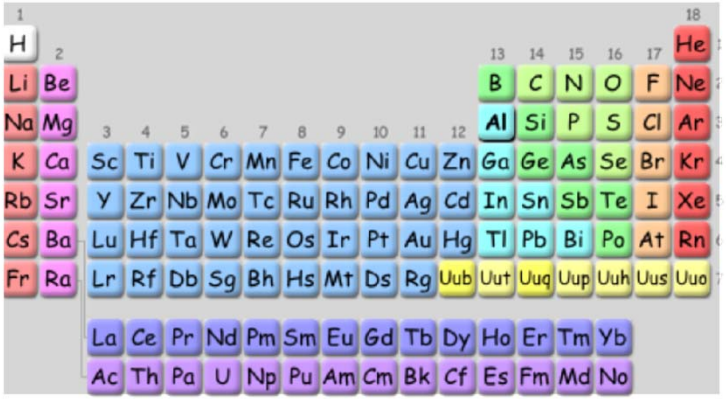
Workshop I: Machine Learning Meets Many-Particle Problems
Part of the Long Program Understanding Many-Particle Systems with Machine Learning
SEPTEMBER 26 - 30, 2016

Real and virtual screening for materials discovery through first principles calculations

Isao Tanaka[1,2,3,4]

[1] Department of Materials Science and Engineering, Kyoto University, JAPAN
[2] Elements Strategy Initiative for Structural Materials, Kyoto University, JAPAN
[3] Center for Materials Research by Information Integration, NIMS, JAPAN
[4] Nanostructure Research Laboratory, Japan Fine Ceramics Center, JAPAN

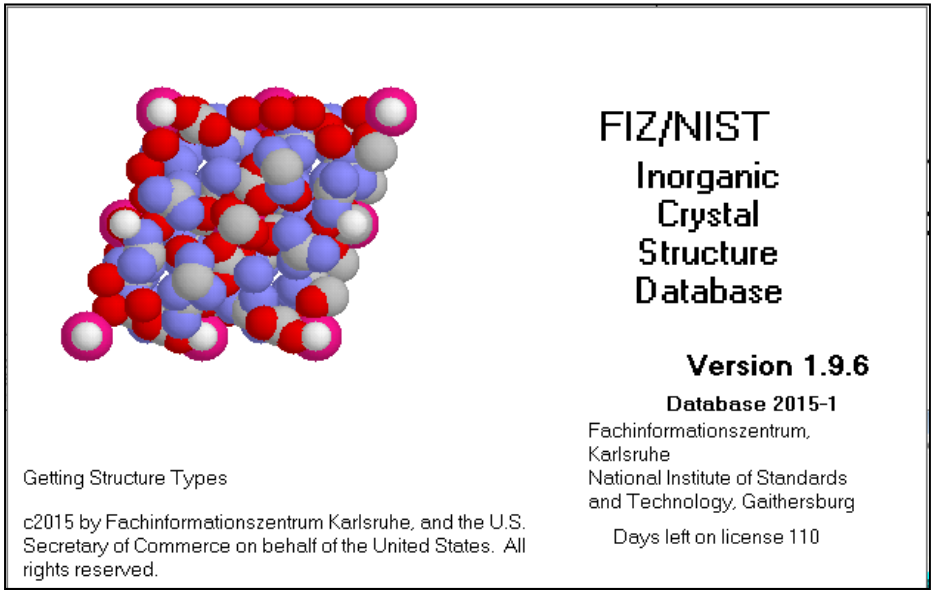
Enormous number of chemical combinations



Periodic Table

Number of chemical elements	Number of chemical combinations (only for simple composition ratio)
1	~100
2	~100,000
3	~10,000,000
4	~1,000,000,000

Inorganic Crystal Structure Database (ICSD)



World largest database
for known inorganic crystals.

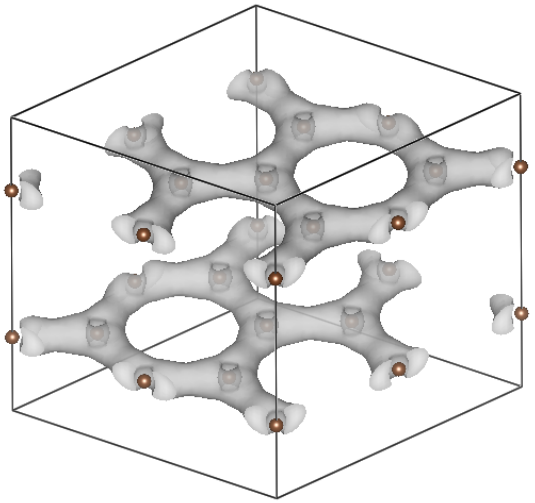
177,000 crystal structures
→ 40,000 structures
excluding duplicates,
incompletes, etc.

Many systems are yet-unexplored !

First Principles Calculations

Useful to fill the gap between enormous number of chemical combinations and experimental database.

Electronic structure calculations entirely based on quantum theory without using empirical parameters.



Total energy

Force/Stress

Electronic/Magnetic structures

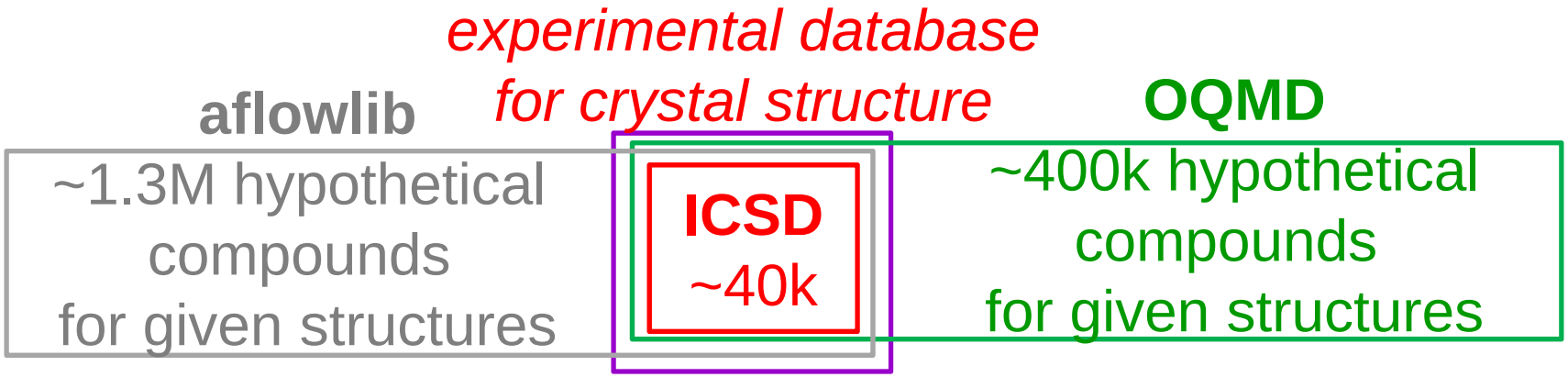
Physical properties

. . . .

***But this is not a magic box* !!**

Vast chemistry space to explore

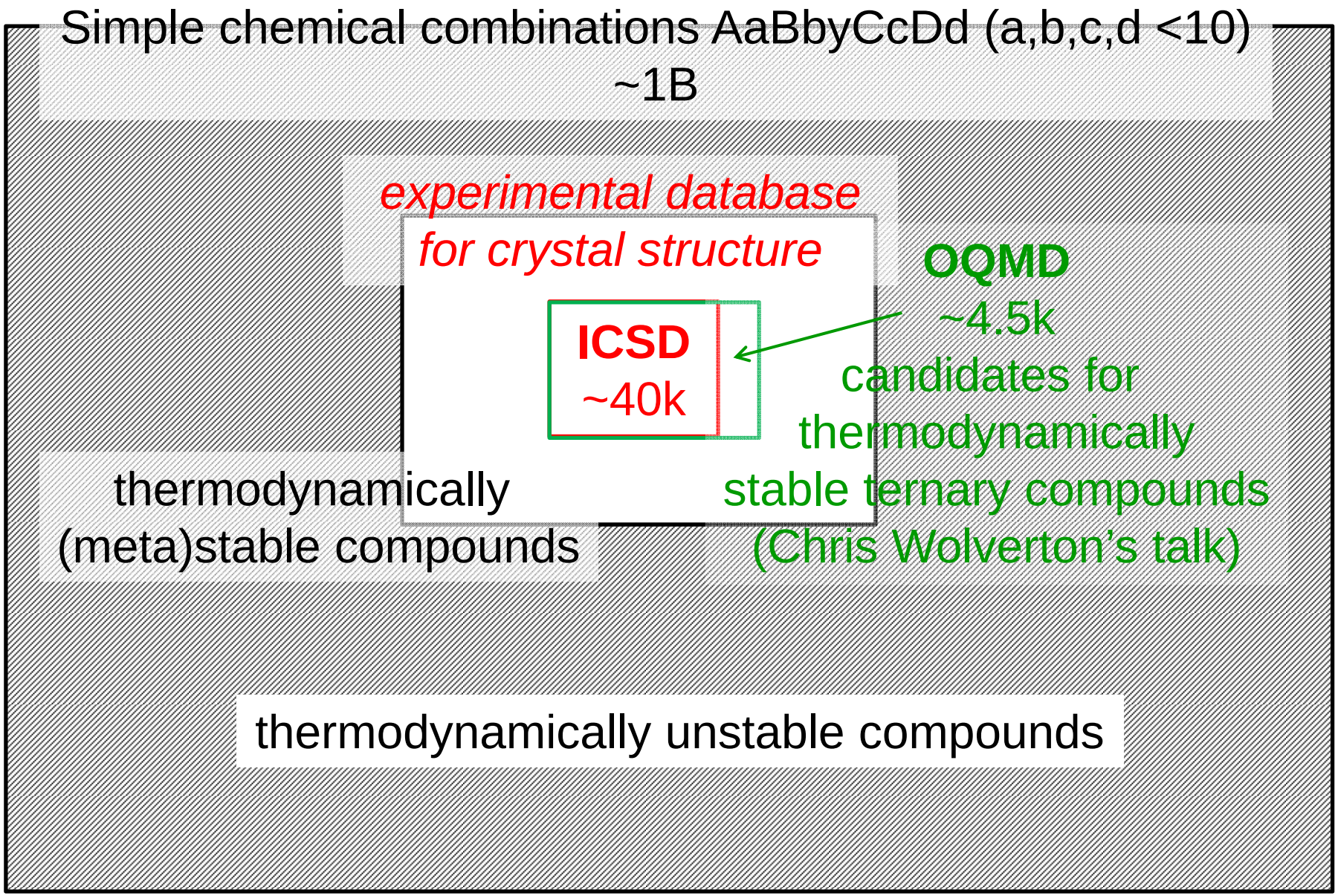
Simple chemical combinations AaBbyCcDd (a,b,c,d <10)
~1B



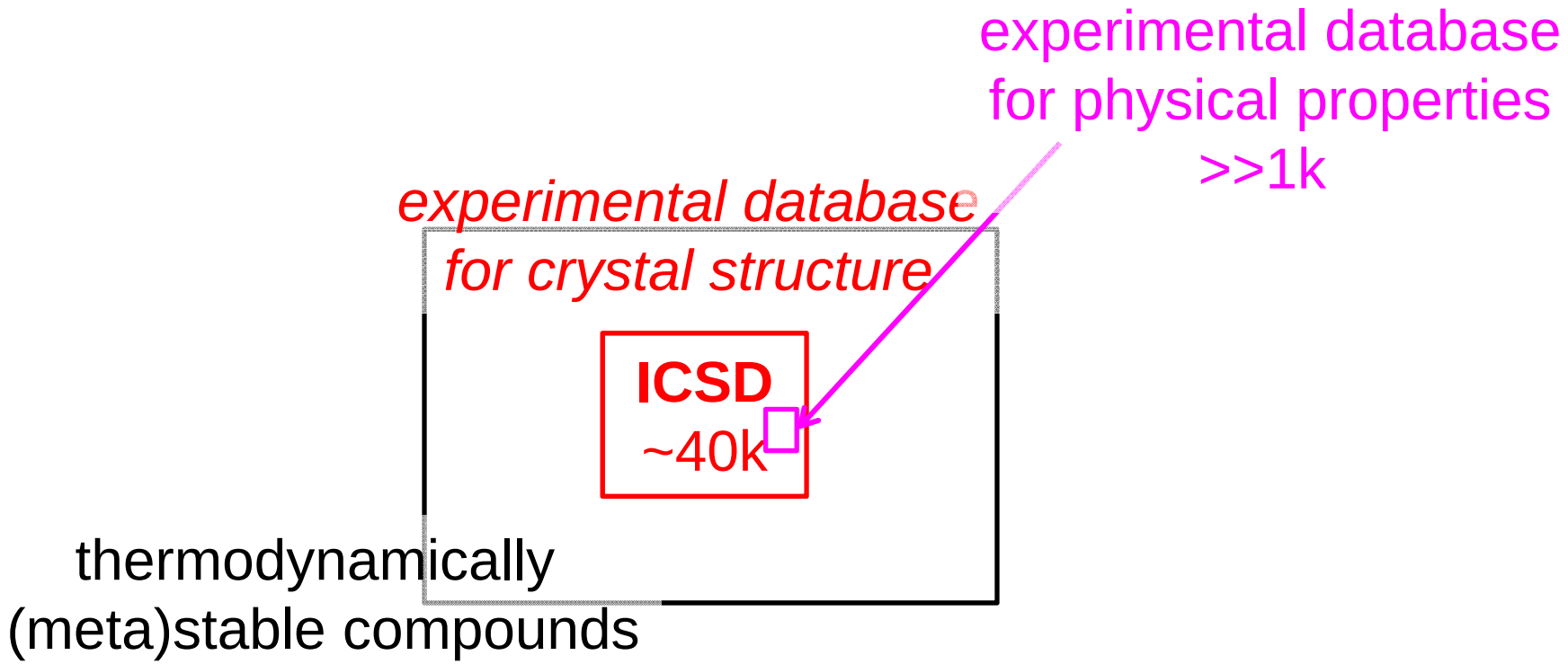
IMPORTANT!

First principles calculation requires an initial “guess” of structure,
otherwise it becomes too expensive to explore the vast space.

Vast chemistry space to explore



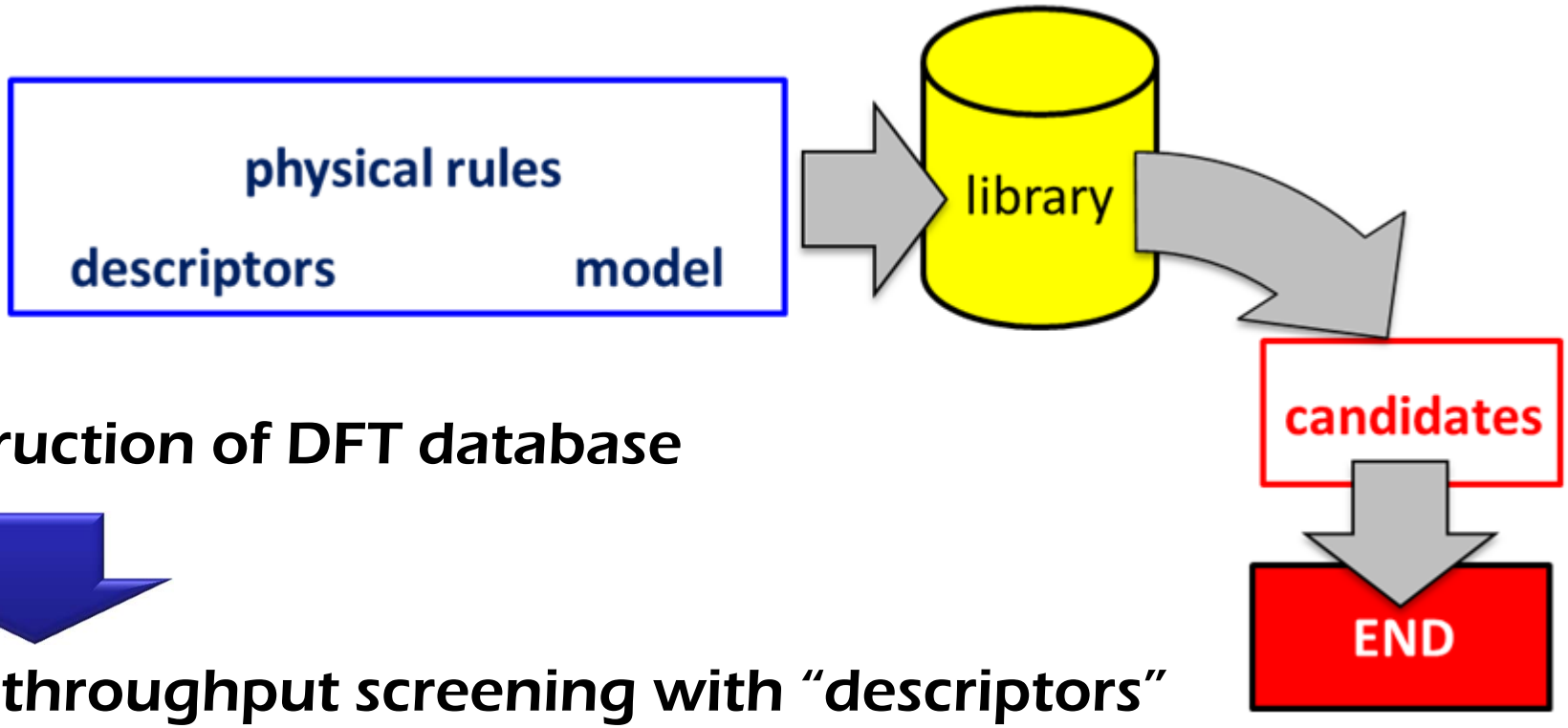
Vast chemistry space to explore



- ✓ **Systematic method to add as-yet-unknown & stable compounds to the inorganics database is desired.**
- ✓ **It is still worthy to explore the space of known compounds.**

