

2022/9/9

# 第二回 シミュレーション技術と機械学習

物質・材料研究機構/東京大学 田村 亮



国立研究開発法人  
物質・材料研究機構



東京大学  
THE UNIVERSITY OF TOKYO



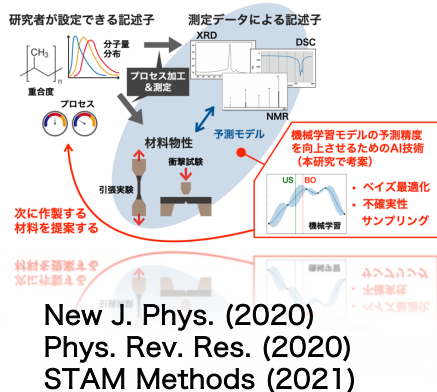
MaDIS  
NIMS MATERIALS DATA and  
INTEGRATED SYSTEM



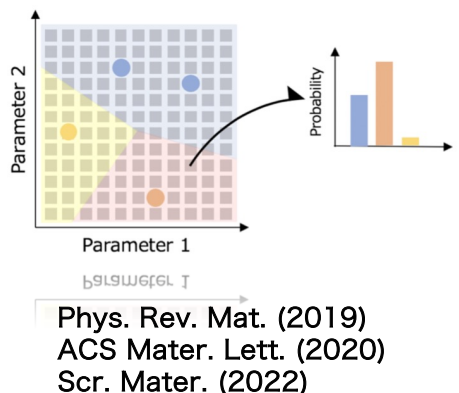
# マテリアルズ・インフォマティクス研究

機械学習の予測を利用した物質設計 レビュー論文  
Acc. Chem. Res. (2021).

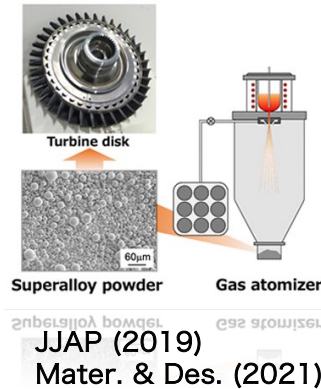
## 材料提案手法



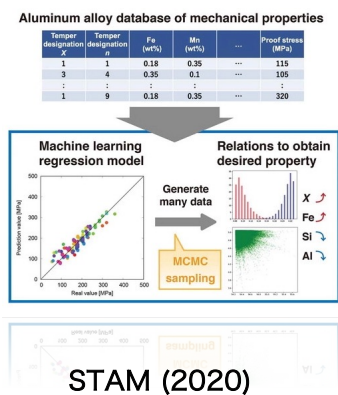
## 相図構築



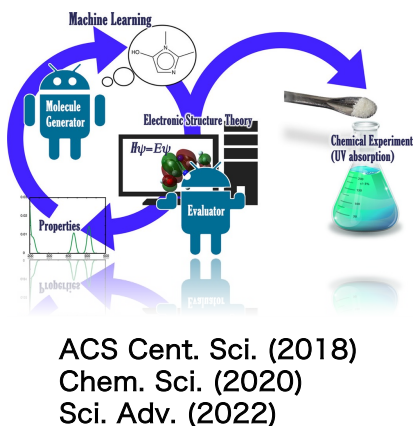
## ガスアトマイズ



## 合金設計



## 有機化合物

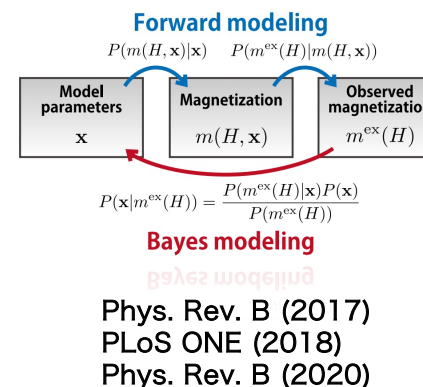


## リチウムイオン伝導

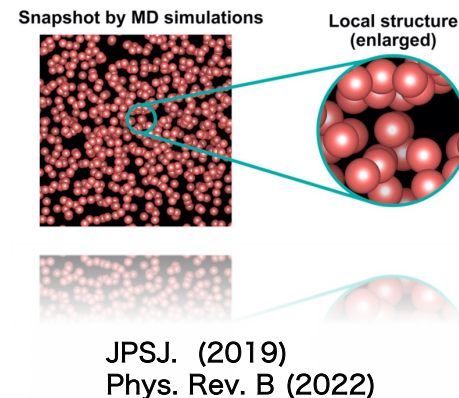


## 計算技術高度化

### ハミルトニアン推定

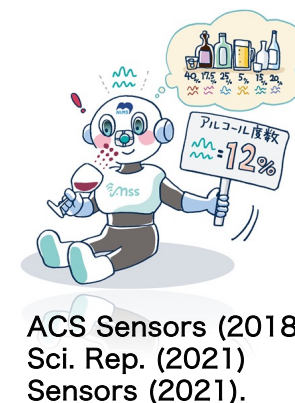


### MD解析

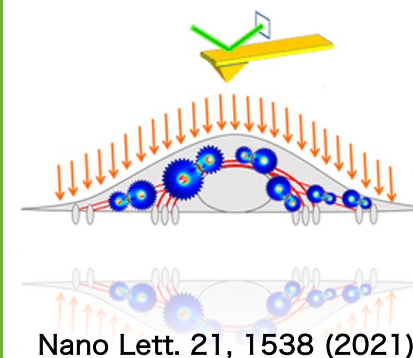


## 測定技術高度化

### 嗅覚センサー



### 原子間力顕微鏡



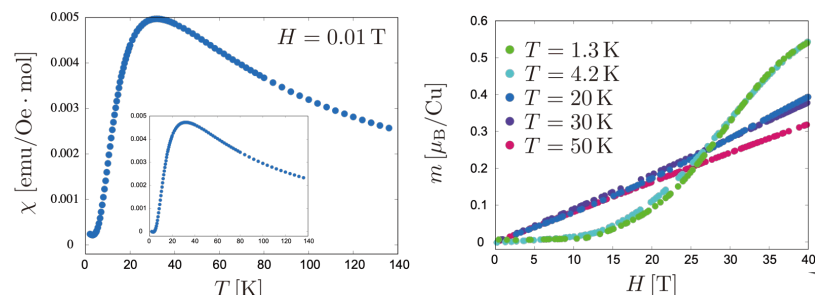
# 有効モデルをデータ駆動で推定する

# 機械学習を用いて有効モデルを推定

機械学習を利用することで、実験データを説明できる有効モデル（ハミルトニアン）を推定したい。



## 実験結果



帯磁率, 磁化, 比熱...

## 候補モデル

$$J_{ij} \mathbf{s}_i \cdot \mathbf{s}_j \quad b_{ij} (\mathbf{s}_i \cdot \mathbf{s}_j)^2 \quad \mathbf{d}_{ij} \cdot [\mathbf{s}_i \times \mathbf{s}_j] \quad D_i (s_i^z)^2$$
$$\frac{\mathbf{s}_i \cdot \mathbf{s}_j}{r_{ij}^3} - 3 \frac{(\mathbf{s}_i \cdot \mathbf{r}_{ij})(\mathbf{s}_j \cdot \mathbf{r}_{ij})}{r_{ij}^5}$$

実験結果を再現するモデルの種類・モデルパラメータを決める。

R. Tamura and K. Hukushima, PRB 95, 064407 (2017).

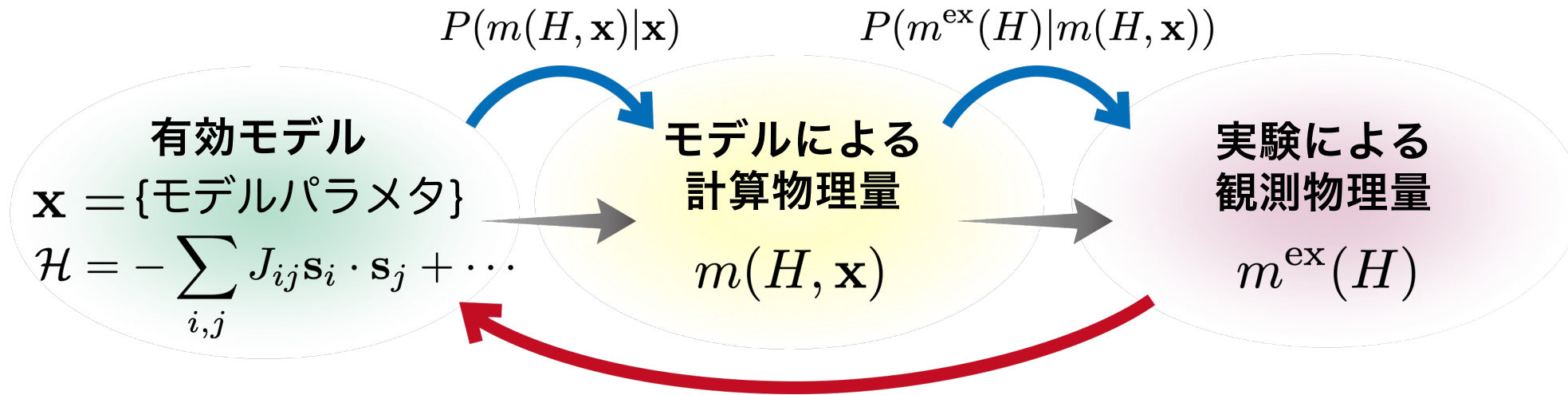
R. Tamura and K. Hukushima, PLoS ONE 13, e0193785 (2018).

R. Tamura, K. Hukushima, A. Matsuo, K. Kindo, M. Hase, PRB 101, 224435 (2020).



# どうやって有効モデルを推定するか？

## フォワードモデリング



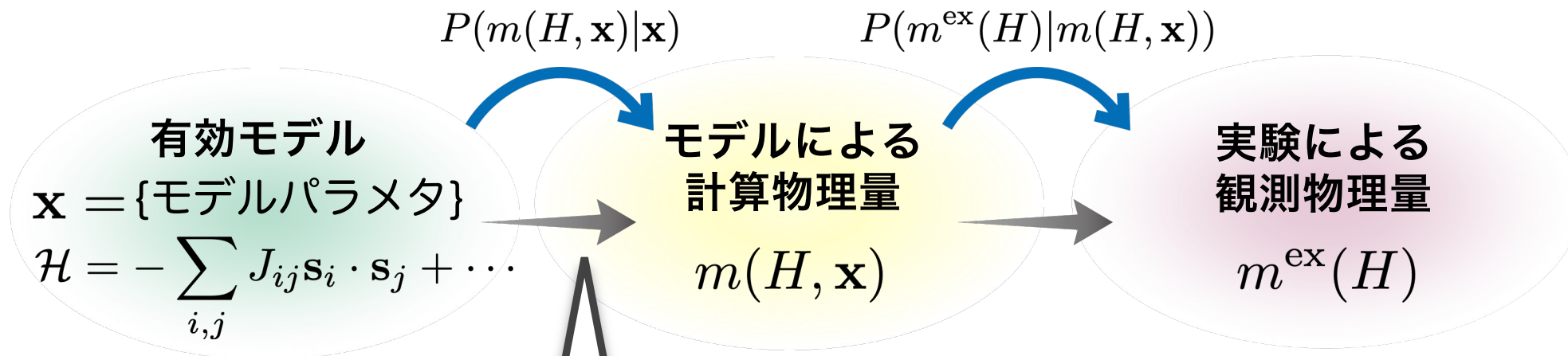
$$P(\mathbf{x}|m^{\text{ex}}(H)) = \frac{P(m^{\text{ex}}(H)|\mathbf{x})P(\mathbf{x})}{P(m^{\text{ex}}(H))} \quad \text{: ベイズの定理}$$

## ベイズモデリング

$P(B|A)$  : 事象Aが観測された後での、事象Bが観測される条件付き確率  
(事後分布)

# フォワードモデリングの第1ステップ

## フォワードモデリング



物理量は熱期待値で計算できる (例えば磁化だと...)

$$\langle \mathbf{s}_i \rangle_{H, \mathbf{x}} = \frac{\text{Tr} \mathbf{s}_i e^{-\beta \mathcal{H}}}{\text{Tr} e^{-\beta \mathcal{H}}} \longrightarrow m(H, \mathbf{x}) = \left| \frac{1}{N|\mathbf{s}|} \sum_{i=1}^N \langle \mathbf{s}_i \rangle_{H, \mathbf{x}} \right|$$

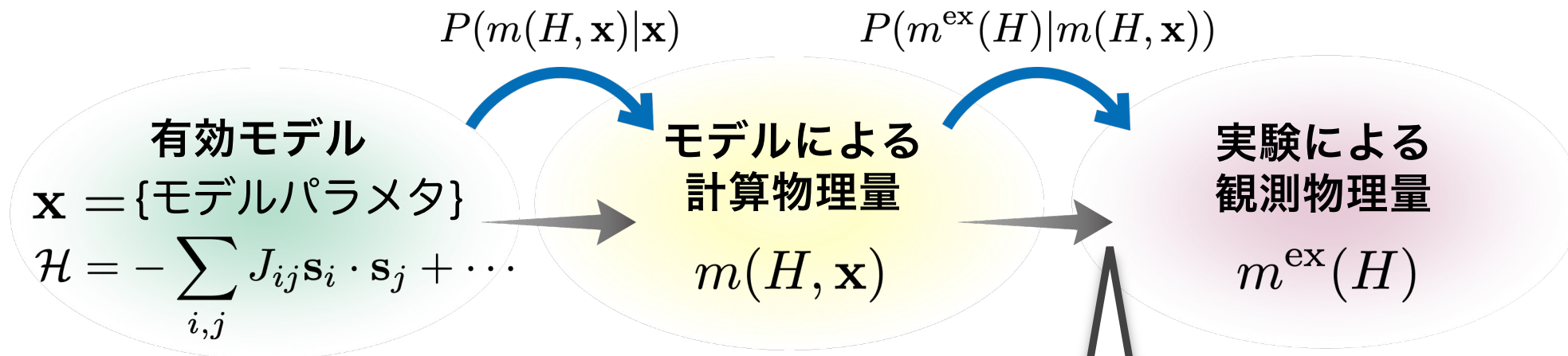
条件つき確率分布

$$P(m(H, \mathbf{x}) | \mathbf{x}) = \delta \left( m(H, \mathbf{x}) - \left| \frac{1}{N|\mathbf{s}|} \sum_{i=1}^N \langle \mathbf{s}_i \rangle_{H, \mathbf{x}} \right| \right)$$

モデルパラメタを与えると磁化が一意に決まる (数値計算だと誤差があるが...)

# フォワードモデリングの第2ステップ

## フォワードモデリング



観測磁化には，計算磁化に観測ノイズがのる

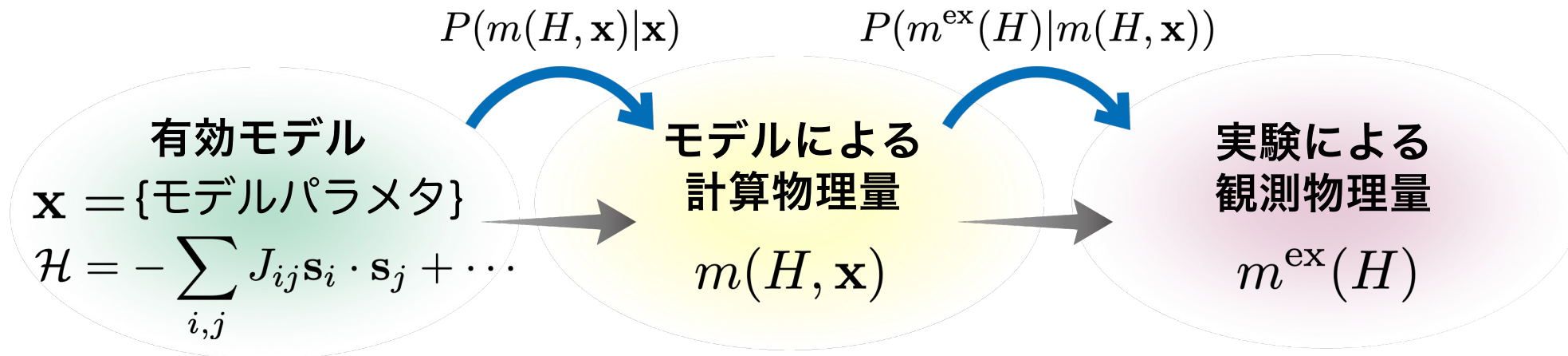
$$m^{\text{ex}}(H) = m(H, \mathbf{x}) + \underbrace{\varepsilon}_{\text{観測ノイズ}} \quad P(\varepsilon) \propto \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right) : \text{観測ノイズはガウス分布 (仮定)}$$

条件つき確率分布

$$P(m^{\text{ex}}(H)|m(H, \mathbf{x})) \propto \exp\left(-\frac{1}{2\sigma^2}(m^{\text{ex}}(H) - m(H, \mathbf{x}))^2\right)$$

# フォワードモデリングの条件付き確率

## フォワードモデリング



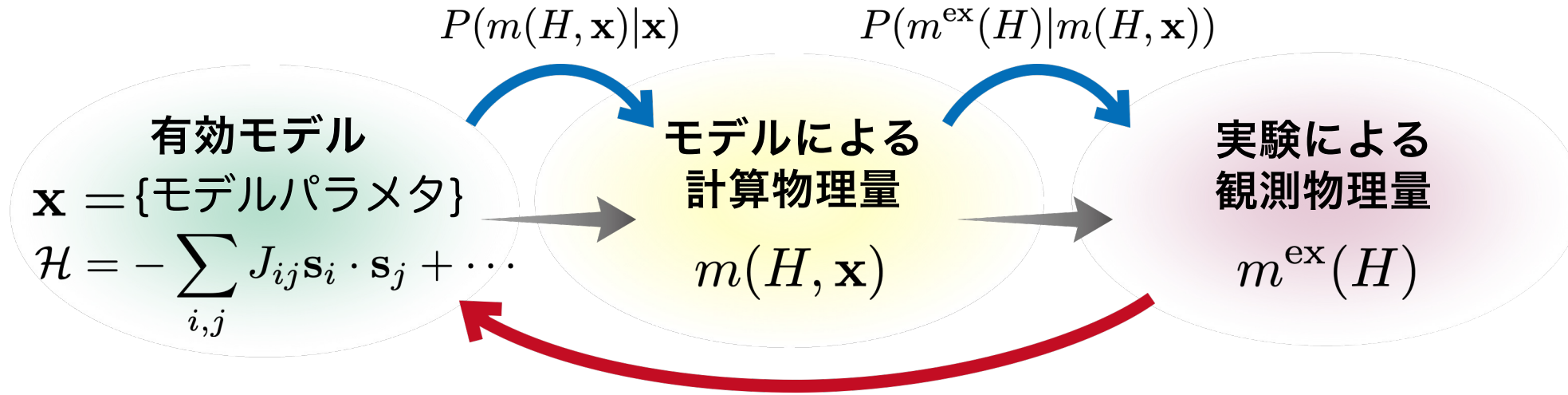
## フォワードモデリングによる条件つき確率分布

$$P(m^{\text{ex}}(H)|\mathbf{x}) \propto \int dm(H, \mathbf{x}) P(m^{\text{ex}}(H)|m(H, \mathbf{x})) P(m(H, \mathbf{x})|\mathbf{x})$$
$$\propto \exp \left[ -\frac{1}{2\sigma^2} \left( m^{\text{ex}}(H) - \left| \frac{1}{N|\mathbf{s}|} \sum_{i=1}^N \langle \mathbf{s}_i \rangle_{H, \mathbf{x}} \right| \right)^2 \right]$$

モデルパラメータを入力した際に、条件つき確率が最大となる物理量が、実験により観測される物理量である。

# ベイズモデリング

## フォワードモデリング



$$P(\mathbf{x}|m^{\text{ex}}(H)) = \frac{P(m^{\text{ex}}(H)|\mathbf{x})P(\mathbf{x})}{P(m^{\text{ex}}(H))} : \text{事後確率分布}$$

## ベイズモデリング

観測物理量を入力した際に、条件つき確率が最大となる有効モデルが、実験を説明できる有効モデルである。



# ベイズモデリングの定式化

実験による測定結果を入力し，事後確率分布を最大（MAP推定）とする  
ハミルトニアンを求める。

事後確率分布（最大化したい）

$$P(\mathbf{x} | \{y^{\text{ex}}(g_l)\}_{l=1, \dots, L}) = \exp[-E(\mathbf{x})]$$

推定したいモデルパラメタ 測定物理量(実験により測定)

エネルギー関数（最小化したい）

$$E(\mathbf{x}) = \frac{1}{2\sigma^2} \sum_{l=1}^L \left[ \underbrace{y^{\text{ex}}(g_l)}_{\substack{\text{測定物理量} \\ \text{(実験により測定)}}} - \underbrace{y^{\text{cal}}(g_l, \mathbf{x})}_{\substack{\text{計算物理量} \\ \text{(モデルから計算)}}} \right]^2 - \log \underbrace{P(\mathbf{x})}_{\substack{\text{事前確率分布} \\ \text{(事前知識)}}$$

R. Tamura and K. Hukushima, Phys. Rev. B 95, 064407 (2017).

# 事前分布の例

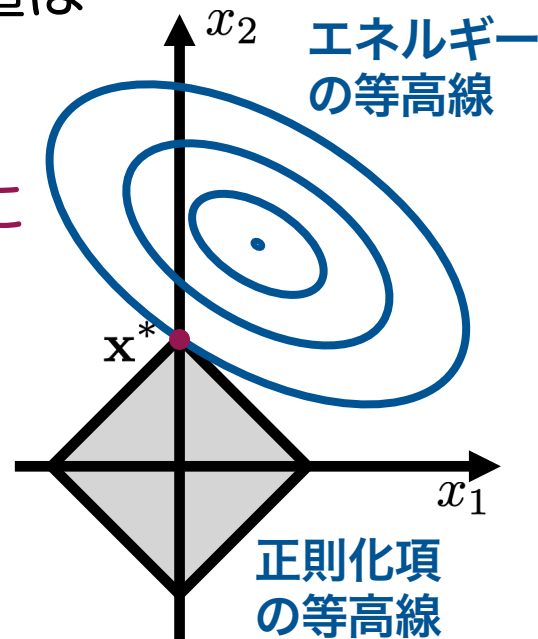
事前確率分布は、モデルパラメタに対する事前知識を与えるため、事前知識が導入できる形式ならどのようなものでも使える。

- L1正則化項

$$P(\mathbf{x}) \propto \exp(-\lambda|\mathbf{x}|)$$

モデルパラメタの値は  
小さい方が良い

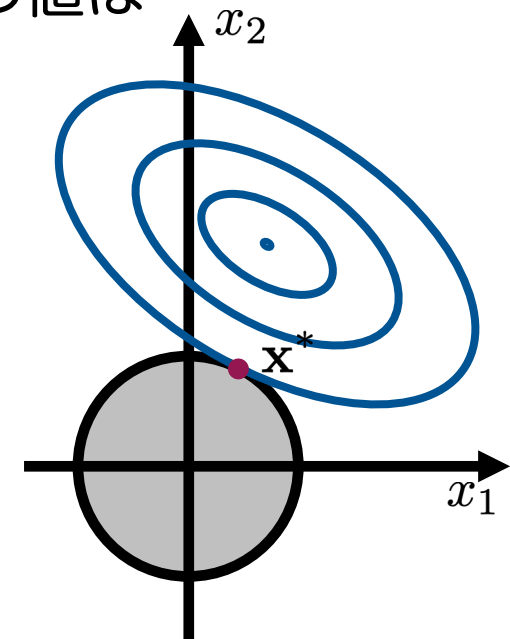
どれかの値がゼロに  
なりやすい



- L2正則化項

$$P(\mathbf{x}) \propto \exp(-\lambda\|\mathbf{x}\|^2)$$

モデルパラメタの値は  
小さい方が良い



# MAP推定に計算時間が膨大になる場合があります…

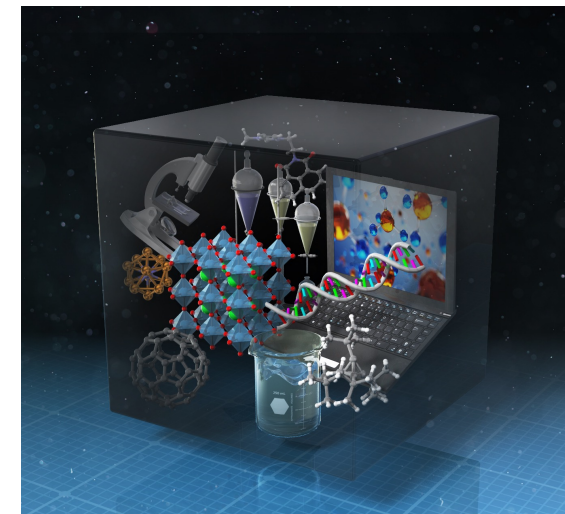
エネルギー関数 (最小化したい)

$$E(\mathbf{x}) = \frac{1}{2\sigma^2} \sum_{l=1}^L \left[ \underbrace{y^{\text{ex}}(g_l)}_{\substack{\text{測定物理量} \\ \text{(実験により測定)}}} - \underbrace{y^{\text{cal}}(g_l, \mathbf{x})}_{\substack{\text{計算物理量} \\ \text{(モデルから計算)}}} \right]^2 - \log \underbrace{P(\mathbf{x})}_{\substack{\text{事前確率分布} \\ \text{(事前知識)}}$$

- モデルパラメタを更新しながら，エネルギー関数が最小となるモデルを探す。
- モデルパラメタを更新すると，物理量の計算が必要となる。
- 物理量の計算が時間がかかる場合，この最適化には長時間必要…
- ベイズ最適化を使ってみたら効率よく最適化できるか？

# ベイズ最適化とは？

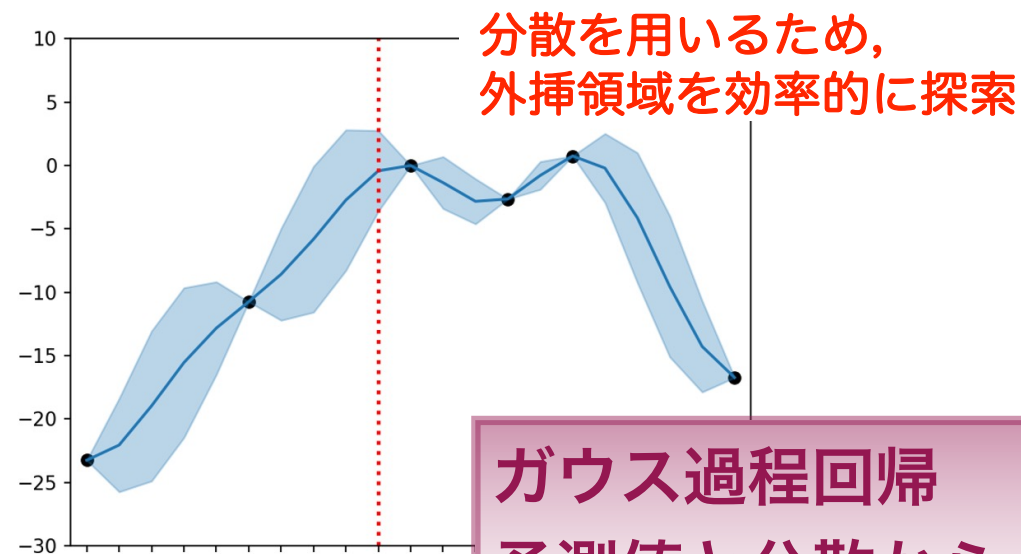
$y = f(\mathbf{x})$   $f$  がわからない場合に有用な最適化手法  
目的変数 (エネルギー)      説明変数 (モデルパラメタ)      → ベイズ最適化



