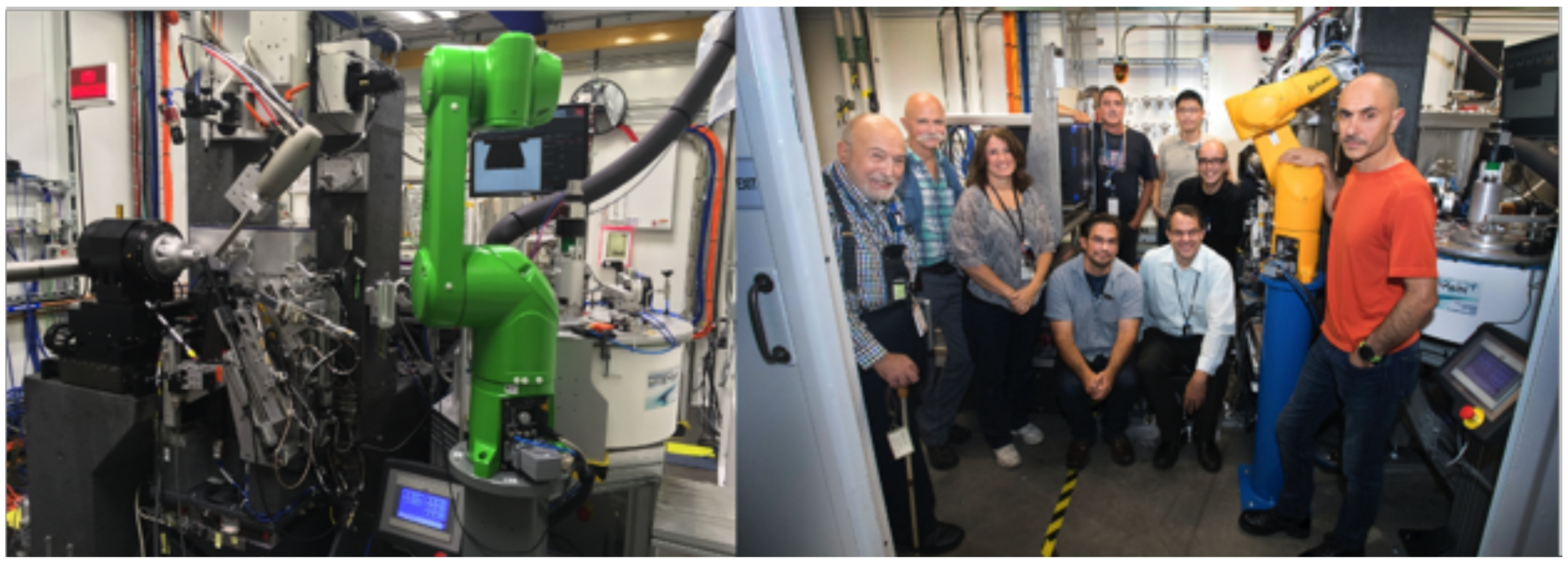
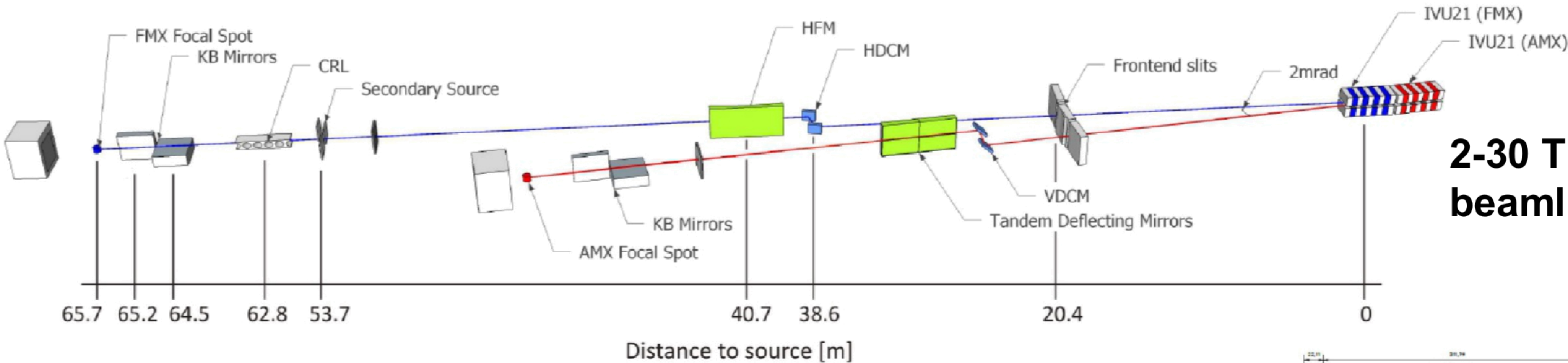
High Data Rate MX (HDRMX) at NSLS-II

Dale Kreitler, NSLS-II, Brookhaven National Laboratory
IUCr XXV HDRMX Workshop 8/14/2021

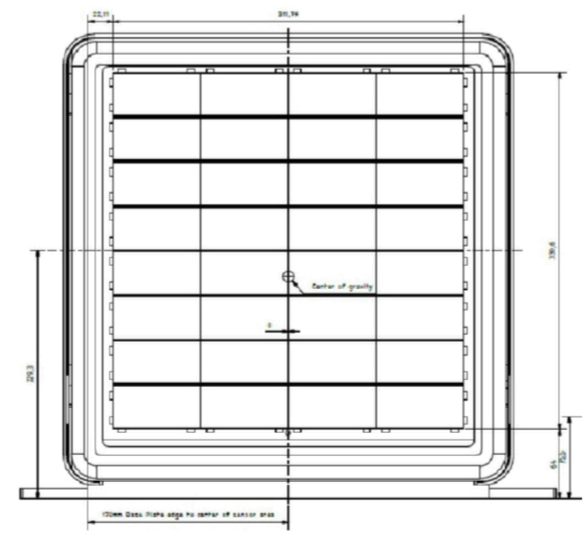


Beamline Characteristics



2-30 TB / day / beamline

Specifications	AMX	FMX
Wavelength range	0.7 – 2.5 Å	0.4 – 2.5 Å
Flux at focus at 12.7 keV	$>4\times10^{12}$ ph/s	3.5×10^{12} ph/s
Focal spot min (H×V)	7×5 μm^2	1.5×1 μm^2
Detector	Eiger 9M (200Hz)	Eiger 16M (100Hz)

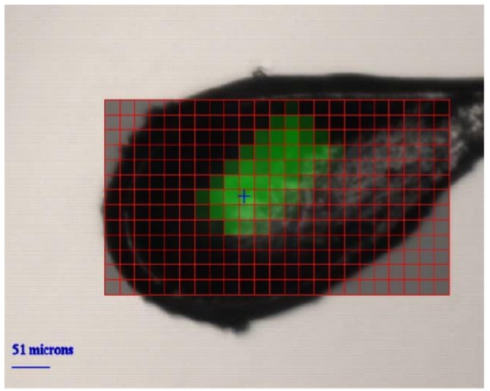


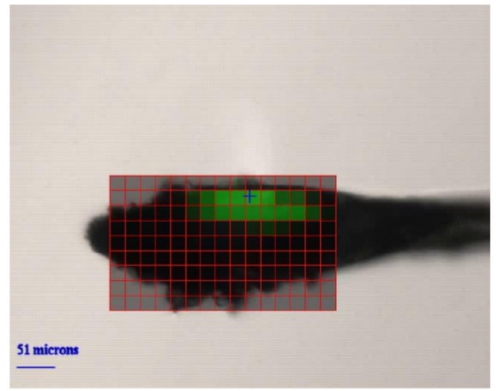
Challenges : data storage, transfer, processing and backup.
Requires advanced software and computing cluster.

