Practical Bayesian optimization

Kentaro KUTSUKAKE

RIKEN, Japan

Kentaro KUTSUKAKE

Postdoctoral Researcher, Dr. Sci.
Center for Advanced Intelligence Project (AIP), RIKEN

Chair, Applied Informatics Group, Japan Society of Applied Physics

Experience

2003-2007 PhD student, *Tohoku University*

2007-2010 Assistant Professor,
Crystal Growth Physics lab. (Prof. Nakajima), IMR, Tohoku University

2010-2017 Assistant Professor,
Crystal Defects Physics lab. (Prof. Yonenaga), IMR, Tohoku University

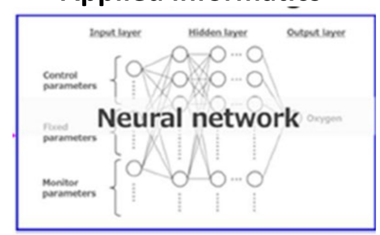
2017-2018 Designated Lecturer,
Info-analysis lab. (Prof. Inoue), IIFS, Nagoya University

2018- Postdoctoral Researcher,
Data-Driven Biomedical Science Team (Prof. Takeuchi), AIP, RIKEN

2018- Visiting Associate Professor,
IMaSS, Nagoya University

Crystal growth and characterization of mc-Si for solar cells

Applied informatics



Research Interest

Application of informatics to crystal engineering

Outline

1. Introduction of Bayesian optimization

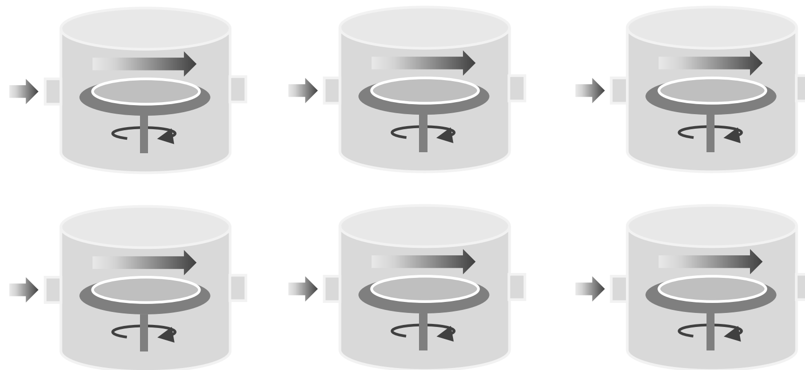
2. Applications to crystal engineering
   - ✓ 2D spatial mapping of micro-beam XRD
   - ✓ Hydrogen plasma treatment
   - ✓ Grinding process of SiC
   - ✓ Epitaxial growth of Si
   - ✓ Short summary

3. Tutorial
   - ✓ 1D Bayesian optimization
   - ✓ 2D Gaussian process regression
   - ✓ 2D Bayesian optimization

4. Summary

Inverse problem

Data



Training a model

Input

f(x)

Output

Process conditions

- Temperature
- Pressure
- Gas flow
- Position and rotation
- Equipment design
- etc

Macroscopic material properties

- Electrical properties
- Mechanical properties
- Magnetic properties
- Impurity concentration, etc.

Production cost

Reproductivity, Material yield, etc.

Inverse problem

Input, **x**

Process conditions

- Temperature
- Pressure
- Gas flow
- Position and rotation
- Equipment design
- etc

f(x)

Inverse problem

Output, **y**

Macroscopic material properties

- Electrical properties
- Mechanical properties
- Magnetic properties
- Impurity concentration, etc.

Production cost

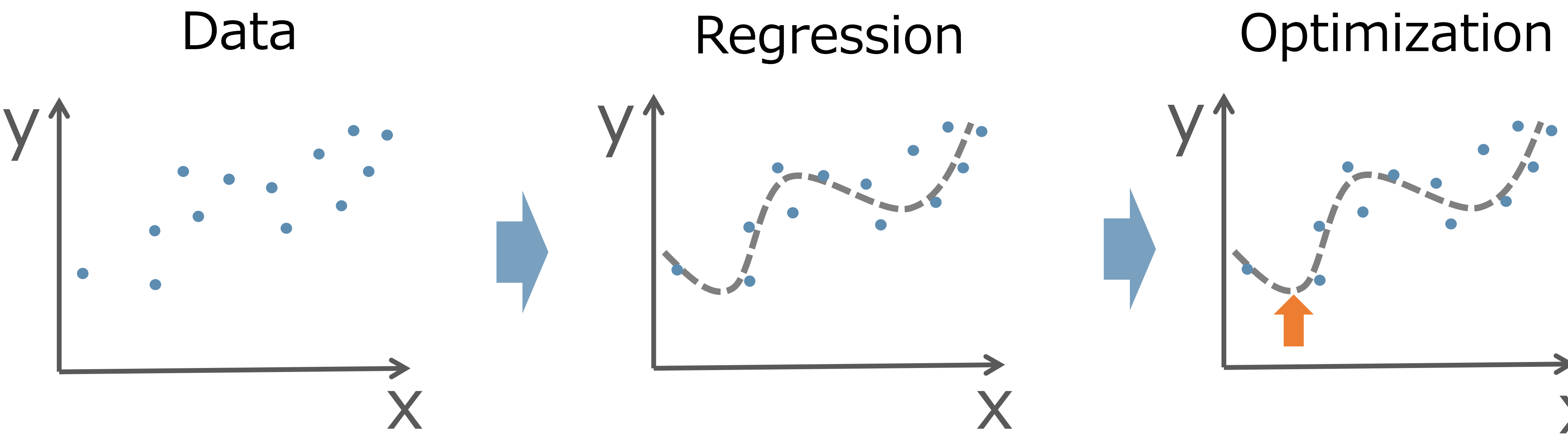
Reproductivity, Material yield, etc.

We want to know
the process condition (**x**) that minimize (maximize) the properties (**y**).

Find **x** that minimize **y**.

Optimization using machine learning model

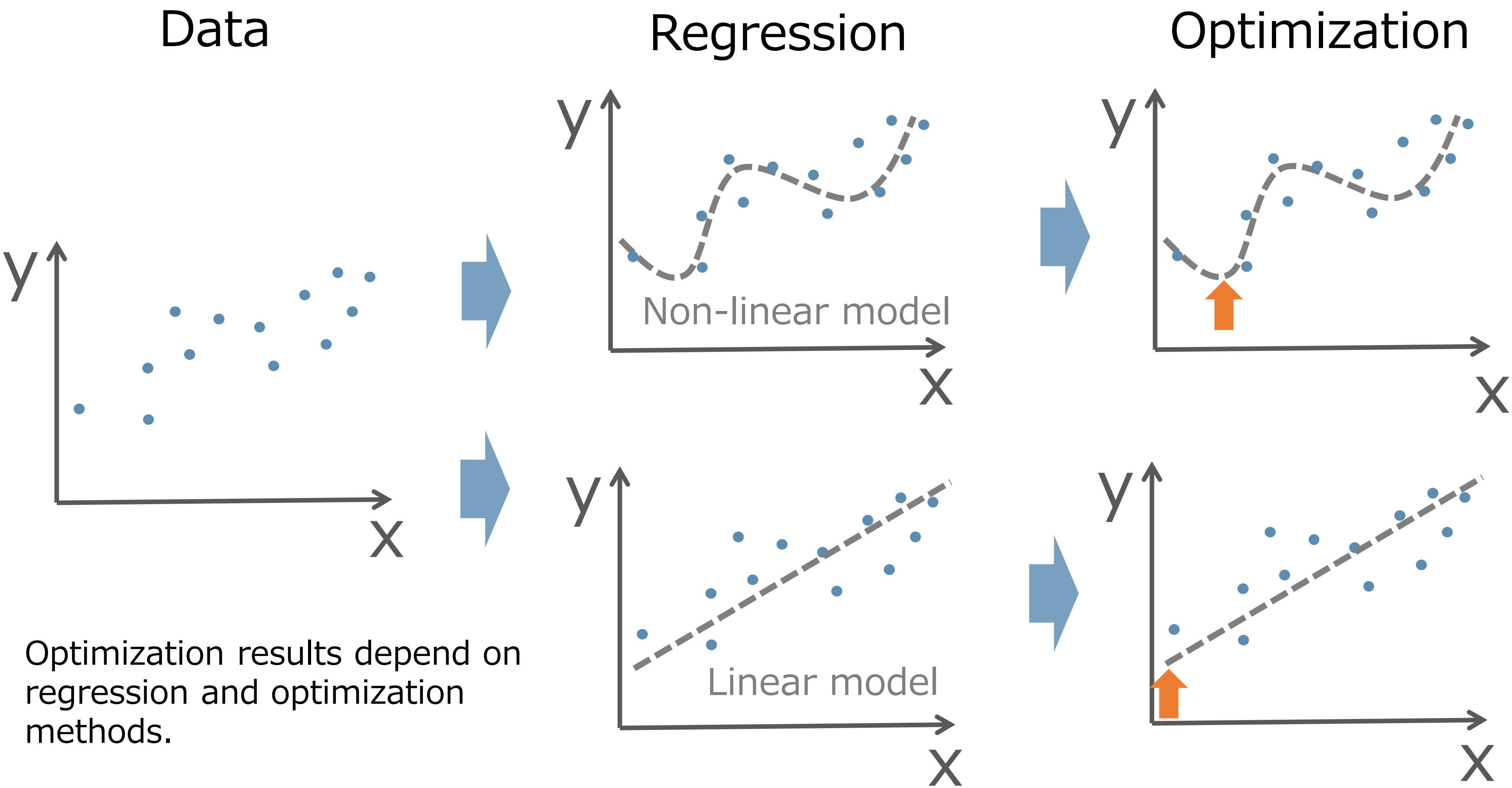
Find x that minimize y.



Optimization results depend on regression and optimization methods.

Optimization using machine learning model

Find x that minimize y.

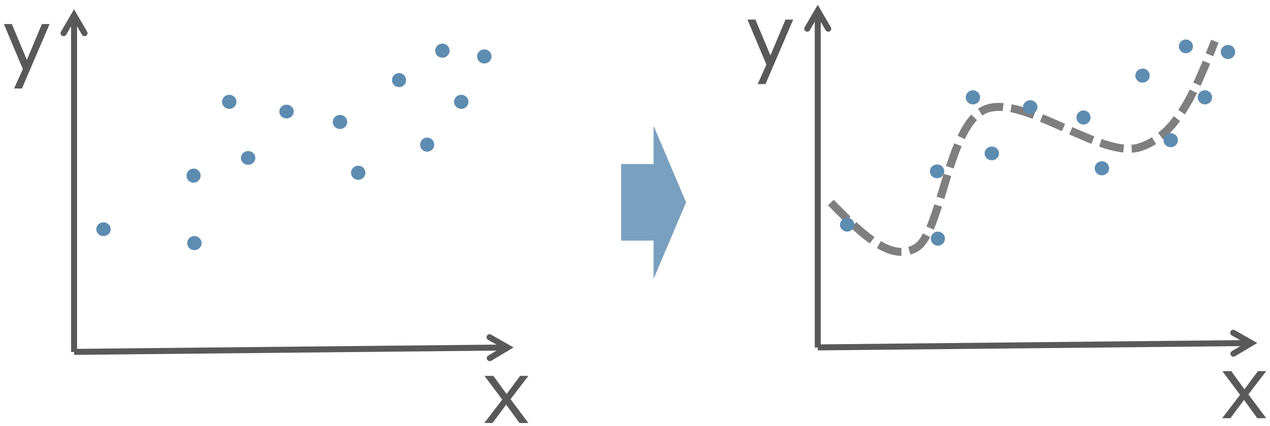


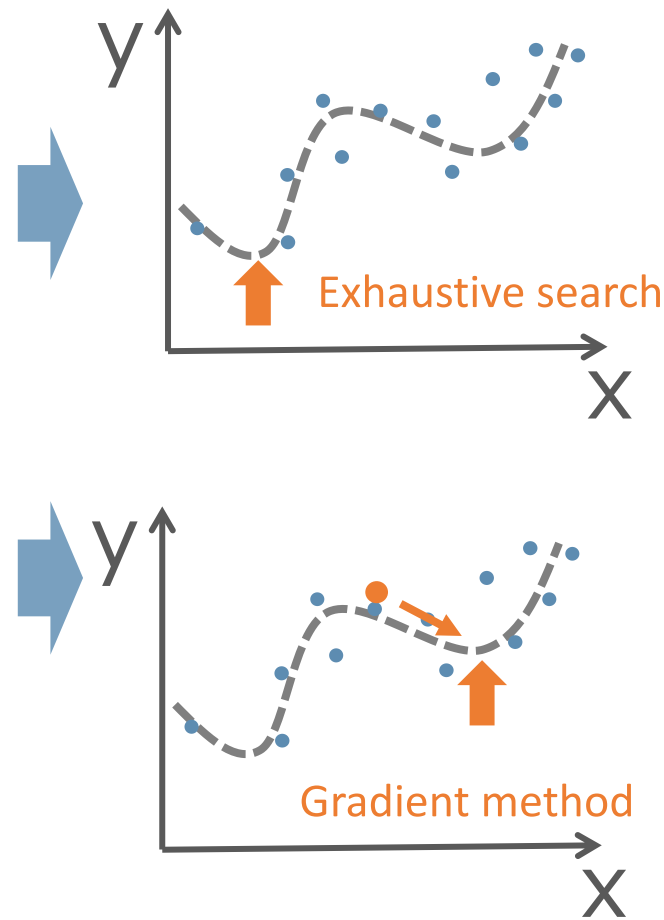
Optimization results depend on regression and optimization methods.

Optimization using machine learning model

Find x that minimize y.

Data　　Regression　　Optimization

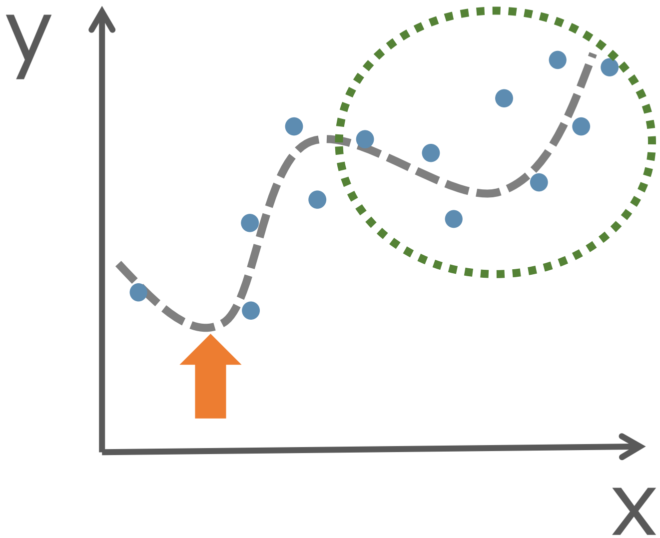




Optimization results depend on regression and optimization methods.

Small data

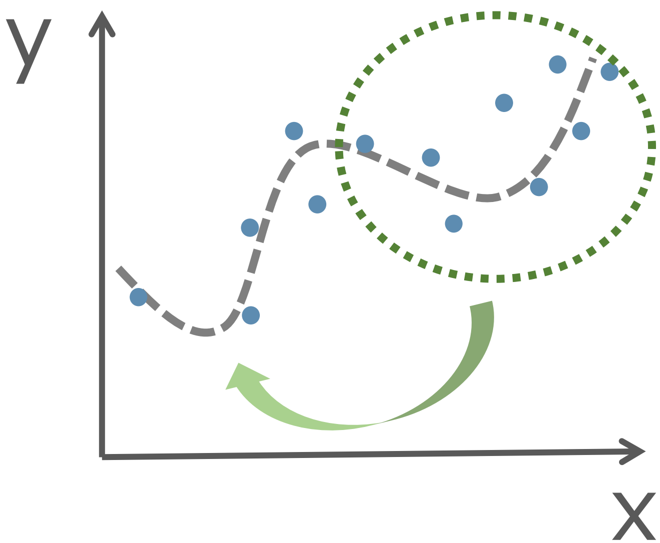
Small data = Limitation of number of experiments



These data are not effective for the optimization.

