Quantifying transfer learning synergies in infinite-layer and perovskite nitrides, oxides, and fluorides

Armin Sahinovic and Benjamin Geisler*

Department of Physics, Universität Duisburg-Essen, Lotharstr. 1, 47057 Duisburg, Germany

E-mail: benjamin.geisler@uni-due.de

Received 31 December 2021, revised 28 February 2022
Accepted for publication 1 March 2022
Published 21 March 2022

Abstract
We combine density functional theory simulations and active learning (AL) of element-embedding neural networks (NNs) to explore the sample efficiency for the prediction of vacancy layer formation energies and lattice parameters in $ABX_n$ infinite-layer ($n = 2$) versus perovskite ($n = 3$) nitrides, oxides, and fluorides in the spirit of transfer learning. Following a comprehensive data analysis from different thermodynamic, structural, and statistical perspectives, we show that NNs model these observables with high precision, using merely $\sim 30\%$ of the data for training and exclusively the $A$-, $B$-, and $X$-site element names as minimal input devoid of any physical \textit{a priori} information. Element embedding autonomously arranges the chemical elements with a characteristic recurrent topology, such that their relations are consistent with human knowledge. We compare two different embedding strategies and show that these techniques render additional input such as atomic properties negligible. Simultaneously, we demonstrate that AL is largely independent of the initial training set, and exemplify its superiority over randomly composed training sets. Despite their highly distinct chemistry, the present approach successfully identifies fundamental quantum-mechanical universalities between nitrides, oxides, and fluorides that enhance the combined prediction accuracy by up to 16\% with respect to three specialized NNs at equivalent numerical effort. This quantification of synergistic effects provides an impression of the transfer learning improvements one may expect for similarly complex materials. Finally, by embedding the tensor product of the $B$ and $X$ sites and subsequent quantitative cluster analysis, we establish from an unbiased artificial-intelligence perspective that oxides and nitrides exhibit significant parallels, whereas fluorides constitute a rather distinct materials class.

Keywords: artificial intelligence, transfer learning, active learning, neural networks, perovskites, infinite-layer, density functional theory

(Some figures may appear in colour only in the online journal)

1. Introduction
Artificial intelligence (AI) algorithms receive increasing attention in computational materials science. Following a decade of high-throughput materials discovery [1–5], which was paralleled by the emergence of different materials databases [6–9], machine learning [10–19] and, more specifically, deep learning techniques [20–23] open the perspective towards intriguing and often unconventional approaches for the identification of novel and possibly exotic stable and metastable materials with optimized properties.

The overall success story of AI is so far dominated by supervised methods, e.g., for the understanding and prediction of chemical trends and physical properties from materials data
However, these techniques are usually highly data expensive, and their accuracy is detrimentally affected by errors in the data. Semi- and unsupervised approaches [26] such as active learning (AL) [11, 27] proved to be more robust and efficient strategies [28]. A further concept that will become increasingly relevant is transfer learning [22], where the AI knowledge gained while solving one problem is exploited in addressing a different, yet related problem, resulting in a boosted sample efficiency and higher accuracy. In the context of materials physics, this requires the existence of underlying universal systematics on the atomic scale, which are generally nontrivial, as quantum mechanics entangles the impact of the constituent chemical elements on the physical properties of a given compound far beyond simple superposition of atomic properties. Nevertheless, the identification of such hidden patterns without external a priori information is one of the key strengths of AI.

In this comprehensive paper, we explore this aspect by complementing our first-principles data on infinite-layer (IL, \(ABX_3\)) and perovskite (P, \(ABX_3\)) oxides [28] by the respective nitrides and fluorides (figure 1). The IL oxides, which are experimentally derived from the P phase by topotactic reduction of the apical X anions, constitute a highly topical materials class [29–49], and contrasting them with their nitride and fluoride analogues is of fundamental interest. Specifically, fluoride perovskites are an important sub-class of the technologically relevant halides [50], and stable nitride perovskites have been predicted just recently [51]. On the other hand, this represents an extremely challenging extension of the already complex transition metal oxides: the strong differences in formal anion oxidation states, Pauling electronegativities (\(\chi_N = 3.04\), \(\chi_O = 3.44\), and \(\chi_F = 3.98\)), and bond dissociation energies of the respective diatomic molecules lead to a highly distinct chemistry in these three materials classes, and unraveling their universalities and synergies by AI in the spirit of transfer learning must be considered a nontrivial task.

We begin with a detailed data exploration for the three materials classes, comparing the relative thermodynamic stability of the P vs the IL phases, quantifying their structural response to the reduction reaction, and discussing chemical stability trends and statistical correlations. Subsequently, we demonstrate that element-embedding neural networks (NNs) are capable of modeling the formation energies of \(VN, V_O,\) and \(VF\) vacancy layers as well as the IL and P lattice parameters with high accuracy, despite their complexity; these observables act as a fingerprint of the reduction reaction, encoding the quantum-mechanical essence of symmetry breaking, four-fold versus six-fold B-site coordination, the redistribution of electrons released by the anion vacancies, the different orbital sequence due to changes of the crystal field, and the modified hybridization. Interestingly, it turns out to be sufficient to provide solely the A-, B-, and X-site element names as input to the NNs. This highly reduced and abstract input is devoid of any physical a priori knowledge. Subsequently, element embedding [24, 28, 52] automatically arranges the chemical elements with a characteristic topology, such that their relations mirror the conventional picture of the periodic table. We compare two different embedding input strategies, and show in how far complementing this approach by atomic properties (atomic radii and electronegativities) of the constituting elements influences the prediction accuracy. Detailed tracking of the prediction accuracy during the iterative AL process shows that a relative error of <1% is achieved by using merely ~30% of the data for NN training. Simultaneously, we demonstrate that AL is largely independent of the initial training set and therefore highly reliable, and highlight its superiority over randomly composed training sets, which becomes particularly evident for the intricate structural anisotropy of the IL phase. We also provide insight into how AL iteratively composes the training sets, which reflects the varying complexity of the considered observables between the different materials classes. By systematically comparing the performance of three specialized NNs trained exclusively on the nitride, oxide, and fluoride data versus a single NN trained on the combined data, we show that AI indeed identifies latent universalities that improve the prediction accuracy, i.e., enhance the sample efficiency. These synergistic effects are found to be particularly pronounced for lattice parameter prediction, with an accuracy improvement up to 16% at equivalent numerical effort, and most relevant for small training set sizes. Finally, we reverse the perspective and exemplify how the unconventional combination of \(BX\) tensor product embedding and subsequent cluster analysis allows to quantify abstract meta-level terms such as the similarity of entire materials classes. This shows that fluorides constitute a rather distinct materials class, whereas oxides and nitrides bear higher similarities.

