

2DMATによる計測データ解析の概要と事例紹介



星健夫(核融合科学研究所, 東大物性研, KEK物構研)

(星の略歴: 出身は, 東京大学工学部物理工学科;
元々の分野は, 物質科学. 2023年8月から現職)

- 背景: 4つのインフォーマティクス, ベイズ推定
- 逆問題データ解析フレームワーク「2DMAT」
- 物質科学・プラズマ科学への応用
- まとめと展望

(*)謝辞

- 2DMAT(<https://github.com/issp-center-dev/2DMAT/>)は,
「2020・2021・2024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, 東大物性研スパコン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>)

背景：4つのインフォマティクス

歴史的経緯：「マテリアルインフォマティクス」から「4つのインフォマティクス」に発展。

(狭義の)マテリアル インフォマティクス

ナニを作れば良いか？
仕様, **what**の問い

プロセス インフォマティクス

ドウ作れば良いか？ **how**の問い。
製造プロセス最適化, プラント制御など

計測 インフォマティクス

情報抽出
「ミエナイ」モノを「ミル」

クライオ電子顕微鏡
(2017年ノーベル賞)

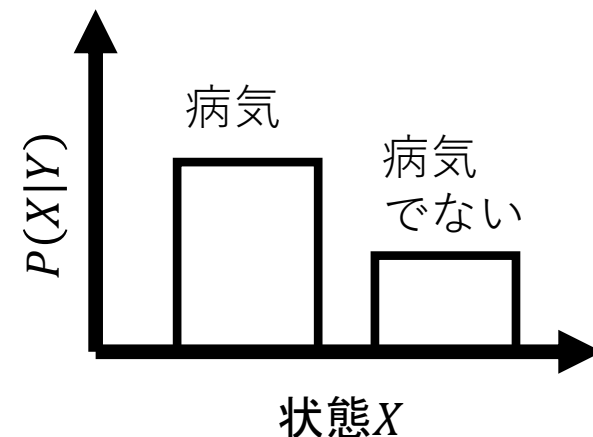
物理 インフォマティクス

演繹的理解。
whyの問い。

背景：計測データ解析におけるベイズ推定

- ・ **ベイズ推定 = 人間の推定を模した, AI型の数理手法**
- ・ 計測データ $Y \equiv (Y_1, \dots, Y_m)$ が与えられた時に,
目的状態量(知りたい量) $X \equiv (X_1, \dots, X_n)$ 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

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



- ・ 例：病気Aには検査Bがあるが、誤判定する場合もある。
「病気Aにかかっているか、否か」が知りたいので、
検査Bを3回行ったところ、3回とも陽性と診断された。
実際に、病気Aにかかっている確率は、何パーセントか？

