

Yonsei University Graduate Class

Energy Materials: Design, Discovery and Data

Exploring the Materials Hyperspace

Dan Davies

PhD Student

University of Bath



<https://wmd-group.github.io>



@danwdavies

About Me

- MChem University of Bath
- PhD student at the CSCT
- Interested in high-throughput computational screening and design of materials

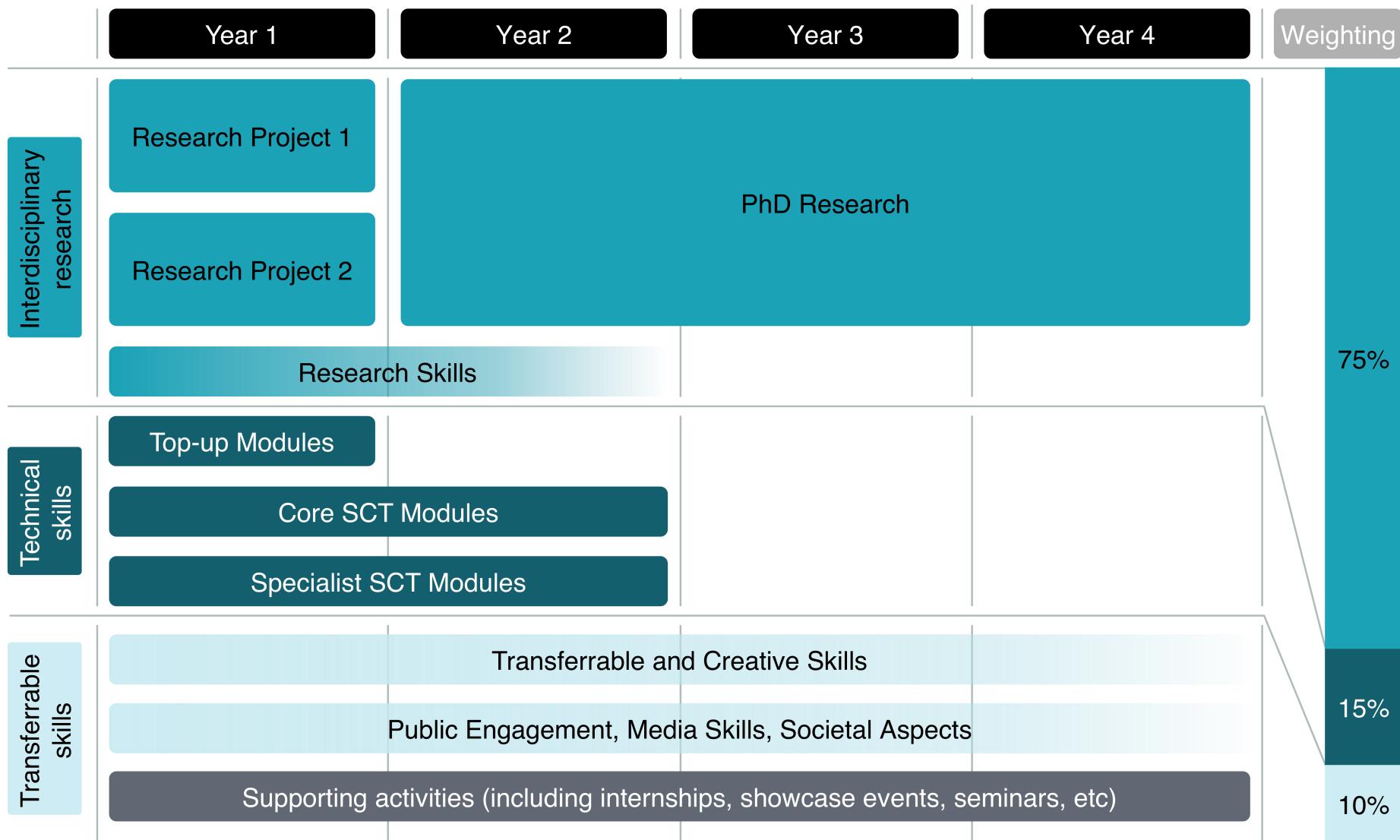


UNIVERSITY OF
BATH



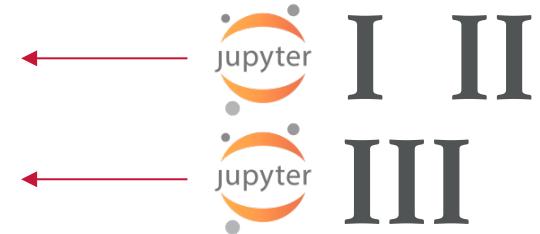
Centre for
Sustainable
Chemical Technologies

About Me



Talk Outline: The Materials Hyperspace

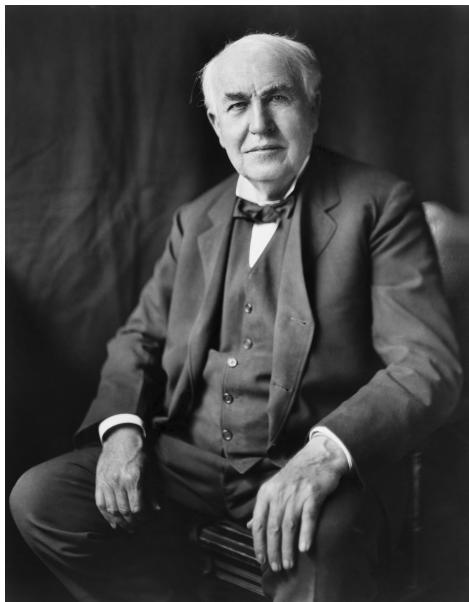
1. Data for Materials Discovery
2. Exploring Chemical Space
3. Interacting with Databases



Class Question

How were the vast majority
of materials we use today
discovered?

Traditional Materials Discovery



- Trial and ~~error~~
- Trial and improvement

“I have not failed, I have just found 10,000 ways that won’t work.”

Traditional Materials Discovery

Teflon®



Safety glass



Silly putty



Traditional Materials Discovery

Teflon®



“A white solid material was obtained, which was supposed to be a polymerized product.”

Traditional Materials Discovery

Teflon®



“It is insoluble in cold and hot water, acetone...”

Traditional Materials Discovery

Teflon®



“It is insoluble in cold and hot water, acetone, Freon 113, ether, petroleum ether, alcohol, pyridine, toluene ethyl acetate...”

