

2023年5月8日

# 富岳を用いたベイズ推定と先端量子ビーム計測への応用

鳥取大学工学部, **KEK**物構研低速陽電子施設(客員), 東大物性研(客員) 星健夫

- ・ 東京大学工学部物理工学科卒業
- ・ 卒論は, 高温超電導体の実験研究. 大学院から計算物質物理学.

## 1. 計算物質物理学の潮流

(1) 計算科学のパラダイムチェンジ

(2) 密度汎関数理論(DFT)の勃興と成熟

## 2. 富岳を用いたベイズ推定と先端量子ビーム計測への応用

# 計算物質物理学の潮流(1): 計算科学のパラダイムチェンジ

## 20世紀

コンピュータの誕生  
→ 数値計算アルゴリズムの整備

パラダイムチェンジ  
(古いプログラムは「死ぬ」?)

## 21世紀

- ・超並列コンピュータ(「京」「富岳」など)
- ・メニーコアCPU
- ・GPU搭載機の躍進(AI需要)
- ・量子計算機への期待

新しい数値手法による学際研究の可能性

### 参考: 20世紀のトップ10アルゴリズム

(F. Sullivan and J. Dongarra, Comp. Sci. Eng. 2, 22 (2000))

- ・ The Metropolis Algorithm (1946)
- ・ Simplex Method (1947)
- ・ Krylov Subspace Method (1950)
- ・ The Decompositional Approach to Matrix Computations (1951)
- ・ The Fortran Optimizing Compiler (1957)
- ・ QR Algorithm (1959)
- ・ Quicksort (1962)
- ・ Fast Fourier Transform (1965)
- ・ Integer Relation Detection (1977)
- ・ Fast Multipole Method (1987)

解説書籍「20世紀のトップ10アルゴリズム」, 2022年→



# 計算物質物理学の潮流(1): 計算科学のパラダイムチェンジ

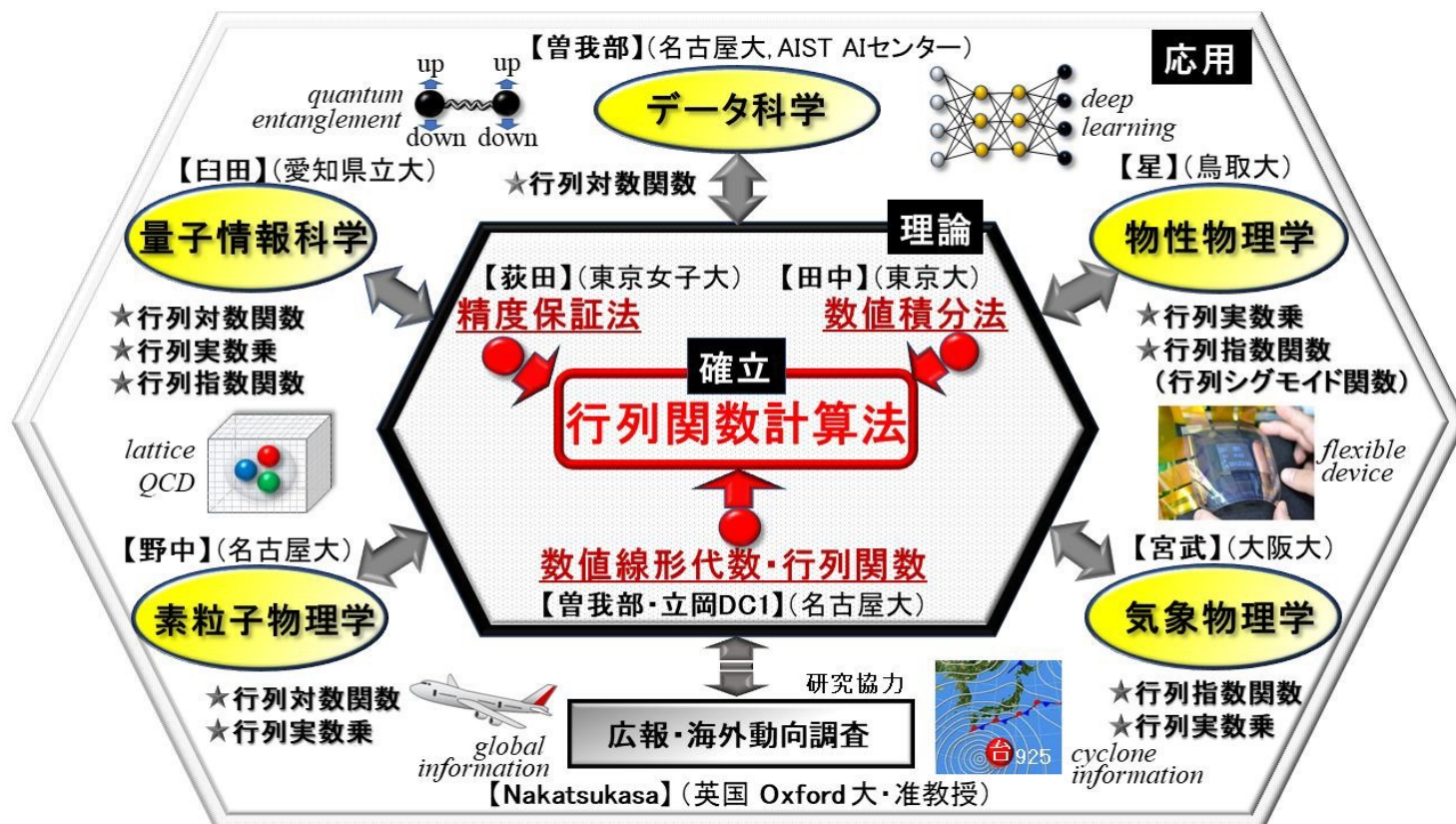
20世紀

21世紀

数理による学際研究例: 科研費基盤A「物理学・情報科学に共通する大規模行列関数の総合的数値計算法の創成」代表: 曾我部知広(名古屋大)

超並列コンピュータ(「京」「富岳」など)  
 一コアCPU  
 搭載機の躍進(AI需要)  
 計算機への期待

数理手法による学際研究の可能性



「...」, 2022年→

# 計算物質物理学の潮流(2): 密度汎関数理論(DFT)の勃興と成熟

密度汎関数理論(Density Functional Theory, 1960-; プラズマ)  
電子密度分布 $\rho(\mathbf{r})$ の汎関数としてエネルギー $E[\rho]$ を定義.