- N個の候補点があり，この中から最大/最小の目的変数を持つものを探したい。

$\mathbf{x}_i$  ( $i = 1, 2, \dots, N$ )

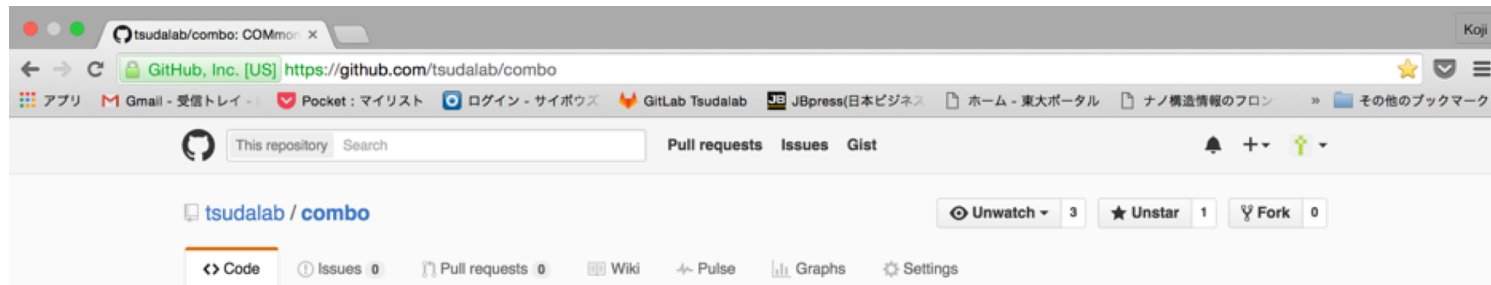
- できるだけ計算回数を少なくしたい。
- M個の候補点に対する計算が終わった。
- 次のM+1個目の候補点を最適に選びたい。
- M個のサンプル点から予測モデルを学習し，それを用いて，残りの候補点をスコアリングし実際に計算するサンプルを選ぶ。



ガウス過程回帰  
予測値と分散から，  
次の候補を選ぶ

# ベイズ最適化パッケージCOMBO

ベイズ最適化では、予測モデルとしてガウス過程を利用



COMmon Bayesian Optimization — Edit

25 commits 2 branches 0 releases 2 contributors

Branch: master New pull request New file Find file HTTPS https://github.com/tsudalab/combo Download ZIP

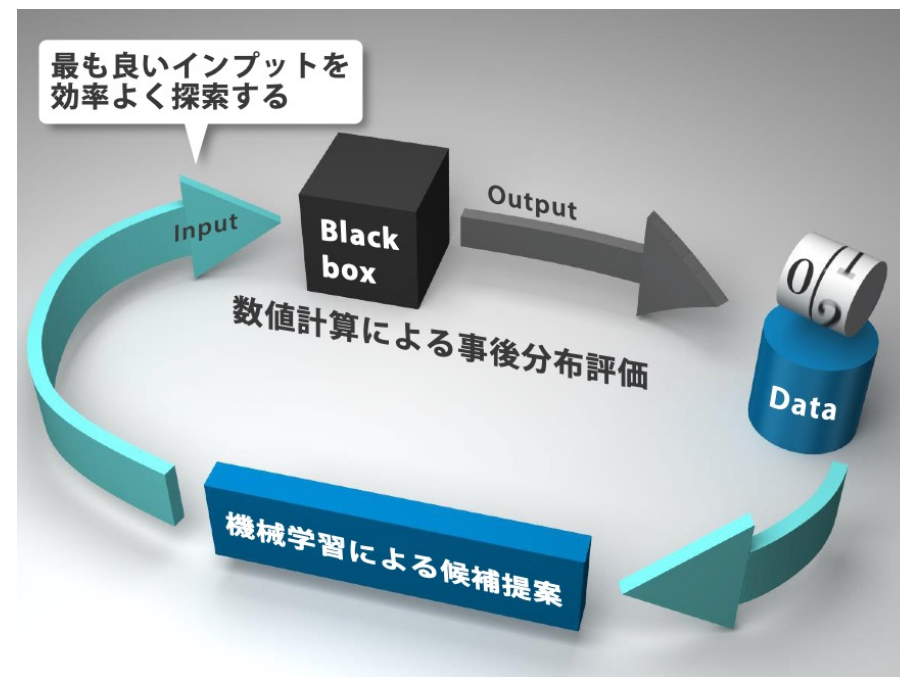
File	Commit	Time
kojitsuda README	Latest commit c9f5e44	6 hours ago
combo	update combo to version 0.1.1	3 days ago
docs	add document	8 hours ago
examples/grain_bound	modify README	9 hours ago
.gitignore	add .gitignore	23 days ago
README.md	README	6 hours ago
setup.py	combo version 0.1.1	3 days ago

COMmon Bayesian Optimization Library ( COMBO )

<https://github.com/tsudalab/combo>

✓ ハイパーパラメタの学習を自動で実行

✓ トレーニングデータに対して線形計算可能







[Top](#) [PHYSBOについて](#) [インストール](#) [ドキュメント](#) [ニュース](#) [お問合せ](#)

## PHYSBO

- ver. 1.0-
  - 田村 亮 (物質・材料研究機構 国際ナノアーキテクトニクス研究拠点)
  - 寺山 慧 (横浜市立大学大学院 生命医科学研究科)
  - 津田 宏治 (東京大学大学院 新領域創成科学研究科)
  - 植野 剛 (東京大学大学院 新領域創成科学研究科)
  - 本山 裕一 (東京大学 物性研究所)
  - 吉見 一慶 (東京大学 物性研究所)
  - 川島 直輝 (東京大学 物性研究所)

# PASUMS

Project for advancement of  
software usability in materials science

## PASUMS

東京大学物性研究所

ソフトウェア開発・高度化プロジェクト

<https://www.pasums.issp.u-tokyo.ac.jp/physbo/>

- **多目的最適化**
- **Python3対応**
- **マニュアル完備**
- **スパコンを用いた並列計算**
- **ライセンスがGPL**

# 量子ハイゼンベルクモデルのパラメタ推定

フォワード計算は厳密対角化で実行  
(事前確率分布は定数と仮定)

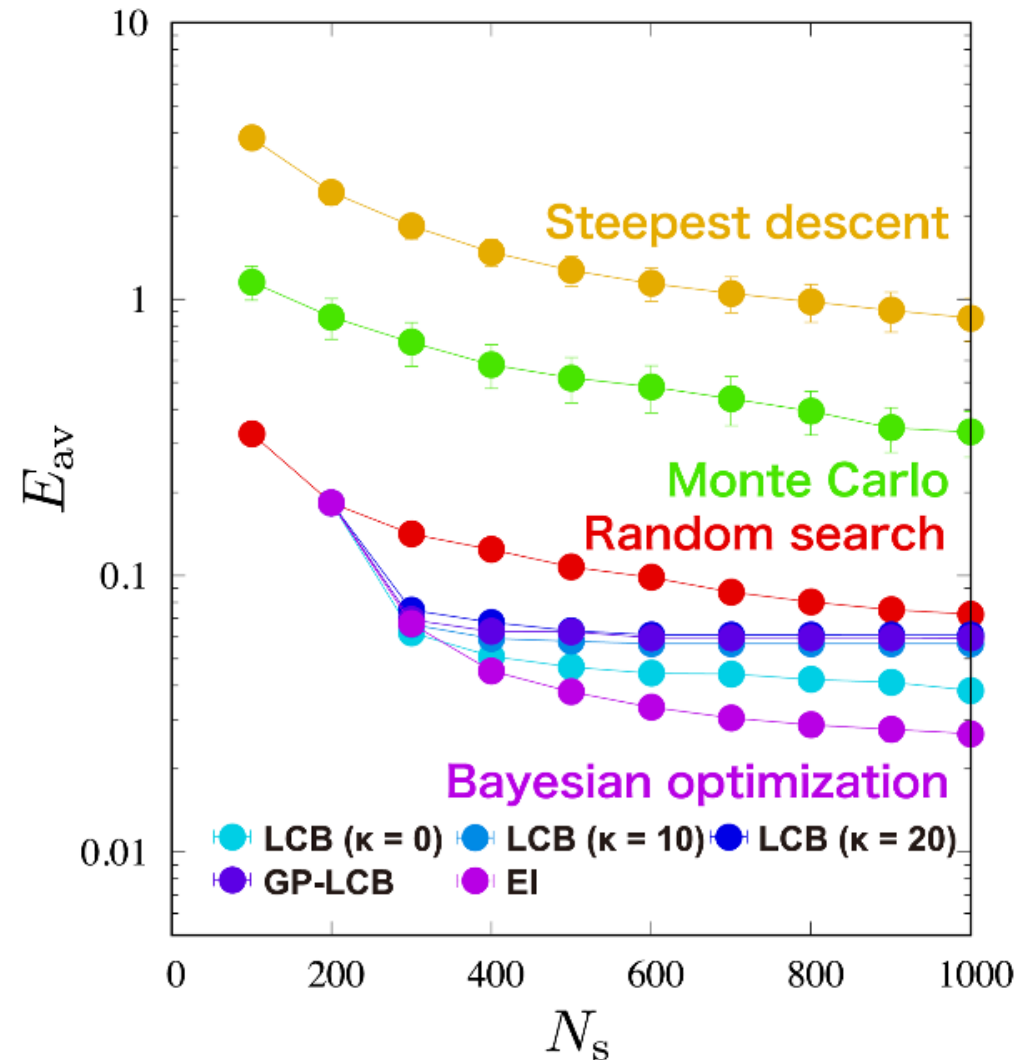
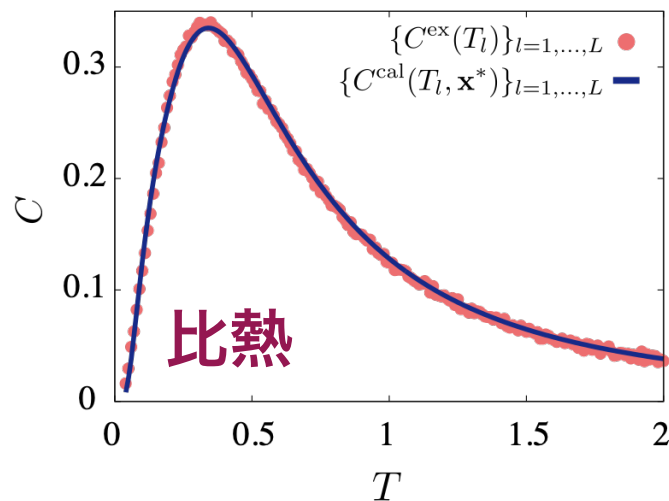
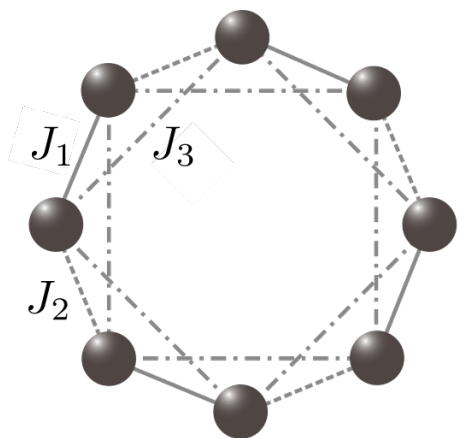
R. Tamura and K. Hukushima, PLoS ONE 13, e0193785 (2018).

$$E(\mathbf{x}) = \frac{1}{2\sigma^2} \sum_{l=1}^L [y^{\text{ex}}(g_l) - y^{\text{cal}}(g_l, \mathbf{x})]^2$$

を最小化するパラメタをベイズ最適化で探索

対象ハミルトニアン：

$$\mathcal{H} = - \sum_{i,j} J_{ij} [s_i^x s_j^x + s_i^y s_j^y + \Delta s_i^z s_j^z] - H \sum_i s_i^z$$



# HΦを用いたハミルトニアン推定

アプリ掲載数 269 件

お問合せ / アプリ掲載依頼

JP / EN

MateriApps  
物質科学シミュレーションのポータルサイト

MateriApps について レビュー募集

Google カスタム検索

もっと詳

NEWS / 講習会・イベント アプリ一覧 アプリ詳細検索 キーワード解説 レビュー 事例 アプリコンシェルジュ

公開度: 3 ★★★★★ ドキュメント充実度: 1 ★☆☆

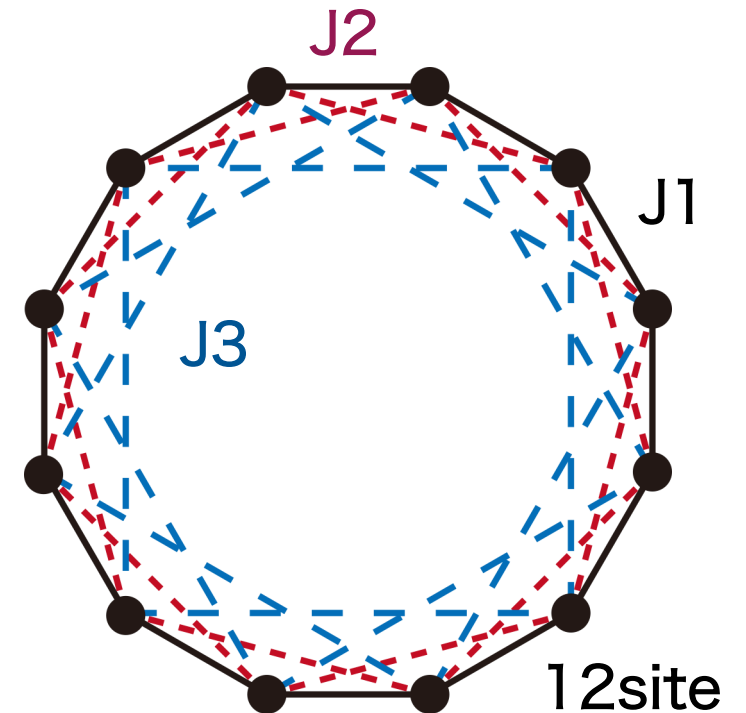
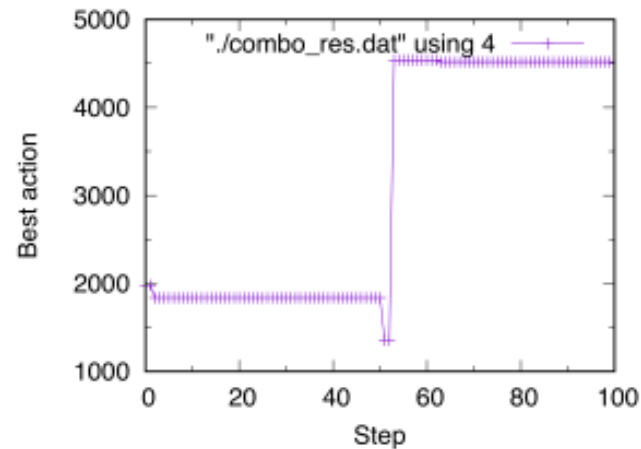
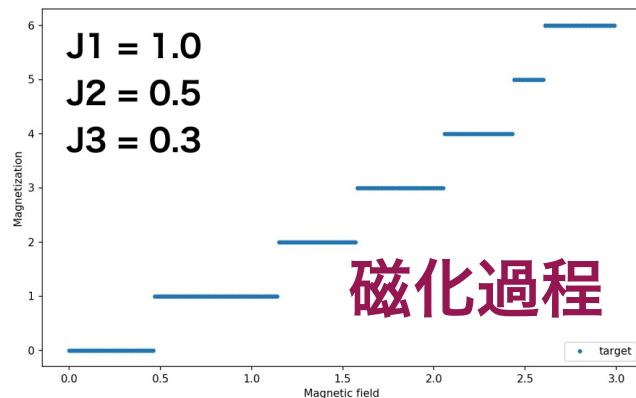
機械学習で使われるベイズ最適化のPythonライブラリ。データ数に対して線形に計算コストが増大するので、大きな特徴空間でベイズ最適化を行うことが可能。ハイパーパラメータは第二種最尤推定に基づいてデータから自動的に学習される。

04 / 06

01 基本情報

量子格子模型ソルバーHΦとベイズ最適化ライブラリCOMBOを組み合わせたモデル推定事例の紹介

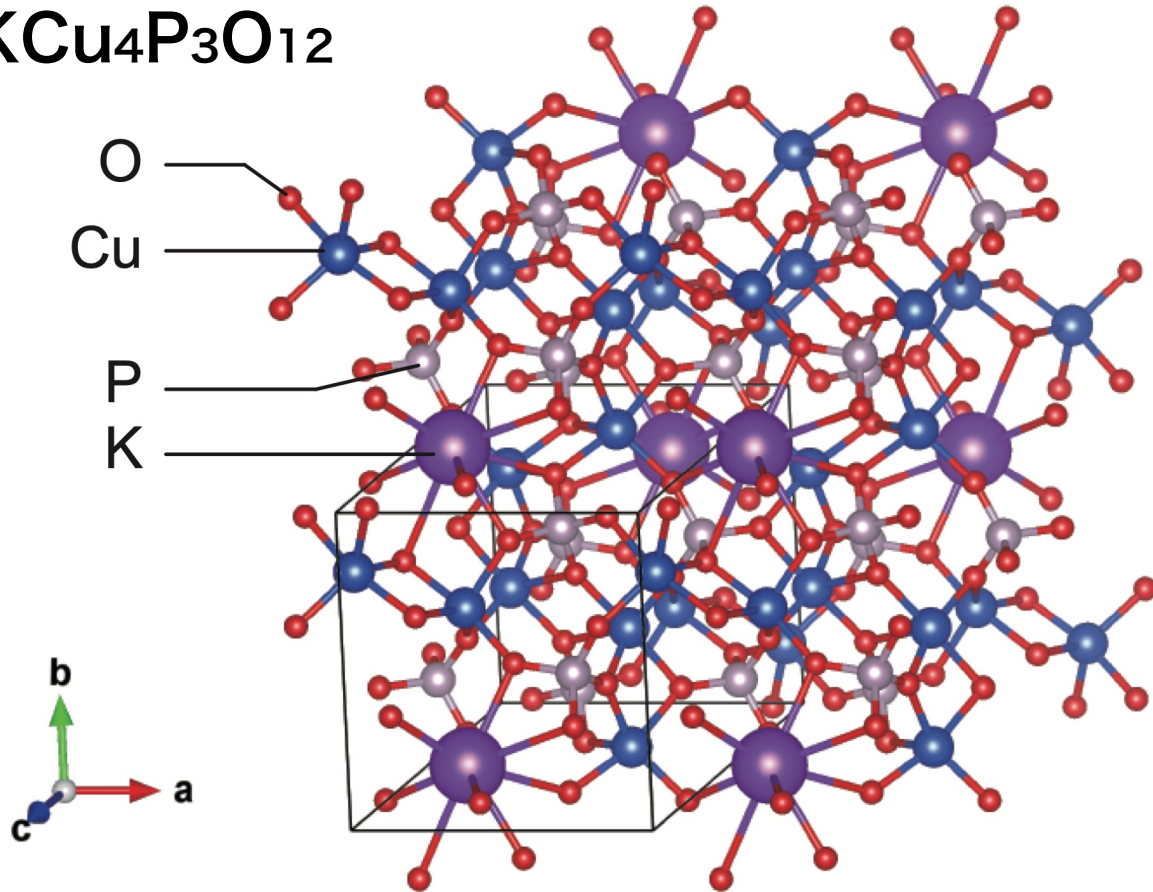
$$H = \sum_{i=1}^{12} J_1 S_i \cdot S_{i+1} + J_2 S_i \cdot S_{i+2} + J_3 S_i \cdot S_{i+3}$$



物性研  
吉見一慶さんと  
共同開発

# 実験データに対するデモンストレーション

$\text{KCu}_4\text{P}_3\text{O}_{12}$



Cuイオン:  $S=1/2$ ハイゼンベルク  
(異方性なし)

g-因子の値: 2.08 (ESR結果)

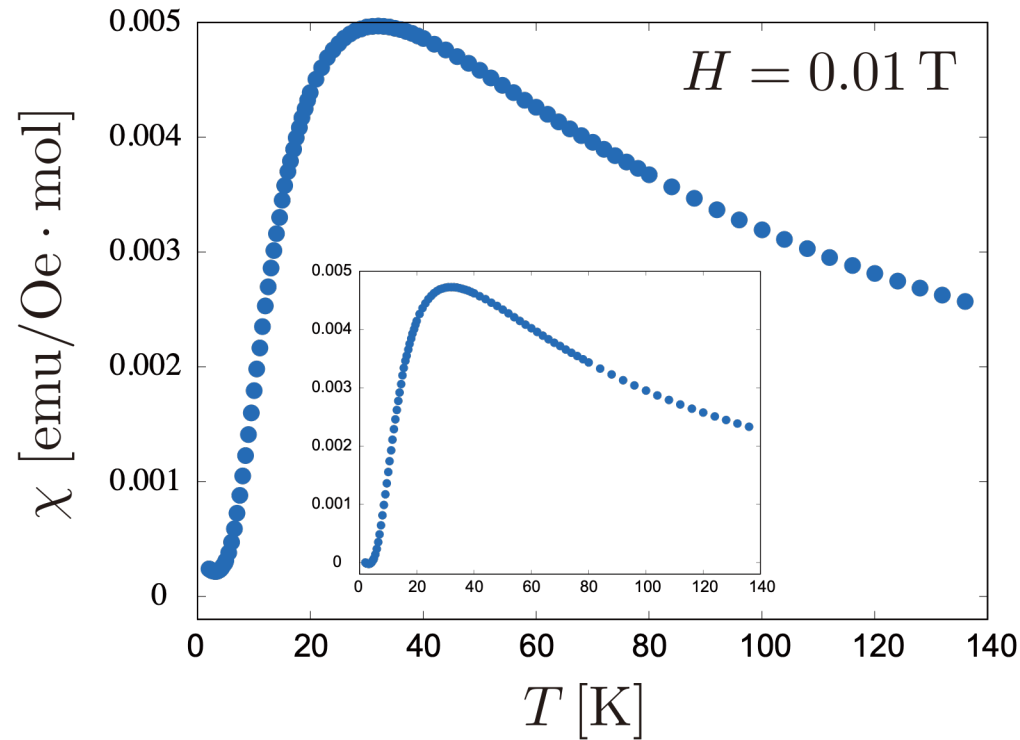
磁性に関する報告はなし

$\text{KCu}_4(\text{PO}_4)_3$ : A compound with two trigonal bipyramidal  $\text{Cu}^{2+}\text{O}_5$  coordination polyhedral, Effenberger H.S., Z. Kristallogr. 180, 43 (1987).

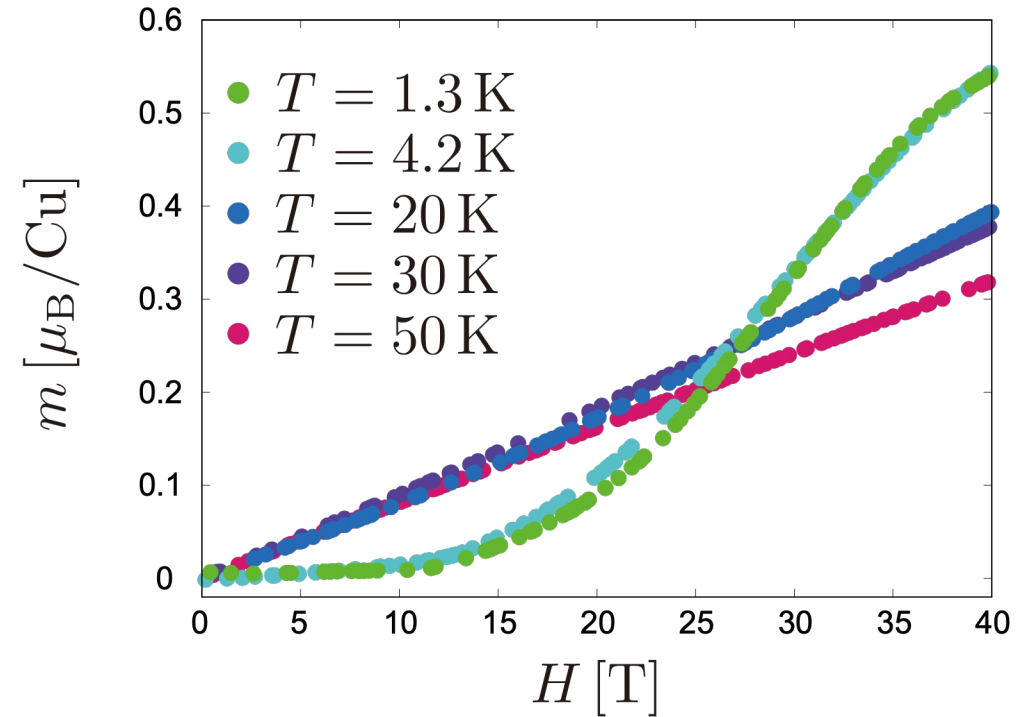


# 磁性に関する測定結果

## 帯磁率



## 磁化



東大物性研究所 国際超強磁場科学研究施設  
によって測定



# ターゲット有効モデル

## 8スピンからなるチェーン上の量子ハイゼンベルクモデル

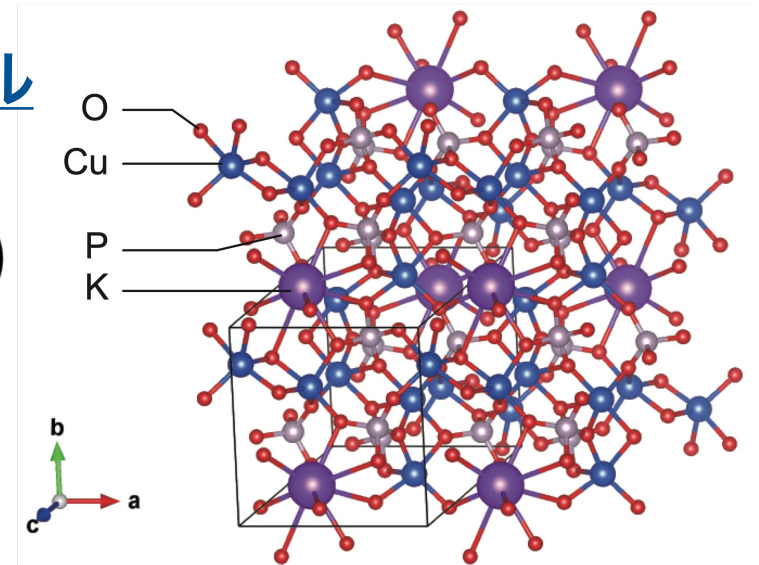
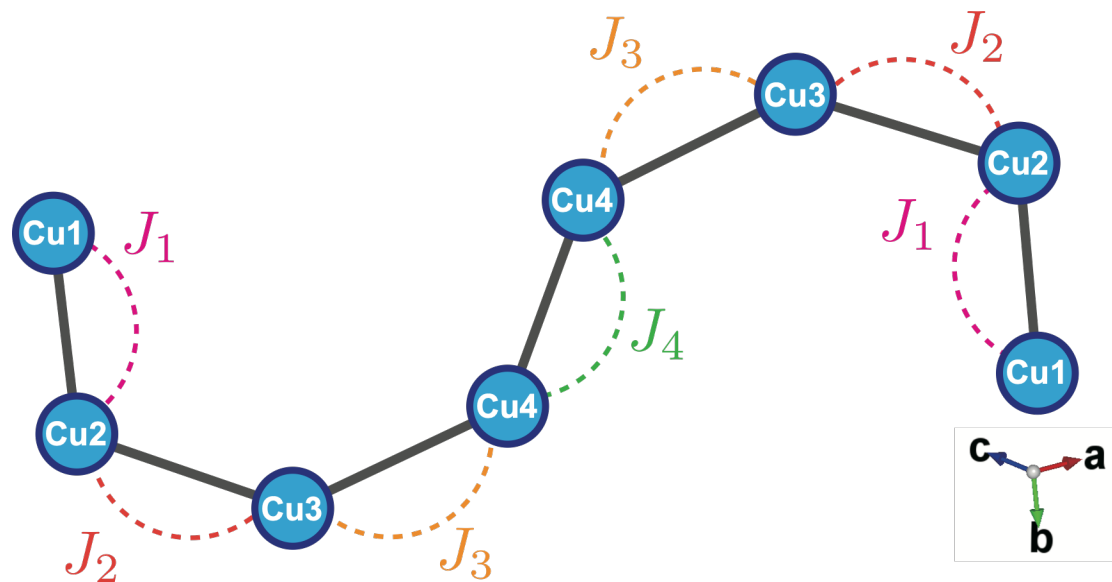
$$\mathcal{H}(\mathbf{x}) = - \sum_{i=1}^7 J_{i,i+1} (\hat{S}_i^x \hat{S}_{i+1}^x + \hat{S}_i^y \hat{S}_{i+1}^y + \hat{S}_i^z \hat{S}_{i+1}^z)$$

$$J_1 = J_{1,2} = J_{7,8}$$

$$J_3 = J_{3,4} = J_{5,6}$$

$$J_2 = J_{2,3} = J_{6,7}$$

$$J_4 = J_{4,5}$$



超交換相互作用(Cu-O-Cuパス)に  
注目すると…



独立な最近接相互作用のみ働く,  
8スピンからなるジグザグスピン  
と近似できる.

