



# 位相的データ解析の基礎と応用

平岡 裕章

Yasuaki Hiraoka

東北大学

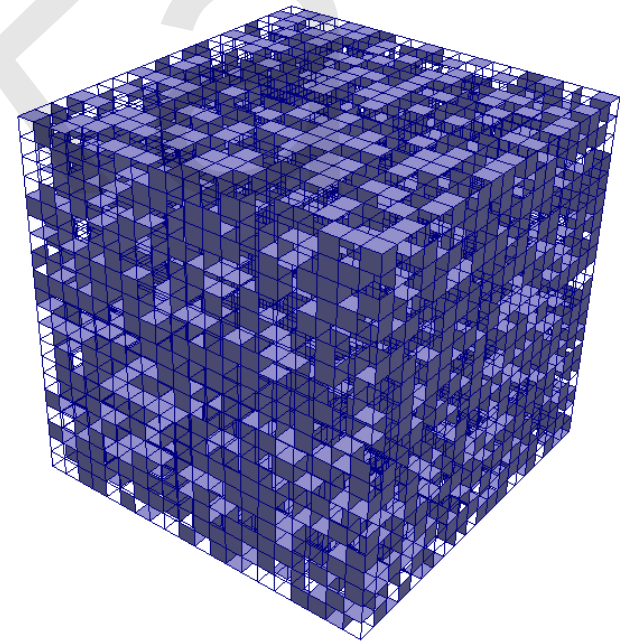
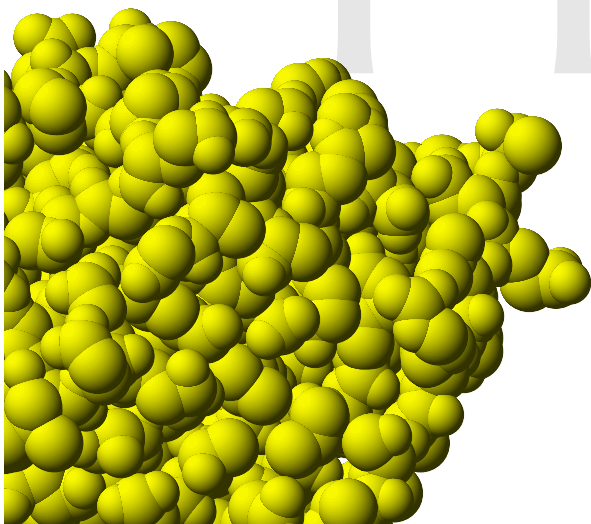
原子分子材料科学高等研究機構

