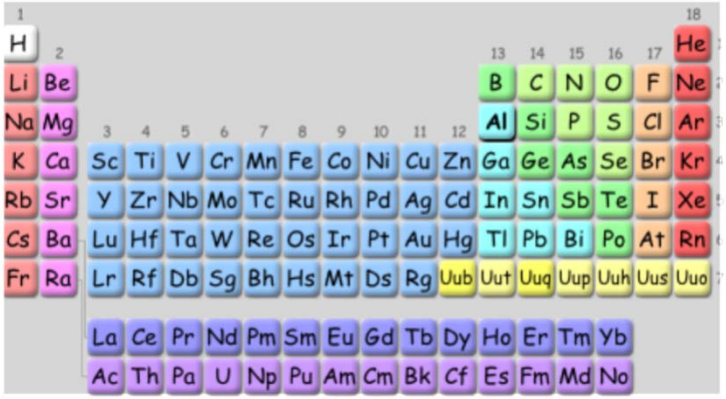
Data driven discovery of new inorganic crystalline materials

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

Vast chemistry space to explore



Periodic Table

Enormous number of chemical combinations

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 (1billion)

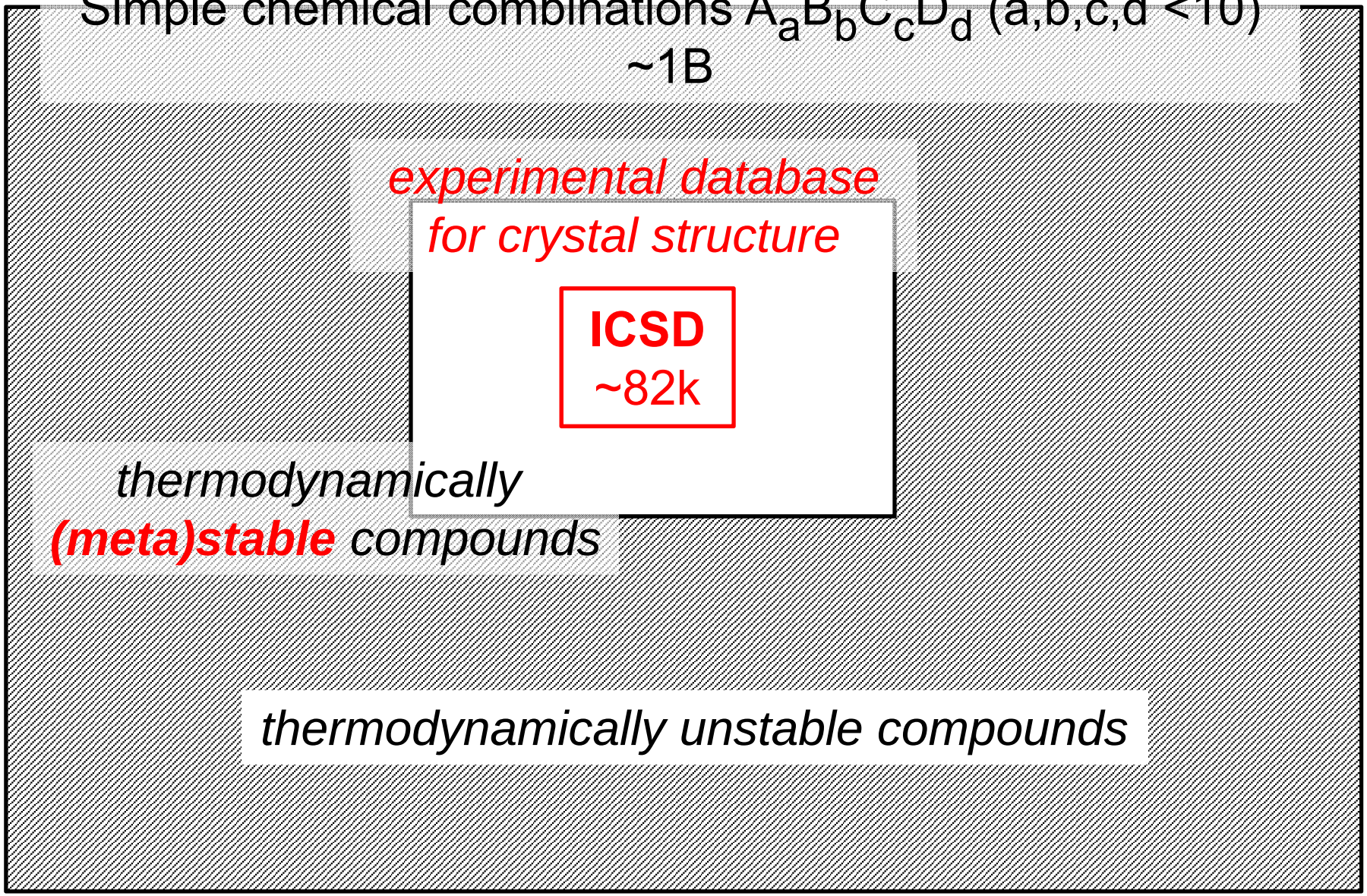
Inorganic Crystal Structure Database (ICSD)

World largest database
for known inorganic crystals.

