Introduction to Crystal Structure Prediction

19 February 2019

34th Computational Materials Design(CMD®) Workshop

Tomoki Yamashita[1, 2]

1. Research and Services Division of Materials Data and Integrated System (MaDIS), National Institute for Materials Science (NIMS), Tsukuba, Japan
2. Institute of Scientific and Industrial Research (ISIR), Osaka University, Ibaraki, Japan

This work is supported by

Materials research by
Information Integration Initiative

Collaborators

Tamio Oguchi[1, 2]
Shinichi Kanehira[2]
Nobuya Sato[3]
Takashi Miyake[1, 3]
Hiori Kino[1]
Kei Terayama[4]
Koji Tsuda[1, 4]

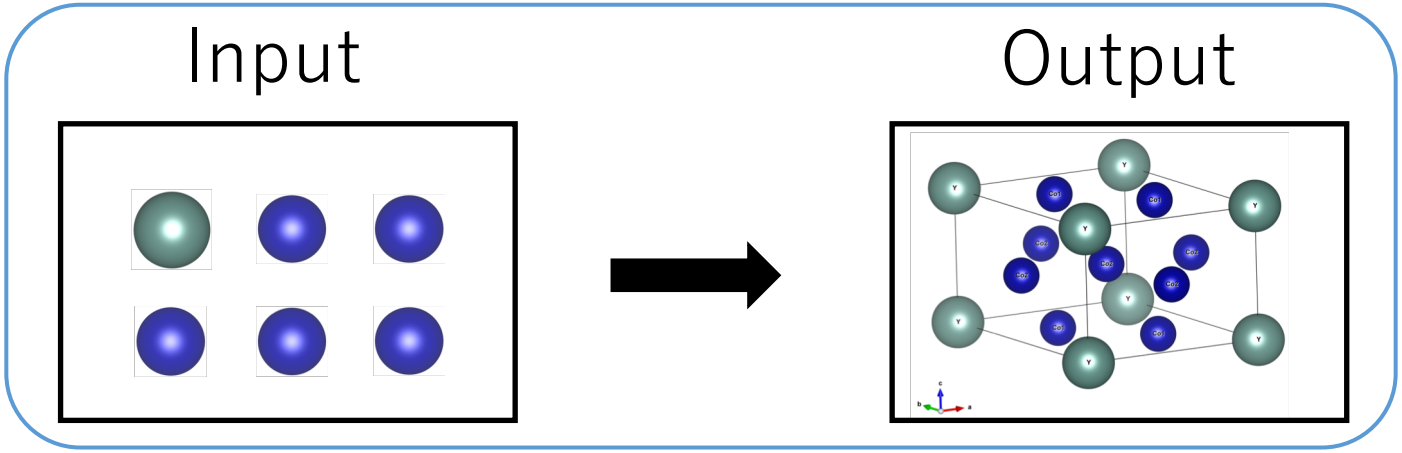
1. MaDIS, NIMS, Tsukuba, Japan
2. ISIR, Osaka University, Ibaraki, Japan
3. National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan
4. The University of Tokyo, Kashiwa, Japan

Outline

1. Introduction
   - Crystal structure prediction (CSP)
   - Propose a selection-type algorithm
2. Bayesian optimization
   - Descriptor
   - Algorithm
   - Searching efficiency
3. LAQA
   - Algorithm
   - Searching efficiency
4. Summary

Crystal structure prediction

Crystal structure prediction (CSP) is a powerful tool to obtain the most stable structure



CSP enables us to

design new materials without experiments

Possible to predict the physical property using first-principles calculations

Crystal structure prediction

CSP is carried out
with structure generation and optimization

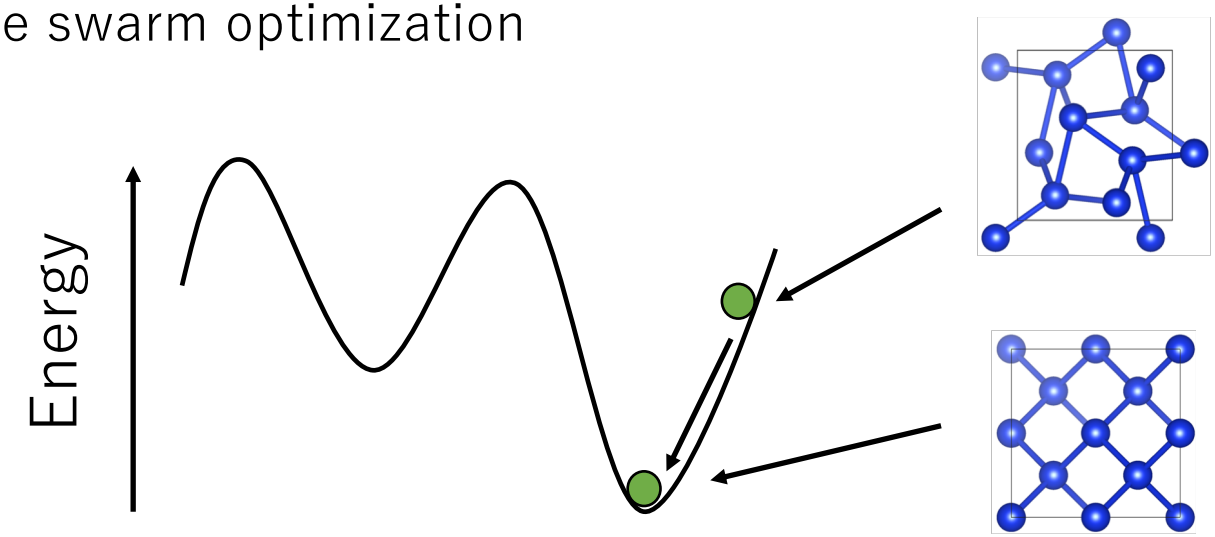
Generation

- Random
- Evolutionary algorithm
- Particle swarm optimization

etc.

Optimization

- First-principles calculation



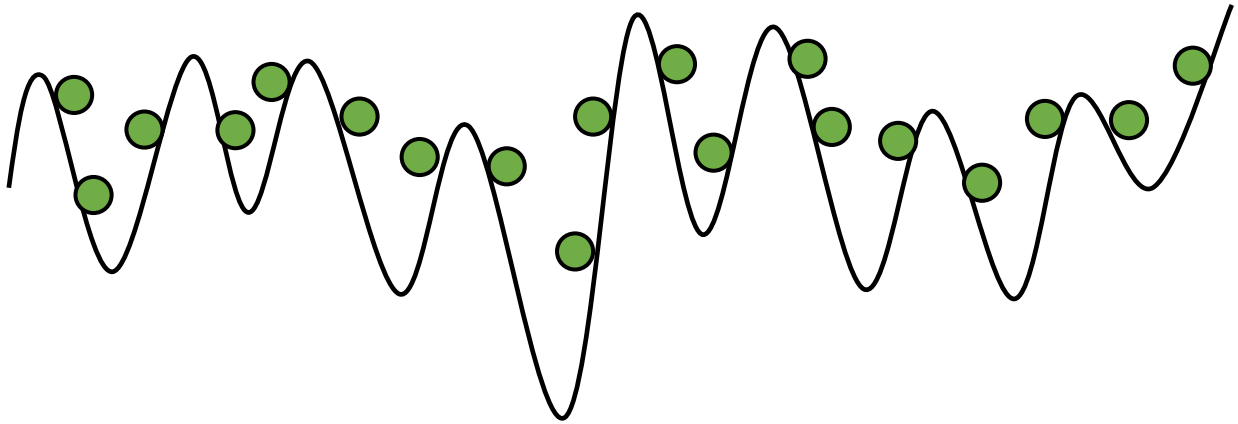
Obtain a local minimum structure by optimizing the initial structure

