National Synchrotron Light Source II

Data Acquisition, Processing and Monitoring at the NSLS-II Life Science Beamlines - an Adaptable Approach for Long-Term Development

Jun Aishima
Data Science and Systems Integration Program
National Synchrotron Light Source II

U.S. DEPARTMENT OF ENERGY | Office of Science

BROOKHAVEN NATIONAL LABORATORY

Outline

- What opportunities do we have?
- How are we seizing the opportunities?
- How are we developing the systems to meet the challenges of the future?

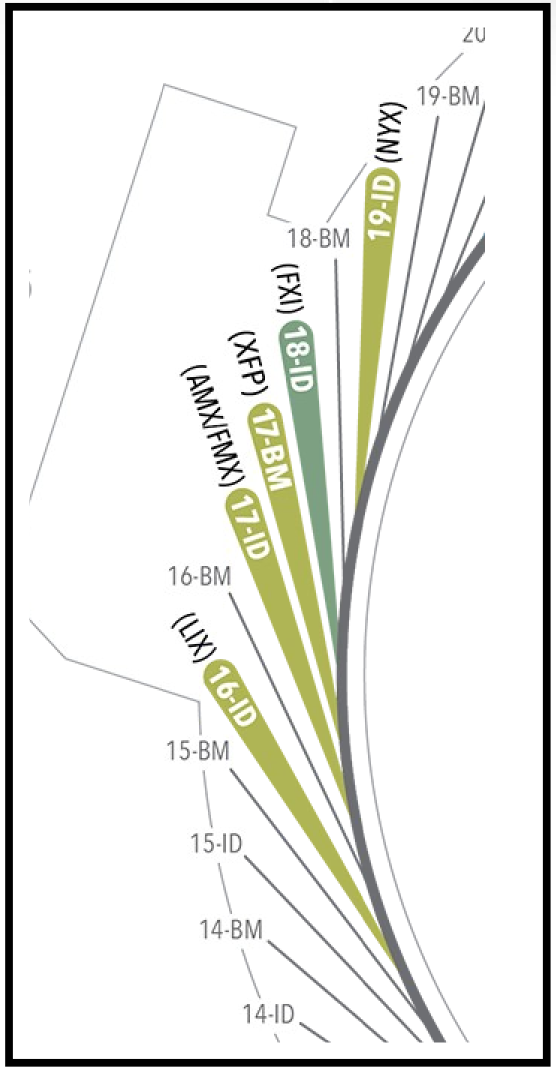
Outline

- What opportunities do we have?
- How are we seizing the opportunities?
- How are we developing the systems to meet the challenges of the future?

- New facility with small, intense beams
- Many beamlines already built with many more to come
- Biology beamlines
  - High degree of automation
  - Heterogenous beamline hardware and software ecosystems
  - Large user community
- LSDC

Bldg.	Port	Name	Contact
742	2-ID	Soft Inelastic X-ray Scattering (SIX)	Valentina Bisogni, Ext. 3163
743	3-ID	Hard X-ray Nanoprobe (HXN)	Yong Chu, Ext. 5582
743	4-BM	X-ray Fluorescence Microprobe (XFM) (XFM)	Ryan Tappero, Ext. 5245
743	4-ID	Integrated In situ and Resonant Hard X-ray Studies (ISR)	Christie Nelson, Ext. 4916
743	5-ID	Submicron Resolution X-ray Spectroscopy (SRX)	Andrew Kiss, Ext. 3569
743	6-BM	Beamline for Materials Measurement (BMM)	Bruce Ravel, Ext. 3613
743	7-BM	Quick x-ray Absorption and Scattering (QAS)	Steven Ehrlich, Ext. 7862
743	7-ID-1	Spectroscopy Soft and Tender (SST-1)	Cherno Jaye, Ext. 5958
743	7-ID-2	Spectroscopy Soft and Tender (SST-2)	Conan Weiland, Ext. 8468
743	8-BM	Tender Energy X-ray Absorption Spectroscopy (TES)	Yonghua Du, Ext. 6234
743	8-ID	Inner-Shell Spectroscopy (ISS)	Eli Stavitski, Ext. 8641
744	10-ID	Inelastic X-ray Scattering (IXS)	Yong Cai, Ext. 7154
744	11-BM	Complex Materials Scattering (CMS)	Masa Fukuto, Ext. 5256
744	11-ID	Coherent Hard X-ray Scattering (CHX)	Andrei Fluerasu, Ext. 4645
744	12-ID	Soft Matter Interfaces (SMI)	Mikhail Zhernenkov, Ext. 5158
745	16-ID	Life Science X-ray Scattering (LiX)	Lin Yang, Ext. 5833
745	17-BM	X-ray Footprinting of Biological Materials (XFP)	Erik Farquhar, Ext. 8174
745	17-ID-1	Highly Automated micro focus Macromolecular Crystallography (AMX)	Jean Jakoncic, Ext. 3930
745	17-ID-2	Frontier Microfocusing Macromolecular Crystallography (FMX)	Martin Fuchs, Ext. 8890
745	18-ID	Full Field X-ray Imaging (FXI)	Wah-Keat Lee, Ext. 4085
745	19-ID	Biological Microdiffraction Facility (NYX)	Kevin Battaile, Ext. 6331
741	21-ID-1	Electron Spectro-Microscopy (ESM-ARPES)	Elio Vescovo, Ext. 7399
741	21-ID-2	Electron Spectro-Microscopy (ESM-XPEEM)	Elio Vescovo, Ext. 7399
741	22-IR-1	Frontier Synchrotron Infrared Spectroscopy (FIS)	Larry Carr, Ext. 2237
741	22-IR-2	Magnetospectroscopy, Ellipsometry and Time-Resolved Optical Spectroscopies (MET)	Larry Carr, Ext. 2237
741	23-ID-1	Coherent Soft X-ray Scattering (CSX)	Claudio Mazzoli, Ext. 8213
741	23-ID-2	In situ and Operando Soft X-ray Spectroscopy (IOS)	Iradwikanari Waluyo, Ext. 3421
742	28-ID-1	Pair Distribution Function (PDF)	Milinda Abeykoon, Ext. 2663
742	28-ID-2	X-ray Powder Diffraction (XPD)	Sanjit Ghose, Ext. 3611
Beamlines Under Development			
744	9-ID	Coherent Diffraction Imaging (CDI)	Garth Williams, Ext. 8747
742	27-ID	High Energy Engineering X-Ray Scattering (HEX)	Michael Drakopoulos, Ext. 6222
742	29-ID-1	Soft X-ray Nanoprobe (SXN)	Andrew Walter, Ext. 4055
742	29-ID-2	NanoARPES and NanoRIXS (ARI)	Andrew Walter, Ext. 4055