Walter Kohn (1923-2016)  
(1998年ノーベル賞)

標準的プログラムの確立(1990年代~2000年代)

例: VASP, Quantum Espresso

計算データライブラリ整備(インフォマティクス)

例: 米国Materials Project (<https://materialsproject.org/>)

物質科学のインフラに(実験家も使うツールに)

未踏領域へのチャレンジ

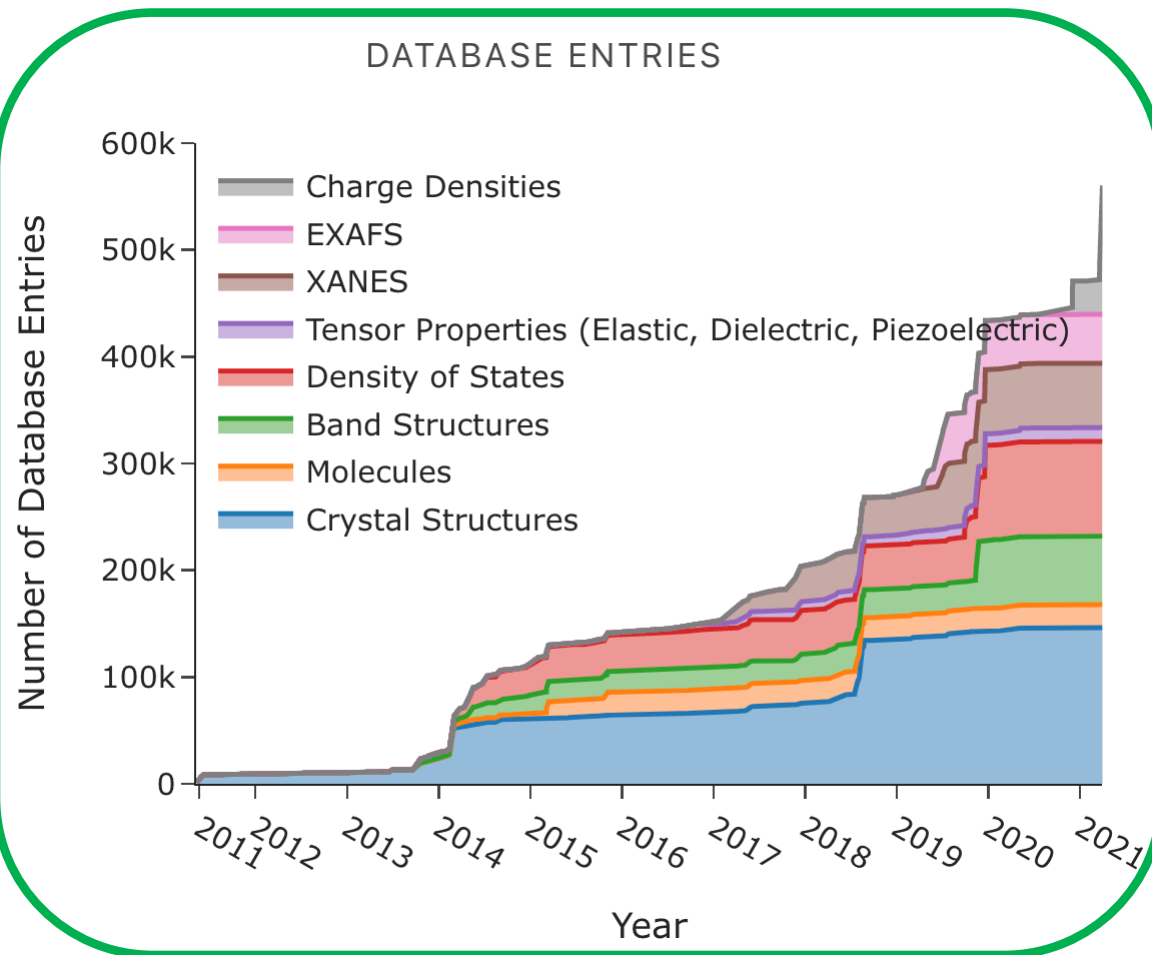
例: 大規模系(>10<sup>3</sup>原子系).

出口志向のマルチスケール理論(例: デバイス)

高密度高温系,

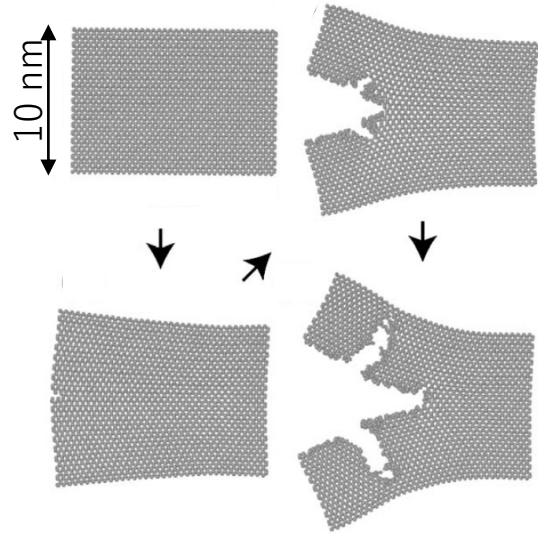
非平衡・非定常状態,

量子ビーム+物質(→計測, 加工)

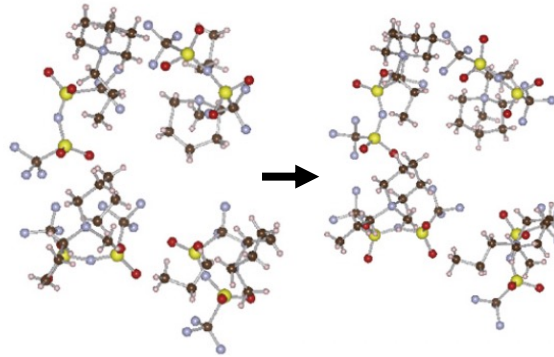


# 我々の研究例: 大規模計算むけ数理手法開発と電子状態計算への応用

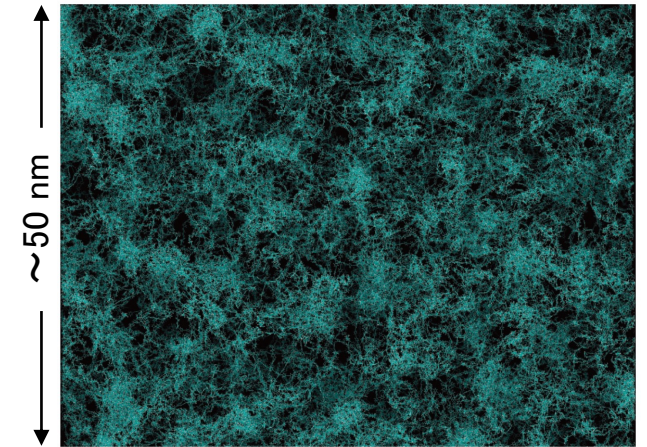
(a) シリコン脆性破壊



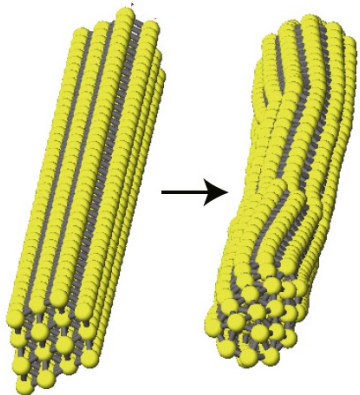
(c) イオン液体 (電池研究)  
(トヨタとの産学連携)



(e) 有機(高)分子集合体  
(フレキシブルデバイス)  
(→住友化学との産学連携)

