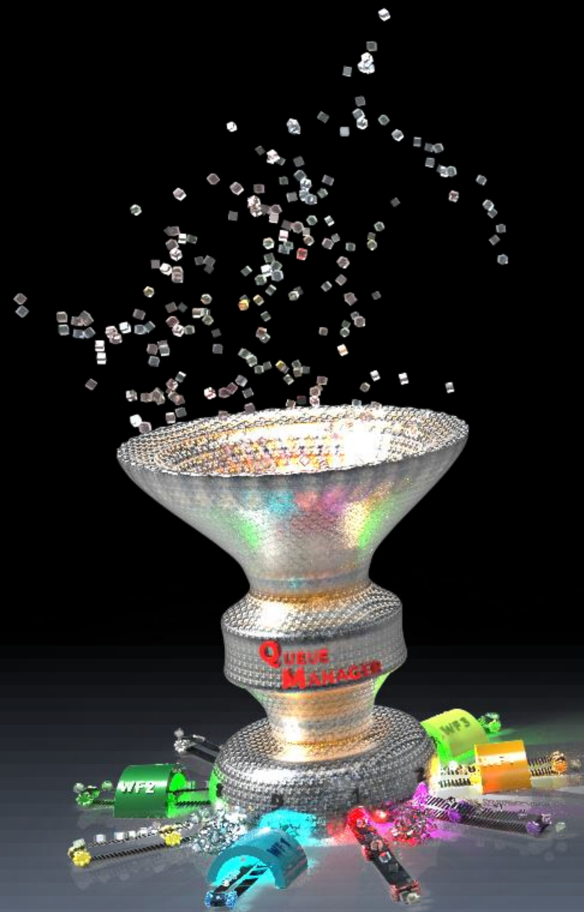
How many materials are left to discover?

An exploration of quaternary space



Presentation by **Michael Sluydts**

Research performed by **Michael Sluydts, Michiel Larmuseau, Karel Dumon, Titus Crepain, Kurt Lejaeghere and Stefaan Cottenier**

Center for Molecular Modeling
Ghent University
Belgium

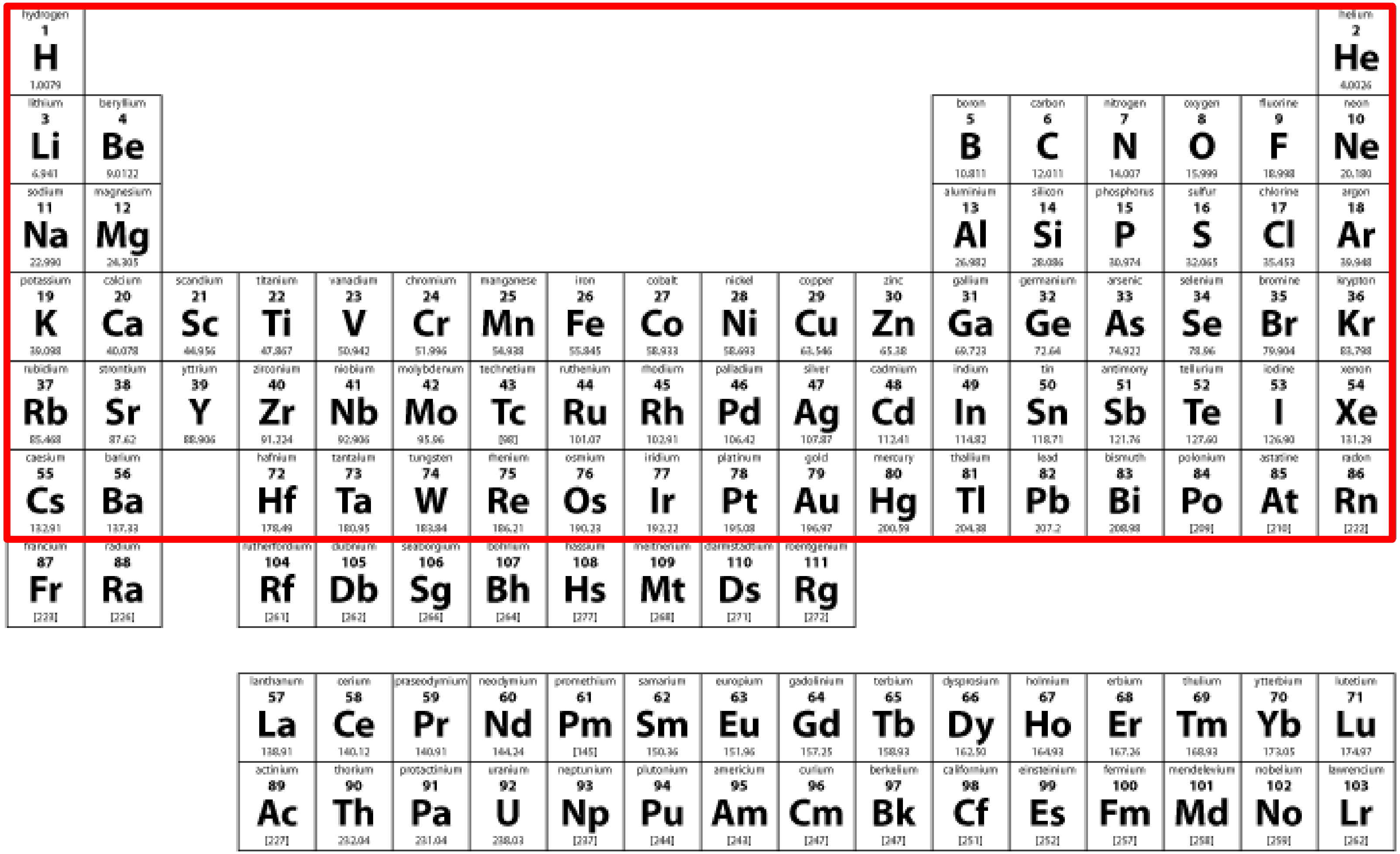
Finding the right material

1) How many materials are there left to discover?

2) How do we find them?