- 29 beamlines in operation, 4 under development
- Storage ring: 792m circumference, 3 GeV, 400 mA operating current designed to deliver an electron beam with very small emittance (0.5 nm-rad H, 8 pm-rad V)
- One of the newest synchrotrons in the world, started operating in 2015 and has a large user community around the US and the world



BL	Min. beam size	Flux	Detector	Max fr. rate	Robotic sample changer	# samples	Data Acquisition GUI	LIMS
AMX	7x5 μm^2	4e12	Eiger 9M	200+	Y	384	LSDC	SynchWeb /ISPyB
FMX	1x1.5 μm^2	3.5e12	Eiger 16M	100+	Y	384	LSDC	SynchWeb /ISPyB
NYX	10x10 μm^2	1e12	ADSC HF-4M	25	Y	240	BluIce	None
LiX	0.4 mm^2 (scattering)		3x Pilatus		Y	54	Custom	None
XFP	0.1 x 0.4 mm^2	1.6e16	N/A		N	96	Custom	None
Cryo-EM							Vendor	None

This talk will concentrate on LSDC (Life Sciences Data Collection)

Outline

- What opportunities do we have?
- How are we seizing the opportunities?
- How are we developing the systems to meet the challenges of the future?

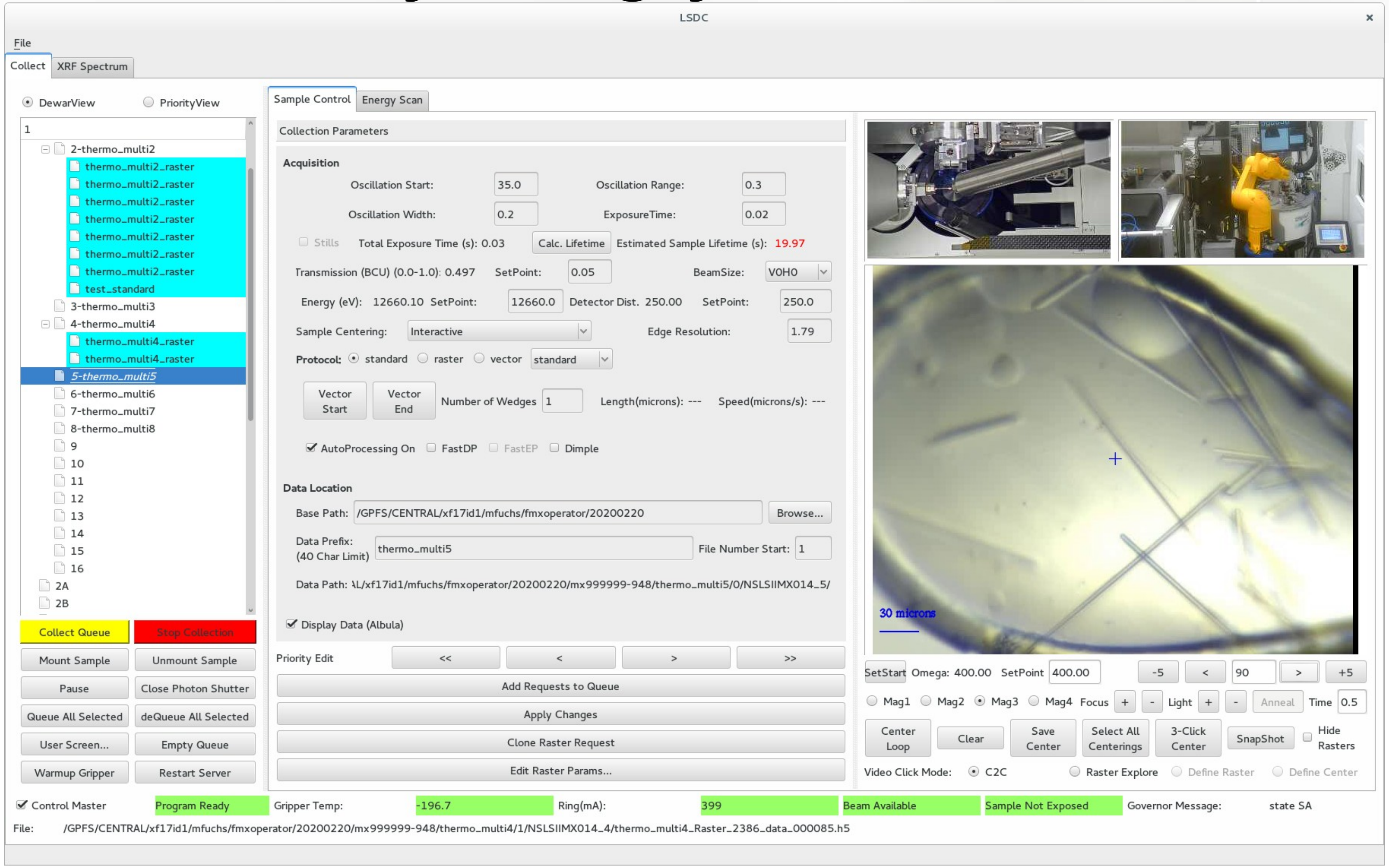
- New facility with small, intense beams
- Many beamlines already built with many more to come
- Biology beamlines
  - High degree of automation
  - Heterogenous beamline hardware and software ecosystems
  - Large user community
- LSDC

Outline

- What opportunities do we have?
- How are we seizing the opportunities?
- How are we developing the systems to meet the challenges of the future?