WPI-AIMR

JST CREST

SIP革新的構造材料

JST イノベーションハブMI<sup>2</sup>I

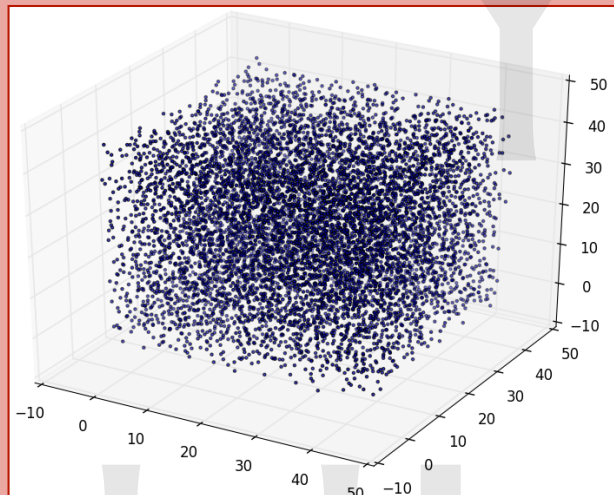


# 背景：Shape of Data

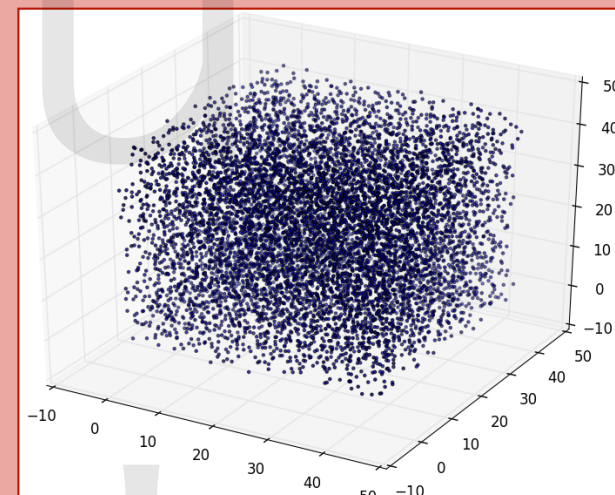
背景①：インターネット・計算機の成長によるデータの氾濫

背景②：データ科学・AI（人工知能）の発展によるデータの潜在価値抽出の必要性

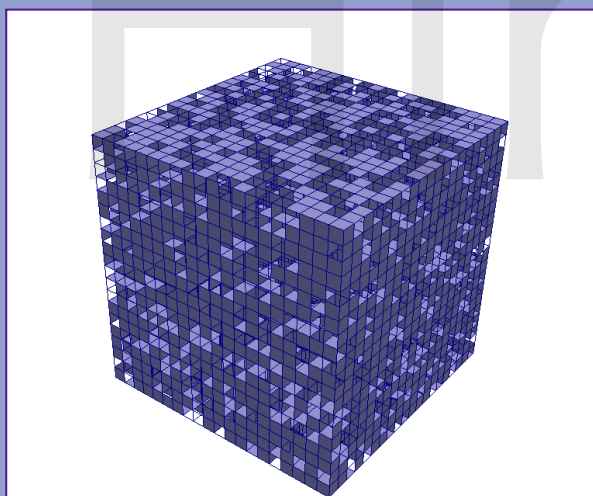
## 材料構造データ



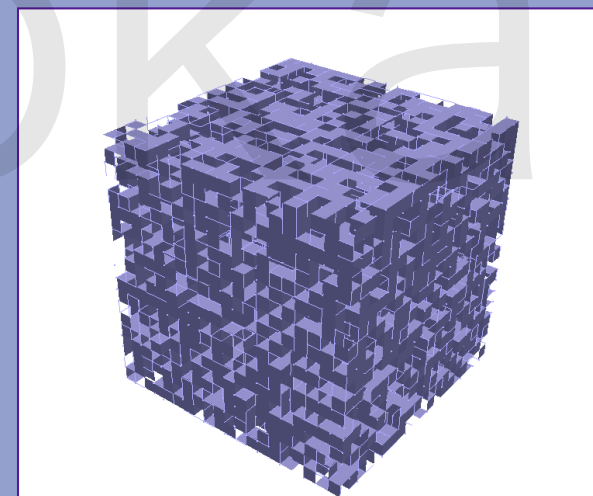
どう違う？



## 3次元画像データ

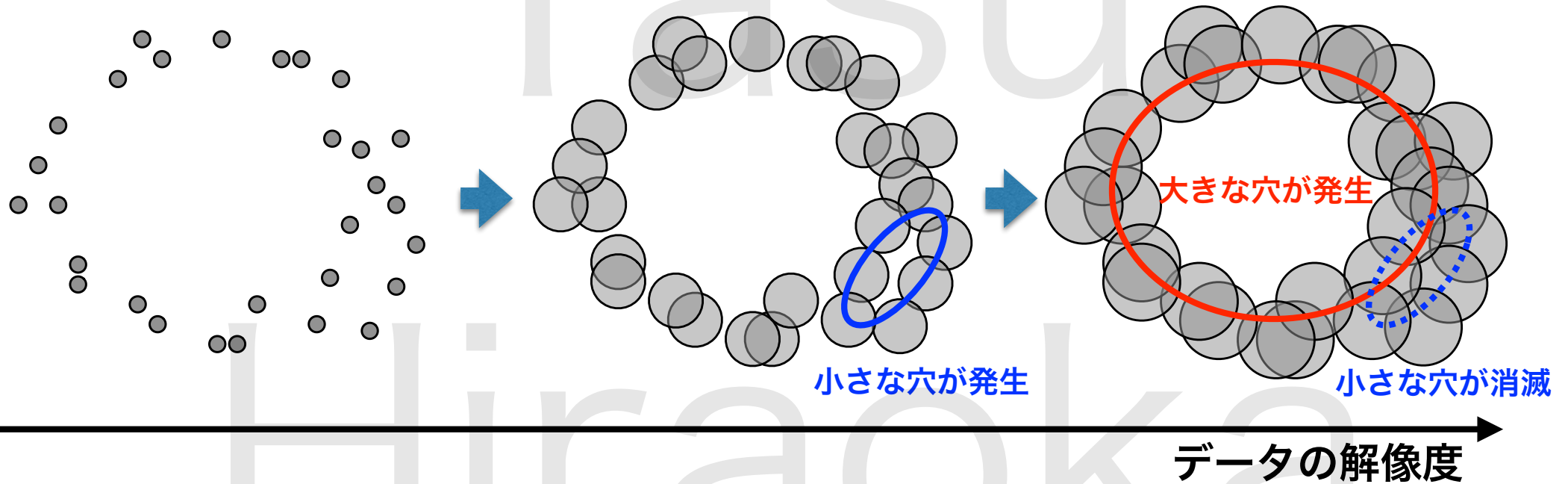


どう違う？

