Pacific Northwest NATIONAL LABORATORY

PNNL-35397

Automated Energy-Dispersive X-ray Spectroscopy Analysis for Multi-Modal Few-Shot Learning

December 2023

1 Michael J. Holden
2 Christina M. Doty
3 Arman H. Ter-Petrosyan
4 Jenna Pope
5 Sarah M. Akers
6 Steven R. Spurgeon

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from
the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062
www.osti.gov
ph: (865) 576-8401
fox: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: http://www.ntis.gov

Automated Energy-Dispersive X-ray Spectroscopy Analysis for Multi-Modal Few-Shot Learning

December 2023

1 Michael J. Holden
2 Christina M. Doty
3 Arman H. Ter-Petrosyan
4 Jenna Pope
5 Sarah M. Akers
6 Steven R. Spurgeon

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Abstract

Scanning transmission electron microscopy (STEM) is a powerful tool that allows for the atomic-scale analysis of a materials' structure, chemistry, and defect domains (Akers et al. 2021). The current generation of microscopes generate vast amounts of data, surpassing the limits of effective manual analysis traditionally performed by domain experts (Spurgeon et al. 2021). While recent strides in machine learning have significantly enhanced the processing of large and intricate datasets acquired through electron microscopy, the prevalent use of proprietary software packages for initial data collection poses a challenge. In many cases, these software packages act as a 'black box', constraining user functionality and hindering the output of data in a format that is conducive to seamless integration into machine learning models.

This work addresses these challenges by adapting HyperSpy, an open-source Python library, for the analysis and quantification of raw energy dispersive spectroscopy (EDS) data acquired through STEM. The modified HyperSpy code successfully facilitates user-defined segmentation of the data, enabling the integration of atomic %, weight %, and raw EDS spectra for each segmented region into an existing few-shot machine learning model. While initial results reveal discrepancies in quantified atomic and weight percentages when compared to proprietary software, ongoing efforts aim to rectify this issue by refining the fit of the HyperSpy model to the EDS spectra.

Overall, this research underscores the potential of open-source tools like HyperSpy to enhance the accessibility of analytical tools, fostering a transparent and user-friendly environment for seamlessly incorporating electron microscopy data into machine learning models.

Acknowledgments

M.J.H. was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI). This research was supported by a Chemical Dynamics Initiative (CDi) Laboratory Directed Research and Development (LDRD) Initiative / Investment project at Pacific Northwest National Laboratory (PNNL). PNNL is a multiprogram national laboratory operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL0-1830.

I would like to express my gratitude to Steven Spurgeon, Jenna Pope, Arman H. Ter-Petrosyan, Christina Doty, and Sarah Akers at PNNL. Steven mentored me throughout this project and ensured I had the opportunity to develop my professional skills and network by arranging lab tours and introducing me to various staff at PNNL. Jenna introduced me to the spec2vec code and worked closely with me on the integration of the EDS data into the few-shot model. Arman assisted me with the process of integrating the EDS data into the few-shot model and helped develop ideas for graphs. Sarah met with me on several occasions to review my initial progress in HyperSpy and provided feedback. Christina walked me through the team's codebase and showed me how to utilize GitLab for this project.

Contents

1.0 Introduction

Scanning transmission electron microscopy (STEM) is a powerful method of characterization that allows for the collection of images and spectroscopic data at the atomic scale. Analysis of this data can provide actionable insights into a materials' structure, chemistry, and defect domains (Akers et al. 2021; Doty et al. 2022). The current generation of electron microscopes generate vast amounts of data (up to ~ 200Tb/hr), surpassing the limits of effective manual analysis traditionally performed by domain experts (Spurgeon et al. 2021; Voyles 2017).

In response to the need for high-throughput data science tools, recent advancements in machine learning models have notably improved the handling of extensive and complex datasets obtained through electron microscopy. One approach to segmenting and classifying images collected via STEM is few-shot machine learning, which demonstrates the ability to effectively categorize images based on a small set of user-provided examples (Akers et al. 2021; Doty et al. 2022). Incorporating the corresponding spectroscopic data alongside the STEM image not only holds the potential to improve the model but also offers additional insights into the rationale behind the model's categorized output. However, a significant challenge arises from the use of proprietary software packages during the initial data collection phase. These software tools often operate as a 'black box', constraining user functionality and hindering the output of data in a format that is conducive to seamless integration into machine learning models (Voyles 2017).

