

NSF BioXFEL STC's SFX data analysis workshop at ACA 2015, Philadelphia  
Presented by Nadia Zatsepin (ASU, BioXFEL)  
[https://www.bioxfel.org/resources/LCLSdata\\_overview](https://www.bioxfel.org/resources/LCLSdata_overview)

# SFX data reduction and preprocessing

## research papers

Journal of  
**Applied  
Crystallography**  
ISSN 1600-5767

Received 14 November 2013  
Accepted 4 April 2014

### ***Cheetah*: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data**

Anton Barty,<sup>a\*</sup> Richard A. Kirian,<sup>a</sup> Filipe R. N. C. Maia,<sup>b,c</sup> Max Hantke,<sup>b</sup> Chun Hong Yoon,<sup>a,d</sup> Thomas A. White<sup>a</sup> and Henry Chapman<sup>a,e</sup>

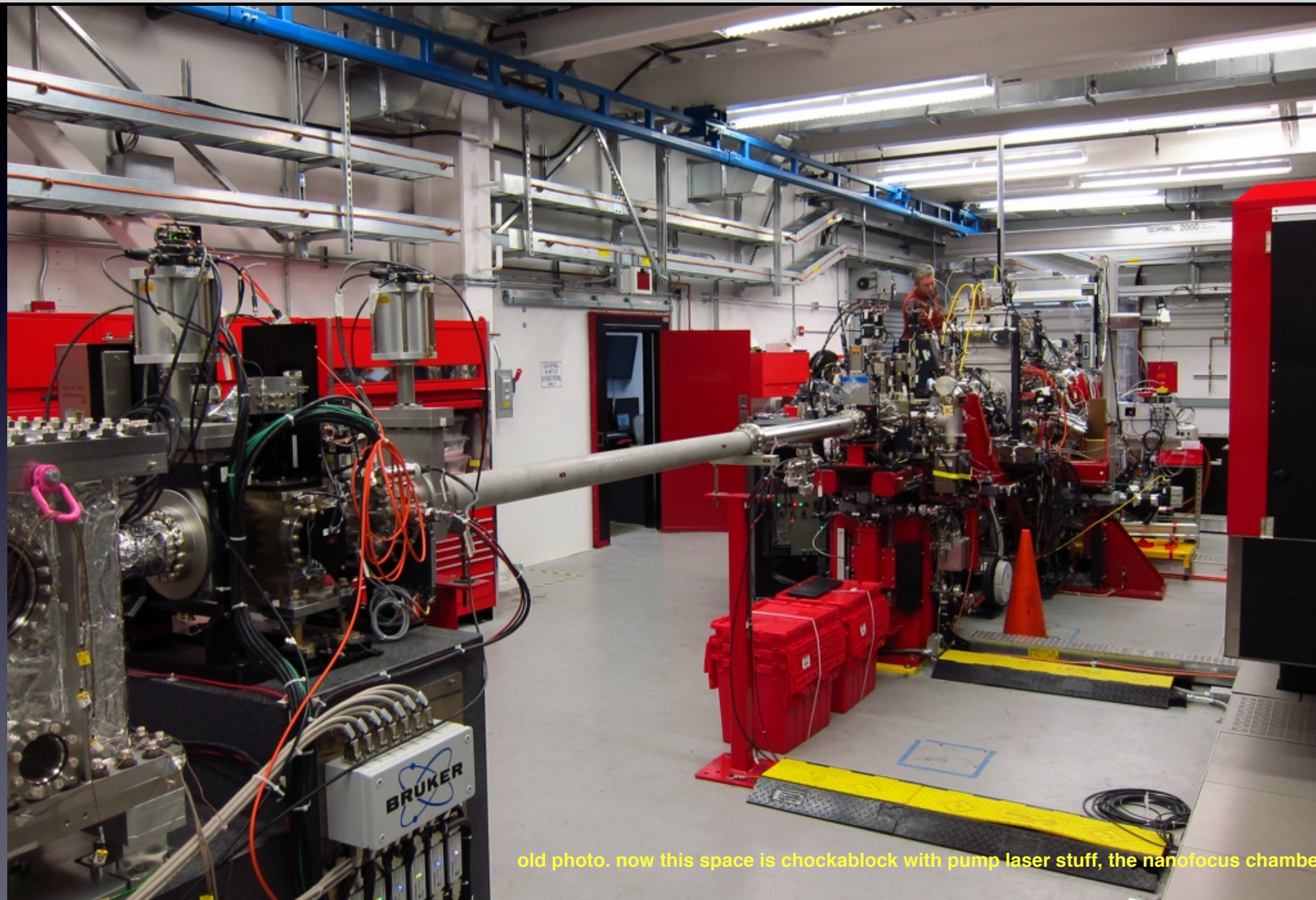
**1118** doi:10.1107/S1600576714007626

*J. Appl. Cryst.* (2014). **47**, 1118–1131

Anton Barty  
Center for Free Electron Laser Science at DESY

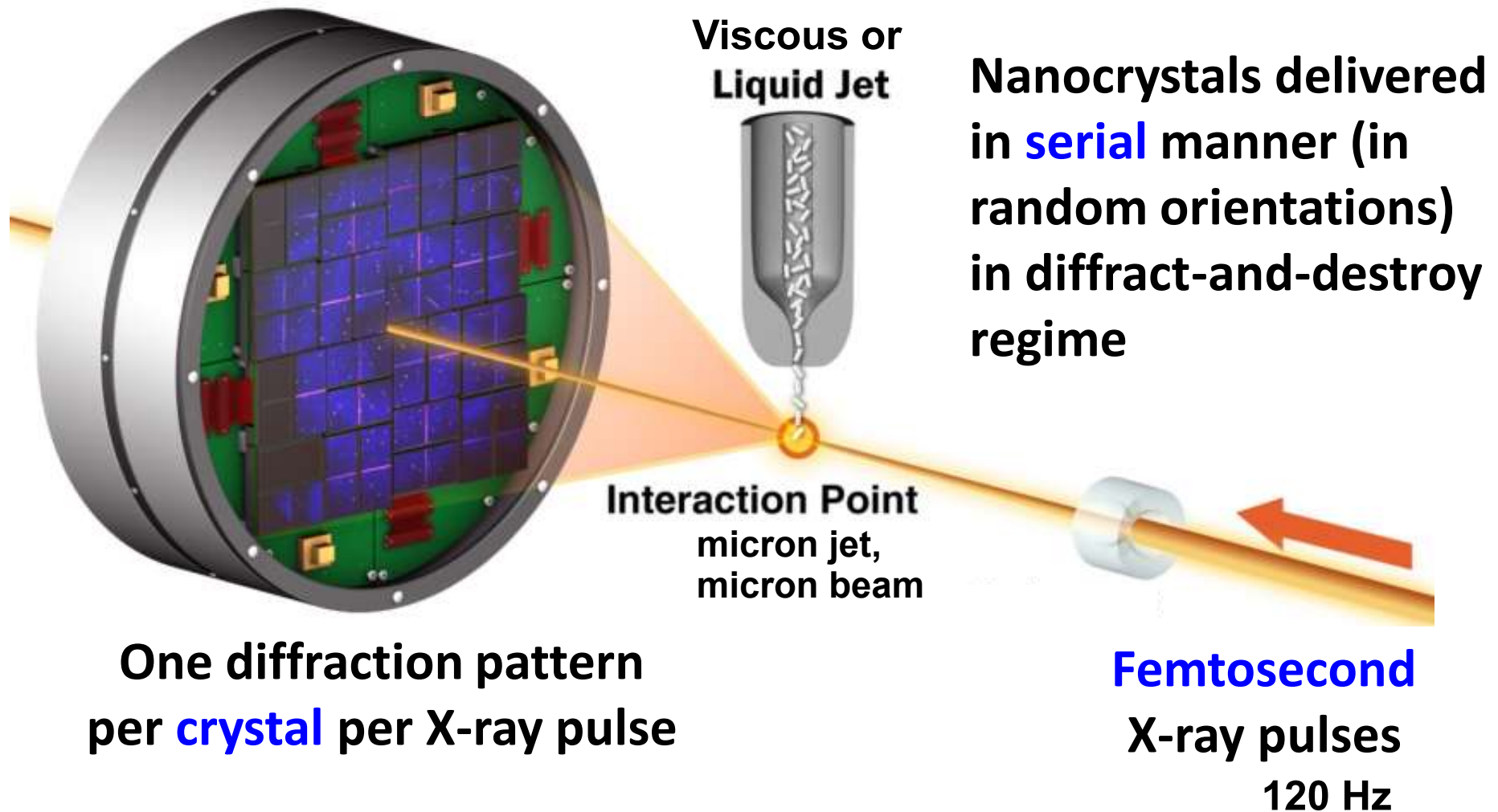


The CXI instrument at LCLS delivers more than  $10^{12}$  photons into a  $1\ \mu\text{m}$  or  $0.1\ \mu\text{m}$  focal spot at 4-9 keV X-ray energy



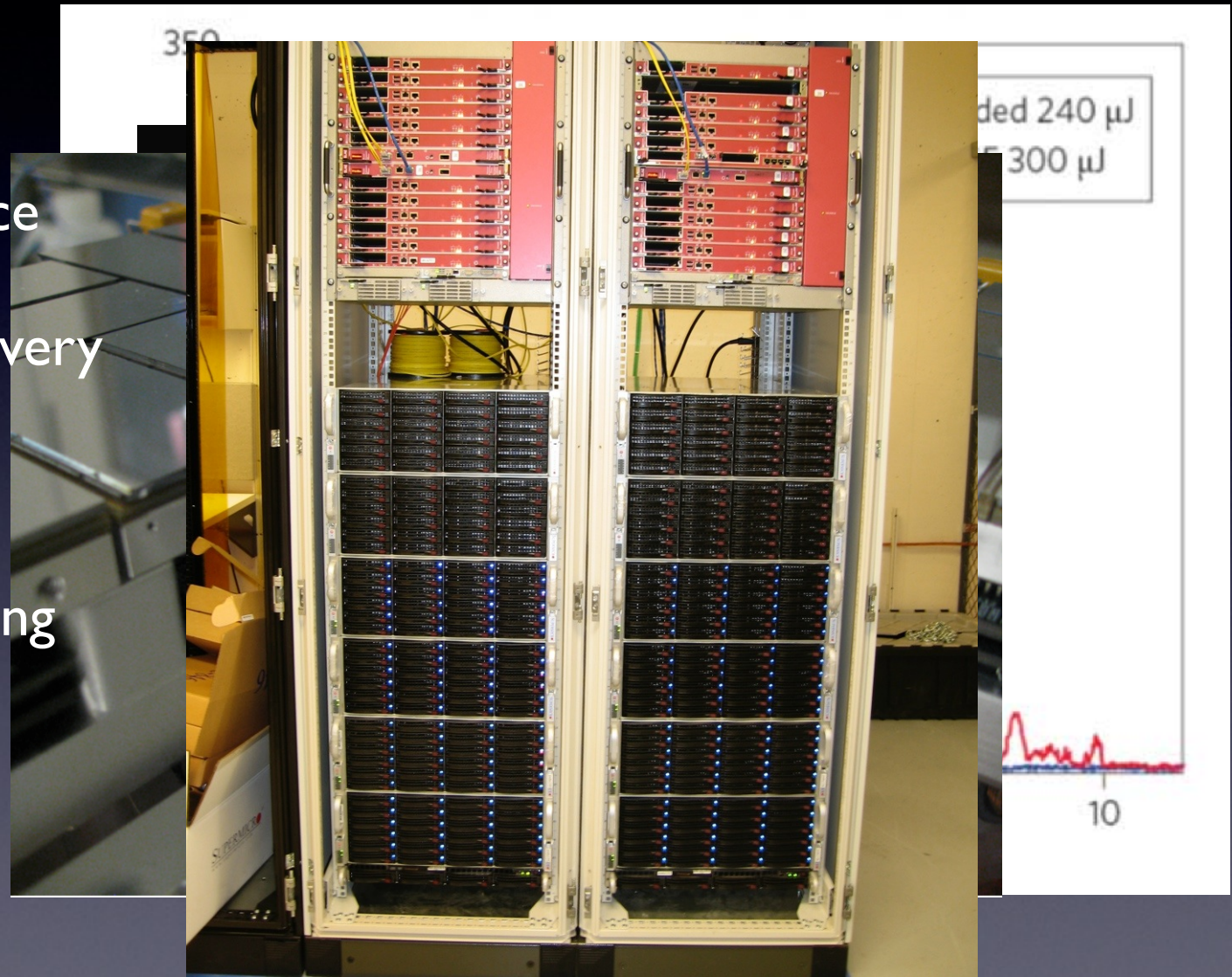
old photo. now this space is chockablock with pump laser stuff, the nanofocus chamber etc.

# Serial femtosecond crystallography (SFX)

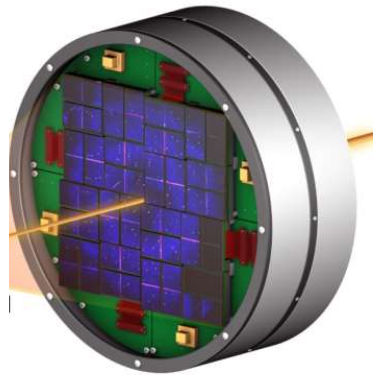


# SFX experiments at LCLS pose some unique challenges

1. X-ray source
2. Sample delivery
3. Detector
4. Data handling



# SFX data analysis pipeline



CSPAD

DAQ: XTC files,  
containing X-ray  
pulse parameters,  
diagnostics, motor  
positions, ..

## Cheetah

1. Hit finding (data reduction)
2. Background estimation, removal
3. Clean diff. pattern & meta data → HDF5
4. Statistics & prelim. analysis

## CCP4, Phenix et al.

Phasing, building,  
refinement,  
validation

## CrystFEL

1. Indexing
2. Integration
3. Merging
4. Post refinement
5. Indexing ambiguity removal

RCSB **PDB**  
PROTEIN DATA BANK

# Why new software?

- New type of data
- Large amount of data
- New, complicated detectors

LCLS fires at **120 Hz**

CSPAD detector at CXI, LCLS:  $2.3 \times 10^6$  pixels, 16 bit

**4.6 MB / frame = 2 TB / hour**

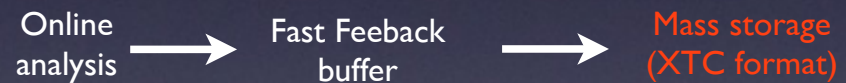
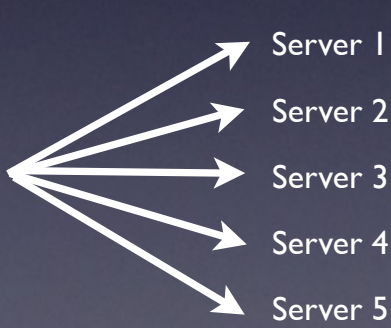
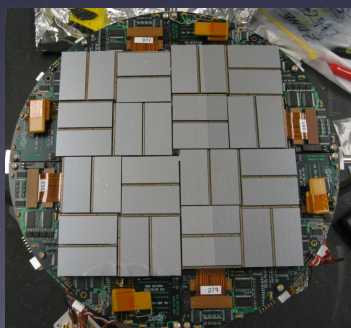
→ Up to ~ **120 TB from a single experiment** (5 shifts) from one detector

European XFEL will run at 27,000 Hz...

# You will soon drown in data

120 frames per second  
432,000 frames per hour

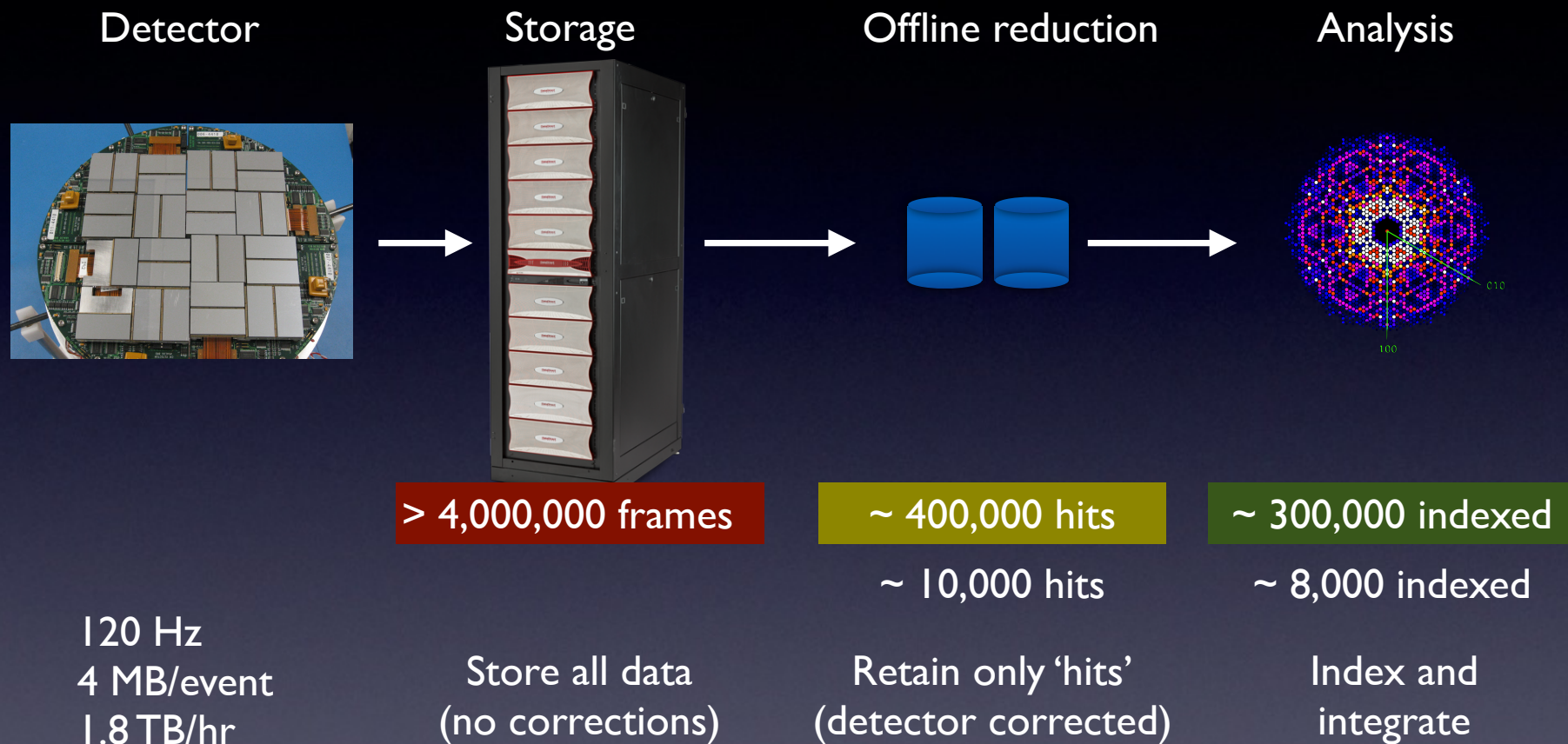
Where is my data ?  
WTF is an XTC ?  
How do I take 100 TB home ?  
How do I read an XTC anyway ?  
What now ?



```
[psexport01:barty]cxi/cxi52212/xtc> ls -alh *0155*  
-r--r-----+ | psdatmgr ps-data 92G Oct 23 11:06 e182-r0155-s02-c00.xtc  
-r--r-----+ | psdatmgr ps-data 20G Oct 23 11:22 e182-r0155-s02-c01.xtc  
...  
-r--r-----+ | psdatmgr ps-data 92G Oct 23 11:06 e182-r0155-s03-c00.xtc  
-r--r-----+ | psdatmgr ps-data 20G Oct 23 11:21 e182-r0155-s03-c01.xtc  
...  
-r--r-----+ | psdatmgr ps-data 93G Oct 23 11:06 e182-r0155-s04-c00.xtc  
-r--r-----+ | psdatmgr ps-data 20G Oct 23 11:22 e182-r0155-s04-c01.xtc  
-r--r-----+ | psdatmgr ps-data 93G Oct 23 11:06 e182-r0155-s05-c00.xtc  
-r--r-----+ | psdatmgr ps-data 20G Oct 23 11:21 e182-r0155-s05-c01.xtc  
-r--r-----+ | psdatmgr ps-data 92G Oct 23 11:06 e182-r0155-s06-c00.xtc  
-r--r-----+ | psdatmgr ps-data 20G Oct 23 11:21 e182-r0155-s06-c01.xtc
```



# Data processing is an exercise in massive data reduction



Automated high volume image processing is essential  
(eg: background correction, weeding useful data from useless data)

