PAUL SCHERRER INSTITUT

PSI

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

Data Processing with XDS

CCP4 / DLS Workshop 2017

30th November 2017

1 - Quick Start on XDS

1.1 - Documentation and Information

- XDS written by Wolfgang Kabsch (MPI Heidelberg) and co–developed by Kay Diederichs (Uni Konstanz)
- Main Site and Download: http://xds.mpimf-heidelberg.mpg.de
- Wiki, GUI, Auxiliary programs: https://strucbio.biologie.uni-konstanz.de/xdswiki
- Questions will also get answered at the CCP4bb: http://www.ccp4.ac.uk/ccp4bb.php

1.2 - How to use XDS

Requirements:

1. Diffraction Images

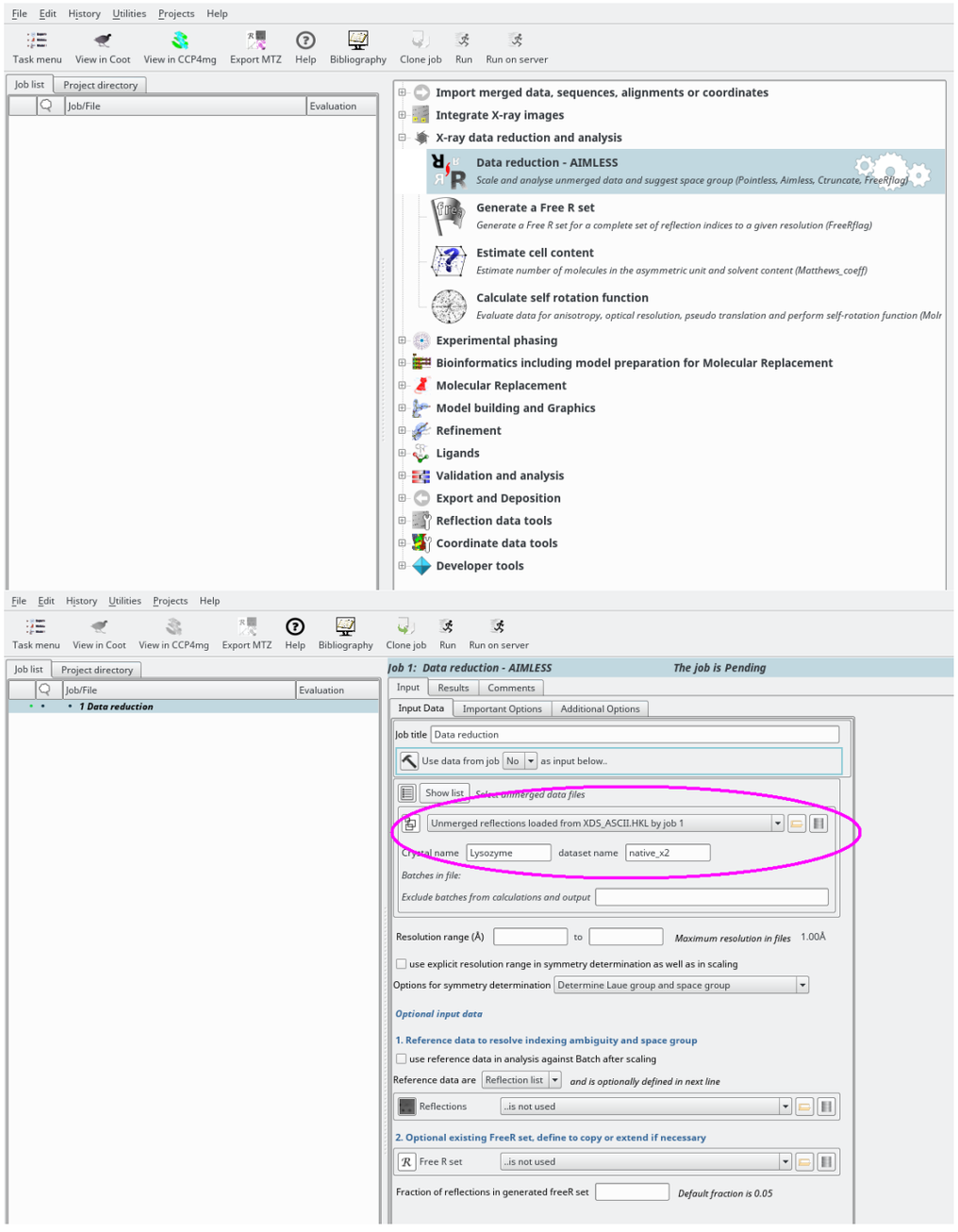
2. XDS.INP: Instruction file for XDS — plain text format

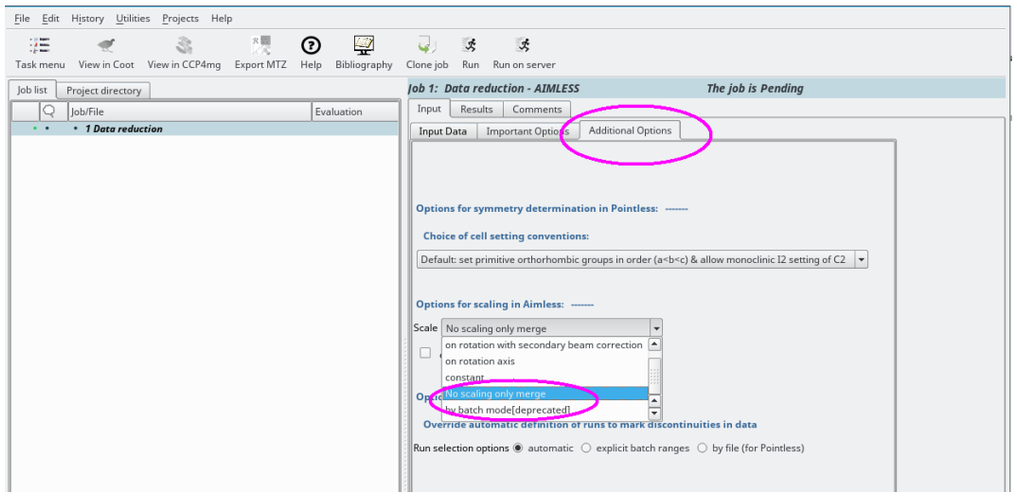
Output:

1. Single Text file `XDS_ASCII.HKL`

   - Scaled and corrected data
   - Contains full description of experiment — ready for import

