

研究紹介 - 位相的データ 解析について

大林一平 [岡山大Cypher, JSTさきがけ]



岡山大学
OKAYAMA UNIVERSITY



数理構造活用

自己紹介



- 大林一平
- 専門は応用数学, 特に位相的データ解析(TDA)や力学系など
 - 最近はパーシステントホモロジー (PH) がメイン
- メインの所属は岡山大学サイバーフィジカル情報応用研究コア (Cypher)
 - 2021年4月から
 - 兼担: 環境理工学部, 工学部, 環境生命科学研究科

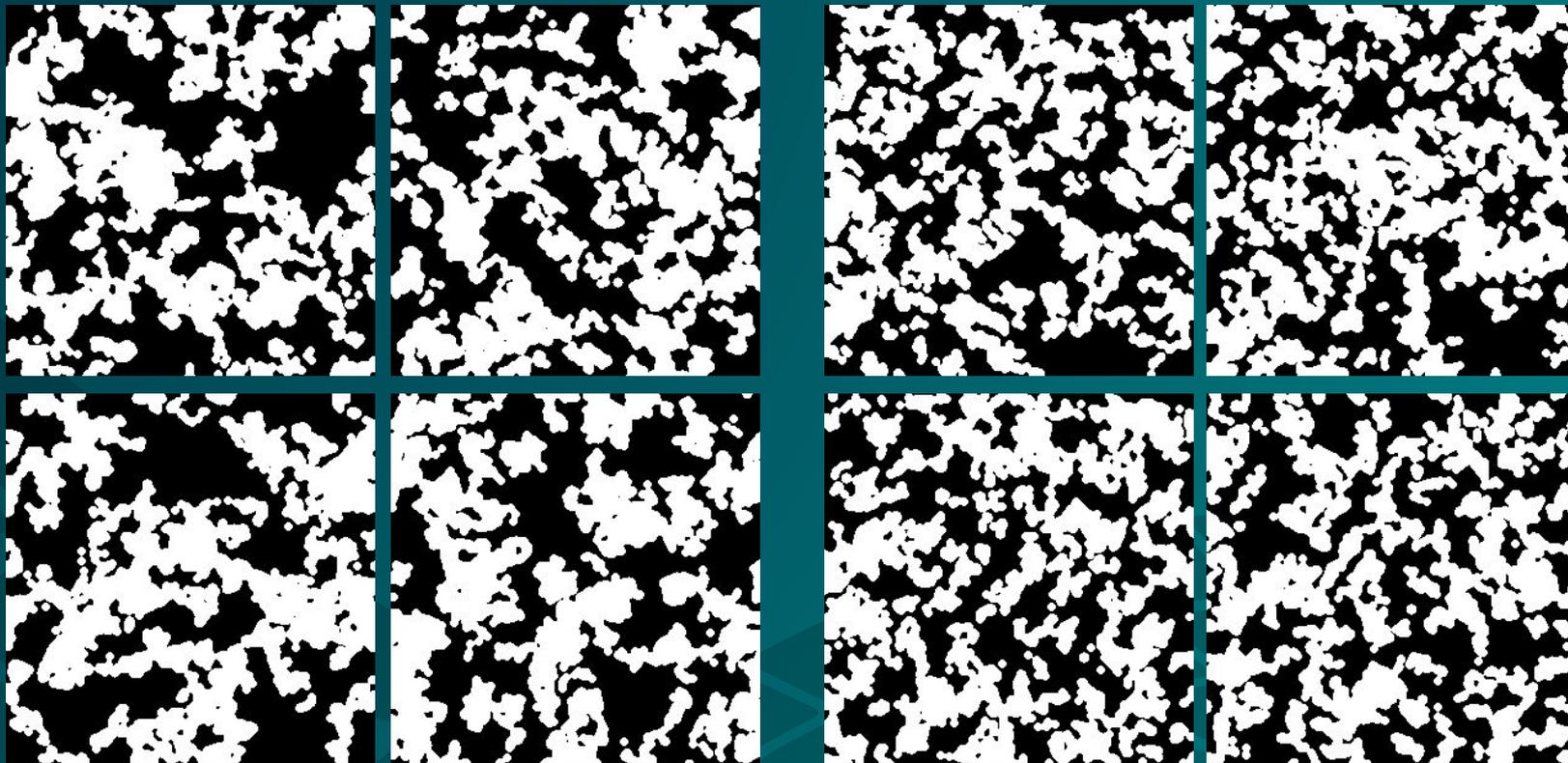