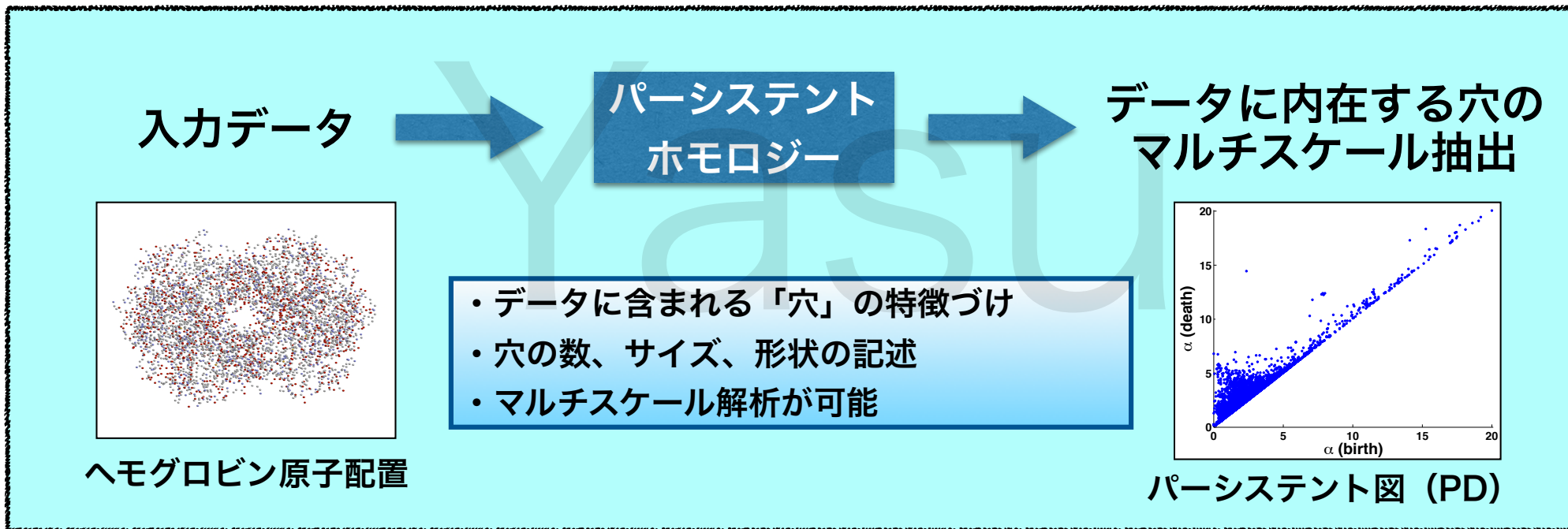


データの形に着目した**数学的**解析手法の開発

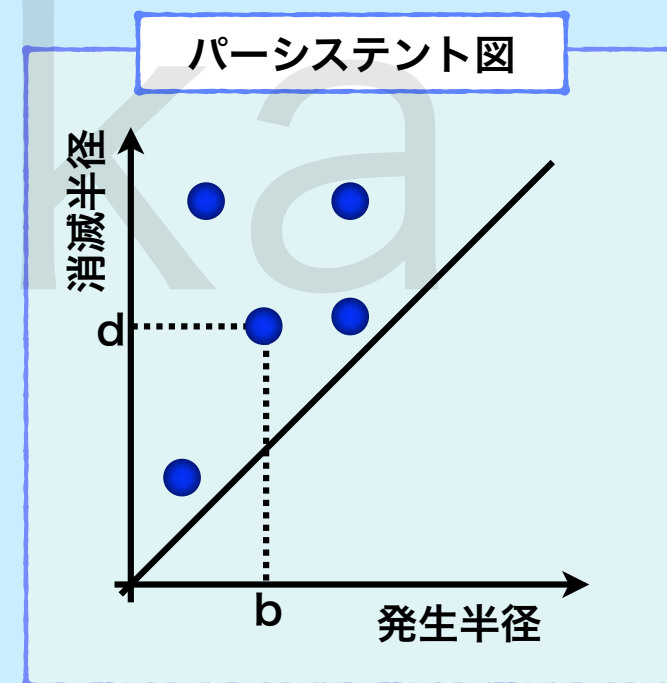
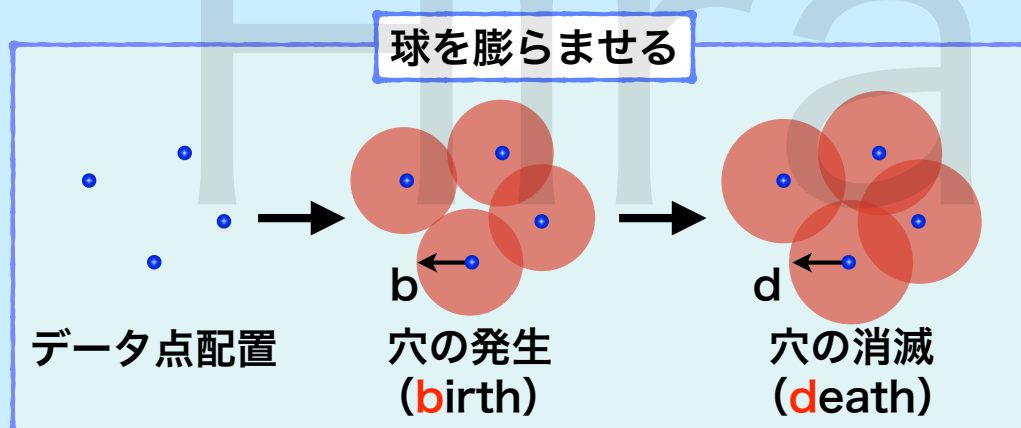
入力データ



- 離散点データを太らせる（解像度を変える）
- データの解像度を変えてマルチスケールで形をとらえる
- 穴の発生と消滅による形の特徴づけ

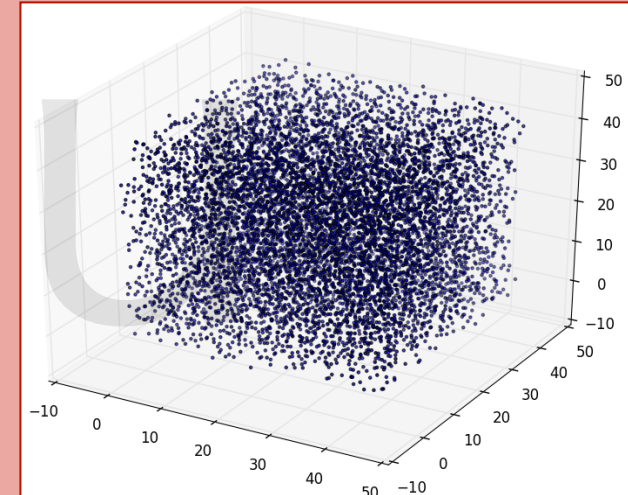
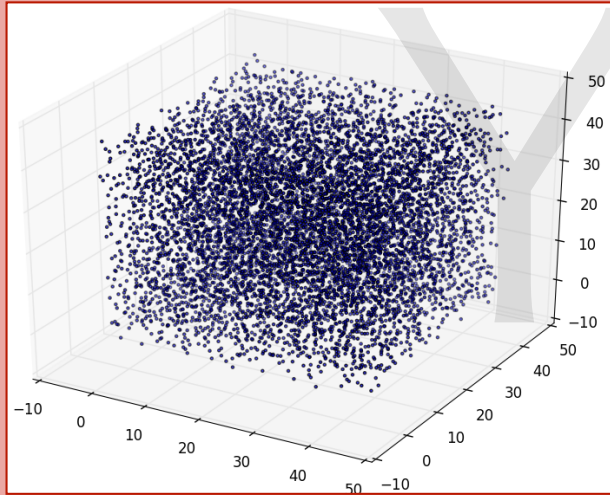


## パーシステント図構成法

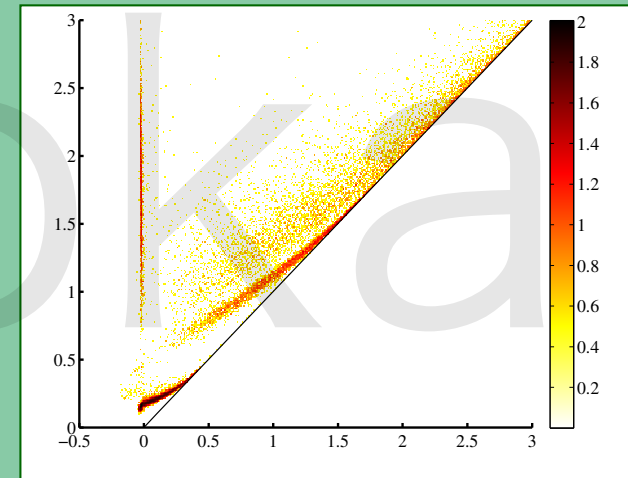
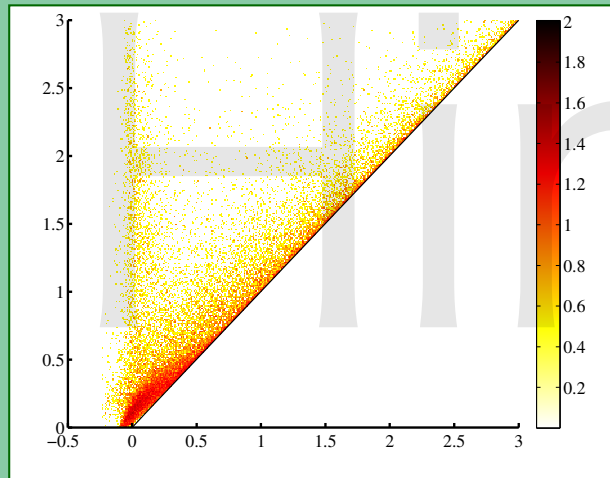


