

マテリアルズ・インフォマティクス 分野の特色と課題

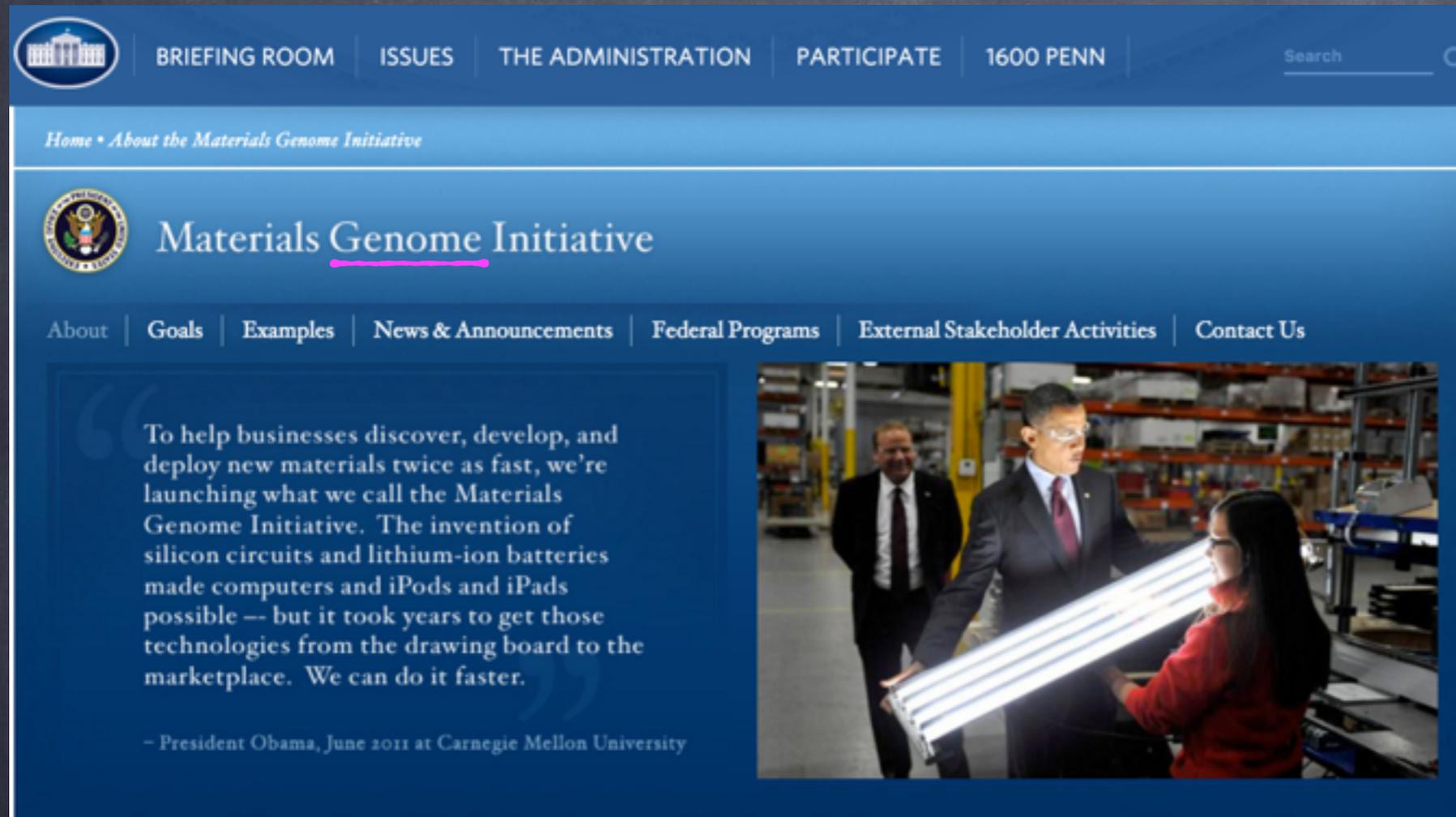
北陸先端科学技術大学院大学

DAM Hieu Chi

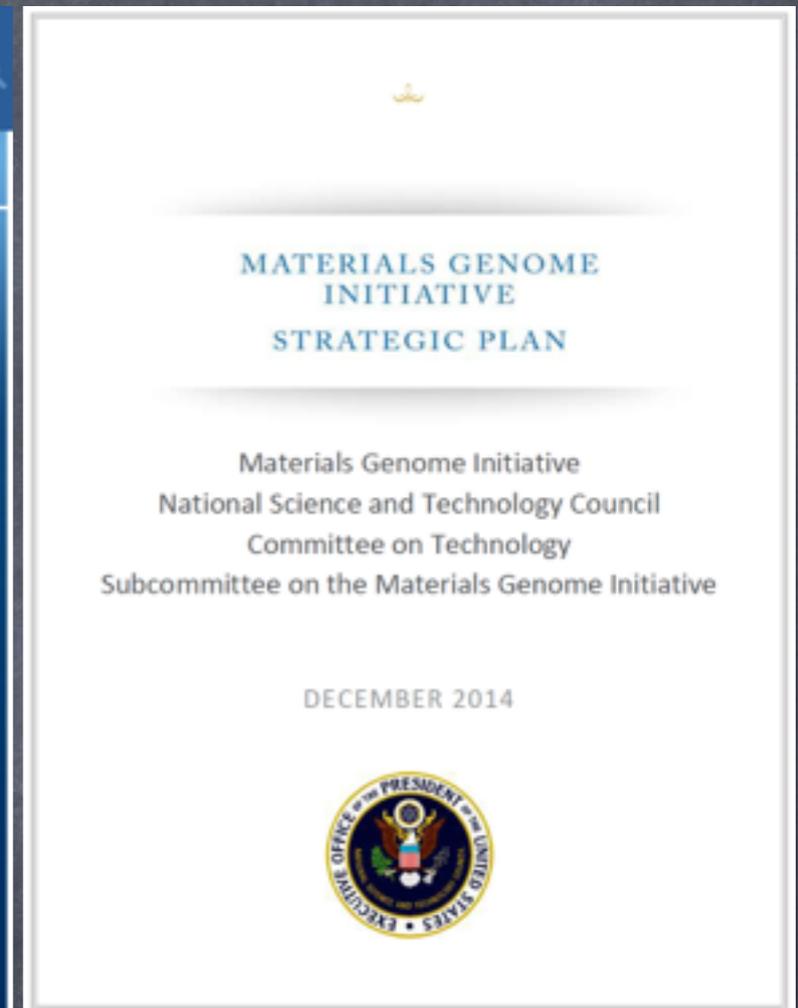
Outline

- イントロ
マテリアルデータ &
マテリアルズ・インフォマティクス
- 材料の構造記述子とそれに基づいた演算
事例と考察
- 結論

マテリアルズ・インフォマティクス



The screenshot shows the homepage of the Materials Genome Initiative website. The top navigation bar includes links for BRIEFING ROOM, ISSUES, THE ADMINISTRATION, PARTICIPATE, and 1600 PENN, along with a search bar. Below the navigation bar, the text "Home • About the Materials Genome Initiative" is displayed. The main header features the Materials Genome Initiative logo and the title "Materials Genome Initiative". A secondary navigation bar includes links for About, Goals, Examples, News & Announcements, Federal Programs, External Stakeholder Activities, and Contact Us. The main content area contains a quote from President Obama: "To help businesses discover, develop, and deploy new materials twice as fast, we're launching what