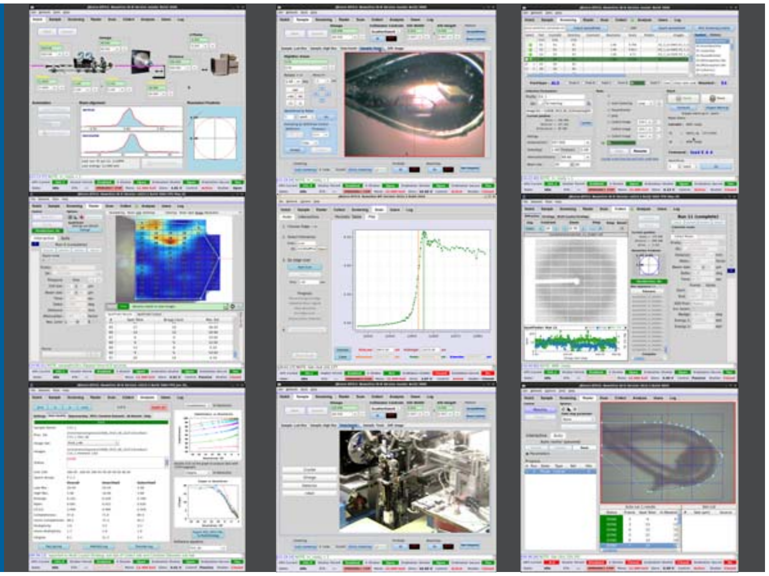
INTEGRATION OF FAST DETECTORS INTO BEAMLINE CONTROLS AT THE GM/CA MACROMOLECULAR CRYSTALLOGRAPHY BEAMLINES AT THE ADVANCED PHOTON SOURCE

