The Journal of Chemical Physics

Mapping XANES spectra on structural descriptors of copper oxide clusters using supervised machine learning

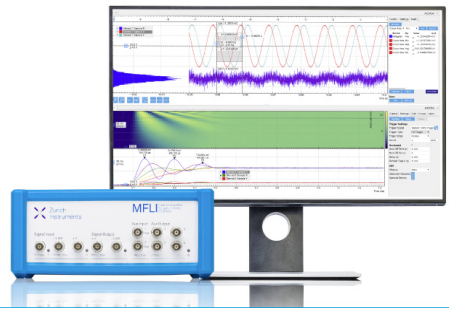
Cite as: J. Chem. Phys. **151**, 164201 (2019); https://doi.org/10.1063/1.5126597
Submitted: 04 September 2019 . Accepted: 01 October 2019 . Published Online: 22 October 2019

Yang Liu, Nicholas Marcella, Janis Timoshenko, Avik Halder, Bing Yang, Lakshmi Kolipaka, Michael. J. Pellin, Soenke Seifert, Stefan Vajda, Ping Liu, and Anatoly I. Frenkel

COLLECTIONS

Note: The paper is part of the JCP Special Topic Collection on Catalytic Properties of Model Supported Nanoparticles.

This paper was selected as an Editor's Pick



The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Mapping XANES spectra on structural descriptors of copper oxide clusters using supervised machine learning

Cite as: J. Chem. Phys. 151, 164201 (2019); doi: 10.1063/1.5126597
Submitted: 4 September 2019 • Accepted: 1 October 2019 •
Published Online: 22 October 2019

Yang Liu,[1,2] Nicholas Marcella,[2] Janis Timoshenko,[2] Avik Halder,[3] Bing Yang,[3] Lakshmi Kolipaka,[3] Michael. J. Pellin,[3] Soenke Seifert,[4] Stefan Vajda,[3,5,6] Ping Liu,[7] and Anatoly I. Frenkel[2,7,a)]

AFFILIATIONS

[1]Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, USA
[2]Department of Materials Science and Chemical Engineering, Stony Brook University, Stony Brook, New York 11794, USA
[3]Materials Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA
[4]X-ray Sciences Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA
[5]Institute for Molecular Engineering, The University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, USA
[6]Department of Nanocatalysis, J. Heyrovský Institute of Physical Chemistry, Czech Academy of Sciences, Dolejškova 3, 18223 Prague 8, Czech Republic
[7]Division of Chemistry, Brookhaven National Laboratory, Upton, New York 11973, USA

Note: The paper is part of the JCP Special Topic Collection on Catalytic Properties of Model Supported Nanoparticles.
[a)]**Author to whom correspondence should be addressed:** anatoly.frenkel@stonybrook.edu

ABSTRACT

Understanding the origins of enhanced reactivity of supported, subnanometer in size, metal oxide clusters is challenging due to the scarcity of methods capable to extract atomic-level information from the experimental data. Due to both the sensitivity of X-ray absorption near edge structure (XANES) spectroscopy to the local geometry around metal ions and reliability of theoretical spectroscopy codes for modeling XANES spectra, supervised machine learning approach has become a powerful tool for extracting structural information from the experimental spectra. Here, we present the application of this method to grazing incidence XANES spectra of size-selective Cu oxide clusters on flat support, measured in *operando* conditions of the methanation reaction. We demonstrate that the convolution neural network can be trained on theoretical spectra and utilized to "invert" experimental XANES data to obtain structural descriptors—the Cu–Cu coordination numbers. As a result, we were able to distinguish between different structural motifs (Cu_2O-like and CuO-like) of Cu oxide clusters, transforming in reaction conditions, and reliably evaluate average cluster sizes, with important implications for the understanding of structure, composition, and function relationships in catalysis.

INTRODUCTION

Metal oxides as heterogeneous catalysts have received considerable attention in both fundamental research and industrial applications.[1–4] For instance, metal oxide catalysts (MOCs) possess high catalytic performance and robustness in the water oxidation reaction.[1,5,6] In industry, MOCs are crucial for the asphaltene adsorption to enhance the oil discovery.[2] The metal oxide nanocatalysts, in particular, display unique electronic properties due to their non-bulk-like coordination geometry and redox properties.[7–9] To understand the activities of metal oxide nanocatalysts, identification of the active sites of the catalysts[10–13] and, importantly, the size and shape of the particles[14–16] are required. The geometric properties of nanoparticles play greater role in their activity

mechanisms because of the larger surface-to-volume ratio,[17] compared to bulk-like particles. Due to the large range of possible structures, the catalytic activities of nanocatalysts exhibit large variation with different geometry.[14–16,18] Besides, the nanocatalysts can undergo agglomeration under reaction conditions,[19,20] affecting their catalytic activity.

In a toolbox of characterization methods tailored for understanding catalytic mechanisms, a prominent place is taken by the *operando* method, in which the structure of the catalysts is analyzed in real time, during the reaction, and the reaction products are detected simultaneously with the structural measurement to build the structure-reactivity relation.[21–23] Due to the formidable challenge that low metal loadings and high reaction temperature and/or pressure present to many techniques, extended X-ray absorption fine structure (EXAFS) spectroscopy,[24–26] the workhorse method for catalytic studies, is limited in its applicability to MOCs. X-ray absorption near-edge structure (XANES) is measured in the same X-ray absorption spectroscopy experiment and has better signal-to-noise ratio than EXAFS; hence, it can be advantageous for use in the *in situ*/*operando* catalytic experiments.[27] XANES is also sensitive to the arrangements of atoms and electronic characteristics[28–31] and is less affected by structural disorder compared to EXAFS.[24,26] For some model catalysts, such as size-selective clusters supported on single crystal surfaces, for which EXAFS data cannot be obtained due to their ultra-low weight loadings, grazing incidence (GI) XANES becomes a unique tool to monitor the transformations in the oxidation state, structure, and/or size of the cluster.[32–34] However, GI XANES has been rarely employed for structural characterization due to the limitations in its quantitative analysis. Recently, we demonstrated that a supervised machine learning-based method enables the establishment of relation between XANES spectral features and structural descriptors of monometallic nanoparticles.[35,36] By employing an artificial neural network (NN) trained on the large set of theoretical XANES, we were able to obtain metal-metal coordination numbers (CNs) and investigate the structure of monometallic nanoparticles and size-selective clusters.[35–39] In all prior cases, we deliberately selected well reduced systems to eliminate metal-nonmetal bonding that would have complicated neural network training and applications. That limitation precludes the broad applicability of our NN-based XANES analysis for *operando* studies, in which the changes in chemical states of the catalysts may occur in real reaction conditions.

In this work, we report the application of the convolution neural network-based method to analysis of the structure and chemical state of size selective copper oxide clusters measured by XANES during their catalytic reaction process. Copper oxide catalysts are known for their good reactivity and selectivity in numerous oxidation and reduction reactions.[40–47] One of the important reactions is CO_2 methanation, which can assist the conversion of CO_2 to chemical feedstock and benefit the inhibition of CO_2 emission.[48,49] We used GI XANES spectra collected for Cu size-selective clusters in the *operando* experiment during the process of catalytic CO_2 methanation to extract information about the oxidation state and size of the clusters. In what follows, we present our method for training and validating NN, describe the experimental data chosen to illustrate its application to MOCs, and demonstrate the applicability of this approach to extract their structural descriptors in *operando* conditions.

