

# 統合機械学習分子動力学システムの構築

---

課題代表者: 奥村雅彦\* (日本原子力研究開発機構)

副代表: 永井佑紀 (日本原子力研究開発機構)、華井雅俊 (東京大学)

共同研究者: 鈴木豊太郎、渡邊聡、沖田泰良、清水康司、芝隼人 (東京大学)、横井達矢 (名古屋大学)、加藤幸一郎 (九州大学)、小林亮 (名古屋工業大学)、島村孝平 (熊本大学)、富谷昭夫 (大阪国際工科専門職大学)、森英喜 (産業技術短期大学)、安藤康伸 (産業総合研究所)、河野秀俊、石田恒 (量子科学技術研究開発機構)、今村穰、浦田新吾、吉田拓未 (AGC 株式会社)、小林恵太、山口瑛子、中村博樹、板倉充洋、町田昌彦 (日本原子力研究開発機構)

\* [okumura.masahiko@jaea.go.jp](mailto:okumura.masahiko@jaea.go.jp)

# 分子動力学法

- 原子や分子の動きをシミュレートする手法

- 本来は量子力学の基礎方程式(シュレーディンガー方程式)を解く必要があるが、解くための計算コストが膨大なため近似を導入

## よく用いられる手法

- 古典分子動力学法

原子	古典物理学の質点に近似
原子間相互作用*	簡単な関数(力場)で近似*

\* 物理法則(第一原理)では決められない「経験的パラメーター」を含む

- 第一原理分子動力学法

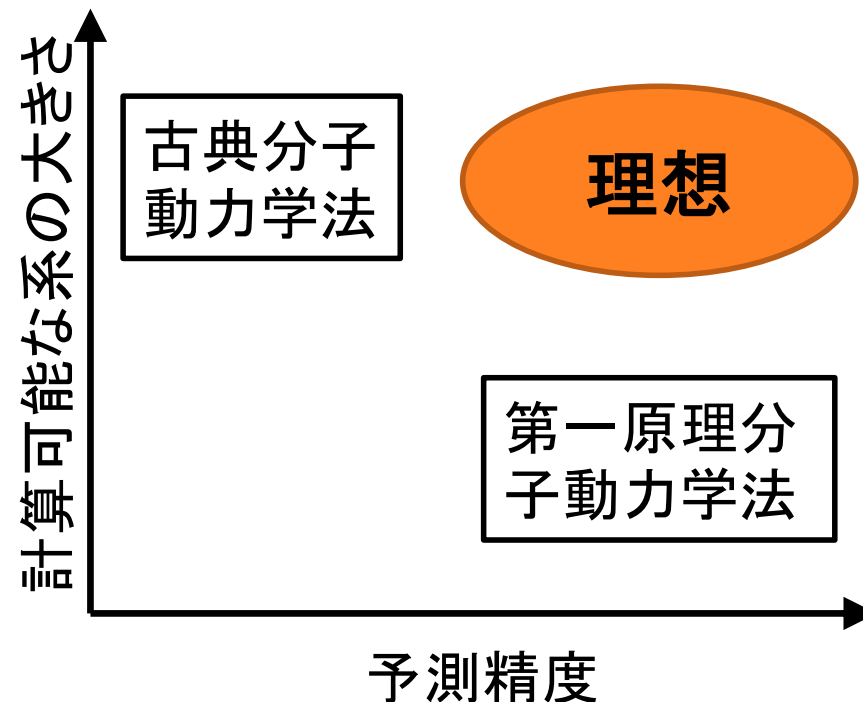
原子	原子核	古典物理学の質点に近似
	電子	量子力学
原子間相互作用**		量子力学を近似して解く**