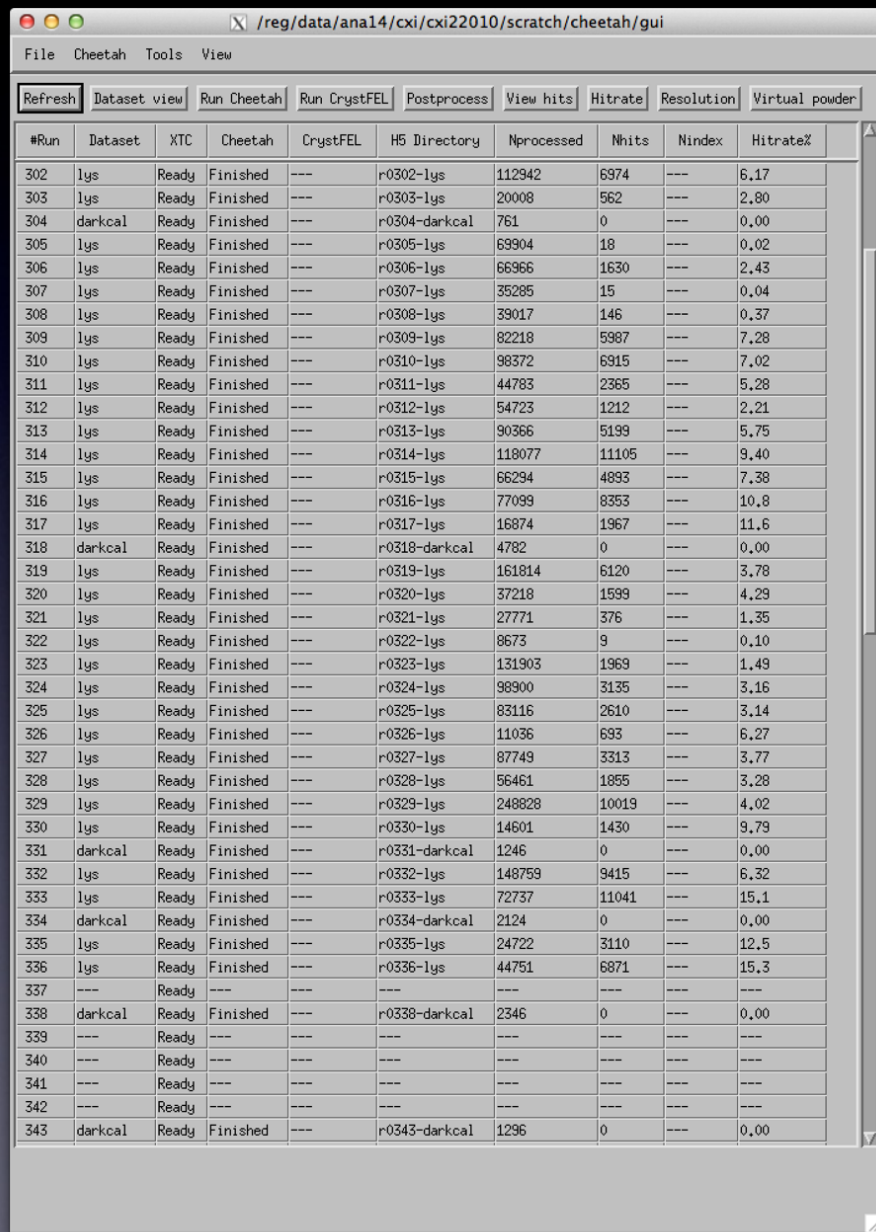
# Step I: Organise your data

A shared spreadsheet is ideal for remembering what is in each run

The screenshot shows a Google Spreadsheet interface with the following data:

Run #	Date	Time	Samp Num	Sample	Sample owner	Comments	Filter/Transmission	Post Sample Transmission	Analysis results
85	27 Jun 2013	02:58:23	4	KI 0.05 M		Calibration for filter checks, catcher z= 3.5, x= 1.1775, y= -0.742 T=100% 25uL/min 200psi			
86	27 Jun 2013	03:00:57	4	KI 0.05 M		Calibration without filters, same catcher pos.			
87	27 Jun 2013	06:58:12		Dark Run			Filter = post attenuation		
88	27 Jun 2013	07:00:24	5	Cat B (eone)	Francesco	1.5 x Concentrated sample 3, no hits	No Filter T=7%		water by mistake
89	27 Jun 2013	07:02:17	5	Cat B (eone)		jet not stable, no hits 10uL/min 1050psi	No Filter T=7%		water by mistake
90	27 Jun 2013	07:08:49	5	Cat B (eone)		chasing crystals	No Filter T=14%		water by mistake
91	27 Jun 2013	07:15:56	5	Cat B (eone)		z=3, out of focus to try to increase hit chance	No Filter T=52%		water by mistake
92	27 Jun 2013	07:32:43	5	Cat B (eone)		inline filter 2um to 10um PEEK, 8uL/min 1000psi	No Filter T=14%		water by mistake
93	27 Jun 2013	07:54:21	6	Lysozyme 4A	ilme	removed Ti/In filters, no hits apparently just running water, checking connections	No Filter T=7%		water by mistake
94	27 Jun 2013	08:13:47		Dark Run					
95	27 Jun 2013	08:21:27	5	Cat B (conc)	Francesco	10uL/min 900psi 10 um PEEK inline filter	No Filter T=14%		
96	27 Jun 2013	08:37:37	5	Cat B (conc)		10uL/min 1060psi 10um inline filter No Ti/In filter	No Filter T=14%		
97	27 Jun 2013	08:47:16	5	Cat B (conc)		15uL/min 1500psi 10um inline filter No Ti/In filter	No Filter T=14%		
98	27 Jun 2013	09:00:34		Dark Run					
End of 2nd shift									
99	27 Jun 2013	23:33:30		Dark Run					
100	27 Jun 2013	23:41:28	7	Cat B (pooled)	Francesco	10uL/min 2um PEEK	No Filter T=3.7%		
101	27 Jun 2013	23:52:58	7	Cat B (pooled)	Francesco	10 -> 15uL/min 2um PEEK	No Filter T=3.7%		
102	28 Jun 2013	00:01:53	7	Cat B (pooled)	Francesco	15uL/min 2um PEEK	No Filter T=3.7%		
103	28 Jun 2013	00:10:51	7	Cat B (pooled)	Francesco	15uL/min 2um PEEK	No Filter T=3.7%		
104	28 Jun 2013	00:19:44	7	Cat B (pooled)	Francesco	15uL/min 2um PEEK 300psi	No Filter T=14%		really short run
105	28 Jun 2013	00:19:48	7	Cat B (pooled)	Francesco	15uL/min 2um PEEK	No Filter T=14%		
106	28 Jun 2013	00:29:54	7	Cat B (pooled)	Francesco	15uL/min 2um PEEK	No Filter T=14%		
107	28 Jun 2013	00:31:12	7	Cat B (pooled)	Francesco	15uL/min 2um PEEK	No Filter T=27%		
108	28 Jun 2013	00:42:38	7	Cat B (pooled)	Francesco	15uL/min 2um PEEK	No Filter T=27%		

# Cheetah is used for data reduction, rapid data evaluation, and translation

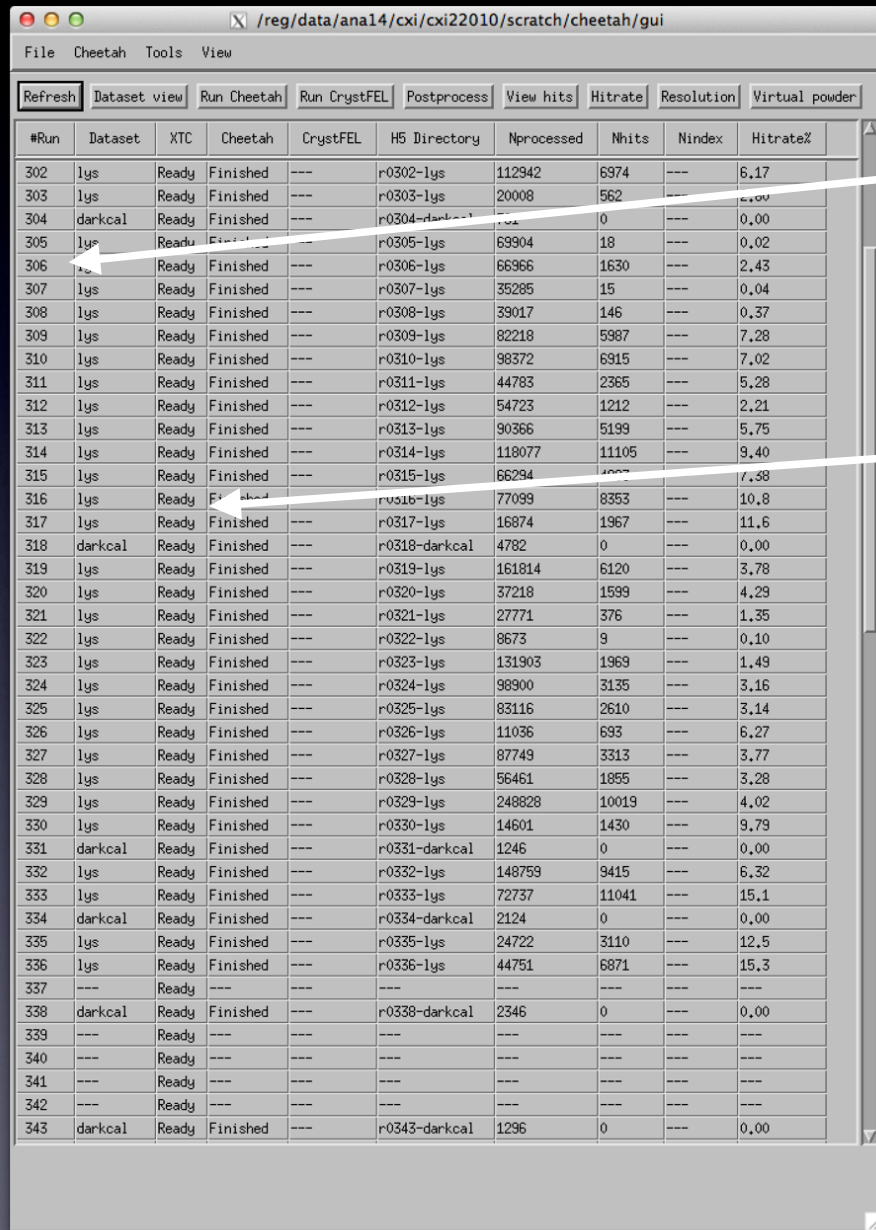


The screenshot shows the Cheetah GUI interface. At the top, there is a menu bar with 'File', 'Cheetah', 'Tools', and 'View'. Below the menu bar is a toolbar with buttons for 'Refresh', 'Dataset view', 'Run Cheetah', 'Run CrystFEL', 'Postprocess', 'View hits', 'Hitrate', 'Resolution', and 'Virtual powder'. The main area is a table with the following columns: #Run, Dataset, XTC, Cheetah, CrystFEL, H5 Directory, Nprocessed, Nhits, Nindex, and Hitrate%. The table contains 34 rows of data, with runs 302 through 343. The status of each run is indicated in the 'Cheetah' column, with most runs being 'Finished' and some being 'Ready'.

#Run	Dataset	XTC	Cheetah	CrystFEL	H5 Directory	Nprocessed	Nhits	Nindex	Hitrate%
302	lys	Ready	Finished	---	r0302-lys	112942	6974	---	6.17
303	lys	Ready	Finished	---	r0303-lys	20008	562	---	2.80
304	darkcal	Ready	Finished	---	r0304-darkcal	761	0	---	0.00
305	lys	Ready	Finished	---	r0305-lys	69904	18	---	0.02
306	lys	Ready	Finished	---	r0306-lys	66966	1630	---	2.43
307	lys	Ready	Finished	---	r0307-lys	35285	15	---	0.04
308	lys	Ready	Finished	---	r0308-lys	39017	146	---	0.37
309	lys	Ready	Finished	---	r0309-lys	82218	5987	---	7.28
310	lys	Ready	Finished	---	r0310-lys	98372	6915	---	7.02
311	lys	Ready	Finished	---	r0311-lys	44783	2365	---	5.28
312	lys	Ready	Finished	---	r0312-lys	54723	1212	---	2.21
313	lys	Ready	Finished	---	r0313-lys	90366	5199	---	5.75
314	lys	Ready	Finished	---	r0314-lys	118077	11105	---	9.40
315	lys	Ready	Finished	---	r0315-lys	66294	4893	---	7.38
316	lys	Ready	Finished	---	r0316-lys	77099	8353	---	10.8
317	lys	Ready	Finished	---	r0317-lys	16874	1967	---	11.6
318	darkcal	Ready	Finished	---	r0318-darkcal	4782	0	---	0.00
319	lys	Ready	Finished	---	r0319-lys	161814	6120	---	3.78
320	lys	Ready	Finished	---	r0320-lys	37218	1599	---	4.29
321	lys	Ready	Finished	---	r0321-lys	27771	376	---	1.35
322	lys	Ready	Finished	---	r0322-lys	8673	9	---	0.10
323	lys	Ready	Finished	---	r0323-lys	131903	1969	---	1.49
324	lys	Ready	Finished	---	r0324-lys	98900	3135	---	3.16
325	lys	Ready	Finished	---	r0325-lys	83116	2610	---	3.14
326	lys	Ready	Finished	---	r0326-lys	11036	693	---	6.27
327	lys	Ready	Finished	---	r0327-lys	87749	3313	---	3.77
328	lys	Ready	Finished	---	r0328-lys	56461	1855	---	3.28
329	lys	Ready	Finished	---	r0329-lys	248828	10019	---	4.02
330	lys	Ready	Finished	---	r0330-lys	14601	1430	---	9.79
331	darkcal	Ready	Finished	---	r0331-darkcal	1246	0	---	0.00
332	lys	Ready	Finished	---	r0332-lys	148759	9415	---	6.32
333	lys	Ready	Finished	---	r0333-lys	72737	11041	---	15.1
334	darkcal	Ready	Finished	---	r0334-darkcal	2124	0	---	0.00
335	lys	Ready	Finished	---	r0335-lys	24722	3110	---	12.5
336	lys	Ready	Finished	---	r0336-lys	44751	6871	---	15.3
337	---	Ready	---	---	---	---	---	---	---
338	darkcal	Ready	Finished	---	r0338-darkcal	2346	0	---	0.00
339	---	Ready	---	---	---	---	---	---	---
340	---	Ready	---	---	---	---	---	---	---
341	---	Ready	---	---	---	---	---	---	---
342	---	Ready	---	---	---	---	---	---	---
343	darkcal	Ready	Finished	---	r0343-darkcal	1296	0	---	0.00

1. 'Control panel' interface to data and LCLS analysis cluster
2. Rapid feedback  
Hit rate, resolution, diffraction quality  
Quickly viewing images
3. Data reduction  
Keeps only useful events crystals  
(ie: frames with crystal diffraction)
4. Data translation  
XTC data is converted to a facility independent format (HDF5)
5. Data organisation  
Summarises what is in each run;  
easy to group data by sample;  
summarises statistics

# Cheetah functionality: XTC monitor



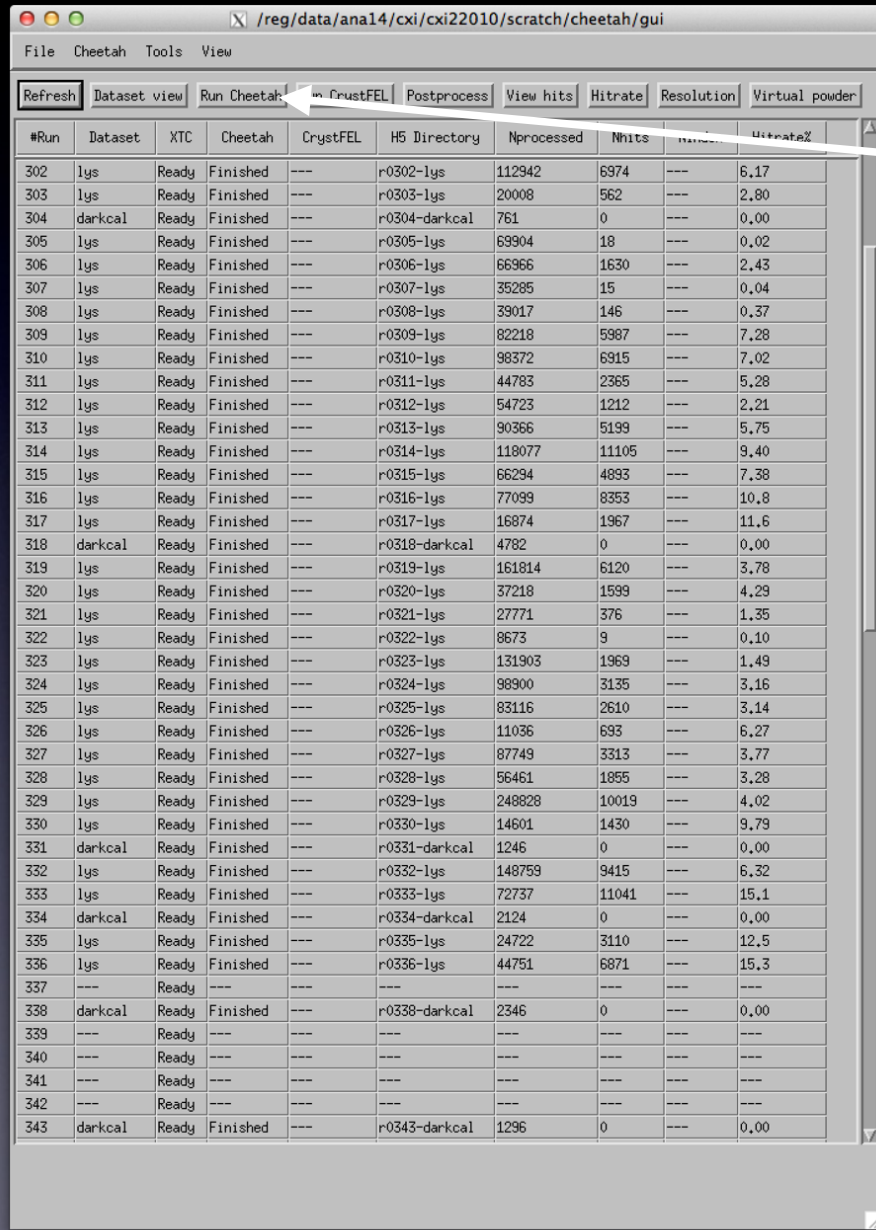
The screenshot shows the Cheetah XTC monitor GUI. The window title is "/reg/data/ana14/cxi/cxi22010/scratch/cheetah/gui". The menu bar includes "File", "Cheetah", "Tools", and "View". Below the menu bar is a toolbar with buttons for "Refresh", "Dataset view", "Run Cheetah", "Run CrystFEL", "Postprocess", "View hits", "Hitrate", "Resolution", and "Virtual powder". The main area is a table with the following columns: #Run, Dataset, XTC, Cheetah, CrystFEL, H5 Directory, Nprocessed, Nhits, Nindex, and Hitrate%. The table contains 44 rows of data, with the first 43 rows having values and the last row (343) having dashes for the XTC, Cheetah, CrystFEL, Nprocessed, Nhits, and Nindex columns.

