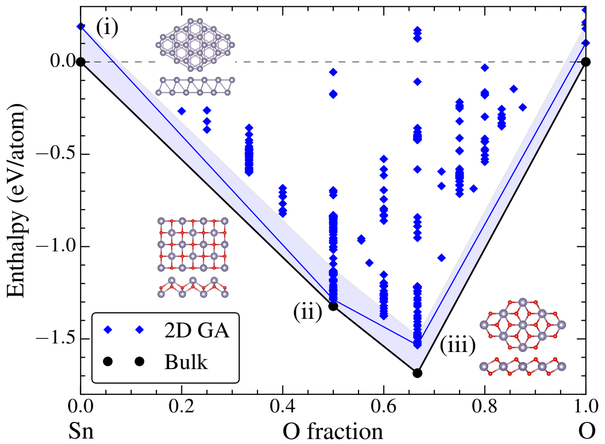
Search of Energy Landscapes by Evolutionary Algorithms and Data Mining

Richard G. Hennig, University of Florida

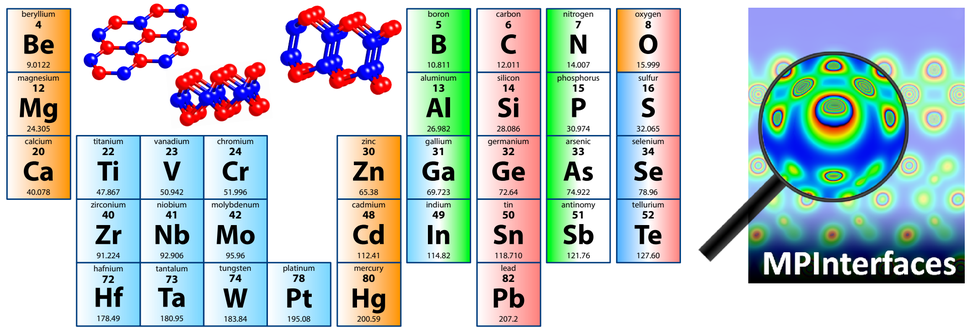
Structure and Energy Landscapes

- Structure representation
- Features of energy landscapes
- Structure searches for 2D materials by datamining, chemical substitution, and evolutionary algorithms

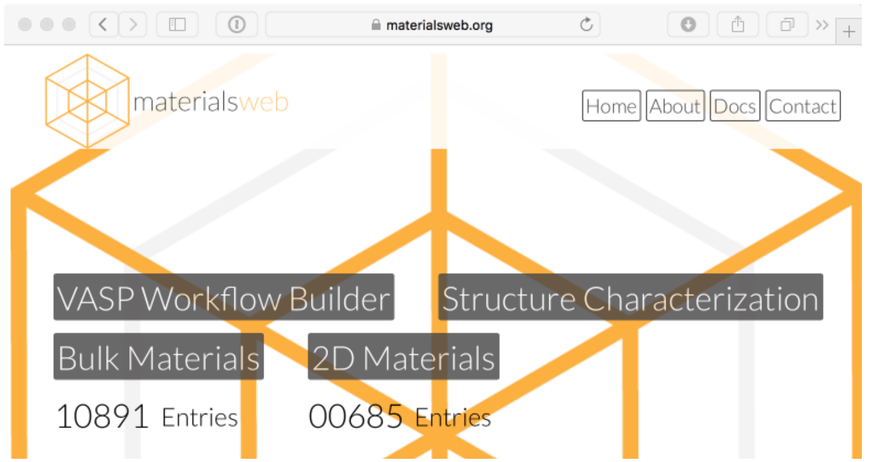


GASP - Genetic algorithm and machine learning for structure predictions

MPInterfaces - High throughput framework for 2D materials



Data available at http://materialsweb.org



Powered by
MPInterfaces & materialsweb

Open source available at https://github.com/henniggroup

Acknowledgment

- MPInterfaces for novel 2D materials: **K. Mathew**, **A. Singh**, **M. Ashton**, **J. Paul**, **D. Gluhovic**, **H. Zhuang**, J. Gabriel, M. Blonsky, M. Johannes, R. Ramanathan, R. Duan, Z. Ziyu
- VASPsol solvation model and nanocrystals: K. Mathew, J. Gabriel, C. Bealing
- GASP genetic algorithm and machine learning: **B. Revard**, **W. Tipton**, **A. Yesupenko**, B. Antonio, S. Honrao
- Financial support by NSF-CAREER, NSF-SSI, NIST, DOE-EFRC
- Computational resources provided by HiPerGator@UF and NSF XSEDE

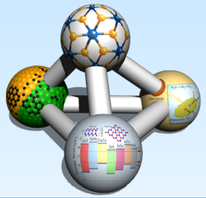
Powered by
MPInterfaces & materialsweb

rhennig@ufl.edu
http://hennig.mse.ufl.edu

Part I: The Energy Landscape

rhennig@ufl.edu
http://hennig.mse.ufl.edu

Materials Structure Space



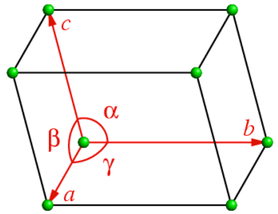
Mixed representation of materials structure space (real space representation)

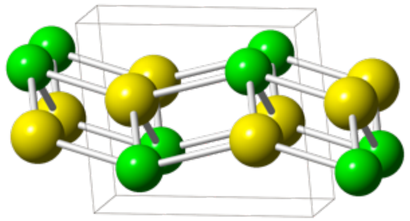
Real variables:

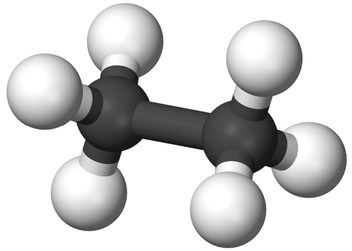
- Atom positions
- Lattice parameters (for crystalline structures)

Integer variables:

- Number of atoms
- Atomic species







Dimensionality of materials structure space

- Determined by number of atoms: $3N_{atom} + 3$ for crystals, $3N_{atom}$ for molecules and clusters

Complexity of materials structure space

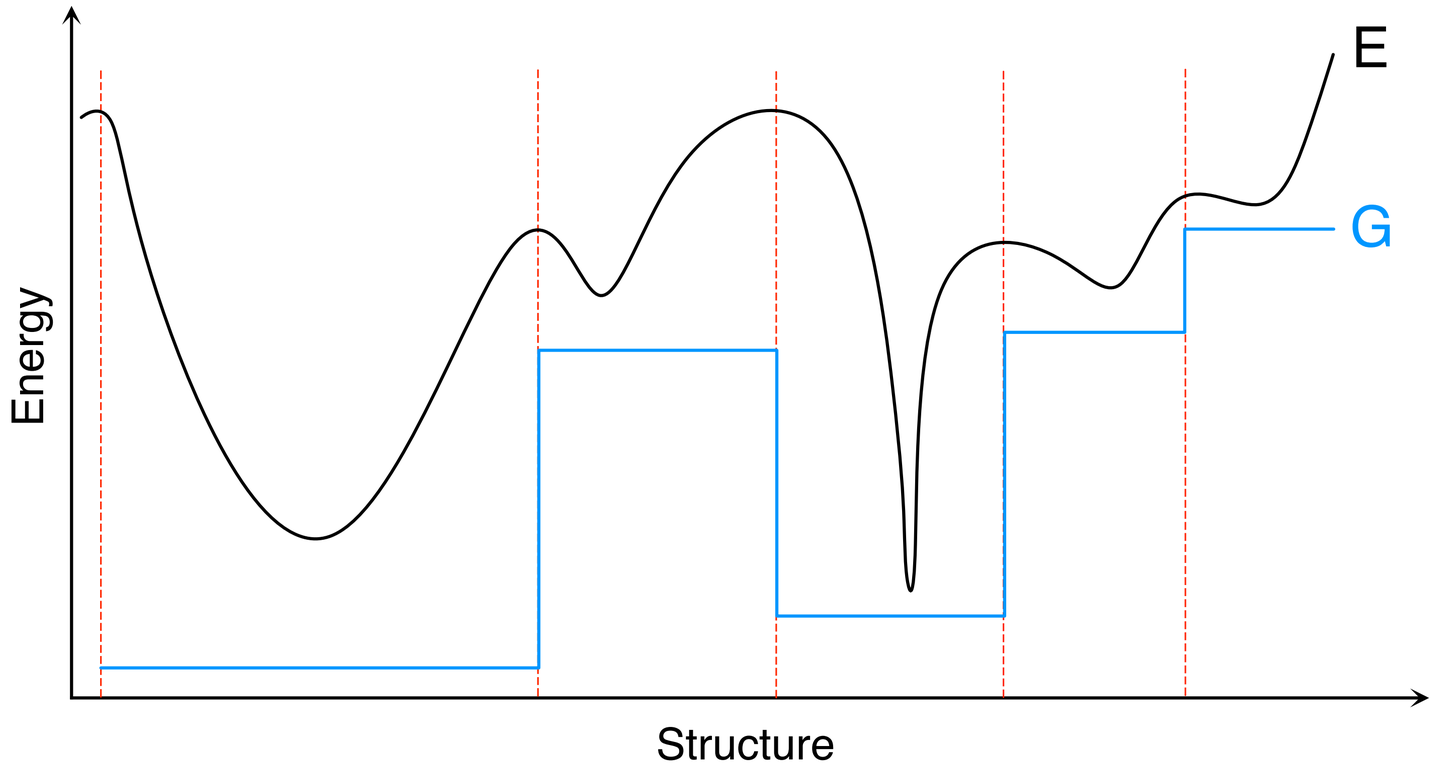
- Exponential increases with number of atoms and number of species
- Reduced by space and point group symmetries

Energy landscape space has continues and integer variables

Minimum Gibbs Free Energy Principle

Computational structure prediction based on optimization

- Stable structure at (p, T) ⇒ Lowest Gibbs free energy, $G(p, T) = E(V, S) - TS + pV$
- Gibbs free energy of a structure is an integral over the basin of the local minimum
  - Expensive to calculate
  - Difficult to explore due to discontinuities
  - Can contain minima not present in the energy landscape

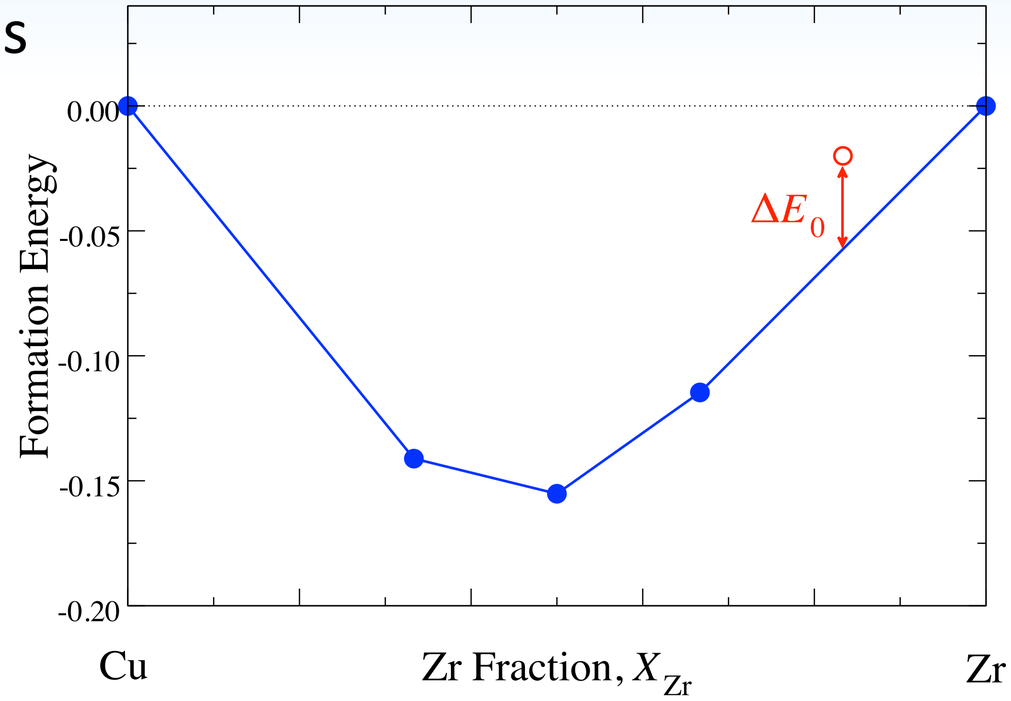


Free energy landscape space is discontinuous and difficult to search

Minimum Energy Principle

Computational structure prediction based on energy minimization

- Combined optimization over real and integer variables
- For multi-component materials, we need to consider stability against competing phases
  - ➡ Concept of convex hull



Energy methods

Interpolative
(Semi-)empirical methods

Extrapolative and predictive
First-principles or *ab-initio* methods

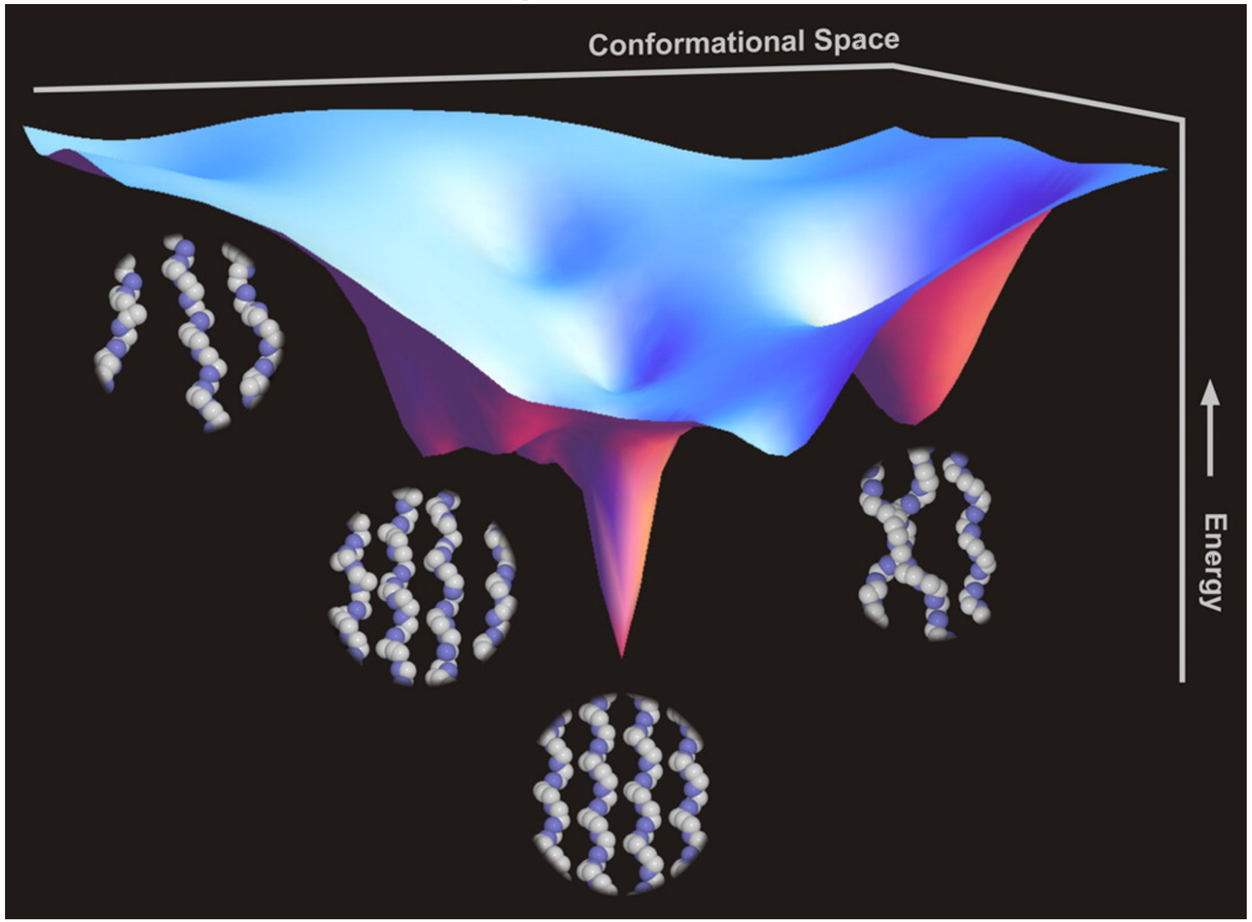
- Density-functional theory offers balance of speed and accuracy
- Pseudopotentials and plane-wave basis (VASP)

Structure for multi-component materials requires knowledge of competing phases

General Features of the Energy Landscape

Bell–Evans–Polanyi principle

- Highly exothermic chemical reactions have low activation energies
- Low-energy basins are expected to occur near other low-energy basins
- These regions are referred to as **‘funnels’**

