

# 分子動力学計算の拡張に向けた 機械学習応用の動向

産業技術総合研究所

機能材料コンピューティショナルデザイン研究センター

主任研究員 安藤 康伸

# 自己紹介

名前：安藤康伸（35）

出身：愛知県西尾市

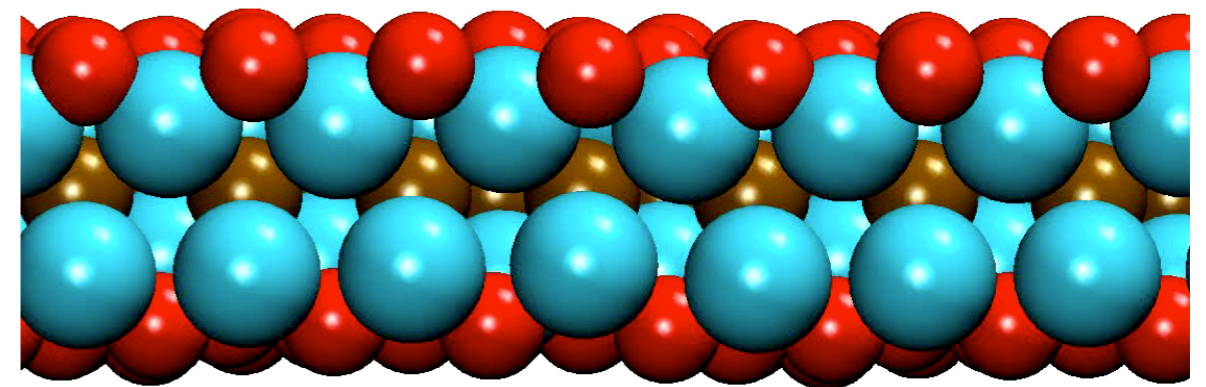
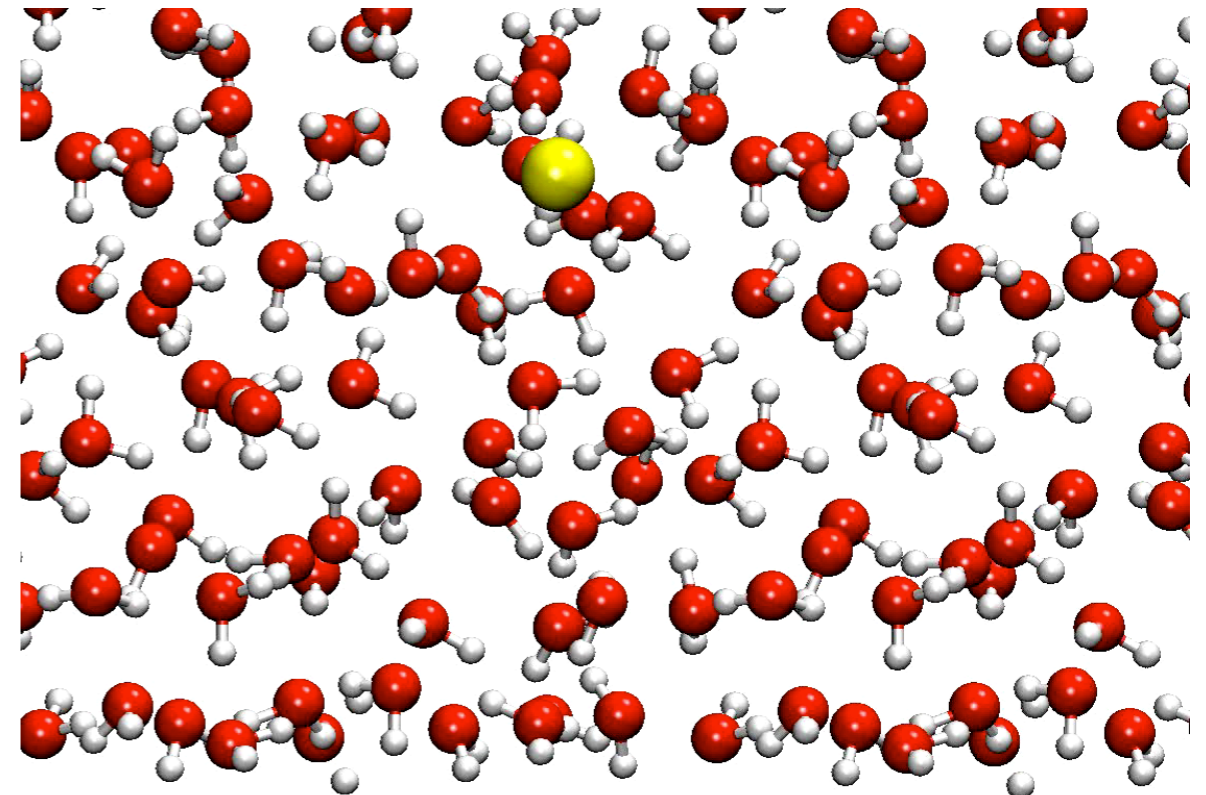
## 略歴

- ✓ 2012.3. 博士（理学） @東大院理
- ✓ 2012.4 ~ 2013.4 ポスドク @AIST
- ✓ 2013.5 ~ 2016.3 助教 @東大マテ工
- ✓ **2016.4** ~ 研究員 @ 機能材料コンピューテーショナルデザイン研究センター (CD-FMat), AIST

## 専門分野

**材料科学 x 情報科学**

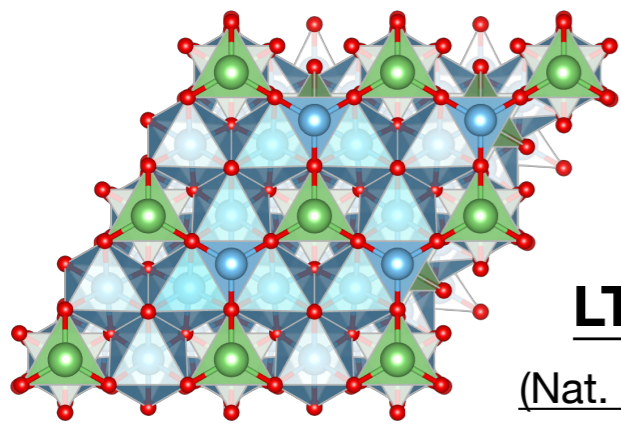
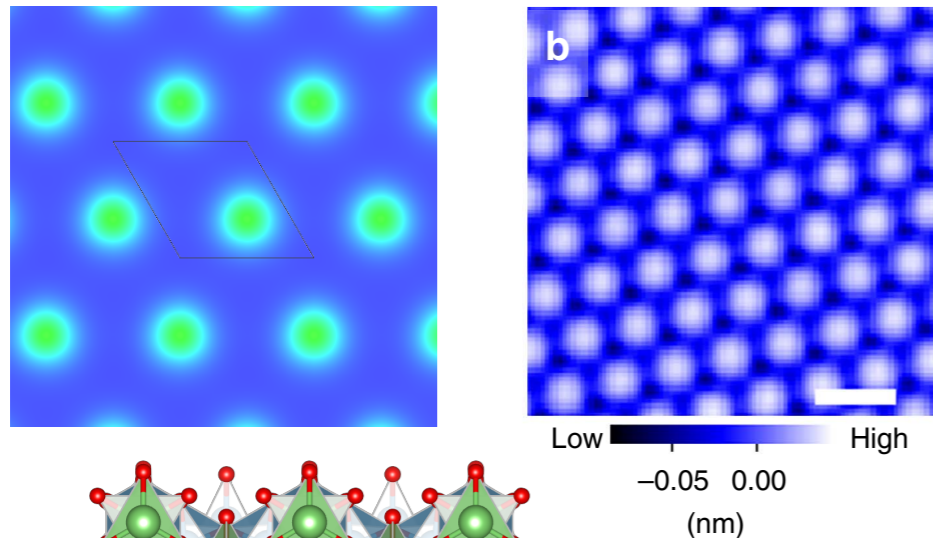
(物理・化学・数学・情報)



目には見えない電池の「内部」の再現

# 原子・分子シミュレーション

## 安定構造探索



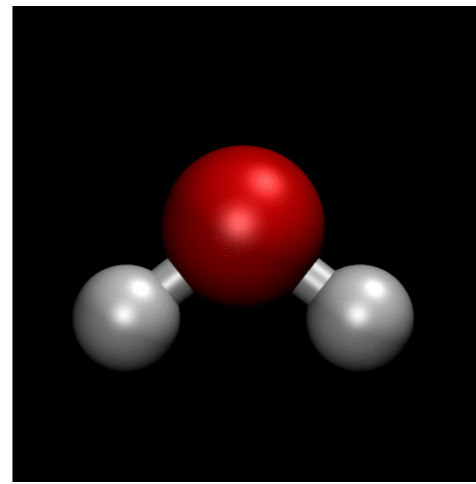
## LTO表面構造の同定

(Nat. Commun. 8, 15975 (2017))

