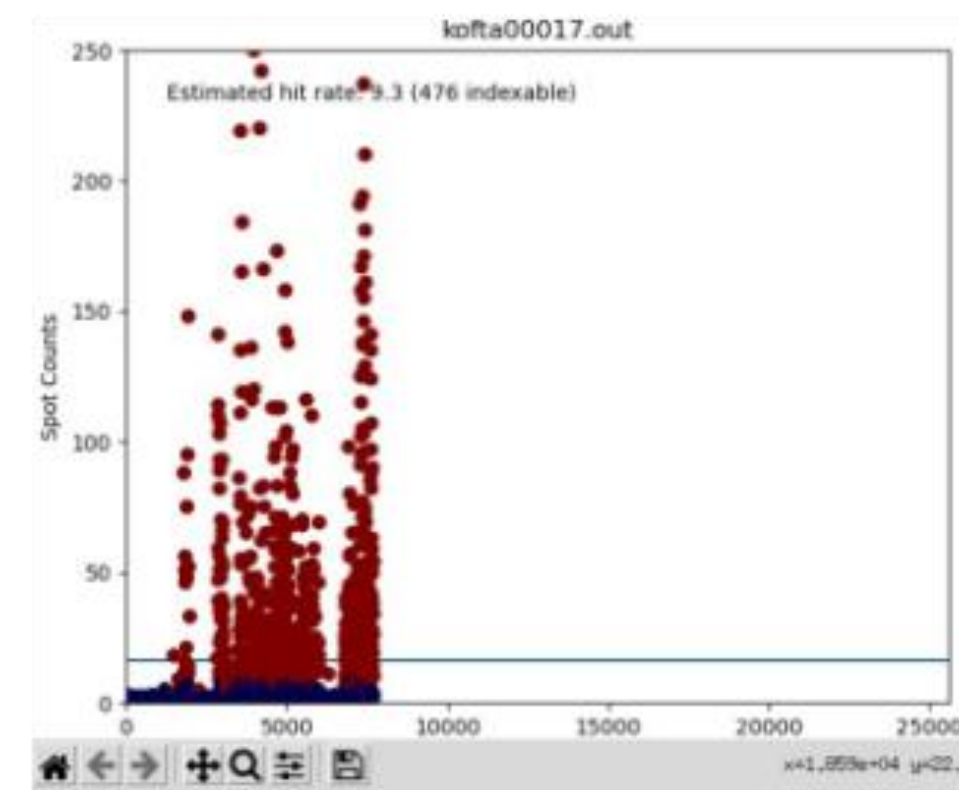
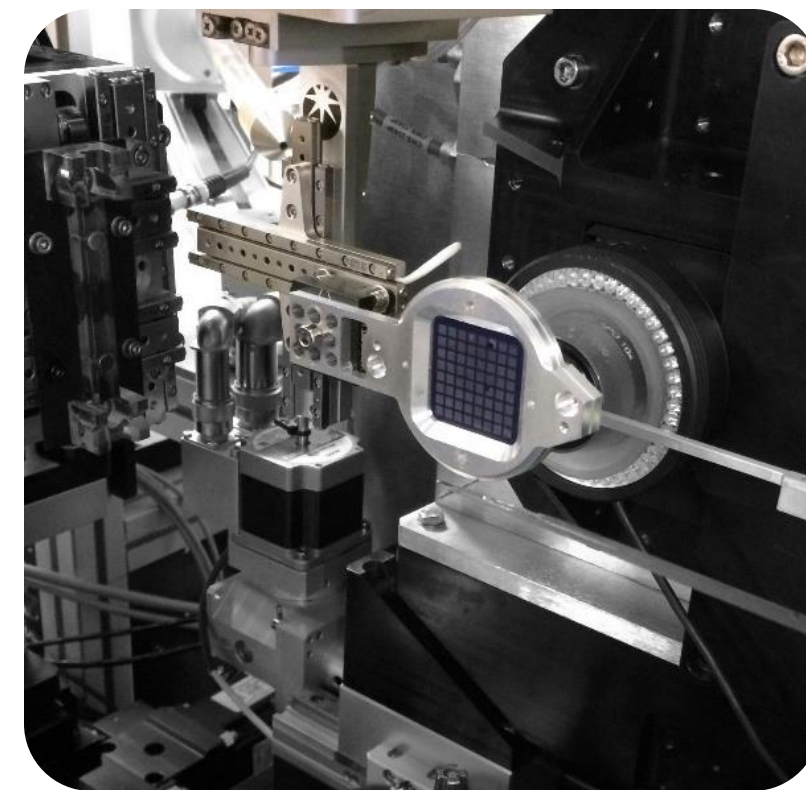
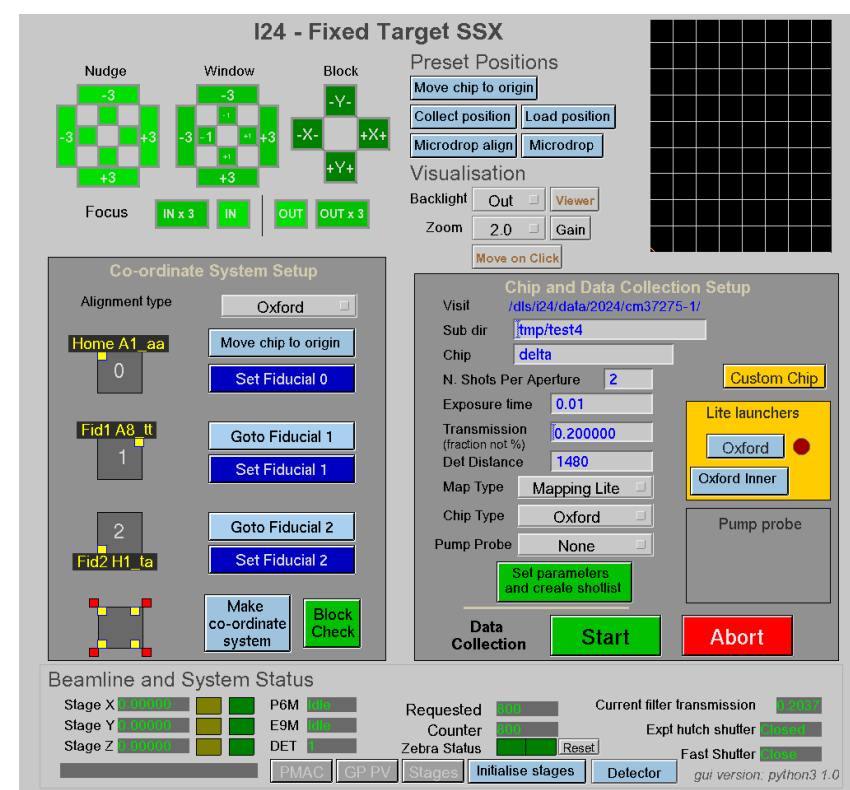


# Automation for serial crystallography data collection and processing at Diamond Light Source



----- Merging -----  
 7850 crystals in total scaled in space group P 43 21 2  
 Median cell: 79.226, 79.226, 38.161, 90.0, 90.0, 90.0  
 Merged mtz file: merged.mtz

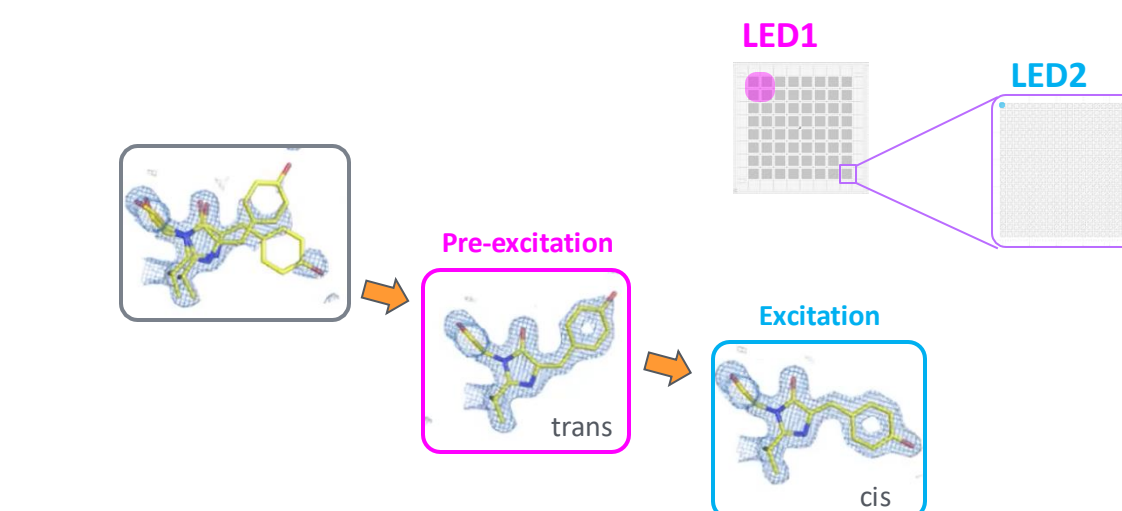
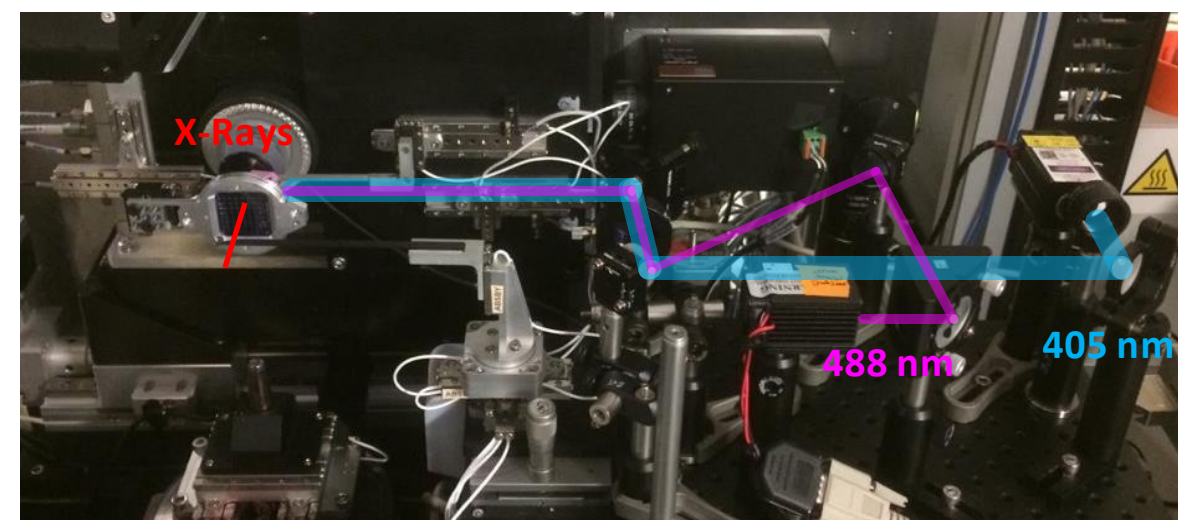
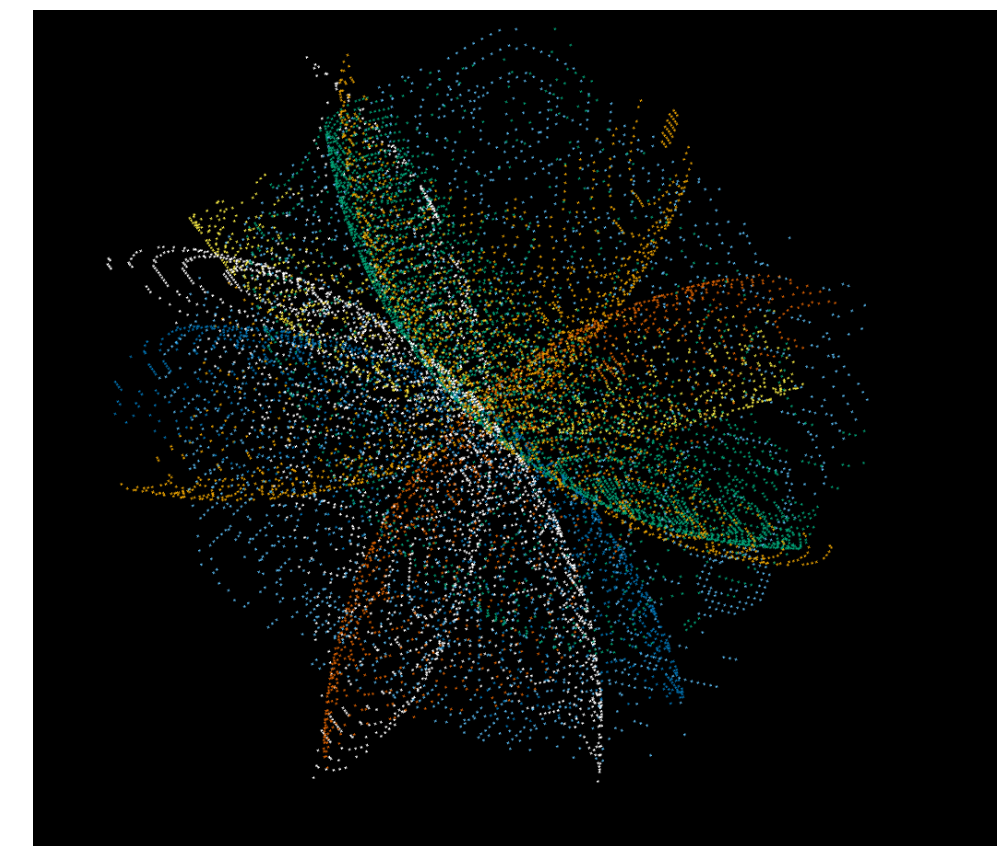
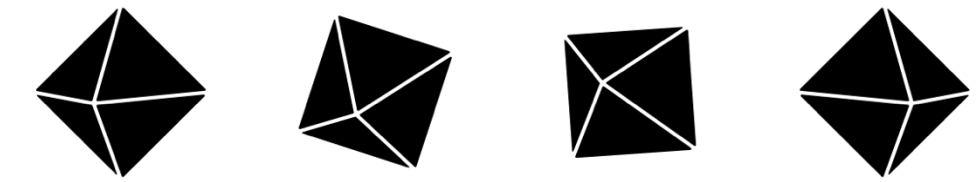
	Suggested	Low	High	Overall
High resolution limit	1.60	4.34	1.60	1.32
Low resolution limit	56.02	56.06	1.63	56.02
Completeness	100.0	100.0	100.0	94.6
Multiplicity	94.1	169.9	39.1	61.5
I/sigma	20.9	69.7	1.4	12.6
Rmerge(I)	0.964	0.662	20.751	1.038
Rmerge(I+/-)	0.962	0.661	20.711	1.035
Rmeas(I)	0.968	0.664	21.022	1.045
Rmeas(I+/-)	0.971	0.665	21.231	1.048
Rpim(I)	0.094	0.052	3.333	0.112
Rpim(I+/-)	0.127	0.068	4.640	0.152
Rsplit(I)	0.077	0.046	1.253	0.195
CC half	0.996	0.995	0.316	0.959
Wilson B factor	9.883			
Anomalous completeness	100.0	100.0	100.0	88.6
Anomalous multiplicity	50.9	103.2	20.6	33.8
Anomalous correlation	-0.017	-0.089	-0.031	-0.004
Anomalous slope	1.515			
dF/F	0.090			
dl/s(dl)	1.572			
Total observations	1558766	158714	31224	1710375
Total unique	16566	934	799	27793