Traditional Materials Discovery

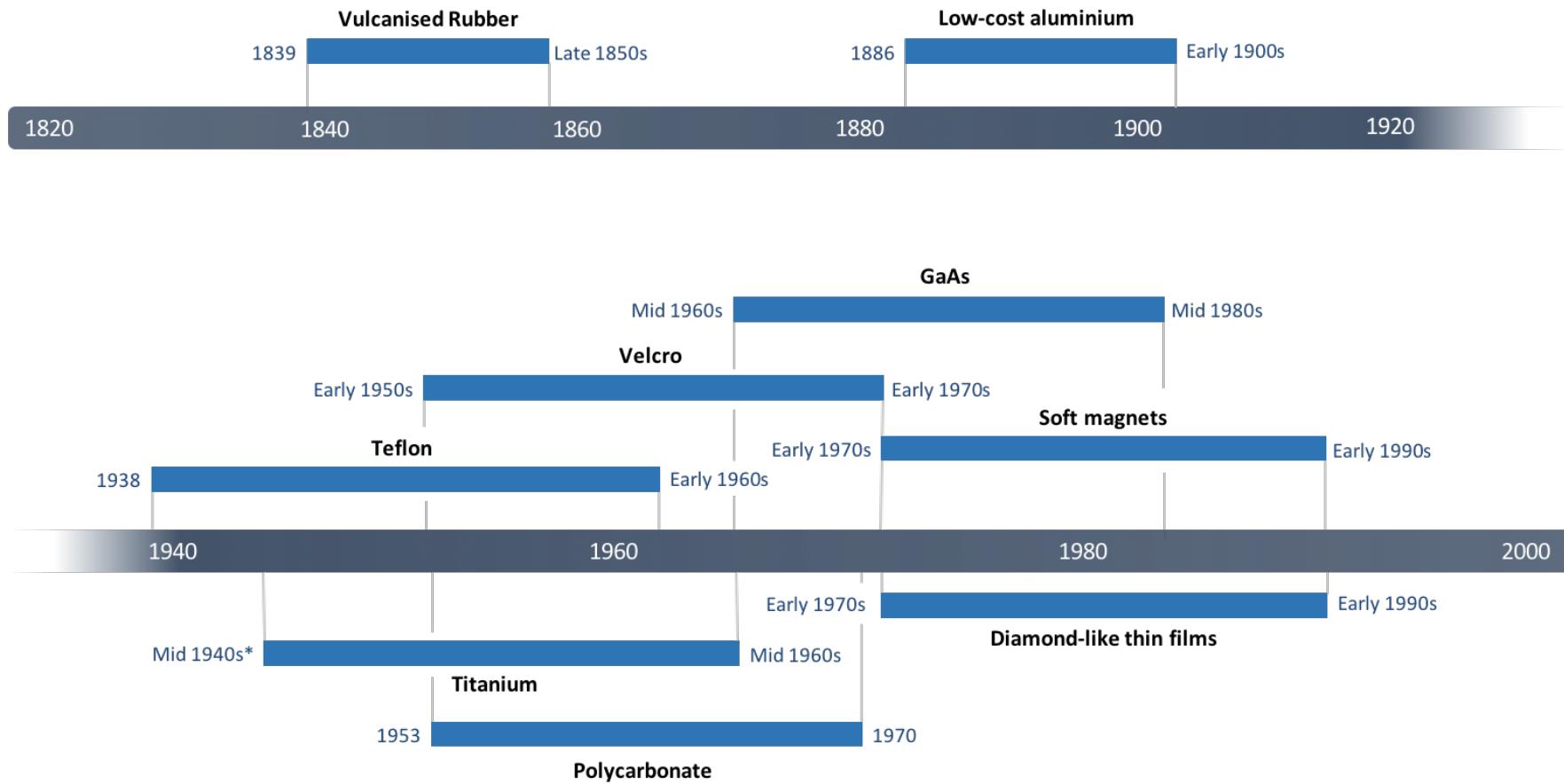
Teflon®



“It is insoluble in cold and hot water, acetone, Freon 113, ether, petroleum ether, alcohol, pyridine, toluene ethyl acetate, concentrated sulfuric acid, glacial acetic acid, nitrobenzene, isoamyl alcohol, ortho-dichlorobenzene, sodium hydroxide, and concentrated nitric acid.”

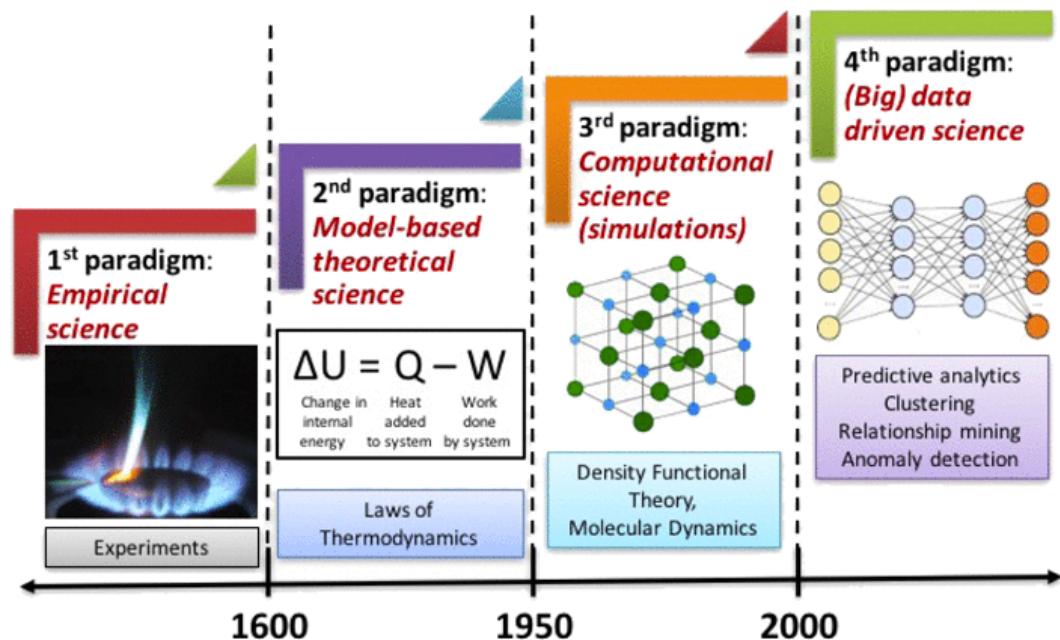
Getting Materials to Market

Time between invention and widespread commercialisation of materials:

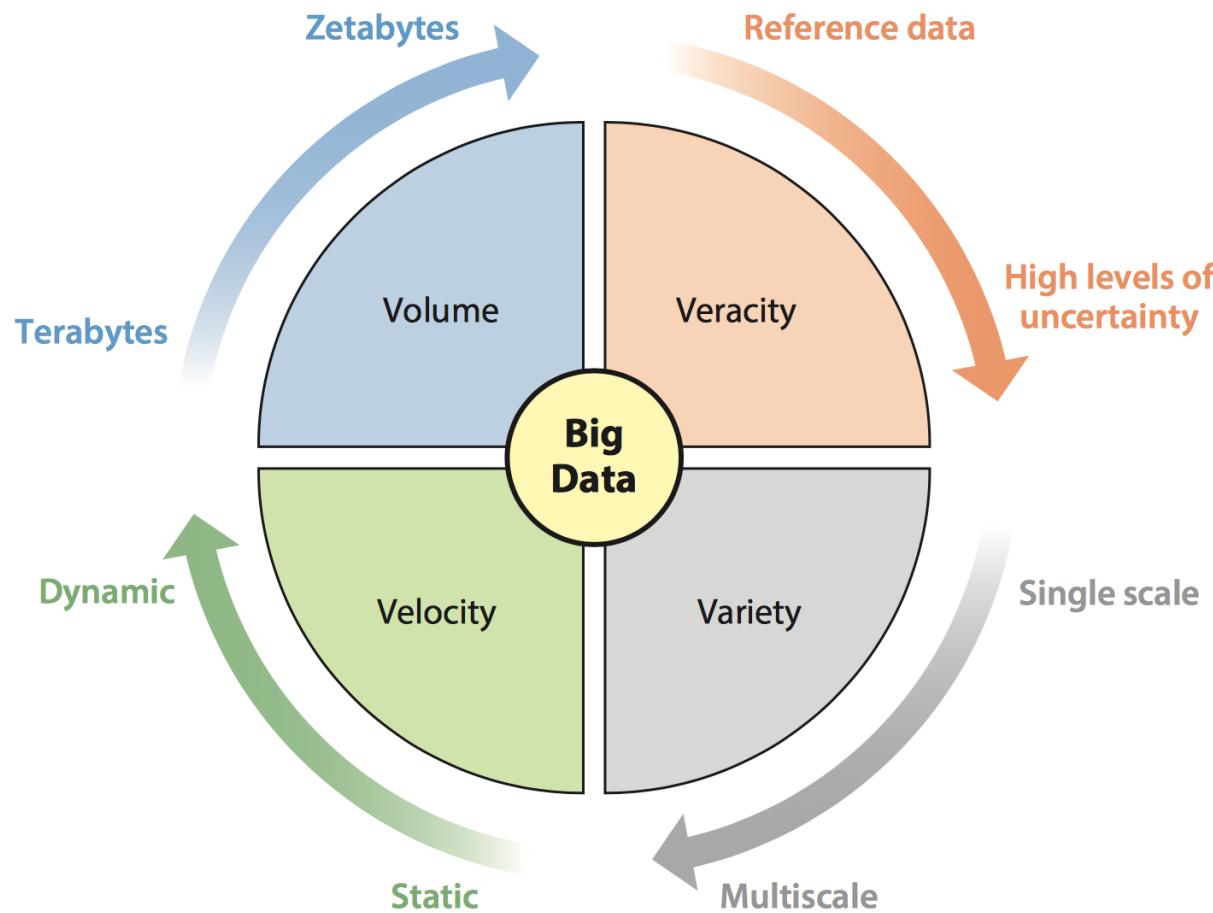


Modern Materials Discovery

- Data from experiment and calculation is now being generated at an incredibly fast rate
- This has allowed for the emergence of “Big Data” driven science



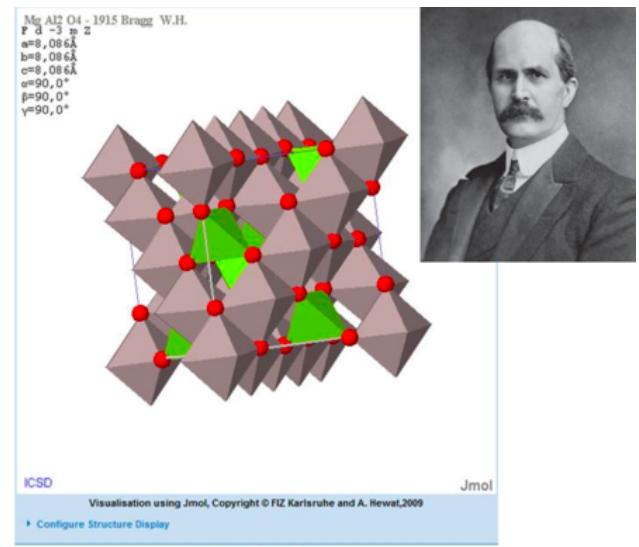
The 4 V of Big Data



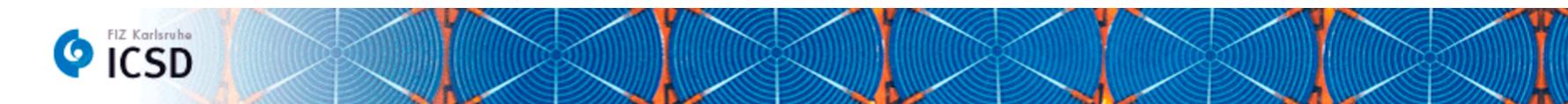
The ICSD – Experimental Data

Inorganic Crystal Structure Database

- Large database crystallographic data
- ~187,000 crystal structures



The ICSD - Experimental Data



FIZ Karlsruhe
ICSD

Home | Contact Welcome to ICSD Web. IP authenticated (135.196.210.216). Royal Society Chemistry i | Close session

Login	Chemistry Search	Search Action
LoginId: <input type="text"/>	Composition: In Sn O <input type="text"/> e.g. Na Cl	Number of Elements: 3 <input type="text"/>
Password: <input type="password"/>	Structural Formula: <input type="text"/> e.g. Pb (W O ₄)	Run Query Clear Query
Login Personalized	Chemical Name: <input type="text"/>	Search Summary
Lost password? Personalize account	Mineral Name: <input type="text"/> e.g. Adamite	Bibliography: - Cell: - Chemistry: 14 Symmetry: - Crystal Chemistry: - Structure Types: - Experimental Info: - DB Info: -
Navigation	Mineral Group: <input type="text"/> e.g. Pyroxene	Combined Results: 14
Basic search & retrieve	ANX Formula: <input type="text"/>	Query History
Advanced search & retrieve	AB Formula: <input type="text"/> Number of Formula Units: <input type="text"/>	Number of queries: 0
Bibliography	Clear Chemistry Search	Clear Query History
Cell	Count Chemistry Search	
Chemistry		
Symmetry		
Crystal Chemistry		
Structure Type		
Experimental Information		

Experimental vs. Computed Properties

“Real”/measured properties of materials

No great databases useful for data-driven approach ☹
(with exception of crystallographic data)

