The 4th High Data-Rate Macromolecular Crystallography (HDRMX) Meeting
Max IV Laboratory, Lund, Sweden, Thurs 16 March 2017
https://indico.maxiv.lu.se/event/233/
Morning session: Data Reduction Software

LABELIT / DISTL Spotfinder

Graeme Winter
Diamond Light Source

On behalf of: Nick Sauter, Berkeley National Laboratory

National Institutes of Health
National Institute of General Medical Sciences
BASIC DISCOVERIES FOR BETTER HEALTH
NIH grant R01-GM117126

BERKELEY LAB
Lawrence Berkeley National Laboratory

U.S. DEPARTMENT OF ENERGY
Office of Science
contract DE-AC02-05CH11231

LABELIT: Initial characterization and indexing

- EIGER support requested by NSLS-II:
  LABELIT produces input files for BEST to compute data collection strategy
- First pass: work within the usual paradigm of
  one file = one image
- Beamline converts HDF5 data to miniCBFEIGER with the program eiger2cbf

```
% labelit.index index_only=True insu_with_bs_labelit_0*.cbf

LABELIT Indexing results:
Beam center x  169.59mm, y  172.31mm, distance  199.83mm ; 80% mosaicity=0.45 deg.

Solution  Metric fit  rmsd  #spots  crystal_system   unit_cell                                  volume
:)  22     0.1155 dg 0.115    256         cubic cI   77.43  77.43  77.43  90.00  90.00  90.00   464160
:)  21     0.1074 dg 0.119    257  rhombohedral hR  109.53 109.53  67.01  90.00  90.00 120.00   696230
:)  20     0.1155 dg 0.134    257  rhombohedral hR  109.53 109.53  67.00  90.00  90.00 120.00   696141
:)  19     0.1074 dg 0.147    256  rhombohedral hR  109.45 109.45  67.13  90.00  90.00 120.00   696478
:)  18     0.1155 dg 0.113    257  rhombohedral hR  109.45 109.45  67.10  90.00  90.00 120.00   696120
:)  17     0.1155 dg 0.136    256    tetragonal tI   77.46  77.46  77.36  90.00  90.00  90.00   464192
:)  16     0.1155 dg 0.134    256  orthorhombic oF   77.36 109.53 109.59  90.00  90.00  90.00   928581
:)  15     0.1155 dg 0.113    256    monoclinic mC   77.36 109.50  67.05  90.00 125.18  90.00   464259
:)  14     0.1074 dg 0.111    255    monoclinic mC   77.39 109.54  67.03  90.00 125.20  90.00   464292
:)  13     0.0716 dg 0.115    258    tetragonal tI   77.47  77.47  77.36  90.00  90.00  90.00   464223
:)  12     0.0716 dg 0.117    258  orthorhombic oF   77.33 109.52 109.61  90.00  90.00  90.00   928267
:)  11     0.0714 dg 0.109    257    monoclinic mC   77.34 109.59  67.02  90.00 125.20  90.00   464178
:)  10     0.0716 dg 0.116    258    monoclinic mC   77.35 109.52  67.03  90.00 125.14  90.00   464302
:)   9     0.0443 dg 0.107    257    tetragonal tI   77.39  77.39  77.52  90.00  90.00  90.00   464295
:)   8     0.0428 dg 0.106    257  orthorhombic oI   77.37  77.41  77.52  90.00  90.00  90.00   464289
:)   7     0.0443 dg 0.106    257  orthorhombic oF   77.51 109.43 109.48  90.00  90.00  90.00   928666
:)   6     0.0441 dg 0.105    257    monoclinic mC   77.52 109.48  67.04  90.00 125.29  90.00   464357
:)   5     0.0428 dg 0.106    258    monoclinic mC  109.51  77.40  77.38  90.00 134.93  90.00   464311
:)   4     0.0443 dg 0.106    257    monoclinic mC   77.51 109.43  67.08  90.00 125.30  90.00   464332
:)   3     0.0328 dg 0.105    257    monoclinic mC  109.53  77.37  77.41  90.00 134.94  90.00   464314
:)   2     0.0274 dg 0.105    258    monoclinic mC  109.43  77.51  77.38  90.00 134.97  90.00   464327
:)   1     0.0000 dg 0.103    257     triclinic aP   67.04  67.06  67.06 109.48 109.53 109.37   232173
```

Code changes needed to support the miniCBF file format

- New dxtbx format class (FormatCBFMiniEiger.py), inheriting from FormatCBFMini
- Labelit also requires old-style format classes in the iotbx/detectors package
- The package now implements a new eiger_minicbf.py class

