Data Acquisition, Management, and Real-Time Analysis at the EMBL PETRA III Beamlines

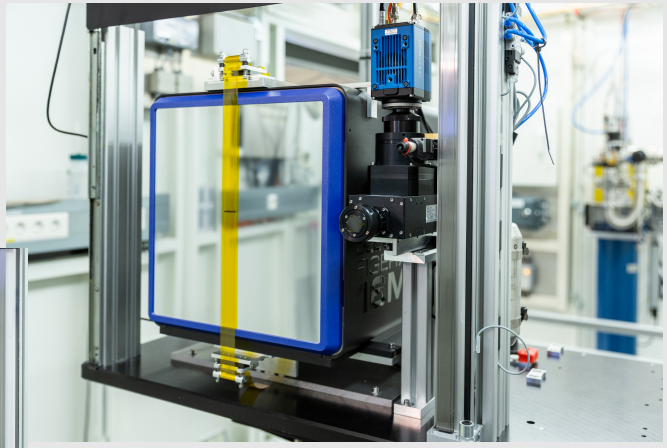
or of Daemons and Data

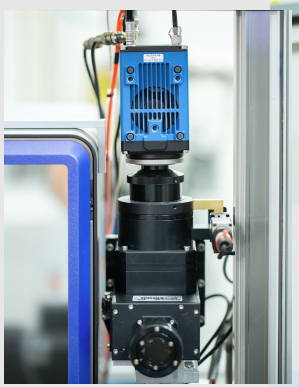
Marina Nikolova

EMBL

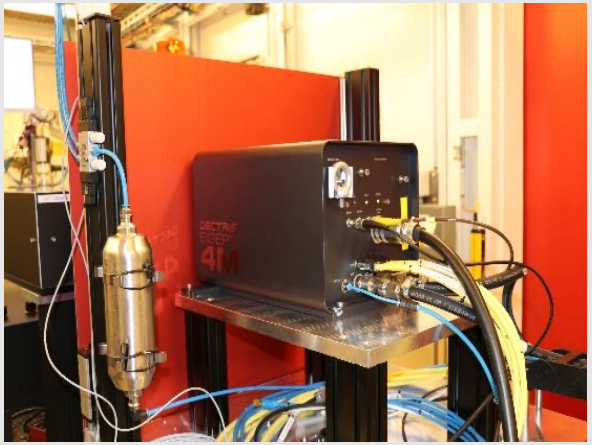
High-Frame-Rate Detectors and Cameras

- Eiger 16M, DECTRIS
- 16 Megapixel frames @133Hz or 4 Megapixel frames @750Hz
- Currently integrated at P14





- PCO Edge Imaging Camera
- 4 Megapixel frames @100Hz
- Integrated at P14



- Eiger 4M, DECTRIS
- 4 Megapixel frames @750Hz
- Moving between TREXX and P12

- Eiger X 16M CdTe, DECTRIS
- 16 Megapixel frames @130Hz or 4 Megapixels @750Hz
- Delivery expected May, 2021
- To be integrated at P14

