MI²I FINAL MEETING
20 FEBRUARY 2020
HITOTSUBASHI-KODO TOKYO

Materials Exploration Group

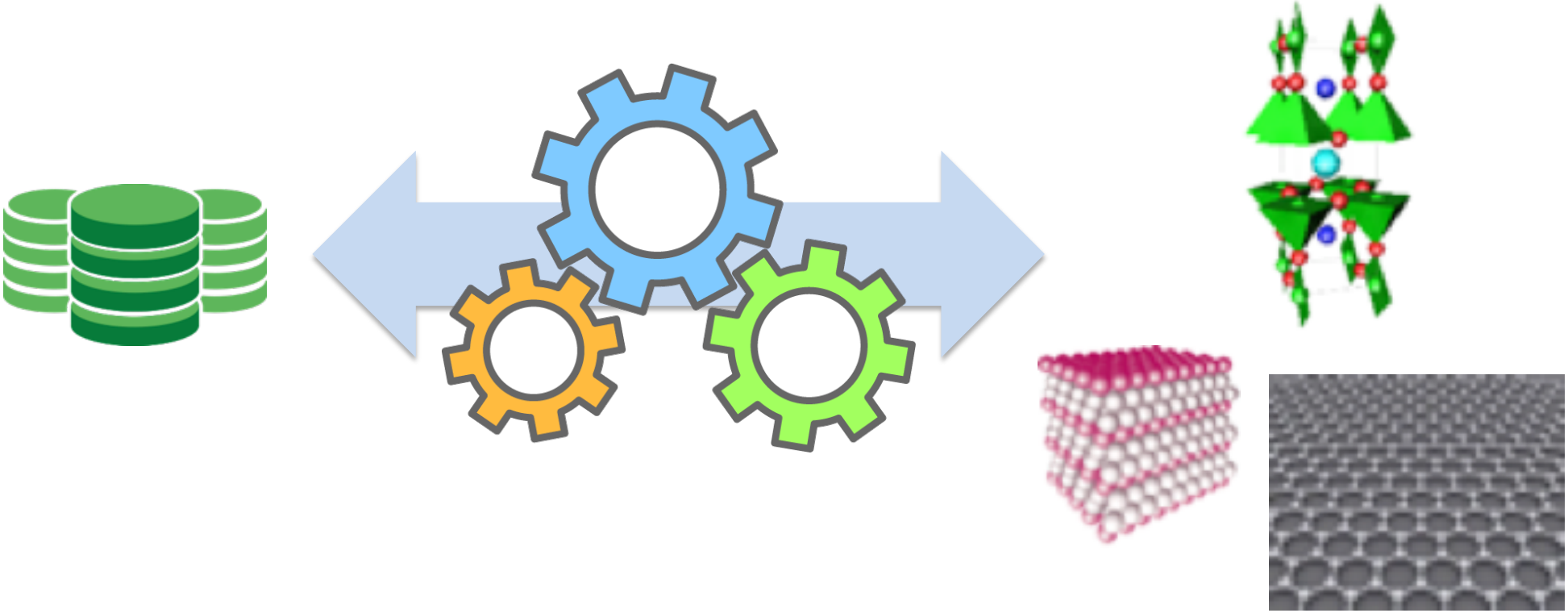
Tamio Oguchi

ISIR, Osaka University

MaDIS, NIMS

Materials Exploration Group

Development and Applications of Data-Driven Tools for Materials Exploration



Outline

1. First-Principles Calculations
2. *Crystal Structure Prediction Tool — CrySPY*
3. *Descriptor Generation Tool — LIDG*
4. Data-Driven Materials Exploration

First-Principles Calculations

- Starting from first principles of quantum mechanics — Density Functional Theory (DFT)
- Some basic approximations assumed
- Calculations of electronic structure and properties for realistic materials systems

✓ General Purpose : Methods do *NOT* depend on materials systems or properties investigated

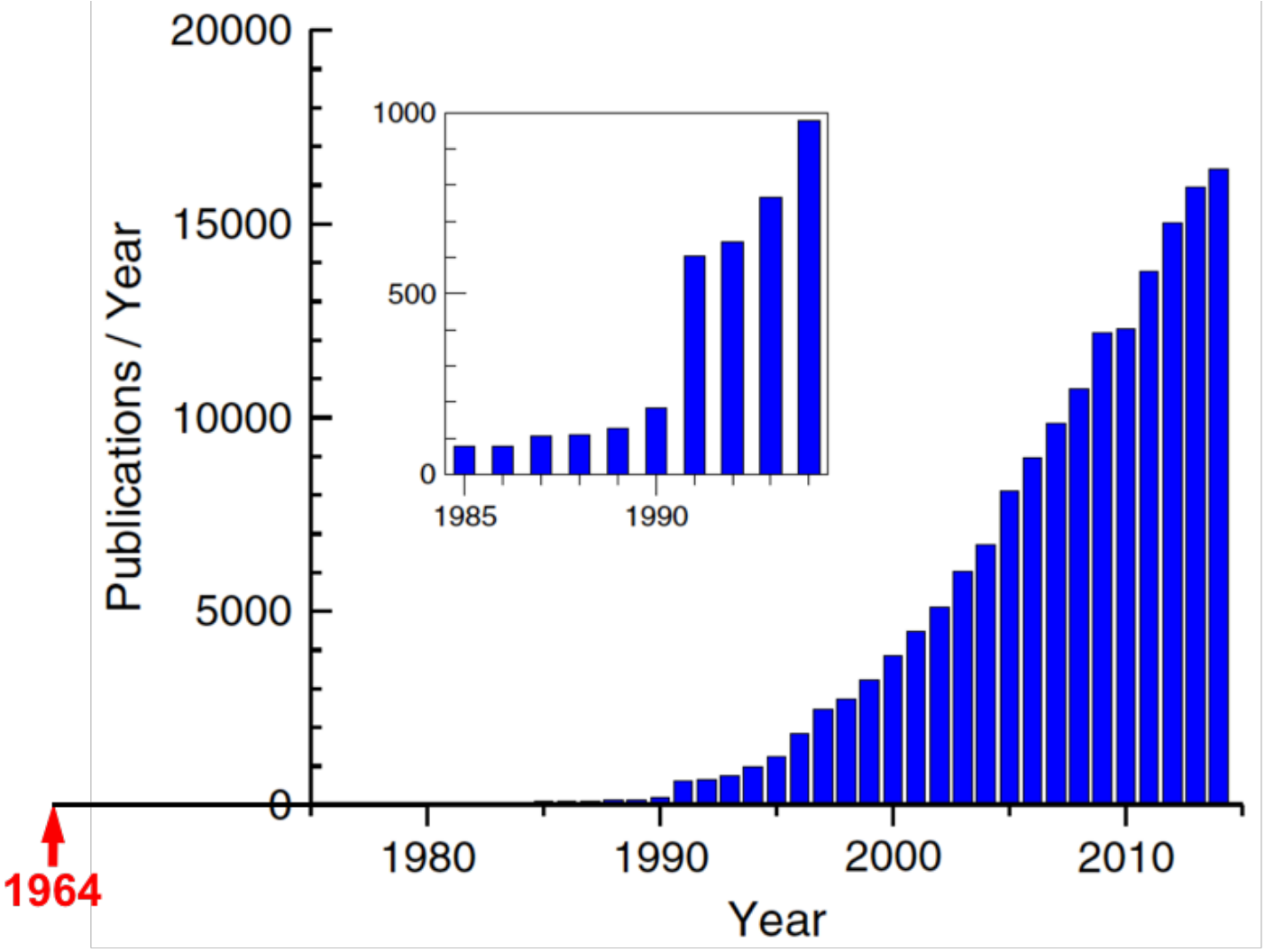
✓ Non-empirical : Methods do *NOT* include empirical parameters as input

