

# Machine Learning the Thermodynamic Stability of Crystal Structures

## Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften

(Dr. rer. nat.)

der

Naturwissenschaftliche Fakultät II

Chemie, Physik und Mathematik

vorgelegt von

Herrn

**Jonathan Schmidt**

geboren am 30.10.1995 in Aachen

Erstgutachter: Prof. Dr. Miguel Marques  
Zweitgutachter: Prof. Dr. Michele Ceriotti  
Drittgutachter: Prof. Dr. Ingrid Mertig

Verteidigt am 5.10.2022

## Abstract

In recent years machine learning is quickly becoming one of the most valuable tools in solid-state material science. In this cumulative thesis we develop two applications of machine learning to solid-state physics and material science. First, we train machine learning exchange-correlation functionals for density functional theory (DFT). In contrast to earlier works we not only train for the exchange-correlation energy but also use automatic differentiation to train for the correct potential as functional derivative of the neural network. Such neural network functionals can be extremely nonlocal while retaining the same computational scaling of local and semilocal functionals and promise to solve some of the non-locality problems plaguing DFT.

We also develop Crystal-Graph Attention networks (CGAT) for the prediction of thermodynamically stable materials. Previous generations of graph neural networks typically use the atomic positions and the atomic species as input. However, one can only obtain the atomic positions of the relaxed crystal structures during high-throughput searches via DFT calculations. Thus making it impossible to apply these networks directly to such studies. We solve this challenge by replacing the atomic distances with embeddings of the graph distances to create networks suitable for high-throughput studies. To train these networks we accumulate and clean one of the largest datasets of DFT calculations with consistent parameters. Applying the dataset and the new network topology to high-throughput searches we have scanned a compound space of over 1 billion materials. As a result we have already discovered more than 23000 materials that are stable relative to the convex hull we started with. Using the hull resulting from our original dataset we perform more than 200k geometry optimizations with

the Perdew-Burke-Ernzerhofer functional for solids (PBEsol) and single point calculations with the "Strongly Constrained and Appropriately Normed" (SCAN) functional to obtain a convex hull and structural information from functionals beyond the PBE. These will allow for a more accurate prediction of stable materials and their properties in the future.



## Abstract

In den letzten Jahren hat sich das maschinelle Lernen schnell zu einem der wertvollsten Werkzeuge in der Festkörpermateriewissenschaft entwickelt. In dieser kumulativen Dissertation entwickeln wir zwei Anwendungen des maschinellen Lernens für die Festkörperphysik und Materialwissenschaft. Erstens trainieren wir mit maschinellem Lernen Austauschkorrelationsfunktionale für Dichtefunktionaltheorie (DFT). Im Gegensatz zu früheren Arbeiten, trainieren wir nicht nur die Austausch-Korrelations-Energie, sondern verwenden automatische Differentiation, um gleichzeitig das korrekte Potential als funktionale Ableitung des neuronalen Netzes zu trainieren. Solche neuronalen Netzwerkfunktionale können extrem nichtlokal sein, während sie rechenkostentechnisch wie lokale und semilokale Funktionale skalieren. Deshalb haben sie das Potential einige der Nichtlokalitätsprobleme der Dichtefunktionaltheorie zu lösen.

Zweitens trainieren, entwickeln, und implementieren wir Crystal-Graph Attention Networks (CGAT) für die Vorhersage von thermodynamisch stabilen Materialien. Frühere Generationen von Graph-Neuronalen Netzwerken verwenden in der Regel die Atompositionen und die Elemente als Eingabe. Die Atompositionen der optimierten Kristallstrukturen können jedoch nur über DFT-Berechnungen ermittelt werden. Dies macht es unmöglich, die Netzwerke direkt auf solche Studien anzuwenden. Wir lösen diese Herausforderung, indem wir die atomaren Abstände durch Repräsentationen der Graphabstände ersetzen, um Netzwerke zu schaffen, die für Hochdurchsatzstudien geeignet sind. Um diese Netzwerke zu trainieren, sammeln und kuratieren wir einen der größten Datensätze von DFT-Berechnungen mit konsistenten Parametern. Durch die Anwendung des Datensatzes und der neuen Netzwerktopologie konnten wir mehr als eine Milliarde Materialien auf ihre Stabilität un-

tersuchen. Als Resultat dieser Hochdurchsatzstudie haben wir bereits mehr als 23000 Materialien entdeckt, die relativ zu der konvexen Hülle, mit der wir begonnen haben, stabil sind. Unter Verwendung der Hülle, die sich aus unserem ursprünglichen Datensatz ergibt, führen wir mehr als 200k Kristallstrukturoptimierungen mit dem Perdew-Burke-Ernzerhofer-Funktional für Festkörper (PBEsol) und Einzelpunktberechnungen mit dem "Strongly Constrained and Appropriately Normed"-Funktional (SCAN) durch, um eine konvexe Hülle und Strukturinformationen von Funktionalen jenseits des PBE zu erhalten. Diese Daten werden in Zukunft eine genauere Vorhersage von stabilen Materialien und ihren Eigenschaften ermöglichen.



# Contents

<b>1</b>	<b>Introduction</b>	<b>7</b>
<b>2</b>	<b>Theoretical Foundation</b>	<b>10</b>
2.1	Many-Body Problem . . . . .	10
2.2	Density Functional Theory . . . . .	11
2.3	Exchange-Correlation Functionals . . . . .	13
2.4	Plane Wave basis and Projector-Augmented-Wave Method	17
2.5	Thermodynamic Stability . . . . .	23
<b>3</b>	<b>Machine Learning</b>	<b>27</b>
3.1	Neural Networks . . . . .	27
3.1.1	Fully Connected Neural Networks . . . . .	28
3.1.2	Hypernetworks . . . . .	31
3.1.3	Message Passing Neural Networks . . . . .	32
3.1.4	Attention Mechanism for Graphs . . . . .	37
3.2	Machine Learning Material Properties . . . . .	42
3.3	Neural Network Functionals . . . . .	46
<b>4</b>	<b>Results and Discussion</b>	<b>50</b>
4.1	Machine Learning the Physical Nonlocal Exchange-Correlation Functional of Density-Functional Theory . . . . .	50
4.2	Crystal-Graph Attention Networks for the Prediction of Stable Materials . . . . .	60
4.3	A Dataset of 175k Stable and Metastable Materials Cal- culated with the PBEsol and SCAN Functionals . . . . .	76
<b>5</b>	<b>Conclusion and Outlook</b>	<b>86</b>
5.1	Outlook . . . . .	88

**List of Publications**

**91**

**Bibliography**

**93**



# 1 Introduction

In the last 15 years, the field of machine learning and its applications in other domains have seen an unprecedented increase in interest. This revolution is a result of the increase in computing power, especially in graphic processing units, the development of new algorithms and the accumulation of large datasets that allow for the application of these algorithms. Machine learning has proven to have superhuman abilities in multiple fields from computer games [12, 164], over chess, Go [144] and self-driving cars [15] to image classification [60] and the prediction of protein folding [72]. While chemistry [114] and genetics [6] already have a long history of data-driven research, theoretical solid-state physics and material science have only recently started to embrace this new paradigm [JSPHD9]. Two of the most active research topics in this field are the discovery of new materials and the development of machine learning density functionals [JSPHD9, JSPHD6].

## Material Discovery

The design of new compounds is one of the principal factors driving progress in material science. Traditionally, experiments based on human ingenuity and even randomness have played the most prominent part in the discovery of new materials [129]. The first computer revolution changed this by introducing computational high-throughput studies based on ab initio methods [105]. These already resulted in a significant speedup for the development of new materials and technologies. However, the first-principle methods, mainly DFT, are still computationally expensive, and the vast majority of DFT calculations is quickly forgotten as they result in unstable or useless materials.

Machine learning allows us to use this forgotten data and speed up the exploration of the chemical space by several orders of magnitude [JSPHD9,

117]. Machine learning models that use the compositions and/or structures as input allow for the direct prediction of the stability and other properties of the materials. The prediction capabilities of these algorithms depend first on the model itself and second on the training data.

Concerning models, graph neural networks based on the atomic positions provide the most precise predictions of material properties [7, 32]. However, these models are not applicable to high-throughput searches of materials as only a crystal structure prototype but not the exact positions are available until DFT calculations are performed [JSPHD6]. Less accurate composition-based models do not present an alternative as they cannot differentiate between polymorphs with the same composition but different structures. In this work, we aim to solve this challenge by proposing a message passing network that relies on imprecise structure representations as input and can therefore be used efficiently in prototype-based high-throughput searches.

We will also work on creating larger and higher fidelity DFT datasets to improve the performance of the machine learning models and the precision of stability predictions via DFT. The accuracy of DFT calculations rests mostly on the quality of the functional that is used. All large scale solid-state databases ( $\geq 10$ k materials) still use the PBE functional [119] developed a quarter-century ago. We will address this problem and create a database of 250k materials calculated with SCAN [152] and PBEsol [121]. These modern functionals profit from decades of functional research and provide more accurate energies, crystal structures and band gaps for solids [178, JSPHD1, 177, JSPHD5].

### **Machine Learning Functionals**

While newer approximations such as SCAN show a significant improvement over old functionals, DFT still suffers from its classical fail-



ure points, in particular the delocalization errors and the locality of most functionals. The number of empirical functionals has already increased in the last decade [92] and neural network functionals have a long history [157]. In recent years more and more machine learning functionals have been suggested. However, most of them either arrive at inaccurate potentials [146] or do not calculate them as functional derivative of the energy functional [111]. At the same time, automatic differentiation frameworks have made significant advances [1, 118]. One of the goals of this thesis will be to use these advances to simultaneously train a neural network exchange-correlation functional and its functional derivative with respect to the density and demonstrate the feasibility of this approach for one-dimensional systems. Furthermore, in contrast to traditional approximations, neural networks should allow us to easily vary the locality of the functional and research different degrees of locality. We will attempt to demonstrate that machine learning functionals have the potential to overcome some of the deficiencies of classical exchange-correlation functionals.

## 2 Theoretical Foundation

We aim to replace or speed up the calculation of precise energies and other properties of crystal structures with machine learning methods. To calculate these energies or other properties we have to solve the many-body problem. In the next section, we will discuss this problem and DFT as an approach to solve it.

### 2.1 Many-Body Problem

The quantum mechanical many-body problem is well-defined by the Schrödinger equation in atomic units:

$$\hat{H}\psi = E\psi, \quad (1)$$

with the following many-body Hamiltonian:

$$\hat{H} = - \sum_j \frac{\nabla_j^2}{2M_j} - \sum_i \frac{\nabla_i^2}{2} + \sum_{j \neq i} \frac{Z_j Z_i}{|R_j - R_i|} + \sum_{j \neq i} \frac{1}{|r_j - r_i|} - \sum_{j,i} \frac{Z_j}{|R_j - r_i|}. \quad (2)$$

Here  $R_j$ ,  $M_j$ ,  $Z_j$  and  $r_i$  denote respectively the coordinates, masses and charges of the nuclei and the coordinates of the electrons. The first two terms of the Hamiltonian are the kinetic energy of the nuclei and the kinetic energy of the electrons. This is followed by the electrostatic energy of the nuclei and the electrons and the electrostatic energy of the nuclei interacting with the electrons.

Commonly, this equation is simplified through the Born-Oppenheimer approximation [16] that separates the wave functions of nuclei and electrons based on the fact that nuclei are at least  $\sim 3$  orders of magnitude heavier than electrons.

$$\psi(R_i, r_i) = \gamma(R_i) \phi_{R_i}(r_i) \quad (3)$$

Unfortunately, even for the resulting simplified electronic Hamiltonian:

$$\hat{H}_e = - \sum_i \frac{\nabla_i^2}{2} + \sum_{j \neq i} \frac{1}{|r_j - r_i|} - \sum_{j,i} \frac{Z_j}{|R_j - r_i|}, \quad (4)$$

no analytic solutions are known for larger systems and it can be easily understood that standard numerical solutions for more than a few particles are impossible. The memory alone to save the solution of a 10 particle problem on a 10 by 10 grid in 3 dimensions would be of the order of  $10^{21}$  gigabytes prohibiting this approach.

## 2.2 Density Functional Theory

Fortunately, Hohenberg and Kohn circumvented this challenge through the development of DFT [62]. DFT replaces the wavefunction as the central object of quantum mechanics with the electronic density:

$$n(r) = N \int dr_2 \dots \int dr_N \psi(r, \dots, r_N) \psi^*(r, \dots, r_N). \quad (5)$$

This replacement is made possible by the Hohenberg-Kohn theorem [62] that we can divide into two parts. The first one states that a one-to-one correspondence between the electronic density and the external potential of a system and consequently its Hamiltonian exists. The second part proves that the ground state electronic density minimizes the energy of a system, which is the variational principle used in DFT. We can write the energy functional as a sum of the energy of an external potential  $v_{\text{ext}}$  and a universal functional  $F[n(r)]$ :

$$E[n(r)] = \int dr n(r) v_{\text{ext}}(r) + F[n(r)], \quad (6)$$

where the latter can be denoted as the kinetic energy plus the Coulomb interaction of the electrons:

$$F[n(r)] = T[n(r)] + U[n(r)]. \quad (7)$$

We can efficiently solve the variational problem by switching to the Kohn-Sham formalism [82]. Here we reduce the many-body problem to a non-interacting system with an effective potential that leads to the same ground state density as the interacting system. Based on this assumption, we arrive at the Kohn-Sham equations:

$$\left( -\frac{\nabla^2}{2} + v_{\text{eff}}([n(r)], r) \right) \Psi_i = E_i \Psi_i. \quad (8)$$

The effective potential  $v_{\text{eff}}([n(r)], r)$  is defined as follows:

$$v_{\text{eff}}([n(r)], r) = v_{\text{ext}}(r) + v_{\text{H}}([n(r)], r) + v_{\text{xc}}([n(r)], r). \quad (9)$$

Here  $v_{\text{ext}}(r)$  is the external potential, parameterized by the positions of the nuclei that were fixed through the Born-Oppenheimer approximation. The Hartree potential is denoted as:

$$v_{\text{H}}([n(r)], r) = \int dr' \frac{n(r')}{|r - r'|}, \quad (10)$$

and

$$v_{\text{xc}}([n(r)], r) = \frac{\delta E_{\text{xc}}}{\delta n(r)} \quad (11)$$

is the exchange-correlation potential that results as the functional derivative of the exchange-correlation energy. The latter potential includes all remaining many-particle interactions. Since two of the potentials are functionals of the electronic density, we have to solve the Kohn-Sham equations in a self-consistent manner. The atomic density can be used as a starting guess to calculate the exchange-correlation and the Coulomb potential. Following the solution of the equations, we can obtain a new density from the resulting Kohn-Sham orbitals:

$$n(r) = \sum_i^N |\psi_i(r)|^2 \quad (12)$$

and use this density to recalculate the potentials. Then we solve the equations again, repeating this process until convergence.

The total energy of the Kohn-Sham system is written as:

$$E_{\text{total}} = T_{\text{KS}}[n(r)] + \int dr n(r) v_{\text{ext}}(r) + E_{\text{H}}[n(r)] + E_{\text{xc}}[n(r)] \quad (13)$$

where  $T_{\text{KS}}[n(r)]$  is the kinetic energy of the non-interacting electrons and  $E_{\text{H}}$  the electrostatic repulsion between electrons:

$$E_{\text{H}}[n(r)] = \frac{1}{2} \int dr \int dr' \frac{n(r) n(r')}{|r - r'|}. \quad (14)$$

While we know the form of the Hartree terms, the same is not true for the exchange-correlation energy. Improving the approximations for the exchange-correlation energy  $E_{\text{xc}}[n(r)]$ , that we only know exactly for simple systems such as the Fermi gas, has been the major challenge since the advent of DFT. Although the problem is not solved, a number of successful approximations have been developed.

### 2.3 Exchange-Correlation Functionals

The simplest of these is the local density approximation:

$$E_{\text{xc}}^{\text{LDA}} = \int \varepsilon_{\text{xc}}[n(r)] n(r) dr. \quad (15)$$

Here  $\varepsilon_{\text{xc}}$  is the exchange-correlation energy per particle of a homogeneous electron gas with charge density  $n(r)$ . Usually we decompose the exchange-correlation energy into an exchange and a correlation energy:

$$E_{\text{xc}} = E_{\text{x}} + E_{\text{c}}. \quad (16)$$

Naturally, we can extend the definition of this functional, and of DFT in general, to consider spin, which results in the local-spin density approximation:

$$E_{\text{xc}}^{\text{LSDA}} = \int (\varepsilon_{\text{x}}[n_{\uparrow}(r), n_{\downarrow}(r)] n(r) + \varepsilon_{\text{c}}[n_{\uparrow}(r), n_{\downarrow}(r)] n(r)) dr. \quad (17)$$

The LSDA represents the first step on Jacob’s ladder (see fig. 1) a hierarchy of exchange-correlation approximations suggested by John Perdew in 2001 [122].



Figure 1: Jacob’s ladder representing the different levels of accuracy of the increasingly more complex categories of exchange-correlation approximations.

Generalized gradient approximations (GGAs) take up the next step on the ladder. In addition to the local density, GGAs also include the gradient of the density:

$$E_{xc}^{\text{GGA}}[n_{\uparrow}, n_{\downarrow}] = \int dr n(r) e_{xc}^{\text{GGA}}(n_{\uparrow}, n_{\downarrow}, \nabla n_{\uparrow}, \nabla n_{\downarrow}). \quad (18)$$

GGAs are the most used functionals in solid-state physics, with the Perdew-Burke-Ernzerhof approximation (PBE) being the foremost func-

tional [119].

Meta-GGAs reach even one step closer to the heaven of chemical accuracy by adding the kinetic energy density  $\tau(r)$ , and  $\nabla^2 n(r)$  to their functional form. As the calculation of  $\tau(r)$  requires the calculation of the derivatives of the Kohn-Sham orbitals, the potentially better accuracy comes at a higher computational cost. The most successful meta-GGA is the strongly constrained and appropriately normed (SCAN) functional [152] that was developed based on all known exact constraints for meta-GGAs.

The next rung is taken up by functionals using information from the occupied orbitals. The most common example are hybrid functionals [11, 150, 120] that combine a portion of the exchange energy from Hartree-Fock with other functionals. The Hartree-Fock exchange energy is calculated as a four-center integral of the orbitals. Consequently, hybrids are nonlocal functionals and scale even slower than meta-GGAs.

Double hybrids extend this concept by including interactions between occupied and unoccupied orbitals. For instance, the PT2 family of functionals [109, 53] adds a correlation term corresponding to the single and double excitations in second order perturbation theory.

In total, more than 500 exchange-correlation functionals have been proposed in the past fifty years [92] even though the majority of them had a very limited impact.

Functionals in general play a major role in this thesis. First, we attempt to skip Jacob's ladder to the heaven of chemical accuracy by taking a "machine learning elevator". We will demonstrate that it is possible to train neural network functionals with a physically correct exchange-correlation potential, that is, the functional derivative of the energy with respect to the density. Furthermore, we show that these

functionals can be extremely nonlocal while maintaining the scaling of traditionally local functionals.

Secondly, we improve existing PBE solid-state datasets by recalculating them with the PBEsol [121] and the SCAN functional [152]. In the following, we introduce the details of these two functionals and the machinery behind solid-state DFT calculations.

### PBEsol and SCAN

Both the PBE, the PBEsol as well as the SCAN functional consider some exact conditions. The PBE functional fulfills 5 conditions for the exchange part [35]:

- exact for uniform densities
- correct scaling of the functional under a uniform density scaling  $n_\lambda(r) = n(\lambda r)$ :

$$E_x [n_\lambda(r)] = \lambda E_x [n(r)] \quad (19)$$

- spin scaling relationship:

$$E_x [n_\downarrow(r), n_\uparrow(r)] = \frac{1}{2} (E_x [2n_\uparrow(r)] + E_x [2n_\downarrow(r)]) \quad (20)$$

- linear response of the LDA for small densities
- local Lieb-Oxford bound [35]:

$$E_x [n(r)] \geq E_{xc}^{\lambda=1} [n(r)] \geq 2.273 E_x^{\text{LDA}} \quad (21)$$

and 3 conditions for the correlation part:

- exact second-order gradient expansion in the slowly varying limit  $\frac{|\nabla n(r)|}{n(r)} \rightarrow 0$
- in the rapidly varying limit  $\frac{|\nabla n(r)|}{n(r)} \rightarrow \infty$  the correlation energy should vanish



- negative constant under uniform scaling to the high-density limit.

Based on these theoretical conditions Perdew, Burke and Ernzerhof fixed the four parameters of the PBE making it fully ab-initio. While it uses the same formula, PBEsol selects different exact conditions leading to changed parameter values. Specifically, PBEsol fixes the density gradient expansion of the exchange resulting in one different parameter and making it impossible to satisfy the gradient expansion for the correlation. A second parameter is also changed, violating the linear response condition but allowing for better surface energies. With these changes, we arrive at a functional that corrects most of the underbinding of PBE, resulting in far superior lattice constants [JSPHD5]. We will later use this fact to obtain more precise ground state geometries while retaining the computational cost of a GGA. The third and last functional used in this thesis is the SCAN [152] functional, a meta-GGA. SCAN was developed by satisfying all 17 exact constraints for meta-GGAs and fulfilling several norms from uniform and slow-varying densities to various gas atoms and the  $Z \rightarrow \infty$  limit of two-electron ions. As a result, it provides lattice constants that are en par with PBEsol [177, 151] while yielding superior formation energies [178, 151] and slightly improved band gaps [JSPHD1]. However, we have to note that SCAN is plagued by some numerical issues that can lead to difficulties converging calculations [38, 8].

## 2.4 Plane Wave basis and Projector-Augmented-Wave Method

While it is possible to solve the Kohn-Sham equations in real space and some codes follow this direction [154, 173], it is usually far more efficient

to represent the Kohn-Sham orbitals in a basis:

$$\phi_i(\mathbf{r}) = \sum_i c_i \chi_i(\mathbf{r}). \quad (22)$$

Here  $c_i$  are the expansion coefficients that are optimized during the SCF-method and  $\chi_i$  is a fixed set of basis functions. When using a basis, the eigenvalue problem transforms to a generalized eigenvalue problem:

$$(H - \epsilon_i S)C = 0 \quad (23)$$

where  $S$  is the overlap matrix of the basis, which vanishes only in the case of an orthogonal basis. In quantum chemistry, Gaussian type orbitals find the widest application as basis sets [36, 153], while solid-state codes are dominated by plane wave bases [85, 84, 45, 50, 31]. The reasons for this are quite simple if we consider the Born–von Kármán periodic boundary conditions of crystal structures:

$$\Psi_{nk}(r) = \Psi_{nk}(r + R) \quad (24)$$

and Bloch’s theorem:

$$\Psi_{nk}(\mathbf{r}) = u_{nk}(\mathbf{r}) \mathbf{e}^{i\mathbf{k}\mathbf{r}} \quad (25)$$

$$u_{nk}(\mathbf{r}) = u_{nk}(\mathbf{r} + \mathbf{R}). \quad (26)$$

We label the states by their Bloch vector  $\mathbf{k}$  and their band index  $n$ ,  $u_{nk}(\mathbf{r})$  are functions following the periodicity of the lattice.  $\mathbf{R}$  is a lattice translation vector composed from the lattice vectors  $a_i$  of the crystal structure  $R = \sum_i n_i a_i$ . The cell periodic functions  $u_{nk}(\mathbf{r})$  are usually expanded in plane waves:

$$u_{nk}(\mathbf{r}) = \frac{1}{\Omega^{\frac{1}{2}}} \sum_G c_{n\mathbf{G}\mathbf{k}} e^{i\mathbf{G}\mathbf{r}} \quad (27)$$

$$\Psi_{nk}(\mathbf{r}) = \frac{1}{\Omega^{\frac{1}{2}}} \sum_G c_{n\mathbf{G}+\mathbf{k}} e^{i(\mathbf{G}+\mathbf{k})\mathbf{r}}. \quad (28)$$

In practice only the plane waves below the energy cutoff  $\frac{1}{2}|G + k|^2 < E_{\text{cutoff}}$  are used which reduces the basis to a finite size. The integrals over the Brillouin zone are evaluated by selecting special sets of k-Points according to some scheme, e.g., the Monkhost-Pack scheme [108]. Plane wave basis sets have many more advantages over atomic basis sets, e.g.,:

- they are orthogonal,
- Laplace operators are diagonal,
- they allow for efficient transformation to real space with fast Fourier transforms,
- they are independent of nuclei positions,
- the precision is systematically improvable by increasing the energy cutoff.

However, the number of plane waves needed to describe strongly localized states or the oscillations of the wave functions near the nucleus quickly becomes impractical. To overcome this problem, we use the frozen core approximation. We calculate core electrons in an atomic environment, leave them frozen during the DFT calculation, and represent them by a pseudopotential. This has the second major advantage of massively reducing the number of electrons in DFT calculations. The main variants of pseudopotentials that are in use are norm-conserving [86] pseudopotentials, ultra-soft pseudopotentials [161], and the Projector-Augmented-Wave (PAW) method [13, 87, 131]. The latter combines ideas from both Vanderbilt-type ultrasoft pseudopotentials [161] and the linearized-augmented-plane-wave (LAPW) method [4]. The only major DFT-code used for this thesis was the Vienna Ab-Initio Simulation Package (VASP) [85, 84] that utilizes the PAW-method.

Therefore, we limit our discussion to this method. For the derivation of the PAW-method we follow the description in Ref. [131]. When using pseudopotentials the wave function can usually be divided into two regions: the interstitial region between atoms where the wave function is relatively smooth, and spheres around the atoms, where approximations are needed. In the PAW-method we use a linear transformation to go from the smooth pseudo-wavefunction  $|\tilde{\Psi}_n\rangle$  to the all-electron Kohn-Sham orbitals:

$$|\Psi\rangle_n = \hat{\mathcal{T}} |\tilde{\Psi}_n\rangle. \quad (29)$$

The linear transformation  $\hat{\mathcal{T}}$  needs to be defined in such a way, that it yields a transformed eigenvalue problem with smooth eigenvectors that are efficient for calculations:

$$\hat{\mathcal{T}}^\dagger \hat{H} \hat{\mathcal{T}} |\tilde{\Psi}\rangle_n = \epsilon_n \hat{\mathcal{T}}^\dagger \hat{\mathcal{T}} |\tilde{\Psi}_n\rangle. \quad (30)$$

We already expect the Kohn-Sham orbitals to be sufficiently smooth in the interstitial regions, therefore we only need to adjust them close to the cores. To simplify the process we can decompose the linear transformation into a set of atom centered transformations  $\hat{\mathcal{T}}_\alpha$  that vanish outside of a cutoff  $|r - R_\alpha| < r_\alpha^{\text{cutoff}}$  centered on each nucleus  $\alpha$ :

$$\hat{\mathcal{T}} = 1 + \sum_\alpha \hat{\mathcal{T}}_\alpha. \quad (31)$$

We name this cutoff augmentation spheres or regions. For each augmentation region, we define a set of partial waves  $|\Phi_i^\alpha\rangle$  and a corresponding set of smooth pseudo partial waves  $|\tilde{\Phi}_i^\alpha\rangle$ . The former is used to expand the all-electron orbitals and has to obey the following relation with the latter:

$$|\Phi_i^\alpha\rangle = (1 + \hat{\mathcal{T}}_\alpha) |\tilde{\Phi}_i^\alpha\rangle \iff \hat{\mathcal{T}}_\alpha |\tilde{\Phi}_i^\alpha\rangle = |\Phi_i^\alpha\rangle - |\tilde{\Phi}_i^\alpha\rangle. \quad (32)$$

As the transformations  $\hat{\mathcal{T}}_\alpha$  are supposed to vanish outside the core regions we arrive at the following:

$$\forall \alpha, i \quad \Phi_i^\alpha(r) = \tilde{\Phi}_i^\alpha(r) \quad \text{for } |r - R_\alpha| > r_\alpha^{\text{cutoff}}. \quad (33)$$

Assuming that each set  $|\tilde{\Phi}_i^\alpha\rangle$  is a complete basis inside its augmentation radius, each all-electron orbital  $|\Psi\rangle_n$  and its smoothed counterpart  $|\tilde{\Psi}_n\rangle$  can be expanded as:

$$|\tilde{\Psi}_n\rangle = \sum_i c_{ni}^\alpha |\tilde{\Phi}_i^\alpha\rangle, \quad \text{for } |r - R_\alpha| < r_\alpha^{\text{cutoff}} \quad (34)$$

$$|\Psi\rangle_n = \hat{\mathcal{T}} |\tilde{\Psi}_n\rangle = \sum_i c_{ni}^\alpha |\Phi_i^\alpha\rangle, \quad \text{for } |r - R_\alpha| < r_\alpha^{\text{cutoff}}. \quad (35)$$

The second equation follows directly from eq. (32).  $\hat{\mathcal{T}}$  is a linear operator and consequently the expansion coefficients are required to be linear in  $|\tilde{\Psi}_n\rangle$ . Therefore, they can be represented as:

$$c_{ni}^\alpha = \langle \tilde{p}_i^\alpha | \tilde{\Psi}_n \rangle. \quad (36)$$

$\tilde{p}_i^\alpha$  are fixed functions that have to be selected and are known as projector functions. The augmentation regions are defined such that there is no overlap between them. Hence, we obtain the following relations inside each augmentation sphere:

$$|\tilde{\Psi}_n\rangle = |\tilde{\Psi}_n^\alpha\rangle = \sum_i |\tilde{\phi}_i^\alpha\rangle \langle \tilde{p}_i^\alpha | \tilde{\Psi}_n \rangle \Rightarrow \sum_i |\tilde{\phi}_i^\alpha\rangle \langle \tilde{p}_i^\alpha | = 1. \quad (37)$$

Consequently, the projector functions and pseudo partial waves are orthonormal to each other inside the augmentation spheres. Using the above completeness relation we arrive at the formulas for the linear operators  $\hat{\mathcal{T}}_\alpha$  and the all-electron Kohn-Sham orbitals:

$$\hat{\mathcal{T}}_\alpha = \sum_i \left( |\Phi_i^\alpha\rangle - |\tilde{\Phi}_i^\alpha\rangle \right) \langle \tilde{p}_i^\alpha | \quad (38)$$

$$|\Psi_n\rangle = |\tilde{\Psi}_n\rangle + \sum_{\alpha i} \left( |\Phi_i^\alpha\rangle - |\tilde{\Phi}_i^\alpha\rangle \right) \langle \tilde{p}_i^\alpha | \tilde{\Psi}_n \rangle. \quad (39)$$

Through this transformation we separated the wave function into a smooth part for the interstitial region and rapidly oscillating parts inside the augmentation spheres that can both be treated efficiently.

By adding the frozen core approximation, we assume that the core electrons are localized in the augmentation spheres and fill their standard atomic orbitals. Consequently, we are only required to treat valence states with the PAW-method. Note that while we primarily use plane waves, the PAW-formalism leads to a non-diagonal overlap operator in the eigenvalue problem.

$|\Phi_i^\alpha\rangle$ ,  $|\tilde{\Phi}_i^\alpha\rangle$  and  $|\tilde{p}_i^\alpha\rangle$  are functions chosen under the previously defined restraints with the goal to be as efficient as possible. They are tabulated for each element, however different codes can vary considerably in their selection of functions. The partial waves  $|\Phi_i^\alpha\rangle$  are usually the eigenstates of the isolated non-spinpolarized atoms. Outside the augmentation spheres, the pseudo partial waves are equal to the partial waves. Inside the spheres we can choose any continuous continuation, e.g., Bessel functions in VASP or polynomials in GPAW. The projector functions are typically calculated according to:

$$|\tilde{p}_i^\alpha\rangle = \left( -\frac{1}{2}\nabla^2 + \tilde{v}_{\text{ks}} - \epsilon_i \right) |\tilde{\phi}_i^\alpha\rangle. \quad (40)$$

for each reference atom, followed by an orthonormalization procedure to enforce the completeness relation. Here  $\tilde{v}_{\text{ks}}$  is the smoothed Kohn-Sham potential that is obtained by applying  $\hat{\mathcal{T}}^\alpha$  to the Kohn-Sham Hamiltonian.

The PAW-method significantly complicates any DFT code as the calculation of most quantities changes due to the dual basis description. For instance, when calculating an energy resulting from an external potential we obtain atomic corrections as well as an additional term from

the pseudo density  $\tilde{n}(r)$ :

$$E_{ext} = \sum_{\alpha} \int dr (n_{\alpha}(r) - \tilde{n}^{\alpha}(r)) v_{ext}(r) + \int dr \tilde{n}(r) v_{ext}(r). \quad (41)$$

## 2.5 Thermodynamic Stability

In the last chapters, we introduced DFT and various exchange-correlation functionals. For us, the most important result of a DFT calculation is the energy of the system. The energy allows us to at least partially answer one of the key questions in this thesis, that is, whether a crystal structure is thermodynamically stable or not.

We define a crystal structure as thermodynamically stable if all possible decomposition channels are higher in energy. For a binary compound with composition  $A_xB_y$ , the conceivable decomposition channels are all linear combinations of elementary materials A, B, or binary compounds  $A_{\beta}B_{\alpha}$ , that result in the composition  $A_xB_y$ . If any of these linear combinations results in a lower energy than the compound, the compound will decompose into the lowest energy combination after a finite amount of time. Thermodynamic stability can be easily visualized and computed using the concept of a convex hull. In fig. 2 we show a convex hull for a binary system. All known stable materials are shown in blue and are part of the convex hull. Any new material (orange) that lies above the hull is noted as having a positive distance to the hull and will decompose into the linear combination that can be read from the hull diagram. Any stable material (pink) resulting from a new calculation will be added to the hull. While materials on the hull have a zero distance to the hull, we use a negative distance in practice. The negative distance is evaluated by removing the material from the hull and then calculating the distance to the remaining hull with the leftover compounds.

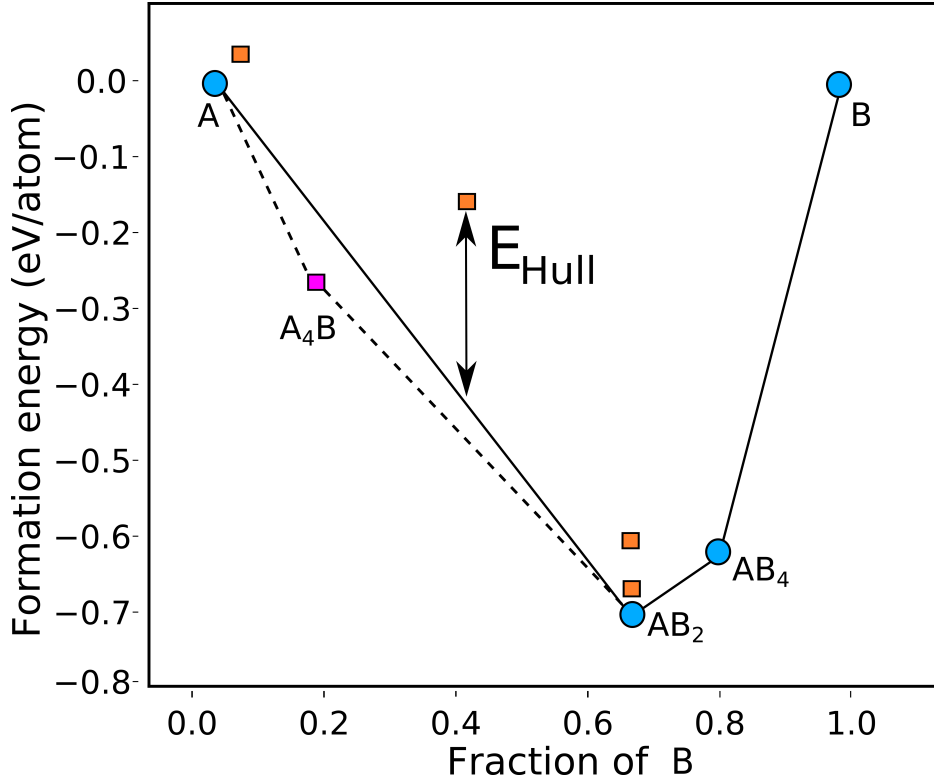


Figure 2: We note stable materials that are already part of the convex hull in blue. Orange materials are unstable and have a positive distance to the convex hull. When we add a new material to the hull (magenta), we need to adjust the hull appropriately. Modified graphic from Ref. [JSPHD6].

Necessarily, our knowledge of the convex hull is incomplete, approximate and constantly expanding. Incomplete, because we are only slowly discovering stable materials, approximate due to the errors associated with DFT and constantly changing and expanding through new high-throughput searches and other calculations. As we compare energies when creating the convex hull, we have to use compatible calculation parameters, e.g., the same exchange-correlation functional, pseudopotentials, etc.. Thus, depending on our data source, the completeness of the hull and the errors associated with the calculation parameters, differ.

The most common data source for calculating distances to the con-



vex hull is the Materials Project [68] which contains only 144k compounds. However, a relatively large number of these are close to stability, including many compounds from the International Crystal Structure Database [2]. Other sources are the Open Quantum Materials Database (OQMD) [133, 77], the Automatic FLOW for Materials Discovery database (AFLOW) [27, 116] as well as works from our group [JS4, JS3, JSPHD10, 166, JSPHD12]. All of these databases rely on the PBE functional [119] as exchange-correlation functional but deviate partially in other parameters. For instance, the OQMD and AFLOW employ different pseudopotentials and k-point grids in some calculations when compared with the Materials Project. In Ref. [JSPHD10] we approach this problem by filtering multiple databases for compatible calculations to obtain a larger consistent convex hull. We tackle the problem of limited accuracy in Ref. [JSPHD11] by creating a convex hull with SCAN calculations [JSPHD11].

Naturally, our goal is to use the convex hull to identify new stable materials and extend the hull through this process. Prototype-based high-throughput searches are one of the most efficient ways to discover new materials. In such a high-throughput search we start by selecting a known crystal structure prototype with a certain composition, e.g., *ABC*. Then we fill up the sites of the structure with all combinations of elemental species. For a ternary prototype, this process typically results in 250k to 500k structures depending on the number of elements that we consider. Traditionally, we calculate the energies of the resulting structures with DFT and then determine the stability of each structure. While this process is feasible for ternaries, it is computationally expensive and quickly becomes impossible for quaternary or quinary compounds.

We will discuss the application of machine learning methods to high-throughput searches in section 3.2. While machine learning has already shown some success to speed up the process, a tailor-made tool, using the latest progress in machine learning, is still missing. We try to overcome this challenge in Ref. [JSPHD10].

## 3 Machine Learning

The goal of machine learning algorithms is to optimize the performance of the algorithm for a specific task based on examples provided to the machine [3]. These examples can both be explicitly labeled data or past experience of the machine. In general, we can divide machine learning into four subcategories, supervised learning, unsupervised learning, self-supervised learning and reinforcement learning. In this thesis, we only employ supervised learning techniques, which are algorithms minimizing a prediction error on a dataset for which the target values/categories are known and provided to the machine. In our case, these tasks are the prediction of exchange-correlation energies and potentials in Ref. [JSPHD7] and the prediction of the thermodynamic stability of crystal structures in Ref. [JSPHD10]. There is a wide variety of supervised machine learning algorithms used in material science [JSPHD9], from support vector machines, decision tree algorithms, linear and kernel-based methods over Gaussian process regression to neural networks. We will limit ourselves to neural networks and discuss their foundations and specialized variants that we employed in this thesis in the following chapter.

