Evolution as a tool for predicting structures of crystals and molecules

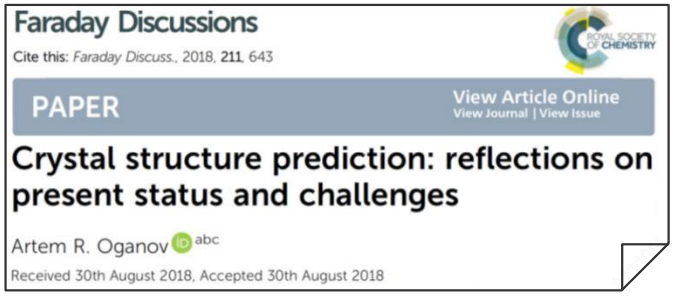
Artem R. Oganov

Skolkovo Institute of Science and Technology, Russia

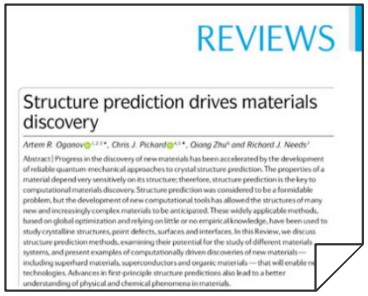
Moscow Institute of Steel and Alloys, Moscow, Russia

Institute of Geochemistry and Analytical Chemistry, Moscow, Russia

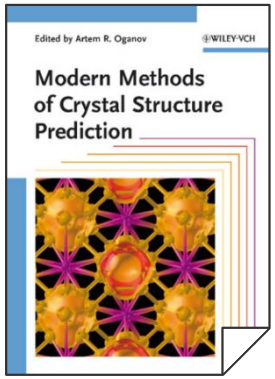
Crystal Structure Prediction



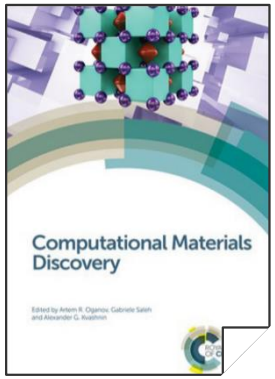
Faraday Discussions (2018)



Nature Reviews Materials (2019)



2011



2018

Until recently, experiment was believed to be the only source of crystal structures

The Nobel Prize in Physics 1914

"for his discovery of the diffraction of X-rays by crystals"

Max von Laue

The Nobel Prize in Physics 1915

"for their services in the analysis of crystal structure by means of X-rays"

Sir William Henry Bragg

William Lawrence Bragg

The Nobel Prize in Chemistry 1985

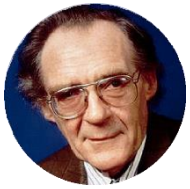
"for their outstanding achievements in the development of direct methods for the determination of crystal structures"

Herbert A. Hauptman

Jerome Karle

(from http://nobelprize.org)

Angelo Gavezzotti



John Maddox

Acc. Chem. Res. **1994,** *27,* 309−314

Are Crystal Structures Predictable?

ANGELO GAVEZZOTTI*

"No": by just writing down this concise statement, in what would be the first one-word paper in the chemical literature, one could safely summarize the present state of affairs

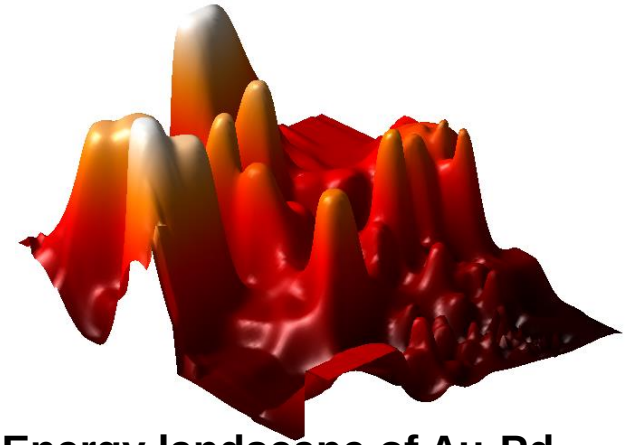
ONE of the continuing scandals in the physical sciences is that it remains in general impossible to predict the structure of even the simplest crystalline solids from a knowledge of their chemical composition. Who, for example, would guess that graphite, not diamond, is the thermodynamically stable allotrope of carbon at ordinary temperature and pressure? Solids such as crystalline water (ice) are still thought to lie beyond mortals' ken.

$$C = \frac{1}{(V/\delta^3)} \frac{(V/\delta^3)!}{[(V/\delta^3) - N]!N!}$$

- Astronomic number of possible structures!

USPEX (Universal Structure Predictor: Evolutionary Xtallography)

- **Combination of evolutionary algorithm and quantum-mechanical calculations.**
- **>8000 users.**
- **Solves “formidable” problem of crystal structure prediction.**



Energy landscape of Au_8Pd_4

uspex-team.org

Samrath Lal Chaplot

RESEARCH NEWS

Crystal structure prediction – evolutionary or revolutionary crystallography?

S. L. Chaplot and K. R. Rao CURRENT SCIENCE, VOL. 91, NO. 11, 10 DECEMBER 2006

- Early article about our work

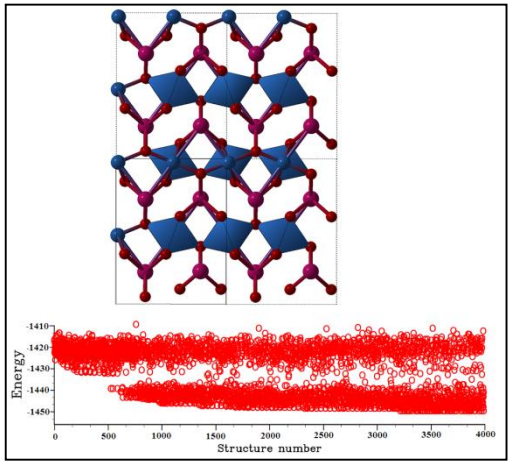
Quantum-mechanical calculations (using density functional theory

$$(-\frac{\nabla^2}{2}+v_{e\text{-}n}[\rho(\mathrm{r})]+v_H[\rho(\mathrm{r})]+v_{xc}[\rho(\mathrm{r})])\phi_i(\mathrm{r})=\varepsilon_i\phi_i(\mathrm{r})$$

W. Kohn
Nobel Prize in Chemistry 1998

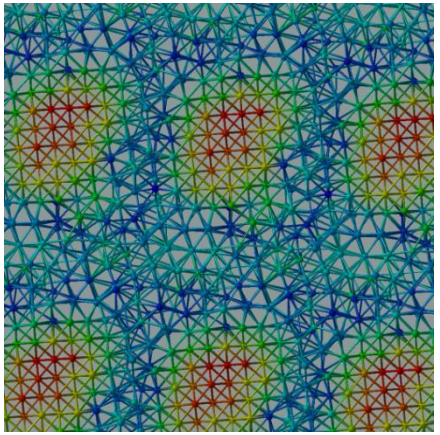
E. Schroedinger
Nobel Prize in Physics 1933

