

機械学習分子動力学法の応用

奥村雅彦

日本原子力研究開発機構
システム計算科学センター
シミュレーション技術開発室



分子動力学法

- 与えられたポテンシャルによって相互作用する原子/分子の動的過程の数値的シミュレーション手法
 - 原子スケールの様々な物理量を計算可能
 - 材料科学や生命科学などへの広い応用
- 手法と特徴
 - 古典分子動力学法
 - 量子性を無視し、原子やイオンを古典粒子として扱う
 - 第一原理分子動力学
 - 原子核と電子をそれぞれ古典粒子、波動関数で記述する

	精度	計算コスト
古典分子動力学法	低	高
第一原理分子動力学法	高	低

分子動力学法

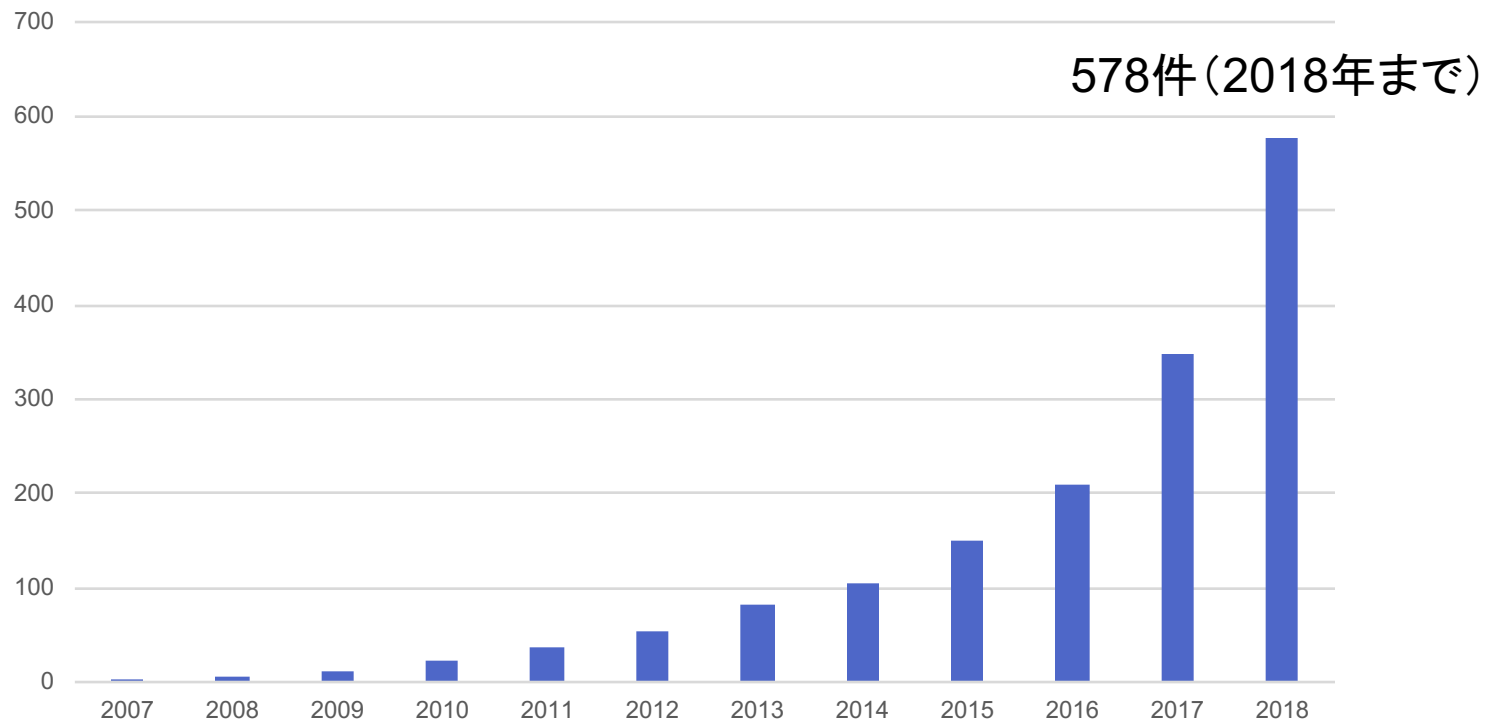
- 与えられたポテンシャルによって相互作用する原子/分子の動的過程の数値的シミュレーション手法
 - 原子スケールの様々な物理量を計算可能
 - 材料科学や生命科学などへの広い応用
- 手法と特徴
 - 古典分子動力学法
 - 量子性を無視し、原子やイオンを古典粒子として扱う
 - 第一原理分子動力学
 - 原子核と電子をそれぞれ古典粒子、波動関数で記述する

	精度	計算コスト
古典分子動力学法	低	高
第一原理分子動力学法	高	低
 機械学習分子動力学法	高	低

機械学習分子動力学法

- 機械学習を用いた分子動力学法
 - 第一原理計算の結果を機械学習してポテンシャルを構成
 - Behler-Parrinelloの方法
 - J. Behler and M. Parrinello, Phys. Rev. Lett **98**, 146401 (2007).