High throughput screening 1

"Real Screening"



Construction of DFT database

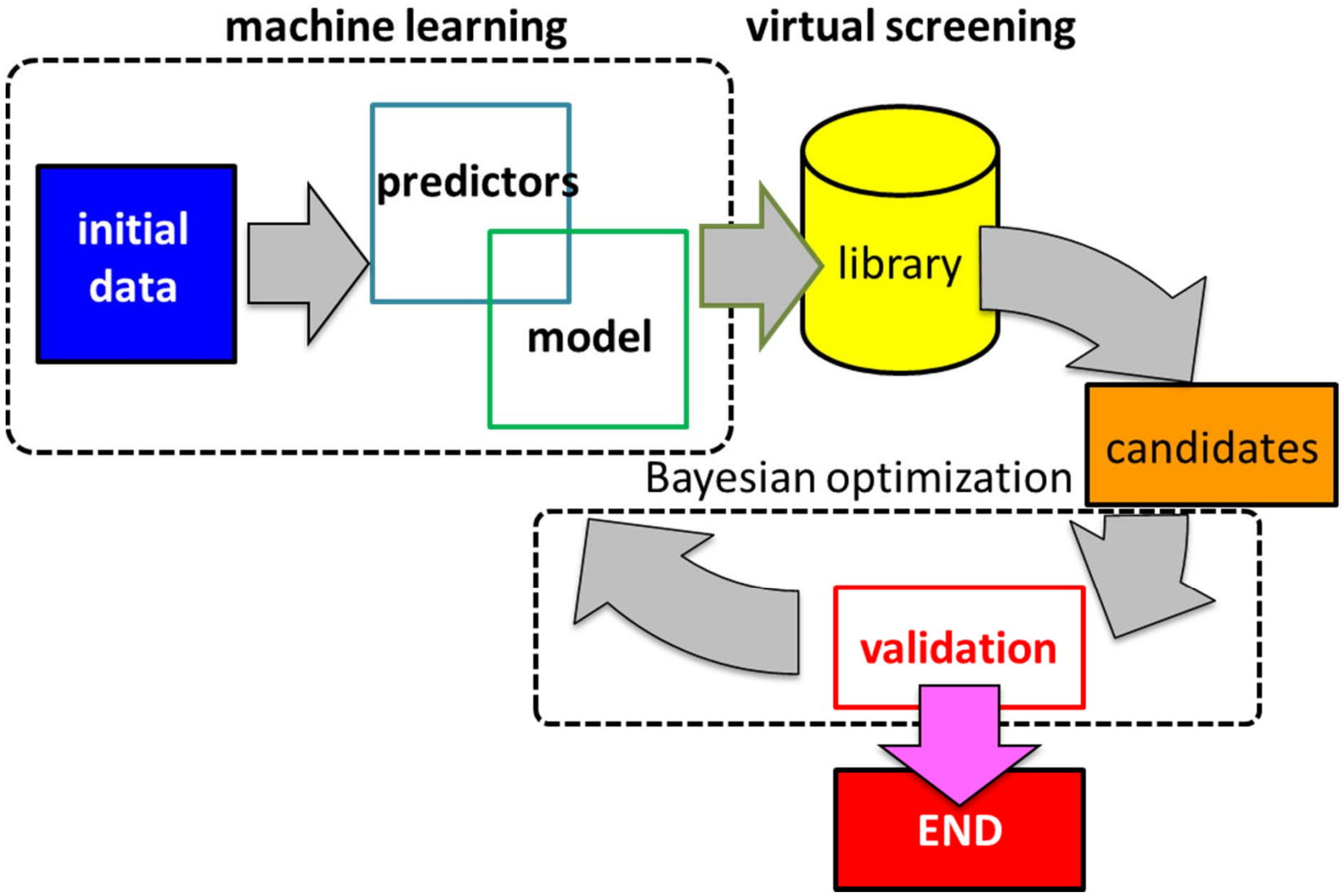
High throughput screening with "descriptors"

Experts' knowledge

Physical Rule

High throughput screening 2

"Virtual Screening"



Contents

Materials discovery through first principles calculations

✓**Discovery of a new Sn(II)-based oxide for daylight-driven photocatalyst**

REAL SCREENING

✓**Discovery of new low thermal conductivity materials from Mat. Proj. database (55,000 crystals)**

VIRTUAL SCREENING

Accelerated discovery of a novel Sn(II)-based oxide for daylight-driven photocatalyst

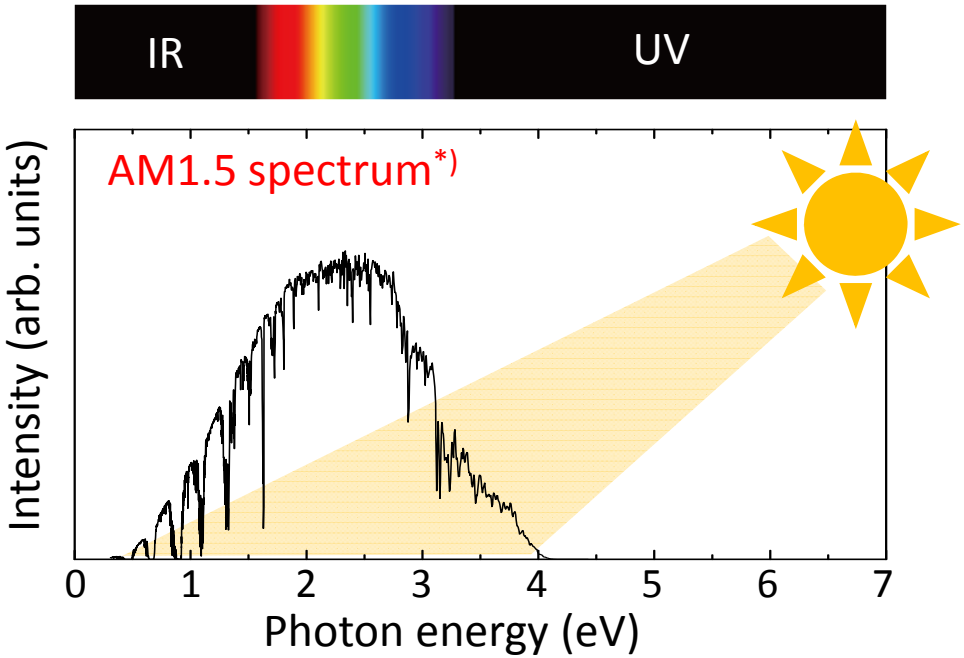
Hiroyuki Hayashi, Shota Katayama, Takahiro Komura, Yoyo Hinuma, Tomoyasu Yokoyama, Kou Mibu, Fumiyasu Oba and IT

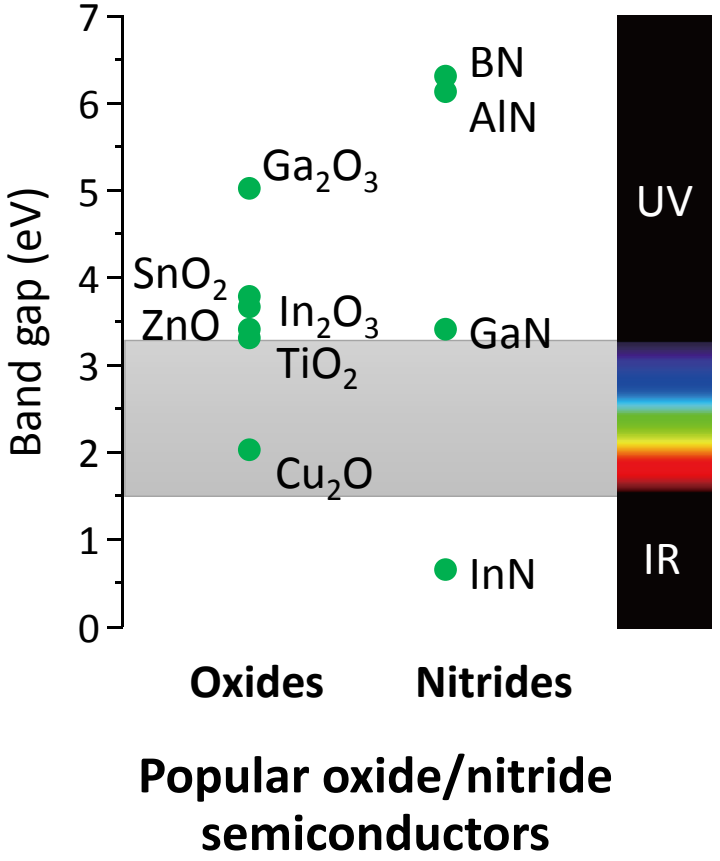
Hiroyuki Hayashi

Advanced Science 9, (2016). 1600246

Sustainable energy: Sunlight

- Solar cells & daylight-driven photocatalysts
  - Target band-gap range
    Solar cell absorber: 1.2 – 2.0 eV
    Photocatalyst: 1.5 – 3.0 eV



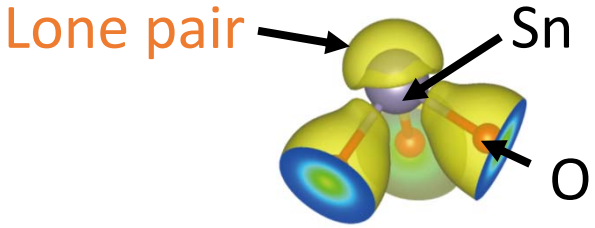


Popular oxide/nitride semiconductors

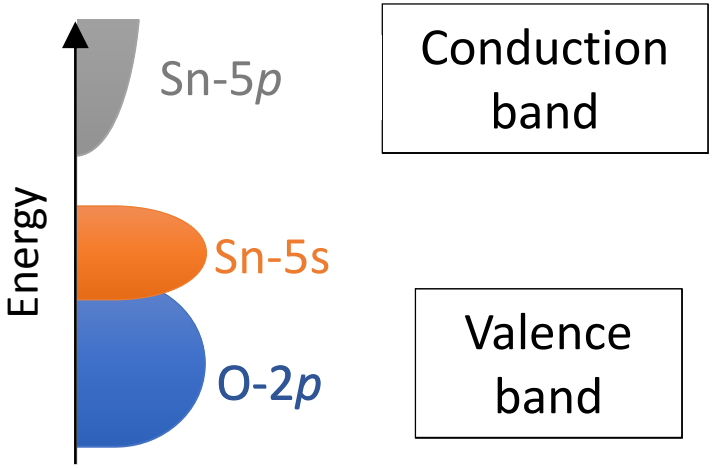
*) http://rredc.nrel.gov/solar/spectra/

Lone pair materials: Sn(II), Pb(II) & Bi(III)

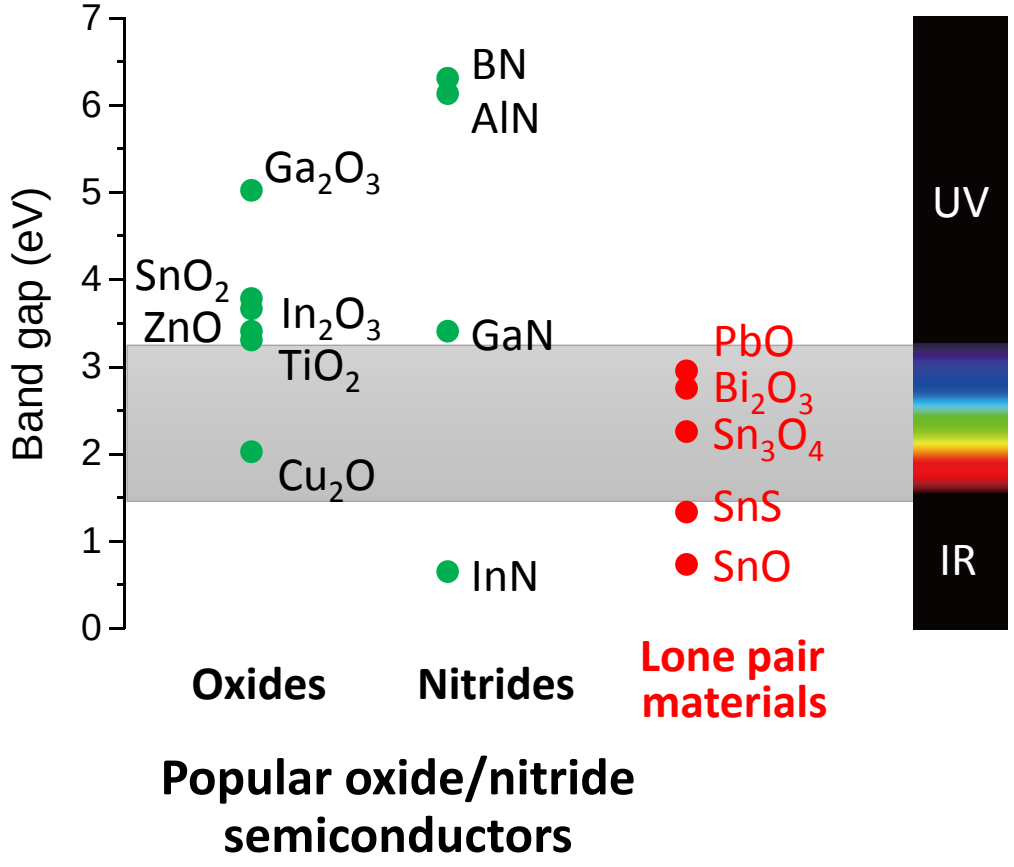
- Characteristic electronic structures



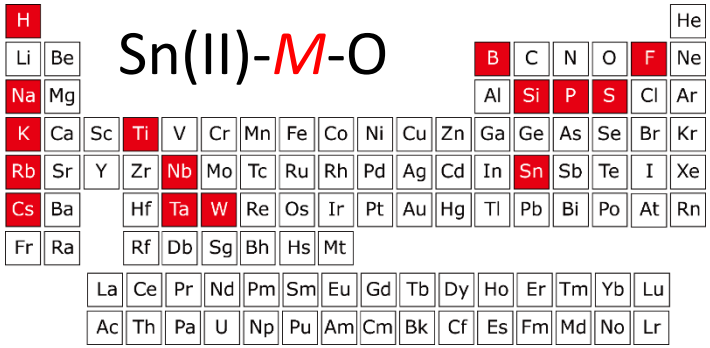
- Narrower band gap
- Wider valence band

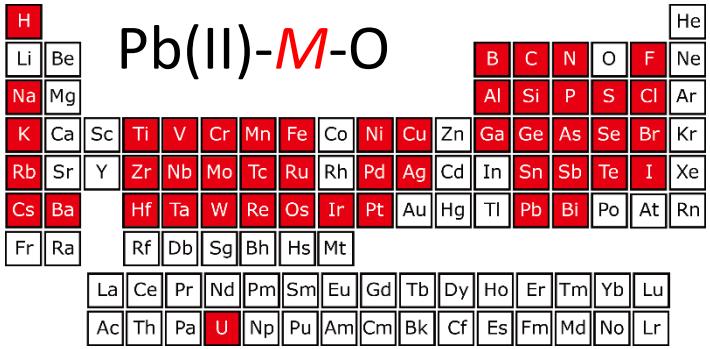


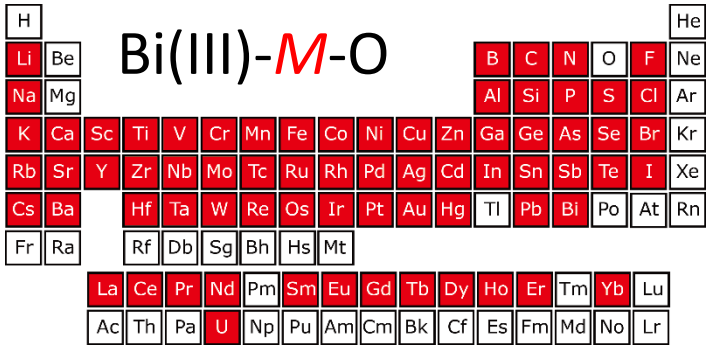
Sn(II) oxides



Known oxides: Sn(II), Pb(II) & Bi(III)



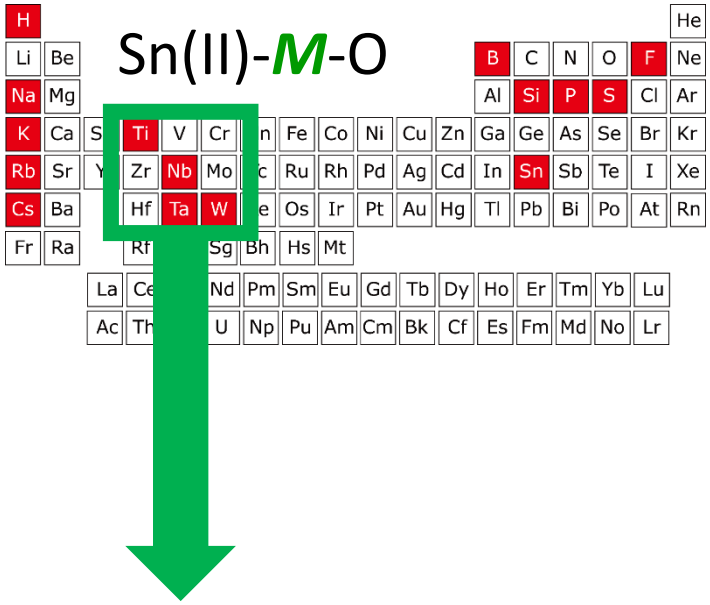




System	Compounds registered in ICSD
Sn(II)-*M*-O	47
Pb(II)-*M*-O	210
Bi(III)-*M*-O	213

Sn(II) oxides are yet-unexplored !