we call the Materials Genome Initiative. The invention of silicon circuits and lithium-ion batteries made computers and iPods and iPads possible -- but it took years to get those technologies from the drawing board to the marketplace. We can do it faster." Below the quote is a photograph of President Obama in a factory setting, holding a glowing material. The text below the photo reads: "- President Obama, June 2011 at Carnegie Mellon University".



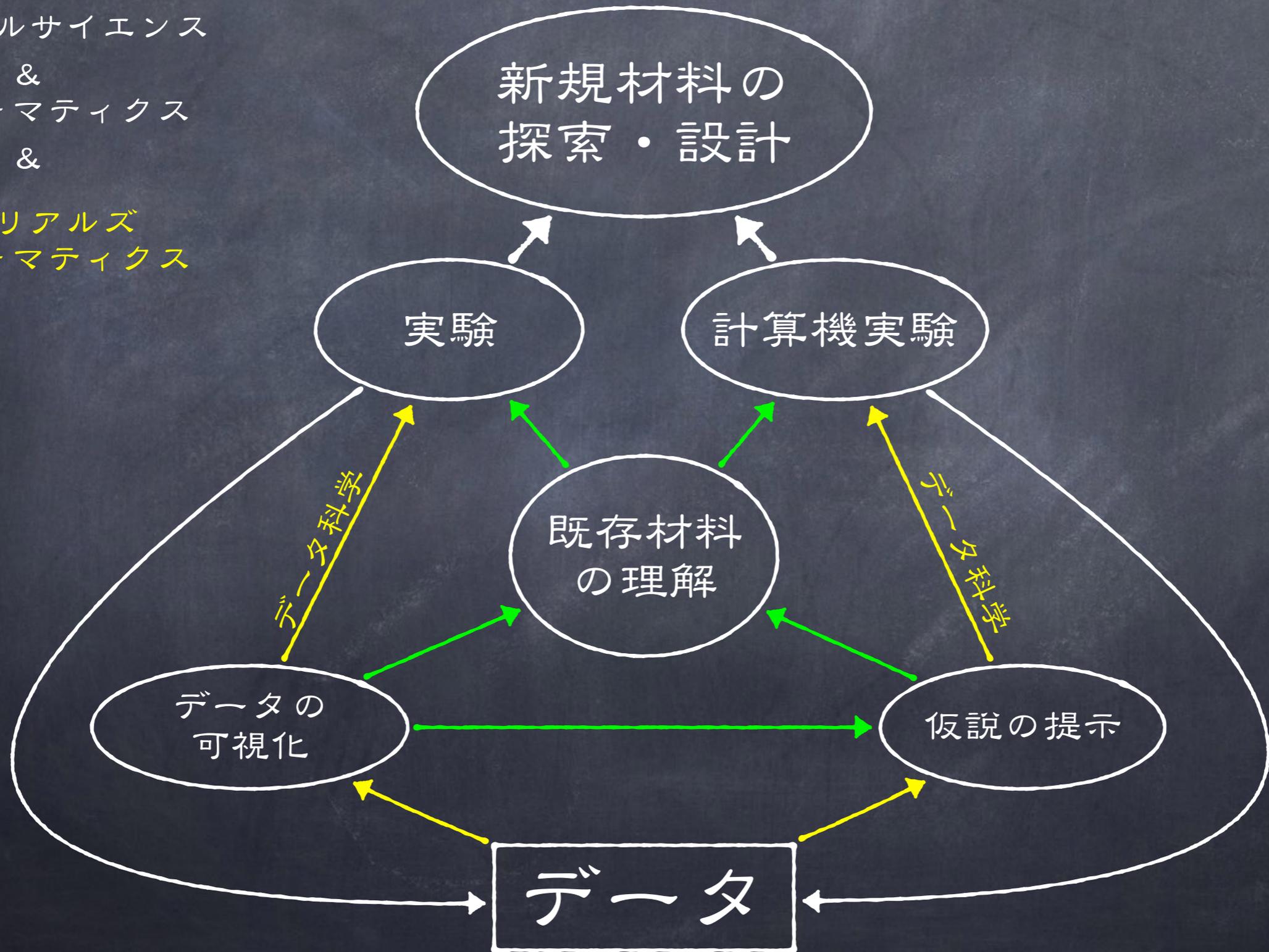
The cover of the Materials Genome Initiative Strategic Plan report features the title "MATERIALS GENOME INITIATIVE STRATEGIC PLAN" at the top. Below the title, the text reads: "Materials Genome Initiative National Science and Technology Council Committee on Technology Subcommittee on the Materials Genome Initiative". The date "DECEMBER 2014" is displayed below the text. At the bottom of the cover is the seal of the Executive Office of the President of the United States.

