

CuXASNet: Rapid and accurate prediction of copper L -edge x-ray absorption spectra using machine learning

Samuel P. Gleason^{1,2}, Matthew R. Carbone³, Deyu Lu⁴, and Jim Ciston¹

¹National Center for Electron Microscopy, *Molecular Foundry*, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

²Department of Chemistry, *University of California, Berkeley*, California, USA

³Computing and Data Sciences Directorate, *Brookhaven National Laboratory*, Upton, New York 11973, USA

⁴Center for Functional Nanomaterials, *Brookhaven National Laboratory*, Upton, New York 11973, USA



(Received 24 November 2024; accepted 28 April 2025; published 10 July 2025; corrected 20 August 2025)

In this work, we have developed CuXASNet, a dense neural network that predicts simulated Cu L -edge x-ray absorption spectra (XAS) from atomic structures. Featurization of the Cu local environment is performed using a component of M3GNet, a graph neural network developed for predicting the potential energy surface. CuXASNet is trained on simulated spectra from FEFF9 at the multiple scattering level of theory, and can predict the L_3 and L_2 edges for Cu sites to quantitative accuracy. To validate our approach, we compare 14 experimental spectra extracted from the literature with the predictions of CuXASNet. The agreement of CuXASNet with experiments is shown by an average mean absolute error of 0.125 and an average Spearman's correlation coefficient of 0.891, which is comparable to FEFF9's values of 0.131 and 0.898 for the same metrics. As such, CuXASNet can rapidly predict a large number of L -edge XAS spectra at the same accuracy as FEFF9 simulations. This can be used as a drop-in replacement for multiple scattering codes for fast screening of candidate atomic structure models of a measured system. This model establishes a general framework for Cu XAS prediction, and can be extended to more computationally expensive levels of theory and to other transition metal L edges.

DOI: [10.1103/xz7f-srky](https://doi.org/10.1103/xz7f-srky)

I. INTRODUCTION

X-ray absorption spectroscopy (XAS) is commonly used to identify atomic and electronic information about functional materials, including local chemical environments and oxidation states (OSs) [1]. In XAS, a core-level photoelectron is excited to the conduction band. The spectrum serves as a fingerprint for the absorbing element's atomic number and is sensitive to core-level orbital energy changes caused by differences in the bonding environment [2]. The information contained in an XAS spectrum is crucial to the mechanistic understanding and rational design of functional materials, such as catalysts [3,4], renewable energy materials [5,6], and batteries [7]. XAS analysis is typically conducted by dividing the spectrum into two components. The first component is the x-ray absorption near-edge structure (XANES), and provides information on OS and the coordination environment of the absorbing site [1]. The second component is the extended x-ray absorption fine structure (EXAFS), and provides information on atomic structure beyond the absorbing site, such as bonding structure and nearest neighbor bond lengths [8]. This work focuses primarily on the generation and analysis of spectra for XANES studies.

Many technologically important functional materials contain $3d$ transition metals, such as Ti, Mn, Fe, Co, Ni, and Cu. This is primarily due to their chemical flexibility, namely that they can adopt several different OSs. Additionally, these elements are relatively highly abundant, allowing for simpler commercial development relative to less abundant elements, such as Pd, Rh, and Ru [11]. $3d$ transition metals are common in catalytic materials, which often contain $3d$ transition metal centers. For instance, Ni, Mn, and Fe are used in artificial photosynthesis [12], Cu is used in CO_2 reduction [13] and the azide-alkene cycloaddition click reaction [14–16], and Cu, Fe, and Ti are used in photovoltaic devices [17–20]. $3d$ transition metals are also common in biotechnology and medical applications. For example, Co has attained wide usage in medical imaging [21], Cu is present in many antimicrobial devices [22], and molecules with Fe centers are studied for a wide range of biomedical applications, including drug delivery and cancer treatment [23,24].

XAS analysis of $3d$ transition metals is typically done using the K , L , and M edges, which correspond to the excitation of core electrons from the $1s$, $2p$, and $3d$ orbitals, respectively [1]. For most materials science applications, the analysis of $3d$ transition metals using XAS is often focused on the L edge [4]. Analysis of L -edge XAS is typically done by matching simulated spectra or experimental standards of known chemical composition and atomic arrangement to an unknown sample. In the case of experimental spectra, a few online databases of experimental spectra exist [25,26]. However, these are often focused on a few specific systems and are

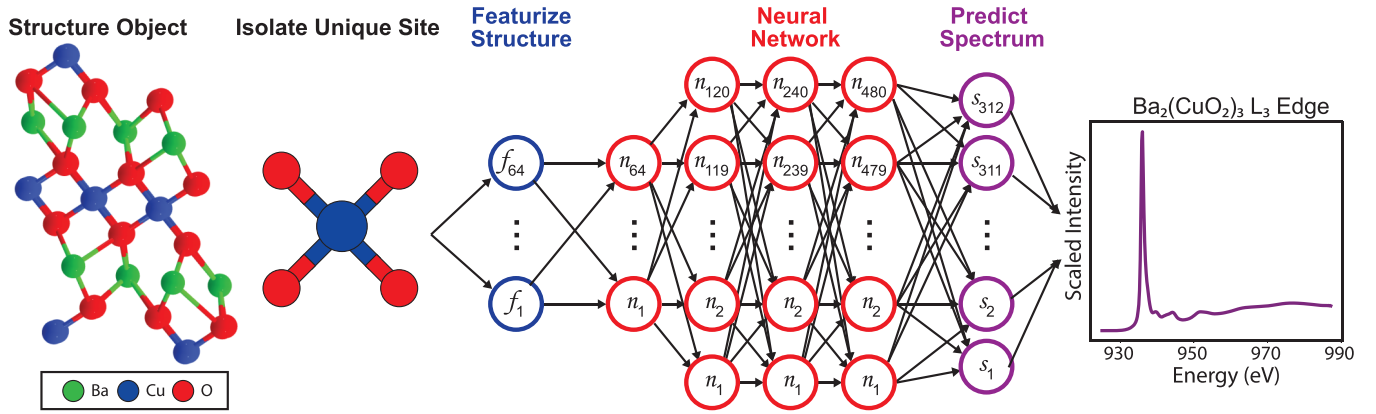


FIG. 1. Outline of the CuXASNet model architecture. A structure object of a Cu containing material from the Materials Project (mpid: mp-615789) is represented using Pymatgen [9]. Then, every symmetrically unique Cu site is extracted using the space group symmetry. Upon isolation of the unique Cu sites, they are featurized using M3GNet which transforms the site into a 64-dimension vector representing its local environment [10]. This vector is then input to our spectral prediction neural network, which contains three fully connected hidden layers, before returning a predicted spectrum that is discretized on an energy grid of 312 dimensions.

too narrow in scope to be used for general experimental spectral analysis. Additionally, they often contain mainly K -edge data for $3d$ transition metals. The limitation of matching unknown experimental samples to simulated spectra is typically the accuracy of the simulations, which often varies depending on the atomic structure and the simulation methods [27–30]. Additionally, the availability of simulated data, particularly for more computationally intensive simulations, is often a significant limitation. To acquire accurate simulations that are useful for experimental analysis, detailed knowledge of the material structure is required [31,32]. Due to the computational cost of simulations, performing extensive searches of candidate atomic structures that match target measured XAS spectra, either through brute force methods or structure sampling algorithms, is computationally prohibitive. Therefore, there is a need for a spectral prediction model that can rapidly predict accurate simulated spectra from an arbitrary atomic structure.

A large collection of simulated XAS spectra is found in the Materials Project [33] Database, which includes K -edge [34] and more recently L -edge [27] XAS spectra using the FEFF9 code [32–35]. However, despite the over $\sim 100k$ new L -edge spectra produced by Chen *et al.* [27], the vast majority of structures in the Materials Project do not have associated L -edge spectra due to the computational cost of XAS simulations. For example, out of the nearly 10k Cu-containing materials in the Materials Project, roughly 8.5k do not have an L -edge spectrum [33]. Therefore, there is a strong demand for models which can predict simulated data quickly enough to process well over a million $3d$ transition metal-containing structures while retaining enough accuracy for application to experimental analysis. Additionally, certain spectral features in the L -edge spectra of Ti and Fe can be better reproduced by levels of theory other than the multiple scattering method, and further investigation of these systems can necessitate more computationally intensive simulations [27–29]. For example, the Ti L edge of TiO_2 is more accurately simulated by the OCEAN simulation code compared to FEFF9 [27,29]. However, the computational cost of OCEAN simulations increases drastically with increasing system size [28,36]. Therefore,

generating large volumes of OCEAN spectra is even more challenging than building a FEFF9 database.