This work addresses these challenges by adapting HyperSpy, an open-source Python library, for the analysis and quantification of raw energy dispersive spectroscopy (EDS) data acquired through STEM (Francisco de la Peña et al. n.d.). The adapted HyperSpy code segments and quantifies EDS data, collected from the cross-section of lanthanum ferrite (LFO) thin films deposited onto a strontium titanate (STO) substrate via pulse laser deposition (PLD). The extracted atomic %, weight %, and raw EDS spectra for each chipped region is exported as a data frame for future integration into an existing few-shot machine learning model (Figure 1).

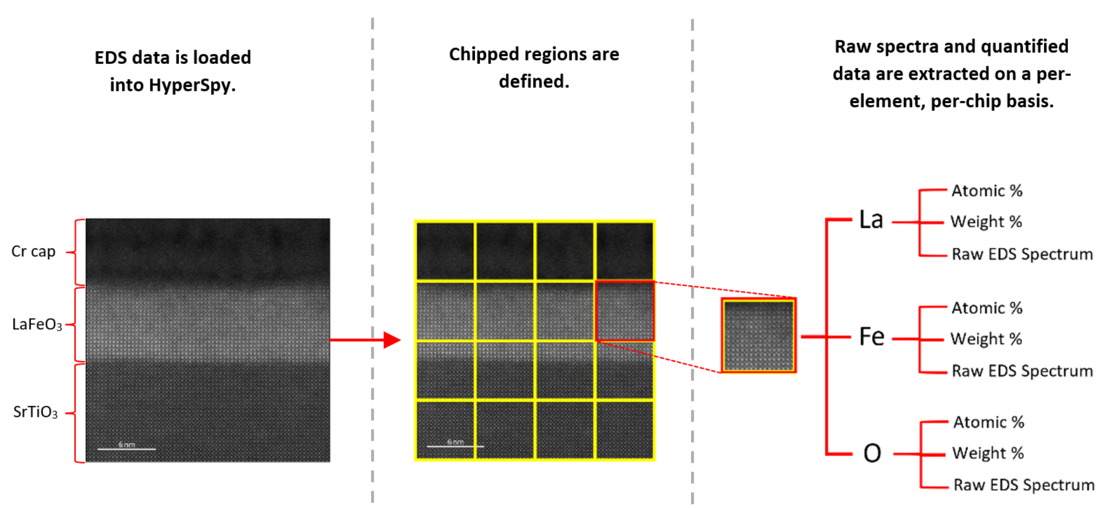

Figure 1. Process of chipping and extracting quantified and raw EDS data for integration with an existing few-shot model

2.0 Methods

EDS analysis was conducted using both HyperSpy and Thermo Fisher Velox, a commonly used commercial EDS analysis package. A comparison was made between the software packages based on two criteria: The ability to facilitate and export data chipping, and the accuracy in quantifying EDS data for each corresponding chip. The resulting atomic percentage values from each process were compared to each other, in addition to the expected stoichiometric ratios of strontium titanate (substrate) and lanthanum ferrite (deposited film).

Due to limitations of manually chipping data using Velox, the dataset used for the comparative study was divided into a simple 8x8 grid, resulting in a total of 64 chips (sample 1 in Table 1). However, when utilizing HyperSpy to chip data for integration with the few-shot model, the existing chipping parameters from the few-shot model were applied to each respective dataset (samples 2-7 in Table 1). Note that sample 7 is the same data set as sample 1 but has undergone a crop on the navigational axis to 500px x 500px and the chip size was reduced to 15px x 15px.

Table 1. EDS samples and their chipped regions. Sample 1 was used for the comparative study between HyperSpy and Velox. Samples 2-7 were chipped in preparation for integration with their respective few-shot models.

