THE JOURNAL OF PHYSICAL CHEMISTRY C
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

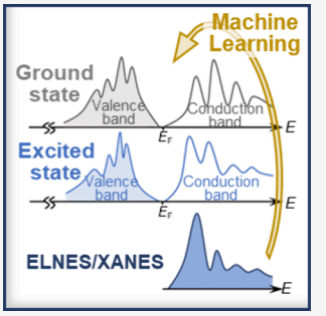
pubs.acs.org/JPCC
Article

Toward the Atomic-Level Analysis of Ground-State Electronic Structures of Solid Materials via Prediction from Core-Loss Spectra: A Computational Study in Si

Izumi Takahara,* Fumihiko Uesugi, Koji Kimoto, Kiyou Shibata, and Teruyasu Mizoguchi*

Cite This: J. Phys. Chem. C 2024, 128, 13500−13507
Read Online

ABSTRACT: Local electronic structure in the ground state is essential for understanding the stability and properties of materials. Core-loss spectroscopy using electron or X-ray provides insights into the local electronic structure, but directly observable information is limited to the partial density of state (PDOS) of the conduction band at the excited state. To overcome this limitation, we developed a machine learning (ML) approach by creating a database of Si-*K* core-loss spectra and corresponding ground-state PDOS for various silicon structures. Using this database, we constructed an ML model capable of predicting the atom-specific ground-state PDOS of the valence and conduction bands from Si-*K* edges. Our model demonstrated the ability of the ML to extract the complex correlation between ground-state PDOS and Si-*K* edges. This study provides crucial insights into achieving atomic-level analysis of ground-state electronic structures, paving the way for a deeper understanding of stability and properties of materials.



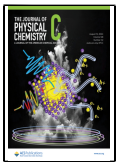
INTRODUCTION

Understanding the local electronic structure is essential for comprehending the structures and properties of materials. One of the approaches to access the local electronic structure is the density functional theory (DFT)[1,2] calculations, where we can analyze the relationship between the atomic and electronic structures in detail. Furthermore, by constructing machine learning (ML) models on DFT data, we can rapidly predict and analyze the electronic structure from the corresponding atomic structures.[3−5] For instance, the prediction of the electronic structure of silicon from local structure has been reported,[3] but when one wants to experimentally analyze the electronic structure, it is difficult to fully determine the atomic structure, making predictions based on the atomic structure challenging. Moreover, it is desirable to directly extract information about the electronic structure from the measurement data without going through the process of structure determination. Among various characterization techniques promising for this purpose, core−electron spectroscopy, specifically electron energy loss near edge structure (ELNES) and X-ray absorption near-edge structure (XANES), stands out as a powerful method.[6,7] ELNES and XANES capture transitions of core electrons to conduction bands, thereby providing insights into the electronic states of specific atoms. This enables the extraction of elemental and chemical information on the atom through the spectrum analysis, such as fingerprint verification and first-principles simulations based on DFT.[7−9] Previous studies have utilized ELNES for atomic-level elemental mapping, oxidization-state mapping, and chemical-bonding mapping, establishing it as a powerful analytical tool for characterizing local states.[10−13] Moreover, XANES provides exceptional detection sensitivity and temporal resolution, enabling the investigation of chemical reactions in battery materials and beyond.[14,15]

On the other hand, the ELNES/XANES spectroscopic method has a significant limitation. The ELNES/XANES reflects the excitation of core electrons to the conduction bands, which is a process that generates core holes and consequently shifts electron orbitals toward lower energy levels. This implies that the orbitals involved in transitions are not in their ground state but in their excited state, making the ELNES/XANES spectral shapes reflect the DOS of the conduction band in the excited state. Furthermore, since the electronic transition obeys the selection rule of dipole transitions, the change in an orbital's angular momentum number is limited to $\Delta l = \pm 1$. Therefore, the ELNES/XANES does not provide information on all orbitals but yields only limited PDOS related to the specific orbitals. In other words, it is not possible to obtain the electronic structure of both the valence and conduction bands for all orbitals in the ground state directly from the ELNES/XANES, while the ground-state electronic structure is more crucial for understanding the stability and properties of the materials.

Traditionally, to obtain information on the electronic states, including both the valence and conduction bands at the

Received: April 30, 2024
Revised: July 3, 2024
Accepted: July 5, 2024
Published: August 2, 2024



https://doi.org/10.1021/acs.jpcc.4c02818
J. Phys. Chem. C 2024, 128, 13500−13507

ground state from ELNES/XANES, it has been necessary to compare the experimental spectra with theoretically simulated spectra or reference spectra followed by further analysis through theoretical calculations. On the other hand, in recent years, the effectiveness of employing ML techniques in materials characterization has been demonstrated.[16] ML has also increasingly been utilized in extracting information from ELNES/XANES spectra, with methods reported for directly predicting atomic structure and material properties from ELNES/XANES.[17−25] Furthermore, it has recently been reported that it is possible to predict the electronic structure of organic molecules in their ground state from C-*K* edge ELNES/XANES.[26]

Building upon prior research employing ML methods, we aim to develop an ML-model to predict the atom-specific electronic structure of all orbitals in their ground state for solid materials from ELNES/XANES, toward the realization of atomic-level analysis of ground-state electronic structures. In this study, we focus on the Si-*K* edge of Si. Si is an indispensable material in industrial applications despite its simplicity, and understanding the electronic structure with a high spatial resolution is of great importance. In Si, the excitation energy of a core electron from the 1*s* orbital to the conduction bands is large, and the Si-*K* edge is located around 1850 eV. It is known that the magnitude of excitation energy correlates with the locality of the information obtained by EELS, and higher excitation energy results in the higher spatial resolution at which the signal can be acquired. Therefore, compared to other typical edges that appear at lower energies, such as the Si-L_{23} edge, the use of the Si-*K* edge is expected to enable the atomic-level analysis of ground-state electronic structures. In order to theoretically investigate the feasibility, we constructed a database of Si-*K* ELNES/XANES and the corresponding ground-state PDOS for various structures of Si by using DFT simulations. Utilizing this database, we developed an ML model based on a one-dimensional convolutional neural network to predict the ground-state PDOS from the Si-*K* edge. Si-*K* edge involves the excitation of the 1*s* core electron to the *p*-orbitals of the conduction band, thus providing the *p*-orbitals of the conduction band at the excited state. We demonstrate that an ML model can capture the complex correlation between Si-*K* edge and the ground-state PDOS, thereby enabling the prediction of PDOS for not only transition-allowed *p*-orbitals but also transition-prohibited *s*-orbitals, not only of conduction band but also of the valence band. Additionally, by conducting a comparison with template matching, we demonstrate the utility of our ML based approach.