Code changes to support spotfinding

- **`distl.signal_strength *.cbf`**
- Default spotfinder parameters:
  minimum_signal_height=4
  (minimum I/sigma value for a pixel to be considered signal)
- Default spotfinder parameters:
  minimum_spot_area=5 (in pixels)
- These EIGER defaults were changed in config_detector.py
- If the defaults don't work for your dataset change them on the command line! **`distl.minimum_spot_area=4`**
- Results are very sensitive to these two thresholds, as we have learned from both synchrotron & XFEL experience.
- Troubleshooting completed to make sure defaults and command line overrides are respected by both indexing and signal_strength
- scanbox.h modified to adapt windowing code to the Eiger sensor dimensions.
- Distl.cpp and libdistl.cpp modified to call the correct class

Running the spotfinder

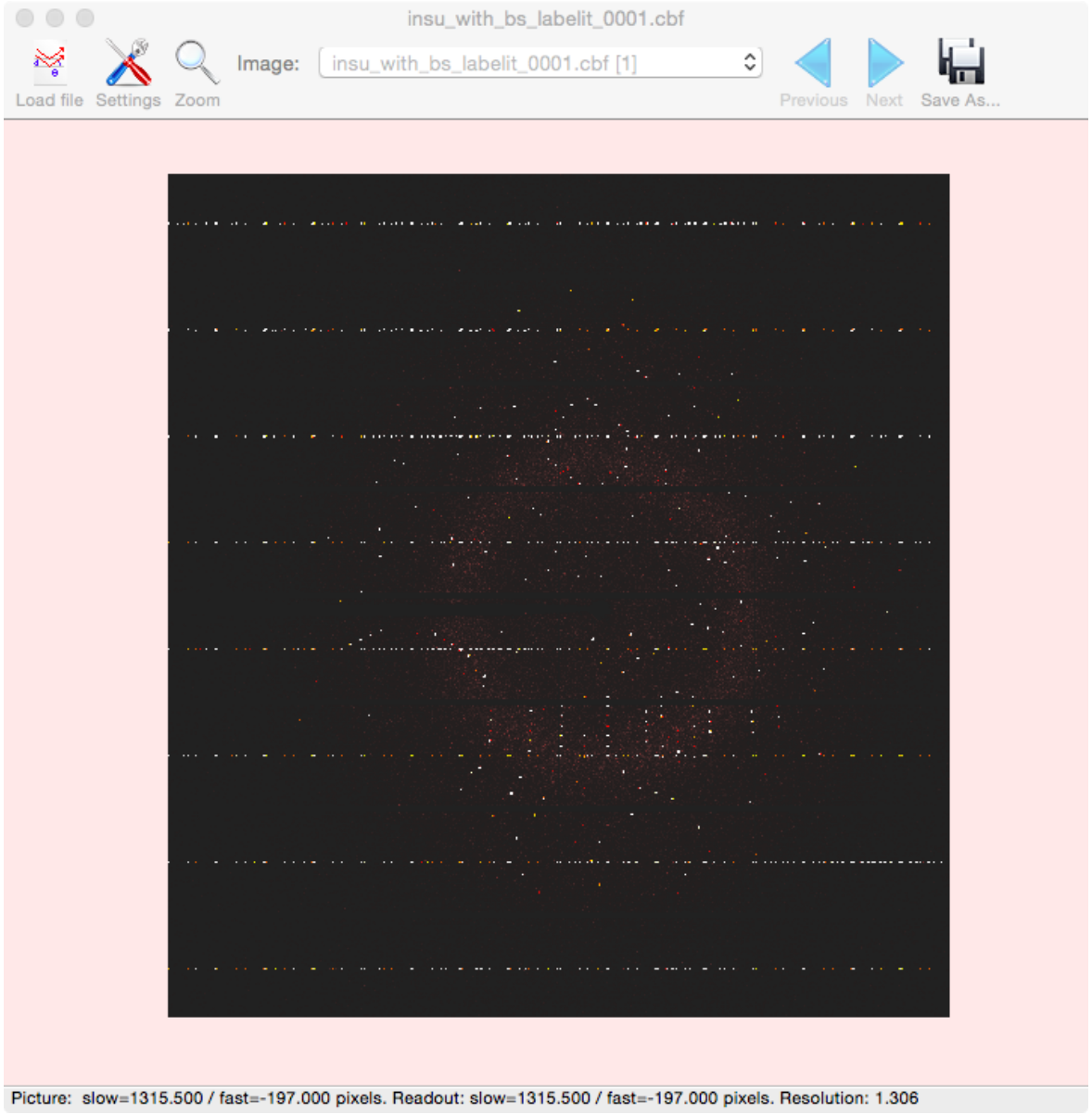
```
% distl.signal_strength insu_with_bs_labelit_0001.cbf
distl.signal_strength: characterization of candidate Bragg spots

                     File : insu_with_bs_labelit_0001.cbf
               Spot Total :    203
              Remove Ice :    199
      In-Resolution Total :    164
    Good Bragg Candidates :    148
                Ice Rings :      0
      Method 1 Resolution :   2.21
      Method 2 Resolution :   2.07
        Maximum unit cell :   81.3
<Spot model eccentricity> :   0.790
%Saturation, Top 50 Peaks :   0.64
In-Resolution Ovrld Spots :      0

Bin population cutoff for method 2 resolution: 20%

Number of focus spots on image #1 within the input resolution range: 164
Total integrated signal, pixel-ADC units above local background (just the good Bragg candidates) 64724
Signals range from 5.2 to 7659.2 with mean integrated signal 481.7
Saturations range from 0.0% to 3.7% with mean saturation 0.2%
```

