small methods
www.small-methods.com

Machine Learning for Quantitative Structural Information from Infrared Spectra: The Case of Palladium Hydride

*Oleg Usoltsev, Andrei Tereshchenko, Alina Skorynina, Elizaveta Kozyr, Alexander Soldatov, Olga Safonova, Adam H. Clark, Davide Ferri, Maarten Nachtegaal, and Aram Bugaev**

Infrared spectroscopy (IR) is a widely used technique enabling to identify specific functional groups in the molecule of interest based on their characteristic vibrational modes or the presence of a specific adsorption site based on the characteristic vibrational mode of an adsorbed probe molecule. The interpretation of an IR spectrum is generally carried out within a fingerprint paradigm by comparing the observed spectral features with the features of known references or theoretical calculations. This work demonstrates a method for extracting quantitative structural information beyond this approach by application of machine learning (ML) algorithms. Taking palladium hydride formation as an example, Pd-H pressure-composition isotherms are reconstructed using IR data collected in situ in diffuse reflectance using CO molecule as a probe. To the best of the knowledge, this is the first example of the determination of continuous structural descriptors (such as interatomic distance and stoichiometric coefficient) from the fine structure of vibrational spectra, which opens new possibilities of using IR spectra for structural analysis.

1. Introduction

Infrared (IR) spectroscopy is a powerful and extensively applied tool to probe the vibrational properties of a molecule or a compound that are determined by its atomic and electronic structure. One can attribute a spectral feature to a specific vibration of the atoms involved, bond orders, and even 3D geometry.[1,2] For example, an experienced researcher can immediately recognize the nature of the observed vibration based on its characteristic wavelength, or discriminate different oxidation states of a metal site based on the vibrational frequency of a probe molecule (e.g., CO),[3–5] or identify the type of metal surface and different geometries of the adsorbates (e.g., linear, bridge, or hollow).[6–8]

In the recent years, it was demonstrate how machine learning (ML) algorithms can help extracting quantitative information from various spectroscopic data.[9–22] For example, by analyzing the relative intensities of the peaks originated from different geometries of CO adsorbed on platinum surfaces, Lansford et al. reconstructed the morphology of platinum particles,[23] which demonstrate how ML can be used to go beyond a fingerprint approach toward quantitative interpretation of IR data. Although powerful, the basis of this approach is not new, since the influence of the particle size on the distribution of surface adsorbed species has been known for decades[24,25] and was also recognized without using ML methods.[8,26,27]

In the current work, we pose a more ambitious question whether ML applied to IR spectra can help to extract fine details of the atomic and electronic structure that are ignored in conventional, human-processed analysis of IR spectra. In particular, we were interested in using IR data for extracting metal–metal bond distances, for which one normally needs X-ray based data. As an example, we consider the process of palladium hydride formation in supported palladium nanoparticles, during which hydrogen is absorbed inside palladium lattice causing expansion of the latter. This process can be tracked directly or indirectly by numerous techniques. For example, volumetric measurements show the amount of absorbed hydrogen from the gas phase per known amount of sample. Palladium lattice expansion can be tracked by extended X-ray absorption fine structure (EXAFS) spectroscopy or X-ray diffraction (XRD).[28–31] Changes in the optical properties of the PdH_x phase allowed its characterization by indirect nanoplasmonic sensing calibrated by XRD data.[32]

Here, we propose a novel quantitative approach utilizing ML applied to diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS) to get insight on the in situ evolution of

O. Usoltsev, A. Skorynina
ALBA Synchrotron
Cerdanyola del Valles, Barcelona 08290, Spain
A. Tereshchenko, A. Soldatov
Southern Federal University
Sladkova 178/24, Rostov-on-Don 344090, Russia
E. Kozyr
University of Turin
Via Giuria 7, Turin 10125, Italy
O. Safonova, A. H. Clark, D. Ferri, M. Nachtegaal, A. Bugaev
Paul Scherrer Institute
Forschungsstrasse 111, Villigen 5232, Switzerland
E-mail: aram.bugaev@psi.ch

