

富岳を用いた高度計測むけデータ駆動科学：2次元物質解析への応用

核融合研究所, 高エネ研 (客員), 東大物性研(客員) 星健夫

(出身：東京大学工学部物理工学科)

- 背景1：21世紀における
計算科学のパラダイムチェンジ
- 背景2：高度計測における
データ解析によるブレークスルー
- 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
- 大局的展望：デジタルツインなど



謝辞：和田 健・望月出海
(高エネ研低速陽電子),
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),
山本有作・工藤周平 (電通大),
岩本晴道, 木下直希 (鳥取大),
中野陽斗, 寺地雄真 (核融合研)

- (*)謝辞
- 富岳は, HPCI課題(hp210267, hp220248, hp230304)などで実施した.
 - 2DMAT(<https://github.com/issp-center-dev/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, 東大物性研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, 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>)

背景 1 : 21世紀における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

計算機の質的变化:
超並列コンピュータ(「京」「富岳」など)

並列(「手分け」)
アルゴリズムによるブレークスルー

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→



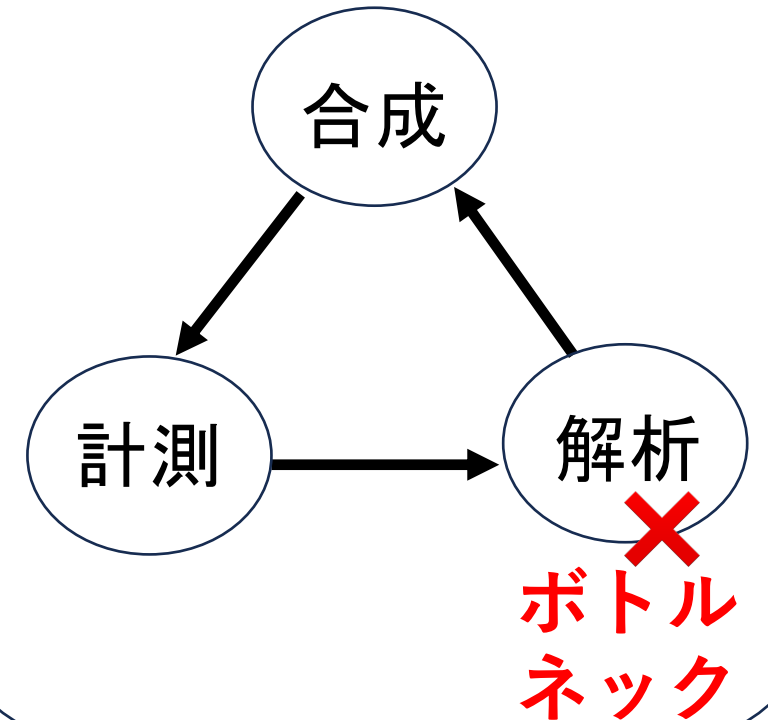
背景 2: 高度計測におけるデータ解析によるブレークスルー

データ解析発展で「見えないモノを見る」（計測限界突破）
（従来計測手法では検出不可能だったモノを検出する）
例：クライオ電子顕微鏡（2017年ノーベル賞）

本研究の対象：先端計測

- ・ 大型実験施設や高度実験技術が必要
- ・ データが「断片的」「一見矛盾」し、「死蔵」されたデータや未解決問題が多数。
→ データ解析による計測革新への期待

研究開発サイクル



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」

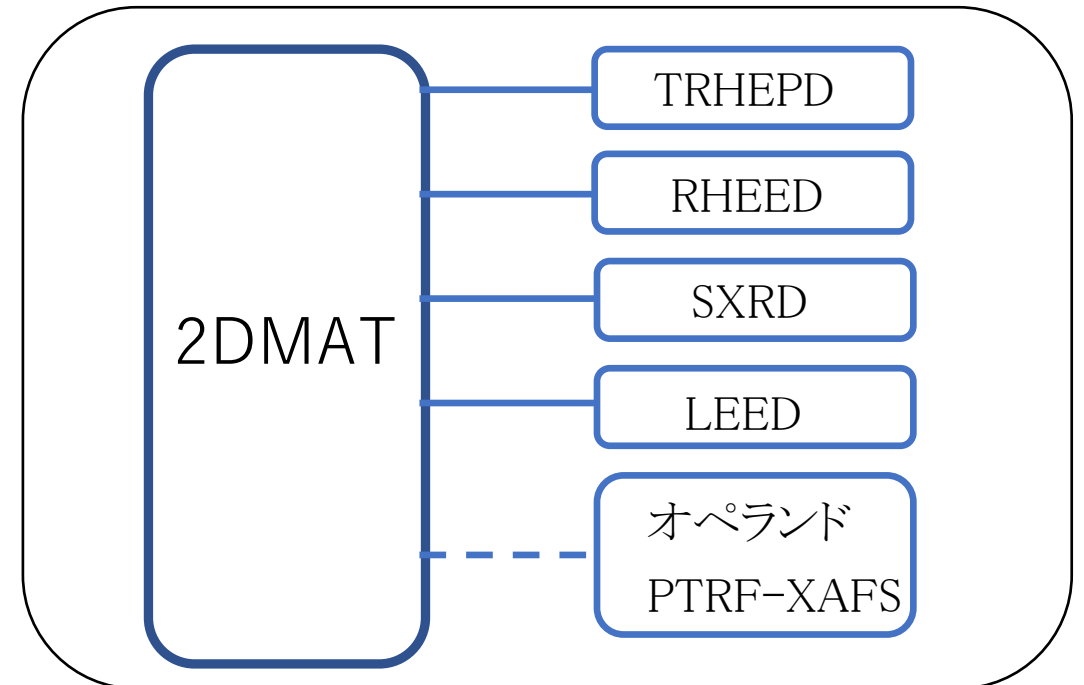
<https://github.com/issp-center-dev/2DMAT/>; Motoyama, *et al.*, Comp. Phys. Commun. 280, 108465 (2022)

