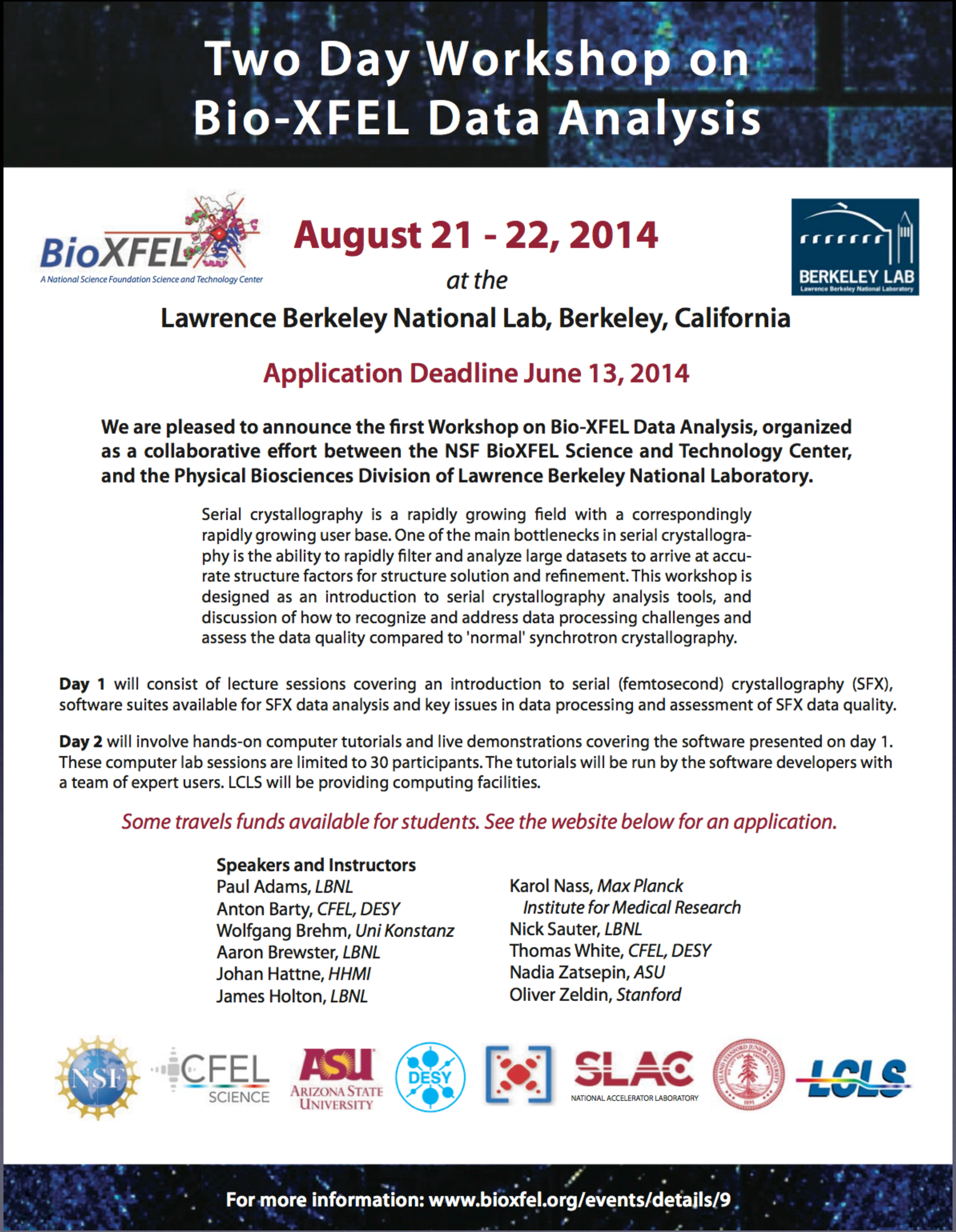


Two Day Workshop on Bio-XFEL Data Analysis

BioXFEL
A National Science Foundation Science and Technology Center

August 21 - 22, 2014

at the

Lawrence Berkeley National Lab, Berkeley, California

BERKELEY LAB
Lawrence Berkeley National Laboratory

Application Deadline June 13, 2014

We are pleased to announce the first Workshop on Bio-XFEL Data Analysis, organized as a collaborative effort between the NSF BioXFEL Science and Technology Center, and the Physical Biosciences Division of Lawrence Berkeley National Laboratory.

Serial crystallography is a rapidly growing field with a correspondingly rapidly growing user base. One of the main bottlenecks in serial crystallography is the ability to rapidly filter and analyze large datasets to arrive at accurate structure factors for structure solution and refinement. This workshop is designed as an introduction to serial crystallography analysis tools, and discussion of how to recognize and address data processing challenges and assess the data quality compared to 'normal' synchrotron crystallography.

Day 1 will consist of lecture sessions covering an introduction to serial (femtosecond) crystallography (SFX), software suites available for SFX data analysis and key issues in data processing and assessment of SFX data quality.

Day 2 will involve hands-on computer tutorials and live demonstrations covering the software presented on day 1. These computer lab sessions are limited to 30 participants. The tutorials will be run by the software developers with a team of expert users. LCLS will be providing computing facilities.

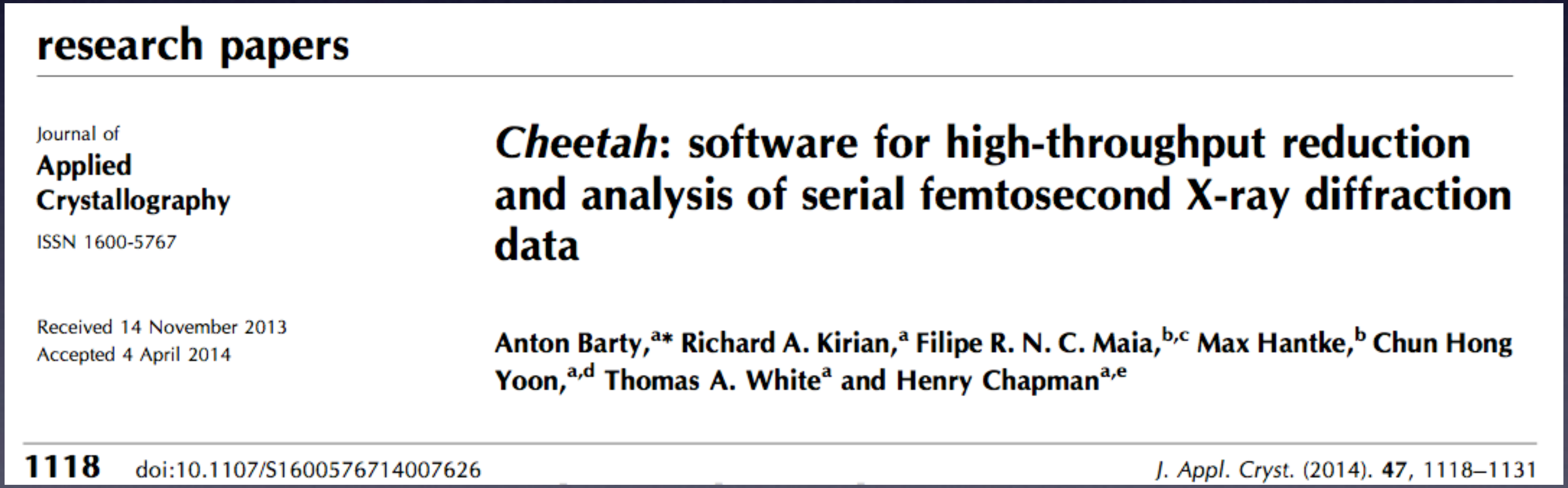
Some travels funds available for students. See the website below for an application.

Speakers and Instructors

Paul Adams, *LBNL*
Anton Barty, *CFEL, DESY*
Wolfgang Brehm, *Uni Konstanz*
Aaron Brewster, *LBNL*
Johan Hattne, *HHMI*
James Holton, *LBNL*
Karol Nass, *Max Planck Institute for Medical Research*
Nick Sauter, *LBNL*
Thomas White, *CFEL, DESY*
Nadia Zatsepin, *ASU*
Oliver Zeldin, *Stanford*

For more information: www.bioxfel.org/events/details/9

Cheetah tutorial



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,[a]* Richard A. Kirian,[a] Filipe R. N. C. Maia,[b,c] Max Hantke,[b] Chun Hong Yoon,[a,d] Thomas A. White[a] and Henry Chapman[a,e]

1118 doi:10.1107/S1600576714007626 *J. Appl. Cryst.* (2014). **47**, 1118–1131

Not necessarily sexy, but a necessary evil

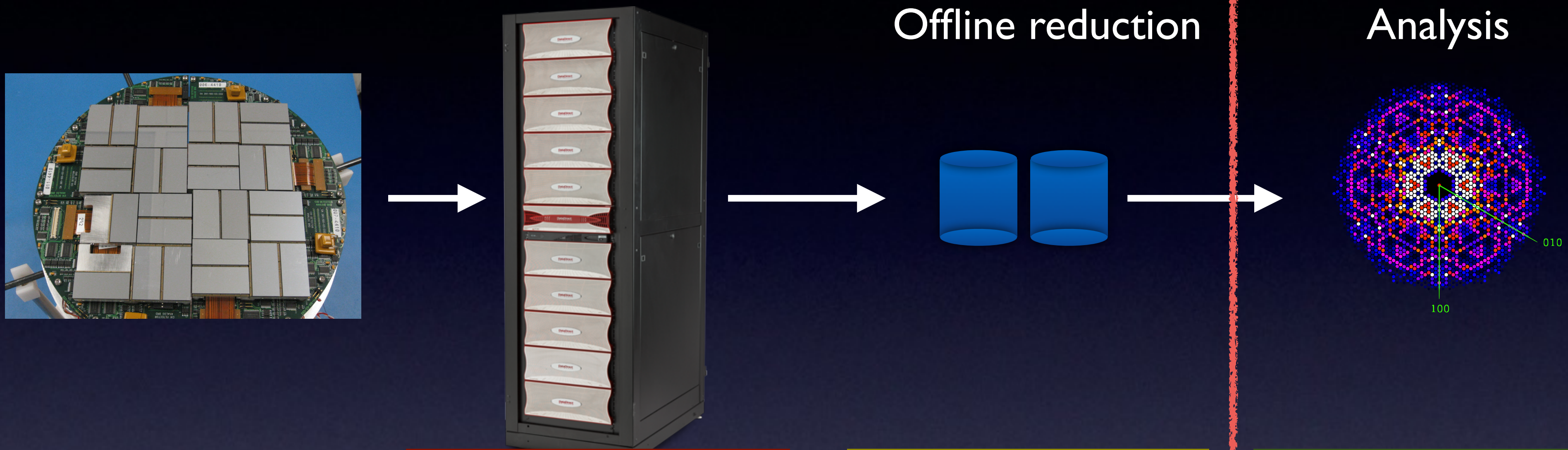
The Cheetah + CrystFEL life cycle

SLAC

Home/Laptop

Offline reduction

Analysis



> 4,000,000 frames

~ 400,000 hits

~ 300,000 indexed

Experiment

Online monitoring

Data reduction

Indexing/ integration

Structure

DAQ

Fast feedback

Cheetah

CrystFEL
CCP4 / Phenix

Please check the following work for you:

