



Enabling Science through European Electron Microscopy

Report on new optimised approaches for open instrument control, random and sparse sampling, machine learning, big data acquisition and automated analysis for multi-dimensional STEM

Deliverable D11.1 – version 2.2

Estimated delivery date: 30th April 2020
Actual delivery date: 29th May 2020
Lead beneficiary: JUM, UCAM
Person responsible: Caterina Ducati
Deliverable type: R DEM DEC OTHER ETHICS ORDP
Dissemination level: PU CO EU-RES EU-CON EU-SEC



THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT NO **823717**



Grant Agreement No:	823717
Funding Instrument:	Research and Innovation Actions (RIA)
Funded under:	H2020-INFRAIA-2018-1: Integrating Activities for Advanced Communities
Starting date:	01.01.2019
Duration:	48 months

Table of contents

Revision history log.....	3
Introduction.....	4
Data acquisition & reconstruction.....	4
Custom scan control.....	4
Binary electron detection.....	6
Data analysis platforms.....	6
HyperSpy extension packages.....	6
pyxem: crystallographic diffraction microscopy.....	7
LumiSpy: analysis of luminescence data.....	8
LiberTEM.....	9
Future development.....	10
Software.....	12
References.....	12

Revision history log

Version number	Date of release	Author	Summary of changes
V1	24/05/2020	Duncan Johnstone	First draft of the deliverable
V1.1	25/05/2020	Johan Verbeeck	Added contribution from U. Antwerpen
V1.2	25/05/2020	Dieter Weber	Updated section on LiberTEM
V2	27/05/2020	Duncan Johnstone	Integrated partner contributions
V2.1	28/05/2020	Lucie Guilloteau	General review
V2.2	29/05/2020	Peter van A. Aken	Final check

Introduction

Data acquisition and analysis in (scanning) transmission electron microscopy ((S)TEM) are undergoing radical transformations resulting from improvements in detector efficiency and computational power. This enables both better control of the microscope and analysis of the recorded data.

This deliverable D11.1 “Report on new optimised approaches for open instrument control, random and sparse sampling, machine learning, big data acquisition and automated analysis for multi-dimensional STEM” summarises recent developments associated with ESTEEM3 towards a paradigm shift in automated analysis for multi-dimensional STEM.

Data acquisition & reconstruction

The development of dose- and time- optimized data acquisition modes yielding ‘just enough data’ to obtain physical insight from multi-dimensional (S)TEM experiments depends on maximising performance of key hardware particularly for scan control and signal detection coupled with efficient reconstruction algorithms.

Custom scan control

A custom scan control module developed at CNRS-LPS laboratory (Laboratoire de Physique des Solides, Orsay, France) enables acquisition of spectrum-images in a random scan operating mode [1]. The pixels are acquired in a shuffled raster order and subsampled random sparse spectrum-images corresponding to specific time frame can then be extracted, as illustrated in Figure 1.

