High-Throughput Intelligent Analysis of High and Low-Loss EELS

Chaitanya Gadre[1], Xingxu Yan[2], Christopher Addiego[3] and Xiaoqing Pan[1]

[1]Department of Physics and Astronomy, University of California, Irvine, CA 92697, Irvine, California, United States, [2]Department of Materials Science and Engineering, University of California - Irvine, Irvine, California, United States, [3]University of California - Irvine, IRVINE, California, United States

Electron Energy Loss Spectroscopy (EELS) is a supremely versatile technique for elemental mapping, electronic structure analysis, and more recently, vibrational spectroscopy. The invention of direct electron detectors (DEDs) coupled with atomic resolution capabilities of the TEM, put it far ahead of competing spectroscopies. DEDs enable low-dose and swift acquisition of spectroscopic data from beam sensitive samples while maintaining respectable signal-to-noise ratio (S/N)1. While there has been much progress in atomic resolution EELS acquisition and data analysis, there is still a need to handle the processing of large amounts of data generated by these DEDs. Additionally, to be able to analyze the fine structure and peak shifts, improving the quality of the data post-acquisition is a must. Using python, one of the most popular programming languages specifically geared towards scientific computing, Principle Component Analysis (PCA) has been applied to intelligentially increase the quality by learning from the dataset and filtering out uncorrelated noise2 but does not natively support the capability to process large volumes of independent data in parallel. Using a python library called DASK3, we achieve full parallelization of EELS data processing, reducing computation times by up to 20x. Our work also demonstrates measurement of oxidations states of perovskite materials and compositional phonon redshifts enabled by the intelligent smoothing of data and massively parallel computation scalable to super clusters.

Initially, the dataset is treated with PCA with the number of components varied as needed to adequately explain the data. This is done so that background subtraction and subsequent analysis is not hindered by uncorrelated noise which may cause instabilities in the fitting. By default, python processes data sequentially unless specific instructions are scheduled for the threads. Fig 1A illustrates the processing workflow for analyzing the fine structure of core loss data. The PCA filtered high loss spectrum is split so that a power law background model can be fit to the pre-edge of each element. Then, the L-shells of transition elements are treated with an arctan background model to eliminate the Hartree–Slater background. In the last step, the background subtracted peaks are then decomposed to obtain their individual peak properties. For vibrational spectra, best results were obtained when PCA was applied after background subtraction. This is because despite standardizing the spectra, the large tail of the zero-loss peak dominates the feature space. Fig 1B shows the task stream, typically 1 process per thread. Taking advantage of the independent nature of each pixel in an EELS dataset, processing of data points can be handled in parallel. Figs. 1C and 1D shows how memory is allocated with respect to each thread and how tasks are scheduled per thread, respectively.

By taking advantage of PCA and DASK, acquisition and processing of large pixel map with short acquisition times becomes possible. Fig 2A illustrates a dramatic increase in S/N enabling the decomposition of the Ti fine structure. It should be stressed that without PCA treatment, peak

decomposition would not be possible as the fitting would either fail or produce large errors. PCA filtered maps in fig. 2D offers an obvious improvement over the untreated data in fig 2C. Mapping the peak properties of the Ti L2 t2g peak, we have a measure of Ti valence state. Just above the interface the onset dips suggesting a lower oxidation state and consequently, the existence of a two-dimensional electron gas (2DEG). Previous studies of free charge carriers using EELS have been limited to single point scans or line scans, for the first time, 2D mapping of oxidation state and charge carriers is possible. Fig. 2F also shows a similar improvement for vibrational spectra as fig. 2A. A main utility of PCA is the ability to separate high variance features. Fig. 2G and 2H demonstrates the uniqueness of the SiGe and Ge vibrations respectively. Additionally, compositional vibrational redshift is possible by increasing the quality of the data via PCA.

Our work demonstrates that by combining the high utility of PCA for denoising data intelligently and DASK for its parallelization capabilities, we enable the rapid processing of EELS features previously beyond the reach of dedicated study. It also becomes possible to spatially study fine structure and correlated changes with atomic precision.
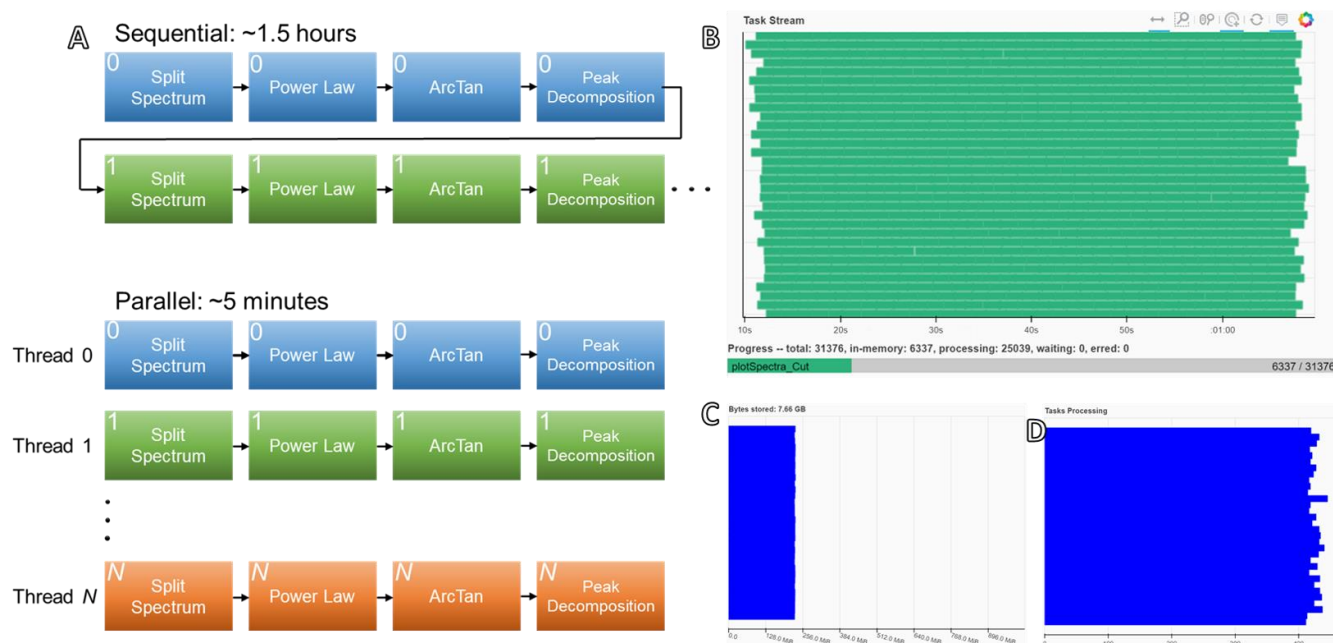