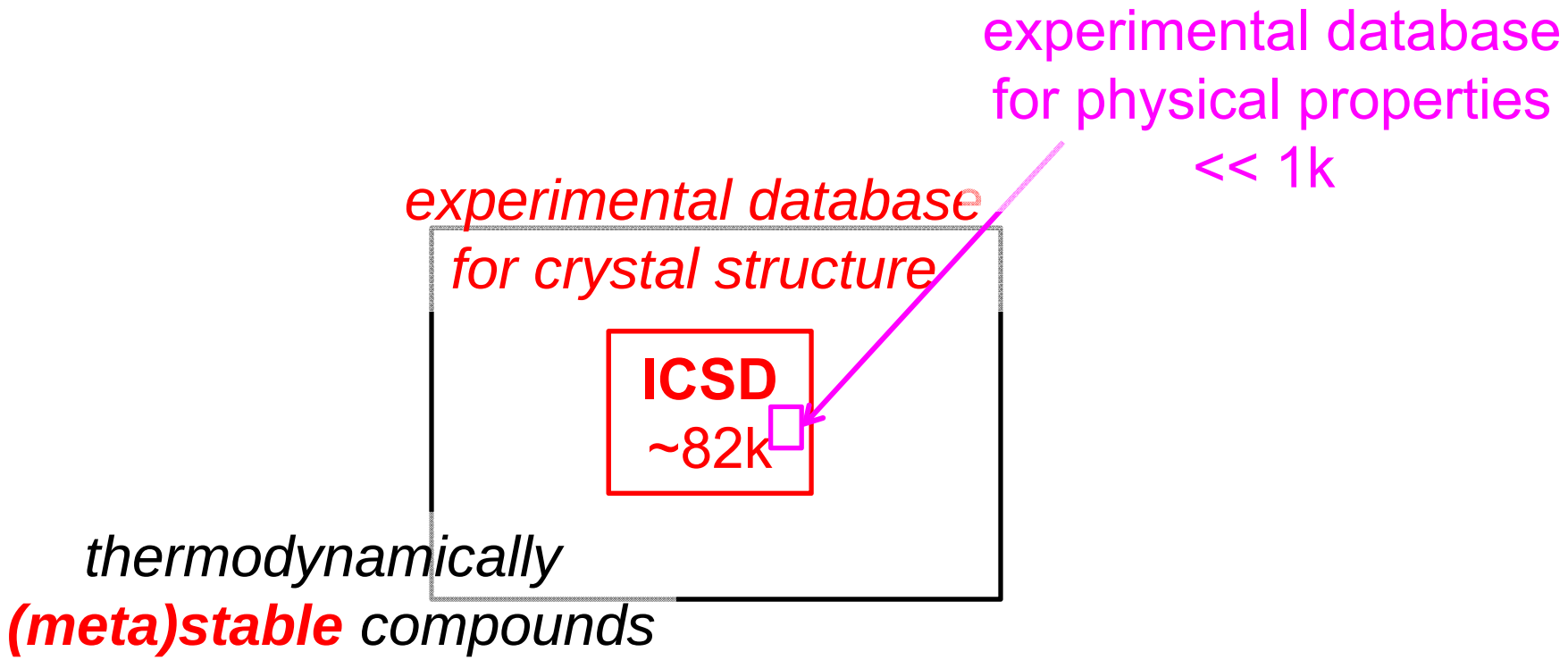
187,000 crystal structures
→ 82,000 structures
excluding duplicates
incompletes, etc.

Many systems are yet-unexplored !

Vast chemistry space to explore



Vast chemistry space to explore



✓ **It is still worthy to explore the space of known compounds.**

Discovery of new low thermal conductivity materials

First-Principles Anharmonic Lattice-Dynamics Calculations and Gaussian Process Regression



Atsuto Seko, Atsushi Togo, Hiroyuki Hayashi,

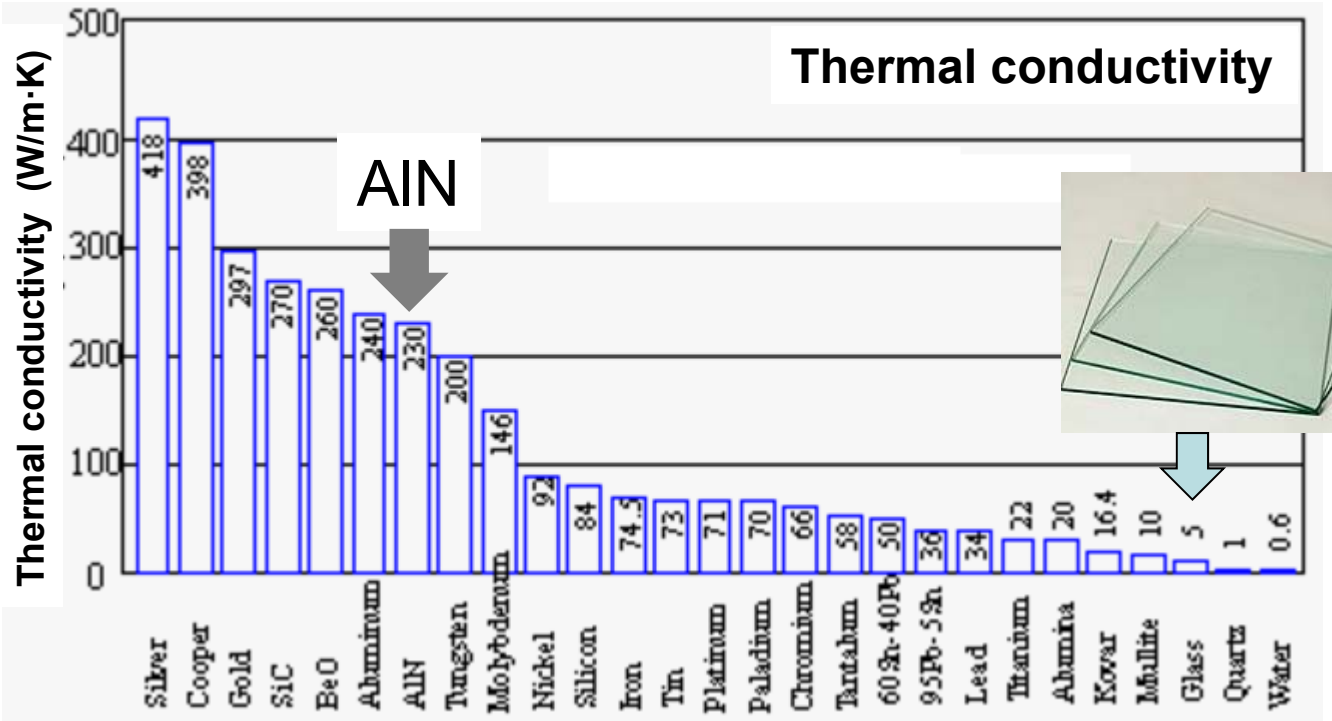
Koji Tsuda, Laurent Chaput, and IT

PRL (2015) 115, 205901.

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

diamond

>2000



Thermoelectric materials

Essential for utilizing otherwise waste heat.

