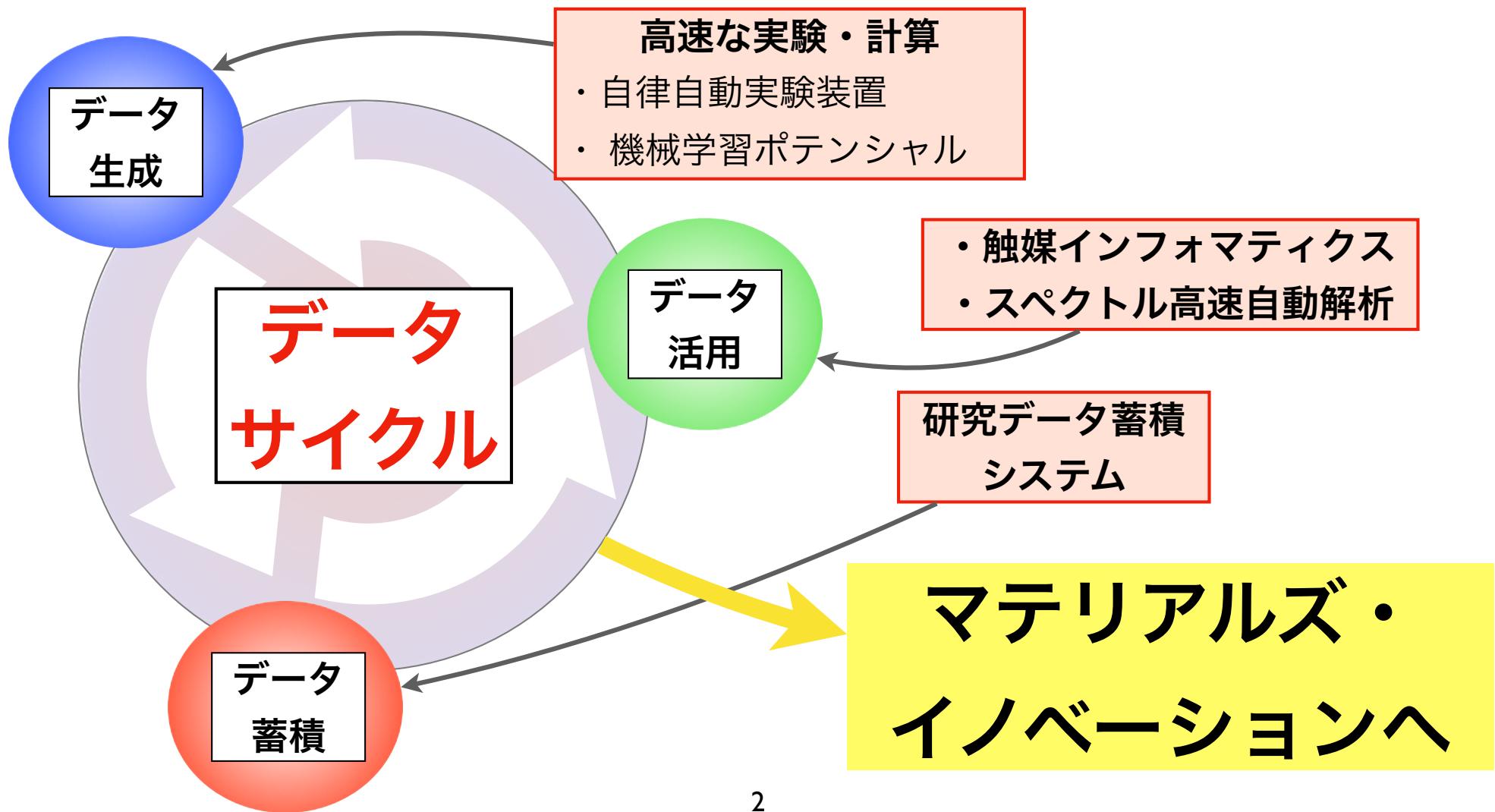


データ生成・蓄積・活用に基づいた マテリアルズ・インフォマティクスの実践

機能材料コンピュテーションナルデザインセンター
材料インフォマティクスチーム
安藤康伸

データ駆動材料研究

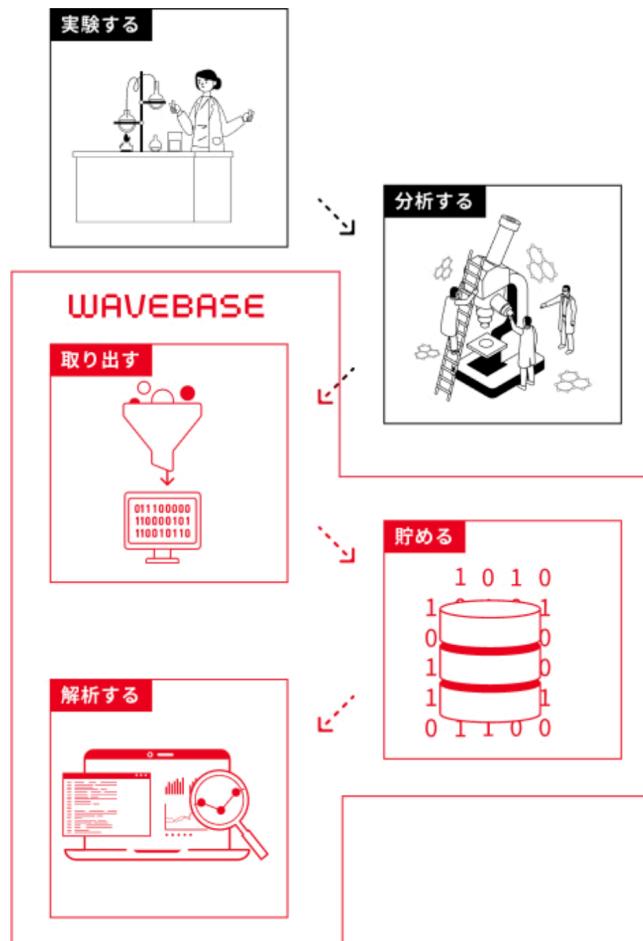
データサイクルを基盤とする新しい材料開発



トヨタ自動車の事業

WAVEBASE

プロジェクトヘッド：庄司 哲也



ABOUT

WAVEBASEとは

- ・データが揃わぬ統計的な処理ができない
 - ・しかし、データ活用の実績がなく、データ取得に投資が出来ない
- という負のループに陥っていないでしょうか？
- 我々は、多種多様な素材を高いレベルで要する自動車開発の中で培った、限られたデータの中からいかに有効な情報を取り出すか、という点に強み持っています。
- まずは、少量のデータで想像以上のデータ活用のメリットを体感頂き、データ活用の文化を醸成すると共に、技術者が従来時間を取られていたデータ解析から開放され、ポテンシャルを最大限引き出すことが出来る環境を、お客様と共に創造します。

