

XPCS Analysis

Summary

XPCS (X-ray Photon Correlation Spectroscopy) is experimental technique used at [Sector 8 of APS](#). The main principle behind the technique is to study the dynamics in materials at nano-scale by studying correlations in time series of area detector images. These correlations involve analyzing the pixel-by-pixel correlations for different time intervals. A multi-tau algorithm is commonly used that defines the time intervals in a logarithmic manner. The current state-of-the-art detector at Sector 8 is capable of acquiring megapixels of frames at a rate of 60 Hz (120 MB/sec). A suite of CCD detectors are used with data rates ranging from 4 MB/sec to 120 MB/sec. The next generation Fast CCD2 detector being built in collaboration between ANL and LBL is projected to run at 200 Hz generating 400 MB/sec of data. The challenge is to compute these correlations at a decent rate that comes within a small factor of the rate of the data acquisition. Doing the analysis at (near) real time saves a lot of effort and time of the scientists.

Team

1. [Faisal Khan \(SSG\)](#)
2. [John Hammonds \(SSG\)](#)
3. [Nicholas Schwarz \(SSG\)](#)
4. [Suresh Narayanan \(XSD\)](#)
5. [Alec Sandy \(XSD\)](#)

Download

Version 0.5 (Released)

Here is a brief list of features in this version:

- Implementaiton of two-times algorithm with optimized data handling.
 - Added secondary sorting to speed up the computation for two-times and reducing number of jobs to 1.
 - Uniform code for two-times and multitau to allow application of same set of pre-processing filters.
- Major code refactoring
 - The file format is independent of key-value processing. This will allow us to plug-in different file formats in future.
 - Job definition is now its own class
- Added a new pre and post processing pixel filters.
- These filters allow parallel implementation of operations like pixel binning within a frame, dark image thresholding, flat field correction, pixel stride, pixel summation et.

Version 0.4.1 (Released March, 04 2014)

Get it from the subversion ([SVN](#))

The version 0.4.1 is the most complete version that is in production since late 2013 at Sector 8.

Here is a brief list of features in this version:

- The complete multitau analysis of data in all three formats :Sparse, Non-Sparse, and Kinetics.
- Process the IMM file stored directly in the Hadoop file system (HDFS).
- Read the job configuration parameters from the HDF5 file (See the confluence page for more information on how the hdf5 parameters are interpreted along with their type:<https://confluence.aps.anl.gov/display/XPCS/Job+Parameters>)
- The final results produced by the MapReduce are converted to a format that can be written as a HDF5 dataset. (See the confluene page<https://confluence.aps.anl.gov/display/XPCS/HDF5+Output+Format> for more details on the output format)
- This version provides a partial fix for the concurrency issues involving processing of input parameters from HDF5 file. Our code reads the input parameters from the fixed location within the HDF5 file. As of this version it is the dataset named '/xpcs'. Before this version the '/xpcs' would be read twice: first, for configuring the job parameters (input, qmap etc), second, for determining the location of the output dataset in the same HDF5 file. The problem arises when an external program (e.g. Suresh's Matlab) modifies the value under '/xpcs' in between these two reads.

We fixed it by reading all the information needed including the location of the output folder. However, this is a problem that require a bit re-engineering in particular when it will come to submitting multiple jobs using the pipeline.

Details

Design of the Hadoop job

Updated getting started guide

Hadoop Cluster (Ask Roger to fill this up?)

Links

- [Design and Development](#)
- [FCCD2 ATCA Analysis](#)
- [HDF5 \(LZ4\) Dataset to IMM Converter](#)
- [Magellan](#)
- [Meeting Minutes](#)
- [Pipe line](#)
- [Plan TRR Toolkit Dec-Feb](#)