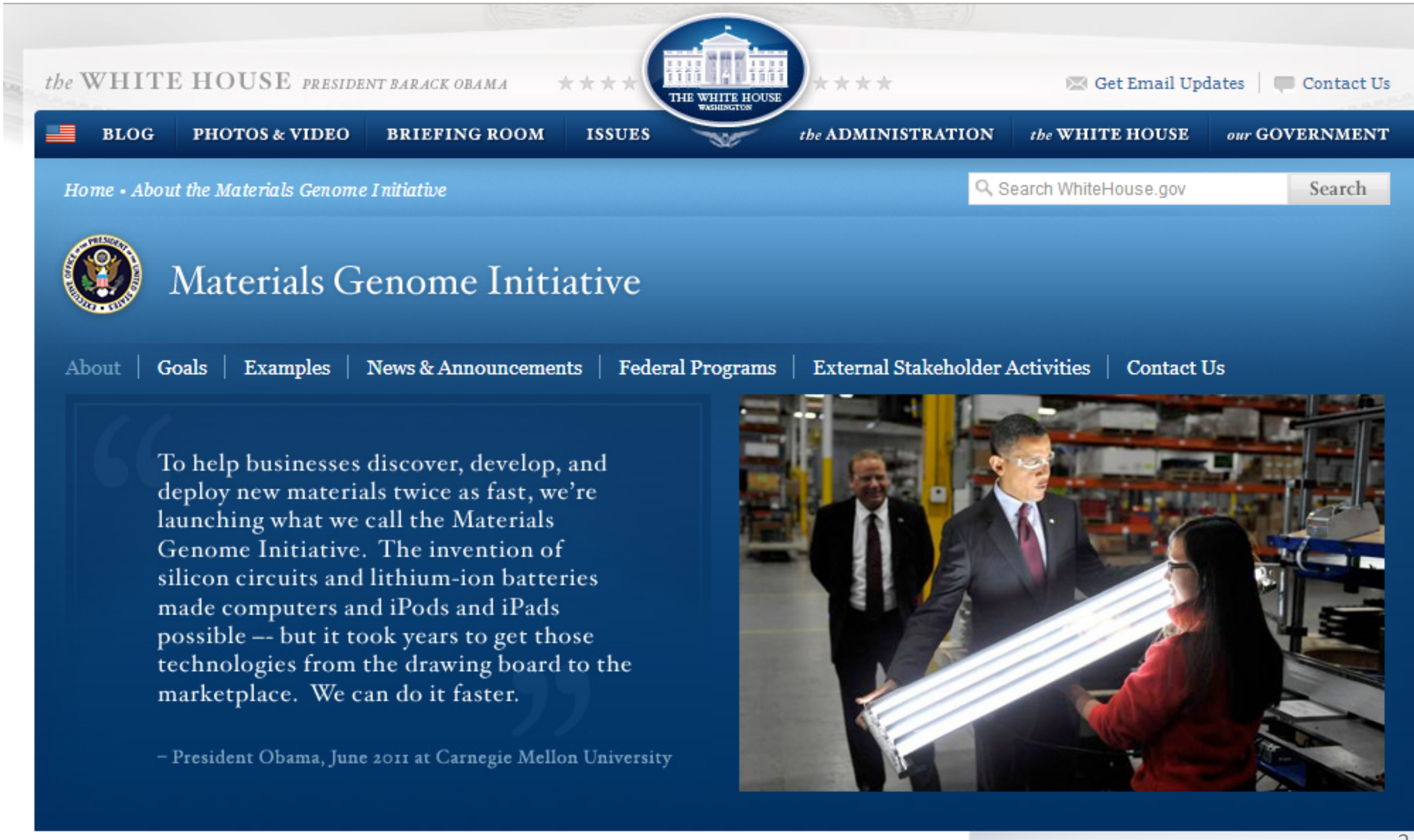
はじめに
ベイズ最適化の紹介

津田　宏治
University of Tokyo / NIMS / RIKEN AIP

Discovery of new functional molecules and materials is of national importance



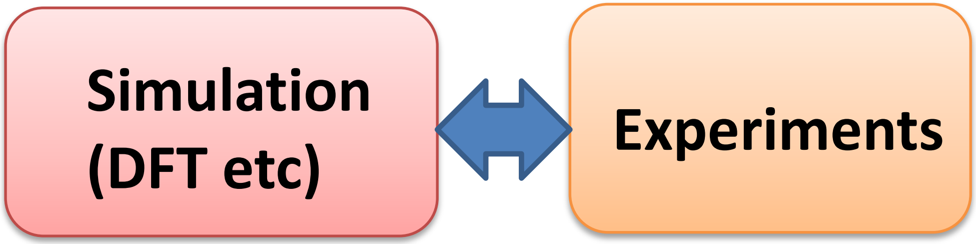
First Principles Calculations

Accurate, Slow

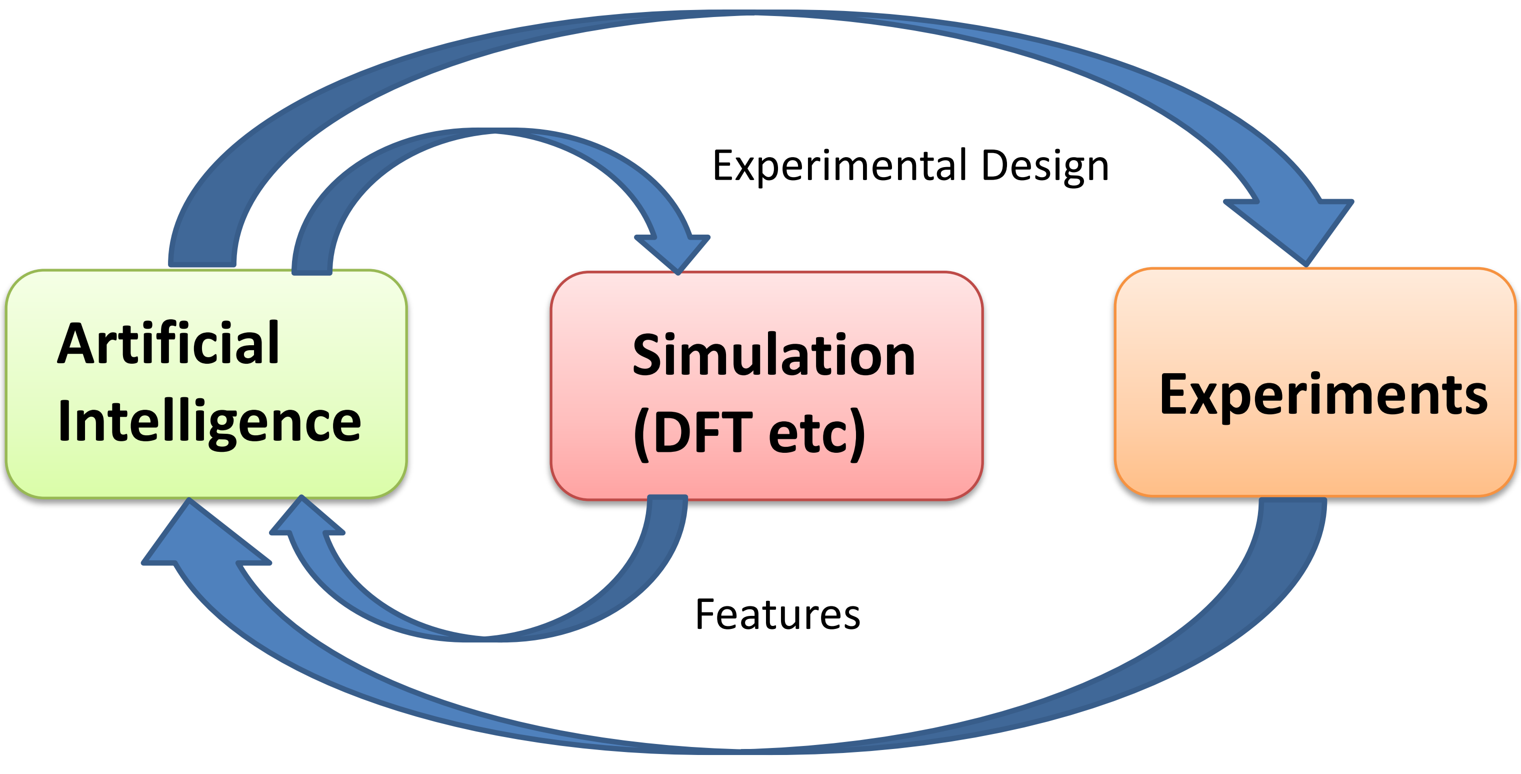
- Full configuration interaction
- Wave function based
- Density functional theory
- Semi-empirical
- Empirical potentials

