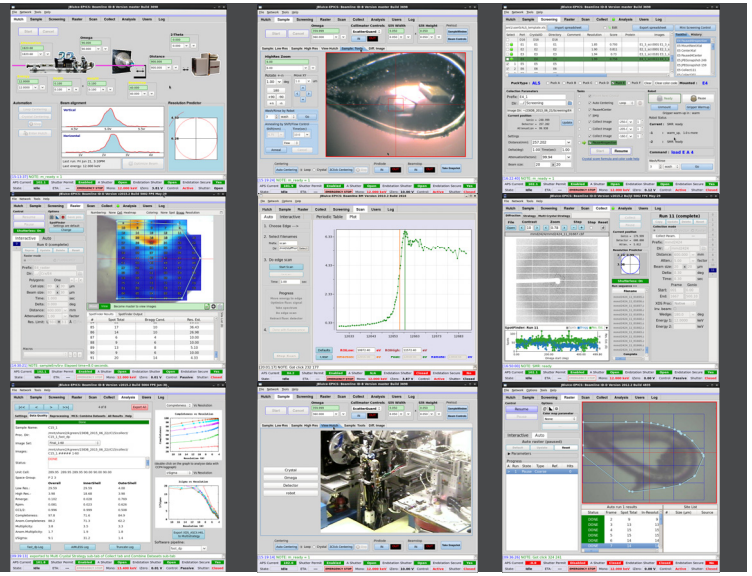
BEAMLINE AND EXPERIMENT AUTOMATIONS FOR THE GENERAL MEDICAL SCIENCES AND CANCER INSTITUTES STRUCTURAL BIOLOGY FACILITY AT THE ADVANCED PHOTON SOURCE (GM/CA@APS)



SERGEY STEPANOV

Argonne National Laboratory,
Advanced Photon Source,
Lemont, IL, USA

October 2017, ICALEPCS-2017, Barcelona, Spain

Coauthors





Oleg Makarov[1], **Mark Hilgart**[1], **Qingping Xu**[1], **Robert Fischetti**[1], **Janet Smith**[2]

[1] Advanced Photon Source, Argonne National Laboratory
[2] Life Sciences Institute, University of Michigan