Sample #	Sample Name	Navigational Axis Length (px)	Chip Size (px)	Number of Columns x Rows	Total Number of Chips
1	HS vs Velox-0dpa-LFOpristine	512 x 512	64 x64	8 x 8	64
2	0dpa-LFOdefectdomains	512 x 512	10 x 10	51 x 51	2601
3	0dpa-LFOdefectdomains	512 x 512	15 x 15	34 x 34	1156
4	0.1dpa-LFOpristine	500 x 500	20 x 20	25 x 25	625
5	0.1dpa-LFOdefectdomains	512 x 512	10 x 10	51 x 51	2601
6	0.1dpa-LFOdefectdomains	512 x 512	12 x 12	42 x 42	1764
7	0dpa-LFOpristine	500 x 500	15 x15	33 x 33	1089

2.1 Data Collection

STEM high-angle annular dark-field (STEM-HAADF) images and EDS maps were collected on a Thermo Fisher Themis Z microscope operating at 300 keV. Data was collected in the native Velox *.emd format and subsequently imported into HyperSpy (Francisco de la Peña et al. n.d.).

Software applications were run on a HP EliteBook equipped with a i7-8665U and 16 GB of 2400 MHz RAM.

2.2 Chipping EDS Data

Prior to quantification in HyperSpy, the 1D EDS spectra and 2D navigational axis were cropped by implementing the ".isig()" and ".inav()" functions, respectively. Data was then chipped using the 'scale' argument of the ".rebin()" function. This allowed for the size of the chips to be defined in pixels and then the navigational axis dimensions were divided by the defined chip size. Any remainder along each navigational axis was automatically cropped from the data set.

In Velox, chipped regions were manually defined using the spectrum integration tool. The position and size of each chipped region were specified in the object properties window. Velox limits the user to a maximum of ten manually defined integration regions. Therefore, the chipping process consisted of defining one row of eight chips, exporting the data, and then moving onto the next row until the entire data set had been chipped.

2.3 Quantification of EDS Data

Once the chipped regions were defined in HyperSpy, a default model was created. The model fits a 6th order background polynomial and a Gaussian for each X-ray line associated with the elements defined in the metadata. The ".multifit()" function was used to iterate over each chipped region to fit the final model to the EDS data. HyperSpy automatically calculates the reduced Chi-squared value to measure the goodness of fit for each chip. After background removal, integration of each x-ray peak yielded the intensities required for quantification. For both Hyperspy and Velox, the Cliff-Lorimer method for qualitative measurement of the atomic percents was implemented. K-factors were extracted from Velox for use in HyperSpy. Finally, the determined atomic percent was converted to weight percent and the results were exported as a .csv data frame.

In Velox, the expected elements were defined, and the default multi-polynomial model with parabolic background order was used. No additional absorption correction or optimized spectrum fit was utilized. Due to the chipping limitations mentioned above, the quantified EDS data was exported in batches of eight chips and manually compiled into an Excel data frame ensuring the chip IDs and locations were conserved.

3.0 Results and Discussion

3.1 Chipping

The chipping process in Velox is cumbersome and prone to errors due to the manual nature and software limitations of defining each chipped region. In its current state, the chipping process in Velox is not a user-friendly experience when defining a simple 8 x 8 grid and is impractical in all respects when attempting to chip a larger grid for integration with the few-shot model.

By contrast, HyperSpy facilitates numerous methods for chipping and export formats. The chipping of data sets can be implemented with a few simple lines of code and the chipped regions can be verified by plotting the data or referencing the axis dimensions via the “. axis_manager” function. Additionally, each pixel in the original data set is indexed in HyperSpy which allows for a per-pixel defined chipping process. This would allow the user to define irregular and asymmetrical shaped chips and greatly increases the functionality of the chipping process when compared to Velox.

3.2 EDS Quantification

The top two rows of figures 2 and 3 illustrate the atomic percent maps generated by HyperSpy and Velox for LFO and STO, respectively. These maps clearly reveal the presence of LFO and STO in the expected regions of the sample. These findings confirm that both HyperSpy can accurately identify the presence and location of the elements of interest in the sample. However, there are variations in the assigned atomic percent values for each element within the chipped regions between the two software applications. The bottom row of images in figures 2 and 3 represents the difference between the quantified atomic percent values from HyperSpy and Velox. Red regions indicate that HyperSpy yields larger values than Velox, blue regions indicate smaller values, and white regions indicate agreement between HyperSpy and Velox.