### 3.1 Neural Networks

We start the section by introducing simple fully connected neural networks and their training process. These will lay the foundation for hypernetworks, and message passing networks in the context of theoretical material science.

### 3.1.1 Fully Connected Neural Networks

A machine learning algorithm has to start with a numeric representation of the input system. In simple cases this input is just a single feature vector. Fully connected neural networks (FCNN) transform such feature vectors in a non-linear fashion to obtain a prediction. We can reduce every FCNN to building blocks of single layers. A layer  $l_i$  consists of a number of neurons which can be represented by a vector  $z_i \in R^N$ , where  $N$  is the number of neurons in the layer. The neurons of each layer are connected with the following layer through a linear transformation, i.e. a matrix multiplication with the weight matrix  $A_{\text{weight}}^i$ , followed by a non-linear activation function. Often a bias vector is added before the activation function is applied:

$$z_{i+1} = f_{\text{activation}}(A_{\text{weight}}^i z_i + \text{bias}_i). \quad (42)$$

Chaining multiple layers together, we arrive at a FCNN. Commonly, we denote the first layer as input layer, the last layer as output layer and all layers in between as hidden layers.

The activation functions can be any form of non-linearity. Traditionally sigmoidal functions such as the hyperbolic tangent were used. However, as gradients of sigmoidal functions quickly saturate to zero resulting in the vanishing gradient problem [83] for deep neural networks they declined in popularity. Nowadays, rectified linear units [113, 49]:

$$\text{ReLU}(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \quad (43)$$

and various functions derived from ReLU such as leaky ReLU [104]:

$$\text{LeakyReLU}(x) = \begin{cases} x & \text{if } x > 0 \\ \epsilon * x & \text{otherwise,} \end{cases} \quad (44)$$

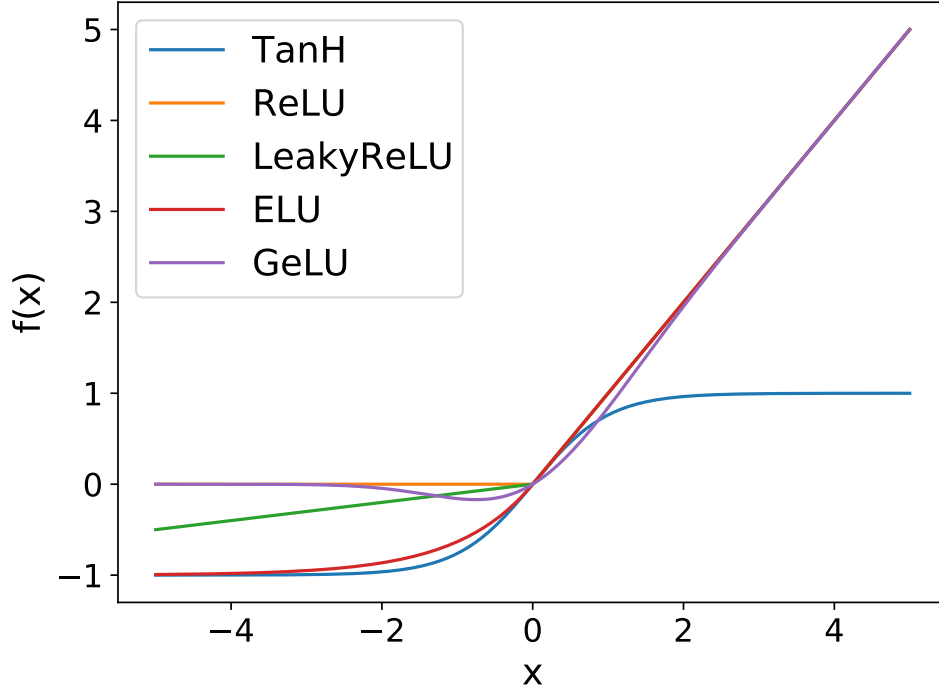


Figure 3: The following parameters were used: LeakyRelu  $\epsilon = 0.1$ , ELU  $\alpha=1.0$ .

exponential linear units (ELU) [25]:

$$\text{ELU}(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha(e^x - 1) & \text{otherwise} \end{cases} \quad (45)$$

or Gaussian error linear unit (GELU) [61]:

$$\text{GELU}(x) = \frac{x}{2} \left( 1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right) \right) \quad (46)$$

are the most common. Here  $\text{erf}(x)$  is the error function. In fig. 3 the main differences between the activation functions are displayed. The gradient of the traditional hyperbolic tangent saturates quickly for negative as well as positive values leading to the vanishing gradient problem. In contrast all the newer activation functions converge to  $f(x) = x$  on the positive side. The gradient of LeakyReLU in difference to ReLU, ELU and GELU does not saturate for negative values. However, ReLU

as well as LeakyReLU have a discontinuous derivative at zero. This needs to be considered in physics applications where potentials are often calculated as derivatives of a neural network.

The parameters of the weight matrices  $A_{\text{weight}}^i$  are usually randomly initialized according to normal or uniform distributions [48]. Naturally, these random parameters do not produce useful predictions as an output of the neural network. To optimize the performance of the neural networks we can apply the gradient descent algorithm to a suitable differentiable loss function, e.g., the mean-squared error:

$$L(y_i, x_i, \theta) = \sum_i (\text{FCNN}_\theta(x_i) - y_i)^2. \quad (47)$$

Here  $\theta$  are the parameters of the neural network contained in its weight matrices and bias vectors,  $x_i$  is the input representation of the  $i_{\text{th}}$  sample and  $y_i$  is the respective target. Following the gradient with respect to the parameters  $\theta$ :

$$\theta_{\text{new}} = \theta - \text{learning rate} \times \frac{\partial L}{\partial \theta} \quad (48)$$

leads to local optima of the loss function where we stop the minimization if we reach a sufficiently small validation error. The gradient steps are reduced by a prefactor called the learning rate. In practice, different variants of stochastic gradient descent are used, which means that the loss function and its gradient are only calculated for one batch of samples at a time followed by a change in the parameters. The most popular variants, i.e. optimizers, are Adam [76] and its corrected implementation AdamW [100] as well as basic stochastic gradient descent with momentum. Adam and AdamW are the main optimizers used in the thesis. The update for the  $j_{\text{th}}$  parameter at the  $t_{\text{th}}$  timestep in Adam

can be denoted as:

$$\delta w_j^t = -\frac{\text{learning rate}}{\sqrt{\frac{1}{1-\beta_2}\text{MA}(g_j^{t^2})}} \times \frac{1}{1-\beta_1}\text{MA}(g_j^t). \quad (49)$$

Here  $\text{MA}(g_j^t)$  and  $\text{MA}(g_j^{t^2})$  are the exponential moving averages of the gradient of the  $j$ th parameter with a decay of  $\beta_1$  and the squared gradient with a decay of  $\beta_2$ , respectively. In contrast to previous algorithms the averages are bias-corrected, i.e. corrected for the initialization bias assuming the moving averages are initialized to 0. Moreover, Adam uses not only an exponential moving average of the squared gradient but also of the gradient itself.

To easily obtain the derivatives of scalar loss functions modern machine learning frameworks such as pytorch [118] or tensorflow [1] use reverse-mode automatic differentiation. This allows for an efficient calculation of the derivative of an arbitrary combination of operations defined within the framework.

### 3.1.2 Hypernetworks

One of these combinations of operations is known as hypernetworks [56, 110]. We briefly introduce them as we employ them in Ref. [JSPhD10]. Hyper networks are neural networks that themselves return a neural network as output which then produces a normal output:

$$\text{HFCNN}_{\theta_H}(\mathbf{x}) = \theta_x \rightarrow \text{FCNN}_{\theta_x} \rightarrow \text{FCNN}_{\theta_x}(\mathbf{x}) \rightarrow \text{output}. \quad (50)$$

Here  $\theta_H$  are the weights of the hypernetwork and  $\theta_x$  the weights obtained as output for the input  $\mathbf{x}$  from the hypernetwork. While hypernetworks exist that output networks of differing topology depending on the input [80], we limit ourselves to hypernetworks that return weights for a fixed topology. Effectively, these are regular fully connected networks

that output a vector of the size of the combined weights of the resulting network. The hypernetworks are trained end-to-end with backpropagation, i.e. the derivative is taken both through the returned network  $\text{FCNN}_{\mathbf{x}}$  and through the hypernetwork  $\text{HFCNN}_{\theta_{\text{H}}}(\mathbf{x})$  with respect to the parameters of the hypernetwork:

$$\frac{\partial L}{\partial \theta_{\text{H}}} = \frac{\partial L}{\partial \theta_{\mathbf{x}}} \times \frac{\partial \theta_{\mathbf{x}}}{\partial \theta_{\text{H}}}. \quad (51)$$

While in most cases they require a larger number of parameters, in the ideal case, we can obtain a suitable network for each input when using a hypernetwork.

### 3.1.3 Message Passing Neural Networks

Message passing neural networks are a relatively general class of neural networks working on graph-like structures. The central idea of message passing networks is to update the representation of each vertex in a graph based on messages sent from neighboring vertices. We discuss message passing networks in the context of crystal and molecular graphs that is relevant to this thesis.

To represent a material as a graph, we denote the atoms as vertices and encode distance and bond information in the edges of the graph. In most cases, we consider undirected graphs. However, directional edge properties can easily be integrated into the message passing framework [79] and are required for some networks [137]. We should also note that some molecular graphs cannot be differentiated by message passing networks lacking angular edge information [127].

Each node  $h_i$  and edge  $e_{ij}$  connecting the nodes  $h_i$  and  $h_j$  are respectively initialized with an embedding vector  $\mathbf{h}_i^0 \in \mathbb{R}^m$  and  $\mathbf{e}_{ij}^0 \in \mathbb{R}^l$ . The starting node embedding  $\mathbf{h}_i^0$  is solely dependent on the atomic species.



In earlier works [172, 20], encodings of basic elemental properties such as the atomic mass were used as features. Later works employed randomly parameterized embeddings that were trained as part of the neural network [141]. An additional alternative are pretrained embeddings learned from text [159]. Of course, we can combine the latter two options by using a fixed pretrained embedding followed by a trainable linear layer. The edge embeddings can contain various information, from atomic distances or chemical bond features [20] to information based on Voronoi-Tesselations [117] or randomly initialized embeddings of the graph distance [JSPHD10].

While the starting embeddings are an important choice they only correspond to the feature vectors in simple FCNNs. In total, the input for a graph network consists of node embeddings, edge embeddings and an adjacency matrix or equivalent information. Once this representation of the graph enters the network, we update the node embeddings based on messages passed from neighbouring nodes to adapt the embeddings to their local graph environment  $\mathcal{N}(i)$ :

$$\mathbf{h}_i^{t+1} = U(\mathbf{h}_i^t, \{\mathbf{h}_j^t, \mathbf{e}_{ji}^t\}, j \in \mathcal{N}(i)). \quad (52)$$

$U$  is the update function of the message passing process [46], the only limitations being that it updates the representation of each node based on its previous embeddings as well as the embeddings of the neighbouring vertices and their connecting edges. While the form of the update function is usually constant through all message passing steps, we use unique weights to parameterize the update function at each step. Earlier works now classified as recurrent graph neural networks used the same weights for each update step [170]. Network topologies or updating functions can differ strongly in their strategy. Crystal graph

convolutions are one successful example [172]:

$$\mathbf{h}_i^{t+1} = \mathbf{h}_i^t + \sum_{j \in \mathcal{N}(i)} \sigma(\mathbf{z}_{i,j}^t \mathbf{W}_f + \mathbf{b}_f) \odot g(\mathbf{z}_{i,j}^t \mathbf{W}_s + \mathbf{b}_s). \quad (53)$$

For crystal structures the neighborhood  $\mathcal{N}(i)$  is usually defined by a cutoff radius.  $\mathbf{z}_{i,j}^t$  denotes the concatenation of the node embeddings  $\mathbf{h}_i^t, \mathbf{h}_j^t$  and the corresponding edge  $\mathbf{e}_{ij}$ ,  $\sigma$  and  $g$  are activation functions, and  $\mathbf{W}_{f/s}$  and  $\mathbf{b}_{f/s}$  are weight matrices and bias vectors, respectively. Effectively, the vector  $\mathbf{z}_{i,j}$  is processed by two single layer FCNNs, and the results are multiplied to create a message from each neighboring atom. Then we sum over the messages from the whole neighborhood to create the update for the node embedding.

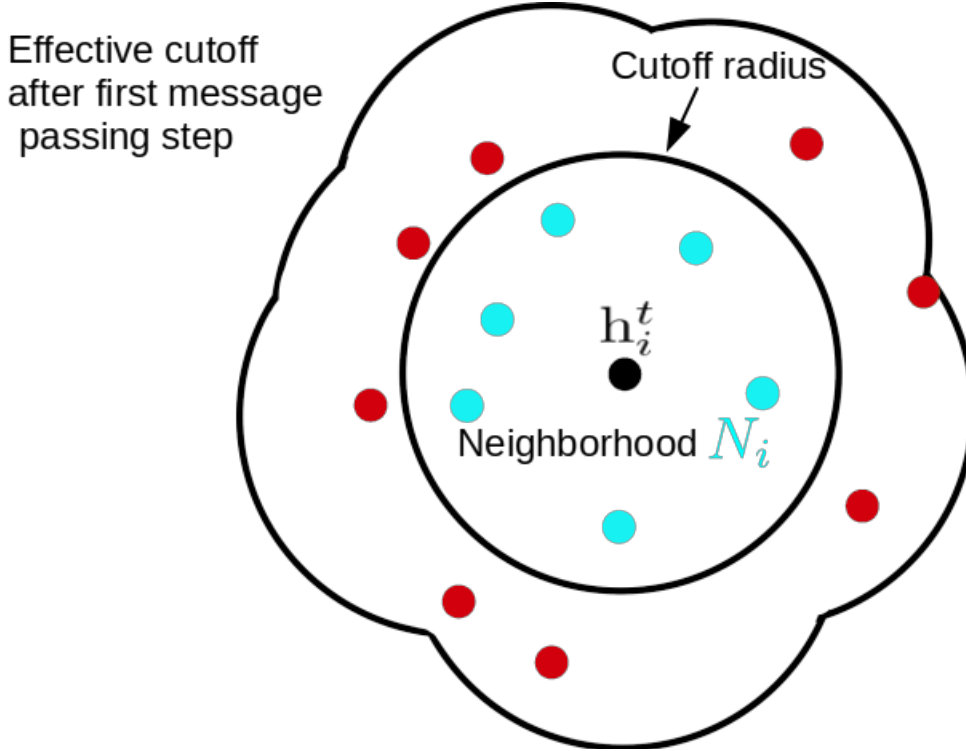


Figure 4: The  $i_{th}$  atom is shown in black with the atoms in its cutoff radius in blue. The atoms in red are inside the cutoff radius of the neighboring atoms. Consequently, the information that was passed to them is part of the second message passing step to the  $i_{th}$  atom.

As we see in fig. 4 the first message passing step is limited to information inside the original cutoff radius. The representations of the atoms in the first cutoff radius are updated based on their own spatially restricted neighborhood. Thus the effective radius from which information is passed to an atom during the second and further message passing steps is gradually extended.

This functionality is very similar to standard convolutional networks that work on a grid, and a large share of message passing networks are actually graph convolutional networks that generalize convolutions to graphs [170].

In material science we typically use three to five updating steps. After the last step, we can predict atomic properties, e.g., atomic charges or magnetic dipoles. To obtain a prediction for the whole crystal structure, we have to combine the atom representations to a single graph embedding through a pooling function. This graph embedding represents the entire crystal structure or molecule and typically enters some form of FCNN to arrive at a prediction. During the last years a large number of message passing network topologies were published that produce impressive results for the prediction of structural properties [172, 46, 41, 20, 51, 117, 143, 71], Hamiltonians [5, 81, 134, 138, 40], and structure generation [175, 28, 160, 89, 98, 106, 70, 42, 43]. We discuss a few of the most relevant and influential examples.

**Crystal Graph Convolutional Neural Networks (CGCNN)** were one of the first message passing networks explicitly designed for crystal structures. Xie *et al.* used elemental properties encoded in one-hot vectors as starting embedding for the atoms. The distances between atoms were encoded in a Gaussian basis. We already noted the update function in eq. (53). It only updates the node embeddings and keeps

the edge embeddings constant. The CGCNN were used to predict a variety of crystal structure properties (formation energy, bandgap, fermi energy, etc.) and according to relatively recent benchmarks [32, 37] they are still competitive with newer models.

With the **SCHNET**-architecture Schütt *et al.* introduced continuous filter convolutions as a message passing function. Here the atoms are represented by randomly initialized embeddings, and the edge embeddings are radial expansions of the distances  $d_{ij}$  between atoms:

$$\mathbf{e}_{ij} = ||_k e^{-\gamma ||d_{ij} - \mu_k||^2}. \quad (54)$$

$||_k$  denotes the concatenation of the  $k$  values. The number of different basis functions and the distances of their centers  $\mu_k$  determine the size and resolution of the filter convolutions. The atomic representations are updated by so-called interaction blocks consisting of fully connected network layers that adapt the representations and the continuous filter convolutions. The edge representations enter a fully connected two layer neural network that yields the filter weights:

$$W^t(\mathbf{e}_{ij}) = \text{FCNN}_c^t(\mathbf{e}_{ij}) \quad (55)$$

for the continuous convolution.

$$\mathbf{h}_i^{t+1} = \mathbf{h}_i^t + \text{FCNN}_1 \left( \sum_{ij} \text{FCNN}_0^t(\mathbf{h}_j^t) \circ W^t(\mathbf{e}_{ij}) \right). \quad (56)$$

Comparing this function with the crystal graph convolutional networks published one year later, we see that the SCHNET convolution operator only depends on the edge information while the CGCNN filters are calculated using both nodes and the connecting edge. Initially, the SCHNET-architecture was only applied to molecular datasets but it can easily be applied to solids and an implementation for solids is available in

the SCHNETPACK software package [140]. Both network architectures retain their starting edge representations during the message passing process. Jorgensen *et al.* [71] expanded the SCHNET-architecture with an edge update function. At each message passing step the edges are updated by a fully connected network using a concatenation of the updated embeddings of the two connected nodes and the edge embedding as input:

$$\mathbf{e}_{ij}^{t+1} = \text{FCNN}(\mathbf{e}_{ij}^t || \mathbf{h}_i^{t+1} || \mathbf{h}_j^{t+1}). \quad (57)$$

This addition leads to a remarkable performance gain of more than 25% for formation energy prediction tasks in the QM9 dataset [130, 132], the Materials Project dataset [68] and the OQMD [133, 77].

However, as these networks and countless other message passing architectures rely on the relative atomic positions as input data, they cannot be used effectively in high-throughput studies to predict properties from unrelaxed structures. For instance, one of the most recent architectures [117] reports several times higher errors for unrelaxed structures.

### 3.1.4 Attention Mechanism for Graphs

The most successful idea in machine learning from the last years is the attention mechanism and its application in natural language processing [162]. In natural language processing, the self-attention mechanism is responsible for directing the attention of the network layers in a transformer [162] to tokens or words in a sentence depending on the sentence itself. Transformers can be considered as graph networks over a fully connected word-graph.

As a result, different versions of the attention mechanism have also found success in message passing architectures [163] and were already applied to material science [51].

In the case of graph networks, the goal is to direct the attention during the message passing process to the nodes with the most relevant messages. The attention is directed by so-called attention coefficients  $a_{ij}$ . For each atom  $i$  the attention coefficient are multiplied with the messages  $\mathbf{m}_{ij}$  from atom  $j$  to atom  $i$ . If the attention coefficient is close to zero, atom  $i$  does "not pay attention" to atom  $j$ . However, there might be multiple important aspects to the chemical environment represented by each message passing step. Thus, we not only calculate one but  $N$  sets of attention coefficients  $a_{ij}^n$  and messages  $\mathbf{m}_{ij}^n$  at each time step. We call each set of networks calculating the messages and attention coefficients an attention head. Ideally, each head learns to direct its focus at different facets of the chemical environment.

In detail each attention head, indexed by  $n$ , at each timestep  $t$ , consists of two fully connected networks  $\text{FCNNN}_a^{t,n}$  and  $\text{FCNNN}_m^{t,n}$  where the first is used in the process of calculating the attention coefficients while the latter directly calculates the messages:

$$\mathbf{s}_{ij}^{t,n} = \text{FCNNN}_a^{t,n}(\mathbf{h}_i^t || \mathbf{h}_j^t || \mathbf{e}_{ij}) \quad (58)$$

$$a_{ij}^{t,n} = \frac{\exp(s_{ij}^{t,n})}{\sum_j \exp(s_{ij}^{t,n})} \quad (59)$$

$$\mathbf{m}_{ij}^{t,n} = \text{FCNN}_m^{t,n}(\mathbf{h}_i^t || \mathbf{h}_j^t || \mathbf{e}_{ij}). \quad (60)$$

Here  $\mathbf{h}_i^t || \mathbf{h}_j^t || \mathbf{e}_{ij}$  denotes the concatenation of the atomic embeddings  $\mathbf{h}_i^t$  and  $\mathbf{h}_j^t$  as well as the edge embedding  $\mathbf{e}_{ij}$ . Based on this information the message  $\mathbf{m}_{ij}^{t,n}$  and the coefficients  $s_{ij}^{t,n}$  are calculated. The latter are normalized with a softmax function to produce the attention coefficients.

We update the representation of each atom using the messages weighted

by the attention coefficients:

$$\mathbf{h}_i^{t+1} = \mathbf{h}_i^t + \text{FCNN}^t \left( \left\| \sum_j^n a_{ij}^n \mathbf{m}_{ij}^h \right\| \right). \quad (61)$$

After we sum over the messages from each atom, we concatenate or average the resulting messages from each attention head before they enter an additional network. A basic version of this attention mechanism was developed in Ref. [51] and another version of it is applied in Ref. [JSPHD10].

”Representation Learning from Stoichiometry” (**Roost**) [51] in contrast to the other models we discussed, is a message passing network for graphs of the composition. For instance, a ternary composition ABC is represented as a fully connected graph with the vertices A, B and C and messages are passed between these nodes. The fractional composition is then used to weigh the embeddings during message passing and pooling. The message passing applies a version of the attention mechanism described above that does not use edge information. Roost uses pretrained embeddings from Ref. [159], that were trained as word2vec [107] embeddings on material science literature, followed by a linear layer. In general, Roost performs superior to previous composition-based machine learning models but worse than message passing models that include the structure as a crystal graph [7].

### **Crystal Graph Attention Networks**

We developed Crystal Graph Attention Networks (CGAT) for the prediction of thermodynamic stability in Ref. [JSPHD10]. They are another message passing architecture for crystal graphs. We use the same fixed pretrained elemental embeddings from Ref. [159] as Ref. [51], followed by an extra linear layer that allows for changes during training. The edge embeddings constitute a significant change in contrast to prior

networks for crystal structures. For the edges, we consider purely the graph distance and assign each graph distance, i.e. first neighbor, second neighbor etc. a vector that is randomly initiated during the network creation. These embeddings are again updated during training to learn sensible representations of the graph distances. We use the previously introduced attention mechanism with some updates as our message passing scheme.

Each message is already a high-dimensional vector. Therefore, we do not only use attention coefficients such as in Ref. [163] and Ref. [51] but introduce attention vectors of the same dimension as the messages. While this change increases the parameter number per attention head, we also see a mean absolute error improvement (MAE) of 11% in comparison to scalar attention coefficients for experiments we performed in Ref. [JSPHD10]. The update function is further revised by changing the last FCNN network in eq. (61) to a hypernetwork.

$$\mathbf{h}_i^{t+1} = \mathbf{h}_i^t + \text{HFCNN}_{\theta_g^t} \left( \left\| \sum_n \mathbf{a}_{ij}^n \mathbf{m}_{ij}^n \right\| \right). \quad (62)$$

Here HFCNN is a hypernetwork along the lines of Ref. [110] that calculates a network based on the representation at each message passing timestep in comparison to the starting time step. While [110] used the same network at each message passing time step, we found separate networks performing better.

Jorgensen *et al.* [71] demonstrated that updating both the edge embeddings as well as the node embeddings yields better results. Thus we also add an edge-update function. We consider this even more essential in our case, as the first neighbor distance between, e.g., an iron and a bismuth atom or an iron and a chloride atom will most likely be very different. And in contrast to an exact distance representation, we ini-



tialize the embeddings for each respective graph distance always with the same vector. Consequently, they need to adapt more based on their environment.

The edge embeddings are updated in a similar manner as the node embeddings at each message passing step:

$$\mathbf{s}_{ij}^{e,n} = \text{FCNNN}_a^n(\mathbf{h}_i^t || \mathbf{h}_j^t || \mathbf{e}_{ij}^t) \quad (63)$$

$$\mathbf{a}_{ij}^{e,n} = \frac{\exp(\mathbf{s}_{ij}^{e,n})}{\sum_n \exp(\mathbf{s}_i^n)} \quad (64)$$

$$\mathbf{m}_{ij}^{e,n} = \text{FCNNN}^n(\mathbf{h}_i^t || \mathbf{h}_j^t || \mathbf{e}_{ij}^t) \quad (65)$$

$$\mathbf{e}_{ij}^{t+1} = \mathbf{e}_{ij}^t + \text{HFCNN}_{\theta_g^t}^{n,t} \left( || \mathbf{a}_{ij}^{e,n} \mathbf{m}_{ij}^{e,n} \cdot \right) \quad (66)$$

After the message passing steps, we have to combine the updated representations of the single atoms into a prediction of a crystal property. The obvious step would be to use an additional attention layer to pool the representations together.

However, in 2021 Louis *et al.* [101] introduced a global composition-based attention pooling layer. Here, the atomic representations are combined contingent on a global context vector. An attention layer is used to combine the representations of the different atoms, but before the representations enter the  $\text{FCNNN}_{\text{Pooling}-a}^n$  and  $\text{FCNNN}_{\text{Pooling}-m}^n$  of the attention heads a vector representing the composition is concatenated. In this way, the entire chemical context can be considered when combining the vectors representing each atom. While Louis *et al.* used a simple vector based on the fractional compositions, we train a ROOST model [51] to output a context vector. Besides the extra context vector, the attention pooling layer is a normal attention layer that considers all

atoms in the unit cell:

$$\mathbf{s}_i^n = \text{FCNN}_a^n(\mathbf{h}_i^t | \mathbf{C}) \quad (67)$$

$$\mathbf{a}_i^n = \frac{\exp(\mathbf{s}_i)}{\sum_i \exp(s_i^n)} \quad (68)$$

$$\mathbf{m}_i^n = \text{FCNN}_m^n(\mathbf{h}_i^t) \quad (69)$$

$$\text{Output} = \text{FCNN}_{RS} \sum_{i,n} \mathbf{m}_i^n \mathbf{a}_i^n. \quad (70)$$

Lastly, a fully connected network with residual connections  $\text{FCNN}_{RS}$  [59] is used to obtain the final output. In Ref. [JSPHD10] we solely use the network to predict distances to the convex hull. However, it can easily be used for other system- or per-atom-properties.

### 3.2 Machine Learning Material Properties

A large section of the machine learning research in material science can be summarized as predicting material properties, i.e. learning structure-property relationships. Using some representation of crystal structures as input, researchers have already attempted to predict more than 40 different properties [JSPHD9, JSPHD6]. The by far most common target properties in solid-state material science are the formation energy or thermodynamic stability of materials [34, JS3, JS4, 117, 71, 97, 75, 139, 172, 97, 174, 20, 179, 51, 110] and the band gap of materials [183, 172, 66, 29, 91, 124, 128, 156, 176, 148, 169, 55, 126, 142, 102, 125, 21, JSPHD1]. As discussed in the introduction, we are specifically interested in the first two. For a machine learning application, the first challenge is selecting an adequate feature set, i.e. input representation. The majority of works [34, 33, 167, JS4, 139] until 2017 used input representations purely built on domain knowledge and human intuition. Typically, a number of atomic properties such as atomic mass and radius were se-

lected and combined to arrive at a feature vector representing the whole composition. In the following, a simple machine learning algorithm, e.g., a decision tree ensemble [17, 44] or a FCNN was applied to the vector to obtain the formation energy or distance to the convex hull. Despite their simplicity, these models were already quite successful, achieving speedups, i.e. reduction in the number of DFT calculations in comparison to a complete high-throughput search, between a factor of 5 to 30 [JS4, JS3, 75]. The next step in model development were deeper fully connected networks or convolutional networks such as ElemNet [69] or Ref. [180, 181], respectively. These do not use atomic properties as features but apply the ability of deep neural networks to learn their own representations [90] from the composition.

ElemNet uses a one-hot vector of length  $N$ , where  $N$  is the number of possible species, for each composition. The corresponding positions in the vector are set to the fraction of the species in the composition (see fig. 5). Then a 7-layer FCNN learns the formation energy from this representation. The representations in Ref. [180, 181] are slightly more sophisticated as they use an "image" of the periodic table filled with the corresponding composition fractions as input, followed by a few convolution layers and then a deep FCNN.

However, if we consider a prototype-based high-throughput search of a binary compound  $AB$ , where  $A$  and  $B$  are not equivalent sites, we quickly see that both models [69, 180] have problems learning. The models cannot differentiate between the structures  $AB$  and  $BA$ . For a single prototype structure this problem can be circumvented by introducing an extra dimension to the input. Then we obtain one vector/periodic table for position  $A$  and one vector/periodic table for position  $B$ . However, as a result the model cannot profit from training

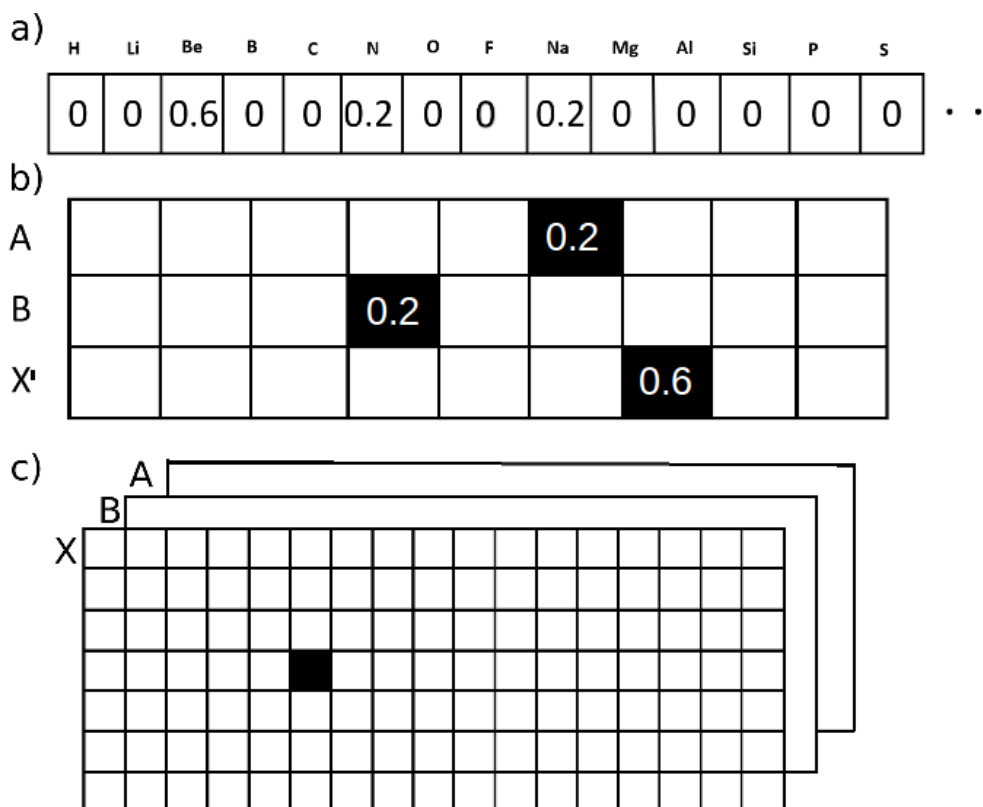


Figure 5: a) shows the ElemNet feature vector for a composition  $\text{NaNiBe}_3$ . b) shows the feature vector for a ternary composition when it is extended to 2-dimensions to apply an ordering of the composition to differentiate between polymorphs of one prototype. c) shows the same method applied to the periodic table "images" used in Ref. [180, 181]. Figure from Ref. [JSPHD10]

data from other prototypes and separate training data need to be calculated for each prototype. The state of the art concerning composition models is Roost [51, 7] that considers the composition as a graph and then applies a message passing mechanism. We discussed it in detail in the previous section 3.1.3. Despite its impressive performance in benchmarks [7], Roost has the common failure point of compositional models as it also cannot differentiate between polymorphs, i.e. compounds with the same composition but a different structure. Isayev *et al.* [66], and Ward *et al.* [167] were some of the first to develop successful structure

sensitive descriptors suitable for high-throughput searches. Besides distinguishing polymorphs, the primary advantage of structure sensitive representations is that all training data of all crystal structures can be used. Ward *et al.* [167, 75] demonstrated that the additional training data can result in significant performance gains. With the advent of deep graph neural networks, the molecular and crystal graphs discussed in the previous section have turned out to be the most successful representation for large data applications [32, JSPHD9].

However, due to their reliance on atomic positions message passing networks have not made an impact in high-throughput studies yet.

To circumvent the problem of unknown exact atomic positions in prototype-based high-throughput searches, we developed crystal graph attention networks. As discussed previously, the edge embeddings are the primary difference to other networks as we solely consider the graph distance and ignore the exact distances. This change brings the advantage that if we simply scale the volume of a structure, the neighbor list, and consequently the network output, stay unchanged. Thus, our prediction is relatively insensitive to a geometry optimization. By constructing prototypes with different cell constant ratios, we can reduce the effect of a changing neighbor list during geometry optimization even more. For instance, mixed perovskites have a tetragonal structure fig. 6, and so we have to consider structures stretched by different ratios to arrive at a minimum energy prediction. This approach is only viable for energies, as we are essentially performing a geometry optimization with the machine learning model by selecting the lowest energy prediction as the ground state prediction. For other properties no such selection criterion is available. Therefore, we use the structure selected during the energy prediction to obtain other properties. We will discuss the

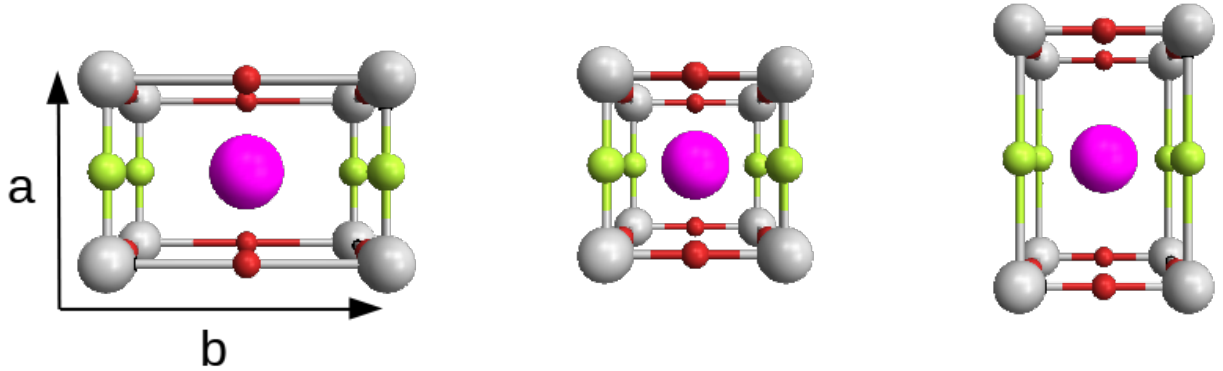


Figure 6: Crystal structures of the mixed perovskites prototype used in Ref. [JSPHD10] with different cell constant ratios  $a/b$  .

details of the application and testing of the CGATs in section 4.2.

### 3.3 Neural Network Functionals

In this section, we give an overview of machine learning works for density functionals up to the time when Ref. [JSPHD7] was published. Tozer *et al.* [157] started the development of machine learning exchange-correlation functionals in 1996, a decade ahead of their time. They used a one hidden-layer FCNN to map the electronic density directly to the exchange-correlation potential. Despite their network containing only 25 parameters, the functional yielded results comparable to the LDA for molecules that were sufficiently represented in the training data.

The machine learning work got rekindled in the early 2010s with, e.g., the work performed by Snyder *et al.* [147, 165, 145, 146, 96]. Starting with a 1D kinetic energy functional [146] based on kernel ridge regression, self-consistent orbital free calculations were possible. As only the functional and not its derivative was trained, the directly calculated functional derivative was inaccurate and needed to be corrected through a projection on the principal components of the most similar training

densities. Snyder *et al.* researched similar kinetic energy functionals in further publications [147, 165, 145, 96]. However, the functionals were always held back by the untrained functional derivatives. As a result, the functional derivatives needed to be constrained to the training data manifold. Yao *et al.* were the first to use convolutional networks for the development of kinetic energy functionals and already mentioned the possibility of using the derivative of the network as a functional derivative but were hindered by the programming effort to realize it. Li *et al.* [94] and Hollingsworth *et al.* [63] also used kernel ridge regression to train kinetic energy functionals for 1D systems. A second way to bypass the Kohn-Sham equations is to train the Hohenberg-Kohn map in the sense of predicting the electronic density directly from the external potential or nuclei positions [18, 14, 54]. Other work [99, 112, 103, 168] focused on training the exchange-correlation functionals that we are interested in. Lundgaard *et al.* [168, 103] fit a standard meta-GGA form with van der Waals corrections to sets of molecules. In Ref. [112] Nagai *et al.* use fully connected neural networks to directly predict exchange-correlation potentials from the electronic density for a one-dimensional two-electron system with a fixed size. Liu *et al.* [99] fit a neural network as the range separation parameter for a long-range-corrected Becke-Lee-Yang-Parr functional [64] improving atomization energies and heats of formation for a set of molecules. Zhou *et al.* train a 3D-convolutional neural network on molecules to predict the exchange-correlation potential [182] directly from a local density. However, this potential is also not calculated as the functional derivative, and the network does not satisfy the correct symmetry requirements. While there are a number of machine-learned density functionals, the majority of them have some problematic features. To train a machine learning functional, we have

to consider a few requirements that are easily fulfilled in most classic functionals. The first requirement is that the network takes purely the electronic density as input to arrive at a physical exchange-correlation potential that fulfills the equation:

$$v_{\text{xc}}([n(r)], r) = \frac{\delta E_{\text{xc}}[n]}{\delta n(r)}. \quad (71)$$

There are a small number of functionals that violate this equation [88, 10, 158]. While these "stray" functionals [39] have found some successful applications, they violate a number of exact conditions limiting the universality and application range of the functionals.

Secondly, the network must be able to treat systems of different sizes and be size-consistent. That means calculations for two infinitely far removed subsystems should arrive at the same result as separate calculations of the subsystems. Quite a few of the early works [111, 146] already do not fulfill the requirement of being able to treat different system sizes. And while some of them used the derivative of the exchange-correlation energy as potential, none of them trained it [JSPHD6].

