Australian Government

ansto

Australia Synchrotron: Facility update

synchrotron.org.au

Australian Synchrotron

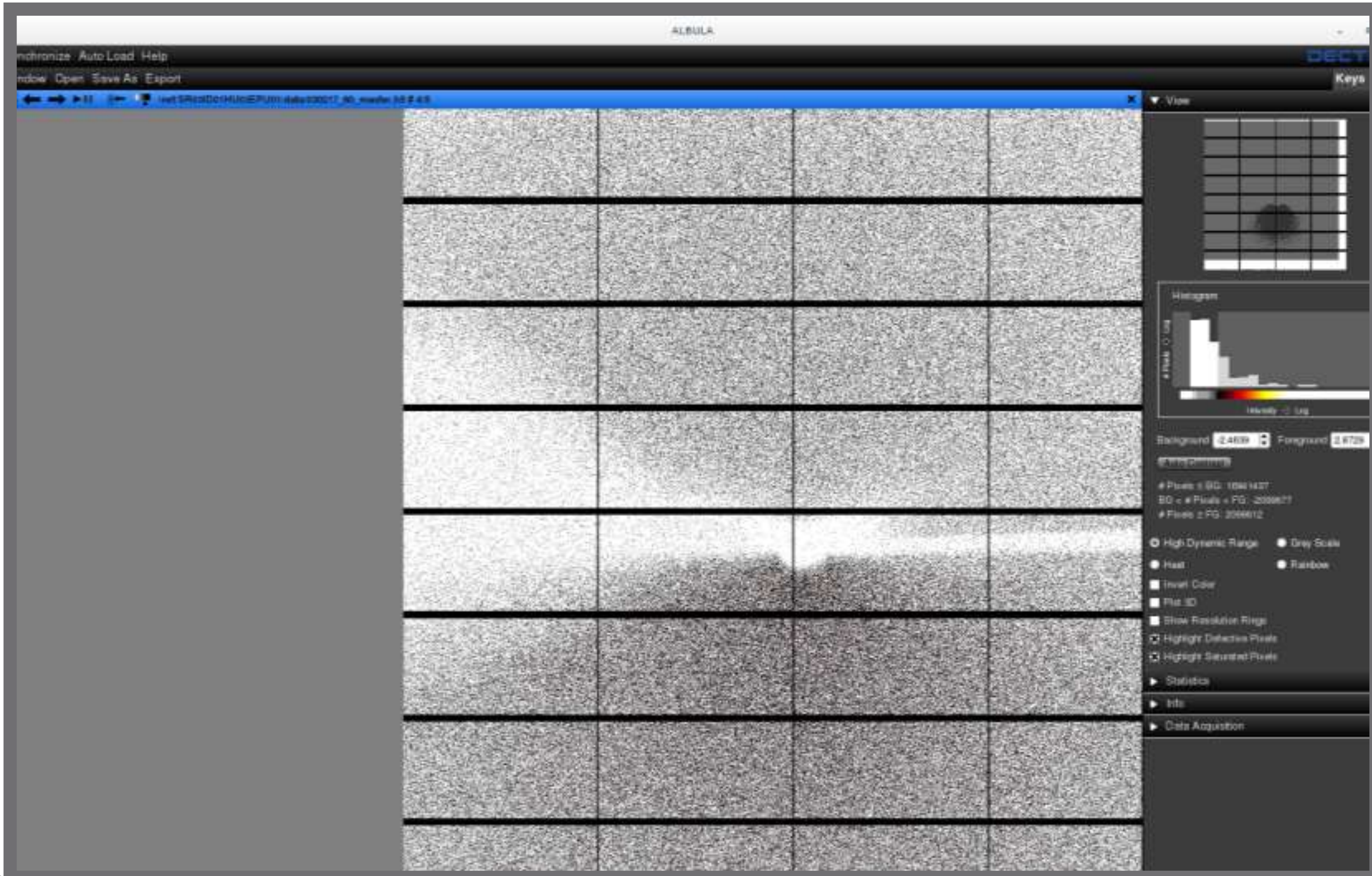
Australian Government

ansto

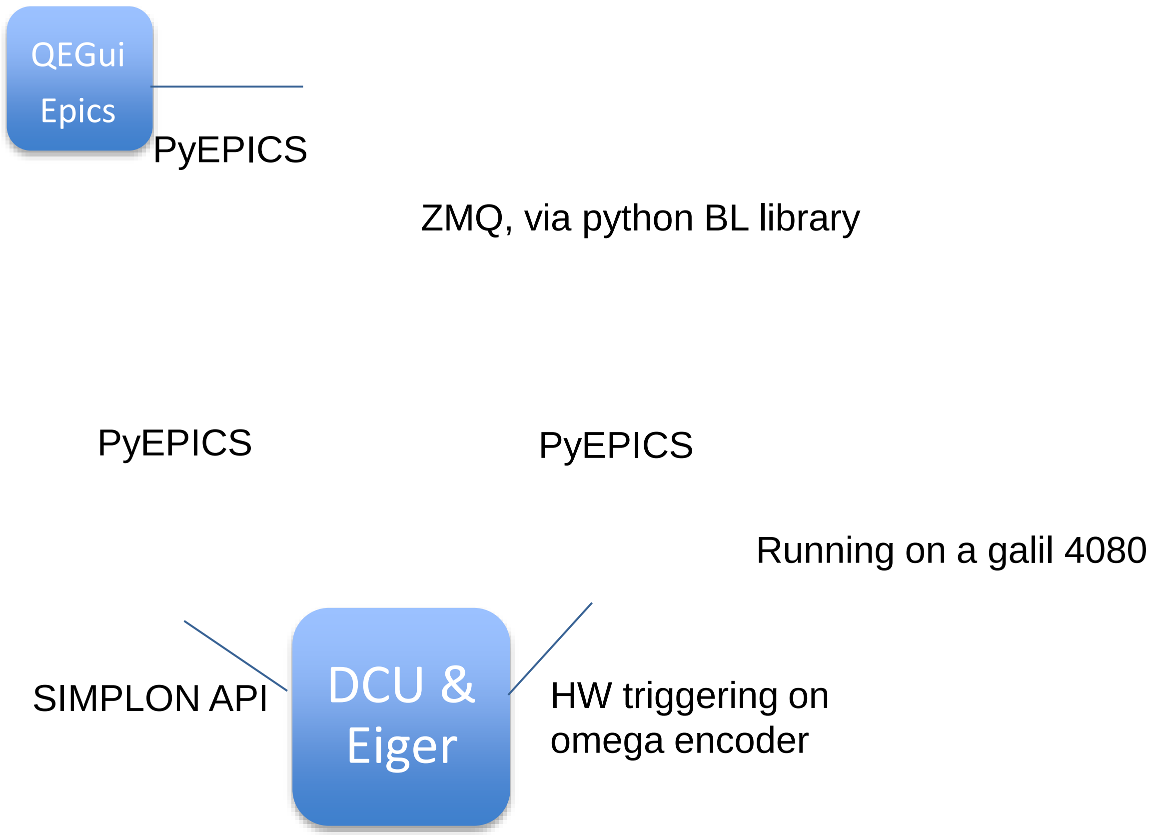