In contrast to Velox's quantified atomic percentages for STO, HyperSpy exhibits slightly lower values for strontium and higher values for titanium. Moreover, HyperSpy yields considerably higher atomic percentages for oxygen in LFO and STO, with a notable discrepancy, in the oxygen atom percentage within the LFO region. This discrepancy persists in the LFO region, where HyperSpy reports higher atomic percentages for iron but significantly underestimates the atomic percentage of lanthanum.

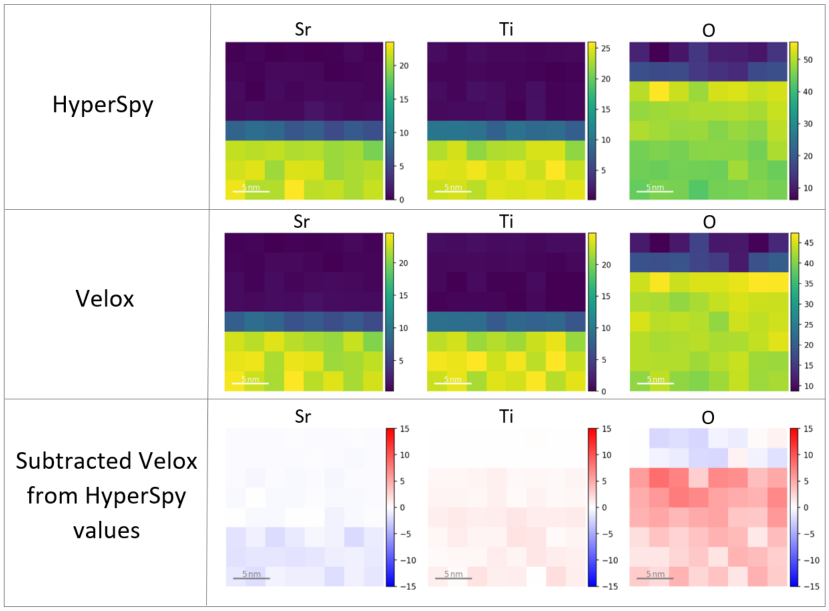

Figure 2. Atomic percent maps of STO quantified in HyperSpy (top), Velox (middle), and the difference in quantified atomic percent values (bottom). Sr_Kα, Ti_ Kα, and O_ Kα ionization peaks were used for quantification.

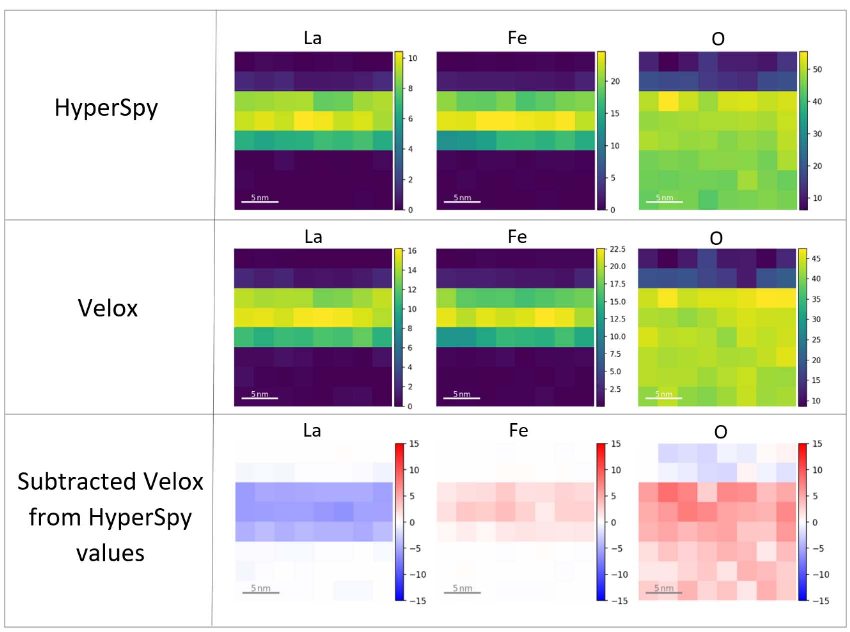

Figure 3. Atomic percent maps of LFO quantified in HyperSpy (top), Velox (middle), and the difference in quantified atomic percent values (bottom). La_Lα, Fe_ Kα, and O_ Kα ionization peaks were used for quantification.

