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

pubs.acs.org/JPCC

Article

How Much Structural Information Could Be Extracted from XANES Spectra for Palladium Hydride and Carbide Nanoparticles

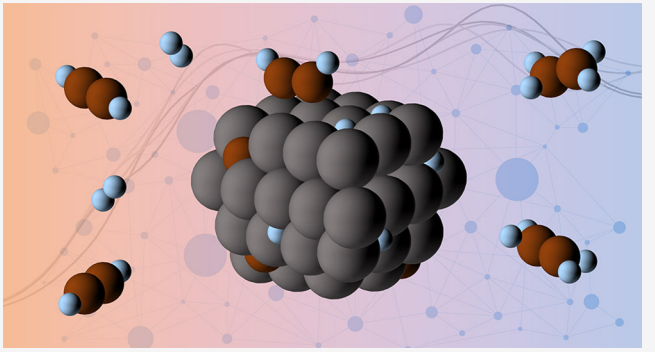
Oleg A. Usoltsev,* Aram L. Bugaev,* Alexander A. Guda, Sergey A. Guda, and Alexander V. Soldatov

Cite This: *J. Phys. Chem. C* 2022, 126, 4921−4928

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: X-ray absorption near-edge structure (XANES) spectroscopy is a powerful characterization technique that is sensitive to both three-dimensional (3D) geometry and the electronic state of the selected element. In this work, we have suggested a set of structural descriptors that can be used to characterize the state of palladium nanoparticles in hydrogenation reactions and explored the possibility of their extraction from Pd K-edge XANES spectra. A theoretical spectral database was calculated for palladium atoms in the bulk and at the (111) surface with variable Pd−Pd interatomic distances. Carbon and hydrogen atoms randomly occupied octahedral interstitial sites for different H/Pd and C/Pd ratios. The presence of hydrogen and hydrocarbon molecules adsorbed at the surface was also considered. The obtained spectral database was subjected to the principal component analysis (PCA) to estimate the number of strongly contributing components and the multivariate curve resolution (MCR) approach to deconvolve the whole set of data into the XANES spectra of "pure" species and their concentrations. The latter were also used as descriptors of spectra, and machine learning (ML) algorithms were then trained to predict them based on the descriptors of structure and vice versa. We have shown that some of the structural parameters, namely, the concentration of surface-adsorbed molecules, have minor effects on the spectra and cannot be predicted. For interatomic distances, their averaged value can be extracted with good prediction quality based on only one MCR concentration, while independent prediction of the distances in the bulk or at the surface gives unsatisfactory results. Finally, we constructed a new set of structural descriptors that have direct relevance to the MCR components.

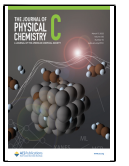


INTRODUCTION

X-ray absorption near-edge structure (XANES) spectroscopy is a powerful characterization technique that is sensitive to both three-dimensional (3D) geometry and the electronic state of the selected element. Since there is no universal method for the extraction of structural information from XANES, numerous approaches have been suggested, including those based on the application of machine learning (ML) algorithms.[1−9] One of the first steps to describe the local atomic environment was suggested by Carbone et al.,[4] who used the FEFF[10] code to calculate the data set for different types of coordination (symmetry) of eight (Ti, V, Cr, Mn, Fe, Co, Ni, and Cu) transition 3d metals and applied supervised ML for classification task, where classifiers were the type of absorbing atom and the type of the first shell coordination. The classification problem was solved with an accuracy of more than 86%; however, the description of experimental (even reference) spectra had several inaccuracies, primarily related to the choice of the method for calculating theoretical spectra. Kiyohara et al. made an attempt to describe the oxidation state of a wide range of elements (46 oxide configurations) based on the ELNES/XANES spectroscopy data of oxygen K-edge.[5] They used hierarchical clustering and decision tree methods choosing classifiers as a possible set of oxide geometries and a database of more than 300 000 theoretical spectra calculated within one-particle density functional theory-generalized gradient approximation (DFT-GGA) in CASTEP software.[11] It should be noted that the work of Timoshenko et al. made great advances in predicting the coordination number and particle shape in metal clusters[1−3] by XANES spectroscopy using FEFF[10] and FDMNES[12,13] for spectra calculations and neural networks as a basic prediction algorithm. In more recent works,[14] the authors used neural networks for the prediction of the coordination numbers and interatomic distances in small Pd and PdH clusters and active sites in bimetallic CuPd clusters.[15]

The application of an ML procedure requires a training set. Each spectrum in the training data set can be characterized by

Received: October 30, 2021
Revised: January 21, 2022
Published: March 7, 2022



https://doi.org/10.1021/acs.jpcc.1c09420
J. Phys. Chem. C 2022, 126, 4921−4928

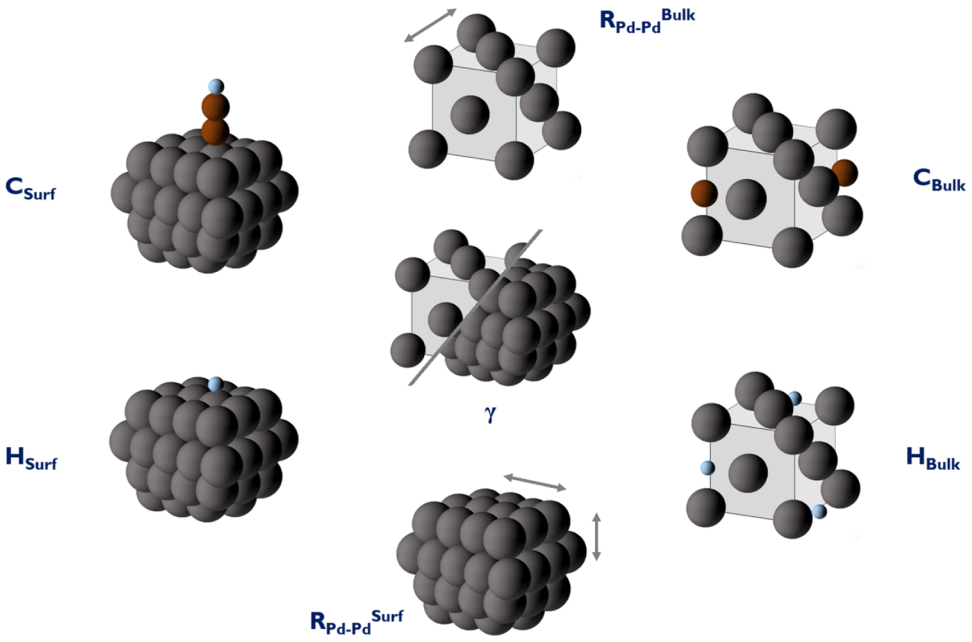


Figure 1. Visual illustration of the structural descriptors.

