パーシステントホモロジーとその応用

平岡　裕章
Yasuaki Hiraoka

東北大学　材料科学高等研究所（AIMR）
理化学研究所AIPセンター
物質・材料研究機構（NIMS）トポロジカル解析グループ

JST CREST
内閣府 SIP革新的構造材料
JST イノベーションハブMI^2I
NEDO 超超プロジェクト

今日と明日の内容

平岡裕章（東北大）：パーシステントホモロジーとその応用

浅芝秀人（静岡大）：クイバーの表現論とパーシステントホモロジー

白井朋之（九州大）：確率論とパーシステントホモロジー

福水健次（統数研）：パーシステント図に対する統計的機械学習

大林一平（東北大）：位相的データ解析ソフトウェアHomCloudの
紹介およびパーシステント図の逆問題について

Content

1. 背景
2. データの幾何モデル
3. パーシステントホモロジーの導入
4. 代数的研究（クイバーの表現論）
5. 安定性定理
6. 計算ソフトウェア
7. 応用（材料科学）

Content

1. 背景

2. データの幾何モデル

3. パーシステントホモロジーの導入

4. 代数的研究（クイバーの表現論）

5. 安定性定理

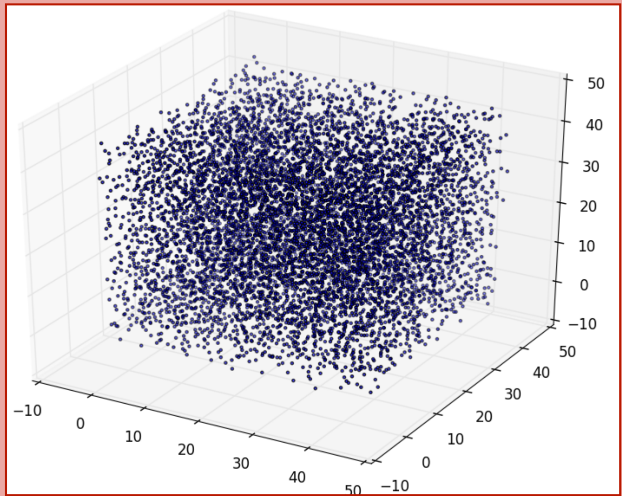
6. 計算ソフトウェア

7. 応用（材料科学）

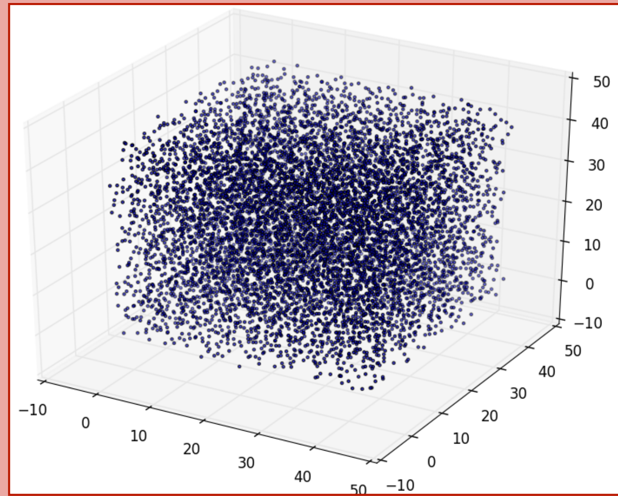
Background : Shape of Data

Data-driven science studies potential values of big and complicated data by machine learning and AI

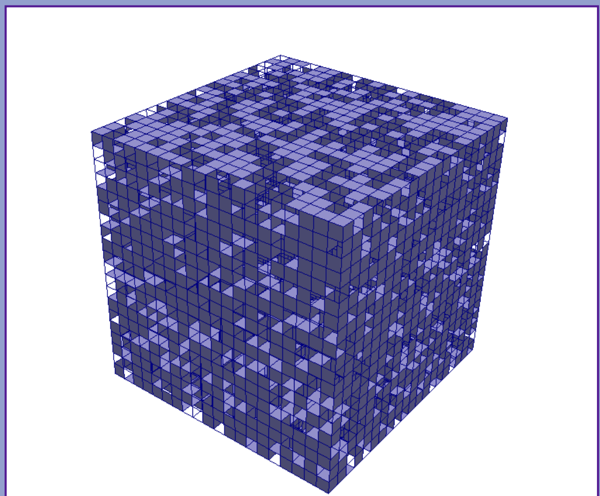
Point cloud data



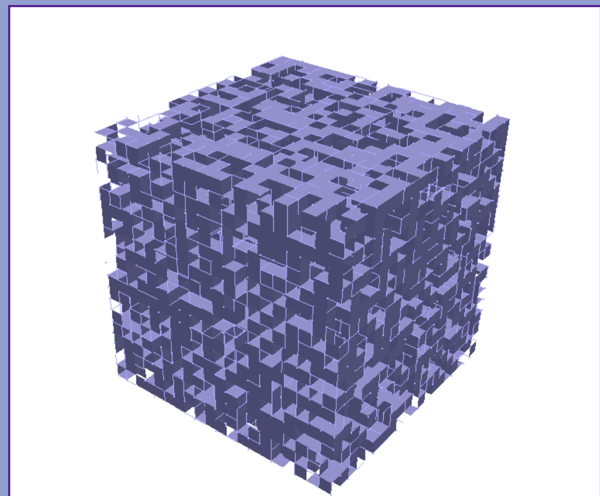
How different?



3D image data

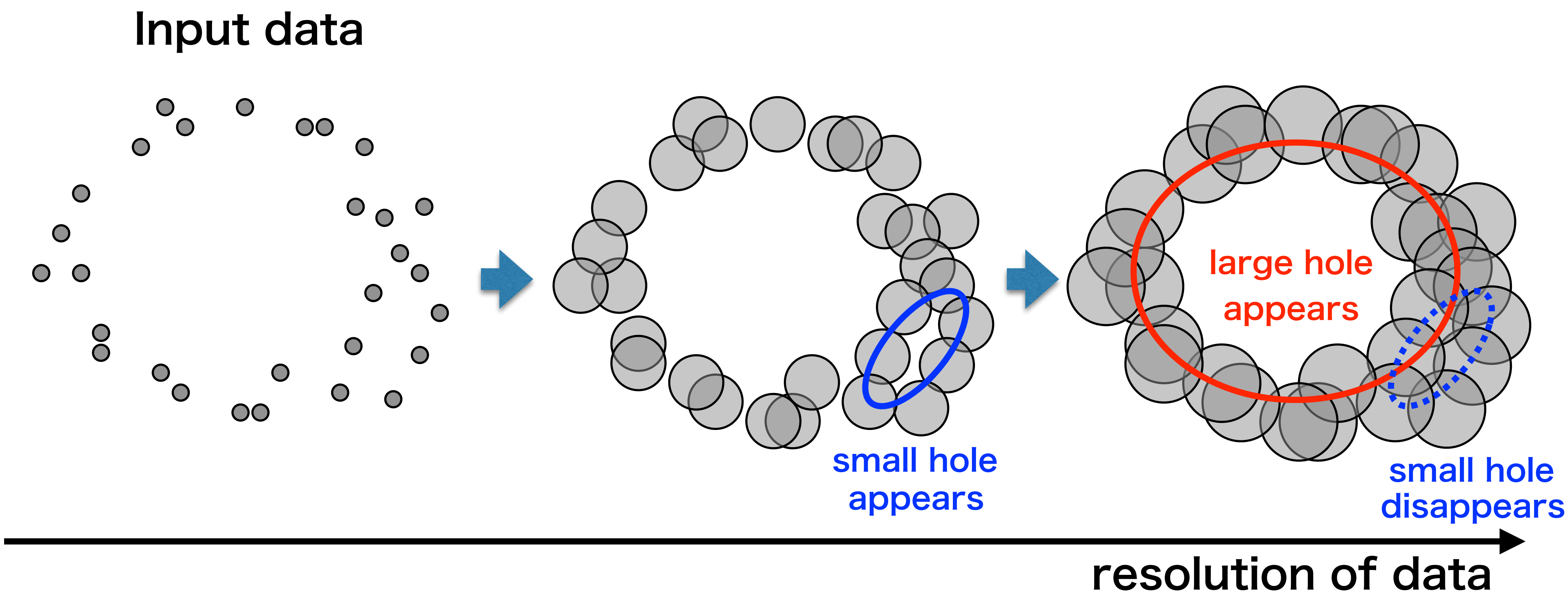


How different?

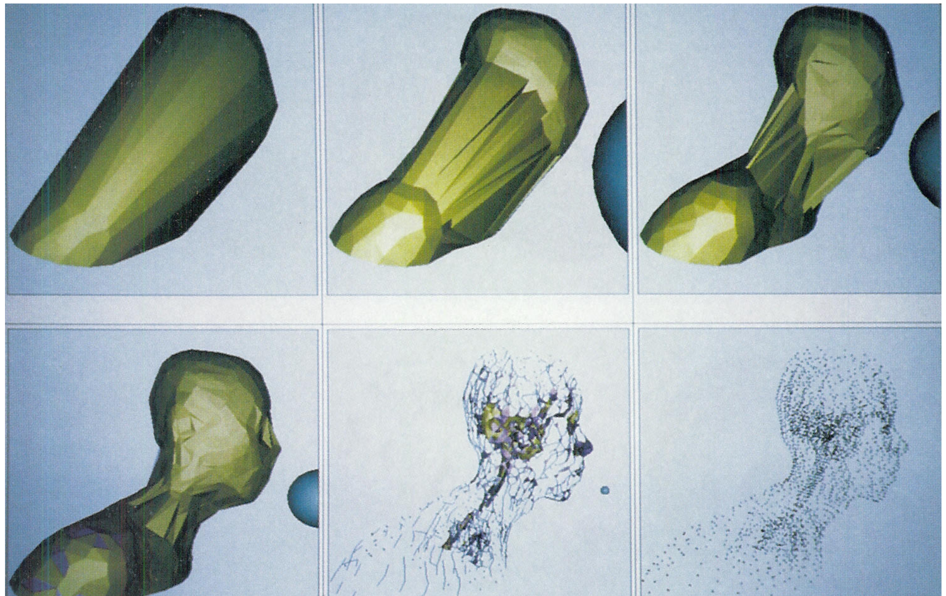


develop mathematical theory for shape of data

Idea : Shape of Data



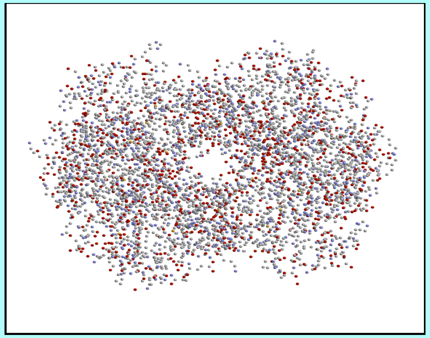
- fattening point data
- changing resolution for multi-scale analysis
- characterization using birth & death of holes



(ref. Edelsbrunner, Mucke)

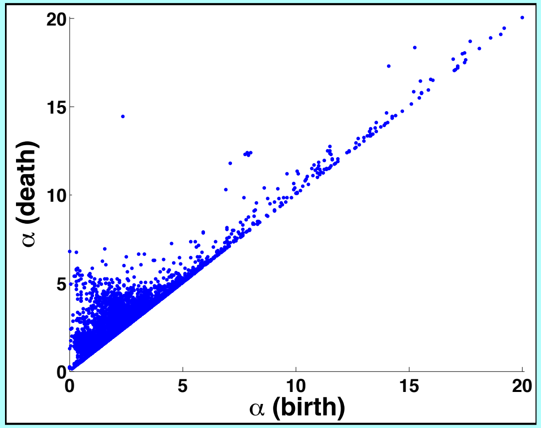
Persistent homology and persistence diagram

Input data → **Persistent Homology** → **Shape of data**



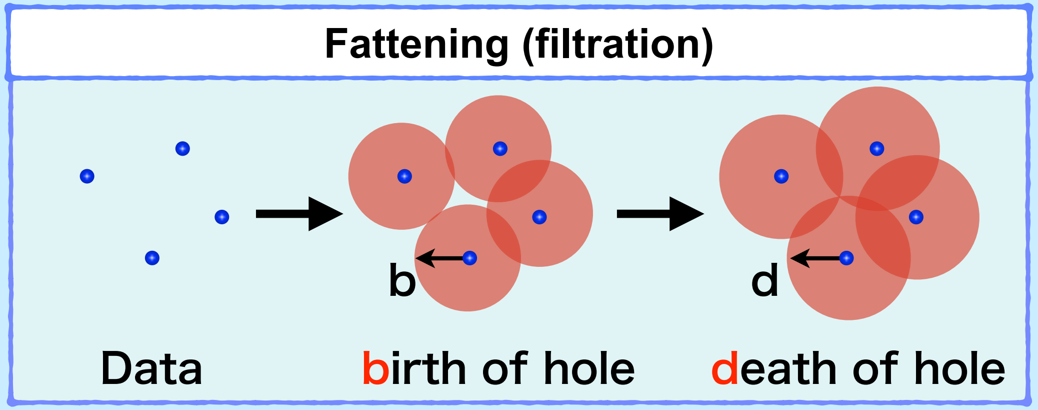
Atomic configuration of hemoglobin

- characterize holes in data
- describe number, size, and shapes
- multi-scale analysis

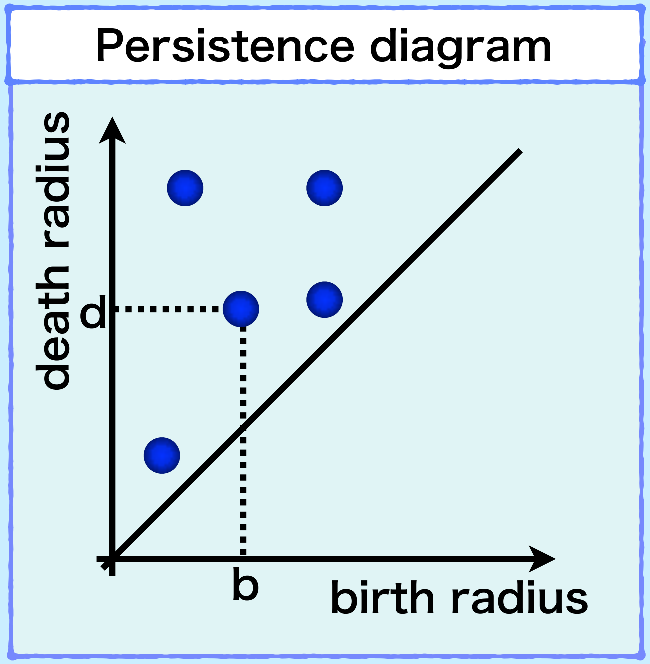


Persistence Diagram (PD)

Persistence diagram of point cloud



inverse



- each point (called generator) in PD expresses a hole in data
- birth & death axes measure shapes of holes
- points close to diagonal are noisy
- points away from diagonal are robust

Note: 2D histogram uncovers further geometry

Content

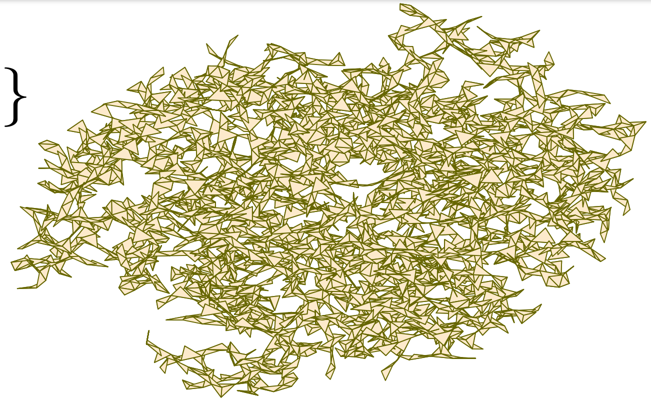
1. 背景
2. データの幾何モデル
3. パーシステントホモロジーの導入
4. 代数的研究（クイバーの表現論）
5. 安定性定理
6. 計算ソフトウェア
7. 応用（材料科学）

Čech complex and nerve theorem

- **Input point cloud** $X=\{x_i\in\mathbf{R}^m \mid i=1,\dots,n\}$
  (e.g., atomic configuration, sensors etc)

- **Čech complex**

$$\mathcal{C}(X,r)=\{|x_{i_0}\cdots x_{i_k}| \mid \cap_{j=0}^{k} B_r(x_{i_j})\neq\emptyset\}$$

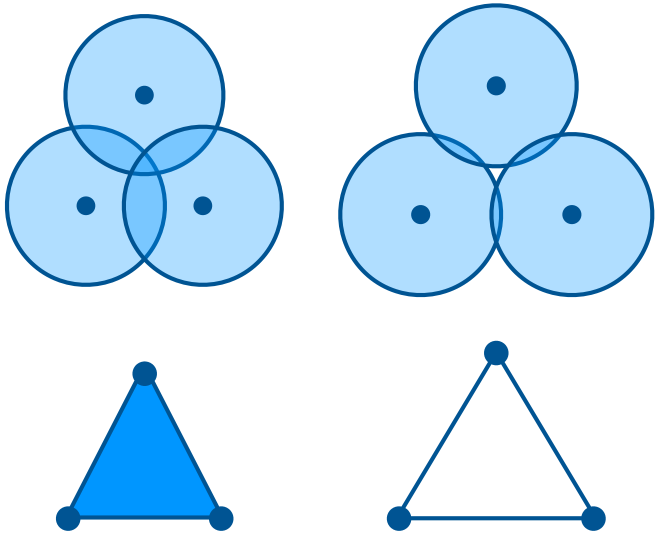


Čech complex model of hemoglobin

<u>example</u>

$\cup_{x\in X}B_r(x)$

$\mathcal{C}(X,r)$



$\exists$ a hole in $\cup_{x\in X}B_r(x)$

⇕

$\exists$ a hole in $\mathcal{C}(X,r)$

Nerve Theorem

$$\cup_{x\in X}B_r(x)\simeq\mathcal{C}(X,r)$$

- homotopy equivalence
- preserve hole information
- LHS: atomic configuration, sensor location, data points
- RHS: easy to treat in computer