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/smtd.202301397

DOI: 10.1002/smtd.202301397

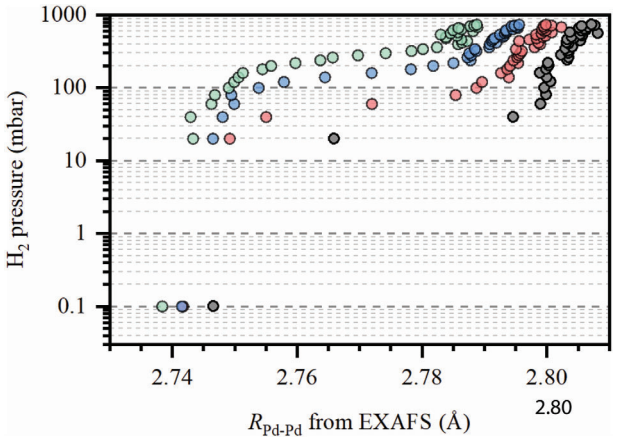

Figure 1. Evolution of Pd–Pd distances obtained from FT-EXAFS data upon increasing H_2 partial pressure at 30 (black), 50 (red), 70 (blue) and 90 (green)°C.

metallic nanoparticle structures supported on alumina, focusing on the formation of the PdH_x phase. We have built a database of DRIFTS and EXAFS spectra collected in identical conditions at various points of Pd-H phase diagram and applied ML algorithms to establish the correlation between the features of DRIFTS spectra and Pd–Pd distances obtained from EXAFS data. We also investigate the effect of the measurement conditions and instrumental parameters on the prediction quality and give the theoretical interpretation of the observed sensitivity of DRIFTS based on density functional theory (DFT) calculations.

2. Results and Discussion

To collect EXAFS and DRIFTS spectra at different points of Pd-H phase diagram for a commercial 5 wt.% Pd/Al_2O_3 sample characterized by a monodispersed Pd nanoparticles of ca. 3 nm in diameter (see Section S1, Supporting Information), an automated setup was built (Figure S1, Supporting Information) allowing for remote control of the sample temperature and partial hydrogen pressure by adjusting the flow of H_2, CO, and Ar. Importantly, exactly the same setup with the same set of gasses and one in situ cell was used for both EXAFS and DRIFTS measurements. The H_2 and He flows were changing stepwise to achieve partial hydrogen pressures from 0 to 0.76 bar according to the same protocol and the same timing for both experiments (see Sections S2 and S3, Supporting Information).

Starting from the in situ reduced metallic Pd particles, the H_2 pressure was increased stepwise reaching the steady-state equilibrium in each point (Figure S2, Supporting Information) with continuous Pd *K*-edge EXAFS data collection. To be consistent with DRIFTS studies (vide infra) where CO was used as probe molecule, 0.5 vol% CO gas was permanently flown through the cell. The measurements were conducted at four different temperatures and the Pd–Pd interatomic distances obtained from by fitting the Fourier transformed (FT) EXAFS spectra demonstrate a typical behavior for PdH_x phase formation (**Figure 1**). Although the commonly reported absolute error in the determination of interatomic distances associated with EXAFS fitting is ≈0.01 Å, the relative changes of this parameter observed from a single experiment under in situ conditions can be determined with much smaller error (even two orders of magnitude).[33]

The DRIFTS data collected for Pd/Al_2O_3 sample in presence of CO are characterized by three main peaks in the region corresponding to C–O bond stretching (**Figure 2a**) assigned to different absorption geometries of CO to the surface of palladium nanoparticles. Minor but noticeable changes were observed in the peak positions and intensities, which can be best appreciated by the red-shift up to 10 cm^{-1} in the position of the peak at 1980 cm^{-1} corresponding to the adsorbed CO species in the bridge geometry. Other spectral features were also changing but with smaller variations. Notably, the logarithm of the H_2 pressure plotted versus the red-shift of the bridged CO (Figure 2b) already resembled the pressure-composition isotherms, similar to those obtained from EXAFS, which allows one to correlate the red-shift of the C—O frequency with the Pd–Pd interatomic distances (Figure 2c). For the quantification of such small changes, it was important to collect the spectra at 0.5 cm^{-1} spectral resolution. The justification and accuracy of such strategy is supported by a monotonous increase of the observed shift with the increasing H_2 pressure (e.g., almost 50 experimental points are homogenously spanned over the rage of less 4 cm^{-1} for 90°C). The excellent correlation between the parameters derived from DRIFTS and EXAFS data demonstrates that vibrational spectroscopy can be applied to estimate the Pd–Pd distance (or similar structural information, such as cell parameters or H/Pd ratio) in palladium nanoparticles using CO as a probe molecule.