Small data

Small data = Limitation of number of experiments



These data are not effective for the optimization

Allocate resources to more important areas.

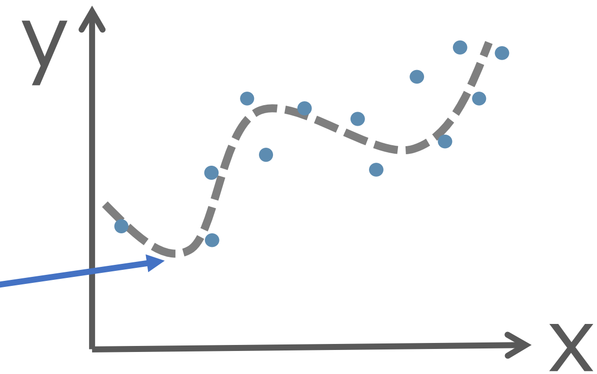
↓

Sequential optimization

Sequential optimization

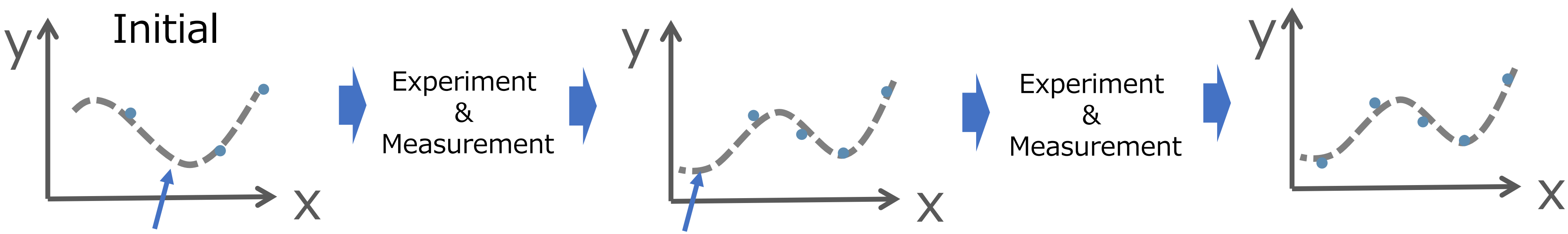
One-time optimization
(sufficient amount of data)

Here is the x that gives the smallest y.



Sequential optimization (small data)

Alternating regression & optimization and experiment & measurement



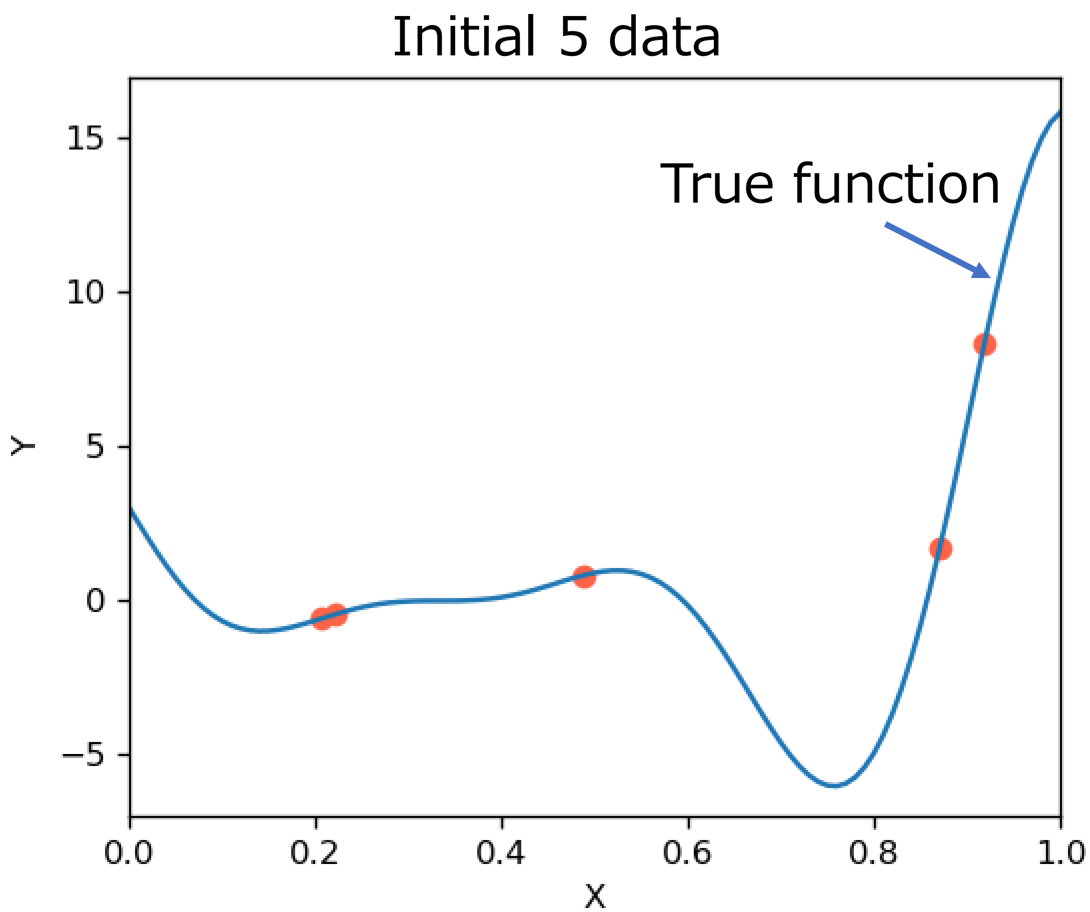
Here may be low.

Next, here may be low.

Global optimization
with small number of experiments
➔ **Bayesian optimization**

Bayesian optimization

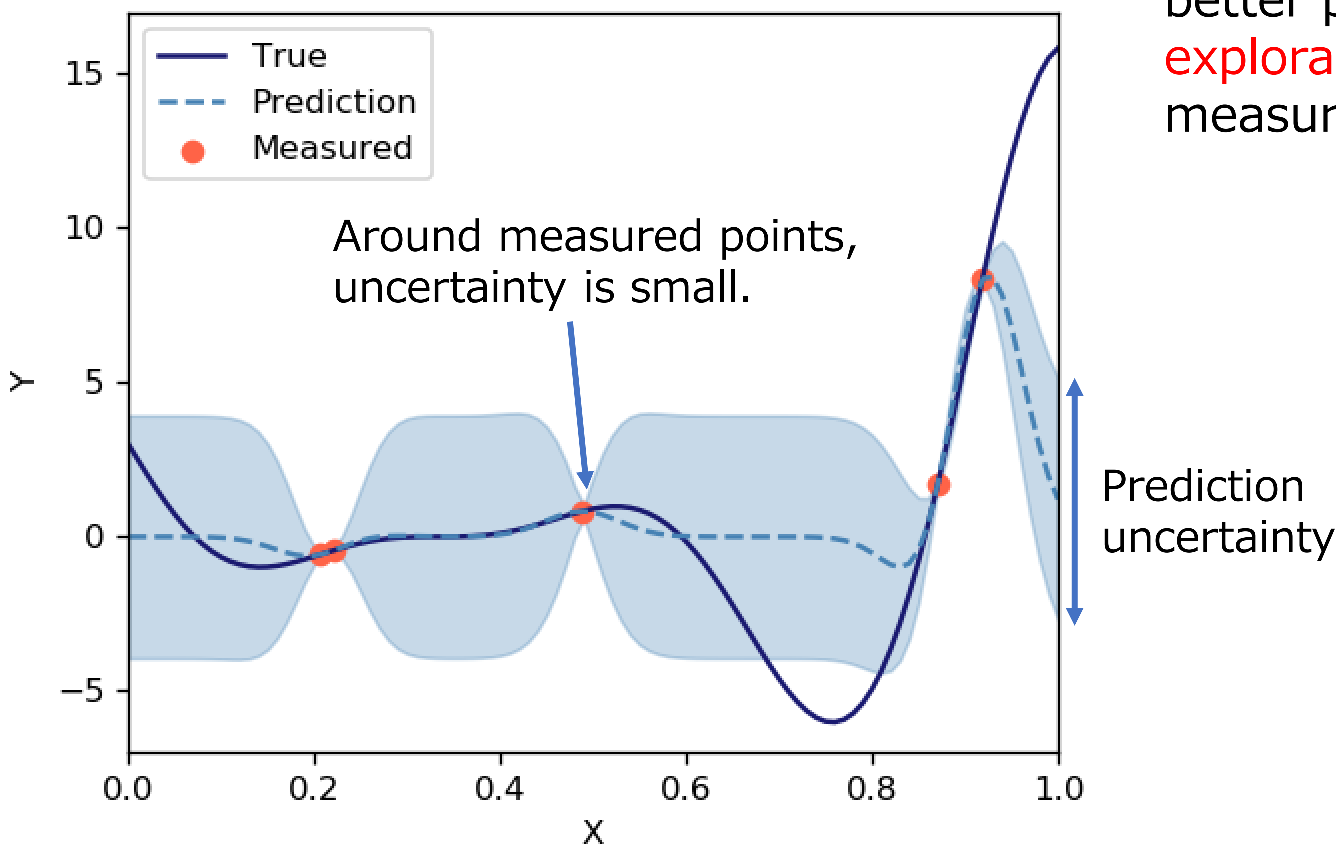
Find X that minimize Y



Bayesian optimization

Find X that minimize Y

Gaussian process regression



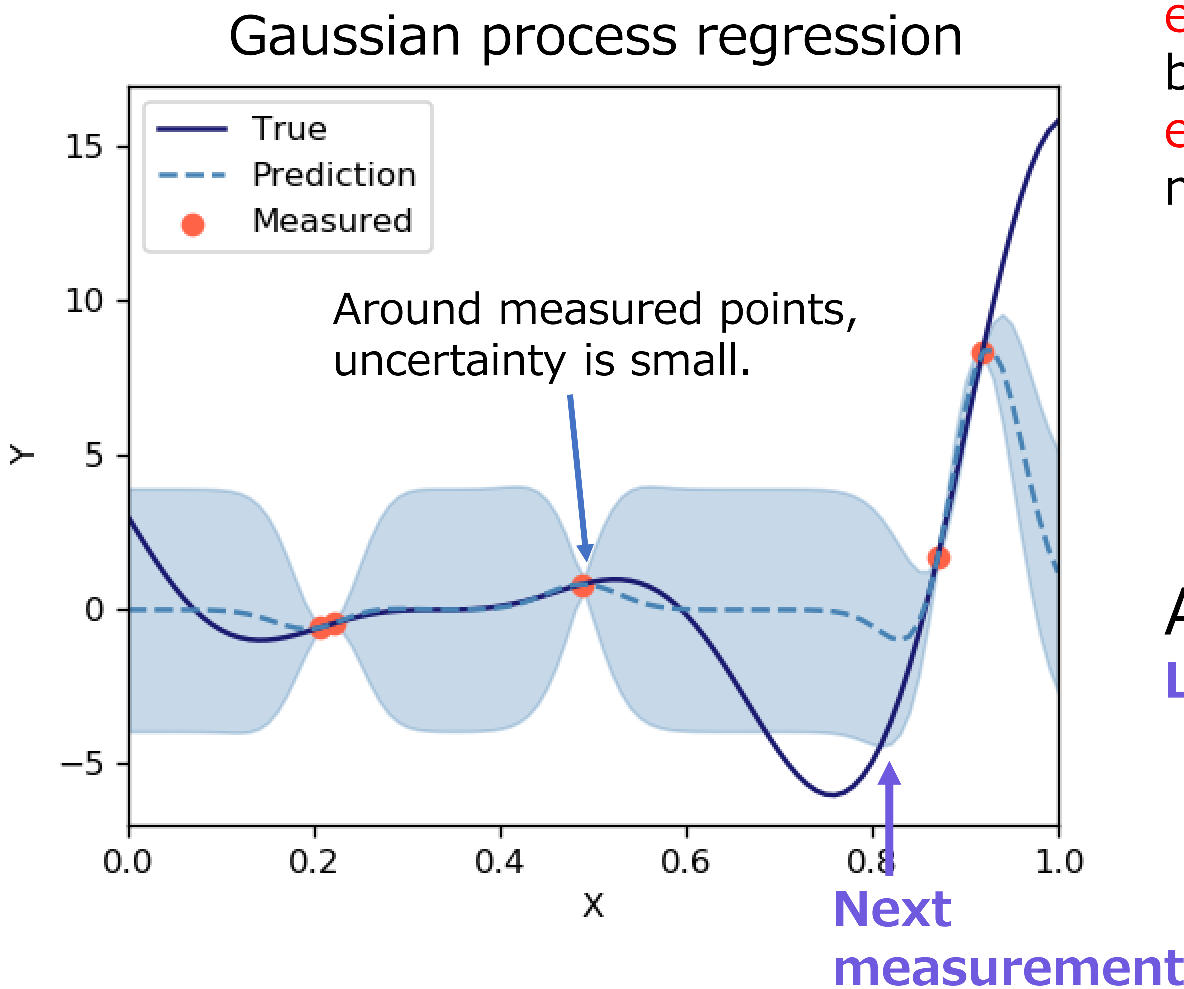
Bayesian optimization

exploitation : conditions with better predicted value

exploration : conditions not yet measured

Bayesian optimization

Find X that minimize Y



Bayesian optimization

exploitation : conditions with better predicted value

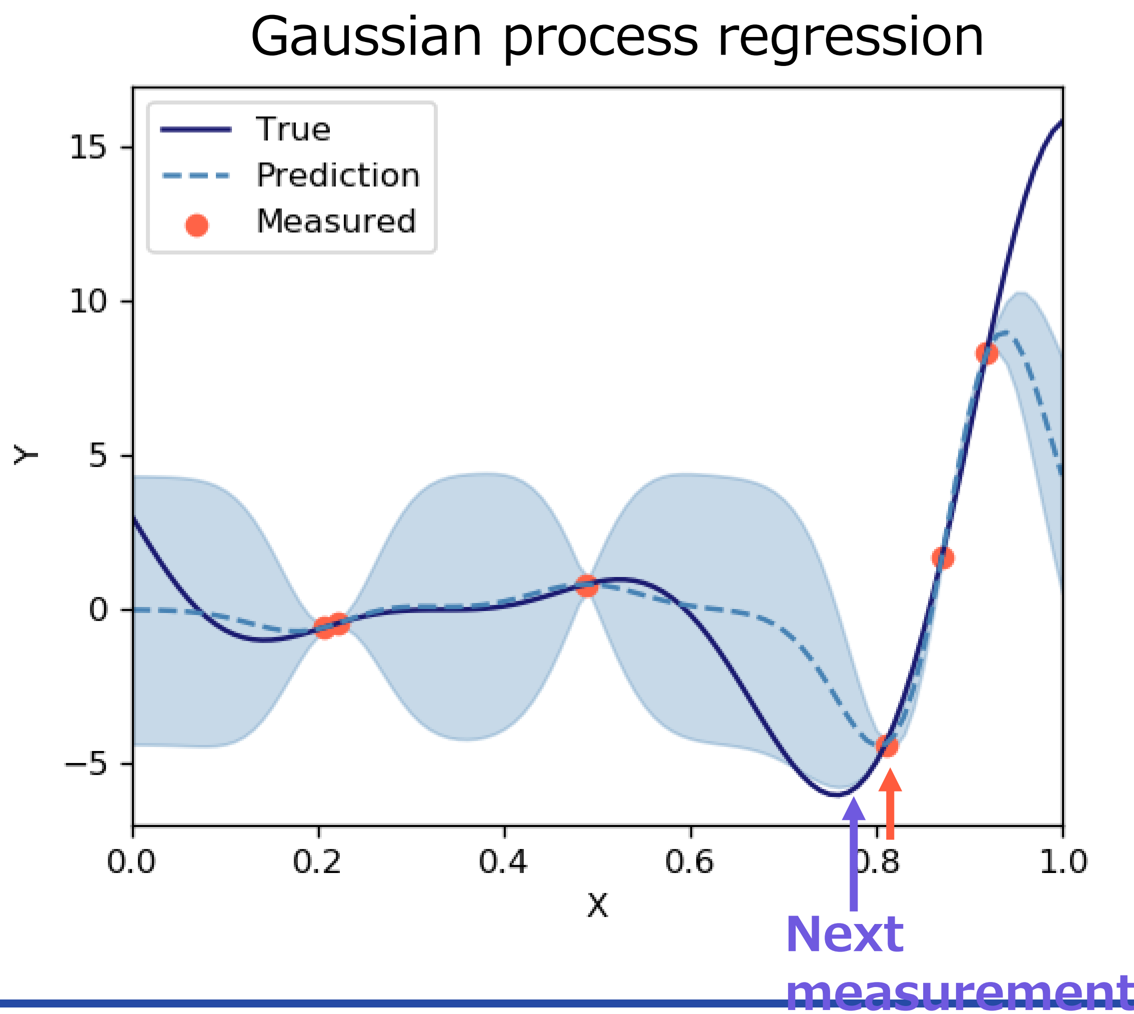
exploration : conditions not yet measured