- ◆ 2011: Materials Genome Initiative (MGI) White Paper
\$250 Million in new R&D MGI
- ◆ 2011-2014: Fact Sheet - Three years of Progress
- ◆ 2014: MGI Strategic Plan

マテリアルズ・インフォマティクス ～ 目標・期待 ～

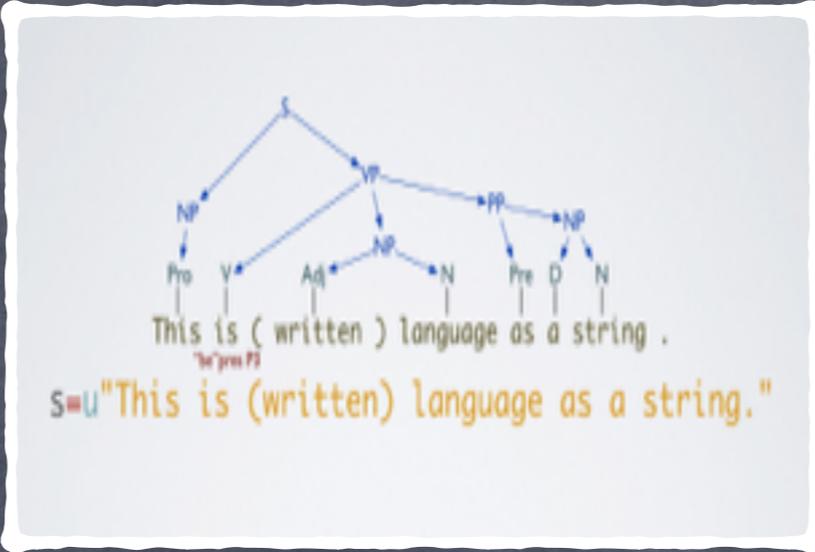
マテリアルサイエンス
&
インフォマティクス
&

マテリアルズ
インフォマティクス

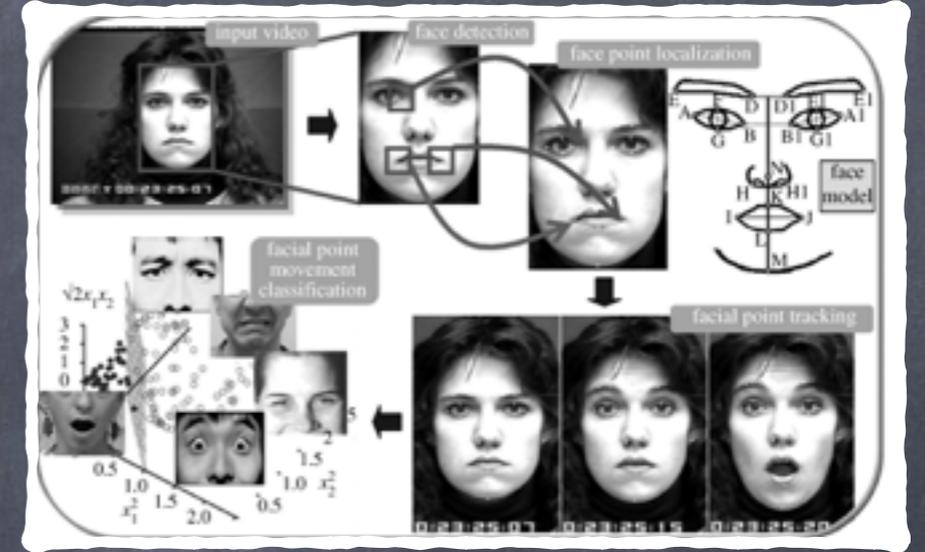


データの記述

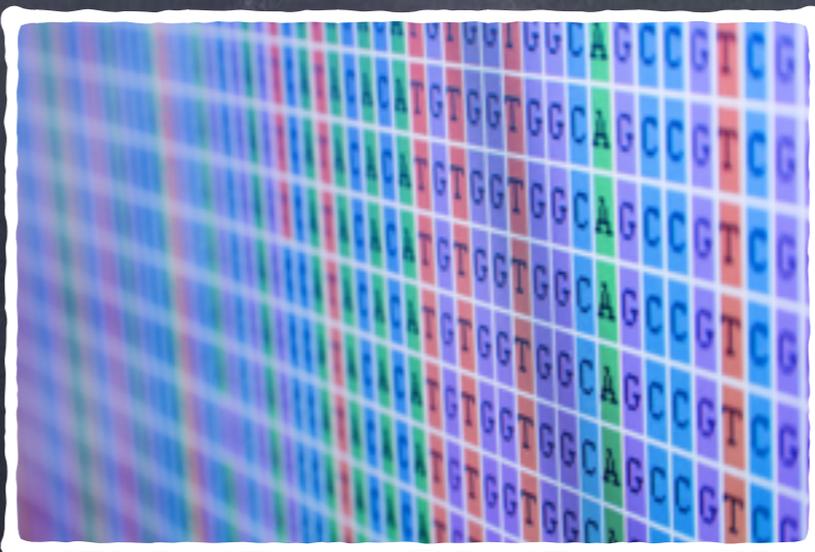
テキストマイニング 自然言語処理



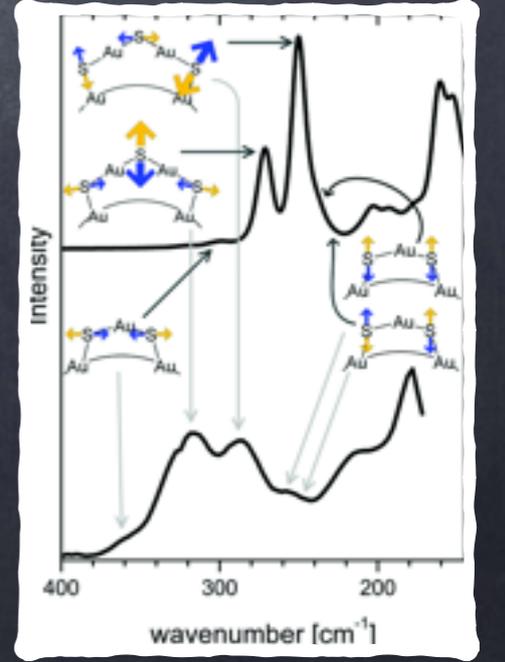
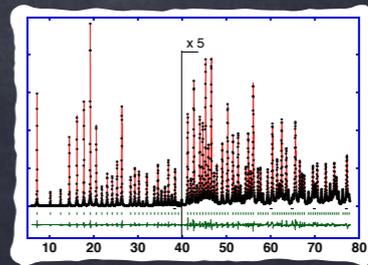
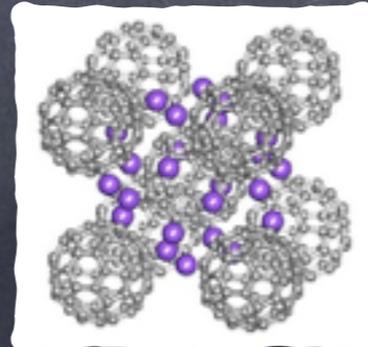
画像処理