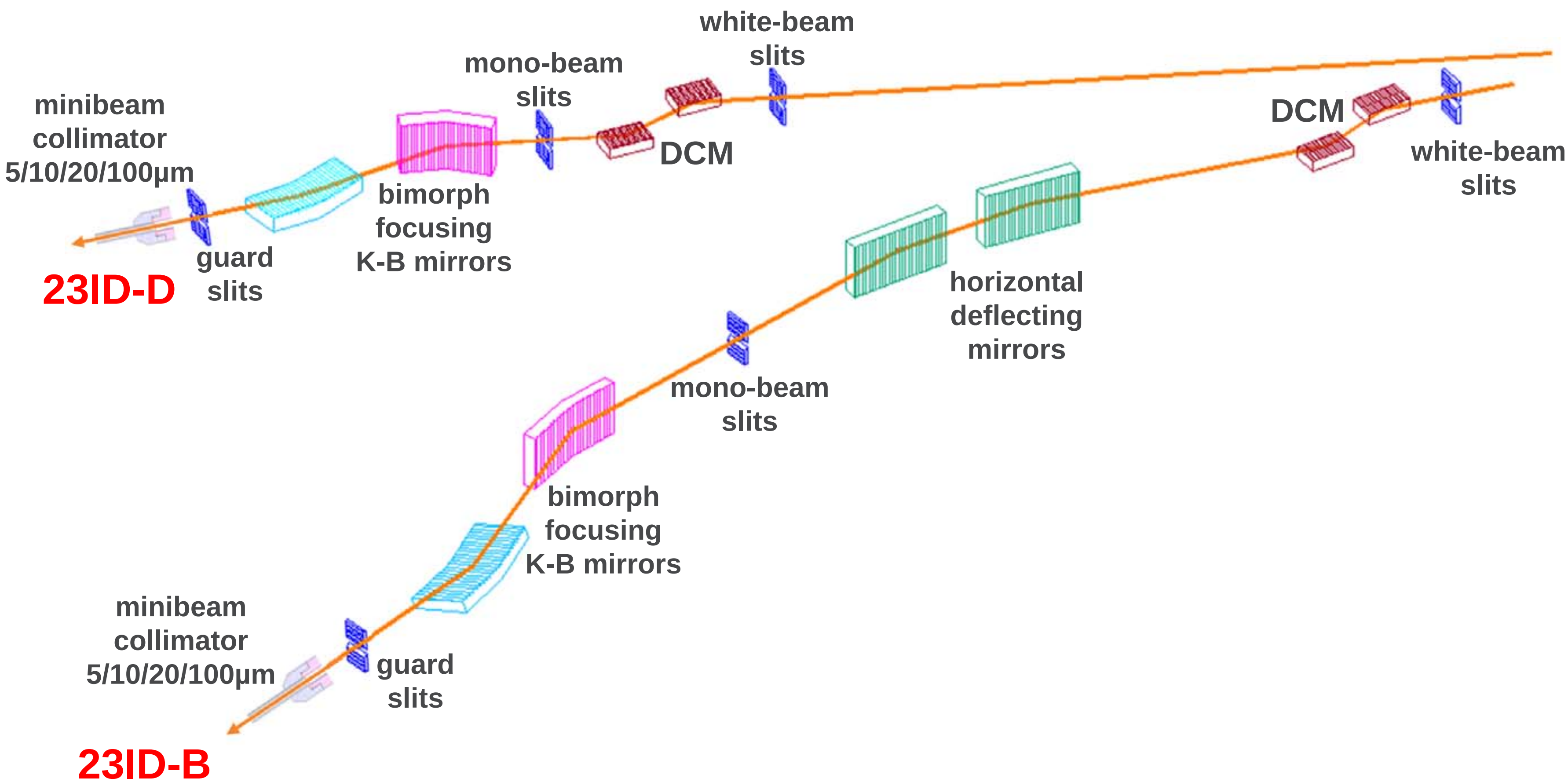


SERGEY STEPANOV

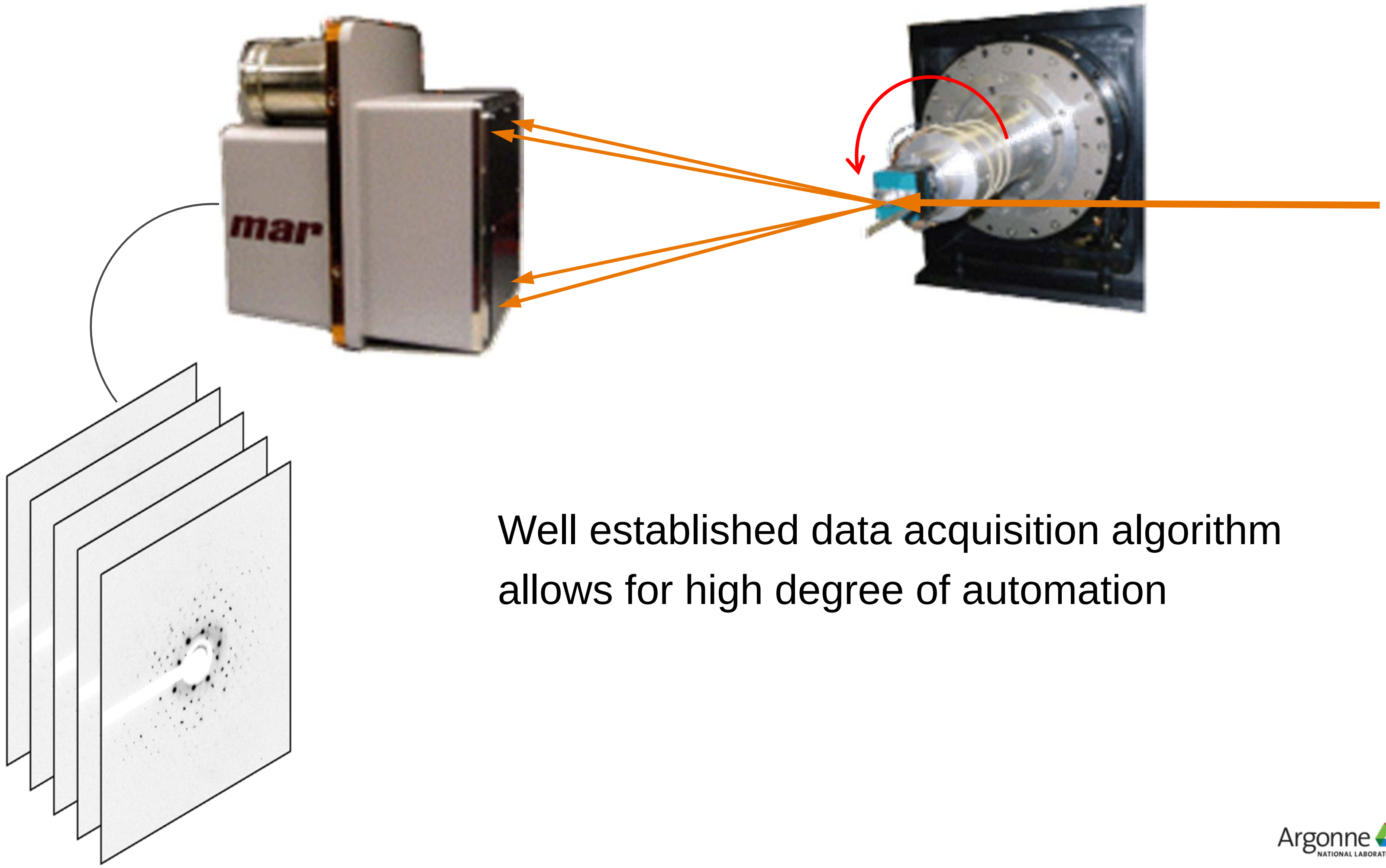
Argonne National Laboratory,
Advanced Photon Source,
Lemont, IL, USA

