www.nature.com/scientificdata

scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Database of ab initio L-edge X-ray absorption near edge structure

Yiming Chen[1], Chi Chen[1], Chen Zheng[1], Shyam Dwaraknath[2], Matthew K. Horton[2], Jordi Cabana[3], John Rehr[4], John Vinson[5], Alan Dozier[6], Joshua J. Kas[4], Kristin A. Persson[7,8] & Shyue Ping Ong[1] ✉

The L-edge X-ray Absorption Near Edge Structure (XANES) is widely used in the characterization of transition metal compounds. Here, we report the development of a database of computed L-edge XANES using the multiple scattering theory-based FEFF9 code. The initial release of the database contains more than 140,000 L-edge spectra for more than 22,000 structures generated using a high-throughput computational workflow. The data is disseminated through the Materials Project and addresses a critical need for L-edge XANES spectra among the research community.

Background & Summary

X-ray absorption spectroscopy (XAS) is a robust and valuable characterization technique for accurate identifications of atomic local environments[1], oxidation states[2], and electronic structures[3], etc. In XAS, core electrons are excited after the absorption of X-ray photons. The edge names are based on the principal quantum number of the electrons excited: K for 1 s, L for 2 s or 2p,etc. The XAS can be further divided into the X-ray absorption near edge structure (XANES) and the extended X-ray absorption fine structure (EXAFS) by energy range. The XANES region often has stronger signal and can be directly correlated with atomic local environments and oxidation states of the absorbing atoms by comparing to known reference spectra. Unfortunately, the availability of reference XANES data is typically less than other X-ray based characterization techniques due to its reliance on high-energy synchrotron sources.

Researchers confront such deficiencies by gathering data from the community to construct calibrated databases[4]. For example, the EELS database (EELS DB)[5] hosted by the European Microscopy Society contains several hundred spectra covering 35 elements. More recently, Diamond Light Source has integrated data processing and collection to improve traceability[6]. In addition, researchers are establishing international collaborations to further enhance XAS data distribution and sharing[6,7].

As a complementary alternative to experimental XAS databases, computational XAS databases are attracting increasing interest. Codes based on multiple scattering[8,9], multiplet[10] and Bethe-Salpeter equation[11–14] methodologies are commonly used in computing the XAS. Unlike semi-empirical multiplet calculations that impose strong restrictions on symmetry options, *ab initio* techniques based on the Bethe-Salpeter equation and multiple scattering are able to account for deviations from the idealized coordination environment around the absorbing atom. Previously, the present authors have used the FEFF9 multiple scattering code[8] to generate the world's largest computational database of K-edge XANES spectra, the XASdb[15,16]. The XASdb currently hosts more than 500,000 K-edge XANES spectra for more than 51,000 materials, while most experimental databases only contain several hundred. The XASdb has already been extensively used in many works, including the development of machine learning models to accelerate the interpretation of XAS. For example, Zheng *et al.*[17] applied random forest models to predict coordination environment labels using XANES spectra from the XASdb. Andrejevic *et al.*[18] demonstrated that neutral network-based models are capable of classifying topological materials solely from XANES spectra by training the models on XASdb. Both works delivered >80% average accuracy among

[1]Department of Nanoengineering, University of California San Diego, La Jolla, CA, 92093, USA. [2]Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA. [3]Department of Chemistry, University of Illinois at Chicago, Chicago, IL, 60607, USA. [4]Department of Physics, University of Washington, Seattle, WA, 98195, USA. [5]Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, 20899, USA. [6]Health Effects Laboratory Division, National Institute for Occupational Safety and Health, Centers for Disease Control, Cincinnati, OH, 45226, USA. [7]Department of Materials Science and Engineering, University of California Berkeley, Berkeley, CA, 94720, USA. [8]Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA. ✉e-mail: ongsp@eng.ucsd.edu

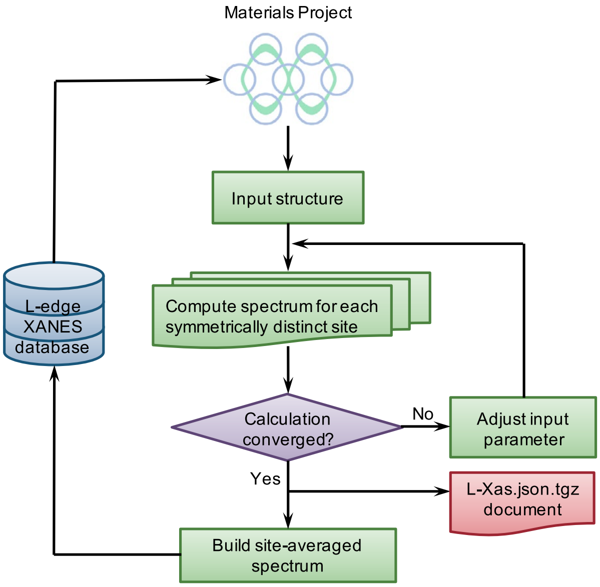


Fig. 1 Schematic diagram of high-throughput workflow for L-edge XANES computation and data processing.

dozens of absorbing elements. Specifically for 3d transition metal elements, researchers have adopted related data from XASdb to predict coordination numbers, Bader charges, mean nearest neighbor distances[19] and local coordination environments[20].

