FLYSCAN: A DISTRIBUTED FAST ACQUISITION SYSTEM FOR MULTI-DETECTOR EXPERIMENTS

F. Langlois, A. Buteau, X. Elattaoui, C. M. Kewish, S. Le, N. Leclercq, P. Martinez, K. Medjoubi, S. Poirier, A. Somogyi, SOLEIL Synchrotron, L'Orme des Merisiers, 91190 Gif-sur-Yvette, France
A. Noureddine, Médiane Système, 54 route de Sartrouville, 78232 Le Pecq, France
C. Rodriguez, GIST, 40 av André Morizet, 92513 Boulogne-Billancourt, France

INTRODUCTION

An increasing number of synchrotron beamlines need to use several detection techniques in parallel, and in fast acquisition modes. For example, a scanning microspectroscopy beamline could simultaneously collect data from two-dimensional (2D) position-sensitive X-ray detectors, one-dimensional (1D) energy-dispersive fluorescence analyzers, and several single-element beam intensity monitors.

Continuous scanning modes for fast data collection are required for scientific experiments conducted on such beamlines. That is, measuring signals from all detectors "on-the-fly" while actuators are moving, in order to avoid significant time overheads that are introduced by motor settling times in step-scan mode, and allow an experiment to be completed within a reasonable time frame. The implementation of this so-called "flyscan mode" is made possible by the progress in detector technology and the high photon fluxes available from 3rd generation synchrotron sources such as Soleil. The large amount of data produced in such an experiment has to be managed, either off-line and on-line, using software tools to analyze the data quality during collection and for categorizing and retrieving information for further scientific analysis.

Relying on our Tango [1] based software distributed architecture, a set of C++/java libraries, Tango devices and GUI applications has been developed to fulfill these requirements: (*i*) modularity, in terms of the number and type of detectors; (*ii*) performance, to handle these large data volumes; and (*iii*) data management, to store and retrieve information in a coherent way.

Taking as an example the challenging data acquisition demands of the future Soleil beamline Nanoscopium, this paper will present the status of development, deployment and operation of this "distributed fast acquisition system for multi-detector experiments".

THE SCIENTIFIC CASE OF THE NANOSCOPIUM BEAMLINE

The Nanoscopium beamline of Soleil will be dedicated to state-of-the-art scanning X-ray imaging techniques in the 5 – 20 keV range. It will provide unique research opportunities by combining the analysis of sample chemistry, *via* X-ray fluorescence and absorption spectroscopy, with structural analysis from coherent diffractive imaging at high spatial resolution (= 30 nm) in two and three dimensions. Elemental distributions and oxidation states can be quantified at the trace level in geological and biological samples for most of the elements starting from Ti. Fluorescence spectroscopy will target elements as light as phosphorous. The beamline is in the construction phase and will be operational in 2013.

These different detection schemes are to be used in flyscan mode in order to measure statistically relevant sample areas (up to several mm^2) and to perform tomographic scans without significant time overhead. Fast data collection will also result in reduced degradation of the cryo-cooled samples during the imaging experiments. Figure 1 shows a typical simultaneous detection mode.

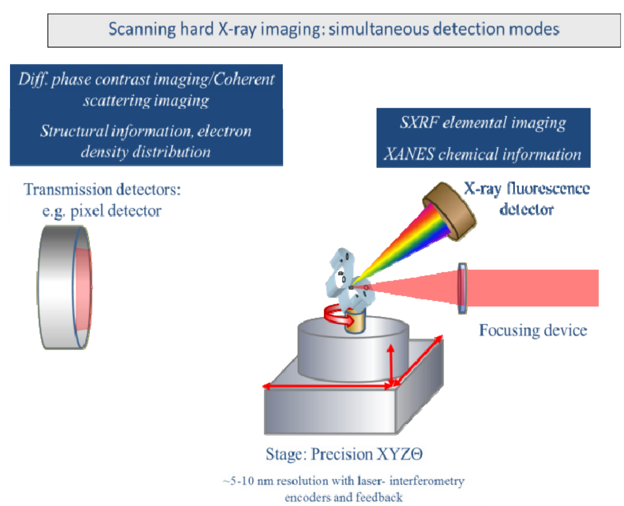

Figure 1: Schematic illustration of a multiple-detector fluorescence microspectroscopy experiment.