Inaccurate, Fast

Old Picture



New Picture


Experimental Design
Artificial Intelligence
Simulation (DFT etc)
Experiments
Features

Materials Design Examples

- Design of Si-Ge nanostructures (Phys Rev X 2017)

- De novo design of organic compounds with desired absorption wavelength

Screening by first principles calculations alone

Mat. 1	Mat. 2	Mat. 3	Mat. 4	Mat. 5	Mat. 6	Mat. 7	Mat. 8	Mat. 9	Mat. 10

First Principles Calc.

Score 1	Score 2	Score 3	Score 4	Score 5	Score 6	Score 7	Score 8	Score 9	Score 10

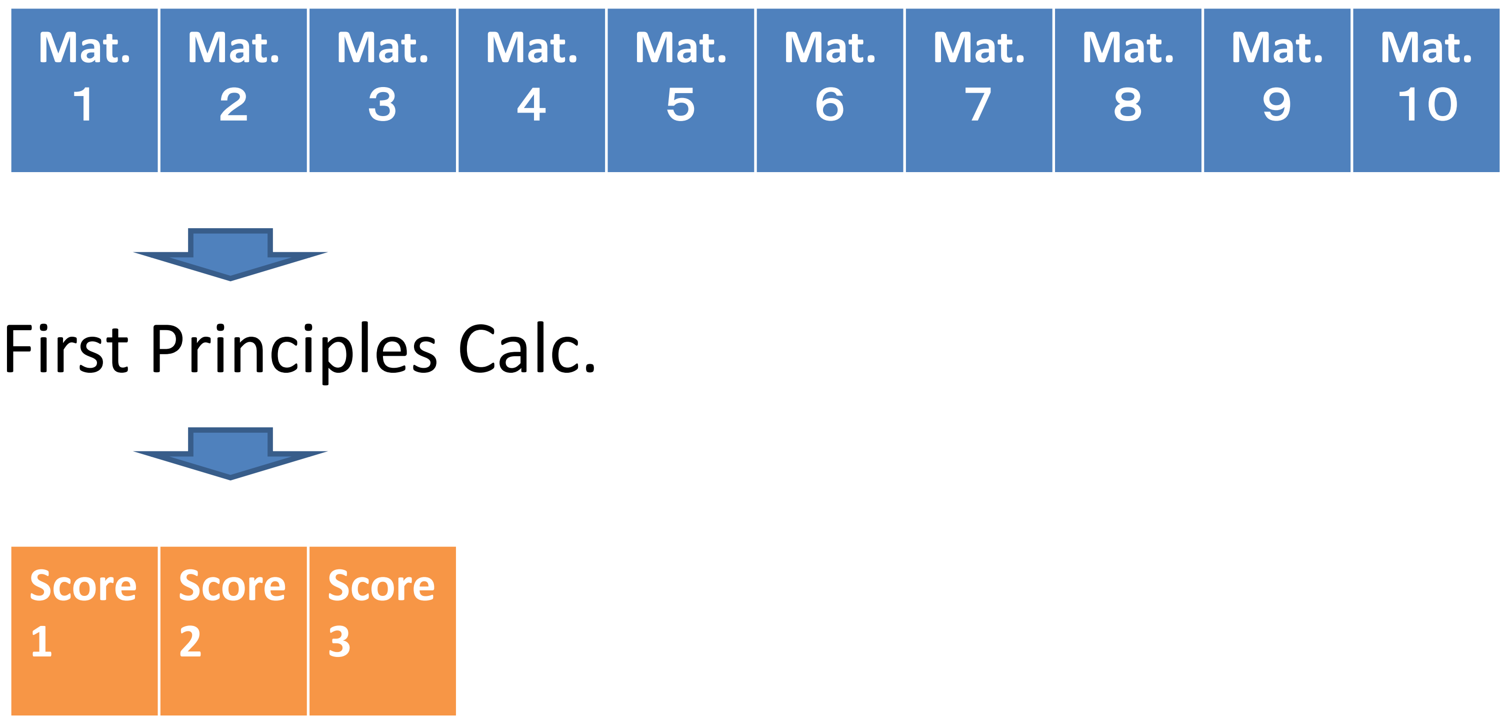
Bayesian Optimization

(Jones et al., 1998)

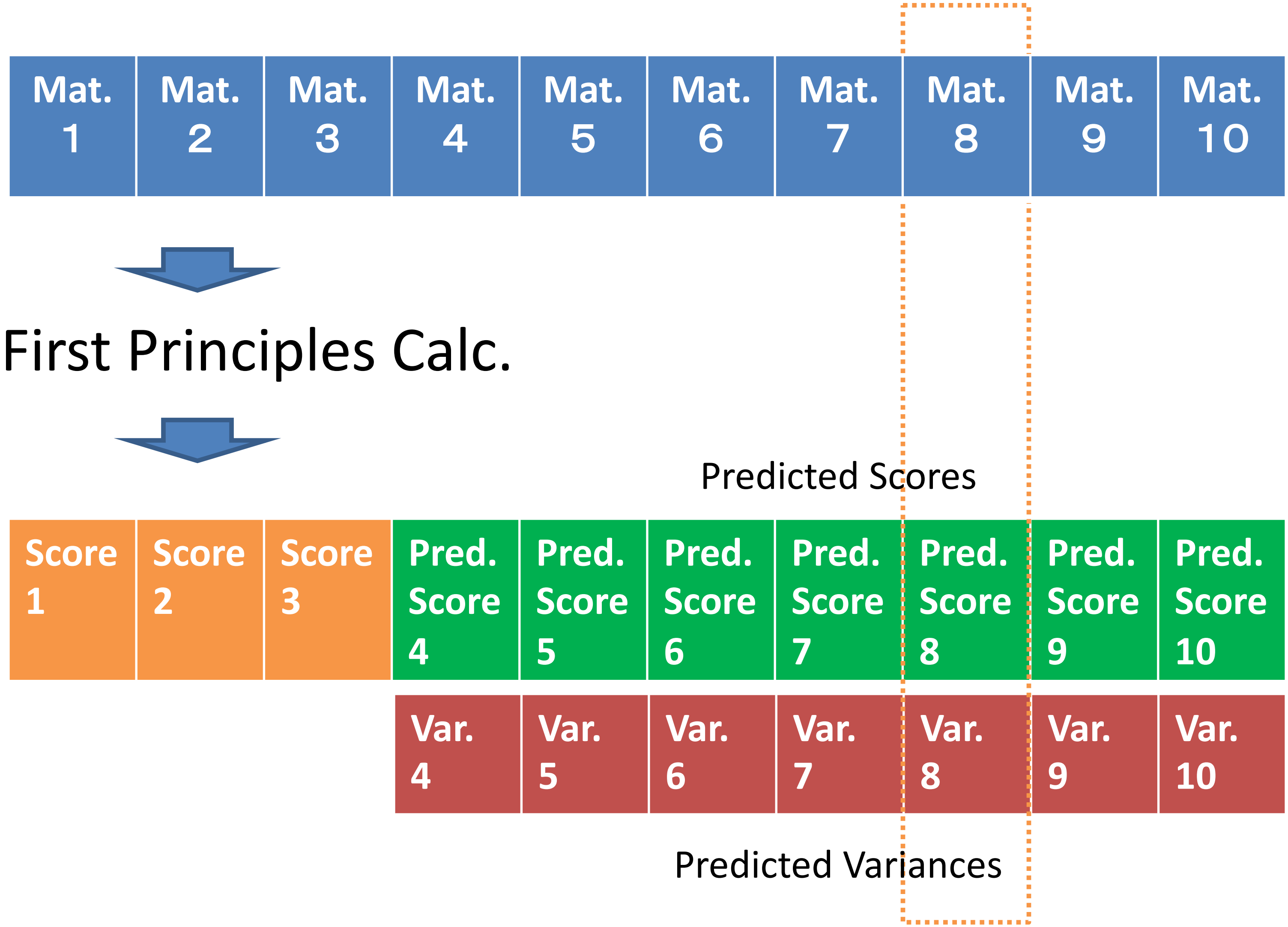
- Find best data points with minimum number of observations

