

# UC Berkeley

## UC Berkeley Previously Published Works

**Title**

High-throughput computational X-ray absorption spectroscopy.

**Permalink**

<https://escholarship.org/uc/item/1gm0h2jt>

**Journal**

Scientific data, 5(1)

**ISSN**

2052-4463

**Authors**

Mathew, Kiran  
Zheng, Chen  
Winston, Donald  
et al.

**Publication Date**

2018-07-01

**DOI**

10.1038/sdata.2018.151

Peer reviewed

# SCIENTIFIC DATA

OPEN

## Data Descriptor: High-throughput computational X-ray absorption spectroscopy

Kiran Mathew<sup>1,\*</sup>, Chen Zheng<sup>2,\*</sup>, Donald Winston<sup>3</sup>, Chi Chen<sup>2</sup>, Alan Dozier<sup>4</sup>, John J. Rehr<sup>5</sup>, Shyue Ping Ong<sup>2</sup> & Kristin A. Persson<sup>1</sup>

Received: 11 December 2017

Accepted: 18 May 2018

Published: 31 July 2018

X-ray absorption spectroscopy (XAS) is a widely-used materials characterization technique. In this work we present a database of computed XAS spectra, using the Green's formulation of the multiple scattering theory implemented in the FEFF code. With more than 500,000 K-edge X-ray absorption near edge (XANES) spectra for more than 40,000 unique materials, this database constitutes the largest existing collection of computed XAS spectra to date. The data is openly distributed via the Materials Project, enabling researchers across the world to access it for free and use it for comparisons with experiments and further analysis.

Design Type(s)	database creation objective
Measurement Type(s)	near-edge X-ray absorption fine structure spectrum
Technology Type(s)	computational modeling technique
Factor Type(s)	chemical entity
Sample Characteristic(s)	

<sup>1</sup>Department of Materials Science, University of California Berkeley, Berkeley, CA 94720, USA. <sup>2</sup>Department of Nanoengineering, University of California San Diego, La Jolla, CA 92093, USA. <sup>3</sup>Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. <sup>4</sup>Division of Applied Research and Technology, National Institute for Occupational Safety and Health, Centers for Disease Control, Cincinnati, OH 45226, USA. <sup>5</sup>Department of Physics, University of Washington, Seattle, WA 98195, USA. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to S.P.O. (email: ongs@eng.ucsd.edu) or to K.A.P. (email: kapersson@lbl.gov).

## Background & Summary

A crucial step in the process of novel materials discovery is the characterization of the synthesized material. There exists a wide array of tools and spectroscopic techniques that are used in the material identification process, e.g. X-ray diffraction (XRD), X-ray emission spectroscopy (XES), and X-ray absorption spectroscopy (XAS). XAS is widely-employed in the characterization of the local structural environment surrounding select elements within a material.

Great progress has been made over the past few years in the development of laboratory-based X-ray spectrometers for high-resolution x-ray absorption near edge structure (XANES) and X-ray emission spectroscopy (XES)<sup>1</sup>. The availability of relatively inexpensive laboratory-based XAFS system (<http://easyxafs.com/>) and third generation synchrotron facilities<sup>2</sup> have established the groundwork for the broad application of high-resolution XAS in characterization of materials. On the other hand, modern computational resources and methodologies have reached a level of maturity and efficiency to complement as well as to fast-track new discoveries. In the case of XAS, a variety of theoretical frameworks including time-dependent density-functional theory (TDDFT)<sup>3,4</sup>, multiple-scattering<sup>5</sup>, and Bethe-Salpeter equation (BSE) based approaches<sup>6</sup> have been implemented, each exhibits its advantages and drawbacks. Leveraging spectroscopic simulation software with large crystal structure databases enables the computation of a large number of reliable theoretical spectra corresponding to well-defined crystal structures<sup>7</sup>, providing a broad reference dataset with clean unique structural fingerprints that can be used for identification purposes. With the help of carefully crafted software tools, these computations can be performed in a high-throughput fashion and can be used to scan the structural phase space for novel materials. In addition, through proper integration with modern database tools, these scans can be saved for future use and leveraged for training machine learning algorithms to assist the characterization process. Some examples of such publicly available spectroscopic database are the EELS Data Base<sup>8</sup>, a compilation of valence and core-loss spectra from EELS and XAS experiments containing 271 spectra that covers 39 elements of the periodic table, and XCOM (<https://www.nist.gov/pml/xcom-photon-cross-sections-database>), which provides photon cross sections for scattering, photoelectric absorption and pair production, as well as total attenuation coefficients, for any element, compound or mixture. Other existing XAS databases<sup>9,10</sup>, i.e. <https://www.cat.hokudai.ac.jp/catdb/> and <http://cars.uchicago.edu/xaslib>, covering a few hundred spectra, are hosted across the world and serve as valuable references for analysis.