- パーシステント図の各点はデータ内の穴
- 発生軸は穴の発生パラメータ
- 消滅軸は穴の消滅パラメータ
- 対角線付近の点はノイズ
- 対角線から離れた点はロバスト

## 材料構造データ



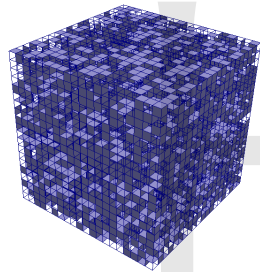
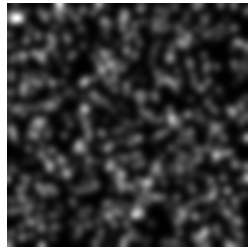
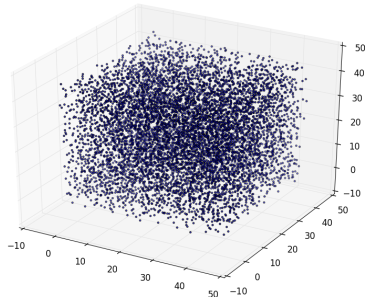
## パーシステント図



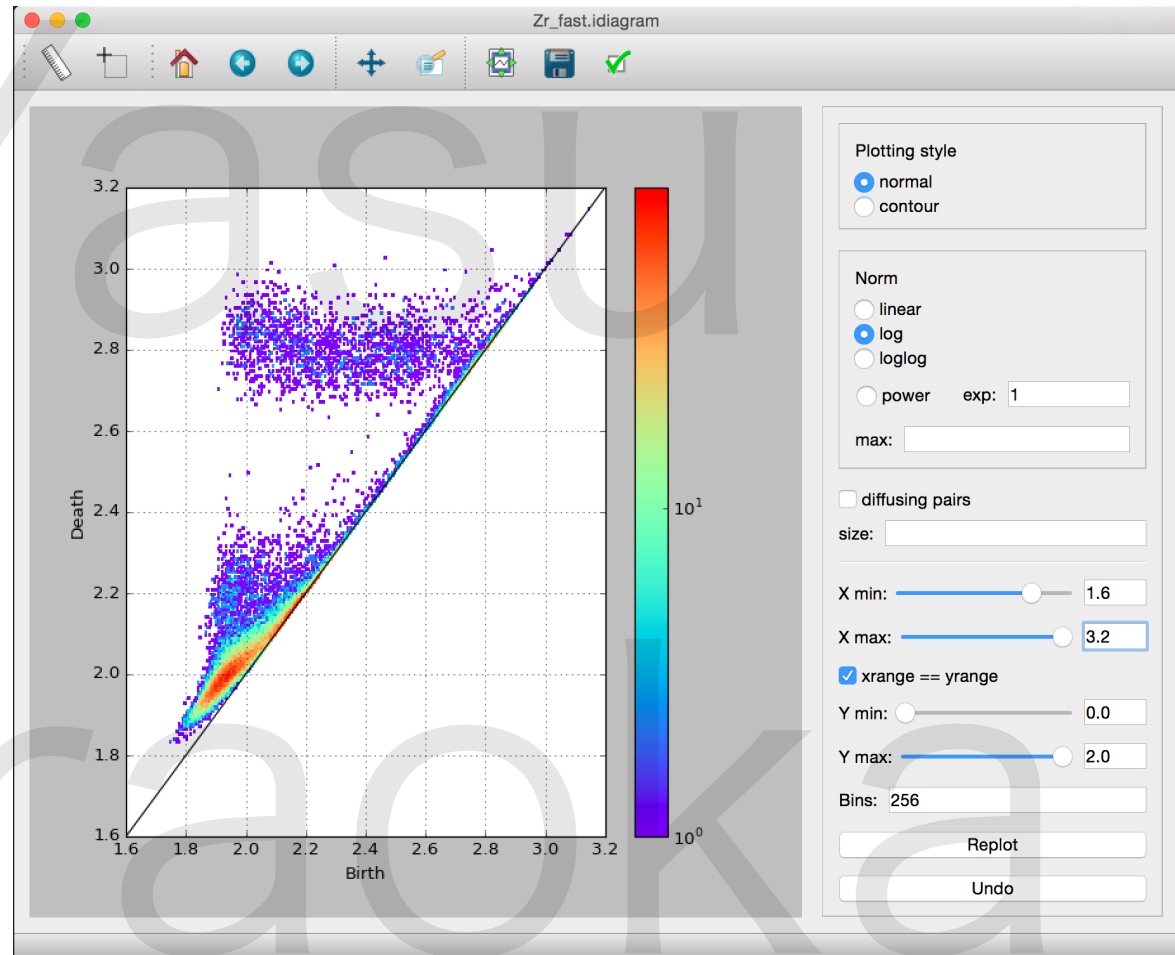
- 曲線の存在による幾何学的秩序の検出
- 逆問題を解くことで曲線に対応する幾何形状の特定
- 材料構造・機能相関を調べる新言語の提案
- 画像データに対しても同様に展開可能
- 機械学習・AIと融合させた新データ解析手法へ発展

応用例：ガラス構造解析、高分子劣化・破壊解析、粉体結晶化現象、画像解析 etc

## 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\\_lab/index.html](http://www.wpi-aimr.tohoku.ac.jp/hiraoka_lab/index.html)

# 位相的データ解析

位相的データ解析 (Topological Data Analysis, TDA)