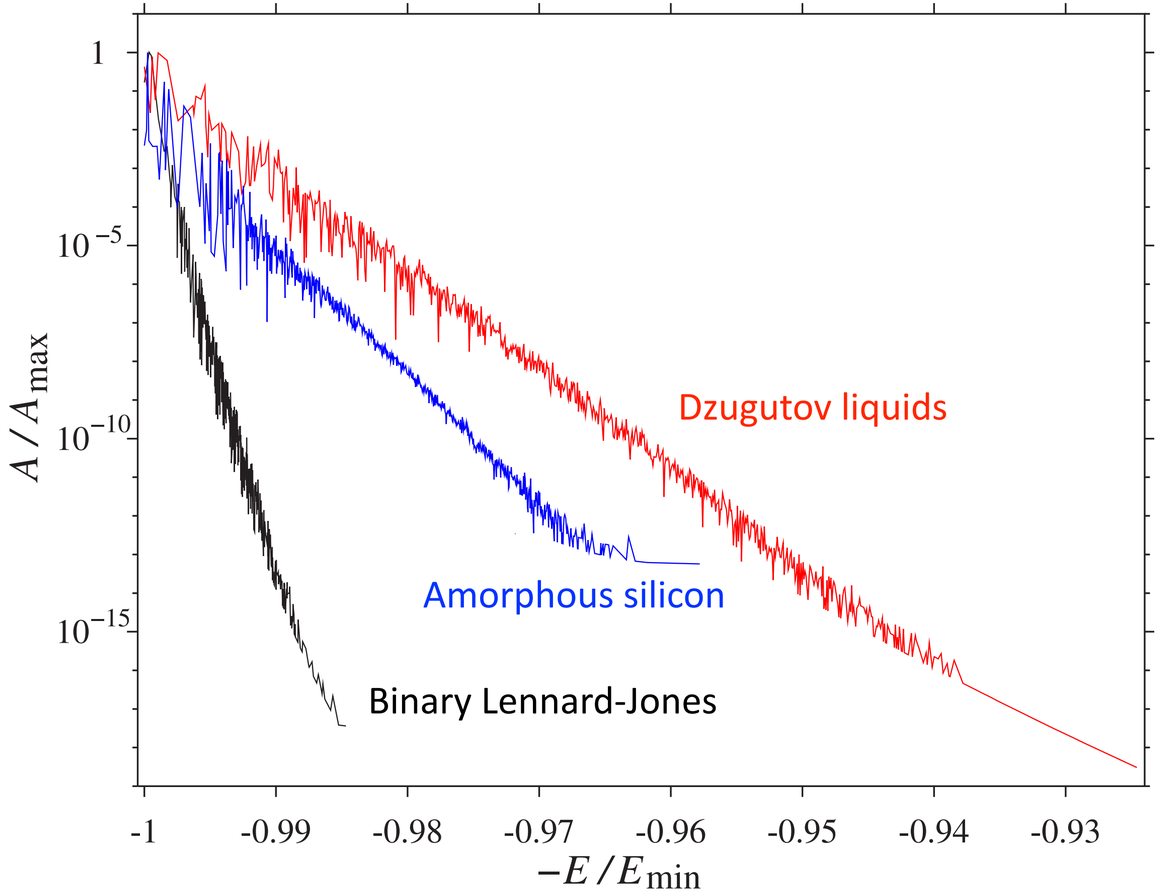


rhennig@ufl.edu
http://hennig.mse.ufl.edu

General Features of the Energy Landscape

Probability distribution of energies of local minima

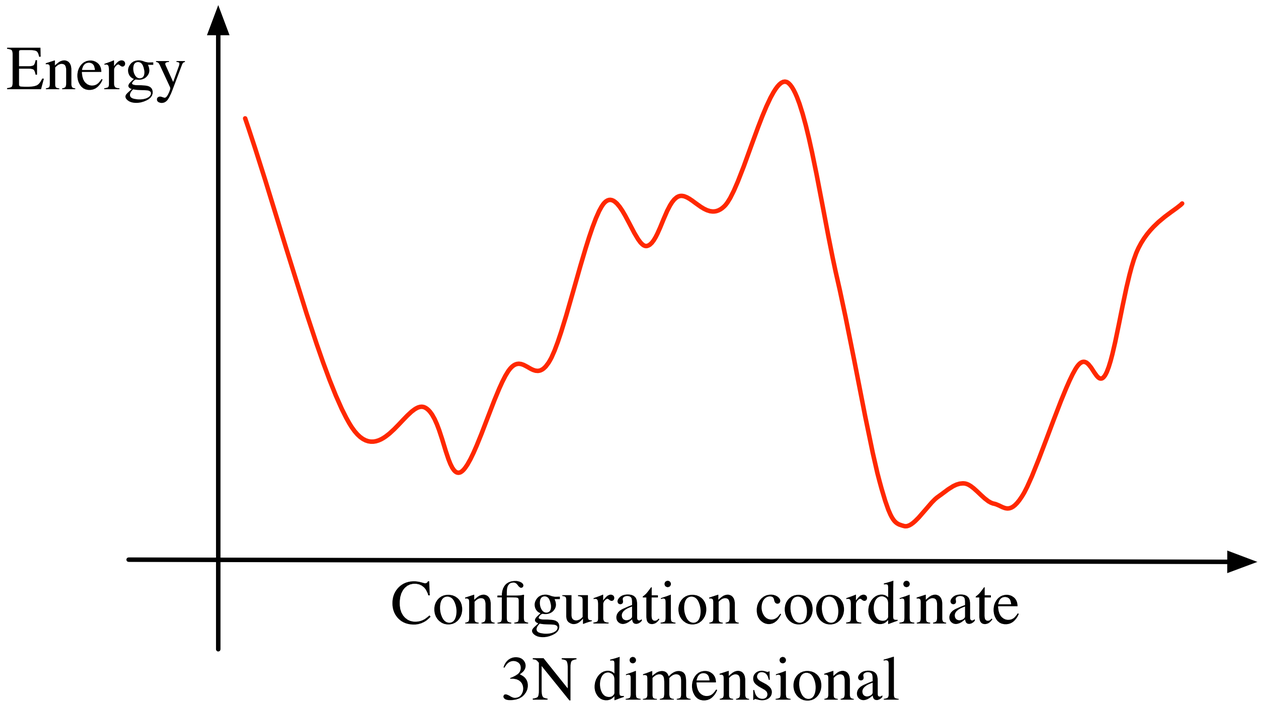
- Basins with lower-energy minima have larger hyper-volumes
- Related to similar elastic constants and vibrational frequencies for different structures
- Power law probability distribution of these hyper-volumes
- Order in the arrangement of basins of different sizes
- Smaller basins filling gaps between larger ones

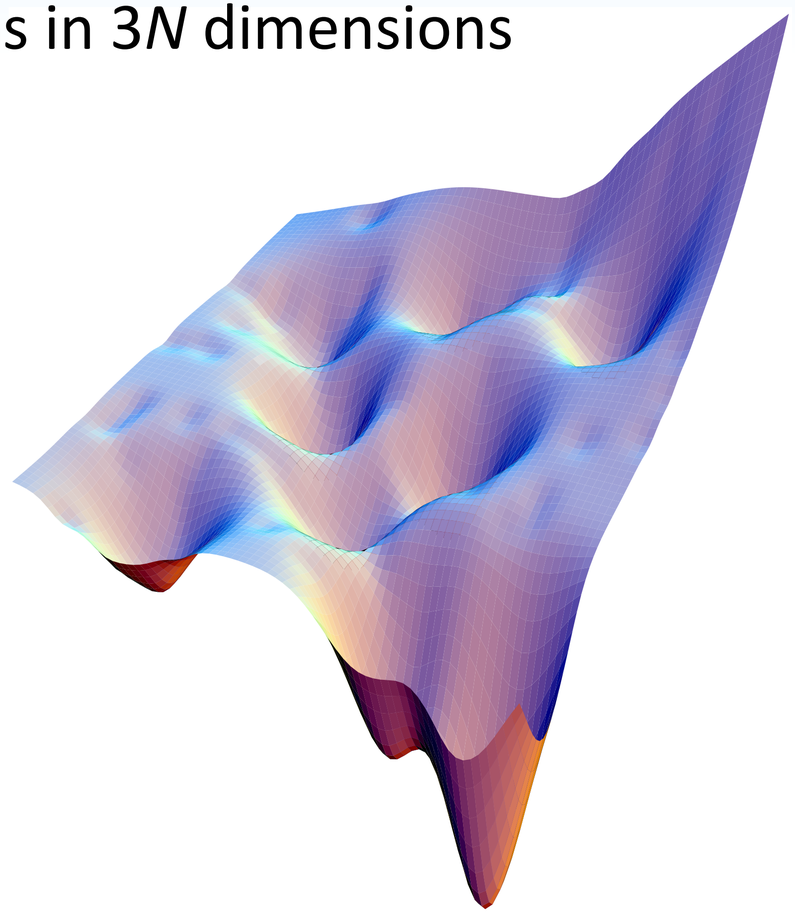


Massen & Doye, Phys. Rev. E **75**, 037101 (2007)

Illustration of Energy Landscape

- 1D and 2D cuts through 3*N*-dimensional configuration space
- Not a good representation of the distribution of basins in 3*N* dimensions

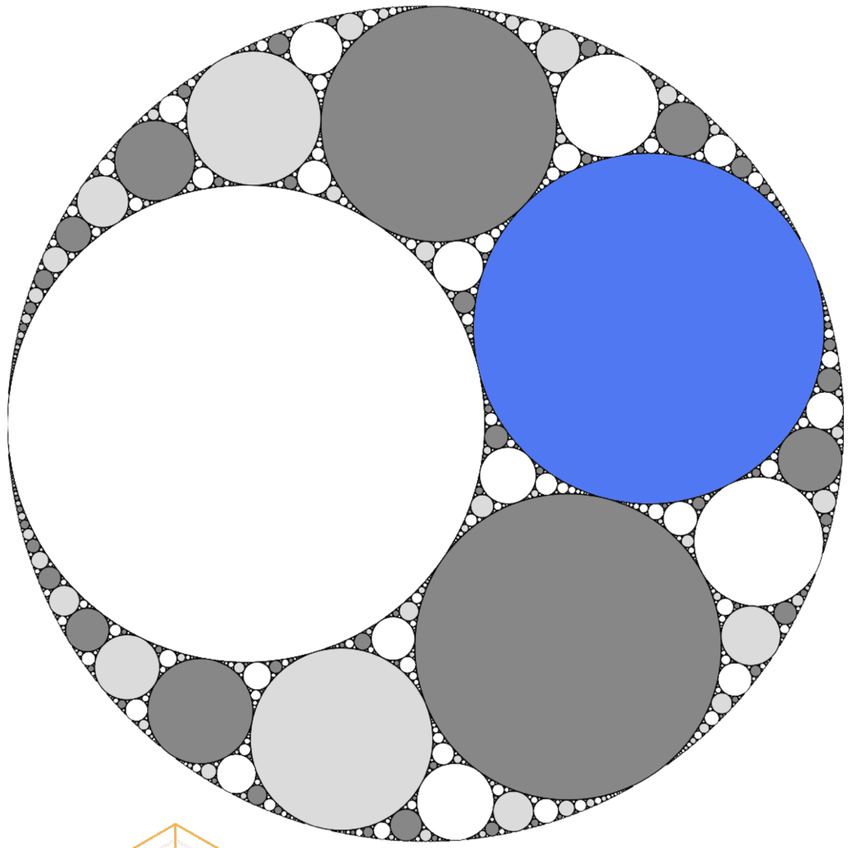


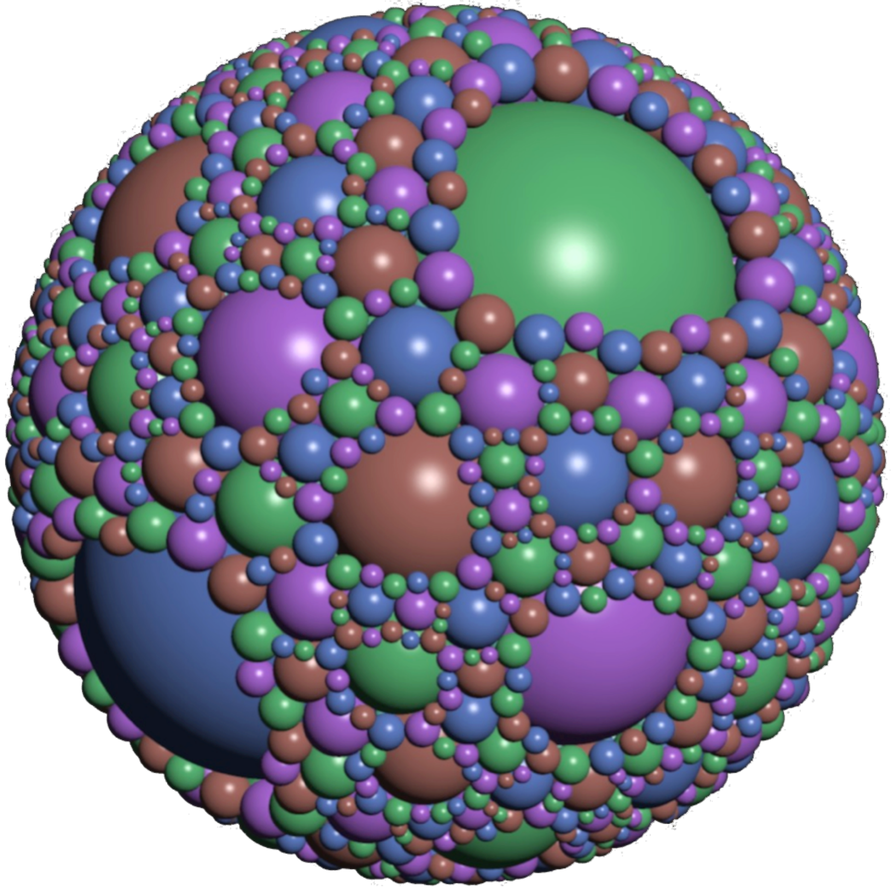


rhennig@ufl.edu
http://hennig.mse.ufl.edu

Illustration of Energy Landscape

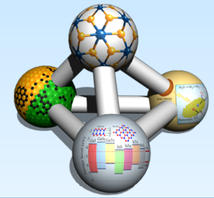
- 1D and 2D cuts through 3*N*-dimensional configuration space
- Not a good representation of the distribution of basins in 3*N* dimensions
- Similarities between Apollonian sphere packings and energy landscapes



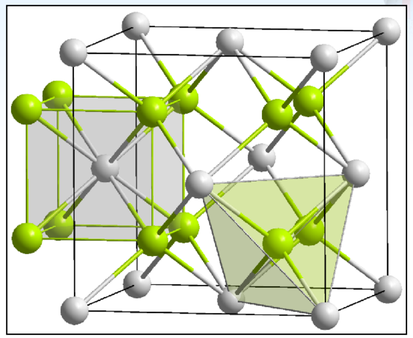


Doye & Massen Phys. Rev. E **71**, 016128 (2005)

Symmetry and Structural Motifs



The Rule of Parsimony: *"The Number of essentially different kinds of constituents in a crystal tends to be small."* (Linus Pauling 1929)

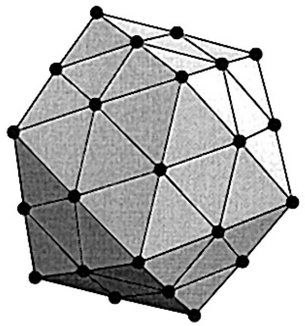


Correlation between energy and symmetry

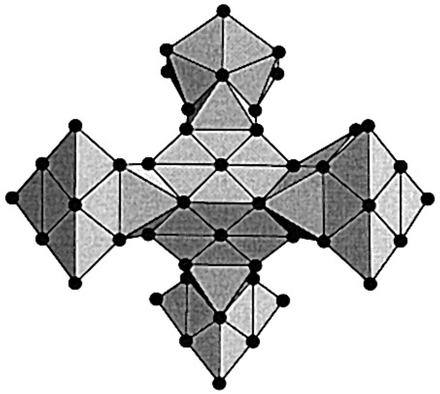
- Low (and high) energy minima tend to correspond to symmetrical structures
- High symmetry of low-energy minima supported by the ubiquity of crystals

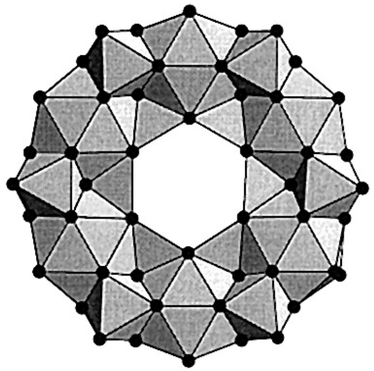
Example: 55-atom Lennard-Jones clusters (D. Wales '98)

Lowest energy



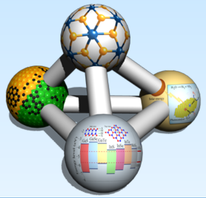
Two highest energy





rhennig@ufl.edu
http://hennig.mse.ufl.edu

Frequency of Space Groups



For crystals of small organic molecules

- 75% of about 30,000 compounds occur in only five space groups
- 29 space groups only have one entry and 35 space groups none at all

Space Group	$P2_1/c$	$P\bar{1}$	$P2_12_12_1$	$P2_1$	$C2/c$
Frequency	36%	14%	12%	7%	7%

Inorganic systems show different space group frequencies

- 67% of about 100,000 compounds occur in only 24 space groups

Space Group	*Pnma*	$P2_1/c$	$Fm\bar{3}m$	$Fd\bar{3}m$
Frequency	7.4%	7.2%	5.6%	5.1%
Examples	Fe_3C, $CaTiO_3$	ZrO_2	Cu	C, Cu_2Mg

- Note: bcc is not among one of the top 24 space groups

Some space groups are much more common than others in crystals

Specific Features of Energy Landscapes

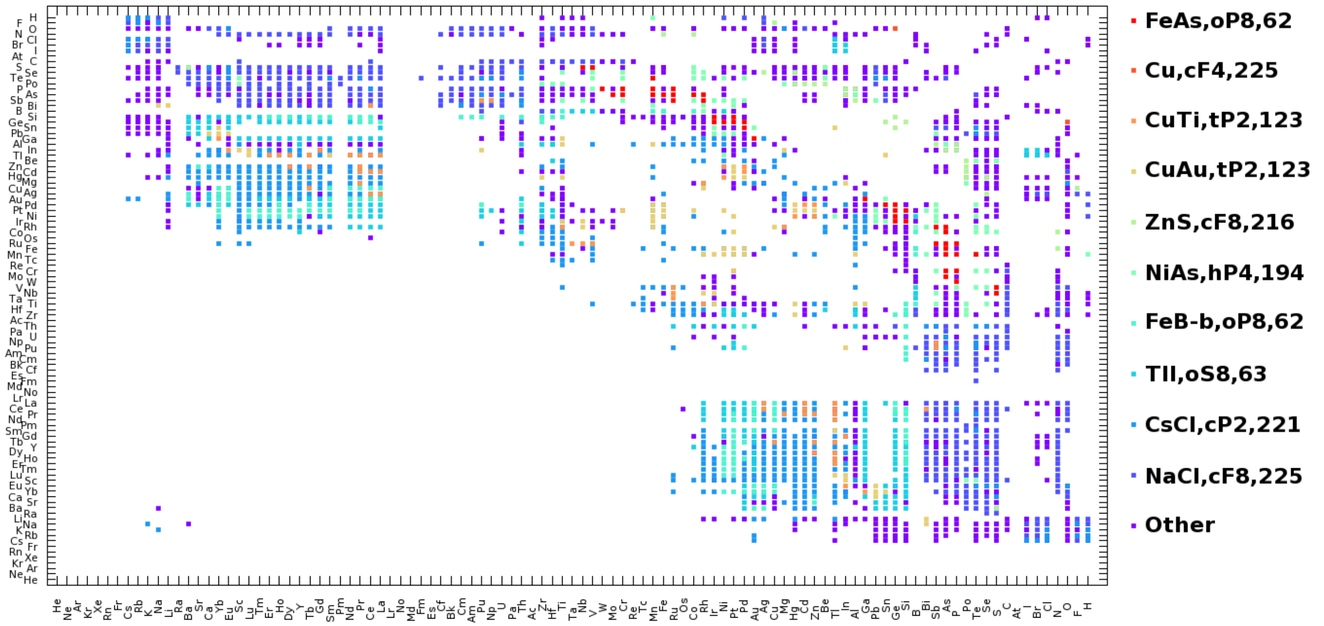
Chemical considerations

- Know a great deal about the chemistry of the systems we study
- Know which atomic types prefer to bond to one another
- Approximate bond lengths
- Likely coordination numbers of the atoms

Resulting empirical rules

- Hume-Rothery rules
- Laves rules for intermetallics
- Pauling rules for ionic materials
- Pettifor structure maps

...



How can we utilize this information to accelerate structure searches?

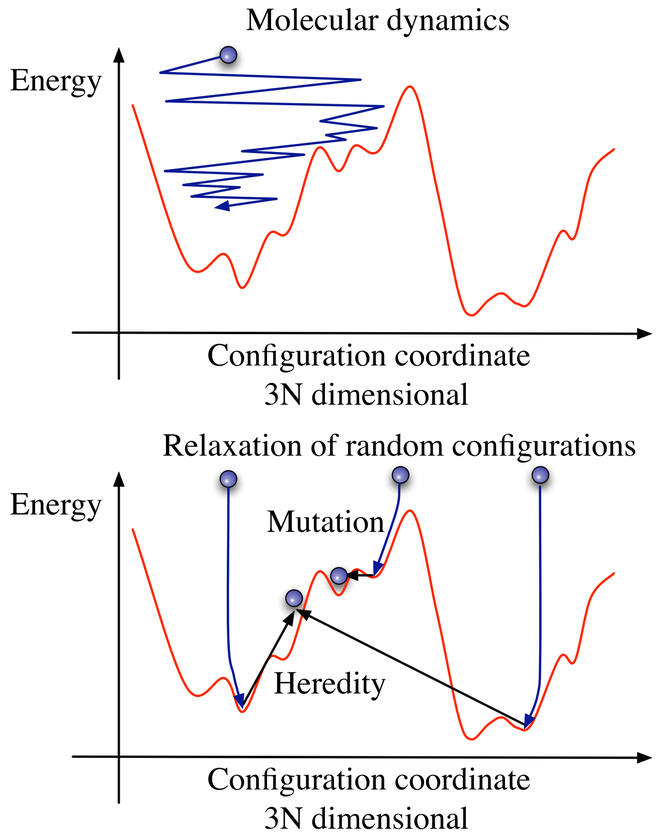
Structure Prediction Methods

Limited success of conventional optimization methods

- Guess of structures based on symmetry and chemistry (trial-and-error)
- Simulated annealing (molecular dynamics, cooling from high temperature)

Advances in optimization methods:

- **Accelerated molecular dynamics**, Voter *et al.* ’97
- **Metadynamics**, Parinello *et al*. ’02
- **Minima hopping**, search for neighboring minima, Goedecker ’04
- **Random search**, randomly sampling phase space, Pickard & Needs ’06
- **Evolutionary algorithms**
  Generation of structures, Bush *et al*. ’95, Woodley *et al*. ’99, Oganov ’06
- **Particle Swarm Optimization**, Ma *et al.* ’10
- **Datamining of structure databases**, Ceder *et al*.