[もっと見る](#)

研究開発データ管理SaaS



現場が感じる強い課題感

研究開発データの管理や取り扱いに
課題を感じていますか？

はい
98%

研究者を対象としたアンケート結果（当社調べ）より

代表：斎藤 耕太郎（中性子実験の専門家が企業）

バラバラだった実験、チーム、
データが整備され、繋がることで新しい価値を生む

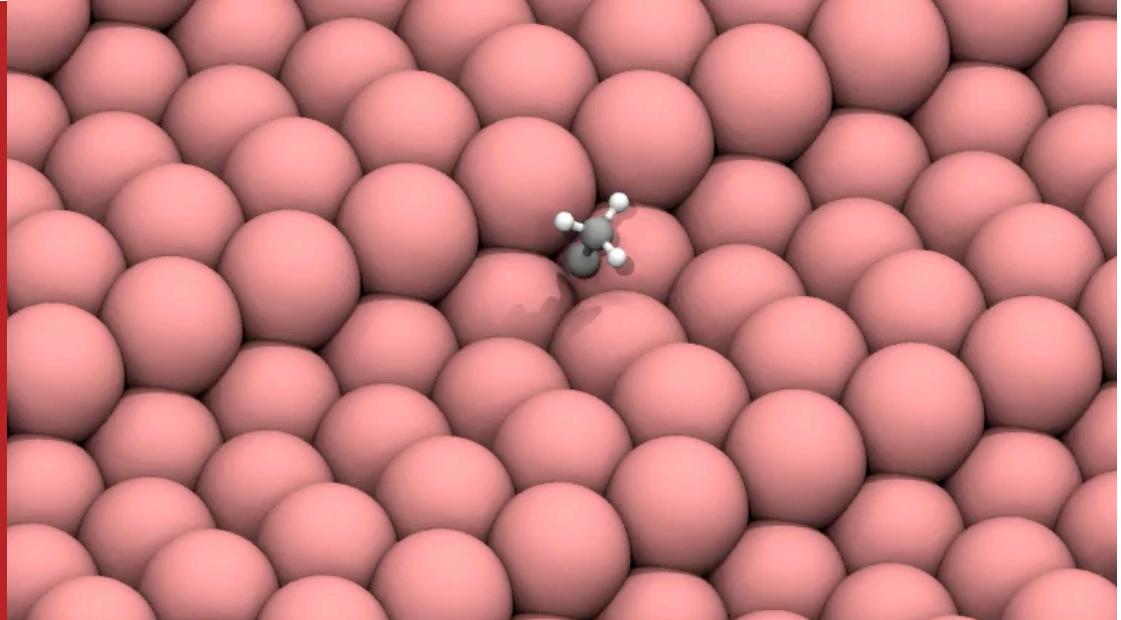


ひとつひとつの実験、ひとりひとりの研究者の成果が、Randeftによるデータジャーニーの変化によって、アップグレードされるだけでなく、その全てが整備され繋がることで、今までになかった分析、実験そのものの効率化、チームのコラボレーションが生まれ、新しい価値の創造につながります。

マテリアルズ・インフォマティクスとその先に生まれる新たなデータ駆動型アプローチにいち早く対応するための基盤を提供します。

引用元：<https://www.randeft.jp/product>

Open Catalyst Project



- ・Meta AI社のFundamental AI Research (FAIR) とカーネギーメロン大学 (CMU) の共同研究
- ・目的は、気候変動への対応に役立つ再生可能エネルギー貯蔵に使用する新しい触媒を、AIを使ってモデル化し発見すること
- ・AIや機械学習を用いることでDFT計算を効率的に近似し効果的な触媒を見つけることができるかもしれない
- ・**MLモデル学習用のOpen Catalyst 2020 (OC20) および2022 (OC22) データセットを公開**
- ・2億6千万回以上のDFT計算による130万個の分子緩和が含まれている
- ・データだけでなく、ベースラインモデルとコードもGithubページでオープンソース化

オンラインサービスMatlantis

Preferred NetworkとENEOSが出資

<https://matlantis.com/ja/>

The screenshot shows the Matlantis website homepage with a dark blue background featuring abstract molecular structures. At the top left is the Matlantis logo. The top navigation bar includes links for Product, Cases, Company, Login, and language switches (JA, EN). A large orange button on the right says "お問い合わせ" with an envelope icon. The main headline reads: "革新的なマテリアルの創出に貢献し、持続可能な世界を実現するために「Matlantis™」は生まれました。" Below the headline is a paragraph about the company's mission and technology, followed by a call-to-action button: "お問い合わせ・ご相談はこちら →".

