



データサイエンスの計算物質科学への応用

-トポロジカルデータ解析・
パーシステントホモロジーを中心に-

大阪大学産業科学研究所
南谷英美

今日の講演内容

前半

パーシステントホモロジーとは何？

ホモロジー群の考え方

パーシステントホモロジー入門

理論

後半

パーシステントホモロジーの応用

具体的な応用例の紹介@材料科学分野

実習

応用

自己紹介

2001～2011

大阪大学工学研究科 応用自然科学科 応用物理学専攻
(学部～博士課程＋学振PD @笠井研究室)

2011～2013

理化学研究所
(基礎科学特別研究員 @Kim表面界面科学研究室)

2013～2019

東京大学工学系研究科 マテリアル工学専攻
(助教・講師 @渡邊研究室)

2019～2022

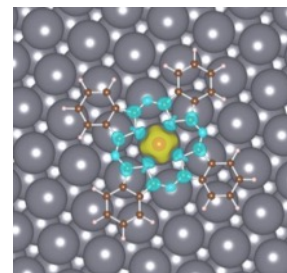
分子科学研究所
(准教授)

2022～

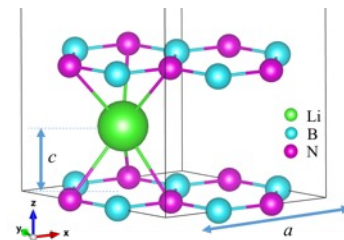
大阪大学産業科学研究所
(教授)

これまでの研究内容

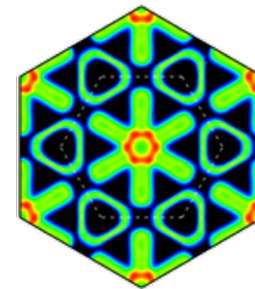
磁性原子・分子における
量子多体効果



新奇層状物質の物性



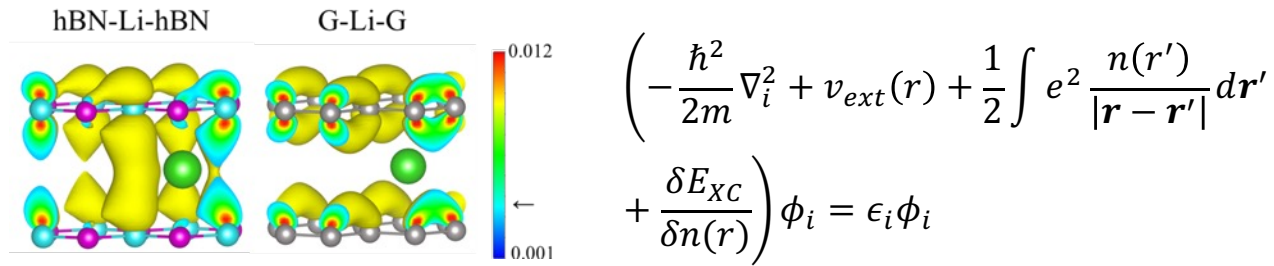
電子フォノン相互作用
の精密計算、
熱物性の第一原理計算



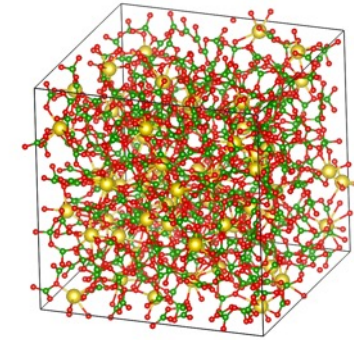
専門領域：物性物理・計算物質科学

ナノスケールでのシミュレーション

密度汎関数理論(Density functional theory) 第一原理計算



分子動力学法(Molecular dynamics)



直接的なシミュレーション

原子構造 → (電子状態) → エネルギー → 力 → 物性

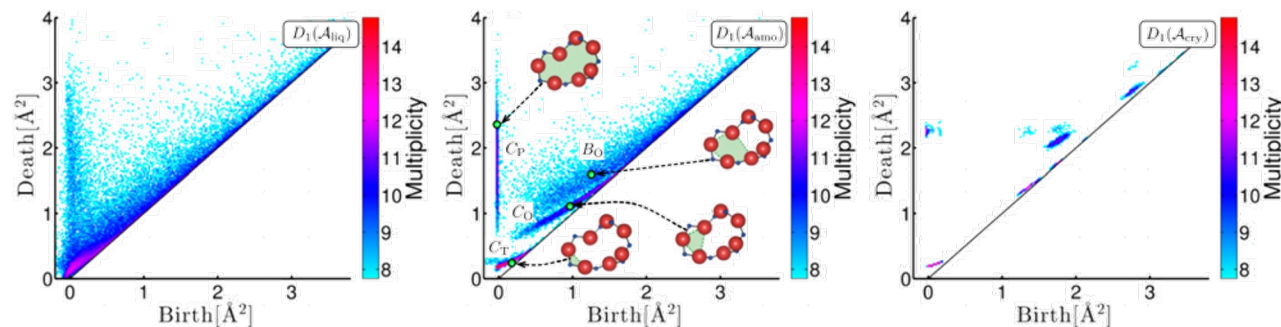
構造特徴の抽出 + データ

原子構造 → 物性

有効モデル → 高効率な計算・構造物性相関の理解

構造解析の新しい切り口：トポロジカルデータ解析

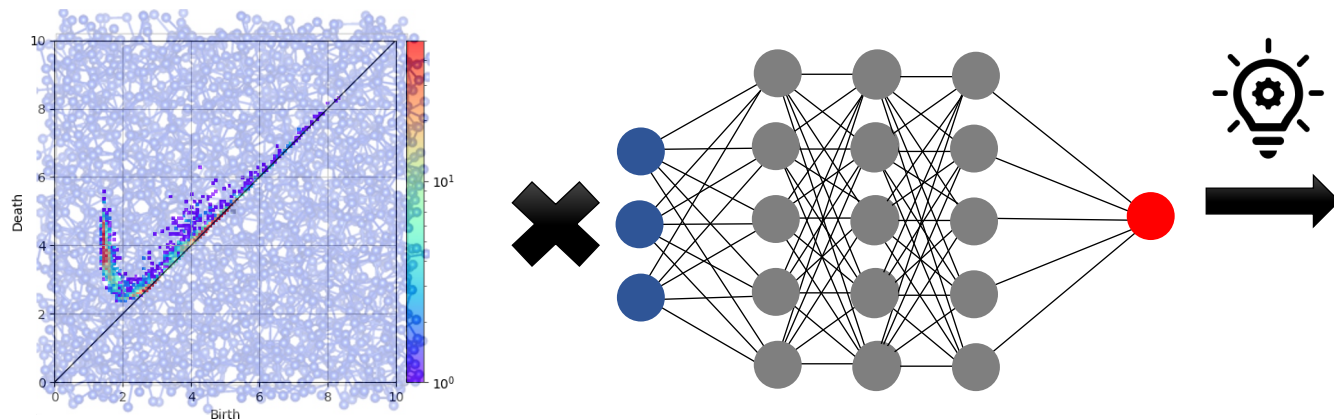
代表的テクニック：パーシステントホモロジー群



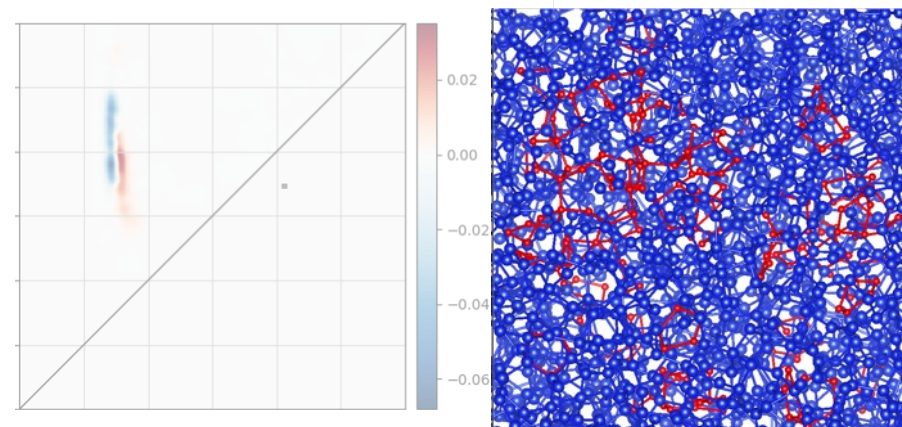
液体ガラスとアモルファス SiO_2 の判別
(Y. Hiraoka et al., PNAS, 113, 7035 (2016))

構造解析だけでなく物性との相関を議論できる？

PH • Simulations • Machine learning



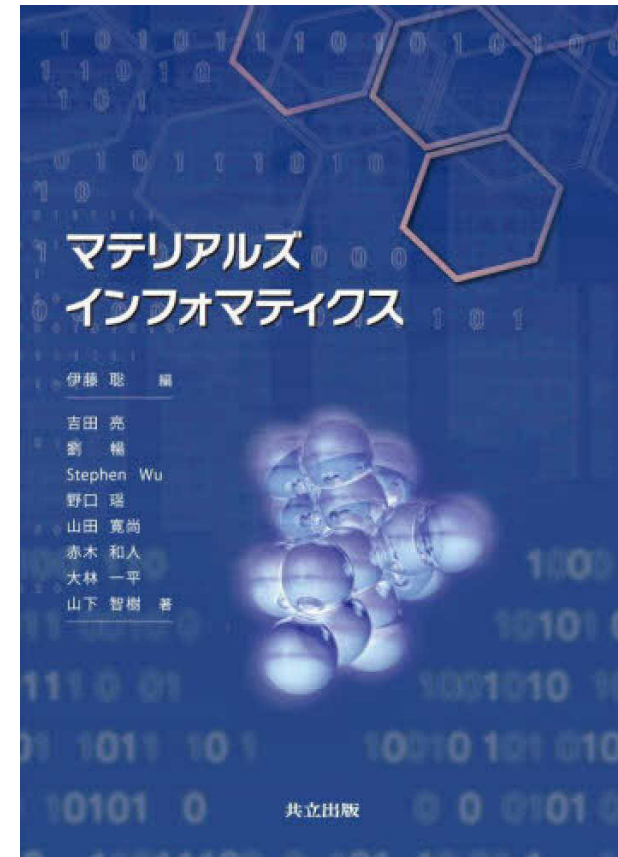
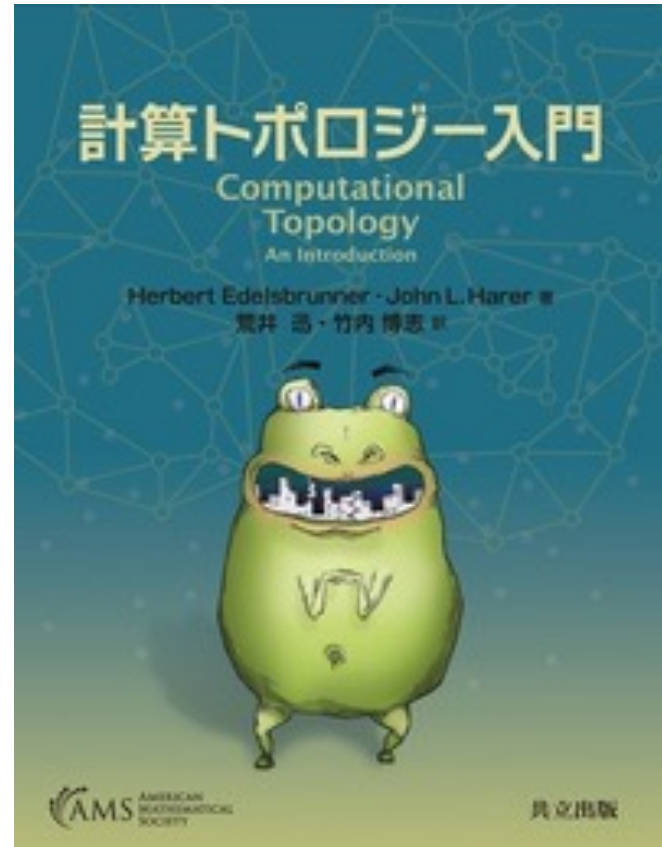
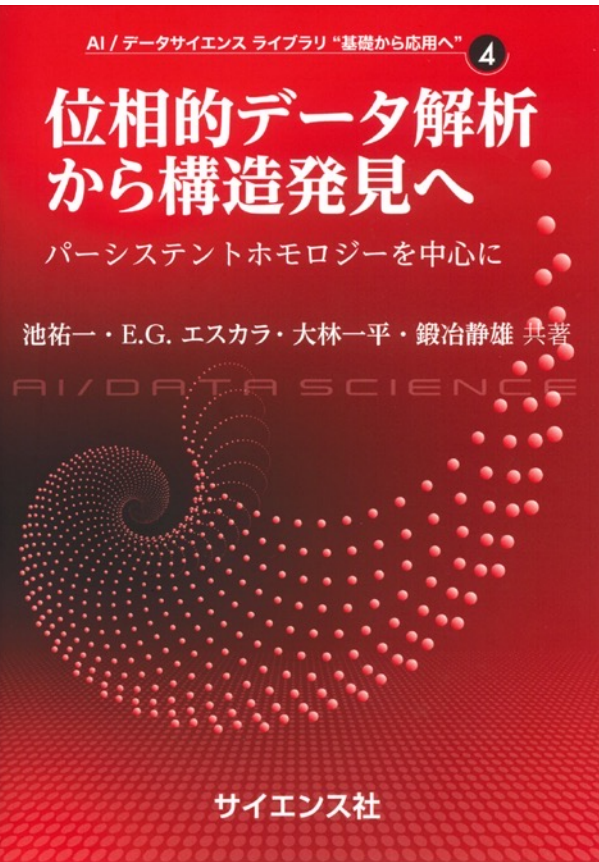
Prediction & structure-properties relationship





パーシステントホモロジー群とは？

おすすめ参考図書



今回の講義担当の南谷は数学の研究者ではないので、
数学的な定義とか言葉の使い方が超適当です

きちんとした説明はこれらの本で確認してください



パーシステントホモロジー群？

そもそもホモロジー群とは？

目的：図形をグループ分けして、この図形はあの図形と同じかどうかを議論したい

群論

トポロジー

合同・相似以外の同じさの
定義を与える

群：特定の規則が成り立つ集合

シンプルな例：整数の集合 $\mathbb{Z}$ 、ルールは加法（+）

- ①演算結果も集合の元である：任意の整数を足し合わせた結果は整数
- ②演算しても結果が変わらない元（単位元）がある：0
- ③演算すると結果が単位元になる元がある（逆元）：ある整数Nに対する-N

群になる集合の部分集合もまた群の性質を保つ場合→グループ分けに使える

部分群とグループ分けの例

群の一部の要素を集めてきた集合がまた群になっているとき→部分群

部分群を使うと、グループ分けが定まる

とても簡単な例

3の倍数全体の集合 $3\mathbb{Z}$

足し算（加法）を考えると

- ① $3N$ と $3M$ を足した結果も3の倍数
- ② 0 が単位元
- ③ $3N$ に対して $-3N$ が逆元

} $\mathbb{Z}$ の部分群

これがグループ分け?? →元々の群の要素と部分群の足し算を考えてみる

部分群とグループ分けの例

$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$$

$$3\mathbb{Z} = \{\dots, -6, -3, 0, 3, 6, \dots\}$$

例えば

$$1+3\mathbb{Z} = 1 + \{\dots, -6, -3, 0, 3, 6, \dots\}$$

$$5+3\mathbb{Z} = 5 + \{\dots, -6, -3, 0, 3, 6, \dots\} = 2 + 3\mathbb{Z}$$

$$4+3\mathbb{Z} = 4 + \{\dots, -6, -3, 0, 3, 6, \dots\} = 1 + 3\mathbb{Z}$$

$$9+3\mathbb{Z} = 9 + \{\dots, -6, -3, 0, 3, 6, \dots\} = 0 + 3\mathbb{Z}$$

$\mathbb{Z}$ の要素について足し算を考えた結果

$$0 + 3\mathbb{Z}$$

$$1 + 3\mathbb{Z}$$

$$2 + 3\mathbb{Z}$$

グループ分け：剰余類

剰余類についての足し算に特定のルールを定めると、群を作ることができる

部分群とグループ分けの例

元々の群	部分群	元々の群の要素に部分群の要素を作用した結果→剰余類
G 要素 a, b, c	H	$cH = \{ch c \in G, h \in H\}$

剰余類の間の足し算をこう定めてやると、剰余類が群をなす

$$aH + bH = (a + b)H$$

$$(0 + 3\mathbb{Z}) + (2 + 3\mathbb{Z}) = 2 + 3\mathbb{Z}$$

$$(1 + 3\mathbb{Z}) + (2 + 3\mathbb{Z}) = 0 + 3\mathbb{Z}$$

$$(2 + 3\mathbb{Z}) + (2 + 3\mathbb{Z}) = 1 + 3\mathbb{Z}$$

群の要素と要素の足し算の結果は群の要素
逆元も自然と定義される
 $0 + 3\mathbb{Z}$ つまり $3\mathbb{Z}$ が単位元

奇妙な足し算は、部分群の要素は単位元と考えていることに相当

商群 (or剰余類群) G/H

図形を集合に変換するには？

群論によるグループ分け：集合+演算（群） & グループ分け用の部分群を決めるルール

図形をどうやって集合として表現するか

構成要素への分解：単体

0-単体

(点)

p_0


(p_0)

1-単体

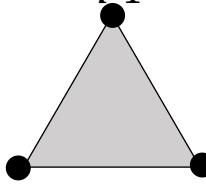
(辺)

p_0  p_1

$(p_0 p_1)$

2-単体

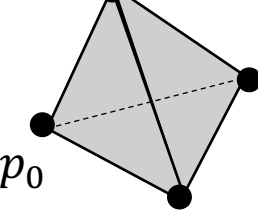
(三角形)

p_1

 p_0 p_2

$(p_0 p_1 p_2)$

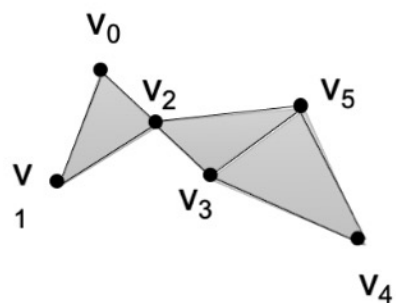
3-単体

(四面体)

p_3

 p_0 p_1 p_2

$(p_0 p_1 p_2 p_3)$

単体的複体



$\{v_0, v_1, v_2, v_3, v_4, v_5,$
 $(v_0v_1), (v_0v_2), (v_1v_2), (v_2v_3), (v_2v_5), (v_3v_5), (v_3v_4), (v_5v_4)$
 $(v_0v_1v_2), (v_2v_3v_5), (v_3v_4v_5)\}$

0単体の集まり

1単体の集まり

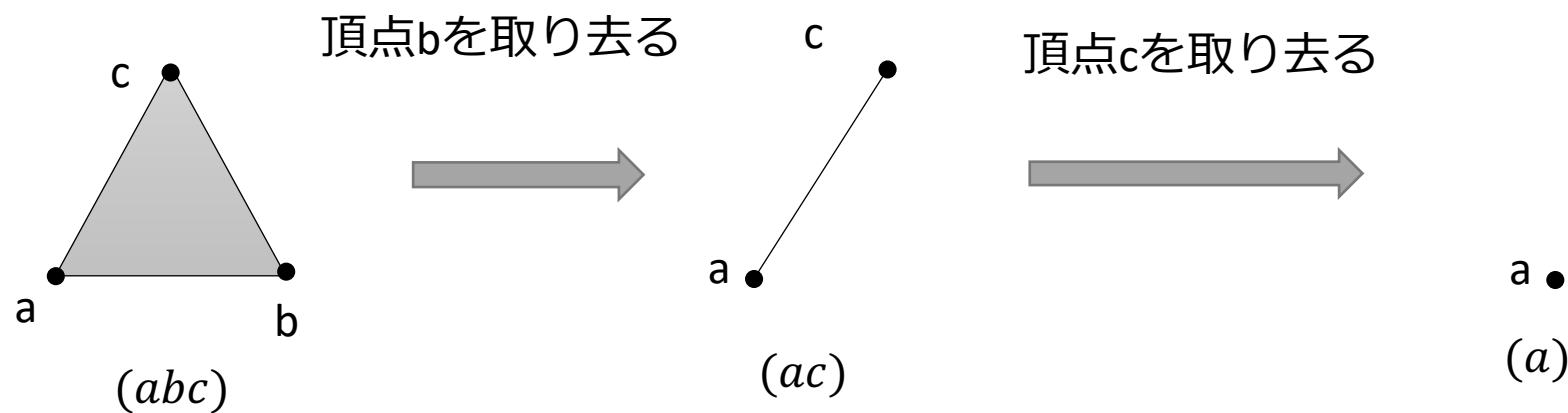
2単体の集まり

様々な図形は各次元の単体の集まりとして表現できる：単体的複体

単体的複体と定義できるためにはいくつかの条件が必要

単体的複体の条件

頂点を取り去ると次元の低い単体ができる→面

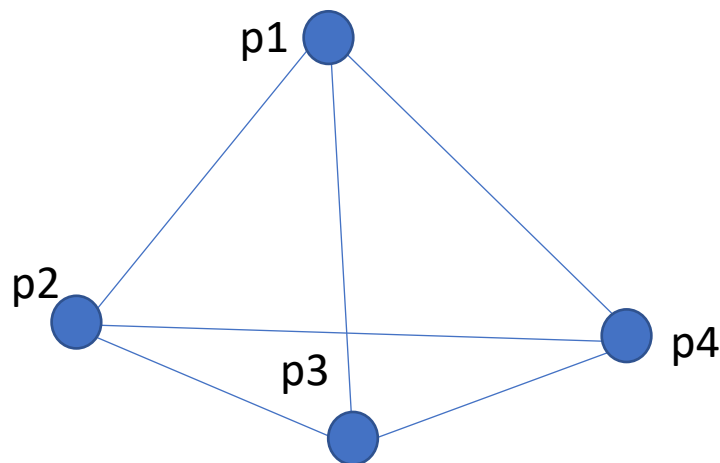


単体的複体である条件

すべての単体の面が要素として含まれていなくてはならない
2つの単体の共通部分は、それらの単体の面として含まれていなくてはならない

図形を群として扱うには？

図形を単体的複体にする & そこに群をいれるには？



$\{p_1, p_2, p_3, p_4$
 $(p_1p_2), (p_1p_3), (p_1p_4), (p_2p_3), (p_3p_4), (p_2p_4)$
 $(p_1p_2p_3), (p_2p_3p_4), (p_1p_2p_4), (p_2p_3p_4),$
 $(p_1p_2p_3p_4)\}$