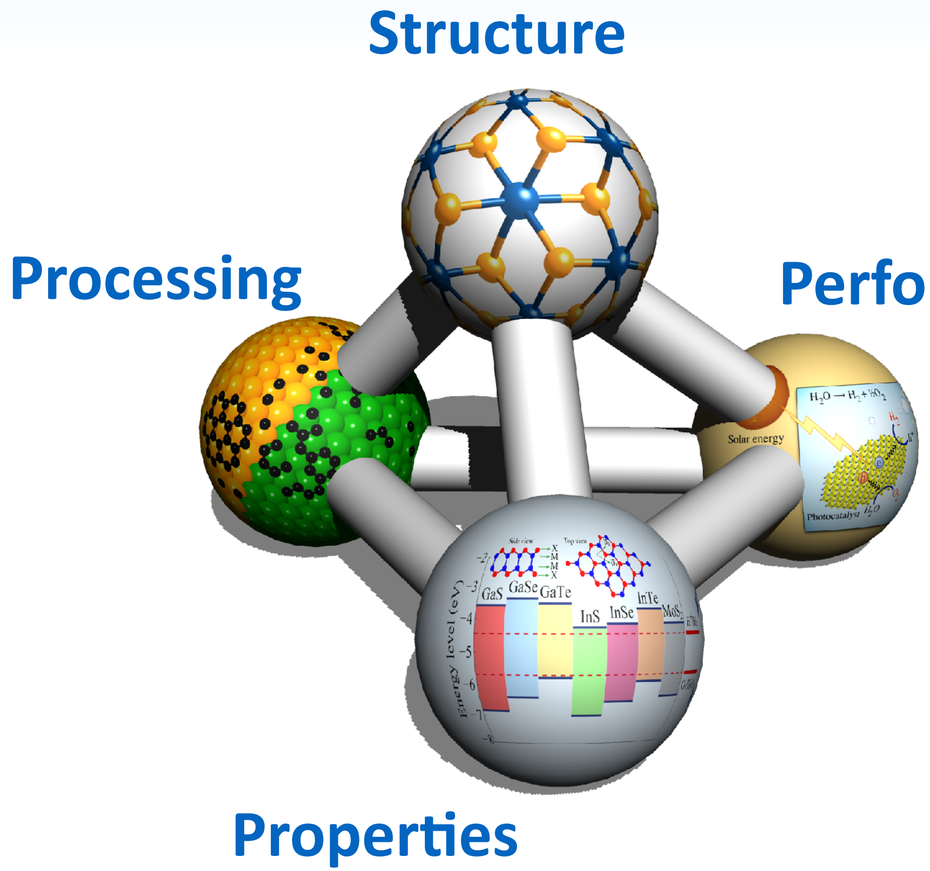


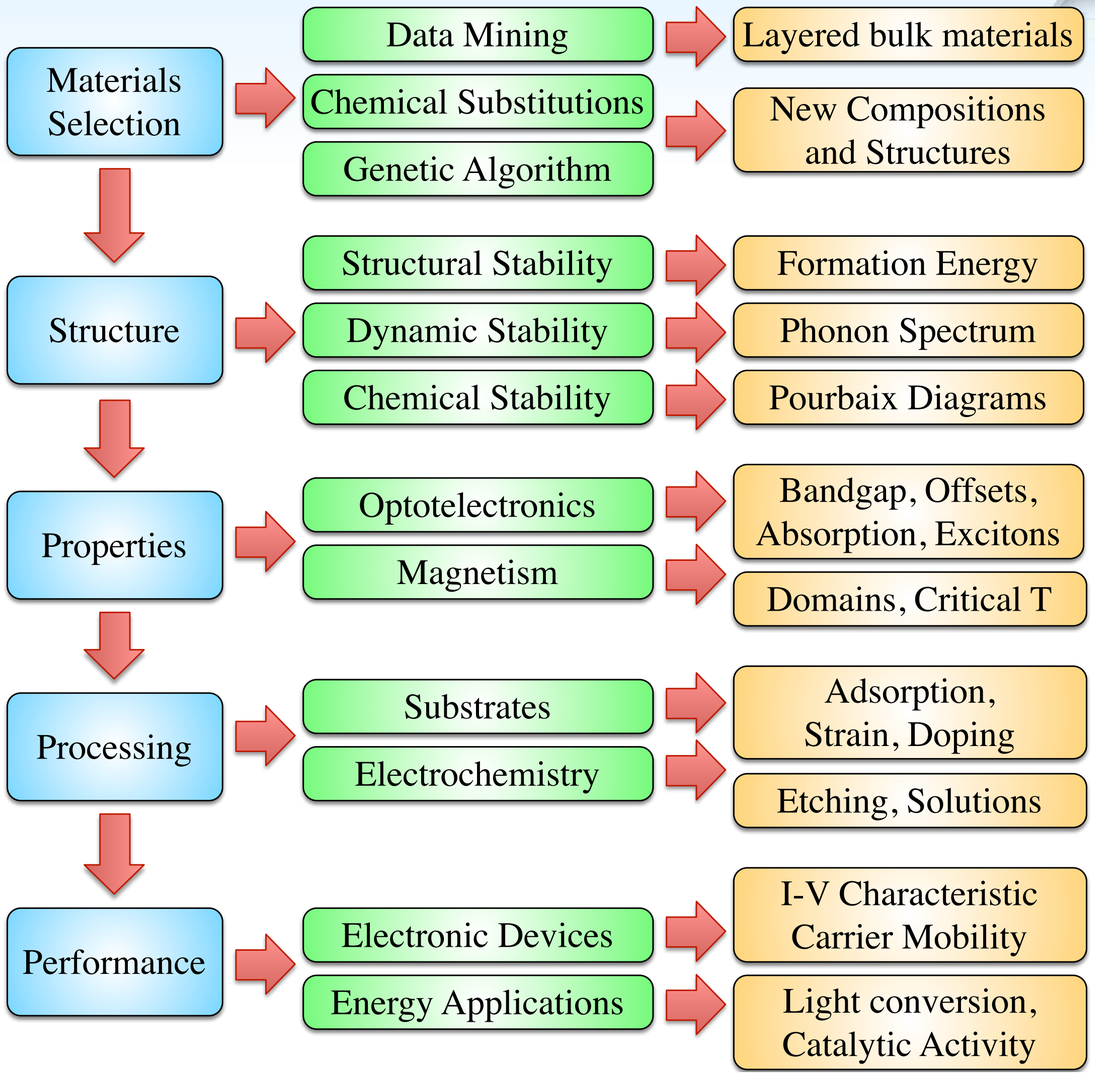
rhennig@ufl.edu
http://hennig.mse.ufl.edu

Part I: Discovery of 2D Materials

rhennig@ufl.edu
http://hennig.mse.ufl.edu

Materials Informatics of 2D Materials

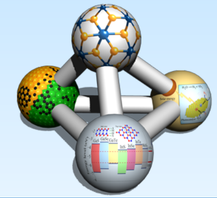


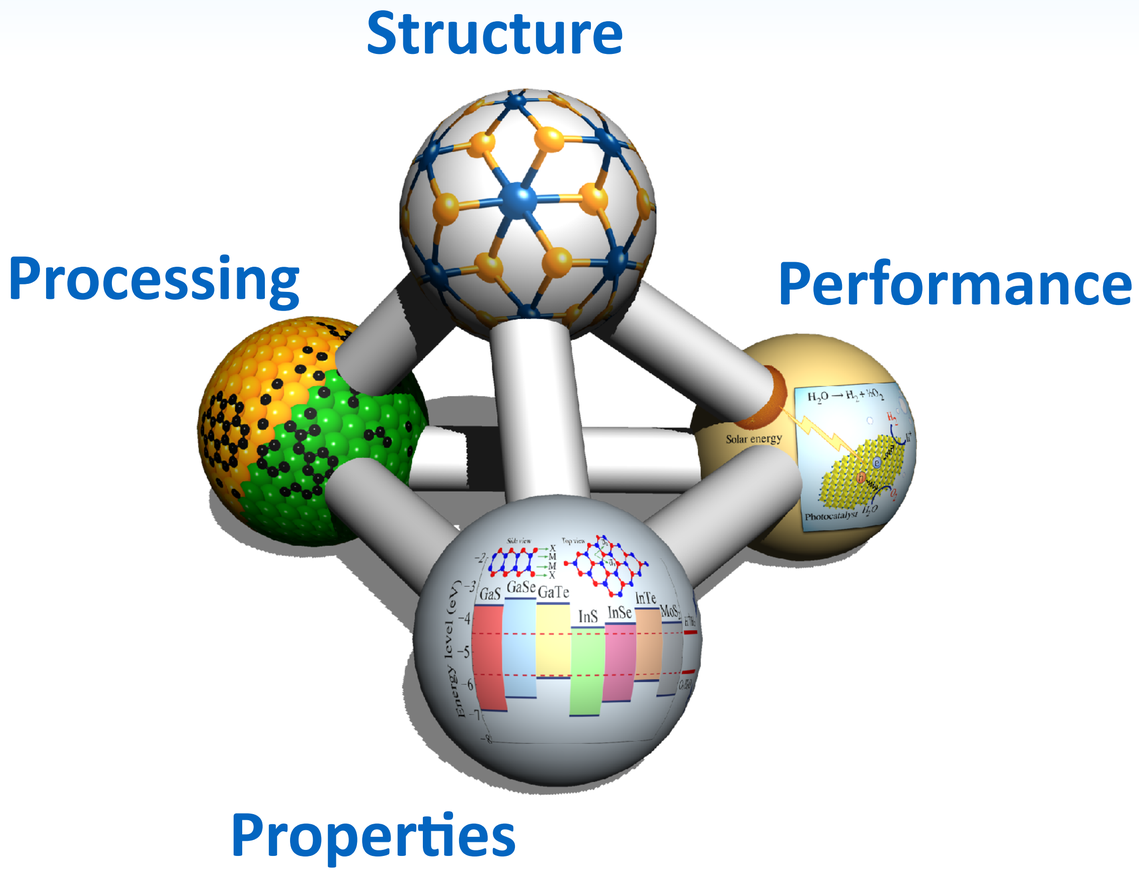


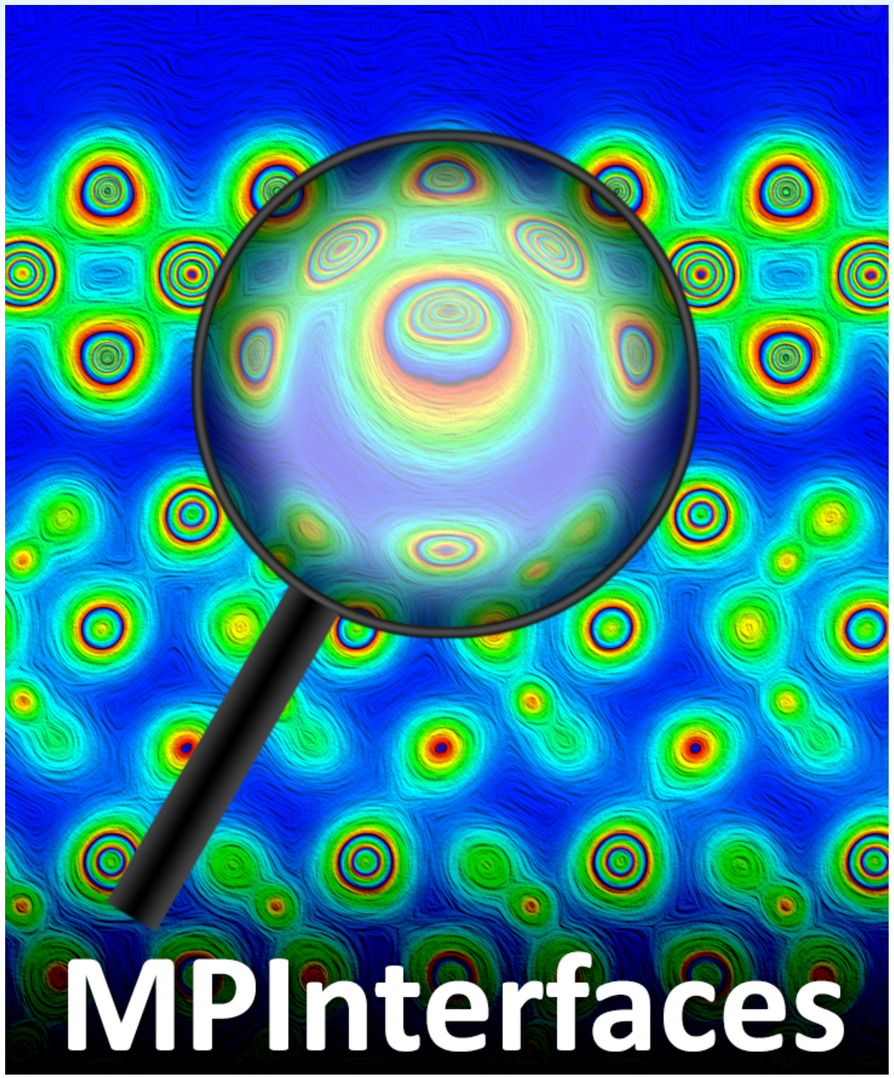
H. L. Zhuang and RGH, JOM 66, 366 (2014)

Powered by
MPInterfaces & materialsweb

Materials Informatics of 2D Materials





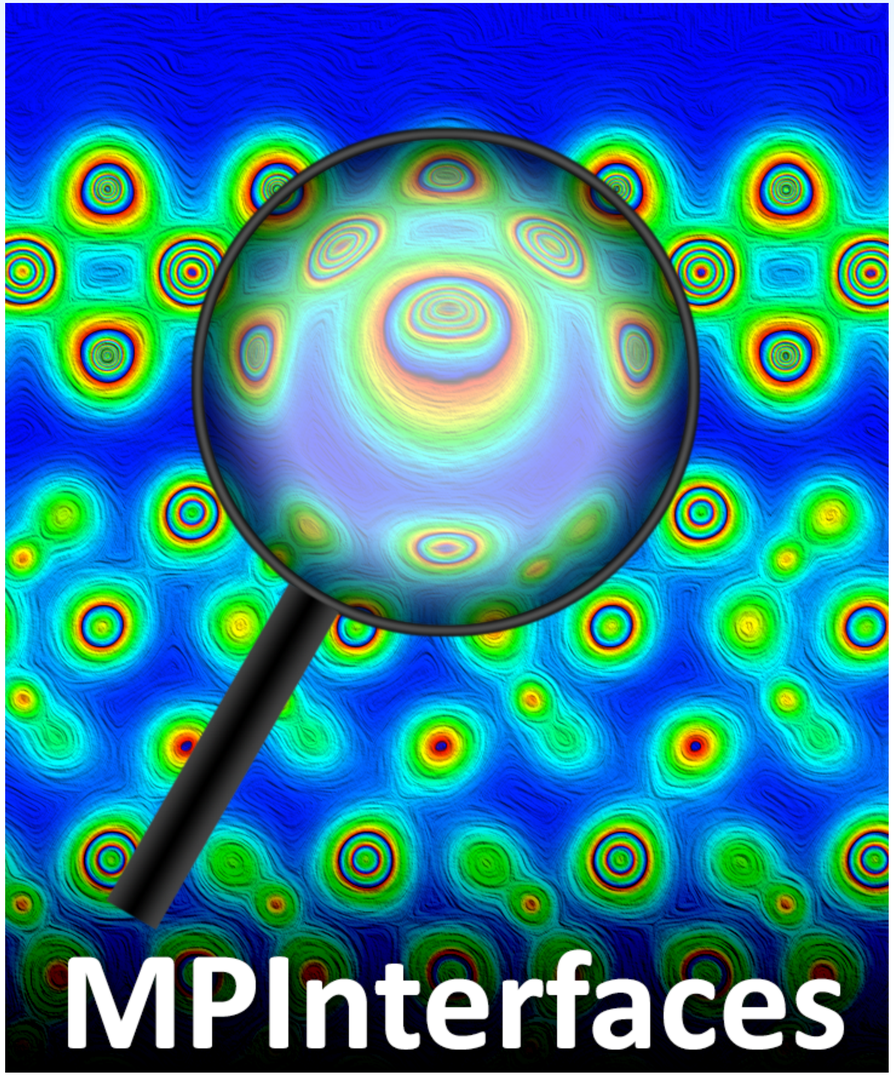


H. L. Zhuang and RGH, JOM 66, 366 (2014)

Powered by
MPInterfaces & materialsweb

Structure and Stability of 2D Materials

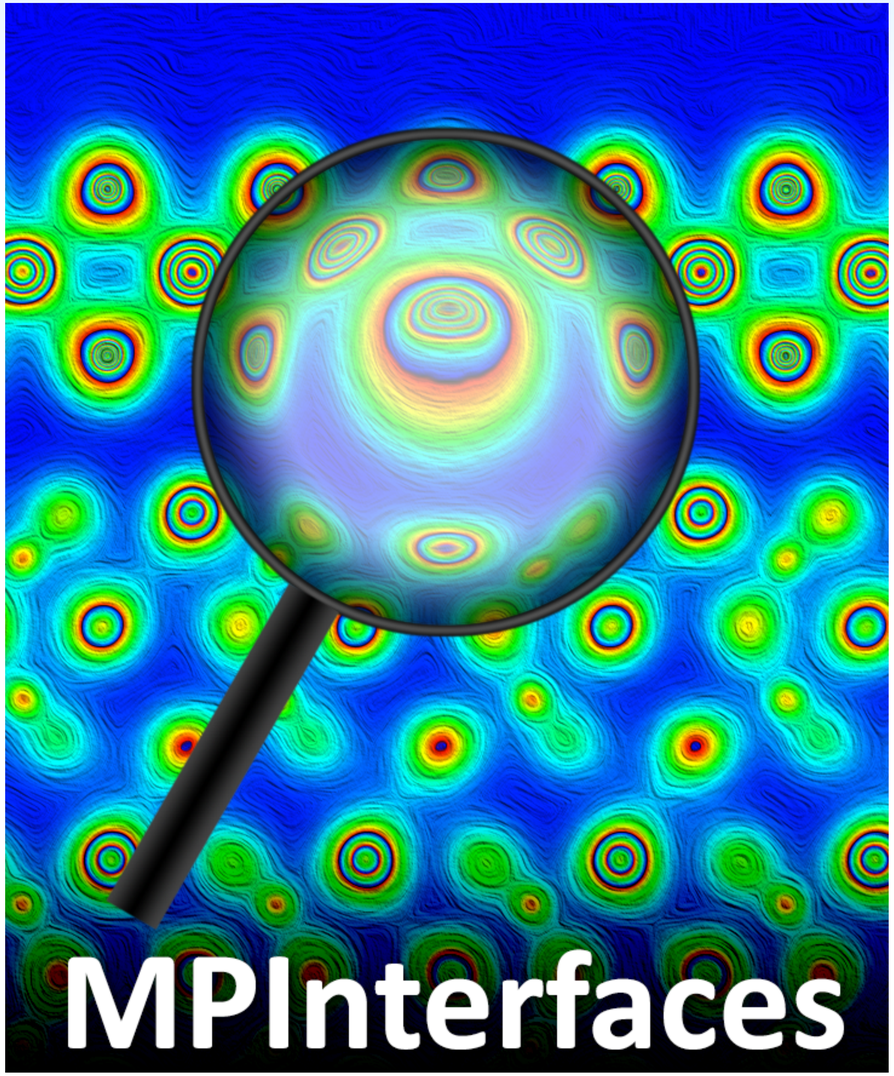
- Classification of 2D materials
- Criteria for stability $\Delta E_f < 200$ meV/atom
- Methods for 2D materials discovery
  - Datamining for exfoliation
  - Chemical substitutions and etching
  - Evolutionary algorithm searches



rhennig@ufl.edu
http://hennig.mse.ufl.edu

Structure and Stability of 2D Materials

- **Classification of 2D materials**
- Criteria for stability $\Delta E_f < 200$ meV/atom
- Methods for 2D materials discovery
  - Datamining for exfoliation
  - Chemical substitutions and etching
  - Evolutionary algorithm searches



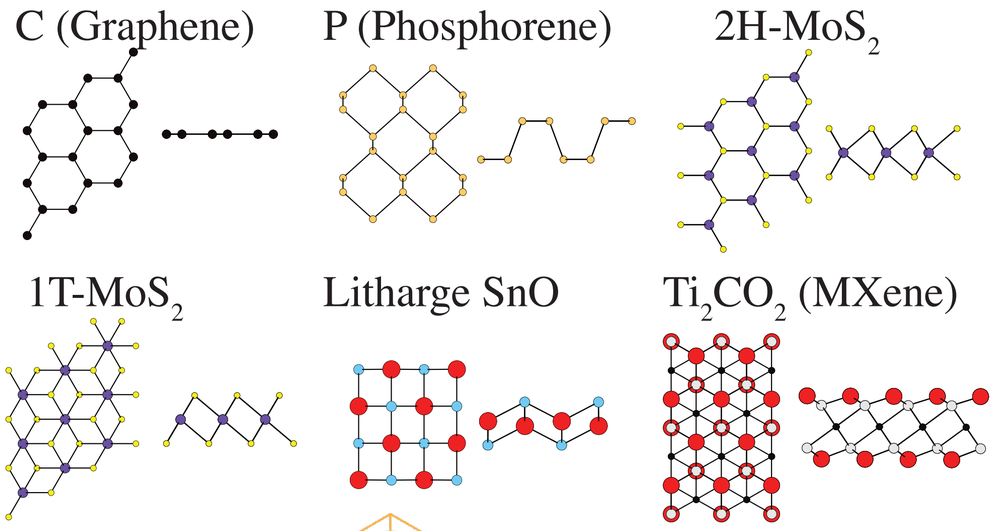
rhennig@ufl.edu
http://hennig.mse.ufl.edu

Classification of 2D Materials

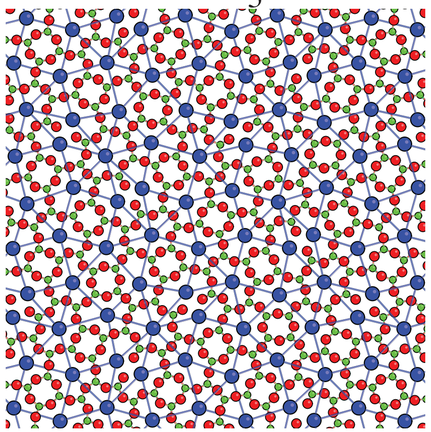
Definition: A 2D material is a material with finite thickness in one dimension and essentially infinite extent in the other two dimensions.

