. Li, L. Hu, Y. Lu, and Z.-M. Su, *Journal of cheminformatics* **8**, 24 (2016).
- [247] Q. Zhang, F. Zheng, R. Fartaria, D. A. Latino, X. Qu, T. Campos, T. Zhao, and J. Aires-de-Sousa, *Chemo-metrics and Intelligent Laboratory Systems* **134**, 158 (2014).
- [248] X. Qu, D. A. Latino, and J. Aires-de-Sousa, *Journal of cheminformatics* **5**, 34 (2013).
- [249] W. Han, Y. Li, Q. Sheng, and J. Tang, *Recent Advances in Scientific Computing and Applications* **586**, 171 (2013).
- [250] H. Z. Li, L. H. Hu, W. Tao, T. Gao, H. Li, Y. H. Lu, and Z. M. Su, *IJMS* **13**, 8051 (2012).
- [251] R. Ramakrishnan, M. Hartmann, E. Tapavicza, and O. A. von Lilienfeld, *Journal of Chemical Physics* **143**, 084111 (2015).



- [252] M. Shaikh, M. Karmakar, and S. Ghosh, *Physical Review B* **101**, 054101 (2020).
- [253] S. Gugler, J. P. Janet, and H. J. Kulik, *Molecular Systems Design & Engineering* **5**, 139 (2020).
- [254] D. Jha, K. Choudhary, F. Tavazza, W.-k. Liao, A. Choudhary, C. Campbell, and A. Agrawal, *Nature communications* **10**, 5316 (2019).
- [255] C.-I. Wang, M. K. E. Braza, G. C. Claudio, R. B. Nellas, and C.-P. Hsu, *Journal of Physical Chemistry A* **123**, 7792 (2019).
- [256] W. Li, W. Miao, J. Cui, C. Fang, S. Su, H. Li, L. Hu, Y. Lu, and G. Chen, *Journal of Chemical Information and Modeling* **59**, 1849 (2019).
- [257] D. M. Wilkins, A. Grisafi, Y. Yang, K. U. Lao, R. A. DiStasio, and M. Ceriotti, *Proc Natl Acad Sci USA* **116**, 3401 (2019).
- [258] E. Iype and S. Urolagin, *Journal of Chemical Physics* **150**, 024307 (2019).
- [259] F. Pereira and J. Aires-de-Sousa, *Journal of cheminformatics* **10**, 43 (2018).
- [260] W. Pronobis, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, *The European Physical Journal B: Condensed Matter and Complex Systems* **91**, 178 (2018).
- [261] T. S. Hy, S. Trivedi, H. Pan, B. M. Anderson, and R. Kondor, *Journal of Chemical Physics* **148**, 241745 (2018).
- [262] R. Wang, *Journal of Physical Chemistry C* **122**, 8868 (2018).
- [263] P. Bleiziffer, K. Schaller, and S. Riniker, *Journal of Chemical Information and Modeling* **58**, 579 (2018).
- [264] F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley, and O. A. von Lilienfeld, *Journal of Chemical Theory and Computation* **13**, 5255 (2017).
- [265] J. P. Janet and H. J. Kulik, *Chemical Science* **8**, 5137 (2017).
- [266] F. Pereira, K. Xiao, D. A. R. S. Latino, C. Wu, Q. Zhang, and J. Aires-de-Sousa, *Journal of Chemical Information and Modeling* **57**, 11 (2016).
- [267] G. Li, Z. Hu, F. Hou, X. Li, L. Wang, and X. Zhang, *Fuel* **265**, 116968 (2020).
- [268] D.-J. Liu, J. W. Evans, P. M. Spurgeon, and P. A. Thiel, *Journal of Chemical Physics* **152**, 224706 (2020).
- [269] V. Venkatraman and B. Alsberg, *Polymers* **10**, 103 (2018).
- [270] M. Eckhoff, K. N. Lausch, P. E. Blöchl, and J. Behler, *Journal of Chemical Physics* **153**, 164107 (2020).
- [271] C. Duan, F. Liu, A. Nandy, and H. J. Kulik, *Journal of Physical Chemistry Letters* **11**, 6640 (2020).
- [272] Q. Zhang, F. Zheng, T. Zhao, X. Qu, and J. Aires-de-Sousa, *Molecular Informatics* **35**, 62 (2015).
- [273] M. Tsubaki and T. Mizoguchi, *Journal of Physical Chemistry Letters* **9**, 5733 (2018).
- [274] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, *Journal of Chemical Theory and Computation* **11**, 2087 (2015).
- [275] R. A. Eremin, P. N. Zolotarev, A. A. Golov, N. A. Nekrasova, and T. Leisegang, *Journal of Physical Chemistry C* **123**, 29533 (2019).
- [276] H. Wu, A. Lorenson, B. Anderson, L. Wittman, H. Wu, B. Meredig, and D. Morgan, *Systematic Organisation of Information in Fuzzy Systems* **134**, 160 (2017).
- [277] R. Juneja and A. K. Singh, *Journal of Materials Chemistry A: Materials for Energy and Sustainability* **8**, 8716 (2020).
- [278] A. Palizhati, W. Zhong, K. Tran, S. Back, and Z. W. Ulissi, *Journal of Chemical Information and Modeling* **59**, 4742 (2019).
- [279] T. D. Rhone, W. Chen, S. Desai, S. B. Torrisi, D. T. Larson, A. Yacoby, and E. Kaxiras, *Scientific Reports* **10**, 15795 (2020).