知りたい情報 X

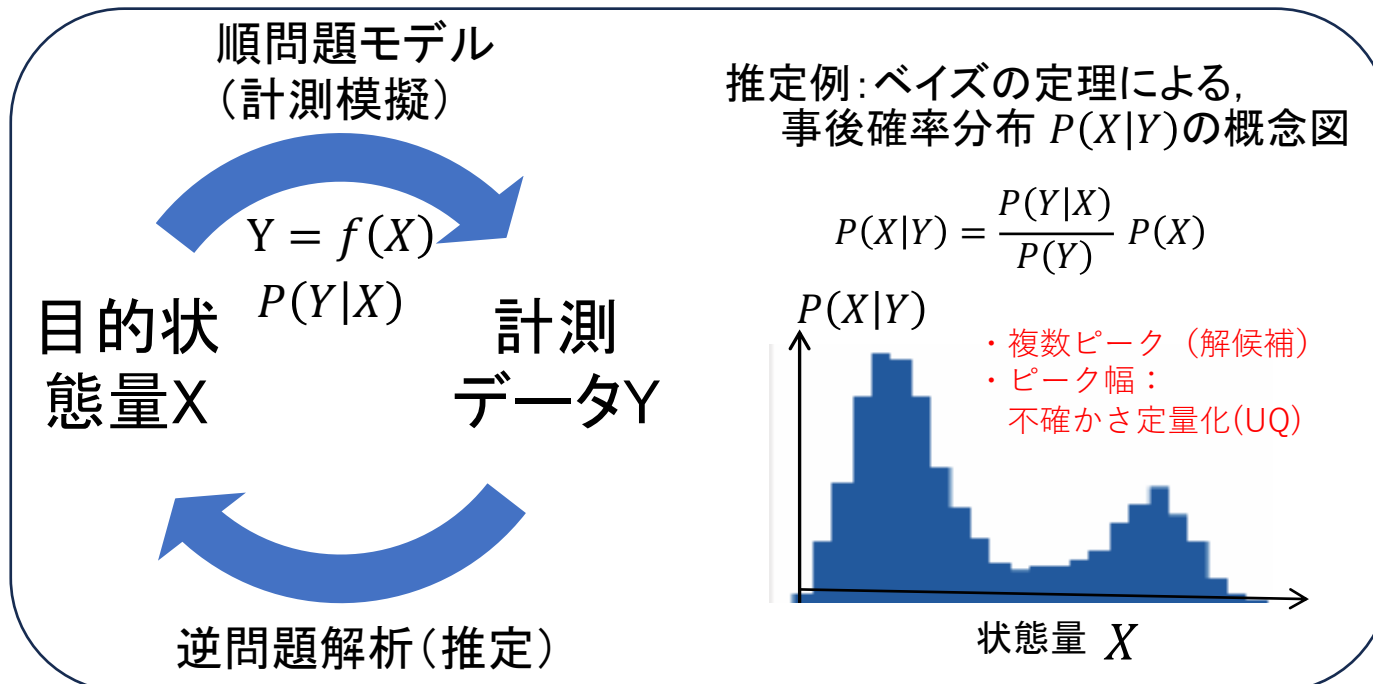
計測データ Y

計測データ解析(逆問題解析)フレームワーク2DMAT

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

- ・データ駆動科学の立場から開発された, 新しいデータ解析フレームワーク
- ・PCからスパコン(「富岳など)を想定して, 高速・高信頼な逆問題型データ解析フレームワーク
- ・順問題モデルを入れ替えることで, 様々な計測に適用可.

逆問題解析の概念図



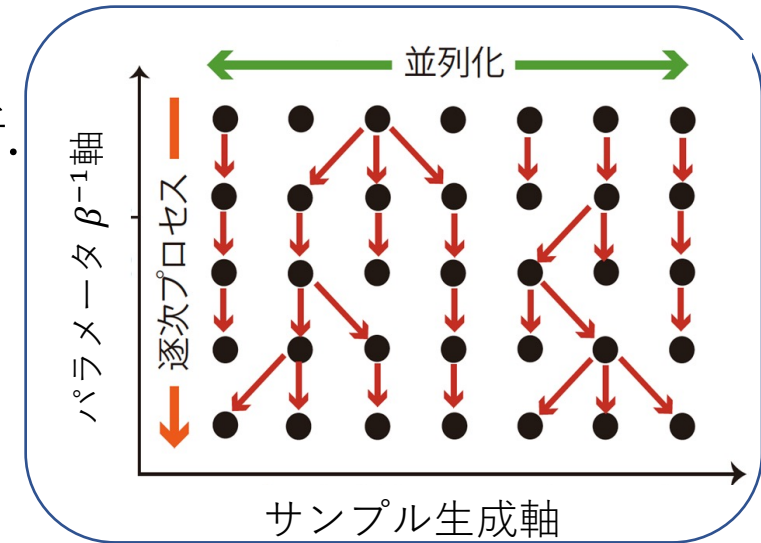
- ・5種解析法が実装
 - (1) 局所探索型最適化(Nelder-Mead法)
 - (2) グリッド型探索
 - (3) 並列化ベイズ最適化
 - (4) レプリカ交換(RE)モンテカルロ(MC)法
 - (5) ポピュレーションアニーリング(PA)モンテカルロ(MC)法**
→ 「富岳」などの超並列計算機に最適

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 ($\tau \equiv \beta^{-1}$ が観測データの不確かさ強度) に依存した事後確率分布 $P(X|Y;\beta)$ 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 $P(X|Y)$ をベイズの定理から計算。

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

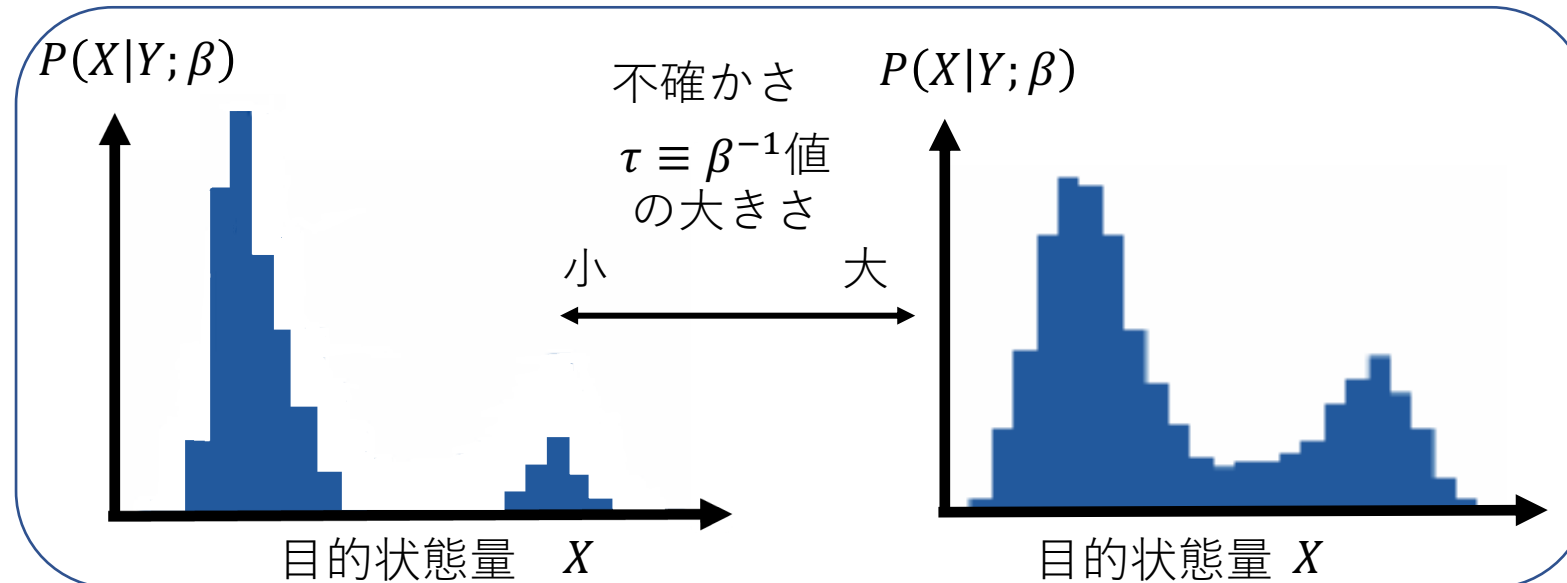
$$P(Y|X;\beta) \propto \exp(-\beta F(X, Y))$$