Simulated spectral data sets can also be used to train machine learning (ML) models for automated analysis of core-level spectroscopy data. Due to the increasing availability of simulated data, ML has seen explosive growth in recent years as a broadly applicable technique to analyze XAS data [33–35]. Several groups have used ML to predict coordination environments and three-dimensional molecular geometries from XAS L -edge and K -edge spectra [30,37–42], extract OS information from XAS [43–45], and determine structural information, such as bond length and angle, from core-level spectroscopy [46–48]. However, the success of these ML approaches requires a large training data set comprised of examples across relevant feature spaces. There are examples of automated core-level spectral analysis models trained exclusively on experimental spectra, such as in Ref. [49], which predicts the OS of Mn. However, data sets for automated analysis of core-level spectra are more often built on simulated spectra due to the significant challenge of collecting a large volume of labeled experimental data. Additionally, the large scale development of core-level spectroscopy models has been limited in part by the smaller amount of available simulated data for the transition metal L -edge relative to K -edge spectra [27,33].

Recently, machine learning models have been used to predict simulated XAS spectra at a fraction of the computational cost of direct simulation [50–53]. However, these studies have mostly focused on K -edge simulations of small molecules/atomic clusters. Therefore, in this work we have developed CuXASNet, an ML model which predicts L -edge XAS spectra of Cu materials using atomic structure as input (see Fig. 1). We envision that CuXASNet can be used to significantly enhance the volume of simulated L -edge data of Cu materials available to the community. CuXASNet is trained on spectra simulated using FEFF9, which utilizes the multiple scattering simulation method [32]. Additionally, the methodology we present here is extendable to more computationally intensive simulations, such as OCEAN or VASP [28,29,36,54], and to methods based on multiplet ligand-field theory [1,55].

II. METHODS

A. Training set generation

Pymatgen was used to extract simulated FEFF9 L -edge XAS spectra of Cu materials from the Materials Project Database [9,33]. Currently, the Materials Project contains L -edge Cu XAS spectra of 1533 materials, which encompass 2300 site spectra from unique Cu absorbers [27,33]. In order to ensure enough training data for training CuXASNet, roughly 2000 additional structures were extracted from the Materials Project by selecting all Cu-containing materials predicted to be stable by density functional theory (DFT) or that had been labeled as experimentally synthesized [33]. By filtering for stable and/or experimentally synthesized materials, we aimed to produce a data set that would be most relevant to experimental studies of Cu materials. Using the *Lightshow* workflow [31,56] and the FEFF9 code [32], we previously simulated an additional 3387 site specific L -edge spectra of Cu absorbers [43]. The unique Cu site(s) in each structure were determined using the space group symmetry via Pymatgen. The data set was then further processed before model training. These processing steps include interpolation to place every spectrum onto a uniform grid with 0.2 eV energy scale and scaling all spectra to have the same energy range, 925–987.2 and 945–1007.2 eV for L_3 and L_2 , respectively. These steps followed the same data-set construction and spectral processing procedure conducted in Ref. [43] and resulted in the same data set, up to a few minor changes described in later sections.

B. Simulated spectral processing

The spectral data set was validated to retain only simulated spectra from converged FEFF9 calculations. Filtering unconverged FEFF9 calculations resulted in the removal of 89 spectra from the full data set. Additionally, all materials used to compare the output of the CuXASNet model to experimental spectra were removed from the training set to ensure the model was not biased by having seen these spectra in the training process. Consequently, these spectra are not in the test set examined in the Results and Discussion section. The resulting data set used for model training and validation contains 5497 site-specific spectra. The simulated spectra are normalized so the highest energy point is equal to 1. This normalization is applied to all the spectra visualized and tested in this work.

CuXASNet was trained using input vectors representing unique absorbing sites in an arbitrary structure object produced by the featurization module of M3GNet [10]. M3GNet was selected as the structure featurizer in this work due to recent results showing it was the most effective model at dimensionality reduction for predicting XAS spectra [10]. This featurization step was performed by extracting node-level features from the graph representation of a material's unit cell before M3GNet's readout layer (see Ref. [10] for details). For each unique site, a feature vector of length 64, the default hidden state size for M3GNet, was produced to represent that specific site. This feature became the input vector for our ML model.

TABLE I. The hyperparameters tested during CuXASNet construction and their associated test set accuracies for L_2 and L_3 edges. The row in bold is the parameters used for CuXASNet in this work.

Layers	Epochs	Batch Size	L_3 Median MAE	L_2 Median MAE
1	500	10	0.047	0.036
3	500	10	0.039	0.032
3	500	100	0.040	0.033
3	500	1000	0.039	0.032
3	50	10	0.044	0.037

C. Dense neural network

In CuXASNet, two deep neural network models were used to predict the L -edge XAS spectra of Cu absorbers from their featurized atomic structure, one for the L_3 edge and one for the L_2 edge with the same hyperparameters (see Table I). The simulated spectral data set was randomly split into train and test segments, with 75% of the data set used for model training while 25% was reserved for testing. The test data set includes only simulated spectra. Each neural network in CuXASNet has an input layer of length 64, equal to the length of the featurized structure. This is followed by three hidden layers of length 120, 240, and 480. The rectified linear unit (ReLU) activation function was used throughout, except for the output layers, where a softplus activation function was used. The output layer has a length of 312, with each point representing an intensity point in the spectrum separated by 0.2 eV. Models were trained using the Adam optimizer and mean squared error loss function for 500 epochs with a batch size of 10 in the training [57]. The number of epochs and batch size was not found to have a significant impact on the model results. Batch sizes from 10 to 1000 did not change the median mean absolute error (MAE) of the test set, 0.0391, by more than 2%, and a batch size of 10 was found to produce the best model (see Table I). The number of epochs had a similarly small impact, with 500 producing the best model (see Table I). Figure S4 of the Supplemental Material [67] shows the relation of the training loss (between the predicted and FEFF9 spectra) versus epoch, where the model has converged by the end of the training process for both L_3 and L_2 edges.

D. Experimental XAS data

Experimental XAS from the literature were used to determine the applicability of the predictions from CuXASNet to experimental analysis. Fourteen experimental spectra were extracted from the literature by using webplotdigitizer [58–65]. The experimental spectra extracted for verification are the following: Cu metal, Cu₂O, CuO, Cu₂S, CuS, CoCu₂S₄, CuBe, CuFeS₂, LaCuOS, LaCuOSe, LaCuOTe, SrCuO₂, ZrCuSiAs, and ZrCuSiP. The experimental spectra were processed using a broadening function that ensured the extraction produced a smoothed spectrum. Our simulated $L_{2,3}$ spectra are set on an energy range 925–1007.2 eV, however, in multiple cases the experimental spectra were measured on a smaller energy range (i.e., 930–975 eV). When this occurred, the simulated spectrum was compared to the experimental spectrum only over the range where experimental data were recorded. Then the simulated spectra were cropped to the same energy range and the mean absolute error (MAE) was

