Towards the autonomation of experiments: quasi-real time analysis of heterogeneous streaming data from scientific experiments

S.J.L. Billinge

Columbia University,

Brookhaven National Laboratory
High level goal

To discover/design better materials
• For sustainable energy
• For health
• For environmental remediation
Scientific Challenge

1. Predict materials with better properties
2. Make them
Scientific Challenge

1. **Predict materials with better properties**
   1. Use quantum mechanical calculations (density functional theory – DFT)
   2. Speed it up with AI, ML and data-mining

2. **Make them**
   1. Predictive synthesis

   “Given this recipe, what dish do I get?” -> “Given this dish, what recipe will make it?”
   - No good solution for materials which have never been made before (Predicted materials!)
Proposed approach

- High throughput, automated and autonomous in situ experiments

Image credit: Christoffer Tyrsted

HTTP://thebillingegroup.com
Adaptive feedback control: autonomation
Computational challenges

1. Instrument control
2. In line rapid data analysis
3. Feedback
   1. Flexible to accommodate different decision policies and experiment designs
Current Stack

- **Epics**: Instrument control with EPICS, but could be anything.
- **Ophyd**: Python interface to Epics (or whatever).
- **Bluesky**: User control and data output.
- **Streaming Data**: Stream is of documents in Bluesky “Event model”.
- **Kafka or 0MQ**: User supplied servers can subscribe to, and publish to, raw or analyzed streams handled by a kafka/0MQ message bus.
Real time streaming data analysis

### Example 3: Asynchronously Monitor During a Scan

<table>
<thead>
<tr>
<th>Event</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Move a motor</td>
<td>Move a motor</td>
</tr>
<tr>
<td>Read motor position and trigger and read detector(s)</td>
<td>Move</td>
</tr>
<tr>
<td>Move and read</td>
<td>Trigger and read</td>
</tr>
<tr>
<td>etc</td>
<td>etc</td>
</tr>
<tr>
<td>Monitor beam current</td>
<td>Record new value</td>
</tr>
<tr>
<td>Record new value</td>
<td>etc</td>
</tr>
</tbody>
</table>

- **Stream**
- **map; inc**
- **Sink; append**
Current Stack: Crystal mapping, dark field microscopy
Streaming data analysis

Python based

• Bluesky event model
• Streamz (native python in a stream)
• SHED (Streaming heterogeneous event data) (Bluesky events in a stream)
  – Also builds in provenance and replay
• Visualization using, for example, Matplotlib
• Raw and analyzed metadata in a MongoDB using Intake for cataloging
• Various object store options for the (large) raw data files themselves
Unsupervised machine learning, e.g., NMF
NMF in a stream example: Chemical components in a discharging battery
=> Spatially Resolved PDFs

- Anton Kovyakh, Soham Banerjee, Chia Hao Liu
Combine PDF and tomography (ctPDF)

ctPDF developed with Simon Jacques, Marco DiMichiel, Andy Beal and Bob Cernik
10,000 2D datasets per image, 30 mins per image ~10Tb/day
Diffpy project
Complex Modeling infrastructure: Diffpy-CMI

DiffPy - Atomic Structure Analysis in Python
A free and open source software project to provide python software for diffraction analysis and the study of the atomic structure of materials.
To the cloud!!!!!
structureMining

Given a measured (or calculated) PDF, structureMining will search databases to find the best structures to fit it.

Please see the structureMining paper for more information. Please cite the structureMining paper if this helps you get a publishable result.

Upload Data

PDF file: Browse... No file selected.

- X-ray
- Neutron

Composition:

Optional Parameter:

Submit
structureMining

Results

SM found total 3 structures and 3 structures with weighted profile agreement factor, \( R_w < 0.5 \).

<table>
<thead>
<tr>
<th>rw</th>
<th>formula</th>
<th>space_group</th>
<th>db</th>
<th>db_id</th>
<th>ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.325028</td>
<td>l4/mmm</td>
<td>MPD</td>
<td>mp-4488</td>
<td>Saha et al. Journal of Physics: Conference Series. 273 (2011) 012104/1-012104/4</td>
</tr>
</tbody>
</table>

Download Full Table
• Structure-mining found the same model as in prior work, MPD No. 1003 (NaFeSi$_2$O$_6$) and COD No. 2983 (NaFeSi$_2$O$_6$), s.g.: C 2/c.
• It also returns some structures with space group C 2, such as MPD No. 998 (Na$_{0.83}$FeSi$_2$O$_6$), which may be viewed as a very similar structure but with a lowered symmetry and deficient atoms at some sites.
• It also returns some structures substituting at Na or Fe sites by other elements. For example, MPD No. 1021 (NaGaSi$_2$O$_6$).
Acknowledgements

• A special thank you to all my current and former students and post-docs
• Also my many wonderful collaborators, mentioned during the talk
• Facilities:
  – APS, CHESS, NSLS, NSLS-II (and people therein)
  – MLNSC, ISIS, SNS (and people therein)
• Funding: DOE-BES and NSF-DMR

HTTP://thebillingegroup.com