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、 新しい測定データ解析フレームワーク「2DMAT」

<https://github.com/issp-center-dev/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

・ 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/>

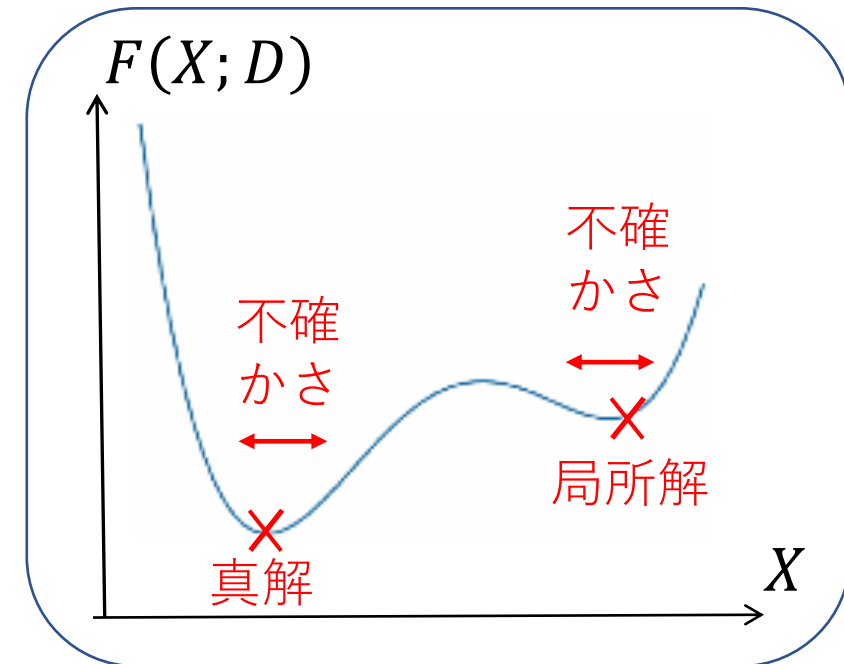
- (4) レプリカ交換(RE)モンテカルロ(MC)法

- (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



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)

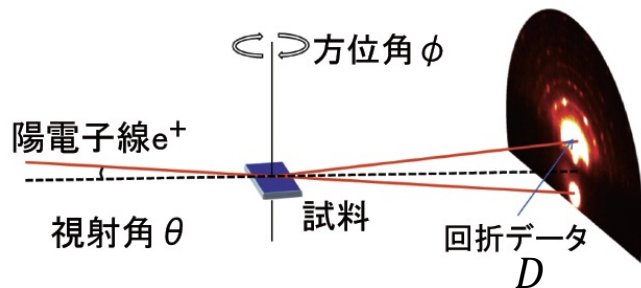
総説 : Fukaya, *et al.*, J. Phys. D 52, 013002 (2019)

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) . 日本発. 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

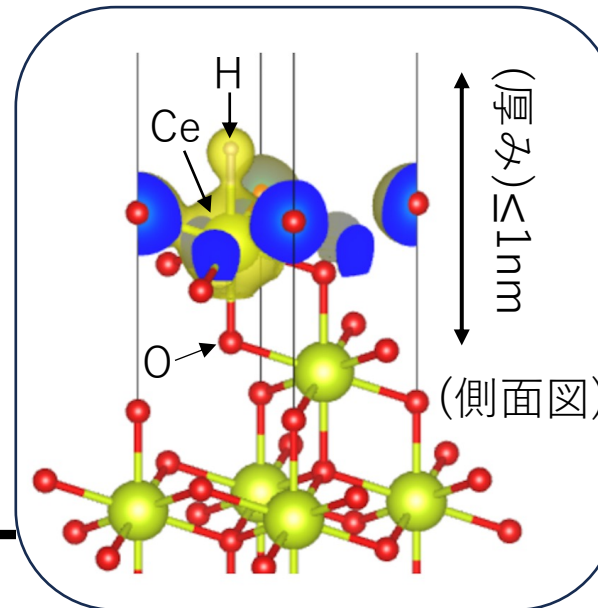
(<https://www2.kek.jp/imss/spf/>)



触媒系 : 水素吸着したセリア(CeO₂)表面,
(Mochizuki *et al.* SLOPOS, 2023)

逆解析
($D \Rightarrow X$)
2DMAT
(解析ソフト)

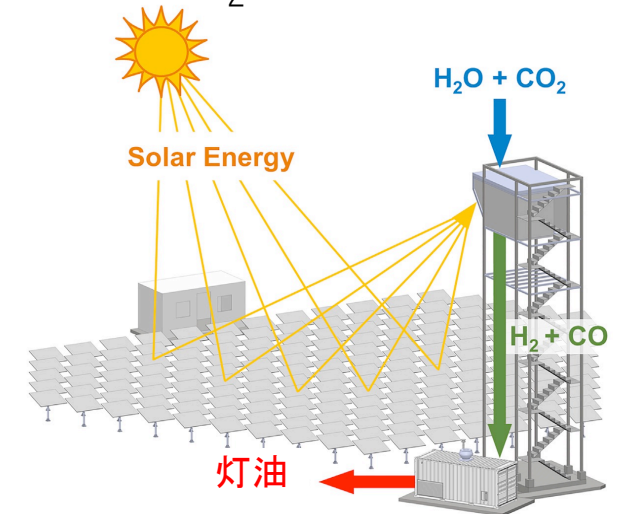
2次元物質(表面) 構造 (原子座標) Xの決定例



出口 (の1つ) :