1.3 - From `XDS_ASCII.HKL` to MTZ–file





Import of `XDS_ASCII.HKL` with the CCP4i2 — activate `ONLYMERGE` in `Aimless`

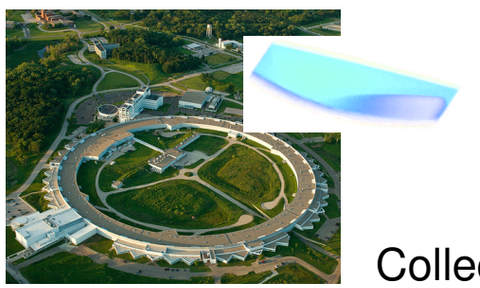
1.4 - Outline

- Purpose and steps of data processing

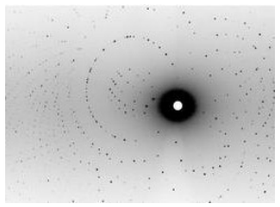
- Setting up and Running XDS

2 - X-ray Diffraction in a Nutshell

2.1 - Model Building & Refinement — Ideal Crystal Data

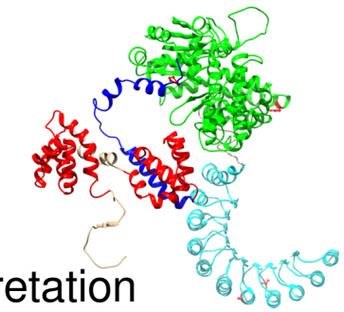


Collection



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

Reduction

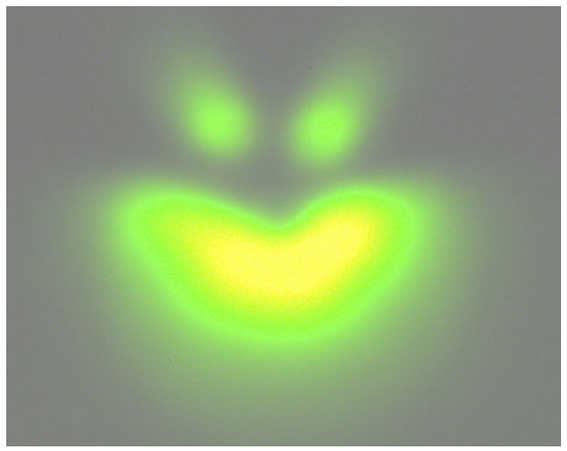


Interpretation

- 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

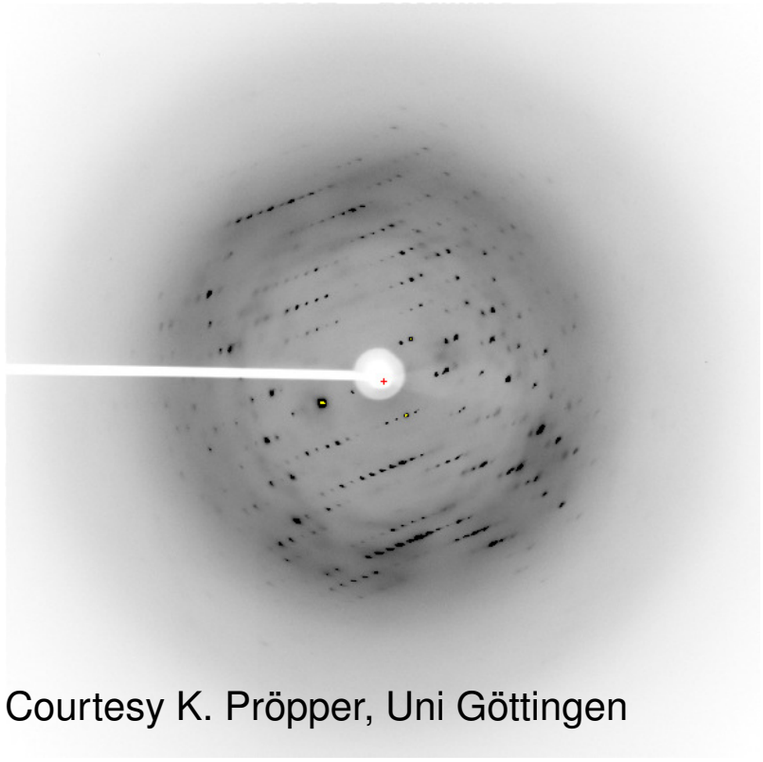
2.2 - Experimental Errors

Unfocused Beam



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

Poorly diffracting crystals



Courtesy K. Pröpper, Uni Göttingen

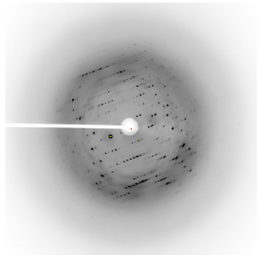
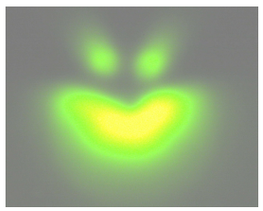
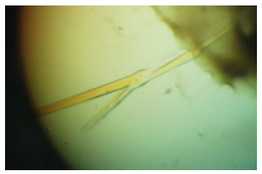
Sources of Errors

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

2.3 - 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"

3 - Data Processing with XDS

3.1 - XDS Characteristics

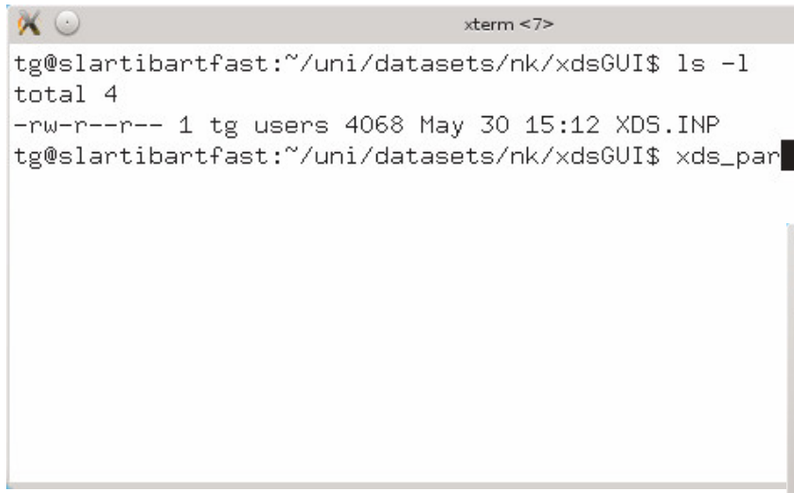
Processing your Data = Getting I_{ideal} from you experiment