It has to be noted, that for the sake of clarity, Figures 1 and 2 show only a fraction of the data, from which the correlations can be visually evidenced. In the full dataset, the initial position of the bridged CO peak differs for different temperatures and even for several isotherms were subsequently acquired at one temperature (see Figure S4 and discussion in Sections S8 and S9, Supporting Information). Thus, the red-shift can be only a semi-quantitative descriptor to follow palladium hydride structure within a single experimental dataset, while the exact quantitative correlation between the C—O frequency and the H/Pd ratio (or Pd–Pd distances) remains unknown. This can be explained by the fact that i) the presence of CO can cause formation of palladium carbide-like phase[29,34–36] and influences the C—O frequency,[37,38] and ii) the temperature has the major effect on the spectra as evidenced by principal component analysis (PCA), as shown in Figure S6 (Supporting Information). EXAFS experiment performed under similar conditions without and with CO highlights the effect of latter on the pressure-composition isotherms (Section S8, Supporting Information).

To develop a method of quantitative extraction of the Pd–Pd distances from DRIFTS data along the formation of PdH_x phase, we further applied the ML procedure (see section S6, Supporting Information) to the experimental datasets. To train the algorithm, each DRIFTS spectrum was labeled by the precise first-shell Pd–Pd distance obtained from EXAFS analysis. Prediction based on the area-normalized spectra in the whole region from 2150 to 1750 cm^{-1}, as shown in Figure 2a, using the ExtraTrees algorithm results in the root-mean-square error on R_{Pd-Pd} of only 0.002 Å (i.e., an order of magnitude smaller than the standard uncertainty of 0.01 Å of the EXAFS analysis) and R^2-score, obtained in a cross-validation (CV) procedure, of 99% (**Figure 3**). Comparable prediction accuracy were obtained when considering

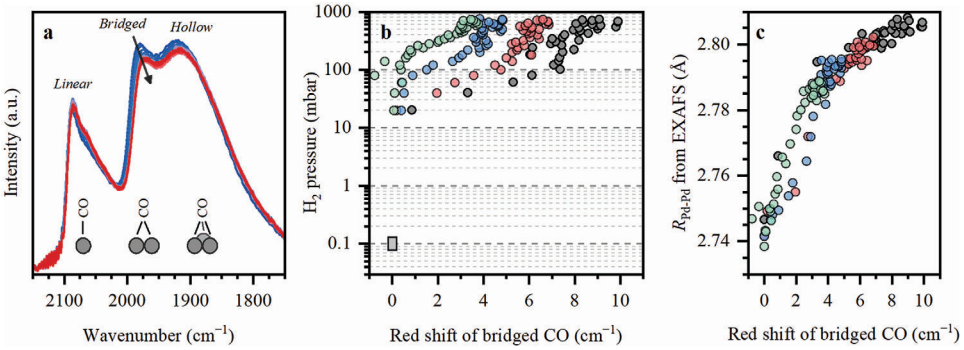

Figure 2. a) Evolution of DRIFTS spectra in the region of the C—O stretch mode of adsorbed CO upon increase of the H_2 partial pressure (from blue to red) at 30 °C. b) Red-shift of bridge CO (1980 cm^{-1}) upon increase of the H_2 partial pressure at 30 (black), 50 (red), 70 (blue) and 90 (green)°C. c) Correlation between the red-shift of bridged CO and the Pd–Pd distance obtained from EXAFS.

solely the positions of the three maxima in the DRIFTS spectra or the parameters derived from spectral deconvolution (see Figure S7, Supporting Information for the comparison of different ML methods and different spectral descriptors).

