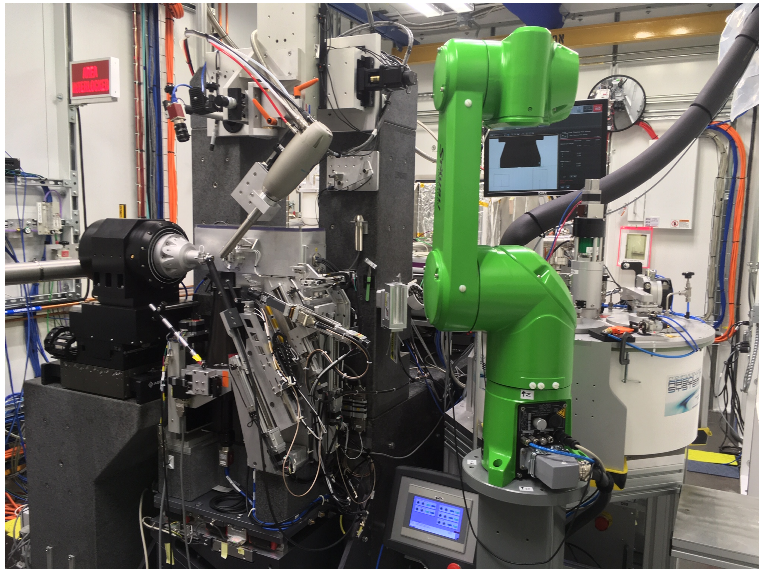
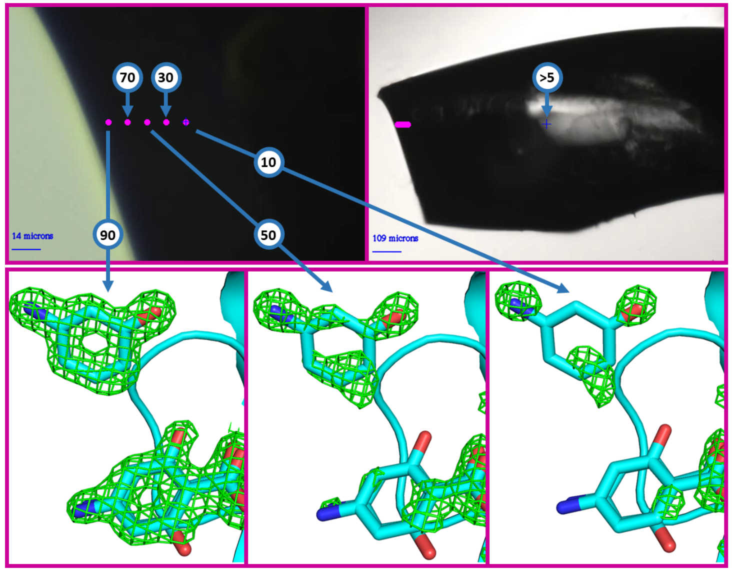
The Gold Standard at NSLS-II