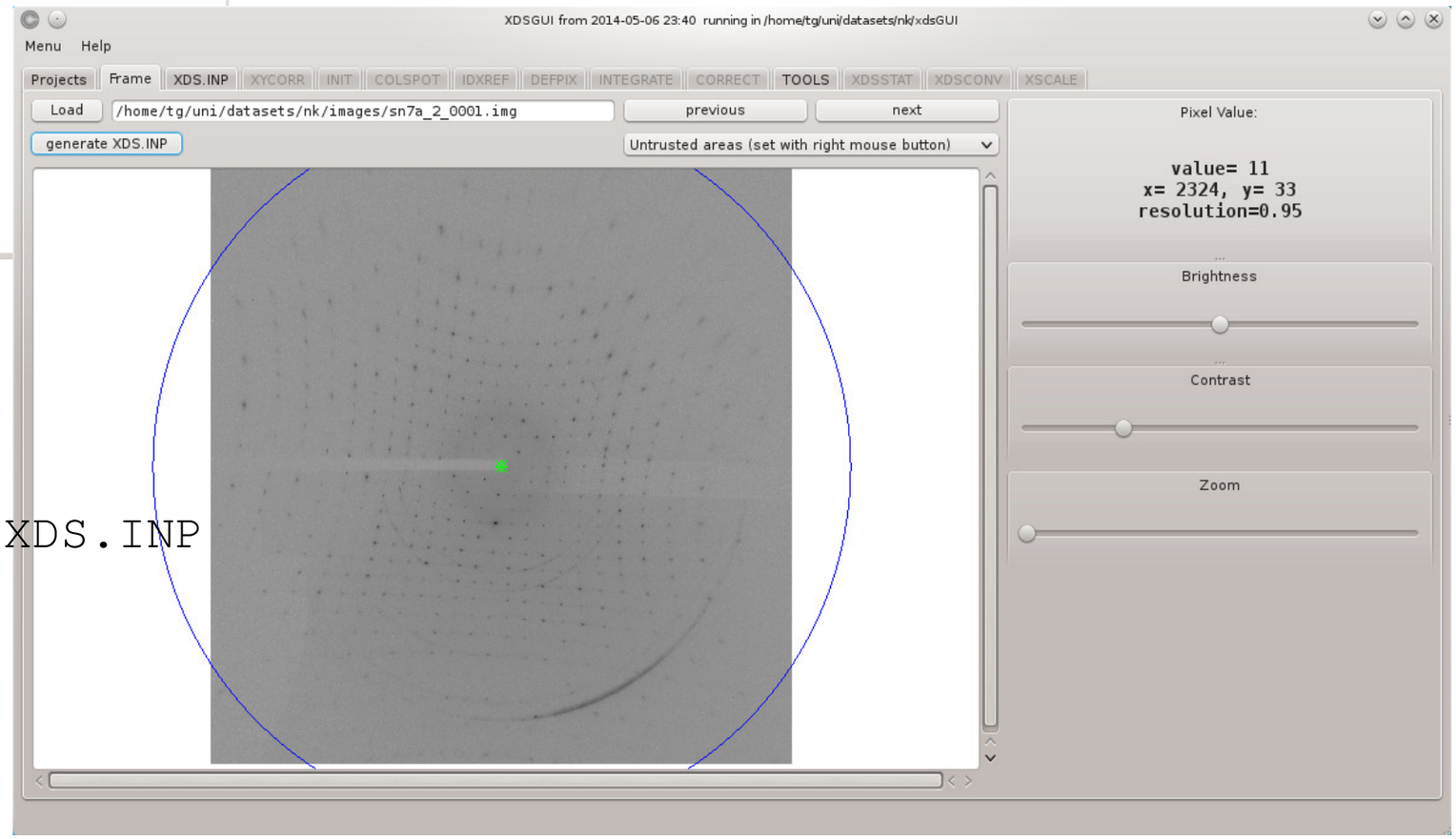
Understanding your Reduction Program(s) = Getting the **best** I_{ideal} from you experiment

Some features of XDS:

- Arbitrary experimental geometry (axes orientation, detector layout, . . .)
- Fast setup: XDS.INP not very complicated; fast runtime
- Exclusion of moving shadows — cf. Diamond's long wavelength beamline I23
- 3-dimensional spot integration
- Correction for Radiation Damage
- Integration of many datasets (Inverse beam, multi–crystal averaging, . . .)
- Close contact with Dectris (Pilatus, Eiger detectors)
- Command-line program
- Clear documentation

3.2 - XDS & xdsGUI





XDS controlled with single input file `XDS.INP`

Command: `xds_par`

Graphical user interface

Command: `xdsgui`

3.3 - 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

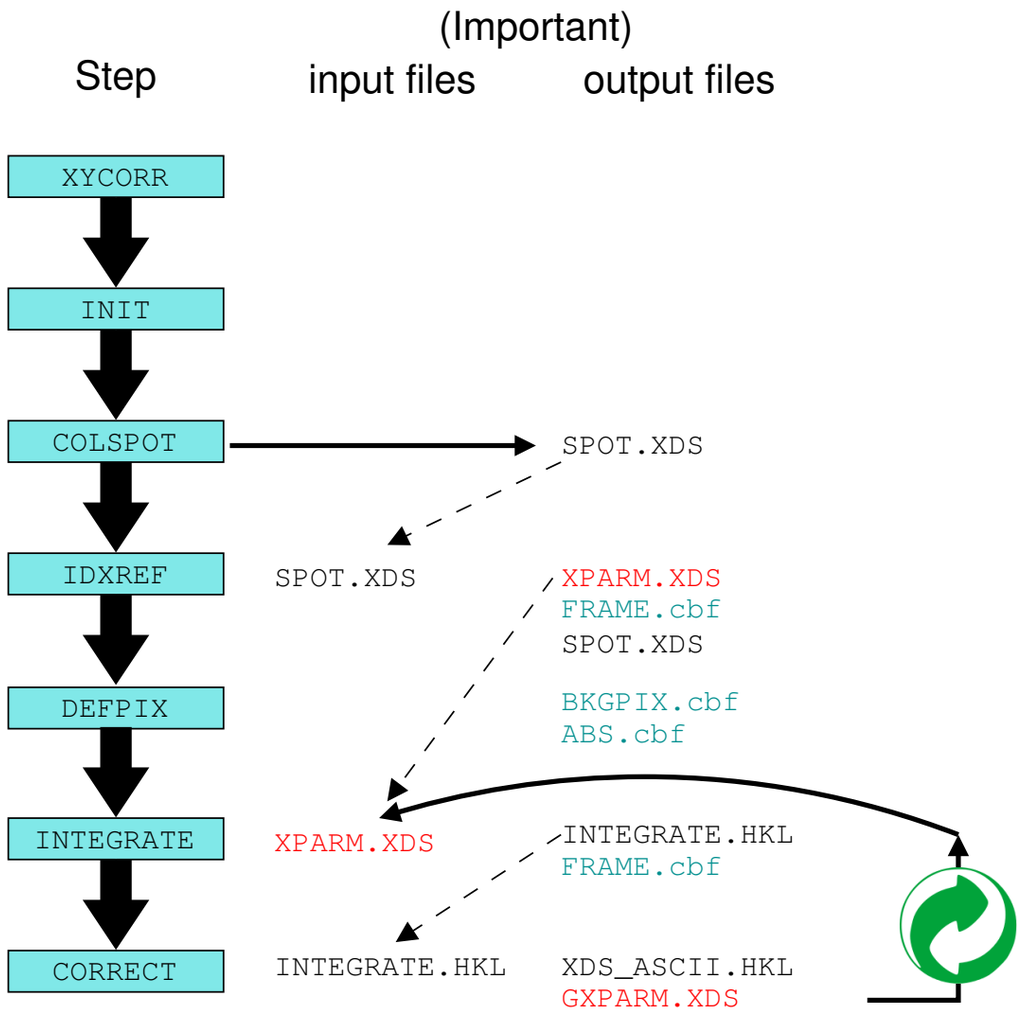
  `KEYWORD=VALUE`
- Only about 10 Keywords must be modified for most data sets (*e.g.* image names, detector distance, number of images, etc.)
- Most important one:

  `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

3.4 - 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

3.5 - 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

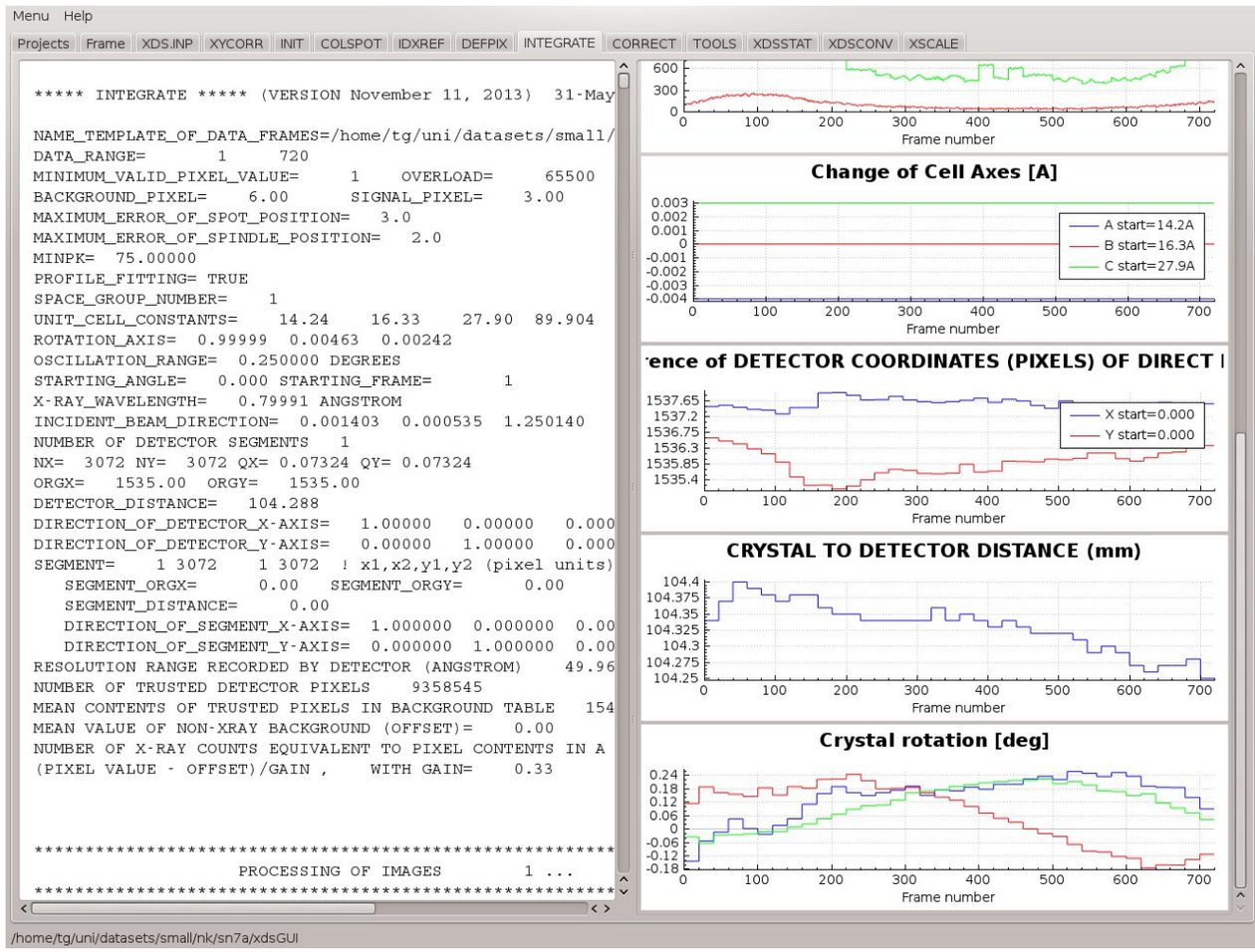
DEFPIX set active dectector area (exclude resolution cut-off, beam stop shadow, ...)

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

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

3.6 - 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



3.7 - 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
```

3.8 - 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.
```

