JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS

Open Access
This article is licensed under CC-BY 4.0

Article

Band Alignment of Oxides by Learnable Structural-Descriptor-Aided Neural Network and Transfer Learning

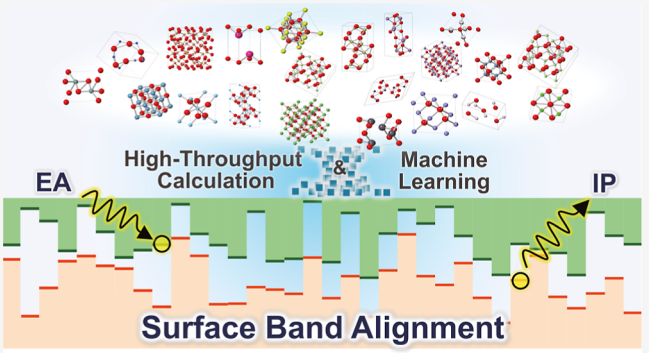
Shin Kiyohara,* Yoyo Hinuma, and Fumiyasu Oba*

Cite This: J. Am. Chem. Soc. 2024, 146, 9697−9708

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The band alignment of semiconductors, insulators, and dielectrics is relevant to diverse material properties and device structures utilizing their surfaces and interfaces. In particular, the ionization potential and electron affinity are fundamental quantities that describe surface-dependent band-edge positions with respect to the vacuum level. Their accurate and systematic determination, however, demands elaborate experiments or simulations for well-characterized surfaces. Here, we report machine learning for the band alignment of nonmetallic oxides using a high-throughput first-principles calculation data set containing about 3000 oxide surfaces. Our neural network accurately predicts the band positions for relaxed surfaces of binary oxides simply by using the information on bulk structures and surface termination planes. Moreover, we extend the model to naturally include multiple-cation effects and transfer it to ternary oxides. The present approach enables the band alignment of a vast number of solid surfaces, thereby opening the way to a systematic understanding and materials screening.



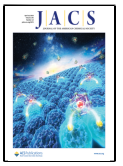
INTRODUCTION

The ionization potential (IP) and electron affinity (EA) of a nonmetallic solid are defined as the energy levels of the valence band maximum (VBM) and conduction band minimum (CBM) with respect to the vacuum level, respectively. They show not only how easily an electron is released or accepted but also information about the relative band positions of materials. Such band alignments or lineups play crucial roles in our understanding, design, and development of materials and devices utilizing surfaces and heterointerfaces, including photocatalysts,[1−3] photovoltaics,[4−6] and all other heterostructured electronic and optoelectronic devices.[7−13] The lattice mismatch, local atomic structure, and charge transfer should also be considered for depicting complete band alignments at heterointerfaces; nevertheless, IPs and EAs are fundamental information for designing interfacial functionalities in addition to surface-related properties.[7]

The IPs and EAs of solids involve contributions of surface dipole moments. Thus, they significantly depend on the atomic and electronic structures in the vicinity of the surfaces and therefore on the surface orientation and composition.[14−16] Computationally, first-principles calculations have successfully quantified IPs and EAs.[15−23] Separate calculations of bulk and surface systems are performed in a typical procedure for the evaluation of IPs and EAs. Bulk calculations offer VBM and CBM positions with respect to a reference level; surface calculations offer the differences between the vacuum and reference levels, which include the surface dipole contributions to electrostatic potentials. IPs and EAs are then obtained by the alignment of the bulk and surface reference levels.

However, an accurate and high-throughput evaluation of IPs and EAs is not straightforward because both processes are time-consuming. Density functional theory calculations using standard local and semilocal functionals produce large errors in the VBM and CBM positions, requiring us to use more elaborate and computationally demanding approaches, e.g., hybrid functionals including nonlocal exchange contributions and *GW* approximations based on many-body perturbation theory.[15−20,24,25] Surfaces have both macroscopic and microscopic degrees of freedom, namely, the surface orientation as given by Miller indices and the location of the termination plane. Substantial reconstruction from ideal structures also takes place at some surfaces,[26−33] which can significantly affect band positions as reported for perovskite oxides and TiO_2.[32,33] Therefore, even one material has a large variety of surface atomic structures, and systematic evaluation of their IPs and EAs requires huge amounts of computation.

Received: December 2, 2023
Revised: February 25, 2024
Accepted: February 27, 2024
Published: March 28, 2024



© 2024 The Authors. Published by American Chemical Society

Recently, machine learning has been widespread in materials science. Virtual screening based on machine learning is an efficient way to explore new materials with desired functions, where a surrogate model enables us to virtually predict material properties.[34−39] In particular, theoretical calculations can generate comparatively large data sets that are prerequisites for constructing accurate surrogate models. Previous studies have reported successful materials screening for various properties by machine learning in combination with theoretical calculations.[35−38,40−44]

In this article, we present a regression model based on an artificial neural network (ANN) using the smooth overlap of atom positions (SOAPs)[45] as input descriptors to predict the IPs and EAs of nonmetallic oxides (see Figure 1). First-principles calculations based on a non-self-consistent dielectric-dependent hybrid functional approach[21] allow for the accurate and efficient evaluation of the IPs and EAs of about 3000 binary and ternary oxide surfaces. The ANN model constructed using this data set accurately predicts the IPs and EAs of binary oxides despite the model using the information before structural relaxation at surfaces as the input, namely, bulk crystal structures and surface termination planes of the oxides of interest. Moreover, we enable the ANN model to be applied to ternary oxides by developing "learnable" SOAPs, which can incorporate atom species varieties while keeping low descriptor dimensions.

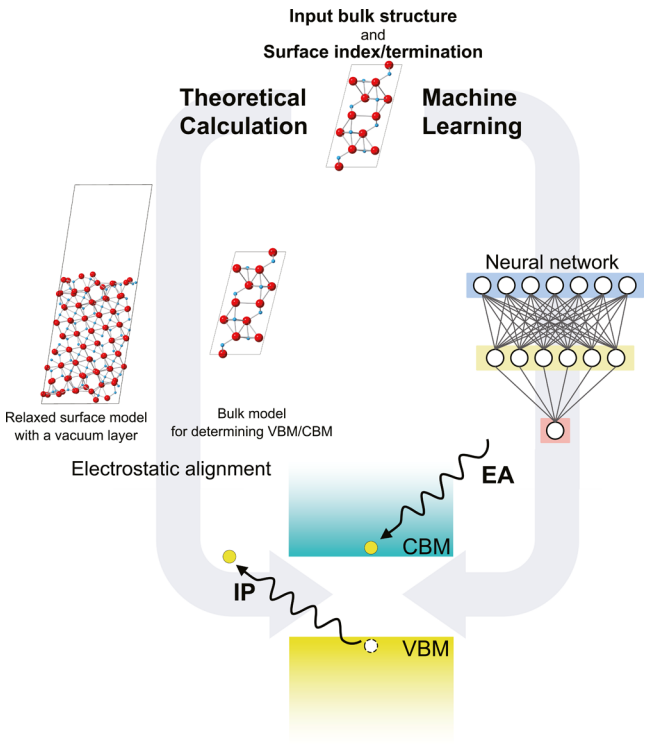


Figure 1. Schematic of prediction of IPs and EAs in nonmetallic solids by theoretical calculations and machine learning. The theoretical calculations from first principles typically use a combination of surface and bulk models to evaluate the energy difference between the vacuum level and the VBM (IP) or CBM (EA). Our ANN predicts the IPs and EAs of relaxed surfaces by simply inputting the information on the bulk crystal structure and the surface index and termination plane.

RESULTS AND DISCUSSION

High-Throughput First-Principles Calculations of IPs and EAs of Binary Oxides. Figure 2a shows distributions of the theoretical IPs and EAs in the binary oxide data set for 2195 nonpolar surfaces; all IP and EA values can be found in the Supporting Information. Here, we focus on nonpolar surfaces because the modeling of polar surfaces requires the consideration of system- and environment-dependent charge-compensation mechanisms[26,27,46] and therefore their high-throughput calculations are not straightforward. The horizontal axis in Figure 2a is the constituent cation species, each of which contains various surfaces of several polymorphs. We find the tendency that the IP and EA values depend largely on the cation species, although the variety in the crystal and surface atomic structures leads to deviations of ∼ ±1 eV from their average values. The calculated IPs and EAs for selected surfaces are shown with available experimental values in Figure 2b. Although the IPs and EAs are, by nature, rather sensitive to the detailed structures and conditions of the experimentally investigated surfaces, reasonable agreement between the experiment and theory is recognized overall. The non-self-consistent dielectric-dependent hybrid functional approach taken here also well-reproduces the experimental IPs and EAs of group IV, III−V, and II−VI semiconductor surfaces.[21] These comparisons with the experimental results demonstrate the accuracy of our high-throughput surface calculation data.