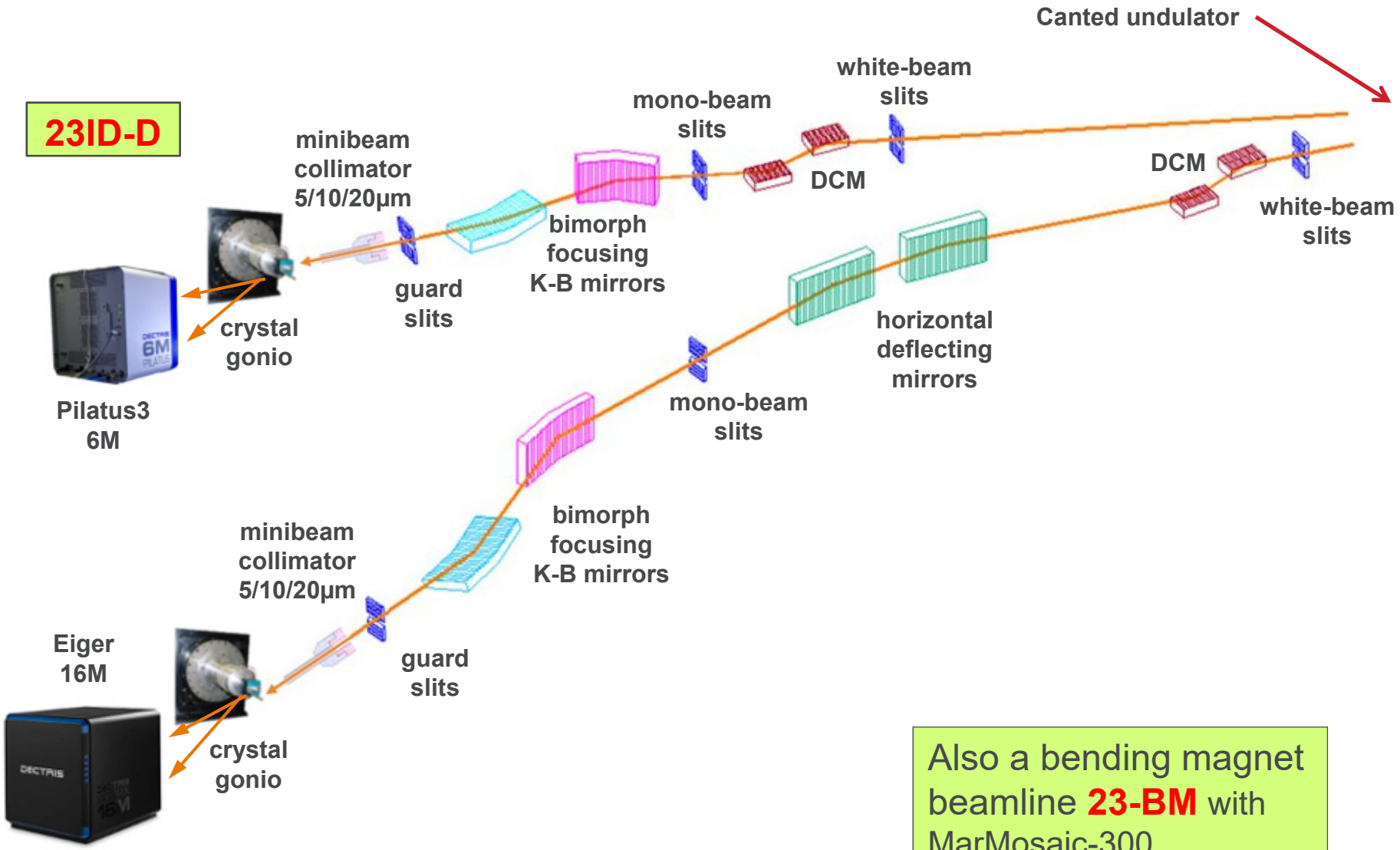
Acknowledgements

All staff of General Medicine and Cancer Institutes Structural Biology Facility at the Advanced Photon Source (GM/CA@APS), Argonne National Laboratory

GM/CA@APS has been funded in whole or in part with Federal funds from the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. The Eiger 16M detector was funded by an NIH–Office of Research Infrastructure Programs, High-End Instrumentation Grant (1S10OD012289-01A1).

Argonne NATIONAL LABORATORY

GM/CA@APS: macromolecular crystallography beamlines

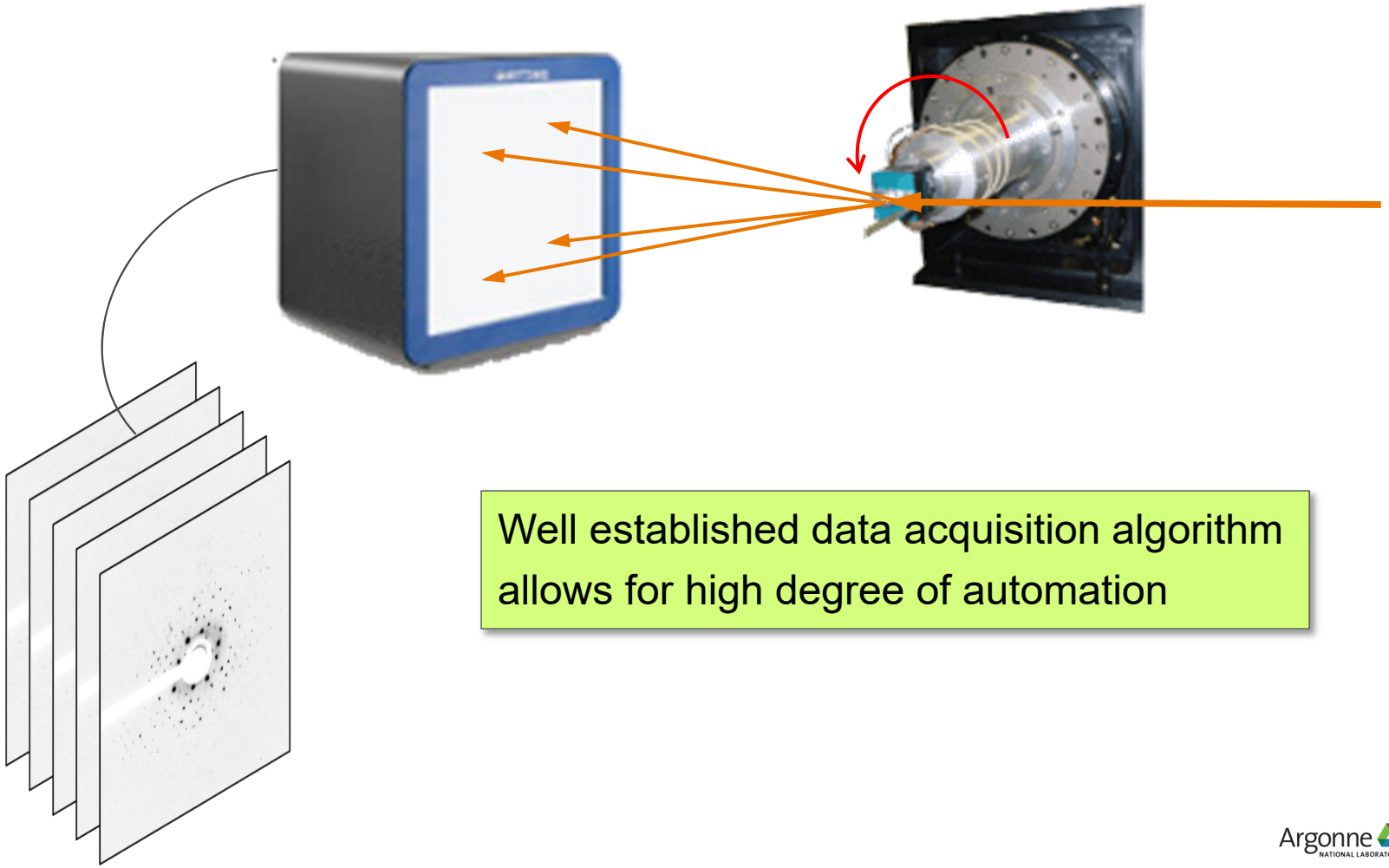


Also a bending magnet beamline **23-BM** with MarMosaic-300

Argonne NATIONAL LABORATORY

MX experiment basics

MarMosaic-300,
Pilatus3-6M,
Eiger-16M



Well established data acquisition algorithm allows for high degree of automation

GMCA@APS: JBluIce software for recording MX data

JBluIce is an advanced EPICS client relying on multiple EPICS servers for complex operations (data collection, automounter, energy scanning…)

