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

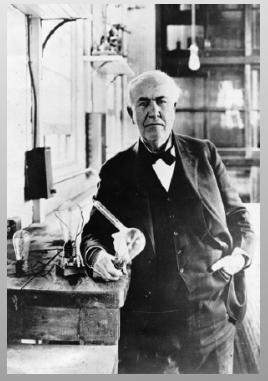
How are New Materials Invented?

"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

PV | TCOs | Permanent magnets | Solar Fuel

Thermoelectrics | Energy Storage | …

18 Years...from the average new materials discovery to commercialization

How to accelerate the innovation and development timeline ?

1930 1940 2000

invented → Teflon ← Lithium ion ← Sony

S. Whittingham

Velcro

Titanium

Polycarbonate

Diamond-like Thin Films

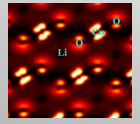
GaAs

Amorphous soft magnets

Materials Data from: Eagar, T.; King, M. Technology Review (00401692) **1995**, 98, 42.

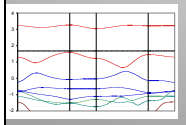
How to compute real world materials properties?

Quantum Mechanics



Optimize

$$\langle \Psi | \text{better battery} | \Psi^* \rangle$$



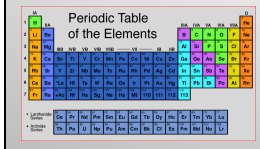
"E = 325.67 kJ"

Engineering Properties

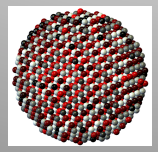
Corrosion, strength, energy density, …

Computational Materials Science and First-Principles Calculations

Aim of *ab initio* calculations



$$\hat{T}\Psi_{\mathrm{MB}} + \hat{V}\Psi_{\mathrm{MB}} = -i\frac{d\Psi_{\mathrm{MB}}}{dt},$$



Atomic Numbers

↓

Solve quantum mechanics for the material

↓

Predict physical and chemical properties of systems

Standard DFT – steady state

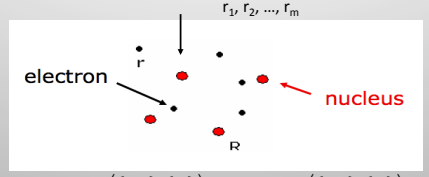
$$\hat{H}\underline{\Psi} = E\underline{\Psi}$$

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

Summary of problem to solve

Assume that the nuclei (Mass M_I) are at: $R_1, R_2, \ldots, R_N$

Assume that the electrons (mass m_e) are at: $r_1, r_2, \ldots, r_m$



$$\hat{H}_{N,e}\Psi_{N,e}\left(\{R_I\},\{r_i\}\right) = E_{N,e}\Psi_{N,e}\left(\{R_I\},\{r_i\}\right)$$

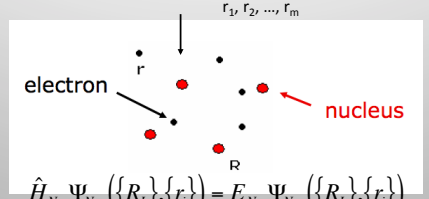
where $\hat{H}_{N,e} = \underbrace{\hat{T}_N + \hat{T}_e}_{\text{kinetic energy}} + \hat{V}_{N-N} + \hat{V}_{N-e} + \hat{V}_{e-e}$

- $\hat{V}_{N-N}$: nucleus – nucleus interaction
- $\hat{V}_{N-e}$: nucleus - electron interaction
- $\hat{V}_{e-e}$: interaction between electrons

Summary of problem to solve

Assume that the nuclei (Mass M_I) are at: $R_1, R_2, \ldots, R_N$

Assume that the electrons (mass m_e) are at: $r_1, r_2, \ldots, r_m$



$$\hat{H}_{N,e}\Psi_{N,e}\left(\{R_I\},\{r_i\}\right) = E_{N,e}\Psi_{N,e}\left(\{R_I\},\{r_i\}\right)$$

where $\hat{H}_{N,e} = \underbrace{\hat{T}_N + \hat{T}_e}_{\text{kinetic energy}} + \underbrace{\hat{V}_{N-N}}_{\text{separable}} + \hat{V}_{N-e} + \underbrace{\hat{V}_{e-e}}_{\text{Non-separable}}$

- $\hat{V}_{N-N}$: nucleus – nucleus interaction
- $\hat{V}_{N-e}$: nucleus - electron interaction
- $\hat{V}_{e-e}$: interaction between electrons

Electrons are difficult!