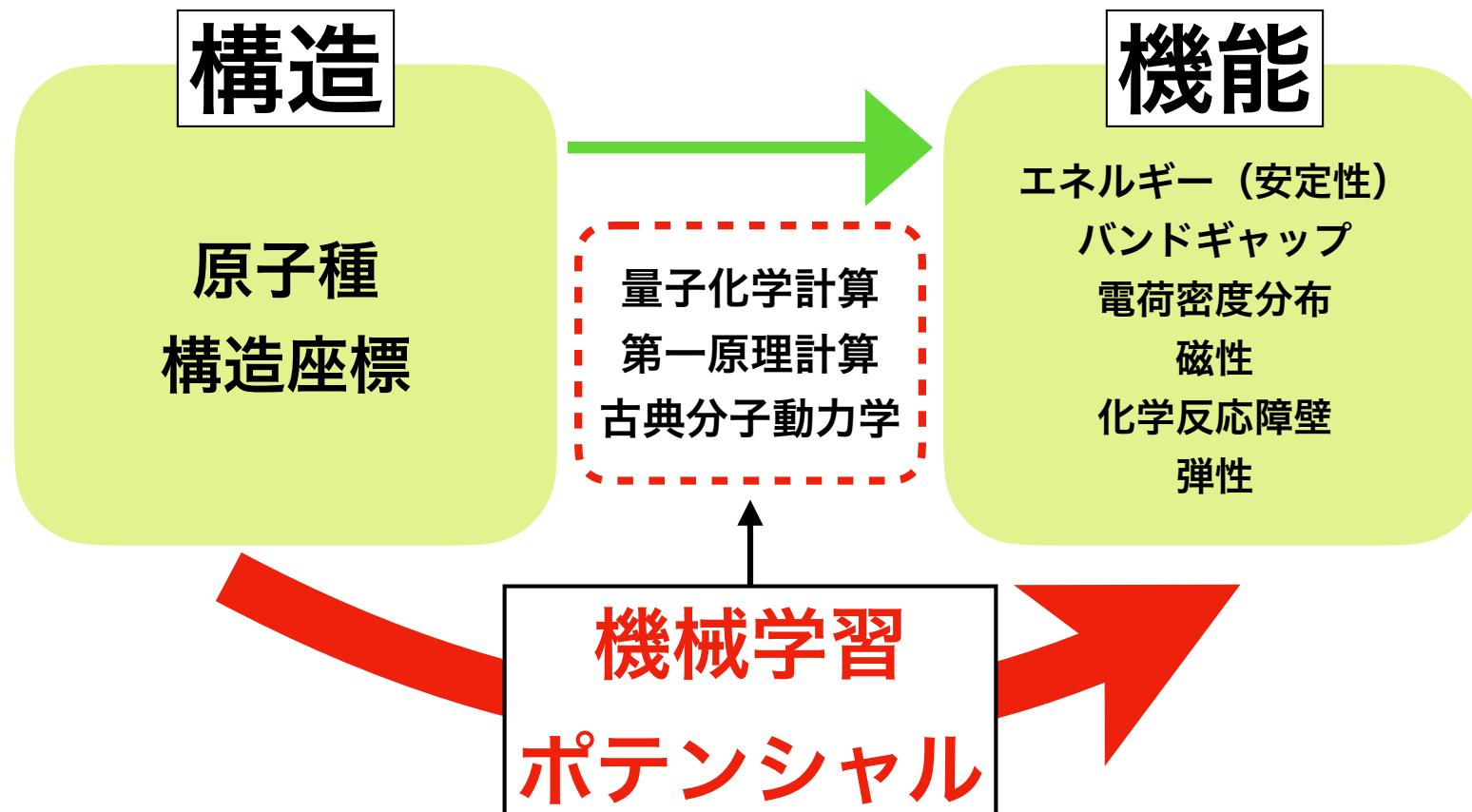
55種類の元素を取り扱える汎用NNPを提供



データ生成

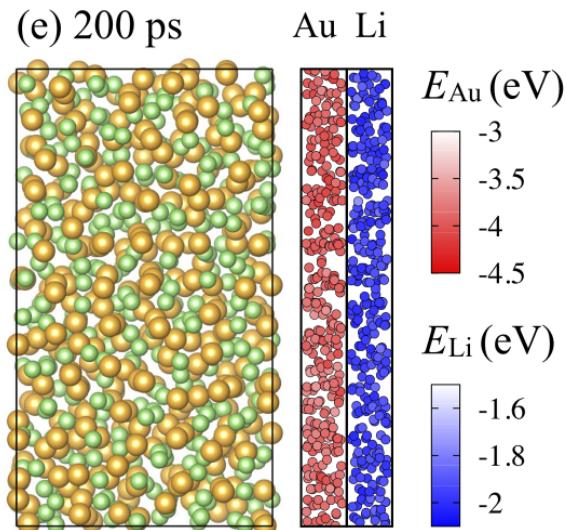
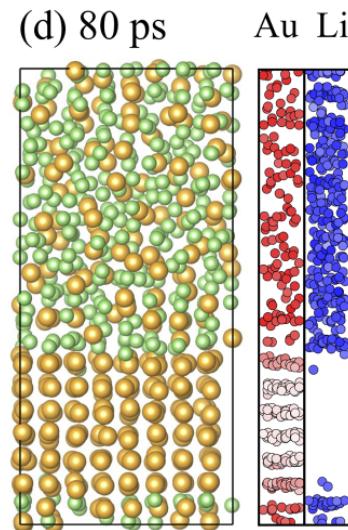
機械学習ポテンシャル

物質シミュレーションを機械学習で代替・高速化する



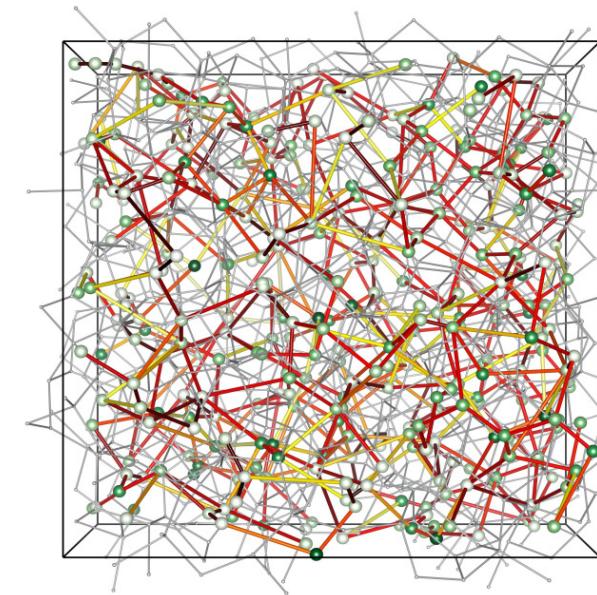
従来できなかった規模・精度のシミュレーションを実現可能に

アモルファス・合金化に関する研究



Au-Li binary alloy

K. Shimizu, E. F. Arguelles, W. Li, Y. Ando, E. Minamitani, and S. Watanabe, Phys. Rev. B **103**, 094112 (2021).



Li diffusion mechanism in amorphous Si

W. Li, and Y. Ando, PHYSICAL REVIEW MATERIALS **4**, 045602 (2020).

Review Paper

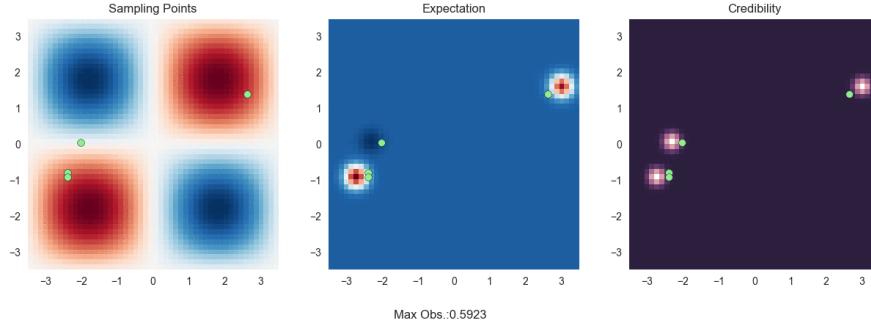
ACCEPTED MANUSCRIPT ·
THE FOLLOWING ARTICLE IS
OPEN ACCESS

High-dimensional neural network atomic potentials for examining energy materials: Some recent simulations