Figure 1. Reduction of the apical X anions in the P structure \((n = 3)\) results in the emergence of the anisotropic IL structure \((n = 2)\). This reaction is associated with the formation energy \(E^{f}\). In the periodic table, the 69 chemical elements combined here at the A and B sites are marked (dark grey), together with the 3 anion elements investigated at the X site (purple).
2. Methodology

We performed first-principles simulations in the framework of density functional theory [53] (DFT) as implemented in the Vienna ab initio simulation package [54, 55] using the generalized gradient approximation as parameterized by Perdew, Burke, and Ernzerhof [56] to construct a database of ground-state energies and optimized lattice parameters for 4692 combinations of different elements at the A and B sites (figure 1) for both the P and the IL geometry, which were modeled by using cubic [17] and tetragonal [34, 36] unit cells, respectively. Exploring $X = N, O,$ and $F$ at the anion sites, this results in 14076 combinations and 28152 simulated compounds in total. Accordingly, we refer to the individual data sets as N, O, and F, whereas NOF = N ∪ O ∪ F denotes the combined data set. Consistent with previous work [28], we largely adopted the DFT + $U$ standards [57] of the Materials Project database [6, 58, 59] and used the respective ground-state crystal structures to construct our elemental bulk references. The NNs were realized in Keras/Tensorflow 2 [60, 61], and the AL algorithm was developed in Python 3. The formation energies of the $V_X$ vacancy layers are determined from DFT ground-state energies via $E_{i}^{V_X} = E(ABX_3) - E(ABX_2) + \mu_X$, where $\mu_X = 4E(X_2)$ models the N-, O-, and F-rich limit. We employ an anion energy correction to mitigate the well-known DFT overbinding of gas-phase $N_2$, $O_2$, and $F_2$ molecules, reproducing the experimental binding energies of 9.79, 5.15, and 1.63 eV, respectively [62–64]. The heats of formation of the P phase from the constituent bulk elements read $E_{i}^P = E(ABX_3) - E(A \text{ bulk}) - E(B \text{ bulk}) - 3\mu_X$. An analogous quantity $E_{i}^{IL}$ can be defined for the IL phase. All energies are given per formula unit.

3. Data exploration: nitrides and oxides versus fluorides

We start our discussion with an overview of the three data sets from a thermodynamic, a structural, and a statistical perspective, identifying similarities and differences between $ABX_n$ nitrides, oxides, and fluorides. Simultaneously,
this provides an impression of the data complexity, and what challenges this poses for subsequent modeling with AI methods.

3.1. Thermodynamic and structural perspective

Figure 2(a) displays the N, O, and F data sets in three $E^X_I$ versus $E^P_I$ phase diagrams, comparing the relative stability of the IL and the P structure as well as their stability with respect to the constituent bulk elements. This perspective at the data is motivated by recent work on IL oxides, specifically superconducting nickelates [29, 35, 41, 45, 49], which are initially stabilized as P films on SrTiO$_3$(001) via heteroepitaxy, followed by a topotactic reduction of the apical X ions. Overall, $E^X_I$ ranges from $-12$ to $+8$ eV, while $E^P_I$ covers $-40$ eV. The individual N, O, and F data clouds exhibit linear trends, i.e., a correlation of the P stability and the corresponding reduction energy to the IL phase. However, the data spreads broadly around the three regression lines:

$$E^X_I = -0.33 E^P_I - 2.32 \text{ eV},$$
$$E^{YO}_I = -0.36 E^P_I - 1.37 \text{ eV},$$
$$E^{YF}_I = -0.32 E^P_I + 0.24 \text{ eV}.$$

Moreover, we observe that the three data clouds are concentrated in different sectors of the phase diagrams, the nitrides (fluorides) occupying predominantly the bottom-right (top-left) corner. The oxides are less compact than the fluorides and appear more balanced around the origin. The circular charts in figure 2(b) quantify the distribution of the compounds across the three stability sectors considered here. Most of the nitrides (68.1%) are labeled as thermodynamically unstable; 31% are located in the IL sector, and only 0.7% are labeled as P. In contrast, 66% of the oxides are located in the P sector, and even 92% of the fluorides. This can be understood from the highly different X$_2$ dissociation energies, the N–N triple bond being much more difficult to break than the O–O and F–F bonds (see above), and the much higher reactivity of F. Moreover, N requires too high cation oxidation states; we will explore this aspect in more detail below. Hence, the oxides emerge as the most balanced of the three materials classes, which is interesting since the P vs IL classification also reflects preferences with respect to the six-fold vs four-fold B-site coordination. Notably, these considerations do not directly imply statements on the absolute stability of the different phases, which require a convex-hull analysis [17, 51]. However, they highlight first fundamental differences between the three materials classes, and some metastable compounds may be accessible via heteroepitaxy.

The structural analysis provided in figure 2(c) confirms our earlier finding for oxides [28] that most materials exhibit the tendency to contract vertically upon reduction with respect to their cubic phase (the $\Delta a_0/a_0^H$ ratios can drop below 0.5, particularly for some nitrides), expanding simultaneously in the plane (up to 20%) with reduced volume. This applies in particular to those materials that feature a low $E^X_I$ [see insets in figure 2(c)]. For some very stable and therefore difficult to reduce compounds, the structural changes are rather modest, as one can observe from the $E^X_I$ maxima that are consistently located close to the centers of the panels. In sharp contrast, some materials expand massively in apical direction for nitrides and oxides, but simultaneously feature more pronounced basal contraction (i.e., extend further to the left of the panel). The nitrides and oxides listed in table 1 are stable as P and IL phases, some metastable compounds may be accessible via heteroepitaxy.

Table 1. Energies and lattice parameters for a selection of ABX$_n$ nitrides, oxides, and fluorides (cf figure 2). $E^X_I < 0$ indicates that the tetragonal IL structure ($n = 2$) is preferred over the cubic P phase ($n = 3$). All listed materials are stable as P and IL phases with respect to A–B cation interchange; the associated energy is reflected exemplarily for the P phase by $\Delta E^P_I = E^P_I(\text{BAX}_n) - E^P_I(\text{ABX}_n)$.