#Run	Dataset	XTC	Cheetah	CrystFEL	H5 Directory	Nprocessed	Nhits	Nindex	Hitrate%
302	lys	Ready	Finished	---	r0302-lys	112942	6974	---	6,17
303	lys	Ready	Finished	---	r0303-lys	20008	562	---	2,80
304	darkcal	Ready	Finished	---	r0304-darkcal	0	0	---	0,00
305	lys	Ready	Finished	---	r0305-lys	69904	18	---	0,02
306	lys	Ready	Finished	---	r0306-lys	66966	1630	---	2,43
307	lys	Ready	Finished	---	r0307-lys	35285	15	---	0,04
308	lys	Ready	Finished	---	r0308-lys	39017	146	---	0,37
309	lys	Ready	Finished	---	r0309-lys	82218	5987	---	7,28
310	lys	Ready	Finished	---	r0310-lys	98372	6915	---	7,02
311	lys	Ready	Finished	---	r0311-lys	44783	2365	---	5,28
312	lys	Ready	Finished	---	r0312-lys	54723	1212	---	2,21
313	lys	Ready	Finished	---	r0313-lys	90366	5199	---	5,75
314	lys	Ready	Finished	---	r0314-lys	118077	11105	---	9,40
315	lys	Ready	Finished	---	r0315-lys	66294	4000	---	7,38
316	lys	Ready	Finished	---	r0316-lys	77099	8353	---	10,8
317	lys	Ready	Finished	---	r0317-lys	16874	1967	---	11,6
318	darkcal	Ready	Finished	---	r0318-darkcal	4782	0	---	0,00
319	lys	Ready	Finished	---	r0319-lys	161814	6120	---	3,78
320	lys	Ready	Finished	---	r0320-lys	37218	1599	---	4,29
321	lys	Ready	Finished	---	r0321-lys	27771	376	---	1,35
322	lys	Ready	Finished	---	r0322-lys	8673	9	---	0,10
323	lys	Ready	Finished	---	r0323-lys	131903	1969	---	1,49
324	lys	Ready	Finished	---	r0324-lys	98900	3135	---	3,16
325	lys	Ready	Finished	---	r0325-lys	83116	2610	---	3,14
326	lys	Ready	Finished	---	r0326-lys	11036	693	---	6,27
327	lys	Ready	Finished	---	r0327-lys	87749	3313	---	3,77
328	lys	Ready	Finished	---	r0328-lys	56461	1855	---	3,28
329	lys	Ready	Finished	---	r0329-lys	248828	10019	---	4,02
330	lys	Ready	Finished	---	r0330-lys	14601	1430	---	9,79
331	darkcal	Ready	Finished	---	r0331-darkcal	1246	0	---	0,00
332	lys	Ready	Finished	---	r0332-lys	148759	9415	---	6,32
333	lys	Ready	Finished	---	r0333-lys	72737	11041	---	15,1
334	darkcal	Ready	Finished	---	r0334-darkcal	2124	0	---	0,00
335	lys	Ready	Finished	---	r0335-lys	24722	3110	---	12,5
336	lys	Ready	Finished	---	r0336-lys	44751	6871	---	15,3
337	---	Ready	---	---	---	---	---	---	---
338	darkcal	Ready	Finished	---	r0338-darkcal	2346	0	---	0,00
339	---	Ready	---	---	---	---	---	---	---
340	---	Ready	---	---	---	---	---	---	---
341	---	Ready	---	---	---	---	---	---	---
342	---	Ready	---	---	---	---	---	---	---
343	darkcal	Ready	Finished	---	r0343-darkcal	1296	0	---	0,00

Newly collected data (new runs) appear automatically ready to process

Status of data collection

# Cheetah functionality: Processing control

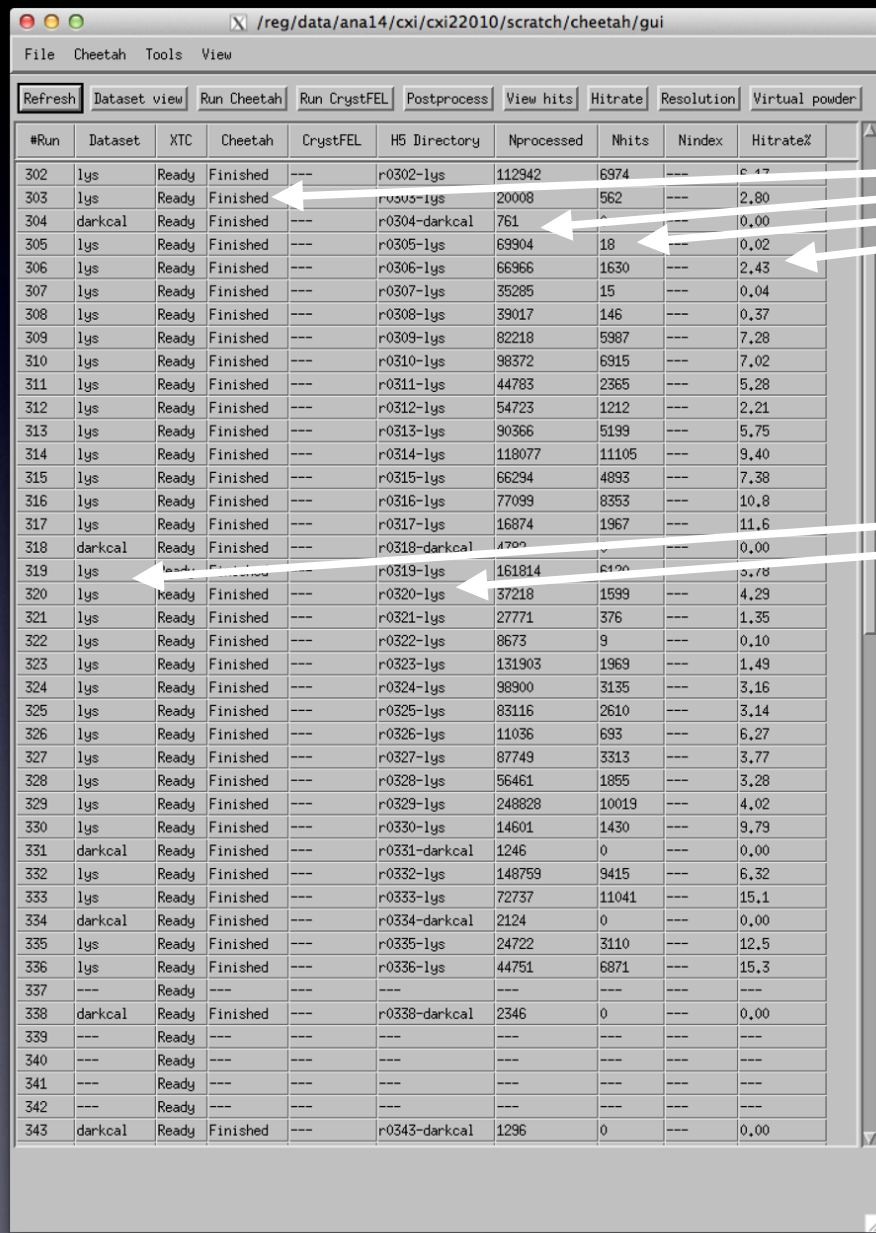


The screenshot shows the Cheetah GUI window with the following menu items: File, Cheetah, Tools, View. The toolbar contains buttons for Refresh, Dataset view, Run Cheetah, CrustFEL, Postprocess, View hits, Htrate, Resolution, and Virtual powder. The 'Run Cheetah' button is highlighted with a white arrow. Below the toolbar is a table with the following columns: #Run, Dataset, XTC, Cheetah, CrystFEL, H5 Directory, Nprocessed, Nhits, Htrate, and Htrate%. The table contains 44 rows of data, with the first row being #302 and the last row being #343.

#Run	Dataset	XTC	Cheetah	CrystFEL	H5 Directory	Nprocessed	Nhits	Htrate	Htrate%
302	lys	Ready	Finished	---	r0302-lys	112942	6974	---	6.17
303	lys	Ready	Finished	---	r0303-lys	20008	562	---	2.80
304	darkcal	Ready	Finished	---	r0304-darkcal	761	0	---	0,00
305	lys	Ready	Finished	---	r0305-lys	69904	18	---	0,02
306	lys	Ready	Finished	---	r0306-lys	66966	1630	---	2,43
307	lys	Ready	Finished	---	r0307-lys	35285	15	---	0,04
308	lys	Ready	Finished	---	r0308-lys	39017	146	---	0,37
309	lys	Ready	Finished	---	r0309-lys	82218	5987	---	7,28
310	lys	Ready	Finished	---	r0310-lys	98372	6915	---	7,02
311	lys	Ready	Finished	---	r0311-lys	44783	2365	---	5,28
312	lys	Ready	Finished	---	r0312-lys	54723	1212	---	2,21
313	lys	Ready	Finished	---	r0313-lys	90366	5199	---	5,75
314	lys	Ready	Finished	---	r0314-lys	118077	11105	---	9,40
315	lys	Ready	Finished	---	r0315-lys	66294	4893	---	7,38
316	lys	Ready	Finished	---	r0316-lys	77099	8353	---	10,8
317	lys	Ready	Finished	---	r0317-lys	16874	1967	---	11,6
318	darkcal	Ready	Finished	---	r0318-darkcal	4782	0	---	0,00
319	lys	Ready	Finished	---	r0319-lys	161814	6120	---	3,78
320	lys	Ready	Finished	---	r0320-lys	37218	1599	---	4,29
321	lys	Ready	Finished	---	r0321-lys	27771	376	---	1,35
322	lys	Ready	Finished	---	r0322-lys	8673	9	---	0,10
323	lys	Ready	Finished	---	r0323-lys	131903	1969	---	1,49
324	lys	Ready	Finished	---	r0324-lys	98900	3135	---	3,16
325	lys	Ready	Finished	---	r0325-lys	83116	2610	---	3,14
326	lys	Ready	Finished	---	r0326-lys	11036	693	---	6,27
327	lys	Ready	Finished	---	r0327-lys	87749	3313	---	3,77
328	lys	Ready	Finished	---	r0328-lys	56461	1855	---	3,28
329	lys	Ready	Finished	---	r0329-lys	248828	10019	---	4,02
330	lys	Ready	Finished	---	r0330-lys	14601	1430	---	9,79
331	darkcal	Ready	Finished	---	r0331-darkcal	1246	0	---	0,00
332	lys	Ready	Finished	---	r0332-lys	148759	9415	---	6,32
333	lys	Ready	Finished	---	r0333-lys	72737	11041	---	15,1
334	darkcal	Ready	Finished	---	r0334-darkcal	2124	0	---	0,00
335	lys	Ready	Finished	---	r0335-lys	24722	3110	---	12,5
336	lys	Ready	Finished	---	r0336-lys	44751	6871	---	15,3
337	---	Ready	---	---	---	---	---	---	---
338	darkcal	Ready	Finished	---	r0338-darkcal	2346	0	---	0,00
339	---	Ready	---	---	---	---	---	---	---
340	---	Ready	---	---	---	---	---	---	---
341	---	Ready	---	---	---	---	---	---	---
342	---	Ready	---	---	---	---	---	---	---
343	darkcal	Ready	Finished	---	r0343-darkcal	1296	0	---	0,00

One-click to start the processing of data sets

# Cheetah functionality: Processing status monitor



The screenshot shows a window titled "/reg/data/ana14/cxi/cxi22010/scratch/cheetah/gui". The window contains a menu bar (File, Cheetah, Tools, View) and a toolbar with buttons: Refresh, Dataset view, Run Cheetah, Run CrystFEL, Postprocess, View hits, Hitrate, Resolution, and Virtual powder. Below the toolbar is a table with the following columns: #Run, Dataset, XTC, Cheetah, CrystFEL, H5 Directory, Nprocessed, Nhits, Nindex, and Hitrate%. The table lists 44 runs, with columns 4, 5, and 6 containing status indicators (---) and directory names. The Hitrate% column shows values ranging from 0.00 to 15.3.

#Run	Dataset	XTC	Cheetah	CrystFEL	H5 Directory	Nprocessed	Nhits	Nindex	Hitrate%
302	lys	Ready	Finished	---	r0302-lys	112942	6974	---	6.17
303	lys	Ready	Finished	---	r0303-lys	20008	562	---	2.80
304	darkcal	Ready	Finished	---	r0304-darkcal	761	---	---	0.00
305	lys	Ready	Finished	---	r0305-lys	69904	18	---	0.02
306	lys	Ready	Finished	---	r0306-lys	66966	1630	---	2.43
307	lys	Ready	Finished	---	r0307-lys	35285	15	---	0.04
308	lys	Ready	Finished	---	r0308-lys	39017	146	---	0.37
309	lys	Ready	Finished	---	r0309-lys	82218	5987	---	7.28
310	lys	Ready	Finished	---	r0310-lys	98372	6915	---	7.02
311	lys	Ready	Finished	---	r0311-lys	44783	2365	---	5.28
312	lys	Ready	Finished	---	r0312-lys	54723	1212	---	2.21
313	lys	Ready	Finished	---	r0313-lys	90366	5199	---	5.75
314	lys	Ready	Finished	---	r0314-lys	118077	11105	---	9.40
315	lys	Ready	Finished	---	r0315-lys	66294	4893	---	7.38
316	lys	Ready	Finished	---	r0316-lys	77099	8353	---	10.8
317	lys	Ready	Finished	---	r0317-lys	16874	1967	---	11.6
318	darkcal	Ready	Finished	---	r0318-darkcal	4799	---	---	0.00
319	lys	Ready	Finished	---	r0319-lys	161814	6190	---	3.78
320	lys	Ready	Finished	---	r0320-lys	37218	1599	---	4.29
321	lys	Ready	Finished	---	r0321-lys	27771	376	---	1.35
322	lys	Ready	Finished	---	r0322-lys	8673	9	---	0.10
323	lys	Ready	Finished	---	r0323-lys	131903	1969	---	1.49
324	lys	Ready	Finished	---	r0324-lys	98900	3135	---	3.16
325	lys	Ready	Finished	---	r0325-lys	83116	2610	---	3.14
326	lys	Ready	Finished	---	r0326-lys	11036	693	---	6.27
327	lys	Ready	Finished	---	r0327-lys	87749	3313	---	3.77
328	lys	Ready	Finished	---	r0328-lys	56461	1855	---	3.28
329	lys	Ready	Finished	---	r0329-lys	248828	10019	---	4.02
330	lys	Ready	Finished	---	r0330-lys	14601	1430	---	9.79
331	darkcal	Ready	Finished	---	r0331-darkcal	1246	0	---	0.00
332	lys	Ready	Finished	---	r0332-lys	148759	9415	---	6.32
333	lys	Ready	Finished	---	r0333-lys	72737	11041	---	15.1
334	darkcal	Ready	Finished	---	r0334-darkcal	2124	0	---	0.00
335	lys	Ready	Finished	---	r0335-lys	24722	3110	---	12.5
336	lys	Ready	Finished	---	r0336-lys	44751	6871	---	15.3
337	---	Ready	---	---	---	---	---	---	---
338	darkcal	Ready	Finished	---	r0338-darkcal	2346	0	---	0.00
339	---	Ready	---	---	---	---	---	---	---
340	---	Ready	---	---	---	---	---	---	---
341	---	Ready	---	---	---	---	---	---	---
342	---	Ready	---	---	---	---	---	---	---
343	darkcal	Ready	Finished	---	r0343-darkcal	1296	0	---	0.00

Status of processing is continually updated

Contents of each run and associated data directory

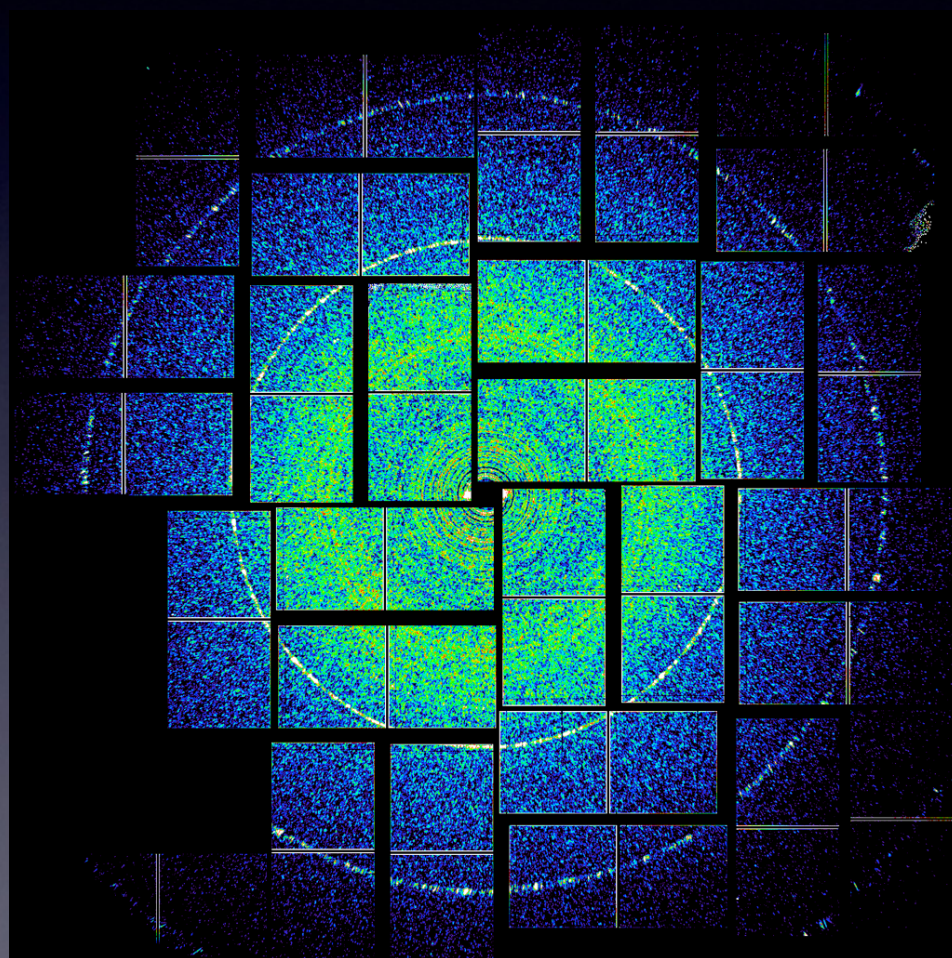
# Cheetah functionality: Run summaries

File Cheetah Tools View

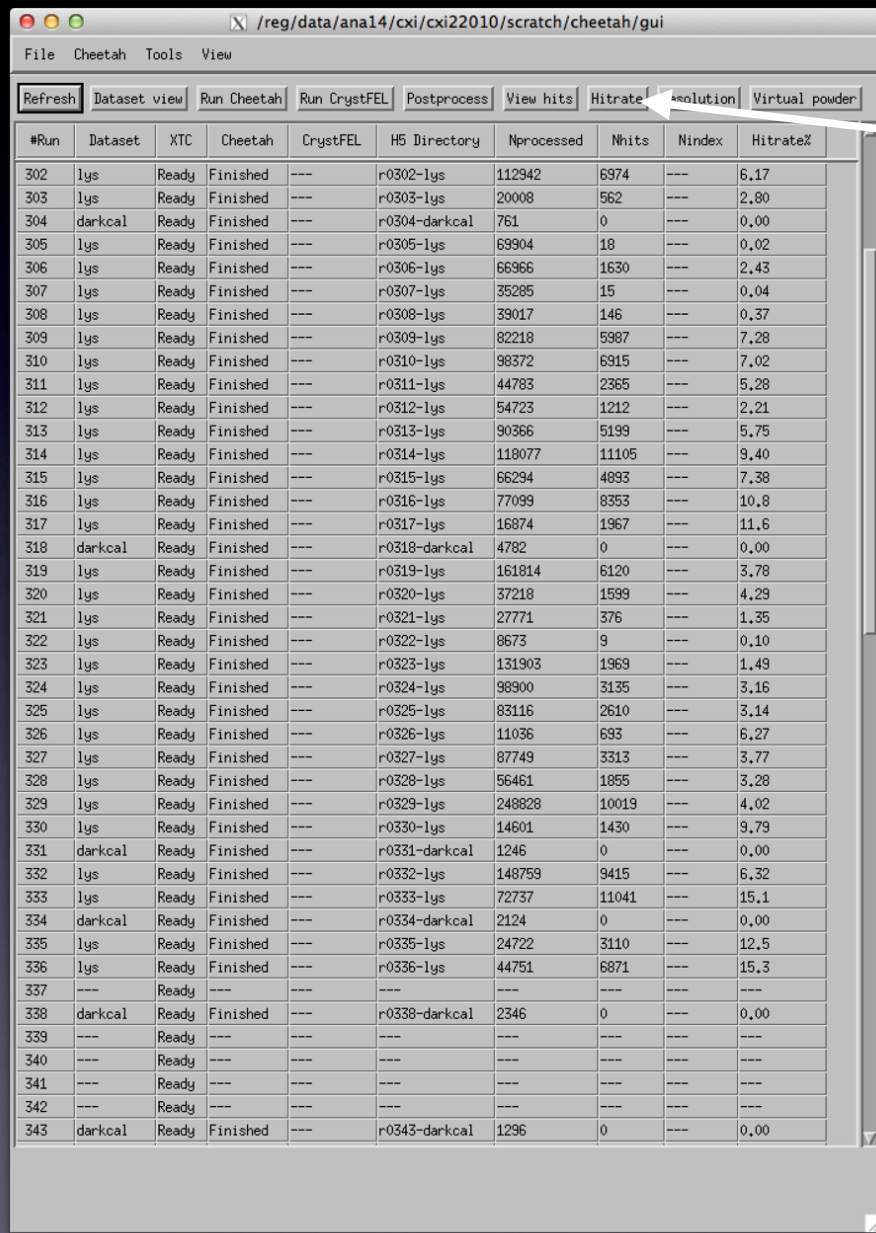
Refresh Dataset view Run Cheetah Run CrystFEL Postprocess View hits Hitrate Resolution Virtual powder