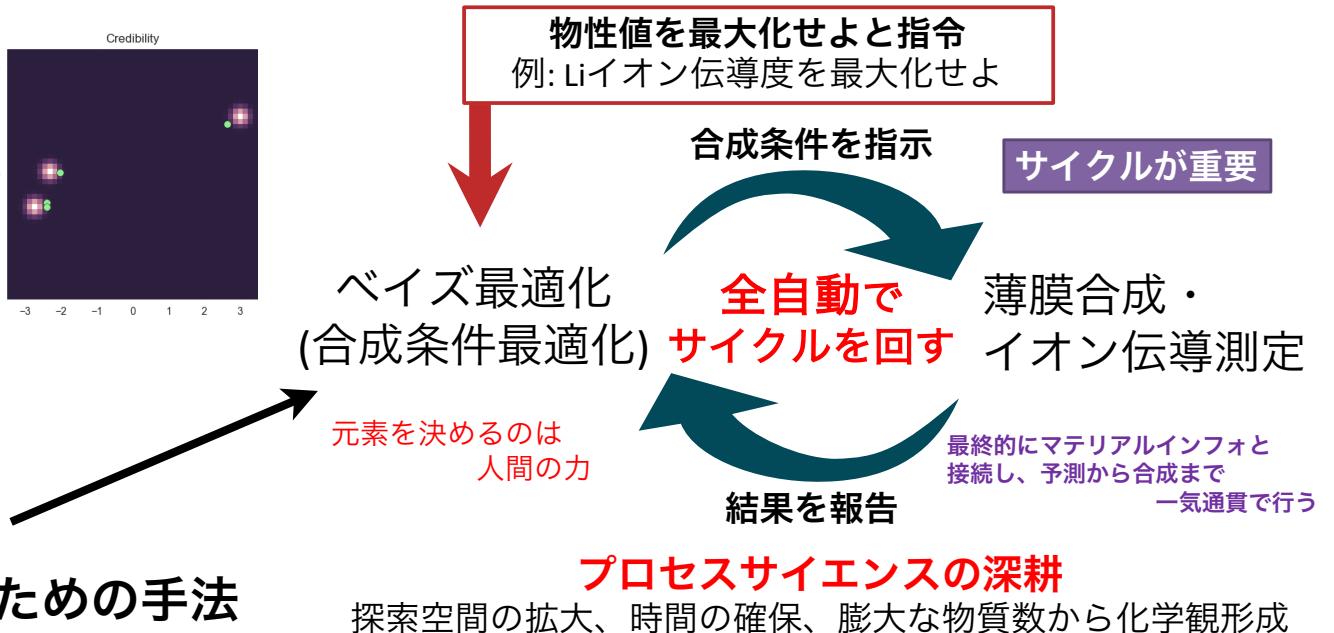
Satoshi Watanabe¹, Wenwen Li², Wonseok Jeong³, Dongheon Lee⁴, Koji Shimizu⁵, Emi Minamitani⁶, Yasunobu Ando² and Seungwu Han³
Accepted Manuscript online 5 November 2020 · © 2020 The Author(s). Published by IOP Publishing Ltd

S. Watanabe, W. Li, W. Jeong, D. Lee, K. Shimizu, E. Mimanitani, Y. Ando, and S. Han, J. Phys. Energy **3**, 012003 (2021).

自律自動実験装置の開発



少ない手数で
最適パラメータを発見するための手法



ハードウェアの価値を高める
ソフトウェア開発の実施



材料合成プロセスの柔軟な最適化の実現

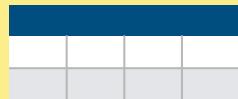
- ベイズ最適化のプログラム開発
- 計測データの高速情報抽出
- データベース構築やデータフローのデザイン
- 実験室におけるIT環境の整備
- 「On-The-Fly解析」による計測モニタリング



データ
蓄積

情報科学の「当たり前」を材料研究へ

テーブル構造

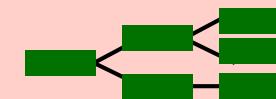


```
angle [degree], count, time[s]
 0, 0, 0.20
 10, 23, 0.19
 20, 110, 0.21
 30, 132, 0.21
 40, 90, 0.20
 50, 20, 0.17
```

CSV

(Comma Separated Values)

ツリー構造



```
<?xml encoding='UTF-8'?>
<experiment>
  <sample id=1>
    <name>BN</name>
    <method>XPS</method>
  </sample >
</experiment>
```

XML
(Extensible Markup Language)

```
{
  "density_unit": "1/cell",
  "H2O": {
    "density": 22,
    "Molecule": "H2O.spc.MOL"
  }
}
```

JSON

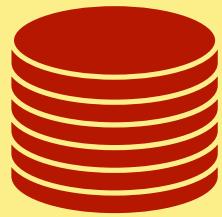
(JavaScript Object Notation)

従来使われてこなかったXML/JSONといった
フォーマットを材料データ蓄積に利用

データベースソフトを開発中

データベース構築目的の 3C

蓄積 (collection)



利用できるよう整理
整頓して収納する

索展 (curation) 理解 (comprehension)