# Serial crystallography

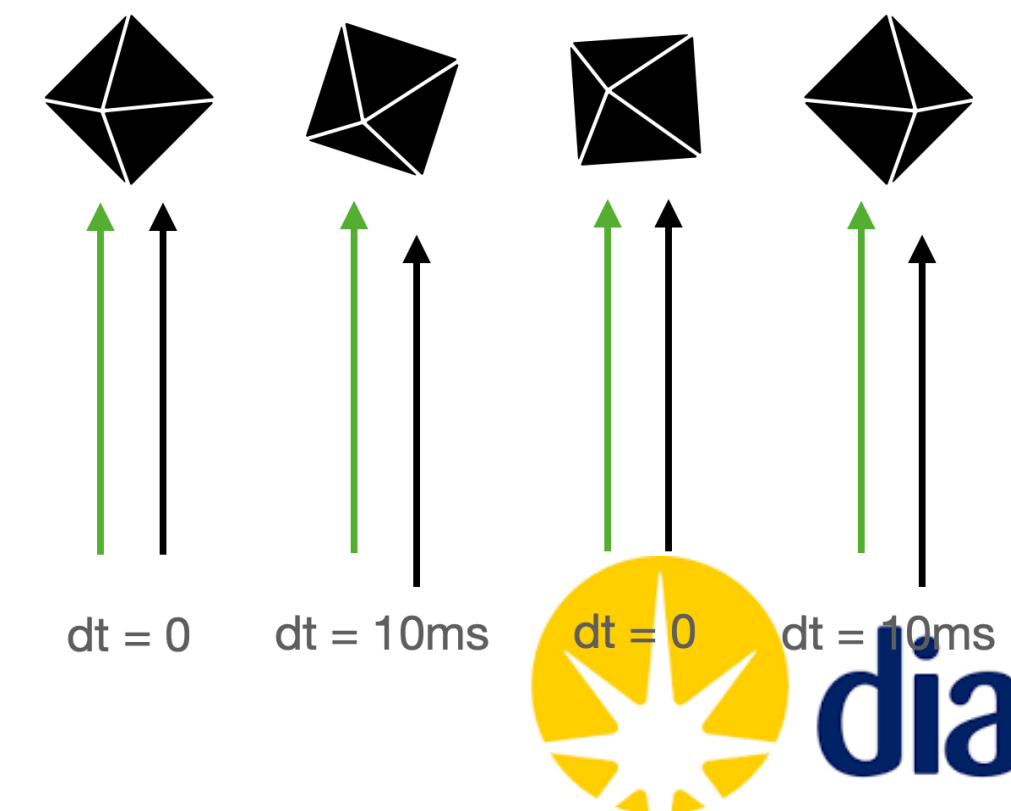
Serial still-shot crystallography is the measurement of 1 to  $N (< \sim 10)$  diffraction patterns from given crystal, with no rotation. A complete, high-resolution dataset requires measurements of 1000s of crystals.

This technique is of particular importance for the study of dynamic biology at room temperature with minimal radiation damage, where processes on the timescale of sub-ms to seconds can be investigated at a synchrotron, or down to femtosecond timescales at an XFEL.

Typical experiments include the investigation of light-activated processes, chemical mixing/binding dynamics, studying of the effect of dose on metalloproteins.



Baxter et al., *J. Phys. Chem. B*, 2022, 126, 45, 9288–9296



# Serial crystallography

## Challenges from an automation perspective

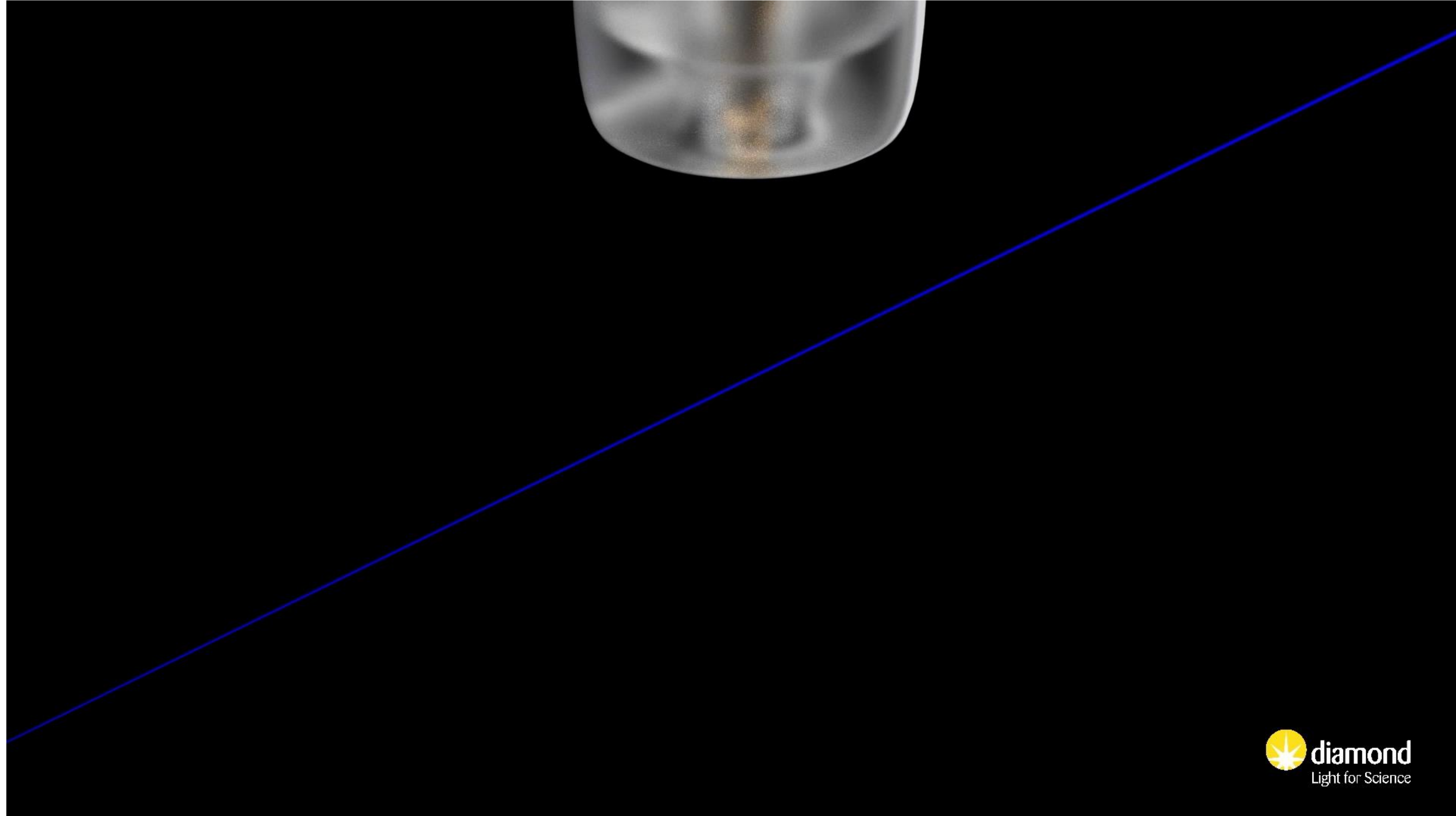
### **Speed**

In theory you could collect new crystals at the detector collection frequency – 500Hz for Eiger, >kHz for newer detectors. Challenge is then the automation of the sample delivery process to achieve high crystal refresh rate. Following on from that, fast data processing and feedback are required to guide the experiment and make best use of available resources.