経歴

- 2010 京都大学 博士(理学) - 京大数学教室 (エルゴード理論的な力学系の研究, 理論から精度保証数値計算まで)
- 2010~2015 京大数学教室でポスドク(力学系の二足歩行モデルへの応用など)
- 2015~2018 東北大AIMR助教 (TDAの材料科学への応用など)
- 2018~2021 理研AIP 研究員 (TDAの理論と応用)
- 2021~ 岡山大学 Cypher 教授

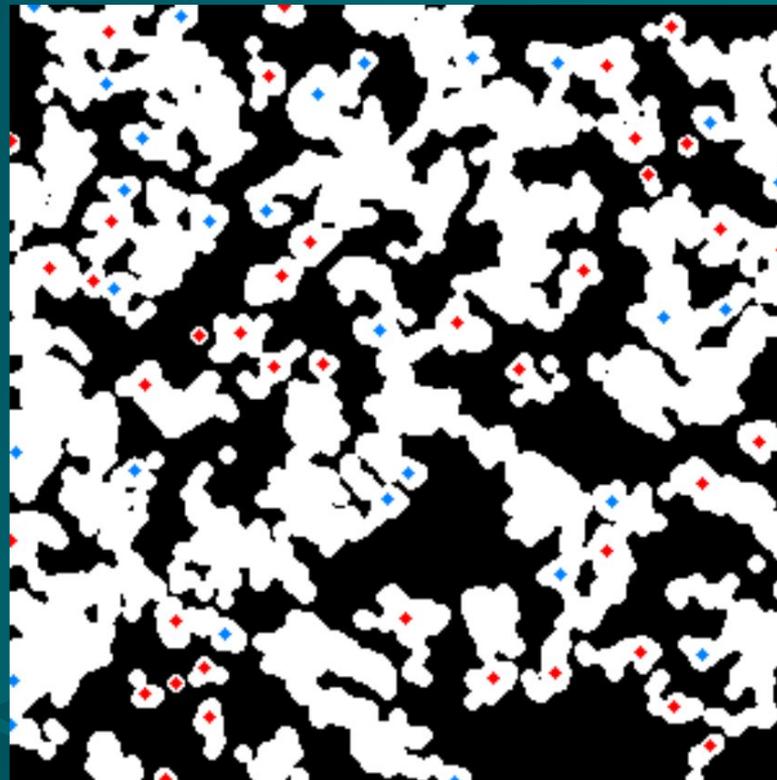
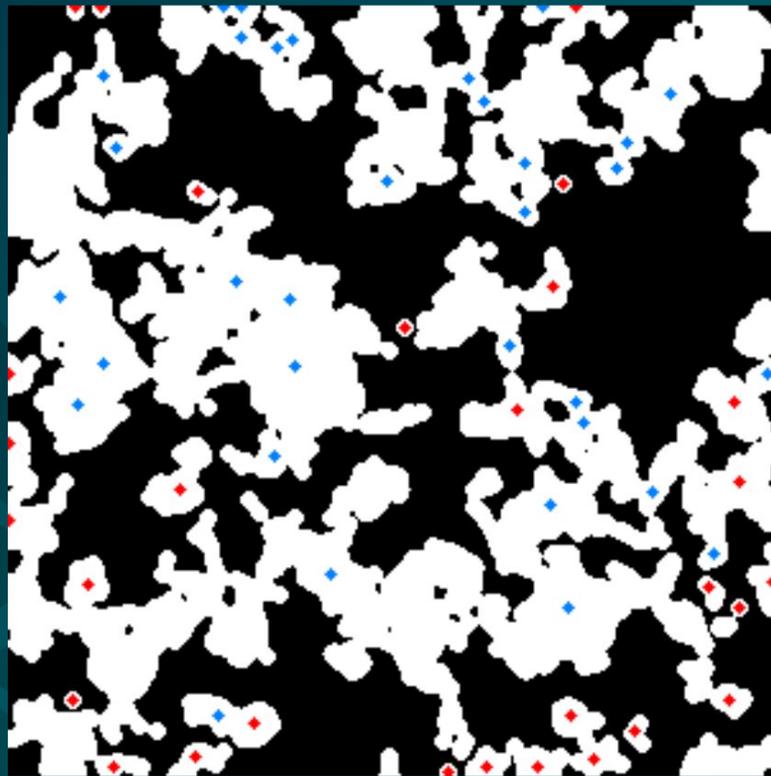
研究概要

