ISSN 1027-4510, Journal of Surface Investigation: X-ray, Synchrotron and Neutron Techniques, 2021, Vol. 15, No. 1, pp. 1–6. © Pleiades Publishing, Ltd., 2021.
Russian Text © The Author(s), 2021, published in Poverkhnost', 2021, No. 1, pp. 3–9.

# Application of Ligand Field Theory for Simulation of the Pre-Edge Structure of X-Ray Absorption Spectra of Amorphous Systems

**D. M. Pashkov[a, b, *], D. S. Rubanik[a], M. V. Kirichkov[a], A. A. Guda[a, **], S. A. Guda[a, b], and A. V. Soldatov[a]**

[a] *International Smart Material Research Institute, Southern Federal University, Rostov-on-Don, 344090 Russia*
[b] *Vorovich Institute of Mathematics, Mechanics, and Computer Sciences, Southern Federal University, Rostov-on-Don, 344090 Russia*
**e-mail: pashkov@sfedu.ru*
***e-mail: guda@sfedu.ru*

Received February 28, 2020; revised March 25, 2020; accepted March 30, 2020

**Abstract**—The theoretical basis of X-ray absorption spectra beyond the Fe *K*-edge in the pre-edge region is constructed, and algorithms for using it to quantitatively analyze experimental X-ray near edge spectroscopy (XANES) spectra are described. Within the framework of the ab initio approach of ligand field theory, 60 theoretical spectra or more for Fe with different coordination numbers, oxidation state, and interatomic distances between the central atom and the nearest O atoms are calculated. Various machine learning (ML) methods for solving problems of determining the oxidation state of Fe and coordination number by means of the obtained spectra are used. The quality of the models are estimated using the Leave-One-Out method, the determination accuracy for the oxidation state is 100%; the determination accuracy for the coordination number is 75% in the case of predictions using the entire dataset and 96 and 68% for datasets with fixed oxidation state of Fe +2 and +3, respectively.

**Keywords:** XANES, pre-edge, Wannier function, MLFT, machine learning, Extra Trees
**DOI:** 10.1134/S1027451021010110

## INTRODUCTION

The study of the local atomic structure around 3*d* metal atoms in amorphous matrices is of interest for a wide range of fields (from geology to catalysis). So, tektite formation conditions determine the oxidation state and the coordination of Fe atoms in silicon dioxide, and the ethylene polymerization reaction depends on the arrangement of individual chromium atoms on the $SiO_2$ surface. Tektites and impaktites are silicate glasses of impact origin [1]. It is accepted to assume that tektites are formed from melt high-speed ejection at early stages of impact-crater formation [2]. The chemical composition of tektites includes significant quantities of silicon dioxide (68–82%), a high content of alumina, a low content of water, and microcavities filled with a mixture of gases, such as carbon dioxide, hydrogen, methane, and noble gases. It was known that parameters, such as the pressure *P*, the temperature *T*, the oxygen volatility $f_{O_2}$, which existed during the glass-formation process, determine the cation coordination. Iron is one of the most widespread elements in impact glasses; therefore, iron can be potentially useful as a sensor of conditions for impact rock formation. For example, the dependence of the oxidized-reduced iron state on *P*, *T*, $f_{O_2}$, and the composition was discussed for melts in [3]. The influence of the parameters *P*, *T*, and $f_{O_2}$ on the ratio $Fe^{3+}/Fe^{2+}$ in impact melts was reported in [4]. Determination of the oxidation state of Fe and the local atomic structure in tektites and impact glasses is an important scientific problem from the point of view of understanding the formation conditions for these objects.