### **Complexity**

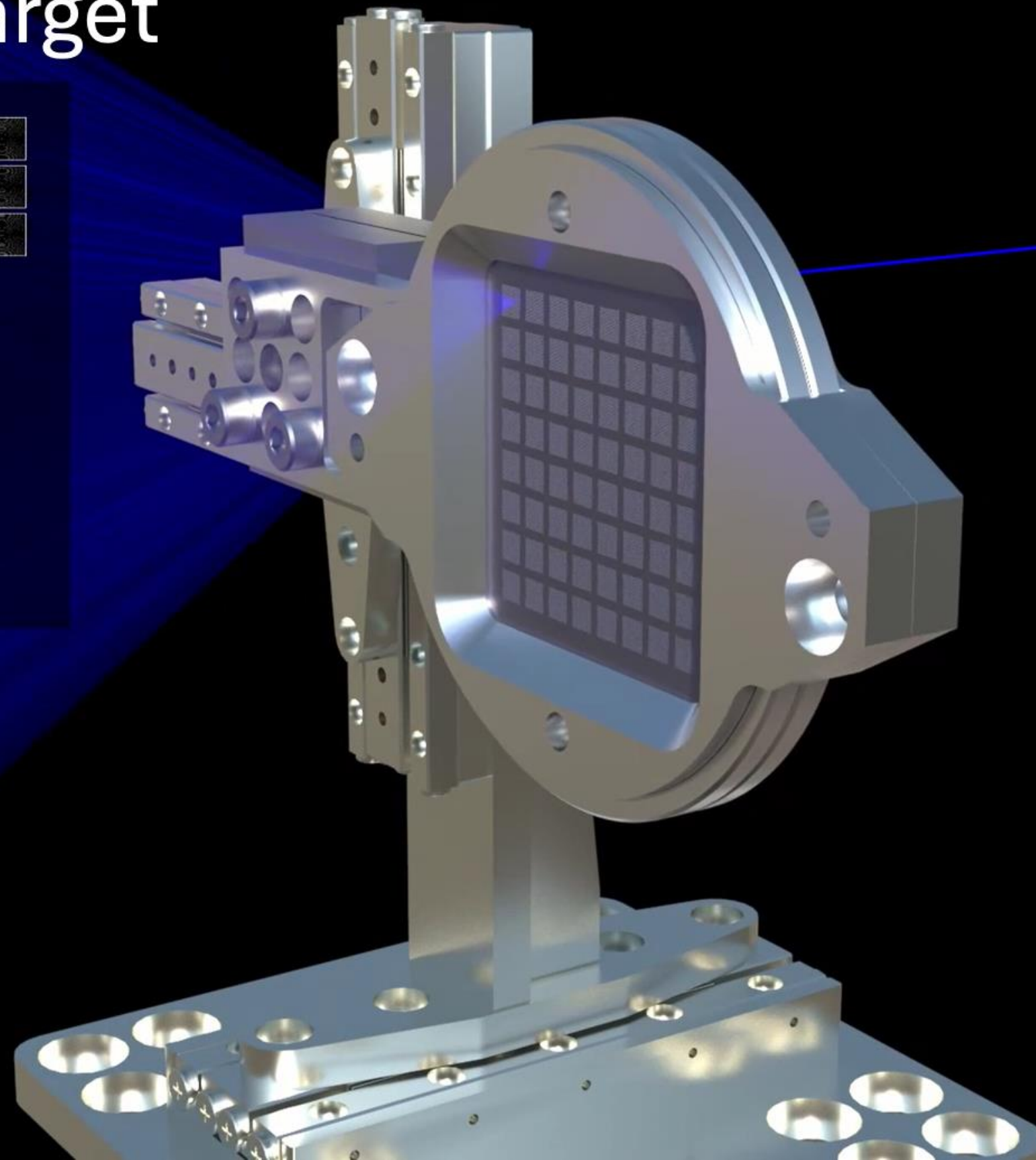
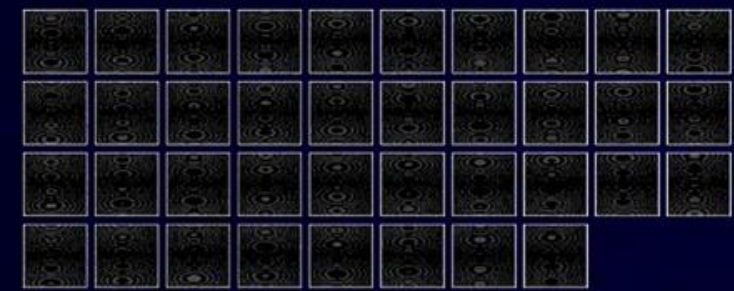
Inherently more complex experiment type compared to rotation crystallography, different hardware options depending on biological question of interest, this extra metadata needs to be handled in processing and display of results.

# Serial crystallography: extruder



# Serial crystallography: Fixed target

## SSX : Fixed Target

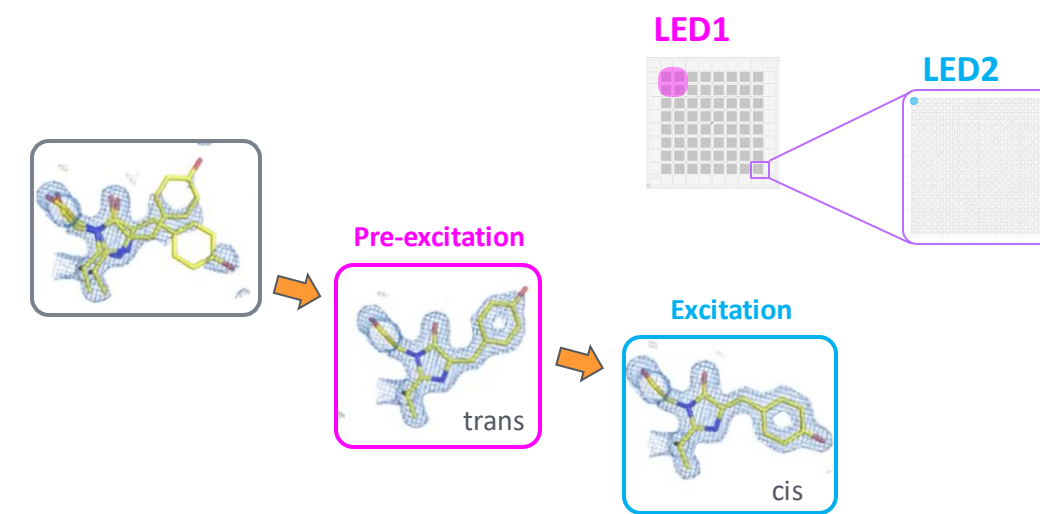
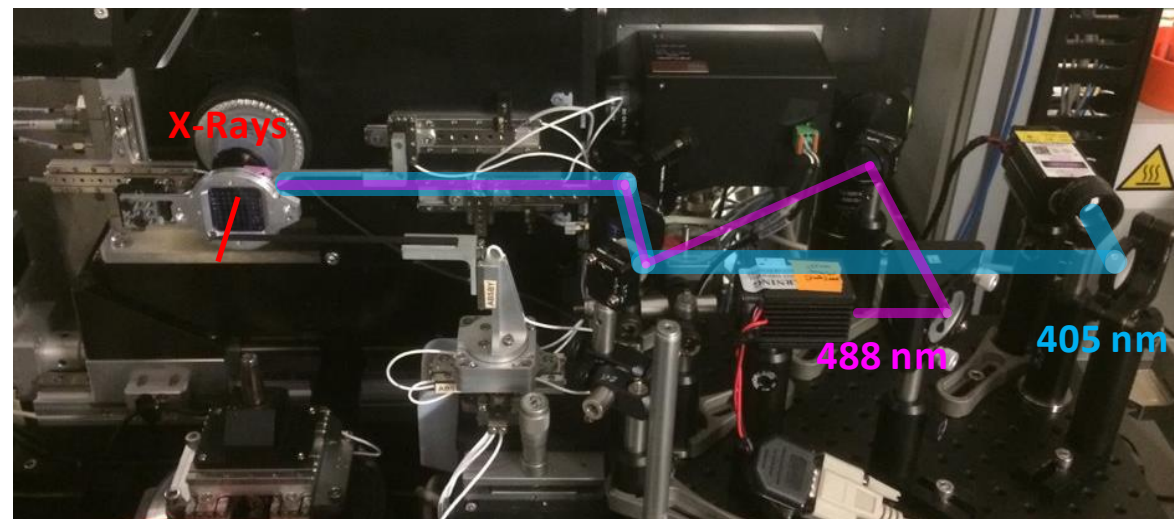


Chip 30 x 30 mm  
8 x 8 city blocks  
400 apertures on each city  
block  
Aperture size 7 – 20  $\mu\text{m}$   
25600 positions  
~ 8 min full data collection

Horrell et al., *JoVE*, 2021, 168, e62200  
Owen et al., *Acta Cryst.*, 2017, D73, 373-378  
Sherrell et al., *J. Synch. Rad.* 2015, 22, 1372-1378