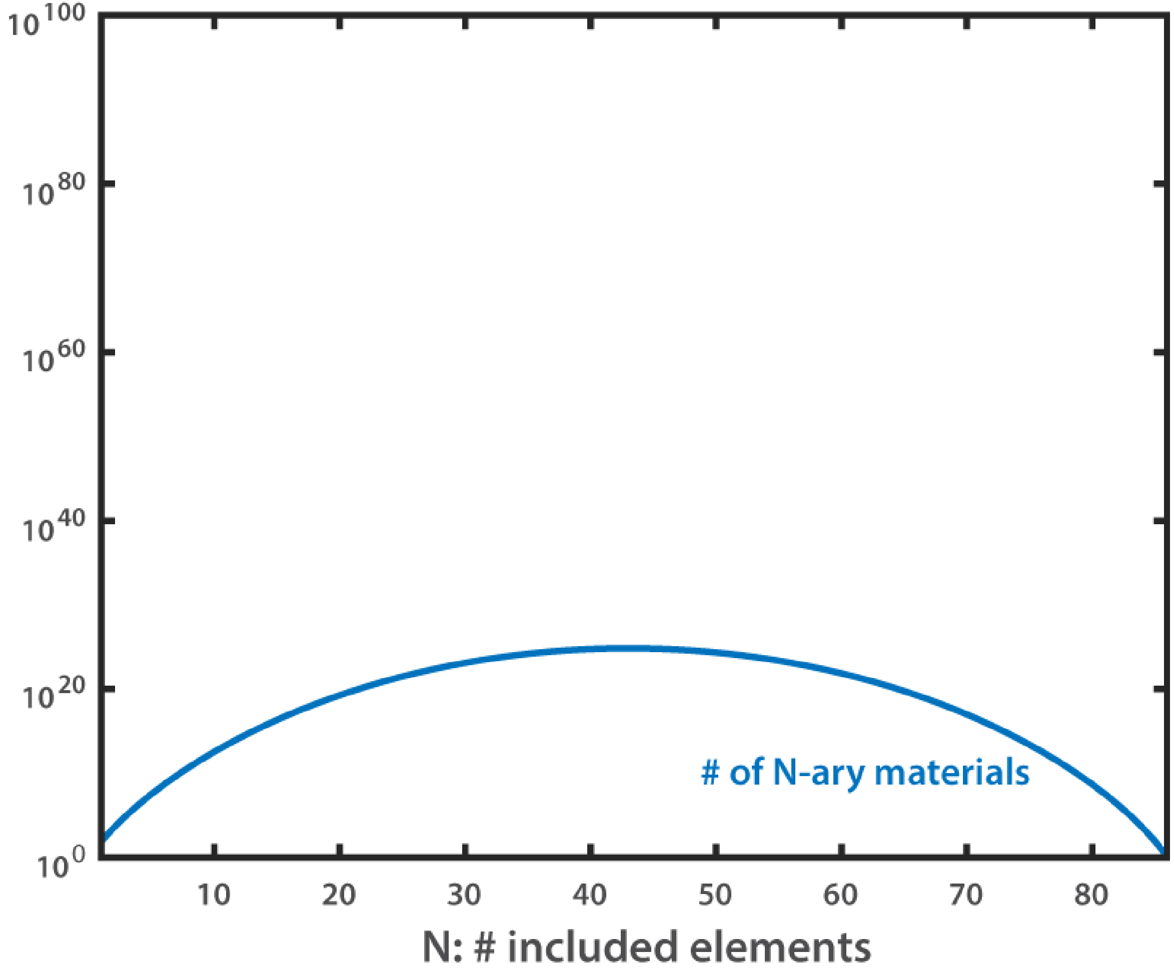
Can we create a Drake equation for materials?

Periodic table



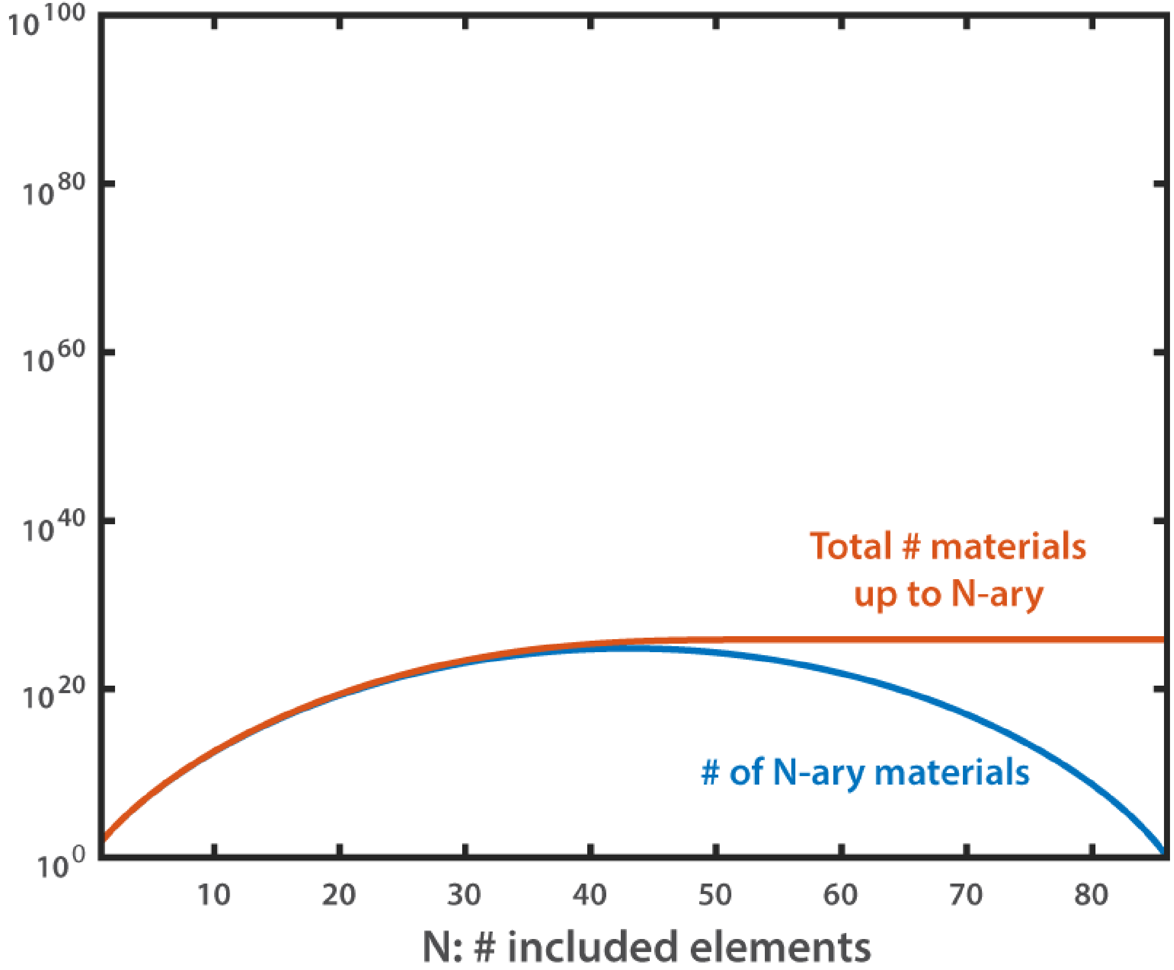
Consider the first 86 elements in the periodic table