- The mathematical difficulty of solving the Schrodinger 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*

Kinetic Energy

Interaction with nucleus

Interaction between electrons

$$H = \sum_{i=1}^{N_e} \nabla_i^2 + \sum_{i=1}^{N_e} V_{nuclear}(r_i) + \frac{1}{2} \sum_{i}^{N_e} \sum_{j \neq i}^{N_e} \frac{1}{|r_j - r_i|}$$

Many electrons

Replace e-e interaction by average potential

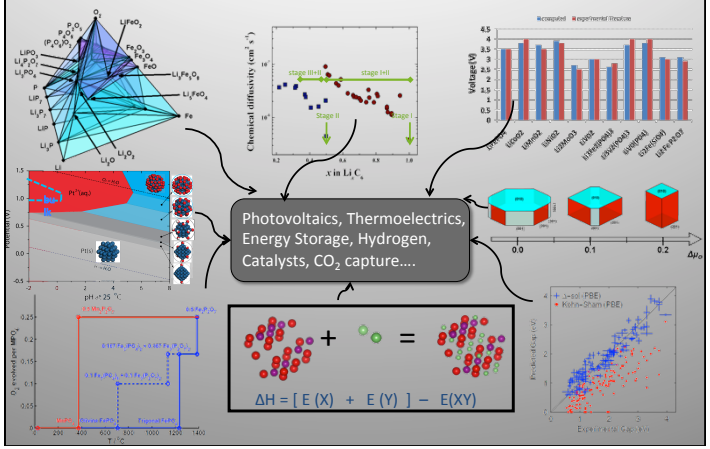
$$H = \sum_{i=1}^{N_e} \nabla_i^2 + \sum_{i=1}^{N_e} V_{nuclear}(r_i) + \sum_{i=1}^{N_e} V_{effective}(r_i)$$

One electron

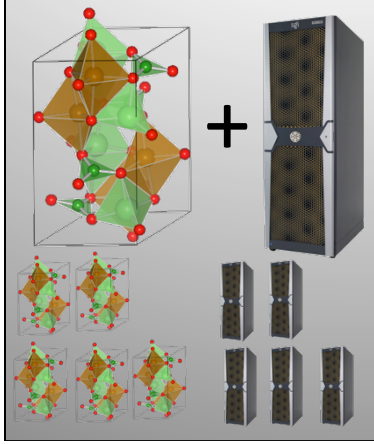
Approximated in the Local Density Approximation (LDA) or Generalized Gradient Approximation (GGA) to Density Functional Theory (DFT)

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

Many properties can be computed



Computations are scalable (or are they?)

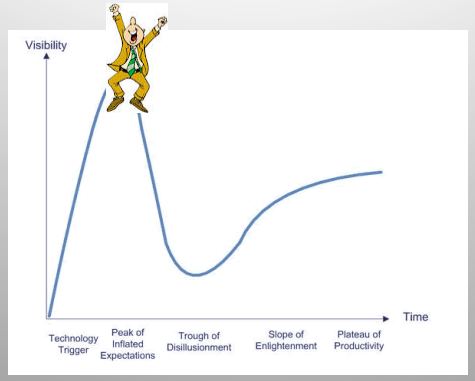


$$\hat{H}\psi = E\psi$$

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.

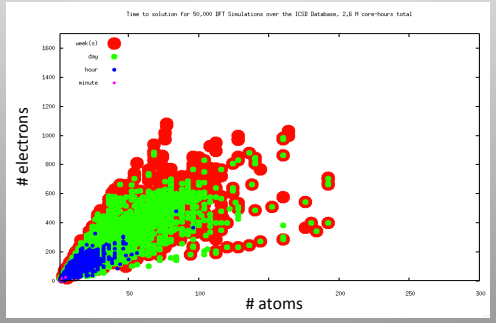
too much communication between nodes!