In this work, we further extend the XASdb by developing a database of L-edge XANES using the FEFF9[8] code. The L-edge is especially useful in the study of transition metal elements where the K-edge energies are too high for facile measurements. For instance, the L-edge XANES has been routinely used to investigate how strain, ligand, and particle size affect the performance of catalysts[21]. The L-edge XANES can be further divided into L_1 and $L_{2,3}$ edges. The L_1-edge representing quadrupole transitions is typically broadened, weak and lacks significant information for interpretation[22]. Thus our focus is on the $L_{2,3}$-edge derived from intense dipole-allowed transitions. Our L-edge XANES database covers around 140,000 $L_{2,3}$-edge XANES for more than 22,000 structures. It has been integrated as part of the XASdb hosted on the Materials Project[23] website for individual spectrum visualization and download. A complete set of L-edge data can be downloaded separately in JSON format. This L-edge dataset serves as a reliable repository of reference spectra and enables machine learning applications using L-edge data.

Methods

Theory. In the perspective of computational XAS, Fermi's golden rule governs the transition probability for a given initial state $|i\rangle$ and final state $|f\rangle$. This transition rate is proportional to X-ray absorption coefficient, and can be expressed in terms of the wavefunction as follows:

$$\mu(E) \propto \sum_f |\langle\psi_i|A(r)\cdot p\,|\psi_f\rangle|^2\delta(E-E_f) \tag{1}$$

where ψ_i and ψ_f are the initial and final eigenstates, and $A(r)\cdot p$ accounts for the coupling of X-ray field. The $L_{2,3}$-edge XANES is due to 2p-to-*n*d transitions.

If written in terms of the one-particle Green's function, μ can be calculated in a more efficient way using the following equation:

$$\mu(E) = -\frac{1}{\pi}Im\langle i|\hat{\varepsilon}\cdot r' G(r, r; E)\hat{\varepsilon}\cdot r|i\rangle \tag{2}$$

where the $\hat{\varepsilon}\cdot r$ is the interaction operator in dipole approximation. $G(r, r; E)$ can be further divided into the contribution from the central atom, G^c and the contribution from the remaining scatters G^{sc} (see ref. [8] and references therein). This theoretical formalism is implemented in the *ab-initio* real-space multiple-scattering XAS code FEFF9 for general calculations of X-ray spectra throughout the periodic table[8]. For a more thorough background in the theory, we direct the interested readers to the review paper by Rehr *et al.*[24].

High-throughput workflow. A high-throughput workflow for L-edge calculations was developed using pymatgen[25], FireWorks[26] and atomate[27], as shown in Fig. 1.

The initial input structure is obtained from relaxed structures in the Materials Project[23] database. For each symmetrically distinct site within the structure, a site-wise spectrum is computed with automatic convergence checks and error recovery using the custodian package[28]. Then the raw site-wise spectra are averaged

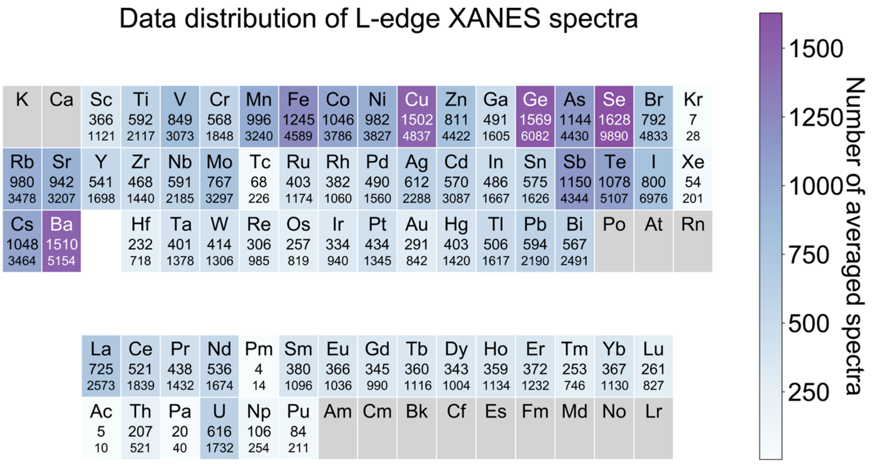


Fig. 2 Data distribution for the L-edge XANES database. The upper and lower numbers for each element correspond to the number of site-averaged (upper) and site-wise count (lower), respectively. One site-averaged count refers to a crystal structure where both its corresponding L_2 and L_3-edge site-wise XANES spectra are complete. One site-wise count refers to either L_2 or L_3 spectrum for a single absorbing site. The figure is colored in terms of the site-averaged count.

to element-wise spectra based on site multiplicity and inserted into the XASdb, which can be accessed via the Materials Project website.

At the present moment, only structures in the Materials Project that originated from the Inorganic Crystal Structure Database (ICSD)[29], i.e., likely to have been experimentally synthesized previously, have been calculated. Further, only elements with atomic number larger than 20 are included because the characterization of lighter elements are typically based on the K-edge XANES.

Data analysis. The cosine similarity is used to evaluate the similarity between spectra quantitatively[30]. It is defined in Eq. 3 as:

$$\text{Cosine similarity}\,(\mathrm{A},\mathrm{B}) = \frac{\overrightarrow{I_A}\cdot\overrightarrow{I_B}}{\|\overrightarrow{I_A}\|\,\|\overrightarrow{I_B}\|} \tag{3}$$

where $\overrightarrow{I_A}$ and $\overrightarrow{I_B}$ represent intensity vectors for the two spectra. It was computed using the scikit-learn[31] package.