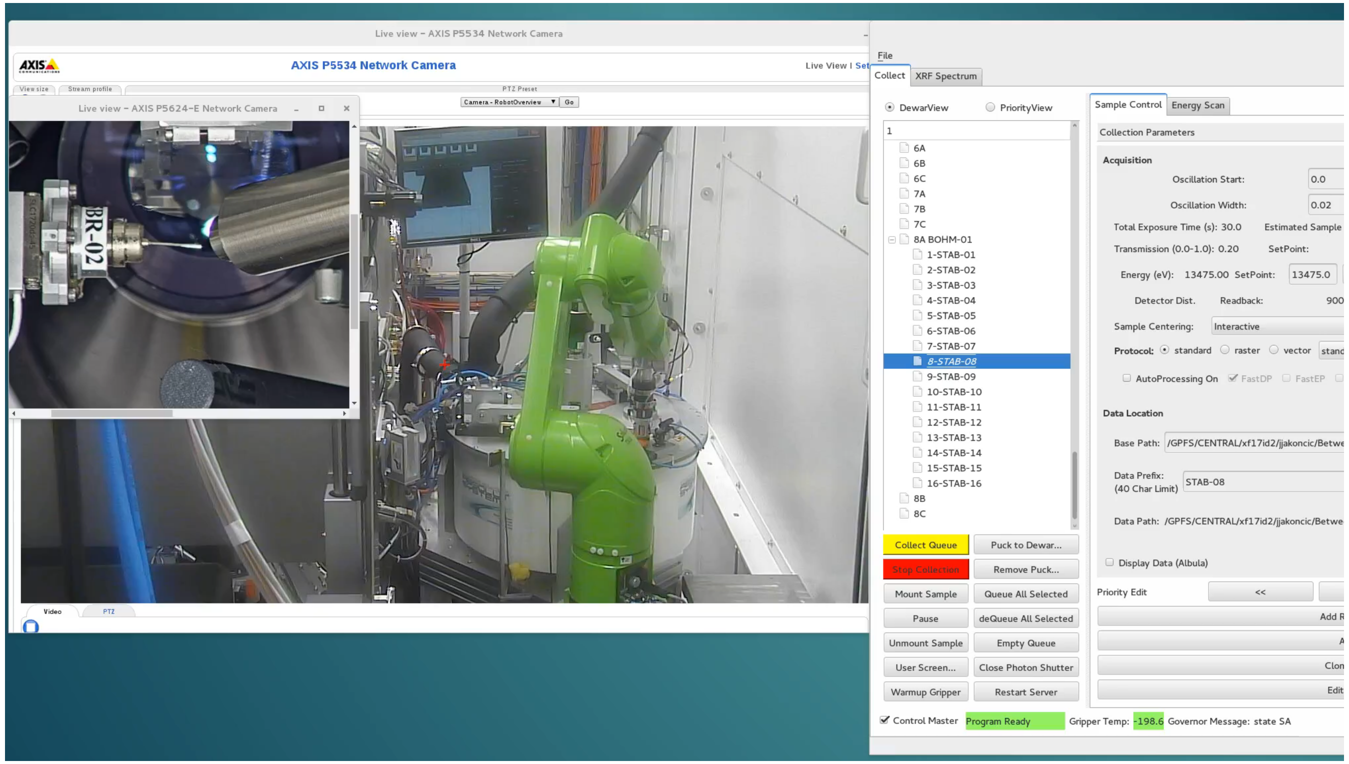
Dale Kreitler: dkreitler@bnl.gov

HDRMX, 22Aug2020

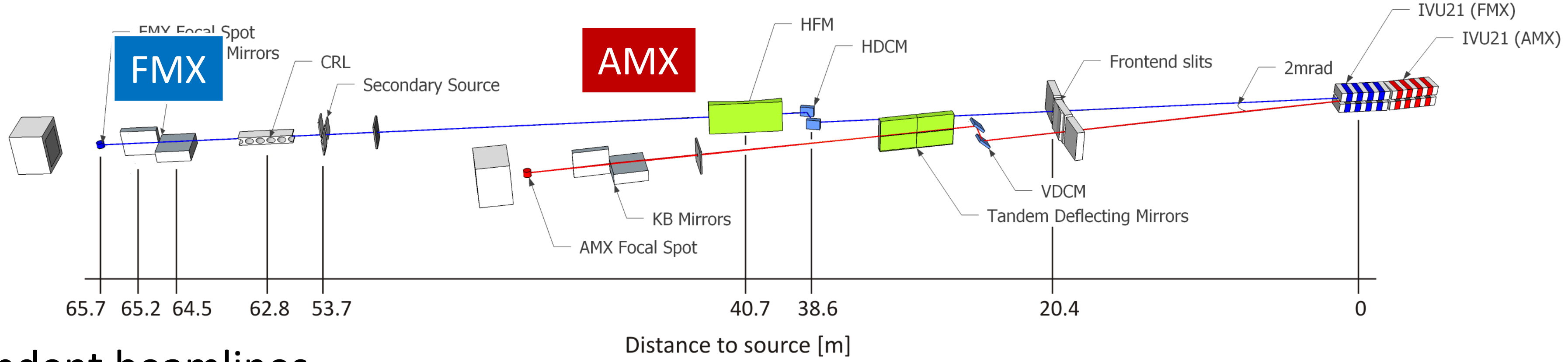
Slides by Jean Jakoncic, Dale Kreitler

What It Is That We Do ?

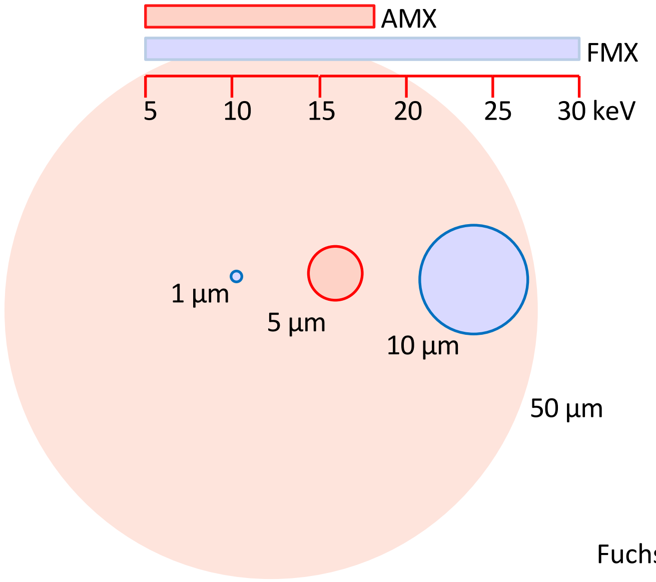
- 2 MX beamlines: AMX and FMX in operation since 2017-01
- Support challenging projects: microcrystals, opaque medium/membrane proteins, complexes
- Each beamline sees up to 2 groups per day
- Local / *Remote* / Automated
- Collection protocols
  - Raster
  - Rotation (standard)
  - Vector (distribute dose)
  - Multiple (small crystals in mesh)
  - Serial (loop/mesh/jets)
- Capacity: 384 Spine samples



