ACS Catalysis

Perspective

Cite This: ACS Catal. 2019, 9, 10192−10211

pubs.acs.org/acscatalysis

"Inverting" X-ray Absorption Spectra of Catalysts by Machine Learning in Search for Activity Descriptors

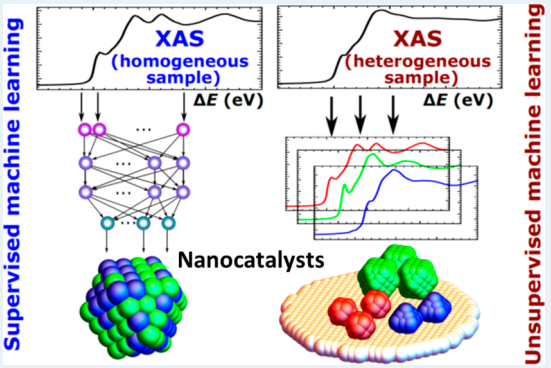
Janis Timoshenko*,† and Anatoly I. Frenkel*,‡,§

†Department of Interface Science, Fritz-Haber-Institute of the Max Planck Society, 14195 Berlin, Germany

‡Department of Materials Science and Chemical Engineering, Stony Brook University, Stony Brook, New York 11794, United States

§Chemistry Division, Brookhaven National Laboratory, Upton, New York 11973, United States

ABSTRACT: The rapid growth of methods emerging in the past decade for synthesis of "designer" catalysts—ranging from the size and shape-selected nanoparticles to mass-selected clusters, to precisely engineered bimetallic surfaces, to single site and pair site catalysts—has opened opportunities for tailoring the catalyst structure for the desired activity and selectivity. It has also sharpened the need for developing approaches to the operando characterization, ones that identify the catalytic active sites and follow their evolutions in reaction conditions. Commonly used methods for determination of the activity descriptors in the nanocatalysts, based on the correlation between the changes in catalyst performance and evolution of its structural and electronic properties, are hampered by the paucity of experimental techniques that can detect such properties with high accuracy and in reaction conditions. Out of many such techniques, X-ray absorption spectroscopy (XAS) stands out as an element-specific method that is very sensitive to the local geometric and electronic properties of the metal atoms and their surroundings and, therefore, is able to track catalyst structure modifications in operando conditions. Despite the vast amount of structure-specific information (such as, e.g., the charge states and radial distribution function of neighbors of selected atomic species) stored in the XAS data of catalysts, extracting it from the spectra is challenging, especially in the conditions of low metal weight loading, nanoscale dimensions, heterogeneous size and composition distributions, and harsh reaction environment. In this Perspective, we discuss the recent developments in XAS data analysis achieved by employing supervised and unsupervised machine learning (ML) methods for structural characterization of catalysts. By benefiting from the sensitivity of ML methods to subtle variations in experimental data, a previously "hidden" relationship between the X-ray absorption spectrum and descriptors of material's structure and/or composition can be found, as illustrated on representative examples of mono-, hetero-, and nonmetallic catalysts. In the case of supervised ML, the experimental spectra can be rapidly "inverted", and the structure of the catalyst can be tracked in real time and in reaction conditions. Emerging opportunities for catalysis research that the ML methods enable, such as high-throughput data analysis, and their applications to other experimental probes of catalyst structure are discussed.



KEYWORDS: *X-ray absorption spectroscopy, heterogeneous catalysis, machine learning, neural networks, principal component analysis, clustering, multivariate curve resolution*

1. INTRODUCTION

Understanding the mechanisms of activity, selectivity, and stability of catalytic materials is an extremely challenging task because of the complex, inherently heterogeneous structure of most catalysts, and their dynamic nature. It has become evident in recent years, given the development of advanced characterization tools and in situ methods, that the structure and properties of the catalysts may change under reaction conditions.[1−3] In the process of restructuring the bonding network, the charge states of active sites and their interaction with reactants, intermediates, and products may change dynamically, directly affecting their catalytic activity, selectivity, and/or stability. Hence, these properties of the catalysts—geometric, dynamic, and electronic—have to be probed experimentally and modeled theoretically at similar, atomistic levels of detail, in order to capture these transformations. For practical reasons, in order to correlate experimental measurements with theoretical simulations, a good approach is to rely on generalized descriptors of the structure, dynamics and electronic properties. Examples of such descriptors include the coordination number (CN) of nearest neighbors in a metal nanoparticle (NP), interatomic distance, the number of (111) facets, the number of missing bonds, surface strain, d-band center, and many others.[4−9] A well-known illustration of this approach includes the volcano plots by Nørskov et al., which link simple descriptors of catalyst structure to activity of its

Received: August 22, 2019
Revised: September 27, 2019
Published: September 27, 2019

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211

interactions with adsorbates.[10] There are also other reports listing several types of descriptors, specific to different catalysts, supports, and reactions.[7,9,11,12] It is often impossible to deduce the values of the descriptors from the knowledge of synthetic parameters and prereaction characterization of the catalyst. One possibility to solve this problem is by using an operando approach, in which not only the structure and property of the catalysts but also their activities are measured in situ, all in the same experiment.[13−17]

A notable challenge with these approaches is the difficulty to experimentally validate the predicted descriptors in working catalysts. There are two main reasons for that. First, for the success of this approach, the ability to synthesize the desired structural motifs on demand is required. The second reason is that in many catalytic reactions with novel catalysts (NPs, clusters, single site, and atomically dispersed catalysts), the detailed determination of the descriptors is complicated because of the paucity of most experimental techniques to reliably characterize these attributes in subnanometer species that may coexist in different structural forms and oxidation states and to follow their transformation in reaction conditions. While machine learning (ML) methods may play quite different roles at different stages of this process (one recent solution is the increasing reliance on the ML methods in materials and chemistry sciences for the task of discovering patterns in complex data sets[18−20]), in this Perspective we summarize their use for the extraction of key structural descriptors of catalysts from the X-ray absorption spectroscopy data, and we discuss the current and future applications of these methods. This topic is a culmination of the efforts started by the authors and their collaborators in 2017, when they demonstrated for the first time that X-ray absorption fine structure (XAFS) data can be decoded by an artificial neural network (NN) and "mapped" onto the nearest neighbor coordination numbers. As a result, important details of the structure of a catalytic material, such as the size and shape of metal NPs, were obtained.[21] In the subsequent works, we demonstrated that the extended X-ray absorption fine structure (EXAFS) can also be analyzed by the NN approach and that atomic pair distribution function can be directly extracted from the data.

The focus of this Perspective is on the applications of ML methods to the task of extracting structural properties of catalysts from their X-ray absorption near-edge structure (XANES) and EXAFS data. The main overarching idea in these approaches is the assumption that there is a unique relationship between the structural, electronic, and compositional descriptors of a given type of a catalyst and its X-ray absorption spectrum. This assumption is in the core of any XAFS analysis and modeling methods, except that conventional approaches rely either (i) on the forward modeling (e.g., in the case of the XANES studies, where one assumes a certain structure, then calculates the corresponding theoretical spectrum) and are, thus, model-dependent or (ii) on the fitting analysis (e.g., in the case of EXAFS studies), which works well for systems with relatively high degree of structural ordering. Neither forward modeling nor fitting methods are reliable for nanoscale catalysts, which, in active states, may have unique structures and/or are too disordered for fitting methods to produce adequate results. In this Perspective, we will show the possibilities of using ML methods and, in particular, the NN approach for solving the inverse problem: given the spectrum, to obtain structural information in a working catalyst. Because of the similarity between the origins of the fine structure in X-ray absorption coefficient and the electron energy loss spectrum (EELS) in the electron microscopy experiment, the discussed ML methods will be applicable for EELS data analysis as well. Specifically, the fine structure data in the energy loss near-edge structure (ELNES) or extended electron energy loss fine structure (EXELFS) spectra can be mapped to the structural and electronic descriptors of catalysts because ELNES and EXELFS contain similar information to XANES and EXAFS spectra, respectively.[22,23]

As an important caveat, the quantitative capability of both XANES and EXAFS, which are inherently ensemble-averaging methods, to solve the local structure around particular type of atoms is hampered by the heterogeneity of atomic distribution in the material. A typical situation is a material in which different species containing the same atom type as X-ray absorbing atoms can coexist, for example, in a reactor compartment containing regions with reduced and oxidized states of the catalyst[24] or in solutions where leaching of metal ions from metal nanoparticles occurs during reaction.[25] While the challenges in the homogeneous and heterogeneous samples are unique and different, both types can be studied by ML methods that can be separated into two classes: unsupervised and supervised ML. The goal of **supervised ML** (SML) is to establish relationship between inputs (spectra in our case) and outputs (descriptors of structure) based on a set of labeled training data (i.e., spectra, for which the correct answer (structure) is known). The goal of **unsupervised ML** is to discover patterns in large sets of experimental data without any labels. Unsupervised ML includes also such methods as principal component analysis (PCA), multivariate curve resolution, and clustering. While the NN approach for XANES and EXAFS data analysis is an example of supervised ML, the task of resolving heterogeneous species is best addressed by methods of unsupervised ML that we also summarize in this Perspective.

As we discuss in the Future Opportunities section, rapid data analysis of experimental data by ML methods enables "on the fly" analysis and, in a more distant future, "reaction on demand" approaches. We note also that, despite all the power and versatility of XAFS method, the complete and unambiguous model of a working catalyst can be constructed only by complementing it with other experimental probes: electron microscopy, IR spectroscopy, mass spectrometry, NMR, and so on. Conceptually, the analysis of the data from these methods can benefit from ML in the same way as it does for XAFS data. In this Perspective, we briefly discuss the recent applications of ML methods for these experimental approaches as well.

2. IF YOU WANT TO UNDERSTAND STRUCTURE, STUDY SPECTRUM

"If you want to understand the function, study structure", wrote Francis Crick. As we emphasized in the Introduction, unlocking the details of the catalyst structure (especially of that in small (<3 nm) NPs) from experimental data is extremely challenging. Crystallographic approaches cannot be used to determine the particle size, shape, and atomic structure of such small NPs that can be very disordered, and the sensitivity of high-resolution electron microscopy methods is not sufficient to resolve in detail the atomic structure.[26,27] XAFS is among a few experimental methods that are useful in this case. The

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211

unique role of XAFS spectroscopy for catalyst studies is based on its element-specificity, sensitivity to the details of local environment around catalyst (or adsorbate) atoms, applicability to broad range of samples, and applicability to in situ and operando studies. An important advantage of the XAFS method is its ability to address very different length-scales that all are relevant for catalysis: from subtle variations in interatomic distances on picometer scale due to interactions with adsorbates and surface-induced stress, to changes in the morphology of materials on nanometer scale (e.g., changes in NPs sizes and shapes), to macroscopic changes in the oxidation state of materials, crystallographic structure, and, finally, distributions of species in chemical reactors. Similarly, very different time-scales are accessible in XAFS experiments: while conventional XAFS measurements take minutes and are thus suitable for monitoring relatively slow changes in catalyst structure (e.g., oxidation/reduction processes) that take place within minutes and hours, other approaches enable investigation of processes at much shorter time scales. For example, millisecond regimes can be nowadays explored with quick-EXAFS and optically dispersed XAFS approaches, while pump-probe approaches provide access to picosecond time scales and allow one to probe, for example, structural transformations in photoexcited photocatalysts.[28] To employ this power of the XAFS method to the full extent, when it is applied to such intrinsically complex systems as catalysts under working conditions, advanced approaches to data analysis are required. Our proposed solution is, therefore, *in order to understand structure, study spectrum*. In the next section, we will describe in more detail the challenges in XANES and EXAFS data analysis and modeling by conventional methods.

3. CHALLENGES IN STRUCTURAL ANALYSIS BY X-RAY ABSORPTION SPECTROSCOPY

XANES and EXAFS constitute two regions of X-ray absorption spectra that encode complementary information. XANES portion of the spectrum corresponds to lower excitation energies and is defined by electronic transitions to unoccupied atomic and hybridized states and contains information, therefore, about the details of spatial arrangements of atoms.[27,29] EXAFS, in turn, is less sensitive to the details of electronic structure (due to higher energy of excited electrons) but contains more detail about the distributions of nearest neighbors around the absorbing atom. For decades, analysis of EXAFS features has been an indispensable tool for quantitative catalyst structure studies. Less attention has been paid to the structural information encoded in XANES, which has unique advantages. While EXAFS data quality can dramatically decrease in reaction conditions because of, for example, high temperature, low metal loading, restrictive reactor environment, and so on,[30−35] XANES remains relatively insensitive to disorder effects and, overall, has better signal-to-noise ratio than EXAFS. It can also be more sensitive to the 3D structure of the material, whereas EXAFS probes mostly radial distribution of atoms around the absorber. Furthermore, in the studies of nanostructured materials, in which the structural disorder is much more pronounced than in their bulk counterparts, XANES could be more sensitive to the structural details than EXAFS. Good-quality XANES data can be collected for samples in harsher experimental conditions, intrinsically disordered materials, samples with a low concentration of absorbing atoms, samples in complex, attenuating sample environments, as well as for samples on attenuating supports and/or crystalline supports that yield strong Bragg reflections that limit the length of usable spectra. Acquisition of XANES spectra can be done faster than acquisition of full EXAFS spectra; therefore, XANES-based studies are attractive for time-dependent investigations as well. XANES (or analogous ELNES) data can also be collected more easily in lab-based setups. In the soft X-ray regime (useful for studies of L and M edges of early transition metals, K-edges of nonmetals like C, N, O, S, and so on), typically only XANES spectra are available. All these factors make XANES analysis an attractive tool for studies of catalysts and other functional materials. Indeed, the sensitivity of XANES spectra to catalyst structure, (e.g., the NPs sizes) have been acknowledged a long time ago, both for metallic[27,36−38] and nonmetallic NPs.[39,40] The main challenge that hinders the usage of XANES for quantitative analysis of nanostructures is the lack of a general methodology that would allow one to extract 3D structural information from experimental data. As discussed in the Introduction, it is particularly desirable that this information is expressed in terms of useful "descriptors" of structure that can be modeled theoretically in order to understand reaction mechanisms and guide the design of novel catalysts. In addition, with the development of instrumentation for in situ, time-resolved, and high-throughput XAS studies,[41,42] there is a growing need for a method that would allow on-the-fly extraction of relevant structural information from hundreds and thousands of acquired spectra in a systematic way. Such methods, beyond conventionally used workhorse XANES analysis methods such as linear combination analysis (LCA) and principal component analysis (PCA), that are useful for chemical speciations of mixtures, but not for structural refinement of unknown structures, are presently lacking.

EXAFS analysis can provide detailed information about the distribution of bond lengths that cannot be extracted from XANES data. While methods for quantitative analysis of EXAFS data are more established than those for XANES data interpretation, ML-based approaches can address several issues that are challenging for conventional methods. Note that EXAFS can be directly linked to the partial radial distribution functions (RDFs) $g_{AB}(R)$, which describe the probability density for finding atom of type B, where symbol B designates any type of a nearest neighbor, including, e.g., A, at distance R from the absorber A. The contribution $\chi_{AB}(k)$ of atomic pair A−B to the total spectrum for absorbing atom A can be expressed as[43,44]

$$\chi_{AB}(k) = S_{0,A}^2 \int_0^{+\infty} g_{AB}(R) F_{AB}(k, R) e^{-2R/\lambda_A(k)} \sin(2kR + \phi_{AB}(k, R)) \frac{dR}{kR^2} \quad (1)$$

Here

$$k = \sqrt{\frac{2m_e}{\hbar^2}(E - E_{0,A})} \quad (2)$$

is photoelectron wavenumber, m_e is electron mass, $\hbar$ is Planck's constant, $E_{0,A}$ is photoelectron reference energy for absorber A, $S_{0,A}^2$ is corresponding amplitude reduction factor due to many-electronic excitations, F_{AB} and ϕ_{AB} are real and imaginary parts of photoelectron scattering functions, and $\lambda_A(k)$ is the effective mean free path for the photoelectron. The total spectrum for absorbing atom A is a sum of all single-

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211

scattering contributions $\chi_{AB}(k)$, but, importantly, also of so-called multiple-scattering (MS) contributions, which depend, e.g., on bonding angles and can be described by an equation, similar to eq 1, but where, instead of RDF $g_{AB}(R)$, a many-atomic distribution function needs to be used. From eq 1, knowing the partial RDF $g_{AB}(R)$, it is straightforward to calculate the corresponding $\chi_{AB}(k)$ contribution to the total EXAFS signal. The inverse problem (determination of $g_{AB}(R)$ from experimental EXAFS) is, however, much more challenging (ill-defined). Conventional EXAFS fitting based on EXAFS equation and nonlinear least-squares procedure[45] requires making an a priori assumption about the shape of bond-length distributions. Typically Gaussian or quasi-Gaussian shape is assumed, although more complex expressions based on asymmetric Γ-function are also used.[32,35,46−48] These approximations work well for well-defined bulk materials at low temperatures. However, for very small NPs or systems at high temperatures (both cases are commonly encountered in heterogeneous catalysis studies), these approximations are inadequate because of the enhanced static and/or thermal disorder, and the analysis results in significant systematic errors in the obtained values of CNs and interatomic distances.[31,32,34,46,49−54] Such asymmetry can be caused by the heterogeneous environments of atoms of the same type that are located at the catalyst surface, in the interior, and in the interface with the support.[33] For example, the atoms near the surface or support interface may be located in the strained environment, while the atoms near the core are in a relaxed environment closer to their bulk sites.[55,56] If only the average interatomic distance is obtained—typical for conventional EXAFS analyses—the ability to detect and analyze the different populations is lost. Another challenge is the analysis of contributions from distant coordination shells: while the more distant neighbors can contribute significantly to the EXAFS spectra of crystalline and nanocrystalline materials and can provide important clues about the 3D geometry of the material,[57,58] interpretation of these contributions in conventional EXAFS fitting approaches is nontrivial. Contributions of the distant neighbors overlap strongly and overlap also with the contributions of MS effects. Corresponding structural parameters thus cannot be obtained without additional constraints and assumptions because of the strong correlation between them.[46,59] Supervised ML-based approaches provide an elegant way to address both these challenges: instead of defining constraints and additional assumptions explicitly, the ML method deduces them automatically from the analysis of the training data set.[60]

As we have briefly outlined in the Introduction, in the case of structural and compositional heterogeneities, structural analysis of multiple coexisting species by ensemble-averaging XANES and EXAFS methods is very challenging. More common is the use of these methods for speciation purposes, using a number of linear algebra-based tools. The most widely used approach is LCA,[61] but more advanced approaches as PCA,[62−65] Blind Signal Separation (BSS)[66−68], and Multivariate Curve Resolution with Alternating Least Squares (MCR-ALS) fitting[69−71] are getting increasingly popular. These latter methods rely on the availability of large sets of experimental data and can be considered as examples of unsupervised ML.