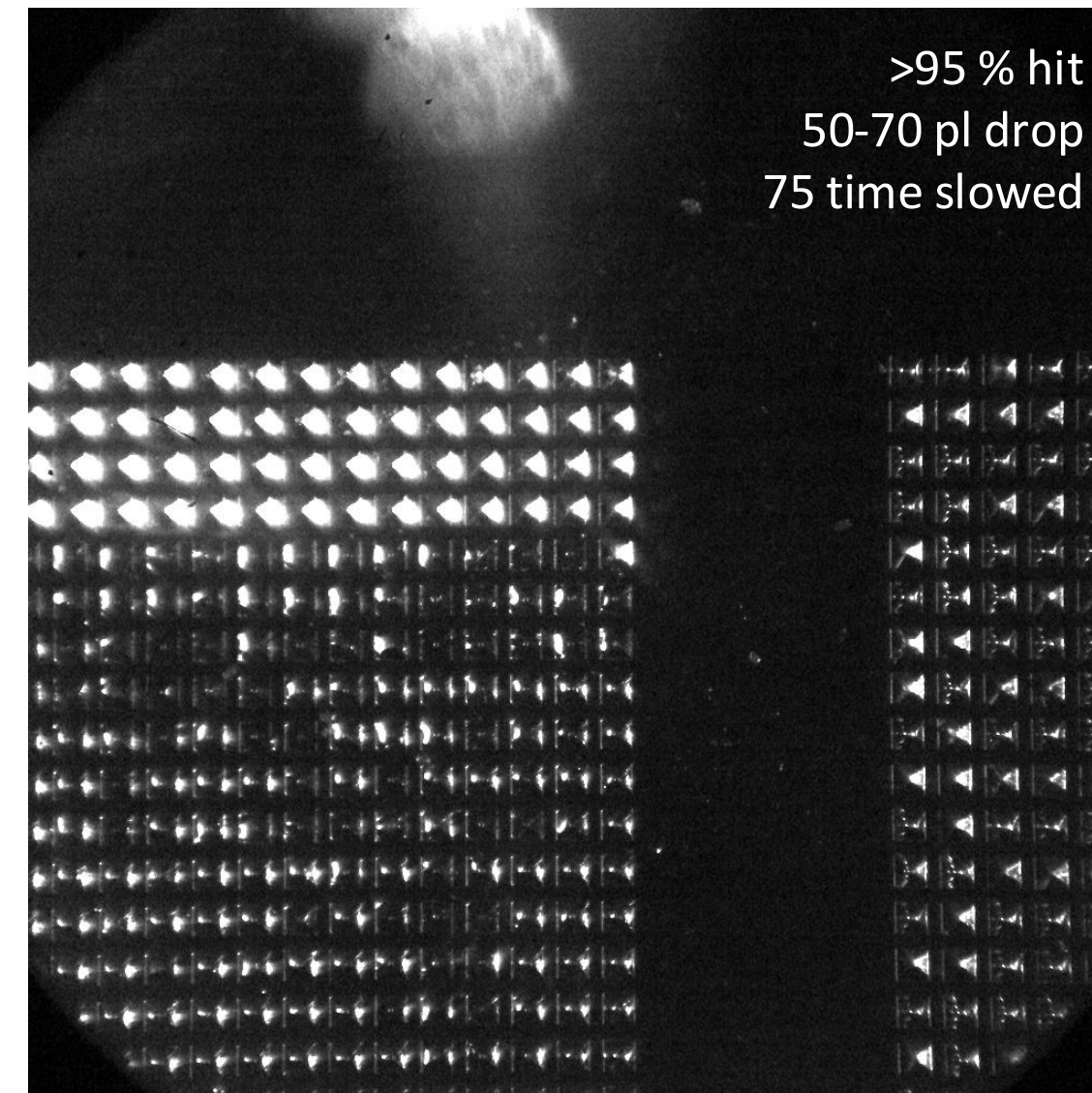
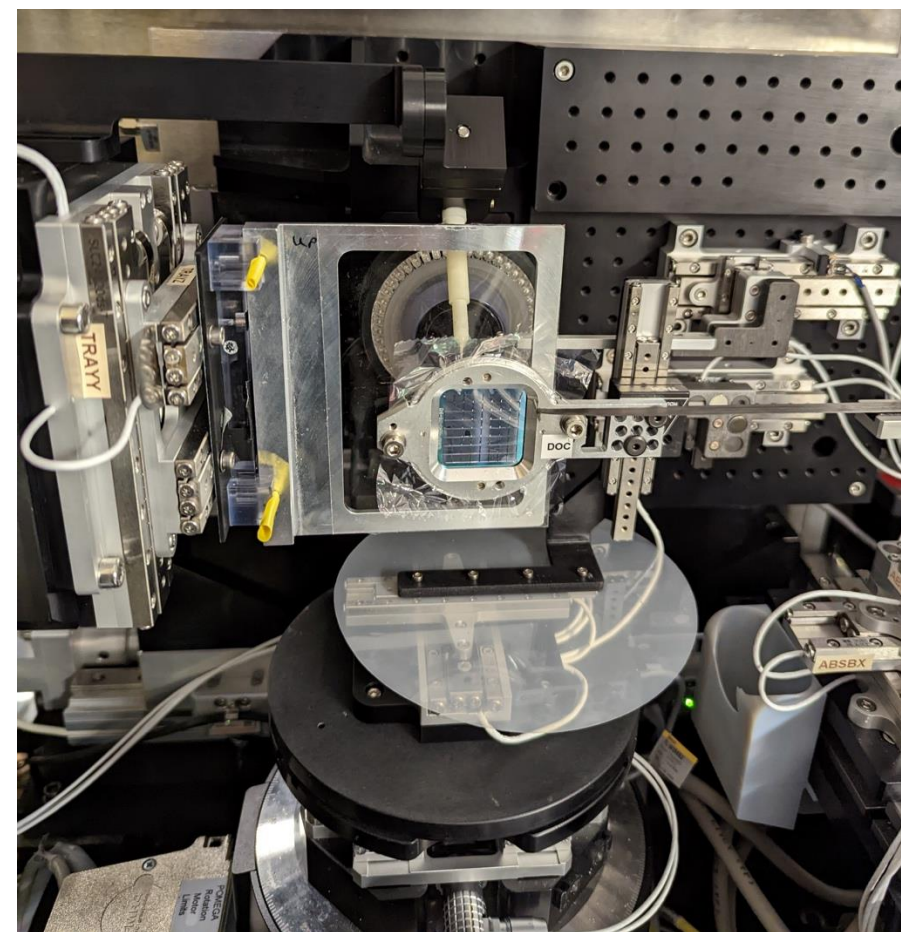


# Serial crystallography: add-ons



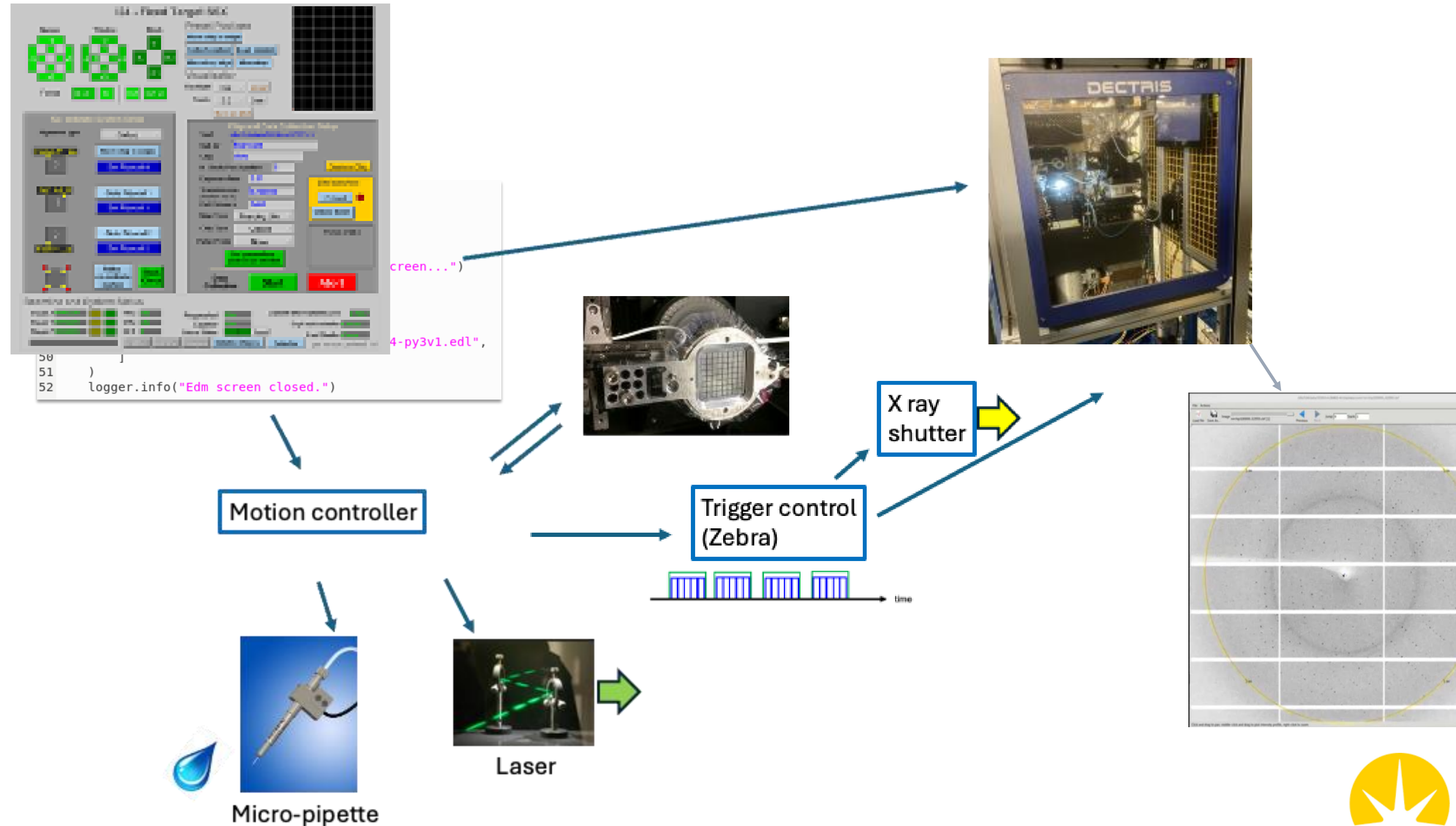
Baxter et al., *J. Phys. Chem. B*, 2022, 126, 45, 9288–9296

Laser excitation.



Rapid mixing with micro-drop ejector.

# Automated data collection: I24 experiment control



# Automated data collection: I24 experiment control

## Fixed target GUI

**I24 - Fixed Target SSX**

**Nudge** **Window** **Block**

Focus: IN x 3, IN, OUT, OUT x 3

**Preset Positions**

Move chip to origin  
Collect position  
Load position  
Microdrop align  
Microdrop

**Visualisation**

Backlight: Out, Viewer  
Zoom: 2.0, Gain  
Move on Click

**Co-ordinate System Setup**

Alignment type: Oxford

Home A1\_aa: 0  
Fid1 A8\_tt: 1  
Fid2 H1\_ta: 2

Move chip to origin  
Set Fiducial 0  
Goto Fiducial 1  
Set Fiducial 1  
Goto Fiducial 2  
Set Fiducial 2  
Make co-ordinate system  
Block Check

**Chip and Data Collection Setup**

Visit: /dls/i24/data/2024/cm37275-1/  
Sub dir: tmp/test4  
Chip: delta  
N. Shots Per Aperture: 2  
Exposure time: 0.01  
Transmission (fraction not %): 0.200000  
Det Distance: 1480  
Map Type: Mapping Lite  
Chip Type: Oxford  
Pump Probe: None

Custom Chip  
Lite launchers: Oxford, Oxford Inner  
Pump probe

Set parameters and create sholist

Data Collection: Start, Abort

**Beamline and System Status**

Stage X: 0.00000, P6M: On, Requested: 300, Counter: 300, Zebra Status: On, Current filler transmission: 0.20000, Expt hutch shutter: Closed, Fast Shutter: Close