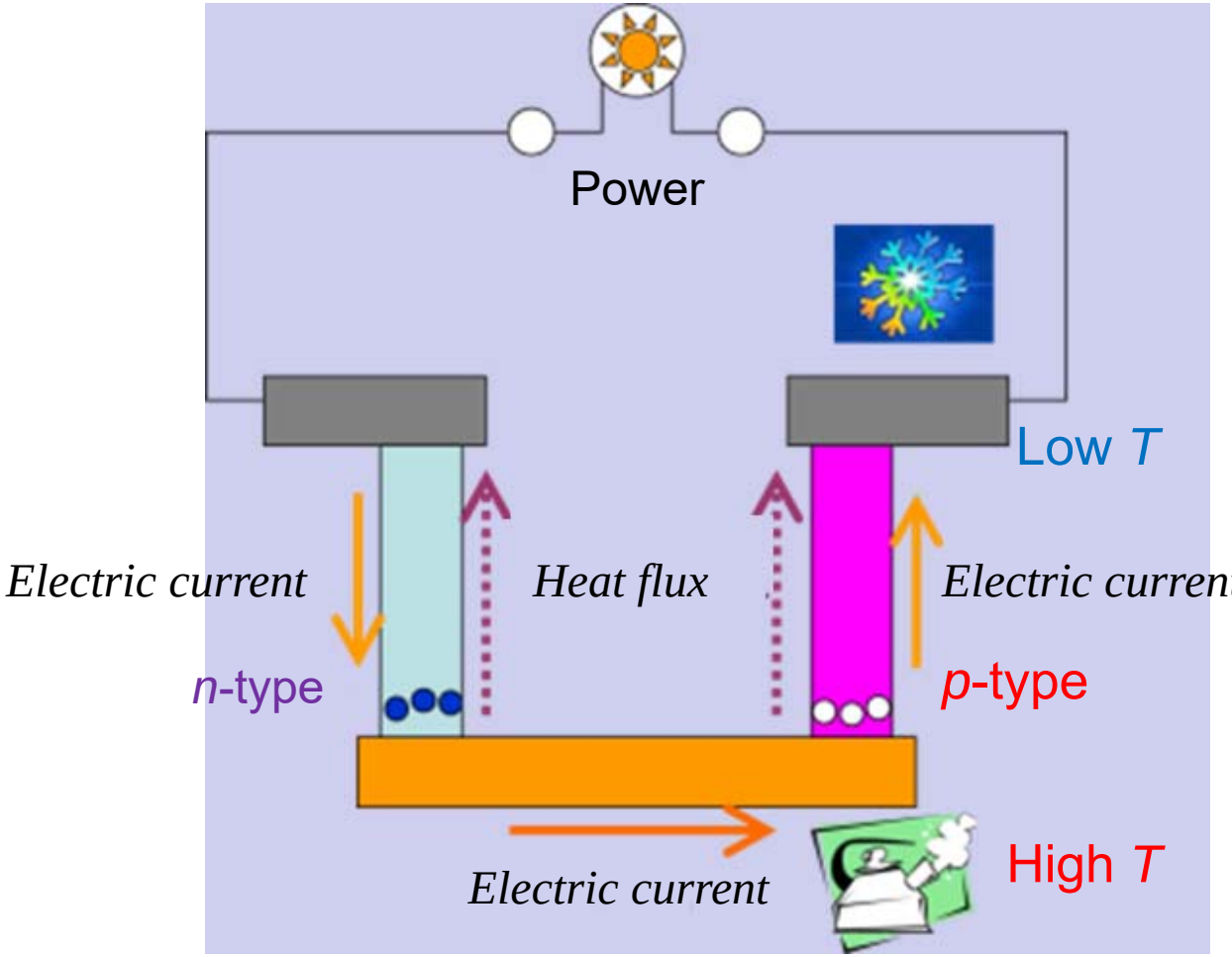


Figure of Merit

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

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

Background

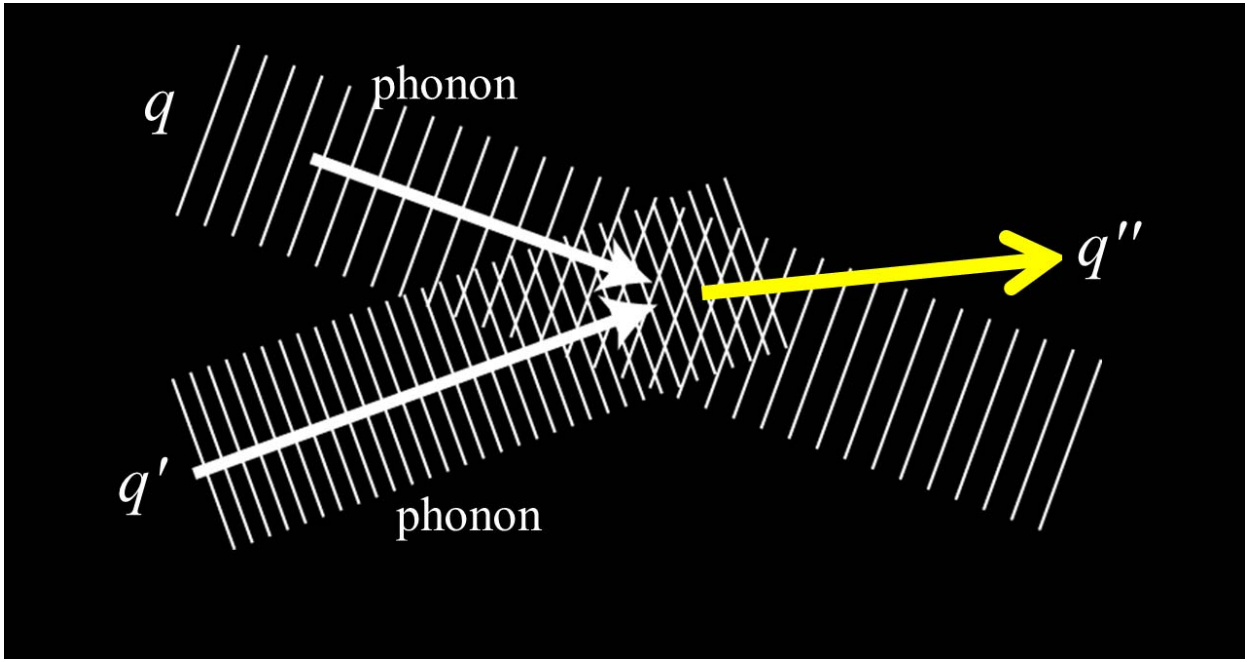
Thermal Conductivity (TC)

$$\kappa = \kappa_{electonic} + \underline{\kappa_{lattice}}$$ Lattice Thermal Conductivity (LTC)

Physical origin of thermal resistivity

Phonon-phonon scattering (phonon anharmonicity)

Harmonic phonons do not interact.