NEURAL NETWORK TRAINING AND VALIDATION

The common route for NN construction is preparing training sets, training the NN, and validating the NN. From our previous works, it is known that for the construction of training set, we need hundreds of thousands XANES data with unique and *a priori* known labels (that is, structural descriptors). It is not feasible to obtain such a large number of labeled data from experimental measurements for this purpose. *Ab initio* XANES simulations could be a good alternative to the experimental spectra, as demonstrated in our prior work.[35,36,39] Before planning the training with theory-generated spectra, it is important to verify that a given method or code used for simulations reliably reproduces standard compounds. For example, FEFF9[50] is adequate for reproducing experimental XANES of bulk Cu oxides, as illustrated in Fig. 1. The details of the XANES simulation are given in Note I of the supplementary material.

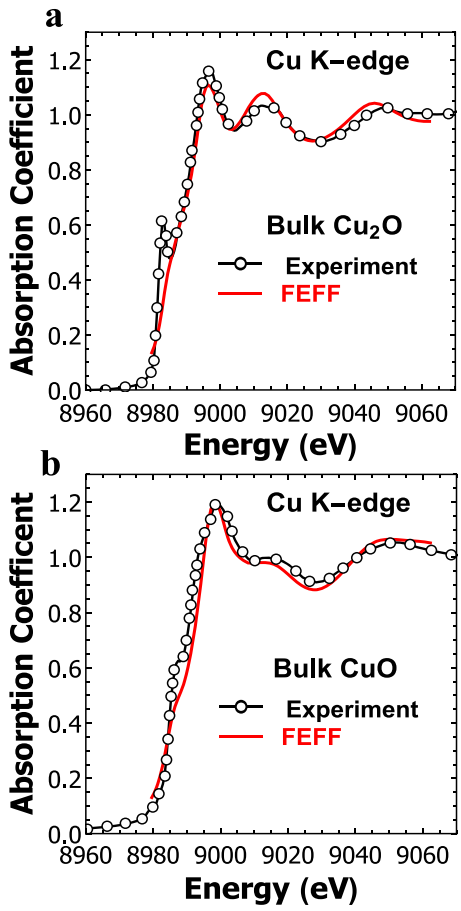


FIG. 1. Experimental and theoretical (calculated with FEFF9) XANES spectra for the bulk Cu_2O (a) and CuO (b) standard compounds.

Verifying the sensitivity of XANES to the size and structure of the nanoscale oxide is the necessary first step for any regression-based method, in general, and NN-based method, in particular, to work. Following the strategy, first implemented in Ref. 36, we first examine the absorption site effect on XANES spectra, as illustrated in Figs. 2(a) and 2(b). Each spectrum in Figs. 2(a) and 2(b) is labeled with two Cu–Cu CNs (for the first coordination shell and second coordination shell) to represent the structure of copper oxide nanoparticles. By comparing the theoretical XANES (calculated by FEFF9) on different sites of CuO and Cu_2O models (indicated by their respective CNs of the 1st and 2nd nearest neighbors), more pronounced features are captured by XANES for the copper atom with the larger Cu–Cu CNs [Figs. 2(a) and 2(b)]. The XANES spectrum for the copper atom in the inner shell of copper oxide model has greater resemblance of the XANES spectrum for the bulk of copper oxide. In contrast, XANES calculated for the copper atom on the surface has relatively more smooth features. After establishing the absorption site dependence, we examined the cluster size effect on XANES by averaging the site-specific spectra over all atoms in the simulated CuO-like and Cu_2O-like clusters of different sizes and stoichiometries [Figs. 2(c) and 2(d)]. That procedure is described in greater detail below. The simulated XANES spectra reveal that particles with larger sizes have more pronounced features compared to the smaller particles, as expected from their difference in the surface to volume ratios. As shown in Fig. 2(e), experimental XANES spectra measured in the Cu oxide clusters show a similar trend to have sharper features for larger sizes as obtained for simulated clusters of the same motif (CuO).

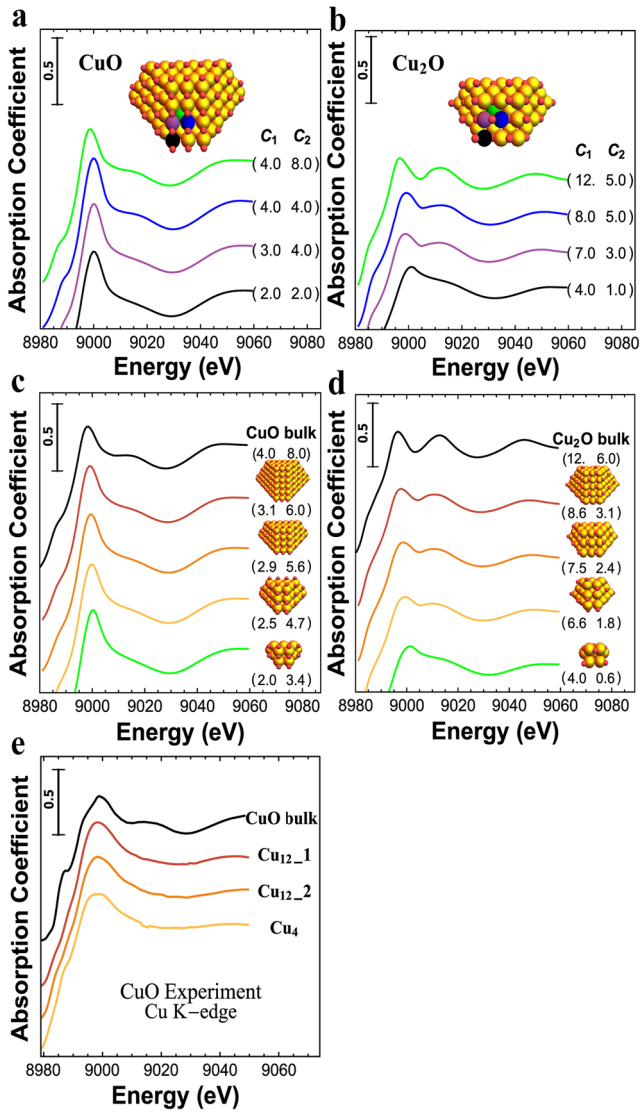


FIG. 2. Absorption site, cluster size, and motif effects on Cu K-edge XANES spectra. Each spectrum in Figs. 2(a)–2(d) is correlated with Cu–Cu CNs on the first and second coordination shells. [(a) and (b)] Site-specific XANES (XANES for the specific atom) spectra for CuO and Cu_2O, calculated with FEFF9. [(c) and (d)] Particle-averaged XANES (averaged over all atoms in the particle) spectra for CuO and Cu_2O, calculated with FEFF9. (e) Experimental XANES of CuO bulk and CuO clusters containing 4 and 12 Cu atoms.

Similarly to our previous work,[36] to build the training set for NN, we first constructed several sets of Cartesian coordinates for atoms residing in the sites that correspond to the crystal structure of bulk CuO and Cu_2O, and truncated the lists of coordinates to simulate clusters with various shapes (tetrahedral, octahedral, and cubic) and sizes. This was accomplished by creating the cluster surfaces using (100) and (111) planes of bulk CuO or Cu_2O. The details of lattice structure information for CuO and Cu_2O are listed in Table S1 of the supplementary material. To generate more models, additional CuO and Cu_2O models were constructed by truncating the previous regular models with (100) and (111) planes. Furthermore, we also constructed the planar structures with one or two layers of (111) plane of CuO and Cu_2O to describe the active structural motif, thin film, which has been reported as an active phase for catalysis.[51–53] As a result, we created 25 CuO models and 30 Cu_2O models to capture the diversity of Cu_xO nanostructures, which are relevant to catalysis.