- Classification into crystalline and amorphous materials
- 2D crystals include **periodic** and **aperiodic** crystals based on point group
- Aperiodic 2D crystals include **incommensurate crystals** and **quasicrystals**

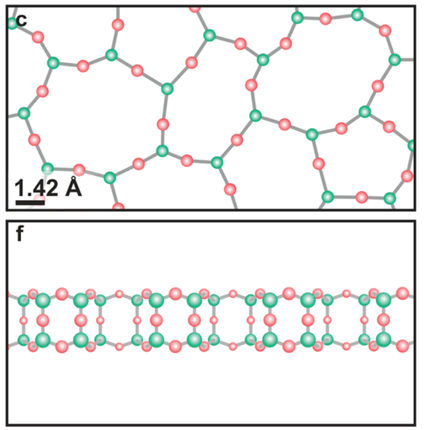
a) Crystalline 2D Materials



b) Quasicrystalline 2D $BaTiO_3$



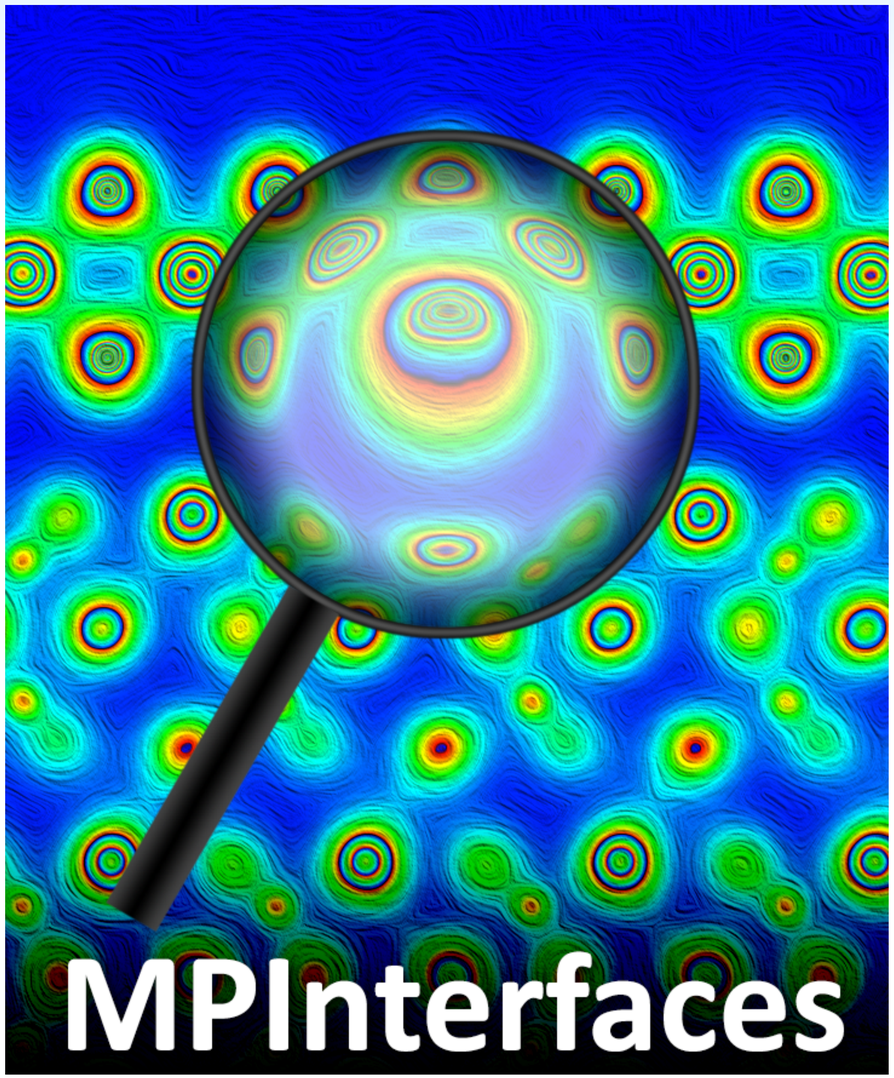
c) Amorphous 2D SiO_2



Paul, Hennig, *et al*. submitted 2017

Structure and Stability of 2D Materials

- Classification of 2D materials
- **Criteria for stability ΔE_f < 200 meV/atom**
- Methods for 2D materials discovery
  - Datamining for exfoliation
  - Chemical substitutions and etching
  - Evolutionary algorithm searches



Thermodynamic Stability of 2D Materials

Formation energy

$$\Delta G_\mathrm{f} = G_\mathrm{2D} - \min_\mathrm{phases} G_\mathrm{3D}$$

$$\text{approximated by } \Delta E_\mathrm{f} = E_\mathrm{2D} - \min_\mathrm{phases} E_\mathrm{3D}$$

min over single or mixture of phases

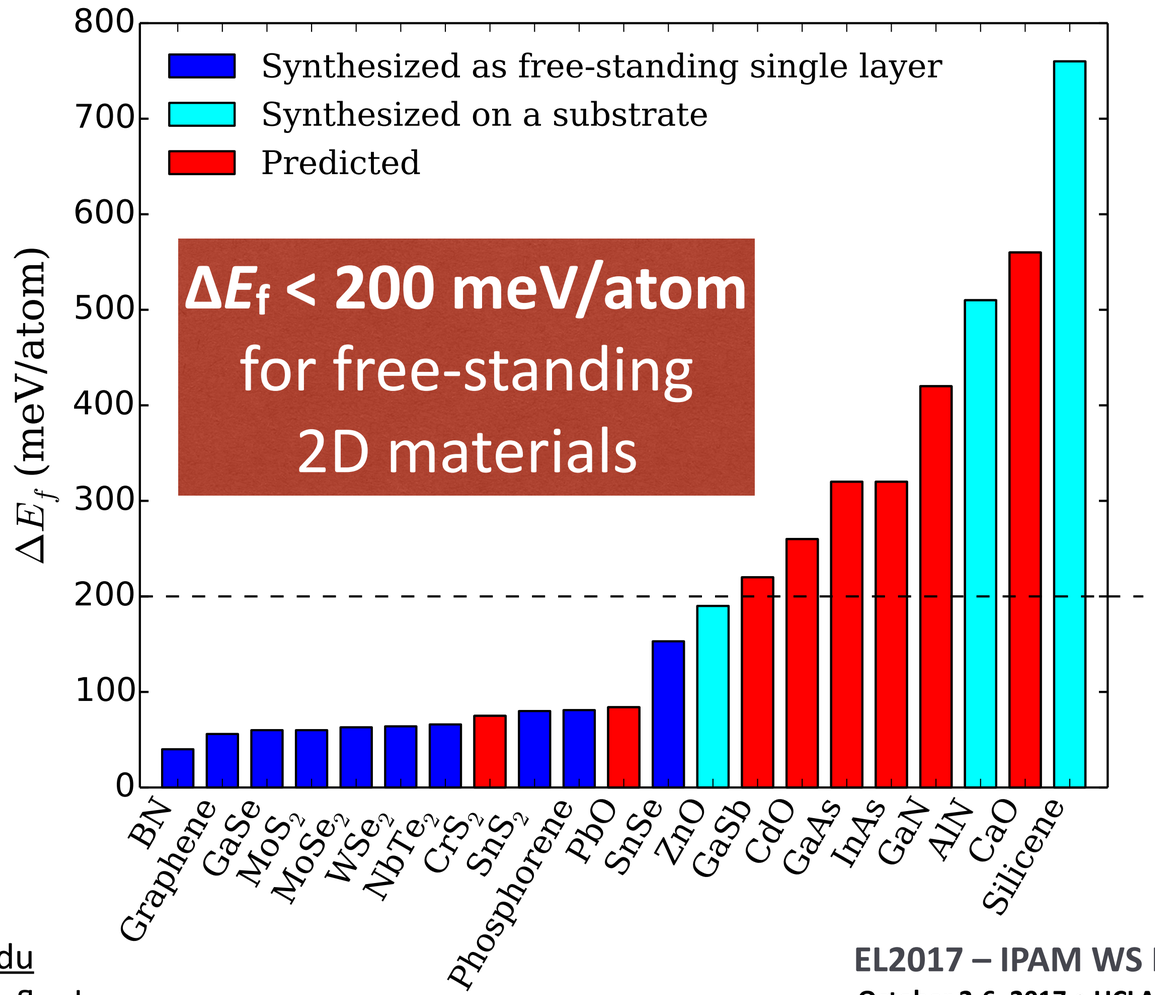
- Comparison of VASP PAW using vdW-DF optB88 with QMC and RPA shows accuracy of ΔE_f of 30 meV/atom

2D materials are metastable

- Energy can always be lowered by van der Waals interactions for multi-layers

Stabilization by substrates

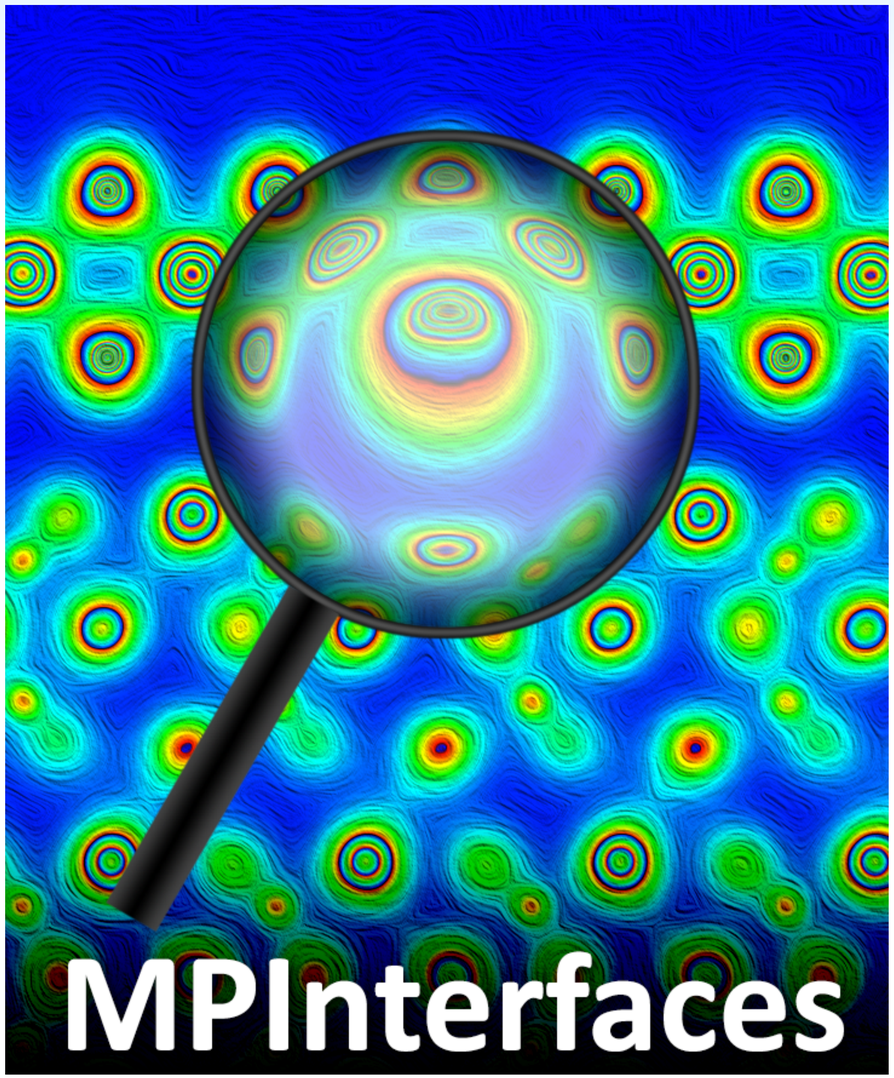
- Substrates used to nucleate and grow 2D materials
- Stabilization through physi- and chemisorption



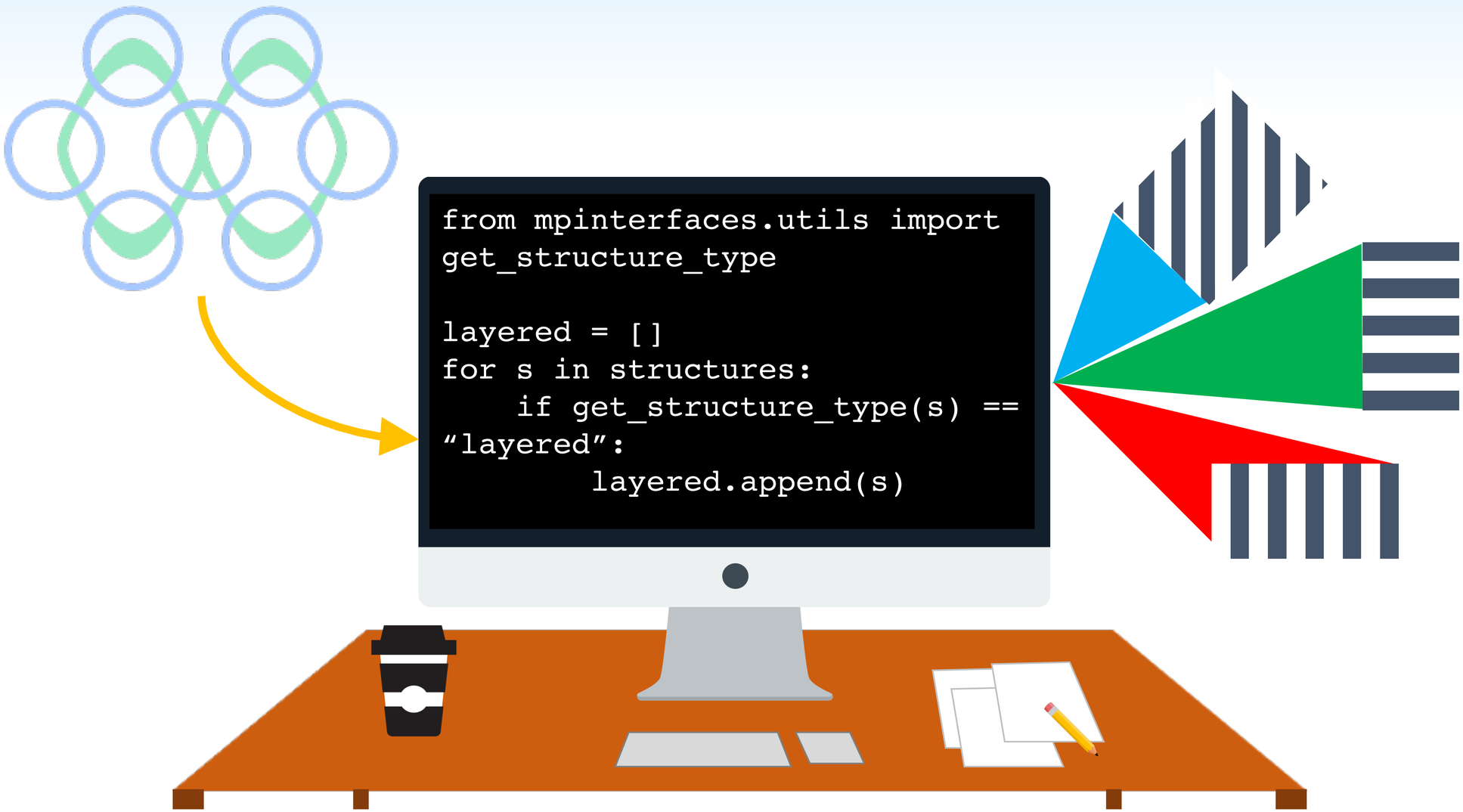
rhennig@ufl.edu
http://hennig.mse.ufl.edu

Structure and Stability of 2D Materials

- Classification of 2D materials
- Criteria for stability $\Delta E_f < 200$ meV/atom
- **Methods for 2D materials discovery**
  - **Datamining for exfoliation**
  - Chemical substitutions and etching
  - Evolutionary algorithm searches

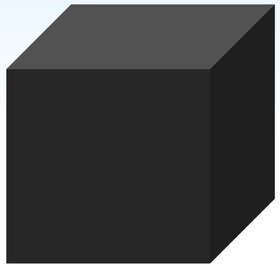


Datamining to Discover 2D Materials



rhennig@ufl.edu
http://hennig.mse.ufl.edu

Dimensionality of Structural Motifs

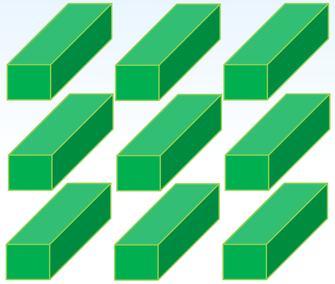


3D

Conventional & intercalated crystals

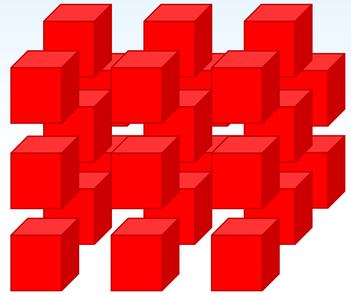
2D

Layered crystals



1D

Chain-like crystals



0D

Molecular & atomic crystals

Large numbers of crystals with each dimensionality exist in most major crystal structure databases. How can we classify them?

rhennig@ufl.edu
http://hennig.mse.ufl.edu

Searching for Layered Compounds

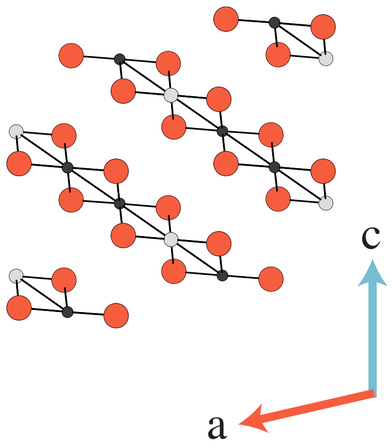
The ideal case

Search for materials with:
- low packing fractions
- gaps along c-axis

You'll find about 100 layered materials in the ICSD[1,2]