カーボンニュートラル向け
(CO₂再資源化)触媒開発

例 : CO₂+水+太陽光で灯油生成



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

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
反射高速電子回折(Reflection high energy positron diffraction)を用いることで、表面領域の構造を非破壊的に計測可能
- 回折データ (各スポットの強度)

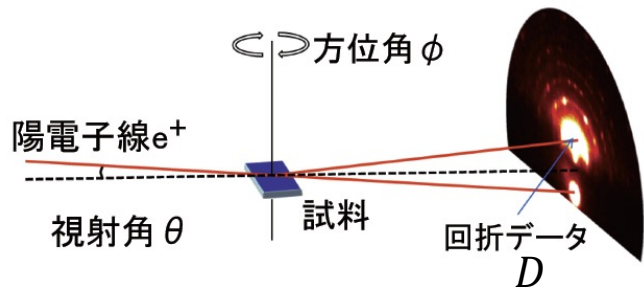
日本語解説記事:

[a] 一宮 彪彦, 「我が国における反射高速電子回折の発展と全反射陽電子回折への展開」, 日本物理学会誌, 2015年

[b] 深谷 有喜, 「全反射高速陽電子回折 (TRHEPD)」, 陽電子学会誌, 2019年

実験装置模式図

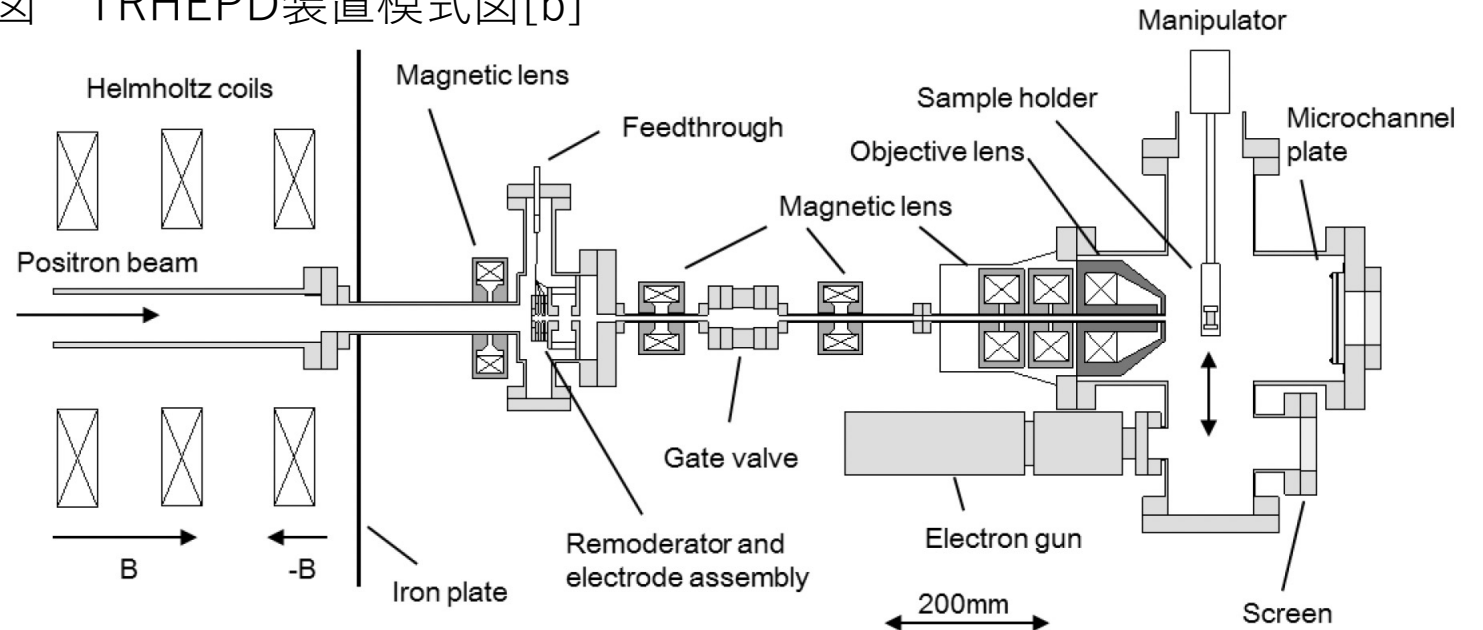
(<https://www2.kek.jp/imss/spf/>)



逆角
(D)
2D
(解析)

触媒系: 水素吸着したセリア(CeO_2)表面
(Mochizuki *et al.* SLOPOS, 2023)

図 TRHEPD装置模式図[b]