Following an excellent prediction quality from the CV procedure, the transferability of the trained algorithm to the datasets measured under different conditions was verified (see Section S7, Supporting Information for details). The choice of spectral descriptors was found to be crucial. Since the most noticeable differences in the spectral shape were induced by temperature (Figures S3 and S6, Supporting Information), we manually excluded the data of the isotherm collected at 70°C from the training set and used it as a test set. The prediction based on the selection of a narrow spectral interval (100 cm^{-1}) or selected spectral descriptors demonstrated significantly better quality (R^2-scores up to 80 %) compared to the full spectral region (R^2-score ≈50%, Figure S10, Supporting Information). The selection of a limited number of proper spectral descriptors further increased the R^2-score to above 90% compared to the case when all deconvolution parameters were used (R^2-score of 33 %, Figure S16, Supporting Information). In addition, the isotherms were also extracted from data measured using a different instrument, a different in situ cell, and different CO content (Figures S13, S18, and S20, Supporting Information). These results also demonstrate the advantage of ML-based analysis and descriptor approach, compared to the utilization of full spectra or using analytical function based on the red-shift (Section S9, Supporting Information).

Finally, DFT calculations were performed to understand the effect of Pd atomic and electronic structure on the vibrational properties of adsorbed CO species (Section S9, Supporting Information). In addition to the possible effect due to the Pd-H charge transfer, elongation of Pd–Pd distance from 2.75 (black line) and 2.81 Å causes itself a red-shift of bridged CO (Figure S26, Supporting Information), comparable with the experimentally observed one. At the same time, the changes in the partial charge of Pd also cause a significant shift of the C—O frequency. To create a homogenous charge distribution, rather than making a local defect, the number of electrons per computational unit cell was artificially changed instead of the explicit addition of H-atoms (see Section S9, Supporting Information). It should be noted, that the electronegativity values for hydrogen and palladium are similar, therefore one can expect a very small charge transfer. Previous theoretical studies reported contradicting results for hydrogen in palladium from being in a highly cationic state to having a small negative charge. Additional attention should be paid to the fact that the surface Pd atoms, that interact with CO, can have different charge states and Pd–Pd distances compared to the bulk ones, which leave space for many speculations on the agreement between theoretical and experimental data. Therefore, we limit ourselves to a qualitative comparison, which, however, evidences the effect of both charge state and Pd–Pd distances on the vibrational spectra.

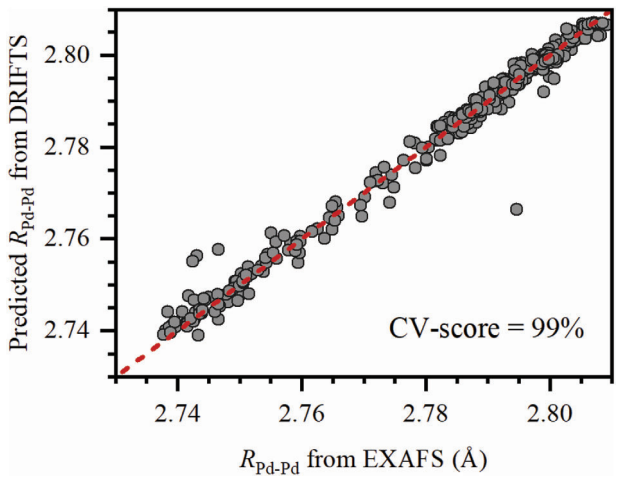

Figure 3. Pd–Pd distance predicted from DRIFTS spectra versus true distance obtained from FT-EXAFS.