\*\* 「経験的パラメーター」は含まない

# 既存手法の特徴と問題点

## • 予測精度 vs 計算コスト

	予測精度	計算コスト
古典分子動力学法	中	低
第一原理分子動力学法	高	高



- どちらの手法も一長一短
- 計算コストが低くて予測精度が高いのが理想
- そんな手法ある？

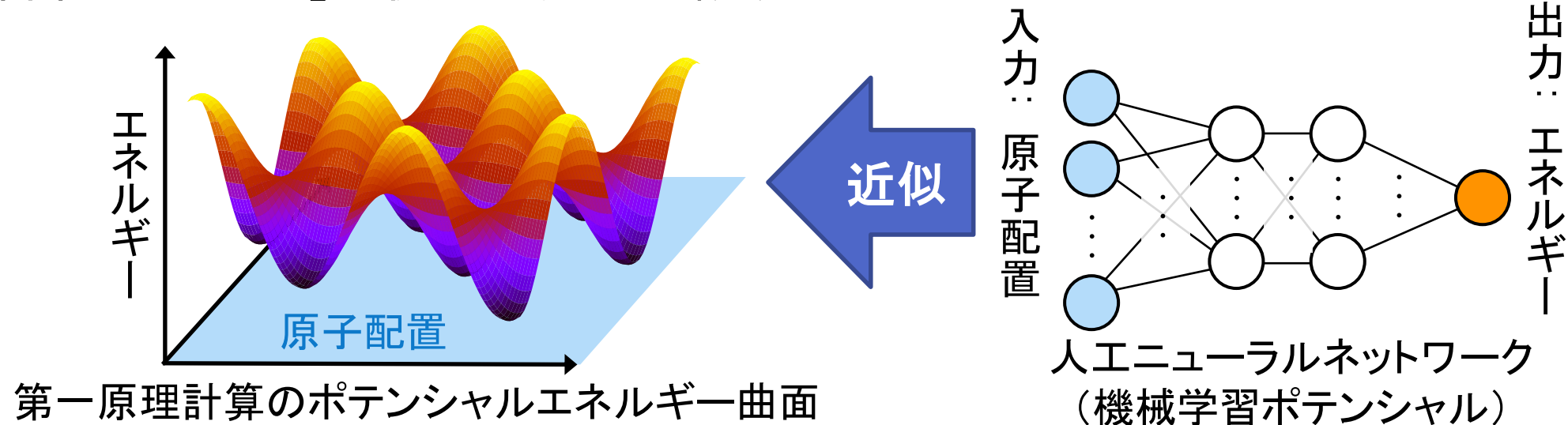


あります！  
**機械学習分子動力学法**

Behler and Parrinello, PRL (2007)

# 機械学習分子動力学法

- 概略: 第一原理計算のポテンシャルエネルギー曲面を人工ニューラルネットワーク等で近似する
  - 機械学習の手法: 人工ニューラルネットワーク、ガウス過程
    - 入力: 原子配置、出力: エネルギー
    - 学習済みの人工ニューラルネットワークを「機械学習ポテンシャル」と呼ぶ
  - 教師データ: 第一原理計算の結果
    - 大量の第一原理計算を実施して教師データを作る
  - 詳細は「まとめ」の後のスライドで説明



