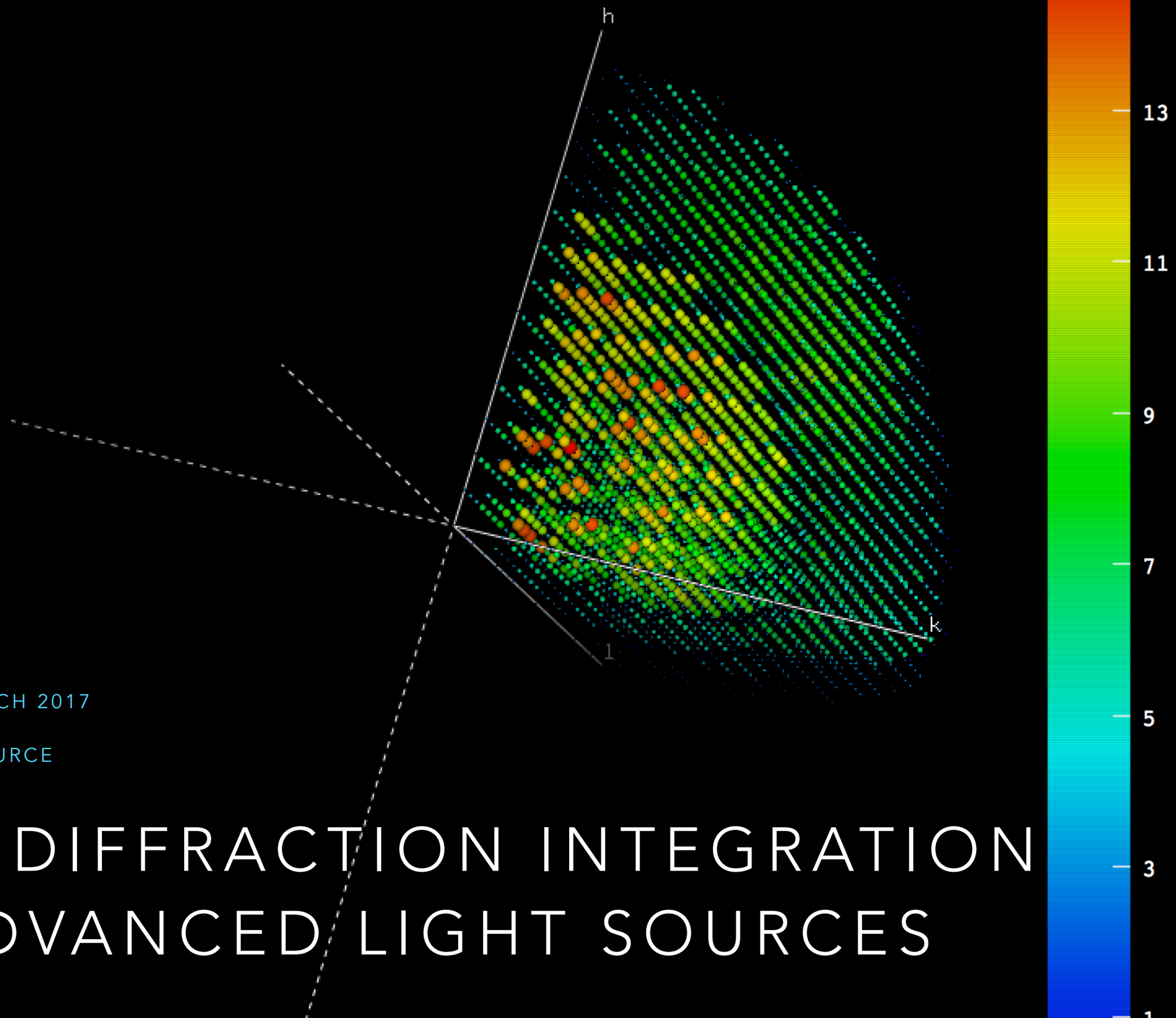
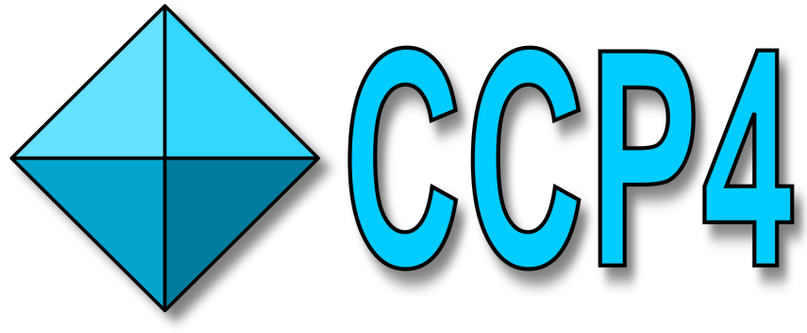


HDRMX 4 LUND MARCH 2017
GRAEME WINTER
DIAMOND LIGHT SOURCE

DIALS: DIFFRACTION INTEGRATION FOR ADVANCED LIGHT SOURCES

PROJECT SUPPORT



EU FP7: #283570 NIH: GM095887 & GM102520 WT: 202933

OPEN SOURCE

dials / **dials**

Unwatch 17 | Star 9 | Fork 3

Code | Issues 77 | Pull requests 0 | Projects 0 | Wiki | Pulse | Graphs | Settings

Diffraction Integration for Advanced Light Sources https://dials.github.io

Edit

crystallography Manage topics

7,494 commits | **12** branches | **7** releases | **18** contributors

Branch: **master** | New pull request | Create new file | Upload files | Find file | Clone or download

jmp1985 Now importing dxtbx.model.experiment.experiment_list from ... Latest commit 9abb12b 5 days ago

algorithms	Now importing dxtbx.model.experiment.experiment_list from	a day ago
array_family	Now importing dxtbx.model.experiment.experiment_list from	a day ago
command_line	Now importing dxtbx.model.experiment.experiment_list from	a day ago
data	More absolute imports, #292	2 months ago
doc	Make installation instructions less wrong. Partial fix for #312	4 days ago
extensions	Move some locations to absolute imports, #292	2 months ago
framework	Move some locations to absolute imports, #292	2 months ago
installer	Remaining conversions to absolute imports. Closes #292	2 months ago
interfaces	Move some locations to absolute imports, #292	2 months ago


DIALS East
DIALS West
Nunavut Territory
ARCTIC CIRCLE
CANADA
Rocky Mountains
Hudson Bay
LABRADOR SEA
Island of Newfoundland
Nova Scotia
ICELAND
Reykjavik
Denmark Strait
NORWAY
SWEDEN
Oslo
Stockholm
NORTH SEA
UNITED KINGDOM
London
Dublin
IRELAND
DENMARK
Copenhagen
Berlin
Amsterdam
GERMANY
POLAND
NETH.
BEL.
LUX.
CZECH REP.
Paris
FRANCE
AUSTRIA
ITALY
Rome
PORTUGAL
Lisbon
SPAIN
Madrid
Algiers
Tunis
TUNISIA
Tripoli
LIBYA
Rabat
MOROCCO
ALGERIA
WESTERN SAHARA
MAURITANIA
MALI
NIGER
NIGERIA
Abuja
CAMEROON
GHANA
TOGO
BENIN
LIBERIA
SIERRA LEONE
GUINEA
SENEGAL
Dakar
CAPE VERDE
UNITED STATES OF AMERICA
Ottawa
Washington
New York
Chicago
San Francisco
Los Angeles
Seattle
Denver
Dallas
Houston
Miami
NORTH ATLANTIC OCEAN
TROPIC OF CANCER
Gulf of Mexico
MEXICO
Mexico
BAHAMAS
Havana
CUBA
HAITI
DOMINICAN REP.
JAMAICA
BELIZE
HONDURAS
GUATEMALA
EL SALVADOR
NICARAGUA
COSTA RICA
PANAMA
CARIBBEAN SEA
VENEZUELA
Caracas
COLOMBIA
Bogota
GUYANA
SURINAME
FRENCH GUIANA (France)
ECUADOR
Quito
PERU
Lima
BRAZIL
Amazon Basin
EQUATOR
Gulf of Guinea
GABON
CONGO
ANGOLA

CORE DEVELOPMENT TEAMS

- DIALS East: PI Gwyndaf Evans; David Waterman, Richard Gildea, James Parkhurst, Markus Gerstel, Luis Fuentes Montero, Graeme Winter - synchrotron rotation data + 2 shortly
- DIALS West: PI Nick Sauter; Aaron Brewster, Iris Young - X-FEL serial crystallography

ACKNOWLEDGEMENTS

research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

XDS

Wolfgang Kabsch

Max-Planck-Institut für Medizinische Forschung, Abteilung Biophysik, Jahnstrasse 29, 69120 Heidelberg, Germany

Correspondence e-mail: wolfgang.kabsch@mpimf-heidelberg.mpg.de

Received 19 August 2009
Accepted 9 November 2009

A version of this paper will be published as a chapter in the new edition of Volume *F* of *International Tables for Crystallography*.