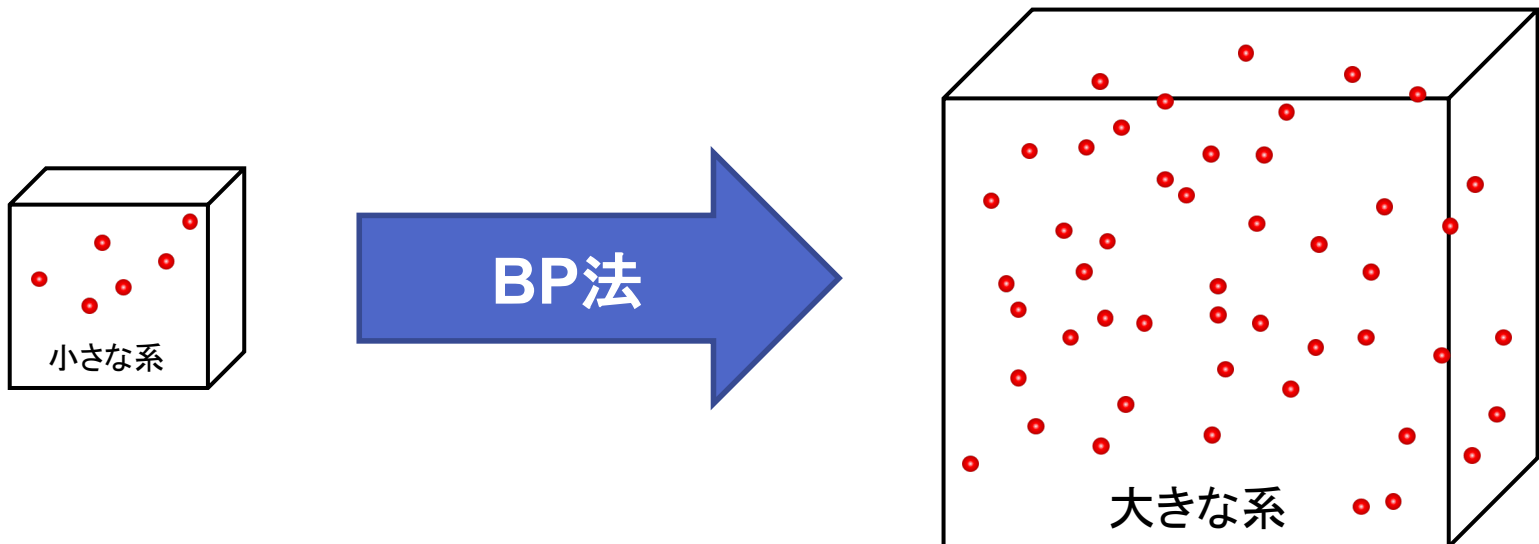
引用数  Scholar



機械学習分子動力学法

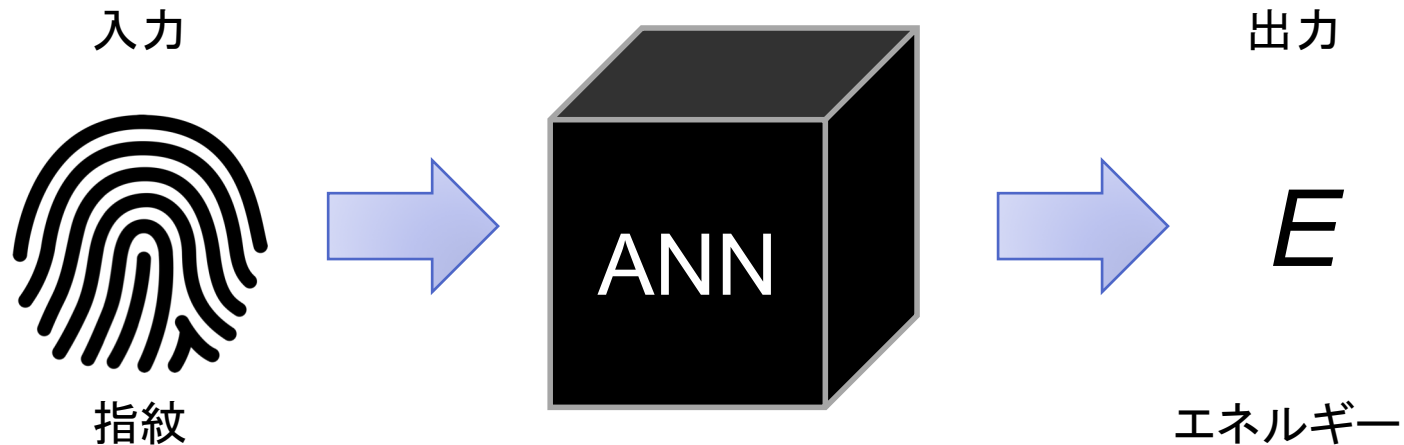
- Behler-Parrinello法の特徴

- 機械学習に人工ニューラルネットワークを使用
 - ガウシアン過程等は多くの学習データを利用できない
 - 人工ニューラルネットワークは複雑な原子配置に対応可能と考えられる
- 学習した系よりも大きな系のシミュレーションが可能
 - 各原子に割り当てられた“部分エネルギー”を定義
 - 部分エネルギーの和として全系のエネルギーを求める



機械学習分子動力学法

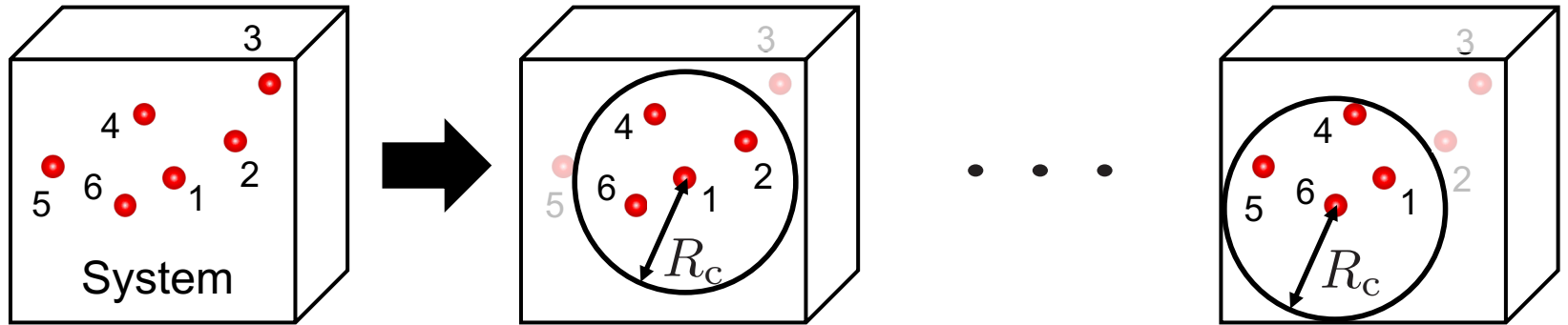
- Behler-Parrinello法
 - 基本構造



指紋とエネルギーの対応を学習

“指紋”

各原子を中心として、cutoff半径 R_c 内の局所原子分布を G_i^{dist} で特徴付ける



$$G_1^{\text{dist}} = \sum_{i \in \Omega_1^{R_c}} e^{-\eta(R_{i1} - R_s)^2}$$

$$\Omega_1^{R_c} = \{2, 4, 6\}$$

$$G_6^{\text{dist}} = \sum_{i \in \Omega_6^{R_c}} e^{-\eta(R_{i6} - R_s)^2}$$

$$\Omega_6^{R_c} = \{1, 4, 5\}$$

- 原子分布が異なれば、 G_i^{dist} の値も異なる。
- この意味で、 G_i^{dist} の値は原子分布の“指紋”である。



...