- Choose next point to observe to discover the best ones as early as possible

Bayesian Optimization (1)



Bayesian Optimization (2)



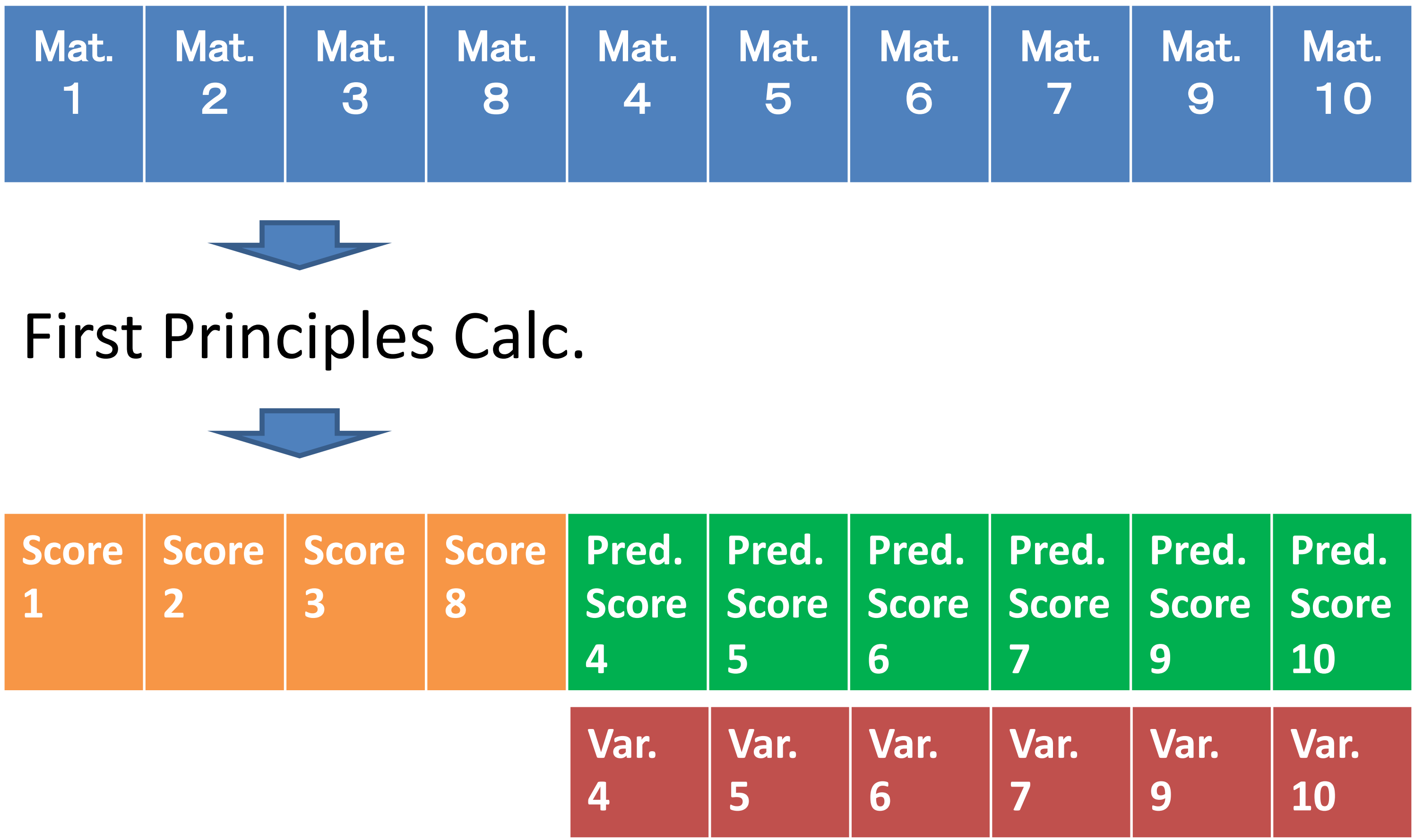
Bayesian Optimization (3)

