The Materials Project: An application of high-throughput computing

NERSC User Day | Feb. 2013

Anubhav Jain
Computational Research Division
Berkeley Lab
Materials development is a key bottleneck to realizing renewable energy

solar PV

electric vehicles

other:
- waste heat recovery (thermoelectrics)
- hydrogen storage
- catalysts/fuel cells
Why Your Smartphone Battery Sucks

If you're hoping your next smartphone will run faster, shine brighter, connect at 4G speed, and last longer on one battery charge, you may be in for a rude surprise. The thirst for battery power in new smartphones and tablets is far outpacing improvements in battery technology. Battery makers are trying to wring the last bits of capacity out of 15-year-old lithium ion technology, while device and app makers seem to be just waking up to the seriousness of the problem. There's an equal share of blame for all parties; meanwhile the immense promise of innovation in mobile devices could come to an early halt due to power limitations, and consumer angst over constantly having to "plug in."

Used to be, you could forget your feature phone's charger at home, go on a long weekend vacation, and—as long as you didn't play hours of Snake—still come home with enough battery life to call a cab. Today, though, we're wedded to our chargers, glaring hawkishly at people who've been hogging airport and coffee shop outlets for too long.

Lithium-ion batteries are both a blessing and a curse when it comes to mobile electronics

The topic of exploding lithium-ion batteries has been debated to death in the wake of massive battery recall over the last couple of years. Amidst the deft public relations maneuvering and finger-pointing, however, the question as to why they explode in the first place is still shrouded in mystery for many.

The most important thing to understand here is that lithium-ion technology is considerably more volatile compared to other forms of rechargeable battery technologies. Defects in the insulating membrane can result in a mini-explosion that rips a battery open to release steam in excess of 600 degrees Fahrenheit.

Manufacturers are aware that it is statistically probable for a lithium-ion to fail, though the calculations employed to sideline the risk are sometimes quite suspect. To determine the mean time between failures (MTBF), manufacturers take a sample of say, 1,000 batteries, which are then used until one fails.
Density Functional Theory is a way to generate materials properties

- \( \text{\sim} 200 \text{ CPU-hours per calc.} \)
- 1-5 calc\'s per material
- Moderately finicky calc\'s.

Examples of real materials design:

G. Hautier, A. Jain, and S.P. Ong,
*From the computer to the laboratory: materials discovery and design using first-principles calculations*
Searching for Li ion battery materials

Li\textsubscript{a} \textit{M} \textsubscript{b} (XY\textsubscript{c})\textsubscript{d}

Li ion source

electron donor / acceptor

structural framework / charge neutrality

examples:
V\textsuperscript{4+/5+}, Fe\textsuperscript{2+/3+}

examples:
O\textsuperscript{-2}, (PO\textsubscript{4})\textsuperscript{3-}, (SiO\textsubscript{4})\textsuperscript{4-}

common cathodes: LiCoO\textsubscript{2}, LiMn\textsubscript{2}O\textsubscript{4}, LiFePO\textsubscript{4}
Compounds screened over time

- No Oxygen (4153)
- Nitrates (61)
- Sulfates (330)
- Carbonates (370)
- Borates (1035)
- Vanadates (1488)
- Phosphates (1609)
- Silicates (1857)

Li Containing Compounds Computed

Jain, Hautier, Moore, Ong, Fischer, Mueller, Persson, Ceder
Comp. Mat. Sci (2011)