(ボルツマン型尤度関数)

$F(X, Y)$ ：目的関数 (例：測定量での実験と計算の2乗残差)

$P(X)$ ：事前確率(例：定義域で一様)

ベイズ事後確率分布の模式図

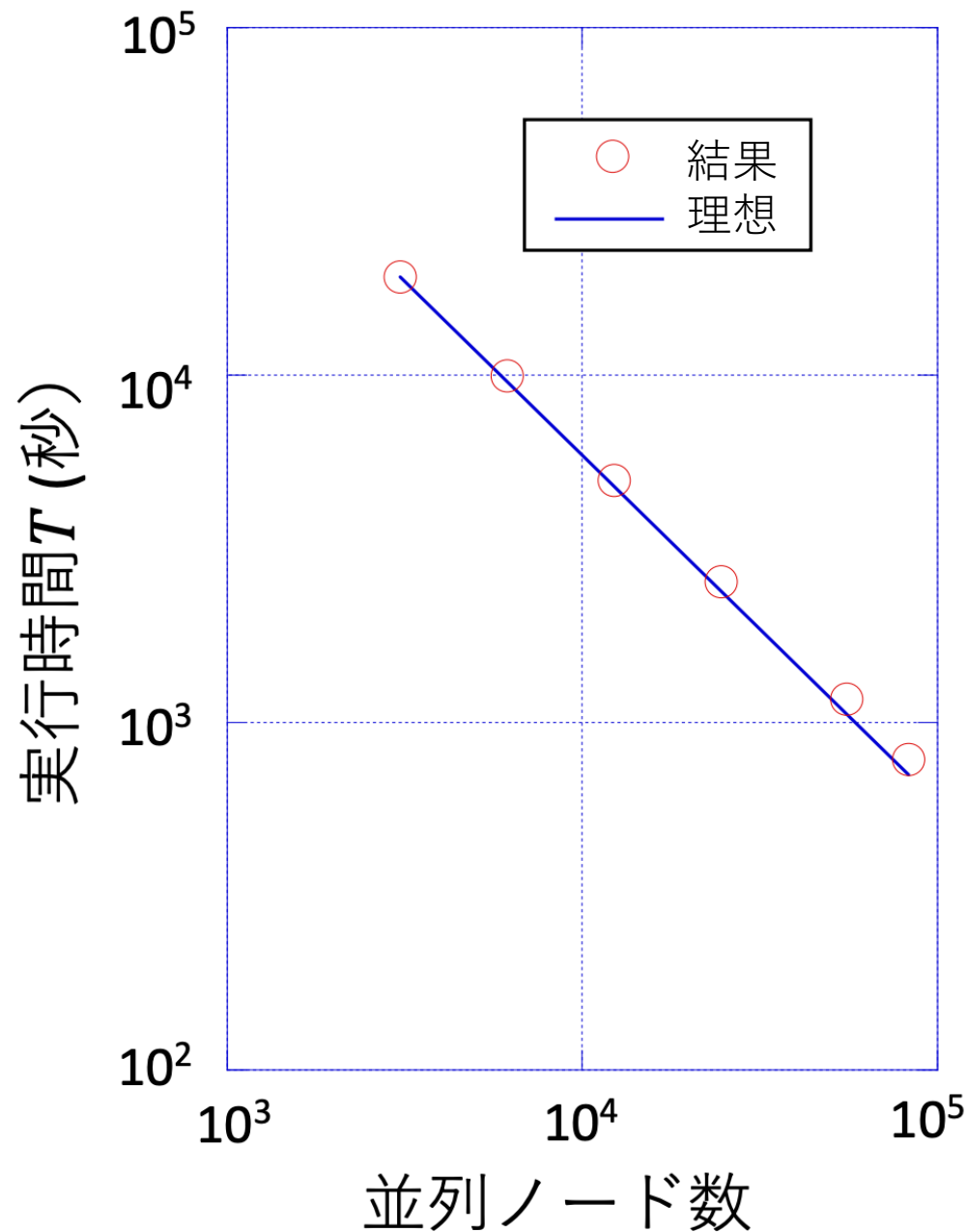


「富岳」での高速計算

- 「富岳」で、PAMC法をコード調整中
- 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 $\alpha = 0.9$ (理想 $\alpha = 1$)