EMBL

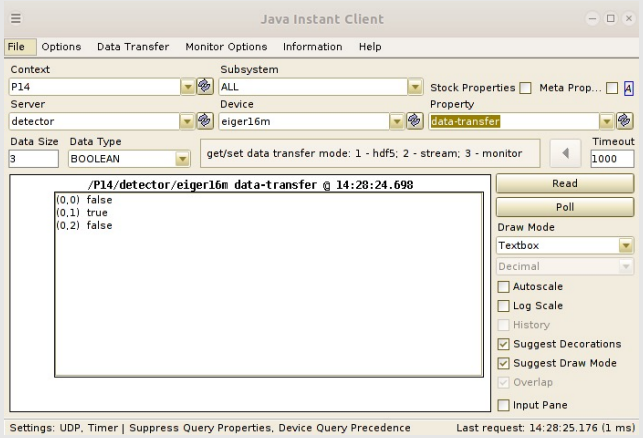
Tsunami Data Waves

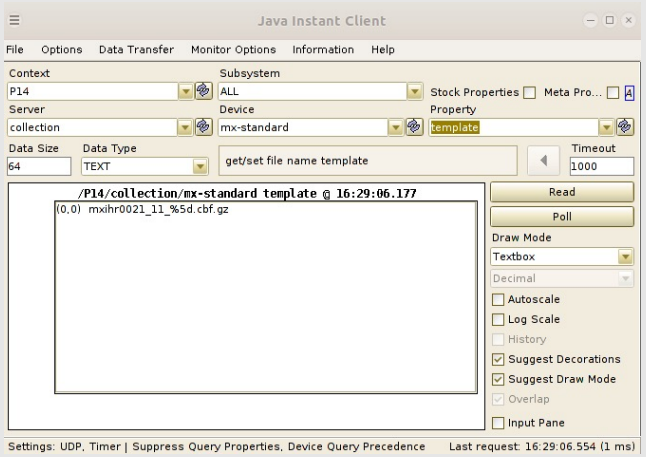
EMBL

The Control-Freak Daemon

Acquisition Control
Detector/Data Collection Server

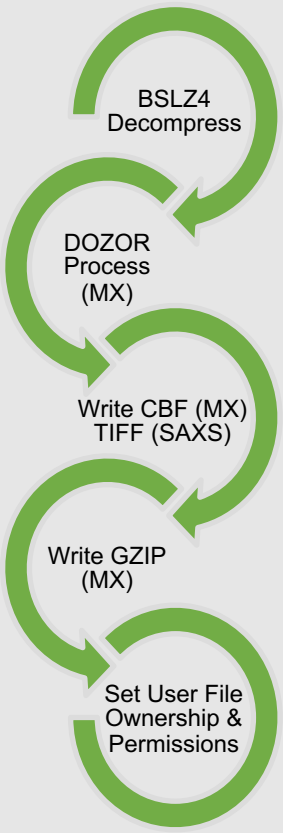
- Unified interface to all X-Ray area detectors
- Presents the experiment as a service
- User Interface point of contact
- Manages the collection sequence
- Creates metadata
- Ensures metadata consistency
- Enables scripting non-standard experiments
- C++; TINE server
- Integrated at P14, TREXX, P13, P12





EMBL

Data Transfer and Real-Time Analysis: TIOGA



- TIOGA: “where it forks”/”swift current”, ZMQ
- 2 88-core and a 36-core DELL servers, 1 machine each for P14, TREXX, P12
- Configurable count of
  - Stream receiving and processing threads, and disk writing threads: 176/44
  - Frame buffers – limited by 256GB RAM
- Connections – 1 InfiniBand + 2 10 Gbit DCU to processing server
- Fully integrated in the TINE Control System
- DOZOR processing via library routines
  - DOZOR results exported every 300ms via TINE
    - Spot Count
    - DOZOR Score
    - Resolution Limit
    - Average Intensity
- No intermediate DOZOR and CBF files – all in memory
- Max count of frames per scan: 2,560,000 – TREXX
- Max frame rate – over 500 DOZOR processed and written fps
- Writing over a 40Gbit InfiniBand
- Data to Parallel Cluster File System Storage
  - 1 PB, 7GB/s BeeGFS

```
extern "C" {
    void dozor_set_defaults_(DATACOL_t*);
    void pre_dozor_(DETECTOR_t*, DATACOL_t*, LOCAL_t*, char*, char*, int*);
    void dozor_bin_image_(DETECTOR_t*, int*, int*, short*);
    void dozor_do_image_(short*, DETECTOR_t*, DATACOL_t*, DATACOL_t*, DATACOL_PICKLE_t*, LOCAL_t*, char*, char*);
    void dozor_get_spot_list_(DETECTOR_t*, DATACOL_t*, DATACOL_PICKLE_t*, reflection_t*);
}
```