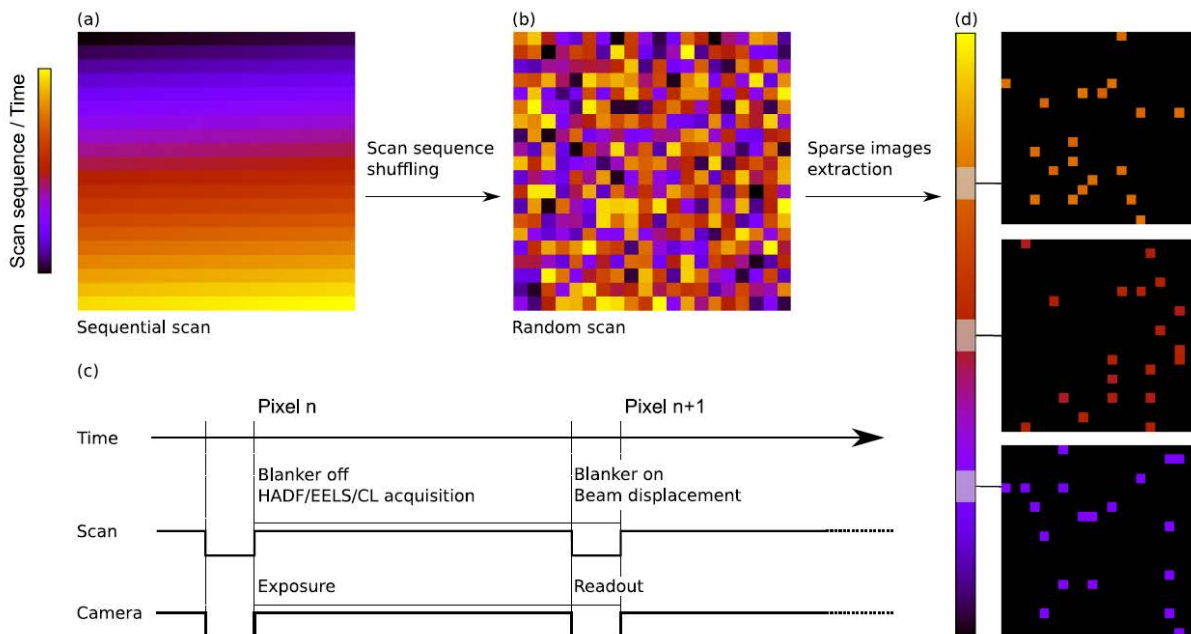


Figure 1: (a) A standard raster scan compared to (b) a random scan sequence, where the colour scale represents the pixel acquisition order. (c) Schematic representation of the synchronization between illumination, beam displacement, blanking, camera exposure, and read-out. (d) Extraction of sparse random images at given time frames. Reproduced from [1].

Fast spectrum-image reconstruction methods were developed by CNRS in collaboration with N. Dobbigeon and coworkers at IRIT/INP-ENSEEIH (Toulouse, France), exploiting the spectral and spatial structures of the spectrum-images in different ways [2,3]. These reconstructions can be used to track sample instabilities or evolution through a time sequence. For sensitive samples, it is possible to consider the spectrum-image corresponding to the first frames, and thus minimizing the total dose received by the sample and the subsequent damage. An example of a drift-corrected image and reconstructed spectrum-image of the same sample obtained using this approach is shown in Figure 2.

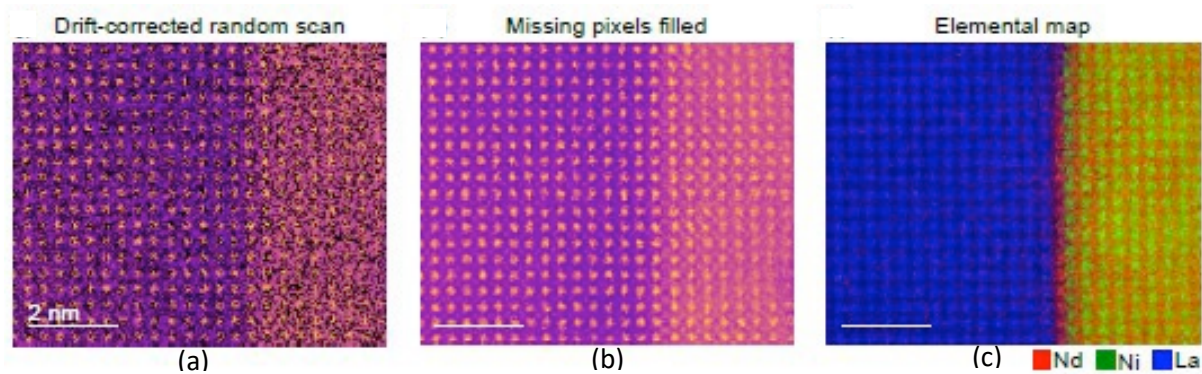


Figure 2: STEM-HAADF image of the NdNiO₃-LaAlO₃ (NNO/LAO) interface, (a) Drift-corrected random-scan HAADF image, and (b) the same image reconstructed. (c) Atomically resolved elemental maps obtained from the core-loss EELS hyperspectral image acquired simultaneously with the random-scan HAADF image. Reproduced from [1].

Drift correction in real time would enable dose-optimised area selection for spectrum imaging and is under current development. This development will be enabled by the acquisition of HAADF images simultaneously to the spectrum-image 5 times faster than previously. The scan hardware is already capable of real time correction and will be possible with a sufficiently fast algorithm to determine the appropriate drift correction.