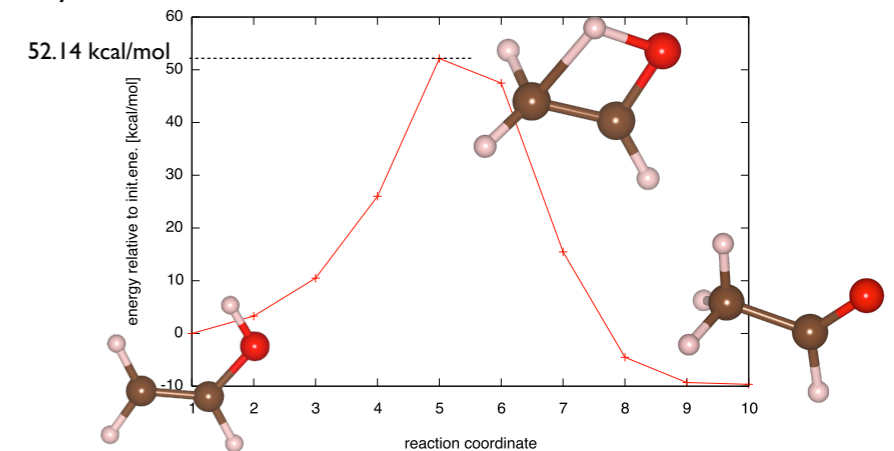
- 静的な安定状態から動的な挙動まで
- 構造と特性・観測データを結びつけるキー
- ナノスケール物理では必須のツール

原子間相互作用 (ポテンシャル)  
の存在が大前提

## 振動モード・反応座標

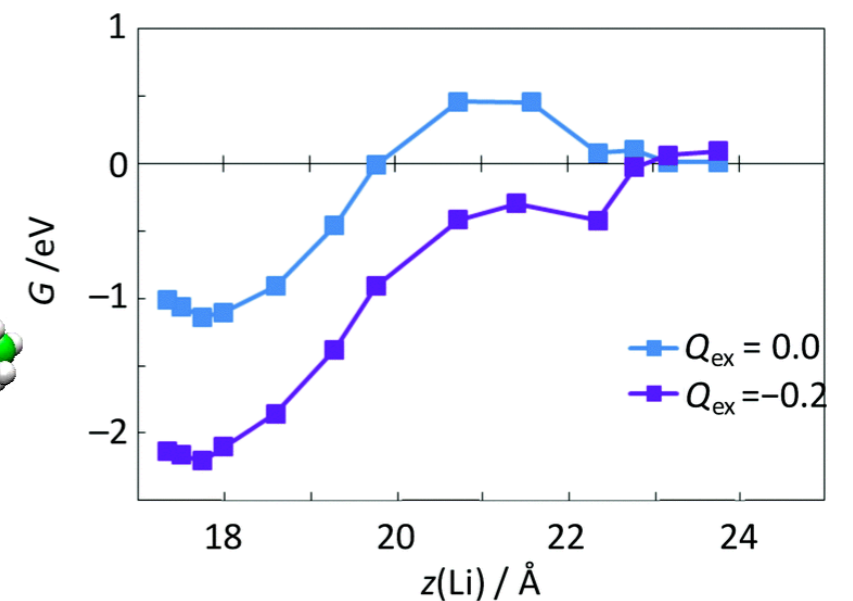
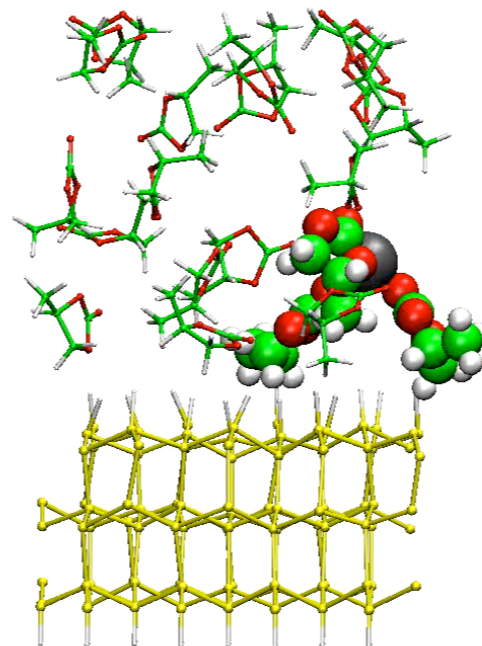


vinyl alcohol to acetaldehyde (NEB method)



DMol3 TS search (LST/QST) 51.473 cal/mol, -10.851 cal/mol (PBE)

## 統計サンプリング(自由エネルギー・エントロピー)



Phys. Chem. Chem. Phys., 2018, 20, 11586-11591

# 原子間ポテンシャル

二体バネモデルなど

## Lennard-Jones Potential

Proc. R. Soc. Lond. A, 106 463(1924).

$$U_{LJ}(r) = -4\epsilon \left( \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right)$$

- 気体粘性の温度依存性を「斥力」「引力」相互作用で説明する事が目的
- 気体の状態方程式の2次のビリアル展開を解析的に議論
- まだ分子の内部構造は十分分かっていない時代のこと

*On the Determination of Molecular Fields.—II. From the Equation of State of a Gas.*

By J. E. JONES, D.Sc., 1851 Exhibition Senior Research Student,  
Trinity College, Cambridge.

$$f(r) = \frac{\lambda_n}{r^n} - \frac{\lambda_m}{r^m},$$

本質をつき且つ解析的に扱える点がすごい

## Stillinger-Weber Potential

F. Stillinger and T. A. Weber, Phys. Rev. B 31, 5262 (1985).

$$f_2(r) = \begin{cases} A(Br^{-p} - r^{-q})\exp[(r-a)^{-1}], & r < a \\ 0, & r \geq a \end{cases}$$

$$f_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) = h(r_{ij}, r_{ik}, \theta_{jik}) + h(r_{ji}, r_{jk}, \theta_{ijk}) \\ + h(r_{ki}, r_{kj}, \theta_{ikj}),$$

$$h(r_{ij}, r_{ik}, \theta_{jik}) = \lambda \exp[\gamma(r_{ij} - a)^{-1} + \gamma(r_{ik} - a)^{-1}] \\ \times (\cos\theta_{jik} + \frac{1}{3})^2;$$

- Siなど共有結合性結晶でのstandard
- 3体項を加えて角度成分を表現
- カットオフ半径の導入
- シンプルな記述

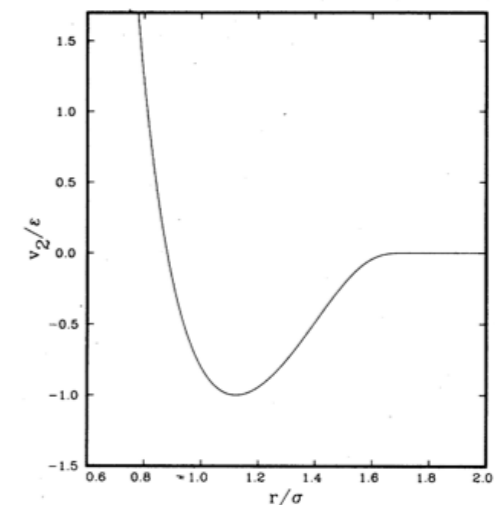


FIG. 2. Reduced pair potential vs distance. This function vanishes identically beyond  $r = 1.80$ .