0単体の集まり
1単体の集まり
2単体の集まり
3単体の集まり

※以下、こういった単体的複体のことを K で表します

それぞれの次元において、係数を0か1にした単体の足し合わせを考える

$$p_1 + p_2 + p_4$$

$$(p_1p_2) + (p_2p_3) + (p_2p_4) \quad \text{など}$$

$$(p_1p_2p_3) + (p_2p_3p_4)$$

あらゆる可能な係数の組み合わせについて
各次元で集めてきたもの
→群をなす

$$C_r(K) = \left\{ \sum_i c_i \sigma_{ri} \mid c_i \in \mathbb{F}_2 \right\} \quad \text{チェイン群}$$

補足：係数による違い

$\mathbb{F}_2$

位数 2 の有限体、元は 0 と 1 の 2 つだけ

+	0	1
0	0	1
1	1	0

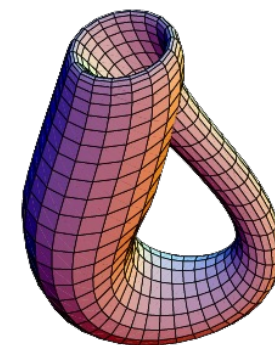
$\times$	0	1
0	0	0
1	0	1

掛け算について逆元が
定義できる $\rightarrow$ 体

数学の教科書ではチェイン群の係数で $\mathbb{Z}$ を使っている場合が多い

$\mathbb{Z}$ の係数を使う場合との違い

計算機での演算は $\mathbb{F}_2$ のほうが行いやすい
境界作用素の結果の定義が異なってくる（向き付けのある単体を使うかどうか）
クラインの壺のような形状の特徴は、 $\mathbb{Z}$ など環の係数でないと捉えられない場合がある



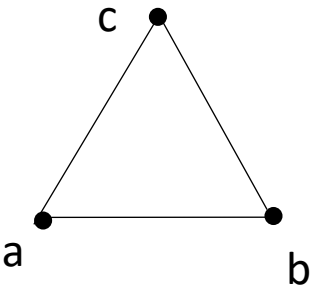
部分群を決めるには？

$r+1$ 個の頂点で決まる r 次元の単体

$$\sigma_r = (p_0 p_1 p_2 \dots p_r)$$

頂点を一つ抜き去って一つ次元が低い単体を作る写像：境界作用素

$$\partial_r \sigma_r = \sum_{i=0}^r (p_0 p_1 p_2 \dots \check{p}_i \dots p_r) \quad \check{p}_i: \text{この頂点を抜くことを表している}$$



$$\partial_1 \{(ca) + (ab) + (bc)\} = a + c + b + a + c + b = 0$$

($\mathbb{F}_2$ の演算では $1+1=0$)

境界作用素によって0になるものがある→**サイクル**

$$C_r(K) = \left\{ \sum_i c_i \sigma_{ri} \mid c_i \in \mathbb{F}_2 \right\}$$

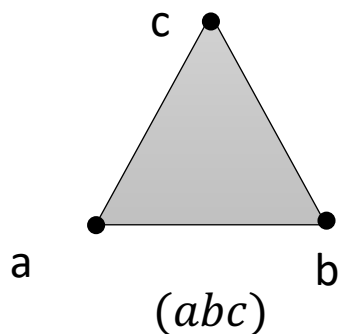
の要素から、境界作用素で0になる部分集合を作ることができる

部分群：**サイクル群** $Z_r(K, \mathbb{F}_2)$

部分群の入れ子構造

サイクル群の要素の中には、ひとつ上の次元の単体での境界作用素の結果になっているものがある

➡ 部分群：バウンダリ群 $B_r(K, \mathbb{F}_2)$



$$\partial_2(abc) = (bc) + (ac) + (ab) = (ca) + (ab) + (bc)$$

中空の三角形をなすサイクルは
三角形のバウンダリでもある

$$B_r(K, \mathbb{F}_2) \subseteq Z_r(K, \mathbb{F}_2) \subseteq C_r(K, \mathbb{F}_2)$$

穴に対応するものと、その中でもなにかの境界になっているものの関係
→これが図形のグループわけに使えるのではないか

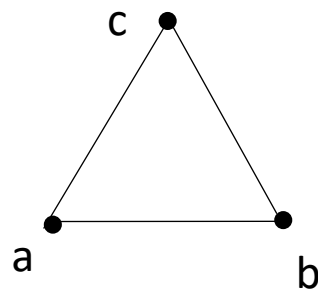
ホモロジー群

$$H_r(K, \mathbb{F}_2) = Z_r(K, \mathbb{F}_2) / B_r(K, \mathbb{F}_2)$$

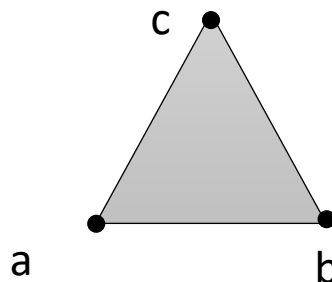
$Z_r(K, \mathbb{F}_2)$ のうち、 $B_r(K, \mathbb{F}_2)$ の元は単位元
と同じとみなすということ

具体的な計算事例

もっとも単純な例



$$\{a, b, c, (ab), (bc), (ca)\}$$



$$\{a, b, c, (ab), (bc), (ac), (abc)\}$$

$$H_1(K, \mathbb{F}_2) = Z_1(K, \mathbb{F}_2) / B_1(K, \mathbb{F}_2)$$

中空の三角形と
中身が詰まった三角形の違いは？

中空の場合

$B_1(K, \mathbb{F}_2) = 0$ (2単体がそもそも単体的複体の要素にない)

$$C_1(K, \mathbb{F}_2) = l(ab) + m(bc) + n(ca) \quad l, m, n \in \mathbb{F}_2 \quad (0 \text{ か } 1)$$

$m = l = n$ が、サイクル群の要素の条件

$$Z_1(K, \mathbb{F}_2) = \{n\{(ab) + (bc) + (ca)\} \mid n \in \mathbb{F}_2\}$$

$$\partial_1 C_1(K, \mathbb{F}_2) = (n + l)a + (l + m)b + (m + n)c \quad \text{なので}$$

$$H_1(K, \mathbb{F}_2) = Z_1(K, \mathbb{F}_2) / B_1(K, \mathbb{F}_2) \cong \mathbb{F}_2$$

具体的な計算事例

中身が詰まった三角形

$$C_2(K, \mathbb{F}_2) = m(abc) \quad m \in \mathbb{F}_2$$

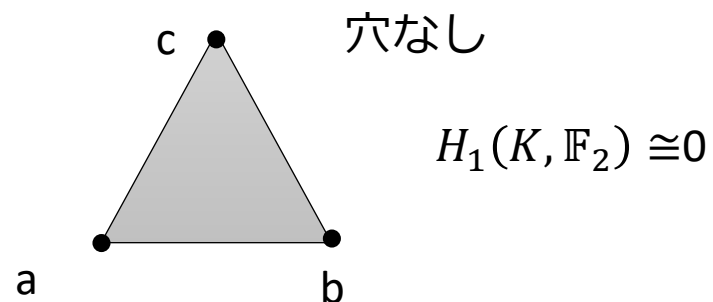
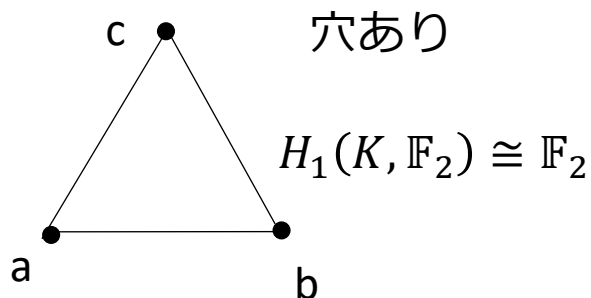
$$\partial_2 C_2(K, \mathbb{F}_2) = m\{(bc) + (ca) + (ab)\} = B_1(K, \mathbb{F}_2)$$



$$Z_1(K, \mathbb{F}_2) = \{n\{(ab) + (bc) + (ca)\} \mid n \in \mathbb{F}_2\}$$

$B_1(K, \mathbb{F}_2)$ と $Z_1(K, \mathbb{F}_2)$ は同じである

$H_1(K, \mathbb{F}_2) = Z_1(K, \mathbb{F}_2)/B_1(K, \mathbb{F}_2) \cong 0$ $B_1(K, \mathbb{F}_2)$ に含まれる要素を単位元とみなすなら、単位元しか残らない

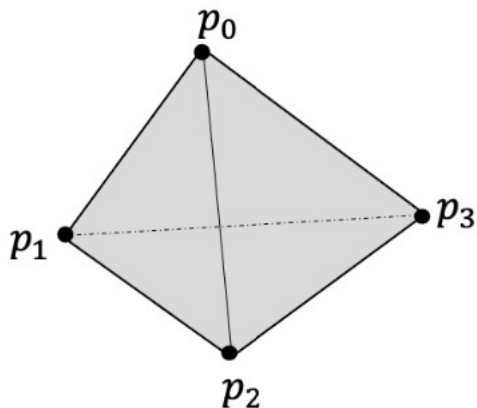


$H_1(K, \mathbb{F}_2) \rightarrow$ 穴の有無を反映

ホモロジー群の数値計算

もっとややこしい構造では手計算では厳しくなる

これぐらいの図形(中空の4面体)でもかなりしんどい



$$\begin{aligned} &\{p_0, p_1, p_2, p_3 \\ &(p_0p_1), (p_0p_2), (p_0p_3), (p_1p_2), (p_1p_3), (p_2p_3) \\ &(p_0p_1p_2), (p_0p_1p_3), (p_0p_2p_3), (p_1p_2p_3)\} \end{aligned}$$

$$C_2(K, \mathbb{F}_2) = a_2(p_0p_1p_2) + b_2(p_0p_1p_3) + c_2(p_0p_2p_3) + d_2(p_1p_2p_3) \quad a_2, b_2, c_2, d_2 \in \mathbb{F}_2$$

$$\begin{aligned} \partial_2 C_2(K, \mathbb{F}_2) = & a_2\{(p_1p_2) + (p_0p_2) + (p_0p_1)\} + b_2\{(p_1p_3) + (p_0p_3) + (p_0p_1)\} \\ & + c_2\{(p_2p_3) + (p_0p_3) + (p_0p_2)\} + d_2\{(p_2p_3) + (p_1p_3) + (p_1p_2)\} \end{aligned}$$

バウンダリやサイクルになる条件は係数の条件としてどう書ける??



バウンダリ行列

2単体→1単体

$$\partial_2 C_2(K) = a_2\{(p_1p_2) + (p_0p_2) + (p_0p_1)\} + b_2\{(p_1p_3) + (p_0p_3) + (p_0p_1)\} \\ + c_2\{(p_2p_3) + (p_0p_3) + (p_0p_2)\} + d_2\{(p_2p_3) + (p_1p_3) + (p_1p_2)\}$$

境界作用素の結果→ $(p_0p_1p_2), (p_0p_1p_3), (p_0p_2p_3), (p_1p_2p_3)$ から
どのように $(p_0p_1), (p_0p_2), (p_0p_3), (p_1p_2), (p_1p_3), (p_2p_3)$ が派生するか

この情報は行列として表現できる

$$\partial_2 \langle (p_0p_1p_2), (p_0p_1p_3), (p_0p_2p_3), (p_1p_2p_3) \rangle = \langle (p_0p_1), (p_0p_2), (p_0p_3), (p_1p_2), (p_1p_3), (p_2p_3) \rangle A_2$$

$$(p_0p_1p_2) \rightarrow (p_1p_2) + (p_0p_2) + (p_0p_1)$$

$$(p_1p_2p_3) \rightarrow (p_1p_2) + (p_1p_3) + (p_2p_3)$$

$$A_2 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

バウンダリ行列

同じように1単体→0単体に対応する行列を考える

$$\partial_1 < (p_0 p_1), (p_0 p_2), (p_0 p_3), (p_1 p_2), (p_1 p_3), (p_2 p_3) > = < p_0, p_1, p_2, p_3 > A_1$$

$$A_1 = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

←実際こうなるか考えてみましょう

$Z_r(K, \mathbb{F}_2)$ と $B_r(K, \mathbb{F}_2)$ の情報はこれらの行列に含まれている

$\uparrow$ $\uparrow$
r次元の境界作用素で常に0になる基底の組み合わせ (核: Ker)
r+1次元の境界作用素の結果がカバーする範囲 (値域: Im)

A_1 や A_2 の値域・核を効率的に求めるには？→標準形の利用

標準形

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

行や列の入れ替え、
他の行や列を足したり引いたりすることを繰り返すと、
1が対角成分にあるブロックと、
その他は0という形にすることができる
(線形代数で習う行列の階数を求めるときの変形)

手計算だとしんどすぎ
→SageMathなどを使いましょう

```
In [ ]: A1=matrix(GF(2),[[1,1,1,0,0,0],[1,0,0,1,1,0],[0,1,0,1,0,1],[0,0,1,0,1,1]])  
A1.smith_form()
```

```
Out [ ]: (
```

	[1 1 0 1 1 0]	
	[0 1 1 1 0 1]	
[1 0 0 0 0 0]	[1 0 0 0]	[0 0 1 0 1 1]
[0 1 0 0 0 0]	[1 1 0 0]	[0 0 0 1 0 0]
[0 0 1 0 0 0]	[1 1 1 0]	[0 0 0 0 1 0]
[0 0 0 0 0 0]	[1 1 1 1]	[0 0 0 0 0 1]

```
)
```

標準形への変換

4面体での2单体（面）→1单体（辺）に対応する行列

標準形

$$A_2 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$



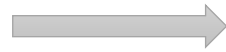
$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

6x4行列で、0でない対角成分があるブロックは3x3

4面体での1单体（辺）→0单体（頂点）に対応する行列

標準形

$$A_1 = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$



$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

4x6行列で、0でない対角成分があるブロックは3x3

標準形とホモロジー群

k次元での境界作用素に対する行列の標準形

k次元単体の数
 n_k

m_k

k-1次元単体の数
 n_{k-1}

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

対角成分が1のブロックが (m_k, m_k) とする

$$Z_k(K, \mathbb{F}_2) \cong (\mathbb{F}_2)^{n_k - m_k}$$

$$B_k(K, \mathbb{F}_2) \cong (\mathbb{F}_2)^{m_{k+1}}$$

$$H_k(K, \mathbb{F}_2) \cong (\mathbb{F}_2)^{n_k - m_k - m_{k+1}}$$

境界作用素の
行列表現の階数



穴の数を決める

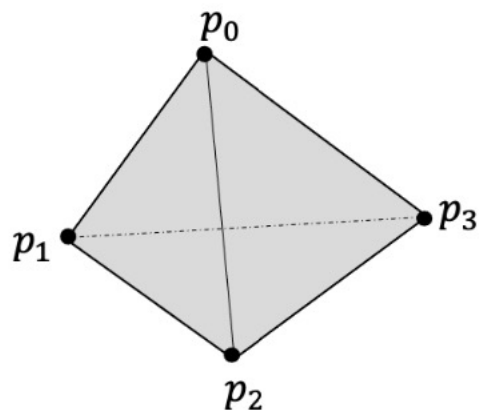
サイクル群：k次元単体の数－k次元境界作用素の標準形で0ではない対角ブロックの行数

バウンダリ群：k+1次元境界作用素の標準形で0ではない対角ブロックの行数

ホモロジー群：サイクル群のもの－バウンダリ群のもの

実際の計算例

中空の四面体の例



$$\begin{aligned} &\{p_0, p_1, p_2, p_3 \\ &(p_0p_1), (p_0p_2), (p_0p_3), (p_1p_2), (p_1p_3), (p_2p_3) \\ &(p_0p_1p_2), (p_0p_1p_3), (p_0p_2p_3), (p_1p_2p_3)\} \end{aligned}$$

これらの $H_1(K, \mathbb{F}_2)$ と $H_2(K, \mathbb{F}_2)$ を計算してみましょう

$$H_1(K)$$

1 次の境界作用素の行列: A_1 $\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$ $m_1 = 3, n_1 = 6$

2 次の境界作用素の行列: A_2 $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ $m_2 = 3, n_2 = 4$

$$H_1(K, \mathbb{F}_2) \cong \mathbb{F}_2^{n_1 - m_1 - m_2} = 0$$

実際の計算例

$H_2(K)$

2 次の境界作用素の行列 : A_2 $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

$$m_2 = 3, n_2 = 4$$

3次の境界作用素の行列？

→ 3 単体（4 面体そのもの）は含まれていないので、 $m_3 = 0, n_3 = 0$

$$H_2(K, \mathbb{F}_2) \cong \mathbb{F}_2^{n_2 - m_2 - m_3} = \mathbb{F}_2$$

4面体内部の空隙を捉えていることに相当

ホモロジー群の計算は、境界作用素を行列で表現することで、
行列変形の問題にすることができる

手計算ではしんどい？ → SageMathなどの数式処理ソフトウェアが使えます

SageMathの利用例

SageMath : フリーで使えるMathematicaのようなソフトウェア



Collaborative Calculation and Data Science

Google CollaboratoryのようにWeb上でSageMathを利用できるCoCalcもある

中身の詰まった4面体と、中空の4面体でのホモロジー群を計算してみましょう

中身が詰まった（3単体になっている）場合を考えてみます。SageMathは最も次数の高い単体を指定して、単体的複体を定義すると、必要な面も全て含んだものが用意されます。

ホモロジー群の計算には、デフォルトではReduced Homologyが計算され、これはスライドで説明したものと少し違うので、定義通りのホモロジー群を計算させるには、`reduced=False` のオプションを使います。

```
In [2]: X=SimplicialComplex([[0,1,2,3]])
        X.homology(reduced=False, base_ring=GF(2))
```

```
Out[2]: {0: Vector space of dimension 1 over Finite Field of size 2,
        1: Vector space of dimension 0 over Finite Field of size 2,
        2: Vector space of dimension 0 over Finite Field of size 2,
        3: Vector space of dimension 0 over Finite Field of size 2}
```

中身が空の四面体ではどうでしょうか？さっきの議論だと、2次のホモロジー群は上の例と違ってF2が答えになるはずです。

```
In [3]: X=SimplicialComplex([[0,1,2],[1,2,3],[0,2,3],[0,1,3]])
        X.homology(reduced=False, base_ring=GF(2))
```

```
Out[3]: {0: Vector space of dimension 1 over Finite Field of size 2,
        1: Vector space of dimension 0 over Finite Field of size 2,
        2: Vector space of dimension 1 over Finite Field of size 2}
```