今世紀に数学者が開発した強力なデータ解析手法

**Data Has Shape, Shape Has Meaning, Meaning Drives Value**

**Gunnar Carlsson (math. Stanford, AYASDI)**

- ビッグデータ, ソーシャルネットワーク, 医療, 金融 etc

**Robert Ghrist (math. UPenn)**

- 情報ネットワーク, センサーネットワーク

**Konstantin Mischaikow (math. Rutgers)**

- 流体解析, データ時系列解析

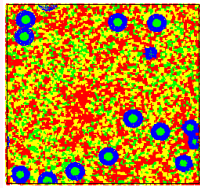
**東北大WPI-AIMR, CREST TDA, SIP, MI<sup>2</sup>**

- 材料科学 (ガラス, 粉体, 高分子, 金属, etc)

# Materials TDA

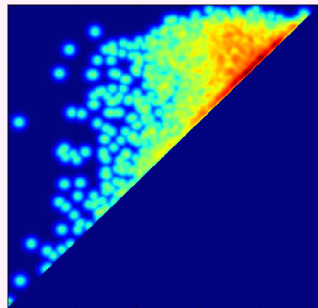
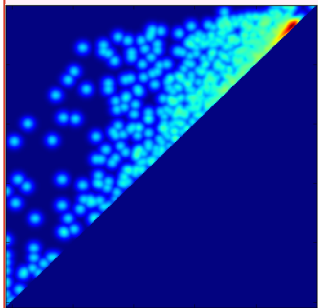
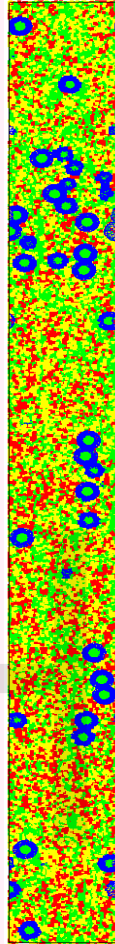
WPI-AIMR, CREST, SIP, MI<sup>2</sup>I, 超超 (NEDO)

## 高分子



AFM image

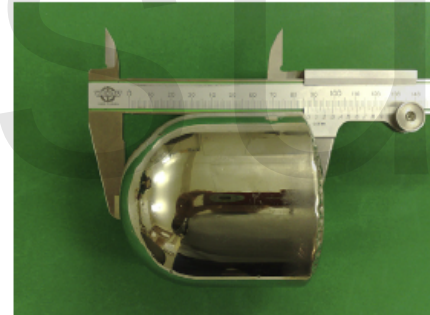
expansion



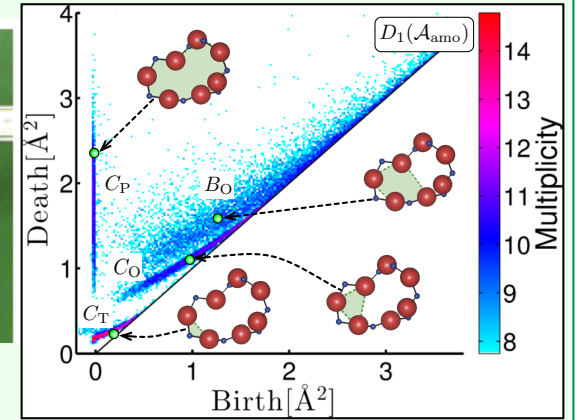
## アモルファス Y.H., et al. PNAS (2016)



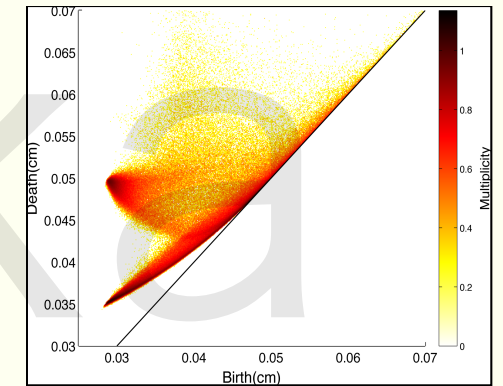
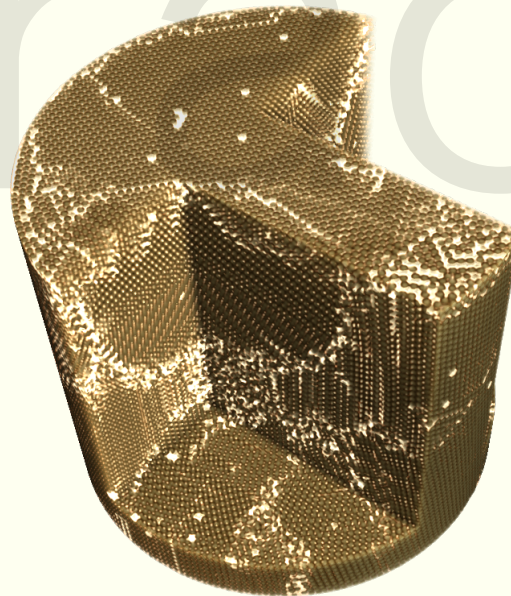
silica



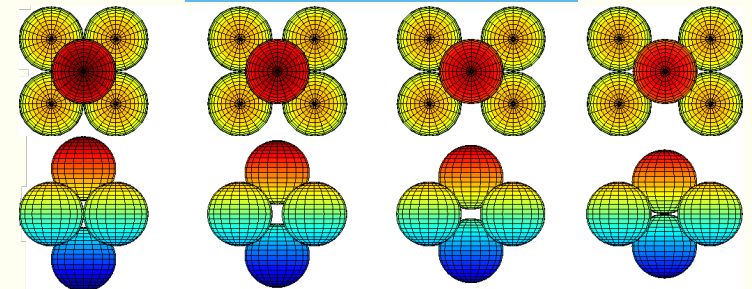
metallic glass



## 粉体 To appear in Nat. Comm.



deformation of octa.