In the next sections of this Perspective, we will focus on the analysis of homogeneous catalytic materials, with narrow particle size and compositional distributions. We start with discussing automated data matching and database search approaches for interpretation of XAFS data and then focus on SML methods (classification and regression), which enable determination of materials properties (discrete or continuous, respectively) even without a perfect agreement between the measured spectrum and one of the entries in the database. For construction of the XAFS spectra database, one relies on the forward modeling of XANES (and/or EXAFS) spectra of individual catalytic species (e.g., NPs or clusters). Methods of unsupervised ML will be then summarized in Section 6. Future opportunities in catalysis research based on the new spectral "inversion" methods will be presented in Section 7.

4. AUTOMATED SPECTRA MATCHING AND SUPERVISED CLASSIFICATION

From the early stages of XAFS method development, the common approach for the interpretation of experimental spectra to determine the electronic and atomistic structure of the material relied on the comparison of obtained XAFS data with those for reference materials. The limitations of this approach are clear. First, spectra for many reference materials need to be collected for the identification of a single spectrum of interest. Second, the result of "visual" comparison between two different spectra may be subjective and prone to bias, especially if the changes in local environment have only a subtle influence on XAFS features. While far from being fully addressed, some progress in dealing with these issues was noticeable during the last years. Part of the solution is the development of publicly available databases, containing experimental and theoretical XAFS data for many relevant reference materials. A necessary second part of the solution relies on the advances in algorithms that enable efficient search of the large databases and automated data matching.

The existing publicly available databases of experimental XAFS and analogous EELS spectra (http://cars.uchicago.edu/xaslib,[72] www.cat.hokudai.ac.jp/catdb,[73] https://eelsdb.eu/[74]), while useful, contain currently just a few hundreds of experimental spectra.[75] This information can be complemented by theoretical XAFS simulations. Note that during the last decades there has been a significant progress in the ab initio methods for XAFS spectra modeling. Current models for EXAFS modeling, implemented, for example, in FEFF[76] and GNXAS[77] codes, provide excellent agreement with experiment. Simulations of EELS and XANES data are more challenging but can be addressed within different approaches, for example, multiple-scattering,[78,79] Bethe-Salpeter equation method,[80,81] time-dependent density-functional theory,[82,83] and others.[84−87] Different levels of theory for XANES data calculations are compared in the recent review.[88]

For the task of generation of large databases, a good starting point is to use the computationally least demanding approaches, based on multiple-scattering approximation. Such an approach is implemented in FEFF[79] code (that carries out XANES simulation in so-called muffin-tin approximation to describe the electron density distribution in the material) and FDMNES[89] code (that can work with muffin-tin, as well as more accurate finite-difference methods). Within these approaches, simulations of a single XANES spectrum takes typically several CPU minutes, giving the possibility to generate databases with thousands of spectra.

Currently, the most impressive accomplishment in this area is the work by Zheng, Mathew et al.,[75,90] where FEFF code[79] was used to generate a database with ca. 800 000 K-edge

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211

XANES spectra for material structures from the Materials Project database.[91−93] Based on this collection, a tool was developed for automatic identification of XANES data. The key here is to define the distance or similarity function that would allow one to assess the difference between two spectra $\mu_1(E)$ and $\mu_2(E)$. Different approaches can be used. The simplest approach is to use the Euclidean distance between discretized XAFS spectra. It is, however, sensitive to the outliers, experimental noise, and systematic errors (e.g., data collected at different beamlines can have different resolution and can be difficult to compare; the details of background subtraction and spectra normalization may affect the results; and alignment of collected spectra in energy space may be affected by X-ray beam instabilities, etc.). Attempts to develop more robust distance functions have been made. Some of the alternative measures of similarity, discussed in the literature, include L_1-norm (Manhattan distance),[87,94] Pearson's correlation coefficient,[87,90,94] cosine distance,[87,90,94] and Spearman correlation test.[94] For matching experimental data with theoretically calculated spectra, the choice of similarity function may strongly affect the result, because due to the systematic errors in XANES data modeling, a perfect match between the input spectra and most appropriate spectrum from the database is not expected. In the algorithm, developed by Zheng, Mathew et al., this problem is addressed by using ensemble learning—different combinations of similarity functions and data preprocessing methods are used simultaneously, and the final result (the structures that match well the given experimental spectrum) is determined on the basis of the combined ranking. An example of the application of this method is shown in Figure 1, where experimental Ni K-edge and Co K-edge XANES spectra for NiO and $LiCoO_2$ materials are shown, together with the 5 most similar entries from the database of generated theoretical spectra. In both cases, the correct structure is among the 5 most likely answers.[75,90]

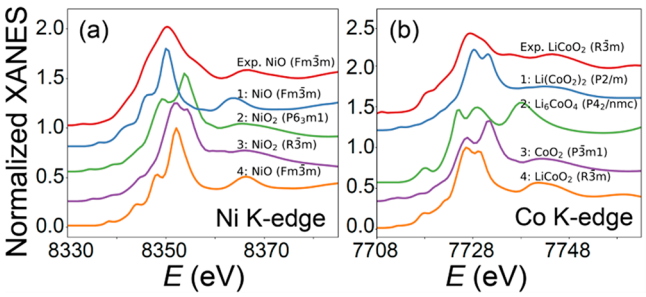


Figure 1. Results of automated spectra matching for (a) Ni K-edge XANES in NiO and (b) Co K-edge XANES in $LiCoO_2$. Experimental spectra are given together with the 5 most similar theoretical spectra for different bulk structures from the database. Theoretical spectra are shown in the order of decreasing similarity. Reproduced with permission from ref 90. Copyright 2018 Springer Nature. Reproduced under Creative Commons Attribution 4.0 International License: https://creativecommons.org/licenses/by/4.0/.

Although this database contains a large number of spectra, it is still limited to well-defined bulk materials; hence, its applicability to analysis of many catalytically relevant systems (heterogeneous catalysts, NPs, bimetallic alloys, etc.) is limited. Some useful information, can, nevertheless, be extracted. For example, even when the algorithm fails to find a correct structure for a given experimental spectrum, the listed possible structure models often have correct oxidation state and/or coordination environment for the absorbing species.[75,90] This paves the road for a more general analysis—instead of looking for an exact match with experimental spectrum, analysis of XAFS features can provide insight into some characteristics of studied material (oxidation state, number of nearest neighbors around absorbing atom and their arrangement, etc.). The interpretation of XANES spectra now can be formulated as a **supervised classification** problem: we attempt to construct a mathematical model $h(\Theta,\mu(E)) \rightarrow p$ that takes as input experimental spectrum $\mu(E)$ and returns a discrete number (label) that characterizes the corresponding property of the sample (e.g., oxidation state of absorbing species). The model depends on parameters Θ, which during the training step are optimized to ensure correct mapping between spectral features and materials properties. For this training step, one uses a set of spectra, for which the correct value of the corresponding property is known. Experimental spectra of well-defined reference spectra can be used for this purpose. For example, recently, Miyazato et al. used 23 K-edge experimental XANES spectra to train different ML routines (different functions $h(\Theta,\mu(E))$), such as support vector machine, random forest, decision trees, logistic regression, etc.) to recognize from the shape of absorption edge, whether the material is oxide or not.[95] However, only simple models can be trained using such small data sets. For more complex problems, much larger training sets are needed, and ab initio simulated spectra can be used for training.

In a recent work by Carbone et al.,[96] a database of 18184 site specific K-edge XANES spectra for transition metal oxides was constructed using structure models from Materials Project[91−93] and FEFF code[79] for XANES modeling. Using this data set, a NN was trained to recognize the local coordination environment around absorbing species (tetrahedral (4 nearest neighbors), square pyramidal (5 nearest neighbors), or octahedral (6 nearest neighbors) geometries). In the case of NNs, the function $h(\Theta,\mu(E))$ is represented as a network of nodes, where the nodes in the first layer are initialized by the values of $x_i^{(1)} = \mu(E_i)$; hence, their number is equal to the number of data points (of the order of a hundred to several hundreds) in the discretized XANES spectrum. The values of the nodes in the following layers are obtained as $x_i^{(n)} = f(z_i^{(n)}) = f\left(\theta_{i0}^{(n)} + \sum_{j=1}\theta_{ij}^{(n)} x_j^{(n-1)}\right)$, where summation is carried out over all nodes in the current layer, and $\theta_{ij}^{(n)}$ are parameters (weights and biases) that are optimized during the training step, and f is some activation function (rectified linear unit (ReLU) is used by Carbone et al.[96]). The output of the nodes in the NN final layer corresponds to the values of the material properties p. The advantage of NNs over other ML methods is that the complexity of the model $h(\Theta,\mu(E))$ can be easily increased by increasing the number of layers and nodes, and if very large training sets are available, quite complex relationships between spectral features and material structure can be constructed. Note here that for small training data sets, other ML methods may be better suited than NN.

A key property of NN that makes its training feasible is that, despite possibly a very complicated structure of NN, it is easy to calculate the derivative of the NN prediction error with respect to all the parameters Θ. Here, the definition of the NN prediction error (cost function) is problem-specific, but in the simplest case (more suitable for regression problems, discussed below, where the NN output can assume continuous values), it can be defined as $\varepsilon = \frac{1}{2K}\sum_{m,k}(x_{m,k}^{(n)} - p_{m,k})^2$ where summa-

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211

tion is carried out over all K training examples or a subset of training examples ("batch"), and $\tilde{p}_{m,k} = x_{m,k}^{(n)}$ and $p_{m,k}$ are the NN-yielded and the true values for parameters p_m, respectively (index m indicates here that NN can have more than one output value). The partial derivatives of ε are then commonly obtained by the so-called back-propagation algorithm. First, for the last NN layer (output layer), $\frac{\partial \varepsilon}{\partial \theta_{ij}^{(n)}}$ terms can be readily calculated. For example, with the above-defined cost function, $\frac{\partial \varepsilon}{\partial \theta_{ij}^{(n)}} = \frac{1}{K}\sum_k \left[(x_{i,k}^{(n)} - p_{i,k})\frac{\partial f}{\partial z_{i,k}^{(n)}}\right] x_{j,k}^{(n-1)}$. Similarly, the partial derivatives of the cost function with respect to the $\theta_{ij}^{(n-m)}$ values for deeper NN layers can be expressed as $\frac{\partial \varepsilon}{\partial \theta_{ij}^{(n-m)}} = \frac{1}{K}\sum_k \frac{\partial \varepsilon}{\partial z_{i,k}^{(n-m)}} x_{j,k}^{(n-m-1)}$. In the back-propagation algorithm, the $\frac{\partial \varepsilon}{\partial z_{i,k}^{(n-m)}}$ terms are recursively calculated layer-by-layer, knowing the corresponding values of derivatives for the weights in the following $(n-m+1)$ layer: $\frac{\partial \varepsilon}{\partial z_{i,k}^{(n-m)}} = \frac{\partial \varepsilon}{\partial x_{i,k}^{(n-m)}} \cdot \frac{\partial f}{\partial z_{i,k}^{(n-m),}}$, where $\frac{\partial \varepsilon}{\partial x_{i,k}^{(n-m)}} = \sum_l \theta_{il}^{(n-m)} \frac{\partial \varepsilon}{\partial z_{l,k}^{(n-m+1)}}$. See ref 97 for more details. Once the partial derivatives for all weights are calculated, it is easy to optimize their values using the standard gradient descent method or some more advanced approaches. For example, Carbone et al. used the so-called ADAM method,[98] which currently is a popular choice for NN training. Using this approach and another set of theoretical spectra for testing, local CN was identified with accuracy better than 90% for tetrahedrally and octahedrally coordinated species. Lower accuracy (ca. 70%) was achieved for 5-coordinated species, which often were mistakenly identified as tetrahedrally or octahedrally coordinated. Another important limitation was that the accuracy of the method, when it was applied to real experimental data, was significantly lower than that achieved in tests with theoretical data. This result is a consequence of significant systematic errors of XANES modeling and demonstrates clearly that extreme caution is needed, when ML-based approaches, trained on theoretical data, are applied for interpretation of experimental data.[96] In principle, the performance of the ML method on experimental data can be improved, if after the training on large sets of theoretical training data, the parameters of NN (e.g., in the last hidden layer only) are slightly reoptimized, using a small set of well-defined experimental data ("transfer learning" approach[96,99]).

It is also important to note that in the above-mentioned works, the different possible labels, that is, different classes of materials (oxide vs nonoxide, or tetrahedrally coordinated vs octahedrally coordinated), are predefined; thus, one assumes from the beginning which material characteristics are most important for explaining XAFS features. For parameters like oxidation state or the symmetry of the coordination environment around the absorbing atom, this approach works well, because these parameters have very large effects on XAFS and their importance is well-documented in the literature. For the interpretation of more subtle relations between spectral features and structural motifs it is desirable, however, to determine the key structural characteristics in an automatic, unbiased way. The data-driven approach by Kiyohara et al.[100] is a demonstration of this idea. In that work, 46 O K-edge ELNES/XANES spectra were generated in theoretical simulations. Spectra were then clustered in a few clusters (see Section 6.3), and then a decision tree was constructed, which describes the structural characteristics that are common for each cluster of spectra (valence state, presence of short or long bonds, etc.), thus automatically constructing a label for each cluster (Figure 2). Interpretation of an experimental XANES/ELNES spectrum is then carried out by just assigning it to the most similar cluster of spectra and reading out the corresponding label.

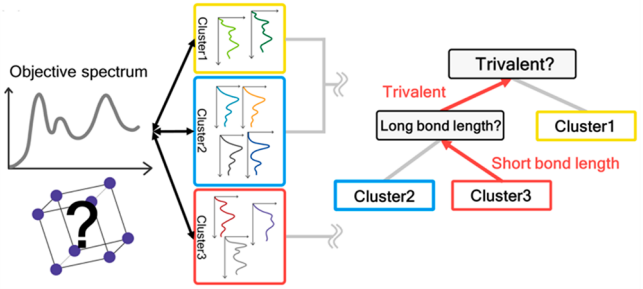


Figure 2. Automatic labeling of spectra using decision tree: generated theoretical spectra are clustered, and a decision tree is constructed, which assigns to each cluster a unique set of corresponding structural characteristics. Reproduced with permission from ref 100. Copyright 2018 Springer Nature. Reproduced under Creative Commons Attribution 4.0 International License: https://creativecommons.org/licenses/by/4.0/.

5. SUPERVISED MACHINE LEARNING: REGRESSION

5.1. New Opportunities for Analysis of X-ray Absorption Spectra. While in the case of classification the output of SML routine is a discrete variable that describes the belonging of input vector to a certain class, in the case of regression the output is a continuous variable (or several variables) that can characterize, for example, the NP size, surface-to-volume ratio, characteristic interatomic distance, bonding angle, or some other structurally relevant (as opposed to an abstract) parameter that affects XAFS spectra. We thus attempt to decipher much subtler differences due to small changes in structure, associated, for example, with in situ changes in NP shape, increase in disorder, interactions with adsorbates, and so on, that are inaccessible to other approaches for XAS data analysis.