#Run	Dataset	XTC	Cheetah	CrystFEL	H5 Directory	Nprocessed	Nhits	Nindex	Hitrate%
302	lys	Ready	Finished	---	r0302-lys	112942	6974	---	6.17
303	lys	Ready	Finished	---	r0303-lys	20008	562	---	2.80
304	darkcal	Ready	Finished	---	r0304-darkcal	761	0	---	0,00
305	lys	Ready	Finished	---	r0305-lys	69904	18	---	0,02
306	lys	Ready	Finished	---	r0306-lys	66966	1630	---	2,43
307	lys	Ready	Finished	---	r0307-lys	35285	15	---	0,04
308	lys	Ready	Finished	---	r0308-lys	39017	146	---	0,37
309	lys	Ready	Finished	---	r0309-lys	82218	5987	---	7,28
310	lys	Ready	Finished	---	r0310-lys	98372	6915	---	7,02
311	lys	Ready	Finished	---	r0311-lys	44783	2365	---	5,28
312	lys	Ready	Finished	---	r0312-lys	54723	1212	---	2,21
313	lys	Ready	Finished	---	r0313-lys	90366	5199	---	5,75
314	lys	Ready	Finished	---	r0314-lys	118077	11105	---	9,40
315	lys	Ready	Finished	---	r0315-lys	66294	4893	---	7,38
316	lys	Ready	Finished	---	r0316-lys	77099	8353	---	10,8
317	lys	Ready	Finished	---	r0317-lys	16874	1967	---	11,6
318	darkcal	Ready	Finished	---	r0318-darkcal	4782	0	---	0,00
319	lys	Ready	Finished	---	r0319-lys	161814	6120	---	3,78
320	lys	Ready	Finished	---	r0320-lys	37218	1599	---	4,29
321	lys	Ready	Finished	---	r0321-lys	27771	376	---	1,35
322	lys	Ready	Finished	---	r0322-lys	8673	9	---	0,10
323	lys	Ready	Finished	---	r0323-lys	131903	1969	---	1,49
324	lys	Ready	Finished	---	r0324-lys	98900	3135	---	3,16
325	lys	Ready	Finished	---	r0325-lys	83116	2610	---	3,14
326	lys	Ready	Finished	---	r0326-lys	11036	693	---	6,27
327	lys	Ready	Finished	---	r0327-lys	87749	3313	---	3,77
328	lys	Ready	Finished	---	r0328-lys	56461	1855	---	3,28
329	lys	Ready	Finished	---	r0329-lys	248828	10019	---	4,02
330	lys	Ready	Finished	---	r0330-lys	14601	1430	---	9,79
331	darkcal	Ready	Finished	---	r0331-darkcal	1246	0	---	0,00
332	lys	Ready	Finished	---	r0332-lys	148759	9415	---	6,32
333	lys	Ready	Finished	---	r0333-lys	72737	11041	---	15,1
334	darkcal	Ready	Finished	---	r0334-darkcal	2124	0	---	0,00
335	lys	Ready	Finished	---	r0335-lys	24722	3110	---	12,5
336	lys	Ready	Finished	---	r0336-lys	44751	6871	---	15,3
337	---	Ready	---	---	---	---	---	---	---
338	darkcal	Ready	Finished	---	r0338-darkcal	2346	0	---	0,00
339	---	Ready	---	---	---	---	---	---	---
340	---	Ready	---	---	---	---	---	---	---
341	---	Ready	---	---	---	---	---	---	---
342	---	Ready	---	---	---	---	---	---	---
343	darkcal	Ready	Finished	---	r0343-darkcal	1296	0	---	0,00

Virtual powder pattern

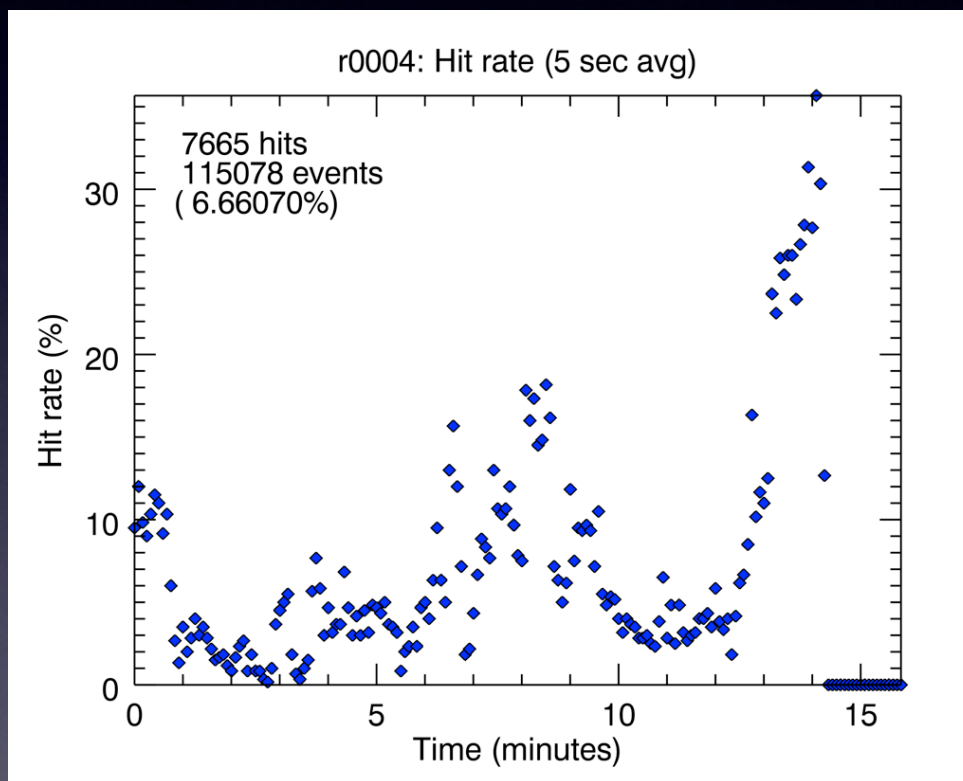


# Cheetah functionality: Hit rates



#Run	Dataset	XTC	Cheetah	CrystFEL	H5 Directory	Nprocessed	Nhits	Nindex	Hitrate%
302	lys	Ready	Finished	---	r0302-lys	112942	6974	---	6.17
303	lys	Ready	Finished	---	r0303-lys	20008	562	---	2.80
304	darkcal	Ready	Finished	---	r0304-darkcal	761	0	---	0,00
305	lys	Ready	Finished	---	r0305-lys	69904	18	---	0,02
306	lys	Ready	Finished	---	r0306-lys	66966	1630	---	2,43
307	lys	Ready	Finished	---	r0307-lys	35285	15	---	0,04
308	lys	Ready	Finished	---	r0308-lys	39017	146	---	0,37
309	lys	Ready	Finished	---	r0309-lys	82218	5987	---	7,28
310	lys	Ready	Finished	---	r0310-lys	98372	6915	---	7,02
311	lys	Ready	Finished	---	r0311-lys	44783	2365	---	5,28
312	lys	Ready	Finished	---	r0312-lys	54723	1212	---	2,21
313	lys	Ready	Finished	---	r0313-lys	90366	5199	---	5,75
314	lys	Ready	Finished	---	r0314-lys	118077	11105	---	9,40
315	lys	Ready	Finished	---	r0315-lys	66294	4893	---	7,38
316	lys	Ready	Finished	---	r0316-lys	77099	8353	---	10,8
317	lys	Ready	Finished	---	r0317-lys	16874	1967	---	11,6
318	darkcal	Ready	Finished	---	r0318-darkcal	4782	0	---	0,00
319	lys	Ready	Finished	---	r0319-lys	161814	6120	---	3,78
320	lys	Ready	Finished	---	r0320-lys	37218	1599	---	4,29
321	lys	Ready	Finished	---	r0321-lys	27771	376	---	1,35
322	lys	Ready	Finished	---	r0322-lys	8673	9	---	0,10
323	lys	Ready	Finished	---	r0323-lys	131903	1969	---	1,49
324	lys	Ready	Finished	---	r0324-lys	98900	3135	---	3,16
325	lys	Ready	Finished	---	r0325-lys	83116	2610	---	3,14
326	lys	Ready	Finished	---	r0326-lys	11036	693	---	6,27
327	lys	Ready	Finished	---	r0327-lys	87749	3313	---	3,77
328	lys	Ready	Finished	---	r0328-lys	56461	1855	---	3,28
329	lys	Ready	Finished	---	r0329-lys	248828	10019	---	4,02
330	lys	Ready	Finished	---	r0330-lys	14601	1430	---	9,79
331	darkcal	Ready	Finished	---	r0331-darkcal	1246	0	---	0,00
332	lys	Ready	Finished	---	r0332-lys	148759	9415	---	6,32
333	lys	Ready	Finished	---	r0333-lys	72737	11041	---	15,1
334	darkcal	Ready	Finished	---	r0334-darkcal	2124	0	---	0,00
335	lys	Ready	Finished	---	r0335-lys	24722	3110	---	12,5
336	lys	Ready	Finished	---	r0336-lys	44751	6871	---	15,3
337	---	Ready	---	---	---	---	---	---	---
338	darkcal	Ready	Finished	---	r0338-darkcal	2346	0	---	0,00
339	---	Ready	---	---	---	---	---	---	---
340	---	Ready	---	---	---	---	---	---	---
341	---	Ready	---	---	---	---	---	---	---
342	---	Ready	---	---	---	---	---	---	---
343	darkcal	Ready	Finished	---	r0343-darkcal	1296	0	---	0,00

Hit rate





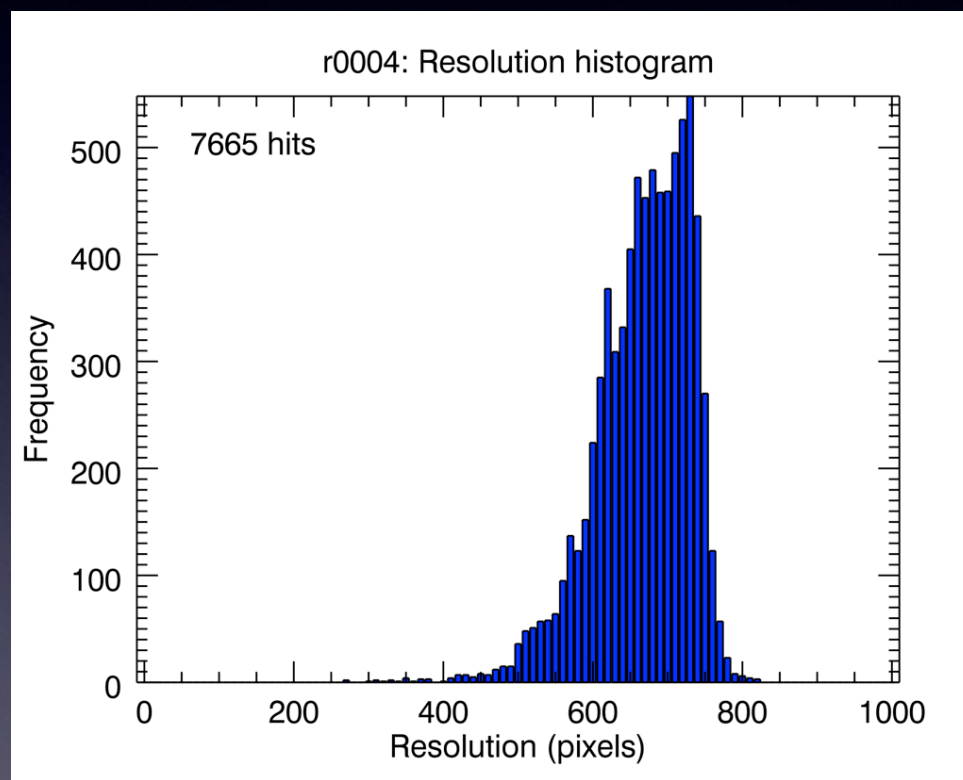
# Cheetah functionality: Resolution

File Cheetah Tools View

Refresh Dataset view Run Cheetah Run CrystFEL Postprocess View hits Hitrate Resolution **Actual powder**

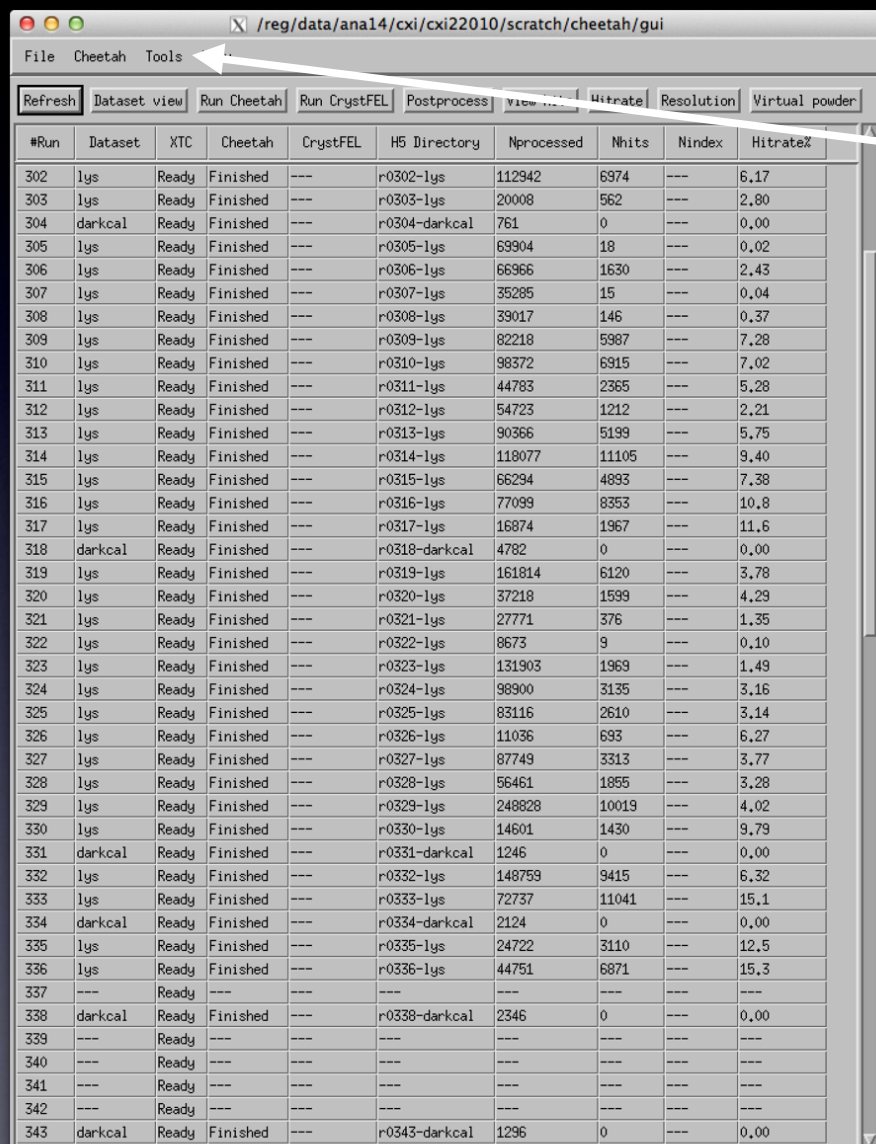
#Run	Dataset	XTC	Cheetah	CrystFEL	H5 Directory	Nprocessed	Nhits	Nindex	Hitrate%
302	lys	Ready	Finished	---	r0302-lys	112942	6974	---	6.17
303	lys	Ready	Finished	---	r0303-lys	20008	562	---	2.80
304	darkcal	Ready	Finished	---	r0304-darkcal	761	0	---	0,00
305	lys	Ready	Finished	---	r0305-lys	69904	18	---	0,02
306	lys	Ready	Finished	---	r0306-lys	66966	1630	---	2,43
307	lys	Ready	Finished	---	r0307-lys	35285	15	---	0,04
308	lys	Ready	Finished	---	r0308-lys	39017	146	---	0,37
309	lys	Ready	Finished	---	r0309-lys	82218	5987	---	7,28
310	lys	Ready	Finished	---	r0310-lys	98372	6915	---	7,02
311	lys	Ready	Finished	---	r0311-lys	44783	2365	---	5,28
312	lys	Ready	Finished	---	r0312-lys	54723	1212	---	2,21
313	lys	Ready	Finished	---	r0313-lys	90366	5199	---	5,75
314	lys	Ready	Finished	---	r0314-lys	118077	11105	---	9,40
315	lys	Ready	Finished	---	r0315-lys	66294	4893	---	7,38
316	lys	Ready	Finished	---	r0316-lys	77099	8353	---	10,8
317	lys	Ready	Finished	---	r0317-lys	16874	1967	---	11,6
318	darkcal	Ready	Finished	---	r0318-darkcal	4782	0	---	0,00
319	lys	Ready	Finished	---	r0319-lys	161814	6120	---	3,78
320	lys	Ready	Finished	---	r0320-lys	37218	1599	---	4,29
321	lys	Ready	Finished	---	r0321-lys	27771	376	---	1,35
322	lys	Ready	Finished	---	r0322-lys	8673	9	---	0,10
323	lys	Ready	Finished	---	r0323-lys	131903	1969	---	1,49
324	lys	Ready	Finished	---	r0324-lys	98900	3135	---	3,16
325	lys	Ready	Finished	---	r0325-lys	83116	2610	---	3,14
326	lys	Ready	Finished	---	r0326-lys	11036	693	---	6,27
327	lys	Ready	Finished	---	r0327-lys	87749	3313	---	3,77
328	lys	Ready	Finished	---	r0328-lys	56461	1855	---	3,28
329	lys	Ready	Finished	---	r0329-lys	248828	10019	---	4,02
330	lys	Ready	Finished	---	r0330-lys	14601	1430	---	9,79
331	darkcal	Ready	Finished	---	r0331-darkcal	1246	0	---	0,00
332	lys	Ready	Finished	---	r0332-lys	148759	9415	---	6,32
333	lys	Ready	Finished	---	r0333-lys	72737	11041	---	15,1
334	darkcal	Ready	Finished	---	r0334-darkcal	2124	0	---	0,00
335	lys	Ready	Finished	---	r0335-lys	24722	3110	---	12,5
336	lys	Ready	Finished	---	r0336-lys	44751	6871	---	15,3
337	---	Ready	---	---	---	---	---	---	---
338	darkcal	Ready	Finished	---	r0338-darkcal	2346	0	---	0,00
339	---	Ready	---	---	---	---	---	---	---
340	---	Ready	---	---	---	---	---	---	---
341	---	Ready	---	---	---	---	---	---	---
342	---	Ready	---	---	---	---	---	---	---
343	darkcal	Ready	Finished	---	r0343-darkcal	1296	0	---	0,00

Resolution



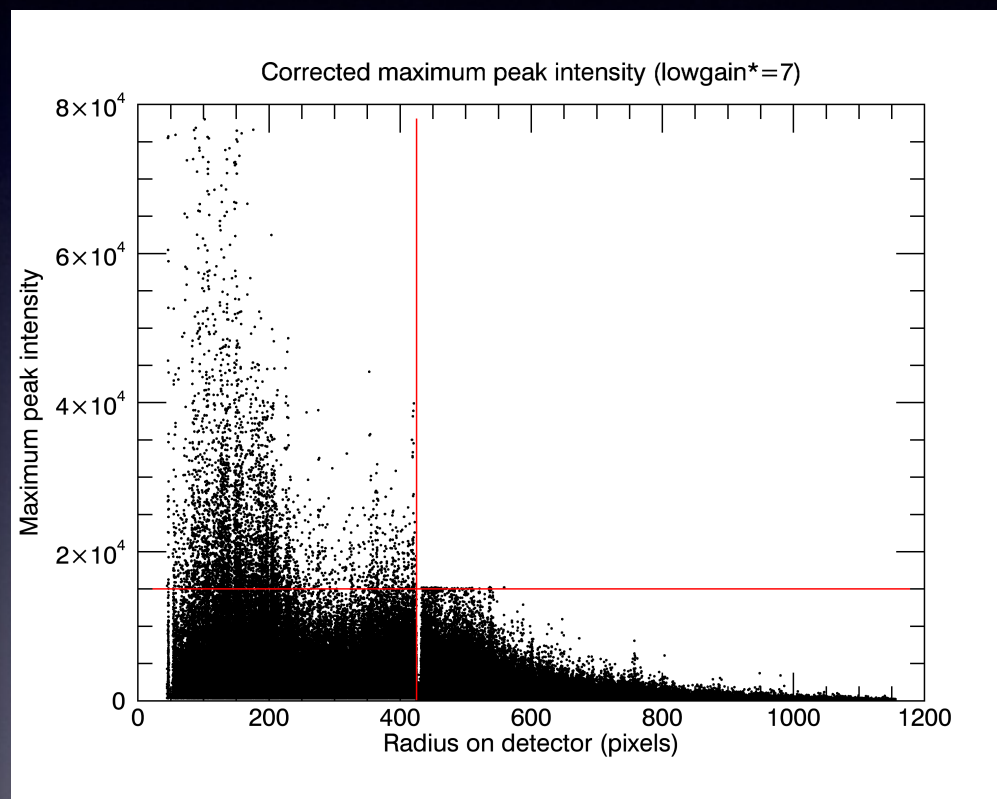
Circle containing 80% of found peaks

# Cheetah functionality: Detector saturation check



#Run	Dataset	XTC	Cheetah	CrystFEL	H5 Directory	Nprocessed	Nhits	Nindex	Hirate%
302	lys	Ready	Finished	---	r0302-lys	112942	6974	---	6.17
303	lys	Ready	Finished	---	r0303-lys	20008	562	---	2.80
304	darkcal	Ready	Finished	---	r0304-darkcal	761	0	---	0.00
305	lys	Ready	Finished	---	r0305-lys	69904	18	---	0.02
306	lys	Ready	Finished	---	r0306-lys	66966	1630	---	2.43
307	lys	Ready	Finished	---	r0307-lys	35285	15	---	0.04
308	lys	Ready	Finished	---	r0308-lys	39017	146	---	0.37
309	lys	Ready	Finished	---	r0309-lys	82218	5987	---	7.28
310	lys	Ready	Finished	---	r0310-lys	98372	6915	---	7.02
311	lys	Ready	Finished	---	r0311-lys	44783	2365	---	5.28
312	lys	Ready	Finished	---	r0312-lys	54723	1212	---	2.21
313	lys	Ready	Finished	---	r0313-lys	90366	5199	---	5.75
314	lys	Ready	Finished	---	r0314-lys	118077	11105	---	9.40
315	lys	Ready	Finished	---	r0315-lys	66294	4893	---	7.38
316	lys	Ready	Finished	---	r0316-lys	77099	8353	---	10.8
317	lys	Ready	Finished	---	r0317-lys	16874	1967	---	11.6
318	darkcal	Ready	Finished	---	r0318-darkcal	4782	0	---	0.00
319	lys	Ready	Finished	---	r0319-lys	161814	6120	---	3.78
320	lys	Ready	Finished	---	r0320-lys	37218	1599	---	4.29
321	lys	Ready	Finished	---	r0321-lys	27771	376	---	1.35
322	lys	Ready	Finished	---	r0322-lys	8673	9	---	0.10
323	lys	Ready	Finished	---	r0323-lys	131903	1969	---	1.49
324	lys	Ready	Finished	---	r0324-lys	98900	3135	---	3.16
325	lys	Ready	Finished	---	r0325-lys	83116	2610	---	3.14
326	lys	Ready	Finished	---	r0326-lys	11036	693	---	6.27
327	lys	Ready	Finished	---	r0327-lys	87749	3313	---	3.77
328	lys	Ready	Finished	---	r0328-lys	56461	1855	---	3.28
329	lys	Ready	Finished	---	r0329-lys	248828	10019	---	4.02
330	lys	Ready	Finished	---	r0330-lys	14601	1430	---	9.79
331	darkcal	Ready	Finished	---	r0331-darkcal	1246	0	---	0.00
332	lys	Ready	Finished	---	r0332-lys	148759	9415	---	6.32
333	lys	Ready	Finished	---	r0333-lys	72737	11041	---	15.1
334	darkcal	Ready	Finished	---	r0334-darkcal	2124	0	---	0.00
335	lys	Ready	Finished	---	r0335-lys	24722	3110	---	12.5
336	lys	Ready	Finished	---	r0336-lys	44751	6871	---	15.3
337	---	Ready	---	---	---	---	---	---	---
338	darkcal	Ready	Finished	---	r0338-darkcal	2346	0	---	0.00
339	---	Ready	---	---	---	---	---	---	---
340	---	Ready	---	---	---	---	---	---	---
341	---	Ready	---	---	---	---	---	---	---
342	---	Ready	---	---	---	---	---	---	---
343	darkcal	Ready	Finished	---	r0343-darkcal	1296	0	---	0.00