➢ Prediction of properties comparable with experiments

➢ Clarification of physical mechanisms

★ Generation of homogeneous data ➡ Integration with ML (Automated HPC)

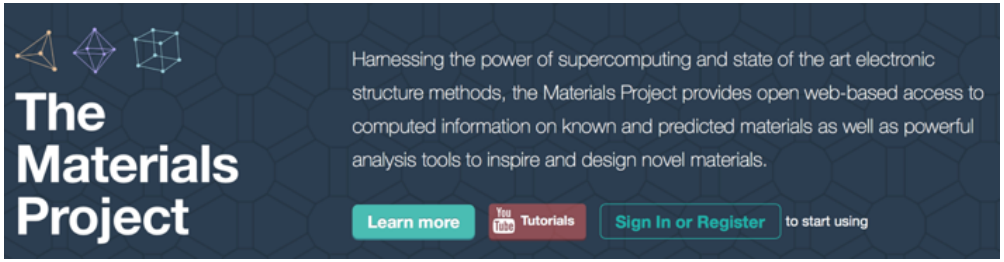
Rapidly Emerging DFT Calculations



Number of publications per year (1975–2014) on topics ("density functional" or "DFT"), according to the Web of Science Core Collection (February 2015)

Jones (2015)

Materials Database by First-Principles Calculations



https://materialsproject.org

OQMD:

The Open Quantum Materials Database

http://oqmd.org

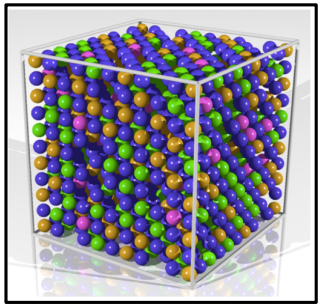
https://nomad-coe.eu

http://aflowlib.org

First-Principles Calculations

Elements & Composition

$A_mB_n...X$

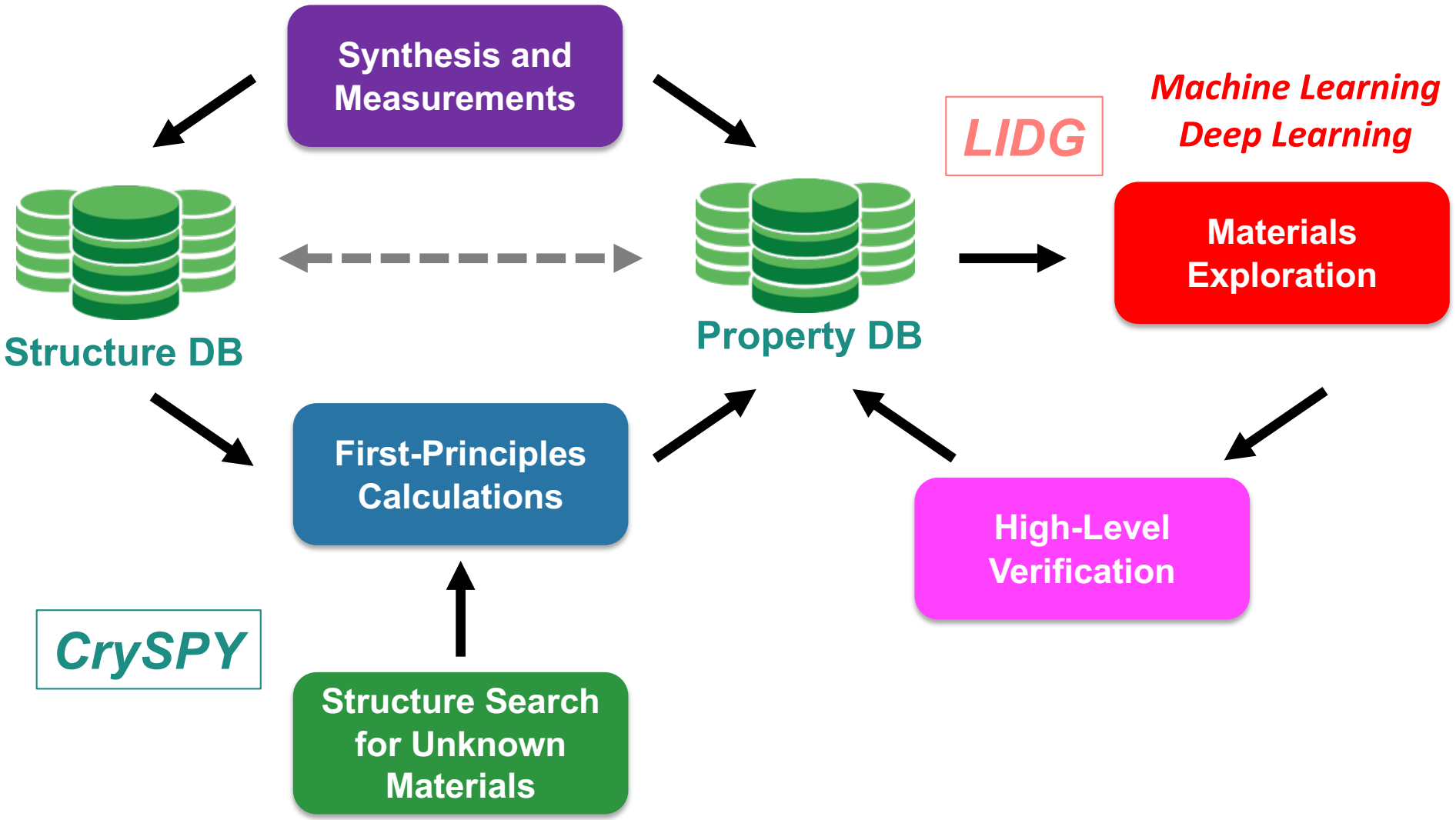


Structure

First-Principles Calculations

Property

Data-Driven Materials Exploration



Crystal Structure Prediction Tool

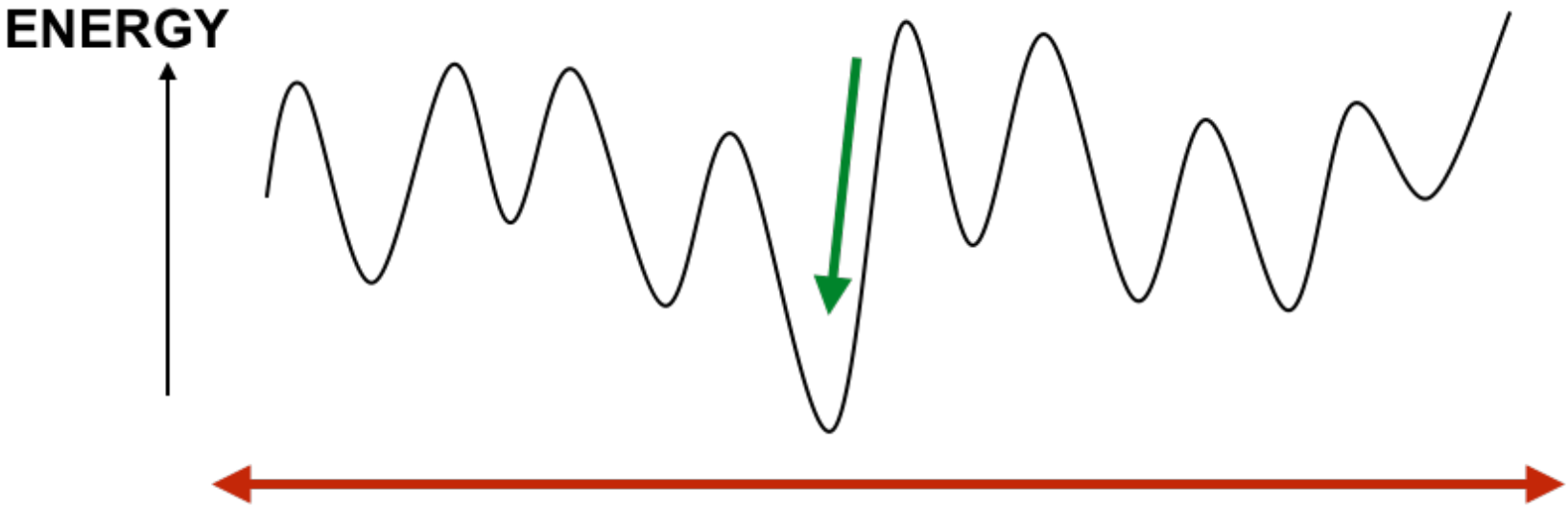
CrySPY

Tomoki Yamashita (Nagaoka U of Technology)

Keys of Structure Prediction

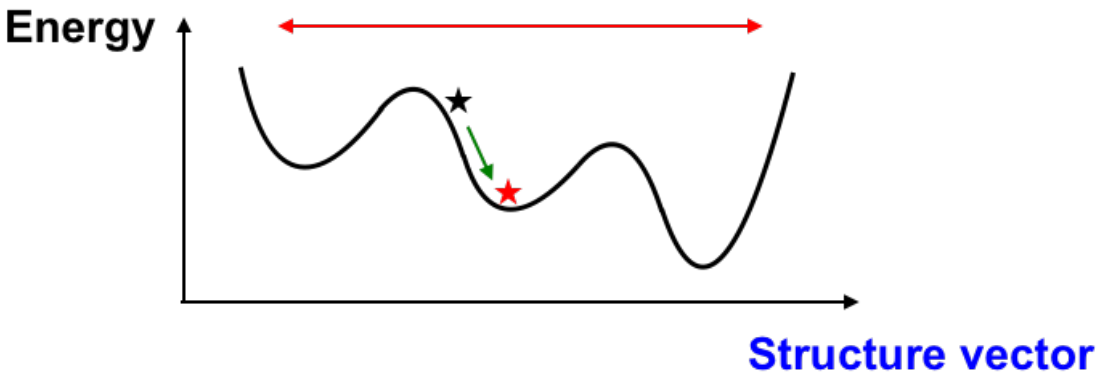
- **Exploration**
  - ✓ to search the structure space as globally as possible

- **Exploitation**
  - ✓ to search the structure space without missing important minima



Algorithm of Structure Prediction

- **Random Search**
  - ✓ Random selection of lattice constants and atomic positions
- **Evolutional-type Search**
  - ✓ Evolutional Algorithm（USPEX）
  - ✓ Particle Swarm Optimization（CALYPSO）
- **Learning-type Search**
  - ✓ Bayesian Optimization
    - ➢ Optimization for global search of an unknown function



Development of Crystal Structure Prediction

CrySPY

https://github.com/Tomoki-YAMASHITA/CrySPY
distributed under the MIT License

- **Algorithm**
  - ✓ Random search
  - ✓ Bayesian optimization
  - ✓ Evolutionary algorithm
  - ✓ LAQA: Look Ahead based on Quadratic Approximation

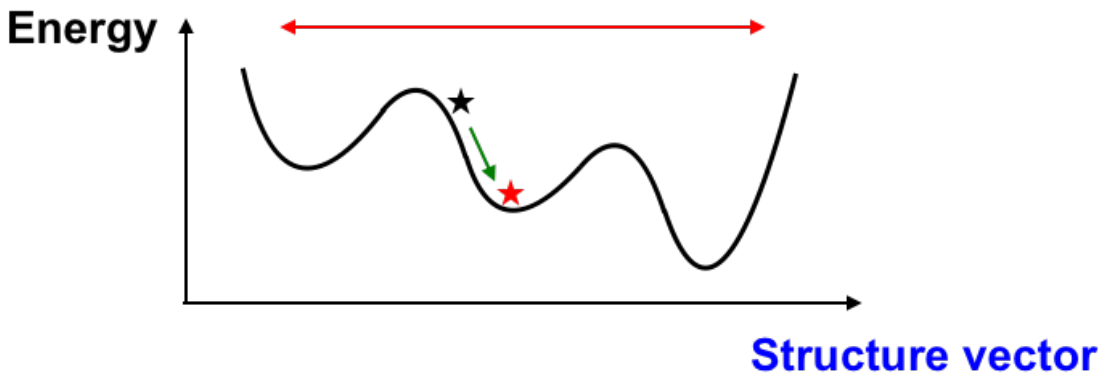
    A fine-grained method to reduce local optimization steps

Space group by find_wy
H. Kino, https://github.com/nims-hrkn/find_wy

Bayesian optimization library, COMBO
T. Ueno *et al.*, Materials Discovery 4, 18 (2016).
https://github.com/tsudalab/combo