3. Conclusion

In conclusion, this work presents a novel and yet unexplored phenomenon of the metal–metal interatomic distances effect on the vibrational spectra of probe molecules. Powered by ML, this phenomenon was turned into a new method for quantitative characterization of nanoparticles from infrared spectra. Considering that vibrational spectroscopy is widespread and easily accessible, we are convinced that the suggested approach can be further extended to various material systems and processes. A special

interest is the possibility of studying materials directly under *operando* reaction conditions with the addition of probe molecules or by exploiting reactive adsorbates originating from the substrate (reactants), or reaction intermediates, or products as probe molecules. In particular, as shown in the present example, this can identify the stoichiometry of the PdH_x phase, which is important in numerous hydrogenation reactions over Pd catalysts, including hydrogenation of hydrocarbons, CO_2 reduction, and electrochemical processes. The work also showed that CO molecules which, from the one hand are commonly used as probes and, from the other hand, can originate as reaction intermediates, have an irreversible effect on the PdH_x phase diagram, highlighting the necessity of in situ characterization techniques.

4. Experimental Section

Pd/Al_2O_3 samples (5 wt.%) used in this study were provided by Chimet S.p.A (Arezzo, Italy). The samples are monodispersed Pd nanoparticles with the average size of 3 nm. The details on sample preparation are given in section S1 (Supporting Information).

DRIFTS and EXAFS data were collected at Paul Scherrer Institute (Switzerland) using the same home-made operando cell and remotely controlled gas setup described in detail in Section S2 (Supporting Information). Both datasets were processed using self-written python codes. For DRIFTSs the procedure included the extraction of data from Bruker Opus format, averaging of the subsequent spectra taken under the same conditions, normalization by area and deconvolution by gaussians in the selected region. For EXAFS the procedure included averaging of raw data collected at 1 s time resolution over the periods in which the conditions were considered to be steady state, energy alignment and normalization, background subtraction, Fourier-Transformation and firsts-shell fitting to extract Pd–Pd distances. All data (averaged and normalized spectra) together with the spectral and structural descriptors are available at https://doi.org/10.5281/zenodo.8387073.

ML analysis was done in a self-written python code in a Jupyter Notebook. Supervised ML algorithms were used within scikit-learn python library for solution of the regression task. The cross-validation (CV) procedure was applied within *KFold* algorithm with number of folds "n_splits" = 10. The evaluation of CV was performed using *cross_val_predict* method implemented in scikit-learn library. Training set consisted of two parts: **X** or the data to fit and **Y** or the target variable for prediction. For **X** part we used either an IR spectrum, its part in the selected range, the parameters of Gussian fit, or peak positions. In addition,a wide range of descriptor combinations was also tested. For **Y**, we always used the Pd–Pd interatomic distances. The comparison of different ML algorithms such as Linear regression, Ridge regression, SVR, Decision tree and Random forest was done in Section S6 (Supporting Information). The Extra Trees Regressor was selected as the best choice. For this approach the minimum number of samples required to split "*min_samples_split*" was set to default value of 2, while the number of trees "*n_estimators*" was varied from 2 to 150 and plotted over CV score Figure S8 (Supporting Information). The CO region of IR spectrum was used for *X* here in the training set. The value *n_estimators* = 100 was selected as optimal and used for all runs.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

Open access funding provided by ETH-Bereich Forschungsanstalten.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are openly available in [Zenodo] at [https://doi.org/10.5281/zenodo.8387073], reference number [8387073].

Keywords

drifts, EXAFS, machine learning, palladium hydrides

Received: October 12, 2023
Revised: January 9, 2024
Published online: January 31, 2024