- [280] A. A. Icazatti, J. M. Loyola, I. Szleifer, J. A. Vila, and O. A. Martin, *PeerJ* **7**, e7904 (2019).
- [281] A. Navarro-Vázquez, *Magnetic resonance in chemistry : MRC* **59**, 414 (2020).
- [282] Z. Chaker, M. Salanne, J.-M. Delaye, and T. Charpentier, *Physical Chemistry Chemical Physics* **21**, 21709 (2019).
- [283] F. M. Paruzzo, A. Hofstetter, F. Musil, S. De, M. Ceriotti, and L. Emsley, *Nature communications* **9**, 4501 (2018).
- [284] P. V. Balachandran, D. Xue, and T. Lookman, *Physical Review B* **93**, 144111 (2016).
- [285] P. V. Balachandran, *Journal of Materials Research* **35**, 890 (2020).
- [286] A. Seko, T. Maekawa, K. Tsuda, and I. Tanaka, *Physical Review B* **89**, 054303 (2014).
- [287] D. C. Elton, Z. Boukouvalas, M. S. Butrico, M. D. Fuge, and P. W. Chung, *Scientific reports* **8**, 9059 (2018).
- [288] T. Morawietz and J. Behler, *Journal of Physical Chemistry A* **117**, 7356 (2013).
- [289] M. Babar, H. L. Parks, G. Houchins, and V. Viswanathan, *Journal of Physics: Energy* **3**, 014005 (2020).
- [290] C. Hong, J. M. Choi, W. Jeong, S. Kang, S. Ju, K. Lee, J. Jung, Y. Youn, and S. Han, *Physical Review B* **102**, 224104 (2020).
- [291] T. E. Gartner, L. Zhang, P. M. Piaggi, R. Car, A. Z. Panagiotopoulos, and P. G. Debenedetti, *Proc Natl Acad Sci USA* **117**, 26040 (2020).
- [292] F. Wu, H. Min, Y. Wen, R. Chen, Y. Zhao, M. Ford, and B. Shan, *Physical Review B* **102**, 144107 (2020).
- [293] M. Stricker, B. Yin, E. Mak, and W. A. Curtin, *Physical Review Materials* **4**, 103602 (2020).
- [294] G. Houchins and V. Viswanathan, *Journal of Chemical Physics* **153**, 054124 (2020).
- [295] D. Lee, K. Lee, D. Yoo, W. Jeong, and S. Han, *Systematic Organisation of Information in Fuzzy Systems* **181**, 109725 (2020).
- [296] A. Rodriguez, Y. Liu, and M. Hu, *Physical Review B* **102**, 035203 (2020).
- [297] K. K. Rao, Y. Yao, and L. C. Grabow, *Adv. Theory Simul.* **3**, 2000097 (2020).
- [298] M. C. Groenenboom, T. P. Moffat, and K. A. Schwarz, *Journal of Physical Chemistry C* **124**, 12359 (2020).
- [299] R. Gaillac, S. Chibani, and F.-X. Coudert, *Chemistry of Materials* **32**, 2653 (2020).
- [300] Q. Hu, M. Weng, X. Chen, S. Li, F. Pan, and L.-W. Wang, *Journal of Physical Chemistry Letters* **11**, 1364 (2020).
- [301] A. Thorn, J. Rojas-Nunez, S. Hajinazar, S. E. Baltazar, and A. N. Kolmogorov, *Journal of Physical Chemistry C* **123**, 30088 (2019).
- [302] T. Wen, C.-Z. Wang, M. J. Kramer, Y. Sun, B. Ye, H. Wang, X. Liu, C. Zhang, F. Zhang, K.-M. Ho, *et al.*, *Physical Review B* **100**, 174101 (2019).
- [303] H.-Y. Ko, L. Zhang, B. Santra, H. Wang, W. E. R. A. DiStasio Jr, and R. Car, *Molecular Physics* **117**, 3269 (2020).

- (2019).
- [304] G. Sosso and M. Bernasconi, *Mrs Bulletin* **44**, 705 (2019).
  - [305] E. Meshkov, I. Novoselov, A. Shapeev, and A. Yanilkin, *Intermetallics* **112**, 106542 (2019).
  - [306] R. Galvelis, S. Doerr, J. M. Damas, M. J. Harvey, and G. De Fabritiis, *Journal of Chemical Information and Modeling* **59**, 3485 (2019).
  - [307] J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev, and A. E. Roitberg, *Nature communications* **10**, 2903 (2019).
  - [308] M. Eckhoff and J. Behler, *Journal of Chemical Theory and Computation* **15**, 3793 (2019).
  - [309] G. P. P. Pun, R. Batra, R. Ramprasad, and Y. Mishin, *Nature communications* **10**, 2339 (2019).
  - [310] Y. Huang, J. Kang, W. A. Goddard, and L.-W. Wang, *Physical Review B* **99**, 064103 (2019).
  - [311] J. Kang, S. H. Noh, J. Hwang, H. Chun, H. Kim, and B. Han, *Physical Chemistry Chemical Physics* **20**, 24539 (2018).
  - [312] S. Rostami, M. Amsler, and S. A. Ghasemi, *Journal of Chemical Physics* **149**, 124106 (2018).
  - [313] Q. Liu, X. Zhou, L. Zhou, Y. Zhang, X. Luo, H. Guo, and B. Jiang, *Journal of Physical Chemistry C* **122**, 1761 (2018).