Open source at: https://www.gmca.aps.anl.gov/jbluice-epics/

Argonne NATIONAL LABORATORY

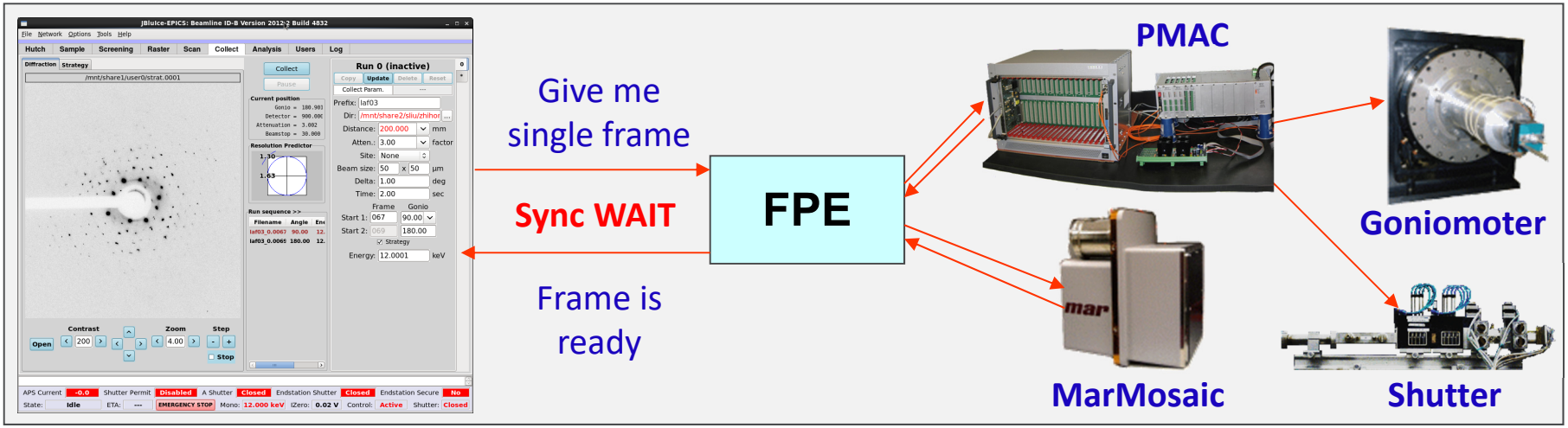
Game change: arrival of fast area detectors

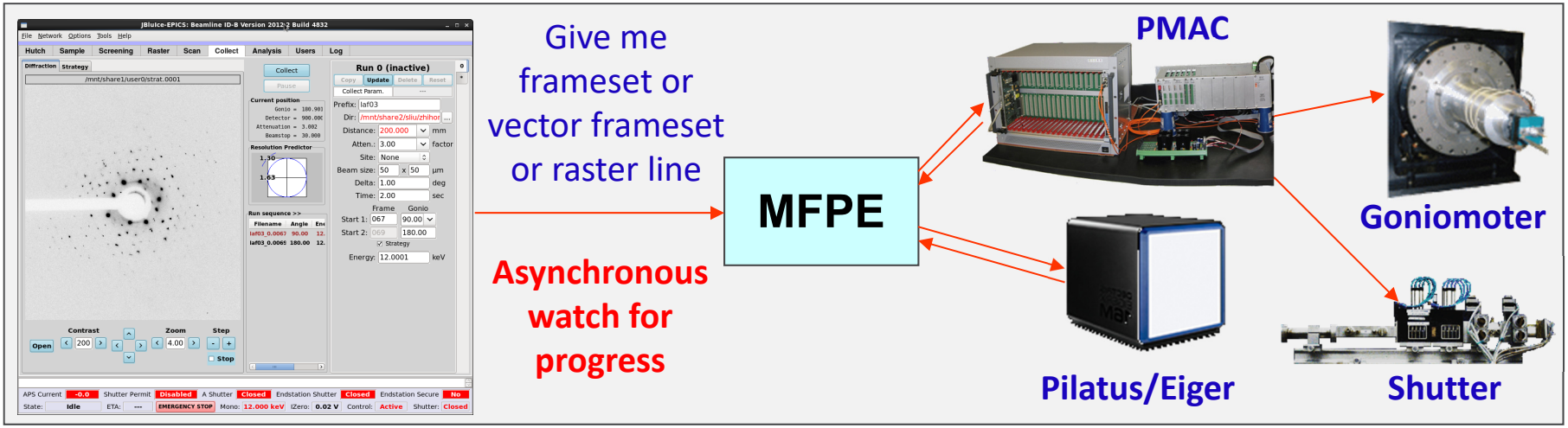
In the past two years GMCA upgraded from two MarMosaic 300 area detectors to Pilatus3 6M and Eiger 16M