- 位相的データ解析, 特にパーシステントホモロジー(PH)
- PHの数学理論
- PHに関するアルゴリズム
- PHと機械学習の組み合わせ
- PHによるデータ解析ソフトウェア, HomCloudの開発
- PHを用いたデータ解析
 - 材料科学, 地質学, 気象学, 医療画像, など

パーシステントホモロジーの例



左右の4ずつの画像の違いは？



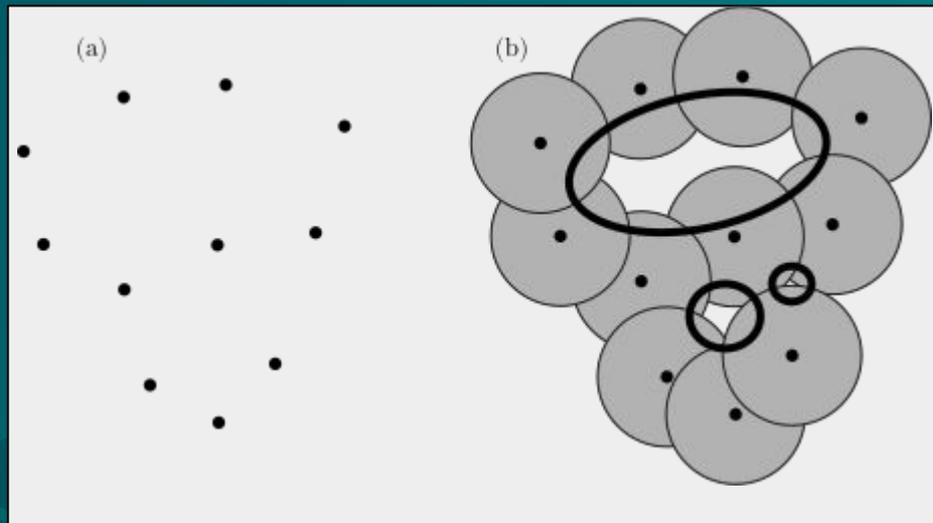
青い点: 左の図に特徴的な構造 / 青い点: 右の図に特徴的な構造 | PHと機械学習の組み合わせで抽出

パーシステントホモロジー

- 位相的データ解析 (Topological Data Analysis, TDA)
 - 数学の「トポロジー」の概念を使ったデータ解析
 - データの「かたち」を定量的に特徴付け
 - 連結成分, リング, 空隙, などを利用
- パーシステントホモロジー (PH) はTDA の中心的道具
 - データのかたちのための良い記述子 (パーシステント図)
 - 21 世紀に急速に発展
 - 数学的な理論やアルゴリズム/ソフトウェア/材料科学やライフサイエンスなどへの応用