David Aragao

Australian Synchrotron

General user operations from 18/2/17

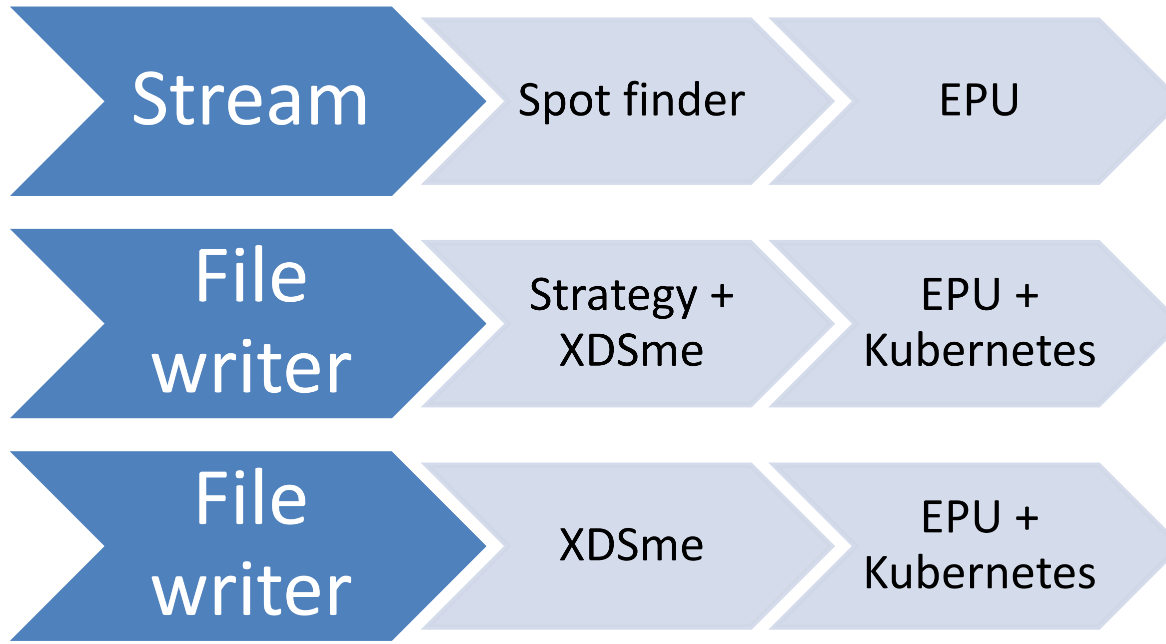


Software services for Eiger collection

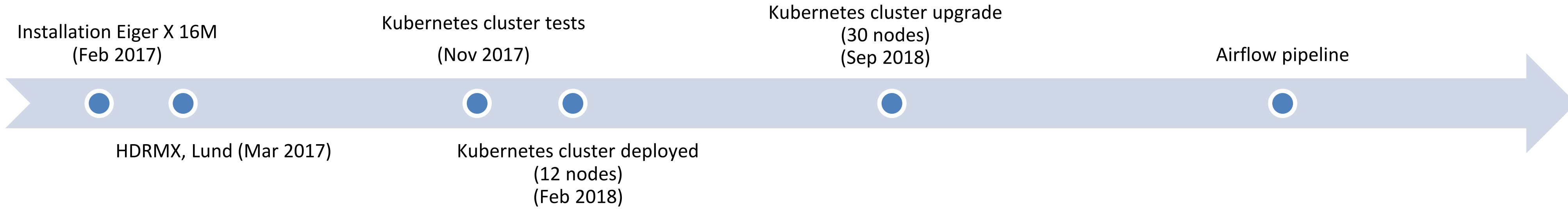


Typical modes of collection

- Raster
  - 0.1 seconds exposure
  - 1 degree wedge
  - 1 Hz / 1 image
- Screen
  - 2 seconds exposure
  - 20 degrees wedge
  - 40 Hz / 80 images
- Dataset
  - 18 seconds exposure
  - 180 degrees wedge
  - 100 Hz / 1800 images