As a third point, the physical symmetries of the density have to be considered. A functional should treat density points independent of their position in space, i.e. the functional has to be translationally invariant. Furthermore, if we rotate or mirror a system, the exchange-correlation energy has to be invariant while the exchange-correlation potential must be equivariant. For 3D systems, these requirements are fulfilled by the Euclidean group in 3 dimensions  $E(3)$ . None of the works we discussed before considered this symmetry directly in the network but some of them calculated invariant features [93] from the density instead. In the one-dimensional case that we treated in Ref. [JSPHD7], the symmetries reduce to a translational invariance and a mirror symmetry. We can meet the latter requirement by adding a mirror symmetry to the weight



matrix of the first layer of a neural network. To achieve the translational invariance and treat different system sizes, we need to use some type of convolutional network. In our case, a sliding-window convolution returned the best results. A sliding-window convolutional network is a simple FCNN that considers only a certain number of density points at once (kernel size) and returns a local energy as output. By scanning over the whole system in this manner, the symmetries of the 1D density are satisfied.  $E(3)$  equivariant networks [9, 155] that solve these challenges in three dimensions have also been developed during the last years, although we are not aware of any application to machine learning functionals yet.

## 4 Results and Discussion

In the previous chapters, we have discussed the theoretical foundations of DFT and machine learning applied to the development of functionals as well as material science. In the following chapter, we present three publications and discuss them in the context of the thesis. Furthermore, we discuss follow up publications as well as important results that are not yet published.

### 4.1 Machine Learning the Physical Nonlocal Exchange-Correlation Functional of Density-Functional Theory

In the following publication ”**Machine Learning the Physical Nonlocal Exchange-Correlation Functional of Density-Functional Theory**” [JSPHD7] we simultaneously train a neural network as an exchange-correlation energy functional and its functional derivative as exchange-correlation potential using automatic differentiation. We demonstrate the feasibility of this approach for one-dimensional systems with two strongly correlated electrons that are normally beyond DFT.

Specifically, we build a local functional that takes  $\kappa$  (kernel size) density points as input to predict a local exchange-correlation energy. Consequently, the network can be used for systems of different sizes and also fulfills the requirement of size-consistency that we discussed in the theory section. By symmetrizing the kernel and using a sliding window convolution we also satisfy the physical symmetries of the systems. We train networks with different kernel sizes on 2 electron systems with one to three randomly placed nuclei. We are able to demonstrate that including the exchange-correlation potential as functional derivative in the loss function leads to massively improved potentials, allowing for

accurate self-consistent calculations.

We also show that increasing the kernel size, i.e. the degree of non-locality, leads to consistently decreasing mean absolute errors for the energy of self-consistent calculations. Furthermore, we evaluate the functional on the dissociating  $\text{H}_2$  molecule in one dimension. Here, we find that increasing the kernel size from one to thirty density points yields a massive improvement in the dissociation curve. However, it has to be noted that kernel sizes larger than thirty prove problematic as they show nonphysical behavior during the dissociation. We also study the homogeneous electron gas and find similar results. Compared with quantum Monte Carlo results, a kernel size of 120 produces the best results, but larger kernel sizes increase the error once again. Finally, we test the best performing functional on systems with four nuclei. We find that its energy error is still eight times smaller than an LDA despite us only training the functional on systems with up to 3 nuclei. While we researched the effects of non-locality, this work was mostly a proof of concept work on one-dimensional model systems.

Indeed, since this work, the number of machine learning functionals have strongly increased [JSPHD6] and a large number of them calculate the exchange-correlation potential as the functional derivative of the neural network energy functional [73, 95, 30, 74, JSPHD3]. Furthermore, we have demonstrated in a follow-up work [JSPHD3], directly based on this publication [**ourfunctional**], that it is possible to correctly include the derivative discontinuity at integer particle numbers in a machine learning functional. We hope that including such exact conditions that go beyond previous functionals might extend the applicability of DFT. For instance, correct derivative discontinuities could result in improved

fundamental gaps from DFT. <sup>1</sup>

---

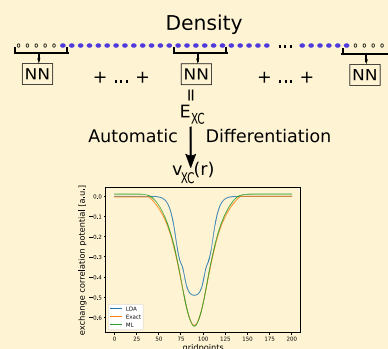
<sup>1</sup>Concerning the following article: Reprinted with permission from Jonathan Schmidt *et al.* J. Phys. Chem. Lett.10.20 (Oct.2019). Copyright 2019 American Chemical Society.

# Machine Learning the Physical Nonlocal Exchange–Correlation Functional of Density-Functional Theory

Jonathan Schmidt, Carlos L. Benavides-Riveros,\*<sup>1</sup> and Miguel A. L. Marques\*<sup>2</sup>

Institut für Physik, Martin-Luther-Universität Halle-Wittenberg, 06120 Halle (Saale), Germany

**ABSTRACT:** We train a neural network as the universal exchange–correlation functional of density-functional theory that simultaneously reproduces both the exact exchange–correlation energy and the potential. This functional is extremely nonlocal but retains the computational scaling of traditional local or semilocal approximations. It therefore holds the promise of solving some of the delocalization problems that plague density-functional theory, while maintaining the computational efficiency that characterizes the Kohn–Sham equations. Furthermore, by using automatic differentiation, a capability present in modern machine-learning frameworks, we impose the exact mathematical relation between the exchange–correlation energy and the potential, leading to a fully consistent method. We demonstrate the feasibility of our approach by looking at one-dimensional systems with two strongly correlated electrons, where density-functional methods are known to fail, and investigate the behavior and performance of our functional by varying the degree of nonlocality.



Nowadays density-functional theory (DFT) is the cornerstone of computational theoretical physics and quantum chemistry, as it provides the prevalent method for the calculation of the electronic structure of both solids and molecules. Based on the Hohenberg–Kohn theorems,<sup>1</sup> DFT re-formulates the quantum many-electron problem as a theory of the ground-state electronic density  $n(\mathbf{r})$ . The success of DFT is to a large extent due to the existence of a system of noninteracting electrons (the Kohn–Sham system) that has the same ground-state density as the interacting electrons. This leads to the Kohn–Sham equations, a set of self-consistent equations for one-particle orbitals.<sup>2</sup> In such a formalism the ground-state (GS) energy can be expressed as

$$E_{\text{GS}} = \sum_i \epsilon_i + E_{\text{xc}}[n] - \int d^3r v_{\text{xc}}(\mathbf{r}) n(\mathbf{r}) - E_{\text{H}}[n] \quad (1)$$

where  $\epsilon_i$  are the eigenvalues of the Kohn–Sham Hamiltonian,  $v_{\text{xc}}(\mathbf{r})$  is the exchange–correlation potential,  $E_{\text{H}}[n]$  is the Hartree energy, and  $E_{\text{xc}}[n]$  is the exchange–correlation energy. The exchange–correlation potential is defined as the functional derivative of the universal exchange–correlation energy functional:

$$v_{\text{xc}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}[n]}{\delta n(\mathbf{r})} \quad (2)$$

Due to the Hohenberg–Kohn existence theorems, if the exact exchange–correlation energy functional  $E_{\text{xc}}[n]$  is known, Kohn–Sham DFT then yields the exact ground-state energy and the exact ground-state electronic density.

Traditionally, “educated” formal expressions of the exchange–correlation energy functional have been proposed by a combination of theoretical insight, highly accurate Monte

Carlo,<sup>3</sup> or quantum chemical simulations or by fitting general expressions to experimental data. In general, functionals can be sorted according to Jacob’s ladder:<sup>4</sup> the lowest rung of the ladder is occupied by local-density approximations (LDA) that use solely single density points as inputs.<sup>5–7</sup> The second rung is occupied by generalized-gradient approximations (GGA) that include the gradient of the density.<sup>8,9</sup> This is followed by the meta-GGAs<sup>10</sup> (that use the kinetic-energy density) and hybrid functionals<sup>11–13</sup> (that mix a fraction of nonlocal Fock exchange) on the subsequent rungs. Note that more than 500 of these functionals have been proposed in the past decades,<sup>14</sup> although most of them with rather limited impact.

In spite of the success of DFT in dealing efficiently with electronic systems, it still suffers from stubborn quantitative and qualitative failures. For instance, barriers of chemical reactions, band gaps of materials, or molecular dissociation energies are usually underestimated.<sup>15</sup> Degenerate or near-degenerate states are also poorly described by DFT. While hybrid functionals can alleviate some of the problems of traditional semilocal functionals, they come at a greatly increased computational cost that limits severely the number and size of systems that can be researched. It is believed that many of these problems originate in the delocalization and static correlation errors which plague approximate functionals.<sup>16–18</sup> Roughly speaking, the delocalization error refers to the tendency of DFT functionals to spread out the electron density, while the static correlation arises from the difficulty of describing degenerated states with a single Slater determinant.<sup>19</sup>

**Received:** August 19, 2019

**Accepted:** October 4, 2019

**Published:** October 9, 2019

More recently, machine learning (ML) has revolutionized many fields of computational sciences, such as image or speech recognition,<sup>20,21</sup> and has found countless applications in material science.<sup>22–24</sup> Within DFT, the application of ML techniques to the formulation of density functionals has already a long history.<sup>25</sup> In 2012, an ML approximation for the kinetic energy functional  $T_s[n]$  was constructed for a system of noninteracting spinless fermions.<sup>26,27</sup> Yao et al.<sup>28</sup> developed a convolutional neural network to reproduce the kinetic energy functional for molecules. They already mentioned the possibility of using the functional derivative of the neural network for minimization purposes. In order to exploit the Hohenberg–Kohn density–potential map, an ML model was later trained to learn the fundamental relation of DFT between external potentials and electronic densities.<sup>29</sup> These works focused mainly on developing functionals for the total energy or the noninteracting kinetic energy to facilitate orbital-free DFT calculations. More recently, some works have addressed the problem of training the exchange–correlation potential.<sup>30–34</sup> However, this line of research has been limited by the fact that the exchange–correlation potential was not obtained from the exchange–correlation energy through the functional derivative of eq 2.

It is true that one can find in the literature a series of approximations to the exchange–correlation functionals that do not fulfill eq 2. For example, the Krieger–Lee–Iafrate approximation<sup>35</sup> breaks this connection in order to simplify the implementations of orbital functionals using the optimized effective potential method.<sup>36,37</sup> Sometimes, it is also convenient to approximate directly the potential (e.g., in the van Leeuwen–Baerends GGA from 1994<sup>38</sup> or the modified Becke–Johnson potential<sup>39,40</sup>), leading again to expressions that do not obey eq 2. These so-called “stray” functionals<sup>41</sup> have found some important applications. For example, the modified Becke–Johnson is one of the most successful functionals to calculate electronic band-gaps.<sup>42</sup> Unfortunately, they are also found to break a series of exact theorems and conditions,<sup>41</sup> severely limiting their universality and range of applicability. By and large, it is highly advantageous to develop consistent functionals that obey the important eq 2.

Modern ML frameworks, like pytorch<sup>43</sup> and tensorflow,<sup>44</sup> allow for automatic differentiation with respect to any parameter. Recently, Nagai et al. used this functionality to train exchange–correlation potentials for molecules.<sup>45</sup> They trained neural networks through a Monte Carlo updating scheme to reproduce accurate energies and densities of molecules. The functionals by Nagai and coauthors follow the traditional approaches of an LDA, GGA, meta-GGA and add a related near-region approximation. Although a clear step forward, using traditional forms for the exchange–correlation functional is unlikely to lead to fundamentally better, disruptive approximations to the exchange–correlation functionals. New paradigms have to be sought in order to unleash the power of ML techniques to its full extent.

In this paper we use the autodifferentiation functionality to train neural-network exchange–correlation functionals through back-propagation. The networks are trained to reproduce not only the correct exchange–correlation energy  $E_{xc}$  but also the exchange–correlation potential  $v_{xc}(r)$  consistently as its functional derivative with respect to the density. Consequently, the resulting functional allows for self-consistent calculations and can easily be integrated into existing Kohn–Sham DFT frameworks. Furthermore, these functionals can be made

highly nonlocal by using the information on the density in a *finite* neighborhood as input to the neural network, allowing for far more nonlocality than traditional LDA or GGA functionals, despite having the same computational scaling with system size. Therefore, this approach promises to alleviate the delocalization problems of DFT and to improve its accuracy without the computational expense of hybrid functionals. To demonstrate the feasibility of this approach, we developed an ML functional for the exchange–correlation energy and exchange–correlation potential based on exact results for two electrons in one-dimension (1D).

The letter presents the details of the data set, training process, and neural networks. The exact dependence of the functional on the degree of locality and its behavior is also discussed, as well as our results for the 1D homogeneous electron gas and the H<sub>2</sub> molecule along the dissociation path. Finally, we finish the letter discussing our conclusions and future research directions.

**Data.** The training data was produced by solving exactly the one-dimensional two-electron problem in the external potential generated by up to three different nuclei. Softening the Coulomb interaction,

$$\frac{1}{r} \rightarrow \frac{1}{\sqrt{1+x^2}} \quad (3)$$

we obtain the 1D Hamiltonian driven by the interaction of the two electrons, namely,

$$H(x_1, x_2) = - \sum_{i=1}^2 \left[ \frac{1}{2} \partial_i^2 + v(x_i) \right] + \frac{1}{\sqrt{1+(x_1-x_2)^2}}$$

where the external potential is given by the superposition of three potentials,

$$v(x) = \sum_{k=1}^3 \frac{Z_k}{\sqrt{1+(x-a_k)^2}} \quad (4)$$

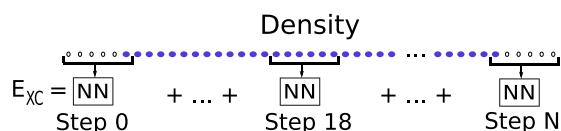
The total charge of the nuclei  $Z = \sum_k Z_k$  is equal to 2 or 3. Qualitatively close to real 3D systems, this 1D model is known as a theoretical laboratory for studying strong correlation and developing exchange–correlation density functionals for DFT.<sup>46</sup> Since the ground-state problem of the Hamiltonian  $H(x_1, x_2)$  can be treated as a one-particle problem in two dimensions, the problem can be solved exactly.

We sampled 20 000 systems and calculated their exact ground-state energy and ground-state electronic density. We used a grid spacing of 0.1 au, and a box size of 20 au, leading to a grid with 201 points. The nuclei positions  $a_i$  in eq 4 were normally distributed with zero mean and variance of 4 au. We then solved the corresponding inverse Kohn–Sham problem in OCTOPUS<sup>47</sup> to find the exact exchange–correlation energy and potential. Since the inversion is known to be numerically unstable,<sup>48</sup> we removed outliers that result from these instabilities. We used up to 12 800 of these systems for training, 6400 for validation during the training, and 2000 systems for the test set. Furthermore, training was considerably improved when removing outliers with  $E_{xc} > -0.55$  au from the training set. No outliers were removed from the test set to allow for a completely unbiased evaluation of the functionals.

In general, one would have to double the data by mirroring the systems to learn the correct symmetry. However, in this specific case one can simply build the symmetry directly into

the neural network functional, as explained in the next subsection.

**Topology of the Neural Network.** Our ML functional scans the density of the total system, as illustrated in Figure 1. The



**Figure 1.** Structure of the ML functional in 1D with the degree of locality equal to  $\kappa = 6$  (see text). At the borders, the density is padded with  $\kappa - 1 = 5$  zeros. Starting at one of the borders, the network calculates the local exchange–correlation energy for  $\kappa$  points. In the next step, the input of the network is moved by one grid point and it is evaluated again. The network itself is a simple fully connected neural network.

density in a neighborhood of the test point is used as the input for a 4 or 5 layer fully connected neural network, that then outputs a *local* exchange–correlation energy. Specifically, the network takes a certain number of density points as input, which we call  $\kappa$ , the *kernel size*. This is the degree of locality of the ML functional. At the borders of the system the density is padded with  $\kappa - 1$  zeros. Starting at one border the network calculates the local exchange–correlation energy. In the next step the input of the network is moved by one grid point, and it is evaluated again. As described in Figure 1, this process continues until the other border is reached. We arrive at the total exchange–correlation energy of the system by summing over all network outputs. The padding and the scanning with a certain kernel size are inspired by standard convolutional neural networks and can also be implemented as such by concatenating the data along the channel-dimensions in between standard convolutional layers. Due to the homogeneity of space, the functional has to be symmetric with respect to its input densities. The symmetry is ensured by initializing the weights of the first layer symmetrically along the spatial dimension. To arrive at the final scalar output we used four- or five-layer fully connected neural networks.

The possible selection of activation functions (i.e., the nonlinearities that follow each multiplication with a weight matrix of a neural network) was rather limited, because typical functions (e.g., rectified linear units) were not usable due to their lacking differentiability at zero (using *relu* actually resulted in piece-wise linear potentials). To avoid this problem, we chose exponential linear functions.<sup>49</sup> Different numbers and sizes of hidden layers were also tested, and we settled on the minimum number of parameters that could be used without underfitting. In this work all networks were built on the basis of the ML framework *pytorch*.<sup>43</sup> The library *Ignite*<sup>50</sup> was used to simplify the training process and *tensorboardX* to integrate *tensorboard*<sup>44</sup> into *pytorch*. The network weights were optimized with *Adam*<sup>51</sup> using default parameters from *pytorch*.

For the loss function (i.e., the cost function which is going to be optimized in the learning process), we should keep in mind that the objective is not only to obtain small errors for the exchange–correlation energy. To arrive at the correct density through the solution of the self-consistent Kohn–Sham equations, also the exchange–correlation potential should be as close as possible to the exact one. Furthermore, we want not only to ensure a small error for the potential but also its smoothness. In addition, the error of the exchange–correlation

energy, as well as the error of the integral  $\int dx v_{xc}(x) n(x)$  (that appears in the expression for the total energy (1)), should be minimized.

In order to achieve all these goals concurrently, we used the following loss function, where  $\theta$  are the parameters of the neural network that have to be optimized:

$$L(\theta, n_i) = \alpha \text{MSE}(E_{xc}) + \beta \text{MSE}(v_{xc}) + \gamma \text{MSE}\left(\frac{dv_{xc}(x)}{dx}\right) + \delta \text{MSE}\left(E_{xc} - \int dx v_{xc}(x) n(x)\right) \quad (5)$$

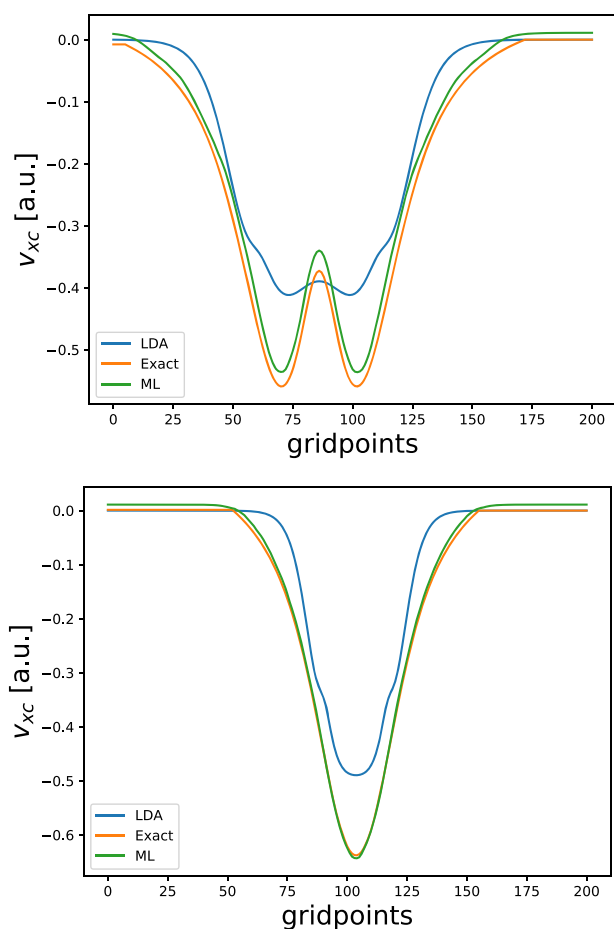
This function is a weighted combination of the mean-squared errors (MSE) of the exchange–correlation energy, the exchange–correlation potential, its numerical spatial derivative, and the difference between the exchange–correlation energy and the integral over the potential. This latter term is part of the formula for the total energy (1) and theoretically allows for some error cancellation. We also attempted to use the integral as a separate term in eq 5. Depending on the network, one or the other term produced better results. Finally, the weights  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  in eq 5 are also optimized as part of the hyperparameter optimization. Usual values for  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$  are 1.0, 100.0, 10.0, 1.0.

The training for the exchange–correlation energy converges quite fast after a few hundred epochs (i.e., one complete pass of the training data). The convergence of the potential can take thousands of epochs depending on the training set and batch size. At each training step the model was saved if it improved the validation error for the potential. The model with the lowest validation error was later used for the self-consistent Kohn–Sham calculations. As the amount of memory that is needed per sample is quite limited, very large batch sizes (e.g., 4096) are possible, allowing for a far more efficient parallelization of the training. Training with larger batch sizes seems to produce better convergence. However, it leads to a strong increase in the error during validation with self-consistent Kohn–Sham calculations. Smaller batch sizes (32, 64, 128) improve the error by up to 50% and provide the best generalization ability of the functionals, in consistency with the literature.<sup>52,53</sup>

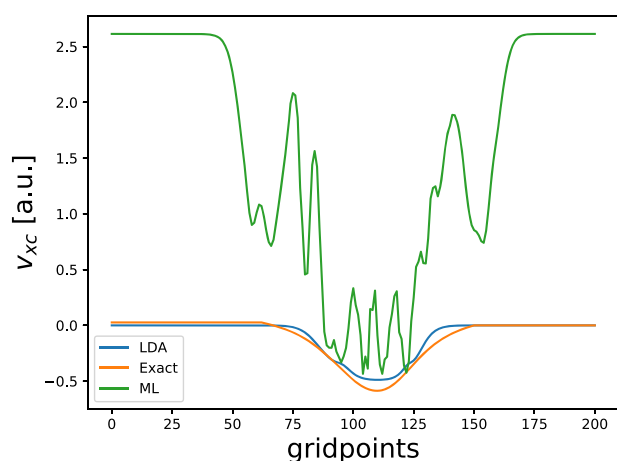
**Evaluation.** We trained neural networks with various kernel sizes and used them within a self-consistent Kohn–Sham calculations for a test set of 2000 systems created with the method described above. The self-consistent Kohn–Sham calculations were run using a self-written code. For all kernel sizes, models with different hyper-parameters were evaluated on a validation set of 250 systems. The training was not completely converged at this stage. Only for the best models of each kernel size did we continue the training and evaluate the models on the test set. To compare various models with respect to the LDA functional of DFT, we chose energy differences, as these are physically more meaningful.

In Figure 2 we plot the exchange–correlation potential resulting from self-consistent calculations with a ML functional trained according to the loss function (5) for two test-systems and compared it with the exact and LDA predicted exchange–correlation potentials. In Figure 3 we plot the same information for one test system; the ML functional is this time trained only on the exchange–correlation energy (i.e., not on the potential). Whenever the machine-learned functional is trained on both the energy and the potential, the exchange–correlation potential presents a great improvement in





**Figure 2.** Comparison of exchange–correlation potentials of an 1D-LDA,<sup>54</sup> the exact potential, and our ML-functional with kernel size 30 for two different systems.



**Figure 3.** Comparison of exchange–correlation potentials of an 1D-LDA,<sup>54</sup> the exact potential, and our ML-functional trained only with the exchange–correlation energy.

comparison to the traditional LDA functional, while the functional trained only on the energy fails completely. Remarkably, the functionals trained with the loss-function (5) also show a qualitatively closer behavior to the exact exchange–correlation potentials. The results for the predicted

total energies of the test set (relative to the energy of H<sub>2</sub> at its equilibrium distance) are presented in Table 1. Our ML-LDA

**Table 1.** Mean Absolute Errors (MAE) for the Total Energy in Self-Consistent Calculations for Various Kernel Sizes of Our ML-DFT Functional, Relative to the Error of the One-Dimensional LDA of Ref 54<sup>a</sup>

kernel size	MAE(ML)/MAE(LDA) [%]
LDA	100
1	38.1
15	21.8
30	8.2
60	8.2
120	7.1
180	6.5

<sup>a</sup>For reference, the mean absolute error of the LDA of ref 54 is  $1.4 \times 10^{-2}$  au.

(i.e., the functional with kernel size  $\kappa = 1$ ) already performs better than the traditional LDA. As the ML-LDA was trained to reproduce the exchange–correlation energy of heterogeneous systems while traditional LDAs are “trained” for constant densities, the difference in performance is not surprising. Increasing the kernel size leads to a monotonical decrease of the error, and improves the results by more than a factor of 6 for the larger sizes. The optimal kernel size is, in our opinion, around 30 (i.e., 3 au), as larger kernels do not provide a significant advantage. Furthermore, some of the functionals with larger kernel sizes also demonstrate unphysical behavior (see below).

Ultimately, the more nonlocal the functional is, the higher the complexity and the larger the number of parameters. This reason, together with the need to represent more long-range interactions that are based on different physical principles (such as van der Waals interaction), makes the training considerably more difficult. One approach to circumvent this problem is to keep the nonlocality limited to ranges on the scale of molecular bonds. This allows for simpler training and still includes most of the nonlocality that is required for the exchange–correlation energy. Another possibility would be to enlarge the nonlocality by increasing the architectural complexity of the functional.

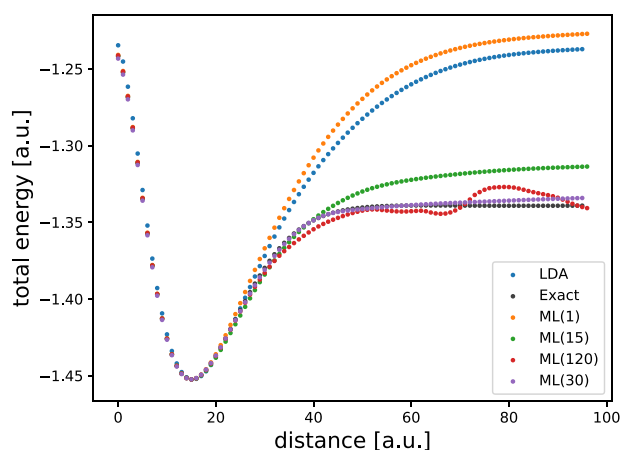
Efforts to decrease the number of training systems for a kernel size of 30 lead to a slightly increased error of 11% using 800 samples for training. Although the scaling of the networks to realistic three-dimensional systems is nontrivial, we expect that both the number of trainable parameters and the density points in the training set will grow cubically when transitioning to three-dimensional systems. In this sense, we expect a similar demand for training data as in 1D. Furthermore, realistic systems are usually far larger and therefore provide more “local” training samples per system for the neural network. Recent research by Nagai and coauthors points in the same direction.<sup>45</sup> Indeed, they only required a few sample molecules and used far more parameters to learn a much more local (and in this sense simpler) functional than the ones used here.

Previously, we tested the ML-functional on sample systems belonging to the distribution of the training data. Now, we go a step further and test how our functionals perform self-consistently on systems outside this distribution as well as a couple of paradigmatic cases. The systems in the training set had external potentials arising from 2 and 3 nuclei. In order to



go beyond these systems, we tested the functional with kernel size 30 also on a test set of 150 systems with 4 nuclei. Using the same functional as in Table 1, we arrive at an error for the total energy more than eight times smaller than with the LDA ( $\text{MAE}(\text{ML})/\text{MAE}(\text{LDA}) = 11.9\%$ ). Naturally, the error increased outside the training distribution. However, considering the different nature of the highly charged systems with 4 nuclei, this hints at a good generalization ability of the functional.

In Figure 4 we present our calculations for the  $\text{H}_2$  molecule in 1D along the dissociation path with functionals of varying



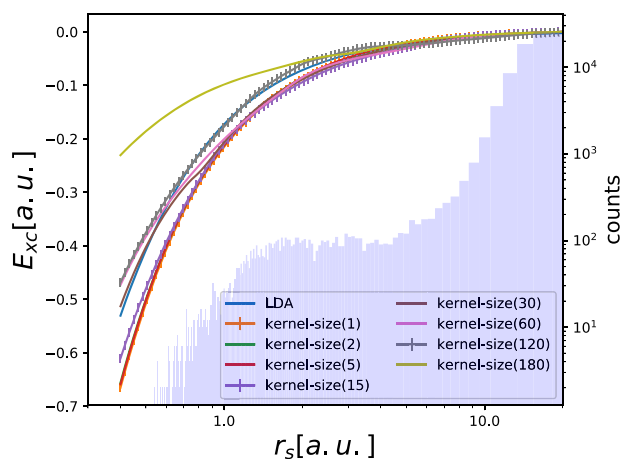
**Figure 4.** Dissociation curves of the 1D  $\text{H}_2$  molecule with ML functionals of varying nonlocality (i.e., kernels of 1, 15, 120, 30) in comparison to the exact and the LDA results.

nonlocality in comparison to the exact result. The curves in Figure 4 are shifted to have the same equilibrium energies. As is well-known, the traditional LDA completely fails to produce the correct dissociation limit.<sup>54</sup> The same behavior is observed for the ML LDA (with a kernel size of 1). Remarkably, using increasingly more nonlocal functionals, we can reduce the relative energy error to 3.2% of the LDA error.

It is obvious that even the functional with a kernel size of 30 will start failing above a certain distance. This is a conceptual problem of local KS-DFT and can only be alleviated and not eliminated in our approach. It can already be considered a success that our functionals are able to reproduce the dissociation curve reasonably well far beyond their own degree of nonlocality.

Yet it has to be noted that not all functionals performed that well. Some of the functionals with larger kernels failed to reproduce a physical behavior with respect to the dissociation distance and produced multiple local minima and maxima. Despite these problems, they still return the correct equilibrium distance and on average far better energies than the LDA. This unphysical behavior of some functionals just stresses the fact that a rigorous validation on a multitude of different systems will be essential to arrive at a working functional. It has to be noted that a larger training set and a longer training time was far more beneficial for this validation than for example the average error. As systems similar to the dissociated molecule are most likely outliers of the training data this is not surprising. Note that the need for more training data can, however, be avoided by active learning and a thoughtful construction of the training set.

Finally, we study the homogeneous-electron gas, a model system that is used in the construction of the majority of exchange–correlation functionals. We can simulate this system with our neural networks by providing them with a constant electronic density as input. The results can then be compared to the numerically exact values for the energy density as obtained, for example, from quantum Monte Carlo simulations.<sup>54</sup> Our results are depicted in Figure 5 as a function of



**Figure 5.** Exchange–correlation energy per unit volume of a 1D homogeneous electron gas from quantum Monte Carlo calculations<sup>54</sup> (curve labeled LDA) is compared with several ML functionals evaluated at constant density. We also plot a histogram of the number of systems of the training set containing more than three grid points with a density within a bin size of 0.01 au. The number of counts can be read on the right axis. Notice that there is basically no system with  $r_s < 1$ . The machine-learned curves are shifted to be exactly zero at zero density.

the Wigner–Seitz radius  $r_s = 1/2n$ . Remarkably, our ML functional with kernel size equal to 120 reproduces the exchange–correlation energy of the 1D homogeneous electron gas, especially for  $r_s > 2$ , where the majority of counts in the training data were located. Furthermore, kernel sizes lower than or equal to 60 are practically indistinguishable for larger  $r_s$  and behave qualitatively very close to the homogeneous electron gas. The largest kernel size (say, 180) underestimates the exchange–correlation energy.

Some difficulties for our neural-network functionals are evident. First, the functionals were trained only for systems with a specific size while the homogeneous electron gas is, in fact, an infinite periodic system. Second, as the histogram in Figure 5 illustrates, the training data does not contain almost any samples with high densities ( $r_s < 1$ ). Naturally, the availability of training data similar to the homogeneous electron gas is even more important for the more nonlocal functionals as they take into account larger regions of space. The first challenge results in the fact that the nonzero biases in each layer cause the neural networks to output a nonzero value for zero density. When training for different system sizes, there are several ways to avoid this failure. First, one could solve the problem by adding systems padded with different amounts of zeros at the border to force the neural network to learn the correct relationship. As a second possibility, one could force all biases of all layers to zero, however, this would severely limit the expressibility of the networks. To circumvent this problem, and in order to compare the behavior of the energy with

respect to the Wigner–Seitz radius, we shifted the curves in Figure 5 to yield zero energy for zero density.

Despite the small amount of training data at high density, functionals with larger kernel sizes still generalize on average far better to the homogeneous electron gas. While this is the case for most models, there are some rare cases, similar to the problems with the H<sub>2</sub> dissociation, where large kernel sizes produce unphysical behavior. It is not obvious whether this result will remain true in three dimensions, it is nevertheless promising that the extra nonlocal information in the larger kernels might help the functionals to be generalized. Constant densities will be an essential feature of a functional for solid-state physics. Fortunately, exact training data in the form of quantum Monte Carlo calculations already exists for this purpose and can be easily incorporated in our training sets.

In conclusion, in this Letter we have demonstrated the viability of learning an exchange–correlation potential, via the differentiation of the exchange–correlation energy in a physically consistent manner. This procedure allows for standard self-consistent Kohn–Sham calculations. From the presented data, it is evident that neural-network functionals trained on the exchange–correlation potential and energy have the potential to be far more precise than previous local DFT functionals. Increasing the nonlocality of the functional allows for an extremely precise treatment of the electronic interaction on the scale of at least a few atomic units and, to a certain extent, even solve long-standing problems in DFT like, e.g., molecular dissociation.

For simplicity, we trained a neural network to the one-dimensional two-electron problem in the external potential generated by up to three nuclei. Training three-dimensional systems will have to be accomplished by using data obtained with coupled-cluster, full configuration-interaction, or quantum Monte Carlo. While sufficient data to train a universal functional still has to be created, exchange–correlation energies and potentials for a few small molecules already exists and can provide a good starting point. The density representation on a grid is unfortunately not feasible for more general systems, as grid sizes and forms will vary. However, we think this can easily be circumvented by representing the density locally in some basis sets (e.g., Gaussians).

Finally, there is already a long history within DFT in the development of empirical functionals.<sup>32,33,55–57</sup> The machine learning paradigm allows us to drastically increase the amount of data used for the training and the complexity of these functionals. Including known exact conditions of the exchange–correlation functional in the learning process as constraints in the minimization will still be helpful<sup>58</sup> and provide further conditions for validation. Furthermore, as the functionals will have to work in practically every density environment, the importance of an extremely in-depth validation cannot be overstated and will be essential to arrive at a widely used functional.

## AUTHOR INFORMATION

### Corresponding Authors

\*E-mail: [carlos.benavides-riveros@physik.uni-halle.de](mailto:carlos.benavides-riveros@physik.uni-halle.de).

\*E-mail: [miguel.marques@physik.uni-halle.de](mailto:miguel.marques@physik.uni-halle.de).

### ORCID

Carlos L. Benavides-Riveros: 0000-0001-6924-727X

Miguel A. L. Marques: 0000-0003-0170-8222

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We acknowledge partial support from the German DFG through project MA-6786/1.

## REFERENCES

- (1) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864.
- (2) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133.
- (3) Ceperley, D. M.; Alder, B. J. Ground State of the Electron Gas by a Stochastic Method. *Phys. Rev. Lett.* **1980**, *45*, 566.
- (4) Perdew, J. P.; Schmidt, K. Jacob's ladder of density functional approximations for the exchange–correlation energy. *AIP Conf. Proc.* **2000**, *577*, 1.
- (5) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **1980**, *58*, 1200.
- (6) Cole, L. A.; Perdew, J. P. Calculated electron affinities of the elements. *Phys. Rev. A: At., Mol., Opt. Phys.* **1982**, *25*, 1265.
- (7) Perdew, J. P.; Wang, Y. Accurate and simple analytic representation of the electron–gas correlation energy. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1992**, *45*, 13244.
- (8) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.
- (9) Perdew, J. P.; Burke, K.; Wang, Y. Generalized gradient approximation for the exchange–correlation hole of a many-electron system. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 16533.
- (10) Sun, J.; Remsing, R. C.; Zhang, Y.; Sun, Z.; Ruzsinszky, A.; Peng, H.; Yang, Z.; Paul, A.; Waghmare, U.; Wu, X.; Klein, M. L.; Perdew, J. P. Accurate first-principles structures and energies of diversely bonded systems from an efficient density functional. *Nat. Chem.* **2016**, *8*, 831.
- (11) Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372.
- (12) Perdew, J. P.; Ernzerhof, M.; Burke, K. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.* **1996**, *105*, 9982.
- (13) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158.
- (14) Lehtola, S.; Steigemann, C.; Oliveira, M. J. T.; Marques, M. A. L. Recent developments in libxc – A comprehensive library of functionals for density functional theory. *SoftwareX* **2018**, *7*, 1.
- (15) Becke, A. D. Perspective: Fifty years of density-functional theory in chemical physics. *J. Chem. Phys.* **2014**, *140*, 18A301.
- (16) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Insights into Current Limitations of Density Functional Theory. *Science* **2008**, *321*, 792.
- (17) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Fractional spins and static correlation error in density functional theory. *J. Chem. Phys.* **2008**, *129*, 121104.
- (18) Skone, J. H.; Govoni, M.; Galli, G. Self-consistent hybrid functional for condensed systems. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 195112.
- (19) Benavides-Riveros, C. L.; Lathiotakis, N. N.; Marques, M. A. L. Towards a formal definition of static and dynamic electronic correlations. *Phys. Chem. Chem. Phys.* **2017**, *19*, 12655.
- (20) Liu, S.-S.; Tian, Y.-T. *Advances in Neural Networks*; Springer: Berlin, Heidelberg, 2010; pp 144–151.
- (21) Waibel, A.; Lee, K.-F., Eds. *Readings in Speech Recognition*; Morgan Kaufmann, 1990.
- (22) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547.

