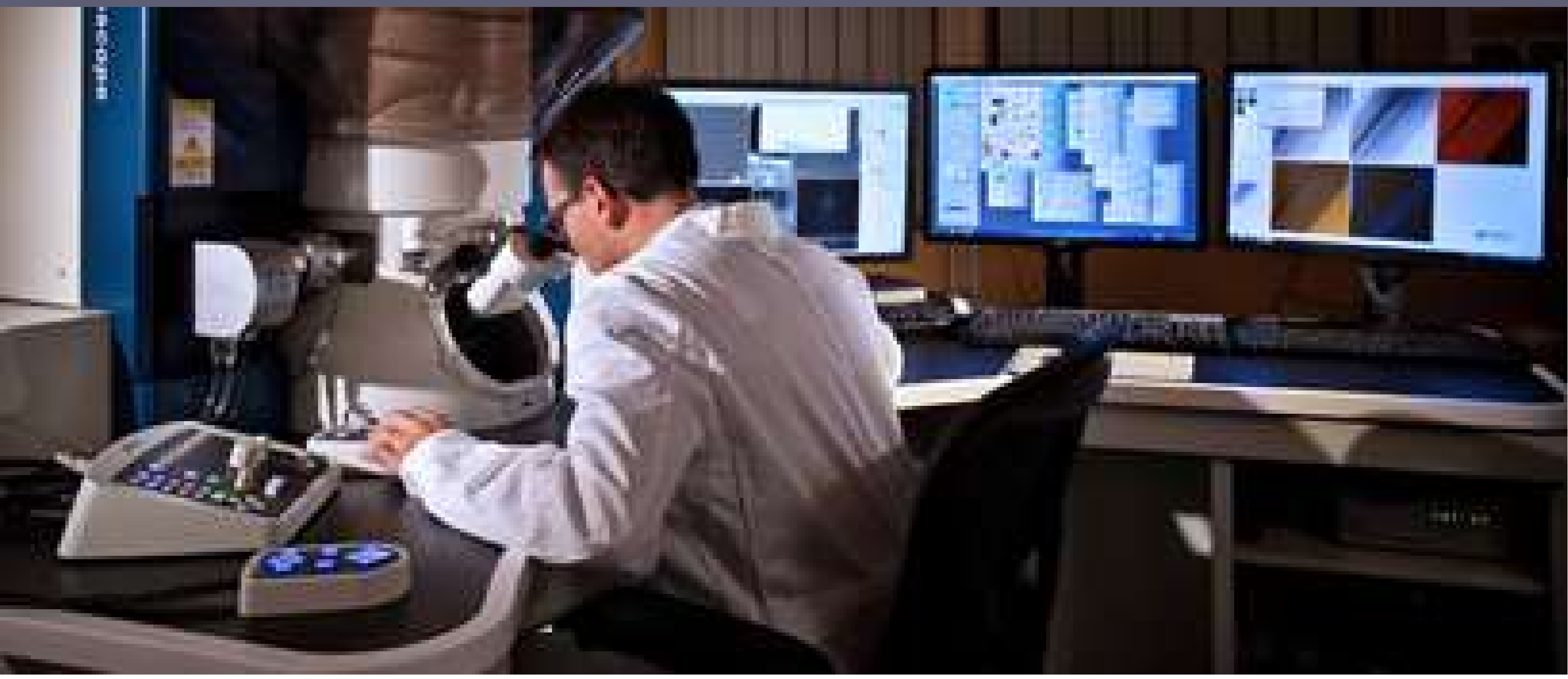
esteem3

Enabling Science through European Electron Microscopy

Report on open source solutions for dose- and time-optimised spectroscopic data, including optimised sampling, compressed sensing, machine learning, handling big data and integration of data acquisition and processing with simulations

Deliverable D11.2 – version 0.2

Estimated delivery date:	30th of June 2021
Actual delivery date:	23rd of June 2021
Lead beneficiary:	JUL
Person responsible:	
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:	823802
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

Version number	Date of release	Author	Summary of changes
V0.1			First draft of deliverable
V0.2			Include feedback from partners
V1	21/06/2021	Peter van Aken	Approval of deliverable

Introduction

This document is a follow-up to report D11.1, since spectroscopic data is similar to the multi-dimensional STEM data that was covered previously. It focuses on new developments since that report and aspects that are specific to spectroscopic data.