John Marcou

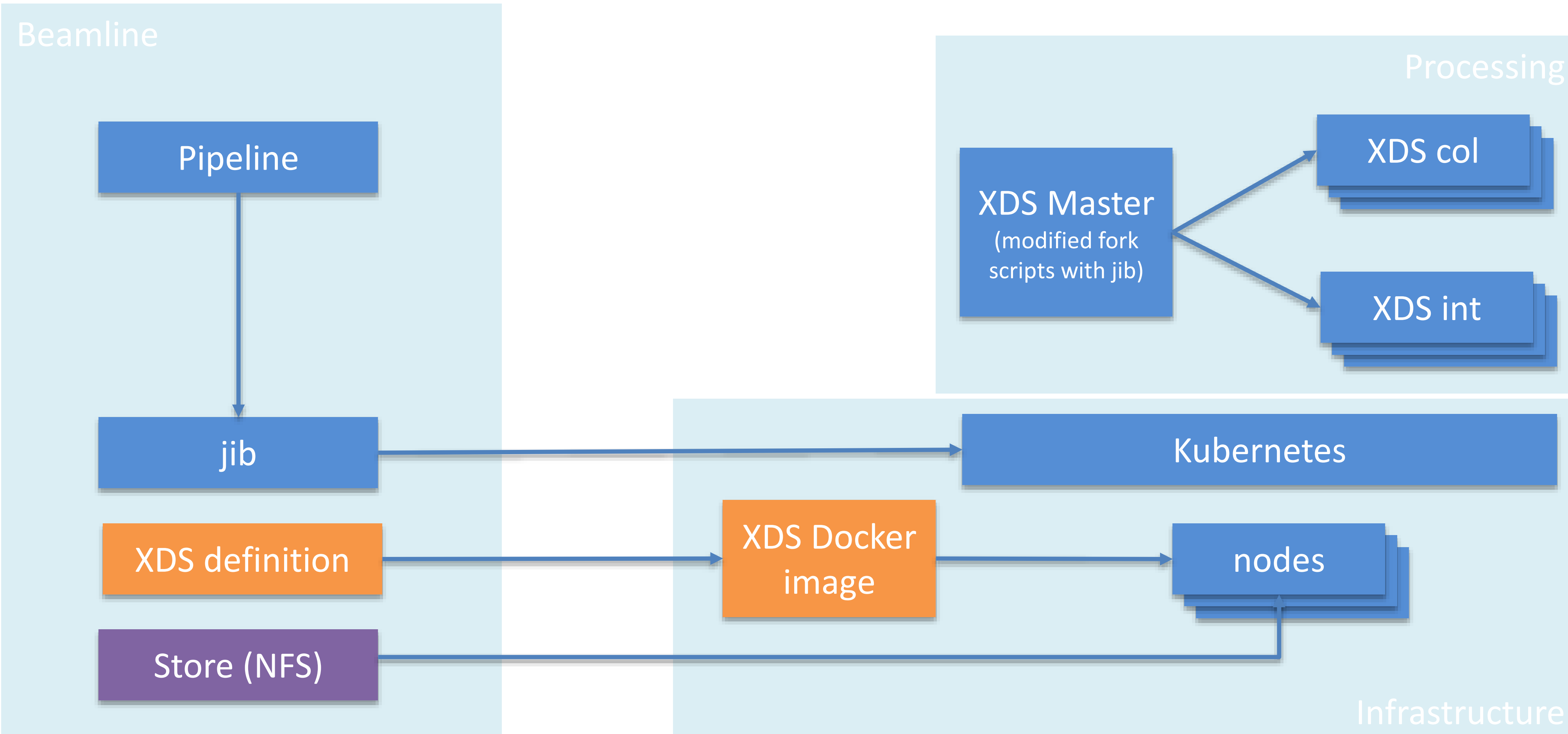


EPU
- Monitor redis queue
- File system host
- 3 redis queue nodes – each may run a processing
- Starting up Kubernetes processing via python script
- Pointless
- Aimless
- Truncate
- (CX) Sadabs, xprep

XDS Master
(Kubernetes Container/Pod)
- All xdsme steps
    - Higher level control of COLSPOT, INTEGRATE steps

XDS Subprocess
(Kubernetes Container/Pod)
- COLSPOT instance
- INTEGRATE instance

MX XDS integration



EPU alone was not sufficient (Feb 2017 – Dec 2018)

Dataset/computer	1x full XDS (Kub) 1x CORRECT P1 (Kub) 1x CORRECT anomalous (Kub) 1x aimless (EPU) 1x truncate (EPU)	Screening (MX2, typical)
Insulin 1800 images / EPU 144 threads	85 to 120 seconds	30 s – 1 min
2x Insulin 1800 images / EPU 144 threads	125 - 180 seconds	30 s – 1 min

- Slow sample changer (~4.5 minutes per change) prevents multiple datasets from queueing – saved us until Feb 2018
- Fast sample changer (25 seconds per change) resulting in multiple datasets very fast → new cluster saved us from Feb 2018

Docker & Kubernetes

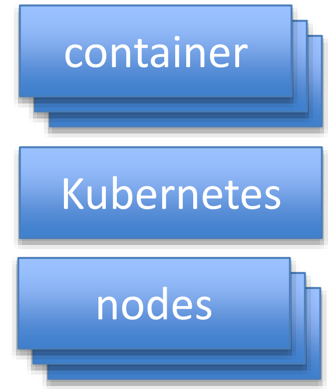
- Docker is a process manager
  - Bundle app + context : docker image
  - Start and manage a process : docker container
  - Portable, repeatability, native performance