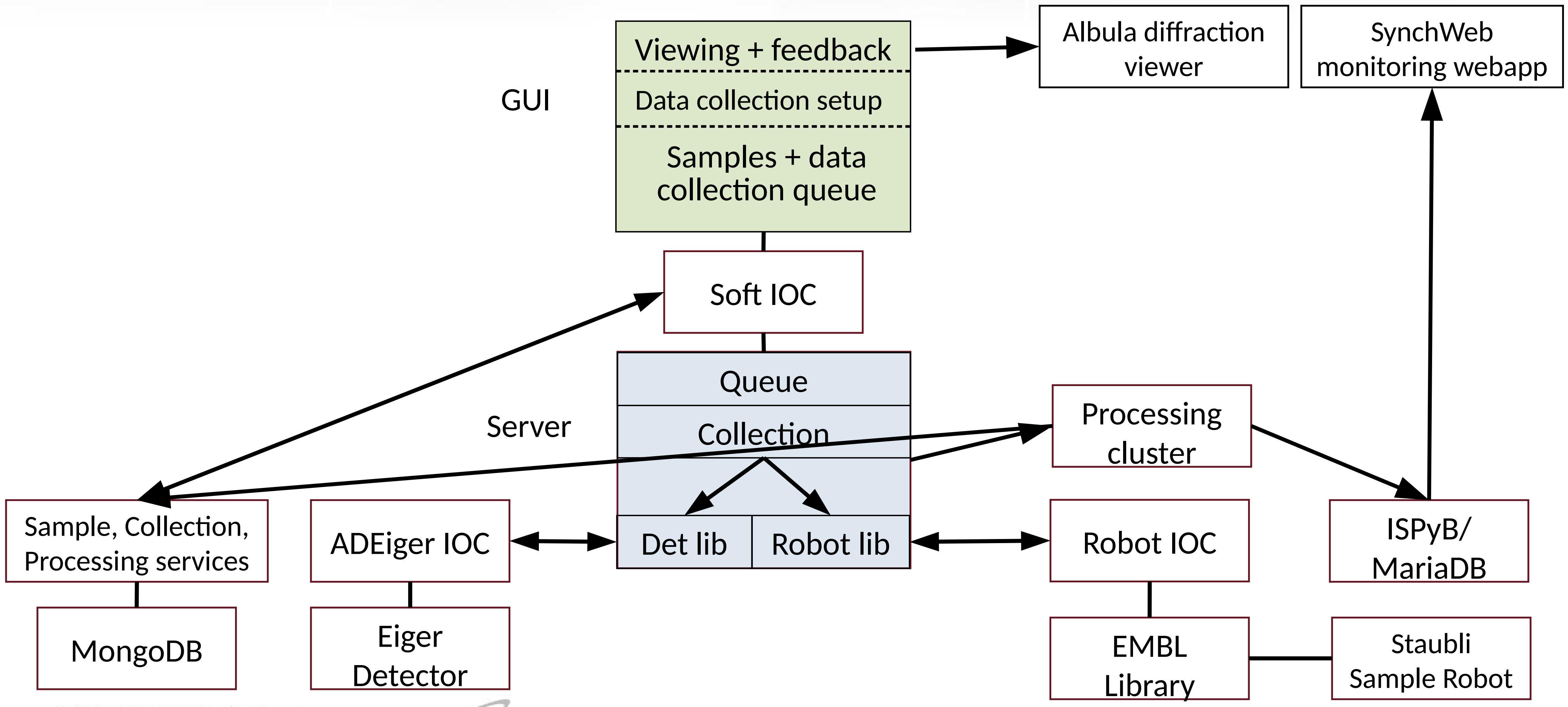
- Experiments have been implemented
- Sample exchange (as we have seen in Edwin Lazo's talk) is reliable
- Remote experiments are running – including automated data collection
- Processing systems effectively providing feedback
- LIMS system integrated

Everything you need to collect remotely



- Layout inspired by MXCube
- Sample viewing/ centering/ rastering feedback area
- Parameters area
- Sample handling and data collection history area