数値計算に向けた設計がされている



# ポテンシャルのパラメータ決定

## Stillinger-Weber Potential

F. Stillinger and T. A. Weber, Phys. Rev. B 31, 5262 (1985).

### 満たすべき条件 (物性・特性)

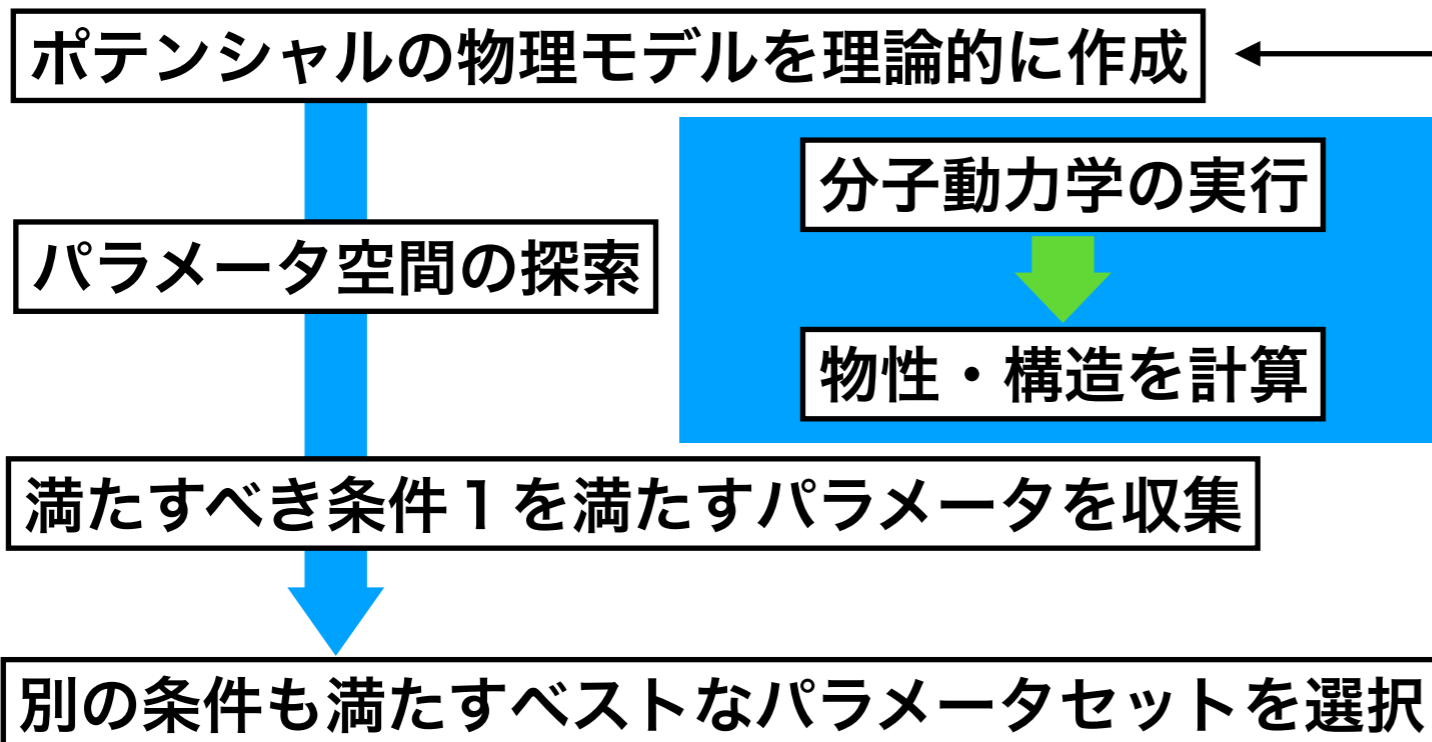
- 低温低圧でdiamond構造を取るように調整
- 融点や高温での液体構造を再現するように調整



Overall, the most satisfactory parameter set thus far discovered is the following:

$$\begin{aligned} A &= 7.049\ 556\ 277, & B &= 0.602\ 224\ 558\ 4, \\ p &= 4, & q &= 0, & a &= 1.80, \\ \lambda &= 21.0, & \gamma &= 1.20. \end{aligned} \quad (2.7)$$

## ポテンシャル構成の流れ



*LJ potential*  
*SW potential*  
*Gay-Berne model*  
*etc.*

モデル選択と多目的最適化を  
専門家が独自に行なっている  
ようなもので、パラメータ発見も非常に高難度 (逆問題)

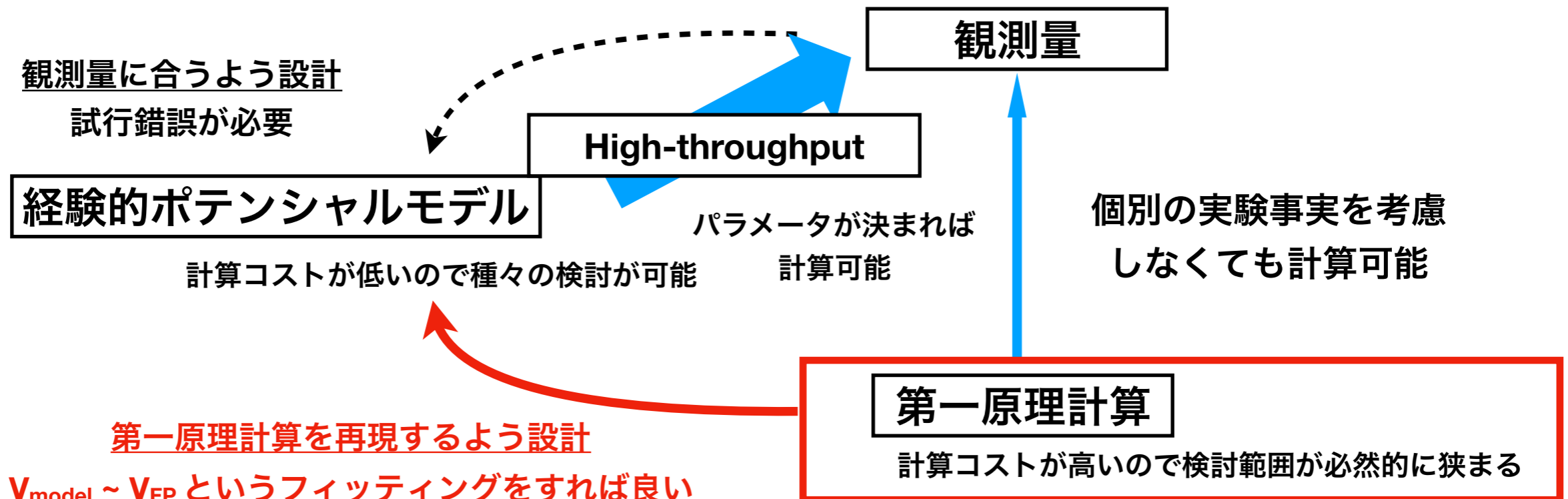
# 逆問題からフィッティングへ

## 第一原理計算

A **first principle** is a basic, foundational, self-evident **proposition** or assumption that cannot be deduced from any other proposition or assumption. (@Wikipedia)

- 基礎方程式と基礎物理量に基づき実験事実に合わせて込んだ経験的パラメータを用いない手法
- 実験結果を仮定せず扱いたい系を支配する物理モデルのうち最も基本的なもののみで物理量を計算する手法
- 密度汎関数法 (DFT) がもっとも主流 (many-body Schrödinger eq.を有効一体問題にマップして解く)
- 固体振動のバネモデルはバネ定数を実験値から決めなければいけないので第一原理的ではない
- DFTからバネ定数を求めて振動モードを計算する方法は第一原理計算
- ハートリー・フォック法, QCDなども第一原理計算

## ポテンシャル設計の流れ



# フィッティングによるポテンシャル作成

## TTAM (Tsuneyuki-Tsukada-Aoki-Matsuda) Potential

Hartree-Fock法で $\text{SiO}_4^{4-} + 4 \text{ point charge}$ モデルを元に、ポテンシャルを作成

Phys. Rev. Lett. **61**, 86 (1988).  
Nature **339**, 209 (1989).

$$= U_{ij}^{\text{Coulomb}}(r) + f_0(b_i + b_j) \exp[(a_i + a_j - r)/(b_i + b_j)] - c_i c_j / r^6,$$

$$U_{ij}^{\text{Coulomb}} = \tilde{Q}_i \tilde{Q}_j [1 - g_{ij}(r)] / r + Q_i Q_j g_{ij}(r) / r,$$

$$g_{ij}(r) = (1 + \zeta r) \exp(-2\zeta r), \quad g_{00}(r) = [1 + 11(\zeta r)/8 + 3(\zeta r)^2/4 + (\zeta r)^3/6] \exp(-2\zeta r).$$

- Si-O間の電荷移動量をモデル化
- 静電相互作用をlong-range, short-rangeに分離
- クラスタモデルから固体系へ移植
- $\text{SiO}_2$ の様々なpolymorphの安定構造を再現
- 6配位, 4配位の差に軌道を仮定しなくてもよかった

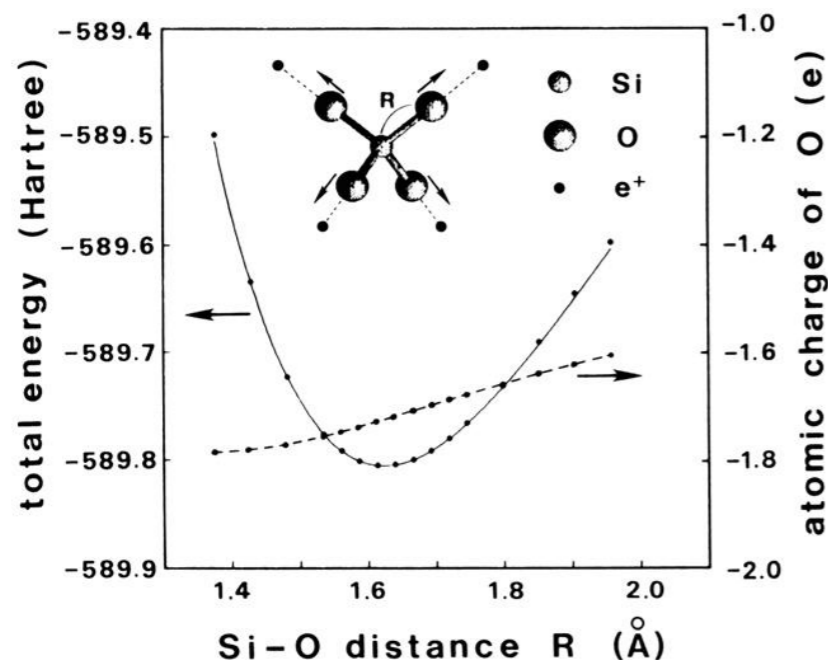


FIG. 1. Total energy and the Mulliken charge on an oxygen atom for  $T_d$  deformation of a  $\text{SiO}_4^{4-} - 4e^+$  cluster shown in the inset. The solid circles are the cluster calculation, full curve is the fitted potential, and the broken curve is a guide to the eye.