テーマに沿って
分かりやすくまとめる



論証に対する根拠を
取り出す

サイズ

バリエーション

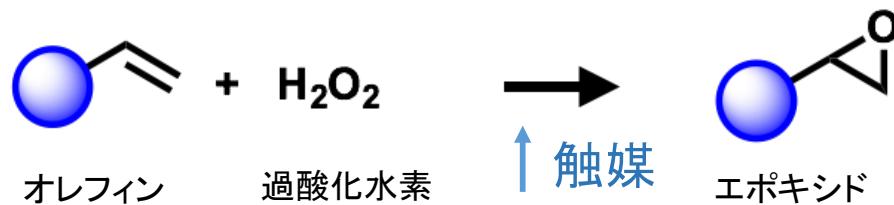
課題・環境に応じて必要な「データベース」をより深く考える必要がある



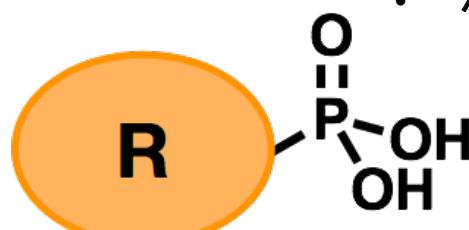
データ
活用

希少データの価値を最大限活かす

高活性ホスホン酸触媒をシミュレーションからデザインすることに成功（矢田・安藤・永田）



- タングステン触媒
- 相間移動触媒
- ホスホン酸触媒

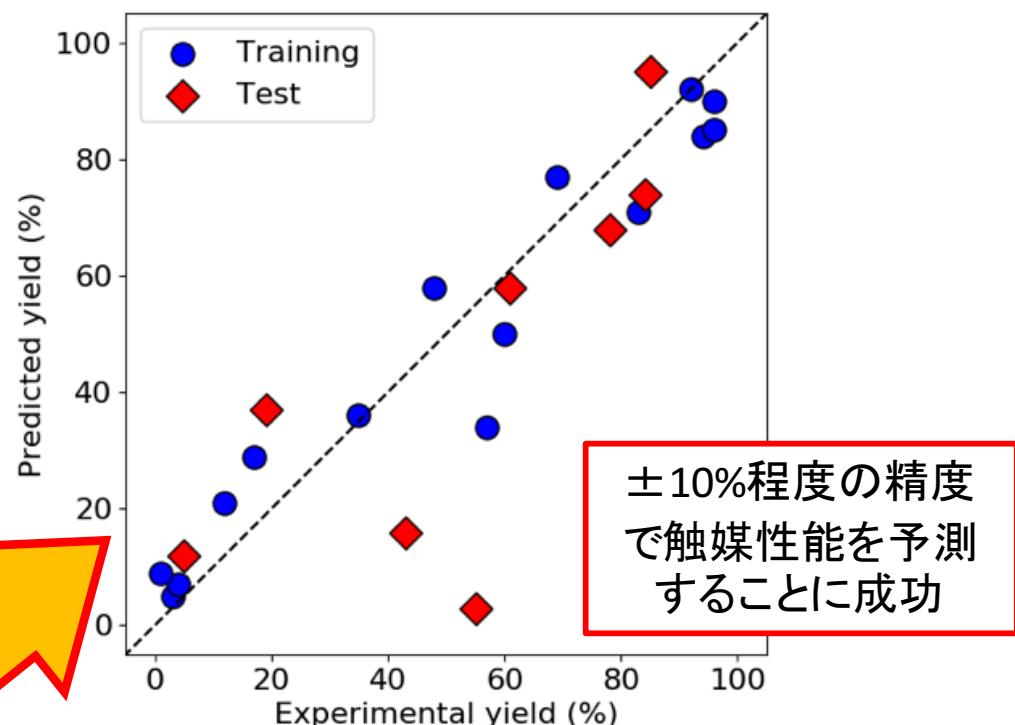


エポキシド高収率な
ホスホン酸の側鎖R
を予測したい

量子化学計算
によって触媒を
数値化

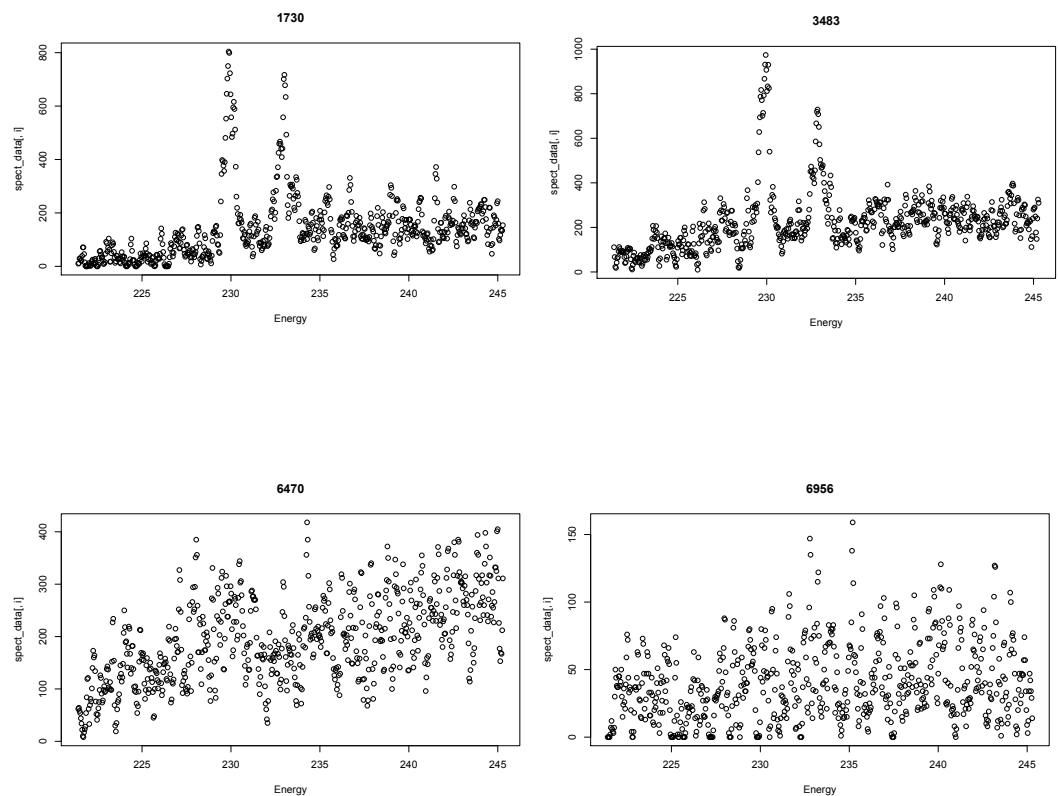
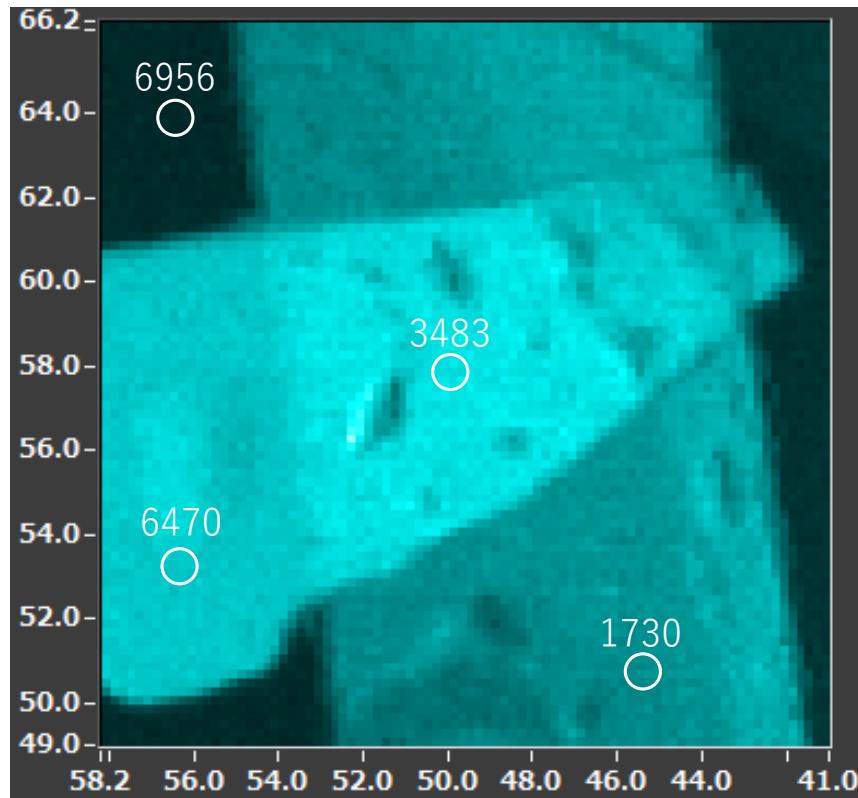
実験によって
触媒性能を評
価