The usage and control of recent modifications of the program package *XDS* for the processing of rotation images are described in the context of previous versions. New features include automatic determination of spot size and reflecting range and recognition and assignment of crystal symmetry. Moreover, the limitations of earlier package versions on the number of correction/scaling factors and the representation of pixel contents have been removed. Large program parts have been restructured for parallel processing so that the quality and completeness of collected data can be assessed soon after measurement.

1. Functional specification

research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The finer things in X-ray diffraction data collection

J. W. Pflugrath

Molecular Structure Corporation, 9009 New Trails Drive, The Woodlands, TX 77381, USA

Correspondence e-mail: jwp@msc.com

Received 6 May 1999
Accepted 5 July 1999

X-ray diffraction images from two-dimensional position-sensitive detectors can be characterized as thick or thin, depending on whether the rotation-angle increment per image is greater than or less than the crystal mosaicity, respectively. The expectations and consequences of the processing of thick and thin images in terms of spatial overlap, saturated pixels, X-ray background and $I/\sigma(I)$ are discussed. The *d*TREK* software suite for processing diffraction images is briefly introduced, and results from *d*TREK* are compared with those from another popular package.

1. Introduction

Two-dimensional position-sensitive detectors have been used for many years in X-ray diffraction data collection. In particular, data from crystals of macromolecules such as proteins, oligonucleotides and their complexes are almost always acquired with an area detector such as film (now virtually obsolete), a multi-wire system, an imaging plate or the recently commercialized charge-coupled device (CCD) coupled to a phosphor-coated fiber-optic taper. With all these detectors, the crystal, centered in the X-ray beam, is rotated or oscillated around a single axis through a small angle of 0.1 to ~2.0°, while counts from diffracted photons are accumulated for a specified time. At the end of the small rotation, the detector is read out and the counts are stored as an image: a two-dimensional array with each array element (pixel) related to a distinct position on the detector and the number of

research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The integration of macromolecular diffraction data

Andrew G. W. Leslie

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

Correspondence e-mail: andrew@mrc-lmb.cam.ac.uk

Received 19 May 2005
Accepted 24 November 2005

The objective of any modern data-processing program is to produce from a set of diffraction images a set of indices (*hkl*s) with their associated intensities (and estimates of their uncertainties), together with an accurate estimate of the crystal unit-cell parameters. This procedure should not only be reliable, but should involve an absolute minimum of user intervention. The process can be conveniently divided into three stages. The first (autoindexing) determines the unit-cell parameters and the orientation of the crystal. The unit-cell parameters may indicate the likely Laue group of the crystal. The second step is to refine the initial estimate of the unit-cell

Centre National de la Recherche Scientifique
Université Paris-Sud

Laboratoire pour l'Utilisation du Rayonnement Electromagnétique

Proceedings

of the EEC Cooperative Workshop

on Position-Sensitive Detector Software

(Phases I & II)

WHAT IS DIALS?

- Indexing, refinement and integration programs
- Scope similar to MOSFLM i.e. does *not* include scaling - use POINTLESS/AIMLESS for scaling (work in progress)
- Integration algorithm similar to XDS
- Does *not* currently include a GUI (in development) though does include a range of visual analysis tools
- Also DIALS is a toolkit, so you can do lots more

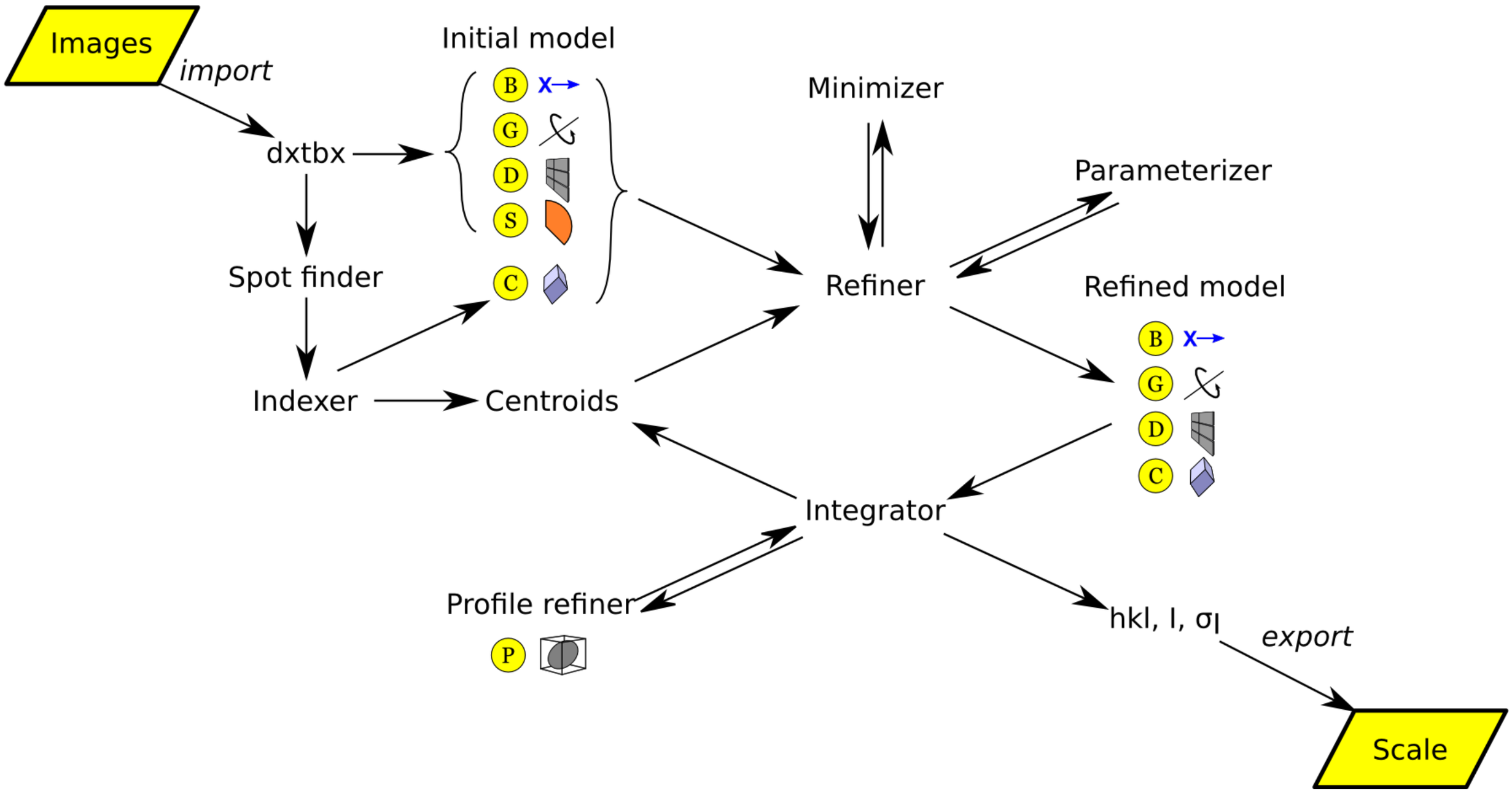
PHILOSOPHY

- Global analysis and model - take data from every image in your data set
- No assumptions made about experimental geometry - completely general / vectorial description
- Treat data as 3D volume not 2D “frames”
- Toolkit - for data processing but you can do fun experiments with a little code

WORKFLOW

- import - read headers, construct model of experiment
- find_spots - find spots on every image; 3D if rotation data
- index - index the reflection list (may be > 1 sweep) in P1 & perform static refinement
- refine_bravais_settings - consider / refine indexed spots in all possible lattices
- reindex - reindex the indexed spot list
- refine - actually refine the model, usually with scan varying model
- integrate - actually integrate the data
- export_mtz - export an MTZ file for scaling with pointless & aimless (CCP4)

