

2024.10.08  
研究データエコシステム構築事業シンポジウム2024



OU MASTER PLAN  
2027  
生きがいを育む社会の創造

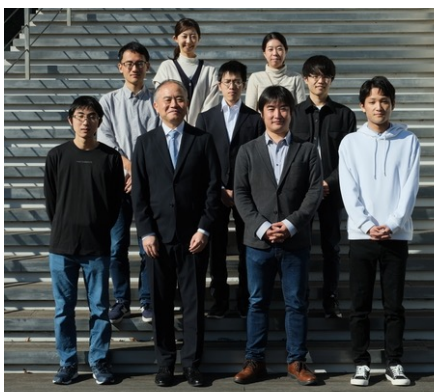
# 大規模実験データの計測・解析・共有・公開を通じた知の創出のためのエコシステム構築

大阪大学 大学院工学研究科 小野 寛太

# 大阪大学 大学院工学研究科 物理学系専攻 応用物理学コース 先端物性工学領域



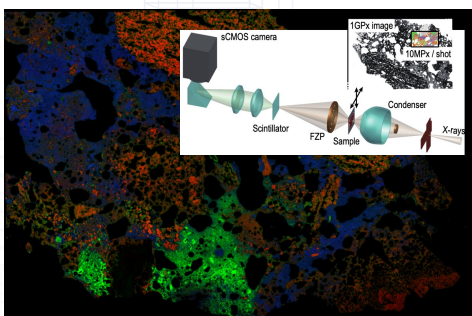
OU MASTER PLAN  
2027  
生きがいを育む社会の創造



## 工学から新しい物理を生み出す

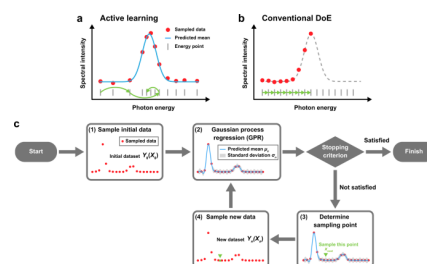
小野寛太  
武市泰男  
博士課程 1  
修士課程 6  
学部学生 3

### 研究テーマ：物性研究の自律化による複雑な材料特性メカニズムの解明



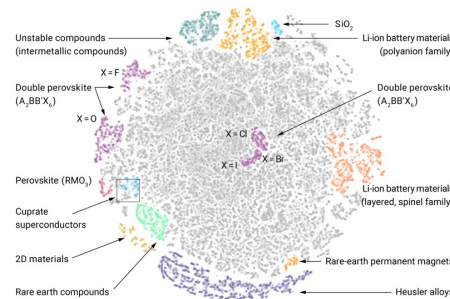
#### X線顕微鏡

広視野・高分解能観察  
可能な新しい顕微鏡の  
開発



#### 計測インフォマティクス

計測の最適化・データ解  
析の自動化で最先端計測  
を目指す



#### マテリアルズ インフォマティクス

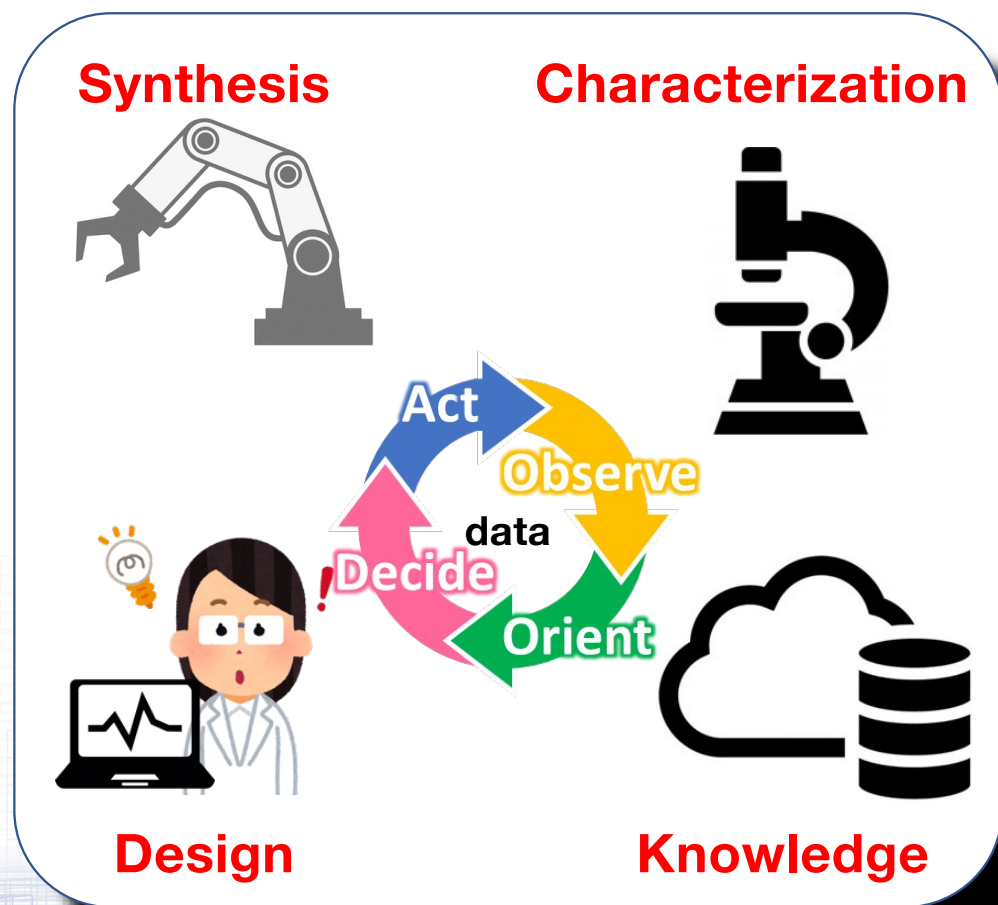
AI 技術を活用した材料研究  
物質空間の地図を作る



#### 物性研究自律化

AI x ロボットによる  
材料研究の自律化

# 物質科学における科学研究の自律化



# 世界の自律実験装置の現状

## A mobile robotic chemist

B. Burger et al., Nature 583, 237–241(2020)



## An autonomous laboratory for the accelerated synthesis of novel materials

N. J. Szymanski et al., Nature 624, 86-91(2023)

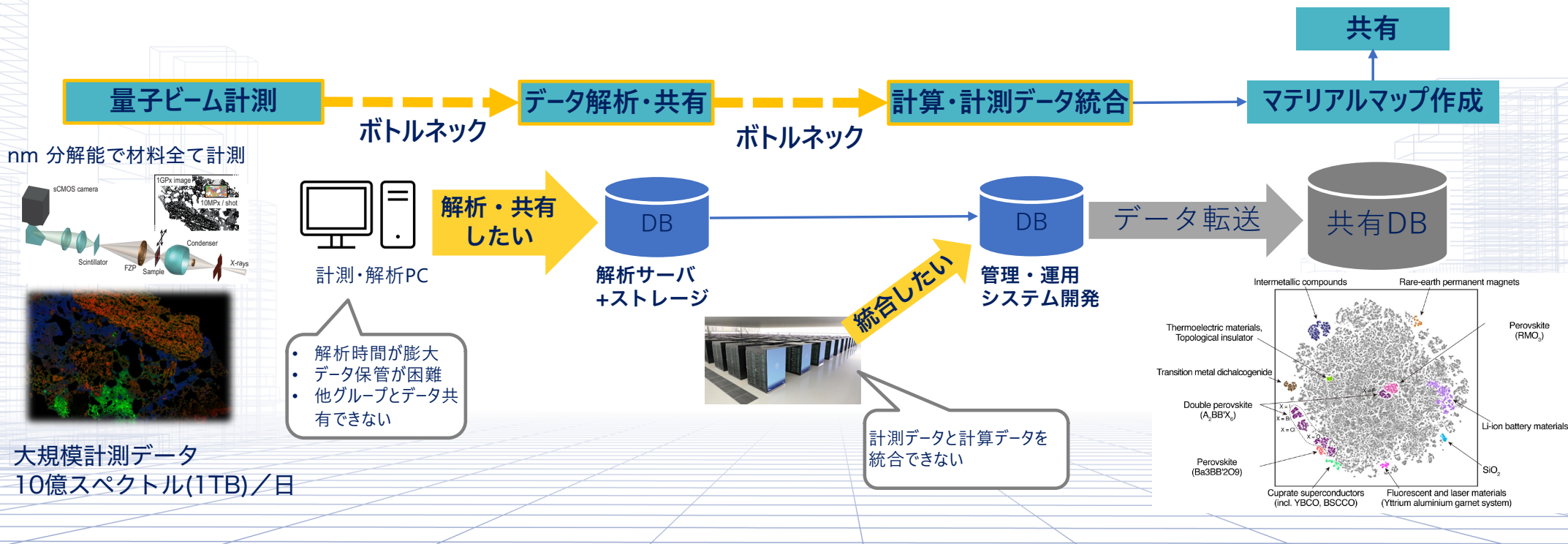


世界では自律実験装置（Self-driving Laboratory）の研究が進んでいる

# 知の創出のためのエコシステム データの計測・解析・共有・公開

研究のボトルネック：大規模計測データの解析・共有が出来ず、共同研究を通じた知の創出が進まない

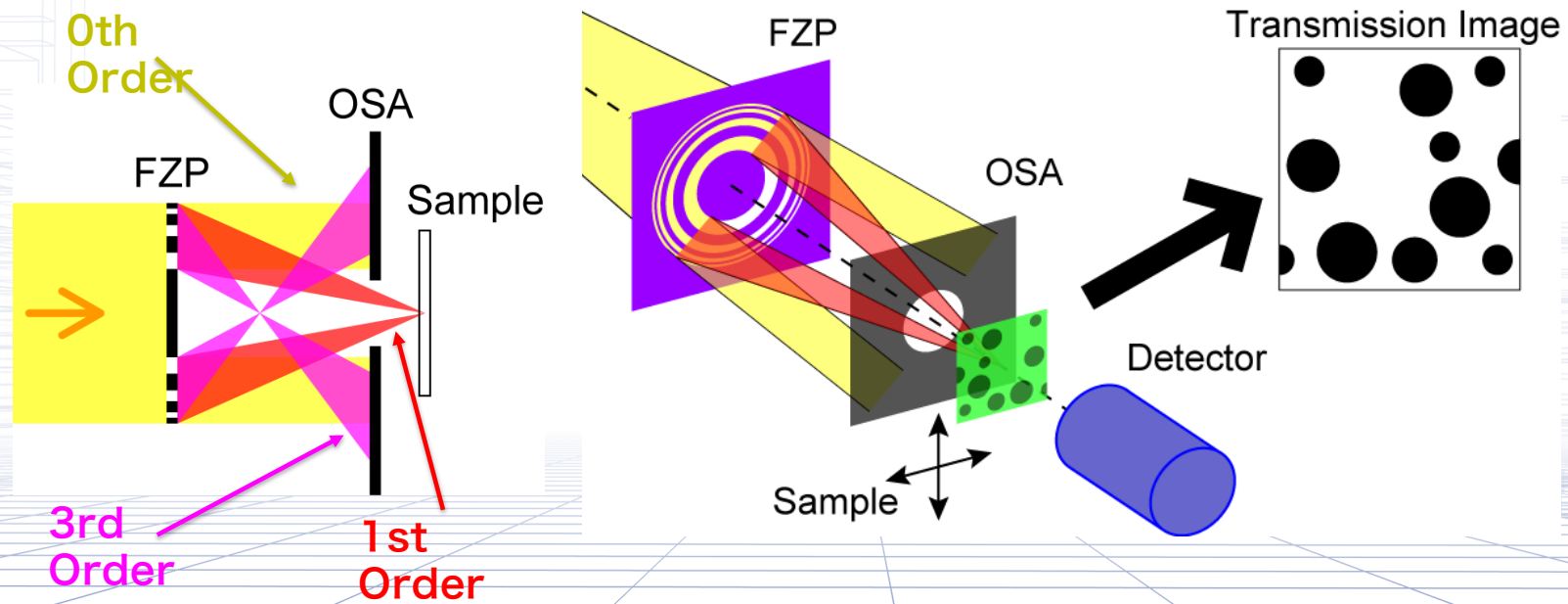
大阪大学の量子ビーム計測技術により、材料の大規模（10億スペクトル：～TB）計測データが創出されるようになったが、解析・共有システムが整備されていないため、計算データとの連携やデータ活用が困難