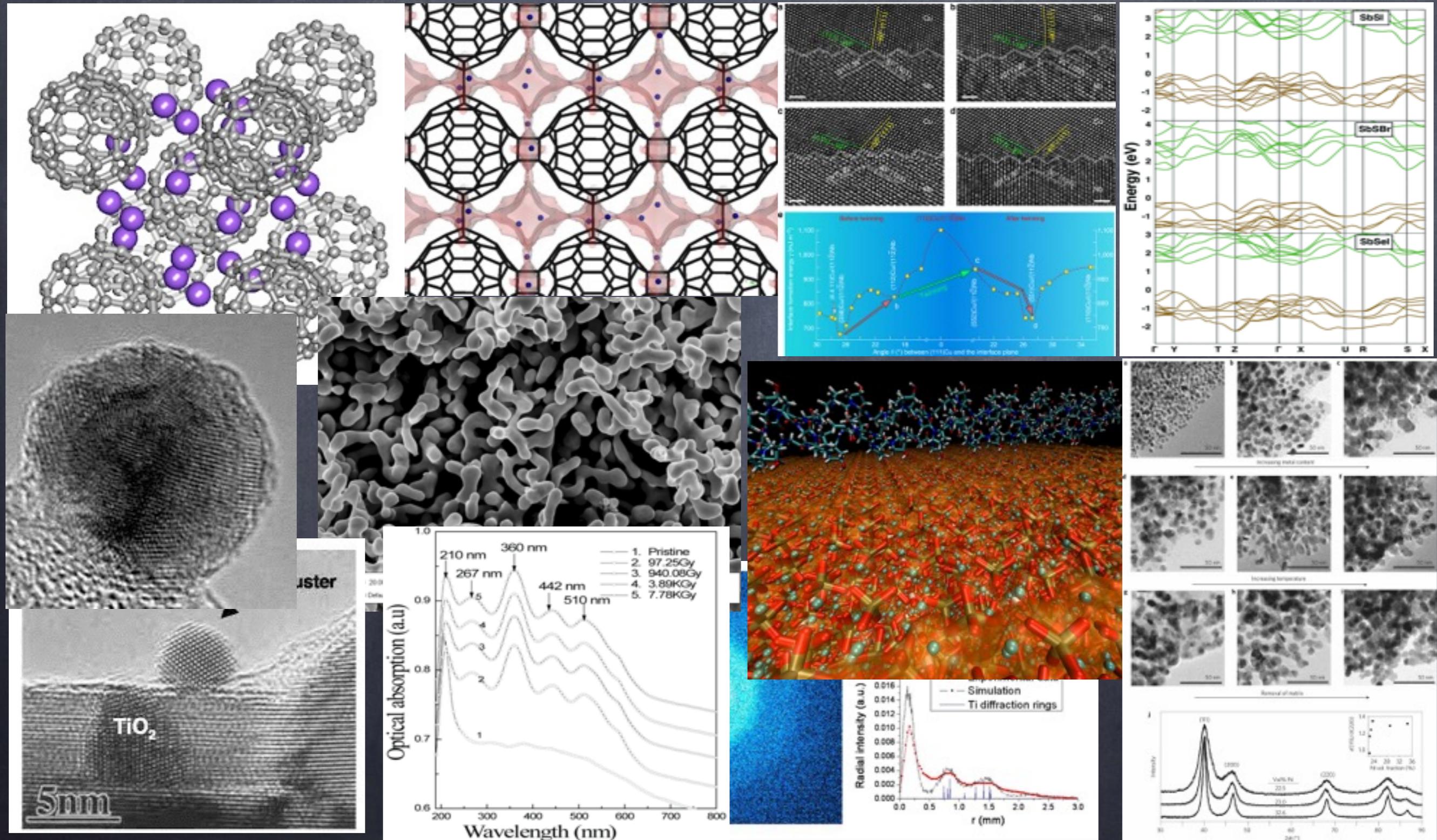
バイオインフォマティクス



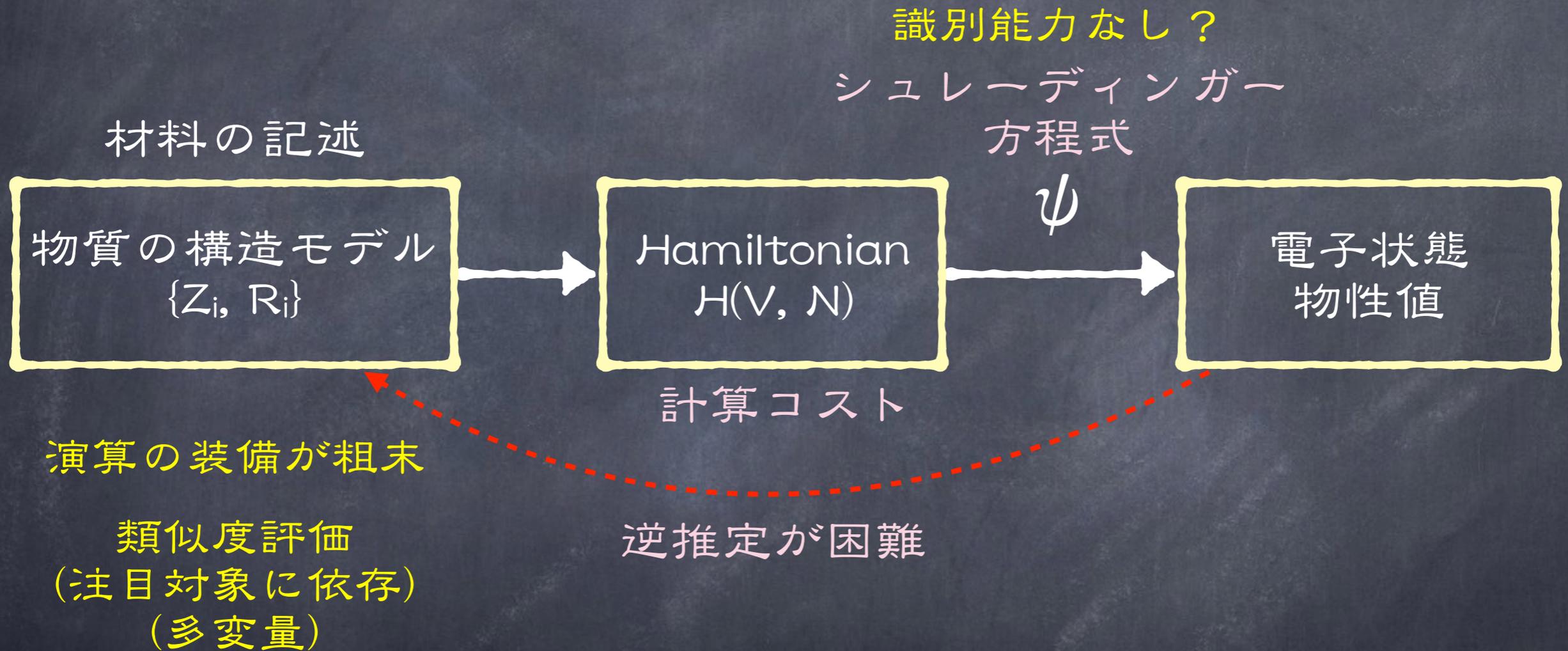
マテリアル サイエンス



材料の記述