In the nanometer-scale nanoparticles and subnanometer clusters, the interatomic distances can deviate from those in their respective bulk compounds[42,54] due to the effects of size, adsorbates, and support. For example, the nearest Cu–Cu distance for the bulk of CuO is 2.93 Å. However, the Cu–Cu distance of the CuO cluster was reported to be longer[54] or shorter[42] compared to the bulk. The shortening of the Cu–Cu distance in size-selected reduced Cu clusters was reported by us earlier.[39] To allow for this effect to be recognized in the process of NN-based analysis, we isotropically stretched or compressed the structures in our theoretical models to generate more training sets. The distance between nearest copper atoms varied from 2.784 Å to 3.077 Å for CuO models and from 2.879 Å and 3.182 Å in Cu_2O models. These ranges bracket the reported Cu–Cu distances for copper oxides available from EXAFS analyses or crystallography data.[42,54,55] To represent the size and shape of the clusters, we choose the first few Cu–Cu coordination numbers as structural descriptors for each unique model. We preferred to rely on the Cu–Cu CNs rather than on Cu–O CNs for this purpose because the latter parameter is not a good descriptor of the cluster size, geometry, and oxidation state in those cases when all Cu atoms are

terminated by oxygens. The copper atoms on the surface have smaller Cu–Cu CNs compared to the inner copper atoms, thus providing the desired sensitivity to the size and shape of the copper oxide clusters. We illustrate the sensitivity of Cu–Cu CNs to the different size and shape of the copper oxide models in Figs. S1 and S2. For different copper oxide models (e.g., Cu_2O vs CuO), the Cu–Cu CNs exhibit unique values. With the increase of the size of the models, the Cu–Cu CNs also get larger. For the models with same Cu–Cu CNs of the first shell, the values of the Cu–Cu CNs of the second shell provides additional sensitivity to the task of classification of different models.

In order to construct a required (large) number of spectra in the training set using FEFF9, we adopted a combinatorial approach, first developed in Ref. 36, relying on randomly mixing several site-specific XANES calculations for CuO and Cu_2O models prepared above. Each spectrum was labeled with first and second Cu–Cu CNs as structural descriptors. The total size of the training set was 100 000 spectra for each of the CuO- and Cu_2O-type models. In order to compensate for the unknown X-ray energy shift between theoretical and experimental XANES spectra for each type of oxide clusters (CuO- or Cu_2O-like), we shifted all the theoretical XANES spectra by ΔE (obtained from the difference in energy between experimental and theoretical XANES spectra for the respective bulk oxides). Such an approach is reasonable because no visible shift was observed in the XANES spectra between different experimental copper oxide clusters. Furthermore, the convolution neural network we used for machine learning has the advantage of shift invariance,[56] which means that the results will not depend strongly on the possible, small (shown to be within a ±1 eV range, as tested in this work) mismatch in the X-ray energy origins used in theory and experiment. An alternative approach, relying on random energy shift between different spectra from the training set, was also recently proposed.[57] After the shift was applied, the spectra were interpolated to the same energy scale from E_{min} = 8981.5 eV to E_{max} = 9059.3 eV. The step size for the energy scale is 0.15 eV near E_{min} and increases to 1.5 eV near E_{max}. Following this step, all spectra were represented as multidimensional vectors, containing 94 data points. Each data point corresponded to the value of absorption coefficient at specific energy.

The NN used in this work was a nonlinear function $f(\mu, \theta) = \{C_1, C_2\}$, where μ represents the preprocessed XANES spectrum (a vector with 94 points) as input and $\{C_1, C_2\}$ represents the first two CNs as output. The parameter space θ consists of the weights and biases in the NN models.[58] The purpose of the training process is to optimize the parameter space θ to accurately correlate input with output. Once the optimal parameters are found, the training process is finished. More details of NN construction and training are described in the supplementary material.

The accuracy of our NN was demonstrated by the theoretical XANES calculated by FEFF9 for particles with different sizes and shapes. Unlike the data set we used for the training, the spectra for validation are particle-averaged spectra (averaged XANES for the particle) corresponding to the real copper oxide models and not used in the NN training process. In Fig. 3, we compare the true Cu–Cu CNs on the first coordination shell with the predicted Cu–Cu CNs for CuO and Cu_2O models. The validation for the Cu–Cu CNs on the second coordination shell is given in Fig. S3 of the supplementary material. According to the comparison result, NN can predict accurate CNs from the theoretical XANES for a large range of particle sizes.

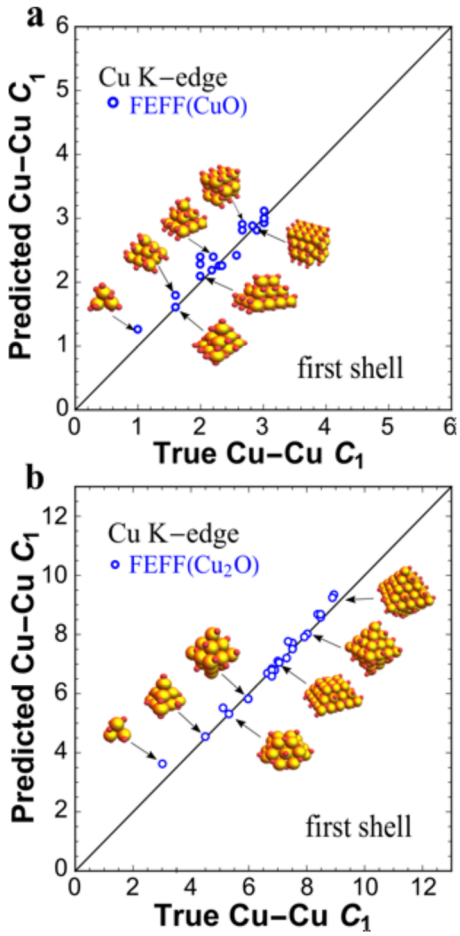


FIG. 3. Validation of CuO (a) and Cu_2O (b) neural networks using theoretical XANES. True Cu–Cu CNs are compared with predicted Cu–Cu CNs of the first coordination shell.

APPLICATION TO EXPERIMENTAL DATA

After the validation of our NNs, we applied them to the unknown structures of supported ultra-small size-selected clusters used in a recent work[59] These copper-based clusters can be used, for example, as catalysts for conversion of CO_2 with hydrogen. The data discussed in this paper were extracted from *in situ* grazing incidence XANES (GI XANES) spectra collected on samples of 4-, 12-, or 20-atom Cu clusters deposited on zirconia support prepared by atomic layer deposition and by supersonic cluster beam deposition, as exposed to CO_2 and H_2 under elevated temperatures reaching 375 °C. More experimental details are given in Note III of the supplementary material.

The data extracted from the *in situ* XANES data were collected and analyzed by multivariate curve resolution with the alternating least squares (MCR-ALS) method to obtain the mixing fraction

of clusters of different oxidation states (CuO, Cu_2O, and Cu).[59] Because our NN method is designed for an idealized, pure metal oxide phase (either CuO- or Cu_2O-like), its test required access to the corresponding phase-pure clusters, which were not found in the series of spectra obtained in our *operando* experiments. We thus selected those spectra collected and analyzed in Ref. 59, which had the highest fractions of CuO or Cu_2O phases based on the results of the MCR-ALS analysis. The fractions of individual copper components (CuO, Cu_2O, and Cu) for the spectra are listed in Tables S2 and S3. The sampling chosen for testing our NN prediction corresponds to those temperatures, for which the XANES data, as analyzed by MCR-ALS, indicated the presence of at least 70% of either CuO or Cu_2O phase. As a justification of validity of this approach, we note that their XANES spectra have similar features with either CuO or Cu_2O bulk XANES spectra, thus validating their designation as tests for NN validation purpose.