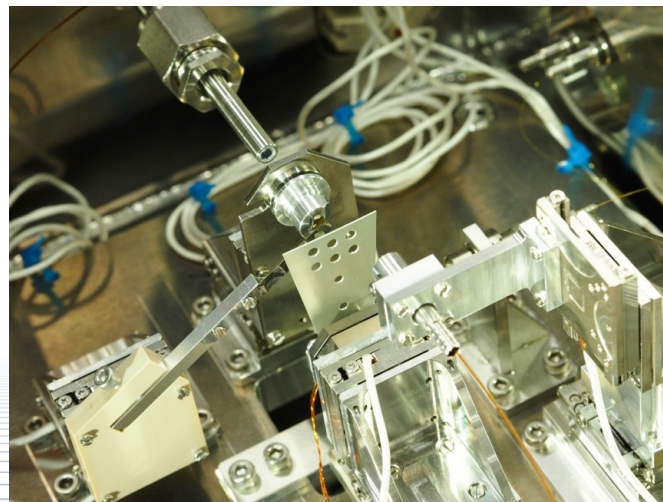
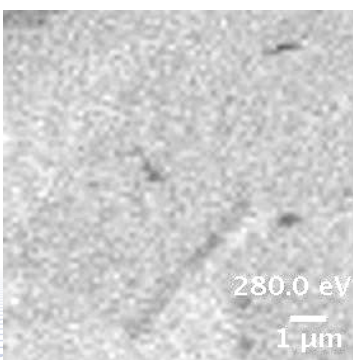
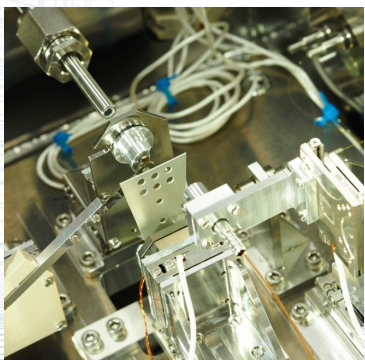
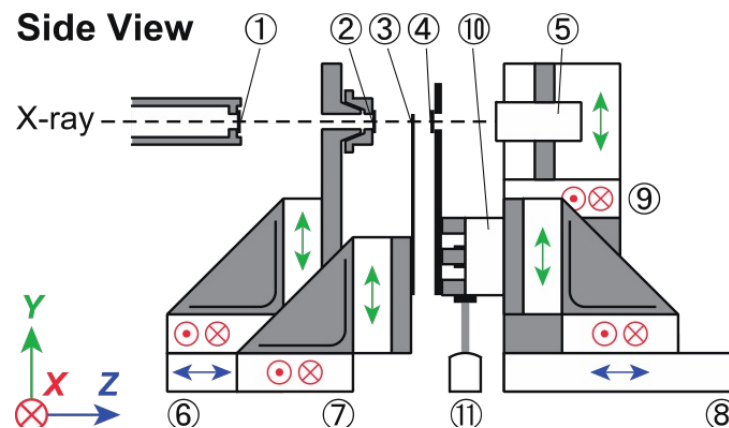
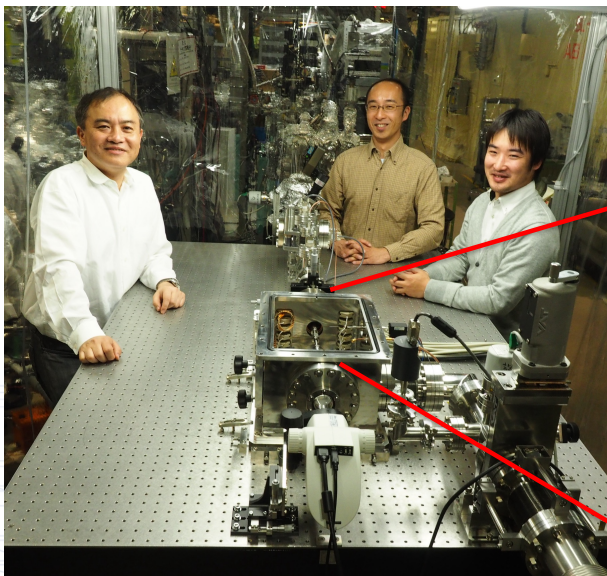
# 走査型透過X線顕微鏡

## STXM (Scanning Transmission X-ray Microscopy)

- ゾーンプレート (FZP) でX線を集光、ピンホール (OSA) で1次回折光のみを選択
- 試料位置をスキャンし、透過X線を検出することで像を得る
- 空間分解能は10 ~ 30 ナノメートル程度

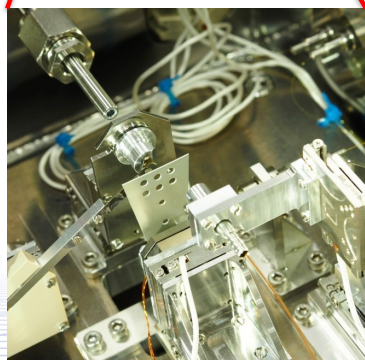
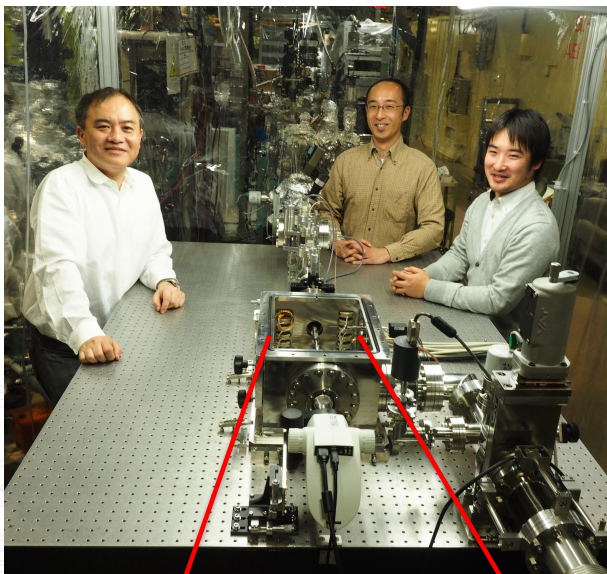