The FEFF framework affords relatively inexpensive calculations compared to other approaches and requires minimum adjustable parameters. It provides an efficient means of generating high quality XAS spectra for a larger amount of chemical systems and structures. Hence, in our study, we selected the FEFF9 (ref. 5) program for the *ab initio* calculation of K-edge X-ray absorption near edge spectra (XANES). Using the parameter settings obtained from recent benchmarking work against experimental spectra<sup>11</sup> and the FEFF workflow infrastructure available in the open source materials science workflow package Atomate<sup>12</sup>, we generate spectra of all the materials available in the publicly accessible and widely used materials database, Materials Project (MP)<sup>7</sup>.

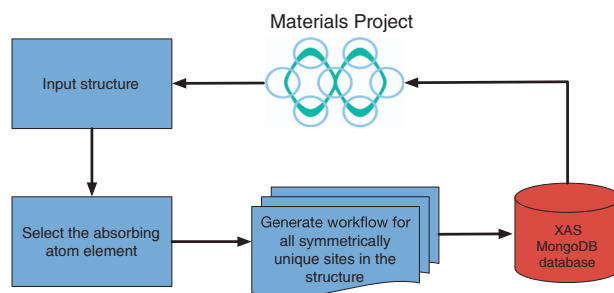
A comprehensive database of computed XAS spectra enables comparison between different spectroscopic signatures across chemical systems and structures such that rapid determination of oxidation states, coordination environment, and other local atomic structure information can be obtained. Furthermore, using matching algorithms<sup>11</sup> or other machine learning methods<sup>13</sup>, the data can be leveraged for on-the-fly characterization. Though the peak positions and amplitudes of the computational K-edge XANES spectral may exhibit differences compared to experimental spectra, theoretically computed XANES spectra provide sufficient information to identify oxidation state and coordination chemistry of the probe atom, and can be highly useful when experimental data are not available or scarce. For example, a previous study by Timoshenko *et al.*<sup>14</sup> showed that *ab initio* XANES spectra provide excellent input data for training supervised machine learning models aimed at reconstructing metal catalyst structures from their experimental XANES. The current authors have also shown in a previous study<sup>11</sup> that an ensemble-learned algorithm to match experimental K-edge XANES spectra in the EELS Data Base to computed spectra can achieve nearly 80% accuracy in identifying the correct oxidation state and coordination environment. In addition, the data is associated with download options and programmatic analyses tools for each structure in the Materials Project database, thereby making it accessible to the broader materials science community. Furthermore, the MP web application enables users to select spectra from the database, upload experimental spectra data and predict the material composition using the matching tool. To date, this is the largest computed XAS dataset available and it is still expanding.

The paper is organized as follows; first we briefly describe the XAS computation methodology as implemented in the FEFF code, and thereafter the high-throughput framework used in the generation of the spectra. We then describe the data storage and dissemination details, followed by the technical validation of the computational methodology and the high-throughput framework.

## Methods

### Theory

The K-edge XANES spectra were computed using the FEFF<sup>5</sup> code which employs the Green's formulation of the multiple scattering theory to compute the spectra<sup>5</sup>. The X-ray absorption  $\mu$  is computed in a manner similar to Fermi's golden rule when written in terms of the projected photoelectron density of



**Figure 1.** Schematic diagram of the high throughput framework employed in the generation of XAS spectra for the Materials Project.

final states or the imaginary part of the one-particle Green's function,  $G(r, r'; E)$ . In terms of the Green's function,  $G(r, r'; E)$ , the absorption coefficient,  $\mu$ , from a given core level  $c$  is given by ref. 15.

$$\mu = -\frac{1}{\pi} \text{Im} \langle c | \epsilon \cdot r G(r, r'; E) \epsilon \cdot r | c \rangle \quad (1)$$

with the Green's function,  $G(r, r'; E)$  given by

$$G(r, r'; E) = \sum_f \frac{\psi_f(r) \psi_f(r')^*}{E - E_f + i\Gamma} \quad (2)$$

where  $\psi_f$  are the final states, with associated energies  $E_f$  and net lifetime  $\Gamma$ , of a one-particle Hamiltonian that includes an optical potential with appropriate core hole screening.