- (23) Schmidt, J.; Marques, M. R. G.; Botti, S.; Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *Npj Comput. Mater.* **2019**, *5*, 83.
- (24) Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M. Big Data of Materials Science: Critical Role of the Descriptor. *Phys. Rev. Lett.* **2015**, *114*, 105503.
- (25) Tozer, D. J.; Ingamells, V. E.; Handy, N. C. Exchange-correlation potentials. *J. Chem. Phys.* **1996**, *105*, 9200.
- (26) Snyder, J. C.; Rupp, M.; Hansen, K.; Müller, K.-R.; Burke, K. Finding Density Functionals with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 253002.
- (27) Snyder, J. C.; Rupp, M.; Hansen, K.; Blooston, L.; Müller, K.-R.; Burke, K. Orbital-free bond breaking via machine learning. *J. Chem. Phys.* **2013**, *139*, 224104.
- (28) Yao, K.; Parkhill, J. Kinetic Energy of Hydrocarbons as a Function of Electron Density and Convolutional Neural Networks. *J. Chem. Theory Comput.* **2016**, *12*, 1139.
- (29) Brockherde, F.; Vogt, L.; Li, L.; Tuckerman, M. E.; Burke, K.; Müller, K.-R. Bypassing the Kohn-Sham equations with machine learning. *Nat. Commun.* **2017**, *8*, 872.
- (30) Liu, Q.; Wang, J. C.; Du, P. L.; Hu, L. H.; Zheng, X.; Chen, G. Improving the Performance of Long-Range-Corrected Exchange-Correlation Functional with an Embedded Neural Network. *J. Phys. Chem. A* **2017**, *121*, 7273.
- (31) Nagai, R.; Akashi, R.; Sasaki, S.; Tsuneyuki, S. Neural-network Kohn-Sham exchange-correlation potential and its out-of-training transferability. *J. Chem. Phys.* **2018**, *148*, 241737.
- (32) Lundgaard, K. T.; Wellendorff, J.; Voss, J.; Jacobsen, K. W.; Bligaard, T. mBEEF-vdW: Robust fitting of error estimation density functionals. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *93*, 235162.
- (33) Wellendorff, J.; Lundgaard, K. T.; Møgelhøj, A.; Petzold, V.; Landis, D. D.; Nørskov, J. K.; Bligaard, T.; Jacobsen, K. W. Density functionals for surface science: Exchange-correlation model development with Bayesian error estimation. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2012**, *85*, 235149.
- (34) Li, L.; Baker, T. E.; White, S. R.; Burke, K. Pure density functional for strong correlation and the thermodynamic limit from machine learning. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *94*, 245129.
- (35) Krieger, J. B.; Li, Y.; Iafate, G. J. Derivation and application of an accurate Kohn-Sham potential with integer discontinuity. *Phys. Rev. Lett. A* **1990**, *146*, 256.
- (36) Sharp, R. T.; Horton, G. K. A Variational Approach to the Unipotential Many-Electron Problem. *Phys. Rev.* **1953**, *90*, 317.
- (37) Talman, J. D.; Shadwick, W. F. Optimized effective atomic central potential. *Phys. Rev. A: At., Mol., Opt. Phys.* **1976**, *14*, 36.
- (38) van Leeuwen, R.; Baerends, E. J. Exchange-correlation potential with correct asymptotic behavior. *Phys. Rev. A: At., Mol., Opt. Phys.* **1994**, *49*, 2421.
- (39) Becke, A. D.; Johnson, E. R. A Simple Effective Potential for Exchange. *J. Chem. Phys.* **2006**, *124*, 221101.
- (40) Tran, F.; Blaha, P. Accurate Band Gaps of Semiconductors and Insulators with a Semilocal Exchange-Correlation Potential. *Phys. Rev. Lett.* **2009**, *102*, 226401.
- (41) Gaiduk, A. P.; Staroverov, V. N. How to tell when a model Kohn-Sham potential is not a functional derivative. *J. Chem. Phys.* **2009**, *131*, 044107.
- (42) Borlido, P.; Aull, T.; Huran, A. W.; Tran, F.; Marques, M. A. L.; Botti, S. Large-Scale Benchmark of Exchange-Correlation Functionals for the Determination of Electronic Band Gaps of Solids. *J. Chem. Theory Comput.* **2019**, *15*, 5069.
- (43) Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. *Automatic differentiation in pytorch. NIPS 2017 Autodiff Workshop: The Future of Gradient-based Machine Learning Software and Techniques.* 2017.
- (44) Abadi, M.; et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.* <https://tensorflow.org/>, 2015.
- (45) Nagai, R.; Akashi, R.; Sugino, O. Completing density functional theory by machine-learning hidden messages from molecules. *arXiv:1903.00238*, **2019**.
- (46) Wagner, L. O.; Stoudenmire, E. M.; Burke, K.; White, S. R. Reference electronic structure calculations in one dimension. *Phys. Chem. Chem. Phys.* **2012**, *14*, 8581.
- (47) Andrade, X.; et al. Real-space grids and the Octopus code as tools for the development of new simulation approaches for electronic systems. *Phys. Chem. Chem. Phys.* **2015**, *17*, 31371.
- (48) Jensen, D. S.; Wasserman, A. Numerical methods for the inverse problem of density functional theory. *Int. J. Quantum Chem.* **2018**, *118*, e25425.
- (49) Clevert, D.-A.; Unterthiner, T.; Hochreiter, S. Fast and accurate deep network learning by exponential linear units (elus). *arXiv:1511.07289*, **2015**.
- (50) Ignite. <https://github.com/pytorch/ignite>, 2018.
- (51) Kingma, D. P.; Ba, J. Adam: A method for stochastic optimization. *arXiv:1412.6980*, **2014**.
- (52) Masters, D.; Luschi, C. Revisiting small batch training for deep neural networks. *arXiv:1804.07612*, **2018**.
- (53) Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press, 2016; <http://www.deeplearningbook.org>.
- (54) Helbig, N.; Fuks, J. L.; Casula, M.; Verstraete, M. J.; Marques, M. A. L.; Tokatly, I. V.; Rubio, A. Density functional theory beyond the linear regime: Validating an adiabatic local density approximation. *Phys. Rev. A: At., Mol., Opt. Phys.* **2011**, *83*, 032503.
- (55) Yu, H. S.; He, X.; Truhlar, D. G. MN15-L: A New Local Exchange-Correlation Functional for Kohn-Sham Density Functional Theory with Broad Accuracy for Atoms, Molecules, and Solids. *J. Chem. Theory Comput.* **2016**, *12*, 1280.
- (56) Yu, H. S.; He, X.; Li, S. L.; Truhlar, D. G. MN15: A Kohn-Sham global-hybrid exchange-correlation density functional with broad accuracy for multi-reference and single-reference systems and noncovalent interactions. *Chem. Sci.* **2016**, *7*, 5032.
- (57) Mardirossian, N.; Head-Gordon, M.  $\omega$ B97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **2016**, *144*, 214110.
- (58) Hollingsworth, J.; Baker, T. E.; Burke, K. Can exact conditions improve machine-learned density functionals? *J. Chem. Phys.* **2018**, *148*, 241743.

## 4.2 Crystal-Graph Attention Networks for the Prediction of Stable Materials

In section 3.2, we have introduced structure- and composition-based graph neural networks in the context of predicting structure-property relations and have discussed the difficulty of predicting properties from unrelaxed structures.

In the next publication ”**Crystal-Graph Attention Networks for the Prediction of Stable Materials**” [JSPHD10] we overcome this challenge by reducing the edge representation to the graph distance between atoms. This imprecise edge representation drastically reduces the dependence on the correct volume or lattice vectors of the input structure and allows for accurate predictions from unrelaxed structures.

To make use of the structure sensitivity of the models, we create a large dataset, accumulating and filtering millions of data points from AFLOW [27], the Materials Project [67] and publications from our group [166, 135, JS3, JS4, JSPHD12]. This results in one of the largest convex hulls and consistent DFT datasets to date.

In the following, we employ this new dataset, and an extra dataset of mixed perovskites calculated for this paper, to train and evaluate the model.

One of the major objectives of the model is its ability to go beyond compositional models and learn to predict the distance to the convex hull from different structure prototypes while not relying on precise structural inputs. To test this hypothesis, we use the large dataset for pretraining and demonstrate that the pretraining can substantially reduce the training data required to achieve a certain MAE for a prototype. For mixed perovskites as an example, the pretraining reduces the training data needed to reach a sufficient MAE for a high-throughput

search by half. We also compare these results to the composition-based 3D-ElemNet model (see section 5) and find that independent of the dataset size, the errors of the 3D-ElemNet model are roughly double that of the pretrained CGAT model. Furthermore, we test the model on a quaternary Heusler dataset from Ref. [75] finding a four times lower test error even without pretraining. In an additional experiment, we demonstrate that the vectorized attention coefficients we introduce in this work lead to a performance gain of more than 10% in the case of the mixed perovskites.

As we have discussed in section 3.2, for most structures, the neighbor lists still change during a geometry relaxation, and thus the CGAT output varies. To predict the minimum energy, we use structures with multiple cell constant ratios as input and the prediction is selected as the one with the minimum energy. We demonstrate for mixed perovskites that just four different cell constant ratios as input are sufficient to reduce the error for the distance to the convex hull to within a few percent of the MAE for relaxed structures.

Finally, we employ the network to perform a high-throughput search of the mentioned mixed perovskites. Searching through 16 million compounds, we discover 8681 compounds below 100 meV/atm, 404 below 5 meV/atm and 325 on the hull. We have found in Ref. [JSPHD12] that distortions, different arrangements of the C and D atoms and configurational entropy can stabilize mixed perovskites by more than 150 meV/atm. In light of this fact, an experimental realisation of far more than just the compounds on the hull might be feasible.

We also find that the performance of the network varies significantly for different prototypes as well as elements. We assume that this is due to a bias of the training set that mainly contains materials result-

ing from high-throughput searches or known stable materials. Consequently, a few prototypes, and elements such as oxygen are very much overrepresented, while elemental groups such as lanthanides are underrepresented.

### **Follow up Work**

CGATs and potentially other more recent architectures [52] seem to be sufficient from a machine learning point of view to solve the scientific question of high-throughput searches for thermodynamically stable materials. While there is still the issue of the bias in the dataset that we will discuss later, the main task now is to perform an enormous number of high-throughput searches [JSPHD6].

The dataset used in Ref. [JSPHD11] consists of roughly 25k prototypes. Here we note every crystal structure as a separate prototype if the structure matcher from pymatgen [115] considers the structures different. We use the default settings for the structure matcher and allow for arbitrary scaling of the volume and coloring of the Wyckoff positions. Even though we do not need to calculate new training data for any prototype, all compounds close to stability still require DFT validation. Consequently, we still have to limit the number of prototypes. We are now searching through all binary, ternary and quaternary prototypes with less than twenty atoms in the unit cell that appear more than ten times in our database and have space group numbers larger than ten. The first two criteria were selected to limit the cost of the DFT calculations, while the last ensures some previous scientific interest in the prototype.

We apply a cutoff of 50 meV/atm distance to the convex hull for the predictions that we validate with DFT. We have already scanned all 639 selected binary and 1829 selected ternary prototypes spanning a search

space of 1.05B compounds using another 463986 DFT calculations.<sup>2</sup> Assuming we found the majority of stable compounds, this is a speedup of approximately a factor of 2200.

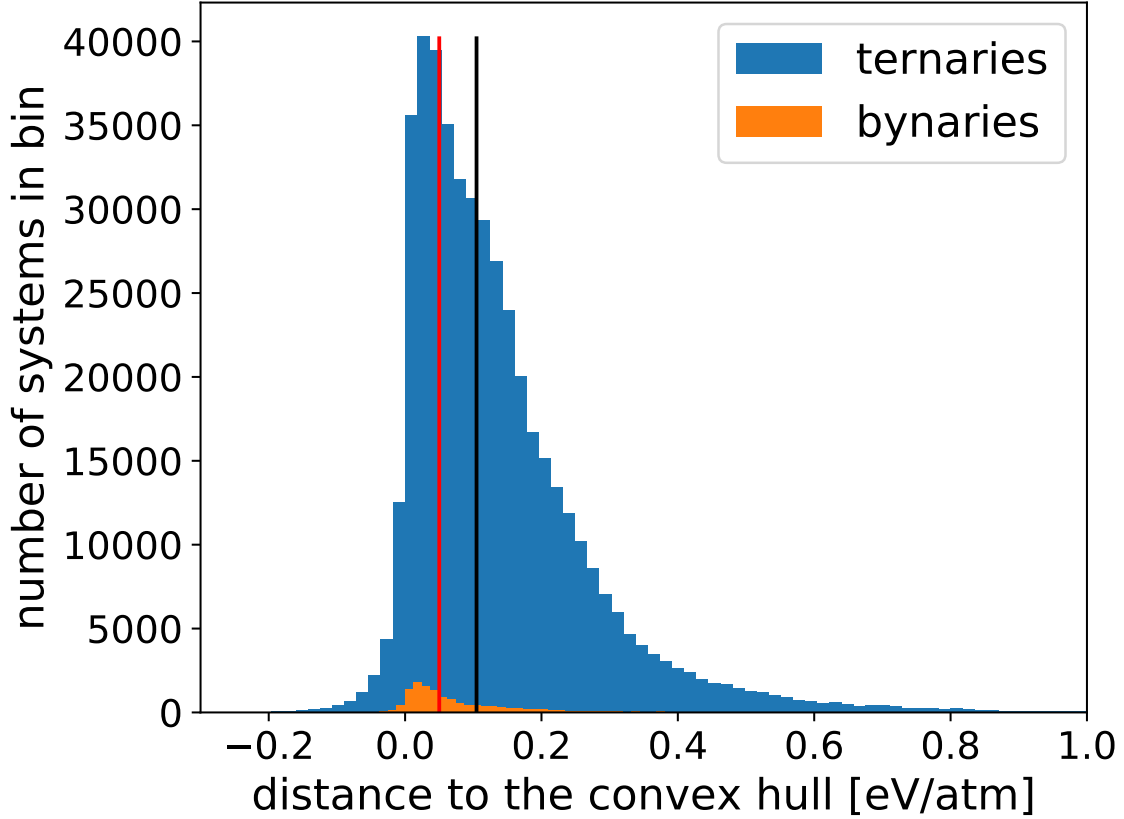


Figure 7: Histogram of the DFT distances to the convex hull for validated predictions of 1829 ternary and 639 binary prototypes. The medians of the binary and ternary dataset are marked with a red and a black vertical line, respectively. All materials predicted to be below 50 meV/atm were validated with DFT calculations.

<sup>2</sup>We performed all calculations discussed in this section with the same parameters as in Ref. [JSPHD10]. The machine learning model was a CGAT network trained with the same hyperparameters as in Ref. [JSPHD10]. We used all the data in Ref. [JSPHD10] including the quaternary perovskites for training, and the resulting calculations from the Garnet high-throughput search were also added to the training data for the large scale high-throughput search of all prototypes. Furthermore, we recalculated all AFLOW compounds identified as outliers in Ref. [JSPHD10] and added them to the hull and training data.



For the validated compounds, the MAE is at 113 meV/atm and the median distance to the convex hull is at 105 meV/atm for the ternary prototypes and 50 meV/atm for the binary prototypes. This results in 41627 ternary and 2035 binary compounds below 10 meV/atm and respectively 22145 and 914 materials below the hull. These stabilities are calculated based on the convex hull before the high-throughput search. We are currently scanning the leftover quaternary prototypes spanning a search space of over 13B compounds. For the quaternary prototypes, we will employ various strategies to reduce the bias in the training set that we saw in Ref. [JSPHD10]. The same bias is also visible in fig. 8. Here we depict the mean and standard deviation of the absolute error for different ternary prototypes from the already performed high-throughput search. It is apparent that a few outlier prototypes have extremely high MAEs. We plan to overcome this challenge by identifying these prototypes and adding corresponding training data through an active learning loop using an ensemble of models.

Lastly, we perform a more detailed high-throughput search of cubic garnets with the composition  $A_{12}B_{20}C_{48}$  using the model trained in Ref. [JSPHD10]. With this search, we demonstrate that even large unit cells of 80 atoms are in the reach of comprehensive high-throughput searches. After DFT validation, we identify 1474 garnets below 100 meV/atm, 391 below 10 meV/atm and 88 below the hull. This includes sulfide and nitride garnets, two previously unknown groups of garnets. A more detailed analysis of these results will be part of future publications, but we think the results already demonstrate the potential of CGATs to expand our knowledge of stable materials significantly. <sup>3</sup>

---

<sup>3</sup>Concerning the following article: Reprinted with permission from Jonathan Schmidt *et al.*, *Sci. Adv.* 7.49 (2021), eabi7948, Copyright © 2021 The Authors, some rights reserved; exclusive license American Association for the Advancement of Science.



## MATERIALS SCIENCE

# Crystal graph attention networks for the prediction of stable materials

Jonathan Schmidt<sup>1</sup>, Love Pettersson<sup>2</sup>, Claudio Verdozzi<sup>2</sup>, Silvana Botti<sup>3\*</sup>, Miguel A. L. Marques<sup>1</sup>

Graph neural networks for crystal structures typically use the atomic positions and the atomic species as input. Unfortunately, this information is not available when predicting new materials, for which the precise geometrical information is unknown. We circumvent this problem by replacing the precise bond distances with embeddings of graph distances. This allows our networks to be applied directly in high-throughput studies based on both composition and crystal structure prototype without using relaxed structures as input. To train these networks, we curate a dataset of over 2 million density functional calculations of crystals with consistent calculation parameters. We apply the resulting model to the high-throughput search of 15 million tetragonal perovskites of composition  $ABCD_2$ . As a result, we identify several thousand potentially stable compounds and demonstrate that transfer learning from the newly curated dataset reduces the required training data by 50%.

## INTRODUCTION

Machine learning methods have found increasing success in materials science and solid-state physics (1–6). Requiring orders of magnitude less computation time than traditional approaches such as density functional theory (DFT), machine learning methods allow for the prediction of material properties with close to ab initio accuracy. In the past few years, machines were developed to predict a plethora of physical properties, ranging from bandgaps (7–9), hardness (10), magnetic transition temperatures (11), etc. A particularly interesting property is the energy that ultimately determines the stability of a given material. Therefore, it is not unexpected that predicting the energy is essential for the challenging task of finding new stable compounds.

The modern theoretical approach to finding new materials involves scanning the whole composition space of one crystal structure, optimizing each crystal with DFT, and then comparing the DFT energy with all possible decomposition channels. Binary composition spaces are easily in the reach of DFT and have been extensively explored in the past (12). However, there are already around  $10^5$  different ternary combinations of chemical elements, and these can exist in a large number of different stoichiometries. Quaternary and higher prototypes are simply out of the reach of systematic DFT searches. Machine learning strategies have shown a lot of promise to speed up this process (13–18), with different approaches being proposed in the past. The main and most efficient approach to high-throughput searches is to calculate the distance to the convex hull of thermodynamic stability for all compositions of a single prototype (16–26). This step can be substantially accelerated by a machine learning model trained for the specific prototype, requiring separate training data for every prototype (16, 17, 20). An alternative is the development of composition-based models that are prototype agnostic (27, 28). These models can determine potentially stable compositions; however, they do not yield any information about the crystal structure of the material.

There are furthermore a large number of message passing networks (MPNs) (4–6, 29–32) that predict formation or absolute energies based on atomic positions and compositions. These networks usually achieve very high accuracy, but, unfortunately, they require a priori knowledge of the crystal structures (both lattice vectors and positions of the atoms) that are, in general, not available when searching for new materials.

In this work, we go beyond these approaches by developing a model that predicts the distance to the convex hull based on both the composition and the generic structure prototype but without requiring knowledge of the precise crystal structure. A previous approach that used the same philosophy is (23), where handcrafted structural features based on the Voronoi tessellations of the unit cell are used as input for random forests. A more recent work (30) also used Voronoi features, but, unfortunately, the error increases markedly for nonrelaxed structures [see the supplementary material of 30]. To circumvent these problems, instead of using handcrafted features, we combine techniques from previous deep MPNs (5, 6, 28, 29, 33, 34).

The goal of this work is to speed up prototype-based high-throughput searches beyond the possibilities of previous machine learning models. We remark that, as our model does not make use of the complete structural information, it does not give us access to forces and stresses and therefore cannot be used as a generic force field.

Just as important as the choice of model is, of course, the dataset. The commonly used datasets for machine learning are obtained from the materials project (35) and the open quantum materials database (OQMD) (36). The former is often used as a benchmark and as a training set for stability prediction [see, e.g., (37)]. However, as it contains mostly stable (or close to stable) compounds, one cannot realistically evaluate the error in the distance to the convex hull (or the formation energy) of a model trained exclusively on this dataset. The OQMD, on the other hand, is hard to combine with other large datasets because of the use of incompatible parameters in the DFT calculations. To construct a large dataset that allows for good transfer learning performance, we accumulated and curated data from various sources. In this way, we obtained a dataset that includes more than 2 million DFT calculations of both stable and unstable materials in a large variety of crystal structures.

Copyright © 2021  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
NonCommercial  
License 4.0 (CC BY-NC).

<sup>1</sup>Institut für Physik, Martin-Luther-Universität Halle-Wittenberg, 06120 Halle (Saale), Germany. <sup>2</sup>Department of Physics, Lund University Box 118, 221 00 Lund, Sweden. <sup>3</sup>Institut für Festkörperteorie und Optik and European Theoretical Spectroscopy Facility, Friedrich-Schiller-Universität Jena, D-07743 Jena, Germany.

\*Corresponding author. Email: silvana.botti@uni-jena.de

The remainder of this manuscript is structured as follows. In next section, we start with the developed model and the accumulated dataset. We then show the power of our model by studying in detail a quaternary family of perovskites. After the discussion of our results, we go over the details of the work in Materials and Methods.

## RESULTS

### Crystal graph attention networks

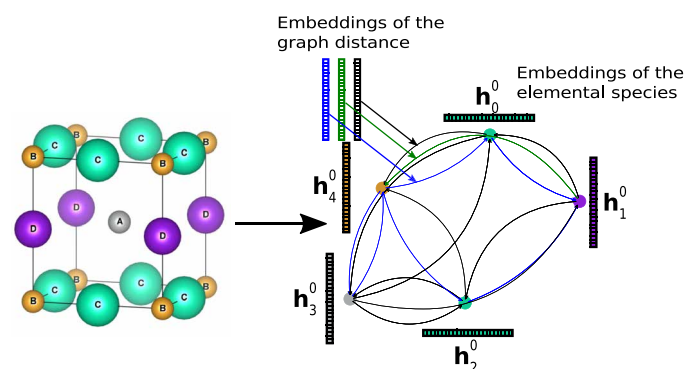
The crystal structure prototype will enter our model as a crystal graph. To incorporate the neighborhood information, each vertex is labeled by an embedding for the elemental species, and each edge by an embedding for the graph distance (see Fig. 1). The edge embeddings are initialized completely randomly, while the vertex embeddings are pretrained embeddings from (38). For the latter, an extra one-layer fully connected network is used to allow for possible changes in the embedding.

As in the MegNet model (5), during the message passing phase, the information of each vertex and its corresponding neighbors and edges is combined to calculate a new representation for each vertex and edge. This process is repeated several times until we arrive at a final representation for each vertex. This representation is then pooled, taking into account a global context vector along the ideas of (29). In the following, we discuss the mathematical details of the message passing phase. Vectors will be denoted in bold letters.

Each material starts with a representation  $\mathbf{h}_i^t$  for each atom  $i$ . This representation is updated through a message passing approach

$$\mathbf{h}_i^{t+1} = U(\mathbf{h}_i^t, \{\mathbf{h}_j, \mathbf{e}_{ij}\}) \quad (1)$$

where  $U$  is an update function that depends on the crystal graph attention networks' previous representation, the vertices of the neighbors, and the edges connecting the neighbors to the vertex. The neighbors are ordered by distance, and the edges are assigned corresponding embeddings for first neighbor, second neighbor, etc. These embeddings start as randomly initialized vectors and are trained together with the rest of the network. Naturally, one has to use periodic boundary conditions and a cutoff radius as we consider solids.



**Fig. 1. The crystal structure is transformed into a graph.** In this case, the crystal structure is a mixed perovskite, and we consider the five nearest neighbors. Here, blue edges represent first neighbors, black edges represent second neighbors, and green edges represent third neighbors. During the message passing steps, each individual edge and vertex embedding is updated based on its neighborhood.

The update function is based on the attention mechanism that has revolutionized natural language processing (39) and has also found application in graph neural networks (28, 33, 40). Previous graph attention networks applied to materials science used simple fully connected neural networks (FCNNs) to calculate a number of coefficients

$$s_{ij}^n = \text{FCNN}_a^n(\mathbf{h}_i^t \parallel \mathbf{h}_j^t \parallel \mathbf{e}_{ij}^t) \quad (2)$$

from the concatenation  $\parallel$  of the two vertex representations and the edge representation. Here, and in the following equations, the index  $n$  counts the number of FCNNs. These coefficients are normalized with a softmax function

$$a_{ij}^n = \frac{\exp(s_{ij}^n)}{\sum_j \exp(s_{ij}^n)} \quad (3)$$

In contrast to these previous works, we use vectors instead of coefficients, which effectively results in a separate attention coefficient for each element of the representation of each node/message. The resulting attention vectors  $\mathbf{a}_{ij}^n$  are used to weight the messages  $\mathbf{m}_{ij}^n$

$$\mathbf{m}_{ij} = \text{FCNN}_m^n(\mathbf{h}_i^t \parallel \mathbf{h}_j^t \parallel \mathbf{e}_{ij}) \quad (4)$$

when the representation is updated by

$$\mathbf{h}_i^{t+1} = \mathbf{h}_i^t + \text{HFCNN}_{\theta_g^t}^t(\parallel \sum_j \mathbf{a}_{ij}^n \mathbf{m}_{ij}^n) \quad (5)$$

Every pair of  $\text{FCNN}_a^n$  and  $\text{FCNN}_m^n$  can be considered one attention head. Ideally, each attention head learns to direct its focus to different features. This message passing procedure is repeated a number of times. At each message passing step, the edge embeddings are updated in a similar manner according to the following equations

$$s_{ij}^{e,n} = \text{FCNN}_a^{e,n}(\mathbf{h}_i^t \parallel \mathbf{h}_j^t \parallel \mathbf{e}_{ij}^t) \quad (6)$$

$$\mathbf{a}_{ij}^{e,n} = \frac{\exp(s_{ij}^{e,n})}{\sum_n \exp(s_{ij}^{e,n})} \quad (7)$$

$$\mathbf{m}_{ij}^{e,n} = \text{FCNN}_m^{e,n}(\mathbf{h}_i^t \parallel \mathbf{h}_j^t \parallel \mathbf{e}_{ij}^t) \quad (8)$$

$$\mathbf{e}_{ij}^{t+1} = \mathbf{e}_{ij}^t + \text{HFCNN}_{\theta_g^t}^{e,n}(\parallel \mathbf{a}_{ij}^{e,n} \mathbf{m}_{ij}^{e,n}) \quad (9)$$

The HFCNNs are hypernetworks along the work of (34), where the parameters of each network depend on the starting state of the node/edge and the state at step  $t$ .

$$\theta_g^t = f(\text{ch}_v^0 + (1 - c)\mathbf{h}_i^t) \quad (10)$$

Recently, (29) suggested using an extra global compositional vector for a last attention-based pooling layer. We follow this approach and use an extra roost (28) model that calculates a representation vector  $\mathbf{C}$  of the total composition also based on a graph attention network where the composition enters as a complete graph. This vector is then concatenated with the representation of each atom  $\mathbf{h}_i^T$  and used to calculate a final representation of the compound according to the following equations

$$s_i^n = \text{FCNN}_a^n(\mathbf{h}_i^T \parallel \mathbf{C}) \quad (11)$$

$$\mathbf{a}_i^n = \frac{\exp(s_i)}{\sum_i \exp(s_i^n)} \quad (12)$$

$$\mathbf{m}_i^n = \text{FCNN}_m^n(\mathbf{h}_i^n) \quad (13)$$

$$\text{Output} = \text{FCNN}_{RS} \sum_{i,n} \mathbf{m}_i^n \mathbf{a}_i^n \quad (14)$$

In this way, the network can once again evaluate the importance of the different elements using learned knowledge of the whole composition. A fully connected network with residual connections  $\text{FCNN}_{RS}$  (41) is used to obtain the final output.

While the study of Bartel *et al.* (37) recommends to learn formation energies, other machine learning studies came to different conclusions (24) and showed that it is advantageous to predict directly the distance to the convex hull. In particular, in (30), the authors obtained lower errors when learning formation energies; however, they showed that stability predictions were more successful when distances to the hull were predicted. For simplicity, we decided to predict directly the quantity we are interested in, the distance to the convex hull of stability.

## Data

As already discussed in Introduction, the datasets most commonly used for machine learning are the materials project (35) and the OQMD (36) as they use internally consistent parameters. The largest freely available database, Automatic FLOW for Materials Discovery (AFLOW) (12), is used less commonly even though it contains over 3 million compounds. Unfortunately, the OQMD can neither be easily combined with the materials project nor AFLOW as its calculation parameters are most often not compatible with the other two.

To construct a consistent dataset, we decided therefore to use all data from the materials project as well as all data from AFLOW that used functionals, pseudopotentials, and Hubbard  $U$ s that were consistent with the materials project. Furthermore, we added about 1.3 million compatible calculations from our group (16, 17, 42).

The final dataset after filtering all nonsuitable materials contained 2,093,838 compounds. Most of these belong to large groups of prototypes, such as cubic ternary perovskites (~230,000 systems),

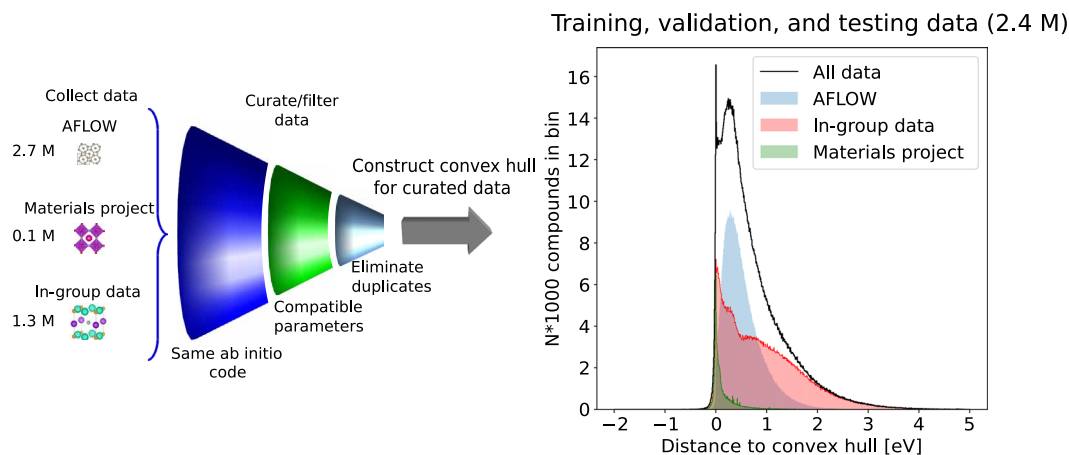
tetragonal mixed perovskites (~340,000 systems), chalcopyrites (~100,000 systems), and delafossites (~30,000 systems). This has to be considered during the evaluation, as the out-of-group prediction error (43) will be larger than for a randomly selected training and validation set.

The distribution of the distances to the convex hull of the final dataset is displayed in Fig. 2. It is evident that the materials project data consist mostly of compounds that are stable or close to stable with a mean, median, and SD of 220, 50, and 530 meV/atom. Because of its consistency in calculation parameters and ease of access, it is very commonly used as a benchmark set for machine learning algorithms. However, because it focuses on stable compounds, the distribution of distances to the convex hull is very different from the distribution of a random sample of compounds. Therefore, it is clear that this dataset is not convenient to train general machines to search for new materials. On the other hand, it is an ideal benchmark set for other properties, such as bandgaps, as these are mostly relevant for stable materials.

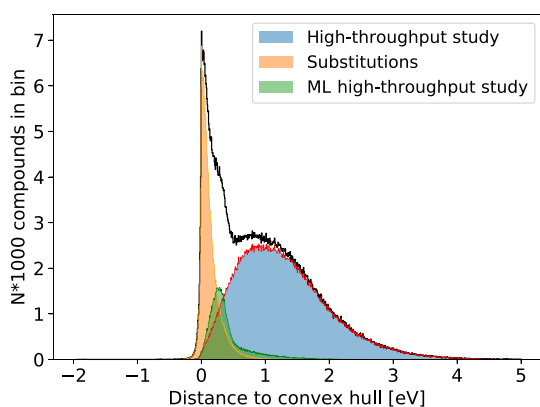
The curated AFLOW follows a typical skewed Gaussian with a mean, median, and SD of 530, 440, and 400 meV/atom.

The data from our group consist of multiple subsets as shown in Fig. 3. The highest peak arises from the data of (42), in which chemical elements were substituted by similar ones in stable materials. As we can see, this method leads to materials very close to stability. The green peak originates from machine learning-guided high-throughput calculations (17) (that resulted in relatively stable compounds) together with ~40,000 compounds from random training sets. The rest of the data consist of various traditional high-throughput searches of perovskites (16), mixed perovskites, chalcopyrites, Heusler compounds, and delafossites.

We will specifically focus on a dataset of mixed perovskites (see Fig. 1 for the crystal structure) as we apply the models developed in this work to a high-throughput search of their compositional space. We used a training set of around 180,000 randomly selected compositions. A further dataset of ~64,000 low-energy mixed perovskites was later selected by the machine learning models. This dataset was not considered in the calculation of the hull to avoid any leak of information into the training set.



**Fig. 2. Schema depicting the workflow for the creation of the dataset and the resulting energy distribution.** A total of 2.7 million calculations from AFLOW, 0.14 million from the materials project, and 1.3 million from our group were accumulated and curated, leaving in the end 2.09 million data points (0.96 million of which from AFLOW, 0.10 million from the materials project, and 1.02 million from our group data). A histogram depicting the distance to the convex hull of the final dataset is shown in the right. M, million.



**Fig. 3. Distribution of the distances to the convex hull for our group's data.** In orange, we show mostly stable or close to stable compounds resulting from substitutions of chemically similar elements into stable structures (43). High-throughput studies of several prototypes with all compositions are in blue/red. Machine learning (ML)-guided high-throughput study including the training set in green.

## Experiment

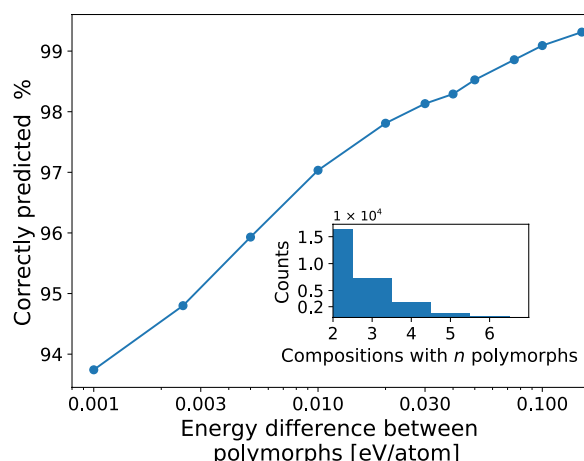
The network was first trained on the newly accumulated large dataset (minus the ~180,000 mixed perovskites used in the next section) with a training/validation/test split of 80%/10%/10%.

The network achieved a mean absolute error (MAE) in the training set of 30 meV/atom for our large dataset. As a large part of the dataset is composed of high-throughput calculations for a small number of crystal structure prototypes, this error does not properly represent the abilities of the model. Consequently, we will do an in-depth validation in the next section using a family of quaternary perovskites.

We also researched the ability of the model to differentiate the stability between different polymorphs. To accomplish this, we selected all structures whose composition appeared at least twice in the test set. We then checked whether the model predicts the correct relative stability for each pair of structures. In Fig. 4, we plot the percentage of correct orderings as a function of the minimal difference in energy between the polymorphs. In the inset, we show the distribution of compositions with a given number of polymorphs in the test set. As expected from a random test set, the majority of repeated compositions only appears twice, with around 7500 appearing three times, roughly 2800 four times, and 850 five times. Our model correctly predicts the relative stability of two polymorphs 93% of the time, increasing to 97% for polymorphs with an energy difference of 10 meV/atom and 99% for 100 meV/atom.

To achieve some comparability with other works, we also tested the model on the materials project dataset for the formation energies from (44). Using a training/validation/test split of 60%/20%/20%, we achieved a MAE of 41 meV/atom in the test set. The learning rate/batch size was changed in comparison to the other datasets because of the different dataset size. For the same set (44), crystal graph convolutional networks achieve an MAE of 33.2 meV/atom, while MEGNet achieves an MAE of 32.7 meV/atom. We have to note that our result is for a single validation split, while in (44), the error was averaged over five different splits. This slightly worse error is not unexpected as our networks do not use the complete information of the optimized crystal structure.

Furthermore, we trained our model on the quaternary Heusler dataset of (24). We used a training/validation/test split of 85%/5%/10% to be compared with 90%/10% of (24). We obtained a test error of



**Fig. 4. Percentage of correctly predicted relative stability between polymorphs versus the minimum difference in energy between the compared polymorphs.**

The main plot shows the percentage of correctly predicted relative stability between polymorphs as a function of the minimum difference in energy between the compared polymorphs. Inset: The distribution of chemical compositions that have a given number of polymorphs is shown. This data include all compositions that appear at least twice in the test set.

9 meV/atom in comparison to the best validation error of 37 meV/atom in (24), demonstrating the quality of our network.

## Validation for mixed perovskites

In this section, we analyze in detail the ability of the model trained in the previous section to be used in high-throughput searches. We concentrate on a family of perovskites that can be obtained by alloying the Wyckoff 3d position of the cubic  $ABX_3$  system, leading to the quaternary  $ABX_2Y$  composition. Among these, we find the well-known mixed-anionic oxynitride and oxyfluoride perovskites (45) that have found interesting applications such as in optoelectronics (46). The basic tetragonal crystal structure of our mixed perovskites is depicted in Fig. 1. We note that the size of the compositional space for the chemical elements we took into account is around 15 million. This includes not only the mixed-anionic systems but also inverted perovskites where the nonmetal occupies the center of the octahedra and the vertices are formed by a mixture of two metals. In view of the large number of stable inverted ternary perovskites (16) and the large number of possibilities for alloying two metals, we expect that the number of stable inverted quaternary perovskites will dwarf the number of mixed-anionic compounds.

All elements up to bismuth with the exception of the noble gases and the lanthanides were used. We did not use traditional stability tolerance factors, like the Goldschmidt factor or charge neutrality, in order not to bias the machine. If we consider compounds close to stability, we discover that the majority are not charge neutral in the standard (most common) oxidation state of the constituent chemical elements. If we use pymatgen to assign oxidation states to the structures below 100 meV/atom, we find that only 1056 cases have a charge neutral configuration, corresponding to 13% of the compounds. Considering these results, we believe that it is important not to bias training data using empirical rules but to rely exclusively on DFT results.

We started by constructing a dataset for random quaternary compositions composed of around 173,900 systems. This dataset was



split into a training set of 139,123 compounds and a validation and test set of 17,390 entries each. The MAE of the general machine discussed in the previous section in this test set is 508 meV/atom, which is considerably higher than the 30 meV/atom obtained for the large training set. This is not unexpected considering the facts that the pretraining dataset is extremely biased toward a few crystal prototypes and has a mean distance to the convex hull of 590 meV/atom, while the mixed perovskites set is extremely unstable with a mean of 1445 meV/atom. If we consider mixed perovskites closer to stability, e.g., below 500 meV/atom, then the MAE becomes a lot more reasonable at 132 meV/atom.

### Transfer learning

A way to improve the general model presented before is to use transfer learning. For this, we create sets containing from 2.5 to 80% of the 173,903 mixed quaternary compounds and use them to retrain our previous general model. In this case, the learning rate was reduced by a factor of 10, and no further hyperparameters were changed or optimized. For comparison, we also trained models only in these training sets (i.e., without pretraining) and with a three-dimensional (3D) ElemNet model (see Materials and Methods). The validation and test sets contained 17,390 compounds each as discussed before.