# L2正則化とノイズ

1, 正則化の強さを決める.

エネルギー関数はエルボーカーブ  
(正則化が効く or 効かないの間)



エルボーポイントを適切な値とする.  
モデルパラメタが決定

2, ノイズを決める.

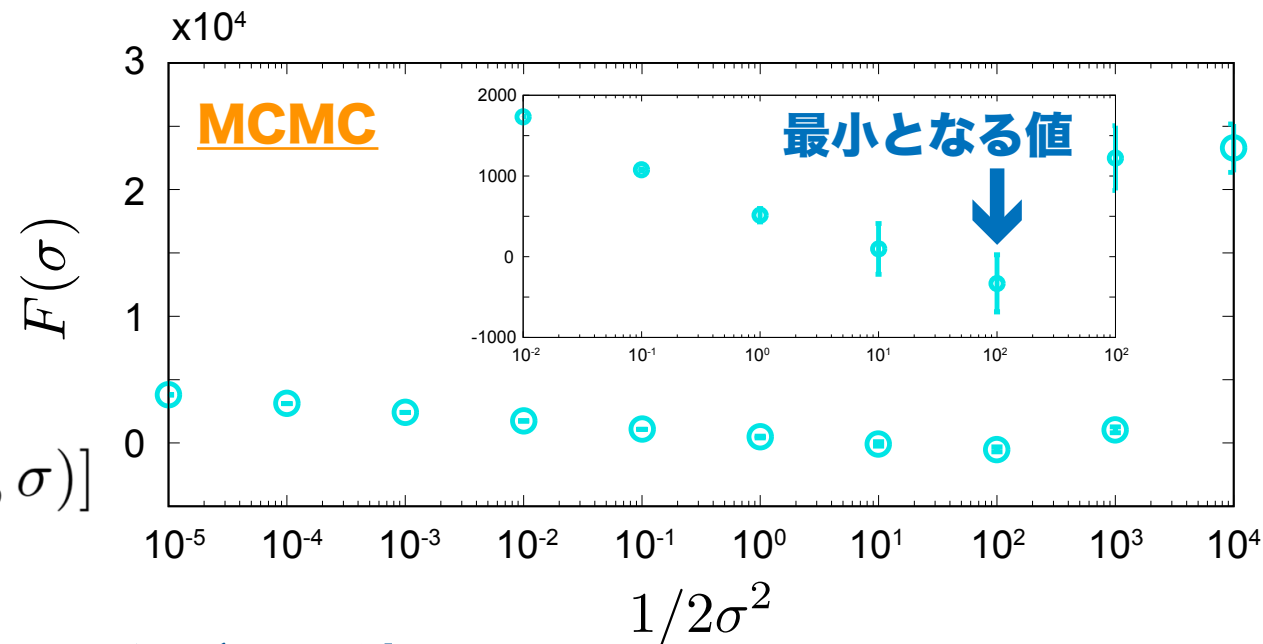
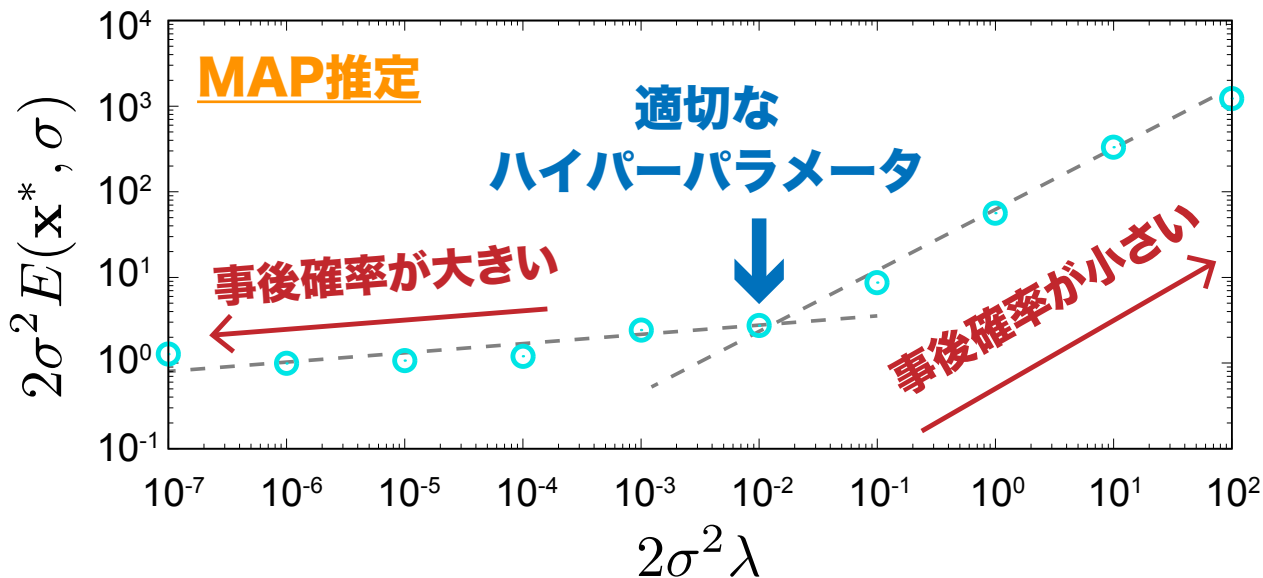
$F(\sigma) := -\log Z(\sigma)$  : ベイズ  
自由エネルギー

$$Z(\sigma) = \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{L}{2}} \int_{\Omega_x} d\mathbf{x} \exp[-E(\mathbf{x}, \sigma)]$$

パラメタの誤差が決定

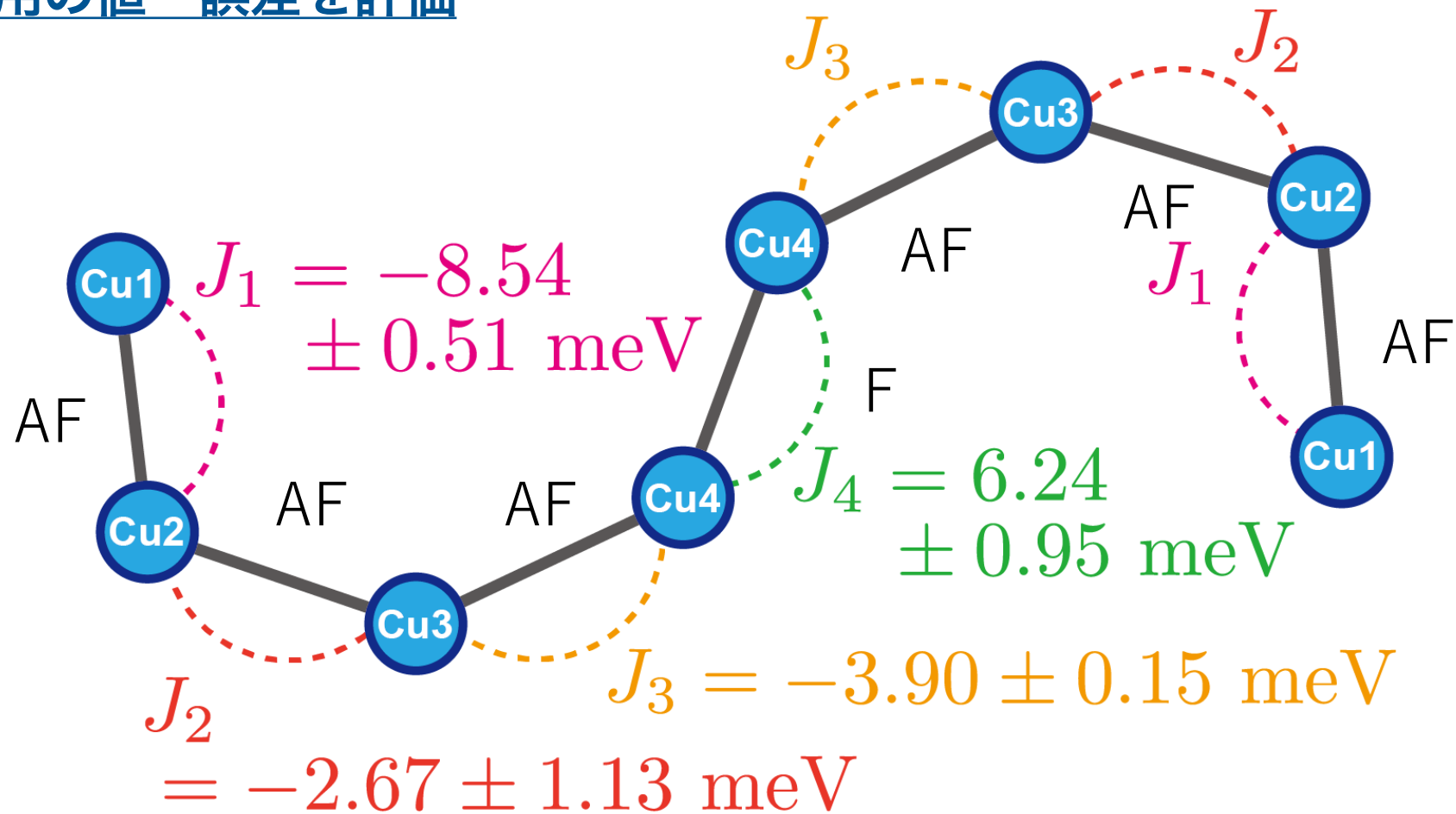
事後確率のパラメタが全て決まる.

$$P(\mathbf{x}) \propto \exp(-\lambda \|\mathbf{x}\|^2)$$

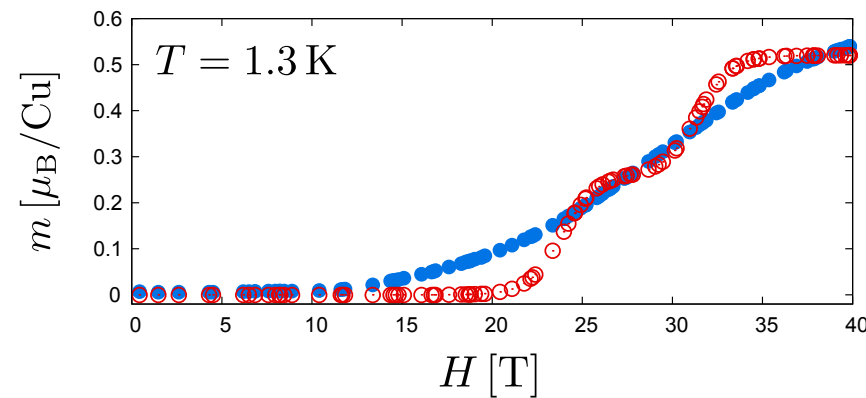
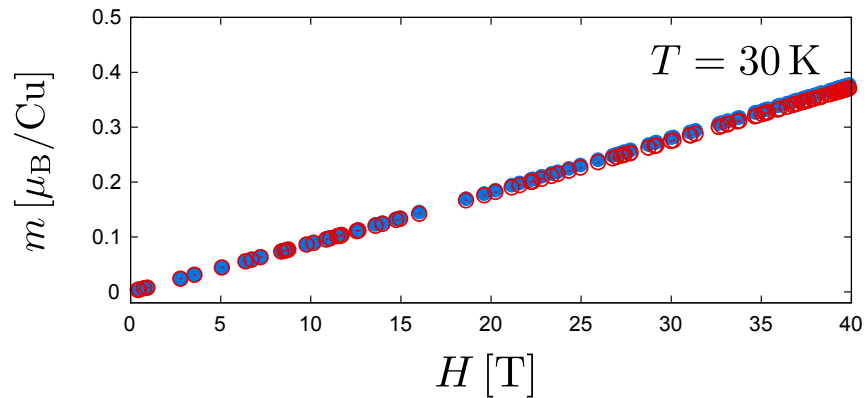
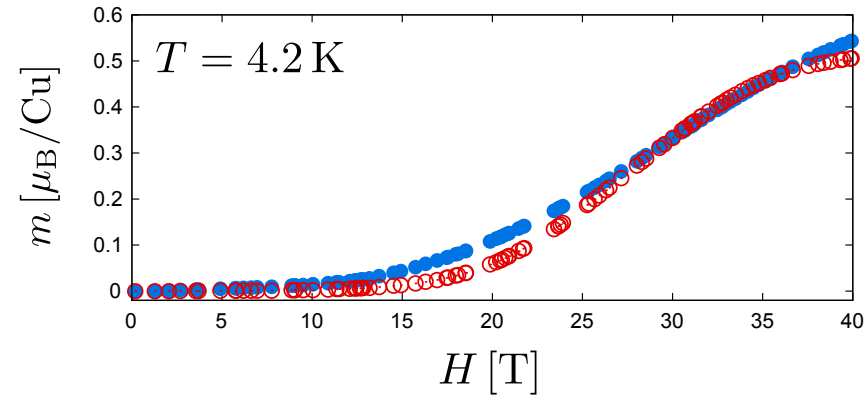
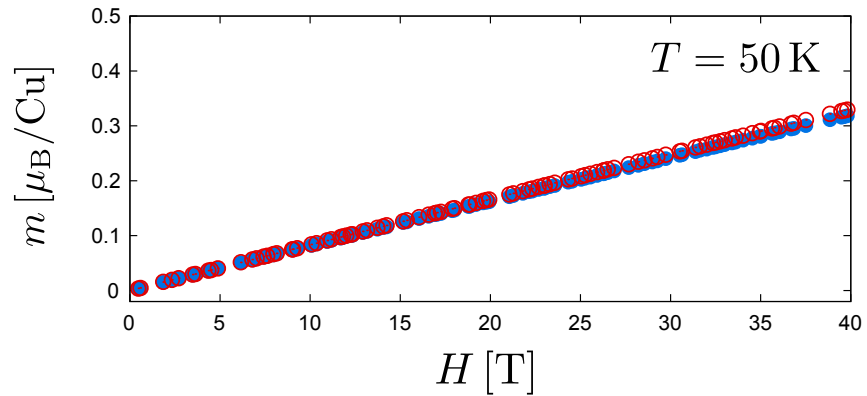
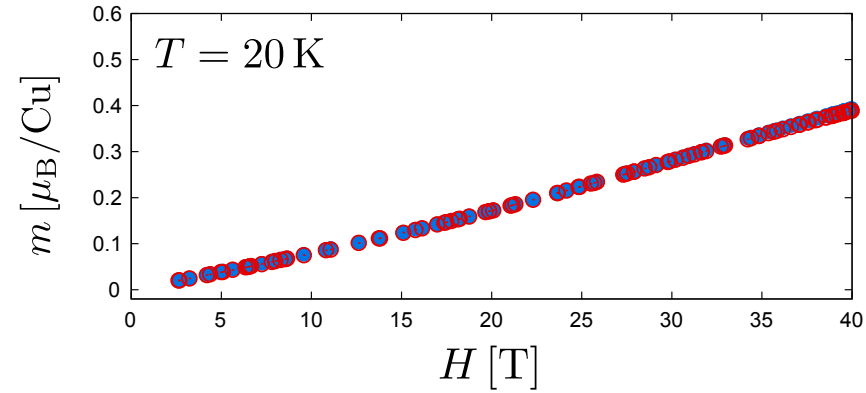
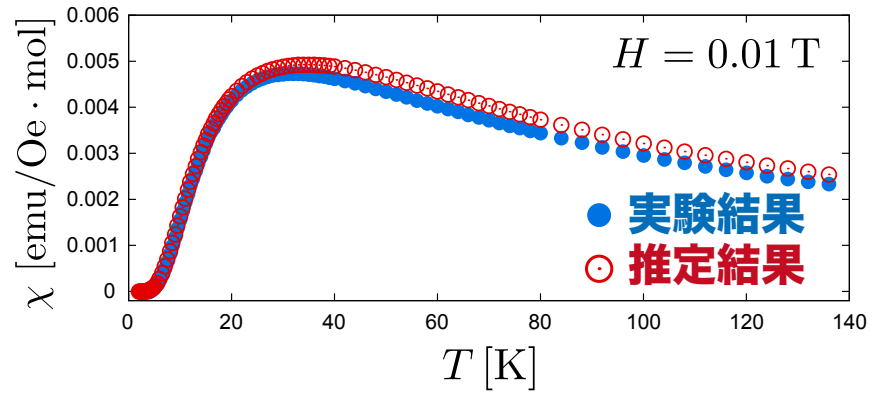


# 推定された相互作用

MCMCにより事後確率分布を求め、分布の平均・分散により相互作用の値・誤差を評価



# 物理量の比較

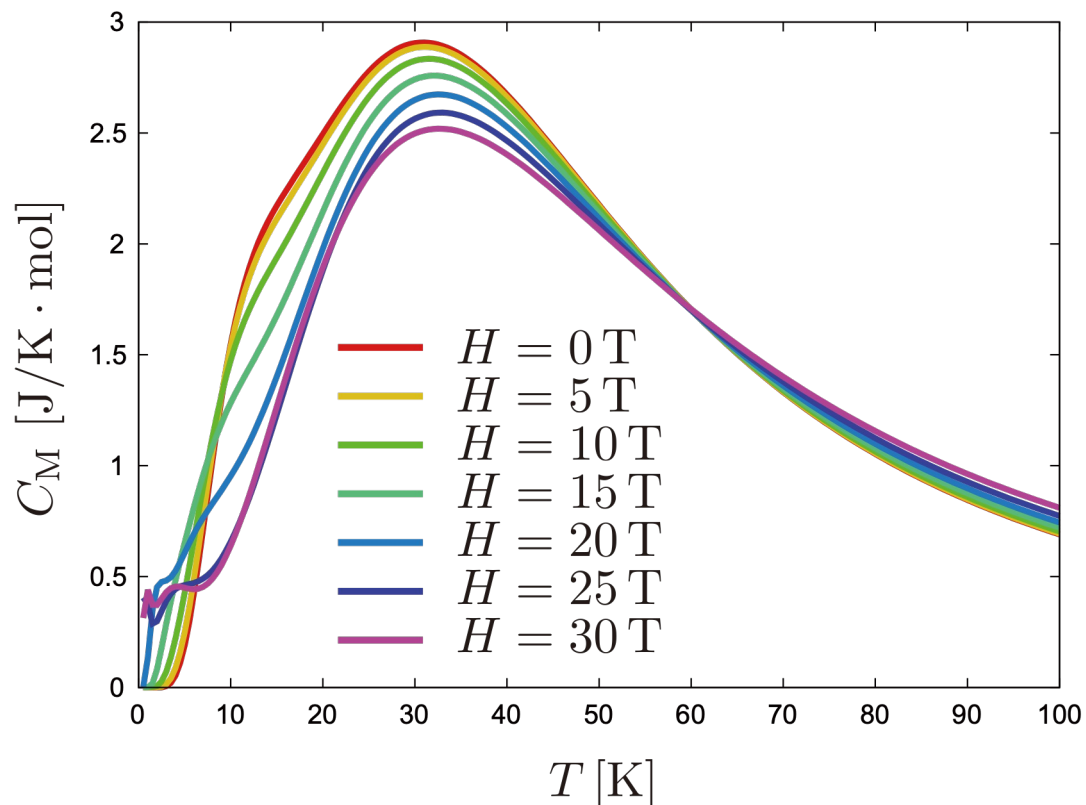


# 推定モデルを用いた特性予測

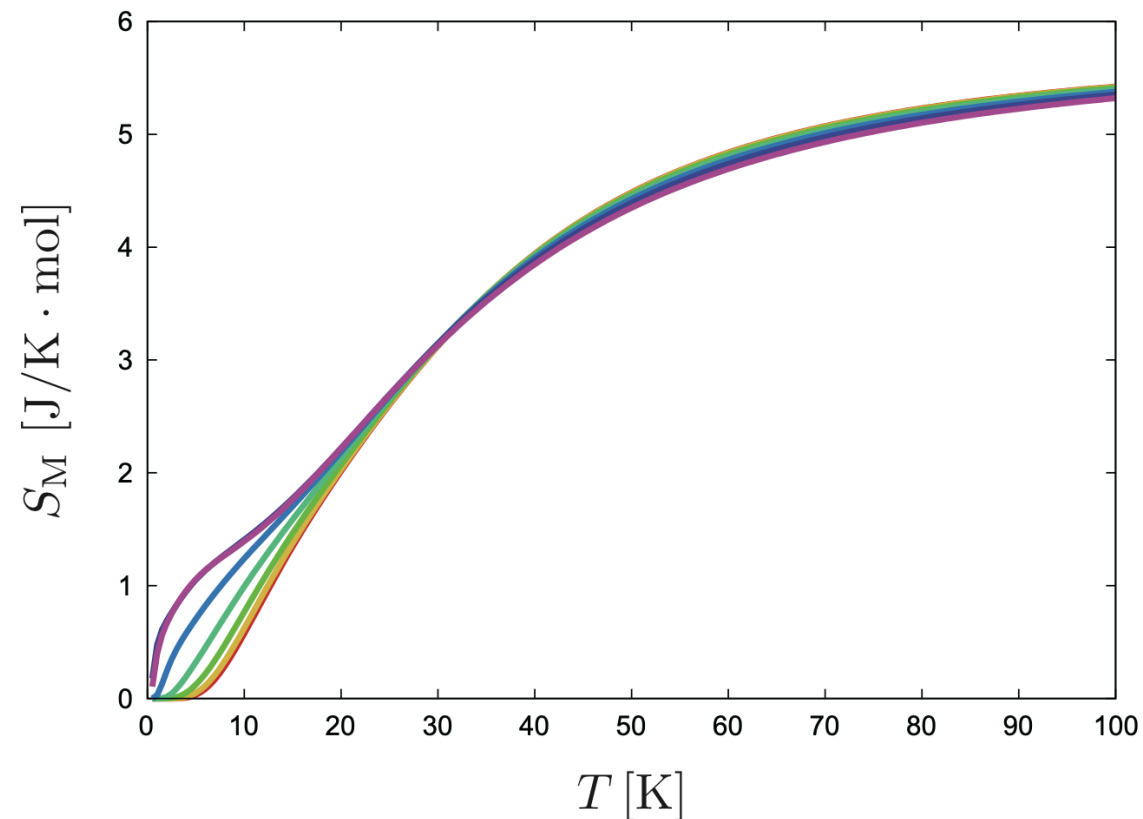
実験が難しい物理量が推定モデルを用いて予測することが可能

スピンギャップ :  $2.87 \pm 0.06$  meV

## 磁気比熱



## 磁気エントロピー

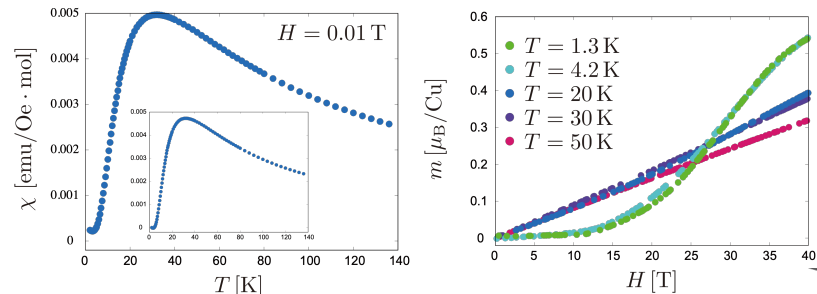




# まとめ

モチベーション：機械学習を利用することで、実験データを説明できる有効モデル（ハミルトニアン）を推定したい。

## 実験結果



帯磁率, 磁化, 比熱...

## 候補モデル

$$J_{ij} \mathbf{s}_i \cdot \mathbf{s}_j \quad b_{ij} (\mathbf{s}_i \cdot \mathbf{s}_j)^2 \quad \mathbf{d}_{ij} \cdot [\mathbf{s}_i \times \mathbf{s}_j] \quad D_i (s_i^z)^2$$
$$\frac{\mathbf{s}_i \cdot \mathbf{s}_j}{r_{ij}^3} - 3 \frac{(\mathbf{s}_i \cdot \mathbf{r}_{ij})(\mathbf{s}_j \cdot \mathbf{r}_{ij})}{r_{ij}^5}$$

実験結果を再現するモデルの種類・モデルパラメータを決める。

R. Tamura and K. Hukushima, PRB 95, 064407 (2017).

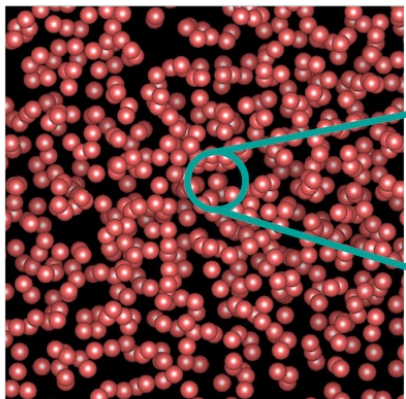
R. Tamura and K. Hukushima, PLoS ONE 13, e0193785 (2018).