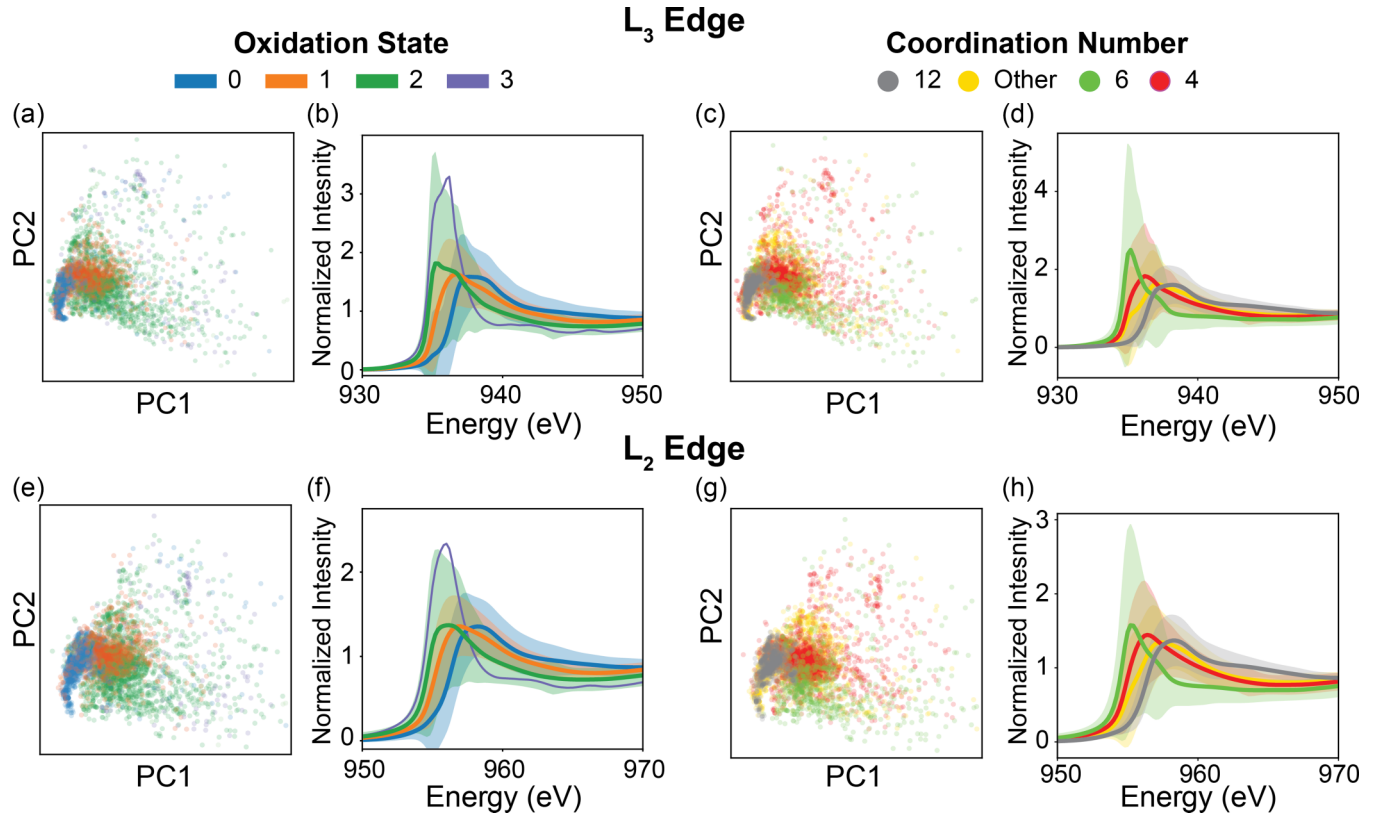


FIG. 2. Illustration of the data spread in our spectral data set. Each row shows a principle component analysis (PCA) and spectral comparison of the L_3 (a)–(d) and L_2 (e)–(h) edges, colored by their OS and CN. The first and third columns show scatter plots of the OS [(a) and (e)] and coordination number [(c) and (g)] of the top two principle components in the PCA conducted on the L_3 and the L_2 edges. The second and fourth columns show the average spectrum (solid line) across each OS [(b) and (f)] and coordination number [(d) and (h)] for L_3 and L_2 edges and spectra within one standard deviation (shaded area around the solid line).

calculated between the simulated/CuXASNet spectrum and the experimental spectrum.

III. RESULTS AND DISCUSSION

A. Spectral database analysis

The spectral database used in this work is comprised of 5497 site-specific Cu $L_{2,3}$ -edge XAS spectra across roughly 3400 Cu containing materials. The spectral features are illustrated in Fig. 2 by first distilling spectral features to a series of component vectors via principle component analysis (PCA) overlaid with averaged spectra across two Cu chemical descriptors, OS and coordination number (CN). PCA linearly decomposes a data set into a set of principal vectors (components) and weights such that any spectrum can be reconstructed via a linear combination of the principal vectors. Plotting the weights of the most significant characteristic vectors for each spectrum and labeling them by known chemical descriptors can elucidate the degree of similarity across chemical systems, such as spectra that share the same OS. A detailed explanation of how PCA can be applied to spectral data can be found in Refs. [42,66].

In Fig. 2(a), the values of the first two principle components are plotted for each L_3 edge and colored by the OS of their corresponding absorbing site. The general trend shows that metallic materials, Cu(0), are highly clustered, indicating

that spectral features of metallic Cu absorbers are relatively similar. As the oxidation state increases, this variability increases, although there are very few Cu(III) materials in this data set due to the rarity of Cu^{3+} ions in nature. Figure 2(b) shows the average of all spectra with a particular OS, with the spread of Cu(III) materials omitted due to their low frequency in the data set. Figure 2(b) demonstrates the increased variance in Cu(II) as compared to Cu(I)/Cu(0), as shown by the larger standard deviation of the averaged Cu(II) spectrum. Figures 2(c) and 2(d) show a similar analysis with color coding and labels corresponding to CN. Figure 2(c) shows the same PCA plot as Figure 2(a) and reveals the following trends: Cu sites with a CN of 12 are strongly associated with metallic Cu in a face-centered-cubic (FCC) lattice, and this is confirmed by the averaged spectrum in Fig. 2(d). Sites with a CN of 6 (e.g., an octahedron motif) are also reasonably well contained, while absorbing sites with CN of 4 (e.g., a tetrahedron motif) produce a wide range of features. Interestingly, there is more variance in the spectra with CN = 4 than the “other” category, which encompasses many CNs that are all present in insufficient numbers to be visualized directly in Fig. 2 (see Fig. S1 [67]). Figures 2(e)–2(h) show the same analysis as Figs. 2(a)–2(d), instead on the L_2 edge. The trends are essentially the same, with the only systematic difference being that the L_2 edge exhibits more uniformity in general, exemplified by the increased clustering in the

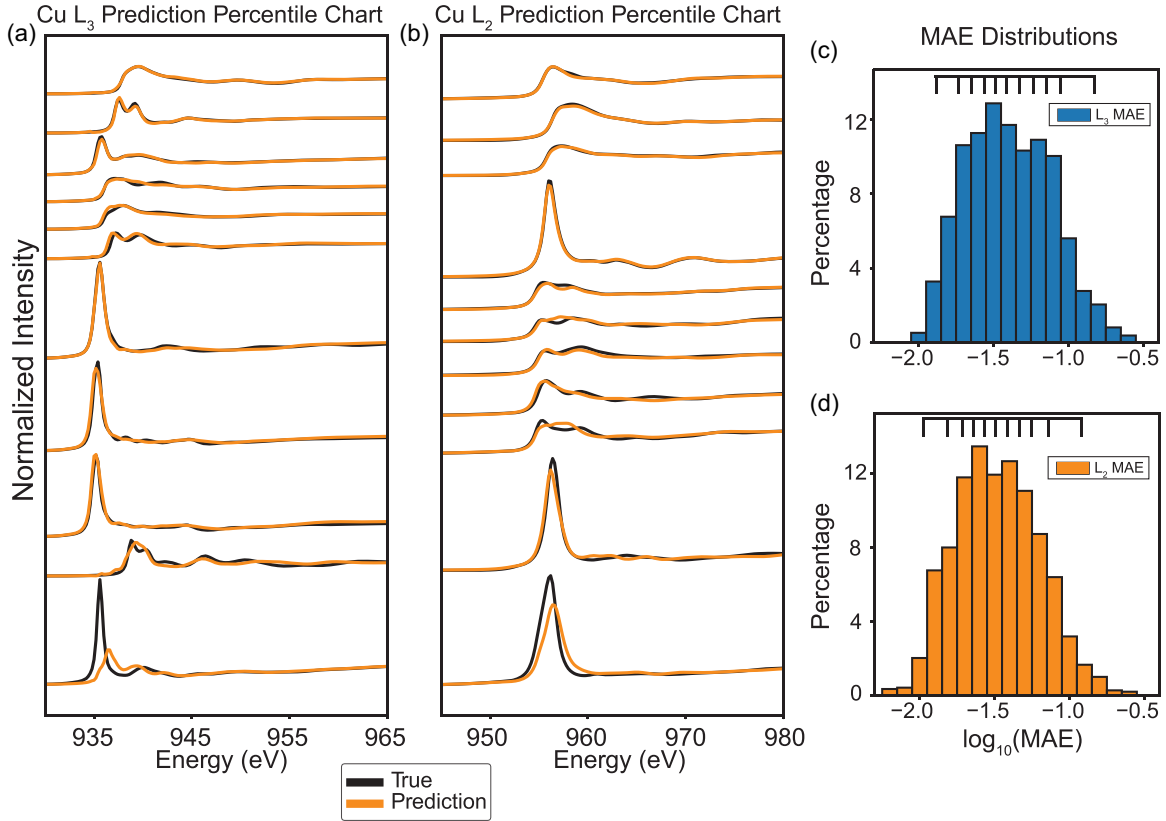


FIG. 3. Waterfall plots showcasing the accuracy of our model at reproducing spectra simulated with FEFF9. Panels (a) and (b) show decile charts with decreasing accuracy moving down the chart. The spectra visualized in these plots reflect the middle 95% of the test set, removing the best and worst 2.5%. Each comparison shows a decile value of CuXASNet’s performance on the middle 95% of our test spectra, with the first comparison showing the most accurate included prediction, the second showing the tenth percentile of this subset of the test data, and subsequent spectra being spaced by ten percentile increments. The L_3 and L_2 edges are shown in (a) and (b), respectively. The MAE distributions are shown in (c) and (d), which have been projected onto a log scale. The horizontal tick marks correspond to spectra in (a) and (b), with moving rightward across (c) corresponding to moving down in (a), and the same being true for (d) and (b).

first two principle components and the smoother averaged spectra.