X-ray absorption spectroscopy is a reliable tool for determining the local atomic and electronic structures of a substance in the condensed state [5, 6]. Wilke et al. [7] showed that the pre-edge structure of the absorption spectra beyond the Fe *K*-edge is more effective for determination of the charged states of Fe atoms. They demonstrated this using the chart of the distribution of the integral intensity of the pre-edge and its center of mass for Fe compounds with different coordination numbers and valences as an example. So, for tektites and impactites, the authors of [8–10] showed that, in the amorphous Si matrix, Fe atoms can occupy positions with coordination numbers between 4 and 6 and with valences in a wide range from $Fe^{2+}$ to $Fe^{3+}$. Difficulties related to the quantitative analysis of spectra of

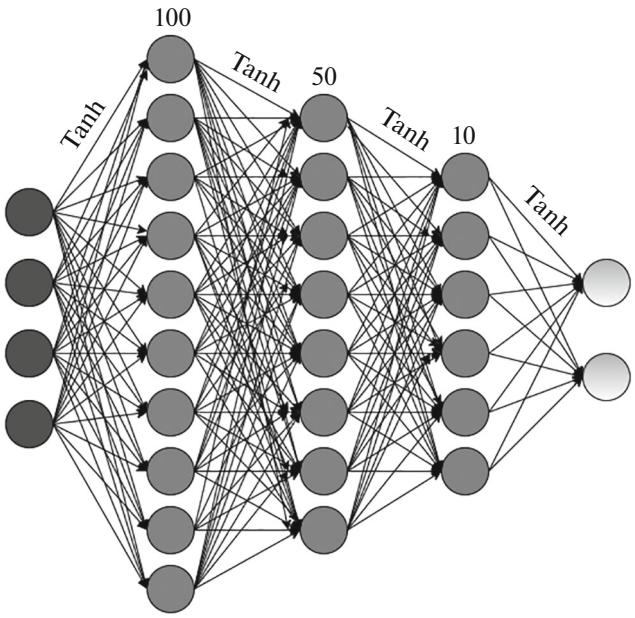


**Fig. 1.** Diagram of the applied multilayer perceptron.

such multicomponent systems make one seek new approaches to the analysis of experimental data.

Recently, machine learning (ML) became a powerful tool for the processing of experimentally obtained data and for theoretical studies. ML methods are able to solve various problems, such as the classification and ranking of objects, the construction of approximations of some functions, the determination of hidden dependences and features in data, etc. The ML methods were already used to analyze spectral data [11–13]. Three-dimensional reconstruction of the local environment of the Fe cation in the $SiO_2$ matrix is possible by analyzing the XANES spectrum because of the contribution of multiple scattering effects in the case of a large mean free path of the photoelectron. This makes it possible to confirm the structural model by comparing the experimental XANES spectrum with the theoretical one constructed for a hypothetical structure [14–17]. The difficulty is that the local environment is a priori unknown; therefore, a large number of test structures [18] can be required to reach the convergence.

As shown recently, the ML methods were successfully used to quantitatively analyze spectral data on the whole [11, 19, 20] and X-ray near edge spectroscopy (XANES) in particular [21, 22]. In this paper, we show the applicability of these methods for determining the structural parameters of the compounds, such as the oxidation state and the coordination number, for the pre-edge region of the XANES spectra.

## EXPERIMENTAL

The pre-edge structure of the X-ray absorption spectra was calculated within the framework of the multiplet theory of ligand fields [23] using Wannier orbitals [24–26] in the XTLS program [27, 28]. At the first stage, the band structure was calculated using the approach of the linearized supplemented basis of plane waves with the total potential, as was implemented in the WIEN2k [29] for cells that contain a full octahedron (or tetrahedron) of oxygen atoms around a 3*d* metal. Then we constructed the local basis of maximally localized Wannier functions, taking all bands from a certain energy window forming the valence band. The typical energy values limiting the wannierization region ranged from –10 to 2 eV with respect to the Fermi level. After the construction of the low-energy basis set of Wannier orbitals, we obtained a local low-energy Hamiltonian containing up to 113 Wannier orbitals. However, a system of such a size cannot be used for multiparticle calculations because their difficulty increased exponentially with the basis size. Thus, we had to restrict ourselves to orbitals related to low-energy system excitations. The minimum local model describing spectra beyond the transition metal *K*-edge contained five Fe 3*d*-like orbitals and five molecular orbitals constructed from the superposition of oxygen orbitals of the *p* type. The reduction of 10 orbitals to the space required applying the block Lanczos procedure to the full matrix of Wannier orbitals, which transformed it into the three-diagonal form in order to construct molecular orbitals with maximum hoppings at 3*d* orbitals.