a set of structural descriptors (i.e., bond distances, bond angles, atomic coordinates, types of ligands, etc.), which can be then predicted by the trained algorithm. It should be noted that in most of the cases, the user is not limited in the choice of structural descriptors, which vary during the calculation of the theoretical training set, expecting to further extract them from the experimental spectra. Some of these descriptors, however, may have the least importance for the shaping of XANES. In particular, in a number of previous reports,[16−19] we demonstrated that both bulk and surface properties of palladium nanoparticles (NPs), such as the formation of hydrides and carbides in the bulk and adsorption of the reactive molecules at the surface, induce changes in XANES, but the amount of information that can be unambiguously extracted from the spectral data has to be examined. We investigate a possible dimensions reduction by the principal component analysis (PCA) and link the pure components to the structural descriptors. In the work of Tetef et al.,[20] unsupervised ML was applied to a database of calculated molecules to investigate better the performance of Variational AutoEncoder, which is alternative to the PCA. However, this method involves a more complicate procedure of understanding parametric space and its correlation with structural descriptors.

In this work, we propose an original approach exemplified on Pd K-edge XANES spectra for palladium NPs, in which a theoretical database is subjected to multivariate curve resolution (MCR) analysis to obtain a set of independent components, whose concentrations are used as the descriptors of spectra. An Extra Trees algorithm implemented in the PyFitIt code[21] was used to establish the correlations between the MCR concentrations and descriptors of the structure. Based on these correlations, a new set of structural descriptors was constructed whose parameters can be predicted using only one MCR component.

METHODS

The database of theoretical XANES spectra included in the training set was calculated within the finite difference method released in the FDMNES code.[22−26] The calculation of one spectrum in the optimized regime on six-core Intel Original Core i7 X6 5930K @3.5 GHz (Haswell) processors took around 1 h. To choose the convolution parameters, the reference spectrum of bulk palladium foil was used. Then, the convolution parameters were fixed for all theoretical spectra. The adaptive energy range was used with 1 eV step in the [−100, −15] eV range, 0.1 eV step in the [−15, 18] eV range, and 2 eV step in the [18, 200] eV region. Then, after optimizing the convolution parameters, only the [−10, 60] eV range interpolated to a 1 eV step was utilized for the ML analysis. The testing of different ML algorithms and principal component analysis (PCA) was performed in the PyFitit code.[21] The Extra trees method was used. The algorithm divided the space of geometric parameters p into non-intersecting rectangles; in each of them, the objective function is approximated by a linear function using the least-squares method. Each node of the decision tree contained a condition $p_j < t$ for one of the geometrical parameters $_{pi}$, which divided the training subset into parts. Initially, the overall training set was randomly divided into training and validation subsets that were used to train and evaluate the prediction quality of the tree, respectively. Each leaf of the tree contained linear approximations of XANES, which were constructed based on the training subset for a given leaf. Thus, a single tree was a form of specifying a piecewise linear function of geometric parameters p_1, ..., p_k.

The multivariate curve resolution (MCR) procedure was performed by means of the *pymcr* library.[27] In this approach, the whole data set initially represented by matrix $\mathbf{D}$ $(m \times n)$, where m is the number of spectra and n is the number of energy points, was decomposed as $\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E}$, where $\mathbf{C}$ $(m \times k)$ is the concentration profiles of k pure components from matrix $\mathbf{S}(n \times k)$ and $\mathbf{E}(m \times n)$ is the error matrix. The MCR analysis was performed fixing the sum of the concentrations to

https://doi.org/10.1021/acs.jpcc.1c09420
J. Phys. Chem. C 2022, 126, 4921−4928

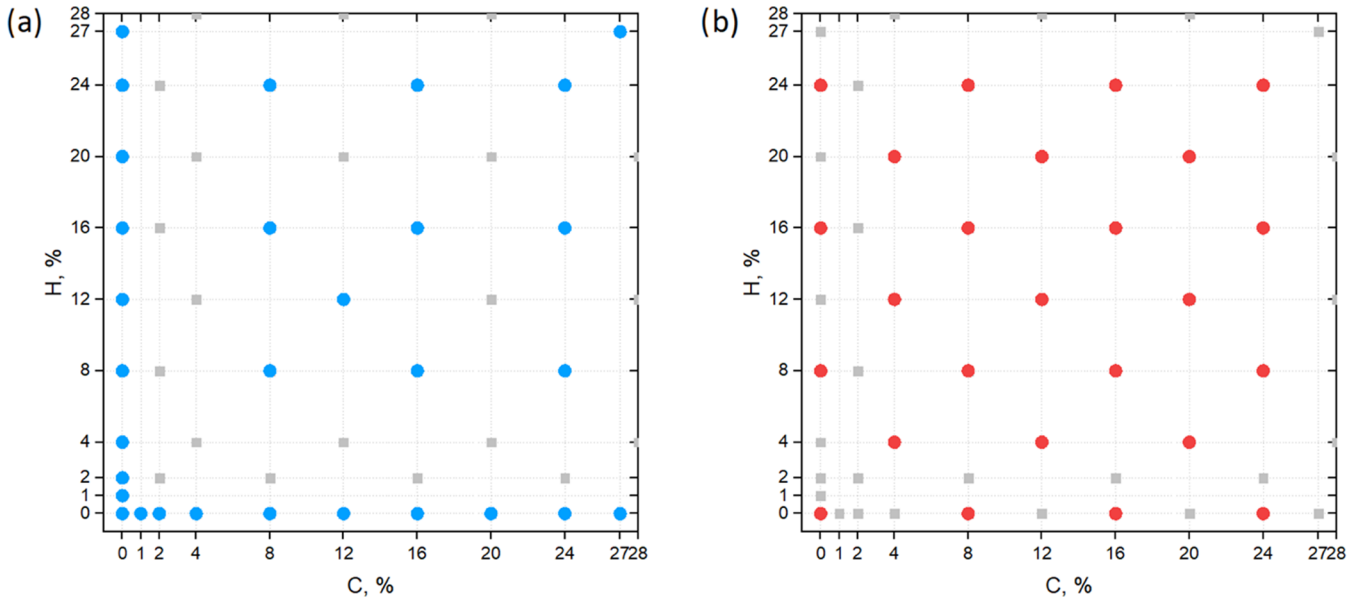


Figure 2. Grids in two-dimensional parameter space for bulk (a) and surface (b) for which the theoretical spectra were obtained for each value of Pd−Pd interatomic distance. Gray squares indicate the direction for the possible augmentation of the database.