Rips complex

Čech complex…… building in higher dimensions is not easy

- **Input point cloud** $X = \{x_i \in \mathbf{R}^m \mid i = 1, \ldots, n\}$
- **Rips complex**

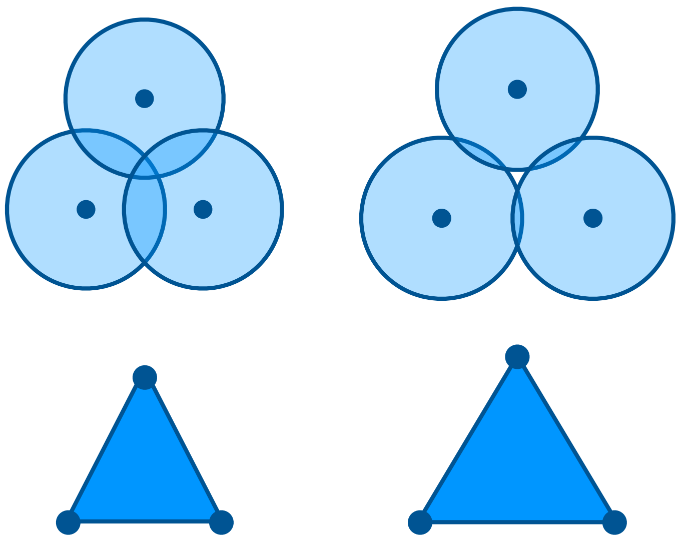
$$\mathcal{R}(X, r) = \{|x_{i_0} \cdots x_{i_k}| \mid B_r(x_{i_s}) \cap B_r(x_{i_t}) \neq \emptyset, \ 0 \leq s < t \leq k\}$$

note: checking pair-wise intersections

<u>example</u>

$\cup_{x \in X} B_r(x)$

$\mathcal{R}(X, r)$



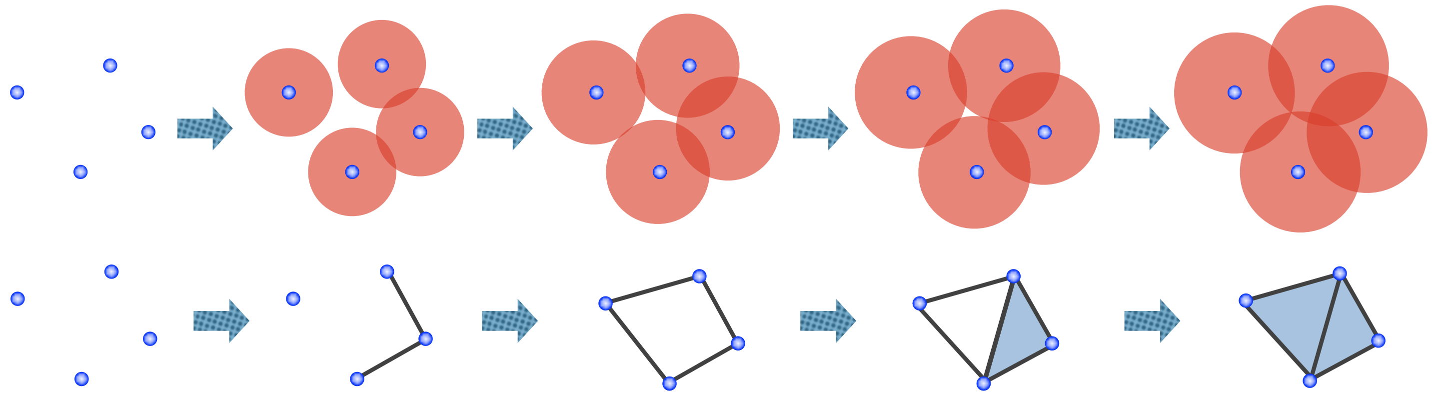
- NOT preserving hole information in general
- computable even in higher dimensions (only requiring distance matrix)

Filtration for multi-scale analysis

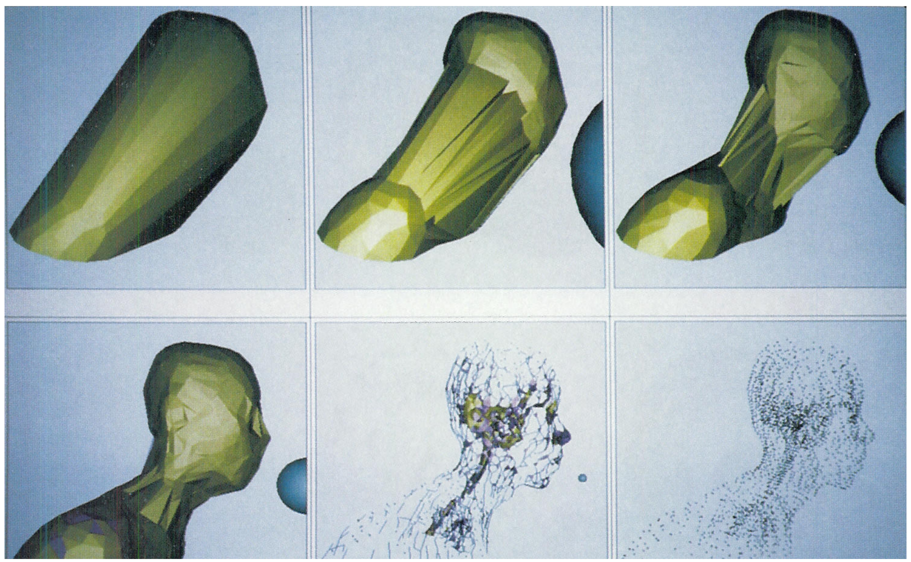
- **Let** $K(X,r) = \mathcal{C}(X,r)$ or $\mathcal{R}(X,r)$

$\Rightarrow$ $K(X,r) \subset K(X,s) \quad \text{for } r \leq s$

($\because$ checking nonempty intersections)



- $\{K(X,r)\}_{r \geq 0}$ **: filtration (fattening sequence)**
- **the parameter r controls a resolution of data**



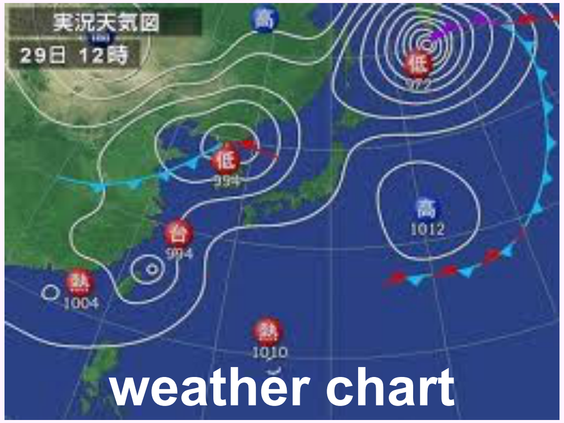
(ref. Edelsbrunner, Mucke)

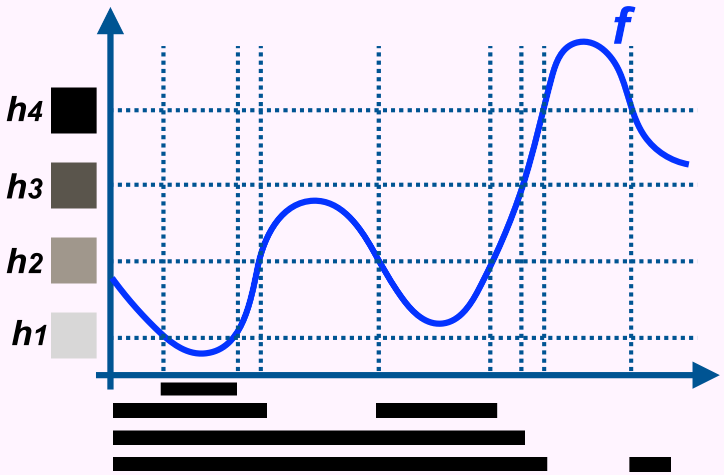
Sublevel set

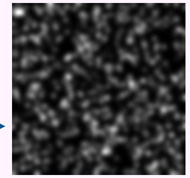
- a map $f : M \to \mathbb{R}$ on a metric space (M, d_M)
- sublevel set $M_h = \{x \in M : f(x) \leq h\}$
- filtration $M_{h_1} \subset M_{h_2} \subset \cdots \subset M_{h_n}$ by $h_1 \leq h_2 \leq \cdots \leq h_n$

Applications

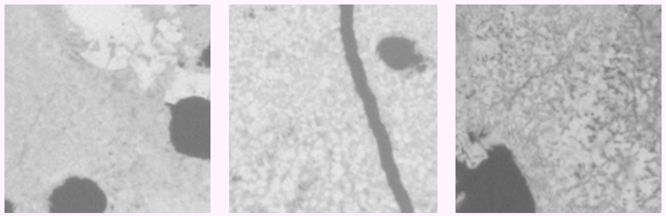
- digital images
- cyclone tracking (Inatsu)







gray scale digital image



slice images of iron ore sinters (Kimura, KEK)

- For a point cloud $X = \{x_i \in M : i = 1, \ldots, K\}$, define

$$\mathrm{dist}_X : M \to \mathbb{R} \quad \text{by} \quad \mathrm{dist}_X(x) := \min_{x_i \in X} d_M(x, x_i)$$

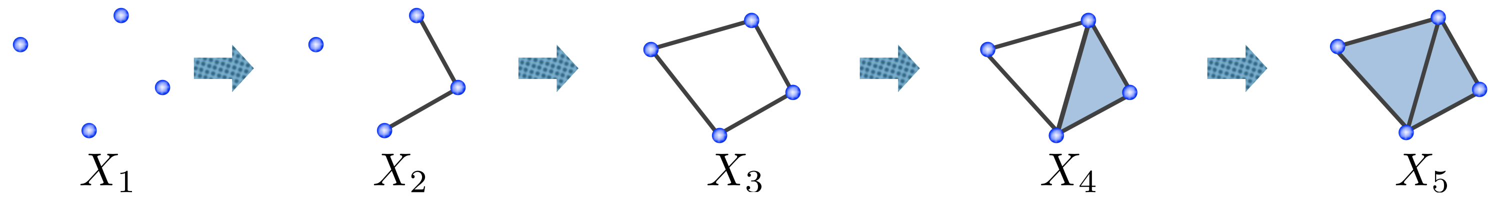
$\Rightarrow$ $M_h = \bigcup_{x_i \in X} B_h(x_i)$ Point clouds can also be studied by sublevel sets

Content

1. 背景
2. データの幾何モデル
3. パーシステントホモロジーの導入
4. 代数的研究（クイバーの表現論）
5. 安定性定理
6. 計算ソフトウェア
7. 応用（材料科学）

Persistent homology and persistence diagram

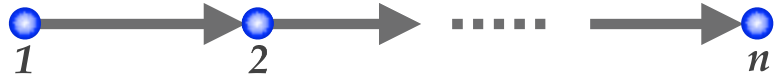
Edelsbrunner, Letscher, Zomorodian, Carlsson, de Silva



$H_1(\mathcal{X}) \simeq I[3,4]$

- **filtration** $\mathcal{X} : X_1 \subset X_2 \subset \cdots \subset X_n$
- **persistent homology** $H_\ell(\mathcal{X}) : H_\ell(X_1) \to H_\ell(X_2) \to \cdots \to H_\ell(X_n)$

representation on A_n



- **interval decomposition (Gabriel's Theorem)**

$$H_\ell(\mathcal{X}) \simeq \bigoplus_{i=1}^{s} I[b_i, d_i]$$

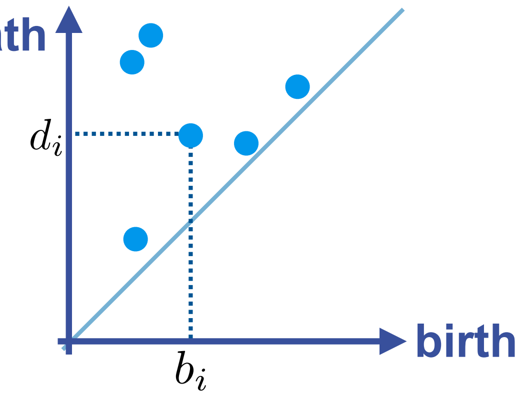
***d* - *b* : lifetime (or persistence)**

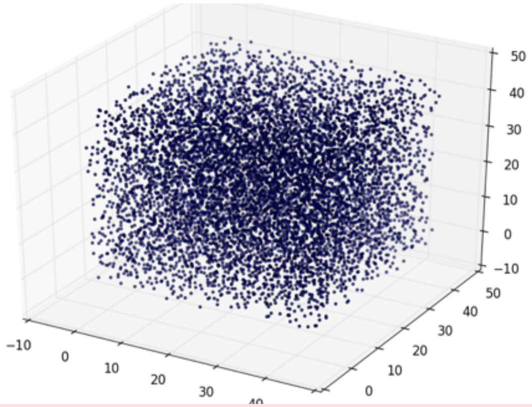
$$I[b,d] : 0 \to \cdots \to 0 \to K \to \cdots \to K \to 0 \to \cdots \to 0$$

at X_b **at X_d**

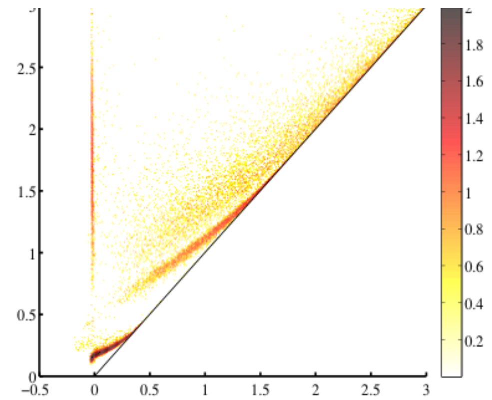
each interval represents birth & death of a topological feature

- **persistence diagram (PD)** $D_\ell(\mathcal{X}) = \{(b_i, d_i) \in \mathbb{R}^2 : i = 1, \ldots, s\}$





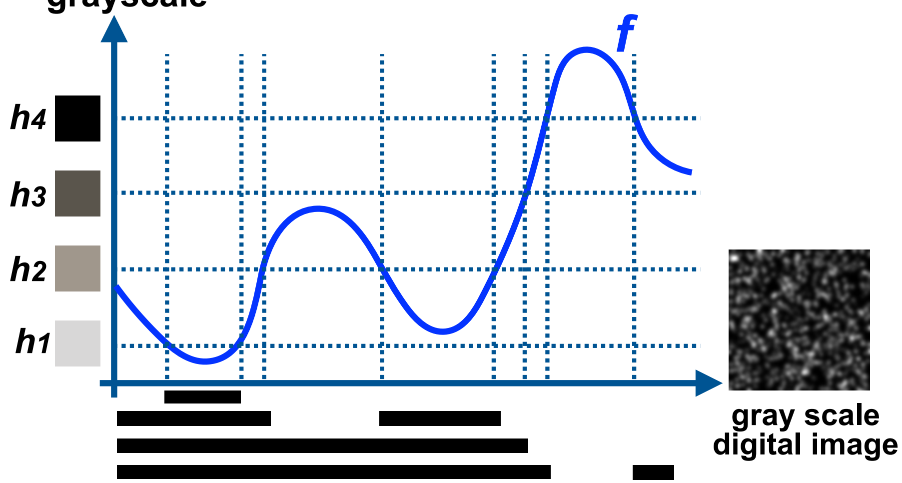
atomic configuration of glass



persistence diagram of glass

Persistent homology of digital image

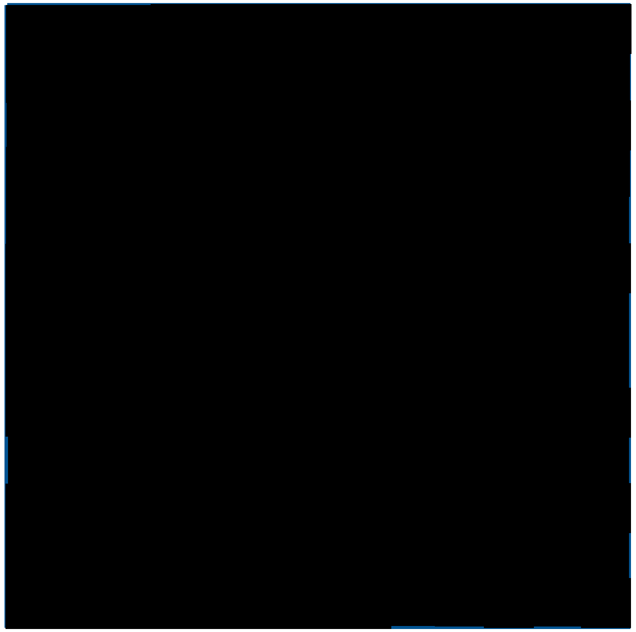
1. Grayscale persistence

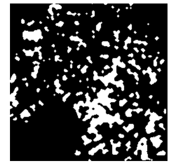


- **sub-level set** $X_h := \{x \in X \mid f(x) \leq h\}$
- **fattening** $X_{h_1} \subset X_{h_2} \subset \cdots \subset X_{h_T}$

  by $h_1 \leq h_2 \leq \cdots \leq h_T$

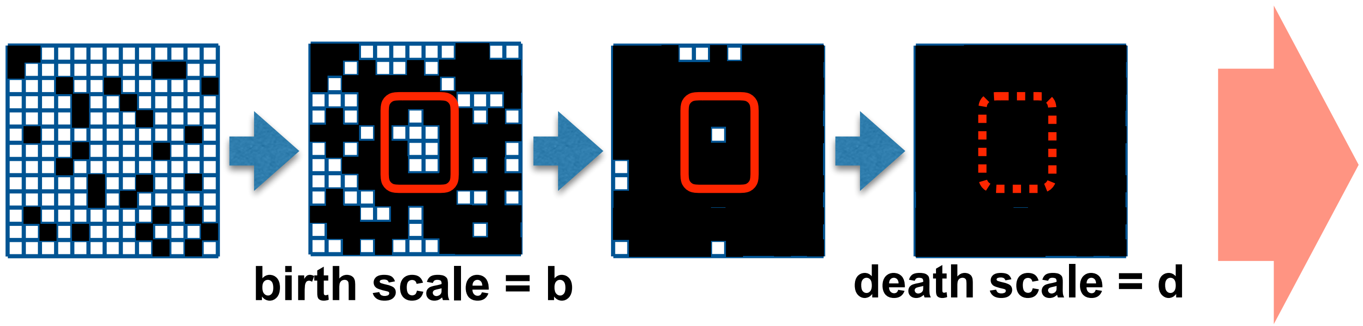
2. Spatial persistence

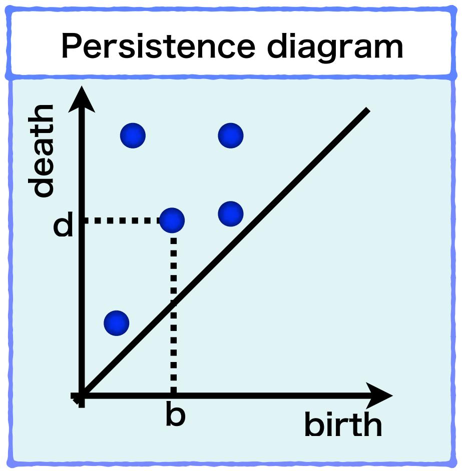




black-white image

Persistence diagram of digital images





Characterize grayscale/spatial persistent holes in images

Historical remarks

Computational homology project ('02-present)