Simulated/predicted properties of materials

Some good emerging databases to choose from ☺

Data From Calculations



COMPUTATIONAL MATERIALS REPOSITORY



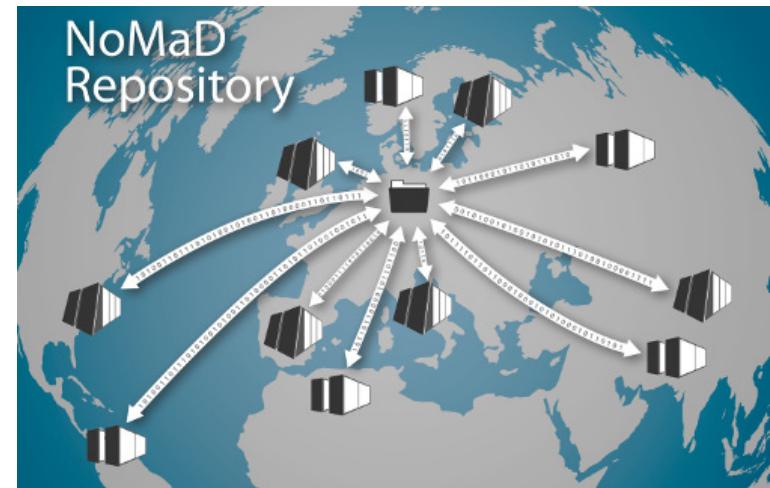
OQMD
@TheOQMD
oqmd.org

CEP – the Harvard
Clean Energy Project | *powered by*
 world community grid

The NoMaD Repository

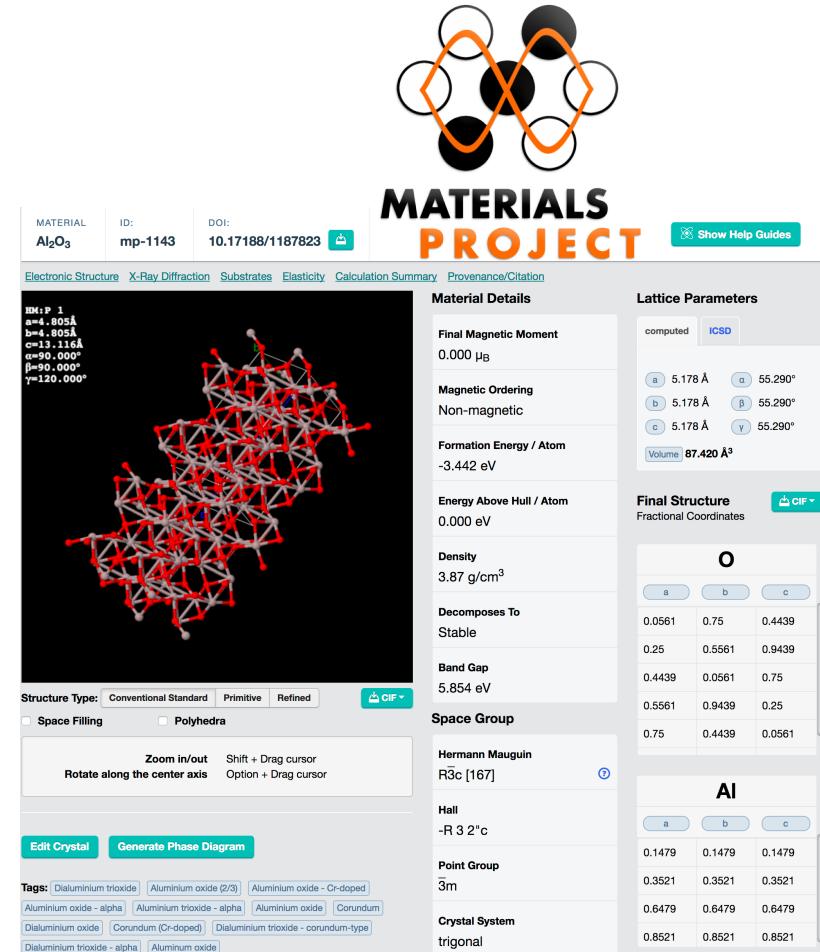
Novel Materials Discovery

- Contains input *and* output files from electronic structure calculations
- 3,300,000 entries
- Anyone can upload → Inhomogeneous data



The Materials Project

- Uses ICSD as primary input source
- 67,000 entries
- All calculations similar
→ homogeneous data



Using the Available Data

Two-Dimensional Metal Dichalcogenides and Oxides for Hydrogen Evolution: A Computational Screening Approach

Mohnish Pandey,[†] Aleksandra Vojvodic,[‡] Kristian S. Thygesen,^{†,§} and Karsten W. Jacobsen^{*,†}

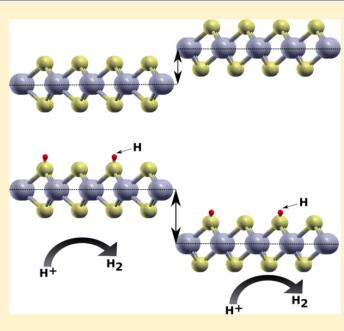
[†]Center for Atomic-Scale Materials Design, Department of Physics, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

[‡]SUNCAT Center for Interface Science and Catalysis, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, United States

[§]Center for Nanostructured Graphene (CNG), Department of Physics, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

Supporting Information

ABSTRACT: We explore the possibilities of hydrogen evolution by basal planes of 2D metal dichalcogenides and oxides in the 2H and 1T class of structures using the hydrogen binding energy as a computational activity descriptor. For some groups of systems like the Ti, Zr, and Hf dichalcogenides the hydrogen bonding to the 2H structure is stronger than that to the 1T structure, while for the Cr, Mo, and W dichalcogenides the behavior is opposite. This is rationalized by investigating shifts in the chalcogenide p levels comparing the two structures. We find that usually for a given material only at most one of the two phases will be active for the hydrogen evolution reaction; however, in most cases the two phases are very close in formation energy, opening up the possibility for stabilizing the active phase. The study points to many new possible 2D HER materials beyond the few that are already known.



2H-MX ₂	ΔH	ΔH _{hull}
RuS ₂	-0.31	-0.70
NiSe ₂	-0.21	-0.34
OsS ₂	0.34	-0.60
TaO ₂	-2.58	-3.00
ReO ₂	-0.91	-1.42
RhS ₂	-0.11	-0.48
PdS ₂	0.01	-0.31
NbS ₂	-1.21	-1.20
ScS ₂	-1.46	-1.46
TiS ₂	-1.23	-1.37
TaTe ₂	-0.32	-0.45
CoS ₂	-0.33	-0.48
IrS ₂	-0.11	-0.48
RhSe ₂	-0.17	-0.45
TaS ₂	-1.24	-1.22
ZrS ₂	-1.55	-1.73
ScO ₂	-2.74	-3.17*
VS ₂	-1.16	-1.14
ScSe ₂	-1.30	-1.25*
CrO ₂	-1.99	-2.15
PdSe ₂	-0.02	-0.33