 Plain Oxides (9204)
New materials found for Li ion batteries

\[
\text{Li}_3(\text{Fe, Mn})\text{PO}_4\text{CO}_3
\]
Hautier, Jain, Moore, Chen, Moore, Ong, Ceder
Journal of Materials Chemistry (2011)
Chen, Hautier, Jain, Moore, Kang, Doe et al.
Chemistry of Materials (2012)
(patent filed)

\[
\text{Li}_9\text{V}_3(\text{P}_2\text{O}_7)_3(\text{PO}_4)_2
\]
Jain, Hautier, Moore, Kang, Lee, Chen, Twu, Ceder
Journal of the Electrochemical Society (2011)
(patent issued allowance)

\[
\text{Li Mn BO}_3
\]
Kim, Moore, Kang, Hautier, Jain, Ceder
Journal of the Electrochemical Society (2011)
(patent filed)

\[
\text{Li (V, Cr) O}_2
\]
Ma, Hautier, Jain, Ceder
Journal of the Electrochemical Society (submitted)
High voltage materials are less safe: but there is lots of scatter
Once the army is in place, you can attack more problems

Jain, Reihani, Fischer, Couling, Ceder, Green, Chemical Engineering Science 65, (2010).

High-throughput diffusion calculations (NEB)
XANES and EXAFS spectra codes
Multiscale modeling / porous materials
Rapid Phase Diagrams
Bulk modulus

collaboration with Cambridge
another collaboration at LBL
www.materialsproject.org

SHOW ME THE MATERIAL!
3500 users and growing!

(*Old chart for 3150+ users)
Design of alloy anodes for Li ion batteries
compared measurements to calculations)

Tran, Obrovac
Journal of the Electrochemical Society (2011)

Computational screening of perovskite metal oxides for optimal solar light capture
crystal structures for candidate stability screening)

Castelli, Olsen, Datta, Landis, Dahl, Thygesen, Jacobsen
Energy & Environmental Science (2011)

DFT Formation Enthalpies of Rare Earth Orthophosphates
find scaling factor for computed heat of formation)

Rustad
American Minerologist (2012)
"Thanks. Your product is astounding. I redid work for a recent paper that took weeks in about 15 minutes! I guess this is truly “transformative” science in the NSF sense!"
“Toruk makto”:
How to get Hopper to run high-throughput?
How we set up our NERSC infrastructure

Rocket launcher
(running on Hopper head node)

thruput job

thruput job

thruput job

thruput job

thruput job

thruput job
What happens when job starts running?

Job wakes up when PBS runs it

Grabs the latest job description from an external DB

Runs the job based on DB description
Need workflow to get multiple properties

Al₂O₃

Structure

Energy/Charge

Band structure

DOS
Workflow handles logic and ‘Plan B’

**CALCULATION PLAN**

- **Structure**
  - Increment mesh until converged
  - Copy structure and charge; Transform structure

- **Energy/Charge**
  - Copy structure and charge

- **Band structure**
  - Copy structure and charge

- **DOS**

How should the system respond if:

- A node crashes?
- The calculation does not converge?
- Early calculations indicate a material is more/less interesting?
- We might have already done this calculation?

Specify **DYNAMIC** workflow!
A calculations superqueue distributes jobs over multiple resources

Perform duplicate check to see if job exists in SuperQueue

Structure/Energy

Static

Band structure

DOS

Job already submitted

Submit!

Will run the job!

work done with M. Kocher
Which workflow do we use?

Developed our own workflow software based on MongoDB

In Production:

FireWorks/Rockets/Engines

Early but active development:

Interested in FireWorks?

http://pypi.python.org/pypi/FireWorks/

or Google ‘PyPi FireWorks’

Open-source
The impact of the workflow system

- We completed over 10 million CPU-hours of high-throughput DFT computations in under 6 months at NERSC

- Used 2300+ CPUs on average, 24/7

- Our army is large!
HPC vs. HT – there’s still a mismatch

Our jobs need to gain weight (and lose height)!