Background

Lattice Thermal Conductivity (LTC)

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

  (1 day/100 cores for 1 simple crystal.)
- ✓ **Insufficient knowledge to predict LTC deductively.**

  (Simple model to determine LTC is available, but....)

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

Ab-initio Lattice Thermal Conductivity (LTC)

https://atztogo.github.io/phono3py/

Atsushi TOGO

Welcome to phono3py

This software calculates phonon-phonon interaction and related properties using the supercell approach. For example, properties are obtained:

- Lattice thermal conductivity
- Phonon lifetime/linewidth
- Imaginary part of self energy
- 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 http://arxiv.org/abs/1501.00691 .

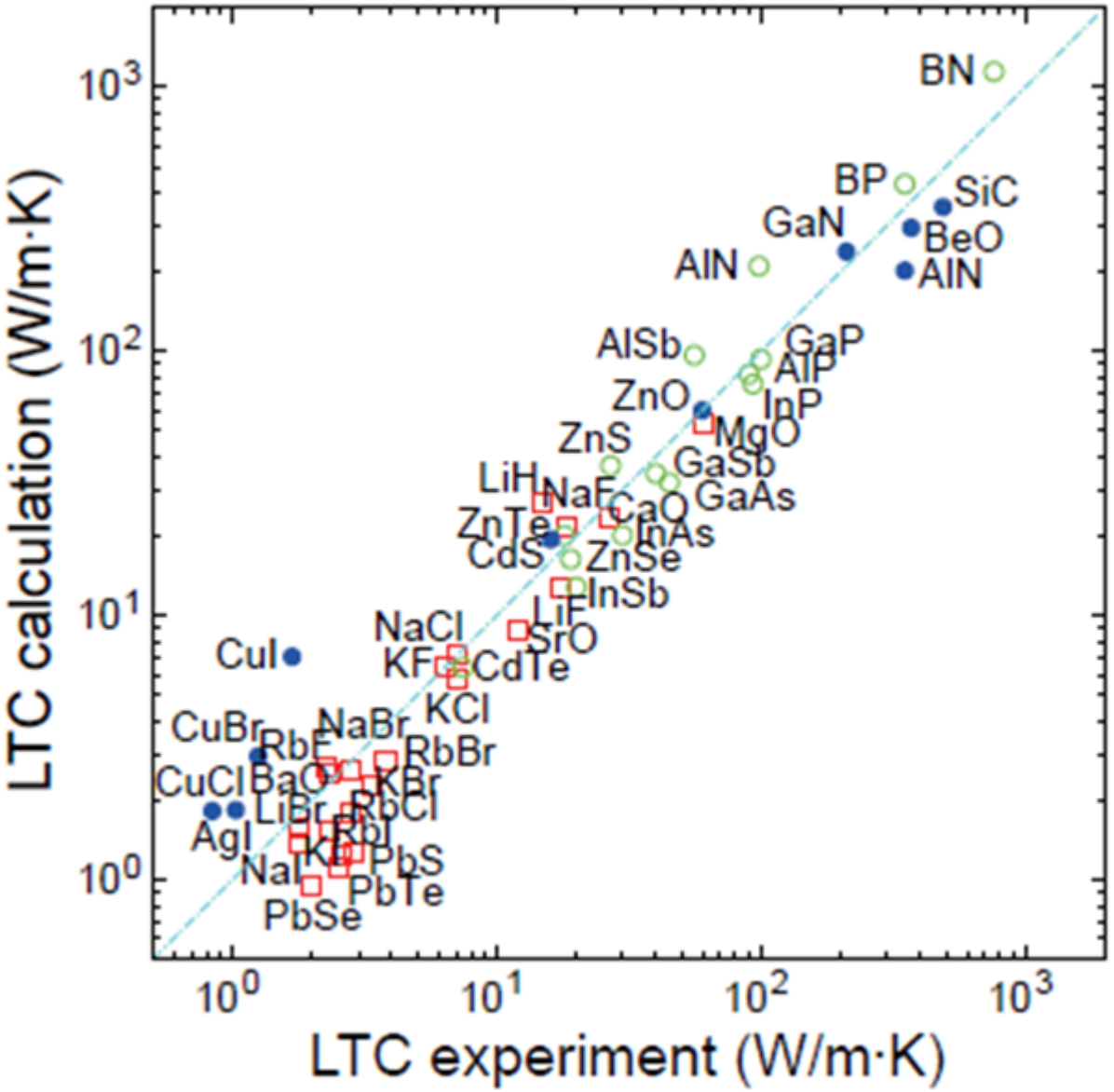
Documentation

- Installation
- Examples
- Interfaces to calculators (VASP, pwscf, CRYSTAL, Abinit)
- Command options / Setting tags
- Output files
- How to read the results stored in hdf5 files
- Auxiliary tools
- Direct solution of linearized phonon Boltzmann equation
- Workload distribution
- Crude force constants calculation by cutoff pair-distance
- External tools
- Tips
- How to cite phono3py
- Change Log