While for classification problems different ML methods can be used, for regression analysis of XANES and EXAFS data, the usage of NNs seems to be the most beneficial. The reasons for that are (i) good scalability of NN-based models, which allows one to construct easily very complex relationships between spectral features and structure descriptors, and (ii) the nonlinear sensitivity of NNs to different features in the input vectors, which allows detection of minor changes in the analyzed spectra and, for example, enhancement of distant coordination shells contribution and minimization of the systematic errors influence. However, the training of complex NN-based regression models requires large training sets with thousands (or, in the case of XAFS data, tens or hundreds of thousands) of spectrum−structure pairs. Additionally, to minimize bias, the NN should be trained to predict structures that can be present in active states during chemical transformations and are unlikely to have analogues in experimentally available spectra. For example, if NN was trained on fcc-type NPs structures only, the NN yielded output will also always correspond to some fcc-type structure,

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211

regardless of the actual structure of the experimentally investigated NP. Ab initio simulations of XAFS data are therefore the only viable approach for constructing such data sets. Generation of training data is the most time-consuming part of NN-XAFS methods. Some automated procedures for selection of training examples for a particular problem may be helpful ("active learning" approach).[101,102]

Before discussing NN-EXAFS and NN-XANES/EELS approaches, let us note here that ML-based regression methods can be, of course, used in the interpretation of other experimental data. For example, ML methods, trained on experimental[103−105] or simulated[106] data are used for determination of chemical shifts in NMR analysis. An interesting (and conceptually similar to our NN-XAFS approach) application of ML methods is also a recent study by Ziatdinov et al.,[107] where NN is used to determine the atomic positions from microscopy images. As in the case of our NN-XAFS, simulated data are used for NN training.[107]

5.2. XANES Analysis. *5.2.1. Method.* The NN approach has been already introduced in Section 4. In our original NN-XANES regression scheme,[21] the output layer of NN produced a vector p_i containing average CNs for the first few coordination shells: $p_i = \{C_1^i, C_2^i, C_3^i, \ldots\}$. Knowledge of CNs allows one to estimate effective NP size, and, for small NPs with narrow size- and shape-distributions, also NP shape: for this purpose, CNs, extracted from NN-XANES analysis, are compared with a database of calculated CNs for particles of different sizes, shapes, and crystallographic structures.[58,108] Moreover, for well-defined reference samples with low disorder, CNs extracted from XANES data can be directly compared with the results of EXAFS analysis to validate the accuracy of our NN-XANES method (Figure 3a).[21] In the follow up study, we extended this approach and used NN to also determine from XANES spectra the effective interatomic distances R.[109] Since for small NPs the interatomic distances depend on NP size, we can use NN-yielded R and C_1 values as two independent characteristics of particle size (Figure 3b).

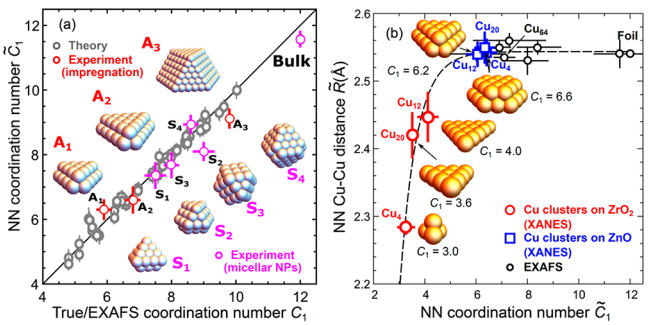


Figure 3. (a) Validation of the accuracy of NN for analysis of Pt L_3-edge XANES data: comparison of NN yielded CNs with their true values for theoretical models and with CNs extracted from conventional analysis of experimental EXAFS data. Shapes for Pt NPs, synthesized via support impregnation with incipient wetness method (samples A_1, A_2, A_3)[114] and for Pt NPs of similar sizes, synthesized using inverse micelle encapsulation (S_1, S_2, S_3, S_4)[112] are shown. Adapted with permission from ref 21. Copyright 2017 American Chemical Society. (b) Numbers of nearest neighbors and effective interatomic distances between them, as extracted by NN from in situ Cu K-edge XANES data for ultrasmall mass-selected Cu clusters under CO_2 conversion conditions. Possible cluster models are shown in the insets. Reprinted with permission from ref 109. Copyright 2018 American Chemical Society.

Other structural and nonstructural parameters that affect XANES spectra can be extracted in a similar manner.[88,101,110,111]

For NN training, we use theoretical XANES spectra for particles of different sizes, shapes, and crystallographic structures, calculated with FEFF[79] and FDMNES[89] codes. Some systematic errors present in both codes are partially compensated, when both of them are used to produce spectra for the NN training. For reliable results, and especially for the analysis of more distant coordination shells, many thousands of training spectra are needed. In principle, one could therefore construct a training set by generating thousands of NP models and calculating particle-averaged XANES spectra for each of them. This is, however, inefficient, because larger NPs have many nonequivalent sites, and several hundreds of site-specific XANES spectra need to be calculated to generate a single particle-averaged spectrum. Moreover, for larger NPs, the particle-averaged spectra are getting close to that of corresponding bulk material, and the contrast between different training examples would be low. That last limitation is the most severe: the degeneracy of a vast number of large-size NP models with respect to the spectra that correspond to them would, effectively, limit this method to generating only several hundreds of nonequivalent spectra, that is not enough for training of a NN for the analysis of XANES or EXAFS data. Instead of particle-averaged spectra, we use random linear combinations of a few site-specific XANES spectra to generate the training sets. From a relatively low number of NP models several hundreds of nonequivalent site-specific spectra can be generated, which then can be combined in practically unlimited number of linear combinations. Thus, a large set of training examples can be generated quickly and with modest computational resources.[21,108,109]

NNs with one or two hidden layers with several hundreds of nodes in each are usually sufficient to address the complexity of the problem.[21,108,109] For the simplest cases (e.g., when one is interested in the number of nearest neighbors in the first coordination shell only), satisfactory performance can be achieved even with the linear regression method (NN without hidden layers).[21] In addition to traditional, fully connected NNs, convolutional NNs can also be employed, which is a powerful method widely used (e.g., in the image processing). In the case of XANES data analysis, our tests do not show, however, any significant advantage of convolutional NNs over fully connected NNs.[99]

We would like to emphasize here again that caution is needed when NN, trained on theoretical data (which contain some unavoidable systematic errors), is applied to the interpretation of real experimental data. Because of the nonlinear nature of the NN function, even small systematic disagreement between experimental and theoretical data can, in principle, result in significant errors. The accuracy of NN predictions thus always needs to be validated using sets of experimental data for well-defined model samples. Moreover, since the accuracy of XANES simulations is different for different materials, a good performance of a NN in the interpretation of one absorption edge data (e.g., Pt L_3-edge) does not guarantee an equally good performance of similar NN for other element/absorption edge (e.g., Au L_3-edge), and it needs to be validated on the case-by-case basis. On the other hand, the nonlinear nature of NN is also an advantage in dealing with systematic errors: during the NN training, NN automatically assigns larger weights to those spectral features

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211

that are most relevant for determination of corresponding structure parameters. If these features are reproduced in simulations reliably, NN will have high accuracy even if systematic errors corrupt some other features in the absorption spectra. This gives significant advantage for NN-based methods over, for example, linear algebra-based methods, which weigh all the data points in the analyzed spectra uniformly, regardless of their relevance for the determination of materials structure. This makes NN-based methods also relatively robust toward the details of data preprocessing, for example, background subtraction and normalization of absorption spectra. While these steps can affect the shape of XANES features, their effect on absorption spectra is quite different from the effect of changes in material structure, and thus, from our experience, often does not influence the performance of NN. One can further minimize their influence, if instead of using absorption spectra $\mu(E)$ directly, NN is trained on theoretical $\Delta\mu(E) = \mu(E) - \mu_{ref}(E)$ data, where $\mu_{ref}(E)$ is a theoretical reference spectrum for corresponding well-defined material (e.g., bulk metal). After training is completed, as input for NN one provides experimental $\Delta\mu(E)$ spectrum, where the role of $\mu_{ref}(E)$ is played by a corresponding experimental reference spectrum for metal foil, for example. In the study of Pt NPs,[21] discussed below, such differential approach allowed us to extend the sensitivity of our NN up to the fourth coordination shell, for example. Clearly, some other steps in data preprocessing can have more critical influence. For example, shifts of spectral features due to a misalignment of experimental data can affect significantly the accuracy of NN predictions. Reference spectra thus always need to be collected together with the spectra of material of interest, to enable accurate alignment of energy scales. Data augmentation approach, discussed in the Section 5.3.1, can also be helpful in addressing the problem of data misalignment.

5.2.2. Examples. In our proof-of-principle work,[21] we applied the above-described approach to a set of γ-Al_2O_3-supported samples of Pt NPs with narrow NP size and shape distributions (Figure 3a). Our first step was to demonstrate the accuracy of our method using theoretical data. By comparing the CNs for the first four coordination shells, yielded by trained NN from the particle-averaged XANES spectra, with the true values for corresponding NP model, we confirmed that (i) our NN training was successful and that (ii) XANES spectra in metal NPs are sensitive not only to the number of nearest neighbors (which was acknowledged before[37,38]) but also to the arrangements of atoms in more distant coordination shells, thus providing the possibility to use XANES analysis for NPs shapes determination.[58] Next, we validated the robustness of our approach toward systematic errors of XANES modeling, by applying NN to experimental data and comparing NN-XANES yielded CNs with the results of conventional EXAFS analysis, which were also available for this well-defined model system. Excellent agreement was obtained. Finally, we used the CNs for the first four coordination shells to determine the average size and shape for experimentally investigated NPs. The obtained NPs sizes were in agreement with TEM results, while the obtained NPs shapes, intriguingly, showed sensitivity toward the NPs synthesis method: preformed Pt NPs, synthesized via inverse-micelle encapsulation method,[112] were more spherical, while NPs prepared via support impregnation with incipient wetness method were found to be flatter, suggesting stronger particle−support interactions. Such difference in NPs shape may have also important implications for understanding thermal[113] and catalytic properties[7] of these model catalysts.

Recently, we have revisited the above-described NN and applied it to analyze polarization-dependent Pt L_3-edge XANES data for Pt NPs on single-crystal supports.[115] Note here that the bonds with different orientation with respect to X-ray polarization contribute differently to the total XAFS spectrum. Thus, by comparing effective CNs collected with different X-ray polarizations (for oriented samples), one can get additional clues about the anisotropy of NPs structure and NP-support interactions.

In the follow-up works, we applied NN-XANES method for interpretation of in situ data. For example, it was used for the interpretation of Ag K-edge XANES data collected during self-assembly of silver clusters in an ionic liquid.[108,116] By following in situ changes in CNs, we concluded that the assembly of clusters occurs without coalescence, and the sizes and shapes of individual clusters are preserved in the assemblies. The work that is perhaps the best illustration of the prospects of NN-XANES method for the studies of catalysts, however, is our investigation of Cu clusters,[109] which are attractive catalysts for CO_2 conversion. Among other factors, the NP size has a decisive importance for the activity and selectivity of these catalysts.[117−120] To investigate the catalytic properties of ultrasmall (with only 4, 12, and 20 copper atoms) clusters, they were prepared in gas phase and soft-landed on a thin layer of oxide deposited by ALD on a silicon wafer. The as-prepared clusters were oxidized. Under reaction conditions (in CO_2 and H_2 mixture and at temperature 375 °C), however, the clusters were reduced, as evidenced by the changes in in situ XANES data, collected in grazing incidence (GI) mode. While the as-prepared clusters were mass-selected, it was not clear, whether they preserved their ultrasmall sizes under reaction conditions or sintered/agglomerated. For powder-supported NPs, this question could be answered by EXAFS data analysis. However, in this case, EXAFS data were not available because of the low loading of metal and strong Bragg reflections from the Si wafer support. Instead, we relied on the NN-based XANES analysis. We considered not only the possible changes in CNs but also in interatomic distances. Because of the expected reduction of interatomic distances in small NPs,[121,122] if we observe significantly reduced CNs simultaneously with the reduced interatomic distances (with respect to the values for bulk copper), we can be confident that the investigated clusters indeed preserve their initially ultrasmall sizes. It indeed was found to be the case for Cu clusters on ZrO_2 support (Figure 3b). For Cu clusters on ZnO support, however, significantly larger average CNs were obtained, as well as longer interatomic distances. This indicates that Cu clusters on ZnO significantly agglomerated.[109] Our efforts now are focused on the extension of the NN approach to nonmetallic clusters. For example, Figure 4 from an ongoing work demonstrates that similar sensitivity to NP size can be observed also in the Cu K-edge XANES data for partially oxidized, size-selective Cu clusters. Using the NN approach, the XANES spectra can be inverted and the information on the cluster size and oxidation state can be extracted, similarly to what has been done for the metallic Cu clusters.

Another good example of application of ML methods for quantitative interpretation of XANES data in catalytically relevant systems are the studies by Guda et al. of local structure around Ni in CPO-27-Ni metalorganic framework (MOF) upon adsorption of CO, CO_2, and NO.[88,111] Interaction with

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211

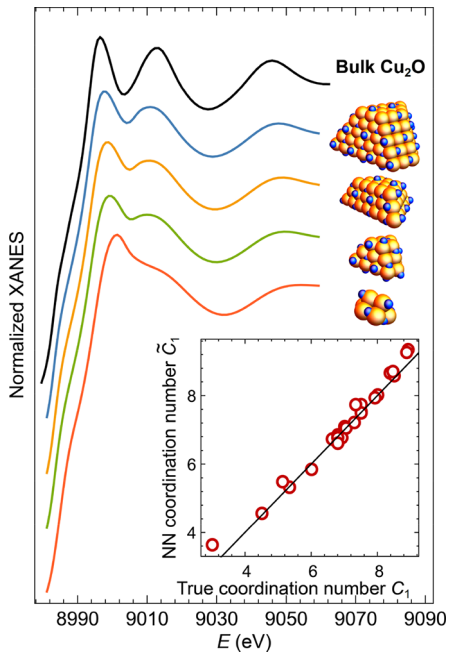


Figure 4. Particle size effect in theoretically calculated[109] Cu K-edge XANES spectra for Cu_2O clusters. Inset shows comparison of Cu−Cu CNs (which characterize NP size), extracted by NN from theoretical XANES data, with the true values for these clusters.

adsorbates affects Ni K-edge XANES, depending on the distance between Ni and adsorbate and the bond angle (e.g., Ni−N−O angle in the case of NO adsorption). The ML method is then trained on theoretical spectra, generated using FDMNES code[89] for DFT-optimized MOF structures. The values of structure parameters then can be successfully extracted from experimental Ni K-edge XANES data. Among other examples, we mention also the recent study by Kiyohara et al.,[110] where NN was used for interpretation of O K-edge ELNES data. In this study, 1171 O K-edge spectra were calculated using CASTEP code[123] for different SiO_2 polymorphs from Materials Project database.[91−93] NN was then trained on these data to predict average Si−O bond length, average bond angle, Voronoi volume, Mulliken charge, excitation energy, and bond overlap. This study thus sets a good example that geometrical parameters as well as parameters of electronic structure can be recognized by NN from near-edge data. NN, trained on theoretical data, was then able to obtain accurate values of these parameters from experimental O K-edge data for α-quartz.

5.3. EXAFS Analysis. *5.3.1. Method.* During the operando transformation of catalysts between different structures, including the strongly disordered ones, a concept of a "coordination shell" is not applicable because the ensemble averaged distribution of neighbors may not feature isolated, well-defined peaks. In these cases, instead of the commonly used (for structural characterization of catalysts) CNs, the structure can be more reliably characterized by partial RDF $g_{AB}(R)$ (eq 1), which has to be extracted from EXAFS data for this purpose. Conceptually, NN for the analysis of EXAFS data works in the same way as for XANES analysis (Figure 5a−d). NN takes EXAFS spectrum as input and yields the description of structure, in this case the entire RDF, parametrized as a histogram (each NN output node yields a height of particular histogram bin). NN training is carried out using a large set of spectra, for which the correct RDF is known.

The main difference between NN-XANES and NN-EXAFS approaches thus lies in the construction of their respective

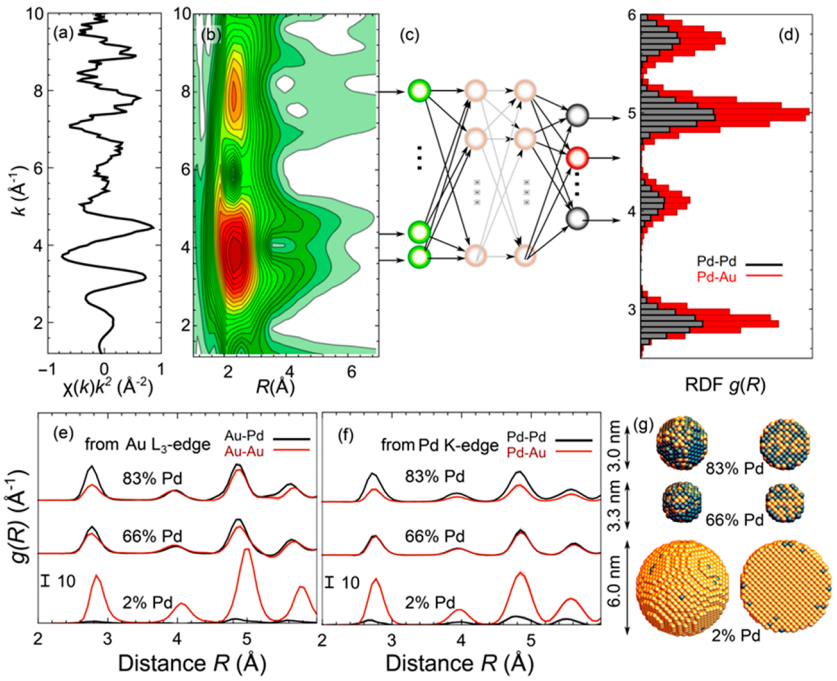


Figure 5. Experimental Pd K-edge EXAFS for PdAu NPs (a) is processed via wavelet transform (b) and fed into NN with several hidden layers (c). The output nodes of NN yield a histogram of bond-lengths (d), which describes RDFs of atoms of different types (Pd and Au) around the absorbing atoms (Pd). Partial RDFs, obtained by NN-EXAFS method for PdAu NPs of different sizes and compositions from experimental Au L_3-edge (e) and Pd K-edge EXAFS data (f). 3D structure models, obtained from the CNs yielded by NN-EXAFS (g). Arrows (in scale) indicate the sizes of NPs, as obtained from TEM. Adapted with permission from ref 137. Copyright 2019 American Chemical Society.