The FEFF code computes the full propagator  $G$  incrementally using matrix factorization and uses the spherical muffin-tin approximation for the scattering potential<sup>15</sup>. For a more detailed description, we direct the readers to the review paper by Rehr *et al.*<sup>15</sup>

### High-throughput Workflow

For the high-throughput XAS spectra generation, we use the FEFF workflow (Fig. 1) available in the open source computational materials science workflow package Atomate<sup>12</sup>. Atomate provides a high-level interface to compose workflows using open source materials science softwares such as pymatgen<sup>16</sup>, FireWorks<sup>17</sup> and Custodian (<https://github.com/materialsproject/custodian>).

Each FEFF calculation involves the following 3 steps:

- Selection of the absorbing site and the cluster of atoms to be included in the scattering calculations.
- Generation of the FEFF input files for each site and its surrounding atomic cluster.
- Execution of the FEFF binary on the generated input files.

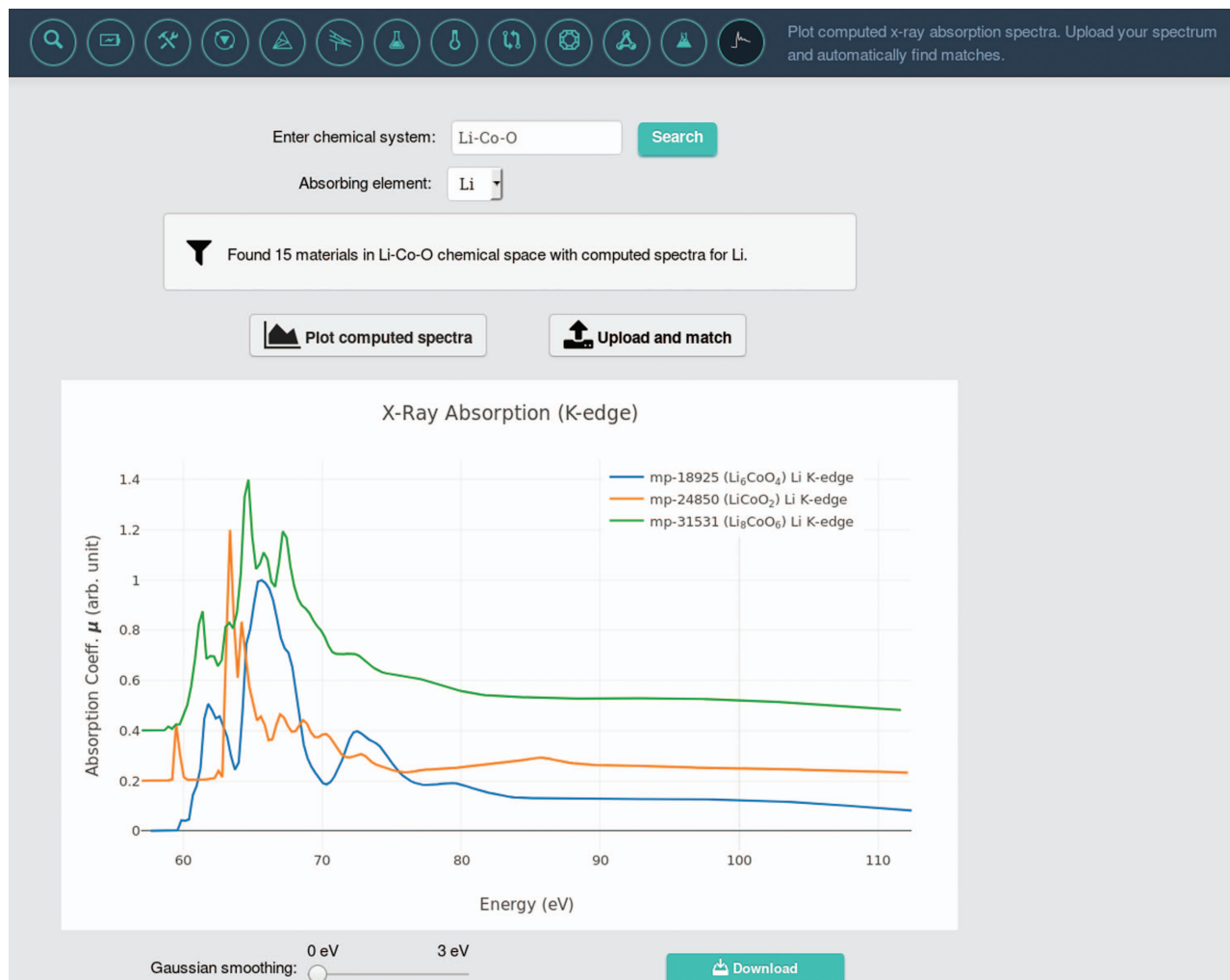
As shown in Fig. 1, the workflow is initiated by importing a structurally optimized compound from MP. Each site in the downloaded structure is a possible absorbing center and FEFF calculation sequence must be initiated for each site. However, the number of calculations can be reduced by considering only the symmetrically unique sites in the structure. The FEFF input files for each such symmetrically unique absorbing site are generated subsequently and the FEFF binary is invoked on each input set. In the final step, the computed spectra from each calculation is inserted into a MongoDB database and disseminated via the Material Project (<https://materialsproject.org/>) website.