Chemical composition of interest



4A – 6A transition metal compounds
- Commonly used for photocatalysts
  ex. TiO_2, WO_3, $NaTaO_3$, TaON, …
- Wide band gaps

Sn(II) oxides
- Relatively narrow band gaps

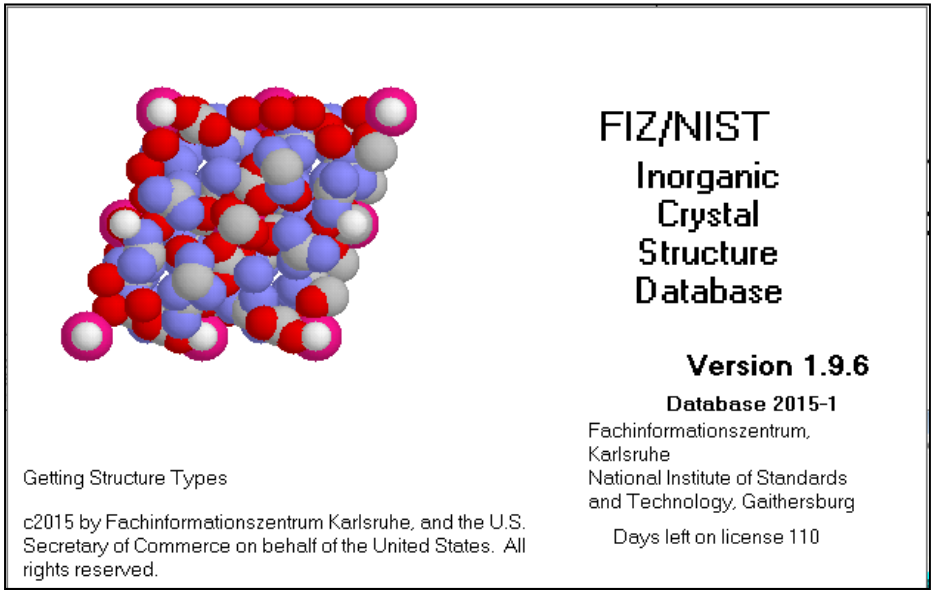
SnO-$MO_{q/2}$ pseudo binary systems

q	*M*	**Known Compounds**
4	Ti, Zr, Hf	$SnTiO_3$, Sn_2TiO_4
5	V, Nb, Ta	$SnNb_2O_6$, $Sn_2Nb_2O_7$, $SnTa_2O_6$, $Sn_2Ta_2O_7$, $SnTa_4O_{11}$
6	Cr, Mo, W	$SnWO_4$, Sn_2WO_5, Sn_3WO_6

Only 10 oxides are known.

Reported as good photocatalyst.

Inorganic Crystal Structure Database (ICSD)



World largest database
for known inorganic crystals.

177,000 crystal structures
→ 40,000 structures
excluding duplicates,
incompletes, etc.

9,100 structure prototypes
(e.g. rock-salt, perovskite, ...)

Number of chemical elements	Number of structure prototypes in ICSD
1	120
2	1,700
3	4,700
4	4,300

Crystal structures

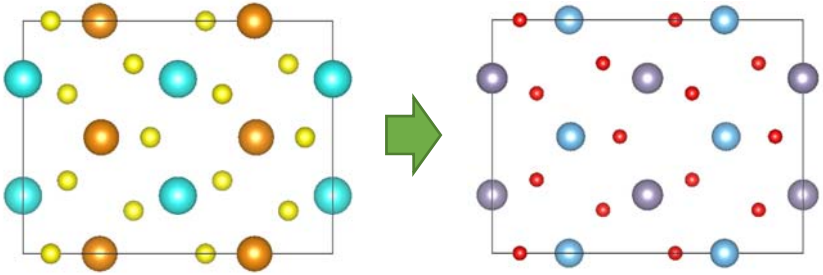
Prototypes of A(II)-*M*(*q*+)-oxides

q	Number of prototype structures in ICSD	Examples
4	586	Pb_3O_4, $AlFeO_3$, $NdYbS_3$, …
5	207	K_2ZrF_6, Bi_2WO_6, …
6	368	$CrUO_4$, β-$SnWO_4$, …

Total : 1,161 crystal structures

SnO-$MO_{q/2}$ pseudo binary systems

$NdYbS_3$ type $NdYbS_3$ $NdYbS_3$ type $SnTiO_3$



Total: 1,163x3 = **3,483** candidate oxides

Screening procedure

1. Candidates generation : DFT calculations
   - 3,483 candidates

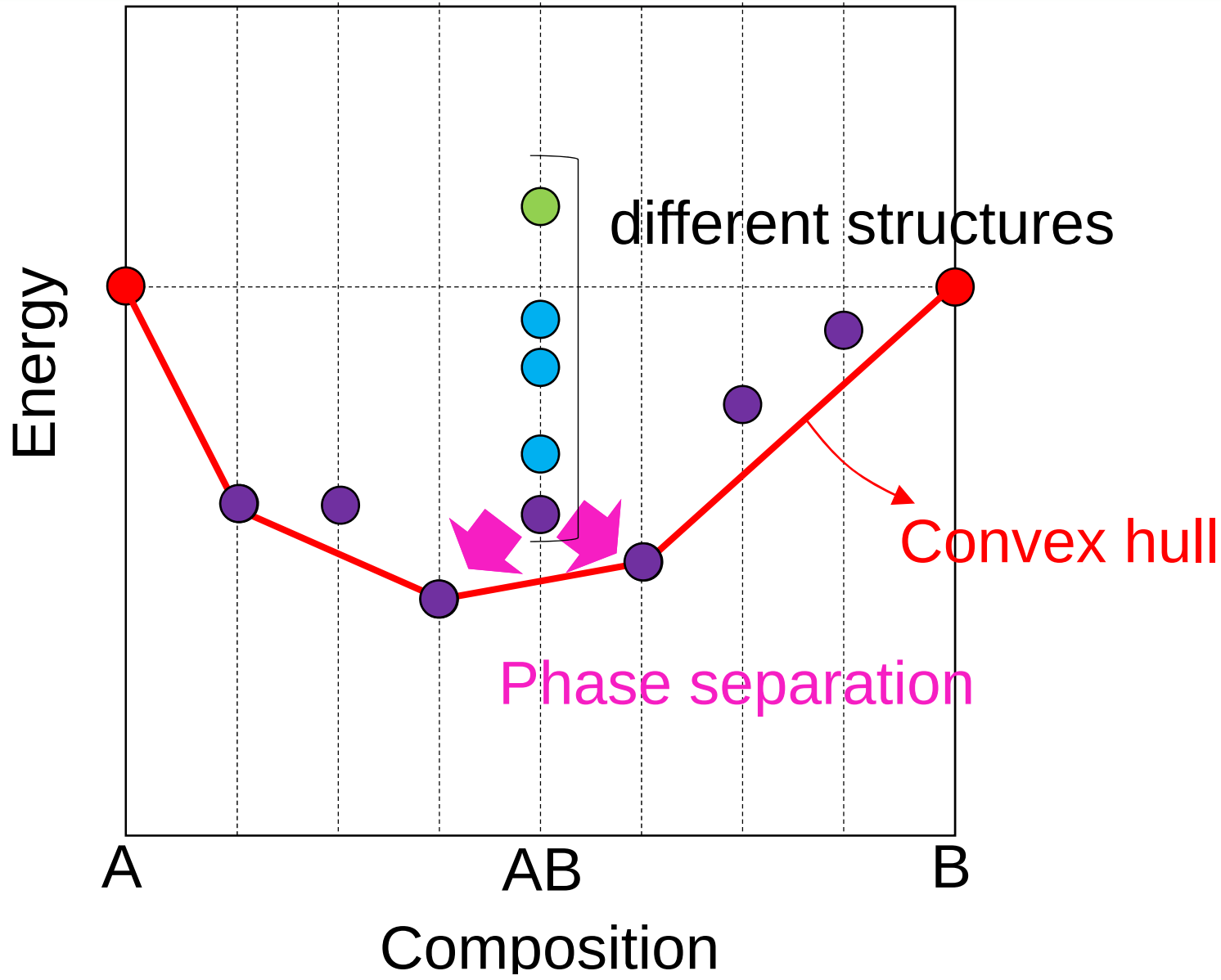
2. Phase stability
   - Formation energy vs. SnO and $MO_{q/2}$

$$E_{form} = E_{\mathrm{Sn}_x M_y \mathrm{O}_{x+yq/2}} - (xE_{\mathrm{SnO}} + yE_{MO_{q/2}})$$

3. Band gap (E_g) by GGA(PBE)
   - Criterion: $E_g > 2$ eV

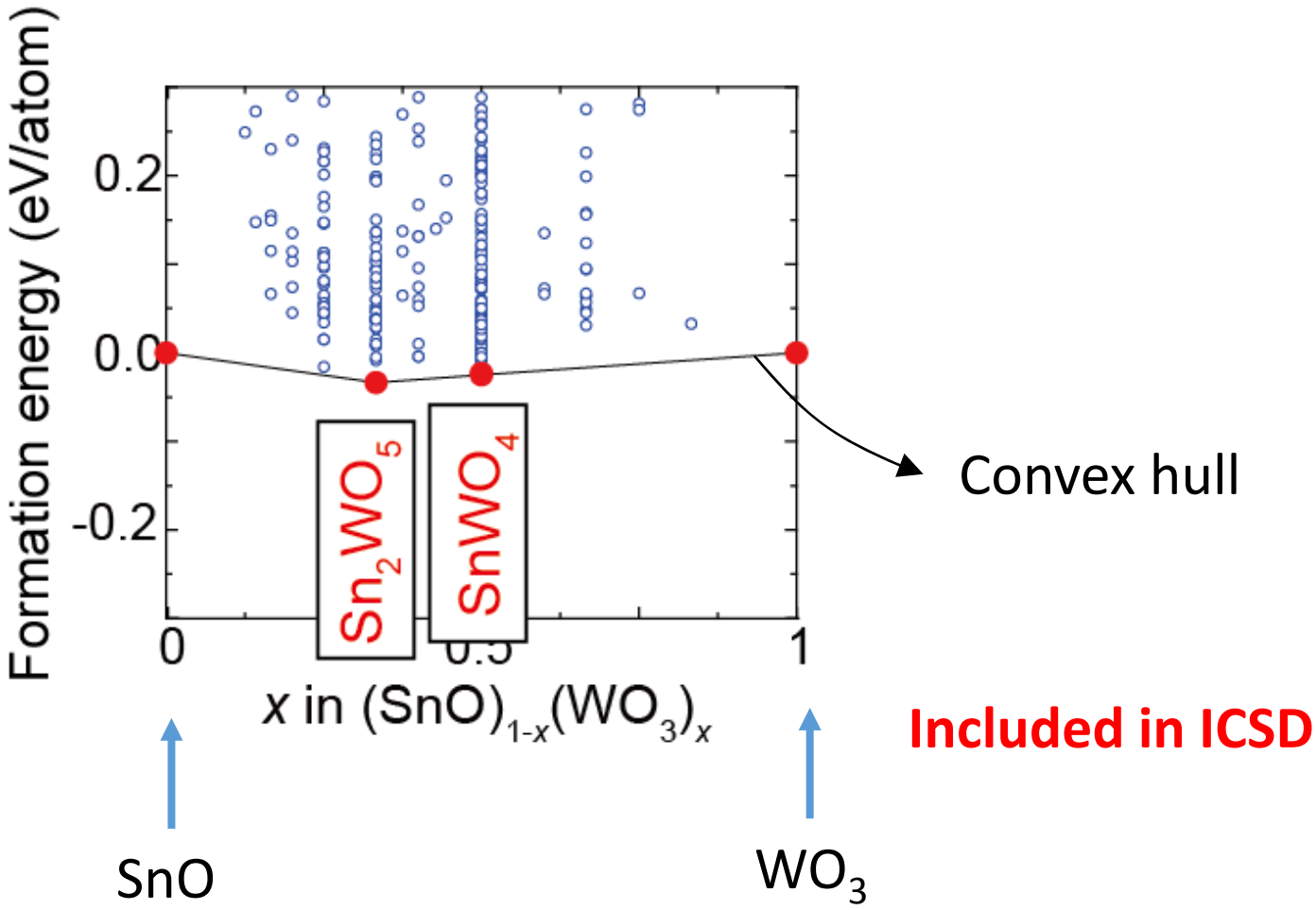
4. Band edge position (ε_{CBM} and ε_{VBM}) by GGA(PBE)
   - Criterion: $\varepsilon_{CBM} > H^+/H_2$ level and $\varepsilon_{VBM} < O_2/H_2O$ level