パーシステントホモロジー群とは？

ホモロジー群の問題

ホモロジー群：行列の変換で計算できて扱いやすい。

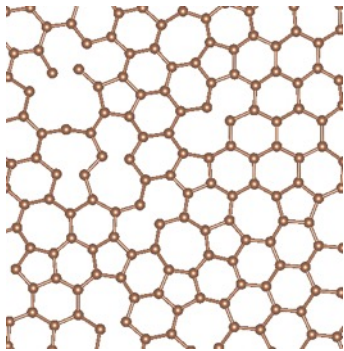
「穴が空いているか」の情報を抽出→図形の同じさの基準

問題点

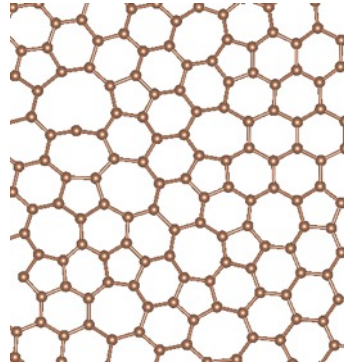
どの頂点がつながっているかが少し変わっただけで劇的に値が変わりうる

どの部分の構造が重要なのかななどを解析しにくい

1.5Åがしきい値



1.8Åがしきい値



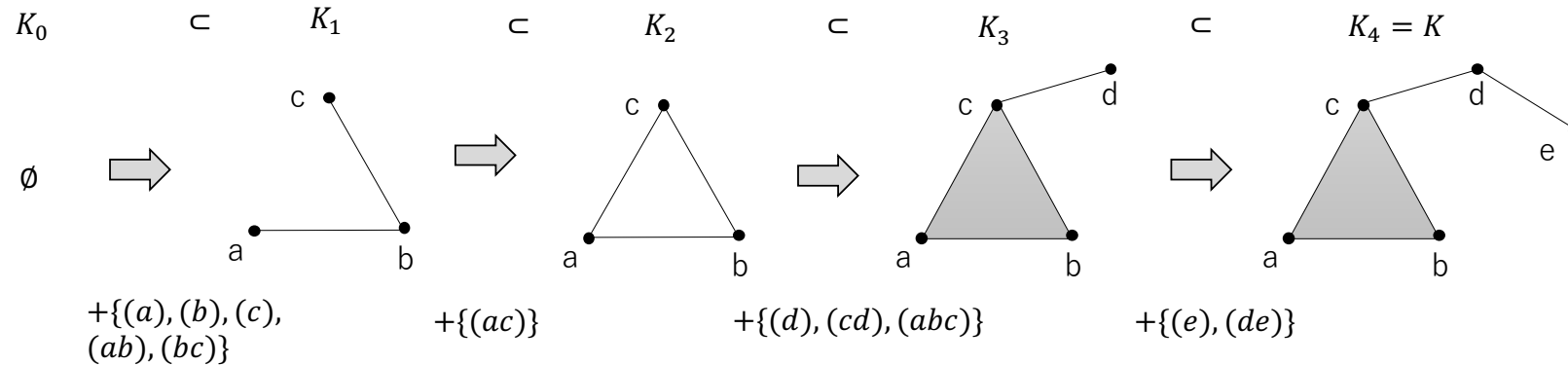
原子構造に用いる場合には特に、
少しの原子間距離の違いで
値が大きく変わると、
似ているものを全く違うと捉えてしまう

解像度や連続性の面で、そのままと材料科学には応用しにくい

パーシステントホモロジー (PH) : 様々なスケールでのホモロジー

スケールが変わっていったときにホモロジー群がどのように移り変わるか

- ・ 何らかのスケール依存性を反映した単体複体の内包関係の系列 (フィルトレーション) を作る

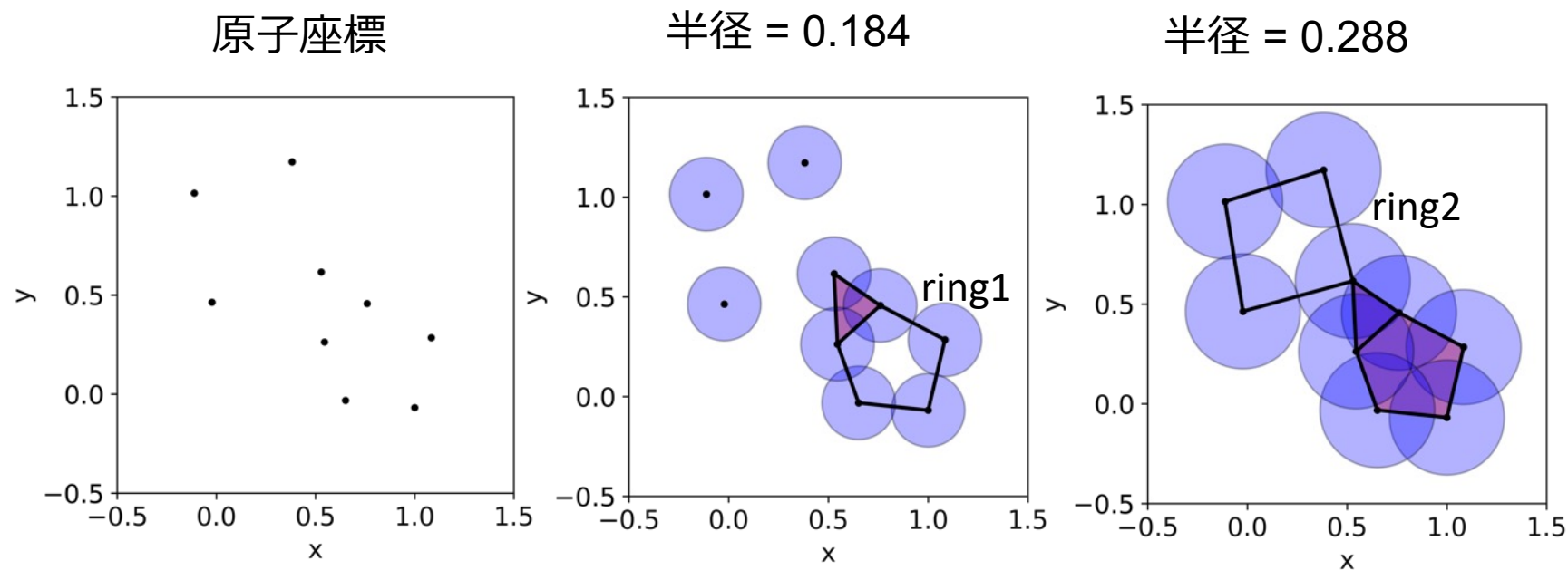


- ・ ホモロジー群として変化する部分・変化しない部分を追跡する

(K_2 でサイクルができるが、それが K_3 でバウンダリになる というような変化を記録する)

原子構造に対するフィルトレーションの構築

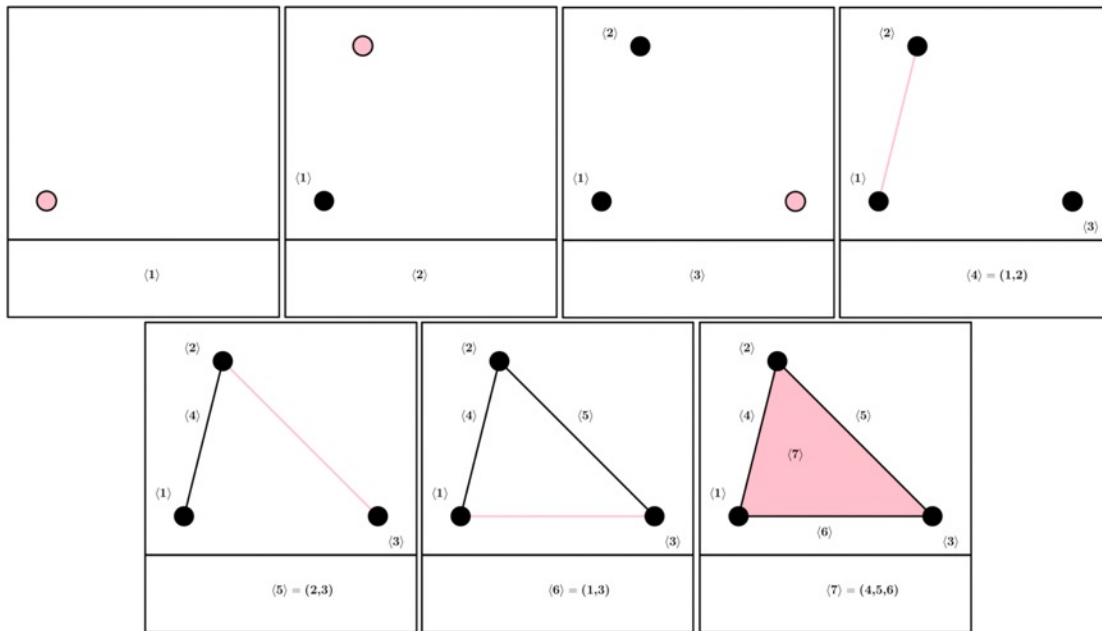
単体的複体の内包関係が成立していればよいが、原子座標に使う場合には、チェック複体やそれと同等なアルファ複体を用いる



パーシステントホモロジーの計算

ホモロジー群の移り変わりを効率的に捉えるには？

フィルトレーションの中で、ある単体がどの単体の面になっているかを行列で表記する



P. Pranov, et al., MNRAS 465, 4281, (2017)

7つの段階でそれぞれ一つずつ単体ができる。
新しくできた単体を<i>のように書く。
その単体になっている面がどれかを行列で表現

たとえば<1>,<2>は<4>の面になっている

$$\begin{array}{c}
 \begin{matrix} <1> <2> <3> <4> <5> <6> <7> \end{matrix} \\
 \begin{pmatrix}
 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0
 \end{pmatrix}
 \end{array}$$

(i,j)要素はiがjの面であれば1

パーシステントホモロジーの計算

Matrix reduction

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

列jで1が入っている最も下方にある行の番号をlow(j)とする

ある $j_0 < j$ で $\text{low}(j_0) = \text{low}(j)$ となる列があれば、 j_0 列をj列に足す

上記の計算を繰り返す

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \textcolor{blue}{1} & \textcolor{blue}{1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$



$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & \textcolor{red}{1} & 1 & \textcolor{red}{1} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$



$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \textcolor{green}{1} & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \textcolor{green}{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \textcolor{green}{1} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$\mathbb{F}_2$ で足し算

パーシステントホモロジーの計算

最終的に得られた行列

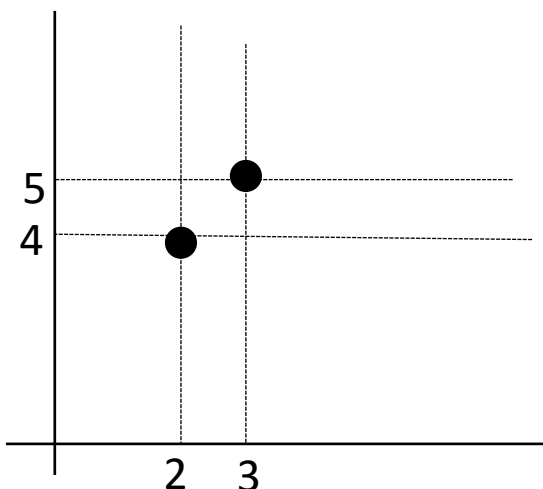
$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

列で一番下に1がある要素の（行番号、列番号）を読み取る

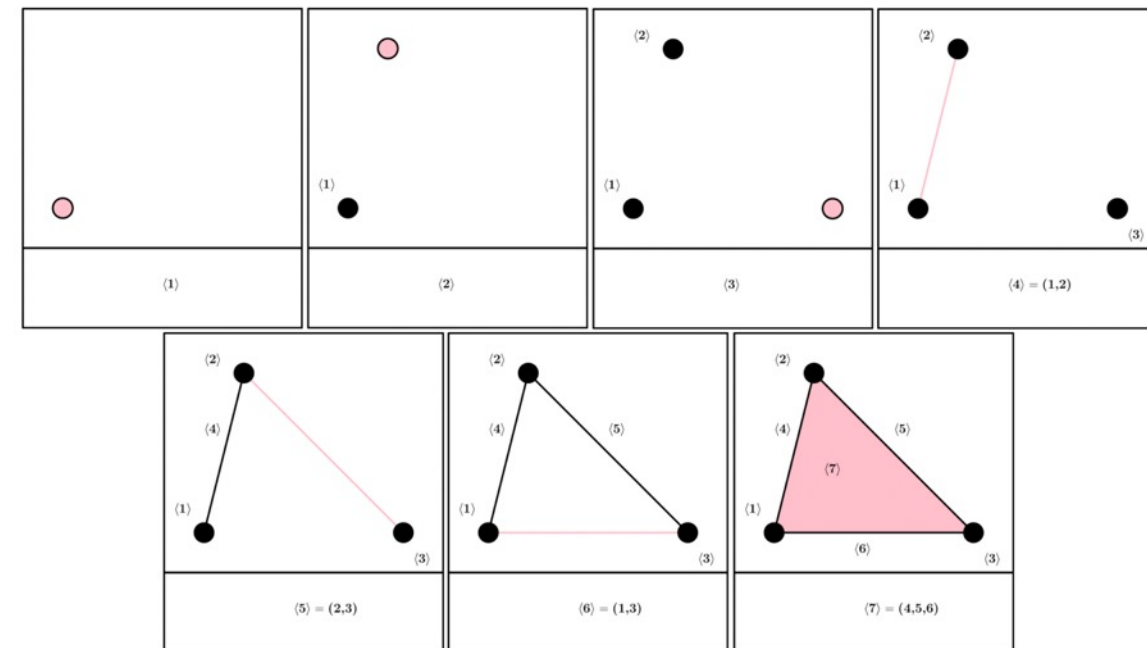
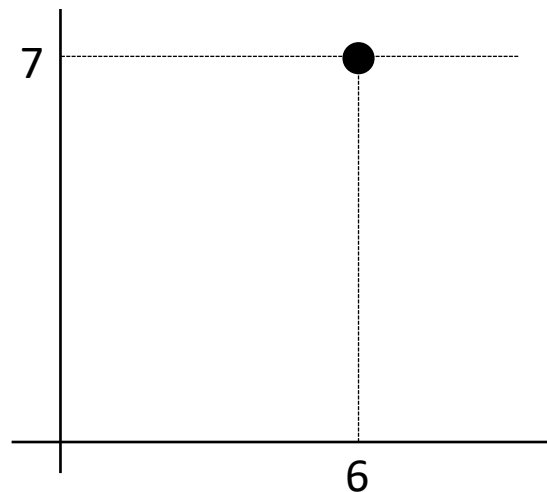
→サイクルが生成したタイミングと、
それがバウンダリになったタイミングに対応する

(2,4) (3,5), (6,7)

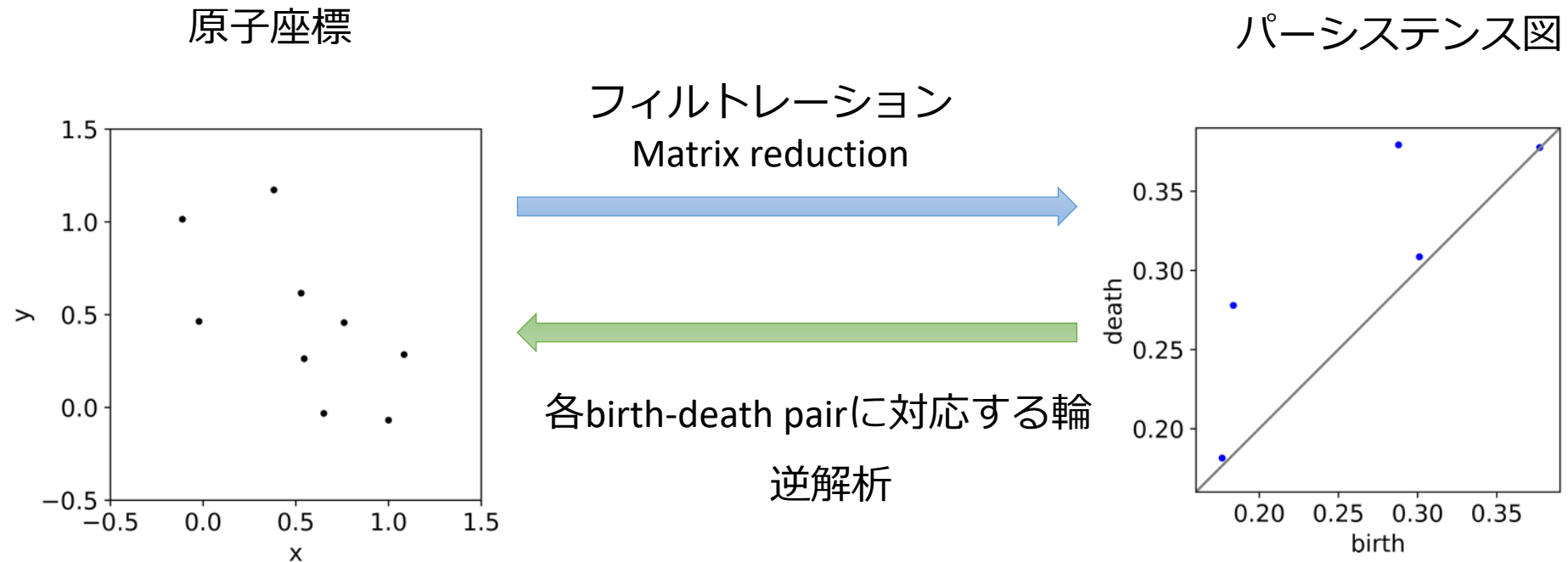
0次元



1次元



逆解析

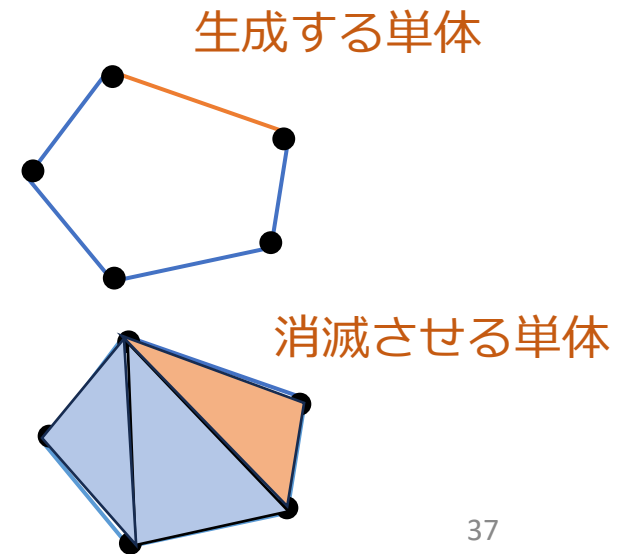


どんな構造が大事か？には逆解析が重要

逆解析の難しいところ：トポロジ的に同じ輪が複数存在しうる

Matrix reductionで読み取る情報：サイクルを生成する単体と消滅させる単体

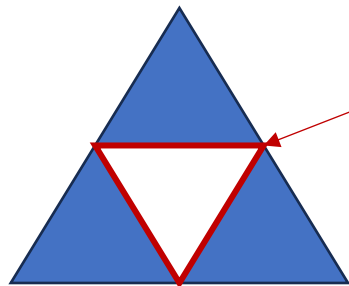
生成する単体と消滅させる単体が定まっても、それ以外の部分をどう取るか、複数の選択肢がありうる



逆解析

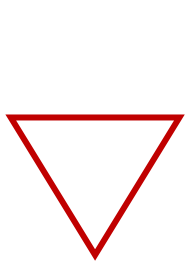
トポロジ的に同じ輪とは？

中身の詰まった三角形に囲まれた中空の三角形

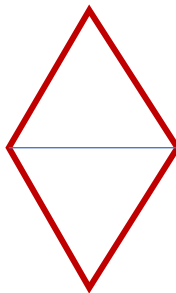


直感的にそれらしく思えるサイクル

ホモロジー群では「バウンダリ群に含まれる要素は単位元とおなじ」
中身の詰まった三角形のバウンダリを足したり引いたりしたものと同じ



と



はトポロジーとしては同じ

どっちを選べばいいの...?

どれを選択するか？

構成する単体の個数が最小のもの：最適サイクル

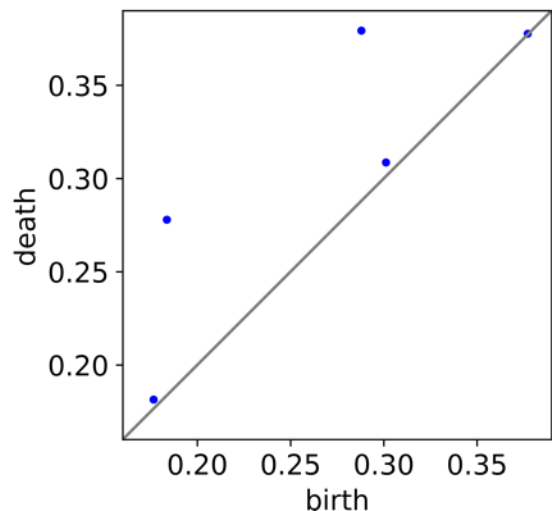
E. G. Escolar, Y. Hiraoka, *Optimization in the Real World* (Springer Japan, 2016), pp. 79–96.

