NSLS-II: updates from AMX and FMX

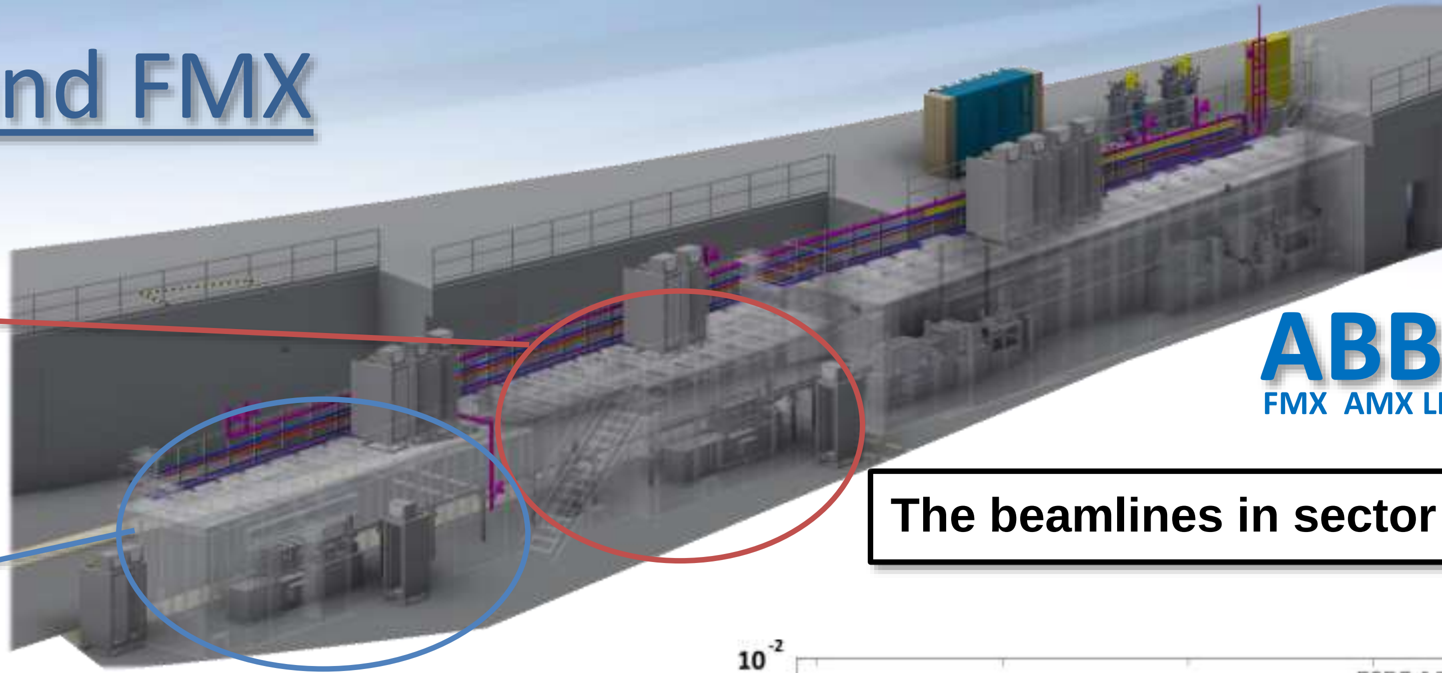
High Data-Rate Macromolecular Crystallography, HDRMX
MAX IV Laboratory, Lund, Sweden, March 15-17 2017