# これまで開発したX線顕微鏡

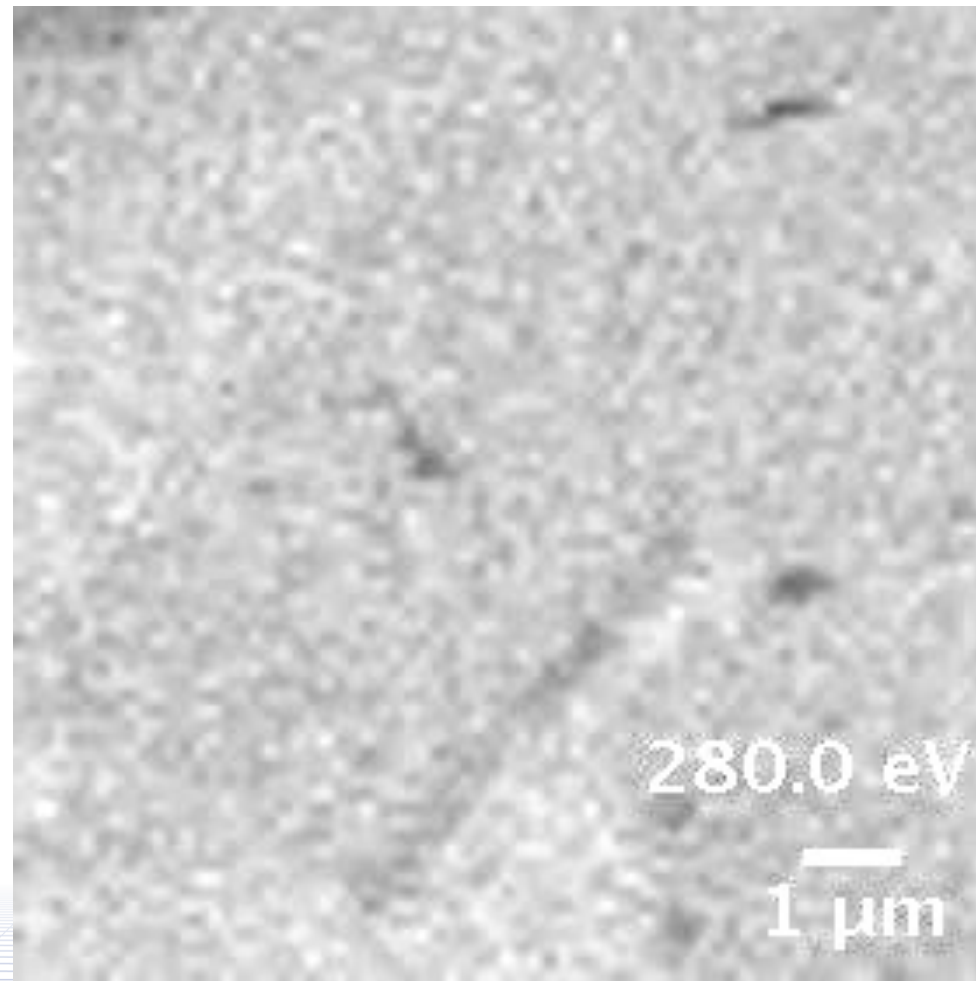


これまでに開発したX線顕微鏡

# これまで開発したX線顕微鏡

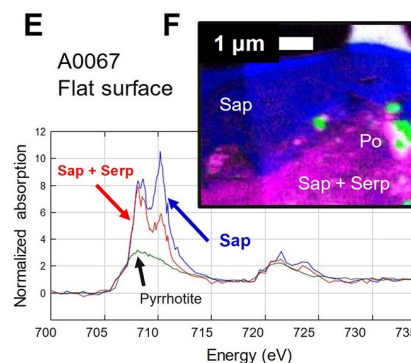
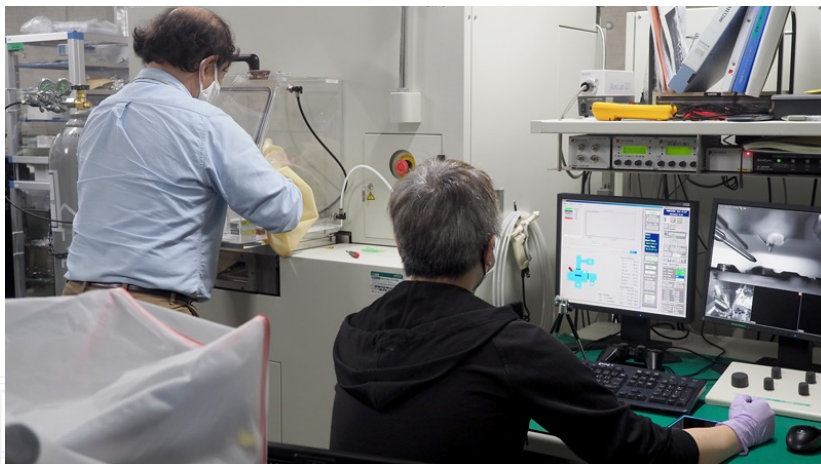


開発したX線顕微鏡

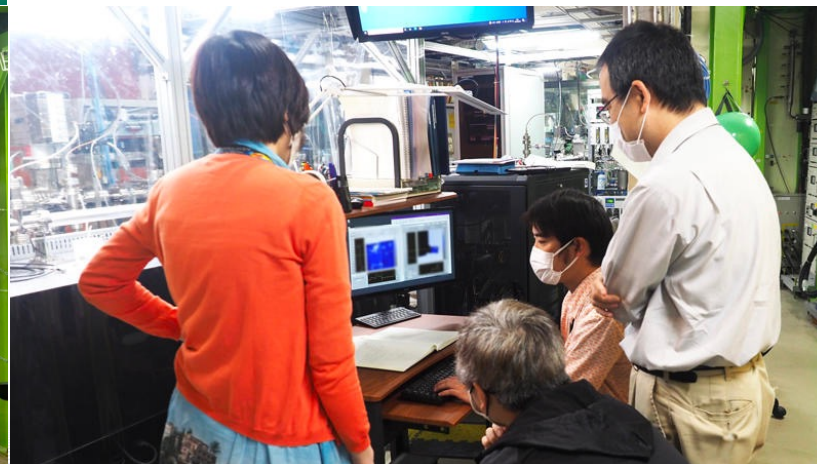




# はやぶさ2帰還リュウグウ試料分析



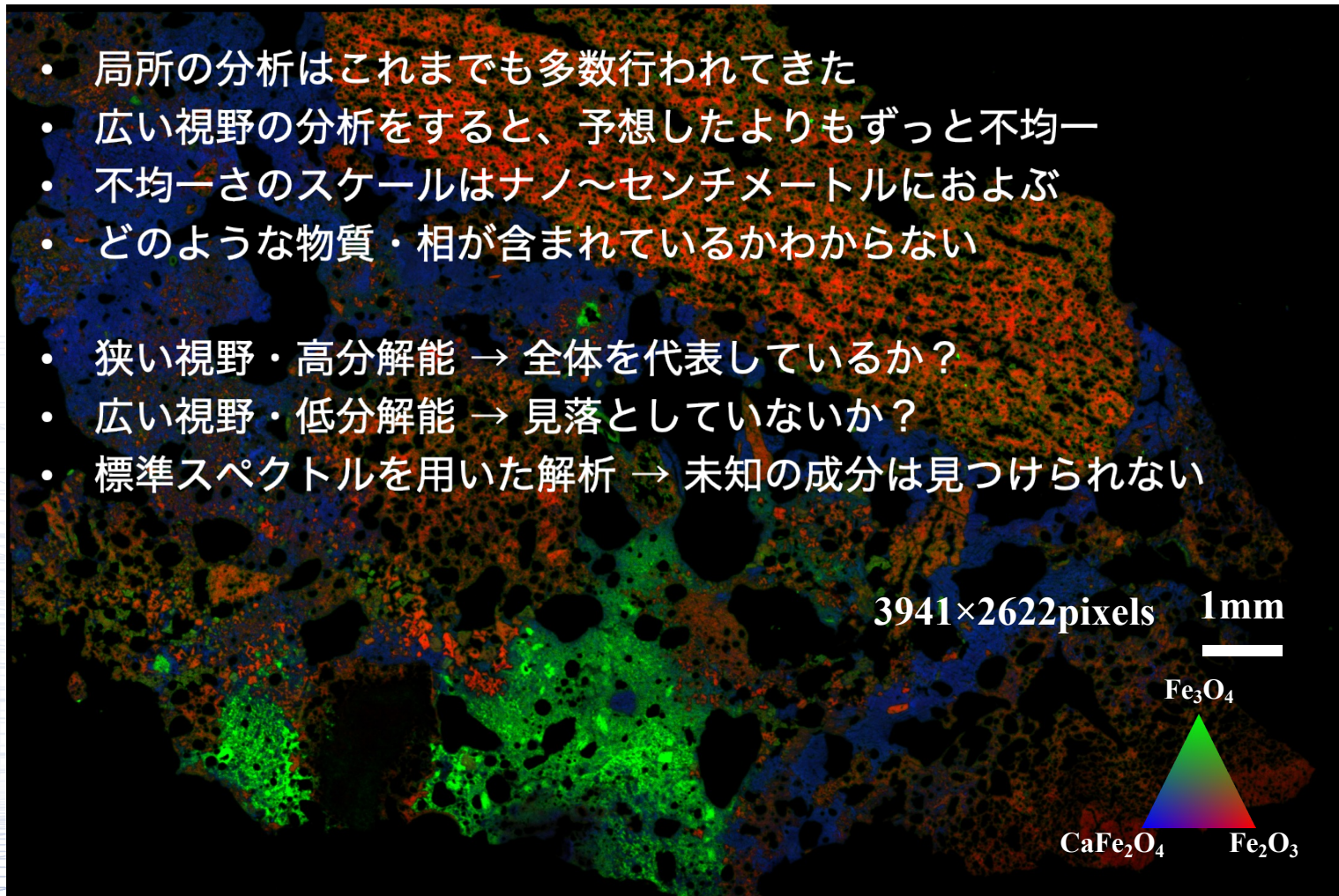
Nakamura et al., *Science* (2022).