詳細: 高速化の論点

- プログラム言語: MPI並列化されたPython.**
高速なbuffer-likeオブジェクト通信を利用.
- 順問題モデルコード(シミュレーター)の一部は
Fortran/Cで書かれており, iso_c_bindingsを用いて
外部ライブラリとして, Pythonにバインド.
- 計算途中でリサンプリング操作 (レプリカ再配分)を実施.
→ Walker's tableアルゴリズムを利用(理想的並列性)
- 「富岳」特有の論点:
Lightweight Layered I/O Accelerator (LLIO) の利用

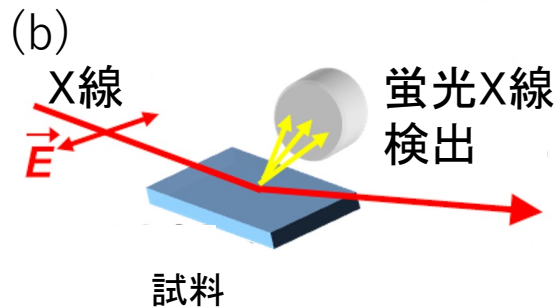
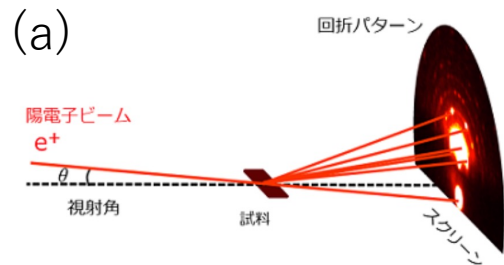


計測データ解析(逆問題解析)フレームワーク2DMAT

物質科学での活用

実験:

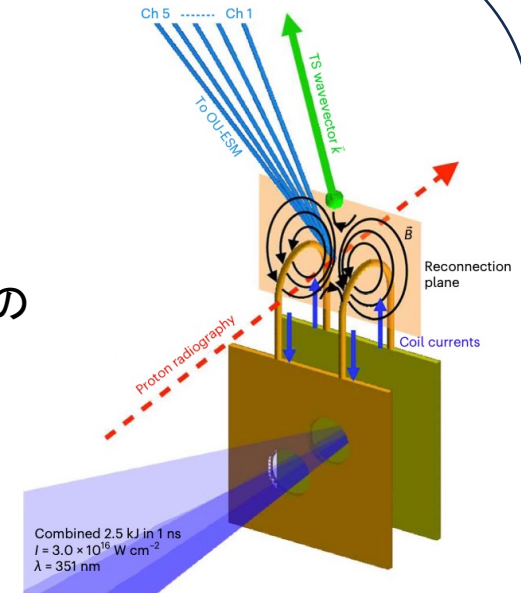
- ・全反射高速陽電子回折 (a),
反射高速電子回折(多数研究施設)
- ・低速陽電子回折(KEK), 低速電子回折(多数施設),
- ・表面X線回折(多数施設)
- ・オペランド偏光全反射蛍光-X線吸収微細構造
(PTRF-XAFS, 北大触媒研)(b)



プラズマ科学での活用(2023年~)

プラズマ核融合学会での例

- ・本講演内容(後述)
- ・小林達哉(NIFS)他
「荷電交換分光計測を用いた
位相空間揺らぎのベイズ推定」
- ・金史良(京大)他
「電子サイクロトロン放射計測への
ベイズ推定の適用」



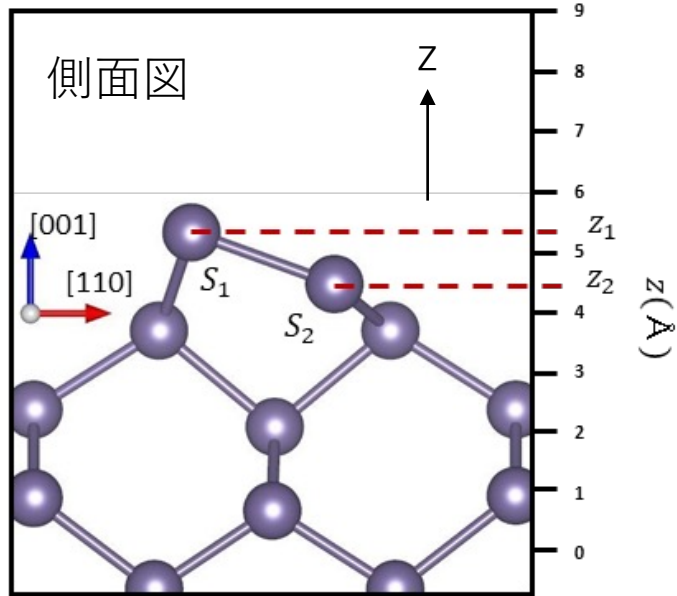
ベイズ事後確率分布のデモ

2変数での分布 $P(z_1, z_2|Y)$ プロット (不確かさ:(a)大, (b)小)