To analyze the pre-edge region of the XANES spectra, we used different ML methods, such as Decision Tree, ExtraTrees [30], the SVM (support vector machine) method, the logistic regression, and multilayer perceptron (MLP). These methods were used to predict the coordination numbers of compounds and the oxidation state of the Fe atom in compounds using their spectra. In our paper, we used the ML methods implemented in the scikit-learn library for the Python program language. The following parameters were used for the ML methods under consideration:

(1) Decision Tree: the maximum depth of the decision tree was 500; the function of the separation quality estimation (the information gain).

(2) ExtraTrees: the number of trees was 300, the function of the separation quality estimation (the Gini criterion).

(3) SVM: the regularization parameter $C$ was 1000.

(4) Logistic regression: we used the L2 regularization with a parameter of $C = 1000$; the optimization algorithm was the quasiNewton limited-memory BFGS algorithm (L-BFGS) [31]. The number of iterations was 4000.

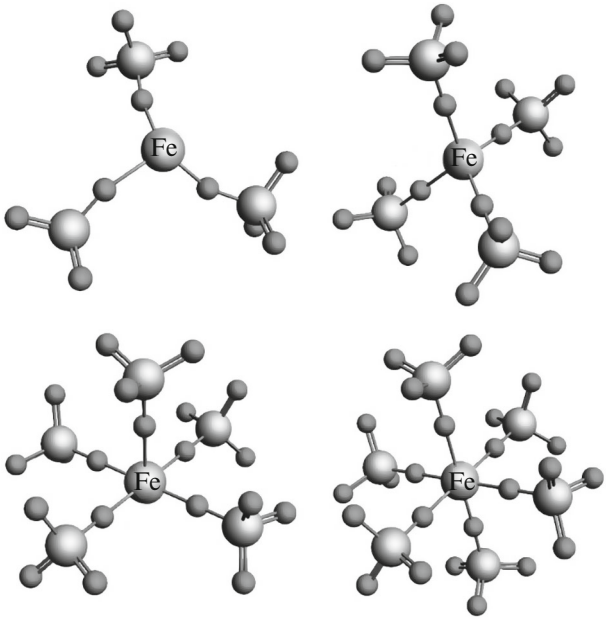


**Fig. 2.** Structural models of the surroundings of Fe ions by $SiO_4$ tetrahedrons.

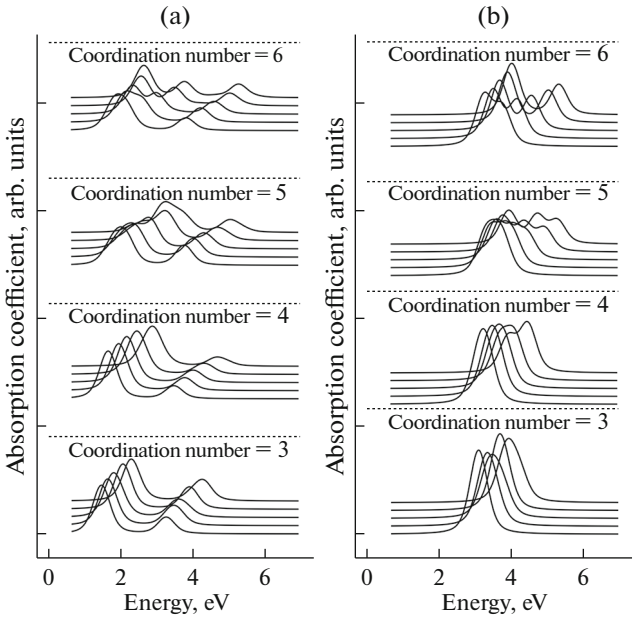


**Fig. 3.** Calculated spectra of the XANES pre-edge region beyond the Fe *K*-edge for compounds with the valences (a) $Fe^{2+}$ and (b) $Fe^{3+}$.