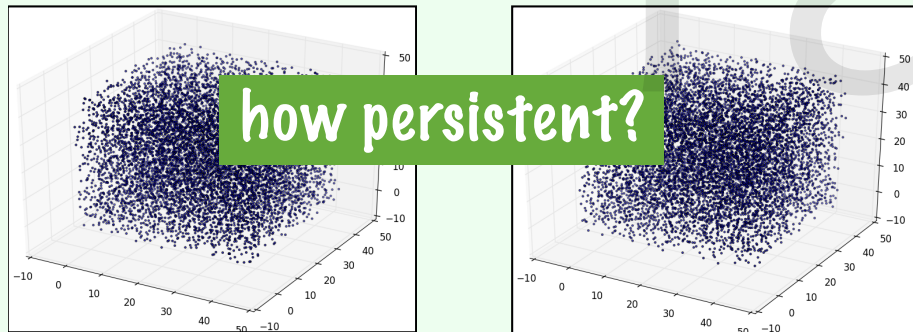


# From materials science to math

representation theory

glass  $\longrightarrow$  compressed

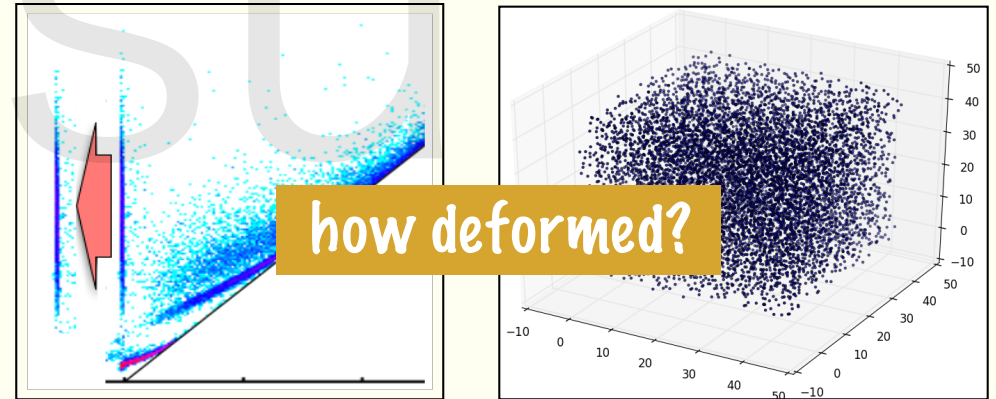
with 浅芝



study multi-persistence (spatiotemporal, compression, etc)

inverse problem

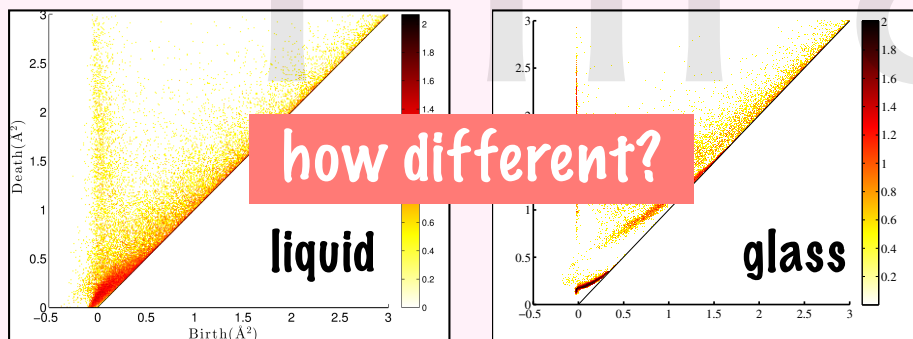
with 大林



study inverse problems (materials design, rigidity, etc)