- ホモロジー高速計算の開発と力学系を中心とした諸問題への応用（Mischaikow, Mrozek, Pilarczyk, 荒井，平岡，國府 etc）
- 入力データの2値化（スケールの固定）が必要

Edelsbrunner, Letscher, and Zomorodian ('02)

- （限定的な）単体複体フィルトレーションに対してベッチ数の変化を調べる
- 区間分解の構成的アルゴリズムを提示

Zomorodian and Carlsson ('05)

- より一般的な設定（persistence modules）での区間分解定理の証明、パーシステント図の導入
- 次数付き加群として鎖複体（境界作用素）を導入する定式化
- Zomorodianはその後企業へ（D. E. Shaw & Co ← 有名なヘッジファンドの一つ）

de Silva and Carlsson ('10)

- *An*クイーバの表現としての定式化

代数的研究の重要課題
multi-parameter persistence

Carlsson and Zomorodian ('09): Gröbner basis
Escolar and H ('16): Auslander-Reiten theory
Justin Curry ('14), Kashiwara-Schapira ('17): (co-)Sheaf, micro-local analysis

その他の流れ

- 安定性定理
- 確率論
- 逆問題
- 統計・機械学習
- 応用

Content

1. 背景

2. データの幾何モデル

3. パーシステントホモロジーの導入

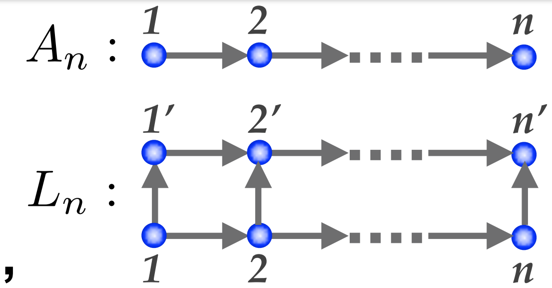
4. 代数的研究（クイバーの表現論）

5. 安定性定理

6. 計算ソフトウェア

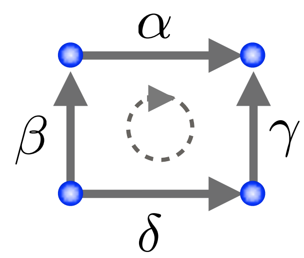
7. 応用（材料科学）

Representation of quivers (associative algebras)



- **Quiver** $Q = (Q_0, Q_1)$
  - **set of arrows** (Q_1)
  - **set of vertices** (Q_0)
- **Path algebra** KQ: K**-vector space spanned by all paths,**
  where the product of two paths is their composition
- **Associative algebra** $A = KQ/I, \quad I = \langle \rho_1, \ldots, \rho_s \rangle$

  a relation $\rho = \sum_{i=1}^{k} c_i w_i, \quad c_i \in K$ (w_i: **path**)



commutative relation
$I = \langle \alpha\beta - \gamma\delta \rangle$

- **A representation** $M = (M_a, \varphi_\alpha)_{a \in Q_0, \alpha \in Q_1}$ **on** Q **(or** A**)**
  - **a vector space** M_a **on each vertex** $a \in Q_0$
  - **a linear map** $\varphi_\alpha : M_a \to M_b$ **for each arrow** $\alpha : a \to b$

  (- $\varphi_\rho = \sum c_i \varphi_{w_i} = 0$**, i.e., composition on relations vanishes)**

  $\varphi_\alpha \varphi_\beta - \varphi_\gamma \varphi_\delta = 0$

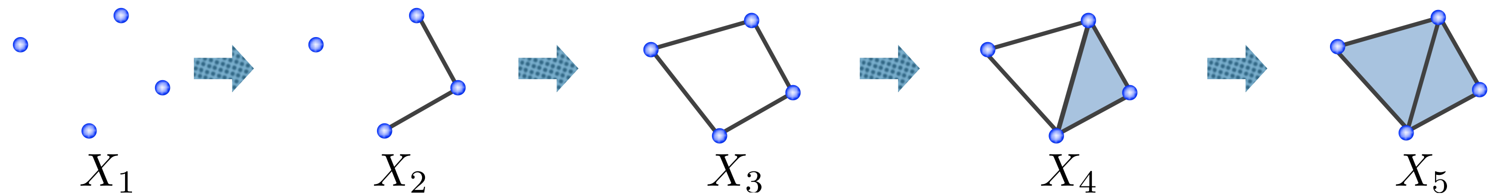
- **An indecomposable representation** M**:** $M = N \oplus N'$ ⇒ $N = 0$ or $N' = 0$
- **Krull-Schmidt Theorem:** $M \simeq N^{(1)} \oplus \cdots \oplus N^{(\ell)}$, $N^{(i)}$**: indecomposable**
- **Gabriel Theorem: For an** A_n**-quiver with arbitrary orientations,**

$$M \simeq \bigoplus_{1 \le b \le d \le n} I[b,d]^{m_{bd}}, \quad m_{bd} \in \mathbb{N}_0 \quad \textbf{(interval decomposition)}$$

$$I[b,d] : 0 \leftrightarrow \cdots \leftrightarrow 0 \leftrightarrow \underset{\textbf{at } b}{K} \leftrightarrow \cdots \leftrightarrow \underset{\textbf{at } d}{K} \leftrightarrow 0 \leftrightarrow \cdots \leftrightarrow 0$$

Persistent homology and persistence diagram

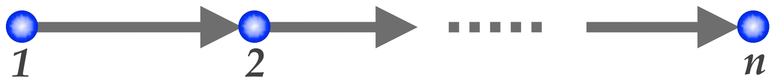
Edelsbrunner, Letscher, Zomorodian, Carlsson, de Silva



$H_1(\mathcal{X}) \simeq I[3,4]$

- **filtration** $\mathcal{X} : X_1 \subset X_2 \subset \cdots \subset X_n$
- **persistent homology** $H_\ell(\mathcal{X}) : H_\ell(X_1) \to H_\ell(X_2) \to \cdots \to H_\ell(X_n)$

representation on A_n



- **interval decomposition (Gabriel's Theorem)**

$$H_\ell(\mathcal{X}) \simeq \bigoplus_{i=1}^{s} I[b_i, d_i]$$

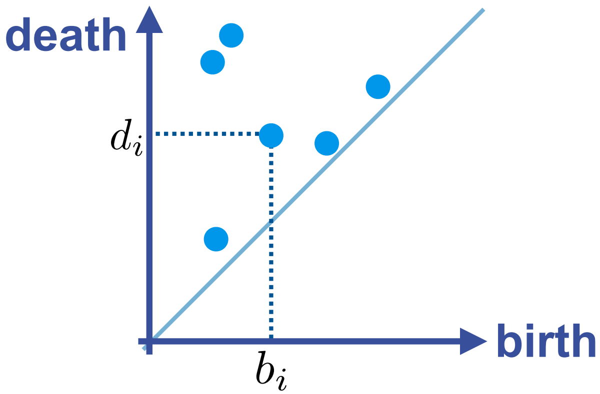
d - b : lifetime (or persistence)

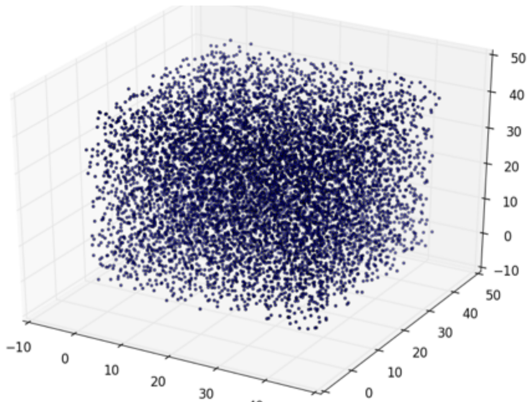
$$I[b,d] : 0 \to \cdots \to 0 \to K \to \cdots \to K \to 0 \to \cdots \to 0$$

at X_b at X_d

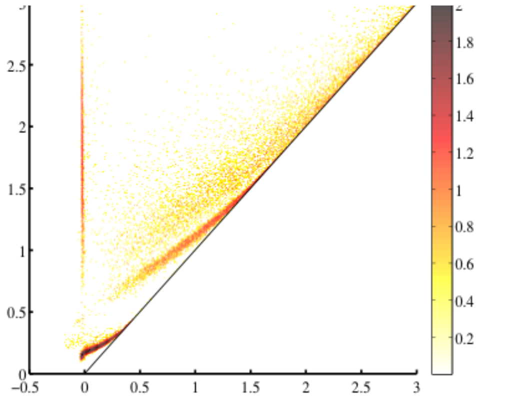
each interval represents birth & death of a topological feature

- **persistence diagram (PD)** $D_\ell(\mathcal{X}) = \{(b_i, d_i) \in \mathbb{R}^2 : i = 1, \ldots, s\}$





atomic configuration of glass



persistence diagram of glass

Zigzag persistence

Merit for data analysis

- **Time series data** $X(t) = \{x_i(t) \in \mathbf{R}^m : i = 1, \ldots, K\}, \quad t = 1, 2, \ldots, T$

  (e.g., protein folding, polymer deformation, moving sensors)

  - time series of Čech complexes $\mathcal{C}(t) := \mathcal{C}(X(t), r), \quad t = 1, \ldots, T$

  but not filtration w.r.t. time t

  - zigzag sequence $\cdots \hookleftarrow \mathcal{C}(t) \hookrightarrow \mathcal{C}(t) \cup \mathcal{C}(t+1) \hookleftarrow \mathcal{C}(t+1) \hookrightarrow \cdots$

  $$\Rightarrow \cdots \leftarrow H\mathcal{C}(t) \rightarrow H(\mathcal{C}(t) \cup \mathcal{C}(t+1)) \leftarrow H\mathcal{C}(t+1) \rightarrow \cdots$$

  $\simeq \oplus_j I[b_j, d_j]$ **(** $\because$ **representation of** A_n **quiver)**

We can study persistent topological features in time series sense

Drawback?

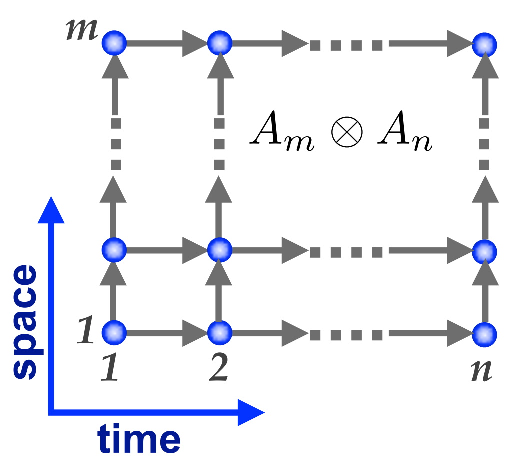
- **The spatial resolution** r **is fixed in** $\mathcal{C}(t) = \mathcal{C}(X(t), r)$ **, and persistent topological features in spatial sense can't be studied!**

What's next?

We want both!! $\Rightarrow$ **persistent homology with multi-parameters**

persistent homology with multi-parameters

- **multi-parameter persistent homology**

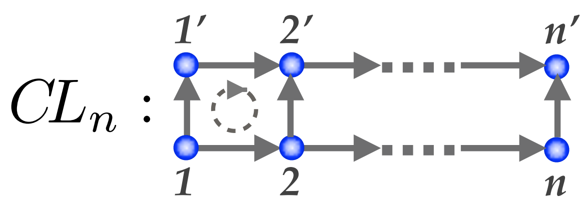


- well-defined as a representation, but its decomposition theory is not understood well.

- developing decomposition theory is very important for applications in TDA

- Carlsson and Zomorodian ('09) apply Gröbner basis to derive (incomplete) invariants

- **persistent homology on commutative ladder**



- Escolar and H ('16) apply Auslander-Reiten theory for decomposition theory of $n \leq 4$

- useful in materials science (detecting robust geometric features in glass under compression process)

- **BOCS representation: 浅芝先生の講演へ**
  (bimodule over a category with a coalgebra structure)

Auslander-Reiten theory and persistence diagrams

Escolar and H. Discrete Comput. Geom. (2016)

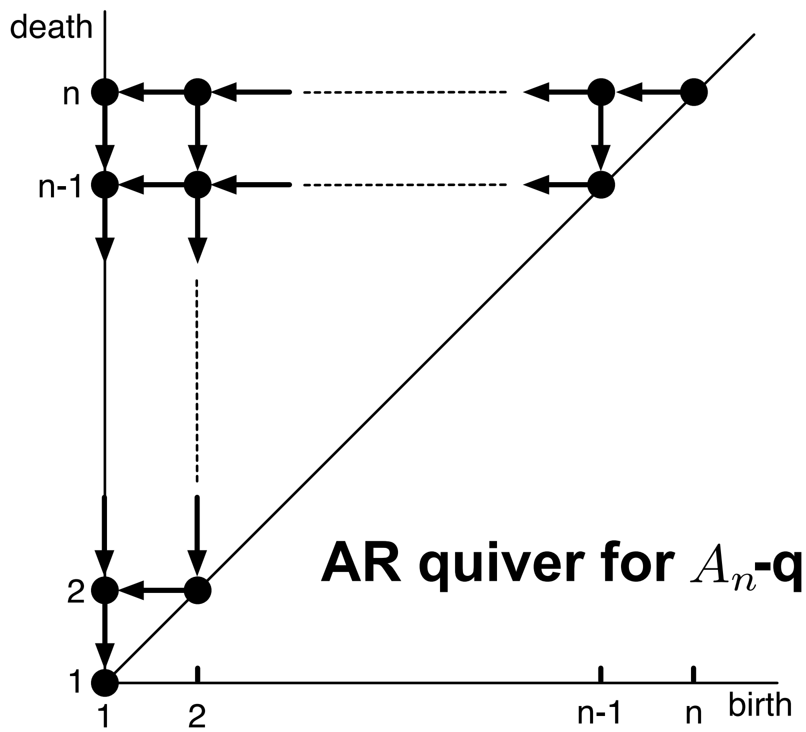
- **Auslander-Reiten quiver** $\Gamma = (\Gamma_0, \Gamma_1)$ **of a quiver** Q **(or** A**)**