All combinations



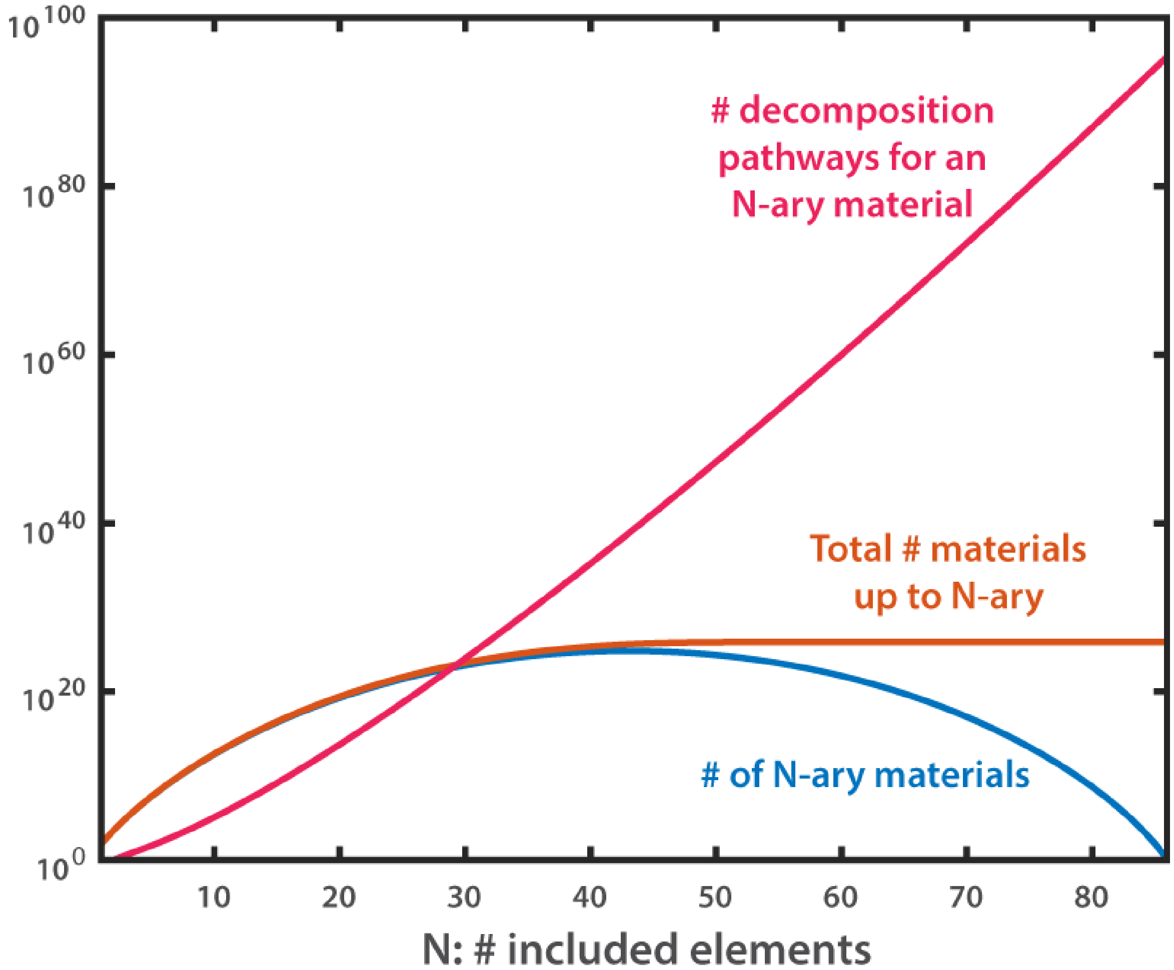
Maximum at 43 elements: 10^{25} materials

All combinations



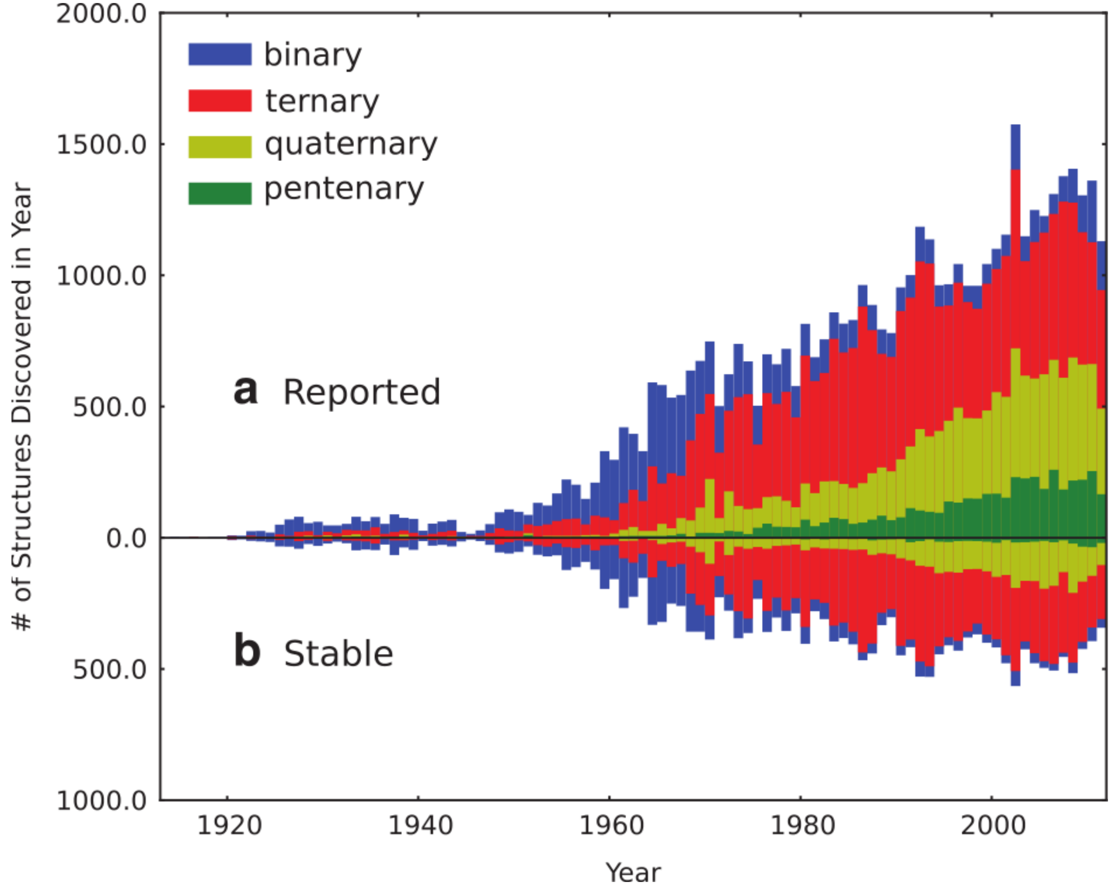
Total dominated by maximum

Decomposition pathways



Decomposition pathways increase even faster

The current status

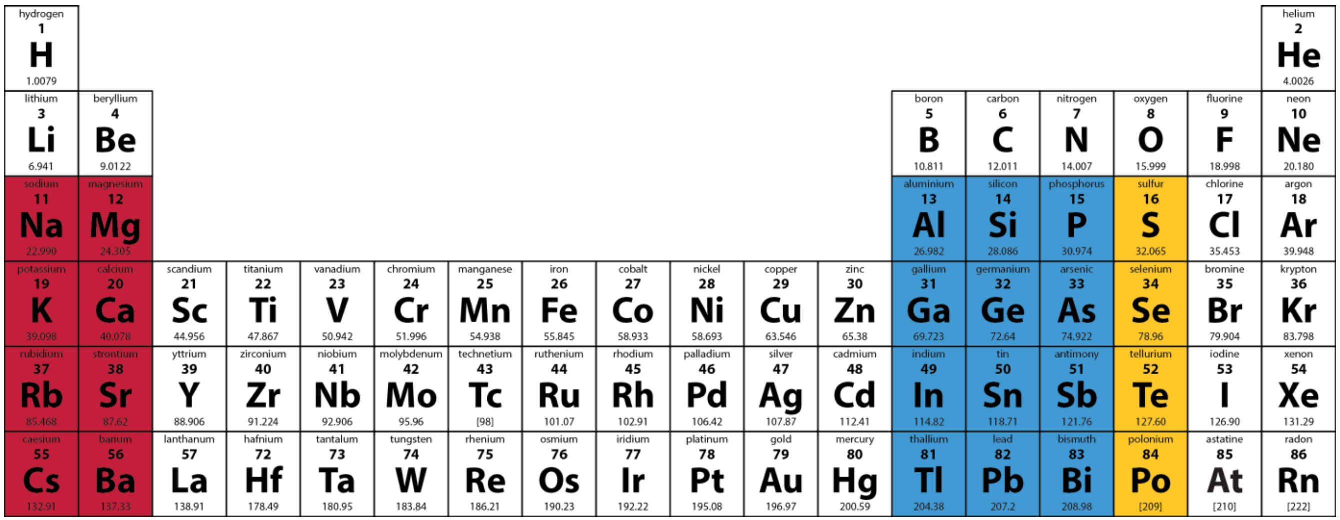


Source: OQMD

Only about 500 quaternaries are found each year

QZP space

Quaternary zintl phases provide an interesting subset of materials



Zintl phases combine covalent and ionic bonding in a single crystal

Visualizing QZP space

Let's create a map of QZP space:

- **Extract all known structures from ICSD**

- **Discard partial occupancies (around 2/3)**

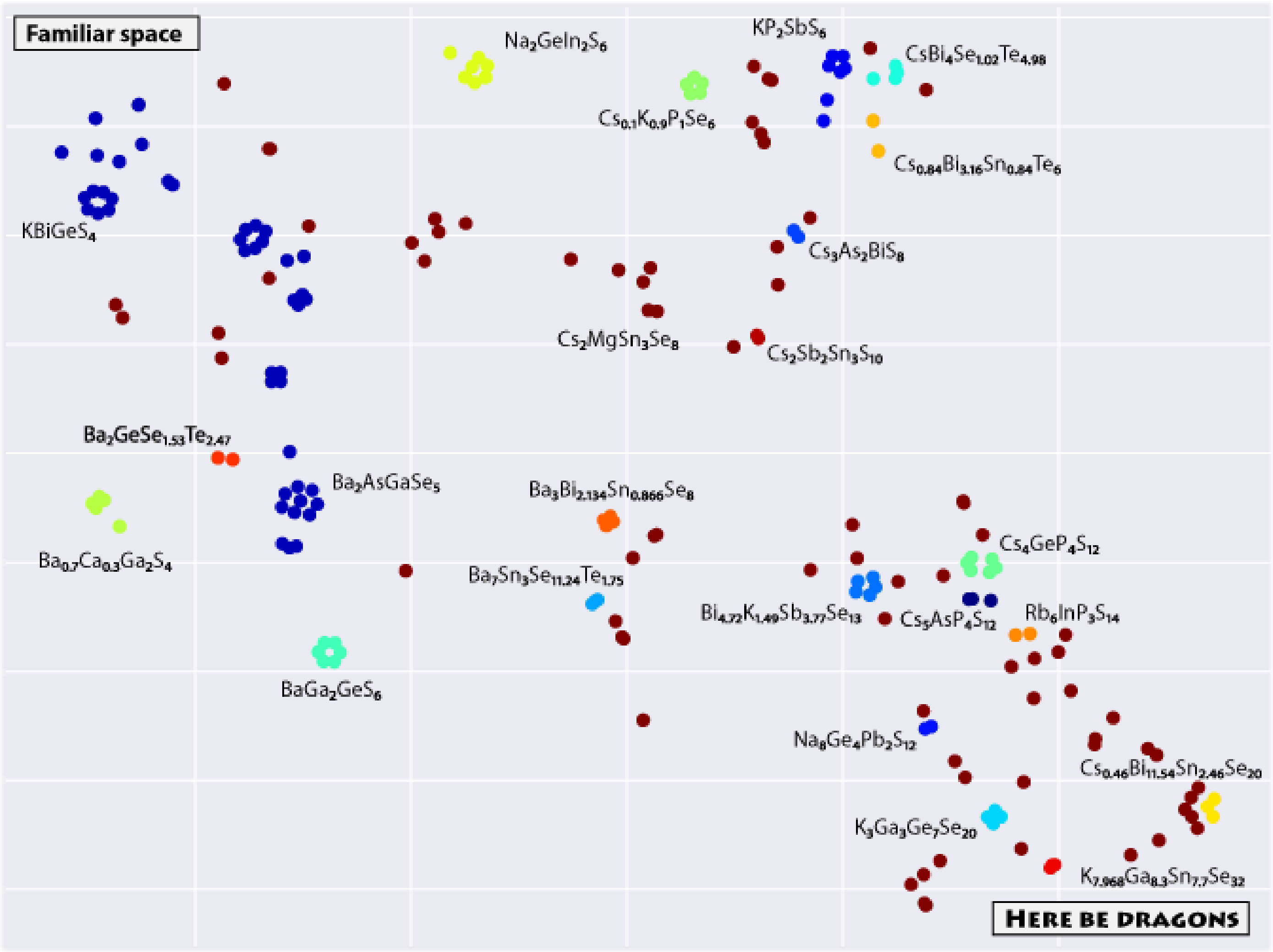
- **Generate features:**

  atomic+ stoichiometry + spacegroup

- **Cluster with DBSCAN**

- **Visualize with t-SNE**

Visualizing QZP space



Similar structures are clustered, but many are unique.

Drake for quaternaries

Estimating the number of materials based on observed QZP is like estimating the number of planets based on the milky way

Simplified Drake equation

$$N_{planets\ with\ life}$$

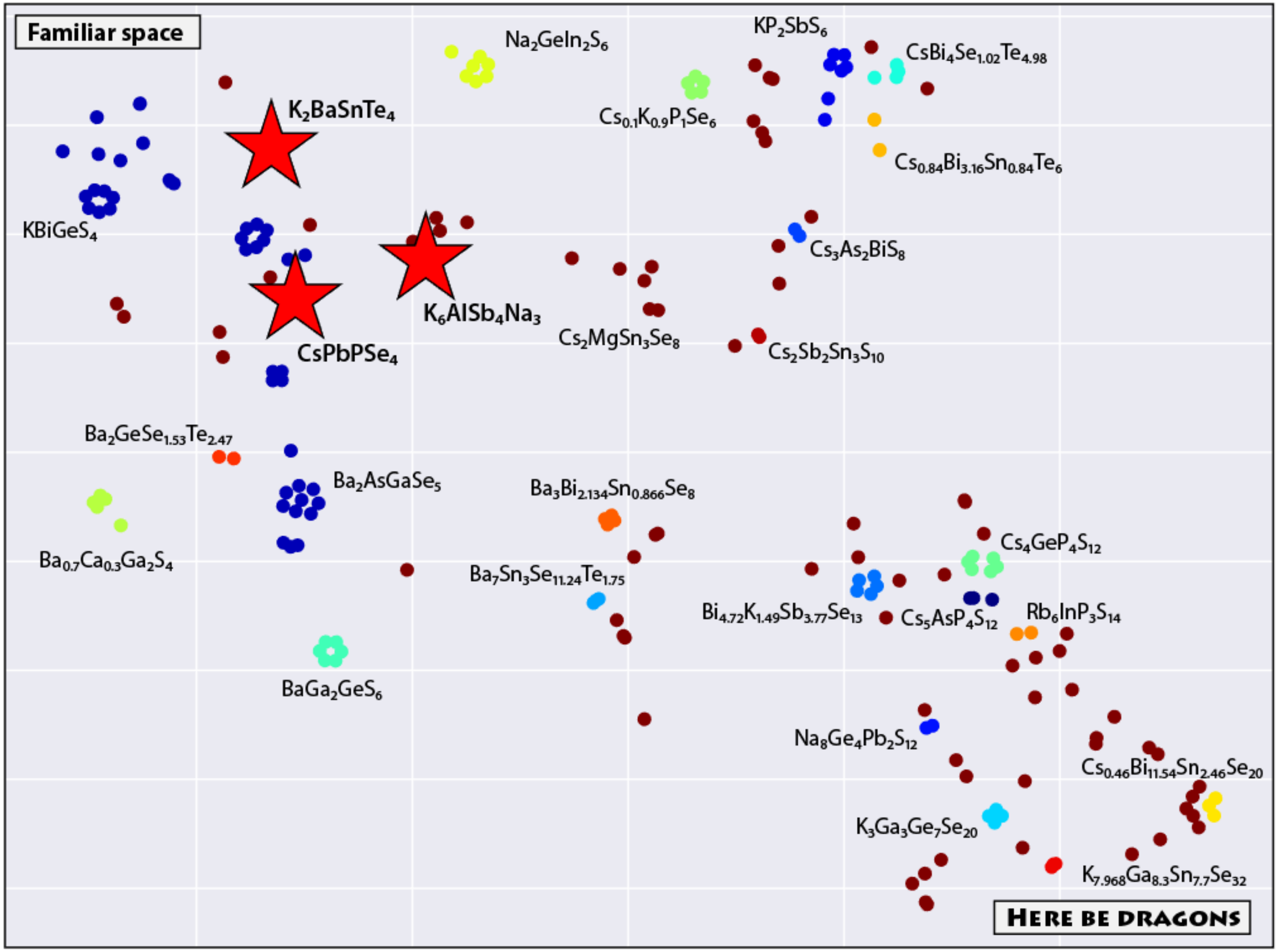
$$= \frac{N_{planets,universe}}{N_{planets,MW}} \cdot N_{solar\ systems,MW} \cdot N_{\frac{observed}{solar\ system}} \cdot N_{\frac{total}{observed}} \cdot f_{life}$$