Mat. 1	Mat. 2	Mat. 3	Mat. 8	Mat. 4	Mat. 5	Mat. 6	Mat. 7	Mat. 9	Mat. 10

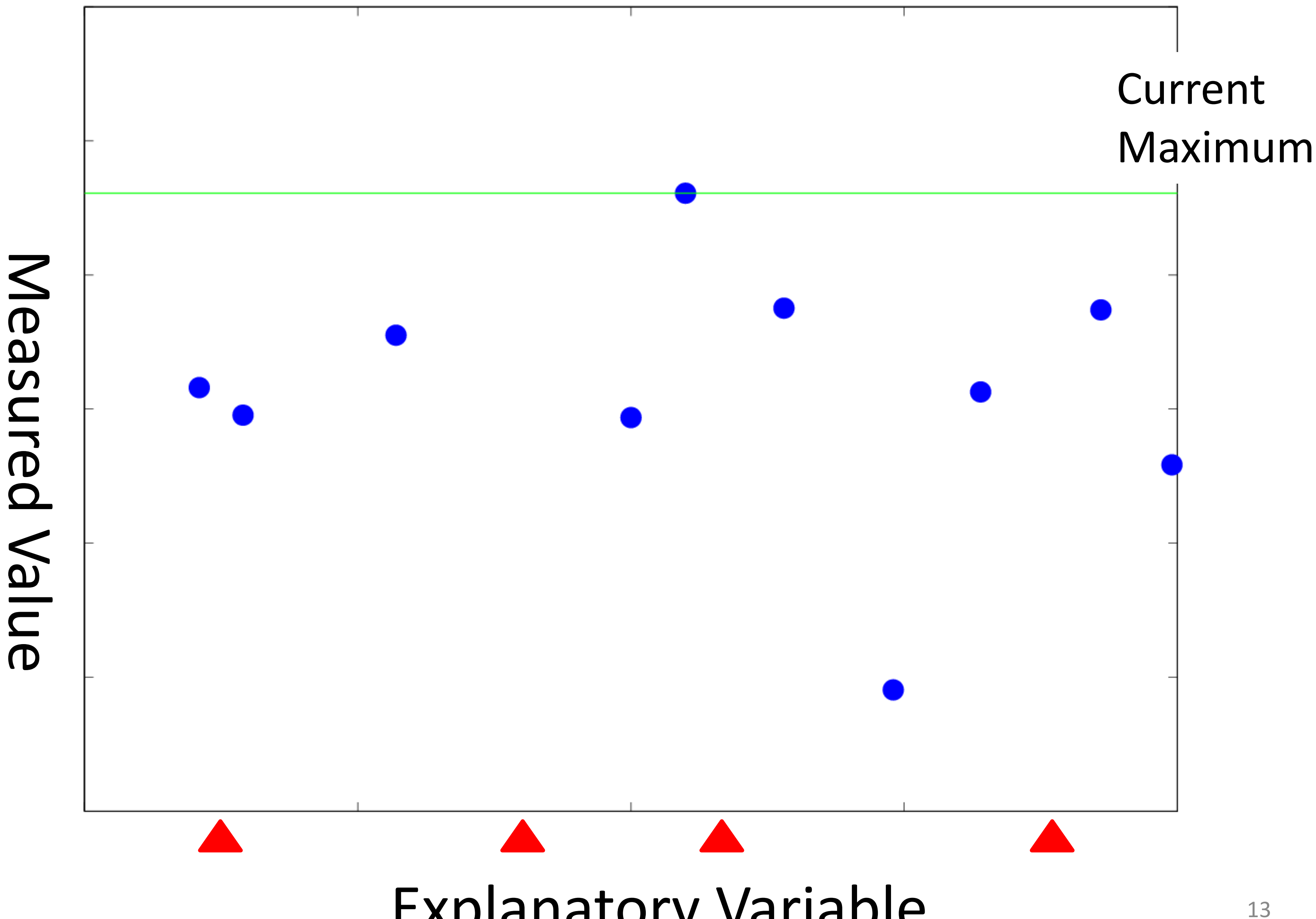
First Principles Calc.

Score 1	Score 2	Score 3	Score 8

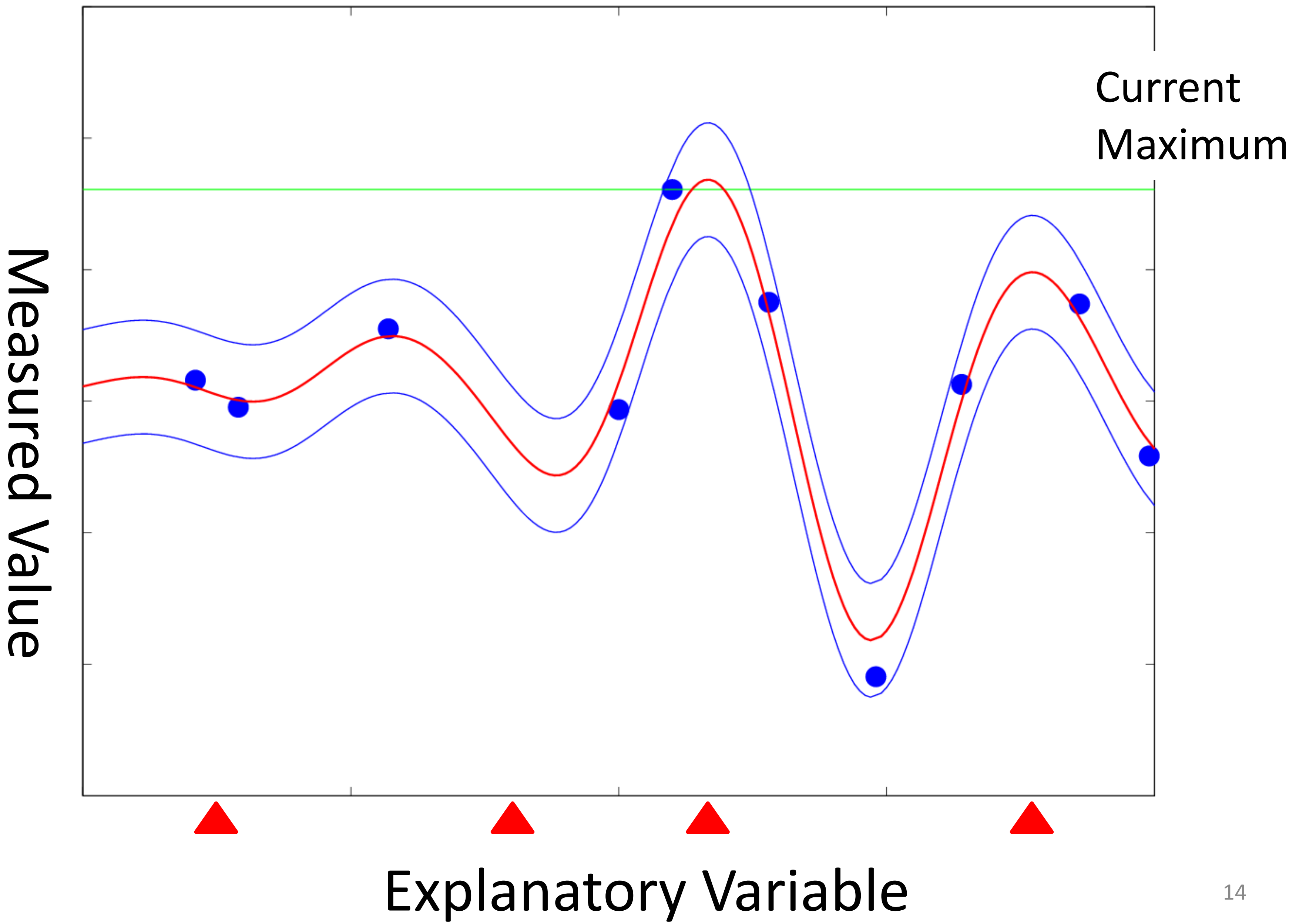
Bayesian Optimization (4)