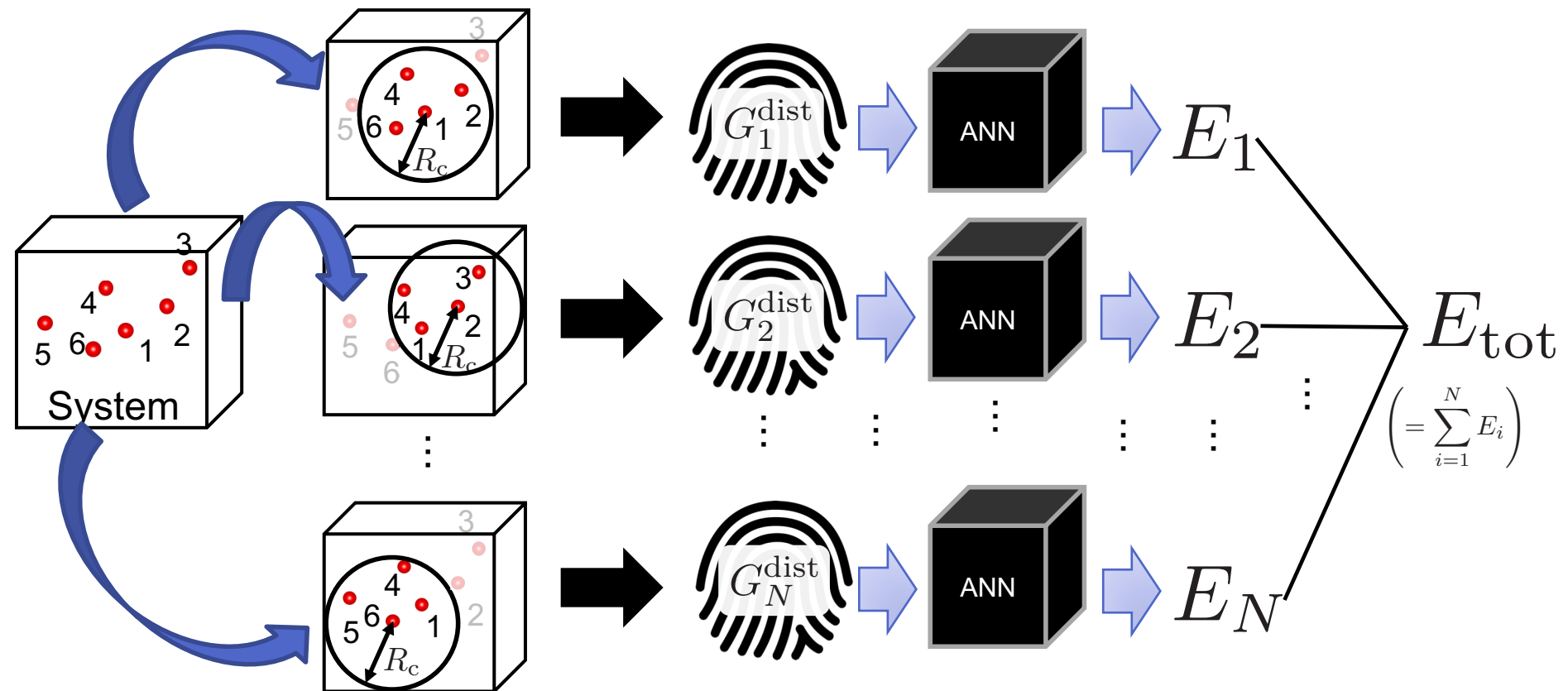


* 通常は、 G_i^{dist} の他に原子間角度を特徴付ける関数も同時に“指紋”として使用される

η, R_s : パラメーター

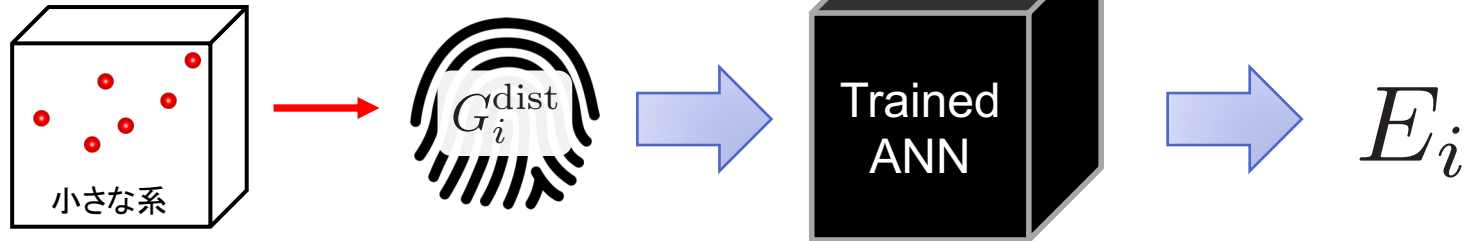
“指紋”と部分エネルギー

- 局所原子分布の“指紋” G_i^{dist} を入力すると、対応する局所部分エネルギー E_i が得られるように訓練する

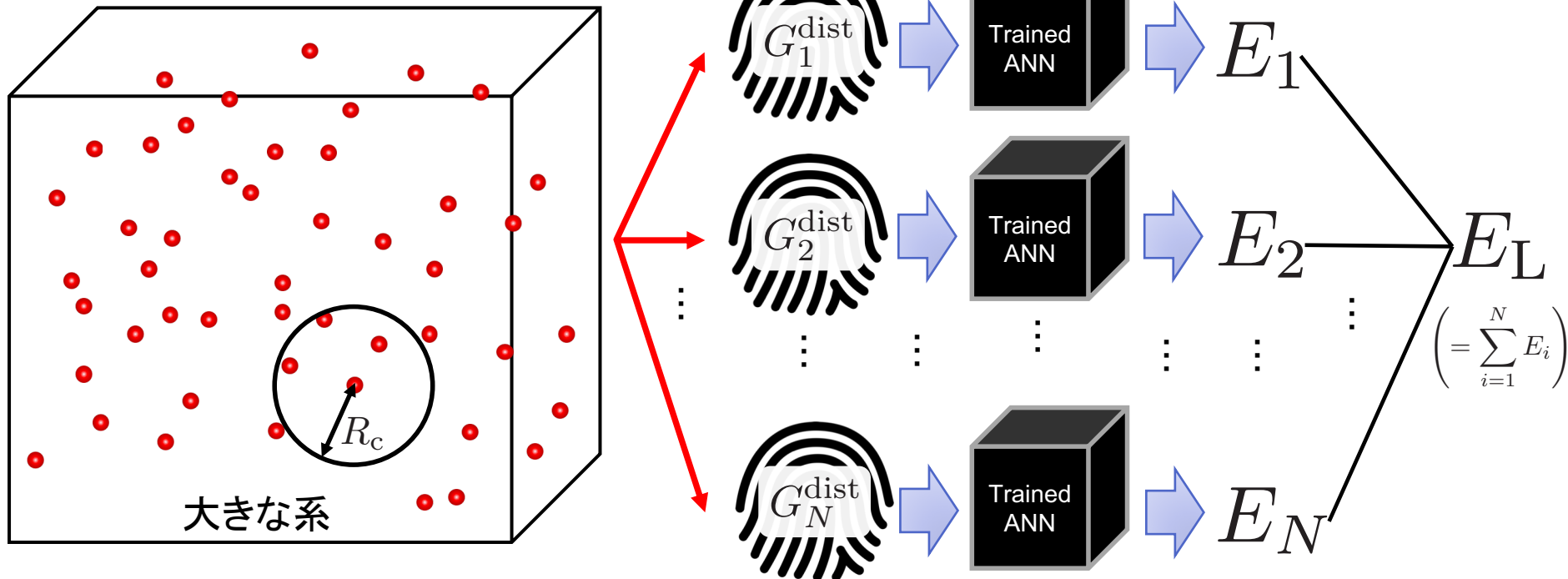


“指紋”と部分エネルギー

- 小さな系で学習

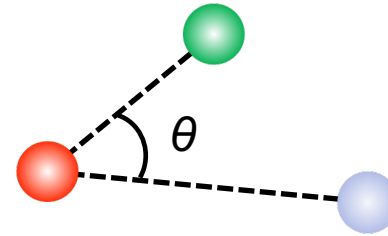


大きな系に拡張



Behler-Parrinello法の弱点と改善手法

- 多原子種が難しい
 - “指紋”が原子種数に依存
 - 例: M 種原子間の角度 $\theta = M \times {}_M C_2$
- Artrith-Urban-Cederの提案
 - 動径分布: 原子種ごとに背の高さを変える
 - Chebyshev多項式で分布を展開
 - 展開係数を“指紋”とする

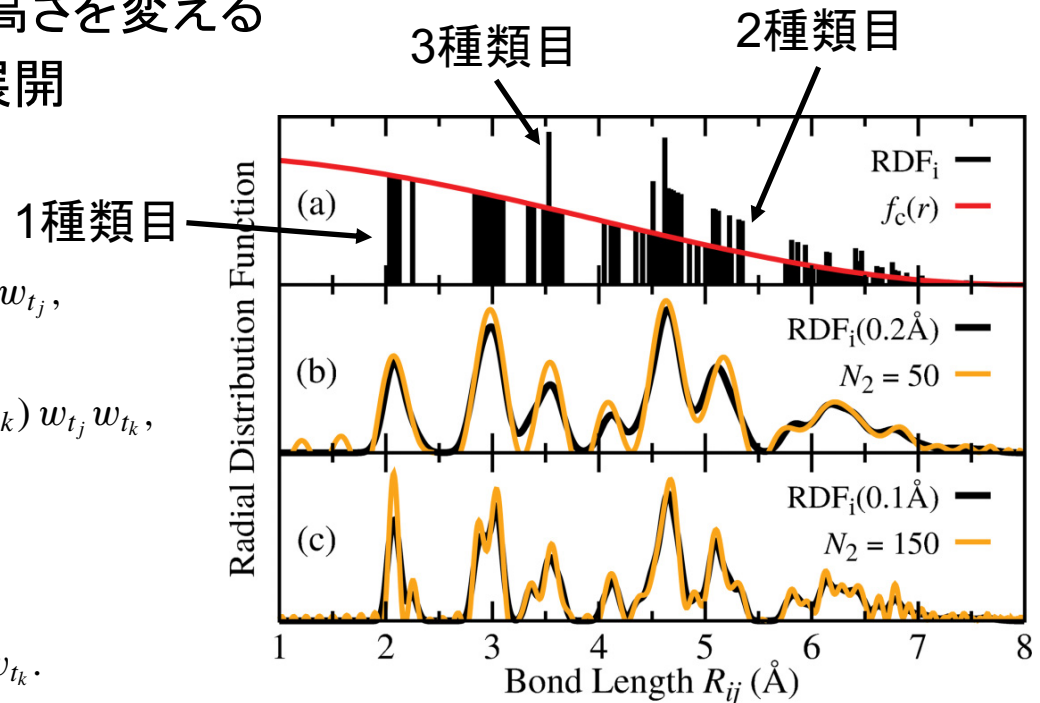


$$\text{RDF}_i(r) = \sum_{\mathbf{R}_j \in \sigma_i^{Rc}} \delta(r - R_{ij}) f_c(R_{ij}) w_{t_j},$$

$$\text{ADF}_i(\theta) = \sum_{\mathbf{R}_j, \mathbf{R}_k \in \sigma_i^{Rc}} \delta(\theta - \theta_{ijk}) f_c(R_{ij}) f_c(R_{ik}) w_{t_j} w_{t_k},$$

$$c_\alpha^{(2)} = \sum_{\mathbf{R}_j \in \sigma_i^{Rc}} \phi_\alpha(R_{ij}) f_c(R_{ij}) w_{t_j},$$

$$c_\alpha^{(3)} = \sum_{\mathbf{R}_j, \mathbf{R}_k \in \sigma_i^{Rc}} \phi_\alpha(\theta_{ijk}) f_c(R_{ij}) f_c(R_{ik}) w_{t_j} w_{t_k}.$$



The Atomic Energy Network (aenet)