Crystal Structure Prediction

Problem

Difficult to predict in large systems

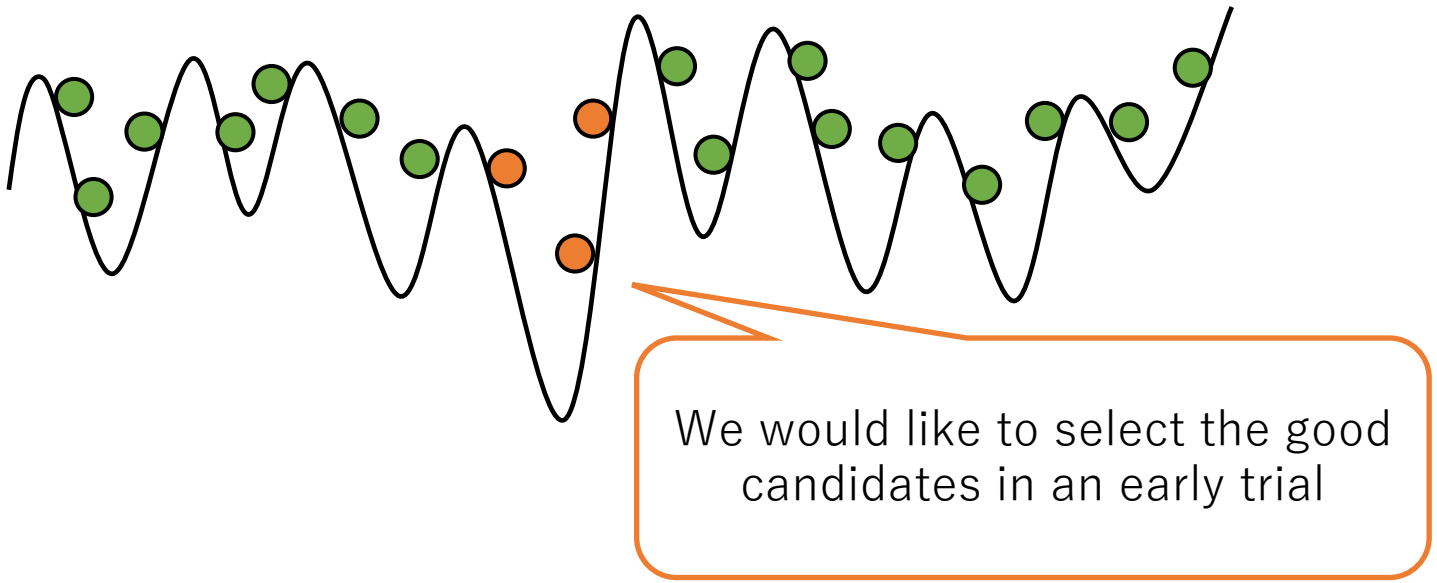
- A vast number of candidates are required
- Optimizing all candidates is computationally expensive



Development of more efficient searching algorithms is highly desired

Selection-type algorithm

Propose new algorithms: selection-type algorithm



- Bayesian optimization (BO)
  - ✓ Well established technique for black-box optimization
- LAQA
  - ✓ Our original algorithm

Code development

CrySPY

https://github.com/Tomoki-YAMASHITA/CrySPY

CrySPY is a Crystal Structure Prediction tool written in Python, distributed under the MIT License

- Structure generation
- Submitting jobs for structure optimization
- Collecting data for structure optimization
- Selecting candidates using machine learning

- Algorithm
  - ✓ Random search
  - ✓ Evolutionary algorithm
  - ✓ **Bayesian optimization**
  - ✓ **LAQA**
- Interface for optimization
  - ✓ VASP
  - ✓ Quantum ESPRESSO
  - ✓ soiap
  - ✓ LAMMPS
- Bayesian optimization
  - ✓ COMBO

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

T. Yamashita, et al., Phys. Rev. Mater. **2**, 013803 (2018).
K.Terayama, T. Yamashita, et al., npj Comput. Mater. **4**, 32(2018).

Outline

1. Introduction
   - Crystal structure prediction (CSP)
   - Propose a selection-type algorithm
2. Bayesian optimization
   - Descriptor
   - Algorithm
   - Searching efficiency
3. LAQA
   - Algorithm
   - Searching efficiency
4. Summary

Descriptor for crystal structures

Fingerprint: based on radial distribution function

A. R. Oganov and M. Valle, J. Chem. Phys. 130, 104504 (2009).

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

$$F_{AB}(0) = -1$$

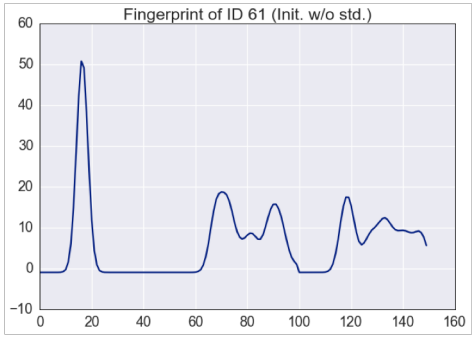
$$F_{AB}(\infty) = 0$$

- ith atoms of type A within the unit cell
- jth atoms of type B within the distance $R_{\max}$ (5 Å)
- discretized over bins of width Δ (50 discretized R points)