Phase stability



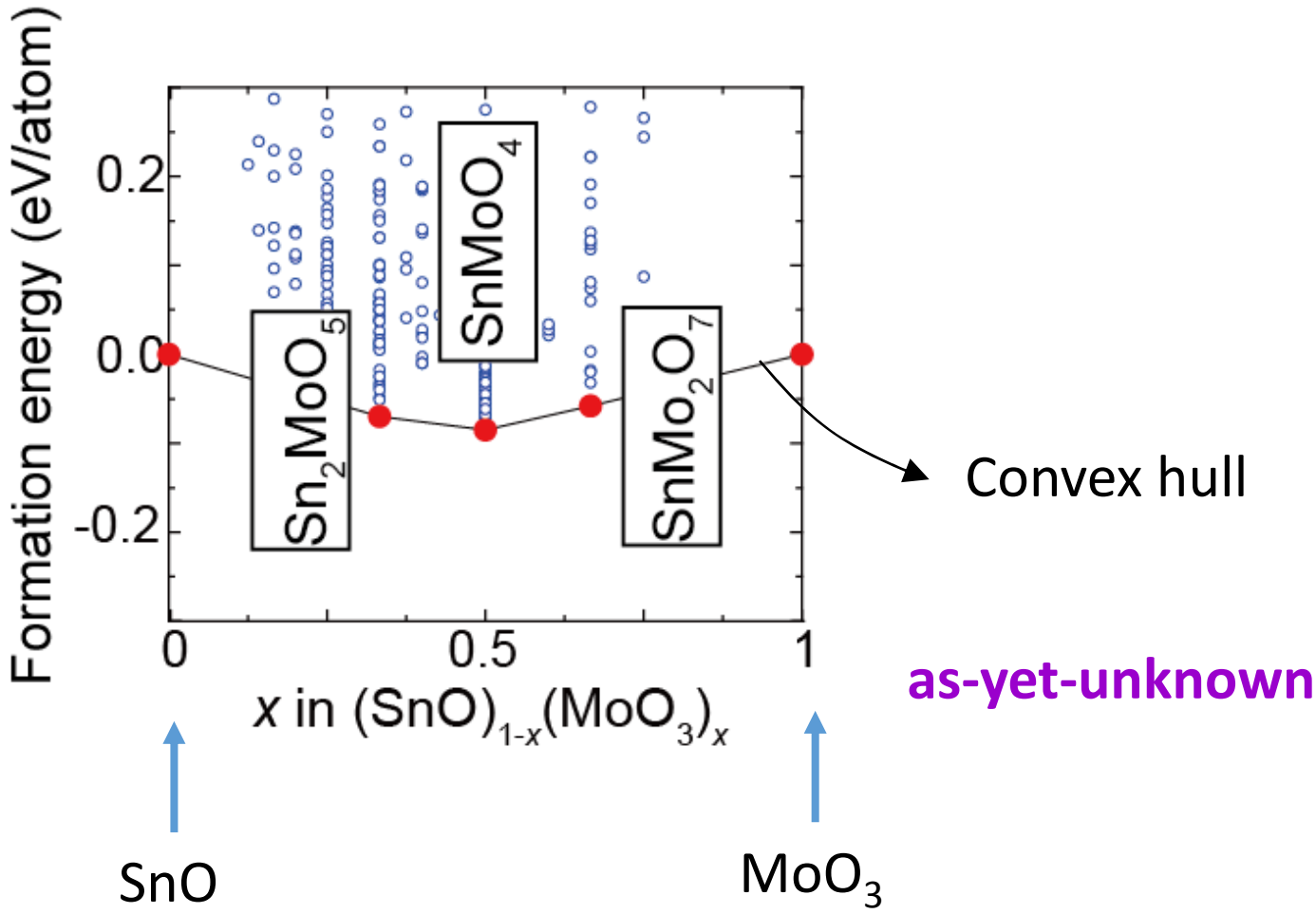
Formation energy

- SnO-WO_3 pseudo binary system



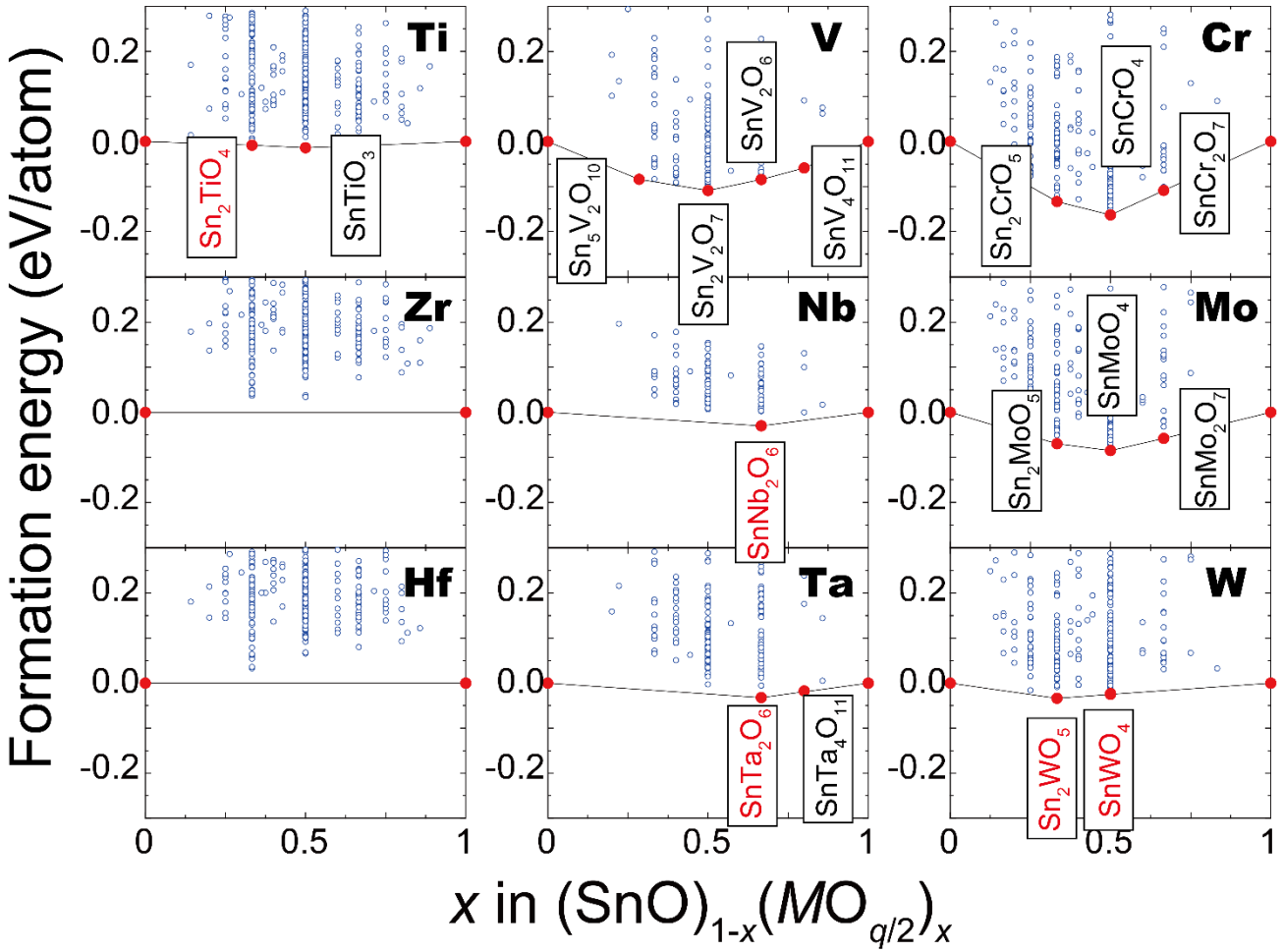
Formation energy

- SnO-MoO_3 pseudo binary system



Formation energy

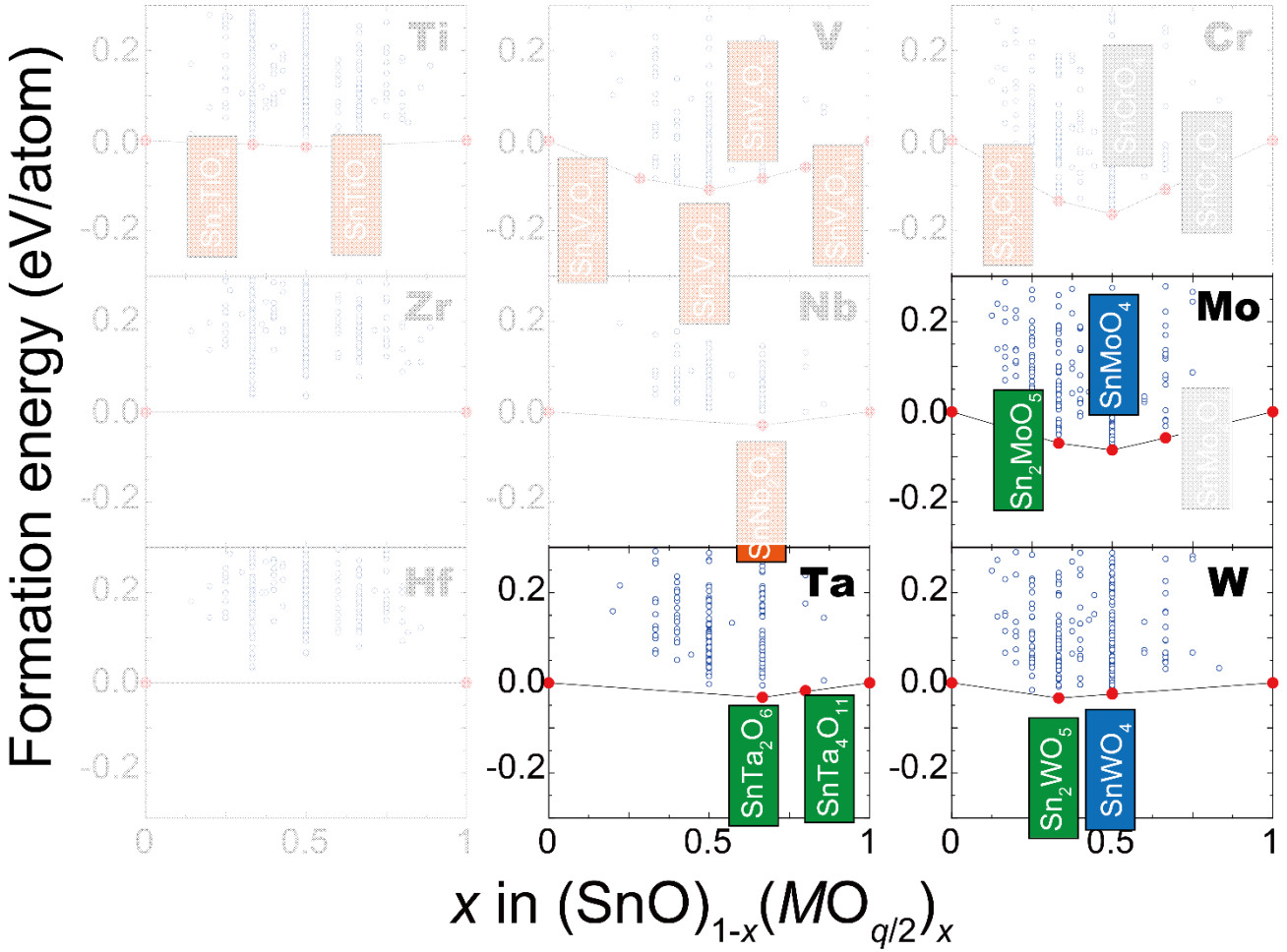
- Convex hull of SnO-$MO_{q/2}$ pseudo binary systems



Reported oxides in ICSD (Red characters) are located on the convex hull.

Band gap

- Band gap of actual photocatalysts ≥ 2 eV (GGA)

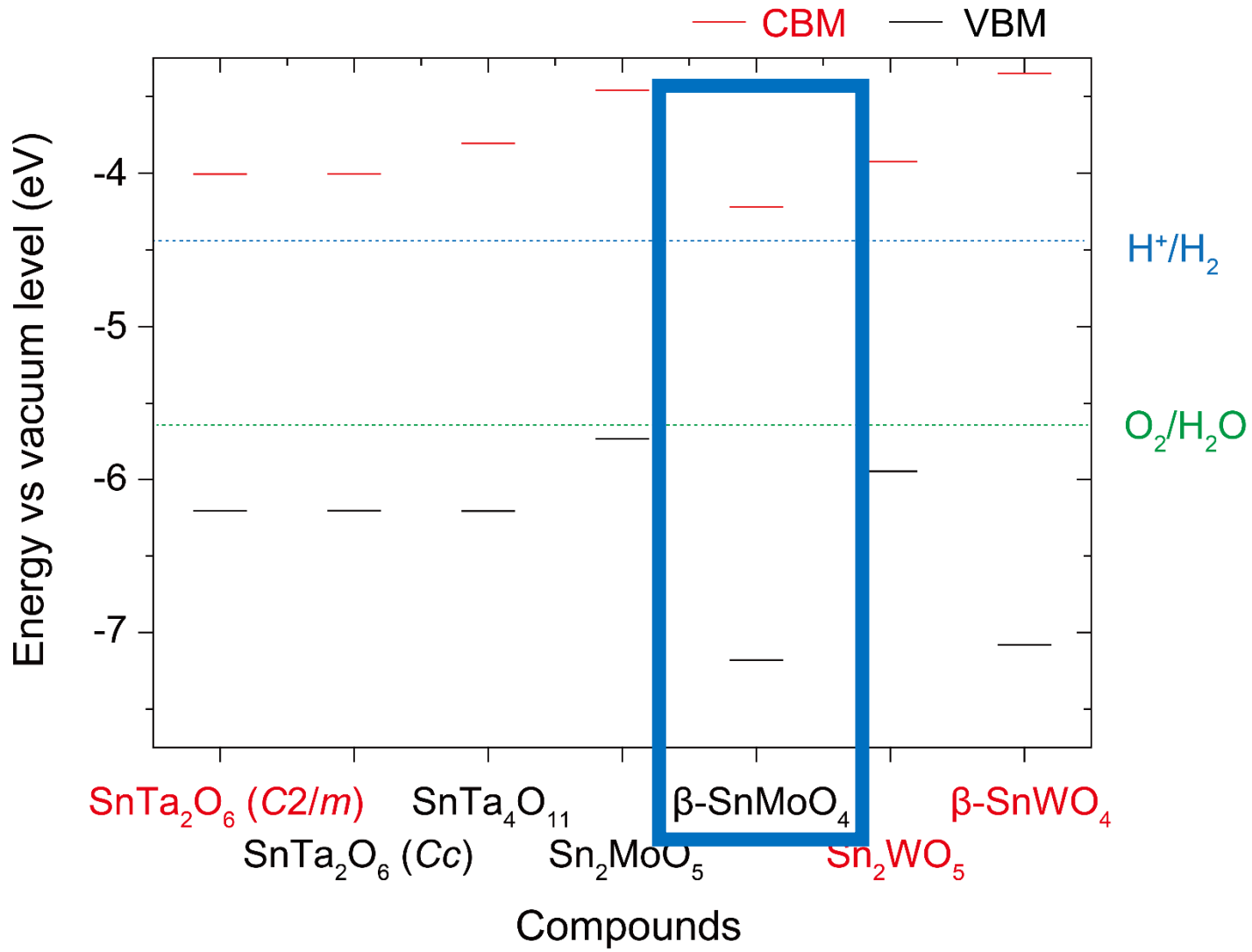


Band gap

- 0 ~ 1 eV
- 1 ~ 2 eV
- 2 ~ 3 eV
- over 3 eV

- $SnO\text{-}Ta_2O_5$
- $SnO\text{-}WO_3$
- $SnO\text{-}MoO_3$

Band edge position



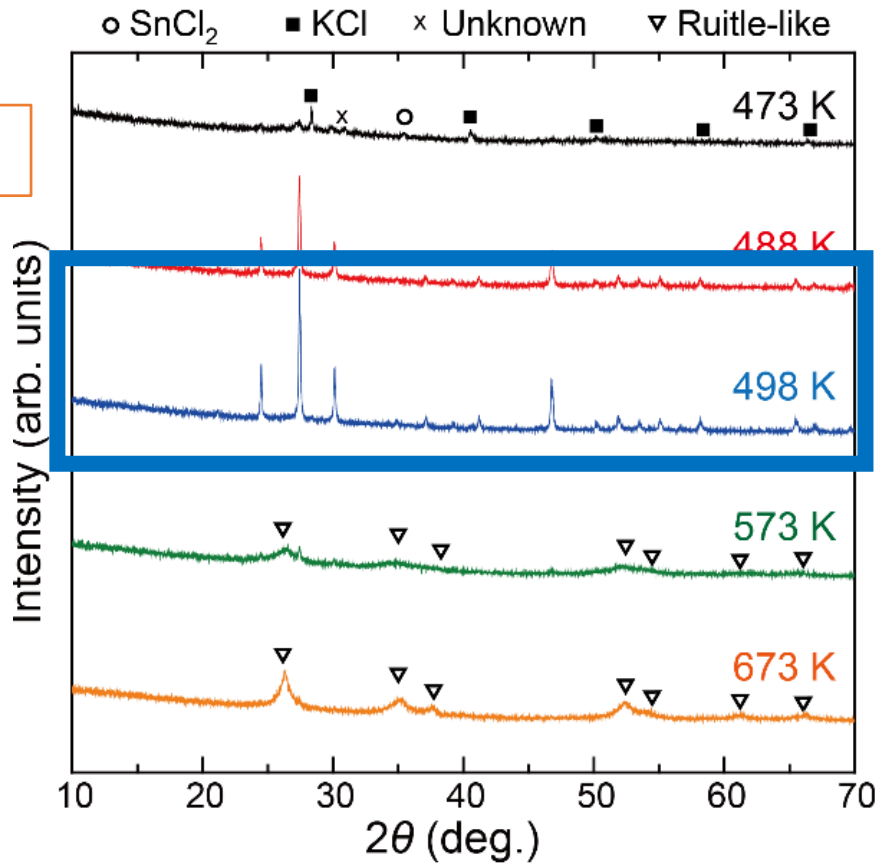
Experimental results

- Synthesis of β-$SnMoO_4$

Mixture of $SnCl_2$ and K_2MoO_4 powders

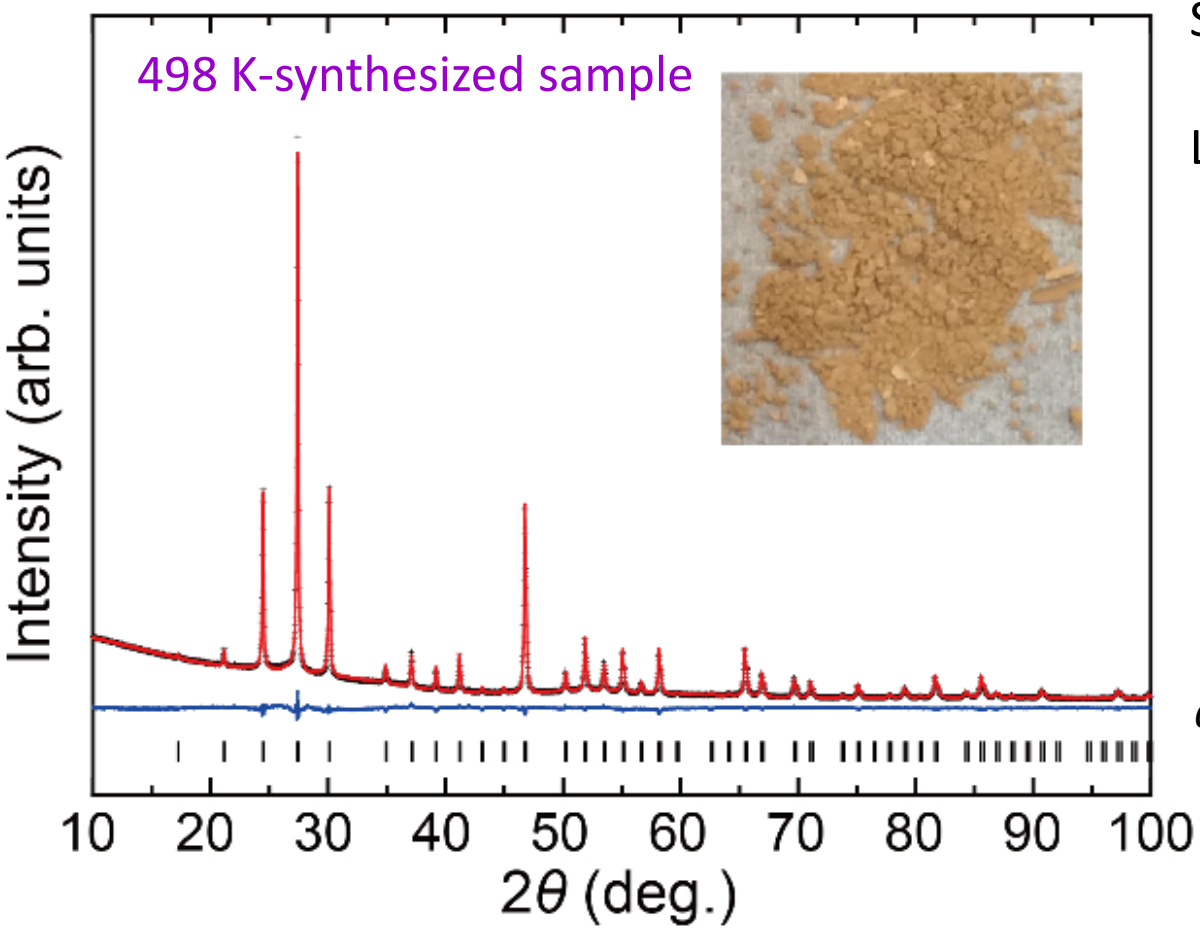
1 hour annealing in Ar gas

Washed and dried



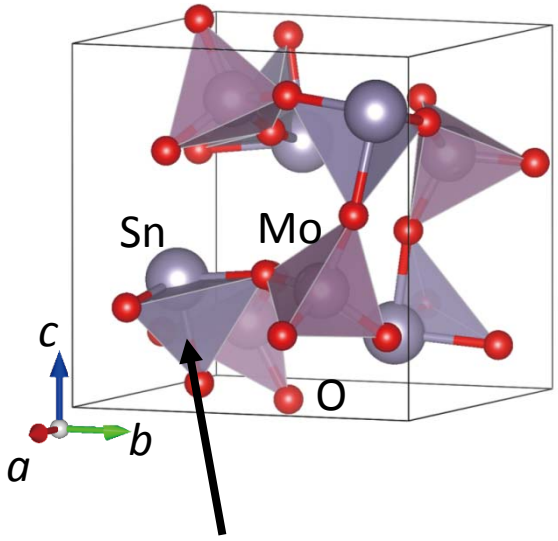
Crystal structure of $SnMoO_4$

- Newly discovered compound



Space group type: $P2_13$ (Cubic)