Spotfinder and indexing results under parameter combinations

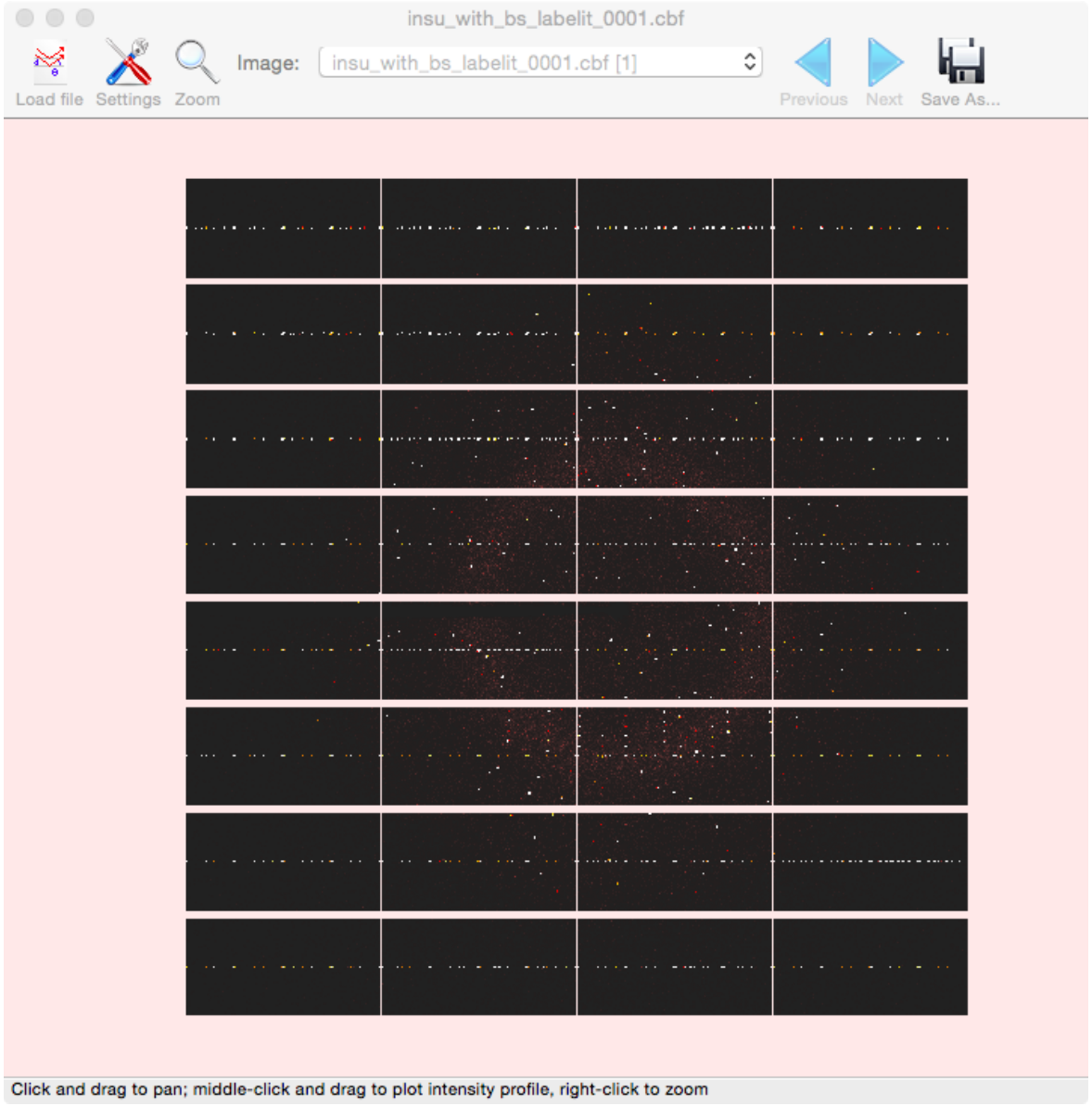
Minimum spot_area	Minimum signal_height	Bragg spot candidates on two images / fit model	Angular deviation from cubic (°)	Cubic setting R.M.S.D. (mm)
3	4.0	619+345 / 282	0.2080	0.601
4	4.0	295+188 / 295	0.0768	0.291
5 (default)	4.0 (default)	148+147 / 256	0.1155	0.115
6	4.0	138+134 / 230	0.0539	0.089
7	4.0	119+122 / 220	0.0596	0.118
5	2.0	1042 + 142 / 194	0.1342	0.379
5	3.0	135 + 178 / 236	0.0470	0.098
5	4.0	148 + 147 / 256	0.1155	0.115
5	5.0	135 + 106 / 208	0.0665	0.083
5	6.0	117 + 113 / 200	0.0555	0.082