1 and using the non-negativity constrain. The cross-validation (CV) procedure was applied to evaluate the quality of prediction. The whole set of spectra was randomly separated into 10 parts and each of the parts were subsequently used as a validation set, while the other 90% were used as a training set. Prediction quality was calculated based on CV R^2-score (see eq 1) multiplied by 100%.

$$R^2 \text{ score} = 1 - \frac{\sum \|X_i - F_i\|^2}{\sum \|X_i - M\|^2} \quad (1)$$

where $\|..\|^2$ corresponds to the L_2 norm of a spectrum, the numerator corresponds to the difference between the actual, X, and predicted, F, value, and the denominator corresponds to the difference between the actual value and the averaged value over the whole training set.

RESULTS AND DISCUSSION

The initial choice of structural descriptors (i.e., the key features that affect the spectral shape) (see Figure 1) was conditioned by the expected structural changes and phases of palladium, which were expected to be formed during hydrogenation reaction, aiming to obtain additional information about the processes that occur in the nanoparticles from XANES data.[16−19] We introduced R_{surf} and R_{bulk} as the first shell Pd−Pd interatomic distances in the surface and bulk, respectively; C_{surf} and C_{bulk} as the relative amount of adsorbed hydrocarbon molecules at the surface and the number of carbon impurities in the bulk, respectively; H_{surf} and H_{bulk} as the respective hydrogen concentrations. The database consisted of two parts, which we will refer to as surface and bulk (examples of the used geometries are shown in Figure S17). XANES spectra were calculated by averaging over all nonequivalent positions for bulk and by averaging over all surface atoms for surface model. The spectra for bulk and surface were then merged using random values of the parameter γ as a fraction of the surface. The interatomic distances grid selection is the most trivial part due to the simple variation of the unit cell volume. Therefore, for each pair of C and H concentrations, 10 different values of R_{Pd-Pd} were taken in [2.72−2.84] Å range and carbon and hydrogen concentrations were varied from 0 to 25% for each R_{Pd-Pd}.

It should be noted that only the (111) surface was used, and the model can therefore account for almost any size of palladium clusters except for the ultrasmall ones, in which the average coordination number is below 9. In the previous work,[8] we have also shown that the theoretical spectra of the palladium nanoparticle calculated by averaging over all inequivalent positions can be approximated by a linear combination of the spectra for "bulk" and "surface" atoms.

The resulting database can be visualized in carbon/hydrogen concentration space as shown in Figure 2. The scatters in the figure correspond to the points for which the averaged spectrum (over all inequivalent positions in the structure) was obtained. It should be noted that it is possible to use random or IHS sampling, but symmetric considerations for specific H/Pd and C/Pd ratios allowed us to reduce the number of calculated spectra. Considering, in addition, that such a grid was constructed for different interatomic distances R_{Pd-Pd} for both bulk (R_{bulk}) and surface (R_{surf}), the data set consisted of over 500 independent spectra (after averaging). The data set of initial spectra (before averaging), which included the calculation of spectra for every nonequivalent atom from the 108 atoms for bulk and 24 atoms for surface for every combination of structural descriptors, contained about 10 000 entries. The number of atoms for the bulk and surface models were chosen to provide a wide range of possible C/Pd and H/Pd ratios (e.g., for a supercell of 108 Pd atoms one can start with the C/Pd concentration below 1%). Finally, the spectra of bulk and surface were mixed together, as a weighted average using γ, which determines the fraction of the surface.

The resulting database was subject to MCR analysis (vide infra).[27−29] This is an efficient chemometric approach that was successfully applied to extract both the spectra of the pure species of the multicomponent systems and their concentration profiles as a function of time, temperature, or gas atmosphere from in situ and operando XAS data.[30,31] The time evolution of the multicomponent system with different behaviors of

https://doi.org/10.1021/acs.jpcc.1c09420
J. Phys. Chem. C 2022, 126, 4921−4928

independent components is the key prerequisite for a successful MCR analysis. If, for example, the two species always evolve in a similar way or if they correspond to similar XAS spectra, they cannot not be separated. The idea behind the present study is beyond the above standard application of the MCR technique. Instead, the subject of the analysis was a theoretical data set of different palladium clusters, in which the known structural parameters were taken instead of variable external conditions in the case of experimental data sets. For each of the local atomic and electronic states considered in the theoretical model, one can expect to find the corresponding MCR component. However, analogous to the experimental data, if the effect on the shaping of XANES is similar, different types of Pd species are indistinguishable. Therefore, the analysis performed below allows us to determine what kind and how many different species or phases one can extract from XANES data sets as independent species.

As a preliminary step for the MCR procedure, the number of independent components was estimated by the PCA applied to determine the number of independent components in the theoretical data set. The Scree plot (Figure 3) has two "elbow"

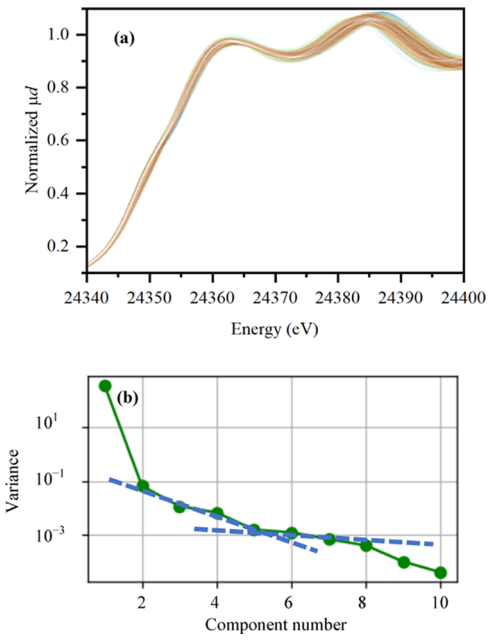


Figure 3. (a) All XANES spectra from the training set and (b) variance described by each component (scree plot) obtained from PCA.

points that can be considered as the number of independent components, 2 and 5. However, if only two components are used, they are responsible only for the amount of γ (i.e., the particle size) and interatomic distances (see Table S2), while it is known from our previous experimental studies that the formation of hydrides and carbides has also a prominent effect on XANES.[8,32−38]

Therefore, based on PCA results, we performed MCR analysis by assuming the presence of five independent species and obtained their concentrations for each of the spectra in the initial database. The MCR concentrations obtained by projecting these spectra on the intermediate database (where bulk and surface were separated) allowed us to perform a visual qualitative attribution of these components to some of the structural parameters (Figure 4). The analysis presented below was also performed for the MCR concentrations obtained by assuming different numbers of pure components (from two to six) and confirmed the correctness of the choice of the five components in terms of the maximal structural information that could be extracted (vide infra).