(5) The MLP had 3 hidden layers with dimensions of 100, 50, and 10 with a hyperbolic tangent as an activation function (Fig. 1). The network learned during 5000 epochs.

## RESULTS AND DISCUSSION

The application of ML methods means the presence of a dataset for algorithm training and testing; therefore, the dataset had to be prepared for the study. At the first stage, it was necessary to create structure files for compounds. The structures were generated using a molecule program module of Py FitIt software [32]. This module made it possible to construct a molecule from other compounds or individual atoms, giving interatomic distances between constituents or rotating them with respect to a certain center through a given angle. Structure files of compounds with different coordination numbers were formed (Fig. 2); the distance from the central Fe atom to the nearest O atoms was varied uniformly from 1.8 to 2.2 Å with a step of 0.1 Å for them.

For each structure, periodic wave functions were found by solving the Schrödinger equation in electron density functional theory, localization using the Wannier method was performed, and spectra beyond the Fe *K*-edge were calculated. The spectra were calculated for different configurations of the *d* shell for Fe atoms, which corresponded to the $Fe^{2+}$ and $Fe^{3+}$ valence (Figs. 3a and 3b). The processes of charge transfer from the metal to the ligand were taken into account in the simulation. 60 spectra were obtained and constituted our training dataset for application of the ML methods.

It follows from the obtained graphs of the spectra (Fig. 3) that, on the graph, the peak heights and widths change noticeably when the coordination number is varied; and, when the oxidation state of Fe is varied, a chemical shift occurs, and the number of peaks can also be changed, which is explained by the different multiplicities of $Fe^{2+}$ and $Fe^{3+}$ ions. As the interatomic distance between the central Fe atom and the nearest O atoms decreases, the overlapping of *d* orbitals with ligand orbitals and the crystal-field parameter decrease. This leads to a change in the energy positions of the spectral features, which is expressed in the form of peak splitting or merging. It was decided that mathematical methods would be used to analyze the data to estimate the amount of structural information.

To successfully use the ML methods, it is required that all objects of the training dataset be represented as a certain set of features that describe each object exhaustively. Each spectrum from our training dataset included values of the absorption coefficient intensity for different energies; in this case, the majority of these values were zero or very small. Moreover, different spectra were determined in the crossed energy intervals rather than in the same ones. For these reasons, the training of the ML methods using the absorption intensities was inefficient. To solve this problem, it was necessary to single out new informative features using existing spectra. To do this, a descriptor was developed; it constructed new features using the pre-edge

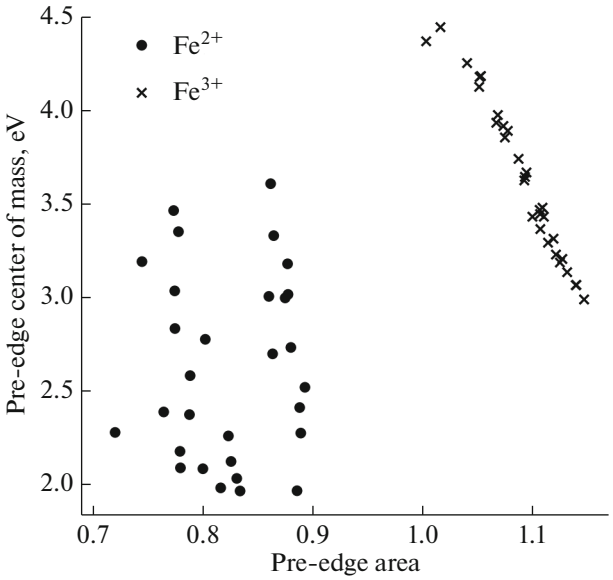


**Fig. 4.** Graph of scattering constructed using two features: the area of the pre-edge region and the pre-edge centroid.

XANES spectra. As features, we chose the peak positions and amplitudes in the graphs of the spectra, the areas under the spectra, the expected value, and the central moments of the random variable up to the fifth order inclusive, where the intensity of the X-ray absorption coefficient was chosen as a random variable. The last characteristics described the center of mass and the moments of inertia, respectively. Using this approach, we reduced the dimension of the space of features from 1500 to 11 and also obtained qualitatively better features of the obtained spectra. Such preliminary processing of the data made it possible to significantly improve the quality of classifying purpose-oriented compound characteristics.