Tasks

A recap of the tasks related to this deliverable:

- The development of standardised file formats and user-friendly open source solutions for spectroscopic data processing (WP5) for handling "big data" and implementing new advanced data analysis schemes
- The development of user-friendly, practical software interfaces for spectroscopic data acquisition, management, archiving, off-site processing, access and cloud storage
- The integration of dose- and time-optimised spectroscopic data acquisition and processing, including sparse and random sampling, compressed sensing, in-painting, vertex component analysis, control interfaces and machine learning within an open control environment
- The integration of spectroscopic data acquisition and processing with optimised simulations

The report is structured roughly along these tasks.

File formats

The larger underlying issue still persists: file formats are not standardized and often don't contain all the metadata that is required for correct interpretation. As a pragmatic approach, open source file readers for the various existing formats are being developed and metadata is documented in other ways for the time being. Publishing executable analysis scripts that transform the raw input data into the final result for publication in an unambiguous way can be an approach to document all required metadata and the correct interpretation of a file format.

LiberTEM

Within the LiberTEM project, the range of supported file formats was extended. A feature that lets the user define the shape of the scan independent of the file metadata and to correct any offset between the recorded data the scan pattern was included. Furthermore, the file reading back-end was restructured to give more flexibility and control over the subdivision of data processing tasks. This greatly improved performance in many relevant applications.

Furthermore, LiberTEM now allows fast on-the-fly correction of detector data to account for the dark current, gain of individual pixels and non-functioning pixels. Embedded correction information that is available in some file formats, as well as user-supplied correction data are supported.

Data analysis, data acquisition, instrument control

These aspects increasingly merge since live data analysis is becoming more common.