サイクルの囲う体積が最小のもの：体積最小サイクル

I. Obayashi, *SIAM Journal on Applied Algebra and Geometry*. **2**, 508–534 (2018). ³⁸

パーシステンス図の機械学習への応用

パーシステンス図



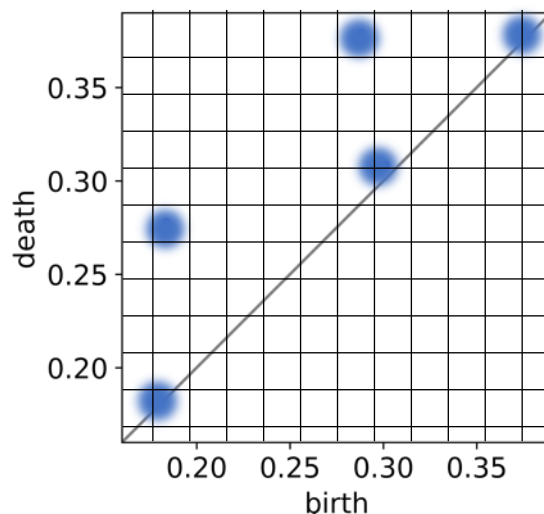
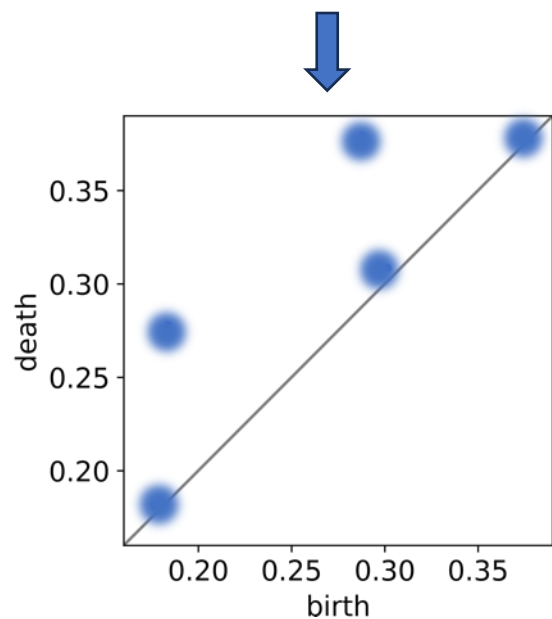
Birth-deathペアの個数は一定ではない
→通常の機械学習の入力としては不向き

固定長のベクトルの形に変換するには？

Persistence image

H. Adams et al., Journal of Machine Learning Research, 18, 1-35 (2017)

I. Obayashi, H. Hiraoka and M. Kimura, Journal of Applied and Computational Topology, 1, 421-449 (2018)



ガウシアンカーネル+重みで分布関数を定義
メッシュ状での値→ベクトル

$$\rho(x, y) = \sum_{k=1}^{\ell} w(b_k, d_k) \exp \left(-\frac{(b_k - x)^2 + (d_k - y)^2}{2\sigma^2} \right),$$

$$w(b, d) = \arctan(C(d - b)^p).$$

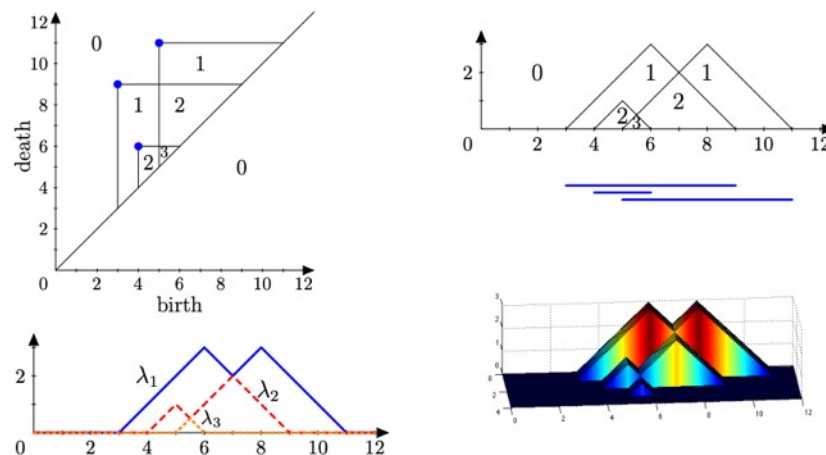
パーシステンス図の機械学習への応用

他のベクトル化手法

Persistence landscape

Birth-deathペアから対角線に線を引く
できた三角形の形状
→birth-deathペアの特徴を反映している

P. Bubenik, *arXiv* 1207.6437(2012)



Persistence imageやpersistence landscapeを変換関数として一般化
重み付けの部分を順序によらない変換を与えるタイプのニューラルネットワークで実装
→**Perslay**

M. Carrière et al., *Arxiv* 1904.09378 (2019)

パーシステントホモロジーの計算ソフトウェア

HomCloud

<https://homcloud.dev/>

日本のグループが開発。逆解析のモジュールがあるのが強力

GUDHI

<https://gudhi.inria.fr/>

機械学習への応用のためのベクトル化や
パーシステンス図間の計量の部分に力を入れている

Ripser

<https://github.com/Ripser/ripser>

Vietoris-Rips複体に特化。高速

Giotto-tda

<https://giotto-ai.github.io/gtda-docs/0.5.1/index.html>

機械学習への応用向けモジュールが豊富
時系列データを取り扱うモジュールもあり

などなど



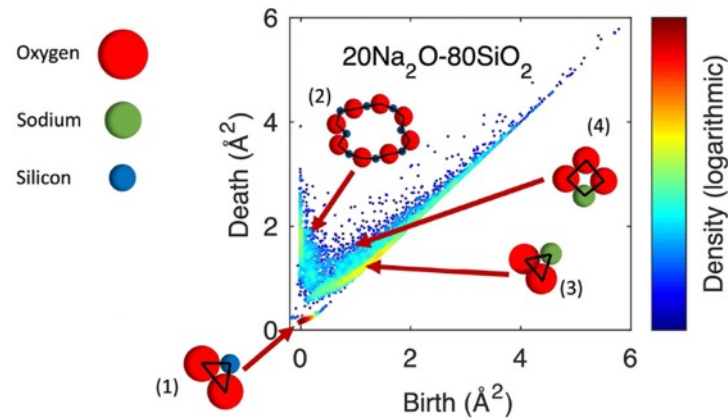
データサイエンスの計算物質科学への応用

-トポロジカルデータ解析・
パーシステントホモロジーを中心に-

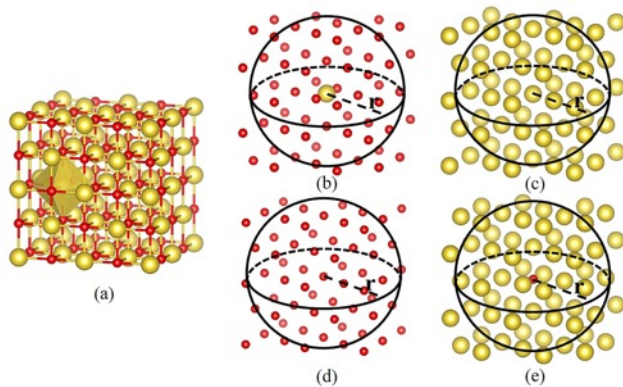
大阪大学産業科学研究所
南谷英美

材料科学分野でどんな応用がされているか

原子構造の特徴抽出



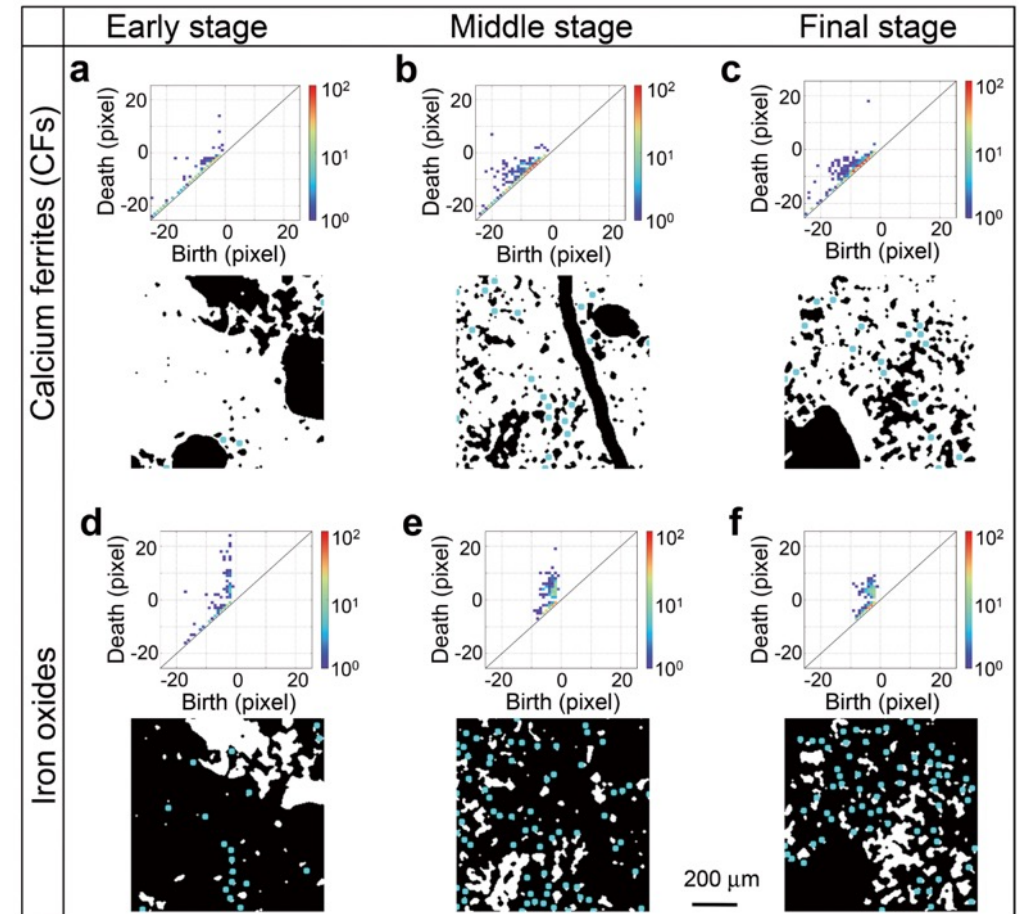
S. S. Sørensen et al., *Science Adv.* **6**, eabc2320 (2020).



Y. Jiang et al., *npj Computational Materials*, **7**, 28 (2021)

画像データの解析

鉄鉱石焼結鉱におけるマイクロクラック



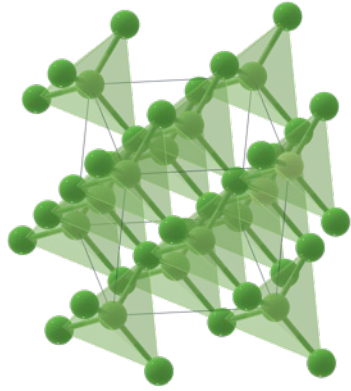
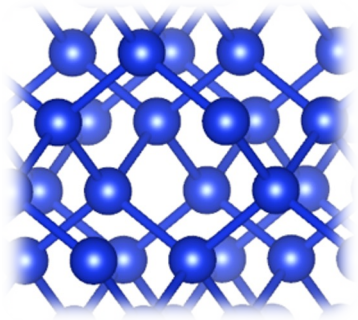
M. Kimura et al., *Sci. Rep.* **8**, 3553 (2018).

ガラスへの応用



非晶質の難しさ

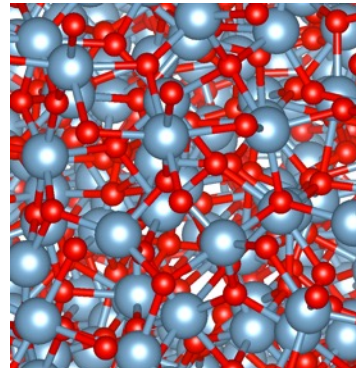
結晶



金属・半導体・絶縁体・トポロジカル絶縁体 etc...

きれいな周期性・対称性
→構造の特徴がわかりやすい
→比較的小さなモデルで記述できる

非晶質



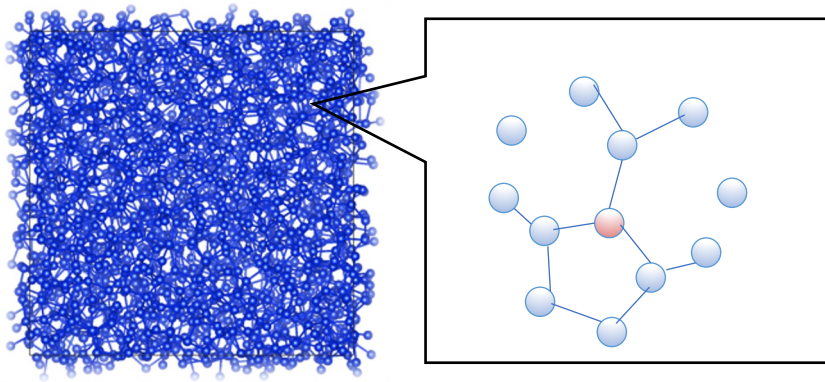
長距離秩序を持たない構造

- ・ ガラス・アモルファス
- ・ ポリマー、ゲル
- ・ 生体組織 etc...

構造の特徴はどう記述すればいいのか？

ガラス・アモルファスの構造特性

非晶質（アモルファス・ガラス）構造は完全にランダムというわけではない



原子間の相互作用の存在
→二面角のような原子位置の多体の相関に制限が加わる

最近接原子間距離を超えたスケールでの何らかの秩序
→中距離秩序：Medium range order（MRO）

中距離秩序の存在やその特性：各種物性と密接に関連するはず

とはいえ、Radial distribution functionなどでは中距離秩序の詳細を議論することが困難

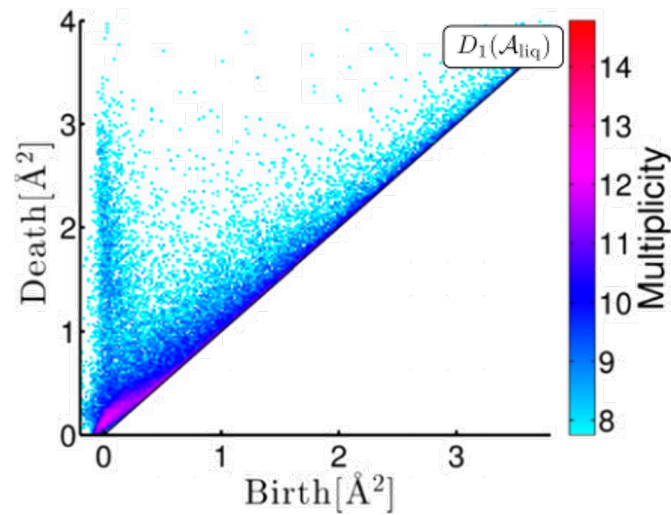
多数の原子の繋がり方、トポロジーの解析を行うと中距離秩序の存在や特性を掴む事ができる