More sophisticated parallellization 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…

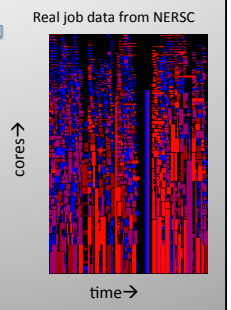


High-Throughput Materials Computing

- 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, …)

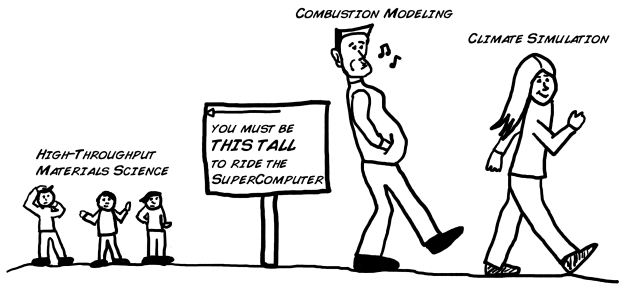
Shared Resources: Some Challenges are Unavoidable

- 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



long queue wait
short queue wait

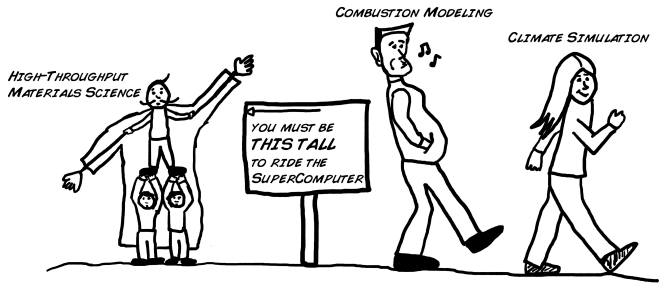
Example of challenge overcome: HPC vs HT barrier



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

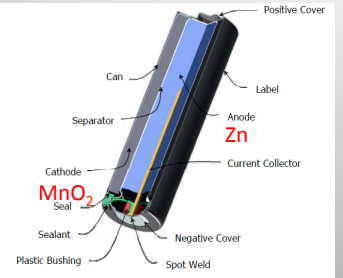
Automatic job packing



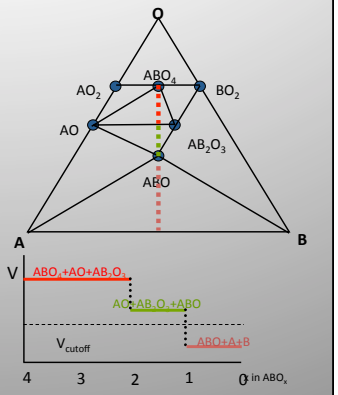
Ok – we fixed the computing – what about new materials???

High-throughput Materials Design: Alkaline batteries

Finding higher energy density cathode



$MnO_2+H_2O+Zn \rightarrow MnOOH + ZnOH_2$



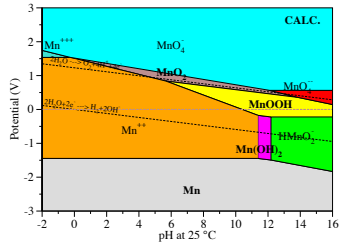
SPECS:

Higher energy density

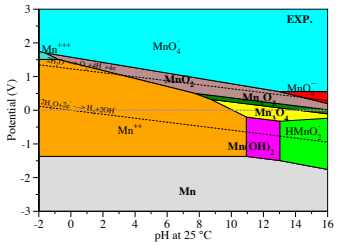
Both reactants and products stable in high molar KOH

Need to model reactions at pH = 15

Our calcs



exp

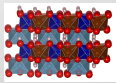


Alkaline cell reaction

$2MnO_2+H_2O+Zn \rightarrow 2MnOOH + ZnO$

M. Pourbaix, "*Atlas of Electrochemical Equilibria in Aqueous Solution*", Pergamon Press, Paris (1966).

The Screening Strategy



Database

Known compounds *30,000* → permutation strategy → New compounds *~ 100,000*

~ 100,000 - 200'000

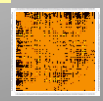
- ✓Solid Stability
- ✓Solubility
- ✓Voltage
- ✓Capacity

Initial screening (phenomenological)