Materials Drake equation

$$N_{useful\ materials}$$

$$= \frac{N_{materials,space}}{N_{materials,QZP}} \cdot N_{material\ clusters,QZP} \cdot N_{\frac{observed}{cluster}} \cdot N_{\frac{total}{observed}} \cdot f_{use}$$

What's left of quaternary space?

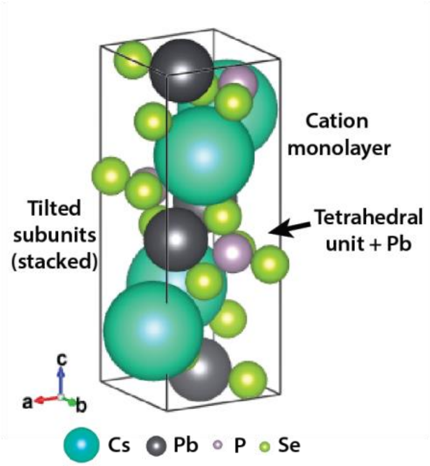


Sample a few interesting families.

Families

Template : $CsPbPSe_4$

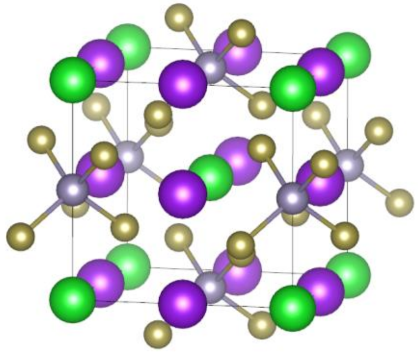
Spacegroup: Pnma
Stoichiometry: 1:1:1:4



6 known

Template : $K_2BaSnTe_4$

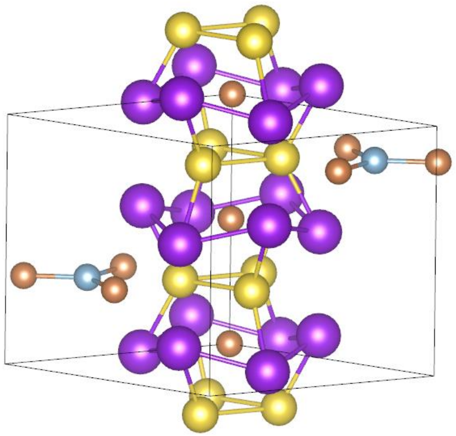
Spacegroup: I-42m
Stoichiometry: 2:1:1:4



1 known

Template : $K_6AlSb_4Na_3$

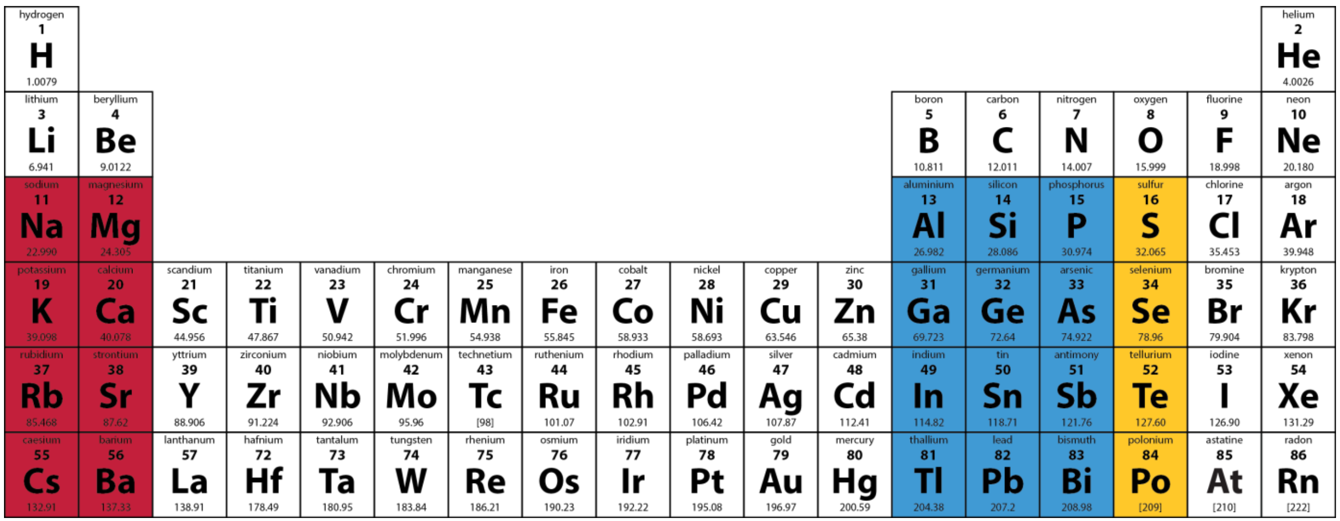
Spacegroup: P63mmc
Stoichiometry: 6:1:4:3