Acquisition function

LCB : lower confidence band

Bayesian optimization

Find X that minimize Y



Bayesian optimization

exploitation : conditions with better predicted value
exploration : conditions not yet measured

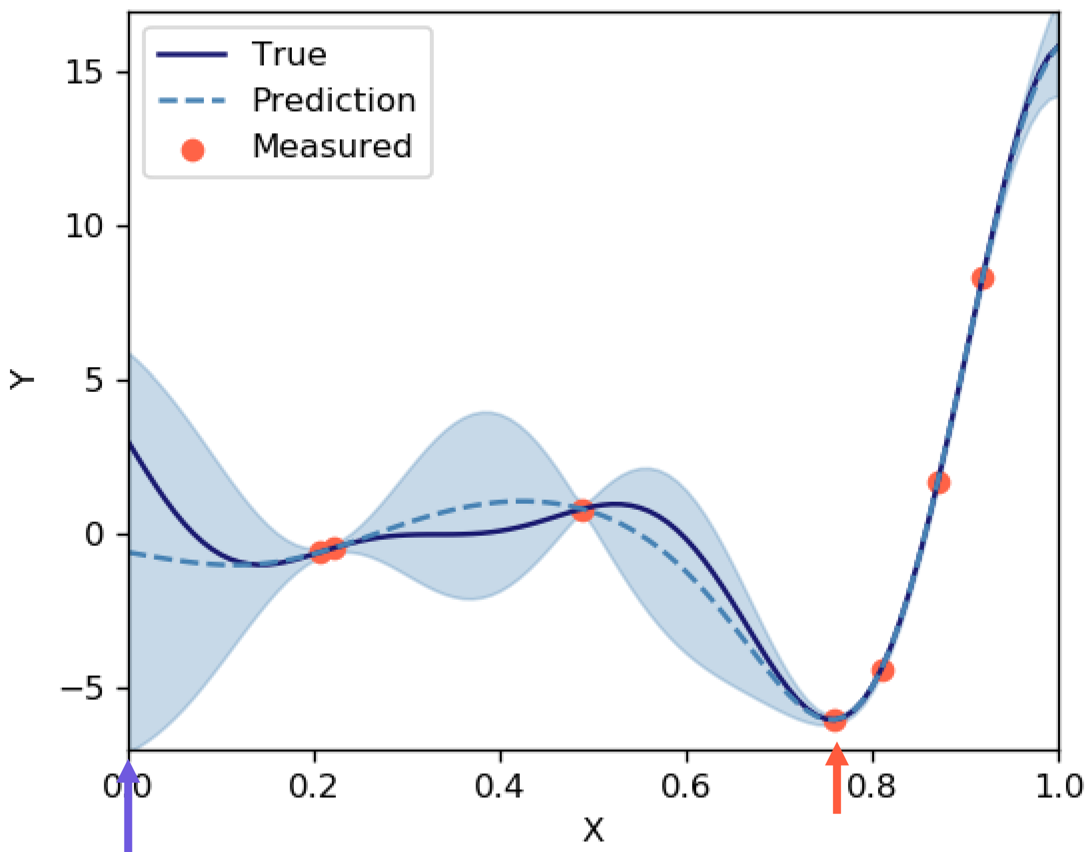
Acquisition function

LCB : lower confidence band

Bayesian optimization

Find X that minimize Y

Gaussian process regression



Next measurement

Bayesian optimization

exploitation : conditions with better predicted value

exploration : conditions not yet measured

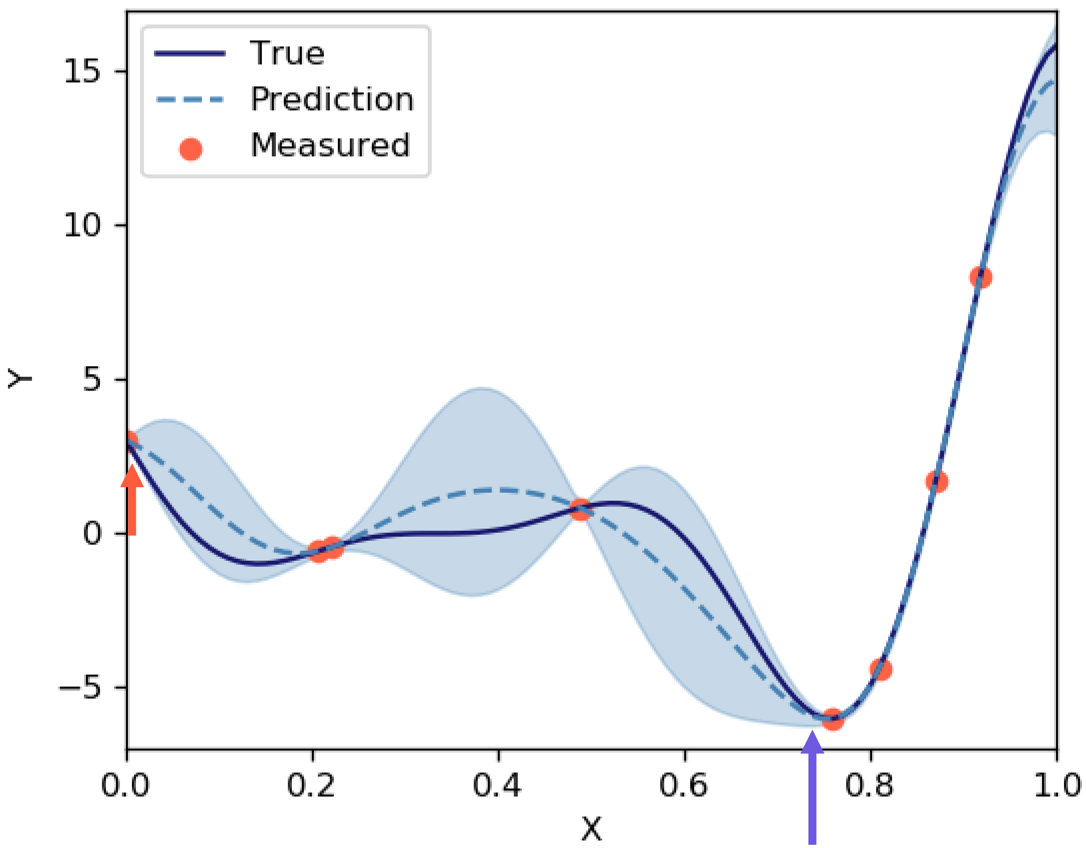
Acquisition function

LCB : lower confidence band

Bayesian optimization

Find X that minimize Y

Gaussian process regression



Next measurement

Bayesian optimization

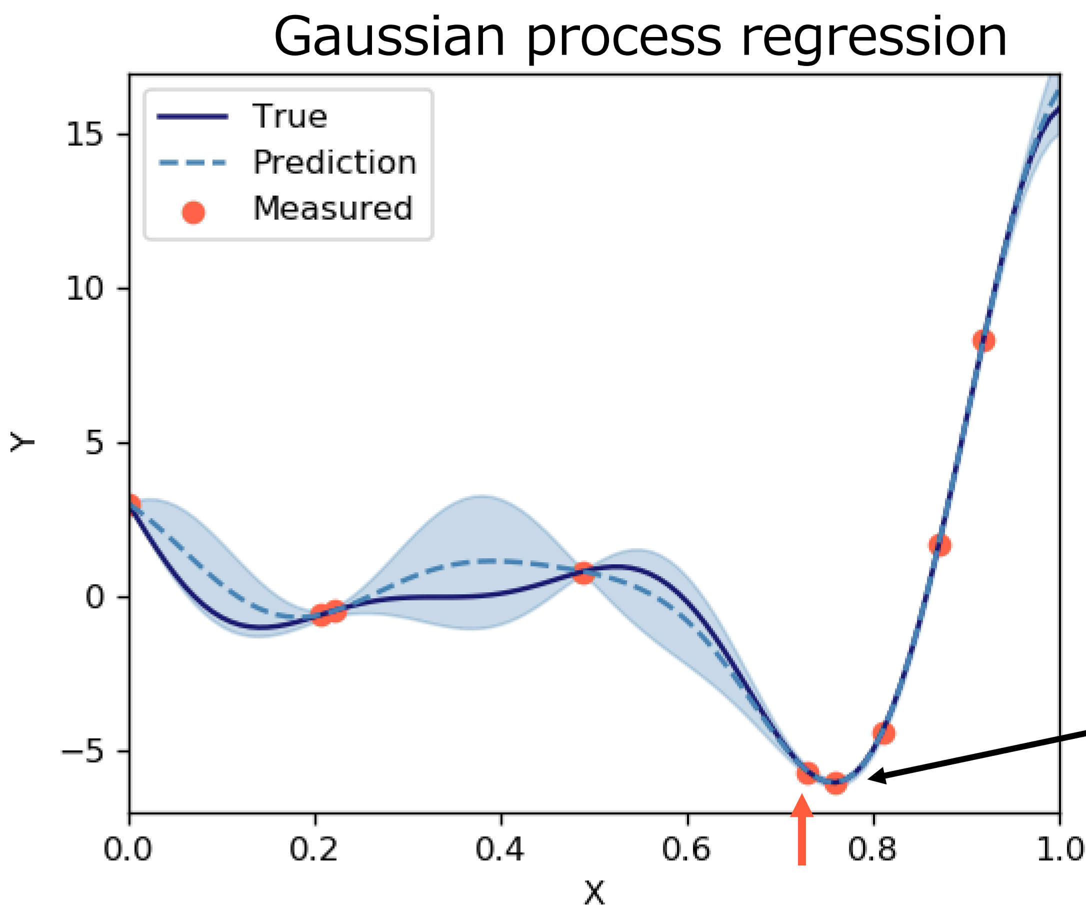
exploitation : conditions with better predicted value

exploration : conditions not yet measured

Acquisition function

LCB : lower confidence band

Find X that minimize Y



Bayesian optimization

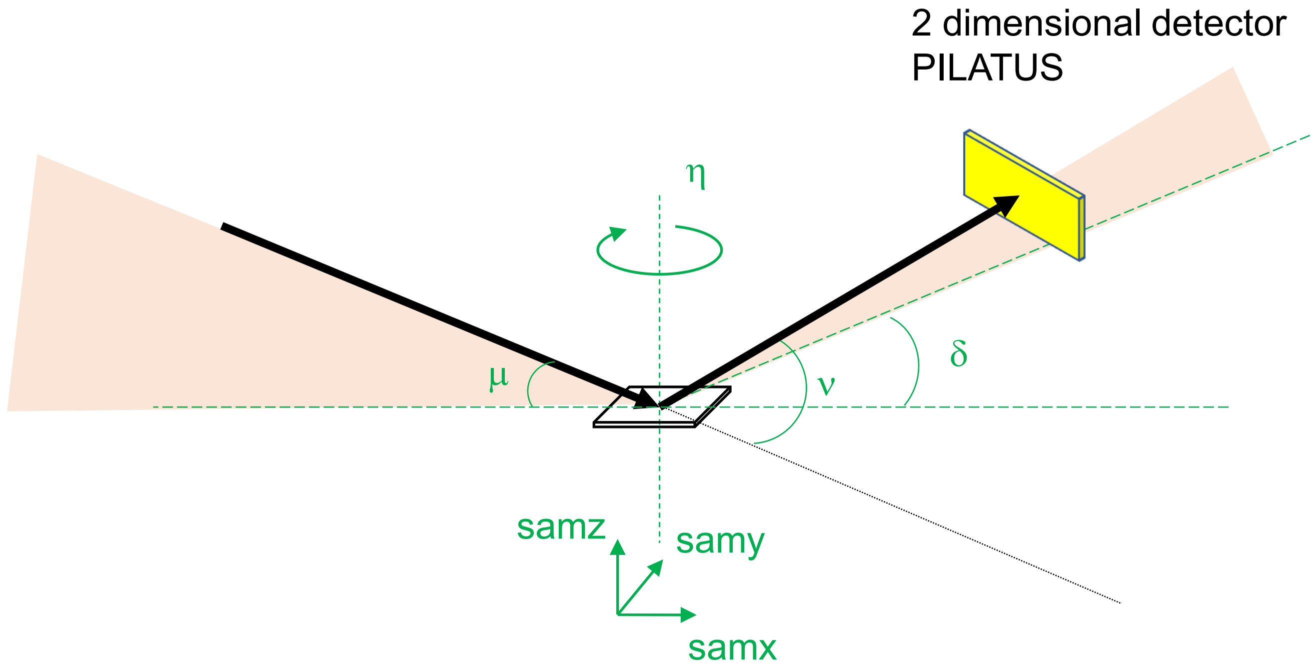
exploitation : conditions with better predicted value
exploration : conditions not yet measured

We have found the condition that minimize Y.