October 2016, NOBUGS Meeting, Copenhagen, Denmark

GMCA@APS: macromolecular crystallography beamlines

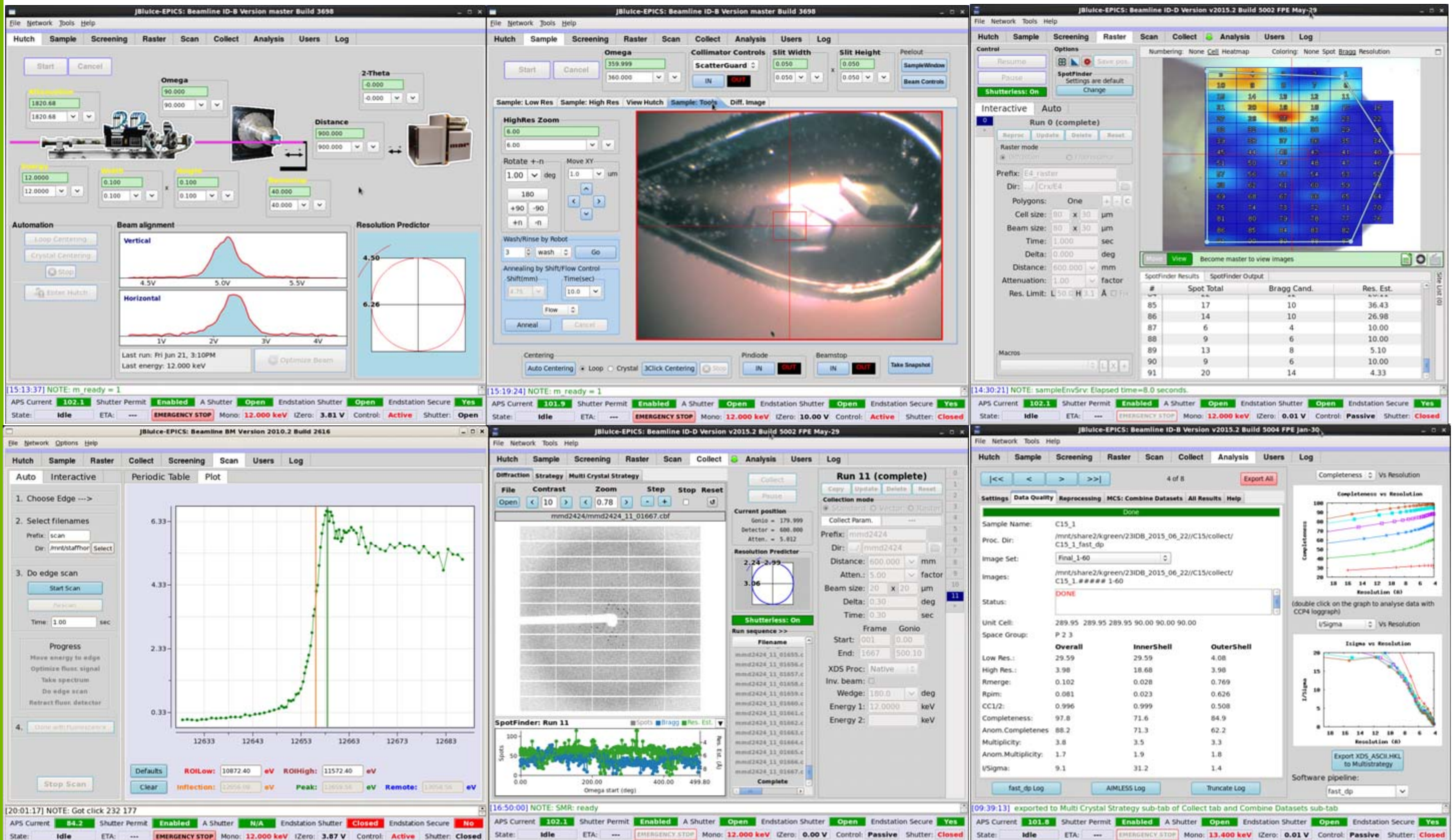


GMCA@APS: MX experiment basics



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 like data collection or energy scanning

Argonne NATIONAL LABORATORY

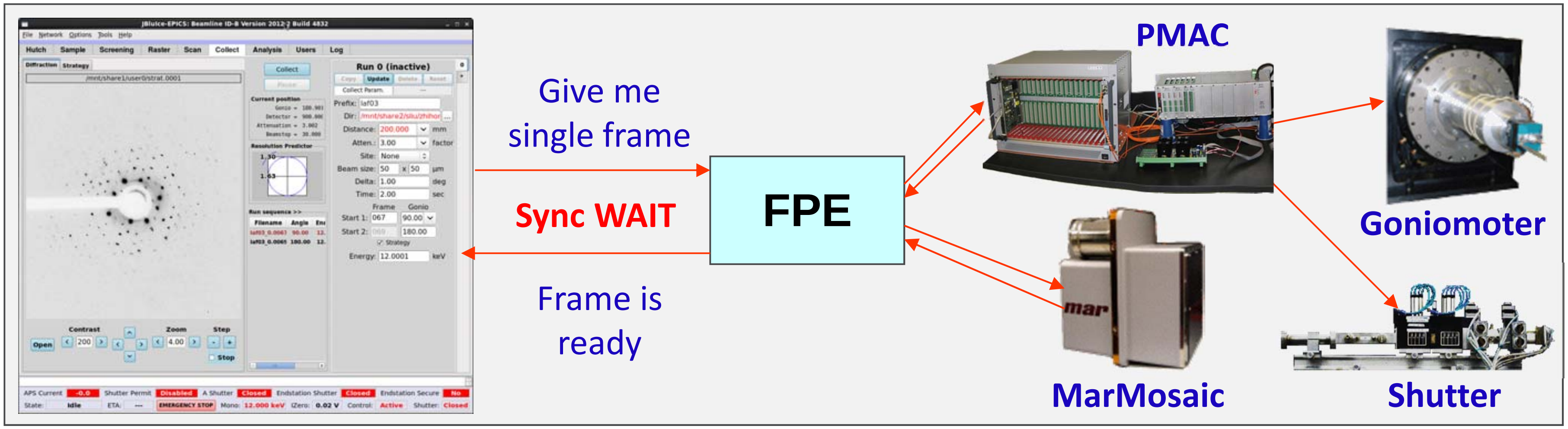
Game change: arrival of fast area detectors