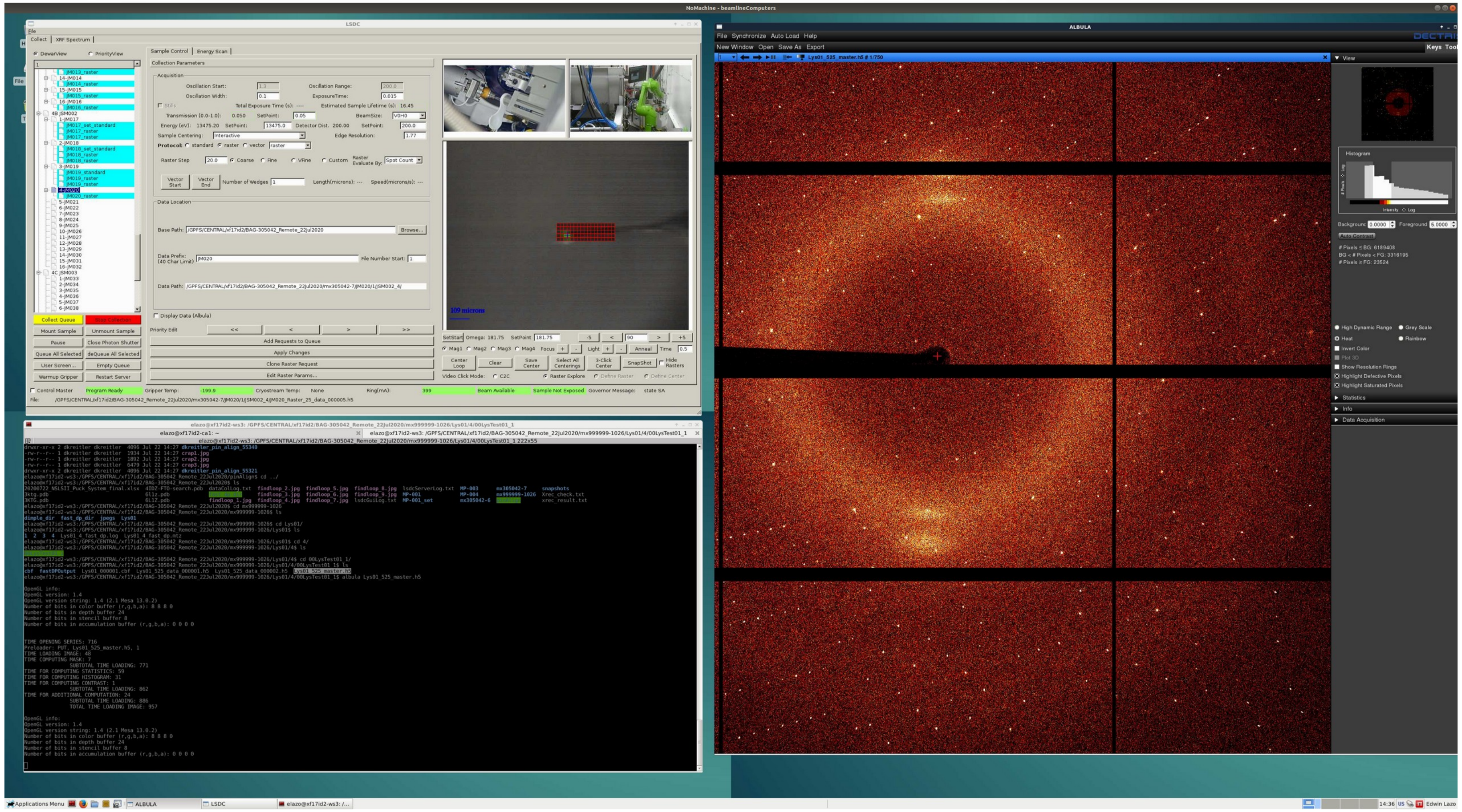
LSDC and related components



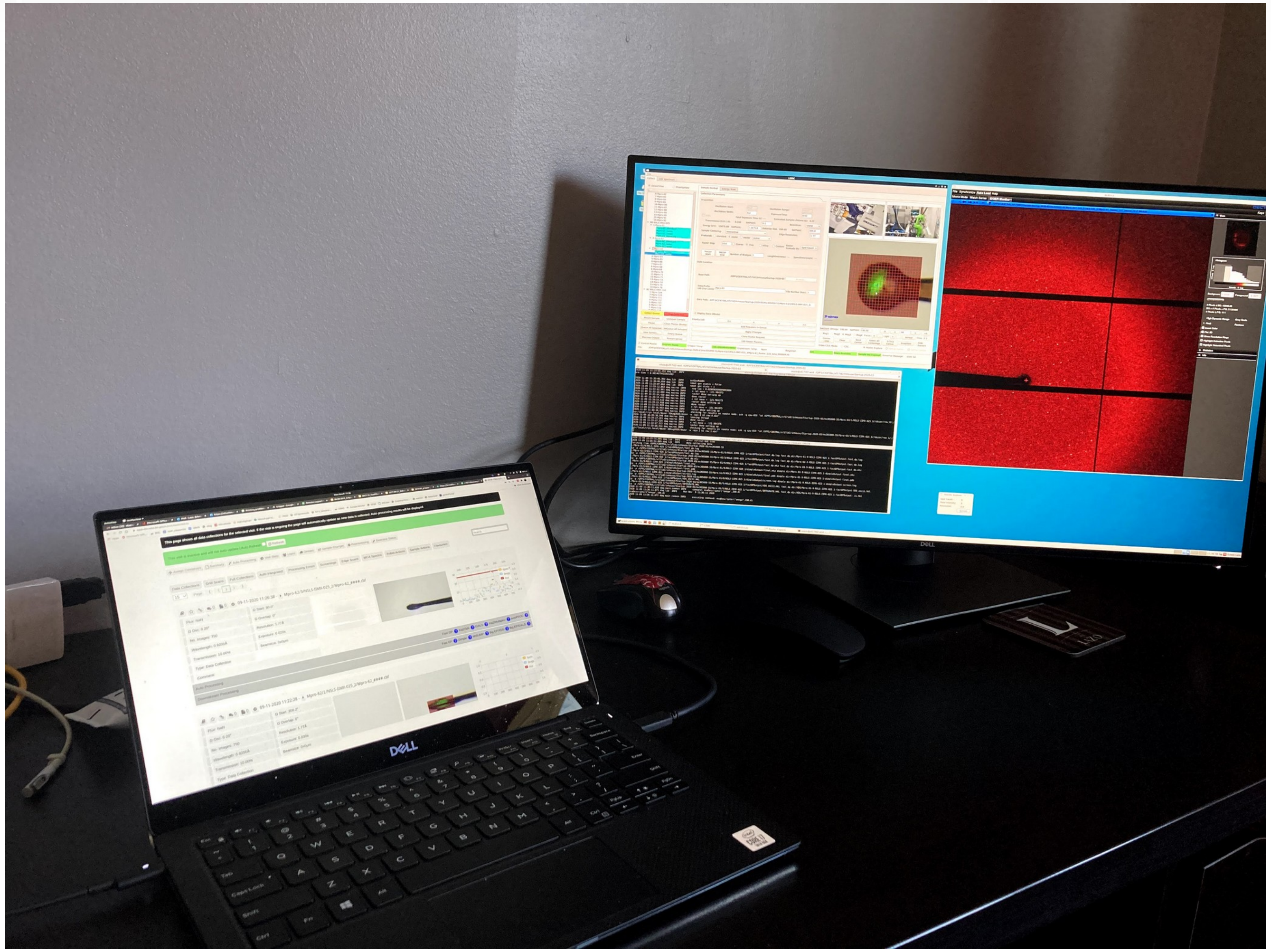
LSDC: current state at a glance

- Currently used at AMX and FMX
- Python (PyQt) GUI client
  - Configure all typical experiments (standard oscillation, diffraction and fluorescence-based rastering, vector/helical, energy scan)
  - GUI-driven sample exchange
  - Rastering heatmaps
  - Albula (Dectris) for diffraction image visualization
- Python server
  - Perform the queued experiments configured by the user
  - Using the libraries below
  - Communication with databases
  - Hardware configuration management
  - Initiating processing
    - FastDP, Xia2
    - Dimple
    - Chooch, Raddose, Dozor, DIALS Spotfinder, etc…
  - Libraries modularizing PyEpics calls to interact with PVs of the EPICS Control System
    - Detector, robot, general PVs
- On both client and server, many custom components including sample centering/rastering definition/raster feedback view, queuing system