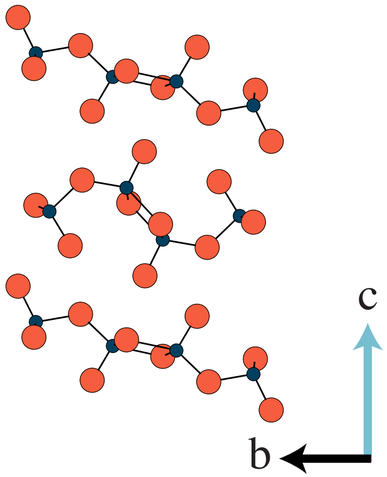
Reality

Co_2NiO_6
mp-765906



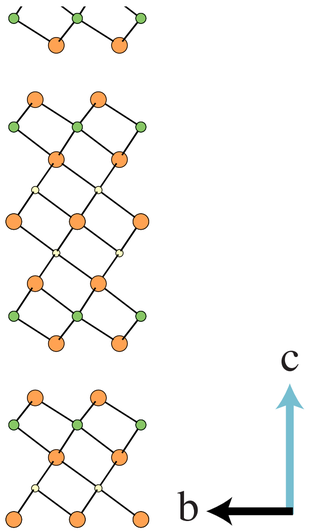
Gaps along other axes

V_2O_5
mp-25643



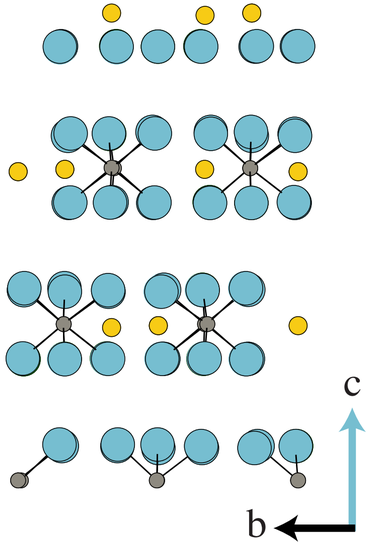
No planar gaps at all

$Ge_2Te_5As_2$
mp-14791



High packing fractions

Li-WCl_6
mp-570512

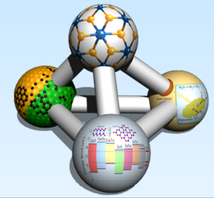


Gaps along multiple axis

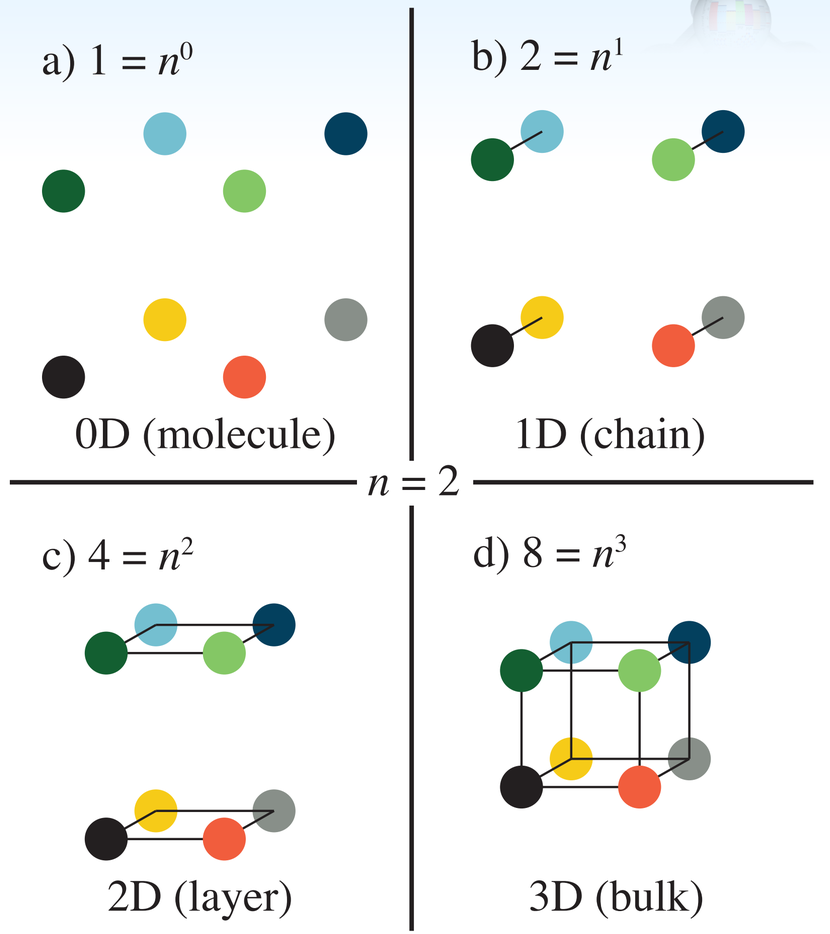
Powered by
MPInterfaces & materialsweb

[1] Björkman, et al. *Phys. Rev. Lett.* 2012
[2] Lebègue, et al. *Phys. Rev. X* 2013

The Topology Scaling Algorithm



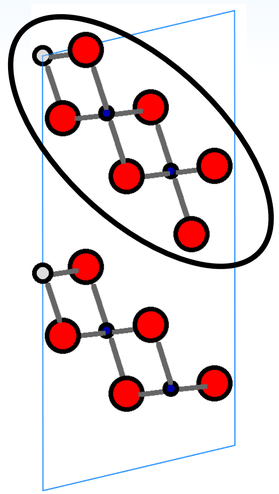
1. Identify bonded atomic clusters in the structure based on covalent radius overlap
2. Make a $n \times n \times n$ supercell of the structure
3. Re-identify bonded atomic clusters, and compare sizes with original cluster sizes



The Topology Scaling Algorithm

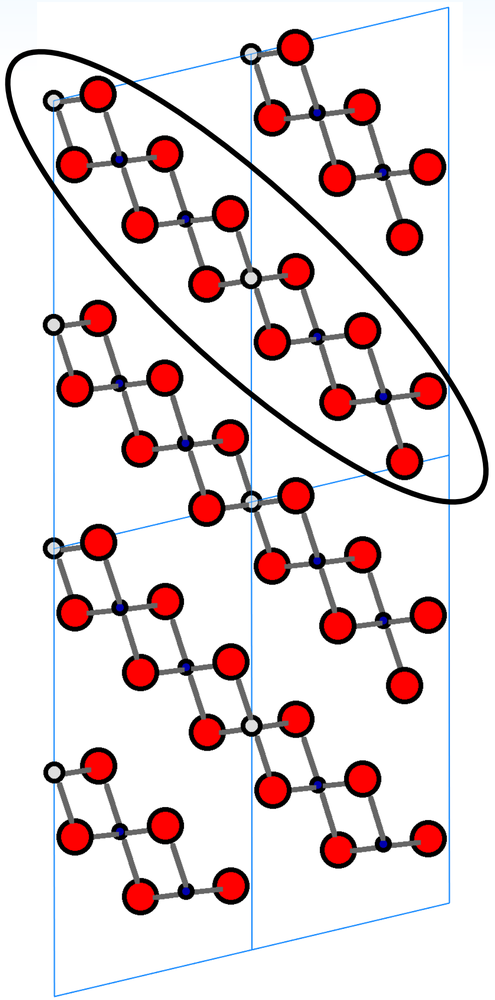
Co_2NiO_6
mp-765906

1×1×1 cell



Cluster of 9 atoms

2×2×2 cell



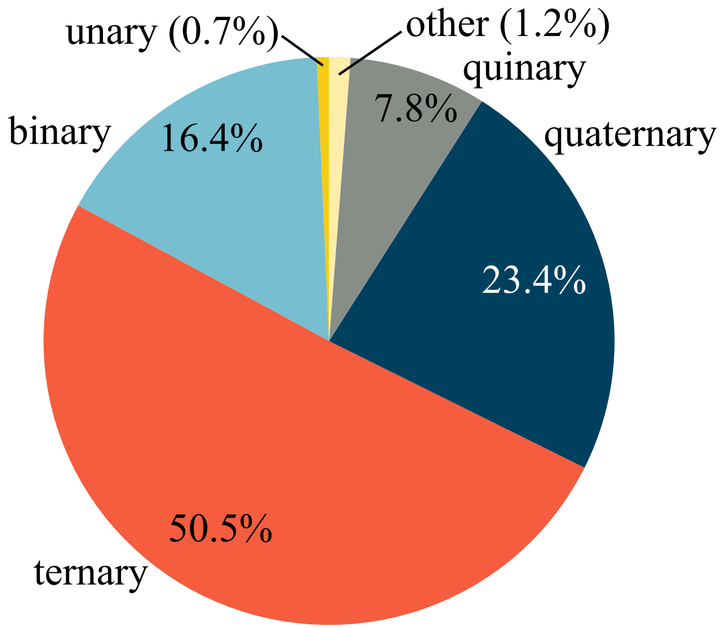
Cluster of 36 atoms

N^2 scaling ⇒ Layered Material

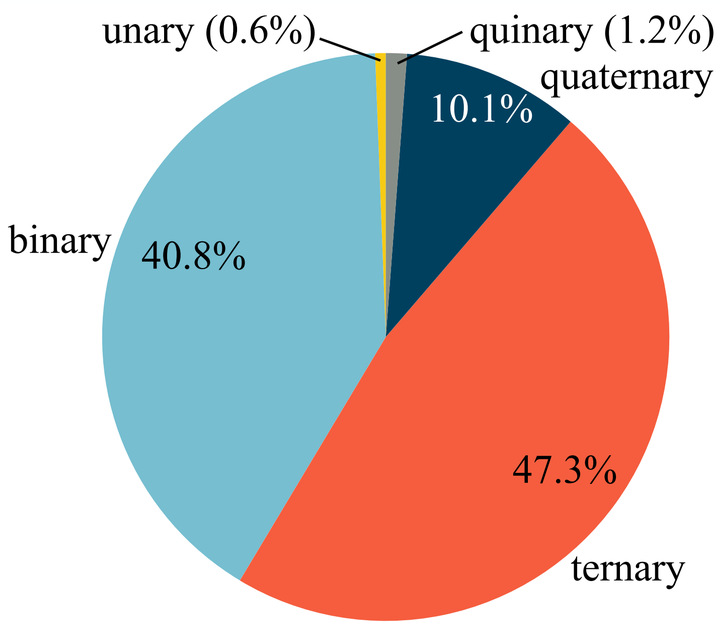
Ashton, Paul, Sinnott & Hennig. *Phys. Rev. Lett.* 2017

Layered Materials in MaterialsProject

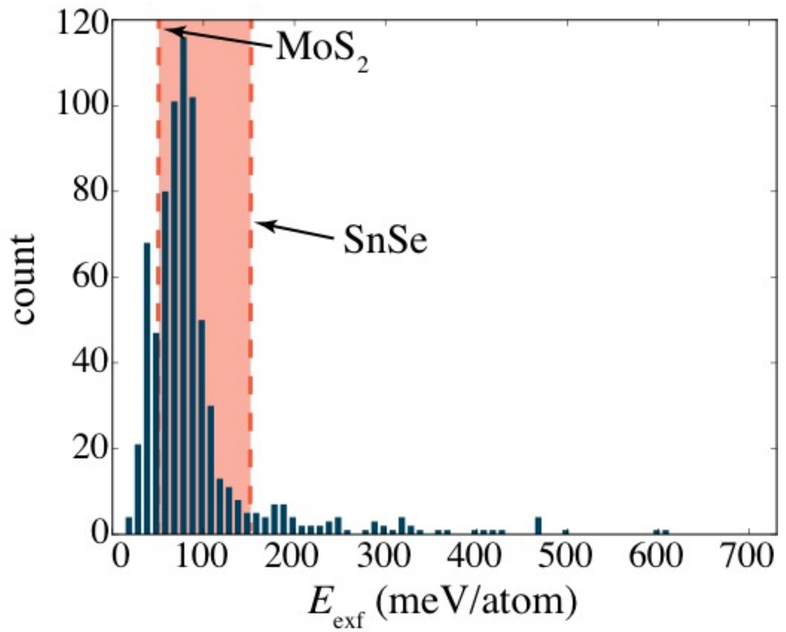
All materials: about 65,000



Layered materials: 1,560 (2%)



Exfoliation energy



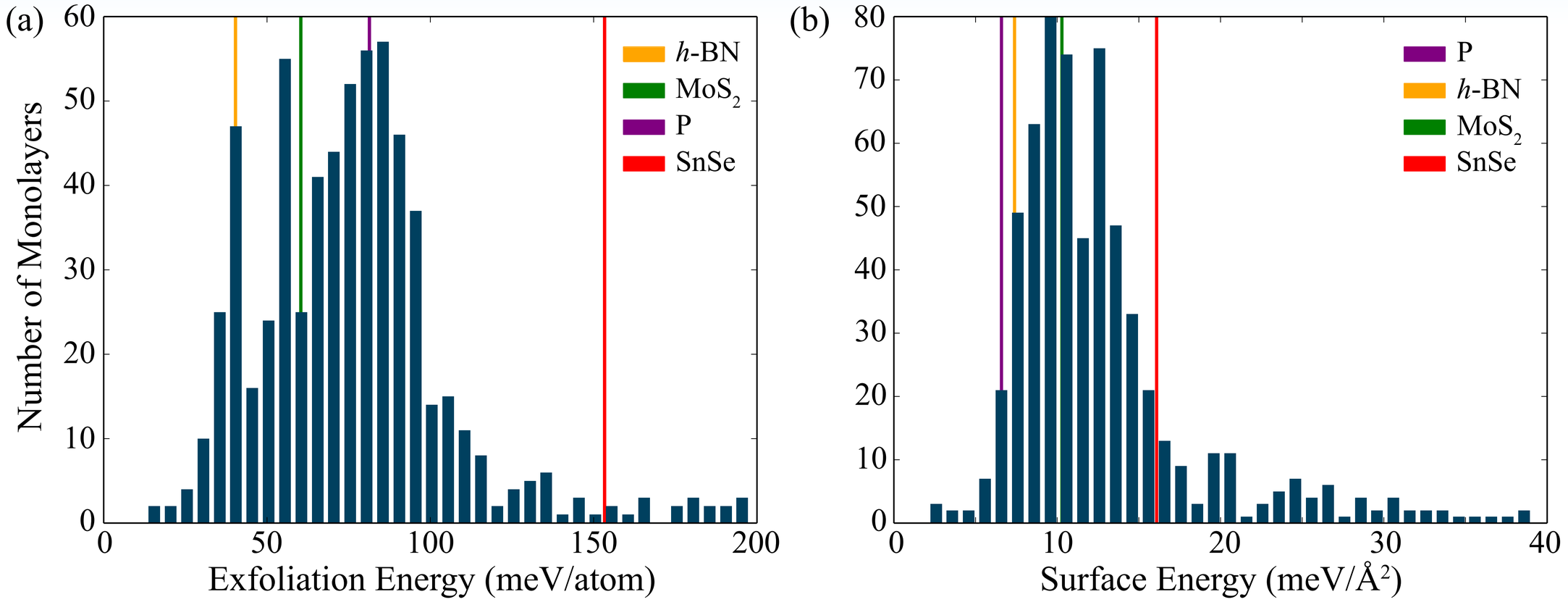
826/1560 have unique monolayers and are nearly stable with E_{hull} < 50 meV/atom.
Identified 625 2D materials with energy below 150 meV/atom
Binary and ternary compounds more likely to be layered.

Powered by
MPInterfaces & materialsweb

Ashton, Paul, Sinnott & Hennig. *Phys. Rev. Lett.* 2017

Stability Criteria for 2D Materials

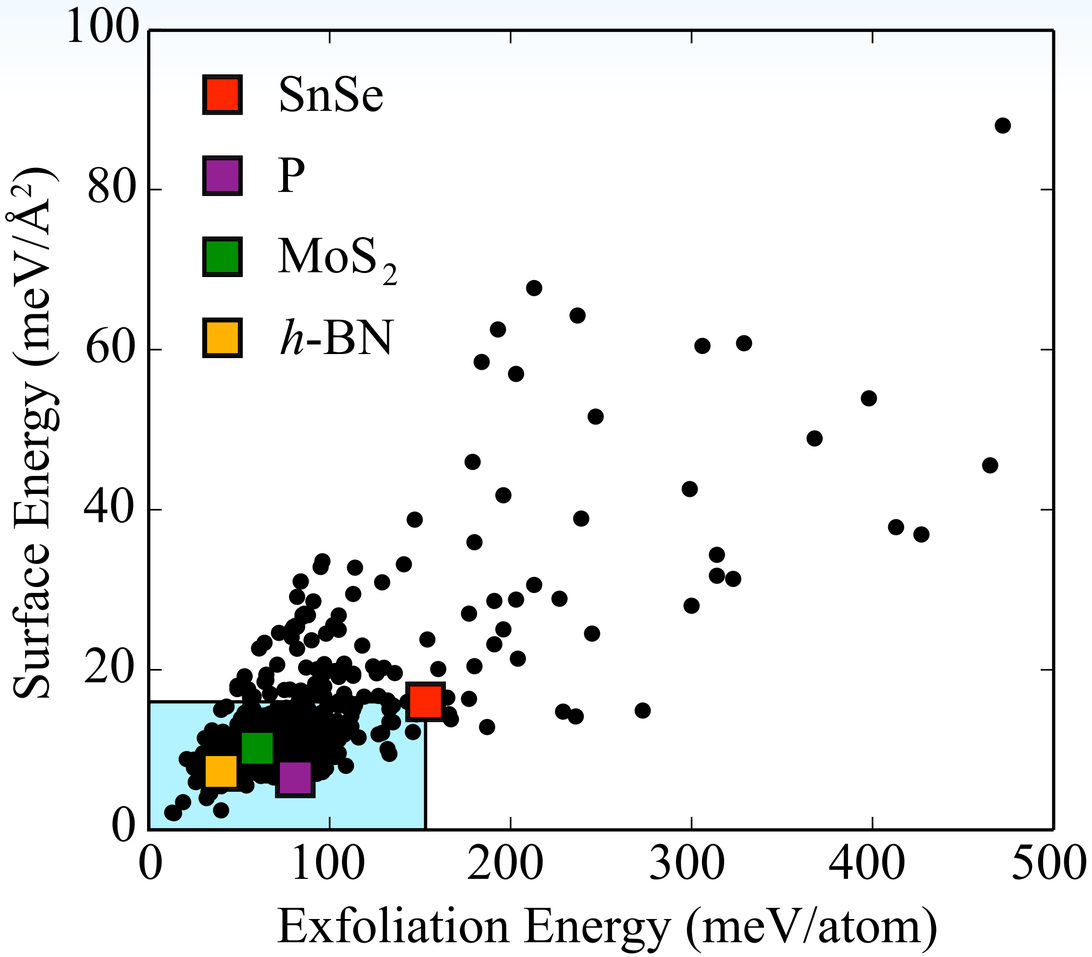
Exfoliation Energy vs. Surface Energy



Ashton, Paul, Sinnott & Hennig. *Phys. Rev. Lett.* 2017

Stability Criteria for 2D Materials

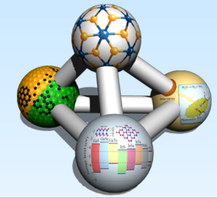
Exfoliation Energy vs. Surface Energy



Alternative empirical measures of stability with similar results.

