Accelerated Materials Design
through High-throughput First-Principles Calculations and Data Mining

Kristin Persson
Environmental and Energy Technology Division (LBNL)

Outline

• Materials Science – to the rescue for a sustainable energy future
• A crash course on density functional theory
• Not a exascale poster child
• We solved the computing (kinda) – does data-driven materials design work??
• The Materials Project - Towards a Materials Genome

Engineered Materials Enable Society

How are New Materials Invented?

“Edison Style”

When looking for a light bulb filament, Edison tried about 3,000 materials

... And he didn't find the best one ...!
Materials Design: Hollywood Style

Materials Innovation Timeline

How to compute real world materials properties?

Computational Materials Science and First-Principles Calculations
Aim of ab initio calculations

- Atomic Numbers
- Solve quantum mechanics for the material
- Predict physical and chemical properties of systems

Standard DFT – steady state

\[ \hat{H}\Psi = E\Psi \]

As you can see, quantum mechanics is "simply" an eigenvalue problem.

Summary of problem to solve

Assume that the nuclei (Mass M) are at: \( R_1, R_2, \ldots, R_N \)
Assume that the electrons (mass \( m_e \)) are at: \( r_1, r_2, \ldots, r_m \)

where

\[ \hat{H}_{N,e} = \hat{T}_N + \hat{T}_e + \hat{V}_{N-N} + \hat{V}_{N-e} + \hat{V}_{e-e} \]

Summary of problem to solve

Assume that the nuclei (Mass M) are at: \( R_1, R_2, \ldots, R_N \)
Assume that the electrons (mass \( m_e \)) are at: \( r_1, r_2, \ldots, r_m \)

where

\[ \hat{H}_{N,e} = \hat{T}_N + \hat{T}_e + \hat{V}_{N-N} + \hat{V}_{N-e} + \hat{V}_{e-e} \]
Electrons are difficult!

• The mathematical difficulty of solving the Schrödinger equation increases rapidly with \( N \)
• The number of computations scales as \( e^N \)
• With modern supercomputers we can solve this directly for a very small number of electrons (maybe 4 or 5 electrons)

Materials contain of the order of \( 10^{26} \) electrons

Quantum power: Density Functional Theory

\[
H = \sum_{i=1}^{N_e} \nabla_i^2 \mathcal{E}_i + \frac{1}{2} \sum_{i \neq j}^{N_e} \frac{1}{r_{ij}}
\]

- Kinetic Energy
- Interaction with nucleus
- Interaction between electrons
- Many electrons
- One electron
- Replace e-e interaction by average potential

Approximated in the Local Density Approximation (LDA) or Generalized Gradient Approximation (GGA) to Density Functional Theory (DFT)

\( V_{\text{eff}} \) = average electrostatic potential from other electrons + exchange effect (Pauli principle) + correlation effects

Many properties can be computed

Photovoltaics, Thermoelectrics, Energy Storage, Hydrogen, Catalysts, CO\(_2\) capture...

Computations are scalable (or are they?)

\[
\hat{H} \psi = E \psi
\]

Total energy
Optimized structure
Magnetic ground state
Charge density
Band structure / DOS
... etc
High Throughput Scientific Computing

So you have access to 150K cores, now what?

Gartner Hype Cycle: graphic representation of the maturity and adoption of technologies and applications

Bummer – ‘exascale’ not working for DFT...

DFT codes are trivially parallelizable over k-points

BUT after every reciprocal k point calculation – all the energies (information) have to be assembled to calculate charge density and total energy… happens hundreds of times per calculation.

too much communication between nodes!

More sophisticated parallelization schemes exist, but fact remains – no DFT code scales better than 30-40 nodes

So what do we do? We run one material per node...no intercommunication needed and large # nodes can be requested
non-predictable walltimes...

- Our computations have rather unpredictable runtimes
- Think swarms of workers – one is very slow, another one terrible efficient...

High-Throughput Materials Computing

- Requires scheduler or run-time tools to facilitate running large numbers of jobs with variable duration time
- Inverse of typical large-scale simulations of inter-connected tasks (climate, astronomy, ...)