- Log on to pslogin.slac.stanford.edu
  ```
  > ssh pslogin.slac.stanford.edu
  ```

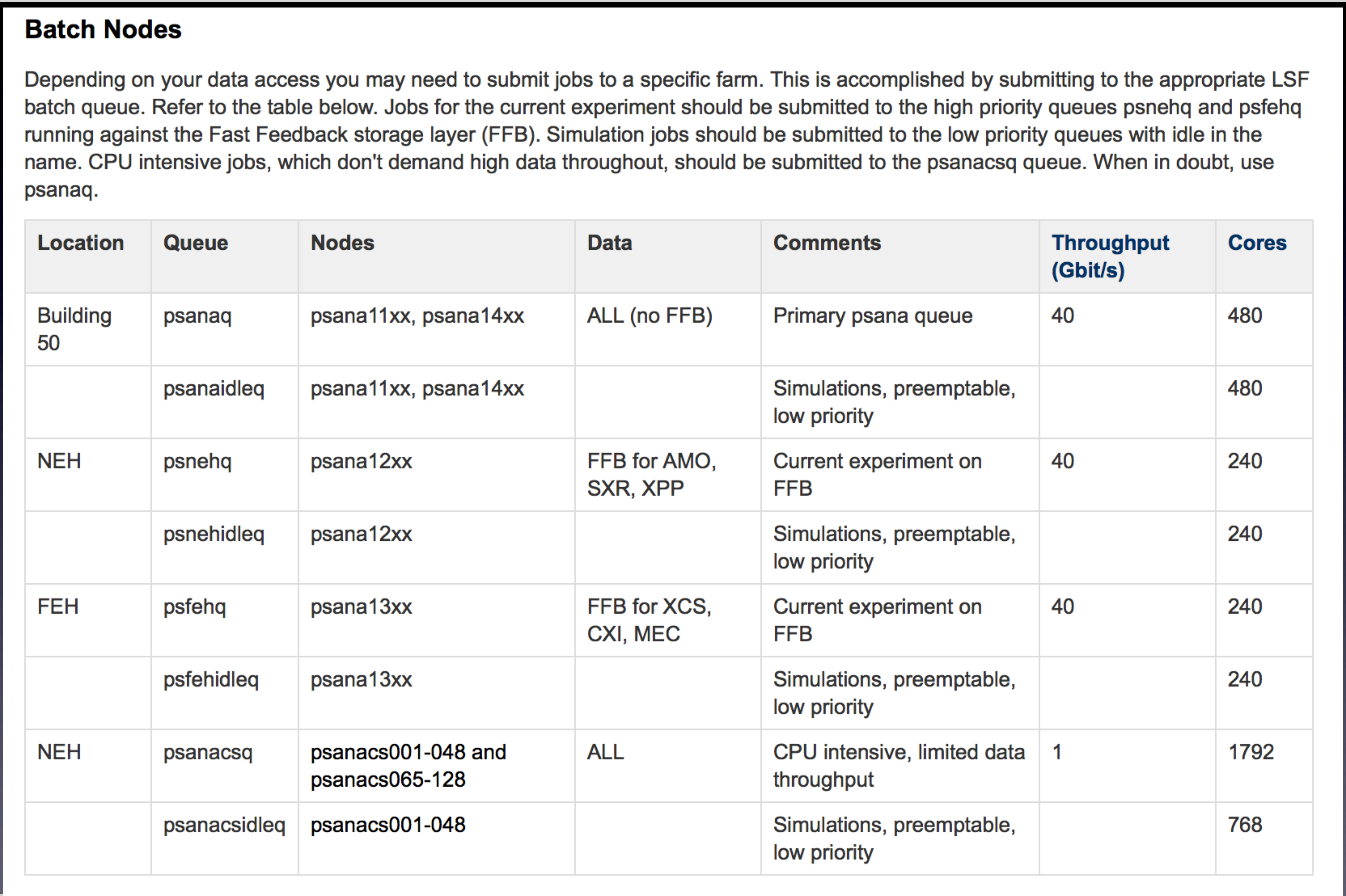
- Access the data
  ```
  > cd /reg/d/psdm/cxi/….
  ```

- Write access to scratch space
  ```
  > touch /reg/d/psdm/cxi/….
  ```

- Submit a job to the batch queue
  ```
  > bsub -q psanaq ls
  ```

- Access Nadia’s spreadsheet of tutorial data
  https://docs.google.com/spreadsheets/d/1rBN8BaW5ys1HpYWUoNsaHldBSbOnVD2_RZyCsSlm__l/edit#gid=1543105999

Batch nodes at LCLS



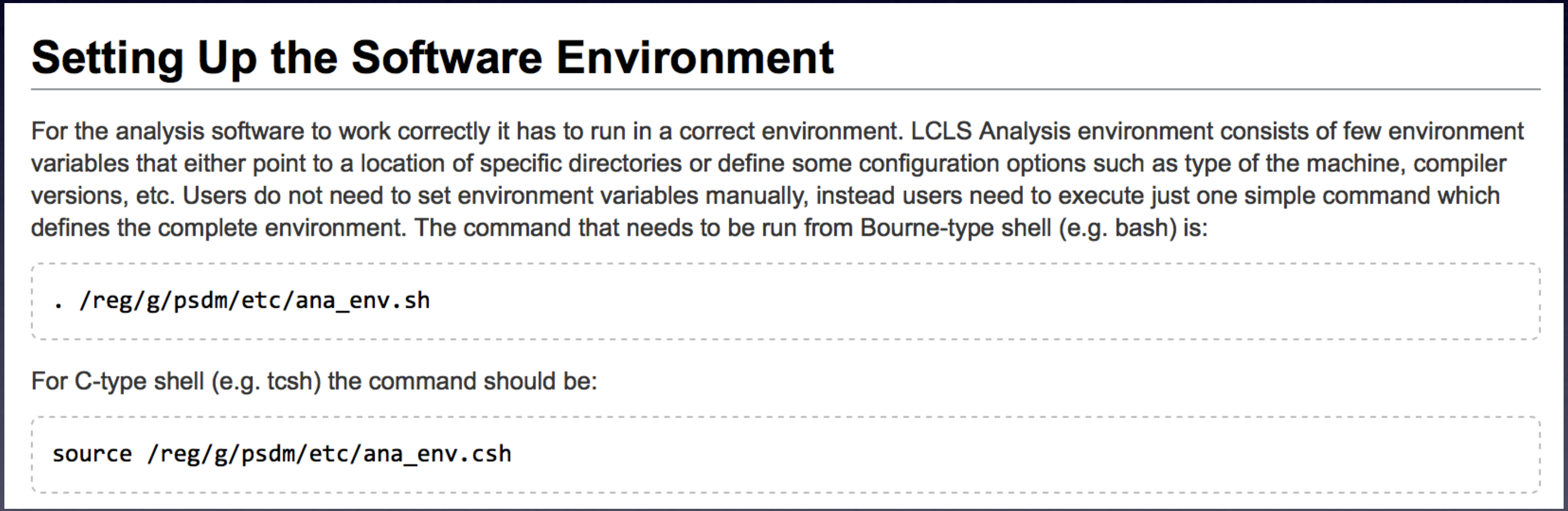
Batch Nodes

Depending on your data access you may need to submit jobs to a specific farm. This is accomplished by submitting to the appropriate LSF batch queue. Refer to the table below. Jobs for the current experiment should be submitted to the high priority queues psnehq and psfehq running against the Fast Feedback storage layer (FFB). Simulation jobs should be submitted to the low priority queues with idle in the name. CPU intensive jobs, which don't demand high data throughout, should be submitted to the psanacsq queue. When in doubt, use psanaq.