Evolutionary approach is extremely powerful in computational materials science



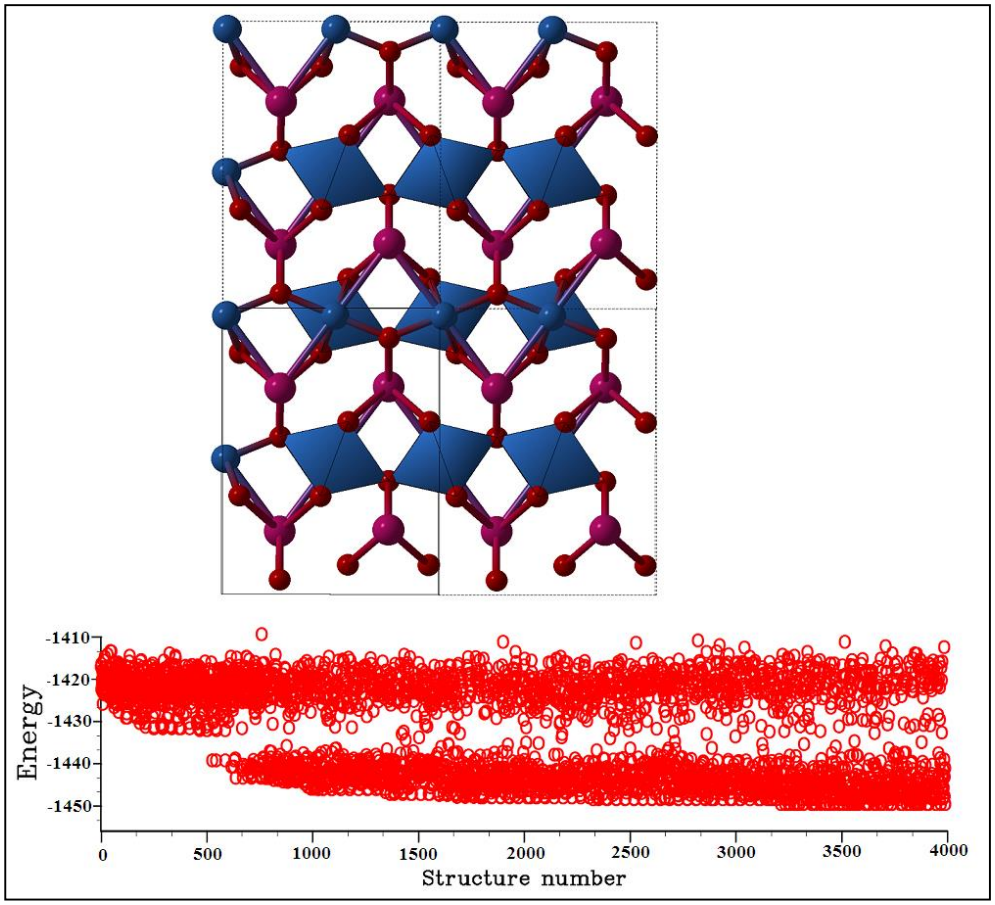
1.Predicting structures by evolution

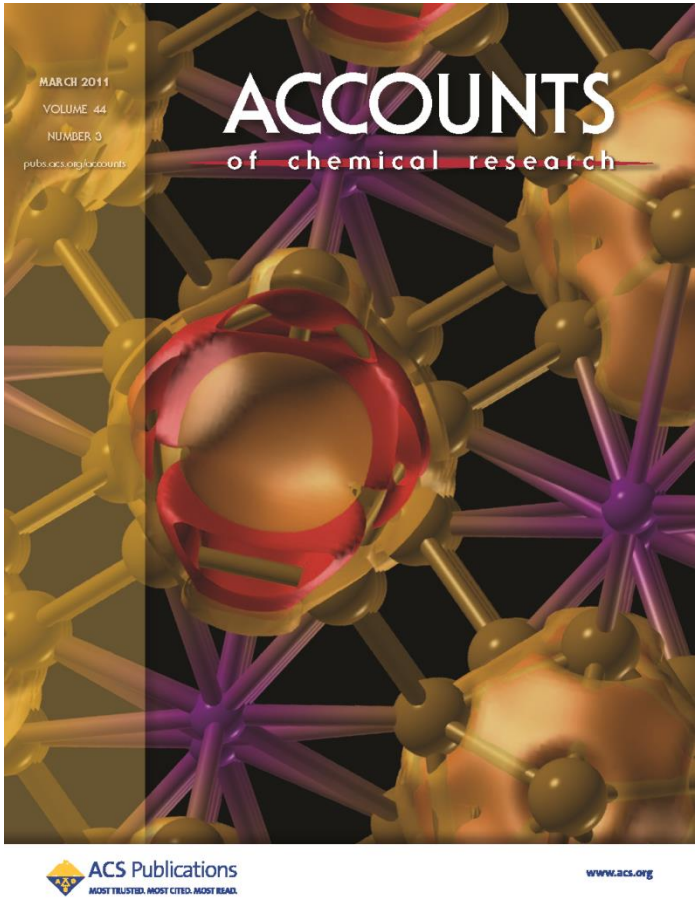
2. When there's more than one solution



3. Looking into the genome

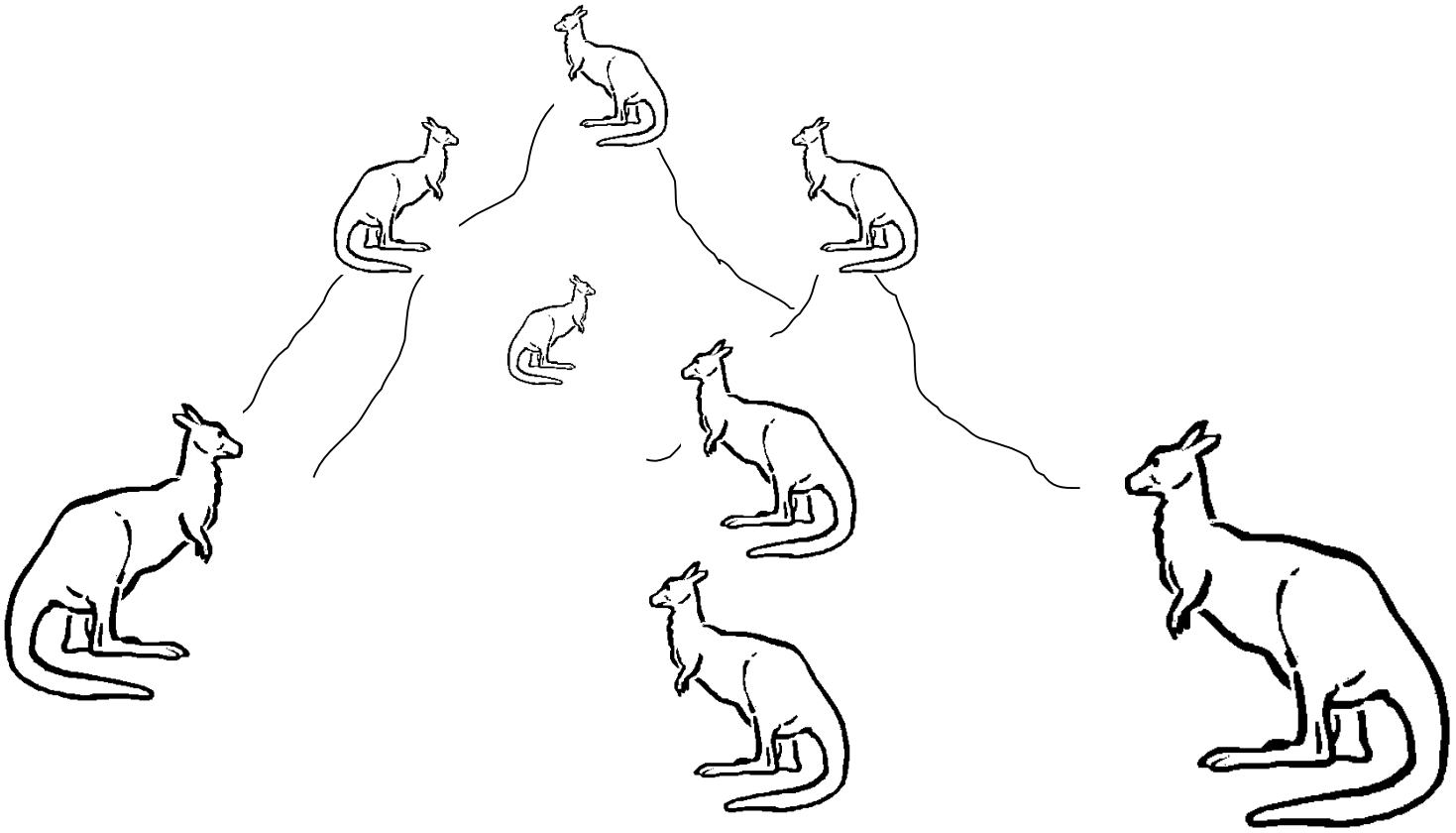
1. Predicting structures by evolution





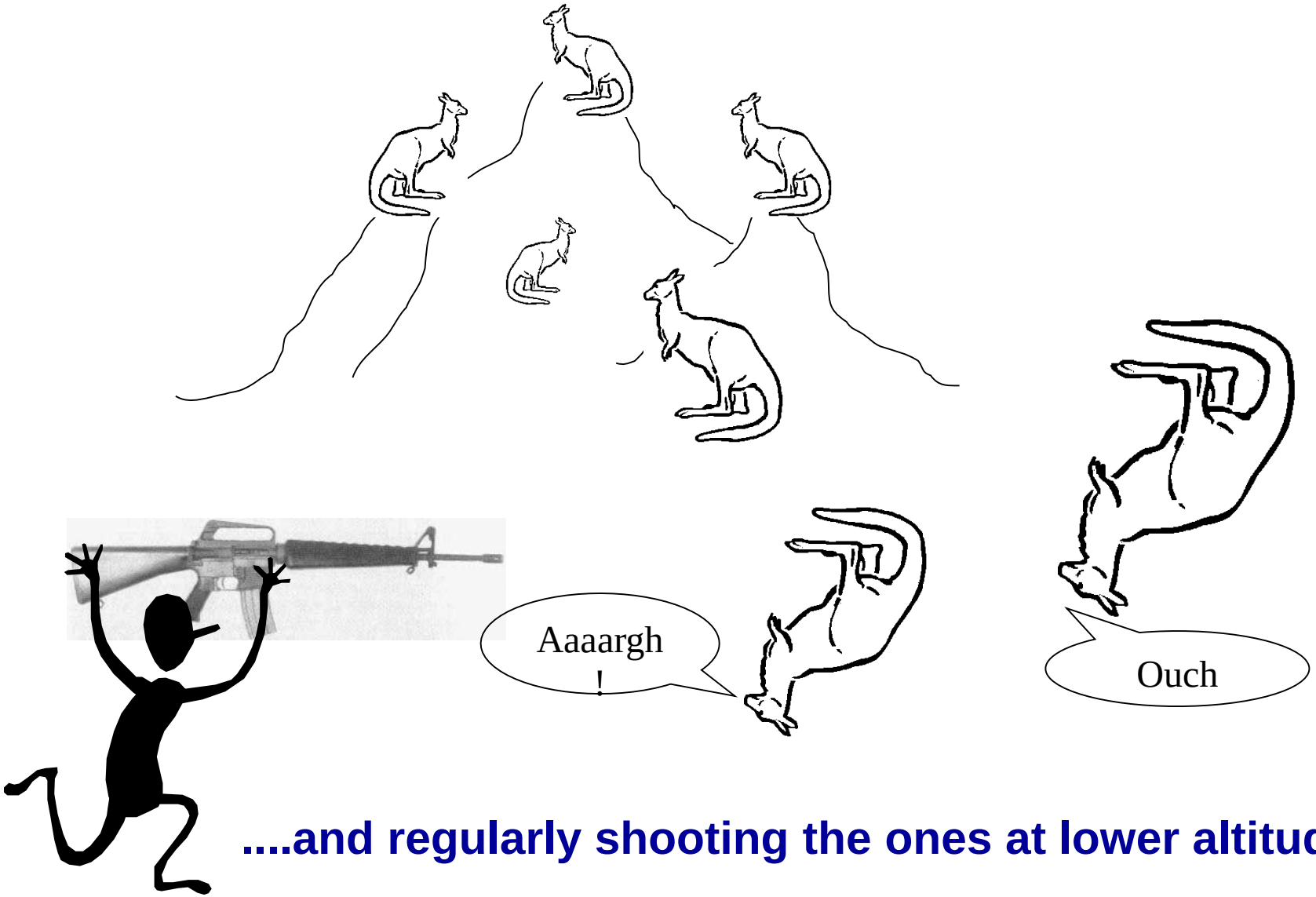
Oganov A.R., Lyakhov A.O., Valle M. (2011).
How evolutionary crystal structure prediction works - and why.
***Acc. Chem. Res.* 44, 227-237.**

Global optimisation methods: Kangaroo's climb to Mt. Everest



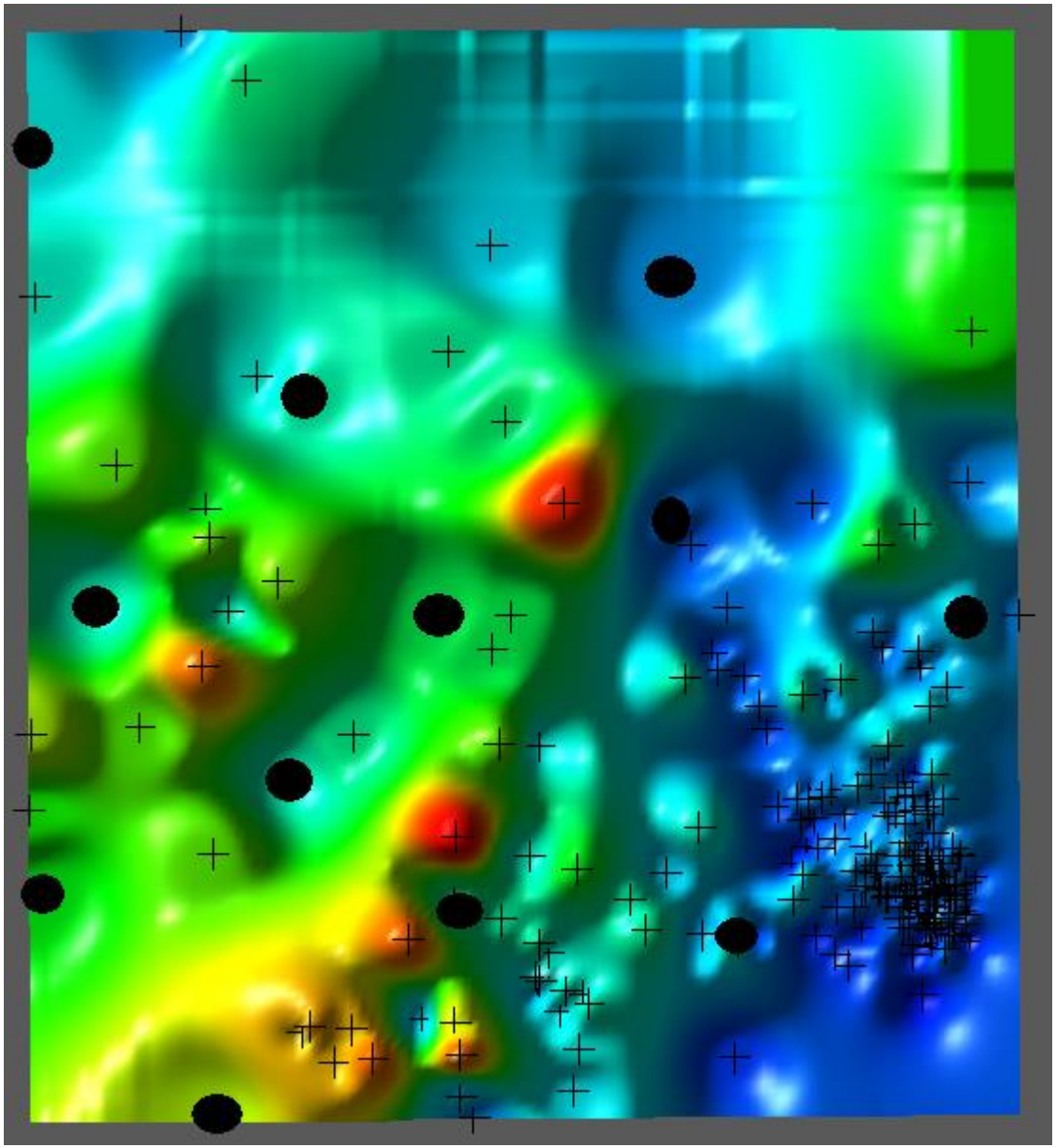
Evolutionary Algorithms are like taking a whole plane load of kangaroo's and letting them reproduce freely (not pictured).....

Global optimisation methods: Kangaroo's climb to Mt. Everest

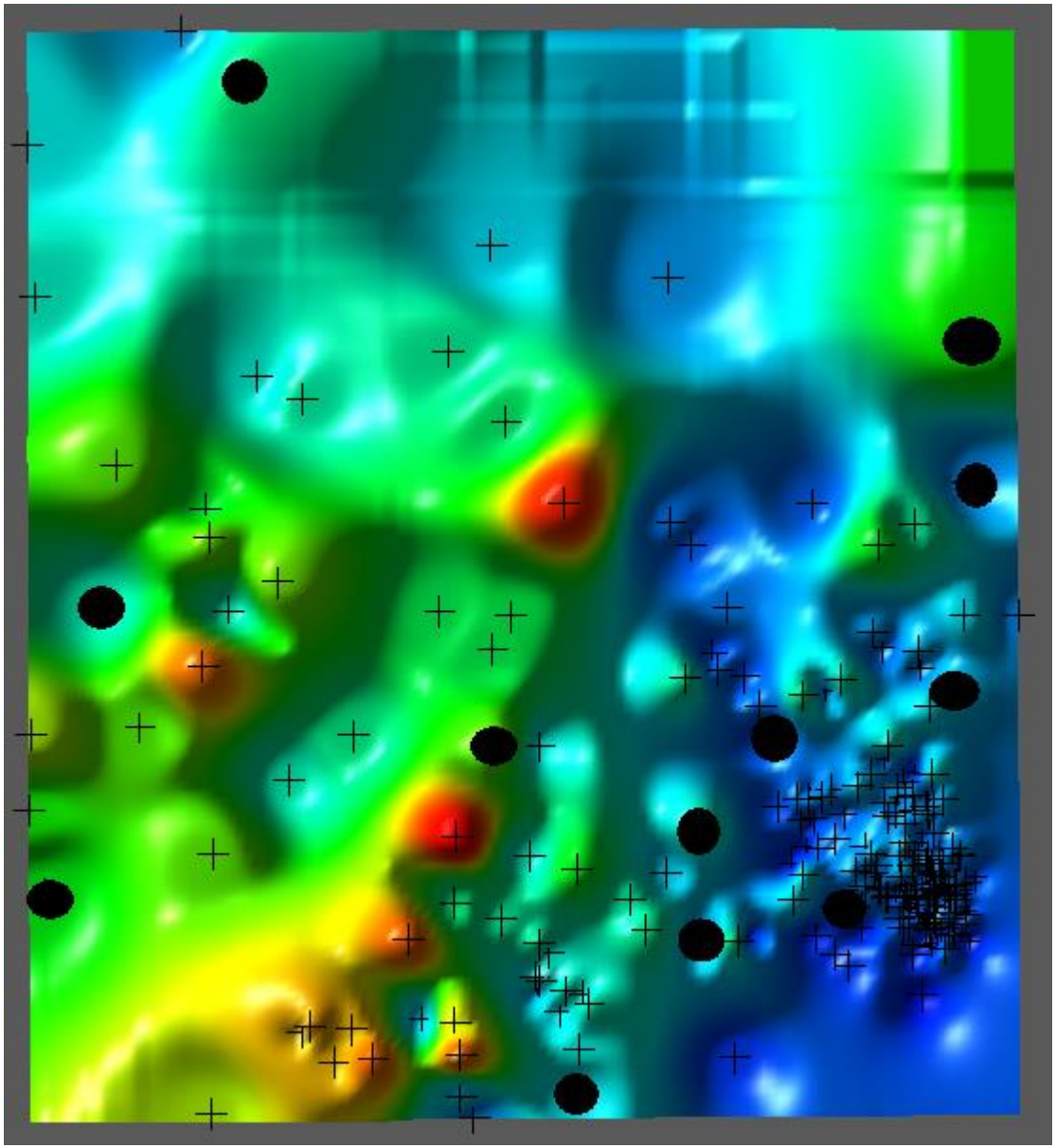


....and regularly shooting the ones at lower altitudes.

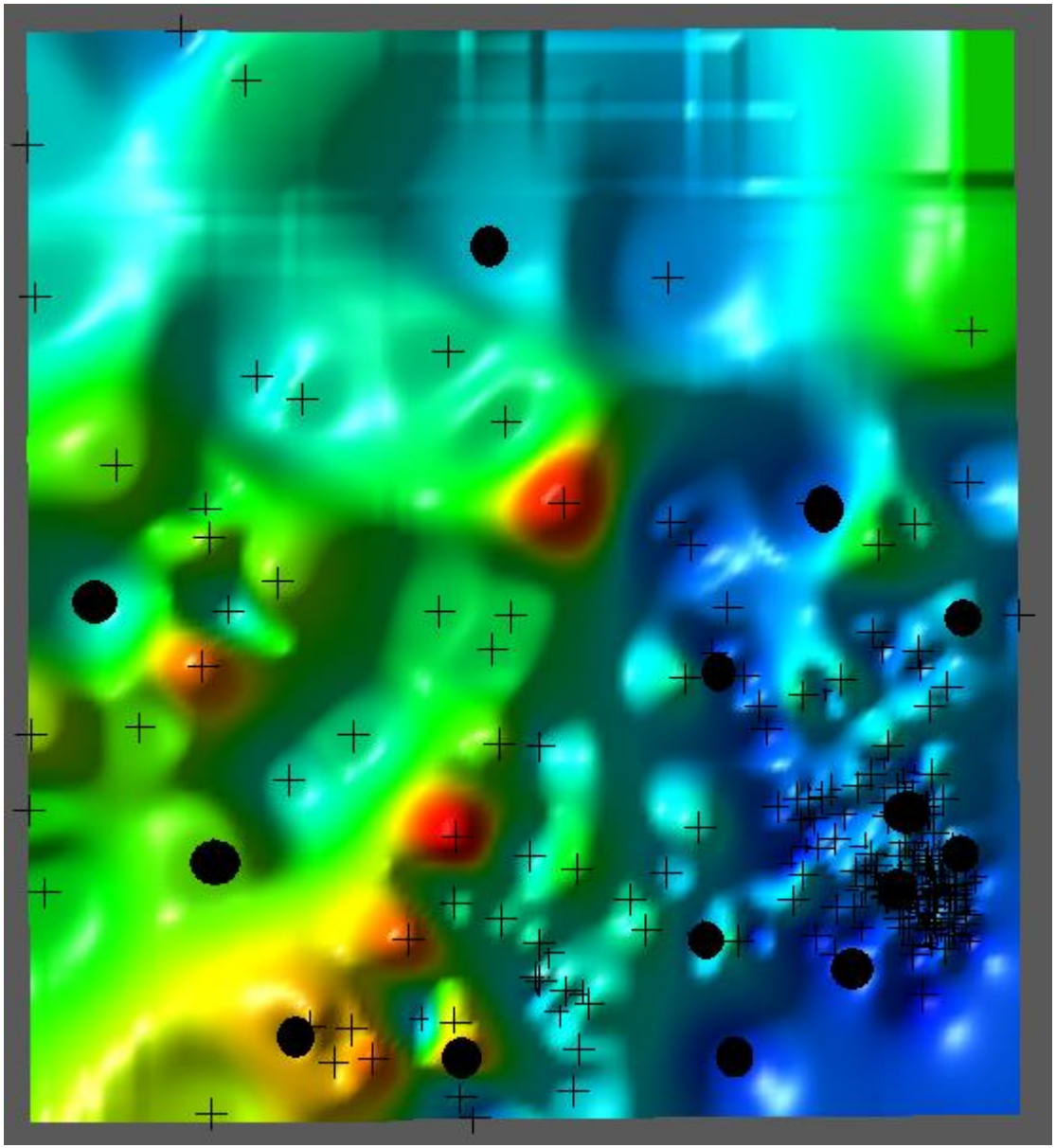
Evolutionary simulations learn & explore the most promising regions of search space