In Table S2, combining the spectra found (by MCR-ALS) to correspond to the CuO-like clusters, we show the application results of the CuO-trained NN model for extracting the first and second Cu–Cu CNs from those XANES spectra. To interpret the results, the correlation between the number of Cu atoms and the first Cu–Cu CNs for the CuO models is shown in Fig. 4(a). All models shown there were selected from the NN training and validation steps. Such a correlation demonstrates that our method can be used for measuring the cluster size, as evident here from the correct detection of the number of atoms (which was known *a priori* from the cluster deposition experiment[60,61]). To check the capability of our NN to distinguish between CuO and Cu_2O motifs, we applied our trained Cu_2O NN model to the XANES spectra and showed the predicted CNs in Table S2. Not surprisingly, the predicted CNs from Cu_2O NN have larger error bars for the first nearest neighbors. It means that the result is unstable. By comparing the predicted CNs from two NN, we obtained that the oxidation state of the sample is consistent with that obtained by an independent chemometric approach, as reported in Table S2.

For the samples with a large fraction of Cu_2O, we performed similar NN-XANES analysis, this time by using our Cu_2O NN model. Table S3 lists samples that are mainly composed of Cu_2O. The first and second Cu–Cu CNs are extracted by our Cu_2O NN model. To validate the results, we present the correlation between the number of Cu atoms and the first Cu–Cu CNs for Cu_2O models in Fig. 4(b). The results demonstrate a correlation between predicted CNs from Cu_2O NN with the sizes of the cluster. Similar to the prior example, we used the CuO-trained NN to check the capability of our method to distinguish between the CuO and Cu_2O motifs. The predicted CNs from CuO NN give much larger error bars for the first nearest neighbors than those obtained from the Cu_2O NN for the same experimental spectra Table S3. Thus, by combining the predicted CNs from two NNs, and comparing them with the known CN values that correspond to the known cluster sizes, we demonstrated that the oxidation state and structural information can be extracted from the spectra for Cu_2O-like clusters (Table S3).

After validating the NNs using experimental spectra for clusters with the known sizes and structures (dominated by either Cu_2O or CuO clusters, as described above), we applied the NNs to analyze the spectra for samples of a nominal size of Cu_{20} and unknown structure and oxidation state. Our trained NNs were applied to answer the question whether the sample was mainly composed of a CuO-like or a Cu_2O-like phase. In Table S4, we present the first and second Cu–Cu CNs extracted by our CuO NN and Cu_2O NN. To analyze the results and determine the oxidation states, we correlate the predicted CNs with the sizes of the cluster and examine the relation between Cu–Cu CNs and the number of copper atoms in Fig. 4. The predicted CNs from CuO NN follow the size-dependent trend in Fig. 4(a). However, the predicted CNs from Cu_2O NN show large error bars when compared with the CuO NN prediction [Fig. 4(b)]. Based on the results for Cu_4 and Cu_{12} clusters, described above and summarized in Tables S2 and S3, larger error bars were always obtained when the incorrect NN models were applied. Therefore, by comparing the predicted CNs and taking into consideration the difference in the error bars (that were demonstrated to be an important factor in discriminating between two possible phases of the copper oxides) from CuO NN and Cu_2O NN, we conclude that the Cu clusters containing 20 atoms were dominated by the CuO phase.

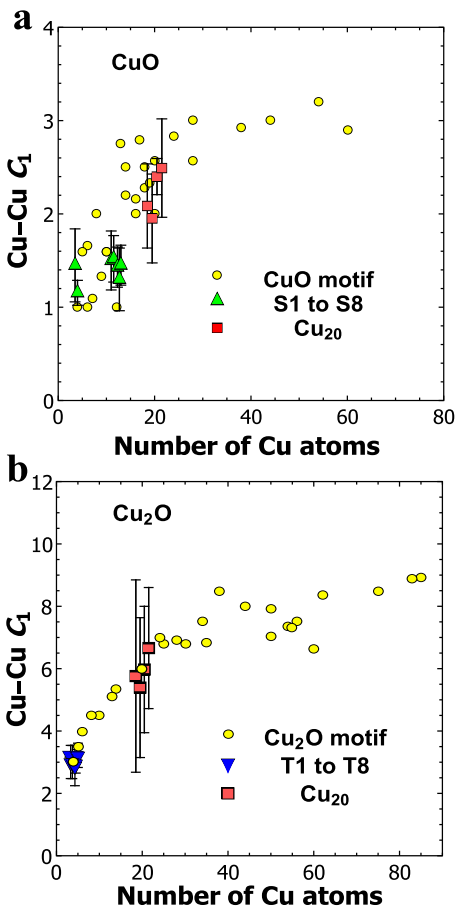


FIG. 4. The correlations between the number of Cu atoms and true Cu–Cu CNs for theoretical CuO (a) and Cu_2O (b) models built during the NN training. The blue, green, and red points are shifted horizontally around their actual values (4, 12, and 20) to show the error bars. S1–S7, T1–T7, and Cu_{20} represent the experimental samples with CNs extracted by the NNs. The detailed description of samples S1–S7, T1–T7, and Cu_{20} is given in Tables S2–S4.

CONCLUSIONS

In summary, a neural network method was utilized to build the relationship between the XANES spectra and structural parameters for copper oxide cluster systems. This method enabled the determination of the average particle size and the oxidation state of metal clusters during the catalytic reaction by "inverting" their XANES spectra. Since the metal clusters acted as important catalysts in many reactions, this method is poised to have many applications. For instance, for the carbon dioxide and nitrogen oxide related reduction reaction where the metal cluster acts as catalyst and gets oxidized,[62] the NN method can be soon utilized to analyze the structure of these metal oxide clusters and thus help decipher the reaction mechanism. At this stage, while our method is an improvement compared to the previously developed NN-based XANES analysis approach, because it is applied, for the first time, not to pure metallic clusters but to the metal oxides, the present NN method still has several important limitations. For example, it relies on the CNs as descriptors and thus cannot distinguish isomers with the same CNs. It is also favoring those speciations where one (CuO, Cu_2O, or Cu) phase dominates and would not be helpful when these phases coexist with similar fractions during a particular reaction step. We envision that such recently developed techniques for XANES analysis as MCR-ALS, linear combination, and principal component analysis will be used in combination with our approach for accurately obtaining both the mixing fractions of different types of clusters and structural characteristics of each type. Our method, after the required training and validation, is also applicable to a wide range of metal oxide cluster catalysts and for the understanding of structure, composition, and function relationships in catalysis.

SUPPLEMENTARY MATERIAL

See the supplementary material for additional details on the XANES calculation, neural network implementation and training, experiment, and results of speciation analysis of clusters used in this work.