| System     | $E^X_I$ (eV) | $E^P_I$ (eV) | $\Delta E^P_I$ (eV) | $a_0^H$ (Å) | $a_0^H$ (Å) | $c_0^H$ (Å) |
|------------|--------------|--------------|---------------------|-------------|-------------|-------------|
| LaTaN$_n$  | 0.3          | -6.2         | 13.8                | 4.11        | 4.15        | 3.50        |
| LaWN$_n$   | 1.1          | -5.8         | 16.1                | 4.04        | 4.09        | 3.54        |
| LaReN$_n$  | 0.4          | -4.6         | 16.0                | 3.97        | 3.99        | 3.58        |
| YReN$_n$   | -0.5         | -3.6         | 14.1                | 3.92        | 3.96        | 3.27        |
| LaNiO$_n$  | 2.8          | -13.2        | 8.3                 | 3.82        | 3.93        | 3.40        |
| PrNiO$_n$  | 2.7          | -12.8        | 8.3                 | 3.80        | 3.91        | 3.35        |
| NdNiO$_n$  | 2.7          | -12.8        | 8.2                 | 3.79        | 3.90        | 3.30        |
| CaCuO$_n$  | -0.01        | -8.5         | 6.7                 | 3.80        | 3.87        | 3.21        |
| SrCuO$_n$  | 0.44         | -8.7         | 7.9                 | 3.89        | 3.95        | 3.49        |
| LaLiO$_n$  | -0.5         | -11.7        | 2.1                 | 3.80        | 4.01        | 3.26        |
| BaHFO$_n$  | 6.8          | -18.9        | 16.7                | 4.20        | 4.21        | 4.05        |
| SrTiO$_n$  | 5.8          | -17.7        | 14.1                | 3.94        | 3.99        | 3.59        |
| LaAlO$_n$  | 6.7          | -18.6        | 10.8                | 3.81        | 3.83        | 3.67        |
| LaScO$_n$  | 5.6          | -18.7        | 7.8                 | 4.07        | 4.15        | 3.60        |
| NaMgF$_n$  | 6.8          | -17.1        | 4.7                 | 3.95        | 4.02        | 3.84        |
| K MgF$_n$  | 7.1          | -17.6        | 7.9                 | 4.05        | 4.12        | 4.26        |
| RbMgF$_n$  | 7.1          | -17.4        | 8.9                 | 4.13        | 4.10        | 4.69        |
| CaCaF$_n$  | 6.8          | -18.6        | 7.3                 | 4.58        | 4.57        | 4.84        |
| SrLiF$_n$  | 6.9          | -18.6        | 2.3                 | 3.88        | 3.92        | 3.90        |
| BaLiF$_n$  | 6.8          | -18.7        | 3.2                 | 4.05        | 4.01        | 4.16        |
| KLiF$_n$   | 0.6          | -12.4        | 3.1                 | 4.03        | 4.01        | 3.67        |
| NaMnF$_n$  | 5.3          | -12.7        | 7.8                 | 3.99        | 4.11        | 2.74        |
| KNiF$_n$   | 3.9          | -12.1        | 7.8                 | 4.01        | 4.05        | 3.74        |
| RbZnF$_n$  | 5.5          | -13.9        | 7.9                 | 4.19        | 4.49        | 3.83        |
| NiLaF$_n$  | 5.2          | -13.2        | 1.7                 | 4.68        | 4.55        | 4.54        |
| ZnLaF$_n$  | 6.1          | -16.4        | 4.5                 | 4.74        | 4.63        | 4.78        |
| HeLaF$_n$  | 6.1          | -17.2        | 7.6                 | 4.75        | 4.63        | 4.95        |
data cloud is formed by the $A^{1+}$ (Ta/W/Re)N$_n$ families, where $A^{1+}$ denotes a group-3 element including the rare-earth metals. This is in line with earlier work that predicted LaReN$_3$, LaW$_N$$_3$, and YReN$_3$ to feature (distorted) perovskite structures as ground state [51], albeit no perovskite ground state was found for LaTa$_N$$_3$. All three families are located near the boundary between the P and IL regimes. The oxide panel highlights the formally $d^0$ IL nickelates and cuprates [36, 40, 46] next to some well-known and highly stable perovskites. The nickelates appear as a compact family in the phase diagram, exhibiting a stable P phase, but being simultaneously close to the IL regime. In contrast, the cuprate family extends widely over the IL region, which reflects the naturally preferred four-coordinate plaquette structure that is a common structural motif in high- $T_C$ cuprate superconductors. In the IL region, the $A^{1+}$ Li$_2$O$_2$ compounds emerge as stable, strongly anisotropic, and insulating materials [28]. For the fluorides, which constitute a sub-class of the technologically important halides [50], we observe a series of well-known P materials in the top-left (i.e., most stable) part of the data cloud: for instance, the $A^{1+}$ (Mg/Ca/Sr)F$_2$ families, where $A^{1+}$ denotes an alkali metal (e.g., neighborite NaMgF$_3$ and parascandolite KMgF$_3$). NaMgF$_3$ is of specific interest due to its role as a low-pressure analog for phase transitions in bridgmanite MgSiO$_3$ [65], which is abundant in the lower part of the earth’s mantle. Of comparable stability are members of the $A^{1+}$ LiF$_n$ family, where $A^{1+}$ denotes an alkaline earth. Characteristic of these materials, often referred to as ‘inverse’ perovskites (not to be confused with antiperovskites), is that the A-site cation has a higher oxidation state than the B-site cation, an example being BaLiF$_3$. In the same part of the phase diagram, we also find the (Zn/Cd/Hg)RF$_n$ families, where $R$ denotes a group-3 element, particularly a rare-earth metal [2]; we will further comment on these compounds below. Near the boundary to the IL regime, we observe the $A^{1+}$ LiF$_n$ family, which is located at a similar position in the phase diagram as the $A^{1+}$ LiO$_n$ oxides.

Table 1 summarizes the DFT-predicted energies and lattice parameters for a selection of nitrides, oxides, and fluorides. As we pointed out earlier [28], the oxide lattice parameters are in close agreement with the experimental and theoretical literature; for instance, NdNiO$_2$ ($\alpha^\text{DFT} = 3.92$, $c^\text{DFT} = 3.28$ Å [30, 66]), PrNiO$_2$ (3.92, 3.30 Å [46]), LaNiO$_2$ (3.96, 3.37 Å [34, 67]), SrCuO$_2$ (3.93, 3.43 Å [68, 69]), CaCuO$_2$ (3.85, 3.18 Å [68, 69]); cubic SrTiO$_3$ ($\alpha^\text{DFT} = 3.94$ Å [70, 71]); and pseudocubic LaNiO$_3$ ($\alpha^\text{DFT} = 3.84$ Å [72, 73]). Our results for nitrides agree perfectly with predictions for pseudocubic LaWN$_3$ ($\alpha^\text{DFT} = 4.04$ Å [51]) and LaReN$_3$ ($\alpha^\text{DFT} = 3.97$ Å [51]). The fluoride perovskite KMgF$_3$ crystallizes in the cubic $Pm\bar{3}m$ structure [50] with $\alpha^\text{DFT} = 4.00$ Å [74], and also NaMgF$_3$ in its high-temperature cubic phase agrees nicely with our results ($\alpha^\text{DFT} = 3.955$ Å [75]), whereas the lattice parameters in the tilted orthorhombic phase are naturally smaller ($\alpha^\text{DFT} \sim 3.834$ Å [76]). The layer reduction energies $\Delta E_{\text{VR}}^3$ can be related to vacancy formation energies. For instance, in SrTiO$_3$, the oxygen vacancy formation energy amounts to $\sim$5.5 eV according to first-principles calculations [77]. In LaAlO$_3$, this energy is even higher (6.9 eV [78]), whereas the values in nickelates and their heterostructures [40, 63] are much lower ($\sim$2.8 eV in bulk LaNiO$_3$ [62]). These theory results align perfectly with table 1. In addition, the energy differences $\Delta E_{\text{VR}}^3 = E_{\text{VR}}^3(BAX_n) - E_{\text{VR}}^3(ABX_n)$ provide an impression of how stable the listed materials are with respect to A–B cation interchange, exemplarily given for the P phase. For instance, the ‘inverse’ perovskite BaLiF$_3$ is more stable than its non-inverse analogue LiBaF$_3$ by 3.2 eV per formula unit.

3.2. Stability trends and statistical correlations

Next, we explore the stability trends of different A–B cation combinations, and particularly how they are modified by varying the X-site anions. In order to highlight the key aspects, we focus on the P phase. The matrices in figure 3 show the averaged P heats of formation,
Assuming the standard oxidation state of 3 for nitrogen, this would imply a very high oxidation state of 15 for group-3 elements at the B site. Since the properties of oxide perovskites are well known, it is interesting to compare these results with the nitrides. As we discussed above, nitride perovskites are rather similar materials classes, whereas fluorides are distinct.