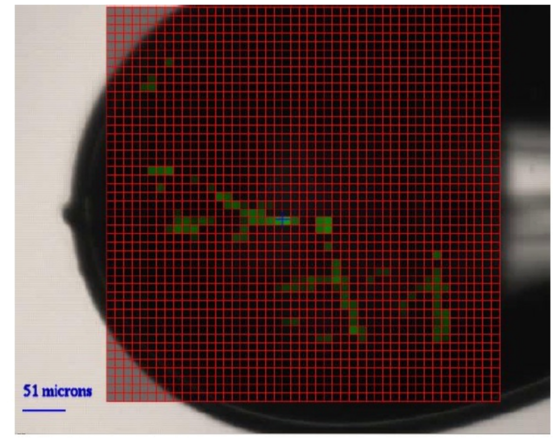
Fuchs, M. R. et al. "NSLS-II biomedical beamlines for micro-crystallography, FMX, and for highly automated crystallography, AMX: New opportunities for advanced data collection *AIP Conf. Proc. SRI2015*, **2016**, *1741*, 030006

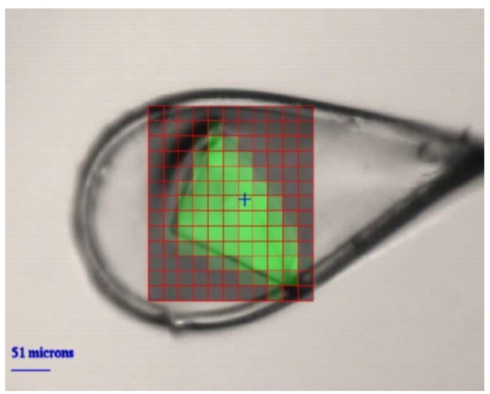
Raster-based centering

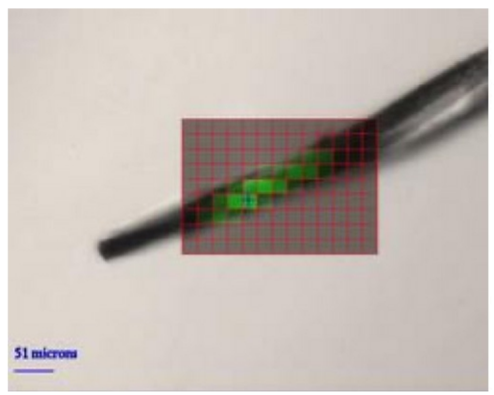
Raster every sample

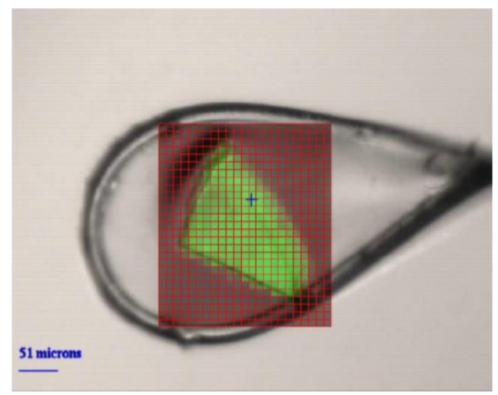


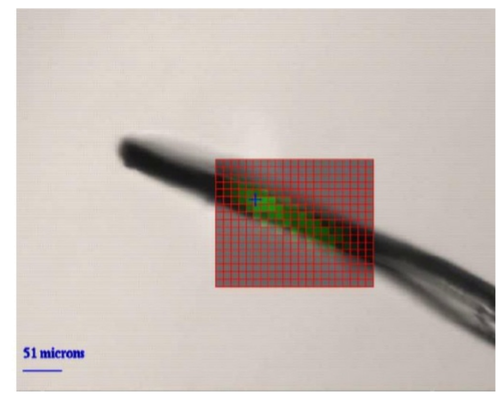




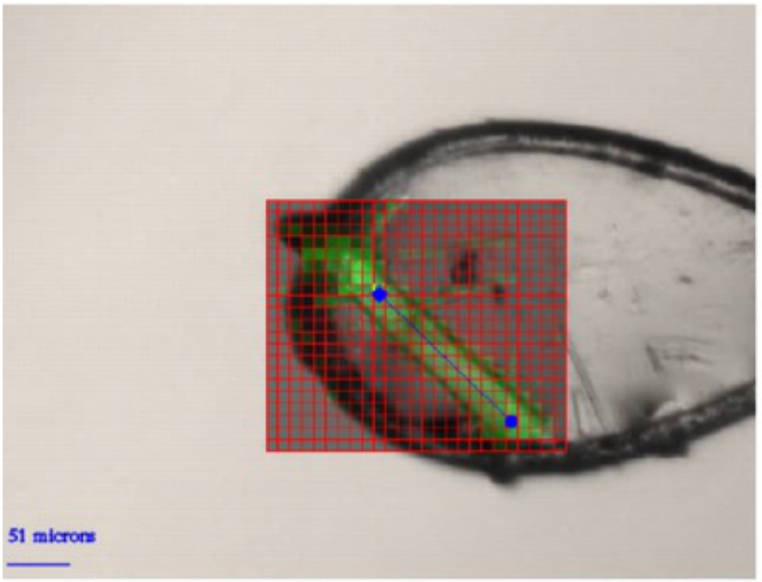


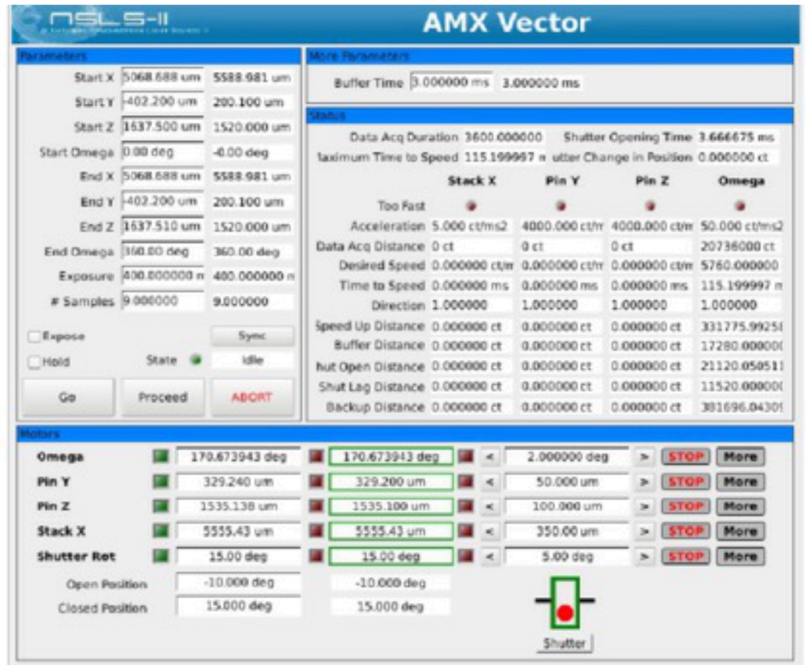


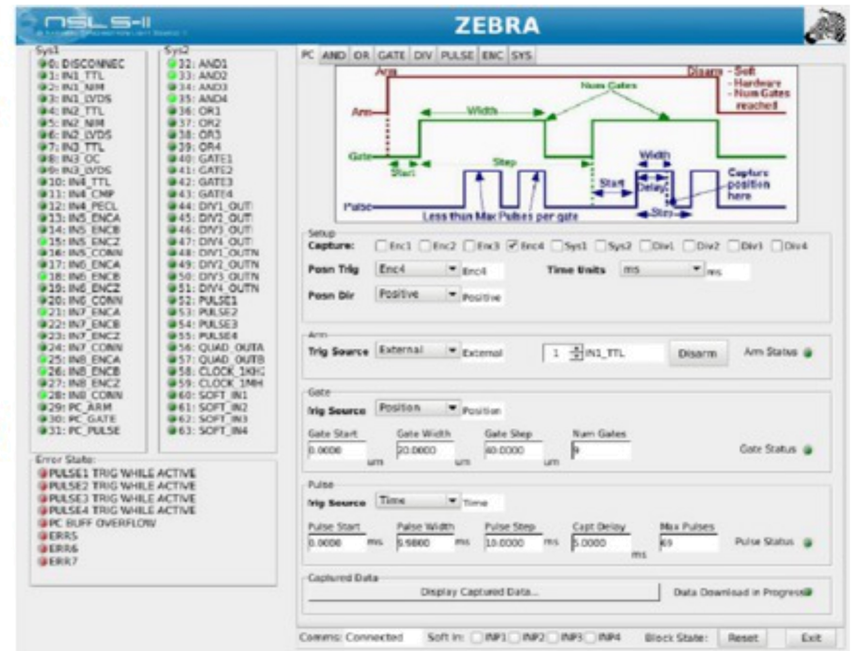