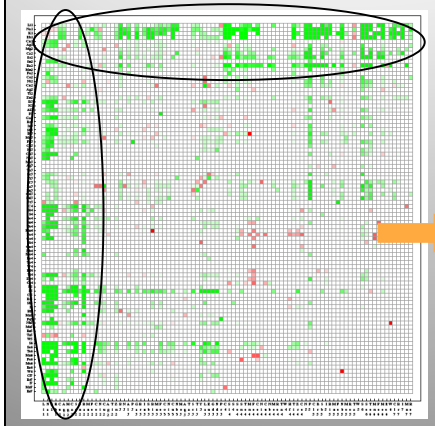
≈ 10,000 - 20'000

Computational Screening → Data mining

Candidate materials



Materials Design: Solid Stability Patterns

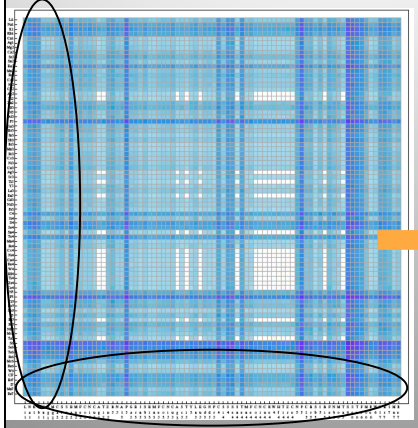


- High negative formation energy
- Negative formation energy
- Positive formation energy
- High positive formation energy

Which cation stabilizes another cation

Design stable solids

Materials Design: KOH Instability



- High dissolution energy
- Moderate dissolution energy
- No dissolution energy

Which cations dissolve

Design stable solids at high pH

Candidate Compounds and Design Rules

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

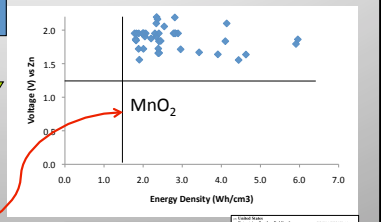
Tier 1

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

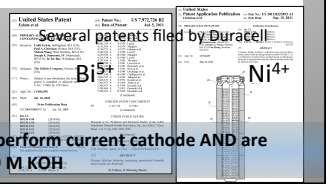
Tier 2

≈ 200 compounds
- ✓ Reactant stable in air
- ✓ Reactant stable in 9N KOH
- ✓ Product stable in 9N KOH

Predicted alkaline cathode material performance



Several patents filed by Duracell



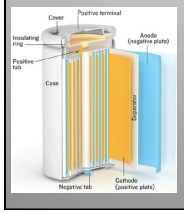
End result – 200 compounds predicted to outperform current cathode AND are predicted stable (through entire reaction) in 9 M KOH

Lithium-ion batteries for electric vehicles

1989

2012

Doc: "No, no, no, no, this sucker's electrical, but I need a nuclear reaction to generate the 1.21 gigawatts of electricity ! "

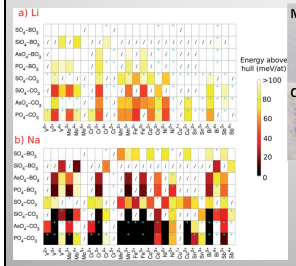


70%-80% of cell cost is materials

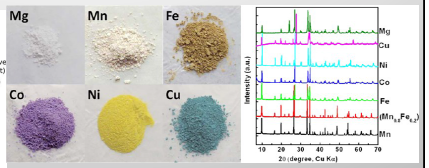
Small car: 4 miles/kWh; 100 miles = $10, 000 battery

Higher kWh/volume reduces cost per kWh of stored energy

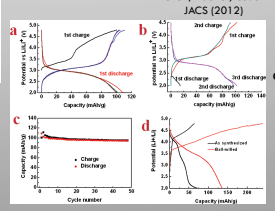
Novel Materials for Li-ion Batteries



Hautier, Jain, Chen, Moore, Ong, Ceder
J. Materials Chemistry (2011)



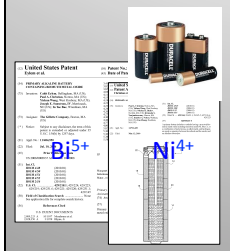
Chen, Hautier, Ceder
JACS (2012)



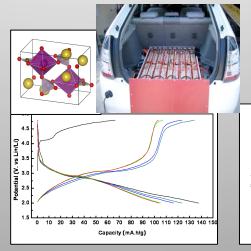
Chen, Hautier, Jain, Moore, Kang, Doe et al.
Chemistry of Materials (2012)