Jean Jakoncic, NSLS-II, BNL, Upton, NY
jjakoncic@bnl.gov

U.S. DEPARTMENT OF ENERGY | Office of Science

BROOKHAVEN NATIONAL LABORATORY | National Synchrotron Light Source II

AMX and FMX



ABBIX
FMX AMX LIX

The beamlines in sector 17-ID

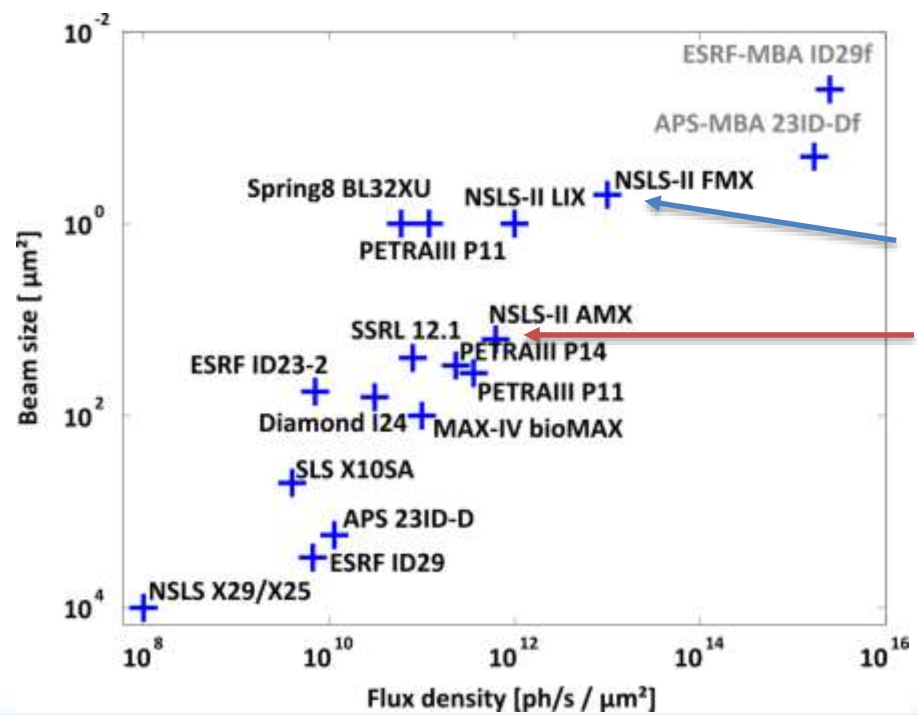
2 canted IVU21 undulators -> 2 independent beamlines

AMX : highly Automated MX
vast number of samples
large assemblies

5-100 µm beam
~8 10^{12} ph.s^{-1} (1Å)
5-18 keV (0.7- 2.5 Å)
EIGER 9M (≤238 Hz)

FMX : Frontier MX
most challenging problems
requiring a µ-focus beam
and / or high X-ray energy

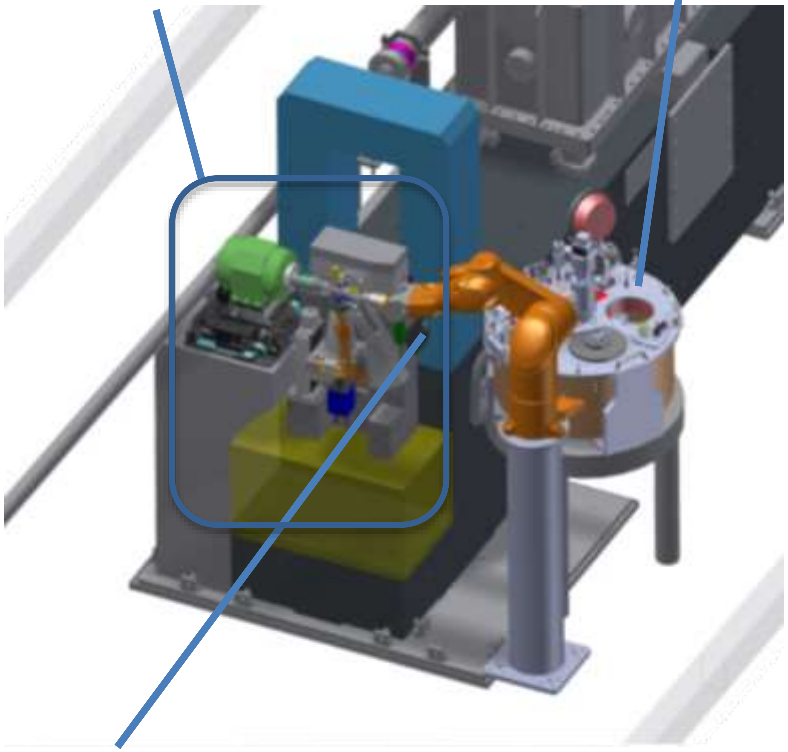
1-20 µm beam
~6 10^{12} ph.s^{-1} (1Å)
5-30 keV (0.4 - 2.5 Å)
EIGER 16M (≤133 Hz)



Automated Sample Mounting (design & implementation)

High Capacity Automated Sample Changer with 24 unipucks capacity (384 bases).