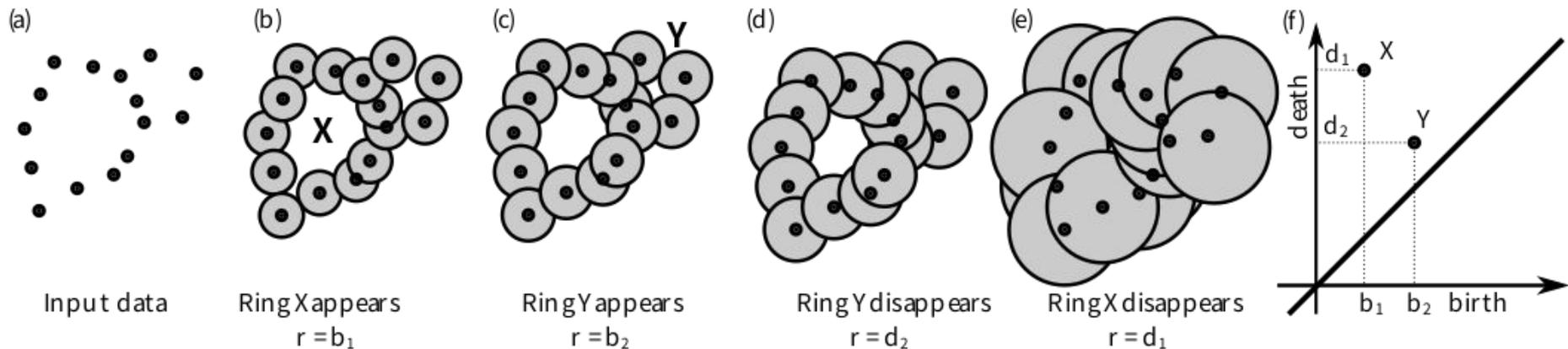
r-Ballモデル

- 図(a)には、リングが2つあるように「見える」
- しかし実際のつながりはない
- そこで、半径 r の円を各点に置いて、構造を作りだす
 - (b)で孔が3つ得られる

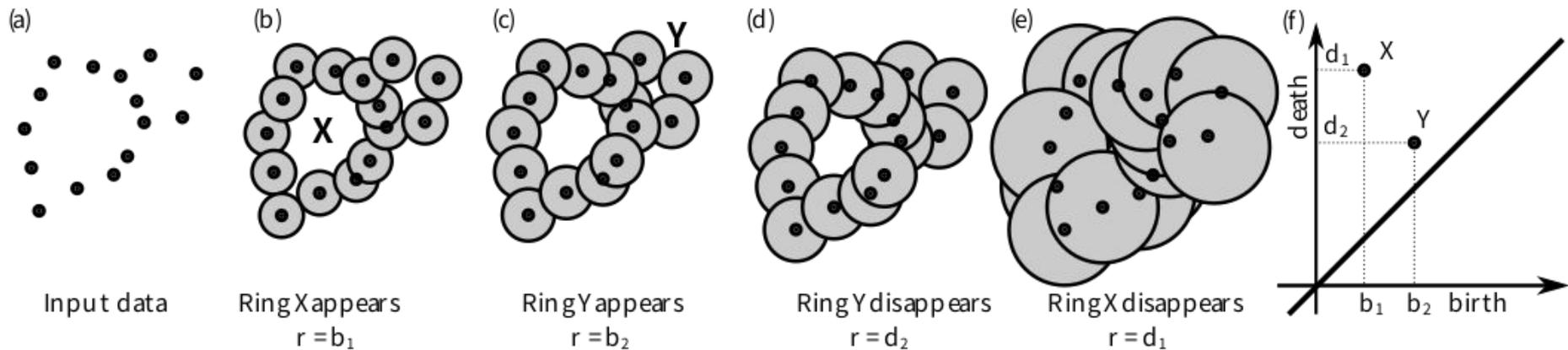


フィルトレーション

半径 r を一つ決めるのではなく、半径を変えていったときの孔の発生消滅を調べるのがPHの中心的アイデア



- PHの理論から, 孔の発生消滅の1対1のペアを作ることができる
- 発生時の半径をbirth time, 消滅時の半径をdeath time, この2つのペアをbirth-death pairと呼ぶ
- birth-death pairの集まりをパーシステント図(PD)と呼び, 散布図や2次元ヒストグラムで可視化する



パーシステントホモロジーについて

- 背景にある数学理論が以下の良い性質を保証している
 - 出力結果(PD)の存在と一意性
 - 入力に対するノイズの出力に対する影響の大きさの上界
- どんな次元のデータにも適用できるが、3次元データが一番活用できそう
- 対角線のそばにあるbirth-death pairは発生してすぐに消滅するので、重要度が低いと考えられる

