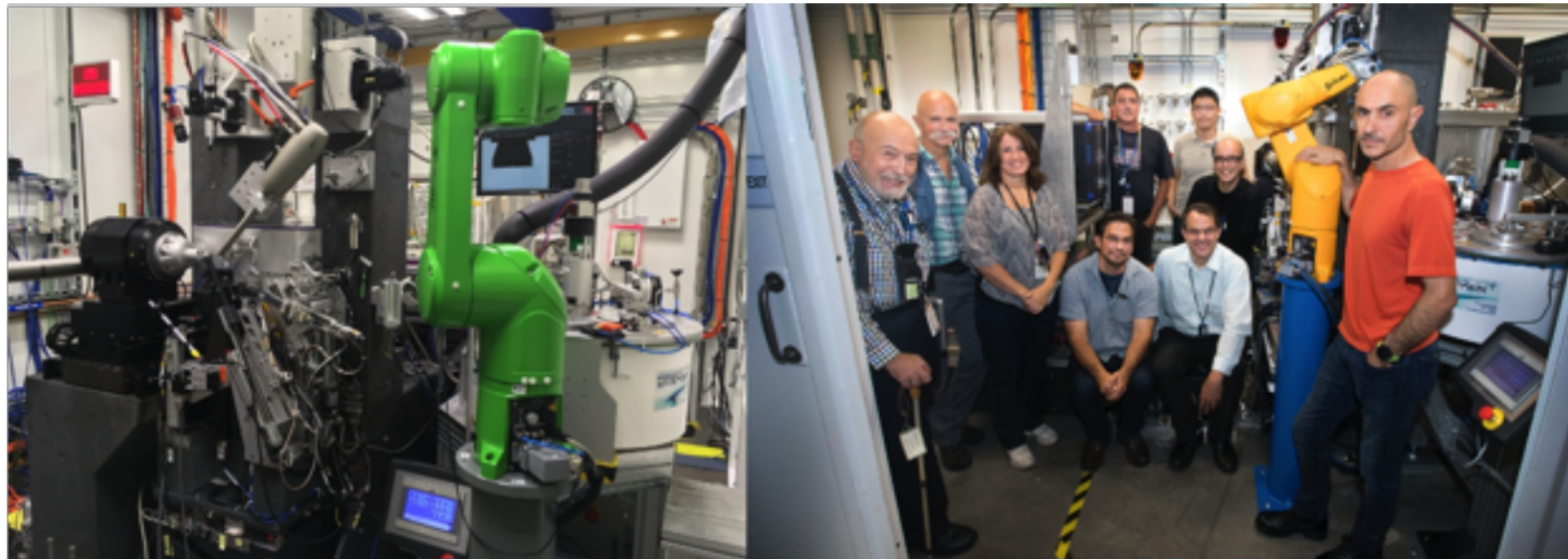
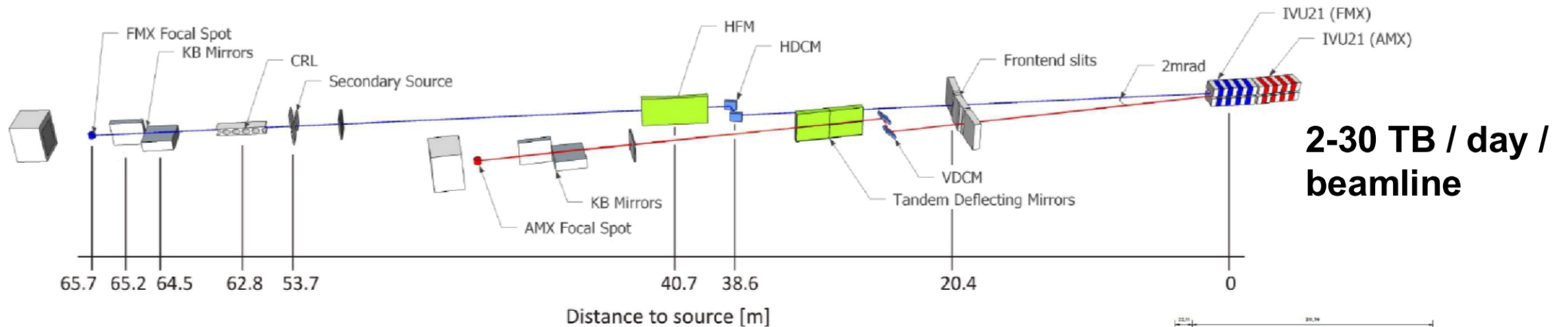