In-House end-station

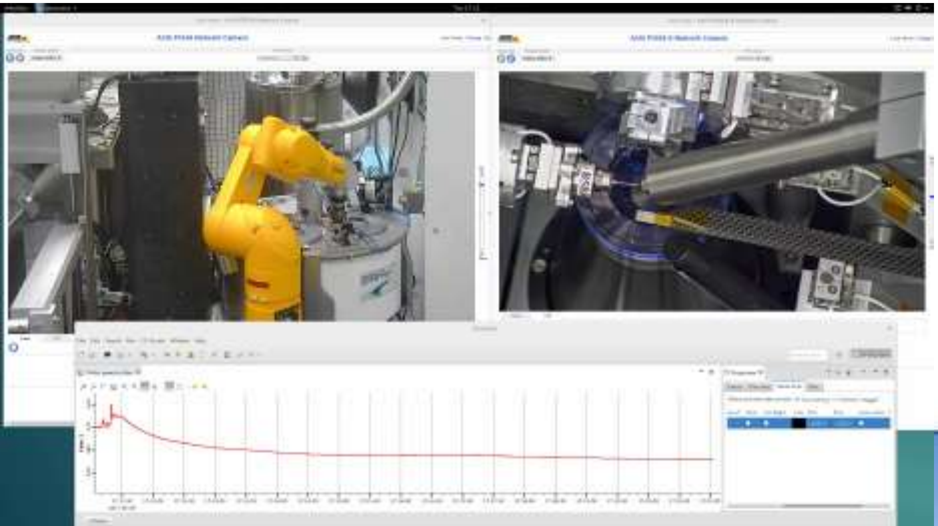


6-axis Staubli TX60/L robotic arm with NSLS-II gripper.

AMX / FMX: Current status (275 mA)

AMX : from 5 to 18 keV
Flux ~ 4 10^{12} ph.s^{-1}
Beam ~ 8 x 7 um^2
Manual sample mounting
Hardware trigger (raster / vector)
Typical data collection:
0.1 deg @ 100 Hz
Data collection : from <1s to 15 secs
Default screening:
10 % transmission & 10 ms/frame
Av data set: 1400 frames
500 frames / data.h5
Compression range: 4.2-25
Aver compression:10