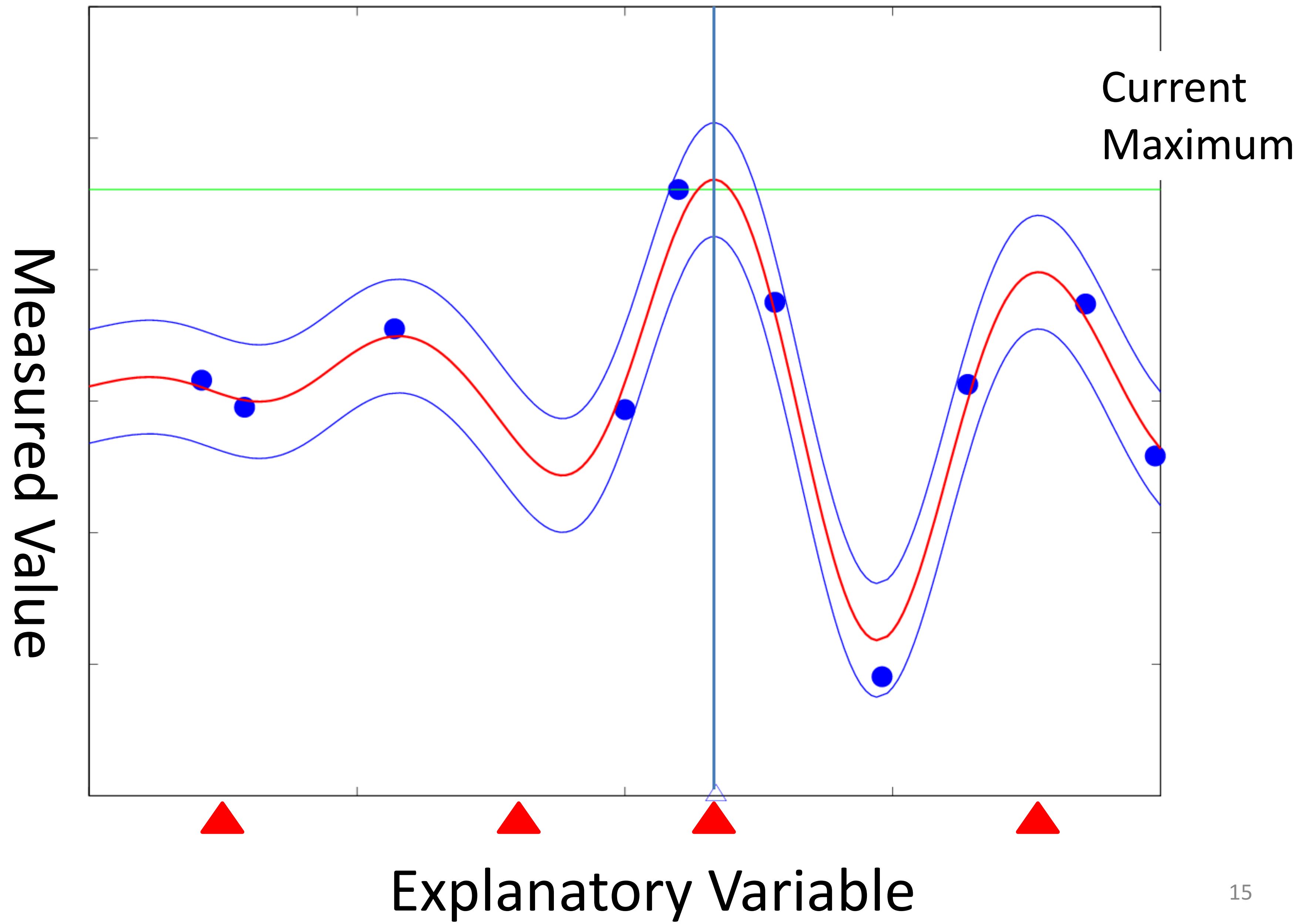
Where to observe next?



Gaussian Process



Maximum probability of improvement



Gaussian Process

Multivariate Gaussian Distribution

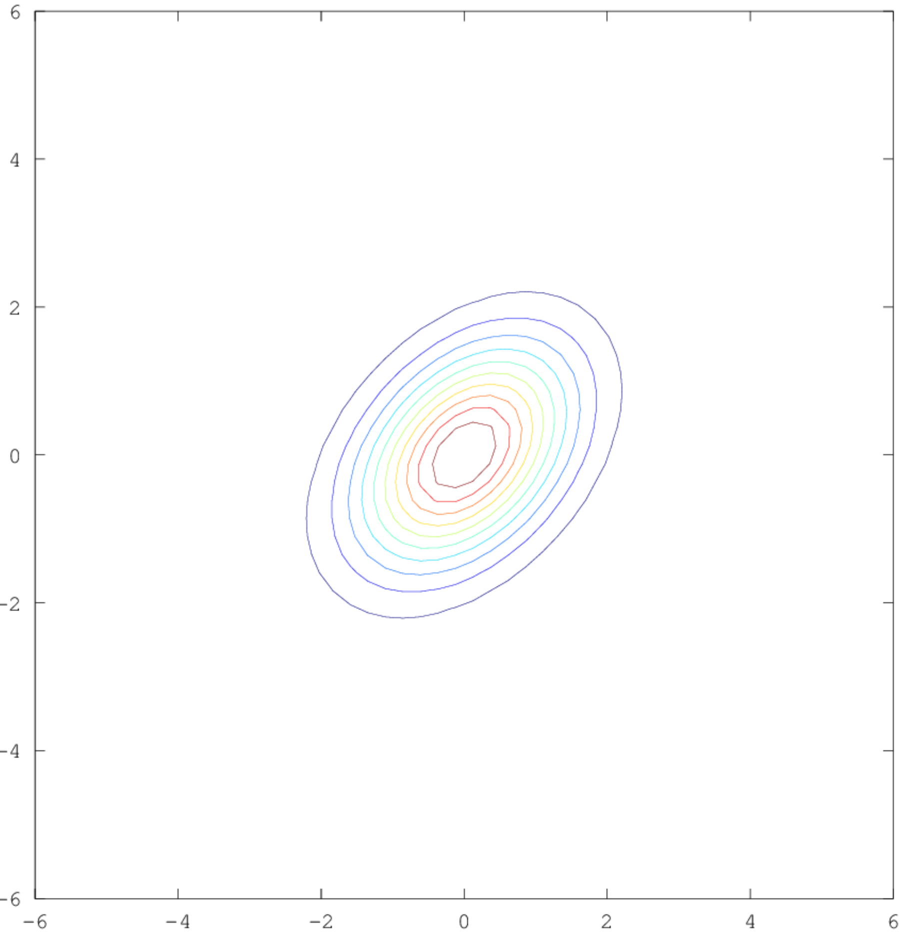
- Probability density function

$$p(\boldsymbol{x} \mid \mu, \Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp(-\frac{1}{2}(\boldsymbol{x} - \mu)^\top \Sigma^{-1}(\boldsymbol{x} - \mu))$$