It's just the beginning…

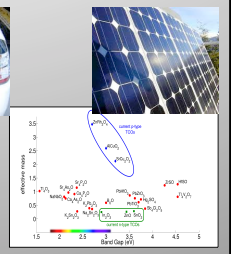
Completely new materials predicted and synthesized based on computational predictions



2005:
Novel stable alkaline batteries

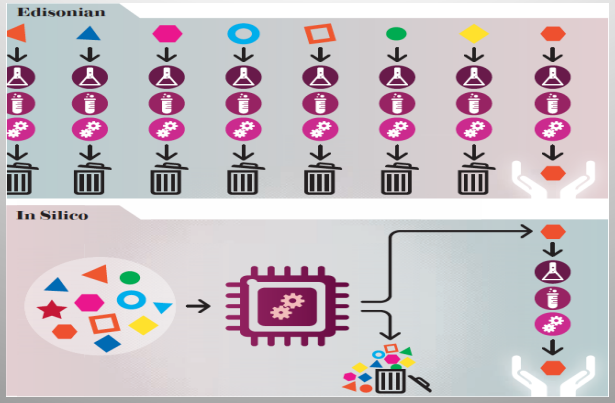


2011:
Novel class of Li ion electrodes



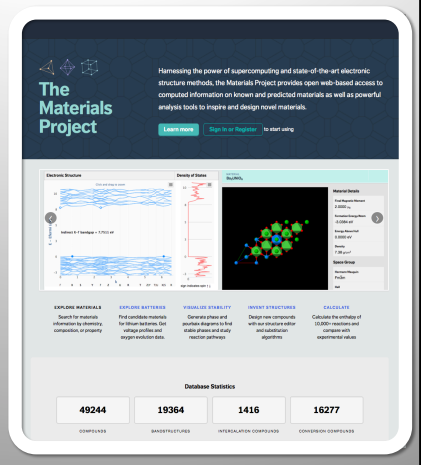
2013:
Improved transparent conductors

The Materials Project: A Growing Public Resource

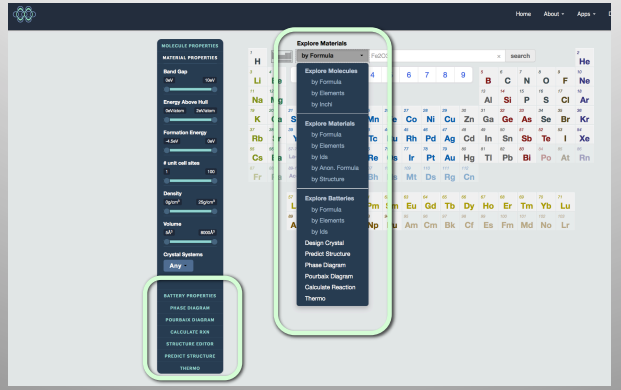


Today's Status:

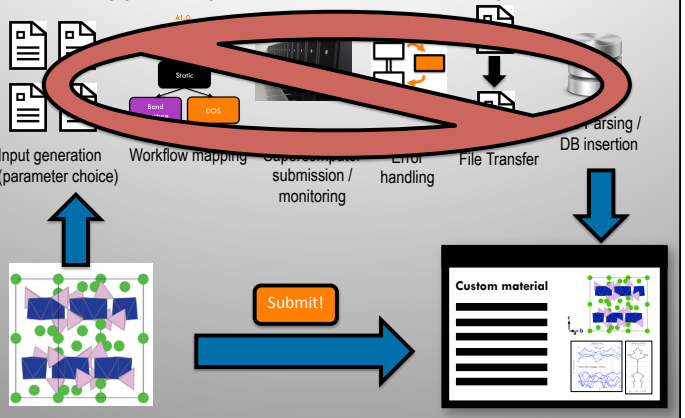
- ❑ Over 59,000 compounds
- ❑ Growing monthly
- ❑ Multiple property sets
- ❑ Multiple tools



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)