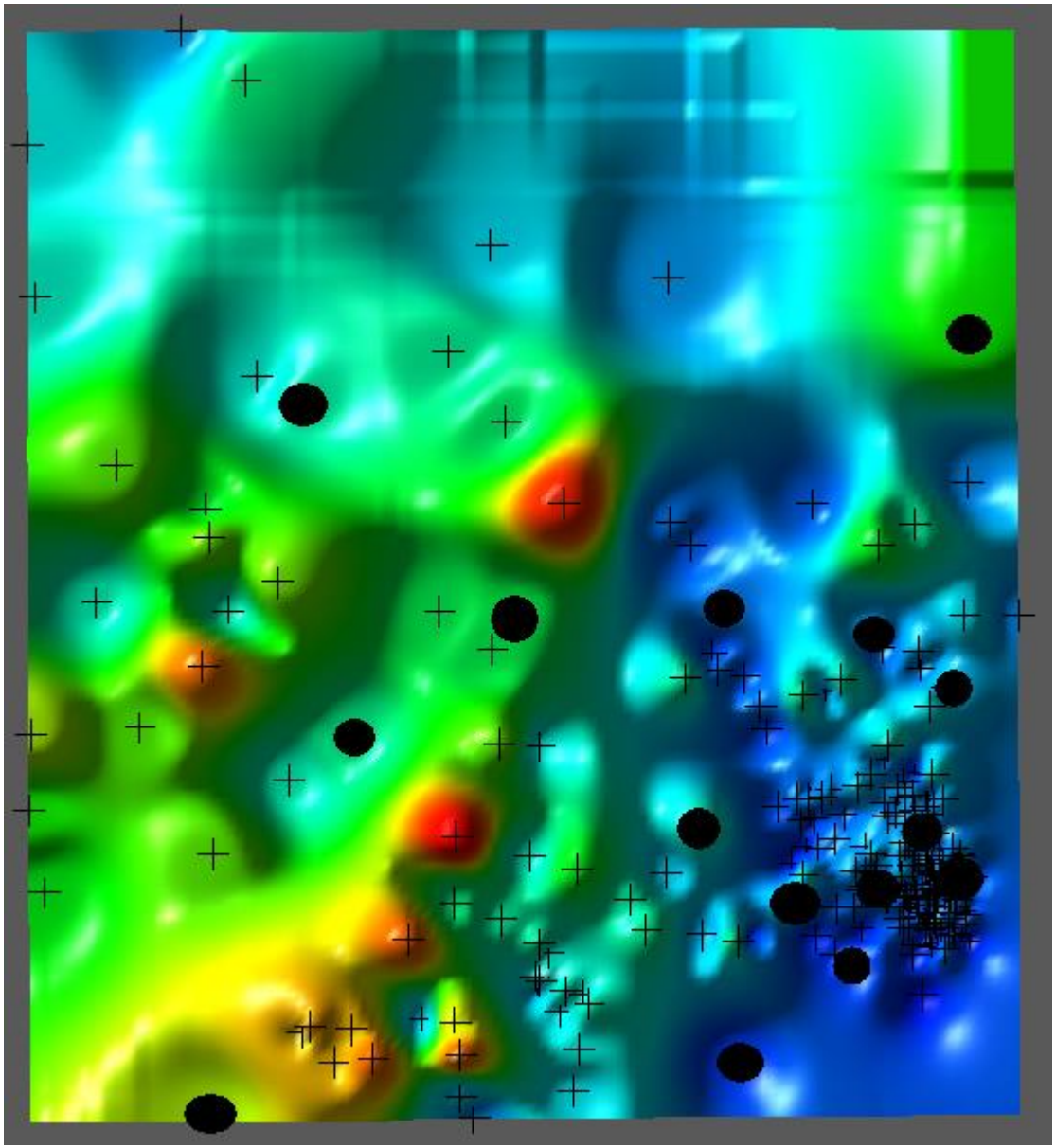
Evolutionary simulations learn & explore the most promising regions of search space



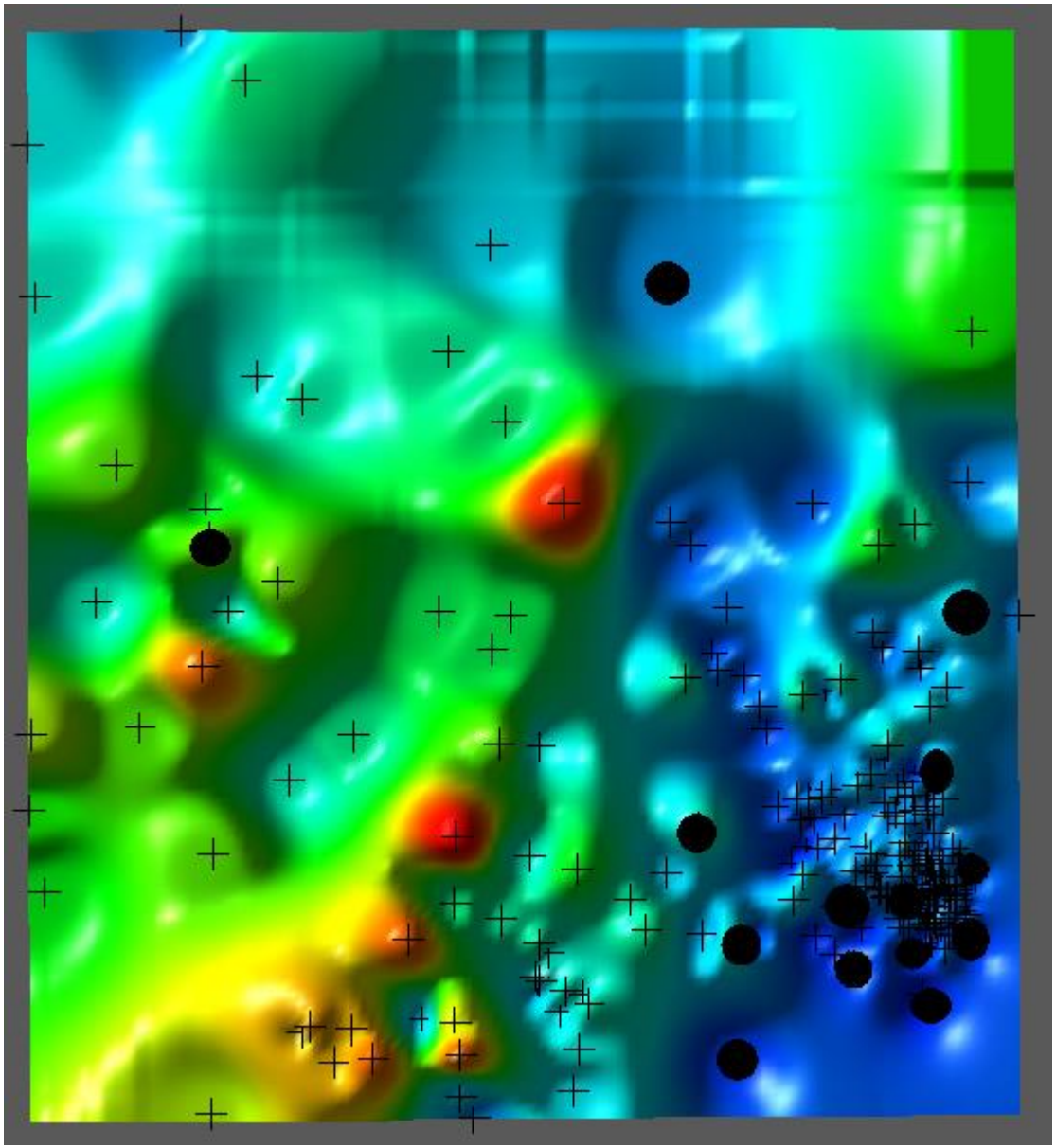
Evolutionary simulations learn & explore the most promising regions of search space



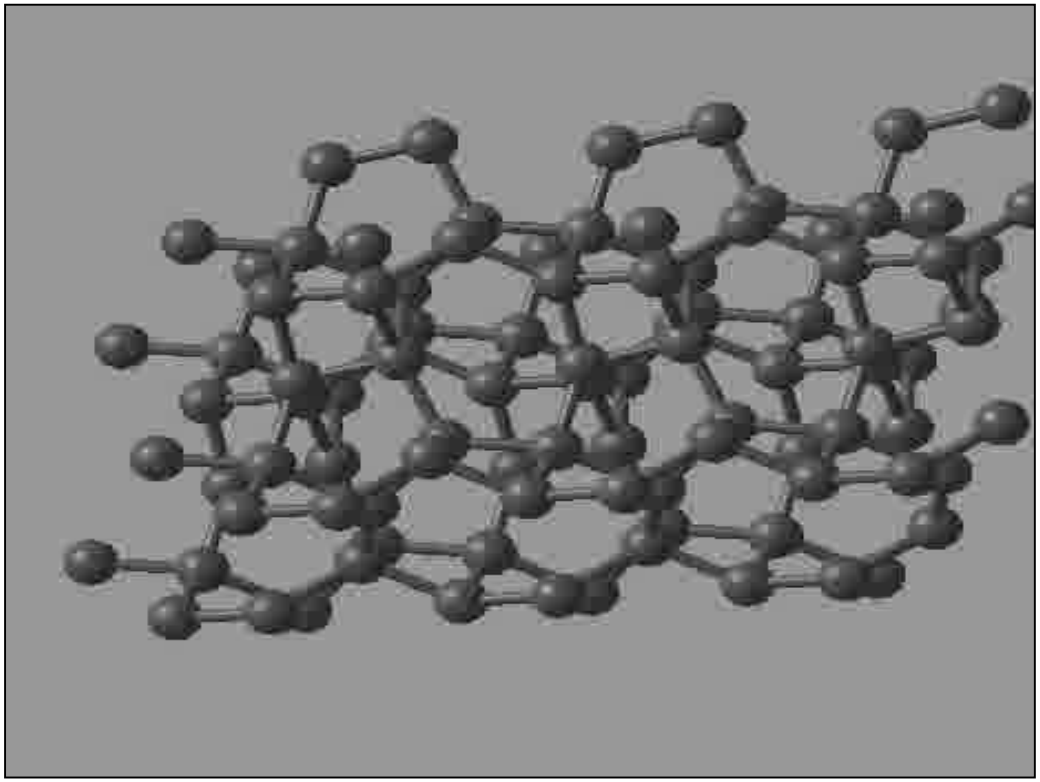
Evolutionary simulations learn & explore the most promising regions of search space



Evolutionary simulations learn & explore the most promising regions of search space

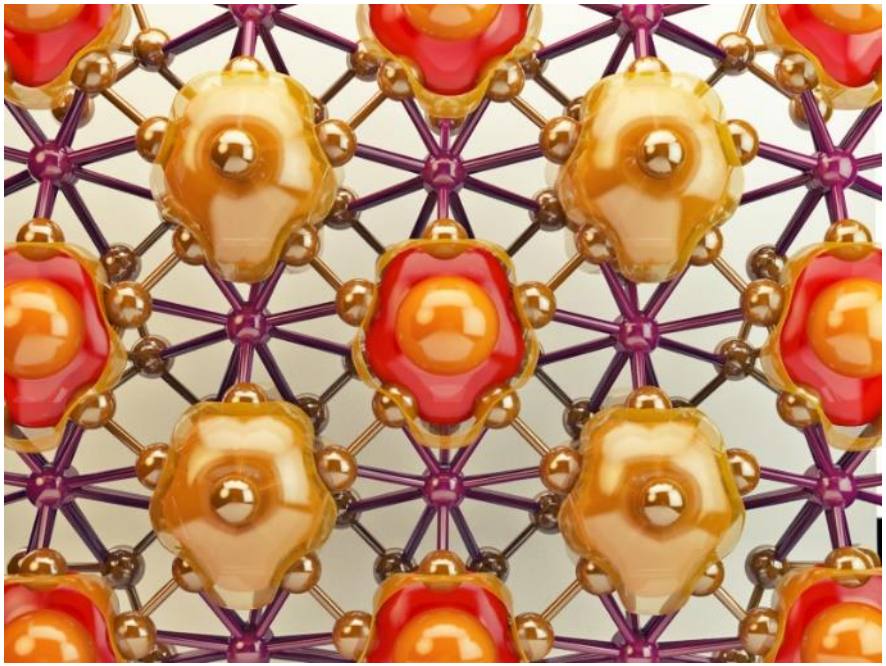


Without any empirical information, method reliably predicts materials



Carbon at 100 GPa – diamond structure is stable

Predicting new crystal structures without empirical information



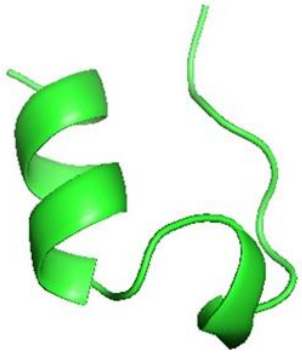
New superhard structure of boron
(Oganov et al., *Nature*, 2009)

High-pressure transparent allotrope of sodium
(Ma, Eremets, Oganov, *Nature*, 2009)

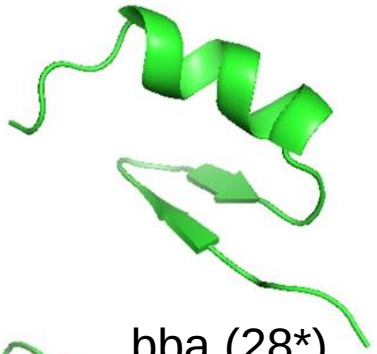
Prediction of very complex structures: proteins

[Rachitsky, Kruglov, Finkelstein, Oganov, submitted]

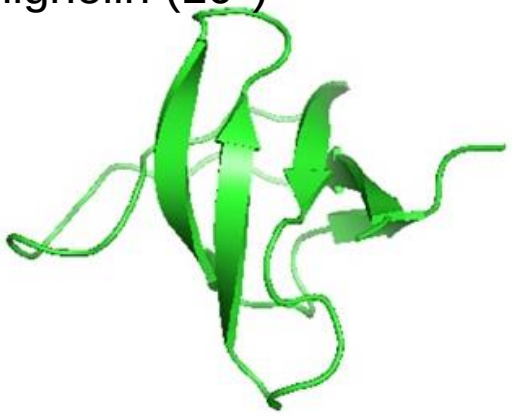
Chignolin (10*)



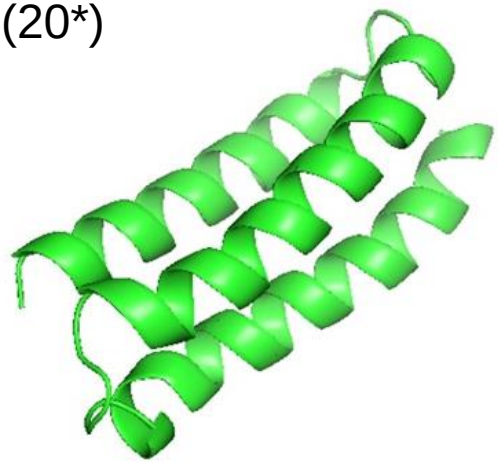
Trp-cage (20*)



bba (28*)



1shf (58*)



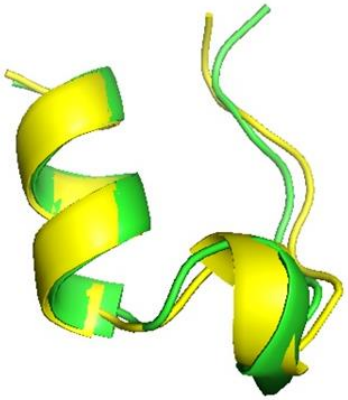
2a3d (73*)

* Number of amino acid residues

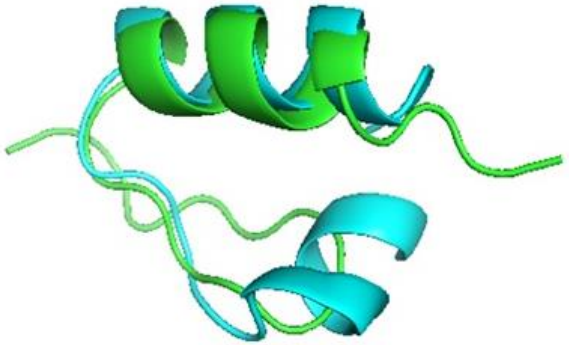
Prediction of very complex structures: proteins

[Rachitsky, Kruglov, Finkelstein, Oganov, submitted]

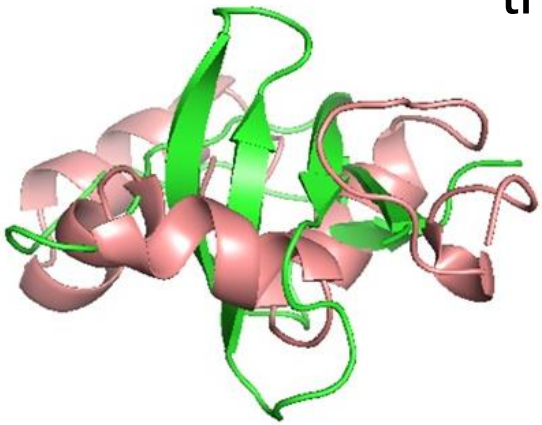
chignolin



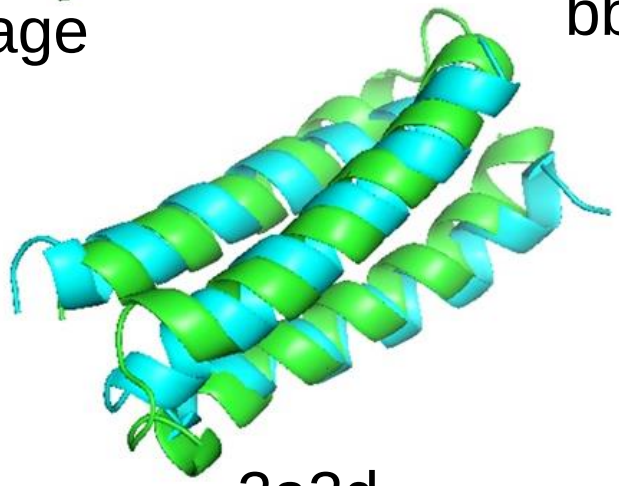
trp-cage



bba



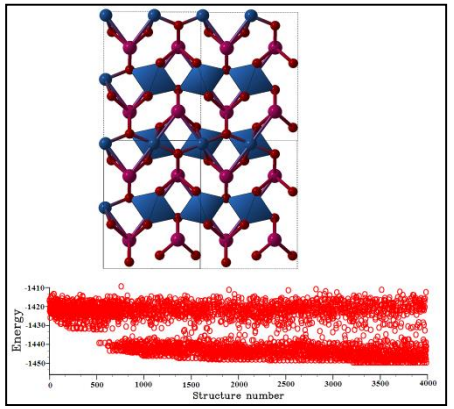
1shf

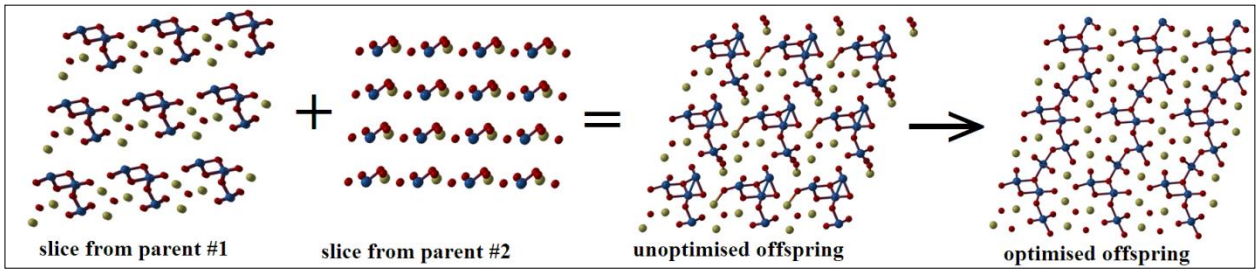


2a3d

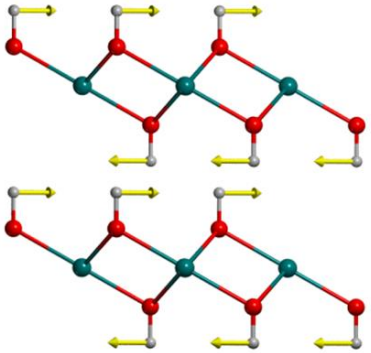
USPEX
(Universal Structure Predictor: Evolutionary Xtallography)

