Agenda

NE-CAT
Automated data analysis @ NE-CAT
Current capabilities
Development of version 2
  Design & implementation
  Problems Opportunities
Acknowledgements

Contributors

Jon Schuermann*
David Neau
Kay Perry
Jim Withrow

NE-CAT Support Staff

Surajit Banerjee
Ali Kaya
Igor Kourinov
N. Sukumar

Support

NIGMS P30 GM124165
NE-CAT Beamlines
APS Sector 24

24-ID-C
Tunable 6.5-20 keV
$1.5 \times 10^{13}$ photons/sec
25 µm H x 10 µm V
Eiger2 X 16M

24-ID-E
Fixed 12,663 eV
$1.0 \times 10^{13}$ photons/sec
35 µm H x 5 µm V
Eiger 16M
Our Goals For Automated Data Analysis

• Help users collect better data faster
• Enable collaboration with off-site users
• Project started in 2009
Current Status – RAPDv1
RAPD Capabilities

- Indexing & Strategy
- Integration, Scaling & Analysis
- Merging
- SAD
- MR
RAPD Capabilities

- Indexing & Strategy – DISTL, Labelit, BEST, Mosflm
- Integration, Scaling & Analysis – XDS, Aimless, Pointless, Phaser, Xtriage
- Merging – Aimless, Pointless
- SAD – SHELX, Phenix
- MR – Phaser
RAPD Index

**Labelit autoindexing summary for:**

Image: /gpfss1/users/cornellchem/7880_C_5704/images/haoyue/snaps/NE38_7PAIR_0_000001.cbf

<table>
<thead>
<tr>
<th>Space Group</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>α (deg)</th>
<th>β (deg)</th>
<th>γ (deg)</th>
<th>mosaicity</th>
<th>resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>P4</td>
<td>57.83</td>
<td>57.83</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
<td>0.10000</td>
<td>1.25</td>
<td></td>
</tr>
</tbody>
</table>

**Detector distance**

<table>
<thead>
<tr>
<th>Resolution limit at edge</th>
<th>Detector distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>150</td>
</tr>
<tr>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>290</td>
</tr>
</tbody>
</table>

**Data collection strategy from BEST**

<table>
<thead>
<tr>
<th>N</th>
<th>Omega Start</th>
<th>Omega End</th>
<th>Rot Range</th>
<th>N of Images</th>
<th>Delta Omega</th>
<th>Exposure time</th>
<th>% Transmission</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>82.00</td>
<td>140.00</td>
<td>58.00</td>
<td>250</td>
<td>0.20</td>
<td>0.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

**ANOMALOUS data collection strategy from BEST**

<table>
<thead>
<tr>
<th>N</th>
<th>Omega Start</th>
<th>Omega End</th>
<th>Rot Range</th>
<th>N of Images</th>
<th>Delta Omega</th>
<th>Exposure time</th>
<th>% Transmission</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>78.00</td>
<td>170.00</td>
<td>92.00</td>
<td>300</td>
<td>0.25</td>
<td>0.2</td>
<td>1.5</td>
</tr>
</tbody>
</table>
RAPD Integrate

Processing Results for NE38_1_1

Images 1-1800

Spacegroup: P 41 21 2
Unit Cell: 57.85 57.85 150.15 90.00 90.00 90.00
SIGMARN (Mosaicity): 0.09329
Asymptotic limit of l Isis (IIs) = 23.81