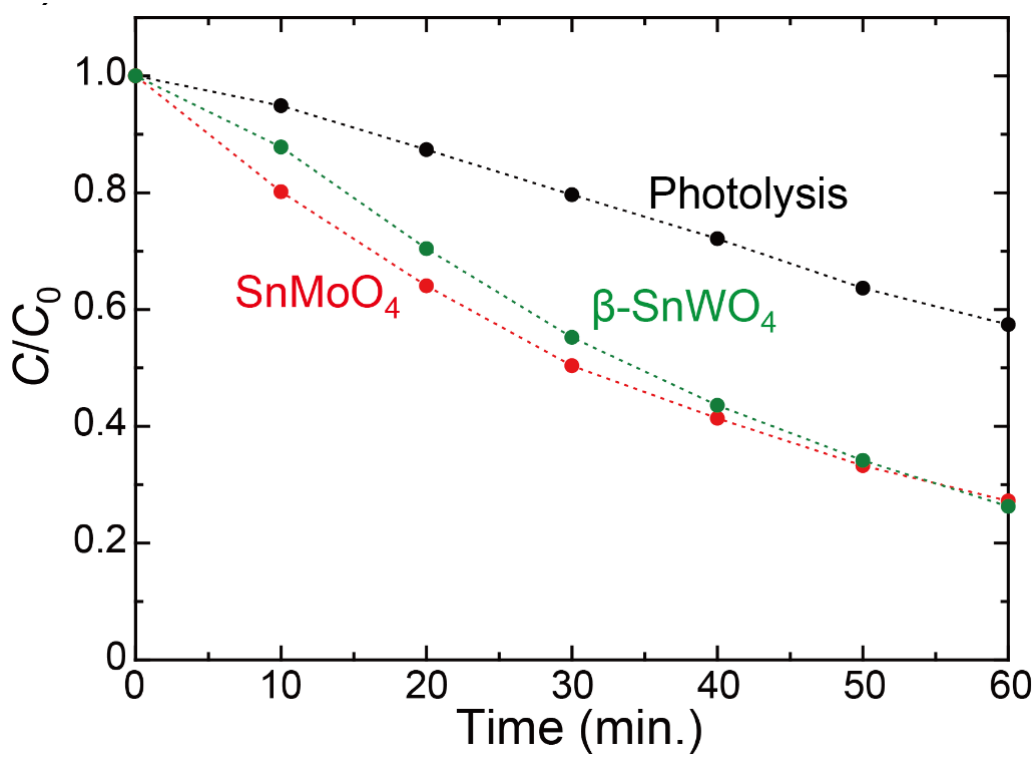
Lattice constant: a = 7.26 Å



Trigonal prism which is characteristic of Sn(II)

Photocatalytic activity of $SnMoO_4$

- Degradation of methylene blue under simulated day-light



Newly-discovered $SnMoO_4$ powder exhibits clear photocatalytic activity.

Discovery of new low thermal conductivity materials

First-Principles Anharmonic Lattice-Dynamics Calculations and Bayesian Optimization



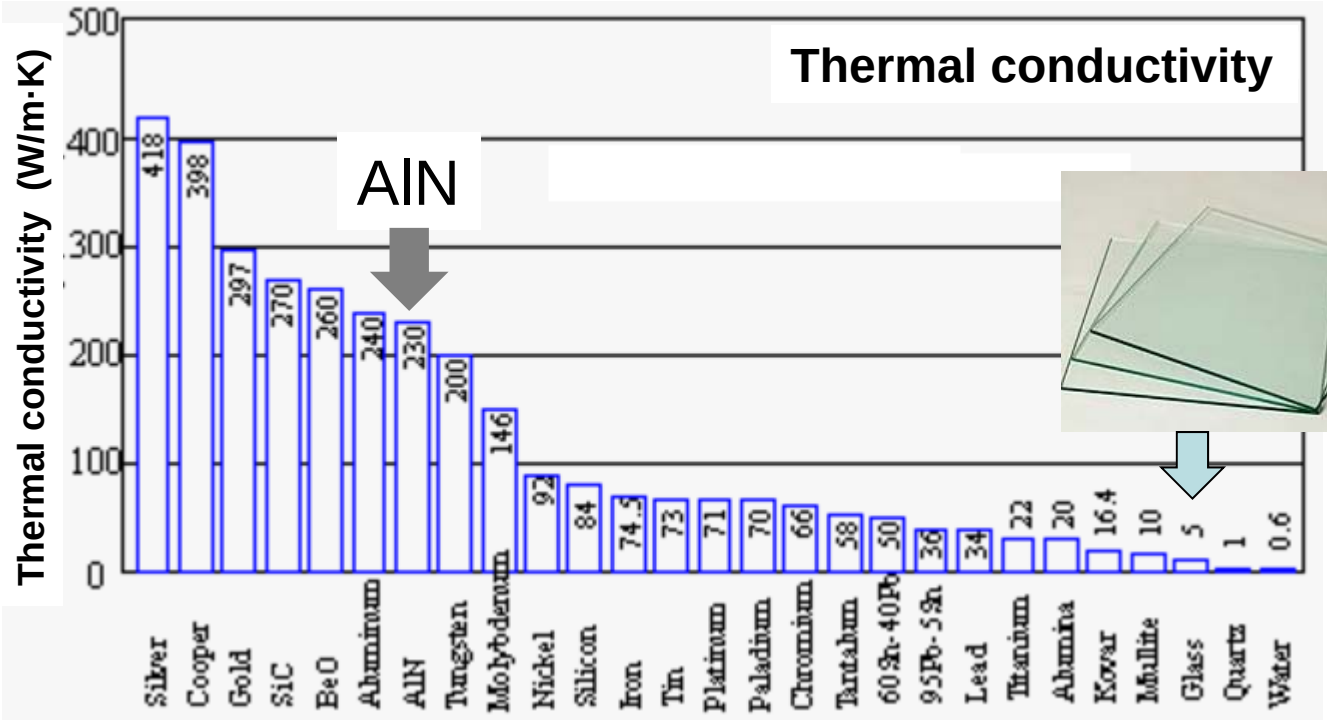
Atsuto Seko, Atsushi Togo, Hiroyuki Hayashi,

Koji Tsuda, Laurent Chaput, and IT

Find ultra-low thermal conductivity materials of 0.1W/m·K level in Mat. Proj. database (55,000 crystals)

diamond

>2000



Background

Thermal Conductivity (TC)

$$\kappa = \kappa_{electonic} + \underline{\kappa_{lattice}}$$

Lattice Thermal Conductivity (LTC)

- ✓ **Reliable experimental dataset is limited (< 100 crystals).**
- ✓ **Reliable first principles calculations are very expensive.**

  (CC5 class : 1 day/100 cores for 1 simple crystal.)
- ✓ **Little knowledge to predict LTC deductively.**

  (Simple rule to determine LTC is not clear.)

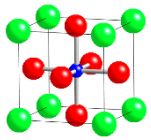
Materials search has been made through "modification" of known compounds showing high/low LTC.

Cost for First Principles Calculations

Computational Cost with 1 core

CC0	1min	Single shot GGA
CC1		Structure opt.
	1hour	
CC2		2x2x2 supercell Single shot GGA
CC3	1day	Phonon frequency
CC4	1 week	Lattice thermal conductivity
CC5		
	1 year	Diffusivity by FPMD simulation

1000 atoms cell Single shot GGA



Computational Cost with 100 cores

1min	
1hour	
CC5	1day
CC6	1 week
CC7	
	1 year

10 nodes

Thermoelectric materials

Essential for utilizing otherwise waste heat.