random topology

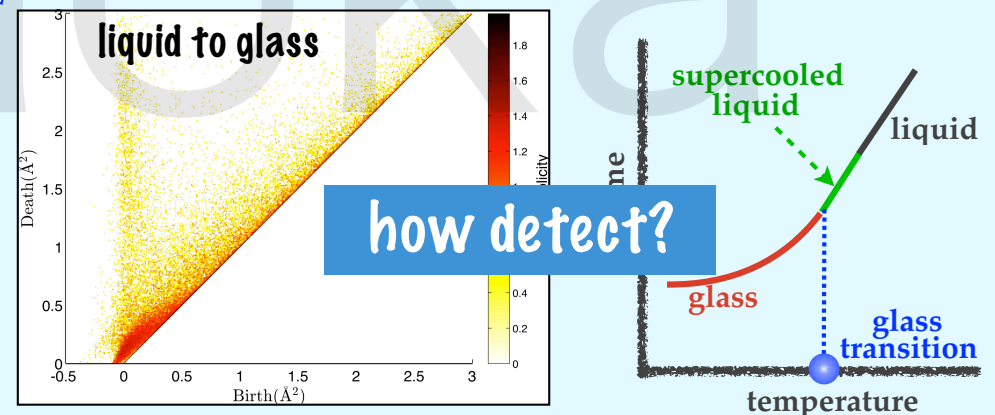
with 白井



random topological features and phase transition

topological statistics

with 福水



detect glass transition and capture principal behaviors

# 講演内容

1. ホモロジー入門
2. パーシステントホモロジー入門
3. 材料科学への応用例の紹介

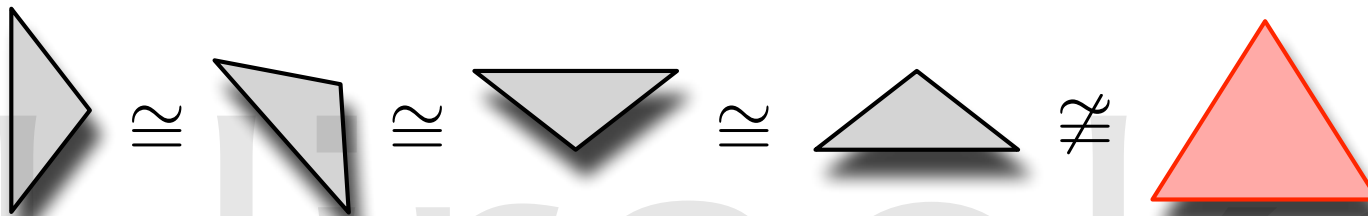
# 1. ホモロジー入門

## 1.1 トポロジーと穴

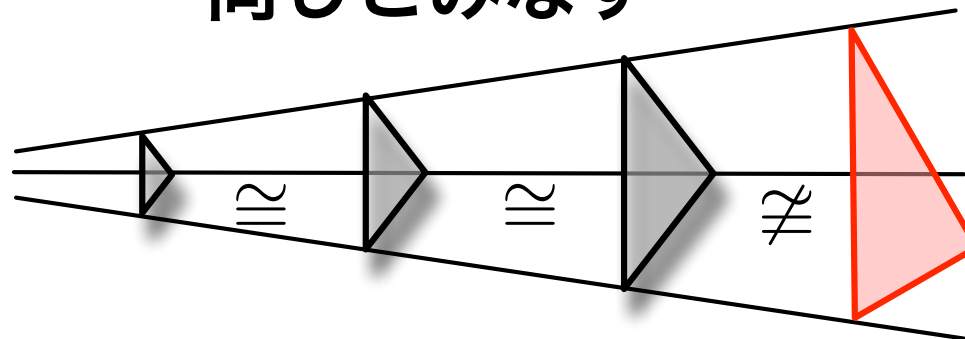
幾何学：ものの形を研究する数学の分野

図形の分類の仕方は**基準**の定め方によって様々

★ 基準「合同」：図形が重なるときは同じとみなす

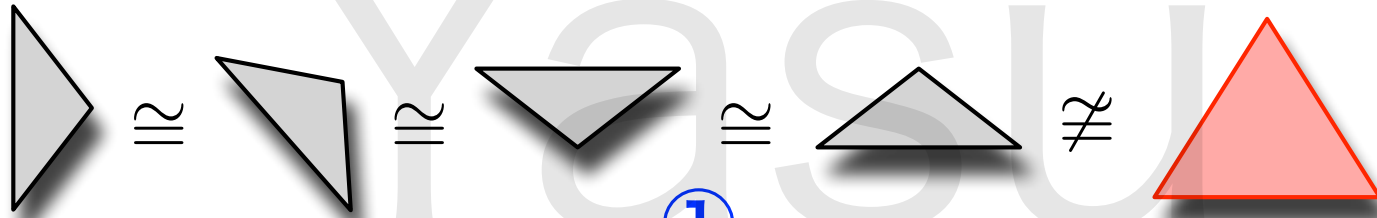


★ 基準「相似」：スケールをかえて重なる図形は同じとみなす



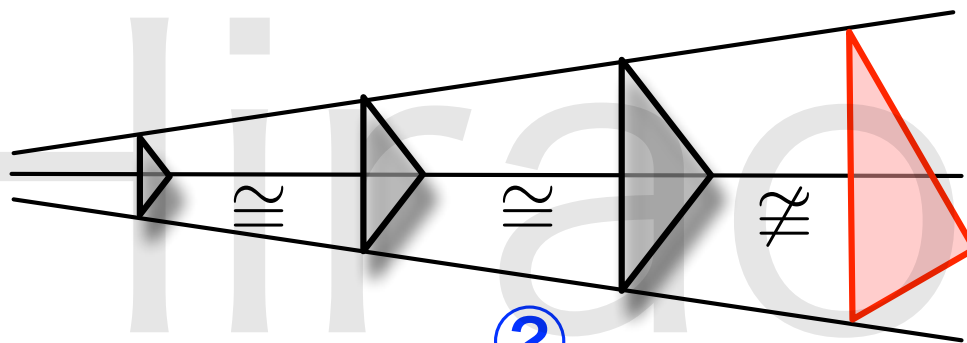
代数的 (定量的) には, , ,

★ 合同 : 図形が重なるときは同じとみなす



↔ ① 3辺の長さが同じ 3角形は同じと見なす

★ 相似 : スケールをかえて重なる図形は同じとみなす



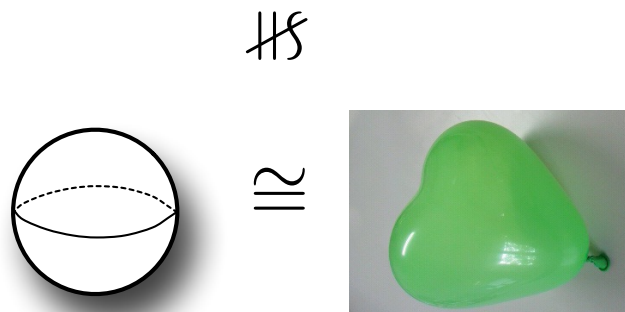
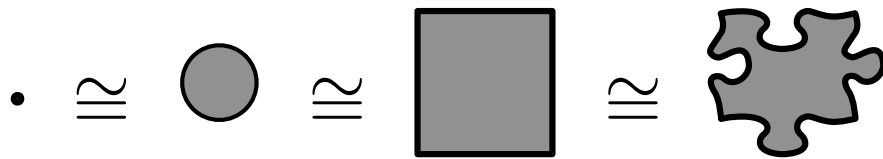
↔ ② 2組の角が同じ 3角形は同じと見なす