To perform a more careful correlation between the obtained MCR components and the structural descriptors, the concentrations of MCR components were regarded as the descriptors of spectra and their prediction was attempted using ML algorithms. The CV procedure was used to quantify the prediction quality of each component based on different combinations of structural descriptors. The prediction quality can be visually appreciated by plotting the predicted vs real MCR concentrations (see Figures S1−S5). We have selected the descriptors for the theoretical data set and trained the Extra Trees algorithm. The Extra Trees method does not suffer from overtraining and works well even for a small number of entries in the training set. The optimal number of randomized trees was adjusted to 70. Increasing the number of trees did not improve the quality.

If the number of structural descriptors used for prediction is equal to two, the prediction results can be visualized in 2D maps for all possible pairs of descriptors (see Figure 5, for the 1st component and Figures S6−S9 for other components). Background color in the figures refers to the predicted concentration of the MCR component for the corresponding pair of structural parameters. The outer part of the circles is colored by the predicted value (from CV), while the real value is shown in the centers of the circles.

For every MCR component, the ML algorithm was trained to predict its concentration by every single descriptor, every possible pair of descriptors, every triple of descriptors, etc. Then, the best combinations were selected for each case. In Table 1, we sorted the prediction results for each MCR concentration by the prediction quality and selected the best sets of n descriptors ($n = \{1,2,3,4,5,6\}$), increasing n until the prediction quality above the user-defined value of 80% was achieved. Analyzing the prediction quality and the set of structural descriptors it was based on, we could conclude which type of species each MCR component corresponded to.

Component 1 depends on C_{bulk} and γ. If C_{bulk} is 0, this component does not depend on γ and its concentration is close to 0 in the whole range of γ. If C_{bulk} is high, its concentration decreases with γ. Finally, if $\gamma = 1$ (i.e., we have only the surface), the component concentration almost does not depend on C_{bulk} (and is close to 0). In the whole range of γ, the concentration of the first MCR component increases with C_{bulk}. Therefore, we can attribute this component to a bulk carbide: if there is no carbide, this concentration is 0, irrespective of particle size; and if the carbide is formed in the bulk, its concentration is bigger for bigger particles.

The concentration of the second component increases with increase in both R_{bulk} and R_{surf}. Also, it also depends on γ: higher the γ, stronger is the dependence on R_{surf}; therefore, decreasing γ increases the dependence on R_{bulk}. Therefore, this component is responsible for the average particle size.

Component 3 is mainly dependent on γ, having higher concentrations for big values of γ. In addition, for big γ (i.e., for small particles), the dependence on R_{surf} also exists: the concentration decreases for big values of interatomic distance. Therefore, this component is attributed to the contribution of small particles (or the surface of the particles).

https://doi.org/10.1021/acs.jpcc.1c09420
J. Phys. Chem. C 2022, 126, 4921−4928

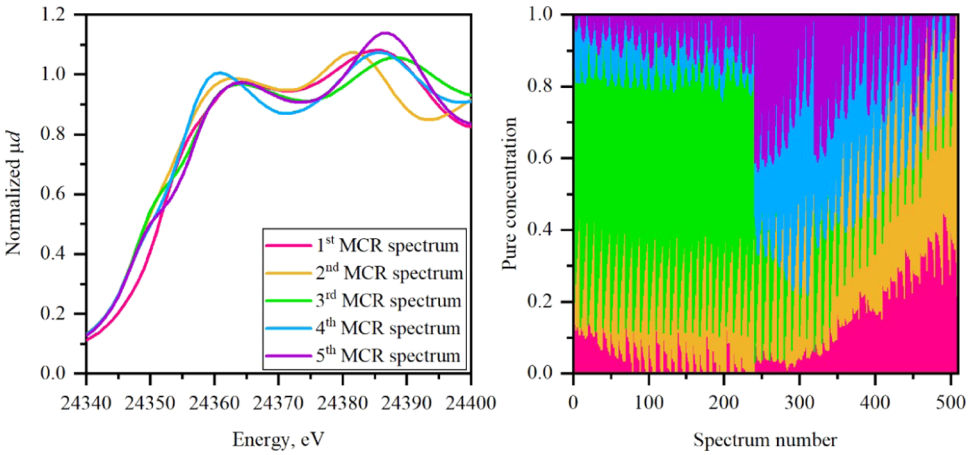


Figure 4. Five spectra determined by the MCR analysis, and the concentration profile obtained by fitting with these spectra the intermediate spectral database (where bulk and surface were separated). The ordering of the data set is as follows: first half of all spectra is the surface part and the second is the bulk; carbon concentration increases within each part from the beginning (0%) to the end (25%); hydrogen concentration increases from 0 to 25% within every 50−60 spectra; and interatomic distances increase from 2.72 to 2.84 within every 10 spectra.