トポロジカルデータ解析の応用

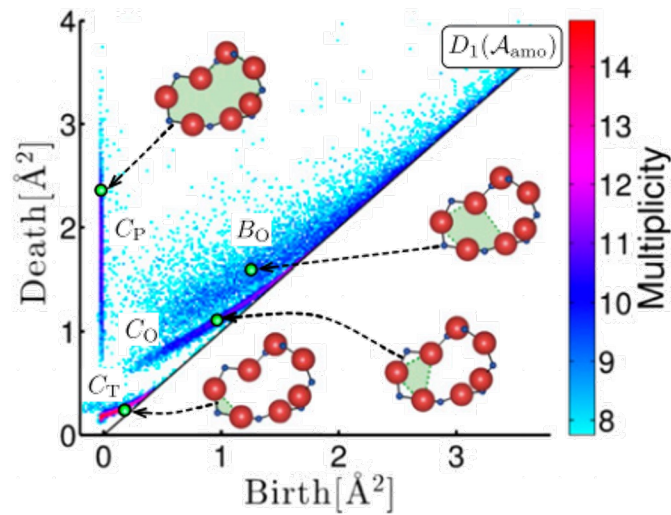
ガラス構造への応用

代表例

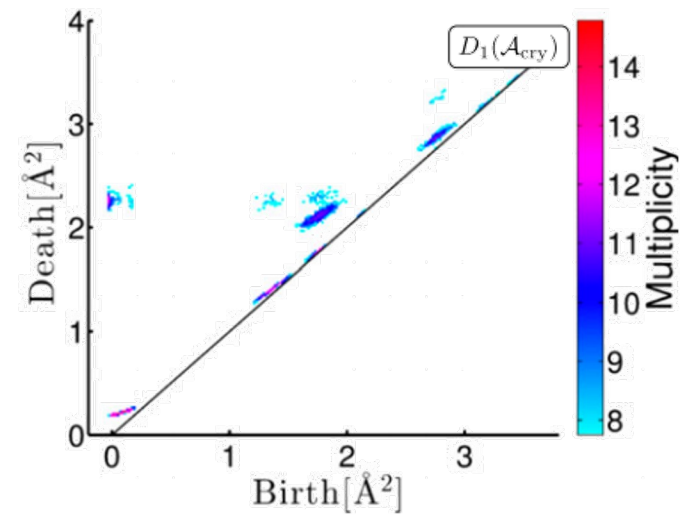
液体 SiO_2



アモルファス SiO_2



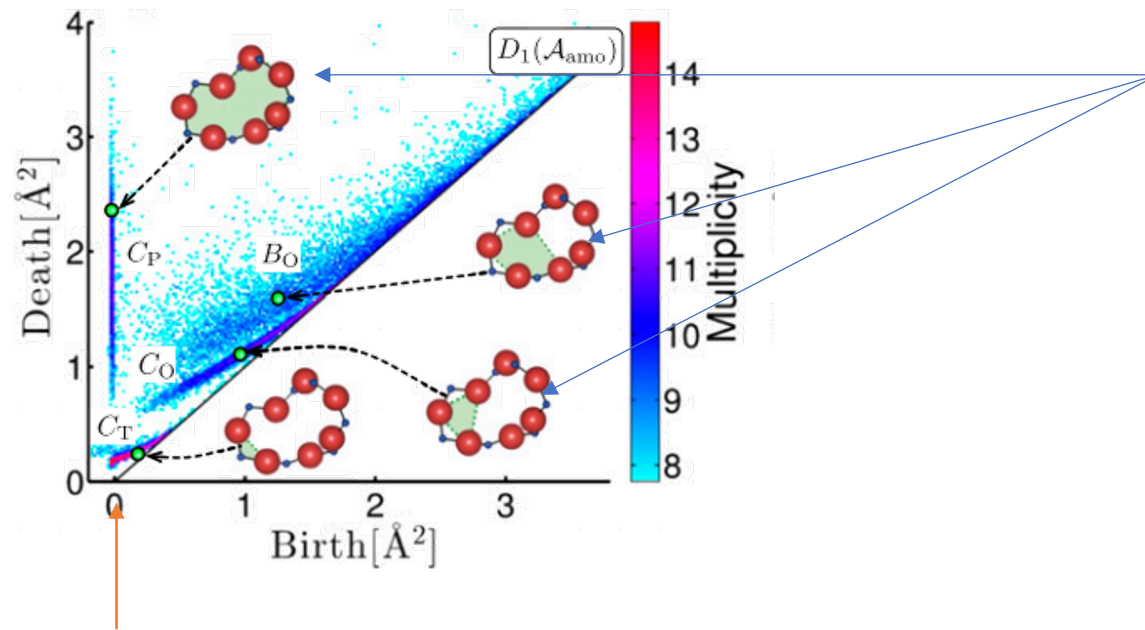
結晶 SiO_2



アモルファスの SiO_2 でのパーシステンス図には、液体・結晶とは異なるバンド状の構造が存在

Y. Hiraoka et al., PNAS, 113, 7035 (2016)

ガラス構造への応用

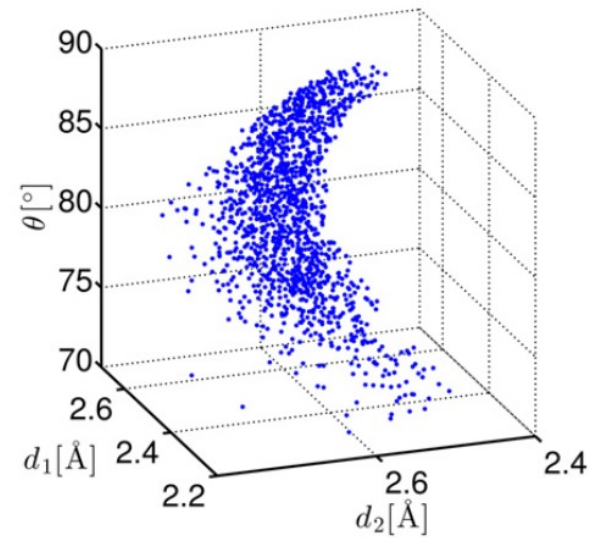


短距離秩序(SRO)に対応する構造

中距離秩序(MRO)に対応する構造

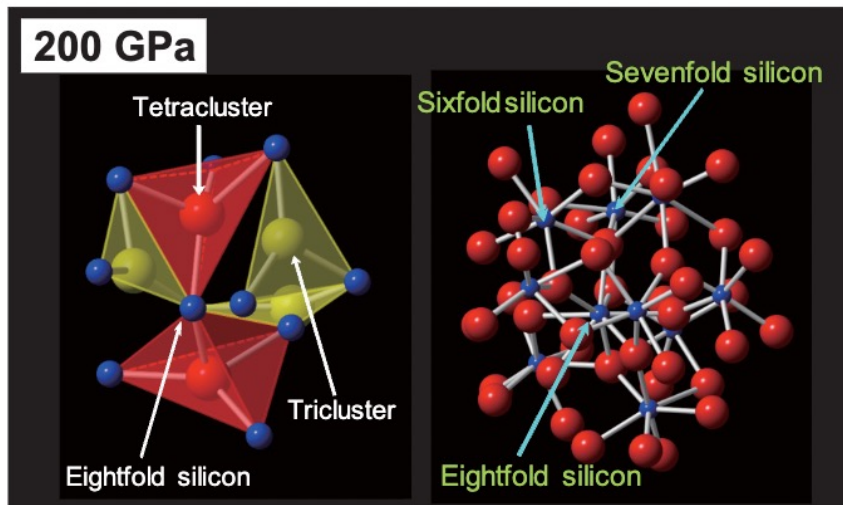
C_0 に対応する三角形

→原子間相互作用に由来する制約を反映している



Y. Hiraoka et al., PNAS, 113, 7035 (2016)

高圧下でのガラス構造の特徴



高圧化でのSiO₂ガラス

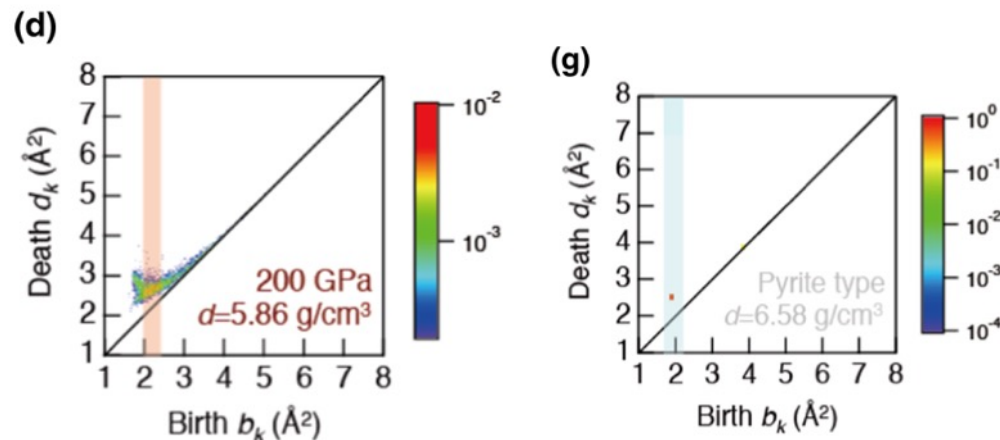
常圧化と異なる形状の特徴がある

常圧下：Siは4つのOと結合を持つ

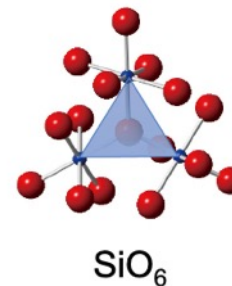
超高圧化：Siがより多数（6,7,8個）のOと結合を持つ

超高圧化での原子構造の特徴とは？

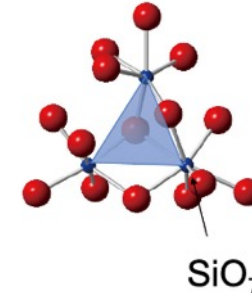
高圧下で安定なSiO₂結晶型（パイライト型）と共通した特徴がある



(a) Pyrite type
($d=6.58 \text{ g/cm}^3$)

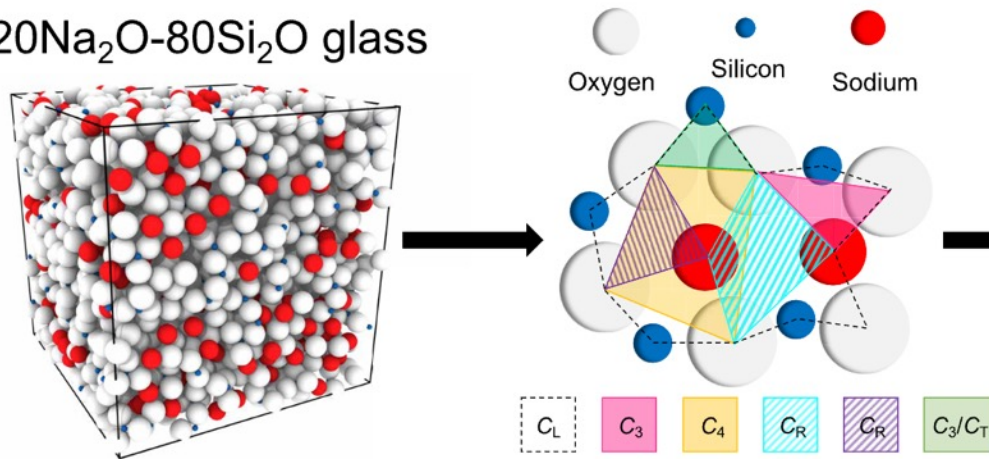


Glass at 200 GPa
($d=5.86 \text{ g/cm}^3$)

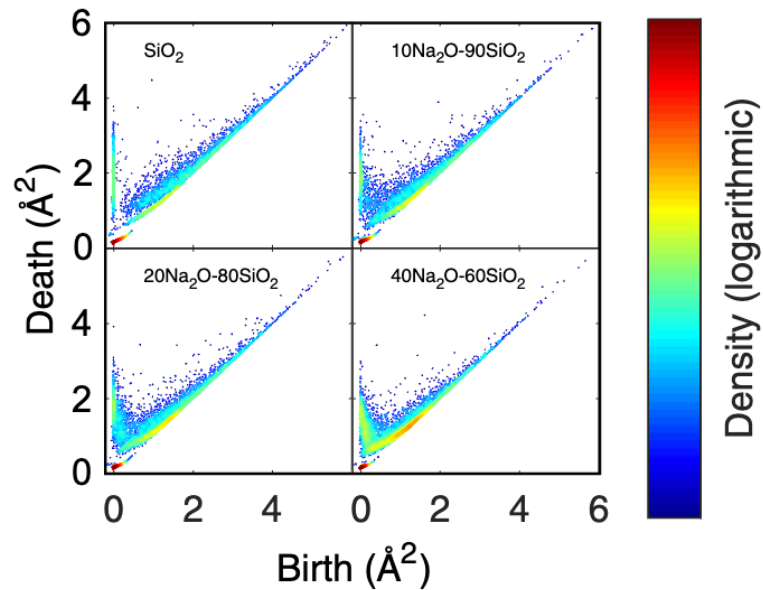
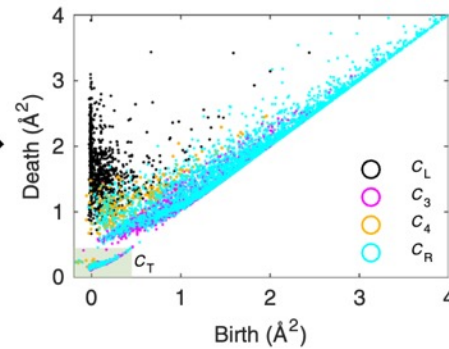


多元素系のガラスへの応用

20Na₂O-80Si₂O glass



Persistence diagram



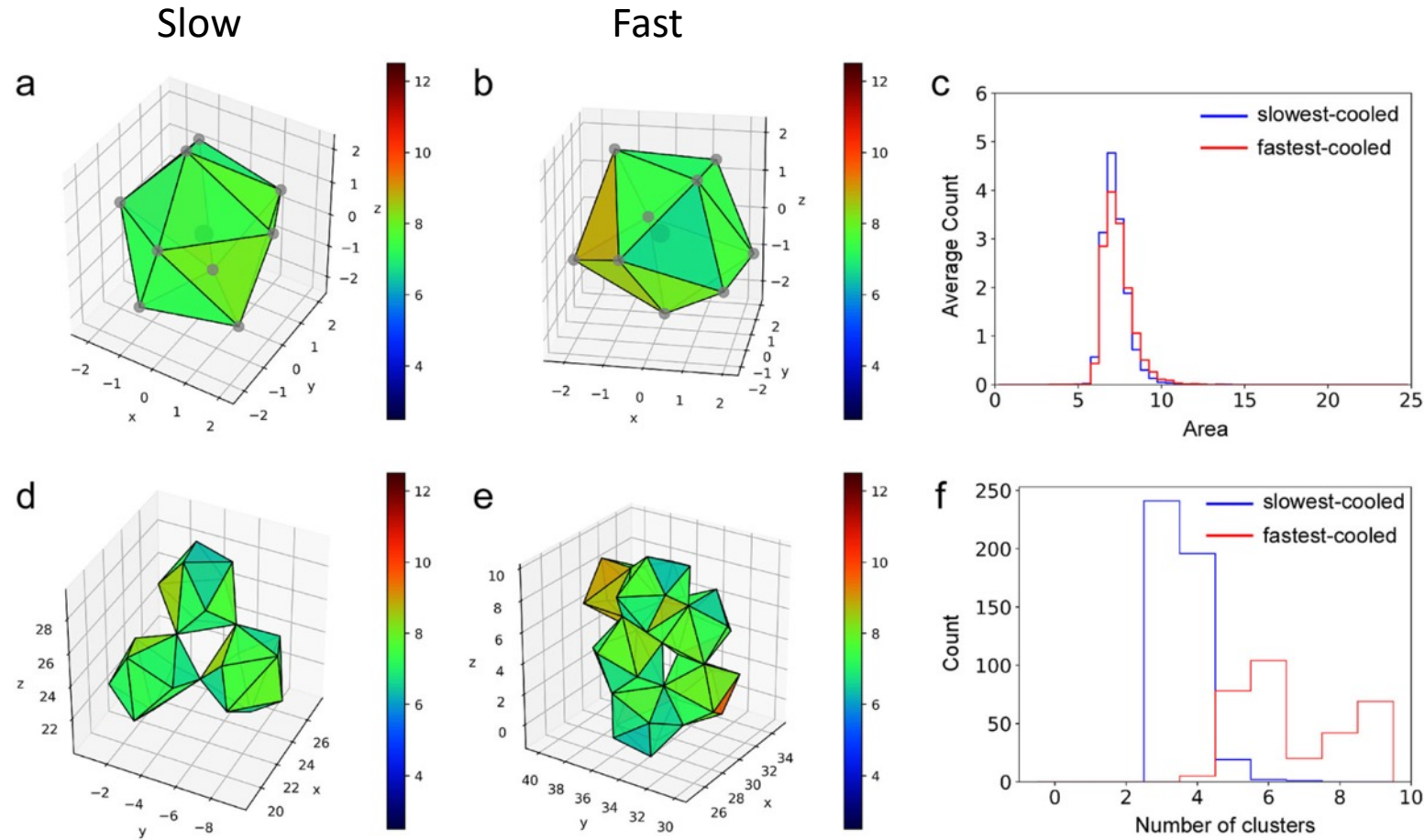
Naが加わると、 c_L のサイクルに対応する部分のペアが減少
→ c_3 , c_4 , c_R などのより小さなサイクルに対応するペアが増加

アルカリ金属の添加による構造変調をよく捉えている

S. S. Sørensen et al., *Science Adv.* **6**, eabc2320 (2020).

金属ガラスへの応用

金属ガラス ($\text{Pd}_{80}\text{Si}_{20}$) において、冷却レートが構造にどのような影響をあたえるか？



Pd原子の座標を使った
2次のパーシステンス図で
主要な点に対応する構造

Si原子の座標を使った
1次のパーシステンス図で
主要な点に対応する構造
(Si原子を囲むPd原子を使って構
造が書かれている)

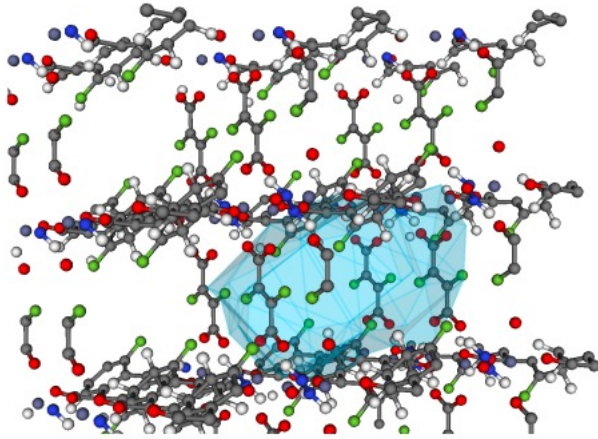
Si原子の成すネットワーク構造が変調

原子構造の記述子としての応用

機械学習との組み合わせ
物性予測へ

Metal Organic Framework(MOF)での例

MOF:多孔質の構造→ガスの分離・貯蔵材料

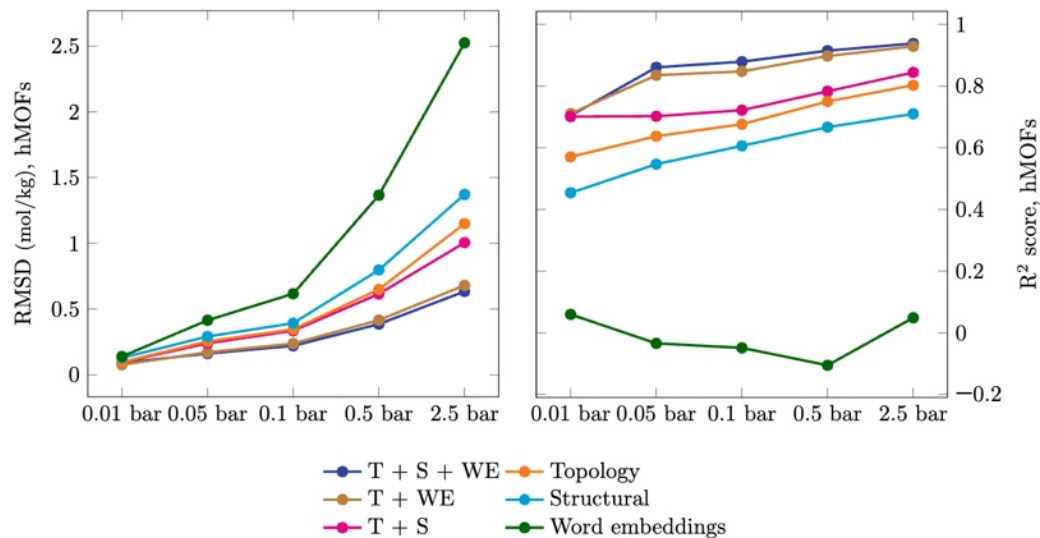


どんなガスがどれくらい吸着するか？

高压での吸着→穴の密度、形状
低压での吸着→化学的性質

幾何学的な記述子、化学組成
それぞれ単独では無理

パーシステントホモロジーによる記述子 (Persistent image)
を加えるとうか？



化学組成：Word Embedding

従来の記述子：Structural

パーシステントホモロジー：Topology



ランダムフォレストによる回帰

→パーシステントホモロジーを加えると性能が向上

分子での例

分子構造を記述する方法：分子化学分野で多数構築されてきた

クーロン行列

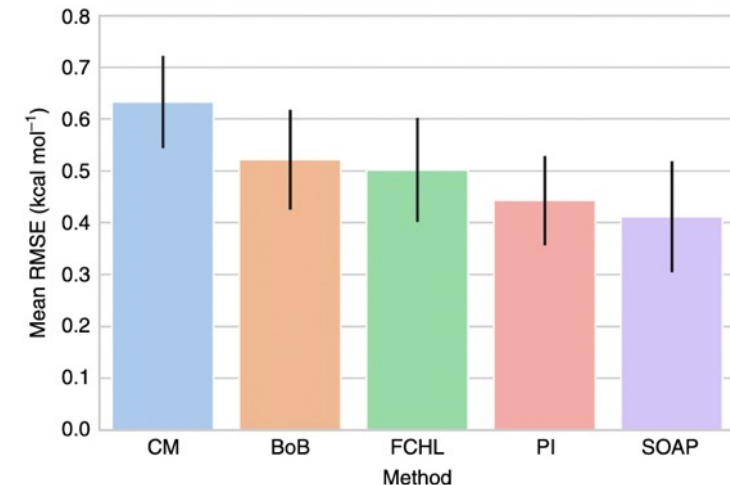
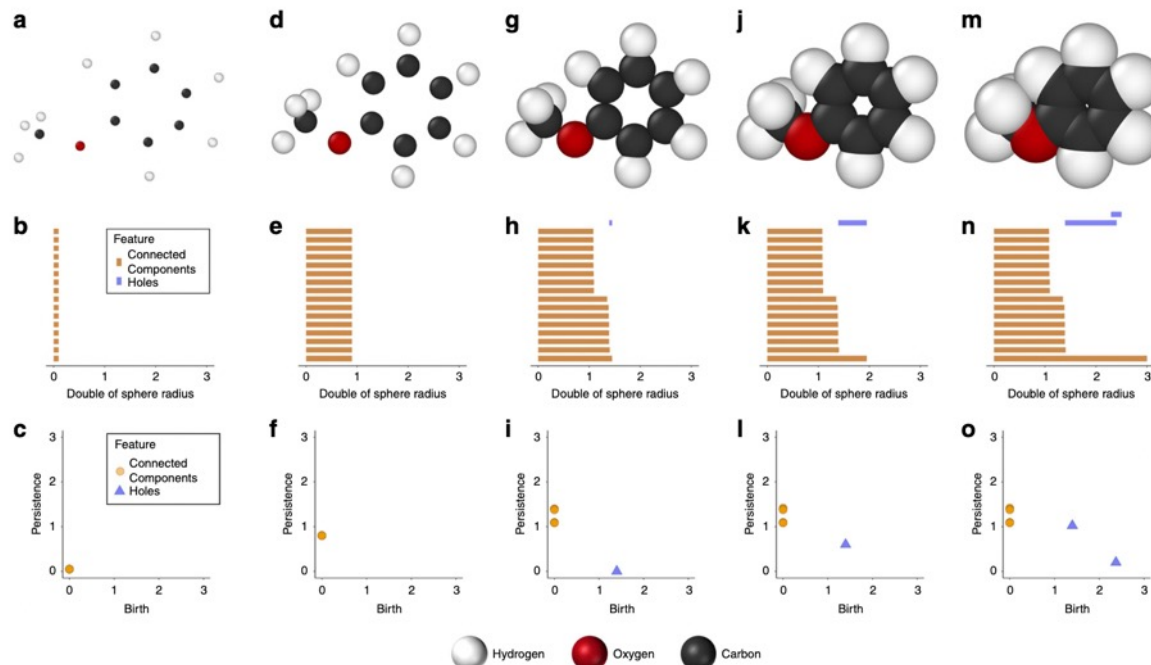
$$M_{ij} = \begin{cases} 0.5Z_i^{2.4} & \text{対角要素} \\ \frac{Z_i Z_j}{R_{ij}} & \text{非対角要素} \end{cases}$$

Smooth overlap of atomic positions (SOAP)

$$p(r_i)_{nn'l}^{Z_1 Z_2} = \pi \sqrt{\frac{8}{2l+1}} \sum_m c_{nlm}^{Z_1}(r_i)^* c_{n'l m}^{Z_2}(r_i)$$

トポロジカルな記述子は分子でも有用か？

CO2との相互作用エネルギー in GDB-9データセット

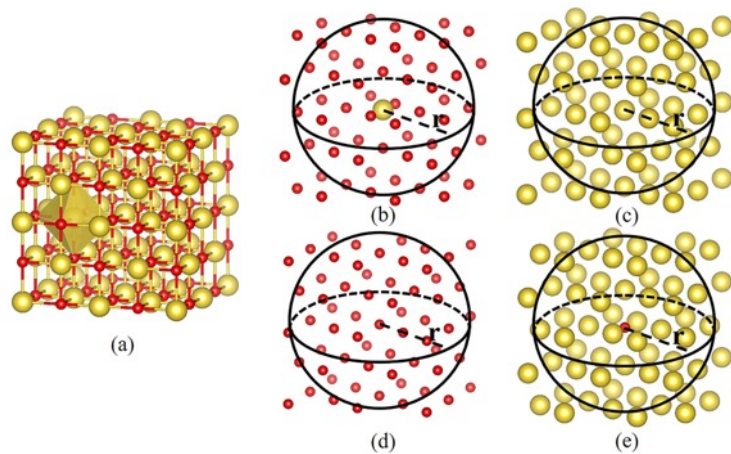


J. Townsend et al., *Nat. Commun.* **11**, 1–9 (2020).

結晶材料での例

パーシステントホモロジーは結晶材料でも有用か？

元素の違いをどのように取り込むか？ → Atom-specific persistent homology(ASPH)