Electron Energy Gain Spectroscopy

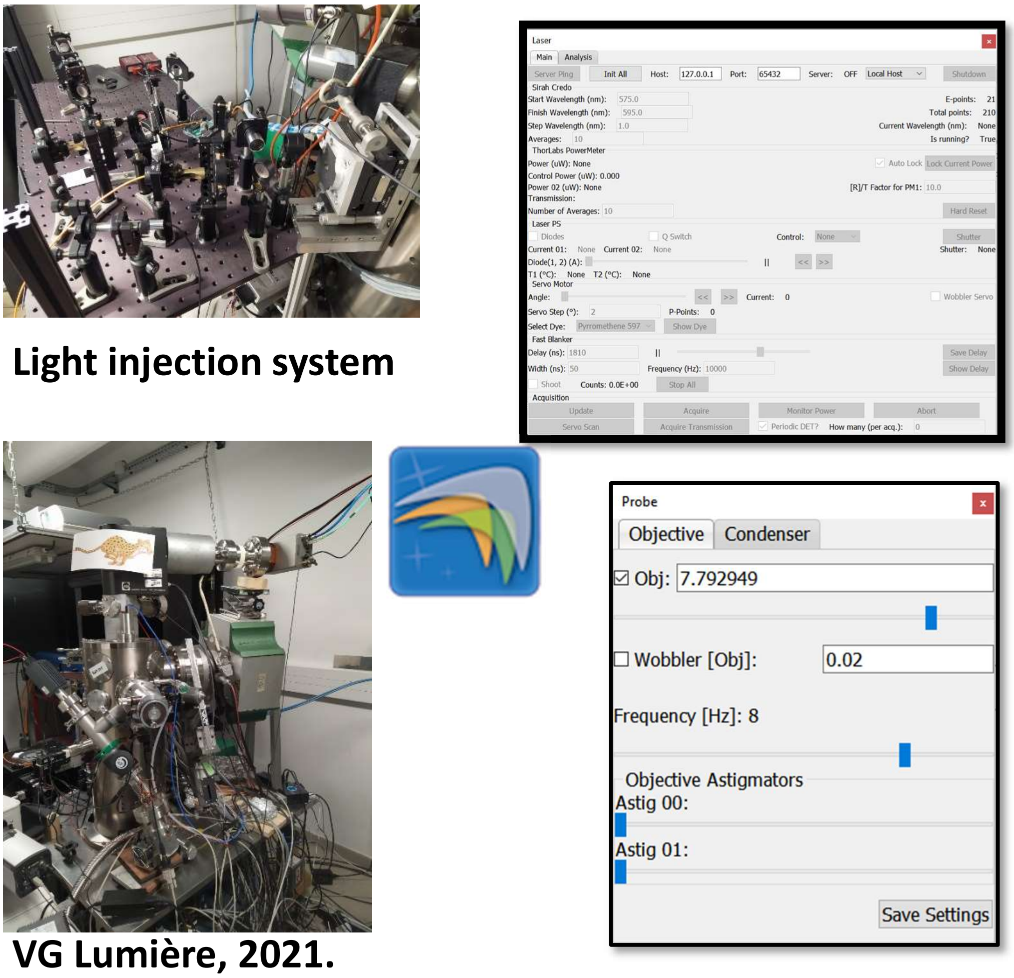


Figure 2: Electron energy gain spectroscopy with Nion Swift (ORS)

An electron energy gain spectroscopy set up has been developed entirely controlled with Nion Swift (ORS, Figure 2). It is compatible with a test bed 40 years old VG that is now entirely driven with Nion Swift. It is compatible with the NION-HERMES 200 (Chromatem) monochromated machine. The various plug-ins (microscope, scan, laser, CL control…) will be released open source. Y. Auad et al., unpublished (2021)

LiberTEM

In D11.1, first results for live data processing with LiberTEM using an early prototype implementation were reported. Since then, the implementation has matured greatly. In particular, it has become a lot more convenient to use. A release as open source software is pending[1].

The core benefit of using LiberTEM for this purpose is disentangling aspects of instrument control, camera interface, data processing and results display by creating well-defined interfaces between the components. This allows users to run an arbitrary set of complex user-defined processing routines on a live data stream or on offline data, using the same implementation for all supported file formats, detector models, microscopes and software environments, provided they can execute Python code.

[1] https://github.com/LiberTEM/LiberTEM-live

As a particular example, this allowed to run ptychography reconstruction, center of mass analysis and conventional STEM analysis live in parallel.

Such an architecture allows to run the same routines within different software environments, including Jupyter notebooks and Gatan Digital Micrograph. Previously, such routines had to be adapted to work with the interfaces of one particular environment.

The architecture of LiberTEM allows to scale the processing to high performance on distributed systems, which enables support for modern high-performance cameras, processing on high-performance systems in connected data centers, and using computationally intensive routines such as ptychography. Currently, the Quantum Detectors Merlin camera is supported. A proof-of-concept for the Gatan K2 IS camera was recently completed in collaboration with Gatan.