B. L_3/L_2 spectra prediction

CuXASNet is comprised of two separate models trained on the simulated data: one to predict the L_3 edge and one to predict the L_2 edge. The accuracy of each of these tasks is shown in Fig. 3. Figures 3(a) and 3(b) shows a decile plot of the middle 95% of the test set performers, discarding the best and worst 2.5%. This range was chosen due to observations that the best and worst 2.5% were significant outliers and did not represent the overall performance of the model. This can be seen by the very low density of the highest and lowest MAEs in Figs. 3(c) and 3(d). Particularly, the least accurate 2.5%, and more specifically the two worst spectra, comprising the worst 0.1% of the test set, were observed to match the FEFF9 simulation very poorly. This is likely caused by the very uncommon features observed in these two FEFF9 spectra (e.g., the energy location of the edge onset is roughly 10 eV higher than most other simulated spectra), and more details on the worst outliers can be found in Fig. S5 of the Supplemental Material. As can be seen from Fig. 3(a), the first 20% of the test set is predicted with essentially perfect accuracy [the

first three predictions in Fig. 3(a) and the first three ticks in Fig. 3(c)]. In the next 70%, the main spectral features are correctly predicted, with some regularization visible around the finer features and a few minor intensity mismatches [the next seven predictions in Fig. 3(a) and the next seven ticks in Fig. 3(c)]. The worst L_3 prediction in the middle 95% of the test set, shown at the bottom of Fig. 3(c), is a more notable failure of the model, where the edge location and edge onset peak shape deviate significantly from the FEFF9 simulation. However, it is interesting to note that the L_2 97.5th percentile is far more accurate than the L_3 , resulting in only a mild edge location and peak shape misidentification as shown in Fig. 3(b). Although the waterfall plots in Figs. 3(a) and 3(b) show different spectra, the error of the L_3 and L_2 prediction for the same absorber is highly correlated across our testing data, as shown in Fig. S2.

C. Edge location accuracy

The previous section demonstrates that CuXASNet can successfully predict the L_2 and L_3 edge of a Cu site in a material. However, an important and perhaps more intuitive metric for the model’s accuracy is the successful prediction of the edge onset of the spectrum. The location of the edge onset

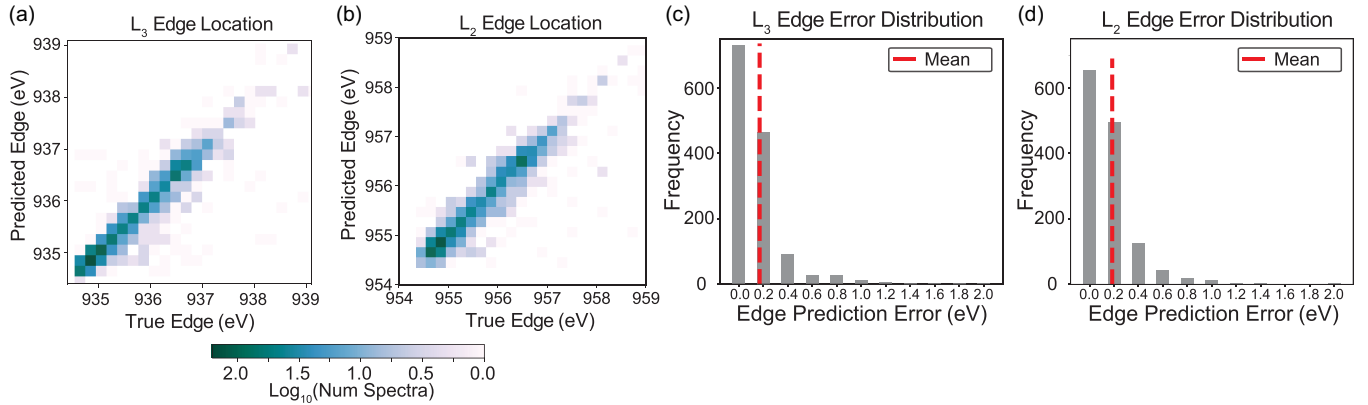


FIG. 4. Accuracy of CuXASNet in predicting the FEFF9 simulated edge onset location of the L_2 and L_3 edge spectra, as determined by finding the location of the maximum in the first derivative in the first edge onset. Two-dimensional histograms of predicted vs true edge location are shown in (a) and (b), where the density is on a log scale to allow visualization of lower accuracy regions. Bar plots of the error in this prediction are shown in (c) and (d) on a 0.2-eV scale corresponding to the energy resolution of our simulated spectra.

is determined by finding the location of the maximum in the first derivative in the first edge onset location in both the L_3 and the L_2 edges. Figures 4(c) and 4(d) show bar plots of the error in the edge location, with roughly half being predicted exactly correct for both L_2 and L_3 edges. As the spectra are set on a 0.2-eV scale, the errors naturally move in units of 0.2 eV, explaining the jumps in the location of error values across the bar plots. The distribution of the errors is very similar between the L_3 and L_2 , with their means occurring at almost identical values. Figures 4(a) and 4(b) show two-dimensional histograms of the edge location accuracy, which show similar trends to Figs. 4(c) and 4(d). Additionally, Figs. 4(a) and 4(b) show that around 300 spectra, out of the 1375 in the test set, are correctly predicted or estimated with only minor errors at between 934.8 and 935 eV for the L_3 edge and 954.8 and 955 eV for the L_2 edge. Errors greater than 0.2 eV are quite rare, and errors of over 1 eV are so uncommon that they are almost impossible to see in Figs. 4(a) and 4(b), despite projecting the density onto a log scale.

D. Comparison to experimental spectra

To determine the applicability of CuXASNet for analyzing experimental spectra, we validate its predictions against experimental spectra of the same materials. Figures 5(a)–5(c)

shows qualitative matching of spectra predicted using CuXASNet (orange) to FEFF9 (blue) and experimental spectra (black). The three materials systems compared in the figure were determined by selecting the best, a middle, and the worst MAE between CuXASNet’s predictions and experiments. By visual inspection, it is quite clear that the minor differences between the FEFF9 simulation and the ML predicted spectrum in Figs. 5(a) and 5(b) often occur when CuXASNet produces a spectrum that is a slightly smoothed version of the FEFF9 spectra, and this causes no systematic difference when comparing to the experimental spectra. Even in the worst example shown in Fig. 5(c), the error is found in the intensity of the L_3 edge rather than a misrepresentation of the features. This is confirmed quantitatively: Fig. 5(d) demonstrates that the MAEs for FEFF9 simulation experiments and CuXASNet experiments are virtually identical across different materials, irrespective of CuXASNet-FEFF9 error. This confirms CuXASNet’s potential to function as a rapid experimental analysis tool for determining the identity of an unknown experimental XAS spectrum.

E. Chemical trends in spectral prediction errors

For the general use of this model, it is essential to understand the conditions under which it can be expected to

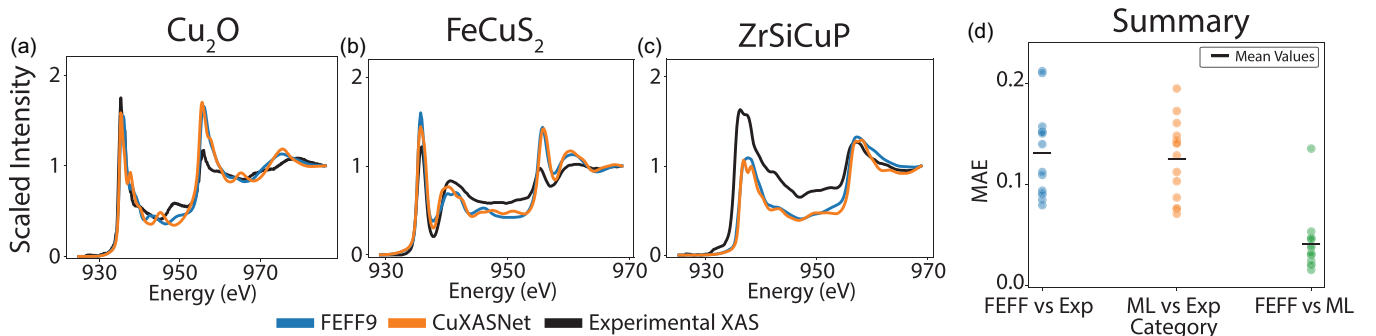


FIG. 5. Performance of CuXASNet when compared to experimental spectra. Comparisons between FEFF9 simulations (blue), CuXASNet predictions (orange), and experimental spectra (black) are shown in (a) Cu_2O , (b) FeCuS_2 , and (c) ZrSiCuP . These three materials were chosen according to the best, a middle, and the worst comparison of CuXASNet to experiments across our literature spectra. (d) Distributions of quantitative comparisons between FEFF9 and experiments (blue), ML and experiments (orange), and FEFF9 and ML (green).