Stage Y: 0.00000, E9M: On  
Stage Z: 0.00000, DET: On

PMAC, GP PV, Stages, Initialise stages, Detector  
gui version: python3 1.0

**Mapping Lite**

Define Chip In PMAC

Load Map File

Presets: All on, All off

Half Chip: 4x4, 6x6

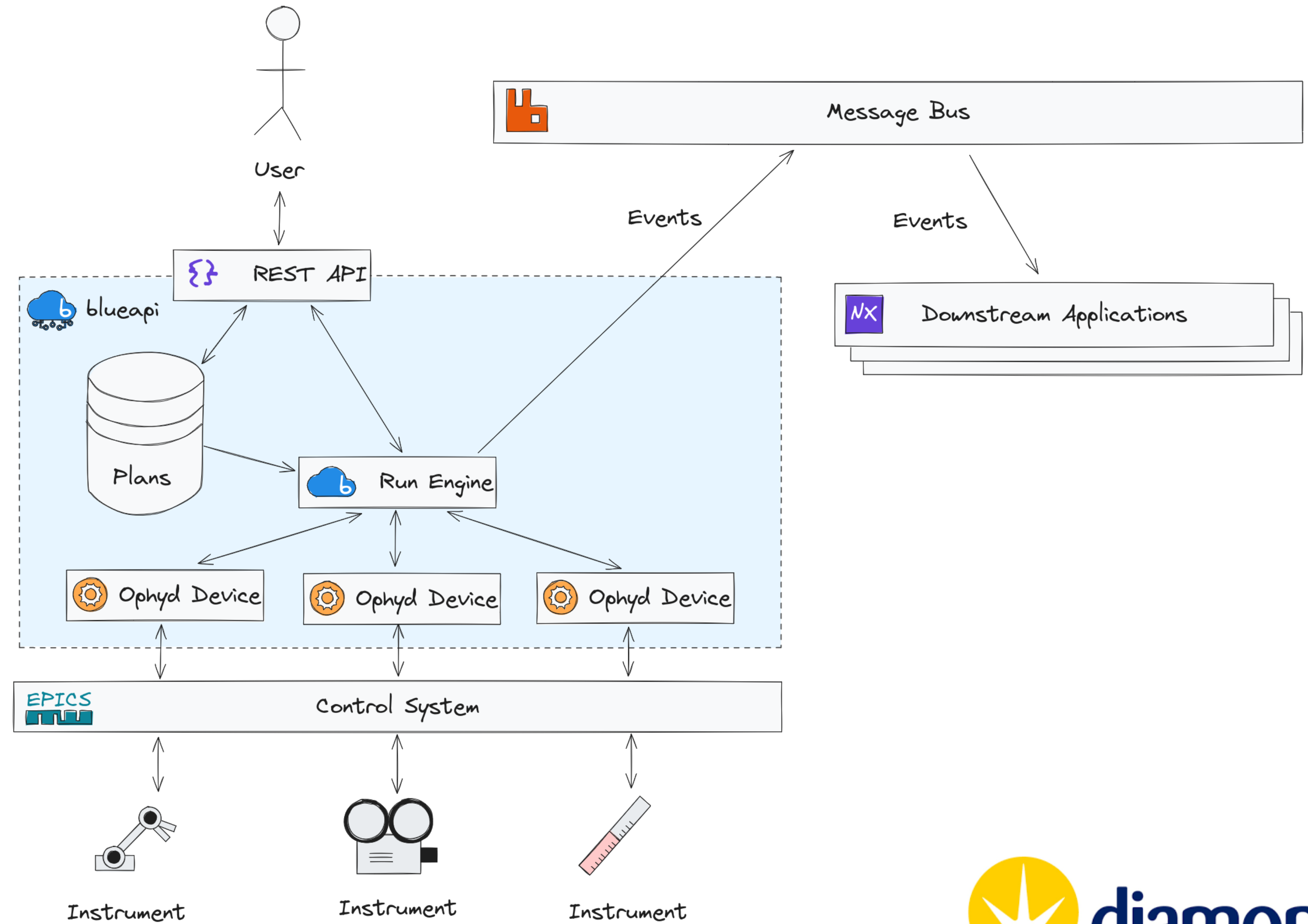
Save Screen Map

Upload Parameters

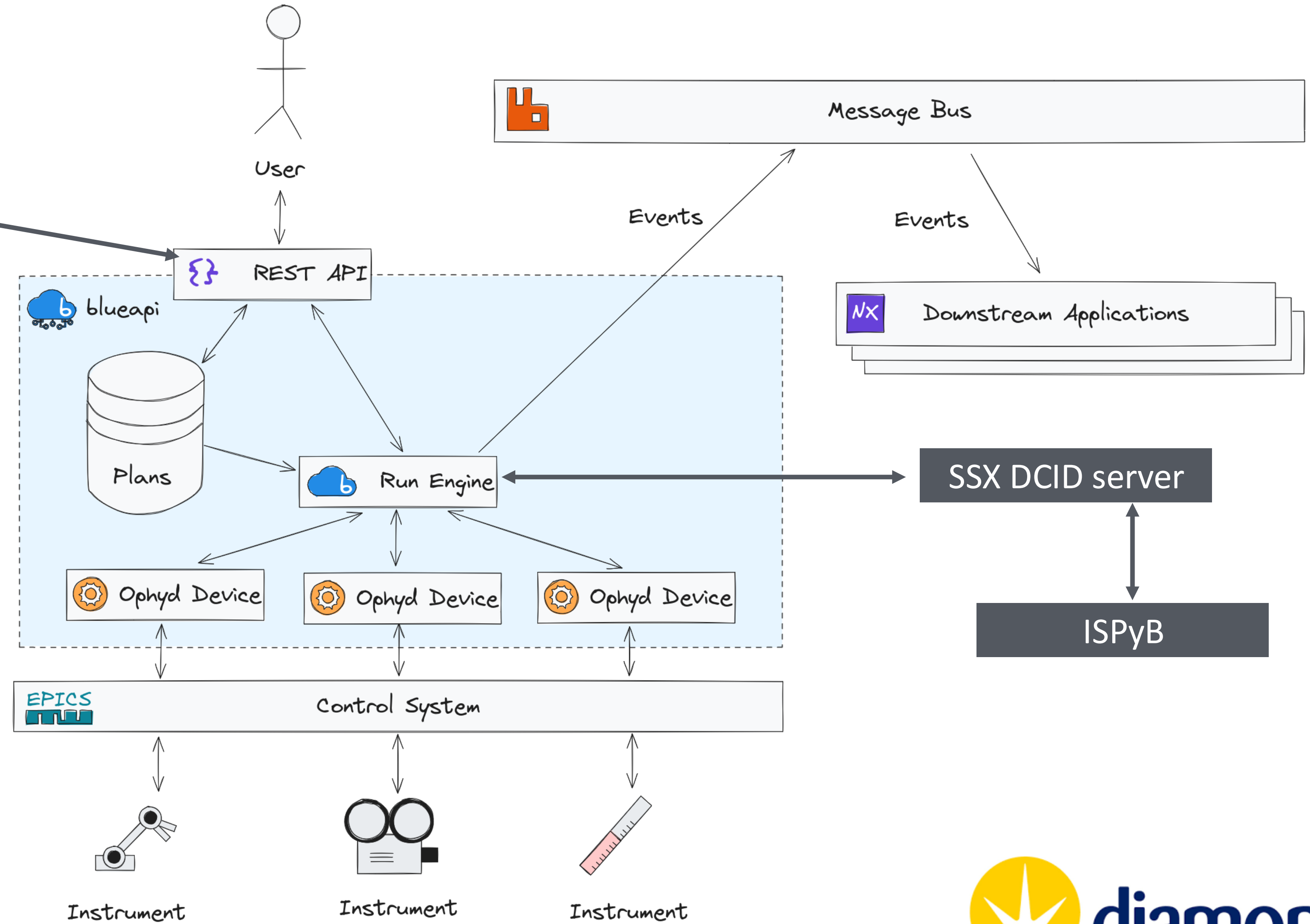
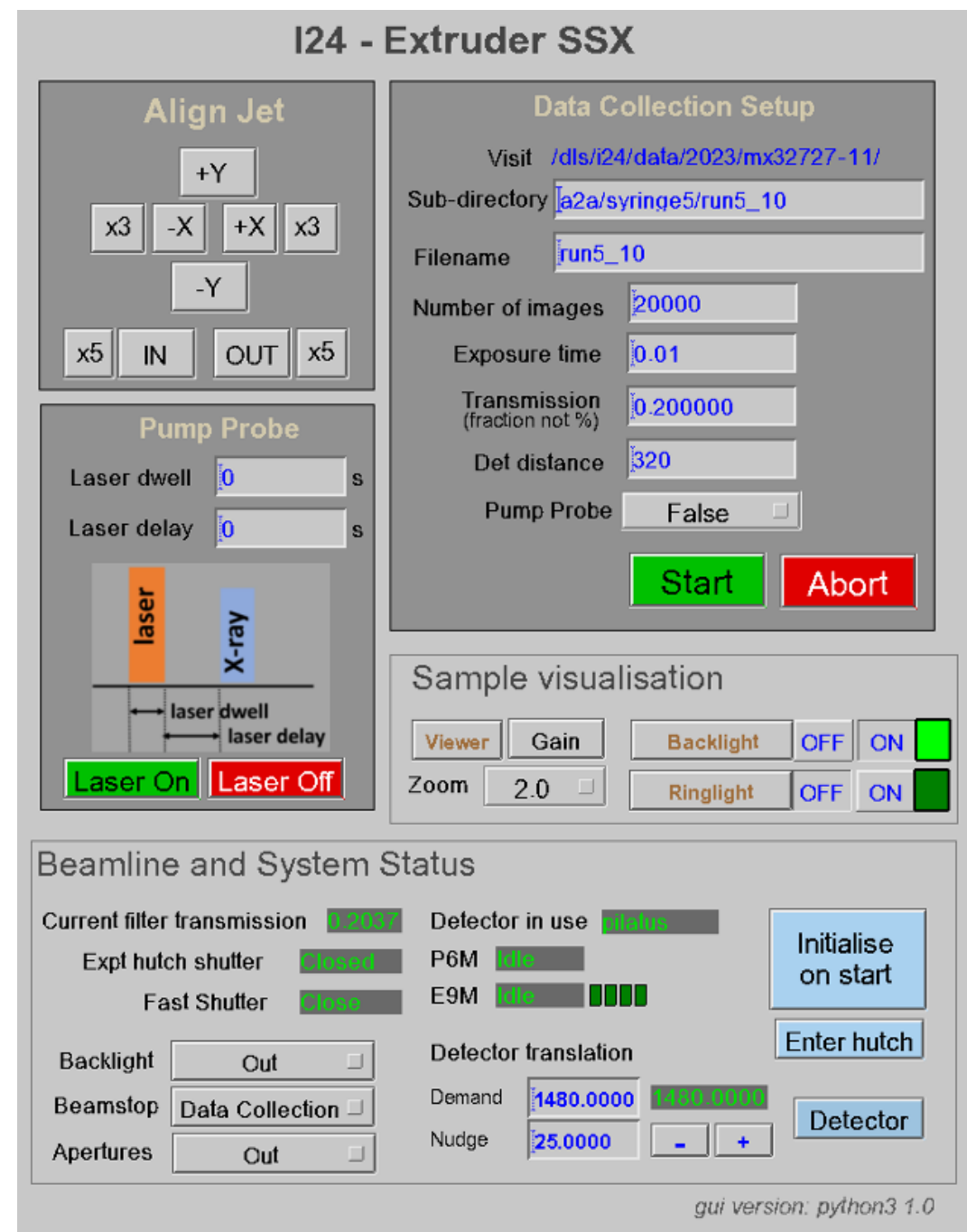
	1	2	3	4	5	6	7	8
A	01	16	17	32	33	48	49	64
B	02	15	18	31	34	47	50	63
C	03	14	19	30	35	46	51	62
D	04	13	20	29	36	45	52	61
E	05	12	21	28	37	44	53	60
F	06	11	22	27	38	43	54	59
G	07	10	23	26	39	42	55	58
H	08	09	24	25	40	41	56	57