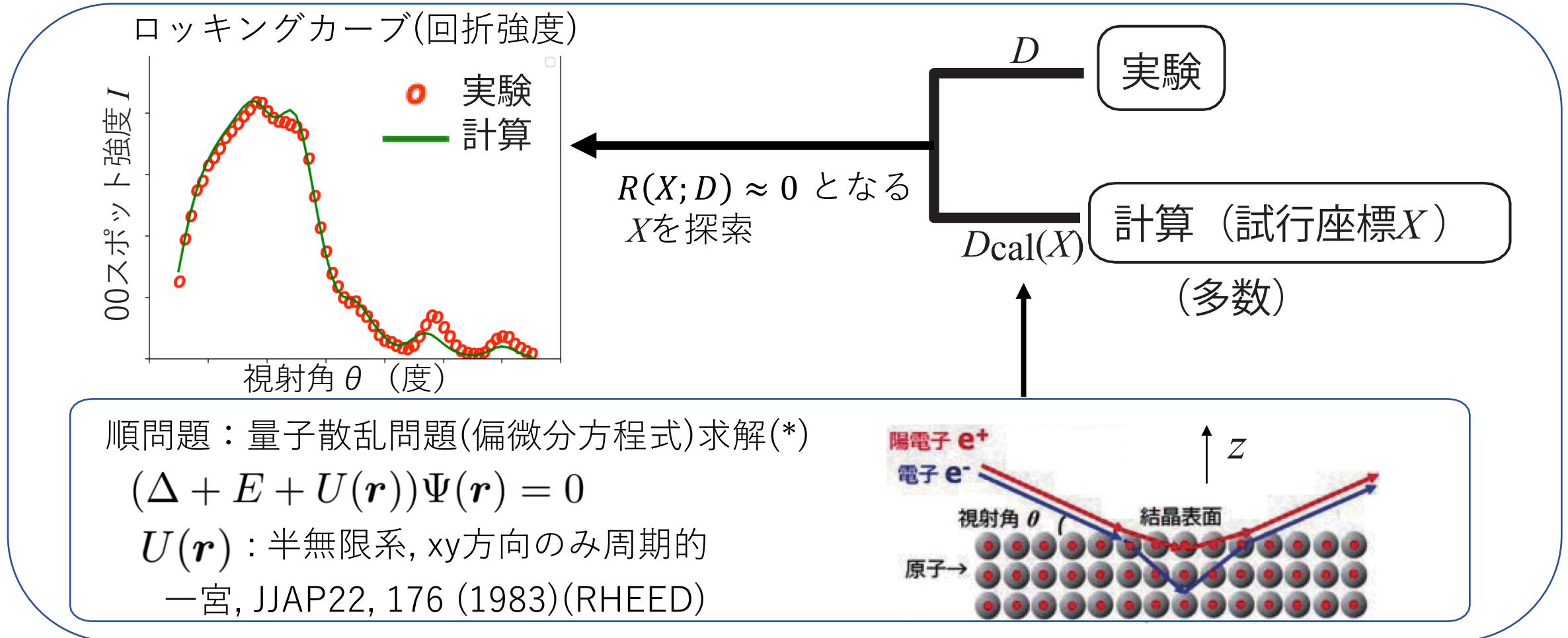


TRHEPDにおけるデータ解析

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))

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。

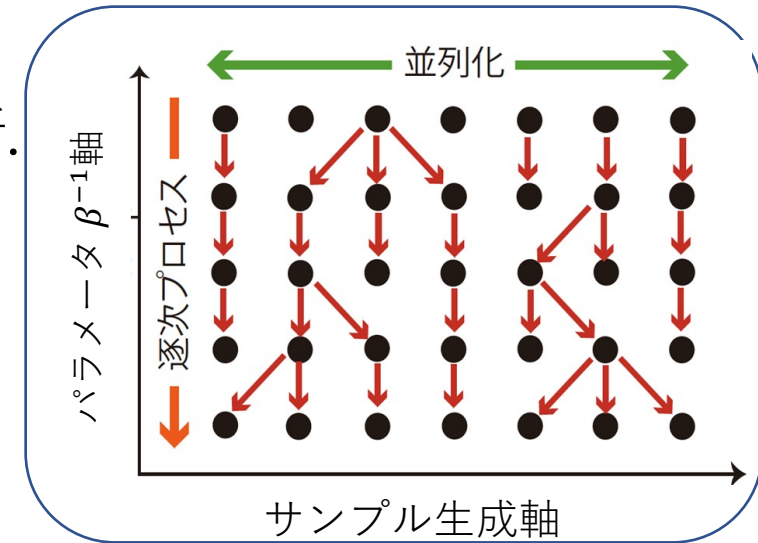
検査Bは、病気Aに罹患していない人のうち、20%を陽性と誤判定する）

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 β (β^{-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。