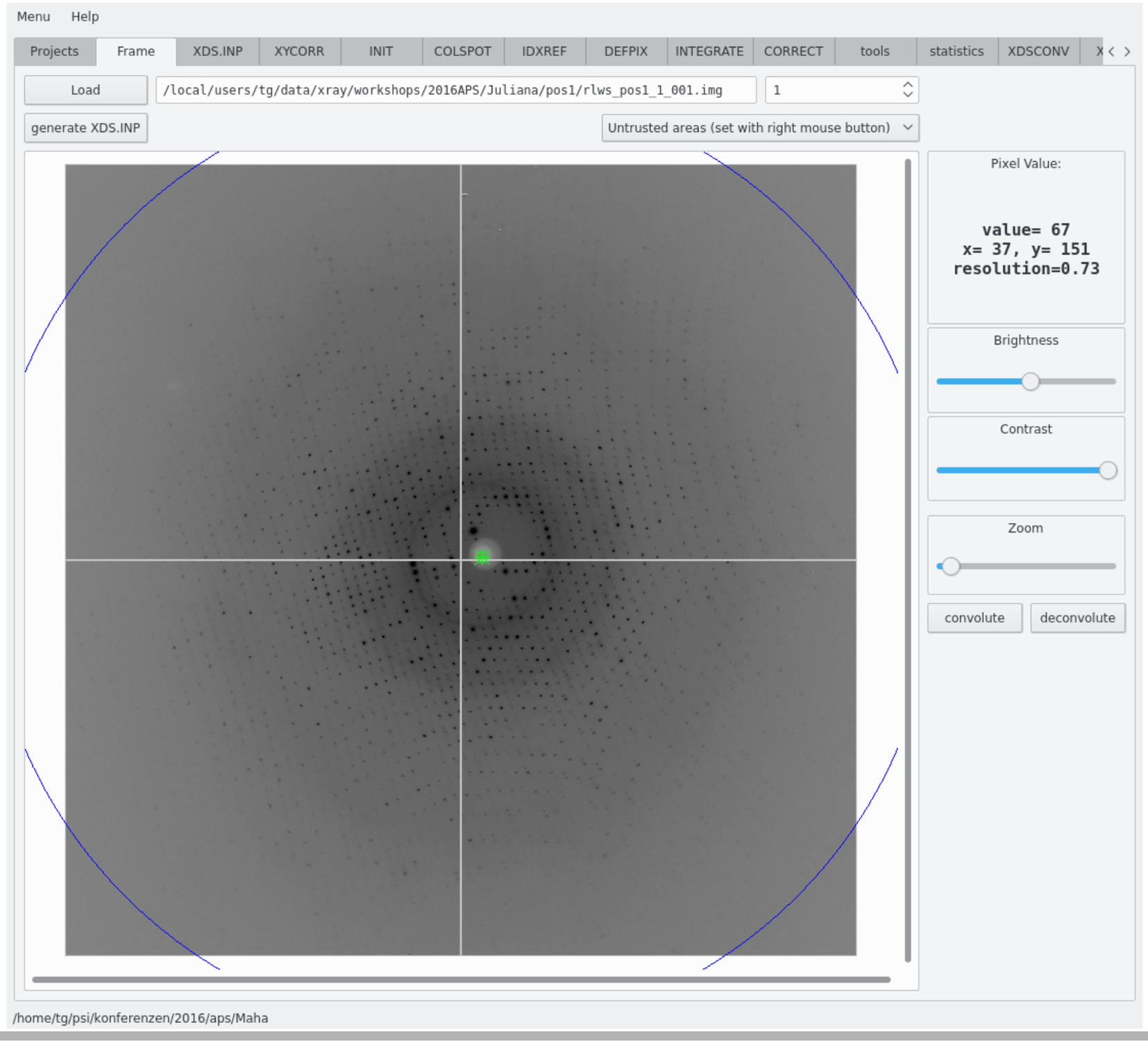
- IDXREF: Miller-Indices according to XPARM.XDS

```
   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
- less than 50% indexed reflections: XDS will stop with the error message

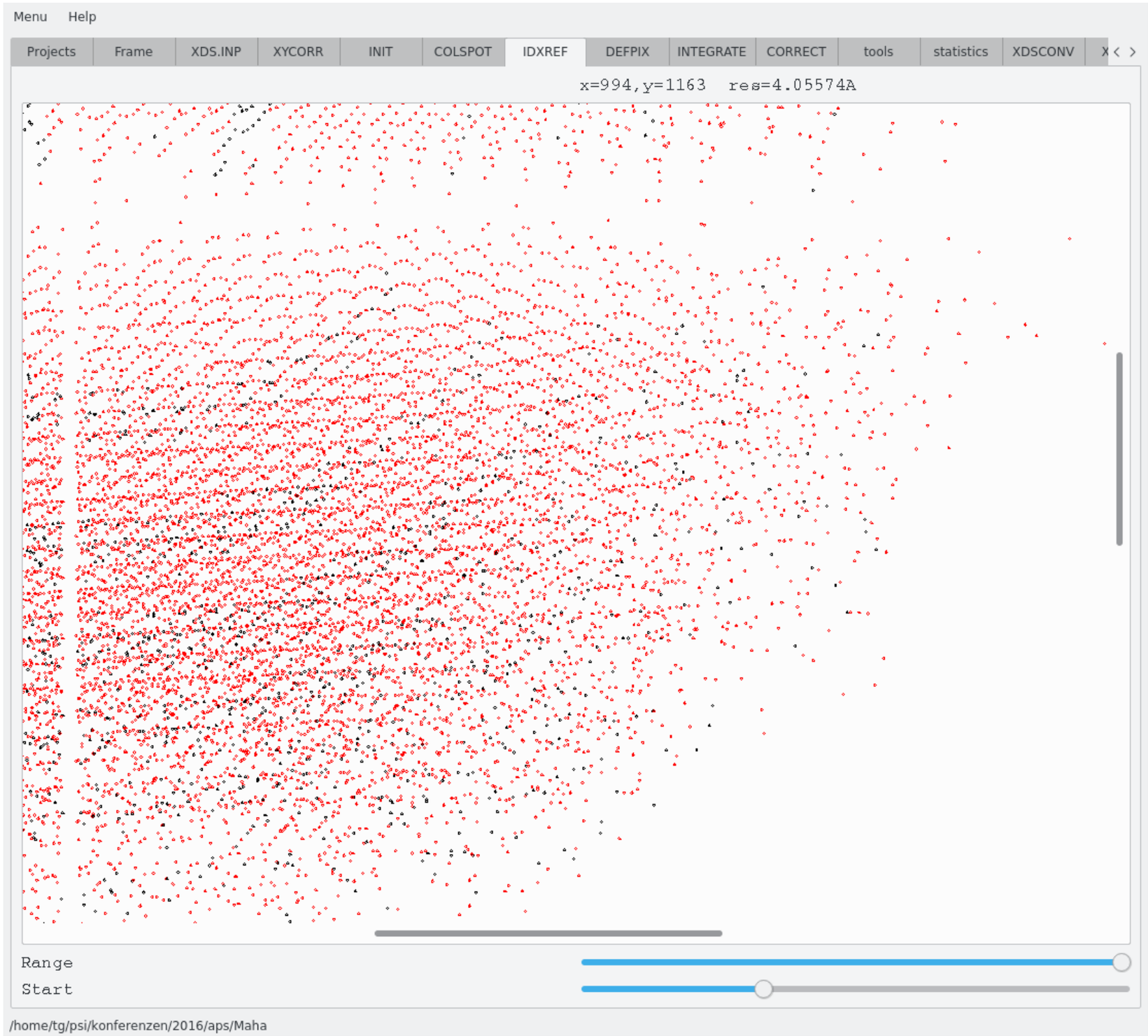
```
!!!  ERROR !!!  SOLUTION IS INACCURATE
```

3.9 - Example for difficult Indexing



- Very high resolution diffraction
- not visible here: Overloads
- not visible here: Ice rings

3.10 - Example for difficult Indexing



- Default: Indexing with all data
- Worms, not lines / lattice

3.11 - Example for difficult Indexing

Menu Help

Projects | Frame | XDS.INP | XYCORR | INIT | COLSPOT | IDXREF | DEFPIX | INTEGRATE | CORRECT | tools | statistics | XDSCONV

```
  0   1 -2      23.3     1.3    1066.6    1036.1 -0.0189   0.0291   1.5714     0.17     0.13     0.76

SELECTED:      INDEX_ORIGIN=   0   0   0


***** REFINED SOLUTION BASED ON INDEXED REFLECTIONS IN SUBTREE # 1 *****

REFINED VALUES OF DIFFRACTION PARAMETERS DERIVED FROM      1178 INDEXED SPOTS
REFINED PARAMETERS:   POSITION BEAM AXIS ORIENTATION CELL
STANDARD DEVIATION OF SPOT    POSITION (PIXELS)      0.62
STANDARD DEVIATION OF SPINDLE POSITION (DEGREES)     3.88
SPACE GROUP NUMBER      1
UNIT CELL PARAMETERS     37.025    48.123    74.695  90.007  89.881  89.986
REC. CELL PARAMETERS   0.027009  0.020780  0.013388  89.993  90.119  90.014
COORDINATES OF UNIT CELL A-AXIS    33.483    -7.243   -14.046
COORDINATES OF UNIT CELL B-AXIS    -2.233   -44.696    17.696
COORDINATES OF UNIT CELL C-AXIS   -31.551   -23.547   -63.478
CRYSTAL MOSAICITY (DEGREES)     0.200
LAB COORDINATES OF ROTATION AXIS  0.999604  0.008365 -0.026859
DIRECT BEAM COORDINATES (REC. ANGSTROEM)   0.017700 -0.067269  1.570293
DETECTOR COORDINATES (PIXELS) OF DIRECT BEAM    1074.98   1013.06
DETECTOR ORIGIN (PIXELS) AT                     1059.30   1072.65
CRYSTAL TO DETECTOR DISTANCE (mm)      142.43
LAB COORDINATES OF DETECTOR X-AXIS  1.000000  0.000000  0.000000
LAB COORDINATES OF DETECTOR Y-AXIS  0.000000  1.000000  0.000000