問題点：200 x 200 ピクセルの計測データすらまともにリアルタイム解析できていない

# 高分解能・広視野X線顕微鏡

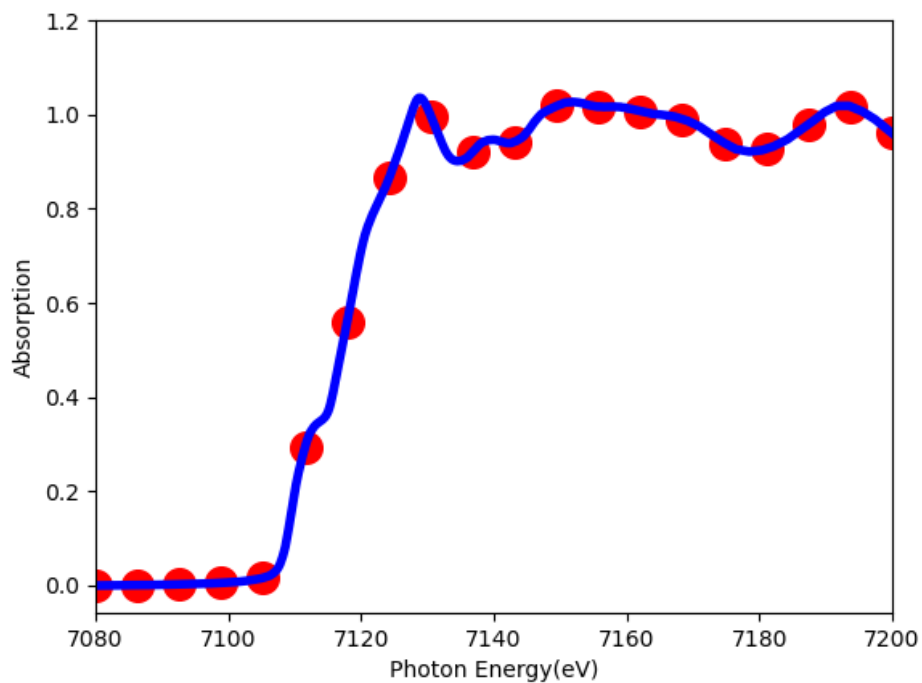
- 局所の分析はこれまでも多数行われてきた
- 広い視野の分析をすると、予想したよりもずっと不均一
- 不均一さのスケールはナノ～センチメートルにおよぶ
- どのような物質・相が含まれているかわからない
- 狭い視野・高分解能 → 全体を代表しているか？
- 広い視野・低分解能 → 見落とししていないか？
- 標準スペクトルを用いた解析 → 未知の成分は見つけれられない