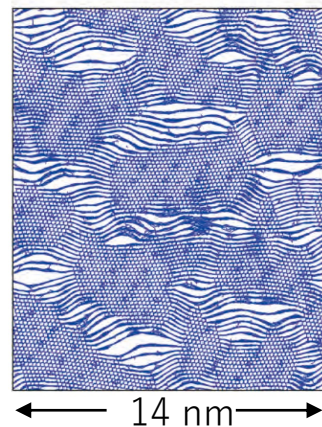


(b) らせん型金ナノワイヤ

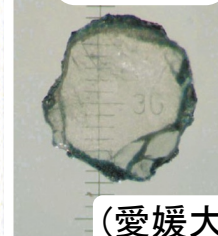


エレクトロニクス, 電子ビーム

(d) ナノ複合カーボン

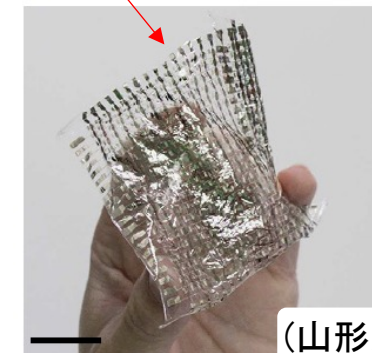


ナノ多結晶  
ダイヤモンド(\*1)  
(~0.1mm)



(愛媛大)

(\*1) Irifune, *et al.*, Nature 421, 599 (2003).



(山形大)

フレキシブルデバイス例(\*2)

(\*2) Fukuda, *et al.*, Nature Comm. 5, 4147 (2014)

# 富岳を用いたベイズ推定と先端量子ビーム計測への応用

鳥取大学工学部, **KEK**物構研低速陽電子施設(客員), 東大物性研(客員) 星健夫

## 目次

- ・ 計測データ解析フレームワーク「2DMAT」
- ・ 超並列モンテカルロ法によるベイズ推定
- ・ 全反射高速陽電子回折(TRHEPD)への応用
- ・ まとめ

謝辞：和田 健・望月出海

(KEK物構研低速陽電子),  
吉見一慶・本山裕一(東大物性研),  
福島孝治(東大総合文化),  
花田貴(東北大金研),  
高草木達・朝倉清高(北大触媒研),  
岩本晴道, 木下直希, 中野陽斗,  
寺地雄真(鳥取大院工; 学生)

(\*)謝辞

- ・ 2DMAT(<https://www.pasums.issp.u-tokyo.ac.jp/2dmat>)は,  
「2020・2021年度 東大物性研 ソフトウェア開発・高度化プロジェクト」など  
の支援を受け開発・公開され, ISSPスーパーコンピュータにプリインストールされている。
- ・ 2DMATの一部データは, ISSP data repositoryにおける以下ディレクトリにアップロードされている;  
2DMAT Gallery (<https://datarepo.mdcl.issp.u-tokyo.ac.jp/repo/17>)

## 基礎：計測データ解析におけるベイズ推定

- 計測データ  $D \equiv (D_1, \dots, D_m)$  が与えられた時に、  
目的状態量(欲しい量)  $X \equiv (X_1, \dots, X_n)$  の確率分布 (ベイズ事後確率分布) を、  
ベイズの定理 (下記) に基づいて得る

$$P(X|D) = \frac{P(D|X)P(X)}{P(D)}$$

- 例：病気Aには検査Bがあるが、誤判定する場合もある。

今、検査Bを3回行ったところ、3回とも陽性と診断された。

計測データ  $D$

実際に、病気Aにかかっている確率は、何パーセントか？

(計算に必要な情報：病気Aの患者の割合は、日本国民の10%。

検査Bは、病気Aに罹患している人のうち、30%を陰性と誤判定する。

検査Bは、病気Aに罹患していない人のうち、20%を陽性と誤判定する)

# データ駆動科学の立場から開発された、 新しい計測データ解析フレームワーク「2DMAT」

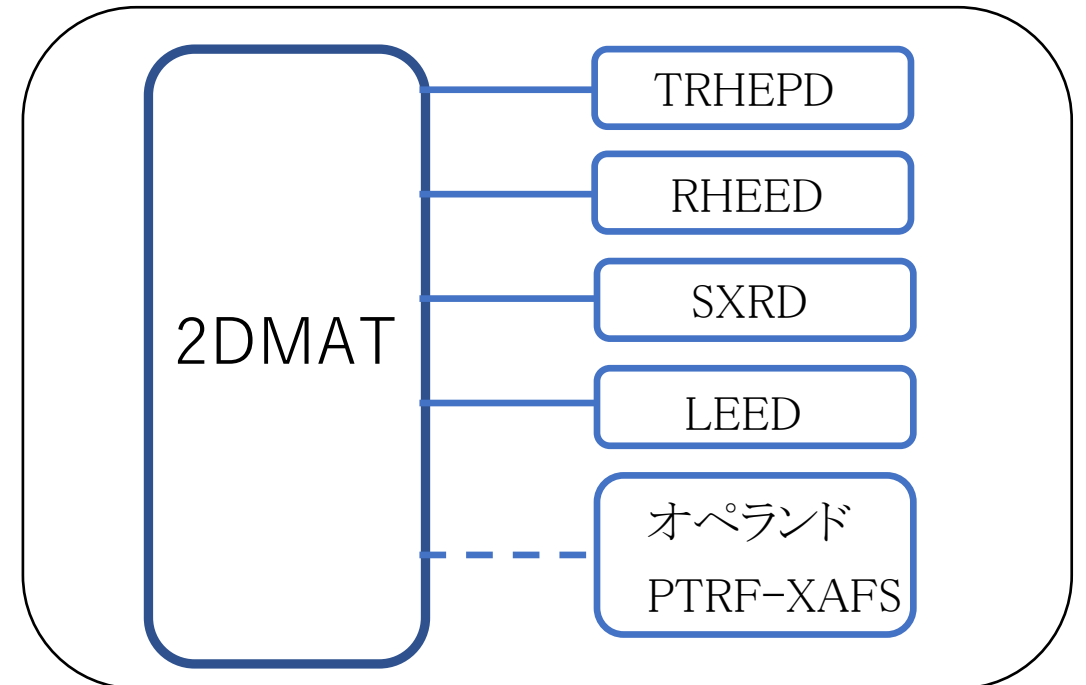
<https://www.pasums.issp.u-tokyo.ac.jp/2DMAT/>; Motoyama, *et al.*, Comp. Phys. Commun. 280, 108465 (2022)

特徴1: 高信頼な計測データ解析を, 高速に実行

- 「見えないものを, 見る」(測定限界突破)  
→ 不確かさ・ノイズに隠された「真の情報」を抽出.
- 人的試行錯誤を伴わない  
→ 大域解析(解のinitial guessは, 不要)
- 並列(「手分け」できる)アルゴリズム(PCからスパコンまで)

特徴2 : 汎用性:さまざまな計測手法に対応

- 当座対象:2次元物質むけ計測
  - (1) 全反射高速陽電子回折(TRHEPD, トレプト)(本発表)
  - (2) 反射高速電子回折(RHEED)
  - (3) 表面エックス線回折(SXRD)
  - (4) 低速電子回折(LEED)
  - (5) (予定) オペランド偏光全反射蛍光(PTRF)  
X線吸収微細構造(XAFS)(高草木・朝倉(北大触媒研))
- 他実験にも対応可.