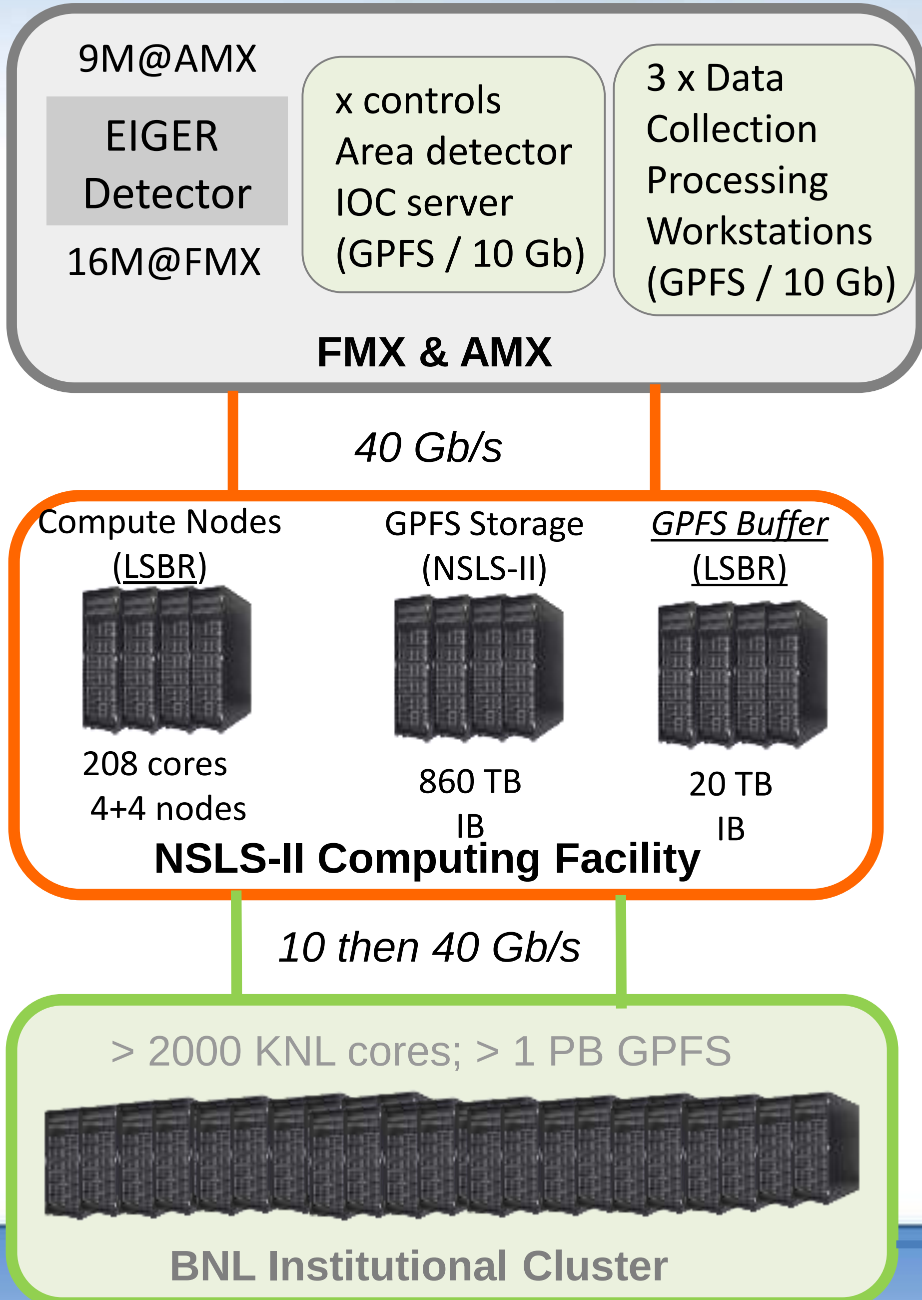
FMX : from 5 to 25 keV
Flux > 2 10^{12} ph.s^{-1}
Beam ~ 10 x 6 um^2
Automated sample mounting
Software trigger (raster / vector)
Typical data collection:
0.1 deg @ 50 Hz
Data collection : from <1s to 15 secs
Default screening:
<10 % transmission & 50 ms/frame
Av data set: 1800 frames
200 frames / data.h5
Compression range: 4.5-32
Aver compression:11.6



40 % of available beamtime for user operation.
To be increased to 80 % gradually.
Operate 2 shifts / day, at the moment.

FMX : up to 30 samples / hour (robot)
AMX: up to 15 samples / hour (manual)

Computing / Network Infrastructures



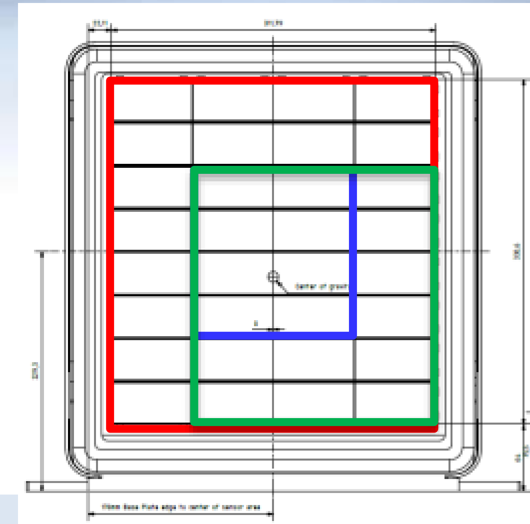
Data collection using LSDC.
Data visualization using Albula or ADSC.
Manual data processing using HKL2000 or other GUIs (ccp4i, phenix, imosflm …),index/Best.
Data transfer onto USB drives.
Direct access to compute nodes over terminal.

Rapid Data Analysis and Reduction on dedicated compute nodes and fast buffer
Pipelines and packages in use: dials spot finders, dozor, XDS, dials, xia2, fast_dp, fast_mr, dimple, fast_ep, fast_ep_weak, fast_ep_NSLS-II, KAMO, and more.
Network backup using dedicated node.

Post processing, Re-processing, Remote processing, Multiple Crystals data reduction and Hierarchical Cluster Analysis.