  Γ_0 **: the set of iso. classes of indecomposable representations**

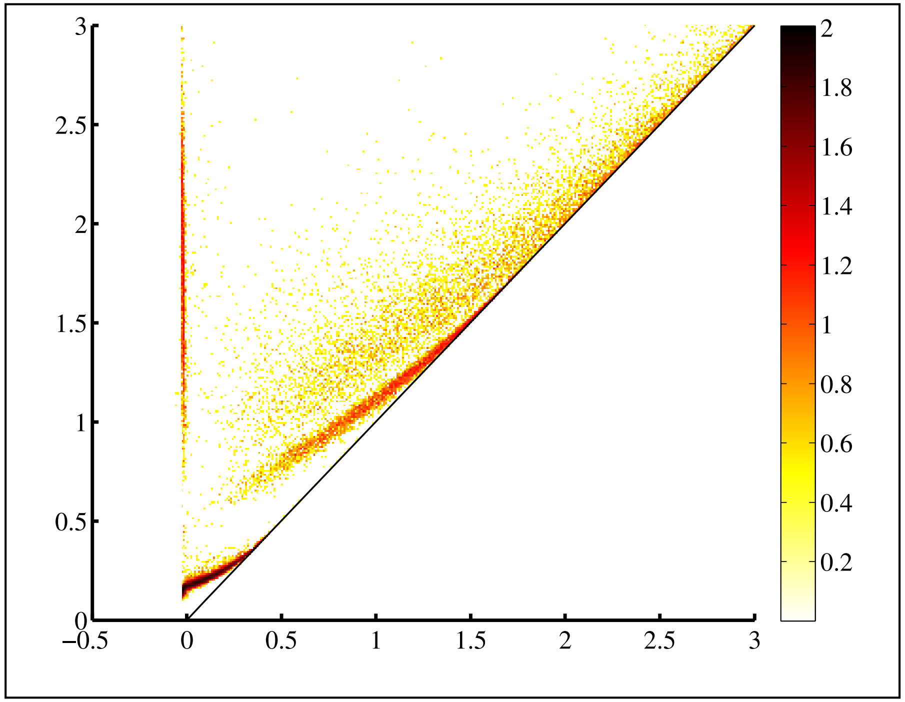
  $\Gamma_1 \ni \varphi : [I] \to [J]$ ⟺ (def) $\exists$ **an irreducible map** $I \to J$

- **From Gabriel's theorem on** A_n**-quiver,** $M \simeq \bigoplus_{1 \le b \le d \le n} I[b,d]^{m_{bd}}, \quad m_{bd} \in \mathbb{N}_0$

→ **PD is defined as the function** $D : \Gamma_0 \ni I[b,d] \to m_{b,d} \in \mathbb{N}_0$



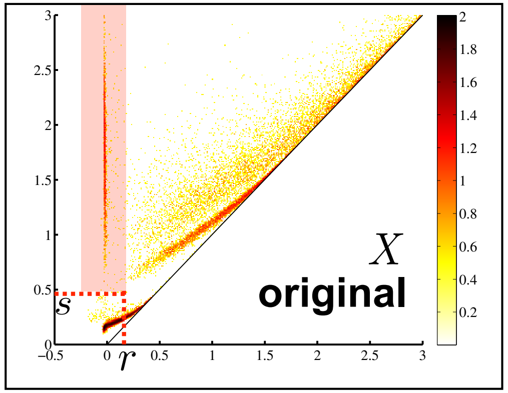
AR quiver for A_n**-quiver**

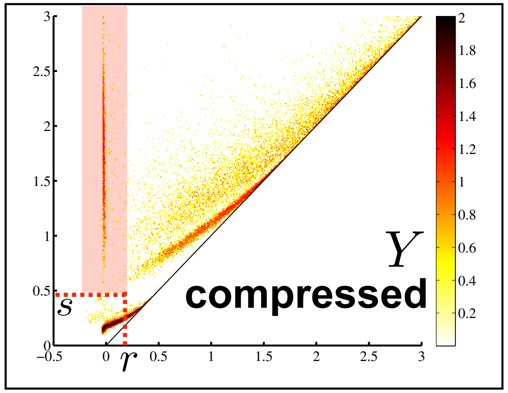


Commutative ladder persistence

Escolar and H. Discrete Comput. Geom. (2016)

Study common and robust top. properties under pressurization of materials

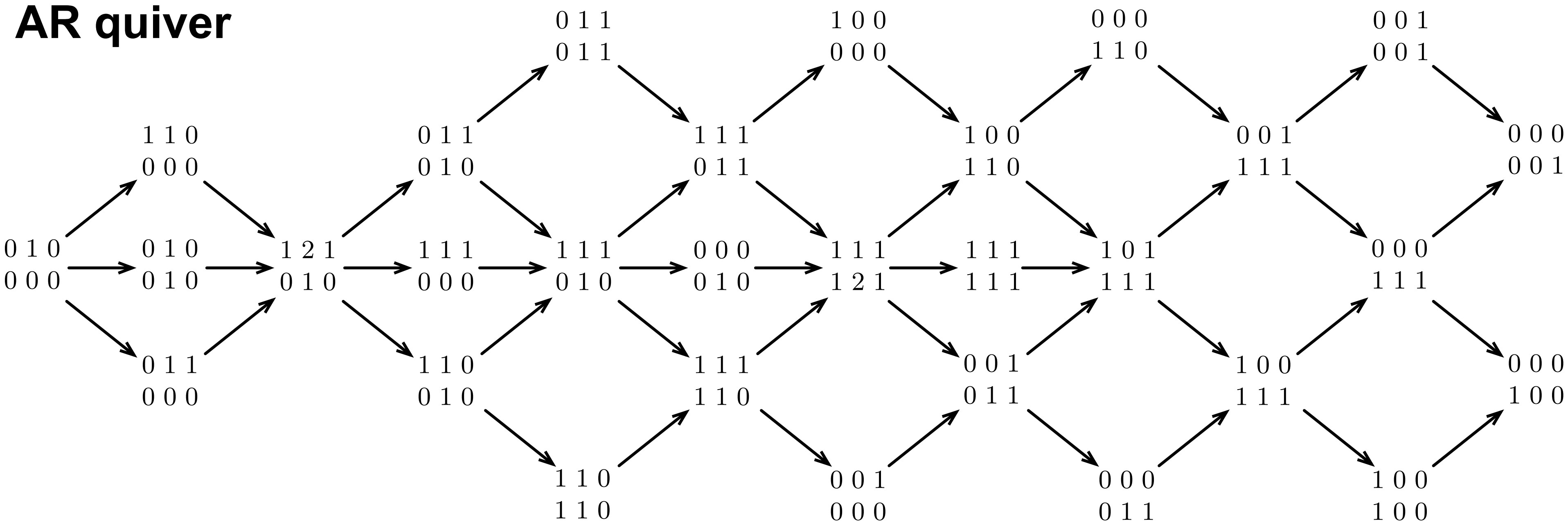




commutative ladder persistence with length 3

$$\begin{array}{ccccc} H_*(X_s) & \rightarrow & H_*(X_s \cup Y_s) & \leftarrow & H_*(Y_s) \\ \uparrow & & \uparrow & & \uparrow \\ H_*(X_r) & \rightarrow & H_*(X_r \cup Y_r) & \leftarrow & H_*(Y_r) \end{array}$$

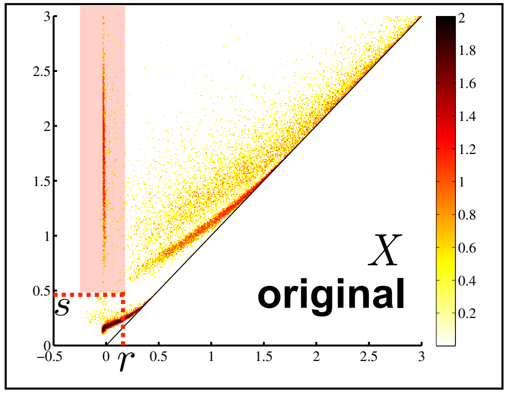
- **AR quiver**

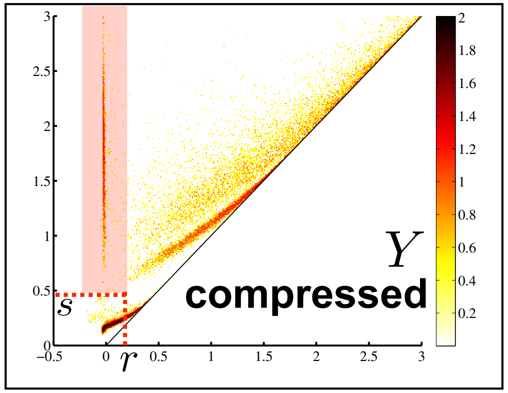


Commutative ladder persistence

Escolar and H. Discrete Comput. Geom. (2016)

Study common and robust top. properties under pressurization of materials

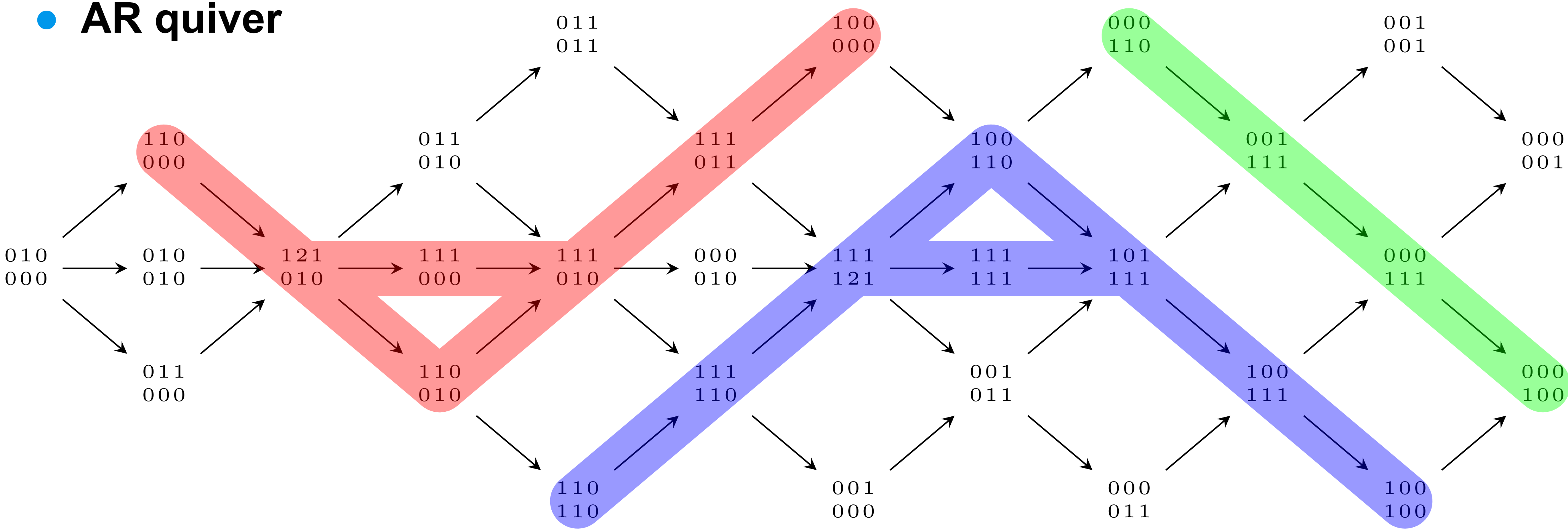




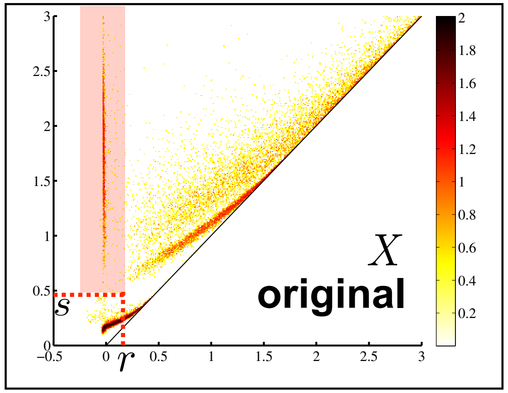
commutative ladder persistence with length 3

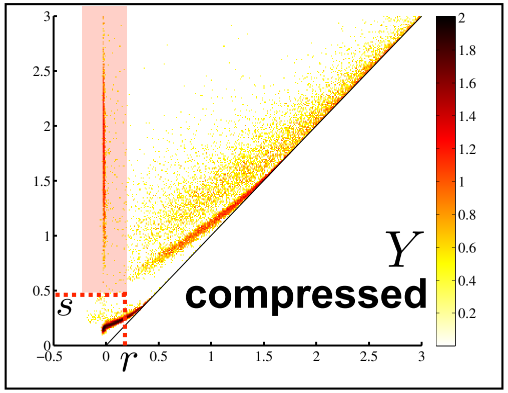
$$
\begin{array}{ccccc}
H_*(X_s) & \rightarrow & H_*(X_s \cup Y_s) & \leftarrow & H_*(Y_s) \\
\uparrow & & \uparrow & & \uparrow \\
H_*(X_r) & \rightarrow & H_*(X_r \cup Y_r) & \leftarrow & H_*(Y_r)
\end{array}
$$

- **AR quiver**



Study common and robust top. properties under pressurization of materials

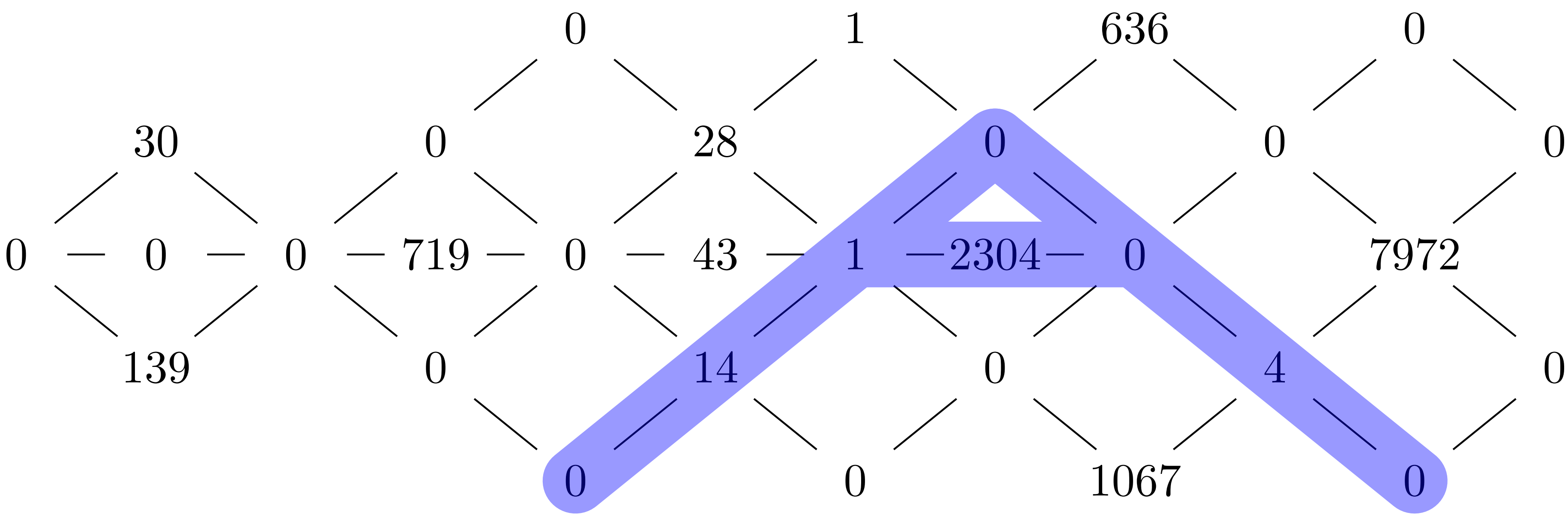




commutative ladder persistence with length 3

$$\begin{array}{ccccc} H_*(X_s) & \rightarrow & H_*(X_s \cup Y_s) & \leftarrow & H_*(Y_s) \\ \uparrow & & \uparrow & & \uparrow \\ H_*(X_r) & \rightarrow & H_*(X_r \cup Y_r) & \leftarrow & H_*(Y_r) \end{array}$$

- **persistence diagram of the pressurization process**



99.18% (≒ 2304/2323) generators persist under pressurization!

