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 3\textsuperscript{rd}-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 Crystallographic Beamline

Key Components

- X-Ray Source (Ring, Mono, Optics, …)
- (Shutter)
- Goni
- Detector
- …
Unfortunately, it is very hard to monitor or evaluate the precision/precision loss of a hardware component in real time.
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)
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.
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)
%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)

XSCALE/XDS (scaling)

CMDXDS, SGXPRO, SPGR4D (evaluation & summarizing)

XDSCONV, CMDXDS, CCP4 (converting output to *.sca files)
Usage: On-Line Help

%cmdxds

...

Usage:

cmdddenzo 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 ...
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
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]:

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

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]:

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

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


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