PAUL SCHERRER INSTITUT

PSI

Dr. Tim Grüne :: Paul Scherrer Institut :: tim.gruene@psi.ch

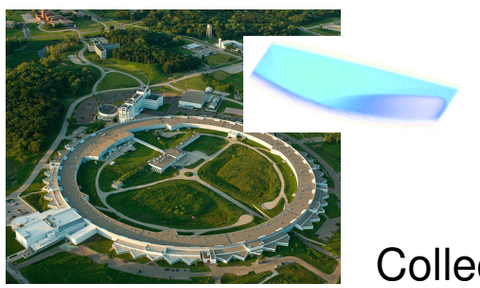
Data Processing with XDS

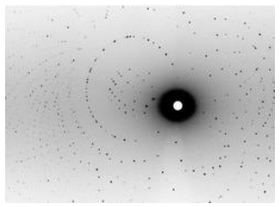
CCP4 / APS School Chicago 2017

19th June 2017

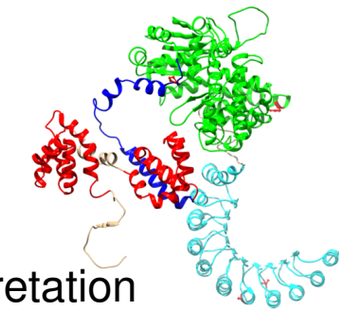
1 - X-ray Diffraction in a Nutshell

Model Building & Refinement — Ideal Crystal Data


Collection


Reduction

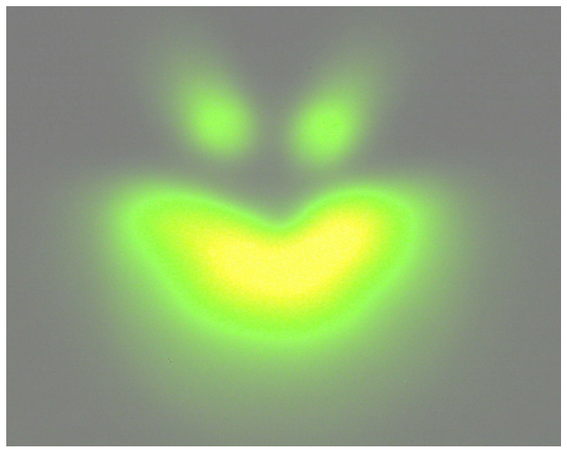
```
 4 -3 -5   390.18 10.15
-5  4 -3   350.72  9.09
67  0 -9   1.2019 0.897
```


Interpretation

- Purpose of crystal structure determination: Molecular Model: atom type, x, y, z, B
- Refinement of model against $h, k, l, F_{\text{ideal}}(hkl), \sigma_F$ or $h, k, l, I_{\text{ideal}}(hkl), \sigma_I$
- Ideal: independent of machine, of wavelength, of crystal shape and size

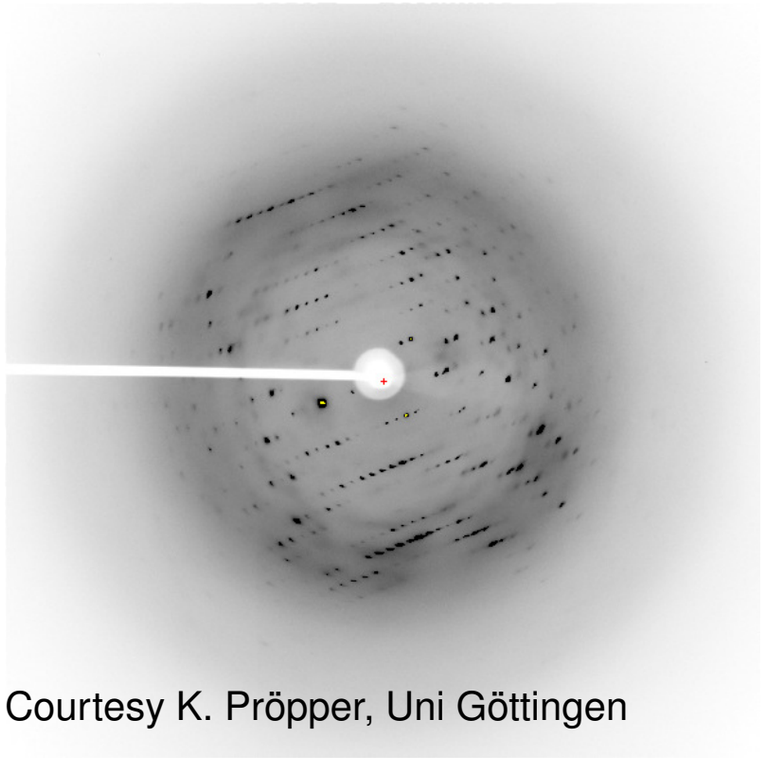
Experimental Errors

Unfocused Beam



Courtesy N. Sanshvili, & S. Corcoran, APS Chicago

Poorly diffracting crystals



Courtesy K. Pröpper, Uni Göttingen

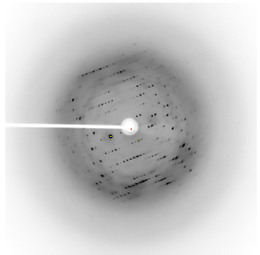
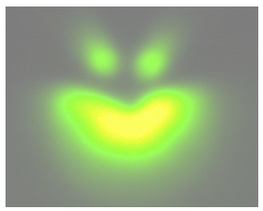
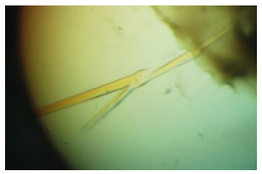
Sources of Errors

- Crystal imperfections
- Radiation damage
- Overloads
- Detector Background Noise
- Missettings (Note down wavelength and distance !!!)

Intensities and Amplitudes — Experiment *vs.* Theory

Intensities $I_{\text{exp}}(hkl)$ are experimental quantities *measured* from the detector.*

$$I_{\text{exp}}(hkl) = \frac{e^4}{{m_e}^2 c^4} \frac{V_{\text{crystal}}}{V_{\text{u.c.}}^2} \underbrace{\lambda^3 I_0 L P T E}_{\text{experimental factors}} I_{\text{ideal}}(hkl)$$


$=$

$\times$

$\times I_{\text{ideal}}(hkl)$

Data Processing:

Data integration: Determination of intensities $I_{\text{exp}}(hkl)$ from frames

Scaling / Merging: Determination of amplitudes $I_{\text{ideal}}(hkl)$ from $I_{\text{exp}}(hkl)$ and experimental settings

*Giacovazzo et el., "*Fundamentals of Crystallography*" (IUCr Texts on Crystallography), 1985, Kapitel "Diffraction Intensities"

2 - Data Processing with XDS

Data Processing

Processing your Data = Getting I_{ideal} from you experiment

Understanding your Reduction Program(s) = Getting the **best** I_{ideal} from you experiment