  - [314] W. Li, Y. Ando, E. Minamitani, and S. Watanabe, *Journal of Chemical Physics* **147**, 214106 (2017).
  - [315] T. Gao and J. R. Kitchin, *Catalysis Today* **312**, 132 (2018).
  - [316] H. Wang, Y. Zhang, L. Zhang, and H. Wang, *Frontiers in Chemistry* **8**, 589795 (2020).
  - [317] W. Liang, G. Lu, and J. Yu, *Adv. Theory Simul.* **3**, 2000180 (2020).
  - [318] N. Wang and S. Huang, *Physical Review B* **102**, 094111 (2020).
  - [319] Y. Nagai, M. Okumura, K. Kobayashi, and M. Shiga, *Physical Review B* **102**, 041124 (2020).
  - [320] G. C. Sosso, G. Miceli, S. Caravati, J. Behler, and M. Bernasconi, *Physical Review B* **85**, 174103 (2012).
  - [321] S. Gabardi, E. Baldi, E. Bosoni, D. Campi, S. Caravati, G. C. Sosso, J. Behler, and M. Bernasconi, *Journal of Physical Chemistry C* **121**, 23827 (2017).
  - [322] G. Pan, P. Chen, H. Yan, and Y. Lu, *Systematic Organisation of Information in Fuzzy Systems* **185**, 109955 (2020).
  - [323] G. Pan, J. Ding, Y. Du, D.-J. Lee, and Y. Lu, *Systematic Organisation of Information in Fuzzy Systems* **187**, 110055 (2021).
  - [324] V. Korolev, A. Mitrofanov, Y. Kucherinenko, Y. Nevolin, V. Krotov, and P. Protsenko, *Scripta Materialia* **186**, 14 (2020).
  - [325] K. Nigussa, *Materials Chemistry and Physics* **253**, 123407 (2020).
  - [326] P. Pattnaik, S. Raghunathan, T. Kalluri, P. Bhimalapuram, C. V. Jawahar, and U. D. Priyakumar, *Journal of Physical Chemistry A* **124**, 6954 (2020).
  - [327] P. Rowe, G. Csányi, D. Alfè, and A. Michaelides, *Physical Review B* **97**, 2469 (2018).
  - [328] A. P. Bartók, M. J. Gillan, F. R. Manby, and G. Csányi, *Physical Review B* **88**, 054104 (2013).
  - [329] K. Konstantinou, J. Mavračić, F. C. Mocanu, and S. R. Elliott, *Physica Status Solidi B* **258**, 2000416 (2020).
  - [330] S. Tovey, A. Narayanan Krishnamoorthy, G. Sivaraman, J. Guo, C. Benmore, A. Heuer, and C. Holm, *Journal of Physical Chemistry C* **124**, 25760 (2020).
  - [331] V. L. Deringer, M. A. Caro, and G. Csányi, *Nature communications* **11**, 5461 (2020).
  - [332] Y.-B. Liu, J.-Y. Yang, G.-M. Xin, L.-H. Liu, G. Csányi, and B.-Y. Cao, *Journal of Chemical Physics* **153**, 144501 (2020).
  - [333] F. L. Thiemann, P. Rowe, E. A. Müller, and A. Michaelides, *Journal of Physical Chemistry C* **124**, 22278 (2020).
  - [334] P. Rowe, V. L. Deringer, P. Gasparotto, G. Csányi, and A. Michaelides, *Journal of Chemical Physics* **153**, 034702 (2020).
  - [335] M. J. Gillan, D. Alfè, A. P. Bartók, and G. Csányi, *Journal of Chemical Physics* **139**, 244504 (2013).
  - [336] C. Zhang and Q. Sun, *Journal of Applied Physics* **126**, 105103 (2019).
  - [337] Q. Tong, L. Xue, J. Lv, Y. Wang, and Y. Ma, *Faraday Discussions* **211**, 31 (2018).
  - [338] F. C. Mocanu, K. Konstantinou, T. H. Lee, N. Bernstein, V. L. Deringer, G. Csányi, and S. R. Elliott, *Journal of Physical Chemistry B* **122**, 8998 (2018).
  - [339] V. L. Deringer, C. Merlet, Y. Hu, T. H. Lee, J. A. Kattirtzi, O. Pecher, G. Csányi, S. R. Elliott, and C. P. Grey, *Chemical Communications* **54**, 5988 (2018).
  - [340] D. Dragoni, T. D. Daff, G. Csányi, and N. Marzari, *Physical Review Materials* **2**, 013808 (2018).
  - [341] V. L. Deringer, G. Csányi, and D. M. Proserpio, *Chemphyschem : a European journal of chemical physics and physical chemistry* **18**, 873 (2017).
  - [342] V. L. Deringer and G. Csányi, *Physical Review B* **95**, 094203 (2017).
  - [343] J.-X. Huang, G. Csányi, J.-B. Zhao, J. Cheng, and V. L. Deringer, *Journal of Materials Chemistry A: Materials for Energy and Sustainability* **7**, 19070 (2019).
  - [344] N. Bernstein, G. Csányi, and V. L. Deringer, *npj computational materials* **5**, 99 (2019).
  - [345] A. V. Shapeev, *Multiscale Modeling & Simulation. A SIAM Interdisciplinary Journal* **14**, 1153 (2016).
  - [346] B. Mortazavi, E. V. Podryabinkin, I. S. Novikov, T. Rabczuk, X. Zhuang, and A. V. Shapeev, *Computer Physics Communications* **258**, 107583 (2021).