- Behler-Parrinello法のオープンソースコード

- N. Artrith, A. Urban, Comp. Mat. Sci. **114**, 135 (2016).
- ライセンス: Mozilla Public License
- 実装の特徴
 - “指紋”(記述子)
 - Behler-Pariinelloのsymmetry function
 - Artrith-Urban-CederのChebyshev多項式展開係数
 - 学習
 - LBFGS法によるバッチ学習(MPI並列)
 - 分子動力学
 - Atomic Simulation Environmentを利用
 - Python
 - シリアル(並列化されていない)

← 多原子種に対応

← スパコン向き

The Atomic Energy Network (aenet)

- Behler-Parrinello法のオープンソースコード

- N. Artrith, A. Urban, Comp. Mat. Sci. **114**, 135 (2016).
- ライセンス: Mozilla Public License
- 実装の特徴
 - “指紋”(記述子)
 - Behler-Pariinelloのsymmetry function
 - Artrith-Urban-CederのChebyshev多項式展開係数
 - 学習
 - LBFGS法によるバッチ学習(MPI並列)
 - 分子動力学
 - Atomic Simulation Environmentを利用
 - Python
 - シリアル(並列化されていない)

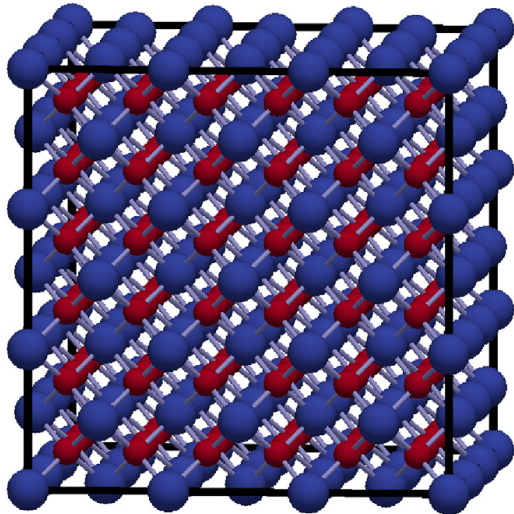
← 多原子種に対応

← スパコン向き

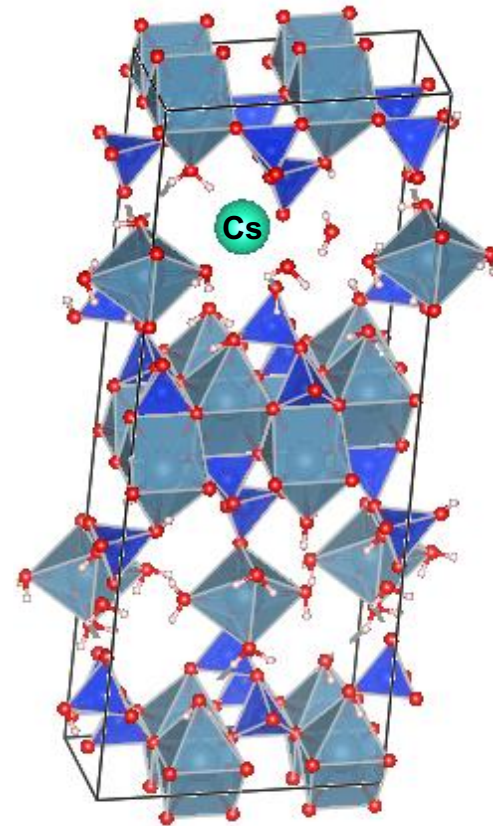
→ 簡易的な並列化

対象物質

- 少数原子種 (2原子種)
 - 二酸化トリウム (ThO_2)
- 多数原子種 (4原子種 + Cs)
 - コンクリート



二酸化トリウム




コンクリート
(トベモライト)

対象物質

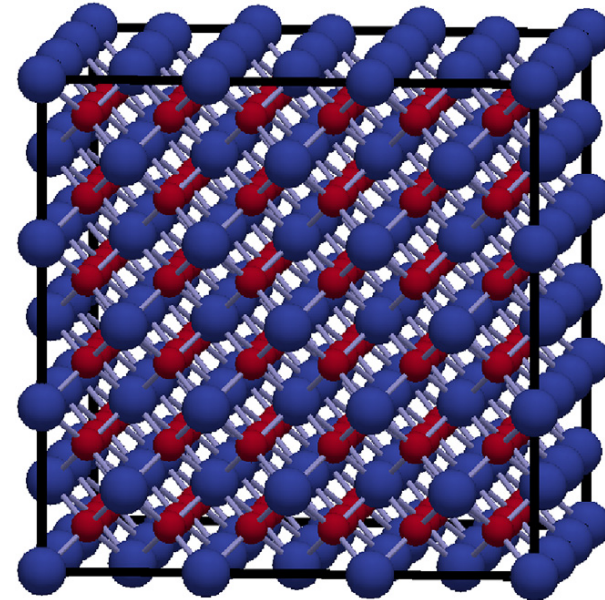
- 少数原子種 (2種類)
 - 二酸化トリウム (ThO_2)
 - UO_2 と比較して
 - 融点が高い
 - 熱伝導率が高い
 - 熱膨張率が低い
 - 燃料として好ましい性質
 - 熱物性の理解が不十分
- 多数原子種 (5種類)
 - コンクリート+セシウム
 - 福島第一原発事故における放射性セシウムによるコンクリート汚染
 - コンクリートにおけるセシウム吸着様態は明らかでない
 - 固液界面におけるイオン吸着シミュレーションが有効



大きな系における高精度なシミュレーションが求められている  **機械学習MD**

二酸化トリウム

- 少数原子種 (2種類)
 - 二酸化トリウム (ThO_2)
 - UO_2 と比較して
 - 融点が高い
 - 熱伝導率が高い
 - 熱膨張率が低い
 - 燃料として好ましい性質
 - 融点等の高温での物性も重要
 - 熱物性の理解が不十分



