CMDDENZO and CMDXDS

Single-Line-Command-Driven User Interfaces for Automated Data Processing at SER-CAT

- Zheng-Qing Fu

- SERCAT, APS, Argonne National Lab., Argonne, IL 60439
  - Department of Biochemistry & Molecular Biology, University Of Georgia, Athens, GA 30602

TWG, APS, ANL, 9/16/2010

To Match the speed offered by Bright X-ray

The much brighter 3rd-generation synchrotron X-ray source has dramatically reduced the time needed for collecting a full data set.

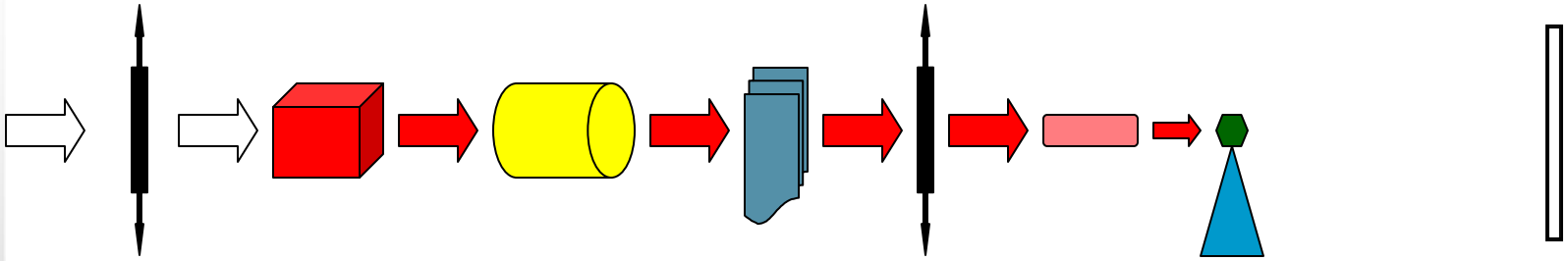
However, to fully take the advantage of quick data acquisition, it requires higher precision of all the hardware components, not only in space but also in time. For example,

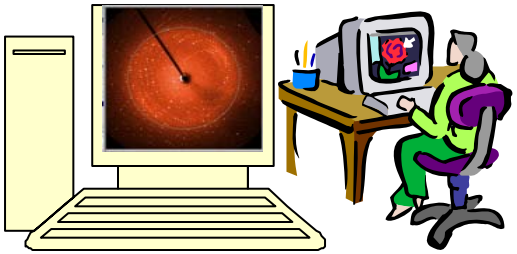
Assuming the (shutter)/goni/detector has a synchronization error of 40ms:

20s Exposure: 0.2%.

2s Exposure: 2%.

Basic Components of a X-ray Crystallograhic Beamline





Key Components

- X-Ray Source (Ring, Mono, Optics, …)
- (Shutter)
- Goni
- Detector
- …

Hard to Monitor Precision/Precision Loss in Real Time

Unfortunately, it is very hard to monitor or evaluate the precision/precision loss of a hardware component in real time.

Monitor Data Quality is Practical with Quick Handy Tool

A practical making up is to evaluate and monitor data quality on-the-fly during data collection.

As part of our efforts to realize real-time quality control, we developed a few Single-Line-Command Driven UIs which exploit several different programs for data processing:

1). **cmddenzo** (DENZO/SCALEPACK, D*TREK, SGXPRO, SPGR4D, X-GEN).
2). **cmdxds** (XDS, CCP4, SGXPRO, SPGR4D).
3). **cmddtrek.py** [Zhongmin Jin] (D*TREK)
4). **xgenproc.py** [Andy Howard] (XGEN)

Purpose

These user interfaces are not intended to match the expert use of each individual program, but to provide a tool to quickly process and characterize a data set by determining:

Spacegroup
Rmerge
I/SigI
Completeness
Redundancy

etc.

Features

To achieve the goal, the UIs should at least have the following two features:

1). Fully functional.

2). Simple to use.