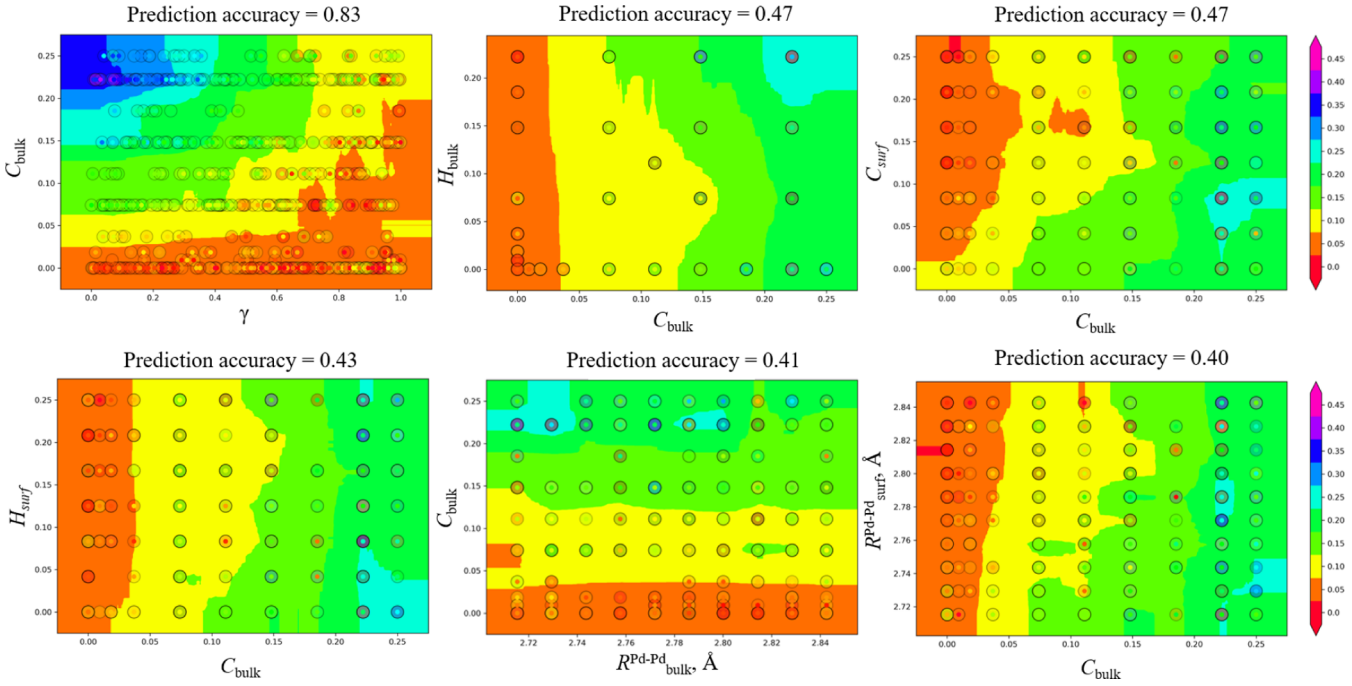


Figure 5. Prediction of the first MCR component by pair of structural parameters sorted by prediction accuracy. Scatters correspond to the real values with the outer shell of each circle colored by the ML-predicted value (during CV procedure) and the center of each circle colored by the real value. The background color corresponds to the result of ML prediction. Similar plots for other MCR concentrations are reported in Figures S6−S9.

Component 4 depends on C_{bulk} and γ; however, unlike that of the first component, its concentration decreases with increasing C_{bulk}, and H_{bulk} should be also included as a descriptor to improve the prediction quality. The concentration increases with increasing H_{bulk} and the dependence is stronger for small γ. Thus, this component corresponds to a bulk palladium hydride.

The concentration of the fifth component decreases with both C_{bulk} and H_{bulk} and, in addition, increases for small R_{bulk}. Obviously, since all of the above structural descriptors are related to the bulk, the dependence is stronger for small γ. In contrast to the third component, component 5 represents big metallic particles (i.e., bulk part of the particles without carbon or hydrogen impurities).

To further prove the made assignment, we also tried to correlate the differences in the shape of MCR components with some known experimental spectra. Clearly, not every species may have the exact experimental reference, especially in the case of nanoparticles, but for MCR components 3−5, the visual differences are in agreement with those observed in experimental data for small palladium nanoparticles, palladium hydride particles, and bulk palladium foil (Figure S16).

Although the initial theoretical model was constructed with a set of seven well-defined structural parameters, two of them, C_{surf} and H_{surf}, did not occur in any of the combinations that determine the concentration of MCR components. This means that their effect on the shaping of XANES spectra is negligible in comparison with other structural parameters. Most of the

https://doi.org/10.1021/acs.jpcc.1c09420
J. Phys. Chem. C 2022, 126, 4921−4928

Table 1. Prediction Quality for the Five MCR Components Using Different Combinations of Structural Descriptors

MCR component	combination of descriptors	prediction quality (%)	interpretation
1	C_{bulk}	41	bulk palladium carbide particles
	$C_{bulk} + \gamma$	83	
2	R_{bulk}	30	particles with big interatomic distances
	$\gamma + R_{bulk}$	51	
	$\gamma + R_{bulk} + R_{surf}$	76	
	$\gamma + R_{bulk} + R_{surf} + C_{bulk}$	91	
3	γ	71	small particles with small averaged interatomic distances
	$\gamma + R_{surf}$	90	
4	C_{bulk}	34	bulk palladium hydride (substituting the carbide)
	$C_{bulk} + \gamma$	70	
	$C_{bulk} + \gamma + H_{bulk}$	84	
5	C_{bulk}	21	bulk metallic particles with small interatomic distances
	$C_{bulk} + \gamma$	44	
	$C_{bulk} + \gamma + R_{bulk}$	65	
	$C_{bulk} + \gamma + R_{bulk} + H_{bulk}$	81	

other parameters should be applied in combinations with each other to establish a correlation between their values and the concentrations of MCR components. This will also complicate the solution of the inverse problem of extraction of these structural parameters from the XANES spectra. Indeed, as can be seen from Table S1, for most of the structural descriptors, except γ, the prediction qualities hardly exceed 50% even if multiple MCR concentrations are used as the descriptors of spectra. For example, although $R_{bulk} + R_{surf}$ affects many MCR components in Table 1, the quality of their prediction is poor even if all five MCR concentrations are used.