Vector (helical) data collections





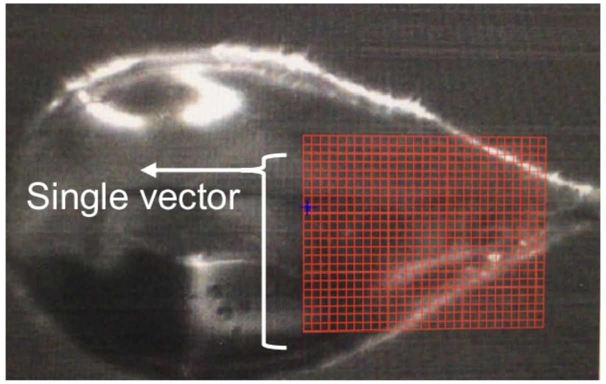


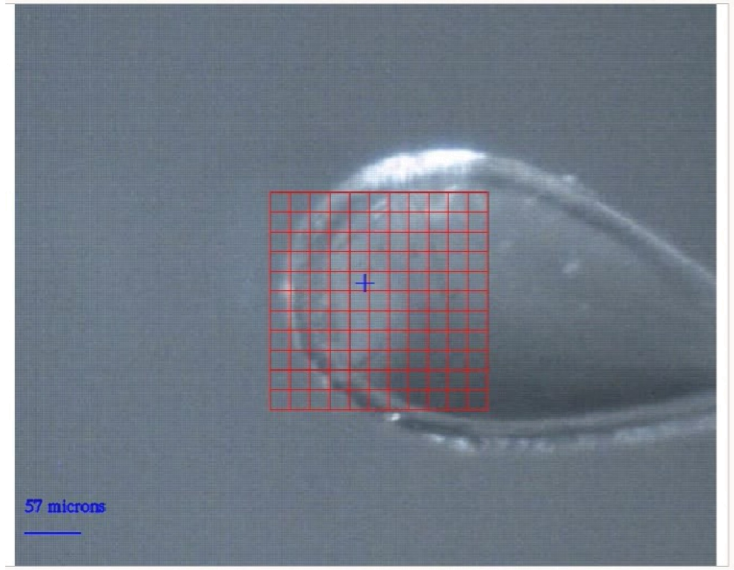
Vector-based rastering and data collection

- Every raster is a 'vector'
- Raster center everything
- Vector collection by row
  - data.h5 file written for each row
  - Only rectangular rasters (single detector arm)
  - Python LSDC server generates dozor input file based on metadata from database request object
  - Dozor processing triggered on remote node (per row, immediately after data acquired)
  - dectris-neggia
  - Also can reprocess with dials.find_spots
  - FileWriter interface