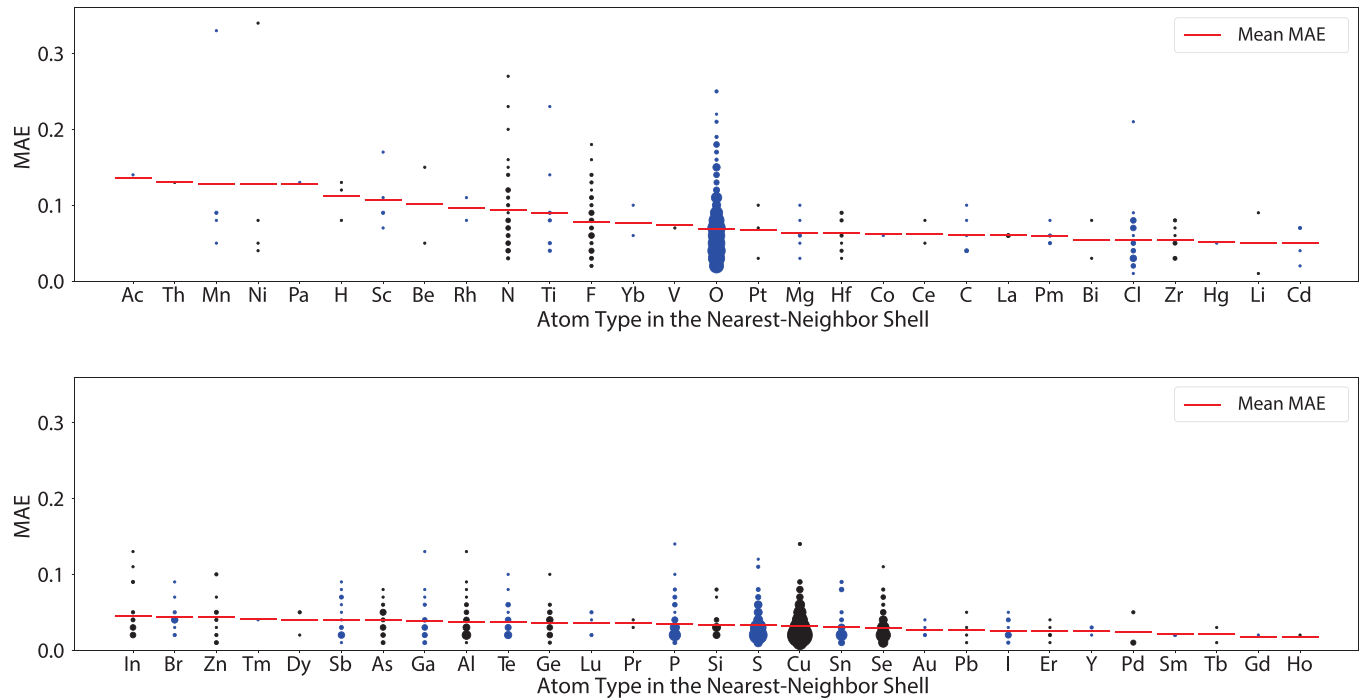


FIG. 6. The accuracy of CuXASNet when predicting Cu spectra with different nearest neighbor elemental configurations. The x axis shows the nearest neighbors of the Cu absorber. Only predictions of the L_3 edge are visualized here. The spot size is proportional to the number of spectra at that point and the red horizontal lines denote the mean MAE for that category. A spectrum can be in more than one category if the Cu atom has multiple different elements in its nearest neighbors.

perform well or struggle to reproduce FEFF9 simulations. Therefore, we explored the model performance vs OS and CN in Fig. S3. When analyzing the performance by OS, the trends reflect those seen in Fig. 2, where the OS containing more spectral uniformity are also predicted with more accuracy on average. A general decrease in accuracy, as measured by the median MAE, is observed as OS increases, although the performance on Cu(0) and Cu(I) is very similar, and there are very few Cu(III) materials in the data set, making their reliable prediction more challenging. An almost identical trend is observed in the L_2 edge, with the main difference being that there are fewer larger errors than observed in the L_3 edge prediction. The accuracy labeled by coordination environment is similarly reflected in Fig. 2, with coordination environments that have a higher diversity of spectral features generally associated with less accurate spectral prediction. Interestingly, CuXASNet performs better on CNs in the “other” category than $CN = 4$, reflecting the wide range of core-level spectral features associated with four-coordinated Cu atoms.

CuXASNet’s performance is broken down by the Cu absorber’s nearest neighbors in Fig. 6. The two worst predictions from this model, which are significant outliers, contain Cu-Ni and Cu-Mn bonds. It is clear, however, that not all Cu-Ni and Cu-Mn bonds produce large errors. Although the sample size is quite low, there are 5 Cu absorbers that contain Cu-Mn bonds and 4 Cu absorbers that contain Cu-Ni bonds, and these two errors are significantly larger than the other examples of this type of bonding. A closer examination of these two structures shows they are antiperovskites of the formula A_3BX , where A is either Mn or Ni [33]. This is the only example of antiperovskite materials in the data set, meaning users should

be wary utilizing this model to predict absorbers in antiperovskite materials. Additionally, these two materials show a specific bonding structure of Cu-metal-N bonding, which may also be less accurate for CuXASNet. However, FEFF9 also predicts a highly unusual spectrum for these materials, as shown in Fig. S5. Therefore, this specific class of materials may be challenging for FEFF9 to model. Additionally, it does not appear to be the case that conventional Cu containing perovskite materials are worse than average for FEFF9 or this model. For example, the L_3 spectrum of the perovskite structure $CsCu_2I_3$ is predicted with a high accuracy, with an MAE of only 0.035. This corresponds to a $\log(\text{MAE})$ of -1.45 , which is slightly better than the median performing spectrum in Fig. 3(c).

Beyond the two outliers discussed above, general trends in performance by elemental composition show that the worst errors among the well represented elements are commonly fluorides and nitrides. While many oxides are predicted quite accurately, there is a nontrivial amount of higher errors in Cu oxides as well. Beyond these, most Cu metallic alloys and Cu phosphides, sulfides, and selenides are predicted accurately. Among the halogens, a clear positive correlation is seen between accuracy and atomic number, with $F < Cl < Br < I$, although iodides and bromides are not very well represented in the data set.

Additionally, we have explored the 20 most inaccurate predictions to determine any significant trends. Through this exploration, it was discovered that of the 20 worst structures, 17 are magnetic and only 3 are not. Of the worst ten, only one is not magnetic. A summary figure describing accuracy on magnetic and nonmagnetic structures in the full test set

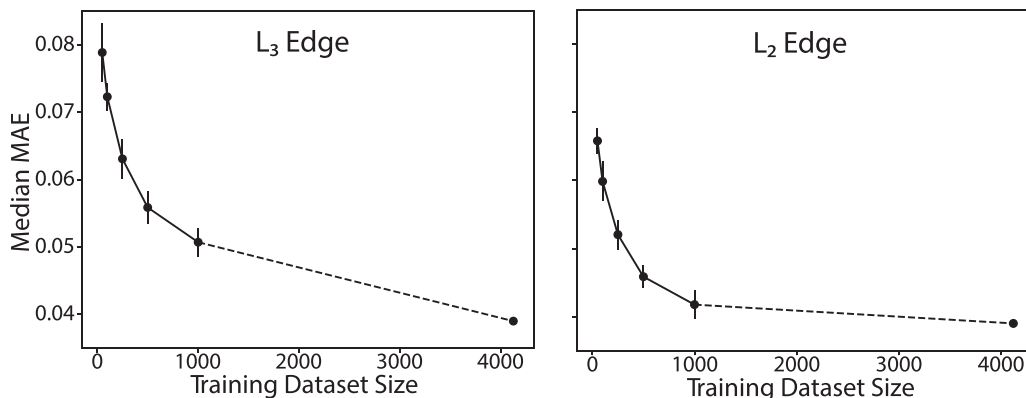


FIG. 7. Accuracy of the model, as determined by the median MAE, when the model is applied to the test set using different volumes of training spectra. The smaller training sets are determined by random sampling the full training set used for the model presented in this work. The vertical bars on each scatter spot show two standard deviations of the median MAE when ten different random samples of training data are used. In each case, the resulting model is predicted on the full test set to determine its accuracy.

is shown in Fig. S6. Although the trend is less clear for the full test set, magnetic structures perform notably worse than nonmagnetic structures in Fig. S6.