To world (1 Gb/s (internet) to 100 Gb/s (EsNet))

EIGER 9M @ AMX and 16M @ FMX



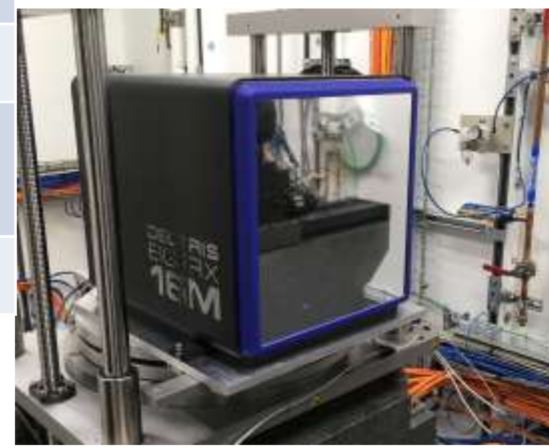
Eiger9M	Eiger16M	Eiger 4M ROI
18 modules	32 modules	8 modules
10 M pixels	18 M pixels	4.5 M pixels
238 Hz	**133 Hz**	**750 Hz**

233x245	311x328	mm^2
238 (750)	133 (750)	Hz
10.2	18.1	M pixel
< 5	< 8	M byte (actual)
200-1700	*200-1700*	MB/s

2-30 TB / day / beamline

NetApp EF560 All-Flash Array

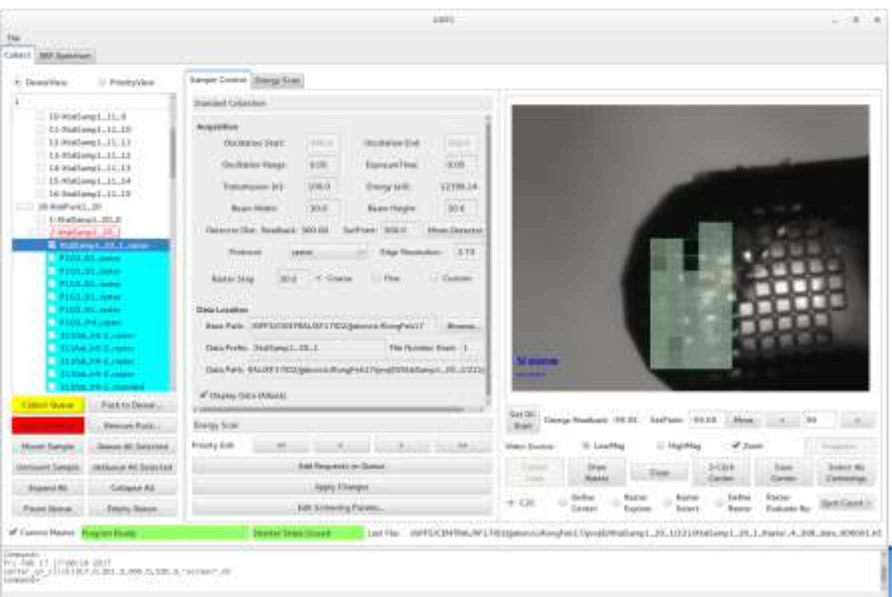
20 x 1.6 TB all flash
6 GB/s (50/50)
2 IB controllers 2 ports each

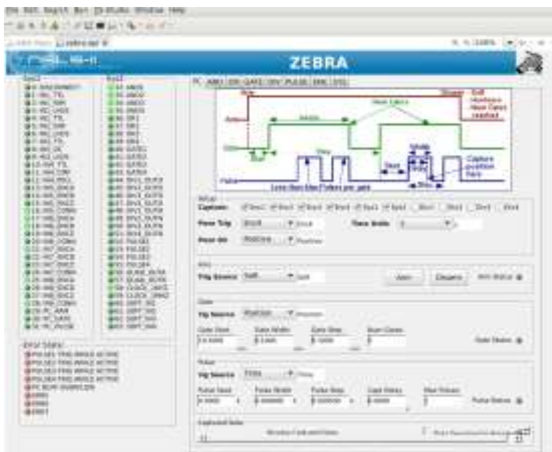


Challenges : data storage, transfer, processing and backup.
Requires advanced software and computing cluster.