Beamlines Characteristics



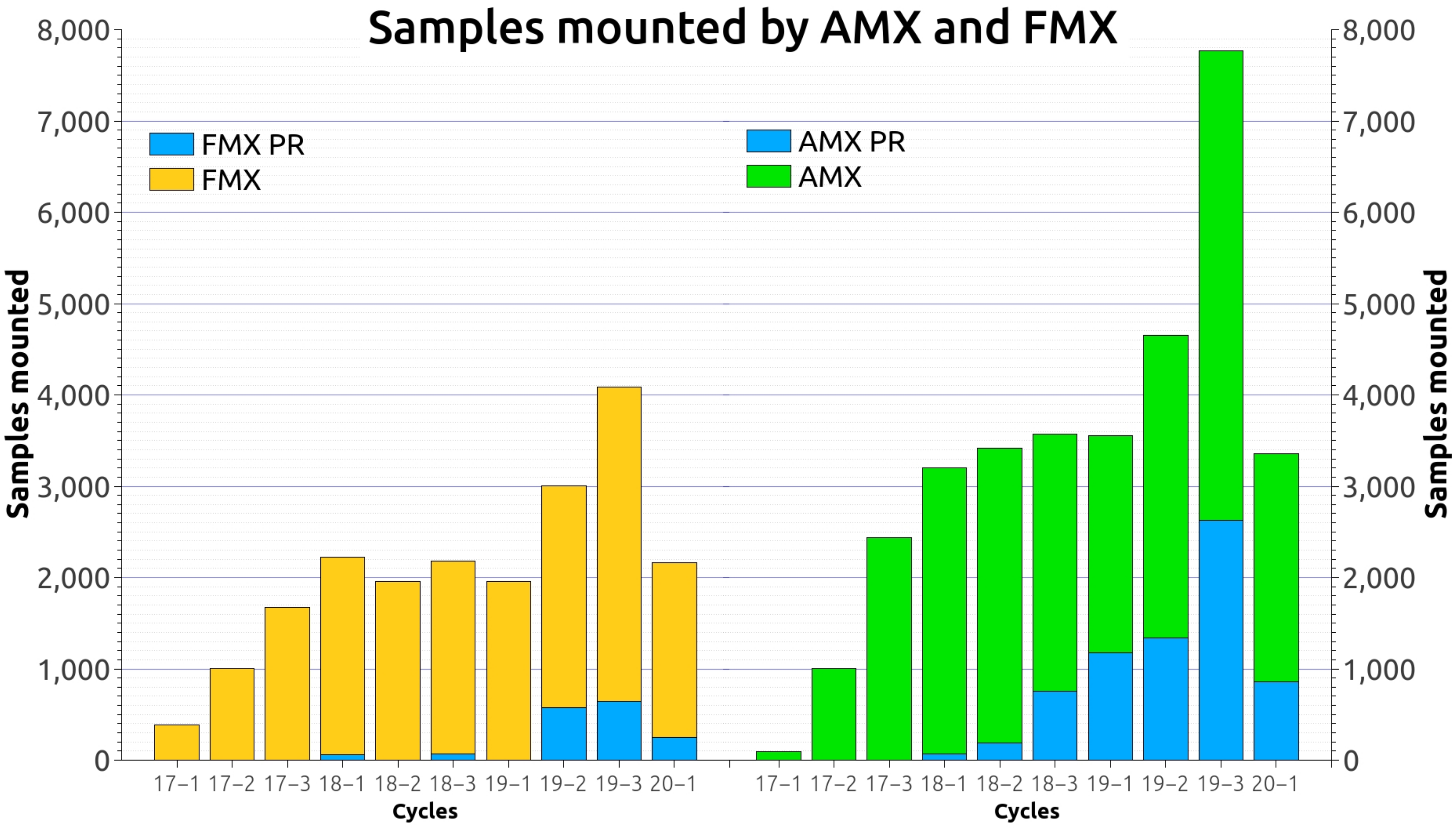
Two independent beamlines with overlapping and complementary capabilities



Specifications	AMX	FMX
Energy range	5 – 18 keV	5 – 30 keV
Wavelength range	0.7 – 2.5 Å	0.4 – 2.5 Å
Flux at focus at 12.7 keV	>4×10^{12} ph/s	3.5×10^{12} ph/s
Focal spot min (H×V)	7 × 5 µm^2	1.5 × 1 µm^2
Focal spot range	NA	1 – 20 µm
Detector	Eiger 9M	Eiger 16M

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

AMX/FMX sample throughput



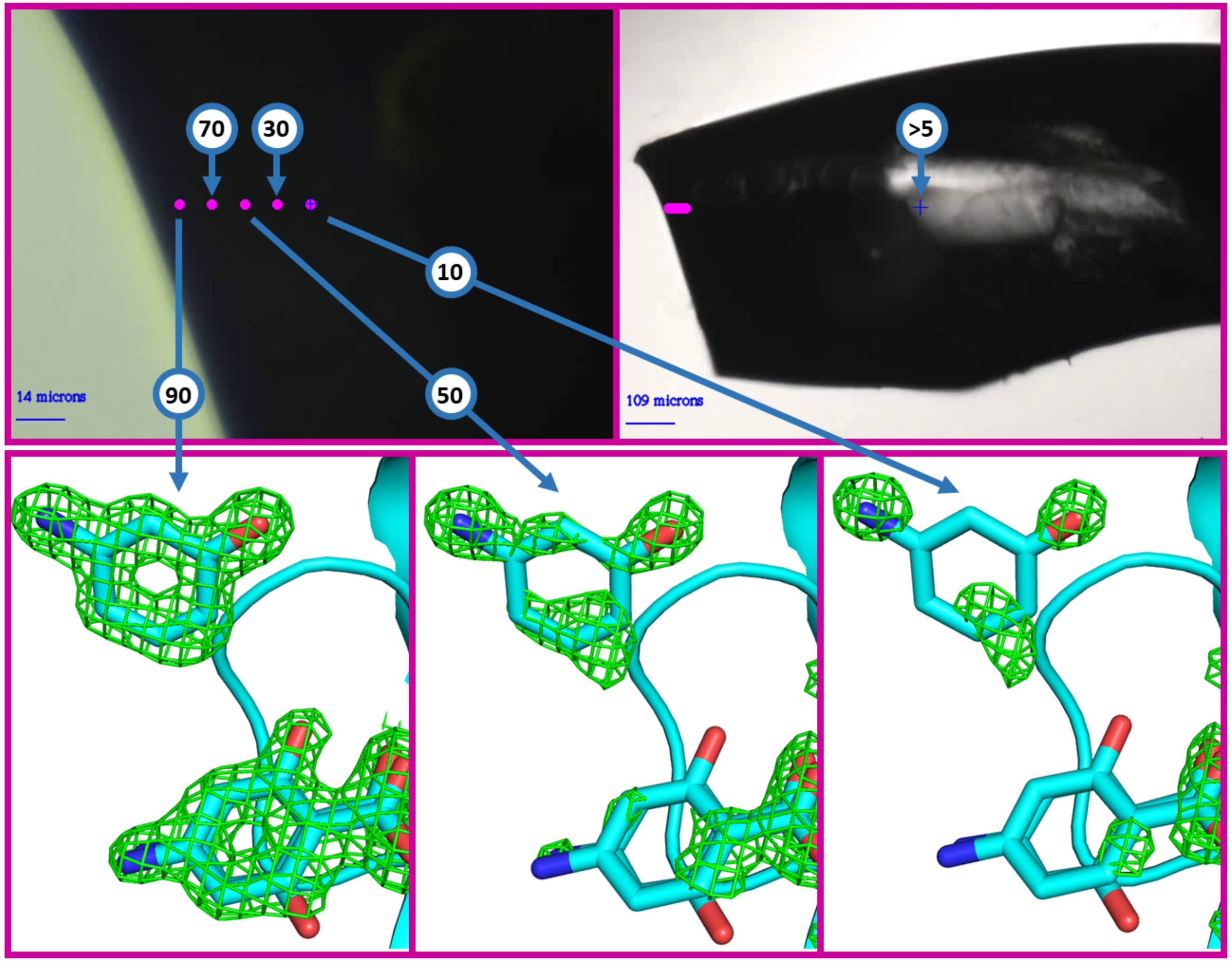
What can you do with a microbeam?