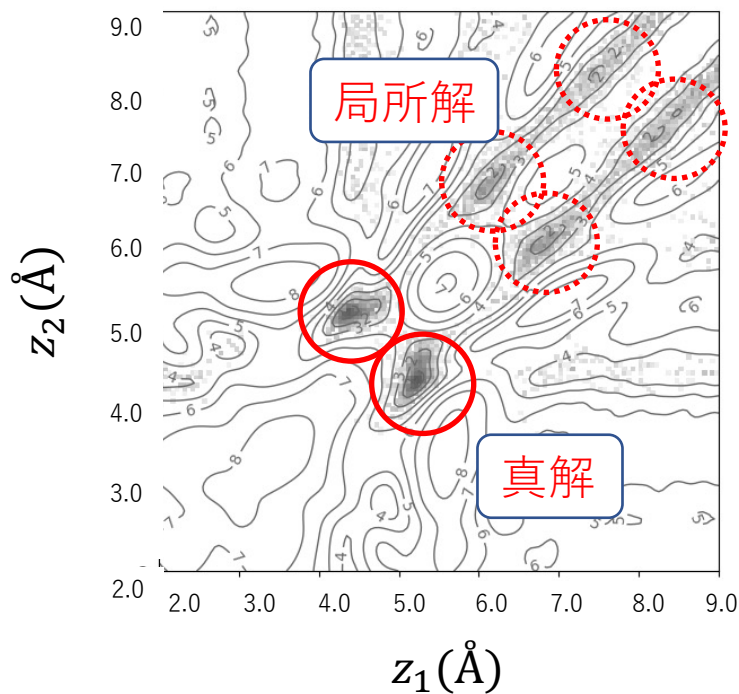
大域探索

→真解・局所解の検出

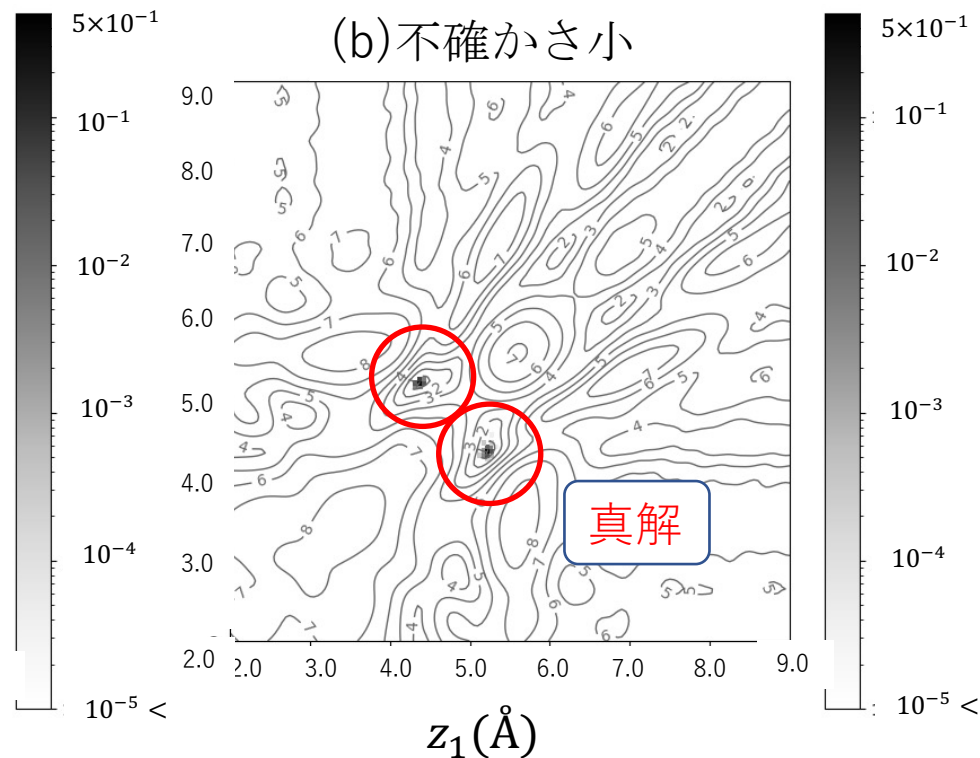
問題:陽電子回折による
半導体Ge表面原子座標 (z_1, z_2)