# データ駆動科学の立場から開発された、 新しい計測データ解析フレームワーク「2DMAT」

<https://www.pasums.issp.u-tokyo.ac.jp/2DMAT/>; Motoyama, *et al.*, Comp. Phys. Commun. 280, 108465 (2022)

## ◎ 大域型逆問題解析

- ・ 内容：測定量  $D \equiv (D_1, \dots, D_m)$  (観測データ) から  
目的状態量 (欲しい量)  $X \equiv (X_1, \dots, X_n)$  を得る.

### ・ 解析原理：

- 1) 物理学により，測定量  $D$  を目的量  $X$  の関数 (順問題ルーチン) で与える

$$D_{\text{cal}} \equiv D_{\text{cal}}(X) \quad (1)$$

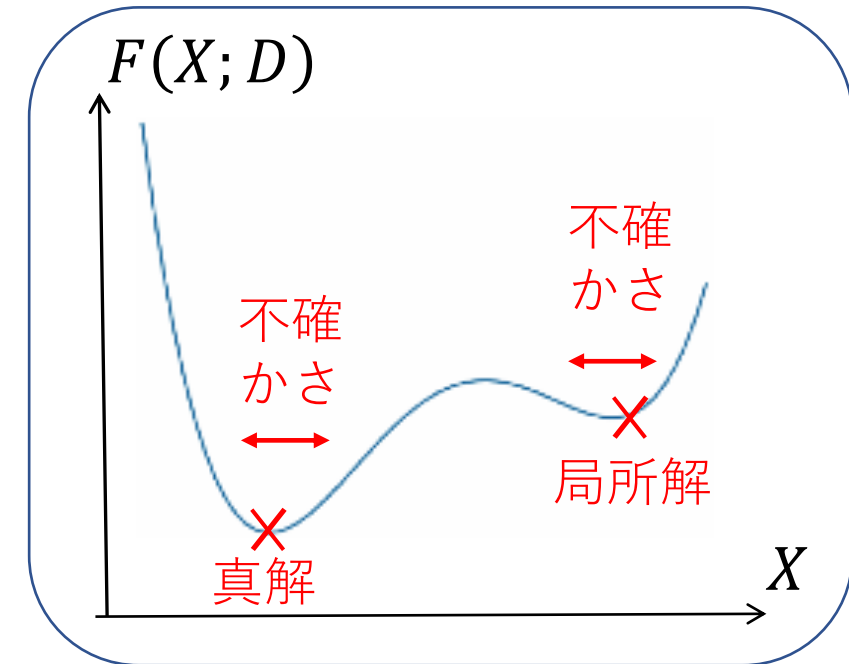
→ 「順問題ルーチン」を取り替えることで様々な実験に対応

- 2) 計算値  $D_{\text{cal}}(X)$  と実測値  $D$  の残差を目的関数  $F(X; D)$  とする.

例えば， $F(X; D) \equiv \sum_i (D_i - D_{\text{cal},i}(X))^2$  (2)

→ 最適化・不確かさ解析を実行

図 目的関数  $F(X; D)$  の概念図



# 対象実験:全反射高速陽電子回折(Total-reflection high-energy positron diffraction; TRHEPD, トレプト)

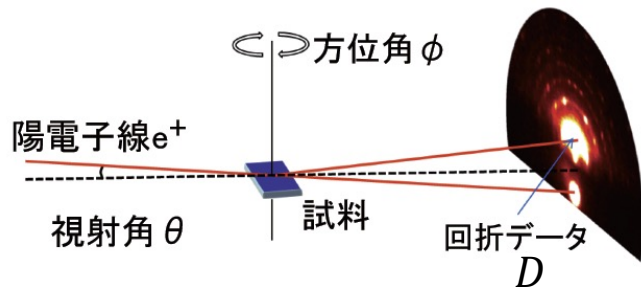
総説 : Fukaya, *et al.*, J. Phys. D 52, 013002 (2019)

解析例 : [1] Hoshi, *et al.*, Comp. Phys. Comm. 271, 108186 (2022)

- 2次元の構造非破壊計測の革新技术法 (図1), KEKで実用的測定可(全国共同利用).  
反射高速電子回折(Reflection high-energy positron diffraction; RHEED)の陽電子版  
陽電子を用いることで, 表面領域 (厚さ < 1nm)のみをプローブできる.
- 回折データ (各スポットの強度) から表面原子配列を決定する

## 実験装置模式図

(<https://www2.kek.jp/imss/spf/>)

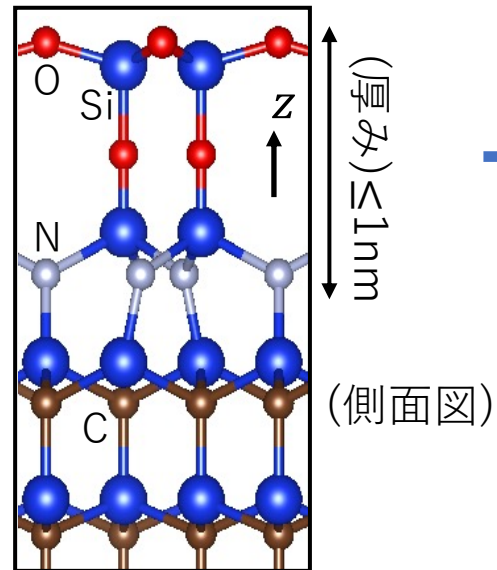


逆問題  
解析

( $D \Rightarrow X$ )

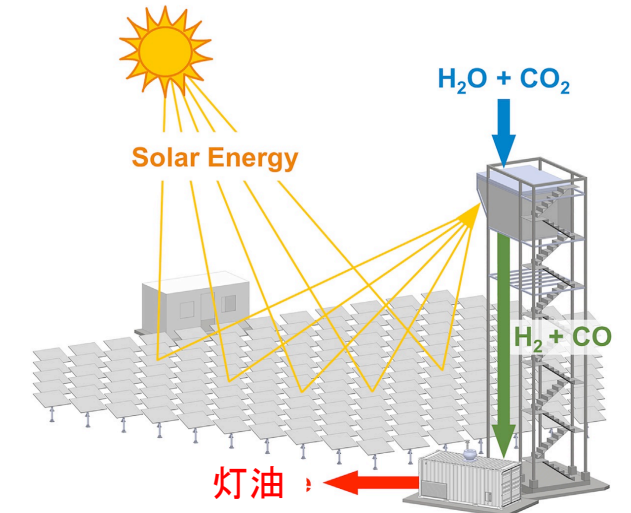
2DMAT  
(解析ソフト)