Lattice Thermal Conductivity (LTC)@300K

First principles calculation vs. Experimental data

phono3py



Computational cost

1 day (wall clock time) per 1 crystal with 100core-PC

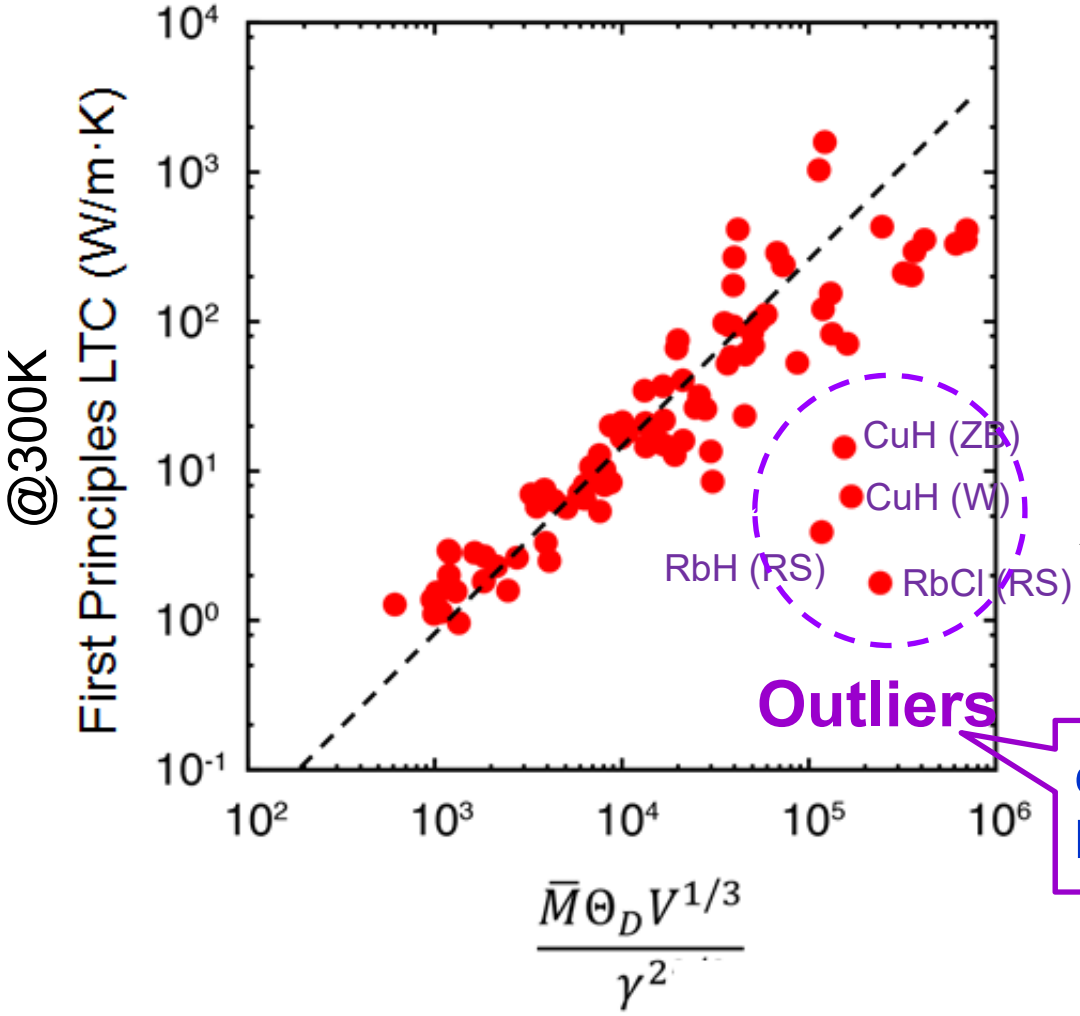
Reliable calculations whose accuracy are comparable to experiments!