Workflow of CMDDENZO

Images

↓

XGEN / D*TREK (harvesting spots for indexing)

↓

DENZO (auto indexing)

↓

CMDDENZO (integration box sizes estimation)

↓

DENZO (integration)

↓

SPGR4D/SGXPRO (space group determination)

↓

3DSCALE/SGXPRO (error model & resolution estimation)

↓

SCALEPACK (scaling)

↓

CMDDENZO (evaluation & summarizing)

Usage: On-Line Help

%cmddenzo

…

Usage:

cmddenzo DefSite LatType SpaceGroup Function ImageFile nFrames -p peaks.file / -g ProgPeakPicking

…

Examples:

…

Usage: CMDDENZO Examples

cmddenzo mar300 u u scalesad image.0001 nFrames

cmddenzo mar225 p3 u scalesad image.0001 nFrames

cmddenzo def.site p3 P622 scalenat image.0001 nFrames

Usage: Demo

… Live Demo …

Workflow of CMDXDS

Images

CMDXDS (extract system configuration parameters from header and ‘def.site’)

CMDXDS (survey the computer to setup **parallel** or **none-parallel** tasks)

XDS (picking spots and indexing)

XDS (integration)

(beam-divergence, resolution, etc.)

XSCALE/XDS (scaling)

CMDXDS, SGXPRO, SPGR4D (evaluation & summarizing)

XDSCONV, CMDXDS, CCP4 (converting output to *.sca files)

Usage: On-Line Help

%cmdxds

…

Usage:

cmddenzo DefSite LatType SpaceGroup Function ImageFile nFrames -p peaks.file / -g ProgPeakPicking

…

Examples:

…

Usage: CMDXDS Examples

cmdxds mar300 image.0001 nFrames

cmdxds mar225 image.0001 nFrames

cmdxds mydef.site image.0001 nFrames

Usage: Demo

… Live Demo …

Output: Indexing

Lattice	Distortion	Symmetrized UnitCell					
P23	8.57%	76.72	76.72	76.72	90.00	90.00	90.00
I23	22.91%	109.32	109.32	109.32	90.00	90.00	90.00
F23	22.97%	134.23	134.23	134.23	90.00	90.00	90.00
R3	8.57%	76.72	76.72	76.72	90.03	90.03	90.03
P3	13.99%	83.80	83.80	62.57	90.00	90.00	120.00
P4	3.16%	83.80	83.80	62.57	90.00	90.00	90.00
I4	12.90%	126.50	126.50	62.57	90.00	90.00	90.00
P222	0.02%	62.57	78.21	89.40	90.00	90.00	90.00
C222	3.16%	118.75	118.81	62.57	90.00	90.00	90.00
I222	12.91%	62.57	118.75	134.25	90.00	90.00	90.00
F222	12.52%	62.57	168.45	189.38	90.00	90.00	90.00
P2	0.01%	62.57	78.21	89.40	90.00	90.04	90.00
C2	3.16%	118.75	118.81	62.57	90.00	90.03	90.00
P1	0.00%	62.57	78.21	89.40	90.03	90.04	90.01

```
PossibleLatticeID: P222
IndexDistortion: 0.02%
UnitCell: 62.570  78.210  89.400  90.00  90.00  90.00
Mosaicity: 0.687
Estimated Lowest Highest Resolutions: 50.0000  1.6826
```

Output: Space Group Determination

```
nBatch is 1

PEEK_scale_major=1  minor=0 /pass
PEEK_scale_major=1  minor=0


Under the lattice type used at the auto-indexing and
integration:
##################################################
#
# Suggested Space Group:          P 21 21 21
#
##################################################


 ... SPGR4D spacegroup determination done ...


 SpaceGroup 'P212121' will be used in scaling.
```

CMDDENZO Final Statistics Output Summary

Statistics by Resolution Shells [range: 50.0000 1.6800]:

--- Reso	ComplInSh	RedunInSh	IsigiInSh	RmergInSh ---
4.56	98.70	9.40	162.49	0.0510
3.62	100.00	10.00	171.66	0.0520
3.16	100.00	10.10	151.58	0.0700
2.87	100.00	10.20	117.72	0.0710
2.67	100.00	10.20	99.36	0.0650
2.51	100.00	10.30	84.88	0.0680
2.38	100.00	10.20	72.14	0.0710
2.28	100.00	10.20	63.67	0.0750
2.19	100.00	10.10	55.74	0.0840
2.12	100.00	10.10	47.98	0.0900
2.05	100.00	10.00	39.44	0.1030
1.99	100.00	10.00	33.19	0.1170
1.94	100.00	9.90	27.84	0.1320
1.89	100.00	9.90	22.67	0.1600
1.85	100.00	9.80	18.14	0.1900
1.81	100.00	9.60	15.06	0.2150
1.77	100.00	9.30	13.00	0.2410
1.74	100.00	8.90	10.52	0.2770
1.71	100.00	8.10	8.53	0.3150
1.68	93.70	6.50	5.68	0.3320
all	99.60	9.60	80.10	0.0710