4 known

Creating new materials

A database of hypothetical materials is created

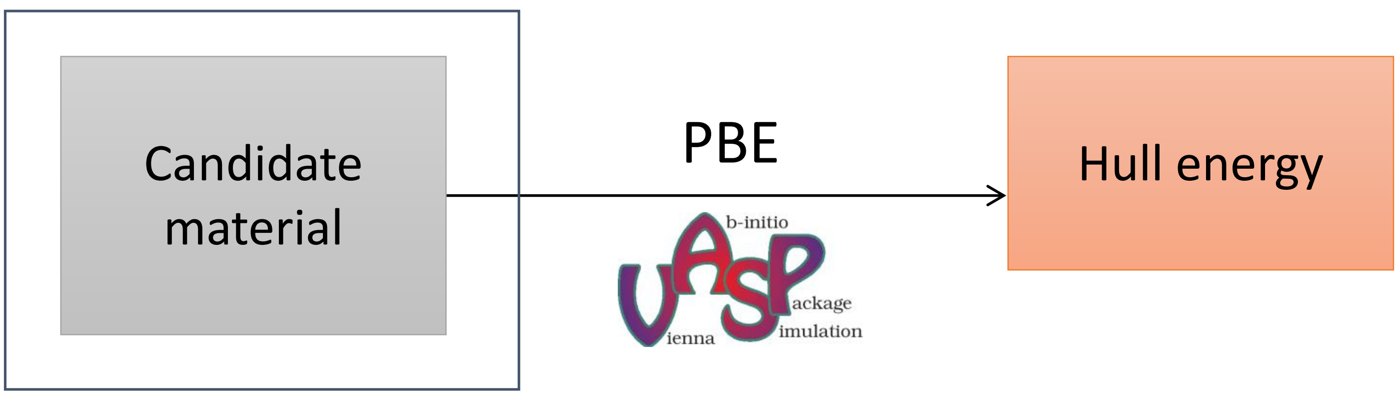


$CsPbPSe_4$ $K_2BaSnTe_4$ $K_6AlSb_4Na_3$

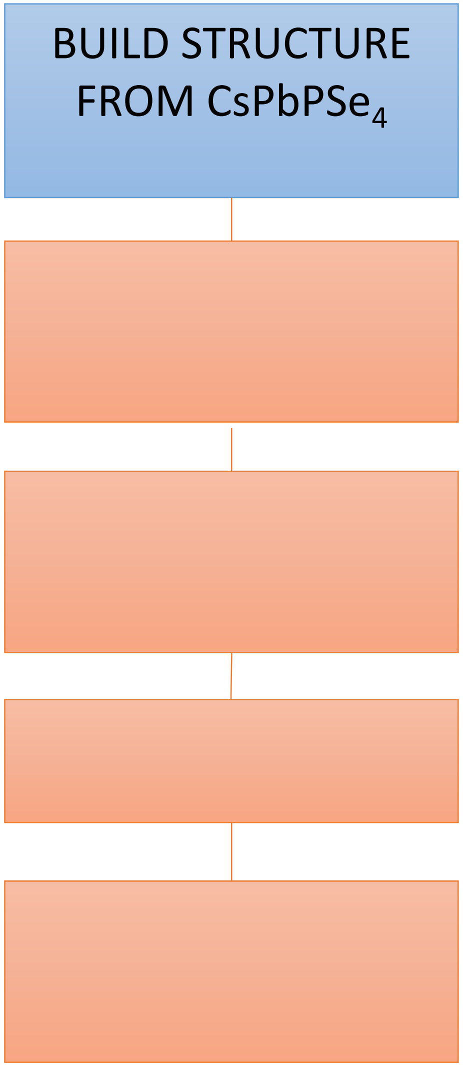
Combinations: 3168 2016 1696

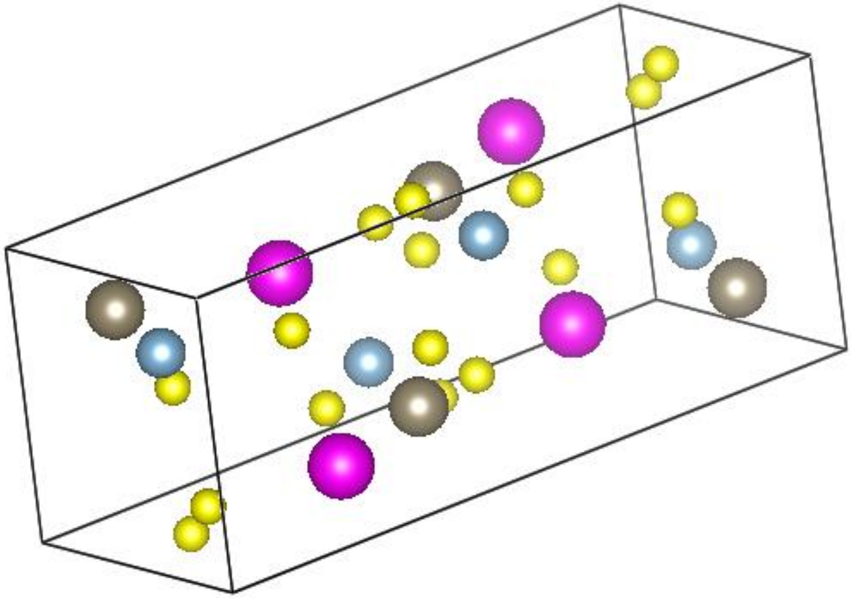
Workflow

High-throughput ab initio screening



Workflow

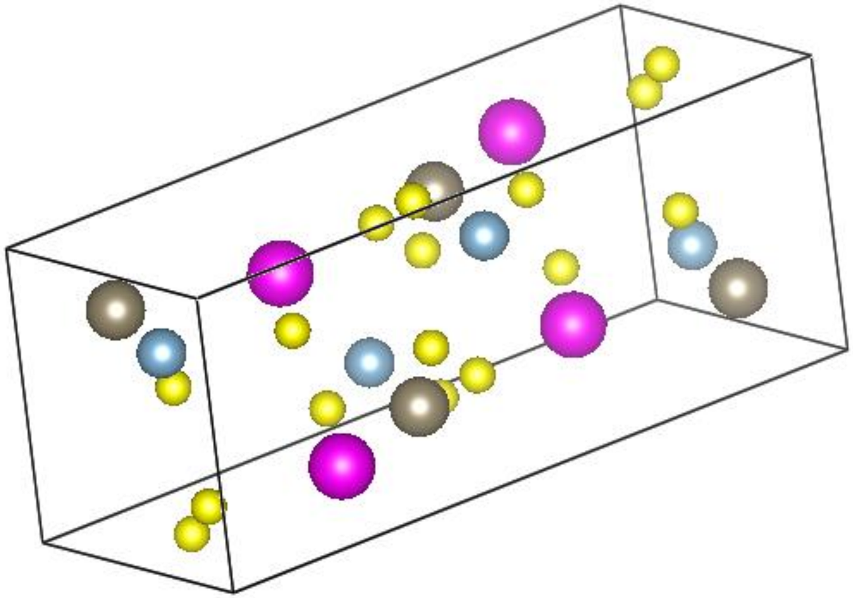




Workflow

BUILD STRUCTURE FROM $CsPbPSe_4$

VOLUME RELAXATION

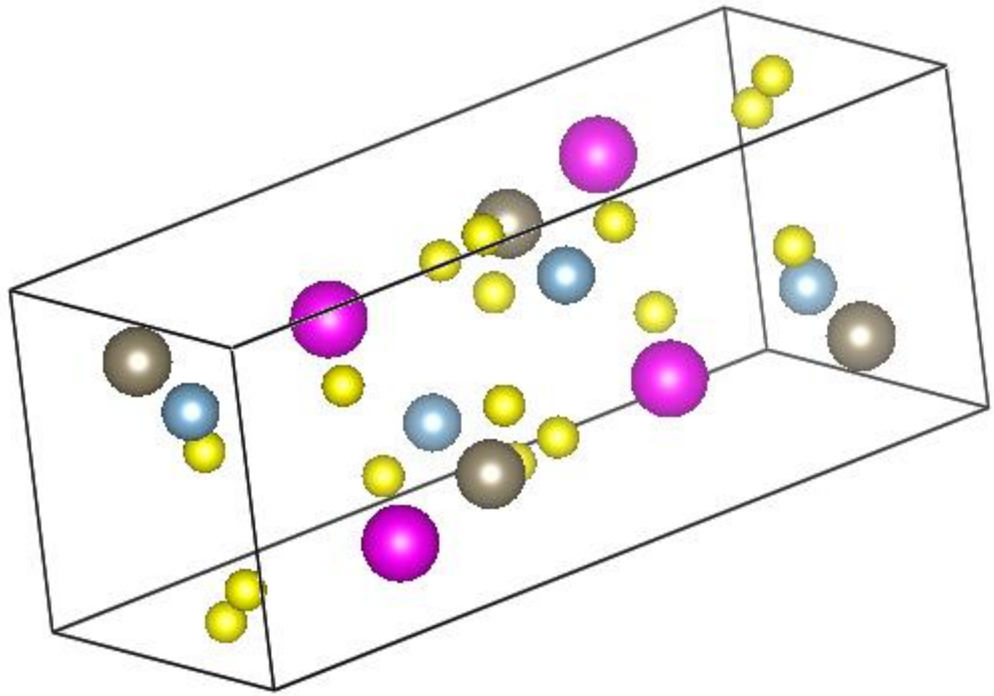


Workflow

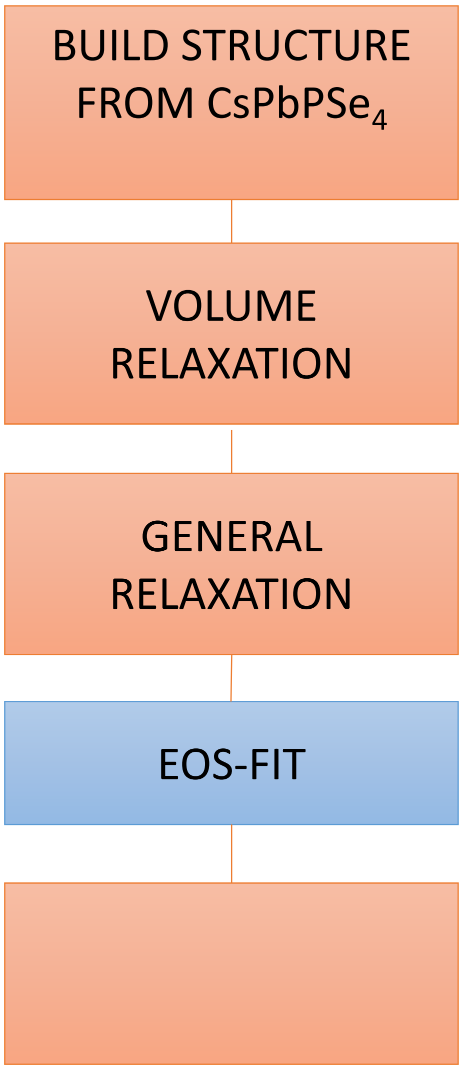