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211

training sets. XANES is relatively insensitive to minor displacements of atoms; therefore, for ML training, one can use simple polyhedral structure models without any disorder. It is not the case for EXAFS, and structure models with realistic structural and thermal disorder (realistic $g_{AB}(R)$) are needed. In our approach, we use for this purpose classical molecular dynamics (MD) simulations. It has been demonstrated that for many systems (bulk metals,[32,46,124] covalent materials,[125−131] NPs,[32,59,124,132−135] etc.), classical MD is a computationally inexpensive way to generate models that agree well with the available EXAFS data and other experimental information.[46,136] By carrying out MD simulations at different temperatures, structure models with different degree of disorder can be generated. Once the structure model is obtained, further generation of theoretical EXAFS spectra is straightforward using codes like FEFF[76] and GNXAS.[77] Similarly, as for NN-XANES analysis, for NN-EXAFS training one can also use either spectra, averaged over all absorbing atoms in the model, or site-specific spectra. The latter approach in beneficial for NP analysis.[137] Finally, as in the conventional EXAFS data analysis, before being analyzed by NN, experimental EXAFS spectra should be preprocessed to remove experimental noise, high-frequency contributions of distant coordination shells that are not included in the theoretical MD-EXAFS models and low-frequency contributions due to the artifacts of background subtraction. Fourier filtering can be used for this purpose. However, we have demonstrated that the usage of more advanced wavelet transform[138,139] can be beneficial, especially for multielement systems, providing better resolution of contributions from different elements and, hence, simplifying the NN analysis.[137] Clearly, the same Fourier- or wavelet filters used for experimental data need to be applied to theoretical data used for NN training.

Once the RDF is known, structure parameters of interest (partial CNs, interatomic distances, disorder factors) can be easily obtained by integration of $g_{AB}(R)$ function,[54,60,137] and used to construct 3D structure model of the material.[137] The influence of nonstructural parameters (E_0 and S_0^2, see eqs 1 and 2) also needs to be considered. S_0^2 correlates directly with the bin heights of $g(R)$ histogram (for a given bin size) and, hence, the values of CNs. Normally it is not a problem, because the value of S_0^2 can be estimated from the analysis of reference compounds. The effect of E_0, however, requires more attention, because it is sensitive to charge transfer, presence of adsorbates, support, temperature, and so on, as well as X-ray beam instabilities. Therefore, it can easily change from one sample to another and can even change in situ during the experiment for the same sample. The chosen value of E_0 can affect the shape of obtained RDF, and the error of ca. 1 eV in the E_0 value can noticeably affect the obtained result.[54] Fortunately, this problem can be addressed quite simply: in ref 54, we have shown that if each of the theoretical EXAFS spectra used for NN training is shifted by a random ΔE_0 value, NN performs much more stably and does not assign any physical meaning to small shifts in the energy scale of experimental data. Similar data augmentation schemes can be helpful in the interpretation of XANES data as well.[96] Other nonstructural parameters (e.g., parameters that define XAFS signal background) can also affect the shapes of spectral features, but their influence appears to be low because of (i) aforementioned Fourier of wavelet filtering that allows one to single out structure-sensitive information from smooth variations of the background and (ii) the nonlinear sensitivity of NNs that, as discussed in Section 5.2.1, automatically assigns large importance only to those spectral features that are relevant for determination of material's structure.

5.3.2. Examples. Our original work on NN-EXAFS method was devoted to the interpretation of Fe K-edge EXAFS data in bulk iron experiencing temperature-induced transition from bcc to fcc phase.[60] While that study did not focus on catalysis, the problem it addressed is relevant for many heterogeneous catalysts: under harsh experimental conditions, the structure of material changes significantly, and this change may be challenging to interpret using conventional approaches. For example, for iron at low temperature, in conventional EXAFS analysis, the bcc structure model fits experimental data noticeably better than the fcc model. At high temperature, however, when distributions of bond lengths are significantly broadened and asymmetric, the difference between fitting results with bcc and fcc models is smaller than the systematic error of the analysis, and the conventional EXAFS fitting cannot distinguish between these models conclusively. Finally, conventional EXAFS analysis methods will be particularly challenged by a spectrum that corresponds to a material measured during the structural phase transformation—for example, neither an fcc nor a bcc model would be adequate to analyze the spectrum of iron measured during such transformation in the example described above.

By using the NN-EXAFS approach, this problem can be addressed. By reconstructing RDFs directly from the experimental data, we are able to access directly the asymmetric shapes of RDF peaks, and integrate $g(R)$ function to extract accurate information on the average density of atom packing, which is different for fcc and bcc structures. Moreover, since the NN-EXAFS method is not limited to the first coordination shell, it allows us to analyze contributions of more distant neighbors, which are informative in even more challenging cases, when, for example, fcc structure needs to be distinguished from hcp structure.[60] This is an important problem in the determination of active species in Co-based Fischer−Tropsch catalysts, for example.[140]

Even more challenging is the determination of RDF in nanocatalysts, where in addition to thermal disorder, structure is affected by strong static disorder effects. As a result, reliable determination of CNs (which can be directly linked to particle size), interatomic distances and other structure parameters for small NPs is challenging even at room temperature. Recently, we have demonstrated[54] that analysis artifacts may contribute to investigation of anomalous thermal properties, observed by EXAFS method in small supported Pt NPs. In particular, it was demonstrated in previous works that for such NPs upon temperature increase, negative thermal expansion of NPs structure can be observed, as well as anomalously high Debye temperatures.[112−114,141] Different explanations for these effects have been proposed (e.g., temperature-dependent interactions between NP and support[114,142,143] or between NP and adsorbates).[113] While strong interaction with support and nonvibrational dynamic motion have been shown to be the main origin of the NTE factors in nm-scale NPs[114,141−143] prepared by support impregnation, recent NN-EXAFS analysis of Pt L_3-edge data for preformed NPs has shown a different explanation of their NTE.[54] We note that particle−support interactions, being significantly weaker in that latter case of preformed NPs, cannot explain the observed NTE in this system. By reconstructing shapes of RDFs in that latter

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211

material, Timoshenko et al. observed that they all were strongly asymmetric and non-Gaussian, due to the pronounced thermal and structural disorder. These effects are challenging to account for using conventional EXAFS analysis approaches. However, when they are considered explicitly using the NN-EXAFS method, one observes positive thermal expansion for those preformed NP samples (in the absence of hydrogen adsorbates and strong particle−support interactions that were shown to be responsible for real, not apparent NTE in the samples prepared by support impregnation[114,141−143]). Similarly, the anomalously large Debye temperatures, also observed for this system, can be explained by the disorder effects that are underestimated by conventional EXAFS fitting approaches at high temperatures.[54] This study thus demonstrates the potentiality of NN-EXAFS method for accurate determination of structure parameters in heterogeneous nanocatalysts at harsh experimental conditions.

Rich structural information can be extracted by NN-EXAFS method for bimetallic and more complex catalysts.[137] Note that discrimination between different heterometallic structural motifs (e.g., between alloyed and segregated NPs) is often based on the analysis of partial CNs and interatomic distances, extracted from EXAFS data.[144] The accurate determination of structural parameters, enabled by NN-EXAFS method, is thus a desired ability for studies of transformations of catalyst structures under reaction conditions. An important advantage of ML-based methods over conventional fitting approaches is also their applicability to the analysis of distant coordination shells, which for heterometallic materials are practically intractable by conventional methods but provides the possibility to discriminate between crystalline and amorphous structures, between close-packed and non-close-packed (e.g., icosahedral) atomic arrangements, as well to detect signatures of subtle ordering (e.g., second-neighbor ensembles[145,146]), which all may be important for the understanding of structure-properties relationship. In our recent work, we applied NN for the analysis of EXAFS data in PdAu NPs of different sizes and compositions.[137] Using both theoretical as well as experimental data, we have demonstrated the accuracy of this method for the determination of structural characteristic in bimetallic NPs. Moreover, by complementing NN-EXAFS analysis with a simple structure optimization algorithm, which processed simultaneously the information from the first four coordination shells extracted from Pd K-edge and Au L_3-edge EXAFS data, 3D models of analyzed NPs were constructed (Figure 5e−g). The sizes and compositions of obtained NP models matched those extracted from other experimental techniques (TEM and ICP-MS). At the same time, unique information about the distributions of different species within the NPs was obtained (in particular, a trend for Pd to segregate to NPs surface in Pd-poor NPs was observed).[137]

6. UNSUPERVISED MACHINE LEARNING: FINDING PATTERNS IN LARGE DATA SETS

6.1. Dimensionality Reduction. PCA. In studies of catalysts under reaction conditions, common is the situation when different species coexist. Interpretation of ensemble-averaged spectroscopic signals using supervised methods in these cases can be challenging, because there is no single label or single structure model that could be unambiguously assigned to the acquired spectrum. In this situation, unsupervised ML methods can be helpful. Two common tasks of unsupervised ML are data clustering and dimensionality reduction.[147,148] The objective of **clustering** is to find a few spectra that could represent the whole data set.[148] Instead of performing the analysis of the whole large data set (which may contain hundreds and thousands of spectra), analysis of these few representative spectra should provide a complete description of the sample. The goal of **dimensionality reduction** is also to provide a simplified description of the available data. One approach here is **feature extraction**,[149] where a few descriptors (e.g., position of the absorption edge, intensity of the white line in XANES, positions of the main peaks in Fourier-transformed EXAFS spectra, etc.) of the obtained spectra are used to track the changes in materials structure.[150] Selection of appropriate features, however, is material- and task-specific and requires human expertise, and therefore, it may be biased. An alternative approach relies on so-called **latent variable analysis**,[147] which assumes that the variations between all measured spectra can be explained by changes in a few variables (e.g., concentrations of constituent species). Measured experimental spectra $\mu_i(E)$ in this case are represented as

$$\mu_i(E) = \sum_{j=1}^{N} w_{ij} s_j(E) \tag{3}$$

where $s_j(E)$ are spectra for pure compounds, N is their number and w_{ij} − their weights. In matrix form, this equation can be rewritten as

$$M = WS \tag{4}$$

In the standard LCA,[61] spectra $s_j(E)$ need to be guessed in advance; thus, LCA is applicable only to studies of relatively simple, well-defined compounds. Much more powerful approach is PCA,[62−65] which does not require any additional human input and provides the most compact representation of the measured data set by finding orthogonal vectors that account for the largest discrepancies between spectra in the data set. Such decomposition of matrix M into matrices W and S can be ensured by singular value decomposition (see, e.g., ref 151 for detailed discussion). For PCA (and other approaches, discussed below) to work reliably, large sets of spectra are required (number of spectra should be much larger than the number of pure components), where weights of different components change systematically. Importantly, if the ratios of some pure components are the same in the whole data set, these components can be separated neither by PCA, nor methods, discussed below. Each of the components should also have a significant contribution to at least a few experimental spectra.[68,152] In this case, only relatively few vectors in matrix S will have nonzero contribution (nonzero w_{ij}) to the measured spectra. The orthogonality of the vectors in matrix S means that these few vectors form orthogonal basis, and the number N (number of significant s_j vectors) is then equal also to the number of independent components that are needed so that every spectrum in the data set can be described as their linear combination and thus will indicate the number of species in the mixture. Vectors in matrix S, however, do not correspond to any particular spectrum of pure compound but are their linear combinations. If the spectra for the pure compounds are known, the spectra for the unknown components can be reconstructed using target transformation.[151]

PCA-XANES is commonly used in the time-resolved in situ studies of catalysts. For example, it was used to identify reaction intermediates from the time-dependent Cu K-edge

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211

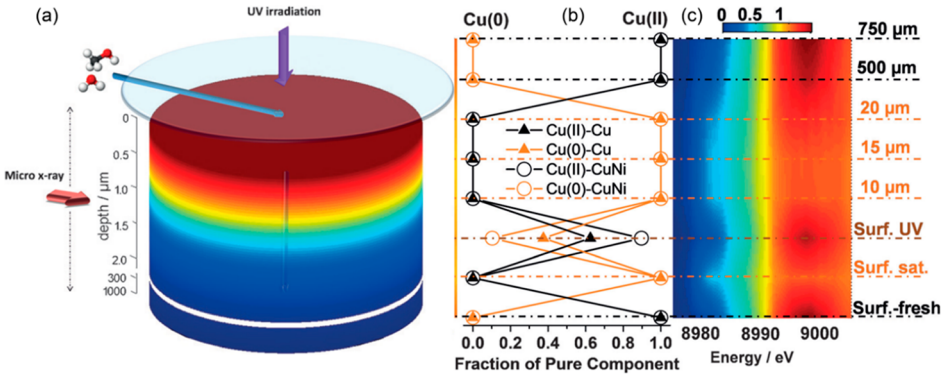


Figure 6. PCA analysis of XANES in CuNi photocatalyst. (a) Experimental scheme: sample is confined in a cell that allows simultaneous gas phase treatment and UV illumination from to top side. Arrows show the direction of the gas flow, UV illumination, and incident X-ray micro beam. (b) Concentration profiles of Cu^0 and Cu^{2+} species in CuNi sample as a function of the distance from the surface, obtained from PCA analysis of Cu K-edge XANES spectra (c) for UV-illuminated CuNi sample. Results for pure Cu are also shown in (b). Adapted with permission from ref 160. Copyright 2018 Wiley-VCH GmbH & Co. KGaA, Weinheim.

XANES spectra for $Ce_{0.8}Cu_{0.2}O_2$ catalyst reduction in hydrogen[153,154] and to study changes in oxidation state of copper in bimetallic CuPd catalyst on zeolite support.[155] PCA was used for speciation of molybdenum oxides (which are active catalysts for oxidation of propene) under reaction conditions,[156] study of Mn-promoted Fe-based Fischer−Tropsch catalysts,[157] and study of the oxidation state of Fe in complex Fe−Mo−Bi catalysts for propylene ammoxidation.[158] In addition to time-dependent studies, PCA is actively used for the interpretation of spatially resolved data of catalysts collected with focused X-ray beam or in full-field mode with 2D detectors.[24,159] For example, in ref 160, the distributions of oxidized and reduced species in illuminated by UV radiation CuNi photocatalyst were obtained by PCA analysis of XANES data collected with a microfocused beam, confirming the presence of Cu^{2+} species only in the illuminated zone of the material (Figure 6).

Clearly, PCA (as well as the approaches discussed below) can provide meaningful information only if the difference between XANES spectra for different species is significant. If the absorbing species in different phases are present in similar environments, or if the spectra are relatively featureless, speciation by XANES method will not be possible.[63] In some cases, this problem can be addressed by performing measurements in the high-energy resolution fluorescence detection (HERFD) mode, which provides superior sensitivity to different XAFS features due to lower spectral broadening.[161,162] For example, PCA-HERFD-XANES was recently used to probe in situ transformation of copper oxidation state in the industrial commercial $Cu/ZnO/Al_2O_3$ catalyst for methanol production.[161] Alternative possibility is to use instead of XANES corresponding EXAFS data that are more sensitive to the details of distance distributions.[64,163,164]

6.2. Blind Signal Separation and MCR-ALS. PCA yielded component are abstract and cannot be readily identified, because the requirement that the spectra $s_j(E)$ should be orthogonal is unphysical. In alternative approaches (blind signal separation methods[66−68]), such as independent component analysis (ICA), non-negative matrix factorization (NNMF)[147] and multivariate-curve resolution with alternating least-squares (MCR-ALS),[69−71] the approximations to matrices W and S are obtained by replacing this condition with other, physically meaningful constraints. As a result, $s_j(E)$ yielded by BSS can often be directly identified with the spectra for pure components. As such constraints, one can use the fact that XANES spectra $s_j(E)$ and corresponding weights w_{ij} should be all non-negative. In this case, the problem of solving eq 4 can be addressed by NNMF.[66] For the MCR-ALS method, in addition to non-negativity, it is also required that the sum of all the weights w_{ij} for a given experimental spectrum $\mu_i(E)$ should be equal to 1. Additional constraints can be given. For example, in ref 68, the BSS method was used for interpretation of Pd K-edge XANES spectra in thiolated Au-Pd clusters, where it was requested that the obtained XANES spectra $s_j(E)$ should not be too different from the theoretically simulated XANES spectra for different Pd placements within the bimetallic cluster.