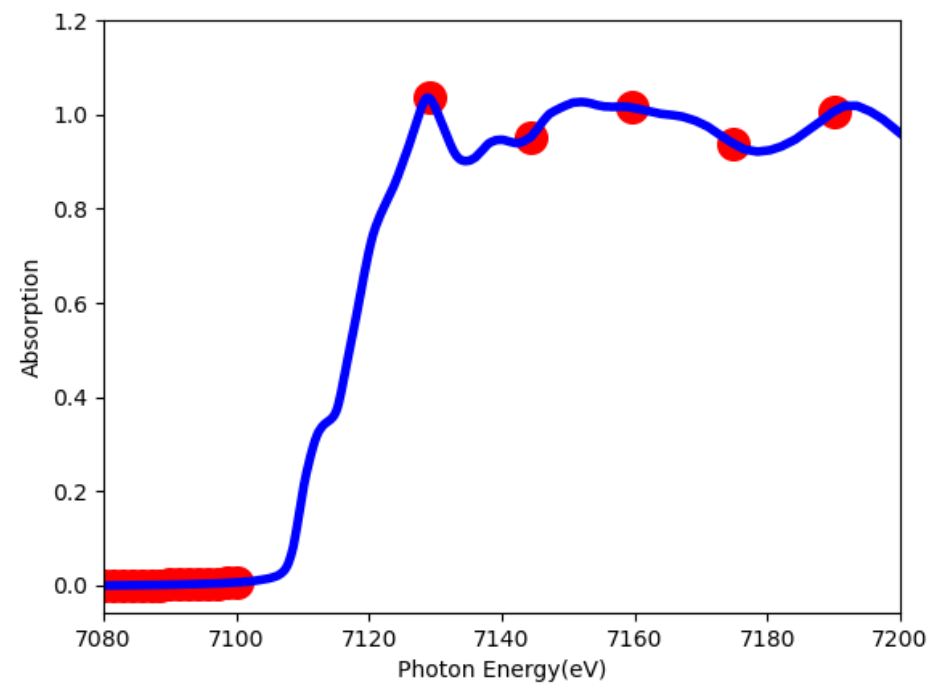
# 最適実験計画

全計測時間（計測点数）を与えた場合の計測点の最適化

どちらが良いですか？



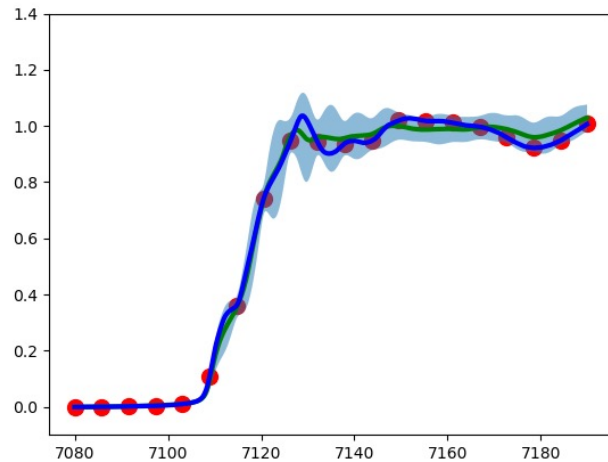
VS



# 最適実験計画

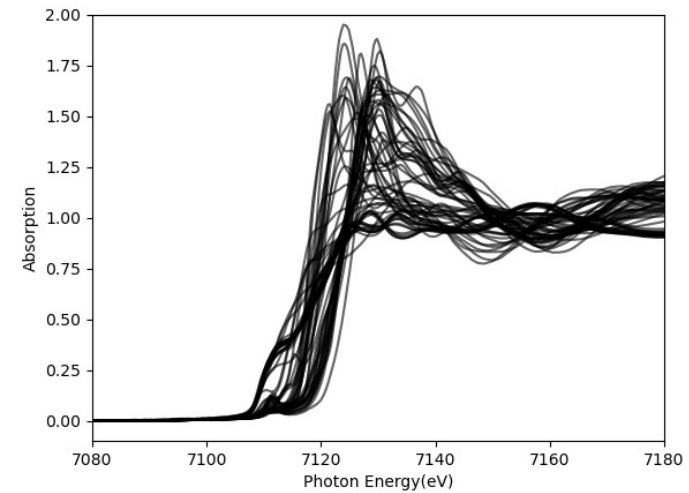
全計測時間（計測点数）を与えた場合の計測点の最適化

## ガウス過程回帰



計測とは  
→ 最適な計測点で真のスペクトルを近似する

## ベイズ実験計画

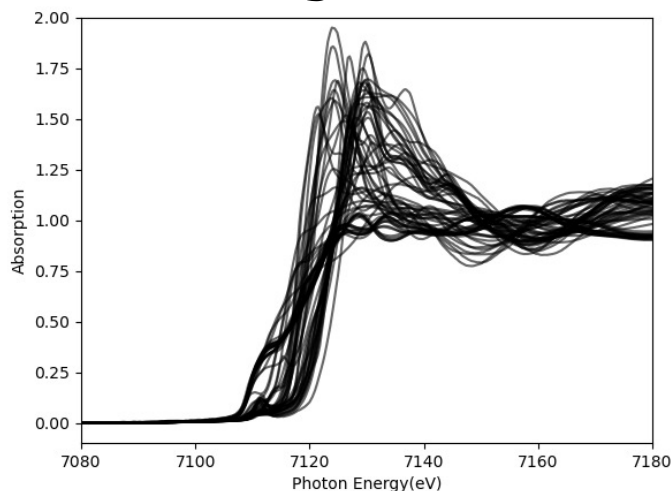


スペクトルが生成される確率分布を考える

# 開発した手法

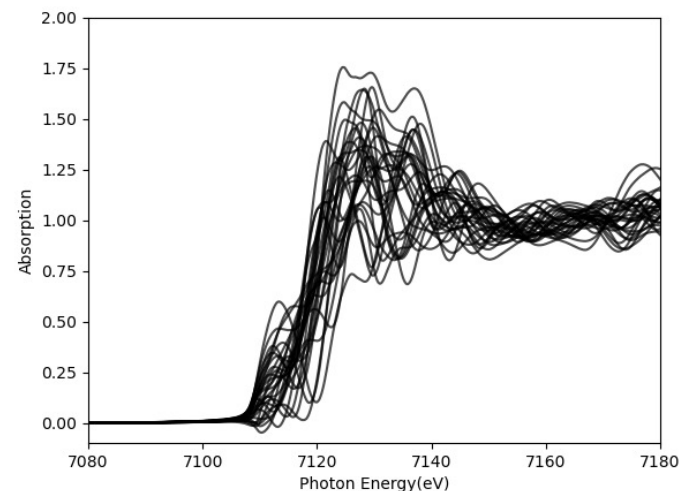
データベースを用いて事前分布を作成  
事前分布からサンプリングを行うことで、取り得るスペクトルセット生成

### Fe-K Edge Database<sup>1</sup>



確率分布 $p(f)$   
に落とし込む

### 確率分布から生成したスペクトル



事前分布からサンプリングしたスペクトルを計測すると見立てて、  