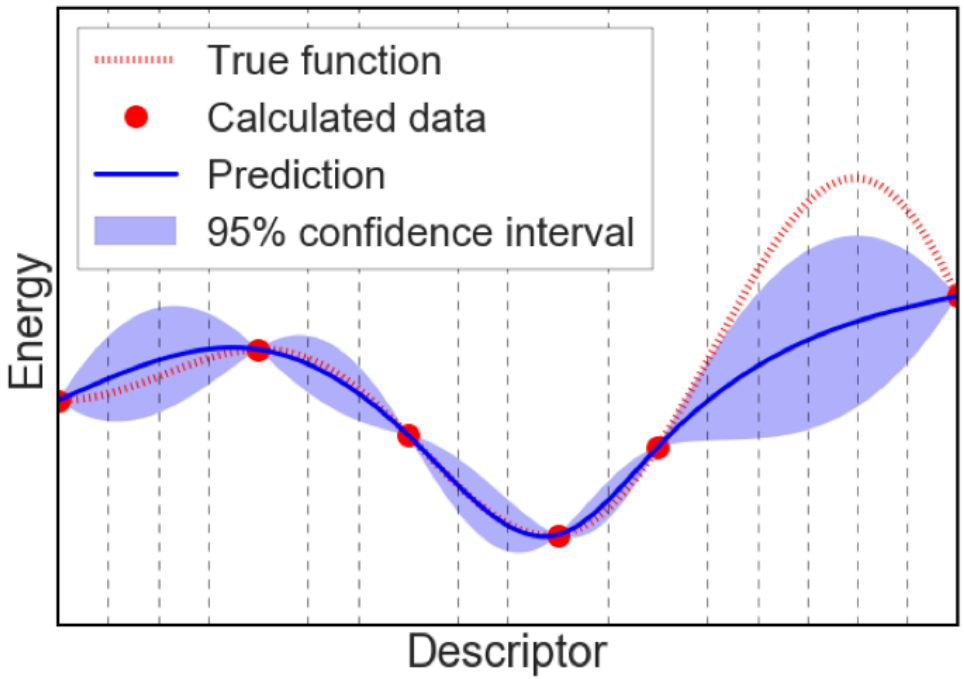
- **Local optimization**

  Interfaced with
  - ✓ VASP
  - ✓ Quantum Espresso
  - ✓ soiap
  - ✓ LAMMPS



Bayesian Optimization

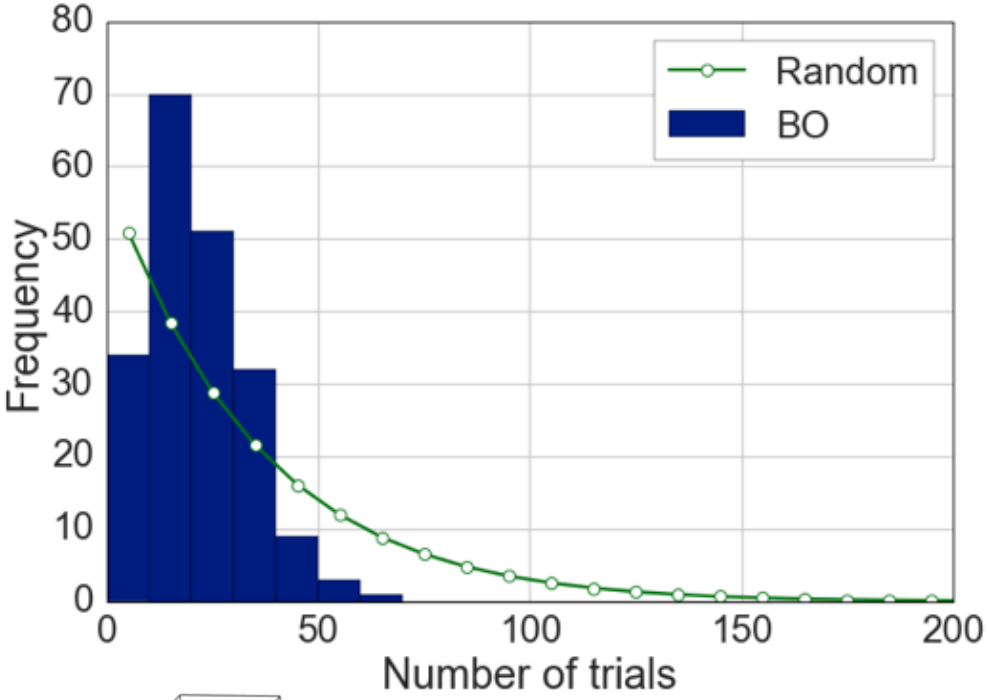
A sequential design strategy for global optimization of black-box functions that doesn't require derivatives. [Wikipedia]

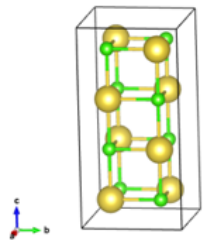


BO enables to balance trade-off between exploration and exploitation of the search algorithm.

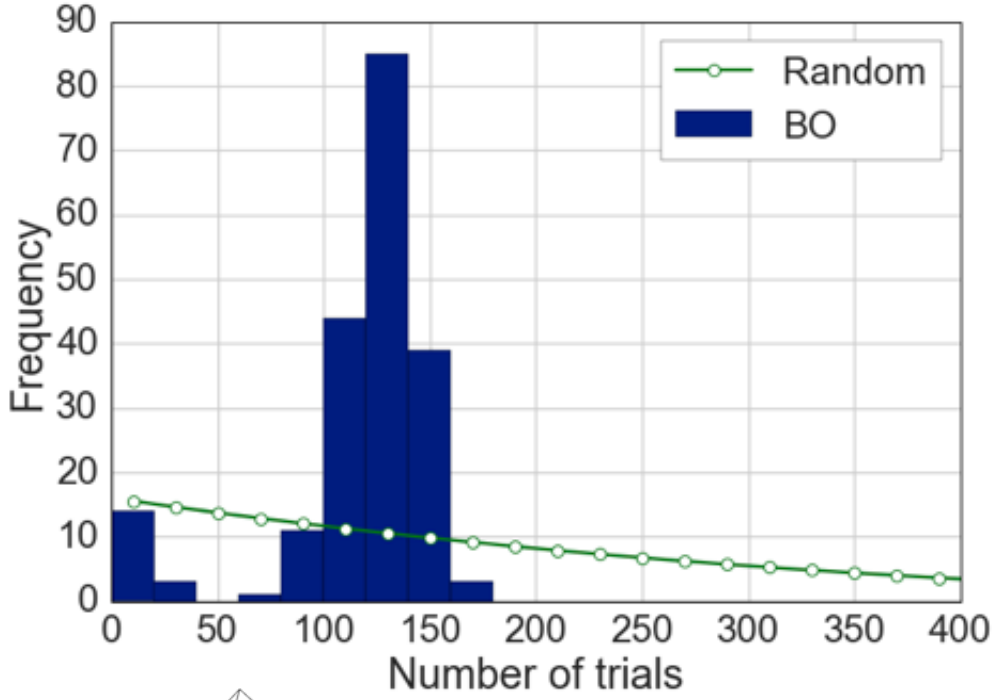
Examples of Structure Prediction

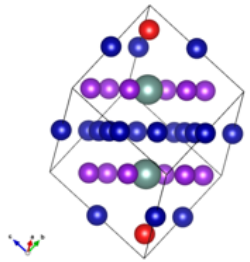
Comparison between random search and Bayesian Optimization（BO)