MCR-ALS is the most popular BSS method for XANES data analysis. In this method, as a first step, the expected number of components N needs to be specified. PCA is normally used for this purpose. Next, initial guess of matrix W (or matrix S) (eq 4) is provided and constructed on the basis of PCA or rough LCA. eq 4 is then solved using the least-squares procedure, with respect to elements of matrix S (matrix W), subject to specified constraints. Next, elements of W (or S) are similarly refined. One continues alternate fitting of S and W for several iterations, until the difference between M and WS product does not improve anymore.[165,166]

To illustrate the accuracy of MCR-ALS method for XANES analysis, in Figure 7a we apply it to a set of synthetic data, constructed as linear combinations of experimental Cu K-edge XANES spectra for standard materials (metallic Cu foil and bulk Cu_2O and CuO). After several iterations, an excellent match between the MCR-ALS yielded approximations of spectra for pure compounds and the true spectra is obtained. The recovered concentration profiles also are in a good agreement with the true values. In Figure 7b, the results yielded by MCR-ALS in the interpretation of real experimental XANES spectra for alumina-supported copper catalysts from ref 71 are shown—the changes observed in Cu K-edge XANES spectra during the thermal activation of the catalyst can be identified with the changes in catalyst oxidation state. It needs to be emphasized here that the association of the spectral components with particular oxidation state and/or structural motif is a separate problem, because MCR-ALS yields only the corresponding XANES spectra for pure components, without any labels. For the interpretation of obtained spectra, one can compare them with the spectra of reference materials either

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211

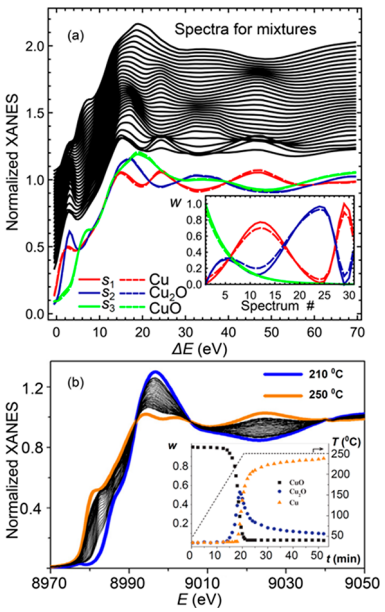


Figure 7. (a) Validation of MCR-ALS method using linear combinations (black solid lines) of Cu K-edge data for Cu foil, Cu_2O and CuO (dashed lines). Corresponding concentration profiles are shown as dashed lines in the inset. Spectra and concentration profiles, reconstructed by MCR-ALS, are shown as solid red, green, and blue lines. (b) Application of MCR-ALS to experimental data. (Adapted with permission from ref 71. Copyright 2014 Elsevier.) Evolution of Cu K-edge XANES for Cu catalyst on Al_2O_3 during activation process (heating in H_2 and He mixture from 50 to 250 °C, followed by measurements at constant 250 °C). Determined concentration profiles for Cu^{2+}, Cu^{+}, and Cu^{0} are shown in the inset.

manually or using automated procedure (see Section 4). Alternatively, ab initio XANES simulations can be used for this purpose, as demonstrated in recent work by Martini et al., where DFT-relaxed structural motifs were used for XANES modeling and then contrasted with the spectra, which were extracted by MCR-ALS from experimental data for Cu-based zeolites.[167] A good example of application of MCR-ALS technique for studies of zeolites-based catalysts is also the recent study by Pappas et al.,[162] where MCR-ALS was used for the interpretation of HERFD XANES data for Cu-based catalyst for methane-to-methanol conversion. Applications of MCR-ALS methods to zeolites were also recently reviewed by Guda et al.[88] Among other recent examples of MCR-ALS method applications, one can mention speciation of Co-based Fischer−Tropsch catalysts from time-resolved QXAFS data,[70,168,169] study of Cr oxidation state changes in Phillips catalyst during ethylene polymerization,[170] in situ study of formation of Ni−Mo catalysts for hydrodesulfurization reaction[171] and Fe−Ni catalyst for furfural hydrogenation,[172] as well as in situ studies of Cu NPs growth.[173,174]

Overall, the PCA and BSS methods provide the biggest advantage for speciation over conventional LCA method, when the structure of the investigated material differs significantly from the structure of well-defined bulk reference materials. Examples of such (catalytically relevant) systems are small NPs,[36,174,175] bimetallic alloys[155,160,172] and zeolites.[152,162,167,176] Also, one can note that PCA and BSS methods are useful, of course, not only for XAFS data interpretation: they are getting increasingly popular for the analysis of overlapping data in EDX microscopy,[177,178] analysis of XRD data[179] and other spectroscopies, including IR and mass spectroscopy.[180]

6.3. Clustering. The most common application of clustering methods in XAFS analysis is processing of spatially resolved spectra.[181−185] Two- or three-dimensional (the latter obtained using tomographic approaches) XANES mapping in some region of a sample is carried out, with the aim to establish spatial distributions of different species. By using clustering methods, two goals are achieved. First, the analysis of obtained data is significantly simplified, because instead of analyzing XANES spectra for each of the hundreds of pixels/voxels, only a few representative spectra need to be interpreted. Second, by monitoring, to which of the clusters each of the pixels/voxels is assigned, information about the distributions of different species can be immediately visualized. Figure 8 shows

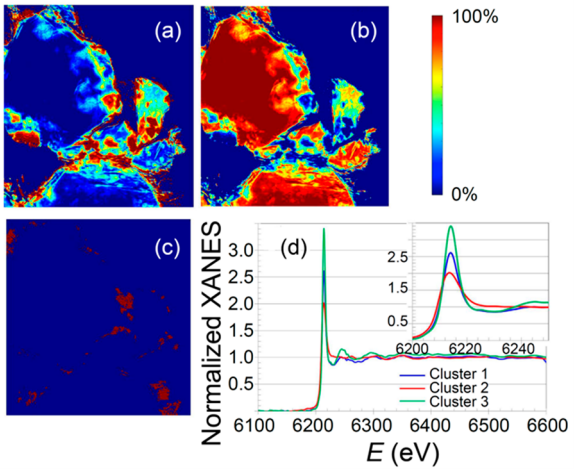


Figure 8. Clustering of XANES spectra for imaging applications. Pixels belonging to three clusters in the Nd L_3-XANES maps for permanent magnet material, $Nd_2Fe_{14}B$ are shown in (a−c). The representative Nd L_3-edge XANES spectra for each of the clusters are shown in (d). Reproduced with permission from ref 186. Copyright 2016 Springer Nature. Reproduced under Creative Commons Attribution 4.0 International License: https://creativecommons.org/licenses/by/4.0/.

an example of such analysis. Here 2D mapping of Nd L_3-edge XANES has been carried out in magnetic $Nd_2Fe_{14}B$ material using full-field imaging.[186] Each image contains 300 000 pixels (i.e., 300 000 XAFS spectra were collected). Clearly, most of the spectra are similar to each other. Using a clustering approach, all the collected spectra can be grouped into three groups. The differences in corresponding XANES spectra (Figure 8d) suggest differences in the Nd environment in these regions. For the interpretation of these differences in terms of CNs and interatomic distances, analysis of full EXAFS spectra was performed.[186] For clustering methods to work, the choice of similarity function is critical.[87,94] Different possible similarity functions are discussed in Section 4. Note also that instead of comparing spectra in energy space, one can first reduce the dimensionality of the data set using PCA, for example.[181−183]

The simplest clustering algorithm is *k*-means clustering,[187] which can be carried out using Lloyd's algorithm.[188] The number of clusters *k* should be specified in the beginning. Next, *k* spectra are picked randomly as initial guesses of representative spectra (cluster centers). Using the defined distance function, for each of the spectra in the data set nearest

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211

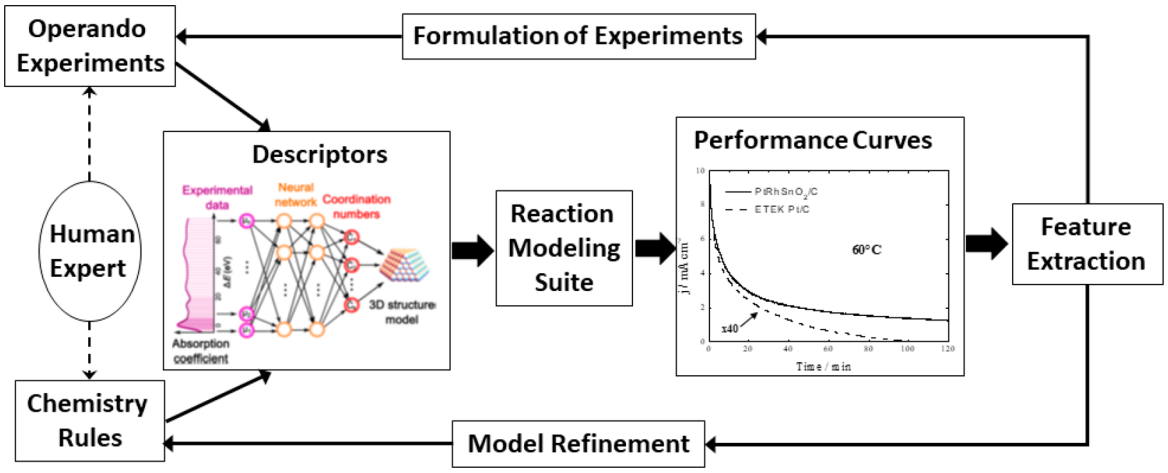


Figure 9. Example of a catalyst design schematic based on an active learning framework. Descriptors of catalytic activity will be identified by extracting the attributes of structure and electronic properties from the spectra and correlating them with catalyst performance.

cluster center is found. The new, updated approximations of representative spectra (cluster centers) are obtained by averaging all the spectra, assigned to each of the k clusters. Process is repeated iteratively, until there are no further changes in cluster centers or changes in cluster centers are smaller than a set threshold. This approach is widely used in the clustering of XAFS data.[183,189−192] However, more advanced clustering approaches, such as density-based spatial clustering are also used.[186]

Among recent examples of clustering-based analysis of XANES data in catalytically relevant materials, one can mention tomographic study of cobalt-based Fischer−Tropsch catalysts,[193] chemical imaging of single particles of Mo−Pt catalyst,[194] and study of the location and speciation of elements in Cu-based zeolites.[192] Without a doubt, such studies of catalysts will get increasingly popular in the future. One can mention that besides analysis of spatially resolved XAFS data, clustering approaches can also be used for other purposes, for example, to simplify ab initio XAFS calculations for large disordered structure models, generated by MD or reverse Monte Carlo (RMC) simulations.[46] In these cases, thousands of contributions of photoelectron scattering paths need to be considered, where many of them have similar (but not identical) geometries. To speed up calculations, contributions of similar paths can be automatically clustered together. Such an approach is implemented in EvAX code for RMC simulations of EXAFS spectra in crystalline and nanocrystalline materials.[32,126]

7. FUTURE OPPORTUNITIES FOR CATALYSIS RESEARCH

ML-assisted methods of data processing and analysis have undergone rapid development in the 2010s, coinciding with similarly expanding palettes of experimental methods, ranging from the high throughput to multimodal to operando modes of characterization. A recent article by Kitchin and co-workers outlines new opportunities that arise in catalysis research when ML is used for image analysis, chemical design, speeding up DFT calculations, and data mining in scientific literature.[18] In this section, we focus on the new opportunities that will become available on the basis of the new methodology summarized in this Perspective, namely, the "inversion" of optical (e.g., X-ray, UV−vis, Raman), electron and other spectra, and extracting practically important descriptors of the catalytic system.

One possible direction in which significant progress can be made is the rational design paradigm. Taking the original "inverse design" approach advanced by Caruthers et al.,[195] the design rules are obtained by linking the formulation of the catalyst and its performance using high-throughput experimentation and a reaction modeling method. From the knowledge of the structural and energetic descriptors, extracted from in situ and operando characterization, one can, in principle, obtain the smaller set that correlates with catalytic performance and, therefore, propose the design rules of the catalysts. This approach requires solving the "forward" problem (given the catalyst structure, predict its activity) and the "inverse" problem (given the desired activity, determine the catalyst structure).[196,197] Similarly, in one recent approach, the classification algorithm (Sure Independent Screening and Sparsifying Operator, or SISSO) was proposed to extract the material's property descriptors out of large feature spaces.[198] Here a desired property of the material (e.g., catalytic performance) is represented as a huge function of all possible structure descriptors and all possible combinations. This function (which may have billions of terms) is then optimized to fit available data for structure−property relationships, but the key point here is the additional condition that as few as possible terms should have nonzero contributions. In that way, the simplest possible relation between structure and properties is obtained. The usefulness of this approach for catalytic applications, however, yet needs to be demonstrated. A systematic approach was also proposed on the basis of using the Reaction Modeling Suite (RMS),[199] which enables automatic reaction kinetic modeling, particularly—formulating a reaction network from chemistry rules and optimizing the catalytic descriptors in the corresponding mathematical model.

A major challenge in these approaches is the difficulty to experimentally validate the predicted descriptors in working catalysts. In nanoscale systems, the descriptors require indirect techniques for their detection, such as spectroscopies, rather than more direct methods based on imaging or scattering. Using ML-based methods, the spectra can be analyzed by the "inversion" method to obtain the descriptors (Figure 9), hence, providing quantitative information for the RMS and closing the design loop described in ref 197.

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211

The ML-assisted analysis of X-ray spectra promises drastic improvements in the way catalytic characterization is done at the synchrotron and other facilities. We envision that the structure of catalysts would be possible to analyze by NN-XANES (and also NN-EXAFS) in harsh reaction conditions, low weight loading, and using in situ/operando reaction cells. Ultrasmall clusters and atomically dispersed catalysts could be characterized by XANES methods, currently underutilized for this purpose. High-energy resolution methods of fluorescence detection will enable much better sensitivity to spectral and, hence, structural features that the NN-XANES method can probe. Importantly, ML-based methods will improve the prospects of the use of lab-based X-ray sources[200] for catalytic studies that are currently unable to collect in situ EXAFS spectra due to very low flux. We also anticipate that "on the fly" data analysis will become possible because of the rapid (on the order of a second or less) analysis of the spectra by the pretrained NN. That capability will enable data analysis in real time while being collected and, in principle, the automated control of the reaction regimes aimed at improving catalytic activity, selectivity, or stability. Finally, our approach can be extended to other spectroscopies, most notably, EELS, Raman, and UV−visible light spectroscopies, where computational capabilities exist for forward modeling of their respective spectra.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: janis@fhi-berlin.mpg.de.
*E-mail: anatoly.frenkel@stonybrook.edu.

ORCID

Anatoly I. Frenkel: 0000-0002-5451-1207

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

A.I.F. acknowledges support by the Laboratory Directed Research and Development Program through LDRD 18-047 of Brookhaven National Laboratory under U.S. Department of Energy Contract No. DESC0012704 for initiating his research in machine learning methods. A.I.F.'s work on monometallic nanoparticles was supported by the U.S. Department of Energy, Office of Basic Energy Sciences under Grant No. DE-FG02-03ER15476. A.I.F.'s work on bimetallic nanoparticles was supported as part of the Integrated Mesoscale Architectures for Sustainable Catalysis (IMASC), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award No. DE-SC0012573. J.T.'s work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement ERC-OPERANDOCAT, No 725915). The authors gratefully acknowledge Drs. J. Aizenberg, S. R. Bare, M. Cargnello, B. Roldan Cuenya, R. M. Crooks, C. M. Friend, G. Henkelman, H. Hövel, M. Knecht, A. Kuzmin, D. Lu, R. G. Nuzzo, J. Purans, J. J. Rehr, S. Vajda, S. Yoo, and their group members for collaboration on multiple projects that contributed examples described in this Perspective.

REFERENCES

(1) Tao, F.; Grass, M. E.; Zhang, Y.; Butcher, D. R.; Renzas, J. R.; Liu, Z.; Chung, J. Y.; Mun, B. S.; Salmeron, M.; Somorjai, G. A. Reaction-Driven Restructuring of Rh-Pd and Pt-Pd Core-Shell Nanoparticles. *Science* **2008**, *322*, 932−934.

(2) Kondrat, S. A.; van Bokhoven, J. A., A Perspective on Counting Catalytic Active Sites and Rates of Reaction Using X-Ray Spectroscopy. *Top. Catal.* **2018**. DOI: 10.1007/s11244-018-1057-4.

(3) Kalz, K. F.; Kraehnert, R.; Dvoyashkin, M.; Dittmeyer, R.; Gläser, R.; Krewer, U.; Reuter, K.; Grunwaldt, J.-D. Future Challenges in Heterogeneous Catalysis: Understanding Catalysts under Dynamic Reaction Conditions. *ChemCatChem* **2017**, *9*, 17−29.

(4) Sinthika, S.; Waghmare, U. V.; Thapa, R. Structural and Electronic Descriptors of Catalytic Activity of Graphene-Based Materials: First-Principles Theoretical Analysis. *Small* **2018**, *14*, 1703609.

(5) Choksi, T. S.; Roling, L. T.; Streibel, V.; Abild-Pedersen, F. Predicting Adsorption Properties of Catalytic Descriptors on Bimetallic Nanoalloys with Site-Specific Precision. *J. Phys. Chem. Lett.* **2019**, *10*, 1852−1859.

(6) Getsoian, A. B.; Zhai, Z.; Bell, A. T. Band-Gap Energy as a Descriptor of Catalytic Activity for Propene Oxidation over Mixed Metal Oxide Catalysts. *J. Am. Chem. Soc.* **2014**, *136*, 13684−13697.

(7) Mostafa, S.; Behafarid, F.; Croy, J. R.; Ono, L. K.; Li, L.; Yang, J. C.; Frenkel, A. I.; Roldan Cuenya, B. Shape-Dependent Catalytic Properties of Pt Nanoparticles. *J. Am. Chem. Soc.* **2010**, *132*, 15714−15719.

(8) Madaan, N.; Shiju, N. R.; Rothenberg, G. Predicting the Performance of Oxidation Catalysts Using Descriptor Models. *Catal. Sci. Technol.* **2016**, *6*, 125−133.

(9) Zhang, S.; Li, Y.; Huang, J.; Lee, J.; Kim, D. H.; Frenkel, A. I.; Kim, T. Effects of Molecular and Electronic Structures in CoO_x/CeO_2 Catalysts on NO Reduction by CO. *J. Phys. Chem. C* **2019**, *123*, 7166−7177.

(10) Seh, Z. W.; Kibsgaard, J.; Dickens, C. F.; Chorkendorff, I.; Nørskov, J. K.; Jaramillo, T. F. Combining Theory and Experiment in Electrocatalysis: Insights into Materials Design. *Science* **2017**, *355*, No. eaad4998.

(11) Williams, W. D.; Shekhar, M.; Lee, W.-S.; Kispersky, V.; Delgass, W. N.; Ribeiro, F. H.; Kim, S. M.; Stach, E. A.; Miller, J. T.; Allard, L. F. Metallic Corner Atoms in Gold Clusters Supported on Rutile Are the Dominant Active Site during Water-Gas Shift Catalysis. *J. Am. Chem. Soc.* **2010**, *132*, 14018−14020.