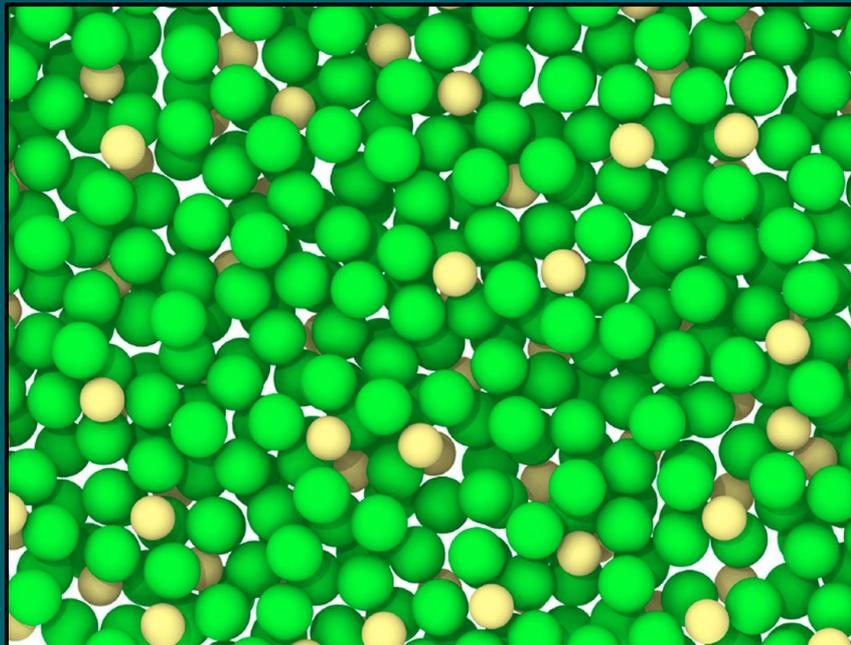
材料科学への応用

A. Hirata, et al, Commun. Mater. 1, 98 (2020).

- 平田(早稲田), 和田(東北), 平岡(京都)との共同研究
- この研究は大木のこれまでの研究の成果を活用した結果
 - 私の研究の全体像を説明するのに都合がよい
- Materials Informaticsと呼ばれる分野

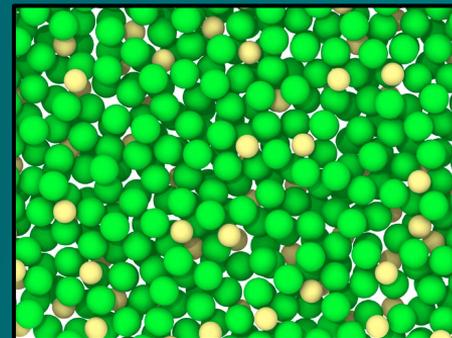
金属ガラス

- ◇ 結晶ではない, 乱れた原子構造を持つ金属
- ◇ 通常の金属とは異なる特性
- ◇ 乱れた原子構造が特性の鍵



Pd-Si ガラスの原子配置, O. Adjaoud and K. Albe, Acta Materialia, 113 284-292 (2016) より引用

問題設定



- 金属ガラスは乱れているが、中距離秩序と呼ばれる何らかの秩序構造を持っていると考えられている
 - 長距離秩序(結晶), 短距離秩序(隣接原子の配置)
- この秩序構造を解明することが最終的な目標
 - この目標はかなりあいまい
- ガラスは高温から急速冷却することで形作られる
 - そこでこの冷却速度を変えたときの構造の差異を調べる
 - 冷却速度が速いほど無秩序性が高くなると考えられている

手法

- 金属ガラスの原子配置は数値計算(分子動力学シミュレーション)で再現
 - $\text{Pd}_{20}\text{Si}_{80}$ という物質
- 原子配置からPDを計算 (Pd原子のみ, Si原子のみ, で計算)
- PDを特徴量, 冷却速度を目的変数とした線形回帰で, 冷却速度の予測に効くbirth-death pairを特定
- 特定したpairから対応する形を抽出
- 抽出した形をさらに解析