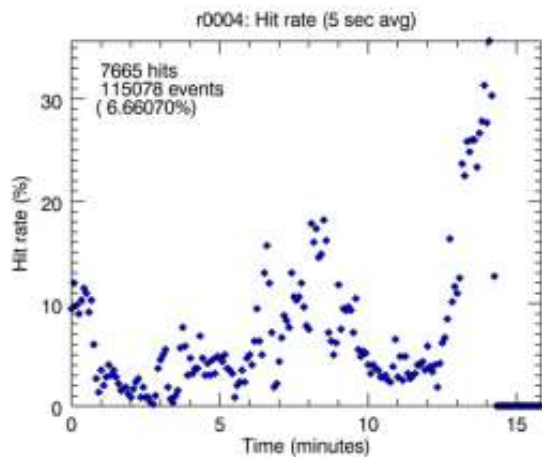
Saturation check



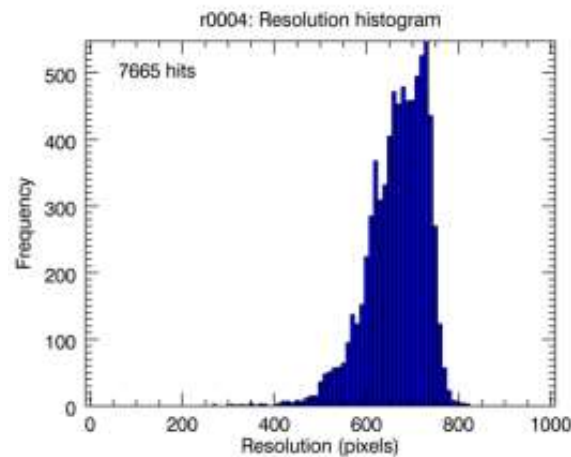
+ support for dual gain mode

# Cheetah: useful diagnostic tool

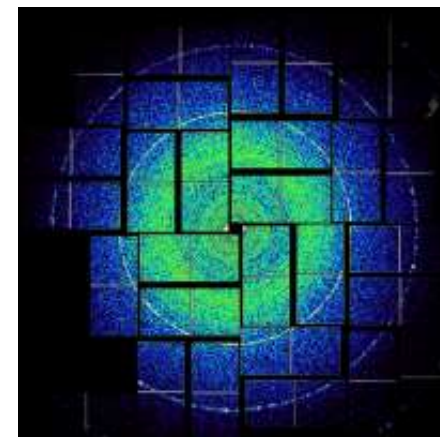
## Hit rates



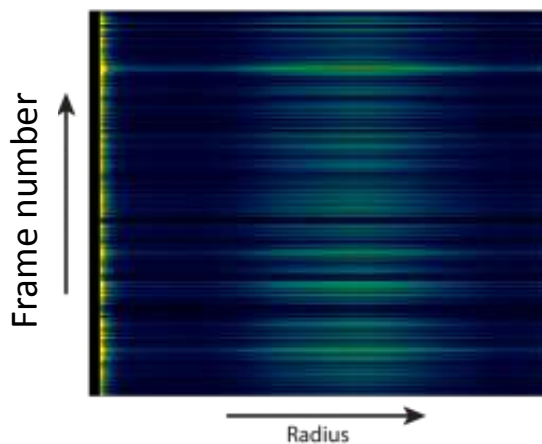
## Resolution



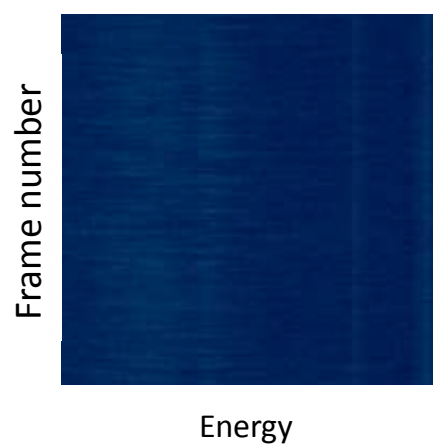
## Virtual powder



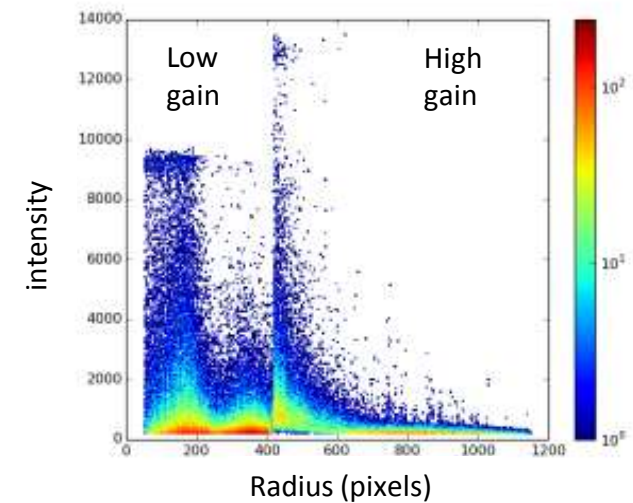
## Radial stacks



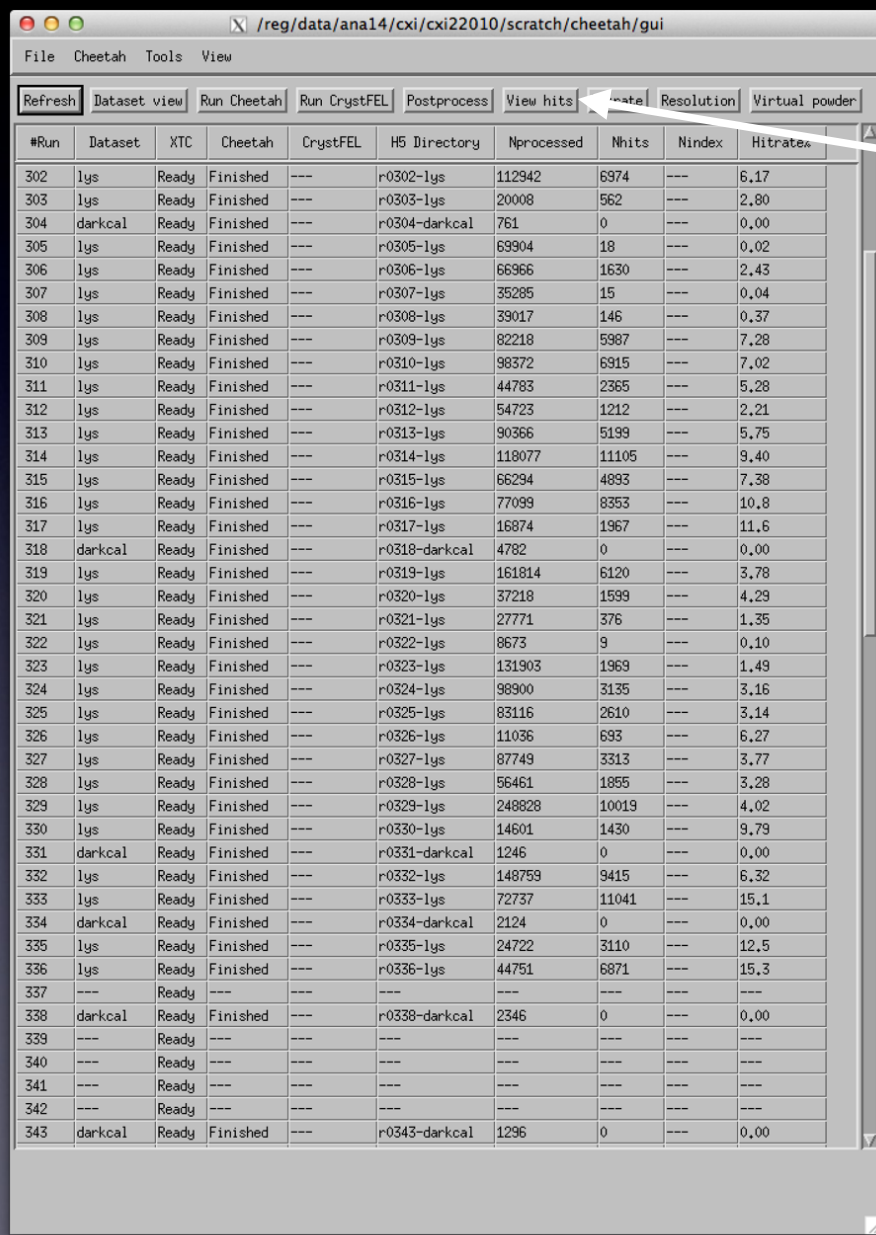
## Spectral stack



## Saturation plots

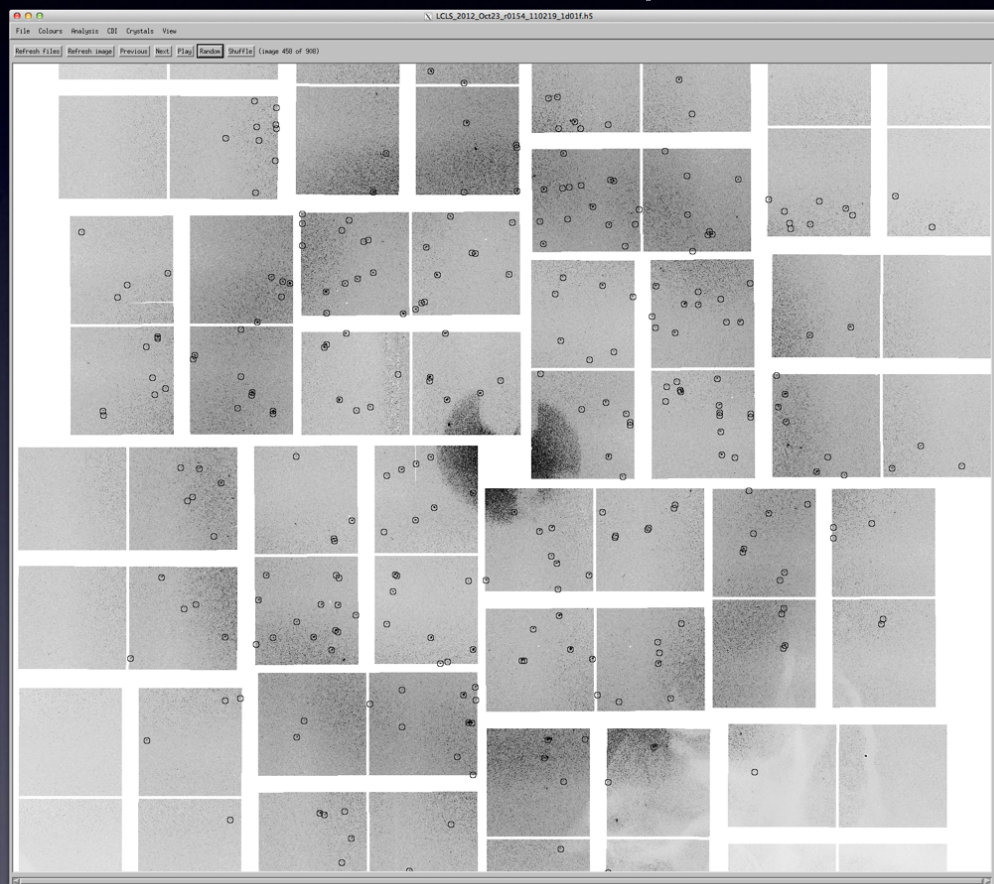


# Cheetah functionality: Data inspection

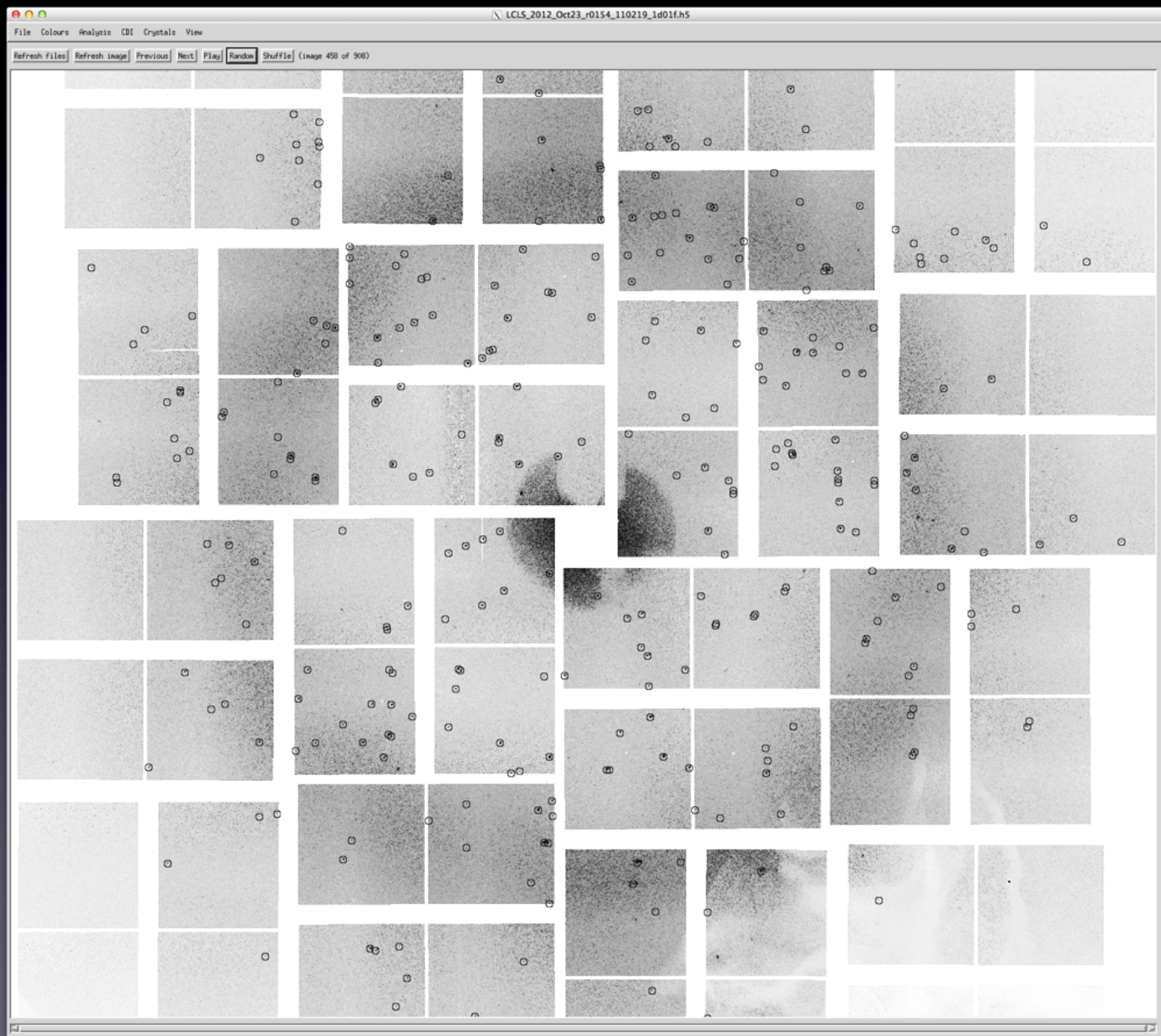


#Run	Dataset	XTC	Cheetah	CrystFEL	H5 Directory	Nprocessed	Nhits	Nindex	Hitratio
302	lys	Ready	Finished	---	r0302-lys	112942	6974	---	6.17
303	lys	Ready	Finished	---	r0303-lys	20008	562	---	2.80
304	darkcal	Ready	Finished	---	r0304-darkcal	761	0	---	0,00
305	lys	Ready	Finished	---	r0305-lys	69904	18	---	0,02
306	lys	Ready	Finished	---	r0306-lys	66966	1630	---	2,43
307	lys	Ready	Finished	---	r0307-lys	35285	15	---	0,04
308	lys	Ready	Finished	---	r0308-lys	39017	146	---	0,37
309	lys	Ready	Finished	---	r0309-lys	82218	5987	---	7,28
310	lys	Ready	Finished	---	r0310-lys	98372	6915	---	7,02
311	lys	Ready	Finished	---	r0311-lys	44783	2365	---	5,28
312	lys	Ready	Finished	---	r0312-lys	54723	1212	---	2,21
313	lys	Ready	Finished	---	r0313-lys	90366	5199	---	5,75
314	lys	Ready	Finished	---	r0314-lys	118077	11105	---	9,40
315	lys	Ready	Finished	---	r0315-lys	66294	4893	---	7,38
316	lys	Ready	Finished	---	r0316-lys	77099	8353	---	10,8
317	lys	Ready	Finished	---	r0317-lys	16874	1967	---	11,6
318	darkcal	Ready	Finished	---	r0318-darkcal	4782	0	---	0,00
319	lys	Ready	Finished	---	r0319-lys	161814	6120	---	3,78
320	lys	Ready	Finished	---	r0320-lys	37218	1599	---	4,29
321	lys	Ready	Finished	---	r0321-lys	27771	376	---	1,35
322	lys	Ready	Finished	---	r0322-lys	8673	9	---	0,10
323	lys	Ready	Finished	---	r0323-lys	131903	1969	---	1,49
324	lys	Ready	Finished	---	r0324-lys	98900	3135	---	3,16
325	lys	Ready	Finished	---	r0325-lys	83116	2610	---	3,14
326	lys	Ready	Finished	---	r0326-lys	11036	693	---	6,27
327	lys	Ready	Finished	---	r0327-lys	87749	3313	---	3,77
328	lys	Ready	Finished	---	r0328-lys	56461	1855	---	3,28
329	lys	Ready	Finished	---	r0329-lys	248828	10019	---	4,02
330	lys	Ready	Finished	---	r0330-lys	14601	1430	---	9,79
331	darkcal	Ready	Finished	---	r0331-darkcal	1246	0	---	0,00
332	lys	Ready	Finished	---	r0332-lys	148759	9415	---	6,32
333	lys	Ready	Finished	---	r0333-lys	72737	11041	---	15,1
334	darkcal	Ready	Finished	---	r0334-darkcal	2124	0	---	0,00
335	lys	Ready	Finished	---	r0335-lys	24722	3110	---	12,5
336	lys	Ready	Finished	---	r0336-lys	44751	6871	---	15,3
337	---	Ready	---	---	---	---	---	---	---
338	darkcal	Ready	Finished	---	r0338-darkcal	2346	0	---	0,00
339	---	Ready	---	---	---	---	---	---	---
340	---	Ready	---	---	---	---	---	---	---
341	---	Ready	---	---	---	---	---	---	---
342	---	Ready	---	---	---	---	---	---	---
343	darkcal	Ready	Finished	---	r0343-darkcal	1296	0	---	0,00