BeeGFS® developed by Fraunhofer

The Baddest Daemon

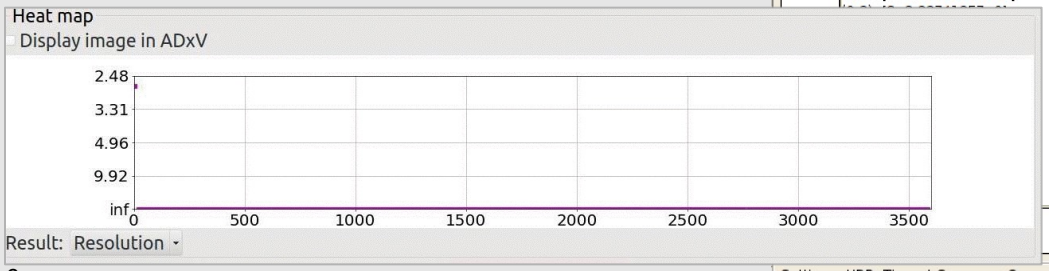
EMBL

Data

- Images
  - CBF/GZ-compressed CBF – P14, TREXX
  - Compressed TIFF – P12

- Spot Files
  - file_name.adx.gz

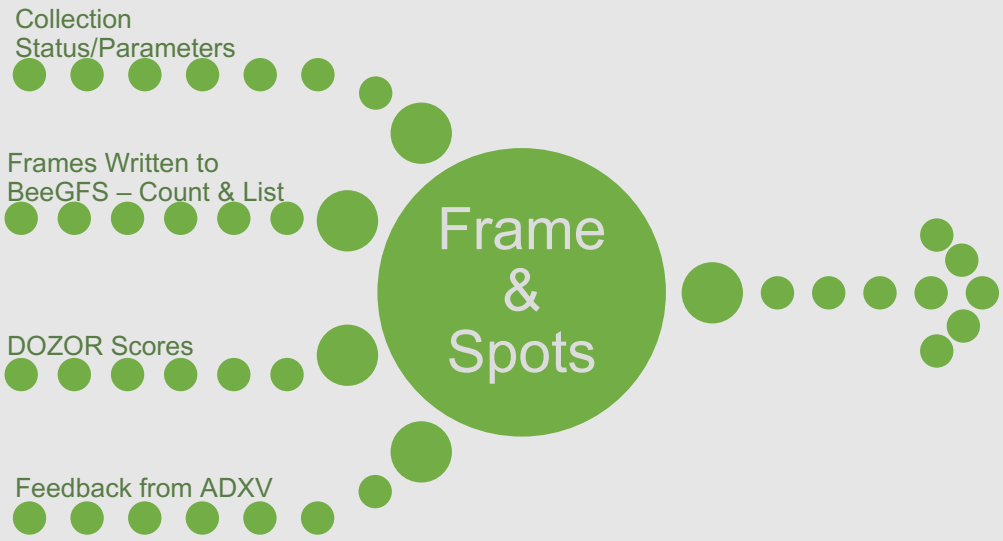
2303.87 1975.71 1 6153.75
1866.46 2055.64 1 1695.58
2328.09 2061.97 1 118.11
…

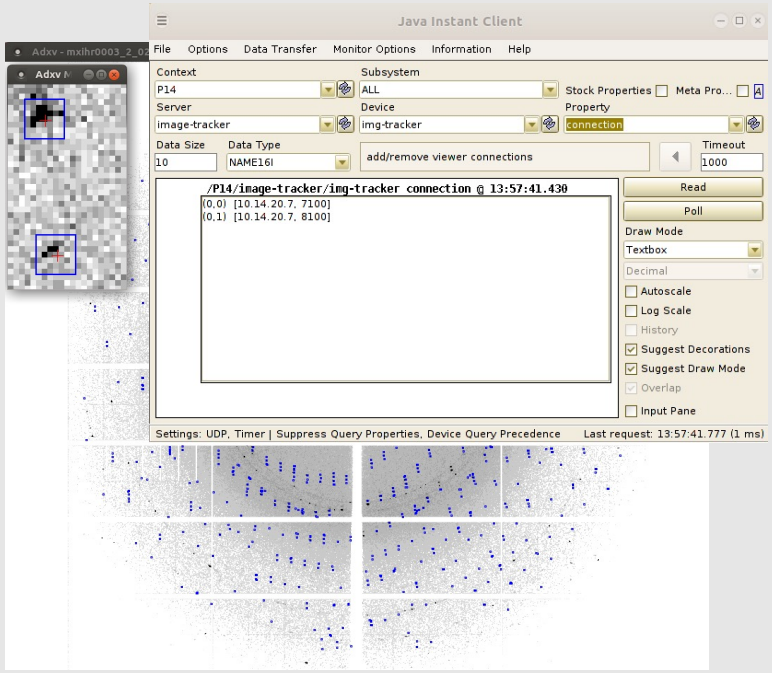


Live Frame Display

The Smart Daemon

- ADXV – data read from BeeGFS on the fly during scan
- Image Tracking Server
  - Talks to Detector/Data Collection Server, and TIOGA via TINE, and to ADXV via socket and NCAT
  - Selects preferentially frames with highest DOZOR score
  - Send Requests to ADXV to display
    - frames
    - spot highlights
  - Heat Map Hovering in GUI: bitmap of frames with diffraction for 32 data collections
  - Can feed multiple ADXV instances
- Advantages for Serial Crystallography Experiments
  - Few(er) empty frames to look at
  - Quick and Easy Spotting of Diffraction