We also trained a Representation Learning from Stoichiometry (ROOST) method (28) to the perovskite data. However, because of the inability of ROOST to differentiate between different compounds with the same overall composition, the error of our trained models remained stubbornly high.

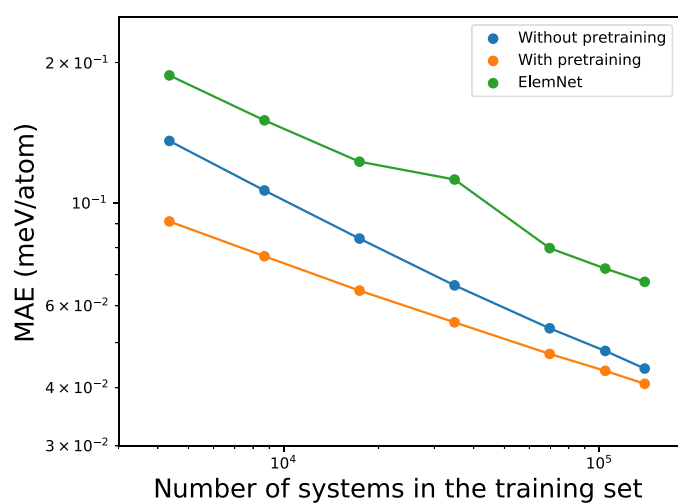
In Fig. 5, we show that the MAEs for the pretrained model are improved for all training set sizes and that the purely composition-based ElemNet is outclassed by the graph networks. However, the advantage of using a pretrained model drops off as the training set size increases and the majority of the information learned during the pretraining is already included in the training data. With pretraining, we arrive at an error of 62 meV/atom with a training set size of 17,400 systems, while more than 35,000 samples are required to achieve the same test error without pretraining. As this error is sufficient to start a machine learning-guided high-throughput search, we see that transfer learning can easily reduce the required training data by a factor of two.

We also trained a model using the scalar attention coefficients of (28, 33) on the mixed-perovskite dataset. The resulting test MAE was 11% higher, proving the superior quality of the vectorized attention operation used in this work.

In Fig. 6, we display the MAE for all mixed perovskites containing a certain chemical element. Note that we left out compounds including vanadium from the training set to investigate the transferability across the periodic table without extra training data (see the next section). As discussed in previous studies (16, 17), magnetic elements like chromium, manganese, and iron as well as some first row elements show a higher MAE. The first is most likely due to the magnetic interactions not being described properly by the DFT calculation, which leads to difficulty in learning, while the second is caused by the well-known first row anomaly (47, 48) of the periodic table.

### High-throughput search

As a further independent validation of our machines, we selected all mixed perovskites predicted to have a distance to the hull below 200 meV/atom. We used two different models: the best pretrained model from the previous section and the ElemNet model to select these compounds.



**Fig. 5. Test MAE versus number of systems in the training set.** We display the MAE for the test set of mixed perovskites in dependence of the number of mixed perovskite for both a pretrained as well as a nonpretrained crystal graph attention network and a 3D ElemNet model.

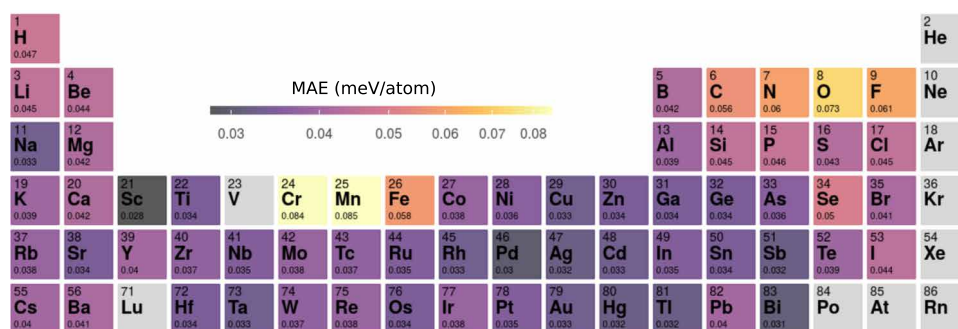
Our mixed perovskites have a tetragonal crystal structure. As such, the list of neighbors that are used as input features to our model depends to some extent on the ratio of the cell constants, which are unfortunately unavailable without performing the actual ab initio calculation. To circumvent this problem, we constructed prototypes with different ratios and used the lowest value of the distance to the convex hull as the actual prediction of the machine learning model. In practice, we used cell constant ratios of 0.85, 0.9, 1.1, and 1.15 for  $c/a$  while keeping  $a = b$  to maintain the symmetry of the system.

In total, 64,914 compounds were selected below the cutoff of 200 meV/atom. This choice is higher than the 70 to 100 meV/atom often used [see, for example, (49)] and is motivated by (45), where it was found that the large majority of experimentally known mixed perovskites are below 200 meV/atom when computed in the five-atom tetragonal unit cell.

As before, we optimized the geometries of these compounds with DFT and calculated the distance to the convex hull of thermodynamic stability. The MAE of the pretrained machine learning model for this new validation set using the DFT-relaxed structures was 33.5 meV/atom, which is comparable to the error in the validation set. The error for the unrelaxed structures decreases with the number of considered structures from 45.7 meV/atom for one unrelaxed structure to 36.5 meV/atom for the four considered structures.

Furthermore, for 88% of the materials, the model predicted correctly whether  $a > c$  or  $a < c$  was the more stable phase, illustrating the structure sensitivity of the model. We are now also in position to determine the generalization error of the machine to vanadium-containing compounds. The MAE for these materials is 87 meV/atom for the best pretrained model. This should be compared (see Fig. 6) 34 meV/atom for Ti, 84 meV/atom for Cr, and 85 meV/atom for Fe. We can see that the error is still perfectly acceptable, showing that the machine can reliably interpolate in the periodic table.

Last, in Fig. 7, we depict the distribution of the distance to the convex hull (calculated with DFT). If we consider various lower bounds, then we arrive at 21,333 materials below 150 meV/atom, 8681 materials below 100 meV/atom, 2405 materials below 50 meV/atom, 404 below 5 meV/atom, and 325 below the convex hull. We note that we demonstrated in a recent work (45) that many mixed



**Fig. 6. MAE for structures containing each element.** MAE for the mixed perovskites in the test set containing each chemical element.

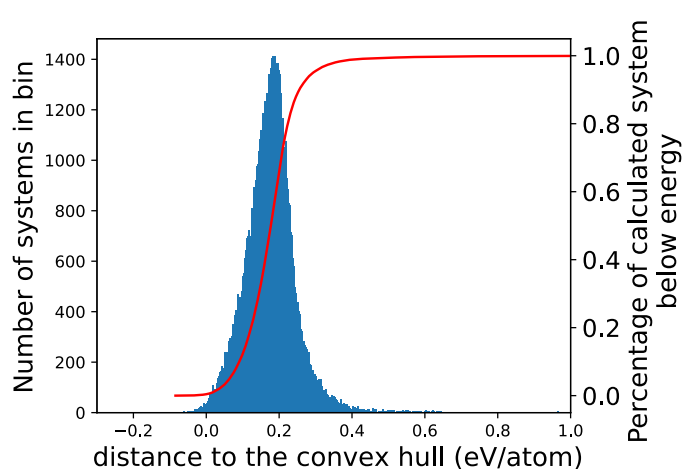
perovskites can be stabilized by a substantial amount through structural distortions (like the rotation and tilting of the octahedra) or by considering different arrangements of the C and D atoms. Some systems are stabilized by more than 150 meV/atom. Furthermore, configurational entropy, which was not considered in (45), is an additional stabilizing factor. Last, the experimentally known mixed perovskite with the highest DFT distance to the convex hull is  $\text{LaZrO}_2\text{N}$  at 260 meV/atom (45). As such, we believe that there is a very good chance that a large majority of the compounds that are below 150 meV/atom can be synthesized experimentally.

In Fig. 8, we show the distribution of elements A, B, C, and D for the potentially stable  $\text{ABC}_2\text{D}$  perovskites. In this case, we plotted all 8681 compounds that were calculated to have a distance to the convex hull smaller than 100 meV/atom. In what concerns the A atom, we find mainly alkali atoms and metals around indium. This is in agreement with the findings of (16) for ternary perovskites. The distribution for the B atom (in the center of the octahedra) is especially notable with 66% of the materials containing either hydrogen, or carbon, or nitrogen, or oxygen. This points to the fact that the majority of discovered materials are inverted perovskites. Last, for the C and D atoms (at the vertices of the octahedra), we find a large collection of metals and also a few halogens. Obviously, the former corresponds to inverted perovskites, and the latter corresponds to normal (noninverted) halide perovskites with two halogens alloyed in the vertices of the octahedra.

The overwhelming majority of the compounds are metallic, with only 8% of the systems exhibiting a bandgap above 0.1 eV. As expected, the largest bandgaps are observed for halide perovskites or for systems with a halogen occupying the C position and H in the D position. On the other hand, more than 15% of the tetragonal compounds exhibit a magnetic polarization, because of the abundance of magnetic 3d metals in the stable compositions.

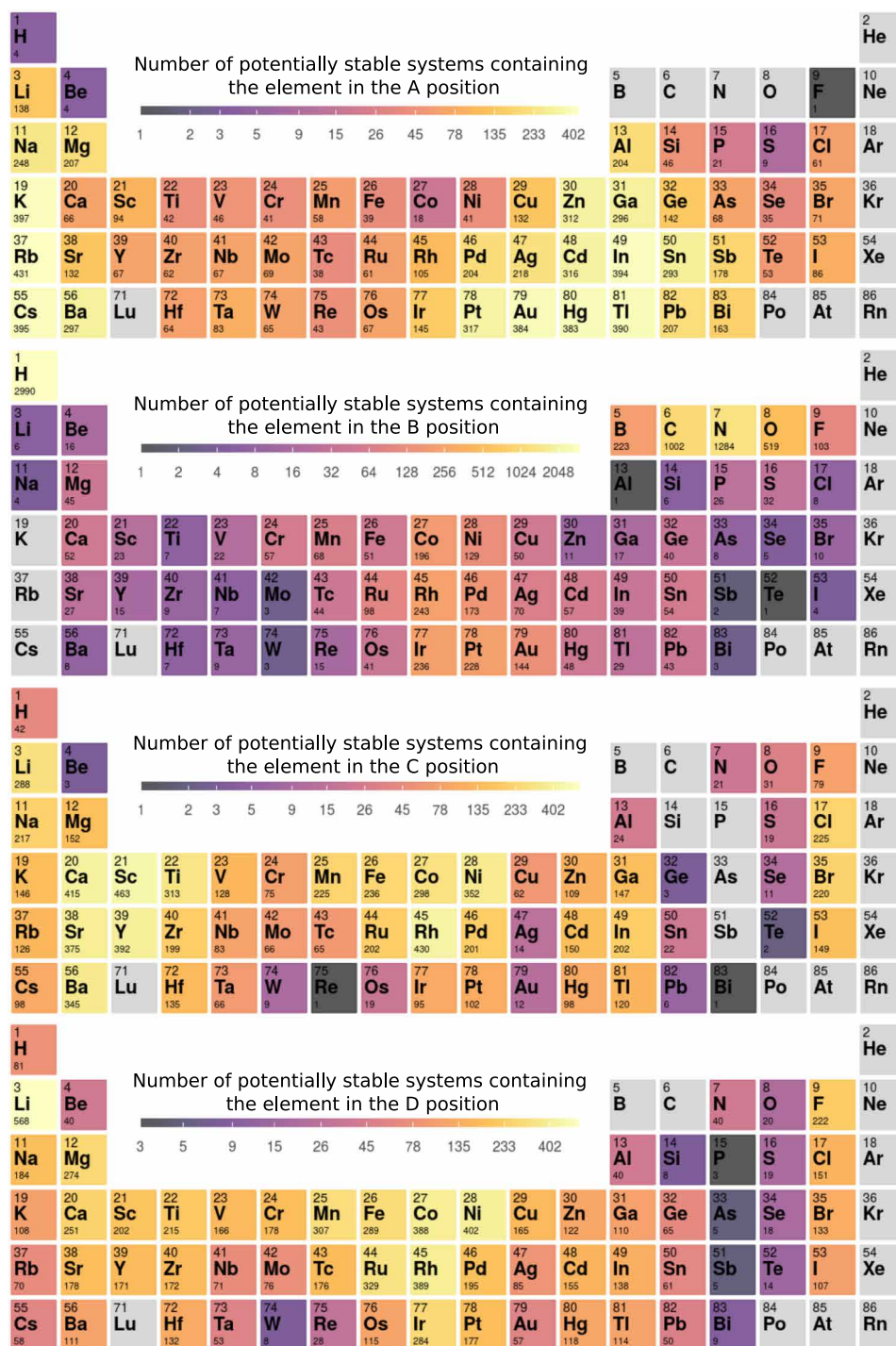
In the following, we look in more detail into some of these compounds. We note that we restrict our discussion to the tetragonal unit cell and that the properties of this crystal phase might be different from a cell including structural deformations (octahedra tilting and rotation, etc.). Moreover, our five-atom unit cell is not capable of describing different positions of the C and D atoms such as the cis arrangement, which is known to be favorable in some oxynitrides, oxyhalides, and oxysulphides (50, 51), or more complex orderings (45). We also do not analyze oxynitride, oxyfluoride, or nitrofluoride compounds as these were discussed in a previous work that performed an exhaustive study of these systems (45).

We start with the normal perovskites, where the C and D atoms belong to the same nonmetallic group. This case is rather important



**Fig. 7. Distances to the convex hull for the mixed perovskites predicted to be below 200 meV/atom.** We show the distances to the convex hull calculated with DFT for the mixed perovskites predicted to be below 200 meV/atom by the machine learning model and the percentage of systems below the energy on the x axis in red.

as alloying the nonmetallic sites allows for the tuning of electronic properties, such as the bandgap. We could not find any system below 100 meV/atom that alloys two pnictogens. This is not unexpected as it is already very difficult to form nitride perovskites because of the very high oxidation state of nitrogen (52, 53). When C and D are elements of the chalcogenide family, we find 12 systems, among which six oxysulfides of compositions  $\text{Ba}\{\text{Zr}, \text{Nb}, \text{Hf}, \text{Ta}\}\text{S}_2\text{O}$  and  $\text{Na}\{\text{Nb}, \text{Ta}\}\text{O}_2\text{S}$  and six S—Se alloys with compositions  $\text{Ba}\{\text{Ti}, \text{Zr}, \text{Hf}\}\text{S}_2\text{Se}$  and  $\text{Ba}\{\text{Ti}, \text{Zr}, \text{Hf}\}\text{Se}_2\text{S}$ . These are mostly nonmagnetic semiconductors with Perdew-Burke-Ernzerhof (PBE) bandgaps that go up to 1.26 eV for  $\text{BaHfS}_2\text{O}$ . We found many more compounds with C and D belonging to the halogen group. We remember that halide perovskites revolutionized research on solar energy with solar power conversion efficiencies that reached up to 22.1% in only 6 years. In modern solar cells, one often alloys I with Br as this was found to improve the stability of these compounds (54, 55). From the 580 compounds we found, 229 contained F, 354 contained Cl, 335 contained Br, and contained 286 I. Obviously, it is easier to alloy Br—Cl than, for example, F—I, and that is exactly what we find in our results with 149 systems for the former and 60 for the latter. The large majority of the compounds turn out to be nonmagnetic semiconductors, with bandgaps going up to 5.7 eV (for  $\text{CsCaCl}_2\text{F}$ ). We note that we find in this list the inorganic perovskites that are



**Fig. 8. Number of potentially stable Perovskite with a specific element in A, B, C, or D position.** Perovskites ABC<sub>2</sub>D that are predicted by DFT to have a distance to the convex hull smaller than 100 meV/atom and that have a certain chemical element in positions A, B, C, or D.

relevant for photovoltaics, such as CsPbCl<sub>2</sub>F and CsPbF<sub>2</sub>Cl for 11 different combinations of the nonmetals, showing the miscibility of the halogens in these compounds.

Another very interesting type of system are hydride perovskites (56). These have attracted some interest recently as possible materials for, e.g., hydrogen storage (56–58). We found a series of exotic hydrogen-containing mixed perovskites, where H is combined with a group 15, 16,

or 17 elements. Up to 100 meV/atom, we find 42 hydronitride systems, the most stable being La{Cr, Mo, Tc, W, Re}N<sub>2</sub>H and La{Mn, Fe, Mo, Tc}H<sub>2</sub>N. Assuming the standard oxidation states of the nonmetals and of La, we see that the metal in the B site should have an oxidation state of +4 and +2, respectively, for the N<sub>2</sub>H and H<sub>2</sub>N perovskites. This is true for many of the systems we found, but we also find less common oxidation states in the data. These systems are



metallic and can be magnetic when the B atom is a magnetic d metal. Looking now at group 16, we find 12 hydroxides (59) and 7 hydrosulfides with mostly Sr or Ba in the A position and a +2 or +3 metal in the B position. There are several semiconductors in this list, reaching a maximum PBE gap of 2.43 eV for BaYO<sub>2</sub>H. Last, we find 49 hydrohalides compounds with a +1 metal in the A position (either an alkali or Tl) and a +2 metal in the B position. There are also several hydride perovskites with similar compositions that were previously studied experimentally and computationally (60), suggesting the possibility of the exchange of the hydrogen by a halide ion. These are nonmagnetic and can have rather large bandgaps reaching a maximum of 5.39 eV (in the PBE approximation) for KMgF<sub>2</sub>H.

We now turn our attention to inverted or antiperovskites. These are materials that are finding diverse applications in battery technology, magnetism, superconductivity, etc. (61). As mentioned before, we find a very large number of such systems, which is not unexpected as the majority of stable ternary perovskites are also inverted (16), and it is relatively easy to alloy two C and D metals.

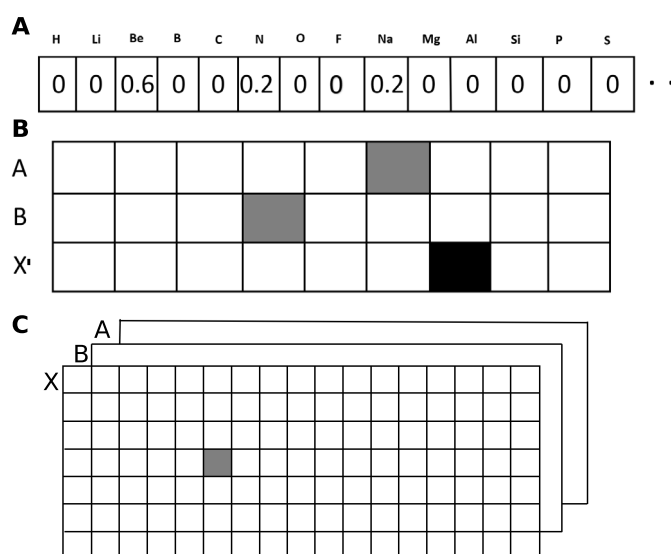
We start by looking at inverted hydrides, of which we find 3007 below 100 meV/atom of the convex hull of stability. The most common elements that we find in the A position in these compounds are noble metals such as Pt (201 systems), Au (163), Ir (125), etc., or an element from groups 13 or 14 such as Sn (148), Ga (143), In (125), Al (100), etc. Naturally, most compounds are metallic, but we find a handful of semiconductors mostly of the form {Se, Te}HA<sub>2</sub><sup>(1)</sup>A<sup>(2)</sup> and {P, As, Sb, Bi}HA<sub>2</sub><sup>(1)</sup>A<sup>(3)</sup>, where A<sup>(1)</sup> and A<sup>(2)</sup> are alkalines and A<sup>(3)</sup> is an alkali earth. We note that the former belong to the family of inverted ternary hydride perovskites with both A and B sites occupied by anions, which were proposed as fast alkali ionic conductors (62). We also found a couple of more exotic systems, such as TeH{Rh, Pd}<sub>2</sub>Li, TeHLi<sub>2</sub>Pd, and TeHPd<sub>2</sub>Sc.

Our list contains 223 inverted boride perovskites, with the A position predominantly filled by Y (31 systems), Sc (30), In (28), Mg (20), etc. On the other hand, the C and D positions contain often a metal like Ru, Rh, Pd, Ir, Pt, etc., or a magnetic metal like Co or Ni. Not unexpectedly, these are all metals.

Carbide antiperovskites have attracted some attention especially since the discovery of superconductivity in MgCNi<sub>3</sub> (63). There are 1002 (metallic) carbide antiperovskites in our list. Mg is the most common metal we find in the A position (with 82 systems), followed by In (76), Zn (74), Cs (72), etc. From the 82 systems with Mg, 12 of which also include Ni in the C or D position (alloyed with another 3d metal or with Li, Rh, and Ir). In the C position, the most common metals are Sc (228 system), Y (200), and La (200), while in the D position, we find mostly a 3d magnetic element or Ru, Ir, Tc, etc.

Inverted nitride perovskites are metal-rich compounds with particularly interesting electronic and magnetic properties (64). We find 1284 such systems with distances to the hull below 100 meV/atom. In the A position, we encounter mostly late transition and noble metals such as Ag, Au, Pt or Zn, Cd, Hg, etc., while in the corners of the octahedra, we find metals such as Ca, Se, La, Ba, etc. (in the C position) or Li, Co, Ni, Rh, etc. (in the D position). These tetragonal systems turn out to be mostly metallic, but with a handful of exceptions when the A position is occupied by another nonmetal such as a chalcogen or a halogen. The maximum PBE gap we find is for SNCa<sub>2</sub>Li with a direct bandgap of 1.04 eV, 39 meV/atom above the convex hull. We also find more than 300 magnetic systems with magnetic 3d elements in the C or D position.

With O in the B position, we find 519 systems. Among the most stable ones, we find series of Li-containing compounds such as {Al, Ga,



**Fig. 9. ElemNet representations with multiple dimensions.** (A) Representation of ElemNet (28). (B and C) ElemNet representation extended to two and three dimensions.

In}{Sc, Y}<sub>2</sub>Li or {Sn, Pb}{Sc, Y}<sub>2</sub>Li, but there is a large variety in the compositions. Recently, ternary oxysilicides and oxygermanides of the type {Si, Ge}O{Ca, Sr}<sub>3</sub> were proposed as candidates for non-toxic infrared semiconductors, as they exhibit sharp absorption edges below 1 eV (65). We find a number of quaternary mixed perovskites of this family, namely, 7 oxysilicides, 11 oxygermanides, and 21 oxystannites, with Sr, Ca, Ba, but also Y, La, etc., alloyed in the C and D positions. All these compounds turn out to be metallic in our calculations, but this is very likely due to the use of the small five-atom tetragonal cell and the PBE approximation. Similarly to the hydrides, we also find a number of semiconducting systems with both the A and B positions occupied by anions. Examples are {S, Se, Te}HA<sub>2</sub><sup>(1)</sup>A<sup>(3)</sup> and {Cl, Br, I}HA<sub>2</sub><sup>(1)</sup>A<sup>(2)</sup>. We find PBE bandgaps in a rather large range up to 2.88 eV (for SOLi<sub>2</sub>Ca).

## DISCUSSION

In this study, we developed a machine learning model that predicts the energy of a material as a function of the composition and the structure prototype. However, and in contrast with previous approaches, our input features do not require the precise knowledge of the geometry, so our model can be used to accelerate the discovery of new materials with high-throughput methods based on DFT. Our machine relies on crystal graph attention neural networks, and during the message passing steps, each individual edge and vertex embedding is updated based on its neighborhood.

To train this machine, we compiled and curated a large dataset of more than 2 million density functional calculations. These include data points from online databases and from our own calculations that were obtained with compatible parameters. Despite its large size, this dataset is somewhat biased as many of the calculations are for a relatively small number of different crystal prototypes.

To circumvent this problem, we propose a transfer learning approach, where our general purpose model is retrained for specific crystal structures. We experimented with this idea for a set of quaternary



perovskites, showing that transfer learning can speed up the training of the model by a factor of two. By omitting vanadium-containing compounds from the training, we also showed that the network can reliably extrapolate to unknown regions of the periodic table.

Last, we used our training model to predict stable quaternary perovskites, predictions that were then validated with DFT. It turns out that there are more than 20,000 materials that have a good chance of being synthesized experimentally. These are mostly inverted perovskites, with hydrogen, carbon, or nitrogen in the center of the octahedra, and with two metals alloyed in the vertices of the octahedra.

In view of the above, we believe that our model combined with transfer learning techniques will allow us to explore a large domain of the vast chemical space in the search for new stable crystalline compounds.

## MATERIALS AND METHODS

### Data accumulation and filtering

As previously discussed, we combined data from the materials project, our group, and AFLOW. Concerning AFLOW, it has to be noted that a number of calculations in the database are “ill-calculated” as noted internally in the code of the AFLOW-CHULL tool (66). Accordingly, all prototypes denoted as “\_DEVIL\_PROTOTYPES\_” and all pseudopotential/prototype combinations that are known to be ill-converged were removed from the AFLOW data. Furthermore, outliers from AFLOW were removed along the strategy explained in (66). Last, all actinides and nobles gases were removed from the data as they are not relevant for most applications, and some of the energies of the former are questionable. Duplicates were removed by checking for structures with the same composition, space group, and total energy (rounded to the fourth digit). The space group was determined with pymatgen, with the “symprec” parameter set to 0.1.

The convex hull was then constructed with pymatgen (67). All energies were corrected according to the materials project compatibility scheme. Furthermore, the distances to the hull were evaluated for each compound by removing the compound itself from the convex hull. Stable systems have then negative distances to the hull instead of being truncated at zero. This, in our opinion, should improve the learning as we are predicting a smoother quantity.

All elementary substances were removed from the dataset as the roost model used for the global pooling only works for multinary structures. A few hundred compounds that did not have 24 neighbors in the cutoff radius were also removed from the dataset.

### Calculation parameters

For all density functional calculations first published in this work, we optimized the geometry and calculated the energy with the code VASP (68). Calculation parameters were chosen to be compatible with the materials project database (35). We used the projector augmented wave (69) datasets of version 5.2 with a cutoff of 520 eV. The Brillouin zone was sampled by  $\Gamma$ -centered  $k$ -point grids with a uniform density calculated to yield 1000  $k$ -points per reciprocal atom. All forces were converged to better than 0.005 eV/Å. Calculations were performed with spin polarization using the Perdew-Burke-Ernzerhof (70) exchange-correlation functional, with the exception of oxides and fluorides containing Co, Cr, Fe, Mn, Mo, Ni, V, and W, where an on-site Coulomb repulsive interaction  $U$  with a value of 3.32, 3.7, 5.3, 3.9, 4.38, 6.2, 3.25, and 6.2 eV, respectively, was added to correct the d states.

### Implementation

The model was implemented in PyTorch (71) and PyTorch geometric (72) using PyTorch lightning (73) for convenience purposes. Code from (28) and (6) was reused, demonstrating the importance of sharing code together with a paper. The code developed here will be distributed on github (<https://github.com/hyllios/CGAT>).

### Crystal graph attention network hyperparameters

The size, number of layers of the FCNNs, the number of attention heads, the number of message passing steps, the size of the embeddings, the learning rate, batch size, optimizer, hyperparameters of the optimizer, the maximum number of neighbors, the cutoff radius for both the graph attention network, and the roost model used for the global composition representation are all hyperparameters that had to be optimized over a number of runs. Because of the high training cost, they were not optimized automatically but rather by hand increasing the number of parameters in terms of attention heads, message passing steps, and embedding sizes until a further increase ceased to efficiently improve the error. Afterward, the batch size and learning rate were optimized. AdamW (74) in combination with a cyclical learning rate scheduler was used for the training of the model. As loss function, we used the expanded MAE used in (28) that includes an estimate of the aleatoric uncertainty. Using hypernetworks instead of normal fully connected networks resulted only in a very small gain in validation error; however, for larger datasets, the validation error converged after less learning rate cycles, reducing the training time.

optimizer	AdamW
learning rate	0.000125
starting embedding	matscholar-embedding
nbr-embedding-size	512
msg-heads	6
batch size	512
max-nbr	24
epochs	390
loss	L1 loss
momentum	0.9
weight decay	1e-06
atom-fea-len	128
message passing steps	5
roost message passing steps	3
other roost parameters	default
vector attention	True
edges	updated
learning rate	cyclical
learning rate schedule	(0.1, 0.05)
learning rate period	130
hypernetwork	3 hidden layers; size, 128
hypernetwork activ. funct.	tahn
FCNN	1 hidden layer, size 512
FCNN activ. funct.	leaky RELU (76)

The hyperparameters used to train the crystal graph attention network were as follows.

The output ResNet consisted of seven hidden layers with sizes 1024, 1024, 512, 512, 256, 256, and 128 and rectified linear unit (76) activation functions (Relu).

### ElemNet

For comparison purposes, we use a model with a composition-based input, in this case a modified version of ElemNet. The original ElemNet was a standard FCNN with dropout and batch normalization layers. Naturally, such a network is not able to predict energies for mixed perovskites as it cannot distinguish between, e.g., BaTiPbO<sub>2</sub> and TiBaPbO<sub>2</sub>. To circumvent this problem, we tried two different representations along the ideas of (77, 78) and used multiple input channels for each crystallographic position, and we ordered the input in the form of a periodic table (Fig. 9). Naturally, this kind of network is fixed to one specific crystal prototype. The ElemNet network started with one 3D convolutional layer [1 input channel, 92 output channels, kernel size (1, 6, 3), stride 1, and padding 0]. The resulting tensor was flattened and input in a 17-layer fully connected network with Relu (76) activation functions and sizes of 5520, 1024, 1024, 1024, 1024, 512, 512, 512, 256, 256, 256, 128, 128, 128, 64, 64, and 32. The hyperparameters for the training of the ElemNet model were as follows.

optimizer	AdamW
learning rate	0.001
batch size	200
epochs	until no improvement for 50 epochs
loss	L2 loss
momentum	0.9
weight decay	1e-06
learning rate	stepwise reduction
step size	50 epochs, reduction by factor 0.5

### REFERENCES AND NOTES

1. K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
2. J. Schmidt, M. R. G. Marques, S. Botti, M. A. L. Marques, Recent advances and applications of machine learning in solid-state materials science. *Npj Comput. Mater.* **5**, 83 (2019).
3. K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, A. Tkatchenko, Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **8**, 13890 (2017).
4. J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, G. E. Dahl, Neural message passing for quantum chemistry, in *Proceedings of the 34th International Conference on Machine Learning-Volume 70* (PMLR, 2017), pp. 1263–1272.
5. C. Chen, W. Ye, Y. Zuo, C. Zheng, S. P. Ong, Graph networks as a universal machine learning framework for molecules and crystals. *Chem. Mater.* **31**, 3564–3572 (2019).
6. T. Xie, J. C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).
7. Y. Zhuo, A. M. Tehrani, J. Brgoch, Predicting the band gaps of inorganic solids by machine learning. *J. Phys. Chem. Lett.* **9**, 1668–1673 (2018).
8. P. Dey, J. Bible, S. Datta, S. Broderick, J. Jasinski, M. Sunkara, M. Menon, K. Rajan, Informatics-aided bandgap engineering for solar materials. *Comput. Mater. Sci.* **83**, 185–195 (2014).
9. J. Lee, A. Seko, K. Shitara, K. Nakayama, I. Tanaka, Prediction model of band gap for inorganic compounds by combination of density functional theory calculations and machine learning techniques. *Phys. Rev. B* **93**, 115104 (2016).
10. Z. Zhang, A. M. Tehrani, A. O. Oliynyk, B. Day, J. Brgoch, Finding the next superhard material through ensemble learning. *Adv. Mater.* **33**, 2005112 (2020).
11. C. J. Court, J. M. Cole, Magnetic and superconducting phase diagrams and transition temperatures predicted using text mining and machine learning. *Npj Comput. Mater.* **6**, 18 (2020).
12. S. Curtarolo, W. Setyawan, G. L. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M. J. Mehl, H. T. Stokes, D. O. Demchenko, D. Morgan, AFLOW: An automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **58**, 218–226 (2012).
13. G. Hautier, C. C. Fischer, A. Jain, T. Mueller, G. Ceder, Finding nature's missing ternary oxide compounds using machine learning and density functional theory. *Chem. Mater.* **22**, 3762–3767 (2010).
14. A. O. Oliynyk, E. Antono, T. D. Sparks, L. Ghadbeigi, M. W. Gaultois, B. Meredig, A. Mar, High-throughput machine-learning-driven synthesis of full-Heusler compounds. *Chem. Mater.* **28**, 7324–7331 (2016).
15. F. Legrain, J. Carrete, A. van Roekeghem, G. K. Madsen, N. Mingo, Materials screening for the discovery of new half-Heuslers: Machine learning versus ab initio methods. *J. Phys. Chem. B* **122**, 625–632 (2018).
16. J. Schmidt, J. Shi, P. Borlido, L. Chen, S. Botti, M. A. L. Marques, Predicting the thermodynamic stability of solids combining density functional theory and machine learning. *Chem. Mater.* **29**, 5090–5103 (2017).
17. J. Schmidt, L. Chen, S. Botti, M. A. L. Marques, Predicting the stability of ternary intermetallics with density functional theory and machine learning. *J. Chem. Phys.* **148**, 241728 (2018).
18. F. A. Faber, A. Lindmaa, O. A. von Lilienfeld, R. Armiento, Machine learning energies of 2 million elpasolite (ABC<sub>2</sub>D<sub>6</sub>) crystals. *Phys. Rev. Lett.* **117**, 135502 (2016).
19. F. Faber, A. Lindmaa, O. A. von Lilienfeld, R. Armiento, Crystal structure representations for machine learning models of formation energies. *Int. J. Quant. Chem.* **115**, 1094–1101 (2015).
20. W. Li, R. Jacobs, D. Morgan, Predicting the thermodynamic stability of perovskite oxides using machine learning models. *Comput. Mater. Sci.* **150**, 454–463 (2018).
21. J. Carrete, N. Mingo, S. Wang, S. Curtarolo, Nanograined half-Heusler semiconductors as advanced thermoelectrics: An ab initio high-throughput statistical study. *Adv. Funct. Mater.* **24**, 7427–7432 (2014).
22. G. Hautier, C. Fischer, V. Ehrlicher, A. Jain, G. Ceder, Data mined ionic substitutions for the discovery of new compounds. *Inorg. Chem.* **50**, 656–663 (2011).
23. L. Ward, R. Liu, A. Krishna, V. I. Hegde, A. Agrawal, A. Choudhary, C. Wolverton, Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations. *Phys. Rev. B* **96**, 024104 (2017).
24. K. Kim, L. Ward, J. He, A. Krishna, A. Agrawal, C. Wolverton, Machine-learning-accelerated high-throughput materials screening: Discovery of novel quaternary Heusler compounds. *Phys. Rev. Mater.* **2**, 123801 (2018).
25. G. R. Schleder, C. M. Acosta, A. Fazio, Exploring two-dimensional materials thermodynamic stability via machine learning. *ACS Appl. Mater. Interfaces* **12**, 20149–20157 (2020).
26. Z. Li, Q. Xu, Q. Sun, Z. Hou, W.-J. Yin, Thermodynamic stability landscape of halide double perovskites via high-throughput computing and machine learning. *Adv. Funct. Mater.* **29**, 1807280 (2019).
27. D. Jha, L. Ward, A. Paul, W.-k. Liao, A. Choudhary, C. Wolverton, A. Agrawal, Elemnet: Deep learning the chemistry of materials from only elemental composition. *Sci. Rep.* **8**, 17593 (2018).
28. R. E. Goodall, A. A. Lee, Predicting materials properties without crystal structure: Deep representation learning from stoichiometry; arXiv:1910.00617 (2019).
29. S.-Y. Louis, Y. Zhao, A. Nasiri, X. Wong, Y. Song, F. Liu, J. Hu, Global attention based graph convolutional neural networks for improved materials property prediction; arXiv:2003.13379 (2020).
30. C. W. Park, C. Wolverton, Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery. *Phys. Rev. Materials* **4**, 063801 (2020).
31. G. Cheng, X.-G. Gong, W.-J. Yin, Crystal structure prediction via combining graph network and bayesian optimization; arXiv:2011.10968 (2020).
32. J. Noh, G. H. Gu, S. Kim, Y. Jung, Uncertainty-quantified hybrid machine learning/density functional theory high throughput screening method for crystals. *J. Chem. Inf. Model.* **60**, 1996–2003 (2020).
33. T. Xie, A. France-Lanord, Y. Wang, Y. Shao-Horn, J. C. Grossman, Graph dynamical networks for unsupervised learning of atomic scale dynamics in materials. *Nat. Commun.* **10**, 2667 (2019).
34. E. Nachmani, L. Wolf, Molecule property prediction and classification with graph hypernetworks; arXiv:2002.00240 (2020).
35. A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. A. Persson, Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).