- (Random) initial population
- Evaluate structures by relaxed (free) energy
- Select lowest-energy structures as parents for new generation
- Standard variation operators:

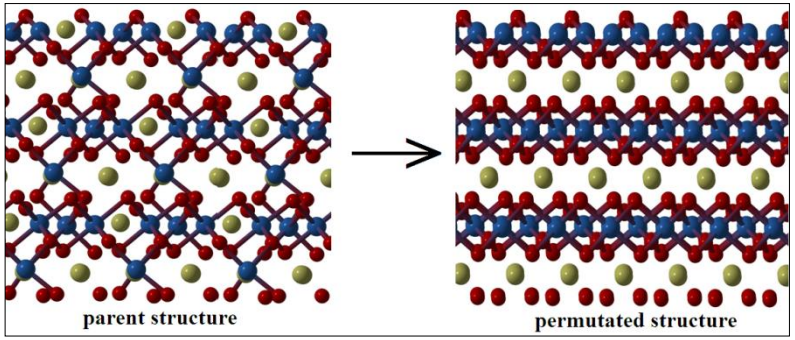




(1) Heredity (crossover)



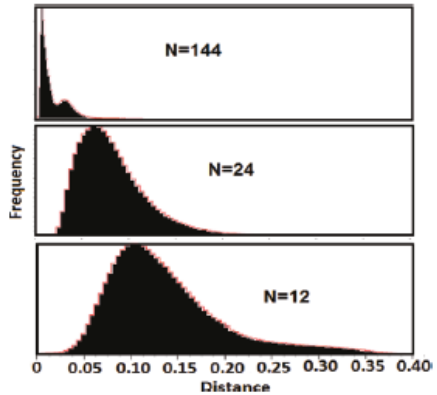
(2) Soft mode mutation



(3) Permutation

+(4) Transmutation, +(5) Rotational mutation, +(6) Soft-mode mutation, +...

Why does USPEX work so well? A few tips and tricks



Distribution of distances between randomly sampled local minima in a binary Lennard-Jones system AB_2.

formal dimensionality of full energy landscape: $d = 3N+3$

intrinsic dimensionality of reduced landscape: $d^* = 3N+3-k$

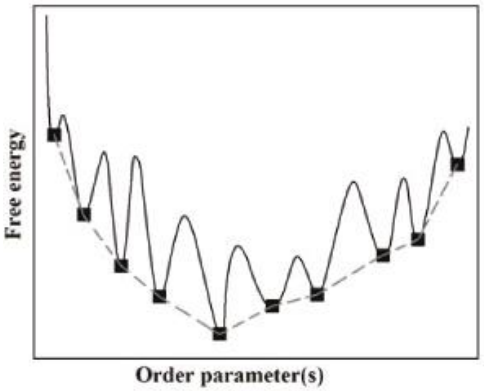
number of distinct structures: $C^* \sim \exp(\beta d^*)$

$d^* = 10.9$ ($d = 39$) for Au_8Pd_4

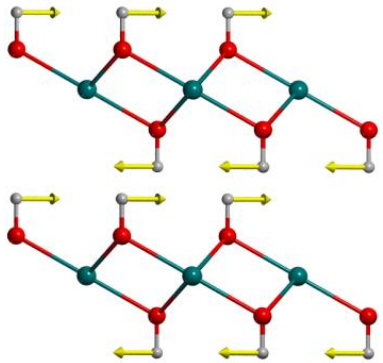
$d^* = 11.6$ ($d = 99$) for $Mg_{16}O_{16}$

$d^* = 32.5$ ($d = 39$) for $Mg_4N_4H_4$

I. Reduction of dimensionality through unbiased symmetric initialization.



II. Reduction of effective dimensionality of problem by structure relaxation (also reduces "noise" and transforms energy landscape to a convenient shape).



III. Variation operators are defined in subspaces of reduced dimensionality and involve cooperative transformations.

First lessons

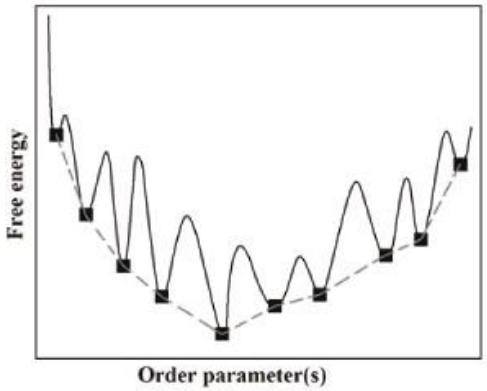
I. Local optimization (as part of global) is essential.

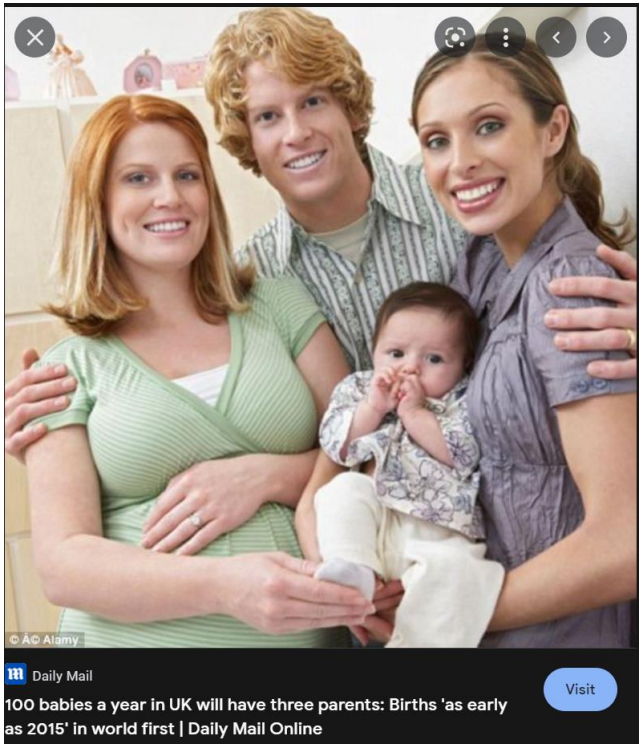
II. Initialization: must have diverse initial population.

III. Selection of potential parents: non-greedy.

IV. Variation operators: low-dimensional (alter a few degrees of freedom at a time).

V. Variation operators: 1-parent (mutations) or 2-parent (crossover). 3-parent crossover creates defective children.



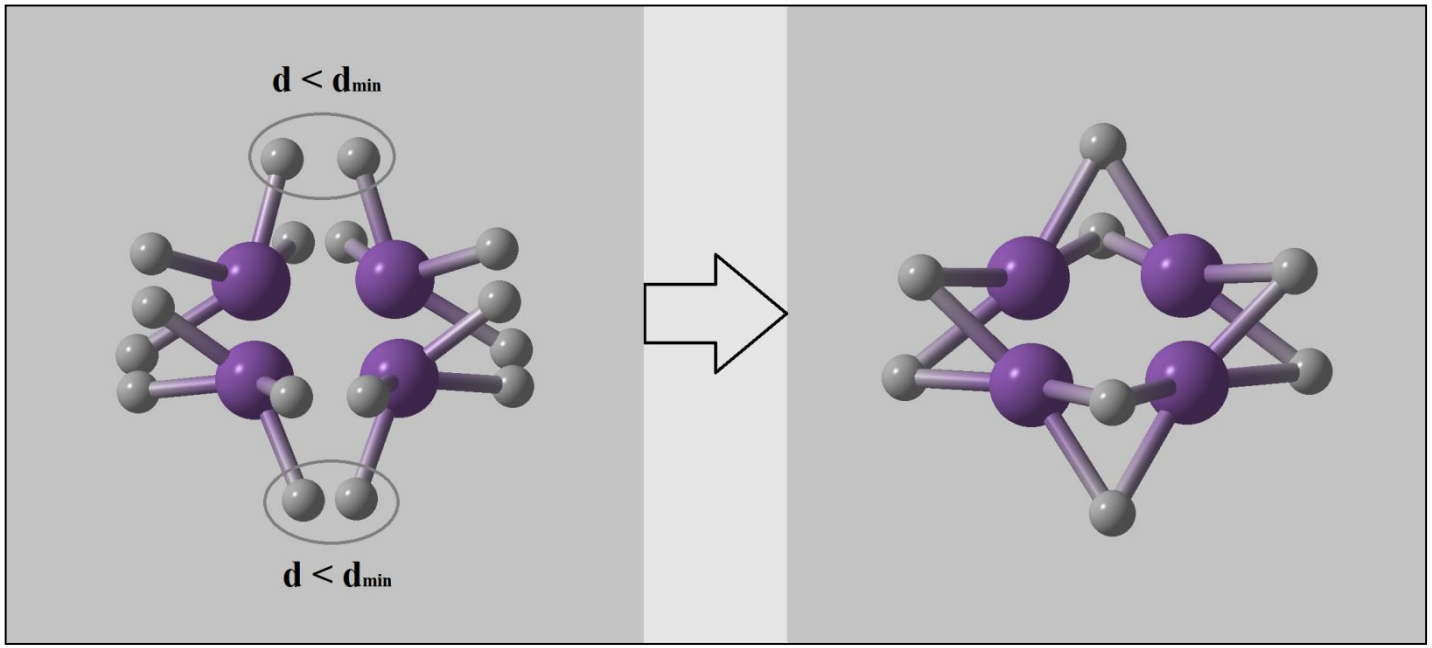


Ways to initialize a diverse population:

1. Random symmetric structure generator

(Lyakhov & Oganov, Comp. Phys. Comm., 2013)

Crystals: 230 space groups
Nanoparticles: point groups



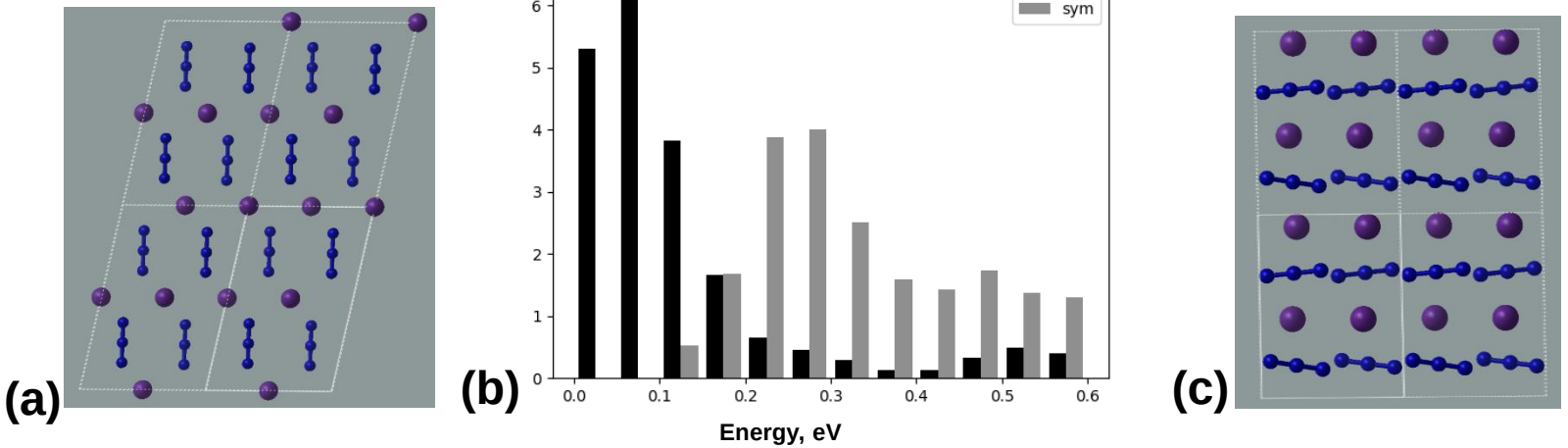
Enables moderately efficient random sampling as one of possible USPEX regimes

Zhu, Oganov, et al, Acta. Cryst. B, 68, 215-226 (2012)

Ways to initialize a diverse population: 2. Random topological structure generator

(Bushlanov, Blatov, Oganov, Comp. Phys. Comm. 2019)

Speedup ~3 times

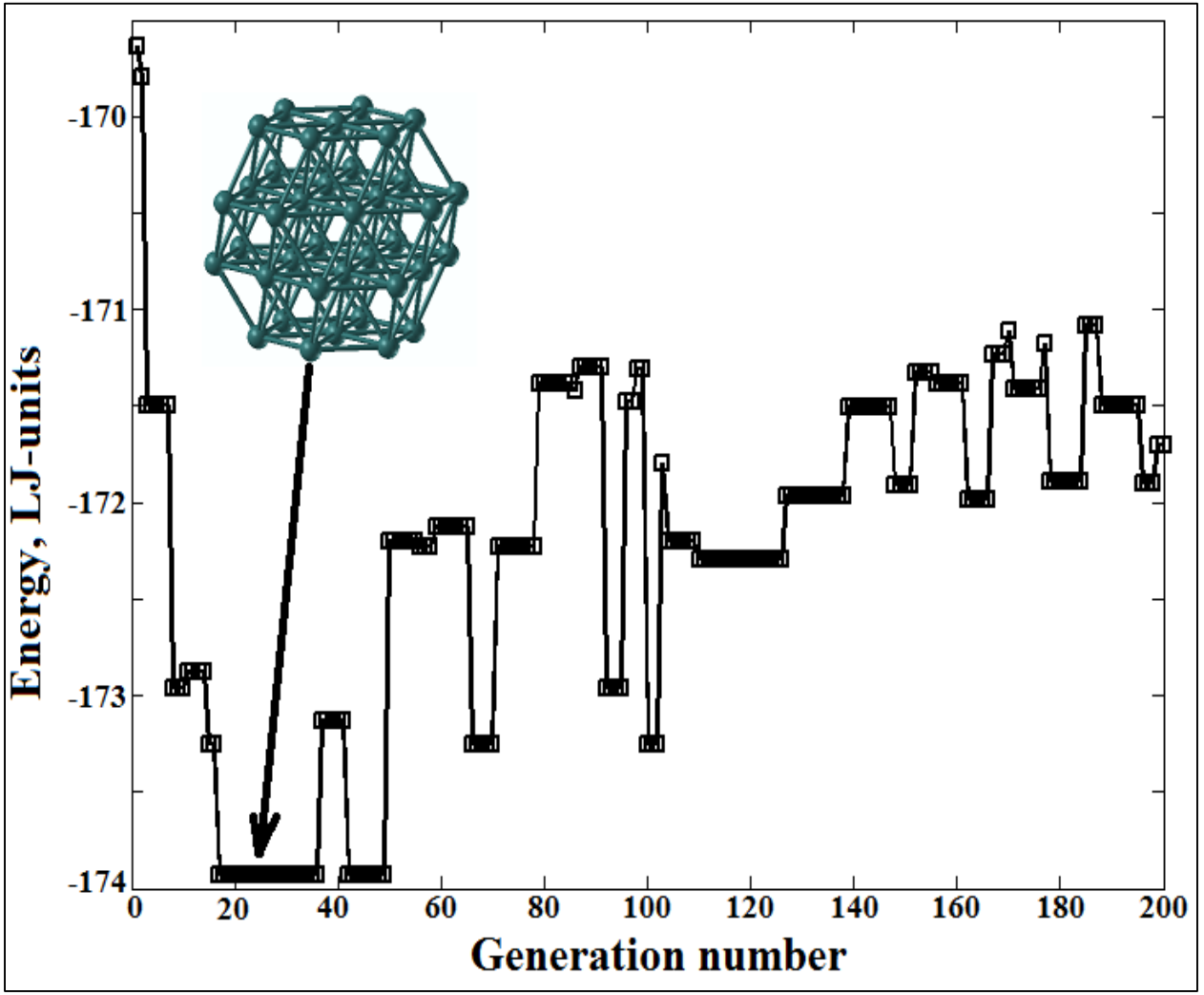


Example of KN_3: (a) topological structure, (c) random symmetric structure, (c) energy distribution of topological (TR) and random symmetric structures