Application to 2D spatial mapping of micro-beam XRD

Spatial mapping of diffraction pattern

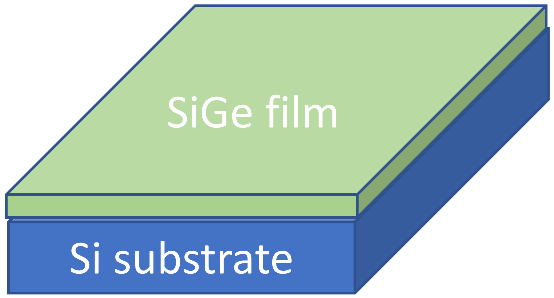
Optical system



Spot size : Φ 1 μm

Sample and diffraction pattern

Sample

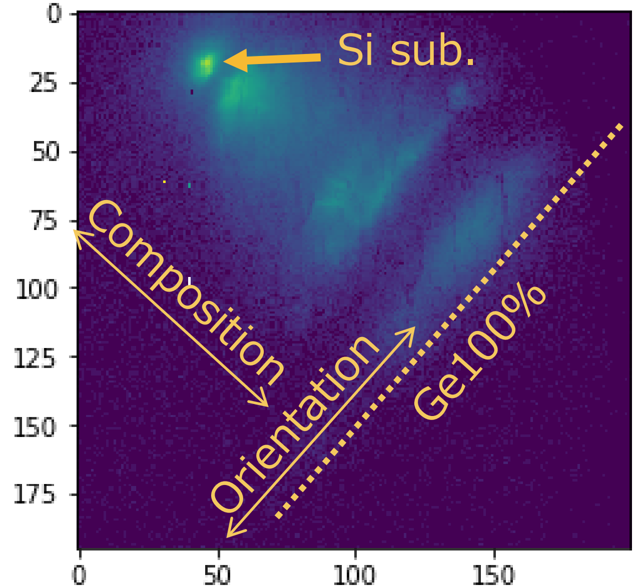


Large variation in crystal orientation and composition of $Si_{1-x}Ge_x$

↓

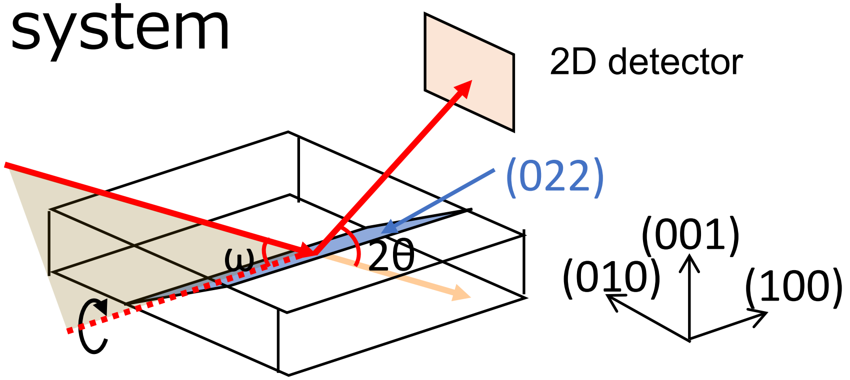
Spatial mapping

Average diffraction pattern



Target
Find the most tilted position

Optical system

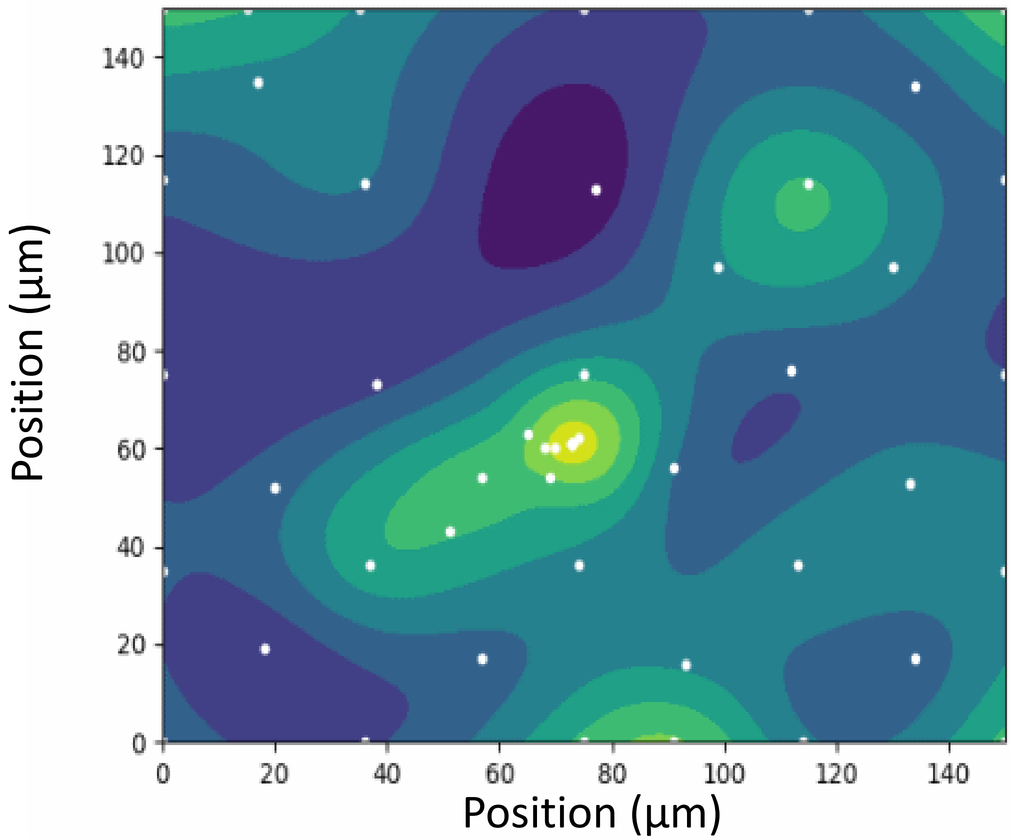


Bayesian optimization: Find the most tilted position

Sequential optimization

Measurement of diffraction pattern
↓
Calculation of tilt angle
↓
Bayesian optimization
↓
Suggestion of next position with highest expected improvement

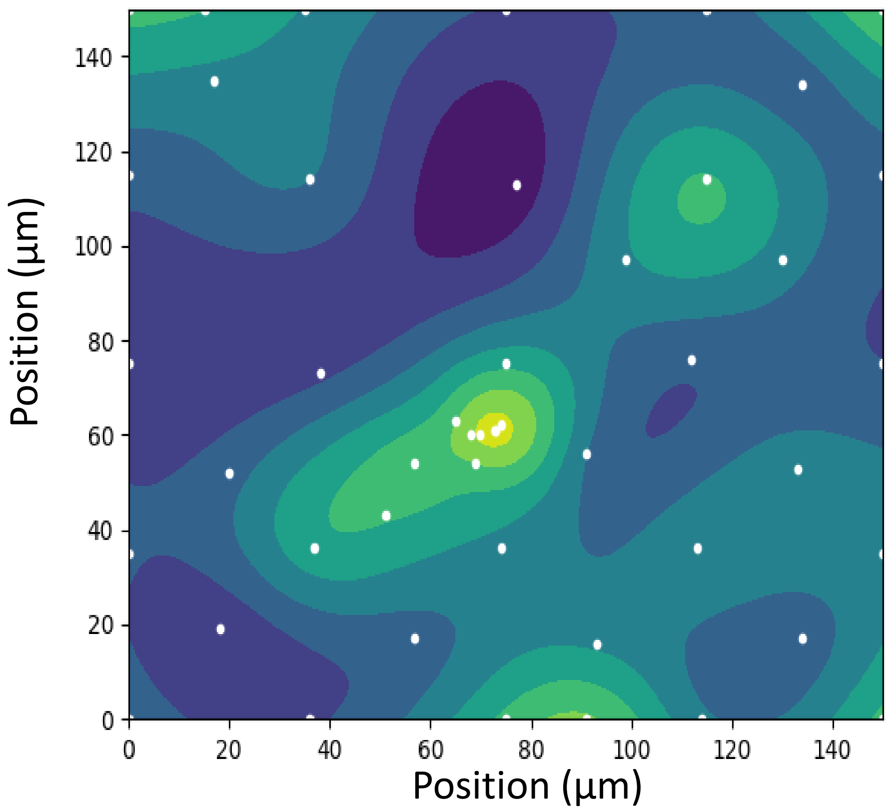
Predicted crystal orientation distribution



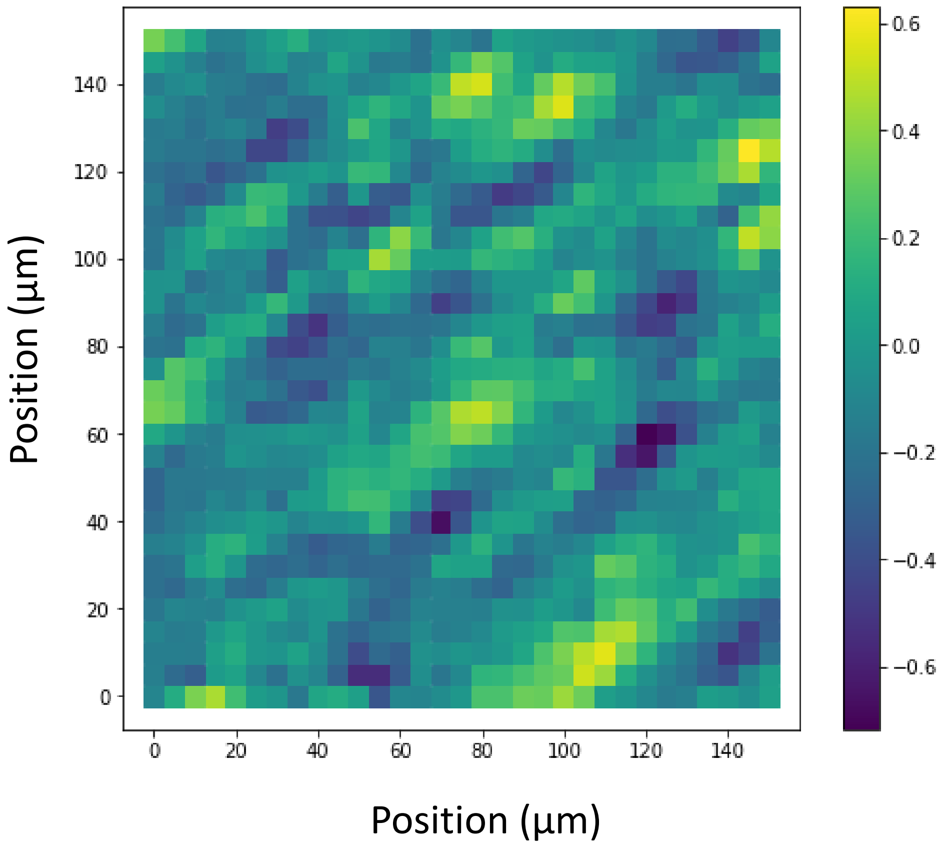
Bayesian optimization

Search for the most tilted position

Crystal orientation distribution predicted with 48 points



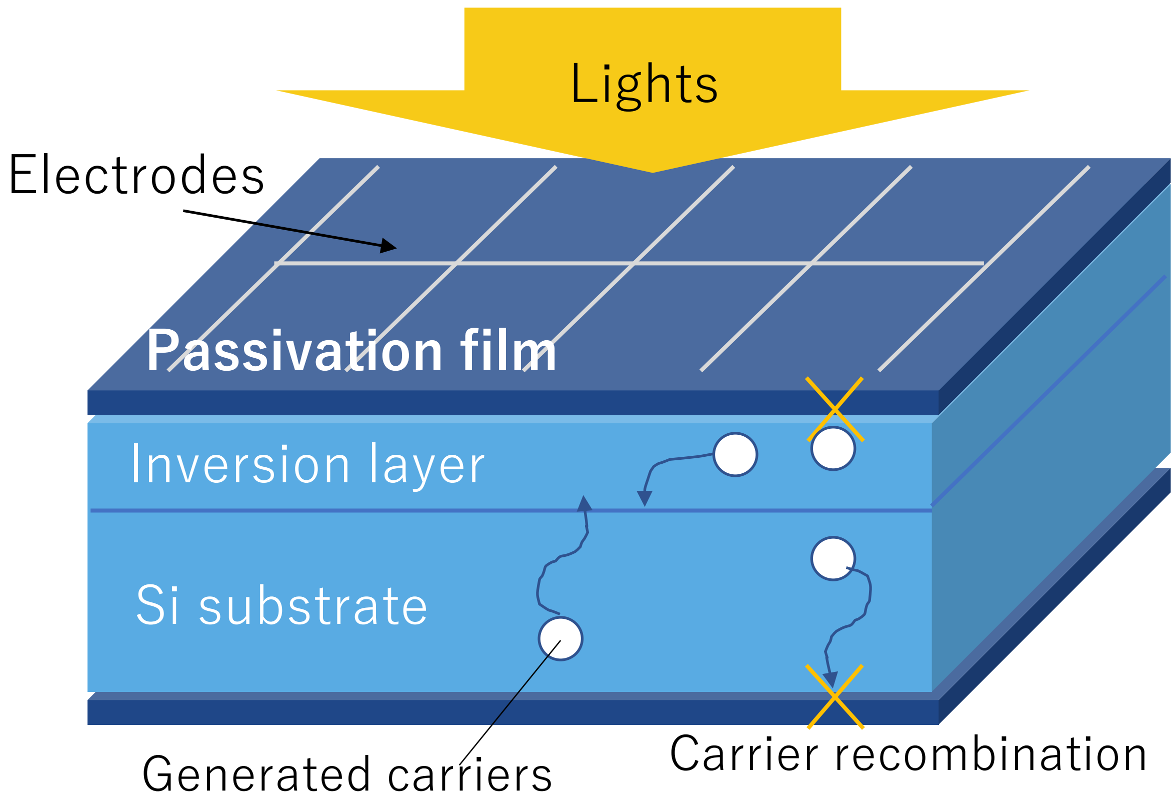
Crystal orientation distribution with 961 points



The most tilted position was obtained with small number of measurement.

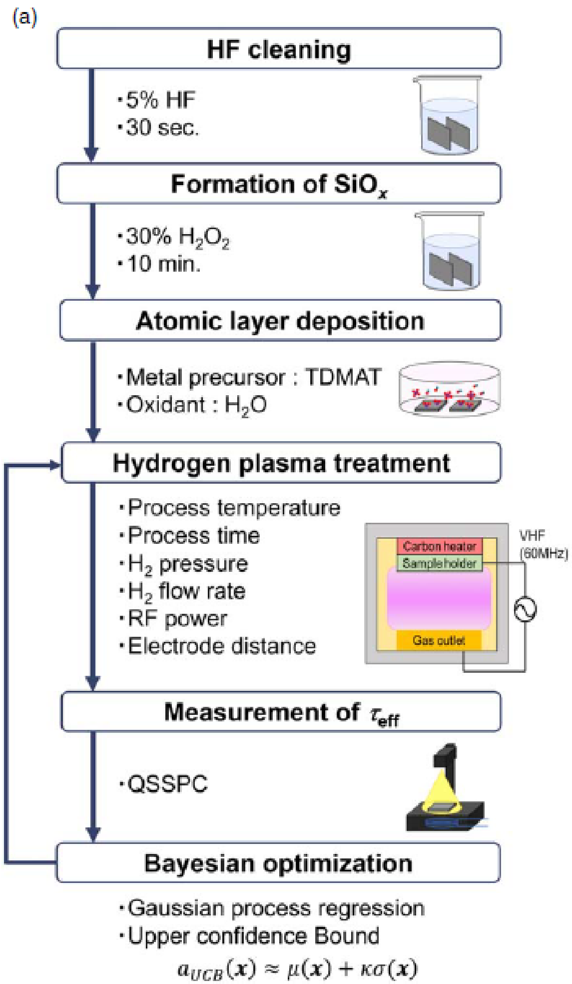
Application to hydrogen plasma treatment

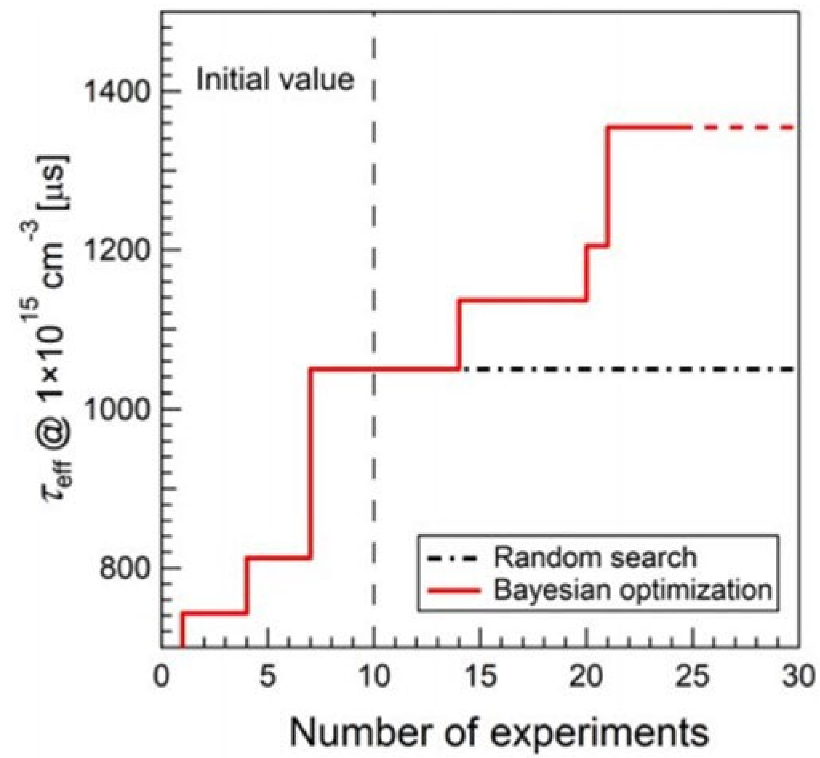
Solar cells and Hydrogen plasma treatment