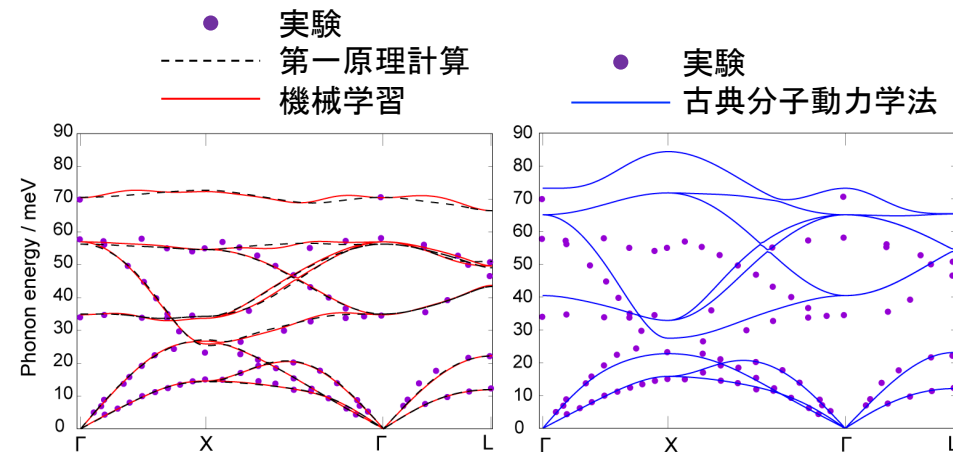
# 機械学習分子動力学法で何ができるか？

## • 高予測精度シミュレーション

### • 例: ThO<sub>2</sub>のフォノンスペクトル

K. Kobayashi *et al.*, *Sci. Rep.* **12**, 9808 (2022).

- 古典分子動力学法では実験結果と合わない部分がある
- 機械学習分子動力学法は実験結果とよく合う

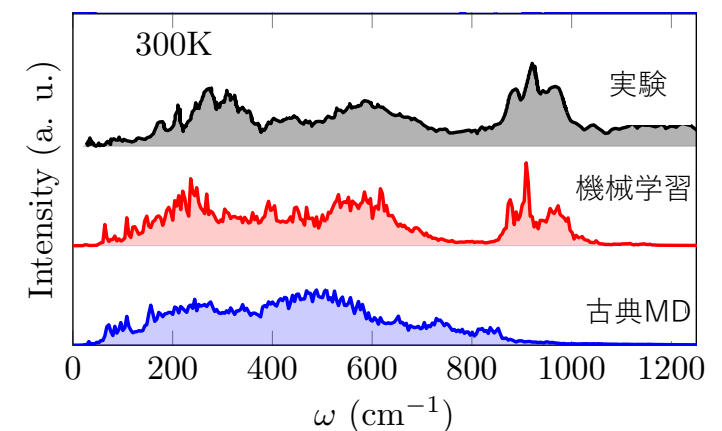


## • 高予測精度・長時間シミュレーション

### • 例: 粘土鉱物カオリナイトのヒドロキシ基振動スペクトル

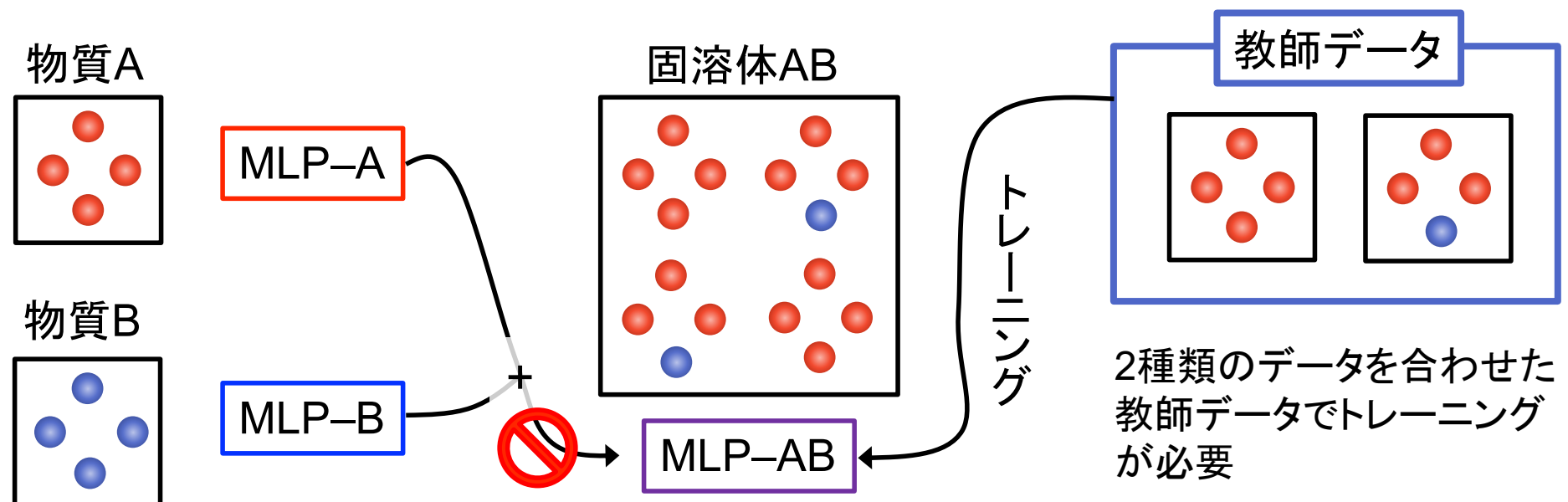
K. Kobayashi *et al.*, *App. Clay Sci.* **228**, 106596 (2022).