Dozor: G. Bourenkov, A. Popov

LSDC: J. Skinner, J. Aishima

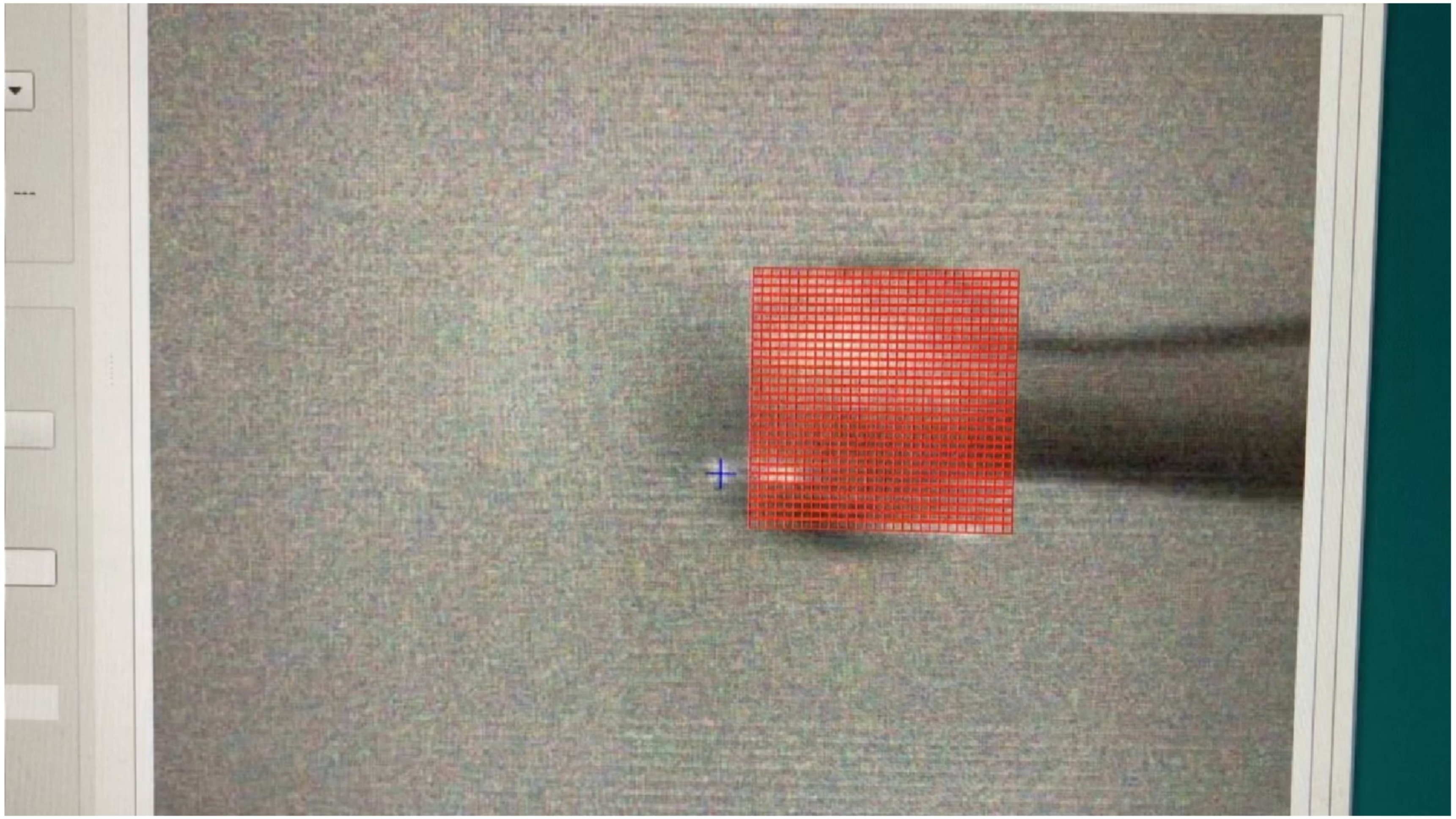




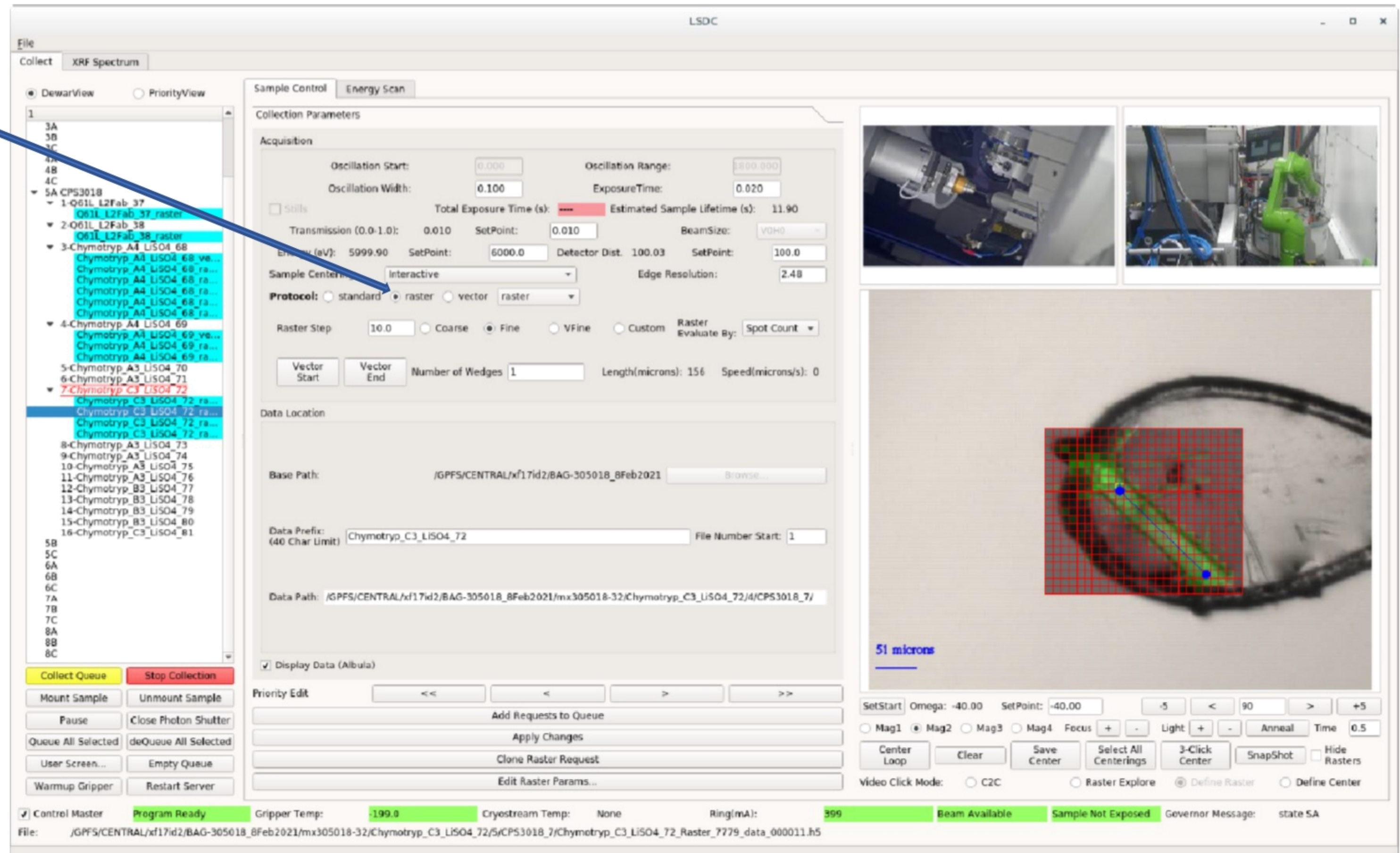

```
PIN_ALIGN: DEBUG enabled
PIN_ALIGN_ROI_WIDTH: 150
PIN_ALIGN_ROI_HEIGHT: 400
PIN_ALIGN_ROI_WIDTH_OFFSET: 410
PIN_ALIGN: Y motor axis is up
PIN_ALIGN: Z motor axis is down
argument 0: /GPFS/CENTRAL/xf17id1/skinnerProjectsBackup/pinAlig
ign_fmx.sh
argument 1: MKP1268_9_4_8_0_PA_0_001.jpg
argument 2: MKP1268_9_4_8_0_PA_90_001.jpg
Task MOUNT isRunning =  false
Running Task PARK true
PIN_ALIGN: DEBUG enabled
PIN_ALIGN_ROI_WIDTH: 150
PIN_ALIGN_ROI_HEIGHT: 400
PIN_ALIGN_ROI_WIDTH_OFFSET: 410
PIN_ALIGN: Y motor axis is up
PIN_ALIGN: Z motor axis is down
PIN_ALIGN: DEBUG enabled
PIN_ALIGN_ROI_WIDTH: 150
PIN_ALIGN_ROI_HEIGHT: 400
PIN_ALIGN_ROI_WIDTH_OFFSET: 410
PIN_ALIGN: Y motor axis is up
PIN_ALIGN: Z motor axis is down
Task PARK isRunning =  false
Running Task UNLATCHROBGOV true
Task UNLATCHROBGOV isRunning =  false
```