METHODOLOGY

The data sets for ELNES/XANES and the corresponding ground-state PDOS were all generated using first-principles calculations based on DFT.[1,2] First, we retrieved all the structures of Si registered on the Materials Project,[27] which resulted in 41 structures. Using the "SpaceGroupAnalyzer" class in the "pymatgen.symmetry" package,[28] we searched for the primitive cell across all these structures and then enumerated all the inequivalent Si sites in all the Si structures, identifying 655 Si sites. To address the scarcity of data, data augmentation was conducted in an *on-the-fly* manner, ultimately yielding data for 1,405 Si sites. Detailed procedure for the data augmentation process is provided in the Supporting Information B.

For all the Si sites, the Si-*K* edge and the corresponding ground-state PDOS were calculated using the CASTEP[29] code and the OptaDOS[30] code. For the relaxation of electronic structures, the plane-wave basis pseudopotential method was employed as implemented in the CASTEP code, and the exchange-correlation interactions were approximated using GGA-PBE.[31] The cutoff-energy for the plane-wave expansion was set to 500 eV. In the calculation of ELNES/XANES, the core-hole effect was considered by generating the pseudopotentials with a full core-hole, and the supercell approach was employed to eliminate the influence of interactions between core-holes.[8] Reflecting the bond distance between Si atoms, supercells with lattice constants larger than 10 Å were used for the calculation of the Si-*K* edge. A *k*-point density of 0.03 $Å^{-1}$ was used for the calculation of both the Si-*K* ELNES/XANES and the ground-state PDOS. The transition energy of Si-*K* ELNES/XANES can be calculated separately based on the total energy difference between the ground-state and excited-state simulations, but for the prediction of the ground-state PDOS in this study, all the spectra were shifted to their spectral onset to be zero. In other words, the prediction was conducted solely based on their spectral features.

All the calculated Si-*K* ELNES/XANES and ground-state PDOS were broadened by a Gaussian function with a standard deviation of 0.5 eV. The Si-*K* ELNES/XANES spectra were compiled into the database as 200-dimensional vectors with a resolution of 0.1 eV spanning from −1 to 19 eV relative to the Fermi level, while the ground-state PDOS spectra were stored in the database as 300-dimensional vectors with a resolution of 0.1 eV spanning from −15 to 15 eV.

We constructed an ML model using the database built through DFT simulations to predict atomic-level ground-state PDOS of a Si atom, including both valence and conduction bands, from the input Si-*K* ELNES/XANES. Given the difficulty in determining the absolute intensity of ELNES/XANES spectra in actual experiments, we normalized each input Si-*K* ELNES/XANES spectrum to have a maximum intensity of 1.0. The ML model is based on a one-dimensional convolutional neural network, designed to directly predict the ground-state PDOS about *s*- and *p*-orbitals separately from the output layer using the Si-*K* ELNES/XANES as the input. The detailed model architecture is described in Supporting Information C. The data was split into training, validation, and test sets with a ratio of 6:2:2.

RESULTS

In Figure 1, we show the result of the prediction for Si with a diamond-type structure, which is a representative crystal structure of Si. Here, the data for the diamond-type Si has been allocated to the test data set. Figure 1a shows the DFT-simulated ground-state PDOS of the Si atom in the diamond-type Si, where the gray dotted line corresponds to the PDOS related to the *s*-orbitals and the gray solid line corresponds to that about *p*-orbitals. The yellow solid line represents the sum of them. In the valence band, three main peaks can be observed. The peak highlighted with a yellow background on the lowest energy side is primarily composed of *s*-orbitals. The second peak consists of hybridized *s*- and *p*-orbitals, and the third peak is primarily composed of *p*-orbitals. In the conduction band, a strong peak composed of *p*- and *s*-orbitals can be observed at the onset.

Figure 1b shows the DFT-simulated Si-*K* edge for a Si atom in diamond-type Si. Since the Si-*K* edge corresponds to the

https://doi.org/10.1021/acs.jpcc.4c02818
J. Phys. Chem. C 2024, 128, 13500−13507

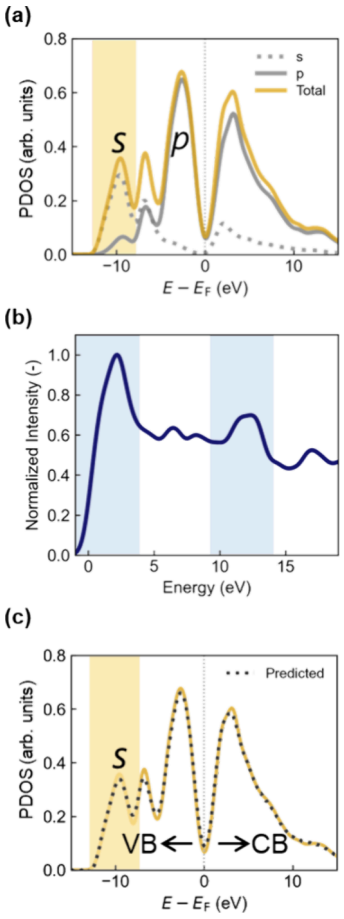


Figure 1. (a) Ground-state PDOS of the valence and conduction bands of Si in diamond-type Si calculated by DFT simulation. (b) DFT-simulated Si-*K* edge in diamond-type Si, where two main peaks can be observed. (c) Result of predicting the ground-state PDOS including both the valence band (VB) and conduction band (CB) of Si atom from the Si-*K* edge. Fermi level (E_F) was aligned to be 0 eV.

transitions of core electrons to the *p*-orbitals of the conduction band in the excited state due to the electric dipole selection rule, the spectral shape of the Si-*K* edge is believed to reflect the PDOS associated with *p*-orbitals of the conduction band in the excited state. Generally, in the case of the Si-*K* edge, two prominent peaks can be observed: a strong peak near the onset and another peak around 13 eV, as indicated by blue hatches in Figure 1b.