ACKNOWLEDGMENTS

A.I.F. acknowledges support by the U.S. Department of Energy, Office of Basic Energy Sciences under Grant No. DE-FG02-03ER15476. A.I.F. acknowledges support by the Laboratory Directed Research and Development Program through Grant No. LDRD 18-047 of Brookhaven National Laboratory under U.S. Department of Energy Contract No. DE-SC0012704 for initiating his research in machine learning methods. The computational work used resources of the Center for Functional Nanomaterials, which is a U.S. DOE Office of Science Facility at Brookhaven National Laboratory under Contract No. DE-SC0012704. This work was supported by resources provided by the Scientific Data and Computing Center (SDCC), a component of the Computational Science Initiative (CSI) at Brookhaven National Laboratory (BNL). The work at the Argonne National Laboratory (A.H., B.Y., L.K., M.J.P., and S.V.) was supported by the U.S. Department of Energy, Office of Science, BES-Materials Science and Engineering, under Contract No. DE-AC-02-06CH11357. This research used resources of the Advanced Photon Source (beamline 12-ID-C—S.S., and 12-BM), a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. S.V. also acknowledges support from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 810310, which corresponds to the J. Heyrovsky Chair project ("ERA Chair at J. Heyrovský Institute of Physical Chemistry AS CR—The institutional approach towards ERA") during the finalization of this paper. The funders had no role in the preparation of the article. P.L. was supported by the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences, Division of Chemical Sciences, Biosciences and Geosciences, Chemical Sciences program, under Contract No. DE-SC0012704.

REFERENCES

[1] M. Zhang, M. de Respinis, and H. Frei, Nat. Chem. **6**, 362 (2014).
[2] N. N. Nassar, A. Hassan, and P. Pereira-Almao, Energy Fuels **25**, 1017 (2011).
[3] J. T. Grant, J. M. Venegas, W. P. McDermott, and I. Hermans, Chem. Rev. **118**, 2769 (2018).
[4] S. Lwin and I. E. Wachs, ACS Catal. **4**, 2505 (2014).
[5] I. Zaharieva, P. Chernev, M. Risch, K. Klingan, M. Kohlhoff, A. Fischer, and H. Dau, Energy Environ. Sci. **5**, 7081 (2012).
[6] S. R. Pendlebury, M. Barroso, A. J. Cowan, K. Sivula, J. Tang, M. Grätzel, D. Klug, and J. R. Durrant, Chem. Commun. **47**, 716 (2011).
[7] J. Rodriguez, S. Ma, P. Liu, J. Hrbek, J. Evans, and M. Perez, Science **318**, 1757 (2007).
[8] G. Melaet, W. T. Ralston, C.-S. Li, S. Alayoglu, K. An, N. Musselwhite, B. Kalkan, and G. A. Somorjai, J. Am. Chem. Soc. **136**, 2260 (2014).
[9] C. Xie, C. Chen, Y. Yu, J. Su, Y. Li, G. A. Somorjai, and P. Yang, Nano Lett. **17**, 3798 (2017).
[10] M. Hermanek, R. Zboril, I. Medrik, J. Pechousek, and C. Gregor, J. Am. Chem. Soc. **129**, 10929 (2007).
[11] L.-P. Wang and T. Van Voorhis, J. Phys. Chem. Lett. **2**, 2200 (2011).
[12] M. Behrens, F. Studt, I. Kasatkin, S. Kühl, M. Hävecker, F. Abild-Pedersen, S. Zander, F. Girgsdies, P. Kurr, B.-L. Kniep, M. Tovar, R. W. Fischer, J. K. Nørskov, and R. Schlögl, Science **336**, 893 (2012).
[13] S. Kattel, P. J. Ramírez, J. G. Chen, J. A. Rodriguez, and P. Liu, Science **355**, 1296 (2017).
[14] T. Kibata, T. Mitsudome, T. Mizugaki, K. Jitsukawa, and K. Kaneda, Chem. Commun. **49**, 167 (2013).
[15] S. Vajda and M. G. White, ACS Catal. **5**, 7152 (2015).
[16] R. Si and M. Flytzanii-Stephanopoulos, Angew. Chem., Int. Ed. **47**, 2884 (2008).
[17] E. Jimenez-Izal, B. C. Gates, and A. N. Alexandrova, Phys. Today **72**(7), 38 (2019).
[18] C. Pan, D. Zhang, and L. Shi, J. Solid State Chem. **181**, 1298 (2008).
[19] D. Zhou, S. W. Bennett, and A. A. Keller, PLos One **7**, e37363 (2012).
[20] A. A. Keller, H. Wang, D. Zhou, H. S. Lenihan, G. Cherr, B. J. Cardinale, R. Miller, and Z. Ji, Environ. Sci. Technol. **44**, 1962 (2010).
[21] M. A. Bañares and I. E. Wachs, J. Raman Spectrosc. **33**, 359 (2002).
[22] M. Guerrero-Pérez and M. Banares, Chem. Commun. **2002**, 1292.
[23] B. M. Weckhuysen, Chem. Commun. **2002**, 97.
[24] S. T. Chill, R. M. Anderson, D. F. Yancey, A. I. Frenkel, R. M. Crooks, and G. Henkelman, ACS Nano **9**, 4036 (2015).
[25] J. Timoshenko and A. I. Frenkel, Catal. Today **280**, 274 (2017).
[26] A. Yevick and A. I. Frenkel, Phys. Rev. B **81**, 115451 (2010).
[27] J. Polte, T. T. Ahner, F. Delissen, S. Sokolov, F. Emmerling, A. F. Thünemann, and R. Kraehnert, J. Am. Chem. Soc. **132**, 1296 (2010).
[28] A. I. Frenkel, J. A. Rodriguez, and J. G. Chen, ACS Catal. **2**, 2269 (2012).
[29] A. Ankudinov, J. Rehr, J. J. Low, and S. R. Bare, J. Chem. Phys. **116**, 1911 (2002).
[30] Y. Dai, T. J. Gorey, S. L. Anderson, S. Lee, S. Lee, S. Seifert, and R. E. Winans, J. Phys. Chem. C **121**, 361 (2016).