2次元物質(表面) 構造  
(原子座標) Xの決定例



出口の1つ:

カーボンニュートラル向け触媒開発  
例:  $\text{CO}_2 + \text{水} + \text{太陽光}$  で灯油生成



Zoller *et al.*, Joule 6, 1606 (2022)

新規 2次元(極薄膜)半導体:

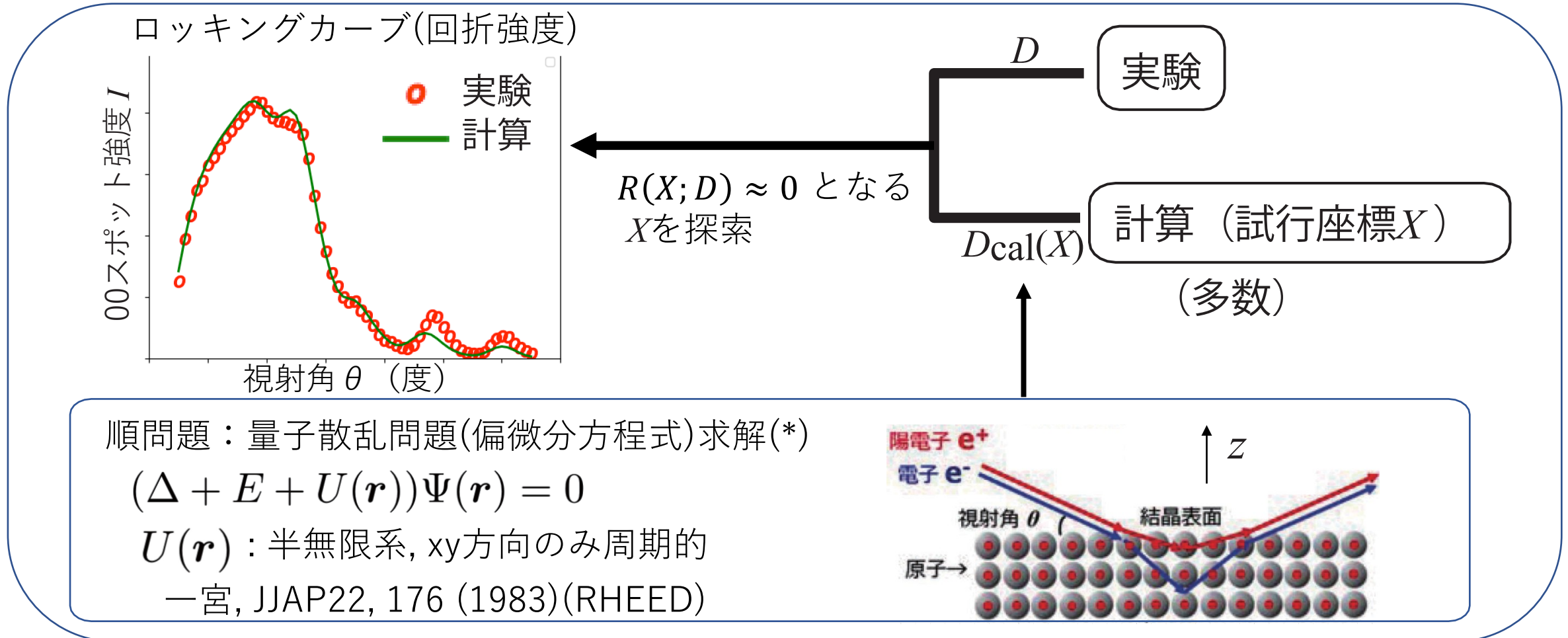
$\text{Si}_4\text{O}_5\text{N}_3 / 6\text{H-SiC}(0001) - (\sqrt{3} \times \sqrt{3}) \text{R}30^\circ$

# TRHEPDにおけるデータ解析

例：[1] Hoshi, *et al*, Comp. Phys. Comm. 271, 108186 (2022)

解析原理：測定量（回折強度）における，実験データ  $D$  と計算データ  $D_{\text{cal}}(X)$  の数値的一致  
→ 「R因子（残差）」  $R(X; D) \equiv |D_{\text{cal}}(X) - D|$  の2乗が，目的関数 ( $F(X; D) \equiv R(X; D)^2$ )

図 TRHEPDにおけるデータ解析の概要図



(\*)利用ルーチン：SIM-TRHEPD-RHEED (<https://github.com/sim-trhepd-rheed/>; Hanada, *et al*, Comp. Phys. Commun. 277, 108371 (2022))

# データ駆動科学の立場から開発された、 新しい測定データ解析フレームワーク「2DMAT」

<https://www.pasums.issp.u-tokyo.ac.jp/2DMAT/>; Motoyama, *et al.*, Comp. Phys. Commun. 280, 108465 (2022)

## ◎ 大域型逆問題解析

- ・ 内容：測定量  $D \equiv (D_1, \dots, D_m)$  (観測データ) から  
目的状態量(欲しい量)  $X \equiv (X_1, \dots, X_n)$  を得る.

### ・ 解析原理：

- 1) 物理学により，測定量  $D$  を目的量  $X$  の関数(順問題ルーチン)で与える

$$D_{\text{cal}} \equiv D_{\text{cal}}(X) \quad (1)$$

→ 「順問題ルーチン」を取り替えることで様々な実験に対応

- 2) 計算値  $D_{\text{cal}}(X)$  と実測値  $D$  の残差を目的関数  $F(X; D)$  とする.

例えば， $F(X; D) \equiv \sum_i (D_i - D_{\text{cal},i}(X))^2 \quad (2)$

→ 最適化・不確かさ解析を実行

### ・ 5種解析法が実装

- (1) 局所探索型最適化(Nelder-Mead法)
- (2) グリッド型探索
- (3) 並列化ベイズ最適化(PHYSBOライブラリ(\*)を利用)