Na_8Cl_8 crystal



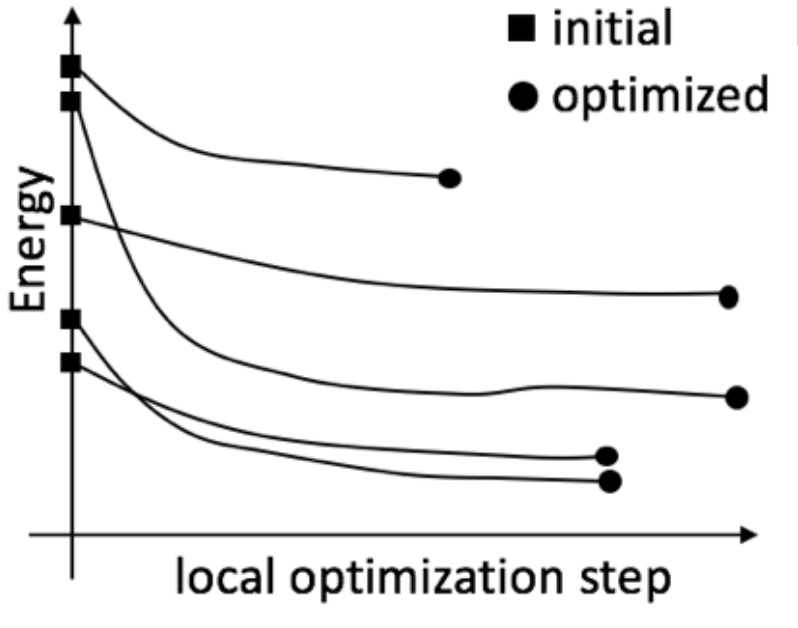


Y_2Co_{17} crystal

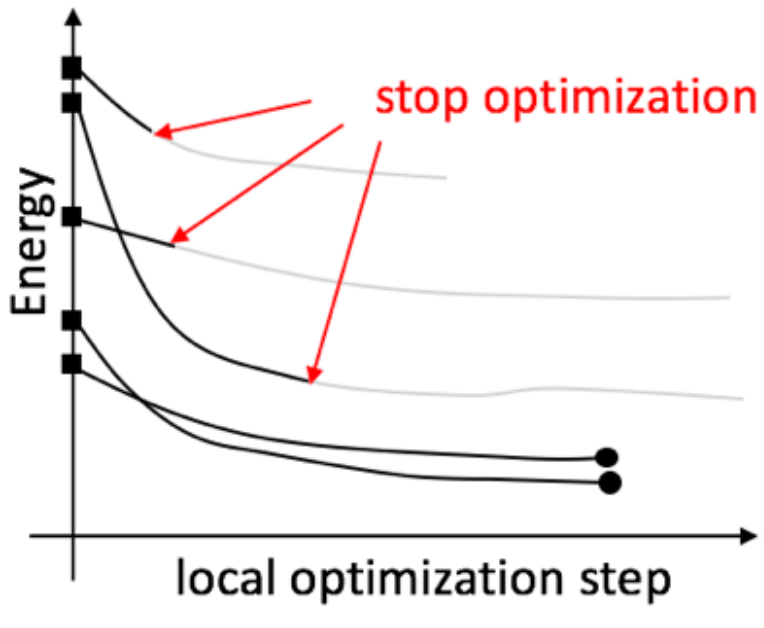
Phys. Rev. Materials 2, 013803 (2018).

Schematic Image of LAQA Algorithm

Standard Method

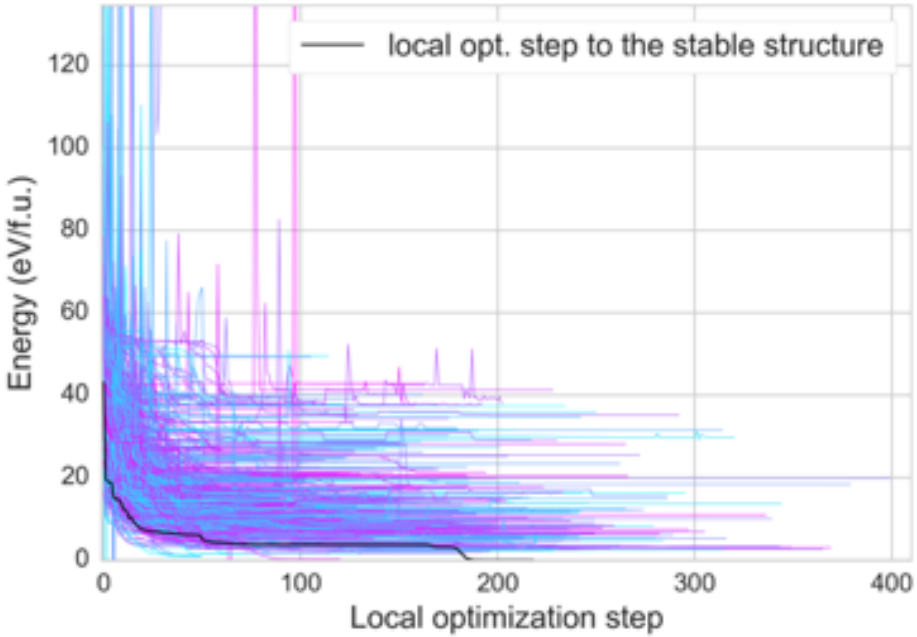


LAQA

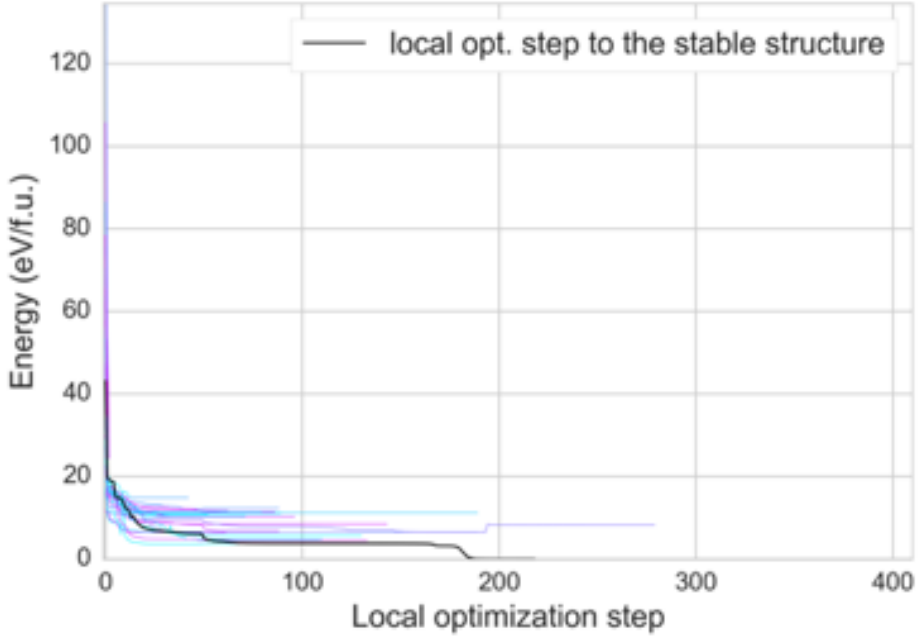


Local Optimization by LAQA Method

DFT calculations for Y_2Co_{17}
Total number of assumed structures: 700



Ordinary Method
91400 steps



LAQA
3300 steps

npj Computational Materials 4, 32 (2018).

Descriptor Generation Tool

LIDG

Hitoshi Fujii (NIMS)

Goals of Materials Informatics

1. **Materials Exploration & Optimization:**
   Searching materials with desired property
   (Not necessarily important to build a model)

2. **Property Prediction:**
   Constructing models with non-linear (black-box function) or linear model (simple function) with appropriate descriptors

3. **Discovery of Underlying Physics:**
   Clarifying physical mechanisms toward the discovery of new concepts or laws → *interpretable modeling*

Meaning of Interpretable Model

- **Interpretation: Clear relation between descriptors and target**

$$Y = F(X)$$

Hard to be obtained by a black-box function

- ✓ **Inefficient for optimizing the function**
- ✓ **Unidentified applicable region of descriptors**

It's easy to guess the behaviors of the obtained function with interpretable models.