Detector	MarMosaic 300	Pilatus3 6M	Eiger 16M
Readout time	2000 ms	0.95 ms	0.003 ms
Shutterless collection	no	yes	yes
Frame rate	0.2 Hz	100 Hz	133 Hz
Megapixels	16M (2x2 binning) 8192 x 8192	6M 2463 x 2527	17M 4150 x 4371
Peak data rates	6.4 MB/s	900 MB/s	2400 MB/s
Typical daily data rates	0.2 TB /day tiff	1 TB /day cbf	4 TB /day hdf5 or cbf
Auto data processing	desired	must	must

Control system needed to be adapted to qualitatively new synchronization requirements between detector/goniometer/shutter, higher data rates & volumes and faster data processing

Gonio-detector synchronization for shutterless data collection: delegating controls to lower level

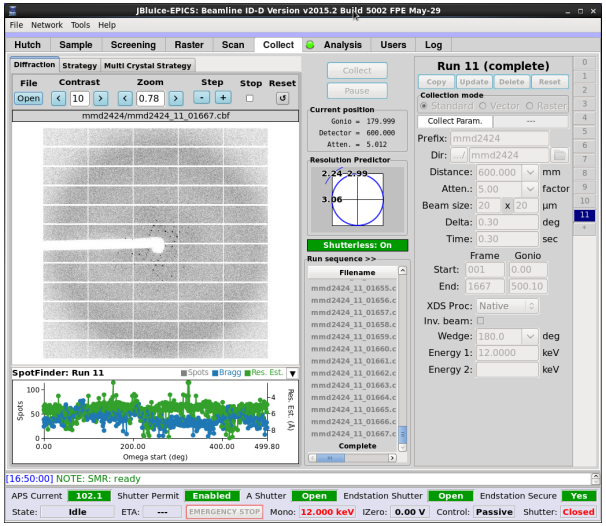


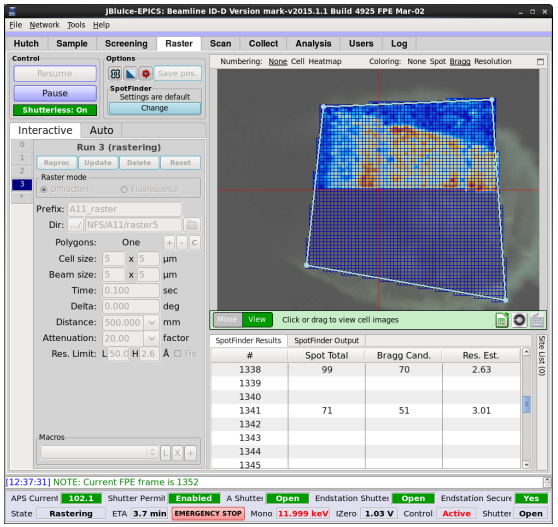


Previously JBluIce was requesting single frame from Frame Processing Engine (FPE). **Now** series of frames are processed by Multi-Frame processing Engine (MFPE). FPE is a State Notation program running in EPICS soft IOC. It re-programs PMAC to sync goniometer, detector and shutter. *Experiment is driven by PMAC* .

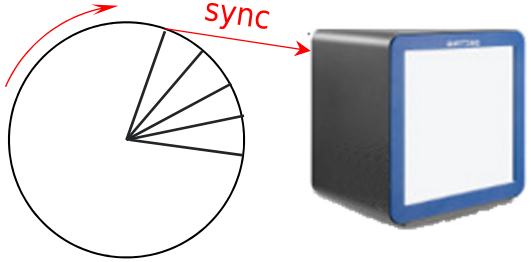
Argonne NATIONAL LABORATORY

Algorithms of shutterless data collection and rastering

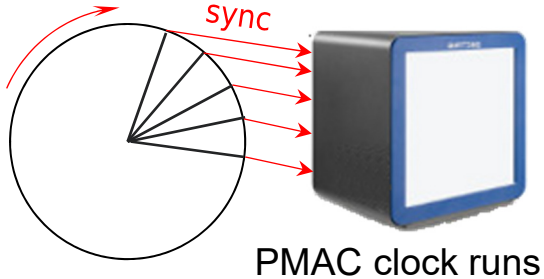




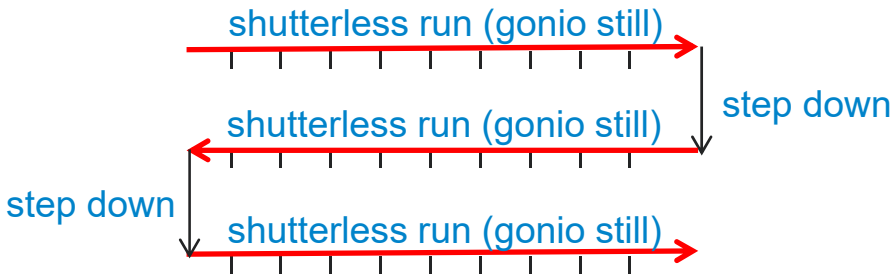
Traditional shutterless data collection



Truly synchronous shutterless collection with external detector advance



Finding promising diffraction spots with rastering:

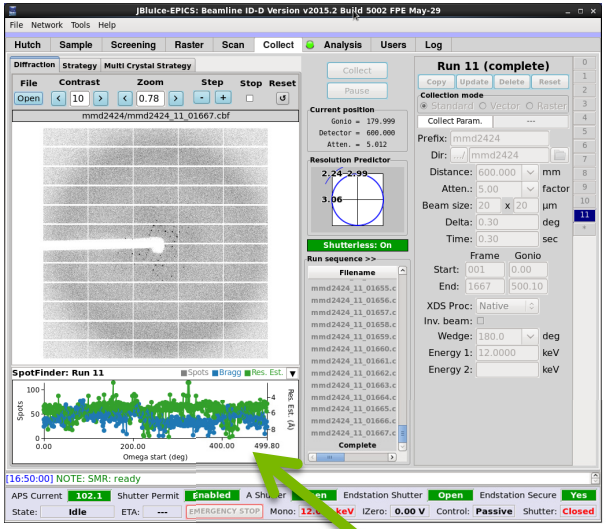


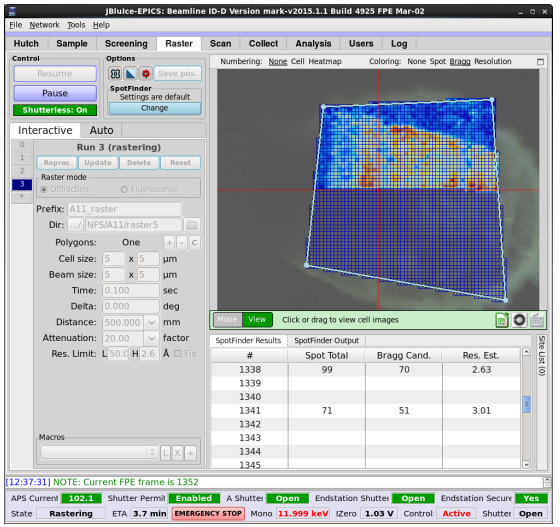
Using external detector advance allows for single-arm detector operation when collecting the whole grid.

PMAC clock runs at 2 kHz; enough to sync 100Hz data collection

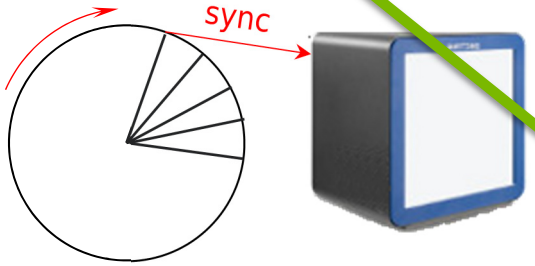
Argonne NATIONAL LABORATORY

Algorithms of shutterless data collection and rastering

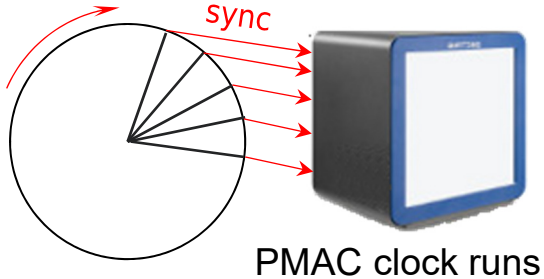




Traditional shutterless data collection



Truly synchronous shutterless collection with external detector advance



Spotfinder graph to complement image viewing: users cannot view all frames @ 100 Hz !

Finding promising diffraction

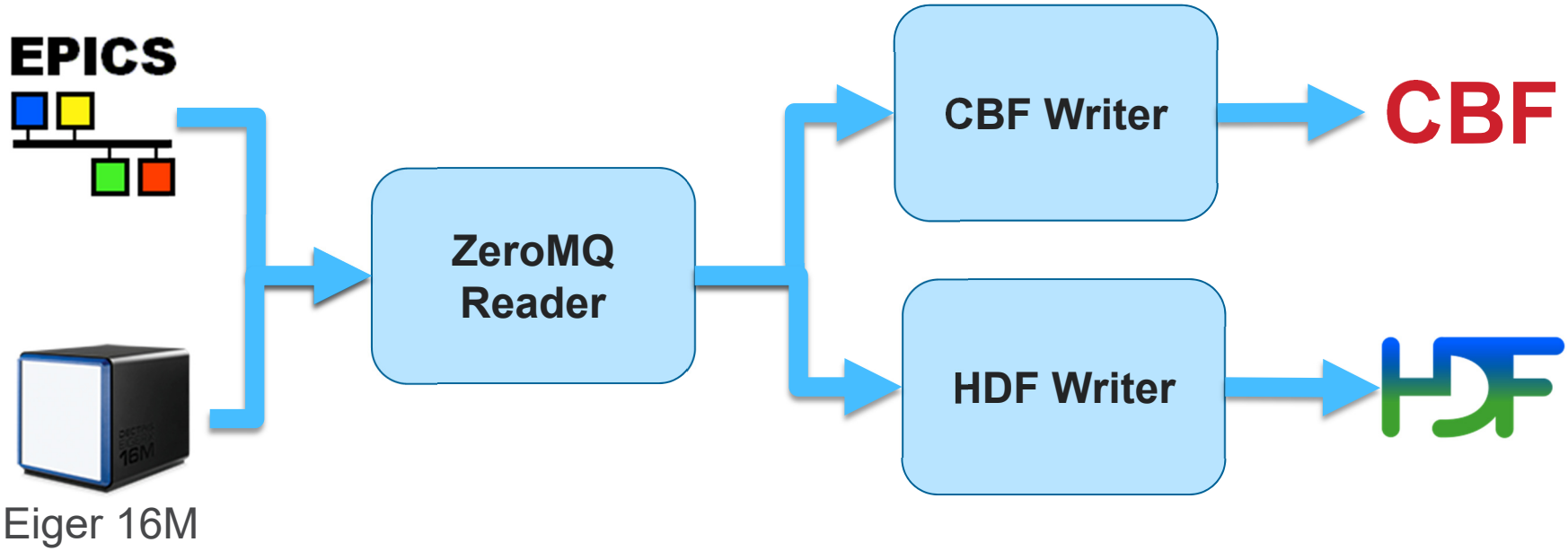
step down

allows for single-arm detector operation when collecting the whole grid.