(\*)<https://www.pasums.issp.u-tokyo.ac.jp/PHYSBO/>

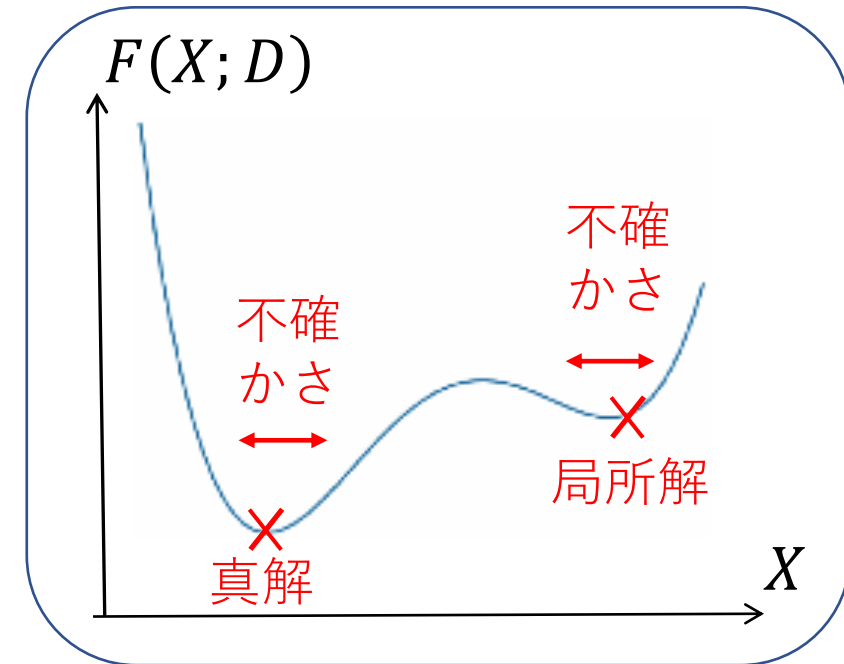
- (4) レプリカ交換(RE)モンテカルロ(MC)法

- (5) **ポピュレーションアニーリング(PA)モンテカルロ(MC)法**

→ 「富岳」などの超並列計算機に最適

ベイズ推定

図 目的関数  $F(X; D)$  の概念図

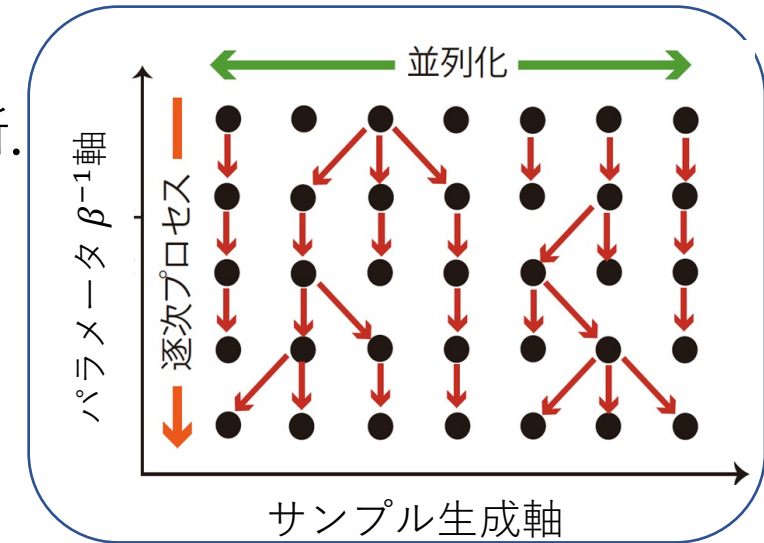


# 使用解析手法：ポピュレーションアニーリングモンテカルロ法(PAMC法)

K. Hukushima and Y. Iba., AIP Conf. Proc. 690, 160 (2003).

- 並列化されたモンテカルロ法の1種で，大域解析を実現。  
→ レプリカと呼ばれる多数サンプルデータを，乱数により生成・更新。  
→ 100万以上の並列自由度。「富岳」に理想的(右図)
- パラメータ $\beta$  ( $\beta^{-1}$ が観測データの不確かさ強度)に依存した事後確率分布 $P(X|D;\beta)$ をヒストグラムとして表現(下図)。

PAMC法の概念図



注：ベイズ事後確率分布 $P(X|D)$ をベイズの定理から計算。

$$P(X|D;\beta) = \frac{P(D|X;\beta)P(X)}{P(D;\beta)}$$

$$P(D|X;\beta) \propto \exp(-\beta F(X, D))$$

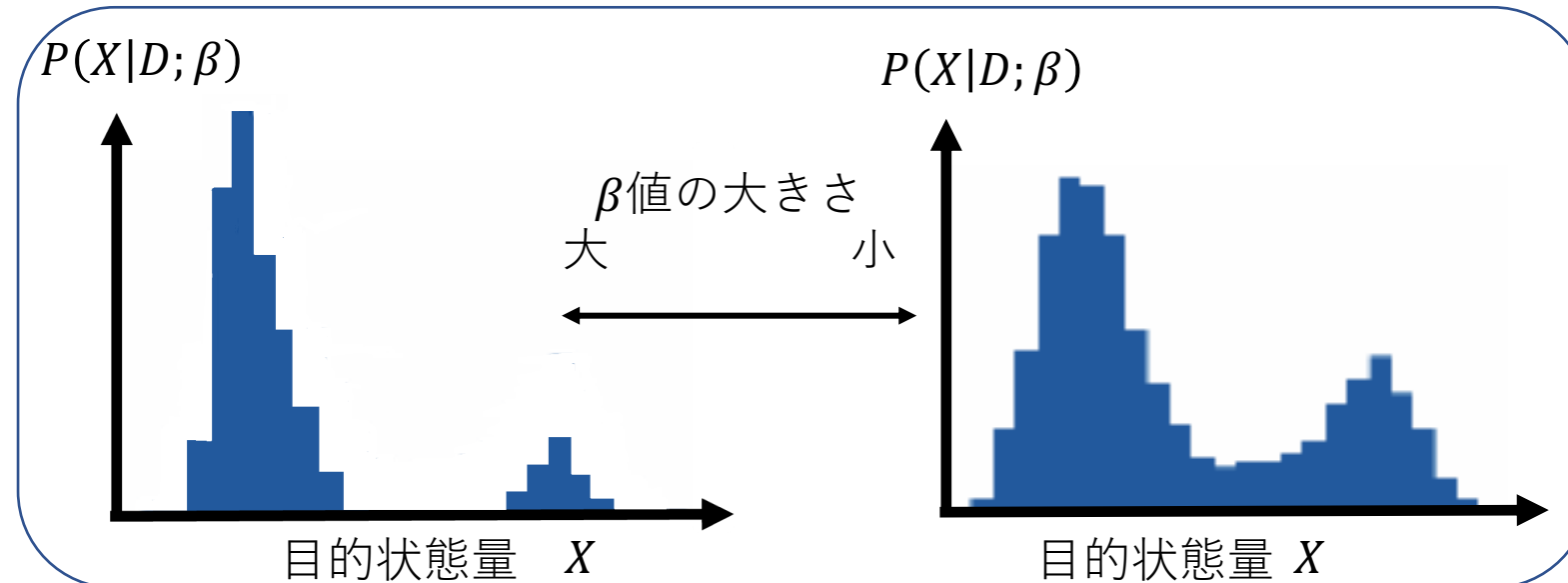
(ボルツマン型尤度関数)

$$F(X, D) \equiv R(X, D)^2:$$

(測定量についての  
実験と計算の2乗残差)