LSDC GUI and Albula (Dectris) on NX



LSDC GUI + Albula, Synchweb



What challenges do data acquisition systems face?

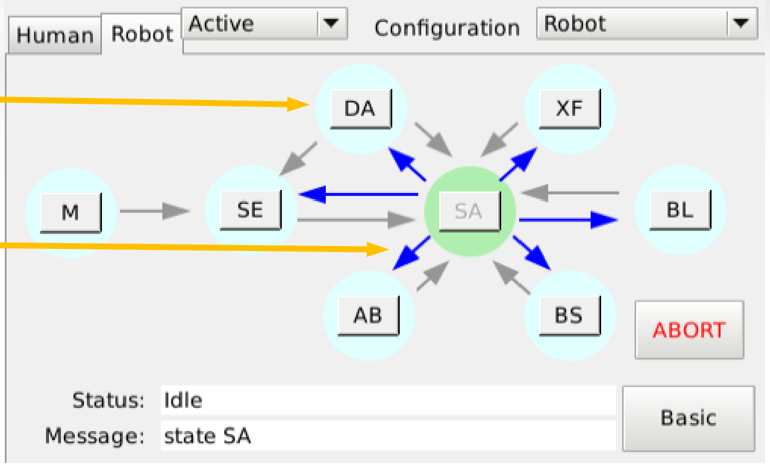
- Evolving infrastructure
- Faster hardware
- Software improvements
- New techniques
- New beamlines
- Knowledge transfer among developers
- Integration within own facility
- Integration with outside beamlines

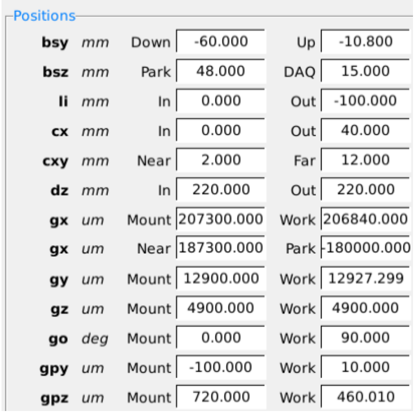
LSDC Core Modules

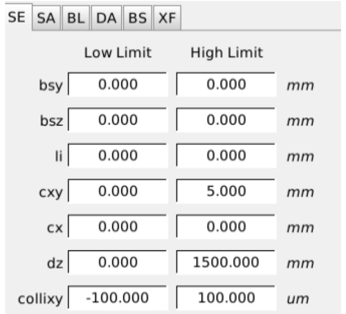
- State machine for safe experiment state transitions – the Governor
- X-ray fluorescence spectrum using standard libraries (Ophyd, Bluesky, Matplotlib, and PyMca)
- Ophyd/Bluesky for MX experiments (in progress)
- Sample, data collection, data processing information storage via web services
- ISPyB-API library (Diamond Light Source) for ISPyB interaction
- Interaction with Albula (Dectris) for diffraction visualization
- EPICS PV interaction code modularized into libraries