Comparison of ab-initio LTC with Slack model

Slack model

$$\kappa = A\frac{\bar{M}\Theta_D V^{1/3}}{\gamma^2 T n^{2/3}}$$

harmonic phonon



M : atomic mass

V: volume per atom

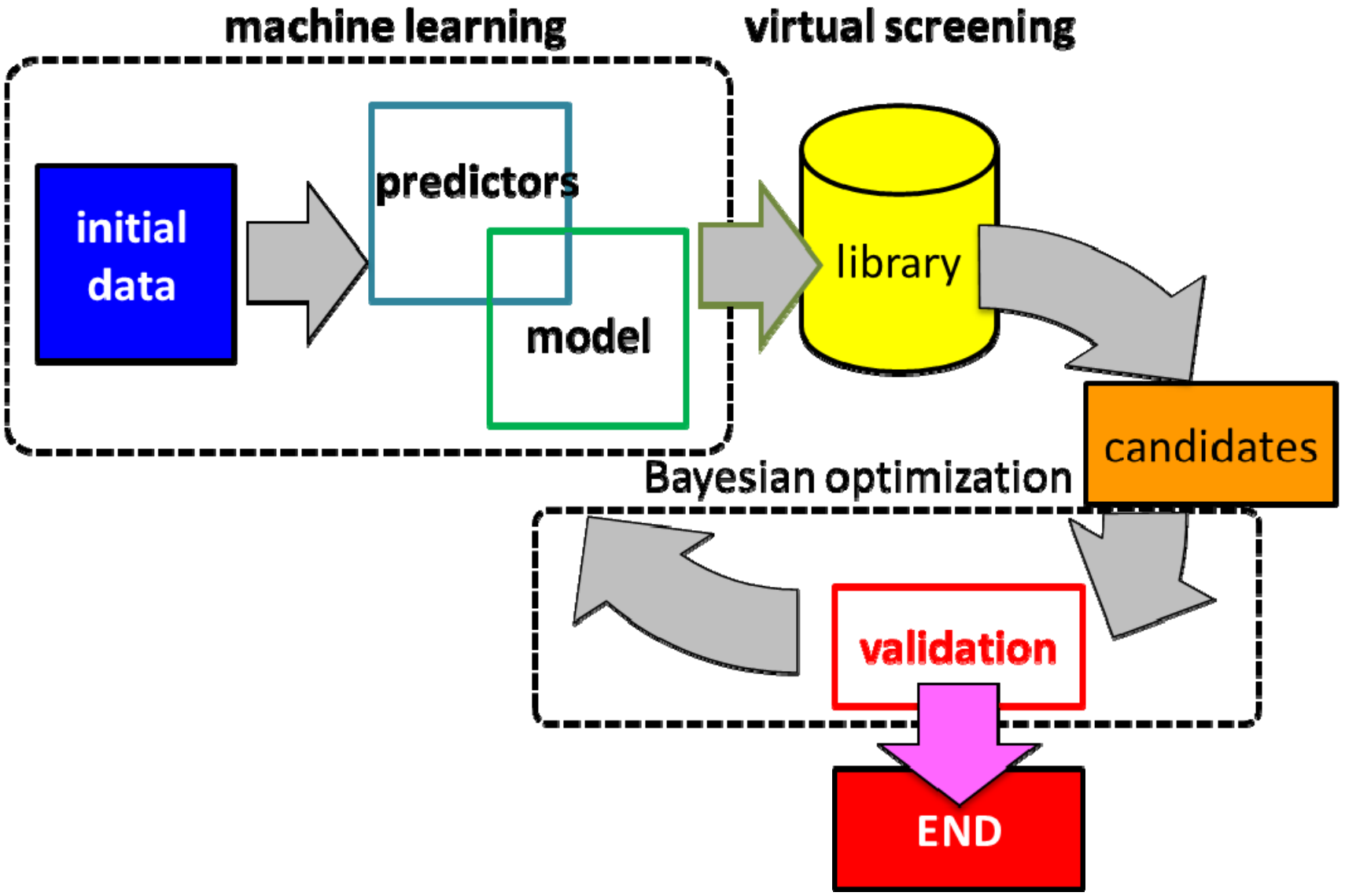
Θ_D: Debye temperature

γ: Grüneisen parameter

ab initio

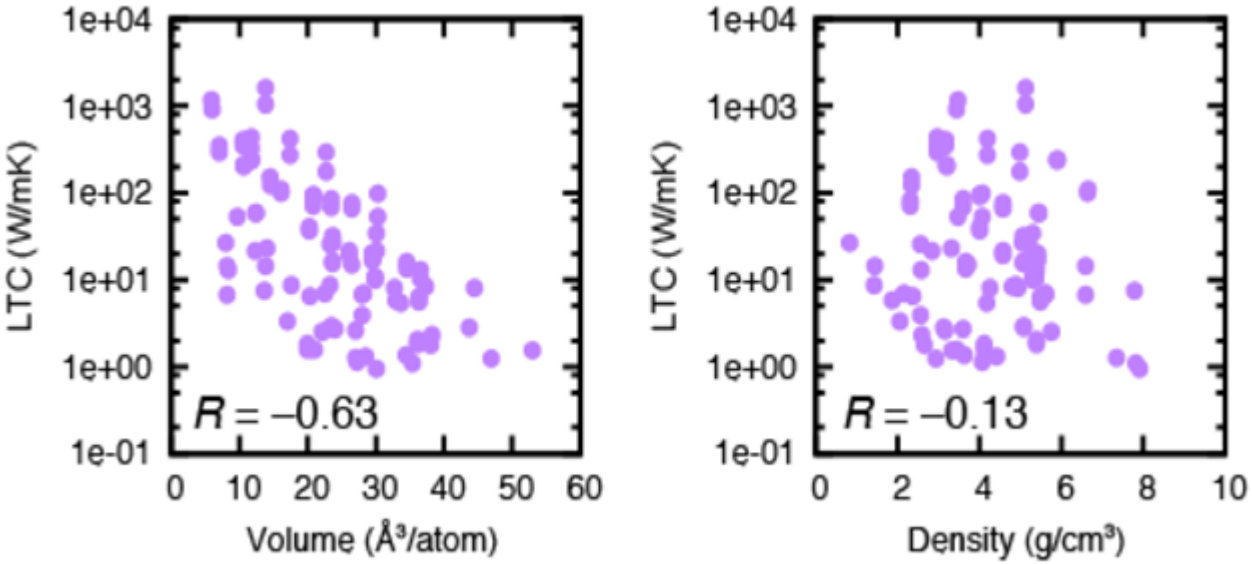
cannot be found just by a model based on harmonic phonon !