Figure 1. Figure 1: Task scheduling of data processing. A. Comparison of sequential and parallel processing schemes B. Task stream of local cluster displayed by DASK C. Memory allocation per thread D. Task scheduling per thread
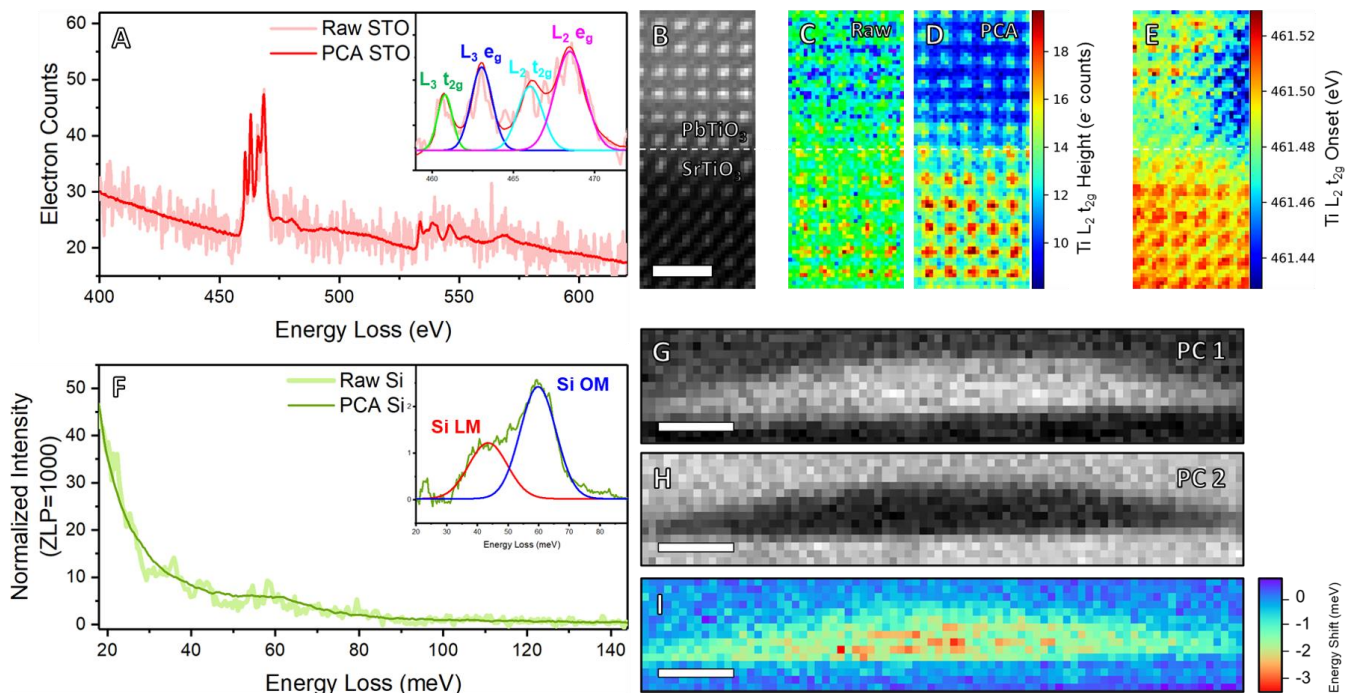
Figure 2. Figure 2: Post data processing high and low loss EELS mapping. A. PCA filtered single-pixel spectrum with inset of peak decomposition of Ti L3 and L2 edges. B. HAADF image of EELS map. C. Ti L2 t2g intensity mapping of untreated data D. Ti L2 t2g intensity mapping of PCA filtered data E. Ti L2 t2g onset mapping obtained from peak decomposition F. PCA filtered single-pixel spectrum with inset of peak decomposition of Si vibrational modes. G-D. Map of loadings of first two principle components. I. Map of Si optical mode (OM) peak shift.

References

[1] Haberfehlner, Georg, et al. "Benefits of direct electron detection and PCA for EELS investigation of organic photovoltaics materials." *Micron* 140 (2021): 102981.

[2] Cueva, Paul, et al. "Data processing for atomic resolution electron energy loss spectroscopy." *Microscopy and Microanalysis* 18.4 (2012): 667-675.

[3] Rocklin, Matthew. "Dask: Parallel computation with blocked algorithms and task scheduling." *Proceedings of the 14th python in science conference*. Vol. 126. Austin, TX: SciPy, 2015.

[4] This work was supported by the Department of Energy (DOE) under Grant DE-SC0014430. The authors acknowledge the use of facilities and instrumentation at the UC Irvine Materials Research Institute (IMRI) supported in part by the NSF MRSEC program (DMR-2011967).