Example of Y_2Co_{17}

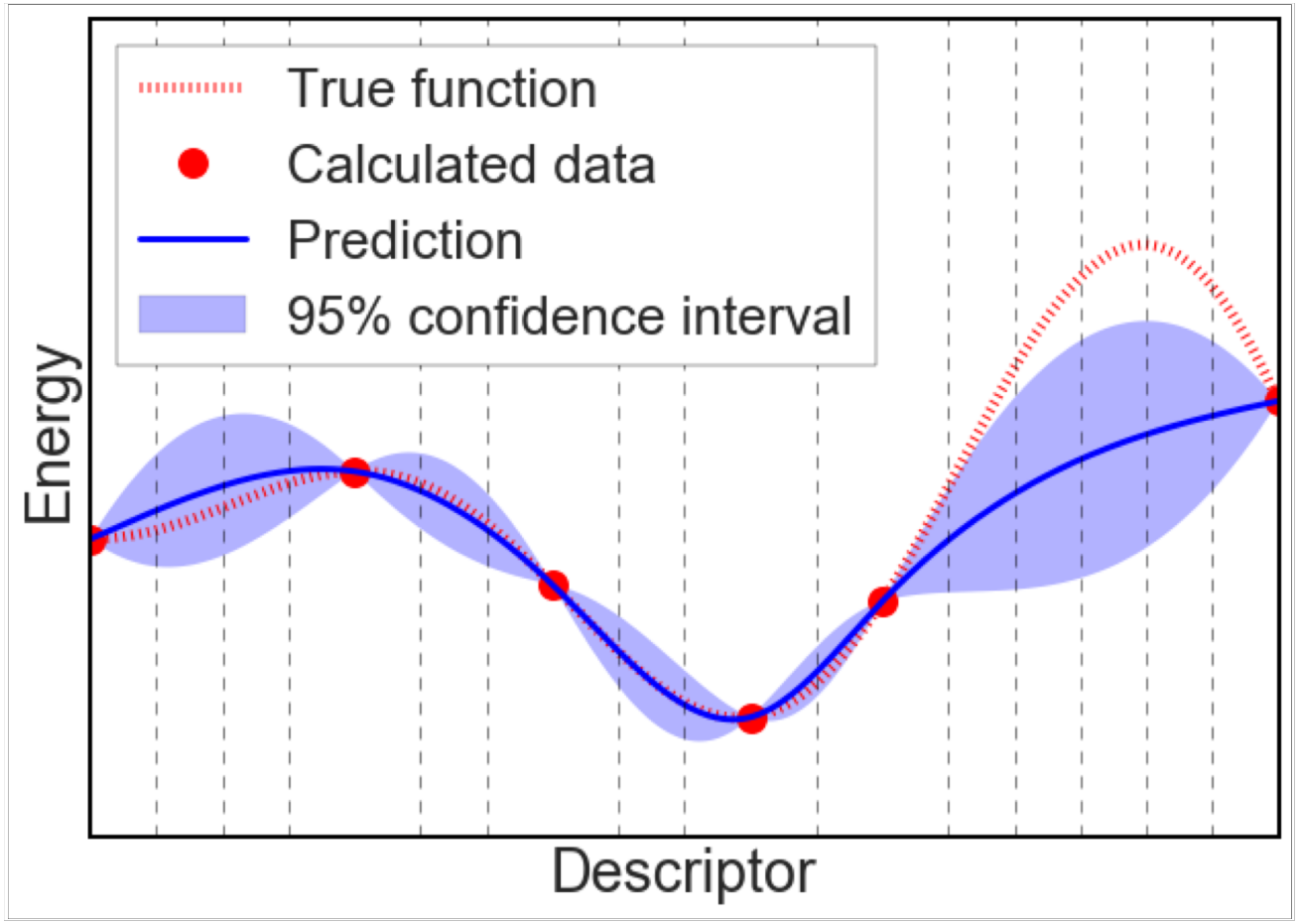
Y-Y, Y-Co, Co-Co

50, 50, 50



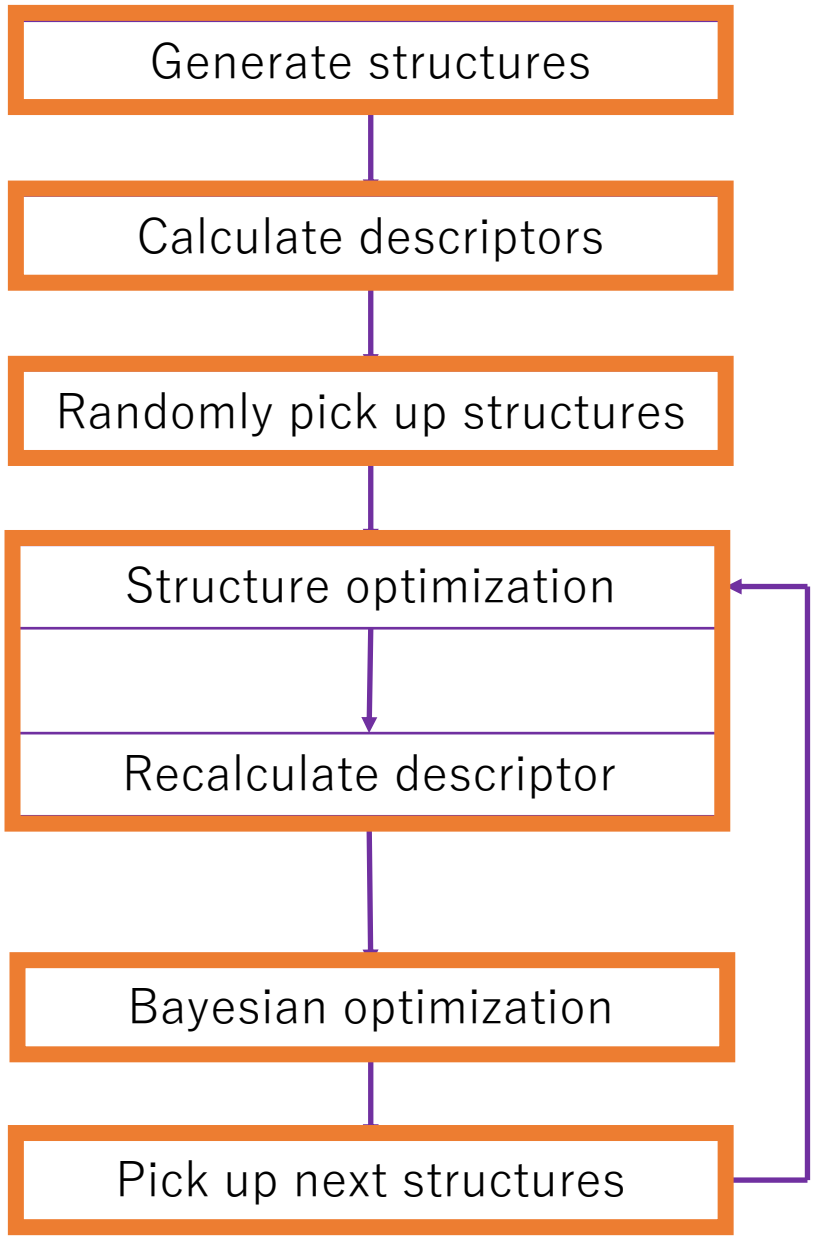
150 dimensional vector

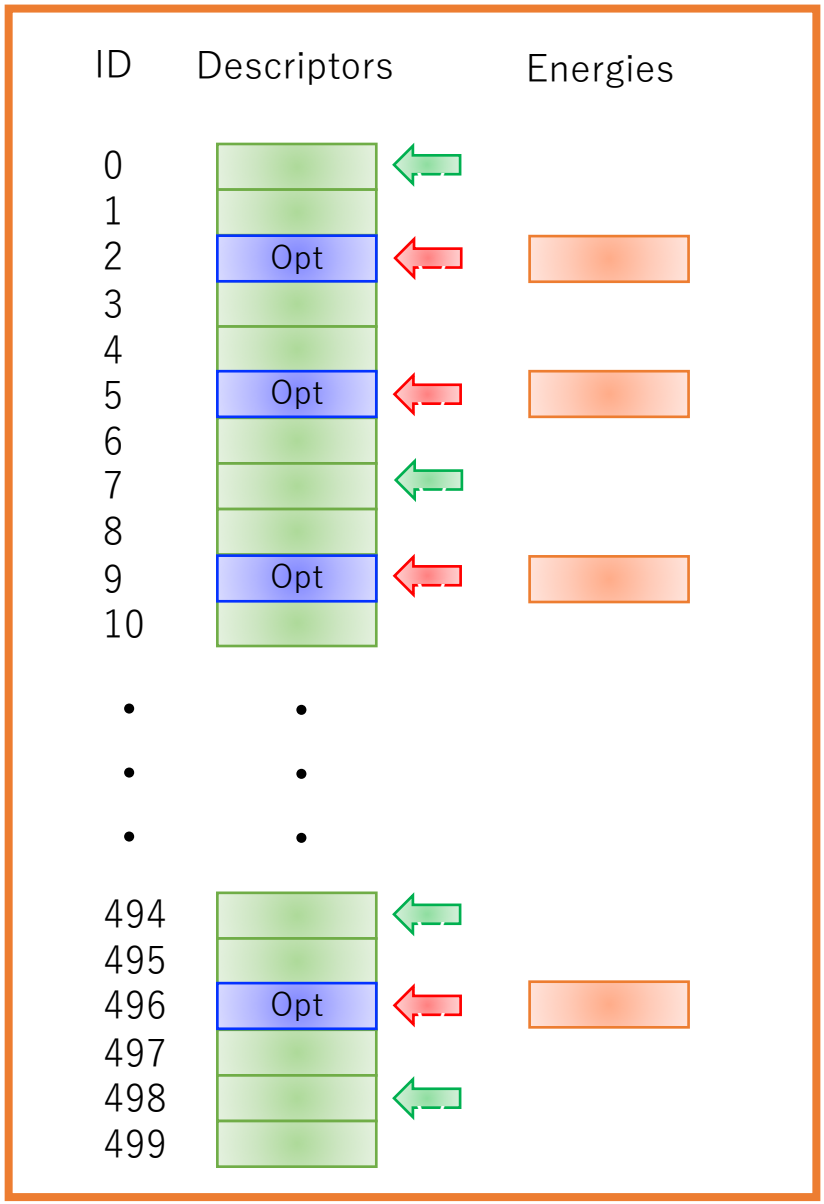
Bayesian optimization (BO)



BO enables us to balance trade-off between exploration and exploitation

Bayesian optimization (BO)



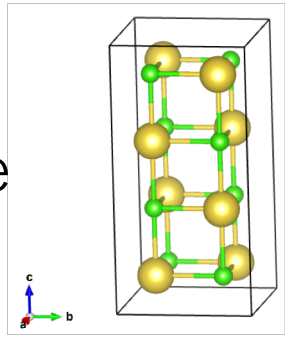


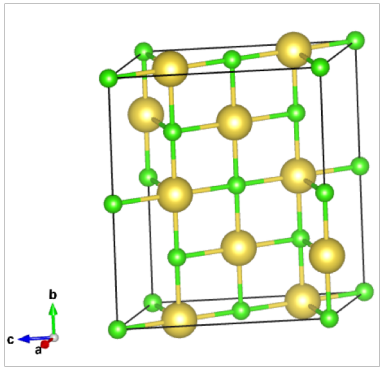
Test simulations

CSP is applied to two known systems using random search and BO for comparison

- Na_8Cl_8 (8 NaCl formula units: 16 atoms)