4. Active learning of element-embedding neural networks

We showed recently that it is possible to arrive at the insights presented so far without explicitly calculating the entire data set, but only a fraction of it, even for such complex materials classes as transition metal oxides [28]. In this context, element-embedding artificial NNs emerged as efficient means to predict the observables of interest for all materials in a parameter space on the basis of just a subset of it. The latter can be iteratively constructed on the basis of entropy and information by AL.
elements’ atomic properties (atomic radii \( r \) and Pauling electronegativities \( \chi \)). This input layer is followed by a sequence of hidden layers, featuring 512 (1024), 256 (512), and 128 (256) densely connected neurons in the case of energy (lattice parameter) prediction, respectively. These numbers emerged from hyperparameter optimization. Finally, the output layer provides energies \( E_{f} \), \( E_{l} \), \( E_{c} \), or lattice parameters \( (a_{0}^{p}, d_{0}^{ll}, e_{0}^{ll}) \) or lattice parameters \( (a_{0}^{c}, d_{0}^{ll}, e_{0}^{ll}) \). We apply error backpropagation on the training set to automatically adapt the weights that connect the individual neurons, until an optimal mapping from input to output is achieved.

Here, we train these NNs within an AL cycle \([11, 27, 28]\). Starting from a small subset of the parameter space (here initially \( \sim 20\% \)), AL iteratively extends the training set in a statistically optimal way and thus assists in approaching a high prediction accuracy rapidly and efficiently. This type of algorithm can be applied to (at least) two scenarios: (i) If extensive data is available, AL separates information-rich data from redundant or erroneous data. (ii) If data has to be generated first (e.g., by time-consuming quantum-chemical or DFT simulations), AL efficiently drives this data generation, predicting and requesting iteratively what materials to simulate next to extend the training set. In either case, one obtains a small but optimal subset of the materials parameter space that contains a maximum of non-redundant information, which is subsequently used by the NNs to accurately predict the properties of all materials in the entire parameter space.

The iterative AL process is illustrated in figure 5(c). The key question in each step is how to optimally increase the training set size to efficiently enhance the prediction accuracy. We solve it by training two NNs of the same type in parallel and subsequently comparing their individual predictions to estimate which materials are attractive candidates: given the observables \( \phi_{i}^{1,2} \) as predicted by NN 1 and NN 2 and the respective DFT ground truth \( x_{i}^{DFT} \), where \( i \) labels either different energies \( (E_{f}^{X}, E_{l}^{X}, E_{c}^{X}) \) or lattice parameters \( (a_{0}^{p}, d_{0}^{ll}, e_{0}^{ll}) \), we define by averaging over \( i \):

\[
\Sigma(A, B, X) = \frac{1}{N} \sum_{i=1}^{N} \phi_{i}^{1,2} \cdot x_{i}^{DFT},
\]

where \( N \) is the number of data points. This input layer is followed by a sequence of hidden layers, featuring 512 (1024), 256 (512), and 128 (256) densely connected neurons in the case of energy (lattice parameter) prediction, respectively. These numbers emerged from hyperparameter optimization. Finally, the output layer provides energies \( E_{f} \), \( E_{l} \), \( E_{c} \), or lattice parameters \( (a_{0}^{p}, d_{0}^{ll}, e_{0}^{ll}) \) or lattice parameters \( (a_{0}^{c}, d_{0}^{ll}, e_{0}^{ll}) \). We apply error backpropagation on the training set to automatically adapt the weights that connect the individual neurons, until an optimal mapping from input to output is achieved.

Here, we train these NNs within an AL cycle \([11, 27, 28]\). Starting from a small subset of the parameter space (here initially \( \sim 20\% \)), AL iteratively extends the training set in a statistically optimal way and thus assists in approaching a high prediction accuracy rapidly and efficiently. This type of algorithm can be applied to (at least) two scenarios: (i) If extensive data is available, AL separates information-rich data from redundant or erroneous data. (ii) If data has to be generated first (e.g., by time-consuming quantum-chemical or DFT simulations), AL efficiently drives this data generation, predicting and requesting iteratively what materials to simulate next to extend the training set. In either case, one obtains a small but optimal subset of the materials parameter space that contains a maximum of non-redundant information, which is subsequently used by the NNs to accurately predict the properties of all materials in the entire parameter space.

The iterative AL process is illustrated in figure 5(c). The key question in each step is how to optimally increase the training set size to efficiently enhance the prediction accuracy. We solve it by training two NNs of the same type in parallel and subsequently comparing their individual predictions to estimate which materials are attractive candidates: given the observables \( \phi_{i}^{1,2} \) as predicted by NN 1 and NN 2 and the respective DFT ground truth \( x_{i}^{DFT} \), where \( i \) labels either different energies \( (E_{f}^{X}, E_{l}^{X}, E_{c}^{X}) \) or lattice parameters \( (a_{0}^{p}, d_{0}^{ll}, e_{0}^{ll}) \), we define by averaging over \( i \):

\[
\Sigma(A, B, X) = \frac{1}{N} \sum_{i=1}^{N} \phi_{i}^{1,2} \cdot x_{i}^{DFT},
\]

where \( N \) is the number of data points. This input layer is followed by a sequence of hidden layers, featuring 512 (1024), 256 (512), and 128 (256) densely connected neurons in the case of energy (lattice parameter) prediction, respectively. These numbers emerged from hyperparameter optimization. Finally, the output layer provides energies \( E_{f} \), \( E_{l} \), \( E_{c} \), or lattice parameters \( (a_{0}^{p}, d_{0}^{ll}, e_{0}^{ll}) \) or lattice parameters \( (a_{0}^{c}, d_{0}^{ll}, e_{0}^{ll}) \). We apply error backpropagation on the training set to automatically adapt the weights that connect the individual neurons, until an optimal mapping from input to output is achieved.

Here, we train these NNs within an AL cycle \([11, 27, 28]\). Starting from a small subset of the parameter space (here initially \( \sim 20\% \)), AL iteratively extends the training set in a statistically optimal way and thus assists in approaching a high prediction accuracy rapidly and efficiently. This type of algorithm can be applied to (at least) two scenarios: (i) If extensive data is available, AL separates information-rich data from redundant or erroneous data. (ii) If data has to be generated first (e.g., by time-consuming quantum-chemical or DFT simulations), AL efficiently drives this data generation, predicting and requesting iteratively what materials to simulate next to extend the training set. In either case, one obtains a small but optimal subset of the materials parameter space that contains a maximum of non-redundant information, which is subsequently used by the NNs to accurately predict the properties of all materials in the entire parameter space.

The iterative AL process is illustrated in figure 5(c). The key question in each step is how to optimally increase the training set size to efficiently enhance the prediction accuracy. We solve it by training two NNs of the same type in parallel and subsequently comparing their individual predictions to estimate which materials are attractive candidates: given the observables \( \phi_{i}^{1,2} \) as predicted by NN 1 and NN 2 and the respective DFT ground truth \( x_{i}^{DFT} \), where \( i \) labels either different energies \( (E_{f}^{X}, E_{l}^{X}, E_{c}^{X}) \) or lattice parameters \( (a_{0}^{p}, d_{0}^{ll}, e_{0}^{ll}) \), we define by averaging over \( i \):

\[
\Sigma(A, B, X) = \frac{1}{N} \sum_{i=1}^{N} \phi_{i}^{1,2} \cdot x_{i}^{DFT},
\]

where \( N \) is the number of data points. This input layer is followed by a sequence of hidden layers, featuring 512 (1024), 256 (512), and 128 (256) densely connected neurons in the case of energy (lattice parameter) prediction, respectively. These numbers emerged from hyperparameter optimization. Finally, the output layer provides energies \( E_{f} \), \( E_{l} \), \( E_{c} \), or lattice parameters \( (a_{0}^{p}, d_{0}^{ll}, e_{0}^{ll}) \) or lattice parameters \( (a_{0}^{c}, d_{0}^{ll}, e_{0}^{ll}) \). We apply error backpropagation on the training set to automatically adapt the weights that connect the individual neurons, until an optimal mapping from input to output is achieved.