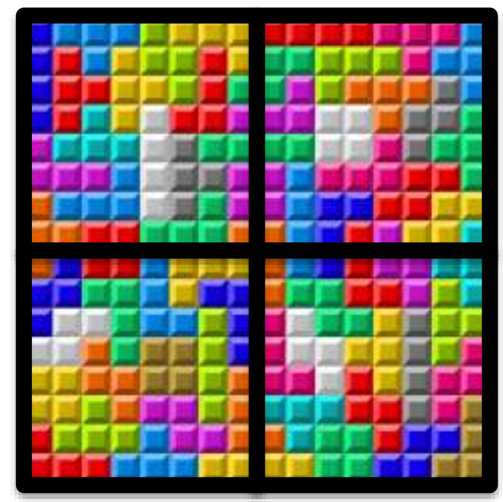
- Kubernetes (K8s) is an orchestrator
  - Is not an HPC scheduler (yet)
  - Manage N nodes
  - Decoupling (data, code, config, infra)
  - Schedule and start docker container on nodes
  - Multiple workloads (deploy/job)
  - Mutualize resources

- MX Kubernetes clusters
  - 30 nodes (AUD$10k each)
  - 64 cores cpu / 128GB ram
  - 10GB network

→

- 1920 cores
- 3840GB ram





Australian Government ansto

MX Kubernetes solution

XDS integration:

- Write a job submitter helper (jib)
- Rewrite the XDS forking bash scripts
- Pipeline is node contacting nodes anymore, but the K8s API (abstraction)
- Reach the performance of max works

Other workloads:

- Remote analysis desktop
- Long run process
- Prioterisation and affinity
- Spotfinder to kubernetes
- Merging datasets to kubernetes
- Redisque pipeline to airflow

Australian Synchrotron

Kubernetes cluster 12 nodes (Feb 2017 – Aug 2018)
Kubernetes cluster 30 nodes (Sep 2018 - now)

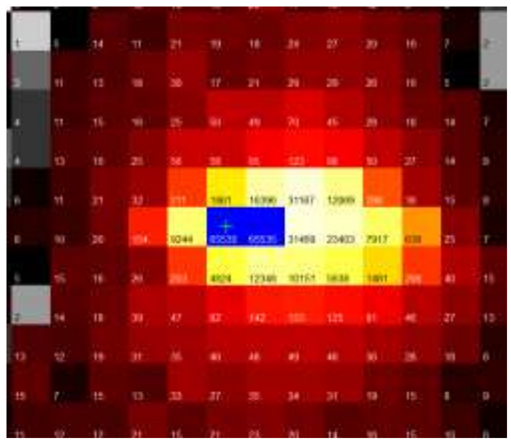
Dataset/computer	1x full XDS (Kub) 1x CORRECT P1 (Kub) 1x CORRECT anomalous (Kub) 1x aimless (EPU) 1x truncate (EPU)	Screening (MX2) 80 images
Insulin 1800 images /EPU + Kubernetes (6 nodes per dataset)	~85 seconds	25 s – 1 min
Insulin 1800 images / EPU 144 threads	85 to 120 seconds	30 s – 1 min
2x Insulin 1800 images /EPU + Kubernetes (6 nodes per dataset)	~85 seconds	25 s – 1 min
2x Insulin 1800 images / EPU 144 threads	125 - 180 seconds	30 s – 1 min