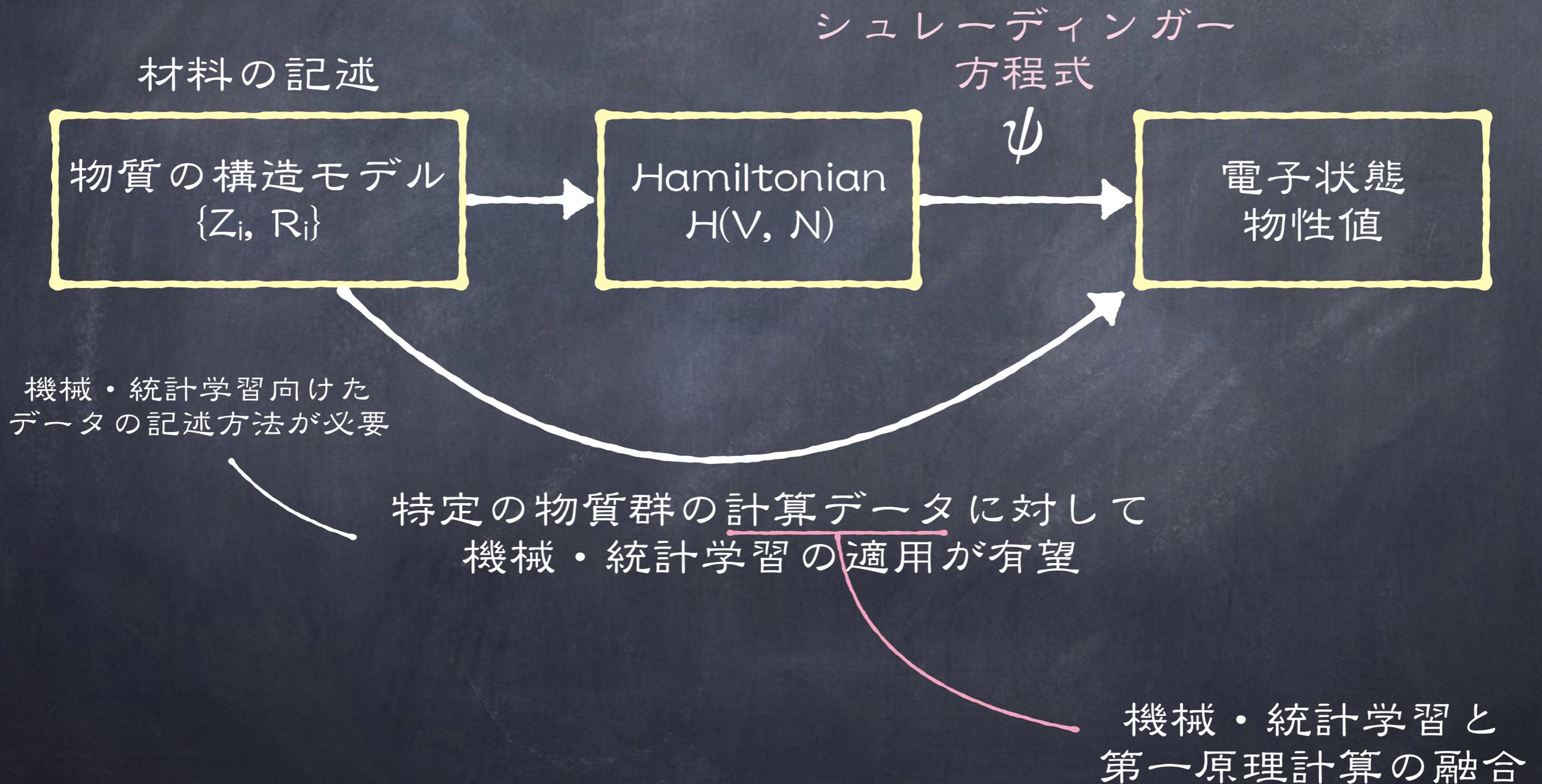
第一原理計算



【視点】

知能処理の最も素朴な処理は任意に与えられた2つ以上の情報表現が等価 (同等) であるかどうかを決定する操作 (単一化操作) である

第一原理計算と計算機実験データ



まとめ (1)