F. Volume of training data necessary for inference

One significant extension of this model is its potential applicability to more computationally intensive simulations, such as the OCEAN or VASP codes [28,29,54]. The model and analysis presented in this work were conducted using a data set of 5497 spectra, with 4123 used for training. However, for systems like Ti and Fe, FEFF9 can struggle to represent the L -edge spectra, necessitating more computationally intensive simulations, the generation of which can be time and resource prohibitive. Therefore, in this work we have examined the volume of data necessary for accurate simulated spectral inference. This is visualized in Fig. 7, which shows the model's accuracy on our test data as a function of training data volume. In this case, the smaller training set is random sampled from the 4000 spectra used for training the full CuXASNet model, with the test set held fixed. Figure 7 shows that the increased performance from adding additional spectra begins diminishing at around 500 spectra, and that the L_2 edge reaches its steady state with fewer training spectra. This is unsurprising, given the relatively fewer significant spectral features in the L_2 edge (Fig. 2) and the model's increased accuracy on the L_2 edge in general as shown in Fig. 3. The relatively low volume of data required to hit reasonable spectral prediction indicates that this model architecture could be used to predict spectra of higher computational cost without requiring a data set that is computationally prohibitive.

G. Predictions beyond crystalline structures

To demonstrate the wide-ranging utility of CuXASNet on more atypical structures not originating from more standard crystalline materials, we have examined the literature for a Cu-containing molecule where an L -edge spectrum for Cu absorption had also been recorded [69]. The molecular complex examined here is a peroxo-Cu(II) species that is created when coupled binuclear copper proteins bind

molecular oxygen [69]. We extracted the structure of the complex from Ref. [68] and compare the predicted Cu $L_{2,3}$ edge from CuXASNet with the experimental spectrum extracted from Ref. [69] in Fig. 8. After manually aligning the L_3 peak energy to match the experimental spectrum, we observe reasonable agreement between CuXASNet and experiments up to the baseline intensity between the L_3 and L_2 peaks. It is important to note that the same manual alignment was applied to both the FEFF9 and CuXASNet spectrum, and the value was chosen to align the FEFF9 L_3 peak to the experimental L_3 peak, which explains why the CuXASNet peak is slightly misaligned. The baseline intensity mismatch between the experimental spectrum and the FEFF9/CuXASNet spectra is an artifact of a processing procedure done in Ref. [69] in order to conduct density of state calculations. Despite this, CuXASNet overestimates the L_2 edge intensity relative to the L_3 edge when comparing to the FEFF9 simulation for this complex as well. However, it should be noted that CuXASNet is better able to predict the small peak between the

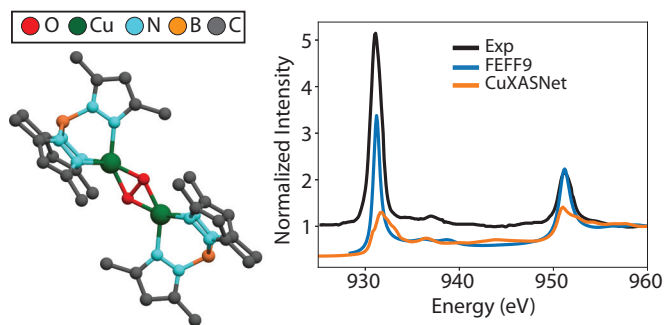


FIG. 8. An application of our CuXASNet model to an uncommon and complicated Cu organometallic structure, extracted from Ref. [68]. A depiction of this structure is shown on the left, focusing on the region of the structure containing the Cu atoms. Hydrogen atoms are removed for clarity. A comparison between an experimental spectrum taken from Ref. [69], a FEFF9 simulation and the spectrum predicted using CuXASNet are shown on the right. The spectra agree up to the baseline intensity between the L_2 and L_3 peaks.

L_3 and L_2 observed in the experimental spectrum, which is rendered much less accurately by FEFF9. This example shows a reasonable approximation for structure far outside the realm of its training data and demonstrates that CuXASNet is a viable tool for fast spectral screening of candidate structures for an unknown system, regardless of structural complexity. This can inform structural candidates for manual simulation studies and help verify an unknown material's composition and structural parameters. To further enhance CuXASNet's ability to predict accurate spectra of unknown complicated structures, CuXASNet can be dynamically refined using an active learning loop.

H. CuXASNet extensions

In this work, we have illustrated the success of CuXASNet at predicting FEFF9 L -edge spectra. However, there are instances in which FEFF9 is unable to accurately model experimental spectra, which can limit the generalizability of CuXASNet. Improving this can be accomplished by adding higher computational cost simulations and integrating experimental spectra into the training data, where a sufficiently sized data set of experimental spectra can be produced via literature mining and high throughput experiments. Future work will seek to improve CuXASNet to the point where its predictions quantitatively surpass the experimental accuracy of FEFF9 using these methods.

IV. CONCLUSION

This work showcases CuXASNet, an ML model which can predict the L -edge XAS spectrum of Cu based on the material's atomic structure. CuXASNet utilizes a graph neural network, M3GNet, to featurize materials or clusters into a fixed length vector representation, which is then fed to the neural network for spectral prediction. CuXASNet predicts the L_3 and L_2 edges for Cu sites with quantitative accuracy to FEFF9. The model shows excellent accuracy on Cu alloys and is highly reliable on most oxide materials. Additionally, our predicted spectra have good agreement with experimental spectra from the literature. Our predicted spectra match experiments with an average Spearman's correlation coefficient of 0.891 and an average MAE of 0.125, matching FEFF9's experimental comparison values of 0.898 and 0.131 for these metrics. We also demonstrate that CuXASNet is able to predict spectra for structures outside of conventional crystalline materials by turning it to a complicated organometallic Cu molecule.

The accuracy of CuXASNet on simulated data, and the viability of the experimental comparison relative to simulated spectra, shows that CuXASNet is a highly valuable tool for rapidly generating more experimentally viable simulated data. There are many interesting potential applications for this model, including structural determination from spectra, targeted materials synthesis, and the rapid and accurate gen-

eration of training data for ML models. The simulated spectra predicted by CuXASNet can be used for theoretical studies of Cu L -edge core-level spectroscopy and for the rational design of materials. CuXASNet predicts an XAS spectrum from a specified structure, which allows large-scale studies of the impact of atomic structural changes on Cu's core-level spectrum. This can be highly relevant in functional material development, especially in semiconductor engineering, where minor structural changes involving specific atom substitution with dopants and the introduction of atomic level defects are a primary method of engineering specific electronic properties. Additionally, through the use of a model that can predict thousands of spectra in minutes from atomic structures [70], rational design of Cu-containing materials can be performed by conducting a bulk search of relevant structures when a material with a specific set of properties is desired. The core-level spectrum can elucidate many applicable properties of a functional material, and CuXASNet can be used to determine structural candidates for specific applications. Additionally, we introduce a model framework that can be extended to predict spectra of more computationally intensive simulations and other transition metal L edges.

ACKNOWLEDGMENTS

This work was primarily funded by the U.S. Department of Energy program "4D Camera Distillery: From Massive Electron Microscopy Scattering Data to Useful Information with AI/ML." Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This research was supported by the U.S. Department of Energy, Office of Science, Office Basic Energy Sciences, Award No. FWP PS-030, and used Theory and Computation resources of the Center for Functional Nanomaterials (CFN), which is a U.S. Department of Energy Office of Science User Facility, at Brookhaven National Laboratory under Contract No. DE-SC0012704.

S.P.G. developed the simulated XAS data set, conducted the machine learning training and analysis, and wrote the manuscript. M.R.C. developed the automated procedure to generate input files for FEFF9 spectral simulation, provided access to the M3GNet model used in this work, and provided advising on the development of the machine learning models. D.L. provided training and expertise necessary to produce the simulated data set and provided project direction and advising. J.C. provided experimental core-level spectroscopy knowledge, led the collaboration, and designed the scope of this work. All authors read, edited, and approved the final manuscript.

DATA AVAILABILITY

The spectral data set and the code to predict and analyze CuXASNet can be found in the GitHub repository [71].