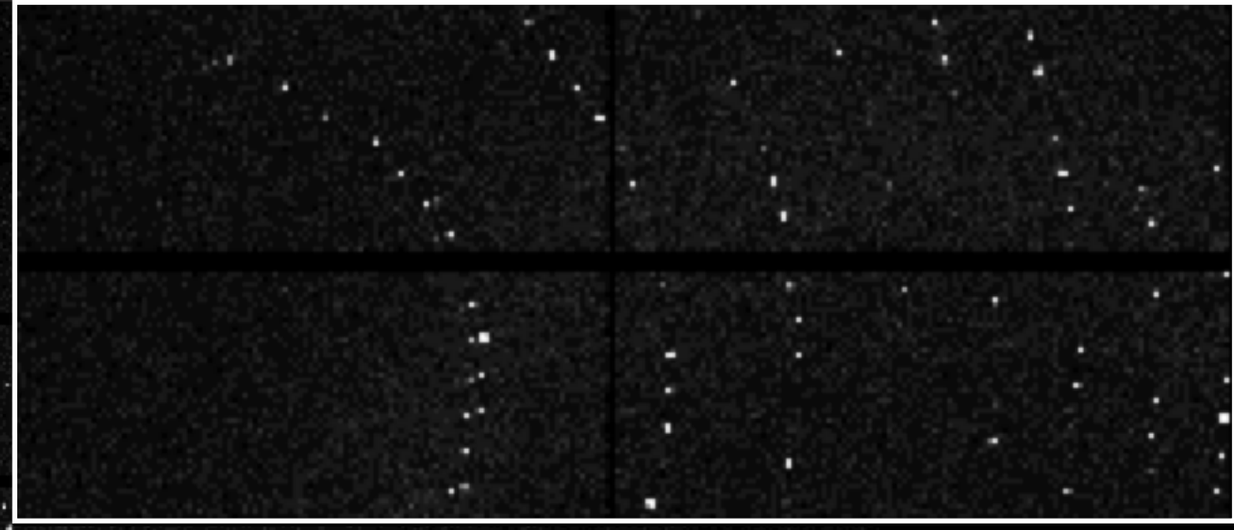
DIALS


Images
import
dxtbx
Initial model
B
G
D
S
C
Minimizer
Parameterizer
Refiner
Refined model
Spot finder
Indexer
Centroids
Integrator
Profile refiner
P
hkl, I, σI
export
Scale