BUILD STRUCTURE FROM $CsPbPSe_4$

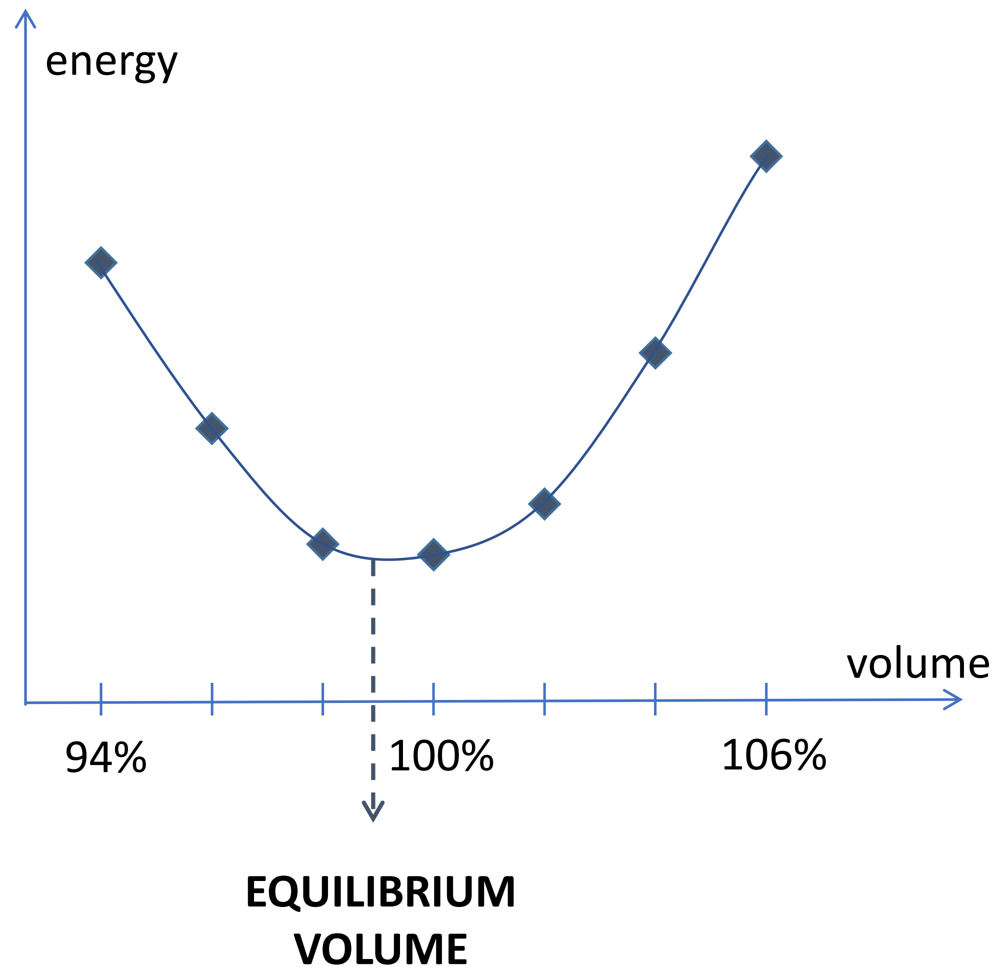
VOLUME RELAXATION

GENERAL RELAXATION

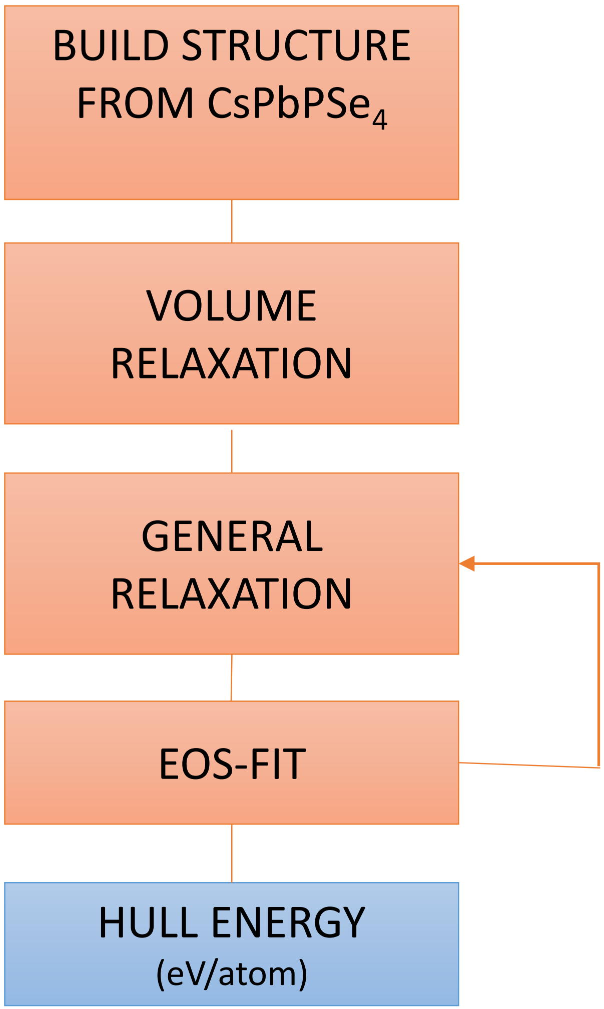


Workflow


BUILD STRUCTURE FROM $CsPbPSe_4$
VOLUME RELAXATION
GENERAL RELAXATION
EOS-FIT


energy
volume
94%
100%
106%
EQUILIBRIUM VOLUME

Workflow



+- 20 calculations per material

> 100,000 calculations

Lost information

Most information is discarded

- Unstable materials: everything
- Stable materials: everything but the Ehull

Recycle with machine learning!