The Governor manages Experiment States

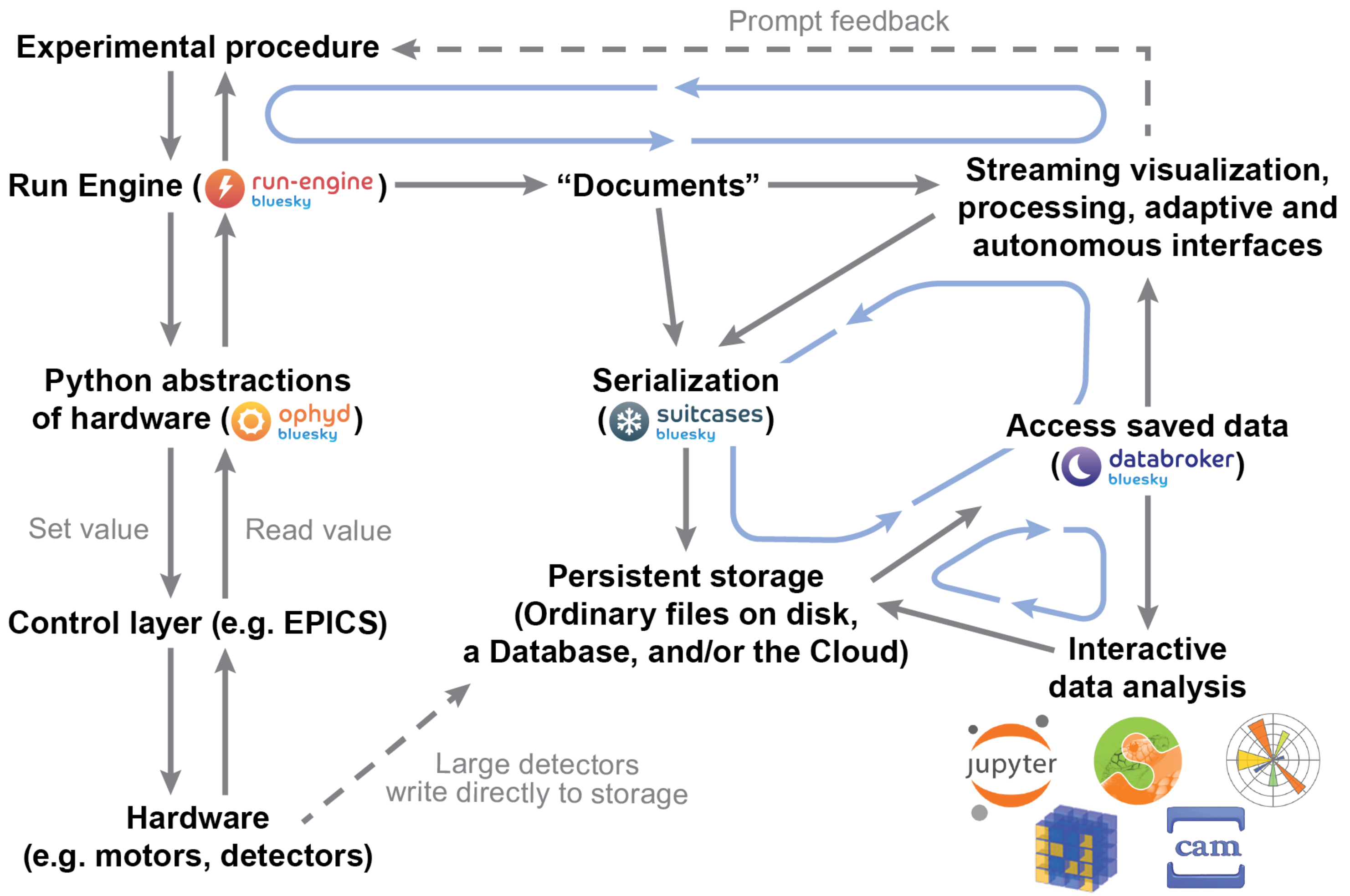
- Define **experiment states** with default positions
- Define **transitions** between states
- Define multiple **positions** per positioner
  - Motors (Near, Far, Park, …)
  - Valves (Open, Close)
- Define allowed **ranges** within states
- Test transitions thoroughly once
- From now on, just call transitions to cycle through experiment states







Working with Bluesky



- Developed at NSLS-II and used at most beamlines
- Used at 5 US DOE Light Sources and the Australian Synchrotron
- Modularized data acquisition system
  - Control system abstraction layer
  - Procedure running code
  - Access layer to data
- http://blueskyproject.io

Outline

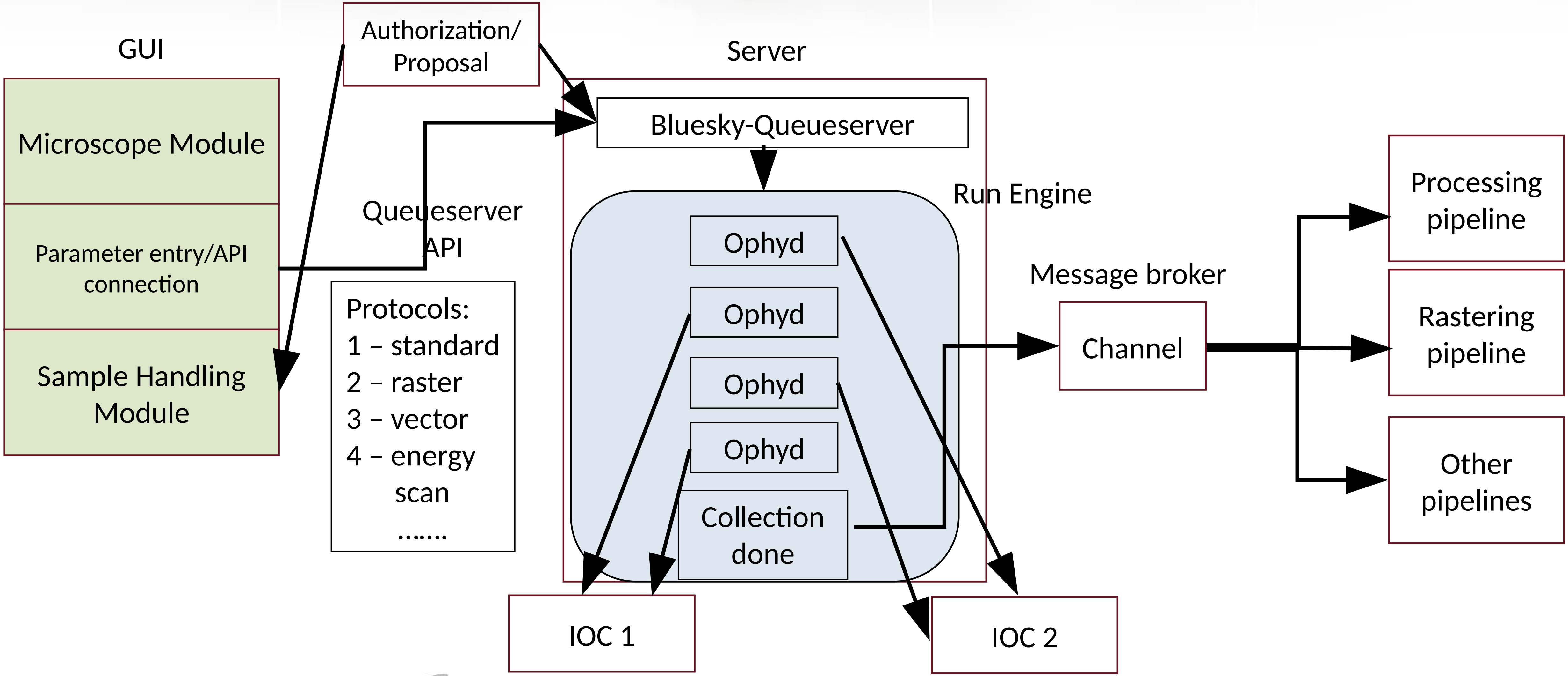
- What opportunities do we have?
- How are we seizing the opportunities?
- How are we developing the systems to meet the challenges of the future?