Data Records

To date, approximately 140,000 site-wise L_2 and L_3 spectra have been computed, which correspond to 40,000 site-averaged $L_{2,3}$-edge XANES spectra for unique crystals. The data distribution for the L-edge XANES dataset is shown in Fig. 2.

Data visualization. The L-edge data were integrated with the previous K-edge data for a comprehensive display of XAS spectra on Materials Project website (https://materialsproject.org/xas). The website allows users to compare the XANES spectra interactively for K, $L_{2,3}$, L_2 and L_3-edge XANES spectra although $L_{2,3}$ functionality is only supported for absorbing elements with $21 \leq Z \leq 30$. Users can also apply Gaussian smoothing to the spectra on-the-fly to account for instrumental broadening.

Data download. The dataset can be downloaded in two ways. For users interested in specific chemical systems (e.g., searching for reference spectra within Ni-O system), an API query (https://api.materialsproject.org/docs#/XAS) through the representational state transfer (REST) endpoint interfaced with the XASdb is recommended. The data obtained through API are site-averaged spectra, which can be directly compared to experimental spectra. The second method is to download a L-XAS.json.tgz file for a complete set of site-wise L-edge data through figshare (https://doi.org/10.6084/m9.figshare.12824513.v1)[32]. The complete dataset is appropriate for data-intensive applications such as machine learning. The detailed data structure is shown in Table 1.

Technical Validation

We performed validation of FEFF9's input parameters based on a comparison between computational and experimental spectra. Unlike the previous K-edge database works[15,16] that covered a more thorough benchmark on FEFF9 input parameters, only critical input parameters that shape the L-edge spectrum, such as the cluster radius in self-consistent field (SCF) calculations and the core-hole treatment, are discussed here. A more detailed discussion about the full multiple scattering (FMS) calculations and the choice of exchange correlation potential, which are kept the same for the L-edge calculations, can be found in previous works on the K-edge calculations[15,16]

Property	Description
xas-id	Unique id for each site-wise spectrum in the form of mpid-siteindex-spectrumtype-edge (e.g., mp-24850-1-XANES-L3)
edge	Absorption edge (L_2 or L_3)
mp-id	Unique materials project ID (e.g., mp-24850)
structure	Pymatgen Structure object in JSON format
absorbing atom	Site index in structure (e.g., 0)
input parameters	FEFF9 input parameters in dictionary form
spectrum	Array of shape (2, 100) where the first row is energies in eV and the second row is absorption coefficients in arbitrary unit

Table 1. Data structure for individual entry in L-XAS.json.tgz.

Formula	Absorbing species	Materials id	Space group	Reference
TiO_2	Ti	mp-2657	$P4_2/mnm$	[36]
V_2O_5	V	mp-754670	Pmmn	[37]
$MgCr_2O_4$	Cr	mp-19202	$Fd\bar{3}m$	[38]
Cr_2O_3	Cr	mp-19399	$R\bar{3}c$	[39]
$LiMn_2O_4$	Mn	mp-25015	$Fd\bar{3}m$	[40]
$MgMn_2O_4$	Mn	mp-32006	$I4_1/amd$	[40]
MnO_2	Mn	mp-714975	$P4_2/mnm$	[41]
FeF_3	Fe	mp-22398	$R\bar{3}c$	[42]
$LiFePO_4$	Fe	mp-19017	Pnma	[43]
$LiCoPO_4$	Co	mp-18915	Pnma	[44]
$LiCoO_2$	Co	mp-24850	$R\bar{3}c$	[45]
CoF_2	Co	mp-556520	$P4_2/mnm$	[46]
NiF_2	Ni	mp-559798	$P4_2/mnm$	[46]
$KNiF_3$	Ni	mp-560976	$Pm\bar{3}m$	[47]
Cu	Cu	mp-30	$Fm\bar{3}m$	[48]
CuO	Cu	mp-1692	$P4_2/mmc$	[48]
Pt	Pt	mp-126	$Fm\bar{3}m$	[49]
CeO_2	Ce	mp-20194	$Fm\bar{3}m$	[50]

Table 2. Details of 18 experimental reference spectra.

for interested readers. While more advanced treatments such as Debye-Waller factors can be performed using FEFF9, we did not benchmark them in this high-throughput work. A total of 18 experimental reference spectra were collected and listed in Table 2. We included diverse absorbing species (both 3d transition metal and heavy 4 f absorbing species) and various local environment to unveil a convergent set of inputs.

The cluster radius in SCF calculations determines the number of atoms to be included in the coordination shell. A sufficiently large cluster radius is necessary to account for all scattering effects but an excessively large one will cause an increase in computation cost and an overestimation of multiple scattering from long-distance atoms[33]. From Fig. 3, we found that a cluster radius of 5.5 Å was sufficient to converge the L-edge spectra to a cosine similarity of 0.98 compared to a cluster radius of 8 Å.