# Automated data collection on I24: Bluesky



# Automated data collection on I24: Bluesky

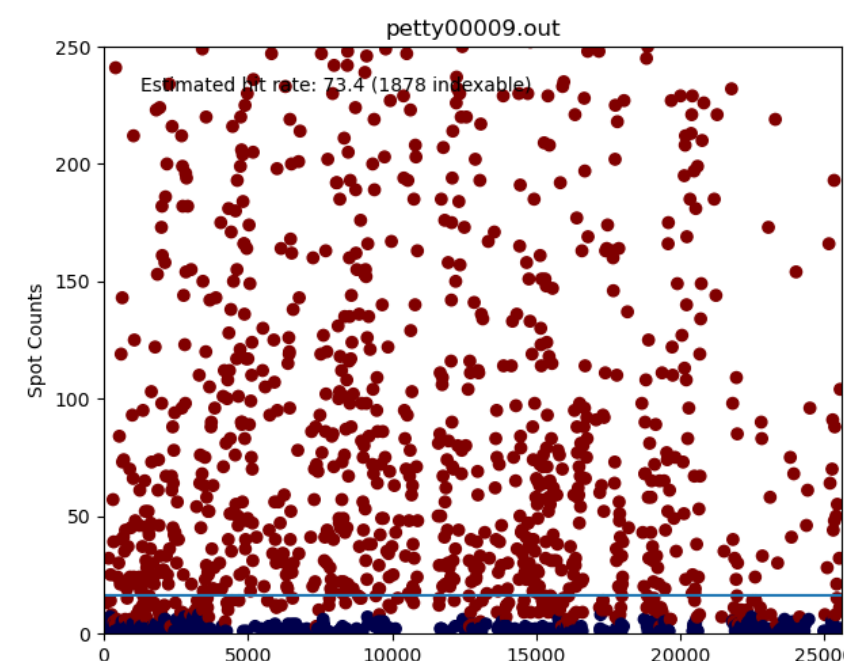
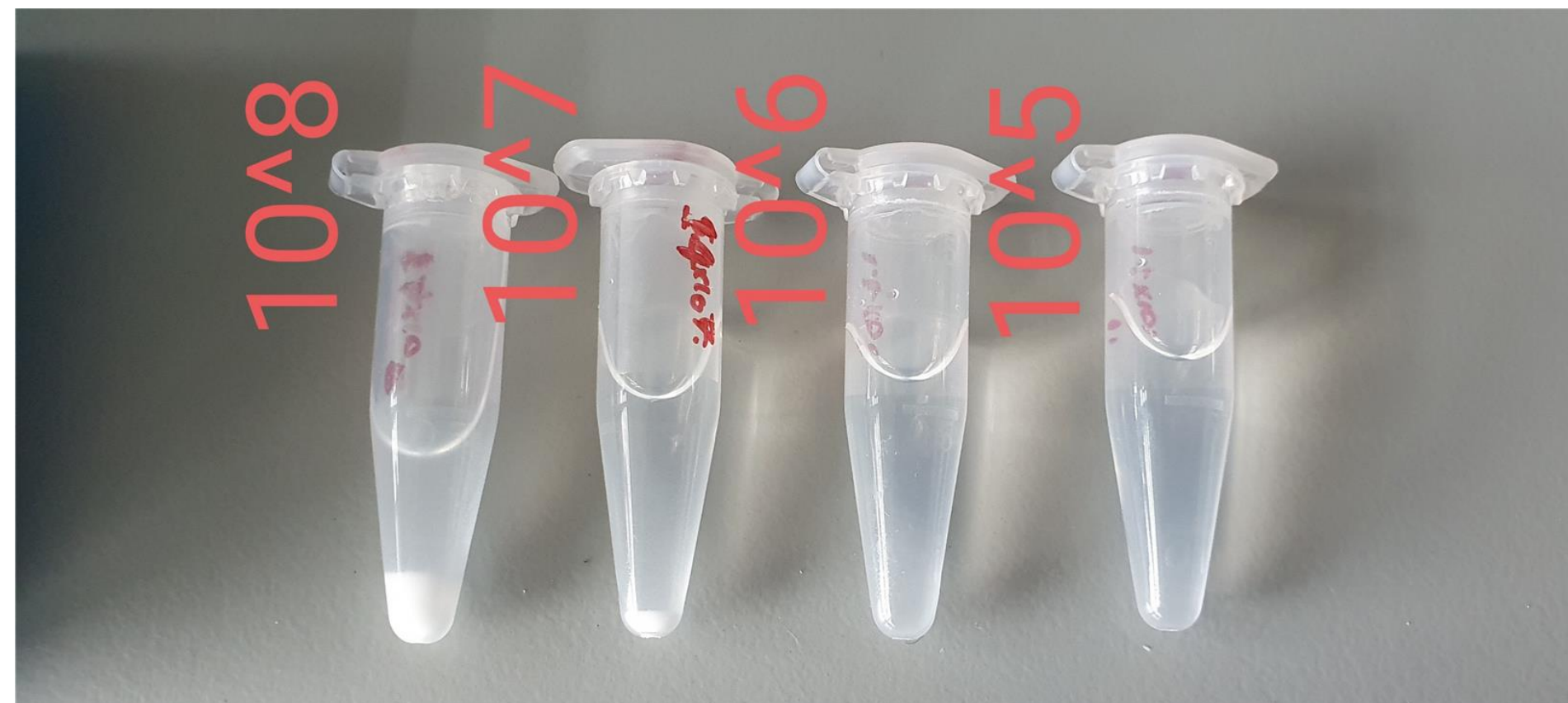


# Automated data processing

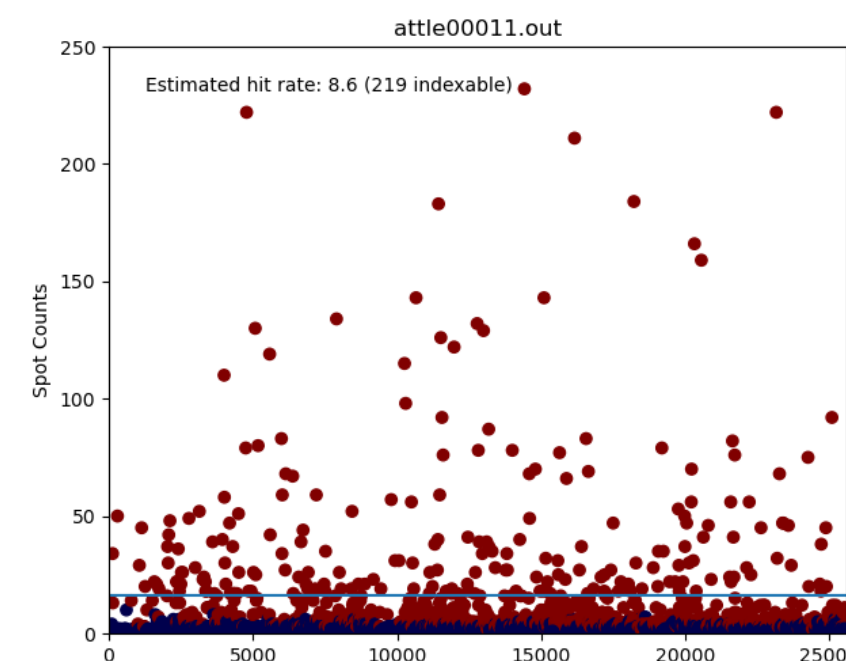
Fast feedback essential for understanding success of sample preparation and data quality.

Hit rate -> indexing success rate -> data quality statistics.