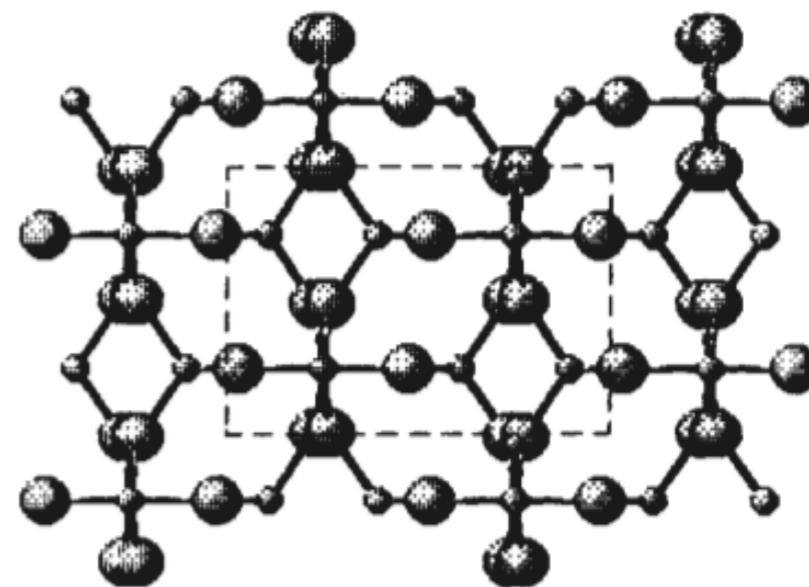


FIG. 2 The new  $Cmcm$  phase predicted by this study, involving both four- and sixfold coordination, as seen from the  $a$  axis, which corresponds to the  $c$  axis of low-cristobalite and stishovite. Small spheres represent silicon, and large spheres represent oxygen. The unit cell is shown by dashed lines.

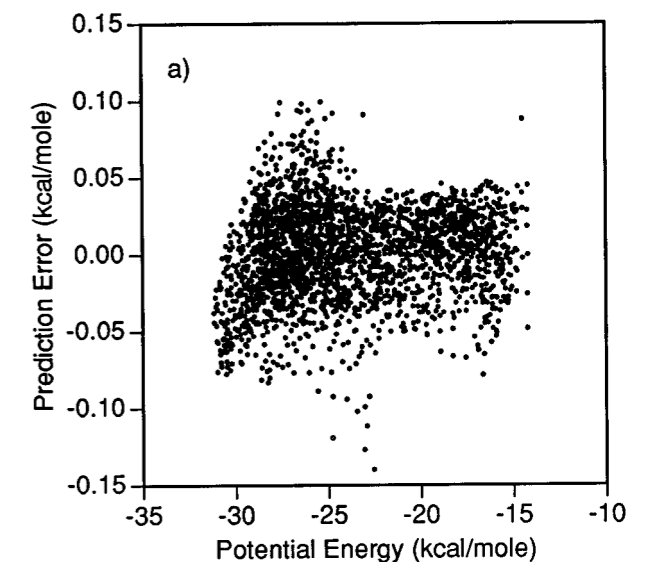
高圧下での $\text{SiO}_2$ 新構造を発見！

# ニューラルネットポテンシャル

(おそらく) 初めてのNNによるフィッティング論文

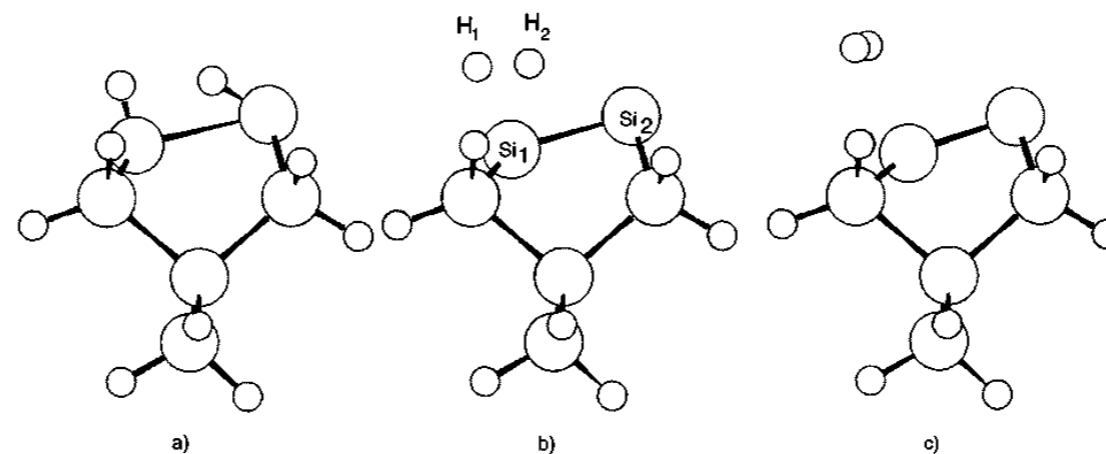
## Classical modelの入力を低次元化する目的

- CO/Ni(111)のCO吸着に関する研究
- COの3次元配置で与えられるポテンシャルを3つの特徴量でモデル化
- 7 x 7 x 11 samples, 3-15-1という構成
- Extend Kalman filterを使って最適化
- 低次元NNで、path-integralを計算し、量子効果を見積もっている

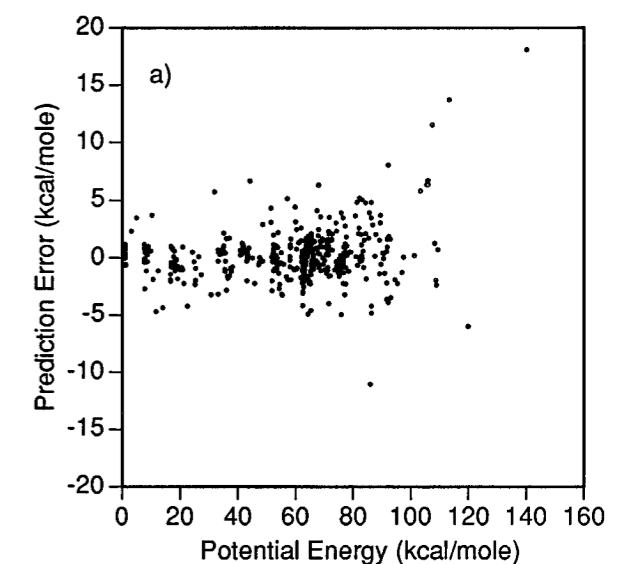


## DFTの結果をモデル化することも試みられている

- Si(100)上のH<sub>2</sub>発生のクラスターモデル計算
- 750 sample, 12-7-1という構成
- 誤差2.1 kcal/mol.とやや大きい



S. Pai and D. Doren, J. Chem. Phys. **103**, 1232 (1995).



**ただし、そもそものシミュレーション手法がそこまで進んでいない。。**



# NNフィッティングの諸問題

## Symmetry-Adapted Functionの設計

fcc(111)の表面形状に合わせてフーリエ変換を元に構成する方法を提案[1]

[1] J. Behler, S. Lorenz, and K. Reuter, J. Chem. Phys. **127**, 014705(2007)

ただし問題に合わせて複雑な関数を設計することは決して容易ではなく  
もっと簡便な表現が必要に。

## 3次元配置を入力にしたNNの問題

- 入力層が原子数に依存するため、異なる系には適用できない（汎用性）
- 同種粒子の入れ替えに対する対称性が破れてしまう
- 同様にポテンシャルが満たすべき対称性が保証されない

**入力表現・汎用性・対称性がキーワード**

# Behler-Parrinelloの方法

## 1. 入力として新しいSymmetry Function (対称性関数) を導入

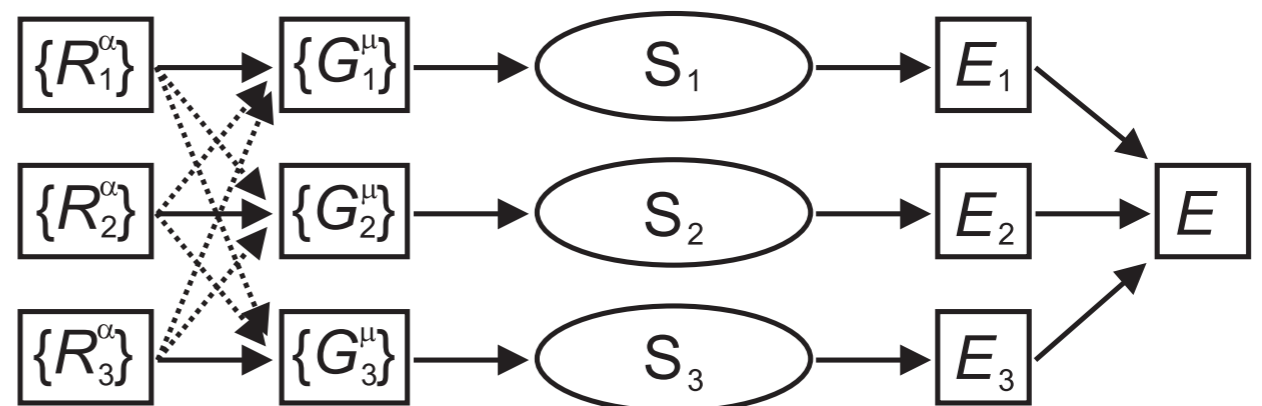
$$G_i^1 = \sum_{j \neq i}^{\text{all}} e^{-\eta(R_{ij} - R_s)^2} f_c(R_{ij}), \quad (2)$$

$$G_i^2 = 2^{1-\zeta} \sum_{j,k \neq i}^{\text{all}} (1 + \lambda \cos \theta_{ijk})^\zeta e^{-\eta(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)} f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk}), \quad (3)$$