Location	Queue	Nodes	Data	Comments	Throughput (Gbit/s)	Cores
Building 50	psanaq	psana11xx, psana14xx	ALL (no FFB)	Primary psana queue	40	480
	psanaidleq	psana11xx, psana14xx		Simulations, preemptable, low priority		480
NEH	psnehq	psana12xx	FFB for AMO, SXR, XPP	Current experiment on FFB	40	240
	psnehidleq	psana12xx		Simulations, preemptable, low priority		240
FEH	psfehq	psana13xx	FFB for XCS, CXI, MEC	Current experiment on FFB	40	240
	psfehidleq	psana13xx		Simulations, preemptable, low priority		240
NEH	psanacsq	psanacs001-048 and psanacs065-128	ALL	CPU intensive, limited data throughput	1	1792
	psanacsidleq	psanacs001-048		Simulations, preemptable, low priority		768

It is necessary to do the following for psana to function

Cheetah is a psana module



Setting Up the Software Environment

For the analysis software to work correctly it has to run in a correct environment. LCLS Analysis environment consists of few environment variables that either point to a location of specific directories or define some configuration options such as type of the machine, compiler versions, etc. Users do not need to set environment variables manually, instead users need to execute just one simple command which defines the complete environment. The command that needs to be run from Bourne-type shell (e.g. bash) is:

```
. /reg/g/psdm/etc/ana_env.sh
```

For C-type shell (e.g. tcsh) the command should be:

```
source /reg/g/psdm/etc/ana_env.csh
```

https://confluence.slac.stanford.edu/display/PSDM/Analysis+Setup

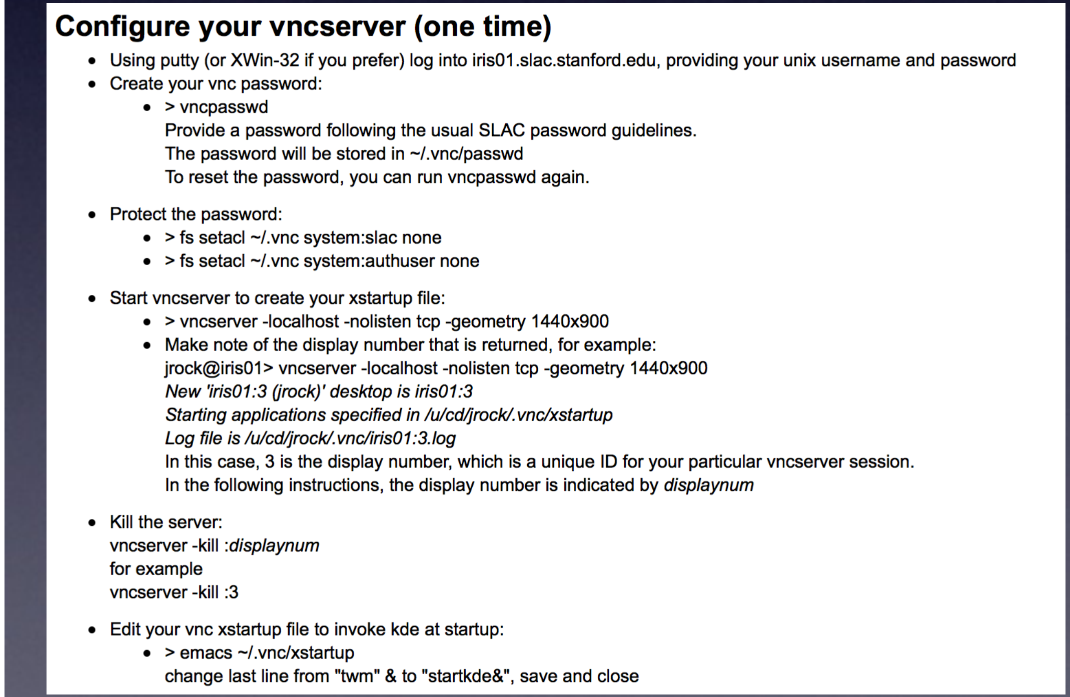
Connect to SLAC

Simple, but slow graphics:

```
> ssh -X <user>@pslogin.slac.stanford.edu
> ssh -X psana
```

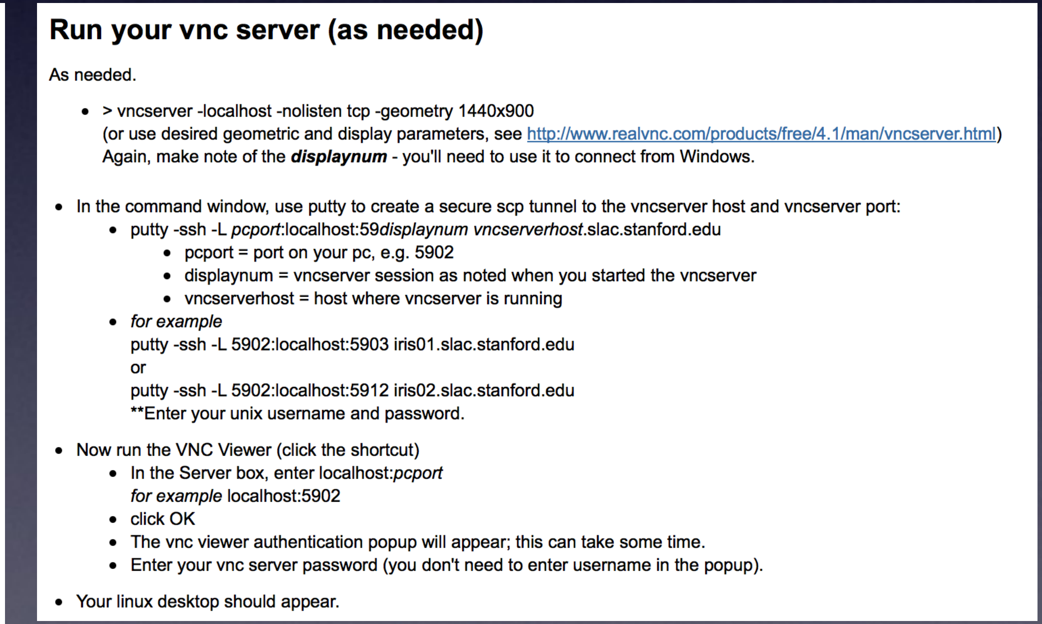
Faster graphics, but more complex:

```
> ssh -X <user>@pslogin.slac.stanford.edu
> vncserver -localhost -nolisten tcp -geometry 2200x1800
```



Configure your vncserver (one time)

- Using putty (or XWin-32 if you prefer) log into iris01.slac.stanford.edu, providing your unix username and password
- Create your vnc password:
  - > vncpasswd
    Provide a password following the usual SLAC password guidelines.
    The password will be stored in ~/.vnc/passwd
    To reset the password, you can run vncpasswd again.
- Protect the password:
  - > fs setacl ~/.vnc system:slac none
  - > fs setacl ~/.vnc system:authuser none
- Start vncserver to create your xstartup file:
  - > vncserver -localhost -nolisten tcp -geometry 1440x900
  - Make note of the display number that is returned, for example:
    jrock@iris01> vncserver -localhost -nolisten tcp -geometry 1440x900
    New 'iris01:3 (jrock)' desktop is iris01:3
    Starting applications specified in /u/cd/jrock/.vnc/xstartup
    Log file is /u/cd/jrock/.vnc/iris01:3.log
    In this case, 3 is the display number, which is a unique ID for your particular vncserver session.
    In the following instructions, the display number is indicated by *displaynum*
- Kill the server:
  vncserver -kill :*displaynum*
  for example
  vncserver -kill :3
- Edit your vnc xstartup file to invoke kde at startup:
  - > emacs ~/.vnc/xstartup
    change last line from "twm" & to "startkde&", save and close



Run your vnc server (as needed)

As needed.

- > vncserver -localhost -nolisten tcp -geometry 1440x900
  (or use desired geometric and display parameters, see http://www.realvnc.com/products/free/4.1/man/vncserver.html)
  Again, make note of the ***displaynum*** - you'll need to use it to connect from Windows.

- In the command window, use putty to create a secure scp tunnel to the vncserver host and vncserver port:
  - putty -ssh -L *pcport*:localhost:59*displaynum vncserverhost*.slac.stanford.edu
    - pcport = port on your pc, e.g. 5902
    - displaynum = vncserver session as noted when you started the vncserver
    - vncserverhost = host where vncserver is running
  - *for example*
    putty -ssh -L 5902:localhost:5903 iris01.slac.stanford.edu
    or
    putty -ssh -L 5902:localhost:5912 iris02.slac.stanford.edu
    **Enter your unix username and password.
- Now run the VNC Viewer (click the shortcut)
  - In the Server box, enter localhost:*pcport*
    for example localhost:5902
  - click OK
  - The vnc viewer authentication popup will appear; this can take some time.
  - Enter your vnc server password (you don't need to enter username in the popup).
- Your linux desktop should appear.

Download the prepared template

Karol’s Gd:lysozyme

```
> tar -xvf /reg/g/cfel/cheetah/XFEL-workshop-gdlys.tar
```