Construction of ANN Models. Using the binary oxide data set, we constructed the ANN (hereinafter called simple-ANN) as shown in Figure 3a. The SOAP descriptors, which sophisticatedly include structural information such as the bond length, bond angle, and coordination number,[45] are evaluated for unrelaxed surfaces and inputted in the first layer to describe the IPs and EAs at relaxed surfaces. The SOAPs in this study were preprocessed as follows. When multiple elements are included in a system, the concatenations of the SOAPs based on the elemental pairs (hereinafter called element-pair SOAPs) are generally used for describing atomic structures in a system[47] as shown in the bottom left panel of Figure 4. Here, we consider 41 atom species simultaneously. Thus, the element-pair SOAPs have several tens of thousands of dimensions because SOAPs of a pair of elements have several tens to several hundreds of dimensions. To reduce the number of descriptor dimensions and prevent overfitting, we dealt with the 40 cation species as the same species when calculating SOAPs and oxygen as an anion, resulting in calculating the SOAPs of three pairs, namely, cation−cation, cation−anion, and anion−anion combinations (hereinafter called cation−anion-pair SOAPs), as shown in the bottom center panel of Figure 4. The cation−anion-pair SOAPs were calculated on each atom site in a system while the IP and EA need one-to-one correspondence with a system. Thus, we averaged the cation−anion-pair SOAPs in a system for inputs. Assuming that the information about the vicinity of the surfaces is critical to the IPs and EAs, we extracted the surface-region atoms by the scheme developed in ref 48 and averaged their SOAPs. More details of the architecture of the simple-ANN and SOAPs are described in "Machine Learning: SOAP Descriptors" and "Procedures for Regression" in Methods. The best combinations of n_{max} and l_{max}, which are hyperparameters for SOAPs (see "Machine Learning: SOAP Descriptors" in Methods), are 3 and 7 based on root mean squared errors (RMSEs) and 5 and 3 based on mean absolute errors (MAEs),

https://doi.org/10.1021/jacs.3c13574
J. Am. Chem. Soc. 2024, 146, 9697−9708

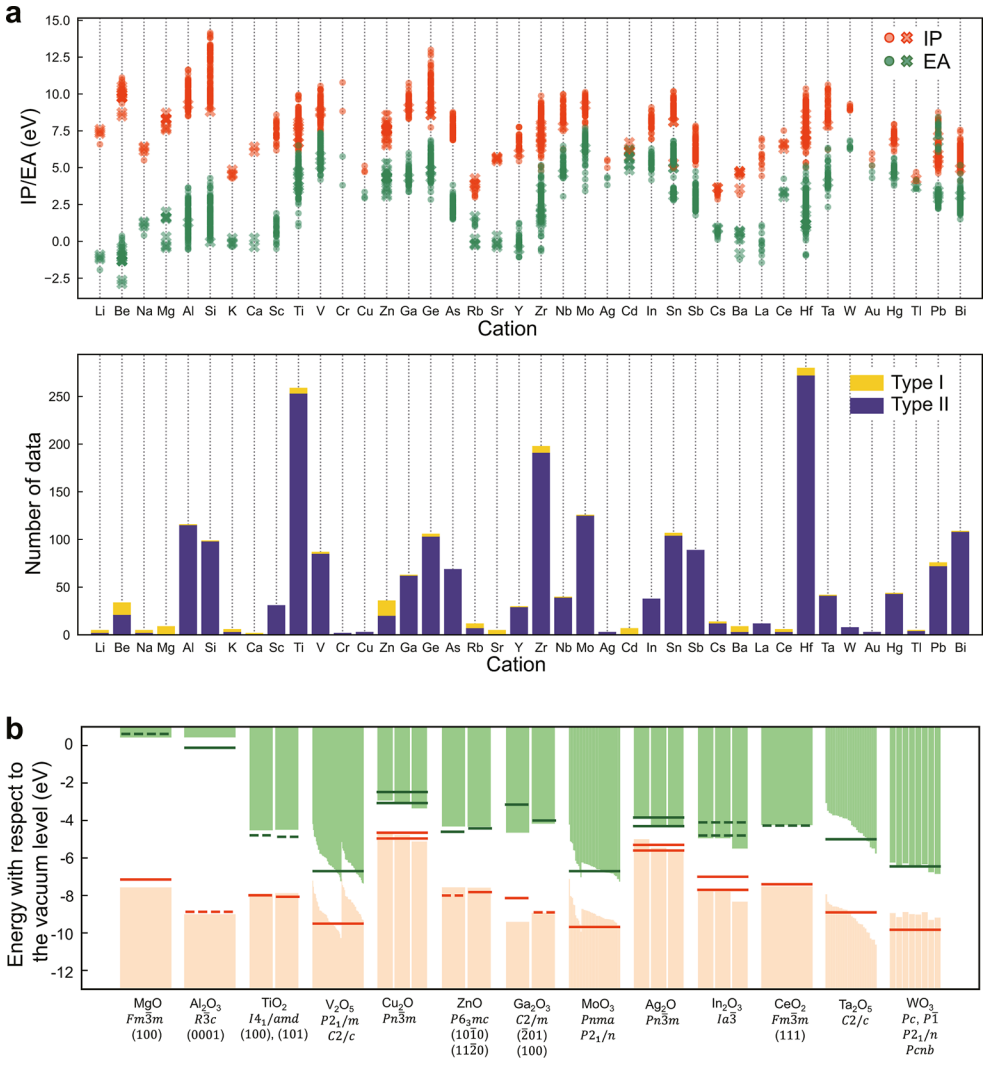


Figure 2. Distribution of theoretical IPs and EAs of binary oxides and comparison with experiments. (a) Upper panel shows the distribution of the IPs and EAs of the respective binary oxides. Orange and green dots are IPs and EAs, respectively. Cross and circle symbols are Tasker's type I and II surfaces,[64] respectively. The bottom panel is the number of surfaces, where yellow and dark blue bars are types I and II, respectively. (b) Theoretical IPs and EAs versus reported experimental values for selected binary oxides. The upper edges of the pale orange bars and the lower edges of the light green bars are calculated VBMs and CBMs with respect to the vacuum level (set at 0 eV), respectively. The orange and green solid lines are experimentally reported IPs and EAs, respectively; the dashed lines are derived by combining experimental IPs or EAs and experimental band gaps. The experimental data are taken from refs 70 and 71 for MgO, refs 72 and 73 for Al_2O_3, ref 74 for TiO_2, ref 75 for V_2O_5, ref 76 for Cu_2O, refs 77−79 for ZnO, refs 80 and 81 for Ga_2O_3, ref 82 for MoO_3, refs 83 and 84 for Ag_2O, refs 85 and 86 for In_2O_3, ref 87 for CeO_2, ref 83 for Ta_2O_5, and ref 88 for WO_3. The surface orientations have not been presented in the experimental reports for V_2O_5, Cu_2O, MoO_3, Ag_2O, In_2O_3, Ta_2O_5, and WO_3. Therefore, all theoretical IPs and EAs of binary oxides with the indicated space groups are depicted in the figure. Note that there are many types of surfaces for V_2O_5, MoO_3, and Ta_2O_5, and the bars for each surface are extremely narrow.

respectively (see Figure S1). Hereinafter, we show the results for n_{max} and l_{max} of 5 and 3, respectively; the other case is shown in Table S1. It should be noted that the results do not depend much on the combinations of n_{max} and l_{max}.

Figure 5a,b show the predicted IPs and EAs of the training data (gray dots) and test data (orange and green dots), respectively. Most of the data points are located near the diagonal lines, which means that the IPs and EAs predicted by the ANN are considerably close to the corresponding theoretical values from first-principles calculations. The coefficient of determination (R^2), RMSE, and MAE of the test data are 0.90, 0.31, and 0.22 eV for the IPs and 0.90, 0.32, and 0.23 eV for the EAs, respectively. The high prediction accuracy indicates that the ANN model has learned to differentiate the cation species through structural information, even though elemental information is not explicitly given. In addition, the errors are much smaller than the aforementioned distributions of the IPs and EAs for each cation species in the data set, which indicates proper learning of the surface structure dependence. It is noteworthy that our ANN model can also learn and predict surface energies at the level of accuracy shown in Table S2, enabling us to screen out unstable or unreasonable surfaces without explicit first-principles calculations.

Although the simple-ANN model shows valuable prediction accuracy, it has a disadvantage; that is, it equally weights surface-region atoms in a system when averaging their SOAPs. To overcome this disadvantage and further enhance the prediction accuracy, we introduced an attention layer into our ANN architecture.[49,50] Figure 3b shows the ANN architecture