While comparing HyperSpy to Velox serves the purpose of assessing HyperSpy's capability to compete with industry-standard software, it falls short of providing an accurate comparison of experimental data against a known ground truth. To address this, the ratio of normalized atomic percentages for each element in each region was compared to the known stoichiometric ratio of their respective materials (20:20:60 for both STO and LFO). Tables 2 and 3 shows large percent errors between the experimentally determined and known ratios of STO for both Hyperspy and Velox, highlighting the qualitative nature of EDS data.

Table 2 reveals that HyperSpy delivers a more accurate representation of atomic percentages in the STO layer. Unfortunately, this trend does not continue in the LFO layer, where HyperSpy significantly underestimates the atomic percent of lanthanum, as demonstrated in Table 3. This issue is consistent across all the samples quantified by HyperSpy, prompting the need for a thorough analysis of how HyperSpy fits the model and quantifies the La_Lα x-ray peak, with the goal of refining and improving its performance for future EDS analysis.

Table 2. Percent error between the quantified atomic percent ratios and expected stoichiometric ratio of STO. Atomic percents were summed and normalized across the chips where STO was detected in the sample.

STO	HyperSpy Sample 1			Velox Sample 1			HyperSpy Sample 7		
Element	Sr	Ti	O	Sr	Ti	O	Sr	Ti	O
Summed Atomic %	21.15	24.12	45.60	22.47	23.22	42.63	20.68	23.49	43.26
Normalized Atomic %	23.27	26.55	50.18	25.44	26.29	48.27	23.65	26.87	49.48
% Error	16.35	32.73	16.36	27.18	31.46	19.55	18.26	34.33	17.53

Table 3. Percent error between the quantified atomic percent ratios and expected stoichiometric ratio of LFO. Atomic percents were summed and normalized across the chips where LFO was detected in the sample.

LFO	HyperSpy Sample 1			Velox Sample 1			HyperSpy Sample 7		
Element	La	Fe	O	La	Fe	O	La	Fe	O
Summed Atomic %	8.40	19.33	49.72	13.51	17.35	44.27	8.37	19.00	46.08
Normalized Atomic %	10.85	24.96	64.19	17.98	23.09	58.92	11.40	25.86	62.74
% Error	45.77	24.81	6.99	10.09	15.47	1.79	43.01	29.32	4.56

Regarding samples 2-7, HyperSpy managed to successfully fit the model and quantify the EDS data despite the low signal to noise ratio attributed to the significantly smaller chip size. Reducing chip size also increased the time to compile the code from ~6 minutes for sample 1 (64 total chips), to ~2 hours for sample 7 (1089 total chips). The resulting maps (Figure 4) shows that the HyperSpy's quantification process successfully identified LFO and STO within the expected regions of the sample. The smaller chip sizes continued to exhibit a similar percent error to the larger chipped data when comparing the normalized atomic % of LFO and STO to the expected stoichiometric ratio (Tables 2 and 3). Since the % error did not significantly fluctuate between the 8x8 and the 33x33 chipped grid, it can be determined that HyperSpy is relatively consistent in its quantification, even when varying the chip size.