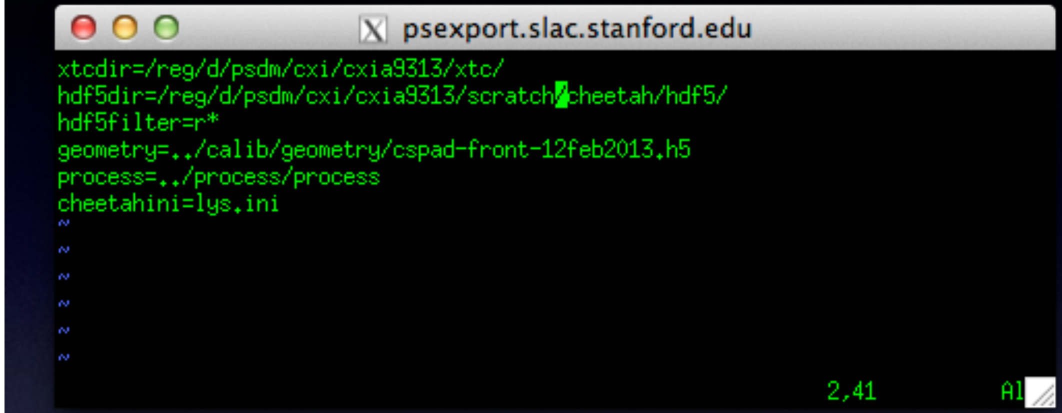
— or —

Boutet lysozyme

```
> tar -xvf /reg/g/cfel/cheetah/XFEL-workshop-boutet.tar
```

Two configuration files need to be edited

gui/crawler.config



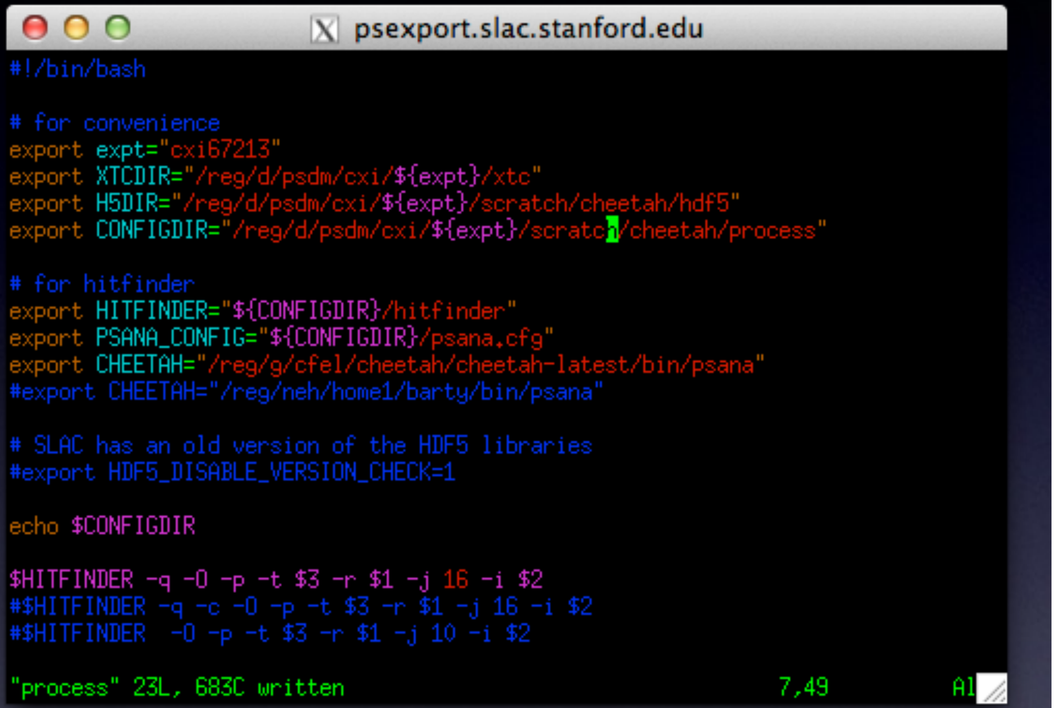
```
xtcdir=/reg/d/psdm/cxi/cxia9313/xtc/
hdf5dir=/reg/d/psdm/cxi/cxia9313/scratch/cheetah/hdf5/
hdf5filter=r*
geometry=../calib/geometry/cspad-front-12feb2013.h5
process=../process/process
cheetahini=lys.ini
~
~
~
~
~
~
                                                            2,41          All
```

Change the line:

hdf5dir=

(point to where tarball unpacked)

process/process



```
#!/bin/bash

# for convenience
export expt="cxi67213"
export XTCDIR="/reg/d/psdm/cxi/${expt}/xtc"
export H5DIR="/reg/d/psdm/cxi/${expt}/scratch/cheetah/hdf5"
export CONFIGDIR="/reg/d/psdm/cxi/${expt}/scratch/cheetah/process"

# for hitfinder
export HITFINDER="${CONFIGDIR}/hitfinder"
export PSANA_CONFIG="${CONFIGDIR}/psana.cfg"
export CHEETAH="/reg/g/cfel/cheetah/cheetah-latest/bin/psana"
#export CHEETAH="/reg/neh/home1/barty/bin/psana"

# SLAC has an old version of the HDF5 libraries
#export HDF5_DISABLE_VERSION_CHECK=1

echo $CONFIGDIR

$HITFINDER -q -O -p -t $3 -r $1 -j 16 -i $2
#$HITFINDER -q -c -O -p -t $3 -r $1 -j 16 -i $2
#$HITFINDER  -O -p -t $3 -r $1 -j 10 -i $2

"process" 23L, 683C written                                 7,49          All
```

Change the lines:

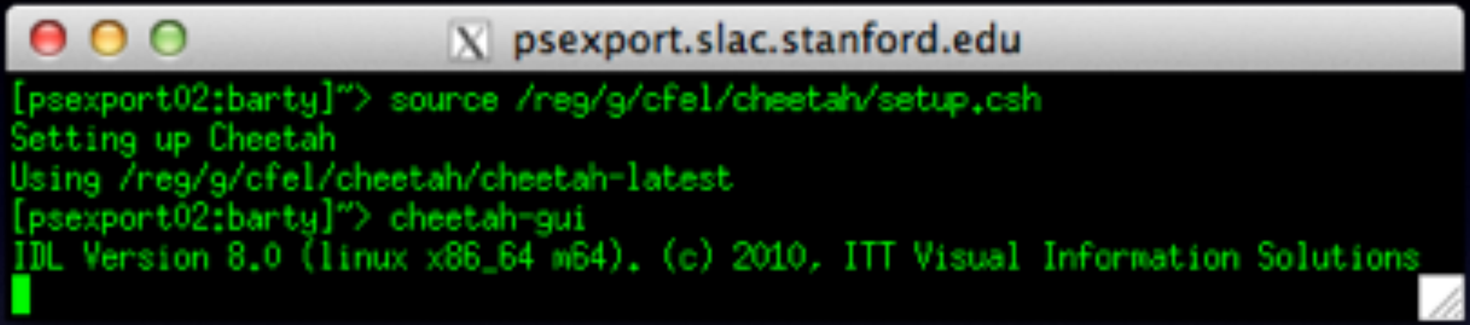
H5DIR=

CONFIGDIR=

(also point to where tarball unpacked)

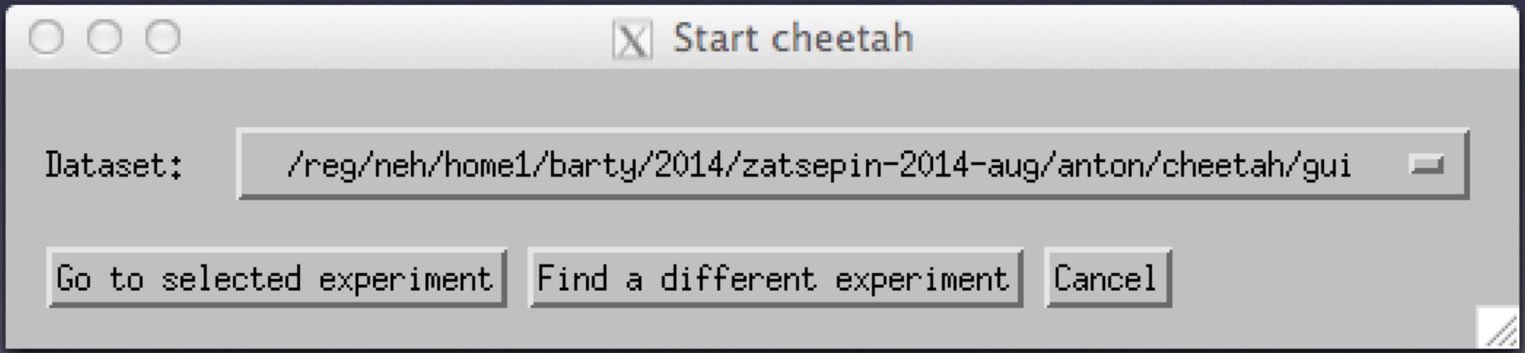
How to start the GUI

1. Reference the central installation



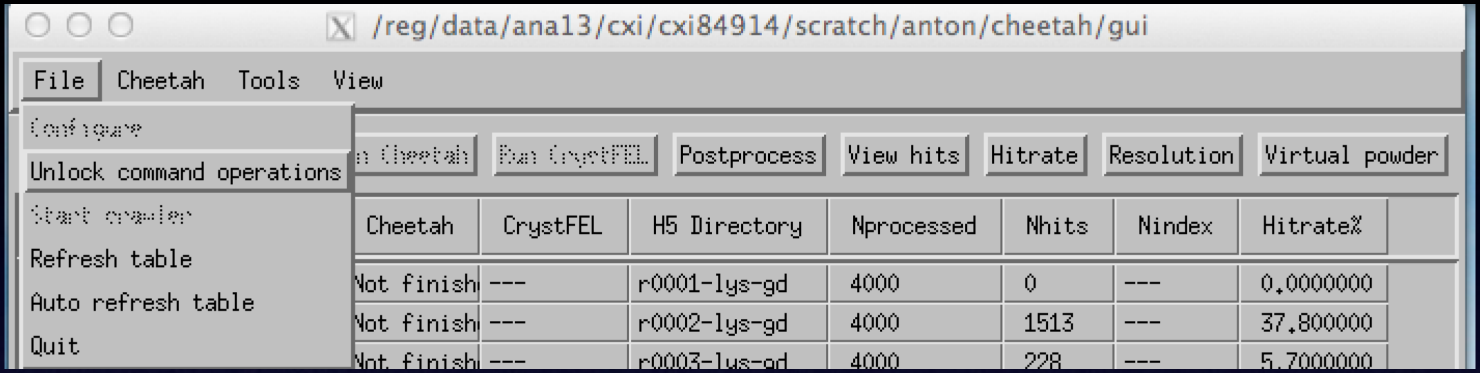
```
[psexport02:barty]~> source /reg/g/cfel/cheetah/setup.csh
Setting up Cheetah
Using /reg/g/cfel/cheetah/cheetah-latest
[psexport02:barty]~> cheetah-gui
IDL Version 8.0 (linux x86_64 m64). (c) 2010, ITT Visual Information Solutions
```