Inhouse test experiment: soak "large" crystal in ligand and collect data using the 7 micron beam at different "depth" (as perceived by ligand). Occupancy: in %

Hypothesis: not all ligands diffuse in crystals during soaking experiments

Assuming sufficient solubility and sufficient affinity.

Can we cluster related data sets based on small differences?

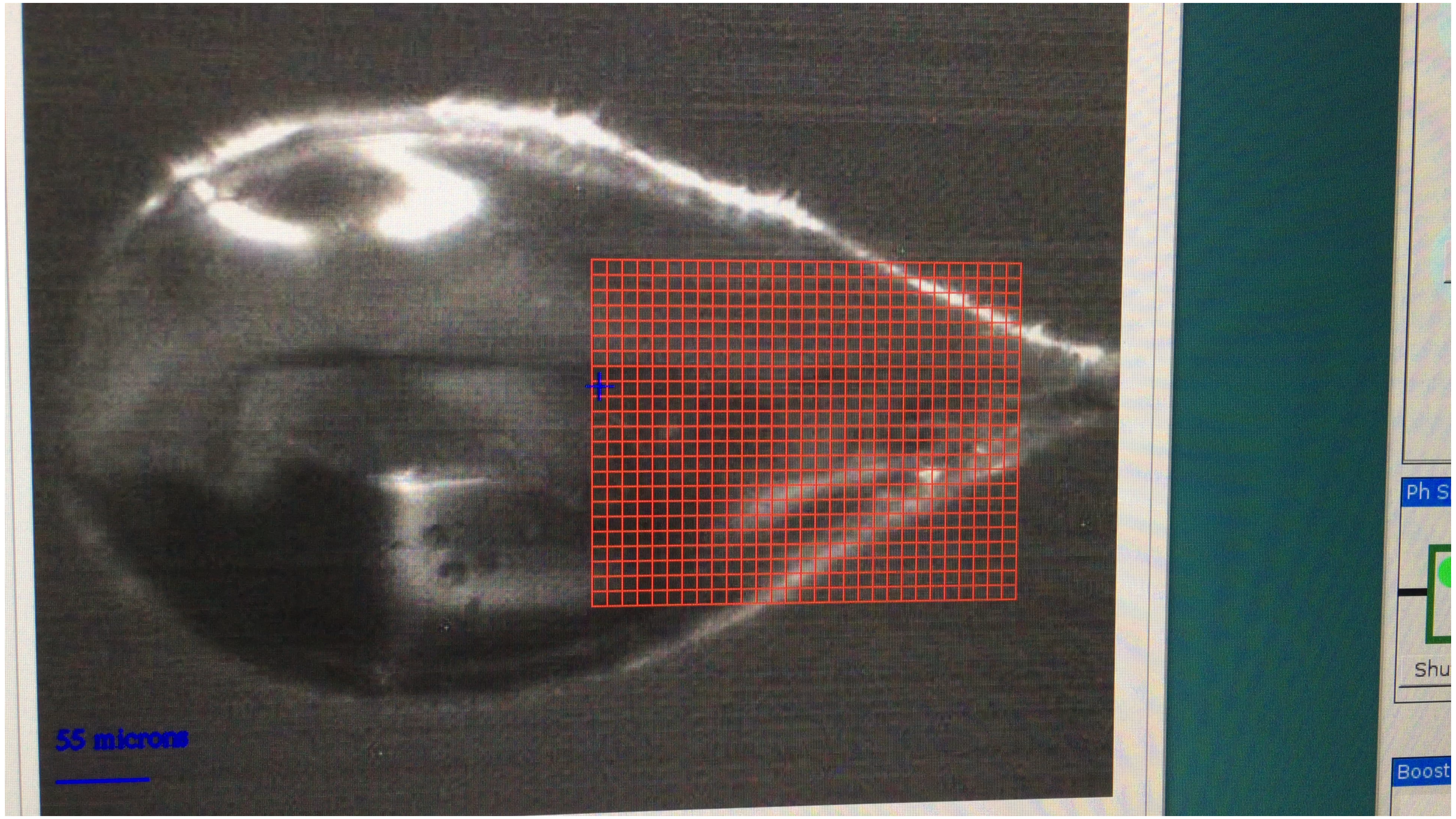


(Soares & Jakoncic, in prep)

Fast Rastering: absolute requirement

- 1 HDF5 data file per row
- LSDC server sends requests to dedicated nodes
- Convert from HDF5 to CBF
- Execute dials.find_spots_client (client/server)
- Count reflections and compute map.
- SSD buffer and GPFS cluster
- Limitations for very large rasters, 10 micron steps or less results in compute bottleneck

➢ Need more nodes
➢ Working to implement improved spot finder, (native HDF5 dozor using durin-plugin.so)
➢ On 480 frame raster test data (single node)
  ➢ dectris-neggia.so : ~**4 seconds**
  ➢ durin-plugin.so : **~2.8 seconds**
  ➢ durin-plugin.so/parallel-ssh
    ➢ 3 nodes: **~1.5 seconds**



Movie Recorded in 07/2018
20 ms exposure per frame
~30 frames/row, 0.6 seconds/row

Thanks Gleb and Sasha!