- 遅い振動モードの計算のためには長時間のシミュレーションが必要
- 古典分子動力学法では実験結果と合わない部分がある
- 機械学習分子動力学法は実験結果とよく合う



# 機械学習分子動力学法の不便な点

- 物質ごとに機械学習ポテンシャルを作る必要がある
  - 既存の機械学習ポテンシャルを組み合わせることはできない
    - 例: 物質A、Bの機械学習ポテンシャル(MLP-A、MLP-B)が公開されている状況で固溶体ABの機械学習ポテンシャル(MLP-AB)を作りたい
      - MLP-AとMLP-Bを組み合わせることはできない
      - 物質A、物質B、物質ABの第一原理計算の結果を混ぜた教師データを用いてトレーニングして作成する必要がある
      - しかし、**教師データはほとんど公開されていないため、計算を全てやり直す必要がある**



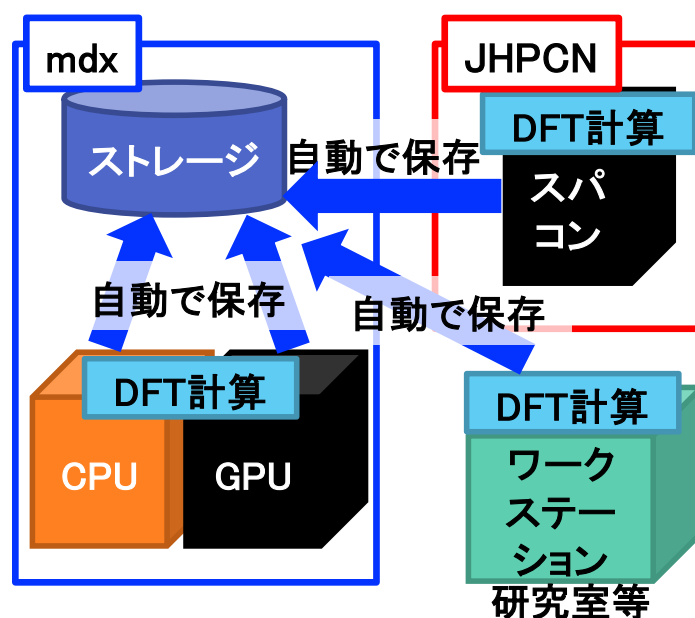
# 統合機械学習分子動力学システム

## 目的

- 機械学習ポテンシャル作成のための第一原理計算結果のデータベース
- 将来的には機械学習ポテンシャルの自動作成を目指す

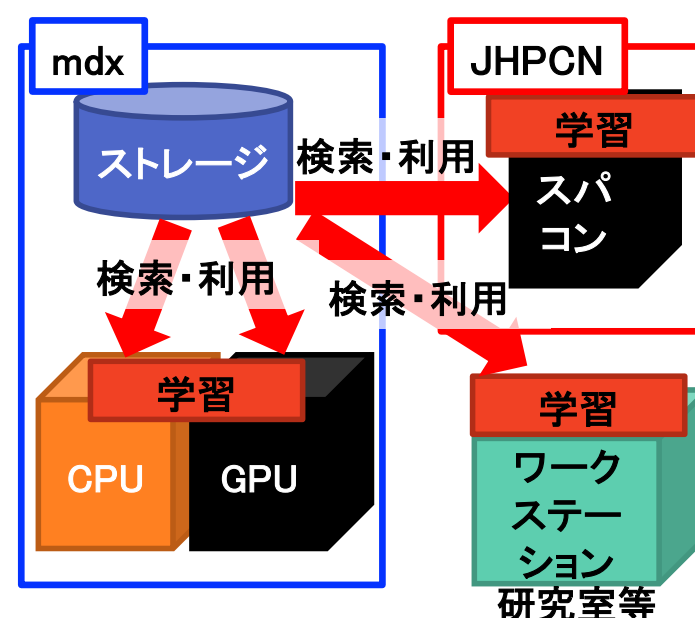