In view of the fact that our dataset was very small, we estimated the quality of the models using the Leave-One-Out method, because this method uses all objects of the dataset to control the quality of predictions of the trained model of the ML:

$$LOO(\mu, X^L) = \frac{1}{L}\sum_{i=1}^{L} Loss\left(\mu\left(X^L \backslash \{x_i\}\right), x_i\right) \to \min.$$

*Prediction of the Oxidation State of Fe Atom*

The problem of predicting the oxidation state of Fe is a problem of binary classification. The scattering graphs were constructed using all pairs of features before solution of this problem. As follows from the obtained graphs (Fig. 4), classes do not penetrate into one another and are linearly separable. Because of the isolation of new features, this problem was solved using all methods under consideration with 100-% classification accuracy.

*Prediction of the Coordination Numbers of Compounds*

The problem of predicting the coordination number is a problem of multiclass classification with six classes. Our preliminary data processing did not guarantee the best isolation of features. The obtained space of features can be redundant for some ML methods. Therefore, principal component analysis (PCA) was also used to improve the quality of classification; it makes it possible to single out features, using which objects of the training dataset are separated in the best way, and omit unnecessary ones. Table 1 contains the accuracies of the results of predicting the coordination number by means of the ML methods under consideration using all features and using ones after the application of PCA. Tables 2 and 3 contain the prediction accuracies for the coordination number in the case of spectra with the same oxidation state of Fe. When working with such a dataset for Fe with the oxidation state +2, the prediction quality increased noticeably. Thus, to solve the problem of analyzing the pre-edge region of the XANES spectra, it is possible to use the cascade of models to consistently predict the oxidation state of Fe and then the coordination number.

The best results were shown by the Extra Trees ensemble method and the multilayer perceptron. It should be noted that the PCA helped to improve the classification quality not for all ML methods. The

**Table 1.** Classification accuracy for the coordination number of the entire dataset by means of the complete set of parameters and after the application of principal component analysis (PCA)

| Machine-learning method | Classification accuracy using all features | Classification accuracy after PCA |
|---|---|---|
| Decision Tree | 0.58 | 0.61 |
| ExtraTrees | 0.55 | 0.65 |
| SVM | 0.75 | 0.66 |
| Logistic Regression | 0.60 | 0.61 |
| MLP | 0.66 | 0.75 |

**Table 2.** Classification accuracy for the coordination numbers of compounds with the oxidation state of Fe of +2 by means of the complete set of parameters and after PCA application

| Machine-learning method | Classification accuracy using all features | Classification accuracy after PCA |
|---|---|---|
| Decision Tree | 0.66 | 0.70 |
| ExtraTrees | 0.86 | 0.96 |
| SVM | 0.83 | 0.80 |
| Logistic Regression | 0.86 | 0.86 |
| MLP | 0.90 | 0.93 |

**Table 3.** Classification accuracy for the coordination numbers of compounds with the oxidation state of Fe +3 by means of the complete set of parameters and after PCA application

| Machine learning method | Classification accuracy using all features | Classification accuracy after PCA |
|---|---|---|
| Decision Tree | 0.33 | 0.60 |
| ExtraTrees | 0.16 | 0.56 |
| SVM | 0.60 | 0.40 |
| Logistic Regression | 0.36 | 0.43 |
| MLP | 0.60 | 0.68 |

SVM showed the best results when training by means of features without application of the PCA.

## CONCLUSIONS

In this paper, we studied the possibilities of the ML methods for analyzing the pre-edge region of the XANES spectra. To do this, we have constructed 60 theoretical spectra in the pre-edge region of the X-ray absorption spectra within the framework of the ab initio method of multiplet ligand field theory for compounds with different coordination numbers, oxidation state, and interatomic distances between the central Fe atom and the nearest O atoms. We have used different ML methods to solve problems of determining the oxidation state of Fe and coordination number using the obtained spectra. When solving problems of determining the oxidation state, all the methods showed 100-% classification quality. When solving the problem of determining the coordination number, we considered two approaches:

(1) the problem was solved for the complete dataset; we obtained a classification accuracy of 75%;

(2) the problem was solved independently for two halves of the dataset with a fixed oxidation state of Fe; we obtained an accuracy of 96% for the subdataset with $Fe^{2+}$ and that of 68% for the subdataset with $Fe^{3+}$.