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>Inner Shell</th>
<th>Outer Shell</th>
</tr>
</thead>
<tbody>
<tr>
<td>High resolution limit</td>
<td>1.27</td>
<td>6.94</td>
<td>1.27</td>
</tr>
<tr>
<td>Low resolution limit</td>
<td>150.15</td>
<td>150.15</td>
<td>1.29</td>
</tr>
<tr>
<td>Completeness</td>
<td>99.9</td>
<td>100.0</td>
<td>98.6</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>25.7</td>
<td>18.2</td>
<td>24.0</td>
</tr>
<tr>
<td>l Isis</td>
<td>30.2</td>
<td>81.7</td>
<td>3.5</td>
</tr>
<tr>
<td>Cc(1/2)</td>
<td>0.999</td>
<td>0.999</td>
<td>0.875</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.102</td>
<td>0.081</td>
<td>1.055</td>
</tr>
<tr>
<td>Rmerge (anomalous)</td>
<td>0.100</td>
<td>0.079</td>
<td>1.037</td>
</tr>
<tr>
<td>Rmeas</td>
<td>0.104</td>
<td>0.083</td>
<td>1.078</td>
</tr>
<tr>
<td>Rmeas (anomalous)</td>
<td>0.104</td>
<td>0.082</td>
<td>1.060</td>
</tr>
<tr>
<td>Rpim</td>
<td>0.020</td>
<td>0.018</td>
<td>0.215</td>
</tr>
<tr>
<td>Rpim (anomalous)</td>
<td>0.028</td>
<td>0.023</td>
<td>0.299</td>
</tr>
<tr>
<td>Anomalous completeness</td>
<td>99.8</td>
<td>100.0</td>
<td>96.8</td>
</tr>
<tr>
<td>Anomalous multiplicity</td>
<td>13.7</td>
<td>12.1</td>
<td>12.7</td>
</tr>
<tr>
<td>Anomalous correlation</td>
<td>-0.131</td>
<td>0.047</td>
<td>-0.089</td>
</tr>
<tr>
<td>Anomalous slope</td>
<td>1.000</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Total observations</td>
<td>1785371</td>
<td>9851</td>
<td>78787</td>
</tr>
</tbody>
</table>
RAPD SAD

SHELXC Results

<table>
<thead>
<tr>
<th>SAD</th>
<th>Resolution</th>
<th>0.73</th>
<th>0.53</th>
<th>2.66</th>
<th>2.17</th>
<th>1.86</th>
<th>1.64</th>
<th>1.47</th>
<th>1.34</th>
<th>1.23</th>
<th>1.14</th>
<th>1.07</th>
</tr>
</thead>
<tbody>
<tr>
<td>9C06</td>
<td>0.96</td>
<td>0.96</td>
<td>2.01</td>
<td>4.35</td>
<td>6.11</td>
<td>9.33</td>
<td>15.99</td>
<td>17.12</td>
<td>19.01</td>
<td>16.68</td>
<td>15.78</td>
<td>14.93</td>
</tr>
<tr>
<td>CNF</td>
<td>1.19</td>
<td>1.02</td>
<td>1.04</td>
<td>1.18</td>
<td>1.35</td>
<td>1.53</td>
<td>1.54</td>
<td>1.51</td>
<td>1.45</td>
<td>1.35</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>&lt;sigma&gt;</td>
<td>182.3</td>
<td>182.0</td>
<td>134.5</td>
<td>107.4</td>
<td>87.3</td>
<td>44.5</td>
<td>26.5</td>
<td>10.7</td>
<td>11.5</td>
<td>5.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Complete</td>
<td>99.9</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>98.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;sigma&gt;*8</td>
<td>3.80</td>
<td>2.37</td>
<td>2.13</td>
<td>2.07</td>
<td>1.72</td>
<td>1.32</td>
<td>1.04</td>
<td>0.85</td>
<td>0.77</td>
<td>0.73</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>CC2(C2)</td>
<td>95.7</td>
<td>93.4</td>
<td>79.4</td>
<td>70.1</td>
<td>52.1</td>
<td>31.7</td>
<td>19.4</td>
<td>10.4</td>
<td>5.1</td>
<td>2.5</td>
<td>-3.0</td>
<td></td>
</tr>
</tbody>
</table>