2. Select the crawler.config file that has just been edited



Dataset: /reg/neh/home1/barty/2014/zatsepin-2014-aug/anton/cheetah/gui

Start the crawler



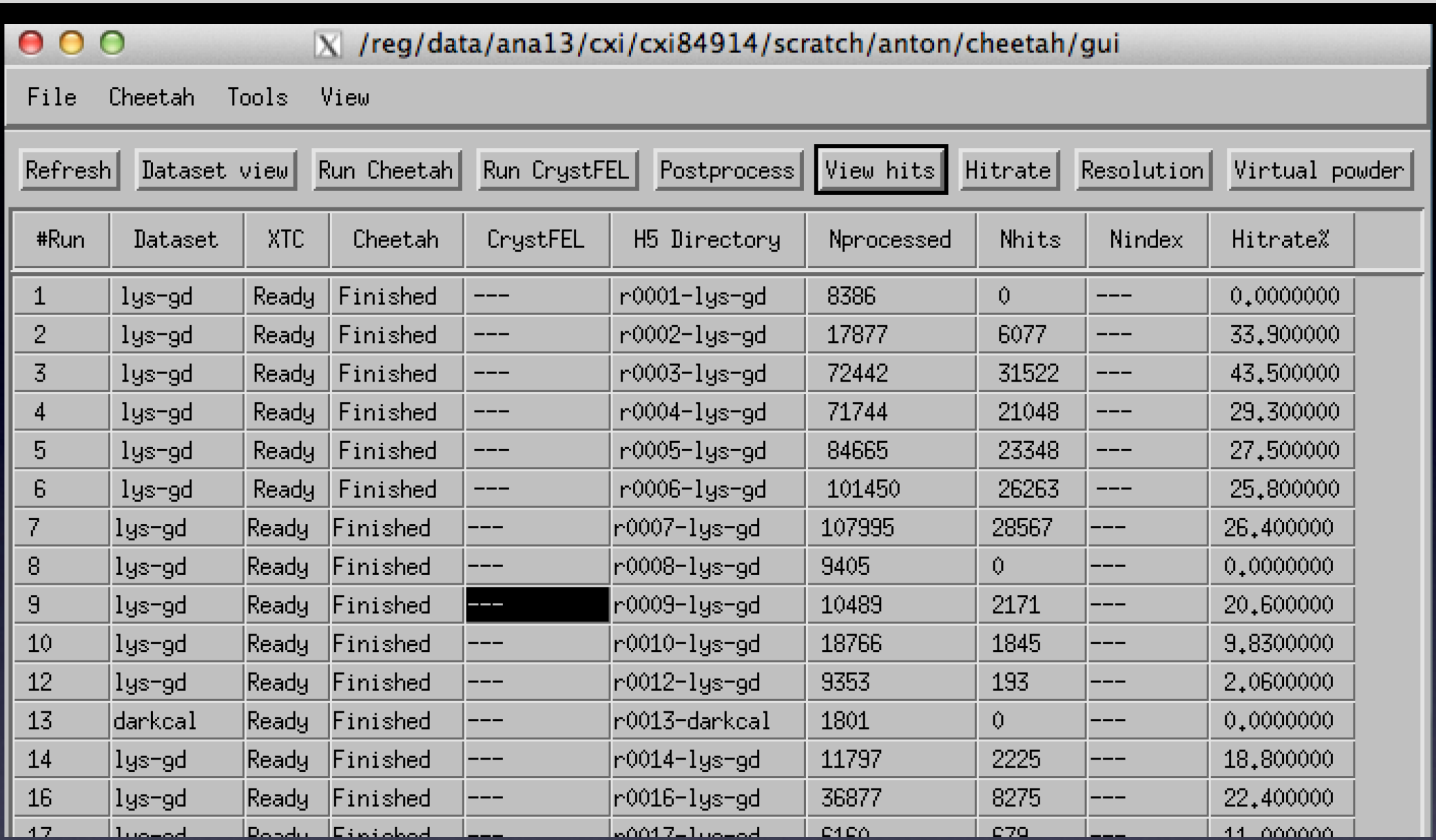
Cheetah	CrystFEL	H5 Directory	Nprocessed	Nhits	Nindex	Hitrate%
Not finish	---	r0001-lys-gd	4000	0	---	0.0000000
Not finish	---	r0002-lys-gd	4000	1513	---	37.800000
Not finish	---	r0003-lys-gd	4000	228	---	5.7000000

1. File / Unlock command operations
2. File / Start crawler

Tip: Many people can view progress of the analysis and view data at the same time. eg: both upstairs and downstairs at the same time.

Only one person should run the crawler (and ideally also start jobs) to avoid simultaneous modification of the database by two processes (hence the "Unlock command operations" step)

Select run 9 and "Run Cheetah"

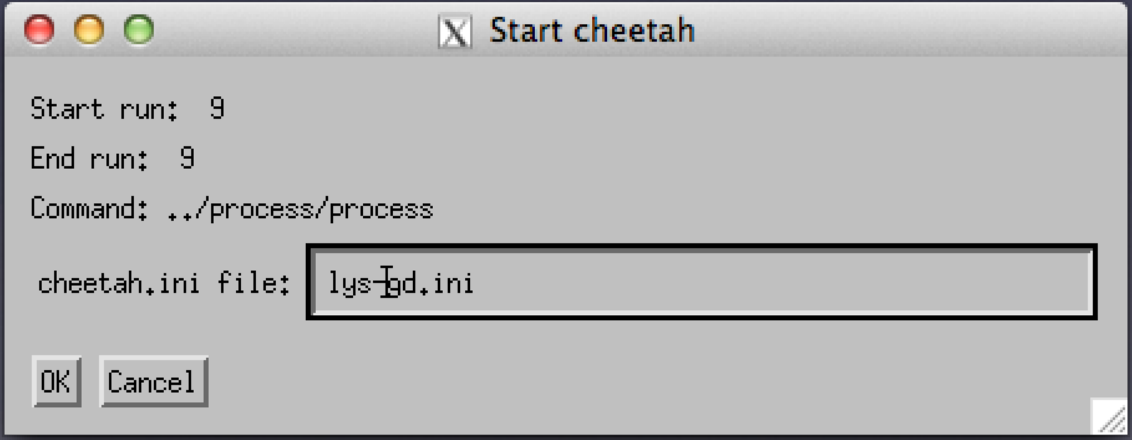


#Run	Dataset	XTC	Cheetah	CrystFEL	H5 Directory	Nprocessed	Nhits	Nindex	Hitrate%
1	lys-gd	Ready	Finished	---	r0001-lys-gd	8386	0	---	0.0000000
2	lys-gd	Ready	Finished	---	r0002-lys-gd	17877	6077	---	33.900000
3	lys-gd	Ready	Finished	---	r0003-lys-gd	72442	31522	---	43.500000
4	lys-gd	Ready	Finished	---	r0004-lys-gd	71744	21048	---	29.300000
5	lys-gd	Ready	Finished	---	r0005-lys-gd	84665	23348	---	27.500000
6	lys-gd	Ready	Finished	---	r0006-lys-gd	101450	26263	---	25.800000
7	lys-gd	Ready	Finished	---	r0007-lys-gd	107995	28567	---	26.400000
8	lys-gd	Ready	Finished	---	r0008-lys-gd	9405	0	---	0.0000000
9	lys-gd	Ready	Finished	---	r0009-lys-gd	10489	2171	---	20.600000
10	lys-gd	Ready	Finished	---	r0010-lys-gd	18766	1845	---	9.8300000
12	lys-gd	Ready	Finished	---	r0012-lys-gd	9353	193	---	2.0600000
13	darkcal	Ready	Finished	---	r0013-darkcal	1801	0	---	0.0000000
14	lys-gd	Ready	Finished	---	r0014-lys-gd	11797	2225	---	18.800000
16	lys-gd	Ready	Finished	---	r0016-lys-gd	36877	8275	---	22.400000
17	lys-gd	Ready	Finished	---	r0017-lys-gd	6160	679	---	11.000000

Hint:
Run 9 is not too long and has a reasonable hit rate

Pre-prepared for quick start

Start Cheetah using:
"lys-gd.ini"