!!! WARNING !!! REFINEMENT DID NOT CONVERGE
                LAST CORRECTION SHIFT WAS   1.5E-02 (should be <  1.0E-03)

***** INDEXING OF OBSERVED SPOTS IN SPACE GROUP #   1 *****
    5889 OUT OF     43780 SPOTS INDEXED.
     107 REJECTED REFLECTIONS (REASON: OVERLAP)
   37784 REJECTED REFLECTIONS (REASON: TOO FAR FROM IDEAL POSITION)
EXPECTED ERROR IN SPINDLE  POSITION     1.108 DEGREES
EXPECTED ERROR IN DETECTOR POSITION      2.20 PIXELS


***** DIFFRACTION PARAMETERS USED AT START OF INTEGRATION *****

REFINED VALUES OF DIFFRACTION PARAMETERS DERIVED FROM      5889 INDEXED SPOTS
```

/home/tg/psi/konferenzen/2016/aps/Maha

- Indexing fails
- Poorly defined rotation axis (*cf.* worms)
- Detector distance refined far from original 110mm
- Only 13% indexed reflections

3.12 - Example for difficult Indexing

Menu Help

Projects | Frame | XDS.INP | XYCORR | INIT | COLSPOT | IDXREF | DEFPIX | INTEGRATE | CORRECT | tools | statistics | XDSCONV | XSCALE

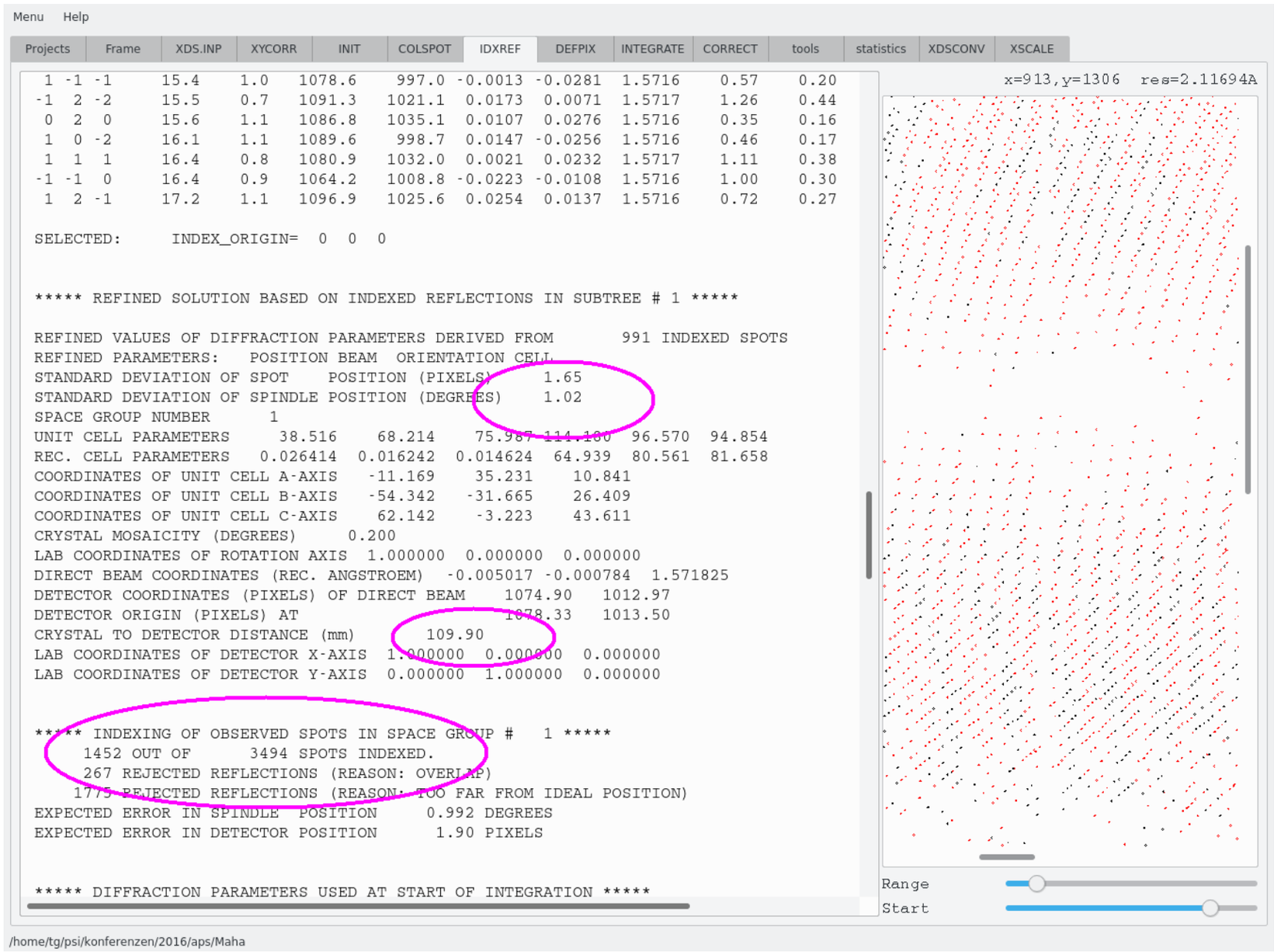
Save | Run XDS | stop xds

```
! written by generate_XDS.INP version 0.73 (18-May-2017)
JOB=  IDXREF !DEFPIX INTEGRATE CORRECT
ORGX= 1079.5 ORGY= 1016.2  ! check these values with adxv !
! Following are possible beam center interpretations for ADSC detectors
! ORGX= 1031.8 ORGY= 1079.5 ! For ESRF, PF, APS Argonne ...
! ORGX= 968.5 ORGY= 1031.8 ! For ALS 5.0.3,..
! ORGX= 1079.5 ORGY= 1016.2 ! For ALS 8.2.2,..
! ORGX= 1079.5 ORGY= 1031.8 ! For SPring-8,..
DETECTOR_DISTANCE= 110.110
OSCILLATION_RANGE= 0.500
X-RAY_WAVELENGTH= 0.63620
NAME_TEMPLATE_OF_DATA_FRAMES=/local/users/tg/data/xray/workshops/2016APS/Juliana/pos1/rlws_pos1_1_???.img
! REFERENCE_DATA_SET=xxx/XDS_ASCII.HKL ! e.g. to ensure consistent indexing
DATA_RANGE=1 300
SPOT_RANGE=100 120
! BACKGROUND_RANGE=1 10 ! rather use defaults (first 5 degree of rotation)

SPACE_GROUP_NUMBER=0                  ! 0 if unknown
UNIT_CELL_CONSTANTS= 70 80 90 90 90 90 ! put correct values if known
INCLUDE_RESOLUTION_RANGE=1.8 0  ! after CORRECT, insert high resol limit; re-run CORRECT
! IDXREF now obeys INCLUDE_RESOLUTION_RANGE and EXCLUDE_RESOLUTION_RANGE to exclude ice-rings

FRIEDEL'S_LAW=FALSE     ! This acts only on the CORRECT step
! If the anom signal turns out to be, or is known to be, very low or absent,
! use FRIEDEL'S_LAW=TRUE instead (or comment out the line); re-run CORRECT