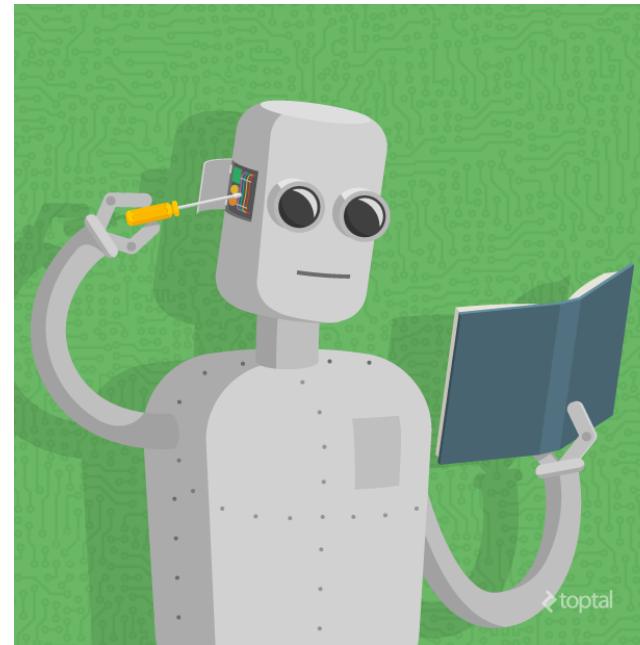
Calculated

From Database

Using the Available Data

Machine learning

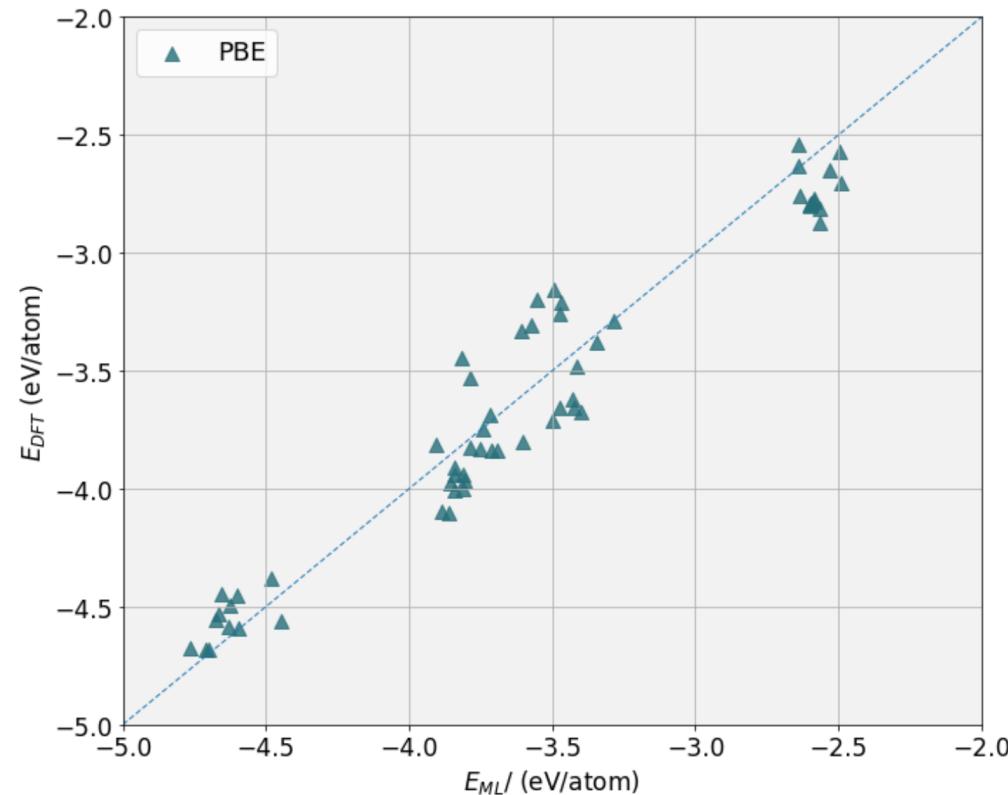
A type of artificial intelligence that provides computers with the ability to learn without being explicitly programmed.



Using the Available Data

Machine learning – Total Energies

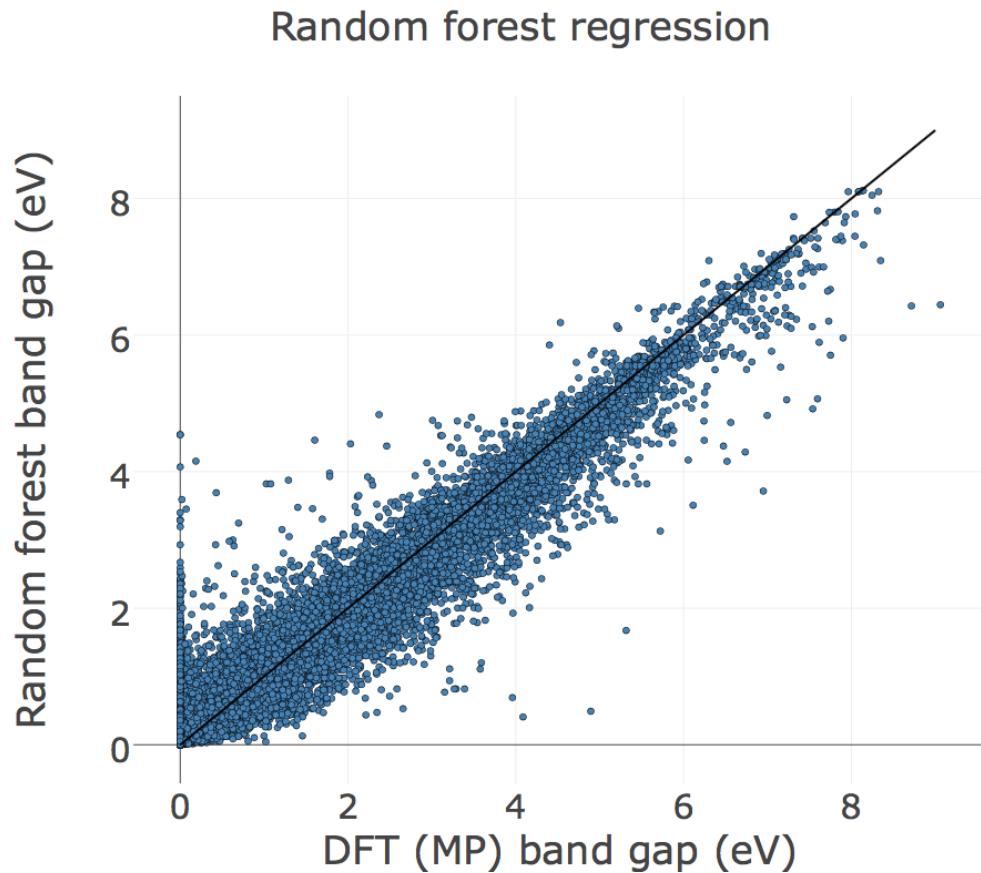
Using element properties such as atomic mass, # valence e⁻, ionisation potential etc. along with connectivity within material



Using the Available Data

Machine learning – Band gaps

Using only
element
composition as
input – no other
information



Talk Outline: The Materials Hyperspace

1. Data for Materials Discovery
2. Exploring Chemical Space
3. Interacting with Databases

Question

How many unique
materials* are known
today?