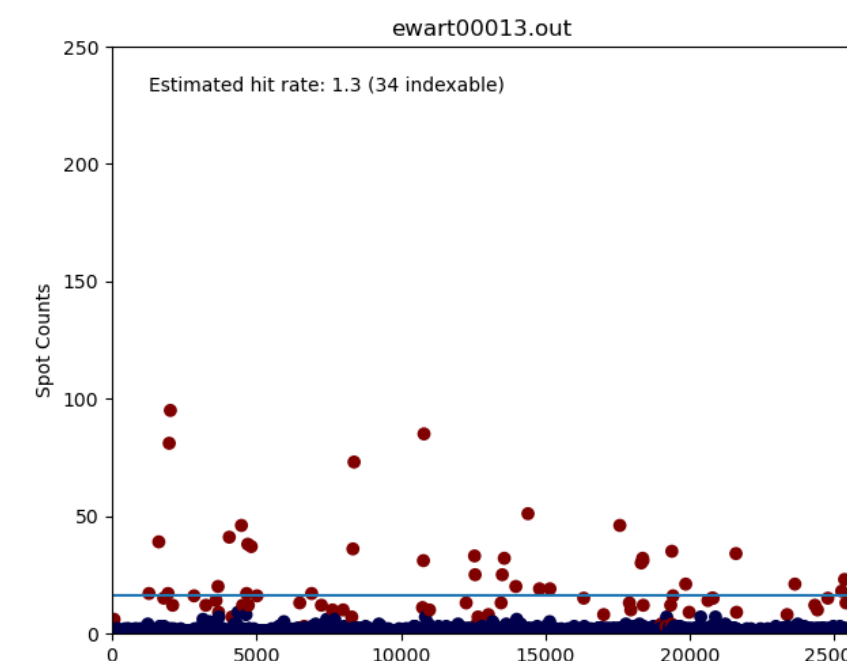
Sam Horrell



75%



9%



1%

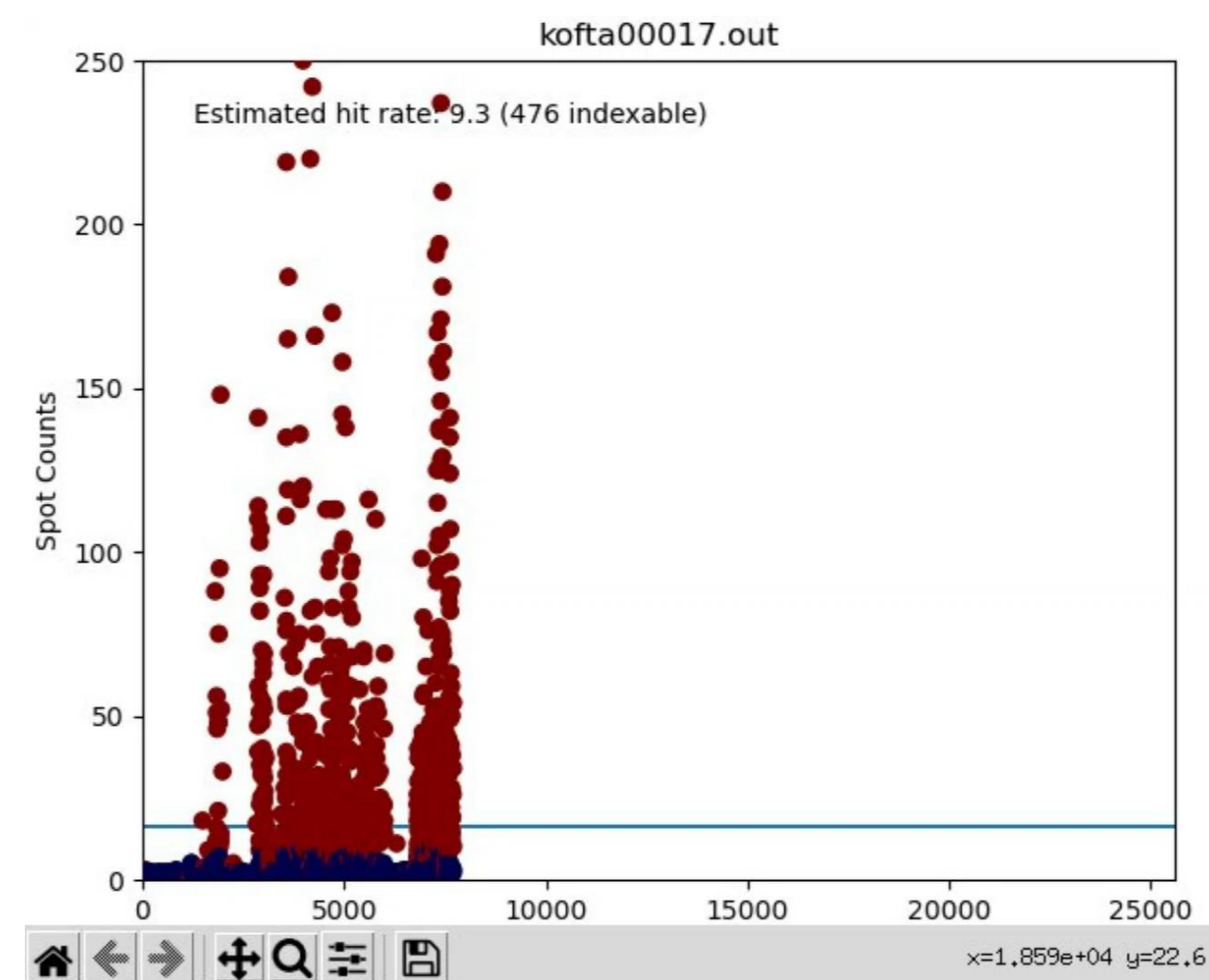
Didn't  
bother to  
collect it

# Automated data processing: fast feedback

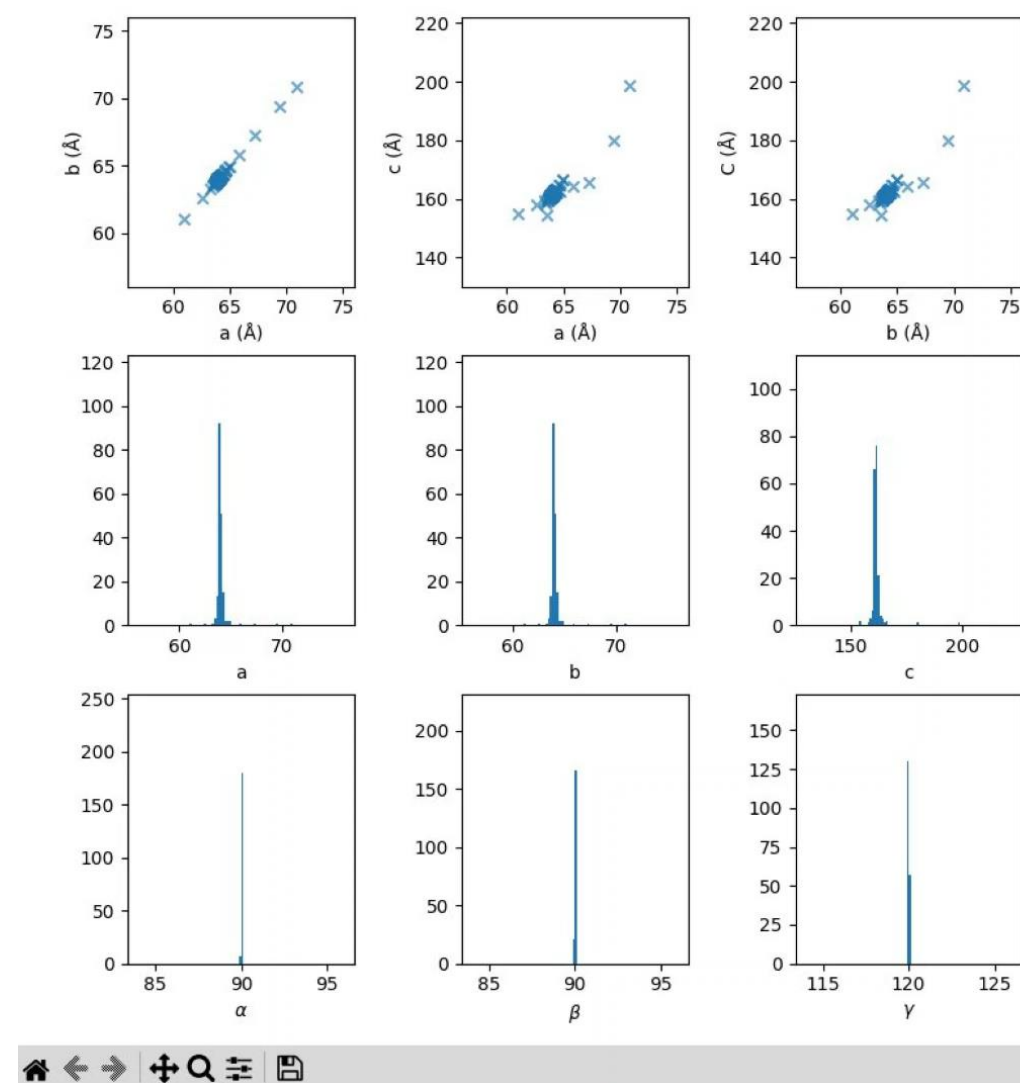
Live feedback from DIALS processing – `dials.find_spots`, `dials.index`



[dials.github.io](https://dials.github.io)



Per-image spot-finding, hit rate estimate



Per-hit indexing – cell parameter distributions



# Automated data processing: infrastructure

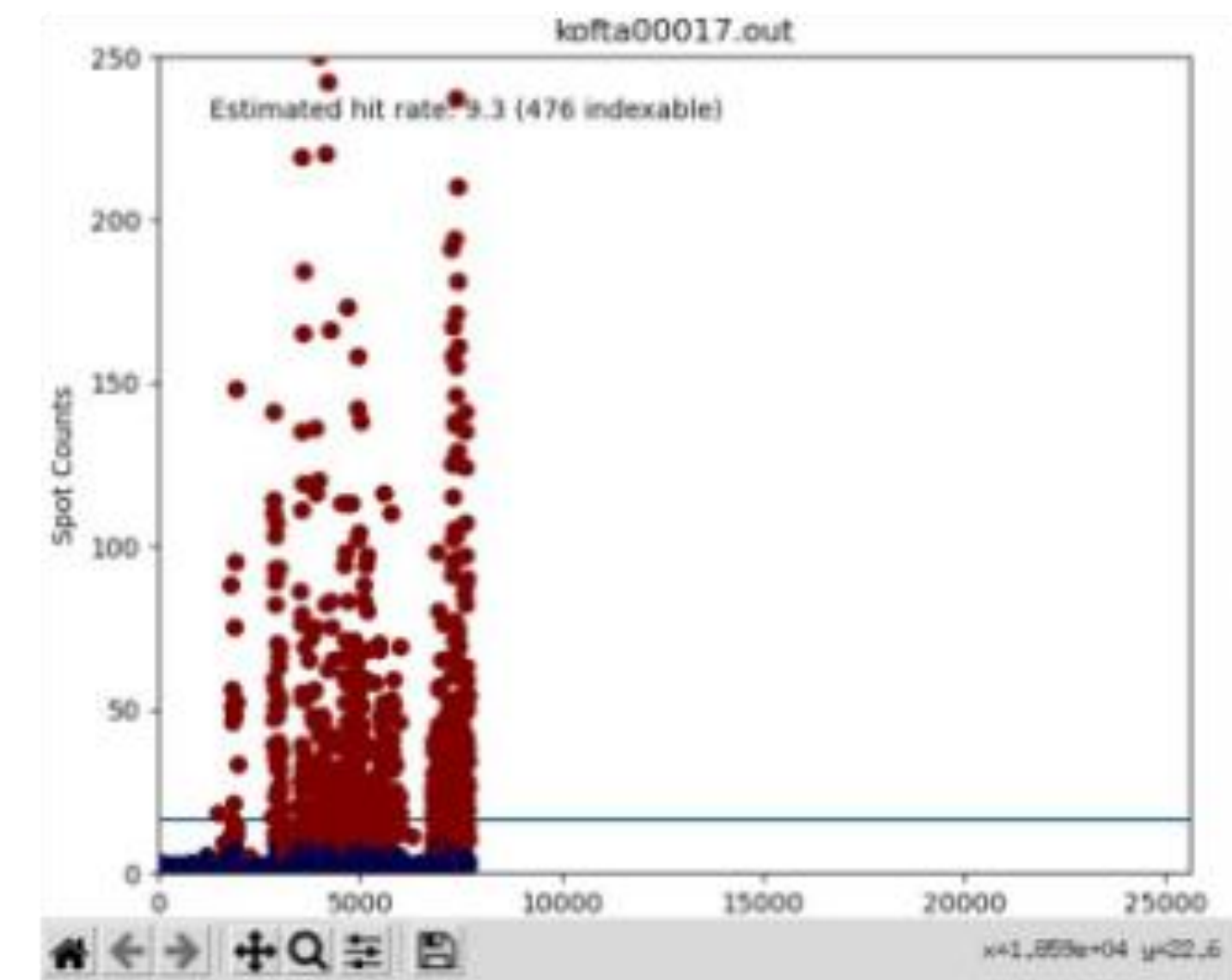
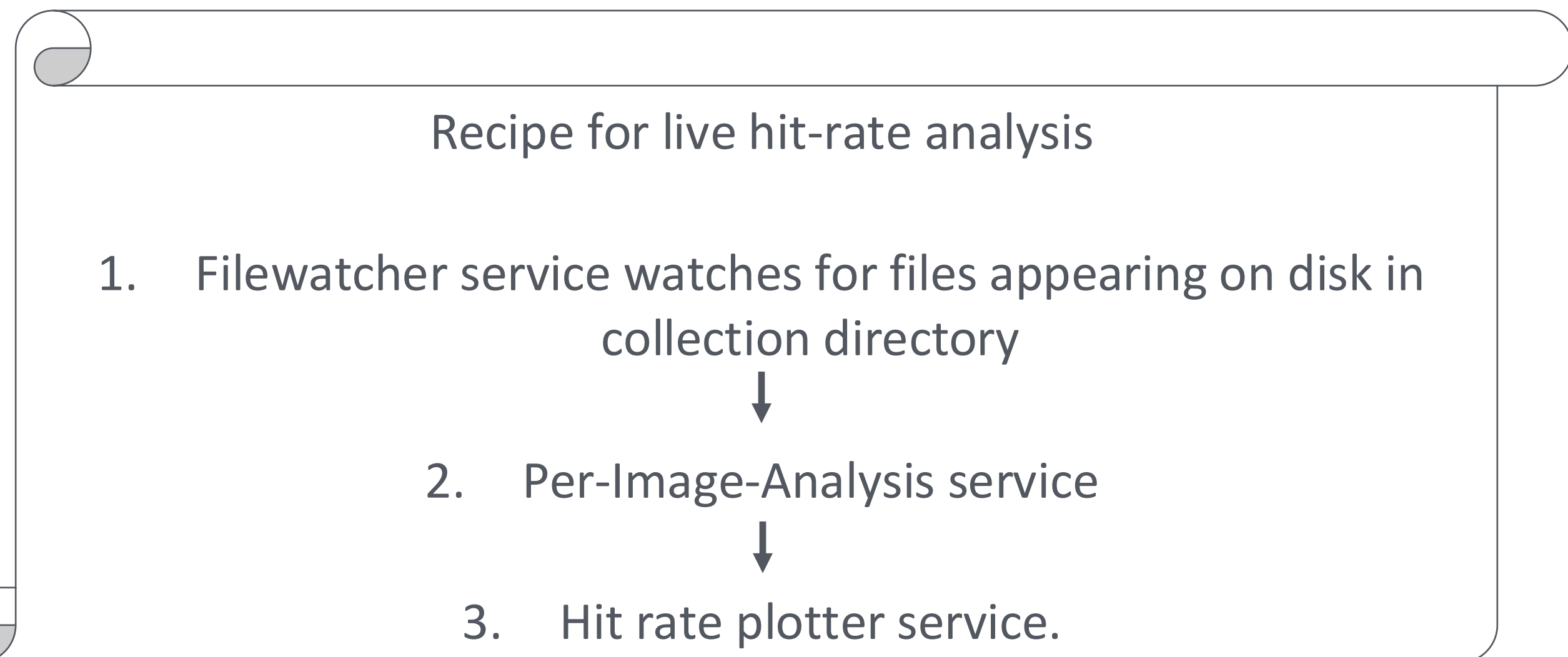
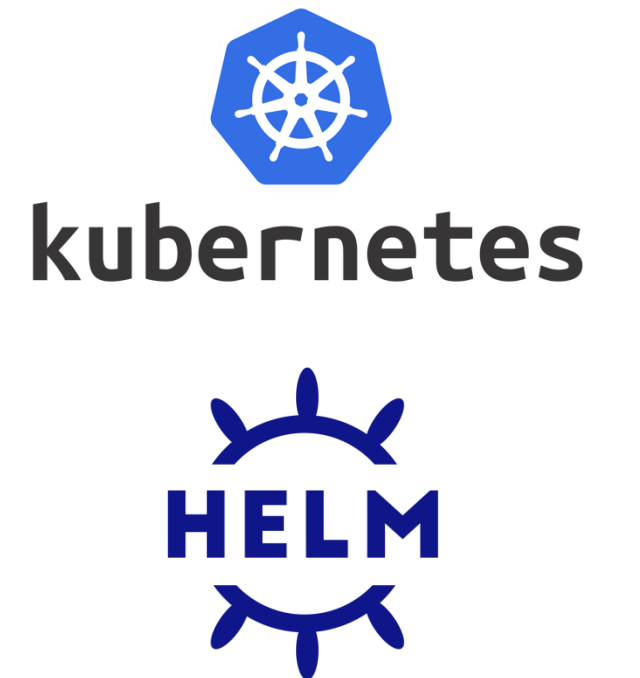
## Analysis infrastructure at Diamond