Rastering

Free-form selected by user
Automated loop detection (X-rec & InHouse) followed by coarse rastering (face ON), then fine rastering and completed by line scan (at 90 degrees).
For now: each row / or column is a collected as a vector (individual data set).
Files are converted to cbf and processed with dials spot finder client / server.
 >> arms for each row or column (~3.5 secs on AMX ~ 10 secs on FMX) : 1 arm for all (in the work)
 >> HW trigger: on omega position and rely on clock on Zebra Box (Quantum Detector).
Will trigger at each omega position (or pin xy positions for stills raster)
<u>Processing at ~ 15 FPS on one node (RAM or GPFS)</u>
Mixed results when ice rings are present.
 >> tuning dials (Nick/Graeme slides), using dozor (Gleb slides).
>>>> optimizing codes for faster results (ROI, HDF5 read, KNL, OMP ……)





fine

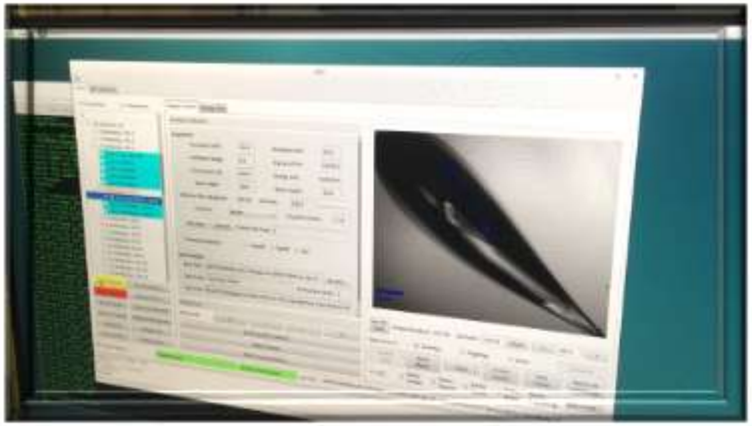
Data Reduction

Fast_dp: on 1 dedicated node (more to be used soon) @ 2-10 FPS.

data set of 900 frames (lyso test). (all data on GPFS)

9M	Total time (s)	FPS
FDP-005	90 (56 RAM)	10 (16)
FDP-002	86 (79 RAM)	10 (11)
FDP-010	83 (50 RAM)	11 (18)
XDS-005	51 (44 RAM)	18 (20)
XDS-002	61 (55 RAM)	15 (16)
XDS-010	72 (43 RAM)	12 (21)
FDP-004-005	TBRD	TBRD
FDP-002-004-005	TBRD	TBRD

XDS: jobs = all, 72 procs, 16 jobs)



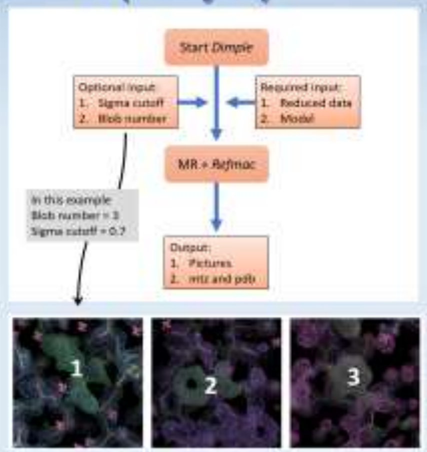
CPUS 004, 005: 36 cores@2.3 GHz (10 Gb)

CPUs 009, 010: 44 cores@2.2 GHz (40 Gb)

CPUs 002, 003, 007, 008: 12 cores@3.4 GHz (10 Gb)

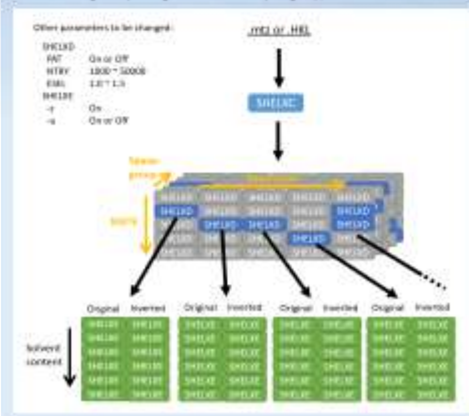
Pipelines

DIMPLE (modified)

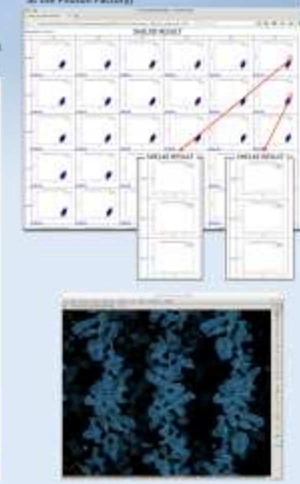


FAST_EP_WEAK

Fast_ep_weak is a modified version of Fast_ep, which is originally developed at Diamond Light Source. Fast_ep_weak is designed to find a structure solution from weak anomalous signals by finer grid searches with spacegroup, nsite and resolution cut off.