Bayesian optimization for hydrogen plasma treatment

Surface passivation in TiOx/SiOy/c-Si heterostructure by hydrogen plasma treatment



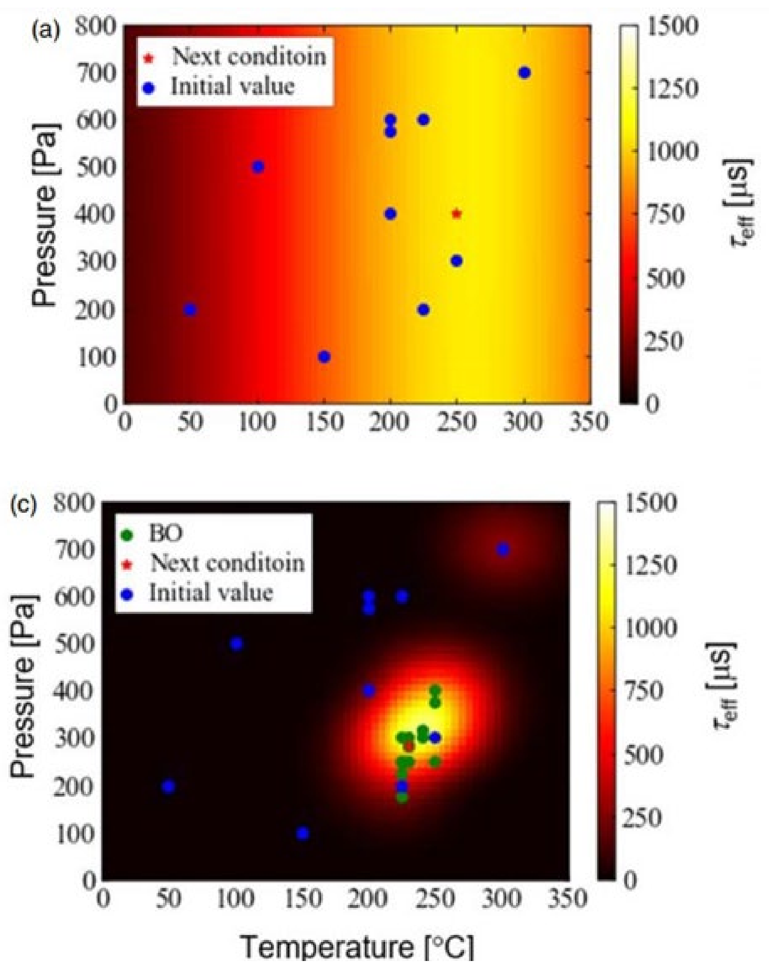


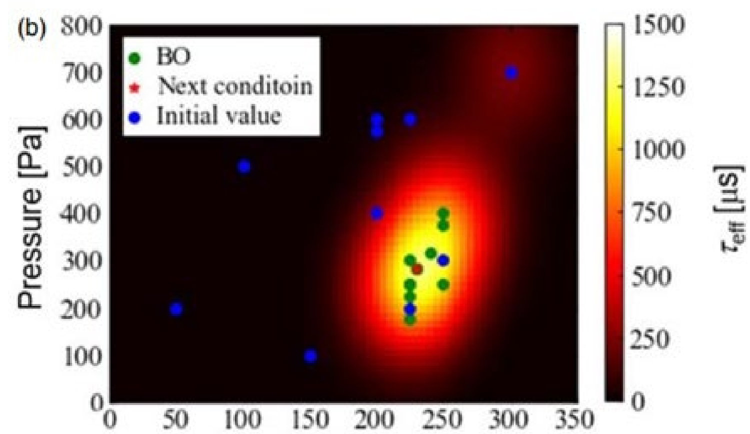
Carrier lifetime 436.3 μs → 1354.6 μs by 25 experiments.

S. Miyagawa, K. Gotoh, K. Kutsukake, Y. Kurokawa and N. Usami, Appl. Phys. Express 14, (2021) 025503.

Bayesian optimization for hydrogen plasma treatment

Surface passivation in TiOx/SiOy/c-Si heterostructure by hydrogen plasma treatment





Response surface of carrier lifetime after (a)10, (b) 20, and (c)25 cycles

S. Miyagawa, K. Gotoh, K. Kutsukake, Y. Kurokawa and N. Usami, Appl. Phys. Express 14, (2021) 025503.

Application to grinding process of SiC

Optimization of grinding process

Process parameters

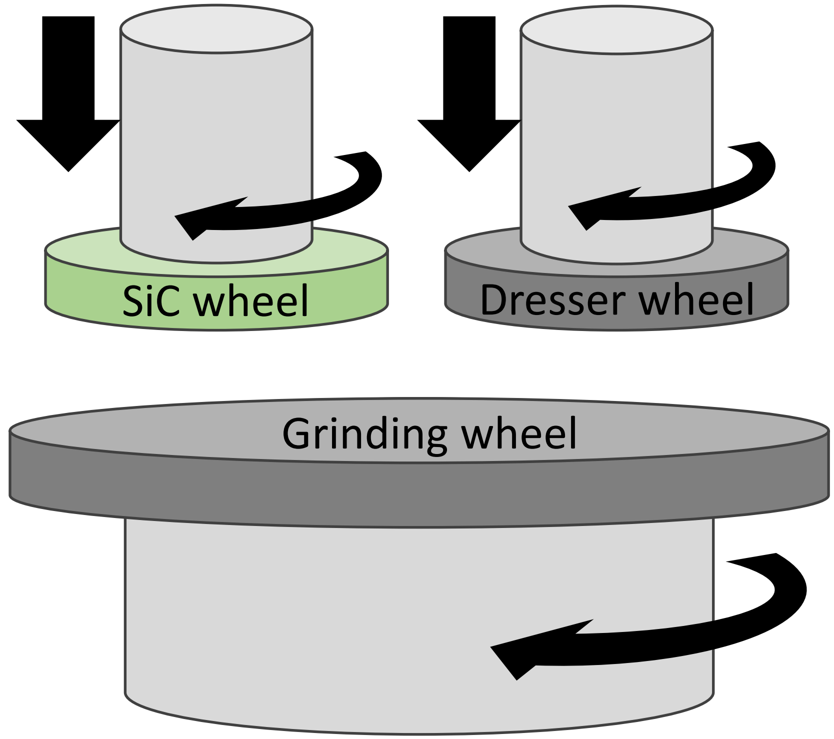
SiC wheel・Dresser wheel :
Rotation rate, Moving speed

Grinding wheel :
Rotation rate,
Hardness, Roughness

Objective parameters

Processing speed, Surface roughness

Difficulty: Complicated relationship between the process parameters and object parameters



SiC wafer grinding

We performed Bayesian optimization on grinding process.

Experimental flow chart

Initial data: 27 experiments at random condition

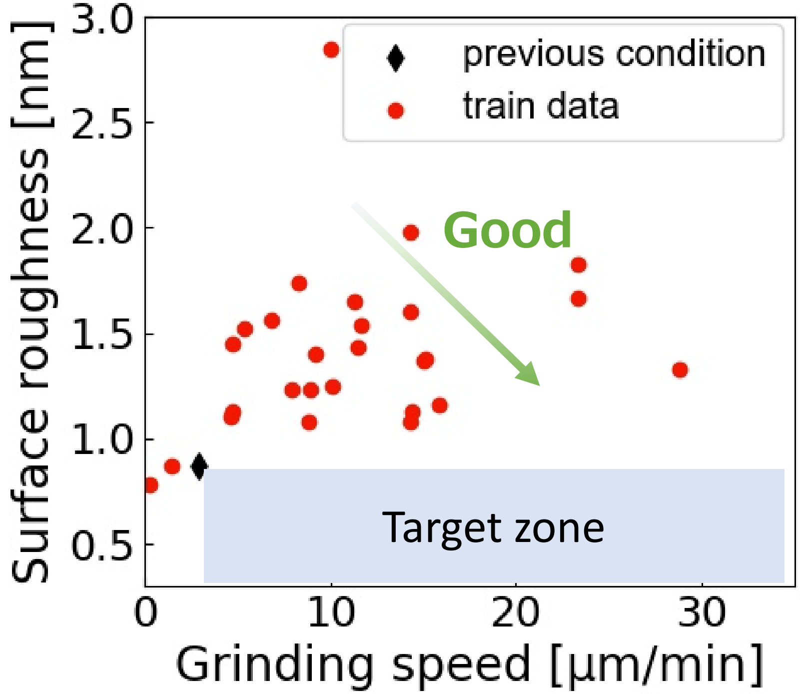
Bayesian optimization

1. **Gaussian process regression for processing speed, Surface roughness**
2. **Calculate EI (expected improvement)**
3. **Suggest next condition by maximize EI**
4. **Experiments with the suggested condition**
5. **Add new data points to the training data**

×32

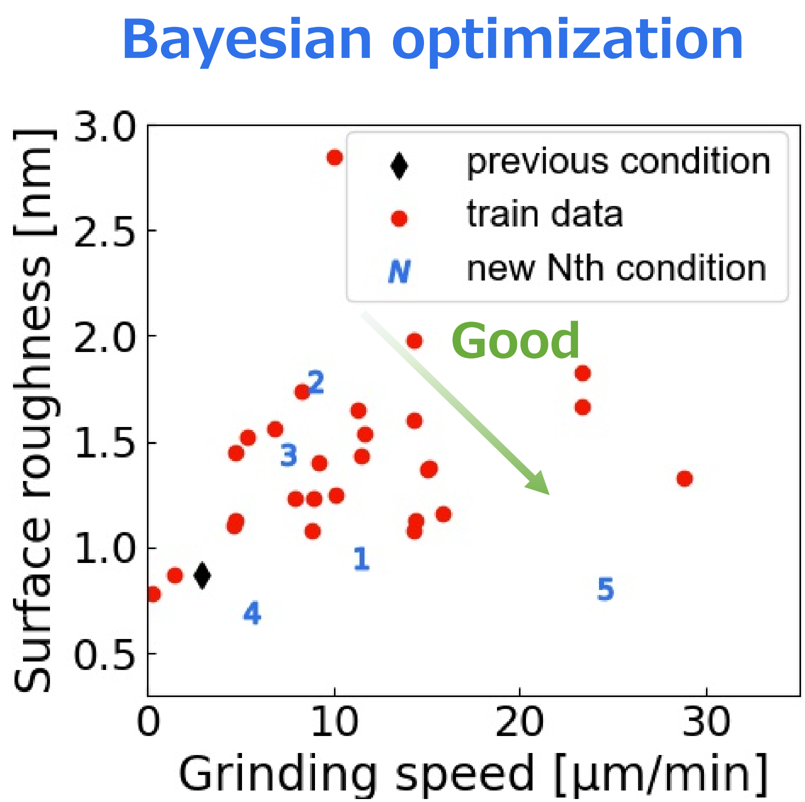
Bayesian optimization of grinding process

Initial data



There are no data in the target zone in the initial data.
→ It is difficult to find appropriate condition by random search.

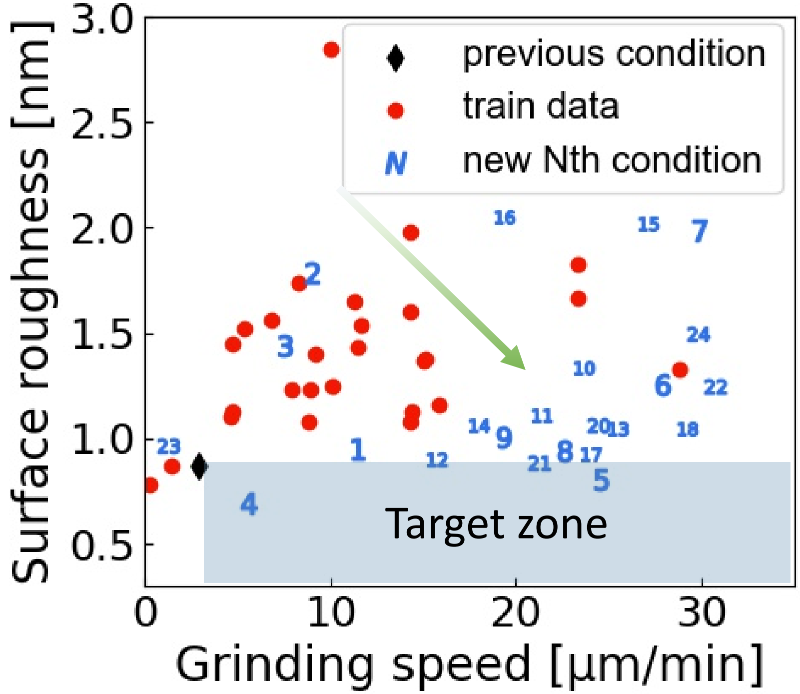
Bayesian optimization of grinding process



At the fifth experiment, grinding speed was significantly improved while maintaining the surface roughness.

Bayesian optimization of grinding process

Bayesian optimization



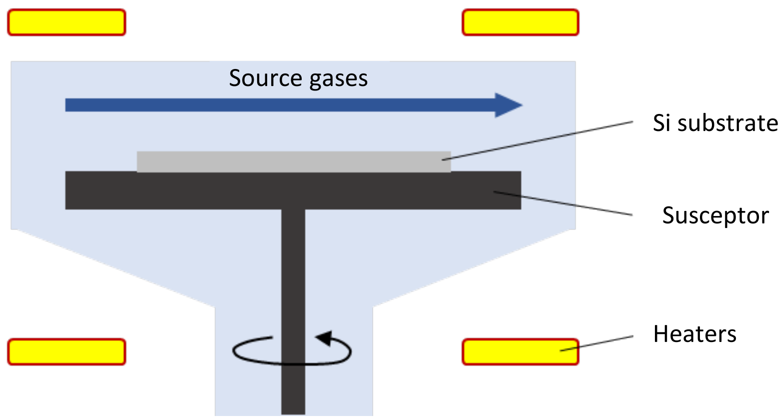
Also in the following experiments, appropriate conditions were obtained.