! remove the "!" in the following line:
! STRICT_ABSORPTION_CORRECTION=TRUE
! if the anomalous signal is strong: in that case, in CORRECT.LP the three
! "CHI^2-VALUE OF FIT OF CORRECTION FACTORS" values are significantly> 1, e.g. 1.5
!
! exclude (mask) untrusted areas of detector, e.g. beamstop shadow :
! UNTRUSTED_RECTANGLE= 1800 1950 2100 2150 ! x-min x-max y-min y-max ! repeat
! UNTRUSTED_ELLIPSE= 2034 2070 1850 2240 ! x-min x-max y-min y-max ! if needed
! UNTRUSTED_QUADRILATERAL= x1 y1 x2 y2 x3 y3 x4 y4 ! see documentation
!
! parameters with changes wrt default values:
TRUSTED_REGION=0.0 1.2 ! partially use corners of detector (0 1.4143: use all pixels)
VALUE_RANGE_FOR_TRUSTED_DETECTOR_PIXELS=6000. 30000. ! often 7000 or 8000 is ok
STRONG_PIXEL=4           ! COLSPOT: only use strong reflections (default is 3)
```

/home/tg/psi/konferenzen/2016/aps/Maha

- Index only on thin wedge (reduce effect of wrong geometrical parameters)
- Only use high resolution data ($d_{\text{min}} < 1.8$Å)
- High order reflections not affected by ice rings, overloads

3.13 - Example for difficult Indexing



- Better defined rotation axis
- Detector distance refines close to input (110mm → 109mm)
- relative (!!) number of indexed reflections increased to 41%

3.14 - DEFPIX: active detector mask

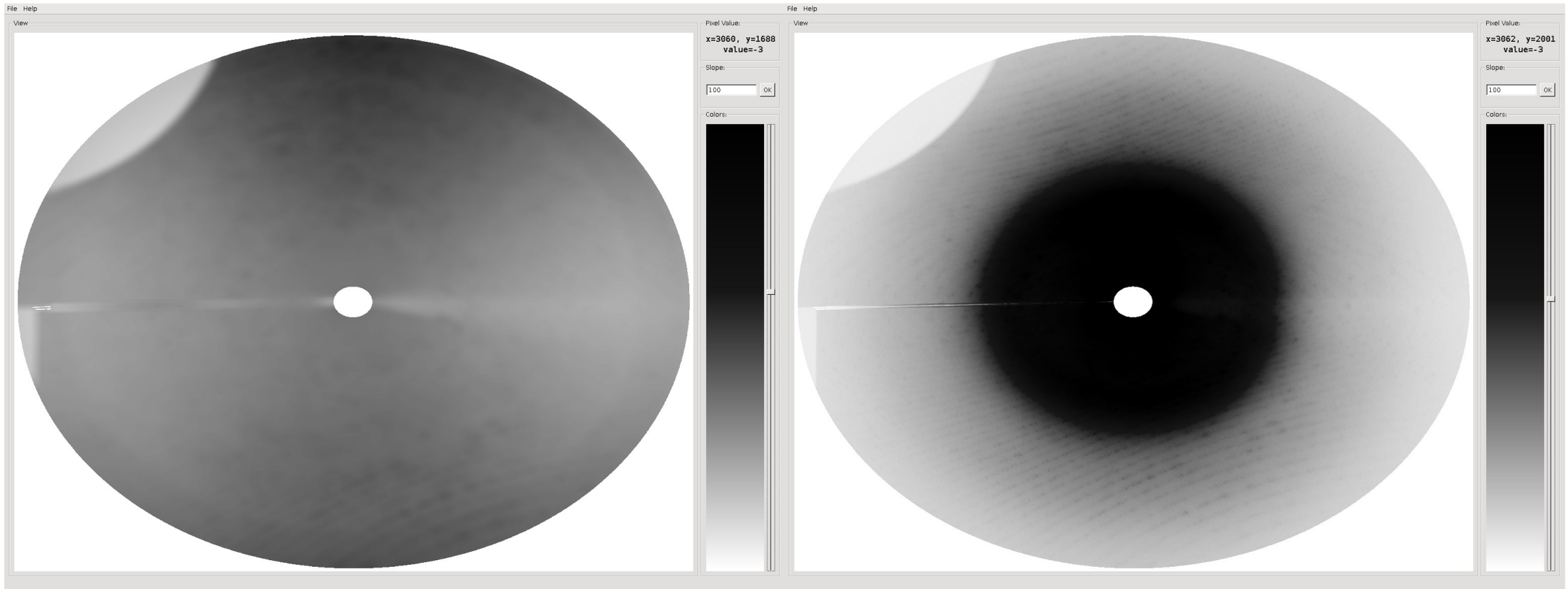
DEFPIX sets the area of the detector which is taken into account during integration. It takes into account:

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