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

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

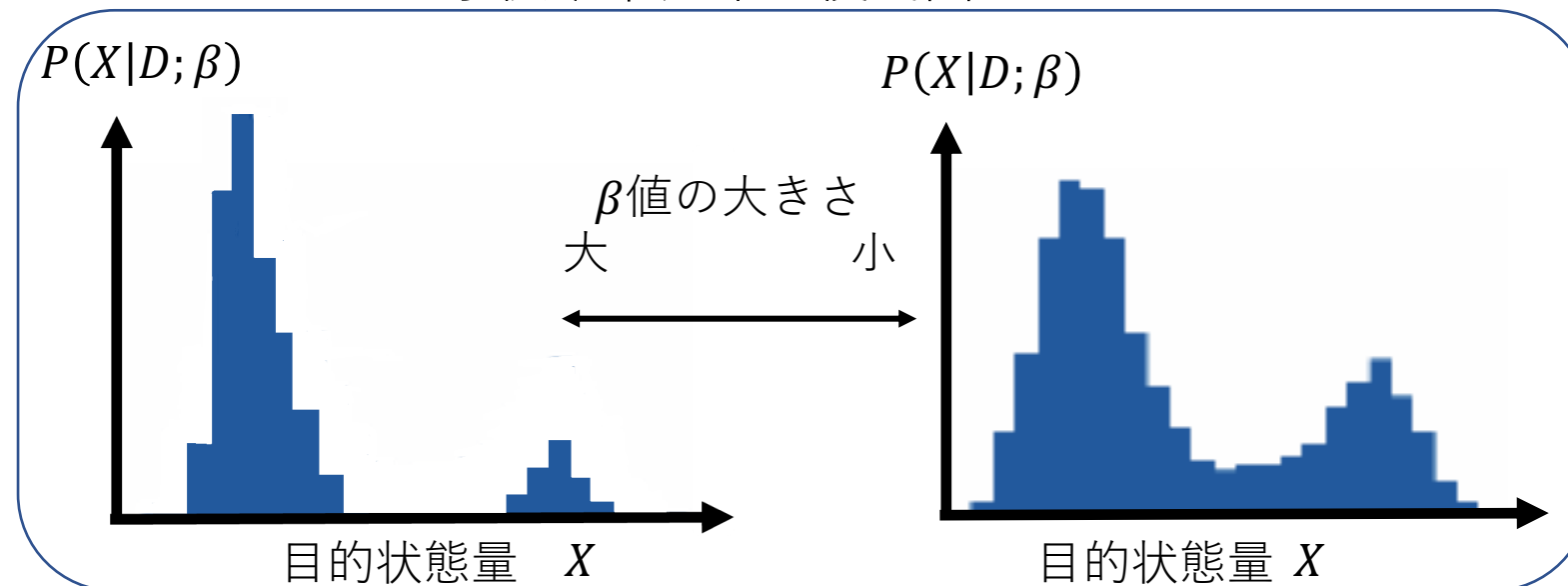
(ボルツマン型尤度関数)

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

(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



「富岳」での高速計算

- 「富岳」で、PAMC法をコード調整中
- [暫定データ]並列ベンチマーク例
(計算ノード数)=3,072- 82,944(全体の約1/2)で,
強スケール型並列効率 $\alpha = 0.91$ (注:理想値は $\alpha = 1$)

詳細:高速化の論点

- (a) プログラム言語は、MPI並列化されたPython. MPI通信は、汎用Pythonオブジェクト(例: `comm.send`, `comm.reduce`)ではなく、高速なbuffer-likeオブジェクト(例: `comm.Send`, `comm.Reduce`)を利用.
- (b) 順問題ルーチン($D_{cal}(X)$ の計算)はFortranで書かれており、`iso_c_binding`を通じて、外部ライブラリとして、Pythonコードにリンク.
- (c) 計算途中でリサンプリング操作(レプリカ再配分)が実施される.
→ Walker's tableアルゴリズムを利用(各MPIプロセスで、理想的ロードバランス)
- (d) 「富岳」特有の論点: Lightweight Layered I/O Accelerator (LLIO) の利用
(データI/Oにおけるデータ通信を最適化)

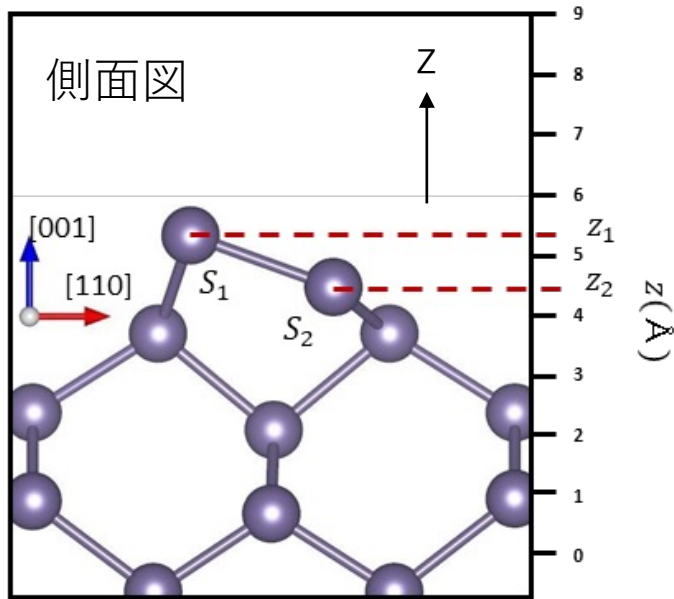
TRHEPD計測におけるベイズ事後確率分布計算のデモ: Ge(001)-c4x2表面構造解析

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

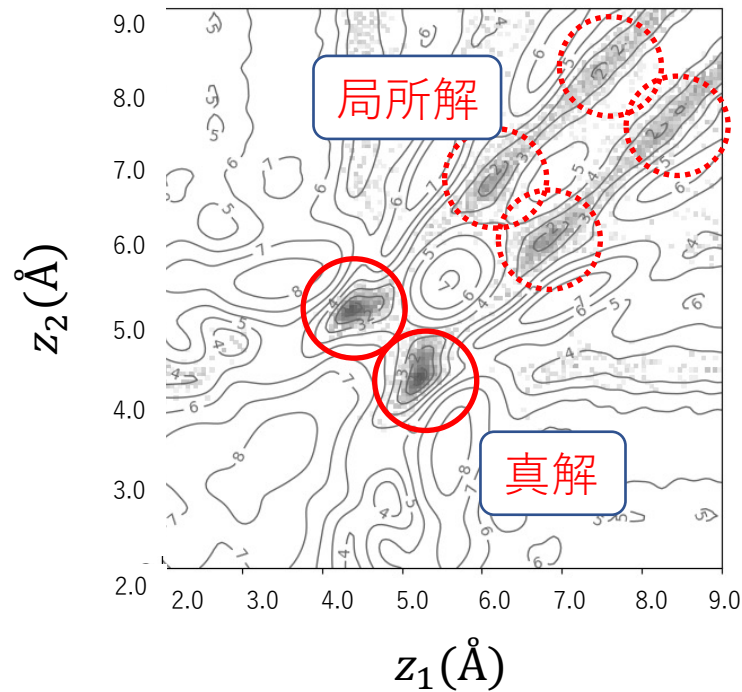
・ 最表面2原子のz座標(z_1, z_2)に対するベイズ事後確率分布 $P(z_1, z_2|D)$