36. J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD). *JOM* **65**, 1501–1509 (2013).
37. C. J. Bartel, A. Trewartha, Q. Wang, A. Dunn, A. Jain, G. Ceder, A critical examination of compound stability predictions from machine-learned formation energies. *Npj Comput. Mater.* **6**, 97 (2020).
38. V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K. A. Persson, G. Ceder, A. Jain, Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **571**, 95–98 (2019).
39. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is All you Need, *Advances Neural Information Processing Systems 30 (NIPS 2017)*, pp. 5998–6008.
40. P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph attention networks. arXiv:1710.10903 (2017).
41. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27 to 30 June 2016.
42. H.-C. Wang, S. Botti, M. A. L. Marques, Predicting stable crystalline compounds using chemical similarity. *Npj Comput. Mater.* **7**, 12 (2021).
43. B. Meredig, E. Antono, C. Church, M. Hutchinson, J. Ling, S. Paradiso, B. Blaiszik, I. Foster, B. Gibbons, J. Hatrick-Simpers, A. Mehta, L. Ward, Can machine learning identify the next high-temperature superconductor? Examining extrapolation performance for materials discovery. *Mol. Syst. Des. Eng.* **3**, 819–825 (2018).
44. A. Dunn, Q. Wang, A. Ganose, D. Dopp, A. Jain, Benchmarking materials property prediction methods: The matbench test set and automatminer reference algorithm. arXiv:2005.00707 (2020).
45. H.-C. Wang, J. Schmidt, S. Botti, M. A. L. Marques, A High-throughput study of oxynitride, oxyfluoride and nitrofluoride perovskites. *J. Mater. Chem. A* **9**, 8501–8513 (2021).
46. M. Sakar, R. M. Prakash, K. Shinde, G. R. Balakrishna, Revisiting the materials and mechanism of metal oxynitrides for photocatalysis. *Int. J. Hydrogen Energy* **45**, 7691–7705 (2020).
47. G. Miessler, P. Fischer, D. Tarr, *Inorganic chemistry: Pearson new international edition* (2013).
48. W. Kutzelnigg, Chemical bonding in higher main group elements. *Angew. Chem.* **23**, 272–295 (1984).
49. S. Kiyohara, H. Oda, T. Miyata, T. Mizoguchi, Prediction of interface structures and energies via virtual screening. *Sci. Adv.* **2**, e1600746 (2016).
50. M. Yang, J. Oro-Solé, J. A. Rodgers, A. B. Jorge, A. Fuertes, J. P. Attfield, Anion order in perovskite oxynitrides. *Nat. Chem.* **3**, 47–52 (2010).
51. G. Pilania, A. Ghosh, S. T. Hartman, R. Mishra, C. R. Stanek, B. P. Uberuaga, Anion order in oxyfluoride perovskites: Origins and implications. *Npj Comput. Mater.* **6**, 71 (2020).
52. R. Sarmiento-Pérez, T. F. T. Cerqueira, S. Körbel, S. Botti, M. A. L. Marques, Prediction of stable nitride perovskites. *Chem. Mater.* **27**, 5957–5963 (2015).
53. J. A. Flores-Livas, R. Sarmiento-Pérez, S. Botti, S. Goedecker, M. A. L. Marques, Rare-earth magnetic nitride perovskites. *J. Phys. Mater.* **2**, 025003 (2019).
54. J. H. Noh, S. H. Im, J. H. Heo, T. N. Mandal, S. I. Seok, Chemical management for colorful, efficient, and stable inorganic–organic hybrid nanostructured solar cells. *Nano Lett.* **13**, 1764–1769 (2013).
55. Y.-Y. Zhang, S. Chen, P. Xu, H. Xiang, X.-G. Gong, A. Walsh, S.-H. Wei, Intrinsic instability of the hybrid halide perovskite semiconductor  $\text{CH}_3\text{NH}_3\text{PbI}_3$ . *Chin. Phys. Lett.* **35**, 036104 (2018).
56. K. Ikeda, T. Sato, S.-i. Orimo, Perovskite-type hydrides – synthesis, structures and properties. *Int. J. Mater. Res.* **99**, 471–479 (2008).
57. A. Reshak,  $\text{NaMgH}_3$  a perovskite-type hydride as advanced hydrogen storage systems: Electronic structure features. *Int. J. Hydrogen Energy* **40**, 16383–16390 (2015).
58. Y. Li, J. S. Chung, S. G. Kang, First-principles computational screening of perovskite hydrides for hydrogen release. *ACS Comb. Sci.* **21**, 736–742 (2019).
59. Y. Kobayashi, O. Hernandez, C. Tassel, H. Kageyama, New chemistry of transition metal oxyhydrides. *Sci. Technol. Adv. Mater.* **18**, 905–918 (2017).
60. H. H. Raza, G. Murtaza, Umm-e-Hani, N. Muhammad, S. M. Ramay, First-principle investigation of  $\text{XSrH}_3$  (X = K and Rb) perovskite-type hydrides for hydrogen storage. *Int. J. Quantum Chem.* **120**, e26419 (2020).
61. Y. Wang, H. Zhang, J. Zhu, X. Lü, S. Li, R. Zou, Y. Zhao, Antiperovskites with exceptional functionalities. *Adv. Mater.* **32**, 1905007 (2019).
62. S. Gao, T. Broux, S. Fujii, C. Tassel, K. Yamamoto, Y. Xiao, I. Oikawa, H. Takamura, H. Ubukata, Y. Watanabe, K. Fujii, M. Yashima, A. Kuwabara, Y. Uchimoto, H. Kageyama, Hydride-based antiperovskites with soft anionic sublattices as fast alkali ionic conductors. *Nat. Commun.* **12**, 201 (2021).
63. T. He, Q. Huang, A. P. Ramirez, Y. Wang, K. A. Regan, N. Rogado, M. A. Hayward, M. K. Haas, J. S. Slusky, K. Inumara, H. W. Zandbergen, N. P. Ong, R. J. Cava, Superconductivity in the non-oxide perovskite  $\text{MgCNi}_3$ . *Nature* **411**, 54–56 (2001).
64. R. Niewa, Metal-rich ternary perovskite nitrides. *Eur. J. Inorg. Chem.* **2019**, 3647–3660 (2019).
65. N. Ohashi, D. Mora-Fonz, S. Otani, T. Ohgaki, M. Miyakawa, A. Shluger, Inverse perovskite oxysilicides and oxygermanides as candidates for nontoxic infrared semiconductor and their chemical bonding nature. *Inorg. Chem.* **59**, 18305–18313 (2020).
66. C. Oses, E. Gossett, D. V. Hicks, F. Rose, M. J. Mehl, E. Perim, I. Takeuchi, S. Sanvito, M. Scheffler, Y. Lederer, O. Levy, C. Toher, S. Curtarolo, AFLOW-CHULL: Cloud-oriented platform for autonomous phase stability analysis. *J. Chem. Inf. Model.* **58**, 2477–2490 (2018).
67. S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, G. Ceder, Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).
68. G. Kresse, J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
69. P. E. Blöchl, Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
70. J. P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
71. A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, A. Lerer, *NIPS 2017 Autodiff Workshop: The Future of Gradient-based Machine Learning Software and Techniques*, Long Beach, CA, US, December 9 2017.
72. M. Fey, J. E. Lenssen, Fast graph representation learning with PyTorch Geometric, in *ICLR Workshop on Representation Learning on Graphs and Manifolds* (2019).
73. W. Falcon, Pytorch lightning. *GitHub. Note*: <https://github.com/PyTorchLightning/pytorch-lightning> **3** (2019).
74. I. Loshchilov, F. Hutter, Decoupled weight decay regularization. arXiv:1711.05101 (2017).
75. S. S. Liew, M. Khalil-Hani, R. Bakhteri, Bounded activation functions for enhanced training stability of deep neural networks on visual pattern recognition problems. *Neurocomputing* **216**, 718–734 (2016).
76. X. Glorot, A. Bordes, Y. Bengio, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, G. Gordon, D. Dunson, M. Dudík, Eds. (JMLR Workshop and Conference Proceedings, Fort Lauderdale, FL, USA, 2011), vol. 15 of *Proceedings of Machine Learning Research*, pp. 315–323.
77. X. Zheng, P. Zheng, R.-Z. Zhang, Machine learning material properties from the periodic table using convolutional neural networks. *Chem. Sci.* **9**, 8426–8432 (2018).
78. X. Zheng, P. Zheng, L. Zheng, Y. Zhang, R.-Z. Zhang, Multi-channel convolutional neural networks for materials properties prediction. *Comput. Mater. Sci.* **173**, 109436 (2020).

#### Acknowledgments

**Funding:** S.B. acknowledges the support of the Volkswagen Foundation through the program Momentum–Funding for Recently Tenured Professors (Project Dandelion) and of the Michael Stifel Center Jena (MSCJ) in the framework of the project “Virtuellen Werkstatt für die Digitalisierung.” C.V. acknowledges the support from the Swedish Research Council and from Crafoordska stiftelsen. We acknowledge support by the Open Access Publication Fund of the Thuringer Universitäts- und Landesbibliothek Jena and the SFB 1357 NOA. **Author contributions:** All authors contributed to designing the research, interpreted the results, and wrote the manuscript. M.A.L.M. performed the DFT calculations. L.P. trained and developed the 3D ElemNet model. J.S. developed, trained, and applied the crystal graph attention networks. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data used in this work are freely available through the Materials Cloud (<https://archive.materialscloud.org/record/2021.128>) and our website (<https://tddf.org/bmg/data.php>). All data needed to evaluate the conclusions in the paper are present in the paper and data repository.

Submitted 31 March 2021  
Accepted 14 October 2021  
Published 3 December 2021  
10.1126/sciadv.abi7948

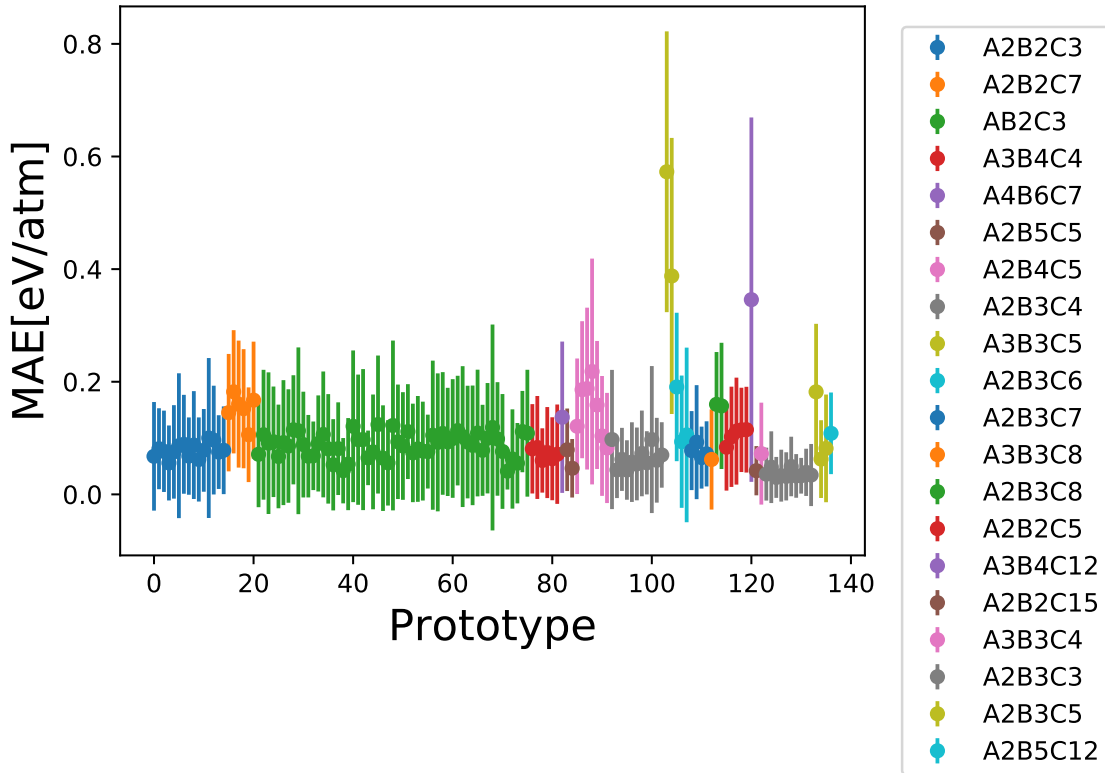


Figure 8: The figure shows the MAE, and the standard deviation of the absolute error as error bar, plotted for a number of ternary prototypes from the new high-throughput search. The prototypes are color-coded according to the matching composition.

### 4.3 A Dataset of 175k Stable and Metastable Materials Calculated with the PBEsol and SCAN Functionals

During the last decade, a number of large databases of theoretical crystal structure properties have been accumulated. As we have discussed in section 2.3, the accuracy of this data largely depends on the functional that was used in its creation.

While there are a number of small databases [58, 47, 23, 24, 22] that go beyond the PBE functional, all large scale databases [27, 133, 77, 67] are calculated with the PBE functional. For most calculated properties,

it is rather simple to improve their accuracy by repeating the calculation with a functional that has proven to have better accuracy with respect to experiment. In the case of thermodynamic stability, the situation is different as we need the energies of all potential decomposition channels calculated with compatible parameters to arrive at a distance to the convex hull. This has so far stopped any functional from establishing itself besides the PBE functional.

In this publication "**A Dataset of 175k Stable and Metastable Materials Calculated with the PBEsol and SCAN Functionals**", we aim to rectify the situation by creating a convex hull with both the SCAN as well as the PBEsol functional. The former has demonstrated a superior accuracy for the determination of thermodynamic stability.

In Ref. [JSPHD10] we have accumulated a rather complete convex hull by combining the AFLOW and Materials Project databases as well as more than a million calculations from our own group. We select the majority of compounds close to and on the convex hull from this dataset to recalculate them with the SCAN functional. However, SCAN calculations can be more than five times as expensive as PBE calculations and sometimes difficult to converge [38, 8]. We can mitigate this problem by minimizing the number of SCAN calculations. The PBEsol functional provides accurate crystal structures on the level of the SCAN functional [26, 65]. Therefore we opt to only perform PBEsol geometry optimizations and single point SCAN calculations at the PBEsol geometries.

We converge nearly all of the calculations successfully, and 176k are already openly available at [136]. Moreover, we have finished an additional 54k calculations of the compounds close to the hull that will be added to the online dataset in the near future. Analysing the dataset, known

trends [57, JSPHD1] concerning the difference in results between the PBE, PBEsol and SCAN functional are again confirmed. This includes the underbinding of the PBE functional that leads to an overestimation of lattice constants in comparison with the PBEsol functional, SCAN functional, and experiment, as well as the strong underestimation of band gaps both by the PBEsol and the PBE functional.

The two primary purposes behind the creation of this dataset are to allow for the stability analysis of materials and the training of machine learning models with the accuracy of the SCAN functional. The first point is relatively simple and just requires an extra PBEsol geometry optimization and single point SCAN calculation to confirm materials found as stable in high-throughput searches. Concerning the second point, we are in the process of applying transfer learning techniques to the PBE-trained models. These should allow us to negate the lack of SCAN/PBEsol training data and create high performing models to predict SCAN distances to the convex hull and PBEsol geometries. For this purpose, we have also performed an extra 50k PBEsol/SCAN calculations of unstable materials to reduce the bias toward stable materials in our training set.





OPEN

DATA DESCRIPTOR

# A dataset of 175k stable and metastable materials calculated with the PBEsol and SCAN functionals

Jonathan Schmidt<sup>1</sup>, Hai-Chen Wang<sup>1</sup>, Tiago F. T. Cerqueira<sup>2</sup>, Silvana Botti<sup>3</sup> & Miguel A. L. Marques<sup>1</sup>✉

In the past decade we have witnessed the appearance of large databases of calculated material properties. These are most often obtained with the Perdew-Burke-Ernzerhof (PBE) functional of density-functional theory, a well established and reliable technique that is by now the standard in materials science. However, there have been recent theoretical developments that allow for increased accuracy in the calculations. Here, we present a dataset of calculations for 175k crystalline materials obtained with two functionals: geometry optimizations are performed with PBE for solids (PBEsol) that yields consistently better geometries than the PBE functional, and energies are obtained from PBEsol and from SCAN single-point calculations at the PBEsol geometry. Our results provide an accurate overview of the landscape of stable (and nearly stable) materials, and as such can be used for reliable predictions of novel compounds. They can also be used for training machine learning models, or even for the comparison and benchmark of PBE, PBEsol, and SCAN.

## Background & Summary

The search for new materials remains one of the most important quests but, unfortunately, also one of the great challenges of materials science. Nowadays, data-driven searching strategies have become the most cost-effective methods to tackle this problem, and the fastest way of finding new materials or study their properties are computational high-throughput searches. After years of data accumulation, there are millions of calculations of materials available in open databases<sup>1,2</sup> that are used as an invaluable reservoir to select and filter promising candidates for further experimental synthesis and characterization.

These high-throughput studies in solid-state material science<sup>3–6</sup> have broadened the exploration of the vast chemical space, while plenty of works have successfully found and predicted promising materials for technological applications. However, nearly all high throughput searches rely on the use of density functional theory (DFT) within the Perdew-Burke-Ernzerhof (PBE) approximation to the exchange-correlation functional<sup>7</sup>. This is a well established and reliable approach that earned its place as the standard technique in solid-state research. However, the PBE functional is now over 25 years old, and more recent (and accurate) functionals have by now been proposed in the literature. For example, the Armiento-Mattson 2005<sup>8</sup> or the PBE for solids<sup>9</sup> functionals consistently lead to superior geometries<sup>10,11</sup>, while the SCAN meta-generalized gradient approximation<sup>12</sup> yields formation energies that are on average better by a factor of two than the PBE<sup>13</sup>. Unfortunately, and in stark contrast with the abundance of PBE data, there are no available comparable large scale datasets calculated with these improved functionals.

There are a number of other databases that use either higher accuracy methods, like  $G_0 W_0$ , or apply density-functionals different from PBE. For example, we can mention the Computational 2D Materials Database<sup>14,15</sup> that provides a dataset of 4000 2D materials calculated with HSE,  $G_0 W_0$ , RPA and the

<sup>1</sup>Institut für Physik, Martin-Luther-Universität Halle-Wittenberg, 06120, Halle (Saale), Germany. <sup>2</sup>CFisUC, Department of Physics, University of Coimbra, Rua Larga, 3004-516, Coimbra, Portugal. <sup>3</sup>Institut für Festkörpertheorie und -optik and European Theoretical Spectroscopy Facility, Friedrich-Schiller-Universität Jena, D-07743, Jena, Germany. ✉e-mail: [miguel.marques@physik.uni-halle.de](mailto:miguel.marques@physik.uni-halle.de)

Bethe-Salpeter equation, while JARVIS contains a large number of calculations with vdW corrections using OptB88vdW<sup>16,17</sup> and performed with the modified Becke-Johnson potential<sup>18–20</sup>.

In a previous work<sup>21</sup> we combined data from the AFLOW database<sup>1</sup>, the Materials Project<sup>2</sup> and from our own group to create a rather complete convex hull of thermodynamic stability at the PBE level. The details of the selection of the dataset can be found in ref. 21. Specifically, we selected all materials that were calculated with the same functional, pseudopotential, as well as *U*-parameters used in the Materials Project. We removed duplicates, i.e. entries with the same space group, composition and total energy (rounded to the 4th digit). Here we determined the space group using pymatgen with the “symprec” keyword set to 0.1. From the AFLOW database we further removed all prototypes labeled “\_DEVIL\_PROTOTYPES\_” and all other combinations of prototypes and pseudopotentials that are noted as ill-converged in the code of ref. 22. As AFLOW is still known to contain outliers<sup>22</sup>, we also removed them from the calculation of the hull following a strategy similar to the one explained in ref. 22. We used the total energies of all the remaining structures to calculate the convex hulls applying the corrections from the materials project workflow to the energies.

From this dataset<sup>23</sup> we selected around 175k compounds that were either stable (i.e., on the convex hull), or close to stable (within 100 meV/atom of the hull). These were then reoptimized with PBEsol<sup>9</sup>. Finally, the total energies were reevaluated with SCAN<sup>12</sup> to create highly accurate formation energies and convex hulls.

## Methods

Our starting point was the dataset used in the machine learning study of ref. 21. This included PBE calculations stemming from the Materials Project database<sup>2</sup>, AFLOW<sup>1</sup>, and our own calculations. These were then filtered to obtain a homogeneous set in what regards the calculation parameters, leading to a dataset containing more than two million compounds. We then constructed the convex hull of thermodynamic stability and extracted entries that were either on the hull or within 0.1 eV/atom<sup>24,25</sup>. The reason for the choice of this cutoff was twofold. First, its value is still below the average error in the formation energies calculated with PBE<sup>26,27</sup>, but is larger than the estimated error in the distances to the convex hull<sup>28,29</sup>. As such, the compounds that were misidentified by the PBE as thermodynamically unstable are likely to be included in the set. Second, there are a number of materials that are metastable, but experimentally accessible (for example for compositions that have more than one polymorph). We can reasonably expect that the cutoff allows for the inclusion of the majority of these cases. We also eliminated materials with unit cells that were too large for our computational resources, leading to a final amount of ~175k compounds.

All calculations were performed using density-functional theory, within the projector augmented wave method (PAW)<sup>30</sup> as implemented in the Vienna *ab initio* simulation package (VASP)<sup>31,32</sup>. We used the PAW setups shipped with version 5.4 of vasp that include information on the kinetic energy density of the core electrons. We mostly followed the recommendations of the Materials Project for the choice of the pseudopotentials. The exception was Cs, for which we used an improved pseudopotential generated by the vasp developers, as the stock PAW setup often led to negative densities during the self-consistent cycle, crashing the calculation. All calculations were performed taking into account spin-polarization, and started from a ferromagnetic configuration (as in the large majority of high-throughput studies). This most likely leads to an incorrect spin configuration for antiferromagnetic systems, resulting in an energy higher than the true ground-state. However, it is well known that in most cases magnetic exchange energies are rather small, so the error in the total energy is limited to a few tens of meV/atom. Methfessel-Paxton order one smearing with a width of 0.2 eV was applied in the integration of the Brillouin zone.

The structures were optimized using the PBEsol<sup>9</sup> approximation, the “High” precision keyword of vasp, and a  $\Gamma$ -centered k-point grid with 2000 k-points per reciprocal atom, until the forces on the atoms were below 5 meV/Å. To detect the structures that required re-optimization we checked that the absolute value of the forces (in meV/Å) and the individual components of the stress tensor (in meV/Å<sup>3</sup>), calculated with PBEsol with stricter convergence criteria (520 eV for the wave-function cutoff and 8000 k-points per reciprocal atom), remained smaller than 0.05. If it was not the case, we increased both the energy cutoff and the number of k-points and performed again the geometry optimization. Around 3% of all structures required this further calculation step. We note that as we had already a reasonable starting point, specifically the PBE geometry, the geometry optimization required a relatively small number of steps in most cases.

A final energy evaluation with the SCAN meta-GGA functional was performed with a cutoff of 520 eV, 8000 k-points per reciprocal atom, and including the non-spherical contributions from the gradient corrections inside the PAW spheres. As expected, SCAN calculations were much more unstable than PBEsol, due to the well known numerical instabilities of this functional<sup>33,34</sup>, which led to a much lower average convergence rate. In any case, we succeeded in converging nearly all calculations. In total, the calculations presented here required around 10 million CPU hours.

## Data Records

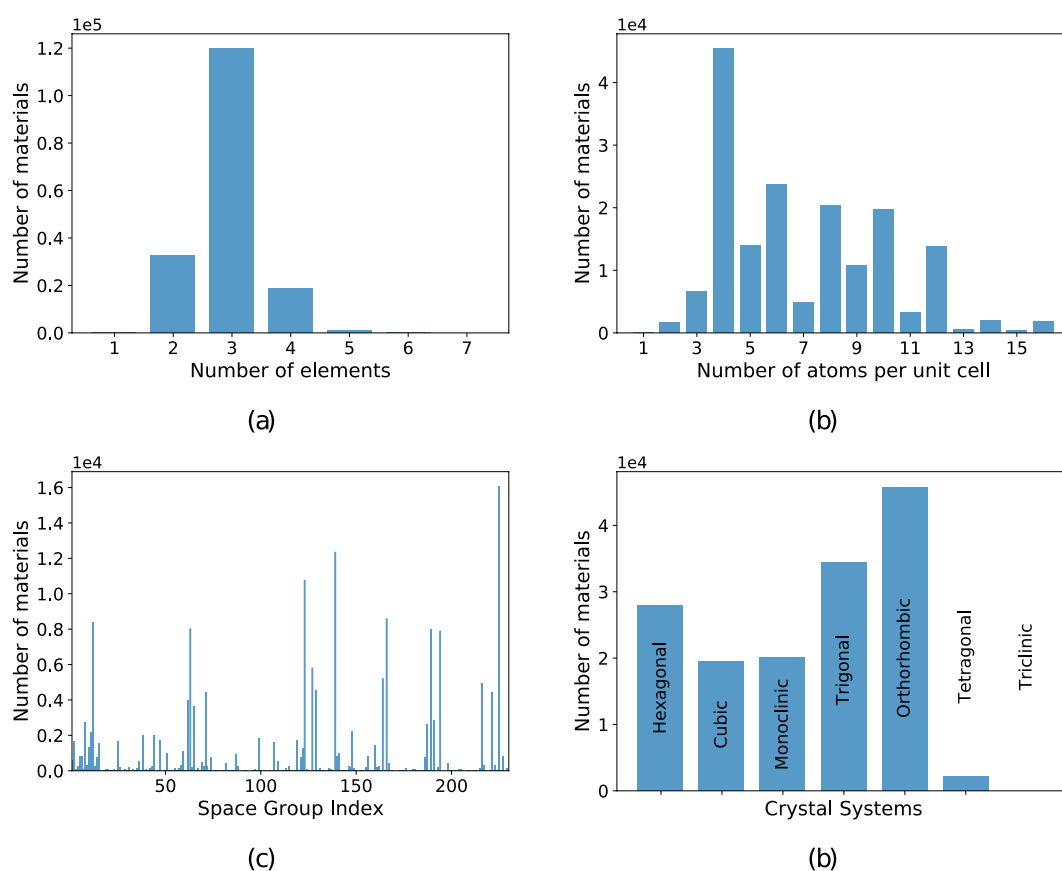
The output files of vasp were collected and processed with the pymatgen library<sup>35</sup>. Each final data record consisted of a *ComputedStructureEntry* that included the chemical composition, the total energy, and the detailed crystal structure of the entry. Two files, containing all entries for each functional, can be freely downloaded from the Materials Cloud repository<sup>36</sup> and can be loaded trivially using the json module of python. For convenience, we also provide a summary of the data in tabulated form, that include the fields listed in Table 1.

In panel a of Fig. 1 we plot the histogram of the number of different chemical elements in our materials. We can clearly see that the dataset is dominated by ternary compounds, followed by binary and quaternary. Relatively few multinary materials with more than 4 different chemical elements are present. We can understand this distribution by considering that the number of permutations increases very rapidly with the number of chemical elements, explaining, e.g., why we have many more binaries than elementary substances or ternary



Composition	Chemical composition of each material.
Number of sites	Number of atoms in the unit cell.
$E_{\text{PBE}}$	Total energy per atom calculated with PBE (extracted from the primary dataset)
$E_{\text{PBEsol}}$	Total energy per atom calculated with PBEsol.
$E_{\text{SCAN}}$	Total energy per atom calculated with SCAN.
$V_{\text{PBE}}$	Volume per atom of the PBE unit cell (extracted from the primary dataset).
$V_{\text{PBEsol}}$	Volume per atom of the PBEsol unit cell.
$\text{Gap}_{\text{PBEsol}}$	Band gap calculated with PBEsol.
$\text{Gap}_{\text{SCAN}}$	Band gap calculated with SCAN.
$M_{\text{PBEsol}}$	Total magnetic moment per atom calculated with PBEsol.
$M_{\text{SCAN}}$	Total magnetic moment per atom calculated with SCAN.
$S_{\text{SCAN}}$	The diagonal elements of the stress tensor calculated with SCAN.

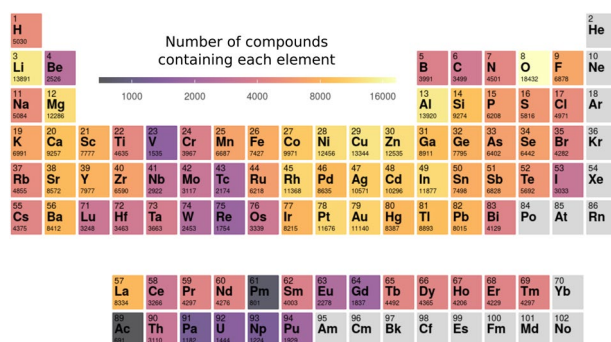
**Table 1.** Fields included in the summary of the data in tabulated form.



**Fig. 1** Distribution of (a) number of different chemical elements per unit cell, (b) number of atoms per unit cell, (c) index of space groups, and (d) crystal systems for all the materials in our dataset.

than binary compounds. However, we have to keep in mind that the complexity of the unit cells also increases and that most of the systematic high-throughput DFT studies were performed for ternary systems<sup>37–42</sup>. This can also be seen in panel (b) of Fig. 1 where we show a histogram of the number of atoms in the primitive unit cell. The distribution is dominated by a peak centered around five atoms per unit cell, arising from the materials stemming from the high-throughput studies of AFLOW<sup>1</sup> and from our group<sup>3,37,38</sup>. Obviously, we do not expect that the true distribution of all possible thermodynamically stable and metastable materials follows the behavior depicted in Fig. 1. Finally, in panels (c) and (d) of the same figure, we plot a histogram of the number of materials as a function of the space group index and of the crystal system. We see that the most represented systems are the trigonal and orthorhombic, while the tetragonal and triclinic are the least represented.

In Fig. 2 we show the distribution of chemical elements for the calculated structures. We considered all elements except noble gases up to bismuth as well as most lanthanides, and actinides up to plutonium. We can observe some obvious trends. Not surprisingly, the most common element is oxygen due to the abundance of



**Fig. 2** Periodic table depicting the chemical elements present in our dataset. The number beneath the chemical symbol is the number of materials present in the dataset that contain the given element.

oxides in our planet and their stability in our oxygen-rich atmosphere. The remaining chalcogens (S, Se, Te) are all equally represented, which is probably an indication of their chemical similarity. For the halogen family we see a decreasing number of materials following the decrease of the electronegativity, with iodine compounds being half as abundant as fluorides. For the pnictogens we witness exactly the opposite behavior, with much fewer nitrides than compounds containing antimony. This can be understood from the high chemical stability of the  $N_2$  molecule and the rather high nominal oxidation state of nitrogen ( $-3$ ) that hinders the synthesis of nitrides. Finally carbon-containing materials are rather scarce due to the absence of organic compounds in our dataset.

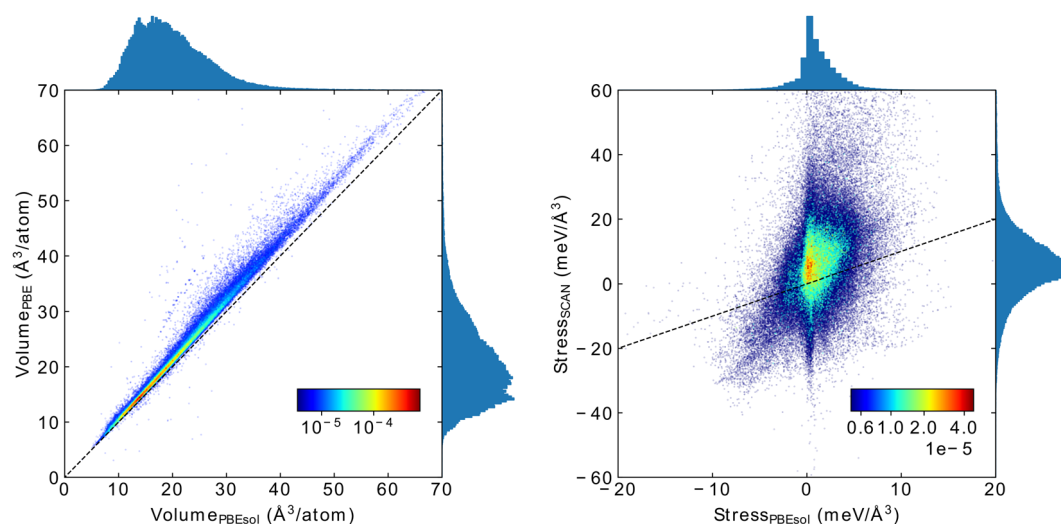
From the metals, the two with the highest number of compounds are aluminum and lithium. The former is within a cluster of highly represented chemical elements (such as nickel, copper, zinc, or indium). The exception in this region of the periodic table is gallium, in spite of its importance in many semiconductors used in electronics and optoelectronics. Lithium compounds, on the other hand, are much more common in our dataset than materials containing any other alkali element, which might be explained by the popularity of studies in lithium compounds in view of their application in battery technologies. Interestingly, in contrast to the other alkali earth elements, beryllium appears in relatively few compounds, probably due to its high toxicity. Another region that exhibits relatively few compounds is the one centered in vanadium and that includes rhenium, hafnium, niobium, etc. We can observe a continuity in the values in this region that might indicate that these chemical elements, often exhibiting the very high oxidation numbers of  $+4$ ,  $+5$  or  $+6$ , have more difficulty producing stable compounds than other metals with lower oxidation numbers. Finally, the least represented groups are unsurprisingly the lanthanides and the actinides, showing how little we know the chemistry of these chemical elements that are essential in a multitude of technologies, such as in hard magnets or in the storage of nuclear waste.

It is also interesting to look at the structural diversity in our dataset. With that objective, we used pymatgen to divide our structures into groups according to structural similarity. In total, our data turned out to contain 24706 prototypes of 4557 different generic compositions. However, and as expected, the distribution of compounds among these prototypes was rather unbalanced. For example, 15399 prototypes only appeared once in the dataset and 2412 only twice. On the other hand, most materials belong to just a few prototypes. The most common was by far the Heusler family of compounds, with almost 12 000 compositions, followed by the double perovskite family with more than 5000 elements. Of course, these numbers reflect not only the chemical stability and size of each one of the families, but also the interest of the community for these compounds.

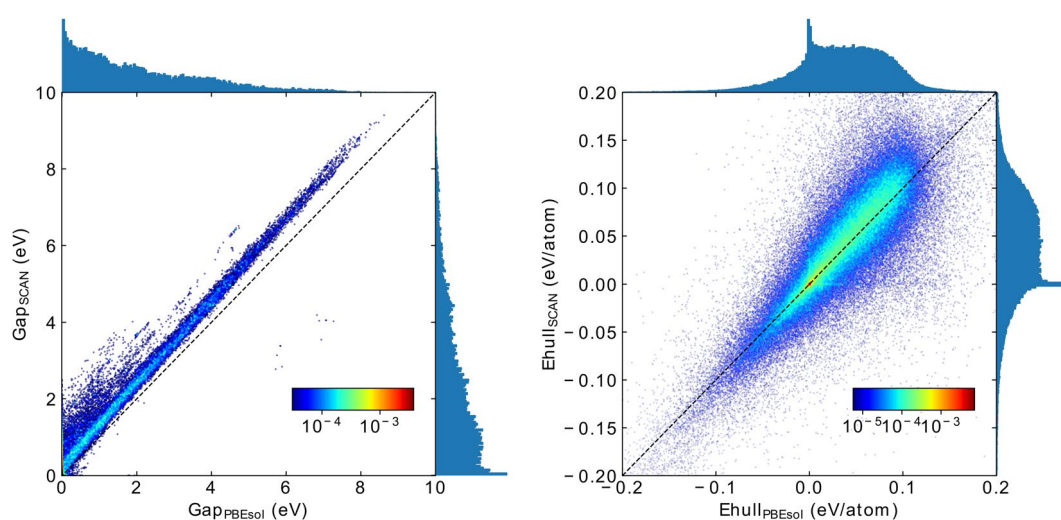
### Technical Validation

In the left panel of Fig. 3 we plot the distribution of the volumes per atom with the PBE (obtained from the primary data) and PBEsol. Both curves look similar, with a peak at around  $15 \text{ \AA}^3$  and skewed towards larger volumes. We can also observe that the PBE data is shifted toward larger volumes with respect to PBEsol. This is due to the well known underbinding of the PBE<sup>43</sup> that leads to lattice constants that are larger than their experimental values by 2–3% (and consequently volumes that are overestimated by  $\sim 10\%$ ). This underbinding is almost totally corrected by PBEsol, yielding smaller volumes in much better agreement with experiment. In the right panel of the same figure we plot a histogram of the three diagonal components of the stress tensor calculated with SCAN at the PBEsol geometry. We can see that the calculations yield rather small stresses, showing that the structures are close to mechanical equilibrium. This is expected because SCAN, like PBEsol, yields high quality geometries in good agreement with experiment<sup>10,44</sup>. This also validates our pragmatic approach of using a single point SCAN calculation at the PBEsol geometries.

In Fig. 4 we show the distribution of (indirect) band gaps calculated with both PBEsol and SCAN. As expected, the curves decay monotonically with the value of the gap, and exhibit a fat tail that extends beyond 10 eV. It is known that PBEsol, as well as PBE, strongly underestimate the value of the band gaps by essentially a factor of two, leading to mean absolute percentage errors bordering the 50%<sup>45,46</sup> with respect to experiment. This is to some extent corrected by SCAN, that increases consistently the band gaps leading to a mean absolute percentage error of around 40%<sup>45,46</sup>. This increase is evident from Fig. 4, where the SCAN band gap distribution is shifted to the right with respect to PBEsol.



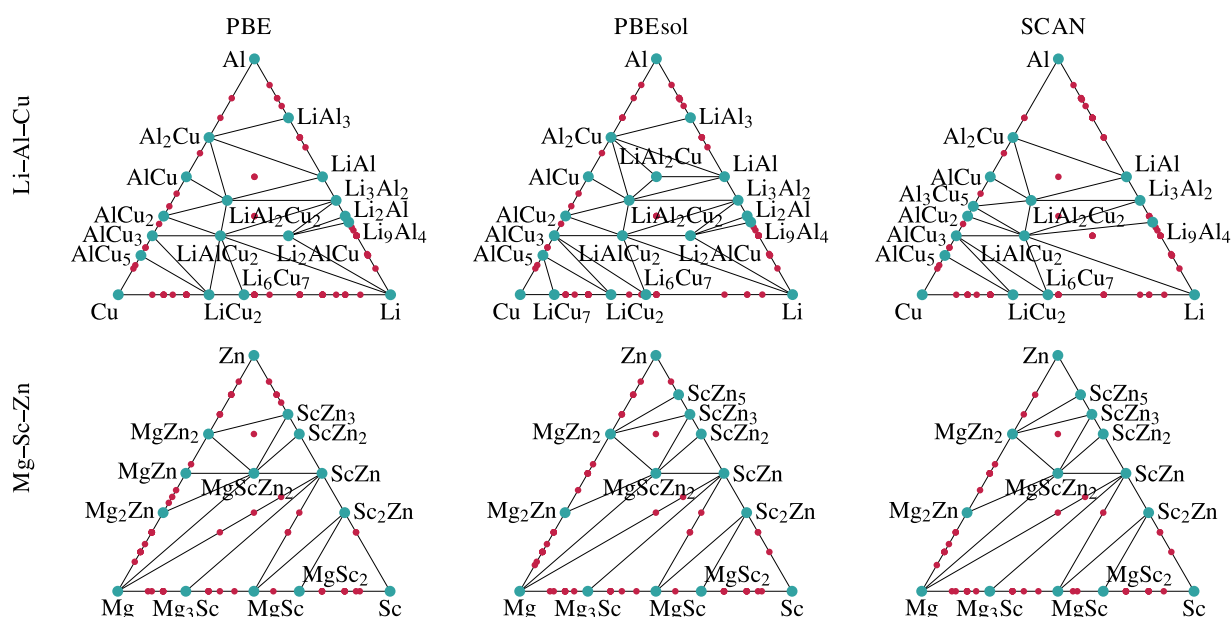
**Fig. 3** Left: Distribution and scatter plots of volumes per atom calculated with PBE (from the primary data) and PBEsol functionals. The width of the bins is  $0.35 \text{ \AA}^3/\text{atom}$ . Right: distribution and scatter plots of the diagonal components of the stress tensor calculated with PBEsol and SCAN at PBEsol geometries. The width of the bins is  $0.6 \text{ meV/\AA}^3$ .



**Fig. 4** Left: Distribution and scatter plots of the (in)direct band gaps calculated with PBEsol and SCAN at the PBEsol geometry. The width of the bins is  $0.1 \text{ eV}$ . Right: Distribution and scatter plots of the distances to the convex hull calculated with PBEsol and SCAN at PBEsol geometries. The corresponding hulls contain 40246 and 38692 materials. The width of the bins is  $2 \text{ meV/atom}$ .

Finally, we performed an analysis of the convex hulls obtained with PBE, PBEsol, and SCAN. The hulls contained 36985, 40246 and 38692 materials respectively. These differences are expected, and mostly stem from relatively small changes in the formation energy for compounds that were close to the hull. The distributions of the distances to the convex hull of thermodynamic stability are plotted in the right panel of Fig. 4. For this plot, we always removed the compound under study from the hull, allowing therefore for negative distances. Although this allows for a better interpretation of the results, we should remember that technically speaking all compounds with negative distances should be placed strictly at zero. Both distributions grow fast for negative distances to the hull, having a marked peak at zero. The main contributors to this behavior are the experimental compounds. The number of materials with positive distance to the hull is relatively constant, until reaching our cutoff at  $\pm 0.2 \text{ eV/atom}$ . This cutoff is obviously smeared for the PBEsol and SCAN calculations.