Statistics (100 runs) of USPEX performance on $MgAl_2O_4$ (28 atoms/cell) at 100 GPa

	Original	On-the-fly adaptation	On-the-fly adaptation AND topology
<No. of structures>	1307	1069	368
Success rate	100%	100%	100%

One more lesson: ‘aging’ technique to prevent “genetic drift”



A metastable state is found first and the ground state is found shortly after

[Lyakhov, ARO, et al. (2013)]

2. When there is more than one optimal solution

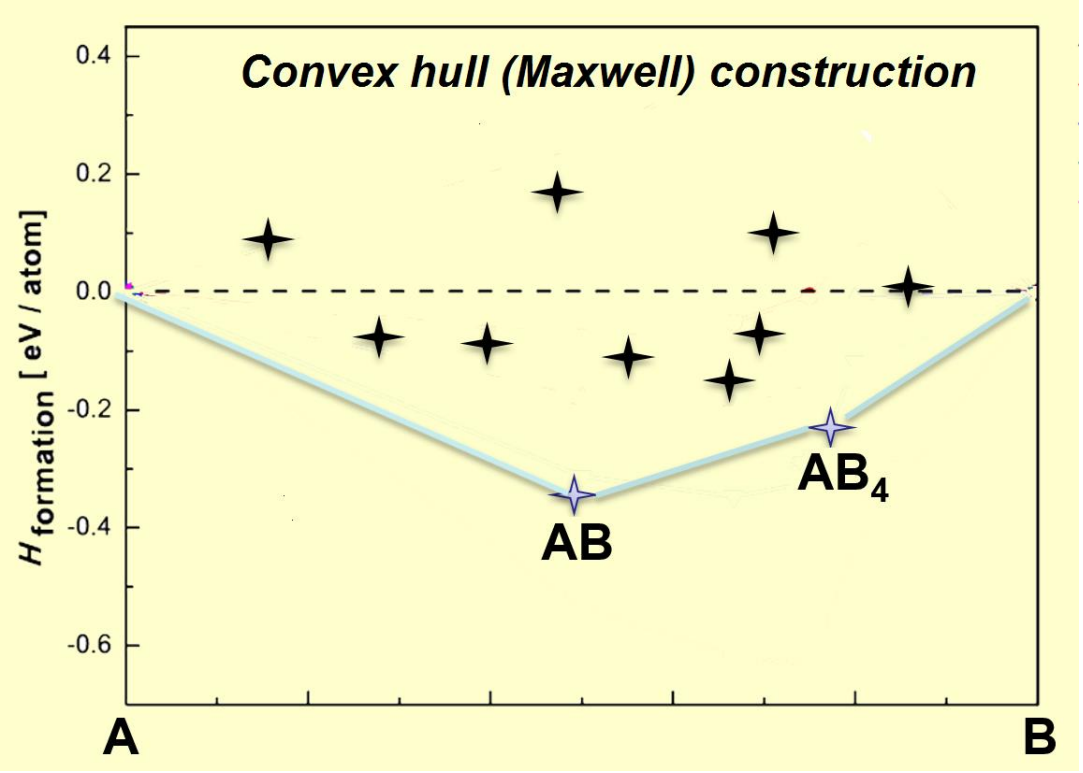
-Pareto optimization
-Variable-composition systems
-Coevolution

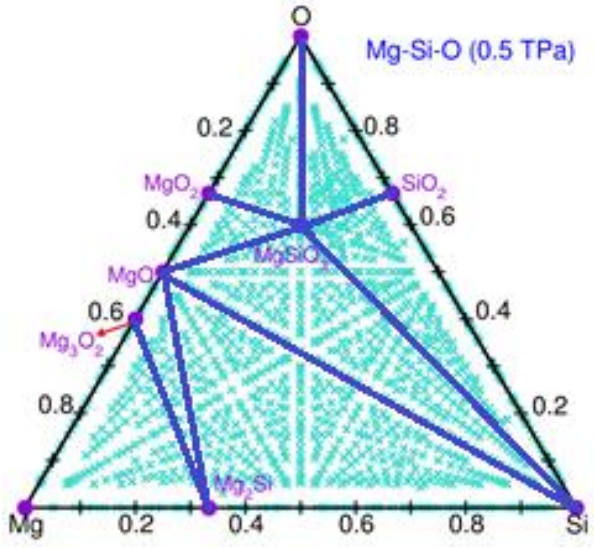
Which compounds of elements A and B are stable?

(*cf* a class of related species?)

To predict thermodynamic stability, we must use the Maxwell construction (the convex hull)

Thermodynamic stability in variable-composition systems

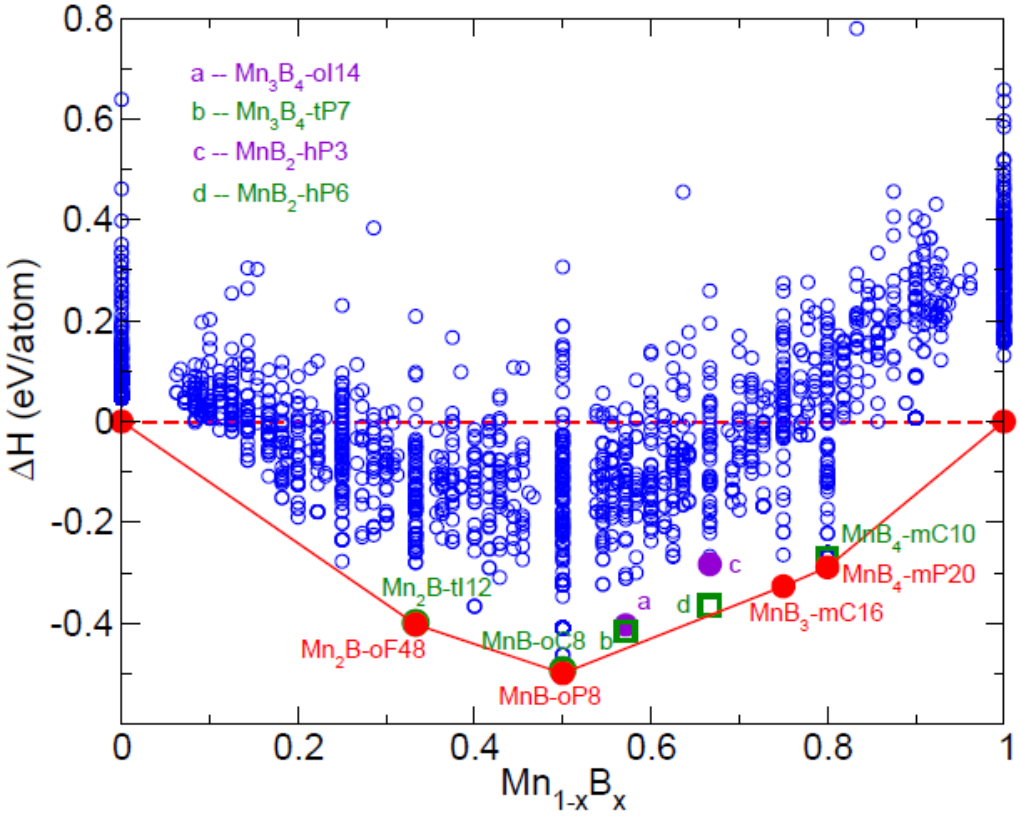


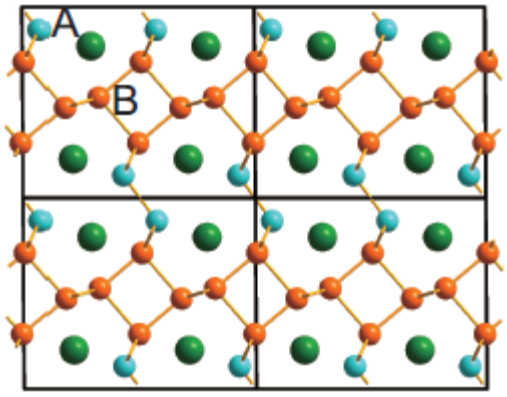


3-component convex hull: Mg-Si-O system at 500 GPa (Niu & Oganov, *Sci. Rep.* 2015)

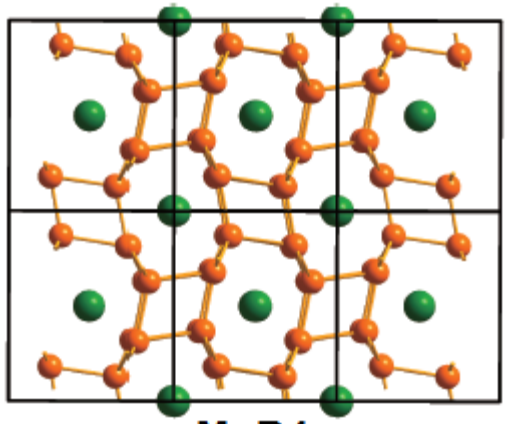
Stable structure must be below all the possible decomposition lines !!

Predicting new compounds: example of “well-known” system Mn-B





MnB3



MnB4

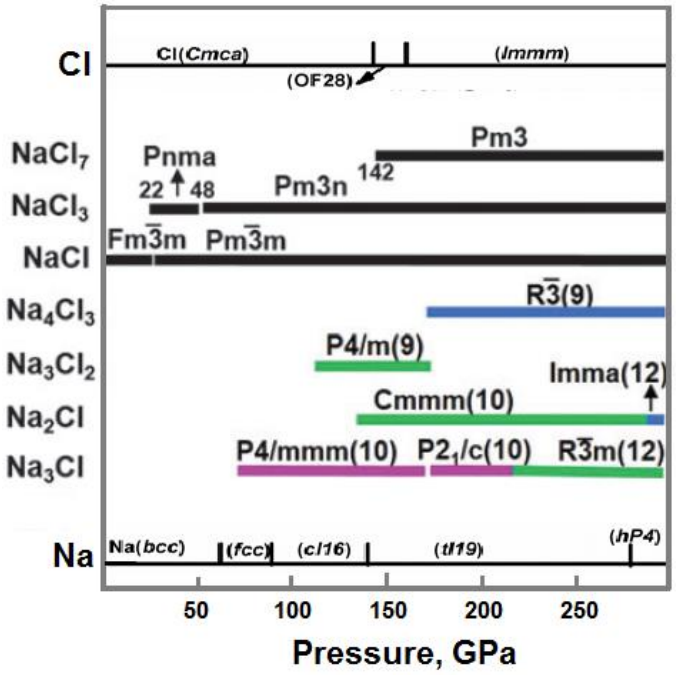
MnB3 was predicted and then synthesized (Niu et al., PCCP, 2014)

1. **New compound discovered – MnB3.**
2. **For MnB4, reported experimental structure was wrong – new experiments confirm our structure.**

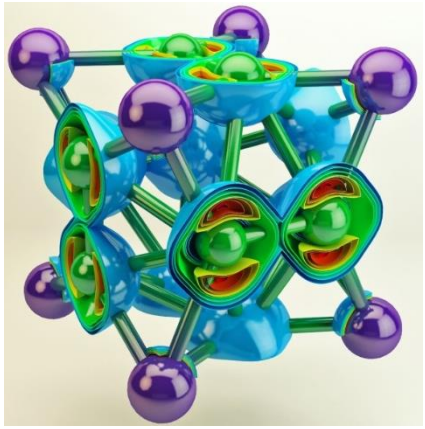
Predictive power of modern methods:

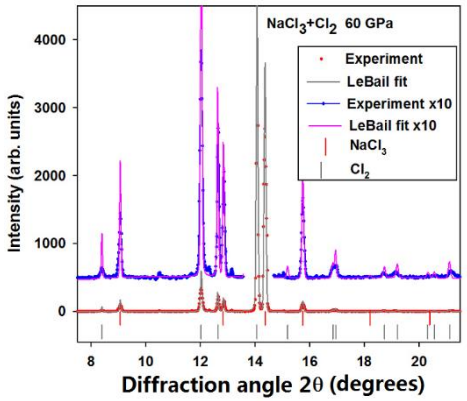
Na-Cl

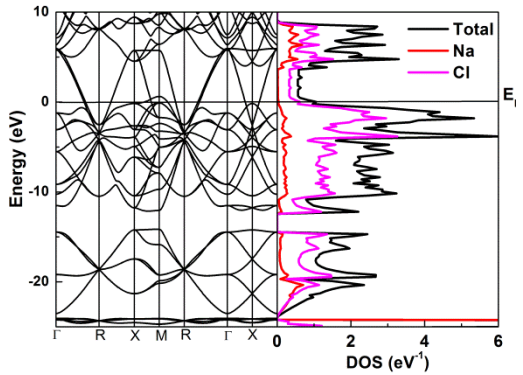
Na_3Cl, Na_2Cl, Na_3Cl_2, NaCl, $NaCl_3$, $NaCl_7$ are stable under pressure [Zhang, Oganov, et al. *Science*, 2013].

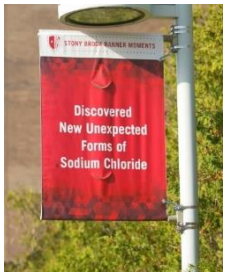


Stability fields of sodium chlorides









$NaCl_3$: atomic and electronic structure, and experimental XRD pattern

Chemical anomalies:

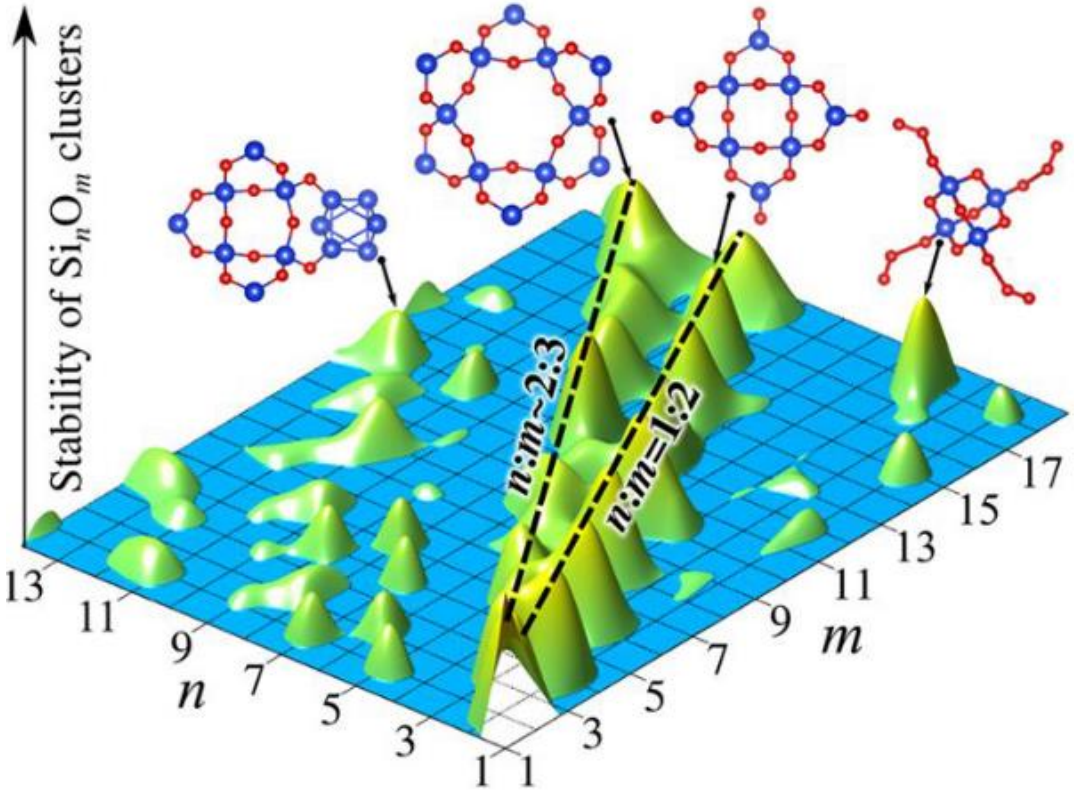
-Divalent Cl in Na_2Cl!
-Coexistence of metallic and ionic blocks in Na_3Cl!
-Positively charged Cl in $NaCl_7$!