EMBL

Raster Scan

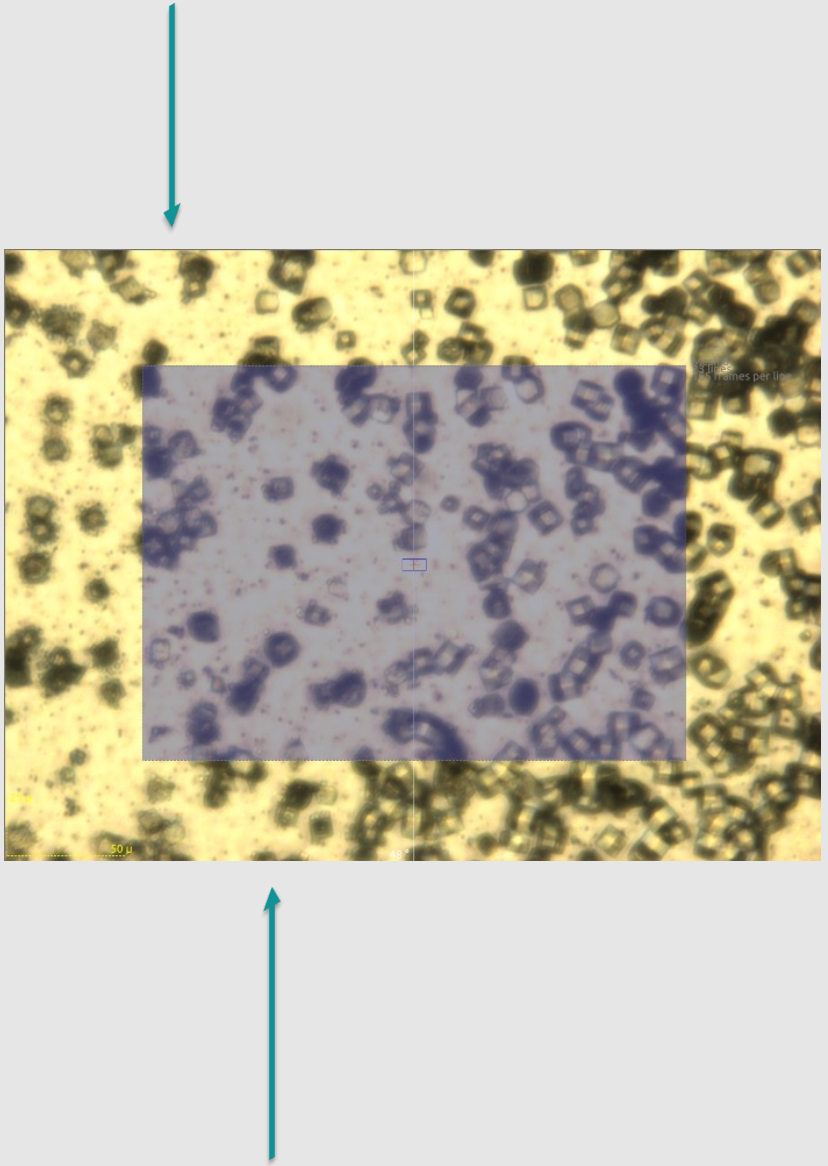
Exposure Time: 0.001343s

ROI Mode: 4M

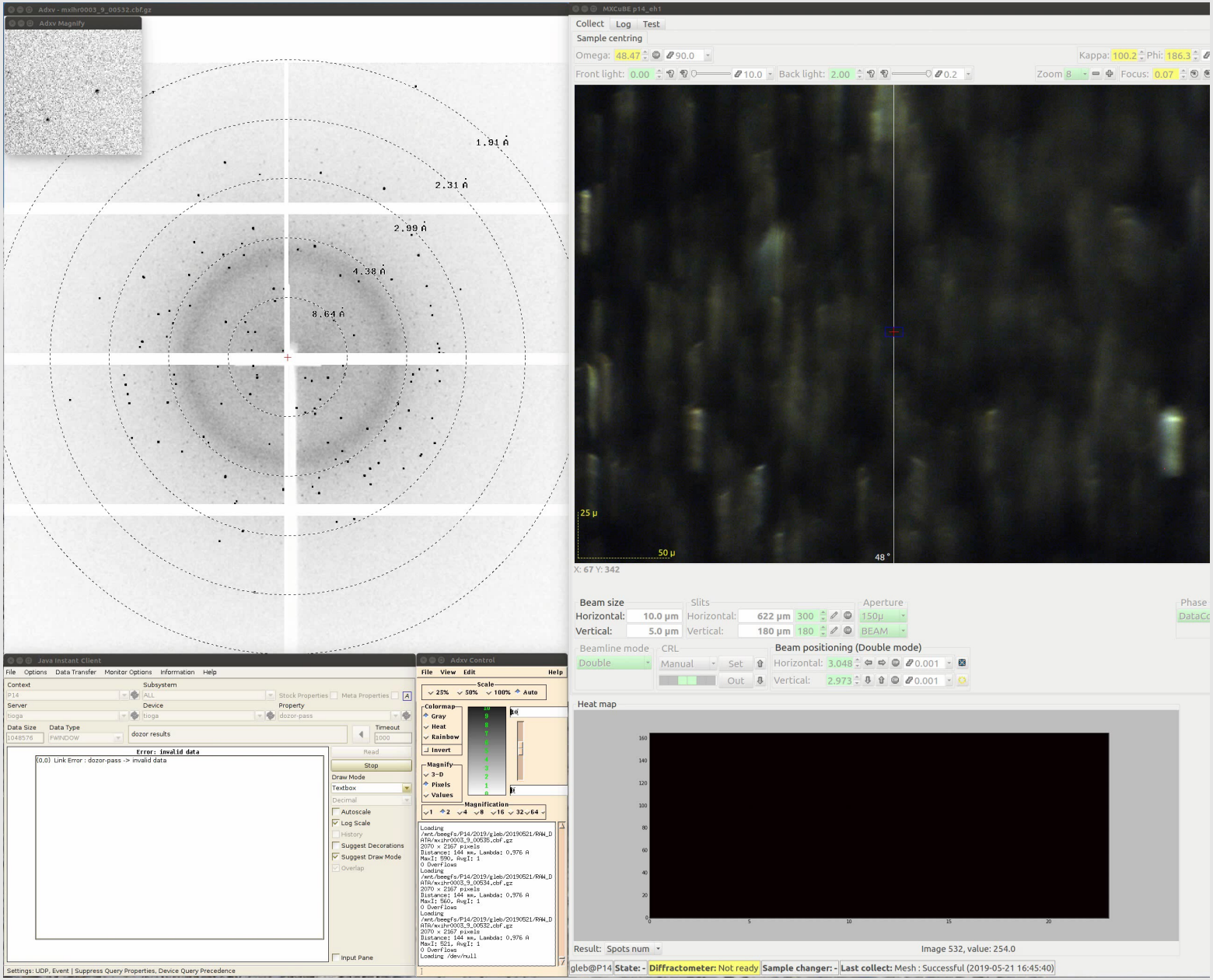
Lines: 23

Frames per Line: 165

Total Frames: 3795



EMBL


Adxv - mxlhr0003_9_00532.cbf.gz
Adxv Magnify
1.91 Å
2.31 Å
2.99 Å
4.38 Å
8.64 Å
MXCuBE p14_eh1
Collect
Log
Test
Sample centring
Omega: 48.47
Kappa: 100.2
Phi: 186.3
Front light: 0.00
Back light: 2.00
Zoom 8
Focus: 0.07
25 µ
50 µ
48 °
X: 67 Y: 342
Beam size
Horizontal: 10.0 µm
Vertical: 5.0 µm
Slits
Horizontal: 622 µm
Vertical: 180 µm
Aperture
150µ
BEAM
Beamline mode
Double
CRL
Manual
Set
Out
Beam positioning (Double mode)
Horizontal: 3.048
Vertical: 2.973
Heat map
Result: Spots num
Image 532, value: 254.0
gleb@P14
State: -
Diffractometer: Not ready
Sample changer: -
Last collect: Mesh : Successful (2019-05-21 16:45:40)
Java Instant Client
Error: invalid data
(0,0) Link Error : dozor-pass -> invalid data
Adxv Control