PMAC clock runs at 2 kHz; enough to sync 100Hz data collection

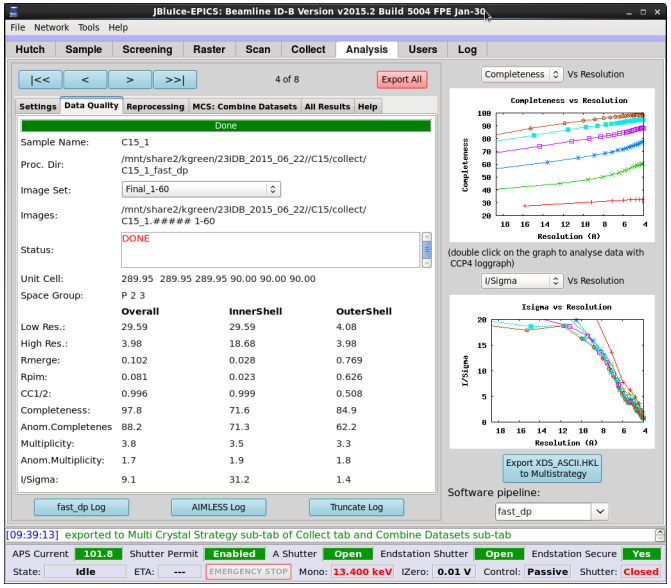
Argonne NATIONAL LABORATORY

StreamWriter for Eiger-16M detector



- Reads data stream via ZeroMQ protocol and writes CBF and/or HDF files to shared BeeGFS storage
- EPICS is used to read filename info and write the on-disk image counter
- Written in C++
- Tested at up to 100 fps with Lysozyme

Automatic data processing



Automatic data processing is the must for efficient use of fast detectors because data collection overruns human brain.

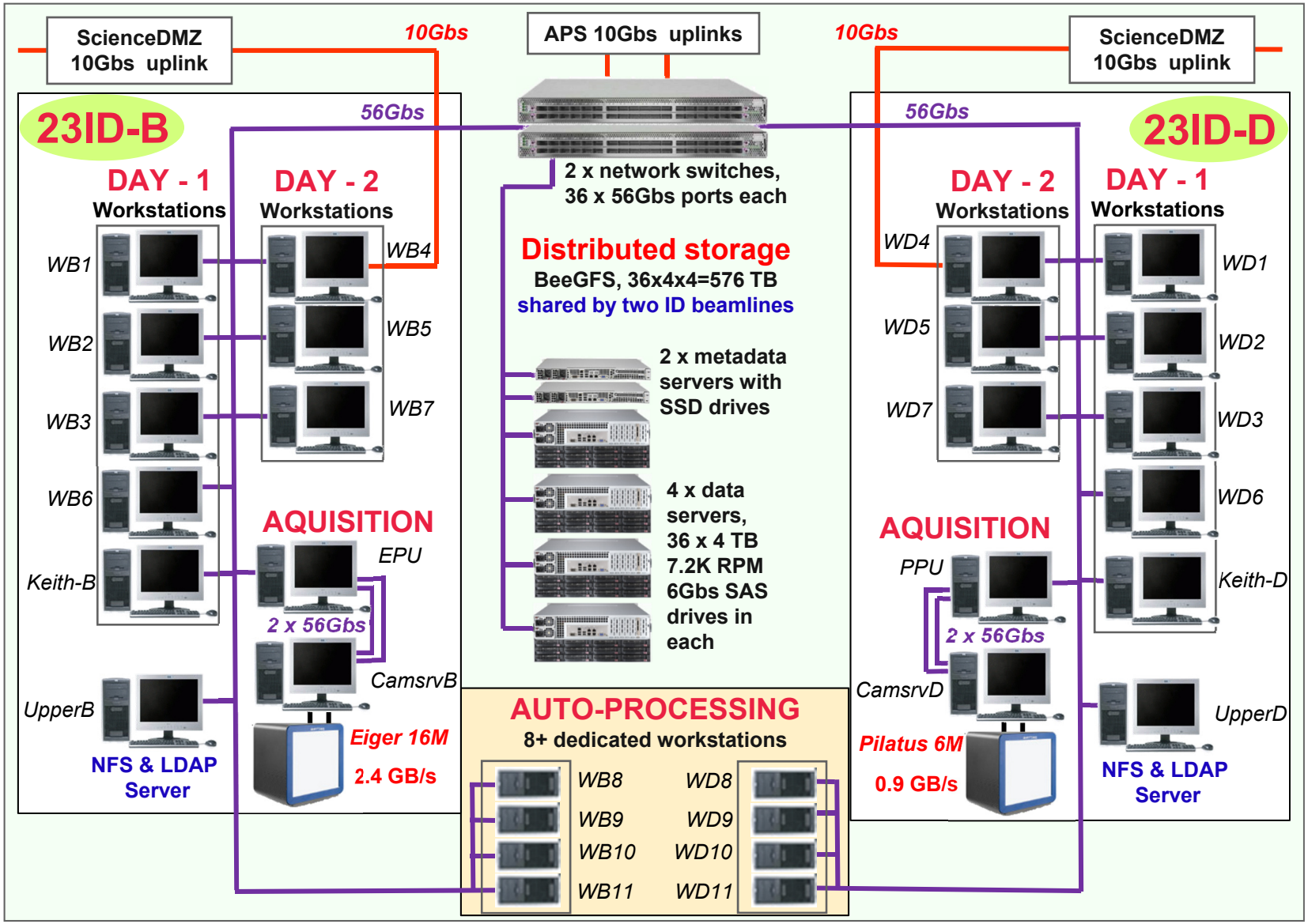
Users need to see the results as soon as the dataset is taken in order to plan next data collection step.