*Inorganic compounds

Counting Known Compounds

Hard to quantify – no definitive list exists



~3,300,000

Lots of duplicates
Many hypothetical



~187,000

All 'real' (mostly)
Some duplicates



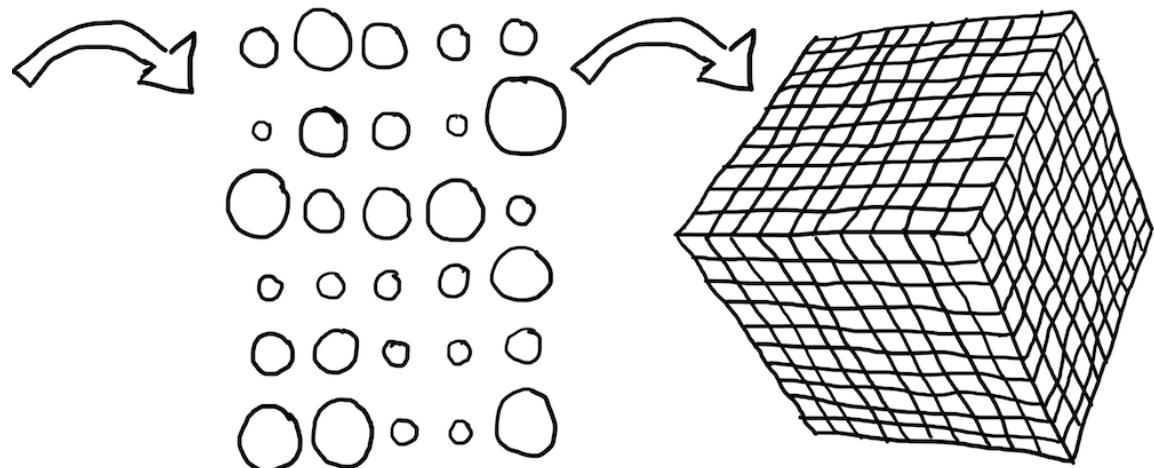
~67,000

Uses ICSD as input
But not exclusively
Not finished

Counting Possible Compounds

- You have a list of **50** chemical elements and a **10x10x10** grid
- You can put **30** atoms of any element anywhere on the grid to make a unit cell

H He Li Be B C N O F Ne
Na Mg Al Si P S Cl Ar K Ca
Sc Ti V Cr Mn Fe Co Ni Cu Zn
Ga Ge As Se Br Kr Rb Sr Y Zr
Nb No Tc Ru Rh Pd Ag Cd In Sn



Counting Possible Compounds

Format statements:

```
print('the value of x is {0}'.format(x))
```

```
print('the value of x is {:.2E}'.format(x))
```

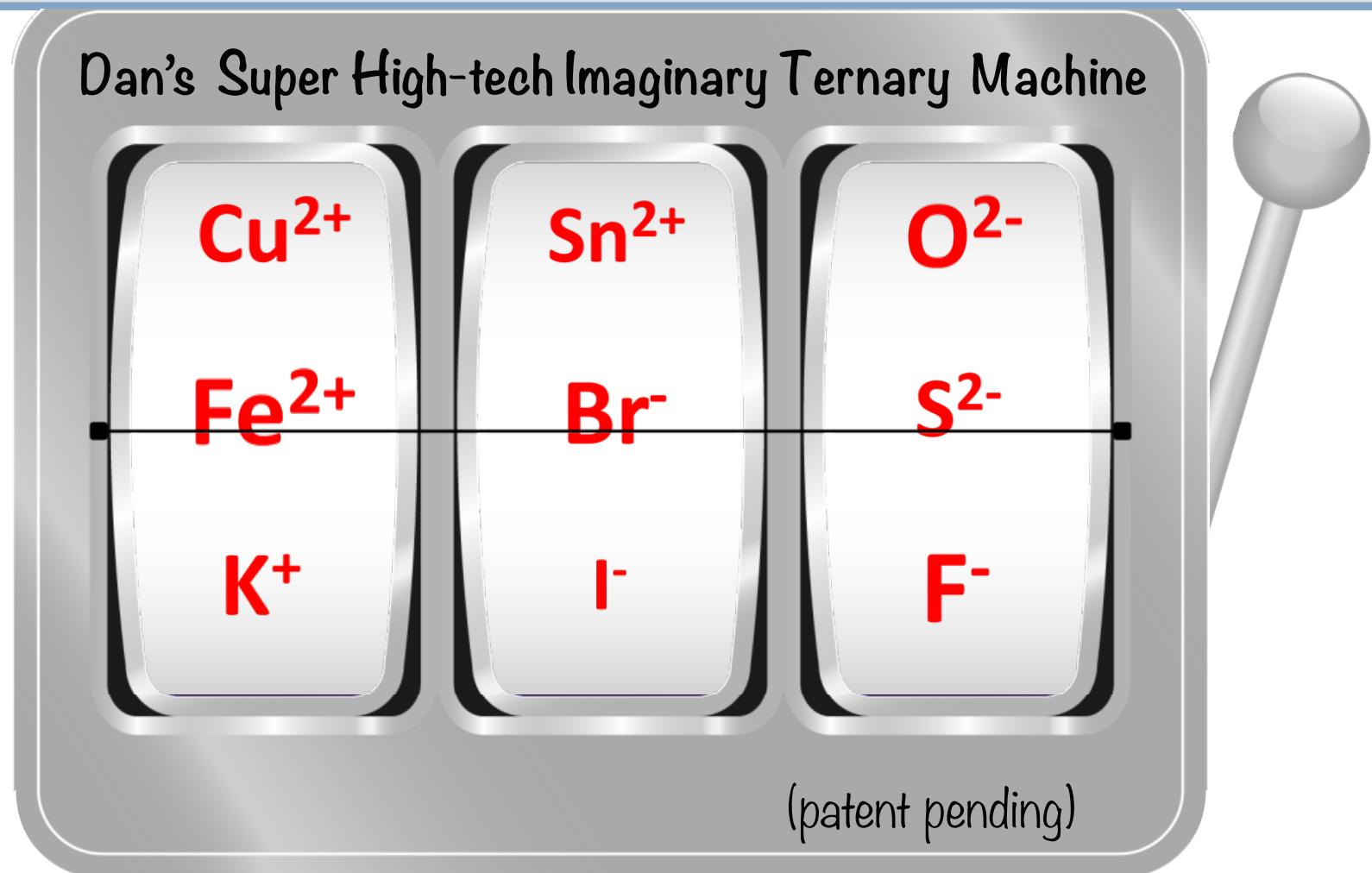
Q. What is the purpose of using the second type of format statement? (Try replacing a {:.2E} with a simple {0} in your notebook to see.)

Mapping Out Chemical Space

- Clearly this approach is a non-starter; we need to simplify things
- One way is to combine elements in their known oxidation states to make binary, ternary, quaternary combinations...