(a)不確かさ大



(b)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

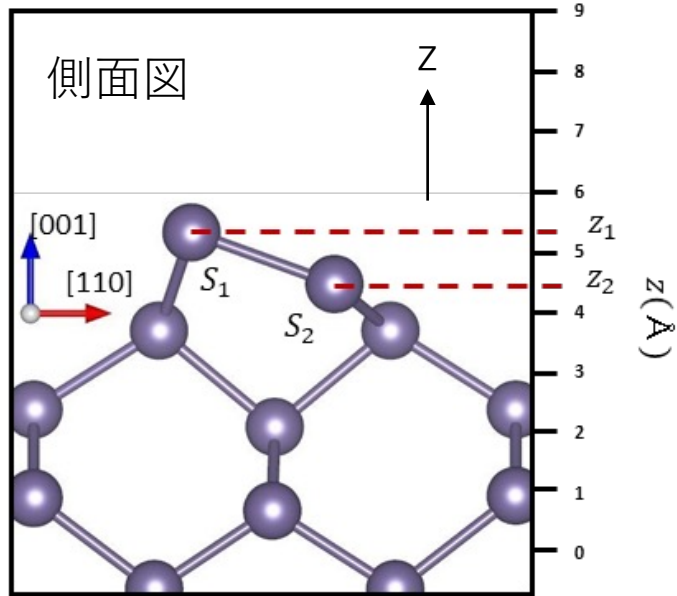
ベイズ事後確率分布のデモ

2変数での分布 $P(z_1, z_2|Y)$ のアニメーション(不確かさ:大→小)

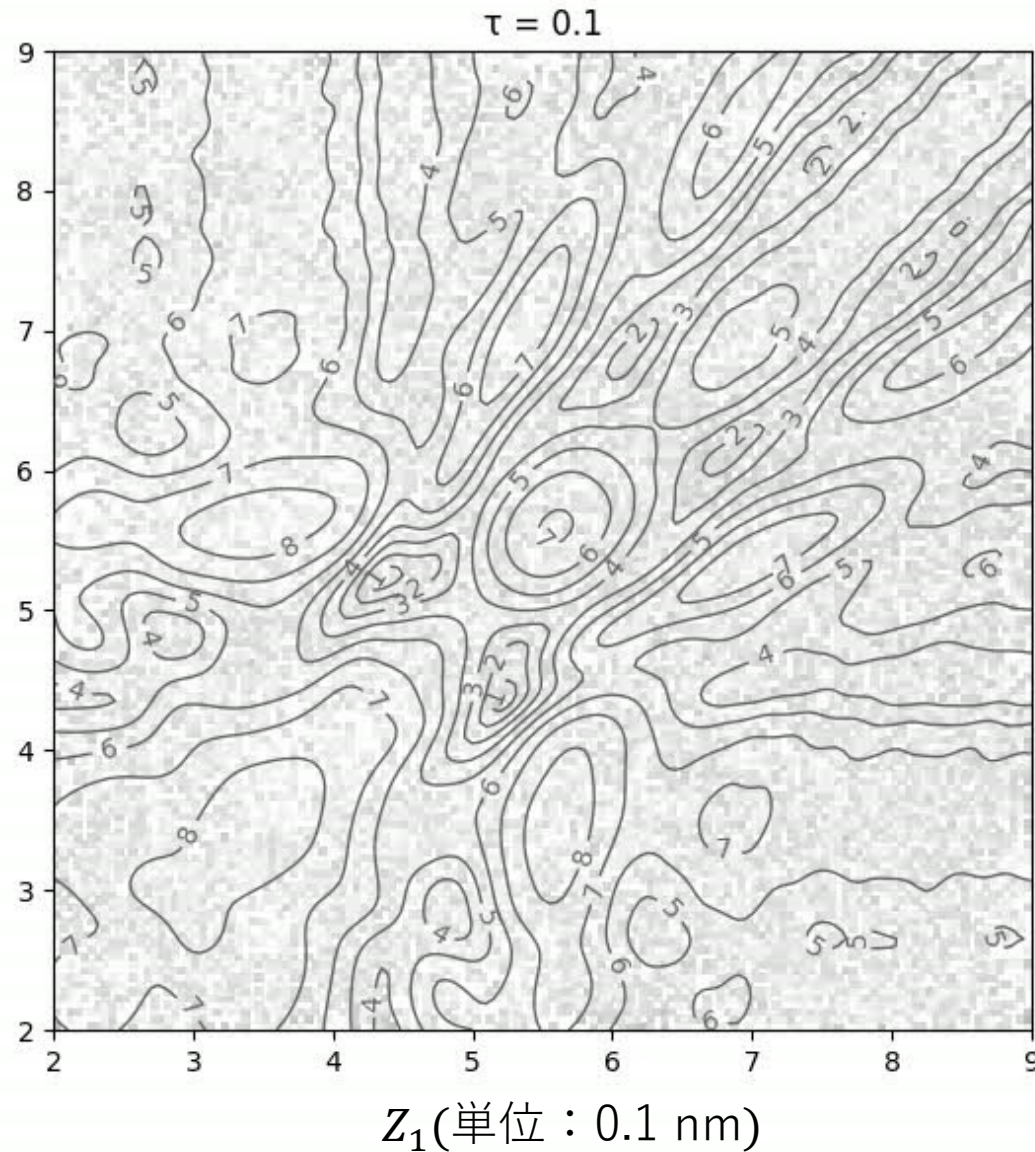
大域探索

→真解・局所解の検出

問題:陽電子回折による
半導体Ge表面原子座標 (z_1, z_2)



z_2 (単位: 0.1 nm)