AI技術で関連付)



- ・実験をせずに高収率触媒を予測・性能を実証
- ・専門家の勘を再現することに成功
- ・AI, 機能材料, 触媒科学3センター連携による成果

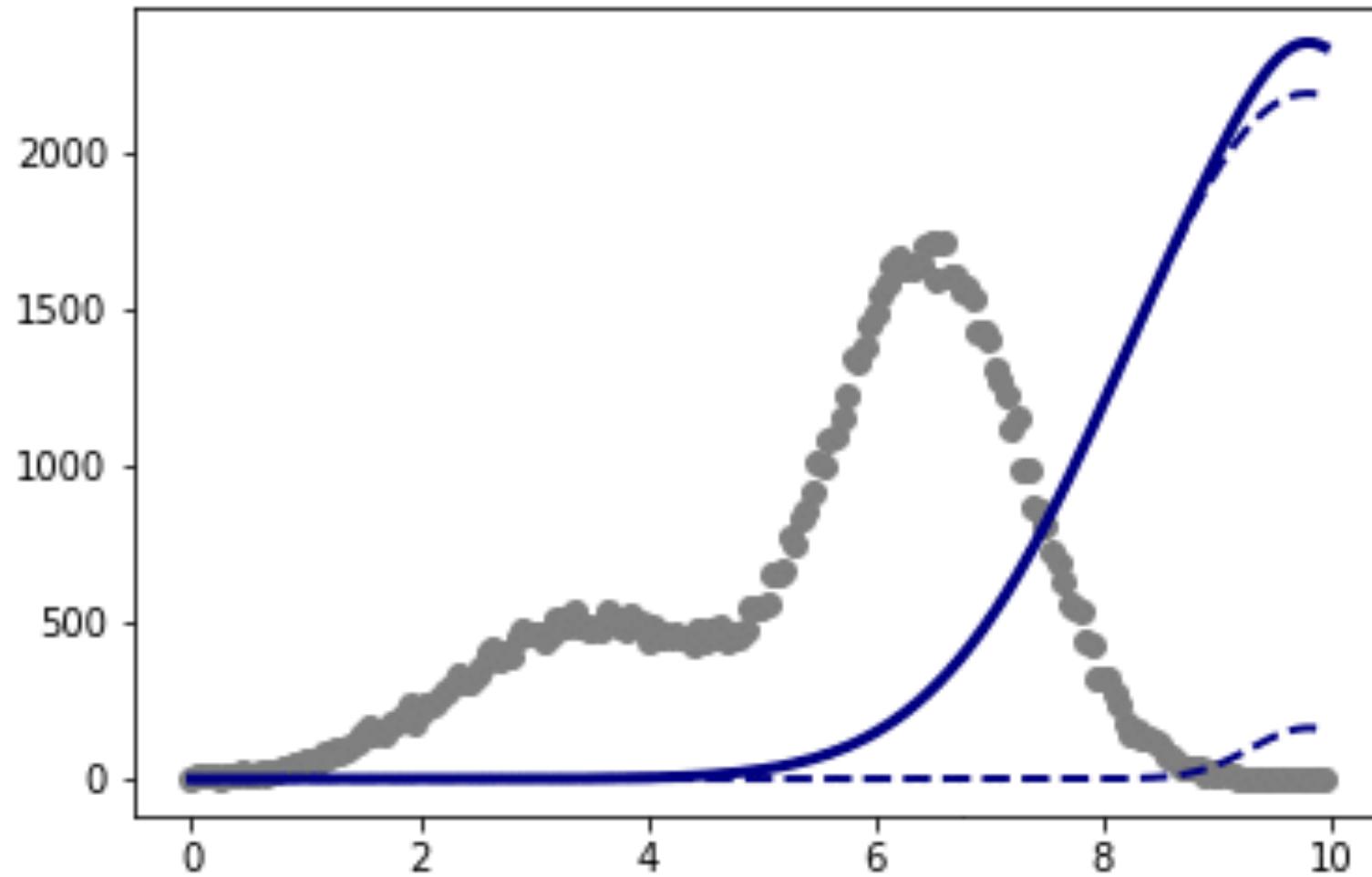
膨大な量の解析を実現する



86 x 86 pixelsの7396 点全てに対応する"スペクトルデータがある"

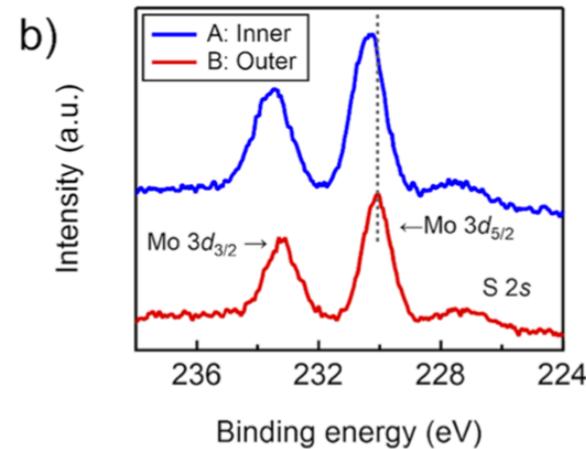
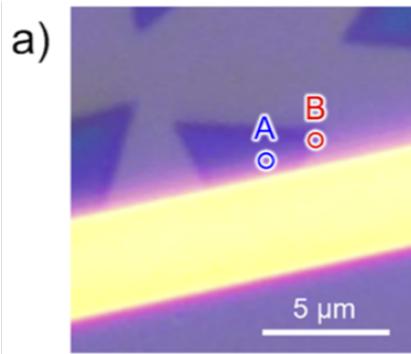
→ 従来とは異なる処理方法が必要不可欠