50 Hz; usage typically 50-200 Hz

500 cell raster, ~2.5-10 sec total exposure

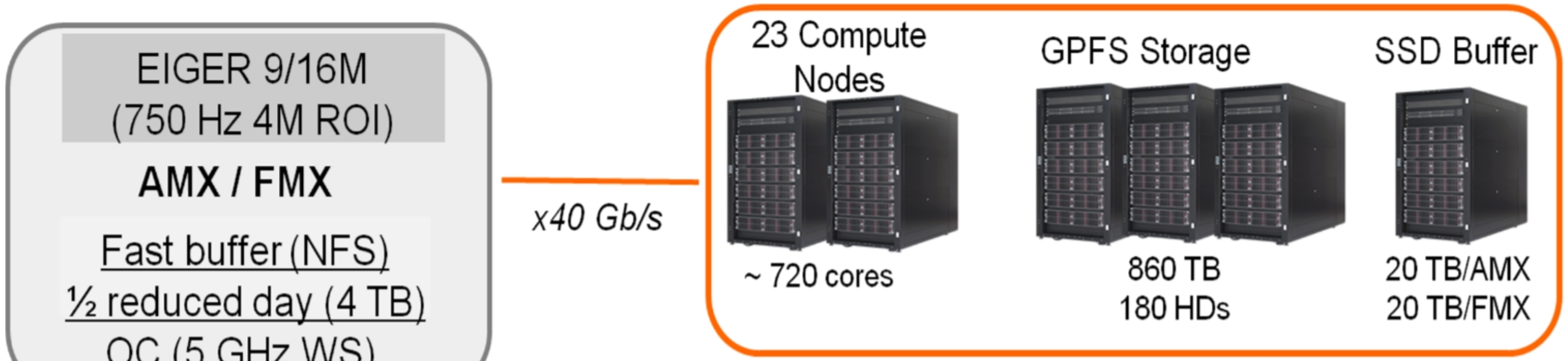


Simple raster queueing



Click 4 corners

AMX/FMX Infrastructure

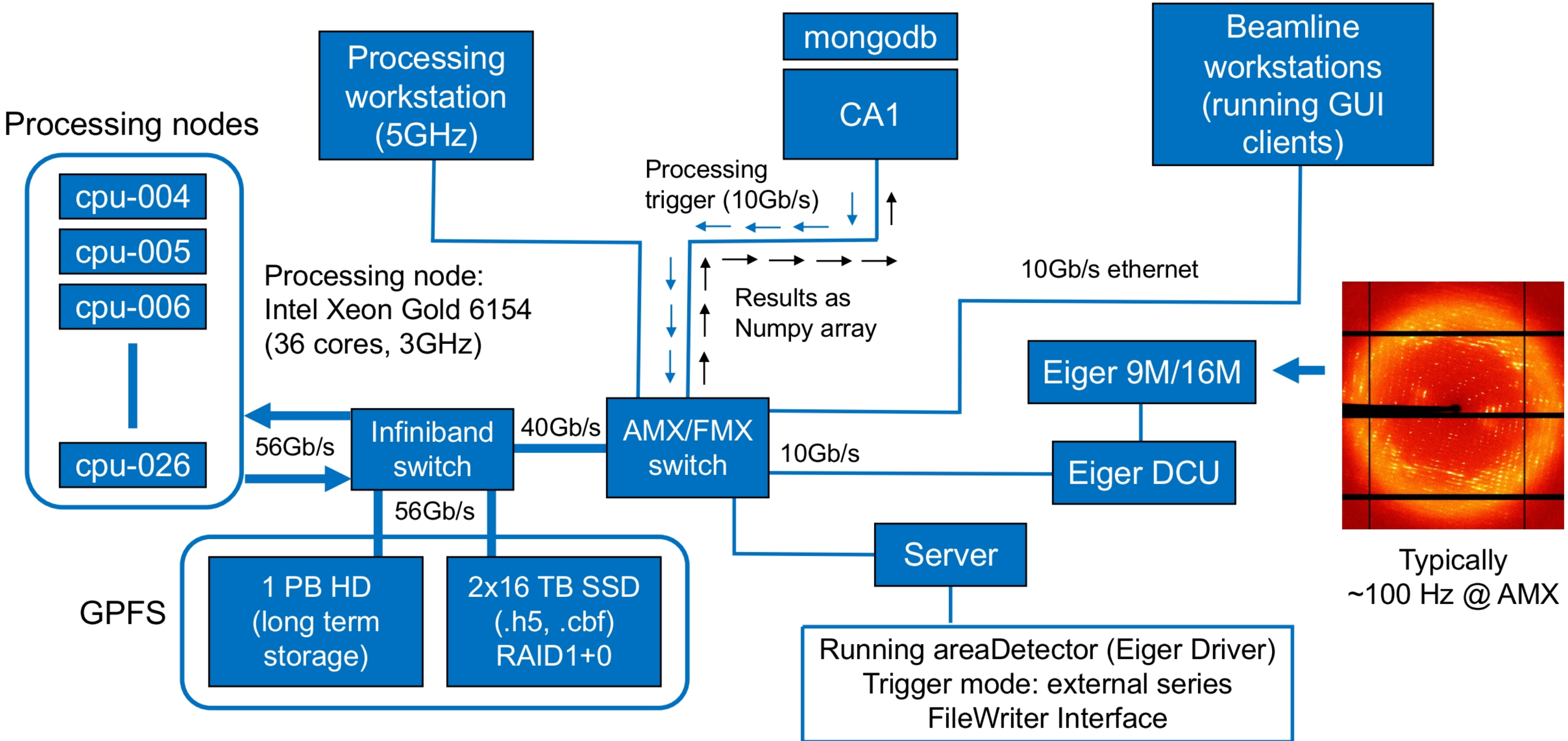


NSLS-II Computing Facility (august 2018)
Scalable Storage and Processing Nodes: sustaining growth