EMBL

Wedges

Multi-trigger Experiments

- Multiple Wedge Data Collections as required by GPhL workflows
- Detector/Data Collection Server prepares EIGER
  - # detector triggers = # wedges
  - # images = # images per wedge
- TIOGA figures out correct headers and file name numbering

```
nikolova@p14mw
File  Edit  View  Search  Terminal  Help
[nikolova@p14mw03 characterisation]$ ls -1 .
mxihr0021_1_00001.cbf.gz
mxihr0021_1_00002.cbf.gz
mxihr0021_1_00003.cbf.gz
mxihr0021_1_00004.cbf.gz
mxihr0021_1_00005.cbf.gz
mxihr0021_1_00006.cbf.gz
mxihr0021_1_00007.cbf.gz
mxihr0021_1_00008.cbf.gz
mxihr0021_1_00009.cbf.gz
mxihr0021_1_00010.cbf.gz
mxihr0021_1_00011.cbf.gz
mxihr0021_1_00012.cbf.gz
mxihr0021_1_00451.cbf.gz
mxihr0021_1_00452.cbf.gz
mxihr0021_1_00453.cbf.gz
mxihr0021_1_00454.cbf.gz
mxihr0021_1_00455.cbf.gz
mxihr0021_1_00456.cbf.gz
mxihr0021_1_00457.cbf.gz
mxihr0021_1_00458.cbf.gz
mxihr0021_1_00459.cbf.gz
mxihr0021_1_00460.cbf.gz
mxihr0021_1_00461.cbf.gz
mxihr0021_1_00462.cbf.gz
mxihr0021_1_00901.cbf.gz
mxihr0021_1_00902.cbf.gz
mxihr0021_1_00903.cbf.gz
mxihr0021_1_00904.cbf.gz
mxihr0021_1_00905.cbf.gz
mxihr0021_1_00906.cbf.gz
mxihr0021_1_00907.cbf.gz
mxihr0021_1_00908.cbf.gz
mxihr0021_1_00909.cbf.gz
mxihr0021_1_00910.cbf.gz
mxihr0021_1_00911.cbf.gz
mxihr0021_1_00912.cbf.gz
mxihr0021_1_01351.cbf.gz
mxihr0021_1_01352.cbf.gz
mxihr0021_1_01353.cbf.gz
mxihr0021_1_01354.cbf.gz
mxihr0021_1_01355.cbf.gz
mxihr0021_1_01356.cbf.gz
mxihr0021_1_01357.cbf.gz
mxihr0021_1_01358.cbf.gz
mxihr0021_1_01359.cbf.gz
mxihr0021_1_01360.cbf.gz
mxihr0021_1_01361.cbf.gz
mxihr0021_1_01362.cbf.gz
mxihr0021_1_01801.cbf.gz
mxihr0021_1_01802.cbf.gz
mxihr0021_1_01803.cbf.gz
mxihr0021_1_01804.cbf.gz
mxihr0021_1_01805.cbf.gz
mxihr0021_1_01806.cbf.gz
mxihr0021_1_01807.cbf.gz
mxihr0021_1_01808.cbf.gz
mxihr0021_1_01809.cbf.gz
mxihr0021_1_01810.cbf.gz
mxihr0021_1_01811.cbf.gz
mxihr0021_1_01812.cbf.gz
```

1st Wedge

2nd Wedge

3rd Wedge

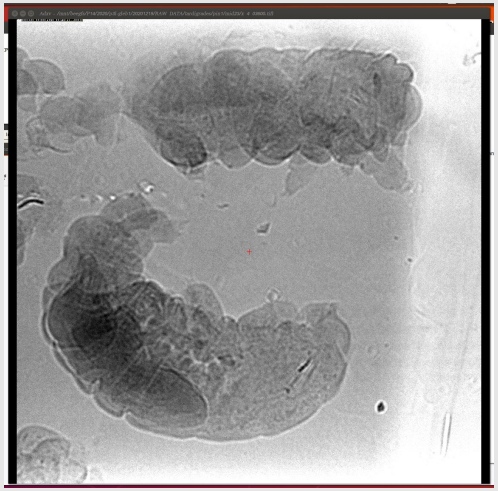
4th Wedge

5th Wedge

EMBL

Imaging for MX

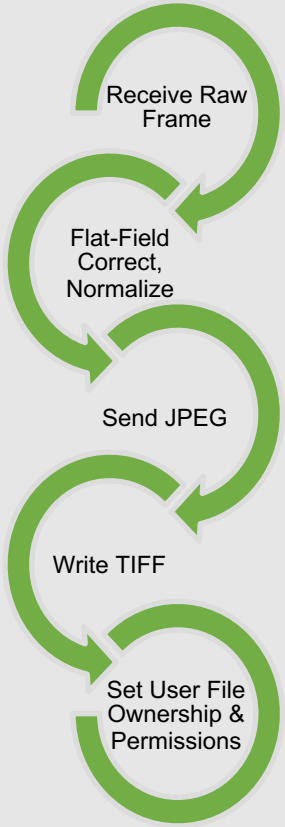
The Down-To-Earth Daemon



PCO-SERVER

Receives settings/commands from detector server

Talks directly to PCO camera
Sends acquired frames to Arapaho
ZMQ over 40Gbit InfiniBand
Keeps a buffer for 360 frames of depth 2 bytes
Reports progress to clients
Runs on a 20-core/40-hyperthread machine with 96GB RAM