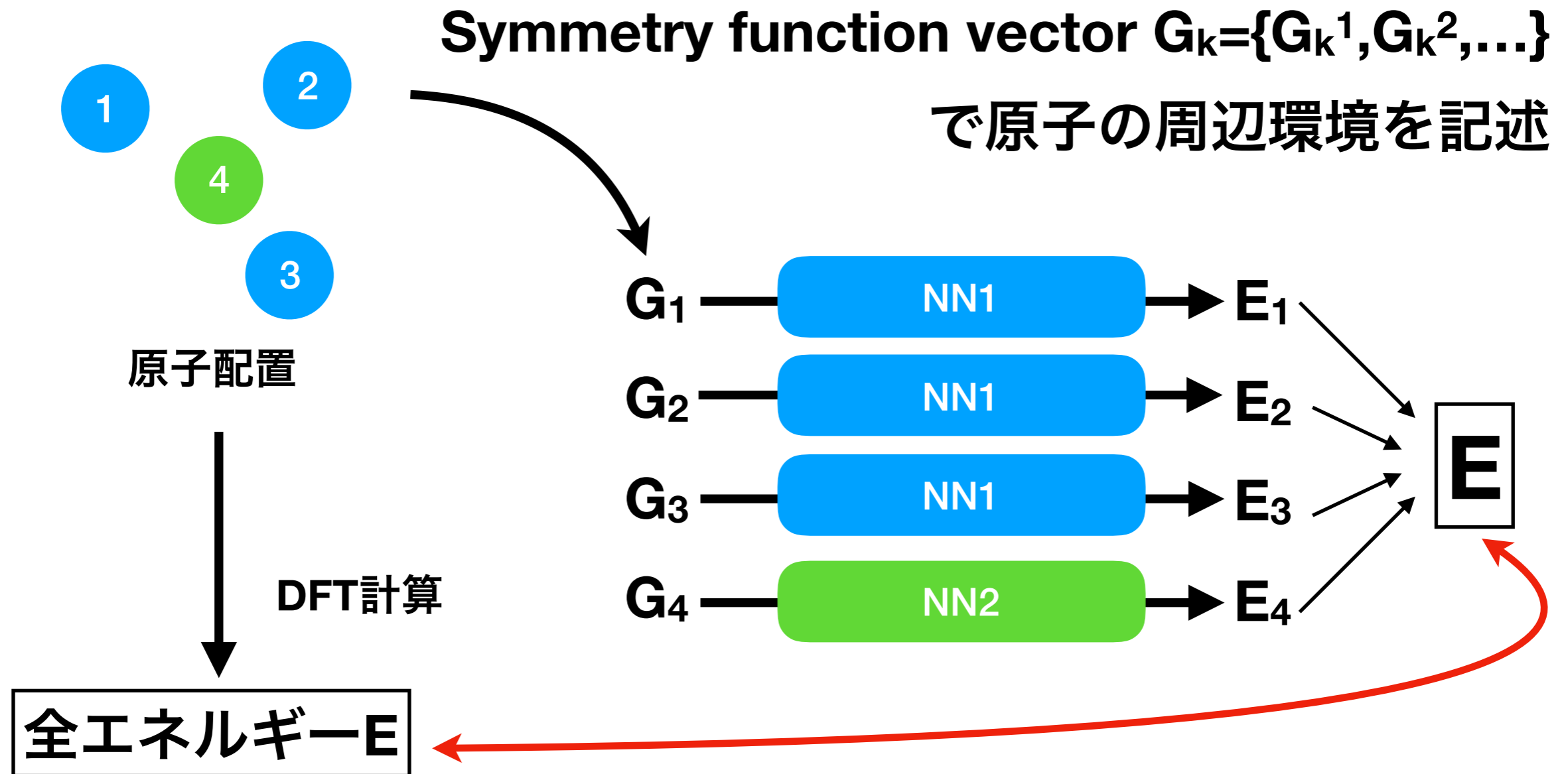
- 各原子周辺環境を2体・3体の指標 (Symmetry function) で数値化
- これらの指標は系の回転・並進で普遍

## 2. 全エネルギーを原子ごとの寄与に分割、原子種ごとに同じNNを構成する

- 同種原子の入れ替え対称性を担保
- 原子数が増えてもシステムサイズで入力層が変わらず学習済みNNのサブネットを増やすだけで良いため

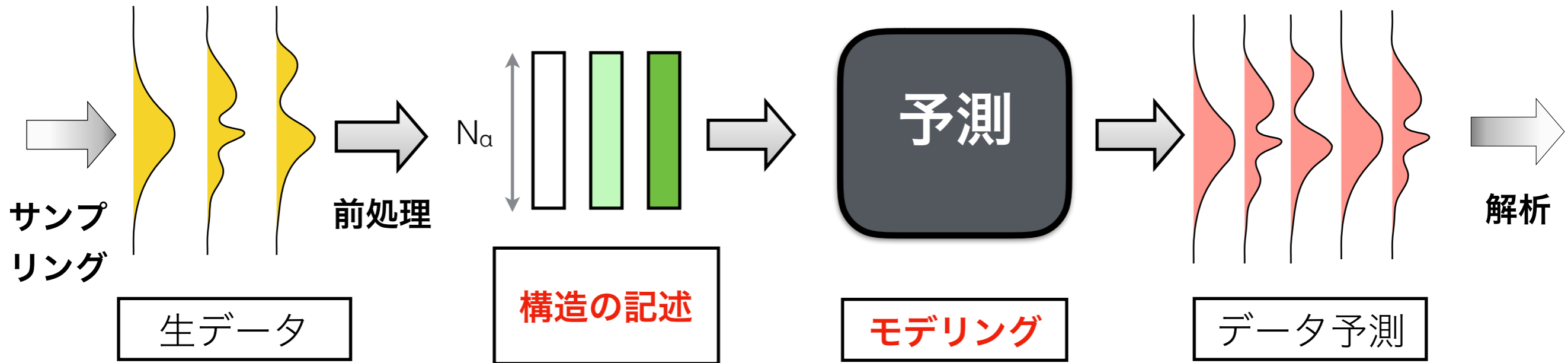


# 汎用NNP構成の流れ



- 正味のINPUTは、サブネットの入力 x 原子数
- 学習すべきNNのパラメータは、原子種ごとに共通。
- また入力ベクトルGの次元は固定できるので原子数を増やすことは容易

# 機械学習ポテンシャル作成の流れ



## DFT計算結果

- (説明変数) 原子の配置
- (目的変数) エネルギー  
(またはforce)