High Data Rate MX (HDRMX) at NSLS-II

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



Beamline Characteristics

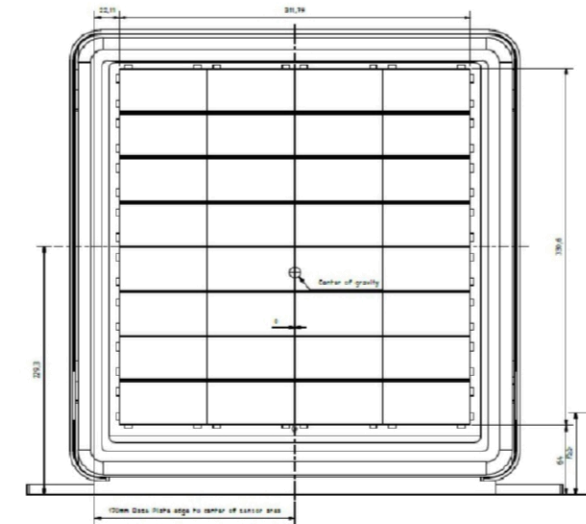


2-30 TB / day / beamline

Specifications

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

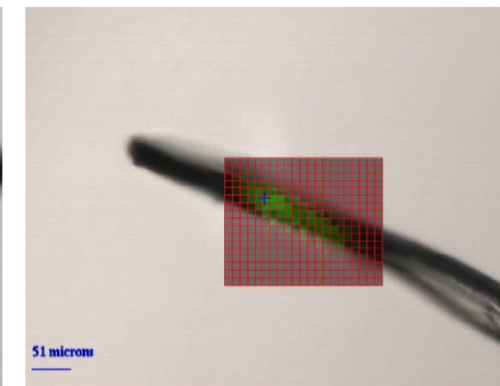
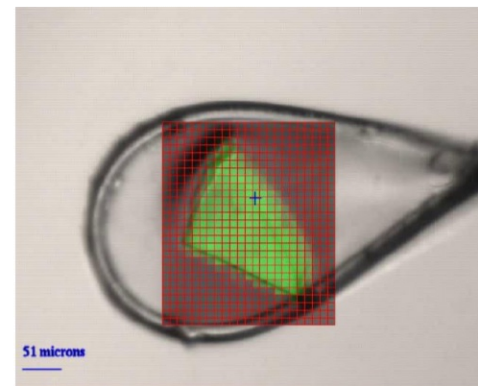
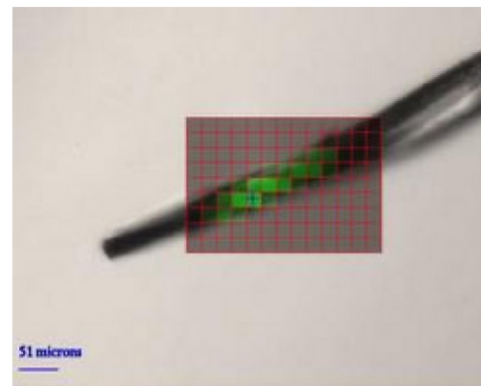
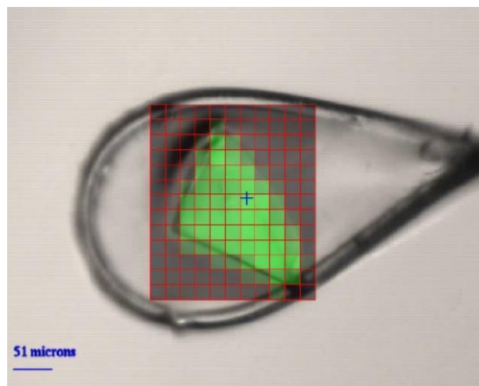
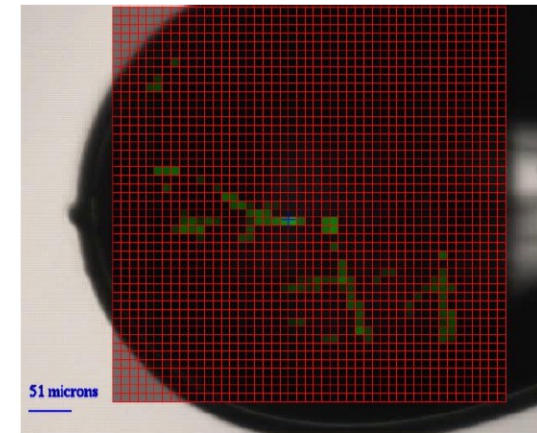
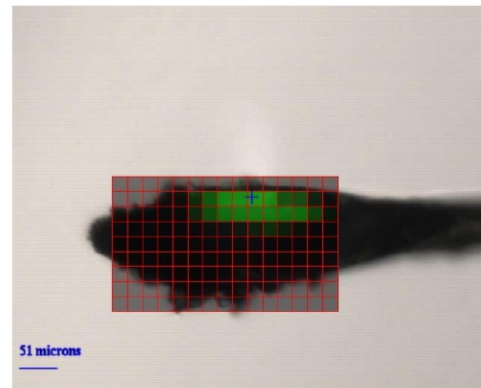
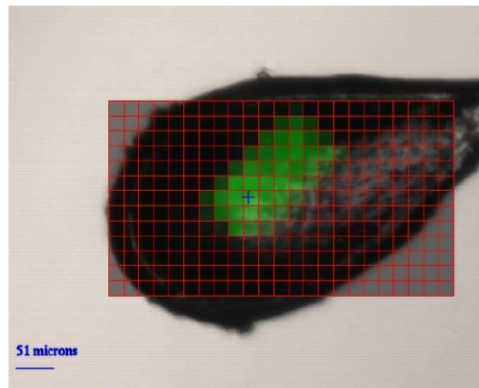
Challenges : data storage, transfer, processing and backup.
Requires advanced software and computing cluster.



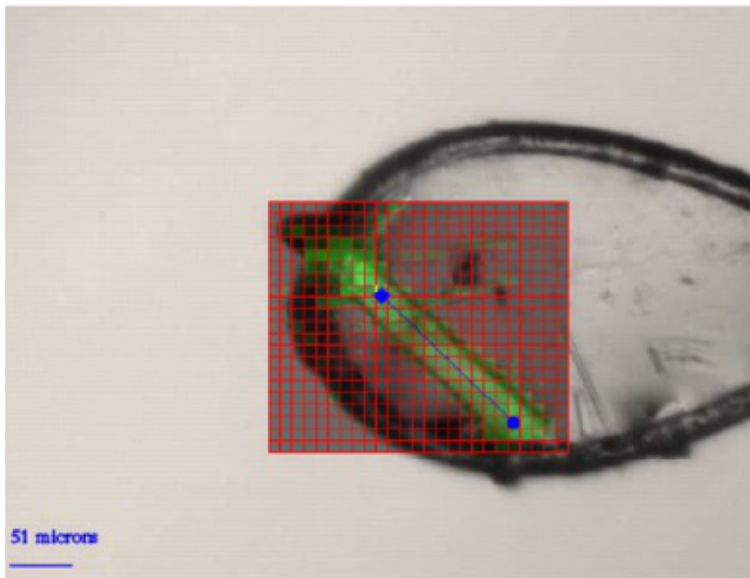
Fuchs, M. R. et al. "NSLS-II biomedical beamlines for micro-crystallography, FMX, and for highly automated crystallography, AMX: New opportunities for advanced data collection *AIP Conf. Proc. SRI2015, 2016, 1741, 030006*

Raster-based centering

Raster every sample



Vector (helical) data collections



AMX Vector

Parameters

Start X	5068.688 um	5588.981 um
Start Y	-402.200 um	200.100 um
Start Z	1637.500 um	1520.000 um
Start Omega	0.00 deg	-4.00 deg
End X	5068.688 um	5588.981 um
End Y	-402.200 um	200.100 um
End Z	1637.510 um	1520.000 um
End Omega	360.00 deg	360.00 deg
Exposure	400.000000 s	400.000000 s
# Samples	9.000000	9.000000

More Parameters

Buffer Time: 3.000000 ms 3.000000 ms

Status

Data Acq Duration: 3600.000000 Shutter Opening Time: 3.666675 ms
 Maximum Time to Speed: 115.199997 m utter Change in Position: 0.000000 ct

	Stack X	Pin Y	Pin Z	Omega
Too Fast				
Acceleration	5.000 ctm/s ²	4000.000 ctm/s ²	4000.000 ctm/s ²	50.000 ctm/s ²
Data Acq Distance	0 ct	0 ct	0 ct	20736000 ct
Desired Speed	0.000000 ctm/s	0.000000 ctm/s	0.000000 ctm/s	5760.000000 ctm/s
Time to Speed	0.000000 ms	0.000000 ms	0.000000 ms	115.199997 m
Direction	1.000000	1.000000	1.000000	1.000000
Speed Up Distance	0.000000 ct	0.000000 ct	0.000000 ct	331775.99251
Buffer Distance	0.000000 ct	0.000000 ct	0.000000 ct	17280.000000
nut Open Distance	0.000000 ct	0.000000 ct	0.000000 ct	21120.050511
Shut Lag Distance	0.000000 ct	0.000000 ct	0.000000 ct	11520.000000
Backup Distance	0.000000 ct	0.000000 ct	0.000000 ct	301696.04301