Another important input parameter to FEFF9 calculations is the core-hole treatment. When the core electron is excited by X-ray photons, it creates a photon-electron or core-hole pair. Among all three core-hole treatments supported by FEFF9, i.e., the final state rule (*FSR*), the random phase approximation (*RPA*) and no core-hole (*NONE*), it was found that calculated spectra without core-hole treatment resulted in the best agreement (highest cosine similarity) with experiments, as shown in Fig. 3. *FSR* sometimes breaks down for L-shell calculations while *RPA* can cause an irregular shake-up in the post-edge region of the XANES spectrum. Further, the more mobile electrons at outer shells (e.g., 2p vs. 1 s electrons) lead to stronger screening of the core-hole effect than can be accounted for with *FSR* or *RPA*.

In Fig. 4, the spectra calculated using the FEFF9 converged input parameters are compared against experimental spectra as well as spectra computed using ocean[11,12], an alternative software based on the Bethe-Salpeter equation for XANES calculation. A horizontal shift to align the L_3 peak and a Gaussian broadening with full width at half maximum (FWHM) of 1.2 eV were applied to the computed spectra to account for the less accurate determination of Fermi level and the instrumental broadening, respectively. In general, the FEFF9 spectra are in relatively good agreement with experimental measurements and the ocean computed spectra in terms of relative peak positions and peak intensities. For Ti in rutile TiO_2 and V in α-V_2O_5, the L_2 peak intensities and the edge separation computed by ocean are in better agreement with experiments than FEFF9, which can be explained by

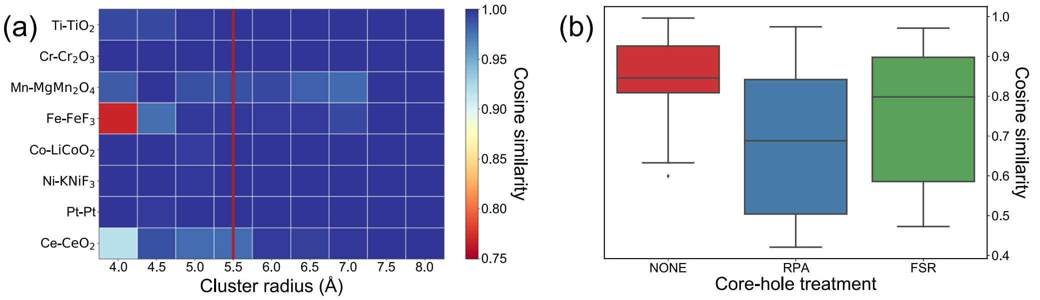
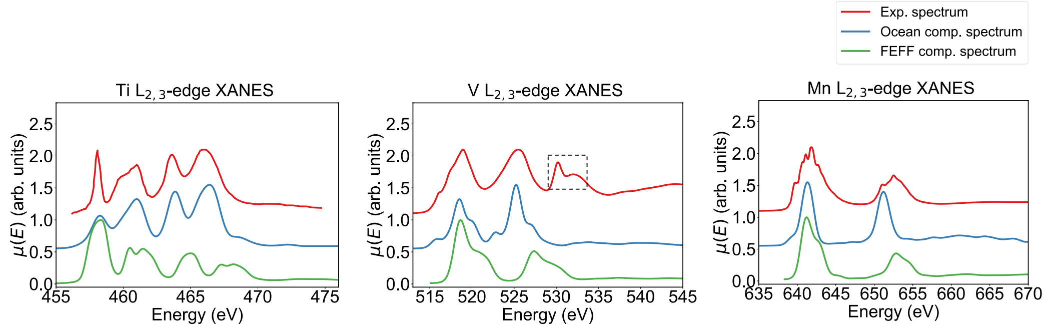
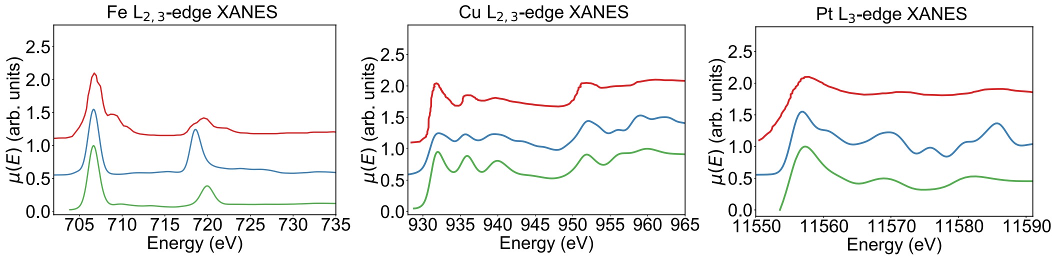


Fig. 3 Benchmarking results for (**a**) cluster radius for SCF, and (**b**) core-hole treatment parameters for FEFF9 L-edge XANES calculations. When benchmarking cluster radius for each chemical system, the cosine similarities are computed with respect to the spectrum that is calculated with a cluster radius of 8 Å. A comparison between $L_{2,3}$-edge XANES is conducted for all absorbing elements except the heavy 5f elements Ce and Pt. For Ce and Pt, only the L_3-edge XANES are compared because of large energy separation between the L_2 and L_3 peaks. For core-hole treatment, the cosine similarities are calculated between the computed spectra and experimental spectra for all chemical systems listed in Table 2.



