

## Electron Microscopy (Big and Small) Data Analysis With the Open Source Software Package HyperSpy

Francisco de la Peña<sup>1,11,13</sup>, Tomas Ostasevicius<sup>1</sup>, Vidar Tonaas Fauske<sup>2</sup>, Pierre Burdet<sup>1</sup>, Petras Jokubauskas<sup>3</sup>, Magnus Nord<sup>2,4</sup>, Mike Sarahan<sup>5</sup>, Eric Prestat<sup>6</sup>, Duncan N. Johnstone<sup>1</sup>, Joshua Taillon<sup>7</sup>, Jan Caron<sup>8</sup>, Tom Furnival<sup>1</sup>, Katherine E. MacArthur<sup>8</sup>, Alberto Eljarrat<sup>9</sup>, Stefano Mazzucco<sup>7</sup>, Vadim Migunov<sup>8</sup>, Thomas Aarholt<sup>10</sup>, Michael Walls<sup>11</sup>, Florian Winkler<sup>8</sup>, Gaël Donval<sup>12,13</sup>, Ben Martineau<sup>1</sup>, Andreas Garmannslund<sup>2</sup>, Luiz-Fernando Zagonel<sup>14</sup> and Ilya Iyengar<sup>1</sup>

1. Department of Materials Science and Metallurgy, University of Cambridge, Cambridge, U.K.
2. Department of Physics, NTNU, Trondheim, Norway
3. Institute of Geochemistry, Mineralogy and Petrology, University of Warsaw, Poland
4. School of Physics and Astronomy, University of Glasgow, Glasgow, U.K.
5. SuperSTEM, STFC Daresbury Laboratories, Warrington, U.K.
6. School of Materials, The University of Manchester, U.K.
7. Material Measurement Laboratory, NIST, Gaithersburg, MD, U.S.A.
8. Forschungszentrum Jülich GmbH, Jülich, Germany
9. Laboratory of Electron NanoScopies, Universitat de Barcelona, Barcelone, Spain
9. Department of Materials, University of Oxford, Oxford, United Kingdom
11. Laboratoire de Physique des Solides, Paris-Sud University, France
12. Institut des Matériaux Jean Rouxel (IMN), Université de Nantes, Nantes, France
13. CEA Grenoble, Grenoble, France
14. Brazilian Nanotechnology National Laboratory, CNPEM, Campinas, Brazil
15. Simula Research Laboratory, Lysaker, Norway

Advances in scientific instrumentation, driven by the evolving needs of science and technology, are dramatically increasing the amount of experimental data generated. Microscopy is following this trend across all domains, posing common challenges for data analysis. Developing new methods of data analysis to harness the wealth of physical insight contained in this data is now the key to maximising the advancement of scientific understanding using these tools. Here, the focus is on scanning transmission electron microscopy, where modern microscopes enable multiple signals with dimensions of energy, reciprocal space and time to be acquired across real space length scales from ~0.1 nm to 10  $\mu\text{m}$ .

Electron microscopy data analysis is mostly performed using dedicated proprietary software at present. These software packages typically offer a familiar and easy to use Graphical User Interface (GUI) for performing common data analysis tasks. However, they are often limited in comparison with the potential to develop innovative data analysis methods based on the platform of modern scientific programming languages. The success of ImageJ [1]—a program for scientific image processing— has shown that open source development of microscopy software can lead to high quality data analysis tools often where no proprietary alternatives exist.

HyperSpy [2] is an open source Python software package for multi-dimensional data analysis that includes a wide range of features for electron microscopy. Python is increasingly recognised as the *lingua franca* of scientific computing. It offers numerous technical advantages as a programming language but the primary advantage for microscopy data analysis is the availability of an outstanding range of high-quality scientific libraries on which to draw. HyperSpy (and its GUI, HyperSpyUI [3]) is

available for Linux, macOS and Windows. Under development since 2007, it has a large community of users and developers in the electron microscopy community and beyond. The main development of HyperSpy is an elegant yet powerful syntax for visualizing, analyzing, accessing and storing multi-dimensional datasets (regardless of their size). HyperSpy complements the scientific Python ecosystem by significantly simplifying the usage of external scientific libraries when operating on multi-dimensional data. The aim is to foster innovation in the field by lowering the entry barrier to advanced interactive data analysis and providing a quality peer-reviewed channel to distribute innovative algorithms. Indeed, many of the HyperSpy developers started as users and, in time, they started contributing new features that they had developed for their research.