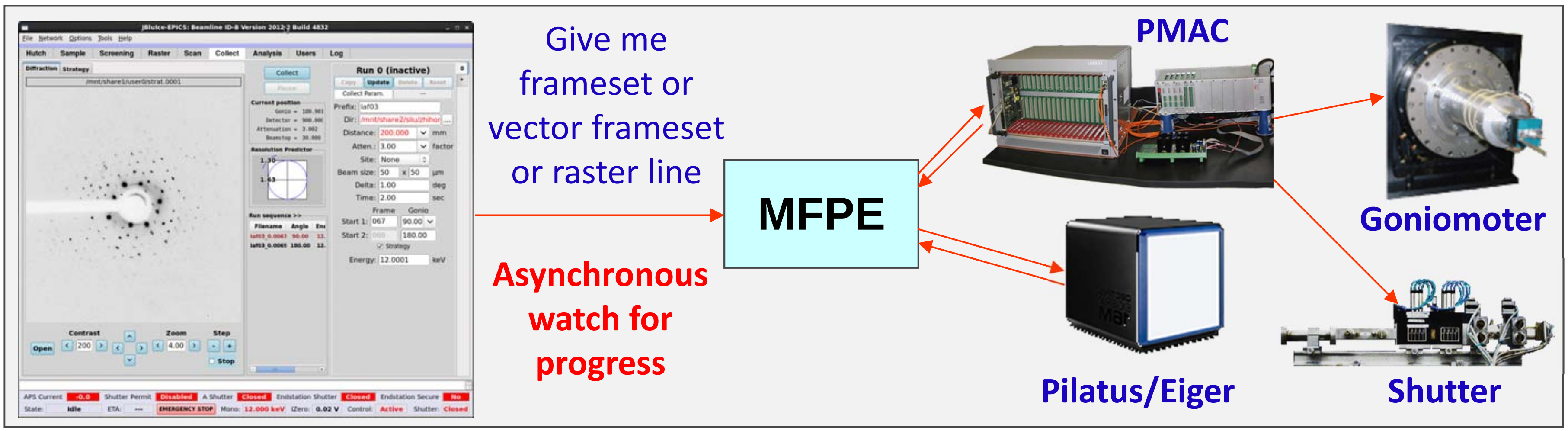
In the past two years GMCA upgraded from two MarMosaic 300 area detectors to Pilatus3 6M and Eiger 16M

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	8 TB /day hdf5+cbf
Auto data processing	desired	must	must

Control system should be adapted to qualitatively new synchronization requirements, data rates, data volumes and processing

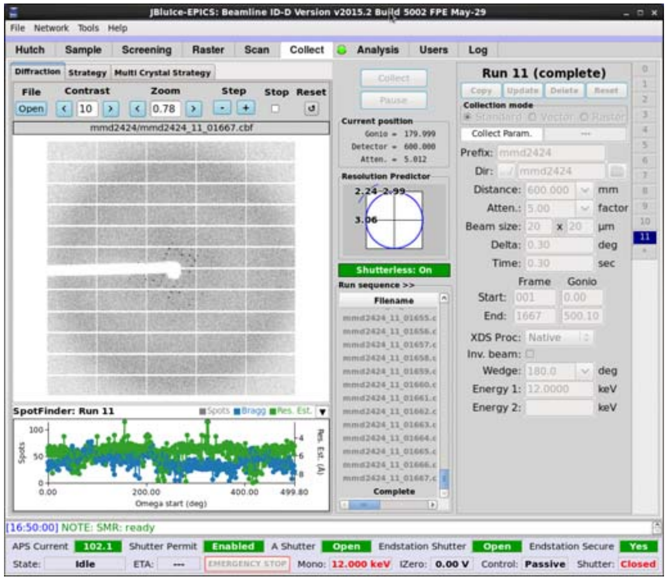
Gonio-detector synchronization for shutterless data collection: delegating controls to lower level

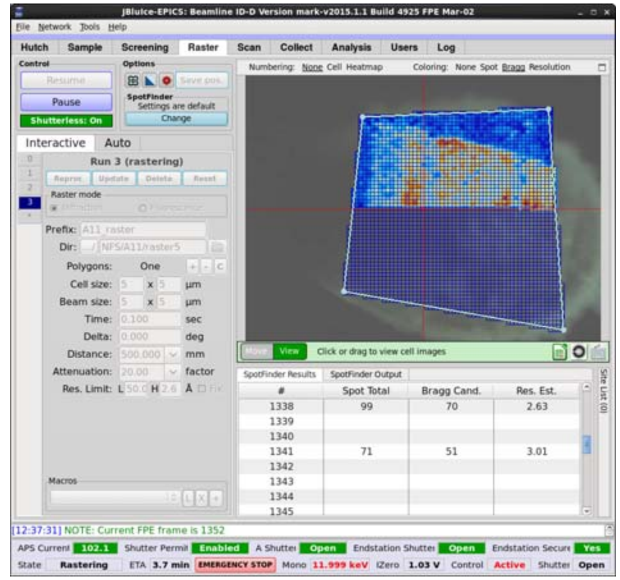




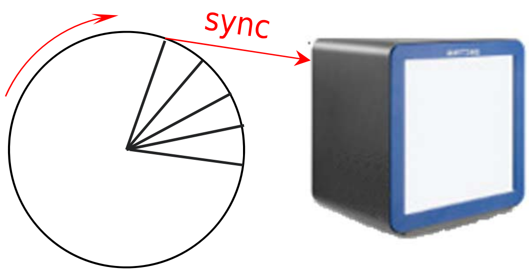
Previously JBluIce was requesting single frame from Frame Processing Engine (FPE). Now series of frames are processed by Multi-Frame processing Engine (MFPE). FPE is a State Notation program running in EPICS soft IOC. It re-programs PMAC to sync goniometer, detector and shutter.