(a) TiO_2 (b) V_2O_5 (c) $MgMn_2O_4$



(d) $LiFePO_4$ (e) Cu (f) Pt

Fig. 4 Comparison between experimental spectra, ocean computed spectra and FEFF9 computed spectra. The first and second major peaks are the L_3 and L_2 peak, respectively. The third peak in experimental V L-edge spectrum in α-V_2O_5, indicated by the black dashed rectangle, is contributed by the oxygen K-edge, which is not accounted for in the computed spectra. The details of experimental data can be found in Table 2. Spectra are shifted vertically and normalized to maximum intensity for ease of visualization.

the inclusion of spin-orbit coupling during ocean calculation. However, the significantly higher computational costs and poorer scalability of ocean calculations make it less suitable for high-throughput generation of spectra.

The $L_{2,3}$-edge XANES spectra for eight 3d transition metal elements in various crystals are plotted in Fig. 5. These elements are predominantly in the octahedral and tetrahedral local environments[34]. Except for Ti and Mn where the octahedral environment dominates, the rest of the 3d transition metals have a more even distribution of octahedral and tetrahedral local environment. The L_3/L_2 branching ratio, defined as the ratio of integrated intensities for L_3 and L_2 peaks, is vital to the spin states because this property is proportional to the number of paired electrons. In theory, the branching ratio for an absorbing atom in tetrahedral environment is smaller than that in octahedral environment. The crystal field splitting for tetrahedral environment is smaller than that of octahedral environment, leading to a preference for high-spin states over low-spin states in the tetrahedral environment. This is consistent with the observation that the L_2 peak intensities for 3d transition metals such as Mn, Fe, Co, Ni and

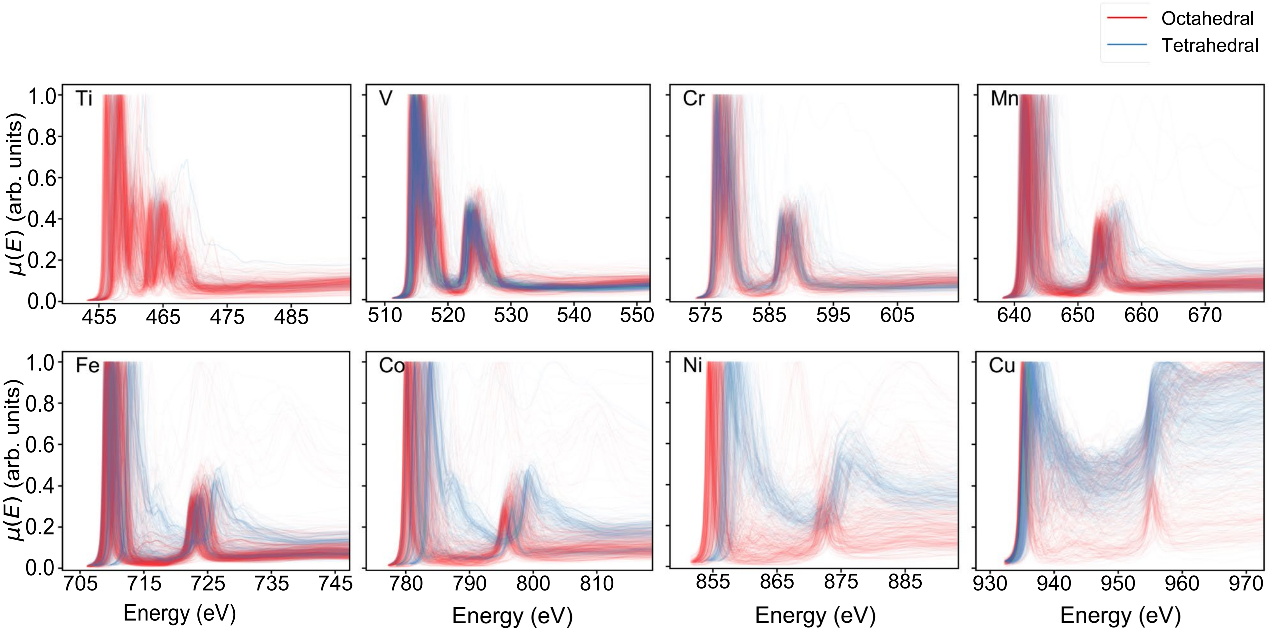


Fig. 5 $L_{2,3}$-edge XANES spectra for 3d transition metal elements. The colors represent the local environments around absorbing atoms where red color refers to octahedral environment and blue is for tetrahedral environment. Spectra are normalized to maximum intensity for ease of visualization.

Cu in tetrahedral environment are significantly larger than in the octahedral environment. Another important finding comes from the decrease of white-line intensities. The white line refers to the intense absorption in the rising edge region of a XANES spectrum. Its intensity is positively related to the number of unoccupied d states. With increasing number of valence states, the white line intensities continue to decrease along with filling of d orbitals. While most FEFF spectra trends are consistent with theory, we observed that a small number of spectra (53 out of around 140,000 site-wise spectra) possess unphysical negative intensities. This could be caused by the instability when directly solving the complex Green's function, which leads to small negative spectra weight since the phase might be wrong. Another possible explanation lies in the numerical precision differences between the atomic background and the scattering contributions, which are calculated separately. While extra caution is needed, it is up to users and use cases to determine how to treat those spectra with negative intensities.