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

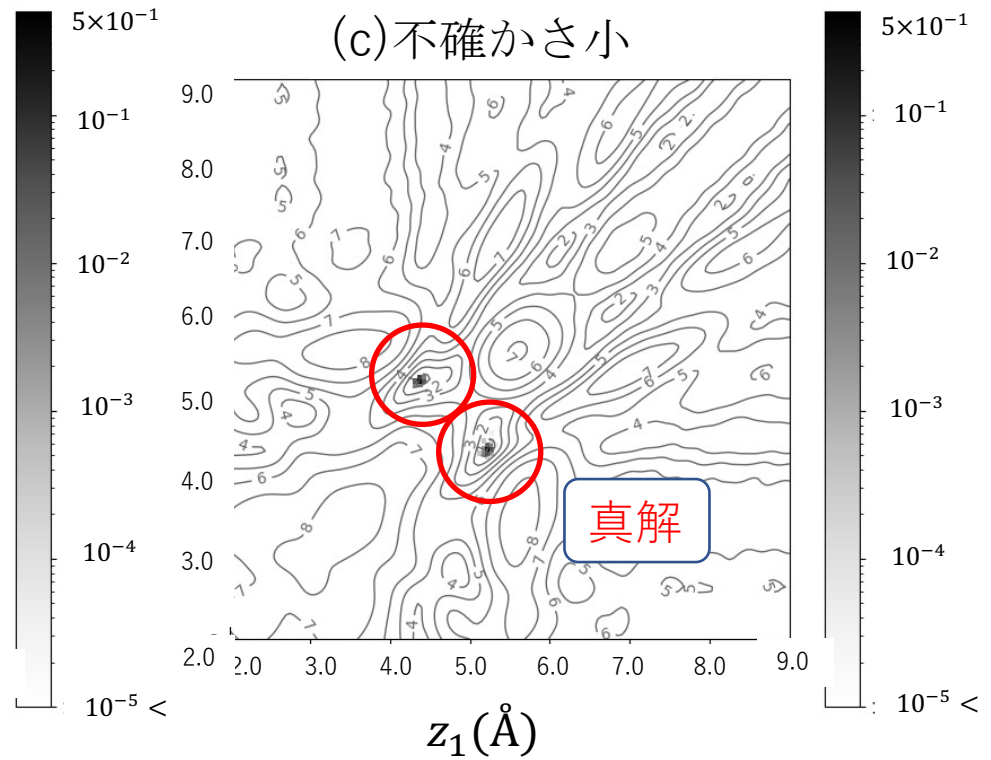
(a) 原子構造



(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

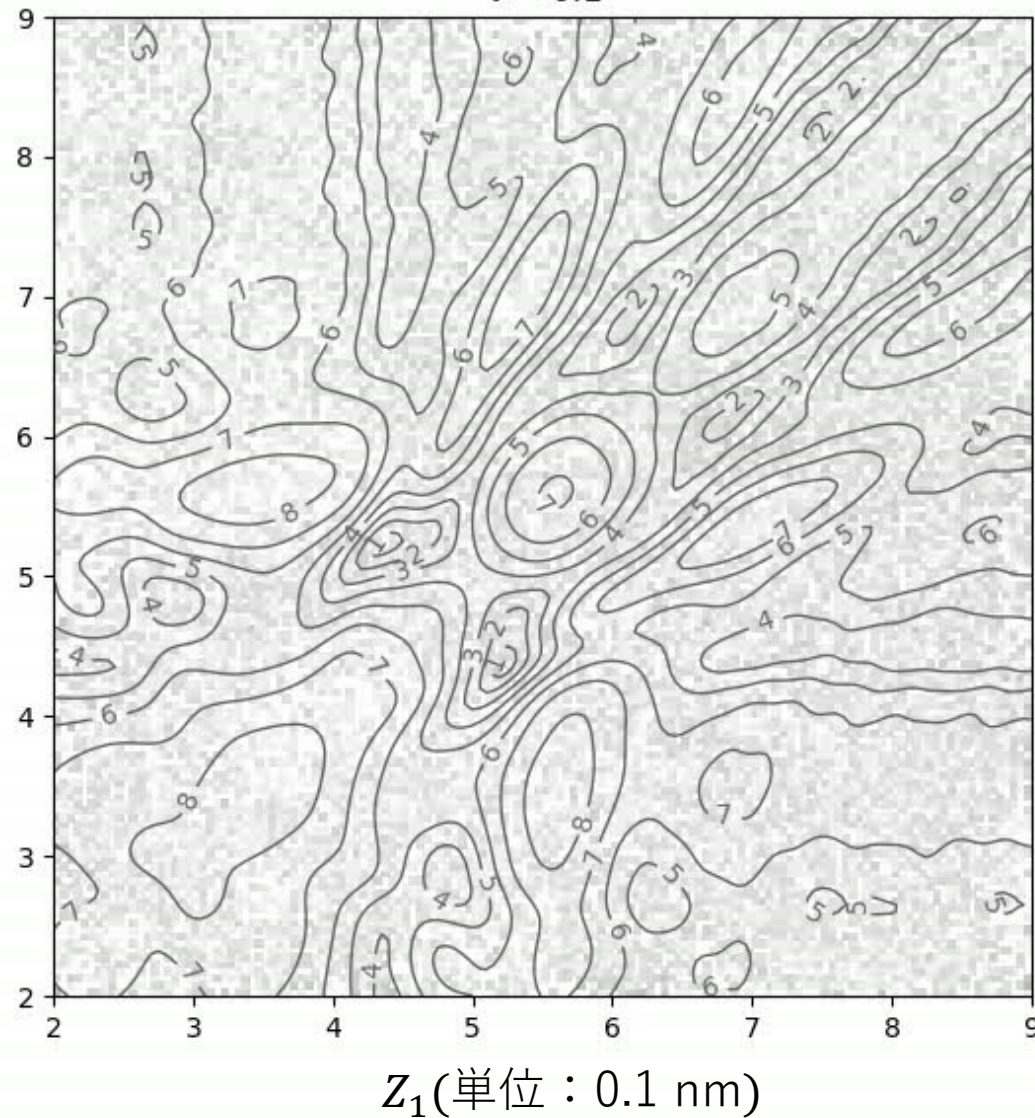
ベイズ事後確率分布 $P(z_1, z_2 | D)$ の不確かさパラメータ依存性 (不確かさ : 大→小)