- ✓ **Rough estimation of good candidates**
- ✓ **Prior identification of applicable region and limitation**
- ✓ **Possible way to go beyond the limitation**

Development of Linearly Independent Descriptor Generation Method

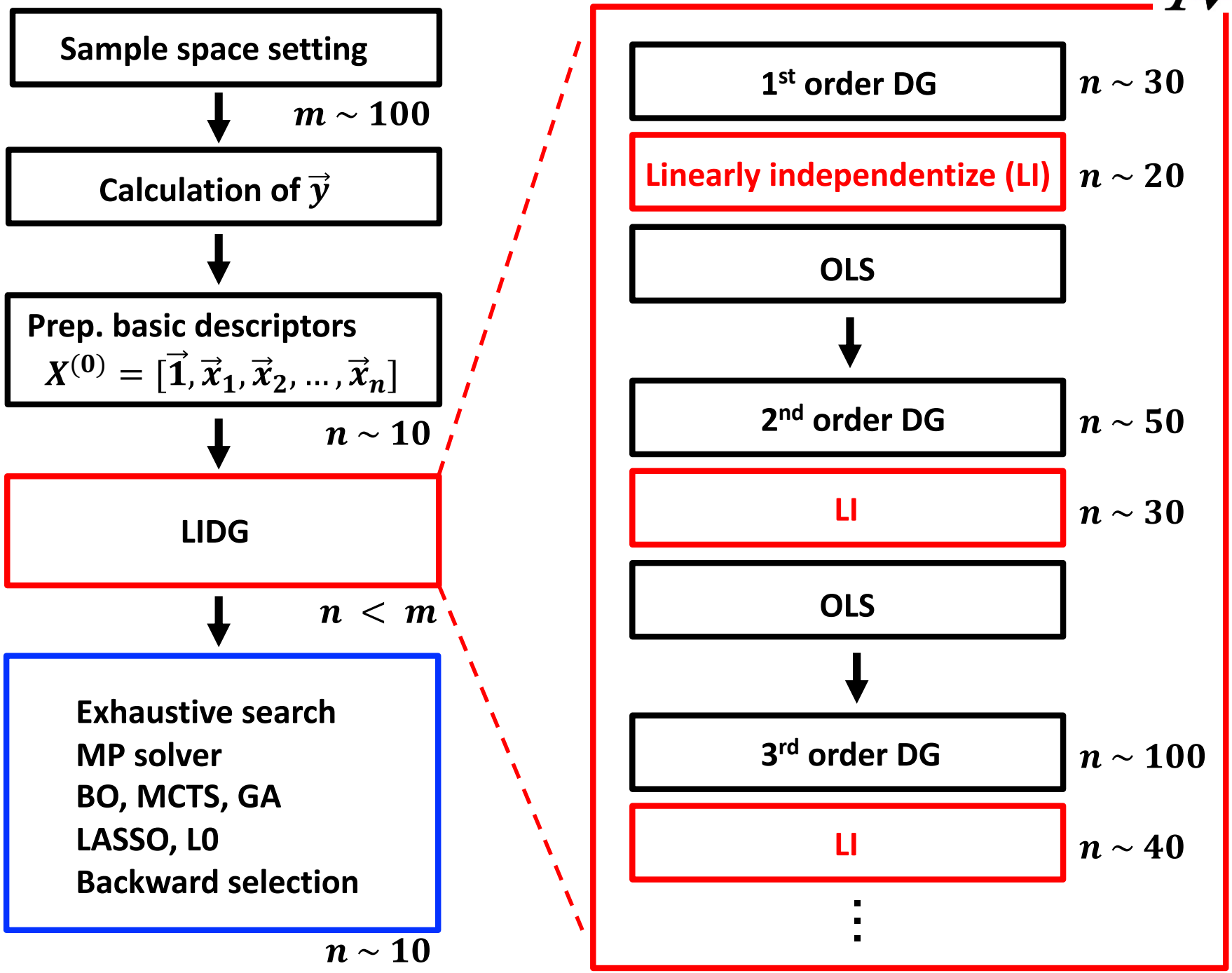
LIDG

https://github.com/Hitoshi-FUJII/LIDG

- **Algorithm**
  - ✓ Generation of higher-order descriptors by products
  - ✓ Symmetrization of descriptors with RDB
  - ✓ Detection and removal of multi-collinearity

- **Advantages in applications**
  - ✓ Step by step generation of descriptors in regression
  - ✓ Construction of simple, interpretable models with sparse modeling method
  - ✓ Usable as a preprocessing of machine learning

LIDG


Sample space setting
$m \sim 100$
Calculation of $\vec{y}$
Prep. basic descriptors
$X^{(0)} = [\vec{1}, \vec{x}_1, \vec{x}_2, \dots, \vec{x}_n]$
$n \sim 10$
LIDG
$n < m$
Exhaustive search
MP solver
BO, MCTS, GA
LASSO, L0
Backward selection
$n \sim 10$
1st order DG
$n \sim 30$
Linearly independentize (LI)
$n \sim 20$
OLS
2nd order DG
$n \sim 50$
LI
$n \sim 30$
OLS
3rd order DG
$n \sim 100$
LI
$n \sim 40$
⋮

Big Data of Materials Science: Critical Role of the Descriptor

Luca M. Ghiringhelli,[1,*] Jan Vybiral,[2] Sergey V. Levchenko,[1] Claudia Draxl,[3] and Matthias Scheffler[1]
[1]*Fritz-Haber-Institut der Max-Planck-Gesellschaft, 14195 Berlin-Dahlem, Germany*
[2]*Department of Mathematical Analysis, Charles University, 18675 Prague, Czech Republic*
[3]*Humboldt-Universität zu Berlin, Institut für Physik and IRIS Adlershof, 12489 Berlin, Germany*
(Received 14 April 2014; revised manuscript received 20 October 2014; published 10 March 2015)

Regression of total energy difference between zinc-blende (ZB) and rock-salt (RS) crystal structures in 82 octet binary compound semiconductors *AB*

$$\Delta E = E(RS) - E(ZB)$$

Extracted important features from the modeling

$$\frac{IP(B) - EA(B)}{r_p(A)^2}, \quad \frac{|r_s(A) - r_p(B)|}{\exp[r_s(A)]}, \quad \frac{|r_p(B) - r_s(B)|}{\exp[r_d(A)]}$$

Sparse Modelling of Chemical Bonding with LIDG

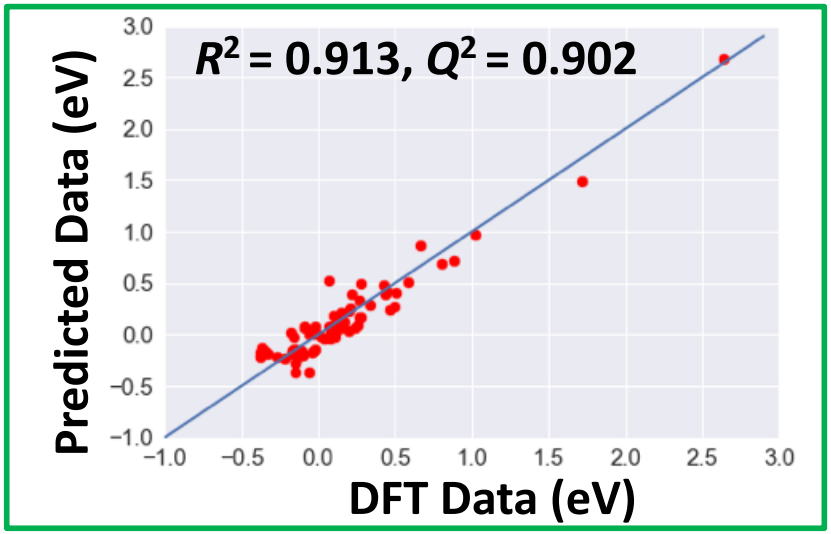
- **Basic features**
  - ✓ **Atomic radius, Electron affinity, Ionization potential, Electronegativity**