### Code availability

Except for the FEFF code, which is proprietary, all the other aforementioned packages used in the high-throughput XAS workflow are open source and can be found at <https://github.com/materialsproject> and <https://github.com/hackingmaterials/atomate>.

### Data Records

All the data described in this section can be found in the file, *xas.json.tgz* (Data Citation 1). The same data is also stored in the Materials Project database and is made freely available to the public. We also provide a user friendly web application called *XAS Matcher* (screenshot shown in Fig. 2) that enables user interaction with the computed data. The app can be reached at <https://materialsproject.org/#apps/xas>. Users can employ the app to search for computed XAS spectra, upload experimental spectra and find structures in the MP database whose computed spectra match that of the uploaded one. Details on spectra matching algorithm employed on Materials Project were published separately<sup>11</sup>.



**Figure 2.** Screen shot of XAS Matcher web application. The web application is hosted at <https://materialsproject.org/#apps/xas>.

### Data Representation

To date, spectra for more than half of the compounds ( $\approx 40000$ ) in the Materials Project database are available, for all the symmetrically unique sites in each structure. Each structure dataset is stored in the database in the binary JavaScript Object Notation (BSON) format. The keys and respective descriptions are summarized in Table 1. Although the workflow yields separate spectra for each unique atomic site, the averaged absorption coefficient over all the sites in the structure with that element is presented on the MP website. This will facilitate comparison with experimental spectra, where the averaging over each element is unavoidable. However, the full data, e.g. spectra for all unique sites, are available to the user for download and further analysis.

### Data Download

The spectral data as well as the input parameters used for the calculations can be downloaded either directly from the Material Project website or using the REST Application Programming Interface (API) available in pymatgen<sup>18</sup>. Data can be downloaded for each element in the selected structure. The downloaded spectrum is provided in a tab separated file format and includes the spectral data for all the symmetrically unique sites of the selected element in the structure. The standard XAS data interchange (XDI) format<sup>19</sup> is also available for download, which can be directly imported into most existing XAFS data analysis programs<sup>20</sup> for further detailed analysis.

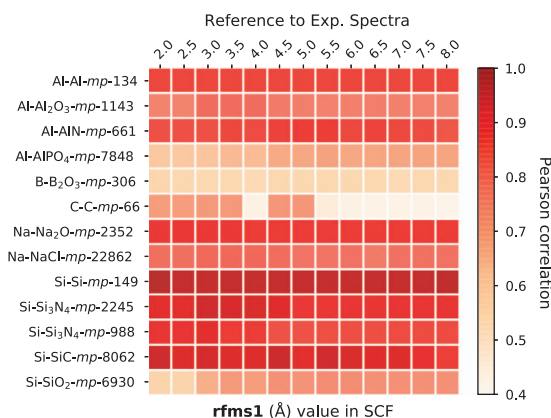
### Technical Validation

#### Verification of the default parameter settings for the workflow.

The workflow described above relies on the default FEFF input parameter settings to generate the K-edge XANES spectra in a high throughput fashion. In this section, we will briefly describe the major FEFF

Key	Data Type	Description
<code>input_parameters</code>	string	the FEFF input settings used in the computation of the spectrum.
<code>xas_id</code>	string	unique id for each spectrum, e.g. 'mp-505011-28-XANES-K'.
<code>spectrum_type</code>	string	type of XAS e.g. 'XANES'.
<code>edge</code>	string	absorption edge e.g. 'K'.
<code>mp_id</code>	string	mp id of the structure.
<code>absorbing_atom</code>	string	site index of the absorbing site in the structure.
<code>structure</code>	string	the structure in dictionary format (can be loaded as a Structure object in pymatgen).
<code>spectrum</code>	float	array of shape (100, 6) where each column means the following(in that order):Energy(eV), Energy with respect to the Fermi level (eV), Wave number, $\mu$ (total absorption coefficient), $\mu_0$ (the background absorption coefficient), $x$ (normalized finestructure)