The newly obtained freedom in scan patterns offered by this development at CNRS-LPS was used at University of Antwerp to implement several novel schemes to optimise strain mapping for e.g. semiconductor devices [4,5] and beam damage reduction in STEM imaging [6] and STEM tomography [7].

Binary electron detection

Modern fast pixelated detectors have enabled imaging of individual convergent beam electron diffraction patterns in a 4D-STEM raster scan at frame rates in the range of 1000–8000 Hz using conventional counting modes. Changing the bit depth of a counting detector, such that only values of 0 or 1 can be recorded at each pixel, allows one to decrease the dwell time and increase the frame rate to 12.5 kHz, reducing the electron exposure of the sample for a given beam current and thus optimising dose and time [8].

Phase retrieval using binary ptychography as a low-dose (ca. 200 $e/\text{\AA}^2$) 4D STEM technique has recently been demonstrated at the University of Oxford using a Merlin-Medipix detector from Quantum Detectors. This approach was demonstrated to yield atomically resolved phase contrast of an aluminosilicate zeolite (ZSM-5) is observed from sparse diffraction patterns with isolated individual electrons, as illustrated in Figure 3.

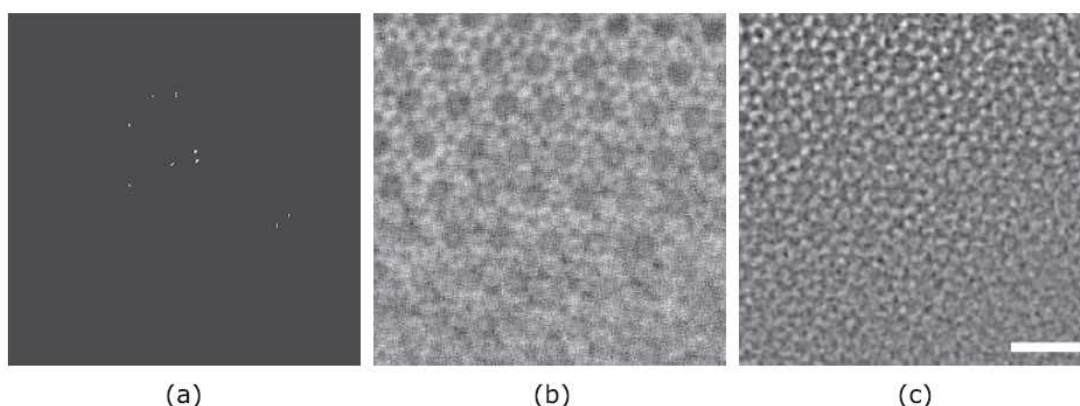


Figure 3: a) Example CBED pattern sequence, (b) iCoM reconstruction, and (c) SSB reconstruction for ZSM-5 using a cumulative electron dose of 200 $e/\text{\AA}^2$. The scale bar for the phase reconstructions is 2 nm. Reproduced from [8].

Data analysis platforms

User-friendly and readily extensible data analysis platforms, harnessing machine learning, are required to glean maximal insight from minimal data thus enabling further dose- and time-optimisation of acquisition. The open-source development of such platforms ensures developments are effectively shared across the community to enable more automated analysis [9]. Here, we report on the developments to extend application specific functionality associated with HyperSpy as a leading open-source library in electron microscopy, and complementary efforts to establish a high-throughput data processing framework for electron microscopy in the LiberTEM project.

HyperSpy extension packages

HyperSpy provides a framework for multi-dimensional data analysis in the electron microscopy community and beyond [A]. Since 2011, HyperSpy has received contributions from numerous ESTEEM3 partner institutions including: CNRS-LPS, University of Cambridge (UCAM), Norwegian University of Science and Technology (TRO), and Forschungszentrum Juelich (JUL). During recent years researchers from these institutions have added functionality for quantitative EELS, EDX and electron holography analysis. HyperSpy development now drives towards the creation of hyperspy-extension packages for specific applications such as *pyxem* and *lumispy* currently being developed by ESTEEM3 partners and described here.

pyxem: crystallographic diffraction microscopy

Pyxem is a hyperspy-extension library written in python and developed for multi-dimensional diffraction microscopy using, for example, 4D-STEM data [B]. The development of pyxem is lead primarily from the University of Cambridge and the Norwegian University of Science and Technology. Objects arising in 4D-STEM data analysis workflows are provided specialized classes in the pyxem library, which are categorized as:

1. **Signals:** Objects associated with raw or processed experimental data.
2. **Generators:** specify simulation parameters or manage operations requiring multiple inputs.
3. **Components:** sub-classes of hyperspy.Component. Objects that are fit to data in models.
4. **Libraries:** Dictionary objects containing simulation results.