Surrogate modeling

Goal: learn the stable compositions as we perform calculations

Atomic information → Gaussian process → **Ehull**

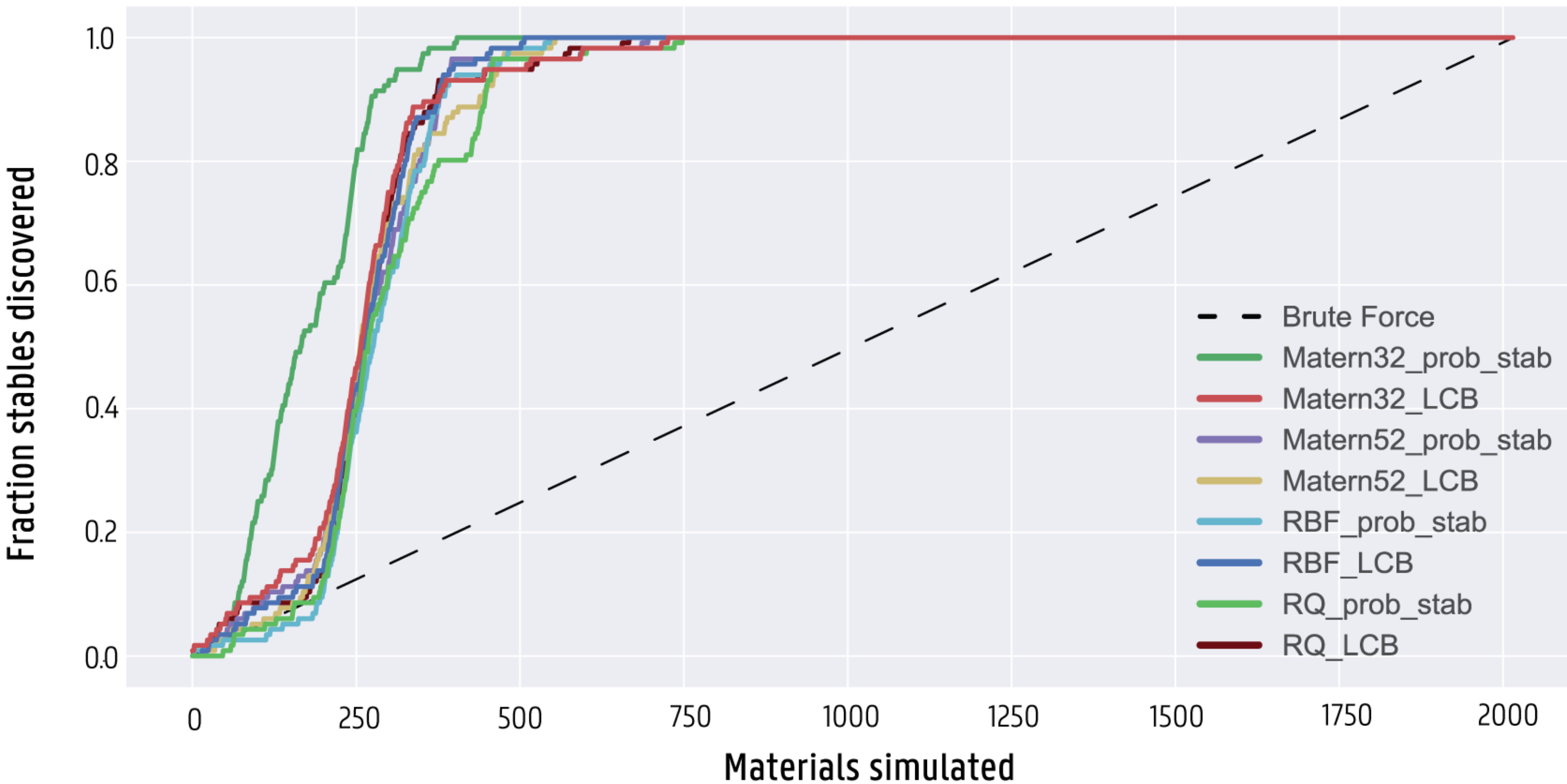
↓

Integrate with Queue Manager to prioritize calculations

Best model

Source: PhD work M. Larmuseau

Best performance for a Matern32 kernel with probability sampling

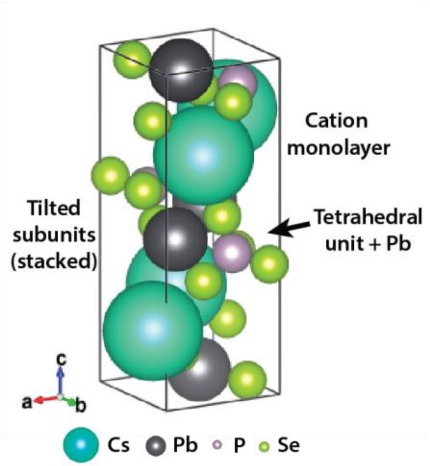


90% of stable materials are found in 300 samples

How many did we find?

Template : $CsPbPSe_4$

Spacegroup: Pnma
Stoichiometry: 1:1:1:4

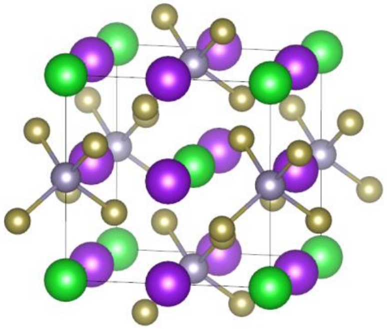


108 stables
3168 total

1 in 30

Template : $K_2BaSnTe_4$

Spacegroup: I-42m
Stoichiometry: 2:1:1:4

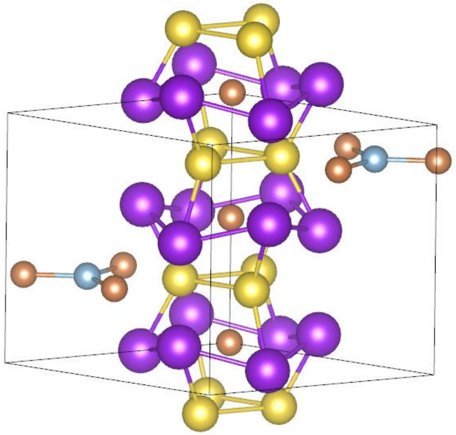


116 stables
2016 total

1 in 20

Template : $K_6AlSb_4Na_3$

Spacegroup: P63mmc
Stoichiometry: 6:1:4:3



115 stables
786 total

1 in 8

Drake for quaternaries

Materials Drake equation

$$N_{useful\,materials} = \frac{N_{materials,space}}{N_{materials,QZP}} \cdot N_{clusters,QZP} \cdot N_{\frac{observed}{cluster}} \cdot N_{\frac{total}{observed}} \cdot f_{use}$$

$\frac{N_{materials,space}}{N_{materials,QZP}}$	$\frac{46399}{509} = 91.15$
$N_{clusters,QZP}$	$\pm\, 25\ clusters + 75\ unique$ $(+partial\ occupancies)$
$N_{\frac{observed}{cluster}}\ N_{\frac{total}{observed}}$	$\pm\, 110$
f_{use}	$1\ (stable)$

The result is around 1 million with 47,000 known and 500 new ones found each year.

Conclusions

- **High-throughput screening allows us to explore large material spaces**
- **Machine learning enables the recycling of information**
- **We can now find more stables in a month than experiment in a year**
- **Experimental synthesis is still needed**
- **Geometric models will allow us to reuse information not only within a family, but also between**