  - [347] B. Mortazavi, E. V. Podryabinkin, S. Roche, T. Rabczuk, X. Zhuang, and A. V. Shapeev, *Materials Horizons* **7**, 2359 (2020).
  - [348] E. V. Podryabinkin, E. V. Tikhonov, A. V. Shapeev, and A. R. Oganov, *Physical Review B* **99**, 064114 (2019).
  - [349] C. Wang, K. Aoyagi, M. Aykol, and T. Mueller, *ACS Applied Materials & Interfaces* **12**, 55510 (2020).
  - [350] E. V. Podryabinkin and A. V. Shapeev, *Systematic Organisation of Information in Fuzzy Systems* **140**, 171 (2017).
  - [351] K. Gubaev, E. V. Podryabinkin, and A. V. Shapeev, *Journal of Chemical Physics* **148**, 241727 (2018).
  - [352] B. Grabowski, Y. Ikeda, P. Srinivasan, F. Körmann, C. Freysoldt, A. I. Duff, A. Shapeev, and J. Neugebauer, *npj computational materials* **5**, 80 (2019).
  - [353] M. Jafary-Zadeh, K. H. Khoo, R. Laskowski, P. S. Branicio, and A. V. Shapeev, *Journal of Alloys and Compounds* **803**, 1054 (2019).
  - [354] V. Ladygin, P. Korotaev, A. Yanilkin, and A. Shapeev, *Systematic Organisation of Information in Fuzzy Systems* **172**, 109333 (2020).

- [355] I. Novoselov, A. Yanilkin, A. Shapeev, and E. Podryabinkin, *Systematic Organisation of Information in Fuzzy Systems* **164**, 46 (2019).
- [356] K. Gubaev, E. V. Podryabinkin, G. L. Hart, and A. V. Shapeev, *Systematic Organisation of Information in Fuzzy Systems* **156**, 148 (2019).
- [357] P. Korotaev, I. Novoselov, A. Yanilkin, and A. Shapeev, *Physical Review B* **100**, 144308 (2019).
- [358] C. Wang, K. Aoyagi, P. Wisesa, and T. Mueller, *Chemistry of Materials* **32**, 3741 (2020).
- [359] I. S. Novikov and A. V. Shapeev, *Materials Today Communications* **18**, 74 (2019).
- [360] I. S. Novikov, A. V. Shapeev, and Y. V. Suleimanov, *Journal of Chemical Physics* **151**, 224105 (2019).
- [361] I. S. Novikov, Y. V. Suleimanov, and A. V. Shapeev, *Physical Chemistry Chemical Physics* **20**, 29503 (2018).
- [362] B. Mortazavi, I. S. Novikov, E. V. Podryabinkin, S. Roche, T. Rabczuk, A. V. Shapeev, and X. Zhuang, *Applied Materials Today* **20**, 100685 (2020).
- [363] B. Mortazavi, E. V. Podryabinkin, I. S. Novikov, S. Roche, T. Rabczuk, X. Zhuang, and A. V. Shapeev, *Journal of Physics: Materials* **3**, 02LT02 (2020).
- [364] B. Mortazavi, F. Shojaei, M. Shahrokhi, M. Azizi, T. Rabczuk, A. V. Shapeev, and X. Zhuang, *Carbon* **167**, 40 (2020).
- [365] M. Raeisi, B. Mortazavi, E. V. Podryabinkin, F. Shojaei, X. Zhuang, and A. V. Shapeev, *Carbon* **167**, 51 (2020).
- [366] C. Zeni, K. Rossi, A. Glielmo, Á. Fekete, N. Gaston, F. Baletto, and A. De Vita, *Journal of Chemical Physics* **148**, 241739 (2018).
- [367] C. Chen, Z. Deng, R. Tran, H. Tang, I.-H. Chu, and S. P. Ong, *Physical Review Materials* **1**, 043603 (2017).
- [368] T. Nishiyama, A. Seko, and I. Tanaka, *Physical Review Materials* **4**, 123607 (2020).
- [369] J. Chapman and R. Ramprasad, *Jom-us.* **72**, 4346 (2020).
- [370] A. Pihlajamäki, J. Hämäläinen, J. Linja, P. Nieminen, S. Malola, T. Kärkkäinen, and H. Häkkinen, *Journal of Physical Chemistry A* **124**, 4827 (2020).
- [371] M. Rupp, M. R. Bauer, R. Wilcken, A. Lange, M. Reutlinger, F. M. Boeckler, and G. Schneider, *PLoS computational biology* **10**, e1003400 (2014).
- [372] A. Seko, A. Takahashi, and I. Tanaka, *Physical Review B* **92**, 054113 (2015).
- [373] X.-G. Li, C. Hu, C. Chen, Z. Deng, J. Luo, and S. P. Ong, *Physical Review B* **98**, 094104 (2018).
- [374] B. Narayanan, H. Chan, A. Kinaci, F. G. Sen, S. K. Gray, M. K. Y. Chan, and S. K. R. S. Sankaranarayanan, *Nanoscale* **9**, 18229 (2017).
- [375] K. Miwa and H. Ohno, *Physical Review B* **94**, 184109 (2016).
- [376] M. A. Wood, M. A. Cusentino, B. D. Wirth, and A. P. Thompson, *Physical Review B* **99**, 184305 (2019).
- [377] M. A. Cusentino, M. A. Wood, and A. P. Thompson, *Journal of Physical Chemistry A* **124**, 5456 (2020).