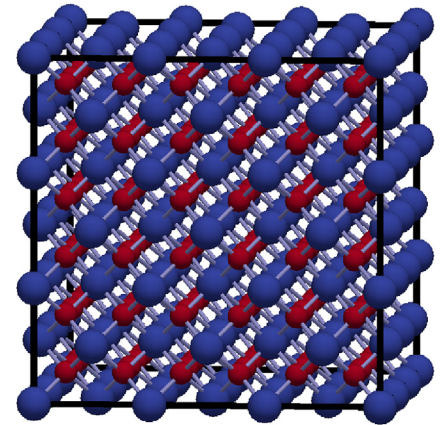
大きな系における高精度なシミュレーションが求められている



機械学習分子動力学法

Bredig転移

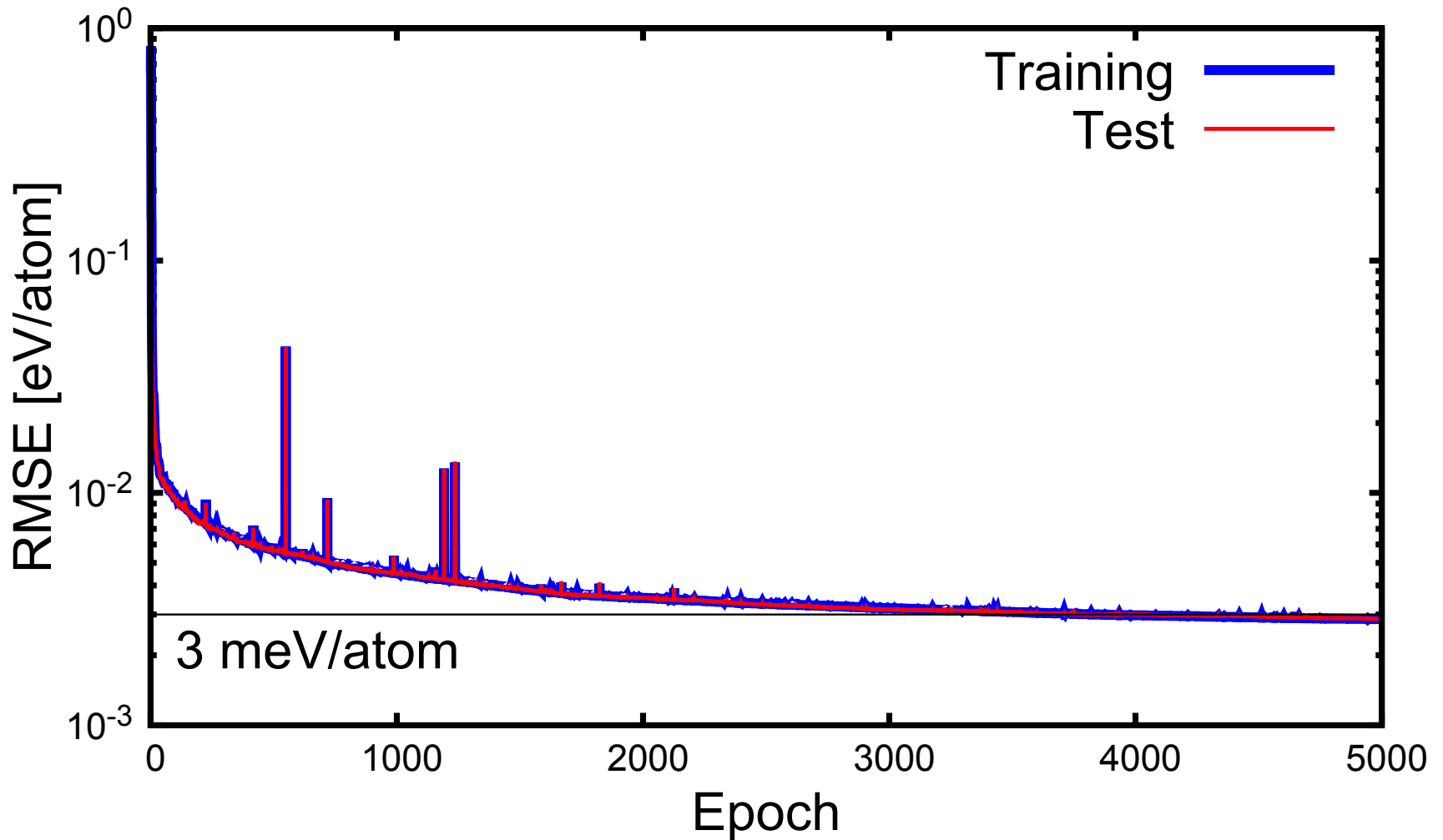
- Bredig転移
 - ホタル石 (CaF_2) 型の化合物は融点より150~400K下で比熱異常を示す
 - 可動イオンの秩序→無秩序転移と考えられている
- 古典分子動力学法によるBredig転移の解析
 - 転移温度は力場のパラメーターに強く依存する
 - S.I. Potashnikov *et al.*, J. Nucl. Mater. **419**, 217 (2011).
 - 対象物質: UO_2
- 密度汎関数法によるBredig転移の解析
 - H. Nakamura and M. Machida, J. Nucl. Mater. **478**, 56 (2016).
 - 対象物質: ThO_2
 - 経験的パラメータなしの解析
 - 転移の傾向が捉えられた
 - 有限サイズ効果の検証が必要(原子数:324個)



二酸化トリウムの機械学習分子動力学

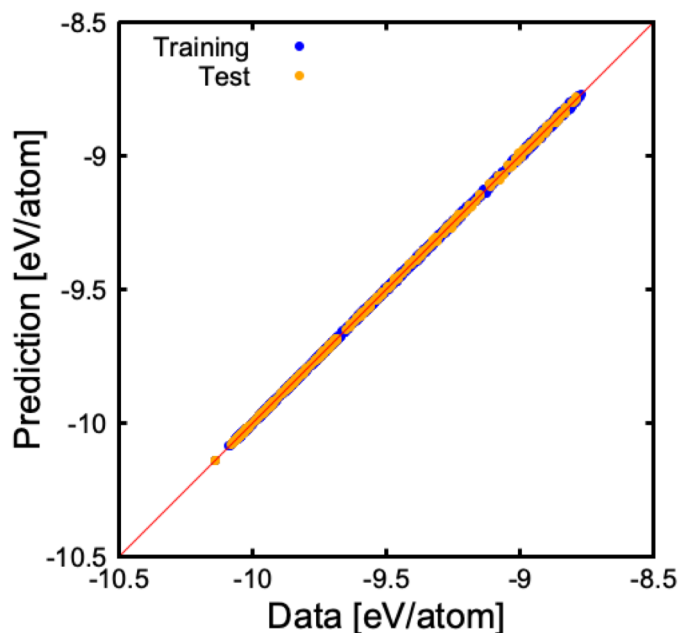
- 学習データセット
 - DFTコード: VASP (Vienna Ab initio Simulation Package)
 - Perdew-Burke-Ernzerhof revised for solids (PBE_sol)
 - Th: 108個、O: 216個
 - $T = 100 \sim 5,000$ K (100Kおき) 各10,000ステップ
 - 各温度1,000個をランダムに選んで学習セットを構成(計50,000個)
- Behler-Parrinello人工ニューラルネットワーク
 - “指紋”: Chebyshev多項式展開
 - 隠れ層: 1層
 - ユニット数: 15個
 - 活性化関数: tanh
 - 最適化法: LBFGS (Limited memory Broyden–Fletcher–Goldfarb–Shannon algorithm)