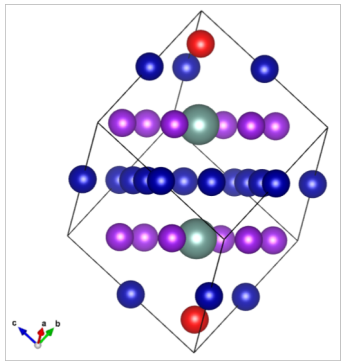
Rocksalt structure





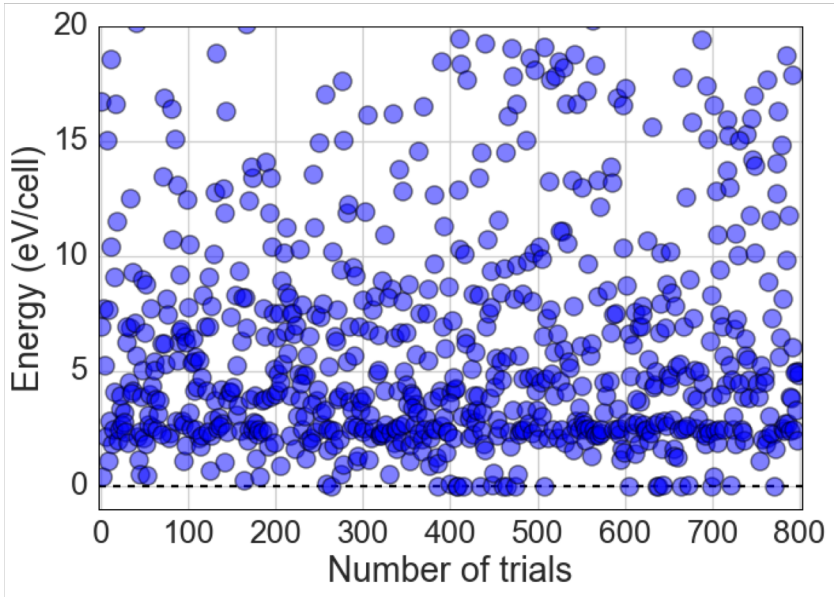
- Y_2Co_{17} (19 atoms)

Th_2Zn_{17}-type structure (R-3m)



Test simulations: NaCl

Random search



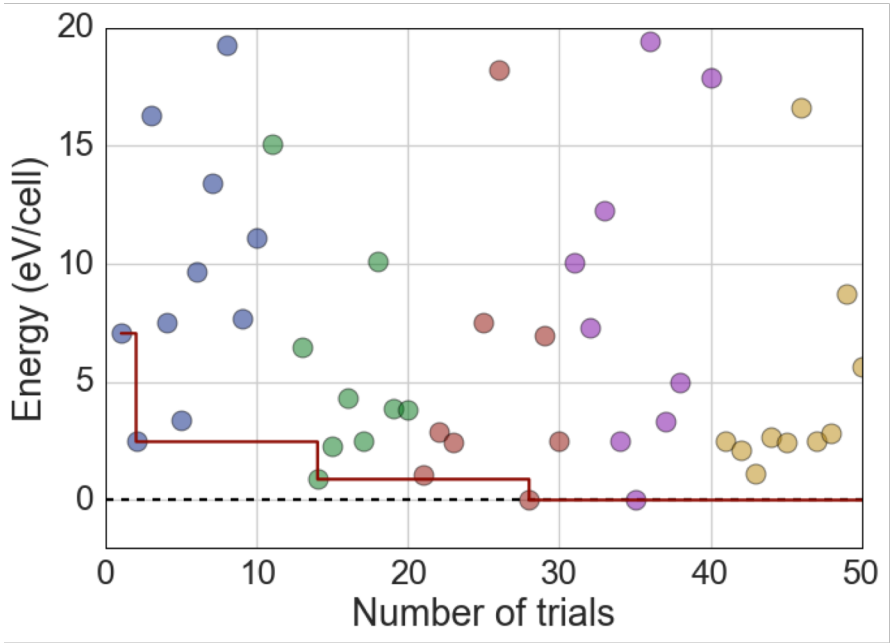
Searched 800 structures
Found out 23 rocksalt structures

Success rate: 23/800

Bayesian optimization

Same data set

- 800 structures
- 23 good candidates

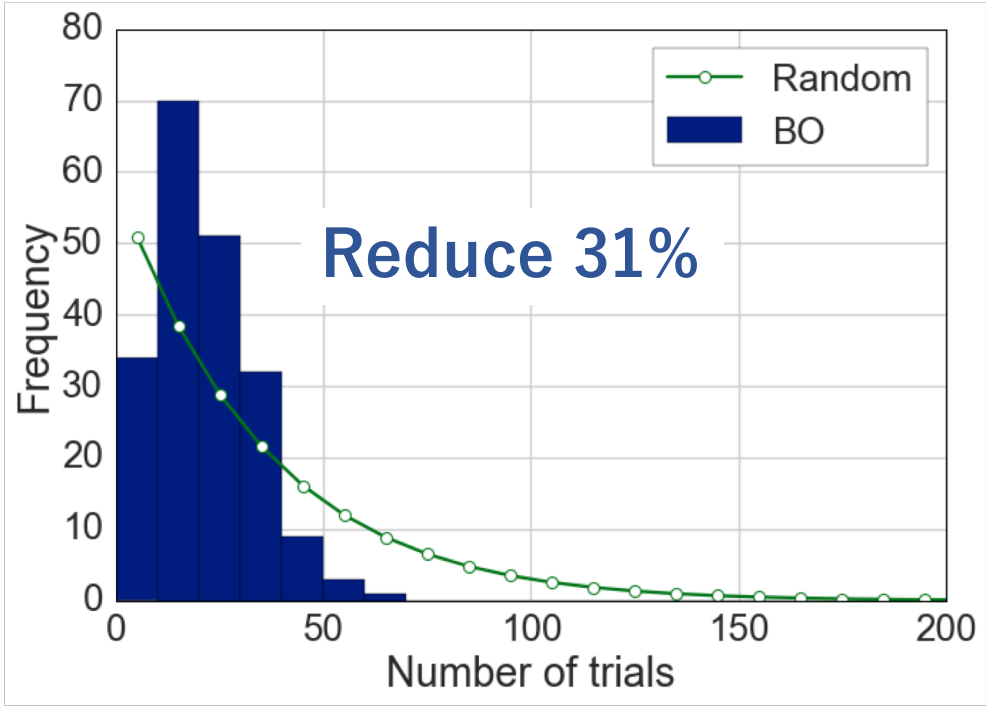


Selection: every 10 trials

Highly efficient!

Test simulations: NaCl

To statistically investigate the efficiency,
200 BO simulations were carried out

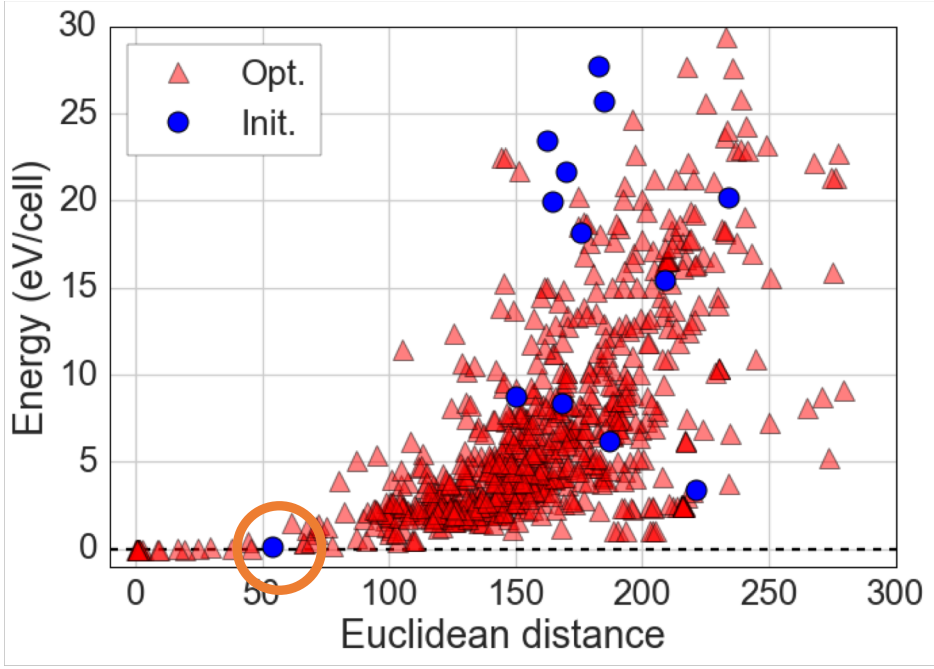


Frequency distribution of number of trials
required to find the rocksalt structure