R. Tamura, K. Hukushima, A. Matsuo, K. Kindo, M. Hase, PRB 101, 224435 (2020).

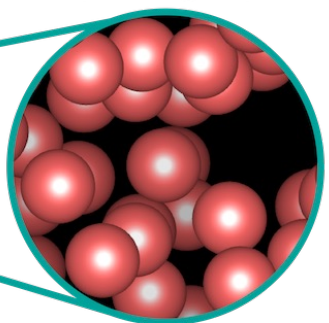
# 不規則構造をデータ駆動で解析する

# 分子動力学の構造解析に機械学習を利用

Snapshot by MD simulations



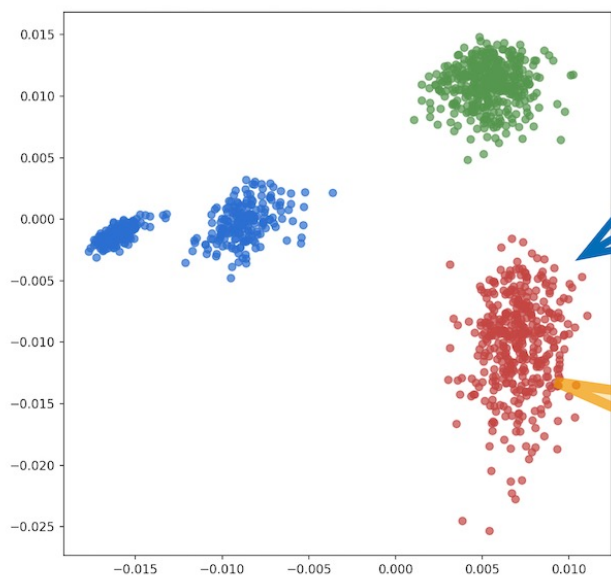
Local structure  
(enlarged)



LAAF descriptor  
(high-dimensional vector)

$$\mathbf{v}_i^{\text{av}} = \begin{pmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}$$

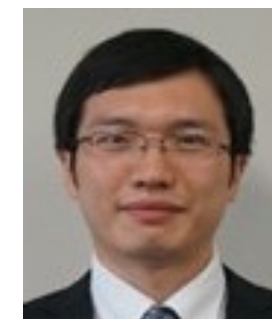
Low-dimensional space



*Dimensionality reduction  
by TS-LPP (linear mapping)*

**Extract**

- relations of atoms
- classification of atoms
- anomalous atoms



R. Tamura, M. Matsuda, J. Lin,  
Y. Futamura, T. Sakurai,  
and Tsuyoshi Miyazaki,  
Physical Review B 105, 075107 (2022).

# モチベーション

分子動力学計算で微視的に何が起きているのか？

原子周りの局所構造から原子の性質を解明



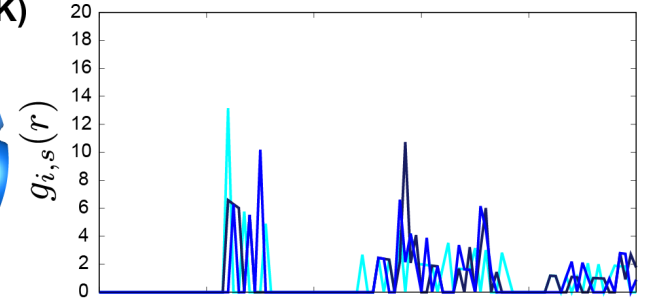
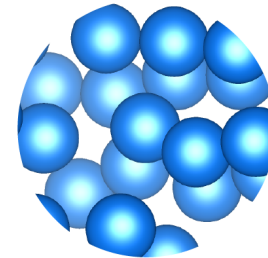
- ① 局所構造による原子の分類
- ② 原子間の相関抽出
- ③ 異常原子の発見

しかし，原子周りの局所構造だけで状態を判定することは**簡単ではない**…  
(局所的なRDFでは全くわからない)

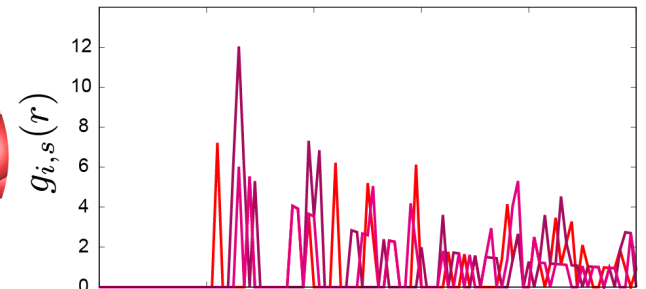
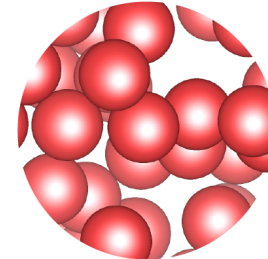
**機械学習による解析で問題を克服！**

Radial distribution of neighbor atoms around one atom

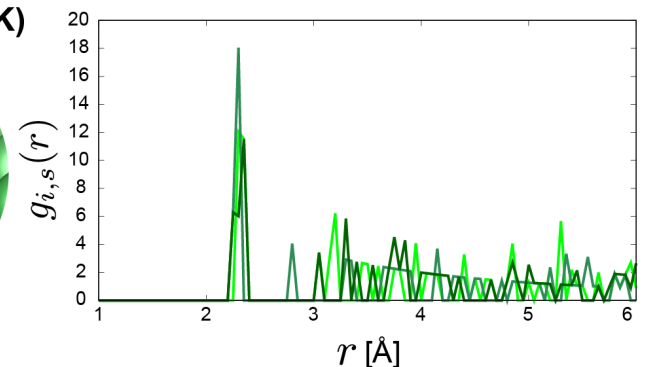
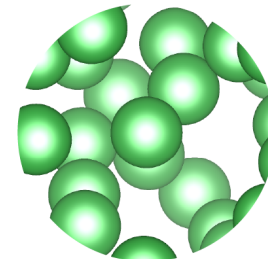
Crystalline (1200 K)



Liquid (3000 K)

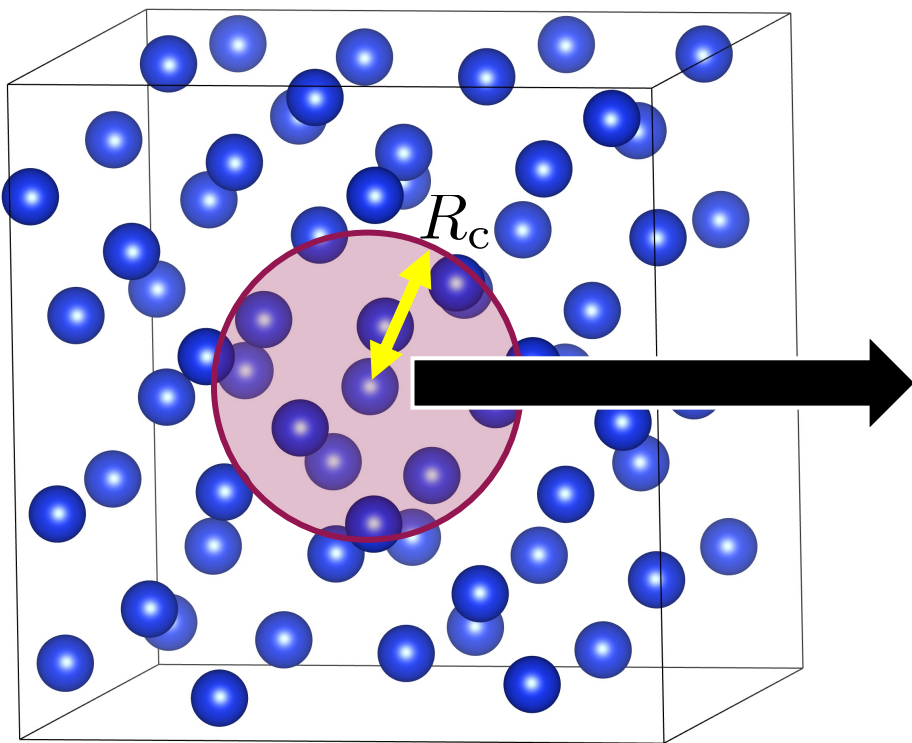


Amorphous (300 K)



# LAAF descriptor

原子周りの局所構造を記述



i番目の原子のdescriptor

$$V_i(\eta_m; R_c) = \sum_{j \neq i} \exp[-(r_{ij}/\eta_m)^2] f(r_{ij}; R_c)$$

$$f(r_{ij}; R_c) = \begin{cases} 0.5[\cos(\pi r_{ij}/R_c)^2 + 1] & (r_{ij} \leq R_c) \\ 0 & (r_{ij} > R_c) \end{cases}$$

複数の  $\eta$  (どの程度の距離の原子を扱うか) を導入し、高次元descriptorを作成

$$\mathbf{V}_i(R_c) = (V_i(\eta_1; R_c), V_i(\eta_2; R_c), \dots, V_i(\eta_M; R_c))$$

近傍の原子の平均 (LAAF descriptor)

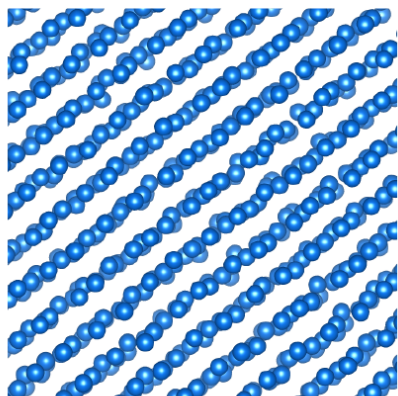
$$\mathbf{V}_i^{\text{av}}(R_c, R_a) = \frac{1}{N_{\in R_a}} \sum_{j \in R_a} \mathbf{V}_j(R_c)$$



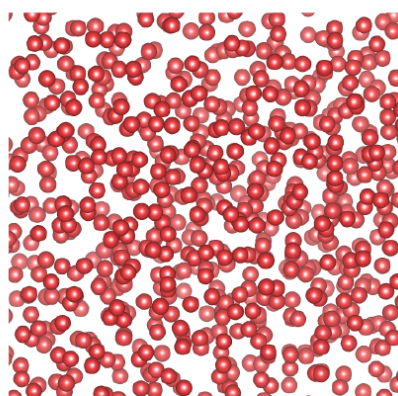
# PCA (教師なし学習) で次元削減 (Si一元系)

Si一元系における固体, 液体, アモルファス状態を第一原理MD計算(CONQUEST)で生成

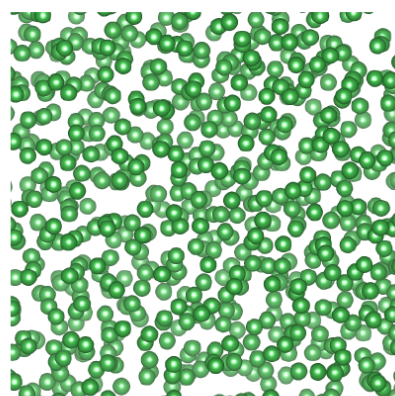
LAAF descriptor(100次元)に対してPCAを利用し2次元空間に次元削減



固体(300K)  
固体(1200K)

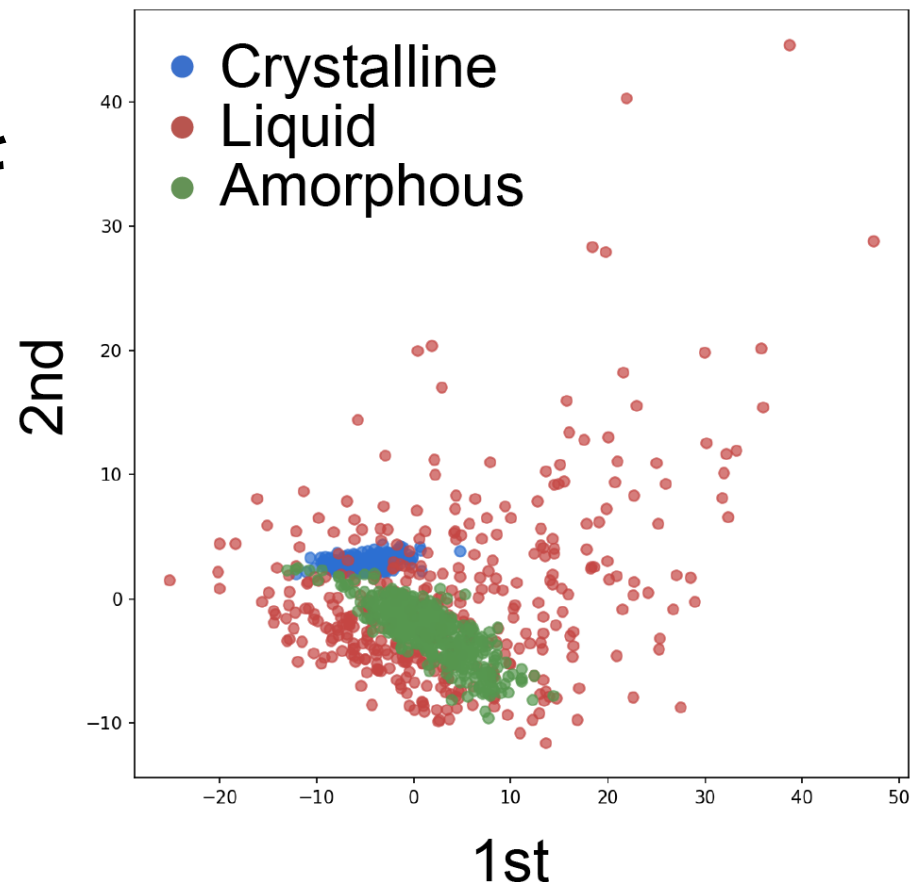


液体(3000K)  
液体(9000K)



アモルファス  
(300K)  
2種類

## PCA

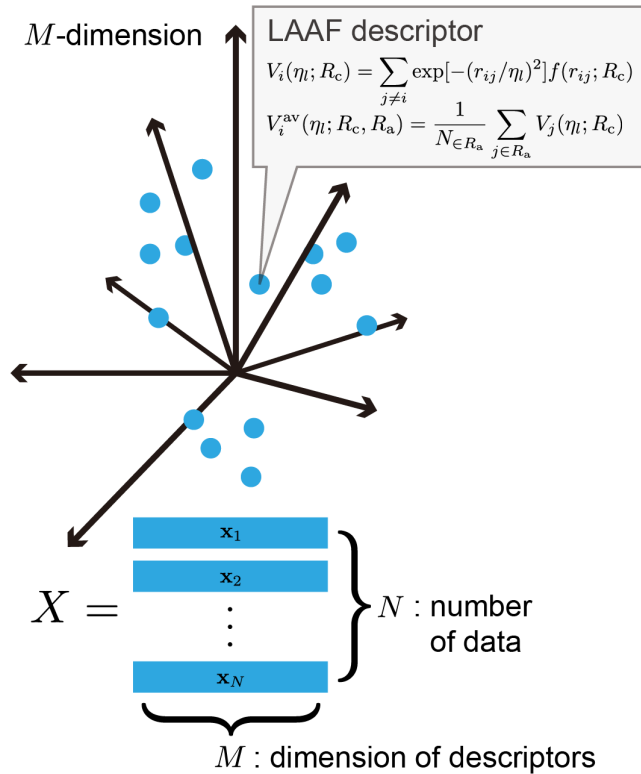


PCAではこれらの原子が分かれて分布する空間を見つけられない。

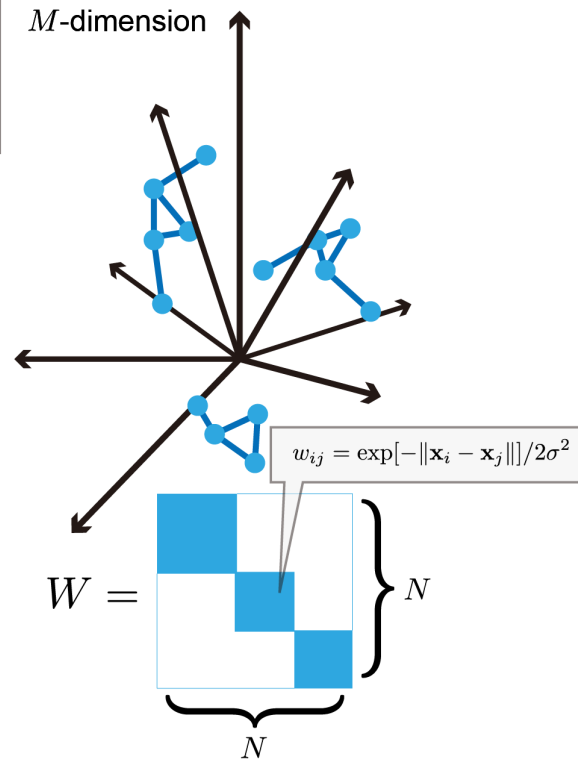
# 次元削減手法TS-LPP (線形次元削減)

次元削減手法Locality Preserving Projections(LPP)を2回行う。

High-dimensional data



Similarity graph w/ adjacency matrix



Linear mapping function

Generalized eigenvalue problem

$$X^T L X \mathbf{y} = \lambda X^T D X \mathbf{y}$$

$$L = D - W$$

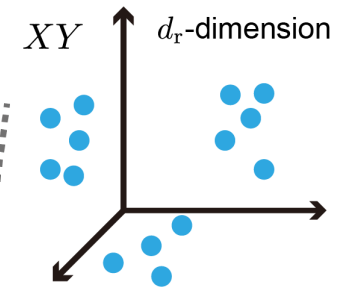
$$D = \text{diag.} \left( \sum_{j=1}^N w_{ij} \right)$$

↓ In ascending order of eigenvalues

$$Y = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \dots & \mathbf{y}_d \end{bmatrix}$$
 }  $M$ 

$d_r$ : reduced dimension

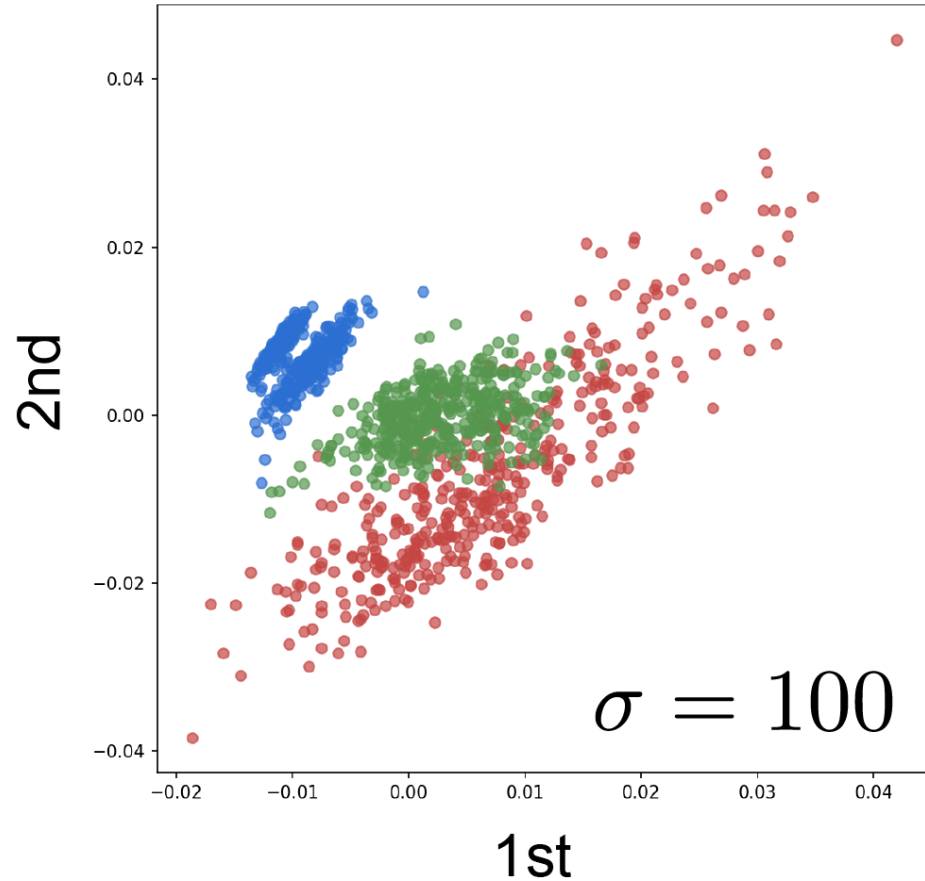
Low-dimensional data



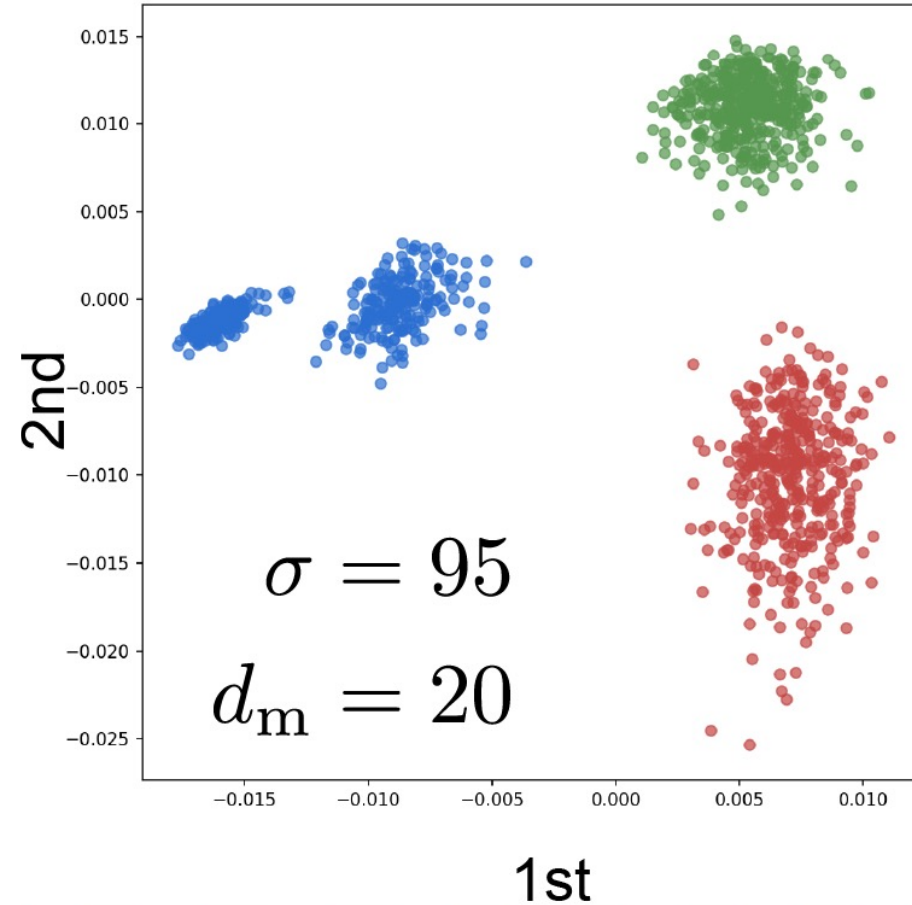
Repeat twice in TS-LPP (First reduced dimension is relatively high.)

# Si一元系へ適用

## LPP



## TS-LPP



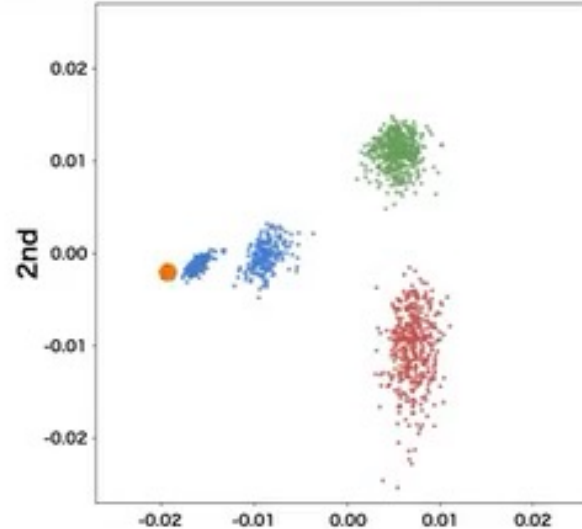
TS-LPP法では，固体，液体，アモルファスが明確に分かれている  
二次元空間を見つけることができる。

# Melt-quench processの解析

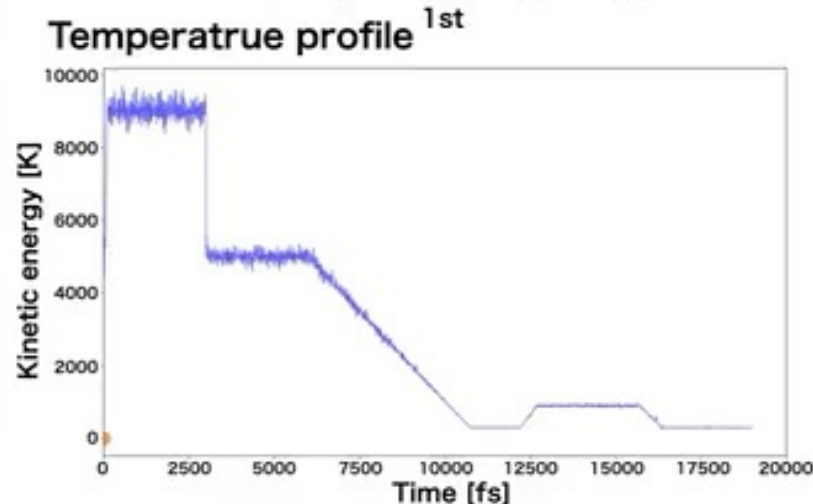
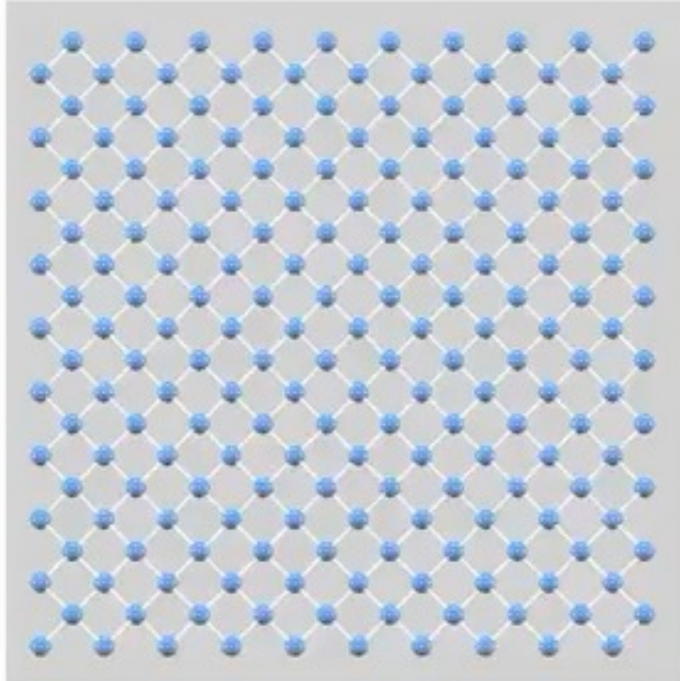
## Melt-quench process 1 in the amorphous formation