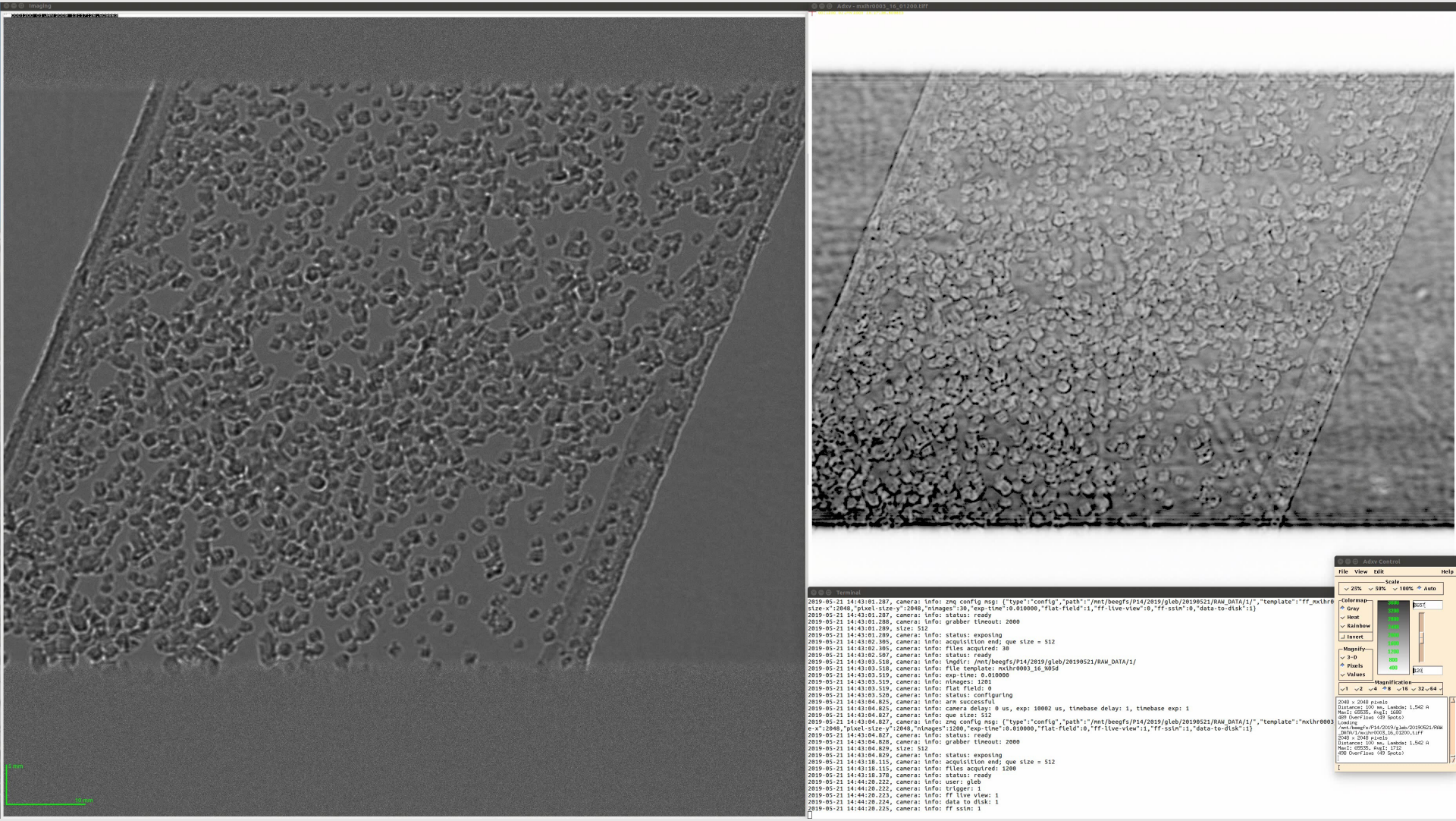
The Copycat Daemon

ARAPAHO

- A TINE, ZMQ server
- Transfers PCO camera frames from pco server to BeeGFS
- Flat field processing based on
  - Structural Similarity Index (SSIM)*
  - MPFR big number library
- Exports flat field SSIM and saturation count to clients
- Sends flat-field-corrected frames to GUI via TINE for display

* Image Quality Assessment: From Error Visibility to Structural Similarity, Wang, Z.
http://www.cns.nyu.edu/pub/eero/wang03-reprint.pdf