- A python-based data processing pipeline with option to choose one or several available engines (gmcaProc, DIALS, fast_dp, autoPROC and their combinations).
- Runs on a cluster with more than 270 CPU cores dedicated to data processing.
- Uses Son of Grid Engine (https://arc.liv.ac.uk/trac/SGE) for distributing data processing load between the cluster nodes.
- More efficient than uploading data to Argonne Leadership Computing Facility

Network and storage upgrade to high data rates & volumes

	Before	After
Network speed	10 Gbps	56 Gbps
Storage size	57 TB x 2 beamlines	576 TB merged
Storage type	Shared, GFS2	Distributed, BeeGFS
Storage link	8 Gbps SAN	56 Gbps Net
Auto processing	1 workstation/beamline	8 workstations shared
Regular Globus GridFTP	Yes	Yes
ScienceDMZ Globus GridFTP	No	Yes

GMCA@APS network/SAN structure, September 2016



Two more, WB12 & WD12
just added (not shown)

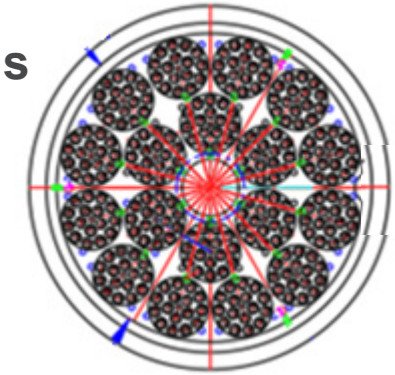
Argonne NATIONAL LABORATORY

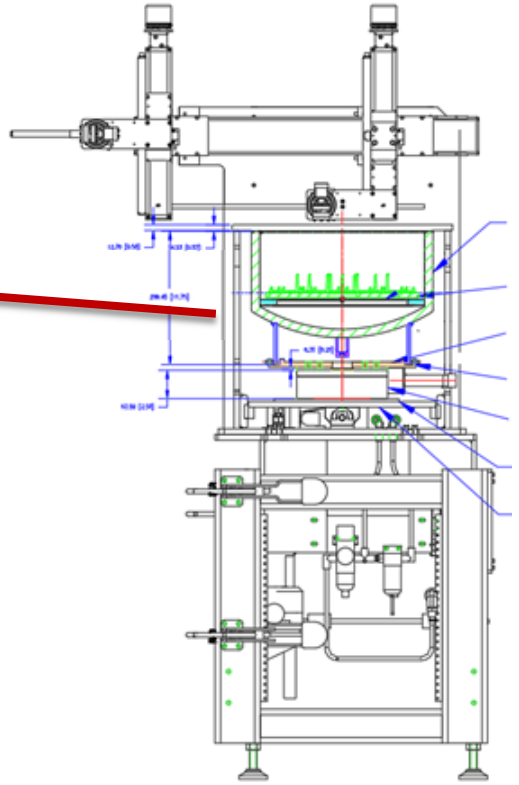
GMCA Remote Access

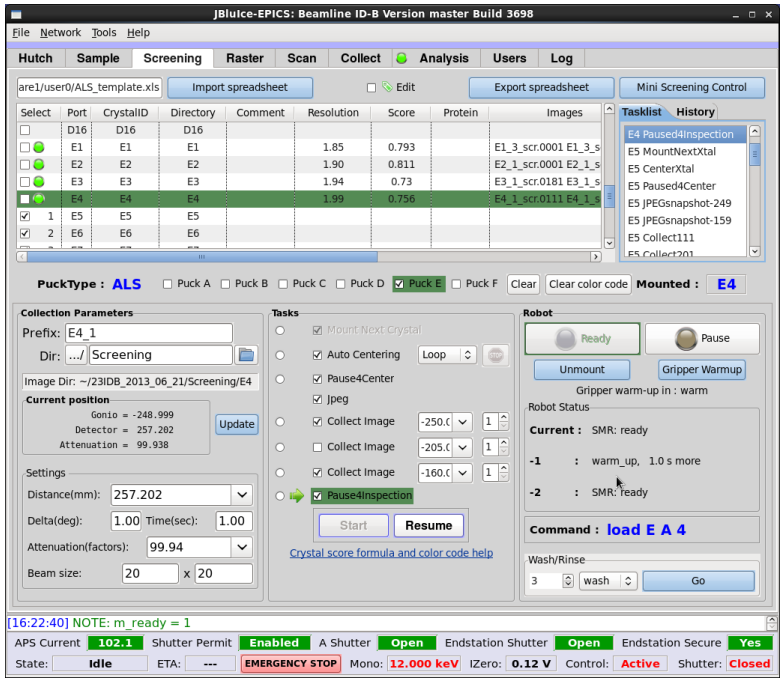
- GMCA now offers Globus Servers at each beamline for fast data transfers to home institutions.
- ScienceDMZ for faster Globus downloads.
- Remote access is provided via a cluster of 7 NOMACHINE servers coordinated by a NOMACHINE Cloud Server. Users can login either with a locally installed NX player or via a regular web browser with no client-side installation. Once logged in, remote users have same Desktop experience as those who choose to travel to GMCA. Therefore, the share of traveling users is permanently shrinking: currently about **80%** experiments are remote.
- The success of remote program is largely due to automounters: **98%** of experiments deploy automounters. Installation of Pilatus and Eiger detectors set demand on increasing automounter Dewars capacity.