Reduced space by TS-LPP



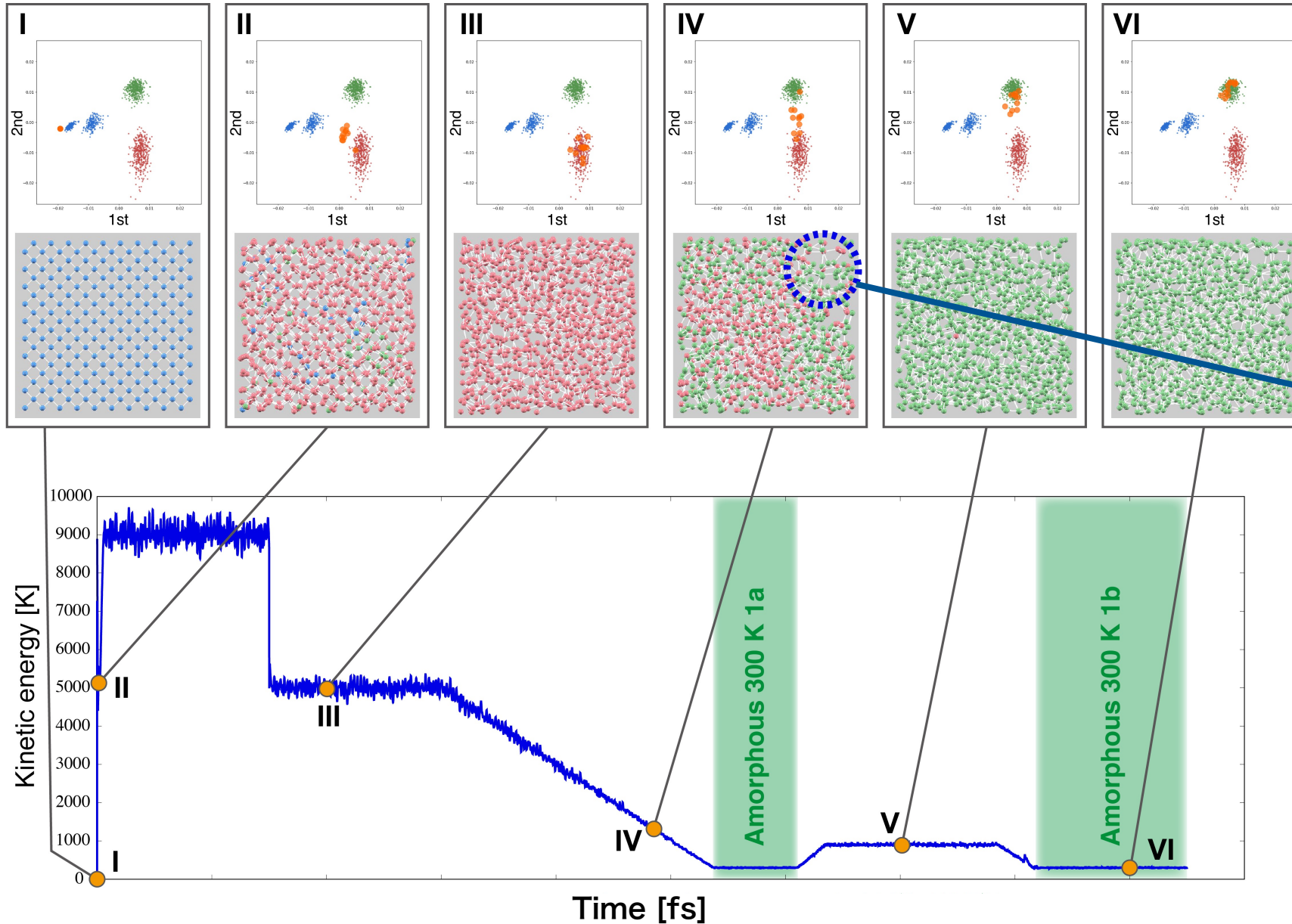
Molecular dynamics simulation results



- TS-LPPでは、新しい原子を低次元にプロットできる。
- Melt-quench processのMD計算を行い、全ての原子を低次元にプロット
- 低次元空間の位置から、各原子がどの相に属しているかを判定 (K-neighbor法)



# Melt-quench processの解析



この辺りを中心に  
アモルファス構造  
が広がっている。

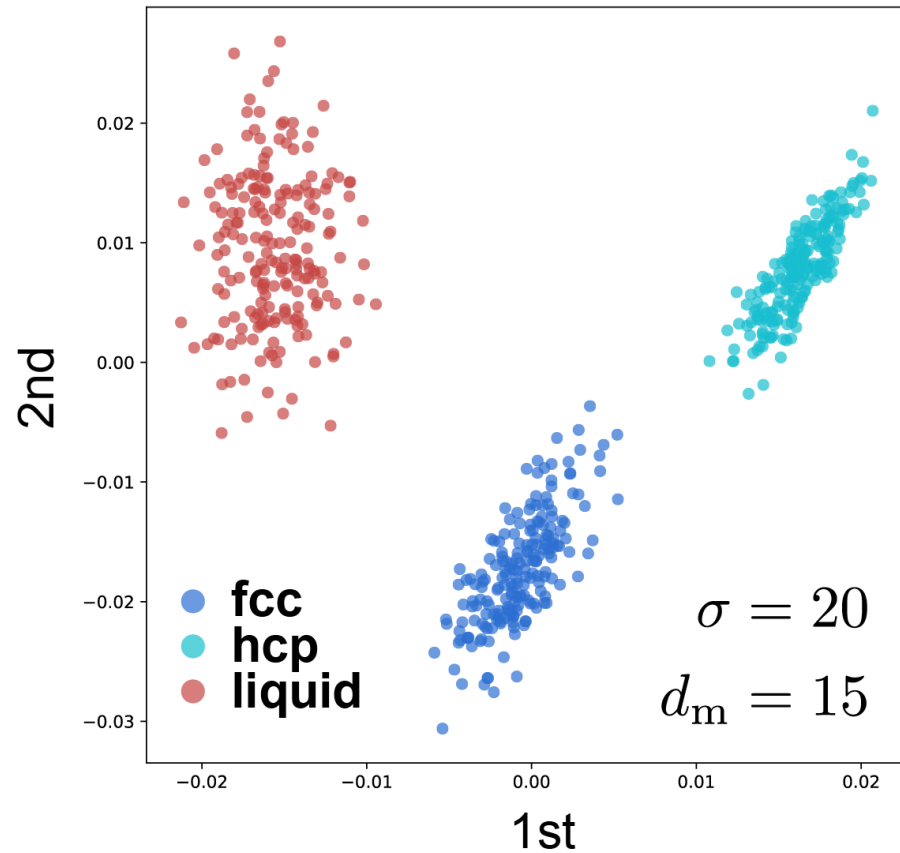
相転移の始まり？



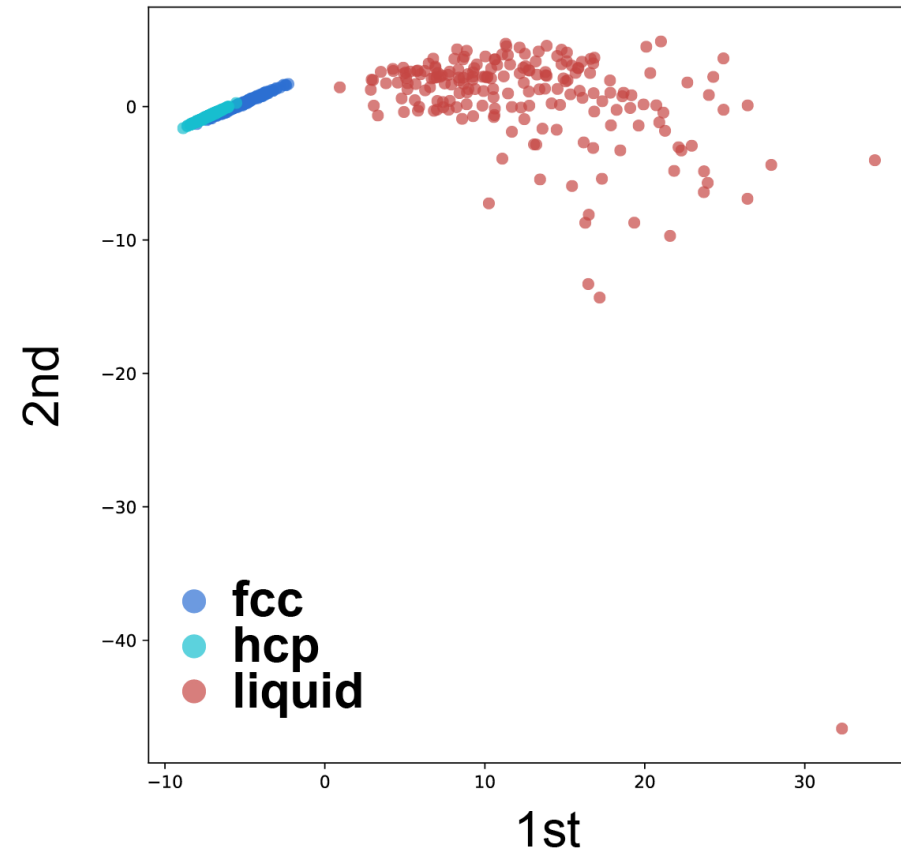
# Cu一元系へ適用

金属系へのTS-LPPの適用 (FCC, HCP, 液体構造)

## TS-LPP

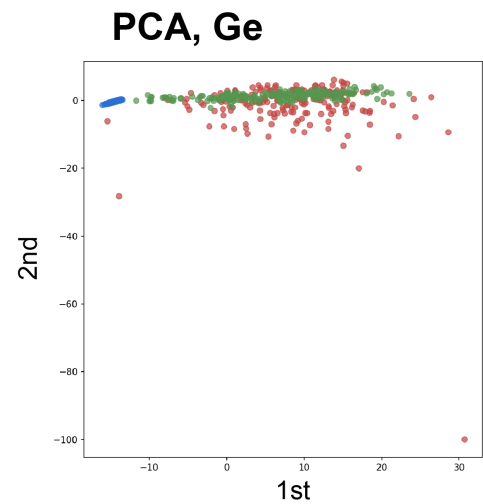
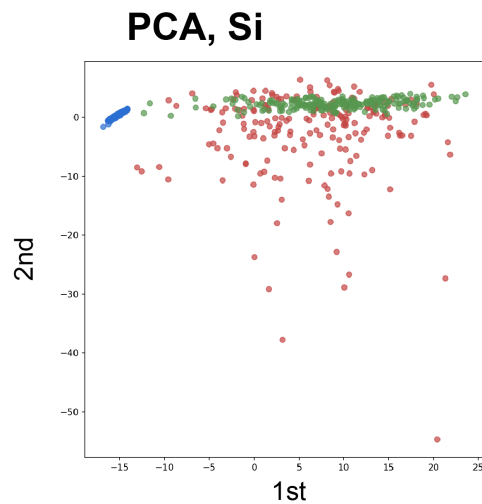
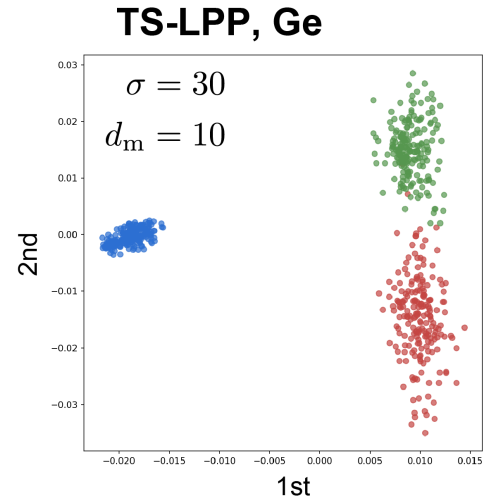
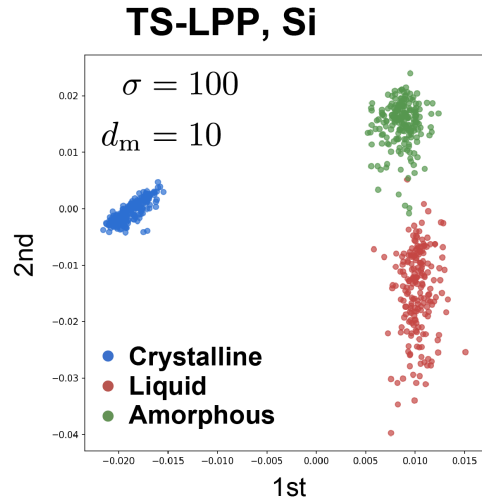


## PCA



# Si-Ge二元系へ適用

## 二元系へのTS-LPPの適用（固体，液体，アモルファル構造）



Si原子に対するdescriptor

Si原子のみのdescriptor

$$V_i(\eta_m; R_c) = \sum_{\substack{j \neq i \\ \text{jはSiのみ}}} \exp[-(r_{ij}/\eta_m)^2] f(r_{ij}; R_c)$$

+

Ge原子のみのdescriptor

$$V_i(\eta_m; R_c) = \sum_{\substack{j \neq i \\ \text{jはGeのみ}}} \exp[-(r_{ij}/\eta_m)^2] f(r_{ij}; R_c)$$

一元系の場合の倍の次元

# Github

The screenshot shows the GitHub interface for the repository 'rtmr / TS-LPP'. At the top, there is a navigation bar with the GitHub logo, a search bar, and links for 'Pull requests', 'Issues', 'Marketplace', and 'Explore'. Below this, the repository name 'rtmr / TS-LPP' is displayed with a 'Public' badge. To the right, there are buttons for 'Pin' and 'Unwatch' (with a count of 2). A secondary navigation bar includes links for 'Code', 'Issues', 'Pull requests', 'Actions', 'Projects', 'Wiki', 'Security', 'Insights', and 'Settings'. Below the navigation, there are buttons for 'Go to file', 'Add file', and 'Code'. The repository details section shows the commit hash 'ad5834c' on '7 Jan' with '18 commits'. A table lists the files: 'README.md' (updated 2 years ago), 'TS-LPP.py' (updated 7 months ago), and 'Training.csv' (added 2 years ago). Below the file list, the 'README.md' file is selected, showing the title 'TS-LPP' and the description 'Two Step Locality Preserving Projections (TS-LPP)'.

Search or jump to... / Pull requests Issues Marketplace Explore

rtmr / TS-LPP Public Pin Unwatch 2

<> Code Issues Pull requests Actions Projects Wiki Security Insights Settings

master 1 branch 0 tags Go to file Add file Code

rtmr Update TS-LPP.py ad5834c on 7 Jan 18 commits

README.md	Update README.md	2 years ago
TS-LPP.py	Update TS-LPP.py	7 months ago
Training.csv	Add files via upload	2 years ago

☰ README.md

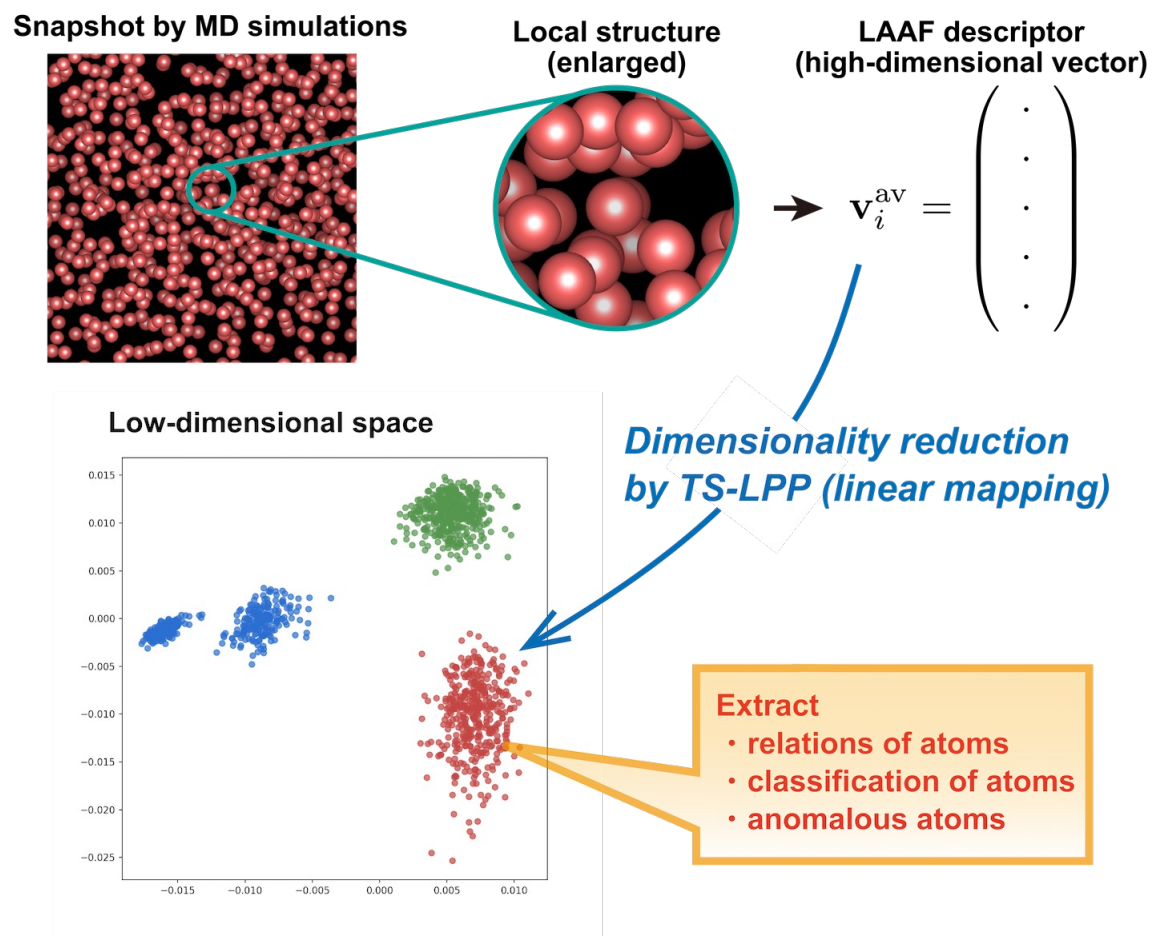
## TS-LPP

Two Step Locality Preserving Projections (TS-LPP)

<https://github.com/rtmr/TS-LPP>

# 分子動力学の構造解析に機械学習を利用

## 原子周りの局所構造から原子の性質を説明する手法を開発



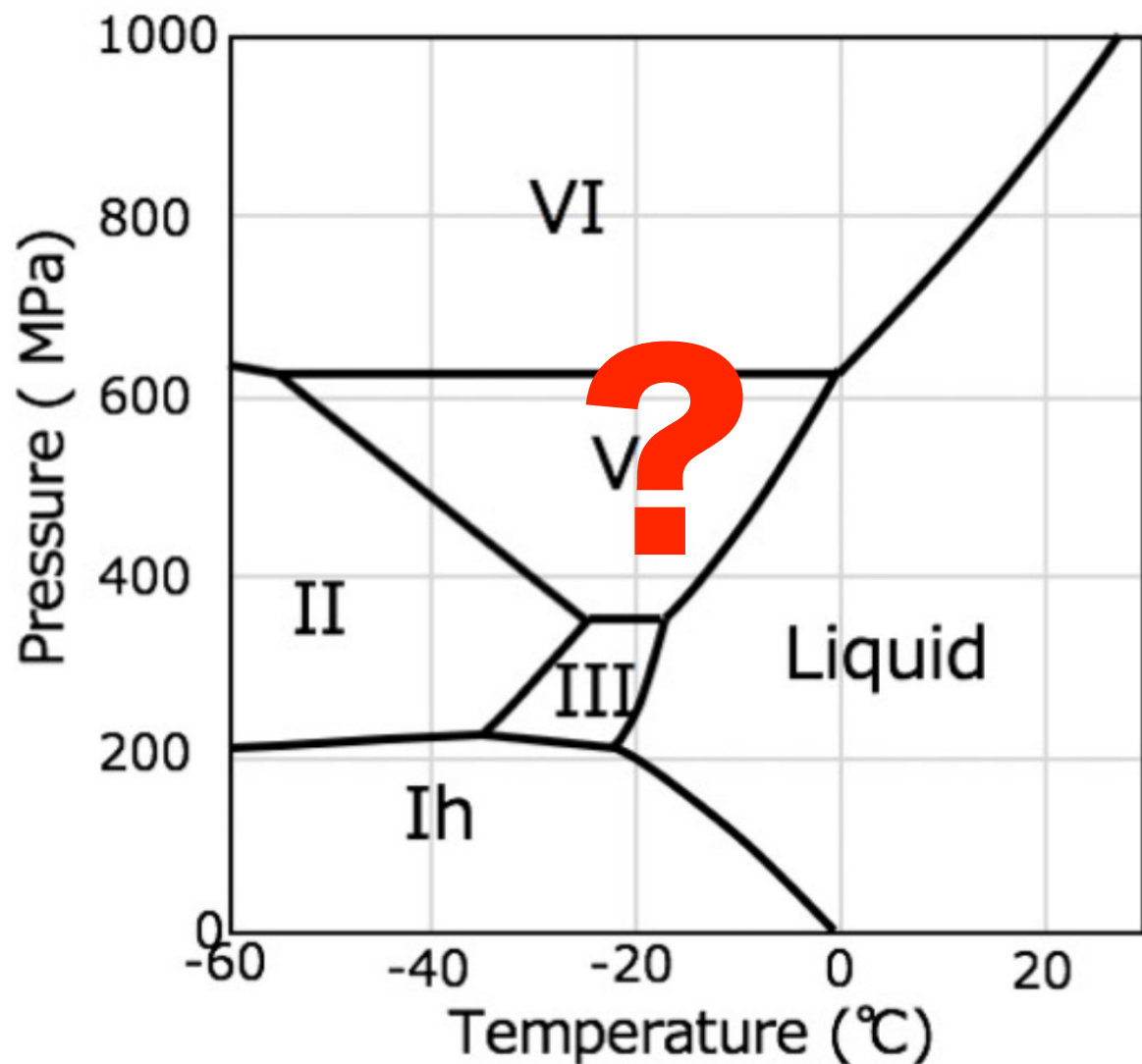
1. 原子の局所構造をdescriptorで表現
2. TS-LPP法を用いた教師なし学習で局所構造の違いを検出できる低次元空間を探索
3. 低次元空間でクラスタリング（解析しているMD結果にどの程度の種類の原子が含まれているか？）
4. 新しい原子をプロットし分類（各原子がどの種類に該当するか？）
5. 分子動力学計算で微視的に何が起きているのかを説明

R. Tamura, M. Matsuda, J. Lin, Y. Futamura, T. Sakurai, and Tsuyoshi Miyazaki, Physical Review B 105, 075107 (2022).

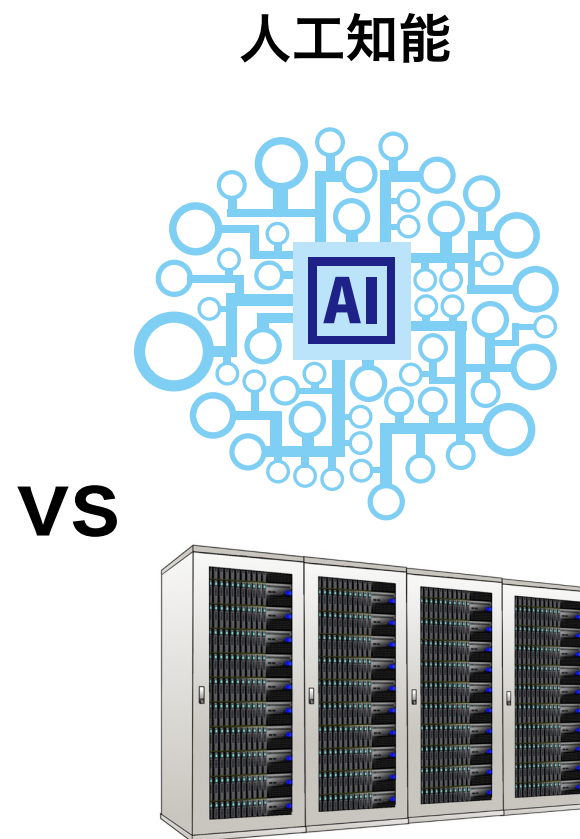
**相図を早く描きたい**



# 人工知能で相図はつくれるか？

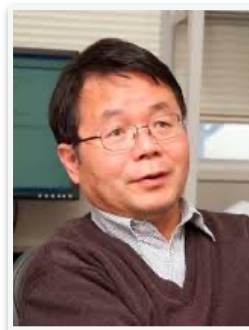
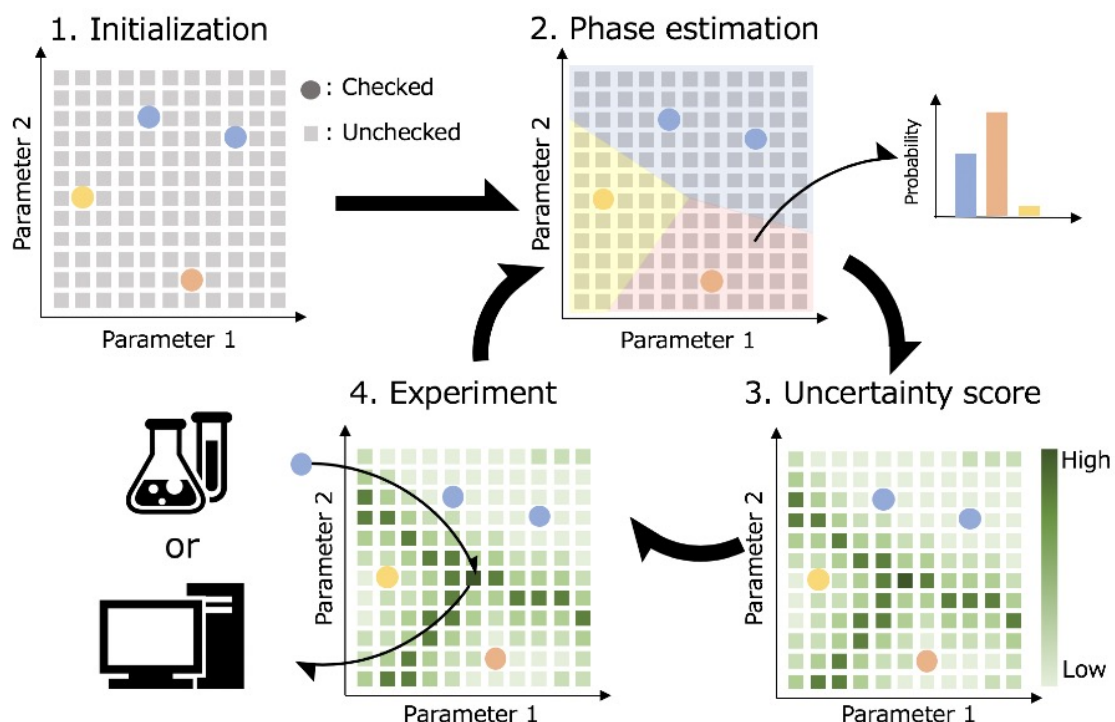


次にどこの実験をしますか？



# 人工知能による相図作成

能動学習（アクティブラーニング）を使って相図作成を効率化するための人工知能手法を開発



K. Terayama, R. Tamura, Y. Nose, H. Hiramatsu, H. Hosono, Y. Okuno, and K. Tsuda, *Physical Review Materials* 3, 033802 (2019).