[1] F. De Groot and A. Kotani, *Core Level Spectroscopy of Solids* (CRC, Boca Raton, 2008)

[2] F. M. F. de Groot, J. C. Fuggle, B. T. Thole, and G. A. Sawatzky, 2p x-ray absorption of 3d transition-metal compounds: An

- atomic multiplet description including the crystal field, *Phys. Rev. B* **42**, 5459 (1990).
- [3] Y. Bai, Y. Wu, X. Zhou, Y. Ye, K. Nie, J. Wang, M. Xie, Z. Zhang, Z. Liu, T. Cheng, and C. Gao, Promoting nickel oxidation state transitions in single-layer NiFeB hydroxide nanosheets for efficient oxygen evolution, *Nat. Commun.* **13**, 6094 (2022).
 - [4] M. Kubin, M. Guo, T. Kroll, H. Lochel, E. Kallman, M. L. Baker, R. Mitzner, S. Gul, J. Kern, A. Fohlsch, A. Erko, U. Bergmann, V. Yachandra, J. Yano, M. Lundberg, and P. Wernet, Probing the oxidation state of transition metal complexes: a case study on how charge and spin densities determine Mn L-edge X-ray absorption energies, *Chem. Sci.* **9**, 6813 (2018).
 - [5] E. Marelli, J. Lyu, M. Morin, M. Leménager, T. Shang, N. S. Yüzbaşı, D. Aegerter, J. Huang, N. D. Daffé, A. H. Clark, D. Sheptyakov, T. Graule, M. Nachtegaal, E. Pomjakushina, T. J. Schmidt, M. Krack, E. Fabbri, and M. Medarde, Cobalt-free layered perovskites $\text{RBaCuFeO}_{5+\delta}$ ($\text{R} = 4\text{f lanthanide}$) as electrocatalysts for the oxygen evolution reaction, *EES Catal.* **2**, 335 (2024).
 - [6] K. Placheta, A. Kot, J. Banas-Gac, M. Zając, M. Sikora, M. Radecka, and K. Zakrzewska, Evolution of surface properties of titanium oxide thin films, *Appl. Surf. Sci.* **608**, 155046 (2023).
 - [7] T. Jenkins, J. A. Alarco, B. Cowie, and I. D. Mackinnon, Direct spectroscopic observation of the reversible redox mechanism in $\text{A}_3\text{V}_2(\text{PO}_4)_3$ ($\text{A}=\text{Li,Na}$) cathode materials for Li-ion batteries, *J. Power Sources* **571**, 233078 (2023).
 - [8] J. Yano and V. K. Yachandra, X-ray absorption spectroscopy, *Photosynth. Res.* **102**, 241 (2009).
 - [9] S. P. Ong, S. Cholia, A. Jain, M. Brafman, D. Gunter, G. Ceder, and K. A. Persson, The Materials Application Programming Interface (API): A simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles, *Comput. Mater. Sci.* **97**, 209 (2015).
 - [10] S. R. Kharel, F. Meng, X. Qu, M. R. Carbone, and D. Lu, A universal deep learning framework for materials x-ray absorption spectra, *Phys. Rev. Mater.* **9**, 043803 (2025).
 - [11] K. M. P. Wheelhouse, R. L. Webster, and G. L. Beutner, Advances and applications in catalysis with earth-abundant metals, *Org. Process Res. Dev.* **27**, 1157 (2023).
 - [12] K. E. Dalle, J. Warnan, J. J. Leung, B. Reuillard, I. S. Karmel, and E. Reisner, Electro- and solar-driven fuel synthesis with first row transition metal complexes, *Chem. Rev.* **119**, 2752 (2019).
 - [13] W. Gao, Y. Xu, L. Fu, X. Chang, and B. Xu, Experimental evidence of distinct sites for CO_2 -to-CO and CO conversion on Cu in the electrochemical CO_2 reduction reaction, *Nat. Catal.* **6**, 885 (2023).
 - [14] M. B. Gawande, A. Goswami, F. X. Felpin, T. Asefa, X. Huang, R. Silva, X. Zou, R. Zboril, and R. S. Varma, Cu and Cu-based nanoparticles: Synthesis and applications in catalysis, *Chem. Rev.* **116**, 3722 (2016).
 - [15] B. Dervaux and F. E. Du Prez, Heterogeneous azide-alkyne click chemistry: Towards metal-free end products, *Chem. Sci.* **3**, 959 (2012).
 - [16] D. Astruc, L. Liang, A. Rapakousiou, and J. Ruiz, Click dendrimers and triazole-related aspects: Catalysts, mechanism, synthesis, and functions. A bridge between dendritic architectures and nanomaterials, *Acc. Chem. Res.* **45**, 630 (2012).
 - [17] J. K. McCusker, Electronic structure in the transition metal block and its implications for light harvesting, *Science* **363**, 484 (2019).
 - [18] Y. Liu, S. C. Yiu, C. L. Ho, and W. Y. Wong, Recent advances in copper complexes for electrical/light energy conversion, *Coord. Chem. Rev.* **375**, 514 (2018).
 - [19] M. B. Frye and L. M. Garten, Reaching the potential of ferroelectric photovoltaics, *Acc. Mater. Res.* **4**, 906 (2023).
 - [20] B. O'Regan and M. Grätzel, A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO_2 films, *Nature (London)* **353**, 737 (1991).
 - [21] A. K. Renfrew, E. S. O'Neill, T. W. Hambley, and E. J. New, Harnessing the properties of cobalt coordination complexes for biological application, *Coord. Chem. Rev.* **375**, 221 (2018).
 - [22] C. Graham, A. Mezzadrelli, W. Senaratne, S. Pal, D. Thelen, L. Hepburn, P. Mazumder, and V. Pruneri, Towards transparent and durable copper-containing antimicrobial surfaces, *Commun. Mater.* **5**, 39 (2024).
 - [23] S. Wang, H. Wu, K. Sun, J. Hu, F. Chen, W. Liu, J. Chen, B. Sun, and A. M. S. Hossain, A novel pH-responsive Fe-MOF system for enhanced cancer treatment mediated by the Fenton reaction, *New J. Chem.* **45**, 3271 (2021).
 - [24] R. Cui, P. Zhao, Y. Yan, G. Bao, A. Damirin, and Z. Liu, Outstanding drug-loading/release capacity of hollow ferrometal-organic framework-based microcapsules: A potential multifunctional drug-delivery platform, *Inorg. Chem.* **60**, 1664 (2021).
 - [25] G. Cibi, D. Gianolio, S. A. Parry, T. Schoonjans, O. Moore, R. Draper, L. A. Miller, A. Thoma, C. L. Doswell, and A. Graham, An open access, integrated XAS data repository at Diamond Light Source, *Radiat. Phys. Chem.* **175**, 108479 (2020).
 - [26] C. Segre, *IXAS XAFS Database* (Center for Synchrotron Radiation Research and Instrumentation at Illinois Institute of Technology, Chicago, 2008).
 - [27] Y. Chen, C. Chen, C. Zheng, S. Dwaraknath, M. K. Horton, J. Cabana, J. Rehr, J. Vinson, A. Dozier, J. J. Kas, K. A. Persson, and S. P. Ong, Database of *ab initio* L-edge x-ray absorption near edge structure, *Sci. Data* **8**, 153 (2021).
 - [28] J. Vinson, J. J. Rehr, J. J. Kas, and E. L. Shirley, Bethe-Salpeter equation calculations of core excitation spectra, *Phys. Rev. B* **83**, 115106 (2011).
 - [29] J. Vinson, Advances in the ocean-3 spectroscopy package, *Phys. Chem. Chem. Phys.* **24**, 12787 (2022).
 - [30] C. Zheng, C. Chen, Y. Chen, and S. P. Ong, Random forest models for accurate identification of coordination environments from x-ray absorption near-edge structure, *Patterns* **1**, 100013 (2020).
 - [31] M. R. Carbone, F. Meng, C. Vorwerk, B. Maurer, F. Peschel, X. Qu, E. Stavitski, C. Draxl, J. Vinson, and D. Lu, Lightshow: a Python package for generating computational x-ray absorption spectroscopy input files, *J. Open Source Softw.* **8**, 5182 (2023).
 - [32] J. J. Rehr, J. J. Kas, F. D. Vila, M. P. Prange, and K. Jorissen, Parameter-free calculations of x-ray spectra with FEFF9, *Phys. Chem. Chem. Phys.* **12**, 5503 (2010).
 - [33] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder *et al.*, Commentary: The Materials Project: A materials genome approach