Usage Notes

Each entry in the database contains site-specific or site-averaged L-edge XANES (as well as K-edge XANES), together with physical and chemical properties such as the elemental oxidation states and coordination environments. These data can be used as references for comparison with experimental spectra to identify the properties of species present or local environment when using a very fine electron energy loss spectroscopy (EELS) probe. While there are discrepancies between the computed and experimental spectra for certain elements (especially the early transition metals), this can be mitigated through computations with other high-level codes such as ocean for select systems.

Another potential usage of this dataset is in the development of machine learning (ML) models to accelerate spectra interpretation. ML models such as random forest[17,19] and neural networks[18,20] have been developed using our previously-constructed K-edge XANES database, with accuracies exceeding 80% having been achieved in classifying the local environment. Potentially, similar or more refined ML models, especially for transition metal systems, can be developed using both the K-edge and L-edge XANES. For instance, the branching ratio in $L_{2,3}$-edge XANES can be used to distinguish between different spin states even in the same complexes[35].

In summary, a large dataset covering 140,000 $L_{2,3}$-edge XANES spectra are made open to the public after an in-depth benchmark, complementing the world's largest computational XAS database. Our benchmark results indicate that multiple scattering codes such as FEFF9 can achieve comparable accuracy when compared to experimental data. We anticipate this work to benefit the whole XAS community through well-curated data and elaborated dissemination.

Code availability

The workflow for FEFF9 calculation including input generation, output parsing and workflow management is available in open-source materials science packages pymatgen[25], FireWorks[26] and atomate[27]. The error handler for automatic error detection and recovery can be found in custodian[28].

Received: 6 January 2021; Accepted: 28 April 2021;
Published online: 11 June 2021