學習

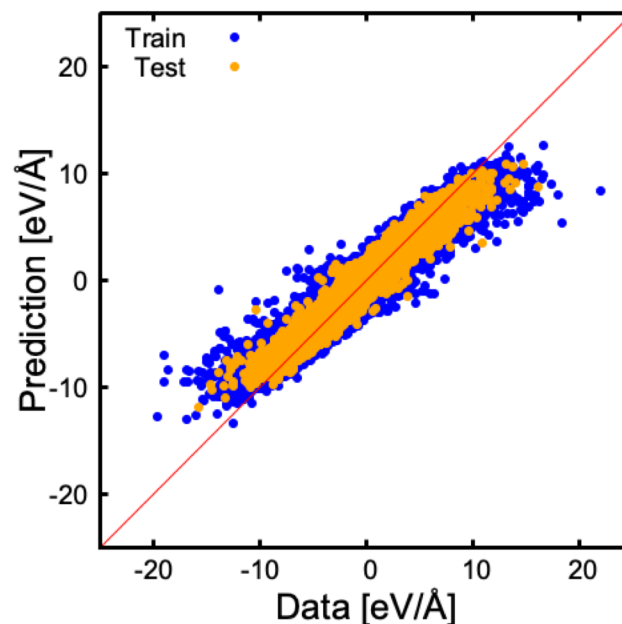


エネルギーと力の精度

- エネルギーの誤差: 約 3 meV/atom
- 力の誤差: 約 0.3 eV/Å

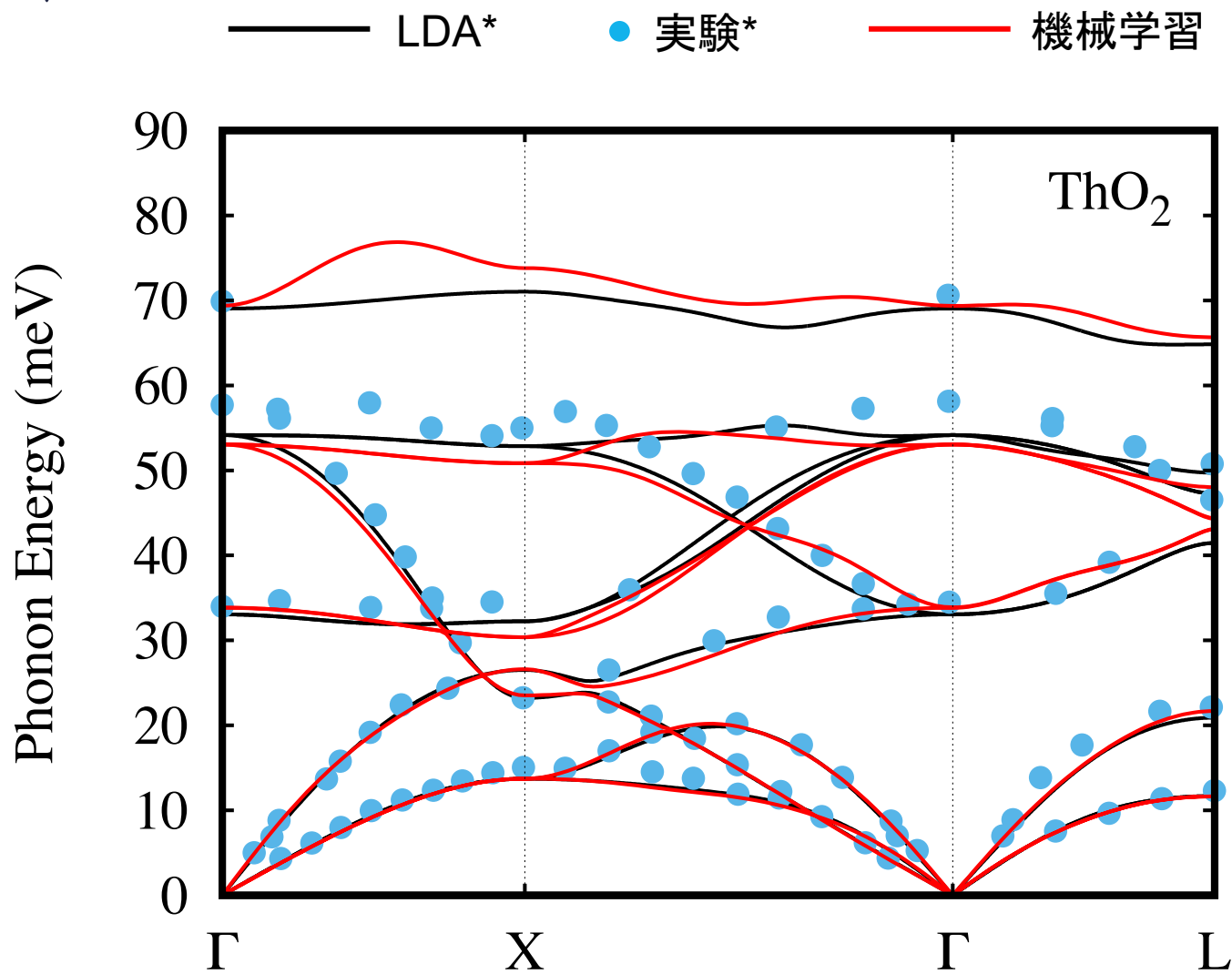


Train: RMSE=0.002678, $R^2=0.999951$
Test : RMSE=0.003457, $R^2=0.999920$



Train: RMSE=0.337745, $R^2=0.957046$
Test : RMSE=0.329458, $R^2=0.957825$

フォノン



実験、DFT計算両者と非常に良く一致

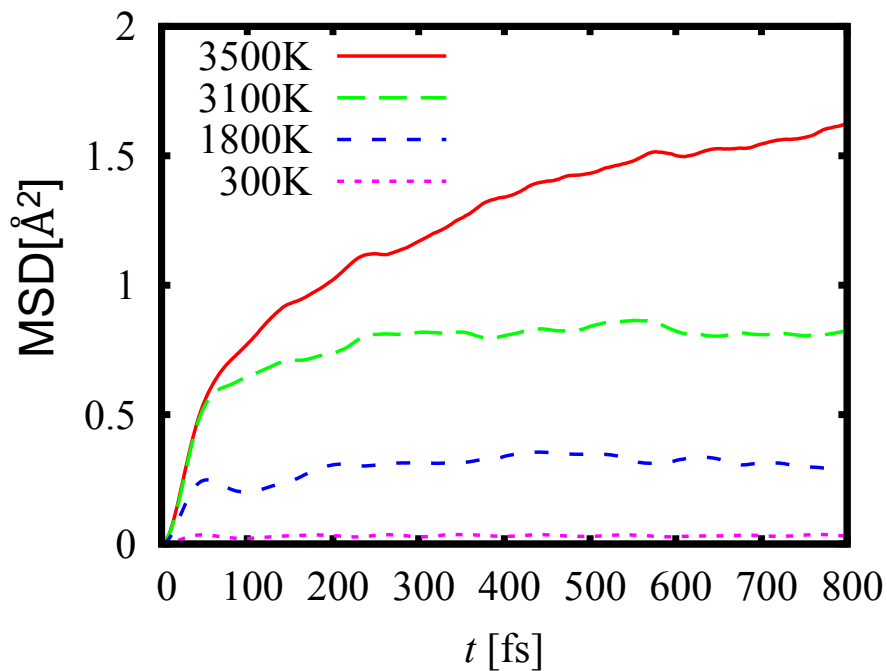
* K. Clausen *et al.*, J. Chem. Soc., Faraday Trans. **2**, 1109 (1987).

Bredig転移の解析 (preliminary)