## システム概略

### 教師データ収集イメージ



- 量子力学計算の結果を自動で保存
- データベースに自動で追加される

### 教師データデータベース利用イメージ

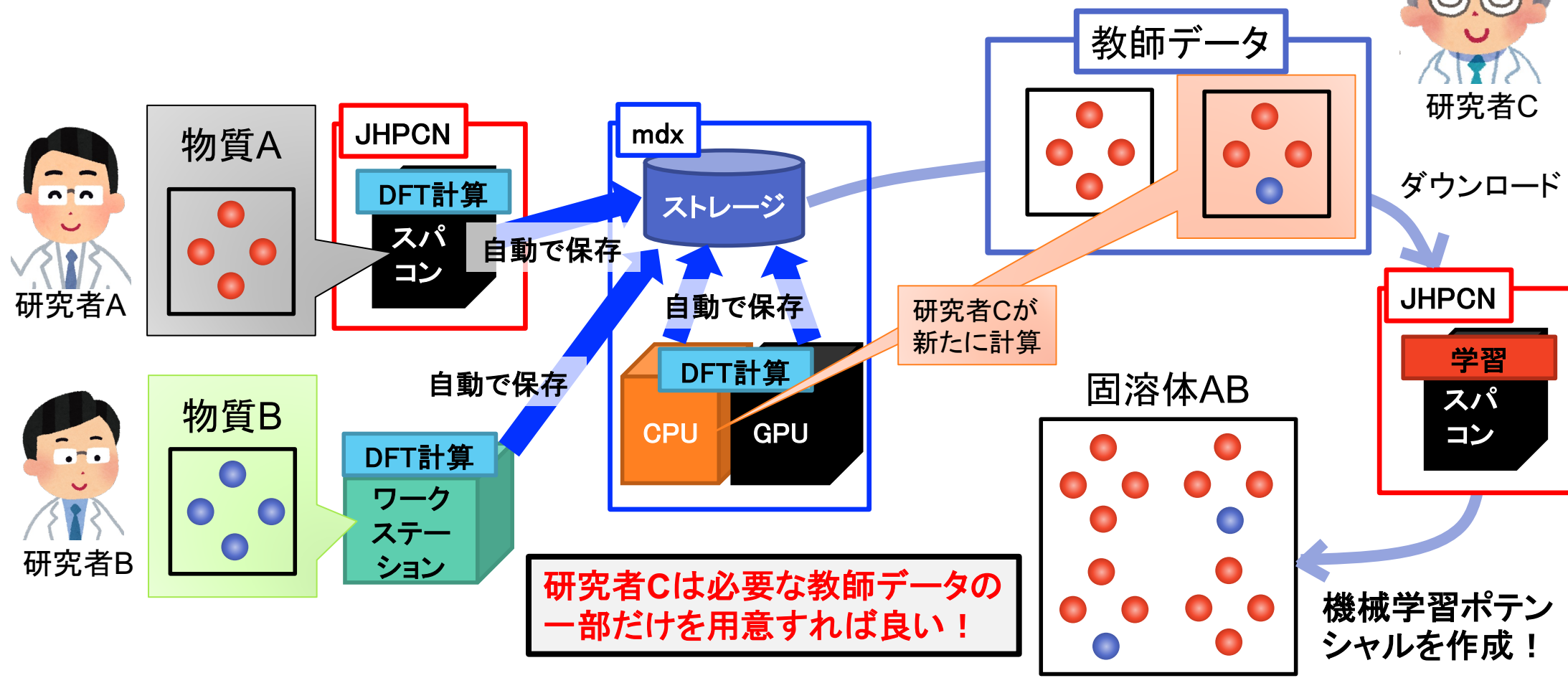


- 物質名や元素名で検索
- 様々な計算機環境で学習に利用

# 統合機械学習分子動力学システム

## ・利用例

- ・ 研究者A: 物質Aの教師データ、機械学習ポテンシャルを作成済み
- ・ 研究者B: 物質Bの教師データ、機械学習ポテンシャルを作成済み
- ・ 研究者C: 固溶体ABの機械学習ポテンシャルを作成したい





# まとめ

- 機械学習分子動力学法の紹介
  - 人工ニューラルネットワーク等で第一原理計算の結果を学習
  - 高予測精度と低計算コストを両立
  - 物質ごとに教師データを用意し、人工ニューラルネットワークのトレーニングが必要
  - 教師データはほとんど公開されていない
- 統合機械学習分子動力学システムの概略紹介
  - 教師データの収集・検索・利用のためのデータベース作成
  - 最も計算資源を必要とするのが教師データ作成であるため、教師データを共有することにより、機械学習ポテンシャルの構築が容易になる
  - 将来的には、機械学習ポテンシャルの自動生成を実現したい