https://doi.org/10.1021/jacs.3c13574
J. Am. Chem. Soc. 2024, 146, 9697−9708

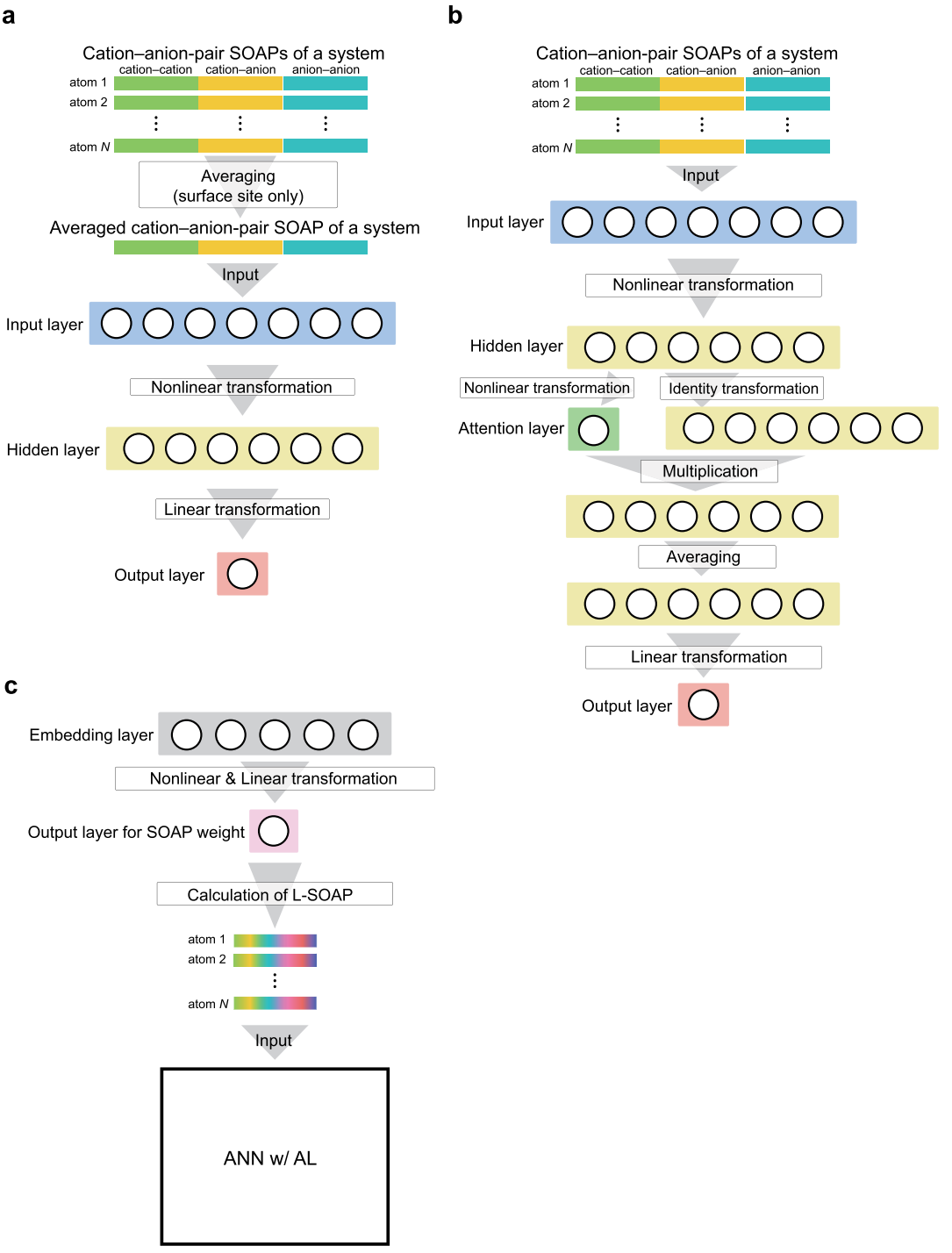


Figure 3. Architecture of ANNs. (a) Simple-ANN, (b) ANN w/AL, and (c) ANN w/L-SOAP. Each circle in the figure is a node where the input and output are scalars. Edges between nodes in two adjacent layers are fully connected but omitted for easy visualization.

with an attention layer (hereinafter called ANN w/AL). In contrast to the simple-ANN model, cation−anion-pair SOAPs of a system are inputted into the ANN without averaging and then averaged after passing through the inserted attention layer just before the output layer. The attention layer outputs a scalar value, and afterward, the value is input into a softmax function. Thus, each SOAP in a system has a weight, the sum of which is one. The transformed cation−anion-pair SOAPs are multiplied with the weights, averaged, and finally input into the output layer. As shown in Figures 5c,d, the prediction accuracy is evidently improved by introducing the attention layer, with an R^2 of 0.90, RMSE of 0.29 eV, and MAE of 0.21 eV for the IPs, and 0.93, 0.27, and 0.19 eV, respectively, for the EAs.

Furthermore, the ANN w/AL can automatically estimate the magnitude of the impact of each atom site on the IPs and EAs. Figure 5e exemplifies this feature for the case of a (001) surface of Sb_2O_3 (space group: *Pccn*). The IP prediction clearly shows the higher importance of the antimony and oxygen atoms in the vicinity of the surface, while the halves of the antimony atoms on the surface are especially relevant to the EA. The weight profiles of some other binary oxides are shown in

https://doi.org/10.1021/jacs.3c13574
J. Am. Chem. Soc. 2024, 146, 9697−9708

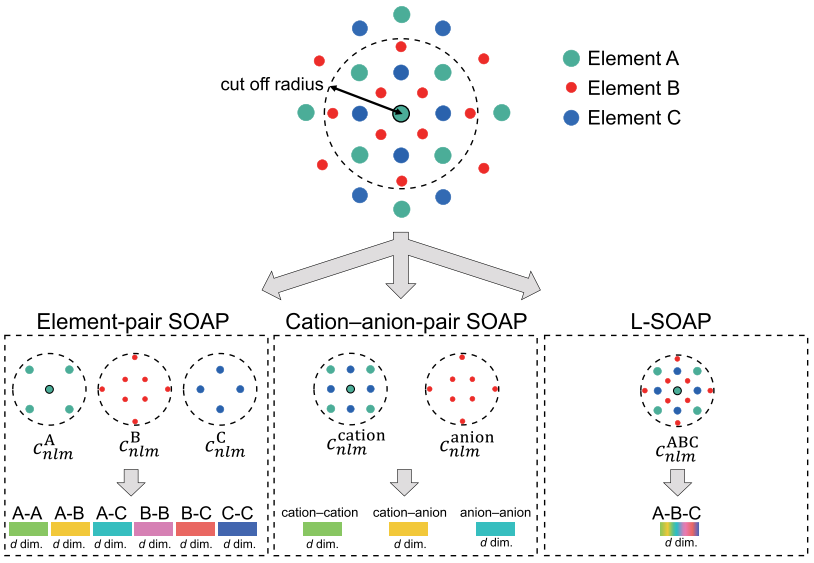


Figure 4. Schematic of conventional and learnable SOAP descriptors. The element-pair SOAPs (bottom left panel) are concatenations of the SOAPs of each elemental pair; the cation−anion-pair SOAPs (bottom center panel) are concatenations of three pairs, namely, cation−cation, cation−anion, and anion−anion combinations; and L-SOAPs (bottom right panel) have element-based learnable weights, which are automatically adjusted during ANN training.

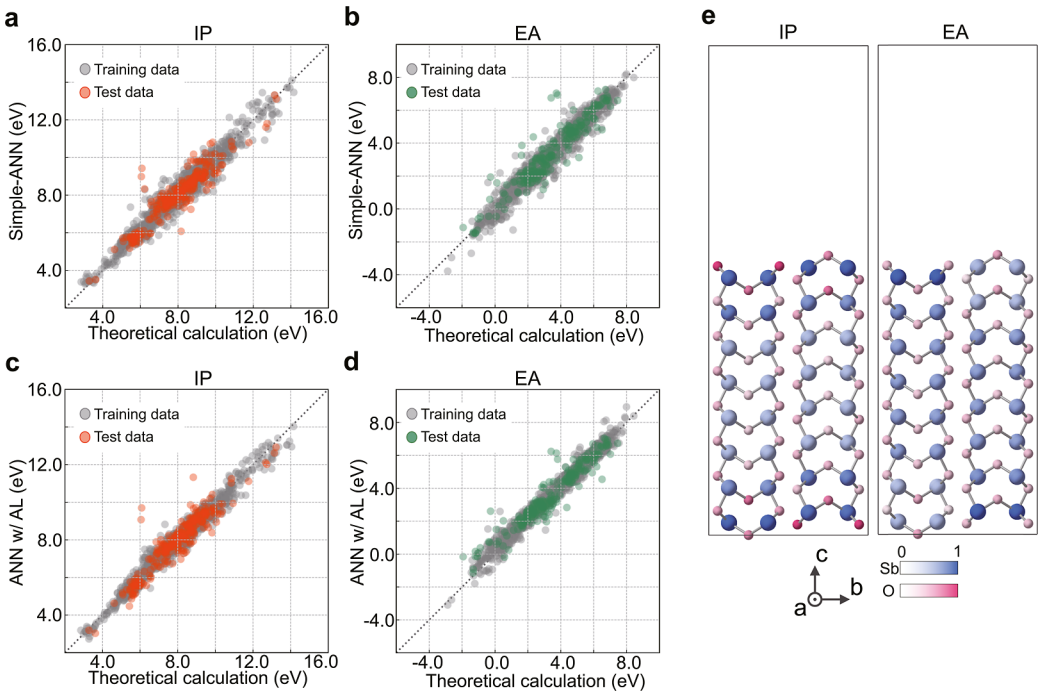


Figure 5. Theoretical and predicted IPs and EAs using simple-ANN and ANN w/AL. (a) IPs and (b) EAs obtained by first-principles calculations versus those predicted by the simple-ANN. (c) IPs and (d) EAs by first-principles calculations versus those predicted by the ANN w/AL. The orange or green and gray dots represent the test and training data, respectively. (e) Atom-site weights from the attention layer in the IP and EA prediction of a (001) surface of Sb_2O_3 whose space group is *Pccn* (index is 20 in Table S3). The frame indicates the surface supercell where the upper vacant region corresponds to the vacuum layer. Larger and smaller circles are the Sb and O atoms, respectively. The weights are represented by the shades of the atom colors: blue for Sb and pink for O. The weights are normalized so that the largest weight is one.

Figure S2. The profiles depend on the oxides and the IP or EA but have features common to the surfaces of the same oxides. Generally, the atoms on the second layers from the surfaces tend to have the smallest weights and the IPs put more importance on the surface atoms than the EAs.