Data analysis managed by zocalo software. [github.com/DiamondLightSource/python-zocalo](https://github.com/DiamondLightSource/python-zocalo)

Zocalo defines a set of services (run on Kubernetes cluster at DLS) – these run continuously and wait for work – via messages managed by RabbitMQ server.

Zocalo also defines wrappers – run processing commands on demand (e.g. xia2 processing job).

Finally have a set of recipes which define workflows based on these services/wrappers as a flowchart/graph.



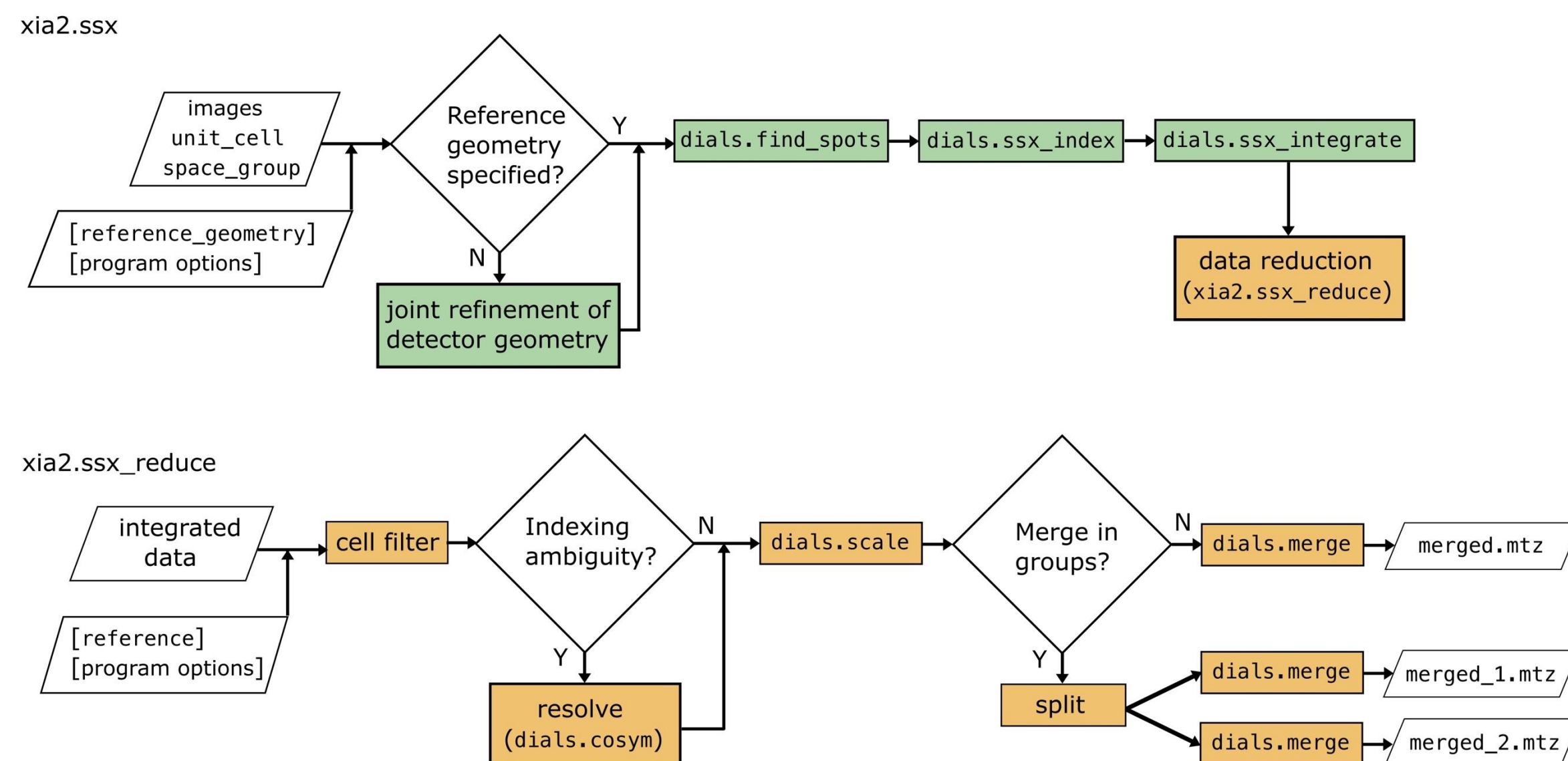
# Automated data processing: full processing

Full post-processing with xia2.ssx triggered automatically after data collection.

Relies on a small configuration (.yaml) file detailing expected unit cell and space group.

xia2.ssx – processes images to merged MTZ running DIALS programs.

xia2.ssx\_reduce – reduces integrated together from multiple collections.



## DIALS merge report

### Merging statistics

Overall			
	Overall	Low resolution	High resolution
Resolution (Å)	56.02 - 1.32	56.07 - 3.57	1.34 - 1.32
Observations	1710375	245915	527
Unique reflections	27793	1620	422
Multiplicity	61.5	151.8	1.2
Completeness	94.62%	100.00%	29.87%
Mean I/σ(I)	12.6	74.7	-0.1
R <sub>merge</sub>	1.038	0.573	-2.070
R <sub>meas</sub>	1.045	0.575	-2.879
R <sub>pim</sub>	0.112	0.048	-1.995
R <sub>split</sub>	0.195	0.042	-2.942
CC <sub>1/2</sub>	0.959	0.996	-0.137

Resolution shells

Xtrriage

✖ 1 serious warning

✖ The merging statistics indicate that the data may be assigned to the wrong space group.

✔ 8 checks passed



# Automated data processing: full processing

Results displayed in SynchWeb on a per-chip basis. Multiple chips of the same condition merged together automatically. Access to log files, html report and merged MTZ file.

Flux: NaN     $\Omega$  Start: 0.0°  
 $\Omega$  Osc: 0.00°     $\Omega$  Overlap: 0.0°  
 No. Images: 25600    First Image: 0  
 Resolution: 1.73Å    Wavelength: 0.9999Å  
 Exposure: 0.0100s    Transmission: 49.30%  
 Beamsize: 7x7 $\mu$ m    Type: Serial Fixed  
 Comment:

Auto Processing xia2.ssx: ✓ xia2.ssx\_reduce: ✓

Type	Resolution	Resolution I/sig(I)=2	Spacegroup	Mn<I/sig(I)>	Rmeas Inner	Rmeas Outer	Completeness	Cell	Status
xia2.ssx	56.00 - 1.32	0.00	P 43 21 2	8.8	0.566	-1.065	88.6	79.20 79.20 38.18 90.00 90.00 90.00	processing successful
3x multi-xia2.ssx_reduce	56.02 - 1.32	0.00	P 43 21 2	12.6	0.575	-2.516	94.6	79.23 79.23 38.16 90.00 90.00 90.00	processing successful

5 check(s) passed  
 3 alert(s)

Beam Centre	X	Y
Start	222.29	225.49
Refined	--	--
$\Delta$	--	--

Space Group	A	B	C	$\alpha$	$\beta$	$\gamma$
P 43 21 2	79.20	79.20	38.18	90.00	90.00	90.00

Shell	Observations	Unique	Resolution	Rmeas	I/sig(I)	CC Half	Completeness	Multiplicity	Anom Completeness	Anom Multiplicity	CC Anom
outerShell	168	160	1.32 - 1.34	-1.065	-0.1	0.5	11.3	1.1	0.0	1.0	0.0
innerShell	66946	1619	3.57 - 56.06	0.566	46.1	1.0	100.0	41.4	1.0	24.4	0.0
overall	475439	25988	1.32 - 56.00	0.990	8.8	0.9	88.6	18.3	0.8	10.4	0.0

Timescale ~15-30 mins



# Automated serial data processing: future directions

- Acceleration of live processing through migration to GPU devices.
- Extension of live processing to cover the full processing, enabling complete metrics on fast timescale.
- Completion of metadata models to cover all serial experiment types (dose/time series).
- Web-based GUI for live display of results
- Migration from reliance on file systems to stream-based workflows.



# Acknowledgements

Data acquisition team, in particular Noemi Frisina, David Perl, Dominic Oram

I24 Beamline team – Robin Owen, Danny Axford, Sofia Jaho, Do-Heon Gu + many others over last 10+ years

Data Analysis/DIALS team, in particular Nick Devenish, James Parkhurst, Graeme Winter

[github.com/DiamondLightSource/blueapi](https://github.com/DiamondLightSource/blueapi)

[nsls-ii.github.io/bluesky/](https://nsls-ii.github.io/bluesky/)

[dials.github.io](https://dials.github.io). [xia2.github.io](https://xia2.github.io)

[github.com/DiamondLightSource/python-zocalo](https://github.com/DiamondLightSource/python-zocalo)

Slide material: Danny Axford, Robin Owen, Sam Horrell



Grant funding from Wellcome trust, CCP4/UKRI, EU grant funding.