Figure 1c shows the predicted ground state PDOS from the Si-*K* edge, where the yellow solid line represents the ground-truth ground-state PDOS and the black dotted line represents the predicted PDOS by the ML model. Despite the Si-*K* edge reflecting the *p*-PDOS of the conduction band in the excited state, PDOS in the ground state was accurately predicted by the model. Furthermore, the valence band PDOS about *s*-orbitals in the ground state (yellow hatched area), which is considered to have little relevance to the conduction band *p*-orbitals, was also predicted successfully.

Figure 2 shows the prediction results for the entire test data set. In Figure 2a, all of the test data were sorted with the mean squared error (MSE) between the ground-truth PDOS and the ML-predicted PDOS. Figure 2b shows the prediction result for the data indicated by 'A' in Figure 2a, which corresponds to the data showing the best prediction accuracy among the entire test data, and the local coordination environment around the

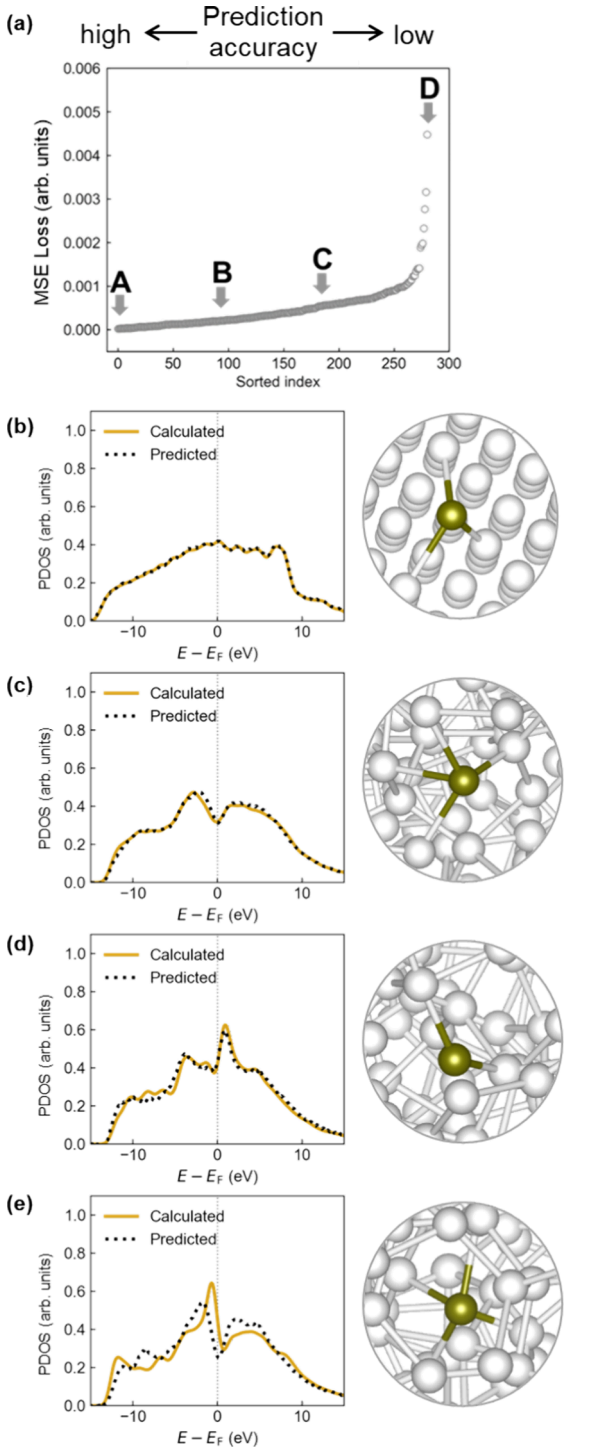


Figure 2. Result of prediction for the entire test data set. (a) Sorted mean squared error (MSE) for all the test data. (b) Result exhibiting the highest prediction accuracy among the test data and corresponding local coordination environment around the excited atom (gold). (c) Result corresponding to the top 33% prediction accuracy among the entire test data. (d) Result with top 67% prediction accuracy and (e) worst prediction accuracy.

corresponding excited atom. Here, the yellow solid line represents the ground-truth PDOS, while the black dotted line represents the predicted PDOS. Additionally, in the illustration depicting the local coordination environment, the excited Si atoms are colored gold, while the surrounding Si

https://doi.org/10.1021/acs.jpcc.4c02818
J. Phys. Chem. C 2024, 128, 13500−13507

atoms are shown in white. The results corresponding to the top 33%, 67%, and the worst predictions are shown in Figure 2c−e, respectively.

The data set includes various crystal structures of Si besides diamond-type Si, resulting in a variety of spectral shapes for the Si-*K* edge and the corresponding ground-state PDOS. From the results in Figure 2b−e, it can be observed that the ground-state PDOS of both valence and conduction bands, including the spectral shapes near the Fermi level, were generally predicted. In the results of Figure 2e, the MSE is large, which is caused by the similarity in the shape of the Si-*K* edge, while the shape of the ground state PDOS differs. The details of the result with low accuracy are described in Supporting Information A.