Mapping Out Chemical Space

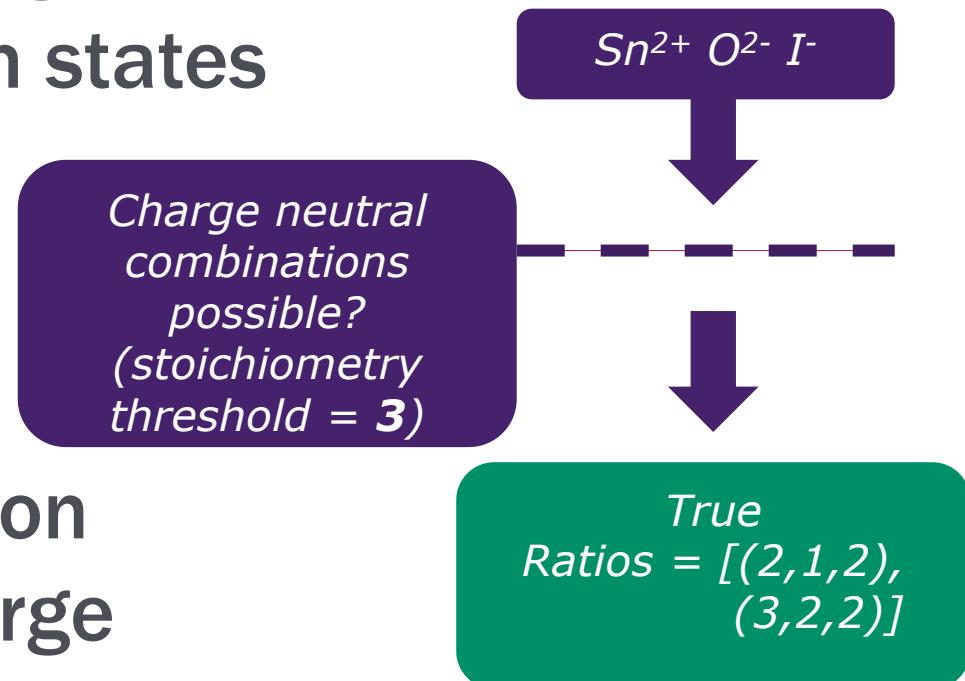


Mapping Out Chemical Space

SMACT

Semiconducting Materials by Analogy and Chemical Theory

- Combine elements together in their known oxidation states exhaustively
- Only allows certain combinations based on certain rules e.g. charge neutrality

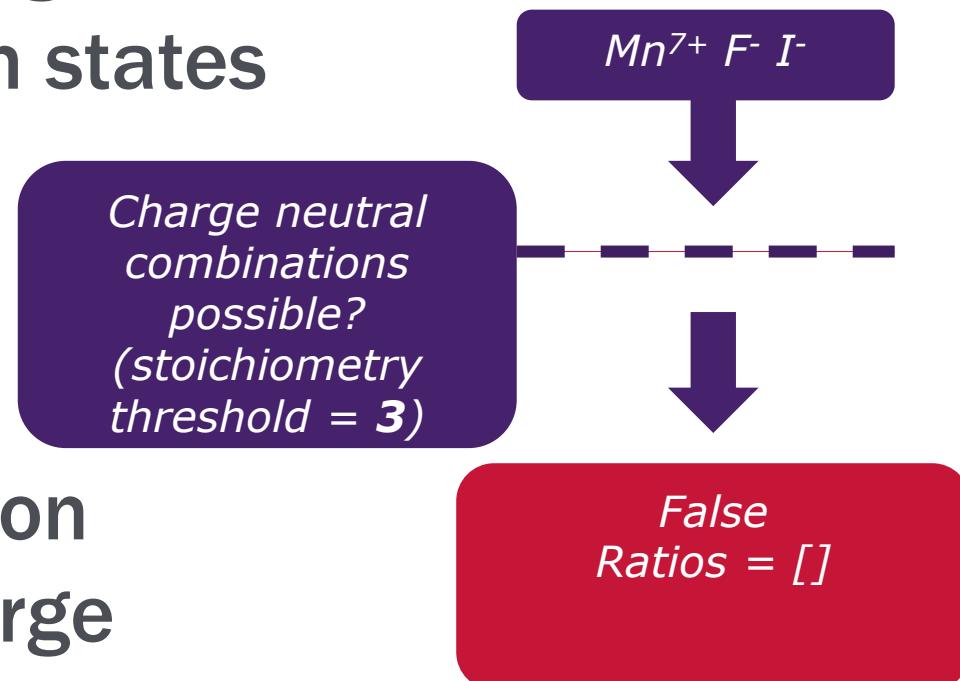


Mapping Out Chemical Space

SMACT

Semiconducting Materials by Analogy and Chemical Theory

- Combine elements together in their known oxidation states exhaustively
- Only allows certain combinations based on certain rules e.g. charge neutrality

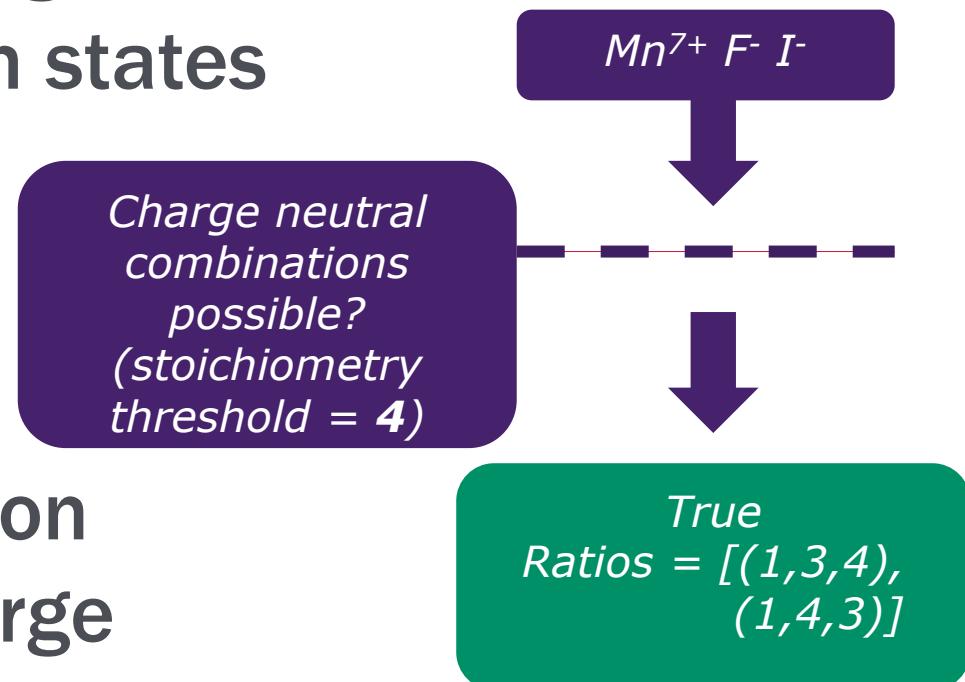


Mapping Out Chemical Space

SMACT

Semiconducting Materials by Analogy and Chemical Theory

- Combine elements together in their known oxidation states exhaustively
- Only allows certain combinations based on certain rules e.g. charge neutrality



Mapping Out Chemical Space

Setting up a dictionary

```
all_elements = smact.element_dictionary()
```

Dictionary of
element objects

Q. Can you spot the other
smact function that
appears in the same
cell?

Built in smact function

```
def element_dictionary(elements=None):
    """
    Create a dictionary of initialised smact.Element objects

    Accessing an Element from a dict is significantly faster than
    repeatedly initialising them on-demand within nested loops.

    Args:
        elements (iterable of strings) : Elements to include. If None,
            use all elements up to 103.
    Returns:
        dict: Dictionary with element symbols as keys and smact.Element
              objects as data
    """
    if elements == None:
        elements = ordered_elements(1,103)
    return {symbol: Element(symbol) for symbol in elements}
```