Here, we train these NNs within an AL cycle \([11, 27, 28]\). Starting from a small subset of the parameter space (here initially \( \sim 20\% \)), AL iteratively extends the training set in a statistically optimal way and thus assists in approaching a high prediction accuracy rapidly and efficiently. This type of algorithm can be applied to (at least) two scenarios: (i) If extensive data is available, AL separates information-rich data from redundant or erroneous data. (ii) If data has to be generated first (e.g., by time-consuming quantum-chemical or DFT simulations), AL efficiently drives this data generation, predicting and requesting iteratively what materials to simulate next to extend the training set. In either case, one obtains a small but optimal subset of the materials parameter space that contains a maximum of non-redundant information, which is subsequently used by the NNs to accurately predict the properties of all materials in the entire parameter space.

The iterative AL process is illustrated in figure 5(c). The key question in each step is how to optimally increase the training set size to efficiently enhance the prediction accuracy. We solve it by training two NNs of the same type in parallel and subsequently comparing their individual predictions to estimate which materials are attractive candidates: given the observables \( \phi_{i}^{1,2} \) as predicted by NN 1 and NN 2 and the respective DFT ground truth \( x_{i}^{DFT} \), where \( i \) labels either different energies \( (E_{f}^{X}, E_{l}^{X}, E_{c}^{X}) \) or lattice parameters \( (a_{0}^{p}, d_{0}^{ll}, e_{0}^{ll}) \), we define by averaging over \( i \):

\[
\Sigma(A, B, X) = \frac{1}{N} \sum_{i=1}^{N} \phi_{i}^{1,2} \cdot x_{i}^{DFT},
\]
Figure 6. Evolution of the MAE with the training set size during AL, comparing element-embedding NNs of type 1 [figure 5(a)] for energy (top) and lattice parameter prediction (bottom) that exclude (left) and include (right) atomic properties as additional input. In each panel, MAE curves for three distinct NNs trained exclusively on the N, O, or F data are contrasted with a MAE curve of a single NN trained on the combined NOF data. Each curve represents the mean of equation (1) over a series of AL runs starting from different random initial training sets, and the colored areas display the corresponding standard deviations. This demonstrates that the reliability and accuracy of AL is largely independent of the initial training set. In parallel, the insets monitor the evolution of the NOF training set composition, starting from a 33% distribution of N, O, and F materials. Particularly for lattice parameter prediction, one can see that the AL algorithm preferentially includes the more challenging fluorides.

\[
\text{MAE}(A,B,X) = \left\langle \frac{\phi_1(A,B,X) + \phi_2(A,B,X)}{2} - x_i^{\text{DFT}}(A,B,X) \right\rangle_i
\]

\[
\Sigma(A,B,X) = \left\langle |\phi_1(A,B,X) - \phi_2(A,B,X)| \right\rangle_i
\]

Site averaging (i.e., integrating over the entire parameter space) yields the overall mean absolute error

\[
\text{MAE} = \left\langle \text{MAE}(A,B,X) \right\rangle_{A,B,X}.
\] (1)

Naturally, if X is fixed, averaging over X is omitted. In the AL cycle [figure 5(c)], the training set is iteratively updated, appending \(\sim 130\) N, O, or F (\(\sim 400\) NOF) materials per step that exhibit the highest \(\Sigma(A,B,X)\), followed by further NN training. Interestingly, \(\Sigma\) can be interpreted as an estimate of the local entropy in the parameter space [28]. In this spirit, the present AL algorithm statistically maximizes the information entailed in the training set. From the definition of \(\Sigma\) it follows that the DFT ground truth beyond the current training set is not required by the AL algorithm to select interesting materials candidates; we use it only \textit{a posteriori} to obtain the MAE and thereby analyze the AL performance [figures 6–8]. Finally, the presented AL algorithm can be stopped as soon as the desired accuracy is reached, establishing the latter as a systematic control parameter. Since only the autonomously selected materials need to be calculated \textit{ab initio} in each iteration, a substantial gain in performance and energy efficiency as compared to conventional high-throughput calculations can be achieved.

4.2. Evolution of the prediction accuracy during active learning

Figure 6 explores how the MAE evolves during AL with the training set size (i.e., the number of AL iterations) by comparing element-embedding NNs of type 1 [cf figure 5(a)] for energy and lattice parameter prediction. Additionally, the impact of providing atomic properties as additional input is
Figure 7. Prediction of $E_{VX}^f$, $a_0^f$, $c_{IL}^f$, and $c_{IL}^0$ in the combined NOF parameter space by single NNs of type 1 [figure 5(a)] versus the DFT ground truth, using $\sim$50% of the data as training set (blue points, seen by the NNs). The red points represent the test set, which has never been presented to the NNs before. The insets display the respective prediction error distributions, including the MAEs. Contrasting results obtained by AL (top and middle row) with those obtained for a randomly chosen training set of equal size (bottom row) reveals the advantages of AL.

analyzed. We note again that the overall MAE [equation (1)] is only accessible since all the data has been calculated in advance in this case. In a ‘real’ AL materials prediction run, this information would not be available. Therefore, we have the opportunity here to explore in very detail the reliability of this AI technique.

First of all, three MAE curves for N, O, and F are shown in each panel of figure 6. They have been obtained from three individual NNs specialized on either N, O, or F prediction; i.e., they are exclusively trained and evaluated on the respective data sets. As one can see from the definition in equation (1), the MAE is a highly integrated measure that averages the NN prediction difference with respect to the DFT ground truth over different observables and the parameter space, and thus condenses a sizable amount of information in a single quantity. Moreover, each curve represents the mean of 25 consecutive and independent AL runs, where each AL run started from a different random initial training set. The colored areas visualize $\pm$ the corresponding standard deviations. Interestingly, these standard deviation corridors are rather small and furthermore narrow rapidly with increasing training set size. This demonstrates that the AL process is largely independent of the initial training set composition, and can therefore be considered highly reliable. We can also see that the accuracy improves systematically, since the MAE curves decrease monotonically with the number of AL iterations.

Second, each panel presents an additional MAE curve that corresponds to a NN of identical architecture which, in contrast, is trained and evaluated on the combined NOF data. In this case, we operate with three times larger training sets, so that the overall number of compounds used for training (and therefore the time and energy invested in DFT calculations) at each step is equal and the accuracy therefore legitimately comparable. Again, each curve constitutes the mean of 25 consecutive AL runs, and is displayed together with the respective standard deviation. At first sight, these NOF curves resemble a simple average of the individual N, O, and F curves. However, the key aspect here is the NOF NNs can identify and learn universal properties of the N, O, and F data sets and benefit from their synergies. We will explore and quantify this important difference below.