Meters

Omega	170.673943 deg	170.673943 deg	< 2.000000 deg >	STOP More
Pin Y	329.240 um	329.200 um	< 50.000 um >	STOP More
Pin Z	1535.138 um	1535.100 um	< 100.000 um >	STOP More
Stack X	5555.43 um	5555.43 um	< 350.00 um >	STOP More
Shutter Rot	15.00 deg	15.00 deg	< 5.00 deg >	STOP More
Open Position	-10.000 deg	-10.000 deg		
Closed Position	15.000 deg	15.000 deg		

Shutter

ZEBRA

Sys1

- #0: DISCONNECT
- #1: IN1_TTL
- #2: IN1_ARM
- #3: IN1_DVDS
- #4: IN2_TTL
- #5: IN2_ARM
- #6: IN2_DVDS
- #7: IN3_TTL
- #8: IN3_OC
- #9: IN3_DVDS
- #10: IN4_TTL
- #11: IN4_CMP
- #12: IN4_PCELL
- #13: IN5_ENCA
- #14: IN5_ENCB
- #15: IN5_ENCC
- #16: IN5_CONN
- #17: IN6_ENCA
- #18: IN6_ENCB
- #19: IN6_ENCC
- #20: IN6_CONN
- #21: IN7_ENCA
- #22: IN7_ENCB
- #23: IN7_ENCC
- #24: IN7_CONN
- #25: IN8_ENCA
- #26: IN8_ENCB
- #27: IN8_ENCC
- #28: IN8_CONN
- #29: PC_ARM
- #30: PC_GATE
- #31: PC_PULSE

Sys2

- #32: AND1
- #33: AND2
- #34: AND3
- #35: AND4
- #36: OR1
- #37: OR2
- #38: OR3
- #39: OR4
- #40: GATE1
- #41: GATE2
- #42: GATE3
- #43: GATE4
- #44: DIV1_OUT
- #45: DIV2_OUT
- #46: DIV3_OUT
- #47: DIV4_OUT
- #48: DIV1_OUTN
- #49: DIV2_OUTN
- #50: DIV3_OUTN
- #51: DIV4_OUTN
- #52: PULSE1
- #53: PULSE2
- #54: PULSE3
- #55: PULSE4
- #56: QUAD_OUTA
- #57: QUAD_OUTB
- #58: CLOCK_1KHz
- #59: CLOCK_1MHz
- #60: SOFT_IN1
- #61: SOFT_IN2
- #62: SOFT_IN3
- #63: SOFT_IN4

Error State:

- PULSE1 TRIG WHILE ACTIVE
- PULSE2 TRIG WHILE ACTIVE
- PULSE3 TRIG WHILE ACTIVE
- PULSE4 TRIG WHILE ACTIVE
- PC_BUFF_OVERFLOW
- ERR5
- ERR6
- ERR7

PC AND OR GATE DIV PULSE ENC SYS

Setup:

Capture: Enc1 Enc2 Enc3 Enc4 Sys1 Sys2 Div1 Div2 Div3 Div4

Posn Trig: Enc1 Enc4 Time Units: ms ms

Posn Div: Positive Positive

Arm:

Trig Source: External External 1 IN1_TTL Disarm Arm Status

Gate:

Trig Source: Position Position

Gate Start	Gate Width	Gate Step	Num Gates	Gate Status
0.0000 um	0.0000 um	0.0000 um	0	

Pulse:

Trig Source: Time Time

Pulse Start	Pulse Width	Pulse Step	Capt Delay	Max Pulses	Pulse Status
0.0000 ms	0.0000 ms	0.0000 ms	0.0000 ms	0	

Captured Data: Display Captured Data... Data Download in Progress

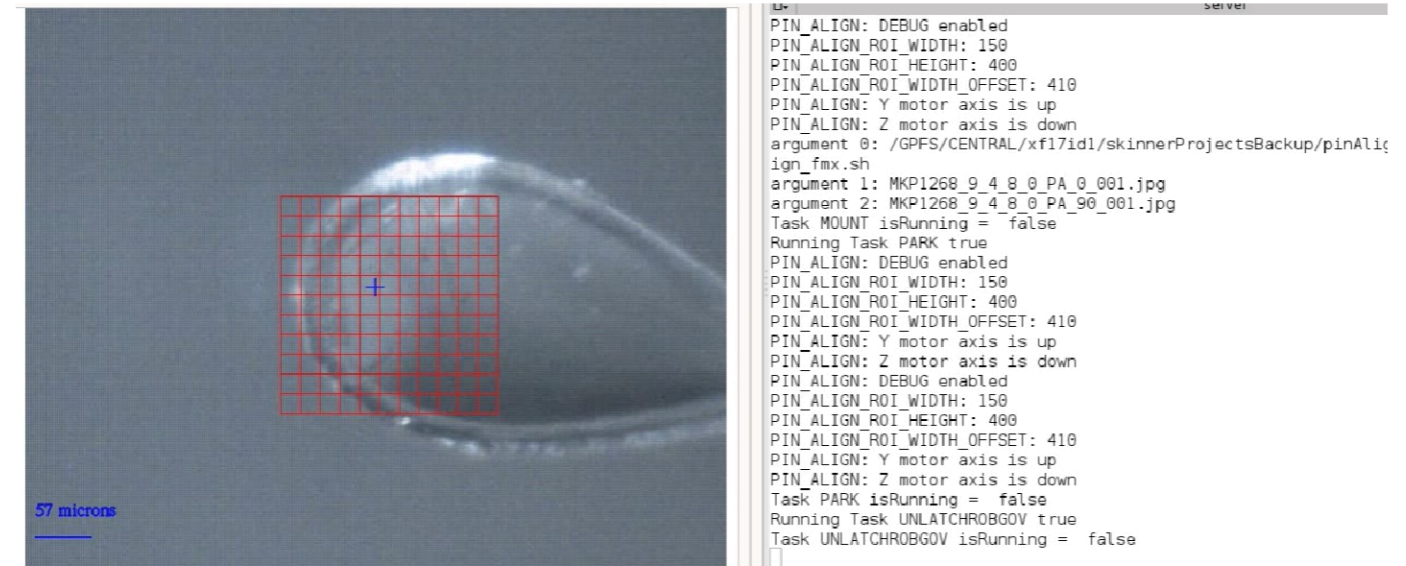
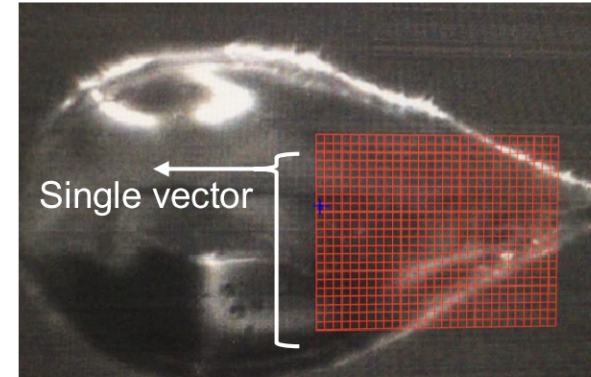
Comms: Connected Soft In: IN1 IN2 IN3 IN4 Block State: Reset Exit