ベイズ的実験計画の枠組みを用いて最適な計測エネルギー一点を決定する。



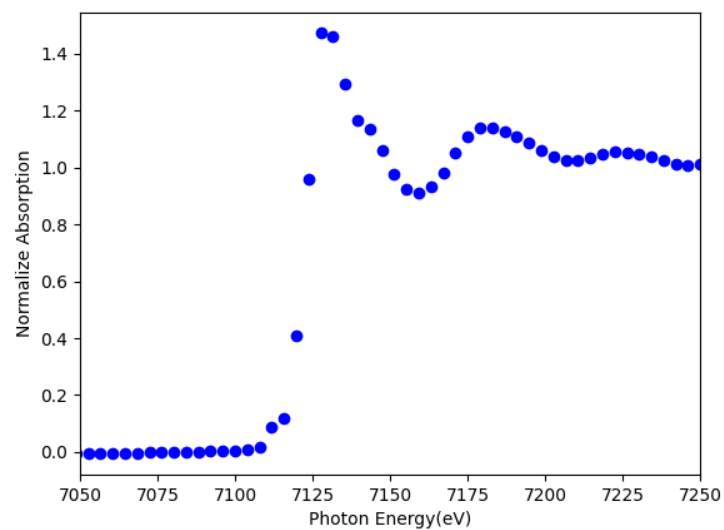
データベースにより

【 ガウス過程回帰モデルに事前情報を入れること  
能動的にではなく、あらかじめ計測エネルギー一点を決定すること

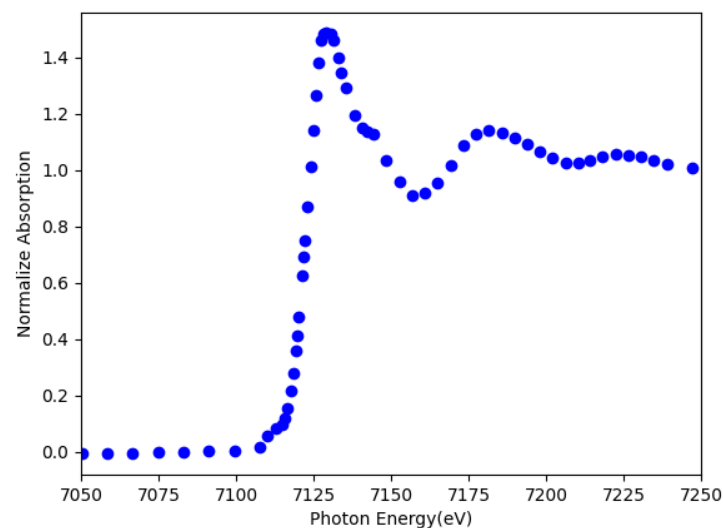
の両方を達成

# スペクトル計測の最適化

## 等間隔サンプリング



## 提案手法



等間隔サンプリングと比較して、  
提案手法では変化が大きいエネルギー範囲を重点的にサンプリング

# 大規模計測データ解析

バルク材料中の極めて小さい領域が材料全体の性質に影響を及ぼすことがある。



材料の理解・デザインのためには、

広い視野かつ高い分解能で化学状態の計測を行う

解決

未知の小さい領域まで見落とすことなく解析を行う

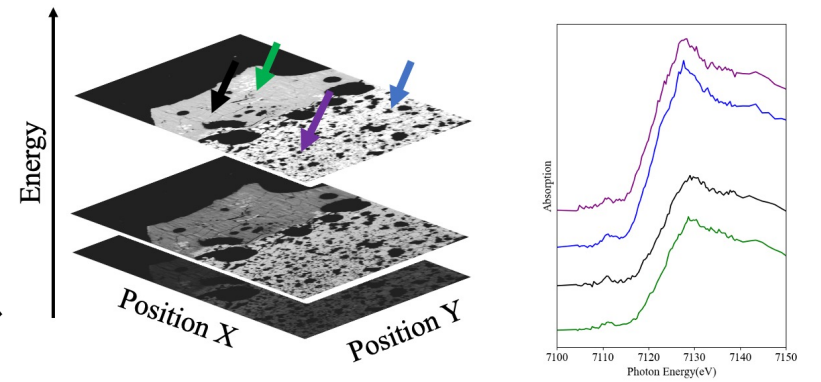
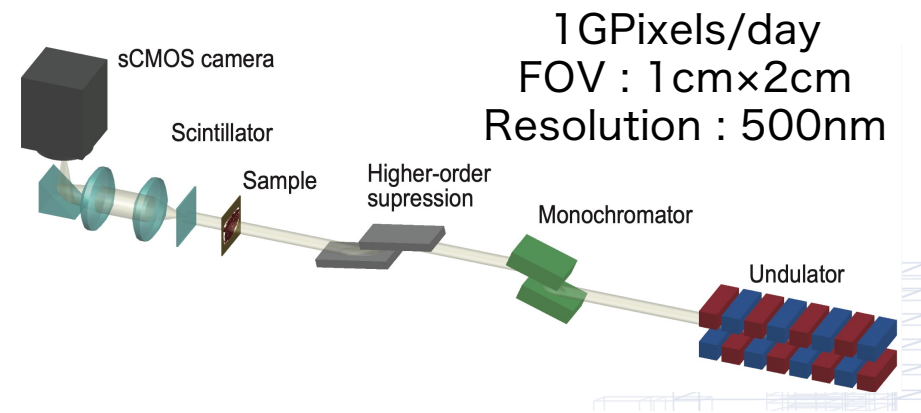
課題 (Needle in a Haystack) の両方が必要不可欠。