SHELXD Results

<table>
<thead>
<tr>
<th>Space Group</th>
<th>Max CCall</th>
<th>Max CCweak</th>
<th>Max CFOM</th>
<th>Max PATTOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>P41212 (P43212)</td>
<td>29.41</td>
<td>26.27</td>
<td>49.66</td>
<td>9.81</td>
</tr>
<tr>
<td>P4122 (P4322)</td>
<td>22.39</td>
<td>14.72</td>
<td>37.11</td>
<td>3.61</td>
</tr>
<tr>
<td>P422</td>
<td>9.13</td>
<td>5.92</td>
<td>14.14</td>
<td>2.94</td>
</tr>
<tr>
<td>P222</td>
<td>8.55</td>
<td>4.53</td>
<td>13.08</td>
<td>3.1</td>
</tr>
<tr>
<td>P42212</td>
<td>10.61</td>
<td>6.00</td>
<td>16.00</td>
<td>2.63</td>
</tr>
</tbody>
</table>

SHELXE Results

SHELXE may have found a solution in P41212 with a solvent fraction of 0.55.
That’s All Great But…

• Very difficult to add new features
• NE-CAT specific code throughout
• PHP UI was outdated
An Interruption to Our Regularly Scheduled Program
RAPD in Remote
Our Goals For **RAPDv2**

- Take the best of **RAPDv1** and throw out the rest
- Borrow knowledge from remote access program
- Flexible design
- Multiple modes of use
- Modern UI
RAPDv1 Design
RAPDv2 Design
Core Python code is beamline-agnostic

Where does specificity come from?
- Configuration file
- Site-specific plugins - detector, gatherers

Data processing plugins now run in CLI

What’s a plugin?
- Connection to databases
- Data gatherers
- Connection to cluster
- Processing pipelines
RAPDv2 Design

Redis-based inter-process communication
Redis is an indispensable tool for us.

- key-value store
- publication-subscribe interprocess communication
- redis.io

Redis-based inter-process communication
RAPDv2 Design

Document-based database
MongoDB
RAPDv2 Design

MongoDB much more flexible than SQL-based solutions

- Stores documents & files
- More flexibility
- Meshes well with modern UI design
- mongodb.com

Document-based database
MongoDB
RAPDv2 Design

NodeJS Application server
RAPDv2 Design

NodeJS offers a lot of power & flexibility

- Powerful, flexible & fast
- Tight integration with UI
- Provides REST & Websocket access to RAPDv2
- nodejs.org
RAPDv2 Design

Angular UI
Modern web development platform

- Very powerful
- Steep learning curve
- Actively developed & large community
- angular.io
RAPDv2 Current Capabilities

- Indexing & Strategy – DISTL, Labelit, BEST, Mosflm
- Integration, Scaling & Analysis – XDS, Aimless, Pointless, Phaser, Xtriage
- Merging – Aimless, Pointless
- SAD – SHELX, Phenix
- MR – Phaser
RAPDv2 Integrate
Our Goals For RAPDv2

Take the best of RAPDv1 and throw out the rest

- Much of the data processing code was adapted
- Some UI elements that we liked were kept
Our Goals For RAPDv2

Borrow knowledge from remote access program

• Use of Redis & MongoDB
• NodeJS as application server
Our Goals For RAPDv2

Flexible design

• Core programs are not site-specific
• Config file determines variables and plugins that define site
Our Goals For RAPDv2

Multiple modes of use

• CLI for data processing & development
• REST interface
• Websocket interface
RAPDv2 Status

Now

• RAPDv2 running as secondary
• Final steps of UI implementation for SAD & MR

Mid 2021

• Replace RAPDv1 as primary
• Built-out operations control

Late 2021

• Merging and Cluster Merging tools
Can I Use RAPDv2?

• If you really like the *bleeding* edge, have at it.

• Waiting until NE-CAT is using as primary (mid-2021) is recommended.
Acknowledgements

Contributors
  Jon Schuermann
  David Neau
  Kay Perry
  Jim Withrow

NE-CAT Support Staff
  Surajit Banerjee
  Ali Kaya
  Igor Kourinov
  N. Sukumar

Support
  NIGMS P30 GM124165