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

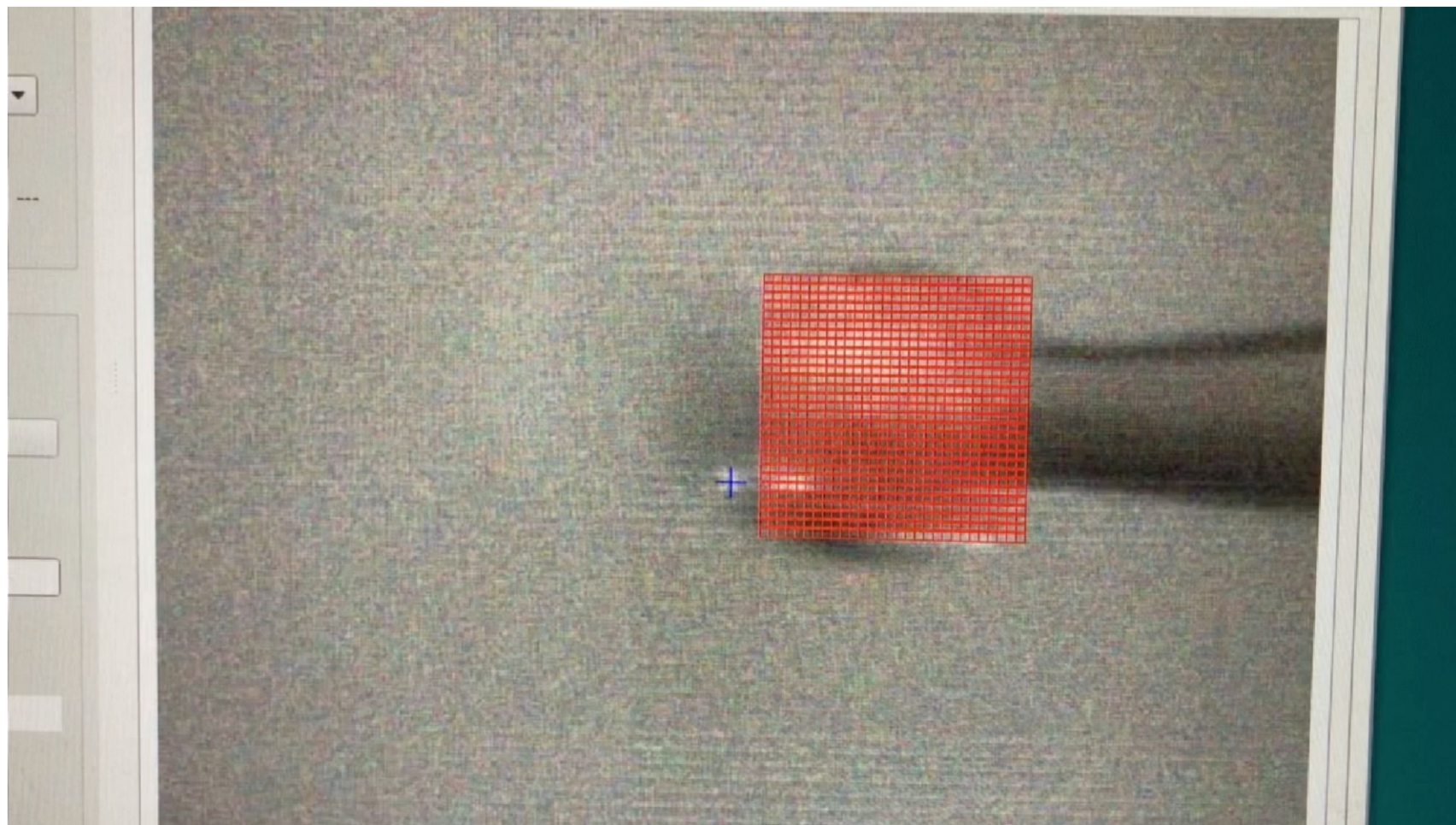
Dozor: G. Bourenkov, A. Popov

LSDC: J. Skinner, J. Aishima

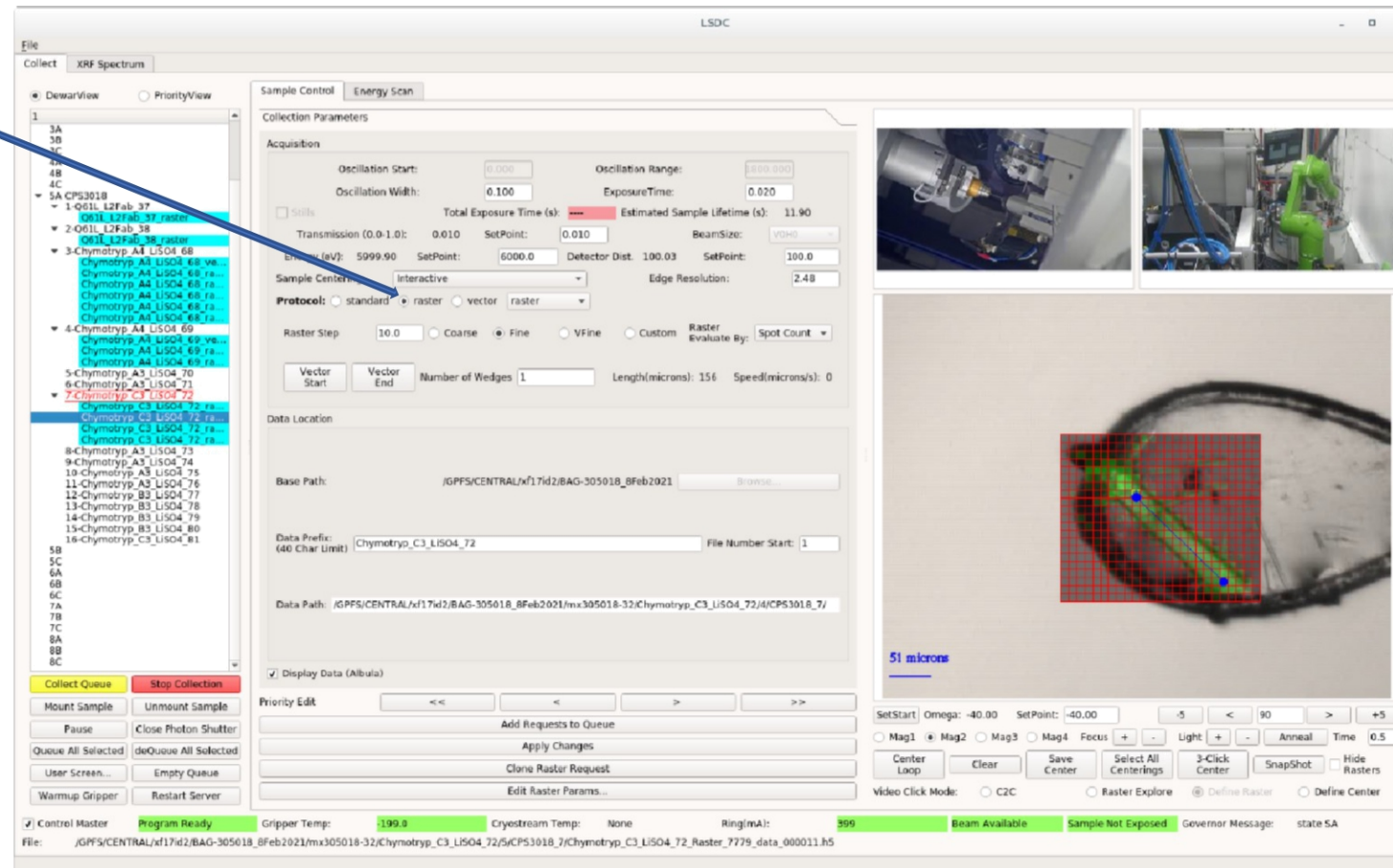


50 Hz; usage typically 50-200 Hz

500 cell raster, ~2.5-10 sec total exposure