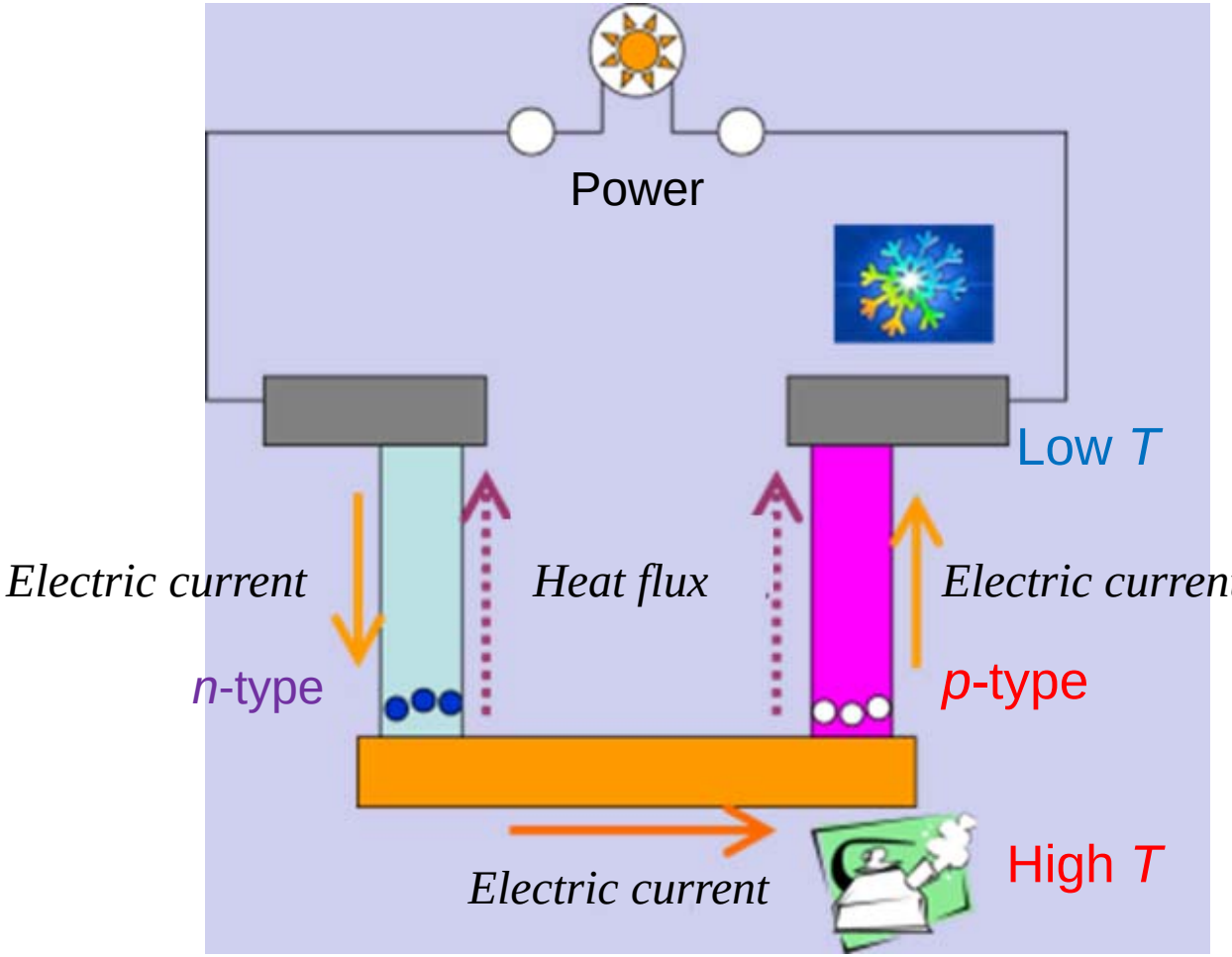


Figure of Merit

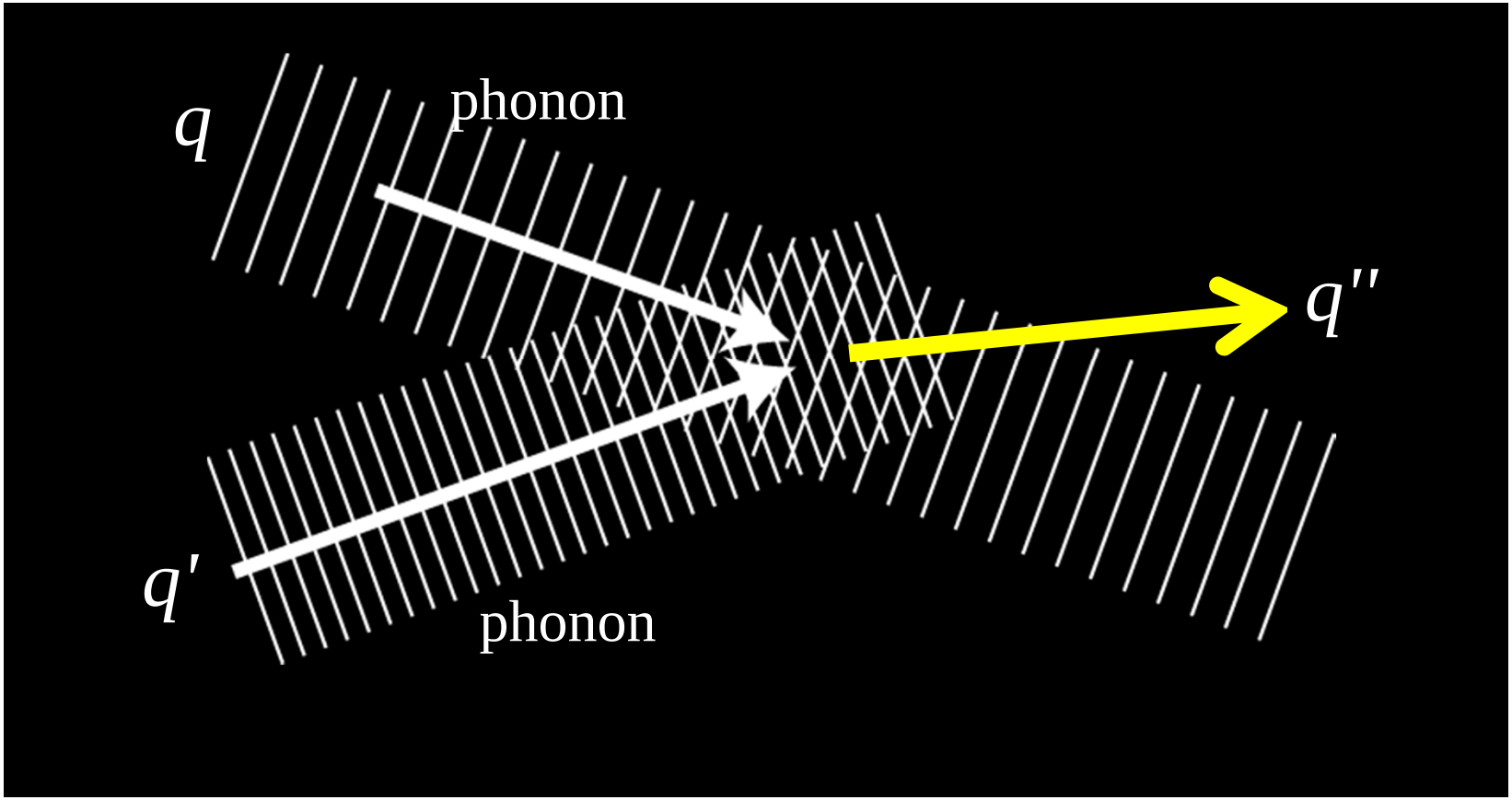
$$ZT = \frac{S^2\sigma}{\kappa} T$$

S : Seebeck coefficient
σ : electrical conductivity
κ : thermal conductivity

Physical origin of thermal resistivity

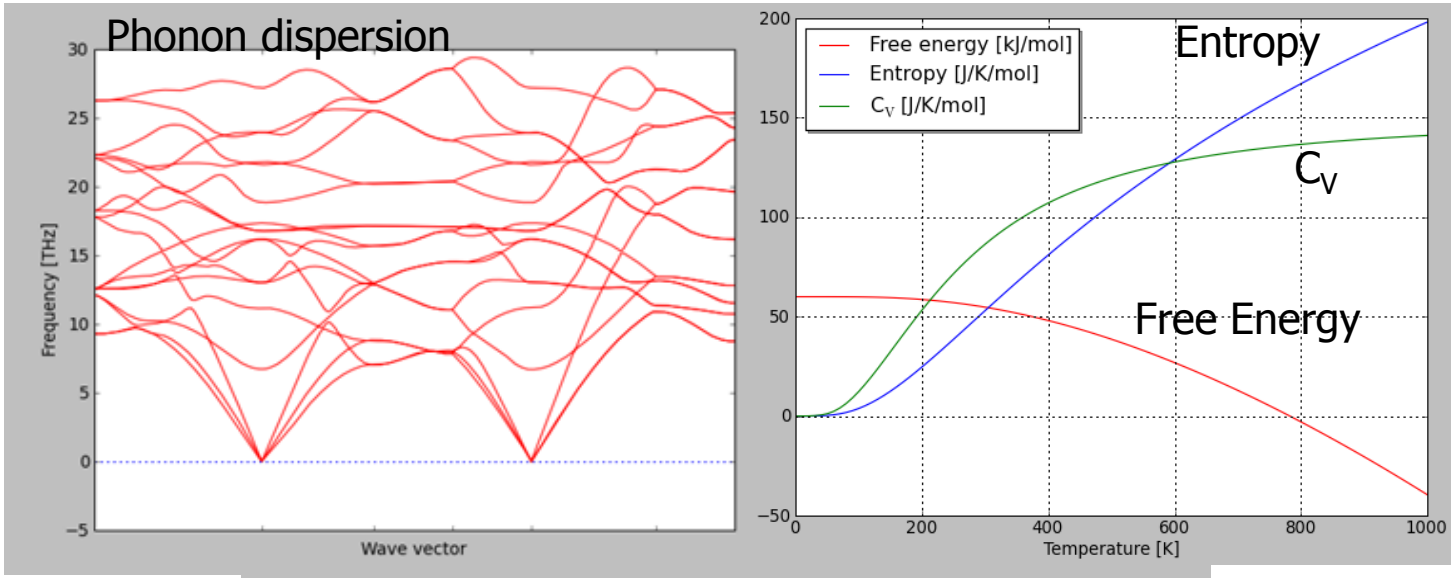
Phonon-phonon scattering (phonon anharmonicity)

Harmonic phonons do not interact.

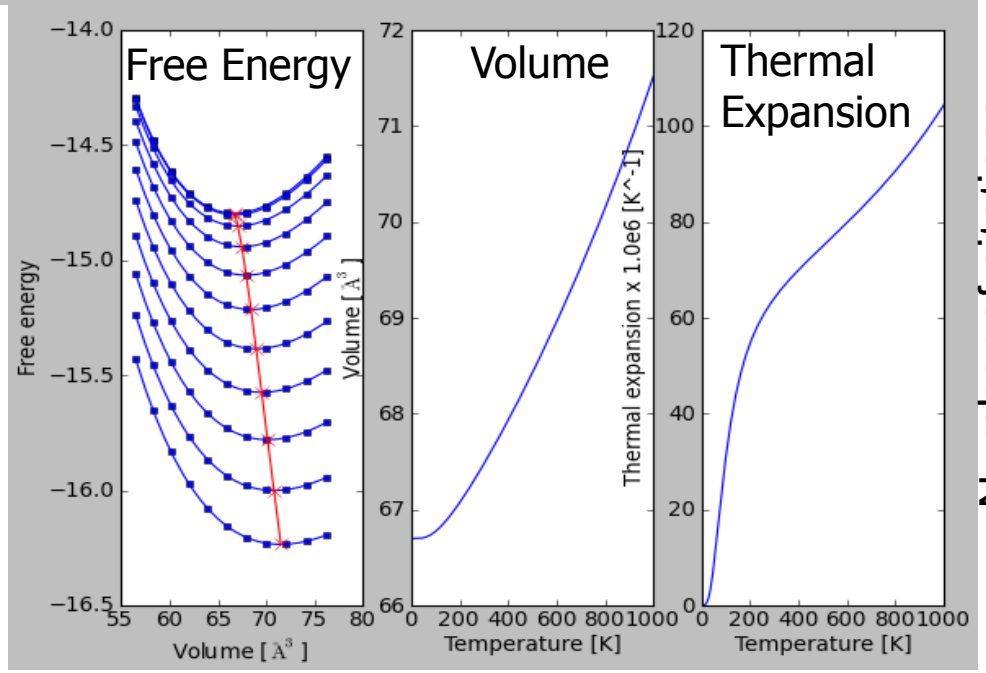


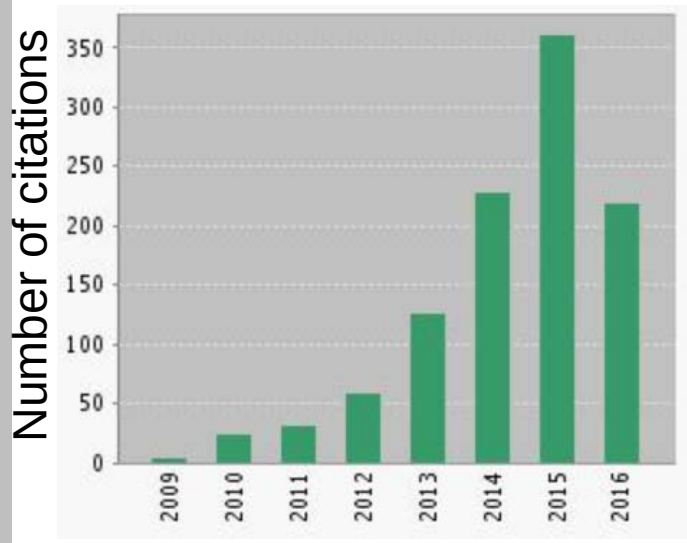
phonopy: open source for ab-initio phonon calcs

http://phonopy.sf.net/



Atsushi TOGO
Kyoto Univ.



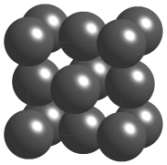


First principles phonon database

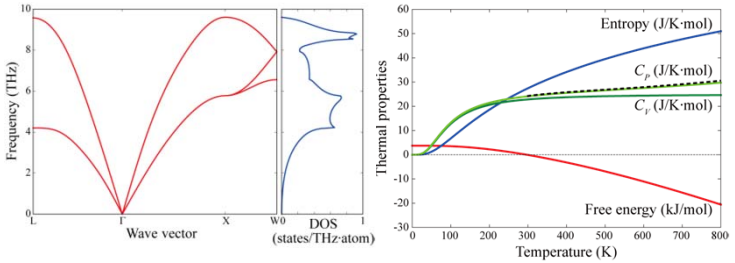
Structure Database
e.g.
ICSD, Atomwork, ...

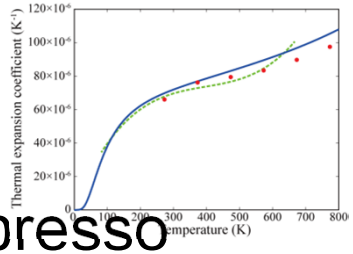
Phonon Database
An open database

DFT Database
e.g.
Mat. Proj.,
OQMD, aflowlib



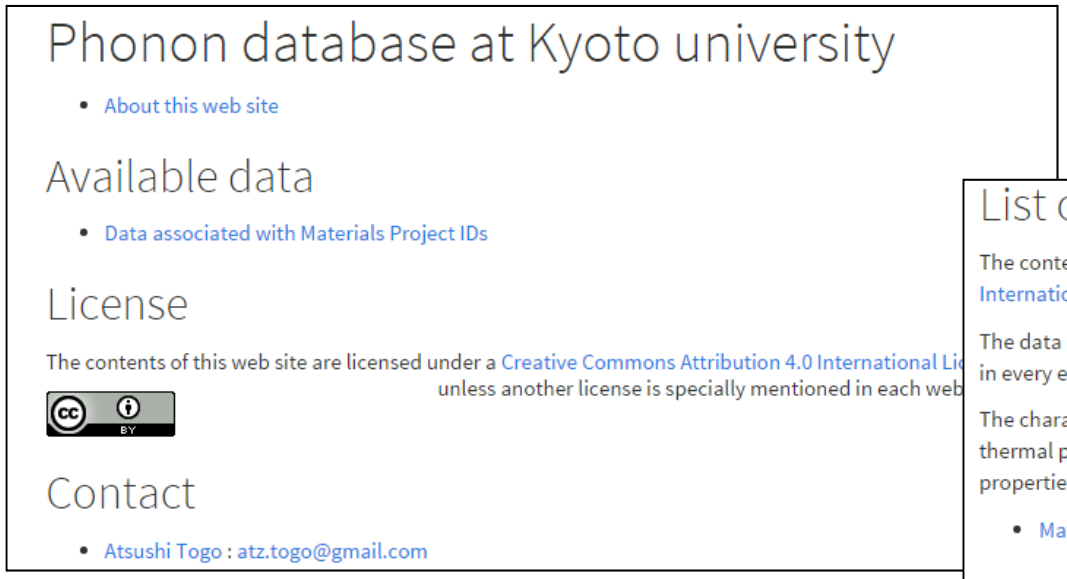
phonopy





First principles engine
e.g. VASP, Quantum Espresso
Abinit, etc...

Phonon Database @ Kyoto University



Phonon database at Kyoto university

- About this web site

Available data

- Data associated with Materials Project IDs

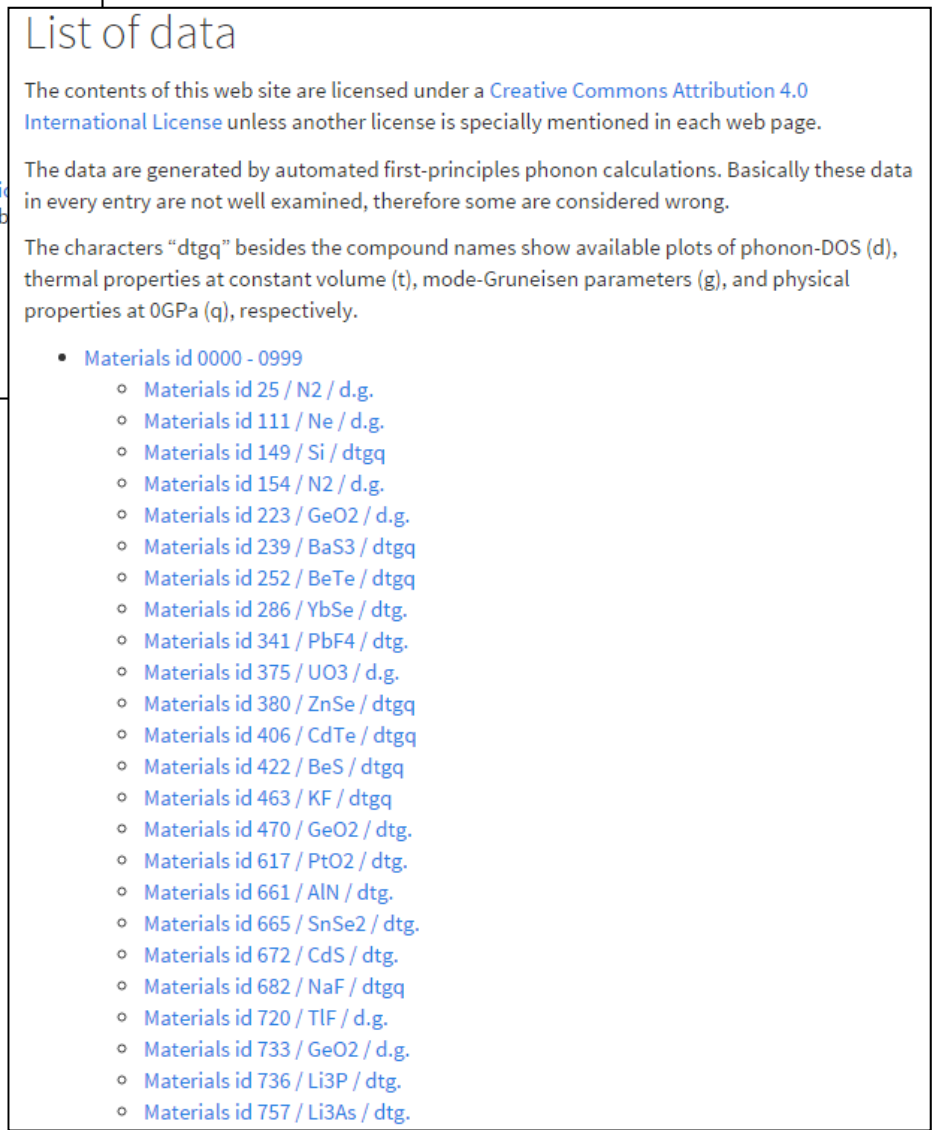
License

The contents of this web site are licensed under a Creative Commons Attribution 4.0 International Li
unless another license is specially mentioned in each web

Contact

- Atsushi Togo : atz.togo@gmail.com

http://phonondb.mtl.kyoto-u.ac.jp/



List of data

The contents of this web site are licensed under a Creative Commons Attribution 4.0 International License unless another license is specially mentioned in each web page.

The data are generated by automated first-principles phonon calculations. Basically these data in every entry are not well examined, therefore some are considered wrong.

The characters "dtgq" besides the compound names show available plots of phonon-DOS (d), thermal properties at constant volume (t), mode-Gruneisen parameters (g), and physical properties at 0GPa (q), respectively.

- Materials id 0000 - 0999
  - Materials id 25 / N2 / d.g.
  - Materials id 111 / Ne / d.g.
  - Materials id 149 / Si / dtgq
  - Materials id 154 / N2 / d.g.
  - Materials id 223 / GeO2 / d.g.
  - Materials id 239 / BaS3 / dtgq
  - Materials id 252 / BeTe / dtgq
  - Materials id 286 / YbSe / dtg.
  - Materials id 341 / PbF4 / dtg.
  - Materials id 375 / UO3 / d.g.
  - Materials id 380 / ZnSe / dtgq
  - Materials id 406 / CdTe / dtgq
  - Materials id 422 / BeS / dtgq
  - Materials id 463 / KF / dtgq
  - Materials id 470 / GeO2 / dtg.
  - Materials id 617 / PtO2 / dtg.
  - Materials id 661 / AlN / dtg.
  - Materials id 665 / SnSe2 / dtg.
  - Materials id 672 / CdS / dtg.
  - Materials id 682 / NaF / dtgq
  - Materials id 720 / TlF / d.g.
  - Materials id 733 / GeO2 / d.g.
  - Materials id 736 / Li3P / dtg.
  - Materials id 757 / Li3As / dtg.

d: phonon-DOS

t: thermal properties at constant volume

g: mode-Grüneisen parameters

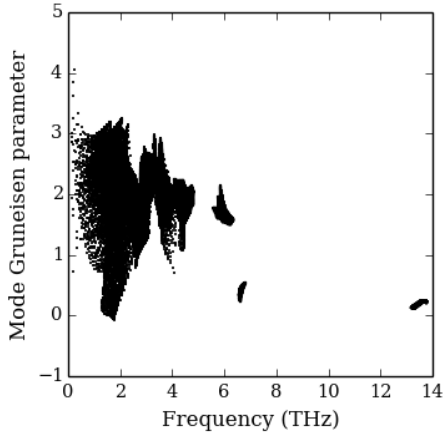
q: physical properties at 0GPa

Phonon Database @ Kyoto University

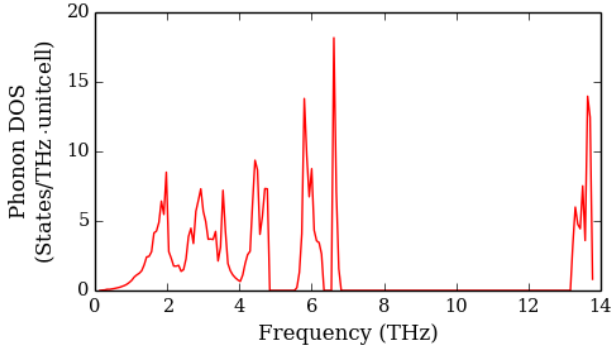
Materials id 239 / BaS3 / dtgq

- Date page updated: 2015-11-24
- Space group type: P-42_1m (113) / P -4 2ab
- Number of formula units (Z): 2
- Phonon raw data: **mp-239.tar.lzma**
- Link to Materials Project: https://www.materialsproject.org/materials/mp-239/