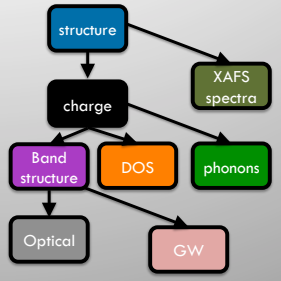
http://pythonhosted.org/FireWorks

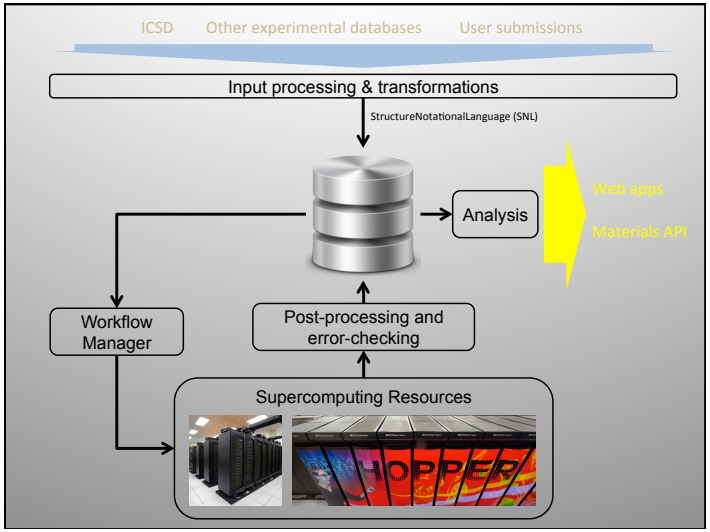
Community can write any workflow in FireWorks

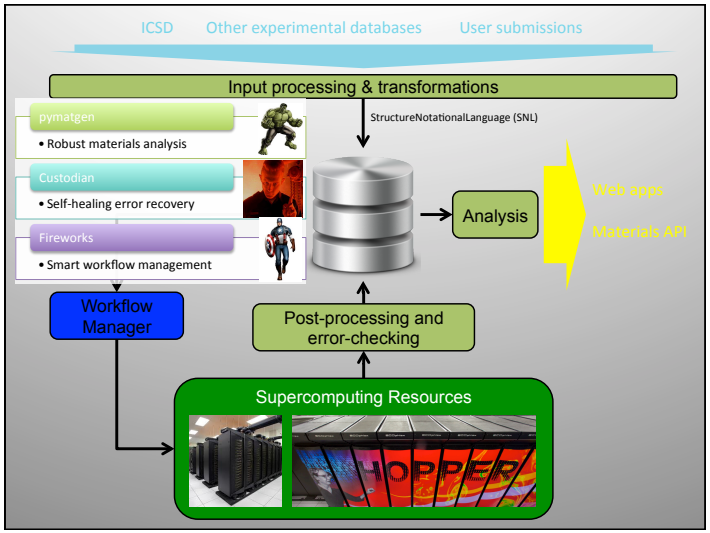
→

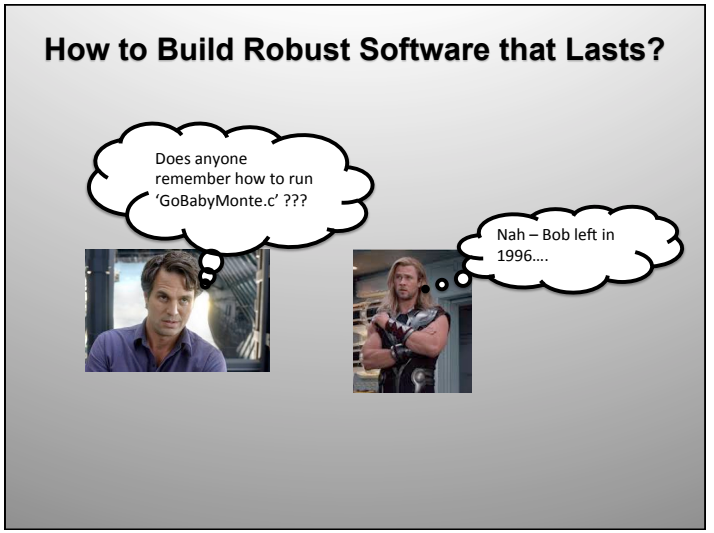
We can automate it over the world's most powerful computers!