- [378] M. A. Wood and A. P. Thompson, *Journal of Chemical Physics* **148**, 241721 (2018).
- [379] Z. Deng, C. Chen, X.-G. Li, and S. P. Ong, *npj computational materials* **5**, 75 (2019).
- [380] J. T. Willman, A. S. Williams, K. Nguyen-Cong, A. P. Thompson, M. A. Wood, A. B. Belonoshko, and I. I. Oleynik, in *SHOCK COMPRESSION OF CONDENSED MATTER - 2019: Proceedings of the Conference of the American Physical Society Topical Group on Shock Compression of Condensed Matter* (AIP Publishing, Portland, OR, USA, 2020) p. 070055.
- [381] A. Seko, A. Takahashi, and I. Tanaka, *Physical Review B* **90**, 024101 (2014).
- [382] H. E. Saucedo, M. Gastegger, S. Chmiela, K.-R. Müller, and A. Tkatchenko, *Journal of Chemical Physics* **153**, 124109 (2020).
- [383] H. E. Saucedo, S. Chmiela, I. Poltavsky, K.-R. Müller, and A. Tkatchenko, *Journal of Chemical Physics* **150**, 114102 (2019).
- [384] S. Chmiela, H. E. Saucedo, I. Poltavsky, K.-R. Müller, and A. Tkatchenko, *Computer Physics Communications* **240**, 38 (2019).
- [385] S. Chmiela, A. Tkatchenko, H. E. Saucedo, I. Poltavsky, K. T. Schütt, and K.-R. Müller, *Science Advances* **3**, e1603015 (2017).
- [386] S. Chmiela, H. E. Saucedo, K.-R. Müller, and A. Tkatchenko, *Nature communications* **9**, 3887 (2018).
- [387] R. Lot, F. Pellegrini, Y. Shaidu, and E. Küçükbenli, *Computer Physics Communications* **256**, 107402 (2020).
- [388] J. S. Smith, O. Isayev, and A. E. Roitberg, *Chemical Science* **8**, 3192 (2017).
- [389] K. Yao, J. E. Herr, D. W. Toth, R. Mckintyre, and J. Parkhill, *Chemical Science* **9**, 2261 (2018).
- [390] A. Khorshidi and A. A. Peterson, *Computer Physics Communications* **207**, 310 (2016).
- [391] I. S. Novikov, K. Gubaev, E. V. Podryabinkin, and A. V. Shapeev, *Machine Learning: Science and Technology* **2**, 025002 (2021).
- [392] N. Bernstein, G. Csanyi, and J. Kermode, “QUIP,” <https://github.com/libAtoms/QUIP> (2019).
- [393] S. Plimpton, *Journal of Computational Physics* **117**, 1 (1995).
- [394] L. Zhang, D.-Y. Lin, H. Wang, R. Car, and W. E, *Physical Review Materials* **3**, 023804 (2019).
- [395] M. Wen and E. B. Tadmor, *npj computational materials* **6**, 124 (2020).
- [396] Q. Liu, J. Wang, P. Du, L. Hu, X. Zheng, and G. Chen, *Journal of Physical Chemistry A* **121**, 7273 (2017).
- [397] X. Lei and A. J. Medford, *Physical Review Materials* **3** (2019), 10.1103/physrevmaterials.3.063801.
- [398] Y. Chen, L. Zhang, H. Wang, and W. E, arXiv:2012.14615 [physics] (2020), arXiv:2012.14615 [physics].
- [399] J. T. Margraf and K. Reuter, *Nature communications* **12**, 344 (2021).
- [400] J. Nelson, R. Tiwari, and S. Sanvito, *Physical Review B* **99**, 075132 (2019).
- [401] J. Schmidt, C. L. Benavides-Riveros, and M. A. L. Marques, *Journal of Physical Chemistry Letters* **10**, 6425 (2019).
- [402] Y. Suzuki, R. Nagai, and J. Haruyama, *Physical Review A: Atomic, Molecular, and Optical Physics* **101**, 050501 (2020).
- [403] M. Yu, S. Yang, C. Wu, and N. Marom, *npj computational materials* **6**, 180 (2020).
- [404] M. Bogojeski, L. Vogt-Maranto, M. E. Tuckerman, K.-R. Müller, and K. Burke, *Nature communications* **11**, 5223 (2020).
- [405] C. D. Griego, L. Zhao, K. Saravanan, and J. A. Keith, *AIChE journal. American Institute of Chemical Engineers* **66** (2020), 10.1002/aic.17041.

- [406] O. Egorova, R. Hafizi, D. C. Woods, and G. M. Day, *Journal of Physical Chemistry A* **124**, 8065 (2020).
- [407] S. Dick and M. Fernandez-Serra, *Nature communications* **11**, 3509 (2020).
- [408] M. Fujinami, R. Kageyama, J. Seino, Y. Iwabata, and H. Nakai, *Chemical Physics Letters* **748**, 137358 (2020).
- [409] R. Nagai, R. Akashi, and O. Sugino, *npj computational materials* **6**, 43 (2020).
- [410] P. D. Mezei and O. A. von Lilienfeld, *Journal of Chemical Theory and Computation* **16**, 2647 (2020).
- [411] L. C. Lentz and A. M. Kolpak, *Journal of Physics: Condensed Matter* **32**, 155901 (2020).
- [412] G. Sun and P. Sautet, *Journal of Chemical Theory and Computation* **15**, 5614 (2019).