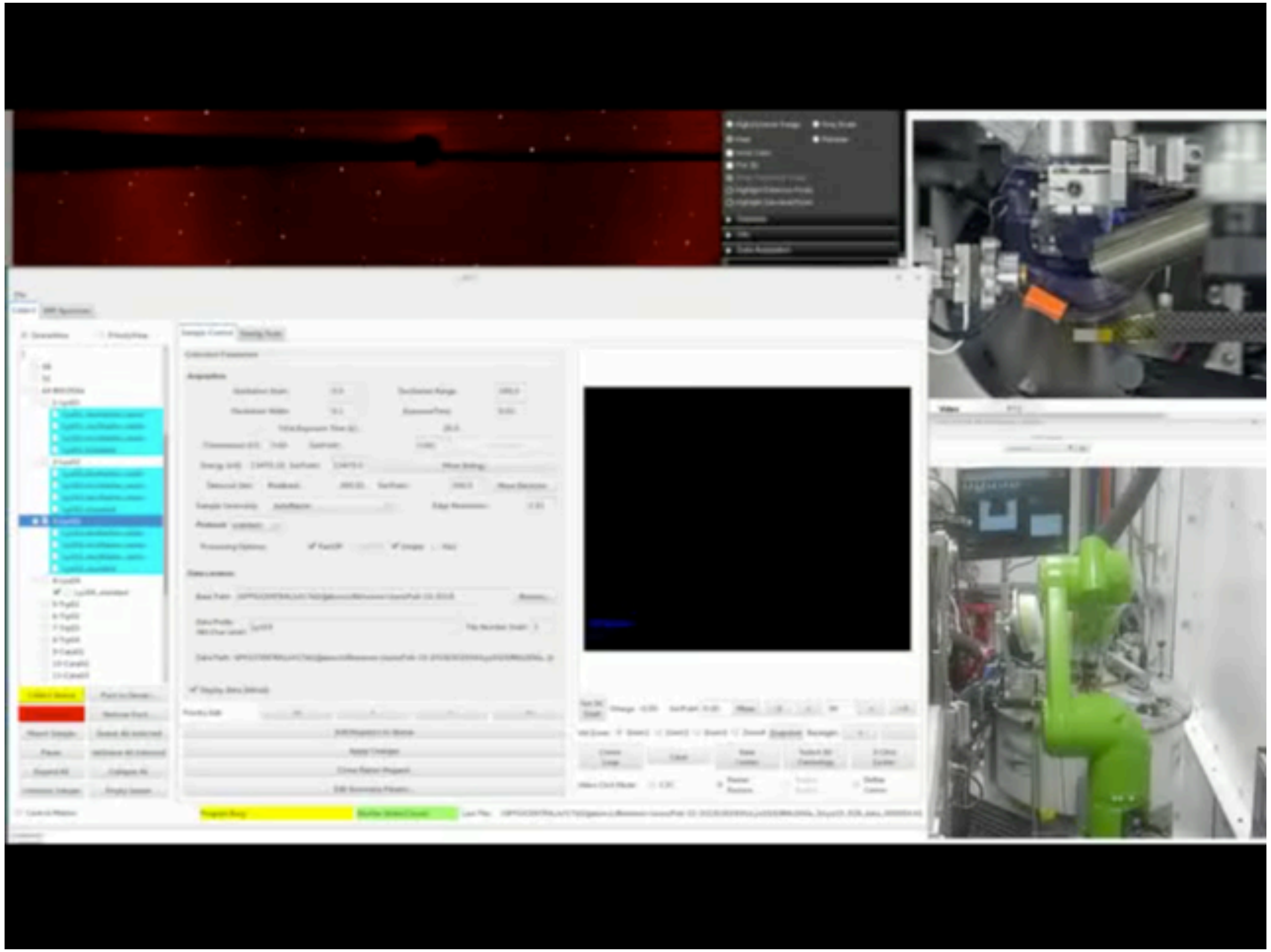
Fully Automated Data Collection

- Auto Raster & Auto Collect
  - loop centering (83 secs/sample)
  - crystal centering (163 sec / sample)
  - **40 samples per hour (loop center)**
  - **22 samples per hour (crystal center)**

Tons of Data
- Fast file system required
- Fast feedback required to achieve high throughout

Needs:
- Fast spotfinding
- More protocols (vector)
- Sorting / Strategy
- Witnessed user this week do 200 vector data sets manually at 4 min / structure