(12) Cargnello, M.; Doan-Nguyen, V. V. T.; Gordon, T. R.; Diaz, R. E.; Stach, E. A.; Gorte, R. J.; Fornasiero, P.; Murray, C. B. Control of Metal Nanocrystal Size Reveals Metal-Support Interface Role for Ceria Catalysts. *Science* **2013**, *341*, 771−773.

(13) Topsøe, H. In situ characterization of catalysts. In *Studies in Surface Science and Catalysis*; Corma, A., Melo, F. V., Mendioroz, S., Fierro, J. L. G., Eds.; Elsevier, 2000; Vol. *130*, pp 1−21.

(14) Bañares, M. A.; Guerrero-Pérez, M. O.; Fierro, J. L. G.; Cortez, G. G. Raman Spectroscopy During Catalytic Operations with On-Line Activity Measurement (Operando Spectroscopy): a Method for Understanding the Active Centres of Cations Supported on Porous Materials. *J. Mater. Chem.* **2002**, *12*, 3337−3342.

(15) Weckhuysen, B. M. Operando Spectroscopy: Fundamental and Technical Aspects of Spectroscopy of Catalysts under Working Conditions. *Phys. Chem. Chem. Phys.* **2003**, *5*, No. vi-vi.

(16) Tinnemans, S. J.; Mesu, J. G.; Kervinen, K.; Visser, T.; Nijhuis, T. A.; Beale, A. M.; Keller, D. E.; van der Eerden, A. M. J.; Weckhuysen, B. M. Combining Operando Techniques in One Spectroscopic-Reaction Cell: New Opportunities for Elucidating the Active Site and Related Reaction Mechanism in Catalysis. *Catal. Today* **2006**, *113*, 3−15.

(17) Bañares, M. A. Operando Spectroscopy: the Knowledge Bridge to Assessing Structure-Performance Relationships in Catalyst Nanoparticles. *Adv. Mater.* **2011**, *23*, 5293−5301.

(18) Kitchin, J. R. Machine Learning in Catalysis. *Nature Catal.* **2018**, *1*, 230−232.

(19) Medford, A. J.; Kunz, M. R.; Ewing, S. M.; Borders, T.; Fushimi, R. Extracting Knowledge from Data through Catalysis Informatics. *ACS Catal.* **2018**, *8*, 7403−7429.

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211

(20) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, *559*, 547−555.
(21) Timoshenko, J.; Lu, D.; Lin, Y.; Frenkel, A. I. Supervised Machine-Learning-Based Determination of Three-Dimensional Structure of Metallic Nanoparticles. *J. Phys. Chem. Lett.* **2017**, *8*, 5091−5098.
(22) Qian, M.; Sarikaya, M.; Stern, E. A. Development of the EXELFS Technique for High Accuracy Structural Information. *Ultramicroscopy* **1995**, *59*, 137−147.
(23) Martin, J. M.; Mansot, J. L. EXELFS Analysis of Amorphous and Crystalline Silicon Carbide. *J. Microsc.* **1991**, *162*, 171−178.
(24) Grunwaldt, J.-D.; Hannemann, S.; Schroer, C. G.; Baiker, A. 2D-Mapping of the Catalyst Structure Inside a Catalytic Microreactor at Work: Partial Oxidation of Methane over Rh/Al_2O_3. *J. Phys. Chem. B* **2006**, *110*, 8674−8680.
(25) Briggs, B. D.; Bedford, N. M.; Seifert, S.; Koerner, H.; Ramezani-Dakhel, H.; Heinz, H.; Naik, R. R.; Frenkel, A. I.; Knecht, M. R. Atomic-Scale Identification of Pd Leaching in Nanoparticle Catalyzed C-C Coupling: Effects of Particle Surface Disorder. *Chem. Science* **2015**, *6*, 6413−6419.
(26) Billinge, S. J. L.; Levin, I. The Problem with Determining Atomic Structure at the Nanoscale. *Science* **2007**, *316*, 561−565.
(27) Ankudinov, A.; Rehr, J.; Low, J. J.; Bare, S. R. Sensitivity of Pt X-ray Absorption Near Edge Structure to the Morphology of Small Pt Clusters. *J. Chem. Phys.* **2002**, *116*, 1911−1919.
(28) Bordiga, S.; Groppo, E.; Agostini, G.; van Bokhoven, J. A.; Lamberti, C. Reactivity of Surface Species in Heterogeneous Catalysts Probed by In Situ X-ray Absorption Techniques. *Chem. Rev.* **2013**, *113*, 1736−1850.
(29) Bugaev, A. L.; Guda, A. A.; Lazzarini, A.; Lomachenko, K. A.; Groppo, E.; Pellegrini, R.; Piovano, A.; Emerich, H.; Soldatov, A. V.; Bugaev, L. A.; et al. In Situ Formation Hydrydes" and Carbides in Palladium Catalyst: When XANES is Better than EXAFS and XRD. *Catal. Today* **2017**, *283*, 119−126.
(30) Matos, J.; Ono, L. K.; Behafarid, F.; Croy, J. R.; Mostafa, S.; DeLaRiva, A. T.; Datye, A. K.; Frenkel, A. I.; Roldan Cuenya, B. In Situ Coarsening Study of Inverse Micelle-Prepared Pt Nanoparticles Supported on γ-Al_2O_3: Pretreatment and Environmental Effects. *Phys. Chem. Chem. Phys.* **2012**, *14*, 11457−11467.
(31) Chill, S. T.; Anderson, R. M.; Yancey, D. F.; Frenkel, A. I.; Crooks, R. M.; Henkelman, G. Probing the Limits of Conventional Extended X-Ray Absorption Fine Structure Analysis Using Thiolated Gold Nanoparticles. *ACS Nano* **2015**, *9*, 4036−4042.
(32) Timoshenko, J.; Frenkel, A. I. Probing Structural Relaxation in Nanosized Catalysts by Combining EXAFS and Reverse Monte Carlo Methods. *Catal. Today* **2017**, *280*, 274−282.
(33) Yevick, A.; Frenkel, A. I. Effects of Surface Disorder on EXAFS Modeling of Metallic Clusters. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *81*, 115451.
(34) Clausen, B. S.; Norskov, J. K. Asymmetric Pair Distribution Functions in Catalysts. *Top. Catal.* **2000**, *10*, 221−230.
(35) Witkowska, A.; Di Cicco, A.; Principi, E. Local Ordering of Nanostructured Pt Probed by Multiple-Scattering XAFS. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2007**, *76*, 104110.
(36) Ankudinov, A.; Rehr, J.; Low, J.; Bare, S. Theoretical Interpretation of XAFS and XANES in Pt Clusters. *Top. Catal.* **2002**, *18*, 3−7.
(37) Bazin, D.; Rehr, J. Limits and Advantages of X-Ray Absorption Near Edge Structure for Nanometer Scale Metallic Clusters. *J. Phys. Chem. B* **2003**, *107*, 12398−12402.
(38) Bazin, D.; Sayers, D.; Rehr, J.; Mottet, C. Numerical Simulation of the Platinum L_{III}Edge White Line Relative To Nanometer Scale Clusters. *J. Phys. Chem. B* **1997**, *101*, 5332−5336.
(39) Kuzmin, A.; Larcheri, S.; Rocca, F. Zn K-Edge XANES in Nanocrystalline ZnO. *J. Phys.: Conf. Ser.* **2007**, *93*, 012045.
(40) Patlolla, A.; Zunino, J.; Frenkel, A. I.; Iqbal, Z. Thermochromism in Polydiacetylene-Metal Oxide Nanocomposites. *J. Mater. Chem.* **2012**, *22*, 7028−7035.
(41) Tsapatsaris, N.; Beesley, A. M.; Weiher, N.; Tatton, H.; Dent, A. J.; Mosselmans, F. J.; Tromp, M.; Russu, S.; Evans, J.; Harvey, I.; et al. High Throughput In Situ XAFS Screening of Catalysts. *AIP Conf. Proc.* **2006**, *882*, 597−599.
(42) Grunwaldt, J.-D.; Frenkel, A. I. Synchrotron Studies of Catalysts: From XAFS to QEXAFS and Beyond. *Synchrotron Radiat. News* **2009**, *22*, 2−4.
(43) Stern, E. A.; Ma, Y.; Hanske-Petitpierre, O.; Bouldin, C. E. Radial Distribution Function in X-Ray-Absorption Fine Structure. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1992**, *46*, 687−694.
(44) Stern, E. EXAFS theory. In *X-Ray Absorption: Principles, Applications, Techniques of EXAFS, SEXAFS and XANES*; Koningsberger, D., Prins, R., Eds.; John Wiley and Sons: New York, 1988.
(45) Ravel, B.; Newville, M. ATHENA, ARTEMIS, HEPHAESTUS: Data Analysis for X-Ray Absorption Spectroscopy Using IFEFFIT. *J. Synchrotron Radiat.* **2005**, *12*, 537−541.
(46) Timoshenko, J.; Duan, Z.; Henkelman, G.; Crooks, R.; Frenkel, A. Solving the Structure and Dynamics of Metal Nanoparticles by Combining X-Ray Absorption Fine Structure Spectroscopy and Atomistic Structure Simulations. *Annu. Rev. Anal. Chem.* **2019**, *12*, 501−522.
(47) Frenkel, A.; Rehr, J. Thermal Expansion and X-Ray-Absorption Fine-Structure Cumulants. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1993**, *48*, 585−588.
(48) Di Cicco, A.; Minicucci, M.; Principi, E.; Witkowska, A.; Rybicki, J.; Laskowski, R. Testing Interaction Models by Using X-Ray Absorption Spectroscopy: Solid Pb. *J. Phys.: Cond. Matt.* **2002**, *14*, 3365.
(49) Clausen, B.; Grabaek, L.; Topsoe, H.; Hansen, L.; Stoltze, P.; Norskov, J.; Nielsen, O. A New Procedure for Particle Size Determination by EXAFS Based on Molecular Dynamics Simulations. *J. Catal.* **1993**, *141*, 368−379.
(50) Clausena, B.; Topsøe, H.; Hansen, L.; Stoltze, P.; Nørskov, J. Determination of Metal Particle Sizes From EXAFS. *Catal. Today* **1994**, *21*, 49−55.
(51) Clausen, B. S.; Topsøe, H.; Hansen, L. B.; Stoltze, P.; Nørskov, J. K. The Effect of Anharmonicity on the EXAFS Coordination Number in Small Metallic Particles. *Jpn. J. Appl. Phys.* **1993**, *32*, 95−97.
(52) Hansen, L. B.; Stoltze, P.; Nørskov, J. K.; Clausen, B.; Niemann, W. Is There A Contraction of the Interatomic Distance in Small Metal Particles? *Phys. Rev. Lett.* **1990**, *64*, 3155−3158.
(53) Yancey, D. F.; Chill, S. T.; Zhang, L.; Frenkel, A. I.; Henkelman, G.; Crooks, R. M. A theoretical and Experimental Examination of Systematic Ligand-Induced Disorder in Au Dendrimer-Encapsulated Nanoparticles. *Chem. Sci.* **2013**, *4*, 2912−2921.
(54) Timoshenko, J.; Ahmadi, M.; Roldan Cuenya, B. Is There A Negative Thermal Expansion in Supported Metal Nanoparticles? An In-Situ X-Ray Absorption Sttudy Coupled with Neural Network Analysis. *J. Phys. Chem. C* **2019**, *123*, 20594−20604.
(55) Huang, W. J.; Sun, R.; Tao, J.; Menard, L. D.; Nuzzo, R. G.; Zuo, J. M. Coordination-Dependent Surface Atomic Contraction in Nanocrystals Revealed by Coherent Diffraction. *Nat. Mater.* **2008**, *7*, 308.
(56) Mays, C. W.; Vermaak, J. S.; Kuhlmann-Wilsdorf, D. On Surface Stress and Surface Tension: II. Determination of the Surface Stress of Gold. *Surf. Sci.* **1968**, *12*, 134−140.
(57) Frenkel, A. Solving the 3D Structure of Metal Nanoparticles. *Z. Kristallogr.* **2007**, *222*, 605−611.
(58) Glasner, D.; Frenkel, A. I. Geometrical Characteristics of Regular Polyhedra: Application to EXAFS Studies of Nanoclusters. *AIP Conf. Proc.* **2006**, *882*, 746−748.
(59) Kuzmin, A.; Chaboy, J. EXAFS and XANES Analysis of Oxides at the Nanoscale. *IUCrJ* **2014**, *1*, 571−589.
(60) Timoshenko, J.; Anspoks, A.; Cintins, A.; Kuzmin, A.; Purans, J.; Frenkel, A. I. Neural Network Approach for Characterizing

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211

Structural Transformations by X-Ray Absorption Fine Structure Spectroscopy. *Phys. Rev. Lett.* **2018**, *120*, 225502.
(61) Benfatto, M.; Meneghini, C. A Close Look Into the Low Energy Region of the XAS Spectra: the XANES Region. In *Synchrotron Radiation Basics, Methods and Applications*; Mobilio, S., Boscherini, F., Meneghini, C., Eds.; Springer: Berlin, 2015; pp 213−240.
(62) Wasserman, S. R. The Analysis of Mixtures: Application of Principal Component Analysis to XAS Spectra. *J. Phys. IV France* **1997**, *7*, C2-203−C2-205.
(63) Frenkel, A. I.; Kleifeld, O.; Wasserman, S. R.; Sagi, I. Phase Speciation by Extended X-Ray Absorption Fine Structure Spectroscopy. *J. Chem. Phys.* **2002**, *116*, 9449−9456.
(64) Wang, X.; Hanson, J. C.; Frenkel, A. I.; Kim, J.-Y.; Rodriguez, J. A. Time-resolved Studies for the Mechanism of Reduction of Copper Oxides with Carbon Monoxide: Complex Behavior of Lattice Oxygen and the Formation of Suboxides. *J. Phys. Chem. B* **2004**, *108*, 13667−13673.
(65) Manceau, A.; Marcus, M.; Lenoir, T. Estimating the Number of Pure Chemical Components In A Mixture by X-Ray Absorption Spectroscopy. *J. Synchrotron Radiat.* **2014**, *21*, 1140−1147.
(66) Lee, D. D.; Seung, H. S. Learning the Parts of Objects by Non-Negative Matrix Factorization. *Nature* **1999**, *401*, 788−791.
(67) Hyvärinen, A.; Oja, E. Independent Component Analysis: Algorithms and Applications. *Neural Networks* **2000**, *13*, 411−430.
(68) Timoshenko, J.; Shivhare, A.; Scott, R. W. J.; Lu, D.; Frenkel, A. I. Solving Local Structure Around Dopants in Metal Nanoparticles with Ab Initio Modeling of X-Ray Absorption Near Edge Structure. *Phys. Chem. Chem. Phys.* **2016**, *18*, 19621−19630.
(69) Ruckebusch, C.; Blanchet, L. Multivariate Curve Resolution: A Review of Advanced and Tailored Applications and Challenges. *Anal. Chim. Acta* **2013**, *765*, 28−36.
(70) Voronov, A.; Urakawa, A.; van Beek, W.; Tsakoumis, N. E.; Emerich, H.; Rønning, M. Multivariate Curve Resolution Applied To In Situ X-Ray Absorption Spectroscopy Data: An Efficient Tool for Data Processing and Analysis. *Anal. Chim. Acta* **2014**, *840*, 20−27.
(71) Cassinelli, W. H.; Martins, L.; Passos, A. R.; Pulcinelli, S. H.; Santilli, C. V.; Rochet, A.; Briois, V. Multivariate Curve Resolution Analysis Applied To Time-Resolved Synchrotron X-Ray Absorption Spectroscopy Monitoring of the Activation of Copper Alumina Catalyst. *Catal. Today* **2014**, *229*, 114−122.
(72) Newville, M.; Carroll, S.; O'day, P.; Waychunas, G.; Ebert, M. A Web-Based Library of XAFS Data on Model Compounds. *J. Synchrotron Radiat.* **1999**, *6*, 276−277.
(73) Asakura, K.; Ikemoto, I.; Kuroda, H.; Kobayashi, T.; Shirakawa, H. Dopant Structure in $FeCl_3$-Doped Polyacetylene Studied by X-Ray Absorption Spectroscopy and X-Ray Photoelectron Spectroscopy. *Bull. Chem. Soc. Jpn.* **1985**, *58*, 2113−2120.
(74) Ewels, P.; Sikora, T.; Serin, V.; Ewels, C. P.; Lajaunie, L. A Complete Overhaul of the Electron Energy-Loss Spectroscopy and X-Ray Absorption Spectroscopy Database: Eelsdb. Eu. *Microsc. Microanal.* **2016**, *22*, 717−724.
(75) Mathew, K.; Zheng, C.; Winston, D.; Chen, C.; Dozier, A.; Rehr, J. J.; Ong, S. P.; Persson, K. A. High-Throughput Computational X-Ray Absorption Spectroscopy. *Scientific Data* **2018**, *5*, 180151.
(76) Ankudinov, A. L.; Ravel, B.; Rehr, J. J.; Conradson, S. D. Real-space Multiple-Scattering Calculation and Interpretation of X-Ray-Absorption Near-Edge Structure. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *58*, 7565−7576.
(77) Filipponi, A.; Di Cicco, A.; Natoli, C. R. X-Ray-Absorption Spectroscopy and N-Body Distribution Functions in Condensed Matter. I. Theory. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1995**, *52*, 15122.
(78) Rehr, J.; Ankudinov, A. Progress in the Theory and Interpretation of XANES. *Coord. Chem. Rev.* **2005**, *249*, 131−140.
(79) Rehr, J. J.; Kas, J. J.; Vila, F. D.; Prange, M. P.; Jorissen, K. Parameter-Free Calculations of X-Ray Spectra with FEFF9. *Phys. Chem. Chem. Phys.* **2010**, *12*, 5503−5513.
(80) Vinson, J.; Rehr, J.; Kas, J.; Shirley, E. Bethe-Salpeter Equation Calculations of Core Excitation Spectra. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2011**, *83*, 115106.
(81) Laskowski, R.; Blaha, P. Understanding the $L_{2,3}$ X-Ray Absorption Spectra of Early 3 D Transition Elements. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *82*, 205104.
(82) Ankudinov, A.; Takimoto, Y.; Rehr, J. Combined Bethe-Saltpeter Equations and Time-Dependent Density-Functional Theory Approach for X-Ray Absorption Calculations. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2005**, *71*, 165110.
(83) Tanaka, I.; Mizoguchi, T. First-Principles Calculations of X-Ray Absorption Near Edge Structure and Energy Loss Near Edge Structure: Present and Future. *J. Phys.: Condens. Matter* **2009**, *21*, 104201.
(84) Stavitski, E.; De Groot, F. M. The CTM4XAS Program for EELS and XAS Spectral Shape Analysis of Transition Metal L Edges. *Micron* **2010**, *41*, 687−694.
(85) Liang, Y.; Vinson, J.; Pemmaraju, S.; Drisdell, W. S.; Shirley, E. L.; Prendergast, D. Accurate X-Ray Spectral Predictions: An Advanced Self-Consistent-Field Approach Inspired by Many-Body Perturbation Theory. *Phys. Rev. Lett.* **2017**, *118*, 096402.
(86) Shirley, E. L. Ab Initio Inclusion of Electron-Hole Attraction: Application to X-Ray Absorption and Resonant Inelastic X-Ray Scattering. *Phys. Rev. Lett.* **1998**, *80*, 794.
(87) Suzuki, Y.; Hino, H.; Kotsugi, M.; Ono, K. Automated Estimation of Materials Parameter from X-Ray Absorption and Electron Energy-Loss Spectra with Similarity Measures. *Npj Computational Materials* **2019**, *5*, 39.
(88) Guda, A. A.; Guda, S. A.; Lomachenko, K. A.; Soldatov, M. A.; Pankin, I. A.; Soldatov, A. V.; Braglia, L.; Bugaev, A. L.; Martini, A.; Signorile, M.; et al. Quantitative Structural Determination of Active Sites from In Situ and Operando XANES Spectra: From Standard Ab Initio Simulations to Chemometric and Machine Learning Approaches. *Catal. Today* **2019**, *336*, 3−21.
(89) Bunău, O.; Joly, Y. Self-Consistent Aspects of X-Ray Absorption Calculations. *J. Phys.: Condens. Matter* **2009**, *21*, 345501.
(90) Zheng, C.; Mathew, K.; Chen, C.; Chen, Y.; Tang, H.; Dozier, A.; Kas, J. J.; Vila, F. D.; Rehr, J. J.; Piper, L. F.; et al. Automated Generation and Ensemble-Learned Matching of X-Ray Absorption Spectra. *Npj Computational Materials* **2018**, *4*, 12.
(91) Ong, S. P.; Cholia, S.; Jain, A.; Brafman, M.; Gunter, D.; Ceder, G.; Persson, K. A. 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.* **2015**, *97*, 209−215.
(92) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A Robust, Open-Source Python Library for Materials Analysis. *Comput. Mater. Sci.* **2013**, *68*, 314−319.
(93) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Materials* **2013**, *1*, 011002.
(94) Iwasaki, Y.; Kusne, A. G.; Takeuchi, I. Comparison of Dissimilarity Measures for Cluster Analysis of X-Ray Diffraction Data from Combinatorial Libraries. *Npj Computational Materials* **2017**, *3*, 4.
(95) Miyazato, I.; Takahashi, L.; Takahashi, K., Automatic Oxidation Threshold Recognition of XAFS Data Using Supervised Machine Learning. *Molecular Systems Design & Engineering* **2019**, DOI: 10.1039/C9ME00043G.
(96) Carbone, M. R.; Yoo, S.; Topsakal, M.; Lu, D. Classification of Local Chemical Environments from X-Ray Absorption Spectra Using Supervised Machine Learning. *Phys. Rev. Materials* **2019**, *3*, 033604.
(97) LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436.