DISCUSSION

We have confirmed the ability to predict ground state PDOS from the Si-*K* edge across the entire test data set. To explore which regions of the Si-*K* edge, namely, how large energy range from the spectral threshold, are crucial for achieving high predictability, we assessed the accuracy of predicting ground-state PDOS while varying the range of the Si-*K* edge used as input. The results are listed in Figure 3. As depicted in Figure 3a, the range of Si-*K* edge from −1 to 19 eV was initially divided into four sections with 5 eV windows: −1 to 4, 4 to 9, 9 to 14, and 14 to 19 eV. The 0 eV corresponds to the spectral onset for each spectrum. Subsequently, we examined the change in the prediction accuracy of the ground state PDOS when incrementally expanding the input Si-*K* region by 5 eV. Resulting changes in the prediction accuracy are shown in Figure 3b. Here, the observation that the MSE of *p*-PDOS is consistently larger than that of *s*-PDOS can be attributed to the ground-state PDOS exhibiting a larger component of *p*-orbitals than *s*-orbitals, as can be found in Figure 1a. From Figure 3b, it can be observed that progressively expanding the input Si-*K* edge region results in a decrease in the MSE for both *s*-PDOS and *p*-PDOS, indicating an enhancement in predictability. However, it can also be observed that additionally including the spectral region from 14 to 19 eV as input does not contribute to the improvement in prediction accuracy. This means that, when predicting the ground-state PDOS in the range of −15−15 eV with respect to the Fermi level, the region from the onset of the Si-*K* edge to 14 eV is more critical. It has been known that the excited electronic structure tends to be localized to the excited atom, resulting in a shift of PDOS toward the lower energy sides compared to the ground state due to the formation of a core-hole. Therefore, it is reasonable that the spectral feature in the lower energy part of ELNES/XANES is more important for PDOS prediction.

As mentioned above, we observed a trend of improved prediction accuracy with an increasing input range. To investigate how each region of the Si-*K* edge contributes to this improvement, we examined the prediction accuracy of the ground-state PDOS when each 5 eV region of the Si-*K* edge was used as an independent input. The results are shown in Figure 3c. From these results, it is evident that higher prediction accuracy in terms of MSE is achieved when using the 4−14 eV region rather than the rising region of the Si-*K* edge for both *s*-PDOS and *p*-PDOS predictions. Furthermore, the highest prediction accuracy for *p*-PDOS was observed when utilizing the 4−9 eV region, while the highest prediction accuracy for *s*-PDOS was achieved with the 9−14 eV region. As explained in more detail later, the Si-*K* edge in the region of

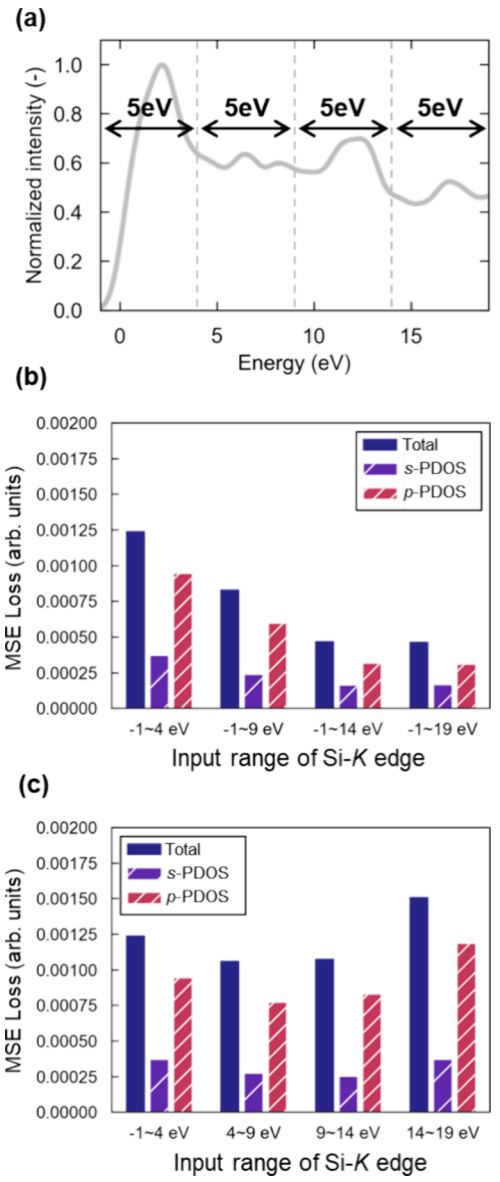


Figure 3. (a) Input Si-*K* edge spectrum, which covers a region of 20 eV, was divided into four sections, each with a range of 5 eV. (b) Change in the MSE as the input region is progressively increased by increments of 5 eV. (c) Variation in the MSE when each 5 eV region is used as an independent input.

4−9 eV contains the most useful information for predicting the *p*-PDOS near the Fermi level, while the Si-*K* edge in the region of 9−14 eV contains the most beneficial information for predicting valence-band *s*-PDOS. In the regions on the lower energy side of 14−19 eV, higher prediction accuracy was obtained regardless of which region is used, and there is no significant change in MSE when using the Si-*K* edge of each region. On the other hand, as can be seen from Figure 3b, the MSE significantly decreased by expanding the region. This suggests that each region of the Si-*K* edge contains information on different characteristics.

To investigate how each region of the Si-*K* edge contributes to the decrease in the MSE of the ground-state PDOS, the ground state PDOS was divided into four regions after predicting the PDOS across the entire region from the Si-*K* edge, and the prediction accuracy for each region was examined. As illustrated in Figure 4a, the 30 eV region from