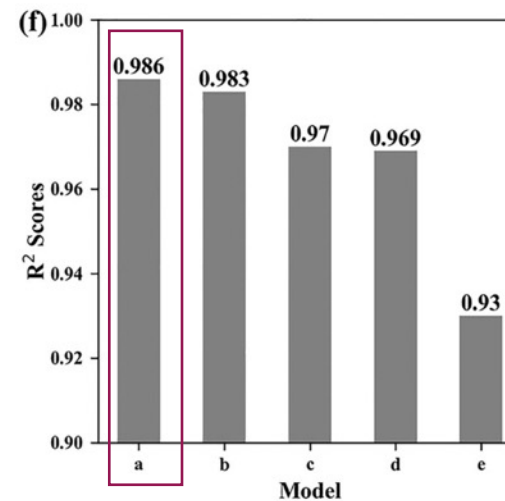
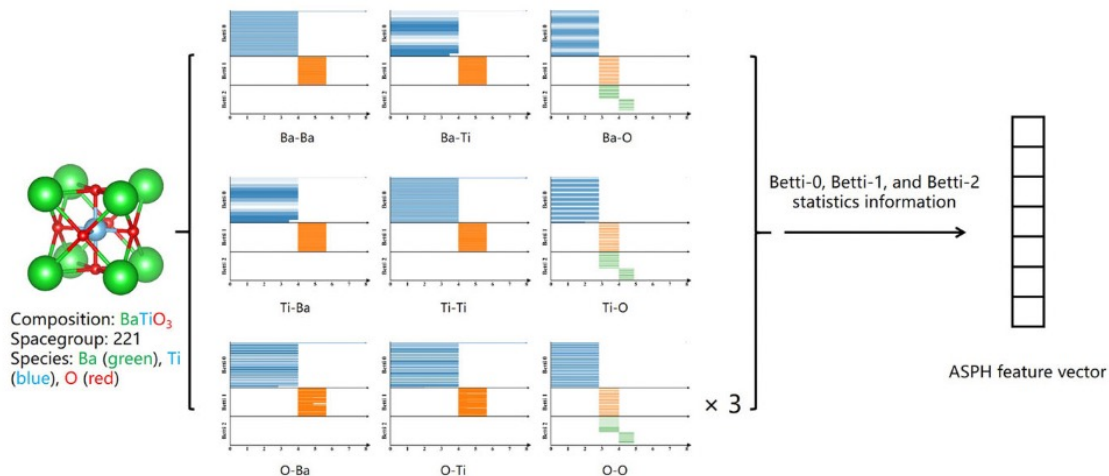
中心原子周りの環境を記述するには？

中心原子の種類: α
周囲の原子の種類: β

$$P_{\alpha i}^{\beta}$$

カットオフ内の原子座標で
birth-death pairを計算

→どの元素との関係を見ているかを反映したパーシステンス図



ASPH + 組成の情報
→ICSDデータベース内の
生成エンタルピーを高精度に予測

Y. Jiang et al.,
npj Computational Materials.
7, 28 (2021)

アモルファスSiにおける 熱伝導率と構造の相関

共同研究者

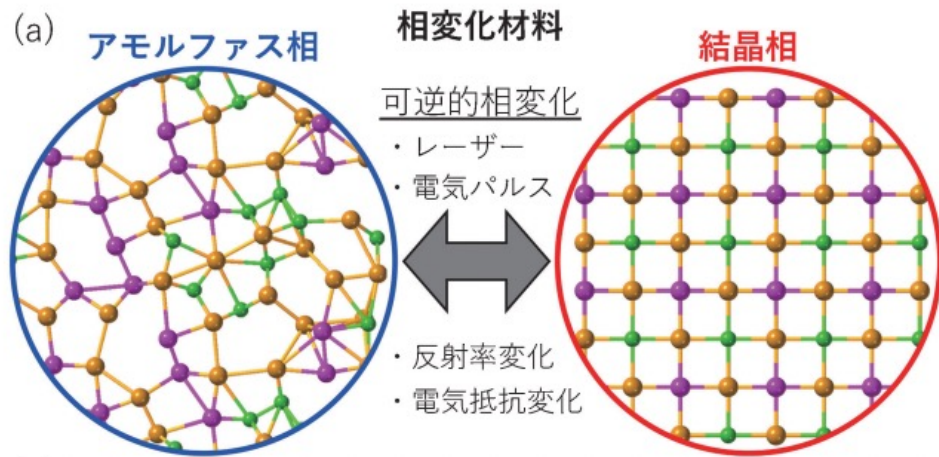
産業技術総合研究所 志賀拓磨
青山学院大学 柏木誠
岡山大学 大林一平

E. Minamitani, T. Shiga, M. Kashiwagi, I. Obayashi,
J. Chem. Phys. 156, 244501 (2022)



アモルファスの様々な物性

可逆的相変化

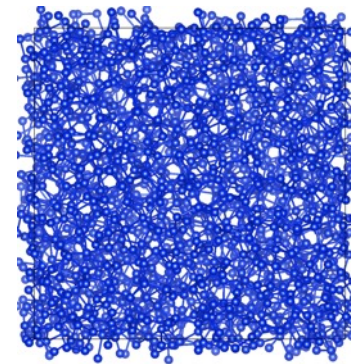


齊藤雄太、まてりあ、59, 387 (2020)

→相変化メモリへの応用

結晶と異なる物性

アモルファスSi



吸光係数



バンドギャップ



移動度



キャリア寿命



→太陽電池への応用

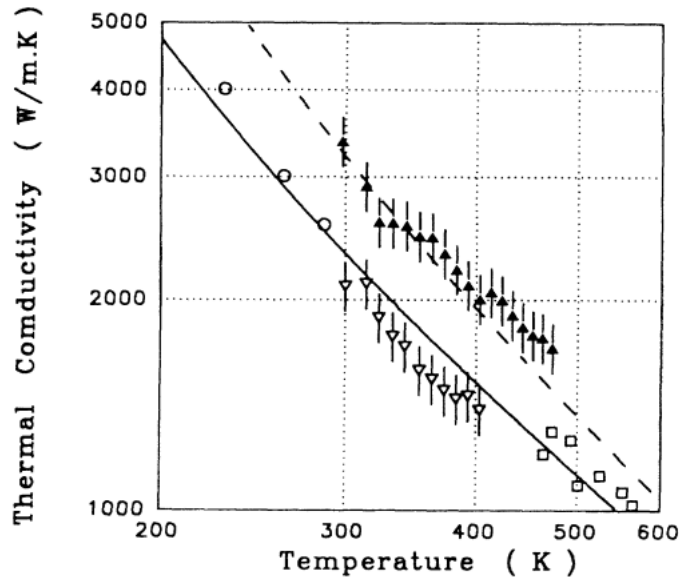
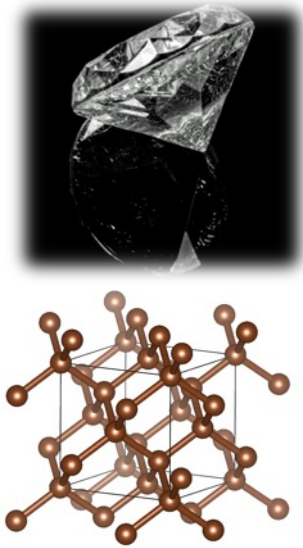
代表的な物性の一つ：熱伝導率

アモルファスと熱伝導率

構成元素が同じでもアモルファスは結晶より劇的に低い熱伝導率を持つ

crystal (diamond)

< 2000 W/m.K @300K

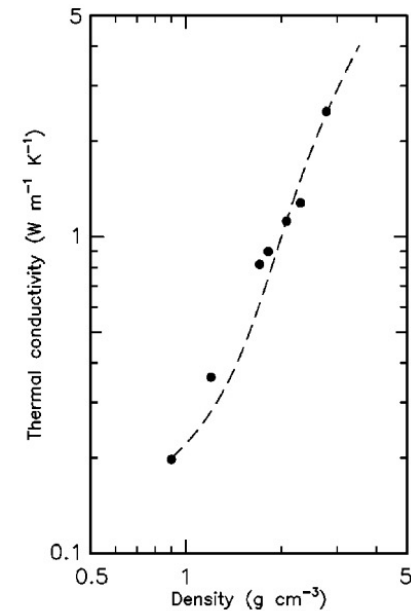
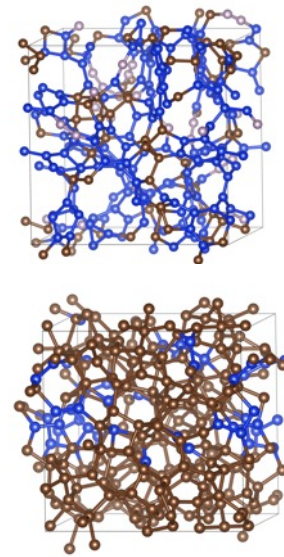


D. G. Onn et al., Phys. Rev. Lett. 68, 2806 (1992)

熱はフォノンが運ぶ

amorphous carbon

0.2 ~ 5 W/m.K @300K



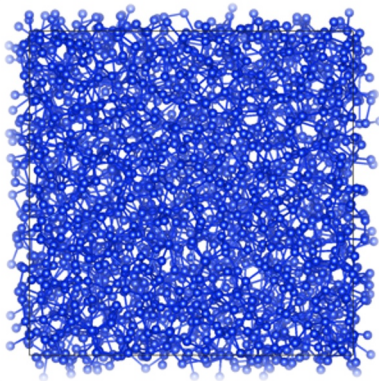
A. J. Bullen et al., J. Appl. Phys. 88, 6317 (2000)

格子振動が乱れで散乱され、平均自由行程が短い
フォノンというよりは格子振動の拡散(diffuson)

アモルファスの構造と熱伝導率の相関

アモルファス内の乱れで格子振動が散乱されることが熱伝導率低下の要因

- どんな局所構造が熱伝導率の低下に最も効くのか？
- **中距離秩序**の影響は？

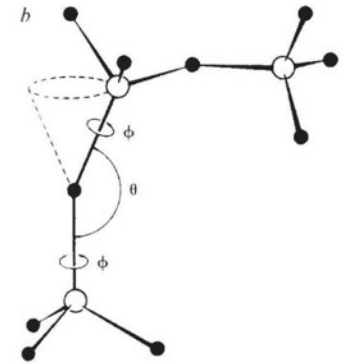


アモルファス構造は完全にランダムな配置ではない

原子間相互作用

→二面角などにも一定の制限が加わる

最近接原子間距離を超えた長さスケールでの
原子位置の相関: 中距離秩序(MRO)



S. R. Elliott, Nature, 354, 445 (1991)

中距離秩序、局所構造、熱伝導率の相関を解明する理論手法開発

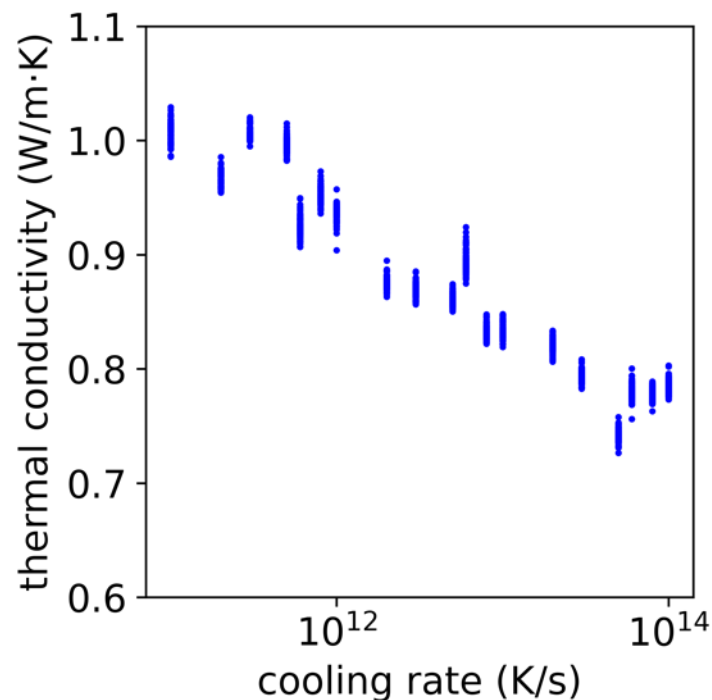
アモルファス構造の特徴をよく抽出する技術が必要

熱伝導率と作成方法の相関

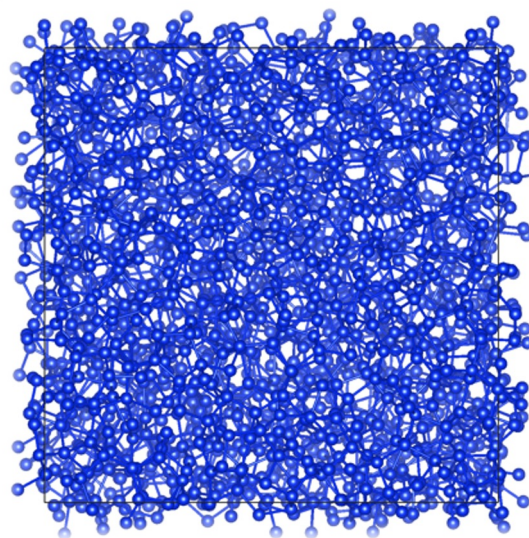
アモルファスSiにおける熱伝導率

古典MDによる溶融急冷法による構造作成
+ Allen-Feldman理論での熱伝導率シミュレーション

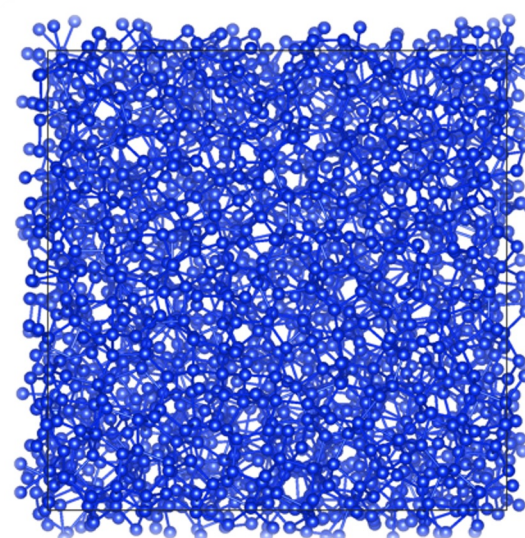
溶融急冷法での冷却レートに
熱伝導率が依存



10¹⁴ K/s で作成



10¹¹ K/s で作成



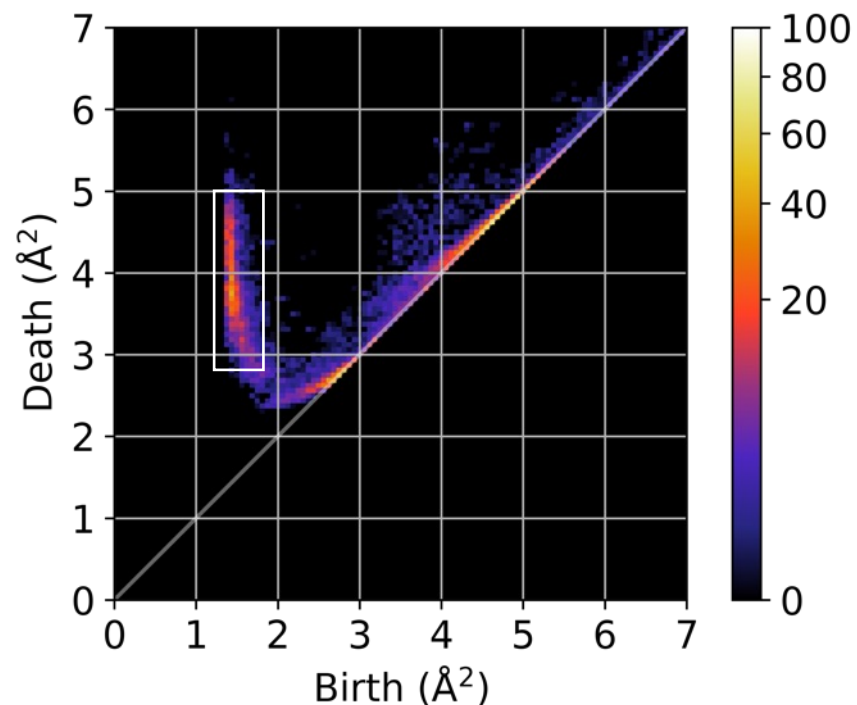
冷却レートを変えて
作成した構造の違いは
一見してもわからないが、
PHのデータを構造の特徴
づけに使えないか??

トポロジカルデータ解析+物性シミュレーション+機械学習を組み合わせ、
局所構造と中距離秩序と熱伝導率の関係を明らかにできる??