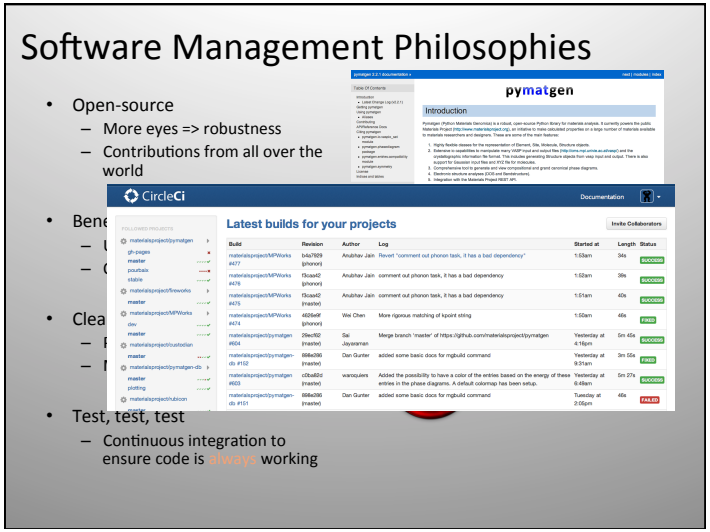
Add more steps later without having to rerun everything

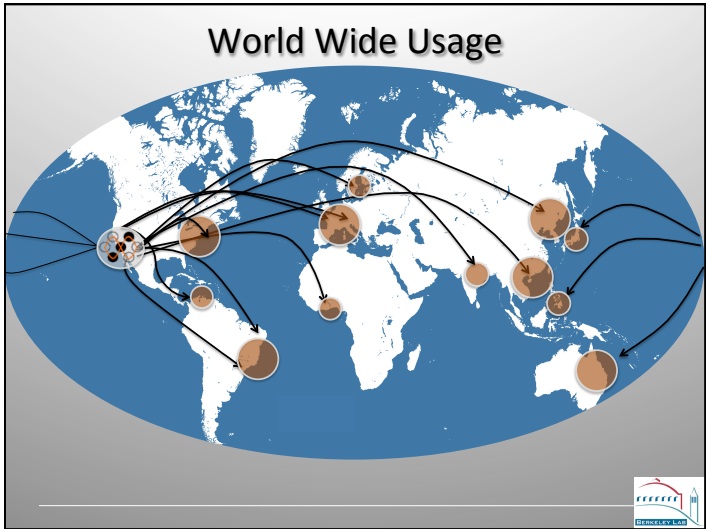


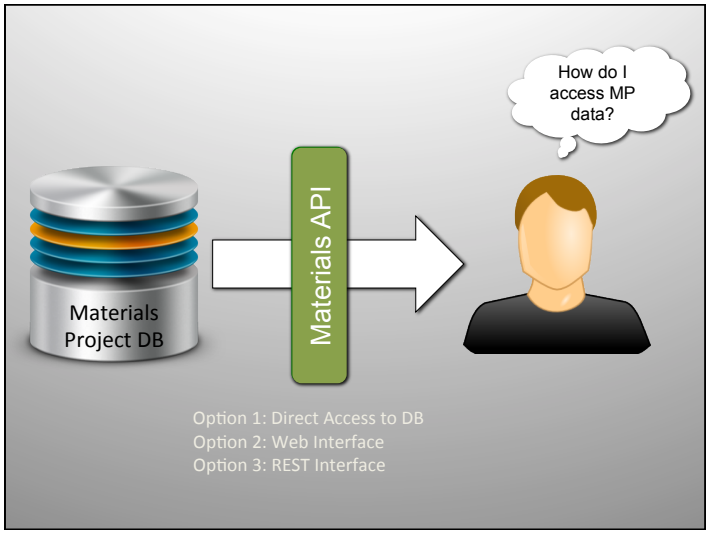



ICSD
Other experimental databases
User submissions
Input processing & transformations
pymatgen
• Robust materials analysis
Custodian
• Self-healing error recovery
Fireworks
• Smart workflow management
StructureNotationalLanguage (SNL)
Analysis
Web apps
Materials API
Workflow Manager
Post-processing and error-checking
Supercomputing Resources
HOPPER


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
Bene
Clea
Test, test, test
Continuous integration to ensure code is always working
pymatgen
Introduction
CircleCi
Documentation
Latest builds for your projects
Invite Collaborators


World Wide Usage



A REST API maps a URL to a resource.