Application to epitaxial growth of Si

Bayesian optimization for epitaxial growth of Si

Maximize the **growth rate** while maintaining the **5 quality parameters**

Epitaxial growth of Si (CVD method)



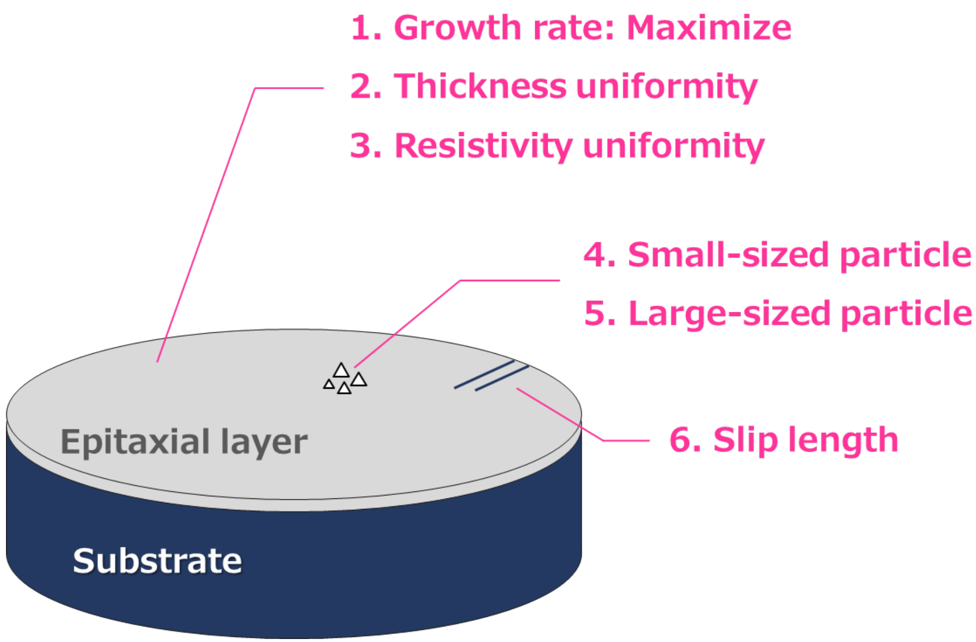
Process parameters (X: 12 parameters)

★ Temperatures (Substrate、Heaters、etc.)

★ Gases (Flux of source gases、etc.)

★ Configurations (Position, Rotation、etc.)

Growth rate and quality paramters



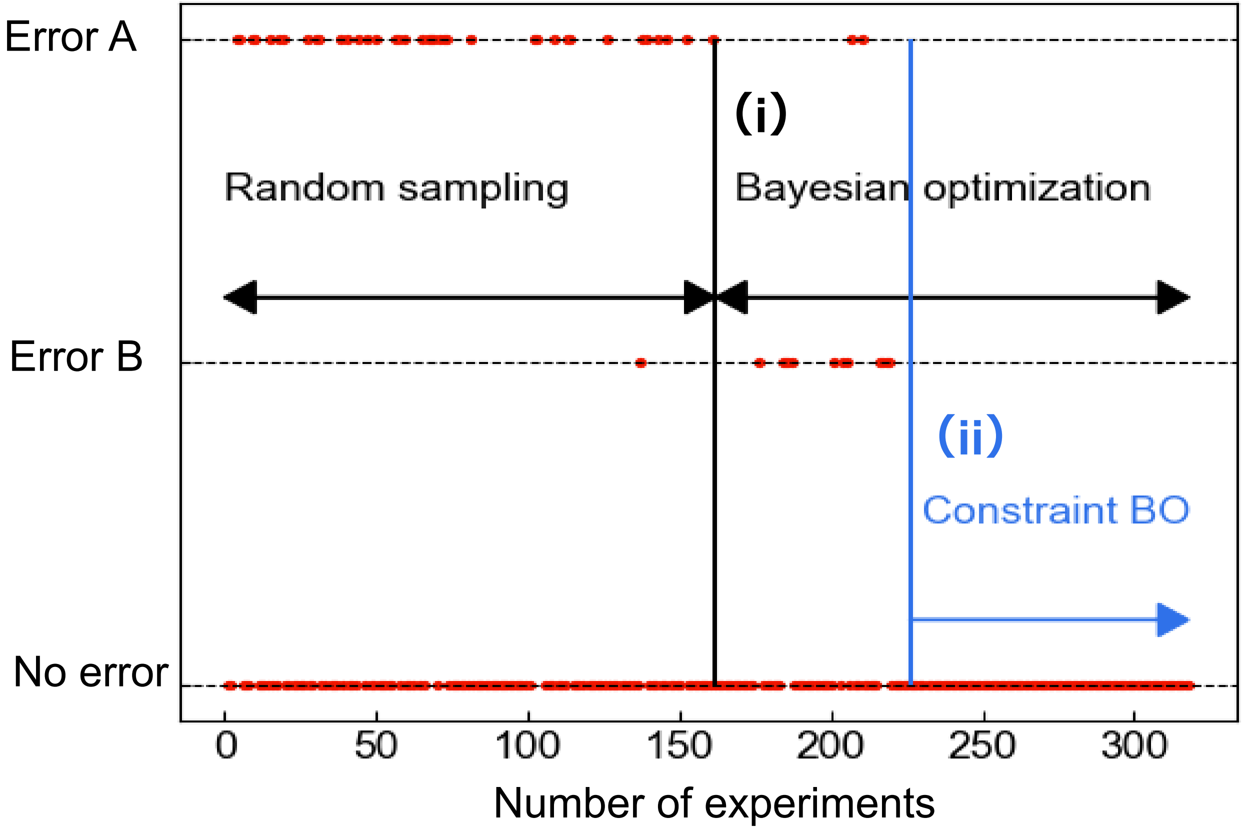
K. Osada, K. Kutsukake, J. Yamamoto, S. Yamashita, T. Kodera, Y. Nagai, T. Horikawa, K. Matsui, I. Takeuchi and T. Ujihra, Materials Today Communications, 25 (2020) 101538.

Bayesian optimization for epitaxial growth of Si

BO ➔ Explore new experimental conditions ➔ Unexpected error

(i) **Error A** : Limit the x parameter range.

(ii) **Error B** : Make a prediction model for the parameter related to error B and use it as a constraint.



Suppression of errors by constraint BO

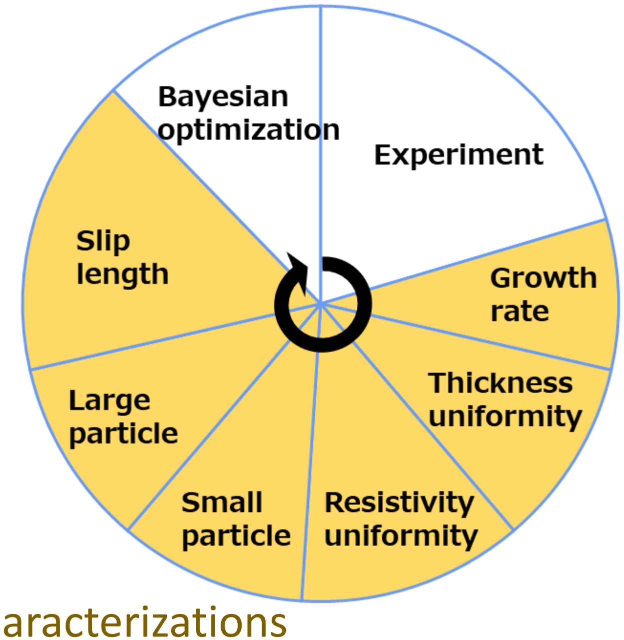
Bayesian optimization for epitaxial growth of Si

General BO: Number of experiments is limited.

This study: The time period for the optimization is limited.

→ Increase the number of data is effective.

Process times in 1 BO cycle



Objective function	Evaluation cost	Importance
1. Growth rate	**Low**	**High**
2. Thickness uniformity	**Low**	**High**
3. Resistivity uniformity	High	Low
4. Small particle	Low	Low
5. Large particle	Low	Low
6. Slip length	High	High

★Procedure 1: Consider only **2 parameters** (growth rate and thickness uniformity)

★Procedure 2: Consider **all the parameters**

Bayesian optimization for epitaxial growth of Si

Acquisition functions

Procedure1: Single Quality Constraint Bayesian Optimization (SQCBO)

$$x^* = \underset{x}{argmax}\, E_f[max\{f(x) - f^*, 0\}] \times E_g\left[max\left\{g_{threshold} - g(x), 0\right\}\right]$$

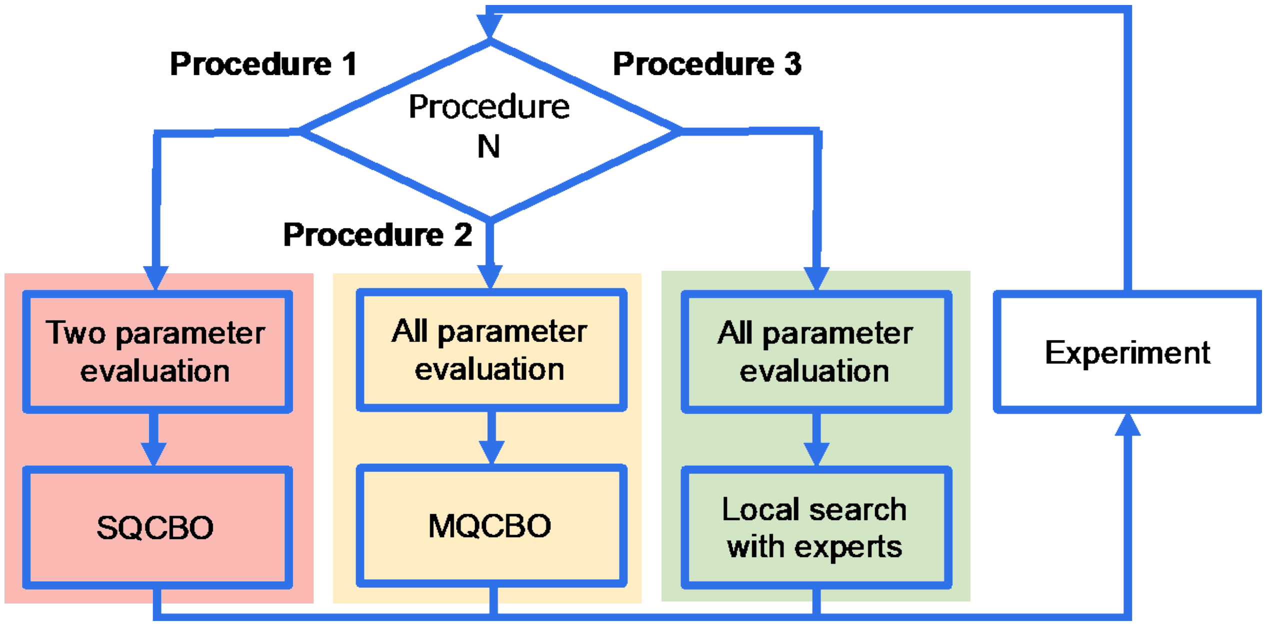
Procedure2: Multi Quality Constraint Bayesian Optimization (MQCBO)

$$x^* = \underset{x}{argmax}\, E_f[max\{f(x) - f^*, 0\}]$$

$$s.t. \quad E_{threshold} - E_g\left[max\left\{g_{threshold} - g(x), 0\right\}\right] \leq 0$$

$E[\cdot]$: *Expected value calculation operator*
f^* : *Maximum value of growth rate in the measured data*
$g_{threshold}$: *Threshold for quality parameters*
$E_{threshold}$: *Threshold fo EI*

Bayesian optimization for epitaxial growth of Si



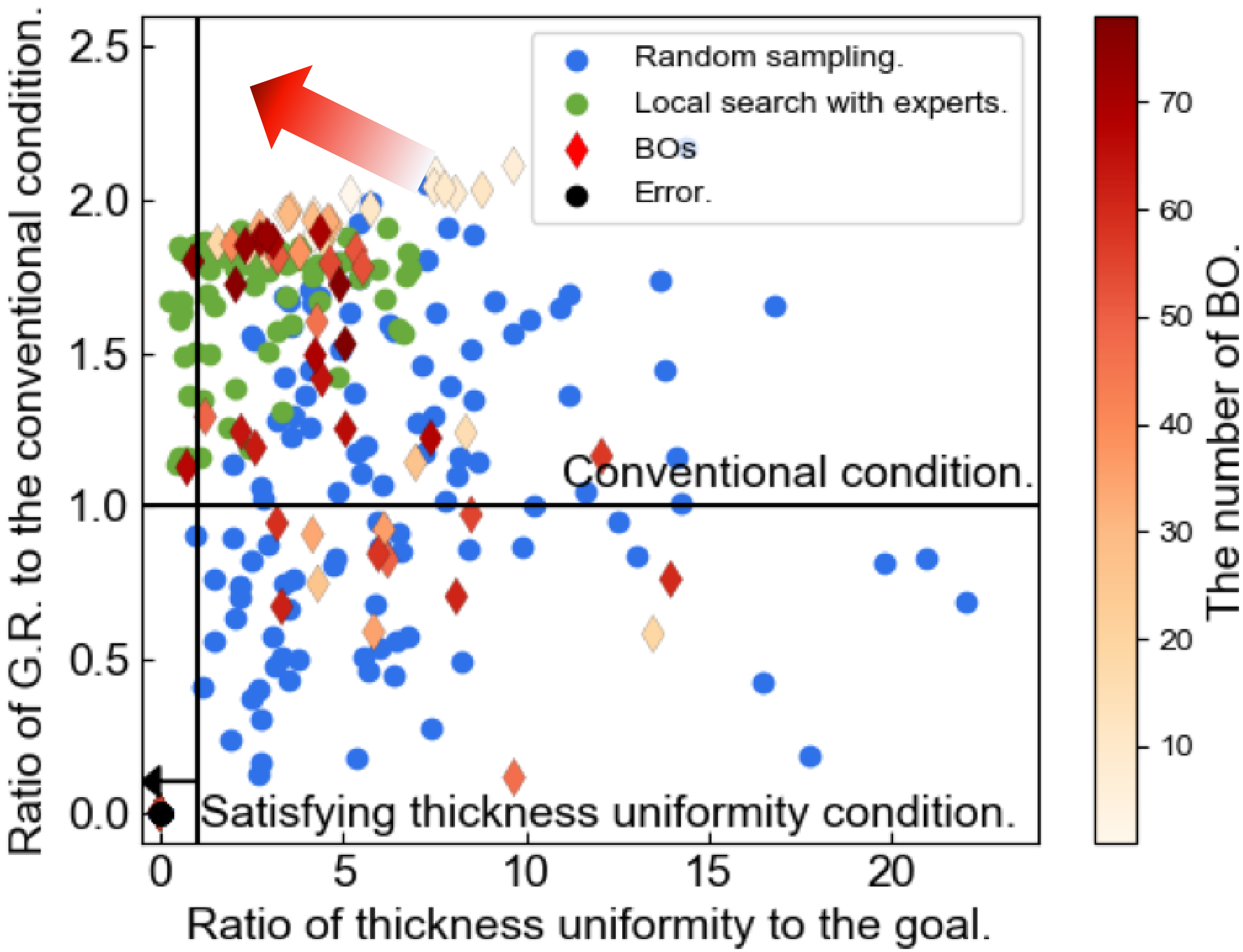
Flowchart of the adaptive constraint BO

K. Osada, K. Kutsukake, J. Yamamoto, S. Yamashita, T. Kodera, Y. Nagai, T. Horikawa, K. Matsui, I. Takeuchi and T. Ujihra, Materials Today Communications, 25 (2020) 101538.

Procedure 1 : SQCBO: Single Quality Constraint Bayesian Optimization
Procedure 2 : MQCBO: Multi Quality Constraint Bayesian Optimization
Procedure 3 : **Local search with experts.**

Bayesian optimization for epitaxial growth of Si

Growth rate v.s. Thickness uniformity



1.8 times higher growth rate was achieved while maintaining the quality parameters within the requirements.