Leveraging existing codes:
- Xrec & C3D

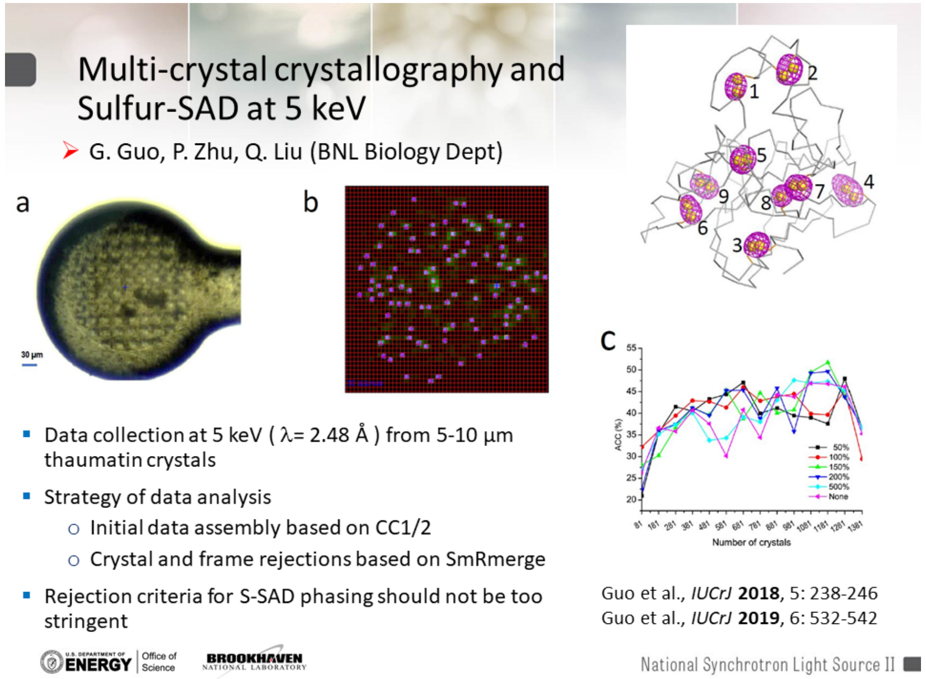


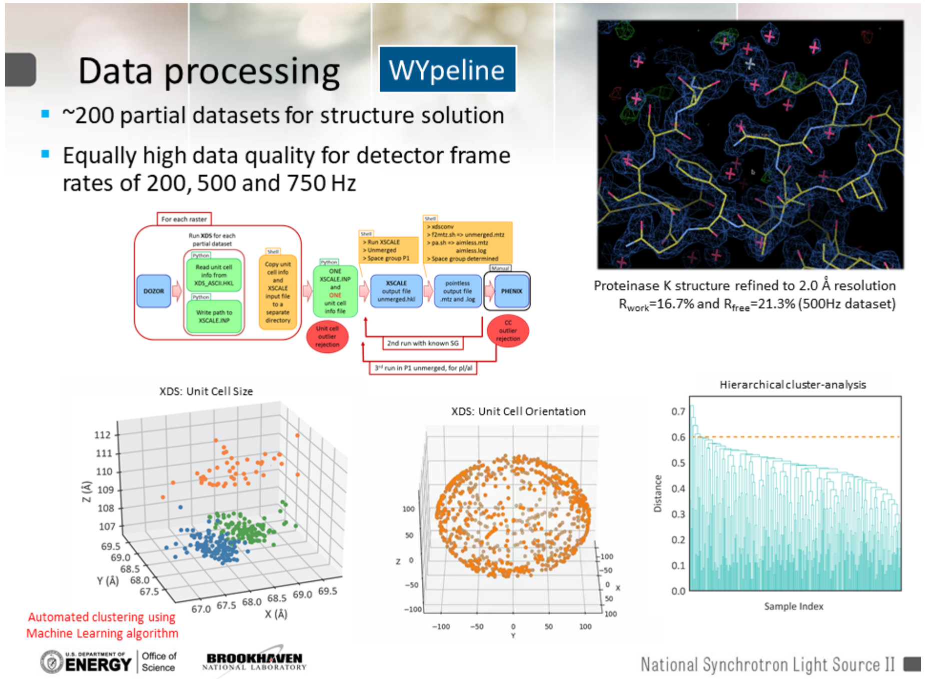
Movie recorded in 02/2018

Automated Data Processing

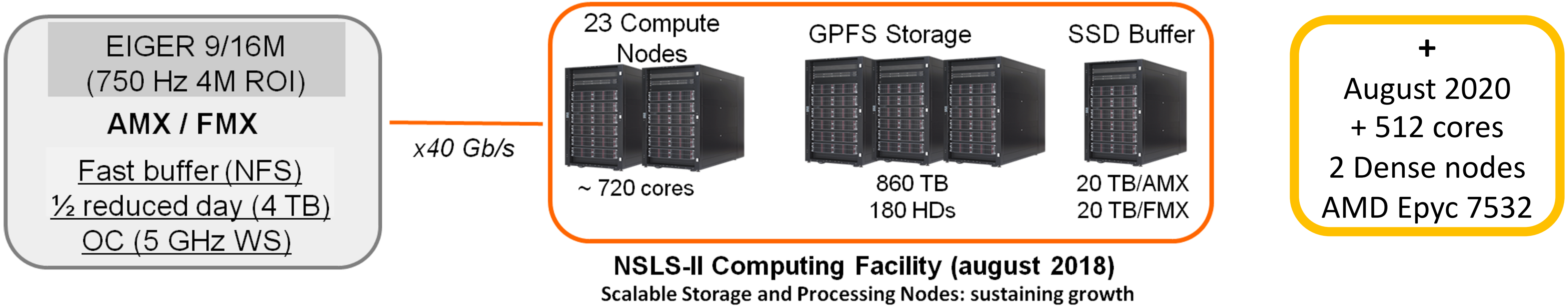
- Spot finder for rastering *(dials, dozor)*
- Data reduction (fast_dp_nsls-ii (XDS)) for all data collection with 5 degrees or more *(+fast_dp_redo)*
- Ligand binding using dimple_nsls-ii (modified for speed and parametrizable): MR + refine + coot map search
- Fast_ep / Fast_ep_weak / Fast_ep_nsls-ii : for SAD structure solution following fast_dp
- Fast_dp_pro (prune data), xmerge (merge and sort) are being tested

For serial crystallography: pipeline developed at FMX (Yuan Gao) and KAMO + pipeline developed by Qun Liu (Bio)