**Table 1.** Keys and their description for each spectra data JSON file.



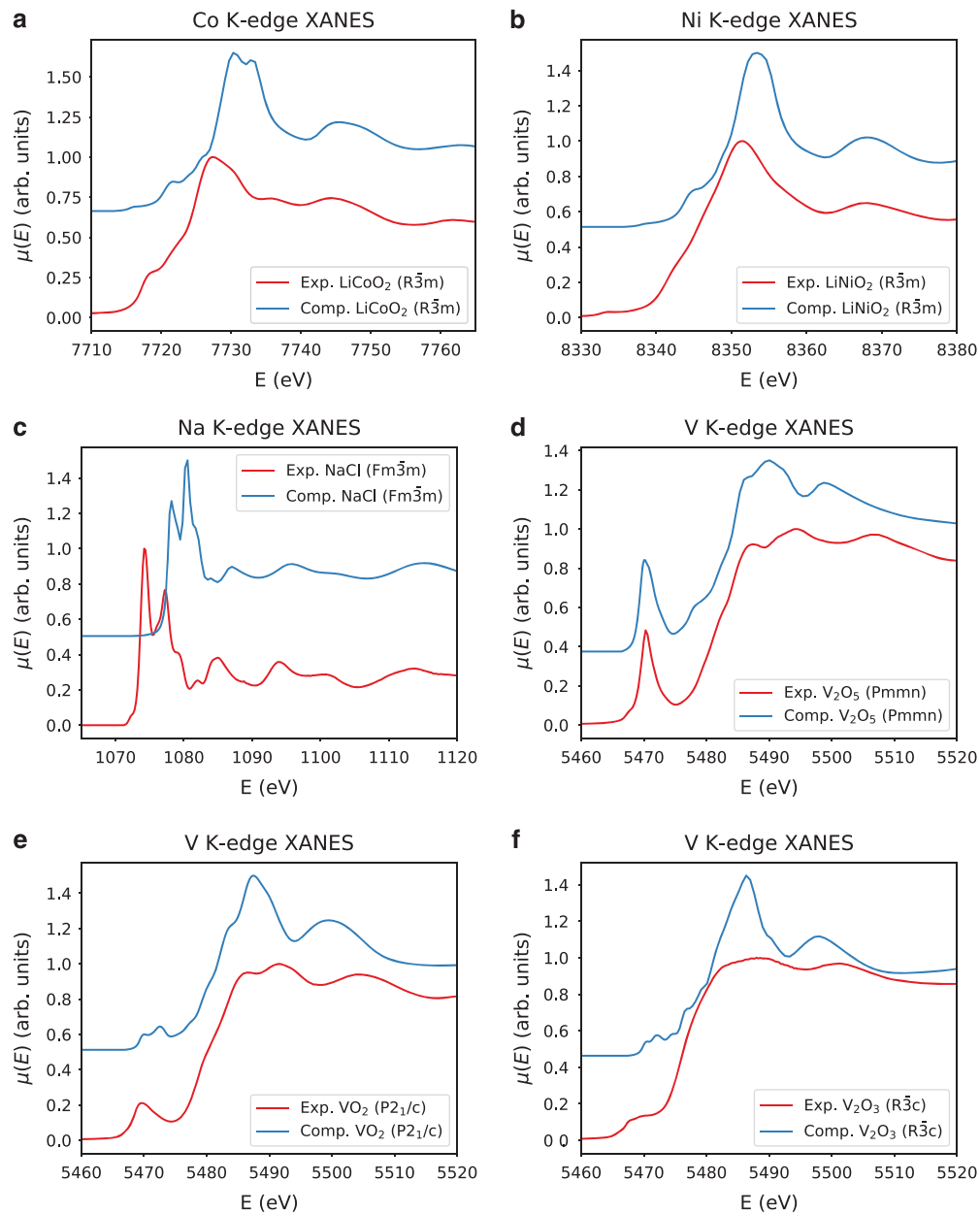
**Figure 3.** Benchmarking results of `rfms1` parameter in the SCF card for K-edge XANES of various materials. Pearson correlation coefficients were calculated between spectra calculated at different `rfms1` and the experimental reference provided by electron energy-loss spectroscopy (EELS) Data Base<sup>8</sup>.

input parameters relevant to the calculation of the XANES spectra, the bench-marking procedure and sample validation cases against experimentally available XANES spectra.