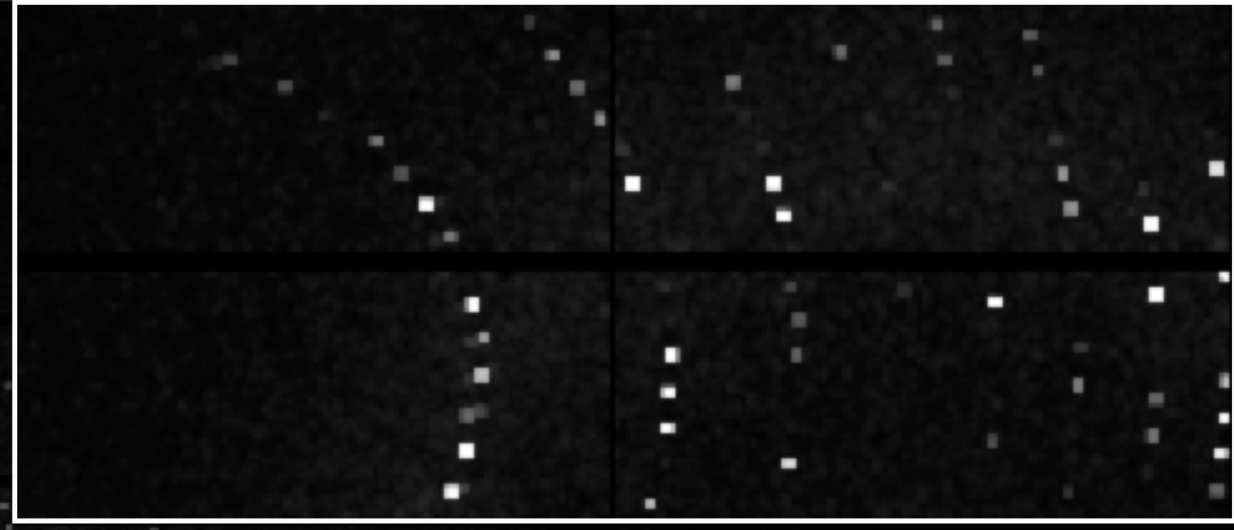
ALGORITHMS

SPOT FINDING

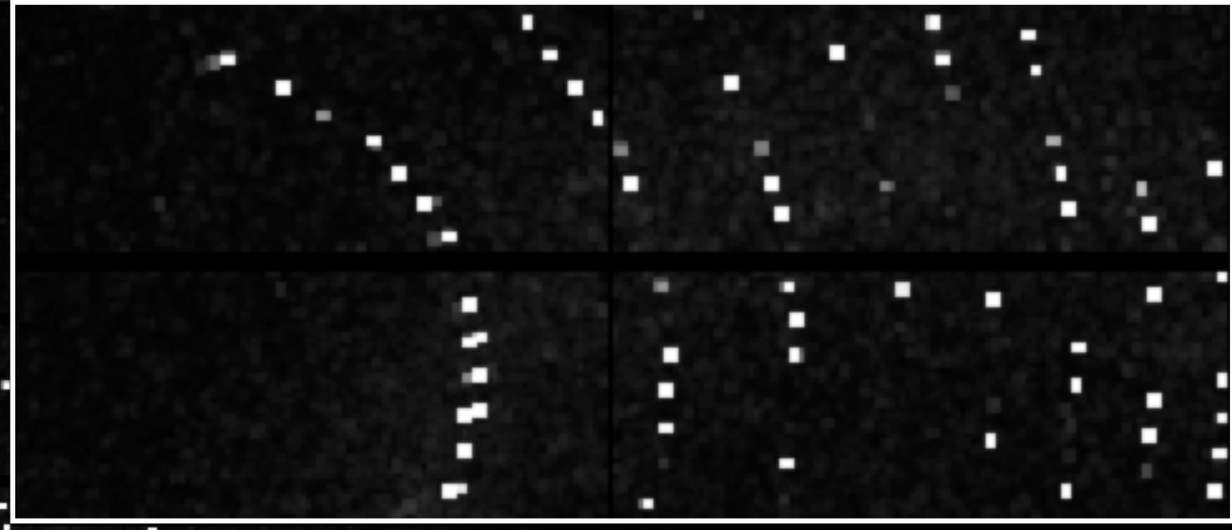
SPOT FINDING



SPOT FINDING

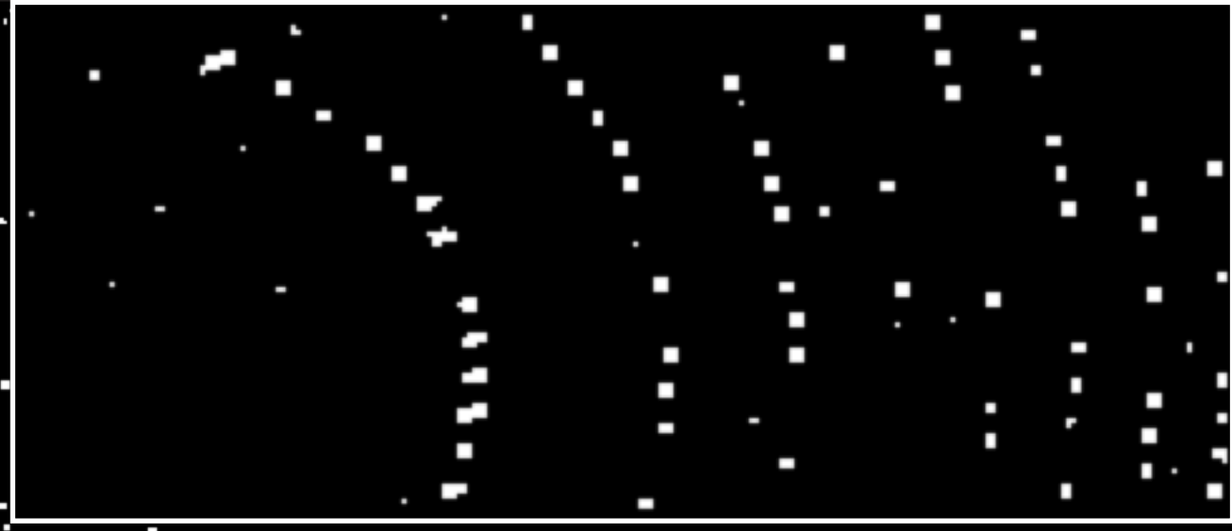


SPOT FINDING

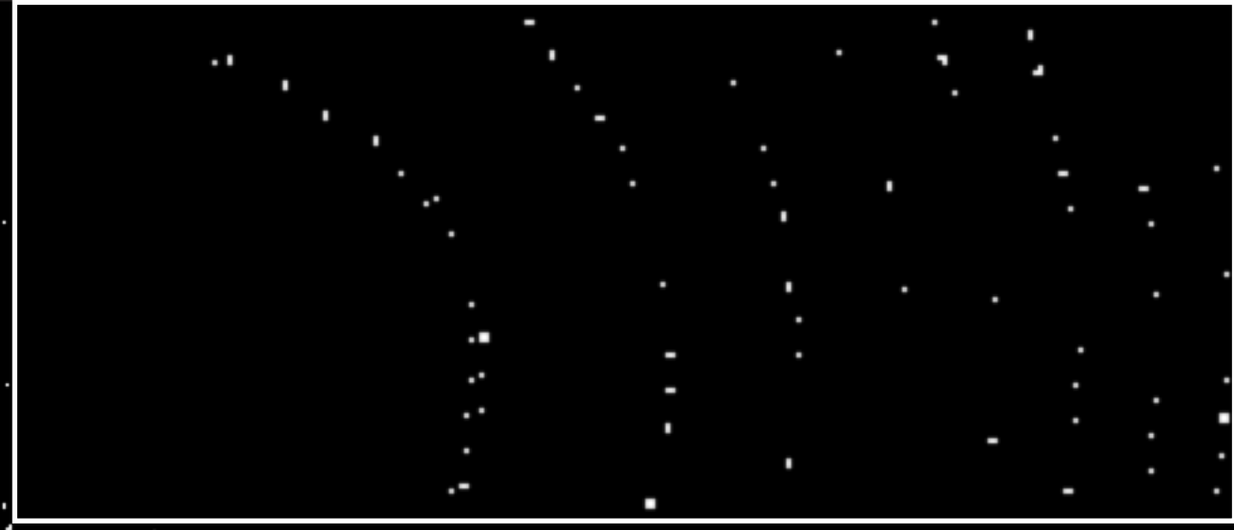


SPOT FINDING

SPOT FINDING



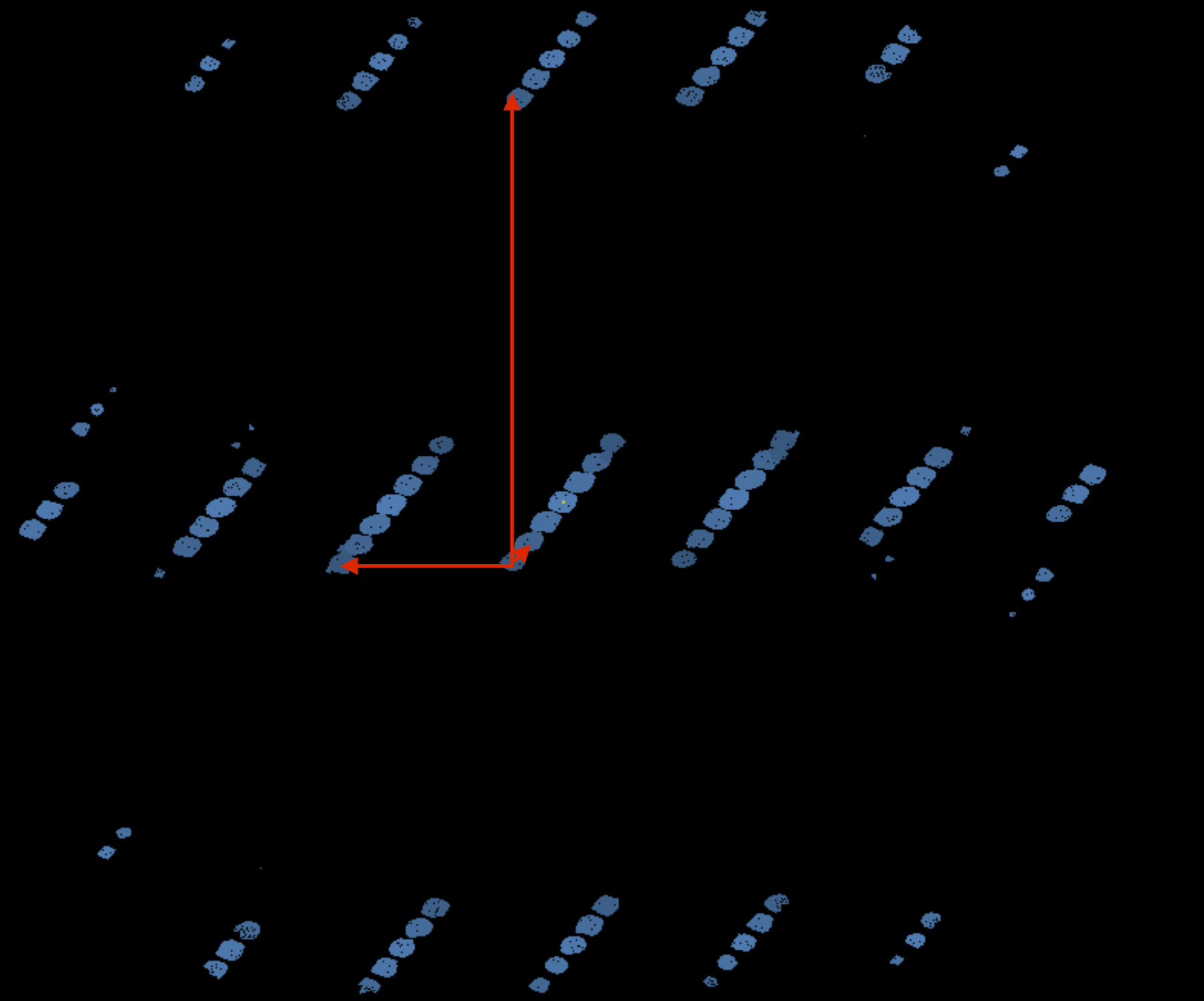
SPOT FINDING



INDEXING

INDEXING

INDEXING



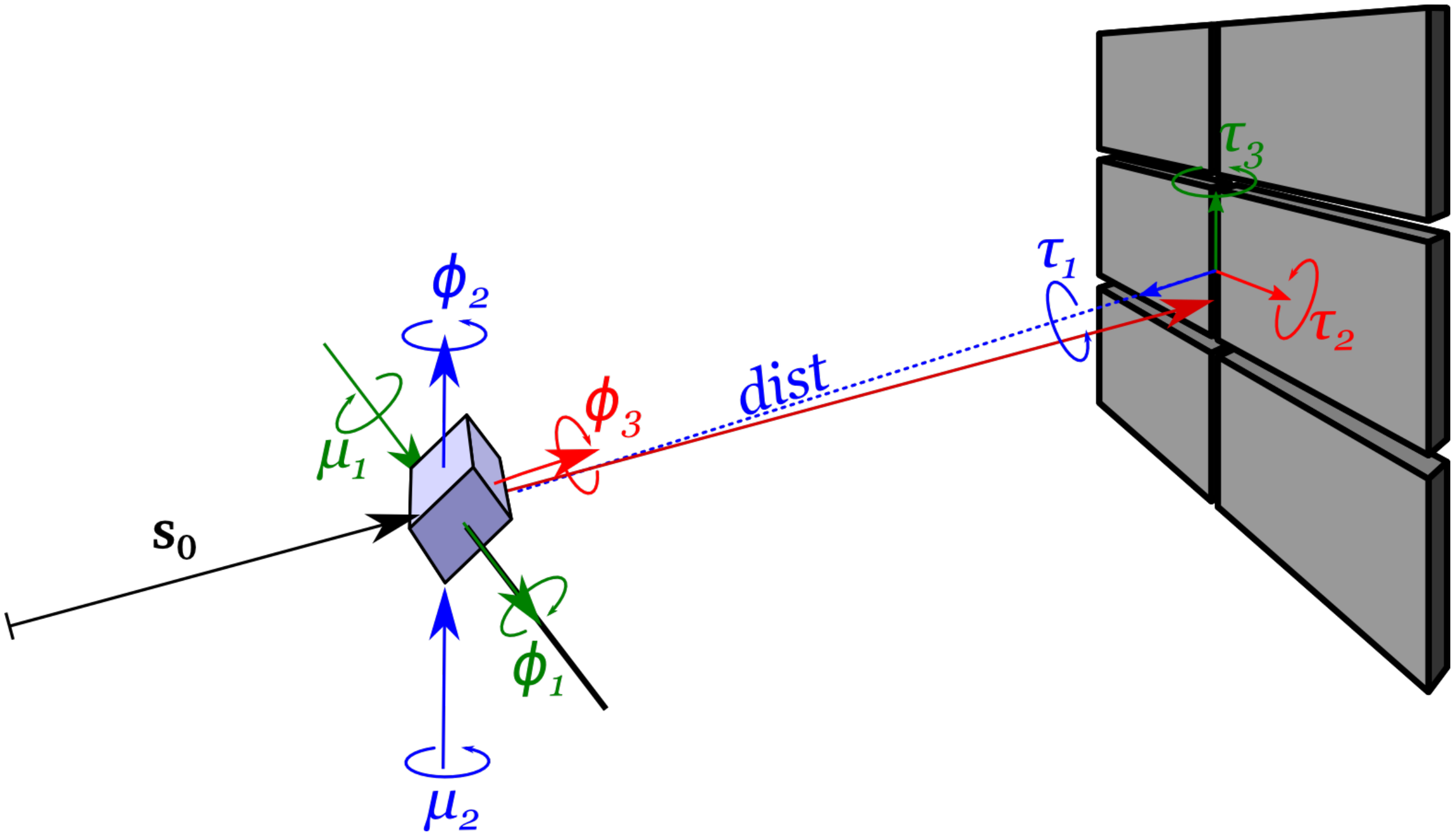
SOLUTION PICKING

Solution	Metric fit	rmsd	min/max cc	#spots	lattice	unit_cell						volume	cb_op
9	1.1465	0.655	0.490/0.650	18000	tI	79.27	79.27	68.72	90.00	90.00	90.00	431793	a+c,a+b,b+c
8	1.1465	0.665	0.487/0.650	18000	oF	68.75	112.00	112.28	90.00	90.00	90.00	864603	b+c,-b+c,2*a+b+c
7	1.1465	0.540	0.487/0.487	18000	mI	65.34	112.03	66.09	90.00	117.08	90.00	430757	a,-b+c,-a-b-c
6	1.1330	0.544	0.493/0.493	18000	mI	64.98	111.74	65.74	90.00	116.83	90.00	425978	-b,-2*a-b-c,-c
5	1.0311	0.392	0.650/0.884	18000	oI	68.51	78.60	79.44	90.00	90.00	90.00	427759	b+c,a+c,a+b
4	1.0310	0.395	0.716/0.716	18000	mI	68.52	78.62	79.46	90.00	89.95	90.00	428065	-b-c,a+c,-a-b