$\tau = 0.1$

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}

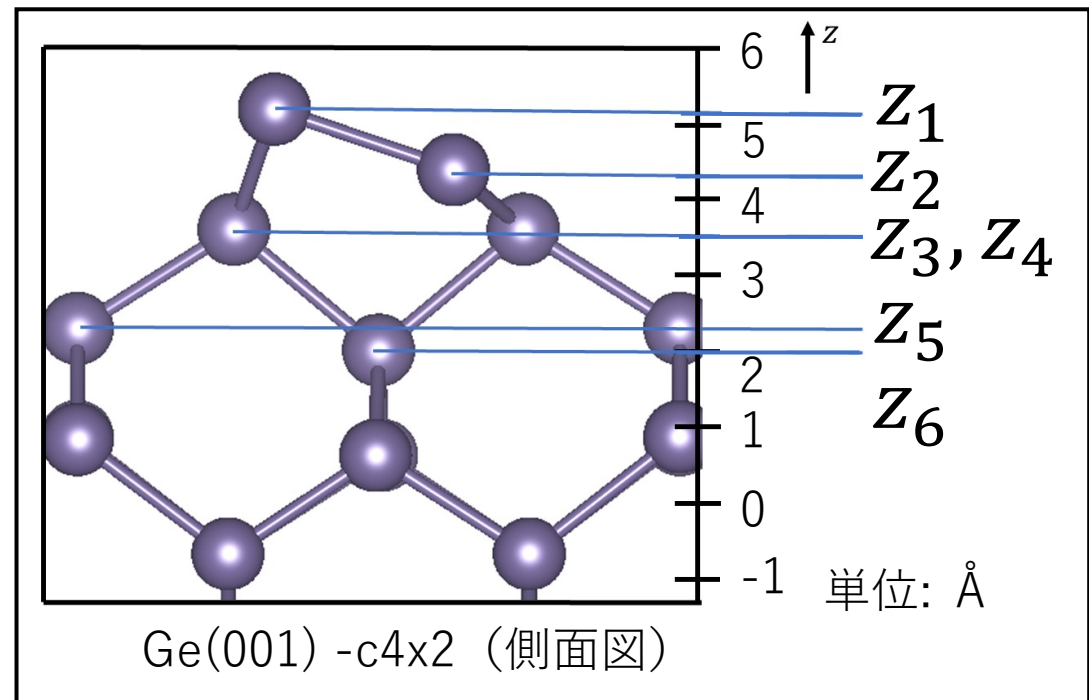
10^{*-4}

$10^{*-5} <$

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. ($X \equiv (z_1, z_2 \dots z_6)$)



(これまでの) 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://github.com/issp-center-dev/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, 真解が得られた.
- 展望:

(i) 「見えないものを, 見る」(測定限界突破)

例: 表面水素の検出 (→ 次世代触媒・電池)

(望月, 招待講演, SLOPOS-16, France, 2023年7月16日)