https://doi.org/10.1021/acs.jpcc.4c02818
J. Phys. Chem. C 2024, 128, 13500−13507

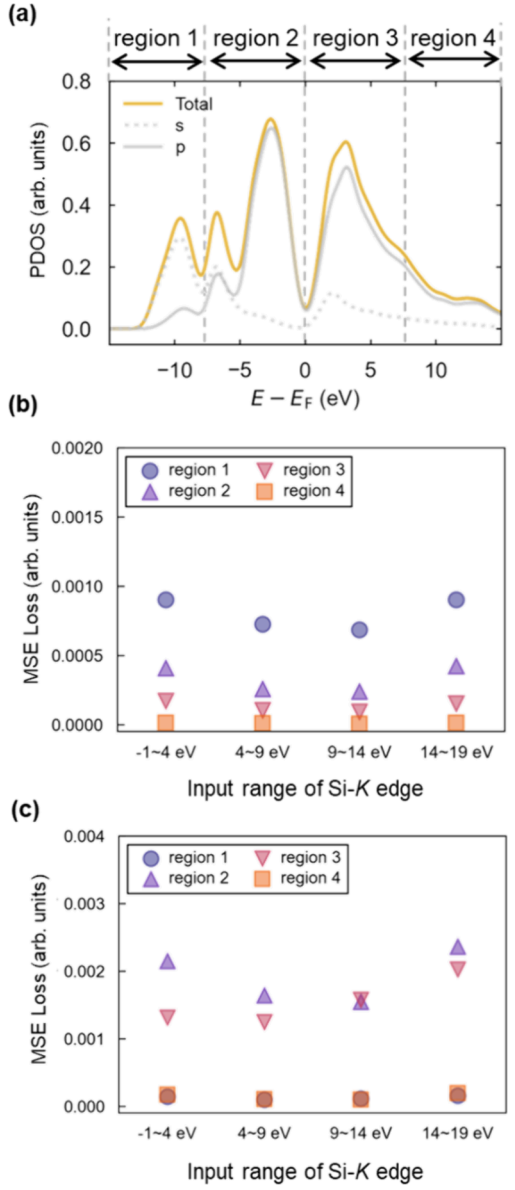


Figure 4. (a) After predicting ground-state PDOS across the entire region by varying the input regions, the ground-state PDOS was segmented into four distinct areas (regions 1 to region 4). Subsequently, the prediction accuracy within each of these areas was investigated. (b) Variation in prediction accuracy for the ground-state PDOS concerning *s*-orbitals within each specified region, in response to the variation in the input Si-*K* edge spectral region. (c) Result for the ground-state PDOS about *p*-orbitals.

−15 to 15 eV was divided into four regions, each spanning 7.5 eV. Here, regions 1 and 2 correspond to the valence band, while regions 3 and 4 correspond to the conduction band. Although the data set used in this work includes PDOSs with various spectral shapes, as an overall trend, characteristic peaks of *s*-PDOS can be observed in region below −7.5 eV (region 1) and that of *p*-PDOS can be observed in region from −7.5 to 7.5 eV (regions 2 and 3).

Figure 4b,c shows the relationship between the input range of Si-*K* edge and the prediction accuracy for each region of *s*-PDOS and *p*-PDOS, respectively. The *s*-PDOS mainly consists of valence band contributions, i.e., regions 1 and 2, and its direct information should be absent in the Si-*K* edge due to the selection rule. It can be observed that the prediction accuracy of such *s*-PDOS is slightly changed by the spectral region, and the highest accuracy for these regions is achieved when using the Si-*K* edge from 9 to 14 eV. As previously mentioned in Figure 3c, the highest prediction accuracy for the *s*-PDOS was achieved by utilizing the Si-*K* edge in the 9−14 eV. Those two results indicate that the 9 to 14 eV region encompasses valuable information for predicting *s*-orbitals throughout the entire valence band. At the present time, the underlying reason for this cannot be elucidated. However, as discussed in Figure 1b, there are two peaks at the Si-*K* edge, and it is implied that the second peak around 13 eV may contain valuable information beneficial for predicting the *s*-orbital in the ground state.

The prediction accuracy for the *p*-PDOS across the different regions differs from that for the *s*-PDOS. The *p*-PDOS primarily exhibits high electronic density of states at the upper edge of the valence band, namely, region 2, and at the lower edge of the conduction band, namely, region 3. Focusing on region 3, which is the lower edge of the conduction band, it is confirmed that high prediction accuracy can be achieved when using the Si-*K* edge with −1 to 9 eV. This seems natural when considering that Si-*K* edge reflects the *p*-orbitals of the conduction band in excited states, and there exists at least some correlation with the *p*-orbitals of the conduction band in the ground state. In contrast, for region 2, similar to the case with the *s*-PDOS, it is found that the highest prediction accuracy is also achieved when using the Si-*K* edge with 9 to 14 eV. This implies that the information related to the hybridization of *s*- and *p*-orbitals is primarily contained within the 9 to 14 eV range. Upon a comprehensive examination of all regions, it can be found that the Si-*K* edge in the energy range of 4−9 eV provides high prediction accuracy for p-PDOS in both region 2 and region 3.

The Si-*K* edge is thought to reflect the *p*-orbital PDOS of the conduction band in the excited state. Since the electronic structures of Si atoms are affected by the hybridization of *s*- and *p*-orbitals, the Si-*K* edge is considered to contain information related to the hybridization to some extent. The results shown in Figure 4 suggest the existence of energy bands containing information related to hybridization and that the ML model successfully extracts information related to the hybridization, allowing the prediction of not only the *p*-orbitals in the conduction band but also *s*-orbitals in the valence band.

In this study, we employed an ML model to predict the ground-state PDOS from the Si-*K* edge. To verify the effectiveness of using ML models, a comparison was conducted with an alternative method, which is template matching. In template matching, given a specific Si-*K* edge, the method involves searching the training data for the Si-*K* edge that is closest to the given one. The ground state PDOS corresponding to the spectrum of the closest match is then utilized as the prediction result. The comparison of the results from predictions made by the ML model and those made through template matching is shown in Figure 5. Here, in this study, MSE is used as the metric for measuring the similarity of the Si-*K* edge in template matching.

In Figure 5b, the values on the horizontal axis correspond to the MSE between the target Si-*K* edge and the Si-*K* edge, matched by template matching. Also, the value on the vertical axis represents the difference between MSE:

$$\Delta\mathrm{MSE} = \mathrm{MSE(Matched)} - \mathrm{MSE(ML\ predicted)}$$

https://doi.org/10.1021/acs.jpcc.4c02818
J. Phys. Chem. C 2024, 128, 13500−13507

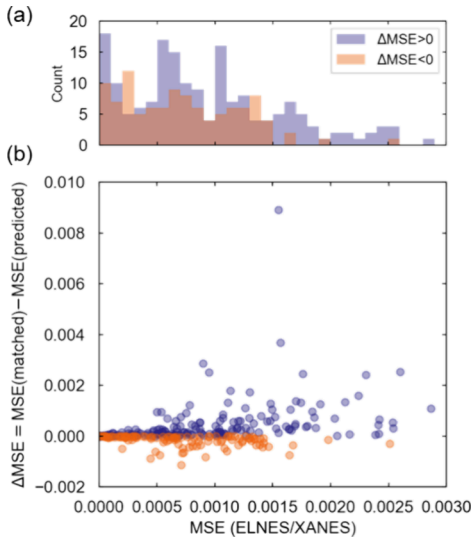


Figure 5. Comparison of predictions using ML vs. template matching. The values on the horizontal axis correspond to the MSE between the original ELNES/XANES and the ELNES/XANES matched by template matching, which is denoted as MSE (ELNES/XANES). (a) Cases where ML performed better compared to template matching are shown in the purple histogram, while the cases where template matching exhibits better accuracy are represented by the orange histogram. (b) Relationship between the differences in MSE of PDOS prediction by ML and MSE of PDOS prediction by template matching, and MSE (ELNES/XANES). The purple dots indicate that predictions made by ML exhibited higher prediction accuracy, while the orange dots signify those predictions made through template matching showed higher prediction accuracy.

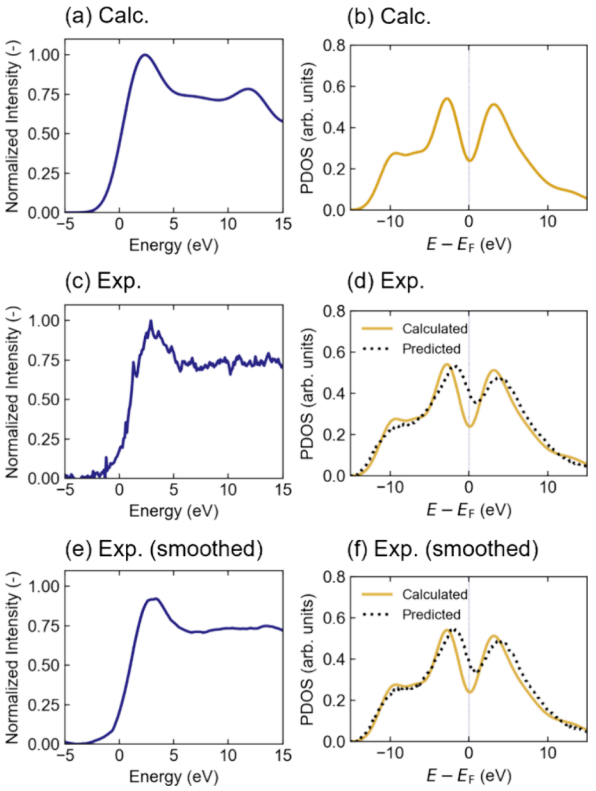


Figure 6. (a) DFT-calculated Si-*K* edge of diamond-type Si. (b) DFT-calculated ground-state PDOS of Si in diamond-type Si. (c) Experimentally obtained Si-*K* edge of diamond-type Si. (d) Ground-state PDOS predicted from the experimental Si-*K* edge of diamond-type Si. (e) Smoothed experimental Si-*K* edge of diamond-type Si. (f) Ground-state PDOS predicted from the smoothed experimental Si-*K* edge of diamond-type Si.

where MSE (Matched) represents the MSE between the ground-truth PDOS and the PDOS predicted by templated-matching, and the MSE (ML predicted) represents the MSE between the ground-truth PDOS and that predicted by the ML. When ΔMSE values are positive, indicating higher prediction accuracy achieved with ML, they are represented by purple dotes. Conversely, when ΔMSE values are negative, indicating higher prediction accuracy achieved with template-matching, they are represented by orange dots. In addition, histograms are displayed in the Figure 5a.

The histogram in Figure 5a reveals that in a large portion of cases ΔMSE is positive, revealing that higher prediction accuracy is achieved with the ML model. Furthermore, two points can be found from Figure 5b. First, the absolute value of ΔMSE tends to be smaller when it is negative, whereas the larger absolute value is often obtained when ΔMSE is positive, indicating that while template matching can be alternative to ML models in some cases, ML generally leads to higher prediction accuracy than the template matching. The second point is that when the MSE of the Si-*K* edge is large, indicating that it is difficult to find spectra with similar shapes within the database, the ΔMSE tends to be positive. This clearly indicates that the ML becomes more effective when there are no similar spectra in the database. Thus, compared to template matching, the utilization of ML models has generally been shown to be effective.

Finally, to investigate the applicability of our approach to experimental spectra, we input the experimental Si-*K* edge of diamond-type Si into the ML model. Figure 6a shows the DFT-calculated Si-*K* edge of diamond-type Si, and Figure 6b shows the corresponding DFT-calculated ground-state PDOS. Here, calculated spectra were broadened by a Gaussian function with a standard deviation of 1.2 eV to match the energy resolution of the experimentally obtained spectrum. The experimental spectrum was obtained using an aberration corrected STEM (JEOL ARM200CF) with a cold-type field emission electron source operated at 200 kV, with an EELS spectrometer (Gatan Quantum system) with an energy dispersion of 0.25 eV/channel. The spectrum was obtained from a single crystal of [110] oriented pure Si. In Figure 6c, we show the experimentally obtained Si-*K* edge of diamond-type Si. Compared to the DFT-calculated Si-*K* edge, it can be found that the experimental spectrum contains noise, and the second peak around 10−15 eV is obscured and cannot be observed.

Figure 6d shows the result of prediction from the Si-*K* edge. Here, the yellow solid line corresponds to the DFT-calculated ground-state PDOS of diamond-type Si and the black dotted line corresponds to the ground-state PDOS predicted by the ML model trained using the spectra with the broadening width of 1.2 eV. The overall shapes of the two PDOSs are similar, but there are differences in the DOS especially near the Fermi level. Figure 6e shows the experimental Si-*K* edge in Figure 6c smoothed using a Savitzky-Golay filter with a filter window size of 40 and a polynomial order of 3, and Figure 6f shows the ground-state PDOS predicted from the smoothed spectrum. Compared to the case using an unsmoothed spectrum, there is a slight change in the spectral shape, but the spectral feature near the Fermi level still differs from the DFT-calculated ground-state PDOS. These results suggest that whether peaks

https://doi.org/10.1021/acs.jpcc.4c02818
J. Phys. Chem. C 2024, 128, 13500−13507

are properly observed in the spectrum has a greater impact on the validity of prediction rather than whether noise smoothing has been performed. Therefore, to better apply to experimental spectra, it is considered necessary to address the factors that contribute to peak obscuration and to improve the energy resolution of Si-*K* edge observations.