Our IT Infrastructure



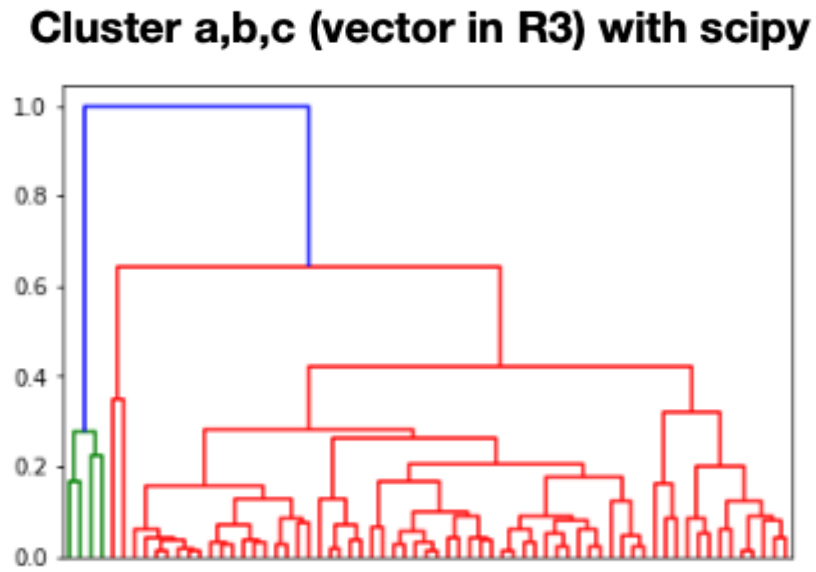
Fast_DP	1 NSLS-II node 44 cores 2.1 GHz	1 SOEIL node 144 cores 2.5 GHz	6 intel nodes 336 cores 2,1 GHz	5 NSLS-II nodes 220 cores 2.1 GHz + 1 AMX WS : 6 cores OC @ 5 GHz
Time (s)	**83**	42	32	**27**

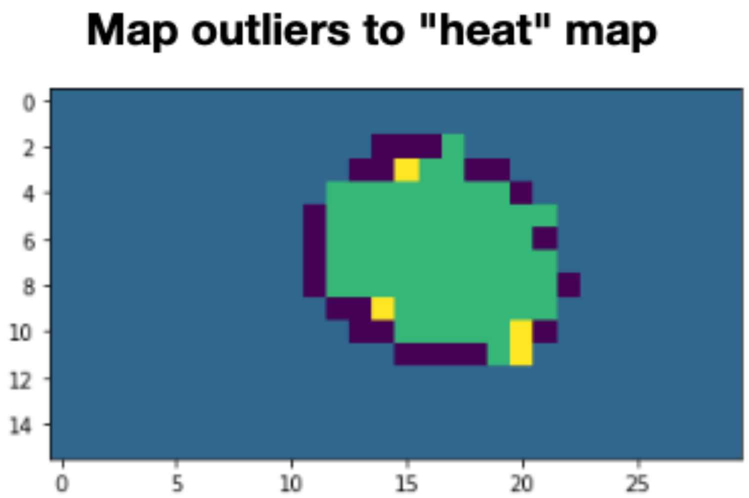
Time (s)	**dimple**	**dimple-fast**	**dimple-faster**
Standard node	**52**	22	18
Optimized workstation	36	15	**12**

Optimizing computing hardware and softwares (inhouse improvements): 3x time speed up for data reduction and ligand binding studies.
Github: https://github.com/nsls-ii-mx

Share resources when not in use, e.g. slurm with VinaLC (computational drug screening)

Currently under development: augmented raster heat maps (in-line indexing)





- Borrow tools from XFEL for single frame indexing
- Efficient algorithms/code that allows for parallelization in distributed memory cluster
- Speed required for inline data-processing (spotfind/index 1000-5000 frames raster < 5 seconds)
- Tomograms for automated vector collection protocols (user this week did several hundred vectors by hand)

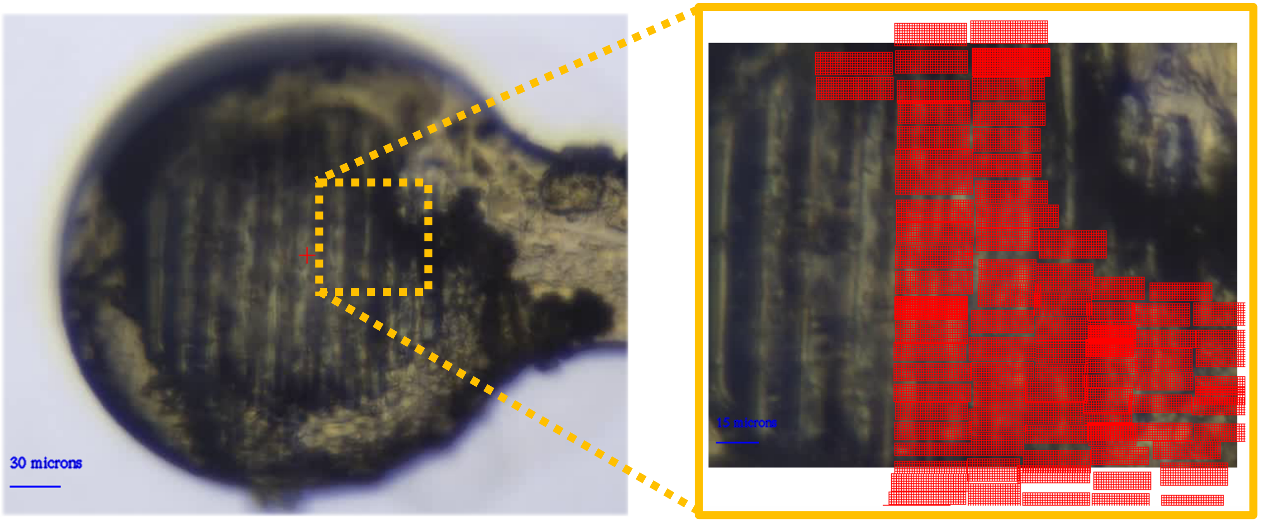
Melnikov et al. 2018, Winter et al. 2018

Future Upgrades / Needs

30 rasters for 1 dataset:
Shutter-open time 18s!