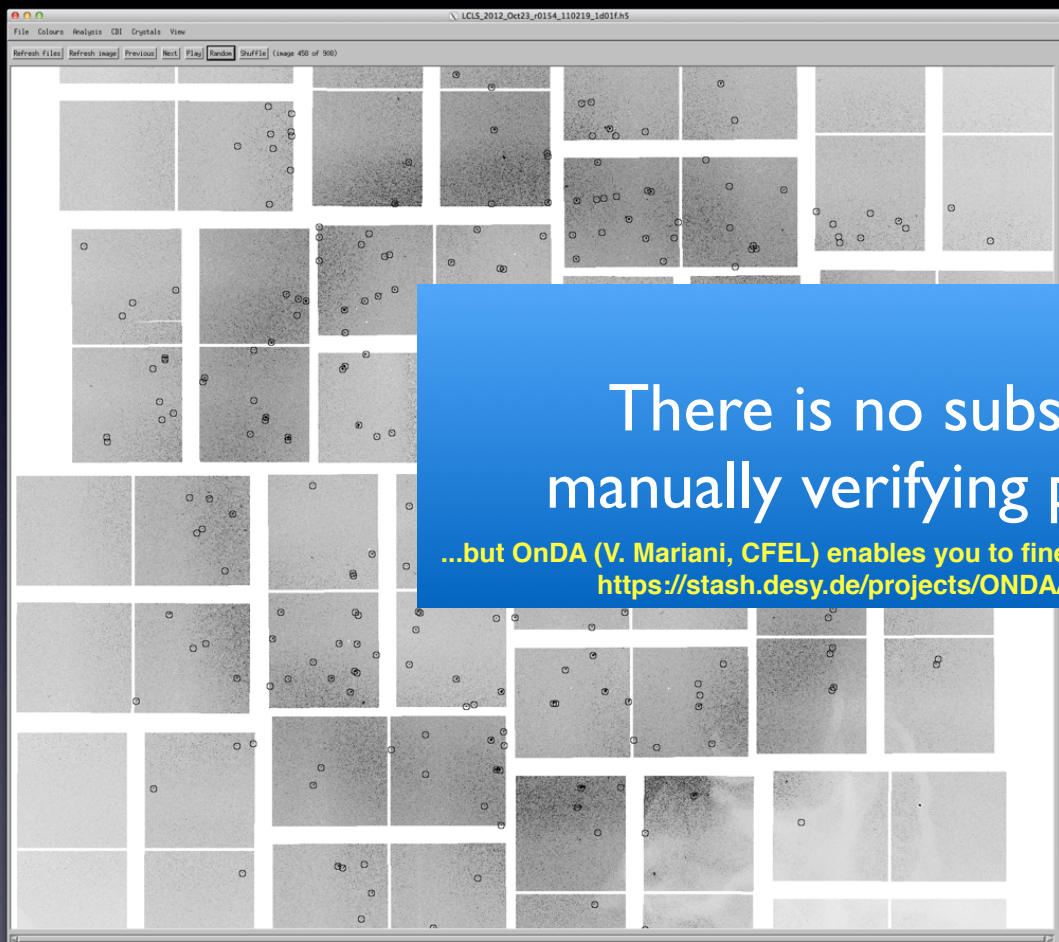
Individual diffraction patterns



# Cheetah functionality: Data viewer for checking peak finding



# It is all boils down to quick but accurate (enough) peak finding



There is no substitute for manually verifying peak finding

...but OnDA (V. Mariani, CFEL) enables you to finetune peak picking interactively:  
<https://stash.desy.de/projects/ONDA/repos/onda/browse>

## “Hitfinder 8”

1. Calculate radial SNR and offset
2.  $\text{thresh}(r) = \text{offset} + s * \text{sigma}$

minADC

an npix connected  
 $\text{thresh}(r)$

1. Require more than n peaks

Critical inputs:

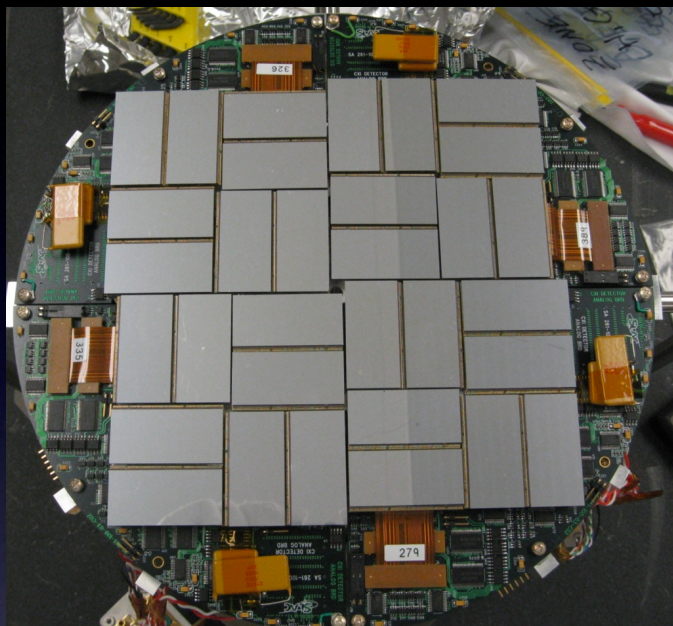
SNR

npix

minADC

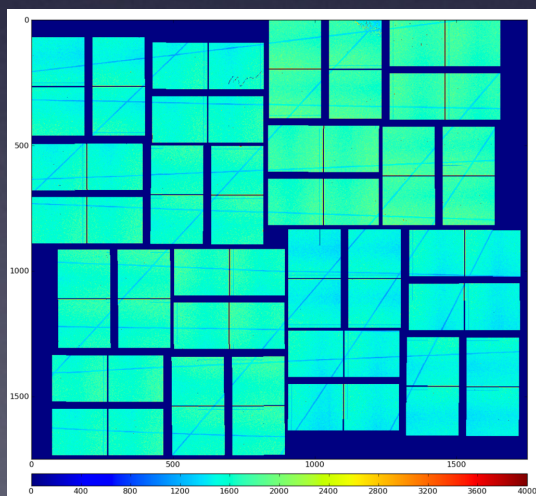
rmin, rmax

# The full cspad detector consists of many tiles of smaller detectors

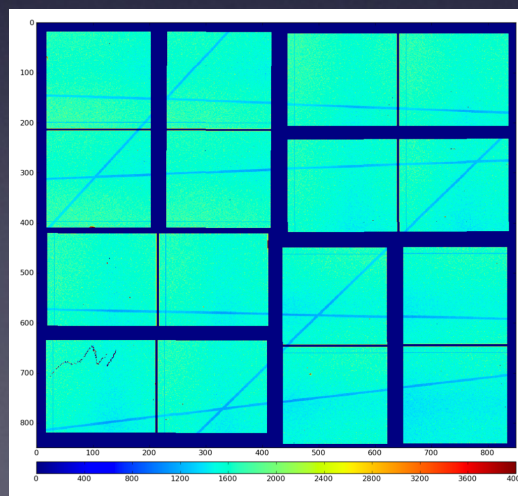


- Each ASIC is a separate detector
- Each quadrant can move independently

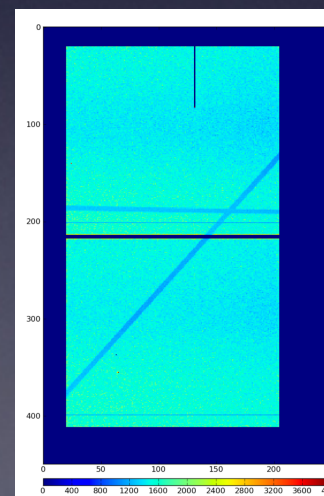
Detector geometry is very important  
(and not a trivial problem)



Full detector



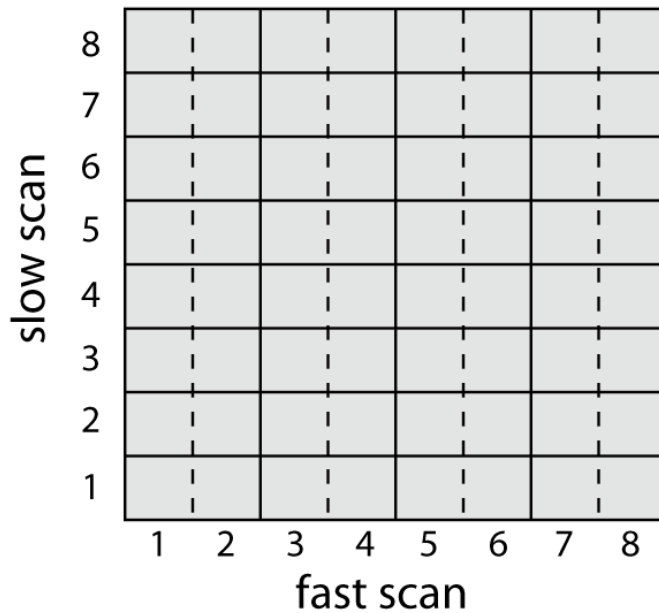
Quadrant



2x1

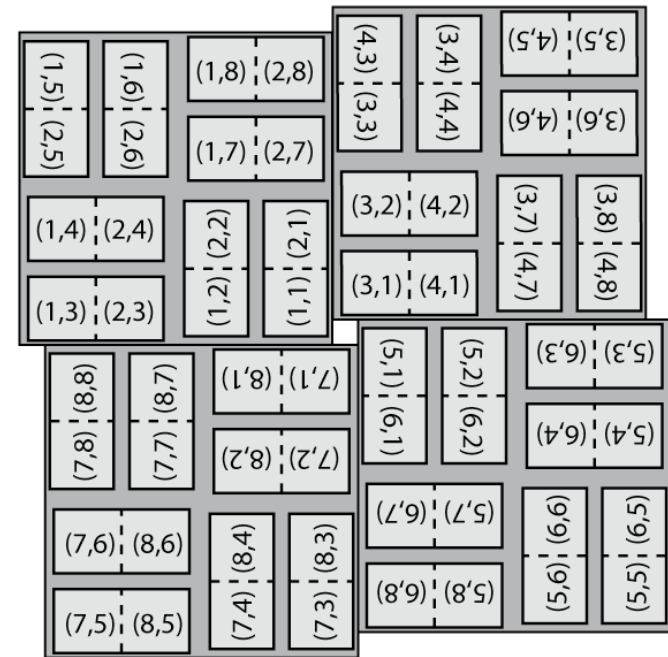
There is no need to assemble one geometrically correct image;  
each module is a collection of pixels placed somewhere in space

a)



Data layout

b)



Physical layout

Geometry is specified in a pixel map:

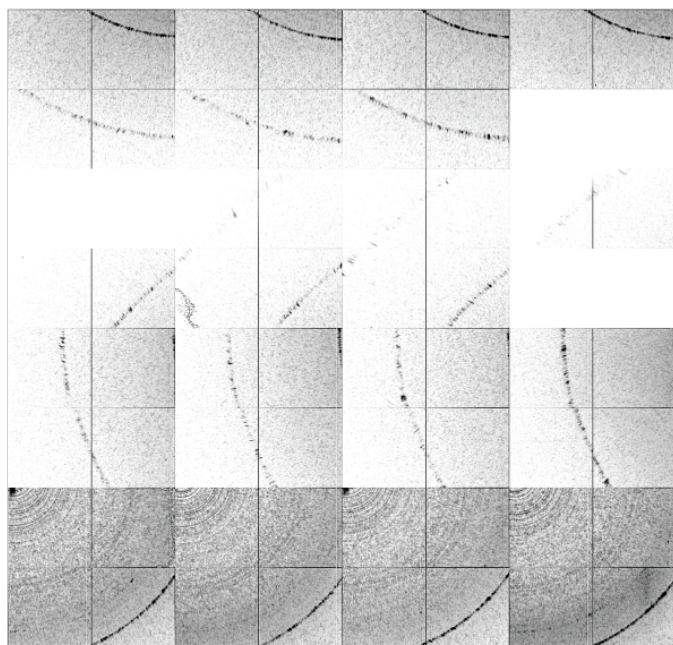
HDF5 file with (x,y,z) coordinate of each pixel in experiment space

as seen when looking downstream (ie: looking at image projected onto front of detector)



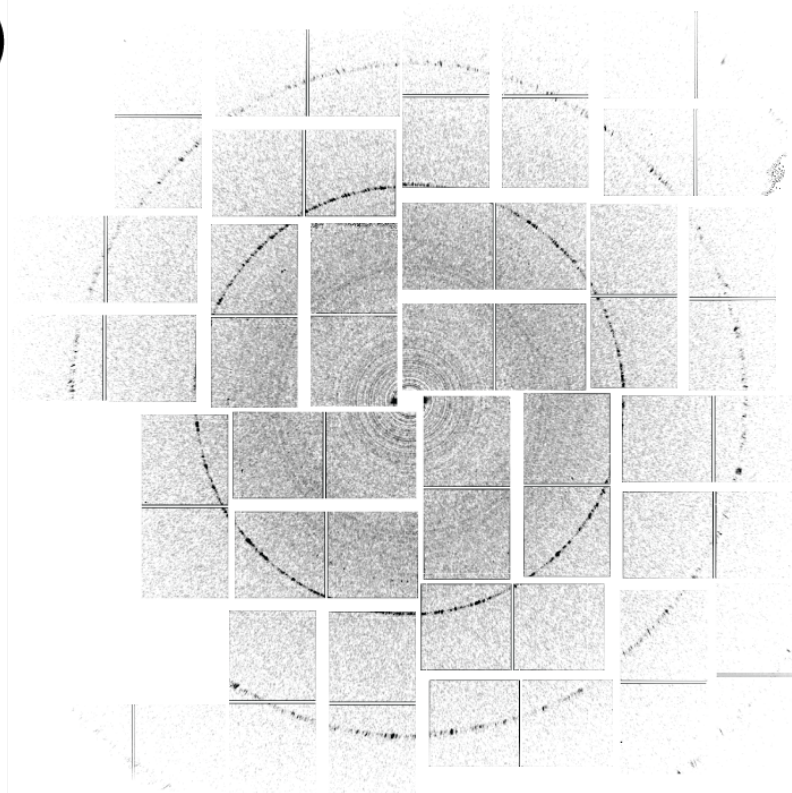
There is no need to assemble one geometrically correct image;  
each module is a collection of pixels placed somewhere in space

a)



Data layout in data file

b)

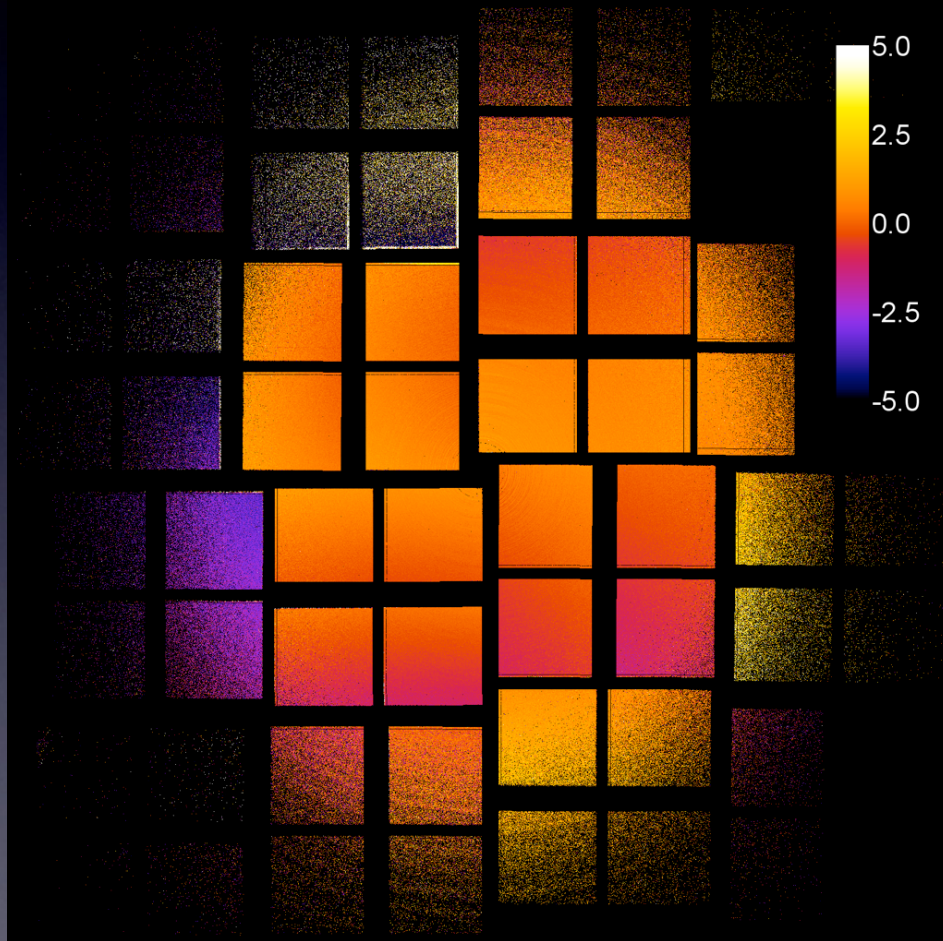


Physical layout

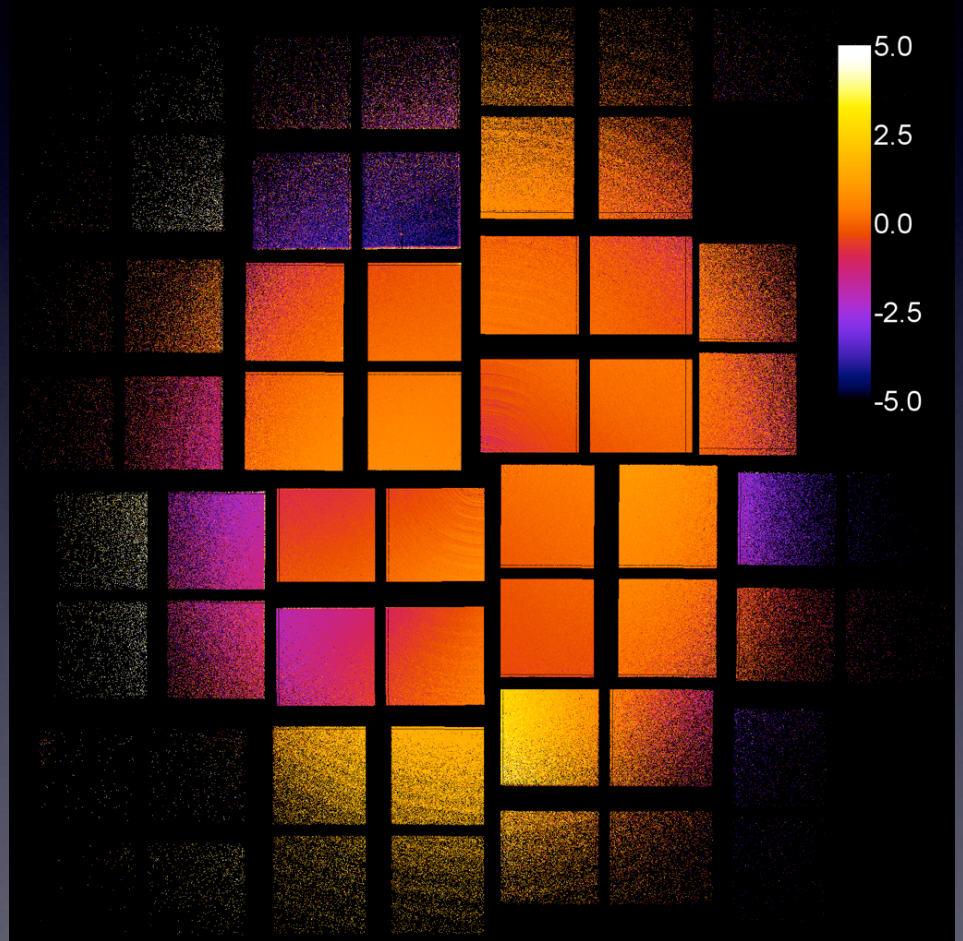
Geometry is specified in a pixel map:  
HDF5 file with (x,y,z) coordinate of each pixel in experiment space  
as seen when looking downstream (ie: looking at image projected onto front of detector)