3	1.0311	0.417	0.650/0.650	18000	mI	78.58	68.52	79.43	90.00	90.19	90.00	427706	a+c,b+c,-a-b
2	0.0165	0.056	0.884/0.884	18000	mI	67.96	79.16	77.97	90.00	91.03	90.00	419397	b+c,a+b,-a-c
1	0.0000	0.055	-/-	18000	aP	64.76	64.74	65.48	117.10	106.47	104.64	209642	a,b,c

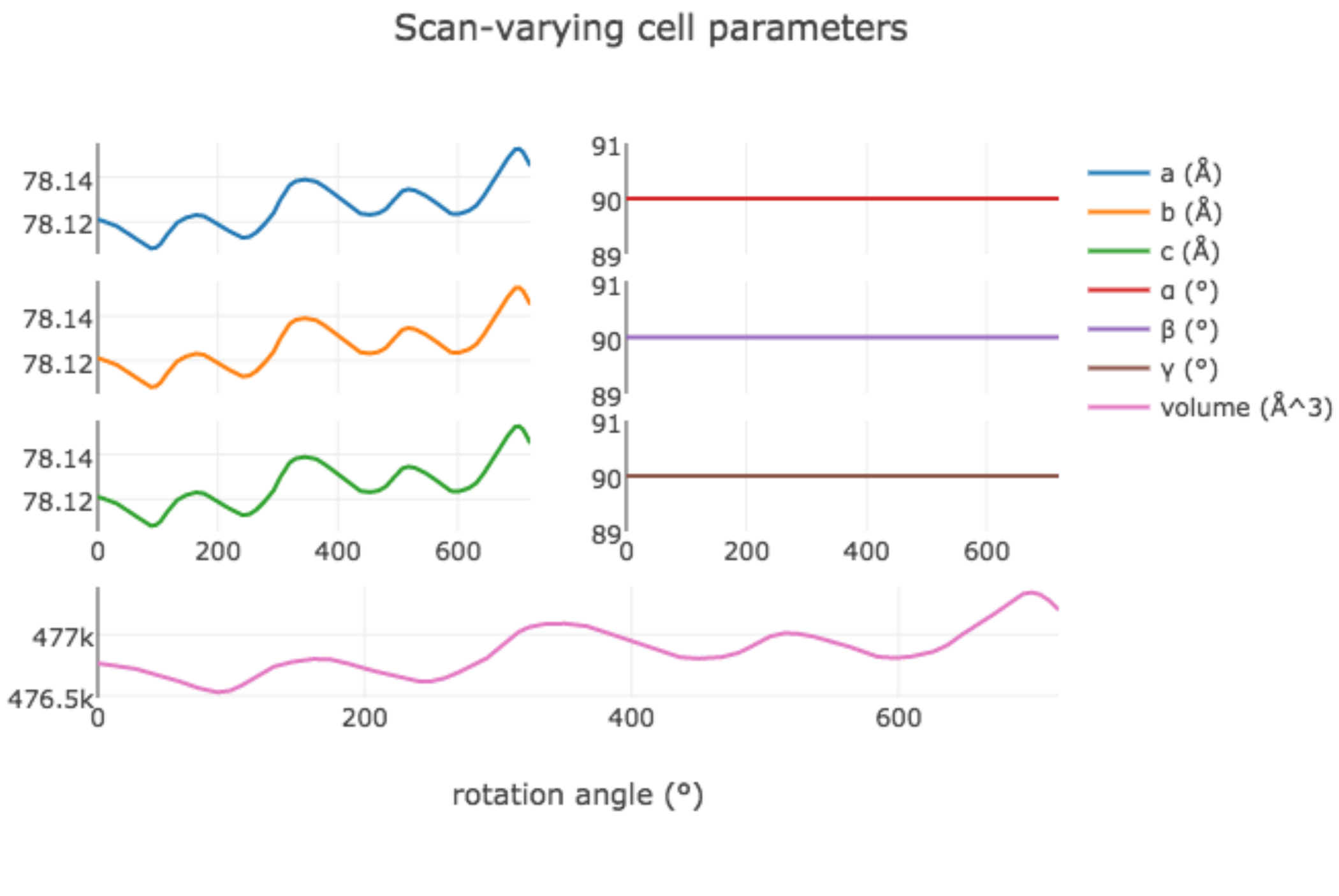
SOLUTION PICKING

Solution	Metric fit	rmsd	min/max cc	#spots	lattice		
9	1.1465	0.655	0.490/0.650	18000	tI	79.27	7
8	1.1465	0.665	0.487/0.650	18000	oF	68.75	11
7	1.1465	0.540	0.487/0.487	18000	mI	65.34	11
6	1.1330	0.544	0.493/0.493	18000	mI	64.98	11
5	1.0311	0.392	0.650/0.884	18000	oI	68.51	7
4	1.0310	0.395	0.716/0.716	18000	mI	68.52	7
3	1.0311	0.417	0.650/0.650	18000	mI	78.58	6
2	0.0165	0.056	0.884/0.884	18000	mI	67.96	7
1	0.0000	0.055	-/-	18000	aP	64.76	6

REFINEMENT



REFINEMENT


Scan-varying cell parameters
78.14
78.12
78.14
78.12
78.14
78.12
0
200
400
600
91
90
89
91
90
89
0
200
400
600
477k
476.5k
0
200
400
600
a (Å)
b (Å)
c (Å)
α (°)
β (°)
γ (°)
volume (Å^3)
rotation angle (°)