Extension of SOAPs and Application to Ternary Oxides. Next, we focus on ternary systems. Figure 6a shows the theoretical IPs and EAs for 718 ternary oxide surfaces, with respect to the constituent cation species. The distributions of IPs and EAs are much larger than those in the binary systems partly because of the inclusion of multiple cation species in the ternary systems, as well as differences in the crystal and surface structures between the binary and ternary systems.

In the element-pair SOAPs, information about elements can be explicitly incorporated by concatenation of the SOAPs of each elemental pair (see bottom left panel of Figure 4), but

https://doi.org/10.1021/jacs.3c13574
J. Am. Chem. Soc. 2024, 146, 9697−9708

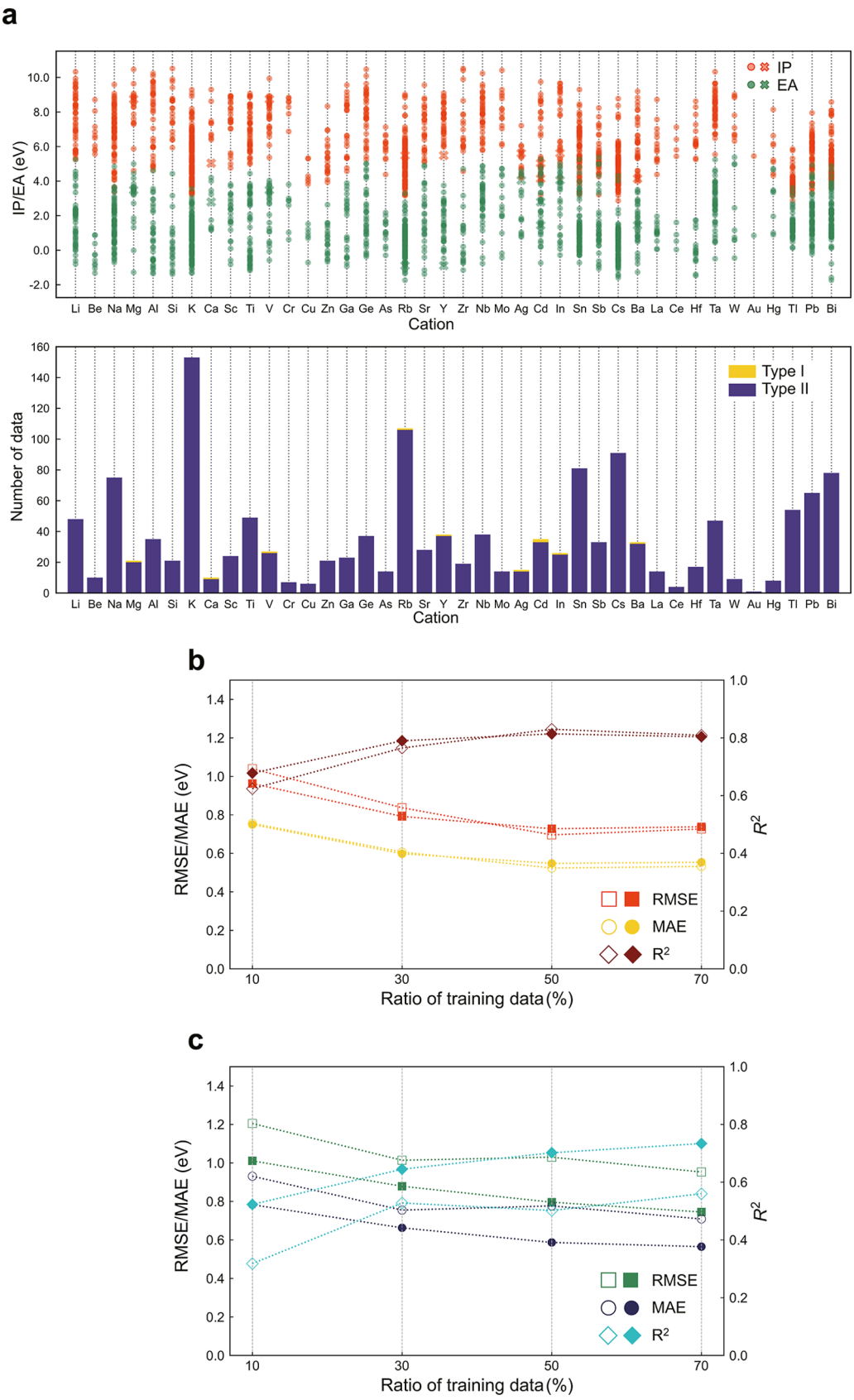


Figure 6. Distribution of theoretical IPs and EAs of ternary oxides and prediction accuracy of transfer learning. (a) Distribution of theoretical IPs and EAs. Ternary oxides include two cation species, and the same data points are shown at both cation species. The other details are the same as those for Figure 2a. (b,c) Prediction accuracy of transfer learning for IPs and EAs, respectively. The filled and open symbols are the results of the ANN w/L-SOAP and the ANN w/AL, respectively. The horizontal axis is the ratio of the ternary data for training to all ternary data.

https://doi.org/10.1021/jacs.3c13574
J. Am. Chem. Soc. 2024, 146, 9697−9708

this results in several thousands of dimensions in multispecies systems (see "Machine Learning: SOAP Descriptors" in Methods). The cation−anion-pair SOAPs have manageable dimensions but cannot differentiate cations in a system, which may be crucial for the ternary systems. Hence, we developed element-based learnable-weighted SOAPs (L-SOAPs), which have sufficient information about atomic species despite the considerably small number of dimensions. The details of the L-SOAPs are described in "Machine Learning: SOAP Descriptors" in Methods.

The new architecture of the ANN with L-SOAP (hereinafter called ANN w/L-SOAP) showed an R^2 of 0.90, RMSE of 0.31 eV, and MAE of 0.23 eV for the IPs, and 0.91, 0.29, and 0.21 eV, respectively, for the EAs of the binary systems; this is comparable to the accuracy of the ANN w/AL. It is noteworthy that the L-SOAP has only one-third of the number of dimensions of the cation−anion-pair SOAP. The number of dimensions of element-pair SOAPs rapidly increases with the number of atomic species, whereas the L-SOAP keeps the same number of dimensions. This feature of the ANN w/L-SOAP model is a great advantage for dealing with multiple-element systems.

The trained model, however, cannot be directly applied to ternary systems, unlike the simple-ANN and the ANN w/AL models, because multiple-cation effects (eq 10 in Methods) are not learned from the binary systems. Thus, we trained the models using the binary oxide data set and afterward retrained them using the ternary oxide data set. In other words, we performed transfer learning from the binary systems to the ternary systems, where all the learning parameters in the ANN w/AL were retrained using the ternary oxide data set. The ratio of the ternary data for training to all ternary data was set at 10, 30, 50, and 70%, where the residues were used as test data. The effectiveness of "transferring" is shown in Figure S3. A clear improvement in the prediction accuracy is found for transfer learning compared with learning from scratch.

Figure 6b shows the performance of the IP prediction with an n_{max} of 5 and l_{max} of 3; see Figure S4 for the parity plots and Figure S5 for the case of an n_{max} of 3 and l_{max} of 7. For comparison, the results of the transferred ANN w/AL are included in Figure 6. The ANN w/L-SOAP and the ANN w/AL exhibit almost the same levels of performance in all the ratios of training and test data. The IPs relate to the VBMs, in which oxygen 2p-orbitals are the major components in most binary oxides. This fact also applies to the ternary oxides although the weak multiple-cation effects on the VBMs exist. Because the ANN w/L-SOAP mainly learns cation−cation interactions here, both of the ANNs show similar performances regarding the IPs.

We now move on to the prediction of the EAs. The main components of the CBMs are cation orbitals that are system-dependent. Because an accurate prediction of the EAs requires learning such CBM characters, one expects that the ANN w/AL, which cannot explicitly incorporate the effects of the variety of cations, would be inaccurate. Indeed, the ANN w/AL shows considerably poor accuracy when the ratio of the training data is 10% (Figure 6c). On the other hand, the ANN w/L-SOAP is clearly more accurate at the same training/test data ratio. Surprisingly, the performance of the ANN w/L-SOAP at the 10% ratio is almost equal to that of the ANN w/AL at the 70% ratio, and the performance at the 70% ratio is comparable to that of the IPs, indicating the successful incorporation of the multiple-cation effects during training. Thus, the ANN w/L-SOAP in combination with transfer learning makes accurate and systematic prediction of both IPs and EAs of ternary systems feasible.

CONCLUSIONS

Our high-throughput first-principles calculations based on the non-self-consistent dielectric-dependent hybrid functional approach have enabled the construction of a large data set of IPs and EAs for 2195 binary and 718 ternary oxide surfaces. The accuracy of the calculated IPs and EAs has been confirmed by comparison with available experimental values for the selected binary oxides as shown in Figure 2b. Using the data set, we constructed the simple-ANN model to predict the IPs and EAs of the binary oxides. The cation−anion-pair SOAPs were used to describe surface atomic structures before structural relaxation, by simply inputting the crystal structures and surface termination planes. The IPs and EAs of relaxed surfaces were accurately predicted using the SOAP structural descriptors even though elemental information was not explicitly given. Furthermore, introducing an attention layer to the simple-ANN model allows for automatically determining relevant atoms in surface regions and enhancing the prediction performance. This feature of the attention layer can collaterally help us to analyze chemical and physical origins of target properties. Finally, we extended the SOAPs to transfer the model to the ternary oxides by introducing the element-based learnable weights to the calculation of the SOAPs and by connecting this process to the ANN w/AL. This transfer learning model based on the ANN w/L-SOAP was able to incorporate the small effects of the cations on the VBMs with the small data set in the IP prediction. Moreover, apparent improvement was demonstrated regarding the EA prediction, indicating that diverse cation effects on the CBMs were well incorporated.