[31]C. Lamberti, S. Bordiga, F. Bonino, C. Prestipino, G. Berlier, L. Capello, F. D'Acapito, F. X. Llabrés i Xamena, and A. Zecchina, Phys. Chem. Chem. Phys. **5**, 4502 (2003).
[32]A. Halder, M. Kilianová, B. Yang, E. C. Tyo, S. Seifert, R. Prucek, A. Panáček, P. Suchomel, O. Tomanec, D. J. Gosztola, D. Milde, H.-H. Wang, L. Kvítek, R. Zbořil, and S. Vajda, Appl. Catal., B **225**, 128 (2018).
[33]S. A. Wyrzgol, S. Schäfer, S. Lee, B. Lee, M. Di Vece, X. Li, S. Seifert, R. E. Winans, M. Stutzmann, and J. A. Lercher, Phys. Chem. Chem. Phys. **12**, 5585 (2010).
[34]S. Lee, S. Lee, D. Gerceker, M. Kumbhalkar, K. M. Wiaderek, M. Ball, M. Mavrikakis, J. Dumesic, and R. E. Winans, Phys. Chem. Chem. Phys. **21**, 11740 (2019).
[35]S. Roese, A. Kononov, J. Timoshenko, A. I. Frenkel, and H. Hövel, Langmuir **34**, 4811 (2018).
[36]J. Timoshenko, D. Lu, Y. Lin, and A. I. Frenkel, J. Chem. Phys. Lett. **8**, 5091 (2017).
[37]M. Ahmadi, J. Timoshenko, F. Behafarid, and B. Roldan Cuenya, J. Phys. Chem. C **123**, 10666 (2019).
[38]J. Timoshenko, S. Roese, H. Hövel, and A. I. Frenkel, "Silver clusters shape determination from in-situ XANES data," Radiat. Phys. Chem. (to be published).
[39]J. Timoshenko, A. Halder, B. Yang, S. Seifert, M. J. Pellin, S. Vajda, and A. I. Frenkel, J. Phys. Chem. C **122**, 21686 (2018).
[40]K. V. Chary, G. V. Sagar, C. S. Srikanth, and V. V. Rao, J. Phys. Chem. B **111**, 543 (2007).
[41]B. White, M. Yin, A. Hall, D. Le, S. Stolbov, T. Rahman, N. Turro, and S. O'Brien, Nano Lett. **6**, 2095 (2006).
[42]S. Grundner, M. A. C. Markovits, G. Li, M. Tromp, E. A. Pidko, E. J. M. Hensen, A. Jentys, M. Sanchez-Sanchez, and J. A. Lercher, Nat. Commun. **6**, 7546 (2015).
[43]P. J. Smeets, M. H. Groothaert, and R. A. Schoonheydt, Catal. Today **110**, 303 (2005).
[44]Z. Zhu, Z. Liu, S. Liu, H. Niu, T. Hu, T. Liu, and Y. Xie, Appl. Catal., B **26**, 25 (2000).
[45]P. W. Park and J. S. Ledford, Appl. Catal., B **15**, 221 (1998).
[46]G. Avgouropoulos, J. Papavasiliou, T. Tabakova, V. Idakiev, and T. Ioannides, Chem. Eng. J. **124**, 41 (2006).
[47]H. Rao, H. Fu, Y. Jiang, and Y. Zhao, Angew. Chem., Int. Ed. **48**, 1114 (2009).
[48]K. Manthiram, B. J. Beberwyck, and A. P. Alivisatos, J. Am. Chem. Soc. **136**, 13319 (2014).
[49]W. Wei and G. Jinlong, Front. Chem. Sci. Eng. **5**, 2 (2011).
[50]J. J. Rehr, J. J. Kas, F. D. Vila, M. P. Prange, and K. Jorissen, Phys. Chem. Chem. Phys. **12**, 5503 (2010).
[51]W. An, A. E. Baber, F. Xu, M. Soldemo, J. Weissenrieder, D. Stacchiola, and P. Liu, ChemCatChem **6**, 2364 (2014).
[52]F. Yang, Y. Choi, P. Liu, D. Stacchiola, J. Hrbek, and J. A. Rodriguez, J. Am. Chem. Soc. **133**, 11474 (2011).
[53]F. Yang, Y. Choi, S. Agnoli, P. Liu, D. Stacchiola, J. Hrbek, and J. A. Rodriguez, J. Phys. Chem. C **115**, 23062 (2011).
[54]M. Swadźba-Kwaśny, L. Chancelier, S. Ng, H. G. Manyar, C. Hardacre, and P. Nockemann, Dalton Trans. **41**, 219 (2012).
[55]K.-S. Lin and H. P. Wang, Environ. Sci. Technol. **34**, 4849 (2000).
[56]Y. LeCun and Y. Bengio, *The Handbook of Brain Theory and Neural Networks* (MIT Press, 1995), p. 3361.
[57]J. Timoshenko, M. Ahmadi, and B. Roldan Cuenya, J. Phys. Chem. C **123**, 20594 (2019).
[58]Y. LeCun, *Connectionism in Perspective* (Elsevier, 1989), Citeseer.
[59]A. Halder, C. Lenardi, B. Yang, J. Timoshenko, L. K. Kolipaka, M. J. Pellin, S. Seifert, A. I. Frenkel, P. Milani, and S. Vajda, "CO_2 methanation on Cu cluster decorated zirconia supports with different morphology: A combined *in situ* XANES and ex-situ XPS study" (unpublished).
[60]B. Yang, C. Liu, A. Halder, E. C. Tyo, A. B. F. Martinson, S. Seifert, P. Zapol, L. A. Curtiss, and S. Vajda, J. Phys. Chem. C **121**, 10406 (2017).
[61]C. Liu, B. Yang, E. Tyo, S. Seifert, J. DeBartolo, B. von Issendorff, P. Zapol, S. Vajda, and L. A. Curtiss, J. Am. Chem. Soc. **137**, 8676 (2015).
[62]O. P. Balaj, I. Balteanu, T. T. J. Roßteuscher, M. K. Beyer, and V. E. Bondybey, Angew. Chem., Int. Ed. **43**, 6519 (2004).

Supplementary Material for "Mapping XANES Spectra on structural descriptors of copper oxide clusters using supervised machine learning"

Yang Liu,[1,2] Nicholas Marcella,[2] Janis Timoshenko,[2] Avik Halder,[3] Bing Yang,[3] Lakshmi Kolipaka,[3] Michael. J. Pellin,[3] Soenke Seifert,[4] Stefan Vajda,[3, 5, 6] Ping Liu,[7] Anatoly I. Frenkel*[,2,7]

[1]Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, USA

[2]Department of Materials Science and Chemical engineering, Stony Brook University, Stony Brook, New York 11794, USA

[3]Materials Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA

[4]X-ray Sciences Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA

[5]Institute for Molecular Engineering, The University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, USA

[6]Department of Nanocatalysis, J. Heyrovský Institute of Physical Chemistry, Czech Academy of Sciences, Dolejškova 3, 18223 Prague 8, Czech Republic

[7]Division of Chemistry, Brookhaven National Laboratory, Upton, New York 11973, USA

Note I. DETAILS OF AB INITIO CALCULATION OF XANES

Similarly to our previous work [1-3], we use *ab initio* code FEFF [4] for XANES simulation. The non-structural parameters for XANES simulations were chosen to ensure the best agreement between the simulated spectrum for the bulk of CuO, Cu_2O and the corresponding experimental XANES data. FEFF version 9.6.4 was used for self-consistent calculation within full multiple scattering (FMS) and muffin-tin (MT) approximations. FMS cluster size was chosen at a large value so that the whole cluster was included in the FMS calculations. Random phase approximation (RPA) was used to model core-hole and use the default value (1.5 Å or 2.0 Å) respectively for Cu_2O or CuO MT radius, as well as complex exchange-correlation Hedin-Lundqvist potential.

Cluster models for XANES calculation were constructed by cutting the bulk of copper oxide with (100) and (111) planes. The details of the lattice parameters for the bulk of CuO and Cu_2O are listed in Table S1. Considering the sensitivity of XANES to the interatomic distance, we constructed additional structure models by stretching and compressing the models from the original set. The distance between the

nearest copper atoms varies from 2.784 and 3.07 Å for CuO. The distance between the nearest copper atom varies from 2.879 and 3.182 Å for Cu_2O. In addition, we constructed the planar structures with one or two layers of (111) planes of CuO and Cu_2O.

The non-equivalent sites in all cluster models were selected for XANES calculation. After that, we obtained the particle-averaged XANES by averaging the site-specific spectra. All theoretical XANES spectra were shifted in energy by ΔE to align the energy scale of theoretical calculations with experimental data. The value of ΔE for Cu_2O is calculated by the alignment of the theoretical XANES for the Cu_2O bulk with experimental spectra with energy scale ranging from E_{min} =8981.2 eV to E_{max} = 9059.3 eV. The value of ΔE for CuO is calculated by the alignment of the theoretical XANES for the CuO bulk with experimental spectra with energy scale ranging from E_{min} =8981.5 eV to E_{max} = 9059.3 eV.