Mapping Out Chemical Space

Using the enumerate function

```
for i, ele_a in enumerate(elements):
```

is the same as

```
for i in range(len(elements)):  
    ele_a = elements[i]
```

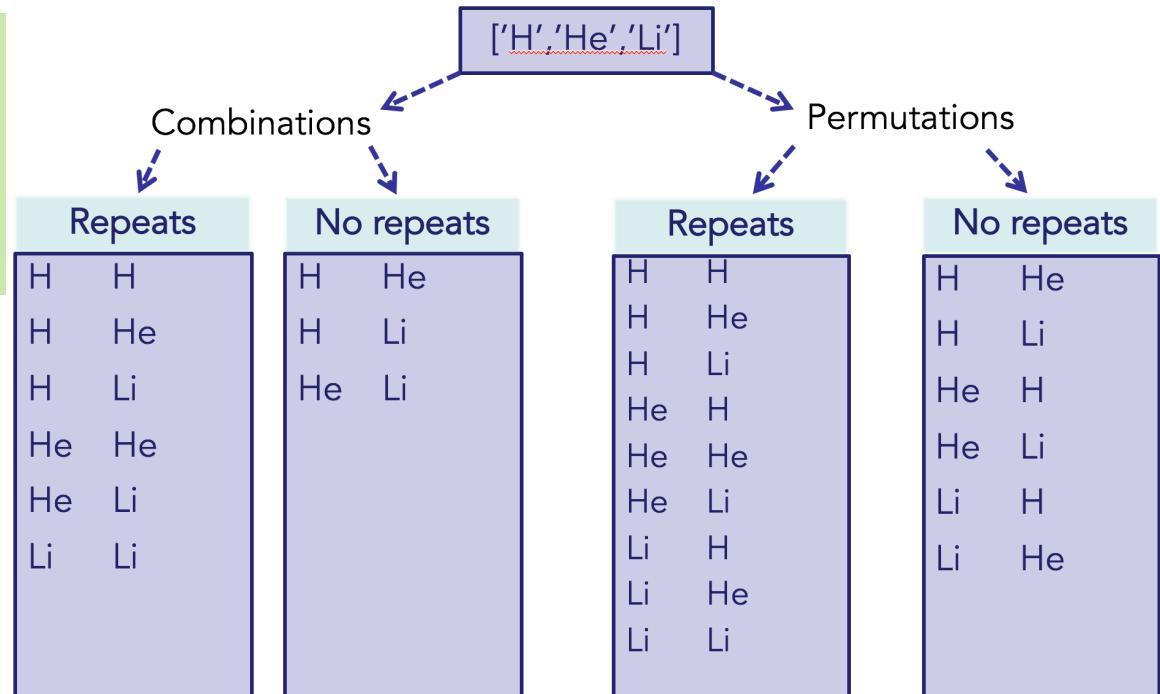
Mapping Out Chemical Space

List slicing

```
for j, ele_b in enumerate(elements[i+1:]) :
```

[where to start]:[where to stop]

Q. What does putting this within our inner loop achieve?



Talk Outline: The Materials Hyperspace

1. Data for Materials Discovery
2. Exploring Chemical Space
3. Interacting with Databases

Class Question

How can we access
materials data?

Accessing Materials Data

- Web Browser e.g. Materials Project
- Data dump e.g. Computational Materials Repository
- Restful API e.g. Materials Project MAPI

RESTful API

- Representational State Transfer Application Programming Interface
- Built around resources and how they are accessed



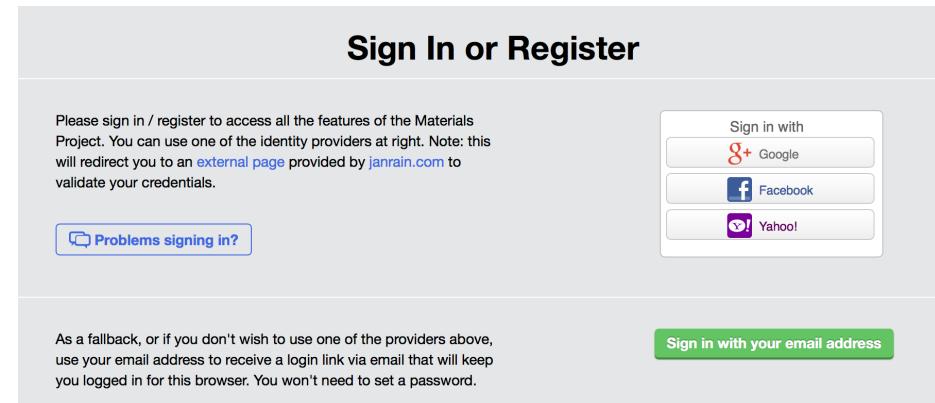
- Resources == objects in object oriented programming!



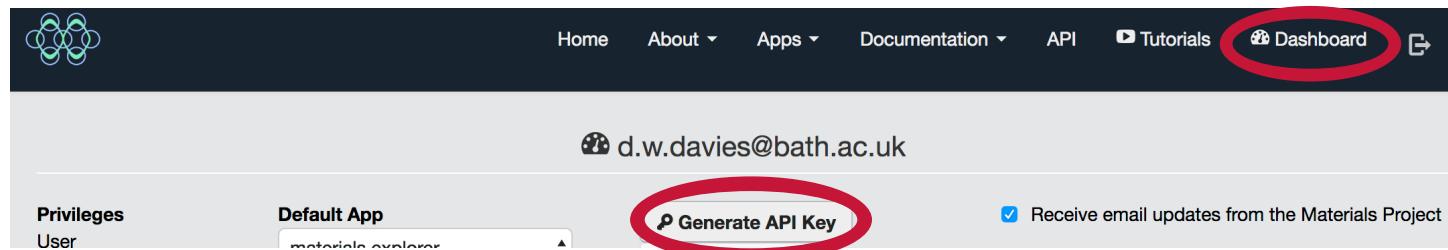
Materials API

- Register:

MaterialsProject.org



- Copy and paste your API key:



The image shows the Materials Project dashboard. At the top, there is a navigation bar with links for Home, About, Apps, Documentation, API, Tutorials, Dashboard (which is circled in red), and a user icon. Below the navigation bar, there is a user profile section with an email address: d.w.davies@bath.ac.uk. Further down, there is a 'Default App' section with a dropdown menu set to 'materials_explorer'. A red circle highlights the 'Generate API Key' button, which is located in the same row as the app dropdown. To the right of the 'Generate API Key' button, there is a checkbox for 'Receive email updates from the Materials Project'.

- `m = MPRester ("")`

Materials API

General usage

```
data = m.query(criteria, properties)
```

API query function

Criteria of the
entries we're
interested in

Properties we want
to get back

Returns a list of dictionaries

```
[ {property_1 : value, property_2: value},  
 {property_1 : value, property_2: value} ]
```

Entry 1

Entry 2

Materials API

Getting clever with criteria

```
criteria = { 'nelements': 2, ←  
             'elements':  
                 { '$in': [ 'Co', 'Fe' ] } }
```

Simple Python
dictionary usage

MongoDB operator
usage (as python strings)

Also accepts \$gt, \$lt, \$eq, \$all, \$nin,
\$exists etc.

Summary – Key Points

- Materials data is being generated at very fast rate
- Emerging efforts to store and organise the data can speed up materials discovery
- Python + modern databases = huge amounts of materials data at your fingertips
- There are vast areas of the chemical landscape that remain totally unexplored