In the future, we plan to calculate the training dataset with a larger size to improve the classification quality and solve other problems, in particular, the problem of determining geometrical parameters of a compound, such as the distances and angles between atoms in the compound, and also use this approach to other classes of compounds.

## FUNDING

The work was supported by the grant of the President of the Russian Federation for young scientists (MK-2730.2019.2).

## REFERENCES

1. C. Koeberl, in *Large Meteorite Impacts and Planetary Evolution,* Ed. by B. O. Dressler, R. A. F. Grieve, and V. L. Sharpton (Geol. Soc. Am., Boulder, CO, 1994), pp. 133–152.
2. N. Artemieva, in *Catastrophic Events Caused by Cosmic Objects,* Ed. by V. V. Adushkin and I. V. Nemchinov (Springer, Berlin, 2008), pp. 267–289. https://doi.org/10.1007/978-1-4020-6452-4_8
3. R. Moretti and G. Ottonello, J. Non-Cryst. Solids **323**, (2003). https://doi.org/10.1016/S0022-3093(03)00297-7
4. O. A. Lukanin and A. A. Kadik, Geochem. Int. **45**, 857 (2007). https://doi.org/10.1134/S0016702907090029
5. Y. Joly and S. Grenier, in *X-Ray Absorption and X-Ray Emission Spectroscopy: Theory and Applications,* Ed. by J. A. van Bokhoven and C. Lamberti (Wiley, Chichester, 2016).
6. G. Bunker, *Introduction to XAFS: A Practical Guide to X-Ray Absorption Fine Structure Spectroscopy* (Cambridge University Press, Cambridge, 2011).
7. M. Wilke, F. Farges, P.-E. Petit, Jr., G. E. Brown, and F. Martin, Am. Mineral. **86**, 714 (2001). https://doi.org/10.2138/am-2001-5-612