Virtual Screening



Simple descriptors for LTC

Model 1: Volume V and Density ρ

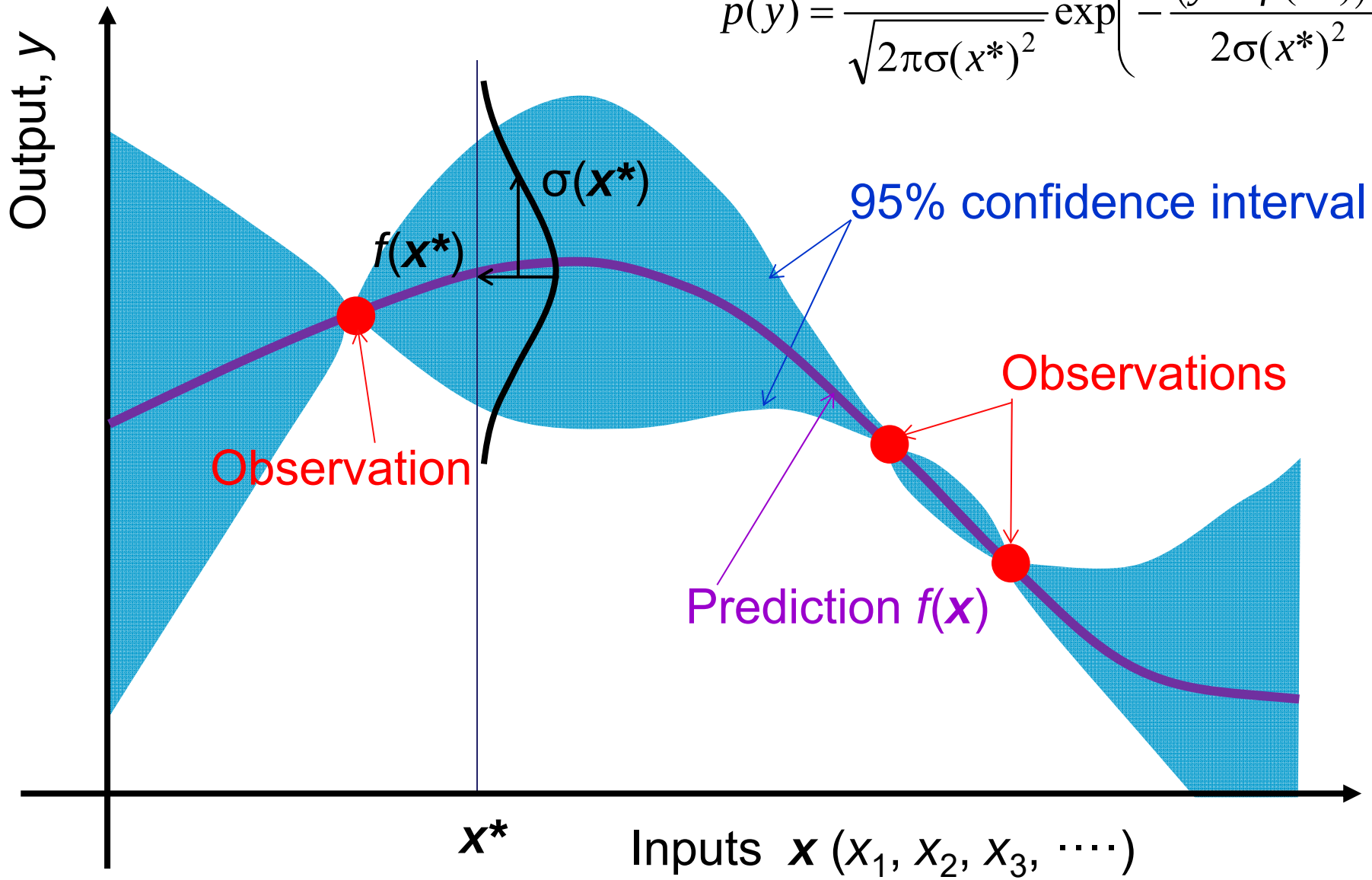


Model 2: Model 1 + 34 primitive elemental descriptors (one-hot)

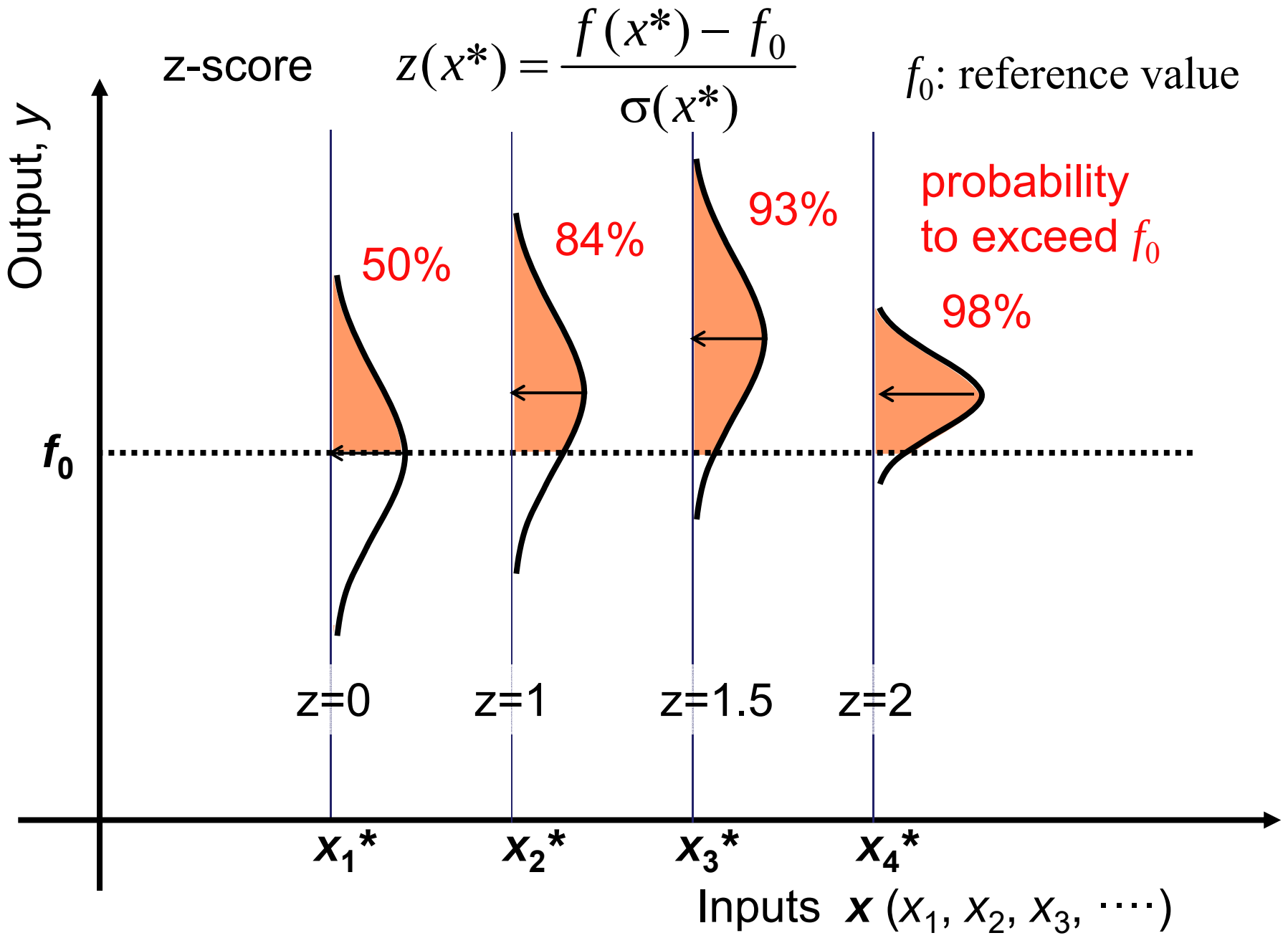
	H	Li	Be	B	C	N	O	F	…
LiH	1	1	0	0	0	0	0	0	
LiF	0	1	0	0	0	0	0	1	
BeO	0	0	1	0	0	0	1	0	
BN	0	0	0	1	0	1	0	0	
…									

Gaussian Process Regression

$$p(y) = \frac{1}{\sqrt{2\pi\sigma(x^*)^2}} \exp\left(-\frac{(y - f(x^*))^2}{2\sigma(x^*)^2}\right)$$