XDS Characteristics

Some features of XDS:

- Arbitrary Geometry (axes orientation, etc.)
- Parallelised: Fast
- 3-dimensional spot integration
- Correction for Radiation Damage available
- Close contact with Dectris (Pilatus, Eiger detectors)
- Command-line program
- Clear documentation

XDS & xdsGUI

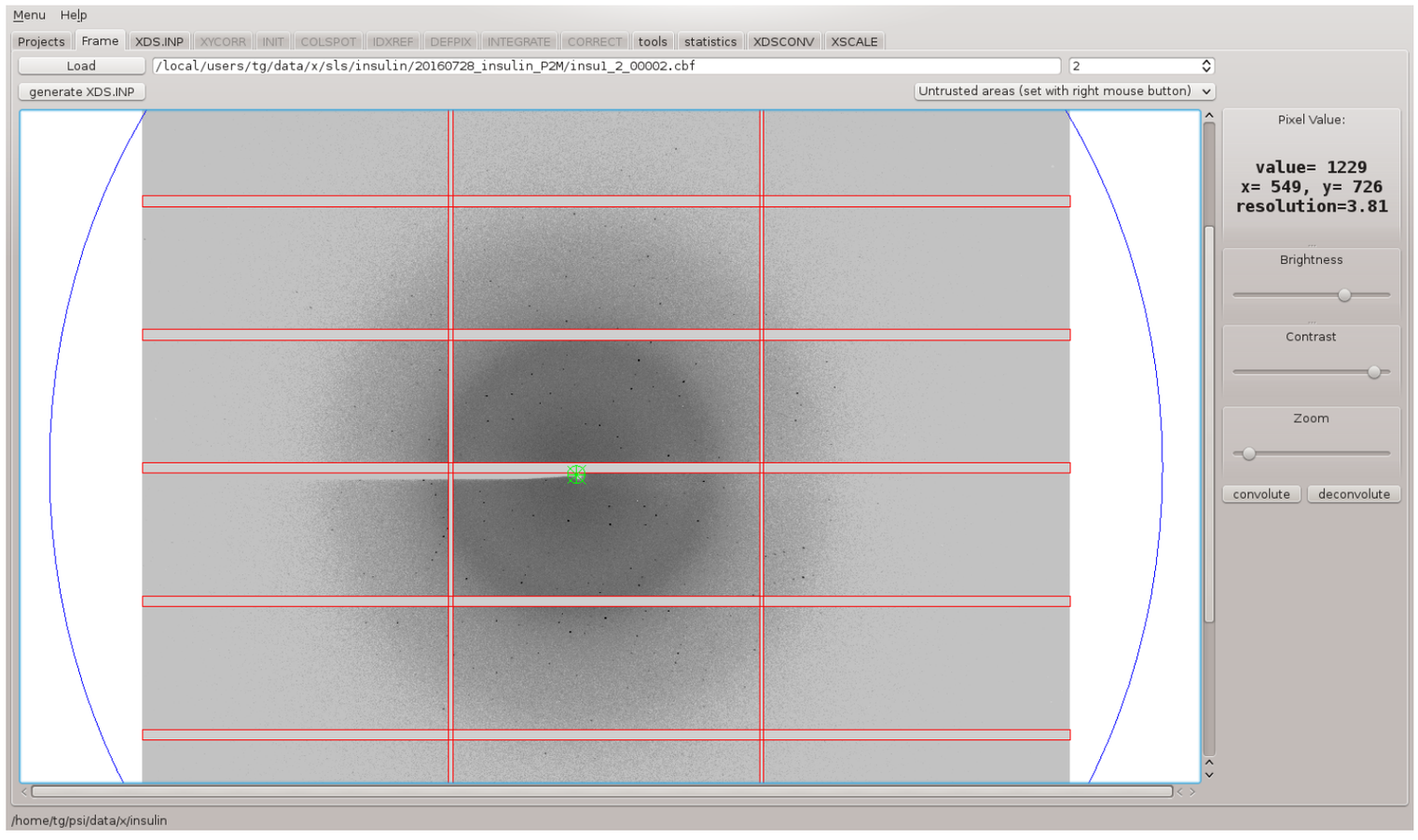
```
tg@hilbert:~/psi/data/x/insulin$ ls -l
total 8
-rw-r--r-- 1 tg tg 5239 Jun 12 15:42 XDS.INP
tg@hilbert:~/psi/data/x/insulin$
```

XDS controlled with single input file `XDS.INP`

Command: `xds_par`

Graphical user interface (by Kay Diederichs)

Command: `xdsgui`)



XDS.INP

XDS is controlled by one single input file: `XDS.INP`.

- Name cannot be changed
- **Each data set must be run in separate directory to avoid overwriting of files.**
- Contains about 100 Keywords [a] of the form

  `KEYWORD=VALUE`
- Only about 10 Keywords must be modified for most data sets (*e.g.* image names, detector distance, number of images, etc.)
- Most important one:

  `JOB= XYCORR INIT COLSPOT IDXREF DEFPIX INTEGRATE CORRECT`

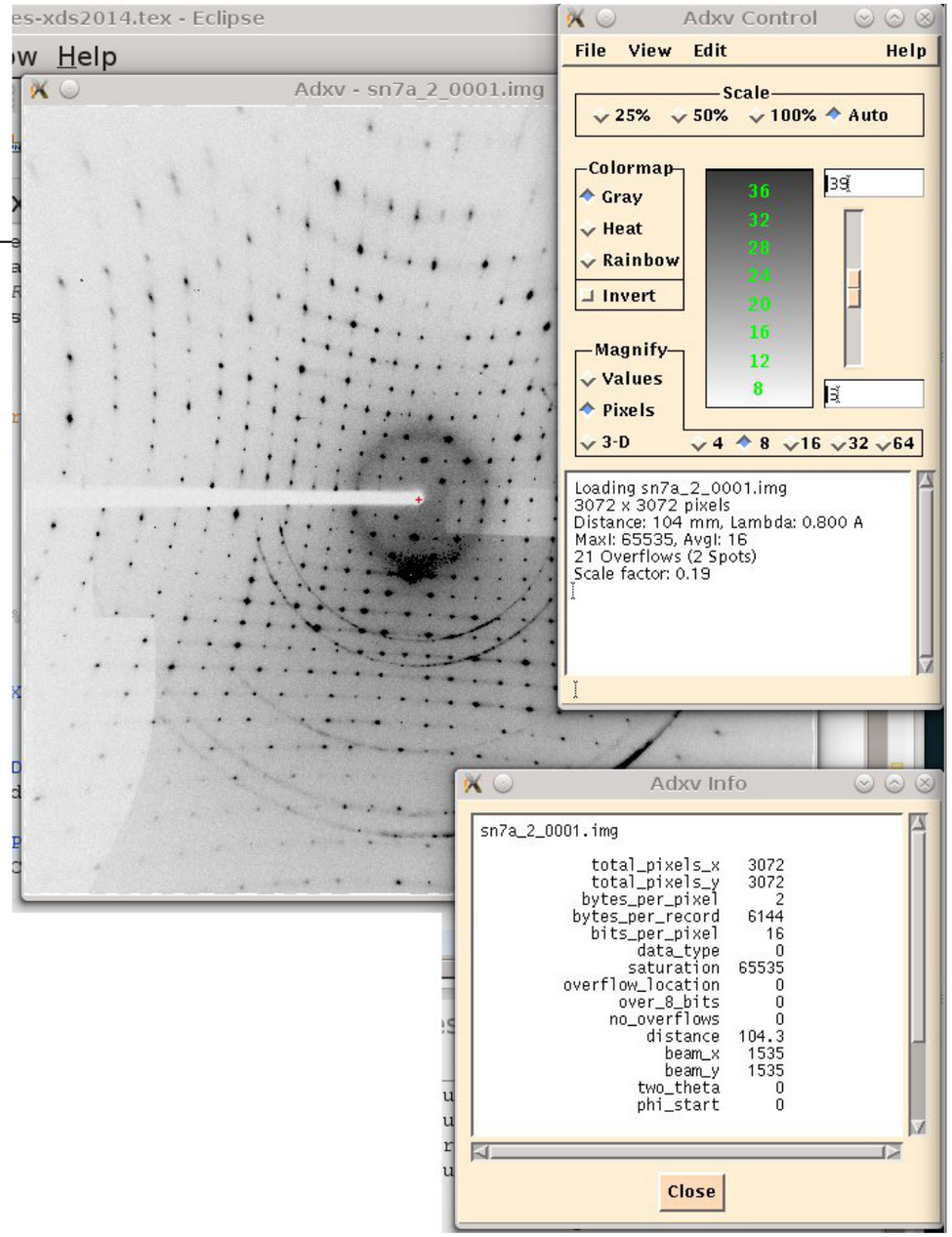
  Each name stands for one of the steps XDS carries out during data integration.
- `xdsgui` sets up `XDS.INP` automatically

[a] also called “cards” for historical reasons

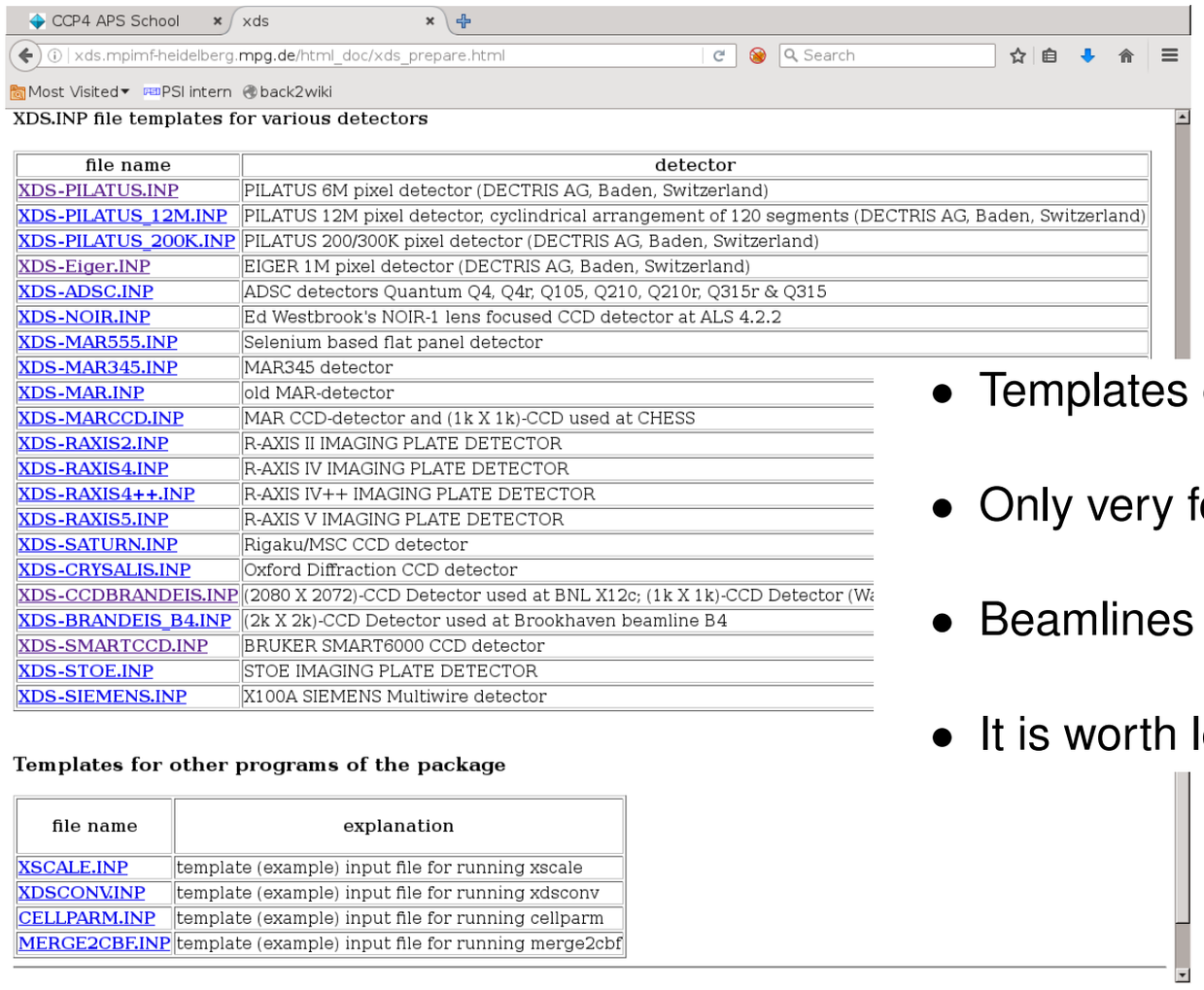
Generating `XDS.INP`

1. `generate_XDS.INP` as part of xdsGUI: from the latest version:

   tested with some datasets from ALS, SSRL, SLS, ESRF, BESSY, SPring-8 and PF; only MAR, ADSC/SMV, PILATUS, Eiger, RAXIS (in-house), Bruker (PHOTON II) detectors;
2. `adxv` Reads virtually all frame formats → manual editing of `XDS.INP`
3. `sfrmtools` Convert Bruker sfrm to XDS coordinate systems (available upon requrest from T.G.)