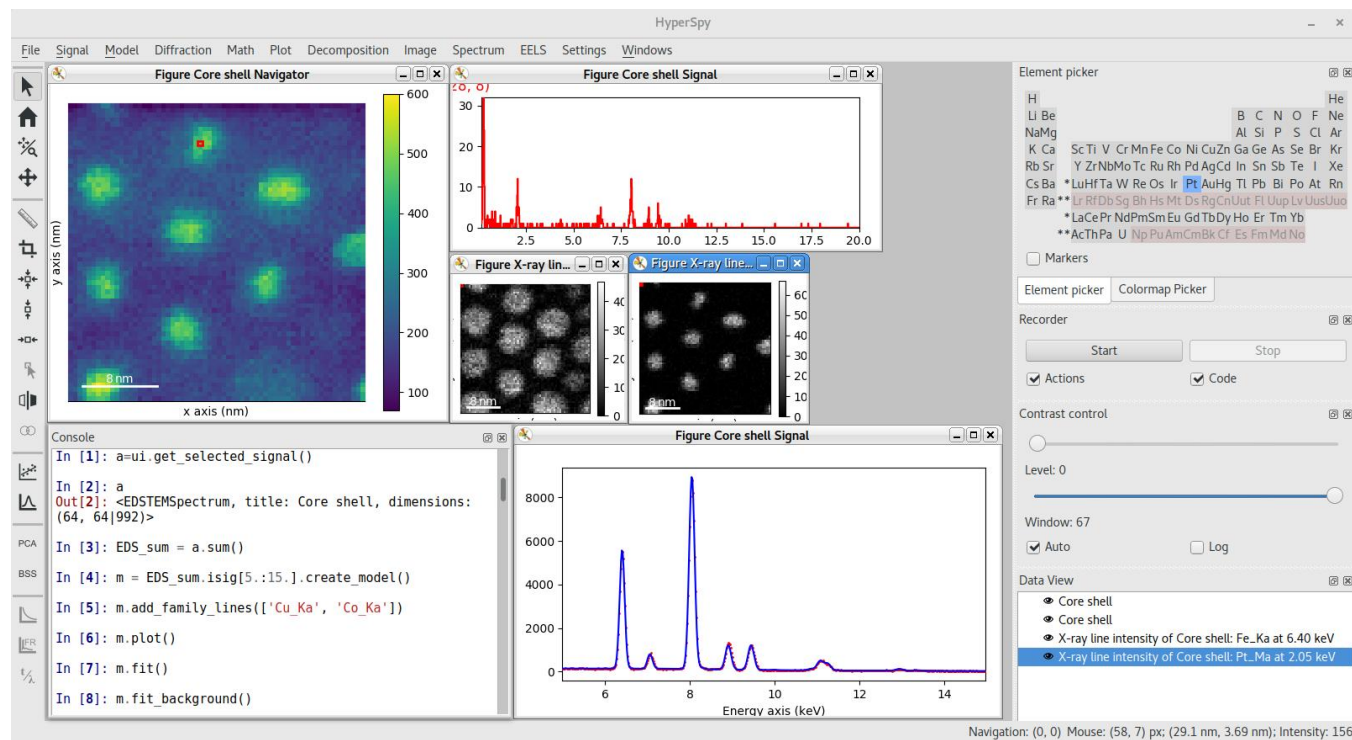
Features specific to electron microscopy include support for most common file formats along with functions for analysing electron energy-loss spectra (EELS), energy dispersive X-ray (EDX) spectra and performing holography. Further, HyperSpy provides easy access to principal component analysis (PCA), blind source separation and multi-dimensional model fitting in 1 and 2 dimensions. Most of its routines (as well as external functions) can be easily run in parallel and using out-of-memory computational schemes enabling seamless big data analysis.

[1] M. Abramoff *et al.*, *Biophotonics international* **11.7** (2004), p. 36.

[2] F. de la Peña *et al.* HyperSpy v.1.1.2 (2016). doi:10.5281/zenodo.60697. <http://hyperspy.org>

[3] V.T. Fauske, HyperSpyUI. <http://hyperspy.org/hyperspyUI/>

[4] The HyperSpy project has not received direct funding and all authors are grateful for support within their respective research groups to make the data analysis tools that they have developed available open-source.



**Figure 1.** Screenshot of HyperSpyUI [3] performing quantification by curve fitting of an EDX spectrum image of FePt core-shell nanoparticles. The embedded IPython console (bottom-left) enables interactive Python scripting.