In order to assess the AL performance, we use an error of 1% as reference, which we calculate for the energies relative to $\sim$20 eV range of $E_{VX}^f$ and for the lattice parameters relative to
the average of the ranges of $a_0^p$, $a_0^{1H}$, and $c_0^{1H}$, which amounts to $\sim$4 Å [see also figures 2(a) and 7]. Figure 6 shows that the N, O, and F energy curves drop below 1% relative error between $\sim$30%–35%, the NOF curve at $\sim$32% training set size. Here, the oxides always exhibit the highest MAE, whereas the fluoride curve is consistently the lowest (most precise). Overall, the curves are close together, in particular for training set sizes beyond 35%. In contrast, the lattice parameter MAE curves exhibit a clearly larger spread. The N, O, and F curves drop below 1% relative error between $\sim$25%–35%, the NOF curve at 29% training set size. The fluoride curve is by far higher than the oxide and nitride curves, the latter being consistently the lowest. This reflects on the one hand how challenging the prediction of energies versus lattice parameters is in general, but simultaneously how the complexity varies between the three classes. Furthermore, we can state that while during the first AL iterations atomic properties improve the accuracy, the impact of this scalar input substantially reduces for training set sizes beyond $\sim$30%. This underlines the strength of element embedding, which renders a priori knowledge largely unnecessary.

Interestingly, in the NOF case, the AL algorithm has the flexibility to dynamically control the relative composition of the training set. Therefore, it is worthwhile to monitor in parallel how it evolves during AL (figure 6, insets). For energy prediction, the contributions from the N, O, and F subsets remain largely balanced around the initial 33%, with a slight preference emerging for O and F at the expense of N. A fundamentally different behavior can be observed for lattice parameter prediction, which, as we observed above, is most challenging for the fluorides. This is automatically identified by the AL algorithm and mitigated by increasing the fraction of fluorides in the training set. Already after 4–5 AL iterations, the training set composition reaches the final values $\sim$40% F versus $\sim$30% O and N. The inclusion of atomic properties as additional NN input does not change this picture significantly.

### 4.3. Observable-resolved prediction accuracy: superiority of active learning

A more detailed perspective on the performance of the NOF NNs with respect to $E_{\text{tot}}^{1Y}$, $E_{\text{tot}}^\text{1H}$, and $c_{\text{IL}}^{\text{1H}}$ prediction is shown in figure 7. In the top row we can see results for single NNs of type 1, suppressing additional atomic properties input. AL-iterating towards a training set size of $\sim$50%, we already obtain a MAE $\sim$0.1 eV for $E_{\text{tot}}^{1Y}$ per vacancy. Hence, we achieve a comparable accuracy on the NOF data set as before on oxides alone (0.072/0.126 eV MAE for $E_{\text{tot}}^{1Y}$ on seen/unseen data [28]), which is also in line with the uniform convergence of the MAE curves in figure 6. Relative to their overall range of $\sim$20 eV, this corresponds to an error of only $\sim$0.5%. The heats of formation $E_{\text{f}}^{1Y}$ and $E_{\text{f}}^{1H}$ are predicted with even higher accuracy, reaching $\sim$30 meV/atom (not shown), which is comparable to recent work on perovskite oxides (20–34 meV/atom [21]) and well within DFT accuracy [20, 79]. For elpasolites, a heat of formation accuracy of 150 meV/atom was obtained [24]. These observations reflect that $E_{\text{tot}}^{1Y}$ is a fingerprint of the complex reduction reaction and consequently more demanding to predict. Among the lattice parameters, the prediction of $a_0^p$ and $a_0^{1H}$ proved to be straightforward, reaching MAEs around 0.01 Å (figure 7). In contrast, $c_{\text{IL}}^{1H}$ turned out to be more challenging, with a higher MAE of $\sim$0.03 Å. This can be traced back to the sparse data available for vertically expanding materials [figure 2(c)] and the almost vanishing correlations of $c_{\text{IL}}^{1H}$ with the other observables (figure 4). Again, we achieve a comparable accuracy on the NOF data set as before on oxides alone (0.01/0.01 Å for $a_0^{1H}$ and 0.02/0.03 Å for $c_{\text{IL}}^{1H}$ on seen/unseen data [28]), which are only slightly better than the MAEs reported here.

As we discussed above, the positive impact of atomic properties in the NN input for small training sets reduces as more and more data becomes available. The results displayed in the middle row are even slightly less accurate (i.e., they exhibit a higher MAE) than those obtained for element embedding alone (top row).

The results shown in the bottom row have been obtained for a randomly chosen training set of equal size ($\sim$50%). While the overall accuracy on the training data is good, the performance on the unseen data is by far inferior. We confirmed that this effect is not related to overfitting. In particular for $c_{\text{IL}}^{1H}$ it becomes obvious that AL considerably enhances the prediction accuracy and thus emerges as superior strategy (figure 7).

Note that the overall numerical effort is quasi identical, since the DFT simulations are the time-consuming step, whereas the AL overhead is negligible. In summary, our results show how difficult these observables are to predict in general for such complex materials classes, even at $\sim$50% training set size, and that a naive random approach is insufficient.

### 5. Transfer learning synergies

Next, we address the fundamental question in how far NNs that are trained on the combined NOF data can identify universal quantum-mechanical systematics and thereby utilize the synergies between nitrides, oxides, and fluorides to enhance their prediction accuracy. Simultaneously, this implies that a smaller training set is sufficient to obtain a desired precision, i.e., it enhances the sample efficiency in the spirit of transfer learning.

Figures 8(a) and (b) compare the prediction accuracy of NOF NNs of type 1 and type 2 (the latter featuring the $B_X$ tensor product embedding), both without atomic properties in the input, to the averaged MAE of three specialized NNs that were trained exclusively on the N, O, and F data:

\[
\text{MAE}_\text{N} + \text{MAE}_\text{O} + \text{MAE}_\text{F} \over 3. 
\]  

Again, in order to provide a statistically representative impression of the reliability, each curve displays the mean of 25 consecutive and independent AL runs (see above), whereas the colored area visualizes ± the corresponding standard deviation. For lattice parameter prediction [figure 8(b)], the NOF NN exhibits a consistently higher accuracy than the three specialized NNs on average, irrespective of the input encoding and even including the standard deviation. For energy prediction [figure 8(a)], the precision of NN type 1 is better than the N,
O, F average below ∼30% training set size, whereas the MAE of NN type 2 is slightly higher. For training set sizes >30%, the two NN types exhibit a quasi equivalent prediction accuracy, despite the fundamental difference in their architecture and physical motivation. NN type 2 may require slightly more data than NN type 1 to establish an accurate representation of the chemical elements in the more complex BX embedding space, but in turn provides highly interesting insights, as we will discuss below.

Figure 8(c) specifically emphasizes the aspect of universality, now focusing exclusively on NNs of type 1, but additionally exploring the impact of the atomic properties input channel. From the curves shown in figure 6, we define the following measure to quantify the synergistic effect:

\[
\Delta \text{MAE} = \frac{3 \text{MAE}_{\text{NOF}}}{\text{MAE}_N + \text{MAE}_O + \text{MAE}_F} - 1
\]

i.e., we use the averaged MAE of the specialized N, O, and F NNs defined above as reference, and \(\Delta \text{MAE} < 0\) indicates a superior performance of the NOF NNs. Indeed, the \(\Delta \text{MAE}\) curves displayed in figure 8(c) demonstrate an overall improvement of the prediction accuracy by up to 16% due to unraveled synergies at equivalent numerical effort, in particular for smaller training sets. For lattice parameter prediction, the curves range from −16% to −2% if element embedding alone is employed, and from −8% to +1% if atomic properties are included in the input. For energy prediction, the results based exclusively on element embedding range from −13% to +8%, and from −4.5% to +9% if atomic properties are included in the input.