Start run: 9

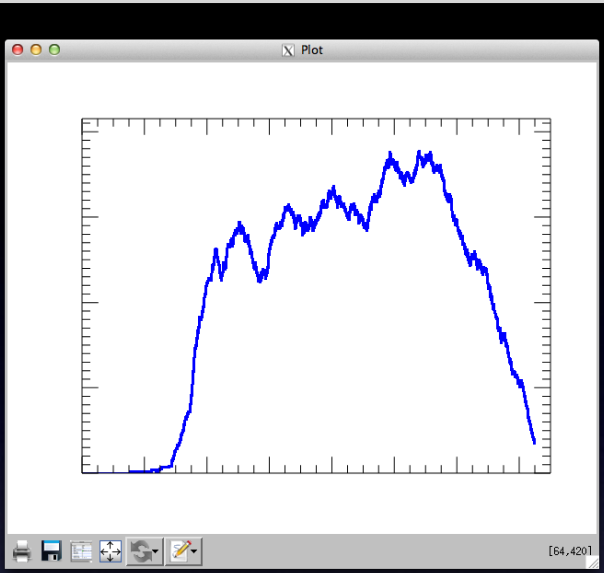
End run: 9

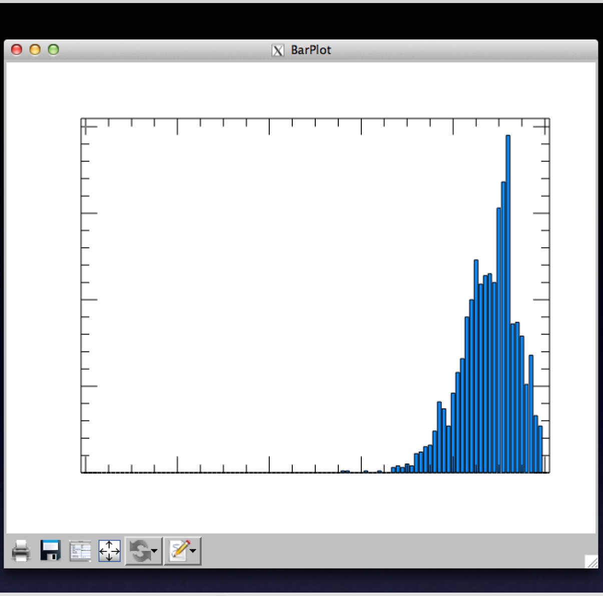
Command: ../process/process

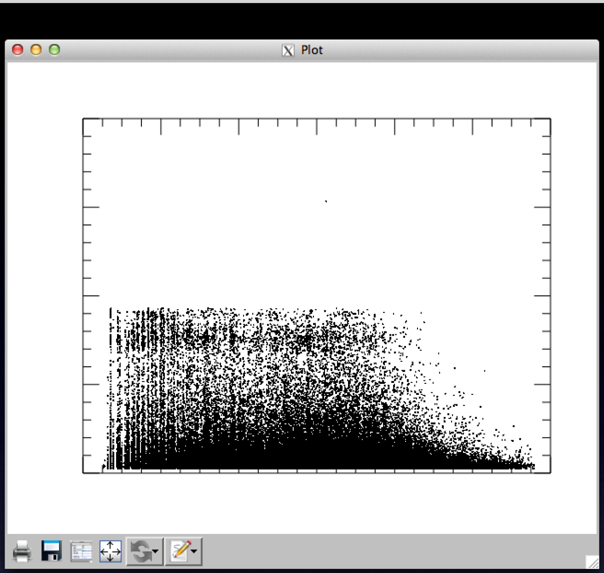
cheetah.ini file: lys-gd.ini

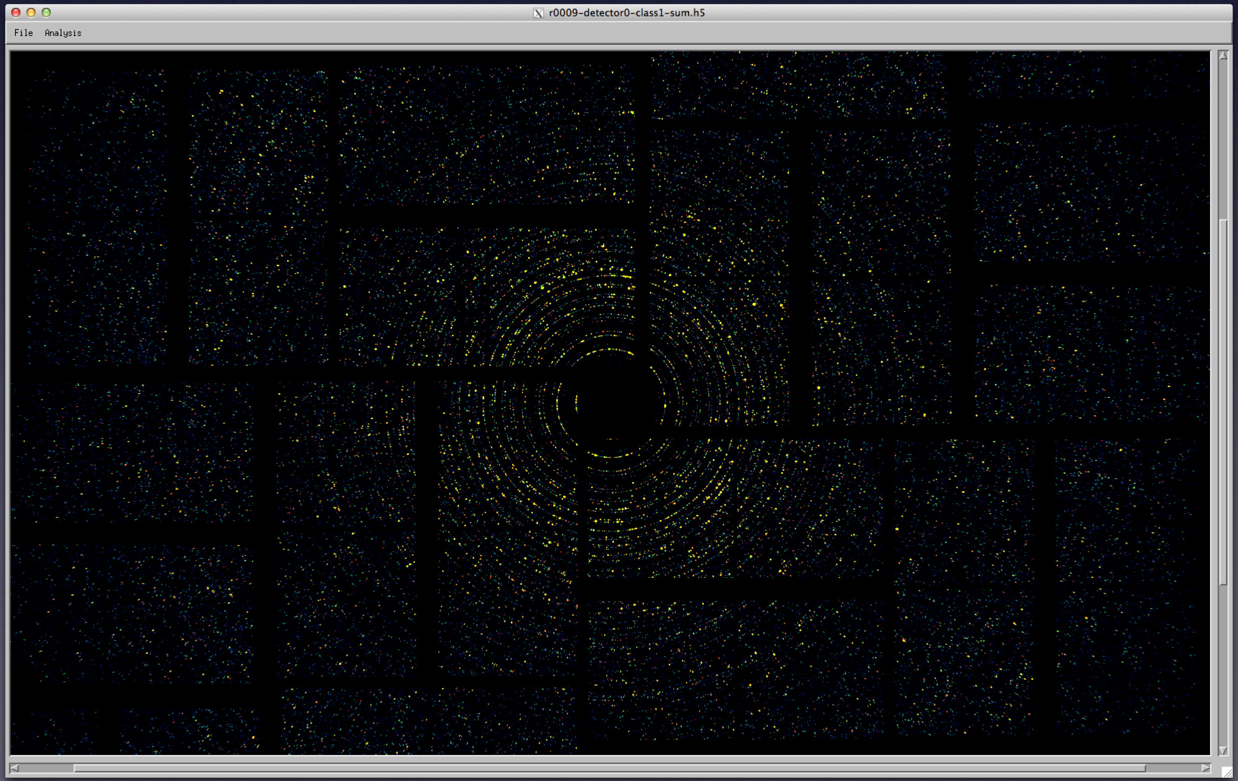
CFEL SCIENCE

Hit rate, resolution, saturation, and virtual powder

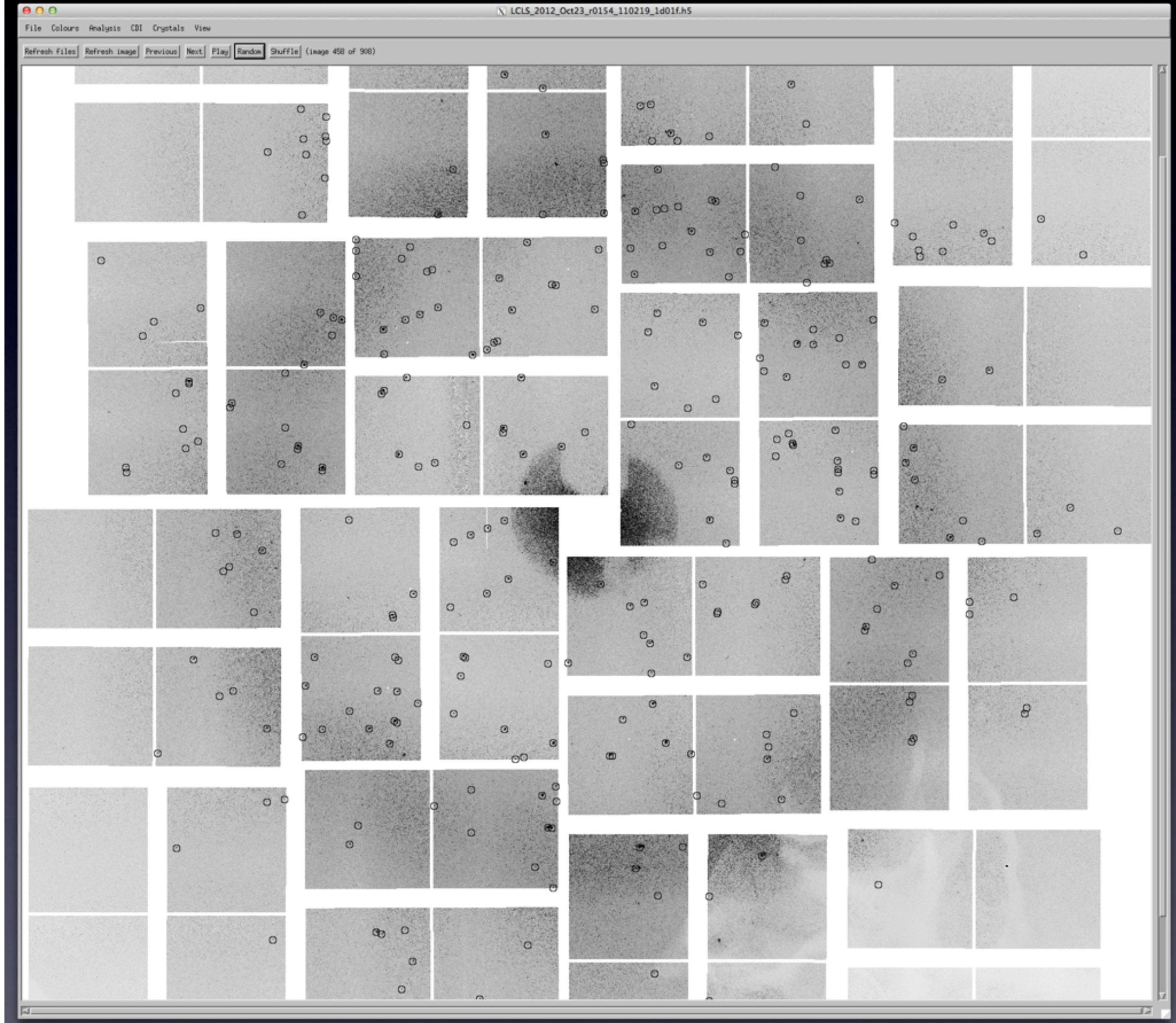








It is all boils down to quick but accurate (enough) peak finding



"Hitfinder 8"

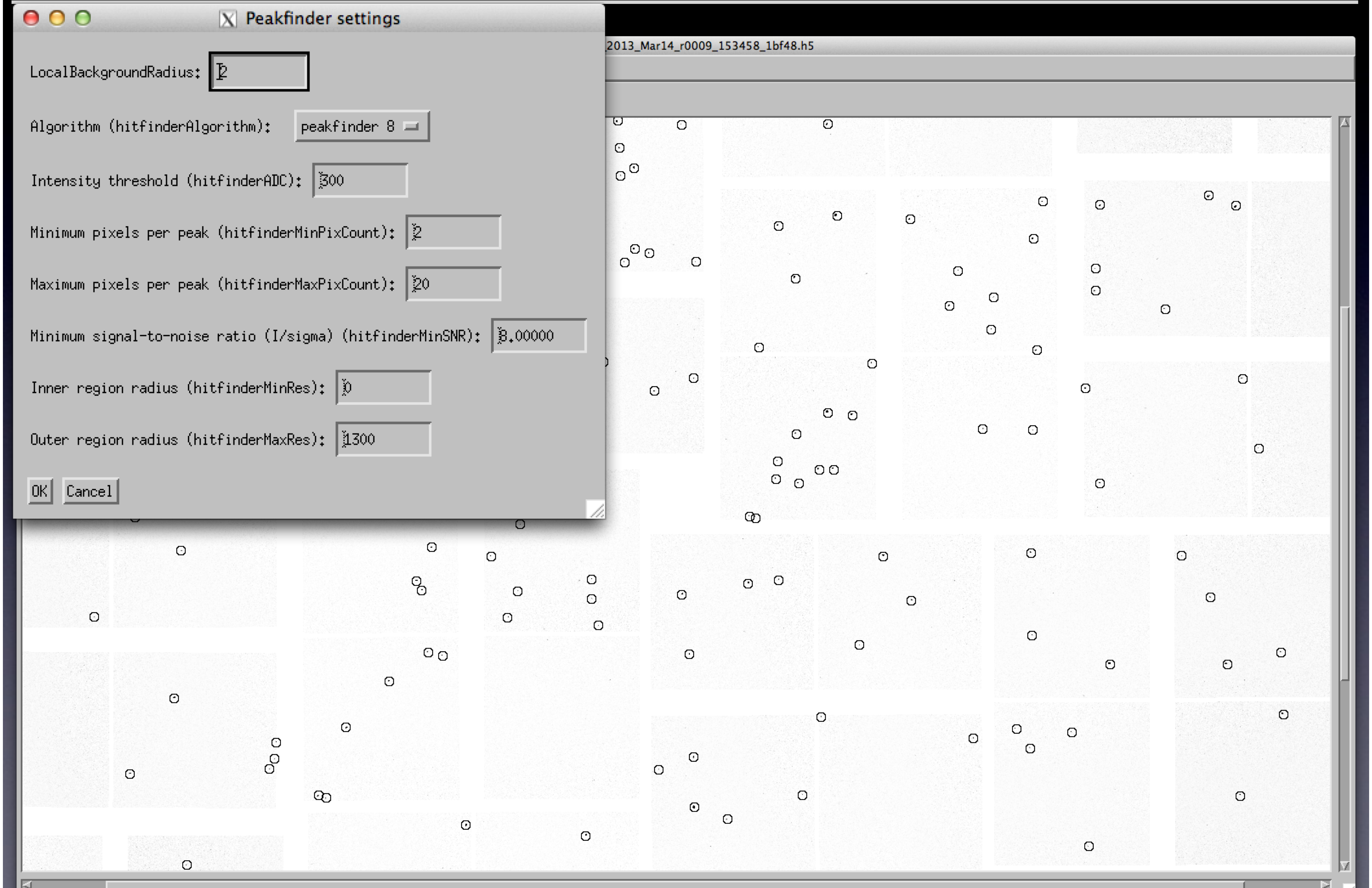
1. Calculate radial SNR and offset
2. thresh(r) = offset + s*sigma
   thresh(r) >= minADC
3. Peak = more than npix connected pixels above thresh(r)
4. Require more than n peaks