- [413] J. Zhu, V. Q. Vuong, B. G. Sumpter, and S. Irle, *MRS Communications* **9**, 867 (2019).
- [414] F. Peccati, E. Desmedt, and J. Contreras-García, *Computational and Theoretical Chemistry* **1159**, 23 (2019).
- [415] J. Hollingsworth, T. E. Baker, and K. Burke, *Journal of Chemical Physics* **148**, 241743 (2018).
- [416] T. Vegge, R. Christensen, C. Dickens, and H. Hansen, American Chemical Society. Abstracts of Papers (at the National Meeting) **255** (2018).
- [417] L. Messina, A. Quaglini, A. Goryaeva, M.-C. Marinica, C. Domain, N. Castin, G. Bonny, and R. Krause, *Nuclear Instruments & Methods in Physics Research, Section B: Beam Interactions with Materials and Atoms* **483**, 15 (2020).
- [418] A. D. Becke, *Physical Review A: Atomic, Molecular, and Optical Physics* **38**, 3098 (1988).
- [419] M. Chiba, T. Tsuneda, and K. Hirao, *Journal of Chemical Physics* **124**, 144106 (2006).
- [420] J.-W. Song, T. Tsuneda, T. Sato, and K. Hirao, *Organic Letters* **12**, 1440 (2010).
- [421] Y. Tawada, T. Tsuneda, S. Yanagisawa, T. Yanai, and K. Hirao, *Journal of Chemical Physics* **120**, 8425 (2004).
- [422] H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao, *Journal of Chemical Physics* **115**, 3540 (2001).
- [423] B. C. Yeo, D. Kim, C. Kim, and S. S. Han, *Scientific reports* **9**, 5879 (2019).
- [424] C. Ben Mahmoud, A. Anelli, G. Csányi, and M. Ceriotti, *Physical Review B* **102**, 235130 (2020).
- [425] F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, *Nature communications* **8**, 2041 (2017).
- [426] M. Tsubaki and T. Mizoguchi, *Physical Review Letters* **125**, 206401 (2020).
- [427] M. Eickenberg, G. Exarchakis, M. Hirn, S. Mallat, and L. Thiry, *Journal of Chemical Physics* **148**, 241732 (2018).
- [428] E. Schmidt, A. T. Fowler, J. A. Elliott, and P. D. Bristowe, *Systematic Organisation of Information in Fuzzy Systems* **149**, 250 (2018).
- [429] J. M. Alred, K. V. Bets, Y. Xie, and B. I. Yakobson, *Composites Science and Technology* **166**, 3 (2018).
- [430] A. Chandrasekaran, D. Kamal, R. Batra, C. Kim, L. Chen, and R. Ramprasad, *npj computational materials* **5**, 22 (2019).
- [431] J. A. Ellis, L. Fiedler, G. A. Popoola, N. A. Modine, J. A. Stephens, A. P. Thompson, A. Cangi, and S. Rajamanickam, *Physical Review B* **104**, 035120 (2021).
- [432] S. Dick and M. Fernandez-Serra, *Journal of Chemical Physics* **151**, 144102 (2019).
- [433] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke, *Physical Review Letters* **108**, 1079 (2012).
- [434] L. Li, T. E. Baker, S. R. White, and K. Burke, *Physical Review B* **94**, 245129 (2016).
- [435] G. Hegde and R. C. Bowen, *Scientific reports* **7**, 2045 (2017).
- [436] A. R. Ferreira, *Physical Review Materials* **4**, 113603 (2020).
- [437] C. Panosetti, A. Engelmann, L. Nemec, K. Reuter, and J. T. Margraf, *Journal of Chemical Theory and Computation* **16**, 2181 (2020).
- [438] J. Ku, A. Kamath, T. Carrington, and S. Manzhos, *Journal of Physical Chemistry A* **123**, 10631 (2019).
- [439] O. Schütt and J. VandeVondele, *Journal of Chemical Theory and Computation* **14**, 4168 (2018).
- [440] T. L. Jacobsen, M. S. Jørgensen, and B. Hammer, *Physical Review Letters* **120**, 026102 (2018).
- [441] P. Gao, J. Zhang, Q. Peng, J. Zhang, and V.-A. Glezhakou, *Journal of Chemical Information and Modeling* **60**, 3746 (2020).
- [442] A. J. Samin, *Journal of Applied Physics* **127**, 175904 (2020).
- [443] V. Venkatraman, S. Abburu, and B. K. Alsberg, *Chemometrics and Intelligent Laboratory Systems* **142**, 87 (2015).
- [444] J. Zhang, K. Terayama, M. Sumita, K. Yoshizoe, K. Ito, J. Kikuchi, and K. Tsuda, *Science and Technology of Advanced Materials* **21**, 552 (2020).
- [445] F. Schmidt, J. Wenzel, N. Halland, S. Güssregen, L. Delafoy, and A. Czich, *Chemical Research in Toxicology* **32**, 2338 (2019).
- [446] K. Choudhary and F. Tavazza, *Systematic Organisation of Information in Fuzzy Systems* **161**, 300 (2019).
- [447] N. C. Frey, J. Wang, G. I. Vega Bellido, B. Anasori, Y. Gogotsi, and V. B. Shenoy, *ACS Nano* **13**, 3031 (2019).
- [448] G. Takasao, T. Wada, A. Thakur, P. Chammingkwan, M. Terano, and T. Taniike, *Acs Catalysis* **9**, 2599 (2019).