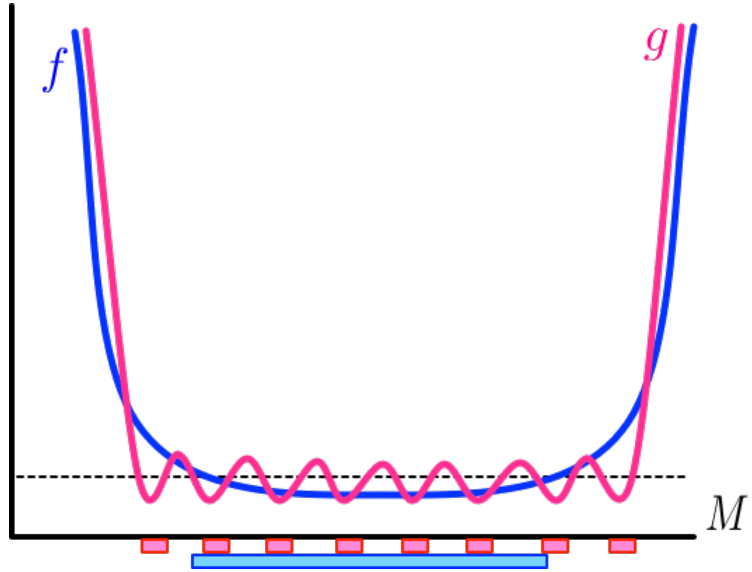
Content

1. 背景
2. データの幾何モデル
3. パーシステントホモロジーの導入
4. 代数的研究（クイバーの表現論）
5. 安定性定理
6. 計算ソフトウェア
7. 応用（材料科学）

Stability of persistence diagrams

Motivation

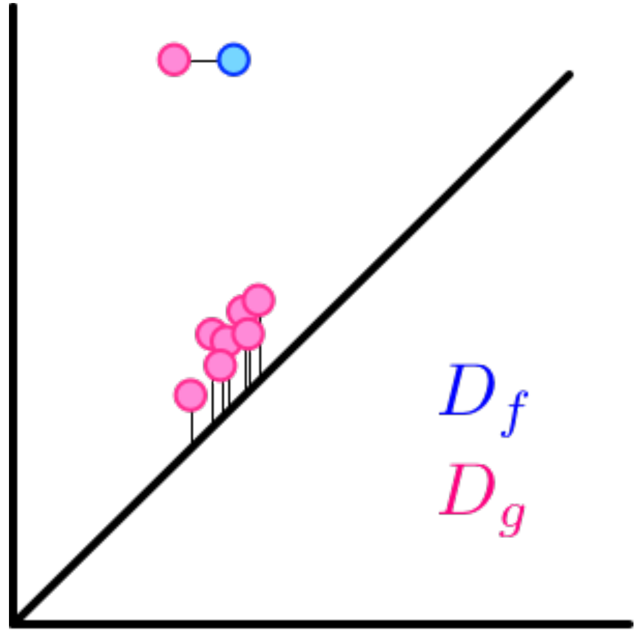
- **In practical applications, input data is usually affected by noise**
- **Homology are NOT stable w.r.t. noise**
- **How about persistent homology?**



Stability Theorem (Cohen-Steiner, et al, '07)

For (tame) continuous functions $f, g : M \to \mathbb{R}$**,**

$$d_{\mathrm{b}}(D_f, D_g) \leq \|f - g\|_\infty$$



- D_f **: PD of the sublevel set filtration for** f
- $d_{\mathrm{b}}(\bullet, \bullet)$**: the bottleneck distance**

$$d_{\mathrm{b}}(D_f, D_g) := \inf_{\gamma} \sup_{p \in \bar{D}_f} \|p - \gamma(p)\|_\infty$$

where $\bar{D} := D \sqcup \Delta$ **and** $\gamma : \bar{D}_f \to \bar{D}_g$ **is a bijection.**

(diagonal)

Algebraic stability theorem (Chazal et al '09, Bauer and Lesnick '14)

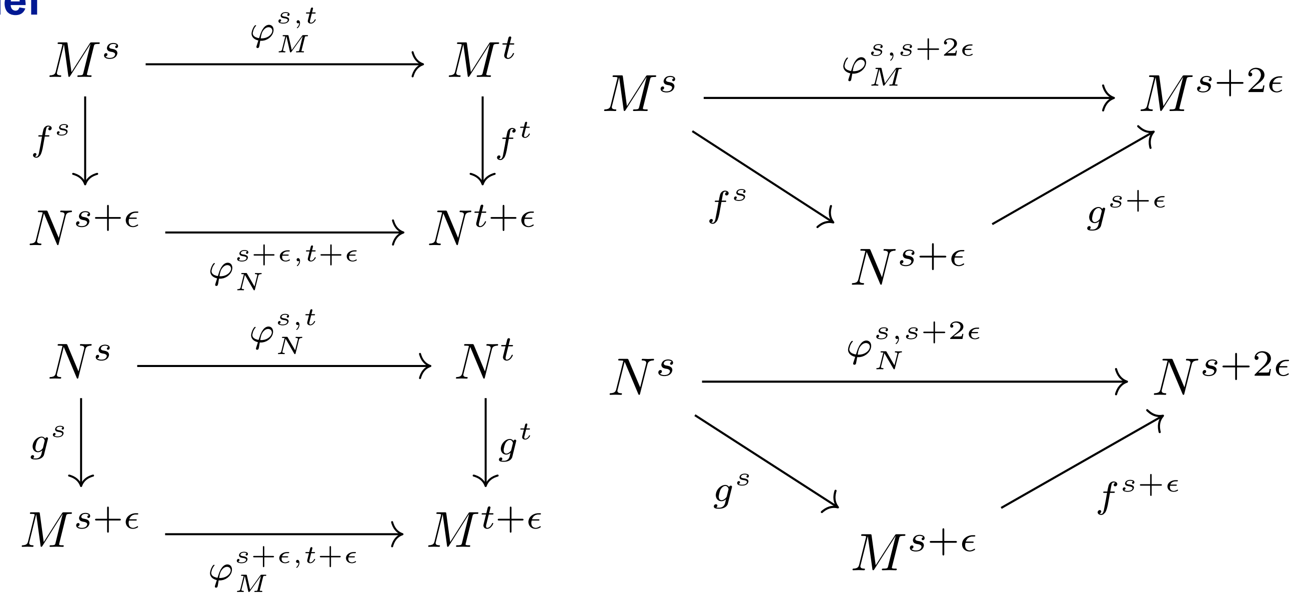
- **persistence module M (PM)** $\overset{\text{def}}{\Longleftrightarrow}$ **functor** $M : \mathbf{R} \to \mathbf{vect}$
  - category of fin. dim. vector spaces
  - as a poset category

$$\varphi_M^{s,t} : M^s \to M^t \quad (s \leq t), \quad \varphi_M^{s,s} = \mathrm{id}$$

Fact: $M \simeq \bigoplus I[b, d]$ **(interval decomposable)** ⇒ **denote its PD by D_M**

- **persistence modules $M, N : \mathbf{R} \to \mathbf{vect}$ are ϵ-interleaving**

$\overset{\text{def}}{\Longleftrightarrow}$ $\exists f : M \to N(\epsilon), \ \exists g : N \to M(\epsilon)$ **s.t.**



- **ϵ-approximation of isomorphism**
- **measures proximity of two modules**

- **interleaving distance** $d_{\mathrm{I}}(M, N) := \inf\{\epsilon \in [0, \infty) : M, N \text{ are } \epsilon\text{-interleaving}\}$

(applicable even to multi-parameter PM)

Algebraic stability theorem $d_{\mathrm{b}}(D_M, D_N) \leq d_{\mathrm{I}}(M, N)$

remark: "=" holds (isometry theorem)

Stability of persistence diagrams

Bottleneck stability for Čech PDs

$$d_{\mathrm{b}}(D(\mathcal{C}(X)), D(\mathcal{C}(Y))) \leq d_{\mathrm{H}}(X, Y)$$

Hausdorff distance: $d_H(X,Y) = \max\{\max_{x \in X} d(x,Y), \max_{y \in Y} d(X,y)\}$

Remarks

- ***r*-Wasserstein distance on PDs and its stability**

$$d_{\mathrm{W}_r}(D, D') = \inf_{\gamma} \left(\sum_{p \in \bar{D}} \|p - \gamma(p)\|_\infty^r \right)^{1/r}$$

➡ **geometry of a set of PDs as a metric space?**

➡ **continuation of point clouds via PDs (Gameiro, Obayashi, H, '16)**

- **Stability theorem using PWGK (Kusano, Fukumizu, H, '16)**

➡ **福水さんの講演へ（Kernel methods on PDs）**

- **PDs as counting measures on the plane**

➡ **白井さんの講演へ（random point process）**

Content

1. 背景
2. データの幾何モデル
3. パーシステントホモロジーの導入
4. 代数的研究（クイバーの表現論）
5. 安定性定理
6. 計算ソフトウェア
7. 応用（材料科学）

Computation

CHomP (Mischaikow, Mrozek, Pilarczyk, etc):
方体ホモロジーの計算

Perseus (Nanda):
離散モース理論を用いたPDの計算

PHAT, DIPHA (Bauer, Kerber, Reininghaus, Wagner):
PDの高速計算

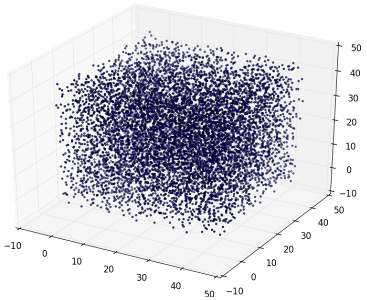
Ripser (Bauer):
Rips PDの高速計算

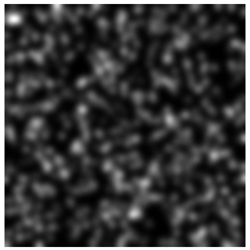
（仮名）Cubical Ripser (阿原・須藤):
方体PDの高速計算

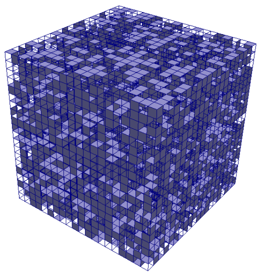
HomCloud (大林):
つぎのスライドで説明

ソフトウェア開発：HomCloud

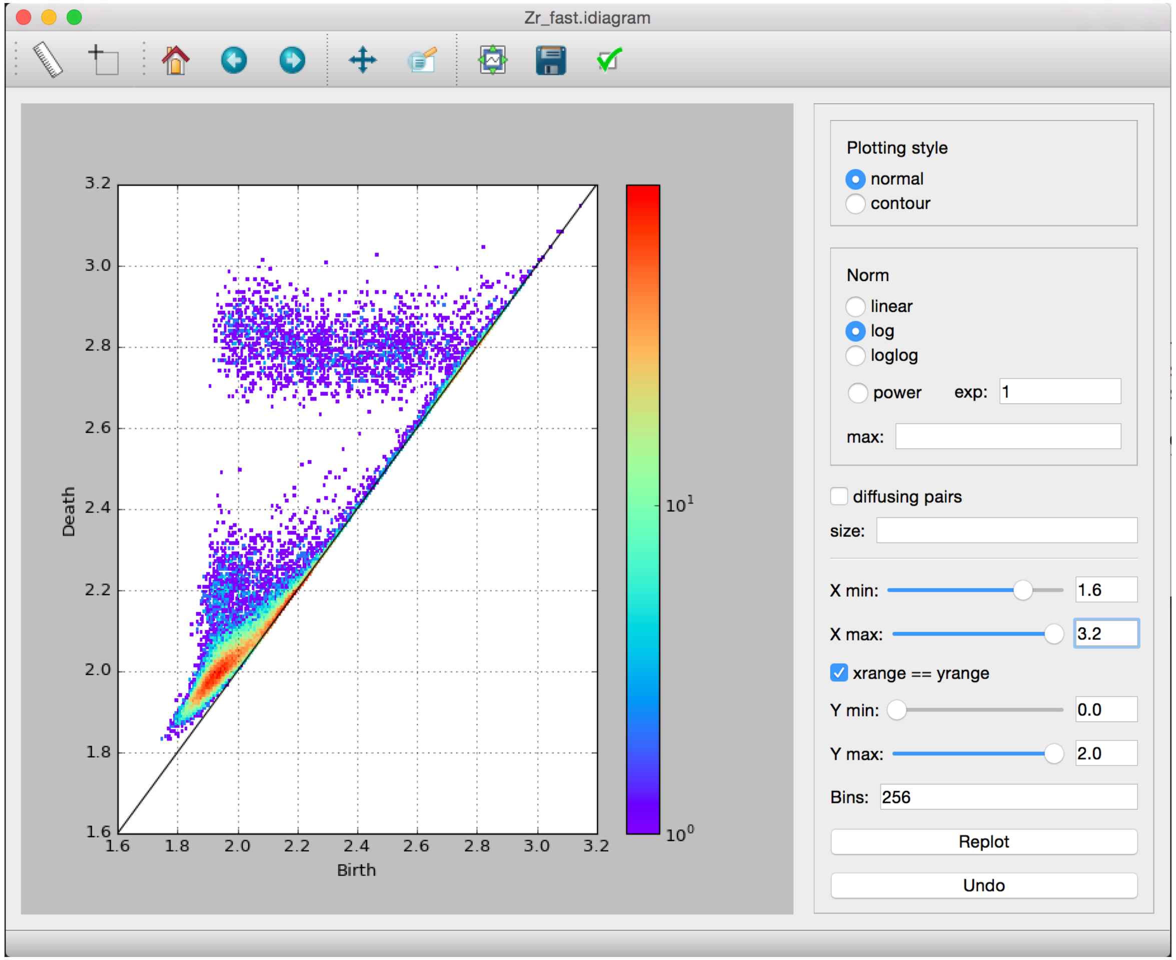
入力データ







パーシステント図（PD）



1)東北大学AIMRで開発するTDAソフトウェア（開発リーダー：大林一平氏）
2)高機能GUIの搭載による汎用性（トポロジーの予備知識は不要）
3)高速PD計算PHAT、DIPHAを搭載
4)空間点データおよび2D/3D画像データ解析
5)PD逆問題、PD統計解析、PDスパース解析（LASSOなど）
http://www.wpi-aimr.tohoku.ac.jp/hiraoka_labo/index.html

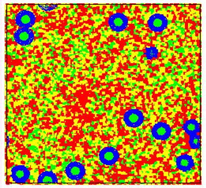
Content

1. 背景
2. データの幾何モデル
3. パーシステントホモロジーの導入
4. 代数的研究（クイバーの表現論）
5. 安定性定理
6. 計算ソフトウェア
7. 応用（材料科学）

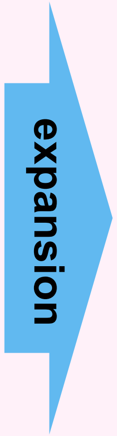
Materials TDA

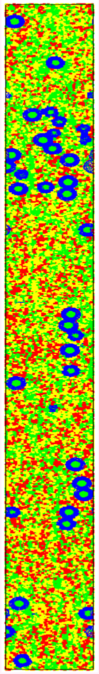
Supported by AIMR, CREST, SIP, MI^2I, NEDO

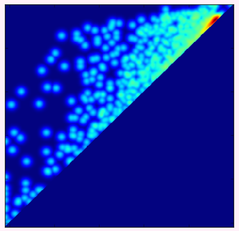
Polymer

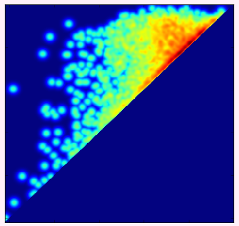


Atomic Force Microscopy image (by Nakajima)

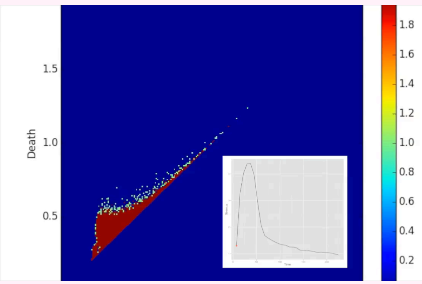








craze formation

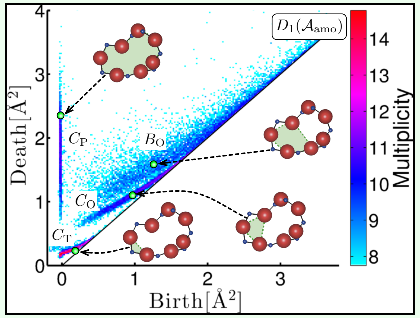


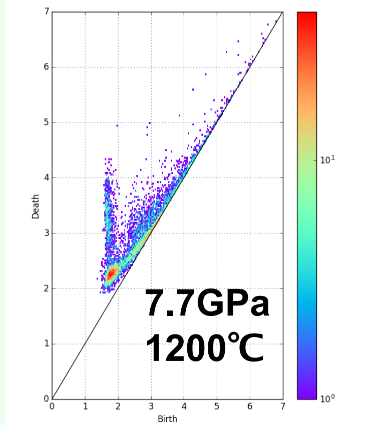
PRE (2017)

Glass

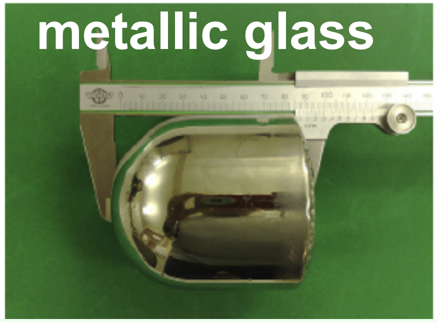
Nanotechnology (2015)

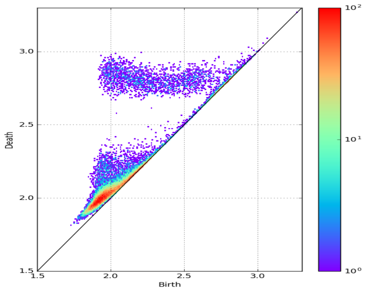
PNAS (2016)

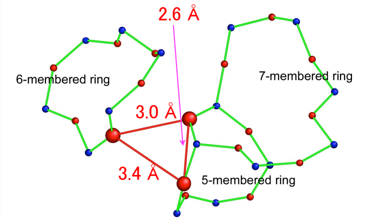




densified silica

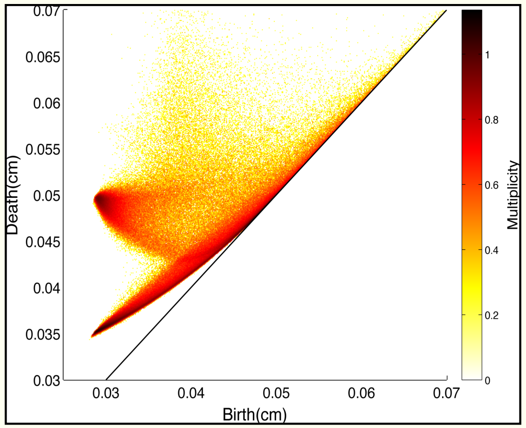


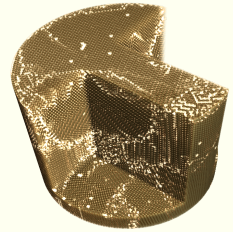




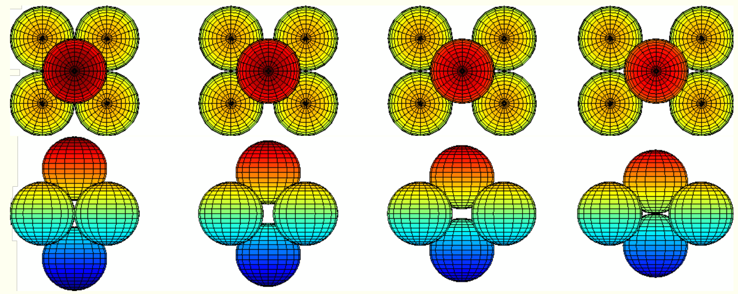
Grain

Nature Communications (2017)





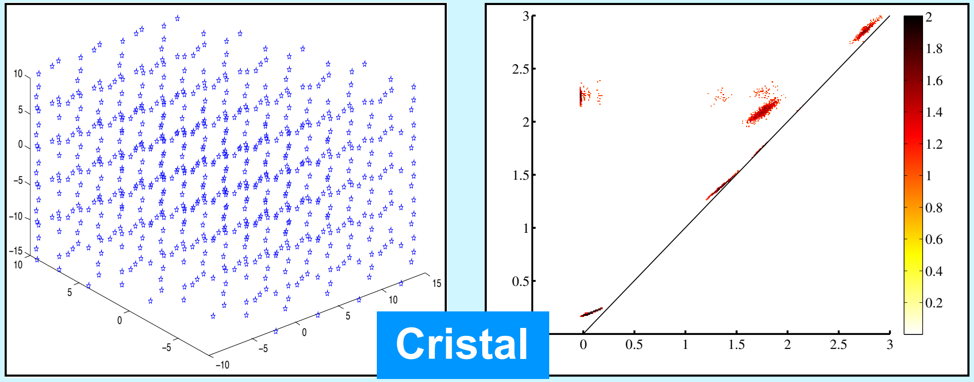
deformation of octa.