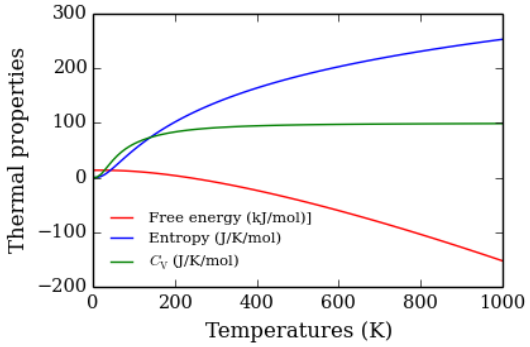
Mode Gruneisen parameter



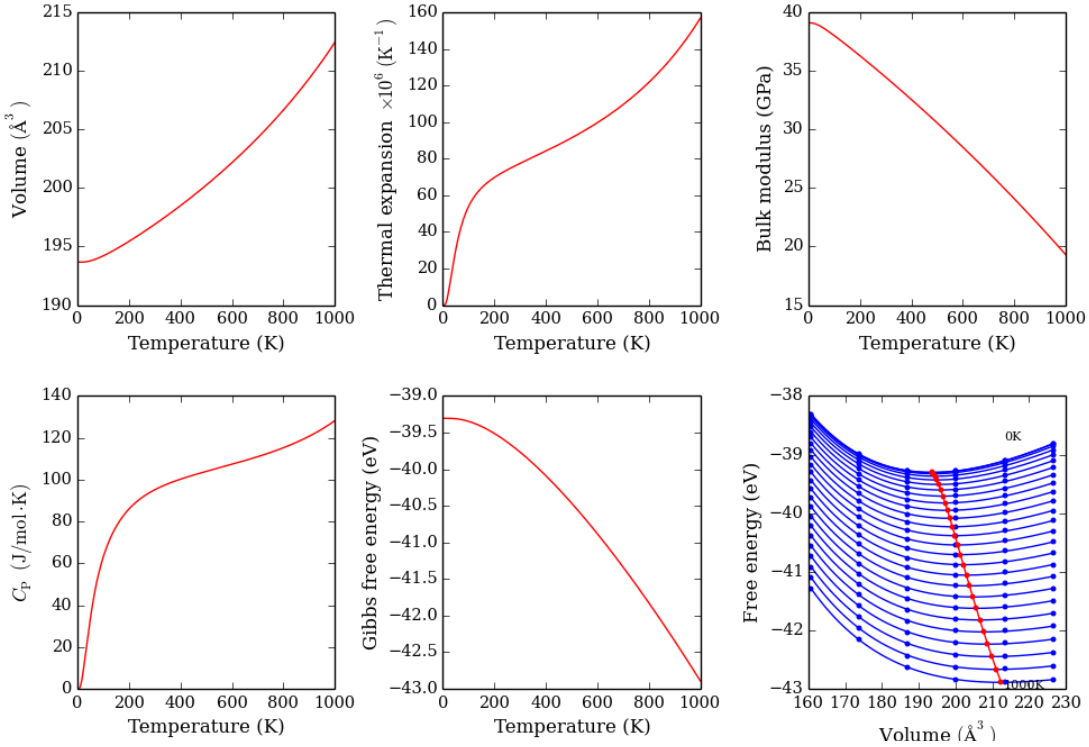
Phonon DOS



Thermal properties at constant volume



Properties at 0GPa under quasi-harmonic approximation



Ab-initio Lattice Thermal Conductivity (LTC)

Phono3py 1.10.5 Site ▾ Page ▾ Work flow »

Welcome to phono3py

This software calculates phonon-phonon interaction related properties:

- Lattice thermal conductivity
- Phonon lifetime/linewidth
- Imaginary part of self energy at the lowest order
- Joint density of states (JDOS) and weighted-JDOS

The theoretical background is summarized in the paper found at
http://dx.doi.org/10.1103/PhysRevB.91.094306 or the draft in arxiv at http://arxiv.org/abs/1501.00691 .

Examples are found in `example-phono3py` directory.

Documentation

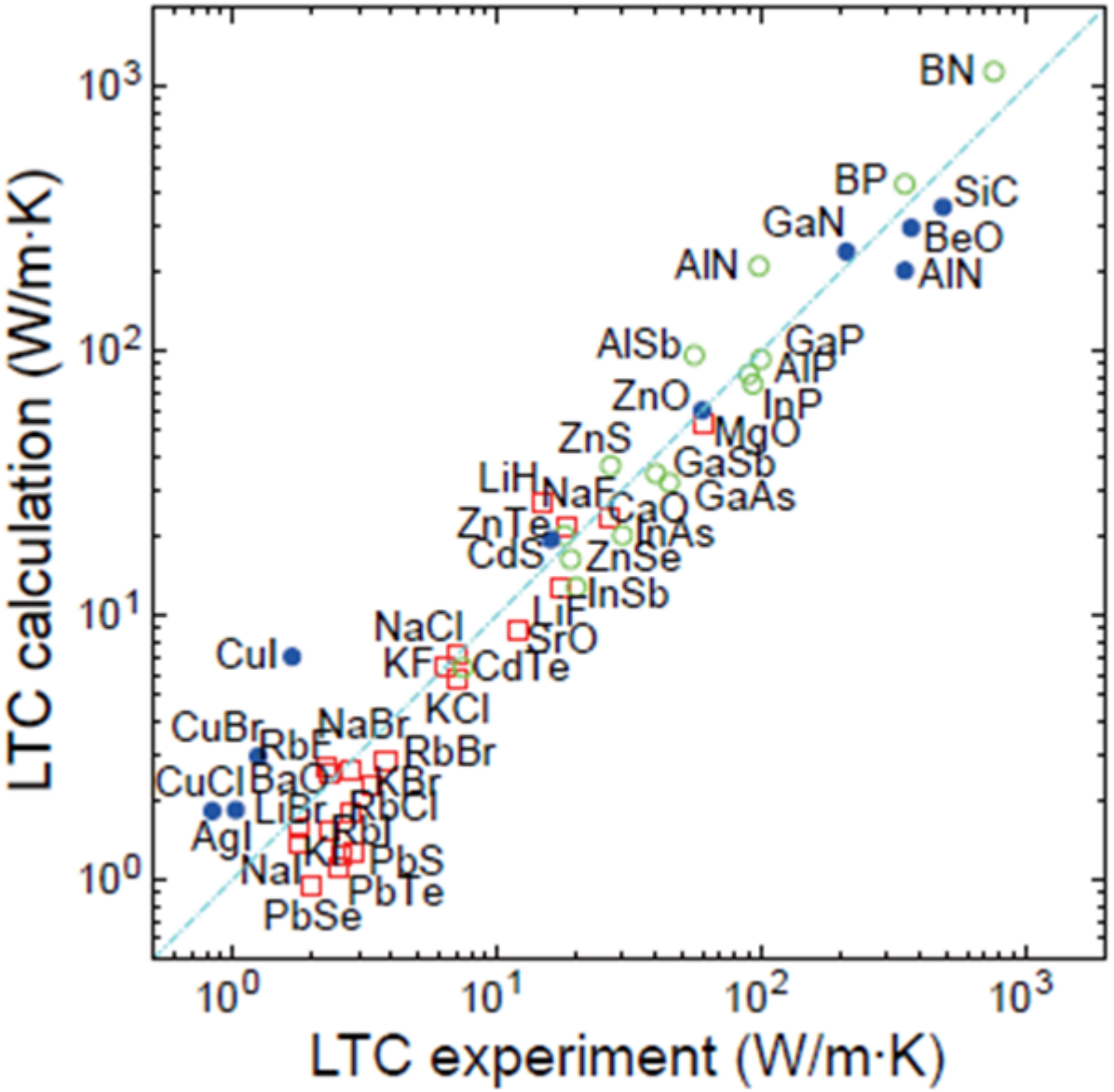
- Work flow
- Installation
- Output files
- Command options
- Interfaces to calculators
- Auxiliary tools
- How to cite phono3py
- Change Log

Atsushi TOGO

Lattice Thermal Conductivity (LTC)@300K

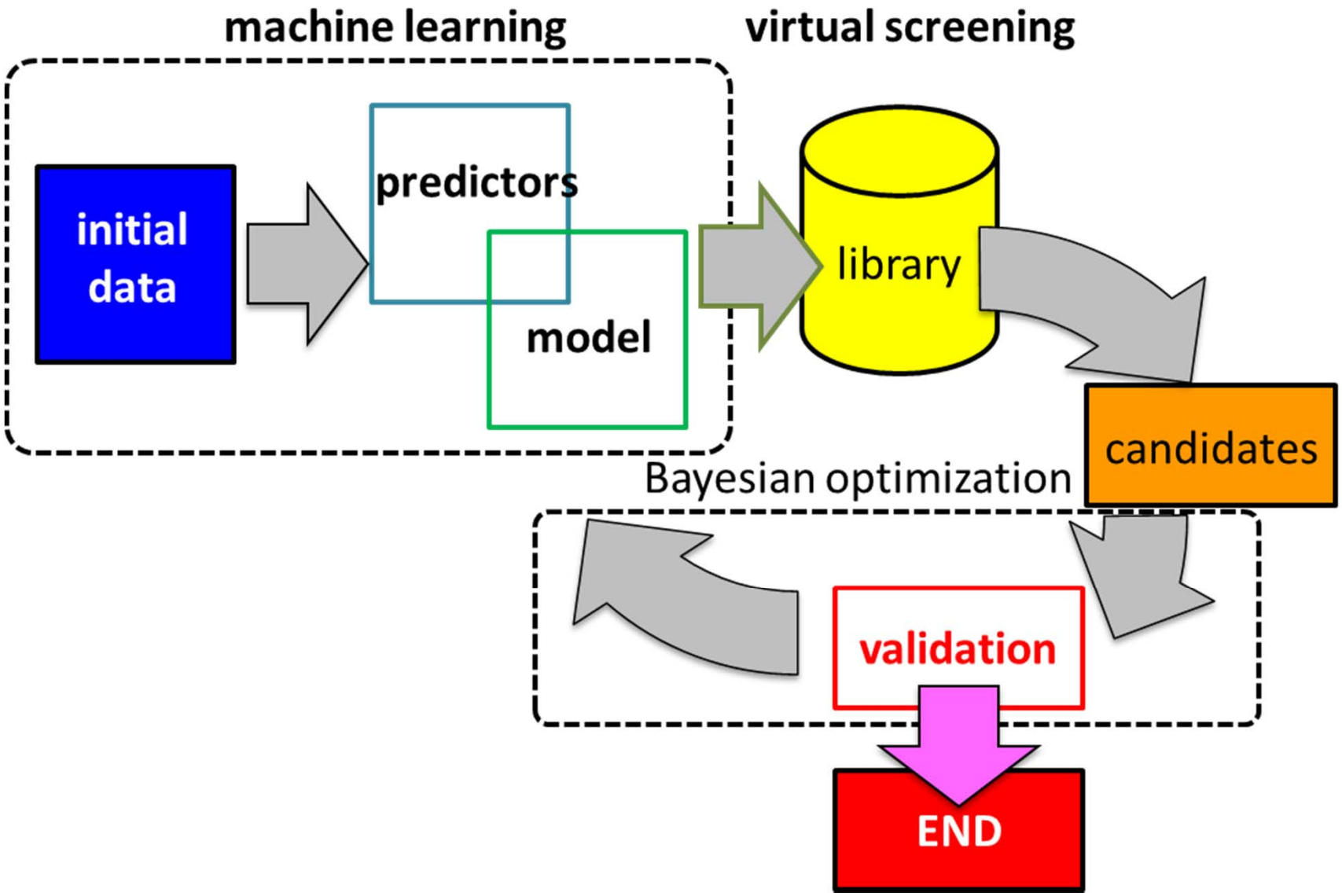
First principles calculation vs. Experimental data

phono3py



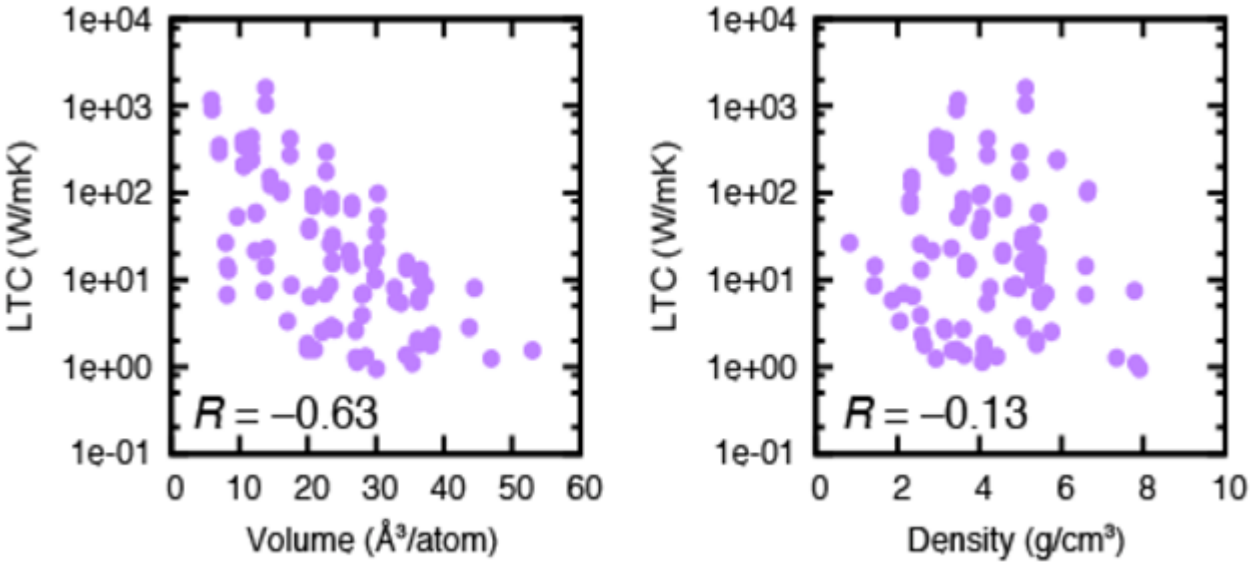
Reliable calculations whose accuracy are comparable to experiments!

Virtual Screening



Candidates of descriptors for LTC

Model 1: Volume V and Density ρ



Model 2: Model 1 + primitive elemental descriptors

	H	Li	Be	B	C	N	O	F	…
LiH	1	1	0	0	0	0	0	0	
LiF	0	1	0	0	0	0	0	1	
BeO	0	0	1	0	0	0	1	0	
BN	0	0	0	1	0	1	0	0	
…									

Ranking of LTC for 54,779 compounds in MPD library

"Virtual screening"

Z-score

$$Z(\boldsymbol{x}^*) = [f(\boldsymbol{x}^*) - f_{\text{best}}]/\sqrt{v(\boldsymbol{x}^*)}$$

$$f = -\log \kappa_L$$

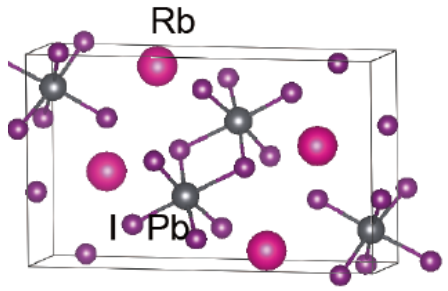
The higher Z-score, the lower predicted LTC !