3.15 - VALUE_RANGE_FOR_TRUSTED_DETECTOR_PIXELS

ABS.cbf

BKGPIX.cbf, all included

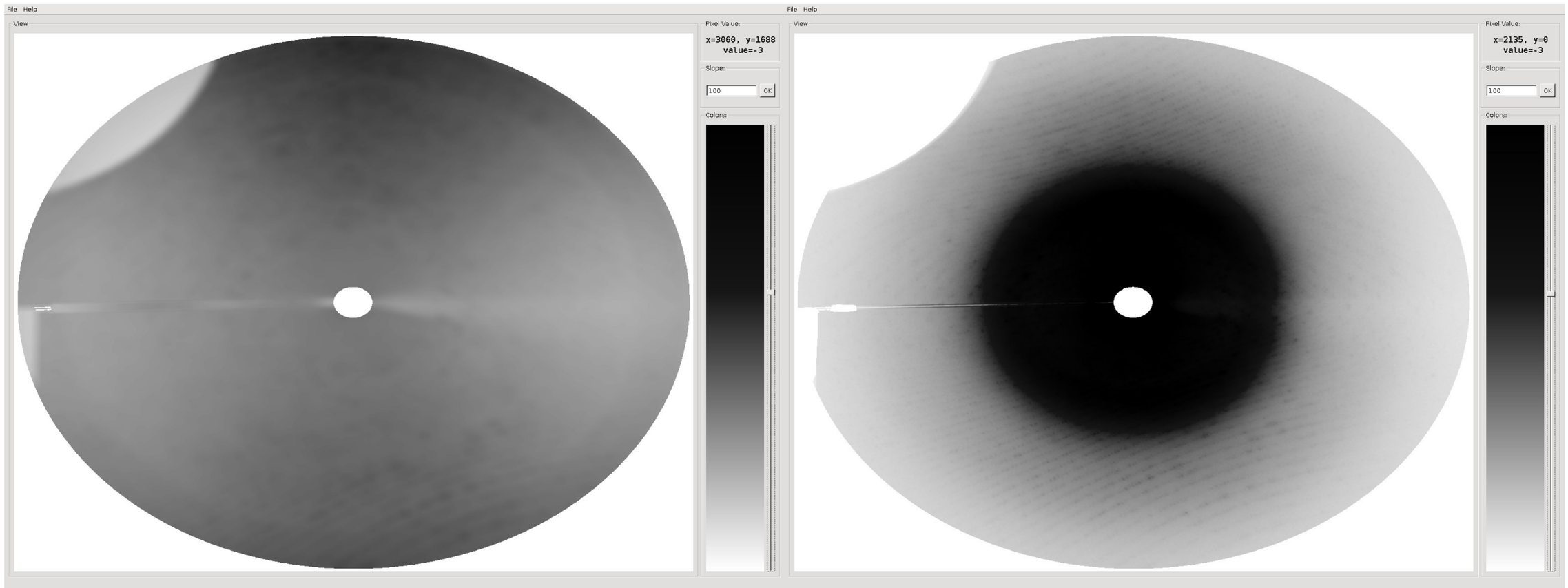


VALUE_RANGE_FOR_TRUSTED_DETECTOR_PIXELS= 1000 30000

3.16 - VALUE_RANGE_FOR_TRUSTED_DETECTOR_PIXELS

ABS.cbf

BKGPIX.cbf, shadows removed



VALUE_RANGE_FOR_TRUSTED_DETECTOR_PIXELS= 6100 30000

3.17 - Recycling

- `CORRECT` writes improved geometrical description to `GXPARM.XDS`
- 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 when symmetry is taken into account
2. Proper **resolution cut–off** can improve data quality
3. Unit cell parameters improve

3.18 - 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.

3.19 - High Resolution Cut-Off

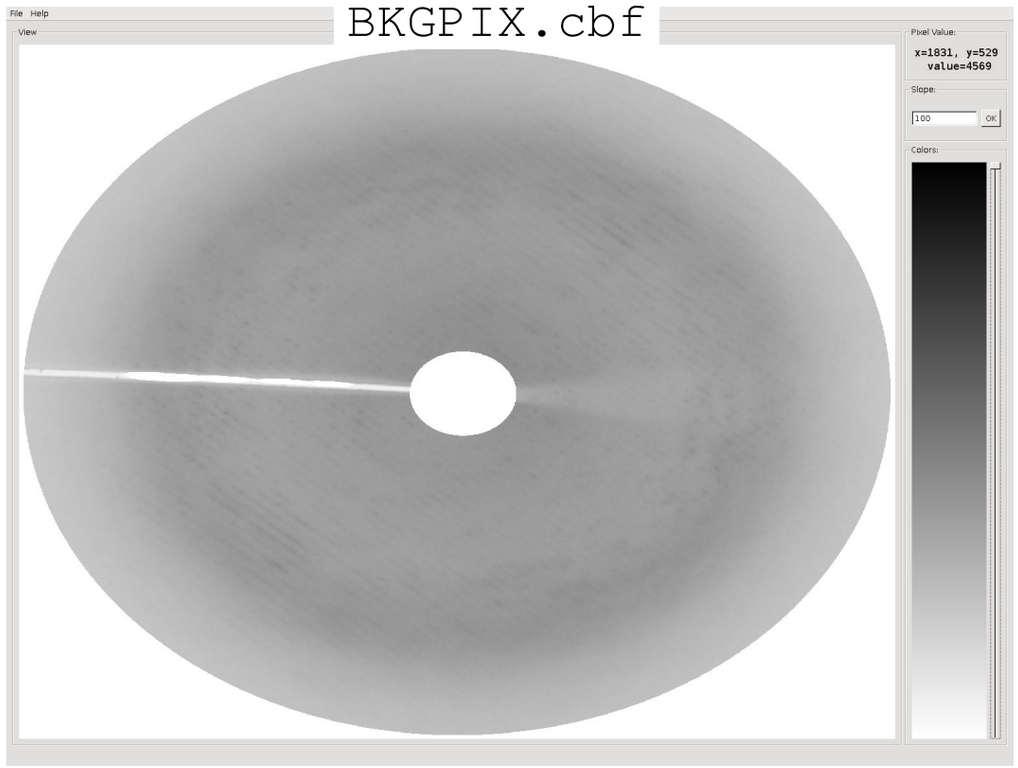
SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

	RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	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*

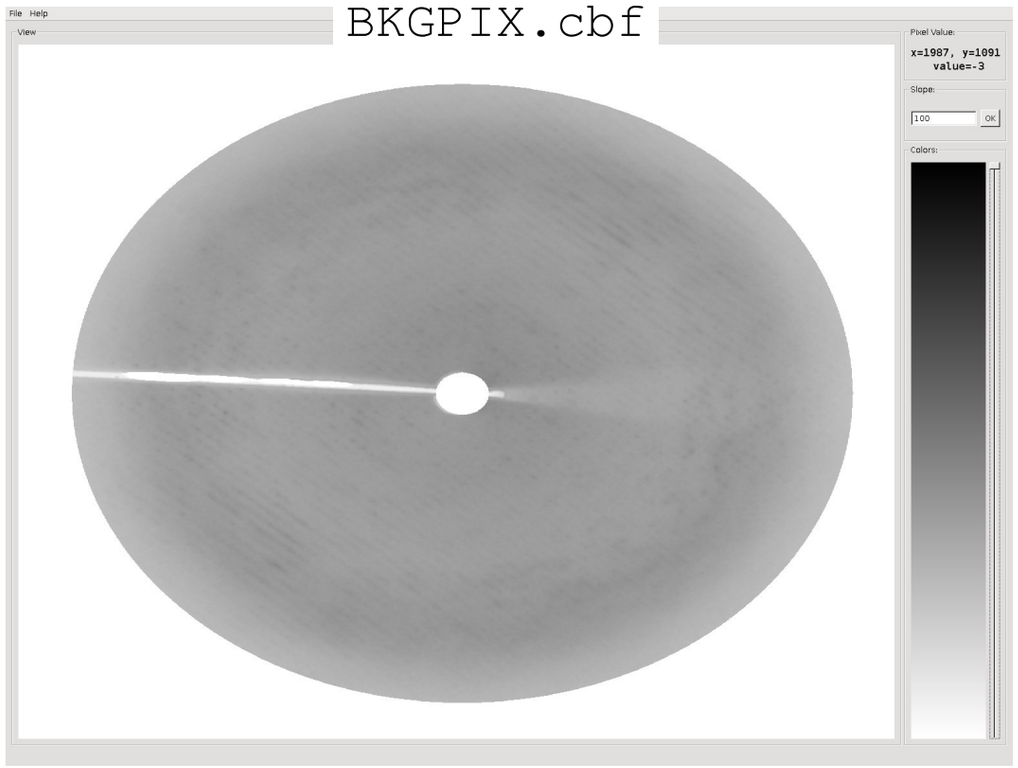
Logfile `CORRECT.LP`

- Modern cut–off: CC1/2 marked by '*': 0.1% significance level
- Conservative cut–off: $I/\sigma_I \geq 2.0$ — should correspond to $CC(1/2) \geq 70\%$
- Personal opinion: use $I/\sigma_I \geq 2.0$ for **reported resolution**

3.20 - 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)

3.21 - 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.

3.22 - 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
```

3.23 - 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%

3.24 - 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