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211

(98) Kingma, D. P.; Ba, J., Adam: A Method for Stochastic Optimization. 2014 *arXiv:1412.6980*. arXiv.org e-Print archive. https://arxiv.org/abs/1412.6980.
(99) Lin, Y.; Topaskal, M.; Timoshenko, J.; Lu, D.; Yoo, S.; Frenkel, A. I., Machine-Learning Assisted Determination of Coordination Numbers of Metallic Nanoparticles - A Benchmark. *Experimental analysis solutions for leading experimental technques*; World Scientific Publisher, 2019.
(100) Kiyohara, S.; Miyata, T.; Tsuda, K.; Mizoguchi, T. Data-Driven Approach for the Prediction and Interpretation of Core-Electron Loss Spectroscopy. *Sci. Rep.* **2018**, *8*, 13548.
(101) Ueno, T.; Hino, H.; Hashimoto, A.; Takeichi, Y.; Sawada, M.; Ono, K. Adaptive Design ofan X-Ray Magnetic Circular Dichroism Spectroscopy Experiment with Gaussian Process Modelling. *Npj Computational Materials* **2018**, *4*, 4.
(102) Ukil, A.; Bernasconi, J. Neural Network-Based Active Learning in Multivariate Calibration. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **2012**, *42*, 1763−1771.
(103) Kvasnička, V. An Application of Neural Networks In Chemistry. Prediction of13c NMR Chemical Shifts. *J. Math. Chem.* **1991**, *6*, 63−76.
(104) Meiler, J. PROSHIFT: Protein Chemical Shift Prediction Using Artificial Neural Networks. *J. Biomol. NMR* **2003**, *26*, 25−37.
(105) Shen, Y.; Bax, A. Protein Backbone Chemical Shifts Predicted from Searching A Database for Torsion Angle and Sequence Homology. *J. Biomol. NMR* **2007**, *38*, 289−302.
(106) Cuny, J.; Xie, Y.; Pickard, C. J.; Hassanali, A. A. Ab Initio Quality NMR Parameters in Solid-State Materials Using a High-Dimensional Neural-Network Representation. *J. Chem. Theory Comput.* **2016**, *12*, 765−773.
(107) Ziatdinov, M.; Dyck, O.; Maksov, A.; Li, X.; Sang, X.; Xiao, K.; Unocic, R. R.; Vasudevan, R.; Jesse, S.; Kalinin, S. V. Deep Learning of Atomically Resolved Scanning Transmission Electron Microscopy Images: Chemical Identification and Tracking Local Transformations. *ACS Nano* **2017**, *11*, 12742−12752.
(108) Timoshenko, J.; Roese, S.; Hövel, H.; Frenkel, A. I., Silver Clusters Shape Determination from In-Situ XANES Data. *Radiat. Phys. Chem.* **2018**. DOI: 10.1016/j.radphyschem.2018.11.003.
(109) Timoshenko, J.; Halder, A.; Yang, B.; Seifert, S.; Pellin, M. J.; Vajda, S.; Frenkel, A. I. Subnanometer Substructures in Nanoassemblies Formed from Clusters under a Reactive Atmosphere Revealed Using Machine Learning. *J. Phys. Chem. C* **2018**, *122*, 21686−21693.
(110) Kiyohara, S.; Tsubaki, M.; Liao, K.; Mizoguchi, T. Quantitative Estimation of Properties from Core-Loss Spectrum Via Neural Network. *Journal of Physics: Materials* **2019**, *2*, 024003.
(111) Guda, A.; Guda, S.; Martini, A.; Bugaev, A.; Soldatov, M.; Soldatov, A.; Lamberti, C. Machine learning Approaches to XANES Spectra for Quantitative 3D Structural Determination: The Case of CO_2 Adsorption on CPO-27-Ni MOF. *Radiat. Phys. Chem.* **2019**, 108430.
(112) Roldan Cuenya, B.; Frenkel, A.; Mostafa, S.; Behafarid, F.; Croy, J.; Ono, L.; Wang, Q. Anomalous Lattice Dynamics and Thermal Properties of Supported Size-And Shape-Selected Pt Nanoparticles. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *82*, 155450.
(113) Roldan Cuenya, B.; Ortigoza, M. A.; Ono, L.; Behafarid, F.; Mostafa, S.; Croy, J.; Paredis, K.; Shafai, G.; Rahman, T.; Li, L.; et al. Thermodynamic Properties of Pt Nanoparticles: Size, Shape, Support, and Adsorbate Effects. *Phys. Rev. B* **2011**, *84*, 245438.
(114) Sanchez, S. I.; Menard, L. D.; Bram, A.; Kang, J. H.; Small, M. W.; Nuzzo, R. G.; Frenkel, A. I. The Emergence of Nonbulk Properties in Supported Metal Clusters: Negative Thermal Expansion and Atomic Disorder in Pt Nanoclusters Supported on γ-Al_2O_3. *J. Am. Chem. Soc.* **2009**, *131*, 7040−7054.
(115) Ahmadi, M.; Timoshenko, J.; Behafarid, F.; Roldan Cuenya, B. Tuning the Structure of Pt Nanoparticles Through Support Interactions: An In Situ Polarized X-Ray Absorption Study Coupled with Atomistic Simulations. *J. Phys. Chem. C* **2019**, *123*, 10666−10676.
(116) Roese, S.; Kononov, A.; Timoshenko, J.; Frenkel, A.; Hoevel, H. Cluster Assemblies Produced by Aggregation of Preformed Ag Clusters in Ionic Liquids. *Langmuir* **2018**, *34*, 4811−4819.
(117) Halder, A.; Curtiss, L. A.; Fortunelli, A.; Vajda, S. Perspective: Size Selected Clusters for Catalysis and Electrochemistry. *J. Chem. Phys.* **2018**, *148*, 110901.
(118) Yang, B.; Liu, C.; Halder, A.; Tyo, E. C.; Martinson, A. B. F.; Seifert, S.; Zapol, P.; Curtiss, L. A.; Vajda, S. Copper Cluster Size Effect in Methanol Synthesis from CO_2. *J. Phys. Chem. C* **2017**, *121*, 10406−10412.
(119) Vajda, S.; White, M. G. Catalysis Applications of Size-Selected Cluster Deposition. *ACS Catal.* **2015**, *5*, 7152−7176.
(120) Reske, R.; Mistry, H.; Behafarid, F.; Roldan Cuenya, B.; Strasser, P. Particle Size Effects in the Catalytic Electroreduction of CO_2 on Cu Nanoparticles. *J. Am. Chem. Soc.* **2014**, *136*, 6978−6986.
(121) Montano, P. A.; Shenoy, G. K.; Alp, E. E.; Schulze, W.; Urban, J. Structure of Copper Microclusters Isolated in Solid Argon. *Phys. Rev. Lett.* **1986**, *56*, 2076−2079.
(122) Frenkel, A.; Nemzer, S.; Pister, I.; Soussan, L.; Harris, T.; Sun, Y.; Rafailovich, M. Size-Controlled Synthesis and Characterization of Thiol-Stabilized Gold Nanoparticles. *J. Chem. Phys.* **2005**, *123*, 184701.
(123) Clark, S. J.; Segall, M. D.; Pickard, C. J.; Hasnip, P. J.; Probert, M. I.; Refson, K.; Payne, M. C. First Principles Methods Using CASTEP. *Z. Kristallogr.* **2005**, *220*, 567−570.
(124) Timoshenko, J.; Keller, K. R.; Frenkel, A. I. Determination of Bimetallic Architectures in Nanometer-Scale Catalysts by Combining Molecular Dynamics Simulations with X-Ray Absorption Spectroscopy. *J. Chem. Phys.* **2017**, *146*, 114201.
(125) Jonane, I.; Lazdins, K.; Timoshenko, J.; Kuzmin, A.; Purans, J.; Vladimirov, P.; Gräning, T.; Hoffmann, J. Temperature-Dependent EXAFS Study of the Local Structure and Lattice Dynamics in Cubic Y2O3. *J. Synchrotron Radiat.* **2016**, *23*, 510−518.
(126) Timoshenko, J.; Kuzmin, A.; Purans, J. EXAFS Study of Hydrogen Intercalation Into ReO_3 Using the Evolutionary Algorithm. *J. Phys.: Condens. Matter* **2014**, *26*, 055401.
(127) Timoshenko, J.; Anspoks, A.; Kalinko, A.; Kuzmin, A. Temperature Dependence of the Local Structure and Lattice Dynamics of Wurtzite-Type ZnO. *Acta Mater.* **2014**, *79*, 194−202.
(128) Timoshenko, J.; Kuzmin, A.; Purans, J. Molecular Dynamics Simulations of EXAFS in Germanium. *Central European J. Phys.* **2011**, *9*, 710−715.
(129) Kalinko, A.; Evarestov, R.; Kuzmin, A.; Purans, J. Interpretation of EXAFS in ReO_3 Using Molecular Dynamics Simulations. *J. Phys.: Conf. Ser.* **2009**, *190*, 012080.
(130) Kuzmin, A.; Timoshenko, J.; Kalinko, A.; Jonane, I.; Anspoks, A. Treatment of Disorder Effects in X-Ray Absorption Spectra Beyond the Conventional Approach. *Radiat. Phys. Chem.* **2018**. DOI: 10.1016/j.radphyschem.2018.12.032.
(131) Kuzmin, A.; Anspoks, A.; Kalinko, A.; Timoshenko, J. The Use of X-Ray Absorption Spectra for Validation of Classical Force-Field Models. *Z. Phys. Chemie* **2016**, *230*, 537−549.
(132) Duan, Z.; Timoshenko, J.; Kunal, P.; House, S. D.; Wan, H.; Jarvis, K.; Bonifacio, C. S.; Yang, J. C.; Crooks, R. M.; Frenkel, A. I.; Humphrey, S. M.; Henkelman, G. Structural Characterization on Heterogeneous Rh-Au Nanoparticles from a Microwave-Assited Synthesis. *Nanoscale* **2018**, *10*, 22520−22532.
(133) Anspoks, A.; Kalinko, A.; Kalendarev, R.; Kuzmin, A. Atomic Structure Relaxation in Nanocrystalline NiO Studied by EXAFS Spectroscopy: Role of Nickel Vacancies. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2012**, *86*, 174114.
(134) Anspoks, A.; Kuzmin, A.; Kalinko, A.; Timoshenko, J. Probing NiO Nanocrystals by EXAFS Spectroscopy. *Solid State Commun.* **2010**, *150*, 2270−2274.
(135) Prasai, B.; Wilson, A.; Wiley, B.; Ren, Y.; Petkov, V. On the Road To Metallic Nanoparticles by Rational Design: Bridging the