EMBL


Imaging
Adxv - mxihr0003_16_01200.tif
Terminal
Adxv Control
File View Edit Help
Scale
25% 50% 100% Auto
Colormap
Gray
Heat
Rainbow
Invert
Magnify
3-D
Pixels
Values
Magnification
1 2 4 8 16 32 64
2048 x 2048 pixels
Distance: 100 mm, Lambda: 1.542 A
Max: 65535, Avg: 1688
489 Overflows (49 Spots)
Loading /mnt/beegfs/P14/2019/gleb/20190521/RAW_DATA/1/mxihr0003_16_01200.tif
2048 x 2048 pixels
Distance: 100 mm, Lambda: 1.542 A
Max: 65535, Avg: 1712
498 Overflows (49 Spots)
2019-05-21 14:43:01.287, camera: info: zmq config msg: {"type":"config","path":"/mnt/beegfs/P14/2019/gleb/20190521/RAW_DATA/1/","template":"ff_mxihr0
size-x":2048,"pixel-size-y":2048,"nimages":30,"exp-time":0.010000,"flat-field":1,"ff-live-view":0,"ff-ssim":0,"data-to-disk":1}
2019-05-21 14:43:01.287, camera: info: status: ready
2019-05-21 14:43:01.288, camera: info: grabber tineout: 2000
2019-05-21 14:43:01.289, size: 512
2019-05-21 14:43:01.289, camera: info: status: exposing
2019-05-21 14:43:02.305, camera: info: acquisition end; que size = 512
2019-05-21 14:43:02.305, camera: info: files acquired: 30
2019-05-21 14:43:02.507, camera: info: status: ready
2019-05-21 14:43:03.518, camera: info: lngdir: /mnt/beegfs/P14/2019/gleb/20190521/RAW_DATA/1/
2019-05-21 14:43:03.518, camera: info: file template: nxihr0003_16_%05d
2019-05-21 14:43:03.519, camera: info: exp-tine: 0.010000
2019-05-21 14:43:03.519, camera: info: nimages: 1201
2019-05-21 14:43:03.519, camera: info: flat field: 0
2019-05-21 14:43:03.520, camera: info: status: configuring
2019-05-21 14:43:04.825, camera: info: arm successful
2019-05-21 14:43:04.825, camera: info: camera delay: 0 us, exp: 10002 us, timebase delay: 1, timebase exp: 1
2019-05-21 14:43:04.827, camera: info: que size: 512
2019-05-21 14:43:04.827, camera: info: zmq config msg: {"type":"config","path":"/mnt/beegfs/P14/2019/gleb/20190521/RAW_DATA/1/","template":"mxihr0003
e-x":2048,"pixel-size-y":2048,"nimages":1200,"exp-tine":0.010000,"flat-field":0,"ff-live-view":1,"ff-ssim":1,"data-to-disk":1}
2019-05-21 14:43:04.827, camera: info: status: ready
2019-05-21 14:43:04.828, camera: info: grabber tineout: 2000
2019-05-21 14:43:04.829, size: 512
2019-05-21 14:43:04.829, camera: info: status: exposing
2019-05-21 14:43:18.115, camera: info: acquisition end; que size = 512
2019-05-21 14:43:18.115, camera: info: files acquired: 1200
2019-05-21 14:43:18.378, camera: info: status: ready
2019-05-21 14:44:20.222, camera: info: user: gleb
2019-05-21 14:44:20.222, camera: info: trigger: 1
2019-05-21 14:44:20.223, camera: info: ff live view: 1
2019-05-21 14:44:20.224, camera: info: data to disk: 1
2019-05-21 14:44:20.225, camera: info: ff ssim: 1

EMBL

Automatic Data Processing

- Beamline Cluster
  - 13 88-core DELL R840 servers
  - Data read from BeeGFS over 56 Gbit InfiniBand

- XDS-ZCBF
  - An XDS and DOZOR plugin for fast in-memory decompression of GZIPped CBFs
  - No intermediate files on disk
  - Deployed for automatic data processing on beamline cluster
  - Usage via line in XDS.INP much like HDF5s NEGGIA and DURIN plugins
    - LIB=/mx-beta/lib/xds-zcbf.so

- https://git.embl.de/nikolova/xds-zcbf

EMBL

Outlook

- Upcoming Software Updates
  - DOZOR Saturated Spot Count
  - Multiple Energies and Inverse Beam via Multi-trigger
  - TIOGA Features also for Pilatus

- New Detector and Camera Integration
  - 16M Eiger CdTe, DECTRIS, @ P14
  - 4M Jungfrau, PSI, @ TREXX
  - 3M Visual Light Motion Camera XStream Mini, Imaging Solutions, @ TREXX

EMBL

Summary

- Instantaneous Availability of Frames for Display and Processing
- Immediate Delivery of Processing Results
- Full Integration of MX Analysis in the Control System
- Streamlined Path for Data
- Data Wave Tamed

EMBL

Acknowledgements

Gleb Bourenkov

Ivars Karpics
Maxim Polikarpov
Andrea d'Amato
Daniel Franke
David von Stetten
Michael Agthe
Thomas R. Schneider
EMBL-HH Instrumentation, and IT Groups
Stefan Fiedler

Andrew Arvai
Phil Duval
Sasha Popov
Arwen Pearson & The University of Hamburg
Kay Diederichs
Global Phasing
EMBL-Grenoble former & current Instrumentation Groups
DECTRIS
BMBF

Bundesministerium für Bildung und Forschung

Thank You!

EMBL