Surface X-ray science analysis of large-scale facility data using scalable data representation

A Python[1] based analysis frame in selected surface X-ray analysis cases

Peter O. Magnusson[1], Uta Hejral[1], Gary Harlow[1], Anders Mikkelsen[1], Edvin Lundgren[1]

[1]*Synchrotron Radiation Research, Department of Physics, Lund University, Lund, Sweden*

peter.magnusson@sljus.lu.se

An analysis software solution is described, as based on requirements from a material scientist user community, performing surface X-ray scattering experiments at synchrotrons. The full acquired data sets from surface X-ray scattering experiments performed at large scale facilities occupy sizes typically exceeding the memory capacity of standard computer platforms. The off-line analysis therefore often in practice involve complex data inspection and reduction procedures. A platform capacity adaptable representation of the data would be beneficial, that allow for n-dim-sliceable data subsets extractions and analyses. An analysis frame solution was developed in Python[1] to cover selected surface X-ray analysis cases. A scalable and instrument-agnostic solution was aimed for. The analysis cases, selected from high-energy surface X-ray diffraction (HESXRD) measurements[2], were static sample temporal resolved measurement and sample rotation measurement. A configuration-class is holding ancillary data acquisition information, and a set of worker classes performs translations to the standardized hierarchical data format HDF5[3] via h5py[4]; q/hkl-space mapping via interface to Binoculars[5]:PyMCA[6] modules; and data visualization, measurement and fitting. Multiple results data sets, as e.g. multiple extracted and mapped crystal truncation rods (CTR), are kept organized in the hierarchical structure of HDF5. Multiple objects for data acquisition setting, data sub-set slicing selection and visualization selections are held by Python-dictionaries. q/hkl-space mapping of the full data set, approaching preserved acquired resolution, can be challenging also at default mode HPC-cluster node performance. Mapping and analysis of the extracted data CTR's or delta-theta-plane sub-stacks could be performed at high resolution with the suggested analysis solution at standard computer capacity, and as n-dim-sliced from the kept full acquired data set. The suggested solution provide a platform capacity adaptive data representation, a configuration and worker class structure, and the full acquired data set and result data sets organized in the hierarchical structure of HDF5. The solution was demonstrated applicable to selected surface X-ray diffraction analysis cases and is extendable to additional analyses cases.

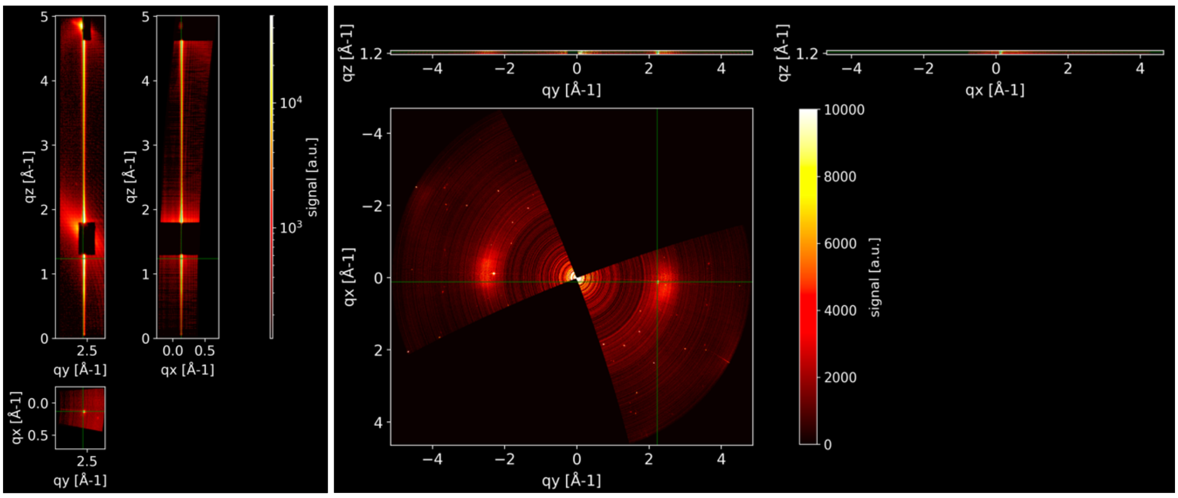

Fig. 1. Orthogonal view of extracts, mapped to q-space at standard computer platform capacity. The extracts examples are crystal truncation rods (CTR) (left, 130x836x162, Mapped resolution: 0.006 $Å^{-1}$) and delta-theta-plane sub-stacks (right, 837x10x779, Mapped resolution: 0.012 $Å^{-1}$) from a Pb-sample high-energy surface X-ray diffraction sample rotation experiment[2].

References:
1. Python Software Foundation, https://www.python.org/
2. Gustafson J. et.al., High-energy surface X-ray diffraction for fast surface structure determination. Science 343 (6172), 758-761, 2014.
3. The HDF Group., http://www.hdfgroup.org/HDF5
4. Collette A., HDF5 for Python, http://h5py.alfven.org, 2008.
5. Roobol S. et.al., J.Appl.Cryst. 48,1324–1329, 2015.
6. Solé V.A. et.al., A multiplatform code for the analysis of energy-dispersive X-ray fluorescence spectra, Spectrochim. Acta Part B 62: 63-68, 2007.