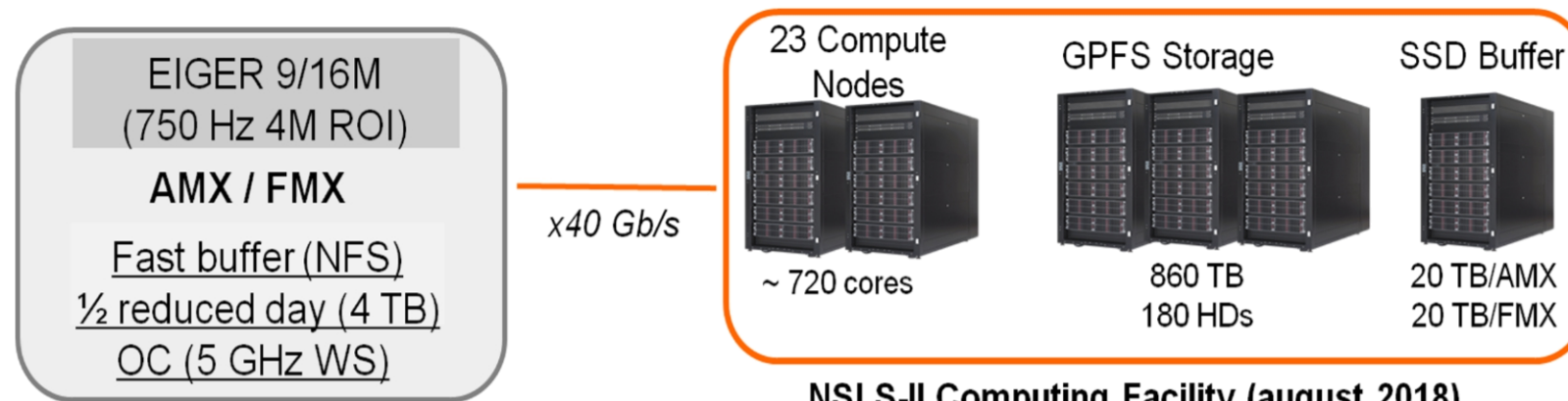


Simple raster queueing



Click 4 corners

AMX/FMX Infrastructure



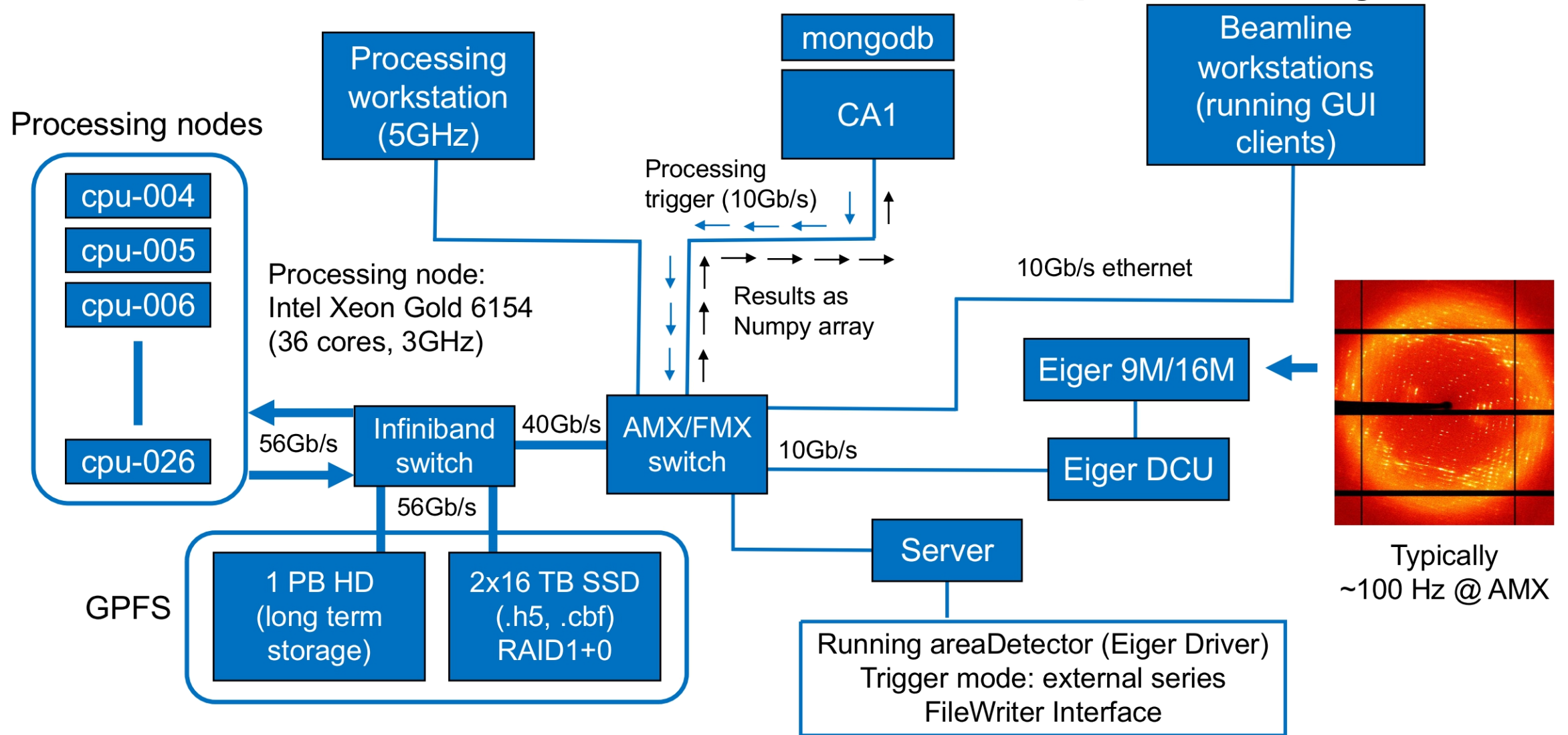
NSLS-II Computing Facility (august 2018)
Scalable Storage and Processing Nodes: sustaining growth