## 説明変数の低次元化

- Symmetry Function
- Structural fingerprint
- SOAP
- Coulomb Matrix
- MBTR

## 回帰モデル

- Neural network
- 多項式展開
- ガウス過程回帰
- Kernel ridge回帰

## High-Throuput MD

- 安定構造探索
- 材料スクリーニング
- MDからの拡散係数計算
- 網羅的NEB計算

機械学習ポテンシャルは (1) 構造の記述 (2) モデリングの違いで大別  
常に使用するメリット・目的を常に念頭におくこと



# SOAP記述子とGAP

## GAP: Gaussian Approximation Potential

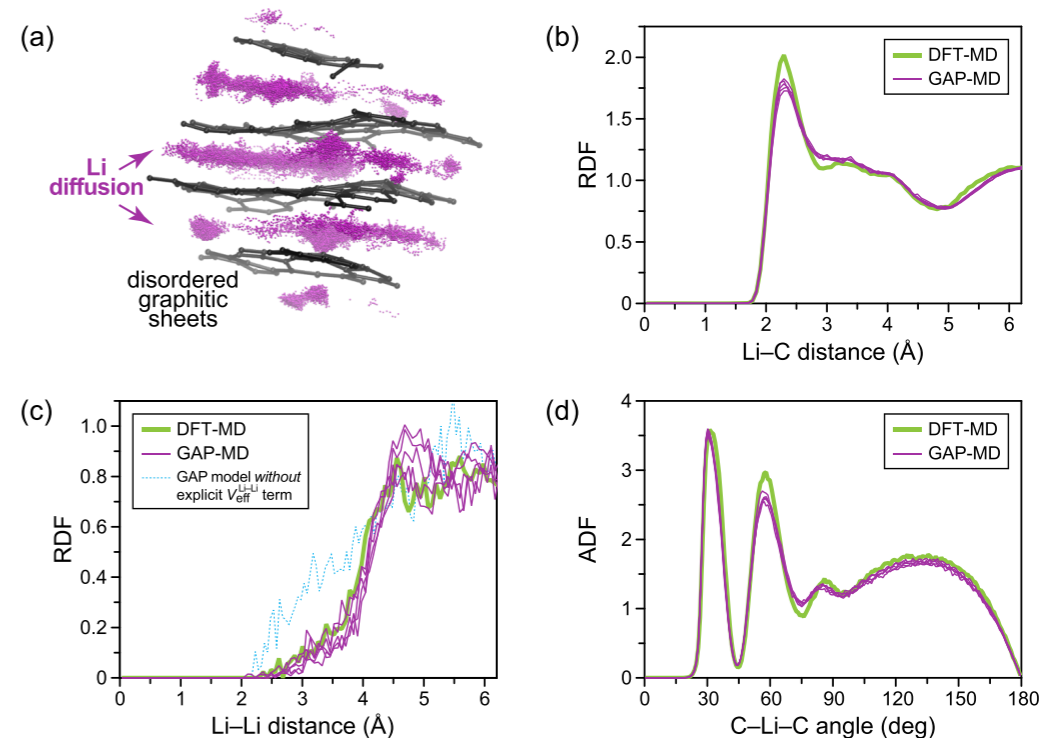
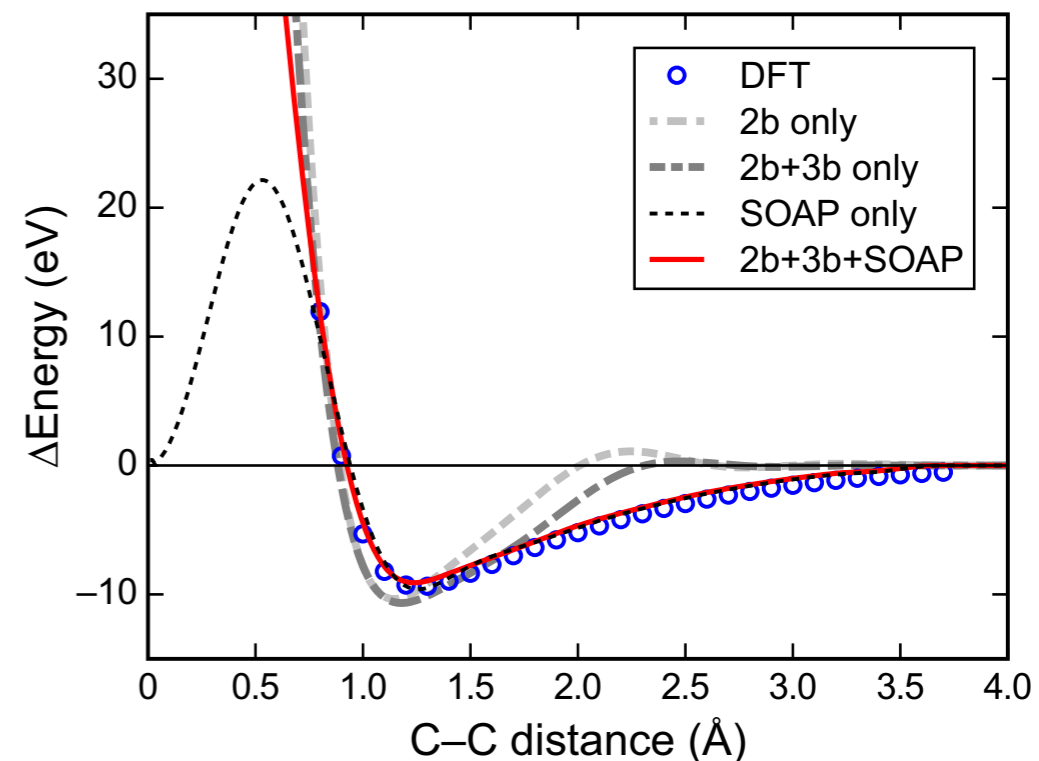
- ✓ **ガウス過程回帰**で相互作用を記述
- ✓ Cの相互作用を2-, 3-, many-bodyと展開
- ✓ 2- and 3-body: Gaussian kernelを利用
- ✓ Many-body: Simple dot product kernel

### Descriptor

2-body:  $q^{(2b)} = |\mathbf{r}_2 - \mathbf{r}_1| \equiv r_{12}$

3-body:  $\mathbf{q}^{(3b)} = \begin{pmatrix} r_{12} + r_{13} \\ (r_{12} - r_{13})^2 \\ r_{23} \end{pmatrix}$

Many-body: **SOAP**  
(Smooth overlap of atomic position)



応用：グラファイト内部のLiイオン拡散シミュレーション

Fujikake et al., J. Chem. Phys. 148, 241714 (2018).

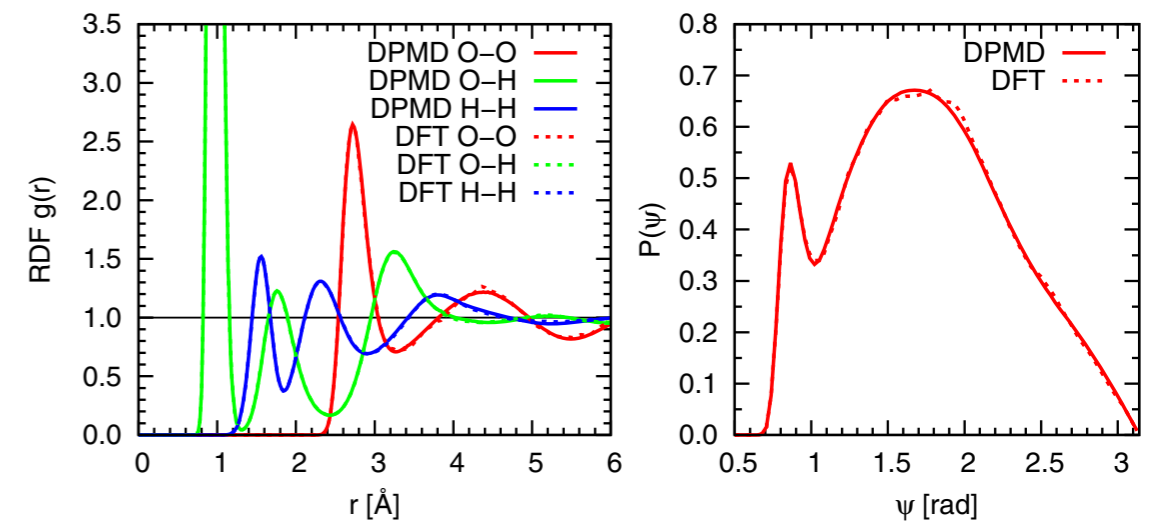
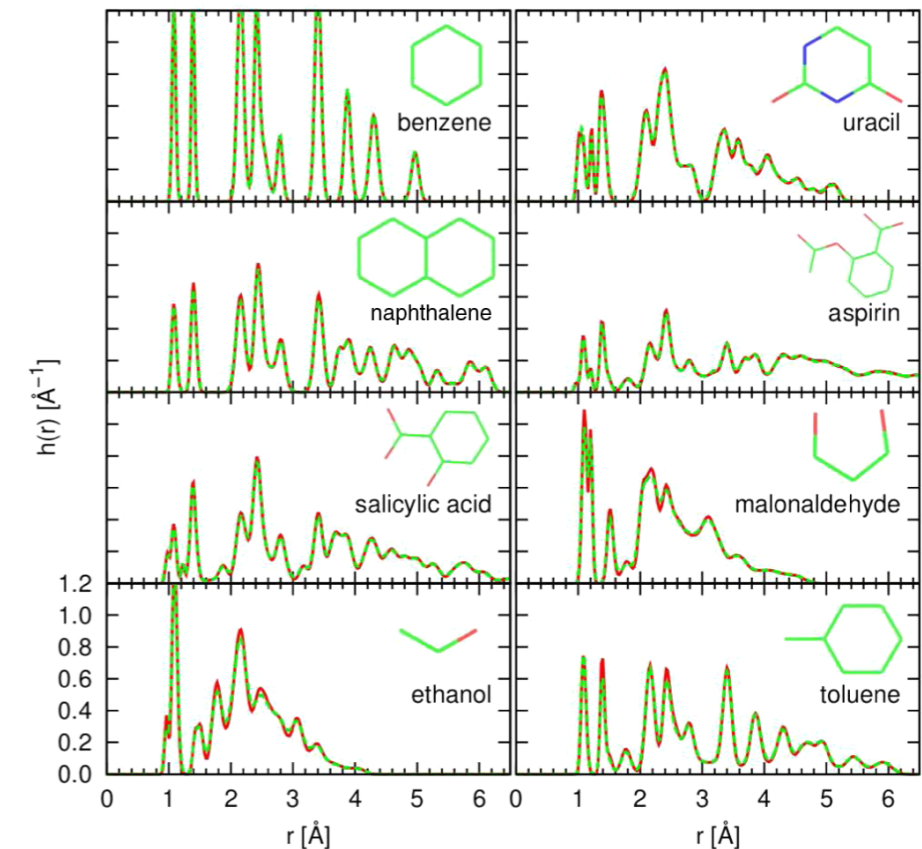
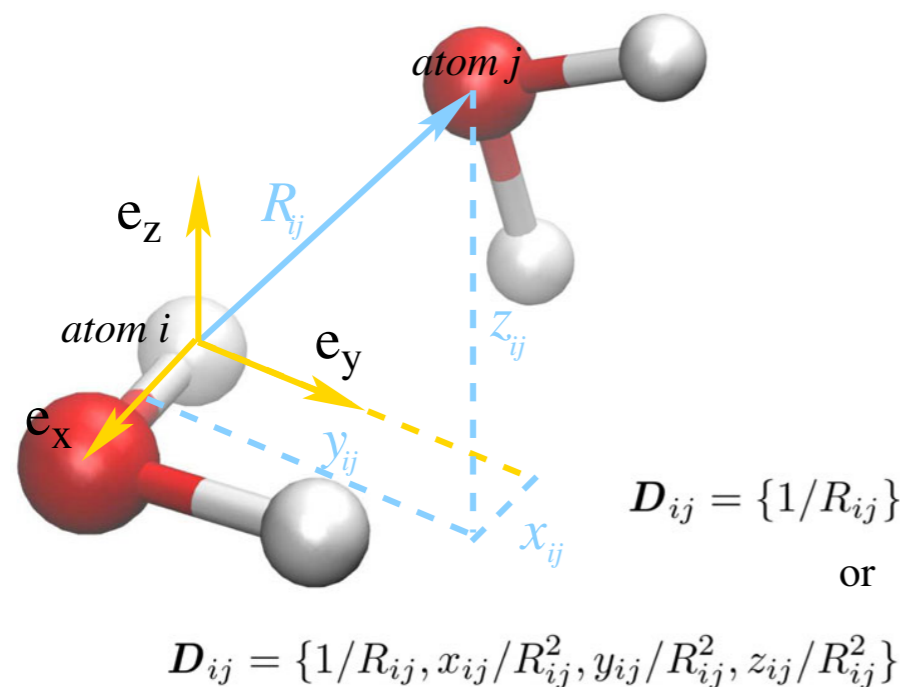
# DeePMD