# Detector geometry is accurately determined by comparing observed and predicted peak locations

Fast scan error

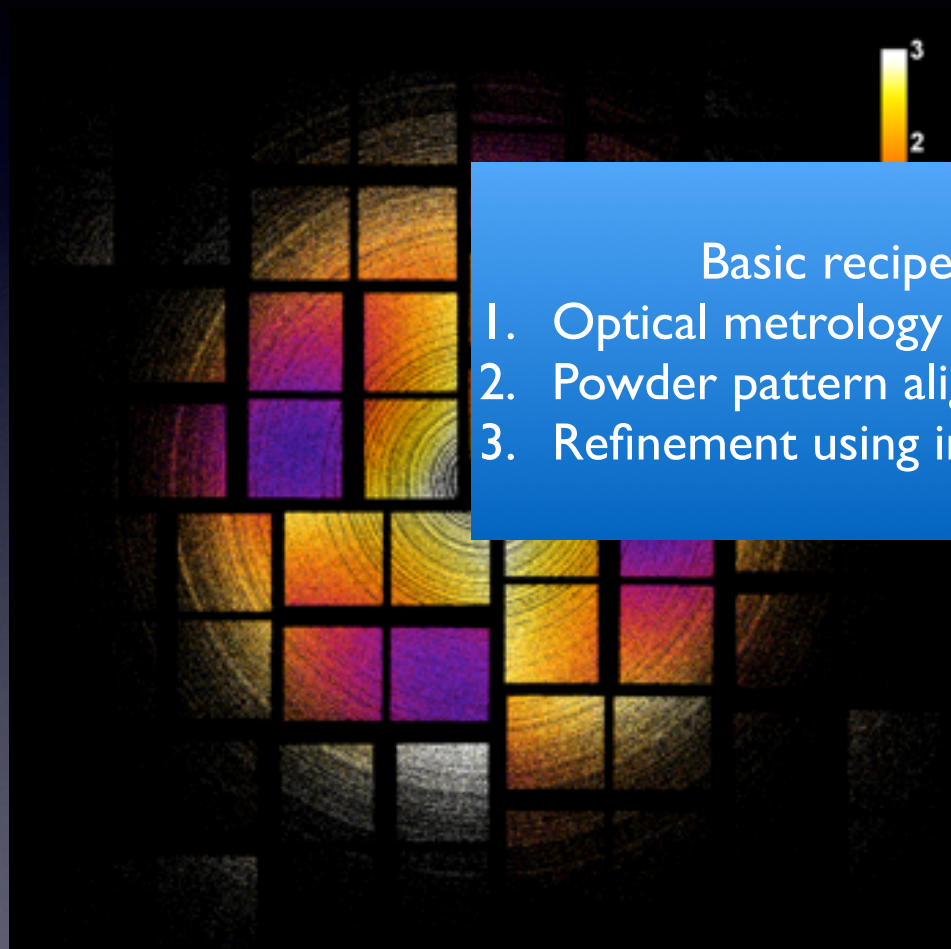


Slow scan error

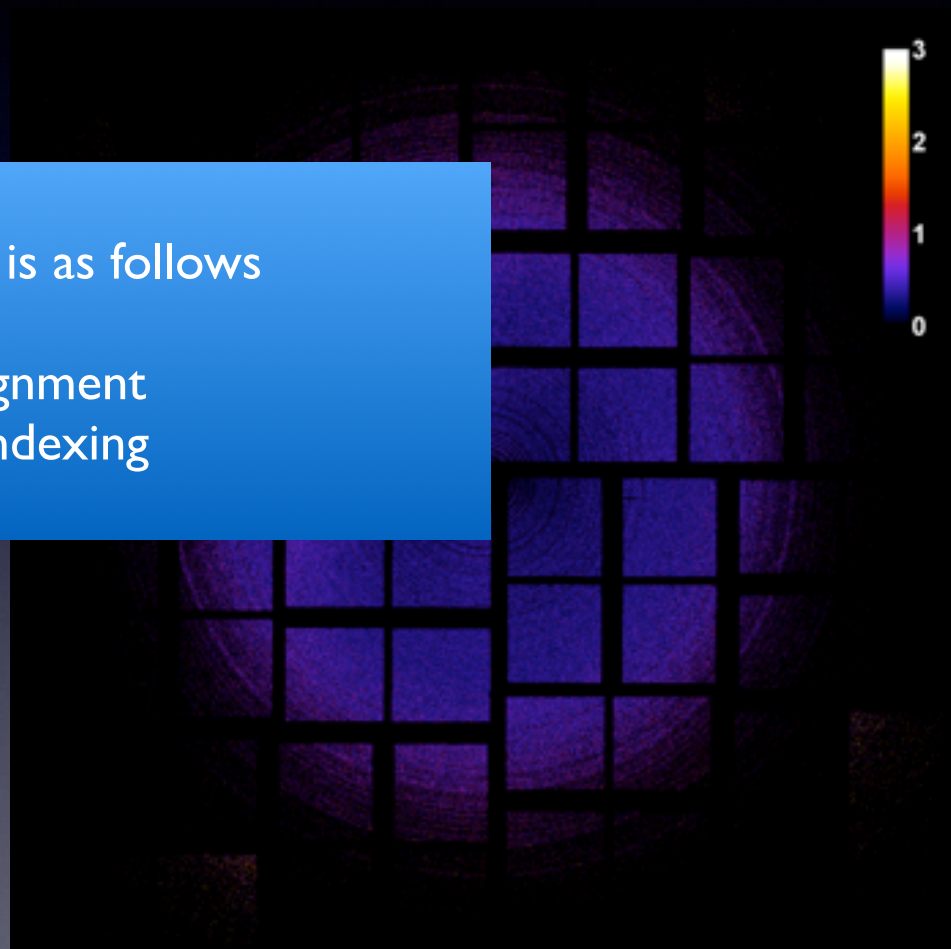


# Detector geometry is accurately determined by comparing observed and predicted peak locations

Error before refinement



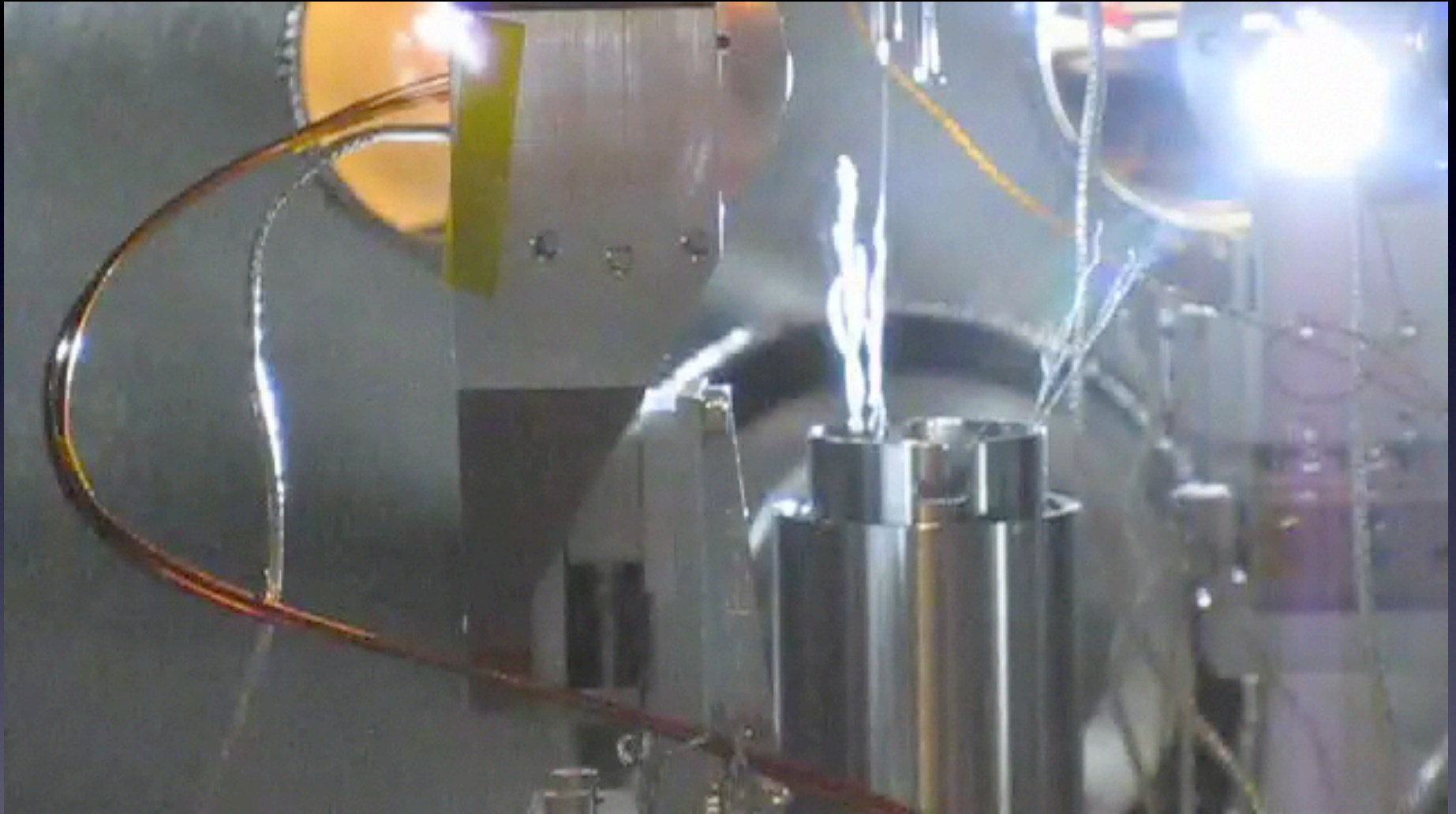
Error after refinement

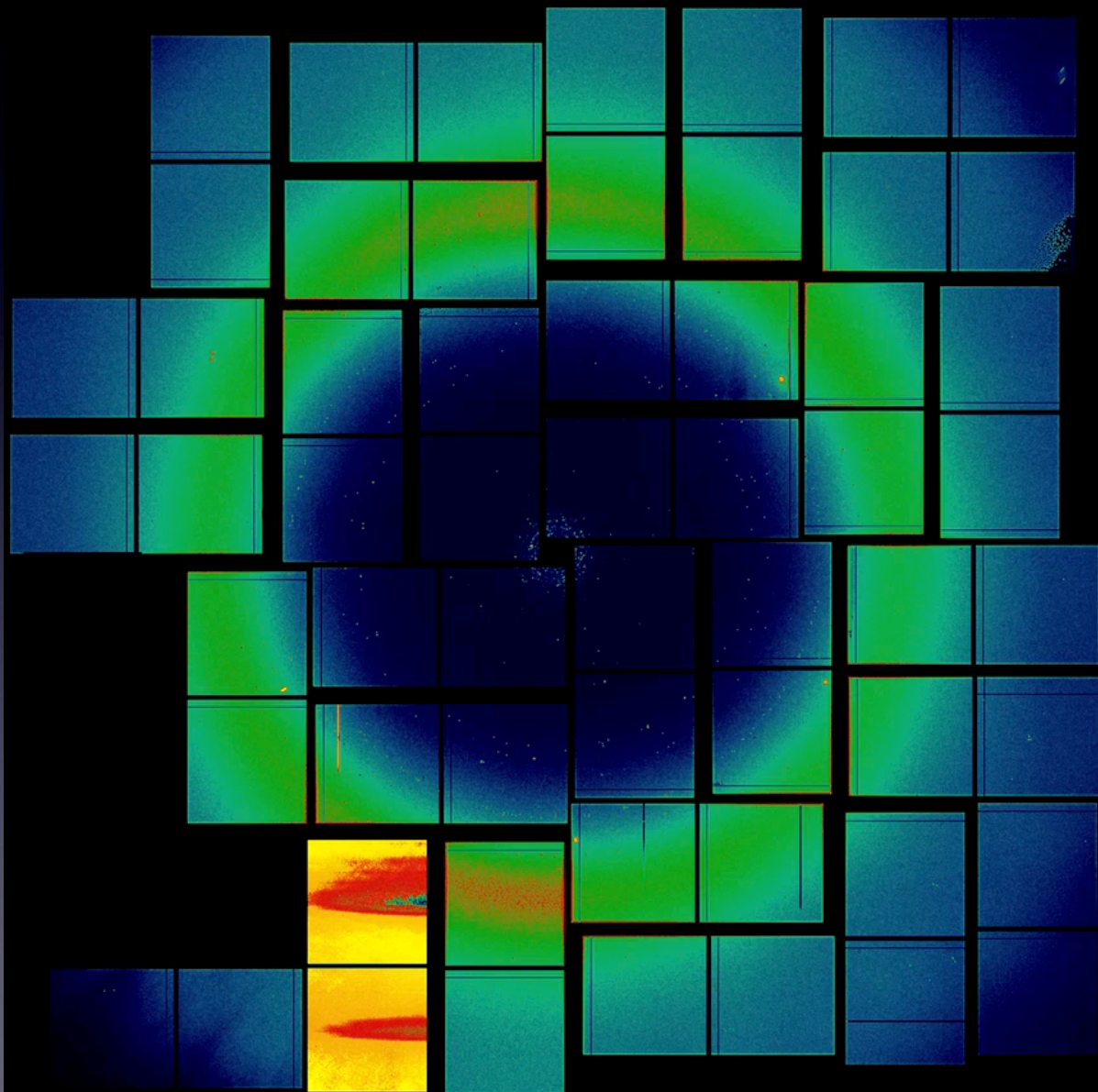


Basic recipe is as follows

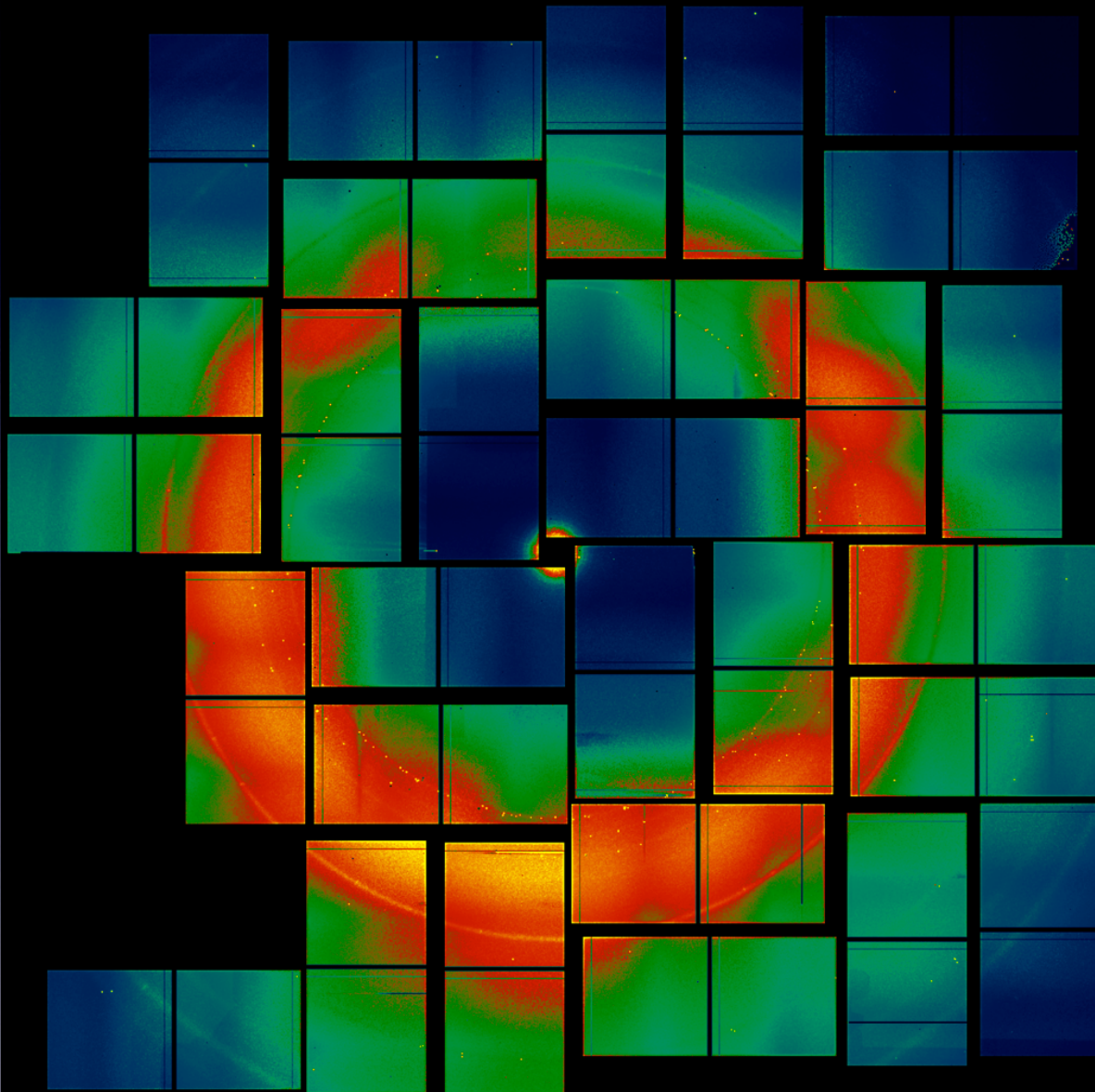
1. Optical metrology
2. Powder pattern alignment
3. Refinement using indexing

Detectors are not disposable, but unfortunately they are easily damaged during the course of an experiment





# Sum of all frames is dominated by water ring background



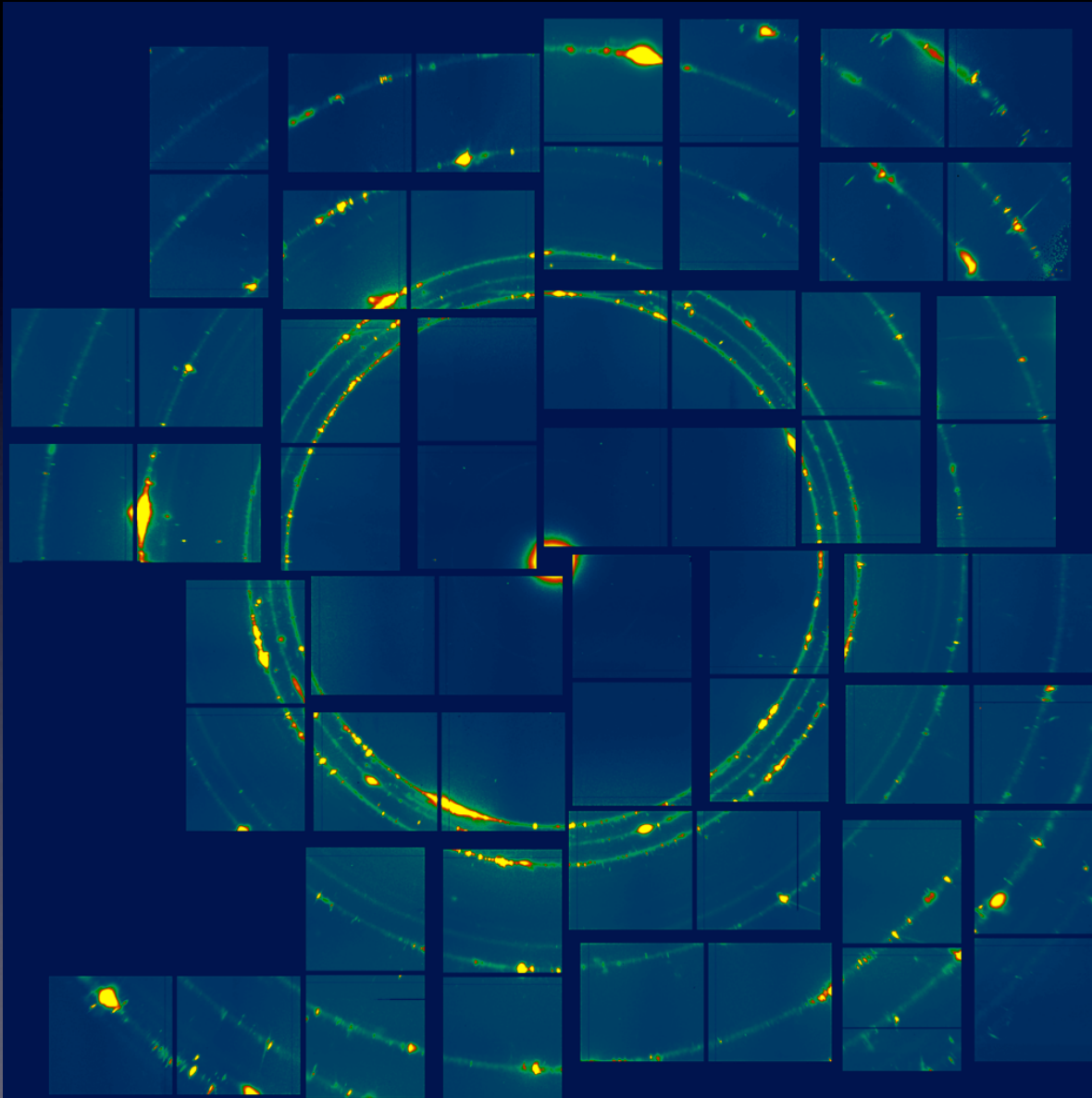
LCLS pulses: 2,292,468

Acquisition time: 5 hr 18 min

Photon energy: 9.4 keV

Up to  $5.2 \times 10^8$  ADU/pixel

# Ice gives rise to strong diffraction peaks on the detector



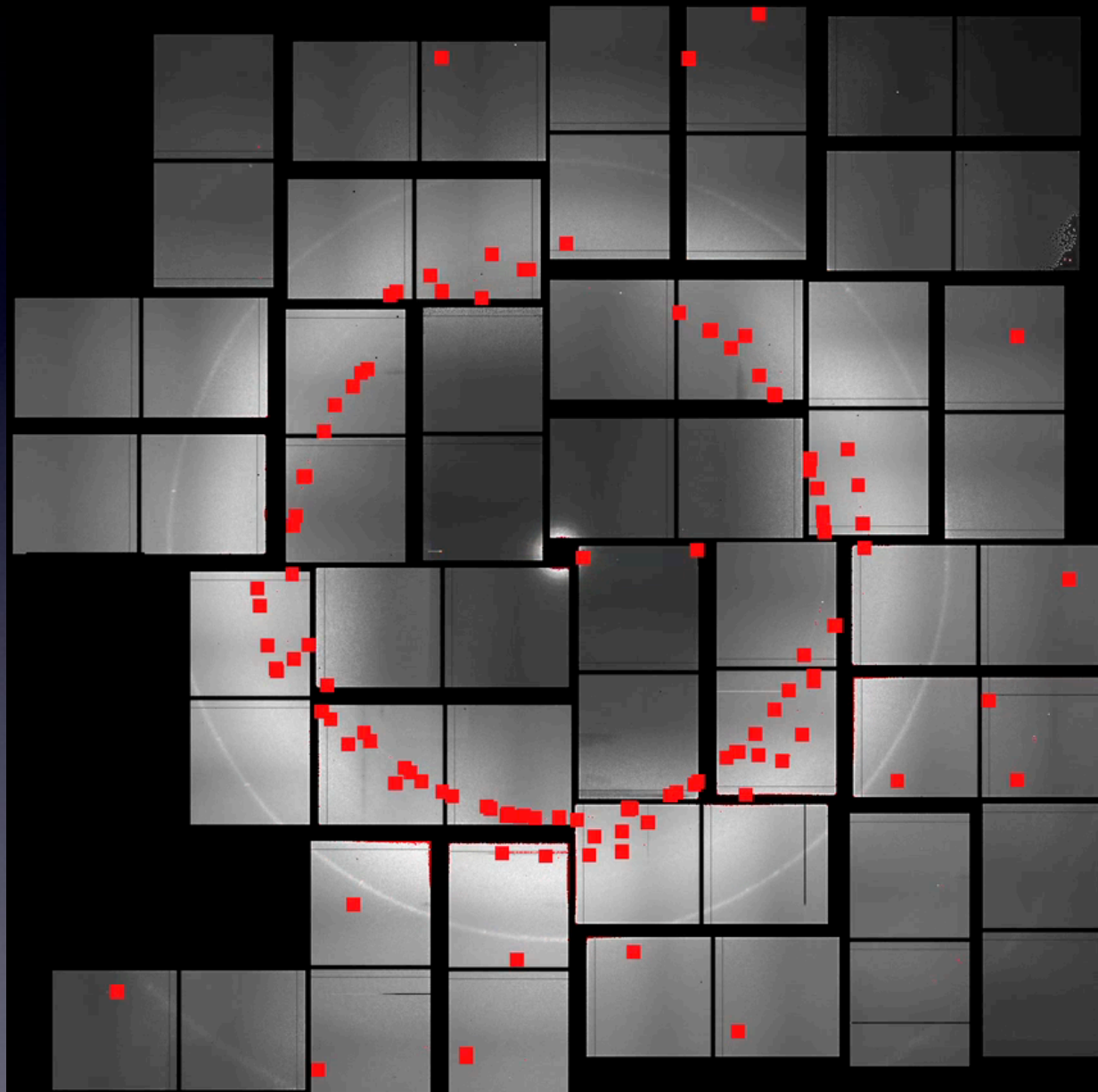
LCLS pulses: 4,293

Acquisition time: 35 seconds

Photon energy: 9.4 keV

Up to  $4.1 \times 10^7$  ADU/pixel

Dead pixels are identified by Cheetah as they accumulate during the course of the experiment