- [449] M. Sumita, X. Yang, S. Ishihara, R. Tamura, and K. Tsuda, *ACS Central Science* **4**, 1126 (2018).
- [450] J. J. Grefenstette, in *Proceedings of the Sixth Annual Conference on Computational Learning Theory - COLT '93*, Vol. 25 (ACM Press, San Francisco, California, United States, 1993).
- [451] L. Tian and W. Yu, *Systematic Organisation of Information in Fuzzy Systems* **186**, 110050 (2021).
- [452] C. J. Bartel, A. Trewartha, Q. Wang, A. Dunn, A. Jain, and G. Ceder, *npj computational materials* **6** (2020), 10.1038/s41524-020-00362-y.
- [453] S. Agarwal, S. Mehta, and K. Joshi, *New Journal of Chemistry* **44**, 8545 (2020).
- [454] F. Hou, Z. Wu, Z. Hu, Z. Xiao, L. Wang, X. Zhang, and G. Li, *Journal of Physical Chemistry A* **122**, 9128 (2018).
- [455] E. M. Collins and K. Raghavachari, *Journal of Chemical Theory and Computation* **16**, 4938 (2020).
- [456] S. Gusarov, S. R. Stoyanov, and S. Siahrostami, *Journal of Physical Chemistry C* **124**, 10079 (2020).
- [457] D. Dickel, D. Francis, and C. Barrett, *Systematic Organisation of Information in Fuzzy Systems* **171**, 109157 (2020).
- [458] S. Laghuvarapu, Y. Pathak, and U. D. Priyakumar, *Journal of computational chemistry* **41**, 790 (2019).



- [459] H. Sahu, W. Rao, A. Troisi, and H. Ma, *Advanced Energy Materials* **8**, 1801032 (2018).
- [460] K. Choudhary, B. DeCost, and F. Tavazza, *Physical Review Materials* **2**, 083801 (2018).
- [461] H. Li, W. Li, X. Pan, J. Huang, T. Gao, L. Hu, H. Li, and Y. Lu, *Journal of Chemometrics* **32**, e3023 (2018).
- [462] R. Jalem, M. Nakayama, Y. Noda, T. Le, I. Takeuchi, Y. Tateyama, and H. Yamazaki, *Science and Technology of Advanced Materials* **19**, 231 (2018).
- [463] M. Deimel, K. Reuter, and M. Andersen, *Acs Catalysis* **10**, 13729 (2020).
- [464] J. Townsend, C. P. Micucci, J. H. Hymel, V. Maroulas, and K. D. Vogiatzis, *Nature communications* **11**, 3230 (2020).
- [465] A. P. Bartók, R. Kondor, and G. Csányi, *Physical Review B* **87**, 184115 (2013).
- [466] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, *Physical Review Letters* **108**, 058301 (2012).
- [467] K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. von Lilienfeld, K.-R. Müller, and A. Tkatchenko, *Journal of Physical Chemistry Letters* **6**, 2326 (2015).
- [468] F. A. Faber, A. S. Christensen, B. Huang, and O. A. von Lilienfeld, *Journal of Chemical Physics* **148**, 241717 (2018).
- [469] R. Drautz, *Physical Review B* **99**, 014104 (2019).
- [470] Y. Lysogorskiy, C. van der Oord, A. Bochkarev, S. Menon, M. Rinaldi, T. Hammerschmidt, M. Mrovec, A. Thompson, G. Csányi, C. Ortner, *et al.*, *npj computational materials* **7**, 1 (2021).
- [471] R. Tamura, J. Lin, and T. Miyazaki, *Journal of The Physical Society of Japan* **88**, 044601 (2019).
- [472] A. A. Peterson, R. Christensen, and A. Khorshidi, *Physical Chemistry Chemical Physics* **19**, 10978 (2017).
- [473] L. Talirz, L. M. Ghiringhelli, and B. Smit, *Living Journal of Computational Molecular Science* **3**, 1 (2021).
- [474] G. Kresse and J. Furthmüller, *Physical Review B* **54**, 11169 (1996).
- [475] G. Kresse and J. Furthmüller, *Systematic Organisation of Information in Fuzzy Systems* **6**, 15 (1996).
- [476] G. Kresse and J. Hafner, *Physical Review B* **47**, 558 (1993).
- [477] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, *et al.*, *Journal of Physics: Condensed Matter* **21**, 395502 (2009).
- [478] P. Giannozzi, O. Basergio, P. Bonfà, D. Brunato, R. Car, I. Carnimeo, C. Cavazzoni, S. de Gironcoli, P. Delugas, F. Ferrari Ruffino, *et al.*, *Journal of Chemical Physics* **152**, 154105 (2020).
- [479] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. Buongiorno Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, *et al.*, *Journal of Physics: Condensed Matter* **29**, 465901 (2017).
- [480] P. Blaha, K. Schwarz, G. K. Madsen, D. Kvasnicka, J. Luitz, *et al.*, “Wien2k: An augmented plane wave+ local orbitals program for calculating crystal properties,” (2001).
- [481] P. Blaha, K. Schwarz, F. Tran, R. Laskowski, G. K. H. Madsen, and L. D. Marks, *Journal of Chemical Physics* **152**, 074101 (2020).
- [482] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, and M. C. Payne, *Zeitschrift für Kristallographie - Crystalline Materials* **220**, 567 (2005).
- [483] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, *et al.*, “Gaussian 16 Revision C.01,” (2016), gaussian Inc. Wallingford CT.