Note II. DETAILS OF NEURAL NETWORK IMPLEMENTATION AND TRAINING

Similarly to our previous works [3], we use *Wolfram Mathematica 11.3.* to construct and train the neural network (NN). Our NN takes discretized XANES spectrum as input and output a vector that describes relevant structural parameters (metal-metal coordination numbers (CNs) for the first two shells (C_1, C_2)). The output layer of our NN contains two nodes and the input layer contains 94 points which are determined by the number of points in the discretized XANES spectrum. The number of nodes in the hidden layers are optimized to ensure optimal performance on validation set. The structure of Cu_2O NN is based on one-dimensional convolution NN which has one convolution layers and two hidden layers. The convolution layer has kernel sizes (64) with 128 channels. The hidden layers have 128 and 64 nodes. The structure of CuO NN is also based on one-dimensional convolution NN which has two convolution layers and one hidden layer. Two convolution layers have the same kernel sizes (16) with 64 channels. The hidden layer has 64 nodes. For each type of the NN, we also add the dropout layer with 0.2 probability after the convolution layer to overcome overfitting. For activation function, we use Rectified Linear Unit for convolution layers and hyperbolic tangent function for the hidden layer. To build the relation between the

XANES and structure parameters, we train NN on theoretical XANES data calculated with FEFF for copper oxide particles with different sizes, shapes and different interatomic distances.

To produce more XANES data, similarly to our previous work [3] for NN training purpose, we constructed artificial dataset by linear combining site-specific theoretical XANES for 25 particles (CuO) and 30 particles (Cu_2O) of different sizes and shapes. For each of site with known CNs $\{C_1, C_2\}$, we calculate the site-specific XANES spectrum by FEFF. Then we select randomly n sites and create corresponding average spectrum as $\mu^i(E) = \sum_{j=1}^{n} \mu_j(E)/n$ where $\mu_j(E)$ are site-specific spectra calculated either with FEFF for j-th of the randomly chosen sites. The corresponding average CNs can be obtained as $\{C_1, C_2\}^i = \sum_{j=1}^{n} \{C_1, C_2\}_j/n$. The selected sites can come from several different spectra but with same Cu-Cu distance R. We chose n=3 for the optimum NN performance.

For NN training, we use "ADAM" optimization algorithm with default parameters (β_1 = 0.9 and β_2 = 0.999). Batch size was 800 for CuO and 1500 for Cu_2O, and the training rounds for NN training depends on the training loss and validation loss. Loss function was defined as the L2-norm between output and target vectors averaged across the batch. Three neural networks were trained to overcome the bias and exhibit the stability of NN.

To validate our NN, we used a set of theoretical spectra that were not used for NN training. The theoretical spectra are particle-averaged XANES spectra obtained by averaging the site-specific XANES at each site of the copper oxide models, which corresponded to the CuO and Cu_2O models with different shapes and sizes. The models for the validation set were also used for the generation of the training set but only site-specific XANES spectra calculated on those models were used in the training set. Results of validation for the first Cu-Cu coordination shell are shown in Figure 3. Results of validation for the second Cu-Cu coordination shells are shown in Figure S3. The parameters of our NN models including the number of nodes, activation function, batch size and number of iterations for training are also optimized according to the performance on the validation set.

Note III. DETAILS OF GI XANES EXPERIMENT

Part 1 (describing the experimental details presented in Table S2 and Table S3)

The data presented in Table S2 and Table S3 were extracted from the data presented in Ref. [6]. In brief, the in-situ GI XANES were performed at beamline 12-ID-C of the Advanced Photon Source at the Argonne National Laboratory. The experimental setup has been reported previously [5]. There were two kinds of support used for the samples, ALD (atomic layer deposition) zirconia (creating zirconia films that were a few monolayers thick) and nano-zirconia (~100 nm thick zirconia film composed of zirconia nanoparticles), prepared by atomic layer deposition and supersonic cluster beam deposition, respectively. On these supports, Cu_4 and Cu_{12} clusters were deposited. The GI XANES experiments were performed under a gas mixture containing pure CO_2 and H_2 in 1:3 ratio, at the Cu K edge (8.987 keV) on a fluorescence detector (Vortex) positioned parallel to the sample surface and perpendicular to the incident beam. The spectra of the Cu_2O, CuO bulk standards were collected at the 12-BM beamline of the Advanced Photon Source in transmission mode as reference spectra.

Part 2 (describing the experimental details presented in Table S4)

Using the same experimental setup as described in Part 1 above, GI XANES data were collected on ALD-ZnO and ZrO_2-supported Cu_{20} clusters, in a gas atmosphere containing 20% CO_2, 60% H_2 and 20% He.

IV. SUPPORTING FIGURES AND TABLES

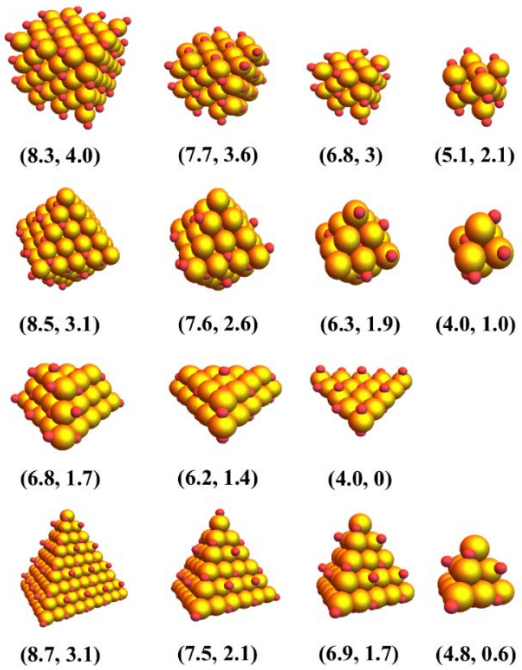


Figure S1. The representation of different Cu_2O models to show the sensitivity of Cu-Cu CNs to the size and shape of the models. Each model is labeled with first and second Cu-Cu CNs.

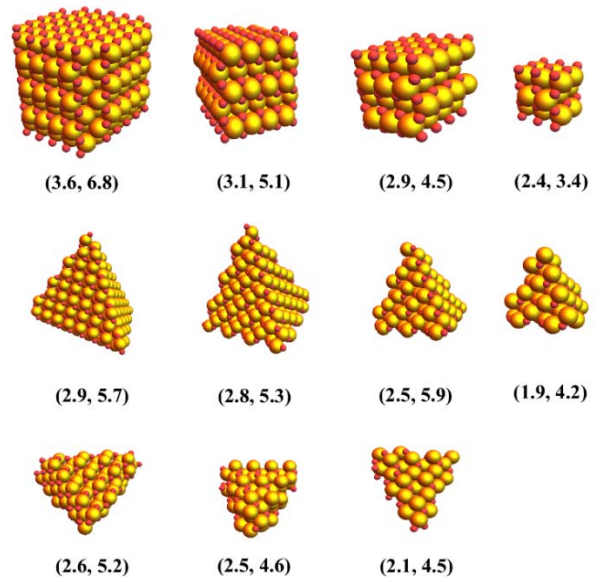


Figure S2. The representation of different CuO models to show the sensitivity of Cu-Cu CNs to the size and shape of the models. Each model is labeled with first and second Cu-Cu CNs.