K. Osada, K. Kutsukake et al., Materials Today Communications, 25 (2020) 101538.

Summary

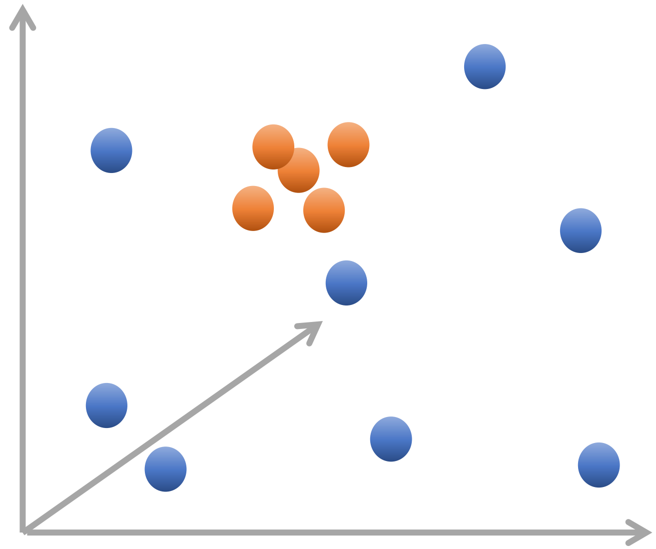
Using Bayesian optimizaiton,

- ✓ **The most tilted position** in XRD measurement of SiGe film.
- ✓ **High passivation performance** in TiOx/SiOy/c-Si heterostructure by hydrogen plasma treatment
- ✓ **Significantly improved condition** in grinding of SiC wafer
- ✓ **1.8 times higher growth rate** in epitaxial growth of Si

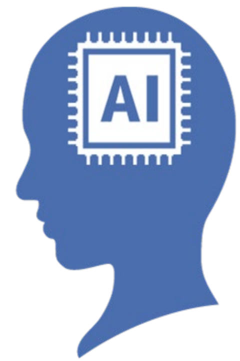
was obtained with a small number of experiments.

Summary

Scientists, Engineers
- ✓ Low dimensional space
- ✓ Domain knowledge
- ✓ Local optimization



Machine learning (Bayesian opt.)
- ✓ High dimensional space
- ✓ Global optimization



Collaboration between scientists and machine learning
➔ Further development in the science and technology.

Acknowledgements

This study was supported by the Japan Society for the Promotion of Science through KAKENHI (JP18K19033), the New Energy and Industrial Technology Development Organization (NEDO), MEXT, Grants-in-Aid for Scientific Research on Innovative Areas "Hydrogenomics" (JP18H05514), and the Center for Advanced Intelligence Project, RIKEN.

The author acknowledges
GlobalWafers Japan Co., Ltd.: Yuta Nagai, Tomoyuki Horikawa, Hironori Banba, Jun Yamamoto, Shigeo Yamashita, Takashi Kodera, Koji Izunome, Takashi Ishikawa, Hisashi Matsumura, and Hiroyuki Tsubota
Tohoku University: Kensaku Maeda
Nagoya University: Toru Ujihara, Keiichi Osada, Takashi Nakano, Noritaka Usami, Shinsuke Miyagawa, Kazuhiro Gotoh, and Yasuyoshi Kurokawa
RIKEN: Ichiro Takeuchi
nitole Corporation: Kiyoshi Narita, Ryong-Seok Doi, and Yukihisa Takeda

Tutorial

Please learn about

× How to write python codes

○ How incorporate your scientific domain knowledge into the machine learning model and optimization flow

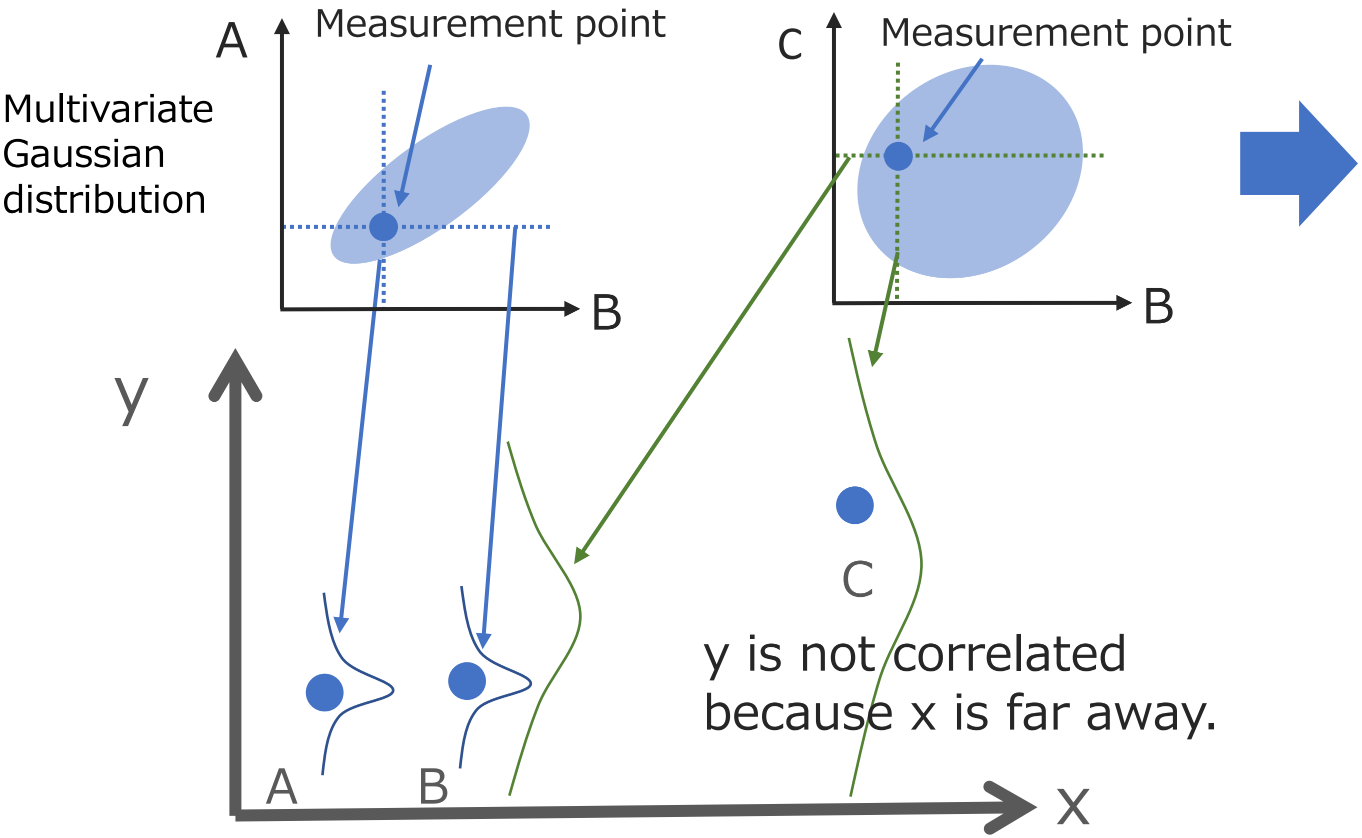
White box optimization

We can incorporate our scientific domain knowledge.

Black box optimization	White box optimization
Automatic optimization libraries (Gpyopt, Optuna, etc.)	Further reduction of the number of experiments
Automatic optimization	We can understand the reason why BO suggested this next condition.
	We can overcome some BO trouble.

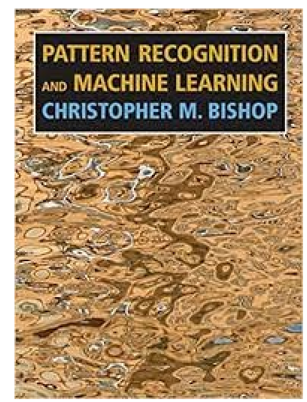
Gaussian process regression (GPR)

Basic idea: if x is similar, y should be similar.
→ Gaussian process: consider the data points to be slices of an infinite-dimensional multivariate Gaussian distribution.



Do this operation between all observation points including x you want to estimate→ Gaussian process regression

Please refer to this textbook for details.



Gaussian process regression (GPR)

Kernel function: a function that gives the degree of similarity (covariance) of data

Gaussian kernel (RBF kernel)

$$k(\mathrm{x}, \mathrm{x}') = \theta_1 \exp\left(-\frac{|x - x'|^2}{\theta_2}\right)$$

Hyper parameters

Constant term
(Effects similar to amplitude) → θ_1

Length scale
(Effects similar to wave length → θ_2

Gaussian process regression adjusts these hyperparameters like weights in other regressions to get a regression result that fits the data.

Acquisition function

Acquisition function: index for determining the next experimental condition

PI : Probability of improvement $$\alpha_{\mathrm{PI}}(\mathbf{x};\mathcal{D}_n) := \mathbb{P}[v > \tau] = \Phi\left(\frac{\mu_n(\mathbf{x}) - \tau}{\sigma_n(\mathbf{x})}\right)$$

EI : Expected improvement
$$\begin{aligned}\alpha_{\mathrm{EI}}(\mathbf{x};\mathcal{D}_n) &:= \mathbb{E}\left[I(\mathbf{x}, v, \theta)\right] \\ &= (\mu_n(\mathbf{x}) - \tau)\Phi\left(\frac{\mu_n(\mathbf{x}) - \tau}{\sigma_n(\mathbf{x})}\right) \\ &\quad + \sigma_n(\mathbf{x})\phi\left(\frac{\mu_n(\mathbf{x}) - \tau}{\sigma_n(\mathbf{x})}\right),\end{aligned}$$

UCB : Upper Confidence Bound $$\alpha_{\mathrm{UCB}}(\mathbf{x};\mathcal{D}_n) := \mu_n(\mathbf{x}) + \beta_n\sigma_n(\mathbf{x})$$

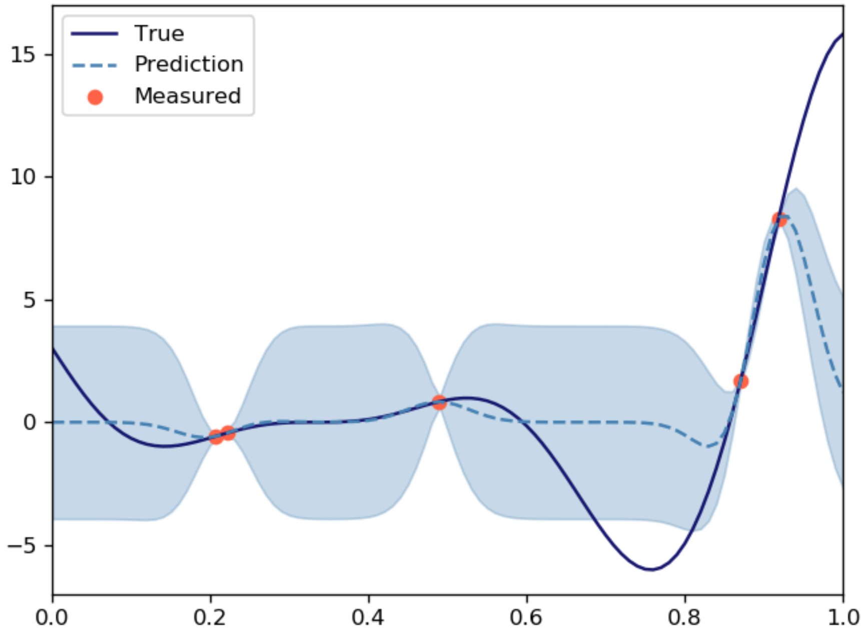
In the tutorial, we use lower confidence bound (LCB).

B. Shahriari, K. Swersky, Z. Wang, R. P. Adams and N. de Freitas, "Taking the Human Out of the Loop: A Review of Bayesian Optimization," in Proceedings of the IEEE, vol. 104, no. 1, pp. 148-175, Jan. 2016, doi: 10.1109/JPROC.2015.2494218.

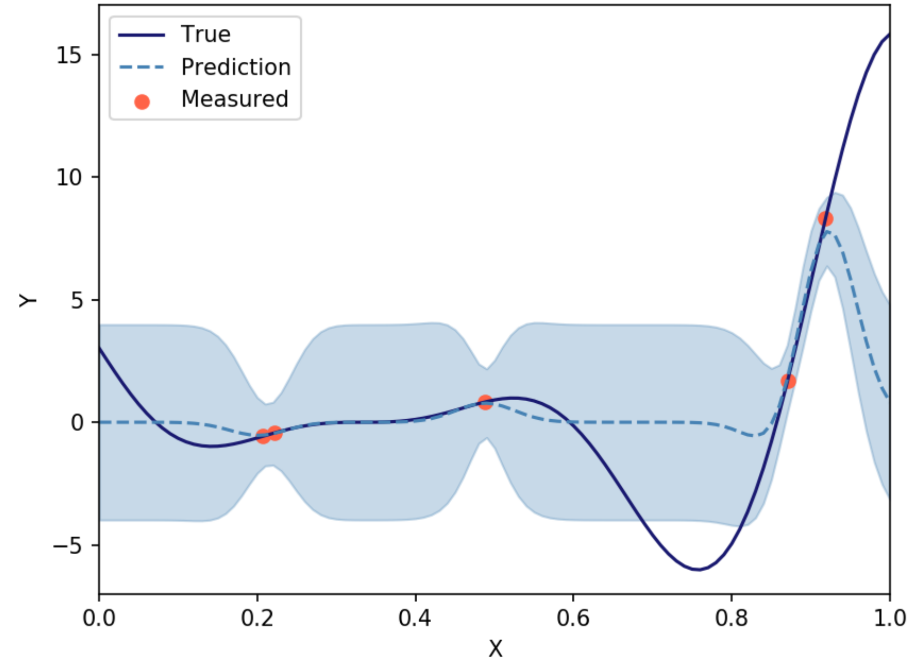
GPR : Influence of hyperparameter range

White noise range

WhiteKernel range: 1E-2～1E2



WhiteKernel range: Set to 1

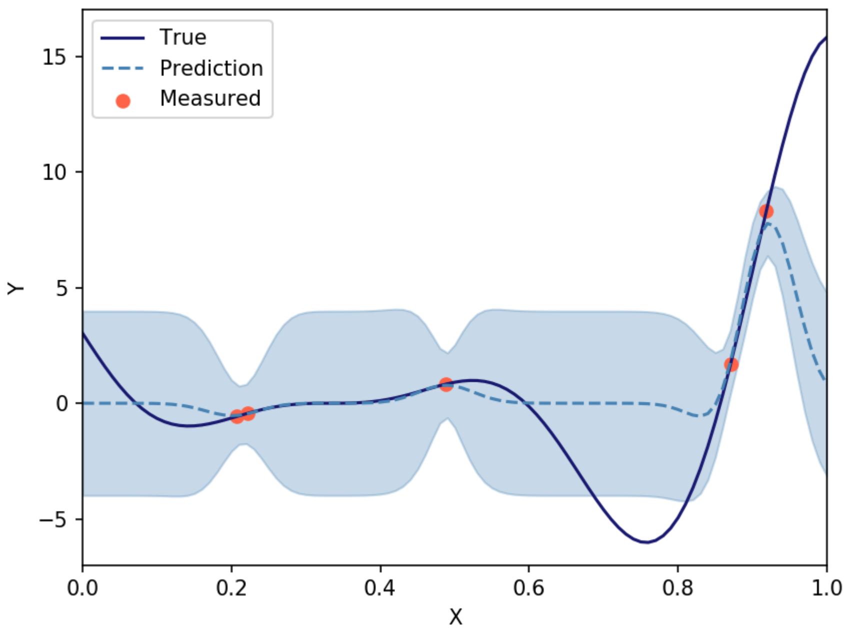


Noise level constraints experimentally obtained or based on expert knowledge lead to more reasonable regression.