Algorithms of shutterless data collection and rastering

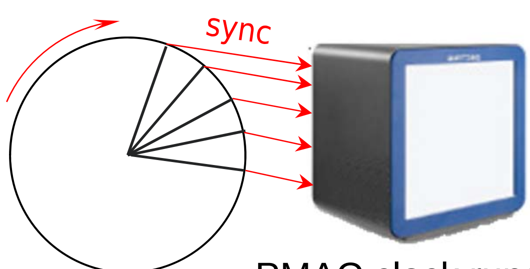




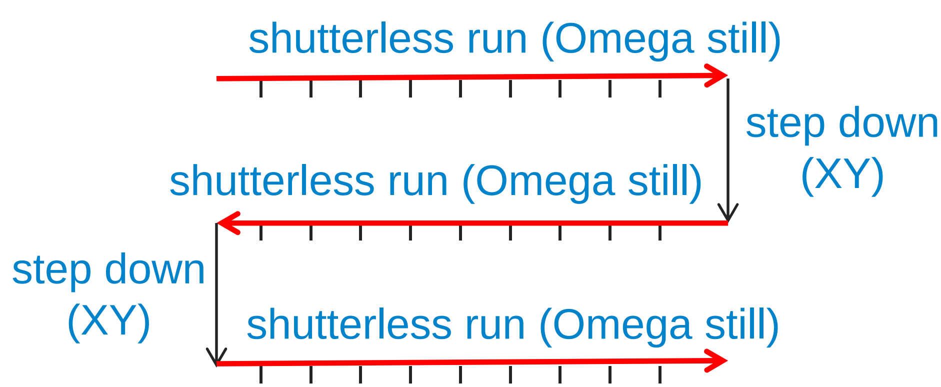
Traditional shutterless data collection



Truly synchronous shutterless collection



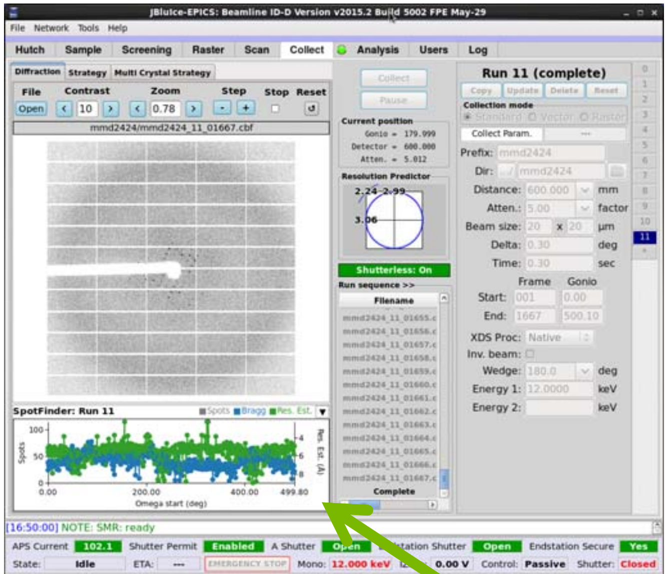
Finding promising diffraction spots with rastering:

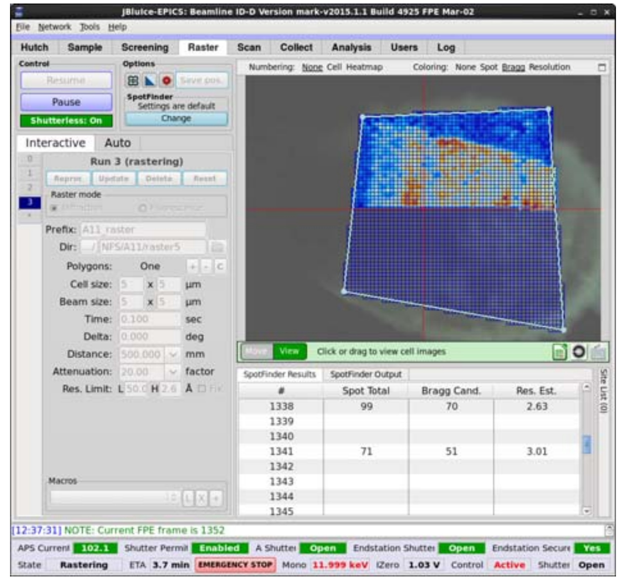


PMAC clock runs at 2 kHz; enough to sync 100Hz data collection

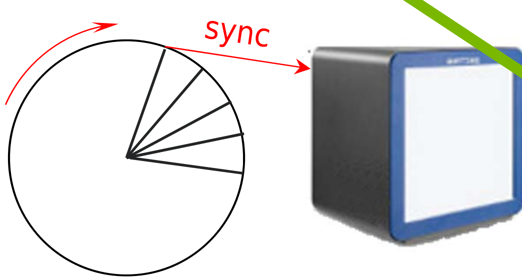
Argonne NATIONAL LABORATORY

Algorithms of shutterless data collection and rastering





Traditional shutterless data collection



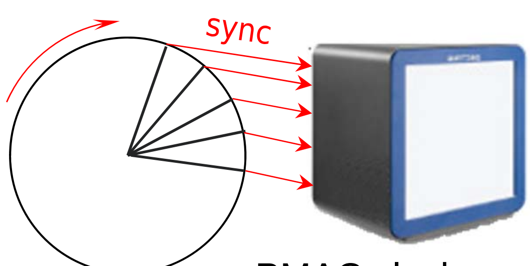
ion

)

tep down
(XY)

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

Truly synchronous shutterless collection



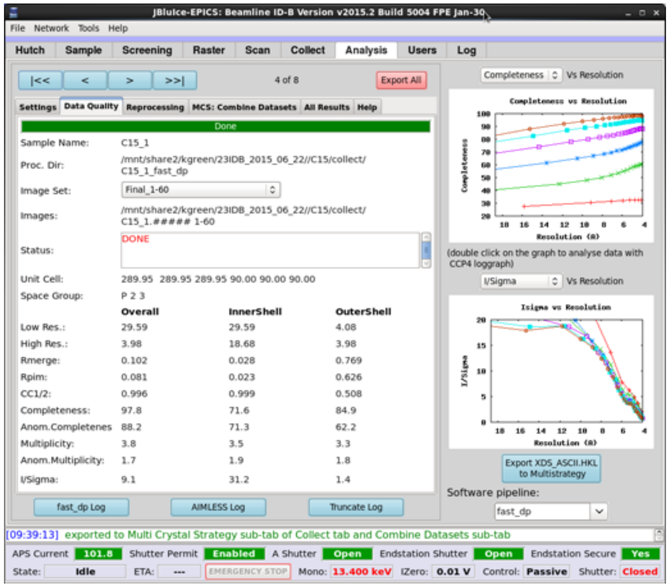
ste

)