To overcome the above issue, we suggest a set of novel structural descriptors that are based on those initially used to construct the spectral database but can be predicted with better quality with the least number of MCR concentrations (see Table 2). The new structural descriptors still possess

Table 2. Prediction Quality for the Structural Descriptors Using Different Combinations of the Concentrations of MCR Components

descriptor	combination of MCR components	prediction quality (%)
γ	MCR_3	77
$\langle R \rangle$	MCR_2	73
M_{bulk}	MCR_5	71
$(1-\gamma)\cdot C_{bulk}$	MCR_1	85
$(H/C)_{bulk}$	MCR_4	77

intuitive interpretation. The first one is the average interatomic distance that mainly depends on the second MCR concentration (Figure S10). It is calculated considering the fraction of surface, γ, and different coordination numbers, N_{bulk} and N_{surf}, in the bulk and at the surface, respectively

$$\langle R \rangle = \frac{N_{bulk}\cdot(1-\gamma)\cdot R_{bulk} + N_{surf}\cdot\gamma\cdot R_{surf}}{N_{bulk}\cdot(1-\gamma) + N_{surf}\cdot\gamma} \quad (2)$$

The next descriptor is the degree of bulk metallic palladium, which is constructed as

$$M_{bulk} = 2\cdot(0.5 - C_{bulk} - H_{bulk})\cdot(1-\gamma)\cdot\frac{2.84 - R_{bulk}}{2.84 - 2.72} \quad (3)$$

The value of M_{bulk} is maximal for bulk metallic particles with small interatomic distances and decreases with decreasing particle size (increase of γ), increasing interatomic distances, or addition of hydrogen or carbon. It mainly depends on the concentration of the fifth component (Figure S11).

Finally, the relative amount of hydrogen with respect to carbon in the bulk of the particle was defined as

$$(H/C)_{bulk} = \frac{H_{bulk}}{C_{bulk}}\cdot(1-\gamma) \quad (4)$$

and characterized by the concentration of the fourth component (see Figure S12).

In addition, since the C_{bulk} descriptor is relevant only to the bulk of the particle, we used a product $(1-\gamma)\cdot C_{bulk}$ as a new descriptor, which is well correlated with the first MCR component.

It is important to note that the absence of C_{surf} and H_{surf} does necessarily mean that they do not affect XANES spectra, which is shown by us in several experimental studies.[8,35,36,39] The main reasons for such a result is that (i) the changes in the XANES spectra due to adsorption of one hydrogen or hydrocarbon molecule are smaller than surrounding a bulk palladium atom by H or C impurities and (ii) the changes produced by surface and bulk Pd−C bonds are similar, in agreement with previous reports; therefore, there is no component (starting from PCA) that can discriminate bulk phase from surface-adsorbed molecules.

Also, the possibility of extracting the average interatomic distance only is highlighted. In the EXAFS analysis, the limitation in discriminating small structural changes (i.e., distortions of structural parameters) around the absorbing centers of the same type is known, and the minimal ΔR that can be resolved can be calculated theoretically.[40] In XANES, there is no such equation; therefore, the procedure suggested in this work provides a useful tool to examine the possibility of the extraction of different structural parameters.

CONCLUSIONS

We have suggested a new strategy for selecting MCR concentration as spectral descriptors for ML algorithms. The application of such an approach for the theoretical spectra allows us to examine how many and which structural parameters can be potentially extracted from the XANES spectra of the system of interest. This is specifically important for the systems in which numerous structural changes can be expected, but their significance for XANES shaping is not evident (e.g., dynamic structure of the catalysts under reaction conditions). The proposed approach was successfully exemplified on Pd K-edge XANES spectra of the palladium nanoparticles and might be potentially extended to any system. The suggested PCA + MCR + ML approach can be used to evaluate the set of structural parameters with the greatest impact on the data. However, the choice of all possible structural deformation and the derivation of new descriptors that can be predicted with better quality definitely differs for every system and also depends on the user choice. Especially, the construction of the equation such as eqs 2−4 cannot be performed automatically at the moment. Here, the theoretical database was constructed to account for the changes in the Pd−Pd distances as well as for the formation of Pd−C and Pd−H bonds in the bulk and at the surface of the particles. For such database, a new set of structural descriptors, which have

https://doi.org/10.1021/acs.jpcc.1c09420
J. Phys. Chem. C 2022, 126, 4921−4928

direct relevance to the determined MCR components, was constructed. It was shown that the particle size can be determined directly based on the concentration of one of the MCR components. For other structural parameters, their combinations, such as average interatomic distance or relative bulk hydrogen concentration with respect to bulk carbon concentration, or the amount of bulk metallic phase, can be extracted. The surface-adsorbed molecules were not extracted first due to their smaller effect on XANES spectra with respect to the modifications induced by bulk insertion of H and C impurities and second because K-edge XANES spectra cannot distinguish whether the Pd−C or Pd−H bonds are formed in the bulk or at the surface of the particle.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcc.1c09420.

Predicted values of the each MCR concentration for the best pair of structural descriptors. Prediction map of each MCR component by a pair of structural parameters sorted by prediction accuracy. Prediction of MCR concentrations obtained with the different numbers of the pure components. Comparison of MCR components with experimental spectra. The geometry models of bulk and surface (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Oleg A. Usoltsev − *The Smart Materials Research Institute, Southern Federal University, 344090 Rostov-on-Don, Russia;* orcid.org/0000-0002-1537-3497; Email: ousolcev@sfedu.ru

Aram L. Bugaev − *The Smart Materials Research Institute, Southern Federal University, 344090 Rostov-on-Don, Russia;* orcid.org/0000-0001-8273-2560; Email: abugaev@sfedu.ru

Authors

Alexander A. Guda − *The Smart Materials Research Institute, Southern Federal University, 344090 Rostov-on-Don, Russia;* orcid.org/0000-0002-6941-4987

Sergey A. Guda − *The Smart Materials Research Institute, Southern Federal University, 344090 Rostov-on-Don, Russia;* orcid.org/0000-0002-2398-1847

Alexander V. Soldatov − *The Smart Materials Research Institute, Southern Federal University, 344090 Rostov-on-Don, Russia*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcc.1c09420

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The research was funded by RFBR Grant No. 20-32-90083 (theoretical XANES calculation, MCR analysis, ML). The measurement of the experimental spectra was supported by the Presidents Grant MK-5853.2021.1.2 (Agreement No. 075-15-2021-499). The authors acknowledge the Center for Collective Use "Nanoscale structure of matter" for providing the resources of the Blokhin cluster for calculating the spectral database.

■ REFERENCES

(1) 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.

(2) 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.

(3) Timoshenko, J.; Roese, S.; Hövel, H.; Frenkel, A. I. Silver clusters shape determination from in-situ XANES data. *Radiat. Phys. Chem.* **2018**, *175*, No. 108049.