INTEGRATION

3.4. Intensity estimation

If an expected intensity distribution $\{p_i | i \in D_0\}$ of the observed profile is given in a domain D_0, the reflection intensity I can be estimated as

$$I = \frac{\sum_{i \in D} (c_i - b_i) p_i / v_i}{\sum_{i \in D} p_i^2 / v_i},$$

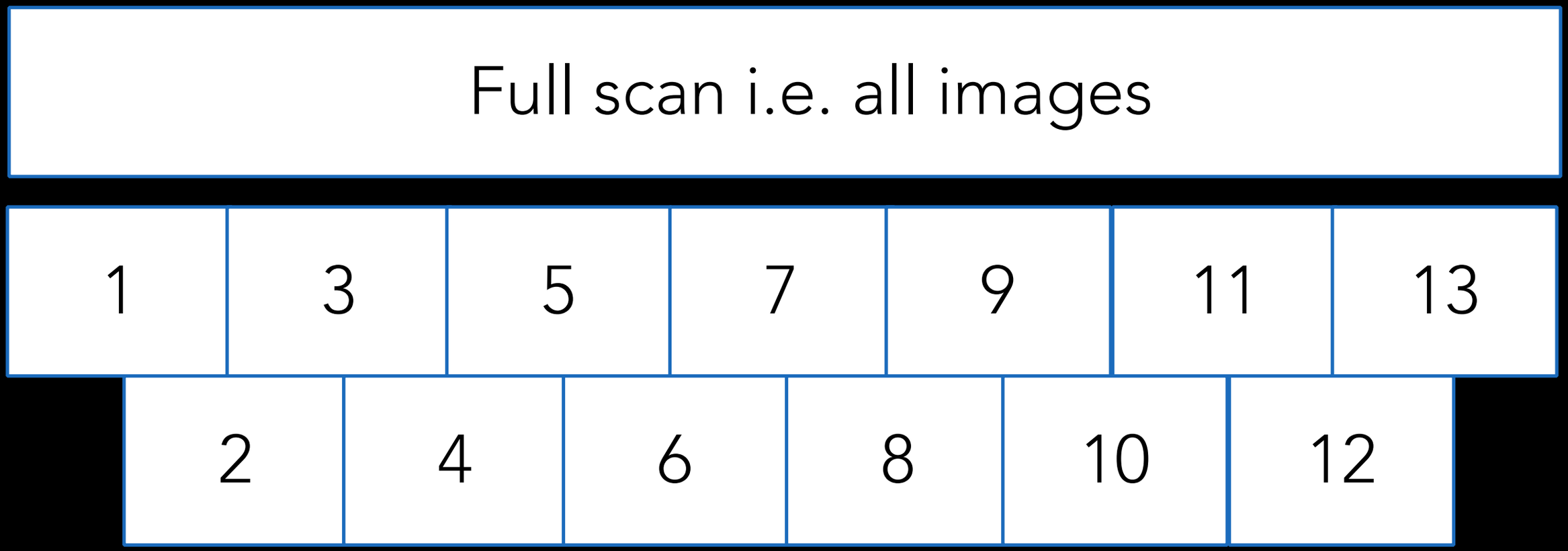
which minimizes the function

$$\psi(I) = \sum_{i \in D} (c_i - I \cdot p_i - b_i)^2 / v_i, \quad \sum_{i \in D_0} p_i = 1.$$

b_i, c_i, v_i $(i \in D)$ are the background, contents and variance of pixels observed in a subdomain $D \subseteq D_0$ of the expected distribution. The background b_i underneath a diffraction spot is often assumed to be a constant which is estimated from the neighbourhood around the reflection.

```
/**
 * Evaluate the next intensity iteration.
 * @ returns The estimate of the intensity
 */
vec2<double>
estimate_intensity(const af::const_ref<FloatType, af::c_grid<3> > &p,
                   const af::const_ref<bool, af::c_grid<3> > &m,
                   const af::const_ref<FloatType, af::c_grid<3> > &c,
                   const af::const_ref<FloatType, af::c_grid<3> > &b,
                   double I) const {
  double sum1 = 0.0;
  double sum2 = 0.0;
  double sumv = 0.0;
  for (std::size_t i = 0; i < p.size(); ++i) {
    if (m[i]) {
      double v = std::abs(b[i]) + std::abs(I * p[i]);
      sumv += v;
      if (v > 0) {
        sum1 += (c[i] - b[i]) * p[i] / v;
        sum2 += p[i] * p[i] / v;
      }
    }
  }
  return vec2<double>(sum2 != 0 ? sum1 / sum2 : 0.0, sumv);
```

REFLECTION PROFILES



Profile for reflection at position x derived from average of strong reflections in block with centre nearest x

INTEGRATION

RASTER / GRID SCAN

TYPICAL RASTER SCAN - 660 IMAGES

20-02-2017 16:18:40 - 20170220/gw/20170220/INS3/grid_INS3_29_3_2_####.cbf

Sample: INS3

Ω Start: -0.0°

Resolution: 1.18Å

Wavelength: 0.9763Å

Exposure: 0.100s

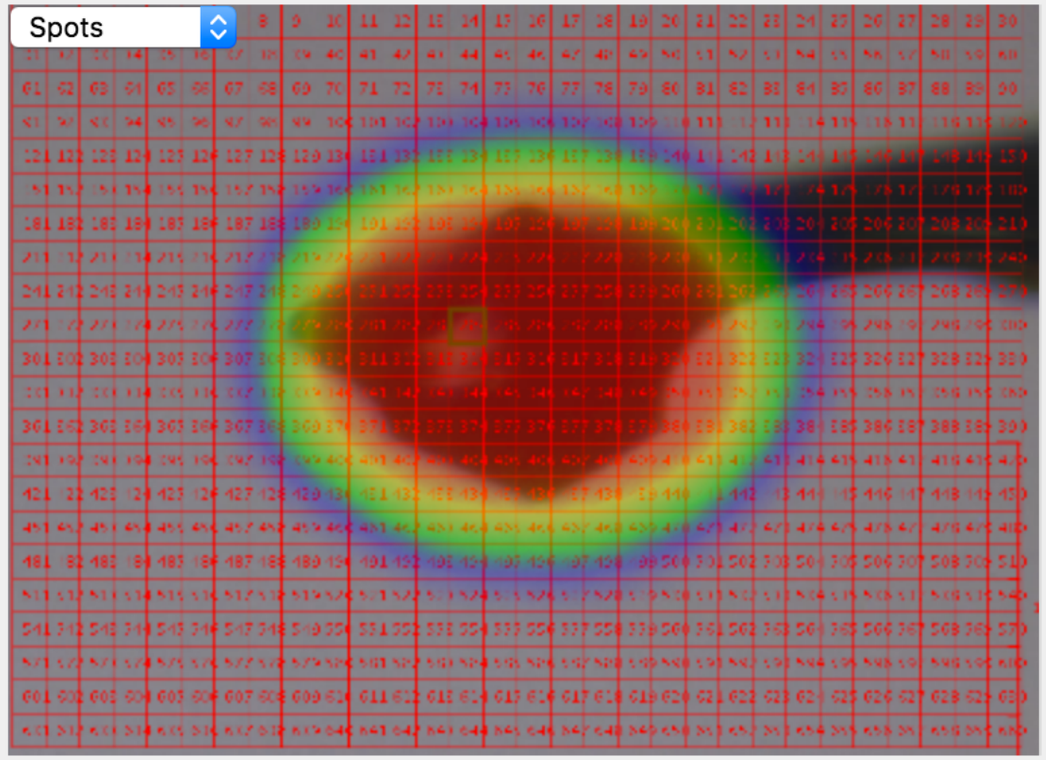
Transmission: 5.00%

Beamsize: 80x20µm

Comment: Diffraction grid scan of 30 by 22 images, Top left [115,72], Bottom right [962,695]