EMアルゴリズムによる自動ピーク検知

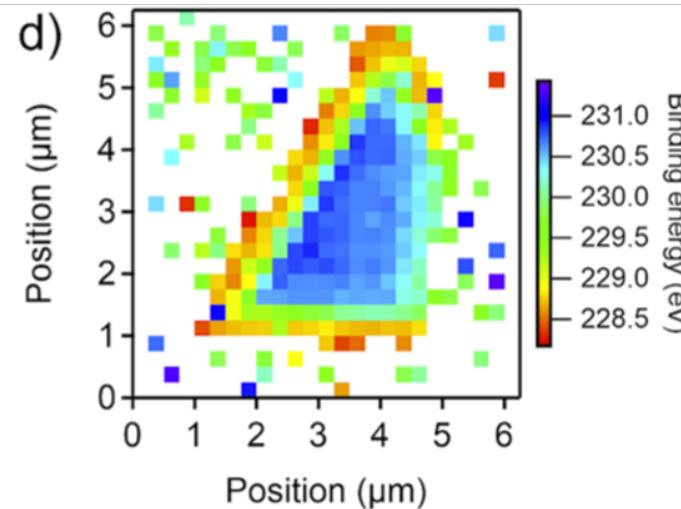
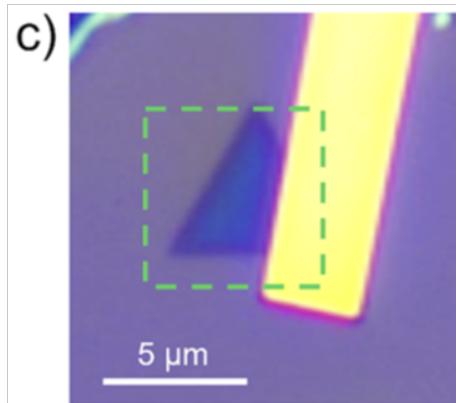
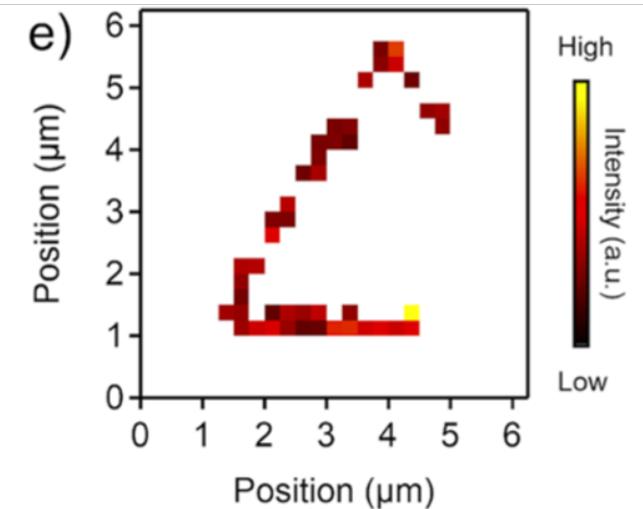