実験室宇宙物理学への応用：非平衡プラズマ解析(予備的結果)

プラズマ核融合学会年会, 2024年11月(招待講演)

星健夫(1,2), 寺地雄真・中野陽斗(1,3), Lan Gao・Hantao Ji(4), 境健太郎(1)

(1)NIFS, (2)東大物性研, (3)鳥取大学, (4) Princeton Plasma Physics Laboratory

動機: 宇宙線(宇宙空間に存在する高エネルギープラズマ粒子)の起源

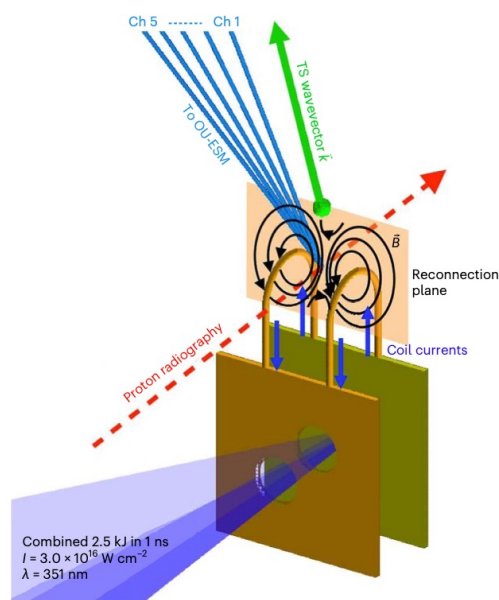
実験: 高出力レーザーをコイルターゲットに照射して発生する磁気リコネクション[1,2]の物理量の測定.

協同トムソン散乱(CTS)スペクトルを計測.

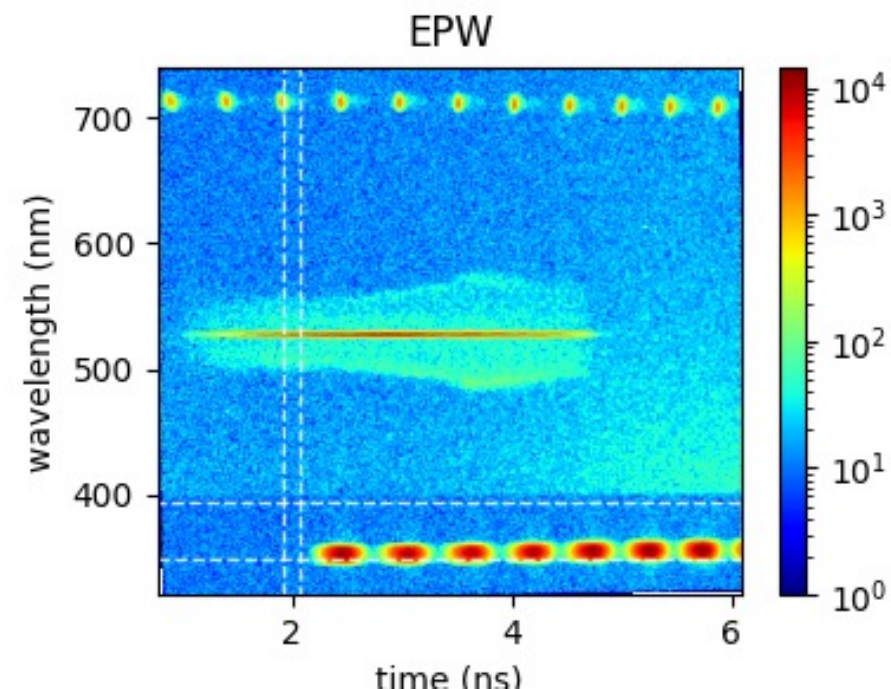
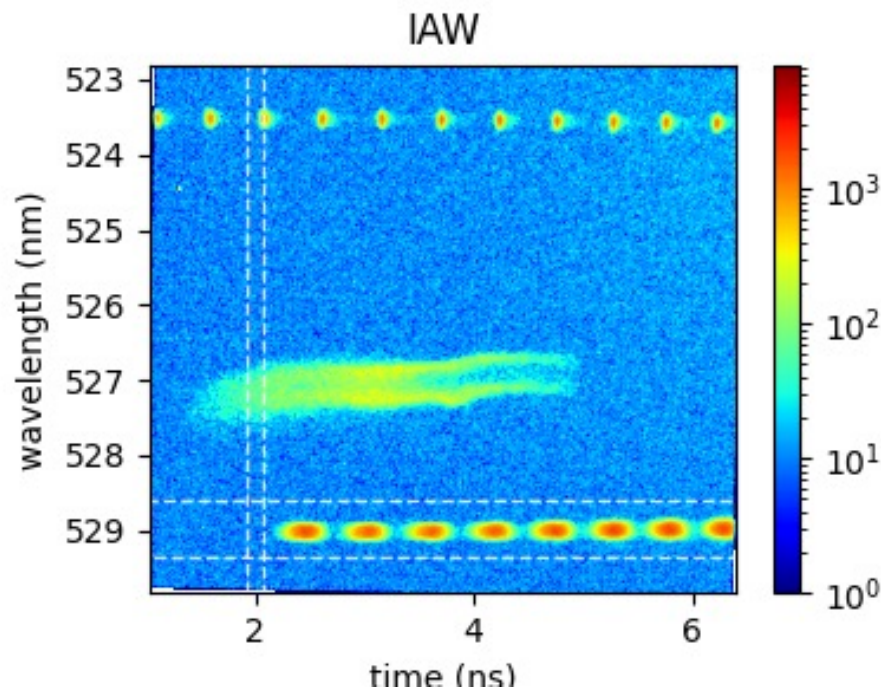
イオン音波 (IAW)と電子プラズマ波 (EPW)と共鳴した信号が得られる[2].