Objects in these pyxem classes may be processed together in various ways to construct a wide range of analysis workflows, as illustrated in Figure 4.

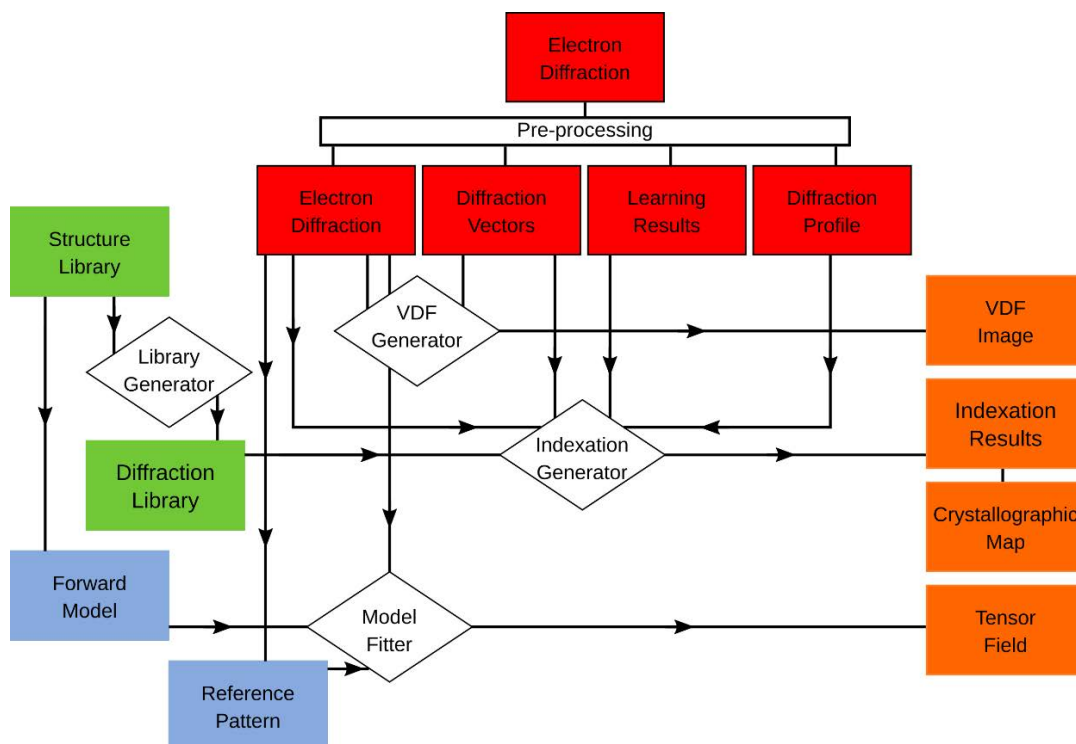


Figure 4: Flow diagram of 4D-SED data analysis workflows using pyxem object classes. Green = libraries, blue = components, white = generators, red = data signals, orange = result signals.

To incorporate prior knowledge of likely crystal phases measured in scanning electron diffraction (SED) data, simulations of the expected electron diffraction patterns, or expected diffraction vectors for all crystal orientations, are required. A separate python library, diffsims, was written to meet this need and integrated into pyxem [C]. This prior knowledge can be used to produce crystallographic maps specifying the crystal phase and orientation at each probe position.

Unsupervised machine learning approaches have been applied directly to SED data to learn component diffraction patterns associated with each distinct crystalline volume in the sample [10]. Using prior knowledge of crystal phases and the expected occurrence of special crystallographic relationships to analyse crystallographic maps directly was explored as a potentially more data efficient route to obtain materials insight [11]. A python library, orix, was created to enable the application of clustering algorithms as unsupervised learning techniques to such maps. This approach enables, for example, special crystallographic relationships to be identified, as illustrated in Figure 5.