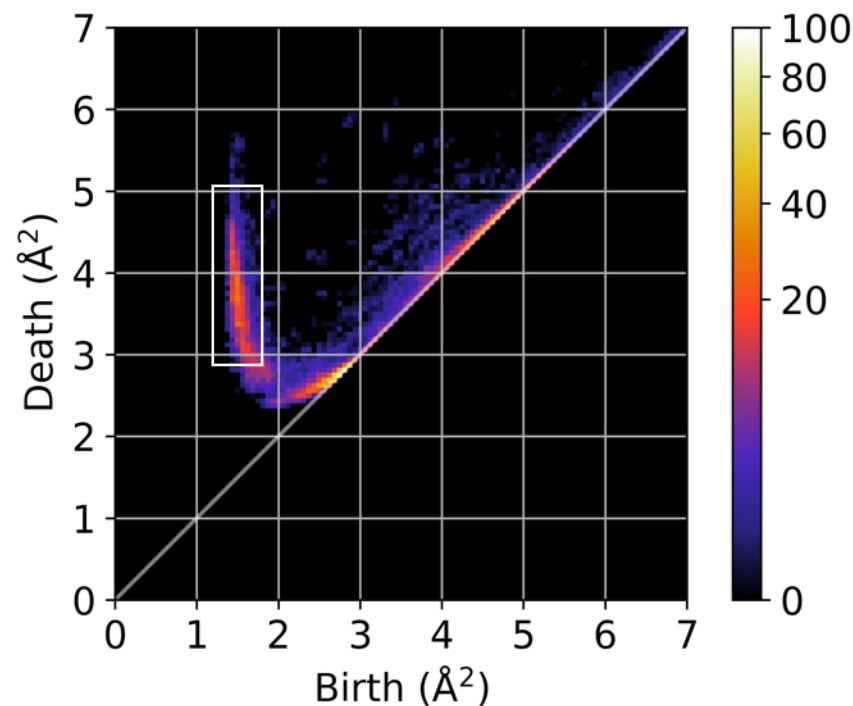
パーシステンス図に現れる違い

1st persistence diagram by HomCloud

10^{11} K/s で作成



10^{14} K/s で作成

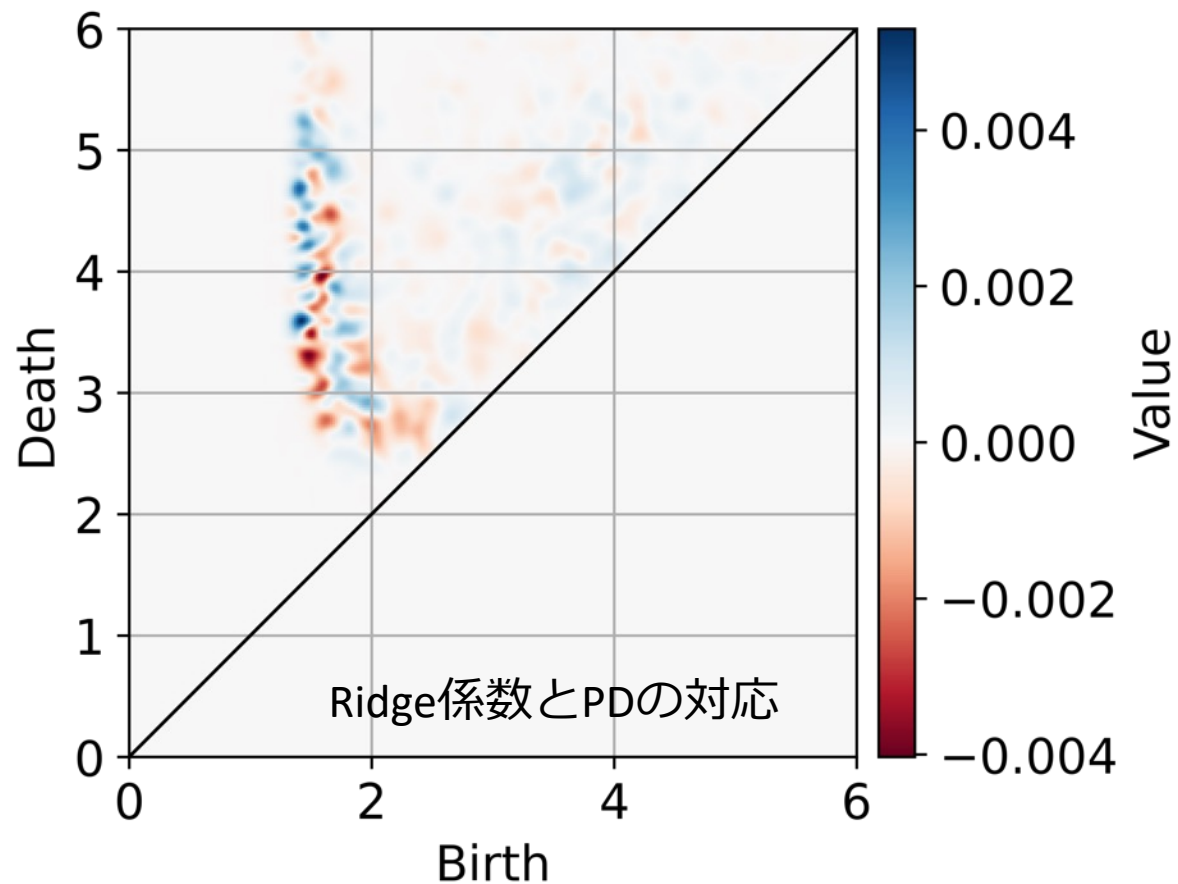
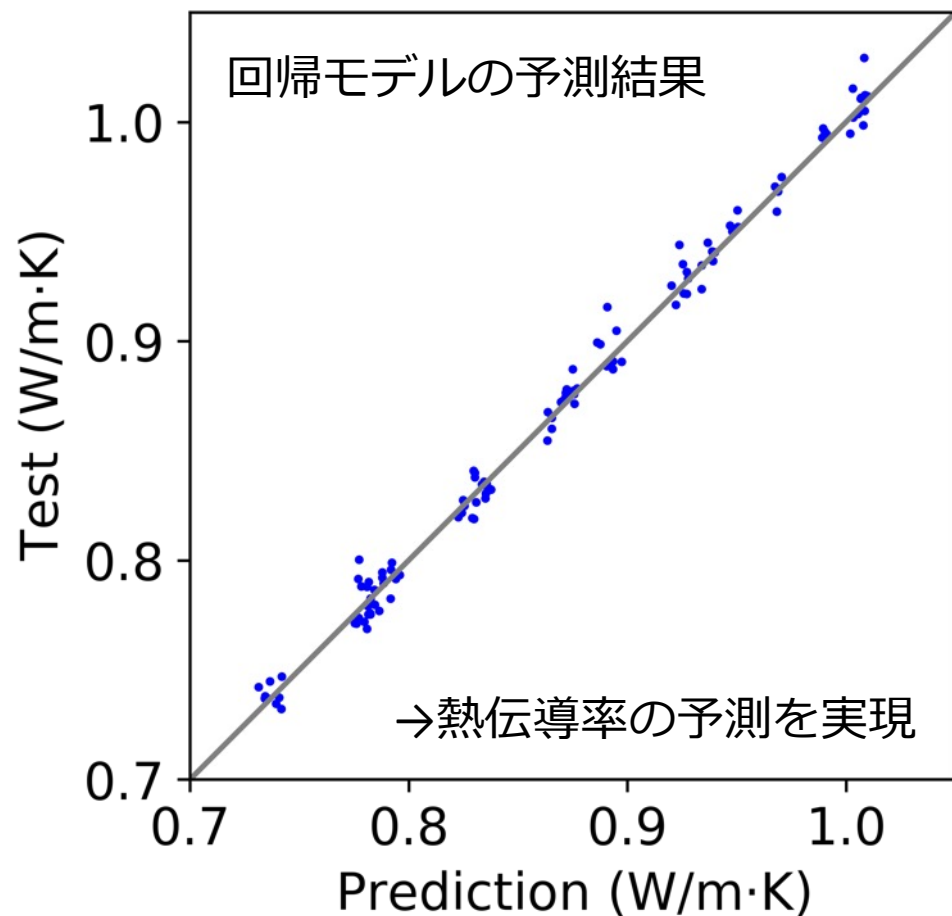


Birth-death ペアの分布には確かに違いが現れる

予測等に結びつけるには？ →ベクトル化 (Persistent Image)

Ridge回帰の結果

作成したベクトルデータを用いて線形回帰モデルを作成（Ridge回帰）



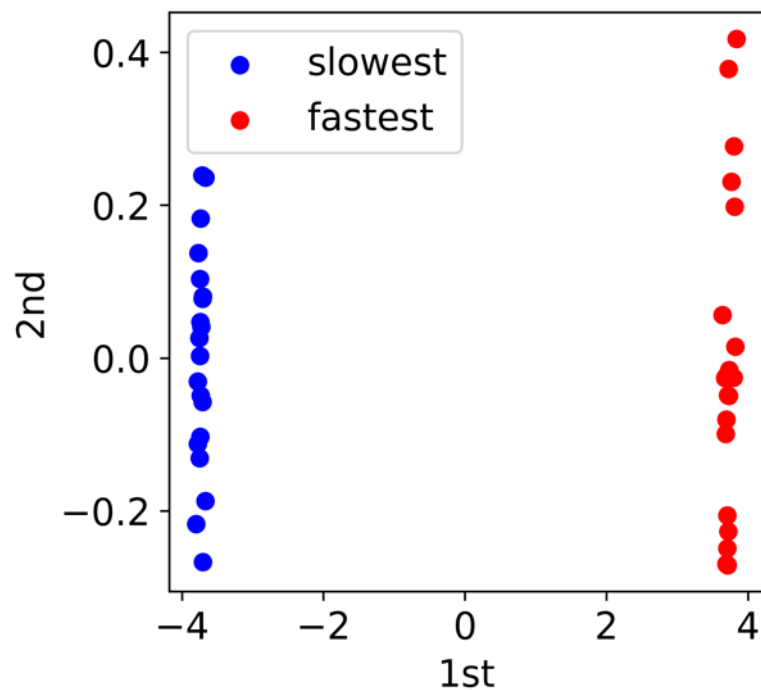
どこの構造が熱伝導率を決めているか→主成分分析で判定

主成分分析 + 逆解析

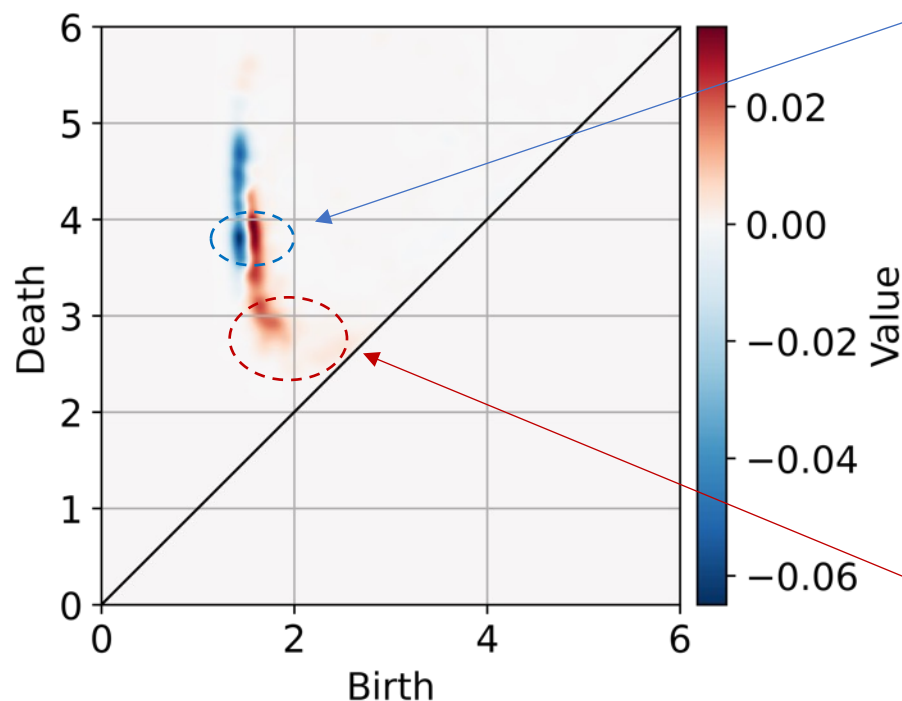
比較的高い熱伝導率になる 10^{11} K/sで生成した構造と、低い熱伝導率になる 10^{14} K/sで生成した構造を典型的な2パターンとして分類問題に焼き直す。

Slowest $\rightarrow 10^{11}$ K/s fastest $\rightarrow 10^{14}$ K/s

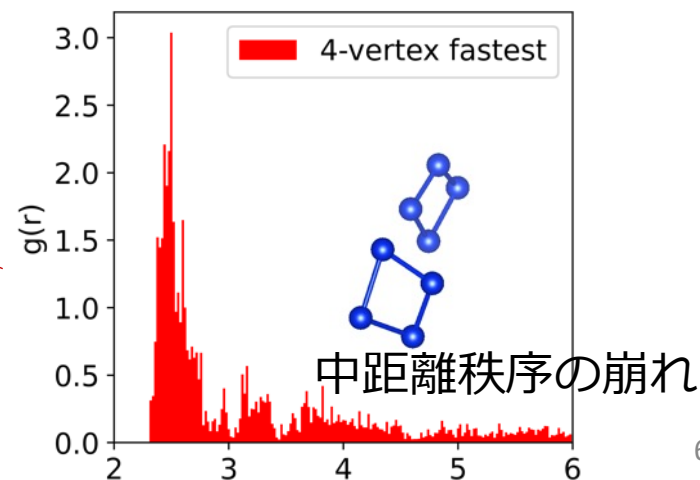
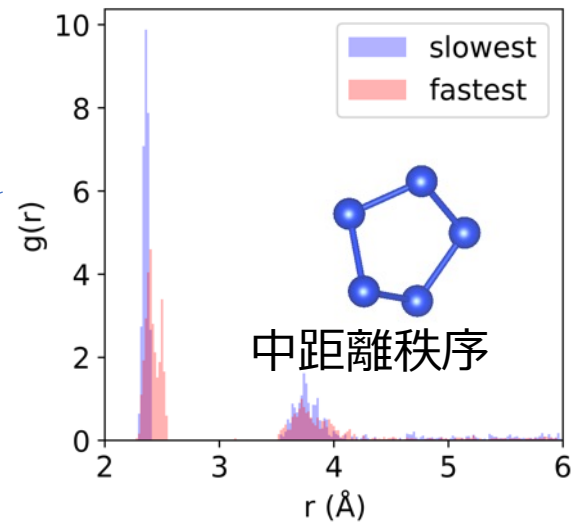
主成分分析



第一主成分の係数から構築したPD



Optimal volume



PHで見つけた構造が熱伝導率に関与するしくみ

熱伝導の理論式

Allen-Feldman理論

量子的な定義

$$\kappa_{\mu} = \frac{V}{T} \int_0^{\beta} d\lambda \int_0^{\infty} dt \langle S_{\mu}(-i\hbar\lambda) S_{\mu}(t) \rangle .$$

S : 熱流演算子

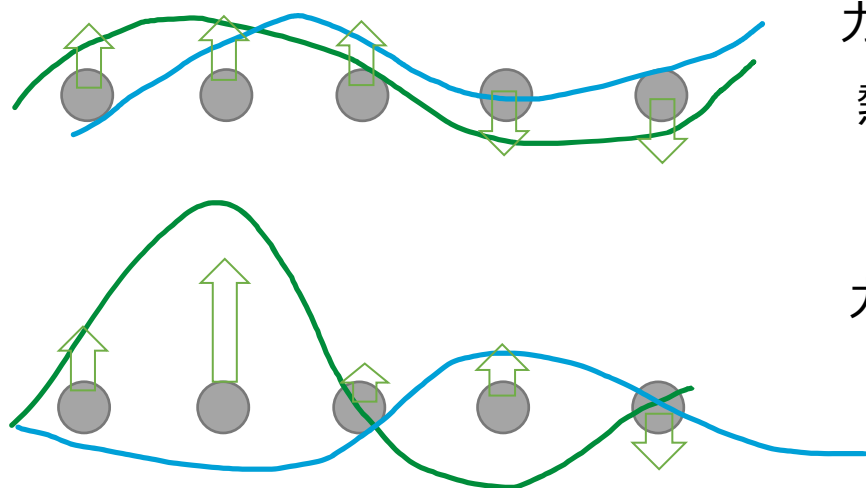
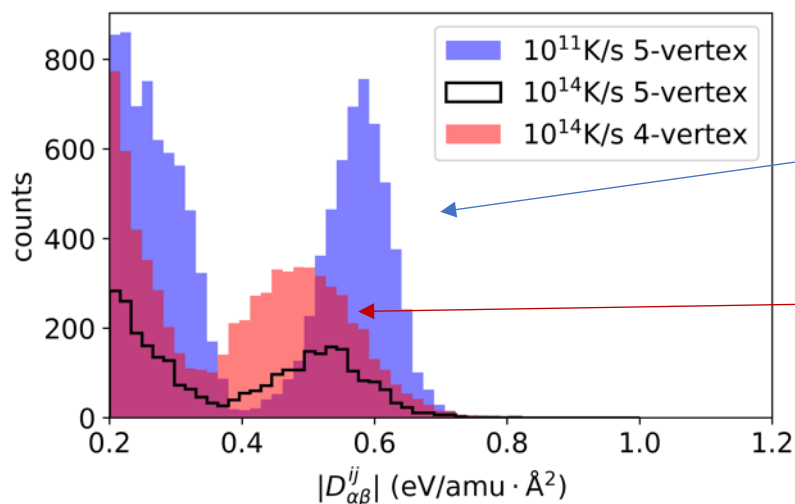
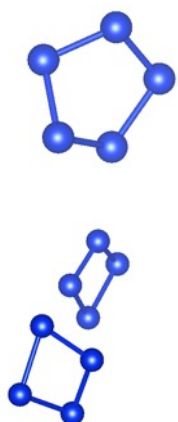
調和近似下での熱流演算子の定義

$$S = \sum_{mn} s_{mn} a_m^{\dagger} a_n , \quad s_{mn} = \frac{\hbar}{2V} V_{mn} (\omega_m + \omega_n) ,$$

$$V_{mn} = \frac{i}{2\sqrt{\omega_m \omega_n}} \sum_{\alpha\beta} \sum_{\gamma\gamma'} (R_{\gamma} - R_{\gamma'}) D_{\alpha\beta}^{\gamma\gamma'} e^{\alpha}(\gamma, m) e^{\beta}(\gamma', n) .$$

P.B. Allen and J.L. Feldman, Phys. Rev. B **48**, 12581 (1993)

振動の固有ベクトルのオーバーラップが重要



力の定数が均一

熱伝導 $\uparrow$

力の定数が不均一

熱伝導 $\downarrow$

より広汎な物性予測にむけて -機械学習ポテンシャルへの応用

共同研究者

岡山大学 大林一平
東京大学 清水康司・渡邊聡

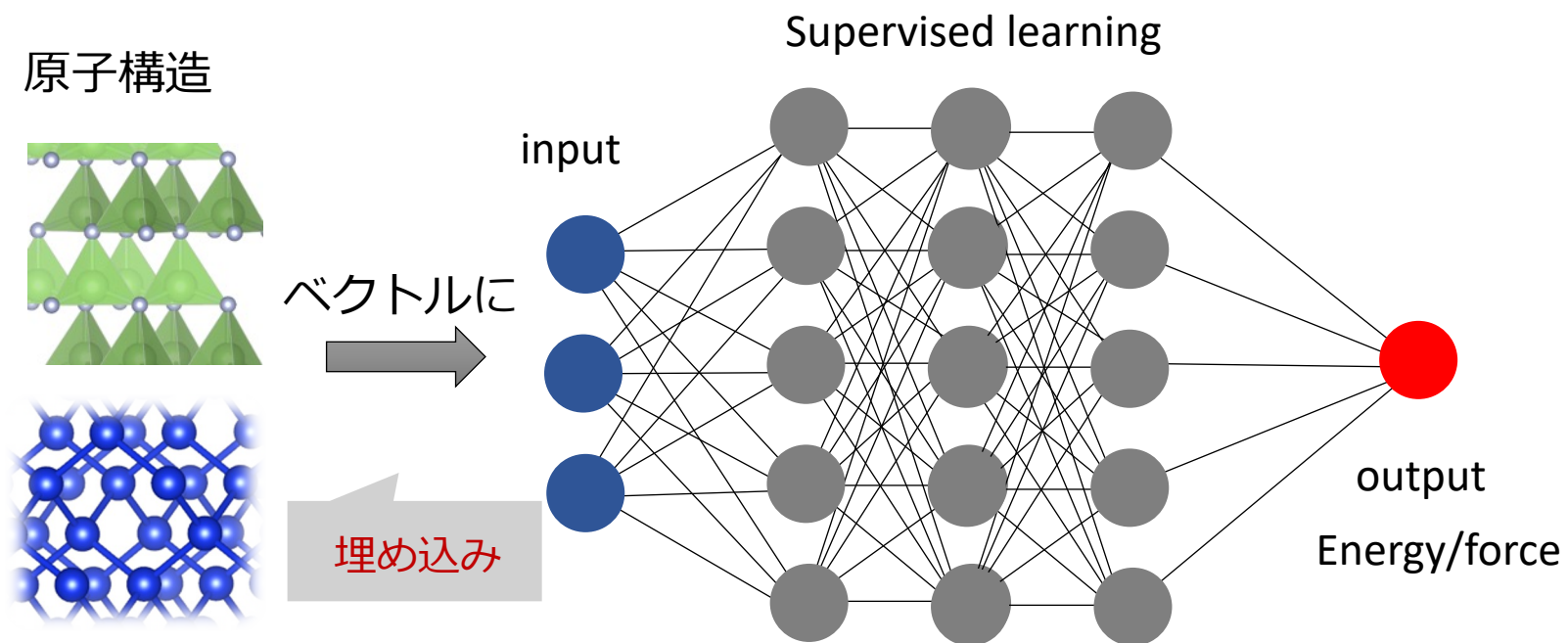
E. Minamitani et al., J. Chem. Phys. 159, 084101 (2023)



機械学習ポテンシャル

量子力学に基づいた精密な第一原理計算→計算コストが高い

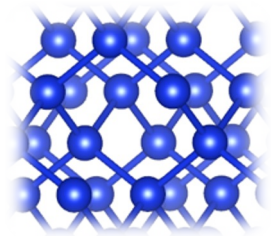
構造の情報を与えると、第一原理計算と同様のエネルギーや力を返すような機械学習モデル
→機械学習ポテンシャル



Jörg Behler and Michele Parrinello, Phys. Rev. Lett. **98**, 146401(2007)

原子座標の埋め込み

直交座標そのままでは、並進・回転・同種原子の入れ替えに対する不変性を保てない



$$f(x_1, x_2, \dots, x_n)$$



$$\begin{pmatrix} 0.5 \\ 0.9 \\ \vdots \\ 0.2 \end{pmatrix}$$

機械学習ポテンシャルが対称操作に対して不変になるよう、原子座標を適切に埋め込む必要がある

既存の方法：対称性関数 (ACSF) , SOAP

多数のハイパーパラメータ

$$G_i^2 = \sum_{j \neq i} e^{-\eta(R_{ij}-R_s)^2} f_c(R_{ij})$$
$$G_i^4 = 2^{1-\zeta} \sum_{j \neq i, k \neq i, j} (1 + \lambda \cos \theta_{ijk}) e^{-\eta(R_{ij}^2 + R_{ik}^2)} f_c(R_{ij}) \cdot f_c(R_{ik})$$

Atom centered symmetry function(ACSF)

J. Behler, J. Chem. Phys. 134, 074106 (2011)

$$p(r_i)_{nn'l}^{Z_1 Z_2} = \pi \sqrt{\frac{8}{2l+1}} \sum_m c_{nlm}^{Z_1}(r_i)^* c_{n'lm}^{Z_2}(r_i)$$

Smooth overlap of atomic positions (SOAP)

A. P. Bartók, R. Kondor and G. Csányi, Phys. Rev. B, 87, 184115 (2013)

ハイパーパラメータをチューニングしなくては、構造特徴をうまく記述できない

最適なハイパーパラメータは系によって異なる...