大林が開発しているPHによるデータ解析ソフト「HomCloud」を利用

- 金属ガラスの原子配置は数値計算で再現
- 原子配置からPDを計算
- PDを特徴量, 冷却速度を目的変数とした線形回帰で, 冷却速度の予測に効くbirth-death pairを特定
- 特定したpairから対応する形を抽出
- 抽出した形をさらに解析

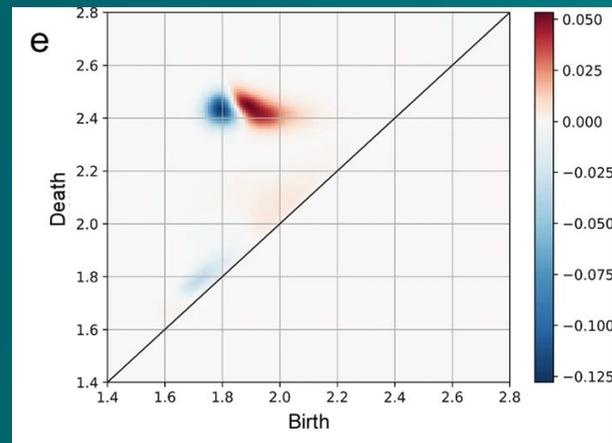
機械学習と位相的データ解析の連携に関する研究成果を活用

大林が開発した数理最適化による形の情報の復元手法を利用

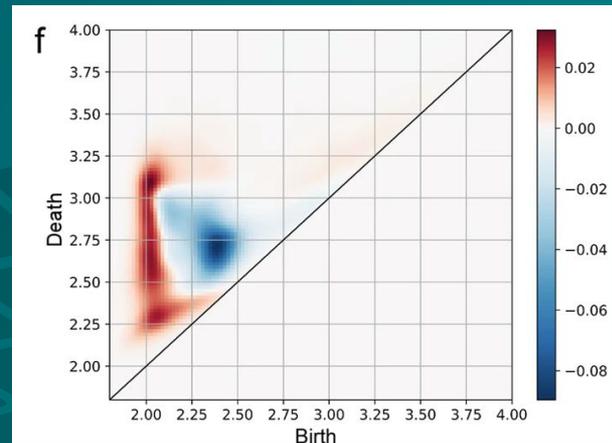
結果

機械学習の利用により, Pdの2次元PDとSiの1次元PDに冷却速度との特徴的関連を発見

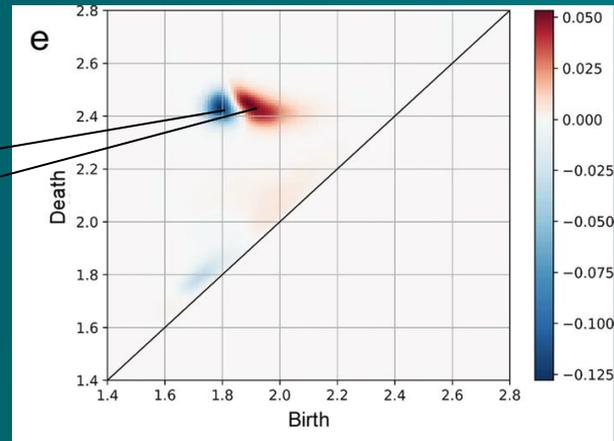
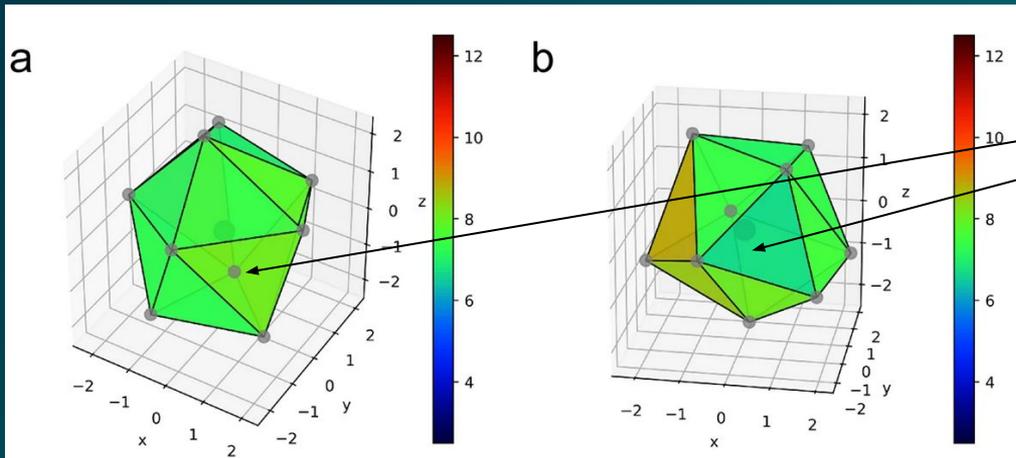
- 赤い領域: 冷却速度に+の相関
- 青い領域: 冷却速度に-の相関