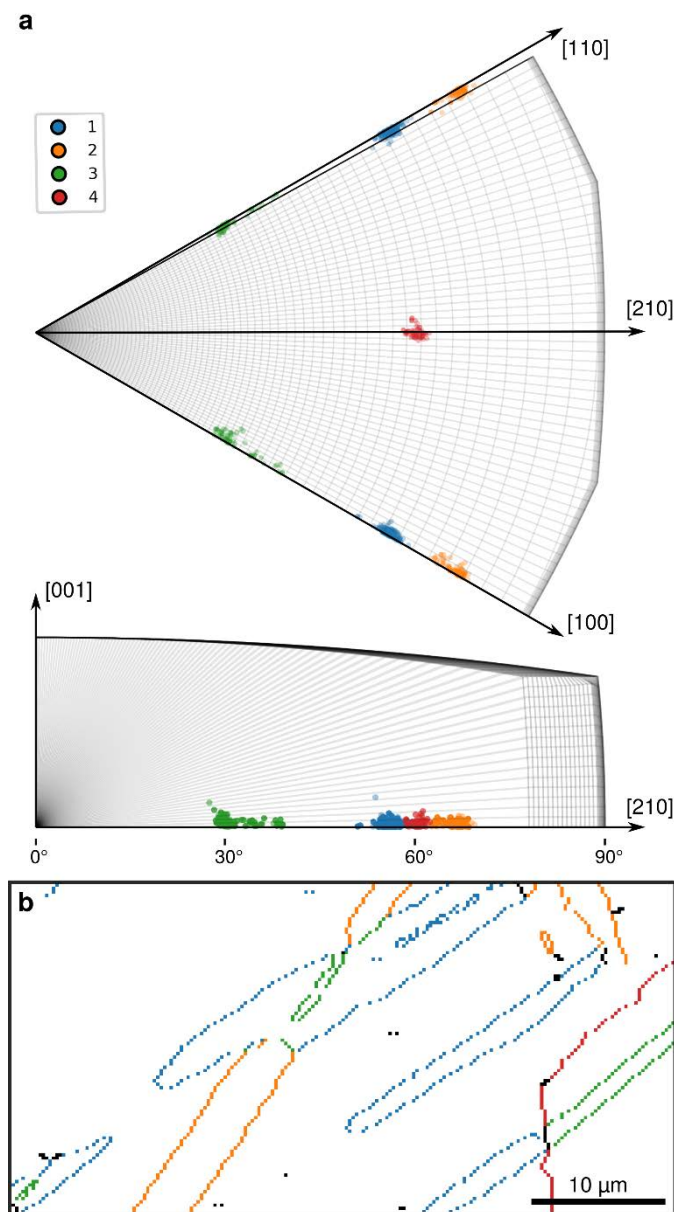


Figure 5: (a) Crystal misorientations plotted in the fundamental zone for the symmetry group pair (622, 622) in axis-angle space and coloured to indicate cluster membership as determined using density-based clustering. Clusters also correspond to special twin boundaries in this instance. Axes are labelled in the crystallographic basis at no rotation. (b) Map of grain boundaries coloured by cluster membership of the misorientation at with each boundary element. Reproduced from [11].

[LumiSpy: analysis of luminescence data](#)

The increasing prevalence of STEM cathodoluminescence experiments has motivated researchers at the University of Cambridge to establish a hyperspy-extension package, written in python, for the analysis of luminescence data [E]. The development of specialized analyses for this package is underway. This work has also motivated the development of the core hyperspy package to support non-linear axes necessary for proper luminescence analysis.

LiberTEM

LiberTEM has been developed specifically as an open-source a platform for efficient, parallel and distributed data processing in the electron microscopy community led by researchers from Forschungszentrum Juelich [F]. This project is also linked to ESTEEM3 Deliverable 1.4a on open software and data, including practicing open data workflows to develop best practices.