Multi-Detector Fast Acquisition System (FLYSCAN) Specifications

- Acquisition subsystems are independent and are synchronized by a common master clock trigger signal;
- Acquisitions must be done while systems are in motion to minimize data collection time;
- It must be possible to add a new acquisition subsystem to the experiment, without having to change existing software;

- Acquisition systems are running on various hosts (or crates), distributed on a TCP/IP network;
- All data collected on these various and independent subsystems must be recorded in a unique set of NeXus [2] files, encapsulating sufficient "metadata" (*e.g.,* experiment and sample identifiers, beamline parameters, time and date stamps) to simplify data reduction and analysis, facilitate reprocessing of the data in the future;
- Data organization within the files must be decoupled between acquisition and data processing to be able to add/remove detectors into the acquisition system without impacting data processing applications, even though the data organization within the files may be different.

SOFTWARE ARCHITECTURE

Data from each detector are collected by independent software applications (in our case Tango Devices), assuming all acquisitions are triggered by a unique master clock. Then, each software device streams its own data onto a common disk space, known as the spool (see Fig. 2). The data are stored in independent temporary NeXus files, with the help of a dedicated high-performance C++ library called NeXus4Tango. A dedicated asynchronous process, called the DataMerger, monitors the spool, and gathers all these temporary files into the final experiment NeXus file stored in SOLEIL common Storage System. Metadata describing the context of the data are also added in the final file by the DataRecorder device. All these actions are sequenced into a measurement process (Matlab, Python, Passerelle…)

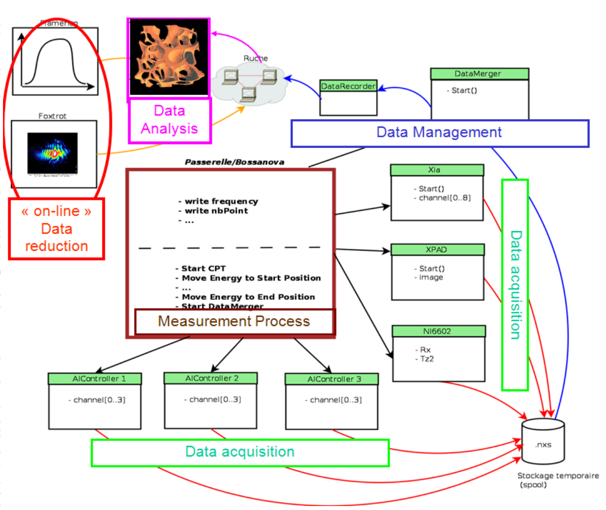

Figure 2: Data flow during the flyscan experiment. Multiple detector processes stream raw files to a temporary storage (spool), which are merged with metadata for permanent storage.

THE ACQUISITION COMPONENTS ON OUR NANOSCOPIUM PROTOTYPE

Acquisition subsystems used on the test bench for Nanoscopium are the following:

- An ADLINK DAQ-2005 analog-to-digital converter (ADC) board receiving signals from Si photodiodes measuring the X-ray beam intensity;
- A National Instrument (NI) PCI-6602 counter acquiring encoder positions of the 2 stages axis;
- A fast 2D photon-counting pixel array, XPAD [3], able to acquire up to 960 lines x 560 columns at 650 frames per second. (the XPAD is controlled with the LImA framework [4])

The master clock is another NI PCI-6602 configured as a trigger generator.

FIRST EXPERIMENTAL RESULTS

The FlyScan prototype has been deployed on the SOLEIL METROLOGIE beamline [5]. Official beamtime of few days has been dedicated to test in real condition this new acquisition system mode in the case of scanning differential phase contrast experiment [6].

In this experiment, a mapping (400 pixels x 40 lines) with a focused beam of a cross-nylon fiber was done in continuous mode. A master clock pulsed at 83 Hz (12 ms period), starts the acquisition of:

- The beam intensity (Fig. 3(a));
- The positions of the sample stages (Fig. 3(b));
- An image of the transmitted defocused beam (Fig. 3(c)) taken with one XPAD module (120 lines x 560 columns).

In order to construct differential phase contrast image, the center of gravity is calculated, on-line or off-line, on each XPAD image, along vertical and horizontal direction (Fig. 3(d) (e)).