In this study, we theoretically examined the predictability of the atomic-level electronic structures at the ground state from ELNES/XANES, focusing on the Si-*K* edge of Si. It was suggested that by training an ML model to learn the correlation between ELNES/XANES and the PDOS of the valence and conduction bands at the ground state, it is possible to predict the electronic structure at the ground state including the valence band from ELNES/XANES. The insights obtained in this study are expected to be in part transferable to the *K* edges of elements with electronic structure similar to Si, such as the C-*K* edge. As in the case of Si, C tends to form hybridized orbitals, and it is expected that the ground-state PDOS, including not only *p*-orbitals but also *s*-orbitals, can be predicted from the C-*K* edge. On the other hand, oxides tend to form more ionic bonds, and the PDOS of O in oxides, unlike in the case of Si and C, has a larger DOS only in valence band. Investigating the applicability of this approach to systems where the correlation between the valence and conduction bands is less apparent is an interesting direction for future research.

CONCLUSIONS

In conclusion, we have developed an ML model capable of predicting atom-specific ground-state PDOS of a Si atom from the corresponding Si-*K* ELNES/XANES. This demonstration has shown the feasibility of predicting PDOS for both the valence and conduction bands, including both *s*- and *p*- orbitals in the ground state of silicon atoms, from the Si-*K* edge, which is traditionally thought to reflect the conduction band PDOS of *p*-orbitals in the excited state. We discovered that the information within the energy region of the Si-*K* edge varies, with the area near the rising edge being particularly effective for predicting *p*-orbitals of the conduction band, which is believed to be directly correlated with the Si-*K* edge. Additionally, the energy region just beyond the rising edge was proved to be effective for predicting the *s*- and *p*-orbitals of the valence band in the ground state. This region is considered to contain information related to the hybridization of *s*- and *p*-orbitals, suggesting that ML models can capture such complex correlations.

Finally, the effectiveness of ML models was confirmed through comparison with template matching. Consequently, the ML models are confirmed to be particularly advantageous over template matching when the database lacks similar shapes of the Si-*K* edge.

Through this study, we have demonstrated the prediction of atom-specific ground-state PDOS from ELNES/XANES with a focus on silicon. Expanding this technique to encompass more complex materials and experimental spectra is anticipated to not only facilitate the realization of atomic-level analysis of ground-state electronic structures but also to deepen our understanding on the stability and properties of materials.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcc.4c02818.

Analysis on the results with low prediction accuracy, detailed procedure for data set augmentation, and detailed model architecture (PDF)

AUTHOR INFORMATION

Corresponding Authors

Izumi Takahara − *Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan;* orcid.org/0009-0004-6022-1945; Email: kougen@iis.u-tokyo.ac.jp

Teruyasu Mizoguchi − *Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan;* orcid.org/0000-0003-3712-7307; Email: teru@iis.u-tokyo.ac.jp

Authors

Fumihiko Uesugi − *Research Network and Facility Services Division, National Institute for Materials Science, Tsukuba 305-0044, Japan*

Koji Kimoto − *Center for Basic Research on Materials, National Institute for Materials Science, Tsukuba 305-0044, Japan;* orcid.org/0000-0002-3927-0492

Kiyou Shibata − *Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan;* orcid.org/0000-0002-0639-5408

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcc.4c02818

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by MEXT/JSPS, KAKENHI (Grant Numbers JP19H05787, JP24K08016, JP24KJ0939) and JST, CREST (Grant Number JP-MJCR1993). This study was partially supported by NEDO (New Energy and Industrial Technology Development Organization), Japan. We acknowledge Mr. Poyen Chen at the University of Tokyo for fruitful discussion on the present study. I.T. was supported by the Program for Leading Graduate Schools (MERIT-WINGS).

REFERENCES

(1) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136* (3B), B864−B871.

(2) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140* (4A), A1133−A1138.

(3) Ben Mahmoud, C.; Anelli, A.; Csányi, G.; Ceriotti, M. Learning the Electronic Density of States in Condensed Matter. *Phys. Rev. B* **2020**, *102* (2), No. 235130.

(4) Chandrasekaran, A.; Kamal, D.; Batra, R.; Kim, C.; Chen, L.; Ramprasad, R. Solving the Electronic Structure Problem with Machine Learning. *npj Comput. Mater.* **2019**, *5* (1), No. 22.

(5) Fung, V.; Ganesh, P.; Sumpter, B. G. Physically Informed Machine Learning Prediction of Electronic Density of States. *Chem. Mater.* **2022**, *34* (11), 4848−4855.

(6) Ikeno, H.; Mizoguchi, T. Basics and Applications of ELNES Calculations. *Microscopy* **2017**, *66* (5), 305−327.

(7) Tanaka, I.; Mizoguchi, T.; Yamamoto, T. XANES and ELNES in Ceramic Science. *J. Am. Ceram. Soc.* **2005**, *88* (8), 2013−2029.

(8) Mizoguchi, T.; Tanaka, I.; Gao, S.-P.; Pickard, C. J. First-Principles Calculation of Spectral Features, Chemical Shift and Absolute Threshold of ELNES and XANES Using a Plane Wave Pseudopotential Method. *J. Phys.: Condens. Matter* **2009**, *21* (10), No. 104204.