- [484] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, *et al.*, *Journal of Computational Chemistry* **14**, 1347 (1993).
- [485] M. S. Gordon and M. W. Schmidt, in *Theory and Applications of Computational Chemistry* (Elsevier, 2005) pp. 1167–1189.
- [486] F. Neese, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2**, 73 (2011).
- [487] F. Neese, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **8** (2017), 10.1002/wcms.1327.
- [488] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, *WIREs Computational Molecular Science* **2**, 242 (2012), <https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.82>.
- [489] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györfy, D. Kats, T. Korona, R. Lindh, *et al.*, “MOLPRO, version 2019.2, a package of ab initio programs,” (2019).
- [490] P. Knowles and N. Handy, *Chemical Physics Letters* **111**, 315 (1984).
- [491] H.-J. Werner and P. J. Knowles, *Journal of Chemical Physics* **82**, 5053 (1985).
- [492] P. J. Knowles and H.-J. Werner, *Chemical Physics Letters* **115**, 259 (1985).
- [493] H.-J. Werner and P. J. Knowles, *Journal of Chemical Physics* **89**, 5803 (1988).
- [494] P. J. Knowles and H.-J. Werner, *Chemical Physics Letters* **145**, 514 (1988).
- [495] R. D. Amos, J. S. Andrews, N. C. Handy, and P. J. Knowles, *Chemical Physics Letters* **185**, 256 (1991).
- [496] P. J. Knowles, J. S. Andrews, R. D. Amos, N. C. Handy, and J. A. Pople, *Chemical Physics Letters* **186**, 130 (1991).
- [497] P. J. Knowles and H.-J. Werner, *Theoretica chimica acta* **84**, 95 (1992).
- [498] P. J. Knowles, C. Hampel, and H.-J. Werner, *Journal of Chemical Physics* **99**, 5219 (1993).
- [499] P. J. Knowles and N. C. Handy, *Computer Physics Communications* **54**, 75 (1989).
- [500] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, *et al.*, in *Advances in Neural Information Processing Systems*, Vol. 32 (Curran Associates, Inc., 2019).
- [501] W. Falcon and The PyTorch Lightning team, “PyTorch Lightning,” (2019).
- [502] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, *et al.*, arXiv:1603.04467 (2016), arXiv:1603.04467 [cs.DC].
- [503] F. Chollet *et al.*, “Keras,” <https://github.com/fchollet/keras> (2015).
- [504] M. Innes, E. Saba, K. Fischer, D. Gandhi, M. C. Rudilosso, N. M. Joy, T. Karmali, A. Pal, and V. Shah, *CoRR abs/1811.01457* (2018), arXiv:1811.01457.
- [505] M. Innes, *Journal of Open Source Software* (2018), 10.21105/joss.00602.
- [506] S. Sonnenburg, G. Rätsch, S. Henschel, C. Widmer, J. Behr, A. Zien, F. d. Bona, A. Binder, C. Gehl, and

- V. Franc, The Journal of Machine Learning Research **11**, 1799 (2010).
- [507] G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa, and A. Mueller, *GetMobile: Mobile Comp. and Comm.* **19**, 29 (2015).
- [508] I. H. Witten and E. Frank, *Acm Sigmod Record* **31**, 76 (2002).
- [509] P. Upchurch, J. Gardner, G. Pleiss, R. Pless, N. Snavely, K. Bala, and K. Weinberger, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 31 (IEEE, 2017).
- [510] A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Duřak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, *et al.*, *Journal of Physics: Condensed Matter* **29**, 273002 (2017).
- [511] S. P. Huber, S. Zoupanos, M. Uhrin, L. Talirz, L. Kahle, R. Häuselmann, D. Gresch, T. Müller, A. V. Yakutovich, C. W. Andersen, *et al.*, *Scientific data* **7**, 1 (2020).
- [512] M. Uhrin, S. P. Huber, J. Yu, N. Marzari, and G. Pizzi, *Computational Materials Science* **187**, 110086 (2021).
- [513] B. Ramsundar, P. Eastman, P. Walters, V. Pande, K. Leswing, and Z. Wu, *Deep Learning for the Life Sciences* (O'Reilly Media, 2019) <https://www.amazon.com/Deep-Learning-Life-Sciences-Microscopy/dp/1492039837>.
- [514] Z. Wu, B. Ramsundar, E. N. Feinberg, J. Gomes, C. Geniesse, A. S. Pappu, K. Leswing, and V. Pande, *Chemical Science* **9**, 513 (2018).
- [515] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, *Jom-us.* **65**, 1501 (2013).
- [516] M. Nakata and T. Shimazaki, *Journal of Chemical Information and Modeling* **57**, 1300 (2017).
- [517] C. Draxl and M. Scheffler, *Mrs Bulletin* **43**, 676 (2018).
- [518] K. Choudhary, K. F. Garritty, A. C. E. Reid, B. DeCost, A. J. Biacchi, A. R. Hight Walker, Z. Trautt, J. Hattrick-Simpers, A. G. Kusne, A. Centrone, *et al.*, *npj computational materials* **6**, 173 (2020).
- [519] Clarivate, “Web Of Science,” <http://www.webofknowledge.com> (1997).
- [520] Google, “Google Scholar,” <https://scholar.google.com/> (2004).
- [521] L. Fiedler, K. Shah, A. Cangi, and M. Bussmann, “Dataset and scripts for A Deep Dive into Machine Learning Density Functional Theory for Materials Science and Chemistry,” (2021).