“Fat” concurrency (16000+ nodes)  “Thin” concurrency (1 node)

Short walltime ~hours

Long walltime ~days

*Fatness not drawn to scale
The height issue – workflows sometimes help

Single script that does multiple tasks in succession

Atomic tasks connected by workflow software
The height issue:  
Need to implement checkpoint/restart
From skinny to fat: PBS autopacking to address concurrency

Need checkpoint/restart to fill leftover time

PBS

Trying to build this into FireWorks in the future
Conclusion / Final thoughts on HT

- Materials Project is an example of ‘Science Gateway’ based on HT-computed data
- At NERSC, viable options for HT already exist
  - We use thruput queue heavily (you may have noticed)
  - new tools under development
- Different groups are developing solutions
  - Will be interesting to know what ‘sticks’, or if equilibrium will be unique workflow per project
  - Would love to discuss it with you!
Thank you!

- **Materials Project**
  - Kristin Persson, Gerbrand Ceder, Shyue Ping Ong, Geoffroy Hautier, Michael Kocher, David Skinner, Dan Gunter, Shreyas Cholia, NERSC staff

- **Funding**
  - DOE, NSF, 2010 LBL LDRD

- **Contact:**
  - ajain@lbl.gov
Backup slides
Deviations from a simple linear workflow

**Detours**
(about 10-20% of jobs fail and must be rerun with different input parameters)

**Branches**
(based on the result of a calculation, the entire workflow might need to be modified)

**Duplicate Job detection**
(if two workflows contain an identical step, ensure that the step is only run once)
Example: Analyzer object decides how to proceed once a VASP calculation finishes

BEGIN

Run the hive import (initialize outputs)

Run all error detection codes

Critical errors detected?

Hive says successful?

Num launches* > 3?

Return TERMINATE

Return DETOUR

Is this the 3rd detour?

Create new FW with specific error fixes

Create new FW with default second parameter set.

Known error fixes?

Num detours > 3?

STOPCAR exists

Return REPEAT

Return TERMINATE
# Examples of high-throughput studies

<table>
<thead>
<tr>
<th>Application</th>
<th>Researcher</th>
<th>Search space</th>
<th>Candidates</th>
<th>Hit rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scintillators</td>
<td>Klintenberg et al.</td>
<td>22,000</td>
<td>136</td>
<td>1/160*</td>
</tr>
<tr>
<td></td>
<td>Curtarolo et al.</td>
<td>11,893</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Topological insulators</td>
<td>Klintenberg et al.</td>
<td>60,000</td>
<td>17</td>
<td>1/3500*</td>
</tr>
<tr>
<td></td>
<td>Curtarolo et al.</td>
<td>15,000</td>
<td>28</td>
<td>1/535*</td>
</tr>
<tr>
<td>High T&lt;sub&gt;C&lt;/sub&gt; superconductors</td>
<td>Klintenberg et al.</td>
<td>60,000</td>
<td>139</td>
<td>1/430*</td>
</tr>
<tr>
<td>Thermoelectrics</td>
<td>Curtarolo et al.</td>
<td>2,500</td>
<td>20</td>
<td>1/125*</td>
</tr>
<tr>
<td>1-photon water splitting</td>
<td>Jacobsen et al.</td>
<td>19,000</td>
<td>20</td>
<td>1/950*</td>
</tr>
<tr>
<td>2-photon water splitting</td>
<td>Jacobsen et al.</td>
<td>19,000</td>
<td>12</td>
<td>1/1585*</td>
</tr>
<tr>
<td>Transparent shields</td>
<td>Jacobsen et al.</td>
<td>19,000</td>
<td>8</td>
<td>1/2375*</td>
</tr>
<tr>
<td>Hg adsorbers</td>
<td>Bligaard et al.</td>
<td>5,581</td>
<td>14</td>
<td>1/400*</td>
</tr>
<tr>
<td>HER catalysts</td>
<td>Greeley et al.</td>
<td>756</td>
<td>1</td>
<td>1/756</td>
</tr>
<tr>
<td>Li ion battery cathodes</td>
<td>Ceder et al.</td>
<td>20,000</td>
<td>4</td>
<td>1/5000</td>
</tr>
</tbody>
</table>

Entries marked with * have not experimentally verified all the candidates. Hit rates are optimistic because the search space is often restricted based on intuition.