The ANN w/L-SOAP developed here enables the band alignment of a vast number of metal oxide surfaces, which is fundamental to the design of various materials and devices. Moreover, it is not restricted to the prediction of surface properties of metal oxides but can be naturally applied to other multicomponent systems and other properties in the same way as crystal graph convolutional neural networks.[38] Thus, we believe that our new model paves the way for the high-throughput prediction of diverse materials and properties.

METHODS

Screening of First-Principles Calculation Data: Binary Oxides. The binary oxides considered in this work are listed in Table S3. The table contains 134 nonmetallic oxides whose band structures are formally constructed by fully occupied or empty atomic orbitals, including 41 types of atom species. To cover a wide variety of stable and metastable polymorphs, experimentally determined crystal structures of binary oxides were used as prototypes, and the cations of the prototypes were substituted with isovalent cations. The details of the cation substitution procedures are described in ref 51.

We excluded the oxides without the special isometry that nonpolar slabs should have.[52] In addition, we screened out binary oxides with rather narrow theoretical band gaps using criteria of band gaps smaller than 0.1 eV and ion-clamped static dielectric constants larger than 20, resulting in the 127 binary oxides with cation species other than those in the parentheses in Table S3. Then, we generated 3383 nonpolar surface models for the 127 binary oxides with relatively small Miller indices as listed in Table S3. To exclude unreasonable surface structures, we selected surfaces that were smoothly optimized during first-principles calculations and those with surface energies higher

https://doi.org/10.1021/jacs.3c13574
J. Am. Chem. Soc. 2024, 146, 9697−9708

than 0 J/m^2. In addition, we excluded the outliers in surface energy, which did not fulfill the following inequality

$$E_{\text{surface,mean}} - 3 \times E_{\text{surface,std}} < E_{\text{surface}} < E_{\text{surface,mean}} + 3 \times E_{\text{surface,std}} \quad (1)$$

where E_{surface} is the surface energy of the objective surface, and $E_{\text{surface,mean}}$ and $E_{\text{surface,std}}$ are the mean and standard deviations of the surface energies, respectively, in which the surfaces are generated from the common bulk system to the objective surface, resulting in the data set with 2195 binary oxide surfaces for machine learning.

Ternary Oxides. We screened and collected ternary oxide data from the Materials Project database[53] in the following manner: (1) extract the most stable binary oxides of A and B, where A and B are any of the 40 cation elements; (2) draw a convex hull of the formation energies of A−B ternary oxides with respect to the A binary oxide and the B binary oxide; (3) select the ternary oxides on the convex hull; and (4) exclude the ternary oxides without the special isometry and with band gaps smaller than 0.1 eV. We then evaluated the band gaps using non-self-consistent dielectric-dependent hybrid functional calculations and screened the ternary oxides in the same way as the binary oxides. We obtained 344 ternary oxides through these processes. Then, we made the nonpolar surface models, where the maximum Miller indices were set to be 2. These surfaces were screened in the same way as the binary cases, resulting in 718 ternary oxide surface data.

First-Principles Calculations of Band Positions. IPs and EAs are estimated as follows[16]

$$\text{IP} = (\varepsilon_{\text{vac}}^{\text{surface}} - \varepsilon_{\text{ref}}^{\text{surface,far}}) - (\varepsilon_{\text{VBM}}^{\text{bulk}} - \varepsilon_{\text{ref}}^{\text{bulk}}) \quad (2)$$

$$\text{EA} = (\varepsilon_{\text{vac}}^{\text{surface}} - \varepsilon_{\text{ref}}^{\text{surface,far}}) - (\varepsilon_{\text{CBM}}^{\text{bulk}} - \varepsilon_{\text{ref}}^{\text{bulk}}) \quad (3)$$

where $\varepsilon_{\text{vac}}^{\text{surface}}$ and $\varepsilon_{\text{ref}}^{\text{surface,far}}$ are the vacuum level and the electrostatic reference level in the bulk-like region far from the surface, respectively. These quantities are obtained by calculations using surface supercells, each of which is composed of a slab of an oxide and a vacuum layer. $\varepsilon_{\text{VBM}}^{\text{bulk}}$, $\varepsilon_{\text{CBM}}^{\text{bulk}}$, and $\varepsilon_{\text{ref}}^{\text{bulk}}$ are the VBM, CBM, and electrostatic reference levels in the corresponding bulk model, respectively. These three values are obtained by calculations using bulk primitive cells.

Computational Procedures for Bulk Systems. The calculations were performed using the projector augmented-wave (PAW) method[54] as implemented in the Vienna *ab initio* simulation package (VASP).[55,56] The $\varepsilon_{\text{VBM}}^{\text{bulk}}$ and $\varepsilon_{\text{CBM}}^{\text{bulk}}$ were calculated by the approach reported in ref 21. In this scheme, the nonlocal exchange mixing parameter in the full-range hybrid functional of the Perdew−Burke−Ernzerhof (PBE0) form[57,58] is set to be the inverse of the ion-clamped static dielectric constant of the system.[59,60] The PBE functional tuned for solids (PBEsol)[61] was used for the semilocal part of the hybrid functional. Non-self-consistent hybrid functional calculations[21,62] were performed on top of PBEsol with Hubbard U corrections to localized orbitals[63] to simultaneously attain low computational cost and high accuracy. This approach is particularly advantageous in the high-throughput evaluation of IPs and EAs as it allows for the direct alignment of the bulk hybrid functional eigenvalues with the vacuum levels in surface supercells obtained using PBEsol(+U) through the common electrostatic reference levels.[21] For the reference level $\varepsilon_{\text{ref}}^{\text{bulk}}$, we used the averaged local potentials at the oxygen sites.

PBEsol(+U) was used for obtaining optimized structures, static dielectric constants, and wave functions for the non-self-consistent dielectric-dependent hybrid functional calculations. The means of the diagonal elements of the ion-clamped static dielectric tensors obtained using the random phase approximation were taken to determine the values of the nonlocal exchange mixing parameter for respective systems.

Computational Procedures for Surfaces. We focus on high-throughput calculations of nonpolar surfaces in this study because specific treatments are necessary for modeling polar surfaces as mentioned above. We prepared slab models for nonpolar surfaces of Tasker's types I and II[64] for the binary and ternary oxides. Their maximum Miller indices were set to be those listed in Table S3 for binary oxides and 2 for ternary oxides. Each slab model has a slab layer thicker than 20 Å and a vacuum layer thicker than 15 Å, which are large enough for the electrostatic potential to converge sufficiently.[21] The convergence test results of IPs, EAs, and surface energies with respect to the vacuum and slab layer thicknesses are shown for selected systems in Figure S6. The slab models were automatically created using the code developed in ref 52. The surfaces were structurally relaxed with their lattice constants fixed by using PBEsol(+U). $\varepsilon_{\text{vac}}^{\text{surface}}$ was estimated as the electrostatic potential at the midpoint of the vacuum layer, and $\varepsilon_{\text{ref}}^{\text{surface,far}}$ was taken to be the average of the local potentials of the oxygen sites around the midpoint of the slab layer.

Computational Details. The PAW data set used in the calculations is detailed in Table S4. The plane-wave cutoff energy was set to 520 eV for the bulk structure optimization, including lattice parameter relaxation, and 400 eV for the other calculations with the lattice parameters fixed. The k-point mesh spacings for structure optimization were set to be smaller than 0.2 Å^{-1} and reduced to be less than 0.1 Å^{-1} in the calculations of the ion-clamped static dielectric tensors. The band path for estimating the VBM and CBM of each bulk system was generated using SeeK-path.[65,66] Hubbard U corrections were applied to localized orbitals by using the parameters listed in Table S4.

Machine Learning: SOAP Descriptors. For input descriptors to describe surface atomic configurations, we employed SOAPs, which are often used to construct machine learning potentials and analyze complicated systems.[67,68] In SOAPs, the positions of atoms labeled with Z (usually atomic species) are smeared with three-dimensional Gaussian functions as

$$\rho^Z(\boldsymbol{r}) = \sum_i^{|Z|} e^{-\frac{1}{2\sigma^2}|\boldsymbol{r}-\boldsymbol{R}_i|} \quad (4)$$

where $\boldsymbol{R}_i$ is the position of the i-th atom surrounding $\boldsymbol{r}$. $|Z|$ is the number of atoms labeled Z within the cutoff radius and σ is a standard deviation of the Gaussian function. The summation of the smeared atom positions is then expanded by linear combinations of radial basis functions with a cutoff radius and spherical harmonics considering the local point of interest. The coefficients are defined as

$$c_{nlm}^Z = \iiint_{\mathbb{R}^3} \mathrm{d}V g_n(|\boldsymbol{r}|) Y_{lm}(\theta, \phi) \rho^Z(\boldsymbol{r}) \quad (5)$$

where g_n and Y_{lm} are radial basis functions and spherical harmonics, respectively. The former generally depends on location $\boldsymbol{r}$ from the center and the latter on polar angle θ and azimuth ϕ in the polar coordinates around the center. n is the degree of the radial basis function; l and m are the degree and order of the spherical harmonics, respectively. The partial power spectrum vector is defined as

$$p_{nn'l}^{Z_1Z_2} = \pi \sqrt{\frac{8}{2l+1}} \sum_m c_{nlm}^{Z_1} c_{n'lm}^{Z_2} \quad (6)$$

where Z_1 and Z_2 are the symbols of elements. The element-pair and cation−anion-pair SOAPs were obtained on the basis of these procedures by using the DScribe code.[47]