(4) Carbone, M. R.; Yoo, S.; Topsakal, M.; Lu, D. Classification of local chemical environments from x-ray absorption spectra using supervised machine learning. *Phys. Rev. Mater.* **2019**, *3*, No. 033604.

(5) 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*, No. 13548.

(6) 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.; Groppo, E.; Piovano, A.; Borfecchia, E.; Lamberti, C. 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.

(7) Martini, A.; Guda, S. A.; Guda, A. A.; Smolentsev, G.; Algasov, A.; Usoltsev, O.; Soldatov, M. A.; Bugaev, A.; Rusalev, Y.; Lamberti, C.; Soldatov, A. V. PyFitit: the software for quantitative analysis of XANES spectra using machine-learning algorithms. *Comput. Phys. Commun.* **2020**, *250*, No. 107064.

(8) Usoltsev, O. A.; Bugaev, A. L.; Guda, A. A.; Guda, S. A.; Soldatov, A. V. Absorption of Hydrocarbons on Palladium Catalysts: From Simple Models Towards Machine Learning Analysis of X-ray Absorption Spectroscopy Data. *Top. Catal.* **2020**, *63*, 58−65.

(9) Martini, A.; Bugaev, A. L.; Guda, S. A.; Guda, A. A.; Priola, E.; Borfecchia, E.; Smolders, S.; Janssens, K.; De Vos, D.; Soldatov, A. V. Revisiting the extended X-ray absorption fine structure fitting procedure through a machine learning-based approach. *J. Phys. Chem. A* **2021**, *125*, 7080−7091.

(10) Rehr, J. J.; Albers, R. C. Theoretical approaches to x-ray absorption fine structure. *Rev. Mod. Phys.* **2000**, *72*, 621−654.

(11) Clark, S. J.; Segall, M. D.; Pickard, C. J.; Hasnip, P. J.; Probert, M. I. J.; Refson, K.; PayneII, M. C. First principles methods using CASTEP. *Z. Kristallogr.* **2005**, *220*, 567−570.

(12) Bunău, O.; Joly, Y. Self-consistent aspects of X-ray absorption calculations. *J. Phys.: Condens. Matter* **2009**, *21*, No. 345501.

(13) Joly, Y.; Bunău, O.; Lorenzo, J. E.; Galéra, R. M.; Grenier, S.; Thompson, B. Self-consistency, spin-orbit and other advances in the FDMNES code to simulate XANES and RXD experiments. *J. Phys.: Conf. Ser.* **2009**, *190*, No. 012007.

(14) Routh, P. K.; Liu, Y.; Marcella, N.; Kozinsky, B.; Frenkel, A. I. Latent Representation Learning for Structural Characterization of Catalysts. *J. Phys. Chem. Lett.* **2021**, *12*, 2086−2094.

(15) Liu, Y.; Halder, A.; Seifert, S.; Marcella, N.; Vajda, S.; Frenkel, A. I. Probing Active Sites in CuxPdy Cluster Catalysts by Machine-Learning-Assisted X-ray Absorption Spectroscopy. *ACS Appl. Mater. Interfaces* **2021**, *13*, 53363−53374.

(16) Armbrüster, M.; Behrens, M.; Cinquini, F.; Föttinger, K.; Grin, Y.; Haghofer, A.; Klötzer, B.; Knop-Gericke, A.; Lorenz, H.; Ota, A.; Penner, S.; Prinz, J.; Rameshan, C.; Révay, Z.; Rosenthal, D.; Rupprechter, G.; Sautet, P.; Schlögl, R.; Shao, L.; Szentmiklósi, L.; Teschner, D.; Torres, D.; Wagner, R.; Widmer, R.; Wowsnick, G. How to Control the Selectivity of Palladium-based Catalysts in Hydrogenation Reactions: The Role of Subsurface Chemistry. *ChemCatChem* **2012**, *4*, 1048−1063.