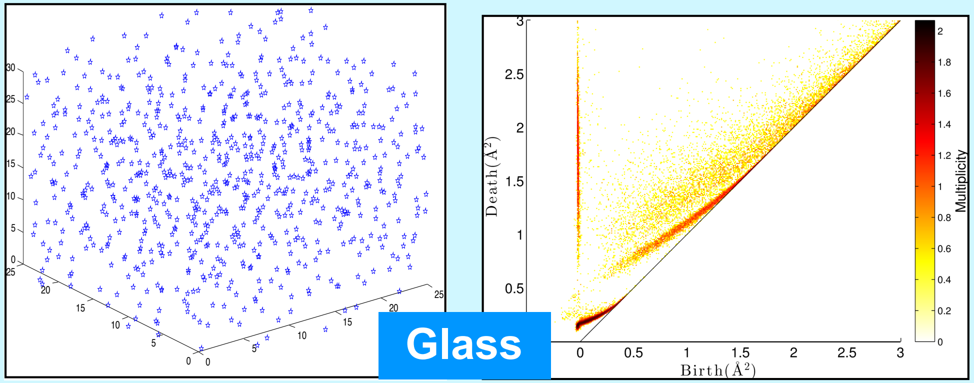


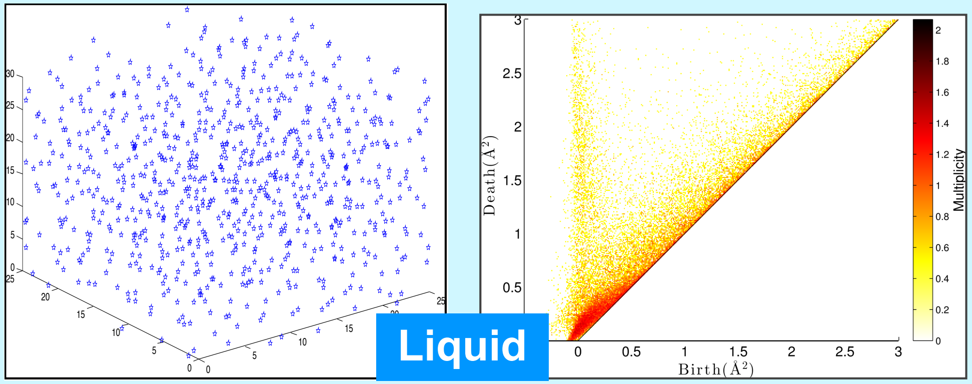
Hierarchical Structural Analysis of Silica Glass

with Nakamura, Hirata, Escolar, Matsue, Nishiura   PNAS (2016)   CREST TDA, SIP

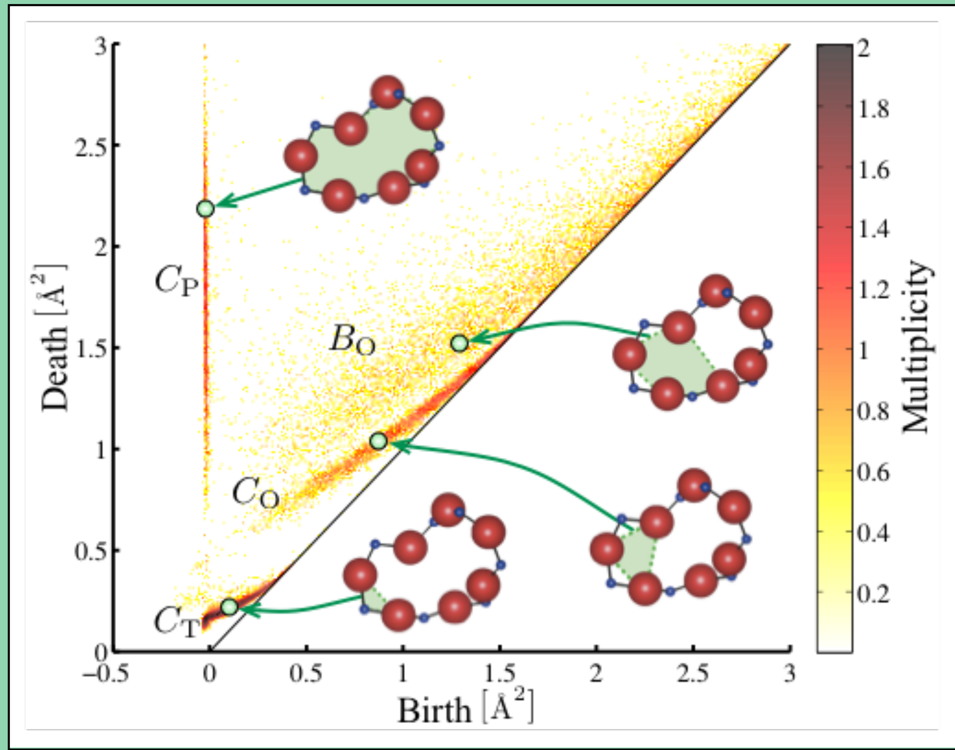
MD and PD1







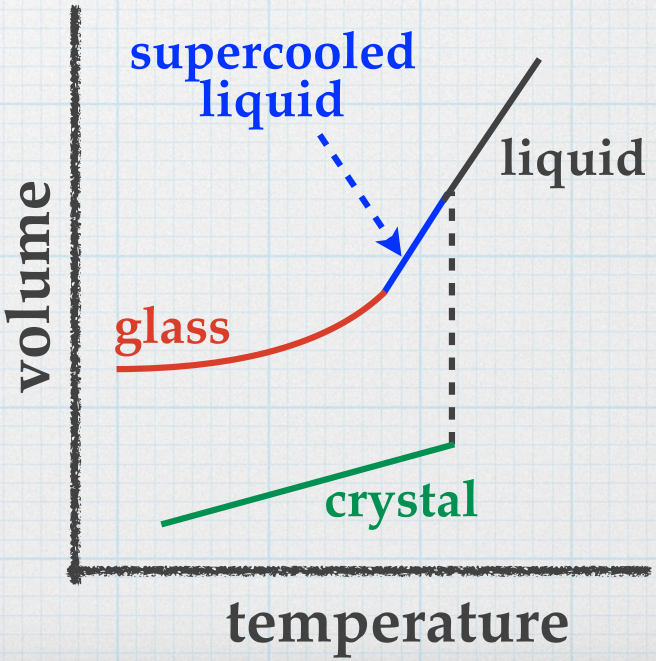
Inverse Analysis



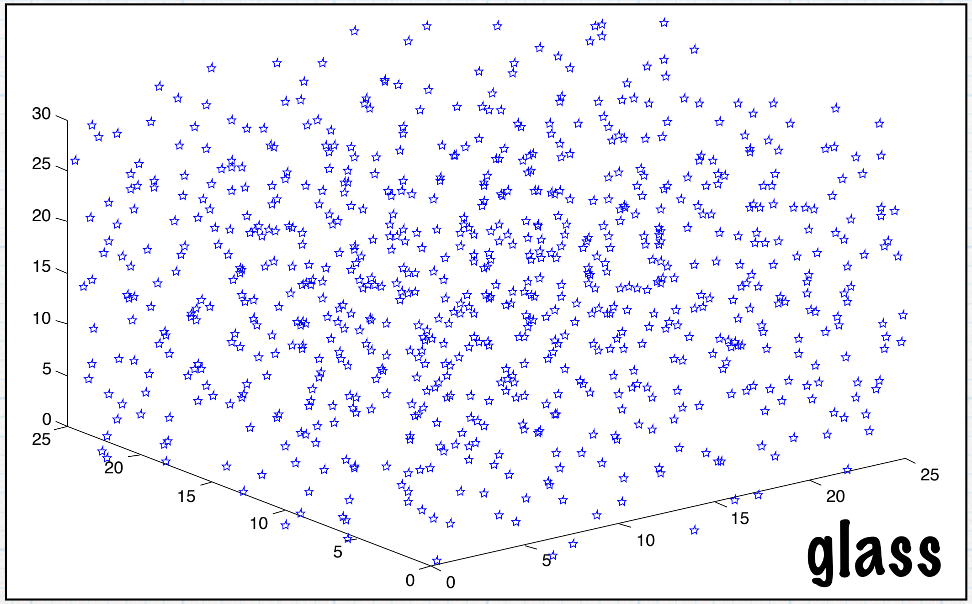
- Glass contains curves in PD
- Curves express geometric constraints (orders) of atomic configurations
- Inverse analysis reveals hierarchical ring structures
- PD multi-scale analysis characterizes inter-tetrahedral O-O orders (curve Co)
- universal tool for structural analysis

What is glass?

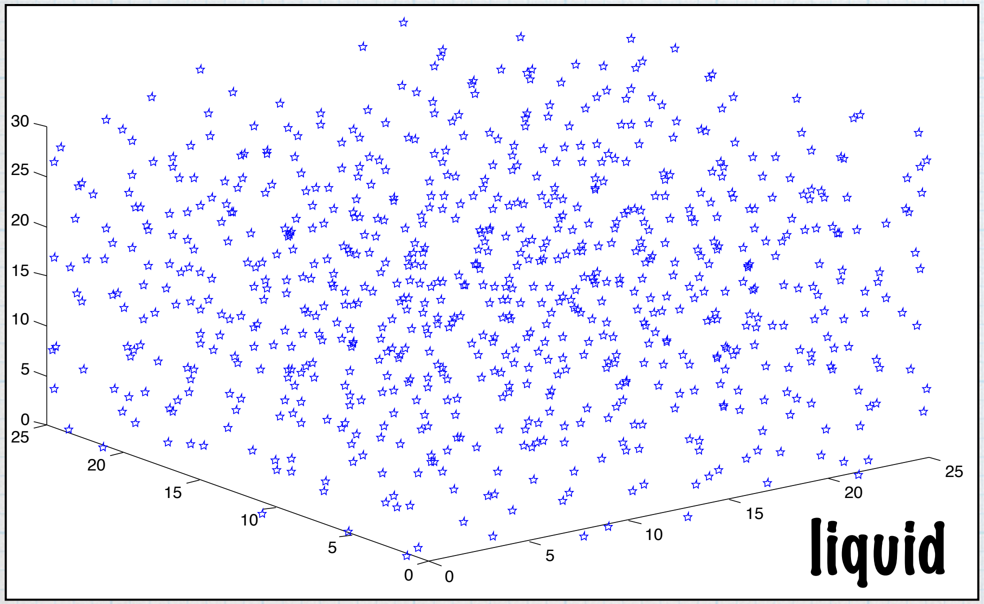
- Not yet fully answered to "what is glass?"
- Not liquid, not solid, but something in-between
- Atomic configuration looks random, but sufficient cohesion to maintain rigidity
- Further geometric understandings of atomic configurations are required
- Solar Energy Glass, DVD, BD, etc.

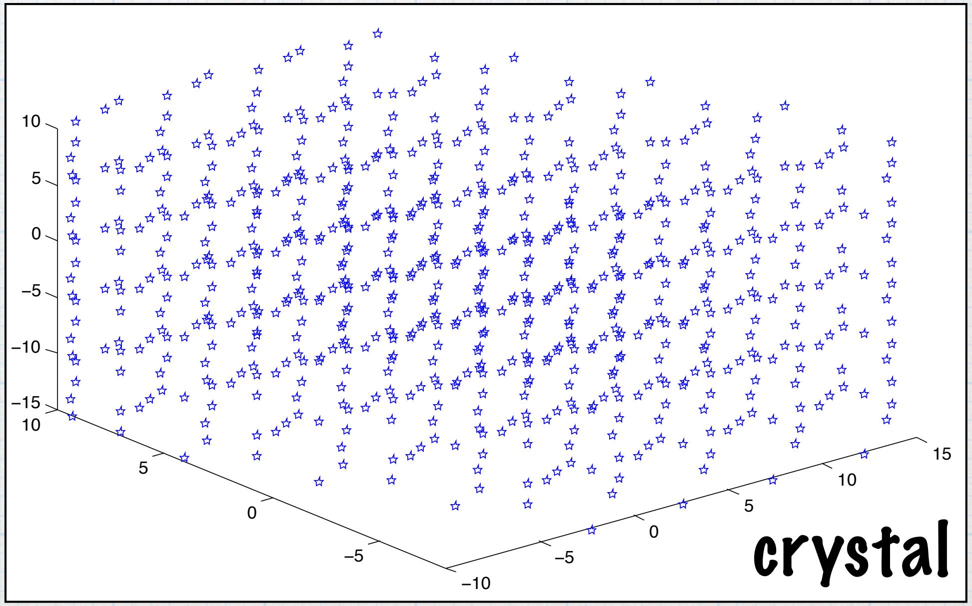


Atomic configurations of silica (SiO2)



obtained by molecular dynamics (MD) simulation

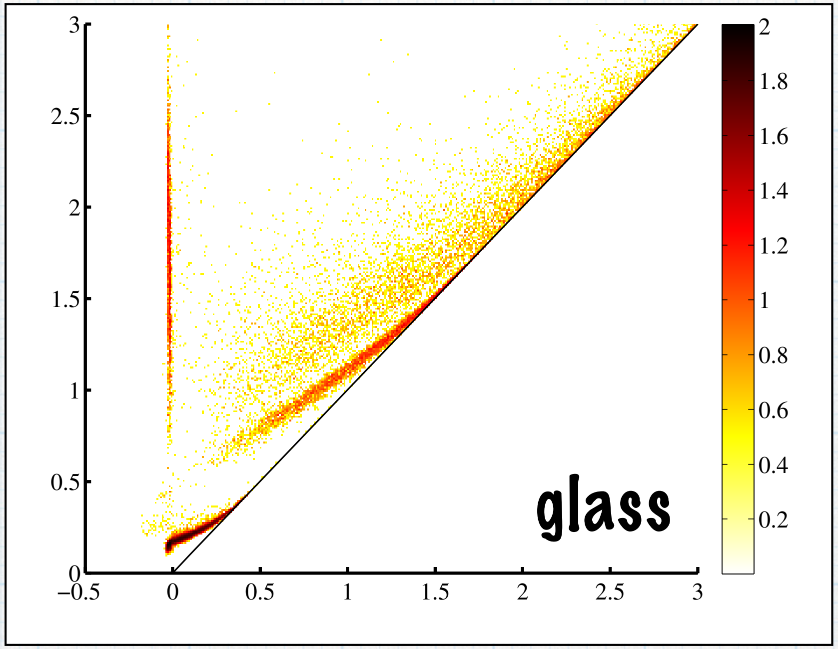




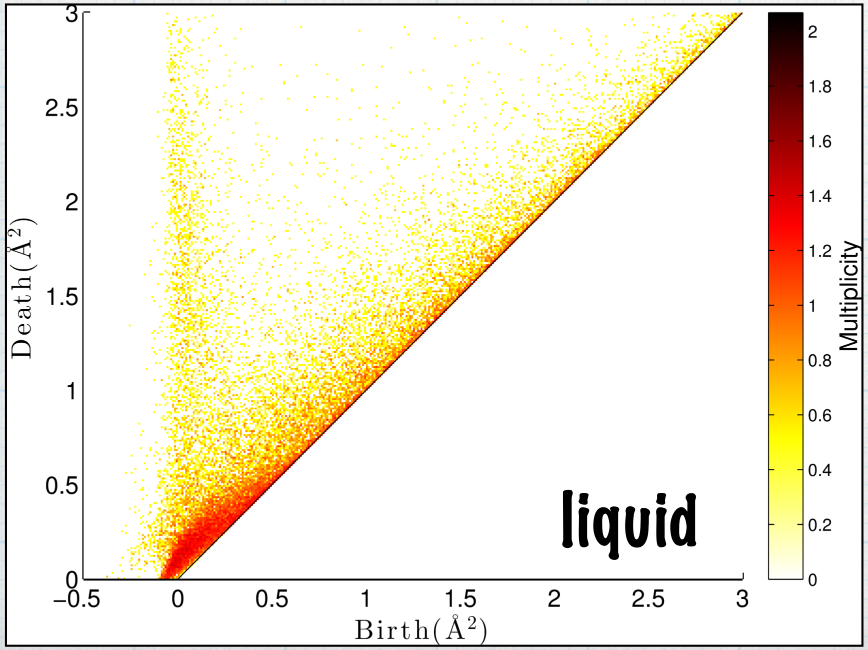
Y.H., et al. PNAS (2016)

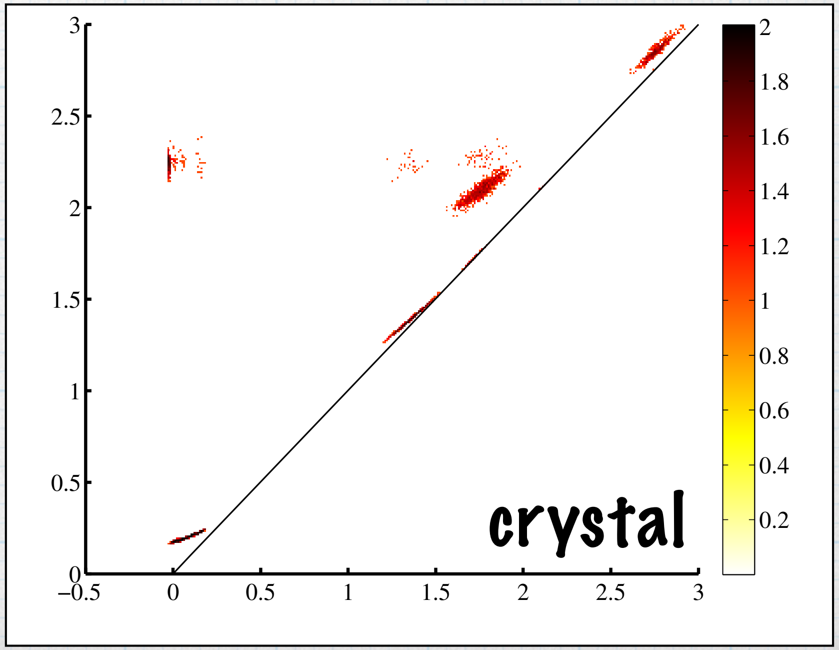
1 dim persistence diagrams of silica

All PD computations today are performed by CGAL, DIPHA (Thanks!)