In the L-SOAP, atomic positions are not separately considered for every atom species, unlike the usual SOAP procedure. Instead, we weighted the broadened atomic positions in eq 4 on every atom as

$$\rho(\boldsymbol{r}) = \sum_i^N w_i e^{-\frac{1}{2\sigma^2}|\boldsymbol{r}-\boldsymbol{R}_i|} \quad (7)$$

where w_i is its weight and N is the number of atoms within the cutoff radius. The bottom right panel of Figure 4 shows the schematic of the

https://doi.org/10.1021/jacs.3c13574
J. Am. Chem. Soc. 2024, 146, 9697−9708

L-SOAP. Here, we replaced the Gaussian function in eq 7 with the delta function for easy calculation, resulting in changing the operation of integration in eq 5 to a summation as

$$c_{nlm} = \sum_{i}^{N} g_n(|\boldsymbol{r}_i|) Y_{lm}(\theta_i, \phi_i) w_i \tag{8}$$

We connected the L-SOAPs to the ANN w/AL (see Figure 3c), which enables the model to learn the weights w_i in eq 8 during training. Atomic species described by one-hot vectors are encoded at an embedding layer to atomic feature vectors as

$$\boldsymbol{v}_i = W_{\text{emb}} \boldsymbol{x}_i \tag{9}$$

where $\boldsymbol{x}_i$ is a one-hot vector of an i-th atom in a system for describing its atomic species. W_{emb} is a weight matrix for encoding the one-hot vectors into the feature vectors $\boldsymbol{v}_i$. Here, the dimensions of the feature vectors were set to 32. Then, the weights for the L-SOAPs in eq 8 are calculated as

$$w_i = f(\boldsymbol{v}_{\text{center}}, \boldsymbol{v}_i, |\boldsymbol{r}_i|) \tag{10}$$

where $\boldsymbol{v}_{\text{center}}$ is the atomic feature vector of the centered atom and $\boldsymbol{v}_i$ is that of the atom located at $|\boldsymbol{r}_i|$ from the center. The function f is a simple neural network with two hidden layers. The coefficients and SOAPs are then calculated according to eqs 6 and 8. The rest of the processes are the same as those for the ANN w/AL.

SOAPs are typically inputted into a kernel function; however, we used nonkernel SOAPs, namely, the partial power spectrum vector in this study. The maxima of n (n_{max}) and l (l_{max}) for the radial basis functions and spherical harmonics were surveyed in combinations 3, 5, and 7. The cutoff radius for the radial basis functions was fixed to 7.0 Å and the standard deviation of the Gaussian function in eqs 4 and 7 was set at 1.0 Å. The relatively large cutoff radius was used to include the surface effects in the SOAPs for atoms on layers below the surfaces, thereby improving prediction accuracy. For instance, a smaller cutoff radius of 5.0 Å in the Simple-NN for the IP prediction shows an MAE of 0.38 eV for the test data, which is clearly less accurate than the case of 7.0 Å, 0.23 eV.

Procedures for Regression. Among various regression techniques, we selected ANNs for predicting IPs and EAs in this study because they offer flexible architecture, automatic extraction of relevant surface-region atoms, as described in Results and Discussion, and transferability to other systems.

The simple-ANN shown in Figure 3a has a typical architecture with input, hidden, and output layers. The input vectors are averaged SOAPs of a system. Its hyperparameters, e.g., the number of hidden layers, are listed in Table S5. The rectified linear unit was used as an activation function, and the dropout rate was fixed at 0.5 in hidden layers that were not linked with the output layer.

The data sets used in the regression were divided into training, validation, and test data sets with a ratio of 8:1:1, where we did not care about the distribution of the cation species in respective data sets. Fivefold cross validation using the validation data sets was performed for tuning the hyperparameters of the ANNs. We used back-propagation based on the Adam scheme[69] to optimize all the learning parameters in the network, thereby minimizing the mean squared errors between the model outputs and the training targets. The learning rate for the optimization was 0.001, and the maximum epoch was 200. The training was terminated when the prediction accuracy of the validation data was converged to be almost constant as exemplified by the IP learning curve for the simple-NN in Figure S7. All of the training converged well within 200 epochs.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.3c13574.

RMSE and MAE of all the considered combinations of the SOAP hyperparameters; profile of the weights from the attention layer; transfer learning versus learning from scratch; theoretical and predicted IPs and EAs for the ternary data sets; convergence of IPs, EAs, and surface energies; learning curve for IPs; prediction accuracies; 134 prototypical binary oxides and surface orientation; PAW datasets and Hubbard U parameters; and hyperparameters of the simple-ANN (PDF)

Theoretical IPs and EAs of the 2195 binary and 718 ternary oxide surfaces and related properties. Database contains the index, system (binary or ternary), chemical formula, space group number, Miller index, surface energy, IP, EA, and band gap (bulk) (XLSX)

Supercells before and after structural optimization for the 2195 binary and 718 ternary oxide surfaces (ZIP)

■ AUTHOR INFORMATION

Corresponding Authors

Shin Kiyohara − *Laboratory for Materials and Structures, Institute of Innovative Research, Tokyo Institute of Technology, Yokohama 226-8501, Japan; Institute for Materials Research, Tohoku University, Sendai 980-8577, Japan;* orcid.org/0000-0003-2890-5760; Email: shin.kiyohara.d6@tohoku.ac.jp

Fumiyasu Oba − *Laboratory for Materials and Structures, Institute of Innovative Research, Tokyo Institute of Technology, Yokohama 226-8501, Japan; MDX Research Center for Element Strategy, International Research Frontiers Initiative, Tokyo Institute of Technology, Yokohama 226-8501, Japan;* orcid.org/0000-0001-7178-5333; Email: oba@msl.titech.ac.jp

Author

Yoyo Hinuma − *Department of Energy and Environment, National Institute of Advanced Industrial Science and Technology (AIST), Ikeda, Osaka 563-8577, Japan;* orcid.org/0000-0003-2272-1178

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.3c13574

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI (grant nos. JP20J00773, JP20H00302, and JP23K13811), KISTEC Project, MEXT Data Creation and Utilization Type Material Research and Development Project (grant no. JPMXP1122683430), and JST CREST (grant no. JPMJCR17J2), Japan. The computing resources of Academic Center for Computing and Media Studies at Kyoto University and Research Institute for Information Technology at Kyushu University were used for this work.

■ REFERENCES

(1) Moniz, S. J. A.; Shevlin, S. A.; Martin, D. J.; Guo, Z.-X.; Tang, J. Visible-Light Driven Heterojunction Photocatalysts for Water Splitting − a Critical Review. *Energy Environ. Sci.* **2015**, *8* (3), 731−759.

(2) Scanlon, D. O.; Dunnill, C. W.; Buckeridge, J.; Shevlin, S. A.; Logsdail, A. J.; Woodley, S. M.; Catlow, C. R. A.; Powell, M. J.; Palgrave, R. G.; Parkin, I. P.; Watson, G. W.; Keal, T. W.; Sherwood, P.; Walsh, A.; Sokol, A. A. Band Alignment of Rutile and Anatase TiO_2. *Nat. Mater.* **2013**, *12* (9), 798−801.