- Improved automation protocols
  - Error reporting/recovery not just data collection algorithms
- Parallel data processing pipelines (off-line)
- *1000 structures in a day project (~1.5/min)*
- Combine spotfinding/indexing for raster centering

Raster data collection



1 µm step raster scans with 1.3 to 5 ms exposure
(~5 µm Proteinase K crystals) (Gao Y. et al, JSR 2018)

Currently: 2 independent beamlines with fast framing detectors
AMX: Dectris Eiger 9M @ > 200 Hz: 200-1500 MB/s
FMX: Dectris Eiger 16 M @ > 100 Hz: 100-1500 MB/s

Multi layer monochromator with 1 % bandpass (~ 120 eV at 12.5 keV): 100x flux increase
Associated with an upgraded fast framing next generation large area detector: ~ 10 MP @ 5000 Hz

From ~ 1 GBs to ~ 50 GBs

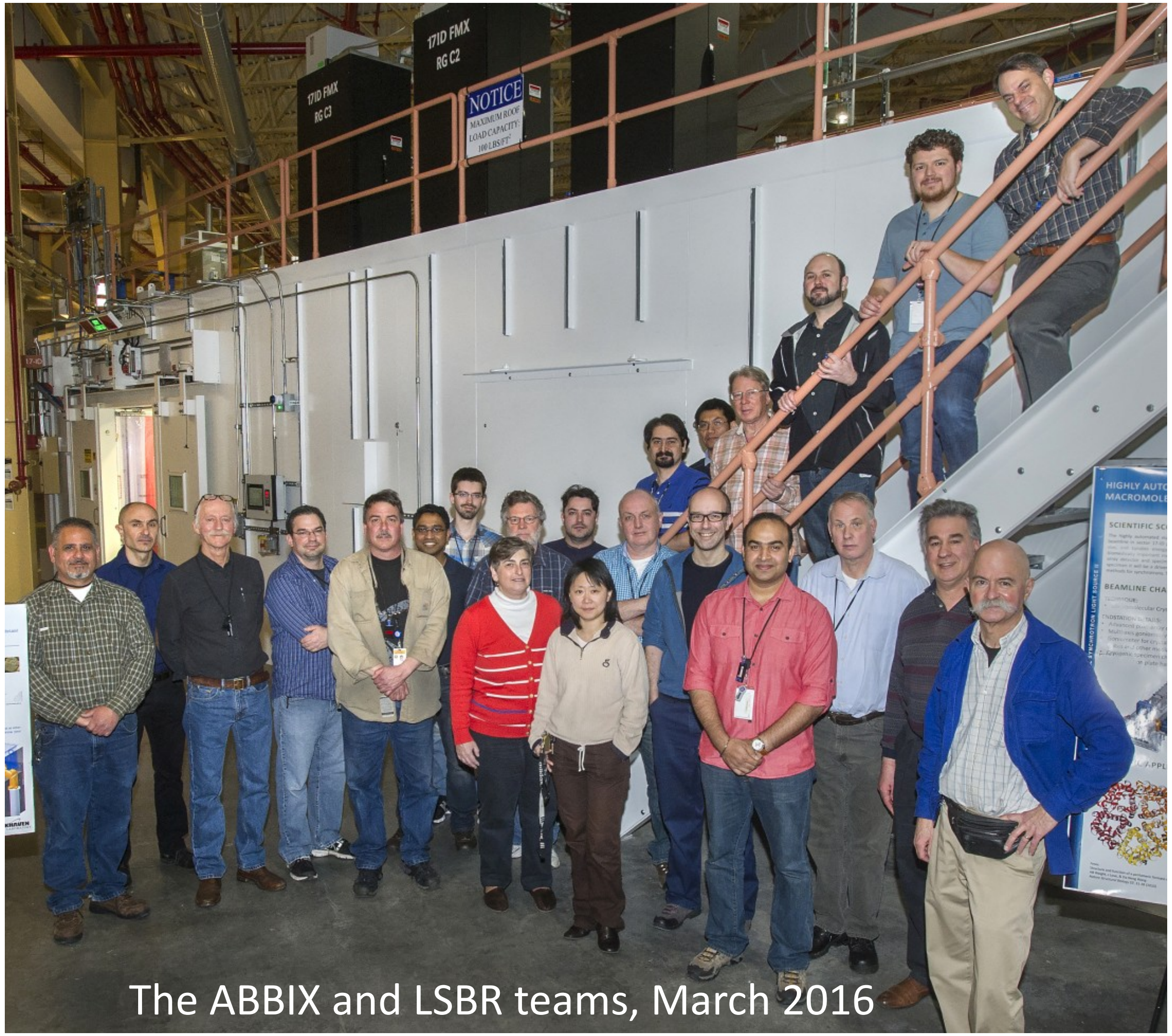
Long Term Goal @ AMX: 1000 crystals / day

- For very high throughput ligand binding studies of well characterized samples
  - fragment based screening
  - difficult target
  - Industrial proprietary research (pharma)

- For studying protein dynamics using: 1000’s of partial or complete data sets

- We are working toward hardware and software deliverables toward that goal and collaborate with groups requiring both.

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
- *Lin Yang*
- Neil Whalen
- Dileep Bhogadi
- Rick Jackimowicz
- Lonny Berman
- Edwin Lazo
- Bob Sweet
- Alexei Soares
- Wuxian Shi
- Yuan Gao
- Jim Magill
- Nicolas Guichard
- *Johnathan DiFabio*
- *Oksana Ivashkevych*
- Herbert Bernstein
- Yusuke Yamada

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