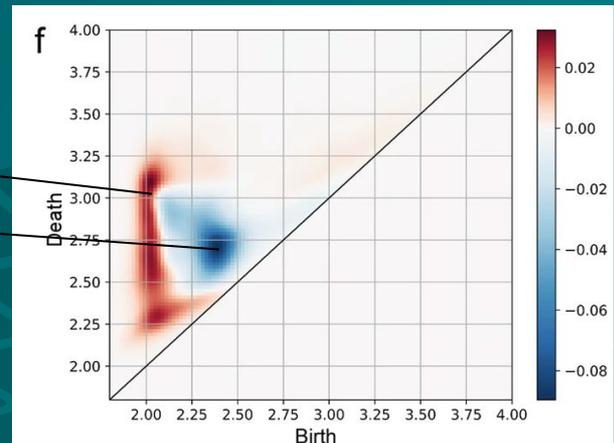
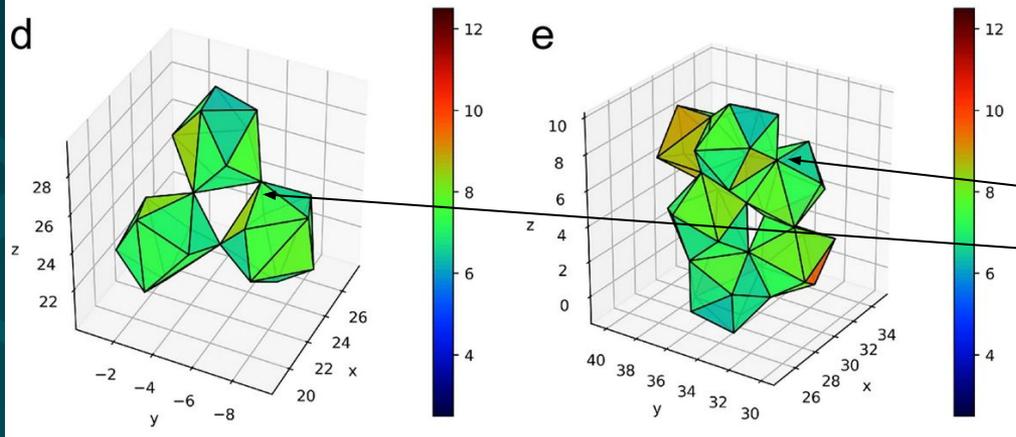
Pdの2次元PD