Effective Resolution Range in Reduced Data: 50.0000 1.6800

Rmerge:	0.0710 [	0.0510,	0.3320]
Redund:	9.60 [	9.40,	6.50]
Complt:	99.60 [	98.70,	93.70]
I/SigI:	80.10 [	162.49,	5.68]
Chi^2:	1.55 [	1.88,	1.25]

CMDDENZO Summary: /home/staff/demo/zzProc_testbm2/zzcmddenzo_sol.log
Scaling Log File: /home/staff/demo/zzProc_testbm2/zzcmddenzo_sca.log
Output *.sca File: /home/staff/demo/zzProc_testbm2/zzcmddenzo_sca.sca

CMDXDS Final Statistics Output Summary

```
... xds/xscale ... cmdxds analyzing ... all done ...

UnitCell: 28.840  28.840  262.590  90.00  90.00  120.00
Space/PointGroup Suggested by XDS/XSCALE: 177        P622
    Warning: The space group from XDS may not be accurate!
             Please check it with 'spgr4d' or 'cmddenzo'.

Statistics by Resolution Shells [range: 50.0000   1.4200]:
--- Reso  ComplInSh  RedunInSh  IsigiInSh  RmergInSh ---
    6.36       83.60       6.99      26.87     0.0290
    4.49       98.60       8.10      27.81     0.0430
    3.67       99.30       7.99      28.57     0.0500
    3.18       99.40       7.25      26.70     0.0550
    2.84       98.50       6.75      25.00     0.0670
    2.60       98.80       6.36      23.70     0.0690
    2.40       98.60       6.03      22.38     0.0780
    2.25       98.50       5.58      20.65     0.0810
    2.12       98.70       5.47      19.30     0.0850
    2.01       96.70       5.33      17.57     0.0920
    1.92       97.10       5.33      14.50     0.1030
    1.83       97.30       5.22      12.94     0.1210
    1.76       97.10       5.13      11.29     0.1330
    1.70       95.40       5.19       9.16     0.1580
    1.64       96.30       4.94       8.54     0.1670
    1.59       94.20       5.00       7.51     0.1860
    1.54       91.50       4.77       6.03     0.2240
    1.50       89.60       4.86       5.37     0.2500
    1.46       81.90       4.05       4.47     0.2860
    1.42       36.90       2.60       2.64     0.3470
     all       90.90       5.45      14.35     0.0670

Effective Resolution Range in Reduced Data: 50.0000   1.4200
Rmerge:     0.0670 [    0.0290,     0.3470]
Redund:       5.45 [      6.99,       2.60]
Complt:      90.90 [     83.60,      36.90]
I/SigI:      14.35 [     26.87,       2.64]

CMDXDS  Summary: /home/staff/apsdemo/zzCmdx_test0/zzCmdx_test0_sol.log
Output Log File: /home/staff/apsdemo/zzCmdx_test0/xs/XSCALE.LP
Output Data XDS File: /home/staff/apsdemo/zzCmdx_test0/xs/zzxscale.ahkl
Output Data SCA File: /home/staff/apsdemo/zzCmdx_test0/xs/test0_xdsAno.sca
Output Data SCA File: /home/staff/apsdemo/zzCmdx_test0/xs/test0_xdsIso.sca
```

More on the Programs

http://www.ser.aps.anl.gov

Beamline User's Guide -> Data Processing

Acknowledgment

Bi-Cheng Wang, John Rose, John Chrzas, Jim Fait, Zhongmin Jin, John Gonczy, Rod Salazar, Unmesh Chinte, Palani Kandavelu, SER-CAT, APS, Argonne National Lab; Department of Biochemistry & Molecular Biology, UGA

Fred Dyda, NIDDK, NIH

Andy Howard, Illinois Institute of Technology