BO can always select the rocksalt structure within 70 trials

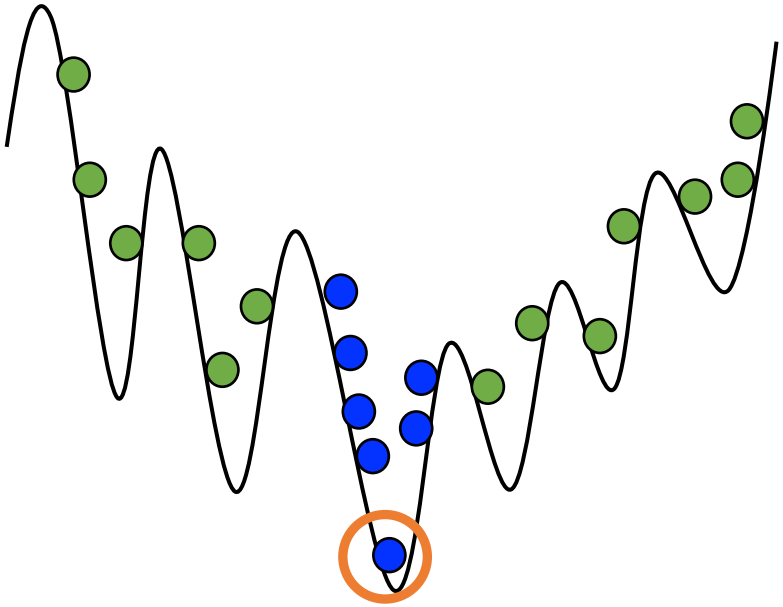
Test simulations: NaCl

800 structure data



Learned correlation between energy and structure distance

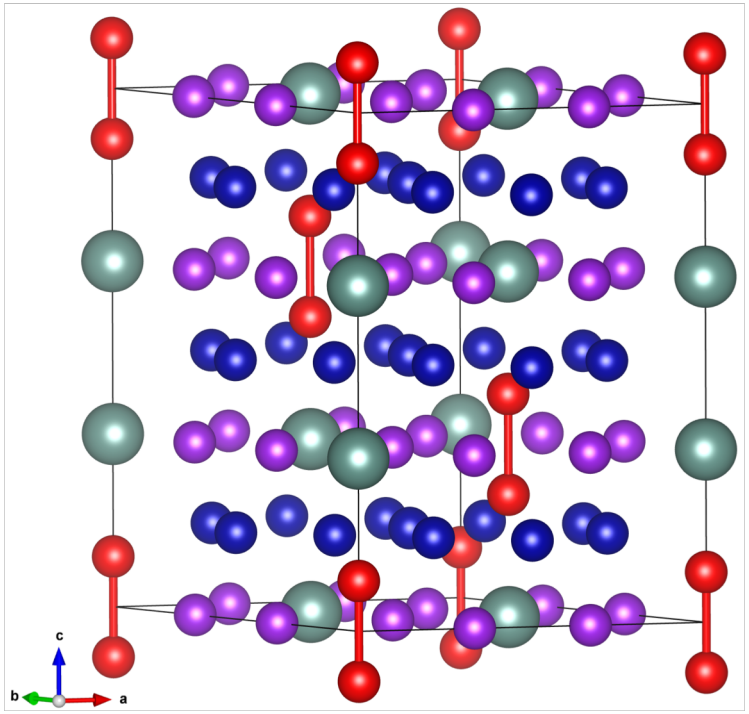
● : 23 good candidates



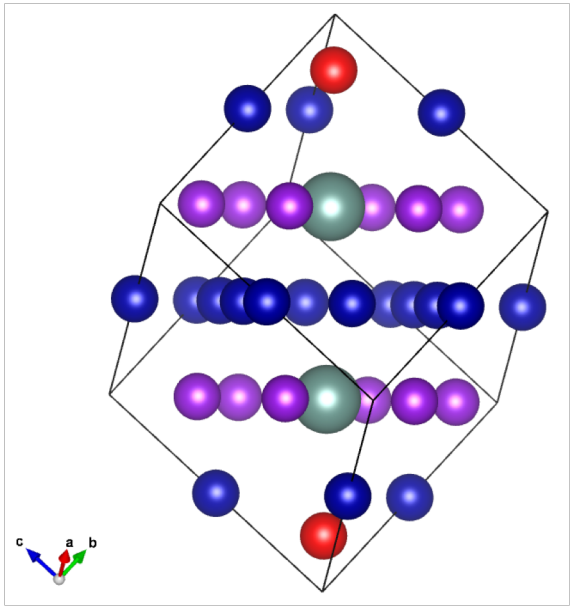
This candidate is always selected within 70 trials, as the exploitation phase of BO worked well.

Test simulation : Y_2Co_{17}

Th_2Zn_{17}-type structure
(R-3m)