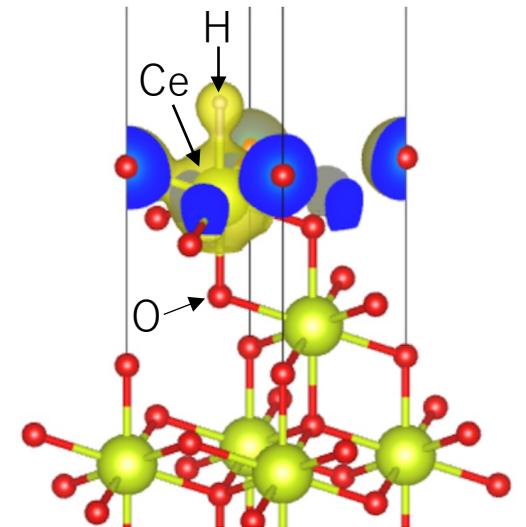
(ii) 種々実験 (TRHEPD, RHEED, SXRD, LEED,

オペランドPTRF-XAFS), DFT計算との連携

→ マルチモーダル (複数手法) 計測における. 統合的データ解析

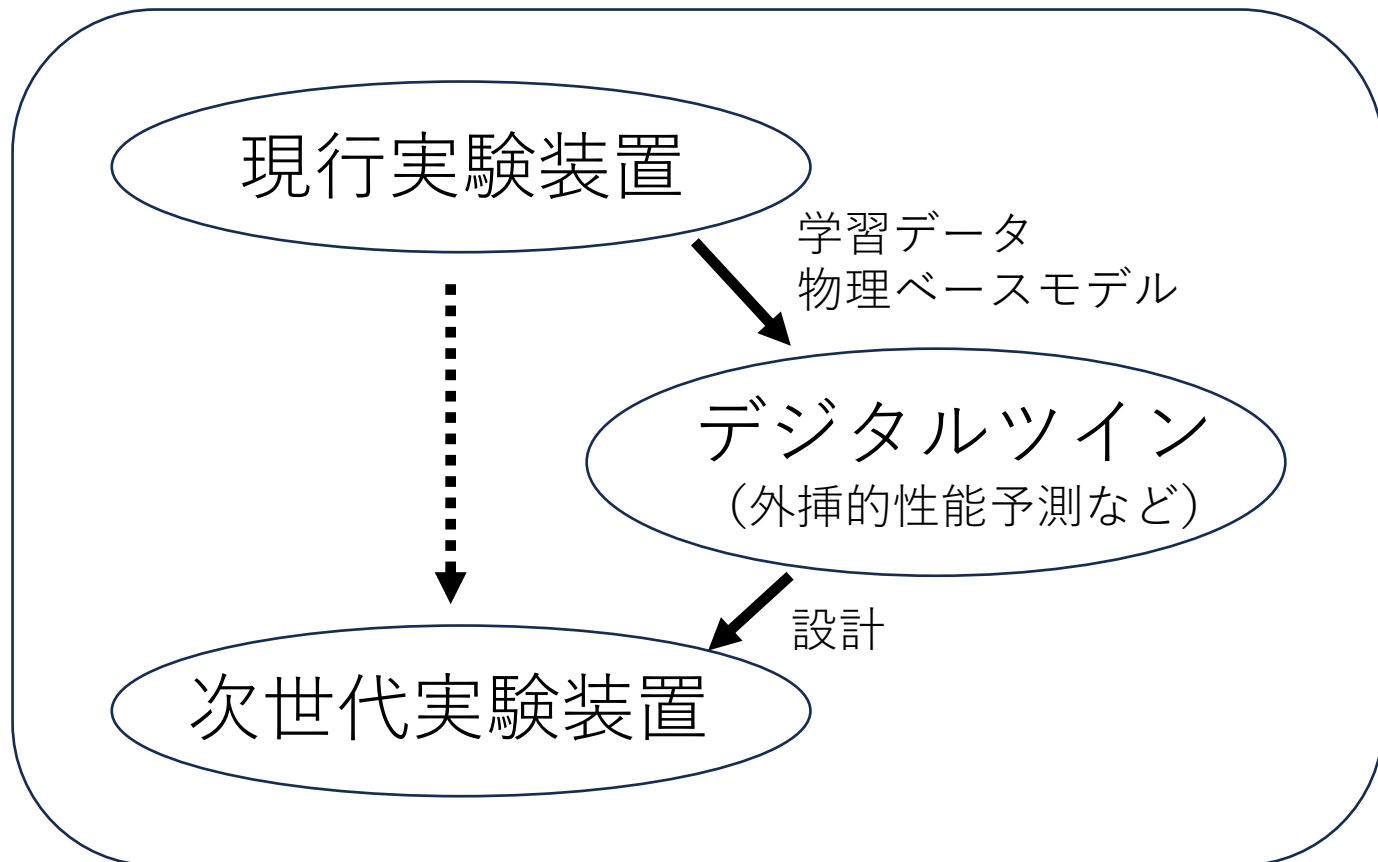
(関連: KEK物構研での試料ケース共通化)

水素吸着した
セリア(CeO_2)表面,



大局的展望：デジタルツインなど

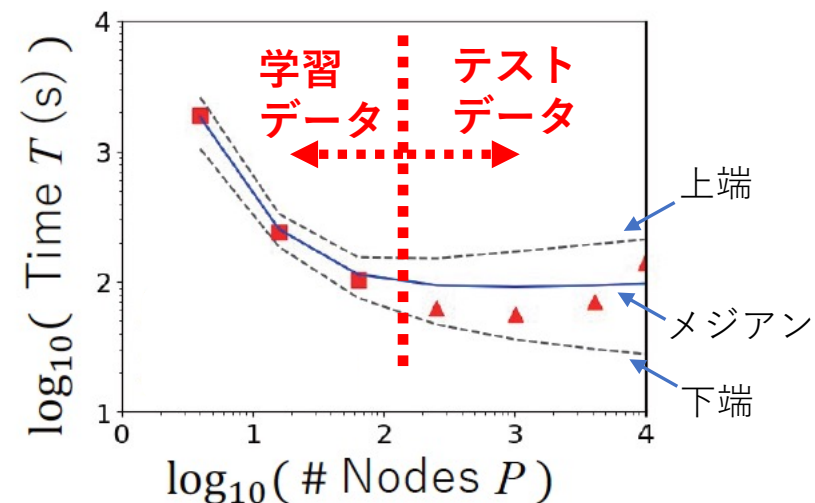
- 計測データ解析は汎用
→物質科学以外への展望（例：核融合科学）
- 大局的展望
→高エネ研・核融合研などでの、次世代実験装置開発にむけたデジタルツイン開発
（要素パラメータが向上したときに、装置性能はどの程度向上するか?）



外挿型予測の例：

スパコンの計算性能予測（ベイズ推定）

[予測：メジアン、信頼性区間の上下端をプロット]



Kohashi *et al.*, JSIAM Letters 14, 13 (2022)