Templates for Manual XDS.INP



CCP4 APS School | xds

xds.mpimf-heidelberg.mpg.de/html_doc/xds_prepare.html

Most Visited | PSI intern | back2wiki

XDS.INP file templates for various detectors

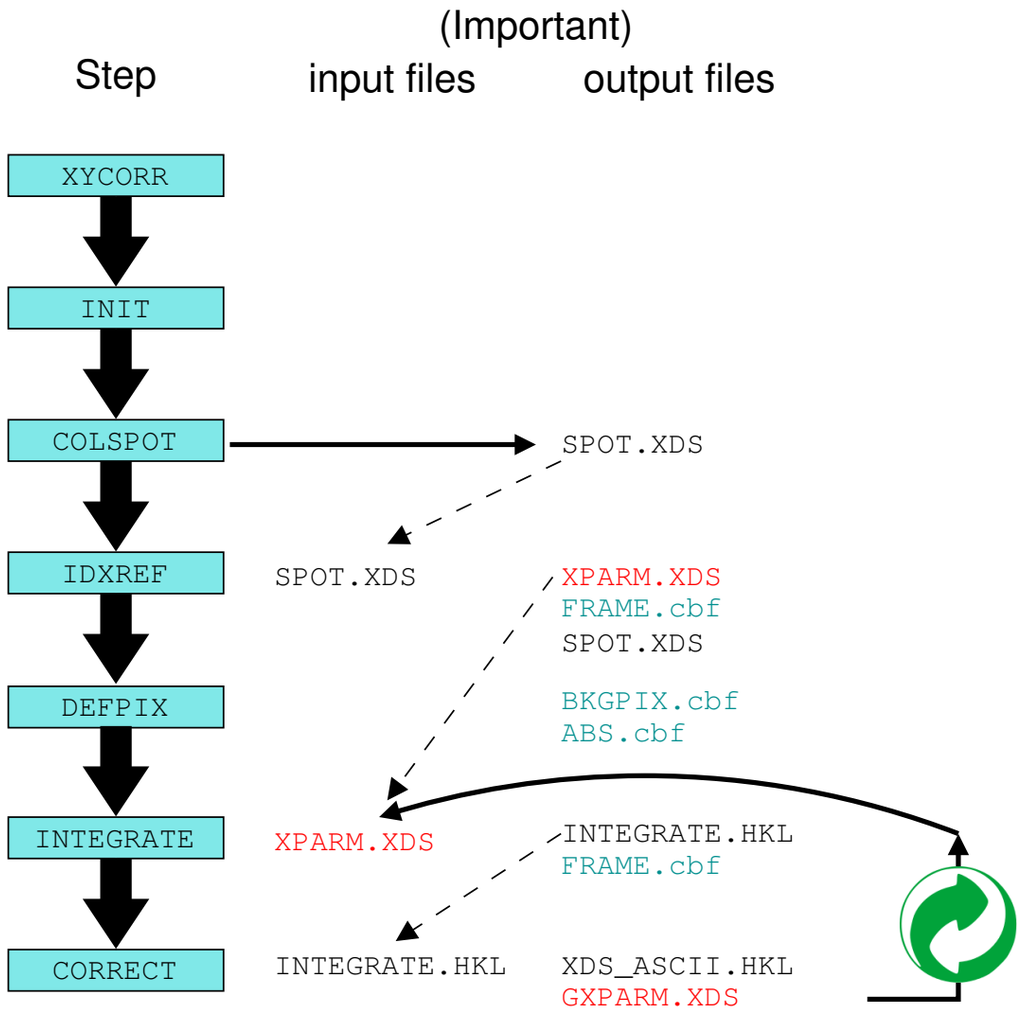
file name	detector
XDS-PILATUS.INP	PILATUS 6M pixel detector (DECTRIS AG, Baden, Switzerland)
XDS-PILATUS_12M.INP	PILATUS 12M pixel detector, cyclindrical arrangement of 120 segments (DECTRIS AG, Baden, Switzerland)
XDS-PILATUS_200K.INP	PILATUS 200/300K pixel detector (DECTRIS AG, Baden, Switzerland)
XDS-Eiger.INP	EIGER 1M pixel detector (DECTRIS AG, Baden, Switzerland)
XDS-ADSC.INP	ADSC detectors Quantum Q4, Q4r, Q105, Q210, Q210r, Q315r & Q315
XDS-NOIR.INP	Ed Westbrook's NOIR-1 lens focused CCD detector at ALS 4.2.2
XDS-MAR555.INP	Selenium based flat panel detector
XDS-MAR345.INP	MAR345 detector
XDS-MAR.INP	old MAR-detector
XDS-MARCCD.INP	MAR CCD-detector and (1k X 1k)-CCD used at CHESS
XDS-RAXIS2.INP	R-AXIS II IMAGING PLATE DETECTOR
XDS-RAXIS4.INP	R-AXIS IV IMAGING PLATE DETECTOR
XDS-RAXIS4++.INP	R-AXIS IV++ IMAGING PLATE DETECTOR
XDS-RAXIS5.INP	R-AXIS V IMAGING PLATE DETECTOR
XDS-SATURN.INP	Rigaku/MSC CCD detector
XDS-CRYSALIS.INP	Oxford Diffraction CCD detector
XDS-CCDBRANDEIS.INP	(2080 X 2072)-CCD Detector used at BNL X12c; (1k X 1k)-CCD Detector (Wa
XDS-BRANDEIS_B4.INP	(2k X 2k)-CCD Detector used at Brookhaven beamline B4
XDS-SMARTCCD.INP	BRUKER SMART6000 CCD detector
XDS-STOE.INP	STOE IMAGING PLATE DETECTOR
XDS-SIEMENS.INP	X100A SIEMENS Multiwire detector

Templates for other programs of the package

file name	explanation
XSCALE.INP	template (example) input file for running xscale
XDSCONV.INP	template (example) input file for running xdsconv
CELLPARM.INP	template (example) input file for running cellparm
MERGE2CBF.INP	template (example) input file for running merge2cbf

- Templates of input scripts for all supported detector formats
- Only very few adjustments necessary to get started
- Beamlines often generate appropriate input scripts
- It is worth learning how to set it up from scratch!

Steps of XDS (JOB=)



red: Text files with parameters or data

cyan: control images (use xds-viewer)

black: data files for further processing

Final data in `XDS_ASCII.HKL`

The Steps

XYCORR writes files for positional corrections of the detector plane. Most modern detectors provide already corrected images so that these to files are normally flat.