Image: 284 Value: 216

X: 769.04 Y: -419.83 Z: 8.43



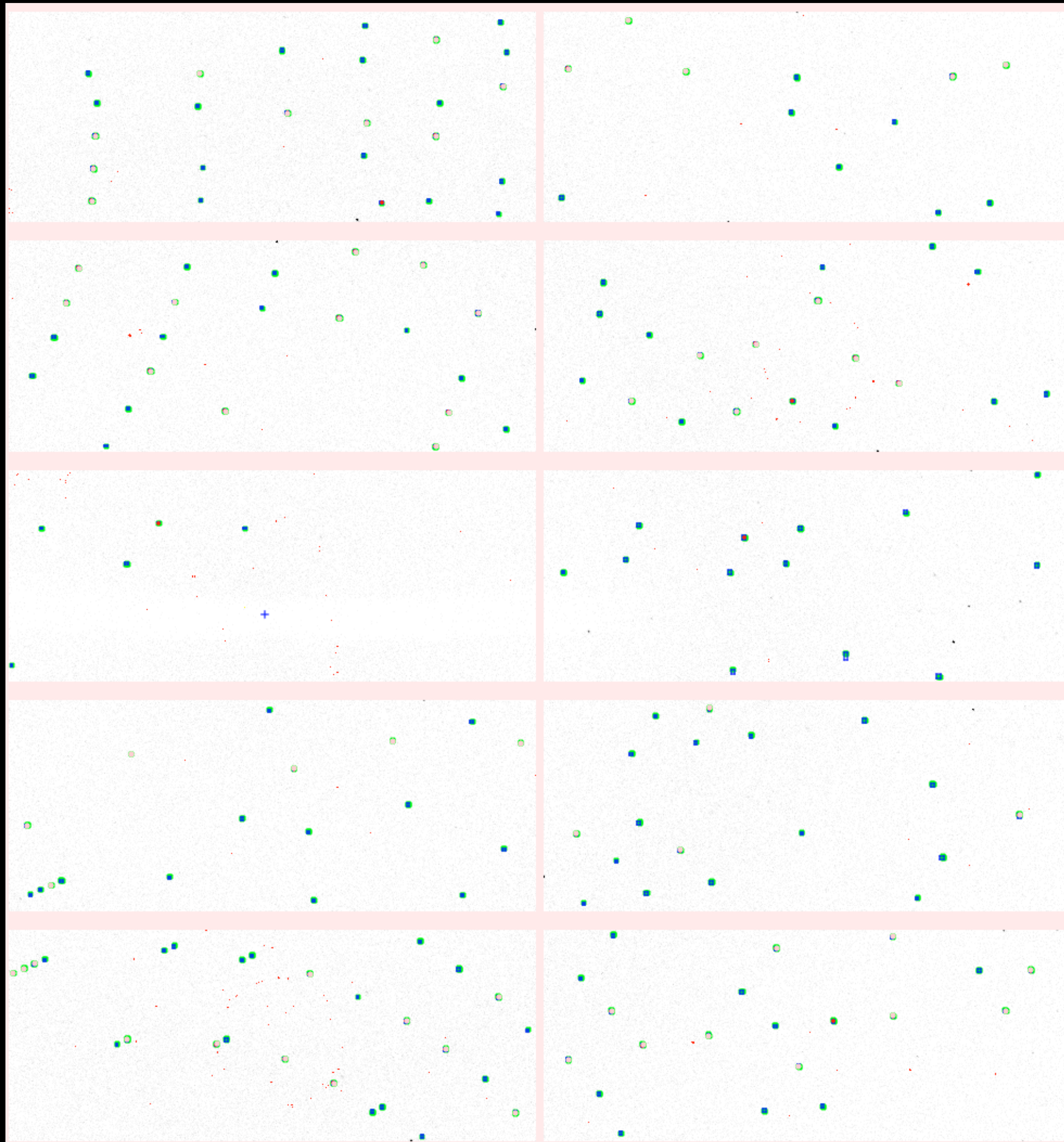
TIMING (SPOT FINDING)

Example set from Martin Savko @ Soleil: 750 frames, Eiger 9M, raster scan - data in ramdisk

- 50s from CBF on 20-core machine - 1.3 frm / core / s
- 60s from HDF5 on 20-core machine - 1.6 frm / core / s