[1] J. K. Gillie, J. Hochlowski, G. A. Arbuckle-Keil, *Anal. Chem.* **2000**, *72*, 71.
[2] E. Stavitski, B. M. Weckhuysen, *Chem. Soc. Rev.* **2010**, *39*, 4615.
[3] G. T. Palomino, S. Bordiga, A. Zecchina, G. L. Marra, C. Lamberti, *J. Phys. Chem. B* **2000**, *104*, 8641.
[4] G. Turnes Palomino, P. Fisicaro, S. Bordiga, A. Zecchina, E. Giamello, C. Lamberti, *J. Phys. Chem. B* **2000**, *104*, 4064.
[5] G. L. Chiarello, Y. Lu, M. Agote-Arán, R. Pellegrini, D. Ferri, *Catalysts* **2021**, *116*, 11.
[6] F. Zaera, *Prog. Surf. Sci.* **2001**, *69*, 1.
[7] O. A. Usoltsev, A. Y. Pnevskaya, E. G. Kamyshova, A. A. Tereshchenko, A. A. Skorynina, W. Zhang, T. Yao, A. L. Bugaev, A. V. Soldatov, *Nanomaterials* **2020**, *10*, 1643.
[8] A. A. Tereshchenko, V. V. Butova, A. A. Guda, O. A. Burachevskaya, A. L. Bugaev, A. N. Bulgakov, A. A. Skorynina, Y. V. Rusalev, I. V. Pankov, V. A. Volochaev, M. Al-Omoush, I. V. Ozhogin, G. S. Borodkin, A. V. Soldatov, *Inorg. Chem.* **2022**, *61*, 3875.
[9] X. Wang, S. Jiang, W. Hu, S. Ye, T. Wang, F. Wu, L. Yang, X. Li, G. Zhang, X. Chen, J. Jiang, Y. Luo, *J. Am. Chem. Soc.* **2022**, *144*, 16069.
[10] A. A. Enders, N. M. North, C. M. Fensore, J. Velez-Alvarez, H. C. Allen, *Anal. Chem.* **2021**, *93*, 9711.
[11] A. A. Skorynina, B. O. Protsenko, O. A. Usoltsev, S. A. Guda, A. L. Bugaev, *Inorg. Chem.* **2023**, *62*, 6608.
[12] O. A. Usoltsev, A. L. Bugaev, A. A. Guda, S. A. Guda, A. V. Soldatov, *J. Phys. Chem. C* **2022**, *126*, 4921.
[13] E. G. Kozyr, A. L. Bugaev, S. A. Guda, A. A. Guda, K. A. Lomachenko, K. Janssens, S. Smolders, D. De Vos, A. V. Soldatov, *J. Phys. Chem. C* **2021**, *125*, 27844.
[14] A. Martini, A. L. Bugaev, S. A. Guda, A. A. Guda, E. Priola, E. Borfecchia, S. Smolders, K. Janssens, D. De Vos, A. V. Soldatov, *J. Phys. Chem. A* **2021**, *125*, 7080.
[15] A. Martini, A. A. Guda, S. A. Guda, A. L. Bugaev, O. V. Safonova, A. V. Soldatov, *Phys. Chem. Chem. Phys.* **2021**, *23*, 17873.
[16] A. A. Guda, S. A. Guda, A. Martini, A. N. Kravtsova, A. Algasov, A. Bugaev, S. P. Kubrin, L. V. Guda, P. Šot, J. A. van Bokhoven, C. Copéret, A. V. Soldatov, *npj Comput. Mater.* **2021**, *7*, 203.
[17] O. A. Usoltsev, A. L. Bugaev, A. A. Guda, S. A. Guda, A. V. Soldatov, *Top. Catal.* **2020**, *63*, 58.
[18] A. Martini, S. A. Guda, A. A. Guda, G. Smolentsev, A. Algasov, O. Usoltsev, M. A. Soldatov, A. Bugaev, Y. Rusalev, C. Lamberti, A. V. Soldatov, *Comput. Phys. Commun.* **2020**, *250*, 107064.
[19] A. A. Guda, S. A. Guda, A. Martini, A. L. Bugaev, M. A. Soldatov, A. V. Soldatov, C. Lamberti, *Radiat. Phys. Chem.* **2020**, *175*, 108430.
[20] K. Choudhary, B. DeCost, C. Chen, A. Jain, F. Tavazza, R. Cohn, C. W. Park, A. Choudhary, A. Agrawal, S. J. L. Billinge, E. Holm, S. P. Ong, C. Wolverton, *npj Comput. Mater.* **2022**, *8*, 59.
[21] Y. Xu, J. Ge, C.-W. Ju, *Energy Advances* **2023**, *2*, 896.