INIT determines initial detector background

COLSPOT Strong reflections for indexing

IDXREF indexing: unit cell dimensions and crystal orientation

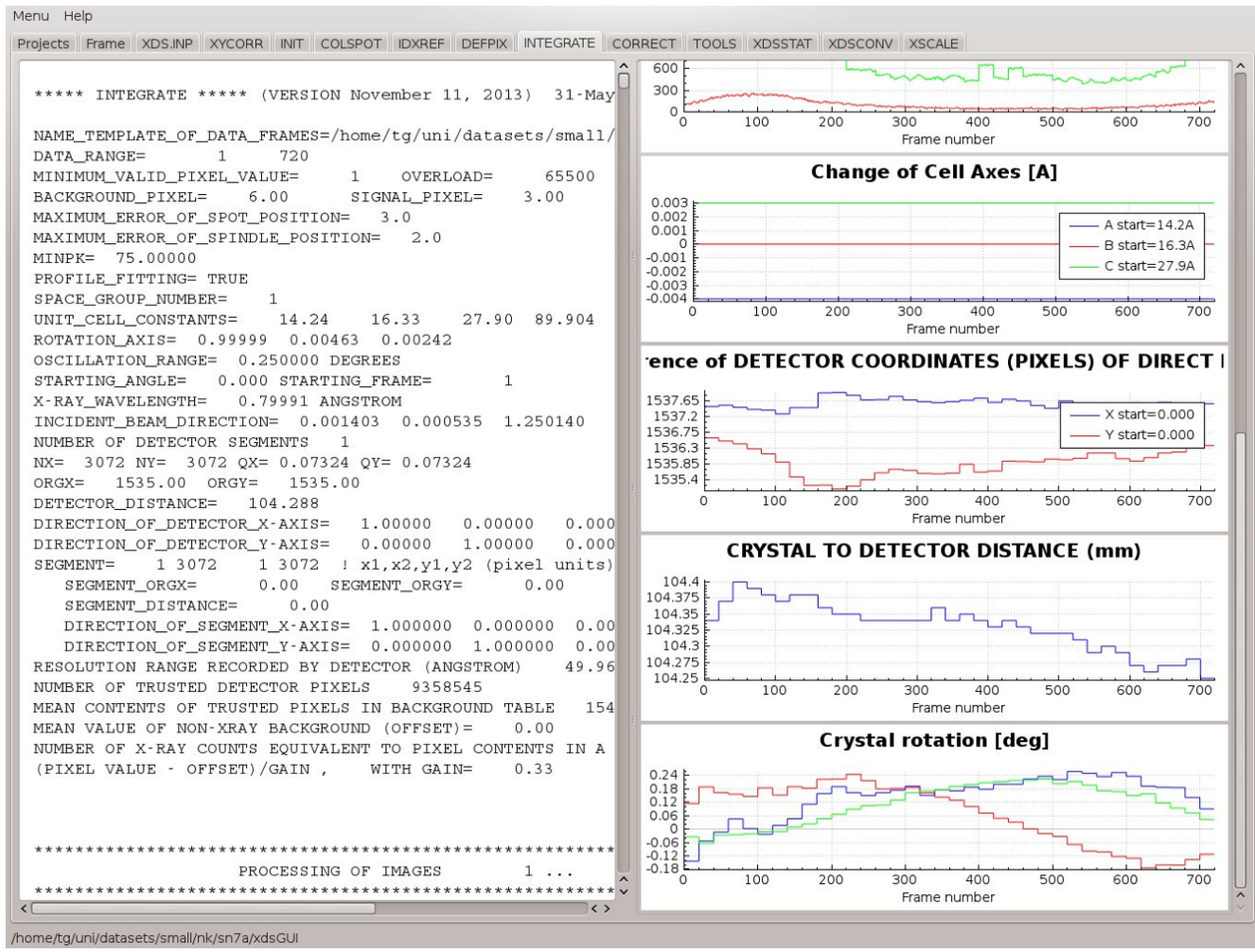
DEFPIX set active dectector area (exclude resolution cut-off, beam stop shadow, ...)

INTEGRATE extract reflection intensities from frames $\rightarrow I_{\text{exp}(hkl)}$

CORRECT applies corrections (polarisation, Lorentz-correction, ...), scales reflections, reports data statistics $\rightarrow I_{\text{ideal}(hkl)}$

Program Flow

- Each step must be passed at least once - the subsequent steps depend on files produced by the previous steps.
- Log-files for each step (`XYINIT.LP`, `INIT.LP`, ...).
- `IDXREF` = main hurdle - once unit cell and crystal orientation are determined, integration usually runs smoothly.
- `CORRECT` scaling and quality tables
- xdsGUI illustrates important graphs



IDXREF

- Indexing step: Find cell parameters and cell orientation.
- First refinement of experimental parameters (Detector distance, ...)
- Writes solution to `XPARM.XDS`

```
 XPARM.XDS     VERSION November 11, 2013
     1          0.0000     0.1000  0.999978  0.006046  0.002667
     0.826568        0.000543        0.001864        1.209820
 199      78.0597      78.0597      78.0597  90.000  90.000  90.000
    77.801140       -0.752043        6.303378
     2.363492       75.369080      -20.179979
    -5.891697       20.303986       75.142204
        1       2463       2527    0.172000    0.172000
  1224.162720    1249.473389     170.145401
     1.000000        0.000000        0.000000
     0.000000        1.000000        0.000000
     0.000000        0.000000        1.000000
        1          1       2463          1       2527
  0.00    0.00     0.00  1.00000  0.00000  0.00000  0.00000  1.00000  0.00000
```

SPOT.XDS

- COLSPOT: Detector coordinates and Intensity of **strong** spots to be used for indexing:

```
   X(pixel)   Y(pixel)    #image     counts
   1056.11    1529.51      15.35      2544.
   1895.52    1525.49       9.19      2481.
   1913.43    1547.90       2.63      1999.
```

- IDXREF: Miller-Indices according to XPARM.XDS

```
   X(pixel)   Y(pixel)    #image     counts         H         K         L
   1056.11    1529.51      15.35      2544.       14        13        18
   1895.52    1525.49       9.19      2481.       -7       -13       -14
   1913.43    1547.90       2.63      1999.        0         0         0   <---
```

- 0 0 0: not consistent with cell

"!!! ERROR !!! SOLUTION IS INACCURATE"

Correct indexing is crucial for data integration. If XDS indexes <50 % of all spots in `SPOT.XDS` it stops with the above error message.

Most common reasons:

1. Wrong `ORGX`, `ORGY`
2. Wrong Parameter settings (Detector distance, wavelength)
3. Twinning, multiple crystals

How to proceed?

The XDS error message `SOLUTION IS INACCURATE` does not necessarily mean that something is seriously wrong. The step IDXREF still refines the experimental parameters and writes them into `XPARM.XDS`. This is all we need to continue.

Check `IDXREF.LP` for

- Number of indexed reflections

```
***** INDEXING OF OBSERVED SPOTS IN SPACE GROUP #   1 *****
   1109 OUT OF    2506 SPOTS INDEXED.
      0 REJECTED REFLECTIONS (REASON: OVERLAP)
   1397 REJECTED REFLECTIONS (REASON: TOO FAR FROM IDEAL POSITION)
```

- Does the cell make sense? (check with *e.g.* adxv)
- refined detector distance and detector origin do not shift too much (<5mm)

Set `JOB = DEFPIX INTEGRATE CORRECT` and integrate your data.