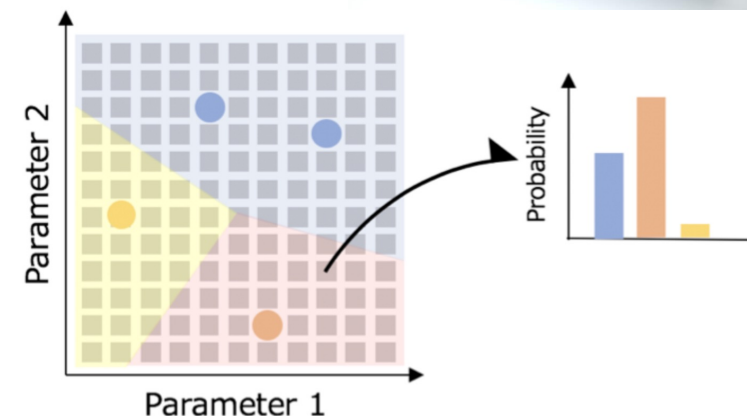
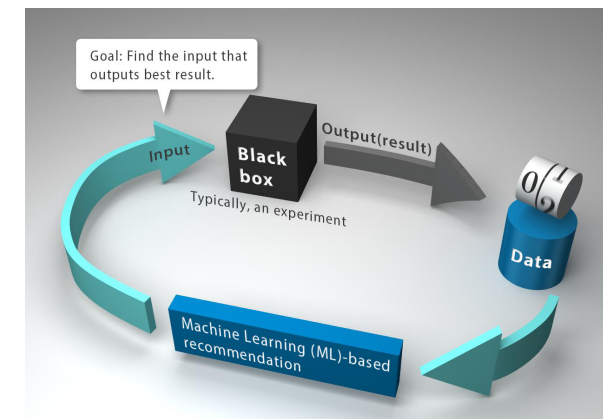
**利点1：サンプリング数を何も考えない  
場合と比べて20%に削減**

**利点2：新しい相を高速に発見**

# 人工知能による相図作成

相図を効率よく描くために次に実験する候補点を提案する。  
(ベイズ最適化と似た手法)

- N個の候補点があり，相図を作成するのに適した点を選びたい。
- できるだけ実験数を少なくしたい。
- M個の候補点に対する実験が終わった。
- 次のM+1個目の候補点を最適に選びたい。
- M個のサンプル点から予測モデルを学習し，それを用いて，残りの候補点をスコアリングし実際に観測するサンプルを選ぶ。

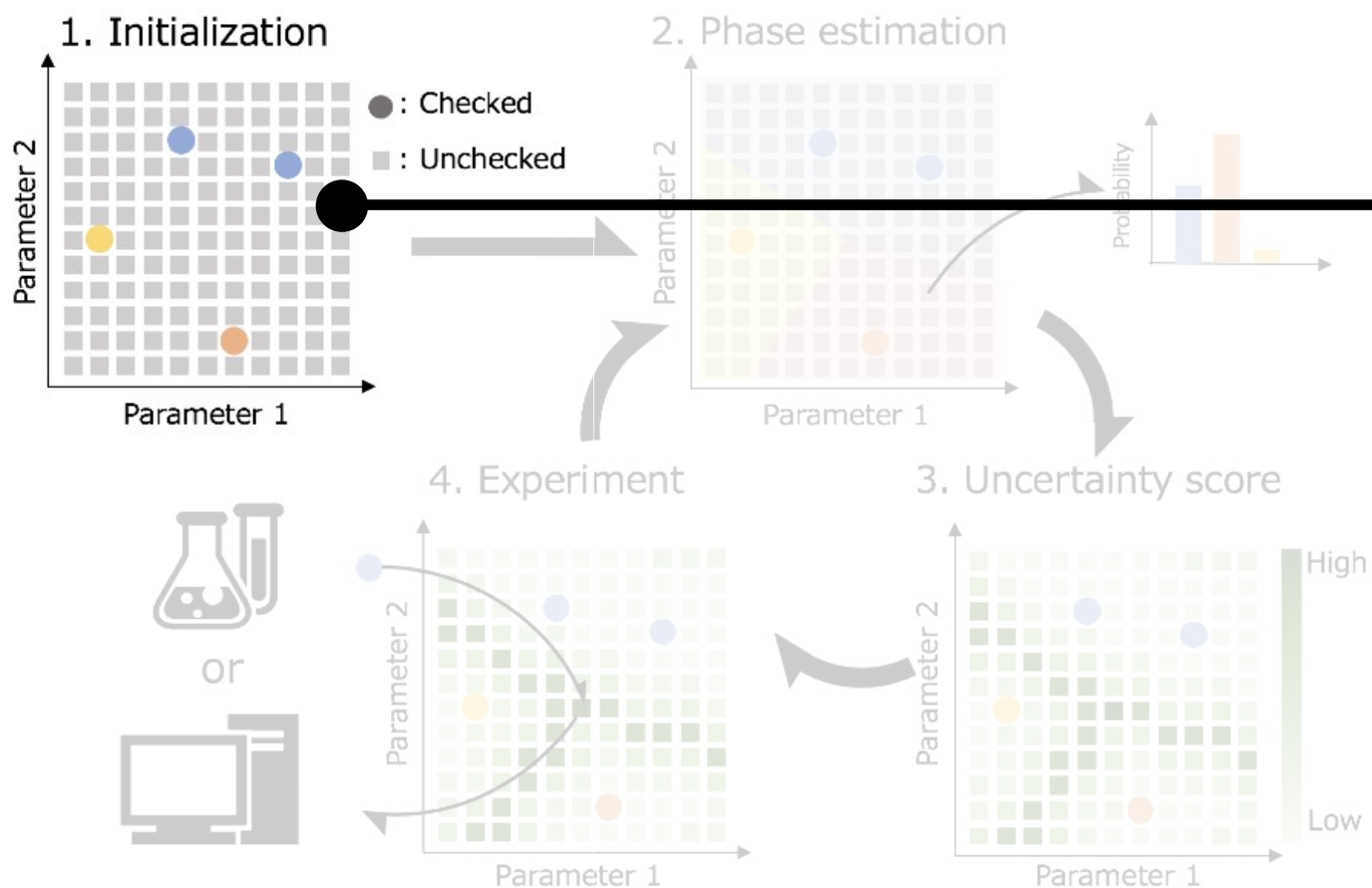






# 1. Initialization

初期点をランダムに用意する。  
(すでに分かっている点からスタートしてもOK)



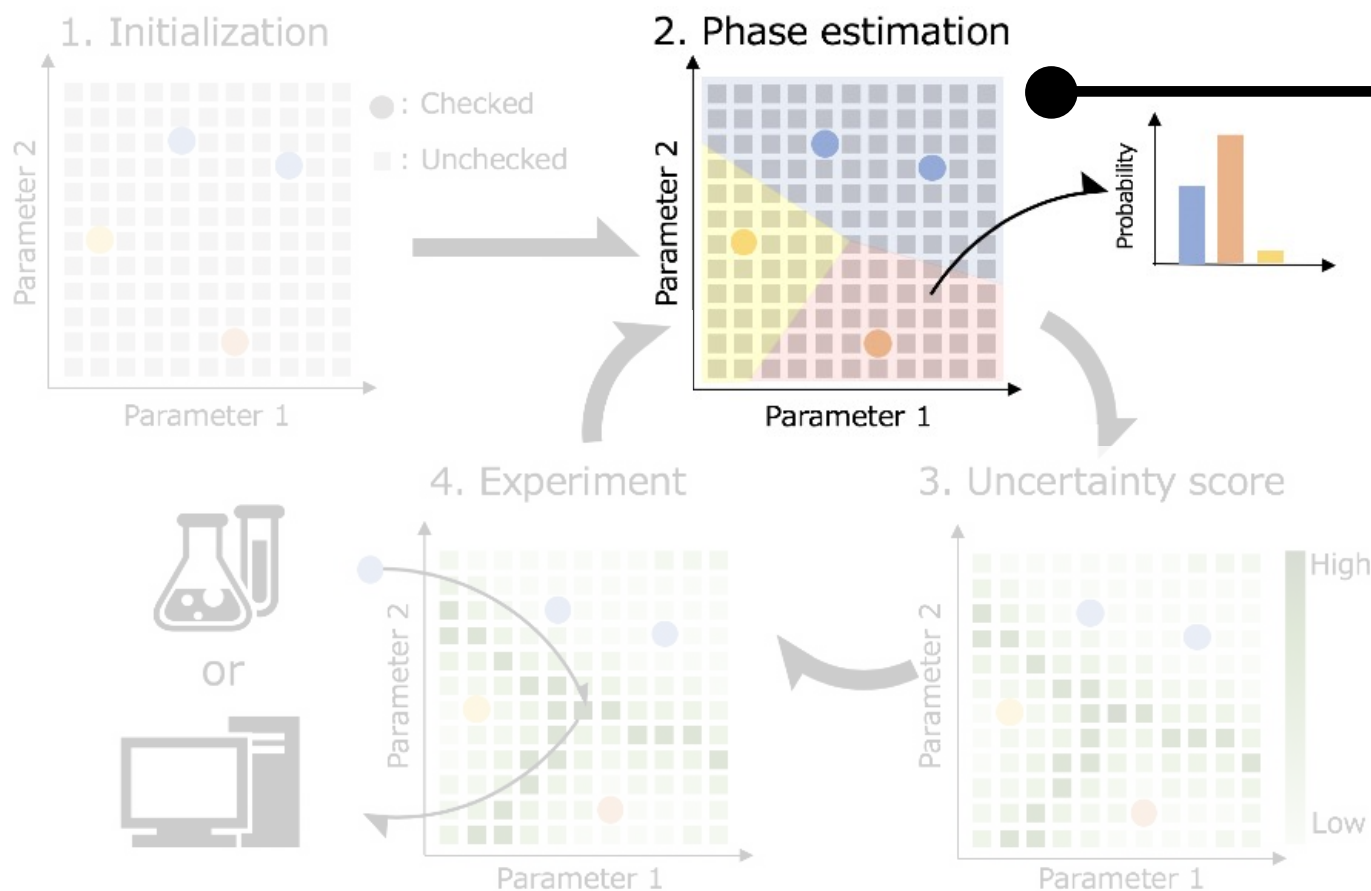
相図として描きたい空間を定義する。

離散化しておく必要がある。

相図の次元は何次元でも扱える。  
(サンプリング数が増える。)

# 2. Phase estimation

各点における所属確率を計算する。  
(すでに見つかっている相に対しての所属確率)



所属確率:

$$P(p|\mathbf{x})$$

相のラベル    相図中の位置

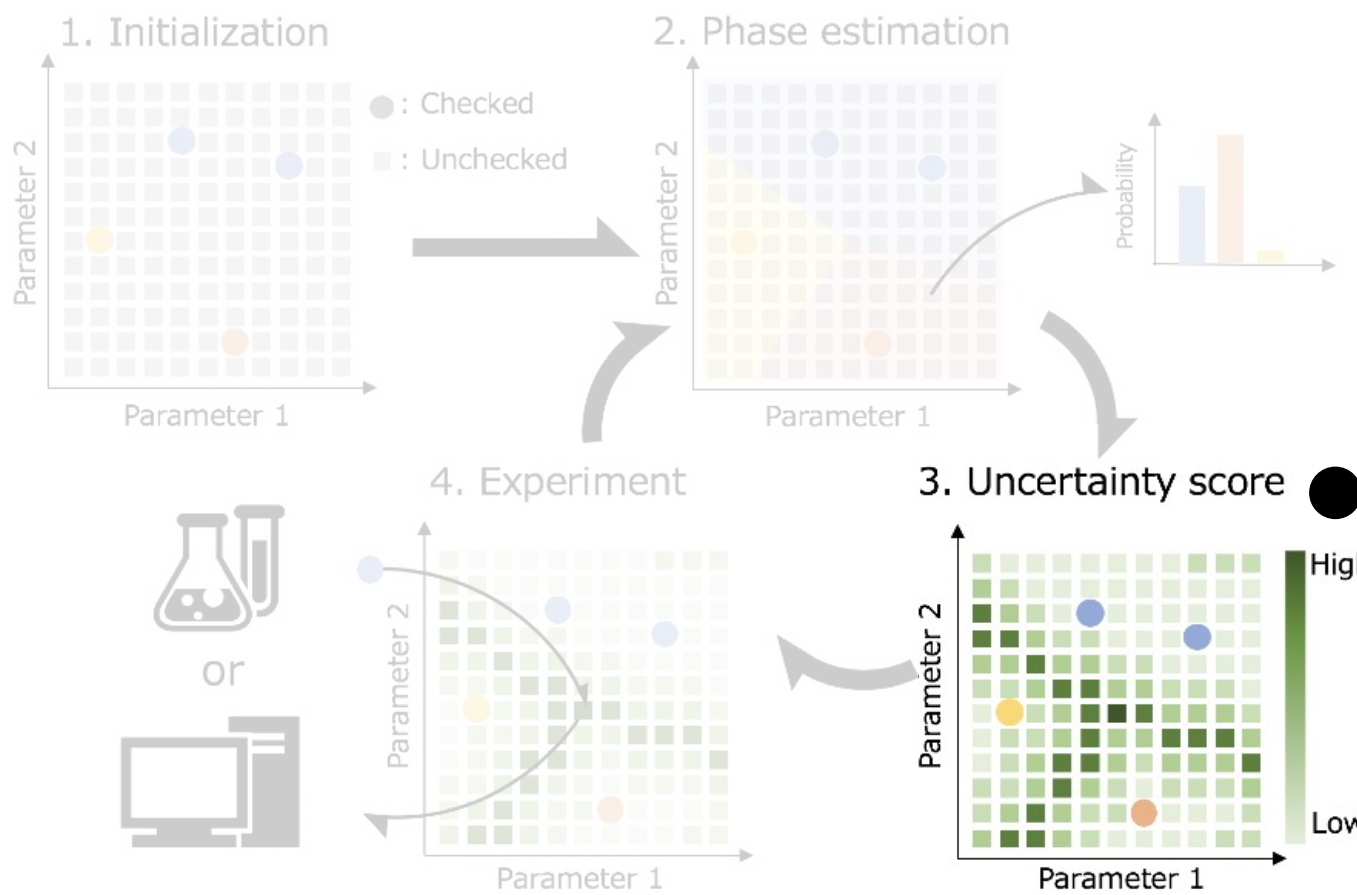
機械学習手法:

- Label propagation (LP)
- Label spreading (LS)

もしこの時点で相図が描きたければ、最大の確率を持つ相を利用すれば良い。

# 3. Uncertainty score

Uncertainty scoreを前ステップで得られた確率を使って評価する。  
(各点がどの程度“不確か”かを表すスコア)

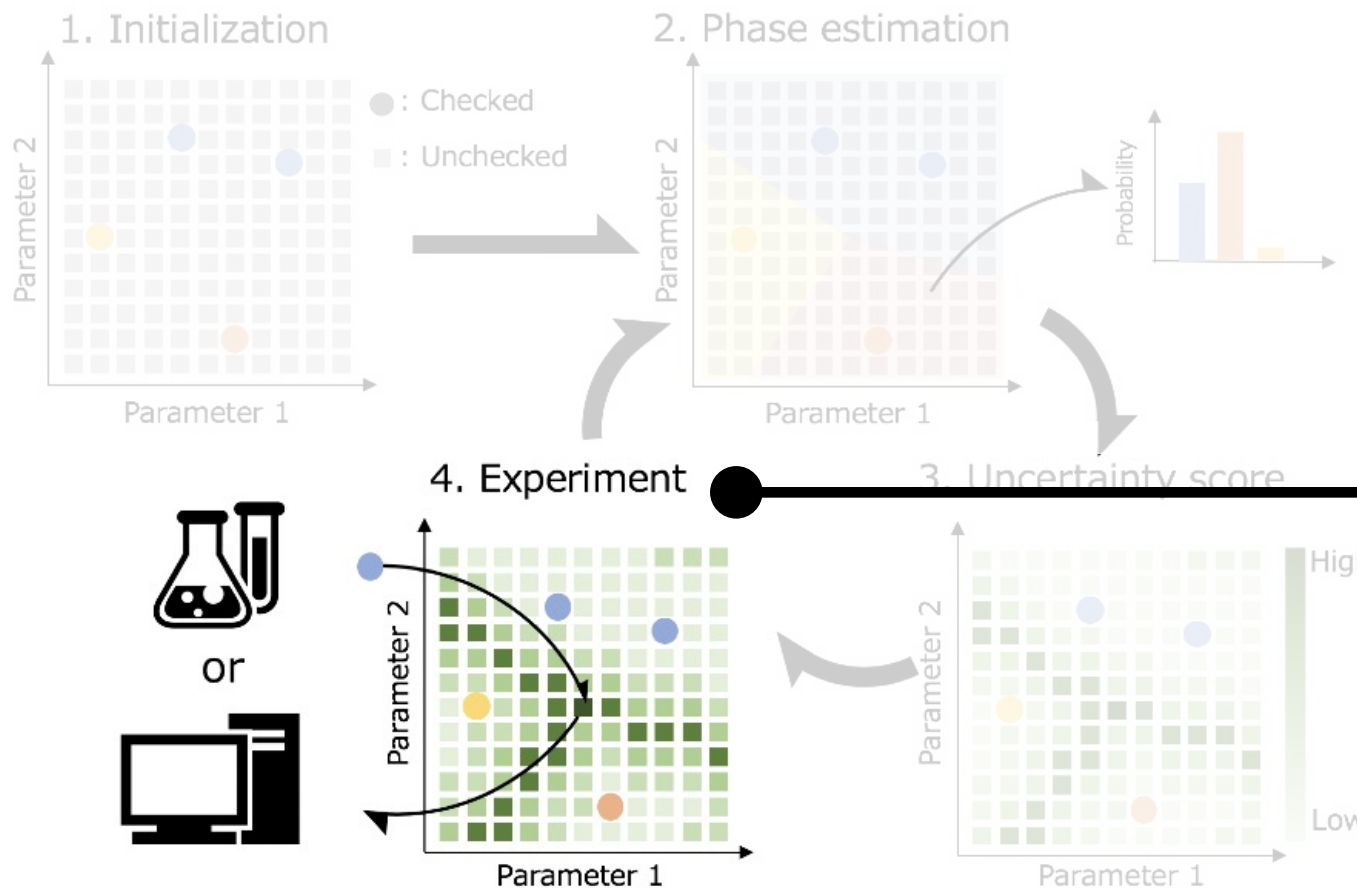


## Uncertainty scores:

- Least Confident (LC)  
$$u_{LC}(\mathbf{x}) = 1 - \max_p P(p|\mathbf{x})$$
- Margin Sampling (MS)  
$$u_{MS}(\mathbf{x}) = 1 - [P(p_1|\mathbf{x}) - P(p_2|\mathbf{x})]$$
- Entropy-based Approach (EA)  
$$u_{EA}(\mathbf{x}) = - \sum_p P(p|\mathbf{x}) \log P(p|\mathbf{x})$$

# 4. Experiments

Uncertainty scoreが最大となる点を，次に測定する候補として選び，実験や計算で，その点の相を特定する。

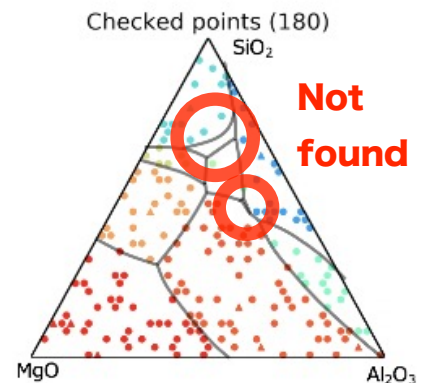
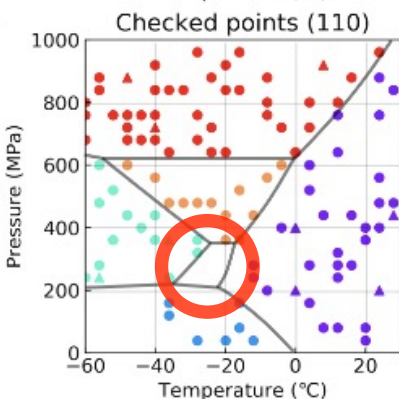
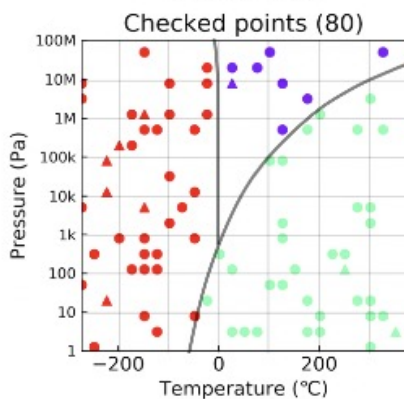
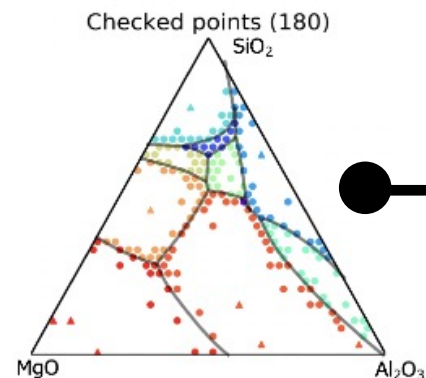
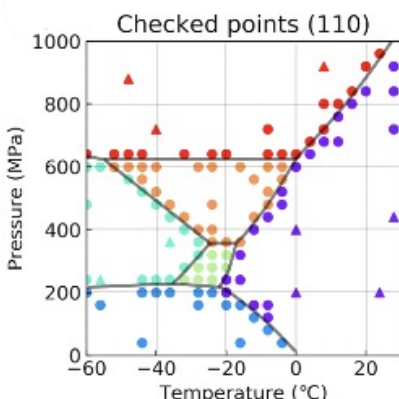
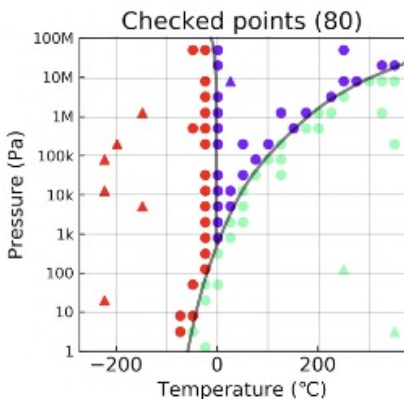
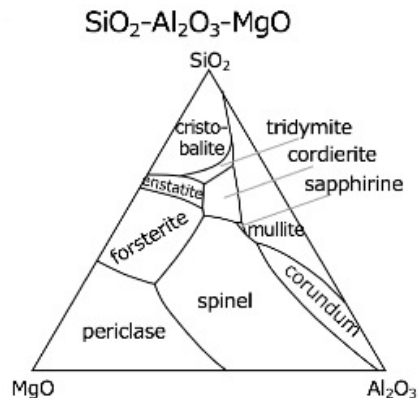
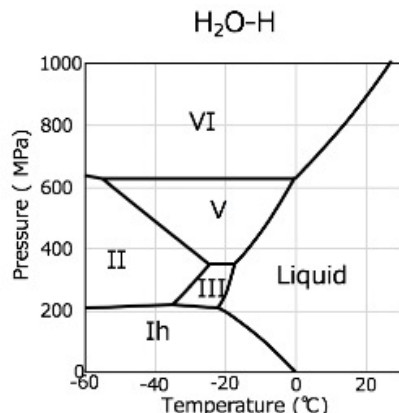
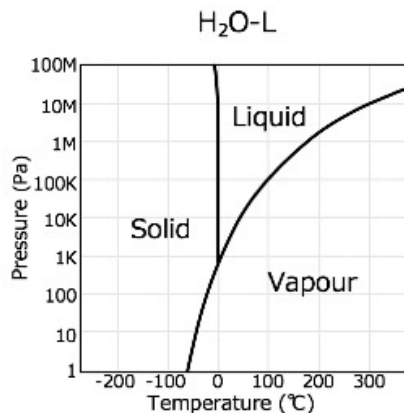


最もUncertainty scoreが高い点を選ぶ。

Uncertainty scoreの性質から，相境界の近傍や，すでに測定された点から遠く離れた点が選ばれる。

# デモンストレーション

True known PD  
 Uncertainty sampling  
 Random sampling



USでは、相境界の近傍を  
 サンプルングする。

USでは小さな相も見つける  
 ことができる。

初期点は9つ (三角)  
 からスタート

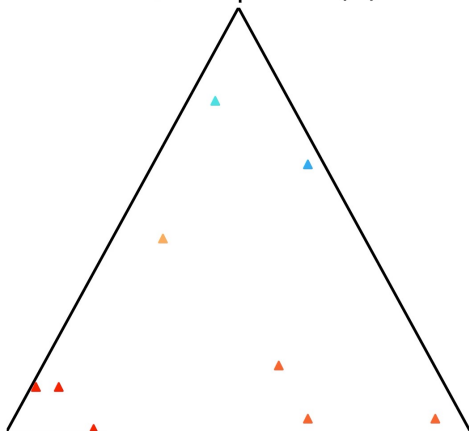


# Sampling movie

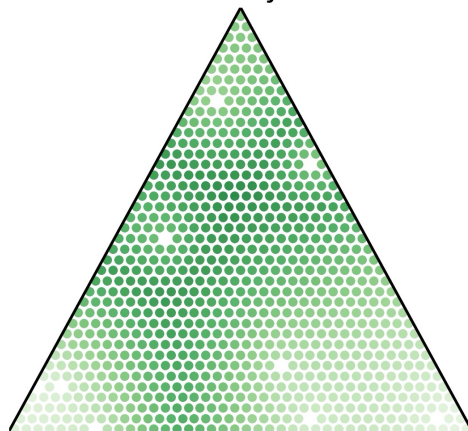
この値が1となるときに、  
予測相図と真の相図が一致。

**Uncertainty sampling** Sampling by LP+LC for  $\text{SiO}_2\text{-Al}_2\text{O}_3\text{-MgO}$

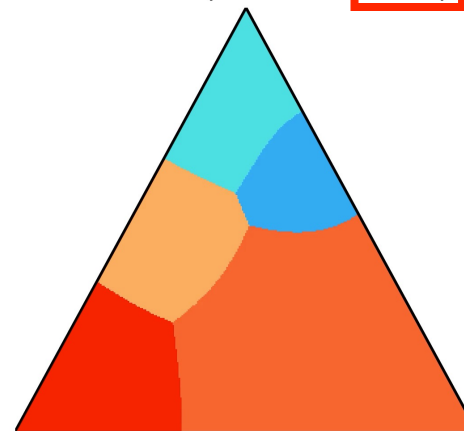
Checked points (9)



Uncertainty score

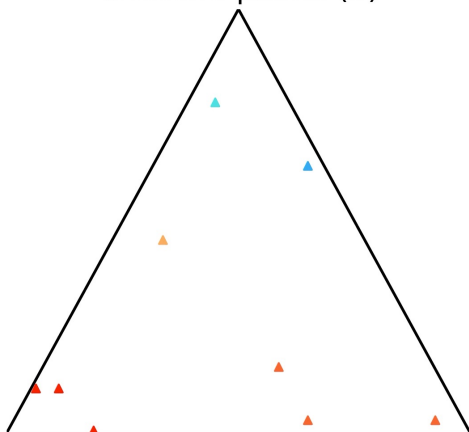


Estimated (Macro-F1) **0.38**

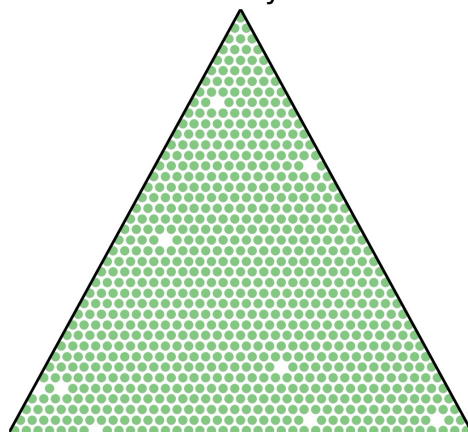


**Random sampling** Sampling by LP+RS for  $\text{SiO}_2\text{-Al}_2\text{O}_3\text{-MgO}$

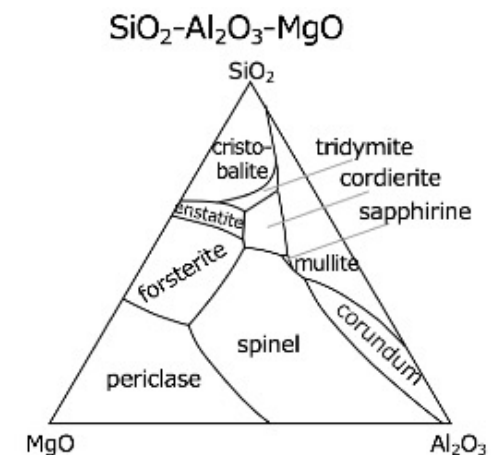
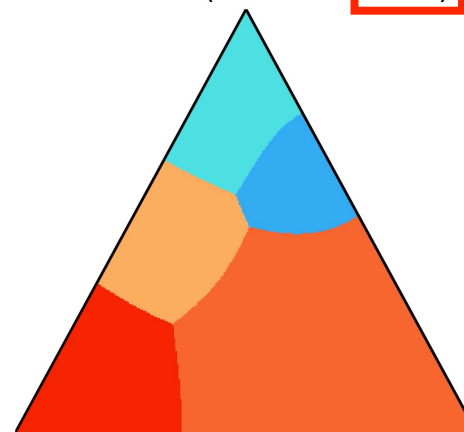
Checked points (9)