The overall positive slope exhibited by all \(\Delta \text{MAE}\) curves in figure 8(c) indicates that the enhancing impact of synergistic effects reduces as more and more training data becomes explicitly available. For energy prediction, we consistently obtain \(\Delta \text{MAE} > 0\) beyond 35% training set size. However, the energy curves also offer a different, yet speculative interpretation. The better accuracy of the specialized NNs for larger training sets may trace back to their higher memory-to-data ratio, as they feature the identical architecture as the NOF NNs. If we tentatively eliminate this aspect by considering the final \(\Delta \text{MAE}\) values as reference (reaching up to ∼+8%), which almost aligns the energy curves with the lattice parameter curves, we can identify considerable synergistic effects of up to ∼20% for small training sets also for energy prediction.

Another interesting observation is that the impact of synergies reduces to some extent if atomic properties are provided as additional input. This evidences that the atomic radii and the electronegativities constitute a sizable part of the underlying systematics. Specifically, one could naively expect that the lattice parameters are largely determined by the atomic radii (despite the complexity of \(c_{IL}\) discussed above; see figure 4). However, particularly for these observables the enhancement is clearly not quenched, which highlights that the AI identifies further, possibly more abstract synergies.

These results imply that the systematics unraveled for one materials class enhance the prediction accuracy for the other two materials classes. This boosted sample efficiency is the very essence of transfer learning. We found that the positive impact of these synergies is largest for smaller training sets. From a practical perspective, this early phase of AL plays an important role, as one seeks to keep the training set (and thus the number of required DFT calculations) as small as possible. This maximizes the advantage and efficiency with respect to a conventional high-throughput approach, where the entire parameter space is explicitly calculated from first principles.
6. Element embedding analysis

As the final step, we analyze how the NNs perceive the materials data. Figure 9 explores the element embedding vectors that emerged automatically in the two types of energy-prediction NNs after being AL-trained towards ~50% of the NOF data. While the clusters of chemically similar elements are not as pronounced and distinct in panels (a) and (b) as in the t-SNE analysis of the embedding vectors trained on oxides alone [28], their overall topology unambiguously exhibits universal trends. For clarity, blue and violet arrows guide the eye by marking the exemplary elements Hf, W, and Hg. Note that here no atomic properties are input to the NNs, so that the AI arranges the chemical elements autonomously in a 16-dimensional vector space without any external a priori knowledge. As indicated by the insets, panels (a) and (b) belong to NN type 1 displayed in figure 5(a), whereas panel (c) refers to NN type 2 displayed in figure 5(b) featuring the BX tensor product embedding (cf table 2). The A-site embedding analysis for NN type 2 is quasi identical to panel (a) and therefore not shown.

![Figure 9](image)

Figure 9. Two-dimensional principal component analysis (PCA) of the element embedding vectors emerging in the two types of energy-prediction NNs after being AL-trained towards ~50% of the NOF data. While the clusters of chemically similar elements are not as pronounced and distinct in panels (a) and (b) as in the t-SNE analysis of the embedding vectors trained on oxides alone [28], their overall topology unambiguously exhibits universal trends. For clarity, blue and violet arrows guide the eye by marking the exemplary elements Hf, W, and Hg. Note that here no atomic properties are input to the NNs, so that the AI arranges the chemical elements autonomously in a 16-dimensional vector space without any external a priori knowledge. As indicated by the insets, panels (a) and (b) belong to NN type 1 displayed in figure 5(a), whereas panel (c) refers to NN type 2 displayed in figure 5(b) featuring the BX tensor product embedding (cf table 2). The A-site embedding analysis for NN type 2 is quasi identical to panel (a) and therefore not shown.

6. Element embedding analysis

As the final step, we analyze how the NNs perceive the materials data. Figure 9 explores the element embedding vectors that emerged automatically in the two types of energy-prediction NNs after being AL-trained towards ~50% of the NOF data. We performed a dimensionality reduction via a principal component analysis (PCA [80]), which represents the element vectors from the 16-dimensional embedding space (see above) by points in two dimensions, while simultaneously retaining as much of their variance as possible. The two dimensions are rather of mathematical character and do not necessarily have a physical correspondence. This technique is less complex than stochastic neighbor embedding (t-SNE), which we employed recently to analyze the embedding vectors trained on oxides alone [28]. Consequently, the clusters of chemically similar elements in figures 9(a) and (b) are not as pronounced, distinct, and well-separated. However, being a linear transformation, the PCA dimension reduction has the advantage of modifying any structure in the data rather gently. One can clearly see in the analysis of the A- and the B-site embedding layers [figures 9(a) and (b)] that the overall topology exhibited by the chemical elements unambiguously shows universal trends. For instance, the transition metal cluster shows an evolution from right to left that aligns largely with the periodic table of elements (from Ti, Zr, and Hf to Zn, Cd, and Hg), the lighter 3d elements being organized more towards the center, the heavier 5d elements more at the outside. A distinct cluster of rare-earth metals forms that interestingly also comprises the group-3 elements Sc and Y. Moreover, the alkali metals are consistently located in the top-left corner (e.g., Rb), next to the alkaline earths (e.g., Ca, Sr, and Ba) which appear more centered. This universal structure is particularly compelling since (i) the A and the B sites have a highly distinct impact on the compounds’ chemistry and (ii) the X site fundamentally modifies this impact, as discussed above. Moreover, we stress that no atomic properties are input to the NNs here, and that the AI is forced to arrange the chemical elements completely autonomously in a 16-dimensional vector space, unbiased by any external a priori knowledge. The NNs develop their individual understanding of the chemical similarity between the different elements, being agnostic about concepts such as the atomic number or the group of a particular element.

An even more fascinating question is how the entangled B and X sites in NN type 2 are interpreted and ordered in their combined embedding space. Albeit rather complex at first sight, the analysis of this BX tensor product embedding displayed in figure 9(c) paints a consistent, yet different and more realistic picture of the (dis-)similarity of the three distinct materials classes than the data clouds in figure 2(a), from which one could naively conclude that the rather stable fluorides and oxides bear higher similarities, whereas the unstable nitrides are consequently different. From top to bottom, figure 9(c) shows an evolution from nitrides over oxides to fluorides, i.e., the materials data is automatically structured into (two to) three distinct clusters. Each of the three materials clusters reflects, in broader strokes and partly modified, internally the universal topology observed in figures 9(a) and (b).