If something **did** seriously go wrong, the file XPARM.XDS would not been written.

DEFPIX: active detector mask

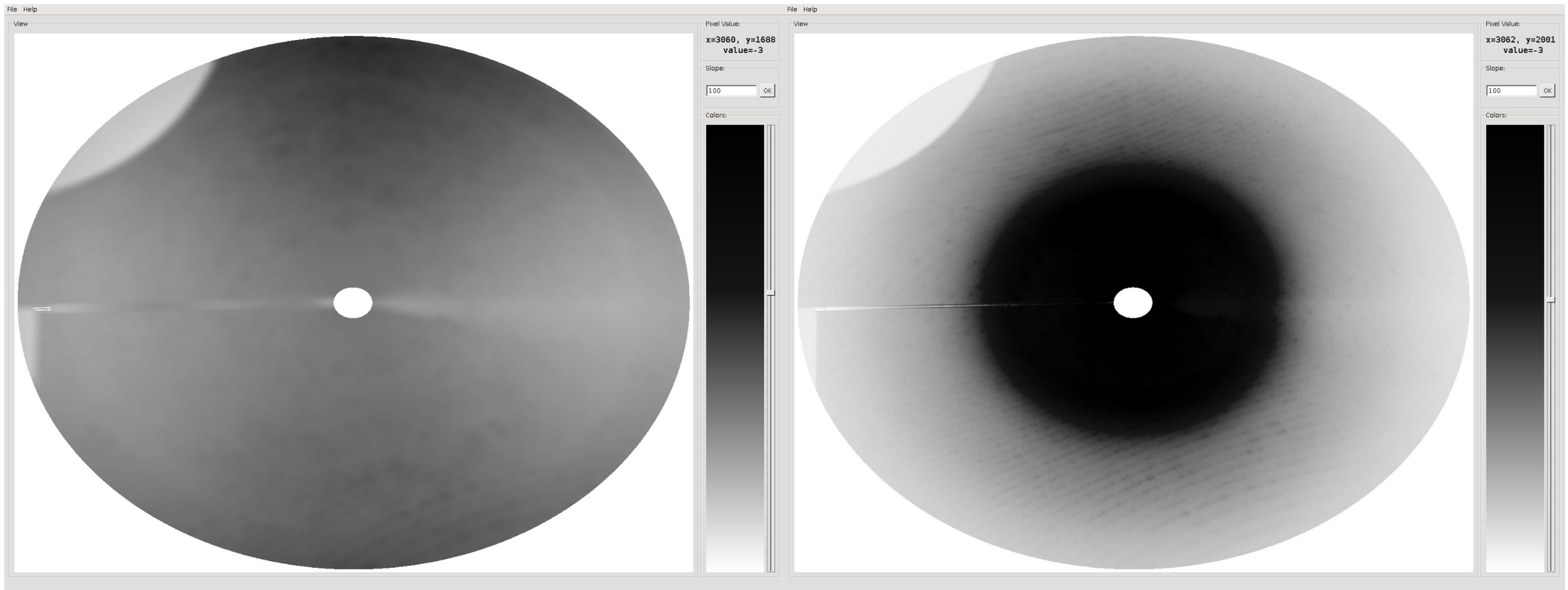
DEFPIX sets the area of the detector which is taken into account during integration. It takes into account:

1. INCLUDE_RESOLUTION_RANGE default: 20 Å to detector edge.
2. VALUE_RANGE_FOR_TRUSTED_DETECTOR_PIXELS exclude shadowed regions, *e.g.* beamstop, cryo stream nozzle
3. UNTRUSTED_RECTANGLE exclude gaps between chips of *e.g.* Pilatus detector
4. EXCLUDE_RESOLUTION_RANGE exclude ice rings from data integration

VALUE_RANGE_FOR_TRUSTED_DETECTOR_PIXELS

`ABS.cbf`

`BKGPIX.cbf`, all included

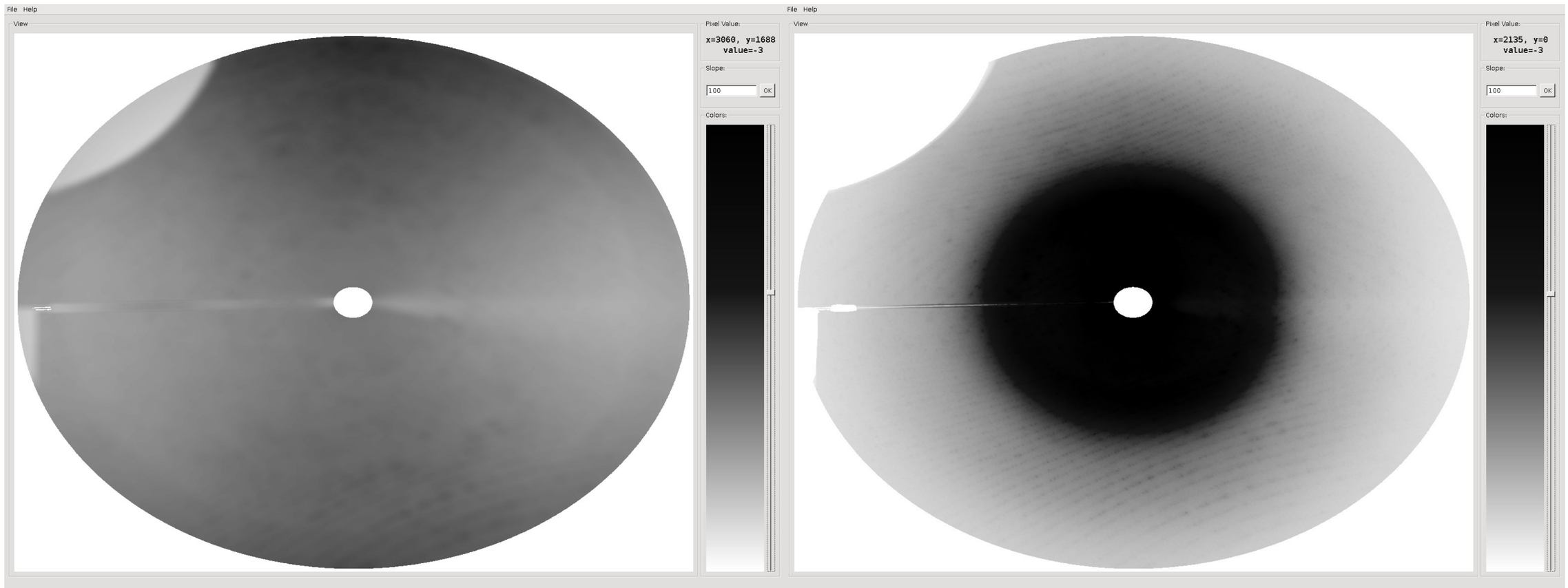



```
VALUE_RANGE_FOR_TRUSTED_DETECTOR_PIXELS= 1000 30000
```