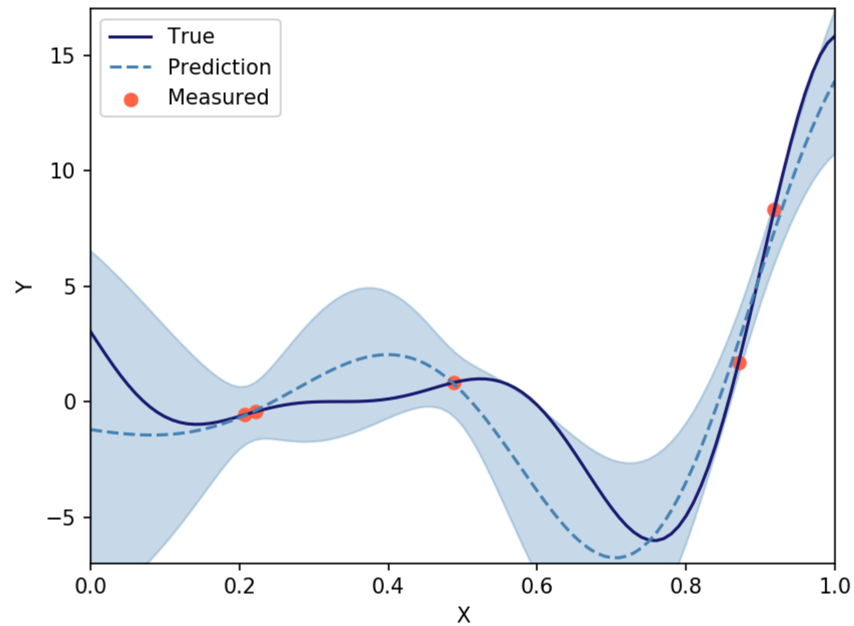
GPR : Influence of hyperparameter range

Length scale range

RBF range : 1E-2～1E2



RBF range : 2E-1～1E2



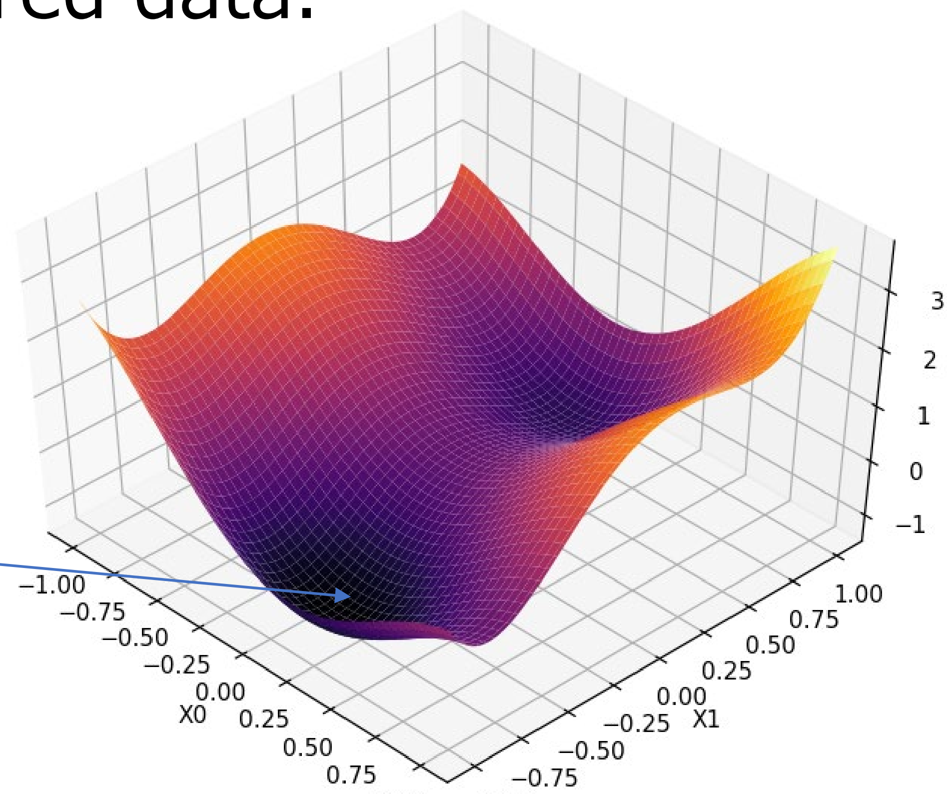
Length scale constraints based on expert knowledge lead to more reasonable regression.

2D Gaussian process regression

2D Gaussian process regression

Find the combination of x0 and x1 that gives the minimum y using the measured data.

Find this position



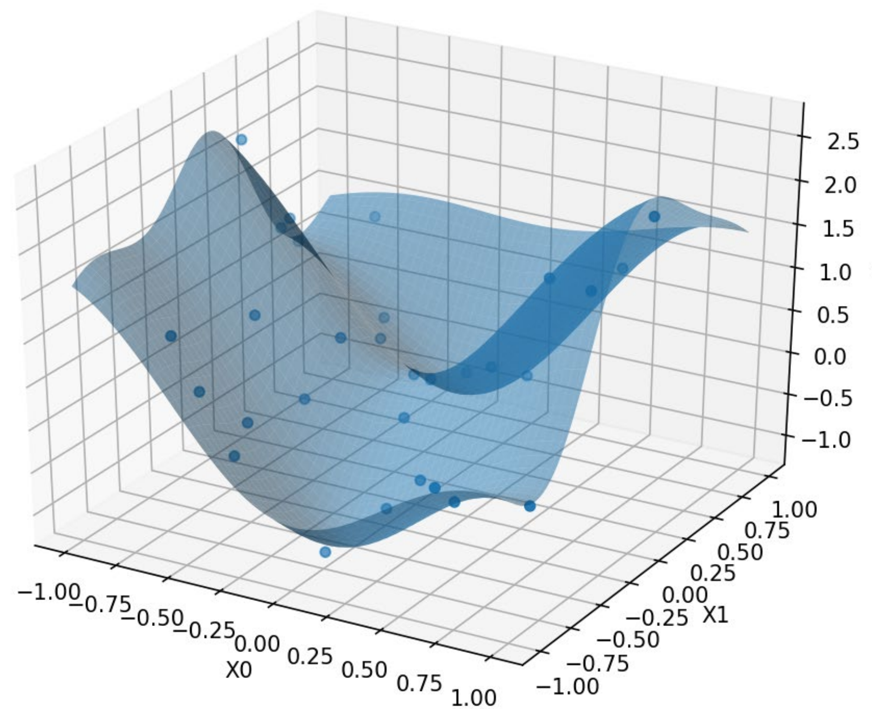
Benchmark function

2D Gaussian process regression

High expression ability
Regression of complex surfaces with high nonlinearity

Important points
1. Does not regress well on high dimensional inputs.

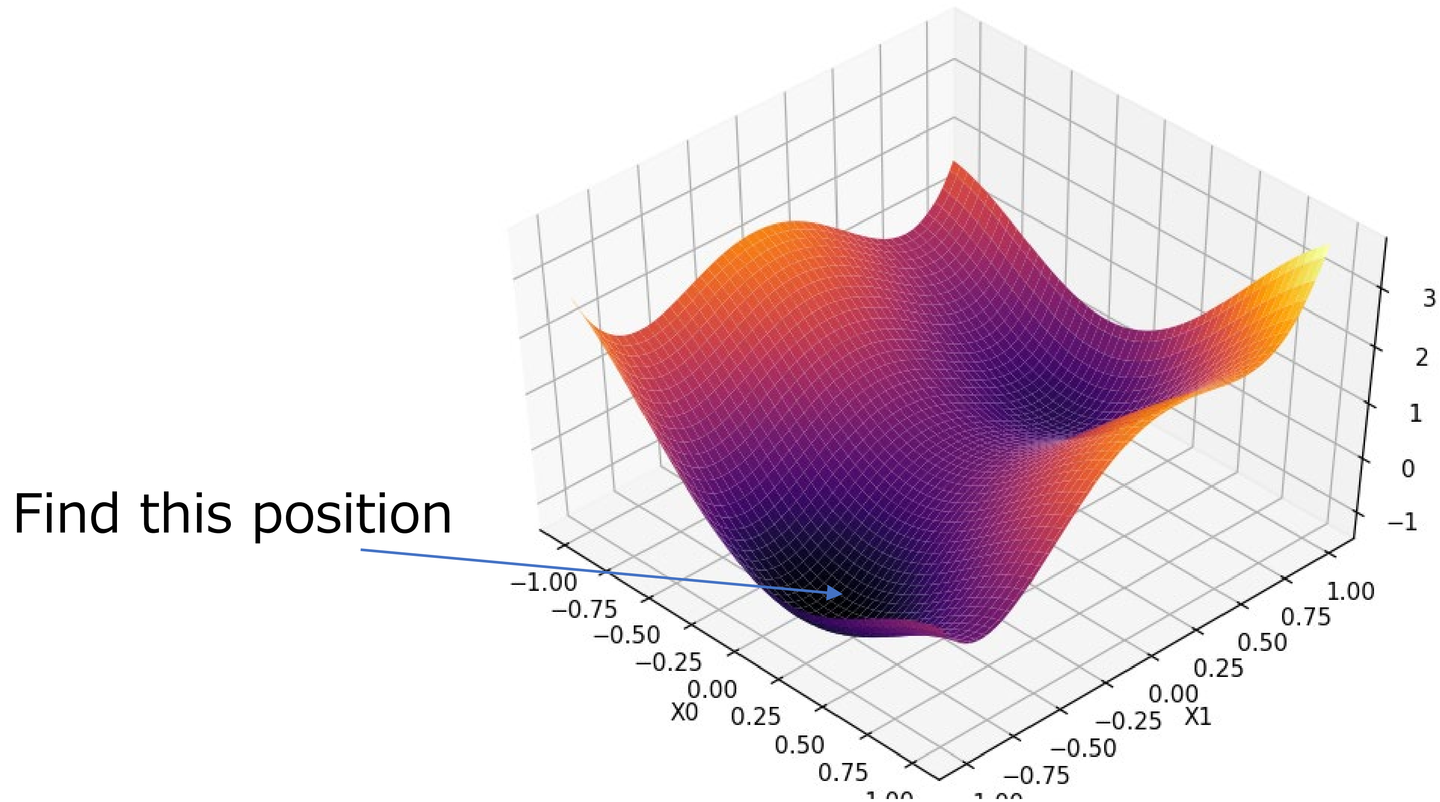
2. The number of computations increases with the cube of the number of data.



2D Bayesian optimization

2D Bayesian optimization

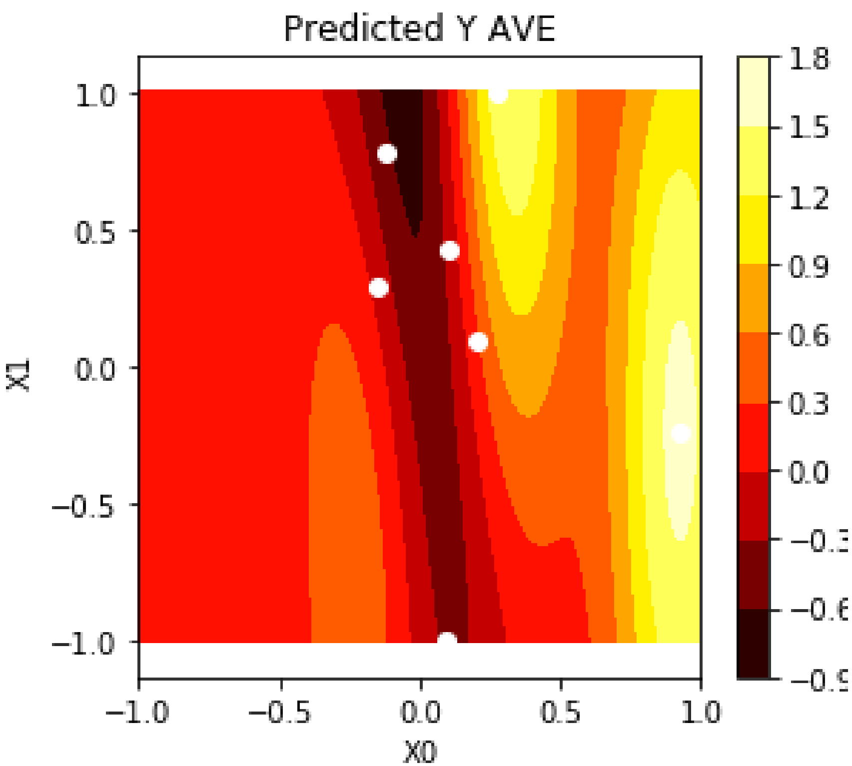
Find the combination of x0 and x1 that gives the minimum y using sequential optimization starting from 5 data.



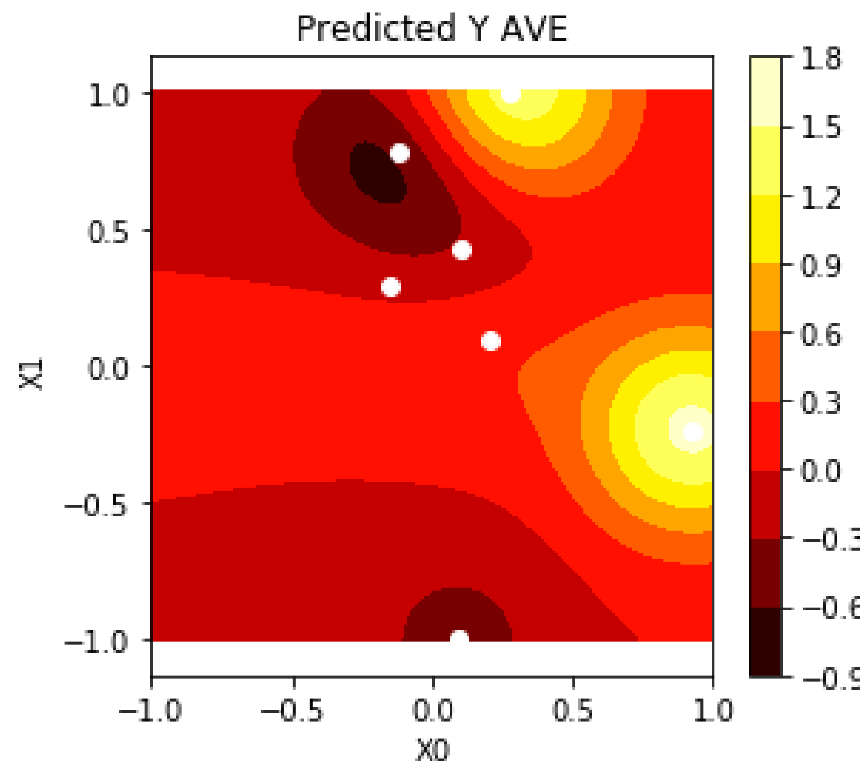
Benchmark function

GPR : Influence of anisotropy

With spatial anisotropy



Without spatial anisotropy



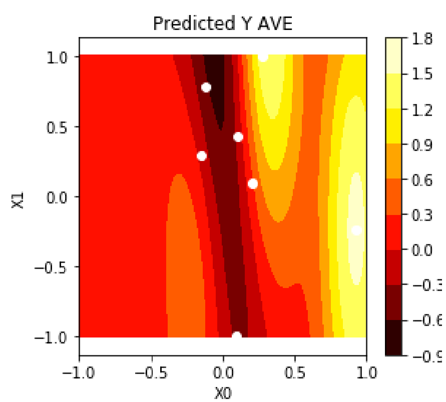
Whether X0 and X1 share the RBF length scale or not

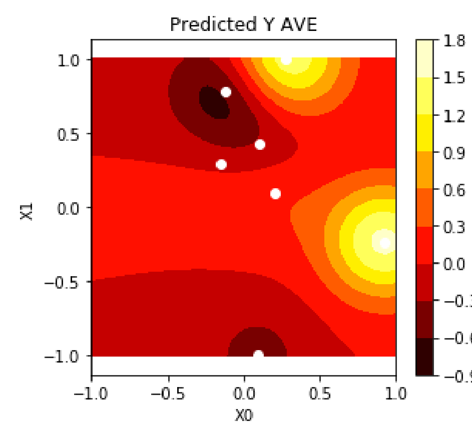
Important points: Parameter settings

Scientific knowledge Ideal result you want

Lead to forgery

Hyperhyperparameters:
Length scale range, Noise level range, Anisotropy, etc.
Hyperparameters:
κ in UCB, normalization of x and y, etc.





Important points: Global prediction

Bayesian optimization is an optimization method and
NOT global prediction method.

We cannot use GPR model obtained through BO for global prediction.