Uncertainty score

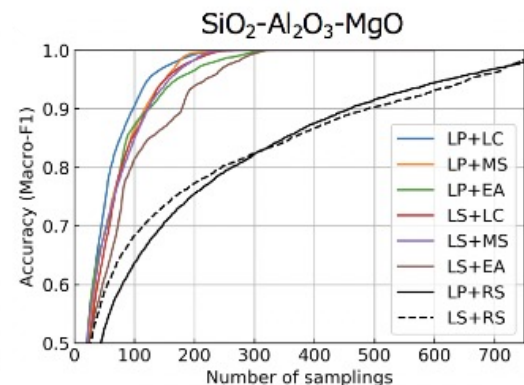
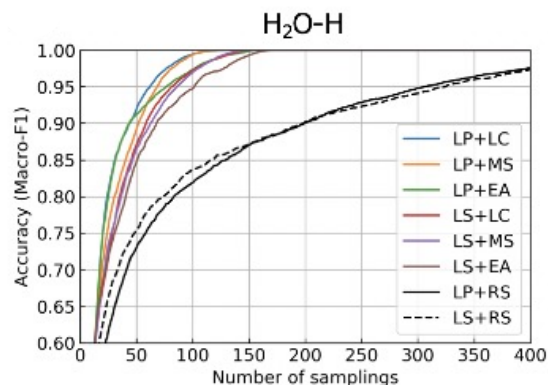
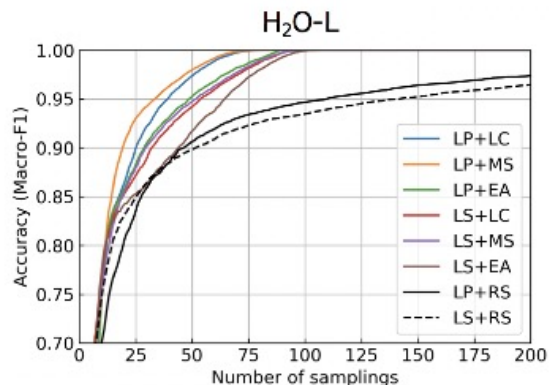


Estimated (Macro-F1) **0.38**



# 定量評価

予測された相図と真の相図がどの程度近いか (Macro-F1) を表す精度のサンプリング数依存性

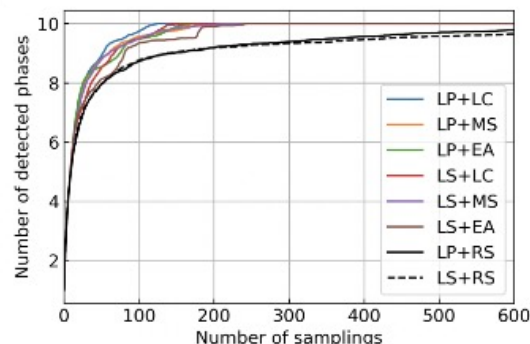
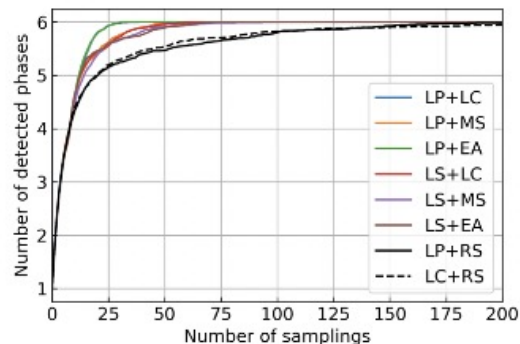
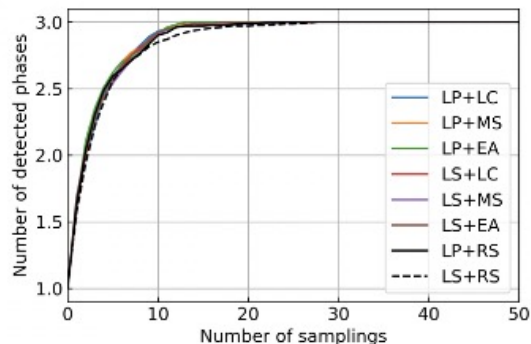


← 真の相図が得られる。  
(Macro-F1=1)

200回の独立試行による  
平均を計算。

USは正確な相図を描くためのサンプリング数を20%に減らすことができる！

見つけた相のサンプリング数依存性



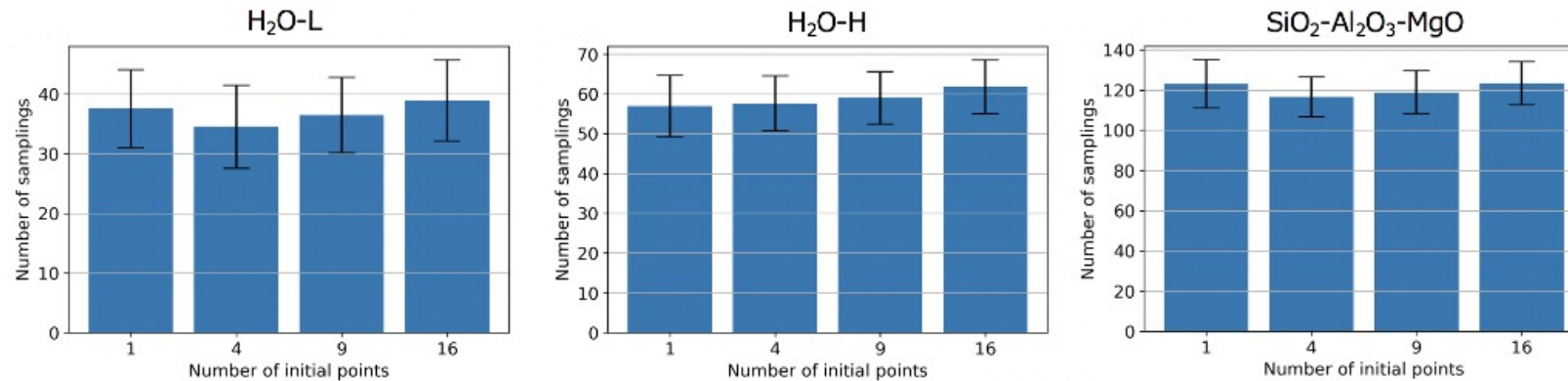
← 全ての相が見つかった。

200回の独立試行による  
平均を計算。

USは見つからない相を早く見つけることができる！

# 定量評価

初期点依存性 (Macro-F1が十分値になるまでのサンプリング数)



初期点数は4つで  
十分

初期点数は少なくてOK (何もデータがないところからでも始めやすい)

Uncertainty samplingは、  
複雑な未知の相図を初めから描く場合に効力を発揮

K. Terayama, R. Tamura, Y. Nose, H. Hiramatsu, H. Hosono, Y. Okuno, and K. Tsuda, Physical Review Materials 3, 033802 (2019).

# GitHub

Why GitHub? ▾ Enterprise Explore ▾ Marketplace Pricing ▾ Search / Sign in Sign up

tsudalab / PDC Watch 7 Star 0 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Insights

Efficient phase diagram construction based on uncertainty sampling

13 commits 1 branch 0 releases 1 contributor

Branch: master ▾ New pull request Find file Clone or download ▾

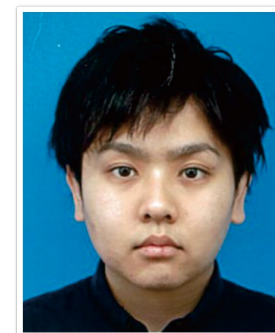
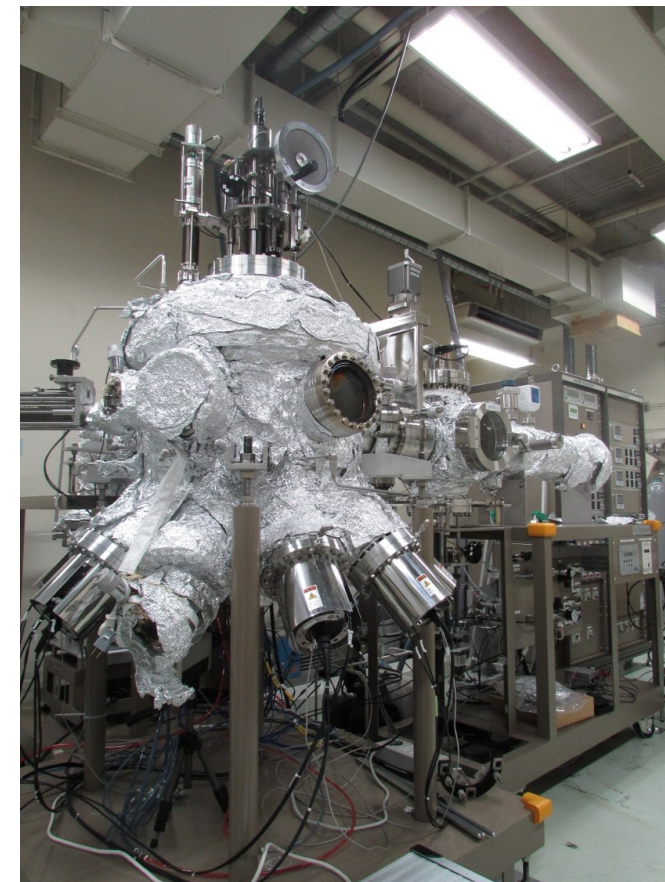
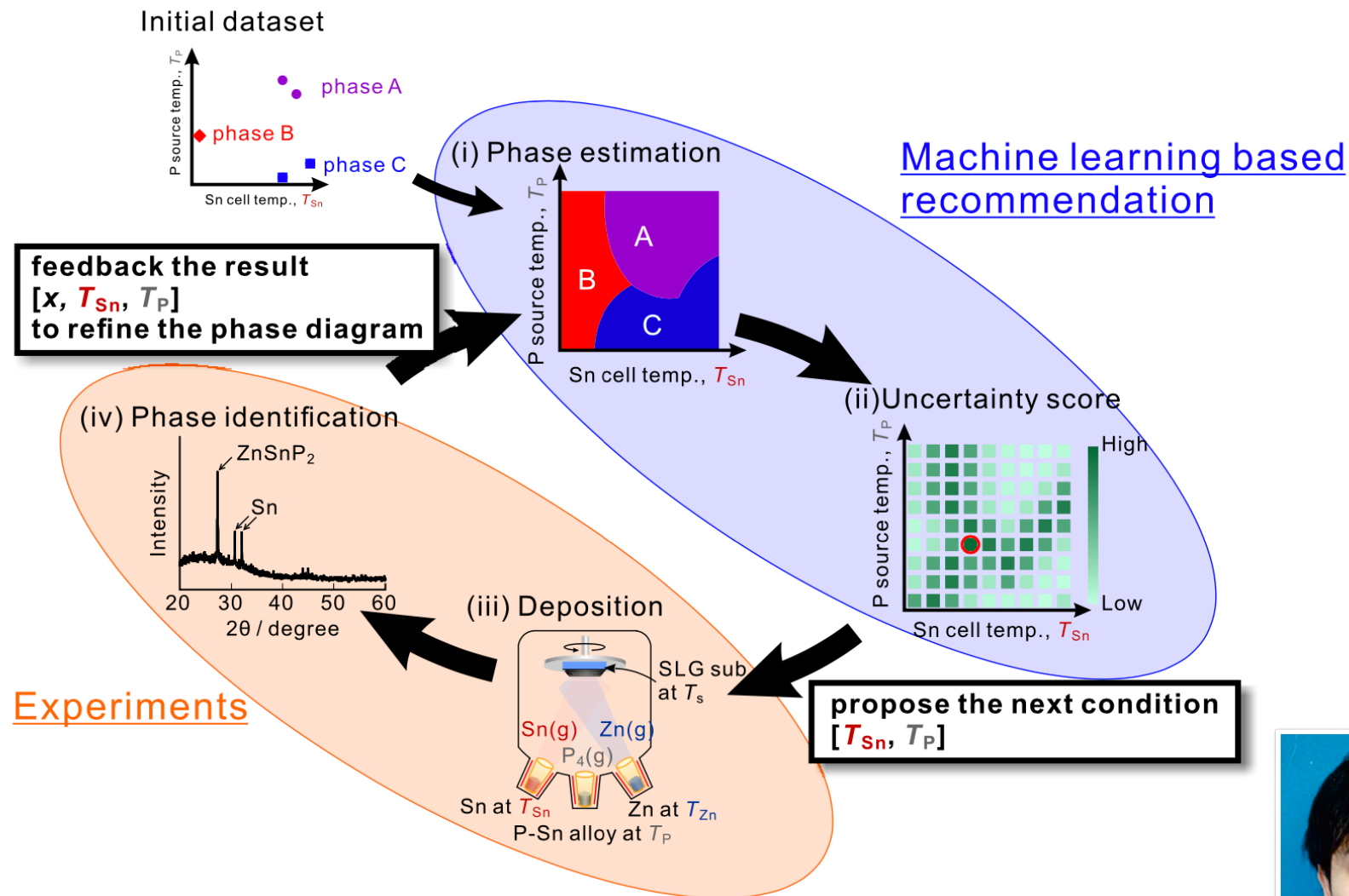
Ktera1988 Update README.md Latest commit b2dbd18 on 16 Dec 2018

PD_examples	Add files via upload	2 months ago
snapshot	Add files via upload	2 months ago
PDC_sampler.py	Add files via upload	a month ago
PDC_sampler_version0.ipynb	Add files via upload	2 months ago
README.md	Update README.md	a month ago
data.csv	Add files via upload	a month ago

<https://github.com/tsudalab/PDC>



# 実験による検証

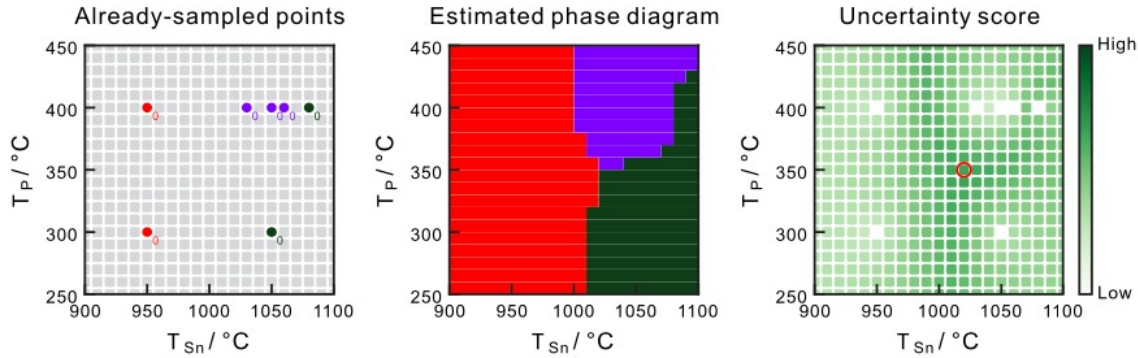


R. Katsube, K. Terayama, R. Tamura, and Y. Nose  
ACS Materials Letters 2, 571 (2020).

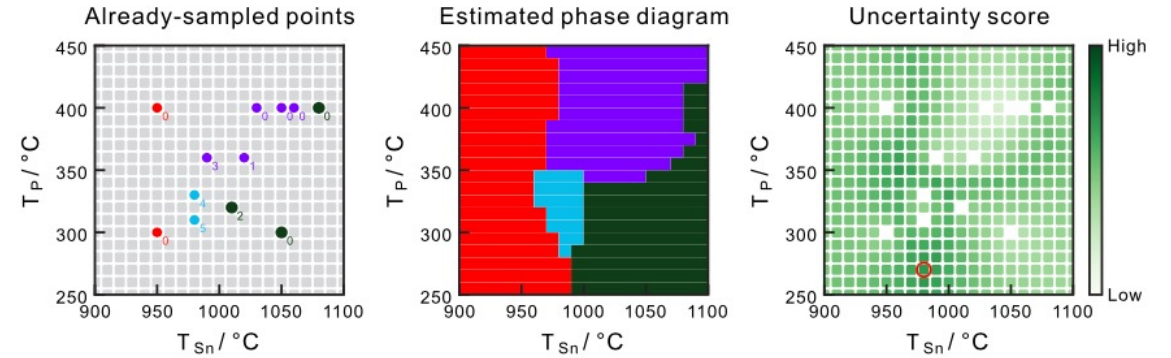


# 実験による検証

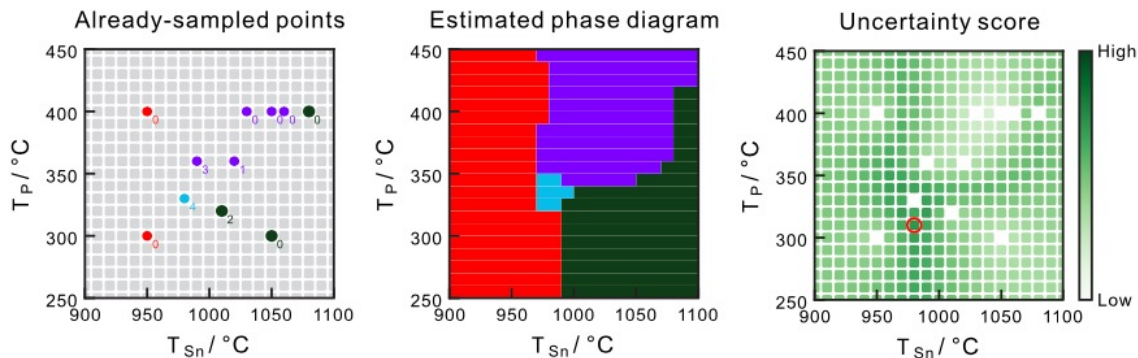
## Initial run using the initial data set



## After 5 cycles

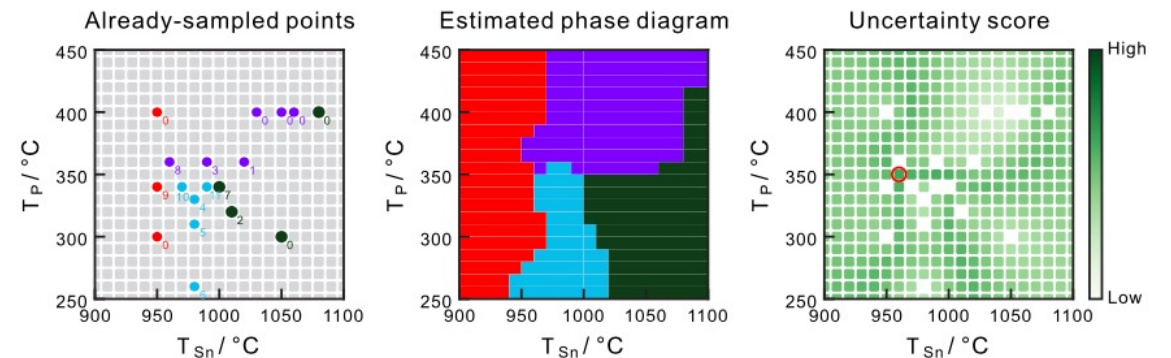


## After 4 cycles



↑  
新しい相が見つかった。  
(4 cycles by PDC)

## After 11 cycles

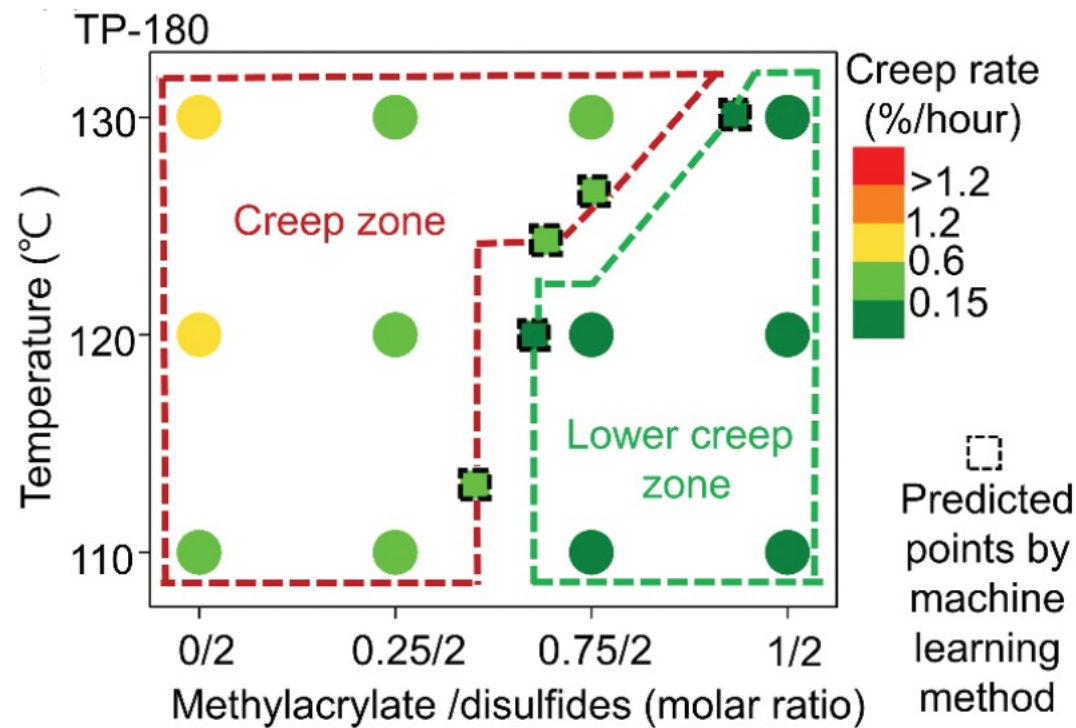
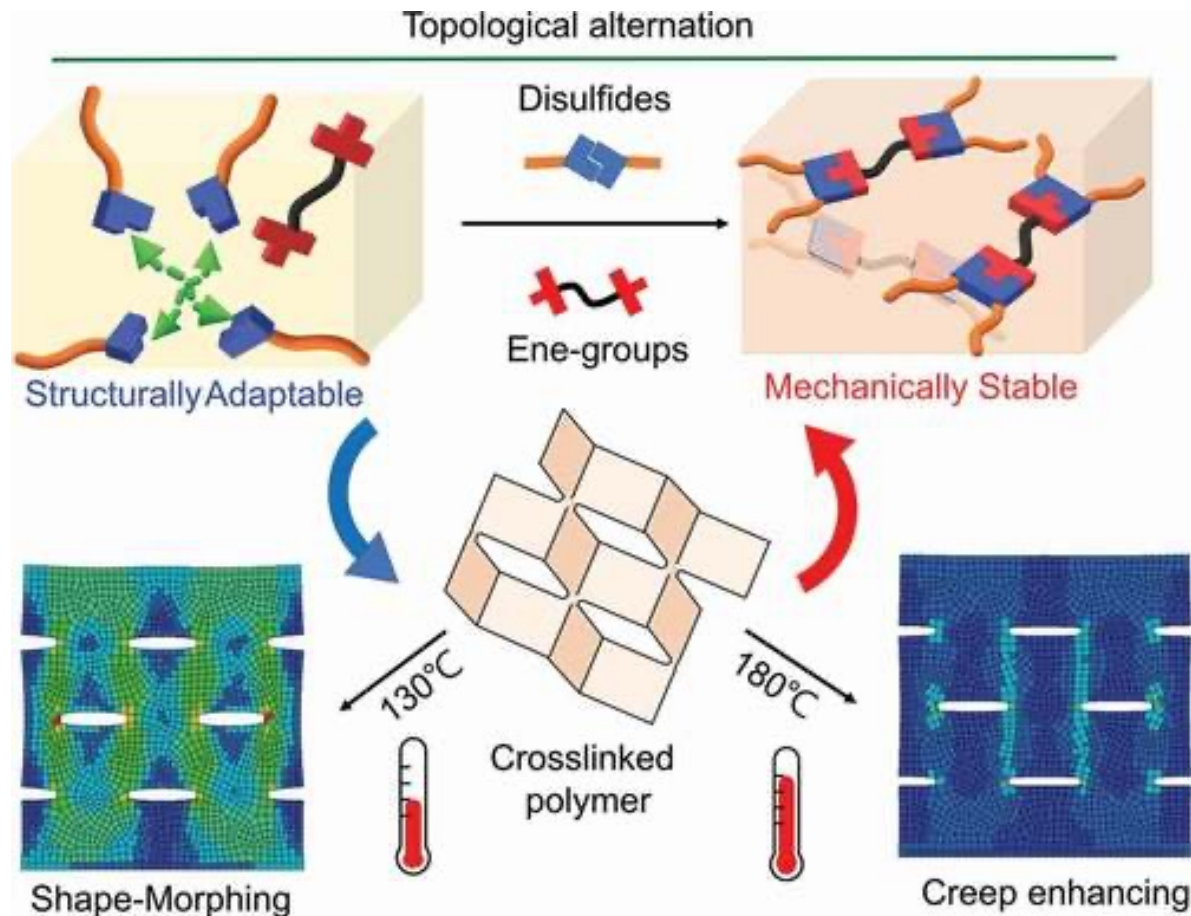


↑  
十分な精度の相図が描けた。  
(11 cycles by PDC)

Uncertainty samplingによる手法は新しい実験に対しても効力を発揮した。

# 架橋ポリマー

W. Hu, T. Chen, R. Tamura, K. Terayama, S. Wang, I. Watanabe, and M. Naito  
 Sci. Technol. Adv. Mater. 23, 66-75 (2022).



Uncertainty samplingは新しい実験に対しても効力を発揮した。



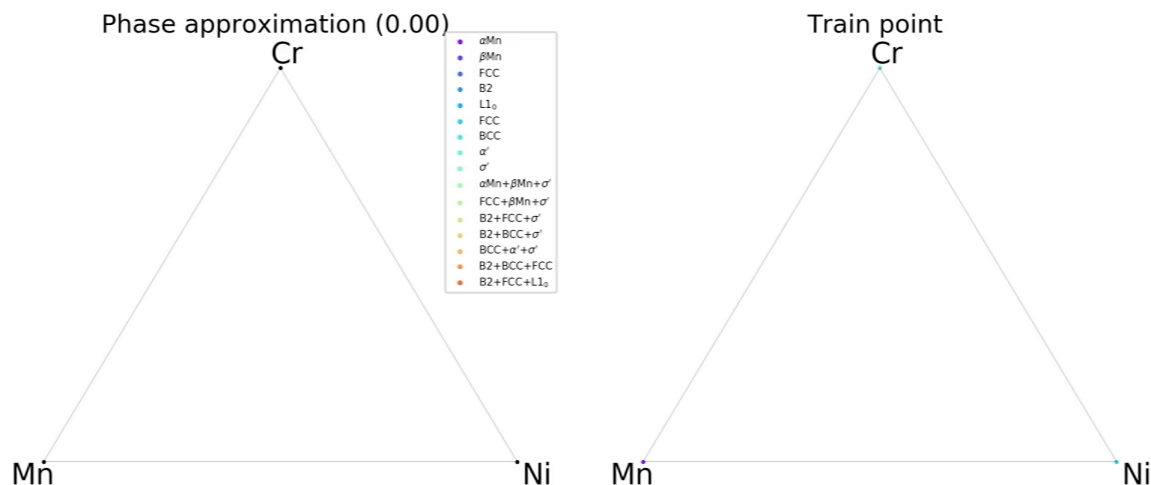
# 手法の高度化1：相図に関する知見を導入

AIを導入することで、相図作成における実験数を減らすことはできたが、  
もっと実験回数を減らしたい！

相図に関する材料研究からの知見（共存相，相律）を利用することで、  
より効率的に相図を作成できないか？

温度固定 3 元系合金相図を対象として、手法の高度化を検討

1 元系の情報しかないところからスタート



利点1：サンプリング数を10%程度に削減

利点2：3相共存相を効率的に探索



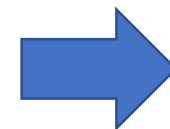
# 手法の高度化具体案 1 : 共存相の扱い

実験で共存相が見つかった際,

1, 单相の情報

2, 2相共存相の midpoint

を学習データに与えて次の実験点をAIで探索する.