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

Argonne NATIONAL LABORATORY

Automatic data processing



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

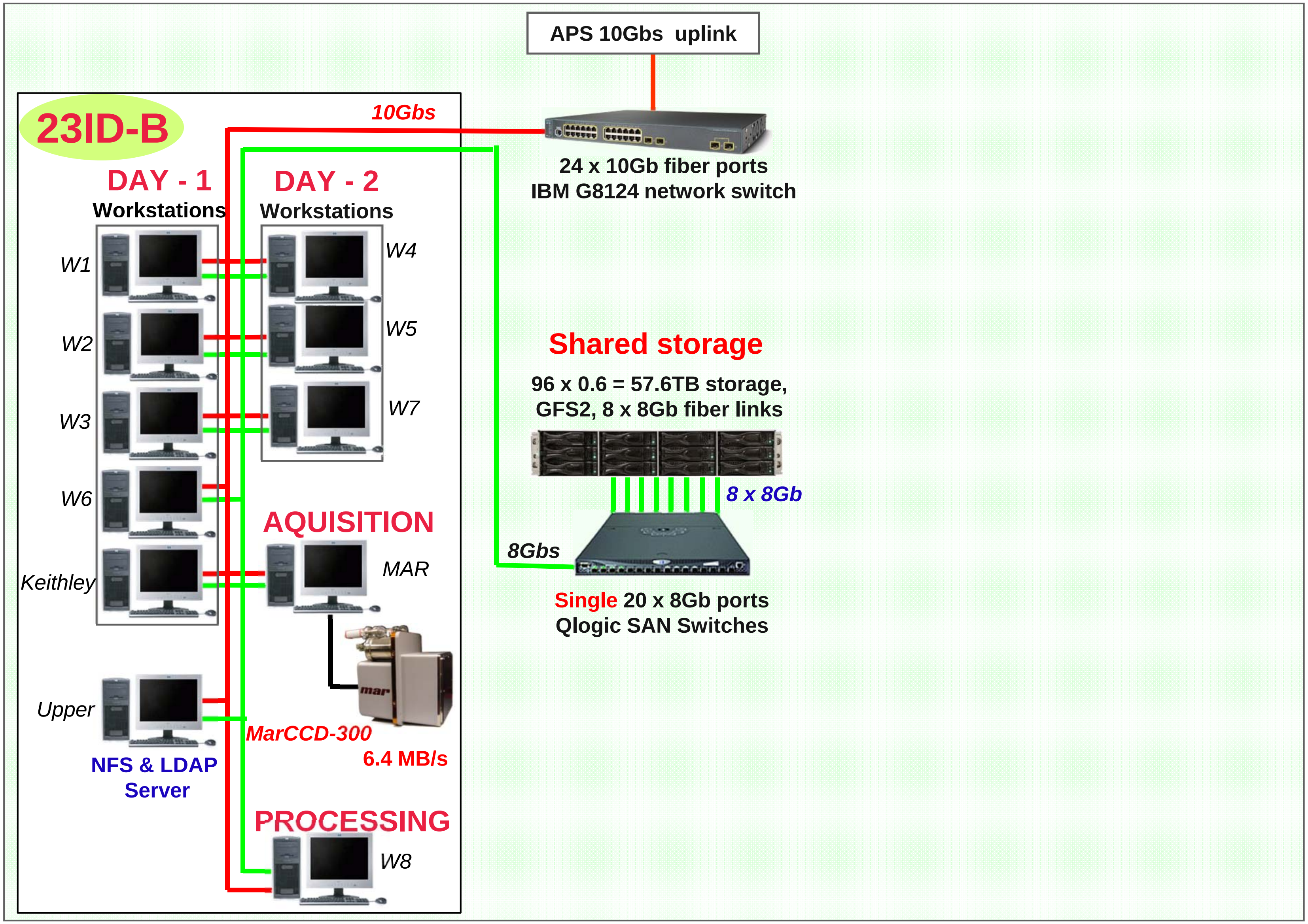
Users need to see the results as soon as the dataset is taken in order to plan next step.

- Runs on cluster using Son of Grid Engine (https://arc.liv.ac.uk/trac/SGE)
- Two pipelines based on XDS: GMCAproc (home-made) and fast_dp (DIAMOND)
- Also considering DIALS and autoPROC (Global Phasing)