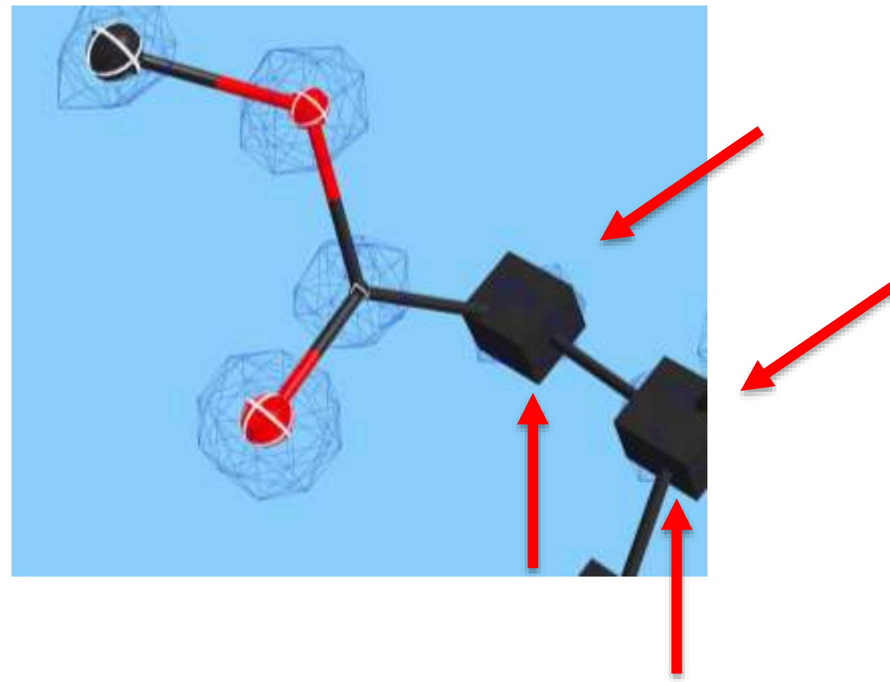
Australian Government

Jun Aishima

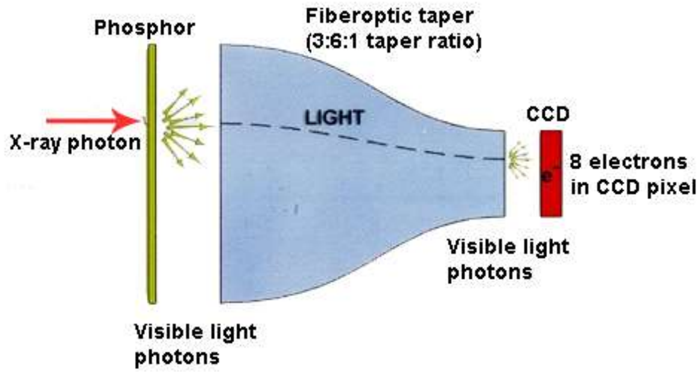
Australian Synchrotron

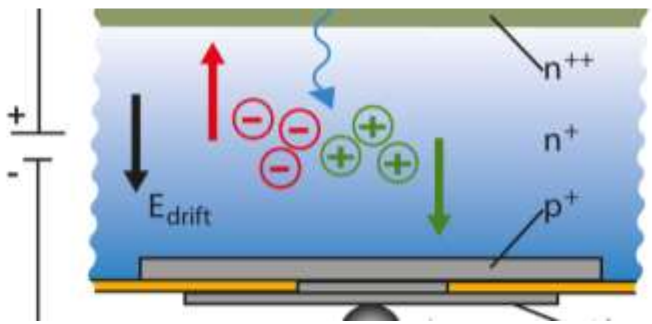

Background
Foreground
Pixels ≤ BG: 10641186
BG < # Pixels < FG: 6205288
Pixels ≥ FG: 91114
High Dynamic Range
Grey Scale
Heat
Rainbow
Invert Color
Plot 3D
Show Resolution Rings
Highlight Defective Pixels
Highlight Saturated Pixels
Statistics
Region of Interest
Max Finder
Find Peaks
X Y Intensity
1388 2511 65535
1445 2217 65535
2991 2012 65535
2330 1528 22870

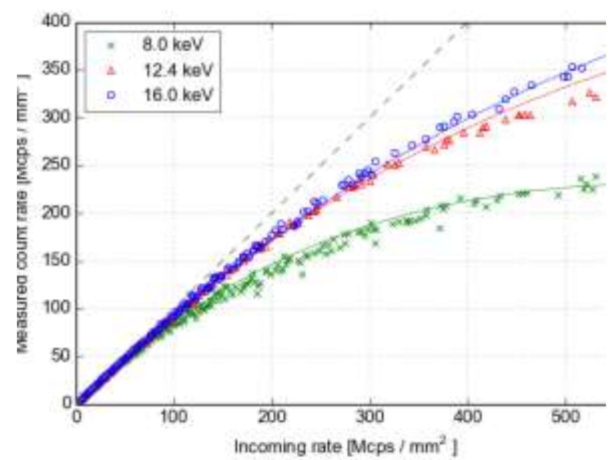




Integrating vs. counting detector







This poster at AsCA on this general topic is being prepared as a manuscript for submission

Australian Government

ansto

Alternative approaches to synchrotron X-ray diffraction data collection

Neil Paterson, Graeme Winter, Richard Gildea
Diamond Light Source, UK

Despite the advances in synchrotron hardware over the years, particularly in the detector field, the way we approach data collection still mainly follows the old paradigm of screening images > strategy determination -> data collection for each individual sample. This 'maximum dose/minimum completeness' approach risks radiation damage to samples that are more sensitive than average and the resulting dataset can be inadequate for weaker experimental phasing cases. The advent of fast photon-counting detectors; common multi axis goniometry; high speed and precise stages; and robust automatic processing of both full and partial data sets mean that this traditional approach can be quite inefficient in terms of beamtime due to the overheads associated with each step.