## [モデリング]

- Behler-ParrinelloのNNフレームワーク

## [記述子]

- 回転対称性を担保するような内部座標
- 距離の逆数で局所環境を表現
- 座標ベースでのNNPを実現



**分子系でも高精度のポテンシャルを実現**

DeePMD-kitが公開されている

# アモルファス材料内部の拡散経路探索

**大問題：高精度に計算できるDFTではコストがかかりすぎる**

( $N^3$ 回の構造最適化 x 計算時間 T)  $N \sim 50$ ,  $T \sim 1$  hour以上 = 13 year

さらに拡散障壁の計算も必要...

$$E_{\text{tot}} = E_{\text{amorph}} + \Delta E_{\text{Cu}} + \Delta E_{\text{opt}}$$

↓

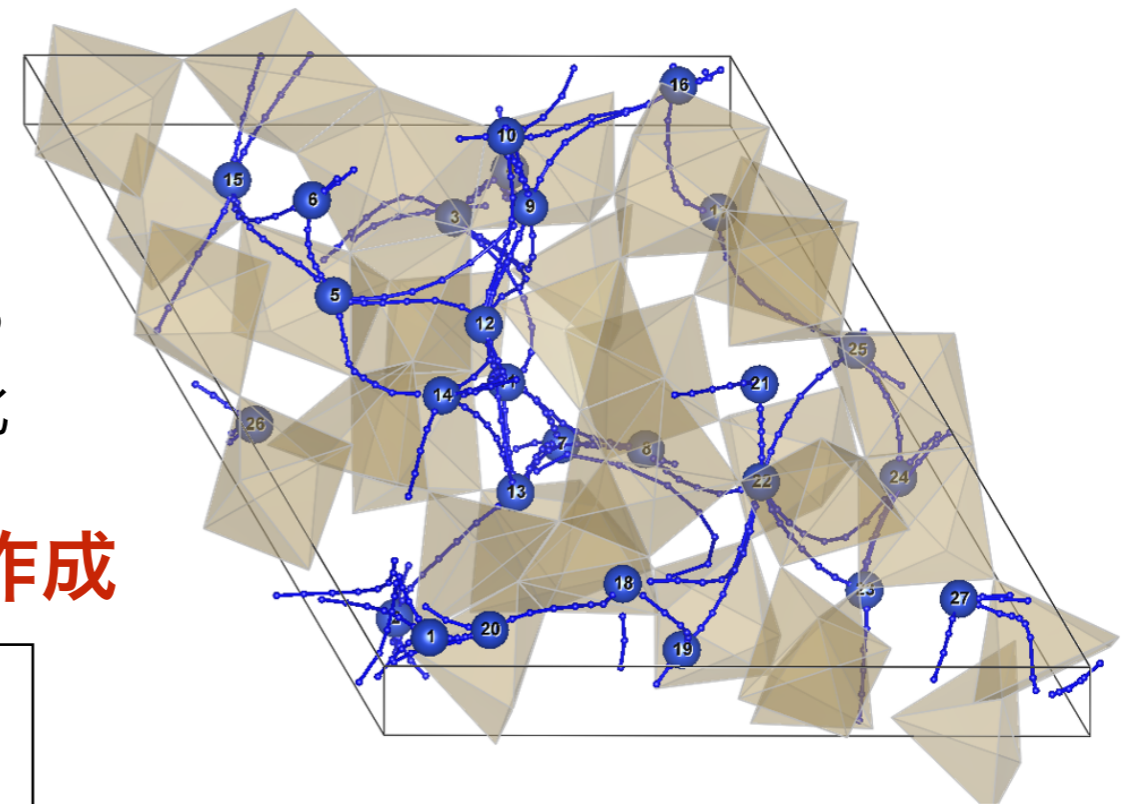
Cu挿入による  
エネルギー変化

↓

構造緩和による  
エネルギー変化

## Behler-Parrinelloの方法でポテンシャル作成

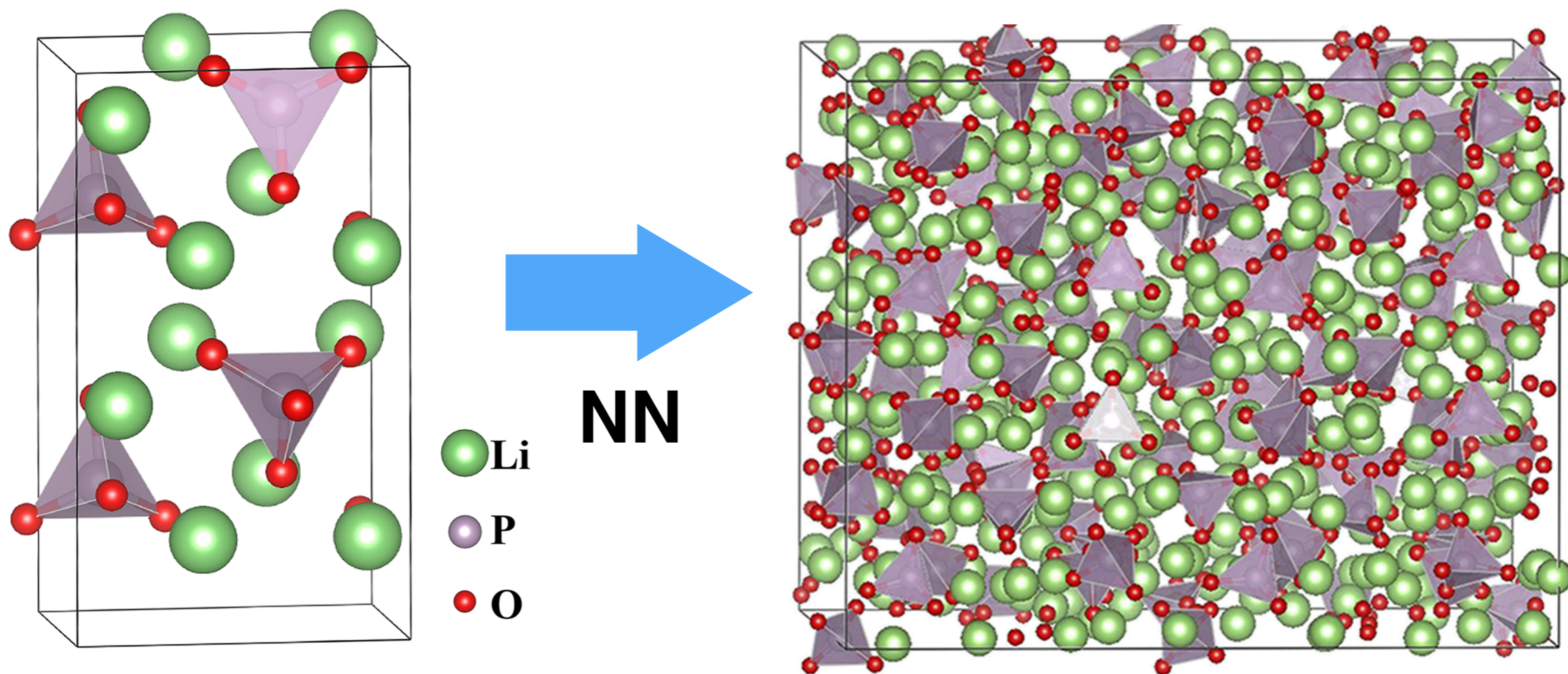
- ✓ アモルファス母体を「場」として取り扱う
- ✓ 学習対象は拡散粒子 (Cu) のみ
- ✓ 3元系でありながら小規模なネットワークで記述可能
- ✓ 事実上1粒子の計算なので、エネルギー評価も容易
- ✓ アモルファスは動かさないなのでMDはできない