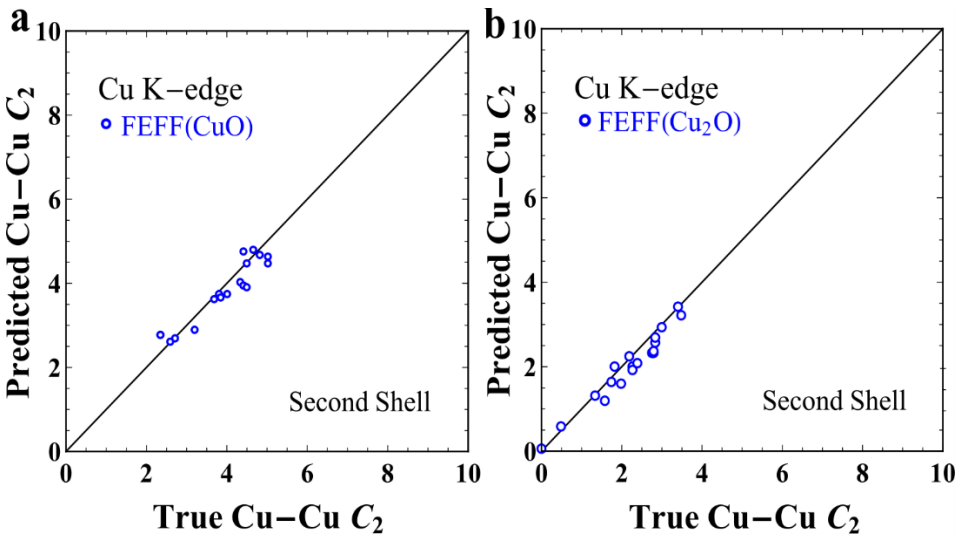


Figure S3. Validation of NN accuracy with experimental XANES and theoretical particle-averaged XANES. Results of NN analysis of XANES data, obtained in FEFF simulation for CuO and Cu_2O particles with different sizes, are compared with the true Cu-Cu coordination number.

Table S1. The lattice parameters for CuO and Cu_2O.

Lattice Parameters	CuO	Cu_2O
a	4.288 Å	2.933 Å
b	4.288 Å	2.933 Å
c	4.288 Å	5.133 Å
α	90°	90°
β	90°	90°
γ	90°	90°

Table S2. First and second shell of Cu-Cu CNs extracted from CuO XANES by CuO NN model. The predicted Cu-Cu CNs by Cu_2O NN model have much larger error bars for the first nearest neighbor shell. The fractions of individual copper components (CuO, Cu_2O and Cu) were obtained from the XANES data as discussed in greater detail in [6].

Name	Size	Support	CuO (%)	Cu_2O (%)	Cu (%)	Temperature (°C)	1st Cu-Cu CNs [a]	2nd Cu-Cu CNs [a]	1st Cu-Cu CNs [b]	2nd Cu-Cu CNs [b]
S1	Cu_{12}	ALD zirconia	74	0	26	25	1.5(3)	2.9(4)	4.5(12)	0.6(4)
S2	Cu_{12}	ALD zirconia	73	0	27	75	1.5(2)	3.3(4)	4.2(10)	0.9(1)
S3	Cu_4	Nano zirconia	70	18	12	225	1.4(4)	3.6(7)	4.7(16)	0.9(1)
S4	Cu_4	Nano zirconia	75	6	19	125	1.2(1)	3.3(4)	4.8(18)	1.0(1)
S5	Cu_{12}	Nano zirconia	73	0	27	225	1.4(2)	3.0(6)	4.8(26)	1.0(3)
S6	Cu_{12}	Nano zirconia	76	1	24	25	1.3(3)	3.3(4)	4.1(23)	2.1(4)
S7	Cu_{12}	Nano zirconia	74	0	26	75	1.5(2)	3.4(4)	3.8(19)	2.3(5)

[a] CNs predicted by CuO NN. [b] CNs predicted by Cu_2O NN. Uncertainties in the last significant digits are given in parentheses.

Table S3. First and second shell of Cu-Cu CNs extracted from Cu_2O XANES by Cu_2O NN model. The predicted Cu-Cu CNs by CuO NN model have much larger error bars for the first nearest neighbor shell. The fraction of individual copper components (CuO, Cu_2O and Cu) was obtained from the XANES data as discussed in greater detail in [6].

Name	Size	Support	CuO (%)	Cu_2O (%)	Cu (%)	Temperature (°C)	1st Cu-Cu CNs [a]	2nd Cu-Cu CNs [a]	1st Cu-Cu CNs [b]	2nd Cu-Cu CNs [b]
T1	Cu_4	ALD zirconia	19	66	19	225	3.2(1)	0.7(3)	1.9(14)	0.8(7)
T2	Cu_4	ALD zirconia	7	68	25	275	3.0(5)	0.6(4)	2.1(12)	0.5(2)
T3	Cu_4	ALD zirconia	8	68	24	325	3.2(2)	0.7(6)	2.0(12)	0.5(2)
T4	Cu_4	ALD zirconia	10	70	20	325	3.2(4)	0.7(6)	2.0(13)	0.5(2)
T5	Cu_4	ALD zirconia	10	70	20	275	3.0(5)	0.7(6)	2.0(12)	0.5(2)
T6	Cu_4	ALD zirconia	11	67	22	225	3.0(3)	0.6(6)	2.0(12)	0.5(2)
T7	Cu_4	ALD zirconia	17	61	22	175	2.9(6)	0.5(4)	2.0(10)	0.6(3)

[a] CNs predicted by Cu_2O NN. [b] CNs predicted by CuO NN. Uncertainties in the last significant digits are given in parentheses.

Table S4. Description of the Cu_{20} oxide cluster samples and their speciation by NN, showing that they are consistent with CuO structure (the results obtained using a Cu_2O - trained NN have much larger error bars).

Name	Size	Support	Reaction Condition	1st Cu-Cu CNs [a]	2nd Cu-Cu CNs [a]	1st Cu-Cu CNs [b]	2nd Cu-Cu CNs [b]
Cu_{20}_1	Cu_{20}	ZrO_2	He	2.1(4)	1.8(3)	6.7(19)	1.1(2)
Cu_{20}_2	Cu_{20}	ZrO_2	Mixture of 20% CO_2, 60% H_2 and 20% He	2.0(4)	2.3(2)	6.0(20)	1.1(2)
Cu_{20}_3	Cu_{20}	ZnO	He	2.4(2)	2.7(5)	5.4(22)	0.6(6)
Cu_{20}_4	Cu_{20}	ZnO	Mixture of 20% CO_2, 60% H_2 and 20% He	2.5(5)	1.9(3)	5.8(31)	0.7(5)

[a] CNs predicted by CuO NN. [b] CNs predicted by Cu_2O NN. Uncertainties in the last significant digits are given in parentheses.

REFERENCES:

[1] S. Roese, A. Kononov, J. Timoshenko, A. I. Frenkel, and H. Hövel, Langmuir **34**, 4811 (2018).
[2] J. Timoshenko, A. Halder, B. Yang, S. Seifert, M. J. Pellin, S. Vajda, and A. I. Frenkel, J. Phys. Chem. C **122**, 21686 (2018).
[3] J. Timoshenko, D. Lu, Y. Lin, and A. I. Frenkel, J. Phys. Chem. Lett. **8**, 5091 (2017).
[4] J. J. Rehr, J. J. Kas, F. D. Vila, M. P. Prange, and K. Jorissen, Phys. Chem. Chem. Phys. **12**, 5503 (2010).
[5] S. Lee, B. Lee, S. Seifert, S. Vajda, and R. E. Winans, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **649**, 200 (2011).
[6] A. Halder, C. Lenardi, B. Yang, J. Timoshenko, L. K. Kolipaka, M. J. Pellin, S. Seifert, A. I. Frenkel, P. Milani, and S. Vajda, (in preparation).