References

1. Hudson-Edwards, K. A. *et al.* Origin and Fate of Vanadium in the Hazeltine Creek Catchment following the 2014 Mount Polley Mine Tailings Spill in British Columbia, Canada. *Environmental Science & Technology* **53**, 4088–4098, https://doi.org/10.1021/acs.est.8b06391 (2019).
2. Ravel, B., Scorzato, C., Siddons, D. P., Kelly, S. D. & Bare, S. R. Simultaneous XAFS measurements of multiple samples. *Journal of Synchrotron Radiation* **17**, 380–385, https://doi.org/10.1107/S0909049510006230 (2010).
3. Boillot, M.-L. *et al.* Pressure-induced spin-state crossovers at room temperature in iron(II) complexes: Comparative analysis; a XANES investigation of some new transitionsDedicated to the memory of Professor Olivier Kahn. *New Journal of Chemistry* **26**, 313–322, https://doi.org/10.1039/b104782p (2002).
4. Lytle, F. W., Wei, P. S. P., Greegor, R. B., Via, G. H. & Sinfelt, J. H. Effect of chemical environment on magnitude of x-ray absorption resonance at *LIII* edges. Studies on metallic elements, compounds, and catalysts. *The Journal of Chemical Physics* **70**, 4849–4855, https://doi.org/10.1063/1.437376 (1979).
5. Ewels, P., Sikora, T., Serin, V., Ewels, C. P. & Lajaunie, L. A Complete Overhaul of the Electron Energy-Loss Spectroscopy and X-Ray Absorption Spectroscopy Database: eelsdb.eu. *Microscopy and Microanalysis* **22**, 717–724, https://doi.org/10.1017/s1431927616000179 (2016).
6. Cibin, G. *et al.* An open access, integrated XAS data repository at Diamond Light Source. *Radiation Physics and Chemistry* 108479, https://doi.org/10.1016/j.radphyschem.2019.108479 (2019).
7. Asakura, K., Abe, H. & Kimura, M. The challenge of constructing an international XAFS database. *Journal of Synchrotron Radiation* **25**, 967–971, https://doi.org/10.1107/S1600577518006963 (2018).
8. Rehr, J. J., Kas, J. J., Vila, F. D., Prange, M. P. & Jorissen, K. Parameter-free calculations of X-ray spectra with FEFF9. *Physical Chemistry Chemical Physics* **12**, 5503, https://doi.org/10.1039/b926434e (2010).
9. Bunău, O. & Joly, Y. Self-consistent aspects of x-ray absorption calculations. *Journal of Physics: Condensed Matter* **21**, 345501, https://doi.org/10.1088/0953-8984/21/34/345501 (2009).
10. Stavitski, E. & de Groot, F. M. The CTM4XAS program for EELS and XAS spectral shape analysis of transition metal L edges. *Micron* **41**, 687–694, https://doi.org/10.1016/j.micron.2010.06.005 (2010).
11. Vinson, J., Rehr, J. J., Kas, J. J. & Shirley, E. L. Bethe-Salpeter equation calculations of core excitation spectra. *Physical Review B* **83**, 115106, https://doi.org/10.1103/PhysRevB.83.115106 (2011).
12. Gilmore, K. *et al.* Efficient implementation of core-excitation Bethe–Salpeter equation calculations. *Computer Physics Communications* **197**, 109–117, https://doi.org/10.1016/j.cpc.2015.08.014 (2015).
13. Blaha, P., Schwarz, K., Madsen, G. K., Kvasnicka, D. & Luitz, J. wien2k. An augmented plane wave + local orbitals program for calculating crystal properties. (2001).
14. Gulans, A. *et al.* Exciting: a full-potential all-electron package implementing density-functional theory and many-body perturbation theory. *Journal of Physics: Condensed Matter* **26**, 363202, https://doi.org/10.1088/0953-8984/26/36/363202 (2014).
15. Mathew, K. *et al.* High-throughput computational X-ray absorption spectroscopy. *Scientific Data* **5**, 180151, https://doi.org/10.1038/sdata.2018.151 (2018).
16. Zheng, C. *et al.* Automated generation and ensemble-learned matching of X-ray absorption spectra. *npj Computational Materials* **4**, 12, https://doi.org/10.1038/s41524-018-0067-x (2018).
17. Zheng, C., Chen, C., Chen, Y. & Ong, S. P. Random Forest Models for Accurate Identification of Coordination Environments from X-Ray Absorption Near-Edge Structure. *Patterns* **1**, 100013, https://doi.org/10.1016/j.patter.2020.100013 (2020).
18. Andrejevic, N., Andrejevic, J., Rycroft, C. H. & Li, M. Machine learning spectral indicators of topology. Preprint at https://arxiv.org/abs/2003.00994 (2020).
19. Torrisi, S. B. *et al.* Random forest machine learning models for interpretable X-ray absorption near-edge structure spectrum-property relationships. *npj Computational Materials* **6**, 109, https://doi.org/10.1038/s41524-020-00376-6 (2020).
20. Carbone, M. R., Yoo, S., Topsakal, M. & Lu, D. Classification of local chemical environments from x-ray absorption spectra using supervised machine learning. *Physical Review Materials* **3**, 033604, https://doi.org/10.1103/PhysRevMaterials.3.033604 (2019).
21. Jia, Q. *et al.* Improved Oxygen Reduction Activity and Durability of Dealloyed PtCo$_x$ Catalysts for Proton Exchange Membrane Fuel Cells: Strain, Ligand, and Particle Size Effects. *ACS Catalysis* **5**, 176–186, https://doi.org/10.1021/cs501537n (2015).
22. Baker, M. L. *et al.* K- and L-edge X-ray absorption spectroscopy (XAS) and resonant inelastic X-ray scattering (RIXS) determination of differential orbital covalency (DOC) of transition metal sites. *Coordination Chemistry Reviews* **345**, 182–208, https://doi.org/10.1016/j.ccr.2017.02.004 (2017).
23. Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **1**, 011002, https://doi.org/10.1063/1.4812323 (2013).
24. Rehr, J. J. & Albers, R. C. Theoretical approaches to x-ray absorption fine structure. *Reviews of Modern Physics* **72**, 621–654, https://doi.org/10.1103/RevModPhys.72.621 (2000).
25. Ong, S. P. *et al.* Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* **68**, 314–319, https://doi.org/10.1016/j.commatsci.2012.10.028 (2013).
26. Jain, A. *et al.* FireWorks: a dynamic workflow system designed for high-throughput applications. *Concurrency and Computation: Practice and Experience* **27**, 5037–5059, https://doi.org/10.1002/cpe.3505 (2015).
27. Mathew, K. *et al.* Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows. *Computational Materials Science* **139**, 140–152, https://doi.org/10.1016/j.commatsci.2017.07.030 (2017).
28. Ong, S. P. *et al.* materialsproject/custodian: v2020.4.27. *Zenodo* https://doi.org/10.5281/zenodo.3770422 (2020).
29. Bergerhoff, G., Hundt, R., Sievers, R. & Brown, I. D. The inorganic crystal structure data base. *Journal of Chemical Information and Modeling* **23**, 66–69, https://doi.org/10.1021/ci00038a003 (1983).
30. Guo, M. *et al.* Fingerprinting Electronic Structure of Heme Iron by Ab Initio Modeling of Metal L-Edge X-ray Absorption Spectra. *Journal of Chemical Theory and Computation* **15**, 477–489, https://doi.org/10.1021/acs.jctc.8b00658 (2019).
31. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).
32. Chen, Y. *et al.* Database of ab initio L-edge X-ray absorption near edge structure. *Figshare* https://doi.org/10.6084/m9.figshare.12824513.v1 (2021).
33. Xu, W. *et al.* Copper *L*-edge spectra: multiplet vs. multiple scattering theory. *Journal of Physics: Conference Series* **430**, 012010, https://doi.org/10.1088/1742-6596/430/1/012010 (2013).
34. Zimmermann, N. E. R., Horton, M. K., Jain, A. & Haranczyk, M. Assessing Local Structure Motifs Using Order Parameters for Motif Recognition, Interstitial Identification, and Diffusion Path Characterization. *Frontiers in Materials* **4**, 34, https://doi.org/10.3389/fmats.2017.00034 (2017).
35. Kubin, M. *et al.* Cr L-Edge X-ray Absorption Spectroscopy of $Cr^{III}(acac)_3$ in Solution with Measured and Calculated Absolute Absorption Cross Sections. *The Journal of Physical Chemistry B* **122**, 7375–7384, https://doi.org/10.1021/acs.jpcb.8b04190 (2018).
36. Laskowski, R. & Blaha, P. Understanding the $L_{2,3}$ x-ray absorption spectra of early 3*d* transition elements. *Physical Review B* **82**, 205104, https://doi.org/10.1103/PhysRevB.82.205104 (2010).
37. Andrews, J. L. *et al.* Reversible Mg-Ion Insertion in a Metastable One-Dimensional Polymorph of V_2O_5. *Chem* **4**, 564–585, https://doi.org/10.1016/j.chempr.2017.12.018 (2018).
38. Hu, L. *et al.* Tailoring the electrochemical activity of magnesium chromium oxide towards Mg batteries through control of size and crystal structure. *Nanoscale* **11**, 639–646, https://doi.org/10.1039/C8NR08347A (2019).