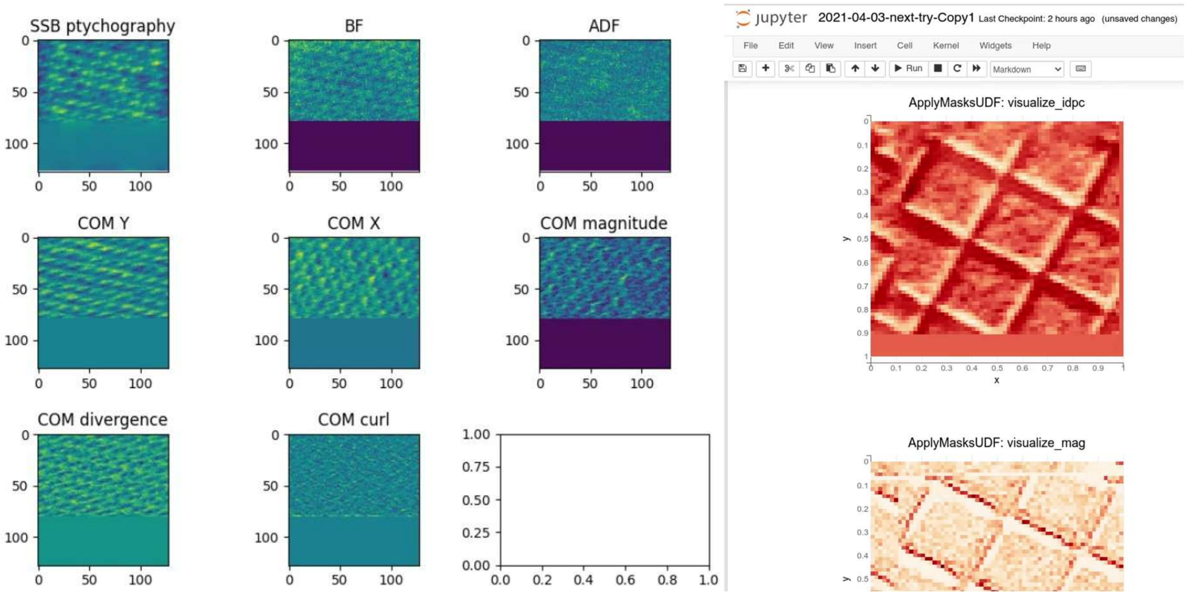


Figure 3: Screenshots of live SSB ptychography and center of mass iDPC with the Gatan K2 IS and LiberTEM.

This is also an example for the positive impact of ESTEEM3 on cooperation with industry. Creating open, standardized interfaces connects components from different vendors. Support for developing the underlying software through ESTEEM3 and other funding sources removes this burden from companies or individual research groups and facilitates cooperation and standardization. Such an infrastructure gives scientists flexibility to develop and share novel processing algorithms independent of instrument modalities, which enables open science. This is closely related with Task 1.4 in ESTEEM3.

Efforts to use LiberTEM as a back-end for processing in Hyperspy as well as supporting Dask arrays as input and output to allow composing different pieces of software into one processing flow are currently underway at JUL.

SerialEM

As part of these efforts to create standardized interfaces that allow applying the same routines and software for data analysis across various modalities, JUL collaborated with SerialEM developers to integrate support for Python scripting in SerialEM. This was recently released.

Digital Micrograph

The efforts to support Python scripting in Gatan Digital Micrograph that were described in D11.1 continue. That includes help and feedback to Gatan to improve the usability and quality of the Python scripting and developing examples that show how to use relevant Python packages within Digital Micrograph, since they need specific adjustments to work in this environment. This also includes support for running LiberTEM within Digital Micrograph, including live plotting.

Distributed matrix decomposition

Previous work by JUL on distributed, streaming matrix decompositions like non-negative matrix factorization and principal components analysis was picked up again within the AIDAS project in cooperation with CEA-Leti in Grenoble. This is a core algorithm for live analysis of spectroscopic data, as opposed to the offline processing that is currently common.

Hyperspy

The backend code associated with applying a function at each navigation position has been rewritten to make more effective use of Dask. This updates the code from its 2017 state. It is hoped this will lead to a seamless improvement in performance across all codes that depend on hyperspy.

Others

Efforts to automate various workflows, for example STEM focus series, data acquisition for ptychography and hologram acquisition using the infrastructure described above, are ongoing as part of various research projects. That includes aligning the quantum sorter using machine learning (JUL, within QSORT project) and automated stabilization of hologram acquisition at TOU, Figure 4.