酸素原子の平均2乗変位 (MSD): $MSD = \langle [\mathbf{r}(t) - \mathbf{r}(0)]^2 \rangle$,

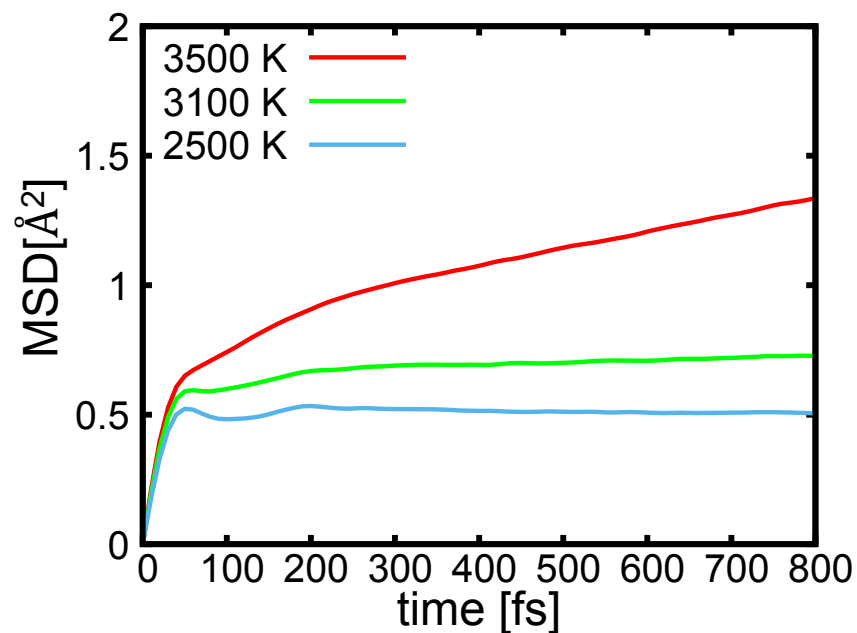
第一原理分子動力学法*

原子数: 324個



機械学習分子動力学法

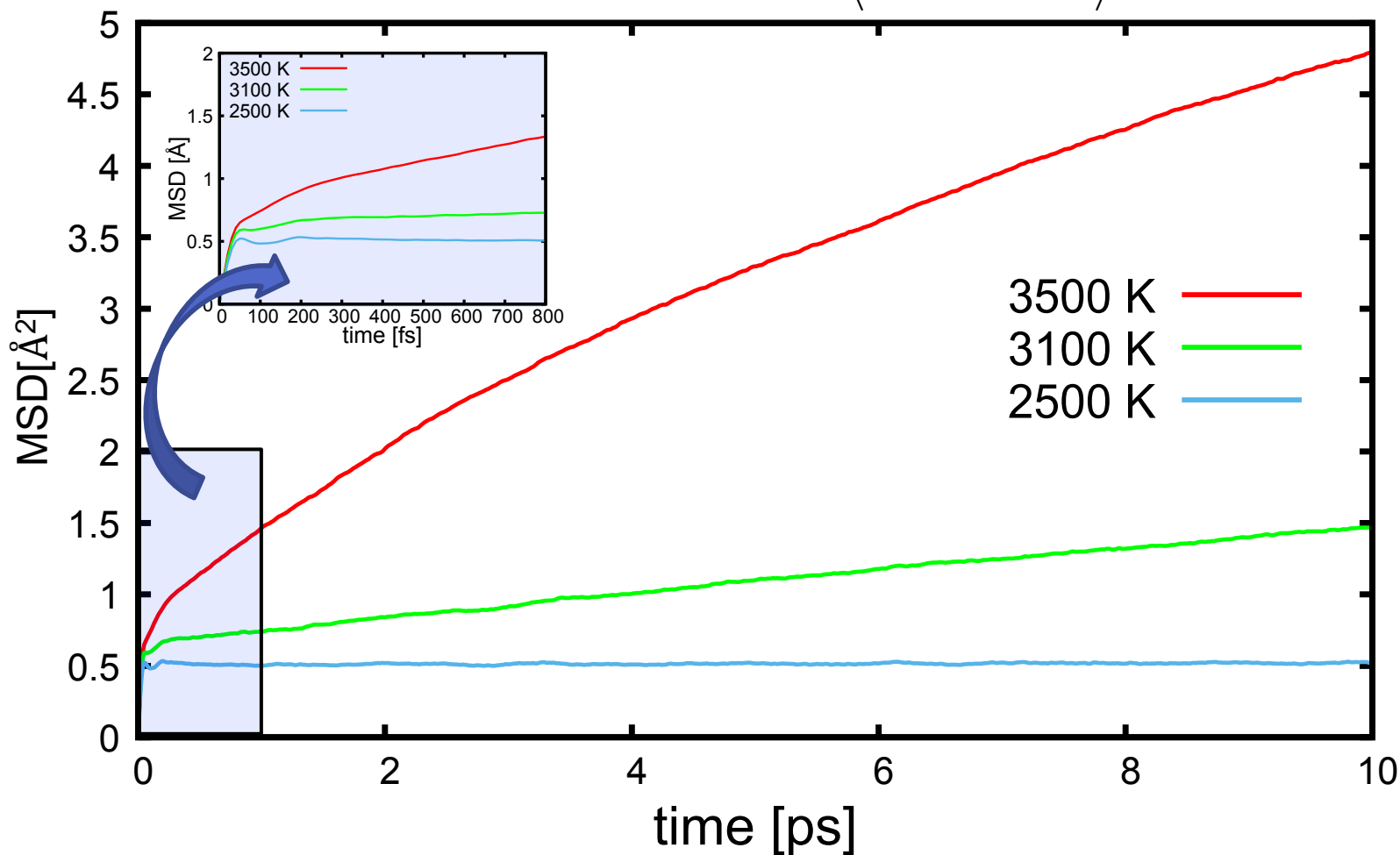
原子数: 2,592個



*H. Nakamura and M. Machida, J. Nucl. Mater. **478**, 56 (2016).

Bredig転移の解析 (preliminary*)

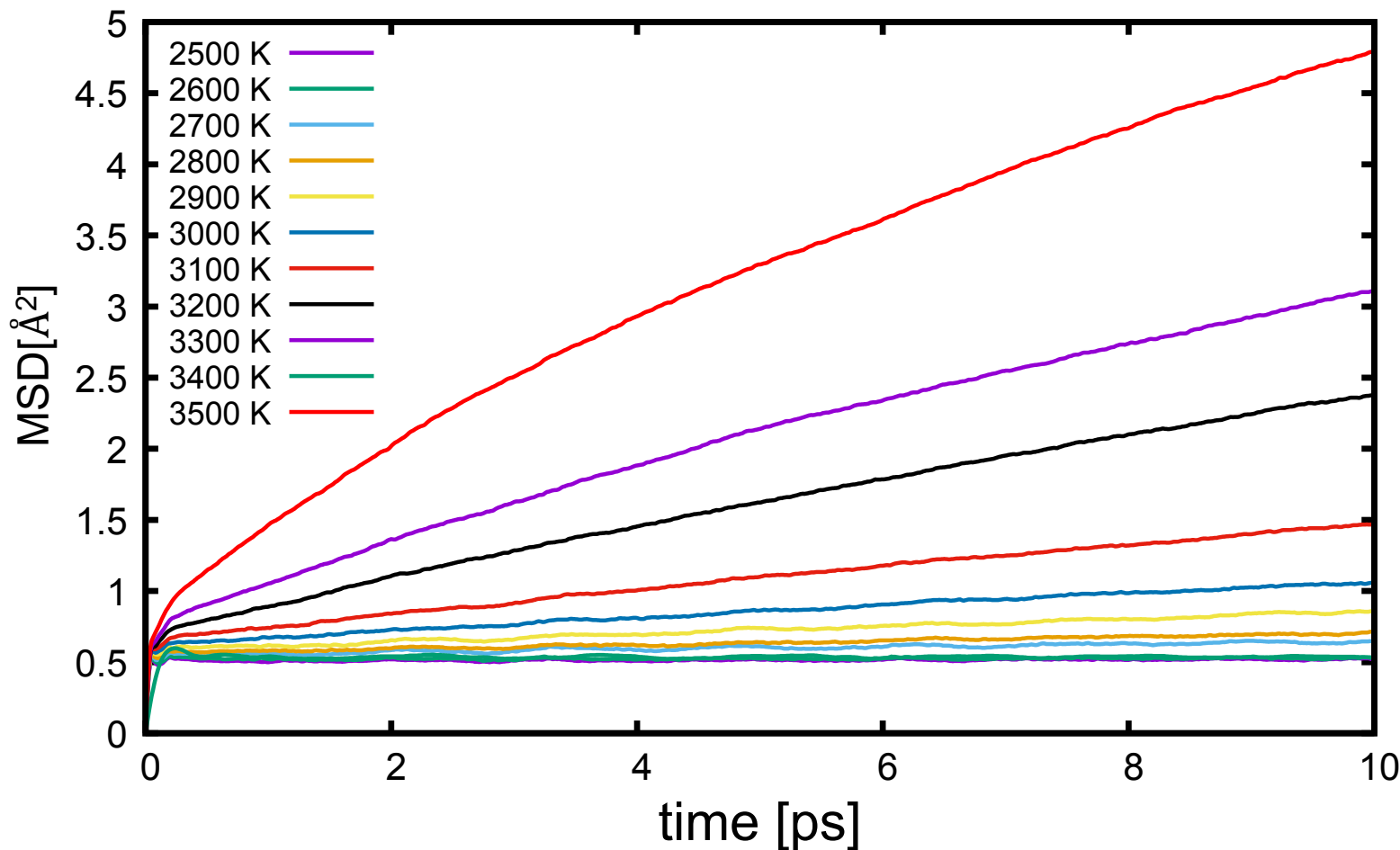
酸素原子の平均2乗変位 (MSD): $MSD = \langle [\mathbf{r}(t) - \mathbf{r}(0)]^2 \rangle$,



*ただし、体積の熱膨張の効果は入っていない

Bredig転移の解析 (preliminary*)

酸素原子の平均2乗変位 (MSD): $MSD = \langle [\mathbf{r}(t) - \mathbf{r}(0)]^2 \rangle$,



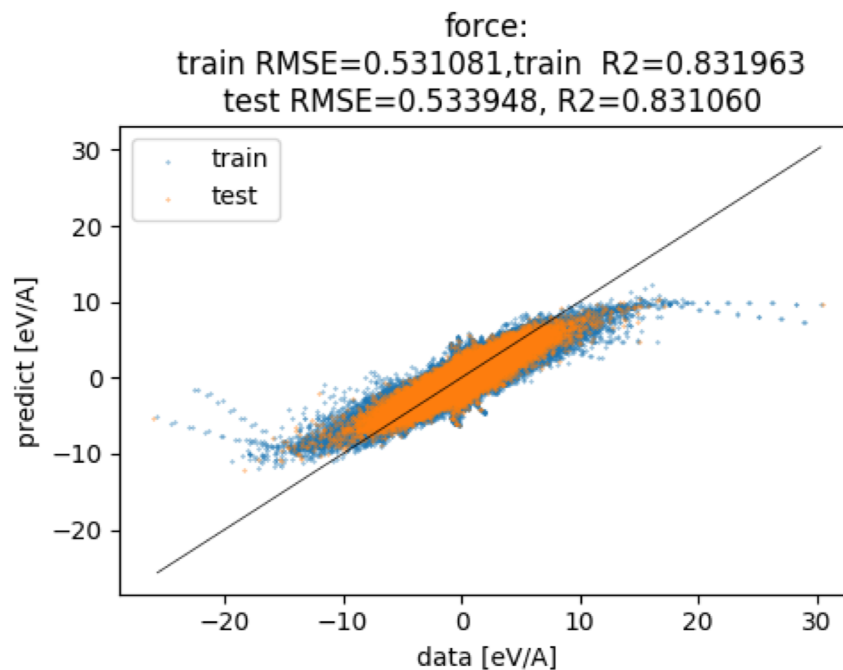
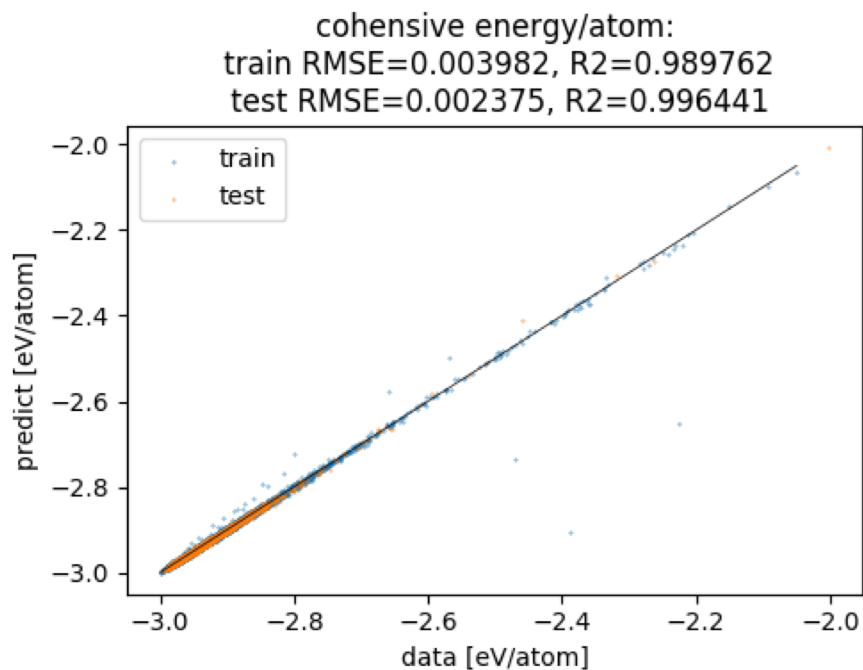
*ただし、体積の熱膨張の効果は入っていない