39. Dedkov, Y. S. *et al*. Correlations in the electronic structure of half-metallic ferromagnetic CrO_2 films: An x-ray absorption and resonant photoemission spectroscopy study. *Physical Review B* **72**, 060401, https://doi.org/10.1103/PhysRevB.72.060401 (2005).
40. Kim, C. *et al*. Multivalent Electrochemistry of Spinel $Mg_x Mn_{3-x} O_4$ Nanocrystals. *Chemistry of Materials* **30**, 1496–1504, https://doi.org/10.1021/acs.chemmater.7b03640 (2018).
41. Gilbert, B. *et al*. Multiple Scattering Calculations of Bonding and X-ray Absorption Spectroscopy of Manganese Oxides. *The Journal of Physical Chemistry A* **107**, 2839–2847, https://doi.org/10.1021/jp021493s (2003).
42. Krasnikov, S. A. *et al*. Electronic structure of FeF_2 and FeF_3 studied by x-ray absorption and fluorescence spectroscopy. *Physica Scripta* **T115**, 1074–1076, https://doi.org/10.1088/0031-8949/2005/T115/326 (2005).
43. Yang, S. *et al*. Soft X-ray XANES studies of various phases related to $LiFePO_4$ based cathode materials. *Energy & Environmental Science* **5**, 7007, https://doi.org/10.1039/c2ee03445j (2012).
44. Lapping, J. G. *et al*. Changes in Electronic Structure upon Li Deintercalation from $LiCoPO_4$ Derivatives. *Chemistry of Materials* **30**, 1898–1906, https://doi.org/10.1021/acs.chemmater.7b04739 (2018).
45. Hong, L. *et al*. Electronic Structure of $LiCoO_2$ Surfaces and Effect of Al Substitution. *The Journal of Physical Chemistry C* **123**, 8851–8858, https://doi.org/10.1021/acs.jpcc.8b11661 (2019).
46. Olalde-Velasco, P., Jiménez-Mier, J., Denlinger, J. & Yang, W.-L. Atomic multiplets at the $L_{2,3}$ edge of 3d transition metals and the ligand K edge in x-ray absorption spectroscopy of ionic systems. *Physical Review B* **87**, 245136, https://doi.org/10.1103/PhysRevB.87.245136 (2013).
47. Plews, M. R. *et al*. Synthesis and X-ray absorption spectroscopy of potassium transition metal fluoride nanocrystals. *CrystEngComm* **21**, 135–144, https://doi.org/10.1039/C8CE01349G (2019).
48. Jiang, P. *et al*. Experimental and theoretical investigation of the electronic structure of Cu_2O and CuO thin films on Cu(110) using x-ray photoelectron and absorption spectroscopy. *The Journal of Chemical Physics* **138**, 024704, https://doi.org/10.1063/1.4773583 (2013).
49. Miller, T. E. *et al*. Actualizing *In Situ* X-ray Absorption Spectroscopy Characterization of PEMFC-Cycled Pt-Electrodes. *Journal of The Electrochemical Society* **165**, F597–F603, https://doi.org/10.1149/2.0151809jes (2018).
50. Francisco, M. S. P., Nascente, P. A. P., Mastelaro, V. R. & Florentino, A. O. X-ray photoelectron spectroscopy, x-ray absorption spectroscopy, and x-ray diffraction characterization of CuO–TiO_2–CeO_2 catalyst system. *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films* **19**, 1150–1157, https://doi.org/10.1116/1.1345911 (2001).

Acknowledgements

This work was intellectually led by the Data Infrastructure Building Blocks (DIBBS) Local Spectroscopy Data Infrastructure (LSDI) project funded by National Science Foundation (NSF), under Award Number 1640899. We also acknowledge the computational resources provided by Triton Shared Computing Cluster (TSCC) at University of California, San Diego and NSF DIBBS funding. Additional support was provided by the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division under contract no. DE-AC02-05CH11231 (Materials Project program KC23MP) at Lawrence Berkeley National Laboratory. The FEFF project is supported by DOE Office of Science BES Grant DE-FG02-97ER45623. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the National Institute for Occupational Safety and Health. Mention of product or company name does not constitute endorsement by the Centers for Disease Control and Prevention. Certain software is identified in this paper in order to specify the procedure adequately. Such identification is not intended to imply recommendation or endorsement by National Institute of Standards and Technology.

Author contributions

Y.C., C.C. and C.Z. designed the workflow, performed calculation and analyzed result. S.D., M.H. and K.A.P. disseminated the data through Materials Project website. J.C. provided experimental data for benchmark. J.J.R., J.V., A.D. and J.J.K. provided theoretical support. Y.C. and C.C. wrote the manuscript. S.P.O. is the primary investigator of this work. All authors contributed to the discussion and writing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.P.O.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

The Creative Commons Public Domain Dedication waiver http://creativecommons.org/publicdomain/zero/1.0/ applies to the metadata files associated with this article.

© The Author(s) 2021