# Reduced data is output in facility independent HDF5 format

The image displays two overlapping windows. The background window is the 'HDF5 Browser' application, showing a file tree for the path `/reg/d/ffb/cxi/temp/cheetah/hdf5/r0003-1ys`. The tree structure includes:

- LCLS
  - data
    - data
    - pixelmask0
    - radialAverage0
    - radialAverageCounter0
    - rawdata
      - rawdata0
  - processing
    - cheetah
      - peakinfo
        - peakinfo-assembled
        - peakinfo-raw
      - hitfinder

The foreground window is a web browser displaying the CXIDB website. The page title is 'CXIDB - Coherent X-ray Imaging Data Bank'. The navigation menu includes: Home, Mission, CXI File Format, Browse Data, Resources, Sponsors, and Contact Us. The main content area features the heading 'The CXI File Format' and a paragraph: 'In this page you can find a complete description of the CXI file format along with many CXI examples files, and code to make and read those files.' Below this, there are two columns of text:

- CXI version 1.3 released**  
Posted by [Filipe Maia](#) on April 20, 2012  
The biggest change is the introduction of the concept of scans to accommodate datasets where one experimental parameter is continuously changed such as wavelength or sample rotation. A new axes attribute was introduced, related with scans. The ptychography example was updated, and now makes use of scans. Several small corrections and clarifications. As usual the document can be found on [github](#).
- Missing features**  
If you feel like there is an important feature (or class) that we are currently missing in CXI files then please send your ideas to [cxi@cxidb.org](mailto:cxi@cxidb.org).

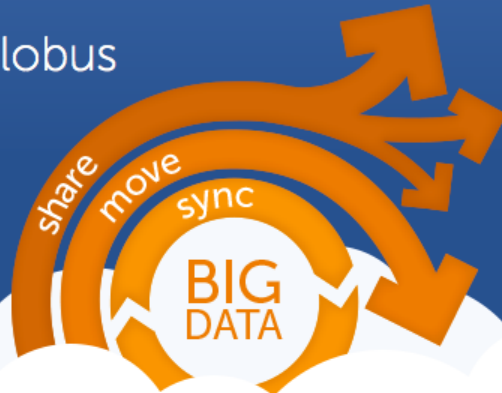
At the bottom of the page, there are two more links: 'CXI version 1.2 released' and 'Example CXI files'.

# Globus online is an excellent tool for transferring terabyte datasets to your home institution

<https://www.globus.org>



[Products](#) ▾ [News](#) ▾ [About](#) ▾ [Support](#) ▾ [Log In](#) [Sign Up](#)



Your research data  
where you need it.

**59,835,963,142** MB  
TRANSFERRED

## Researchers

Focus on your research, not IT problems. We make it easy to move, manage, and share big data.

[LEARN MORE](#) >

[READ CASE STUDIES](#) >



## Resource Providers

Globus gives you more control over your data infrastructure, while providing excellent ease-of-use for your researchers.

[LEARN MORE](#) >

[GLOBUS PROVIDER PLANS](#) >



## How It Works

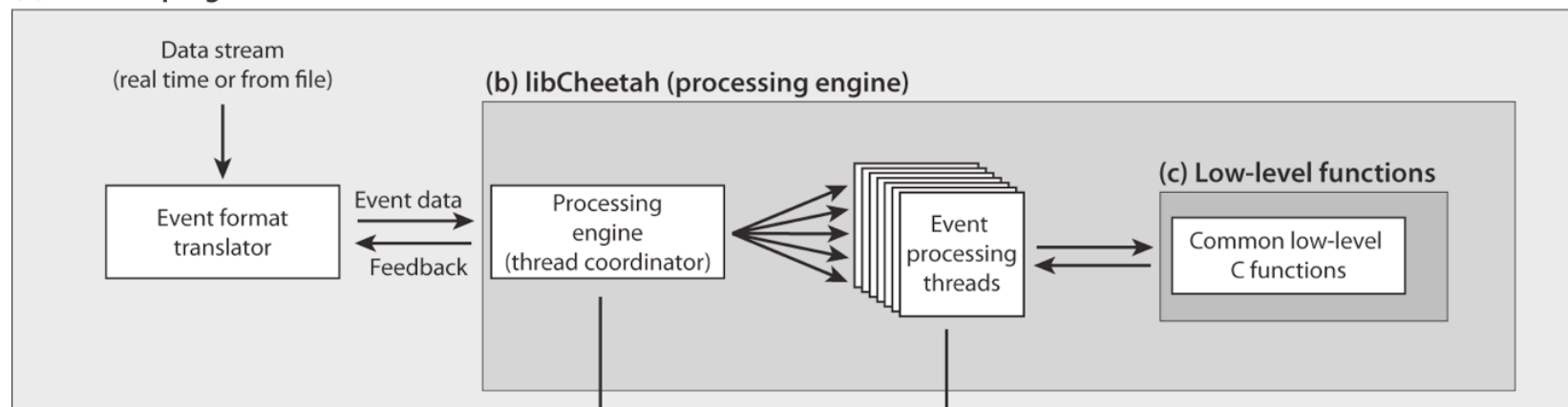
Globus' tools and services help connect people and HPC resources, so that no researcher is an island.

[LEARN MORE](#) >

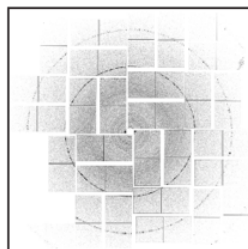


# Cheetah can easily be deployed at other facilities

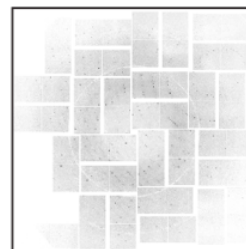
(a) Cheetah program



Processing summaries:  
• Hit lists, run statistics  
• Virtual powder patterns  
• Sorted radial stacks  
• etc.



Processed event data:  
• Background corrected data  
• Bragg peak locations  
• On-the-fly indexing  
• etc.



Has been used at:  
SACLA (.h5)  
ESRF (.edf)  
Petra III (.cbf)

- Output format is always the same
- CrystFEL does not have to change, and is free from facility dependencies
- The only change is in the file reader

# Cheetah is modular by design and open for code reuse

```
/*
 * Function prototypes
 */
void *worker(void *);

// detectorCorrection.cpp
void subtractDarkcal(cEventData*, cGlobal*);
void applyGainCorrection(cEventData*, cGlobal*);
void applyBadPixelMask(cEventData*, cGlobal*);
void cspadModuleSubtract(cEventData*, cGlobal*);
void cspadModuleSubtract2(cEventData*, cGlobal*);
void cspadModuleSubtract(cEventData*, cGlobal*, int);
void cspadSubtractUnbondedPixels(cEventData*, cGlobal*);
void cspadSubtractBehindWires(cEventData*, cGlobal*);
void calculateHotPixelMask(cGlobal*);
void identifyHotPixels(cEventData*, cGlobal*);
void applyHotPixelMask(cEventData*, cGlobal*);
void calculateHaloPixelMask(cGlobal*);
void updateHaloBuffer(cEventData*, cGlobal*, int);

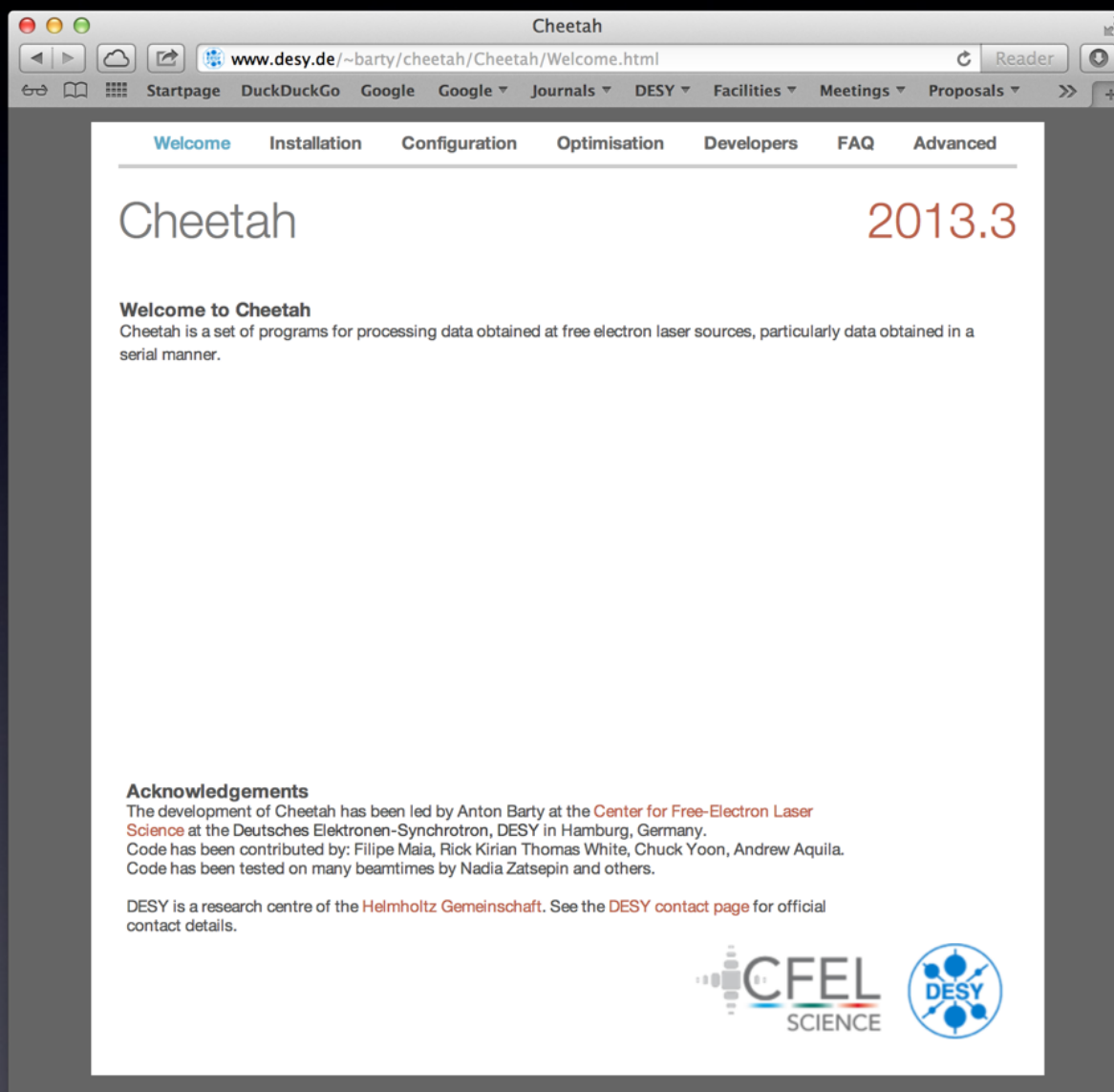
void subtractDarkcal(float*, float*, long);
void applyGainCorrection(float*, float*, long);
void applyBadPixelMask(float*, uint16_t*, long);
void cspadModuleSubtract(float*, uint16_t*, float, long, long, long, long);
void cspadSubtractUnbondedPixels(float*, uint16_t*, long, long, long, long);
void cspadSubtractBehindWires(float*, uint16_t*, float, long, long, long, long);
long calculateHotPixelMask(uint16_t*, int16_t*, long, long, long);
long calculateHaloPixelMask(uint16_t*, float*, float, long, long);
```

```
// assemble2DImage.cpp
void assemble2DImage(cEventData*, cGlobal*);
void assemble2Dmask(cEventData*, cGlobal*);
void assemble2DImage(int16_t*, float*, float*, float*, long, long, long, int);
void assemble2Dmask(uint16_t*, uint16_t*, float*, float*, long, long, long, int);
void downsample(cEventData*, cGlobal*);
void downsampleImage(int16_t*, int16_t*, long, long, long, long);
void downsampleMask(uint16_t*, uint16_t*, long, long, long, long);
```

Resources can be found on the web

NSF BioXFEL STC's guide for SFX data analysis at LCLS:  
[https://www.bioxfel.org/resources/LCLSdata\\_overview](https://www.bioxfel.org/resources/LCLSdata_overview)  
"LCLS serial femtosecond crystallography data analysis:  
everything you need to know"

<http://www.desy.de/~barty/cheetah/>



# Cheetah is pre-installed at SLAC

/reg/g/cfel/cheetah

```
psexport@slac.stanford.edu
http://www.slac.stanford.edu/comp/unix/public-machines.html
=====

Please log on to psananeh or psanafeh for code buildings, and testings.

Please use LSF to submit any analysis job.

We appreciate your cooperation to keep psexport up and running. Thank you!

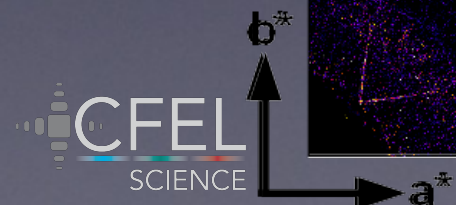
+++++

[psexport02;barty]"> source /reg/g/cfel/cheetah/setup.csh
[psexport02;barty]"> cheetah-gui
IDL Version 8.0 (linux x86_64 m64), (c) 2010, ITT Visual Information Solutions

Working directory: /reg/d/ffb/cxi/temp/cheetah/gui-demo
XTC directory: /reg/d/ffb/cxi/temp/xtc/
HDF5 directory: /reg/d/ffb/cxi/temp/cheetah/hdf5-demo/
HDF5 run filter: r*
Process script: ../process/process
Geometry file: ../calib/geometry/cspad-front-12feb2013-naz.h5
Default cheetah.ini: lys.ini
Refreshing table
█
```

- > source /reg/g/cfel/cheetah/setup.csh
- > cheetah-gui

The end result is an accurate set of reflection intensities for structure determination



Resolution = 0.18 nm

# A very good summary of cctbx.xfel vs Cheetah + CrystFEL

## Protein crystal structure obtained at 2.9 Å resolution from injecting bacterial cells into an X-ray free-electron laser beam

Michael R. Sawaya<sup>a,b,1</sup>, Duilio Cascio<sup>a,b,1</sup>, Mari Gingery<sup>a,b,1</sup>, Jose Rodriguez<sup>a,b</sup>, Lukasz Goldschmidt<sup>a,b</sup>, Jacques-Philippe Colletier<sup>c,d,e</sup>, Marc M. Messerschmidt<sup>f,2</sup>, Sébastien Boutet<sup>f</sup>, Jason E. Koglin<sup>f</sup>, Garth J. Williams<sup>f</sup>, Aaron S. Brewster<sup>g</sup>, Karol Nass<sup>h</sup>, Johan Hattne<sup>g</sup>, Sabine Botha<sup>h</sup>, R. Bruce Doak<sup>h,i</sup>, Robert L. Shoeman<sup>h</sup>, Daniel P. DePonte<sup>f</sup>, Hyun-Woo Park<sup>i,3</sup>, Brian A. Federici<sup>j,k</sup>, Nicholas K. Sauter<sup>g</sup>, Ilme Schlichting<sup>h</sup>, and David S. Eisenberg<sup>a,b,l,4</sup>

<sup>a</sup>UCLA–DOE Institute for Genomics and Proteomics, <sup>b</sup>Department of Biological Chemistry, and <sup>l</sup>Howard Hughes Medical Institute, University of California, Los Angeles, CA 90095-1570; <sup>c</sup>Université Grenoble Alpes, <sup>d</sup>Centre National de la Recherche Scientifique, and <sup>e</sup>Commissariat à l’Energie Atomique, Institut de Biologie Structurale, F-38044 Grenoble, France; <sup>f</sup>Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, CA 94025; <sup>g</sup>Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; <sup>h</sup>Max Planck Institute for Medical Research, 69120 Heidelberg, Germany; <sup>i</sup>Department of Physics, Arizona State University, Tempe, AZ 85287; and <sup>j</sup>Department of Entomology and <sup>k</sup>Graduate Program in Cell, Molecular and Developmental Biology, University of California, Riverside, CA 92521

Contributed by David S. Eisenberg, July 23, 2014 (sent for review April 22, 2014)

It has long been known that toxins produced by *Bacillus thuringiensis* (3, 9). The crystals for this study were *not* grown in artificial crys-

has published results on nine different macromolecular systems since its inception in 2009 (Table 1). One system in particular, cathepsin B, marks an advancement toward in vivo crystallography

<sup>1</sup>To whom correspondence should be addressed. Email: david@mbi.ucla.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1413456111/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1413456111/-DCSupplemental).

[www.pnas.org/cgi/doi/10.1073/pnas.1413456111](http://www.pnas.org/cgi/doi/10.1073/pnas.1413456111)

PNAS Early Edition | 1 of 6



# The LCLS data retention policy retains all data for 3 / 6 months: Users will want to take their data home

## Storage Classes

July 2015

Space	Size	Backup	Lifetime	Storage class	Comment
xtc	Unlimited	Tape archive	6 months	Short-term	Raw data
usr	Unlimited	Tape archive	6 months	Short-term	Raw data from users' DAQ systems
hdf5	Unlimited	Tape archive	6 months	Short-term	Data translated to HDF5
scratch	Unlimited	None	3 months	Short-term	Temporary data
xtc/hdf5	10TB	n/a	2 years	Medium-term	Selected XTC and HDF5 runs
ftc	10TB	None	2 years	Medium-term	Filtered, translated, compressed
res	1TB	Tape	2 years	Medium-term	Analysis results
User home	20GB	Disk + tape	Indefinite		User code
Tape archive	Unlimited	Two copies	10 years	Long-term	Raw data

Refining hit finding for a whole 5x12hr beamtime takes time.

Move your scripts and analysis to **/res** within **3** months

Move your cheetah/hdf5 output to **/ftc** within **6** months

# Worldwide XFEL capacity is growing



- Wide geographical distribution (Asia, Europe, USA)
- In hard x-ray region, by 2025 could see increase from 1 undulator today serving 4 stations serially to ~ 30-50 serving multiple stations simultaneously (including *serial* serial crystallography)

# Differences between 3<sup>rd</sup> generation synchrotron and XFEL

	Synchrotron	XFEL
Photon flux	$10^{11} - 10^{13}$ / second	$10^{12} - 10^{13}$ / pulse (avg)
Pulse duration	Down to 100 ps, generally longer	10 – 200 fs
X-ray wavelengths	0.5 – 2.0Å (5-15 keV)	1.5 – 6.2 Å (2-9 keV) at LCLS
Transverse (spatial) coherence	10's of $\mu\text{m}$	Fully spatially coherent
Bandwidth	Tunable: pink 2-3% bw, 0.001% bw with Si 311 mono	SASE: 0.1 % bw Seeded beam: $10^{-4}$
X-ray focus	$\sim 25$ nm	$\sim 100$ nm