# 機械学習分子動力学法詳細 1/3

## • 特徴

- 機械学習を用いて、高予測精度かつ低計算コストを実現
- 小さい系で学習して大きな系を計算可能

## • 仮定、その他

- 全エネルギーは各原子に付随する「原子エネルギー」に分割
- 原子エネルギーはカットオフ半径内に存在する周辺原子の配置で決まる
  - 長距離力は十分にスクリーンされている
- 周辺原子配置は並進・回轉變換普遍的な「記述子」と呼ばれる特徴量の関数

$$E_{\text{tot}} = \sum_{i=1}^N \epsilon \left( \mathbf{D}_{r_c}^{(i)} \right)$$

記述子  $\mathbf{D}_{r_c}^{(i)}$

- カットオフ半径  $r_c$  内に存在する周辺原子配置の特徴量
- 並進・回轉變換不変

全エネルギー  $E_{\text{tot}}$

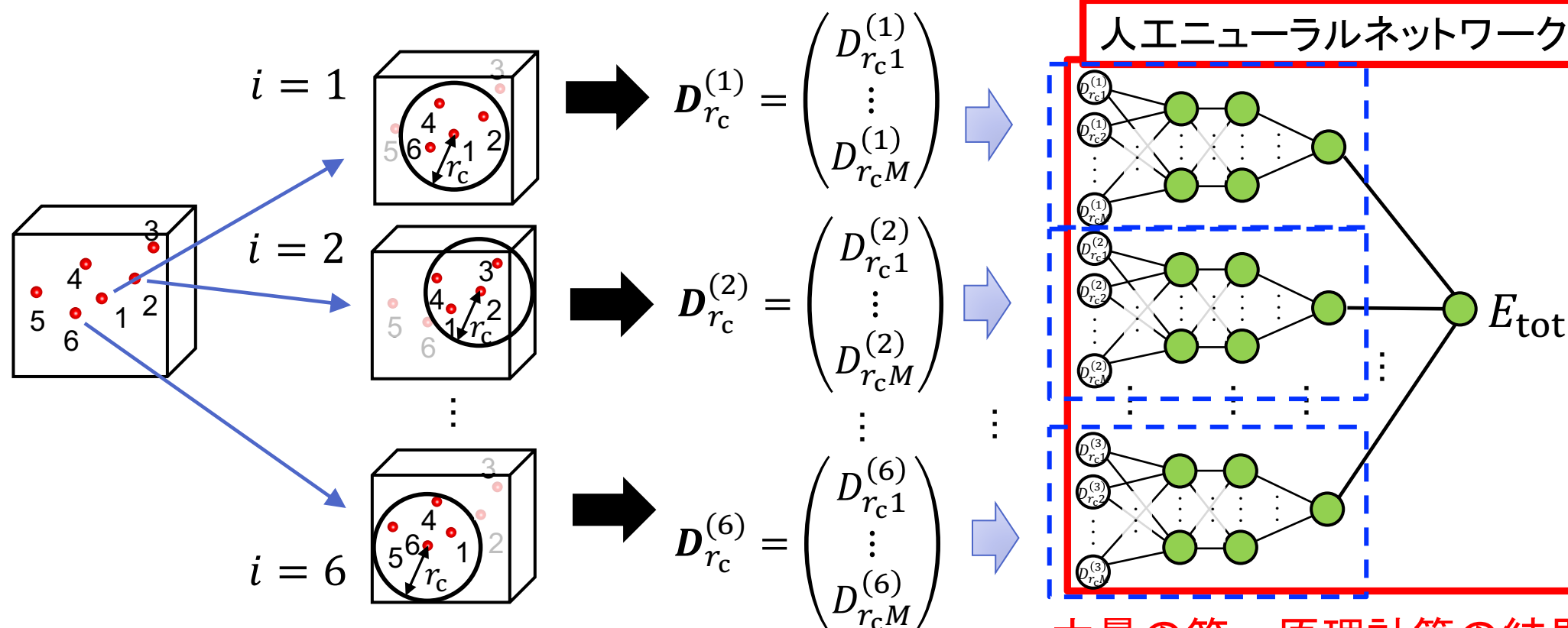
原子エネルギー  $\epsilon$

入力  $\mathbf{D}_{r_c}^{(i)}$  の非線形関数

人工ニューラルネットワークで近似

# 機械学習分子動力学法詳細 2/3

- 例: 6個の原子で構成される系



記述子  $D_{r_c}^{(i)}$  ( $M$ 次元)