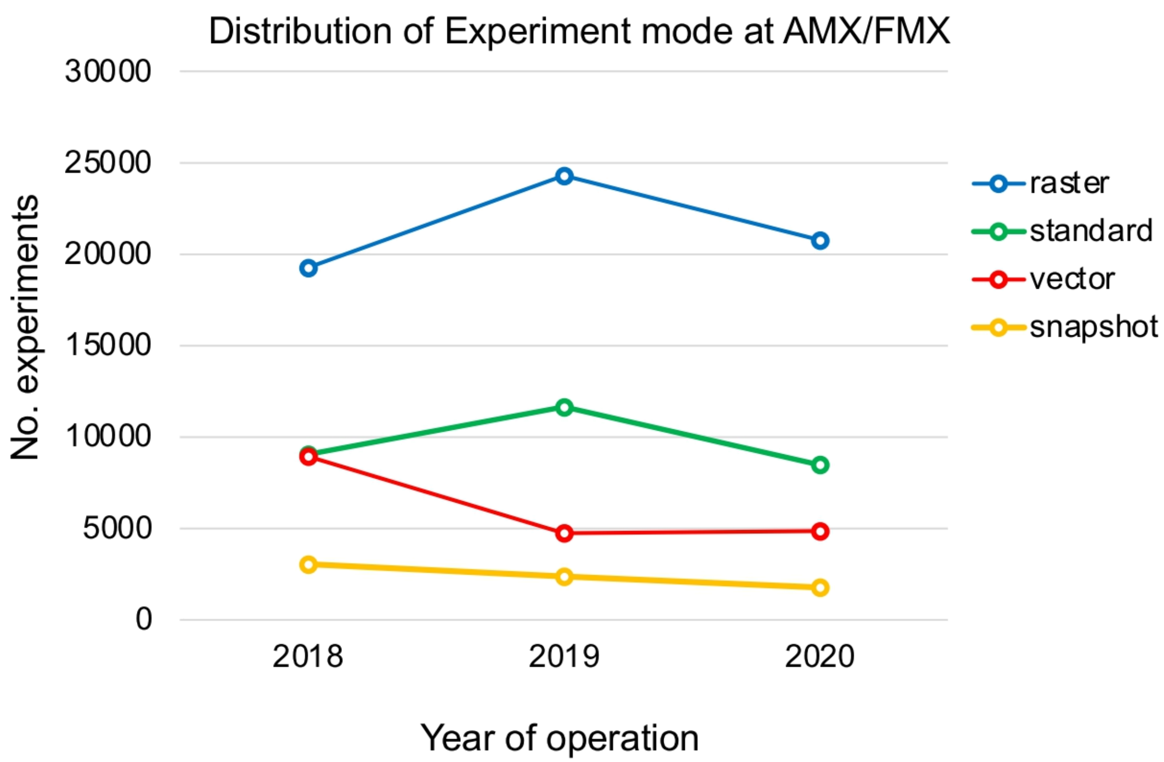
- @ beamlines (2x): 3 WS, 1 data backup WS, 1 OC WS (single processes) and CAs SRVs … (all on 10 Gbs and on GPFS)
- @ NSLS-II central facility: 23 nodes: ~ 720 cores (56 Gb/s IB to GPFS; 10 Gbs/ node to node)
- @ NSLS-II central facility: 2 x 16 TB SSD fast buffer (NSDs) GPFS

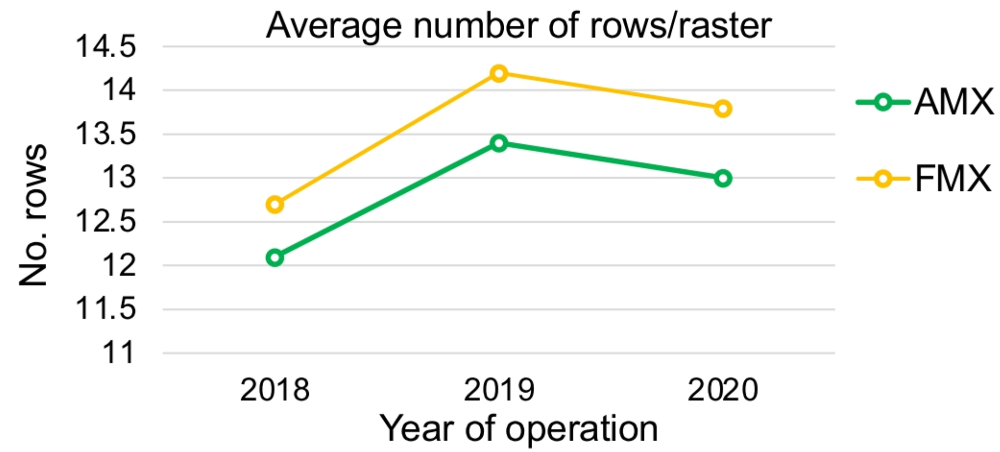
Data transfer to users with globus (https://www.globus.org), no hard drives

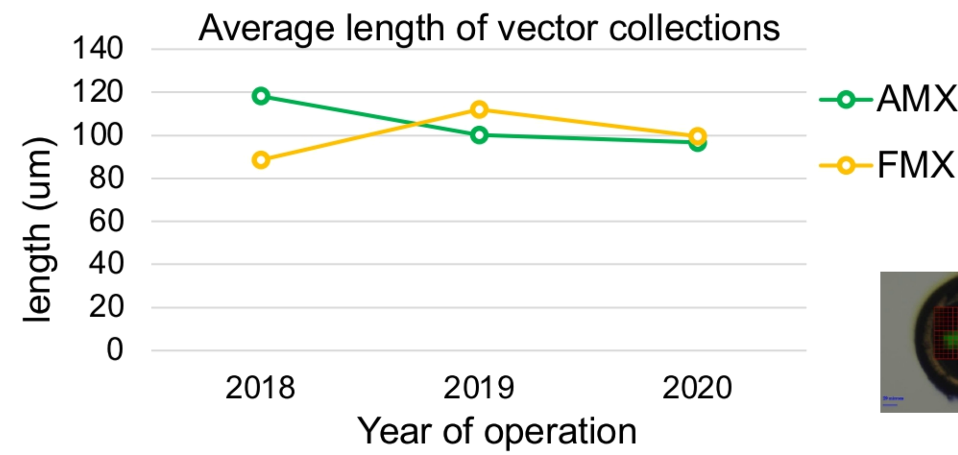
Network architecture for raster processing