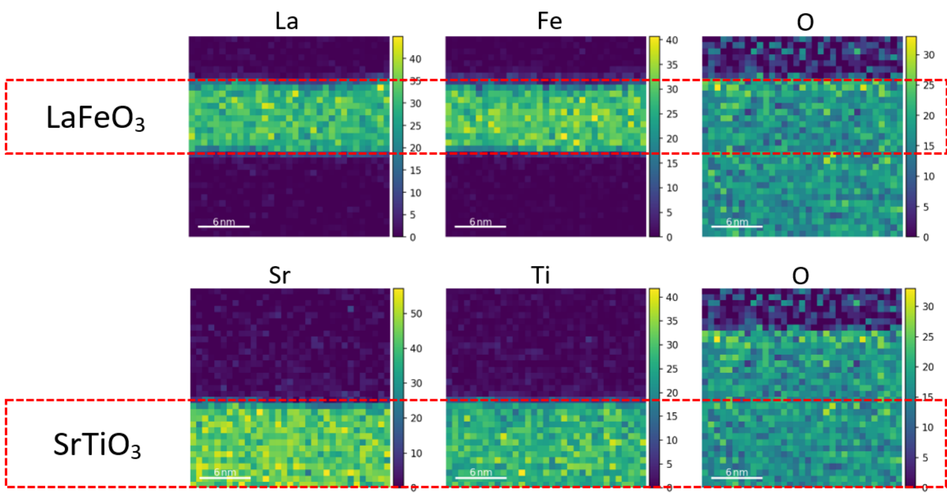

Figure 4. Atomic percent maps of LFO and STO in sample 7, quantified in HyperSpy

4.0 Conclusions

Overall, the comparison between Velox and HyperSpy highlights significant differences in their chipping processes. Velox's manual and limited chipping capabilities make it less practical for segmenting large arrays of chips and integrating quantified data with few-shot models. By contrast, HyperSpy offers a more efficient and flexible chipping process, enabling users to define irregular and asymmetrical shaped chips. Additionally, HyperSpy also facilitates the ability to easily export the resulting data in several formats for use with few-shot models.

Regarding EDS quantification, both HyperSpy and Velox can accurately identify the presence and location of elements in the sample. However, variations in assigned atomic percent values exist between the two software applications, with differences particularly notable in oxygen and Lanthanum percentages for LFO. While HyperSpy generally exhibits a more accurate representation of stoichiometric ratios in the STO layer, discrepancies in specific elements, such as lanthanum in LFO, highlight the need for further refinement of the HyperSpy model.

Finally, considerations must be made by the user when determining the ideal chip size for their application. Since HyperSpy iterates over each chip during the model building and fitting process, smaller chips will incur a penalty of increased computational time. Furthermore, a decrease in chip size will increase resolution, however, smaller chips have a lower signal to noise ratio than larger chips, which makes it difficult to fit a model to the data set. This work has shown that HyperSpy can produce consistent results even with a reduced to chip size, but its limitations have not yet been fully investigated. It is important to note that the chip framework is a stepping stone to true semantic (pixel-level) segmentation; given the sparse nature of X-ray data, it will nevertheless be necessary to apply some form of averaging / kernel to the data prior to fitting.

In conclusion, our script based on HyperSpy has proven to be robust and amenable to the chipping process, offering relatively consistent quantification of EDS data. This research shows the potential of open-source tools like HyperSpy to enhance the accessibility of analytical tools, fostering a transparent and user-friendly environment for seamlessly incorporating electron microscopy data into machine learning models.

References

Akers, Sarah, Elizabeth Kautz, Andrea Trevino-Gavito, Matthew Olszta, Bethany E. Matthews, Le Wang, Yingge Du, and Steven R. Spurgeon. 2021. “Rapid and Flexible Segmentation of Electron Microscopy Data Using Few-Shot Machine Learning.” *Npj Computational Materials* 7 (1). https://doi.org/10.1038/s41524-021-00652-z.

Doty, Christina, Shaun Gallagher, Wenqi Cui, Wenya Chen, Shweta Bhushan, Marjolein Oostrom, Sarah Akers, and Steven R. Spurgeon. 2022. “Design of a Graphical User Interface for Few-Shot Machine Learning Classification of Electron Microscopy Data.” *Computational Materials Science* 203 (February). https://doi.org/10.1016/j.commatsci.2021.111121.

Francisco de la Peña, Eric Prestat, Vidar Tonaas Fauske, Pierre Burdet, Jonas Lähnemann, Petras Jokubauskas, Tom Furnival, et al. n.d. “Hyperspy: V1.7.6.” Accessed December 2, 2023. https://doi.org/10.5281/ZENODO.10151686.

Spurgeon, Steven R., Colin Ophus, Lewys Jones, Amanda Petford-Long, Sergei V. Kalinin, Matthew J. Olszta, Rafal E. Dunin-Borkowski, et al. 2021. “Towards Data-Driven next-Generation Transmission Electron Microscopy.” *Nature Materials*. Nature Research. https://doi.org/10.1038/s41563-020-00833-z.

Voyles, Paul M. 2017. “Informatics and Data Science in Materials Microscopy.” *Current Opinion in Solid State and Materials Science*. Elsevier Ltd. https://doi.org/10.1016/j.cossms.2016.10.001.

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354

1-888-375-PNNL (7665)

www.pnnl.gov