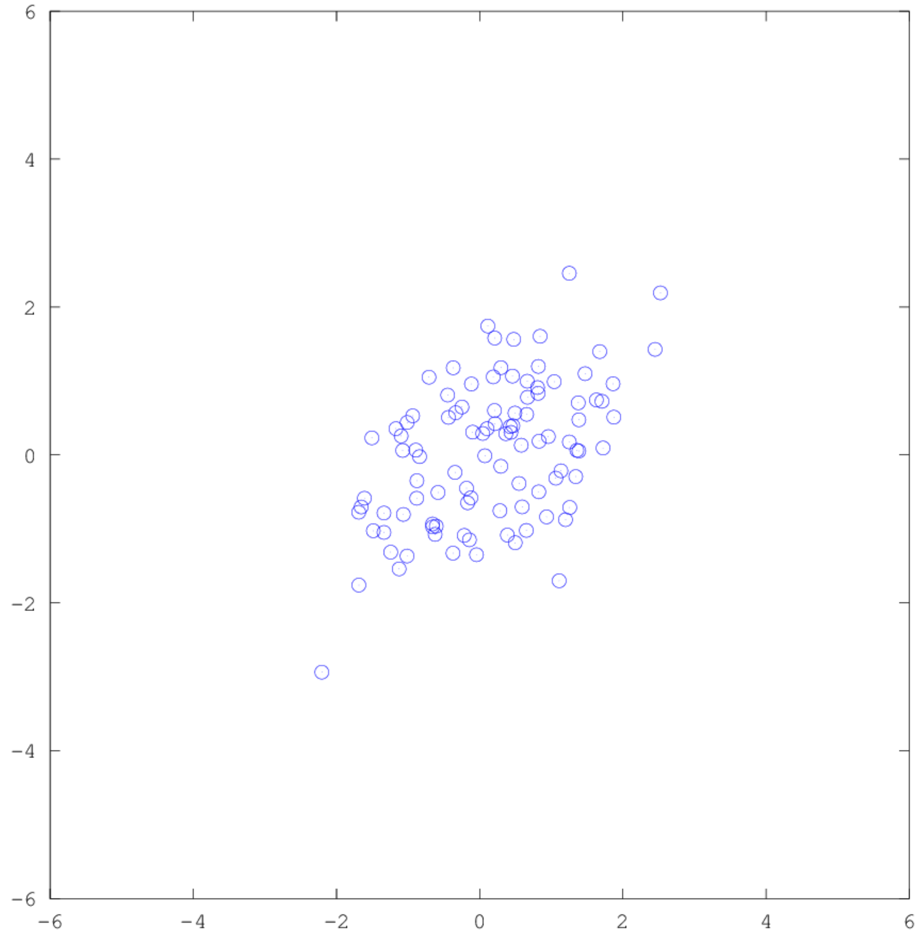
$\boldsymbol{\mu}$ Mean

Σ Covariance Matrix

Probability Density



100 Samples



$$\boldsymbol{\mu} = (0, 0)^\top \qquad \Sigma = \begin{pmatrix} 0.4 & 1 \\ 1 & 0.4 \end{pmatrix}$$

Conditional Distribution

Mean | Covariance

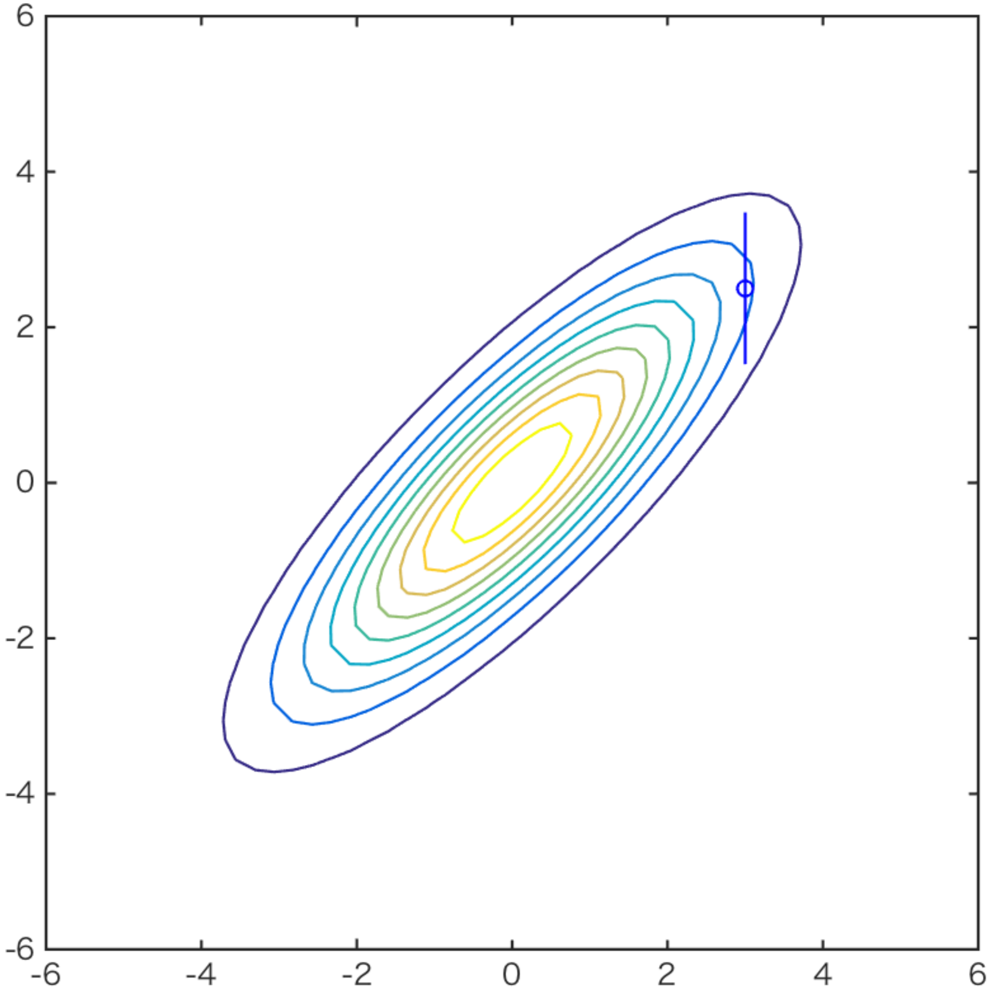
$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$$P(x_1 \mid x_2 = a) = \mathcal{N}(\mu_c, \Sigma_c)$$

$$\mu_c = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(a - \mu_2)$$

$$\Sigma_c = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

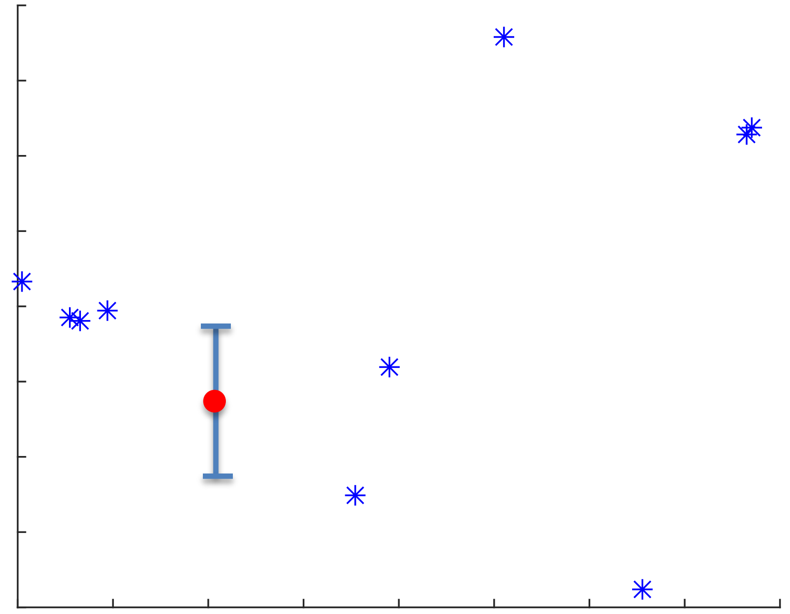
Conditional distribution at X=3



$$\Sigma = \begin{pmatrix} 3 & 2.5 \\ 2.5 & 3 \end{pmatrix}$$

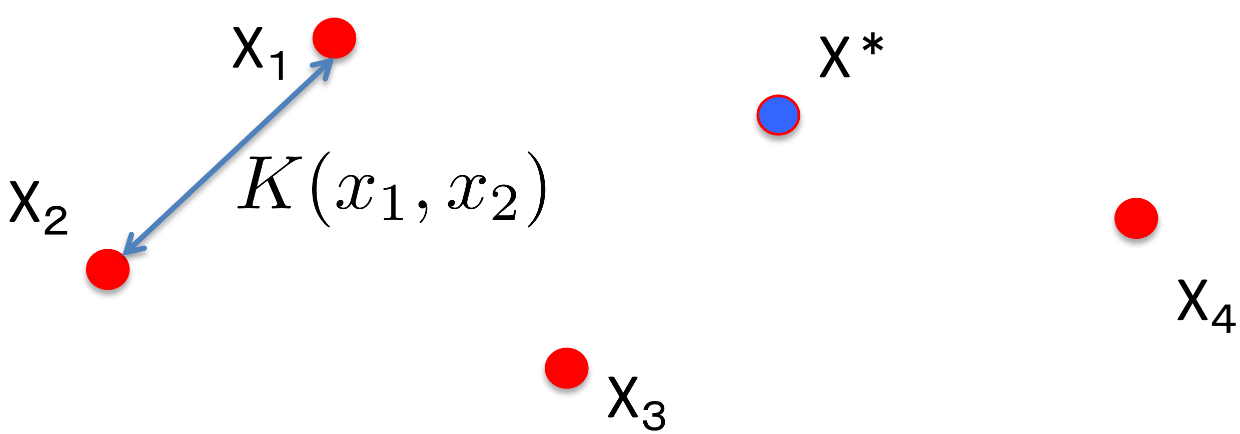
Gaussian Process

- Kernel method for regression
- Provides predictive variance in addition to regression function