Critical inputs:

SNR
npix
minADC
rmin, rmax

CFEL SCIENCE

Optimisation of peak finding parameters using the GUI



LocalBackgroundRadius: 2

Algorithm (hitfinderAlgorithm): peakfinder 8

Intensity threshold (hitfinderADC): 300

Minimum pixels per peak (hitfinderMinPixCount): 2

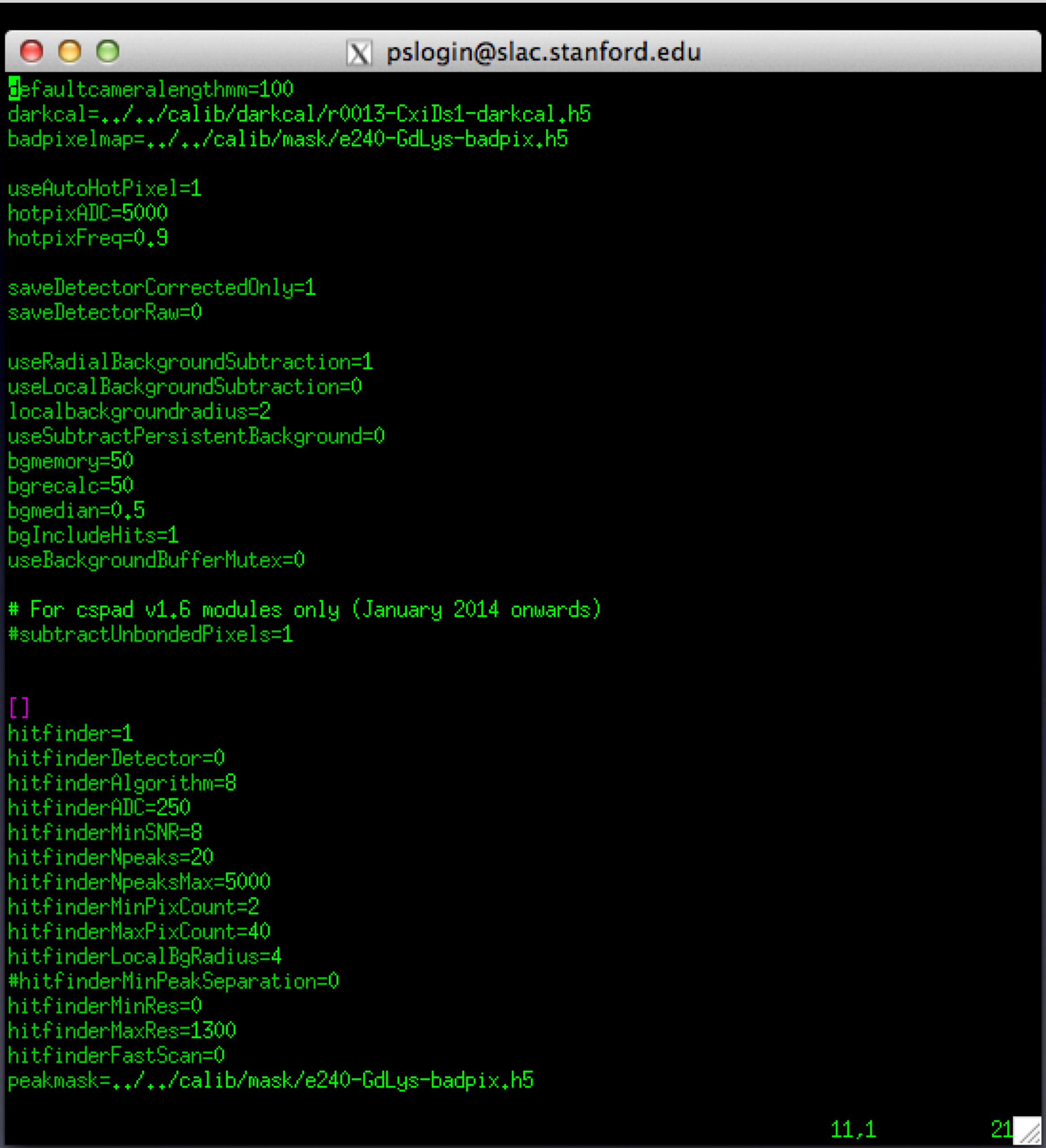
Maximum pixels per peak (hitfinderMaxPixCount): 20

Minimum signal-to-noise ratio (I/sigma) (hitfinderMinSNR): 8.00000

Inner region radius (hitfinderMinRes): 0

Outer region radius (hitfinderMaxRes): 1300

Edit lys-gd.ini

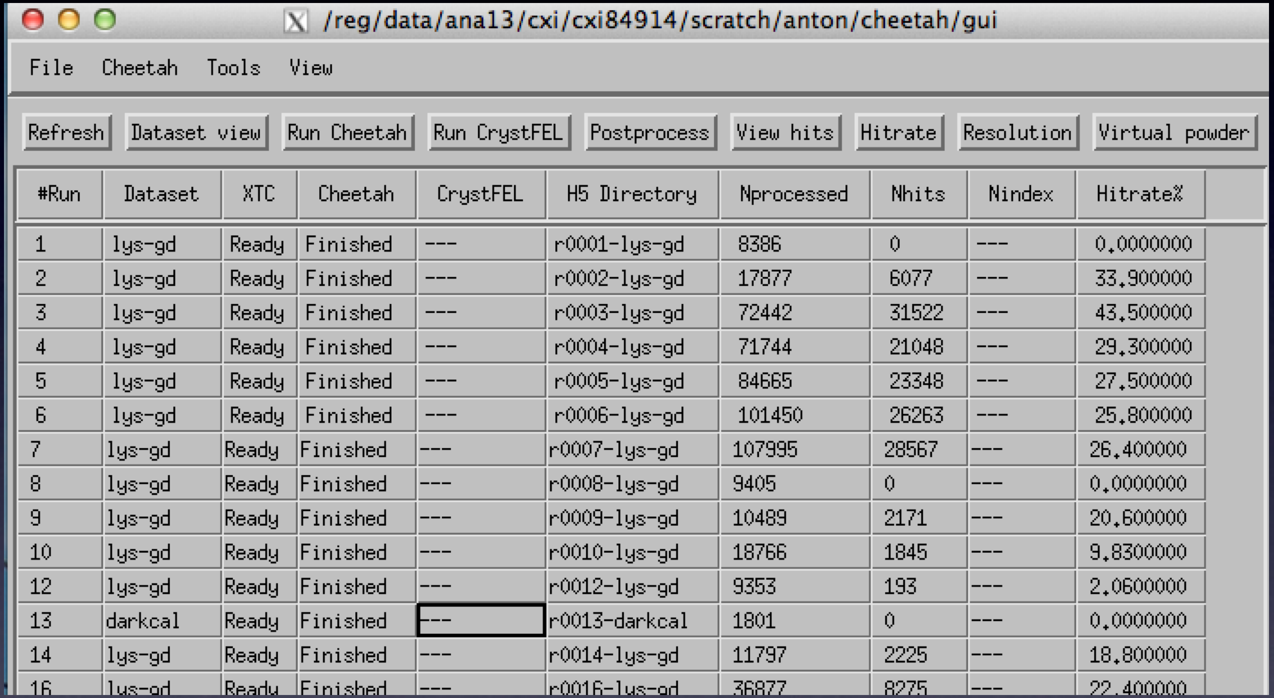


```
defaultcameralengthmm=100
darkcal=../../calib/darkcal/r0013-CxiDs1-darkcal.h5
badpixelmap=../../calib/mask/e240-GdLys-badpix.h5

useAutoHotPixel=1
hotpixADC=5000
hotpixFreq=0.9

saveDetectorCorrectedOnly=1
saveDetectorRaw=0

useRadialBackgroundSubtraction=1
useLocalBackgroundSubtraction=0
localbackgroundradius=2
useSubtractPersistentBackground=0
bgmemory=50
bgrecalc=50
bgmedian=0.5
bgIncludeHits=1
useBackgroundBufferMutex=0

# For cspad v1.6 modules only (January 2014 onwards)
#subtractUnbondedPixels=1


[]
hitfinder=1
hitfinderDetector=0
hitfinderAlgorithm=8
hitfinderADC=250
hitfinderMinSNR=8
hitfinderNpeaks=20
hitfinderNpeaksMax=5000
hitfinderMinPixCount=2
hitfinderMaxPixCount=40
hitfinderLocalBgRadius=4
#hitfinderMinPeakSeparation=0
hitfinderMinRes=0
hitfinderMaxRes=1300
hitfinderFastScan=0
peakmask=../../calib/mask/e240-GdLys-badpix.h5

                                                        11,1          21
```