原子座標の埋め込み

グラフニューラルネットワーク(GNN)の応用

Message Passing Neural Network

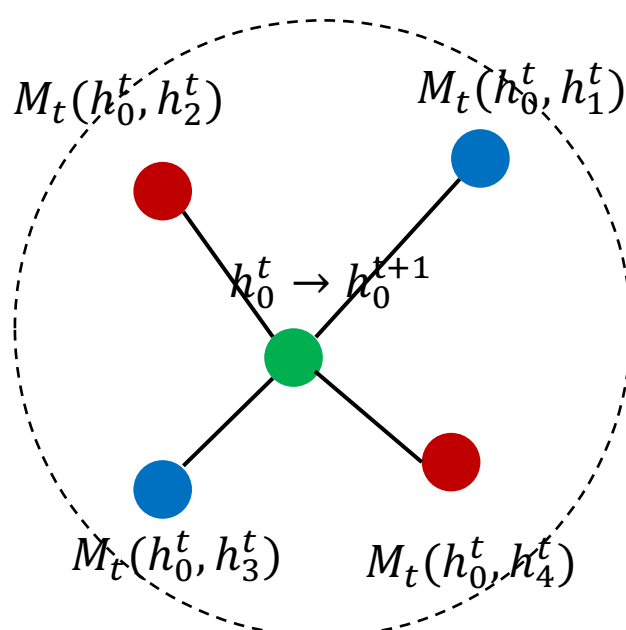
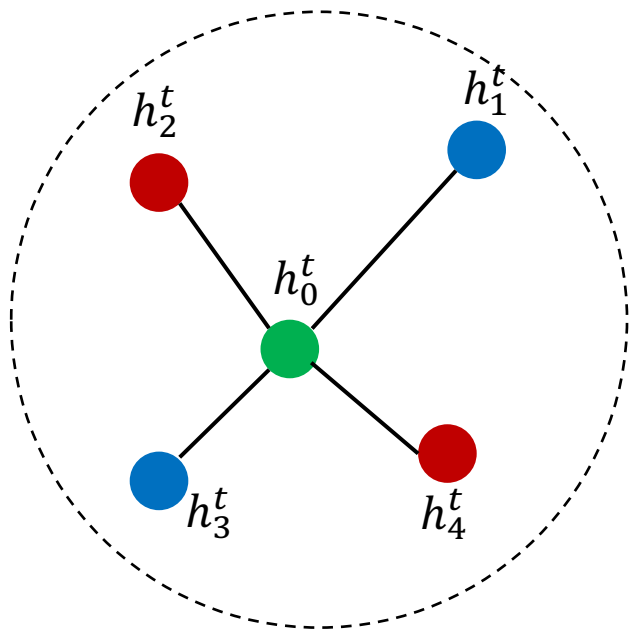
J. Gilmer et al., Arxiv:1704.01212v2

$$m_v^{t+1} = \sum_{w \in N(v)} M_t(h_v^t, h_w^t, e_{vw})$$

近傍ノード、中心ノード、エッジの情報を使って「メッセージ」を作成

$$h_v^{t+1} = U_t(h_v^t, m_v^{t+1})$$

「メッセージ」によって中心ノードを更新



近さ遠さ・原子の種別の情報が自動的に
取り込まれたノード特徴量を作成できる

Schnet Kristof T. Schütt et al., Arxiv: 1706. 08566

Dimenet J. Gasteiger et al., Arxiv: 2003.03123v2

PFP S. Takamoto et al., Nat. Commun., 13, 2991 (2022)

第3の方向性

従来型の記述子

ACSF etc...



わかりやすい数式・シンプルなモデル



多量のハイパーパラメータ

GNN

SchNet, DimeNet, PFP etc...

記述子についてのハイパーパラメータが少ない

モデルの複雑さ・使用メモリが莫大

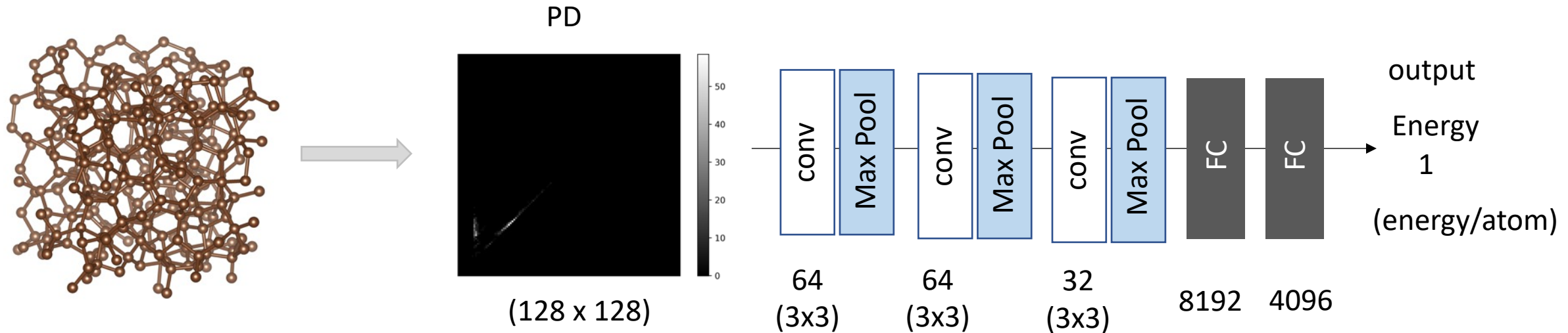
パーシステントホモロジー

パーシステント図は原子座標の並進・回転・入れ替えに対して不変
→よい埋め込みになっているはず

ハイパーパラメータの少なさとモデルのシンプルさを両立できるのでは？

PHを用いた機械学習ポテンシャル

パーシステント図≡ グレースケール画像→畳み込みニューラルネットワーク



Dataset:

Amorphous carbon dataset

Density: 2.0~3.4 g/cm³

Sampling from ab-initio melt-quench simulation (VASP, potential: PAW-LDA)

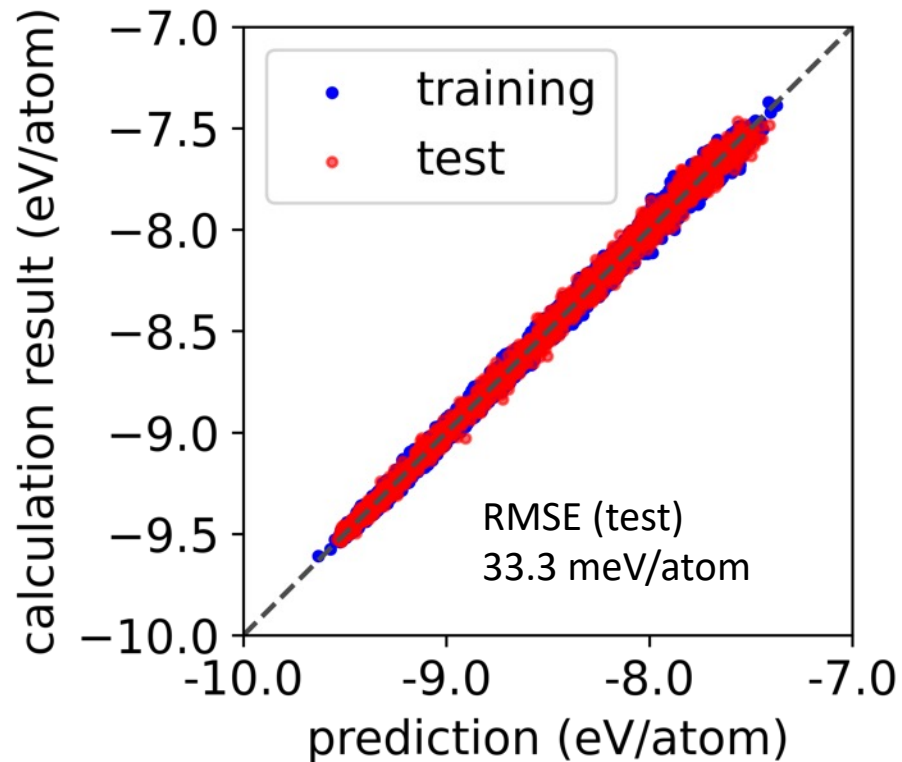
216 C atoms/unitcell

Number of data: 12,960

Including liquid and amorphous configuration

アモルファスカーボンでの結果

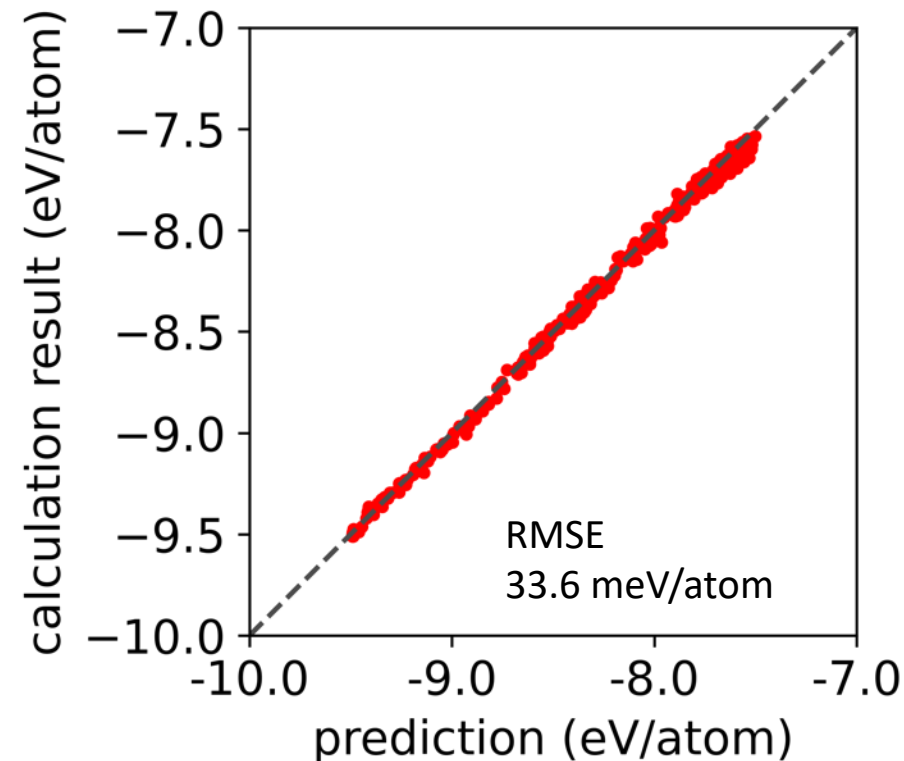
Training/test @ 216-atom systems



系のサイズに関する
スケーラビリティ

学習データとして用いた
系のサイズより大きいも
のでも予測できるか？

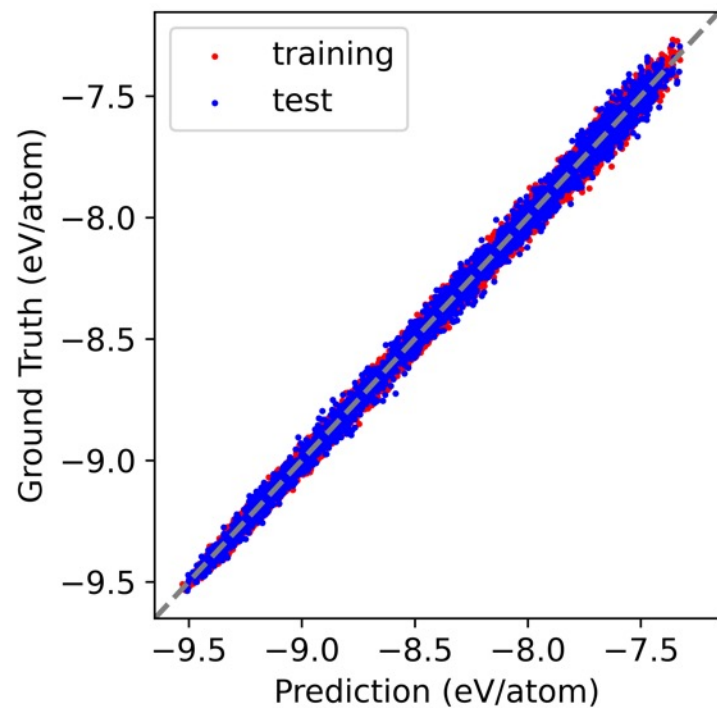
Prediction @ 512-atom systems



パーシステンス図は機械学習ポテンシャル用の新しい構造の記述子になり得る

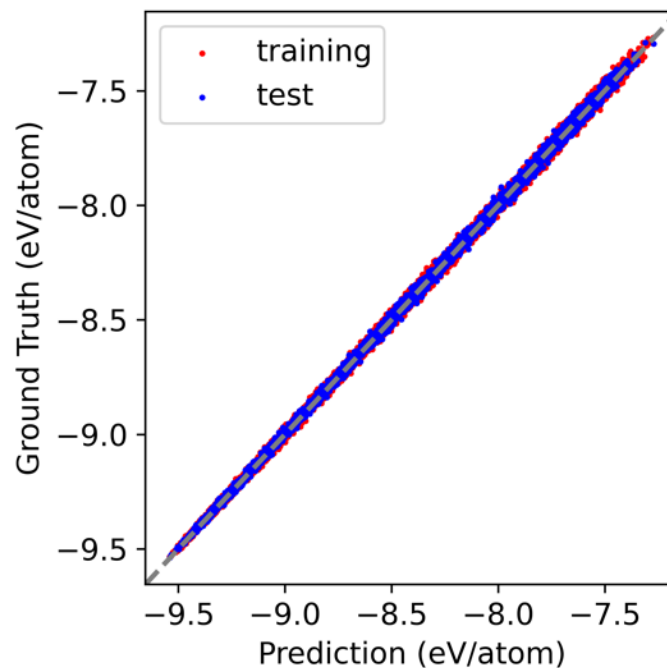
他のモデルとの比較

PH + CNN



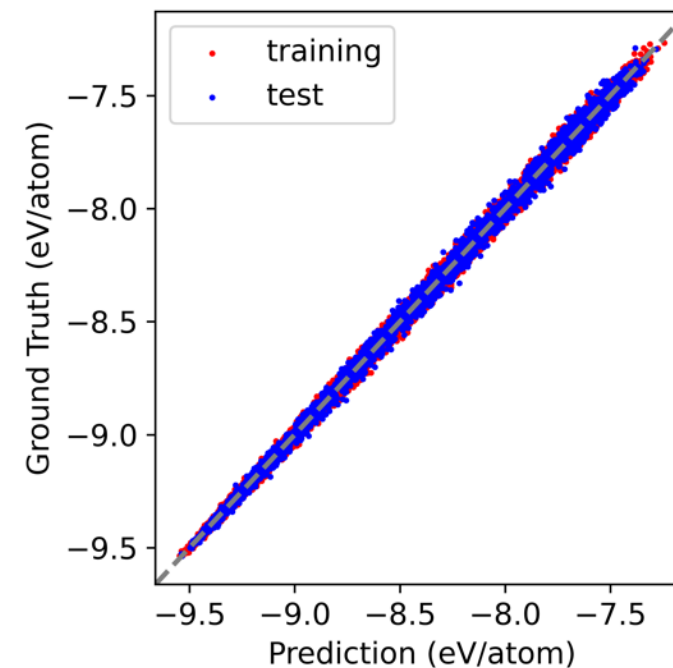
RMSE 40.7 meV/atom

SOAP + NN



RMSE 16.4 meV/atom

SchNet(GNN)



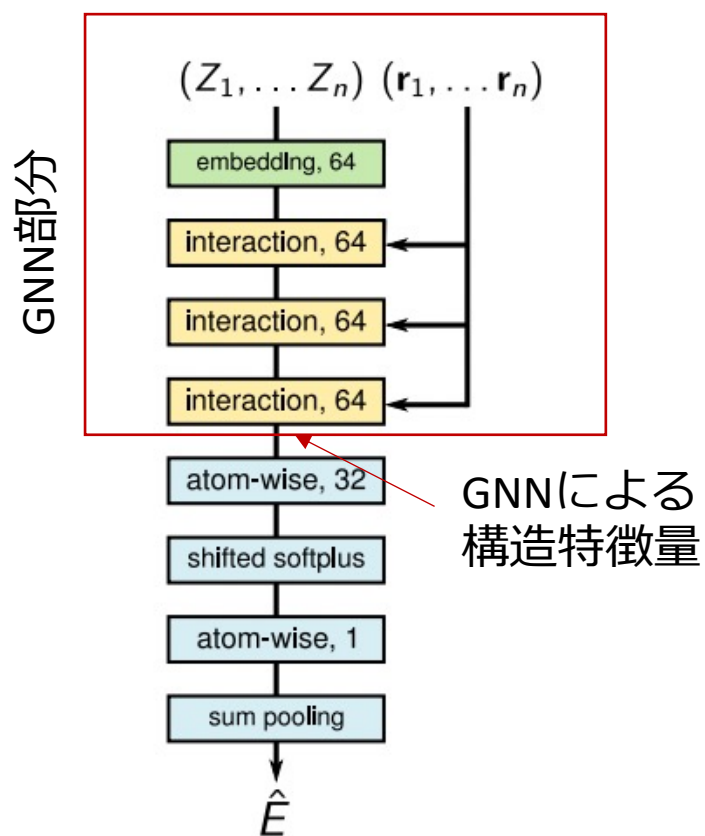
RMSE 30.5 meV/atom

抜群に優れている訳では無いが、既存手法と同程度の精度が出ている

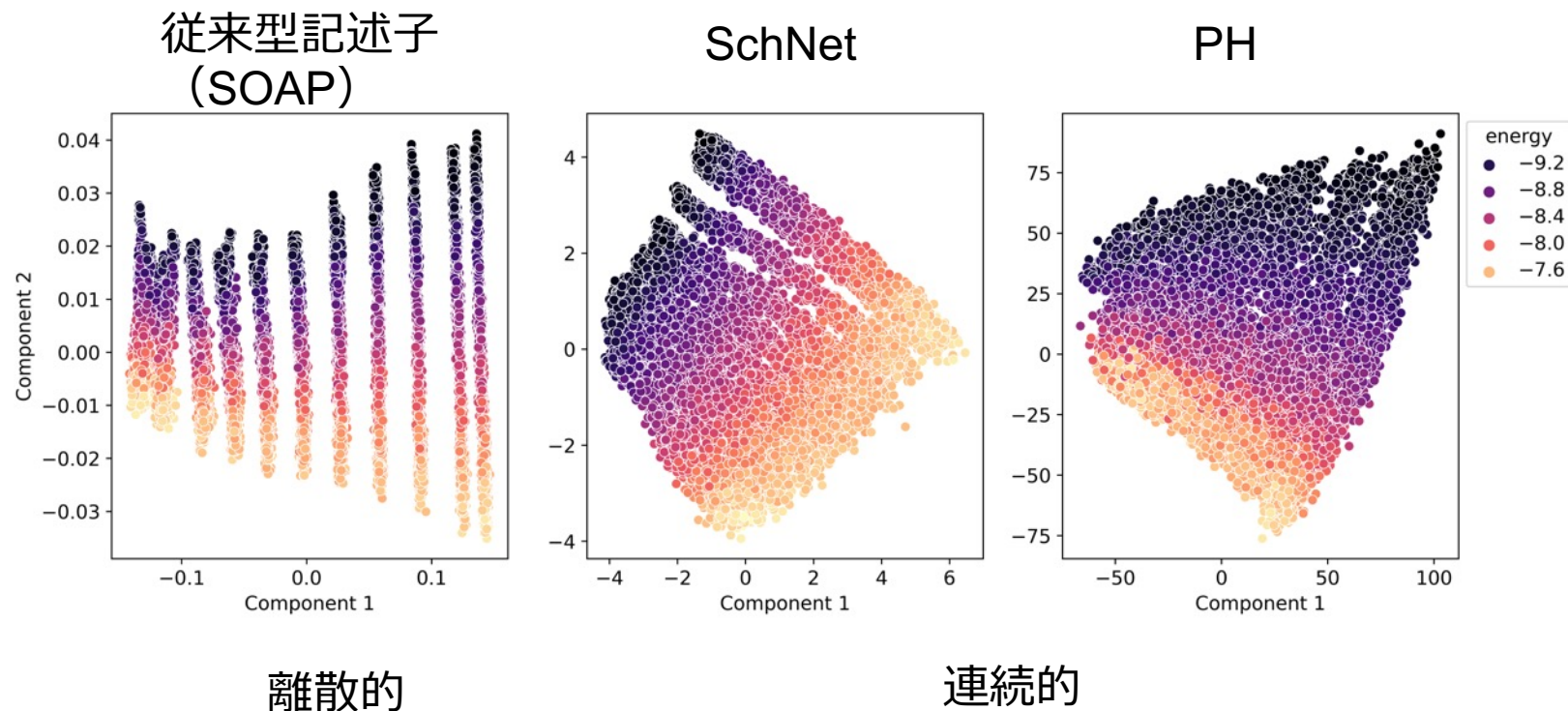
構造特徴の捉え方の比較

グラフニューラルネットワークによる機械学習ポテンシャル

SchNet



構造特徴量を主成分分析にて次元削減して可視化した結果



PHによる構造特徴の捉え方がGNNと類似している可能性

パーシステントホモロジー

数学のホモロジー群がベース
複雑な構造の特徴をうまく抽出できる



乱れた構造の解析
構造と物性の関係

物質・材料科学者にも使いやすいツールが多数整備されている

HomCloud, GUDHI, Ripserなど

注意点

解析結果を解釈するには、物質についてのドメイン知識が必要

なんにでもうまくいく silver bullet というわけではない

「ハンマーを持つ人にはすべてが釘に見える」に気をつけましょう（自戒）

Special thanks to 大林一平先生（岡山大学）

実際のアモルファス物質の構造データをつかって、パーシステントホモロジーの解析と、それから物性予測を行うモデルを作ってみましょう

使用するデータ：アモルファスカーボンデータセット

<https://zenodo.org/records/7905585>

Files (258.5 MB) ▼		
Name	Size	 Download all
aCdataset-zenodo.zip md5:f5af013080390cdaa8431d1555d66f80 	258.5 MB	 Preview  Download

↑かなり重いファイルなので、
事前のダウンロードをおすすめします。

この中の、dataset_train_test_split/test.xyz を今回は使います（全体を使うとデータが重すぎるので）

必要な環境

- python仮想環境
 - HomCloud
 - ASE
 - scikit-learn
- Jupyter notebook (VSCodeのJupyter notebookアドオンなど)

コード・使い方の説明一式

https://github.com/eminamitani/spring_school_2024

こちらに、必要な分だけのデータも入れてあるので、これをcloneするのが一番手っ取り早いです