(3) Zhang, Z.; Yates, J. T. Band Bending in Semiconductors: Chemical and Physical Consequences at Surfaces and Interfaces. *Chem. Rev.* **2012**, *112* (10), 5520−5551.
(4) Polman, A.; Knight, M.; Garnett, E. C.; Ehrler, B.; Sinke, W. C. Photovoltaic Materials: Present Efficiencies and Future Challenges. *Science* **2016**, *352* (6283), aad4424.
(5) Green, M. A.; Bremner, S. P. Energy Conversion Approaches and Materials for High-Efficiency Photovoltaics. *Nat. Mater.* **2017**, *16* (1), 23−34.
(6) Chen, J.; Park, N.-G. Materials and Methods for Interface Engineering toward Stable and Efficient Perovskite Solar Cells. *ACS Energy Lett.* **2020**, *5* (8), 2742−2786.
(7) Robertson, J. Band Offsets, Schottky Barrier Heights, and Their Effects on Electronic Devices. *J. Vac. Sci. Technol., A* **2013**, *31* (5), 050821.
(8) Ponce, F. A.; Bour, D. P. Nitride-Based Semiconductors for Blue and Green Light-Emitting Devices. *Nature* **1997**, *386* (6623), 351−359.
(9) Mannhart, J.; Schlom, D. G. Oxide Interfaces—An Opportunity for Electronics. *Science* **2010**, *327* (5973), 1607−1611.
(10) Janotti, A.; Van de Walle, C. G. Absolute Deformation Potentials and Band Alignment of Wurtzite ZnO, MgO, and CdO. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2007**, *75* (12), 121201.
(11) Li, Y.-H.; Walsh, A.; Chen, S.; Yin, W.-J.; Yang, J.-H.; Li, J.; Da Silva, J. L. F.; Gong, X. G.; Wei, S.-H. Revised *Ab Initio* Natural Band Offsets of All Group IV, II-VI, and III-V Semiconductors. *Appl. Phys. Lett.* **2009**, *94* (21), 212109.
(12) Cho, S.; Kim, S.; Kim, J. H.; Zhao, J.; Seok, J.; Keum, D. H.; Baik, J.; Choe, D.; Chang, K. J.; Suenaga, K.; Kim, S. W.; Lee, Y. H.; Yang, H. Phase Patterning for Ohmic Homojunction Contact in $MoTe_2$. *Science* **2015**, *349* (6248), 625−628.
(13) Walzer, K.; Maennig, B.; Pfeiffer, M.; Leo, K. Highly Efficient Organic Devices Based on Electrically Doped Transport Layers. *Chem. Rev.* **2007**, *107* (4), 1233−1271.
(14) Cahen, D.; Kahn, A. Electron Energetics at Surfaces and Interfaces: Concepts and Experiments. *Adv. Mater.* **2003**, *15* (4), 271−277.
(15) Stevanović, V.; Hartman, K.; Jaramillo, R.; Ramanathan, S.; Buonassisi, T.; Graf, P. Variations of Ionization Potential and Electron Affinity as a Function of Surface Orientation: The Case of Orthorhombic SnS. *Appl. Phys. Lett.* **2014**, *104* (21), 211603.
(16) Hinuma, Y.; Grüneis, A.; Kresse, G.; Oba, F. Band Alignment of Semiconductors from Density-Functional Theory and Many-Body Perturbation Theory. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *90* (15), 155405.
(17) Chen, W.; Pasquarello, A. Band-Edge Positions in *GW*: Effects of Starting Point and Self-Consistency. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *90* (16), 165133.
(18) Moses, P. G.; Miao, M.; Yan, Q.; Van de Walle, C. G. Hybrid Functional Investigations of Band Gaps and Band Alignments for AlN, GaN, InN, and InGaN. *J. Chem. Phys.* **2011**, *134* (8), 084703.
(19) Komsa, H.-P.; Broqvist, P.; Pasquarello, A. Alignment of Defect Levels and Band Edges through Hybrid Functionals: Effect of Screening in the Exchange Term. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *81* (20), 205118.
(20) Oba, F.; Kumagai, Y. Design and Exploration of Semiconductors from First Principles: A Review of Recent Advances. *Appl. Phys. Express* **2018**, *11* (6), 060101.
(21) Hinuma, Y.; Kumagai, Y.; Tanaka, I.; Oba, F. Band Alignment of Semiconductors and Insulators Using Dielectric-Dependent Hybrid Functionals: Toward High-Throughput Evaluation. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2017**, *95* (7), 075302.
(22) Butler, K. T.; Hendon, C. H.; Walsh, A. Electronic Chemical Potentials of Porous Metal−Organic Frameworks. *J. Am. Chem. Soc.* **2014**, *136* (7), 2703−2706.
(23) Jacobs, R.; Booske, J.; Morgan, D. Understanding and Controlling the Work Function of Perovskite Oxides Using Density Functional Theory. *Adv. Funct. Mater.* **2016**, *26* (30), 5471−5482.
(24) Grüneis, A.; Kresse, G.; Hinuma, Y.; Oba, F. Ionization Potentials of Solids: The Importance of Vertex Corrections. *Phys. Rev. Lett.* **2014**, *112* (9), 096401.
(25) Ping, Y.; Rocca, D.; Galli, G. Electronic Excitations in Light Absorbers for Photoelectrochemical Energy Conversion: First Principles Calculations Based on Many Body Perturbation Theory. *Chem. Soc. Rev.* **2013**, *42* (6), 2437.
(26) Deacon-Smith, D. E. E.; Scanlon, D. O.; Catlow, C. R. A.; Sokol, A. A.; Woodley, S. M. Interlayer Cation Exchange Stabilizes Polar Perovskite Surfaces. *Adv. Mater.* **2014**, *26* (42), 7252−7256.
(27) Setvin, M.; Reticcioli, M.; Poelzleitner, F.; Hulva, J.; Schmid, M.; Boatner, L. A.; Franchini, C.; Diebold, U. Polarity Compensation Mechanisms on the Perovskite Surface $KTaO_3$ (001). *Science* **2018**, *359* (6375), 572−575.
(28) Enterkin, J. A.; Subramanian, A. K.; Russell, B. C.; Castell, M. R.; Poeppelmeier, K. R.; Marks, L. D. A Homologous Series of Structures on the Surface of $SrTiO_3$ (110). *Nat. Mater.* **2010**, *9* (3), 245−248.
(29) Lazzeri, M.; Selloni, A. Stress-Driven Reconstruction of an Oxide Surface: The Anatase TiO_2 (001)-(1 × 4) Surface. *Phys. Rev. Lett.* **2001**, *87* (26), 266105.
(30) Zhu, Q.; Li, L.; Oganov, A. R.; Allen, P. B. Evolutionary Method for Predicting Surface Reconstructions with Variable Stoichiometry. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87* (19), 195317.
(31) Wanzenböck, R.; Arrigoni, M.; Bichelmaier, S.; Buchner, F.; Carrete, J.; Madsen, G. K. H. Neural-Network-Backed Evolutionary Search for $SrTiO_3$ (110) Surface Reconstructions. *Digital Discovery* **2022**, *1* (5), 703−710.
(32) Mochizuki, Y.; Sung, H.-J.; Gake, T.; Oba, F. Chemical Trends of Surface Reconstruction and Band Positions of Nonmetallic Perovskite Oxides from First Principles. *Chem. Mater.* **2023**, *35* (5), 2047−2057.
(33) Kim, S.; Sinai, O.; Lee, C.-W.; Rappe, A. M. Controlling Oxide Surface Dipole and Reactivity with Intrinsic Nonstoichiometric Epitaxial Reconstructions. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2015**, *92* (23), 235431.
(34) Stanev, V.; Oses, C.; Kusne, A. G.; Rodriguez, E.; Paglione, J.; Curtarolo, S.; Takeuchi, I. Machine Learning Modeling of Superconducting Critical Temperature. *npj Comput. Mater.* **2018**, *4* (1), 29.
(35) Kiyohara, S.; Oda, H.; Miyata, T.; Mizoguchi, T. Prediction of Interface Structures and Energies via Virtual Screening. *Sci. Adv.* **2016**, *2* (11), 1600746.
(36) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K. R. SchNet − A Deep Learning Architecture for Molecules and Materials. *J. Chem. Phys.* **2018**, *148* (24), 241722.
(37) Freeze, J. G.; Kelly, H. R.; Batista, V. S. Search for Catalysts by Inverse Design: Artificial Intelligence, Mountain Climbers, and Alchemists. *Chem. Rev.* **2019**, *119* (11), 6595−6612.
(38) Xie, T.; Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.* **2018**, *120* (14), 145301.
(39) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, *559* (7715), 547−555.
(40) Kang, S.; Jeong, W.; Hong, C.; Hwang, S.; Yoon, Y.; Han, S. Accelerated Identification of Equilibrium Structures of Multicomponent Inorganic Crystals Using Machine Learning Potentials. *npj Comput. Mater.* **2022**, *8* (1), 108.
(41) Shen, C.; Li, T.; Zhang, Y.; Xie, R.; Long, T.; Fortunato, N. M.; Liang, F.; Dai, M.; Shen, J.; Wolverton, C. M.; Zhang, H. Accelerated Screening of Ternary Chalcogenides for Potential Photovoltaic Applications. *J. Am. Chem. Soc.* **2023**, *145* (40), 21925−21936.
(42) Hwang, S.; Jung, J.; Hong, C.; Jeong, W.; Kang, S.; Han, S. Stability and Equilibrium Structures of Unknown Ternary Metal Oxides Explored by Machine-Learned Potentials. *J. Am. Chem. Soc.* **2023**, *145* (35), 19378−19386.