The two images used are 0.1° oscillation shots of cubic insulin, taken 90° apart. DIALS refinement on the full rotation sweep would be expected to produce much better R.M.S.D. values.

cctbx.image_viewer



Unbreak the active area discrimination



- Inactive areas of the detector are now shaded pink (was broken for Pilatus too)
- Inactive pixel values (-1) are not figured into the contrast function

Supporting the native H5 format

- The iotbx/detectors package now implements the class eiger.py
- The dxtbx/format package handles file I/O (with class FormatHDFEigerNearlyNexus), while iotbx provides lightweight wrappers (no I/O) that adapt to the legacy code, spotfinder and labelit
- Changed the paradigm from “one-image-per-file” to “one-scan-per-file”
- Supporting the new paradigm only took 47 new lines of code, but a whole day to figure it out.
- How to express the new concept for distl spotfinder?

Many thanks to Martin Savko @ Soleil for example data

```
distl.signal_strength 200Hz_0.1dpf_1_master.h5 range=10,12
```

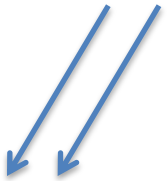
New “range” parameter for HDF5 data only, gives 0-based image index, or Python-style range (0,10 means the first 10 images).
Default (range=None) is to report spotfinder spots on the 0th image

- Metadata are only read once; image data are read only when needed
- Performance is fair; 3.8 sec / Eiger-9M image on a Mac laptop

Supporting the native H5 format in Labelit indexing

- Code added to the Labelit repository to support HDF5, but only needed 30 new lines. The real work was already done by James Parkhurst in dxtbx.

No range keyword for Labelit. The 0-based indices for the images used for indexing are given directly on the command line, similar to the existing syntax for one-image-per-file (non-HDF5) data.



```
$ labelit.index soleil-hdf5-eiger-9m/200Hz_0.1dpf_1_master.h5 0 900 index_only=True \
distl.minimum_signal_height=3
200Hz_0.1dpf_1_master.h5
/Users/nksauter/xtaldata/soleil/soleil-hdf5-eiger-9m/200Hz_0.1dpf_1_master.h5
/Users/nksauter/xtaldata/soleil/soleil-hdf5-eiger-9m/200Hz_0.1dpf_1_master.h5


LABELIT Indexing results:
Beam center x  124.23mm, y  114.93mm, distance  126.81mm ; 80% mosaicity=0.45 deg.

Solution  Metric fit  rmsd  #spots  crystal_system   unit_cell                                    volume
:)    5     0.0871 dg 0.081    290  orthorhombic oP   43.03  64.15  85.05  90.00  90.00  90.00   234817
:)    4     0.0850 dg 0.081    290    monoclinic mP   64.14  43.03  85.06  90.00  89.98  90.00   234811
:)    3     0.0871 dg 0.081    290    monoclinic mP   43.04  85.05  64.15  90.00  89.99  90.00   234818
:)    2     0.0246 dg 0.088    291    monoclinic mP   43.05  64.16  85.03  90.00  89.92  90.00   234835
:)    1     0.0000 dg 0.080    289     triclinic aP   43.05  64.16  85.03  89.98  89.92  89.99   234837
```