- Experiments have been implemented
- Sample exchange (as we have seen in Edwin Lazo’s talk) is reliable
- Remote experiments are running – including automated data collection
- Processing systems effectively providing feedback
- LIMS system integrated
- Modularization well under way

Outline

- What opportunities do we have?
- How are we seizing the opportunities?
- How are we developing the systems to meet the challenges of the future?
  - Hardware abstraction to enable new beamlines (NYX in progress, more in the future?)
  - New experiments
  - Minimize disruptions due to infrastructure changes
  - Rapid development of processing systems
  - Continued modularization opportunities to incorporate/supply to other beamlines and facilities
  - Improved infrastructure for better user and staff experience

Proposed future LSDC system - changes



How LSDC is being prepared for long-term development

- Improving the core scanning code with Ophyd/Bluesky
  - Hardware abstraction layer
  - Offline testing
  - Better modularization
- Replace custom collection queueing code with Bluesky-Queueserver
- Modularization of GUI code
  - Specify a new "microscope" module that can be used at many beamlines to view samples and control experiments
- Software engineering improvements
- Improved processing systems triggered via Kafka message broker
  - More flexible, less hard-coded processing – adding multi-crystal, multi-dataset (from Dale's talk)
  - Reprocessing

LSDC is on Github!
https://github.com/NSLS-II/lsdc

https://github.com/NSLS-II/lsdc

sanitize_sheet.py	add tests, fix sanitize_sheet to make things testable	13 months ago
start_bs.py	Revert "Add vector program"	3 months ago
sweepDump.py	remove commented-out code	13 months ago
test_sanitize_sheet.py	add tests, fix sanitize_sheet to make things testable	13 months ago
wrapGui.py	add docstrings for the main programs	12 months ago

README.md

LSDC - Life Sciences Data Collection

Author: John Skinner, formerly of National Synchrotron Light Source II, Brookhaven National Laboratory, NY

Currently used at the NSLS-II MX beamlines - AMX, FMX

Offline fork of https://github.com/jskinner53/lsdc

Needs more documentation as only a module diagram currently exists.

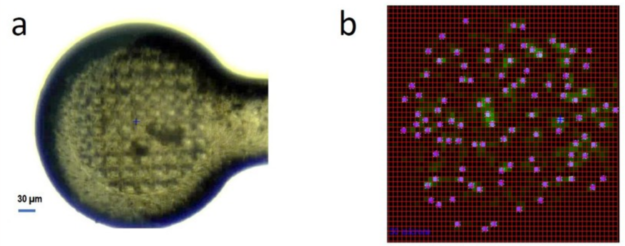
- Future
  - Continuous Integration
    - Unit testing
  - pip installation
  - Docker

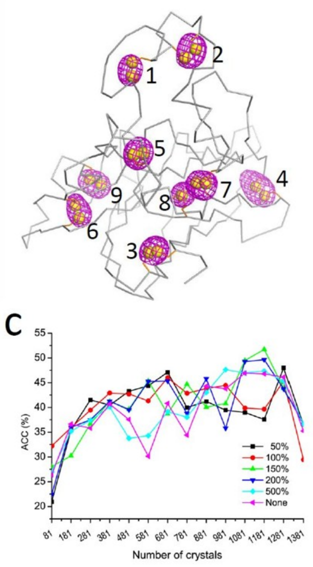
Common systems being developed

- NSLS-II as a whole
  - Guacamole to complement NX
  - Single sign-on
  - Direct access to web applications once signed in
  - Better integration of proposal system
- LSDC modules
  - Bluesky queueing system
  - Bluesky GUI experiment configuration module
  - Microscope module
  - Sample management module
  - Externalize processing with Kafka message broker system (Dale's talk from yesterday)

PyMDA: microcrystal data assembly using Python – Multi-crystal crystallography and Sulfur-SAD at 5 keV

➢ G. Guo, P. Zhu, Q. Liu (BNL Biology Dept)





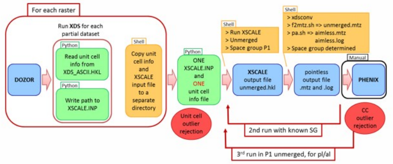
- Data collection at 5 keV (λ= 2.48 Å) from 5-10 µm thaumatin crystals
- Strategy of data analysis
  - Initial data assembly based on CC1/2
  - Crystal and frame rejections based on SmRmerge

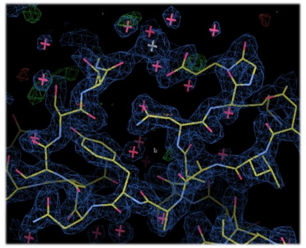
Takemaru *et al.*, *J. Appl. Crystallogr.* **53**, 277 (2020)
Guo *et al.*, *IUCrJ* **6**, 532 (2019)
Guo *et al.*, *IUCrJ* **5**, 238 (2018)

Ultrafast raster-scanning serial crystallography data processing

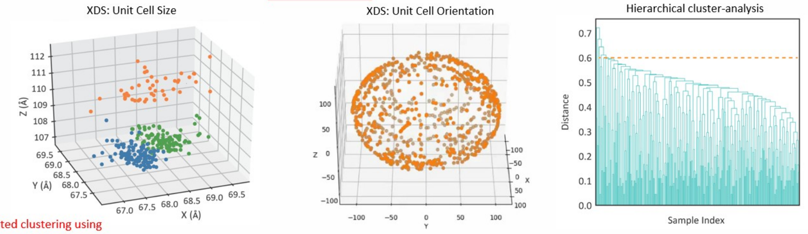
WYpeline

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





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



Automated clustering using Machine Learning algorithm

Gao, Y. *et al.*, *J. Synchrotron Rad.*, (2018), **25**, 1362-1370

Outline

- What opportunities do we have?
- How are we seizing the opportunities?
- How are we developing the systems to meet the challenges of the future?