Y



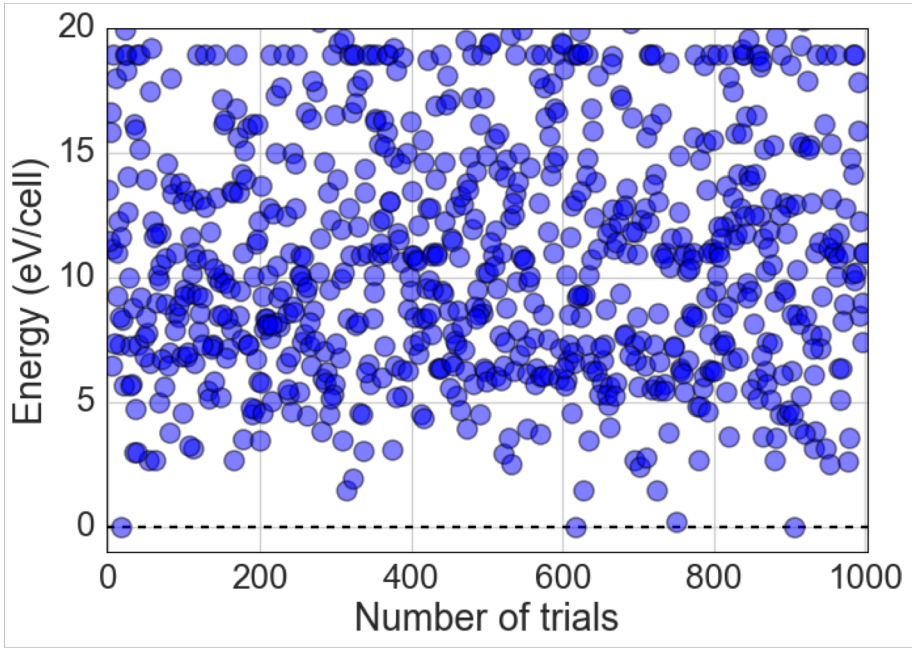
[Conventional]
Hexagonal

[Primitive]
Rhombohedral

19 atoms/cell

Test simulation : Y_2Co_{17}

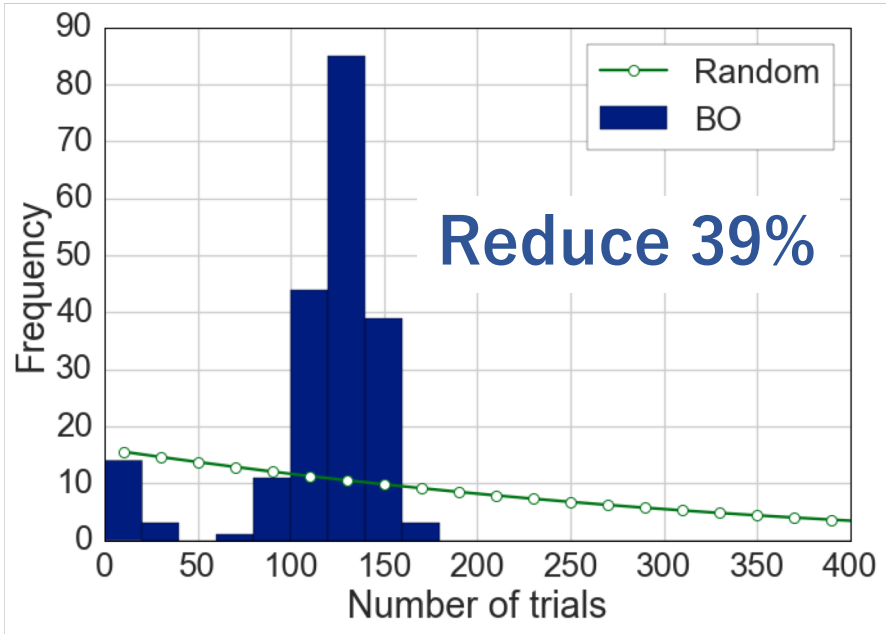
Random search



Searched 1000 structures
Found out 4 stable structures

Success rate: 4/1000

200 BO simulations

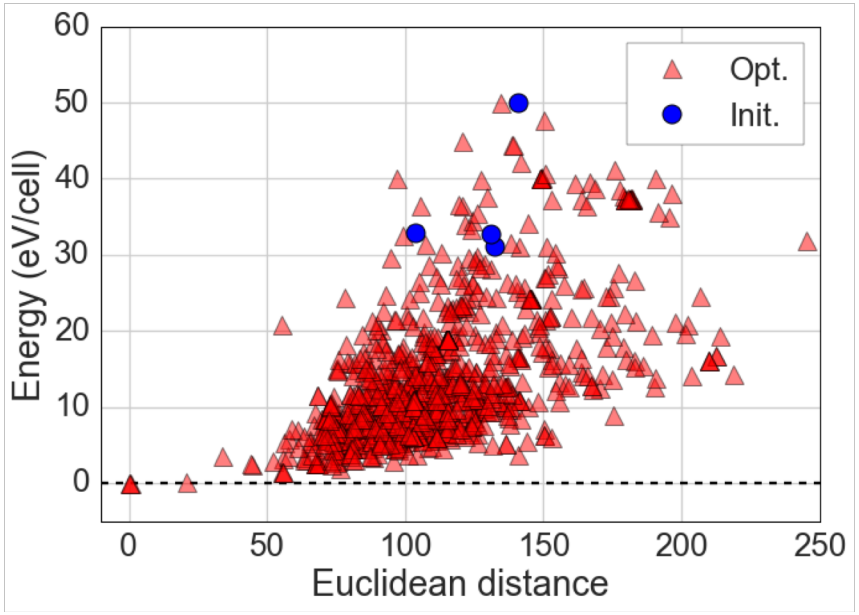


Selection: every 20 trials

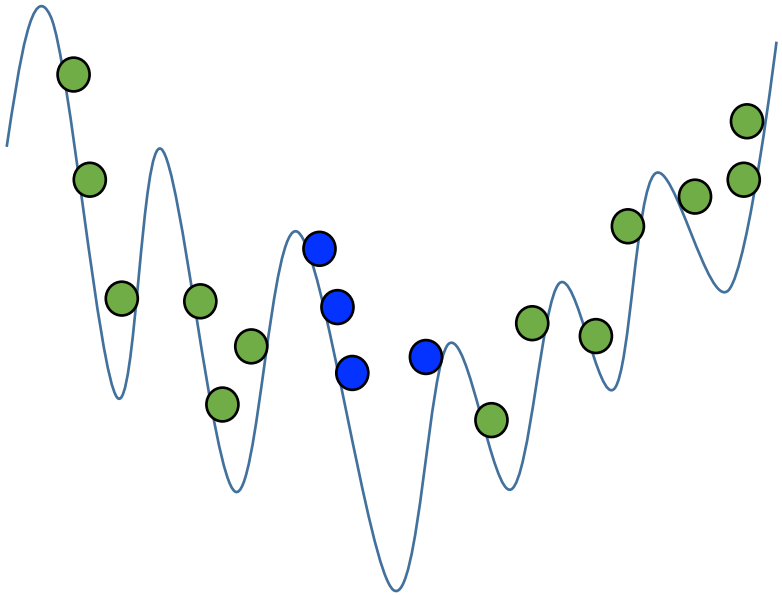
BO is clearly better than random selection

Test simulation : Y_2Co_{17}

1000 structure data



● : 4 good candidates



Difficult to predict, however …
the exploration phase of BO worked well.

➡ Leads to less number of trials

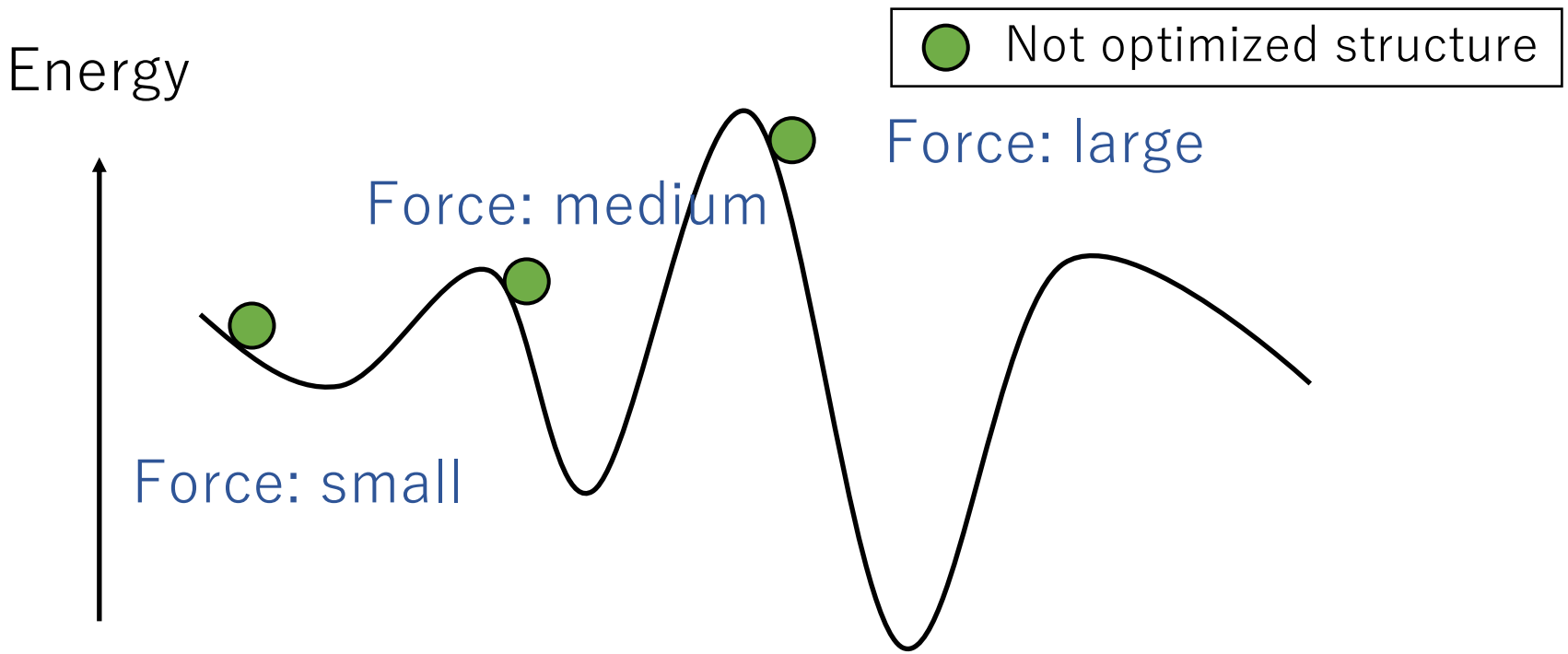
Outline

1. Introduction
   - Crystal structure prediction (CSP)
   - Propose a selection-type algorithm
2. Bayesian optimization
   - Descriptor
   - Algorithm
   - Searching efficiency
3. LAQA
   - Algorithm
   - Searching efficiency
4. Summary

LAQA

Look **A**head based on **Q**uadratic **A**pproximation

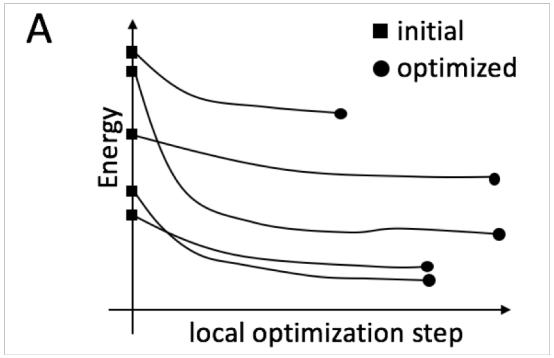
K.Terayama, T. Yamashita, et al., npj Comput. Mater. **4**, 32(2018).



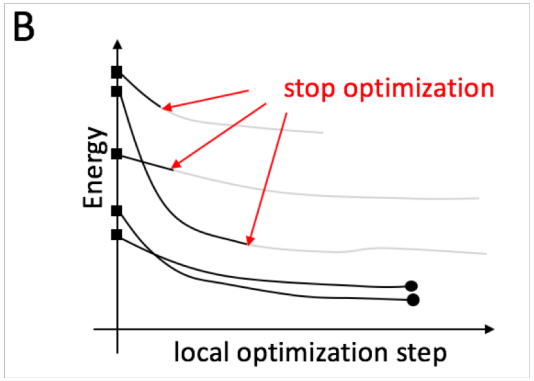
Which candidate do you select?