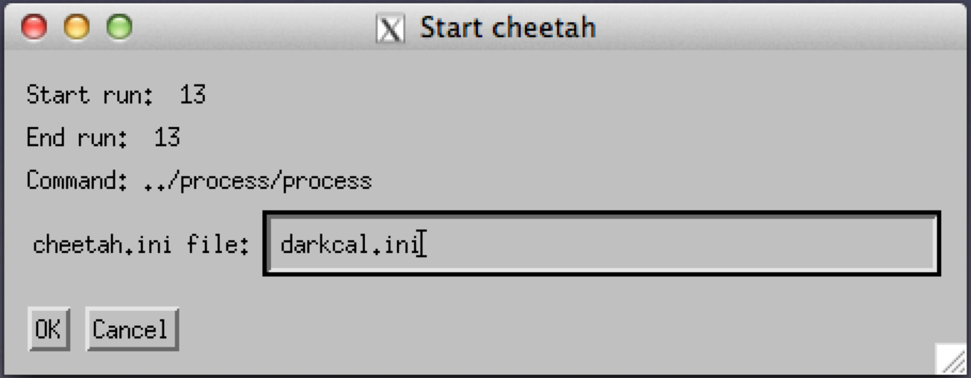
This is where all the processing options get changed

Dark calibration

Run 13 and 39 are dark runs.
Select run 13, click "Run Cheetah" and use darkcal.ini



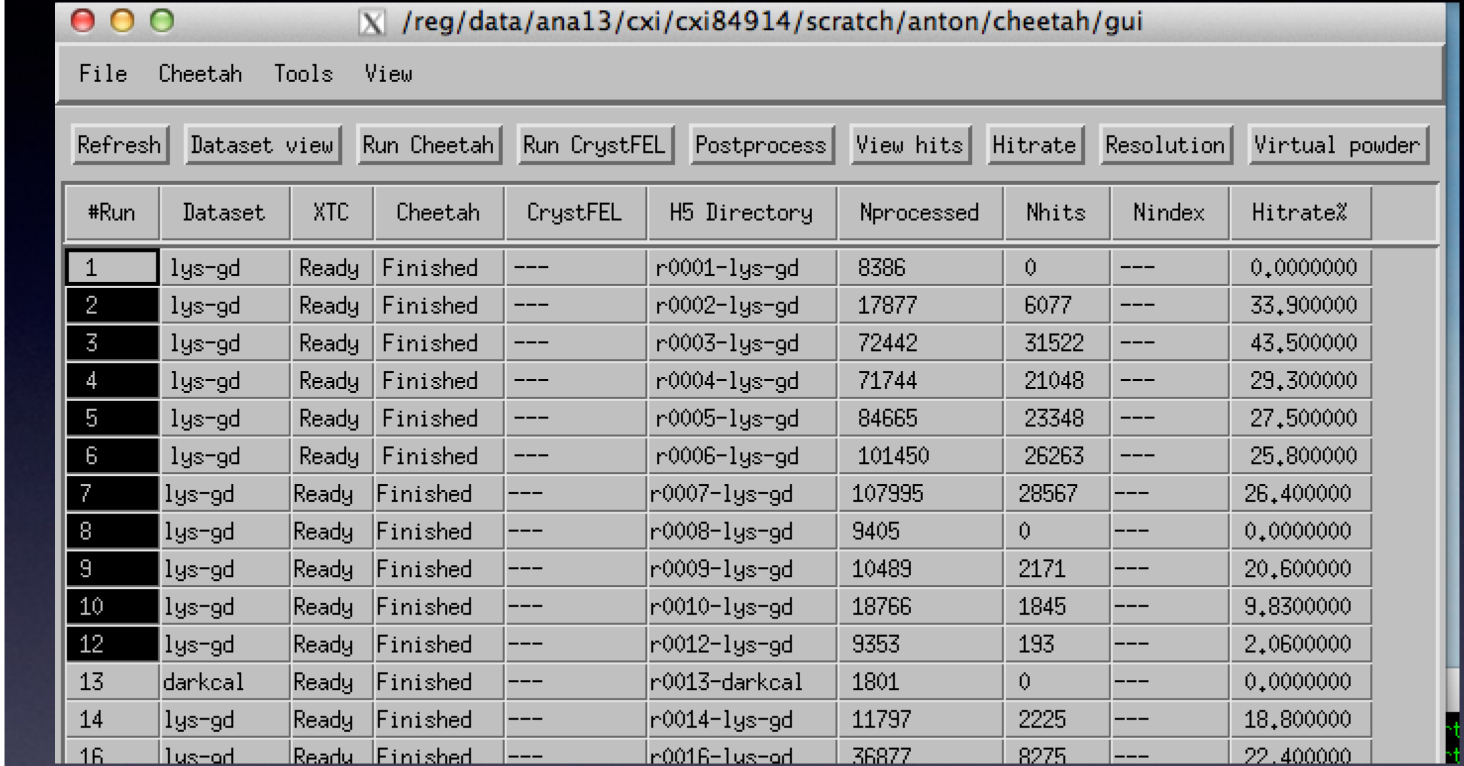
#Run	Dataset	XTC	Cheetah	CrystFEL	H5 Directory	Nprocessed	Nhits	Nindex	Hitrate%
1	lys-gd	Ready	Finished	---	r0001-lys-gd	8386	0	---	0.0000000
2	lys-gd	Ready	Finished	---	r0002-lys-gd	17877	6077	---	33.900000
3	lys-gd	Ready	Finished	---	r0003-lys-gd	72442	31522	---	43.500000
4	lys-gd	Ready	Finished	---	r0004-lys-gd	71744	21048	---	29.300000
5	lys-gd	Ready	Finished	---	r0005-lys-gd	84665	23348	---	27.500000
6	lys-gd	Ready	Finished	---	r0006-lys-gd	101450	26263	---	25.800000
7	lys-gd	Ready	Finished	---	r0007-lys-gd	107995	28567	---	26.400000
8	lys-gd	Ready	Finished	---	r0008-lys-gd	9405	0	---	0.0000000
9	lys-gd	Ready	Finished	---	r0009-lys-gd	10489	2171	---	20.600000
10	lys-gd	Ready	Finished	---	r0010-lys-gd	18766	1845	---	9.8300000
12	lys-gd	Ready	Finished	---	r0012-lys-gd	9353	193	---	2.0600000
13	darkcal	Ready	Finished	---	r0013-darkcal	1801	0	---	0.0000000
14	lys-gd	Ready	Finished	---	r0014-lys-gd	11797	2225	---	18.800000
16	lys-gd	Ready	Finished	---	r0016-lys-gd	36877	8275	---	22.400000



Start run: 13
End run: 13
Command: ../process/process
cheetah.ini file: darkcal.ini

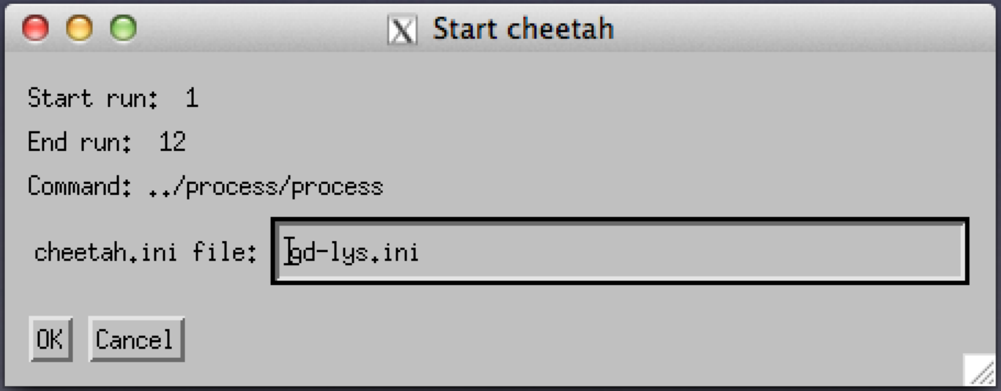
Process multiple runs

Select runs 1-12 and “Run Cheetah”



#Run	Dataset	XTC	Cheetah	CrystFEL	H5 Directory	Nprocessed	Nhits	Nindex	Hitrate%
1	lys-gd	Ready	Finished	---	r0001-lys-gd	8386	0	---	0.0000000
2	lys-gd	Ready	Finished	---	r0002-lys-gd	17877	6077	---	33.900000
3	lys-gd	Ready	Finished	---	r0003-lys-gd	72442	31522	---	43.500000
4	lys-gd	Ready	Finished	---	r0004-lys-gd	71744	21048	---	29.300000
5	lys-gd	Ready	Finished	---	r0005-lys-gd	84665	23348	---	27.500000
6	lys-gd	Ready	Finished	---	r0006-lys-gd	101450	26263	---	25.800000
7	lys-gd	Ready	Finished	---	r0007-lys-gd	107995	28567	---	26.400000
8	lys-gd	Ready	Finished	---	r0008-lys-gd	9405	0	---	0.0000000
9	lys-gd	Ready	Finished	---	r0009-lys-gd	10489	2171	---	20.600000
10	lys-gd	Ready	Finished	---	r0010-lys-gd	18766	1845	---	9.8300000
12	lys-gd	Ready	Finished	---	r0012-lys-gd	9353	193	---	2.0600000
13	darkcal	Ready	Finished	---	r0013-darkcal	1801	0	---	0.0000000
14	lys-gd	Ready	Finished	---	r0014-lys-gd	11797	2225	---	18.800000
16	lys-gd	Ready	Finished	---	r0016-lys-gd	36877	8275	---	22.400000

Do the same for runs 14-30 and 32-40



Start run: 1

End run: 12

Command: ../process/process

cheetah.ini file: gd-lys.ini