- @ 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

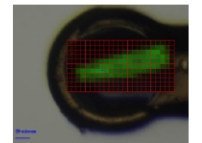
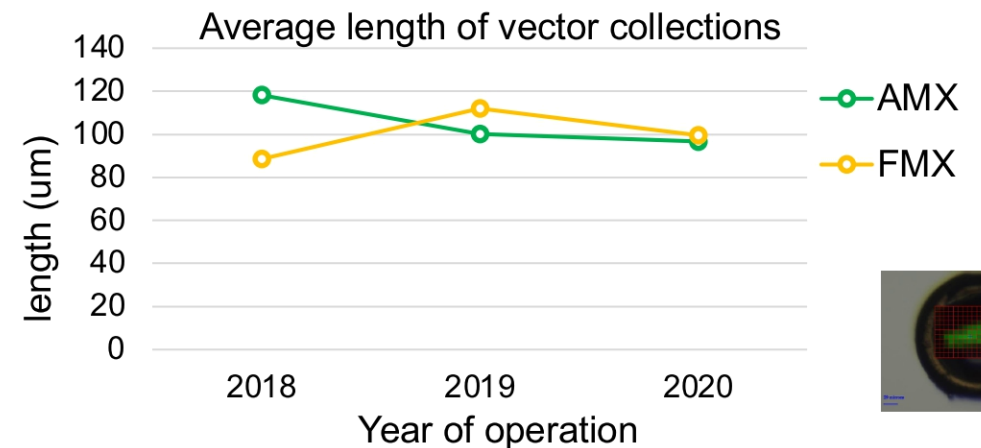
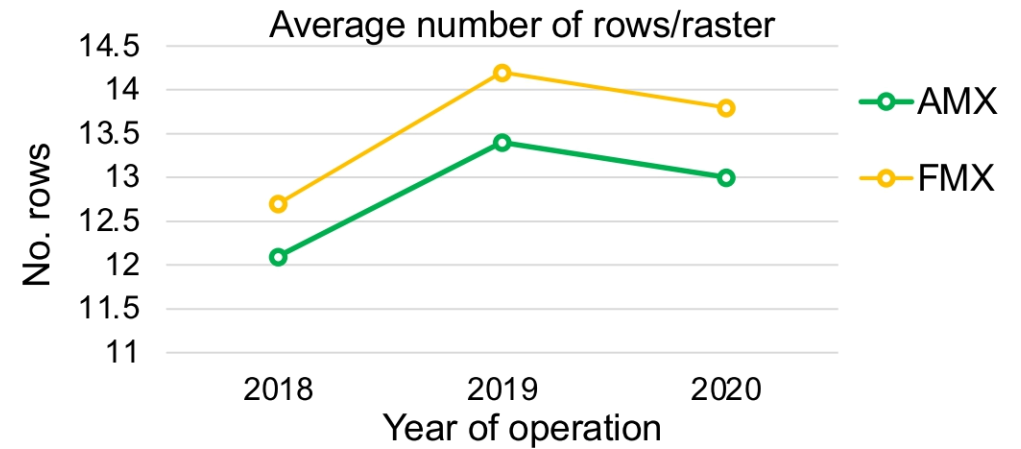
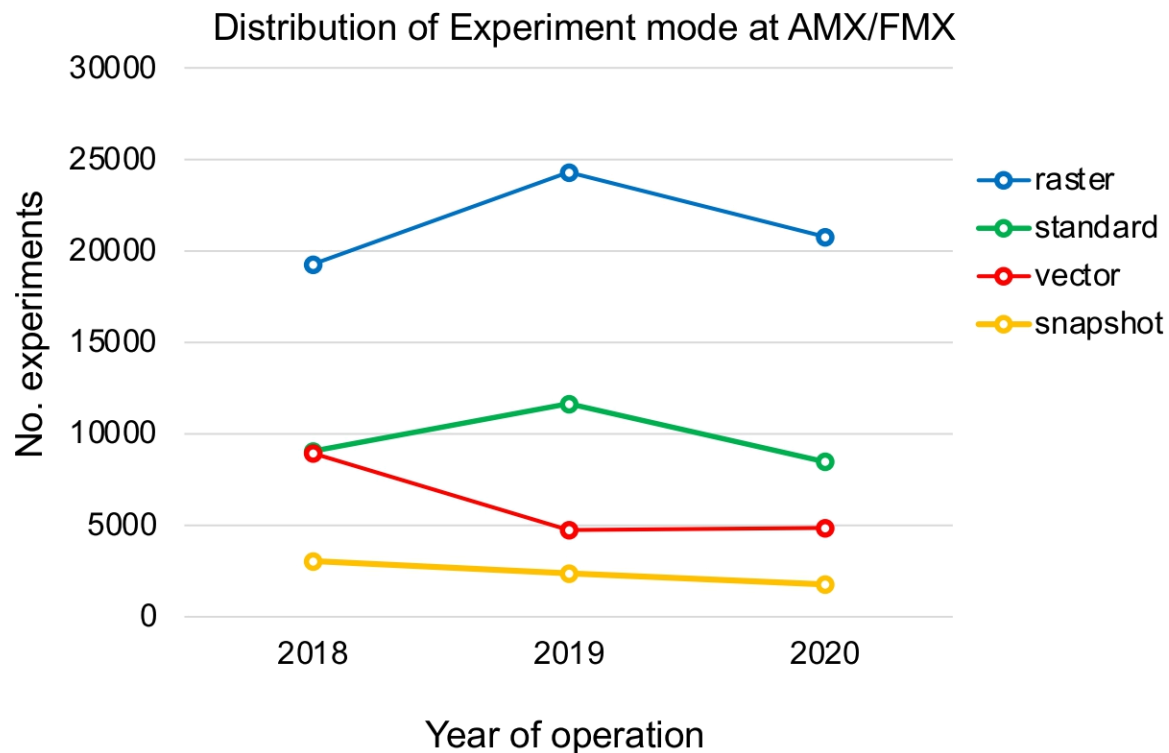
Data transfer to users with globus (<https://www.globus.org>), no hard drives