- カットオフ半径  $r_c$  内に存在する周辺原子配置の特徴量
- 並進・回轉變換不変

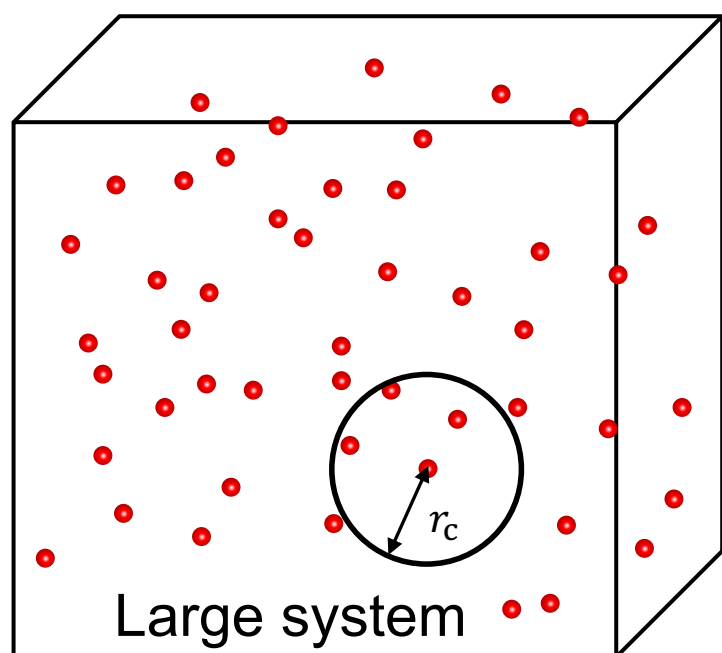
$$E_{\text{tot}} = \sum_{i=1}^6 \epsilon \left( D_{r_c}^{(i)} \right)$$

大量の第一原理計算の結果を教師データとしてトレーニング

非線形関数  $\epsilon \left( D_{r_c}^{(i)} \right)$  が得られる

# 機械学習分子動力学法詳細 3/3

- 大きな系への適用

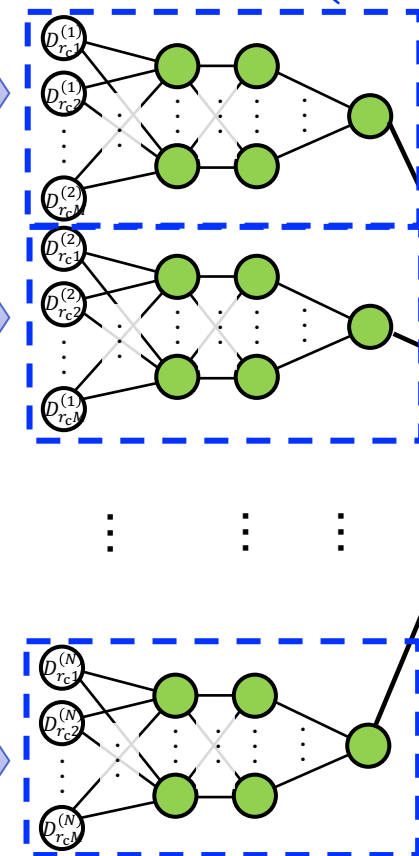


$$\mathbf{D}_{r_c}^{(1)} = \begin{pmatrix} D_{r_c1}^{(1)} \\ \vdots \\ D_{r_cM}^{(1)} \end{pmatrix}$$

$$\mathbf{D}_{r_c}^{(2)} = \begin{pmatrix} D_{r_c1}^{(2)} \\ \vdots \\ D_{r_cM}^{(2)} \end{pmatrix}$$

$$\mathbf{D}_{r_c}^{(N)} = \begin{pmatrix} D_{r_c1}^{(N)} \\ \vdots \\ D_{r_cM}^{(N)} \end{pmatrix}$$

トレーニング済みの  
非線形関数  $\epsilon(\mathbf{D}_{r_c}^{(i)})$



$$E_{\text{tot}}^{\text{large}} := \sum_{i=1}^N \epsilon(\mathbf{D}_{r_c}^{(i)})$$

トレーニング済みの  $\epsilon(\mathbf{D}_{r_c}^{(i)})$  を組み合わせて、大きな系の  
全エネルギーを表す人工ニューラルネットワークを構築する

全エネルギーを原子位置で微分  
すれば力が得られる  
→ **分子動力学計算が可能!**