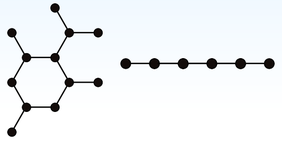
Ashton, Paul, Sinnott & Hennig. *Phys. Rev. Lett.* 2017

Structure Prototypes

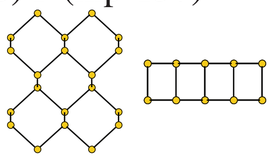


Elementary 2D Materials

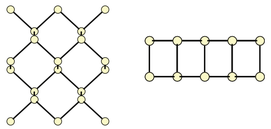
a) C (mp-48)



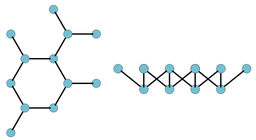
b) P (mp-157)



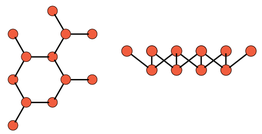
c) As (mp-158)



d) Sb (mp-104)

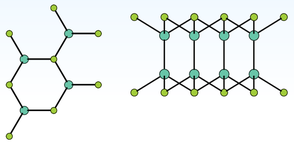


e) Bi (mp-23152)

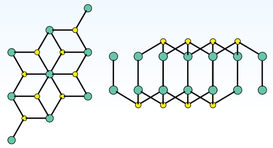


Binary A-B 2D Materials

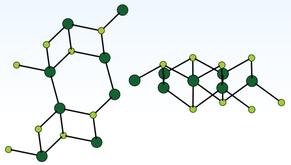
a) GaSe (mp-568263)



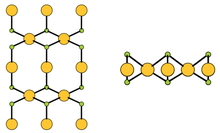
b) GaS (mp-9889)



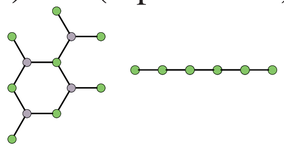
c) InSe (mp-21405)



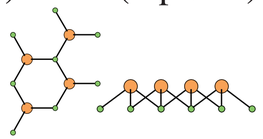
d) AuSe (mp-2793)



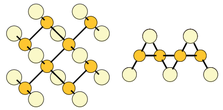
e) BN (mp-604884)



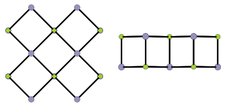
f) GeTe (mp-938)



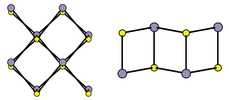
g) AuBr (mp-505366)



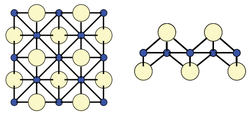
h) SnSe (mp-8936)



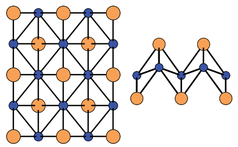
i) SnS (mp-2231)



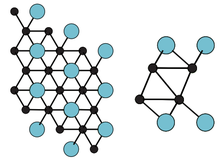
j) CuBr (mp-22917)



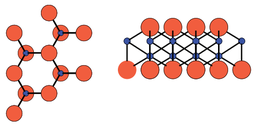
k) CuTe (mp-20826)



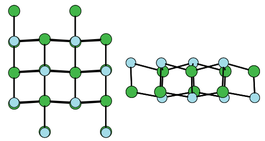
l) ZrCl (mp-27440)



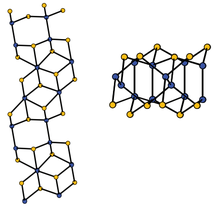
m) CuI (mp-570136)



n) TlF (mp-720)



o) SiP (mp-2798)

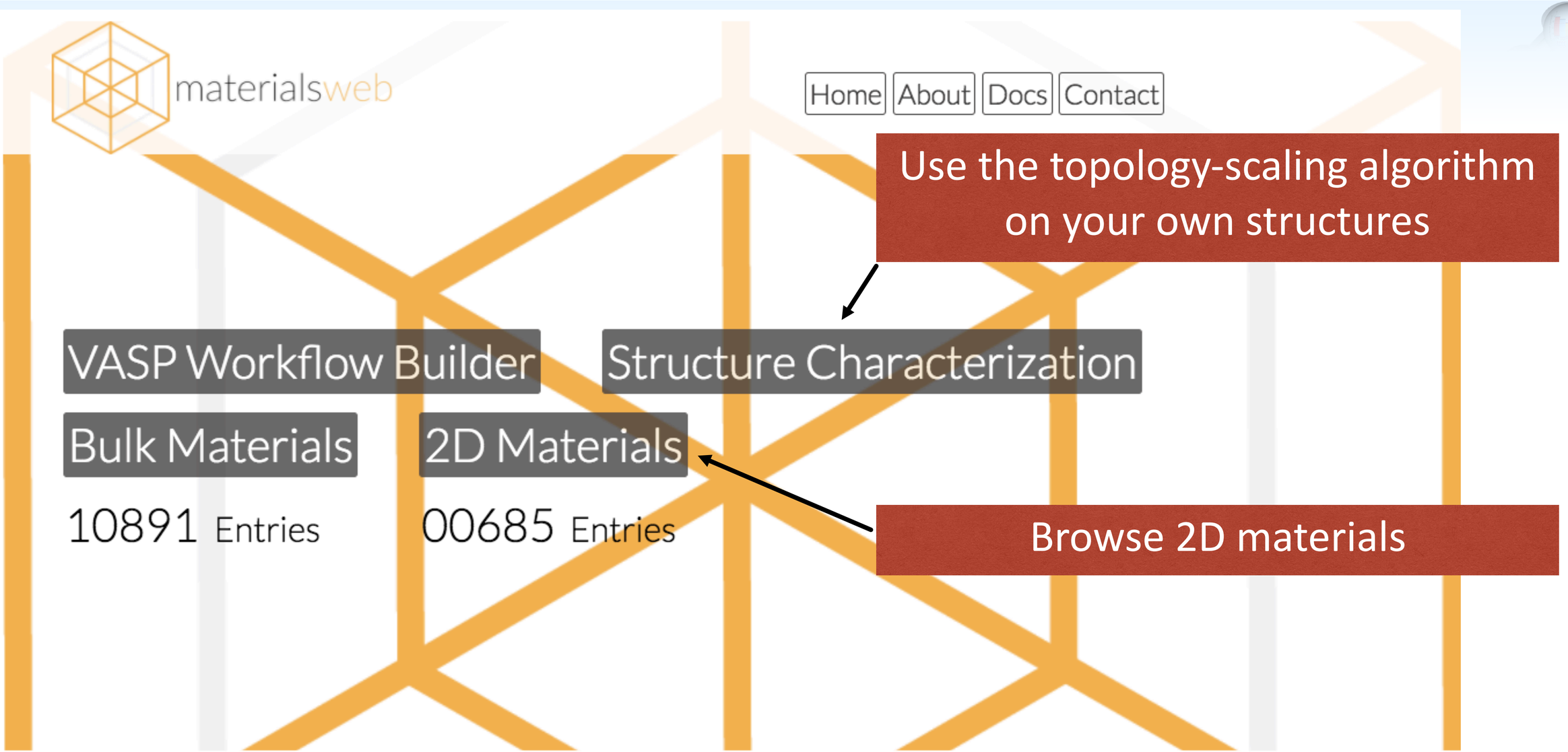


...

Powered by
MPInterfaces & materialsweb

Ashton, Paul, Sinnott & Hennig. *Phys. Rev. Lett.* 2017

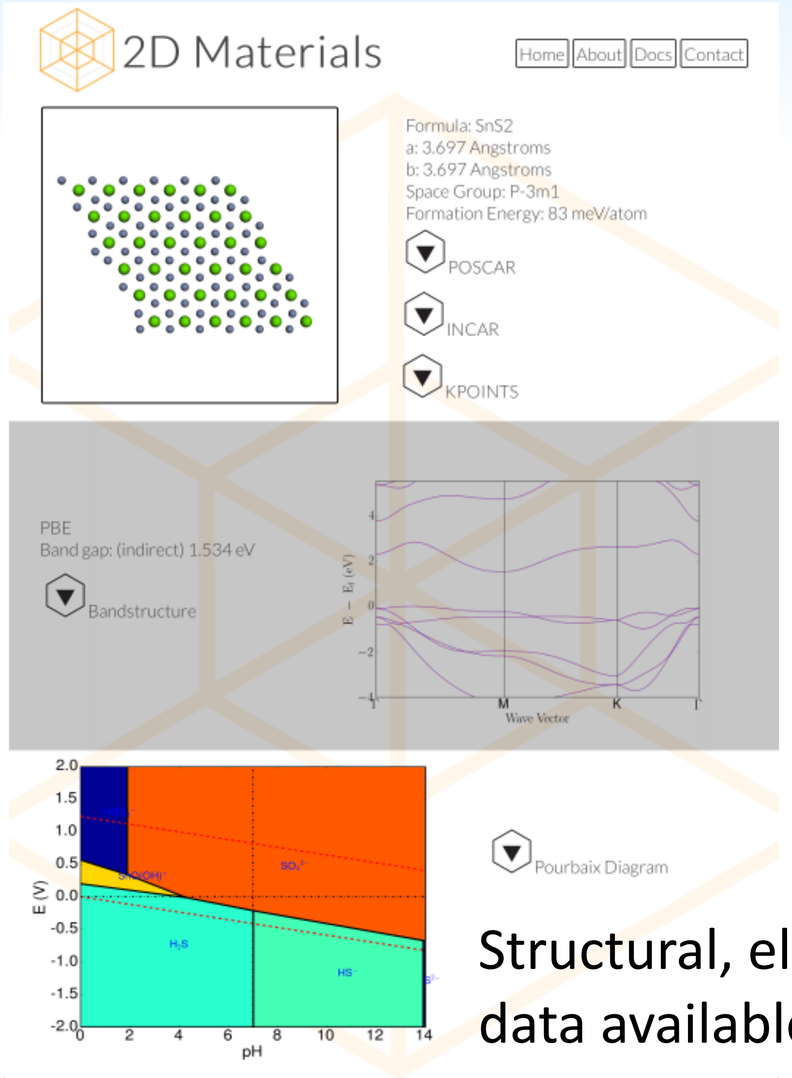
Online Database: https://materialsweb.org

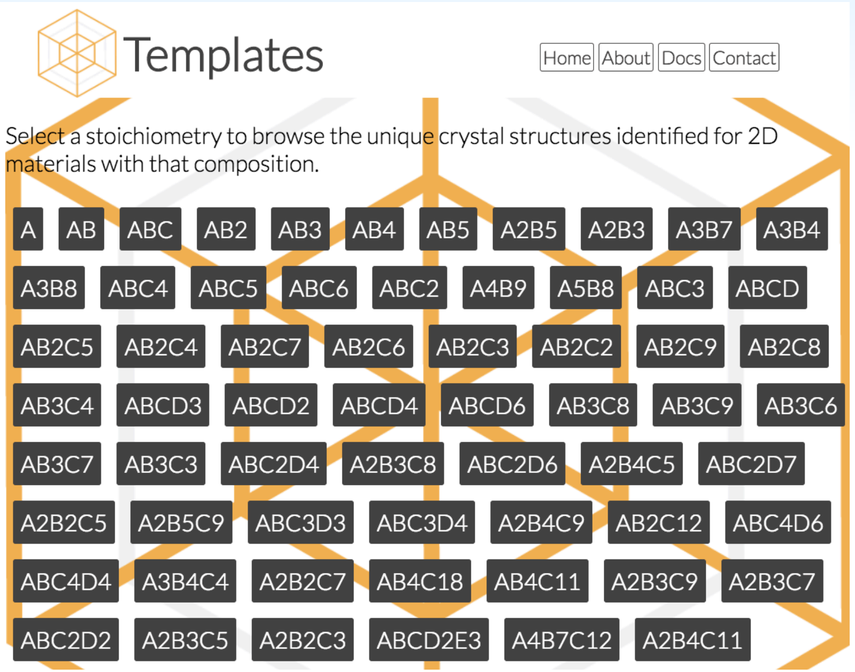


Powered by
MPInterfaces & materialsweb

Ashton, Paul, Sinnott & Hennig. *Phys. Rev. Lett.* 2017

Online Database: https://materialsweb.org





Unique crystal structures for each 2D material stoichiometry in the DB

Structural, electronic and thermodynamic data available for all materials

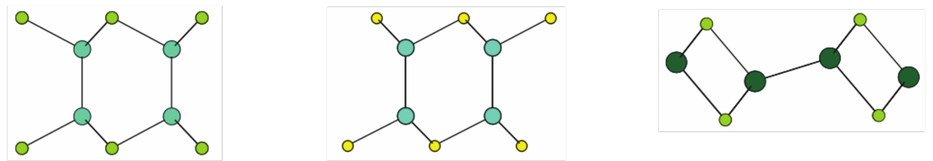
Powered by
MPInterfaces & materialsweb

Ashton, Paul, Sinnott & Hennig. *Phys. Rev. Lett.* 2017

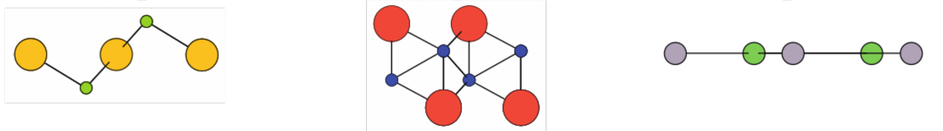
Using Crystal Structure Templates

Unique AB crystal structures

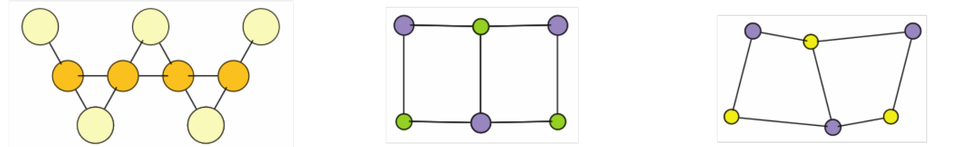
a) GaSe (mp-568263) b) GaS (mp-9889)* c) InSe (mp-21405)*



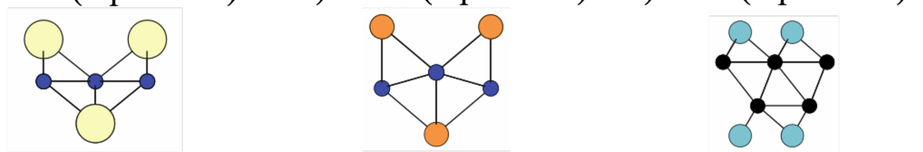
d) AuSe (mp-2793) e) CuI (mp-570136) f) BN (mp-604884)



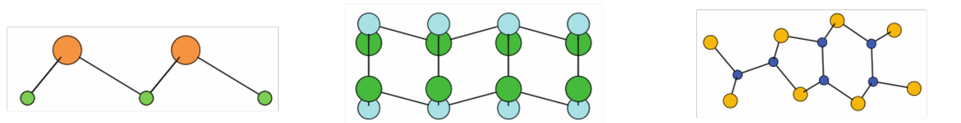
g) AuBr (mp-505366) h) SnSe (mp-8936) i) SnS (mp-2231)



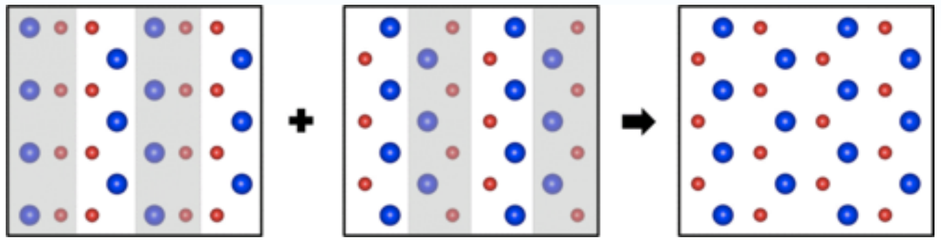
j) CuBr (mp-22917) k) CuTe (mp-20826) l) ZrCl (mp-27440)



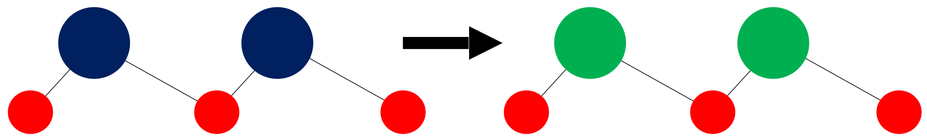
m) GeTe (mp-938)† n) TlF (mp-720) o) SiP (mp-2798)



Genetic algorithms



Element substitution

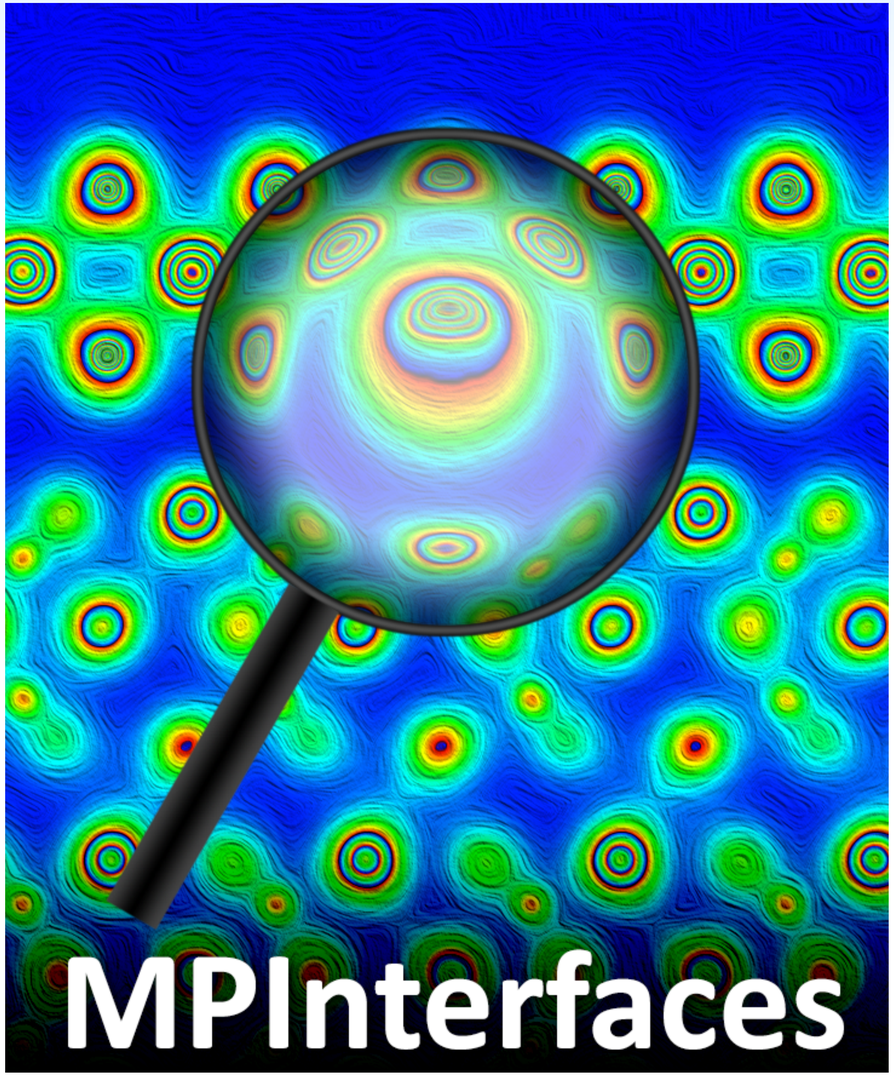


Powered by
MPInterfaces & materialsweb

Ashton, Paul, Sinnott & Hennig. *Phys. Rev. Lett.* 2017

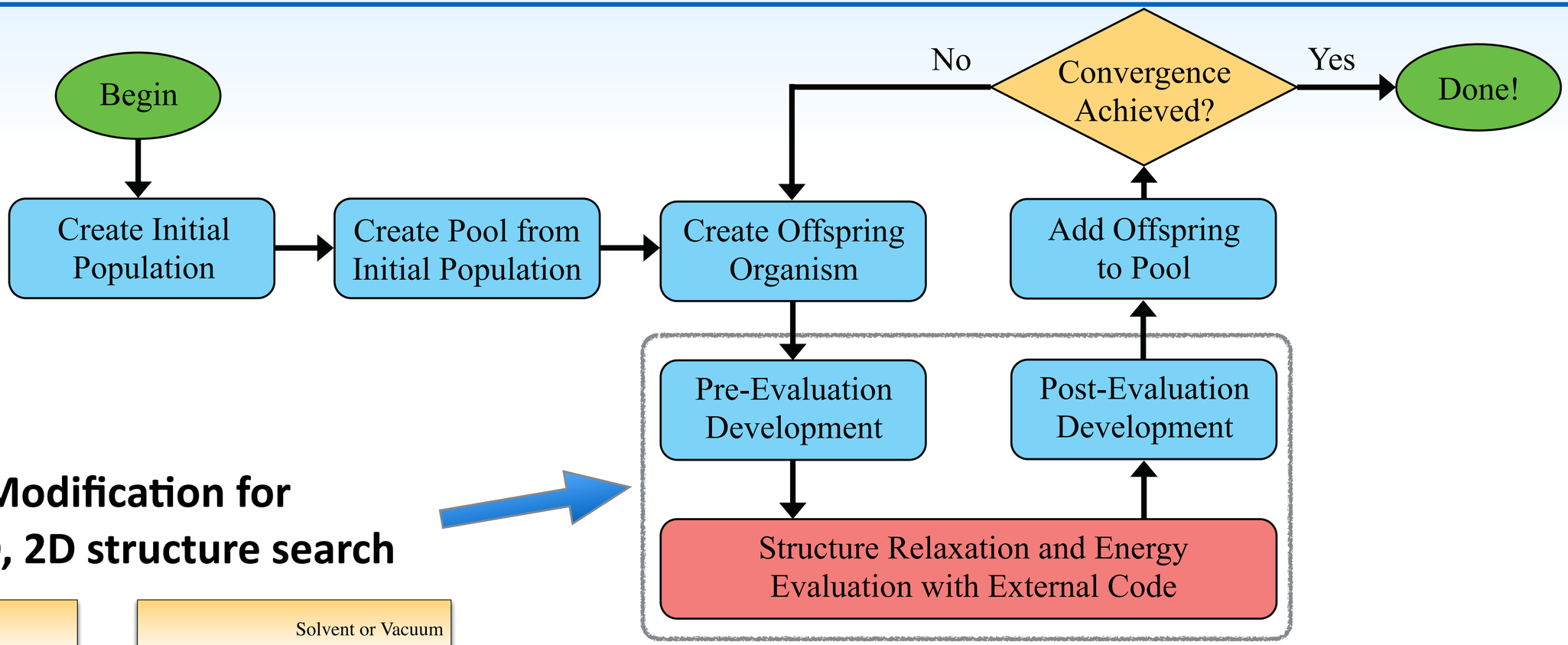
Structure and Stability of 2D Materials

- Classification of 2D materials
- Criteria for stability $\Delta E_f < 200$ meV/atom
- **Methods for 2D materials discovery**
  - Datamining for exfoliation
  - Chemical substitutions and etching
  - **Evolutionary algorithm searches**