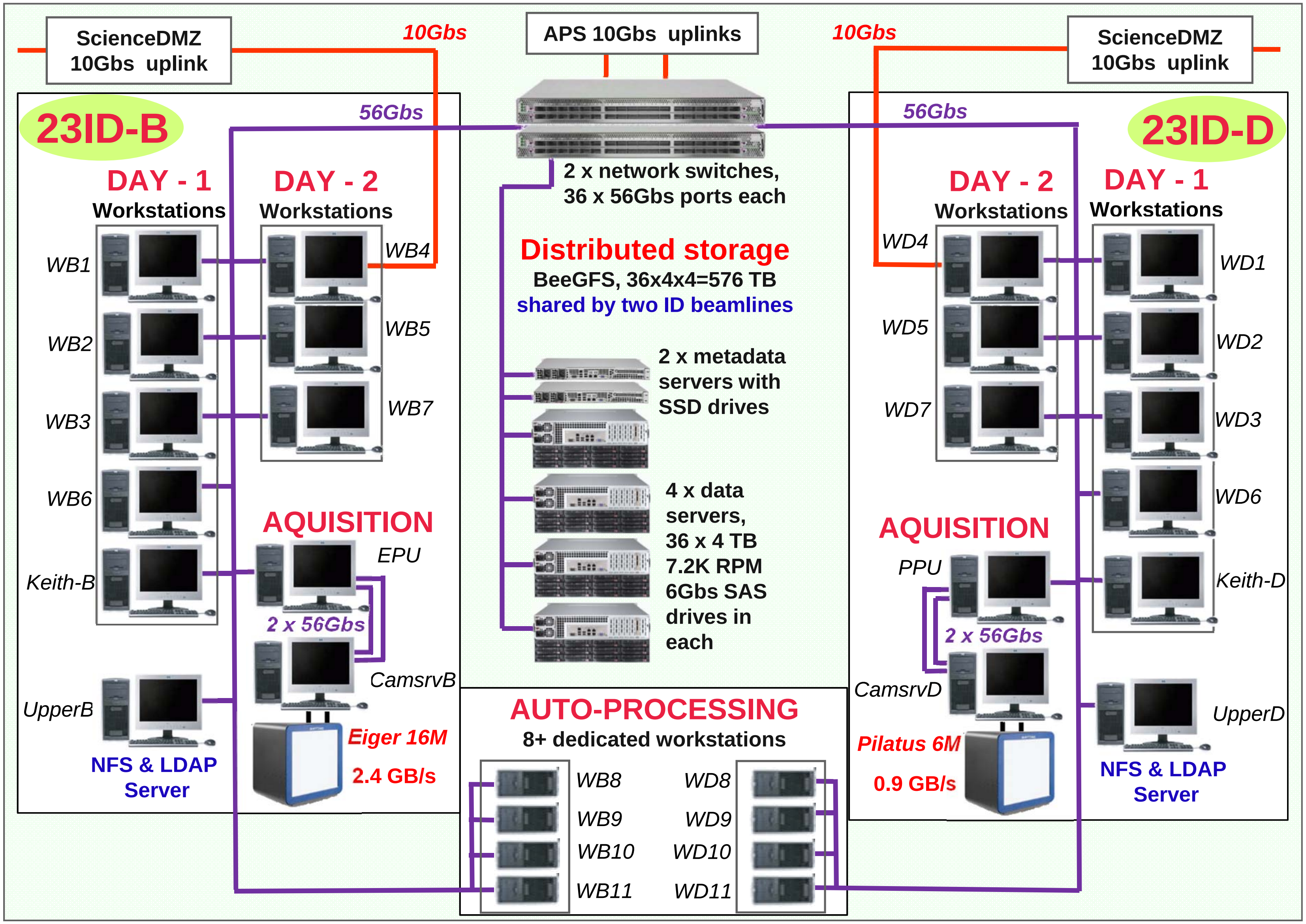
Network and storage upgrade to high data rates & volumes