https://doi.org/10.1021/jacs.3c13574
J. Am. Chem. Soc. 2024, 146, 9697−9708

(43) Kim, M.; Yeo, B. C.; Park, Y.; Lee, H. M.; Han, S. S.; Kim, D. Artificial Intelligence to Accelerate the Discovery of N_2 Electroreduction Catalysts. *Chem. Mater.* **2020**, *32* (2), 709−720.
(44) Chen, C.; Ong, S. P. A Universal Graph Deep Learning Interatomic Potential for the Periodic Table. *Nat. Comput. Sci.* **2022**, *2* (11), 718−728.
(45) Bartók, A. P.; Kondor, R.; Csányi, G. On Representing Chemical Environments. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87* (18), 184115.
(46) Dulub, O.; Diebold, U.; Kresse, G. Novel Stabilization Mechanism on Polar Surfaces: ZnO (0001)-Zn. *Phys. Rev. Lett.* **2003**, *90* (1), 016102.
(47) Himanen, L.; Jäger, M. O. J.; Morooka, E. V.; Canova, F. F.; Ranawat, Y. S.; Gao, D. Z.; Rinke, P.; Foster, A. S. DScribe: Library of Descriptors for Machine Learning in Materials Science. *Comput. Phys. Commun.* **2020**, *247*, 106949.
(48) Hinuma, Y.; Kamachi, T.; Hamamoto, N. Algorithm for Automatic Detection of Surface Atoms. *Trans. Mater. Res. Soc. Jpn.* **2020**, *45* (4), 115−120.
(49) Goodall, R. E. A.; Lee, A. A. Predicting Materials Properties without Crystal Structure: Deep Representation Learning from Stoichiometry. *Nat. Commun.* **2020**, *11* (1), 6280.
(50) Wang, A. Y. T.; Kauwe, S. K.; Murdock, R. J.; Sparks, T. D. Compositionally Restricted Attention-Based Network for Materials Property Predictions. *npj Comput. Mater.* **2021**, *7* (1), 77.
(51) Hinuma, Y.; Hayashi, H.; Kumagai, Y.; Tanaka, I.; Oba, F. Comparison of Approximations in Density Functional Theory Calculations: Energetics and Structure of Binary Oxides. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2017**, *96* (9), 094102.
(52) Hinuma, Y.; Kumagai, Y.; Oba, F.; Tanaka, I. Categorization of Surface Polarity from a Crystallographic Approach. *Comput. Mater. Sci.* **2016**, *113*, 221−230.
(53) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, *1* (1), 011002.
(54) Blöchl, P. E. Projector Augmented-Wave Method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *50* (24), 17953−17979.
(55) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for *Ab Initio* Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54* (16), 11169−11186.
(56) Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59* (3), 1758−1775.
(57) Perdew, J. P.; Ernzerhof, M.; Burke, K. Rationale for Mixing Exact Exchange with Density Functional Approximations. *J. Chem. Phys.* **1996**, *105* (22), 9982−9985.
(58) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110* (13), 6158−6170.
(59) Skone, J. H.; Govoni, M.; Galli, G. Self-Consistent Hybrid Functional for Condensed Systems. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89* (19), 195112.
(60) Gerosa, M.; Bottani, C. E.; Caramella, L.; Onida, G.; Di Valentin, C.; Pacchioni, G. Electronic Structure and Phase Stability of Oxide Semiconductors: Performance of Dielectric-Dependent Hybrid Functional DFT, Benchmarked against *GW* Band Structure Calculations and Experiments. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2015**, *91* (15), 155201.
(61) Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E.; Constantin, L. A.; Zhou, X.; Burke, K. Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces. *Phys. Rev. Lett.* **2008**, *100* (13), 136406.
(62) Tran, F. On the Accuracy of the Non-Self-Consistent Calculation of the Electronic Structure of Solids with Hybrid Functionals. *Phys. Lett. A* **2012**, *376* (6−7), 879−882.
(63) Dudarev, S. L.; Botton, G. A.; Savrasov, S. Y.; Humphreys, C. J.; Sutton, A. P. Electron-Energy-Loss Spectra and the Structural Stability of Nickel Oxide: An LSDA+*U* Study. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *57* (3), 1505−1509.
(64) Tasker, P. W. The Stability of Ionic Crystal Surfaces. *J. Phys. C Solid State Phys.* **1979**, *12* (22), 4977−4984.
(65) Hinuma, Y.; Pizzi, G.; Kumagai, Y.; Oba, F.; Tanaka, I. Band Structure Diagram Paths Based on Crystallography. *Comput. Mater. Sci.* **2017**, *128*, 140−184.
(66) Togo, A.; Tanaka, I. Spglib: A Software Library for Crystal Symmetry Search. *arXiv* **2018**, arXiv:1808.01590 [cond-mat.mtrl-sci].
(67) Bartók, A. P.; Kermode, J.; Bernstein, N.; Csányi, G. Machine Learning a General-Purpose Interatomic Potential for Silicon. *Phys. Rev. X* **2018**, *8* (4), 041048.
(68) De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M. Comparing Molecules and Solids across Structural and Alchemical Space. *Phys. Chem. Chem. Phys.* **2016**, *18* (20), 13754−13769.
(69) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv*, **2016**, arXiv:1412.6980[cs.LG].
(70) Jaouen, T.; Jézéquel, G.; Delhaye, G.; Lépine, B.; Turban, P.; Schieffer, P. Work Function Shifts, Schottky Barrier Height, and Ionization Potential Determination of Thin MgO Films on Ag (001). *Appl. Phys. Lett.* **2010**, *97* (23), 232104.
(71) Roessler, D. M.; Walker, W. C. Electronic Spectrum and Ultraviolet Optical Properties of Crystalline MgO. *Phys. Rev.* **1967**, *159* (3), 733−738.
(72) Käämbre, H. A comment on "The Effective Electron Affinity Estimation from the Simultaneous Detection of Thermally Stimulated Luminescence and Exoelectronic Emission. Application to an α-Alumina Single Crystal". *J. Phys. D Appl. Phys.* **1997**, *30*, 1961−1962.
(73) Innocenzi, M. E.; Swimm, R. T.; Bass, M.; French, R. H.; Villaverde, A. B.; Kokta, M. R. Room-Temperature Optical Absorption in Undoped α-Al_2O_3. *J. Appl. Phys.* **1990**, *67* (12), 7542−7546.
(74) Kashiwaya, S.; Morasch, J.; Streibel, V.; Toupance, T.; Jaegermann, W.; Klein, A. The Work Function of TiO_2. *Surfaces* **2018**, *1* (1), 73−89.
(75) Meyer, J.; Hamwi, S.; Kröger, M.; Kowalsky, W.; Riedl, T.; Kahn, A. Transition Metal Oxides for Organic Electronics: Energetics, Device Physics and Applications. *Adv. Mater.* **2012**, *24* (40), 5408−5427.
(76) Chu, C.-Y.; Huang, M. H. Facet-Dependent Photocatalytic Properties of Cu_2O Crystals Probed by Using Electron, Hole and Radical Scavengers. *J. Mater. Chem. A* **2017**, *5* (29), 15116−15123.
(77) Jacobi, K.; Zwicker, G.; Gutmann, A. Work Function, Electron Affinity and Band Bending of Zinc Oxide Surfaces. *Surf. Sci.* **1984**, *141* (1), 109−125.
(78) Mönch, W. *Semiconductor Surfaces and Interfaces*, 3rd ed.; Springer Series in Surface Sciences; Springer Berlin Heidelberg: Berlin, Heidelberg, 2001; Vol. *26*.
(79) McLeod, J. A.; Wilks, R. G.; Skorikov, N. A.; Finkelstein, L. D.; Abu-Samak, M.; Kurmaev, E. Z.; Moewes, A. Band Gaps and Electronic Structure of Alkaline-Earth and Post-Transition-Metal Oxides. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *81* (24), 245123.
(80) Swinnich, E.; Hasan, M. N.; Zeng, K.; Dove, Y.; Singisetti, U.; Mazumder, B.; Seo, J. H. Flexible β-Ga_2O_3 Nanomembrane Schottky Barrier Diodes. *Adv. Electron. Mater.* **2019**, *5* (3), 1−8.
(81) Mohamed, M.; Irmscher, K.; Janowitz, C.; Galazka, Z.; Manzke, R.; Fornari, R. Schottky Barrier Height of Au on the Transparent Semiconducting Oxide β-Ga_2O_3. *Appl. Phys. Lett.* **2012**, *101* (13), 132106.
(82) Kröger, M.; Hamwi, S.; Meyer, J.; Riedl, T.; Kowalsky, W.; Kahn, A. P-Type Doping of Organic Wide Band Gap Materials by Transition Metal Oxides: A Case-Study on Molybdenum Trioxide. *Org. Electron.* **2009**, *10* (5), 932−938.
(83) Greiner, M. T.; Helander, M. G.; Tang, W.-M.; Wang, Z.-B.; Qiu, J.; Lu, Z.-H. Universal Energy-Level Alignment of Molecules on Metal Oxides. *Nat. Mater.* **2012**, *11* (1), 76−81.

https://doi.org/10.1021/jacs.3c13574
J. Am. Chem. Soc. 2024, 146, 9697−9708

(84) Lei, Y.; Lu, X. The Decisive Effect of Interface States on the Photocatalytic Activity of the Silver(I) Oxide/Titanium Dioxide Heterojunction. *J. Colloid Interface Sci.* **2017**, *492*, 167−175.
(85) Hohmann, M. V.; Ágoston, P.; Wachau, A.; Bayer, T. J. M.; Brötz, J.; Albe, K.; Klein, A. Orientation Dependent Ionization Potential of In_2O_3: A Natural Source for Inhomogeneous Barrier Formation at Electrode Interfaces in Organic Electronics. *J. Phys.: Condens. Matter* **2011**, *23* (33), 334203.
(86) Walsh, A.; Da Silva, J. L. F.; Wei, S.-H.; Körber, C.; Klein, A.; Piper, L. F. J.; DeMasi, A.; Smith, K. E.; Panaccione, G.; Torelli, P.; Payne, D. J.; Bourlange, A.; Egdell, R. G. Nature of the Band Gap of In_2O_3 Revealed by First-Principles Calculations and X-Ray Spectroscopy. *Phys. Rev. Lett.* **2008**, *100* (16), 167402.
(87) Wardenga, H. F. *Surface Potentials of Ceria and Their Influence on the Surface Exchange of Oxygen*; Technische Universität Darmstadt: Darmstadt, Germany, 2019.
(88) Meyer, J.; Kröger, M.; Hamwi, S.; Gnam, F.; Riedl, T.; Kowalsky, W.; Kahn, A. Charge Generation Layers Comprising Transition Metal-Oxide/Organic Interfaces: Electronic Structure and Charge Generation Mechanism. *Appl. Phys. Lett.* **2010**, *96* (19), 1−4.

https://doi.org/10.1021/jacs.3c13574
J. Am. Chem. Soc. 2024, 146, 9697−9708