拡散ネットワークの全体図



# 原子拡散計算の大規模化



DFT for small cell

Simulation for large cell

小規模・高精度計算の  
結果を学習

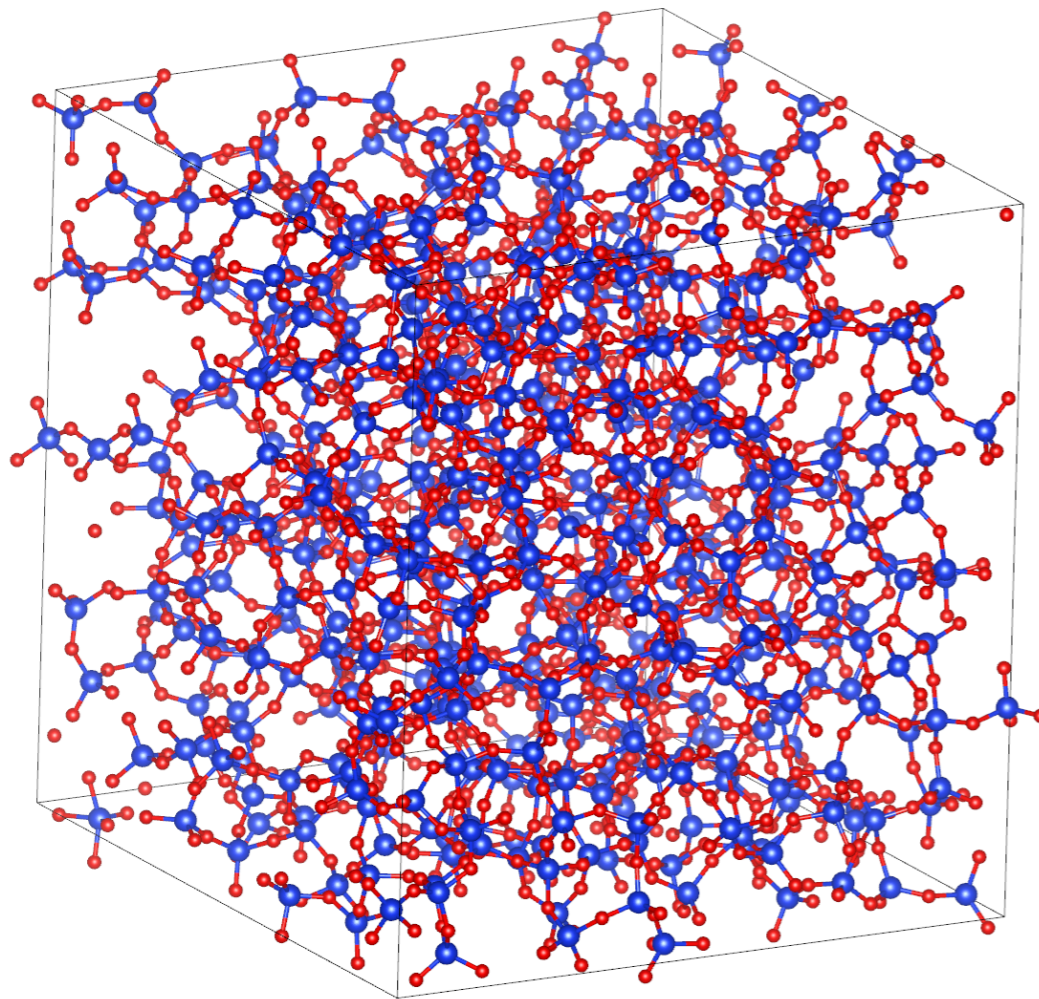
学習結果を使って大規模化

**DFTと同様の同型分布・拡散係数の実験値を再現**



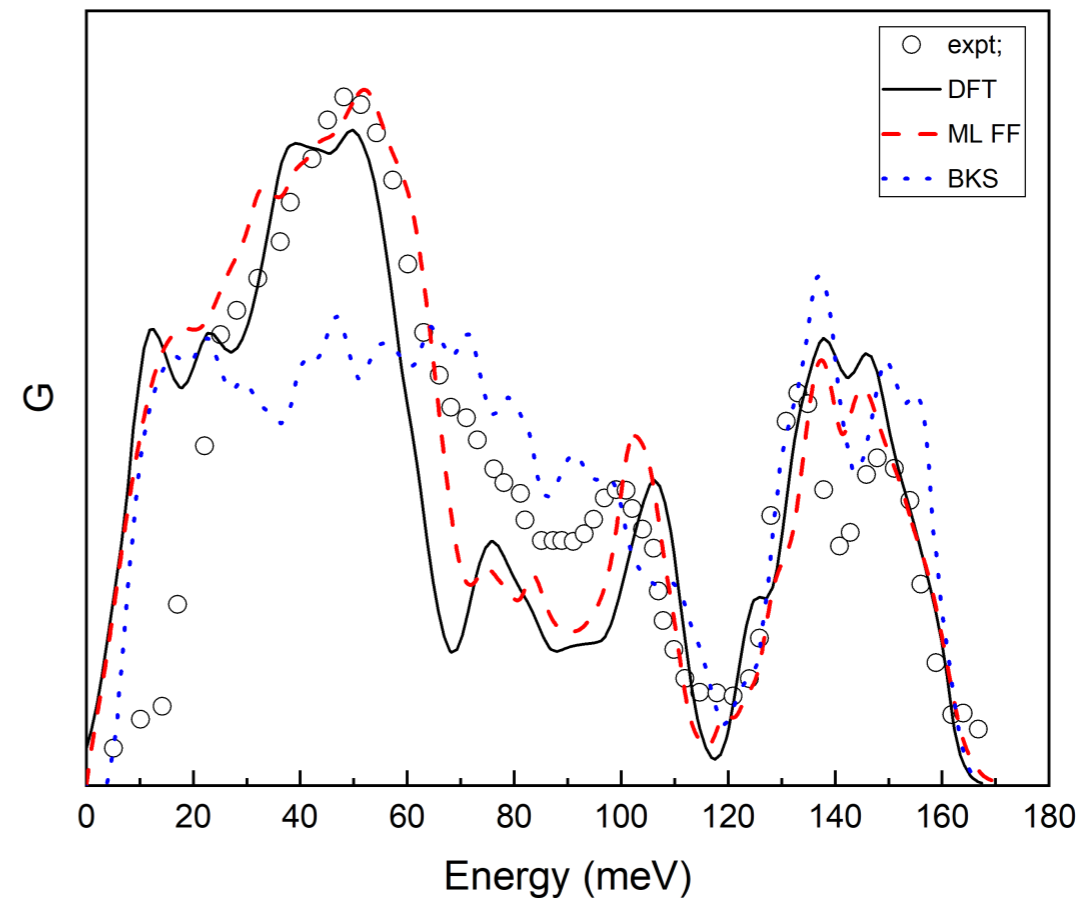
# アモルファスの振動物性評価と構造同定

## Amorphous model



1728 atoms

## Phonon density of state



The phonon density of state (DOS) calculated from ML force fields.  
[Science, 1997, 275, 1925–1927; Europhys. Lett., 2002, 60, 269–275; Phys. Rev. Lett., 1985, 54, 441–443]

**MLFFを使うことで大規模系をゆっくりアニール出来る (0.01K/fs)**

**Si-Oネットワークの環構造をDFTより改善でき、フォノン状態密度も良い結果が得られた**

# 機械学習ポテンシャルの課題

- 多成分系（4元系以上）に弱い
- 第一原理計算の結果がないと何もできない
- 構造データが十分かどうか、の検証方法が少ない
- 学習モデルの軽量・高速化, 入力の低次元化
- 利用できるプログラムがまだ少ない
- ポテンシャルの妥当性をきちんと評価する必要あり

# まとめ

- 原子分子シミュレーションは原子間相互作用を決めることが根幹にある
- 原子間相互作用の標識は技術革新に伴い時代とともに変化してきたが実験結果に合うようなパラメータを逆問題的に決定することが必須であった
- 第一原理計算の登場により、実験結果ではなくシミュレーション結果をフィットすることでポテンシャルが構成できるようになった
- Behler-Parrinelloの方法により汎用的なNNが構成可能となった
- 現在ではGAP, DeePMDなどいくつかの手法が提案されている
- アモルファス系に関するこれまでの研究事例：拡散経路, 小規模から大規模へ繋ぐ, アモルファスの振動特性と構造作成