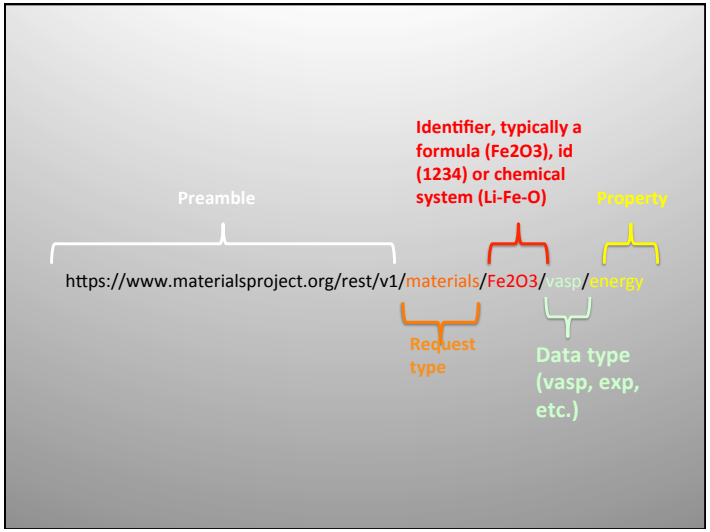
Example:

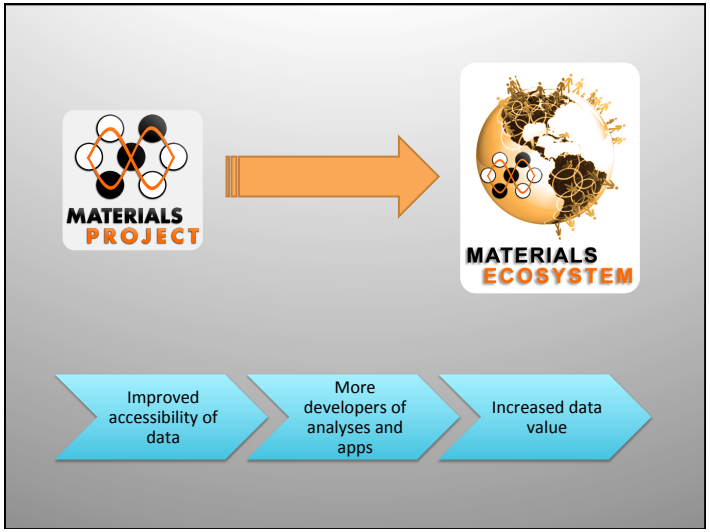
GET https://api.dropbox.com/1/account/info

Returns information about a user's account.

Methods: GET, POST, PUT, DELETE, etc.

Response: Usually JSON or XML or both

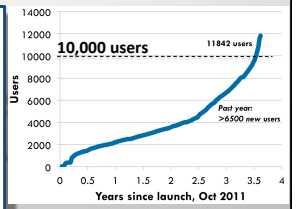


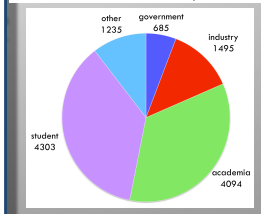


The Materials Project Impact

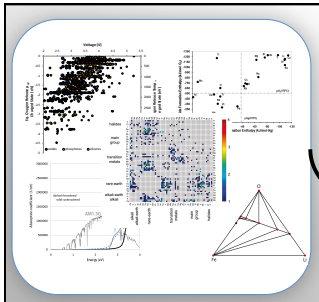
- **Launched online Oct 2011**
- **Users**: ~12,000 registered; **~300 users log on every day**
- **31 published internal papers**
- **55+ publications from non-funded PIs/groups citing significant use of Materials Project information/ algorithms in paper**
- **277** publications citing the **Materials Project as a community resource;**
- 8 million records downloaded through API with >100 distinct users
- **Active Partnerships:** JCAP, JCESR, NREL-EFRC, Minnesota MGI Center on Porous Materials; U Leuven, U Wisconsin-Madison (MaterialsHub), EPFL, …
- **Course-ware** : UC San Diego, UC Davis, UC Irvine, U Michigan, Johns-Hopkins, Cornell, MIT,UC Santa Barbara…

- **Companies:** Toyota, Sony, Bosch, 3M, Honda, Samsung, LG Chem, Dow Chemicals, GE Global Research, Intermolecular, Applied Materials, Energizer, Advanced Materials, General Motors, Corning, DuPont, Nippon Steel, L'Oreal USA, Caterpillar, HP, Unilever, Lockheed Martin, Texas Instruments, Ford, Bose, Sigma-Aldrich, Siemens, Raytheon, Umicore, Seagate, …





From single entities…



… towards a materials genome

Thanks for your attention!