FEFF9 is capable of achieving quantitative agreement with XAS experimental results with a minimal set of adjustable parameters. The development and implementation of parameter-free models within the FEFF9 code permit consistent calculations across different chemical systems and constitute the main advantage for high-throughput calculations. In the benchmarking process, we included 13 unique compounds and their corresponding high-quality K-edge XAS spectra available in the open EELS/XAS database<sup>8</sup>, supplemented by 6 experimental XANES spectra of  $V_2O_5$ ,  $V_2O_3$ ,  $VO_2$ ,  $LiNiO_2$ ,  $LiCoO_2$ , and  $NiO$  from previous studies<sup>21,22</sup>. Compounds included in the earliest commentary<sup>5</sup> of FEFF9 software were also evaluated. The benchmark compound dataset has a high structural diversity and covers a wide chemical space. Detailed benchmark information is provided in a previous publication<sup>11</sup>. For benchmark compounds that contain detailed structural information, we used structures from the Materials Project (<https://materialsproject.org/>) database that exhibit an optimized geometry with the same space group as the benchmark compound. For benchmark compounds without provided structural information, MP ground state structures with identical chemical compositions were used<sup>11</sup>.

The following input fields in FEFF9 were subjected to convergence and optimization tests:

- *Self-consistent field (SCF)*: The SCF card controls FEFF automated self-consistent potential calculations. The self-consistent potential calculation is required in the XANES calculation for the Fermi level  $E_0$  estimation. In the convergence test, we varied the number of atoms included in the self-consistent potential calculations through changing the `rfms1` value from 2 Å to 8 Å at 1 Å interval.
- *Full multiple scattering (FMS)*: The FMS card is required in the XANES calculation as the multiple scattering (MS) expansion's convergence might not be stable in the XANES calculation<sup>5</sup>. To identify the effect of `rfms` field on XANES calculation results, we varied the `rfms` value from 3 Å to 11 Å at 1 Å interval.
- *EXCHANGE*: The EXCHANGE card specifies the exchange correlation potential model used for XANES calculation. The Hedin-Lundqvist self-energy is chosen as previously recommended for most applications<sup>23</sup>.



**Figure 4.** Sample comparisons of FEFF computed K edge XANES spectra with the corresponding experimental ones for six different compounds. Computational spectra are shifted upwards by 0.5. (a)  $\text{LiCoO}_2$ , (b)  $\text{LiNiO}_2$ , (c)  $\text{NaCl}$ , (d)  $\text{V}_2\text{O}_5$ , (e)  $\text{VO}_2$ , (f)  $\text{V}_2\text{O}_3$ .

- COREHOLE:** The COREHOLE card is used for specifying how the core is treated during XANES calculation. The default choice in FEFF treats the core-hole interaction based on the Final State Rule (FSR), which could overestimate the strength of the core-hole and excludes the core-hole mixing effect<sup>24</sup>. To overcome these deficiencies and avoid possible break down of FSR for the L-shell metals<sup>25</sup>, the random phase approximation (RPA) is used to approximate the core-hole interactions in our high-throughput K-edge XANES calculations.

Through the benchmarking study, a set of optimized FEFF parameters were determined to achieve the best balance between the computational cost and convergence performance. The Pearson correlation coefficient is used to compare spectra calculated using different parameters. The Pearson correlation coefficient between two same energy grid spectra,  $X_i$  and  $Y_i$ , is calculated using the following expression:

$$S_{\text{Pearson}}(X, Y) = \frac{\sum_{i=1}^D (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{(\sum_{i=1}^D (X_i - \bar{X})^2)(\sum_{i=1}^D (Y_i - \bar{Y})^2)}}, \quad (3)$$

where  $X_i$  and  $Y_i$  are the corresponding absorption coefficients.



We noticed that the Pearson correlation coefficients between FEFF computed spectra and experimental spectra obtained from EELS Data Base are above 0.85 in general (see Fig. 3). For C and  $B_2O_3$ , the FEFF-computed spectra are not in good agreement with experimental spectra. Possible solutions include the adoption of other higher-level real-space full-potential multiple scattering theory or first principles approaches<sup>26</sup>, which are not ideal for high-throughput implementation due to their high computational cost. Figure 4 depicts some sample comparisons between the computed and the experimental K-edge XANES spectra. We note that the computed spectra match with that of the experimental ones only up to a constant shift in the energy. The computed K-edge XANES spectra of vanadium oxides given in Fig. 4d–f show a strong change in their first peak intensity. Reasonably good agreement between computational and experimental spectra was found.