Network architecture for raster processing



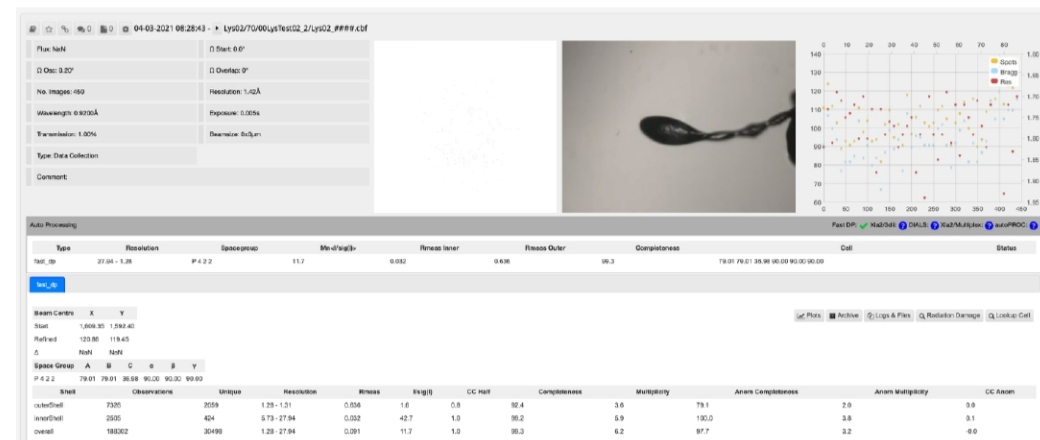
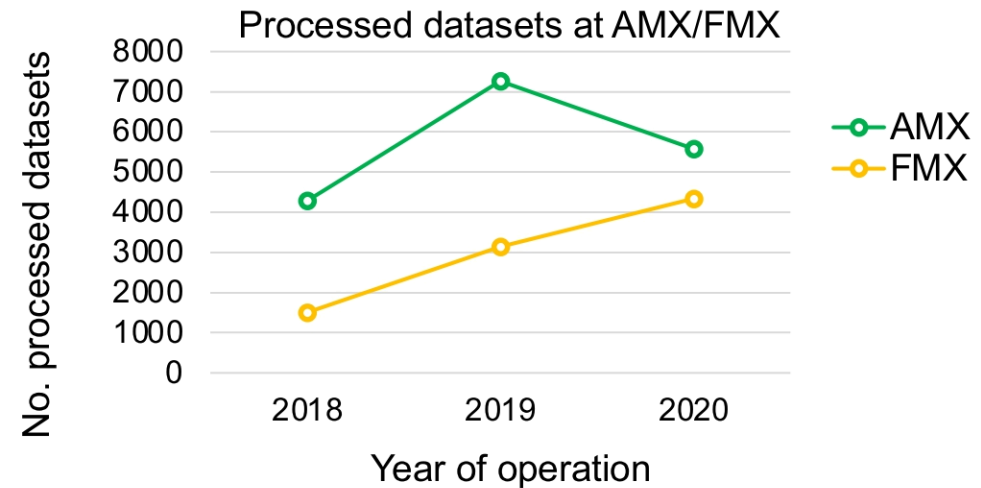
Database statistics of AMX/FMX collections



- ~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

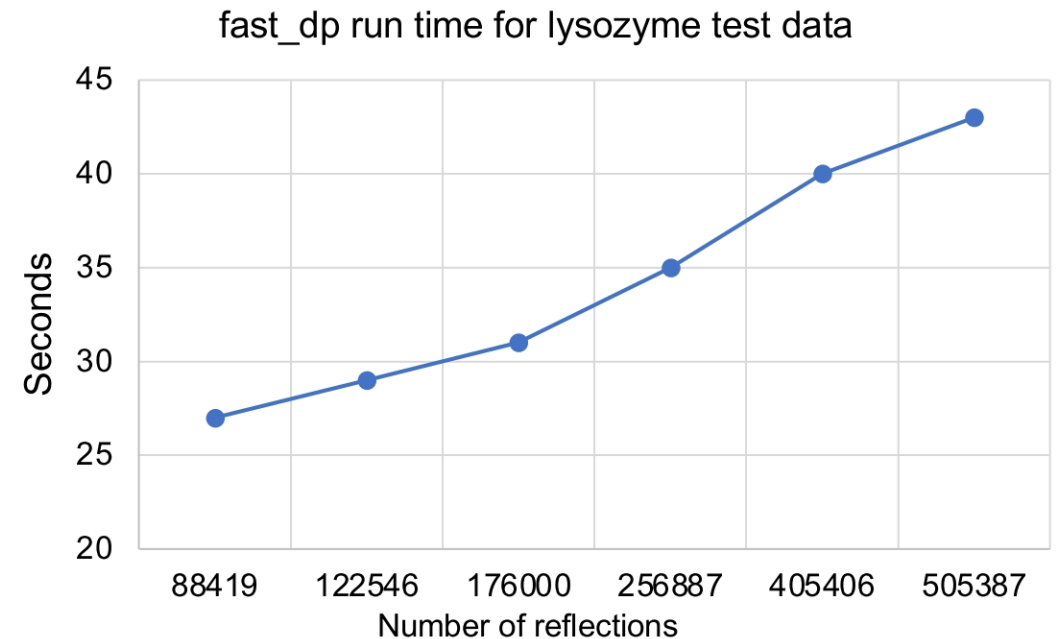
- Autoprocessing of all standard, vector datasets triggered after collection (XDS/fast_dp)
- Typical dataset ~10 seconds (180 deg, 0.2 deg, 100Hz)
- In automated mode, ~1 unipuck/hour
- Well diffracting samples (ligand binding or dynamics) — ~4 min/dataset including sample exchange
- 2 dedicated 36 core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