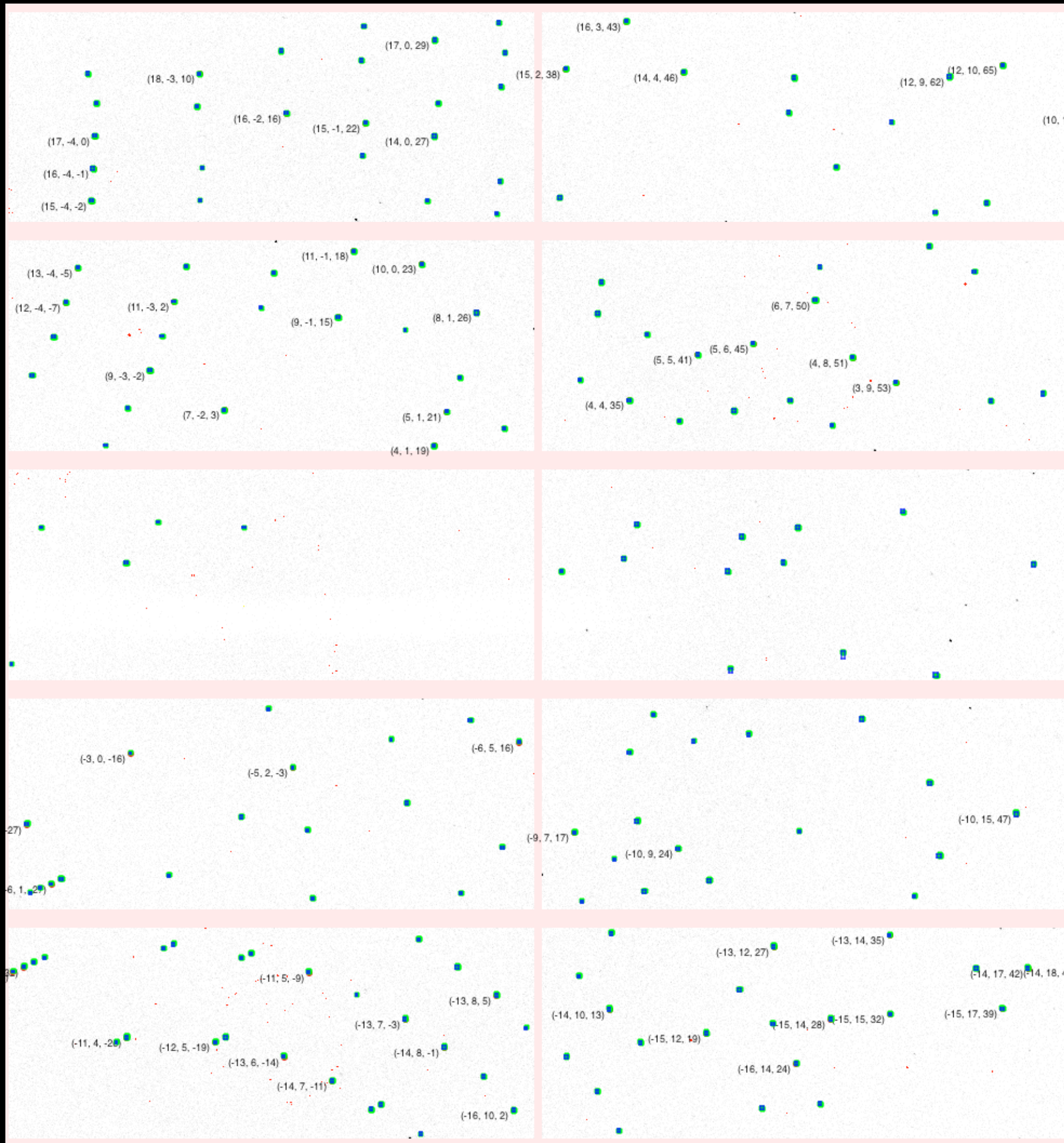
RASTER METRICS

- #spots
- resolution
- shape
- I/sig(I)
- #lattices
- …



VMXI ANALYSIS

- #spots
- #lattices
- unit cell
- resolution
- …



BENEFITS AND CHALLENGES

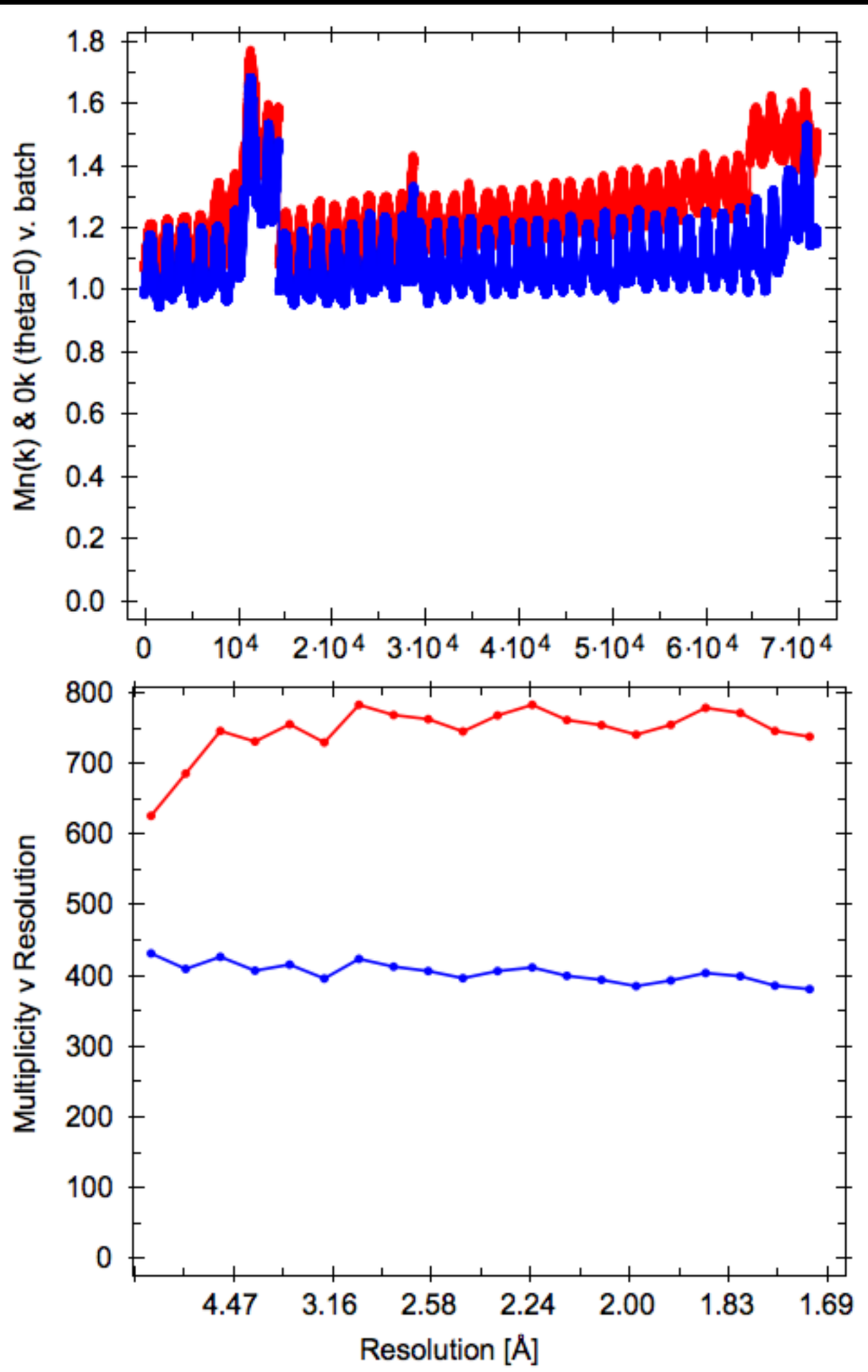
- Much greater computational cost - spot finding ~ 1s, indexing ~ 10s, … :-(
- Problem is embarrassingly parallel - never need to bring the data together :-)
- May converge on synchrotron serial crystallography :-|
- To provide useful feedback to user may require more advanced analytics :-(

ROTATION DATA

PROCESSING

HIGH MULTIPLICITY LOW DOSE

- Rational mode of data collection
- Take “standard” collection
- Record 10 x as many frames at 0.1 x dose / frame
- *Post mortem* decide where to stop experiment



BENEFITS AND CHALLENGES

- Always confident of getting complete data set :-)
- Making excellent use of low noise / photon counting detector :-)
- Avoid principle problem of MX collection :-)
- Have massive #frames :-(
- Sparse data compress well :-)
- Can apply parallelism i.e. integrate every run separately :-)

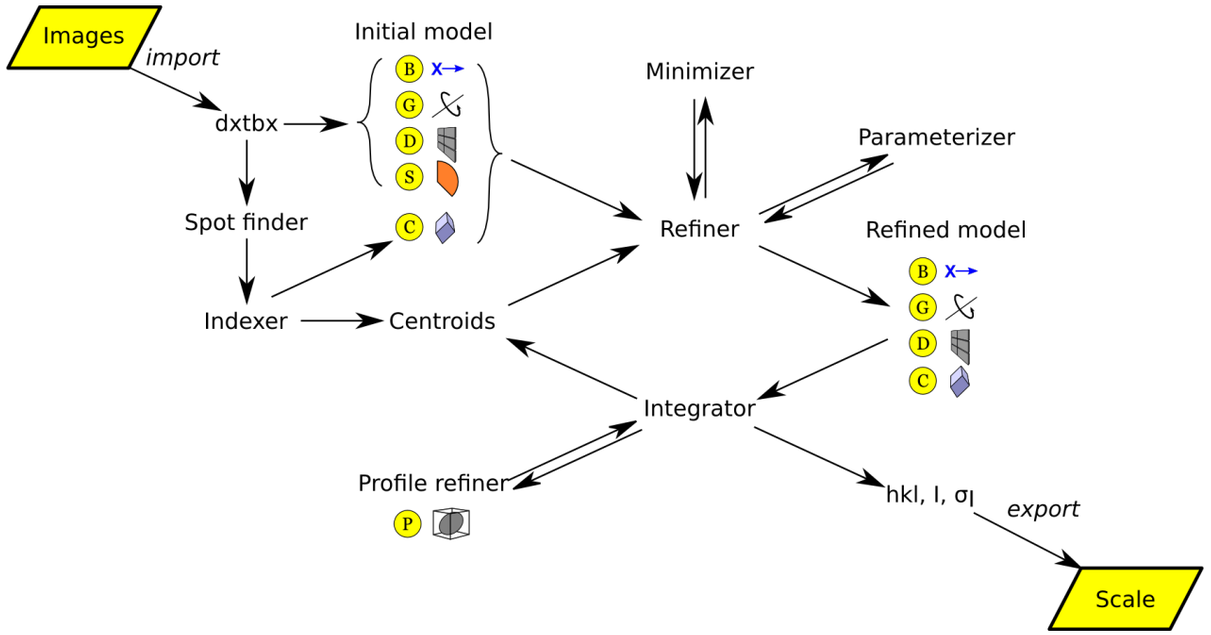
BIGGEST CHALLENGES

- Scaling :-(
- Deciding post-mortem where was best
- As soon as you have to look at every measurement time goes through the roof (aimless ~ 4 hours) :-(
- Find more parallel friendly algorithm - XSCALE much quicker :-)
- Also have to be sure you are integrating carefully - very small numbers of photons / spot likely :-(

AUTOMATION

XIA2

- Embed indexing, integration, scaling
- Embed expertise on e.g. cluster usage
- Allows direct comparison with XDS (also inside xia2)
- Hides complexity of DIALS usage
- Scales well in high throughput environment




```
process — gw56@cs03r-sc-serv-16:/dls/tmp/gw56/c3a/dials/DEFAULT/scale...
[Graemes-MBP-5:process graeme$ xia2 image=~/data/i04-bag-training/th_8_2_0001.cbf]
.gz
Environment configuration...
Python => /Users/graeme/svn/cctbx/build/../base/Python.framework/Versions/2.7/Re
sources/Python.app/Contents/MacOS/Python
CCTBX => /Users/graeme/svn/cctbx/modules
CCP4 => /Applications/ccp4-7.0
CLIBD => /Applications/ccp4-7.0/lib/data
CCP4_SCR => /var/folders/hs/gvssfcd910s0jczjhjczj6h80000gn/T/tmpEfqKjV
Starting directory: /private/tmp/process
Working directory: /private/tmp/process
Free space:        26.41 GB
Contact: xia2.support@gmail.com
XIA2 0.5.263-g6fd2d14
DIALS 1.dev.1215-g9abb12b
CCP4 7.0.021
Command line: xia2 image=/Users/graeme/data/i04-bag-training/th_8_2_0001.cbf.gz
-------------------- Spotfinding SWEEP1 --------------------
```

STATUS

- DIALS now used routinely at Diamond Light Source for automated data processing *via* xia2 - will be in CCP4 7.0
- xia2 is the "friendly" DIALS user interface for synchrotron data, and is bundled with DIALS
- Tools developed using DIALS framework for image inspection, strategy evaluation on I19, detector metrology, …

FUTURE

- Implement scaling - new methods and new software
- Address errors - critical for high multiplicity data
- Port to many-core architecture e.g. xeon phi - address more complex serial crystallography challenges

ACKNOWLEDGEMENTS

- Development teams & users, CCTBX, Diamond staff, …
- Funding - EU, NIH, Wellcome, Diamond, CCP4, …
- 21st century coding environment
- Users providing feedback and example data