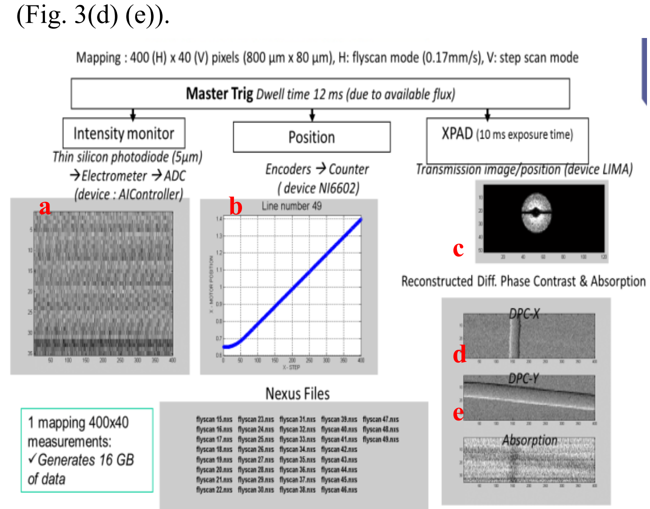

Figure 3: 1st results.

The architecture proved to be modular in terms of number and type of detectors. We have been able during the tests to easily integrate another 2D Detector (a Basler CCD camera).

DATA MANAGEMENT

Data Size and Network Throughput

There are two factors that have to be carefully taken into account: the volume of data and the speed of acquisition. A square mapping of 400 lines by 400 columns, with the same dwell time of 12 ms per point, generates 160 GB in 31 minutes (neglecting overheads incurred due to transferring the data over Ethernet and storing the files on hard disk). We estimate that once network and data storage performances are optimized (*using recent technologies like SSD disks and 10 Gbits Ethernet*), a single beamline such as Nanoscopium, working in flyscan mode with 2ms dwell time per point, could generate several hundred terabytes of data per year. This will present challenges to the data storage infrastructure and data processing capabilities.

Data Access

The software architecture proved to be modular and flexible at the Data Acquisition level. Nevertheless, the internal organisation of data with NeXus files changes depending on the number and types of detectors integrated in the data acquisition system. To hide these data organisation details to data analysis applications, SOLEIL develops with ANSTO the CommonDataModelAccess library [7].

CDMA is made of a core API that access data through a data format plugins mechanism and scientific applications definitions (i.e. sets of logically organized keywords defined by scientists for each experimental technique). Using an innovative “mapping” system between applications definitions and physical data organizations, the CDMA allows to develop data reduction applications regardless data files formats AND organizations.

Online Data Reduction

Because of the very large amount of data, it became evident already during the first experiments that data reduction applications should provide on-line data quality indicators to ensure whether the experiment is running as expected. The java COMETE [8] framework developed at SOLEIL allows us to reduce in “real-time” the large stack of raw images and spectrums to a small subset of reduced data, like for example reducing 160,000 XPAD images to the four images presented in this paper.

CONCLUSION/PERSPECTIVE

The following technical objectives are planned to be fulfilled before the end of 2011:

- stabilize all software components at the acquisition system level *(XPAD detector library , Tango device, Nexus4Tango library)*
- add the fluorescence XIA detector
- enrich existing COMETE data reduction applications
- integrate the CDMA C++ port within MATLAB and python to start integration in scientific data analysis applications

After the realisation, stabilisation, and test of the Nanoscopium prototype, an important project management phase will be to deploy FlyScans measurements on SOLEIL beamlines on a large scale basis, taking into account the impact on our data storage infrastructure and processing capabilities *(cluster and GPU)*.

We believe that this project is an important step for SOLEIL to prepare our computing resources to the scientific data deluge announced since years, but which is now becoming a reality.

REFERENCES

[1] TANGO framework: http://www.tango-controls.org/.
[2] NeXus file format : http://www.nexusformat.org.
[3] K. Medjoubi et al., Performances and applications of the CdTe- and Si-XPAD3 photon counting 2D detector, JINST 6 (2011) C01080.
[4] LiMA : https://forge.epn-campus.eu/projects/show/lima.
[5] M. Idir, P. Mercère, T. Moreno, A. Delmotte, P. Da Silva, and M. H. Modi, Metrology and Tests beamline at SOLEIL Design and first results. *AIP Conf. Proc.* 1234, *485-488* (2010).
[6] A. Menzel, C. M. Kewish, P. Kraft, B. Henrich, K. Jefimovs, J. Vila-Comamala, C. David, M. Dierolf, P. Thibault, F. Pfeiffer and O. Bunk, “Scanning transmission X-ray microscopy with a fast framing pixel detector,” Ultramicroscopy, 110(9), 1143-1147 (2010).
[7] CommonDataModelAccess: ICALEPCS 2011 THCHAUST03.
[8] COMETE: http://comete.sourceforge.net.