Siの1次元PD

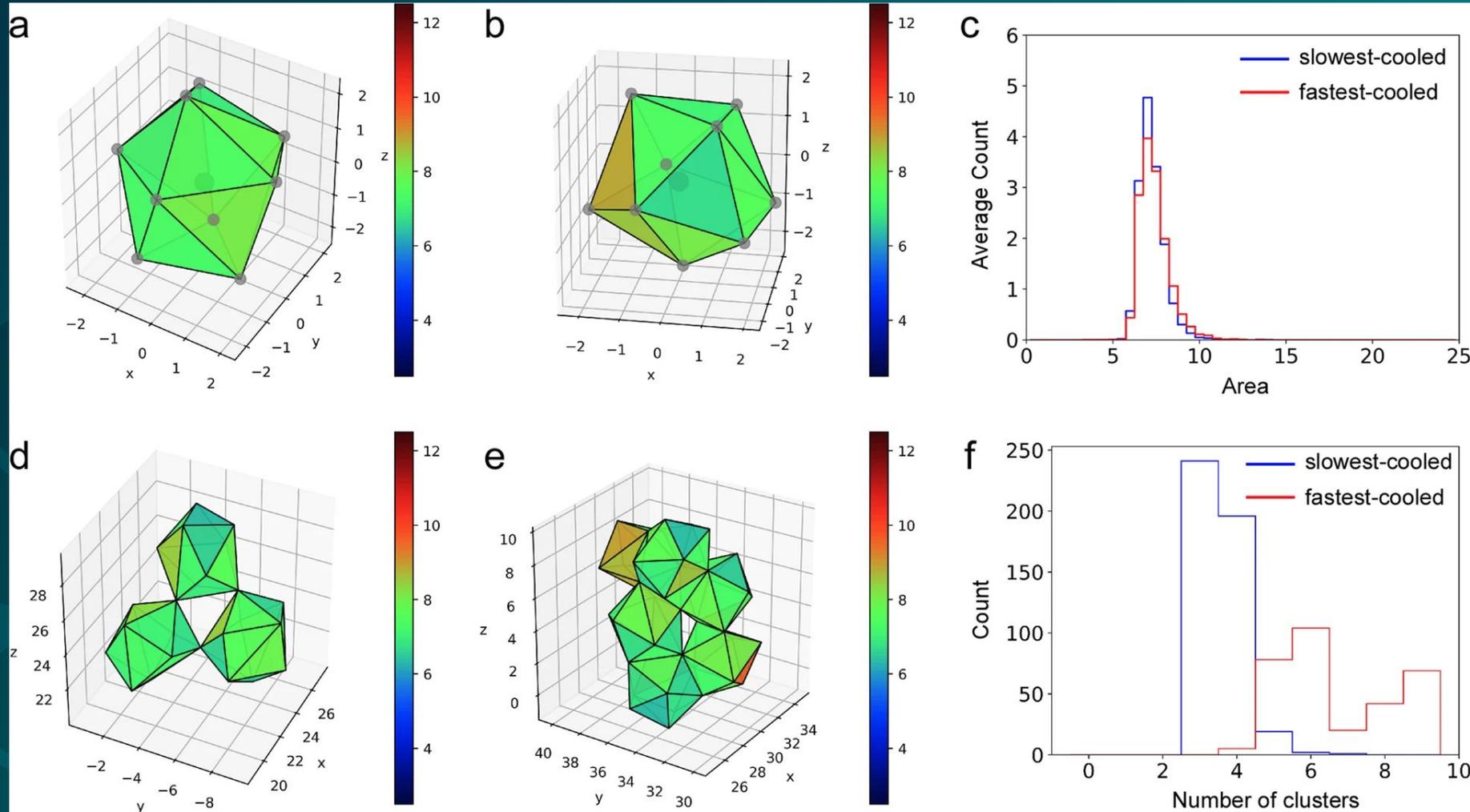


Pdの2次元PD



Siの1次元PD

機械学習で得られたパターンに対応する形を抽出



金属ガラスへの応用-まとめ

- 金属ガラスの冷却速度の変化による構造の違いをPHと機械学習の組み合わせ, 構造復元手法(逆解析と呼んでいる)で特定した
- この構造の違いを調べることで, 冷却速度の変化による無秩序性の変化の説明に成功した

PHが得意とするデータ

- リング構造や空洞構造を特徴付けに使うので、そういった構造の重要性に関する予想/仮説があるのが良い
 - ガラスでは「ネットワーク構造」「パッキング構造」が重要と言われている
 - 特に高次元データは、この仮説が重要
- 3次元空間データは人間による解釈のしやすさ、データ解析の難しさ、のバランスで使いやすい

まとめ

- パーシステントホモロジーはデータの形を定量的に特徴付けることを可能とする
- ポイントクラウド, 画像などに適用できる
- 機械学習の手法との組み合わせも有用
- HomCloud
 - 逆解析などの特徴的機能
- 材料科学を始め, 様々な応用が進んでいる
- 共同研究等のご相談があれば歓迎いたします