Gaussian Process Regression



Ranking of LTC for 54,779 compounds in MPD library

$$z(x^*) = \frac{f(x^*) - f_0}{\sigma(x^*)} \qquad f = -\log \kappa_L$$

κ_L : lattice thermal conductivity

$\sigma(\boldsymbol{x}^*)$: standard deviation of prediction

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

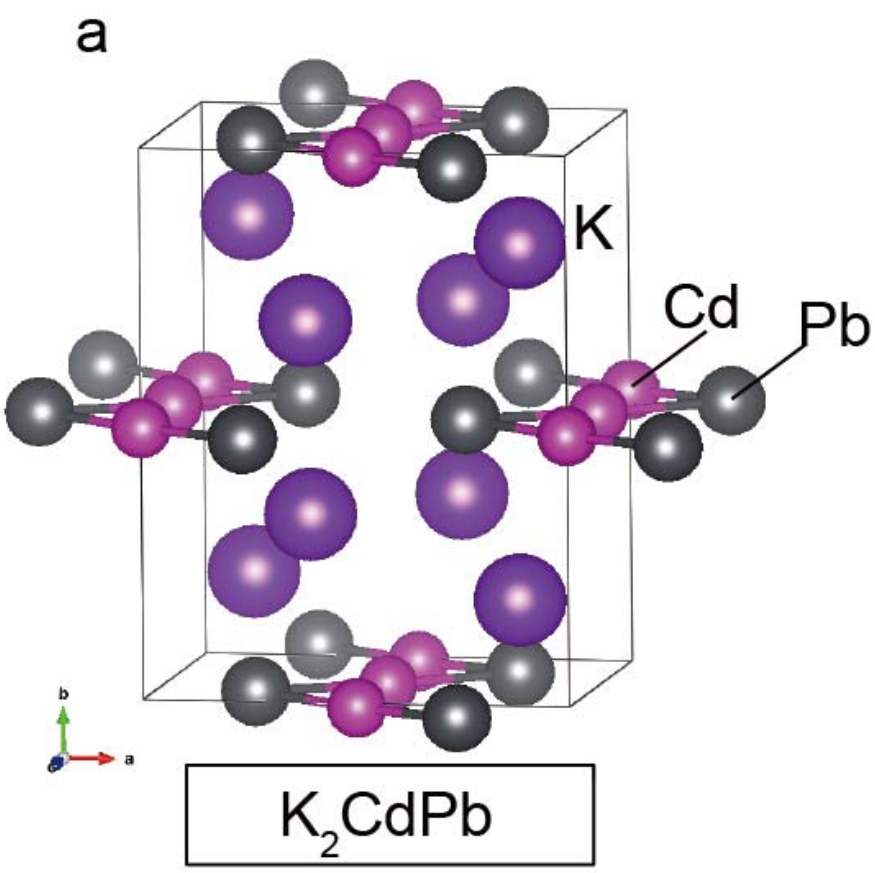
Virtual screening of 54,779 compounds in MPD library

ranking	Z-score	compound		*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

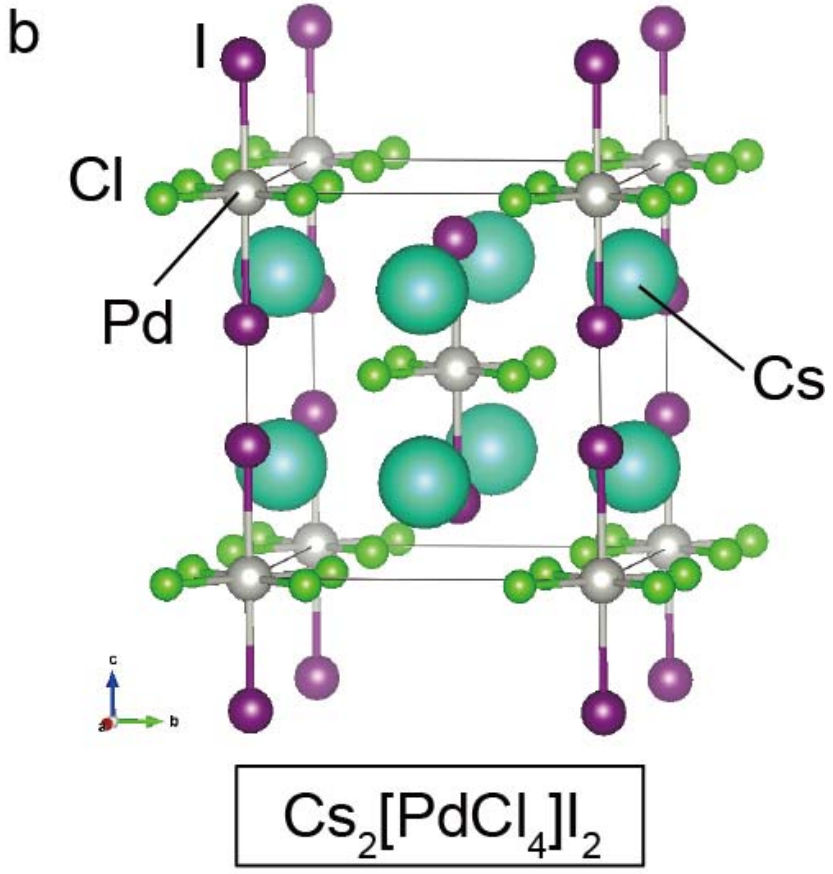
Newly discovered candidates for thermoelectrics

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.

Modelling of Lattice Thermal Conductivity

Kernel Ridge Regression (KRR)

Seko et al. PRB 95 (2017) 144110

Descriptors	RMS Error (log κ)
V, ρ	0.545
V, ρ + Primitive elemental descriptors	0.220
Elemental descriptors **+ GRDF**	0.102
Elemental descriptors **+ BOOP**	0.096

GRDF: Generalized Radial Distribution Function

BOP: Bond-Orientational Order Parameter

Test data = Randomly-selected 10 % of whole data, 200 trials

(Whole data = 101 compounds)

Lattice Thermal Conductivity (LTC)@300K

First principles calculation vs. Experimental data

phono3py