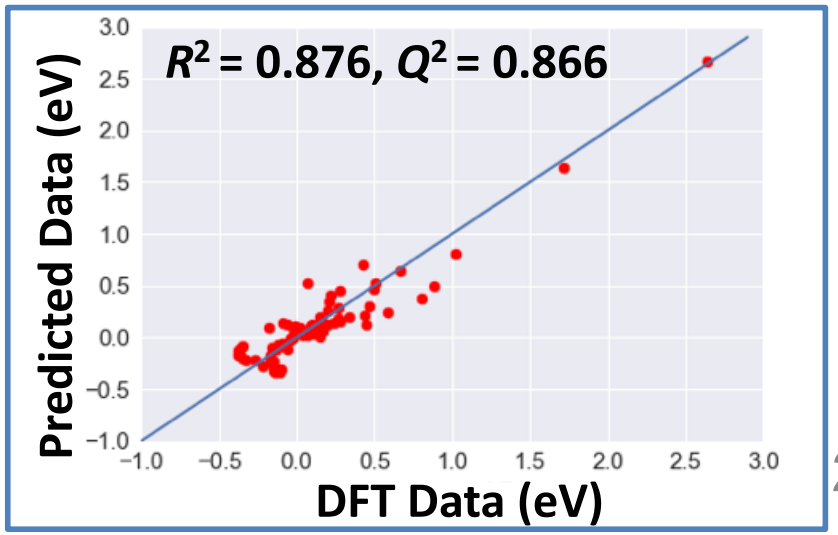
★ **Best model with 3 descriptors up to 2nd order of all basic features**

$$E(RS) - E(ZB) \cong 0.59\left|\mathrm{EN}_A - \mathrm{EN}_B\right| - 1.95\frac{\left|\mathrm{EN}_A - \mathrm{EN}_B\right|}{r_A + r_B} + 6.15\frac{1}{\left(r_A + r_B\right)^2} - 0.75$$

★ **Best model with 2 descriptors up to 4th order of one basic feature**

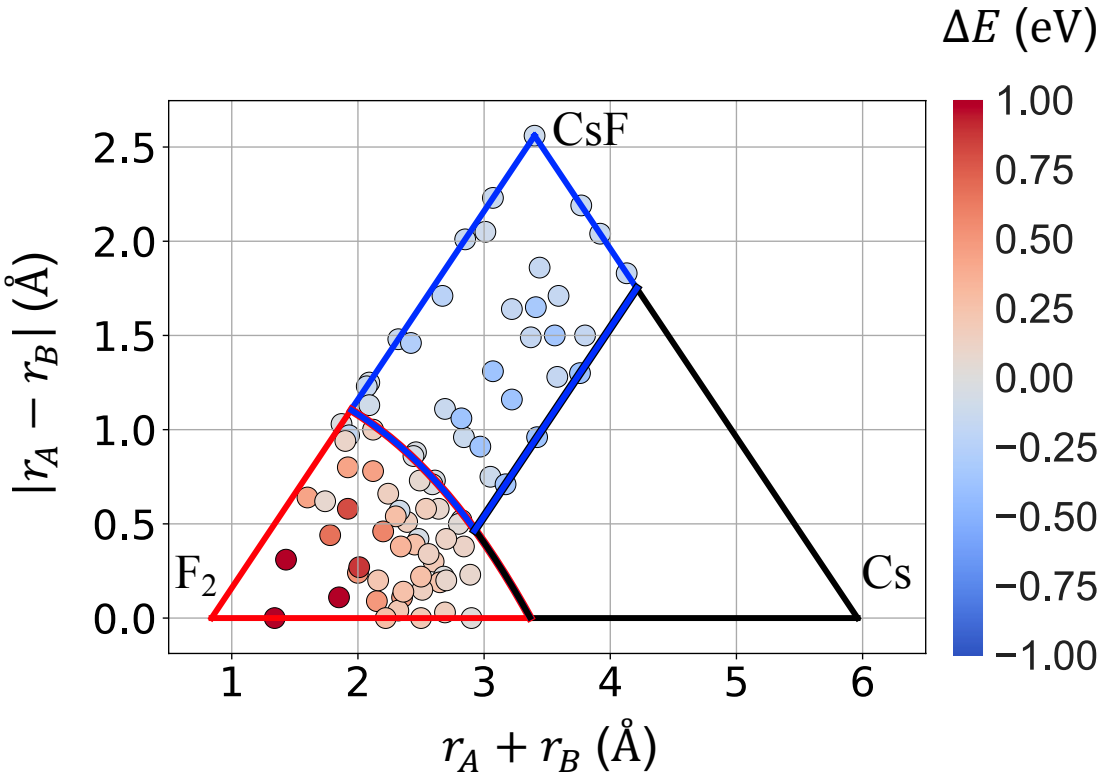
$$E(RS) - E(ZB) \cong 6.87\frac{1}{\left(r_A + r_B\right)^3} - 5.02\frac{\left|r_A - r_B\right|}{\left(r_A + r_B\right)^3} - 0.18$$





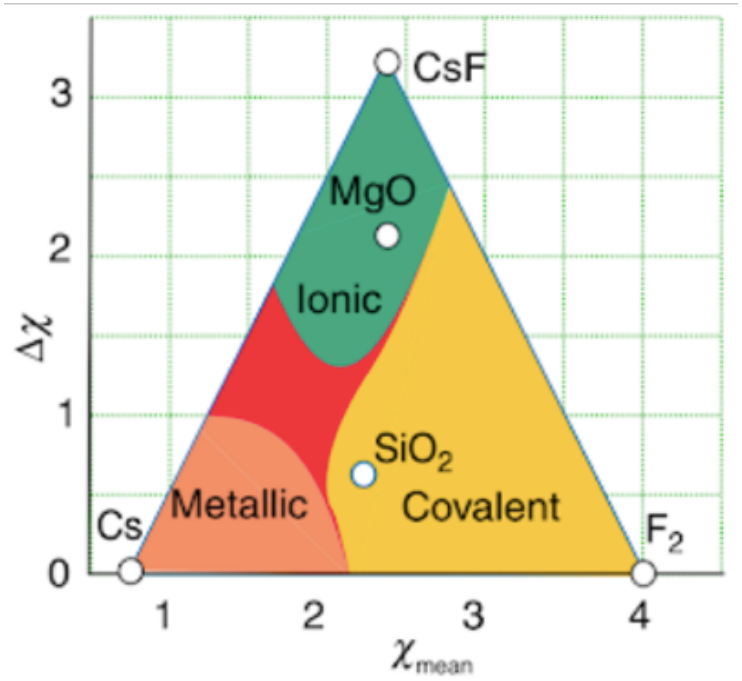
Sparse Modelling of Chemical Bonding with LIDG

$$E(RS) - E(ZB) \cong 6.87\frac{1}{(r_A + r_B)^3} - 5.02\frac{|r_A - r_B|}{(r_A + r_B)^3} - 0.18$$