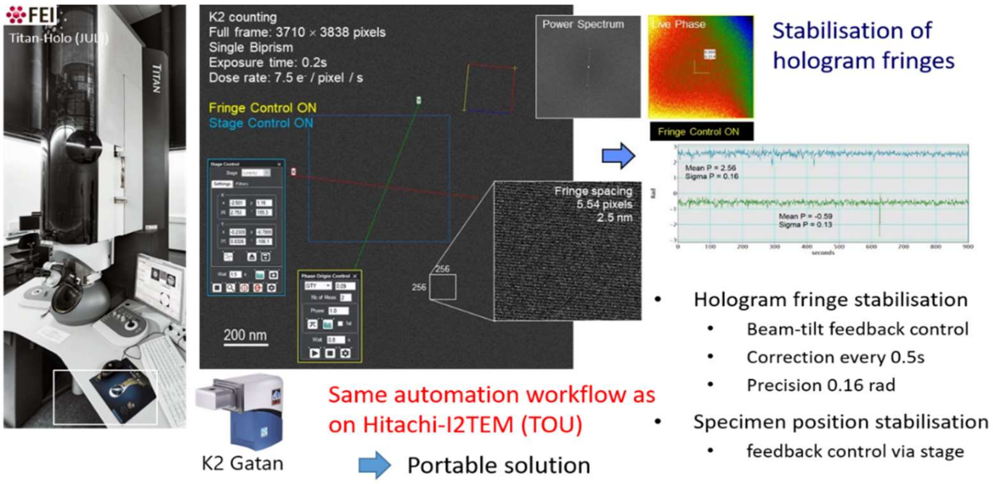


Figure 4: Hologram fringe stabilization (TOU)

Real-time drift correction for EELS hyperspectrum acquisition

Taking advantage of the high-speed of the new Medipix3RX Merlin EM detector, hyper-spectra can be recorded in sequences while adjusting the scanned area at ORS. This is applied to drift correction for mapping of Sc L and Dy M signals (Figure 5).

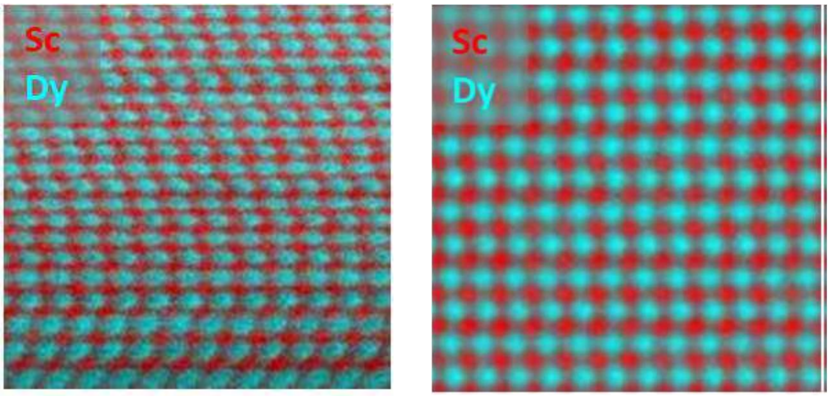


Figure 5: Conventional acquisition on the left, with drift correction by ORS on the right

Data management

In D11.1, requirements analysis and specification for a data management system were reported. By now, a prototype is operational at JUL and near ready for a wider rollout. Requirements include:

- Handle data and access many users, including guests from TA projects
- Data sharing inbound and outbound
- Allow defining and enforcing access restrictions
- Ingest from instrument computers
- Fast access from data processing systems
- Fast access on user devices
- Standardized interfaces: Nextcloud, WebDAV, network file systems
- Backup, integrity, long-term archival
- Extensible
- Integration with electronic lab books

Currently, the system is being tested and prepared for going online on the internet. The design and experiences so far are already being shared upon request and will be publicised once the system is mature and rolled out successfully. The development also includes suitable policies for data management.

Integration with simulation

Diffsims

This simulation framework was developed as part of pyxem originally but has since been split out to increase the modularity of the code. Capabilities are currently limited to kinematical electron diffraction simulation, but the package is design with extensibility in mind.

LiberTEM

LiberTEM allows to limit data processing to an arbitrary subset of the input data through a feature called “region of interest” (ROI). That can be used to simulate custom scan patterns, including sparse and random sampling. This was used successfully to investigate the characteristics of ptychography for live data processing and to test “divide and conquer” strategies.

Furthermore, the previously described efforts to use Dask arrays to connect different pieces of software that work on big data will allow to integrate on-the-fly simulations seamlessly into a processing workflow.

Outreach

ePSIC hosted a successful virtual Hyperspy workshop in April with over 100 attendees. A similar online workshop is planned for late June as part of NordTEMHub project.