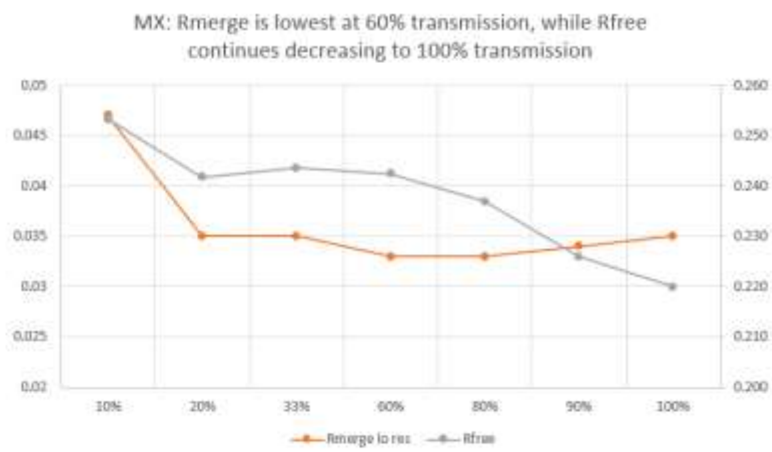
At Diamond, we are exploring the idea of using low dose, high multiplicity measurements to build up each dataset and relying on counting statistics to improve signal/noise. This involves the collection of multiple weak sweeps of data that are combined to yield a dataset with comparable intensity to the traditional approach but with higher precision and a more even radiation damage spread. For sensitive samples, radiation damaged data can be excluded without compromising overall completeness. This removes the need for indexing and lifetime estimation prior to data collection. In combination with automated X-ray centring of samples, this approach provides a path to providing fast and efficient, fully automated data collection.

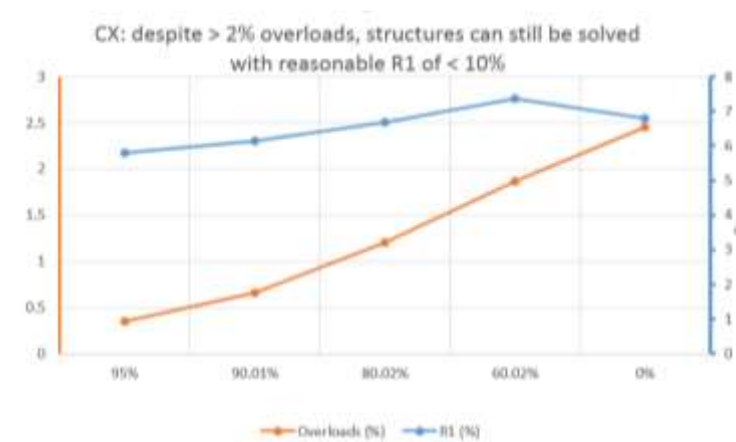
Key conclusions:

- Very low transmission high redundancy collections are good
- Using high transmission for weak data not necessary – good ligand density still visible with low transmission
- Do not just look at visible spots – collect data carefully using multiple passes, start with low transmission, and monitor stats

We tried overexposing MX, CX crystals but data was still ok

Australian Government





Australian Synchrotron

“Counting detectors love weak data”, Andreas Förster, Dectris Ltd.

- Expose less – our standard recommendation has been to use 0% attenuation. **Try starting with 50%**
- Wean users from adxv and simulated 1 degree images, use Albula to view individual images and use its tools to monitor the number of overloads and high intensity peaks
- For our standard collection, aim for a maximum of 10k counts – may be tricky to get this with fewer diffraction spots and diffraction anisotropy if just examining the first few images of a dataset
- Clarify situations where unattenuated beam is necessary to optimize high resolution data

Future work – more tools to help users collect better data

- Multi-pass data collection – develop tools to collect and merge a low resolution, high attenuation dataset with a high resolution, unattenuated dataset
- Provide feedback to users of percentage of overloads when it may start becoming a problem
- Make these tools open source and methods published so that other beamline scientists and electron detector manufacturers can take advantage of them
- Get Albula to follow data collections (like adxv autoload) to be more similar to current adxv functionality
- Better understand the issues resulting in CX non positive displacements – how do overloads and count rate correction result in small structure factors?