未知かつ微小にしか含まれない成分を見つけ出すことは人間が行うことは不可能・既存の機械学習手法を用いても難しい。



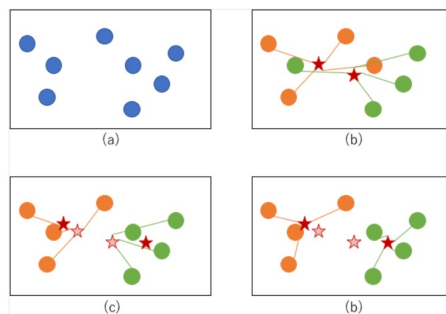
目的：未知・微小領域まで見落とさないような解析を実現し、材料の理解を進める。

G-XRM 計測セットアップ



# 既存の分類手法

## K-means法

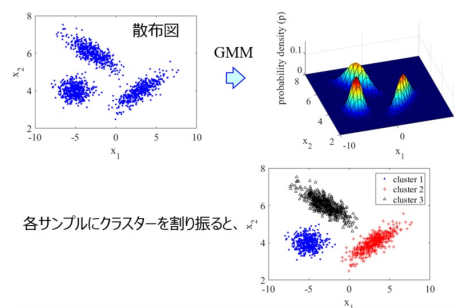


$$J = \sum_{k=1}^K \sum_{n=1}^N r_{nk} \|x_n - \mu_k\|^2$$

微量(微小領域)しか含まれないスペクトルを一つのクラスターと分類しても**損失関数**への寄与はわずかなため、見つけることは困難

**BIC**などを用いて分類数を決めることも同じ理由で困難

## 混合ガウスモデル



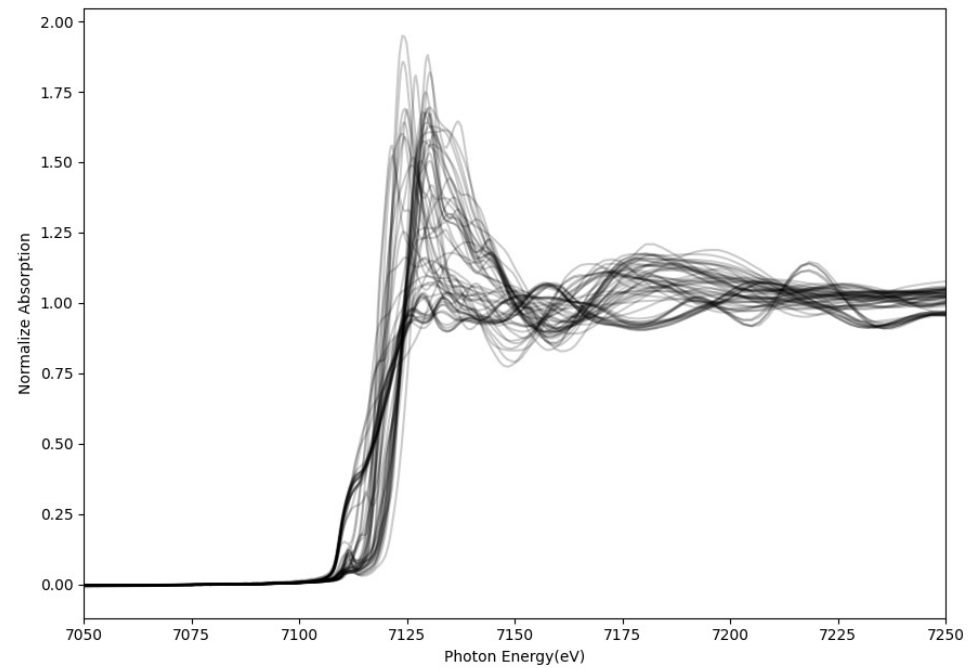
既存の手法では { 少量しか含まれないスペクトルの分類  
ノイズへの頑健性 } を両立することが困難