### Usage Notes

We present a database of K-edge XANES spectra computed using FEFF. The data is made freely available to all researchers via the Materials Project ([www.materialsproject.org](http://www.materialsproject.org)). Users can also download the data using the REST API that is part of pymatgen. All the codes used to create the high throughput are made freely available at Github (<https://github.com/materialsproject> and <https://github.com/hackingmaterials/atomate>). We hope that the users will find the data to be useful and will find novel ways to employ the data to accelerate their research. One such use case would be using machine learning techniques to predict structures from the experimentally measured spectra.

For users of FEFF and the spectra resulting in this study, it should be noted that K-edge XANES spectra computed by FEFF are more accurate for the investigation of elements in the periodic table up to the fifth-row. For excitations in heavier elements, e.g., the rare earth elements and 5d elements, L-edge XANES spectra are primarily used. FEFF is also applicable for the simulation of L-edge XANES spectra, though in certain cases<sup>27–29</sup> ground-state DFT methodologies need to be used for better agreement between computed spectra and experimental results. A detailed study of high-throughput FEFF calculation and implementation of L-edge XANES is currently being conducted by our research group. Furthermore, the analysis of XANES is recommended for the identification of oxidation state and coordination chemistry of the absorbing atom<sup>30</sup>. We note that the quantitative accuracy of XANES calculations is not comparable to EXAFS in identification of the distances, coordination number, and species of the neighbors of the absorbing atom. The accurate and precise interpretation of EXAFS is routinely conducted coupled with well-established software packages<sup>31</sup> using the FEFF calculated EXAFS. The FEFF calculated K-edge EXAFS of all the materials available in the Materials Project database is underway, and a significant portion will be released in parallel with this publication.

### References

- Seidler, G. T. *et al.* A laboratory-based hard x-ray monochromator for high-resolution x-ray emission spectroscopy and x-ray absorption near edge structure measurements. *Rev. Sci. Instrum.* **85**, 113906 (2014).
- Bilderback, D. H., Elleaume, P. & Weckert, E. Review of third and next generation synchrotron light sources. *J. Phys. B At. Mol. Opt. Phys.* **38**, S773–S797 (2005).
- Ankudinov, A. L., Takimoto, Y. & Rehr, J. J. Combined Bethe–Salpeter equations and time-dependent density-functional theory approach for x-ray absorption calculations. *Phys. Rev. B* **71**, 165110 (2005).
- Tanaka, I. & Mizoguchi, T. First-principles calculations of x-ray absorption near edge structure and energy loss near edge structure: present and future. *J. Phys. Condens. Matter* **21**, 104201 (2009).
- Rehr, J. J., Kas, J. J., Vila, F. D., Prange, M. P. & Jorissen, K. Parameter-free calculations of X-ray spectra with FEFF9. *Phys. Chem. Chem. Phys.* **12**, 5503 (2010).
- Laskowski, R. & Blaha, P. Understanding the  $L_{2,3}$  x-ray absorption spectra of early 3d transition elements. *Phys. Rev. B* **82**, 205104 (2010).
- Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater* **1**, 011002 (2013).
- Ewels, P., Sikora, T., Serin, V., Ewels, C. P. & Lajaunie, L. A Complete Overhaul of the Electron Energy-Loss Spectroscopy and X-Ray Absorption Spectroscopy Database: eelsdb.eu. *Microsc. Microanal.* **22**, 717–724 (2016).
- Kiyotaka, A., Isao, I., Haruo, K., Takaaki, K. & Hideki, S. Dopant structure in fecl3-doped polyacetylene studied by x-ray absorption spectroscopy and x-ray photoelectron spectroscopy. *Bull. Chem. Soc. Jpn* **58**, 2113–2120 (1985).
- Newville, M., Carroll, S. A., O'Day, P. A., Waychunas, G. & Ebert, M. A web-based library of XAFS data on model compounds. *J. Synchrotron Radiat.* **6**, 276–277 (1999).
- Zheng, C. *et al.* Automated generation and ensemble-learned matching of X-ray absorption spectra. *npj Comput. Mater* **4**, 12 (2018).
- Mathew, K. *et al.* Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows. *Comput. Mater. Sci.* **139**, 140–152 (2017).
- Takigawa, I., Shimizu, K.-i., Tsuda, K. & Takakusagi, S. Machine-learning prediction of the d-band center for metals and bimetallics. *RSC Adv* **6**, 52587–52595 (2016).
- 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.* **8**, 5091–5098 (2017).
- Rehr, J. J. & Albers, R. C. Theoretical approaches to x-ray absorption fine structure. *Rev. Mod. Phys.* **72**, 621–654 (2000).
- Ong, S. P. *et al.* Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).
- Jain, A. *et al.* FireWorks: a dynamic workflow system designed for high-throughput applications. *Concurr. Comput. Pract. Exp* **27**, 5037–5059 (2015).
- Ong, S. P. *et al.* The Materials Application Programming Interface (API): A simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles. *Comput. Mater. Sci.* **97**, 209–215 (2015).