- to accelerating materials innovation, *APL Mater.* **1**, 011002 (2013).
- [34] C. Zheng, K. Mathew, C. Chen, Y. Chen, H. Tang, A. Dozier, J. J. Kas, F. D. Vila, J. J. Rehr, L. F. Piper, K. A. Persson, and S. P. Ong, Automated generation and ensemble-learned matching of X-ray absorption spectra, *npj Comput. Mater.* **4**, 12 (2018).
- [35] K. Mathew, C. Zheng, D. Winston, C. Chen, A. Dozier, J. J. Rehr, S. P. Ong, and K. A. Persson, Data descriptor: High-throughput computational x-ray absorption spectroscopy, *Sci. Data* **5**, 180151 (2018).
- [36] K. Gilmore, J. Vinson, E. L. Shirley, D. Prendergast, C. D. Pemmaraju, J. J. Kas, F. D. Vila, and J. J. Rehr, Efficient implementation of core-excitation Bethe-Salpeter equation calculations, *Comput. Phys. Commun.* **197**, 109 (2015).
- [37] S. B. Torrisi, M. R. Carbone, B. A. Rohr, J. H. Montoya, Y. Ha, J. Yano, S. K. Suram, and L. Hung, Random forest machine learning models for interpretable x-ray absorption near-edge structure spectrum-property relationships, *npj Comput. Mater.* **6**, 109 (2020).
- [38] J. Timoshenko, D. Lu, Y. Lin, and A. I. Frenkel, Supervised machine-learning-based determination of three-dimensional structure of metallic nanoparticles, *J. Phys. Chem. Lett.* **8**, 5091 (2017).
- [39] S. Tetef, N. Govind, and G. T. Seidler, Unsupervised machine learning for unbiased chemical classification in x-ray absorption spectroscopy and x-ray emission spectroscopy, *Phys. Chem. Chem. Phys.* **23**, 23586 (2021).
- [40] S. Tetef, V. Kashyap, W. M. Holden, A. Velian, N. Govind, and G. T. Seidler, Informed chemical classification of organophosphorus compounds via unsupervised machine learning of x-ray absorption spectroscopy and x-ray emission spectroscopy, *J. Phys. Chem. A* **126**, 4862 (2022).
- [41] M. Higashi and H. Ikeno, Extraction of local structure information from x-ray absorption near-edge structure: A machine learning approach, *Mater. Trans.* **64**, 2179 (2023).
- [42] M. R. Carbone, P. M. Maffettone, X. Qu, S. Yoo, and D. Lu, Accurate, uncertainty-aware classification of molecular chemical motifs from multimodal x-ray absorption spectroscopy, *J. Phys. Chem. A* **128**, 1948 (2024).
- [43] S. P. Gleason, D. Lu, and J. Ciston, Prediction of the Cu oxidation state from EELS and XAS spectra using supervised machine learning, *npj Comput. Mater.* **10**, 221 (2024).
- [44] M. Chatzidakis and G. A. Botton, Towards calibration-invariant spectroscopy using deep learning, *Sci. Rep.* **9**, 2126 (2019).
- [45] D. del-Pozo-Bueno, D. Kepaptsoglou, F. Peiró, and S. Estradé, Comparative of machine learning classification strategies for electron energy loss spectroscopy: Support vector machines and artificial neural networks, *Ultramicroscopy* **253**, 113828 (2023).
- [46] T. David, N. K. Nik Aznan, K. Garside, and T. Penfold, Towards the automated extraction of structural information from x-ray absorption spectra, *Digital Discovery* **2**, 1461 (2023).
- [47] S. A. Guda, A. S. Algasov, A. A. Guda, A. Martini, A. N. Kravtsova, A. L. Bugaev, L. V. Guda, and A. V. Soldatov, Search for analytical relations between x-ray absorption spectra descriptors and the local atomic structure using machine learning, *J. Surf. Investig.* **15**, 934 (2021).
- [48] A. A. Guda, S. A. Guda, A. Martini, A. N. Kravtsova, A. Algasov, A. Bugaev, S. P. Kubrin, L. V. Guda, P. Šot, J. A. van Bokhoven, C. Copéret, and A. V. Soldatov, Understanding x-ray absorption spectra by means of descriptors and machine learning algorithms, *npj Comput. Mater.* **7**, 203 (2021).
- [49] Z. Ji, M. Hu, and H. L. Xin, MnEdgeNet for accurate decomposition of mixed oxidation states for Mn XAS and EELS L2,3 edges without reference and calibration, *Sci. Rep.* **13**, 14132 (2023).
- [50] Y. Chen, C. Chen, I. Hwang, M. J. Davis, W. Yang, C. Sun, G. H. Lee, D. McReynolds, D. Allan, J. Marulanda Arias, S. P. Ong, and M. K. Chan, Robust machine learning inference from x-ray absorption near edge spectra through featurization, *Chem. Mater.* **36**, 2304 (2024).
- [51] C. D. Rankine, M. M. Madkhali, M. M. Madkhali, and T. J. Penfold, A deep neural network for the rapid prediction of x-ray absorption spectra, *J. Phys. Chem. A* **124**, 4263 (2020).
- [52] C. D. Rankine and T. J. Penfold, Accurate, affordable, and generalizable machine learning simulations of transition metal x-ray absorption spectra using the XANESNET deep neural network, *J. Chem. Phys.* **156**, 164102 (2022).
- [53] L. Watson, C. D. Rankine, and T. J. Penfold, Beyond structural insight: a deep neural network for the prediction of Pt L2/3-edge x-ray absorption spectra, *Phys. Chem. Chem. Phys.* **24**, 9156 (2022).
- [54] J. Hafner, *Ab-initio* simulations of materials using VASP: Density-functional theory and beyond, *J. Comput. Chem.* **29**, 2044 (2008).
- [55] M. W. Haverkort, M. Zwierzycki, and O. K. Andersen, Multiplet ligand-field theory using Wannier orbitals, *Phys. Rev. B* **85**, 165113 (2012).
- [56] F. Meng, B. Maurer, F. Peschel, S. Selcuk, M. Hybertsen, X. Qu, C. Vorwerk, C. Draxl, J. Vinson, and D. Lu, Multicode benchmark on simulated Ti K-edge x-ray absorption spectra of Ti-O compounds, *Phys. Rev. Mater.* **8**, 013801 (2024).
- [57] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization* (3rd International Conference for Learning Representations, San Diego, 2014).
- [58] F. Marin, A. Rohatgi, and S. Charlot, WebPlotDigitizer, a polyvalent and free software to extract spectra from old astronomical publications: application to ultraviolet spectropolarimetry (2017), <https://automeris.io>.
- [59] S. W. Goh, A. N. Buckley, R. N. Lamb, R. A. Rosenberg, and D. Moran, The oxidation states of copper and iron in mineral sulfides, and the oxides formed on initial exposure of chalcopyrite and bornite to air, *Geochim. Cosmochim. Acta* **70**, 2210 (2006).
- [60] R. Liu, J. Chen, P. Nachimuthu, R. Gundakaram, C. Jung, J. Kim, and S. Lee, Evidence for electron-doped (n-type) superconductivity in the infinite-layer (Sr_{0.9}La_{0.1})CuO₂ compound by x-ray absorption near-edge spectroscopy, *Solid State Commun.* **118**, 367 (2001).
- [61] A. N. Buckley, W. M. Skinner, S. L. Harmer, A. Pring, and L. J. Fan, Electronic environments in carrollite, CuCo₂S₄, determined by soft x-ray photoelectron and absorption spectroscopy, *Geochim. Cosmochim. Acta* **73**, 4452 (2009).
- [62] M. Glaser, F. Ciccullo, E. Giangrisostomi, R. Ovsyannikov, A. Calzolari, and M. B. Casu, Doping and oxidation effects under ambient conditions in copper surfaces: A “real-life” CuBe surface, *J. Mater. Chem. C* **6**, 2769 (2018).
- [63] B. W. Rudyk, P. E. Blanchard, R. G. Cavell, and A. Mar, Electronic structure of lanthanum copper oxychalcogenides LaCuOCh (Ch=S, Se, Te) by x-ray photoelectron

- and absorption spectroscopy, *J. Solid State Chem.* **184**, 1649 (2011).
- [64] P. E. Blanchard, R. G. Cavell, and A. Mar, Electronic structure of ZrCuSiAs and ZrCuSiP by x-ray photoelectron and absorption spectroscopy, *J. Solid State Chem.* **183**, 1536 (2010).
- [65] M. Grioni, J. F. Ancker, M. T. Czyzyk, and J. C. Fuggle, Unoccupied electronic structure and core-hole effects in the x-ray-absorption spectra of Cu₂O, *Phys. Rev. B* **45**, 3309 (1992).
- [66] M. R. Carbone, S. Yoo, M. Topsakal, and D. Lu, Classification of local chemical environments from x-ray absorption spectra using supervised machine learning, *Phys. Rev. Mater.* **3**, 033604 (2019).
- [67] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/xz7f-srky> for six additional figures providing additional analysis on cuxasnet's performance and training data.
- [68] N. Kitajima, K. Fujisawa, C. Fujimoto, Y. Moro-oka, S. Hashimoto, T. Kitagawa, K. Toriumi, K. Tatsumi, and A. Nakamura, Oxygen complexes and oxygen activation by transition metals, *J. Am. Chem. Soc.* **114**, 1277 (1992).
- [69] M. F. Qayyum, R. Sarangi, K. Fujisawa, T. D. P. Stack, K. D. Karlin, K. O. Hodgson, B. Hedman, and E. I. Solomon, L-edge x-ray absorption spectroscopy and DFT calculations on Cu 2O₂ species: Direct electrophilic aromatic attack by side-on peroxo bridged dicopper(II) complexes, *J. Am. Chem. Soc.* **135**, 17417 (2013).
- [70] H. Kwon, T. Hsu, W. Sun, W. Jeong, F. Aydin, J. Chapman, X. Chen, M. R. Carbone, D. Lu, F. Zhou, and T. A. Pham, Spectroscopy-guided discovery of three-dimensional structures of disordered materials with diffusion models, [arXiv:2312.05472](https://arxiv.org/abs/2312.05472).
- [71] <https://github.com/smglsl12/CuXASNet>.
- Correction:* The formatting in row 4 of Table I has been fixed. Typographical errors in the citations of Figs. 4(a), 4(b), 4(c), and 4(d) in Sec. III C have been fixed.