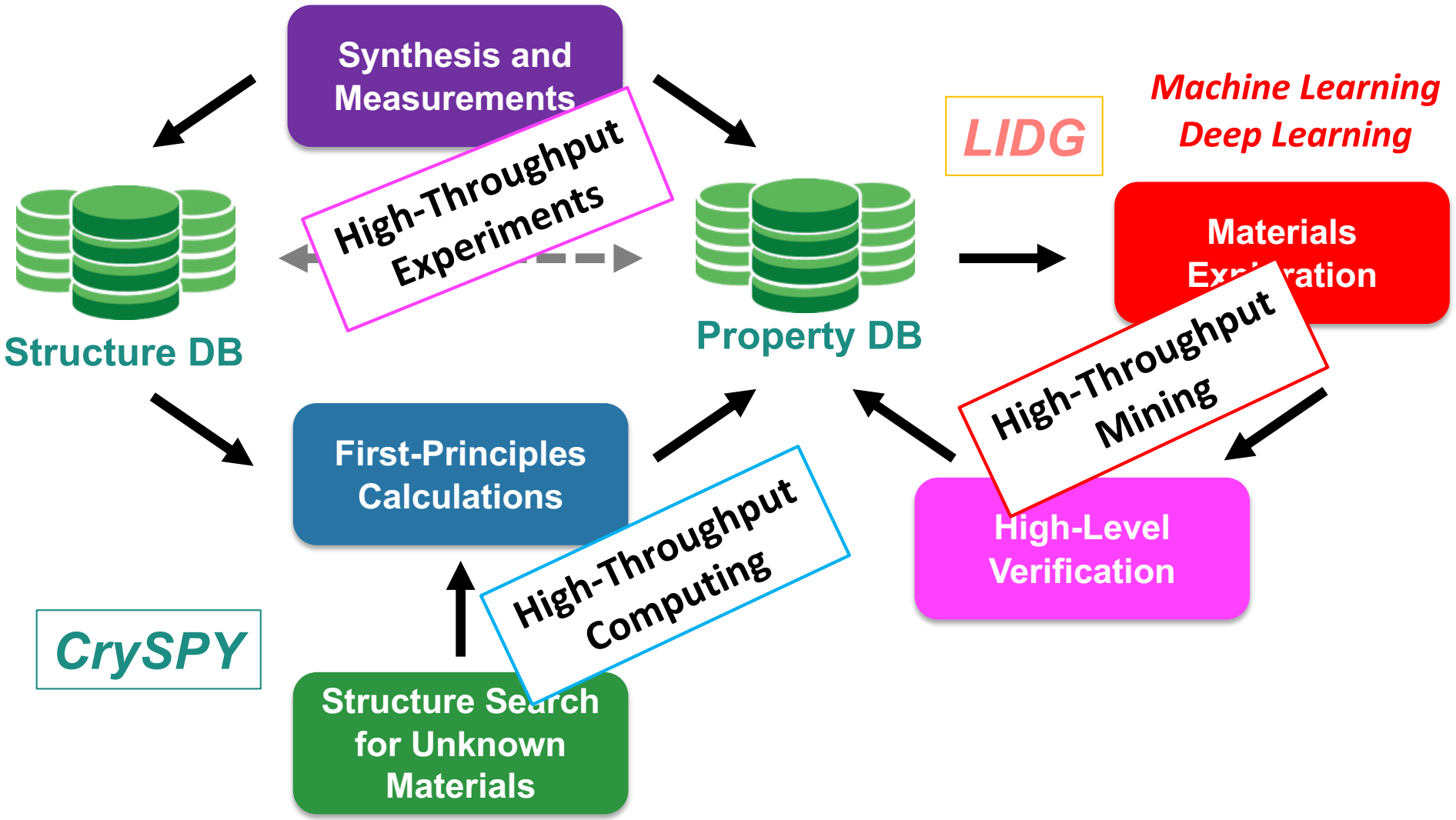
Sci. Tech. Adv. Mater. 20, 1178 (2019).

van Arkel-Ketelaar Triangle for Chemical Bonding



Inorganic Chemistry. 7th ed. Oxford University Press, 2018.

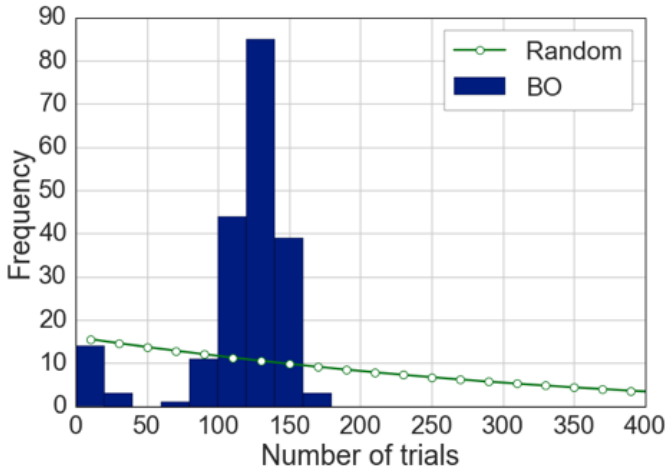
Data-Driven Materials Exploration



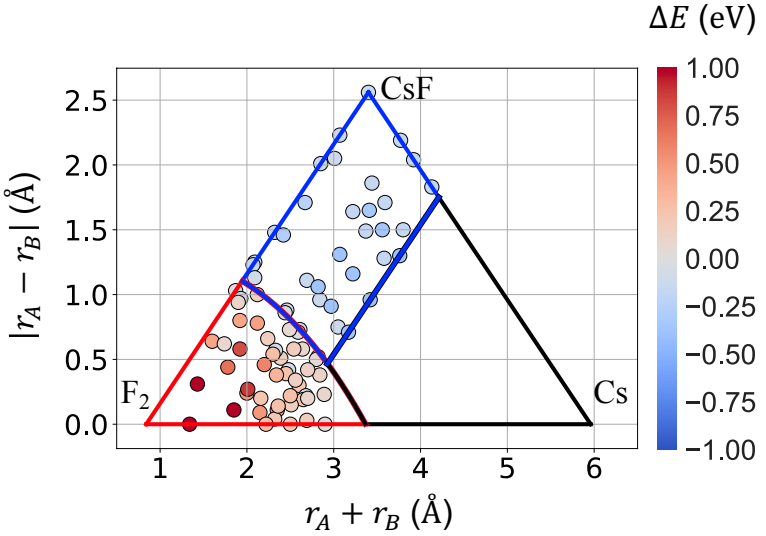
MI²I FINAL MEETING
20 FEBRUARY 2020
HITOTSUBASHI-KODO TOKYO

Thank you for your attention

CrySPY



LIDG



ISIR OSAKA UNIVERSITY

Mi²i

MaDIS
NIMS MATERIALS DATA and INTEGRATED SYSTEM

NIMS