1回の実験で  
多くの情報を  
AIに取り込む

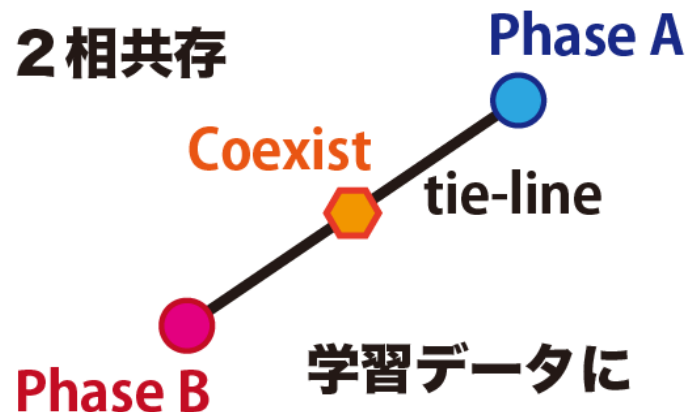
## 单相, 2相共存, 3相共存が見つかった場合の扱い

実験点が  
单相

● Phase A

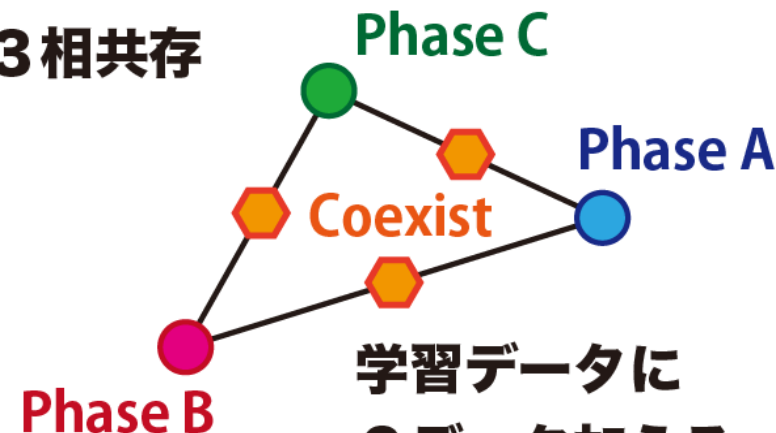
学習データに  
1データ加える

実験点が  
2相共存



学習データに  
3データ加える

実験点が  
3相共存

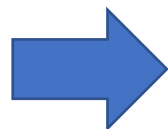


学習データに  
6データ加える

# 手法の高度化具体案2：相律の扱い

## ギブスの相律

$$\text{自由度} : F = \underbrace{C}_{\substack{\text{成分の数} \\ \text{(今回は3)}}} - \underbrace{P}_{\substack{\text{相の数}}}$$



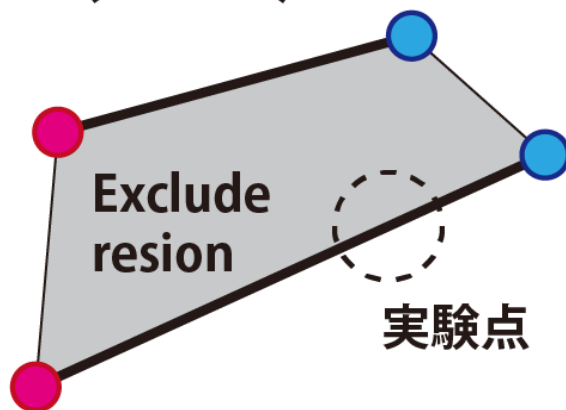
予測された相図の中で  
自由度が少ない領域  
は探索範囲から除外し  
効率的探索を実現する

単相, 2相共存, 3相共存と予測された領域に対する除外領域

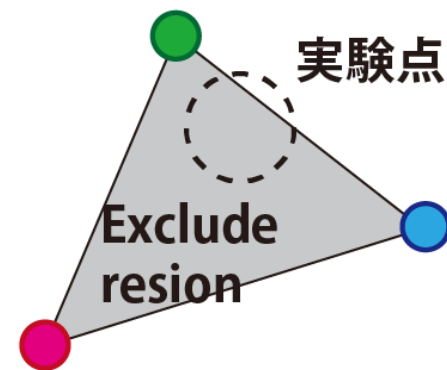
単相を除外  
( $F = 2$ )



2相共存を除外  
( $F = 1$ )

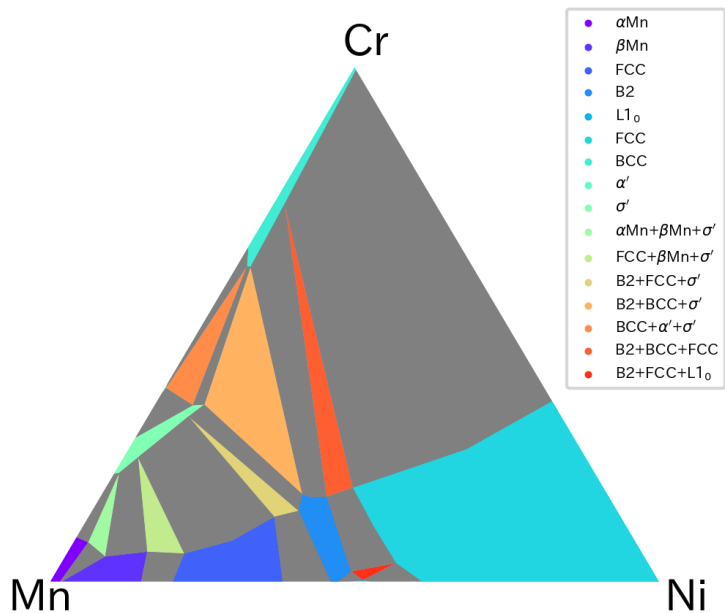


3相共存を除外  
( $F = 0$ )

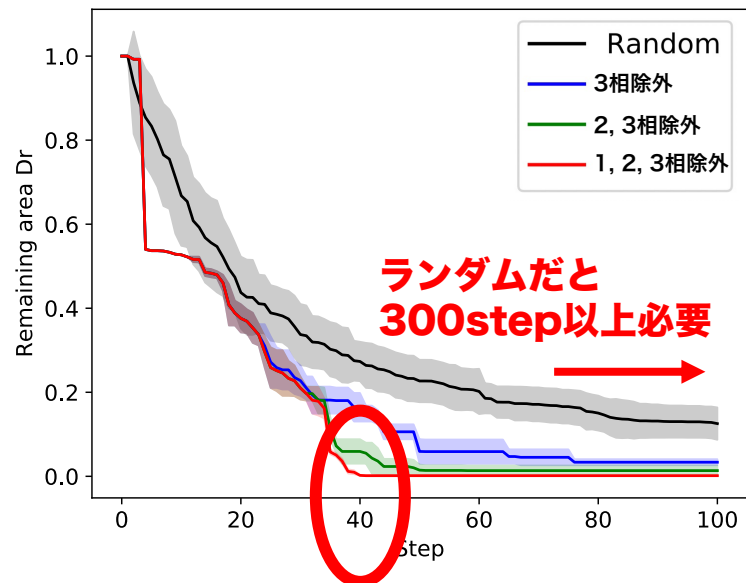




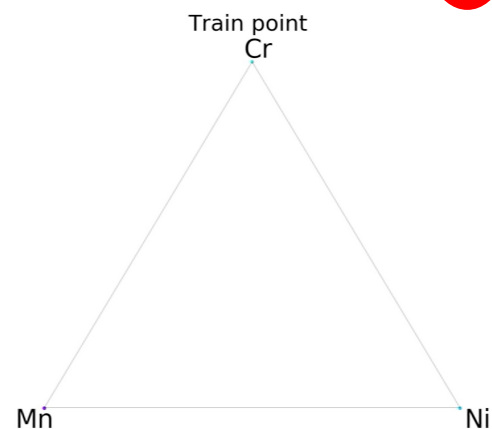
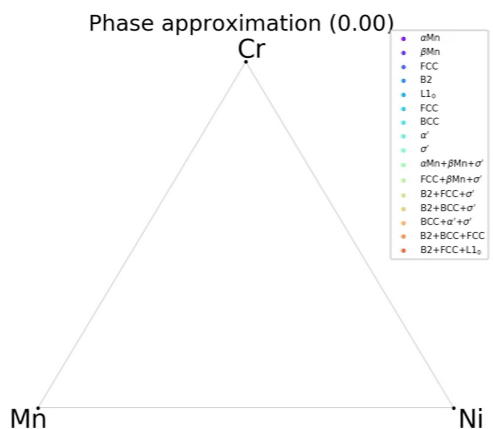
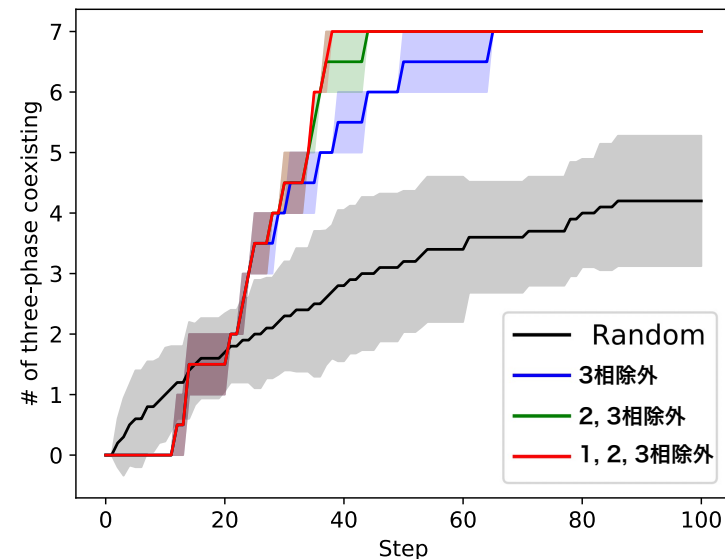
# Mn-Cr-Niを対象としたデモンストレーション



確定していない領域



見つかった3相共存の数



ランダム探索と比べて

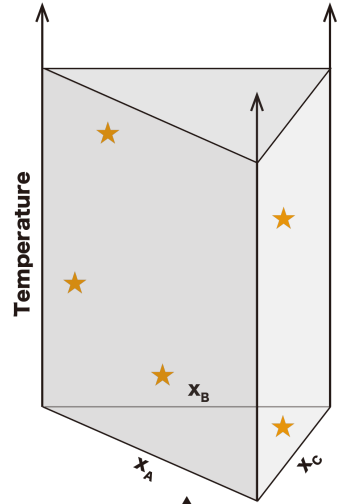
1. サンプル数を10%程度に削減
2. 3相共存相を効率的に探索

これまでのAIと比べても2倍程度高速

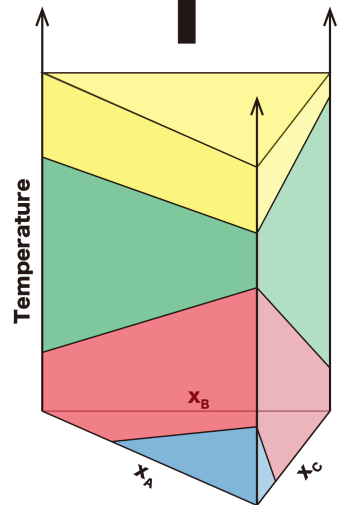
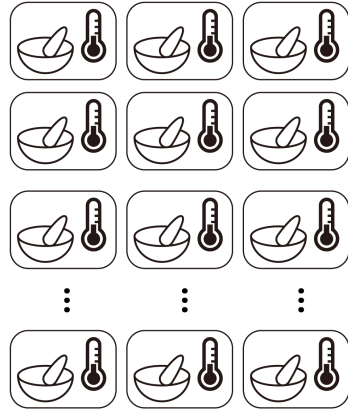
K. Terayama, K. Han, R. Katsube, I. Ohnuma, T. Abe, Y. Nose, and R. Tamura, Scripta Materialia 208, 114335 (2022).

# 手法の高度化2：並列実験

Multiple suggestions



Batch experiments

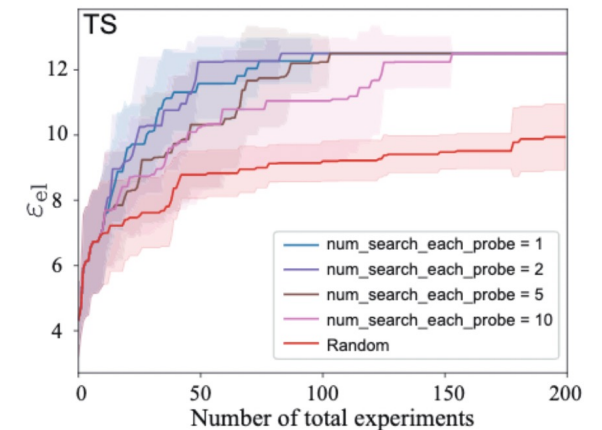
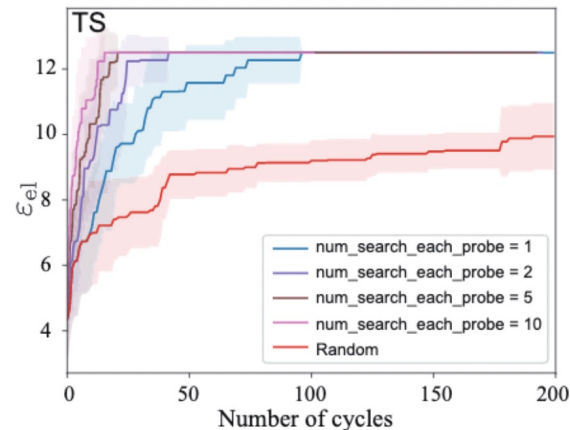


Phase estimation

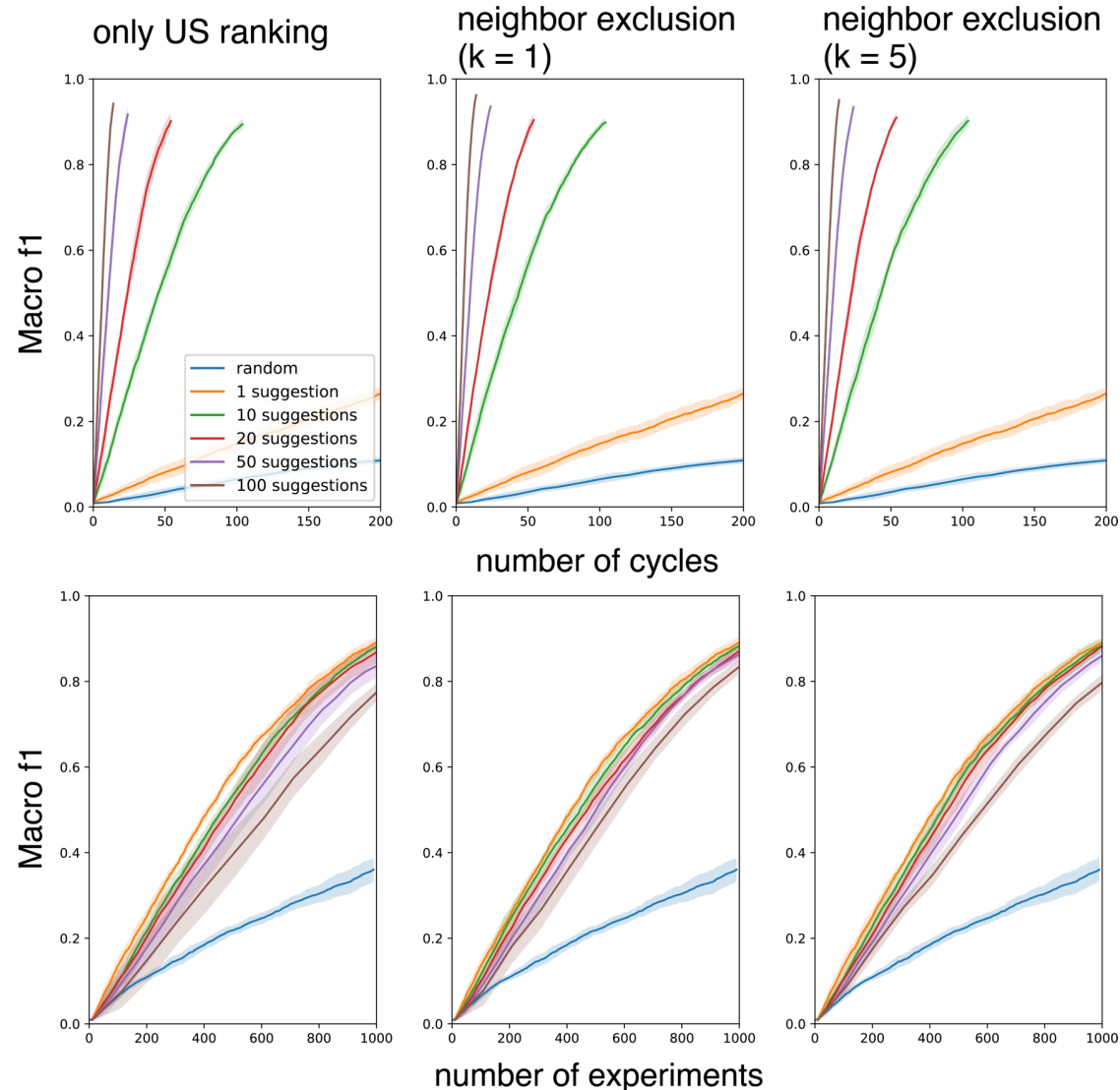
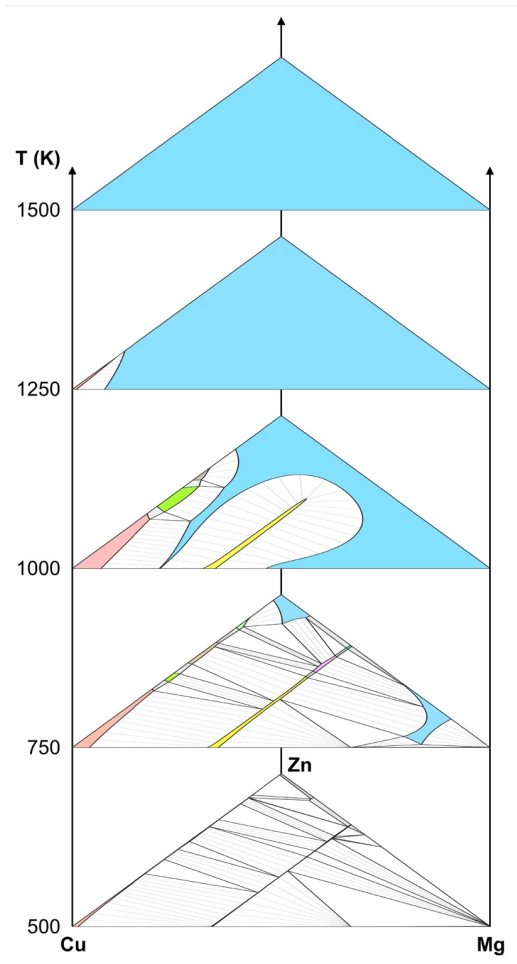
Database

1. PDCで複数提案を行った際の精度はどうか？
2. 複数提案用アルゴリズムを複数検討
3. 複数提案で相図作成の効率化が実現可能（コストは変化せず，時間が短縮可能）

## ベイズ最適化の場合



# 並列実験用PDC



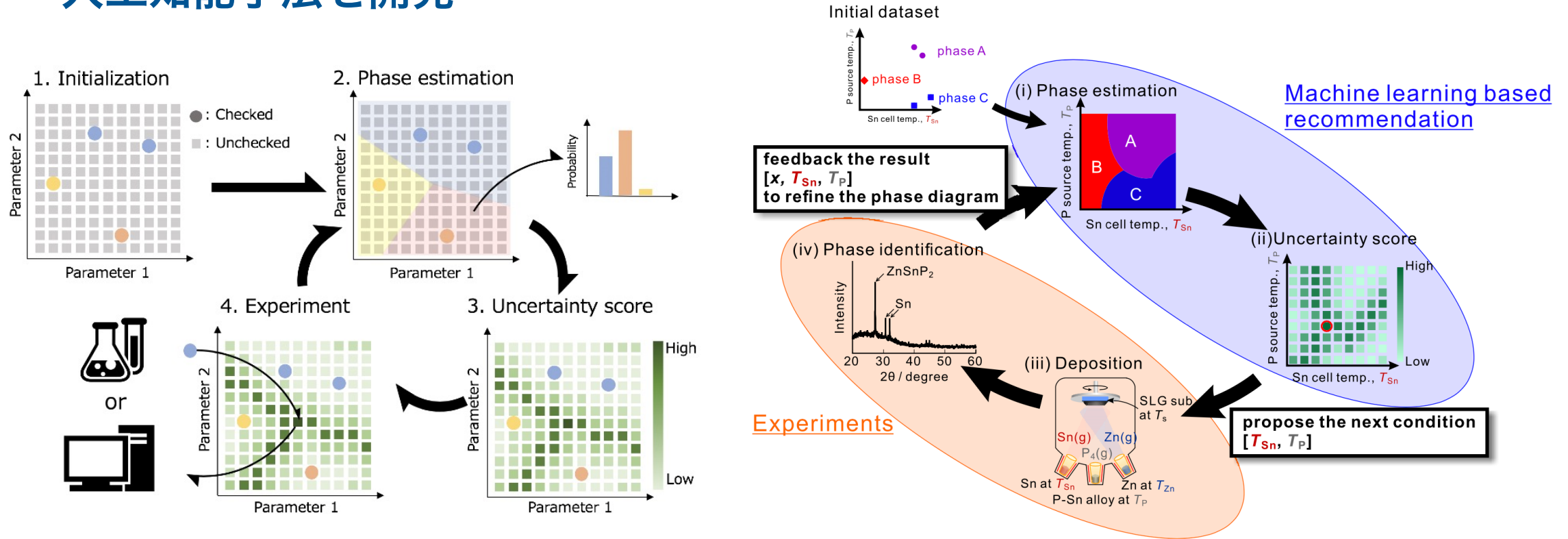
サイクル数は削減可能  
実験回数は変化せず

PDCは、  
並列実験に向けた手法  
である！

R. Tamura, G. Deffrennes, K. Han, T. Abe, H. Morito, Y. Nakamura, M. Naito, R. Katsube, Y. Nose, and K. Terayama, submitted to STAM Methods 2, 153-161 (2022).

# 人工知能による相図作成

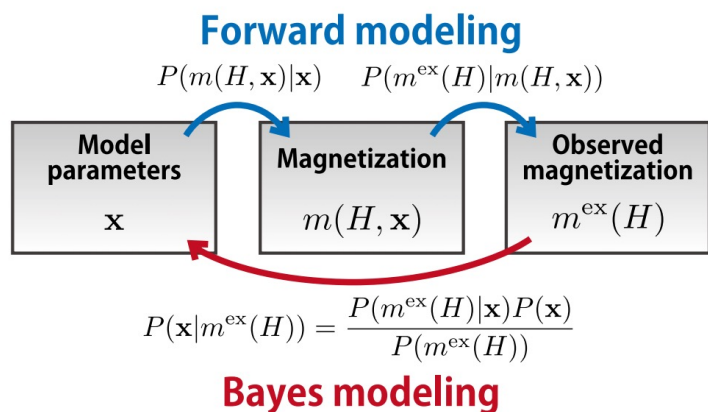
能動学習（アクティブラーニング）を使って相図作成を効率化するための人工知能手法を開発



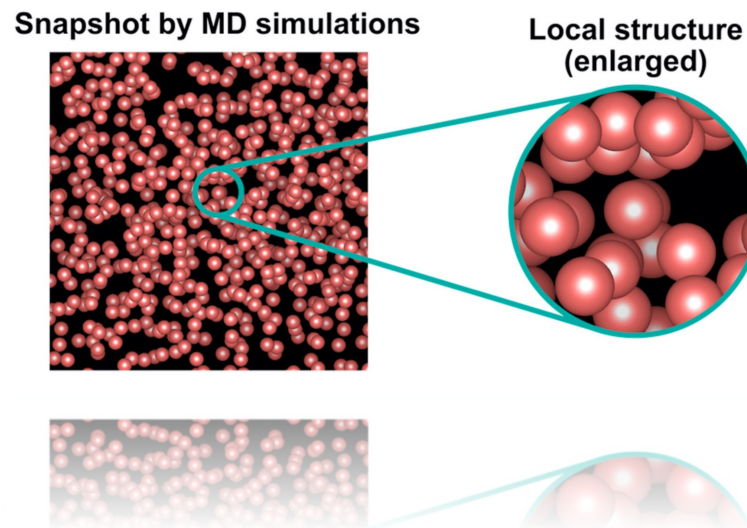
田村亮, 寺山慧, 勝部涼司, 野瀬嘉太郎 "機械学習による相図作成の効率化", (応用物理, 第91巻, 第2号, p.96, 2022).

# まとめ

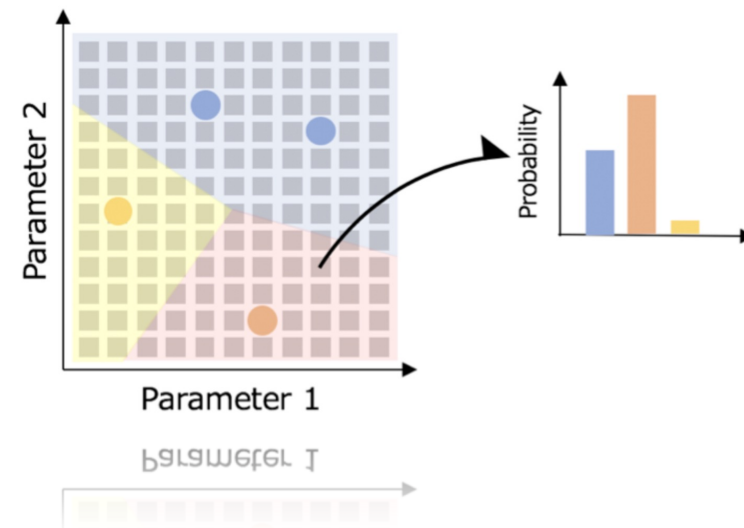
## ハミルトニアン推定



## MD解析



## 相図構築



機械学習は物理学における様々な問題を扱うことができる。

ニーズに適した手法を適用・開発することが重要であり、  
うまく利用できれば、今までできなかったことができるようになる！