Gaussian Process（No noise）

- Training points $\{x_i\}_{i=1,\ldots,n}$、Test point x^*
- Observed outcomes y_i, y^* are subject to n+1 dim Gaussian
- Mean of y_i is 0.
- Covariances are given as $K(x_i,x_j)$



Covariance matrix by Gaussian kernel

$$\begin{pmatrix} k(\mathbf{x}^*, \mathbf{x}^*) & \mathbf{k}^{*\top} \\ \mathbf{k} & K \end{pmatrix}$$

$$K(x, x') = \exp(-\|x - x'\|^2/\eta)$$

Gaussian Process （No noise）

- K：Kernel matrix for training points
- **y**: Observed outcomes for training points
- Predicted outcome at ***x****

$$E[y^*] = \mathbf{k}^{*\top} K^{-1} \mathbf{y}$$

- Predicted variance

$$V[y^*] = k(\mathbf{x}^*, \mathbf{x}^*) - \mathbf{k}^{*\top} K^{-1} \mathbf{k}^*$$

Gaussian process with noise

- Observed outcome include mean 0, variance σ^2 noise
- Covariance matrix

$$\begin{pmatrix} k(\mathbf{x}^*, \mathbf{x}^*) + \sigma^2 & \mathbf{k}^{*\top} \\ \mathbf{k} & K + \sigma^2 I \end{pmatrix}$$

Gaussian Process（with noise）

- K： Kernel matrix for training points
- **y**: Observed outcomes for training points
- Predicted outcome at ***x****

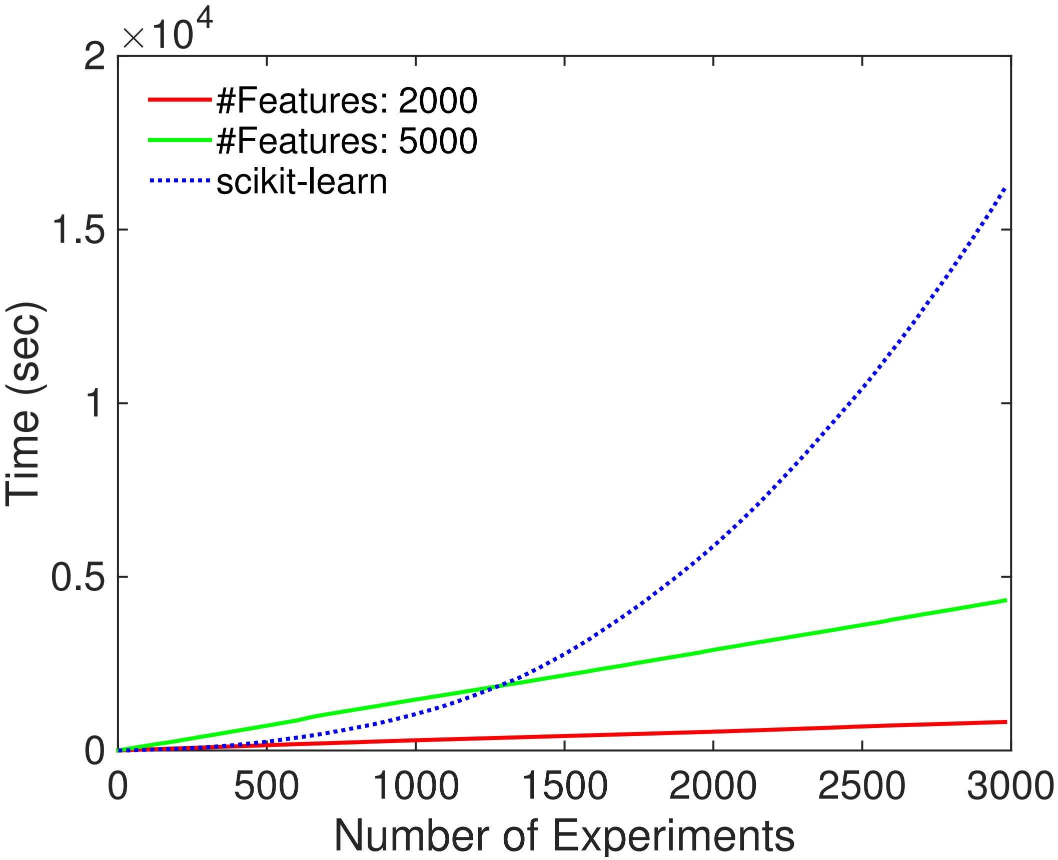
$$E[y^*] = \mathbf{k}^{*\top}(K + \sigma^2 I)^{-1}\mathbf{y}$$

- Predicted variance

$$V[y^*] = k(\mathbf{x}^*, \mathbf{x}^*) + \sigma^2 - \mathbf{k}^{*\top}(K + \sigma^2 I)^{-1}\mathbf{k}^*$$

PHYSBO (COMBO)

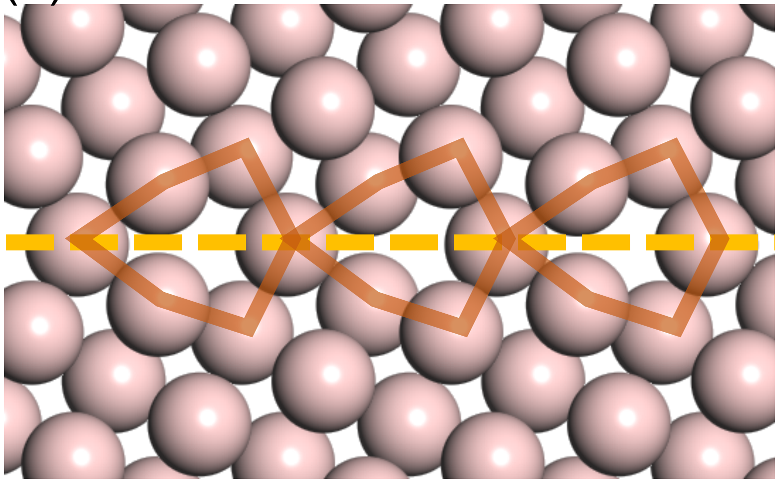
- Fast learning by random feature maps
- Automatic hyperparameter initialization & update



ベイズ最適化による粒界構造決定の高速化

Cu [001] (210) Σ5 粒界

(a)