(9) Mizoguchi, T.; Tanaka, I.; Yoshioka, S.; Kunisu, M.; Yamamoto, T.; Ching, W. Y. First-Principles Calculations of ELNES and XANES of Selected Wide-Gap Materials: Dependence on Crystal Structure and Orientation. *Phys. Rev. B* **2004**, *70* (4), No. 045103.
(10) Kimoto, K.; Asaka, T.; Nagai, T.; Saito, M.; Matsui, Y.; Ishizuka, K. Element-Selective Imaging of Atomic Columns in a Crystal Using STEM and EELS. *Nature* **2007**, *450* (7170), 702−704.
(11) Bosman, M.; Keast, V. J.; García-Muñoz, J. L.; D'Alfonso, A. J.; Findlay, S. D.; Allen, L. J. Two-Dimensional Mapping of Chemical Information at Atomic Resolution. *Phys. Rev. Lett.* **2007**, *99* (8), No. 086102.
(12) Tan, H.; Turner, S.; Yücelen, E.; Verbeeck, J.; Van Tendeloo, G. 2D Atomic Mapping of Oxidation States in Transition Metal Oxides by Scanning Transmission Electron Microscopy and Electron Energy-Loss Spectroscopy. *Phys. Rev. Lett.* **2011**, *107* (10), No. 107602.
(13) Bugnet, M.; Ederer, M.; Lazarov, V. K.; Li, L.; Ramasse, Q. M.; Löffler, S.; Kepaptsoglou, D. M. Imaging the Spatial Distribution of Electronic States in Graphene Using Electron Energy-Loss Spectroscopy: Prospect of Orbital Mapping. *Phys. Rev. Lett.* **2022**, *128* (11), No. 116401.
(14) Cuisinier, M.; Cabelguen, P.-E.; Evers, S.; He, G.; Kolbeck, M.; Garsuch, A.; Bolin, T.; Balasubramanian, M.; Nazar, L. F. Sulfur Speciation in Li−S Batteries Determined by Operando X-Ray Absorption Spectroscopy. *J. Phys. Chem. Lett.* **2013**, *4* (19), 3227−3232.
(15) Iglesias-Juez, A.; Chiarello, G. L.; Patience, G. S.; Guerrero-Pérez, M. O. cy—XAS,XANES. *Can. J. Chem. Eng.* **2022**, *100* (1), 3−22.
(16) Houhou, R.; Bocklitz, T. Trends in Artificial Intelligence, Machine Learning, and Chemometrics Applied to Chemical Data. *Anal. Sci. Adv.* **2021**, *2* (3−4), 128−141.
(17) Timoshenko, J.; Lu, D.; Lin, Y.; Frenkel, A. I. Supervised Machine-Learning-Based Determination of Three-Dimensional Structure of Metallic Nanoparticles. *J. Phys. Chem. Lett.* **2017**, *8* (20), 5091−5098.
(18) Kiyohara, S.; Tsubaki, M.; Liao, K.; Mizoguchi, T. Quantitative Estimation of Properties from Core-Loss Spectrum via Neural Network. *J. Phys. Mater.* **2019**, *2* (2), No. 024003.
(19) Zheng, C.; Chen, C.; Chen, Y.; Ong, S. P. Random Forest Models for Accurate Identification of Coordination Environments from X-Ray Absorption Near-Edge Structure. *Patterns (N Y)* **2020**, *1* (2), No. 100013.
(20) Kiyohara, S.; Mizoguchi, T. Radial Distribution Function from X-Ray Absorption near Edge Structure with an Artificial Neural Network. *J. Phys. Soc. Jpn.* **2020**, *89* (10), No. 103001.
(21) Kikumasa, K.; Kiyohara, S.; Shibata, K.; Mizoguchi, T. Quantification of the Properties of Organic Molecules Using Core-loss Spectra as Neural Network Descriptors. *Adv. Intell. Syst.* **2022**, *4* (1), No. 2100103.
(22) Mizoguchi, T.; Kiyohara, S. Machine Learning Approaches for ELNES/XANES. *Microscopy* **2020**, *69* (2), 92−109.
(23) Guda, A. A.; Guda, S. A.; Martini, A.; Kravtsova, A. N.; Algasov, A.; Bugaev, A.; Kubrin, S. P.; Guda, L. V.; Šot, P.; van Bokhoven, J. A.; Copéret, C.; Soldatov, A. V. Understanding X-Ray Absorption Spectra by Means of Descriptors and Machine Learning Algorithms. *npj Comput. Mater.* **2021**, *7* (1), No. 203.
(24) Carbone, M. R.; Yoo, S.; Topsakal, M.; Lu, D. Classification of Local Chemical Environments from X-Ray Absorption Spectra Using Supervised Machine Learning. *Phys. Rev. Mater.* **2019**, *3* (3), No. 033604.
(25) Torrisi, S. B.; Carbone, M. R.; Rohr, B. A.; Montoya, J. H.; Ha, Y.; Yano, J.; Suram, S. K.; Hung, L. Random Forest Machine Learning Models for Interpretable X-Ray Absorption near-Edge Structure Spectrum-Property Relationships. *npj Comput. Mater.* **2020**, *6* (1), No. 109.
(26) Chen, P.-Y.; Shibata, K.; Hagita, K.; Miyata, T.; Mizoguchi, T. Prediction of the Ground-State Electronic Structure from Core-Loss Spectra of Organic Molecules by Machine Learning. *J. Phys. Chem. Lett.* **2023**, *14* (20), 4858−4865.
(27) 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), No. 011002.
(28) Togo, A.; Tanaka, I. *Spglib: A Software Library for Crystal Symmetry Search. arXiv [cond-mat.mtrl-sci]*, 2018. http://arxiv.org/abs/1808.01590.
(29) Clark, S. J.; Segall, M. D.; Pickard, C. J.; Hasnip, P. J.; Probert, M. I. J.; Refson, K.; Payne, M. C. First Principles Methods Using CASTEP. *Zeitschrift für Kristallographie - Crystalline Materials* **2005**, *220* (5−6), 567−570.
(30) Morris, A. J.; Nicholls, R. J.; Pickard, C. J.; Yates, J. R. OptaDOS: A Tool for Obtaining Density of States, Core-Level and Optical Spectra from Electronic Structure Codes. *Comput. Phys. Commun.* **2014**, *185* (5), 1477−1485.
(31) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77* (18), 3865−3868.

https://doi.org/10.1021/acs.jpcc.4c02818
J. Phys. Chem. C 2024, 128, 13500−13507