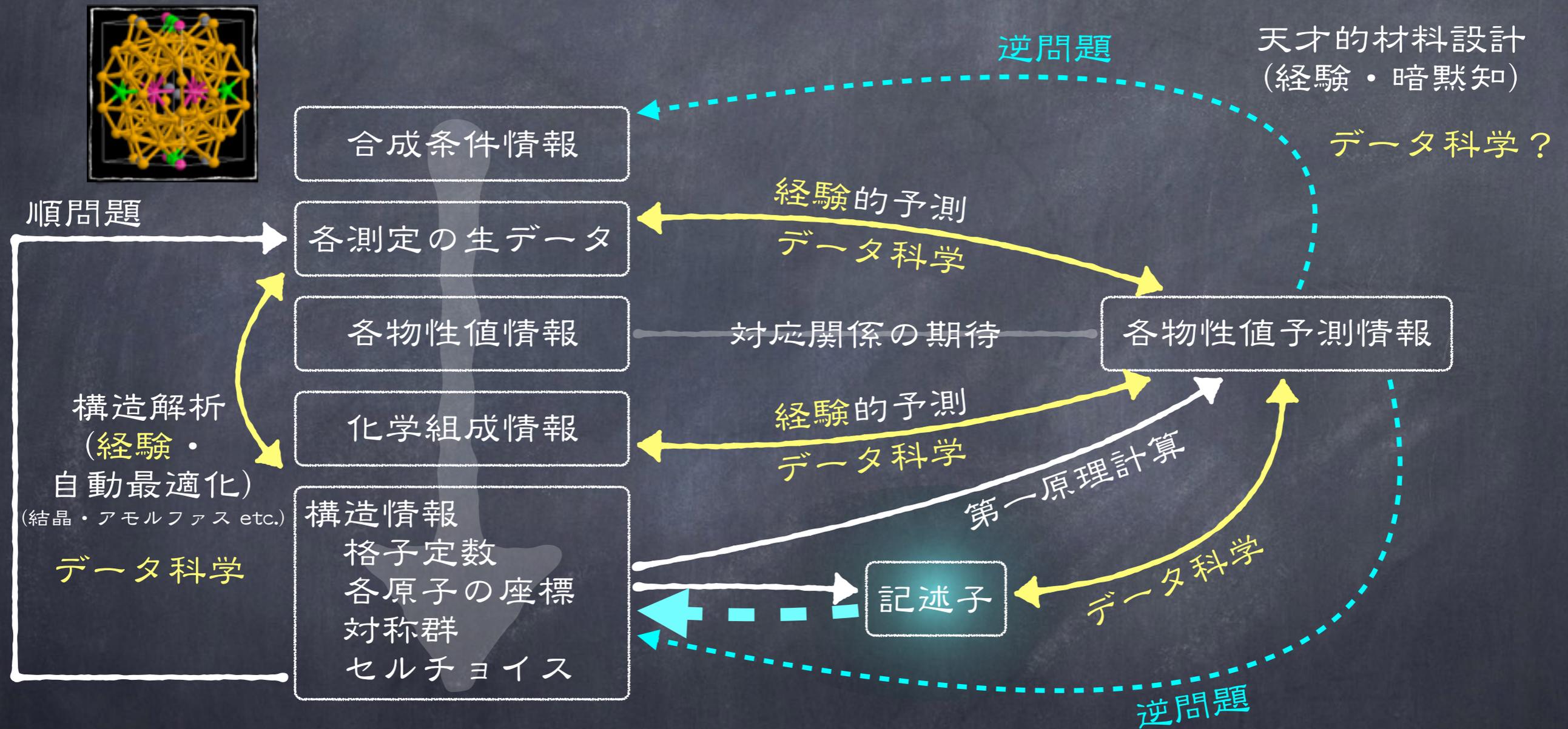
マテリアルズインフォマティクスの特徴：

1. マテリアルの記述が簡単のようで難しい
2. 算計算機実験によるデータ収集が可能

Outline

- イントロ
マテリアルデータ &
マテリアルズ・インフォマティクス
- 材料の構造記述子とそれに基づいた演算
事例と考察
- 結論

マテリアルズ・インフォマティクス ～ データ科学への期待 ～



データ科学への期待：予測精度と新たな知識・洞察の獲得
材料データに対して如何にデータマイニング手法を適切に活用するか？

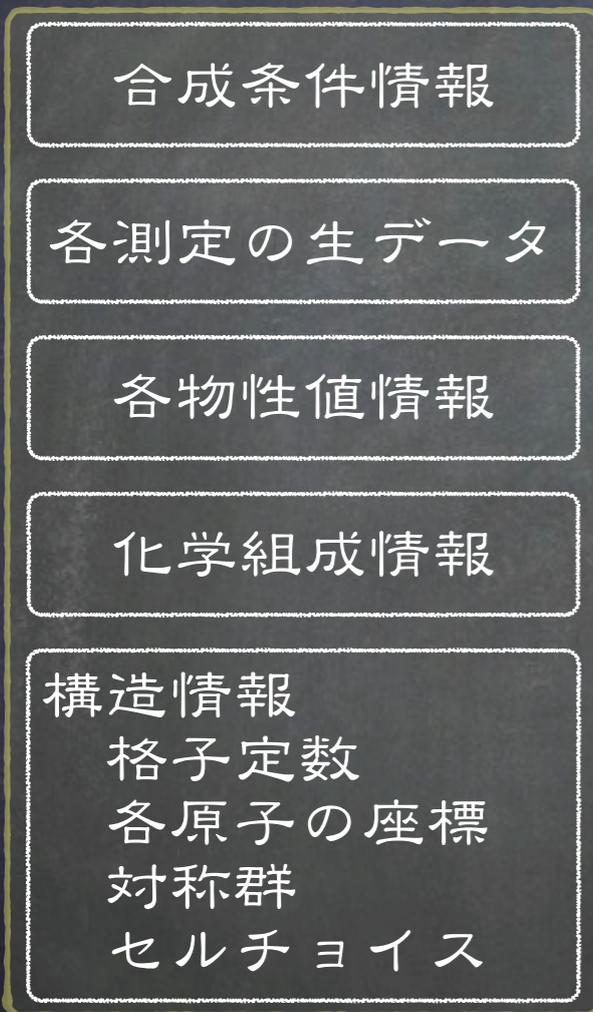
マイニングマテリアルデータ

応用性

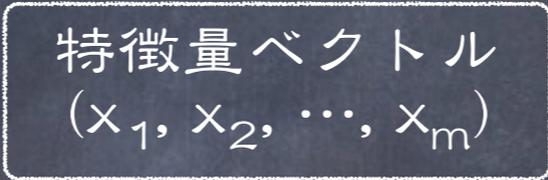
線形モデル
混合線形モデル
サポートベクトルマシン
ガウス過程
ニューラルネットワーク

全エネルギー
磁気モメント
磁気相転移温度

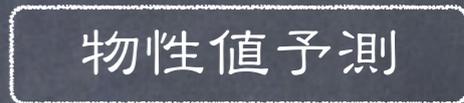
材料のデータ



記述子による表現

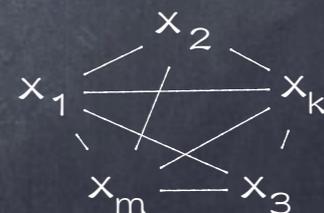


予測解析



$y = f(x_1, x_2, \dots, x_m)$
離散的情報 → 関数として知識化
解析的な探索・解析が可能

構造物性相関構造を
整理・可視化



記述解析



線形・非線形次元削減
Graphical LASSO
非線形回帰による
変数組み合わせ選択

潜在変数抽出
特徴量間相関構造
(線形・非線形相関)

新たな知識・洞察

マイニングマテリアルデータ

材料のデータ

合成条件情報

各測定の実データ

各物性値情報

化学組成情報

構造情報

格子定数

各原子の座標

対称群

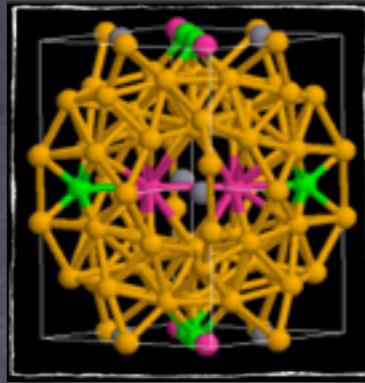
セルチョイス

記述子による表現

特徴量ベクトル
(x_1, x_2, \dots, x_m)