[Zhang, Oganov, et al., *Science* (2013)]
[Saleh & Oganov, *PCCP* (2015)]

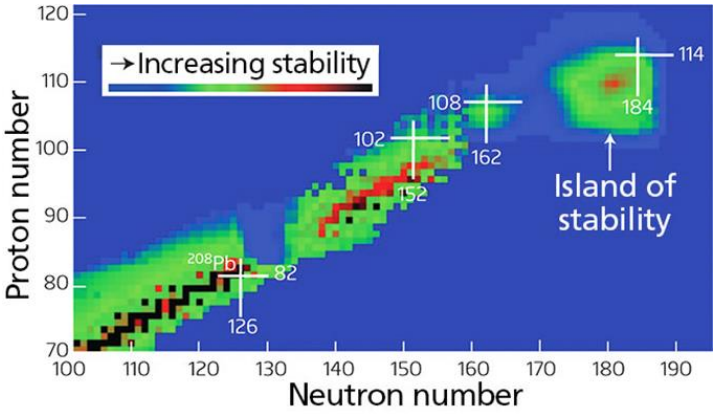
Map of stability of Si-O clusters

[Lepeshkin & Oganov, *J. Phys. Chem. Lett.* 2019]

Si-O



Ridges of stability: SiO_2, Si_2O_3
Islands of stability: e.g., Si_4O_{18}



«Magic» nuclei: with filled proton or neutron shells (2, 8, 20, 28, 50, 82, 126 p or n)
($1s^2/2p^6/3d^{10}2s^2/4f^8/4f^63p^65g^{10}/5g^84d^{10}3s^26h^{12}$)

Magic numbers of electrons = 2, 10, 18, 36, 54, 86, 118)

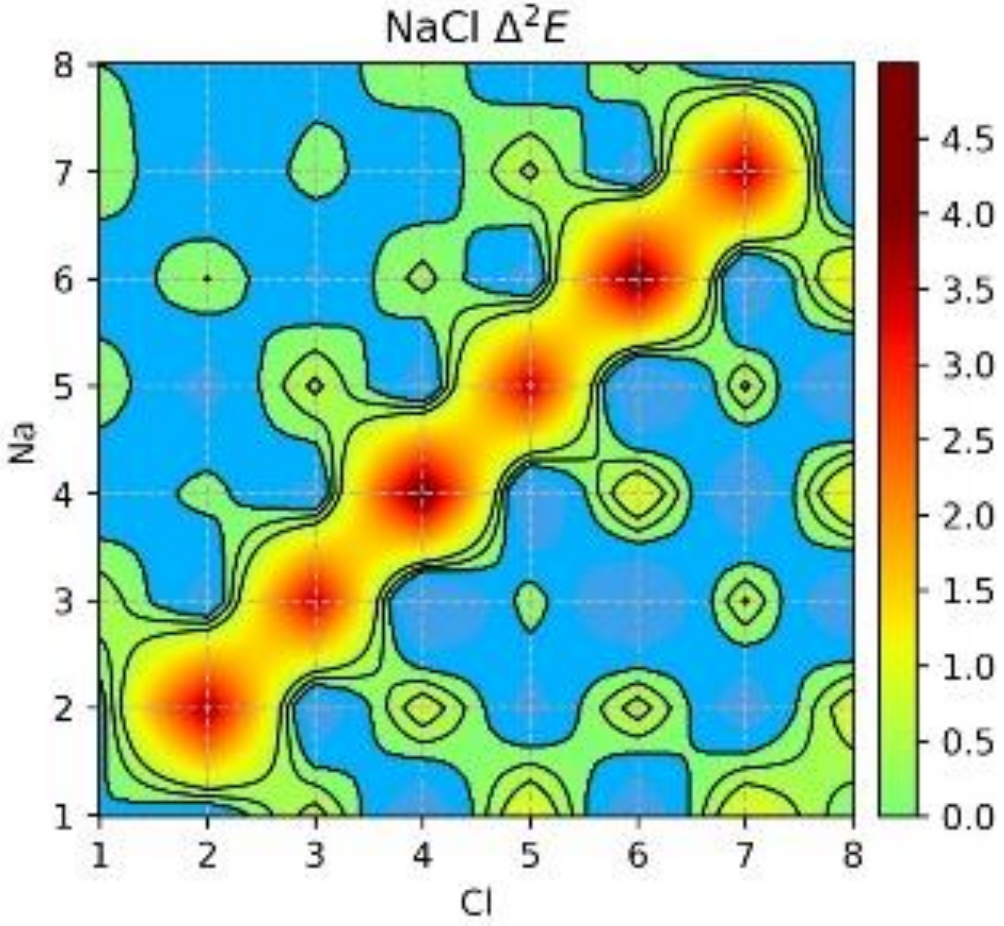
Analogy with magic atomic nuclei

Na-Cl

How to predict stability of NaCl molecules

$(NaCl)_n$ ridge of stability.

Numerous islands of stability.

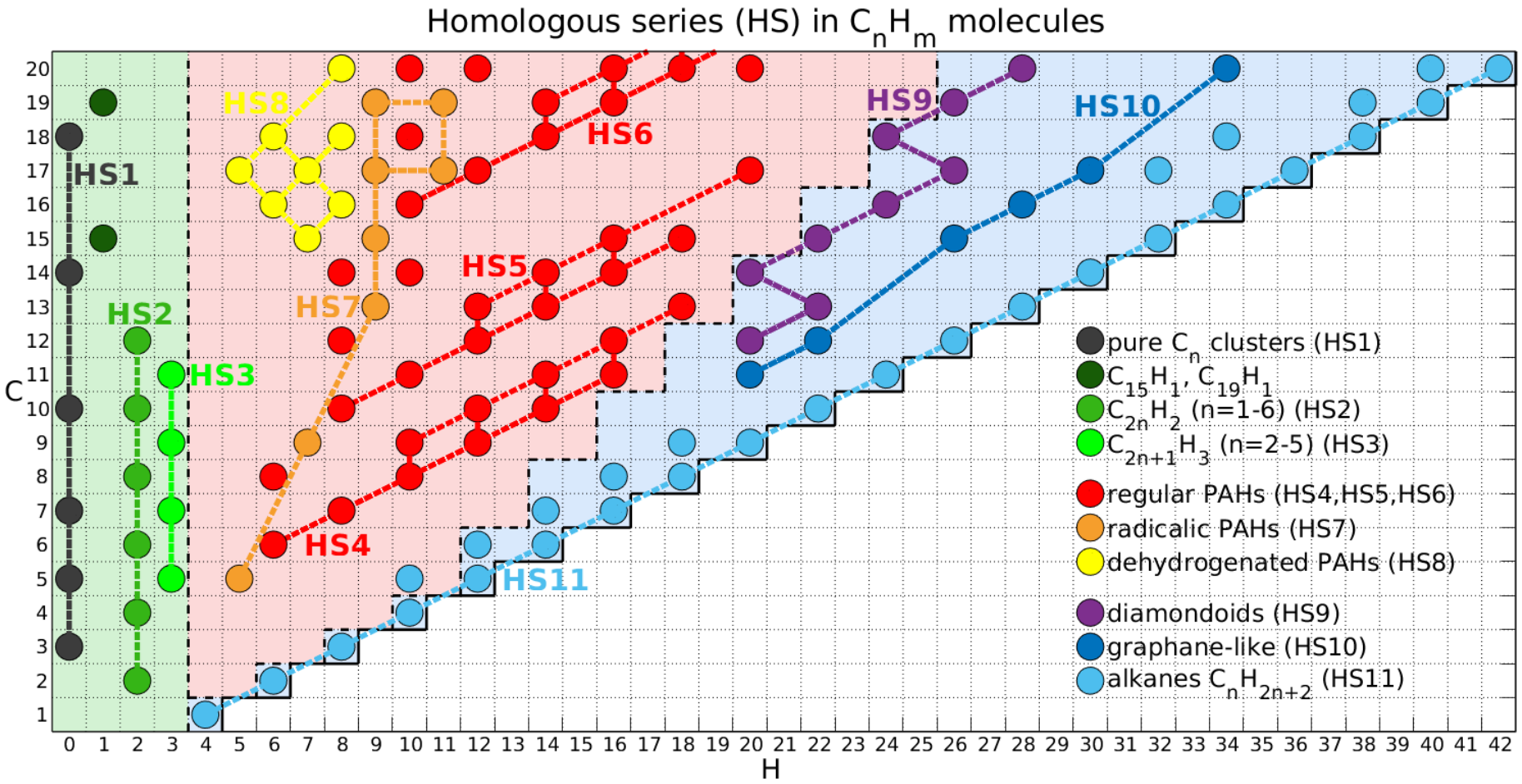


(result of M. Fedyaeva and S.V. Lepeshkin)

Which C-H molecules are stable?

[Lepeshkin, Oganov, submitted]

-Huge diversity, explaining the richness of organic chemistry.



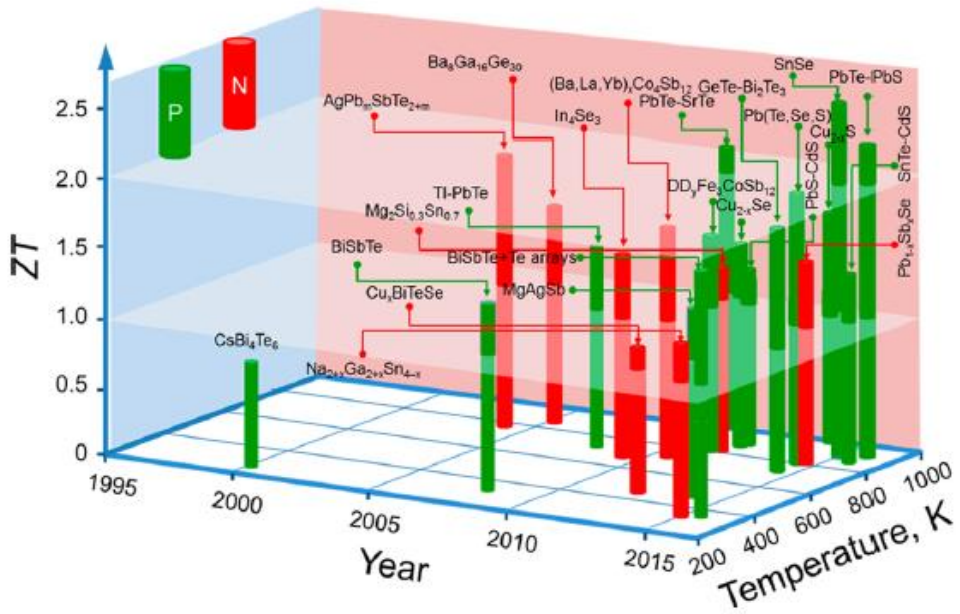
Which compounds have an optimal combination of several properties?

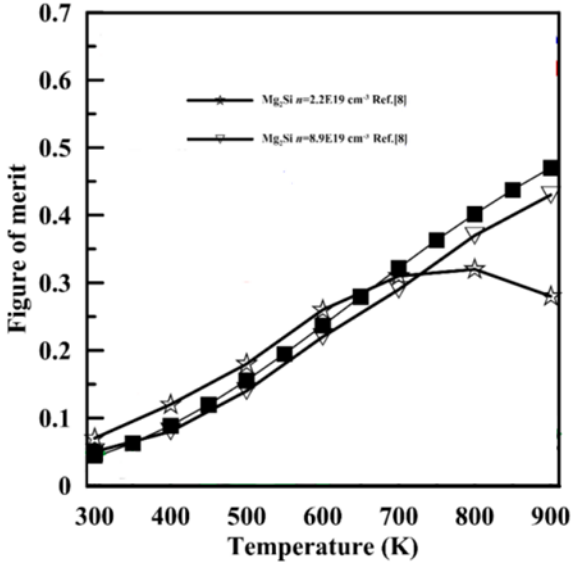
(Pareto optimization – *cf* ecological niches)

How to improve efficiency of thermoelectric devices?

"One shouldn't work on semiconductors, that is a filthy mess; who knows whether any semiconductors exist"

-W. Pauli, letter to R. Peierls (1931)





[Fan & Oganov (2018)]

$$ZT = \frac{\sigma S^2 T}{\kappa}$$

$$S = \frac{8\pi^2 k_B^2}{3eh^2} m^* T \left(\frac{\pi}{3n}\right)^{2/3}$$

$$\eta = \frac{\Delta T}{T_H} \frac{\sqrt{1+zT}-1}{\sqrt{1+zT}+\frac{T_C}{T_H}}$$

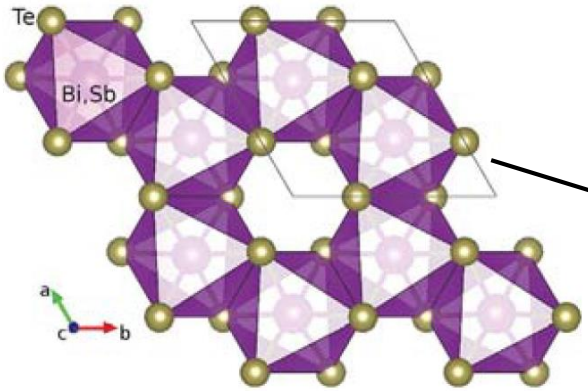
- efficiency

Multiobjective (Pareto) optimization finds a new thermoelectric polymorph of Bi_2Te_3

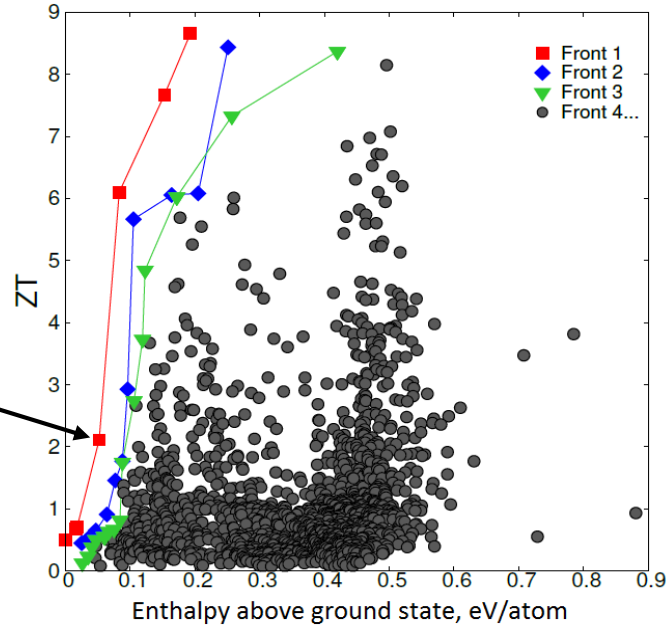
Computer Physics Communications 222 (2018) 152–157

Efficient technique for computational design of thermoelectric materials

Maribel Núñez-Valdez , Zahed Allahyari , Tao Fan , Artem R. Oganov



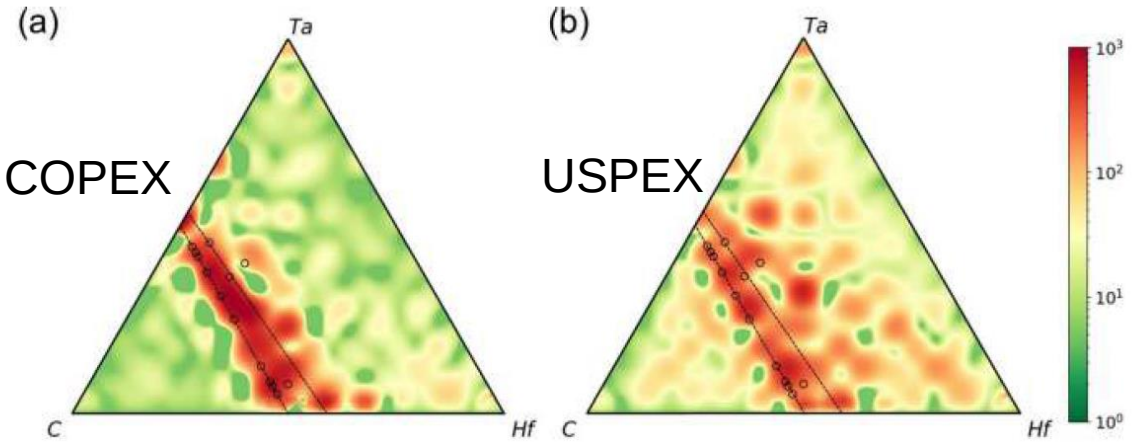
Predicted *P*6$_3$*cm* structure of Bi_2Te_3