$P(X)$ ：定義域で一様分布

ベイズ事後確率分布の模式図



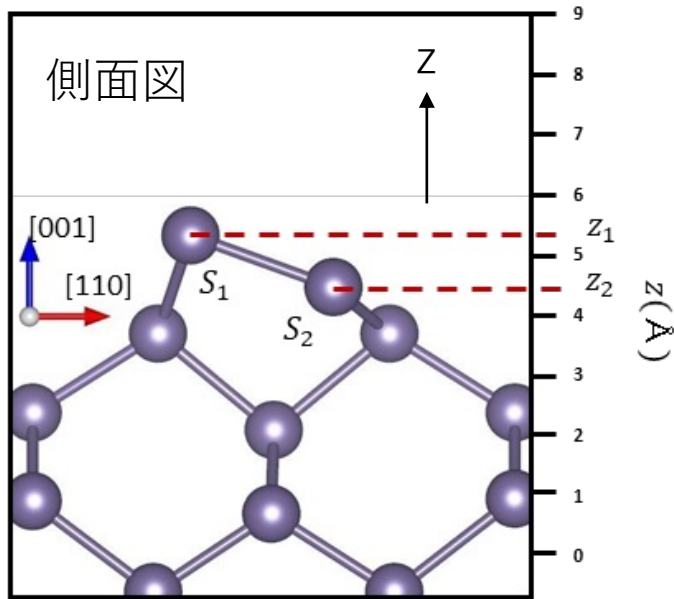
# TRHEPD計測におけるベイズ事後確率分布計算のデモ: Ge(001)-c4x2表面構造解析

Motoyama, et al., Comp. Phys. Commun. 280, 108465 (2022)

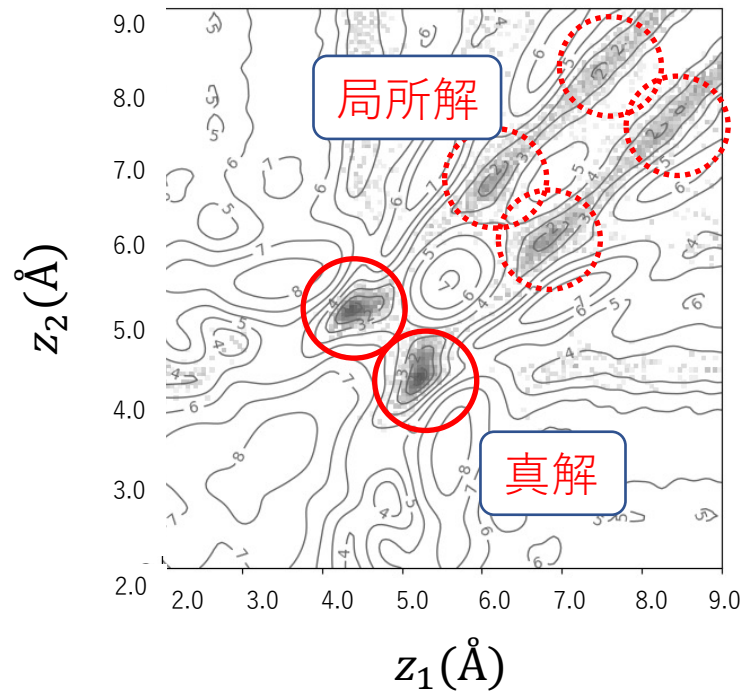
・ 最表面2原子のz座標( $z_1, z_2$ )に対するベイズ事後確率分布 $P(z_1, z_2|D)$

→ 大域解析(initial guess不要). 不確かさパラメータ値 $\beta$ により, 真解・局所解を検出・区別できる.