構造情報

格子定数
各原子の座標
対称群
セルチョイス



材料の構造記述子 (1)

記述子の必要条件

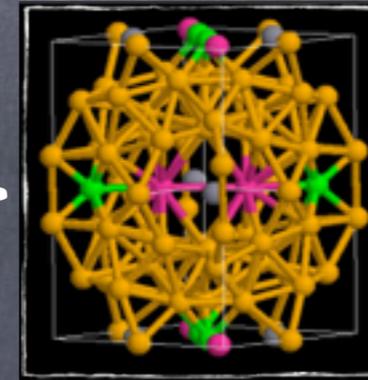
- ① 不変性
 - 並進・回転の不変性
 - 同種原子入れ替えの不変性
 - 結晶単位構造の取り方の不変性
- ② 情報量の制御
 - 完全性 (解析の目的のために十分な情報を担保する)
 - 単射性 (別の二つの構造が異なった記述ベクトルに写像される)

望ましい性質

- ① 構造に逆変換・推定可能
 - 最適化された記述子の情報を基にして、設計すべき材料の構造を推定できる
- ② 物理化学の意味付け可能
 - 記述子で記述されるデータから得られる知識や洞察が既存の物理化学の知識表現で説明できる

材料の構造記述子：原子座標の集合

格子定数
各原子の座標


$$\left\{ \begin{array}{l} a, b, c, \alpha, \beta, \gamma \\ atom_1, x_1, y_1, z_1 \\ atom_2, x_2, y_2, z_2 \\ \vdots \\ atom_n, x_n, y_n, z_n \end{array} \right\}$$

✓ 不変性

✓ 情報量の制御

✓ 構造に逆変換・推定可能

✗ 物理化学の意味付け可能

✗ 高度演算が可能

材料の構造記述子：原子座標の集合

物質A

$$\left\{ \begin{array}{l} a^A, b^A, c^A, \alpha^A, \beta^A, \gamma^A \\ atom_1^A, x_1^A, y_1^A, z_1^A \\ atom_2^A, x_2^A, y_2^A, z_2^A \\ \vdots \\ atom_n^A, x_n^A, y_n^A, z_n^A \end{array} \right\}$$

集合の代数
 $\cup \cap \equiv \neq$



物質B

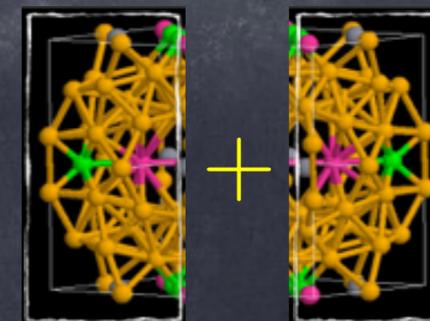
$$\left\{ \begin{array}{l} a^B, b^B, c^B, \alpha^B, \beta^B, \gamma^B \\ atom_1^B, x_1^B, y_1^B, z_1^B \\ atom_2^B, x_2^B, y_2^B, z_2^B \\ \vdots \\ atom_m^B, x_m^B, y_m^B, z_m^B \end{array} \right\}$$

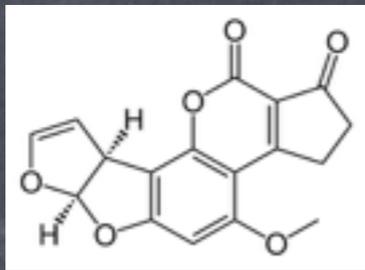
成功事例

OganovらのUSPEX

1. 無理矢理な構造演算の実行
2. 第一原理計算による構造の修正

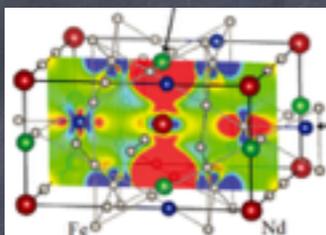
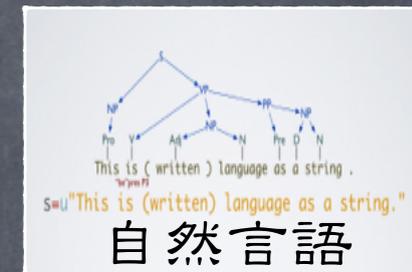
構造の演算





Chemo-informatics

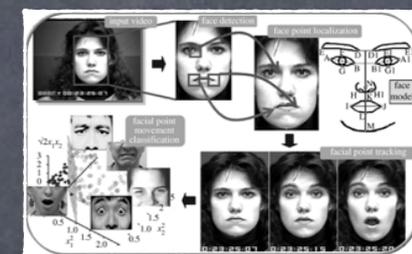
O1C=C[C@H]([C@H]1O2)c3c2cc(OC)c4c3OC(=O)C5=C4CCC(=O)5



密度汎関数法
画像処理技術
深層学習?