Shared Resources: Some Challenges are Unavoidable

- Long “small” jobs and short “large” jobs are natural enemies, hard to co-schedule
- Don’t get mad, get even (or get things done)
- Get the throughput you want
- Read the queue policies

Example of challenge overcome: HPC vs HT barrier

Combustion Modeling
High-Throughput Materials Science
Climate Simulation

You must be this tall to ride the SuperComputer
Why does this matter?
• NERSC offers us 40 million CPU-hours / year
  – A 16-core workstation would take 285 years to produce that data....
• But we can only leverage that time if we follow their policies
• We will never get close with small jobs
  – queued job limit
  – walltime limit
• Need to play by their rules, which are designed for massive single simulations

Automatic job packing

Ok – we fixed the computing – what about new materials???
High-throughput Materials Design: Alkaline batteries

Finding higher energy density cathode
SPECS:
Higher energy density
Both reactants and products stable in high molar KOH
Need to model reactions at pH = 15

Alkaline cell reaction

2MnO₂ + H₂O + Zn → 2MnOOH + ZnO


The Screening Strategy

Materials Design: Solid Stability Patterns

Materials Design: KOH Instability
Candidate Compounds and Design Rules

- >130,000 compounds considered
  - >30,000 known from ICSD
  - >100,000 new generated

Tier 1
- >1500 compounds
  - >1.1 V < Ave voltage < 2.2 V
  - Energy density > 1.7Wh/cc

Tier 2
- >700 compounds
  - Reactant stable in air
  - Reactant stable in 9N KOH
  - Product stable in 9N KOH

Predicted alkaline cathode material performance

End result - 200 compounds predicted to outperform current cathode AND are predicted stable (through entire reaction) in 9M KOH

Novel Materials for Li-ion Batteries

- Novel stable alkaline batteries
- Novel class of Li ion electrodes
- Improved transparent conductors

Lithium-ion batteries for electric vehicles

1989
- Doc: “No, no, no, this sucker’s electrical, but I need a nuclear reaction to generate the 1.21 gigawatts of electricity!”

2012
- 70%-80% of cell cost is materials
- Small car: 4 miles/kWh; 100 miles = $10,000 battery
- Higher kWh/volume reduces cost per kWh of stored energy

It’s just the beginning...

Completely new materials predicted and synthesized based on computational predictions
The Materials Project: A Growing Public Resource

Today’s Status:
- Over 66,000 compounds
- Growing monthly
- Multiple property sets
- Multiple tools

MP Capabilities
What typically slows down computations?

Dynamic workflows can be programmed

Detours
(based on result, insert a new step before resuming)

Branches/Additions
(based on result, modify or add to the workflow)

Duplicate Job detection
(if two workflows contain an identical step,
ensure that the step is only run once)

http://pythonhosted.org/FireWorks

Community can write any workflow in FireWorks
→
We can automate it over the world’s most powerful computers!
Add more steps later without having to rerun everything
How to Build Robust Software that Lasts?

Does anyone remember how to run `GoBabyMonte.c` ???

Nah – Bob left in 1996....

Software Management Philosophies

- Open-source
  - More eyes => robustness
  - Contributions from all over the world
- Benevolent dictators
  - Unified vision
  - Quality control
- Clear documentation
  - Prevent code rot
  - More users
- Test, test, test
  - Continuous integration to ensure code is always working

• Robust materials analysis
• Self-healing error recovery
• Smart workflow management
• Pymatgen
• Custodian
• Fireworks
World Wide Usage < 18,000 registered users

MP Going Interactive

US Professor: “I was wondering if you ever considered a way for the community to contribute to the Materials Project? … I really like the approach and I think that this is exactly where (especially computational) materials science has to go.”

MPComplete: Crowd-sourcing MP

Launched Sept 2015

- **Motivation:** new compounds supplied directly by community
- Users suggest structures; MP checks for uniqueness and runs full suite of calculations.
- Ensures user-relevant materials with consistent provenance

Powered by XSEDE

MPComplete usage; every day

>500 submissions by >100 users since launch, resulting in >300 new materials
MPComplete Use Cases

- One-at-a-time, or bulk:
  - One user contributed four new lead halide PV compounds, one at a time. (design)
  - One user submitted 64 ABN$_3$ perovskites, associating with publication. (data sharing)
  - Another user submitted 131 structures to check for stability against all MP compounds (validation and data sharing)

World-Wide Resource

Haynes International: "The correlation between softening temperature and your calculated data is high – can we add novel alloy compositions?"

Cymbet: "I am so incredibly happy an effort like this exists now... Please please don't stop growing!"

Toyota: "Materials Project is a wonderful project. Please accept my appreciation to you to release it free and easy to access."

Micron: Your project is inspirational. Please come to Boise and educate our technical community"

Materials Project Growth

- Press releases in 2011
- Release of Pourbaix App and new Web interface
- Release of elastic constants, NMGC App, band structures...
- Past year: ≈ 19 new users/day
- ≈ 7,000 new users
- > 10,000 users

From single entities...
Thanks for your attention!