# スペクトル空間

計測可能なスペクトルはどのくらいある？

Fe の K-edge 吸収スペクトル



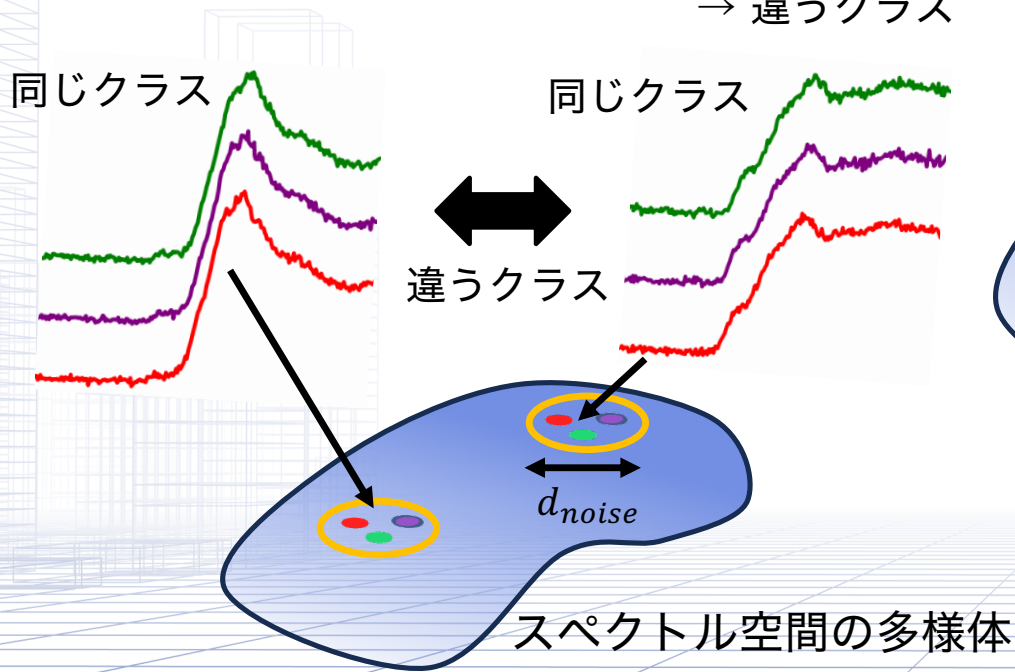
同

# スペクトル空間

計測したスペクトルの教師なし分類を行うことで、未知・微小な領域であっても発見することができる。

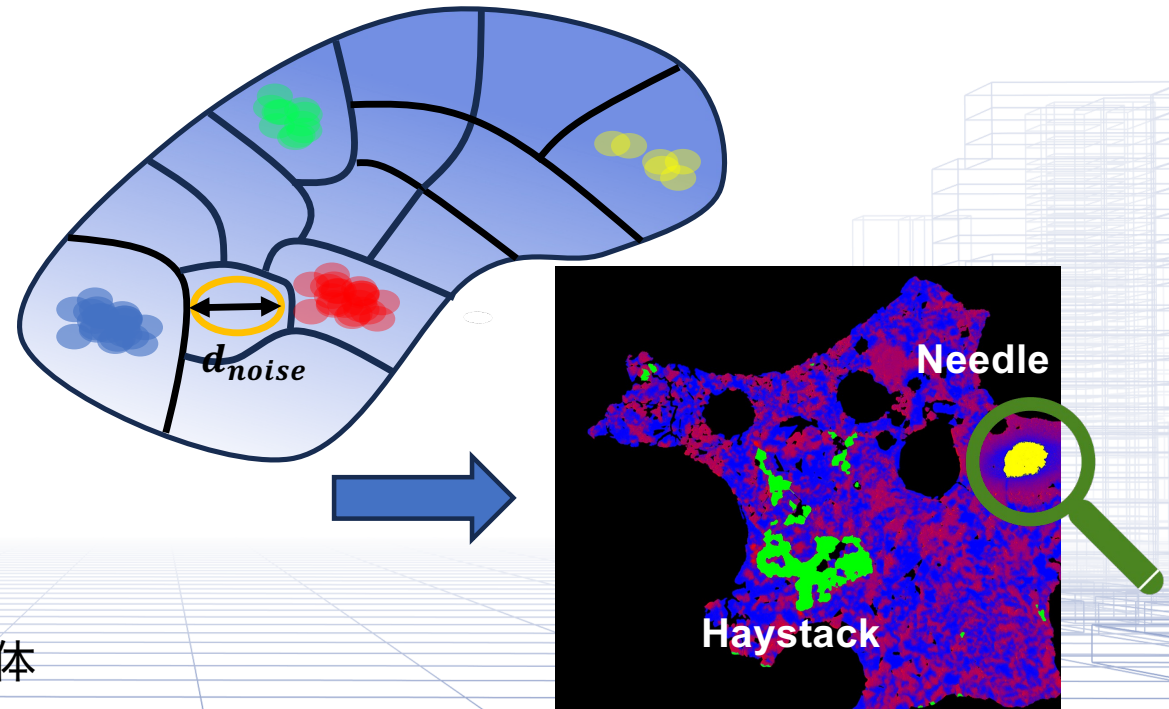
## 理想的な分類

- 計測ノイズで識別不可能となるスペクトル  
→ 同じクラス
- 計測ノイズを加味しても識別可能なスペクトル  
→ 違うクラス

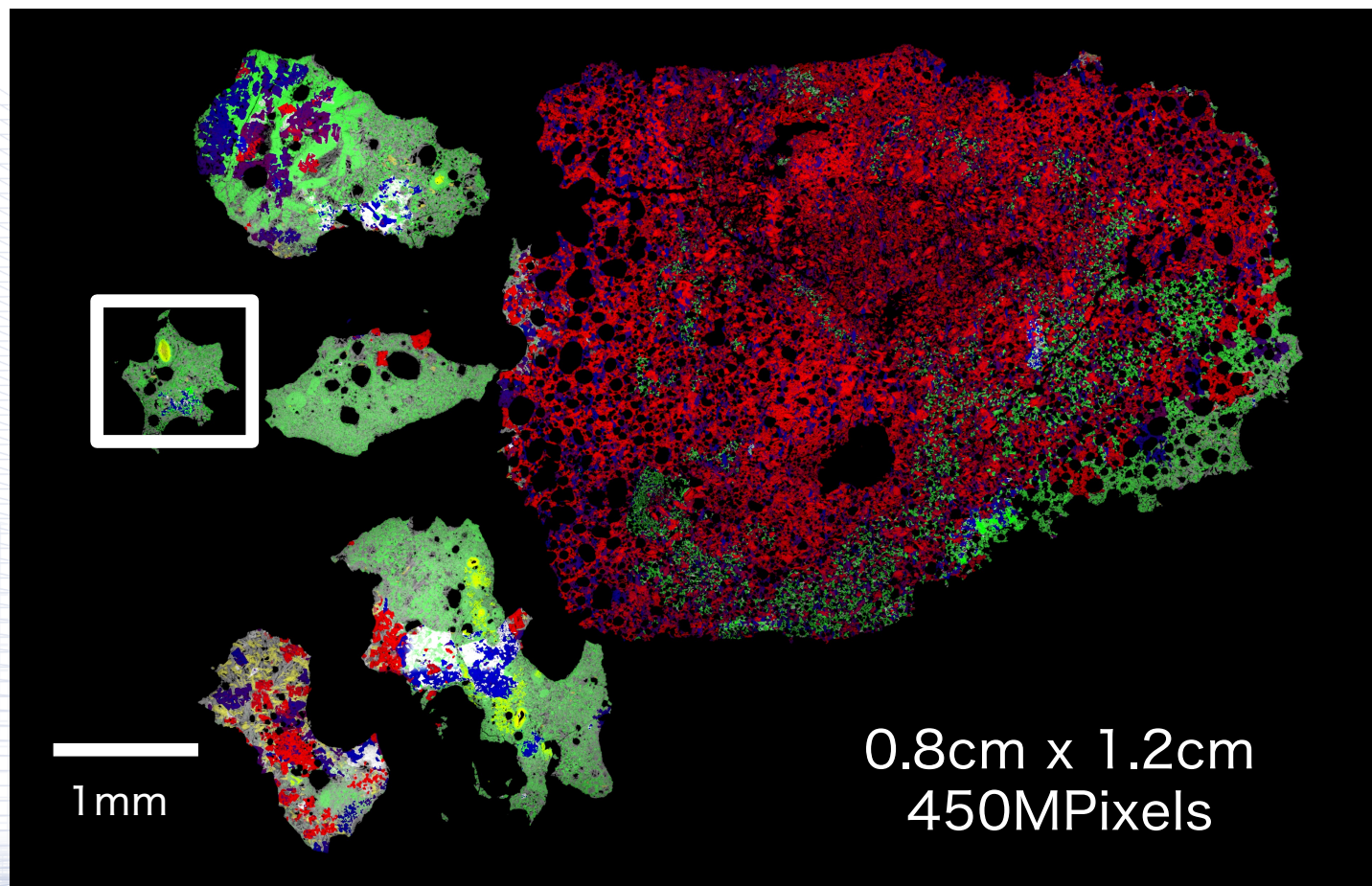


## スペクトル空間

スペクトルが分布している空間を構成出来れば、計測ノイズ  $d_{noise}$  程度の間隔で分割することで実現可能。



# クラスタリングの結果



ピクセル数:450MPixels  
測定時間:10時間  
解析時間:20時間  
相の数:5相

赤,青,緑:単純鉄酸化物  
( $\text{Fe}_2\text{O}_3$ ,  $\text{Fe}_3\text{O}_4$ )  
灰:SFCA相  
黄:SFCA相?

# 知の創出のためのエコシステム データの計測・解析・共有・公開

研究のボトルネック：大規模計測データの解析・共有が出来ず、共同研究を通じた知の創出が進まない

大阪大学の量子ビーム計測技術により、材料の大規模（10億スペクトル：～TB）計測データが創出されるようになったが、解析・共有システムが整備されていないため、計算データとの連携やデータ活用が困難