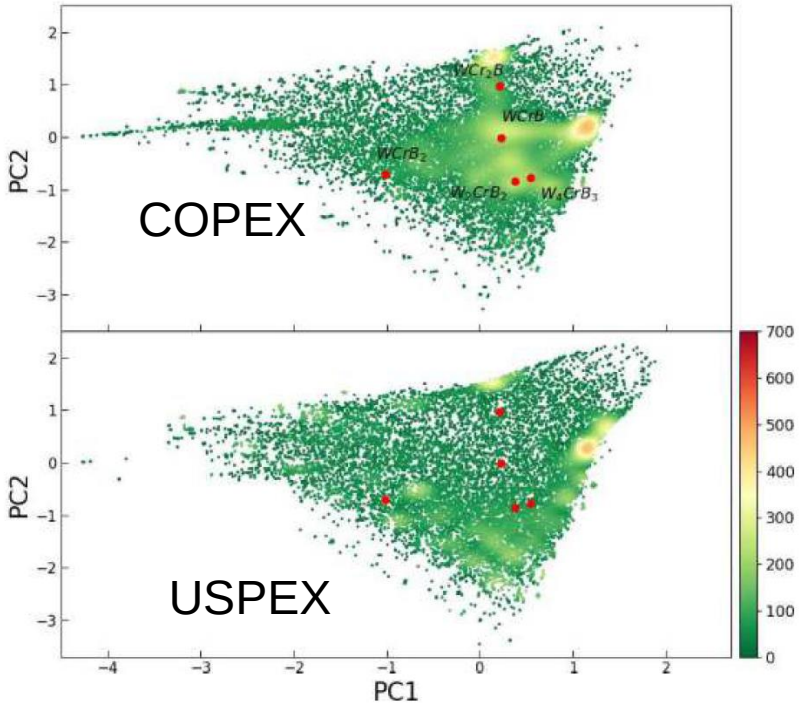
Pareto optimization of ZT and stability in the Bi-Te system

Coevolution: sampling very wide search spaces

Coevolution to enhance sampling of the chemical space: COPEX method (Liu, Niu, Oganov, 2021)



Hf-Ta-C system

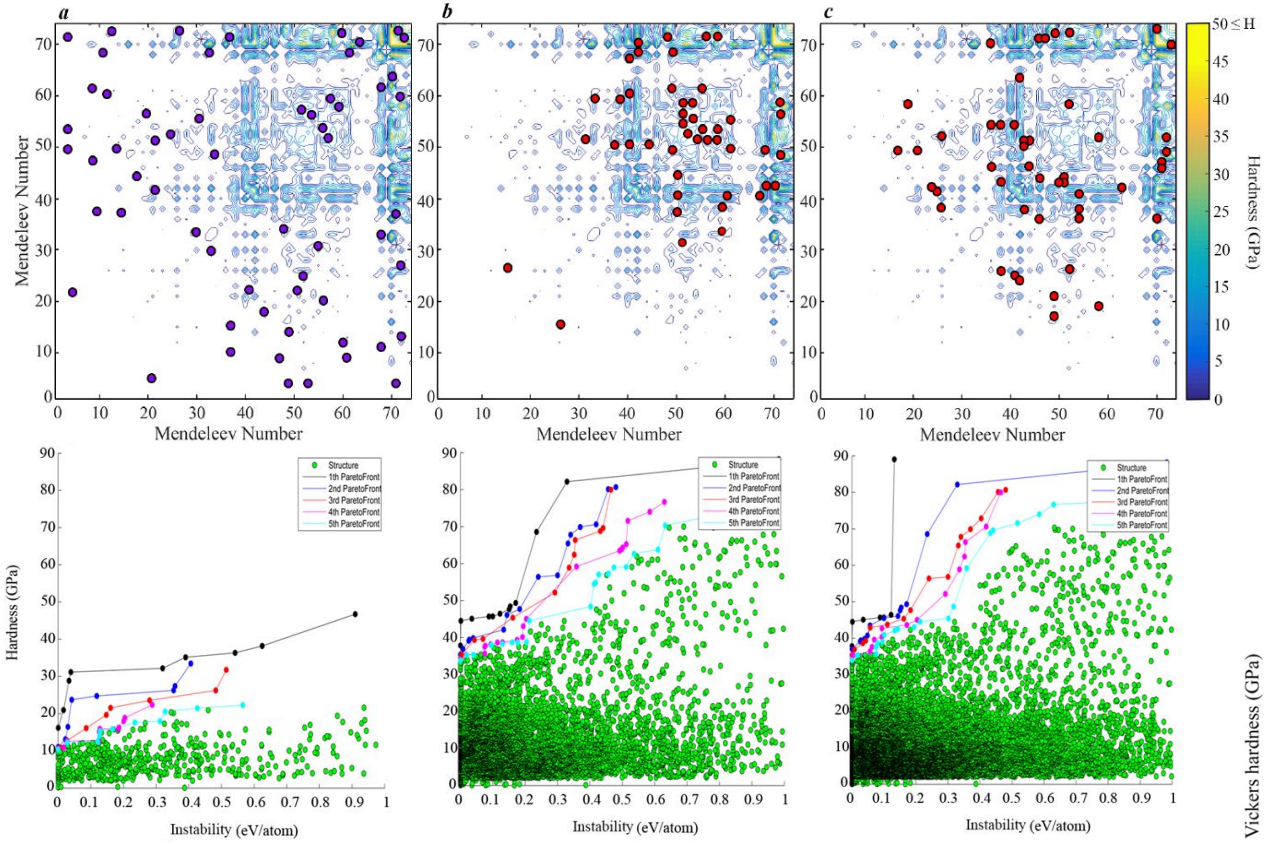


W-Cr-B system
(PCA map)

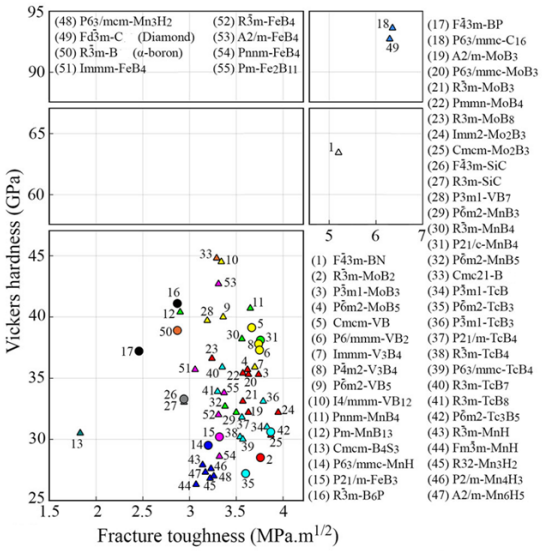
Extending the concept further: Mendelevian search predicts optimal materials among all possible compounds

[Allahyari & Oganov, *npj Comp. Mat.*, 2020]

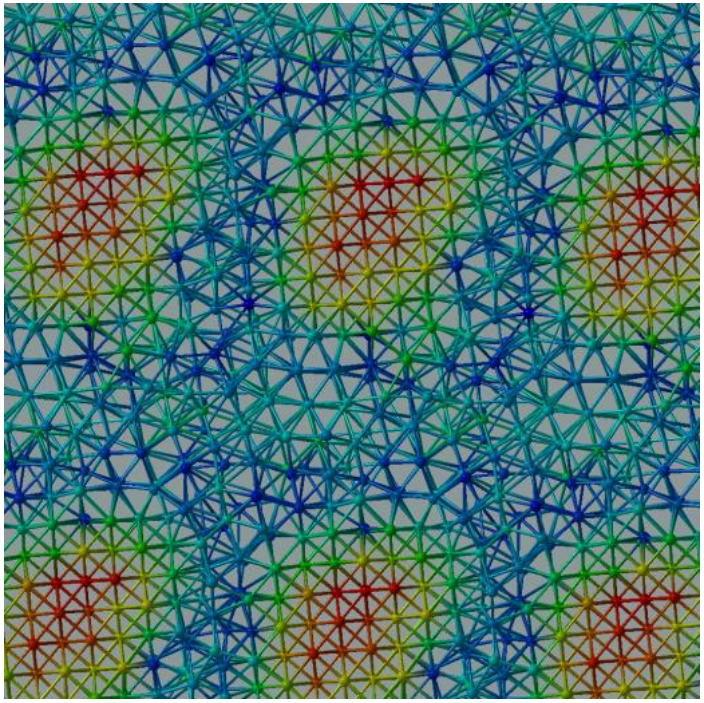
1st generation **5th generation** **10th generation**



Diamond is the hardest possible crystalline material!

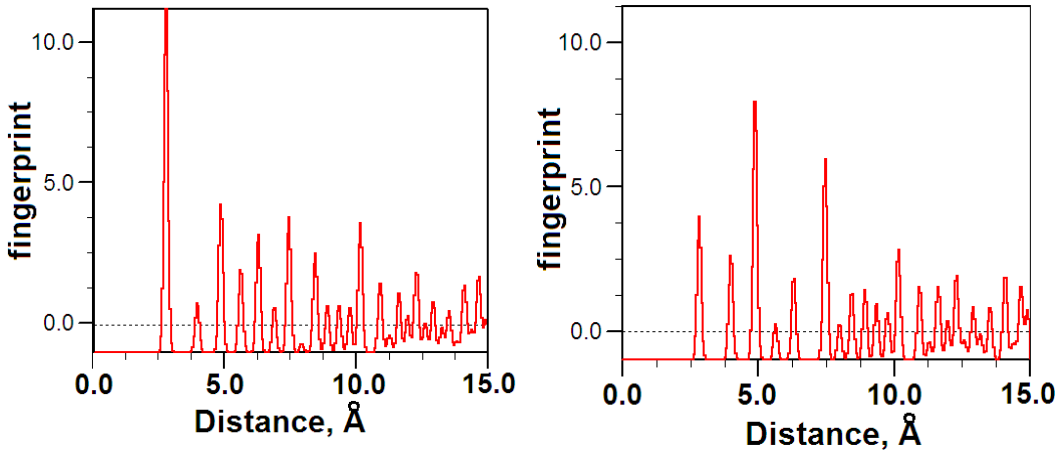


3. Looking into the genome



Fingerprint theory is the basis of our analysis

Fingerprint function is a 1D-descriptor of the structure: diffraction spectrum, PCF, ...



Fingerprint is related to the pair correlation function

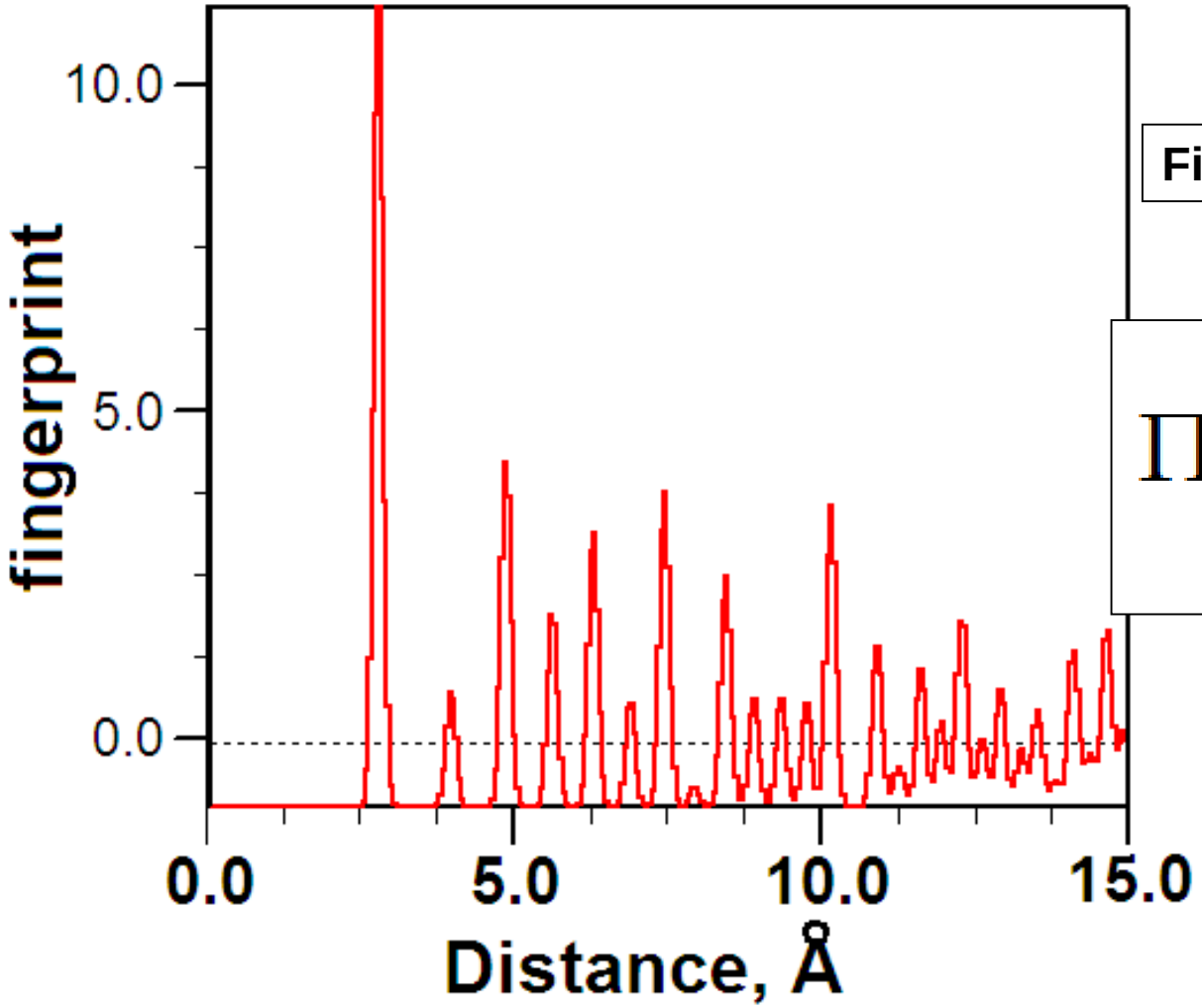
$$F_{AB}(R) = \sum_{A_i,\text{cell}} \sum_{B_j} \frac{\delta(R - R_{ij})}{4\pi R_{ij}^2 \frac{N_A N_B}{V} \Delta} - 1 = g_{AB}(R) - 1$$

Difference between 2 structures is given by „distance“, e.g.: $D = \int (f_1 - f_2)^2 dR$

“Cosine distance” $D = \frac{1}{2}\left(1 - \frac{f_1 f_2}{|f_1||f_2|}\right)$**, always in the range [0;1]**

[ARO & Valle, J. Chem. Phys. 130, 104504 (2009)]

Structure fingerprints



Fingerprint f=0 for ideal gas

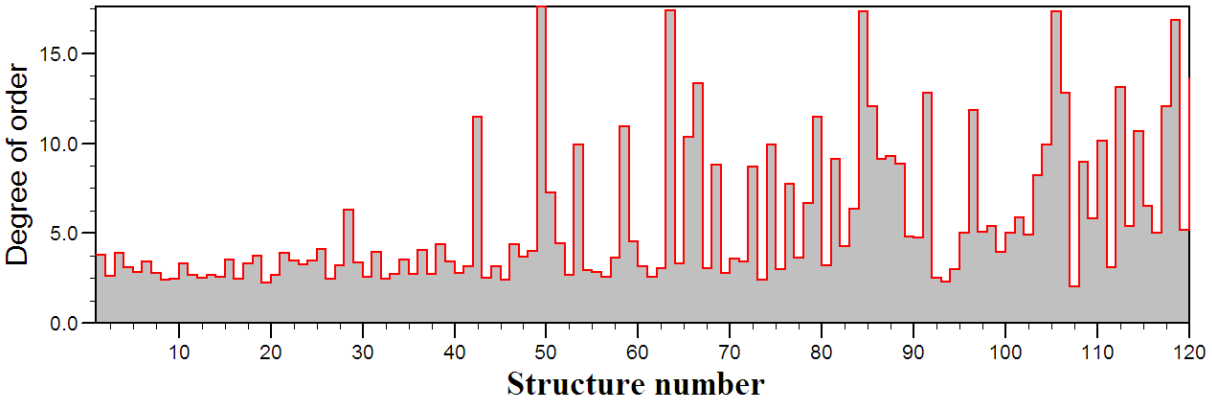
Degree of order :

$$\Pi = \frac{1}{V^{1/3}} \int_0^{\infty} f^2 \, dR$$

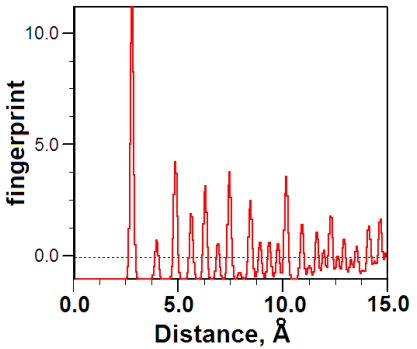
[ARO & Valle, J. Chem. Phys. 130, 104504 (2009)]