コンクリート

- 多数原子種 (4種類+Cs)
 - コンクリート+セシウム
 - 福島第一原発事故における放射性セシウムによるコンクリート汚染
 - コンクリートにおけるセシウム吸着様態は明らかでない
 - 固液界面におけるイオン吸着シミュレーションが有効
- データセット
 - トベモライト等 27構造
 - 学習データ: 55,245データ
- 人工ニューラルネットワーク
 - 隠れ層: 1層
 - ユニット数: 15個
 - “指紋”: Chebyshev多項式

エネルギーと力の精度

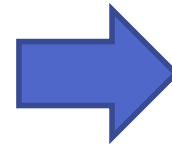
- エネルギーの誤差: 約 2~4 meV/atom
- 力の誤差: 約 0.5 eV/Å



まとめ

- 二酸化トリウム

- フォノンを精度よく計算可能
- 1万原子程度の分子動力学を実行
- Bredig転移の解析 (preliminary)



ThO₂の熱物性解明へ

- コンクリート

- 機械学習力場を作成
 - 4原子種 (Ca, Si, O, H) + Cs
- 精度検証
 - 良い精度



**コンクリート汚染メカニ
ズム解明へ**

- コード (aenet)

- 簡易的な並列化



独自コード開発

謝辞

- 小林恵太 (RIST)
- 中村博樹 (CCSE, JAEA)
- 永井祐紀 (CCSE, JAEA)
- 板倉充洋 (CCSE, JAEA)
- 志賀基之 (CCSE, JAEA)
- 町田昌彦 (CCSE, JAEA)

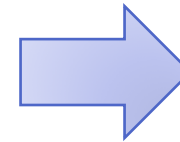
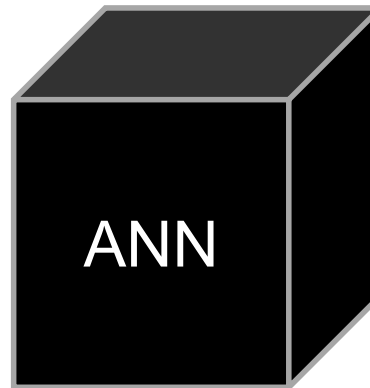
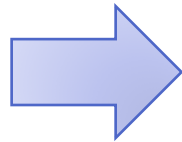


ご静聴ありがとうございました

“指紋”と部分エネルギー

- 局所原子分布の“指紋” G_i^{dist} を入力すると、対応する局所部分エネルギー E_i が得られるように訓練する

i 番目の原子を中心とした
局所原子分布の
“指紋”

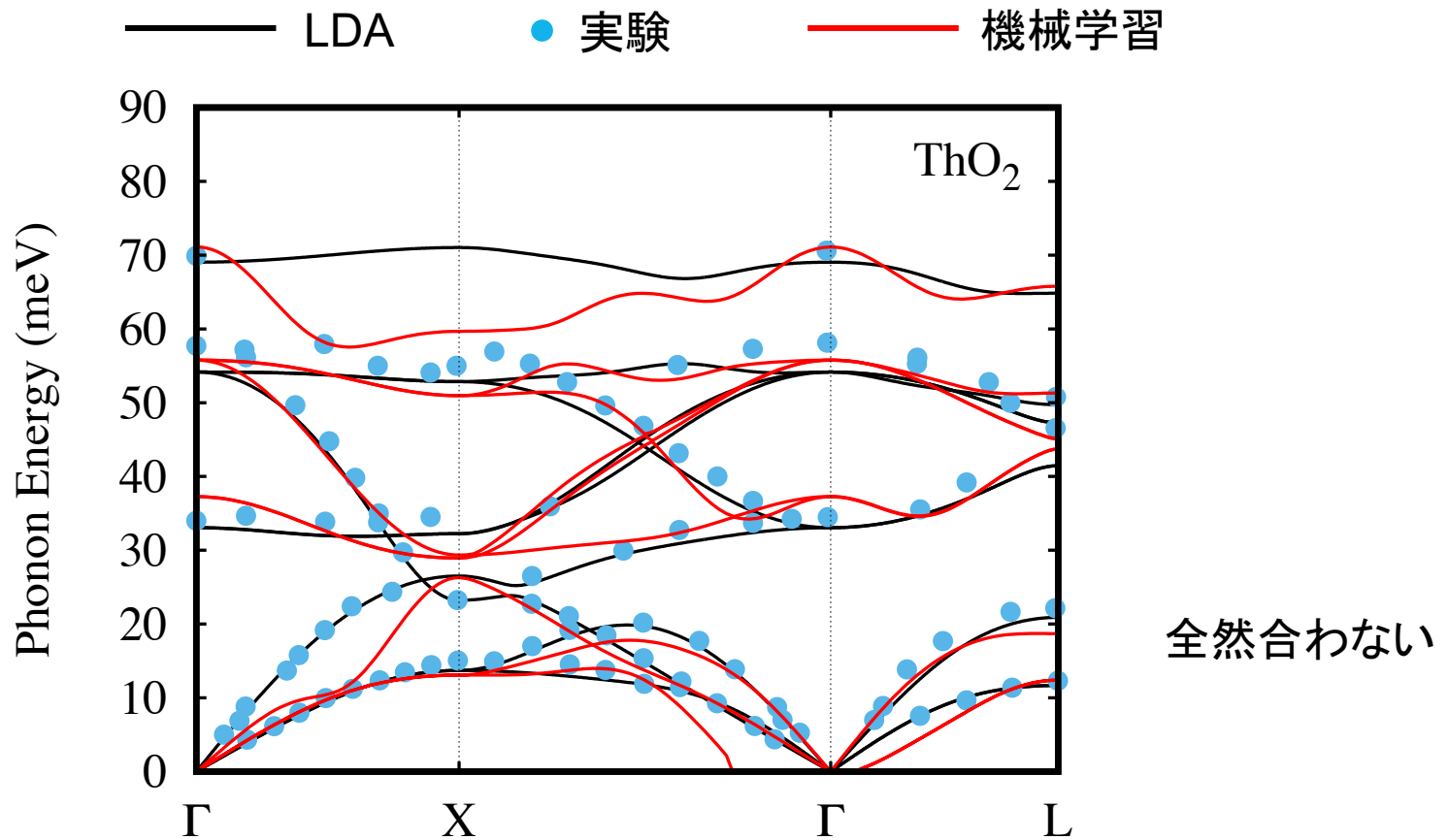


“局所部分エネルギー”

E_i

フォノンの評価 1/3

- 学習データセットの違い
 - 学習データセット: $T=100\text{K}$ の第一原理MD10,000データ



フォノンの評価 2/3

- 学習データセットの違い
 - 学習データセット: $T=100\sim 500\text{K}$ の第一原理MD 50,000データ