Interestingly, the fluoride cluster is almost entirely separated from oxides and nitrides, while the oxide and nitride clusters exhibit substantial overlap. This aspect is numerically analyzed in table 2, where we alternatingly compare two of the three clusters. In order to quantify their similarity, which represents the similarity of the respective materials classes, two different measures are employed: the Euclidean distance and the silhouette coefficient. The (dimensionless) Euclidean
Table 2. Numerical analysis of the three materials clusters emerging during the BX tensor product embedding quantifies the (dis-)similarity of nitrides, oxides, and fluorides by employing different techniques. Alternately, two of the three clusters are compared. Smaller (larger) Euclidean distances and silhouette coefficients indicate a higher (dis-)similarity. While ‘16 dim.’ denotes that the analysis has been performed in the entire embedding space, ‘2 dim.’ refers to an analysis subsequent to the PCA dimensionality reduction [cf figure 9(c)].

| Measure | F vs N | F vs O | O vs N |
|---------|--------|--------|--------|
| Euclidean distance, 16 dim. | 1.031 | 0.918 | 0.616 |
| Euclidean distance, 2 dim. | 0.999 | 0.829 | 0.357 |
| Silhouette coefficient, 16 dim. | 0.238 | 0.187 | 0.098 |
| Silhouette coefficient, 2 dim. | 0.445 | 0.340 | 0.079 |

distances of the centroids (i.e., the geometric ‘centers of mass’) clearly show that fluorides are much more different from nitrides and oxides (reaching values ∼1) than oxides and nitrides are from each other (0.616). Simultaneously, fluorides are slightly less distant from oxides than from nitrides. This holds both in the abstract 16-dimensional embedding space as well as in the PCA-reduced two-dimensional space shown in figure 9(c). Moreover, we can infer that the similarity of oxides and nitrides is artificially enhanced by the PCA, as the Euclidean distance is reduced from 0.616 to 0.357. Thus, in 16 dimensions it becomes more clear that three (and not only two) clusters are constructed by the AI. Hence, it is indispensable to concomitantly explore the results obtained from the full 16-dimensional data. A complementary and more involved characteristic is the (averaged) silhouette coefficient [80, 81]. It is based on the clusters’ tightness and separation and ranges from −1 to +1, where a high positive value is indicative of disentangled and distant clusters, whereas a near-zero value corresponds to overlapping clusters. (Negative values are of minor relevance here, as they arise predominantly when more than two clusters are compared.) Overall, the silhouette coefficient confirms the (dis-)similarities unraveled by the Euclidean distance measure in 16 and two dimensions, and also that PCA tunes the contrast to some extent (table 2). We conclude from these observations that (i) the two-dimensional PCA paints a representative picture of the situation in the abstract 16-dimensional embedding space, and (ii) that fluorides constitute a rather distinct materials class, whereas oxides and nitrides bear higher similarities.

7. Summary

Combining DFT simulations and AL of element-embedding NNs, we investigated the sample efficiency for the prediction of vacancy layer formation energies and lattice parameters in infinite-layer versus perovskite nitrides, oxides, and fluorides in the spirit of transfer learning. First, we provided a detailed data analysis from different thermodynamic, structural, and statistical perspectives. Subsequently, we showed that NNs model these observables with high precision, using just ∼30% of the data for training and based exclusively on the constituent element names as minimal input devoid of any physical a priori knowledge. Element embedding autonomously arranges the chemical elements with a characteristic and recurrent topology, such that their relations align with the conventional picture of the periodic table. We compared two different embedding strategies and showed that these techniques render additional scalar input such as atomic properties negligible. AL was found to be largely independent of the initial training set and thereby emerged as reliable and robust algorithm. Additionally, we exemplified its superiority over randomly composed training sets. The present methodology successfully identified fundamental quantum-mechanical universalities between the three materials classes despite their largely distinct chemistry that enhanced the combined prediction accuracy by up to 16% with respect to three specialized NNs that were trained exclusively on the isolated data sets, highlighting an increased sample efficiency. This quantification of synergistic effects provides an impression of the transfer learning improvements one may expect for other materials of similar complexity. Finally, traditional data analysis and unconventional artificial-intelligence methodology in the form of BX tensor product embedding combined with subsequent quantitative cluster analysis converged on the picture that fluorides constitute a rather distinct materials class, whereas oxides and nitrides bear substantial parallels.

Acknowledgments

BG acknowledges financial support of this project provided by the Excellent Early Career Researchers Funding Competition of the University of Duisburg-Essen and a Grant by the Department of Physics of the University of Duisburg-Essen.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Benjamin Geisler Ⓢ https://orcid.org/0000-0003-0864-2908

References

[1] Curtarolo S, Hart G L W, Nardelli M B, Mingo N, Sanvito S and Levy O 2013 Nat. Mater. 12 191
[2] van Roekeghem A, Carrete J, Oses C, Curtarolo S and Mingo N 2016 Phys. Rev. X 6 041061
[3] Mouvet N et al 2018 Nat. Nanotechnol. 13 246
[4] Marzari N, Ferretti A and Wolverton C 2021 Nat. Mater. 20 736
[5] Singh A K, Montoya J H, Gregoire J M and Persson K A 2019 Nat. Commun. 10 443
[6] Jain A et al 2013 APL Mater. I 011002
[7] Saal J E, Kirklin S, Aykol M, Meredig B and Wolverton C 2013 JOM 65 1501
[8] Curtarolo S et al 2012 Comput. Mater. Sci. 58 218
[9] Draxl C and Scheffler M 2018 MRS Bull. 43 676
[10] Butler K T, Davies D W, Cartwright H, Isayev O and Walsh A 2018 Nature 559 547
[11] Schmidt J, Marques M R G, Botti S and Marques M A L 2019 npj Comput. Mater. 5 83
[12] Wang A Y-T, Murdock R J, Kauwe S K, Oliynyk A O, Gurlro A, Brgoch J, Persson K A and Sparks T D 2020 Chem. Mater. 32 4954
[13] Ward L, Agrawal A, Choudhary A and Wolverton C 2016 npj Comput. Mater. 2 16028
[14] Faber F A, Lindmaa A, von Lilienfeld O A and Armiento R 2016 Phys. Rev. Lett. 117 135502
[15] Ghiringhelli L M, Vybiral J, Levechenko S V, Draxl C and Scheffler M 2015 Phys. Rev. Lett. 114 105503
[16] Bartel C J, Sutton C, Goldsmith B R, Ouyang R, Musgrave B H, Wang Y, Kang C J, Miao H and Kotliar G 2020 npj Comput. Mater. 6 102
[17] Schmidt J, Shi J, Borbild P, Chen L, Botti S and Marques M A L 2017 Chem. Mater. 29 5090
[18] Belviso F et al 2019 Inorg. Chem. 58 14939
[19] Lopez-Bezanilla A and Littlewood P B 2020 MRS Commun. 10 1
[20] Xie T and Grossman J C 2018 Phys. Rev. Lett. 120 145301
[21] Ye W, Chen C, Wang Z, Chu I-H and Ong S P 2018 Nat. Commun. 9 3800
[22] Jha D, Choudhary K, Tavazza F, Liao W-K, Choudhary A, Faber F A, Lindmaa A, von Lilienfeld O A and Armiento R 2016 Phys. Rev. Lett. 117 135502
[23] Zhou Q, Tang P, Liu S, Pan J, Yan Q and Zhang S-C 2018 Proc. Natl Acad. Sci. USA 115 E6411
[24] Jha D, Ward L, Paul A, Liao W-K, Choudhary A, Wolverton C and Agrawal A 2018 Sci. Rep. 8 17593
[25] Tshitoyan V, Dagdelen J, Weston L, Dunn A, Rong Z, Kononova O, Persson K A, Ceder G and Jain A 2019 Nature 571 95
[26] Lookman T, Balachandran P V, Xue D and Yuan R 2019 npj Comput. Mater. 5 21
[27] Malashevich A and Ismail-Beigi S 2015 Phys. Rev. B 91 075137
[28] Kohn W and Sham L J 1965 Phys. Rev. 140 A1133
[29] Kresse G and Joubert D 1999 Phys. Rev. B 59 1758
[30] Liu J et al 2013 Nat. Commun. 4 2714