19. Ravel, B., Hester, J. R., Solé, V. A. & Newville, M. Towards data format standardization for X-ray absorption spectroscopy. *J. Synchrotron Radiat.* **19**, 869–874 (2012).
20. Ravel, B. & Newville, M. XAFS Data Interchange: A single spectrum XAFS data file format. *J. Phys. Conf. Ser.* **712**, 1–5 (2016).
21. Rana, J. *et al.* Local structural changes in LiMn<sub>1.5</sub>Ni<sub>0.5</sub>O<sub>4</sub> spinel cathode material for lithium-ion batteries. *J. Power Sources* **255**, 439–449 (2014).
22. Rana, J. *et al.* On the structural integrity and electrochemical activity of a 0.5Li<sub>2</sub>MnO<sub>3</sub>·0.5LiCoO<sub>2</sub> cathode material for lithium-ion batteries. *J. Mater. Chem. A* **2**, 9099 (2014).
23. Moreno, M. S., Jorissen, K. & Rehr, J. J. Practical aspects of electron energy-loss spectroscopy (EELS) calculations using FEFF8. *Micron* **38**, 1–11 (2007).
24. Rehr, J. J., Soininen, J. A. & Shirley, E. L. Final-state rule vs the bethe-salpeter equation for deep-core x-ray absorption spectra. *Phys. Scr.* **2005**, 207 (2005).
25. Vinson, J. & Rehr, J. J. Ab initio Bethe-Salpeter calculations of the x-ray absorption spectra of transition metals at the L-shell edges. *Phys. Rev. B* **86**, 1–6 (2012).
26. Xu, J. *et al.* X-ray absorption spectra of graphene and graphene oxide by full-potential multiple scattering calculations with self-consistent charge density. *Phys. Rev. B* **92**, 125408 (2015).
27. Alperovich, I. *et al.* Understanding the electronic structure of 4d metal complexes: From molecular spinors to L-edge spectra of a di-Ru catalyst. *J. Am. Chem. Soc.* **133**, 15786–15794 (2011).
28. Newville, M. EXAFS analysis using FEFF and FEFFIT. *J. Synchrotron Radiat.* **8**, 96–100 (2001).
29. Shishido, T., Asakura, H., Yamazoe, S., Teramura, K. & Tanaka, T. Structural analysis of group V, VI, VII metal compounds by XAFS and DFT calculation. *J. Phys. Conf. Ser.* **190**, 012073 (2009).
30. Newville, M. Fundamentals of XAFS. *Rev. Mineral. Geochemistry* **78**, 33–74 (2014).
31. Ravel, B. & Newville, M. ATHENA, ARTEMIS, HEPHAESTUS : data analysis for X-ray absorption spectroscopy using IFEFFIT. *J. Synchrotron Radiat.* **12**, 537–541 (2005).

## Data Citation

1. Mathew, K. *et al.* Figshare <https://doi.org/10.6084/m9.figshare.c.3946561> (2018).

## 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, which support K.M., C. Z., C. C., D. W., S. P. O and K.A.P. The authors thank J. Rehr and A. Dozier for their contributions towards the FEFF workflow. Computational resources were provided by the NSF DIBBs funding as well as the Triton Shared Computing Cluster (TSCC) at the University of California, San Diego. The online data, and search capabilities were funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division under Contract No. DE-AC02-05-CH11231: Materials Project program KC23MP.

## Author Contributions

S.P.O. and K.A.P. conceived the study. K.M., C.Z. extracted and processed input data. K.M., C.Z., D.W. and C.C. assembled data products and generated figures and tables. A.D., J.J.R. helped the simulations of XAS spectra. K.M., C.Z. and C.C. wrote the manuscript. S.P.O., K.A.P. reviewed the manuscript and contributed to revisions.

## Additional Information

**Competing interests:** The authors declare no competing interests.

**How to cite this article:** Mathew, K. *et al.* High-throughput computational X-ray absorption spectroscopy. *Sci. Data* 5:180151 doi: 10.1038/sdata.2018.151 (2018).

**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 made available in this article.

© The Author(s) 2018