LAQA (Look Ahead based on Quadratic Approximation)

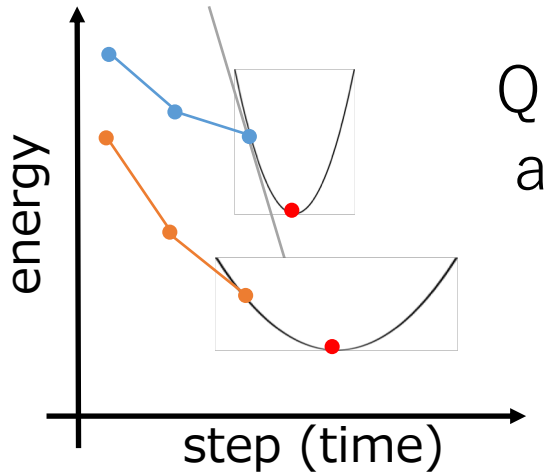
Conventional



Concept of LAQA



Gradient: force



Quadratic approximation

$$\text{score} = E - c\frac{F^2}{2\Delta F}$$

Select a candidate which has lower energy and larger force

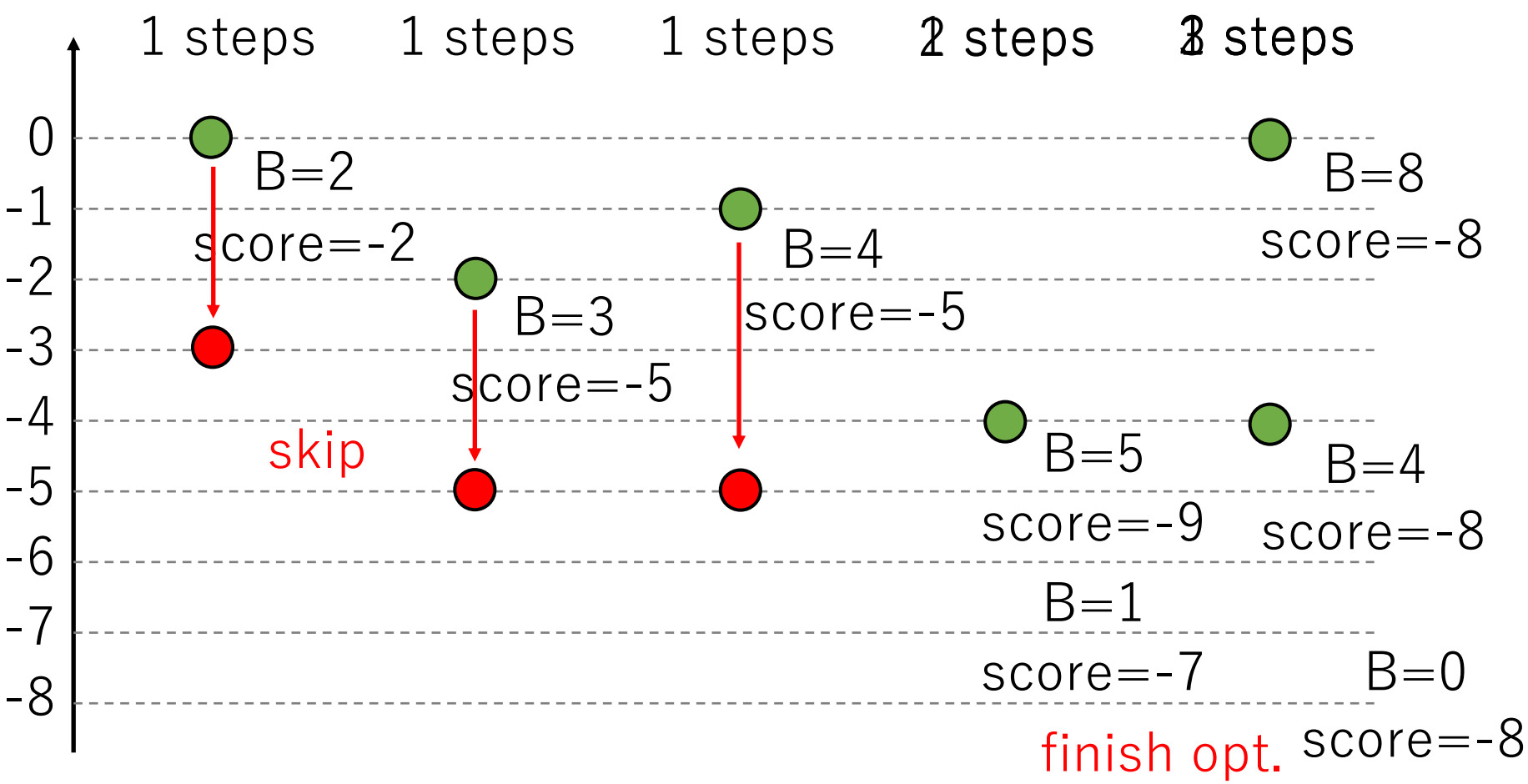
Descriptor is not required!

LAQA (Look Ahead based on Quadratic Approximation)

Procedure in LAQA

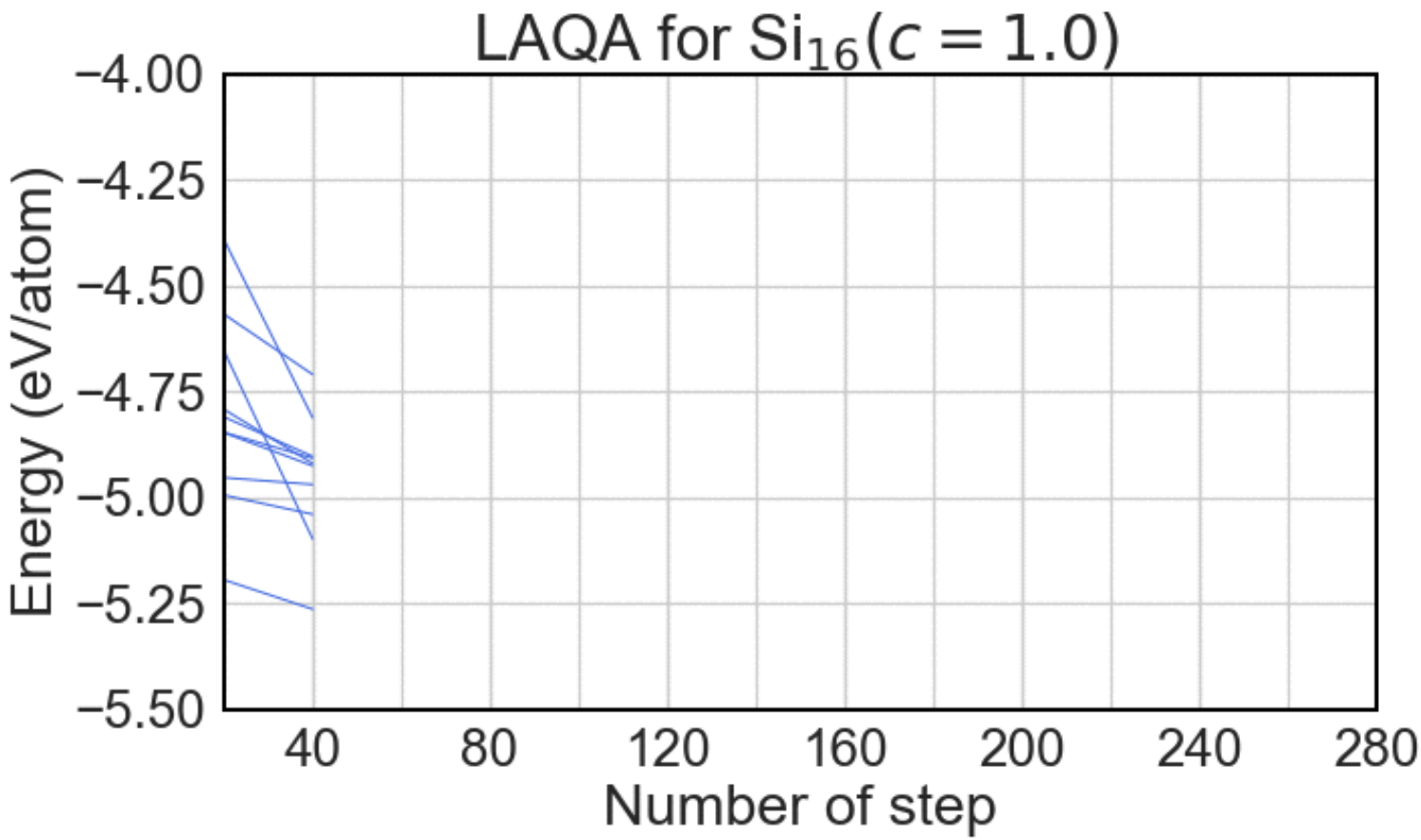
$$\text{score} = E - B$$

$$B = c\frac{F^2}{2\Delta F}$$



LAQA (Look Ahead based on Quadratic Approximation)