P-SAD on a DNA crystal (Data sets were collected at BL1A at the Photon Factory)



FAST_EP_NSLS2 (NSLS-II ADPSSP)

Fast_ep_nsls2 is a single crystal SAD phasing pipeline from a diffraction data set to a model with fast_dp, fast_ep, fast_ep_weak and other crystallographic software.

Simple report of a processing result

Thermolysin Zn-SAD at AMX at 11 keV

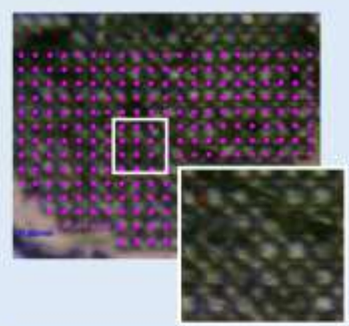
fast_dp (data reduction)	< 100 s	fast_ep	< 200 s
Space group	P6122	[illegible]	[illegible]
Unit cell parameters	[illegible]	[illegible]	[illegible]
Mosaicity spread (°)	[illegible]	Best space group	[illegible]
Resolution (Å)	[illegible]	[illegible]	[illegible]
R_{merge}	[illegible]	CC (%)	[illegible]
I/sigma	21.20 [illegible]	Chains	[illegible]
Completeness (%)	[illegible]	Residues	[illegible]
Multiplicity	[illegible]	arpwarp (1 building cycles)	[illegible]
CC1/2	[illegible]	Residues / water	[illegible]
Anom. Completeness	[illegible]	R_{work} / R_{free}	[illegible]
Resolution (Å)	[illegible]	arpwarp (3 building cycles)	[illegible]
R_{merge}	[illegible]	Residues / water	[illegible]
I/sigma	21.20 [illegible]	R_{work} / R_{free}	[illegible]

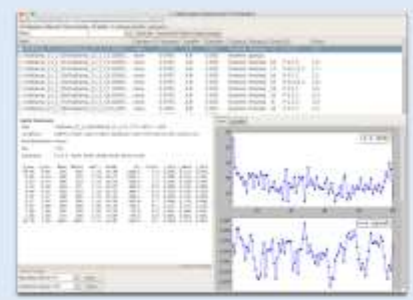
All processing performed on a single node : total time < 15 mins