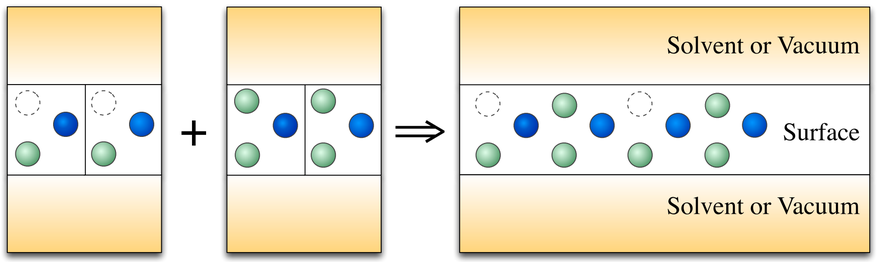


rhennig@ufl.edu
http://hennig.mse.ufl.edu

Genetic Algorithm Search for 2D Materials



Modification for
0D, 1D, 2D structure search



Grand canonical genetic algorithm
Variable number of atoms and composition

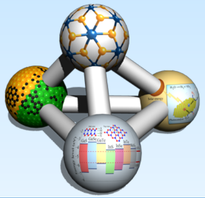
https://github.com/henniggroup/gasp-python

W. W. Tipton, RGH, J. Phys.: Cond. Matter 25, 495401 (2013)

B. C. Revard, W. W. Tipton, A. Yesypenko, R. G. Hennig, PRB 93, 054117 (2016)

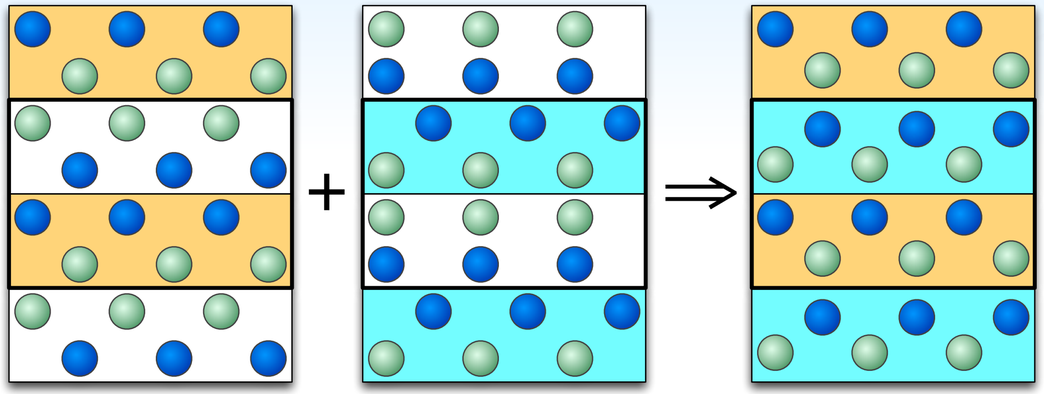
Powered by
MPInterfaces & materialsweb

Genetic Operations



Mating:

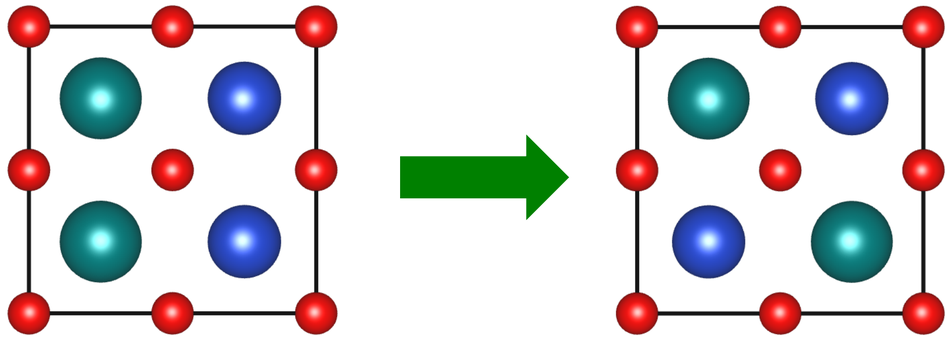
- combines two structures
- preserves local structural motifs
- handles different lattices



Mutation:

1. Structural Mutation
   - lattice strain
   - atomic coordinate displacement
2. Permutation
   - swap atomic locations
3. Number of atoms mutation
   - add or remove atoms

$$\vec{a}' = (\mathbb{I} + \epsilon_{ij})\vec{a}$$



https://github.com/henniggroup/gasp-python

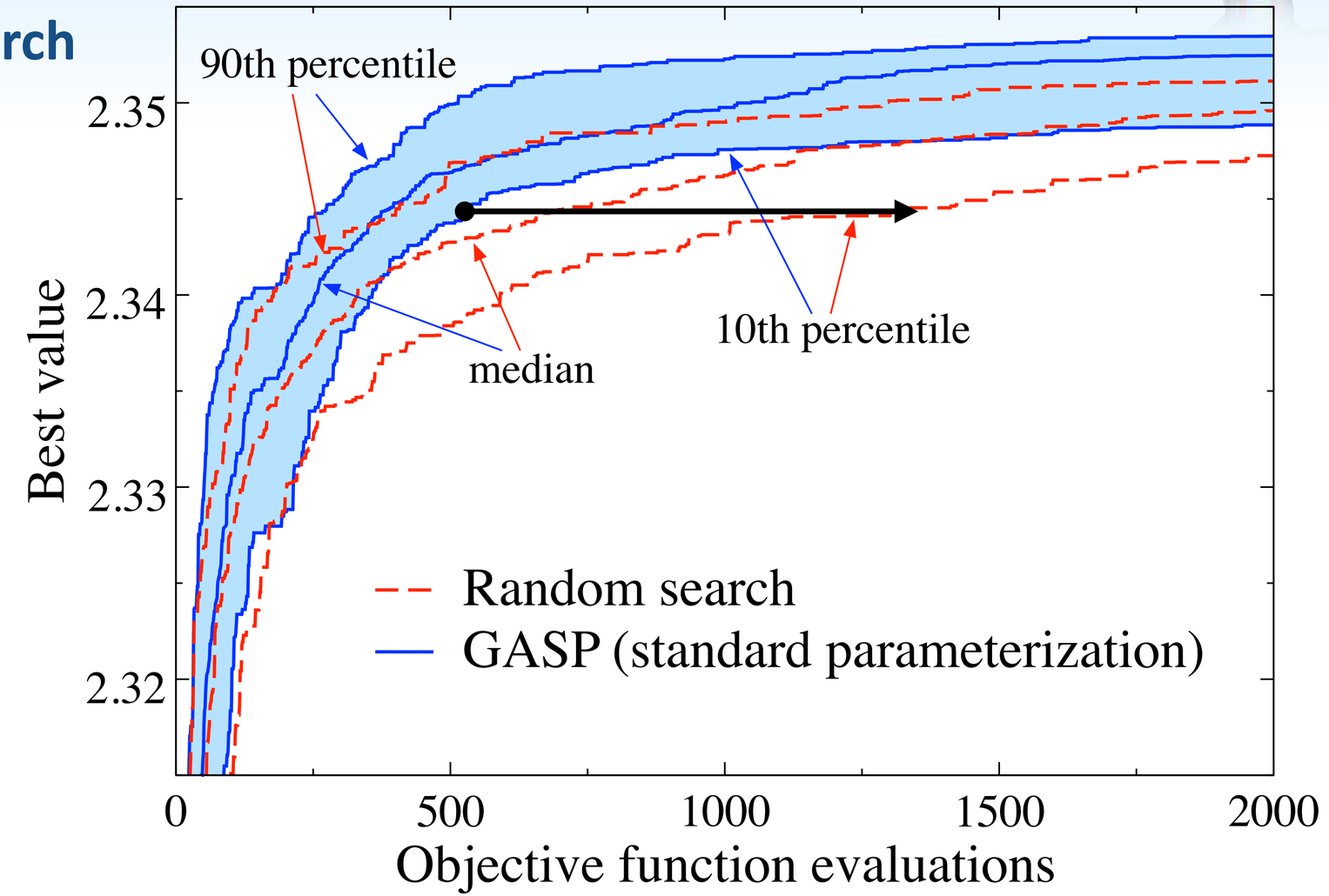
W. W. Tipton, RGH, J. Phys.: Cond. Matter 25, 495401 (2013)

B. C. Revard, W. W. Tipton, A. Yesypenko, R. G. Hennig, PRB 93, 054117 (2016)

Efficiency of Genetic Algorithm

Efficiency compared to random search

- Random search requires 2-3x more structure relaxations
- Genetic algorithm learns from previous structures

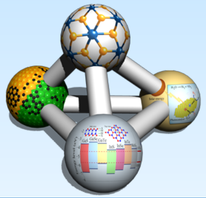


https://github.com/henniggroup/gasp-python

W. W. Tipton, RGH, J. Phys.: Cond. Matter 25, 495401 (2013)

B. C. Revard, W. W. Tipton, A. Yesypenko, R. G. Hennig, PRB 93, 054117 (2016)

Phase Diagram Searching

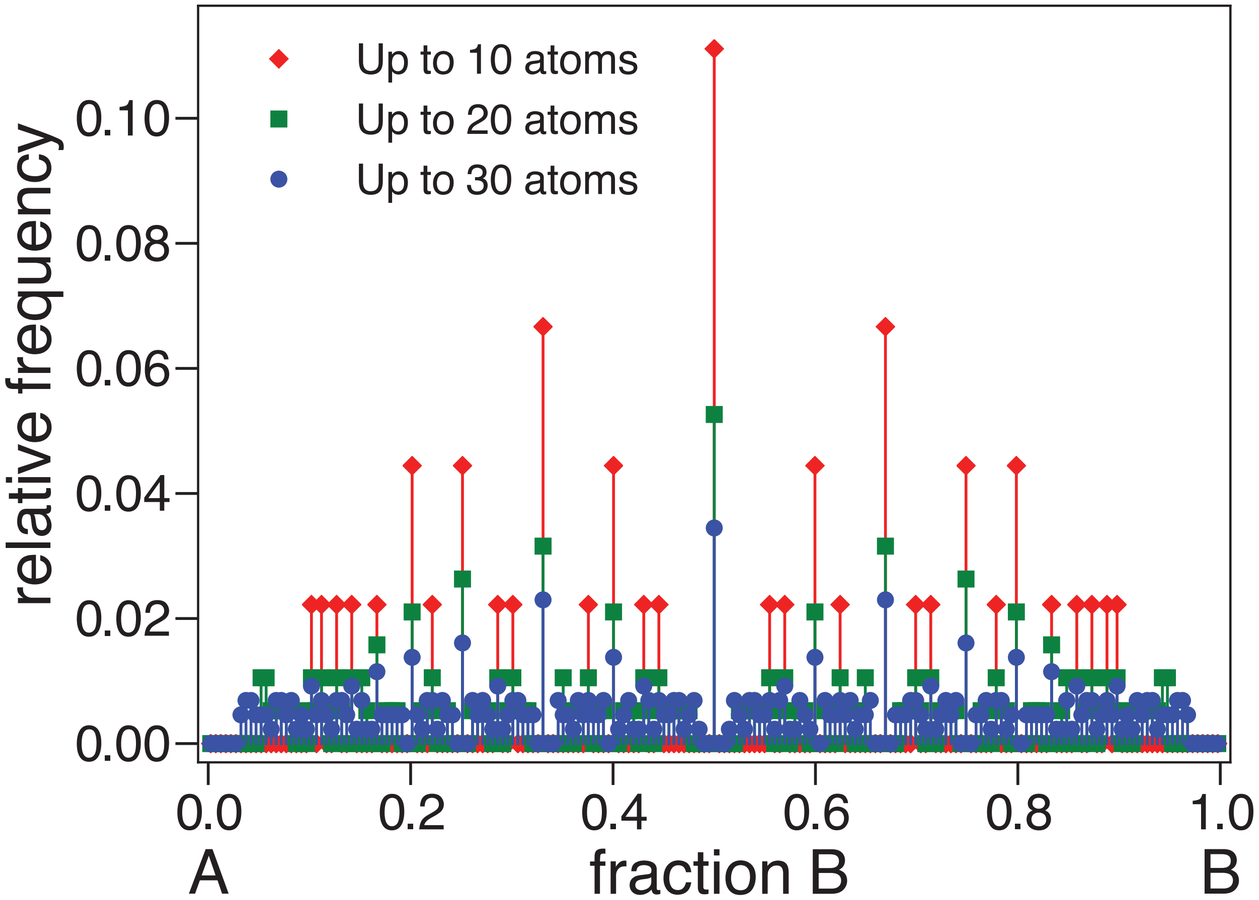


Problem:

- Naïve algorithm oversamples average compositions

Solutions

1. Use larger endpoint structures
   - Works but expensive
2. Preferentially select parents with similar compositions
   - Needs metric for distance of structures in composition space



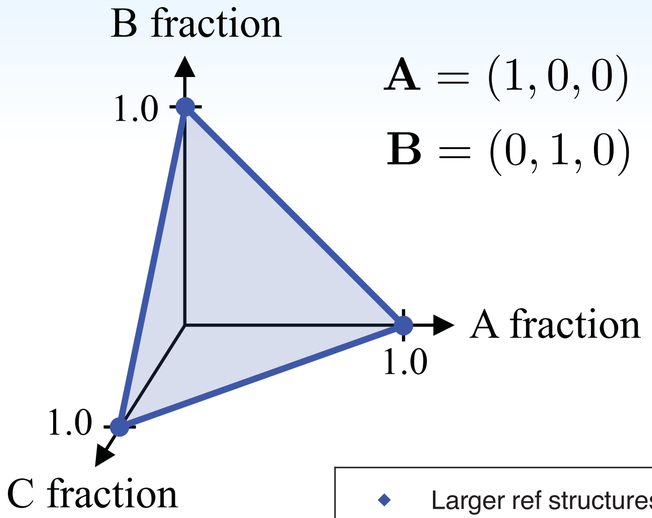
Phase Diagram Searching

Metric for distance in composition space

- Express composition as a vector
- Use L_1 Norm to define distance:

$$d_{XY} = \frac{1}{2}\left|\left|\mathbf{X} - \mathbf{Y}\right|\right|_1$$

$$\left|\left|\mathbf{A} - \mathbf{B}\right|\right|_1 = \left|\left|(1, -1, 0)\right|\right|_1 = |1| + |-1| + |0| = 2$$

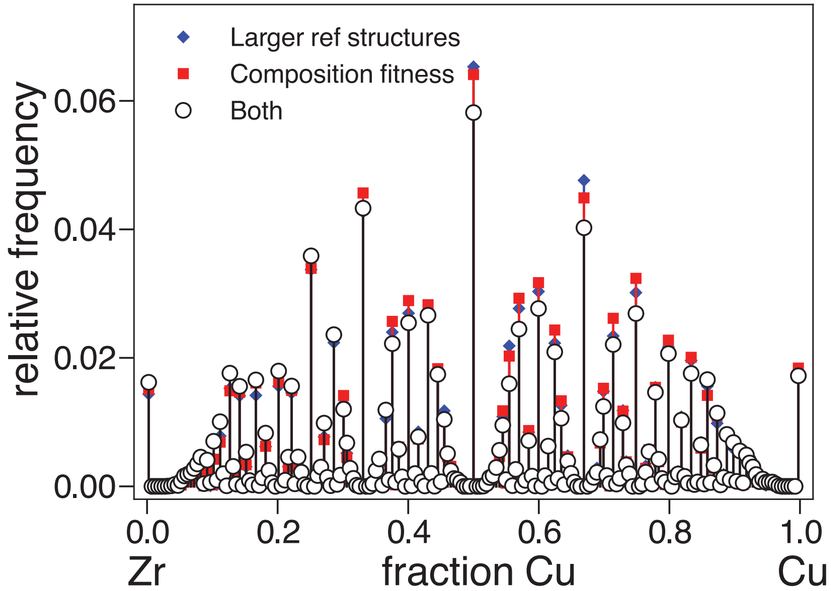


Fitness for structures

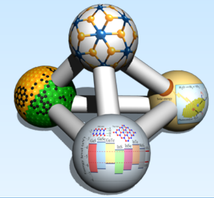
- Composition fitness $f_{comp} = 1 - d$
- Relative fitness $f_{rel} = w_{comp} f_{comp} + (1 - w_{comp}) f_{reg}$

Fitness for structures

- Sampling distribution improved but not uniform
- Use partial phase diagram searches if needed

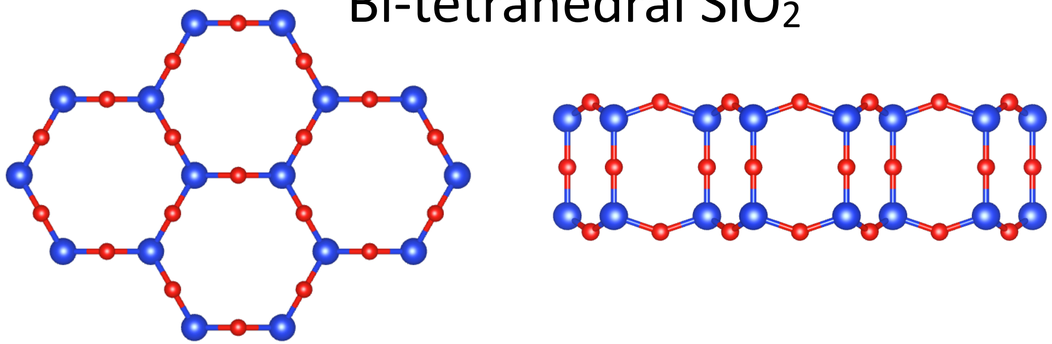


Genetic Algorithm – 2D Group-IV Dioxides



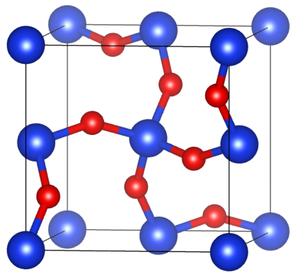
- SiO_2 recently synthesized in 2D form

Bi-tetrahedral SiO_2

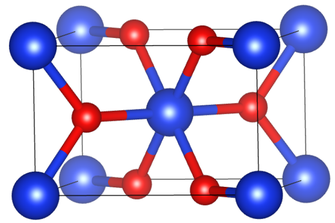


- Bulk structures not layered

α-Quartz (SiO_2)