As an illustration, we plot in Fig. 5 the ternary phase diagrams of Li–Al–Cu and Mg–Sc–Zn. We see that the three functionals agree to a large extent on which are the stable materials. However, some differences are also clear. For example, MgZn is stable with the PBE functional but not with PBEsol or SCAN, or LiAl<sub>3</sub> is unstable with SCAN but stable with the other two functionals. In any case, compounds that are stable with one functional do appear stable, or at least metastable with the other functionals. Of course, SCAN is the most accurate of the



**Fig. 5** Ternary phase diagrams of the Li–Al–Cu (upper panel) and Mg–Sc–Zn (lower panel) systems, calculated with the PBE (left), PBEsol (middle) and SCAN (right). The blue points indicate compositions on the convex hull, while red points denote materials that are within 50 meV/atom from the hull.

three functionals in what concerns formation energies and distances to the convex hull, so the SCAN diagrams should on average have the highest accuracy.

### Usage Notes

The data can be downloaded from the Materials Cloud repository<sup>36</sup>. The energies, compositions, and structures for each material are formatted as *ComputedStructureEntries* and stored as compressed json files. They can therefore be trivially loaded in python and analyzed with pymatgen. We note that we used version v2019.10.2 of pymatgen, but the data should be compatible with other versions.

### Code availability

All data can be easily processed with publicly available tools such as json and pymatgen<sup>35</sup>. An example usage is provided with the data. The dataset was generated with VASP, the bash and python scripts to generate input files or manage the output files can be downloaded from github repository: [https://github.com/hyllios/utills/tree/main/ht\\_pd\\_scan](https://github.com/hyllios/utills/tree/main/ht_pd_scan).

Received: 16 April 2021; Accepted: 25 January 2022;

Published online: 02 March 2022

### References

1. Curtarolo, S. *et al.* AFLOW: An automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **58**, 218–226, <https://doi.org/10.1016/j.commatsci.2012.02.005> (2012).
2. Jain, A. *et al.* Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002, <https://doi.org/10.1063/1.4812323> (2013).
3. Körbel, S., Marques, M. A. L. & Botti, S. Stable hybrid organic–inorganic halide perovskites for photovoltaics from ab initio high-throughput calculations. *J. Mater. Chem. A* **6**, 6463–6475, <https://doi.org/10.1039/c7ta08992a> (2018).
4. Grauzinytė, M., Botti, S., Marques, M. A. L., Goedecker, S. & Flores-Livas, J. A. Computational acceleration of prospective dopant discovery in cuprous iodide. *Phys. Chem. Chem. Phys.* **21**, 18839–18849, <https://doi.org/10.1039/c9cp02711d> (2019).
5. Flores-Livas, J. A., Sarmiento-Pérez, R., Botti, S., Goedecker, S. & Marques, M. A. L. Rare-earth magnetic nitride perovskites. *J. Phys. Mater.* **2**, 025003, <https://doi.org/10.1088/2515-7639/ab083e> (2019).
6. Wang, H.-C., Pistor, P., Marques, M. A. L. & Botti, S. Double perovskites as p-type conducting transparent semiconductors: a high-throughput search. *J. Mater. Chem. A* **7**, 14705–14711, <https://doi.org/10.1039/c9ta01456j> (2019).
7. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865, <https://doi.org/10.1103/PhysRevLett.77.3865> (1996).
8. Armiento, R. & Mattsson, A. E. Functional designed to include surface effects in self-consistent density functional theory. *Phys. Rev. B* **72**, 085108, <https://doi.org/10.1103/PhysRevB.72.085108> (2005).
9. Perdew, J. P. *et al.* Restoring the density-gradient expansion for exchange in solids and surfaces. *Phys. Rev. Lett.* **100**, 136406, <https://doi.org/10.1103/PhysRevLett.100.136406> (2008).
10. Csonka, G. I. *et al.* Assessing the performance of recent density functionals for bulk solids. *Phys. Rev. B* **79**, 155107, <https://doi.org/10.1103/PhysRevB.79.155107> (2009).
11. Tran, F., Stelzl, J. & Blaha, P. Rungs 1 to 4 of DFT Jacob's ladder: Extensive test on the lattice constant, bulk modulus, and cohesive energy of solids. *J. Chem. Phys.* **144**, 204120, <https://doi.org/10.1063/1.4948636> (2016).

12. Sun, J., Ruzsinszky, A. & Perdew, J. P. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.* **115**, 036402, <https://doi.org/10.1103/PhysRevLett.115.036402> (2015).
13. Zhang, Y. *et al.* Efficient first-principles prediction of solid stability: Towards chemical accuracy. *Npj Comput. Mater.* **4**, <https://doi.org/10.1038/s41524-018-0065-z> (2018).
14. Hastrup, S. *et al.* The computational 2D materials database: high-throughput modeling and discovery of atomically thin crystals. *2D Materials* **5**, 042002, <https://doi.org/10.1088/2053-1583/aacfc1> (2018).
15. Gjerding, M. N. *et al.* Recent progress of the computational 2d materials database (c2db). *2D Materials* **8**, 044002, <https://doi.org/10.1088/2053-1583/ac1059> (2021).
16. Thonhauser, T. *et al.* Van der waals density functional: Self-consistent potential and the nature of the van der waals bond. *Phys. Rev. B* **76**, 125112, <https://doi.org/10.1103/PhysRevB.76.125112> (2007).
17. Klimeš, J. C. V., Bowler, D. R. & Michaelides, A. Van der waals density functionals applied to solids. *Phys. Rev. B* **83**, 195131, <https://doi.org/10.1103/PhysRevB.83.195131> (2011).
18. Choudhary, K. *et al.* The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design. *npj Computational Materials* **6**, <https://doi.org/10.1038/s41524-020-00440-1> (2020).
19. Choudhary, K. *et al.* Computational screening of high-performance optoelectronic materials using OptB88vdw and TB-mBJ formalisms. *Scientific Data* **5**, <https://doi.org/10.1038/sdata.2018.82> (2018).
20. Choudhary, K., Cheon, G., Reed, E. & Tavazza, F. Elastic properties of bulk and low-dimensional materials using van der Waals density functional. *Phys. Rev. B* **98**, 014107, <https://doi.org/10.1103/PhysRevB.98.014107> (2018).
21. Schmidt, J., Pettersson, L., Verdozzi, C., Botti, S. & Marques, M. A. L. Crystal graph attention networks for the prediction of stable materials. *Science Advances* **7**, eabi7948, <https://doi.org/10.1126/sciadv.abi7948> (2021).
22. Oses, C. *et al.* Aflow-chull: Cloud-oriented platform for autonomous phase stability analysis. *J. Chem. Inf. Model.* **58**(12), 2477–2490, <https://doi.org/10.1021/acs.jcim.8b00393> (2018).
23. Schmidt, J. & Pettersson, L. Crystal-graph attention networks for the prediction of stable materials. *Materials Cloud* <https://archive.materialscloud.org/record/2021.222> (2021).
24. Emery, A. A. & Wolverton, C. High-throughput DFT calculations of formation energy, stability and oxygen vacancy formation energy of ABO<sub>3</sub> perovskites. *Sci. Data* **4**, <https://doi.org/10.1038/sdata.2017.153> (2017).
25. Sun, W. *et al.* The thermodynamic scale of inorganic crystalline metastability. *Sci. Adv.* **2**, e1600225, <https://doi.org/10.1126/sciadv.1600225> (2016).
26. Sarmiento-Pérez, R., Botti, S. & Marques, M. A. L. Optimized exchange and correlation semilocal functional for the calculation of energies of formation. *J. Chem. Theory Comput.* **11**, 3844–3850, <https://doi.org/10.1021/acs.jctc.5b00529> (2015).
27. Stevanović, V., Lany, S., Zhang, X. & Zunger, A. Correcting density functional theory for accurate predictions of compound enthalpies of formation: Fitted elemental-phase reference energies. *Phys. Rev. B* **85**, 115104, <https://doi.org/10.1103/PhysRevB.85.115104> (2012).
28. Hautier, G., Ong, S. P., Jain, A., Moore, C. J. & Ceder, G. Accuracy of density functional theory in predicting formation energies of ternary oxides from binary oxides and its implication on phase stability. *Phys. Rev. B* **85**, 155208, <https://doi.org/10.1103/PhysRevB.85.155208> (2012).
29. Bartel, C. J., Weimer, A. W., Lany, S., Musgrave, C. B. & Holder, A. M. The role of decomposition reactions in assessing first-principles predictions of solid stability. *Npj Comput. Mater.* **5**, 1–9, <https://doi.org/10.1038/s41524-018-0143-2> (2019).
30. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979, <https://doi.org/10.1103/PhysRevB.50.17953> (1994).
31. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50, [https://doi.org/10.1016/0927-0256\(96\)00008-0](https://doi.org/10.1016/0927-0256(96)00008-0) (1996).
32. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186, <https://doi.org/10.1103/PhysRevB.54.11169> (1996).
33. Bartók, A. P. & Yates, J. R. Regularized SCAN functional. *J. Chem. Phys.* **150**, 161101, <https://doi.org/10.1063/1.5094646> (2019).
34. Furness, J. W., Kaplan, A. D., Ning, J., Perdew, J. P. & Sun, J. Accurate and numerically efficient r2SCAN meta-generalized gradient approximation. *J. Phys. Chem. Lett.* **11**, 8208–8215, <https://doi.org/10.1021/acs.jpcclett.0c02405> (2020).
35. Ong, S. P. *et al.* Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319, <https://doi.org/10.1016/j.commatsci.2012.10.028> (2013).
36. Schmidt, J., Wang, H.-C., Cerqueira, T. F. T., Botti, S. & Marques, M. A. L. A new dataset of 175k stable and metastable materials calculated with the PBEsol and SCAN functionals. *Materials Cloud* <https://archive.materialscloud.org/record/2021.164> (2021).
37. Schmidt, J. *et al.* Predicting the thermodynamic stability of solids combining density functional theory and machine learning. *Chem. Mater.* **29**, 5090–5103, <https://doi.org/10.1021/acs.chemmater.7b00156> (2017).
38. Schmidt, J., Chen, L., Botti, S. & Marques, M. A. L. Predicting the stability of ternary intermetallics with density functional theory and machine learning. *J. Chem. Phys.* **148**, 241728, <https://doi.org/10.1063/1.5020223> (2018).
39. Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD). *JOM* **65**, 1501–1509, <https://doi.org/10.1007/s11837-013-0755-4> (2013).
40. Hautier, G., Fischer, C. C., Jain, A., Mueller, T. & Ceder, G. Finding nature's missing ternary oxide compounds using machine learning and density functional theory. *Chem. Mater.* **22**, 3762–3767, <https://doi.org/10.1021/cm100795d> (2010).
41. Shi, J. *et al.* High-throughput search of ternary chalcogenides for p-type transparent electrodes. *Sci. Rep.* **7**, <https://doi.org/10.1038/srep43179> (2017).
42. Olynyk, A. O. *et al.* High-throughput machine-learning-driven synthesis of full-heusler compounds. *Chem. Mater.* **28**, 7324–7331, <https://doi.org/10.1021/acs.chemmater.6b02724> (2016).
43. Haas, P., Tran, F. & Blaha, P. Calculation of the lattice constant of solids with semilocal functionals. *Phys. Rev. B* **79**, <https://doi.org/10.1103/physrevb.79.085104> (2009).
44. Isaacs, E. B. & Wolverton, C. Performance of the strongly constrained and appropriately normed density functional for solid-state materials. *Phys. Rev. Mater.* **2**, 063801, <https://doi.org/10.1103/PhysRevMaterials.2.063801> (2018).
45. Borlido, P. *et al.* Exchange–correlation functionals for band gaps of solids: benchmark, reparametrization and machine learning. *Npj Comput. Mater.* **6**, <https://doi.org/10.1038/s41524-020-00360-0> (2020).
46. Borlido, P. *et al.* Large-scale benchmark of exchange–correlation functionals for the determination of electronic band gaps of solids. *J. Chem. Theory Comput.* **15**, 5069–5079, <https://doi.org/10.1021/acs.jctc.9b00322> (2019).

## Acknowledgements

The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. ([www.gauss-centre.eu](http://www.gauss-centre.eu)) for funding this project by providing computing time on the GCS Supercomputer SuperMUC-NG at the Leibniz Supercomputing Centre under the project pn68wa. We would also like to thank Georg Kresse for providing us with an improved PAW setup for Cs.



### Author contributions

J.S., H.W. and M.A.L.M. performed the calculations. All authors contributed to the analysis of the results and to the writing the manuscript.

### Funding

Open Access funding enabled and organized by Projekt DEAL.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to M.A.L.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



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

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2022

## 5 Conclusion and Outlook

In three publications, we have presented various advancements for machine learning applications in theoretical solid-state physics and material science.

In the first publication, we have successfully introduced and implemented the idea of using automatic differentiation to simultaneously train a neural network as an exchange-correlation functional and its functional derivative. In contrast to only training the energy functional, this approach allows us to accurately reproduce the exact exchange-correlation potentials for strongly correlated two-electron systems. For random as well as specifically selected test systems the functional significantly outperforms the classical LDA in one dimension. The neural networks make it possible to arbitrarily scale the locality of the trained functionals, which alleviates some of the traditional challenges of

DFT. Specifically, increasing the non-locality allows the functional to correctly reproduce the one-dimensional  $\text{H}_2$  dissociation process for considerably larger distances than possible with classical local functionals.

In the second publication, we have suggested and implemented a modification to graph networks that allows us to use them in prototype-based high-throughput searches. This development has enabled us to perform high-throughput searches of already more than 2000 prototypes spanning a compound space of one billion materials. As a result, we can suggest a large number of potentially stable mixed perovskites, garnets and more than ten thousand other materials.

As a secondary result of this work, we have curated a dataset of more than 2 million crystal structures with consistent calculation parameters. This dataset already allows most researchers to upgrade their machine learning models solely by providing more training data than was previously available to them. Furthermore, the dataset provides a rather complete convex hull, increasing the accuracy of thermodynamic stabilities.

Machine learning supported high-throughput studies are already a relatively mature field with successful applications over the last seven years. However, due to the need for separate training data for most prototypes, high-throughput searches have been limited in scope. We think that the works presented in this thesis can significantly expand this scope. Considering that prototype-based high-throughput publications were mostly limited to one ternary or quaternary prototype, and we already performed high-throughput searches of more than a thousand ternary prototypes just this year, the scope has already massively expanded.

In the third publication, we have upgraded our convex hull by recal-



culating the majority of compounds close to the convex hull with the PBEsol and the SCAN functional. To the best of our knowledge, the resulting convex hull is the first reasonably complete hull calculated with a functional more precise than the PBE functional. It also provides more precise geometries and bandgaps for interested experimentalists, allowing them to better identify promising candidates for synthesis and further study.

## 5.1 Outlook

We are in the process of addressing two of the challenges high-throughput searches are facing right now. First, we are attempting to lessen the biases in terms of prototypes and elements in our data sets to achieve a more consistent error across the chemical space. Secondly, we are upgrading the machine learning models via transfer learning to the accuracy of SCAN and PBEsol. It seems that the number of DFT-geometry optimizations performed during a comprehensive high-throughput search cannot be significantly reduced at the moment. However, there is substantial potential to reduce the number of steps during each geometry optimization through machine learning force fields. So far solid-state machine learning force fields have been limited in structure and stoichiometry space[JSPHD6].

We plan to harness the millions of force and energy data points that result from our DFT geometry optimizations each year and combine them with state of the art graph neural networks. This should allow us to cut the number of DFT geometry optimization steps and consequently the computational cost by at least a factor of three. For this purpose, we are already saving geometry optimization steps from the large-scale high-throughput search discussed in section 4.2 and the cal-

culations in Ref. [JSPHD11]. Most recently, work in this direction on a smaller scale has already been published by Chen *et al.* [19].

Our work on machine learning functionals was mostly a proof of concept work demonstrating that training accurate exchange correlation potentials is possible through automatic differentiation. Due to the abundance of high-fidelity training and validation data produced by coupled cluster, full-CI, and quantum Monte Carlo (QMC) calculations, the future of machine learning functionals in chemistry is quite promising. Using this data and classical functional forms enhanced with machine learning, the most recent works were able to produce functionals [78] that consistently improve over traditional functionals although at a higher computational cost.

There are now several research directions to proceed from here. Traditional functionals use a very limited amount of data and parameters. Continuing on the path of using the entire machine learning toolbox to fit traditional functional forms enhanced by some machine learning should allow us to progress beyond the existing functionals. While this approach will be useful, it seems unlikely to produce real innovations. A promising research direction could be to replace the expensive hybrid parts of functionals with semi-local machine-learned terms to improve the calculation speed, while retaining the accuracy of existing hybrids. A further research direction is to add exact conditions through the data, and functional form, that are mostly ignored in traditional functionals, e.g., the derivative discontinuity [JSPHD3]. Removing such a failure point of DFT from functionals has the potential to make a major difference for the applicability and accuracy of DFT.

In solid-state DFT the picture is different to quantum chemistry as high-quality data exists mainly for the electron gas. Consequently, as

a first step, the research community has to find a suitable calculation method to produce high fidelity benchmark and training data. This would allow us to progress to the point where quantum chemistry is right now. The calculation method will most likely have to be a type of QMC, e.g., diffusion- or auxiliary field QMC. Machine learning based variational and diffusion Monte Carlo approaches have also shown some promise for this problem in the last two years [123, 171, 149]. However, it seems, without such data, progress can neither be achieved nor even be properly evaluated.

## Publications during PhD

- [JSPHD1] Pedro Borlido, **Jonathan Schmidt**, Ahmad W. Huran, Fabien Tran, Miguel A. L. Marques, and Silvana Botti. “Exchange-correlation functionals for band gaps of solids: benchmark, reparametrization and machine learning”. In: *npj Comput. Mater.* 6.1 (2020). DOI: [10.1038/s41524-020-00360-0](https://doi.org/10.1038/s41524-020-00360-0).
- [JSPHD2] Pedro Borlido, **Jonathan Schmidt**, Haichen Wang, Silvana Botti, and Miguel A. L. Marques. “Computational screening of materials with extreme gap deformation potentials”. In: *accepted in npj Comput. Mater.* (2022).
- [JSPHD3] Johannes Gedeon, **Jonathan Schmidt**, Matthew J.P. Hodgson, Jack Wetherell, Carlos L. Benavides-Riveros, and Miguel A. L. Marques. “Machine learning the derivative discontinuity of density-functional theory”. In: *Mach. learn.: sci. technol.* (2021). DOI: [10.1088/2632-2153/ac3149](https://doi.org/10.1088/2632-2153/ac3149).
- [JSPHD4] Noah Hoffmann, Tiago F. T. Cerqueira, **Jonathan Schmidt**, and Miguel A. L. Marques. “Superconductivity in antiperovskites”. In: *accepted in principle in npj Comput. Mater.* (2022).
- [JSPHD5] Robert Hussein, **Jonathan Schmidt**, Tomás Barros, Miguel AL Marques, and Silvana Botti. “Machine-learning correction to density-functional crystal structure optimization”. In: *MRS Bull.* (2022). DOI: [10.1557/s43577-022-00310-9](https://doi.org/10.1557/s43577-022-00310-9).
- [JSPHD6] Heather Kulik, Thomas Hammerschmidt, **Jonathan, Schmidt**, et al. “Roadmap on Machine Learning in Electronic Structure”. In: *Electron. Struct.* (2022). DOI: [10.1088/2516-1075/ac572f](https://doi.org/10.1088/2516-1075/ac572f).
- [JSPHD7] **Jonathan Schmidt**, Carlos L. Benavides-Riveros, and Miguel A. L. Marques. “Machine Learning the Physical Nonlocal Exchange–Correlation Functional of Density-Functional Theory”. In: *J. Phys. Chem. Lett.* 10.20 (2019), pp. 6425–6431. DOI: [10.1021/acs.jpcllett.9b02422](https://doi.org/10.1021/acs.jpcllett.9b02422).
- [JSPHD8] **Jonathan Schmidt**, Matteo Fadel, and Carlos L. Benavides-Riveros. “Machine learning universal bosonic functionals”. In: *Phys. Rev. Res.* 3 (3 2021), p. L032063. DOI: [10.1103/PhysRevResearch.3.L032063](https://doi.org/10.1103/PhysRevResearch.3.L032063).

- [JSPHD9] **Jonathan Schmidt**, Mário R. G. Marques, Silvana Botti, and Miguel A. L. Marques. “Recent advances and applications of machine learning in solid-state materials science”. In: *npj Comput. Mater.* 5.1 (2019). DOI: [10.1038/s41524-019-0221-0](https://doi.org/10.1038/s41524-019-0221-0).
- [JSPHD10] **Jonathan Schmidt**, Love Pettersson, Claudio Verdozzi, Silvana Botti, and Miguel A. L. Marques. “Crystal graph attention networks for the prediction of stable materials”. In: *Sci. Adv.* 7.49 (2021), eabi7948. DOI: [10.1126/sciadv.abi7948](https://doi.org/10.1126/sciadv.abi7948).
- [JSPHD11] **Jonathan Schmidt**, Hai-Chen Wang, Tiago F. T. Cerqueira, Silvana Botti, and Miguel A. L. Marques. “A new dataset of 175k stable and metastable materials calculated with the PBEsol and SCAN functionals”. In: *Sci. Data* 12.1 (2022). DOI: [10.1038/s41597-022-01177-w](https://doi.org/10.1038/s41597-022-01177-w).
- [JSPHD12] Haichen Wang, **Jonathan Schmidt**, Silvana Botti, and Miguel AL Marques. “High-throughput study of oxynitride, oxyfluoride and nitrofluoride perovskites”. In: *J. Mater. Chem. A* (2021). DOI: [10.1039/D0TA10781F](https://doi.org/10.1039/D0TA10781F).

## Publications before PhD

- [JS1] **Jonathan Schmidt**, Carlos L. Benavides-Riveros, and Miguel A. L. Marques. “Reduced density matrix functional theory for superconductors”. In: *Phys. Rev. B* 99 (22 2019), p. 224502. DOI: [10.1103/PhysRevB.99.224502](https://doi.org/10.1103/PhysRevB.99.224502).
- [JS2] **Jonathan Schmidt**, Carlos L. Benavides-Riveros, and Miguel A. L. Marques. “Representability problem of density functional theory for superconductors”. In: *Phys. Rev. B* 99.2 (2019). DOI: [10.1103/physrevb.99.024502](https://doi.org/10.1103/physrevb.99.024502).
- [JS3] **Jonathan Schmidt**, Liming Chen, Silvana Botti, and Miguel A. L. Marques. “Predicting the stability of ternary intermetallics with density functional theory and machine learning”. In: *J. Chem. Phys.* 148.24 (2018), p. 241728. DOI: [10.1063/1.5020223](https://doi.org/10.1063/1.5020223).

- [JS4] **Jonathan Schmidt**, Jingming Shi, Pedro Borlido, Liming Chen, Silvana Botti, and Miguel A. L. Marques. “Predicting the Thermodynamic Stability of Solids Combining Density Functional Theory and Machine Learning”. In: *Chem. Mater.* 29.12 (2017), pp. 5090–5103. DOI: [10.1021/acs.chemmater.7b00156](https://doi.org/10.1021/acs.chemmater.7b00156).

## Copyright

Publication JSPhd5: Reprinted with permission from Jonathan Schmidt *et al.* *J. Phys. Chem. Lett.* 10.20 (Oct.2019). Copyright 2019 American Chemical Society.

Publication JSPhd8: Reprinted with permission from Jonathan Schmidt *et al.*, *Sci. Adv.* 7.49 (2021), eabi7948, Copyright © 2021 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science.

## References

- [1] Martín Abadi, Ashish Agarwal, Paul Barham, et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2011.
- [2] F. H. Allen, G. Gergerhoff, and R. Sievers, eds. *Crystallographic databases*. International Union of Crystallography, Chester, 1987.
- [3] Ethem Alpaydin. *Introduction to Machine Learning*. The MIT Press, 2014. 640 pp.
- [4] O. Krogh Andersen. “Linear methods in band theory”. In: *Phys. Rev. B* 12 (8 1975), pp. 3060–3083. DOI: [10.1103/PhysRevB.12.3060](https://doi.org/10.1103/PhysRevB.12.3060).
- [5] Hexin Bai, Peng Chu, Jeng-Yuan Tsai, et al. “Graph neural network for hamiltonian-based material property prediction”. In: *arXiv preprint arXiv:2005.13352* (2020).

- [6] Pierre Baldi and Soren Brunak. *Bioinformatics: the machine learning approach*. The MIT Press, 2001. 476 pp.
- [7] Christopher J. Bartel, Amalie Trewartha, Qi Wang, Alexander Dunn, Anubhav Jain, and Gerbrand Ceder. “A critical examination of compound stability predictions from machine-learned formation energies”. In: *npj Comput. Mater.* 97 (1 2020). DOI: [10.1038/s41524-020-00362-y](https://doi.org/10.1038/s41524-020-00362-y).
- [8] Albert P. Bartók and Jonathan R. Yates. “Regularized SCAN functional”. In: *J. Chem. Phys.* 150.16 (2019), p. 161101. DOI: [10.1063/1.5094646](https://doi.org/10.1063/1.5094646).
- [9] Simon Batzner, Albert Musaelian, Lixin Sun, et al. “Se (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials”. In: *arXiv preprint arXiv:2101.03164* (2021).
- [10] A. D. Becke and E. R. Johnson. “A Simple Effective Potential for Exchange”. In: *J. Chem. Phys.* 124.22 (2006), p. 221101. ISSN: 0021-9606, 1089-7690. DOI: [10.1063/1.2213970](https://doi.org/10.1063/1.2213970).
- [11] Axel D. Becke. “Density-functional thermochemistry. III. The role of exact exchange”. In: *J. Chem. Phys.* 98.7 (1993), pp. 5648–5652. DOI: [10.1063/1.464913](https://doi.org/10.1063/1.464913).
- [12] Christopher Berner, Greg Brockman, Brooke Chan, et al. “Dota 2 with large scale deep reinforcement learning”. In: *arXiv preprint arXiv:1912.06680* (2019).
- [13] P. E. Blöchl. “Projector augmented-wave method”. In: *Phys. Rev. B* 50 (24 1994), pp. 17953–17979. DOI: [10.1103/PhysRevB.50.17953](https://doi.org/10.1103/PhysRevB.50.17953).
- [14] Mihail Bogojeski, Felix Brockherde, Leslie Vogt-Maranto, et al. “Efficient prediction of 3D electron densities using machine learning”. In: *arXiv preprint arXiv:1811.06255* (2018).
- [15] Mariusz Bojarski, Davide Del Testa, Daniel Dworakowski, et al. “End to End Learning for Self-Driving Cars”. In: *arXiv preprint ()*, arXiv:1604.07316.
- [16] M. Born and R. Oppenheimer. “Zur Quantentheorie der Molekeln”. In: *Ann Phys.* 389.20 (1927), pp. 457–484. DOI: [10.1002/andp.19273892002](https://doi.org/10.1002/andp.19273892002).
- [17] Leo Breiman. “Random forests”. In: *Mach. Learn.* 45.1 (2001), pp. 5–32. DOI: [10.1023/a:1010933404324](https://doi.org/10.1023/a:1010933404324).

- [18] Felix Brockherde, Leslie Vogt, Li Li, Mark E. Tuckerman, Kieron Burke, and Klaus-Robert Müller. “Bypassing the Kohn-Sham equations with machine learning”. In: *Nat. Commun.* 8.1 (2017), p. 872. DOI: [10.1038/s41467-017-00839-3](https://doi.org/10.1038/s41467-017-00839-3).
- [19] Chi Chen and Shyue Ping Ong. “A Universal Graph Deep Learning Interatomic Potential for the Periodic Table”. In: *arXiv preprint arXiv:2202.02450* (2022).
- [20] Chi Chen, Weike Ye, Yunxing Zuo, Chen Zheng, and Shyue Ping Ong. “Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals”. In: *Chem. Mater.* 31.9 (2019), pp. 3564–3572. DOI: [10.1021/acs.chemmater.9b01294](https://doi.org/10.1021/acs.chemmater.9b01294).
- [21] Chi Chen, Weike Ye, Yunxing Zuo, Chen Zheng, and Shyue Ping Ong. “Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals”. In: *Chem. Mater.* 31.9 (2019), pp. 3564–3572. DOI: [10.1021/acs.chemmater.9b01294](https://doi.org/10.1021/acs.chemmater.9b01294).
- [22] Kamal Choudhary, Gowoon Cheon, Evan Reed, and Francesca Tavazza. “Elastic properties of bulk and low-dimensional materials using van der Waals density functional”. In: *Phys. Rev. B* 98 (1 2018), p. 014107. DOI: [10.1103/PhysRevB.98.014107](https://doi.org/10.1103/PhysRevB.98.014107).
- [23] Kamal Choudhary, Kevin F. Garrity, Andrew C. E. Reid, et al. “The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design”. In: *npj Comput. Mat.* 6.1 (2020). DOI: [10.1038/s41524-020-00440-1](https://doi.org/10.1038/s41524-020-00440-1).
- [24] Kamal Choudhary, Qin Zhang, Andrew C.E. Reid, et al. “Computational screening of high-performance optoelectronic materials using OptB88vdW and TB-mBJ formalisms”. In: *Sci. Data* 5.1 (2018). DOI: [10.1038/sdata.2018.82](https://doi.org/10.1038/sdata.2018.82).
- [25] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. “Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)”. In: *arXiv preprint* (2015), arXiv:1511.07289.



- [26] Gábor I. Csonka, John P. Perdew, Adrienn Ruzsinszky, et al. “Assessing the performance of recent density functionals for bulk solids”. In: *Phys. Rev. B* 79 (15 2009), p. 155107. DOI: <https://doi.org/10.1103/PhysRevB.79.155107>.
- [27] Stefano Curtarolo, Wahyu Setyawan, Gus L.W. Hart, et al. “AFLOW: An automatic framework for high-throughput materials discovery”. In: *Comput. Mater. Sci.* 58 (2012), pp. 218–226. ISSN: 0927-0256. DOI: <https://doi.org/10.1016/j.commatsci.2012.02.005>.
- [28] Nicola De Cao and Thomas Kipf. “MolGAN: An implicit generative model for small molecular graphs”. In: *arXiv preprint arXiv:1805.11973* (2018).
- [29] Partha Dey, Joe Bible, Somnath Datta, et al. “Informatics-aided bandgap engineering for solar materials”. In: *Comput. Mater. Sci.* 83 (2014), pp. 185–195. DOI: [10.1016/j.commatsci.2013.10.016](https://doi.org/10.1016/j.commatsci.2013.10.016).
- [30] Sebastian Dick and Marivi Fernandez-Serra. “Machine learning accurate exchange and correlation functionals of the electronic density”. In: *Nat. Commun.* 11.1 (2020). DOI: [10.1038/s41467-020-17265-7](https://doi.org/10.1038/s41467-020-17265-7).
- [31] Marie Dumaz, Reese Boucher, Miguel AL Marques, and Aldo H Romero. “Authorship and citation cultural nature in Density Functional Theory from solid state computational packages”. In: *Scientometrics* 126.8 (2021), pp. 6681–6695. DOI: [10.1007/s11192-021-04057-z](https://doi.org/10.1007/s11192-021-04057-z).
- [32] Alexander Dunn, Qi Wang, Alex Ganose, Daniel Dopp, and Anubhav Jain. “Benchmarking Materials Property Prediction Methods: The Matbench Test Set and Automatminer Reference Algorithm”. In: *arXiv preprint arXiv:2005.00707* (2020).
- [33] Felix Faber, Alexander Lindmaa, O. Anatole von Lilienfeld, and Rickard Armiento. “Crystal structure representations for machine learning models of formation energies”. In: *Int. J. Quantum Chem.* 115.16 (2015), pp. 1094–1101. DOI: [10.1002/qua.24917](https://doi.org/10.1002/qua.24917).
- [34] Felix A. Faber, Alexander Lindmaa, O. Anatole von Lilienfeld, and Rickard Armiento. “Machine Learning Energies of 2 Million Elpasolite (ABC<sub>2</sub>D<sub>6</sub>) Crystals”. In: *Phys. Rev. Lett.* 117.13 (2016), p. 135502. DOI: [10.1103/physrevlett.117.135502](https://doi.org/10.1103/physrevlett.117.135502).

- [35] Carlos Fiolhais, Fernando Nogueira, and Miguel AL Marques. *A primer in density functional theory*. Vol. 620. Springer Science & Business Media, 2003.
- [36] M. J. Frisch, G. W. Trucks, H. B. Schlegel, et al. *Gaussian~16 Revision C.01*. Gaussian Inc. Wallingford CT. 2016.
- [37] Victor Fung, Jiaxin Zhang, Eric Juarez, and Bobby G. Sumpter. “Benchmarking graph neural networks for materials chemistry”. In: *npj Comput. Mat.* 7.1 (2021). DOI: [10.1038/s41524-021-00554-0](https://doi.org/10.1038/s41524-021-00554-0).
- [38] James W. Furness, Aaron D. Kaplan, Jinliang Ning, John P. Perdew, and Jianwei Sun. “Accurate and Numerically Efficient rSCAN Meta-Generalized Gradient Approximation”. In: *J. Phys. Chem. Lett* 11.19 (2020). PMID: 32876454, pp. 8208–8215. DOI: [10.1021/acs.jpcllett.0c02405](https://doi.org/10.1021/acs.jpcllett.0c02405).
- [39] A. P. Gaiduk and V. N. Staroverov. “How to tell when a model Kohn–Sham potential is not a functional derivative”. In: *J. Chem. Phys.* 131.4 (2009), p. 044107. DOI: [10.1063/1.3176515](https://doi.org/10.1063/1.3176515).
- [40] Michael Gastegger, Adam McSloy, M Luya, Kristof T Schütt, and Reinhard J Maurer. “A deep neural network for molecular wave functions in quasi-atomic minimal basis representation”. In: *J. Chem. Phys.* 153.4 (2020), p. 044123. DOI: [10.1063/5.0012911](https://doi.org/10.1063/5.0012911).
- [41] Michael Gastegger, Kristof T Schütt, and Klaus-Robert Müller. “Machine learning of solvent effects on molecular spectra and reactions”. In: *Chem. Sci.* 12.34 (2021), pp. 11473–11483. DOI: [10.1039/D1SC02742E](https://doi.org/10.1039/D1SC02742E).
- [42] Niklas WA Gebauer, Michael Gastegger, Stefaan SP Hessmann, Klaus-Robert Müller, and Kristof T Schütt. “Inverse design of 3d molecular structures with conditional generative neural networks”. In: *arXiv preprint arXiv:2109.04824* (2021).
- [43] Niklas WA Gebauer, Michael Gastegger, and Kristof T Schütt. “Symmetry-adapted generation of 3d point sets for the targeted discovery of molecules”. In: *arXiv preprint arXiv:1906.00957* (2019).
- [44] Pierre Geurts, Damien Ernst, and Louis Wehenkel. “Extremely randomized trees”. In: *Mach. Learn.* 63.1 (2006), pp. 3–42. DOI: [10.1007/s10994-006-6226-1](https://doi.org/10.1007/s10994-006-6226-1).

- [45] P Giannozzi, O Andreussi, T Brumme, et al. “Advanced capabilities for materials modelling with QUANTUM ESPRESSO”. In: *J. Condens. Matter Phys.* 29.46 (2017), p. 465901. DOI: [10.1088/1361-648X/aa8f79](https://doi.org/10.1088/1361-648X/aa8f79).
- [46] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. “Neural message passing for quantum chemistry”. In: *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org. 2017, pp. 1263–1272.
- [47] Morten Niklas Gjerding, Alireza Taghizadeh, Asbjørn Rasmussen, et al. “Recent progress of the computational 2D materials database (C2DB)”. In: *2d Mater.* 8.4 (2021), p. 044002. DOI: [10.1088/2053-1583/ac1059](https://doi.org/10.1088/2053-1583/ac1059).
- [48] Xavier Glorot and Yoshua Bengio. “Understanding the difficulty of training deep feedforward neural networks”. In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings. 2010, pp. 249–256.
- [49] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. “Deep Sparse Rectifier Neural Networks”. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Ed. by Geoffrey Gordon, David Dunson, and Miroslav Dudík. Vol. 15. Proceedings of Machine Learning Research. Fort Lauderdale, FL, USA, 2011, pp. 315–323.
- [50] Xavier Gonze, Bernard Amadon, Gabriel Antonius, et al. “The Abinit-project: Impact, environment and recent developments”. In: *Computer Physics Communications* 248 (2020), p. 107042. DOI: [10.1016/j.cpc.2019.107042](https://doi.org/10.1016/j.cpc.2019.107042).
- [51] Rhys EA Goodall and Alpha A Lee. “Predicting materials properties without crystal structure: Deep representation learning from stoichiometry”. In: *arXiv preprint arXiv:1910.00617* (2019).
- [52] Rhys EA Goodall, Abhijith S Parackal, Felix A Faber, Rickard Armiento, and Alpha A Lee. “Rapid Discovery of Novel Materials by Coordinate-free Coarse Graining”. In: *arXiv preprint arXiv:2106.11132* (2021).
- [53] Stefan Grimme. “Semiempirical hybrid density functional with perturbative second-order correlation”. In: *J. Chem. Phys.* 124.3 (2006), p. 034108. DOI: [10.1063/1.2148954](https://doi.org/10.1063/1.2148954).

- [54] Andrea Grisafi, Alberto Fabrizio, Benjamin Meyer, David M. Wilkins, Clemence Corminboeuf, and Michele Ceriotti. “Transferable Machine-Learning Model of the Electron Density”. In: *ACS Cent. Sci.* 5.1 (2018), pp. 57–64. DOI: [10.1021/acscentsci.8b00551](https://doi.org/10.1021/acscentsci.8b00551).
- [55] Tianhong Gu, Wencong Lu, Xinhua Bao, and Nianyi Chen. “Using support vector regression for the prediction of the band gap and melting point of binary and ternary compound semiconductors”. In: *Solid State Sci.* 8.2 (2006), pp. 129–136. DOI: [10.1016/j.solidstatesciences.2005.10.011](https://doi.org/10.1016/j.solidstatesciences.2005.10.011).
- [56] David Ha, Andrew Dai, and Quoc V Le. “Hypernetworks”. In: *arXiv preprint arXiv:1609.09106* (2016).
- [57] Philipp Haas, Fabien Tran, and Peter Blaha. “Calculation of the lattice constant of solids with semilocal functionals”. In: *Phys. Rev. B* 79.8 (2009). DOI: <https://doi.org/10.1103/physrevb.79.085104>.
- [58] Sten Haastrup, Mikkel Strange, Mohnish Pandey, et al. “The Computational 2D Materials Database: high-throughput modeling and discovery of atomically thin crystals”. In: *2d Mater.* 5.4 (2018), p. 042002. DOI: [10.1088/2053-1583/aacfc1](https://doi.org/10.1088/2053-1583/aacfc1).
- [59] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778. DOI: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [60] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification”. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2015. DOI: [10.1109/iccv.2015.123](https://doi.org/10.1109/iccv.2015.123).
- [61] Dan Hendrycks and Kevin Gimpel. “Gaussian error linear units (gelus)”. In: *arXiv preprint arXiv:1606.08415* (2016).
- [62] P. Hohenberg and W. Kohn. “Inhomogeneous Electron Gas”. In: *Phys. Rev.* 136.3B (1964), B864–B871. DOI: [10.1103/physrev.136.b864](https://doi.org/10.1103/physrev.136.b864).
- [63] Jacob Hollingsworth, Thomas E. Baker, and Kieron Burke. “Can exact conditions improve machine-learned density functionals?” In: *J. Chem. Phys.* 148.24 (2018), p. 241743. DOI: [10.1063/1.5025668](https://doi.org/10.1063/1.5025668).