Cartesian Automounter with high 288 samples capacity

18 pucks * 16 pins/puck = **288 pins**



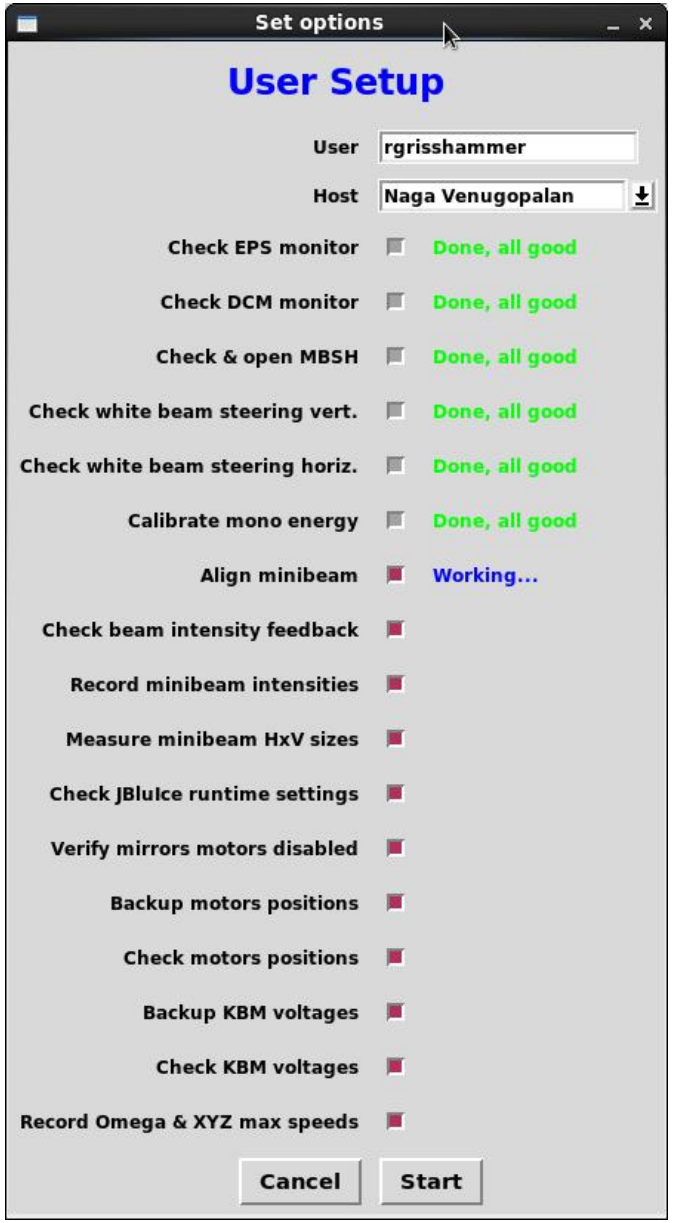




- Automounter is controlled by an EPICS state notation program (EPICS server)
- It is interfaced from the JBluIce Screening tab
- JBluIce needs to send only a few commands to the server, basically **mount** and **unmount**.

Last but not least: beamline automations

- New fast detectors are capable of producing tons of data. With data rates of up to 2.4 GB/s the current daily rates of 4 TB show that there is still room to increase beamline productivity.
- Scheduling ½ **day shifts** or even ¼ shifts is another direction we are moving towards.
- Also, the more automated are any beamline operations, the more time users can have to collect data. We have developed **automatic beamline setup** software which also provides quality assurance, for example checks the beam intensity and beam size at the sample against the reference values.



Conclusions

Upgrading beamlines to fast detectors goes far beyond the task of interfacing the detector API and synchronizing detectors with other beamline equipment.

One should consider a comprehensive approach to improve all parts of the experiment:

- handling high data rates with proper network/SAN architecture,
- storing large data volumes,
- efficiently copying data to users institutions with Globus & ScienceDMZ,
- displaying and auto-processing data at up to 100 Hz,
- providing efficient remote access
- increasing the capacity of automounters
- improving beamline automations to shorten beamline setup time and thus raise the share of beamtime when the detectors are actually in use.