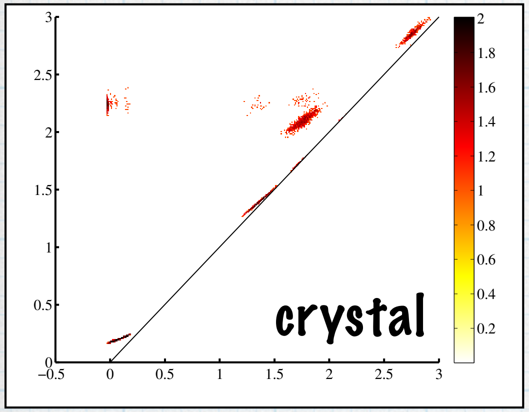


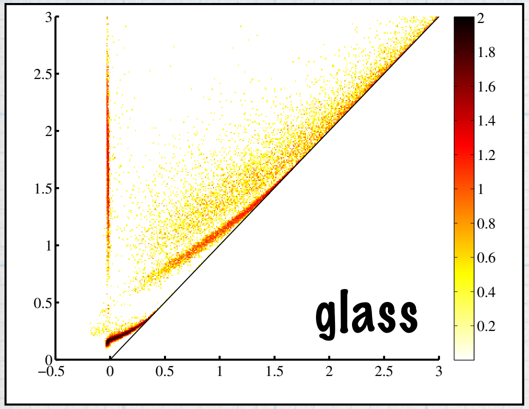
PD1 of weighted alpha filtrations

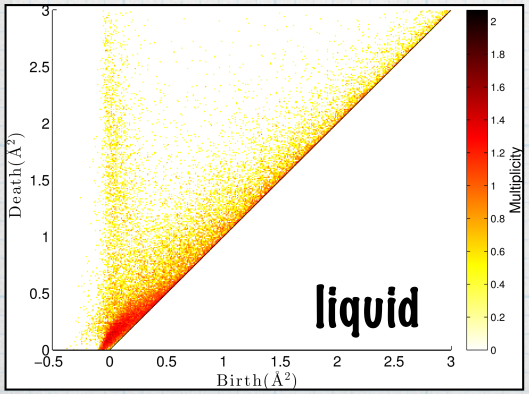




Support dim and order parameter







0-dim support results from **periodic** atomic locations of crystals

transition

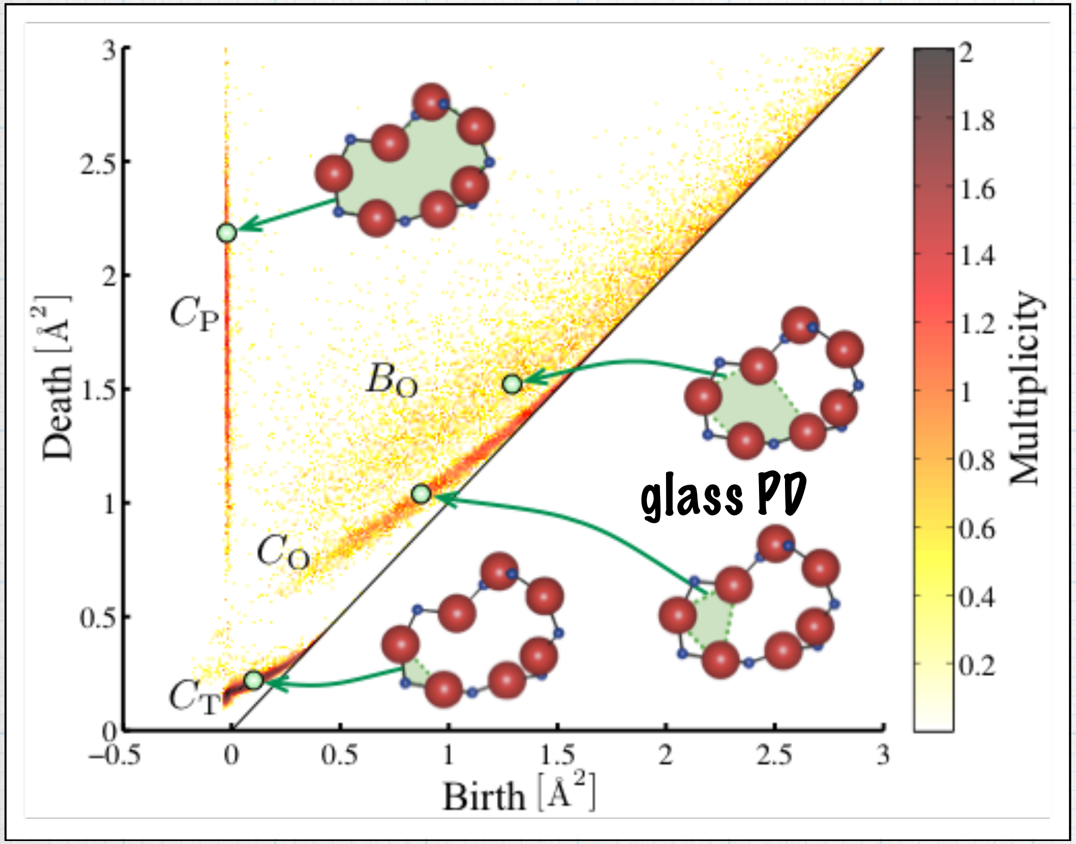
1-dim support (curves!) appears

transition

2-dim support results from **random** atomic locations of liquids

Remark: limiting PD in glass and its support may be related to phase transition

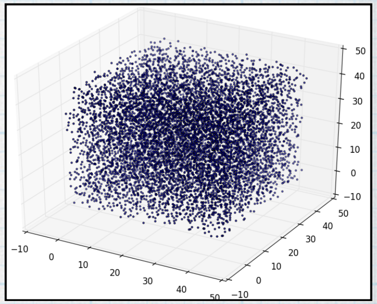
Geometric origins of curves: inverse problem



what characterizes glass structures?

what is the geometric origin of the curves?

inverse problem!

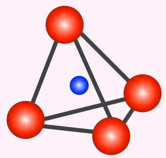


- optimal cycle
  Dey, et.al, 2011, Escolar and H, 2015
- continuation
  Gameiro, Obayashi, H. Physica D, 2015

hierarchical ring structure

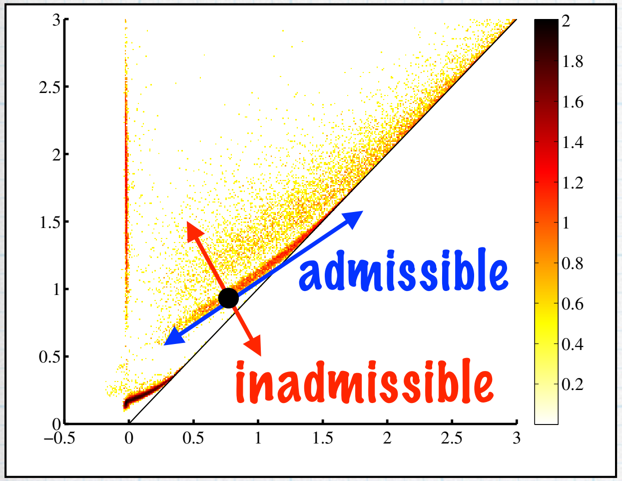
C_P: primary rings generating the others → **C_O: three oxygen rings**

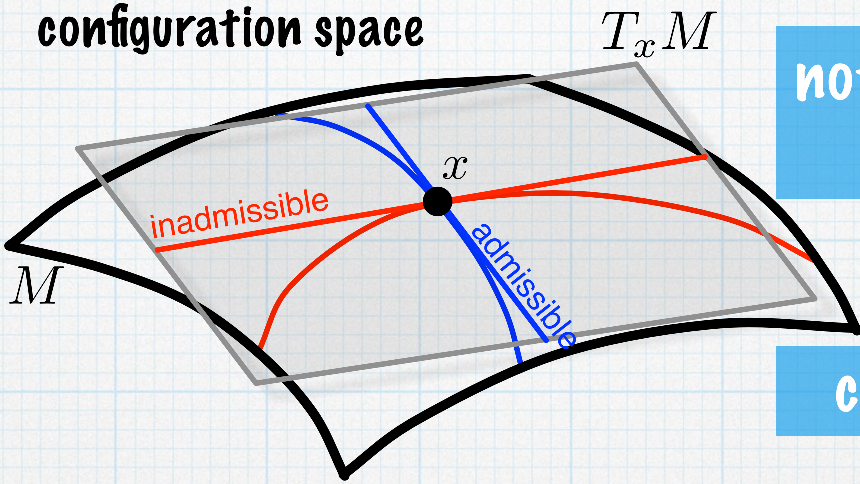
C_T: triangles on tetrahedra



B_O: oxygen rings ($\geq$ four)

Curves and constrains



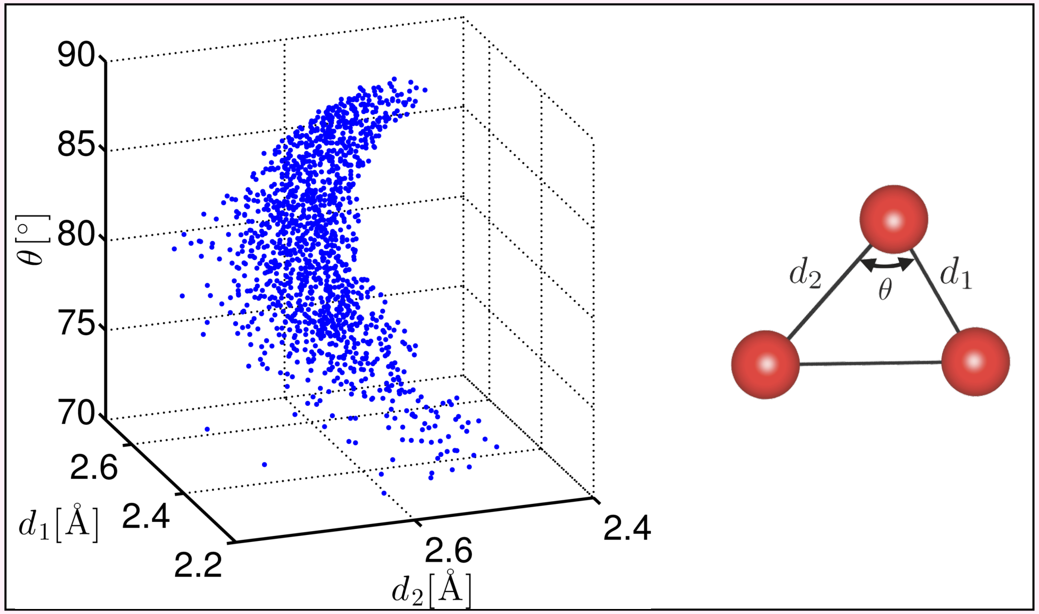


not allowed to move in the normal direction

rigidity

constrains of the rings

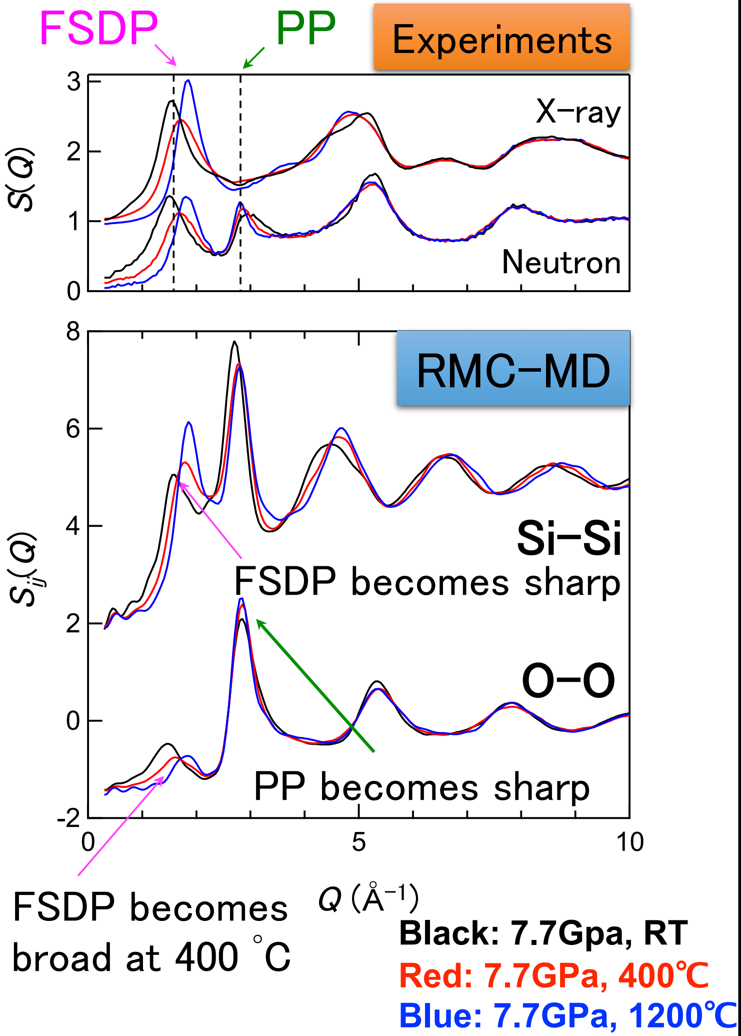
O-O-O structures (MRO)



- **O-O-O ring constrains are discovered**
- **necessary to study both distance and angle distributions simultaneously (conventional methods cannot detect)**

Densified silica glass in high pressure and temperature

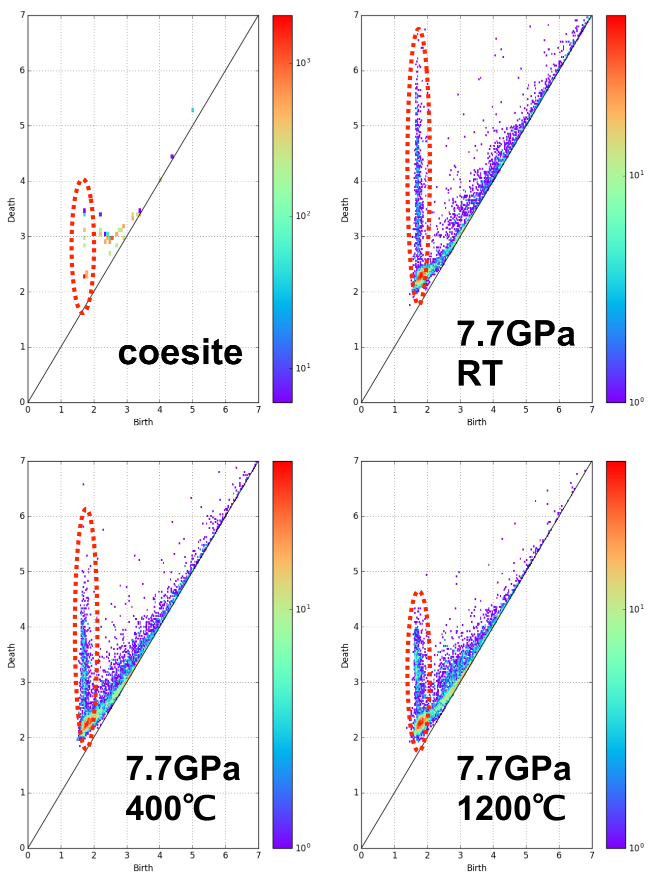
with Kohara (NIMS), Hirata, Obayashi (AIMR)   MI^2I (Innovation Hub), CREST TDA

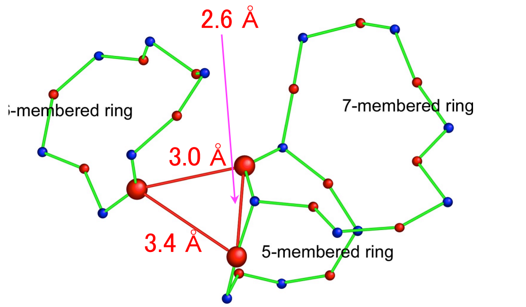


- PP of O-O correlation becomes sharp with increasing temperature
- conventional methods could not explain this behavior

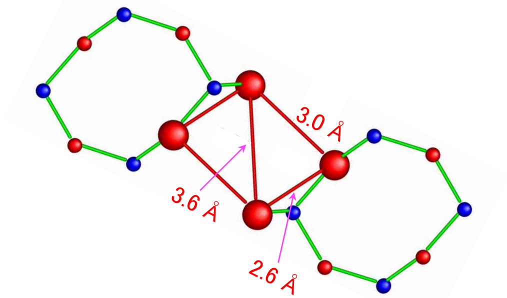
➡ what is the geometric origin?