ISPyB

Data processing timing with fast_dp

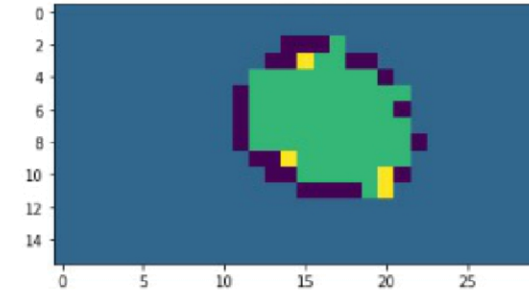
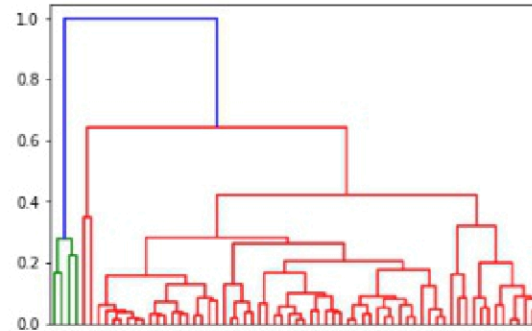
- From lysozyme start up data
- 0.2 deg, 150 deg (750 frames)
- Eiger 9M
- 1 Angstrom resolution
- 8 nodes (XDS):
 - Intel Xeon Gold 6154
 - 36 cores, 3 GHz
- AMX workstation (pointless/aimless):
 - Intel Core i7-8700K
 - 6 cores
 - overclocked @ 5 GHz
- Dimple (tetragonal lysozyme; 12-36 seconds)
-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

Multi-crystal crystallography and Sulfur-SAD at 5 keV
 G. Guo, P. Zhu, Q. Liu (BNL Biology Dept)

a **b**

c

- Data collection at 5 keV ($\lambda = 2.48 \text{ \AA}$) from 5-10 μm thaumatin crystals
- Strategy of data analysis
 - Initial data assembly based on CC1/2
 - Crystal and frame rejections based on SmRmerge
- Rejection criteria for S-SAD phasing should not be too stringent

Guo et al., *IUCr* 2018, 5: 238-246
 Guo et al., *IUCr* 2019, 6: 532-542

Office of Science | BROOKHAVEN NATIONAL LABORATORY | National Synchrotron Light Source II

Data processing WYpipeline

- ~200 partial datasets for structure solution
- Equally high data quality for detector frame rates of 200, 500 and 750 Hz

Protease K structure refined to 2.0 Å resolution
 $R_{\text{work}}=16.7\%$ and $R_{\text{free}}=21.3\%$ (500Hz dataset)

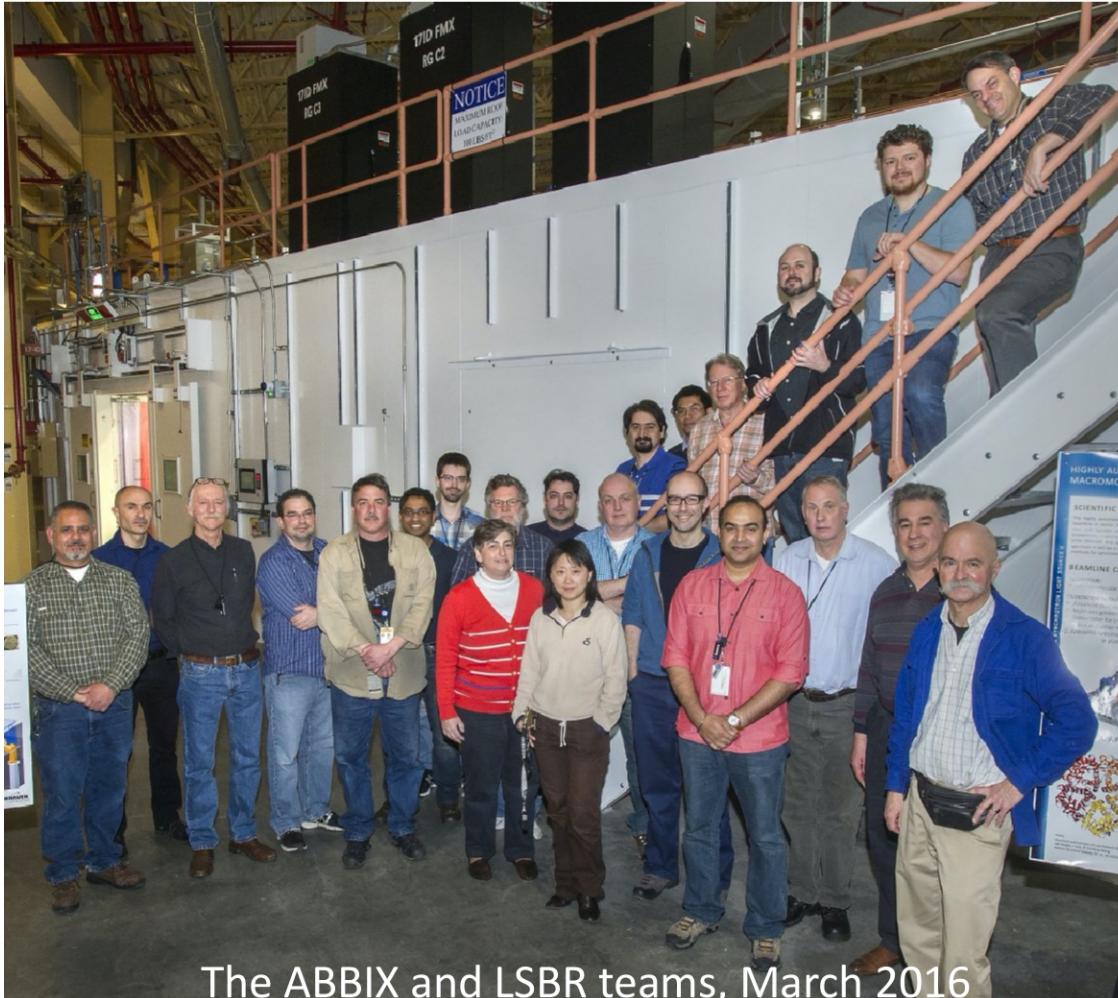
Automated clustering using Machine Learning algorithm

Office of Science | BROOKHAVEN NATIONAL LABORATORY | National Synchrotron Light Source II

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