チューニングの必要なく～1s 程度の処理を実現

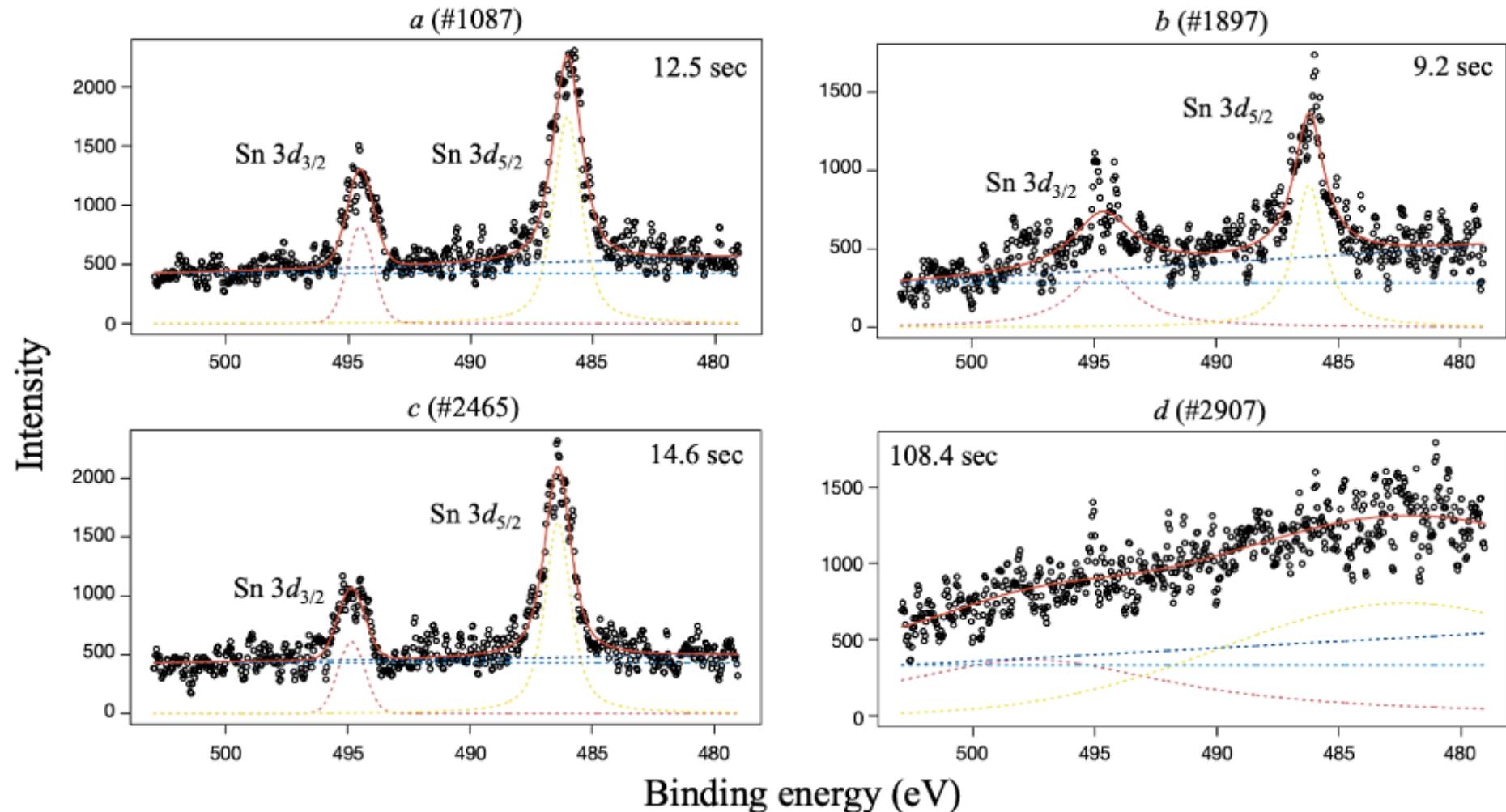
応用：X線顕微分光データ解析



- MoS₂-MoS₂:Nb への3D nano-ESCAで得られたスペクトルデータに提案手法を適用
- Mo $3d_{5/2}$ ピークマッピングからエッジ方向へのBinding energyの減少トレンドを検出
- 625セットのスペクトルデータを解析(6ピーク分解)
並列処理で実質6~7時間で解析可能
- ベースライン処理を含めた一連の解析作業を自動化

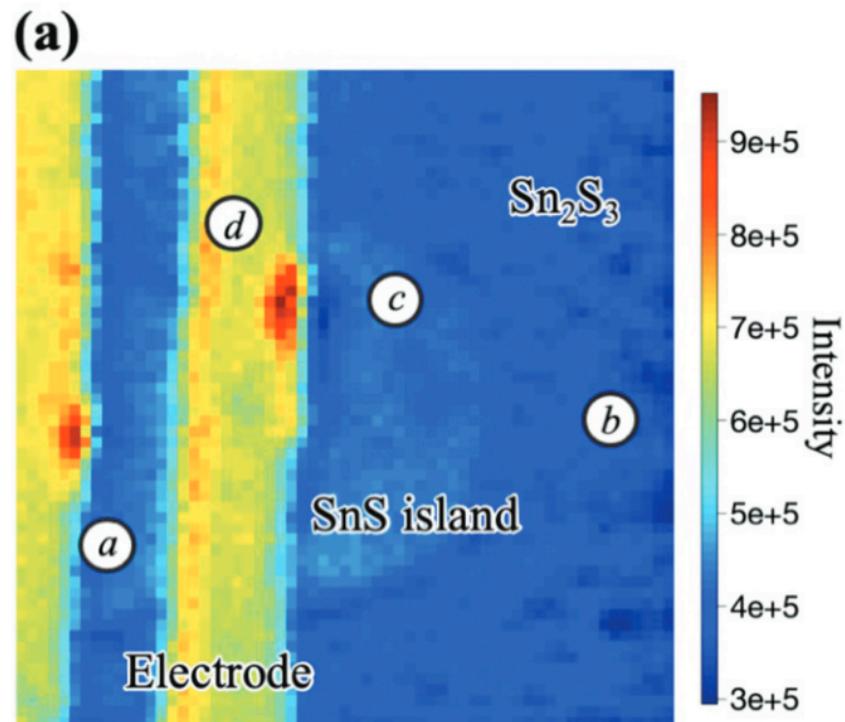
Peak-position mapping image of Mo⁴⁺ 3d_{5/2}Peak intensity mapping image of Nb⁴⁺ 3d_{5/2}.

線形バックグラウンドモデリング

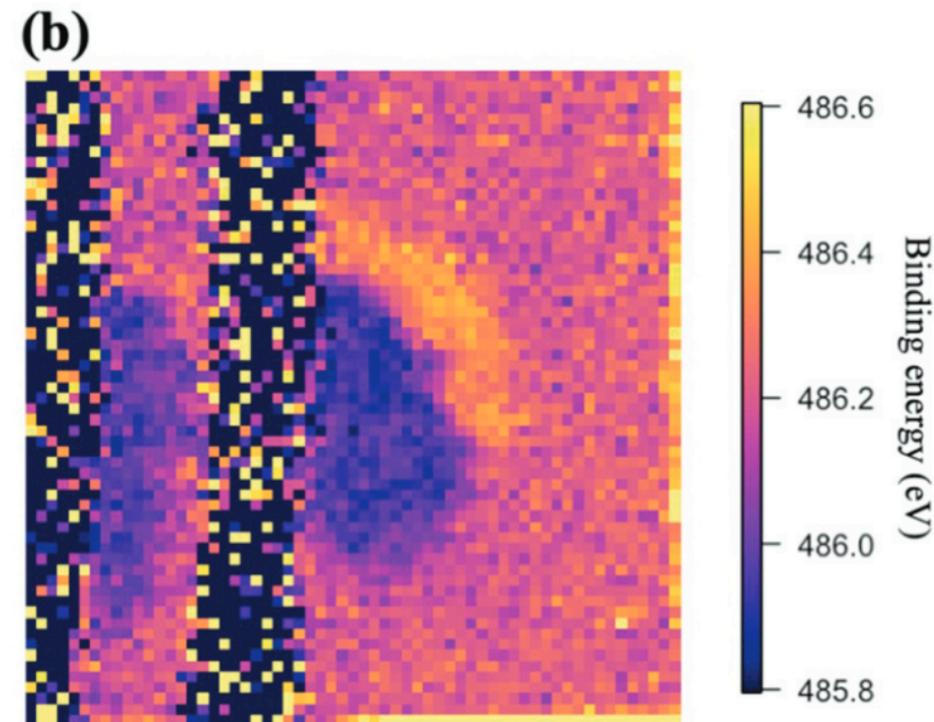


区間一様分布・三角分布の尤度を考慮することで自動モデリングに成功

スペクトルモデリングの応用結果



強度マップ



XPSピーク位置マップ

0.4 eV程度のケミカルシフトを検出・Snの価数分布のマッピングに成功

まとめ：データ駆動科学で実現したいこと

計算・機械学習を駆使して「現代」でしかできない研究をしたい

- 100年前に天才はたくさんいたがコンピュータはない
- 新しいことは今の時代できることから生まれる

優秀なソフト・ハードの価値をもっと引き出したい

- すごい実験装置やシミュレーターはたくさん開発されている
- そこで取得したデータをどこまで活用できているだろうか？

作業効率を上げて”研究”に割く時間を増やしたい

- 24時間つきっきりの実験現場を楽にしたい
- 手作業で淡々と行われる解析作業を自動化したい