O-PD1





inter-tetrahedral O-O ordering in 7.7GPa, 1200℃

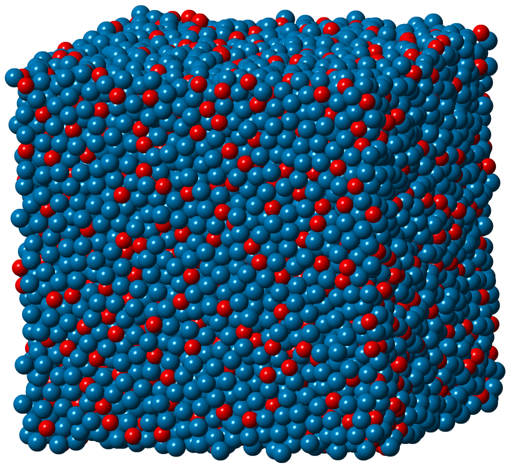


ordering of oxygen in coesite

- PDs become sharper like PP, and show the increase of packings of oxygens at high temp.
- Oxygen PDs ascribe for the first time O-O ordering between different SiO4 tetrahedra to PP
- The geometric origin of PP ordering is coesite-like rings

Metallic Glass: geometric origin of distorted icosahedra

with Hirata, Obayashi, Takeuchi (AIMR)   CREST TDA

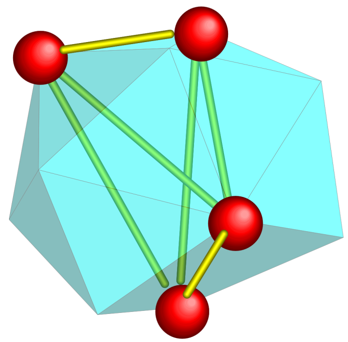


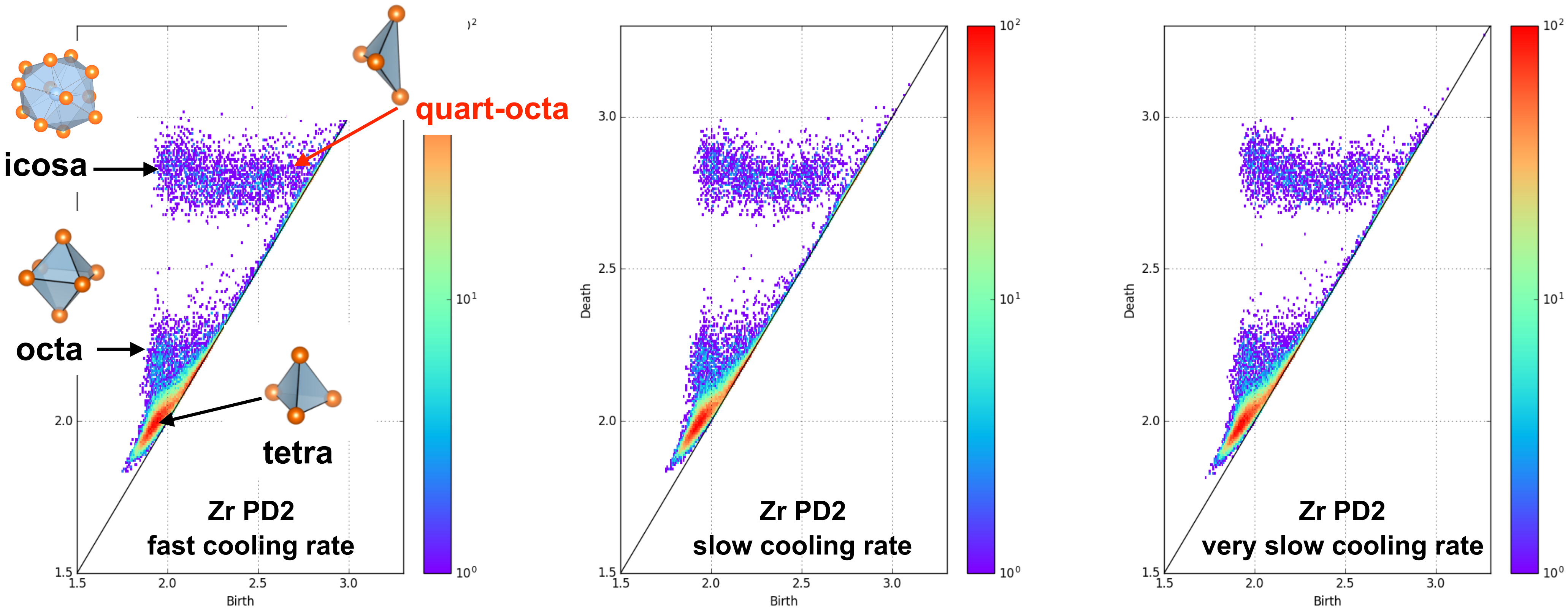
Zr80Pt20

# of atoms = 12,000

- deformation of icosa is detected as a curve
- distorted icosa can be characterized by quart-oct
- the slower the cooling rate, the less the distorted icosa and quart-octahedra

distorted icosahedron with a quart-octahedron



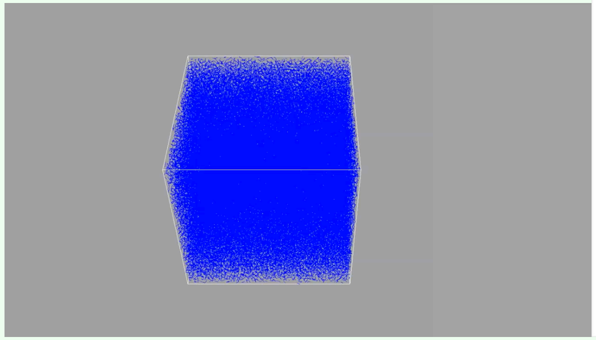


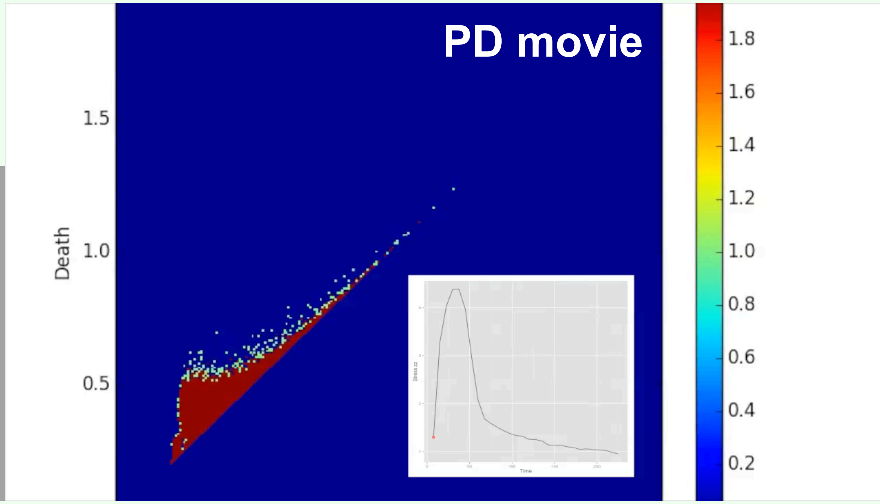
Craze formation of polymers

with Ichinomiya, Obayashi PRE (2017) SIP, NEDO

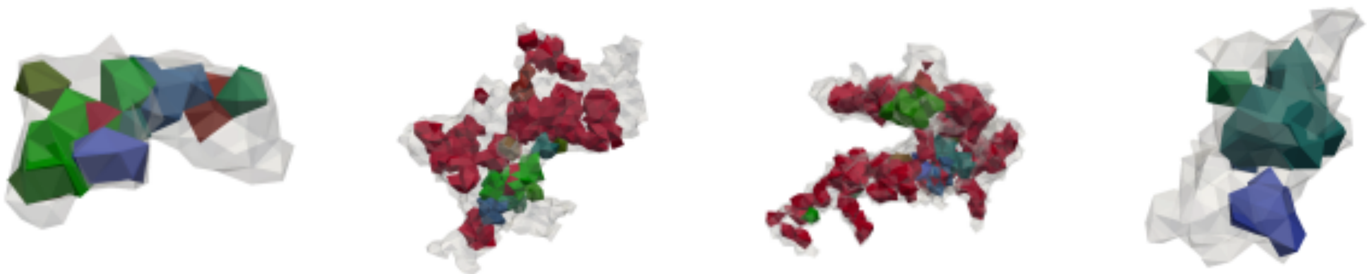
Kremer-Grest model

uniaxial deformation



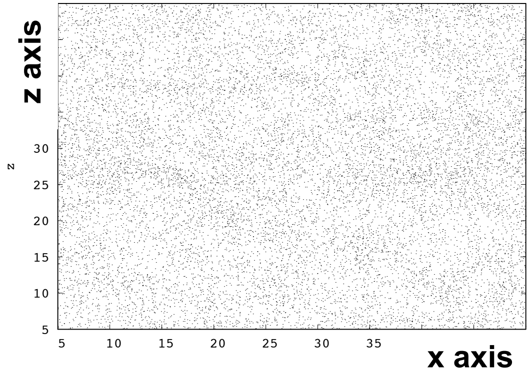


void coalescence during craze formation

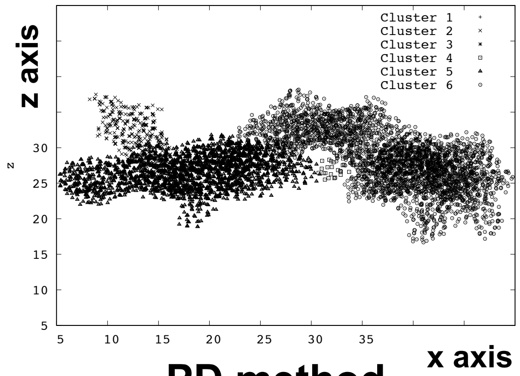


- gray voids are large voids observed after yielding
- colored voids are initial micro voids generating large voids

craze position



Voronoi volume (conventional)



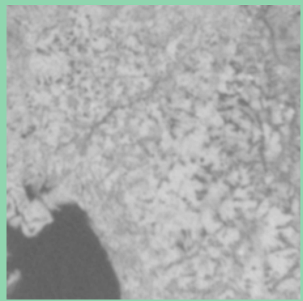
PD method

- detect large voids from PD movie as generators with large death values
- explore initial config. of large voids by reversing time with inverse PD method
- large voids are generated by coalescence of micro voids (void percolation)

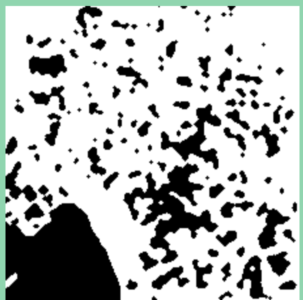
Materials Informatics: Machine Learning on PDs

with Kimura (KEK), Obayashi (AIMR) SIP, CREST TDA

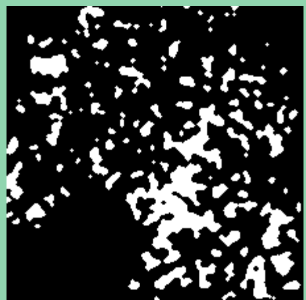
X-CT of iron-ore sinters



original



iron oxide



calcium ferrite (CF)

Trigger site of micro cracks are supposed to be related to hetero-structure of iron oxide and CF. No descriptors have been developed so far.

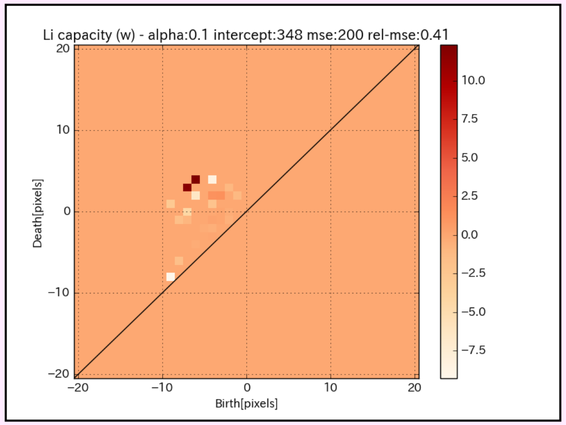
background

- **large amount of experimental images are available**
- **want to find a compact descriptor to connect images to materials properties (cracks, elasticity, conductivity etc)**

➡ **develop a method of image analysis using big data**

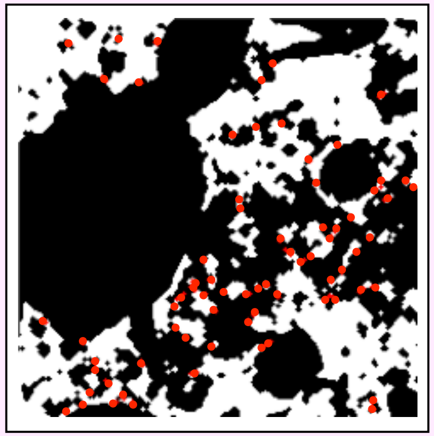
our approach

- **PD for compact descriptor of images**
- **ML for combining with big data**



LASSO (Sparse PD)

inverse



detected trigger site of cracks

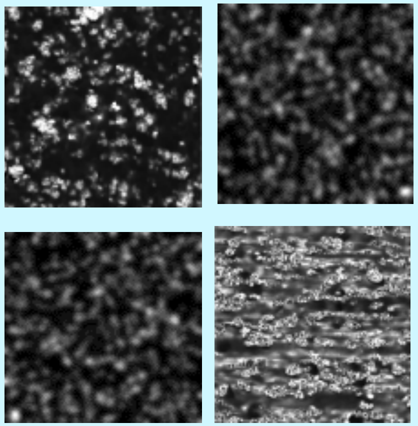
Statistical inverse analysis on persistence diagram

with Obayashi (AIMR) arXiv:1706.10082 CREST TDA, SIP, NEDO, MI^2

Background

- **PDs are good descriptors in materials science**
- **Want to extract statistical features in the dataset of PDs**
- **Vectorization of PDs are necessary for applying machine learnings (persistence landscape, persistence image, PSSK, PWGK, etc)**
- **Want to study the original data space (inverse problems)**

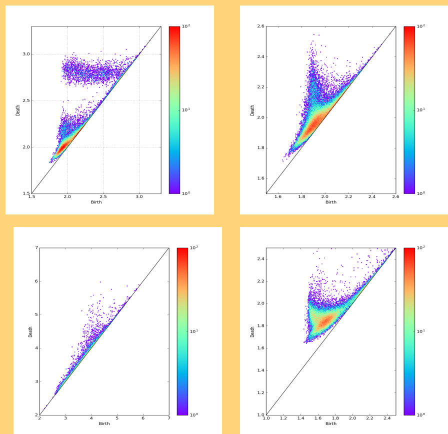
Dataset of inputs



direct →

← inverse

Dataset of PDs



direct →

← inverse

Machine learning

- SVM
- PCA
- Regression
- Anomaly detection
- Time series
- etc

Study machine learning models based on persistence diagrams

Vectorization: persistence image
ML: Logistic regression, Linear regression (LASSO/RIDGE)

今後の展望

代数

- multiparameter persistent homology の理論整備
- 表現論 (浅芝, 吉脇, Escolar), sheaf (Curry), microlocal analysis (Kashiwara-Schapira)の展開

確率論

- パーシステント図に対する極限定理（白井，Duy，角田）
- multiparameter persistent homologyへの確率論的視点

幾何

- 測度距離空間としてのパーシステント図の集合の幾何構造（PD空間）

力学系

- 有限サンプル点上で定まる力学系解析（竹内，Edelsbrunner, Jabłoński, Mrozek）

統計・機械学習

- 時系列解析とmultiparameter persistent homology
- 機械学習の性能解析（福水，草野，大林）

逆問題

- 実現可能なパーシステント図（realizable PD）は？PD空間と逆問題
- 最小生成元（optimal cycle）の高速計算（大林）

ソフトウェア開発

- 高速化、多機能化（大林、須藤、阿原）

応用

- より踏み込んだ応用（材料、生命、脳、気象、医療、経済 *etc*）．数学へのフィードバック

参考文献

- 平岡裕章．位相的データ解析とパーシステントホモロジー. 日本数学会『数学』68, 361-380 (2016).
- 平岡裕章，タンパク質構造とトポロジー: パーシステントホモロジー群入門，共立出版 (2013).
- Carlsson. Topology and Data. Bulletin of AMS. 2009.
- Edelsbrunner and Harer. Computational Topology. AMS. 2009.
- Ghrist. Elementary Applied Topology. CreateSpace Independent Publishing Plat- form, 2014.
- AYASDI. http://www.ayasdi.com