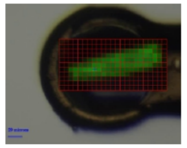


Database statistics of AMX/FMX collections





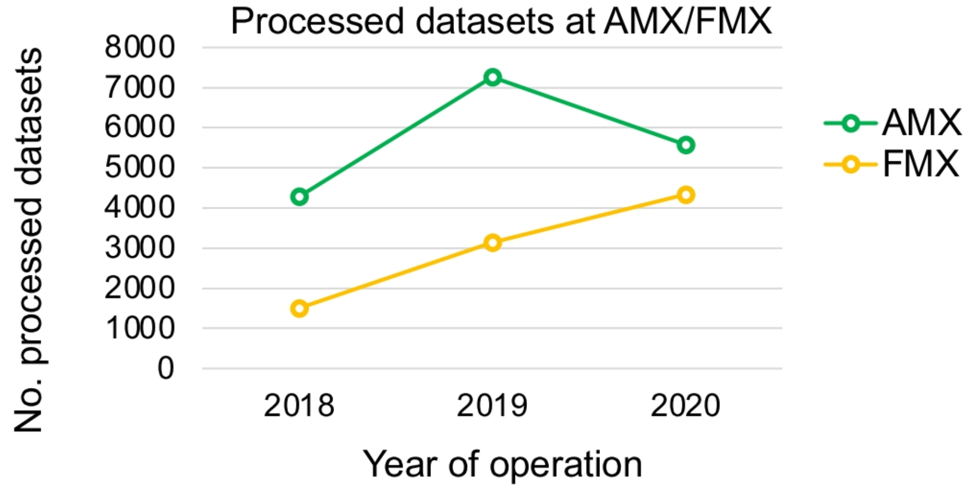


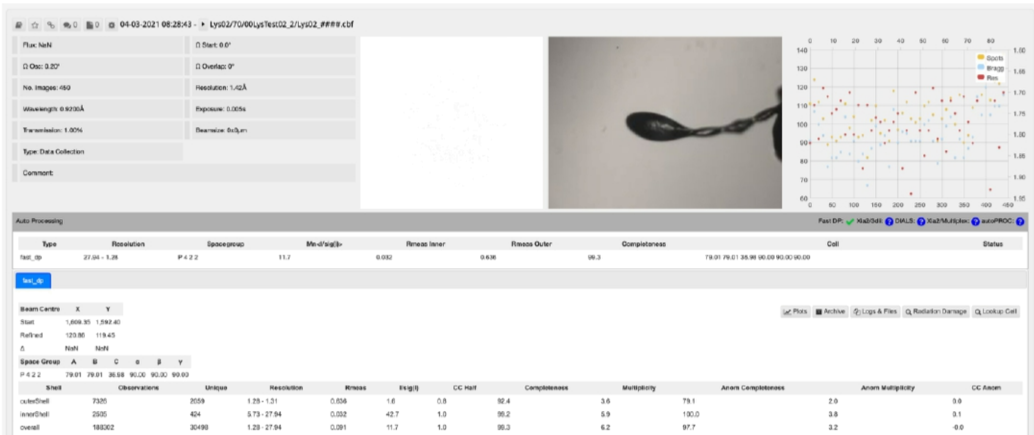


- ~270k vector collections in 2020
- ~2 rasters/dataset
- 10 sec raster time improvement → ~55 h of annual beamtime

Pipelines: In-line processing with fast_dp, dimple

- Autoprocessing of all standard, vector datasets triggered after collection (XDS/fast_dp)
- Typical dataset ~10 seconds (180 deg, 0.2 deg, 100Hz)
- In automated mode, ~1 unipuck/hour
- Well diffracting samples (ligand binding or dynamics) — ~4 min/dataset including sample exchange
- 2 dedicated 36 core nodes for spotfinding, indexing, integration
- pointless/aimless run on overclocked workstation

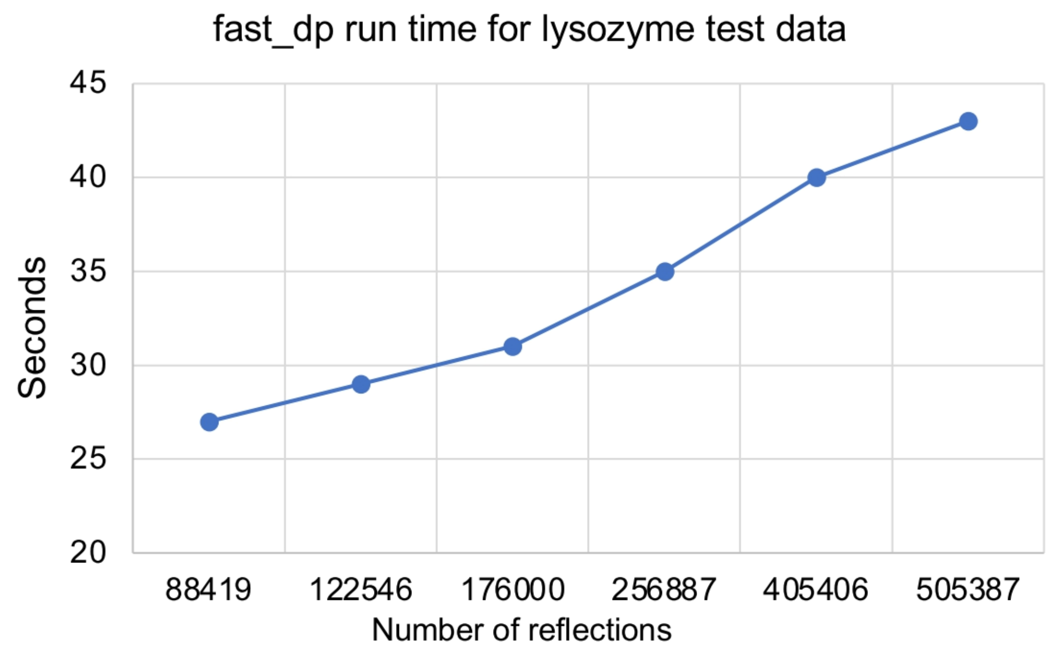




ISPyB

Data processing timing with fast_dp

- From lysozyme start up data
- 0.2 deg, 150 deg (750 frames)
- Eiger 9M
- 1 Angstrom resolution
- 8 nodes (XDS):
  - Intel Xeon Gold 6154
  - 36 cores, 3 GHz
- AMX workstation (pointless/aimless):
  - Intel Core i7-8700K
  - 6 cores
  - overclocked @ 5 GHz
- Dimple (tetragonal lysozyme; 12-36 seconds)
- If achieving 1000 samples/day need, processing must keep pace (1 dataset/min)



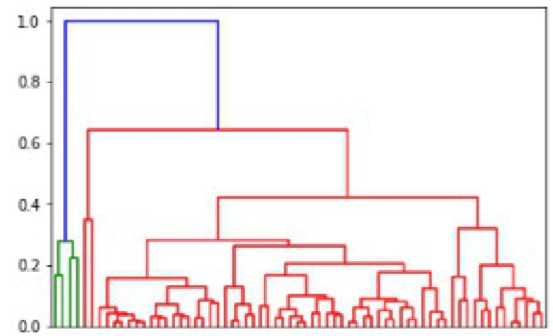
To do list

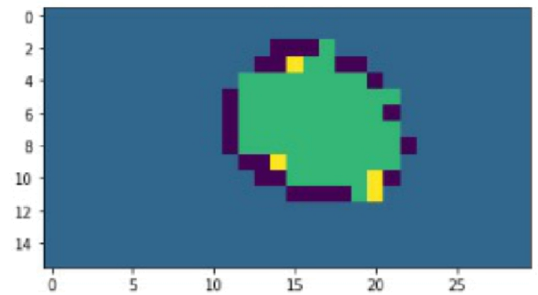
In-line processing:

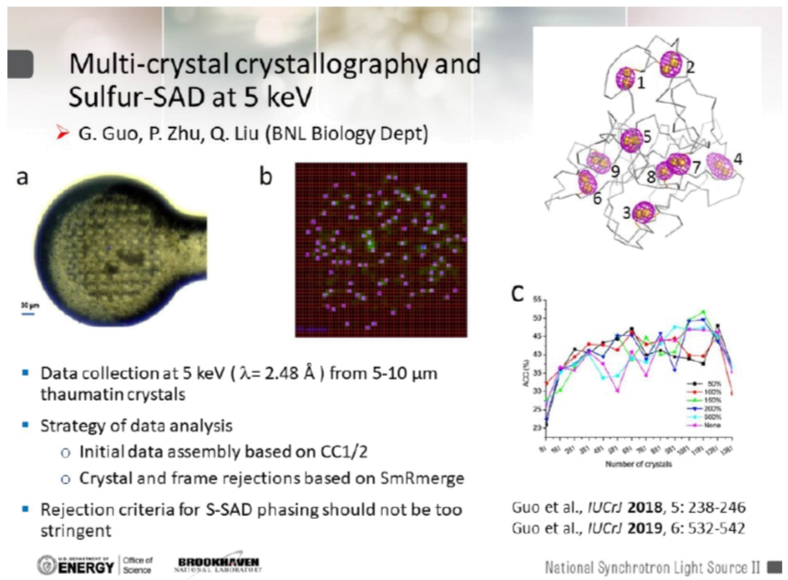
- indexing/clustering of heat maps
- Detection/scoring of multiple lattices for improved heat maps

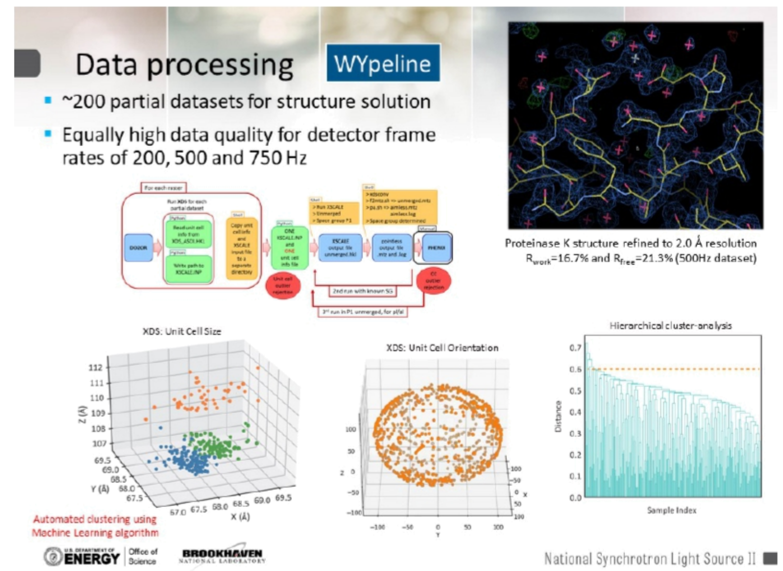
Off-line processing:

- Additional pipelines (autoPROC, DIALS, KAMO)
- Model building (phenix, Rosetta)





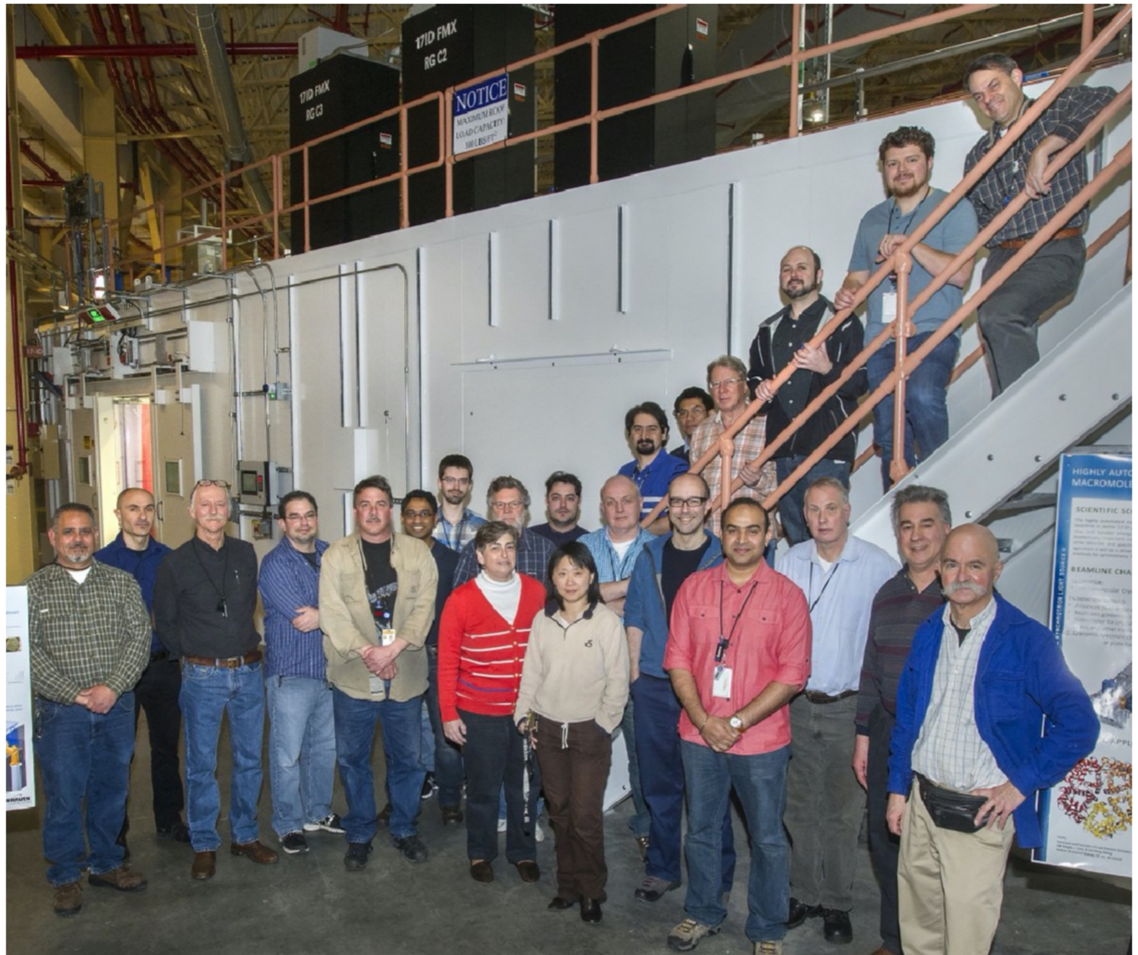




Considerations for the future

- How to keep pace with throughput without compromising quality?
- Wealth of new information, but avoid analysis paralysis
- Automation vs. manual collection
- New types of experiments enabled by HDRMX: clustering, dynamics, time-resolved, etc.
- How do we use these new tools?
- Precision of sample stages/goniometers, stable optics required by faster detectors

Acknowledgments



The ABBIX and LSBR teams, March 2016

- John Lara
- Jean Jakoncic
- Dieter Schneider
- Stu Myers
- Tom Langdon
- *Shirish Chodankar*
- Bruno Seiva Martins
- Vivian Stojanoff
- John Skinner
- *Hugo Slepicka*
- Kun Qian
- Sean McSweeney
- Martin Fuchs
- Babak Andi
- Dale Kreitler

This work is supported by the US National Institutes of Health and the US Department of Energy

Software:

Dozor -- Gleb Bourenkov, Alexander Popov

fast_dp

developers/maintainers Graeme Winter, Markus Gerstel, Richard Gildea, Nicholas Devenish, and others