	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 2015



GMCA@APS network/SAN structure, September 2016

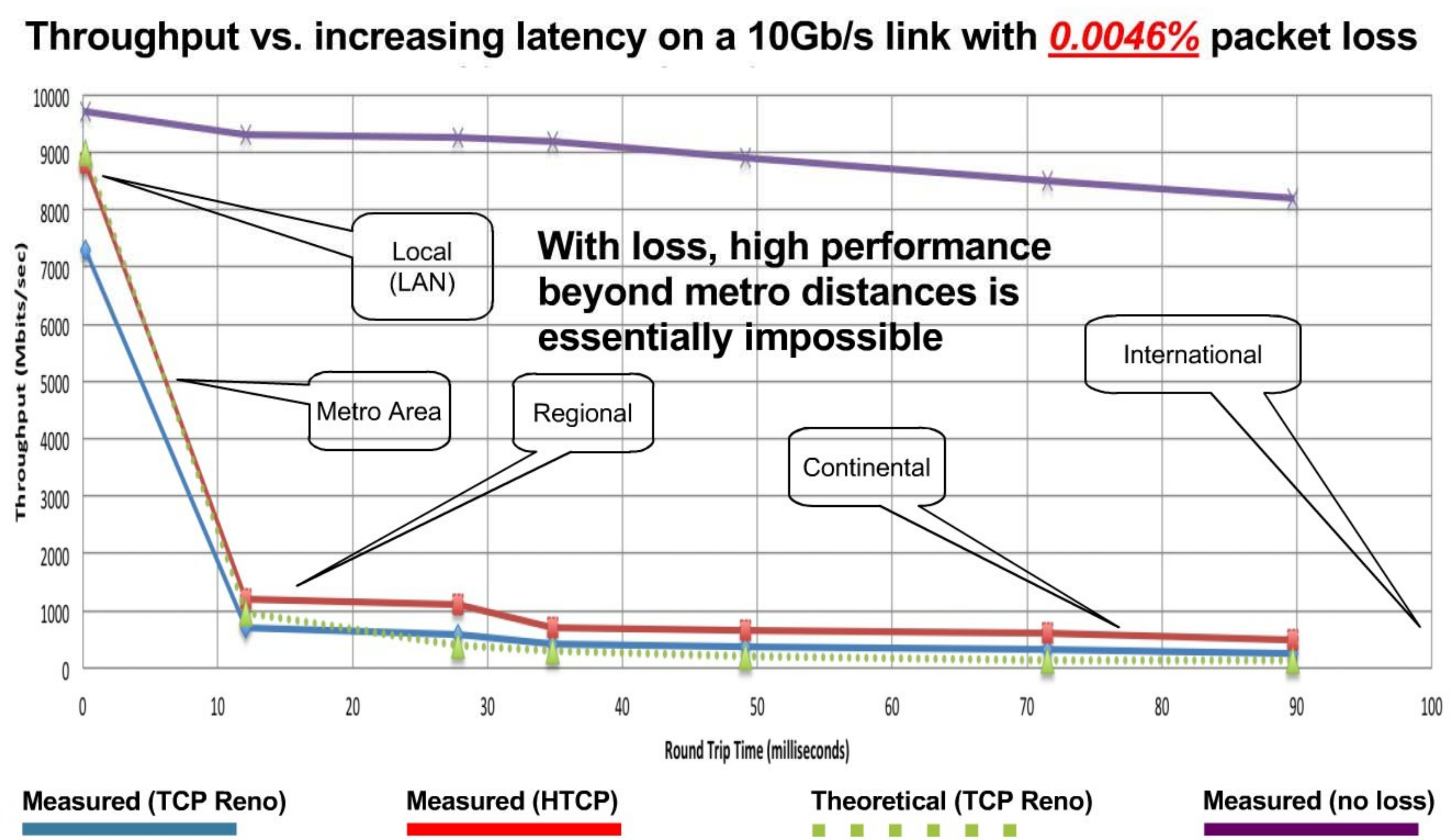


ScienceDMZ project in collaboration with esNET

- When users were producing 200 GB of data per one-day experiment, they were typically taking them home on portable USB drives, although GMCA was offering Globus GridFTP servers at each beamline.
- Upgrading data volumes from 0.2 TB to 8 TB changed the game. Globus became the must and the speed of data transfers via Globus became critical.
- The idea of ScienceDMZ is to replace general-purpose firewalls filtering tons of traffic types and therefore often overloaded and loosing packets by specialized firewalls permitting the Globus traffic only.

ScienceDMZ: packet loss effect

A small amount of packet loss makes a huge difference in TCP performance



Courtesy Jason Zurawski, esNET

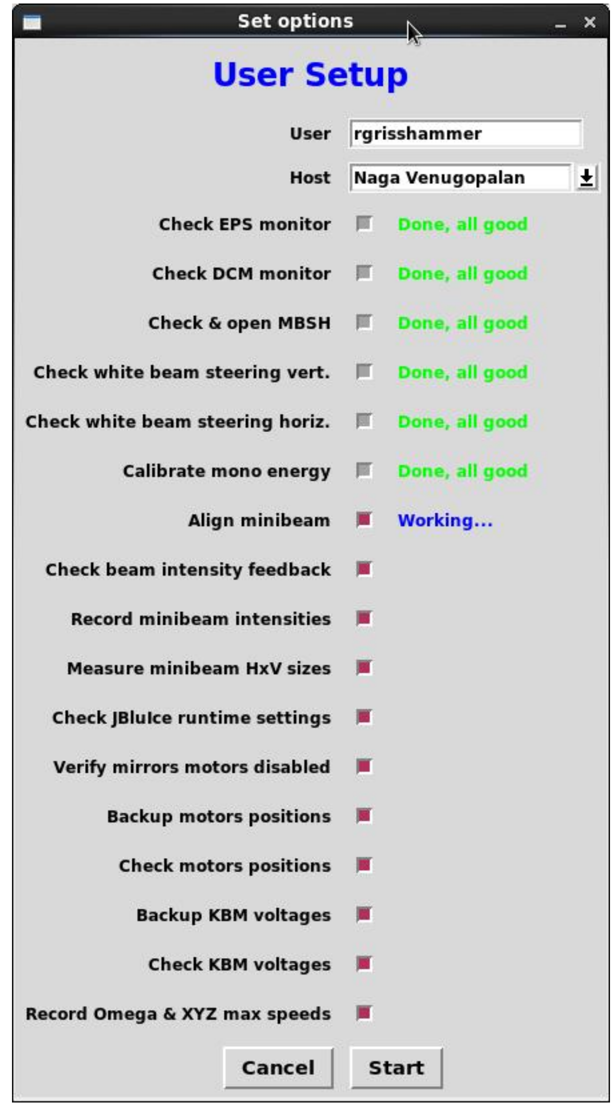
GMCA Globus Servers

- GMCA now offers one regular Globus Server and one ScienceDMZ Globus server at each beamline.
- The results are promising, but for full advantage remote clients should optimize their institution's network too (preferably install ScienceDMZ) and the full path needs to be optimized.
- Installing perfSONAR servers on both ends for monitoring network performance helps a lot.

	150GB (medium files & shallow dirs)			*500GB (large files)*		
	Regular	ScienceDMZ	x Better	Regular	ScienceDMZ	x Better
GMCA->BNL (MB/s)	81	266	3.3	140	456	3.3
GMCA->LBL (MB/s)	46	122	2.7	109	442	4.1

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 8 TB show that there is a lot of room for improvement.
- Larger sample Dewars are definitely required. We are working on it.
- 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.



Argonne NATIONAL LABORATORY

Conclusions

Upgrading beamlines to fast detectors goes far beyond the task of interfacing the detector API and synchronizing it with other beamline equipment. One should also thing about handling high data rates, storing large data volumes, efficiently copying data to users institutions, displaying and processing data, and improving beamline automations to raise the share of beamtime when the detectors are actually in use.

Argonne NATIONAL LABORATORY

Acknowledgements

Oleg Makarov, Mark Hilgart, Sudhir Pothineni and Robert Fischetti

Advanced Photon Source, Argonne National Laboratory

Jason Zurawski

Energy Sciences Network (ESnet)

Janet Smith

Life Sciences Institute, University of Michigan

Argonne NATIONAL LABORATORY