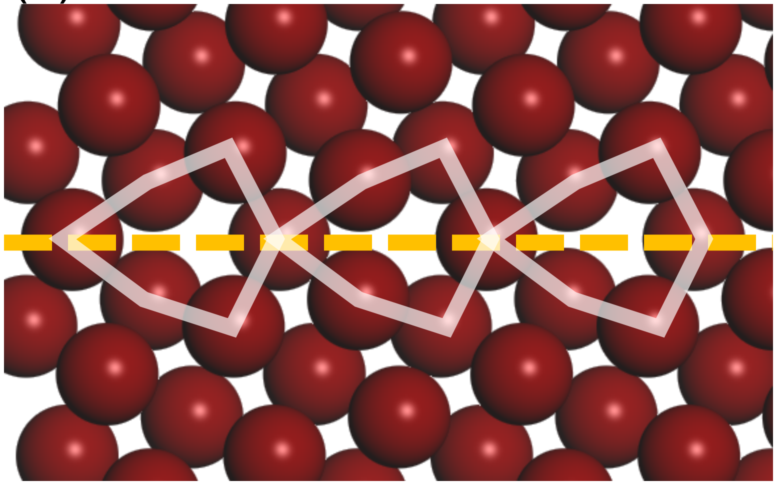


網羅的計算により決定

GB energy=0.96J/m²

計算回数＝16,983回

(b)

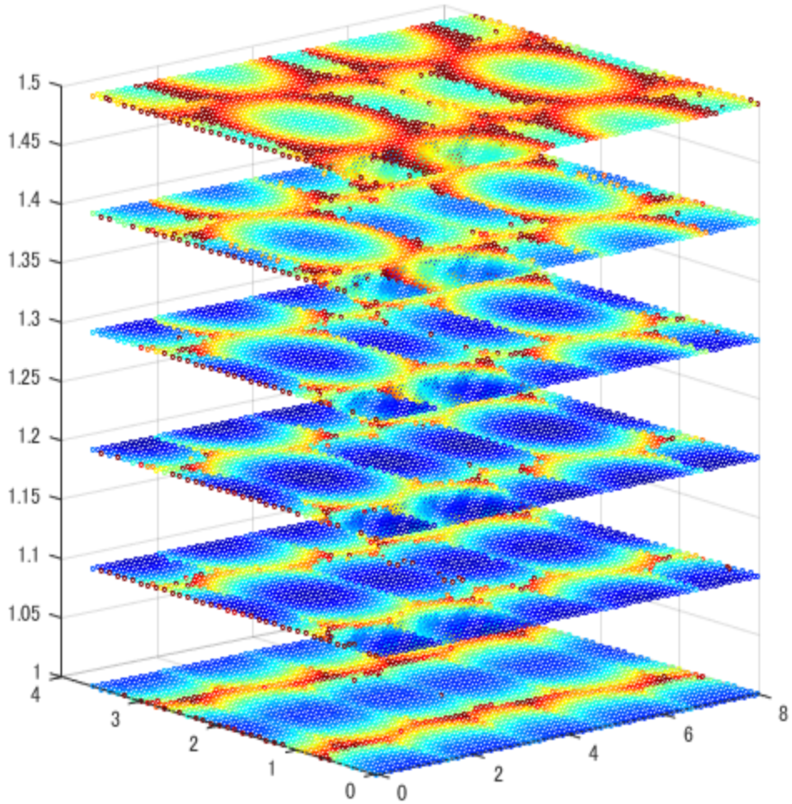


ベイズ最適化により決定

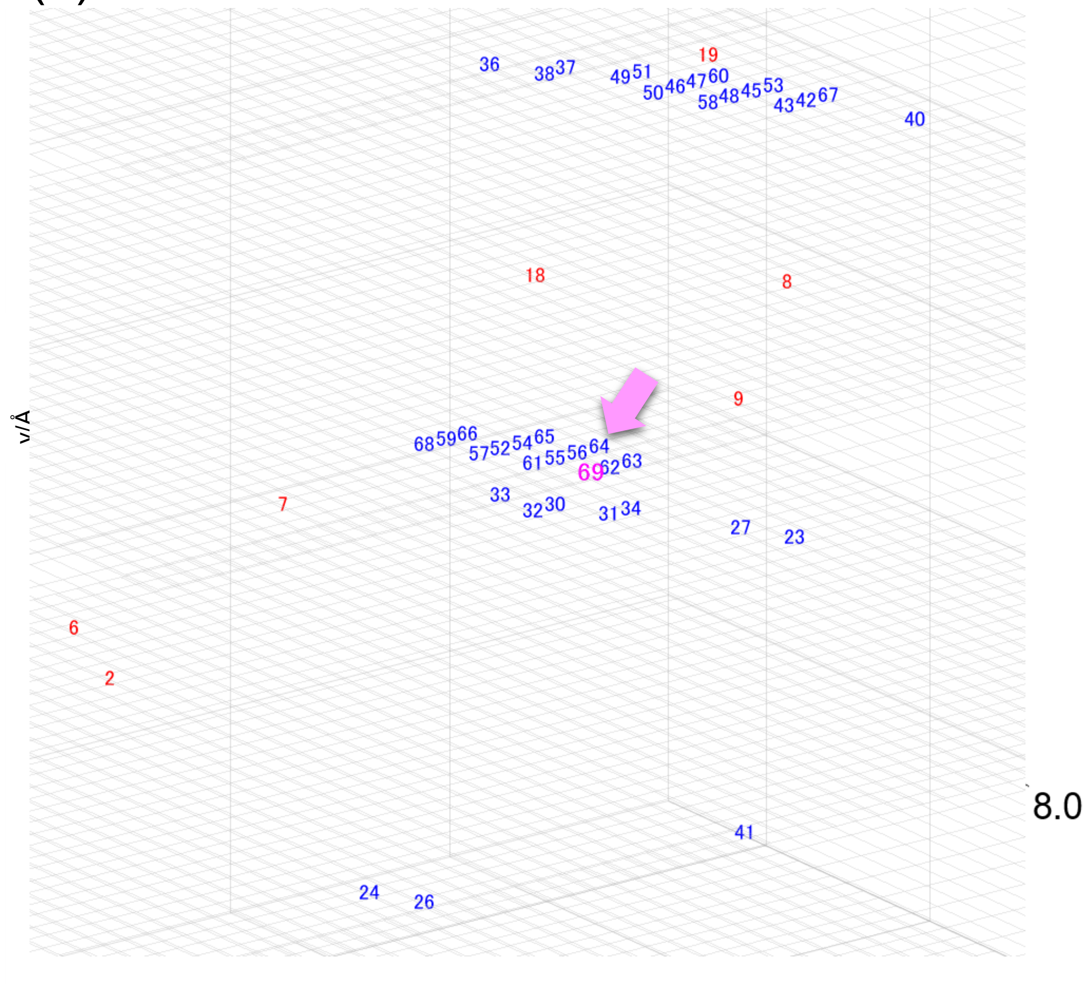
GB energy=0.96J/m²

計算回数＝69回
(20回のinitial trial含む）

東大生研　溝口准教授との共同研究　S. Kiyohara et al., Jpn. J. Appl. Phys., 2016


1.5
1.45
1.4
1.35
1.3
1.25
1.2
1.15
1.1
1.05
1
4
3
2
1
0
0
2
4
6
8

(b)


y/Å
8.0
36
38 37
49 51
19
50 46 47 60
58 48 45 53
43 42 67
40
18
8
9
68 59 66
57 52 54 65
61 55 56 64
69
62 63
33
32 30
31 34
7
27
23
6
2
41
24
26