Evolution leads to the loss of entropy

$$\Pi = \frac{1}{V^{1/3}} \int_0^{\infty} f^2 dR$$ - Degree of order

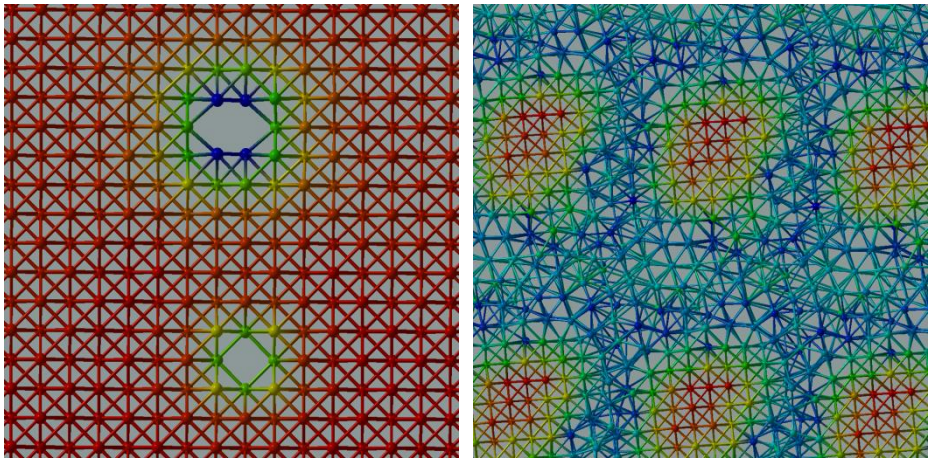


Emergence of order from chaos during evolutionary simulation



Fingerprint: can be computed for whole structure or for each atomic site

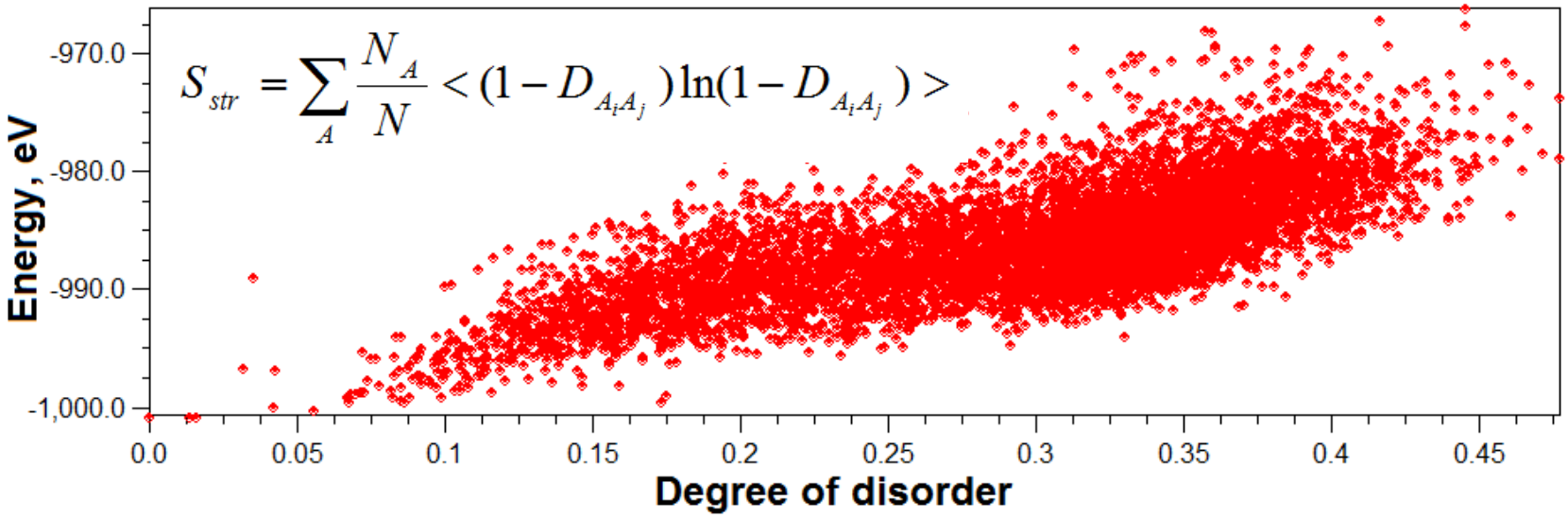
Local degree of order – indicates defects and low-symmetry sites



[ARO & Valle (2009), Lyakhov et al. (2010)]

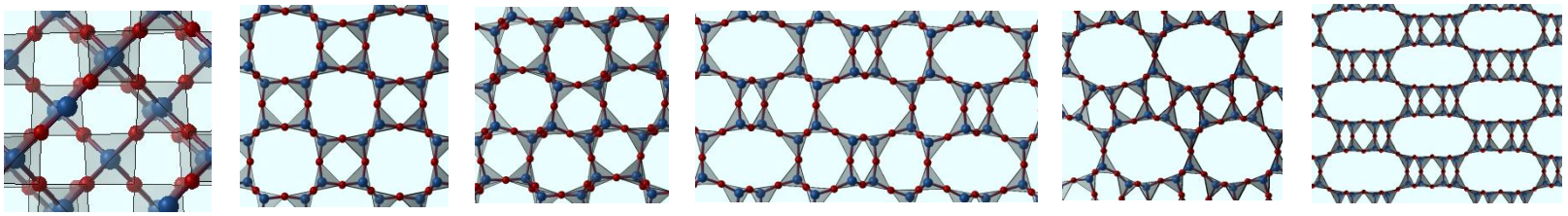
Stability vs complexity.
Pauling‘s 5th rule:

„The number of essential structural elements of stable structures tends to be small“



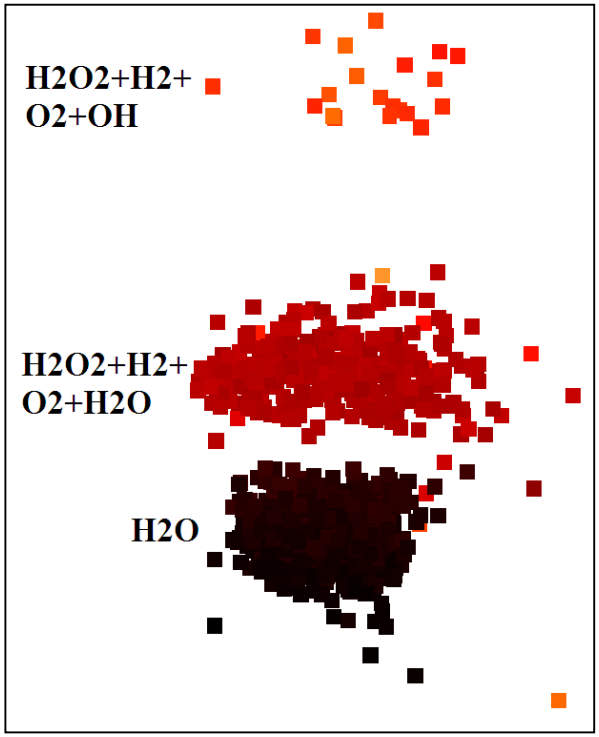
Correlation plot for 6900 structures of SiO_2 with 24 atoms/cell

Some of the (many) remarkable silicate frameworks:

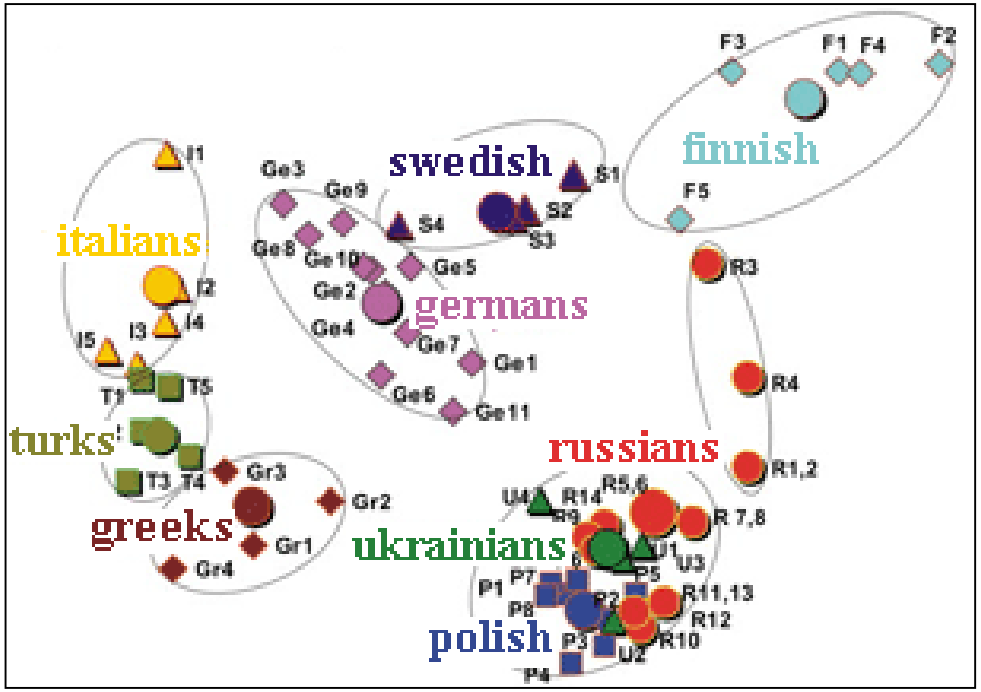


Increasing energy

Grouping structures into similarity classes: quest for more insight in complex systems



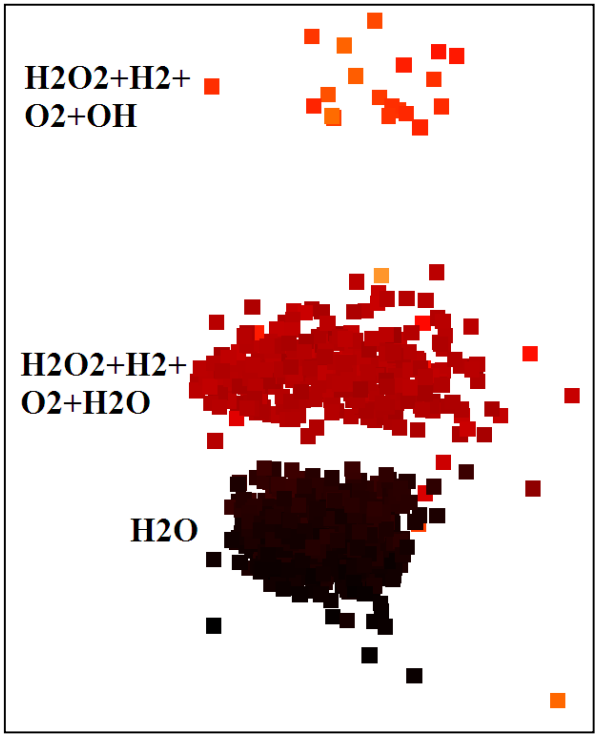
Distance-preserving mapping of crystal structures of H_2O (*darker* – lowest E, *lighter* – higher E).



DNA grouping in Europe

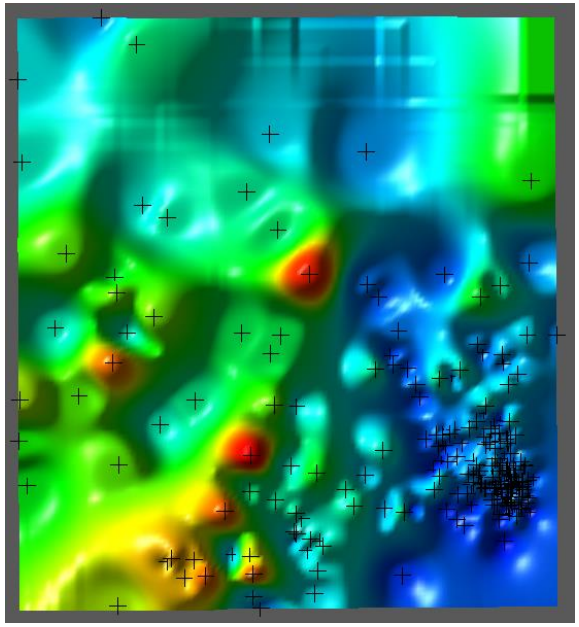
[ARO & Valle, J. Chem. Phys. 130, 104504 (2009)]

Grouping structures into similarity classes: quest for more insight in complex systems

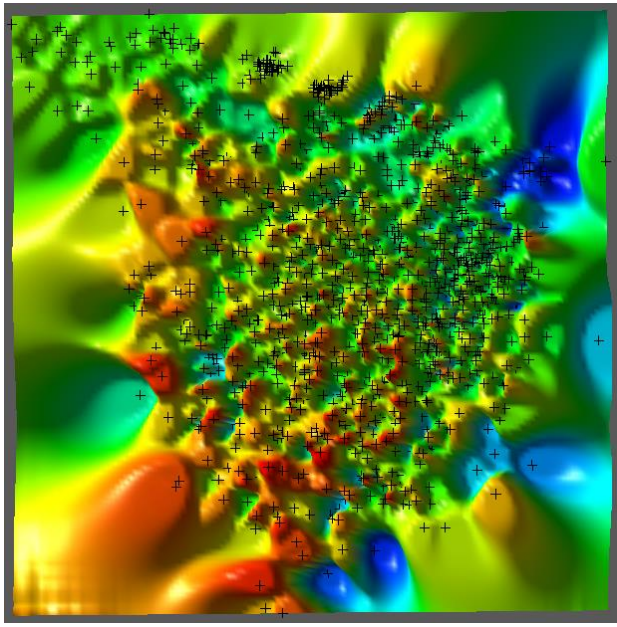


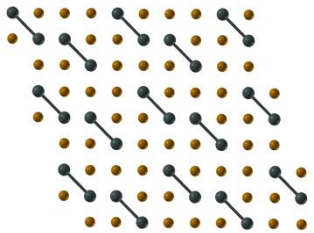
Distance-preserving mapping of crystal structures of H_2O (*darker* – lower E, *lighter* – higher E)

Au_8Pd_4 - simple

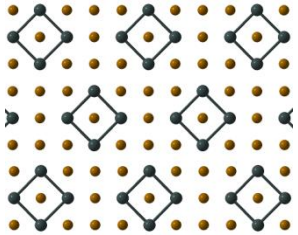


L_4J_8 - complex

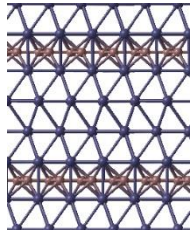




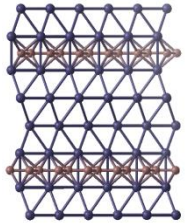
-61.960 eV



-61.957 eV



-99.12ε



-99.05ε

[ARO & Valle, J. Chem. Phys. 130, 104504 (2009)]

Evolution is nature's preferred way of solving complex problems

The analogy between biological and computational evolution is limited and profound at the same time

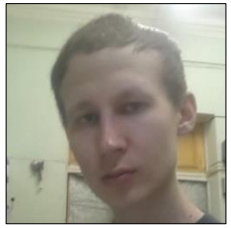
USPEX Computational Materials Discovery

Our team. Where great minds do NOT think alike.

Artem R. Oganov
Professor, head of laboratory

Zahed Allahyari
Postdoc

Pavel Bushlanov
Postdoc



Sergey Lepeshkin
Postdoc

Vladimir Baturin
Postdoc

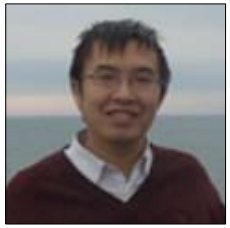
A. Goncharov
Experimental confirmation

Dmitrii Semenok
PhD student

Efim Mazhnik
PhD student

Tao Fan
PhD student

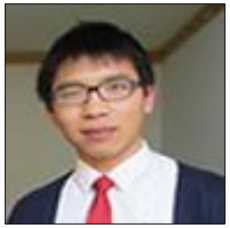
V. Blatov



X. Dong

Q. Zhu

X. F. Zhou



H. Niu

A. Kvashnin

I. Troyan
Experimental confirmation