8. G. Giuli, G. Pratesi, C. Cipriani, and E. Paris, Geochim. Cosmochim. Acta **66**, 4347 (2002).
9. G. Giuli, S. G. Eeckhourt, E. Paris, C. Koeberl, and G. Pratesi, Meteorit. Planet. Sci. **40**, 1575 (2005).
10. A. N. Kravtsova, L. V. Guda, A. A. Guda, A. L. Trigub, D. D. Badyukov, and A. V. Soldatov, Radiat. Phys. Chem. **175**, 17 (2020). https://doi.org/10.1016/j.radphyschem.2018.12.017
11. K. Ghosh, A. Stuke, M. Todorovic, P. B. Jorgensen, M. N. Schmidt, A. Vehtari, and P. Rinke, Adv. Sci. **6**, 1801367 (2019). https://doi.org/10.1002/advs.201801367
12. M. Cianciosa, K. J. H. Law, E. H. Martin, and D. L. Green, J. Quant. Spectrosc. Radiat. Transfer **240**, 37 (2020). https://doi.org/10.1016/j.jqsrt.2019.106671
13. J. Padarian, B. Minasny, and A. B. McBratney, Geoderma Region. **16**, 56 (2019). https://doi.org/10.1016/j.geodrs.2018.e00198
14. A. Martini, E. Borfecchia, K. A. Lomachenko, I. A. Pankin, C. Negri, G. Berlier, P. Beato, H. Falsig, S. Bordiga, and C. Lamberti, Chem. Sci. **8**, 6836 (2017). https://doi.org/10.1039/C7SC02266B
15. L. Braglia, E. Borfecchia, L. Maddalena, S. Oien, K. A. Lomachenko, A. L. Bugaev, S. Bordiga, A. V. Soldatov, K. P. Lillerud, and C. Lamberti, Catal. Today **283**, 89 (2017). https://doi.org/10.1016/j.cattod.2016.02.039
16. Y. Tulchinsky, C. H. Hendon, K. A. Lomachenko, E. Borfecchia, B. C. Melot, M. R. Hudson, J. D. Tarver, M. D. Korzynski, A. W. Stubbs, J. J. Kagan, C. Lamberti, C. M. Brown, and M. Dinca, J. Am. Chem. Soc. **139**, 5992 (2017). https://doi.org/10.1021/jacs.7b02161
17. 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, and C. Lamberti, Catal. Today **283**, 119 (2017). https://doi.org/10.1016/j.cattod.2016.02.065
18. 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 **122**, 12029 (2018). https://doi.org/10.1021/acs.jpcc.7b11473
19. M. Chatzidakis and G. A. Botton, Sci. Rep. **9**, 2126 (2019). https://doi.org/10.1038/s41598-019-38482-1
20. P. Torrione, L. M. Collins, and K. D. Morton, in *Laser Spectroscopy for Sensing*, Ed. by M. Baudelet (Woodhead, Cambridge, 2014), p. 125. https://doi.org/10.1533/9780857098733.1.125
21. J. Timoshenko, A. Anspoks, A. Cintins, A. Kuzmin, J. Purans, and A. I. Frenkel, Phys. Rev. Lett. **120**, 225502 (2018). https://doi.org/10.1103/PhysRevLett.120.225502
22. C. Zheng, K. Mathew, C. Chen, Y. M. Chen, H. M. Tang, A. Dozier, J. J. Kas, F. D. Vila, J. J. Rehr, L. F. J. Piper, K. A. Persson, and S. P. Ong, Comput. Mater. **4**, 12 (2018). https://doi.org/10.1038/s41524-018-0067-x
23. M. W. Haverkort, M. Zwierzycki, and O. K. Andersen, Phys. Rev. B: Condens. Matter Mater. Phys. **85**, 165113 (2012). https://doi.org/10.1103/PhysRevB.85.165113
24. J. Kunes, R. Arita, P. Wissgott, A. Toschi, H. Ikeda, and K. Held, Comput. Phys. Commun. **180**, 1888 (2010). https://doi.org/10.1016/j.cpc.2010.08.005
25. A. A. Mostofi, J. R. Yates, Y. S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, Comput. Phys. Commun. **178**, 685 (2008).
26. A. A. Mostofi, J. R. Yates, G. Pizzi, Y. S. Lee, I. Souzo, D. Vanderbilt, and N. Marzari, Comput. Phys. Commun. **185**, 2309 (2014). https://doi.org/10.1016/j.cpc.2014.05.003
27. M. Kobayashi, H. Niwa, Y. Takeda, A. Fujimori, Y. Senba, H. Ohashi, A. Tanaka, S. Ohya, P. N. Hai, M. Tanaka, Y. Harada, and M. Oshima, Phys. Rev. Lett. **112**, 107203 (2014). https://doi.org/10.1103/PhysRevLett.112.107203
28. A. Tanaka and T. Jo, J. Phys. Soc. Jpn. **64**, 2248 (1995).
29. K. Schwarz and P. Blaha, Comput. Mater. Sci. **28**, 259 (2003). https://doi.org/10.1016/s0927-0256(03)00112-5
30. P. Geurts, D. Ernst, and L. Wehenkel, Mach. Learn. **63**, 3 (2006). https://doi.org/10.1007/s10994-006-6226-1
31. M. M. Najafabadi, T. M. Khoshgoftaar, F. Villanustre, and J. Holt, J. Big Data **4**, 22 (2017). https://doi.org/10.1186/s40537-017-0084-5
32. A. Martini, S. A. Guda, A. A. Guda, G. Smolentsev, A. Algasov, O. Usoltsev, M. A. Soldatov, A. L. Bugaev, Yu. Rusalev, C. Lamberti, and A. V. Soldatov, Comput. Phys. Commun. **250**, 107064 (2020). https://doi.org/10.1016/j.cpc.2019.107064

*Translated by L. Kulman*