VALUE_RANGE_FOR_TRUSTED_DETECTOR_PIXELS

ABS.cbf

BKGPIX.cbf, shadows removed



VALUE_RANGE_FOR_TRUSTED_DETECTOR_PIXELS= 6100 30000

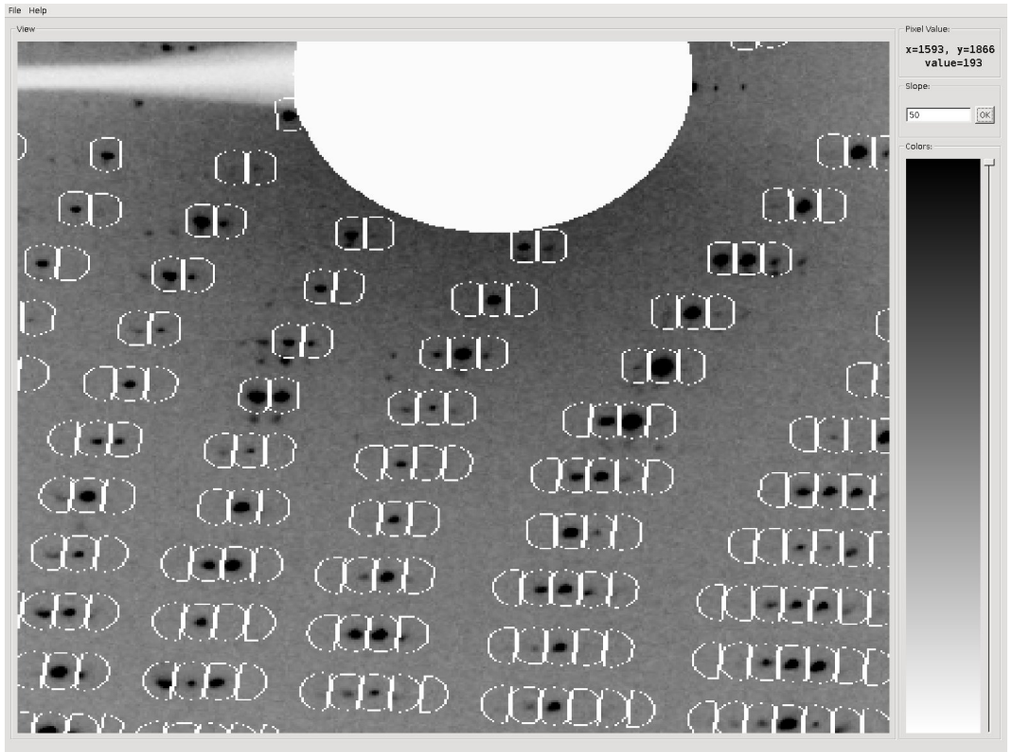
INTEGRATE and CORRECT: Intensity determination and fine tuning

INTEGRATE determine spot intensities based on parameters (saved in XPARM.XDS)

CORRECT experimental corrections (*e.g. Lorentz- and Polarisation-correction*), refine parameters and determine space group (saved to GXPARM.XDS)

CORRECT writes FRAME.cbf:

- predicted spot positions encircled
- check correctness of predictions



Recycling

- Parameters (in `GXPARM.XDS`) depend on measured intensities
- Intensities (including corrections) depend on Parameters

$\Rightarrow$ rename `GXPARM.XDS` to `XPARM.XDS` and rerun XDS (`JOB = DEFPIX INTEGRATE CORRECT`) to improve results.

Implications:

1. Scaling in XDS improves with higher symmetry than $P1$
2. Proper **resolution cut–off** can improve data quality
3. Unit cell parameters improve

Resolution Cut-Off

The default resolution range in XDS is 20 Å to the detector edge

```
INCLUDE_RESOLUTION_RANGE=20.0 0.0
```

- Medium to low resolution data: increase 20 Å to 30 Å or even 50 Å (check `BKGPIX.cbf`)
- Once confident about space group (technically: Laue group): determine high-resolution cut-off.

High Resolution Cut-Off

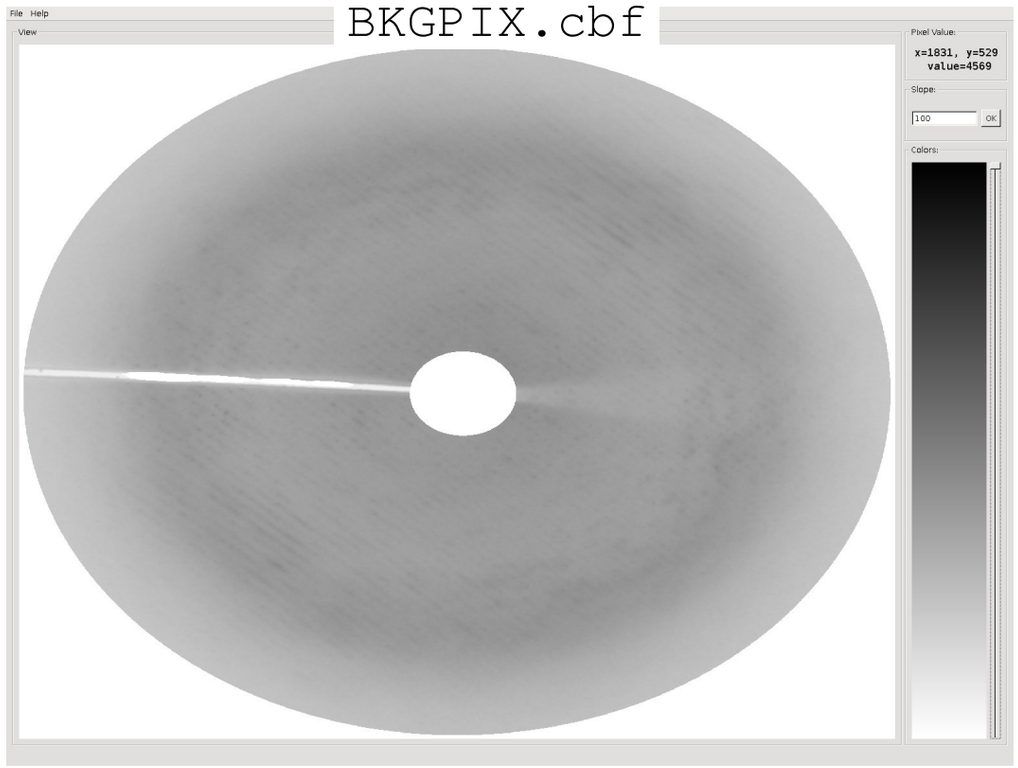
SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

	RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR COMPARED expected		I/SIGMA	R-meas	CC(1/2)
	3.23	12757	2442	2447	99.8%	1.9%	2.0%	12753	72.35	2.1%	99.9*
	2.29	22608	4464	4470	99.9%	2.8%	2.6%	22607	49.33	3.2%	99.9*
	1.87	30231	5714	5714	100.0%	5.1%	4.7%	30230	28.89	5.6%	99.8*
	1.62	33623	6803	6809	99.9%	12.2%	12.1%	33598	12.30	13.7%	99.0*
	1.45	40406	7684	7684	100.0%	27.5%	28.3%	40402	5.72	30.5%	95.0*
-->	1.32	41510	8518	8524	99.9%	59.3%	62.4%	41482	2.50	66.5%	79.1*
									^^^^^^^		
	1.22	45003	9238	9243	99.9%	124.7%	131.8%	45000	1.14	139.9%	45.4*
	1.15	37687	9934	9940	99.9%	210.4%	228.0%	37484	0.52	244.9%	16.7*
	1.08	11083	6440	10653	60.5%	256.8%	279.8%	7939	0.22	336.4%	9.4
	total	274908	61237	65484	93.5%	5.2%	5.3%	271495	11.89	5.8%	100.0*

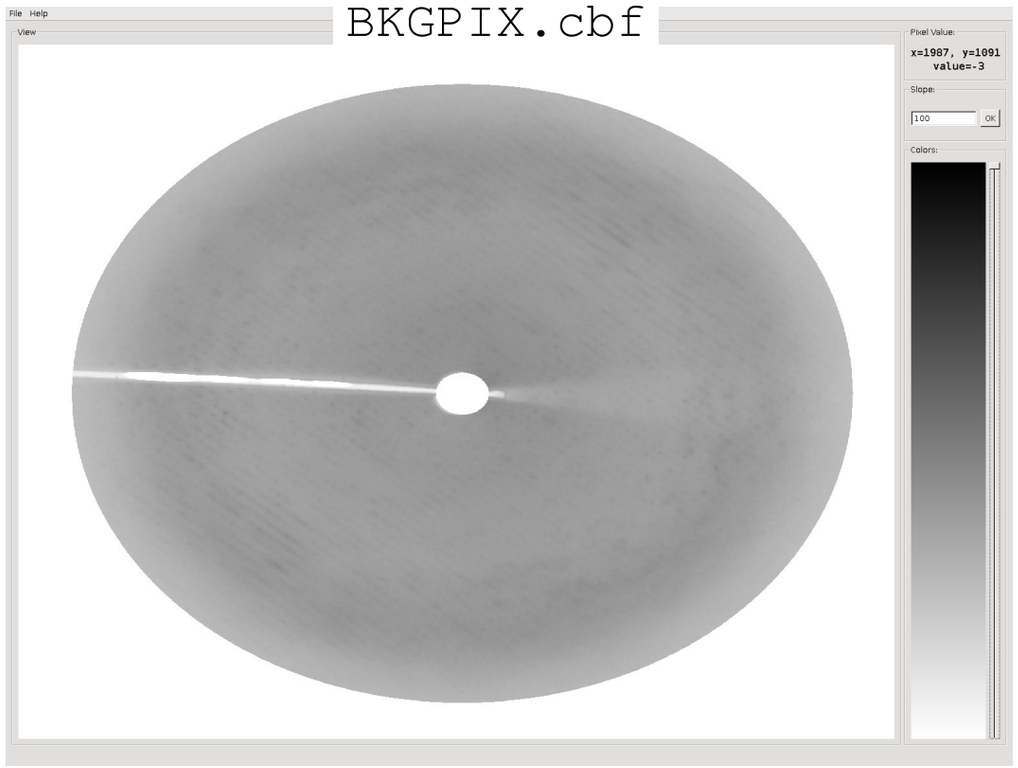
Logfile `CORRECT.LP`

- Conservative cut–off: $I/\sigma_I \geq 2.0$ — should correspond to $CC(1/2) \geq 70\%$
- If resolution matters: $I/\sigma_I \geq 1.0$ — ensure $CC(1/2) \approx 50\%$

Low Resolution Cut-Off



Default resolution range (20 Å - edge)
- includes noise at edge
- loose low resolution reflections



Adjusted resolution range (40 Å - 2.85 Å)
- $I/\sigma_I \approx 2$ in outer shell
- low resolution reflection important for “shape” of molecule (MR, refinement)

Multiple Datasets: REFERENCE_DATA_SET

In many space groups, indexing is not unique. $I2_13$: 24 possibilities.

```
     #     1   2   3   4     5   6   7   8     9 10 11 12
     1     1   0   0   0     0   1   0   0     0  0  1  0
     2     0   0   1   0     1   0   0   0     0  1  0  0
...
    24     0  -1   0   0     0   0  -1   6     1  0  0  0
```

`REFERENCE_DATA_SET= ../A1/XDS_ASCII.HKL` in `XDS.INP` takes care of everything.

Otherwise: data sets do not merge.

Scaling Multiple Datasets

- `XDS_ASCII.HKL` **already scaled** and ready to use

- Multiple data sets (e.g. inverse beam): Use `XSCALE`

- Alternative: `pointless` from CCP4

Command: `xscale_par`; Control: `XSCALE.INP`

```
OUTPUT_FILE=insulin.HKL
INPUT_FILE=../A1/XDS_ASCII.HKL
INPUT_FILE=../A2/XDS_ASCII.HKL
```

XSCALE.LP

```
 1  0.3887E+01      6622       0  ../k001/XDS_ASCII.HKL
 2  0.7979E+01      9383       0  ../k004/XDS_ASCII.HKL
 3  0.4293E+01      9871       0  ../k005/XDS_ASCII.HKL
 4  0.4267E+01      5138       0  ../k006/XDS_ASCII.HKL
```

DATA SETS #i	#j	NUMBER OF COMMON REFLECTIONS	CORRELATION BETWEEN i,j	RATIO OF COMMON INTENSITIES (i/j)	B-FACTOR BETWEEN i,j
1	2	461	0.976	0.4670	0.1388
1	3	172	0.972	1.1661	-0.0818
2	3	302	0.979	2.3026	-0.1889
1	4	136	0.405	0.6497	0.6766
2	4	200	0.349	1.8650	0.0965
3	4	73	0.214	1.2222	1.1729

- Correlation with data set #4 low
- Check for unit cell, indexing ambiguities, etc.
- Even poor data sets usually have CC>95%

The XDS Resources

availability:

XDS http://xds.mpimf-heidelberg.mpg.de/ main program suite

Wiki and auxiliary programs http://strucbio.biologie.uni-konstanz.de/xdswiki/

GUIs

- xdsGUI (http://strucbio.biologie.uni-konstanz.de/xdswiki/index.php/XDSgui)
- Xdsapp (http://www.helmholtz-berlin.de, search for “xdsapp”)

ADXV Andrew Arvai, http://www.scripps.edu/~arvai/adxv.html