[22] S. F. Acaru, R. Abdullah, D. T. C. Lai, R. C. Lim, *Energy Advances* **2023**, *2*, 1204.
[23] J. L. Lansford, D. G. Vlachos, *Nat. Commun.* **2020**, *11*, 1513.
[24] R. van Hardeveld, F. Hartog, in *Advances in Catalysis*, (Eds.: D. D. Eley, H. Pines, P. B. Weisz) Academic Press, Cambridge, **1972**, *22*, pp. 75.
[25] L. L. Sheu, Z. Karpinski, W. M. H. Sachtler, *J. Phys. Chem.* **1989**, *93*, 4890.
[26] Q. Fan, S. He, L. Hao, X. Liu, Y. Zhu, S. Xu, F. Zhang, *Sci. Rep.* **2017**, *7*, 42172.
[27] A. Tereshchenko, A. Guda, V. Polyakov, Y. Rusalev, V. Butova, A. Soldatov, *Analyst* **2020**, *145*, 7534.
[28] A. L. Bugaev, A. A. Guda, K. A. Lomachenko, V. V. Srabionyan, L. A. Bugaev, A. V. Soldatov, C. Lamberti, V. P. Dmitriev, J. A. van Bokhoven, *J. Phys. Chem. C* **2014**, *118*, 10416.
[29] A. L. Bugaev, A. A. Guda, A. Lazzarini, K. A. Lomachenko, E. Groppo, R. Pellegrini, A. Piovano, H. Emerich, A. V. Soldatov, L. A. Bugaev, V. P. Dmitriev, J. A. van Bokhoven, C. Lamberti, *Catal. Today* **2017**, *283*, 119.
[30] A. L. Bugaev, A. A. Guda, K. A. Lomachenko, V. V. Shapovalov, A. Lazzarini, J. G. Vitillo, L. A. Bugaev, E. Groppo, R. Pellegrini, A. V. Soldatov, J. A. van Bokhoven, C. Lamberti, *J. Phys. Chem. C* **2017**, *121*, 18202.
[31] A. L. Bugaev, O. A. Usoltsev, A. Lazzarini, K. A. Lomachenko, A. A. Guda, R. Pellegrini, M. Carosso, J. G. Vitillo, E. Groppo, J. A. van Bokhoven, A. V. Soldatov, C. Lamberti, *Faraday Discuss* **2018**, *208*, 187.
[32] C. Langhammer, E. M. Larsson, B. Kasemo, I. Zoric, *Nano Lett.* **2010**, *10*, 3529.
[33] R. F. Pettifer, O. Mathon, S. Pascarelli, M. D. Cooke, M. R. Gibbs, *Nature* **2005**, *435*, 78.
[34] A. L. Bugaev, A. A. Guda, I. A. Pankin, E. Groppo, R. Pellegrini, A. Longo, A. V. Soldatov, C. Lamberti, *Catal. Today* **2019**, *336*, 40.
[35] A. A. Skorynina, A. A. Tereshchenko, O. A. Usoltsev, A. L. Bugaev, K. A. Lomachenko, A. A. Guda, E. Groppo, R. Pellegrini, C. Lamberti, A. Soldatov, *Radiat. Phys. Chem.* **2020**, *175*, 108079.
[36] A. L. Bugaev, O. A. Usoltsev, A. A. Guda, K. A. Lomachenko, I. A. Pankin, Y. V. Rusalev, H. Emerich, E. Groppo, R. Pellegrini, A. V. Soldatov, J. A. van Bokhoven, C. Lamberti, *J. Phys. Chem. C* **2018**, *122*, 12029.
[37] O. A. Usoltsev, A. A. Skorynina, B. O. Protsenko, V. Martin-Diaconescu, R. Pellegrini, A. V. Soldatov, J. van Bokhoven, A. L. Bugaev, *Appl. Surf. Sci.* **2023**, *614*, 156171.
[38] A. L. Bugaev, V. V. Srabionyan, A. V. Soldatov, L. A. Bugaev, J. A. van Bokhoven, *J. Phys.: Conf. Ser.* **2013**, *430*, 012028.