→ 文書のような表現

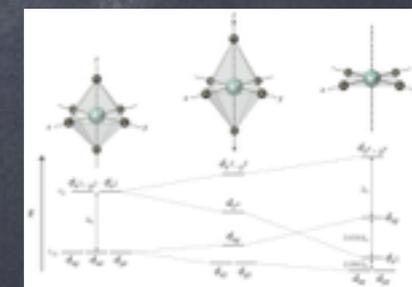
→ 画像のような表現



材料の構造記述子

従来のやり方
次元削減
スパースモデリング

→ 経験による特徴量



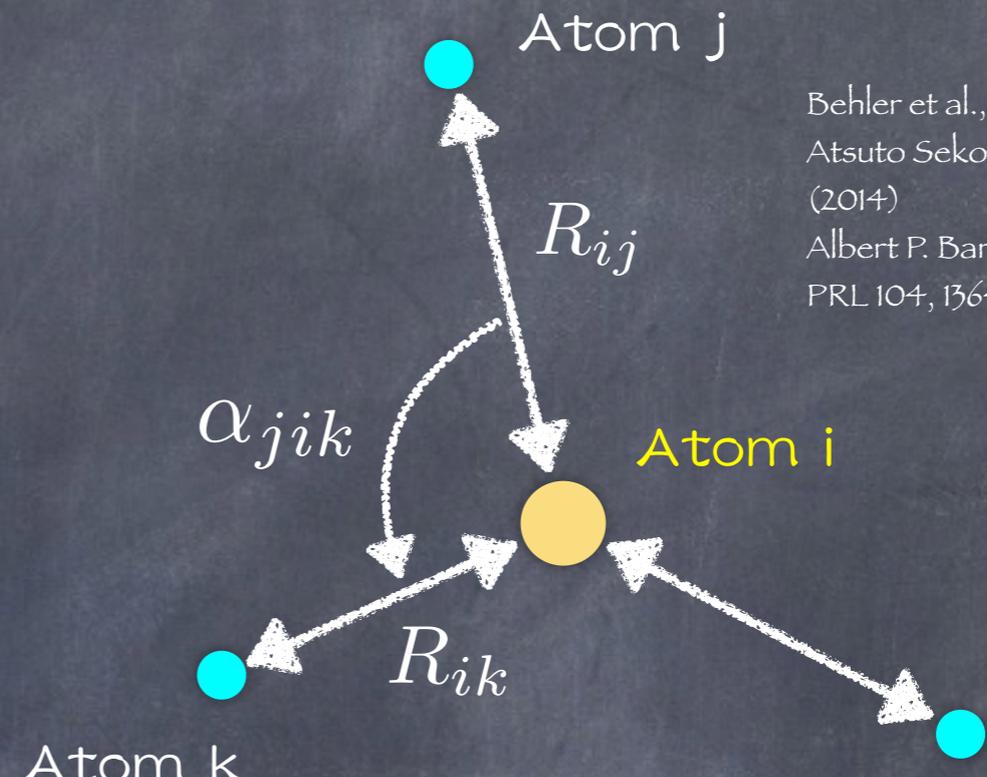
→ マテインフォ独自



材料の構造記述子：原子環境の集合

原子座標の集合

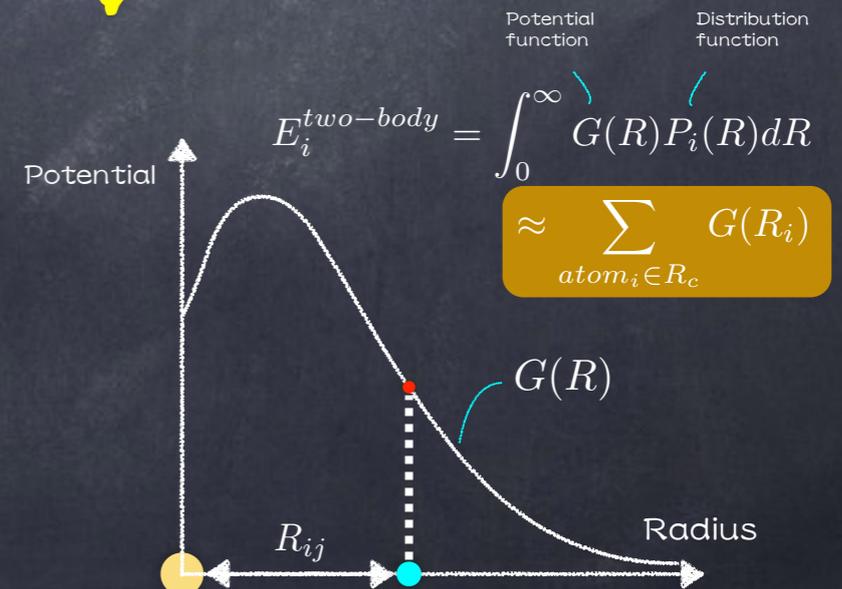
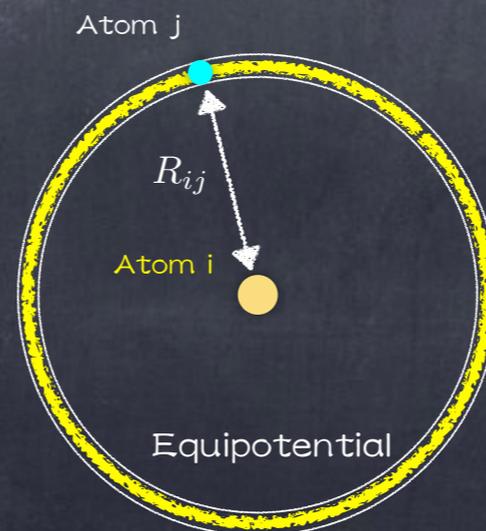
$$\left\{ \begin{array}{l} a, b, c, \alpha, \beta, \gamma \\ atom_1, x_1, y_1, z_1 \\ atom_2, x_2, y_2, z_2 \\ \vdots \\ atom_i, x_i, y_i, z_i \\ \vdots \\ atom_n, x_n, y_n, z_n \end{array} \right\}$$



Behler et al., PRL 98, 146401 (2007);
 Atsuto Seko, et.al. PRB 90, 024101 (2014)
 Albert P. Barto'k and Mike C. Payne, PRL 104, 136403 (2010)

原子環境の集合

$$\left\{ \begin{array}{l} atom_1, (x_1^1, \dots, x_1^m) \\ atom_2, (x_2^1, \dots, x_2^m) \\ \vdots \\ atom_i, (x_i^1, \dots, x_i^m) \\ \vdots \\ atom_n, (x_n^1, \dots, x_n^m) \end{array} \right\}$$



材料の構造記述子：原子環境の集合

原子座標の集合

$$\left\{ \begin{array}{l} a, b, c, \alpha, \beta, \gamma \\ atom_1, x_1, y_1, z_1 \\ atom_2, x_2, y_2, z_2 \\ \vdots \\ atom_i, x_i, y_i, z_i \\ \vdots \\ atom_n, x_n, y_n, z_n \end{array} \right\}$$



$$\left\{ \begin{array}{l} atom_1, (x_1^1, \dots, x_1^m) \\ atom_2, (x_2^1, \dots, x_2^m) \\ \vdots \\ atom_i, (x_i^1, \dots, x_i^m) \\ \vdots \\ atom_n, (x_n^1, \dots, x_n^m) \end{array} \right\}$$

原子環境の集合



不変性



情報量の制御



構造に逆変換・推定可能



物理化学の意味付け可能



高度演算が可能

結論

マテリアルズ・インフォマティクス

- ◎ 計算機シミュレーション

 - ➡ 計算機実験データ

- ◎ 記述子の設計 → 記述子の学習

 - ➡ 知識リーチなデータの記述

- ◎ データマイニング・機械学習

 - ➡ 知識抽出

… + 人間の閃きとの共創