- [64] Hisayoshi Iikura, Takao Tsuneda, Takeshi Yanai, and Kimihiko Hirao. “A long-range correction scheme for generalized-gradient-approximation exchange functionals”. In: *J. Chem. Phys.* 115.8 (2001), pp. 3540–3544. DOI: [10.1063/1.1383587](https://doi.org/10.1063/1.1383587).
- [65] Eric B. Isaacs and Chris Wolverton. “Performance of the strongly constrained and appropriately normed density functional for solid-state materials”. In: *Phys. Rev. Mater.* 2 (6 2018), p. 063801. DOI: <https://doi.org/10.1103/PhysRevMaterials.2.063801>.
- [66] Olexandr Isayev, Corey Oses, Cormac Toher, Eric Gossett, Stefano Curtarolo, and Alexander Tropsha. “Universal fragment descriptors for predicting properties of inorganic crystals”. In: *Nat. Commun.* 8 (2017), p. 15679. DOI: [10.1038/ncomms15679](https://doi.org/10.1038/ncomms15679).
- [67] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, et al. “Commentary: The Materials Project: A materials genome approach to accelerating materials innovation”. In: *APL Mater.* 1.1 (2013), p. 011002. DOI: [10.1063/1.4812323](https://doi.org/10.1063/1.4812323).
- [68] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, et al. “The Materials Project: A materials genome approach to accelerating materials innovation”. In: *APL Mater.* 1.1 (2013), p. 011002. ISSN: 2166532X. DOI: [10.1063/1.4812323](https://doi.org/10.1063/1.4812323).
- [69] Dipendra Jha, Logan Ward, Arindam Paul, et al. “ElemNet: Deep Learning the Chemistry of Materials From Only Elemental Composition”. In: *Sci. Rep.* 8.1 (2018), p. 17593. ISSN: 2045-2322. DOI: [10.1038/s41598-018-35934-y](https://doi.org/10.1038/s41598-018-35934-y).
- [70] Wengong Jin, Regina Barzilay, and Tommi Jaakkola. “Junction Tree Variational Autoencoder for Molecular Graph Generation”. In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. PMLR, 2018, pp. 2323–2332.
- [71] Peter Bjørn Jørgensen, Karsten Wedel Jacobsen, and Mikkel N Schmidt. “Neural message passing with edge updates for predicting properties of molecules and materials”. In: *arXiv preprint* (2018), arXiv:1806.03146.

- [72] John Jumper, Richard Evans, Alexander Pritzel, et al. “Highly accurate protein structure prediction with AlphaFold”. In: *Nature* 596.7873 (2021), pp. 583–589. DOI: [10.1038/s41586-021-03819-2](https://doi.org/10.1038/s41586-021-03819-2).
- [73] Bhupalee Kalita, Li Li, Ryan J. McCarty, and Kieron Burke. “Learning to Approximate Density Functionals”. In: *Acc. Chem. Res.* 54.4 (2021), pp. 818–826. DOI: [10.1021/acs.accounts.0c00742](https://doi.org/10.1021/acs.accounts.0c00742).
- [74] M. F. Kasim and S. M. Vinko. “Learning the Exchange-Correlation Functional from Nature with Fully Differentiable Density Functional Theory”. In: *Phys. Rev. Lett.* 127 (12 2021), p. 126403. DOI: [10.1103/PhysRevLett.127.126403](https://doi.org/10.1103/PhysRevLett.127.126403).
- [75] Kyoungdoc Kim, Logan Ward, Jiangang He, Amar Krishna, Ankit Agrawal, and C. Wolverton. “Machine-learning-accelerated high-throughput materials screening: Discovery of novel quaternary Heusler compounds”. In: *Phys. Rev. Mater.* 2.12 (2018), p. 123801. DOI: [10.1103/physrevmaterials.2.123801](https://doi.org/10.1103/physrevmaterials.2.123801).
- [76] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).
- [77] Scott Kirklin, James E Saal, Bryce Meredig, et al. “The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies”. In: *npj Comput. Mater.* 1 (2015), p. 15010. DOI: [10.1038/npjcompumats.2015.10](https://doi.org/10.1038/npjcompumats.2015.10).
- [78] James Kirkpatrick, Brendan McMorrow, David HP Turban, et al. “Pushing the frontiers of density functionals by solving the fractional electron problem”. In: *Science* 374.6573 (2021), pp. 1385–1389. DOI: [10.1126/science.abj6511](https://doi.org/10.1126/science.abj6511).
- [79] Johannes Klicpera, Janek Groß, and Stephan Günnemann. “Directional message passing for molecular graphs”. In: *arXiv preprint arXiv:2003.03123* (2020).
- [80] Boris Knyazev, Michal Drozdal, Graham W Taylor, and Adriana Romero Soriano. “Parameter Prediction for Unseen Deep Architectures”. In: *Advances in Neural Information Processing Systems* 34 (2021).

- [81] Dmitrii Kochkov, Tobias Pfaff, Alvaro Sanchez-Gonzalez, Peter Battaglia, and Bryan K Clark. “Learning ground states of quantum Hamiltonians with graph networks”. In: *arXiv preprint arXiv:2110.06390* (2021).
- [82] W. Kohn and L. J. Sham. “Self-Consistent Equations Including Exchange and Correlation Effects”. In: *Phys. Rev.* 140.4A (1965), A1133–A1138. DOI: [10.1103/physrev.140.a1133](https://doi.org/10.1103/physrev.140.a1133).
- [83] John F. Kolen and Stefan C. Kremer. “Gradient Flow in Recurrent Nets: The Difficulty of Learning Long-Term Dependencies”. In: *A Field Guide to Dynamical Recurrent Networks*. Wiley-IEEE Press, 2001. DOI: [10.1109/9780470544037.ch14](https://doi.org/10.1109/9780470544037.ch14).
- [84] G. Kresse and J. Furthmüller. “Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set”. In: *Phys. Rev. B* 54 (16 1996), pp. 11169–11186. DOI: [10.1103/PhysRevB.54.11169](https://doi.org/10.1103/PhysRevB.54.11169).
- [85] G. Kresse and J. Furthmüller. “Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set”. In: *Comput. Mater. Sci.* 6.1 (1996), pp. 15–50. ISSN: 0927-0256. DOI: [http://dx.doi.org/10.1016/0927-0256\(96\)00008-0](http://dx.doi.org/10.1016/0927-0256(96)00008-0).
- [86] G Kresse and J Hafner. “Norm-conserving and ultrasoft pseudopotentials for first-row and transition elements”. In: *J. Condens. Matter Phys.* 6.40 (1994), pp. 8245–8257. DOI: [10.1088/0953-8984/6/40/015](https://doi.org/10.1088/0953-8984/6/40/015).
- [87] G. Kresse and D. Joubert. “From ultrasoft pseudopotentials to the projector augmented-wave method”. In: *Phys. Rev. B* 59 (3 1999), pp. 1758–1775. DOI: [10.1103/PhysRevB.59.1758](https://doi.org/10.1103/PhysRevB.59.1758).
- [88] J. B. Krieger, Y. Li, and G. J. Iafrate. “Derivation and application of an accurate Kohn-Sham potential with integer discontinuity”. In: *Phys. Lett. A* 146.5 (1990), p. 256. ISSN: 0375-9601. DOI: [10.1016/0375-9601\(90\)90975-T](https://doi.org/10.1016/0375-9601(90)90975-T).
- [89] Youngchun Kwon, Jiho Yoo, Youn-Suk Choi, Won-Joon Son, Dongseon Lee, and Seokho Kang. “Efficient learning of non-autoregressive graph variational autoencoders for molecular graph generation”. In: *J. Cheminform.* 11.1 (2019), pp. 1–10. DOI: [10.1186/s13321-019-0396-x](https://doi.org/10.1186/s13321-019-0396-x).
- [90] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep learning”. In: *Nature* 521.7553 (2015), pp. 436–444. DOI: [10.1038/nature14539](https://doi.org/10.1038/nature14539).



- [91] Joohwi Lee, Atsuto Seko, Kazuki Shitara, Keita Nakayama, and Isao Tanaka. “Prediction model of band gap for inorganic compounds by combination of density functional theory calculations and machine learning techniques”. In: *Phys. Rev. B* 93.11 (2016), p. 115104. DOI: [10.1103/physrevb.93.115104](https://doi.org/10.1103/physrevb.93.115104).
- [92] Susi Lehtola, Conrad Steigemann, Micael JT Oliveira, and Miguel AL Marques. “Recent developments in libxc—A comprehensive library of functionals for density functional theory”. In: *SoftwareX* 7 (2018), pp. 1–5. DOI: [10.1016/j.softx.2017.11.002](https://doi.org/10.1016/j.softx.2017.11.002).
- [93] Xiangyun Lei and Andrew J. Medford. “Design and analysis of machine learning exchange-correlation functionals via rotationally invariant convolutional descriptors”. In: *Phys. Rev. Mater.* 3 (6 2019), p. 063801. DOI: [10.1103/PhysRevMaterials.3.063801](https://doi.org/10.1103/PhysRevMaterials.3.063801).
- [94] Li Li, Thomas E. Baker, Steven R. White, and Kieron Burke. “Pure density functional for strong correlation and the thermodynamic limit from machine learning”. In: *Phys. Rev. B* 94.24 (2016). DOI: [10.1103/physrevb.94.245129](https://doi.org/10.1103/physrevb.94.245129).
- [95] Li Li, Stephan Hoyer, Ryan Pederson, et al. “Kohn-Sham Equations as Regularizer: Building Prior Knowledge into Machine-Learned Physics”. In: *Phys. Rev. Lett.* 126.3 (2021). DOI: [10.1103/physrevlett.126.036401](https://doi.org/10.1103/physrevlett.126.036401).
- [96] Li Li, John C. Snyder, Isabelle M. Pelaschier, et al. “Understanding machine-learned density functionals”. In: *Int. J. Quantum Chem.* 116.11 (2015), pp. 819–833. DOI: [10.1002/qua.25040](https://doi.org/10.1002/qua.25040).
- [97] Wei Li, Ryan Jacobs, and Dane Morgan. “Predicting the thermodynamic stability of perovskite oxides using machine learning models”. In: *Comput. Mater. Sci.* 150 (2018), pp. 454–463. DOI: [10.1016/j.commatsci.2018.04.033](https://doi.org/10.1016/j.commatsci.2018.04.033).
- [98] Yibo Li, Liangren Zhang, and Zhenming Liu. “Multi-objective de novo drug design with conditional graph generative model”. In: *J. Cheminform* 10.1 (2018), pp. 1–24. DOI: [10.1186/s13321-018-0287-6](https://doi.org/10.1186/s13321-018-0287-6).



- [99] Qin Liu, JingChun Wang, PengLi Du, LiHong Hu, Xiao Zheng, and GuanHua Chen. “Improving the Performance of Long-Range-Corrected Exchange-Correlation Functional with an Embedded Neural Network”. In: *J. Phys. Chem. A* 121.38 (2017), pp. 7273–7281. DOI: [10.1021/acs.jpca.7b07045](https://doi.org/10.1021/acs.jpca.7b07045).
- [100] Ilya Loshchilov and Frank Hutter. “Decoupled weight decay regularization”. In: *arXiv preprint arXiv:1711.05101* (2017).
- [101] Steph-Yves Louis, Yong Zhao, Alireza Nasiri, et al. “Global Attention based Graph Convolutional Neural Networks for Improved Materials Property Prediction”. In: *arXiv preprint arXiv:2003.13379* (2020).
- [102] Shuaihua Lu, Qionghua Zhou, Yixin Ouyang, Yilv Guo, Qiang Li, and Jinlan Wang. “Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning”. In: *Nat. Commun.* 9.1 (2018), p. 3405. DOI: [10.1038/s41467-018-05761-w](https://doi.org/10.1038/s41467-018-05761-w).
- [103] Keld T. Lundgaard, Jess Wellendorff, Johannes Voss, Karsten W. Jacobsen, and Thomas Bligaard. “mBEEF-vdW: Robust fitting of error estimation density functionals”. In: *Phys. Rev. B* 93.23 (2016). DOI: [10.1103/physrevb.93.235162](https://doi.org/10.1103/physrevb.93.235162).
- [104] Andrew L Maas, Awni Y Hannun, Andrew Y Ng, et al. “Rectifier nonlinearities improve neural network acoustic models”. In: *Proc. icml*. Vol. 30. 1. Citeseer. 2013, p. 3.
- [105] Richard M. Martin. *Electronic Structure: Basic Theory and Practical Methods*. Cambridge University Press, 2008. 650 pp.
- [106] Rocío Mercado, Tobias Rastemo, Edvard Lindelöf, et al. “Graph networks for molecular design”. In: *Mach. learn.: sci. technol.* 2.2 (2021), p. 025023. DOI: [10.1088/2632-2153/abcf91](https://doi.org/10.1088/2632-2153/abcf91).
- [107] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. “Efficient estimation of word representations in vector space”. In: *arXiv preprint arXiv:1301.3781* (2013).
- [108] Hendrik J. Monkhorst and James D. Pack. “Special points for Brillouin-zone integrations”. In: *Phys. Rev. B* 13 (12 1976), pp. 5188–5192. DOI: [10.1103/PhysRevB.13.5188](https://doi.org/10.1103/PhysRevB.13.5188).

- [109] Paula Mori-Sánchez, Qin Wu, and Weitao Yang. “Orbital-dependent correlation energy in density-functional theory based on a second-order perturbation approach: Success and failure”. In: *J. Chem. Phys.* 123.6 (2005), p. 062204. DOI: [10.1063/1.1904584](https://doi.org/10.1063/1.1904584).
- [110] Eliya Nachmani and Lior Wolf. “Molecule property prediction and classification with graph hypernetworks”. In: *arXiv preprint arXiv:2002.00240* (2020).
- [111] Ryo Nagai, Ryosuke Akashi, Shu Sasaki, and Shinji Tsuneyuki. “Neural-network Kohn-Sham exchange-correlation potential and its out-of-training transferability”. In: *J. Chem. Phys.* 148.24 (2018), p. 241737. DOI: [10.1063/1.5029279](https://doi.org/10.1063/1.5029279).
- [112] Ryo Nagai, Ryosuke Akashi, Shu Sasaki, and Shinji Tsuneyuki. “Neural-network Kohn-Sham exchange-correlation potential and its out-of-training transferability”. In: *J. Chem. Phys.* 148.24 (2018), p. 241737. DOI: [10.1063/1.5029279](https://doi.org/10.1063/1.5029279).
- [113] Vinod Nair and Geoffrey E. Hinton. “Rectified Linear Units Improve Restricted Boltzmann Machines”. In: *ICML’10 Proceedings of the 27th International Conference on International Conference on Machine Learning*. Omnipress, 2010, pp. 807–814.
- [114] Jan H. Noordik. *Cheminformatics Developments: History, Reviews and Current Research*. IOS Press, 2004.
- [115] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, et al. “Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis”. In: *Comput. Mater. Sci.* 68 (2013), pp. 314–319. DOI: [10.1016/j.commatsci.2012.10.028](https://doi.org/10.1016/j.commatsci.2012.10.028).
- [116] Corey Oses, Eric Gossett, David V. Hicks, et al. “AFLOW-CHULL: Cloud-Oriented Platform for Autonomous Phase Stability Analysis.” In: *J. Chem. Inf. Model.* 58 12 (2018), pp. 2477–2490. DOI: [10.1021/acs.jcim.8b00393](https://doi.org/10.1021/acs.jcim.8b00393).
- [117] Cheol Woo Park and Chris Wolverton. “Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery”. In: *Phys. Rev. Mater.* 4 (6 2020), p. 063801. DOI: [10.1103/PhysRevMaterials.4.063801](https://doi.org/10.1103/PhysRevMaterials.4.063801).

- [118] Adam Paszke, Sam Gross, Soumith Chintala, et al. “Automatic differentiation in pytorch”. In: *NIPS 2017 Autodiff Workshop: The Future of Gradient-based Machine Learning Software and Techniques*. 2017.
- [119] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. “Generalized Gradient Approximation Made Simple”. In: *Phys. Rev. Lett.* 77 (18 1996), pp. 3865–3868. DOI: [10.1103/PhysRevLett.77.3865](https://doi.org/10.1103/PhysRevLett.77.3865).
- [120] John P. Perdew, Matthias Ernzerhof, and Kieron Burke. “Rationale for mixing exact exchange with density functional approximations”. In: *J. Chem. Phys.* 105.22 (1996), pp. 9982–9985. DOI: [10.1063/1.472933](https://doi.org/10.1063/1.472933).
- [121] John P. Perdew, Adrienn Ruzsinszky, Gábor I. Csonka, et al. “Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces”. In: *Phys. Rev. Lett.* 100 (13 2008), p. 136406. DOI: <https://doi.org/10.1103/PhysRevLett.100.136406>.
- [122] John P. Perdew and Karla Schmidt. “Jacob’s ladder of density functional approximations for the exchange-correlation energy”. In: *AIP Conference Proceedings* 577.1 (2001), pp. 1–20. DOI: [10.1063/1.1390175](https://doi.org/10.1063/1.1390175).
- [123] David Pfau, James S. Spencer, Alexander G. D. G. Matthews, and W. M. C. Foulkes. “Ab initio solution of the many-electron Schrödinger equation with deep neural networks”. In: *Phys. Rev. Res.* 2 (3 2020), p. 033429. DOI: [10.1103/PhysRevResearch.2.033429](https://doi.org/10.1103/PhysRevResearch.2.033429).
- [124] G. Pilania, J.E. Gubernatis, and T. Lookman. “Multi-fidelity machine learning models for accurate bandgap predictions of solids”. In: *Comput. Mater. Sci.* 129 (2017), pp. 156–163. DOI: [10.1016/j.commatsci.2016.12.004](https://doi.org/10.1016/j.commatsci.2016.12.004).
- [125] G. Pilania and X.-Y. Liu. “Machine learning properties of binary wurtzite superlattices”. In: *J. Mater. Sci.* 53.9 (2018), pp. 6652–6664. DOI: [10.1007/s10853-018-1987-z](https://doi.org/10.1007/s10853-018-1987-z).
- [126] G. Pilania, A. Mannodi-Kanakkithodi, B. P. Uberuaga, R. Ramprasad, J. E. Gubernatis, and T. Lookman. “Machine learning bandgaps of double perovskites”. In: *Sci. Rep.* 6.1 (2016), p. 19375. DOI: [10.1038/srep19375](https://doi.org/10.1038/srep19375).
- [127] Sergey N Pozdnyakov and Michele Ceriotti. “Incompleteness of graph convolutional neural networks for points clouds in three dimensions”. In: *arXiv preprint arXiv:2201.07136* (2022).

- [128] Arunkumar Chitteth Rajan, Avanish Mishra, Swanti Satsangi, et al. “Machine-Learning-Assisted Accurate Band Gap Predictions of Functionalized MXene”. In: *Chem. Mater.* 30.12 (2018), pp. 4031–4038. DOI: [10.1021/acs.chemmater.8b00686](https://doi.org/10.1021/acs.chemmater.8b00686).
- [129] Krishna Rajan. “Materials informatics”. In: *Mater. Today* 8.10 (2005), pp. 38–45. DOI: [10.1016/S1369-7021\(05\)71123-8](https://doi.org/10.1016/S1369-7021(05)71123-8).
- [130] Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. “Quantum chemistry structures and properties of 134 kilo molecules”. In: *Sci. Data* 1 (2014), p. 140022. DOI: [10.1038/sdata.2014.22](https://doi.org/10.1038/sdata.2014.22).
- [131] Carsten Rostgaard. “The projector augmented-wave method”. In: *arXiv preprint arXiv:0910.1921* (2009).
- [132] Lars Ruddigkeit, Ruud van Deursen, Lorenz C. Blum, and Jean-Louis Reymond. “Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17”. In: *J. Chem. Inf. Model.* 52.11 (2012). PMID: 23088335, pp. 2864–2875. DOI: [10.1021/ci300415d](https://doi.org/10.1021/ci300415d).
- [133] James E. Saal, Scott Kirklin, Muratahan Aykol, Bryce Meredig, and C. Wolverton. “Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD)”. In: *JOM* 65.11 (2013), pp. 1501–1509. ISSN: 1543-1851. DOI: [10.1007/s11837-013-0755-4](https://doi.org/10.1007/s11837-013-0755-4).
- [134] Alvaro Sanchez-Gonzalez, Victor Bapst, Kyle Cranmer, and Peter Battaglia. “Hamiltonian graph networks with ode integrators”. In: *arXiv preprint arXiv:1909.12790* (2019).
- [135] Rafael Sarmiento-Pérez, Tiago F. T. Cerqueira, Sabine Körbel, Silvana Botti, and Miguel A. L. Marques. “Prediction of Stable Nitride Perovskites”. In: *Chem. Mater.* 27.17 (2015), pp. 5957–5963. DOI: [10.1021/acs.chemmater.5b02026](https://doi.org/10.1021/acs.chemmater.5b02026).
- [136] Jonathan Schmidt, Hai-Chen Wang, Tiago F. T. Cerqueira, Silvana Botti, and Miguel A. L. Marques. “A new dataset of 175k stable and metastable materials calculated with the PBEsol and SCAN functionals”. In: *Materials Cloud* (2021). DOI: <https://archive.materialscloud.org/record/2021.164>.

- [137] Kristof Schütt, Oliver Unke, and Michael Gastegger. “Equivariant message passing for the prediction of tensorial properties and molecular spectra”. In: *International Conference on Machine Learning*. PMLR. 2021, pp. 9377–9388.
- [138] KT Schütt, Michael Gastegger, Alexandre Tkatchenko, K-R Müller, and Reinhard J Maurer. “Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions”. In: *Nat. Commun.* 10.1 (2019), pp. 1–10. DOI: [10.1038/s41467-019-12875-2](https://doi.org/10.1038/s41467-019-12875-2).
- [139] K. T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K. R. Müller, and E. K. U. Gross. “How to represent crystal structures for machine learning: Towards fast prediction of electronic properties”. In: *Phys. Rev. B* 89.20 (2014), p. 205118. DOI: [10.1103/physrevb.89.205118](https://doi.org/10.1103/physrevb.89.205118).
- [140] K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko, and K.-R. Müller. “SchNetPack: A Deep Learning Toolbox For Atomistic Systems”. In: *J. Chem. Theory Comput.* 15.1 (2019), pp. 448–455. DOI: [10.1021/acs.jctc.8b00908](https://doi.org/10.1021/acs.jctc.8b00908).
- [141] K. T. Schütt, H. E. Saucedo, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller. “SchNet – A deep learning architecture for molecules and materials”. In: *J. Chem. Phys.* 148.24 (2018), p. 241722. DOI: [10.1063/1.5019779](https://doi.org/10.1063/1.5019779).
- [142] Wahyu Setyawan, Romain M. Gaume, Stephanie Lam, Robert S. Feigelson, and Stefano Curtarolo. “High-Throughput Combinatorial Database of Electronic Band Structures for Inorganic Scintillator Materials”. In: *ACS Comb. Sci.* 13.4 (2011), pp. 382–390. DOI: [10.1021/co200012w](https://doi.org/10.1021/co200012w).
- [143] Zeren Shui and George Karypis. “Heterogeneous Molecular Graph Neural Networks for Predicting Molecule Properties”. In: *2020 IEEE International Conference on Data Mining (ICDM)*. 2020, pp. 492–500. DOI: [10.1109/ICDM50108.2020.00058](https://doi.org/10.1109/ICDM50108.2020.00058).
- [144] David Silver, Aja Huang, Chris J. Maddison, et al. “Mastering the game of Go with deep neural networks and tree search”. In: *Nature* 529.7587 (2016), pp. 484–489. DOI: [10.1038/nature16961](https://doi.org/10.1038/nature16961).

- [145] John C. Snyder, Matthias Rupp, Katja Hansen, Leo Blooston, Klaus-Robert Müller, and Kieron Burke. “Orbital-free bond breaking via machine learning”. In: *J. Chem. Phys.* 139.22 (2013), p. 224104. DOI: [10.1063/1.4834075](https://doi.org/10.1063/1.4834075).
- [146] John C. Snyder, Matthias Rupp, Katja Hansen, Klaus-Robert Müller, and Kieron Burke. “Finding Density Functionals with Machine Learning”. In: *Phys. Rev. Lett.* 108.25 (2012), p. 253002. DOI: [10.1103/physrevlett.108.253002](https://doi.org/10.1103/physrevlett.108.253002).
- [147] John C. Snyder, Matthias Rupp, Klaus-Robert Müller, and Kieron Burke. “Nonlinear gradient denoising: Finding accurate extrema from inaccurate functional derivatives”. In: *Int. J. Quantum Chem.* 115.16 (2015), pp. 1102–1114. DOI: [10.1002/qua.24937](https://doi.org/10.1002/qua.24937).
- [148] Taylor Sparks, Steven Kauwe, and Taylor Welker. “Extracting Knowledge from DFT: Experimental Band Gap Predictions Through Ensemble Learning”. In: *ChemRxiv preprint* (2018). DOI: [10.26434/chemrxiv.7236029.v1](https://doi.org/10.26434/chemrxiv.7236029.v1).
- [149] James S Spencer, David Pfau, Aleksandar Botev, and W Matthew C Foulkes. “Better, faster fermionic neural networks”. In: *arXiv preprint arXiv:2011.07125* (2020).
- [150] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. “Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields”. In: *J. Phys. Chem.* 98.45 (1994), pp. 11623–11627. DOI: [10.1021/j100096a001](https://doi.org/10.1021/j100096a001).
- [151] Jianwei Sun, Richard C Remsing, Yubo Zhang, et al. “SCAN: An efficient density functional yielding accurate structures and energies of diversely-bonded materials”. In: *arXiv preprint arXiv:1511.01089* (2015).
- [152] Jianwei Sun, Adrienn Ruzsinszky, and John P Perdew. “Strongly constrained and appropriately normed semilocal density functional”. In: *Phys. Rev. Lett.* 115.3 (2015), p. 036402.
- [153] Qiming Sun, Timothy C Berkelbach, Nick S Blunt, et al. “PySCF: the Python-based simulations of chemistry framework”. In: *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 8.1 (2018), e1340. DOI: [10.1002/wcms.1340](https://doi.org/10.1002/wcms.1340).

- [154] Nicolas Tancogne-Dejean, Micael J. T. Oliveira, Xavier Andrade, et al. “Octopus, a computational framework for exploring light-driven phenomena and quantum dynamics in extended and finite systems”. In: *J. Chem. Phys.* 152.12 (2020), p. 124119. DOI: [10.1063/1.5142502](https://doi.org/10.1063/1.5142502).
- [155] Nathaniel Thomas, Tess Smidt, Steven Kearnes, et al. “Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds”. In: *arXiv preprint arXiv:1802.08219* (2018).
- [156] Robert Tibshirani. “Regression Shrinkage and Selection via the Lasso”. In: *J. Roy. Statist. Soc. Ser. B* (1996). DOI: [10.1111/j.2517-6161.1996.tb02080.x](https://doi.org/10.1111/j.2517-6161.1996.tb02080.x).
- [157] David J. Tozer, Victoria E. Ingamells, and Nicholas C. Handy. “Exchange-correlation potentials”. In: *J. Chem. Phys.* 105.20 (1996), pp. 9200–9213. DOI: [10.1063/1.472753](https://doi.org/10.1063/1.472753).
- [158] F. Tran and P. Blaha. “Accurate Band Gaps of Semiconductors and Insulators with a Semilocal Exchange-Correlation Potential”. In: *Phys. Rev. Lett.* 102.22 (2009), p. 226401. ISSN: 0031-9007, 1079-7114. DOI: [10.1103/physrevlett.102.226401](https://doi.org/10.1103/physrevlett.102.226401).
- [159] Vahe Tshitoyan, John Dagdelen, Leigh Weston, et al. “Unsupervised word embeddings capture latent knowledge from materials science literature”. In: *Nature* 571.7763 (2019), pp. 95–98. DOI: [10.1038/s41586-019-1335-8](https://doi.org/10.1038/s41586-019-1335-8).
- [160] Yutaka Tsujimoto, Satoru Hiwa, Yushi Nakamura, Yohei Oe, and Tomoyuki Hiroyasu. “L-MolGAN: An improved implicit generative model for large molecular graphs”. In: *ChemRxiv*. (2021). DOI: [10.26434/chemrxiv.14569545.v3](https://doi.org/10.26434/chemrxiv.14569545.v3).
- [161] David Vanderbilt. “Soft self-consistent pseudopotentials in a generalized eigenvalue formalism”. In: *Phys. Rev. B* 41 (11 1990), pp. 7892–7895. DOI: [10.1103/PhysRevB.41.7892](https://doi.org/10.1103/PhysRevB.41.7892).
- [162] Ashish Vaswani, Noam Shazeer, Niki Parmar, et al. “Attention is all you need”. In: *Adv Neural Inform Process Syst.* 2017, pp. 5998–6008. DOI: [10.5555/3295222.3295349](https://doi.org/10.5555/3295222.3295349).

- [163] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. “Graph attention networks”. In: *arXiv preprint arXiv:1710.10903* (2017).
- [164] Oriol Vinyals, Igor Babuschkin, Wojciech M. Czarnecki, et al. “Grandmaster level in StarCraft II using multi-agent reinforcement learning”. In: 575.7782 (2019), pp. 350–354. DOI: [10.1038/s41586-019-1724-z](https://doi.org/10.1038/s41586-019-1724-z).
- [165] Kevin Vu, John C. Snyder, Li Li, et al. “Understanding kernel ridge regression: Common behaviors from simple functions to density functionals”. In: *Int. J. Quantum Chem.* 115.16 (2015), pp. 1115–1128. DOI: [10.1002/qua.24939](https://doi.org/10.1002/qua.24939).
- [166] Hai-Chen Wang, Silvana Botti, and Miguel A. L. Marques. “Predicting stable crystalline compounds using chemical similarity”. In: *npj Comput. Mater.* 7.1 (2021). DOI: [10.1038/s41524-020-00481-6](https://doi.org/10.1038/s41524-020-00481-6).
- [167] Logan Ward, Ruoqian Liu, Amar Krishna, et al. “Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations”. In: *Phys. Rev. B* 96.2 (2017), p. 024104. DOI: [10.1103/physrevb.96.024104](https://doi.org/10.1103/physrevb.96.024104).
- [168] Jess Wellendorff, Keld T. Lundgaard, Andreas Møgelhøj, et al. “Density functionals for surface science: Exchange-correlation model development with Bayesian error estimation”. In: *Phys. Rev. B* 85.23 (2012). DOI: [10.1103/physrevb.85.235149](https://doi.org/10.1103/physrevb.85.235149).
- [169] L. Weston and C. Stampfl. “Machine learning the band gap properties of kesterite  $I_2-II-IV-V_4$  quaternary compounds for photovoltaics applications”. In: *Phys. Rev. Mater.* 2.8 (2018), p. 085407. DOI: [10.1103/physrevmaterials.2.085407](https://doi.org/10.1103/physrevmaterials.2.085407).
- [170] Oliver Wieder, Stefan Kohlbacher, Méline Kuenemann, et al. “A compact review of molecular property prediction with graph neural networks”. In: *Drug Discovery Today: Technologies* 37 (2020), pp. 1–12. DOI: [10.1016/j.ddtec.2020.11.009](https://doi.org/10.1016/j.ddtec.2020.11.009).
- [171] Max Wilson, Saverio Moroni, Markus Holzmann, et al. “Wave function Ansatz (but Periodic) Networks and the Homogeneous Electron Gas”. In: *arXiv preprint arXiv:2202.04622* (2022).



- [172] Tian Xie and Jeffrey C. Grossman. “Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties”. In: *Phys. Rev. Lett.* 120.14 (2018), p. 145301. DOI: [10.1103/physrevlett.120.145301](https://doi.org/10.1103/physrevlett.120.145301).
- [173] Qimen Xu, Abhiraj Sharma, Benjamin Comer, et al. “SPARC: Simulation package for ab-initio real-space calculations”. In: *SoftwareX* 15 (2021), p. 100709. DOI: [10.1016/j.softx.2021.100709](https://doi.org/10.1016/j.softx.2021.100709).
- [174] Weike Ye, Chi Chen, Zhenbin Wang, Iek-Heng Chu, and Shyue Ping Ong. “Deep neural networks for accurate predictions of crystal stability”. In: *Nat. Commun.* 9.1 (2018), p. 3800. DOI: [10.1038/s41467-018-06322-x](https://doi.org/10.1038/s41467-018-06322-x).
- [175] Jiaxuan You, Bowen Liu, Zhitao Ying, Vijay Pande, and Jure Leskovec. “Graph Convolutional Policy Network for Goal-Directed Molecular Graph Generation”. In: *Advances in Neural Information Processing Systems* 31 (2018).
- [176] Bo Zhang, Xin-Qi Zheng, Tong-Yun Zhao, Feng-Xia Hu, Ji-Rong Sun, and Bao-Gen Shen. “Machine learning technique for prediction of magnetocaloric effect in  $\text{La}(\text{Fe},\text{Si}/\text{Al})_{13}$ -based materials”. In: *Chin. Phys. B* 27.6 (2018), p. 067503. DOI: [10.1088/1674-1056/27/6/067503](https://doi.org/10.1088/1674-1056/27/6/067503).
- [177] Guo-Xu Zhang, Anthony M Reilly, Alexandre Tkatchenko, and Matthias Scheffler. “Performance of various density-functional approximations for cohesive properties of 64 bulk solids”. In: *New J. Phys.* 20.6 (2018), p. 063020. DOI: [10.1088/1367-2630/aac7f0](https://doi.org/10.1088/1367-2630/aac7f0).
- [178] Yubo Zhang, Daniil A. Kitchaev, Julia Yang, et al. “Efficient first-principles prediction of solid stability: Towards chemical accuracy”. In: *npj Comput. Mater.* 4.1 (2018). DOI: [10.1038/s41524-018-0065-z](https://doi.org/10.1038/s41524-018-0065-z).
- [179] Xiaolong Zheng, Peng Zheng, and Rui-Zhi Zhang. “Machine learning material properties from the periodic table using convolutional neural networks”. In: *Chem. Sci.* 9.44 (2018), pp. 8426–8432. DOI: [10.1039/c8sc02648c](https://doi.org/10.1039/c8sc02648c).
- [180] Xiaolong Zheng, Peng Zheng, and Rui-Zhi Zhang. “Machine learning material properties from the periodic table using convolutional neural networks”. In: *Chem. Sci.* 9 (44 2018), pp. 8426–8432. DOI: [10.1039/C8SC02648C](https://doi.org/10.1039/C8SC02648C).

- [181] Xiaolong Zheng, Peng Zheng, Liang Zheng, Yang Zhang, and Rui-Zhi Zhang. “Multi-channel convolutional neural networks for materials properties prediction”. In: *Comput. Mater. Sci.* 173 (2020), p. 109436. ISSN: 0927-0256. DOI: [10.1016/j.commatsci.2019.109436](https://doi.org/10.1016/j.commatsci.2019.109436).
- [182] Yi Zhou, Jiang Wu, Shuguang Chen, and GuanHua Chen. “Toward the Exact Exchange–Correlation Potential: A Three-Dimensional Convolutional Neural Network Construct”. In: *J. Phys. Chem. Lett* 10.22 (2019), pp. 7264–7269. DOI: [10.1021/acs.jpcllett.9b02838](https://doi.org/10.1021/acs.jpcllett.9b02838).
- [183] Ya Zhuo, Aria Mansouri Tehrani, and Jakoah Brgoch. “Predicting the Band Gaps of Inorganic Solids by Machine Learning”. In: *J. Phys. Chem. Lett.* 9.7 (2018), pp. 1668–1673. DOI: [10.1021/acs.jpcllett.8b00124](https://doi.org/10.1021/acs.jpcllett.8b00124).



## Acknowledgement

Completing the work for this thesis and writing it would never have been possible without the help of the people who accompanied and supported me through the last years. First, I would like to express my deep gratitude to Professor Marques for the amazing opportunity to work in his group throughout the last years. He introduced me to the topic of machine learning, gave me an uncountable number of opportunities and shared a great amount of knowledge and support not only concerning science but also life in general with me.

I would like to thank the permanent and previous members of the group Haichen, Ahmad, Thomas, Mario and Carlos who were a constant source of support, serious, as well as fun discussion, and where always available to answer any question no matter how trivial.

Furthermore, I would like to thank Dominik, Marius and Michael from our neighboring offices for all the fun discussion over lunch and coffee breaks. Then, I would like to thank all the students Mattheo, Paul, Johannes, Noah, Jacob and Martin who I had the pleasure to work with during my PhD-thesis and through whom I learned a lot. Next, I would like to thank Katja for helping us navigate the bureaucratic jungle of the university.

Finally, I want to thank my friends, my parents, my brother Benjamin, and my girlfriend Sarah whose support and encouragement made all this possible and the whole experience a lot more fun.

## **Eidstaatliche Erklärung**

Hiermit erkläre ich, Jonathan Schmidt, gemäß §5 der Promotionsordnung der Naturwissenschaftlichen Fakultäten I, II und III der Martin-Luther-Universität Halle-Wittenberg vom 13.07.2016, dass ich die vorliegende Arbeit "Machine Learning the Thermodynamic Stability of Crystal structures" selbstständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Ich erkläre die Angaben wahrheitsgemäß gemacht, keine vergeblichen Promotionsversuche unternommen und keine Dissertation an einer anderen wissenschaftlichen Einrichtung zur Erlangung eines akademischen Grades eingereicht zu haben. Ich bin weder vorbestraft noch sind gegen mich Ermittlungsverfahren anhängig.

Halle (Saale), den 03.05.2022

# Jonathan Schmidt

Google Scholar: <https://scholar.google.com/citations?user=OG1vKocAAAAJ&hl=en>

Born: October 30, 1995—Aachen, Germany  
Nationality: German

## Current position

Researcher, Martin-Luther Universität Halle-Wittenberg, Halle

## Education

- since Dec 2018 PhD-student at Martin-Luther-Universität Halle-Wittenberg with Prof. Miguel Marques
- 2016-2018 Physics (Master - 120 LP) at Martin-Luther-Universität Halle-Wittenberg  
Degree: Master of Science in physics  
Master thesis: Reduced Density Matrix Functional Theory for Superconductors
- 2013-2016 Physics (Bachelor - 180 LP) at Martin-Luther-Universität Halle-Wittenberg  
Degree: Bachelor of science in physics  
Bachelor thesis: Machine Learning Prediction of the Stability of Perovskites
- 2009 – 2013 Georg-Cantor Gymnasium Halle Degree: general qualification for university entrance
- 2006-2009 Deutschhaus-Gymnasium Würzburg