Si_{16} : 100 candidates



Red lines: top 3

Summary

- **We have developed a tool for crystal structure prediction**

Open source software: CrySPY

https://github.com/Tomoki-YAMASHITA/CrySPY

- **Selection-type algorithms have been developed**

  - **Bayesian optimization**
  - **LAQA**

Gaussian process

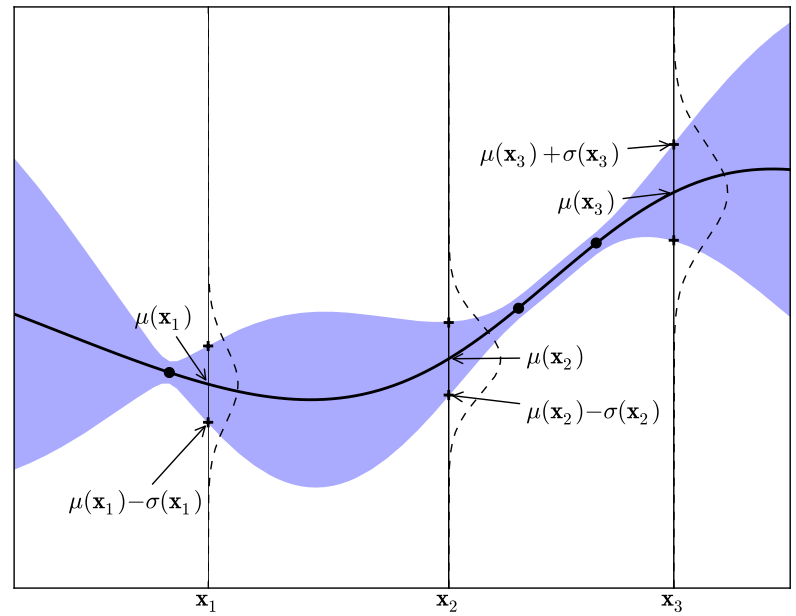
$$m(\mathbf{x}_{N+1}) = \mathbf{k}^{\mathrm{T}}\mathbf{C}_N^{-1}\mathbf{t}_N$$

$$\sigma^2(\mathbf{x}_{N+1}) = c - \mathbf{k}^{\mathrm{T}}\mathbf{C}_N^{-1}\mathbf{k}$$

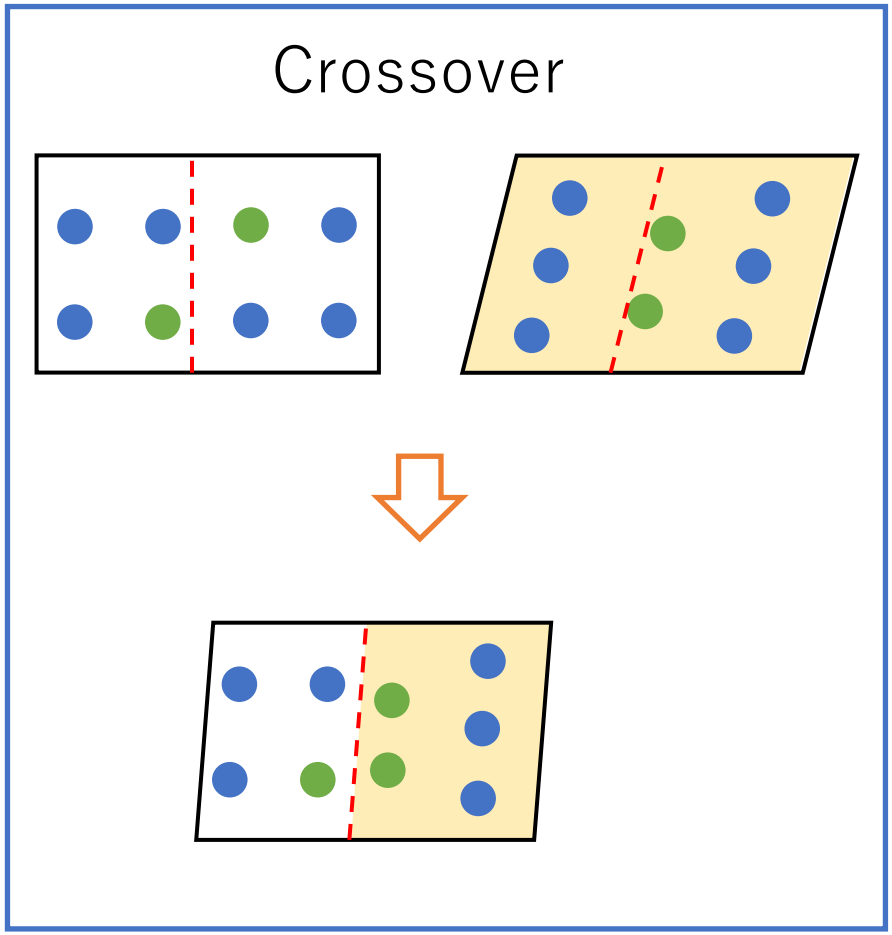
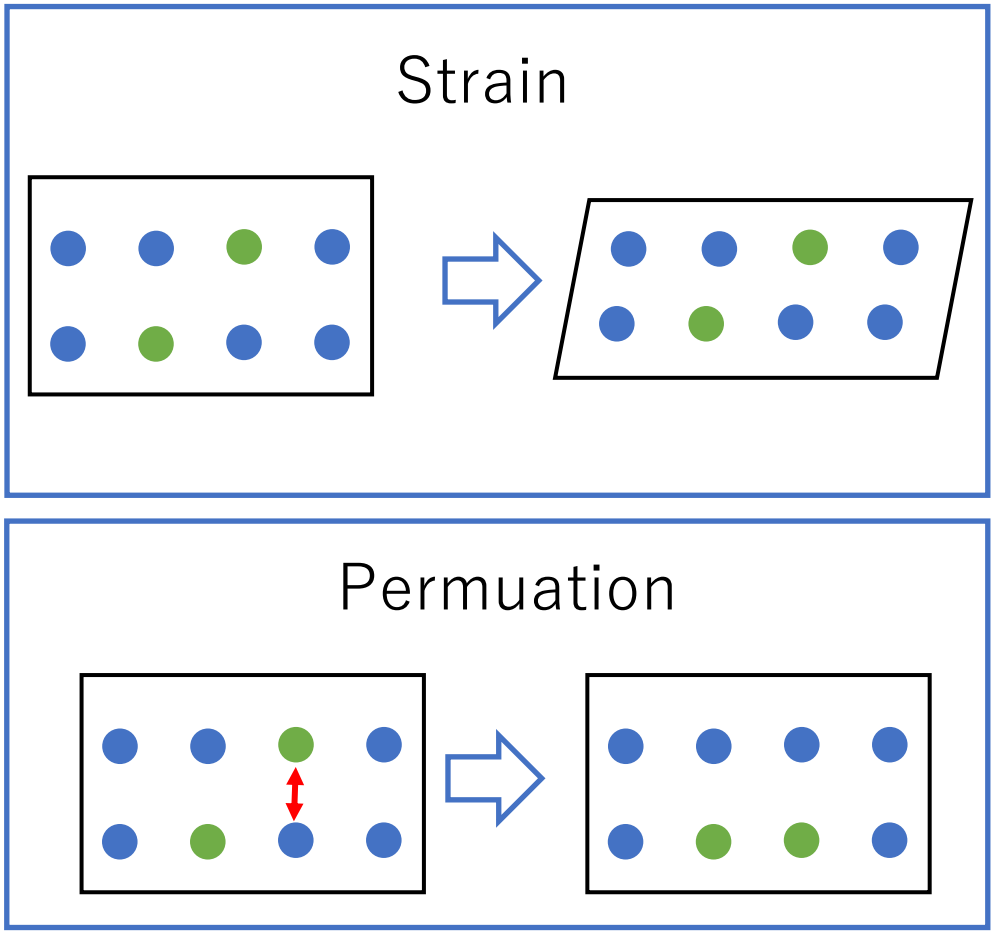
$$\mathbf{C}_{N+1} = \begin{pmatrix} \mathbf{C}_N & \mathbf{k} \\ \mathbf{k}^{\mathrm{T}} & c \end{pmatrix}$$

$$\mathrm{C}(\mathbf{x}_n, \mathbf{x}_m) = k(\mathbf{x}_n, \mathbf{x}_m) + \beta^{-1}\delta_{nm}$$

$$c = k(\mathbf{x}_{N+1}, \mathbf{x}_{N+1}) + \beta^{-1}$$



Gaussian kernel

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right)$$

Evolutionary algorithm