A. Soares, Y. Yamada.
G.W., H.B.

KAMO

Multiple small wedge data collection at AMX

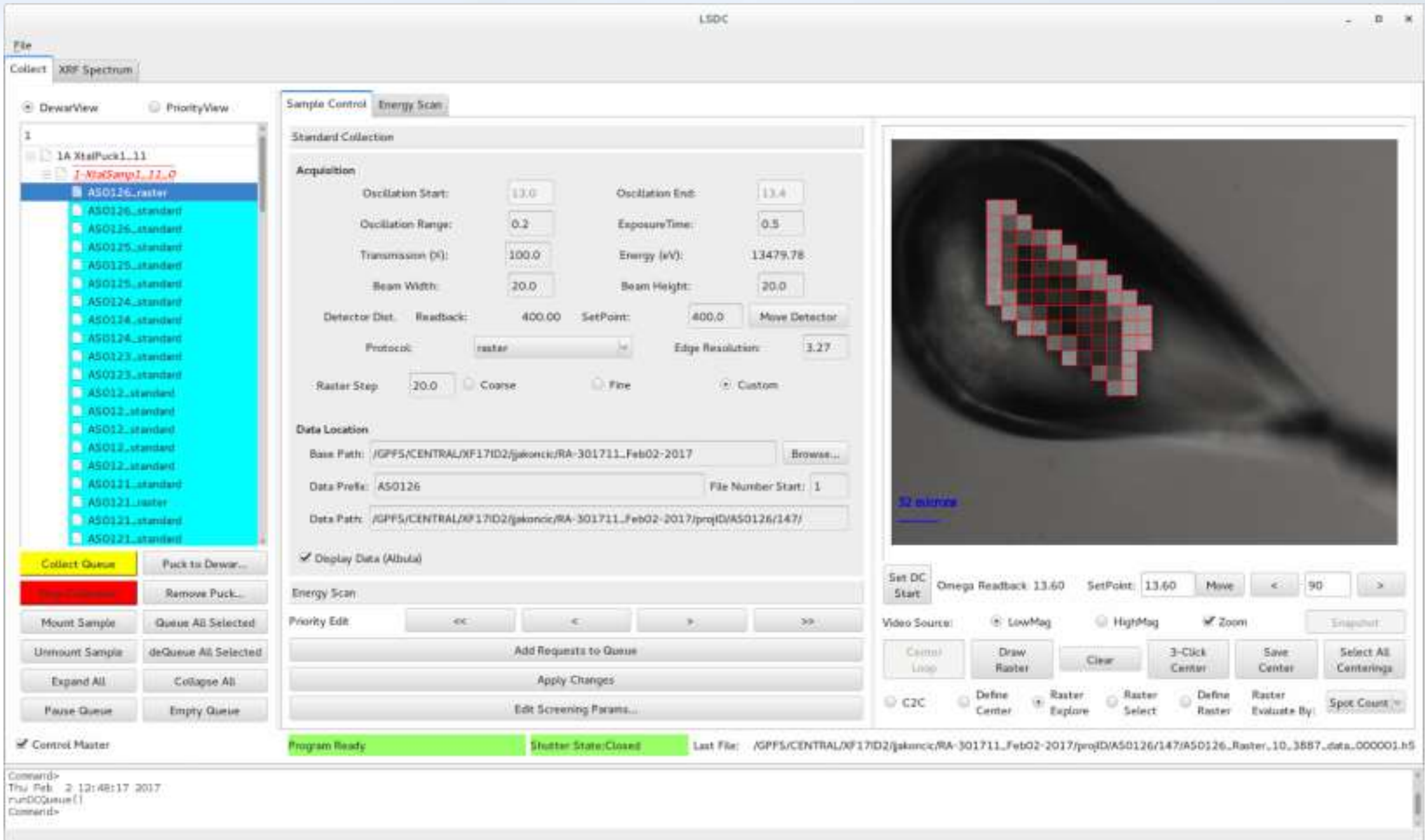




Energy	8
Flux	5.3 x 10^{11} ph/sec (85 % attenuation)
Total oscillation	5 deg./spot
Osc. width	0.2 deg./frame
Exp. time	0.04 sec./frame (P6M)

(multi.collect on LSDC)
Raster vs raster + threshold

LSDC User Interface (under continuous development)



- Layout derived from MxCuBE
- Automatic loop centering, shape recognition (XREC) + In-House code.
- User-defined raster areas & helical scanning data collection (NSLS-II: bluesky; CSS)

MxCuBE: J. Gabadinho et al. *J. Synchrotron Rad.* (2010) **17**, 700-707 XREC: S. B. Pothineni et al. *Acta Cryst. D,* (2006) **62**, 1358-1368

Conclusions & Acknowledgements

GPFS works well with 2 writes but not so well with x concurrent reads.
Currently implementing the SSD buffer to be tested in the upcoming weeks.
Will update nodes network as needed (40 or IB)

With beams 10 microns or less, we see more head scratching.

Data collection from multiple micro-crystals more and more used.

Not all users are prepared to handle the amount of data and reprocessing data : database a must and general cluster with remote access would help.

Learning as we go and implement new features (automation, remote, Multiple crystals data collections, in-situ ...

Good timing for next meeting > 6 months from now !

Using KNLs: spot finder, dozor, xds ? We are testing on Intel cluster: ACA workshop with Intel

ABBIX, LSBR, NSLS-II controls staff.
J. Skinner, B. Martins, R. Petkus, M. Fuchs, A. Soares, E. Lazo.
N. Sauter, G. Winter, S. Popov, G. Bourenkov.
W. Shi, H. Bernstein, Y. Yamada[#].