(17) Bond, G.; Dowden, D.; Mackenzie, N. The selective hydrogenation of acetylene. *Trans. Faraday Soc.* **1958**, *54*, 1537−1546.
(18) Borodziński, A.; Bond, G. C. Selective Hydrogenation of Ethyne in Ethene-Rich Streams on Palladium Catalysts, Part 2: Steady-State Kinetics and Effects of Palladium Particle Size, Carbon Monoxide, and Promoters. *Catal. Rev.* **2008**, *50*, 379−469.
(19) Teschner, D.; Borsodi, J.; Kis, Z.; Szentmiklosi, L.; Revay, Z.; Knop-Gericke, A.; Schlogl, R.; Torres, D.; Sautet, P. Role of Hydrogen Species in Palladium-Catalyzed Alkyne Hydrogenation. *J. Phys. Chem. C* **2010**, *114*, 2293−2299.
(20) Tetef, S.; Govind, N.; Seidler, G. Unsupervised Machine Learning for Unbiased Chemical Classification in X-ray Absorption Spectroscopy and X-ray Emission Spectroscopy. *Phys. Chem. Chem. Phys.* **2021**, *23*, 23586−23601.
(21) Martini, A.; Guda, S. A.; Guda, A. A.; Smolentsev, G.; Algasov, A.; Usoltsev, O.; Soldatov, M. A.; Bugaev, A.; Rusalev, Y.; Lamberti, C.; Soldatov, A. V. PyFitit: The software for quantitative analysis of XANES spectra using machine-learning algorithms. *Comput. Phys. Commun.* **2020**, *250*, No. 107064.
(22) Joly, Y. X-ray absorption near-edge structure calculations beyond the muffin-tin approximation. *Phys. Rev. B* **2001**, *63*, No. 125120.
(23) Guda, S. A.; Guda, A. A.; Soldatov, M. A.; Lomachenko, K. A.; Bugaev, A. L.; Lamberti, C.; Gawelda, W.; Bressler, C.; Smolentsev, G.; Soldatov, A. V.; Joly, Y. Optimized finite difference method for the full-potential XANES simulations: Application to molecular adsorption geometries in MOFs and metal−ligand intersystem crossing transients. *J. Chem. Theory Comput.* **2015**, *11*, 4512−4521.
(24) Guda, A. A.; Guda, S. A.; Soldatov, M. A.; Lomachenko, K. A.; Bugaev, A. L.; Lamberti, C.; Gawelda, W.; Bressler, C.; Smolentsev, G.; Soldatov, A. V.; Joly, Y. Finite difference method accelerated with sparse solvers for structural analysis of the metal-organic complexes. *J. Phys.: Conf. Ser.* **2016**, *712*, No. 012004.
(25) Guda, A.; Guda, S.; Soldatov, M.; Lomachenko, K.; Bugaev, A.; Lamberti, C.; Gawelda, W.; Bressler, C.; Smolentsev, G.; Soldatov, A. Finite difference method accelerated with sparse solvers for structural analysis of the metal-organic complexes. *J. Phys.: Conf. Ser.* **2016**, *712*, No. 012004.
(26) Guda, S. A.; Guda, A. A.; Soldatov, M. A.; Lomachenko, K. A.; Bugaev, A. L.; Lamberti, C.; Gawelda, W.; Bressler, C.; Smolentsev, G.; Soldatov, A. V.; et al. Optimized finite difference method for the full-potential XANES simulations: Application to molecular adsorption geometries in MOFs and metal−ligand intersystem crossing transients. *J. Chem. Theory Comput.* **2015**, *11*, 4512−4521.
(27) Camp, C. H., Jr. PyMCR: A python library for multivariatecurve resolution analysis with alternating regression (MCR-AR). *J. Res. Natl. Inst. Stand. Technol.* **2019**, *124*, 1−10.
(28) Ahmadi, G.; Tauler, R.; Abdollahi, H. Multivariate calibration of first-order data with the correlation constrained MCR-ALS method. *Chemom. Intell. Lab. Syst.* **2015**, *142*, 143−150.
(29) De Juan, A.; Jaumot, J.; Tauler, R. Multivariate Curve Resolution (MCR). Solving the mixture analysis problem. *Anal. Methods* **2014**, *6*, 4964−4976.
(30) Liu, Y.; Halder, A.; Seifert, S.; Marcella, N.; Vajda, S.; Frenkel, A. I. Probing Active Sites in Cu_xPd_y Cluster Catalysts by Machine-Learning-Assisted X-ray Absorption Spectroscopy. *ACS Appl. Mater. Interfaces* **2021**, *13*, 53363−53374.
(31) Bugaev, A. L.; Skorynina, A. A.; Braglia, L.; Lomachenko, K. A.; Guda, A.; Lazzarini, A.; Bordiga, S.; Olsbye, U.; Lillerud, K. P.; Soldatov, A. V.; et al. Evolution of Pt and Pd species in functionalized UiO-67 metal-organic frameworks. *Catal. Today* **2019**, *336*, 33−39.
(32) Bugaev, A. L.; Srabionyan, V. V.; Soldatov, A. V.; Bugaev, L. A.; van Bokhoven, J. A. The role of hydrogen in formation of Pd XANES in Pd-nanoparticles. *J. Phys.: Conf. Ser.* **2013**, *430*, No. 012028.
(33) Bugaev, A. L.; Guda, A. A.; Lomachenko, K. A.; Lazzarini, A.; Srabionyan, V. V.; Vitillo, J. G.; Piovano, A.; Groppo, E.; Bugaev, L. A.; Soldatov, A. V.; Dmitriev, V. P.; Pellegrini, R.; van Bokhoven, J. A.; Lamberti, C. Hydride phase formation in carbon supported palladium hydride nanoparticles by in situ EXAFS and XRD. *J. Phys.: Conf. Ser.* **2016**, *712*, No. 012032.
(34) Bugaev, A. L.; Guda, A. A.; Lomachenko, K. A.; Kamyshova, E. G.; Soldatov, M. A.; Kaur, G.; Øien-Ødegaard, S.; Braglia, L.; Lazzarini, A.; Manzoli, M.; Bordiga, S.; Olsbye, U.; Lillerud, K. P.; Soldatov, A. V.; Lamberti, C. Operando study of palladium nanoparticles inside UiO-67 MOF for catalytic hydrogenation of hydrocarbons. *Faraday Discuss.* **2018**, *208*, 287−306.
(35) Bugaev, A. L.; Usoltsev, O. A.; Guda, A. A.; Lomachenko, K. A.; Pankin, I. A.; Rusalev, Y. V.; Emerich, H.; Groppo, E.; Pellegrini, R.; Soldatov, A. V.; van Bokhoven, J. A.; Lamberti, C. Palladium carbide and hydride formation in the bulk and at the surface of palladium nanoparticles. *J. Phys. Chem. C* **2018**, *122*, 12029−12037.
(36) Bugaev, A. L.; Guda, A. A.; Pankin, I. A.; Groppo, E.; Pellegrini, R.; Longo, A.; Soldatov, A. V.; Lamberti, C. The role of palladium carbides in the catalytic hydrogenation of ethylene over supported palladium nanoparticles. *Catal. Today* **2019**, *336*, 40−44.
(37) Usoltsev, O. A.; Pnevskaya, A. Y.; Kamyshova, E. G.; Tereshchenko, A. A.; Skorynina, A. A.; Zhang, W.; Yao, T.; Bugaev, A. L.; Soldatov, A. V. Dehydrogenation of ethylene on supported palladium nanoparticles: A double view from metal and hydrocarbon sides. *Nanomaterials* **2020**, *10*, 1643.
(38) Bugaev, A. L.; Usoltsev, O. A.; Guda, A. A.; Lomachenko, K. A.; Brunelli, M.; Groppo, E.; Pellegrini, R.; Soldatov, A. V.; van Bokhoven, J. Hydrogenation of ethylene over palladium: Evolution of the catalyst structure by operando synchrotron-based techniques. *Faraday Discuss.* **2021**, *229*, 197−207.
(39) Bugaev, A. L.; Usoltsev, O. A.; Lazzarini, A.; Lomachenko, K. A.; Guda, A. A.; Pellegrini, R.; Carosso, M.; Vitillo, J. G.; Groppo, E.; van Bokhoven, J. A.; Soldatov, A. V.; Lamberti, C. Time-resolved operando studies of carbon supported Pd nanoparticles under hydrogenation reactions by X-ray diffraction and absorption. *Faraday Discuss.* **2018**, *208*, 187−205.
(40) Bugaev, L. A.; Avakyan, L. A.; Srabionyan, V. V.; Bugaev, A. L. Resolution of interatomic distances in the study of local atomic structure distortions by energy-restricted x-ray absorption spectra. *Phys. Rev. B* **2010**, *82*, No. 064204.

https://doi.org/10.1021/acs.jpcc.1c09420
J. Phys. Chem. C 2022, 126, 4921−4928