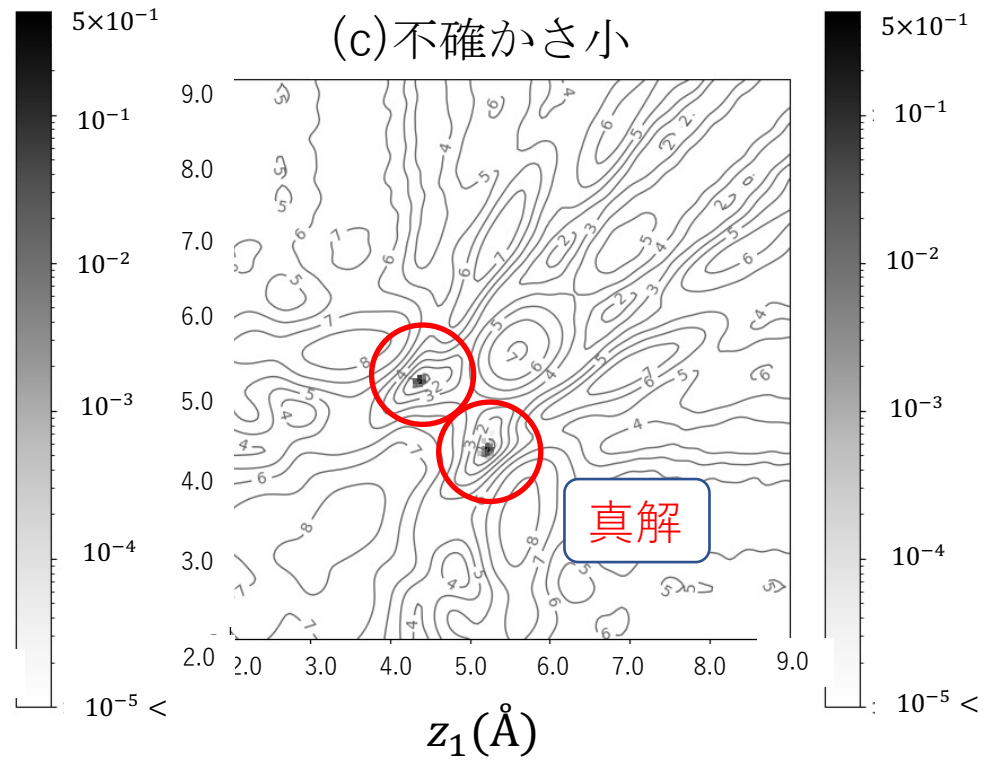
(a) 原子構造



(b) 不確かさ大



(c) 不確かさ小



注: 実測定データの代わりに, 人工データを利用

# 結果 : Ge(001)-c4x2表面構造解析 2変数探索

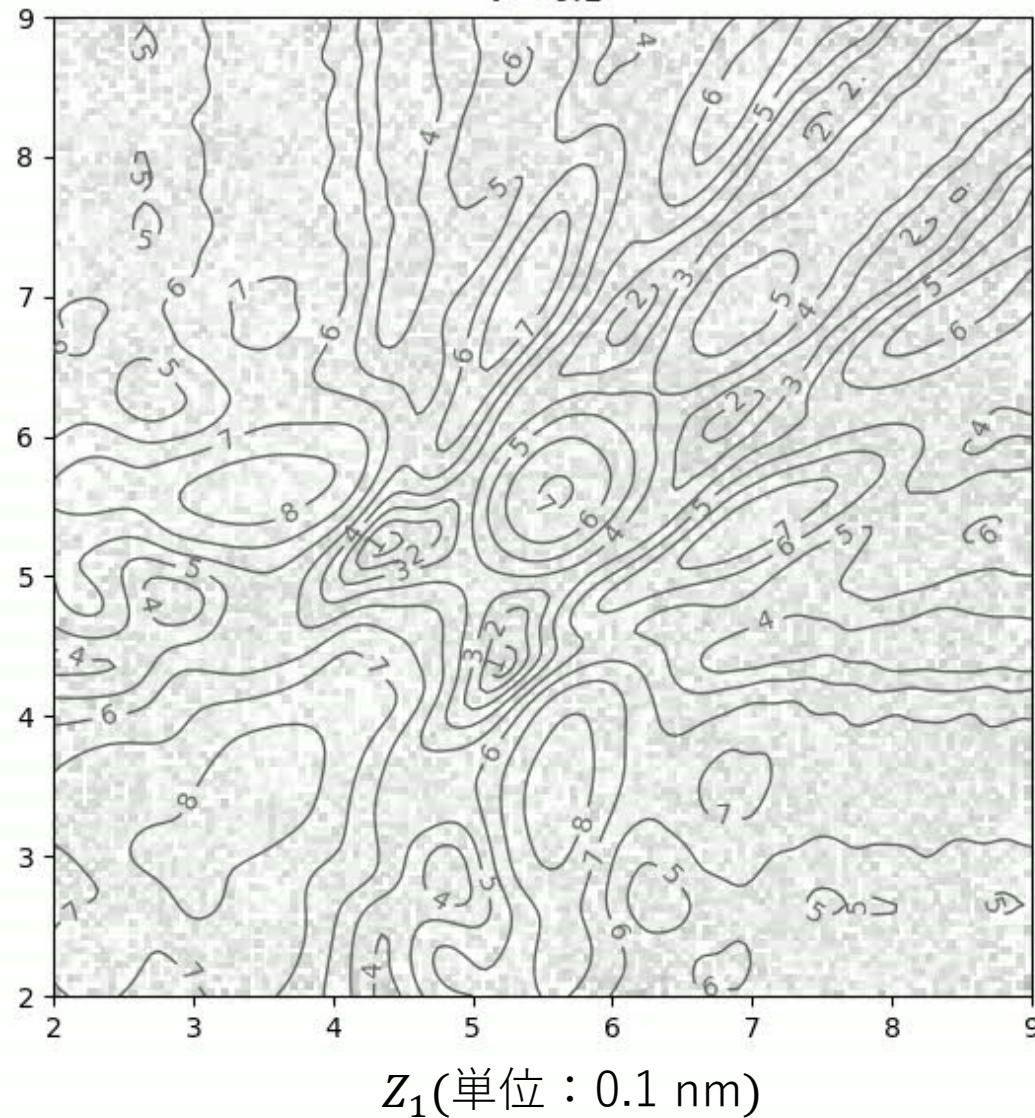
ベイズ事後確率分布 $P(z_1, z_2 | D)$ の不確かさパラメータ依存性 (不確かさ : 大→小)

$\tau = 0.1$

等高線図の値 :

R因子(残差) $R(z_1, z_2) \times 100$

$z_2$ (単位 : 0.1 nm)



$5 \times 10^{*-1}$

$10^{*-1}$

$10^{*-2}$

$10^{*-3}$

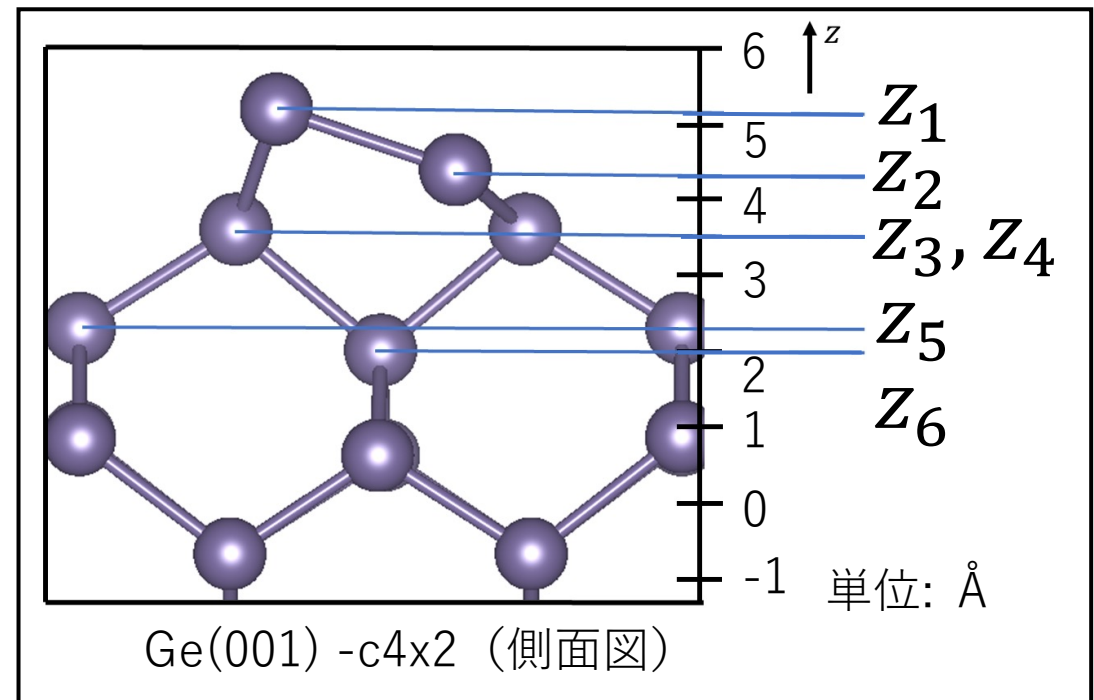
$10^{*-4}$

$10^{*-5} <$

# Ge(001)-c4x2構造への応用： 表面1~6層のz座標の大域解析

- 真解が既知である構造を使った，手法テスト
- 一波条件(z座標のみの決定)
- 側面図（右図）で，  
 $0 < z < 9 \text{ \AA}$  での大域解析

→ 事後確率分布 $P(X|D; \beta)$ をヒストグラムとして生成. ( $X \equiv (z_1, z_2 \dots z_6)$ )





まとめ：

- ・ 「2DMAT」：データ駆動科学の立場から開発された，データ解析フレームワーク  
「2DMAT」 (<https://www.pasums.issp.u-tokyo.ac.jp/2DMAT/>)
- ・ 超並列モンテカルロ法(PAMC法)によるベイズ推定（ベイズ事後確率分布計算）  
→「富岳」などの超並列計算機に理想的.
- ・ 全反射高速陽電子回折実験(TRHEPD,トレプト)で実践.  
→回折データ $D$ から，表面付近（厚さ：サブナノメートル）の原子座標 $X$ を推定.  
→大域解析(解へのinitial guess不要)で，真解（正しい原子座標・原子数）が得られた.
- ・ 展望：
  - (i) 「見えないものを，見る」(測定限界突破)  
例：TRHEPDによる表面水素の検出（→次世代触媒・電池）  
（望月，招待講演，SLOPOS-16, Orléans, France, 2023年7月16日）
  - (ii) 種々実験（TRHEPD, RHEED, SXRD, LEED, オペランドPTRF-XAFS), DFT計算との連携  
→マルチプローブ計測（同一サンプルを多数手法で計測）における，  
データ駆動科学の立場からの統合的データ解析  
（関連：KEK物構研での試料ケース共通化）