	Materials Id	Formula	Z-score
1	mp-23517	RbPbI3	1.90
2	mp-571465	PbIBr	1.76
3	mp-28564	Rb4PbBr6	1.56
4	mp-23053	PbICl	1.56
5	mp-22997	PbBrCl	1.56
6	mp-23043	RbPb2Br5	1.45
7	mp-22883	PbI2	1.44
8	mp-567503	PbI2	1.43
9	mp-540839	CsPbI3	1.40
10	mp-569595	PbI2	1.34
11	mp-567246	PbI2	1.34
12	mp-23436	Cs4PbBr6	1.34
13	mp-505148	ThPbI6	1.34
14	mp-22893	PbI2	1.34
15	mp-567199	PbI2	1.32
16	mp-580202	PbI2	1.32
17	mp-600089	CsPbBr3	1.31
18	mp-640058	PbI2	1.31
19	mp-574189	PbI2	1.30
20	mp-567542	PbI2	1.30
21	mp-540789	PbI2	1.30
22	mp-567178	PbI2	1.29
23	mp-567629	CsPbBr3	1.29
24	mp-672671	PbI2	1.28
25	mp-561320	PbS	1.27
26	mp-680205	PbI2	1.25
27	mp-29883	Rb3PbCl5	1.25
28	mp-567681	CsPbBr3	1.24
29	mp-608081	Rb3Pb4Au	1.18
30	mp-27552	TlPbI3	1.17
31	mp-674972	Rb6Pb5Cl16	1.15
32	mp-31508	KPb2Br5	1.15
33	mp-771814	CaPbI6	1.14
34	mp-23037	CsPbCl3	1.14
35	mp-675524	CsPbCl3	1.14
36	mp-675022	CsPbCl3	1.14
37	mp-23475	Rb2PbCl6	1.13
38	mp-29212	Tl4PbI6	1.12
39	mp-581775	Cs5(KPb6)3	1.09
40	mp-23380	Tl3PbI5	1.08
41	mp-630851	Cs3NaPb4	1.07
42	mp-574070	Cs4Pb9	1.05
43	mp-674339	Tl6PbI10	1.04
44	mp-23425	Cs2PbCl6	1.03
45	mp-771691	CaPbI6	1.01
46	mp-756313	CaPbI4	0.99
47	mp-673703	Rb3Bi7Pb3(IO)10	0.99
48	mp-680463	Rb4Pb9	0.97
49	mp-755943	CaPbI4	0.96
50	mp-622294	Hg2Pb(SBr)2	0.96
51	mp-21525	RbPb	0.96
52	mp-571638	Rb2Cu(BrCl)2	0.90
53	mp-756136	CaPbI4	0.90
54	mp-31317	La3PbI3	0.88
55	mp-756451	CaPbI4	0.88
56	mp-753670	CaPbI4	0.87
57	mp-569879	Cs2Ta6PbCl18	0.86
58	mp-570753	Tl3PbBr5	0.86
59	mp-27451	Tl3PbBr5	0.85
60	mp-569238	Cs3LiI4	0.85
61	mp-755977	CaPbI4	0.84
62	mp-680159	CsTa6PbCl18	0.84
63	mp-557719	PbS	0.82
64	mp-755056	CaPbI4	0.82
65	mp-771827	Ca3Pb3I14	0.81
66	mp-754540	CaPbI4	0.81
67	mp-622106	Ba5Pb3	0.81
68	mp-771877	Ca3Pb3I14	0.80
69	mp-621612	AgPb2Br5	0.80
70	mp-21246	Ba2Pb	0.78
71	mp-37163	Ag2HgI4	0.78
72	mp-569465	Cs2Nb6PbCl18	0.77
73	mp-554116	BaPb2BrF5	0.76
74	mp-674993	KPb2Cl5	0.73
75	mp-607267	KPb2Cl5	0.72
76	mp-3	Cs	0.71
77	mp-27662	CsI2Br	0.71
78	mp-20136	BaPb	0.70
79	mp-573579	Cs	0.70
80	mp-672241	Cs	0.70
81	mp-11832	Cs	0.70
82	mp-1	Cs	0.69
83	mp-639727	Cs	0.69
84	mp-569225	Mo6PbI14	0.67
85	mp-541112	ZrI4	0.66
86	mp-28077	PbBr2	0.66
87	mp-569866	Cs6K7	0.64
88	mp-31288	La5Pb3I	0.64
89	mp-554245	BaPb2IF5	0.64
90	mp-613652	PbClF	0.63
91	mp-5811	CsPbF3	0.61
92	mp-570355	PbCl4	0.61
93	mp-20282	CsPbF3	0.60
94	mp-579536	CsCu2ICl2	0.60
95	mp-22969	PbIF	0.59
96	mp-559470	AgPbBrO	0.58
97	mp-674359	Tl3PbCl5	0.58
98	mp-30519	Tl3PbCl5	0.58
99	mp-23066	Pb5(SI3)2	0.56

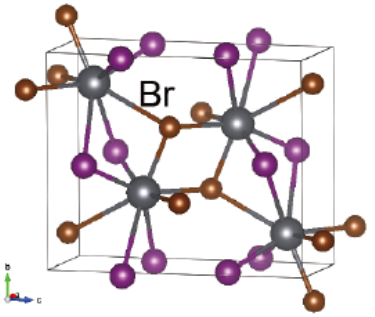
Top 10 lowest LTC compounds among 54,779

Virtual screening of 54,779 compounds in MPD library

ranking	Z-score	compound	
		formula	space group
1	1.90	$PbRbI_3$	*Pnma*
2	1.76	PbIBr	*Pnma*
3	1.56	$PbRb_4Br_6$	*R*-3*c*
4	1.56	PbICl	*Pnma*
5	1.56	PbClBr	*Pnma*
7	1.44	PbI_2	*R*-3*m*
8	1.43	PbI_2	*P*63*mc*

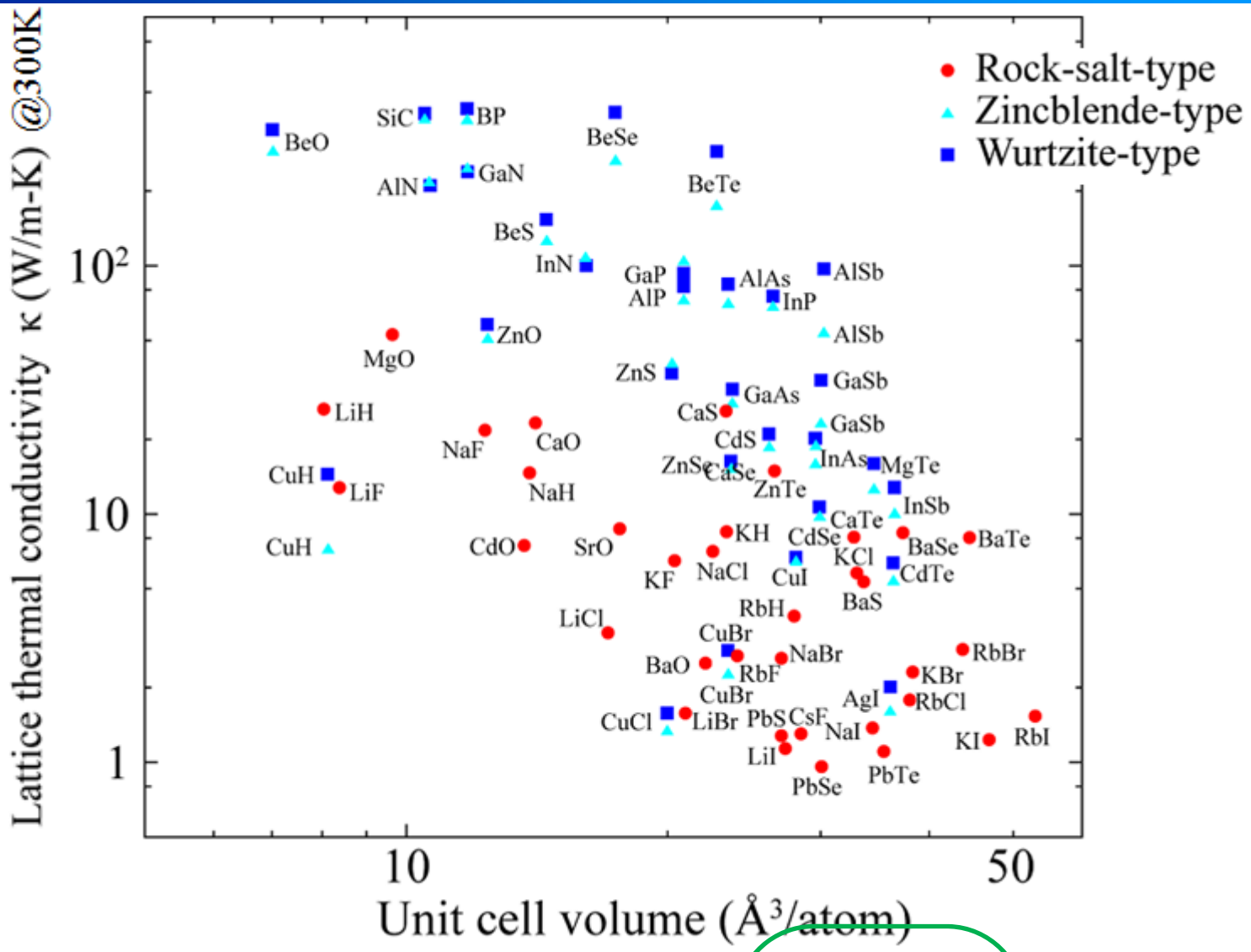


$PbRbI_3$

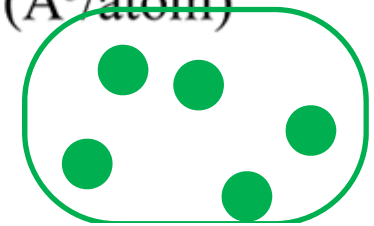


PbIBr

Lattice Thermal Conductivity (LTC)



Newly discovered !!



Thermoelectric materials

Essential for utilizing otherwise waste heat.

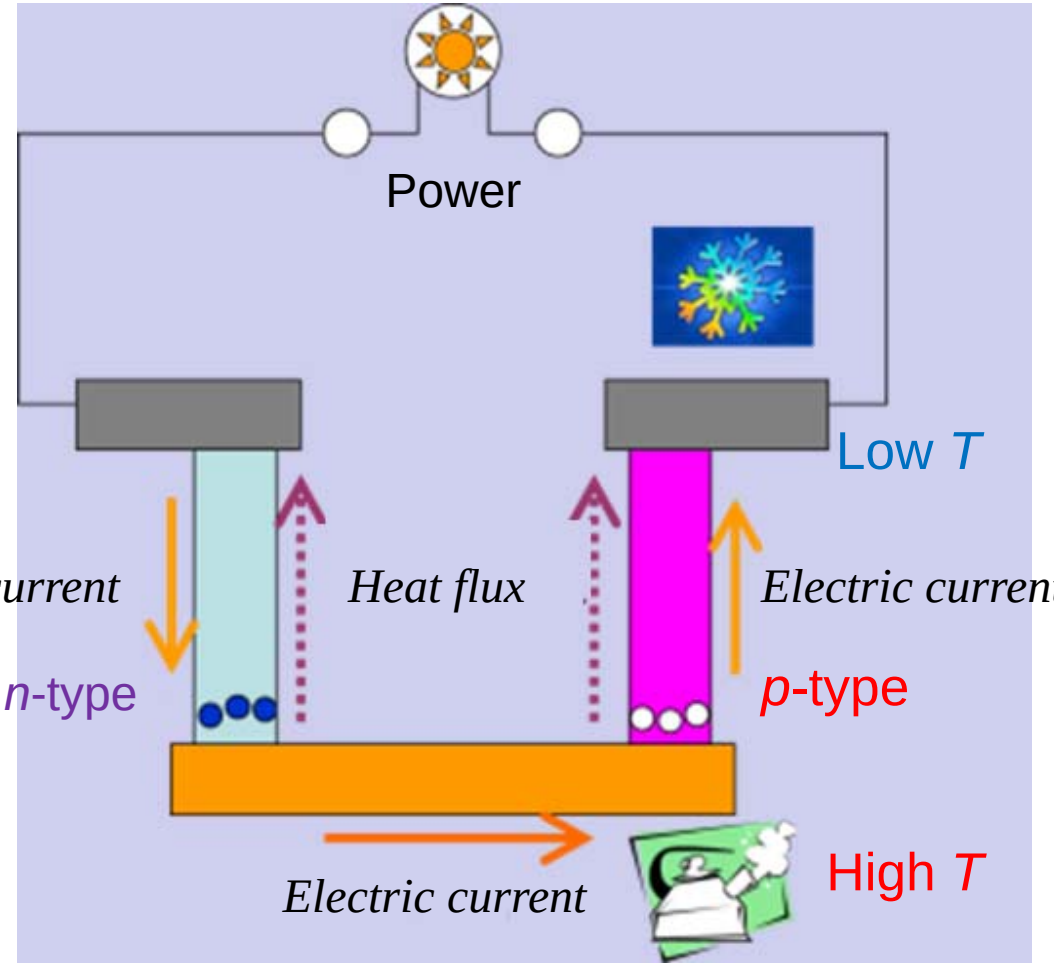


Figure of Merit

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

S : Seebeck coefficient

σ : electrical conductivity

κ : thermal conductivity

Top 10 lowest LTC compounds among 54,779

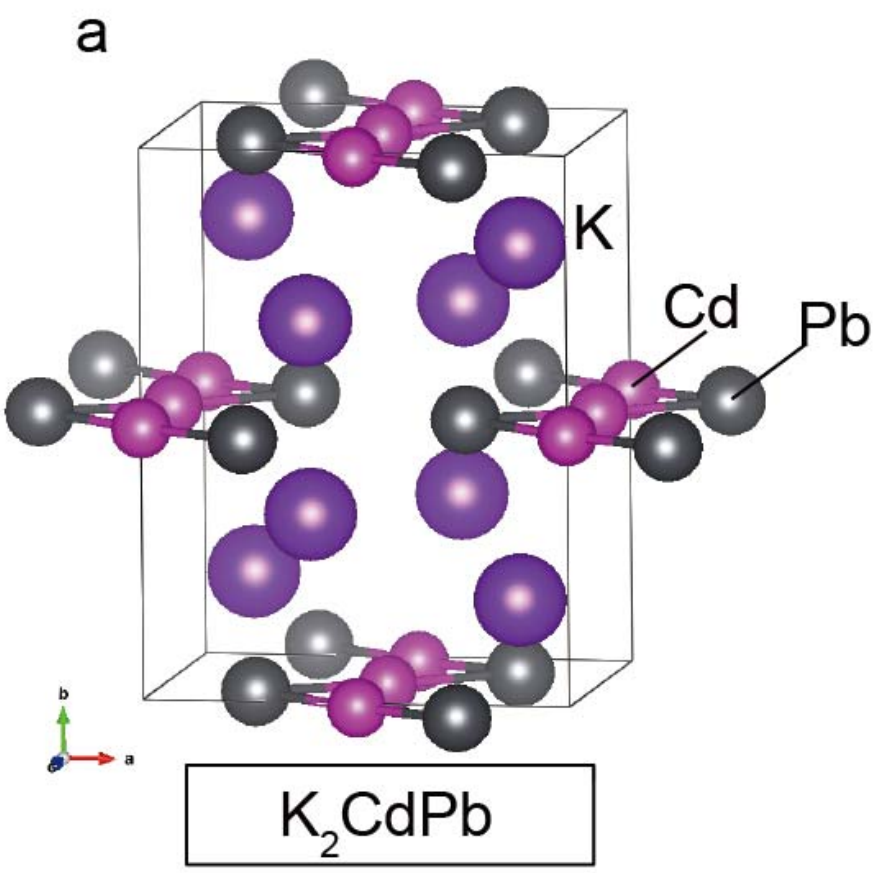
Virtual screening of 54,779 compounds in MPD library

ranking	Z-score	compound		*ab initio* LTC (W/m·K)	Band Gap (eV)
		formula	space group		
1	1.90	$PbRbI_3$	*Pnma*	0.10	2.46
2	1.76	PbIBr	*Pnma*	0.13	2.56
3	1.56	$PbRb_4Br_6$	*R*-3*c*	0.08	3.90
4	1.56	PbICl	*Pnma*	0.18	2.72
5	1.56	PbClBr	*Pnma*	0.09	3.44
7	1.44	PbI_2	*R*-3*m*	0.29	2.42
8	1.43	PbI_2	*P*63*mc*	0.29	2.45
121	0.39	K_2CdPb	*Ama*2	0.45	0.18
144	0.29	$Cs_2[PdCl_4]I_2$	*I*4/*mmm*	0.31	0.88

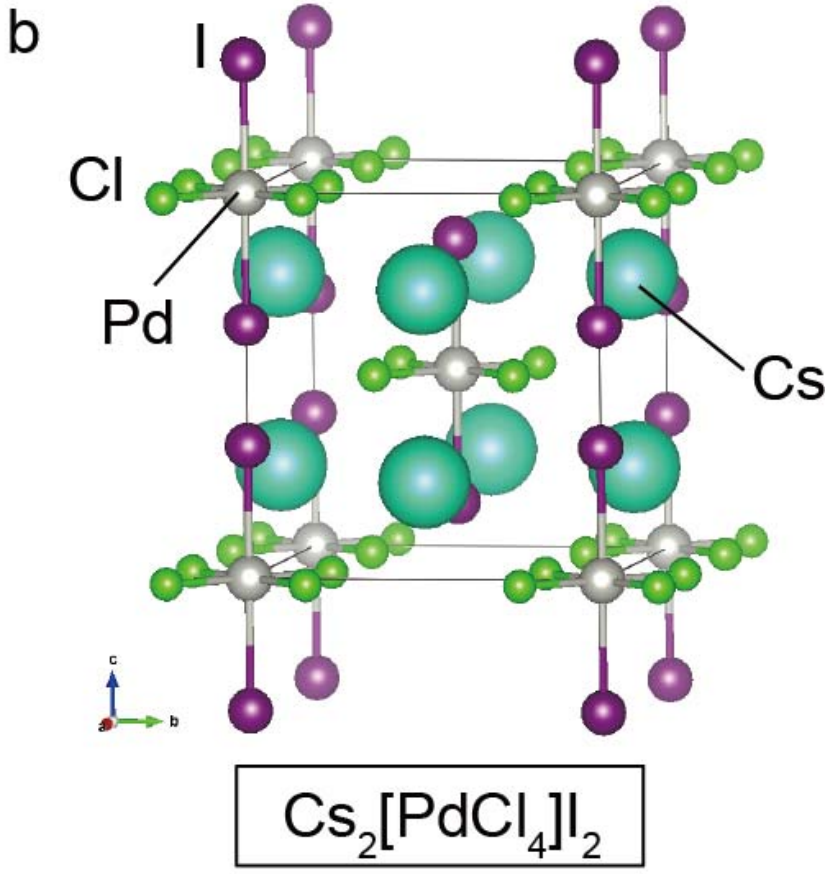
Newly discovered candidates for thermoelectrics

with low LTC of <0.5 W/mK (@300 K)

and narrow band gap of <1 eV.



K_2CdPb



$Cs_2[PdCl_4]I_2$

A. Seko, A. Togo, H. Hayashi, K. Tsuda, L. Chaput, and IT PRL (2015) 115, 205901.

Summary

Materials discovery through first principles calculations

✓**Discovery of a new Sn(II)-based oxide for daylight-driven photocatalyst**

REAL SCREENING

✓**Discovery of new low thermal conductivity materials from Mat. Proj. database (55,000 crystals)**

VIRTUAL SCREENING

Fin