- Hardware abstraction to enable new beamlines (NYX in progress, more in the future?)
- New experiments
- Minimize disruptions due to infrastructure changes
- Rapid development of processing systems
- Continued modularization opportunities to incorporate/supply to other beamlines and facilities
- Improved infrastructure for better user and staff experience

What challenges do data acquisition systems face?

- Evolving infrastructure
- Faster hardware
- Software improvements
- New techniques
- New beamlines
- Knowledge transfer among developers
- Integration within own facility
- Integration with outside beamlines

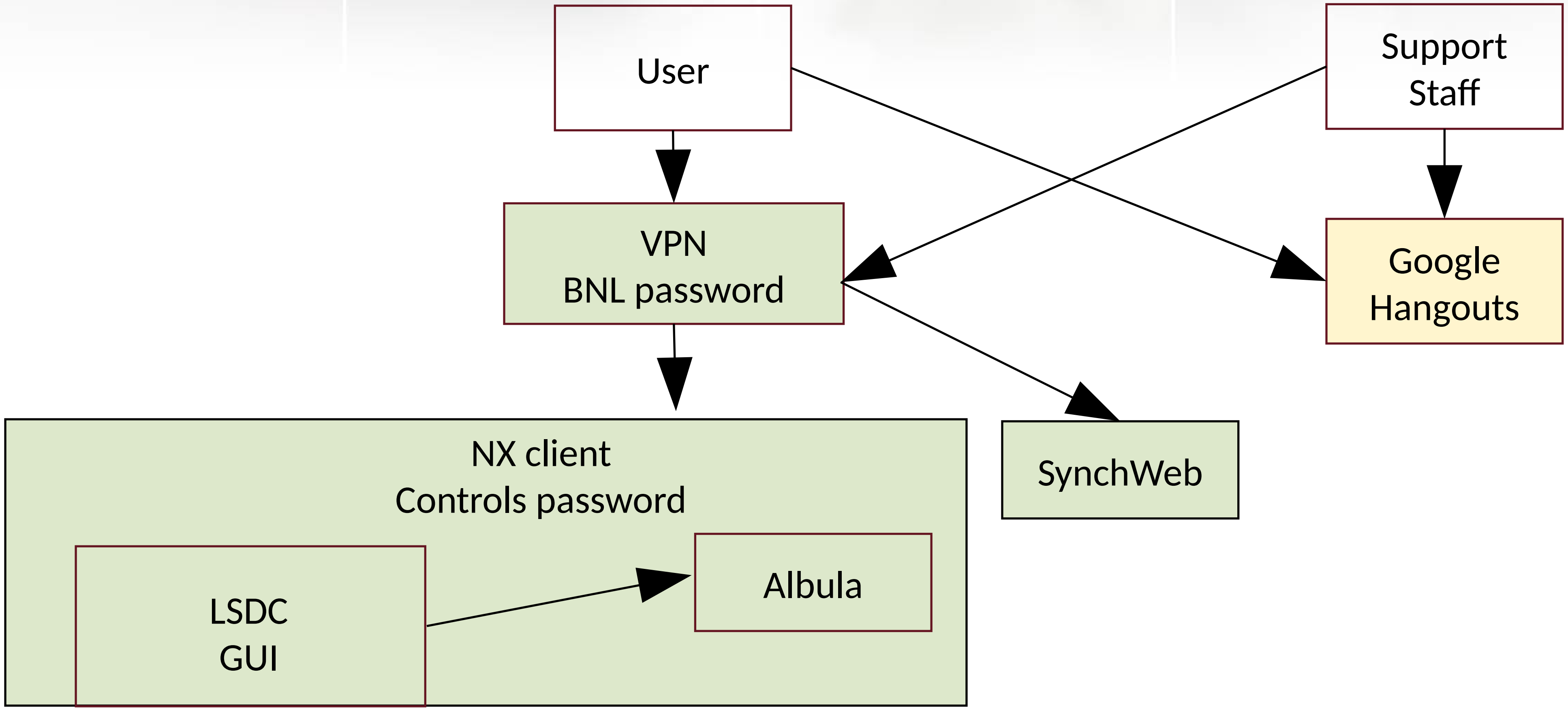
The future is bright for the NSLS-II beamlines!

- More common infrastructure systems at NSLS-II
  - Good for users
  - Good for staff!
- Development of new cross-beamline software projects
- Bluesky as both a cross-beamline and cross-facility project
- LSDC as a framework for GUI-based data acquisition
- All while keeping the existing users happy!

Thanks!

- John Skinner and Bob Sweet for their long-term work on LSDC and its predecessors
- Beamline staff – Stu Myers, Martin Fuchs, Jean Jakoncic, Dale Kreitler, Edwin Lazo in particular
- Data Science and Systems Integration group
  - Tom Caswell, Dan Allan – Bluesky originators
  - Marcus Hanwell – microscope module
  - Maksim Rakitin – CI, packaging, and Ophyd/Bluesky integration
- SynchWeb, ispyb-api developers (Diamond Light Source)
- Albula library (Dectris)

Current MX beamline situation – what the user sees



Future MX beamline User View