The LiberTEM project has so far:

- Developed a user-defined function (UDF) interface to specify algorithms in a way that they can achieve high throughput on distributed systems and are suitable for processing live data. It follows a MapReduce programming model adapted for binary data.
- Implemented I/O pipeline for high-throughput distributed data processing on various file formats related to electron microscopy.
- Optimized implementation of a number of algorithms relevant for electron microscopy using the user-defined function interface, including cross-correlation, strain mapping, virtual detectors, standard deviation mapping, gain and dark reference correction, generate feature vectors for clustering.
- Interfaced with existing solutions including Nion Swift and Gatan Microscopy Suite.
- Designed back-end support for large-scale distributed live ptychography.
- Requirements analysis and system specification for TEM data management at Forschungszentrum Juelich based on open source components
- Requirements analysis and rough specification for future high-speed TEM automation

These developments position LiberTEM as a critical platform for high throughput electron microscopy data processing particularly focused around the user-defined function interface illustrated in Figure 6. This interface will enable other packages in the wider electron microscopy community to benefit from improved computational efficiency.

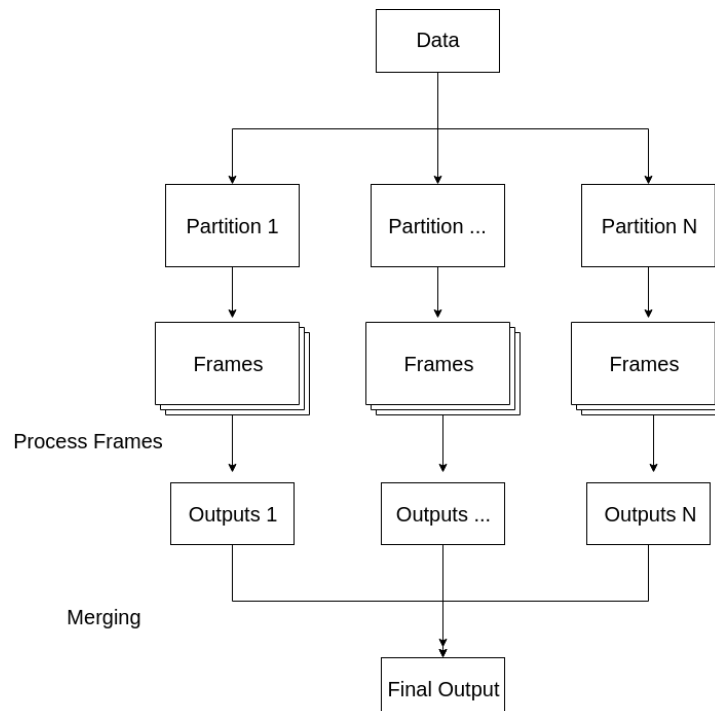


Figure 6: Schematic illustration of LiberTEM user-defined interface for efficient partitioning and processing of data to obtain a final output.

Future development

Future development of data acquisition and analysis in STEM will build on the developments in scan control, signal detection and software platforms described in this report. Figure 7 shows a recent prototype running a list of LiberTEM user-defined functions on live data from a Quantum Detectors Merlin camera to achieve live ptychographic phase reconstruction. The progressing results are plotted in real time as the data comes in from the detector.