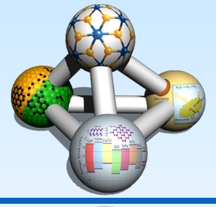


Rutile (GeO_2, SnO_2, PbO_2)



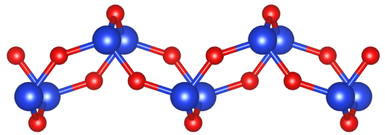
- Cations are
  - 4-fold coordinated in α-quartz structure
  - 6-fold coordinated in rutile structure

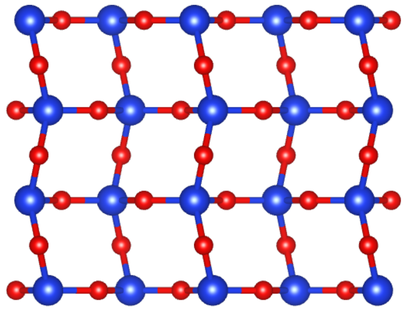
Genetic Algorithm – 2D Group-IV Dioxides



- For 2D SiO_2, the algorithm predicted an orthorhombic structure

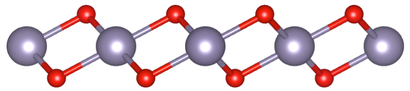
Orthorhombic SiO_2

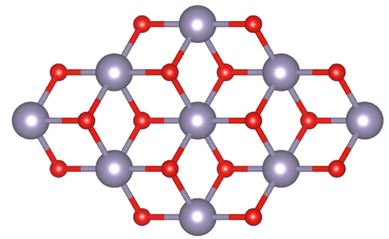




- For GeO_2, the algorithm found the 1*T* structure

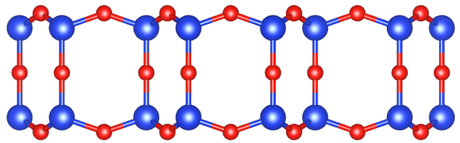
1*T* GeO_2

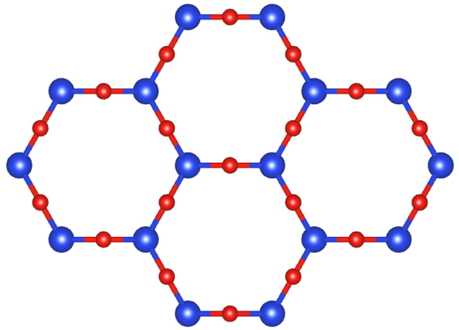




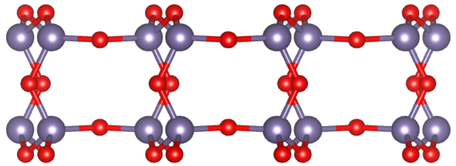
- Seeded both searches with the bi-tetrahedral structure
- Found lower energy monoclinic structure of 2D GeO_2

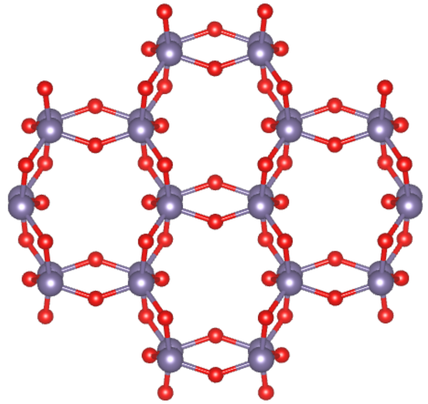
Bi-tetrahedral SiO_2





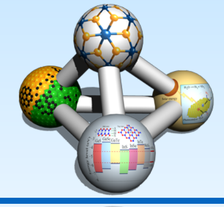
Monoclinic GeO_2



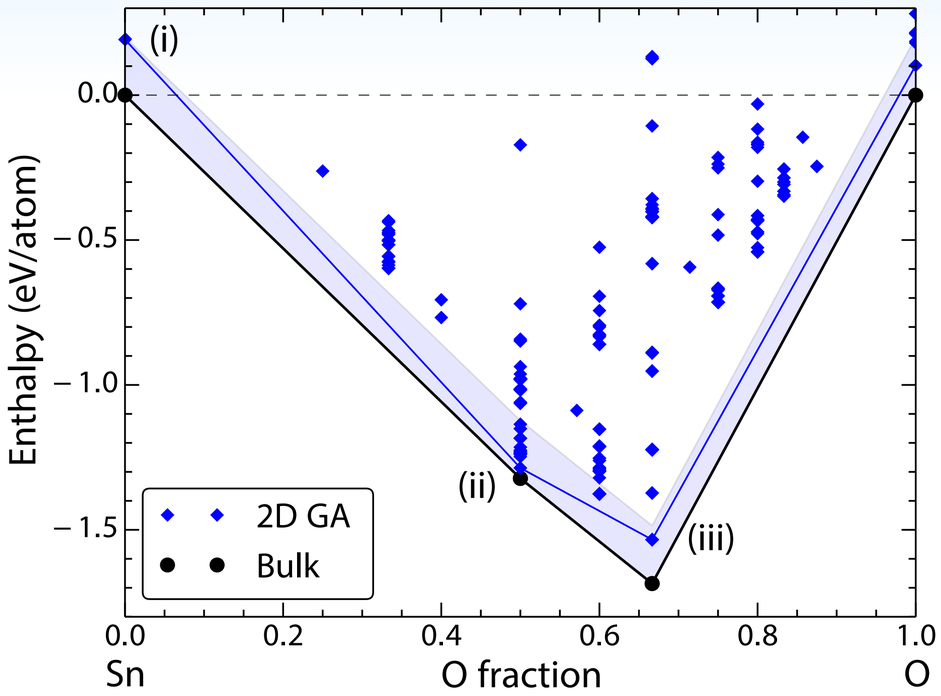


Seeding search with known structures improves performance

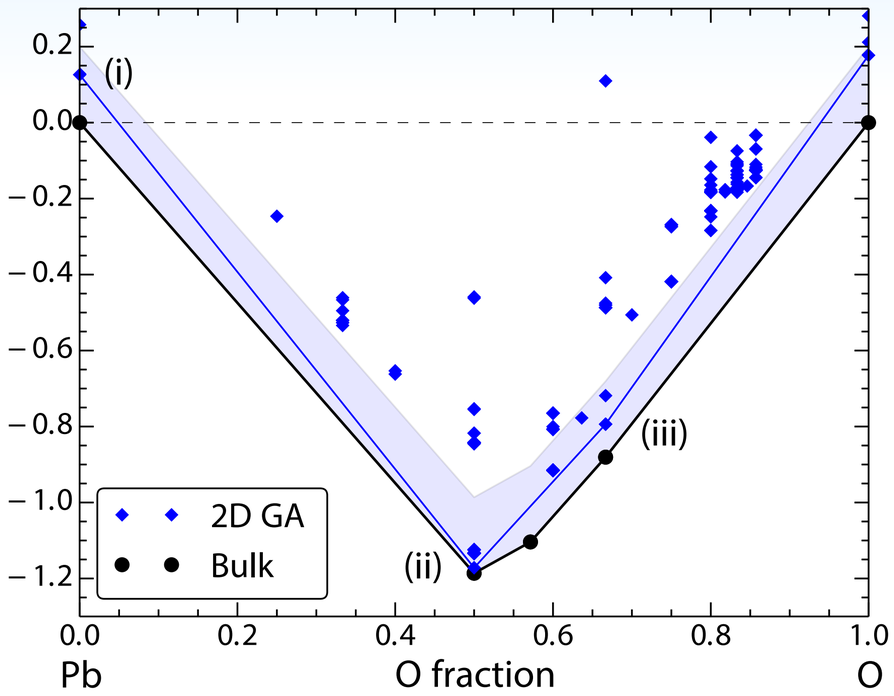
Genetic Algorithm – 2D Tin and Lead Oxides

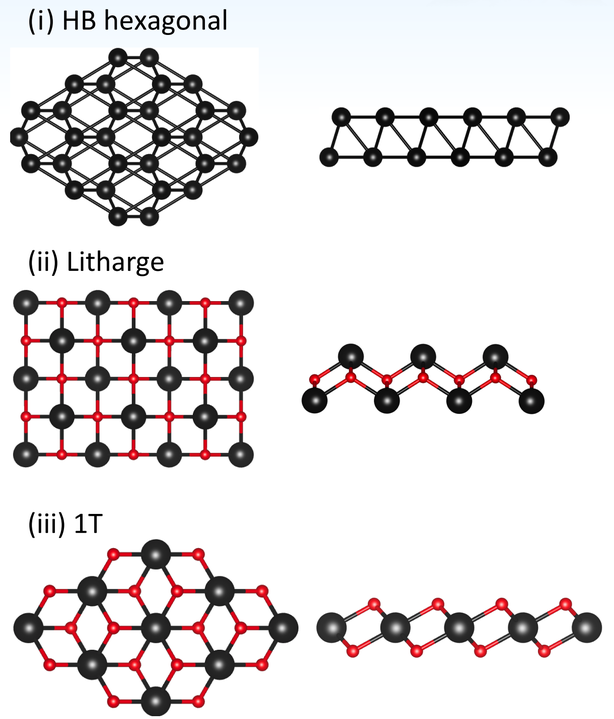


2D Sn-O Phases



2D Pb-O Phases





Confirmation of our previous prediction of litharge structure for 2D SnO and PbO

New prediction of 1*T* structure for 2D SnO_2 and PbO_2

https://github.com/henniggroup/gasp

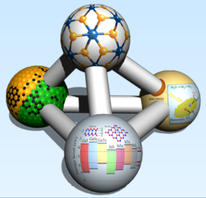
W. W. Tipton, RGH, J. Phys.: Cond. Matter 25, 495401 (2013)

B. C. Revard, W. W. Tipton, RGH, Topics in Current Chemistry (2014)

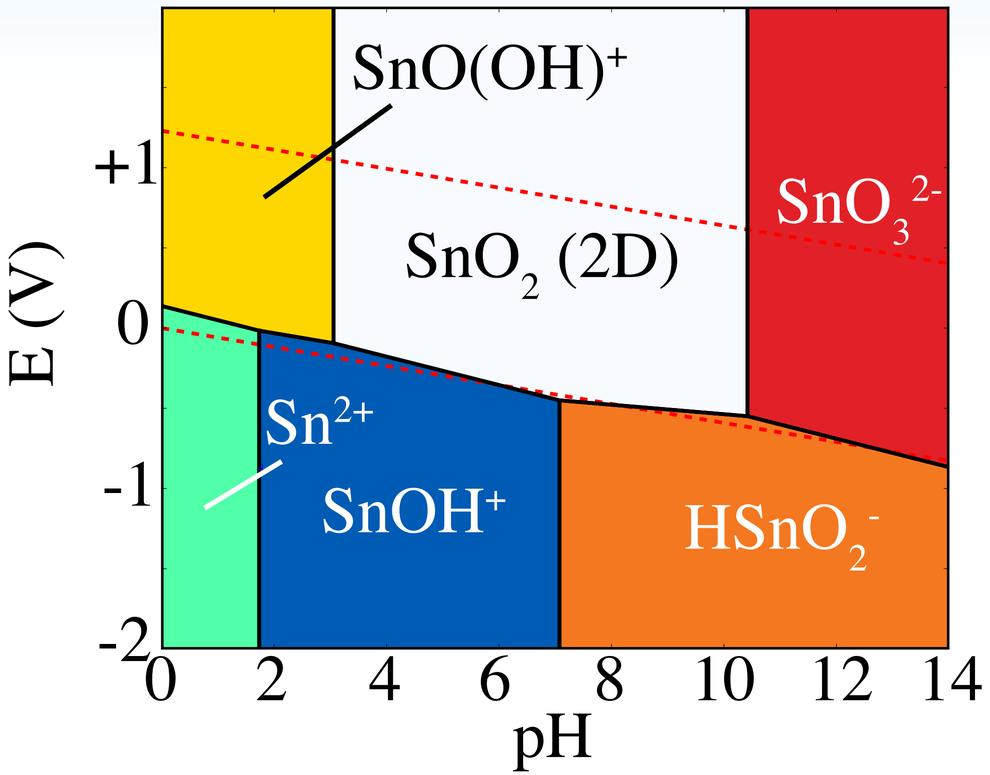
B. C. Revard, W. W. Tipton, A. Yesypenko, R. G. Hennig, PRB 93, 054117 (2016)

Powered by
MPInterfaces & materialsweb

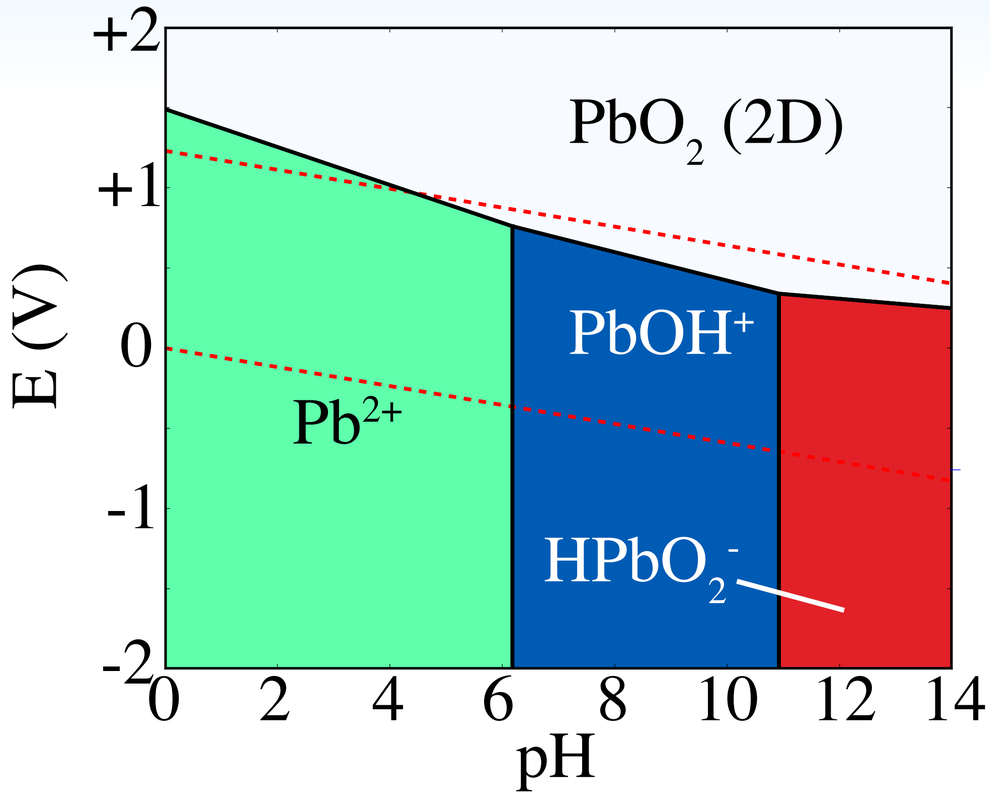
Genetic Algorithm – 2D Tin and Lead Oxides



2D Sn-O Phases



2D Pb-O Phases



The 2D SnO_2 and PbO_2 1T phases are stable in water for ionic concentrations of 10^{-6}!

https://github.com/henniggroup/gasp

W. W. Tipton, RGH, J. Phys.: Cond. Matter 25, 495401 (2013)

B. C. Revard, W. W. Tipton, RGH, Topics in Current Chemistry (2014)

B. C. Revard, W. W. Tipton, A. Yesypenko, R. G. Hennig, PRB 93, 054117 (2016)

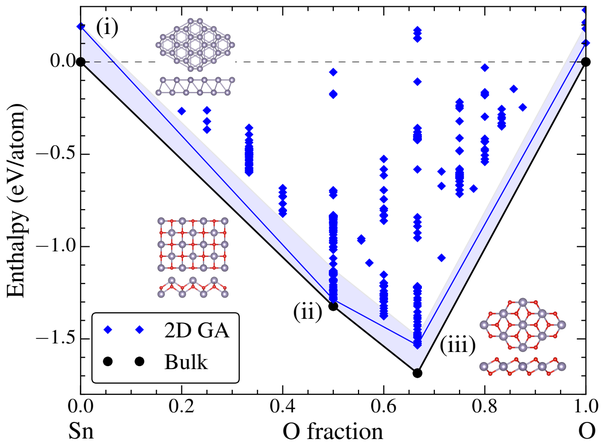
Powered by
MPInterfaces & materialsweb

Search of Energy Landscapes by Evolutionary Algorithms and Data Mining

Richard G. Hennig, University of Florida

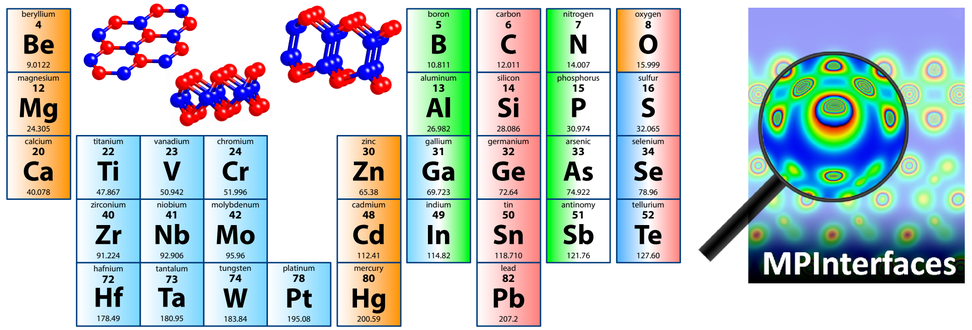
Structure and Energy Landscapes

- Structure representation
- Features of energy landscapes
- Structure searches for 2D materials by datamining, chemical substitution, and evolutionary algorithms

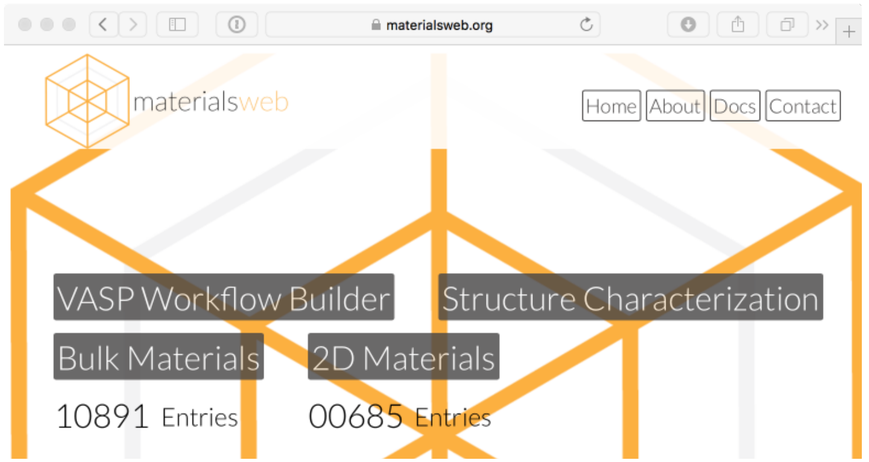


GASP - Genetic algorithm and machine learning for structure predictions

MPInterfaces - High throughput framework for 2D materials



Data available at http://materialsweb.org



Powered by
MPInterfaces & materialsweb

Open source available at https://github.com/henniggroup