[1] Zhang, *et al.*, Nature Physics 19, 909 (2023). [2] Ji, *et al.*, Phys. Plasmas 31, 102112 (2024)

実験系模式図



(以下では, $t=2.9\text{ns}$ のデータを抽出)



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
<https://github.com/issp-center-dev/2DMAT>
- 5種のアプローチ(最適化, ベイズ推定)
- 超並列モンテカルロ法(PAMC法)による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.
- 物質科学, プラズマ科学(実験室宇宙物理)への応用例
- 一般的展望→次ページ

参考: ムーンショット目標10(フュージョンエネルギー)でも活用予定

プロジェクト名: 超次元状態エンジニアリングによる未来予測型デジタルシステム
代表 (Project Manager): 星健夫, 2024年12月~

ここから、新・未来へ



一般的展望

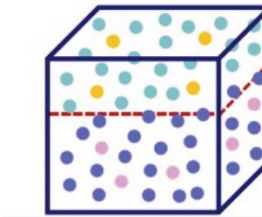
- ・データ解析フレームワークを共有した, 先端計測インフォマティクス拠点形成へ
- ・共通ニーズ例: マルチモーダル解析(異種データ統合)

NIFS: LHDデータ解析, および,
次世代装置設計での活用

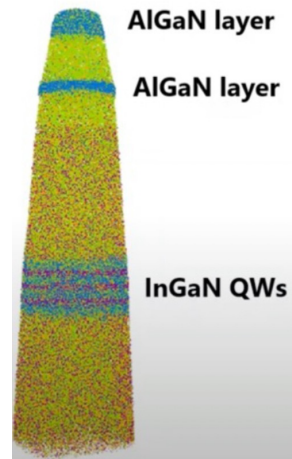
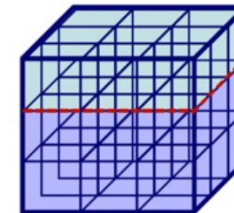
KEK物構研
:4種量子ビーム計測連携
(放射光・中性子・陽電子・
ミュオン)

半導体産業

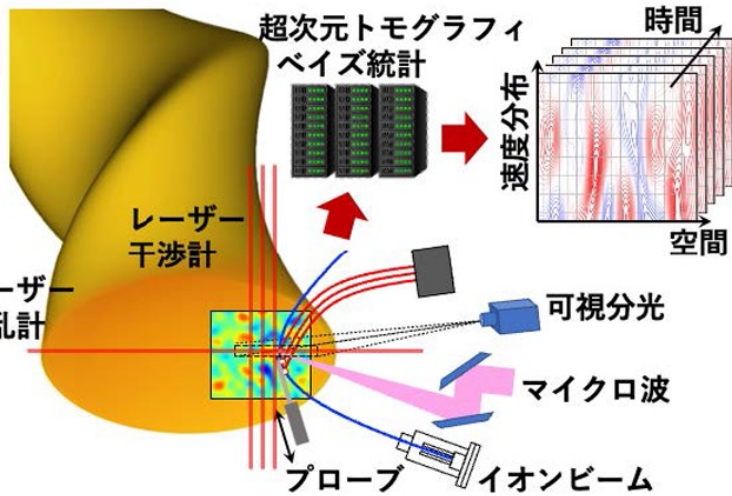
3D原子プローブ



電子線トモグラフィ



作図: 富谷茂隆 (NAIST)



作図: 小林達哉 (NIFS)

量子ビーム横断的試料搬送システム