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211

Gap Between Atomic-Level Theoretical Modeling and Reality by Total Scattering Experiments. *Nanoscale* **2015**, *7*, 17902−17922.
(136) Kuzmin, A.; Evarestov, R. Quantum Mechanics-Molecular Dynamics Approach To the Interpretation of X-Ray Absorption Spectra. *J. Phys.: Condens. Matter* **2009**, *21*, 055401.
(137) Timoshenko, J.; Wrasman, C. J.; Luneau, M.; Shirman, T.; Cargnello, M.; Bare, S. R.; Aizenberg, J.; Friend, C. M.; Frenkel, A. I. Probing Atomic Distributions in Mono- and Bimetallic Nanoparticles by Supervised Machine Learning. *Nano Lett.* **2019**, *19*, 520−529.
(138) Timoshenko, J.; Kuzmin, A. Wavelet Data Analysis of EXAFS Spectra. *Comput. Phys. Commun.* **2009**, *180*, 920−925.
(139) Funke, H.; Scheinost, A.; Chukalina, M. Wavelet Analysis of Extended X-Ray Absorption Fine Structure Data. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2005**, *71*, 094110.
(140) Khodakov, A. Y.; Chu, W.; Fongarland, P. Advances in the Development of Novel Cobalt Fischer- Tropsch Catalysts for Synthesis of Long-Chain Hydrocarbons and Clean Fuels. *Chem. Rev.* **2007**, *107*, 1692−1744.
(141) Kang, J. H.; Menard, L. D.; Nuzzo, R. G.; Frenkel, A. I. Unusual non-bulk Properties in Nanoscale Materials: Thermal Metal-Metal Bond Contraction of γ-Alumina-Supported Pt Catalysts. *J. Am. Chem. Soc.* **2006**, *128*, 12068−12069.
(142) Vila, F. D.; Rehr, J. J.; Nuzzo, R. G.; Frenkel, A. I. Anomalous Structural Disorder in Supported Pt Nanoparticles. *J. Phys. Chem. Lett.* **2017**, *8*, 3284−3288.
(143) Vila, F.; Rehr, J.; Kas, J.; Nuzzo, R.; Frenkel, A. Dynamic Structure in Supported Pt Nanoclusters: Real-Time Density Functional Theory and X-Ray Spectroscopy Simulations. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *78*, 121404.
(144) Frenkel, A. I. Applications of Extended X-Ray Absorption Fine-Structure Spectroscopy to Studies of Bimetallic Nanoparticle Catalysts. *Chem. Soc. Rev.* **2012**, *41*, 8163−8178.
(145) Yuan, D.; Gong, X.; Wu, R. Atomic Configurations of Pd Atoms in PdAu (111) Bimetallic Surfaces Investigated Using The First-Principles Pseudopotential Plane Wave Approach. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2007**, *75*, 085428.
(146) Gotsis, H.; Rivalta, I.; Sicilia, E.; Russo, N. Atomic Configurations of Pd Atoms in PdAu (111) and PdAu (100) Surface Alloys: Ab Initio Density Functional Calculations. *Chem. Phys. Lett.* **2009**, *468*, 162−165.
(147) Mueller, T.; Kusne, A. G.; Ramprasad, R. Machine Learning in Materials Science: Recent Progress and Emerging Applications. *Reviews in Computational Chemistry* **2016**, *29*, 186−273.
(148) Swann, E.; Sun, B.; Cleland, D.; Barnard, A. Representing Molecular and Materials Data for Unsupervised Machine Learning. *Mol. Simul.* **2018**, *44*, 905−920.
(149) Suzuki, Y.; Hino, H.; Ueno, T.; Takeichi, Y.; Kotsugi, M.; Ono, K. Extraction of Physical Parameters from X-Ray Spectromicroscopy Data Using Machine Learning. *Microsc. Microanal.* **2018**, *24* (S2), 478−479.
(150) Chang, S.; Molleta, L.; Booth, S.; Uehara, A.; Mosselmans, J.; Ignatyev, K.; Dryfe, R.; Schroeder, S. Automated Analysis of XANES: A Feasibility Study of Au Reference Compounds. *J. Phys.: Conf. Ser.* **2016**, *712*, 012070.
(151) Ressler, T.; Wong, J.; Roos, J.; Smith, I. L. Quantitative Speciation of Mn-Bearing Particulates Emitted from Autos Burning (Methylcyclopentadienyl) Manganese Tricarbonyl-Added Gasolines Using XANES Spectroscopy. *Environ. Sci. Technol.* **2000**, *34*, 950−958.
(152) Martini, A.; Alladio, E.; Borfecchia, E. Determining Cu-Speciation in the Cu-CHA Zeolite Catalyst: The Potential of Multivariate Curve Resolution Analysis of In Situ XAS Data. *Top. Catal.* **2018**, *61*, 1396−1407.
(153) Wang, Q.; Hanson, J. C.; Frenkel, A. I. Solving the Structure of Reaction Intermediates by Time-Resolved Synchrotron X-Ray Absorption Spectroscopy. *J. Chem. Phys.* **2008**, *129*, 234502.
(154) Gamarra, D.; Fernández-García, M.; Belver, C.; Martínez-Arias, A. Operando DRIFTS and XANES Study of Deactivating Effect of CO_2 on a $Ce_{0.8}Cu_{0.2}O_2$CO-PROX Catalyst. *J. Phys. Chem. C* **2010**, *114*, 18576−18582.
(155) Fernandez-Garcia, M.; Marquez Alvarez, C.; Haller, G. XANES-TPR Study of Cu-Pd Bimetallic Catalysts: Application of Factor Analysis. *J. Phys. Chem.* **1995**, *99*, 12565−12569.
(156) Ressler, T. Application of Time-Resolved In-Situ X-Ray Absorption Spectroscopy in Solid-State Chemistry. *Anal. Bioanal. Chem.* **2003**, *376*, 584−593.
(157) Campos, A.; Lohitharn, N.; Roy, A.; Lotero, E.; Goodwin Jr, J. G.; Spivey, J. J. An activity and XANES Study of Mn-Promoted, Fe-based Fischer−Tropsch Catalysts. *Appl. Catal., A* **2010**, *375*, 12−16.
(158) Wu, L. B.; Wu, L. H.; Yang, W. M.; Frenkel, A. I. Study of the Local Structure and Oxidation State of Iron in Complex Oxide Catalysts for Propylene Ammoxidation. *Catal. Sci. Technol.* **2014**, *4*, 2512−2519.
(159) Grunwaldt, J. D.; Wagner, J. B.; Dunin-Borkowski, R. E. Imaging Catalysts at Work: A Hierarchical Approach from the Macro- To the Meso-And Nano-Scale. *ChemCatChem* **2013**, *5*, 62−80.
(160) Muñoz-Batista, M. J.; Motta Meira, D.; Colón, G.; Kubacka, A.; Fernández-García, M. Phase-Contact Engineering in Mono-and Bimetallic Cu-Ni Co-catalysts for Hydrogen Photocatalytic Materials. *Angew. Chem., Int. Ed.* **2018**, *57*, 1199−1203.
(161) Kleymenov, E.; Sa, J.; Abu-Dahrieh, J.; Rooney, D.; Van Bokhoven, J. A.; Troussard, E.; Szlachetko, J.; Safonova, O. V.; Nachtegaal, M. Structure of the Methanol Synthesis Catalyst Determined by In Situ HERFD XAS and EXAFS. *Catal. Sci. Technol.* **2012**, *2*, 373−378.
(162) Pappas, D. K.; Martini, A.; Dyballa, M.; Kvande, K.; Teketel, S.; Lomachenko, K. A.; Baran, R.; Glatzel, P.; Arstad, B.; Berlier, G.; et al. The Nuclearity of the Active Site for Methane To Methanol Conversion in Cu-Mordenite: A Quantitative Assessment. *J. Am. Chem. Soc.* **2018**, *140*, 15270−15278.
(163) Wasserman, S. R.; Allen, P. G.; Shuh, D. K.; Bucher, J. J.; Edelstein, N. M. EXAFS and principal Component Analysis: A New Shell Game. *J. Synchrotron Radiat.* **1999**, *6*, 284−286.
(164) Kim, J. Y.; Rodriguez, J. A.; Hanson, J. C.; Frenkel, A. I.; Lee, P. L. Reduction of CuO and Cu_2O with H_2: H Embedding and Kinetic Effects in the Formation of Suboxides. *J. Am. Chem. Soc.* **2003**, *125*, 10684−10692.
(165) Tauler, R. Multivariate Curve Resolution Applied To Second Order Data. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 133−146.
(166) Jaumot, J.; de Juan, A.; Tauler, R. MCR-ALS GUI 2.0: New Features and Applications. *Chemom. Intell. Lab. Syst.* **2015**, *140*, 1−12.
(167) Martini, A.; Borfecchia, E.; Lomachenko, K.; Pankin, I.; Negri, C.; Berlier, G.; Beato, P.; Falsig, H.; Bordiga, S.; Lamberti, C. Composition-Driven Cu-Speciation and Reducibility in Cu-CHA Zeolite Catalysts: A Multivariate XAS/FTIR Approach To Complexity. *Chem. Sci.* **2017**, *8*, 6836−6851.
(168) Hong, J.; Marceau, E.; Khodakov, A. Y.; Gaberová, L.; Griboval-Constant, A.; Girardon, J.-S.; Fontaine, C. L.; Briois, V. Speciation of Ruthenium As A Reduction Promoter of Silica-Supported Co Catalysts: A Time-Resolved In Situ XAS Investigation. *ACS Catal.* **2015**, *5*, 1273−1282.
(169) Tsakoumis, N. E.; Walmsley, J. C.; Rønning, M.; van Beek, W.; Rytter, E.; Holmen, A. Evaluation of Reoxidation Thresholds for γ-Al_2O_3-Supported Cobalt Catalysts Under Fischer−Tropsch Synthesis Conditions. *J. Am. Chem. Soc.* **2017**, *139*, 3706−3715.
(170) Barzan, C.; Piovano, A.; Braglia, L.; Martino, G. A.; Lamberti, C.; Bordiga, S.; Groppo, E. Ligands Make the Difference! Molecular Insights Into Cr^{vi}/SiO_2 Phillips Catalyst During Ethylene Polymerization. *J. Am. Chem. Soc.* **2017**, *139*, 17064−17073.
(171) Rochet, A. l.; Baubet, B.; Moizan, V.; Devers, E.; Hugon, A.; Pichon, C.; Payen, E.; Briois, V. r. Influence of the Preparation Conditions of Oxidic $NiMo/Al_2O_3$Catalysts on the Sulfidation Ability: A Quick-XAS and Raman Spectroscopic Study. *J. Phys. Chem. C* **2015**, *119*, 23928−23942.
(172) Shi, D.; Yang, Q.; Peterson, C.; Lamic-Humblot, A.-F.; Girardon, J.-S.; Griboval-Constant, A.; Stievano, L.; Sougrati, M. T.; Briois, V.; Bagot, P. A.; et al. Bimetallic Fe-Ni/SiO_2 Catalysts for

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211

Furfural Hydrogenation: Identification of the Interplay Between Fe and Ni During Deposition-Precipitation and Thermal Treatments. *Catal. Today* **2019**, *334*, 162−172.
(173) Nayak, C.; Bhattacharyya, D.; Jha, S. N.; Sahoo, N. K. In Situ XAS Study on Growth of PVP-Stabilized Cu Nanoparticles. *ChemistrySelect* **2018**, *3*, 7370−7377.
(174) Staniuk, M.; Zindel, D.; Van Beek, W.; Hirsch, O.; Kränzlin, N.; Niederberger, M.; Koziej, D. Matching the Organic and Inorganic Counterparts During Nucleation and Growth of Copper-Based Nanoparticles-In Situ Spectroscopic Studies. *CrystEngComm* **2015**, *17*, 6962−6971.
(175) Bazin, D.; Rehr, J. Comment on "Operando DRIFTS and XANES Study of Deactivating Effect of CO_2 on a $Ce_{0.8}Cu_{0.2}O_2$CO-PROX Catalyst. *J. Phys. Chem. C* **2011**, *115*, 23233−23236.
(176) Possato, L. G.; Acevedo, M. D.; Padró, C. L.; Briois, V.; Passos, A. R.; Pulcinelli, S. H.; Santilli, C. V.; Martins, L. Activation of Mo and V Oxides Supported on ZSM-5 Zeolite Catalysts Followed by In Situ XAS and XRD and Their Uses in Oxydehydration of Glycerol. *Molecular Catalysis* **2018**.110158
(177) Rossouw, D.; Burdet, P.; de la Peña, F.; Ducati, C.; Knappett, B. R.; Wheatley, A. E. H.; Midgley, P. A. Multicomponent Signal Unmixing from Nanoheterostructures: Overcoming the Traditional Challenges of Nanoscale X-ray Analysis via Machine Learning. *Nano Lett.* **2015**, *15*, 2716−2720.
(178) Jany, B. R.; Janas, A.; Krok, F. Retrieving the Quantitative Chemical Information at Nanoscale from Scanning Electron Microscope Energy Dispersive X-ray Measurements by Machine Learning. *Nano Lett.* **2017**, *17*, 6520−6525.
(179) Corma, A.; Serra, J.; Serna, P.; Moliner, M. Integrating High-Throughput Characterization Into Combinatorial Heterogeneous Catalysis: Unsupervised Construction of Quantitative Structure/Property Relationship Models. *J. Catal.* **2005**, *232*, 335−341.
(180) Ievlev, A. V.; Belianinov, A.; Jesse, S.; Allison, D. P.; Doktycz, M. J.; Retterer, S. T.; Kalinin, S. V.; Ovchinnikova, O. S. Automated Interpretation and Extraction of Topographic Information from Time of Flight Secondary Ion Mass Spectrometry Data. *Sci. Rep.* **2017**, *7*, 17099.
(181) Lerotic, M.; Jacobsen, C.; Schäfer, T.; Vogt, S. Cluster Analysis of Soft X-Ray Spectromicroscopy Data. *Ultramicroscopy* **2004**, *100*, 35−57.
(182) Lerotic, M.; Jacobsen, C.; Gillow, J.; Francis, A.; Wirick, S.; Vogt, S.; Maser, J. Cluster Analysis in Soft X-Ray Spectromicroscopy: Finding The Patterns in Complex Specimens. *J. Electron Spectrosc. Relat. Phenom.* **2005**, *144*, 1137−1143.
(183) Lerotic, M.; Mak, R.; Wirick, S.; Meirer, F.; Jacobsen, C. MANTiS: a Program for the Analysis of X-Ray Spectromicroscopy Data. *J. Synchrotron Radiat.* **2014**, *21*, 1206−1212.
(184) Neu, T. R.; Manz, B.; Volke, F.; Dynes, J. J.; Hitchcock, A. P.; Lawrence, J. R. Advanced Imaging Techniques for Assessment of Structure, Composition and Function in Biofilm Systems. *FEMS Microbiol. Ecol.* **2010**, *72*, 1−21.
(185) Liu, Y.; Meirer, F.; Williams, P. A.; Wang, J.; Andrews, J. C.; Pianetta, P. TXM-Wizard: a Program for Advanced Data Collection and Evaluation in Full-Field Transmission X-Ray Microscopy. *J. Synchrotron Radiat.* **2012**, *19*, 281−287.
(186) Duan, X.; Yang, F.; Antono, E.; Yang, W.; Pianetta, P.; Ermon, S.; Mehta, A.; Liu, Y. Unsupervised Data Mining in Nanoscale X-Ray Spectro-Microscopic Study of NdFeB Magnet. *Sci. Rep.* **2016**, *6*, 34406.
(187) Agarwal, P. K.; Procopiuc, C. M. Exact and Approximation Algorithms for Clustering. *Algorithmica* **2002**, *33*, 201−226.
(188) Lloyd, S. Least Squares Quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129−137.
(189) Brinza, L.; Schofield, P. F.; Hodson, M. E.; Weller, S.; Ignatyev, K.; Geraki, K.; Quinn, P. D.; Mosselmans, J. F. W. Combining μXANES and μXRD Mapping To Analyse the Heterogeneity in Calcium Carbonate Granules Excreted by the Earthworm Lumbricus Terrestris. *J. Synchrotron Radiat.* **2014**, *21*, 235−241.
(190) Morrell, A. P.; Floyd, H.; Mosselmans, J. F. W.; Grover, L. M.; Castillo-Michel, H.; Davis, E.; Parker, J. E.; Martin, R. A.; Addison, O. Improving our Understanding of Metal Implant Failures: Multiscale Chemical Imaging of Exogenous Metals in Ex-Vivo Biological Tissues. *Acta biomaterialia* **2019**, DOI: 10.1016/j.actbio.2019.05.071.
(191) Boesenberg, U.; Meirer, F.; Liu, Y.; Shukla, A. K.; Dell'Anna, R.; Tyliszczak, T.; Chen, G.; Andrews, J. C.; Richardson, T. J.; Kostecki, R.; Cabana, J. Mesoscale Phase Distribution in Single Particles of $LiFePO_4$ Following Lithium Deintercalation. *Chem. Mater.* **2013**, *25*, 1664−1672.
(192) Schmidt, J. E.; Ye, X.; van Ravenhorst, I. K.; Oord, R.; Shapiro, D. A.; Yu, Y. S.; Bare, S. R.; Meirer, F.; Poplawsky, J. D.; Weckhuysen, B. M. Probing the Location and Speciation of Elements in Zeolites with Correlated Atom Probe Tomography and Scanning Transmission X-Ray Microscopy. *ChemCatChem* **2019**, *11*, 488−494.
(193) Beale, A. M.; Jacques, S. D.; Di Michiel, M.; Mosselmans, J. F. W.; Price, S. W.; Senecal, P.; Vamvakeros, A.; Paterson, J. X-Ray Physico-Chemical Imaging During Activation of Cobalt-Based Fischer−Tropsch Synthesis Catalysts. *Philos. Trans. R. Soc., A* **2018**, *376*, 20170057.
(194) Price, S.; Ignatyev, K.; Geraki, K.; Basham, M.; Filik, J.; Vo, N.; Witte, P.; Beale, A.; Mosselmans, J. Chemical imaging of Single Catalyst Particles with Scanning μ-XANES-CT and μ-XRF-CT. *Phys. Chem. Chem. Phys.* **2015**, *17*, 521−529.
(195) Caruthers, J. M.; Lauterbach, J. A.; Thomson, K. T.; Venkatasubramanian, V.; Snively, C. M.; Bhan, A.; Katare, S.; Oskarsdottir, G. Catalyst Design: Knowledge Extraction from High-Throughput Experimentation. *J. Catal.* **2003**, *216*, 98−109.
(196) Venkatasubramanian, V.; Chan, K.; Caruthers, J. M. Computer-Aided Molecular Design Using Genetic Algorithms. *Comput. Chem. Eng.* **1994**, *18*, 833−844.
(197) Venkatasubramanian, V. The Promise of Artificial Intelligence in Chemical Engineering: Is It Here, Finally? *AIChE J.* **2019**, *65*, 466−478.
(198) Ouyang, R.; Curtarolo, S.; Ahmetcik, E.; Scheffler, M.; Ghiringhelli, L. M. SISSO: A Compressed-Sensing Method for Identifying the Best Low-Dimensional Descriptor in an Immensity ofoffered Candidates. *Phys. Rev. Materials* **2018**, *2*, 083802.
(199) Katare, S.; Caruthers, J. M.; Delgass, W. N.; Venkatasubramanian, V. Intelligent System for Reaction Kinetic Modeling and Catalyst Design. *Ind. Eng. Chem. Res.* **2004**, *43*, 3484−3512.
(200) Seidler, G. T.; Mortensen, D. R.; Remesnik, A. J.; Pacold, J. I.; Ball, N. A.; Barry, N.; Styczinski, M.; Hoidn, O. R. A Laboratory-Based Hard X-Ray Monochromator for High-Resolution X-Ray Emission Spectroscopy and X-Ray Absorption Near Edge Structure Measurements. *Rev. Sci. Instrum.* **2014**, *85*, 113906.

DOI: 10.1021/acscatal.9b03599
ACS Catal. 2019, 9, 10192−10211