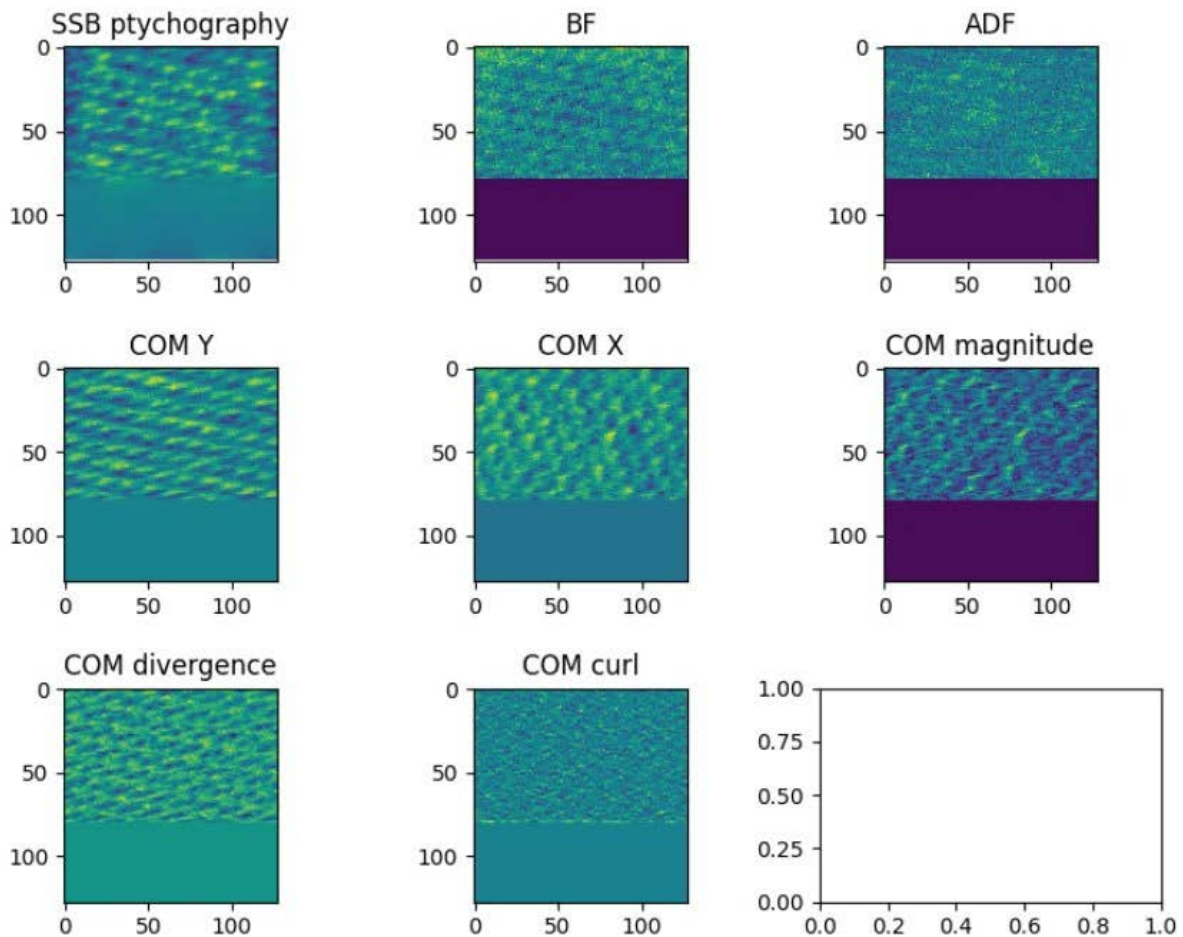


Figure 7: Simultaneous live data processing for single side band ptychography, bright field virtual detector, annular dark field virtual detector and center of mass analysis using a recent prototype for running LiberTEM user-defined functions on live data. It accepts data from a Quantum Detectors Merlin camera and runs an arbitrary list of user-defined functions on the data portions as they come in. The results can be plotted in a live-updating display, including continuous scanning.

Integrating advanced analysis with efficient computation and acquisition control will enable further dose- and time- optimised acquisition as well as greater material insights. Incorporating standardized file formats and robust data management solutions will complete the picture, as illustrated schematically in Figure 8. This total integrated acquisition and analysis framework will enable greater materials insight via smart microscopy achieved both live and long term.

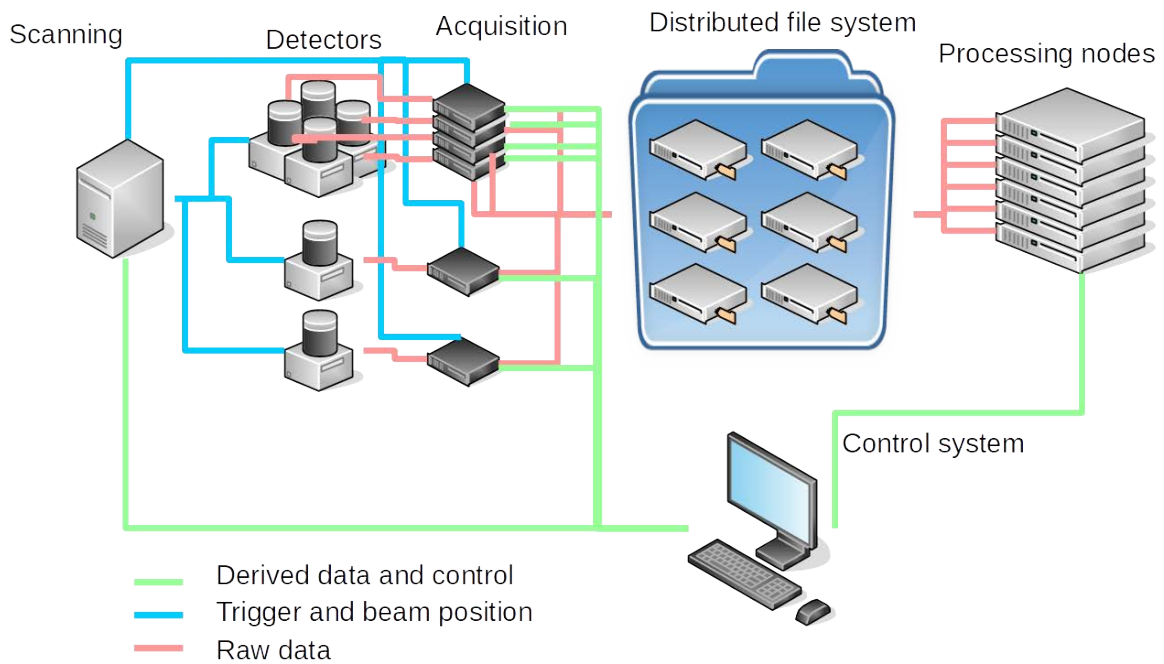


Figure 8: Schematic illustration of an integrated acquisition, analysis, and data management framework for real time and long-term smart microscopy.

Software

[A] www.github.com/hyperspy/hyperspy

[B] www.github.com/pyxem/pyxem

[C] www.github.com/pyxem/diffsims

[D] www.github.com/pyxem/orix

[E] www.github.com/LumiSpy/lumispay

[F] www.github.com/LiberTEM/LiberTEM

References

[1] Zobelli, A., Woo, S. Y., Tararan, A., Tizei, L. H., Brun, N., Li, X., ... & Tencé, M. (2019). Spatial and spectral dynamics in STEM hyperspectral imaging using random scan patterns. *Ultramicroscopy*, 112912.

[2] Monier, E., Oberlin, T., Brun, N., Tencé, M., de Frutos, M., & Dobigeon, N. (2018). Reconstruction of Partially Sampled Multiband Images—Application to STEM-EELS Imaging. *IEEE Transactions on Computational Imaging*, 4(4), 585-598.

[3] Monier, E., Oberlin, T., Brun, N., Li, X., Tencé, M., & Dobigeon, N. (2020). Fast reconstruction of atomic-scale STEM-EELS images from sparse sampling. *Ultramicroscopy*, 112993.

[4] Velazco, A., Nord, M., Béch e, A., Verbeeck, J. (2020) Evaluation of different rectangular scan strategies for STEM imaging, arXiv:2002.09370

[5] Prabhakara, V., Jannis, D., B ech e, A., Bender, H., Verbeeck, J., (2019). Strain measurement in semiconductor FinFET devices using a novel moir e demodulation technique, *Semiconductor Science and Technology*, 35 (3), 034002.

[6] Van den Broeck, W., Reed, B.W., B ech e, A., Verbeeck, J., Koch, C.T., (2019). Viability of Compressed Sensing as a Dose Reduction Strategy in STEM, *Microscopy and Microanalysis* 25 (S2), 1686-1687.

[7] Vanrompay, H., B ech e, A., Verbeeck, J., Bals, S. (2019). Experimental evaluation of undersampling schemes for electron tomography of nanoparticles, *Particle and Particle Systems Characterization*, 36 (7), 1900096.

[8] O’Leary, C.M., Allen, C.S., Huang, C., Kim, J.S., Liberti, E., Nellist, P.D., Kirkland, A., (2020). Phase reconstruction using fast binary 4D-STEM data. *Applied Physics Letters*, 116, 124101.

[9] Nord, M. and Verbeeck, J. Open source development tools for robust and reproducible electron microscopy data analysis, *Microscopy and Microanalysis*, 25 (S2), 138-139.

[10] Martineau, B.H., Johnstone, D.N., van Helvoort, A.T.J., Midgley, P.A., Eggeman, A.S., (2019). Unsupervised machine learning applied to scanning precession electron diffraction data. *Advanced Structural and Chemical Imaging*, 5, 1-14.

[11] Johnstone, D.N. Martineau, B.H., Crout, P., Midgley, P.A., Eggeman, A.S., (2020). Density-based clustering of crystal orientations and misorientations and the orix python library. arXiv:2001.02716.