- ①や②は代数的不変量とよばれるものの例
- それぞれの代数的不変量を調べることで図形の分類が可能

トポロジー：幾何学の1分野でその基準は

「切り貼りしない連続変形で重なり合う図形は同じもの」

で与えられます



# トポロジー = 切り貼りしない連続変形はOK!

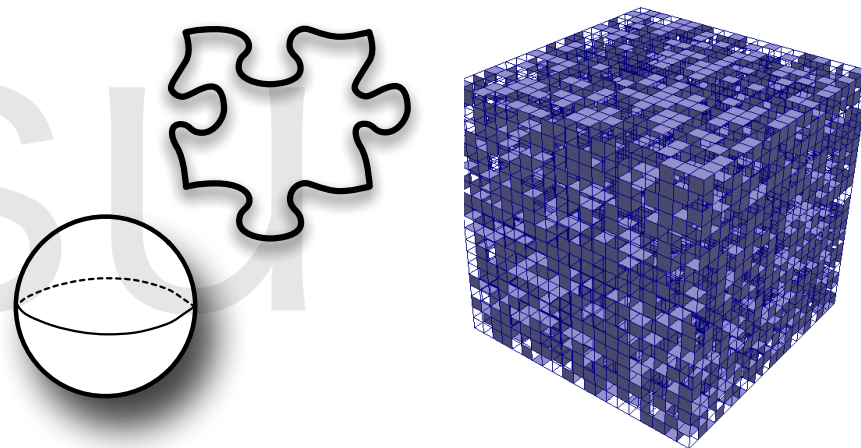
疑問：何でもありって感じの世界だけど不変な量ってあるの？

	連結成分	リング	空洞
	1	0	0
	2	0	0
	1	1	0
	1	0	1

連結成分, リング, 空洞の数は**不変量**

★ 連結成分, リング, 空洞 : 大域的なつながり方を表す量

- 0次元の「穴」 : 連結成分
- 1次元の「穴」 : リング
- 2次元の「穴」 : 空洞



トポロジー : 連続変形で重なり合う図形の  
穴の数は不変

★ 不満 : あまりにも「穴」の定義があいまい、、、  
正確な定義 = 代数化

→ ホモロジー

# ★ ホモロジーとは

幾何学的対象  $X$  に対して

$$H_0(X) = \mathbf{R}^{k_0}$$

$k_0$  : 連結成分の数

$$H_1(X) = \mathbf{R}^{k_1}$$

$k_1$  : リングの数

$$H_2(X) = \mathbf{R}^{k_2}$$

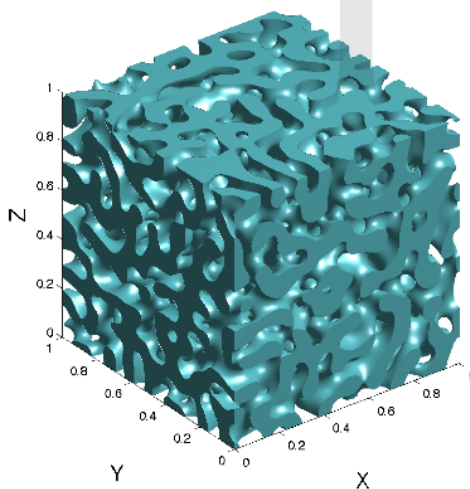
$k_2$  : 空洞の数

⋮

$$H_q(X) = \mathbf{R}^{k_q}$$

$k_q$  :  $q$ 次元の「穴」の数

を表す, 代数的な道具 (現代数学に必須!)



$$H_0(\text{cube}) = \mathbf{R}$$

$$H_1(\text{cube}) = \mathbf{R}^{847}$$

$$H_2(\text{cube}) = 0$$



$$H_0(\text{西浦先生}) = \mathbf{R}^{10}$$

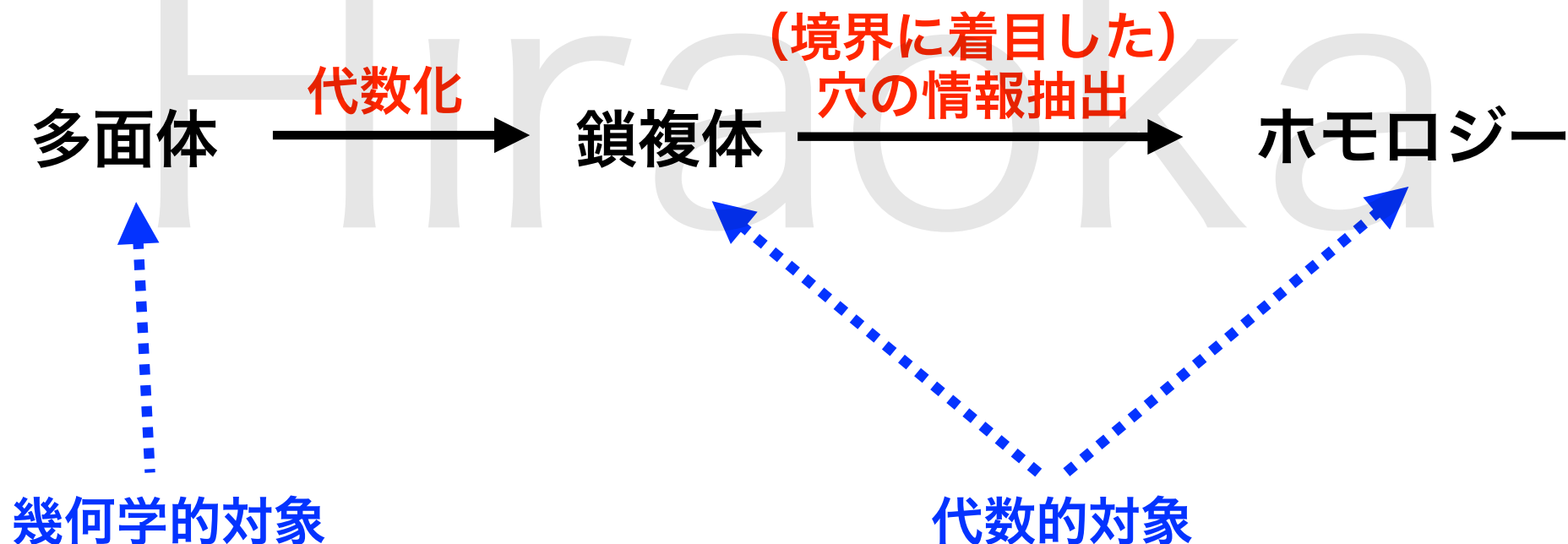
$$H_1(\text{西浦先生}) = \mathbf{R}^{36}$$



## 1.2. 穴を定義する

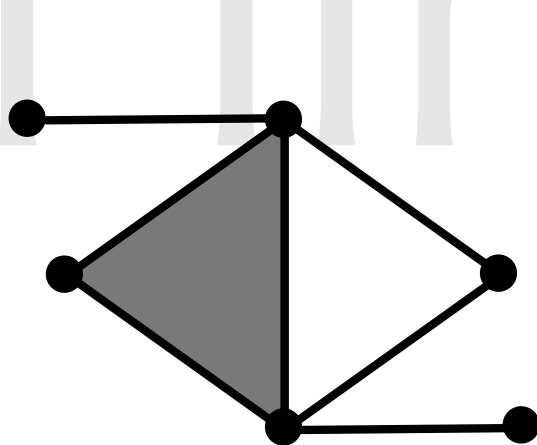
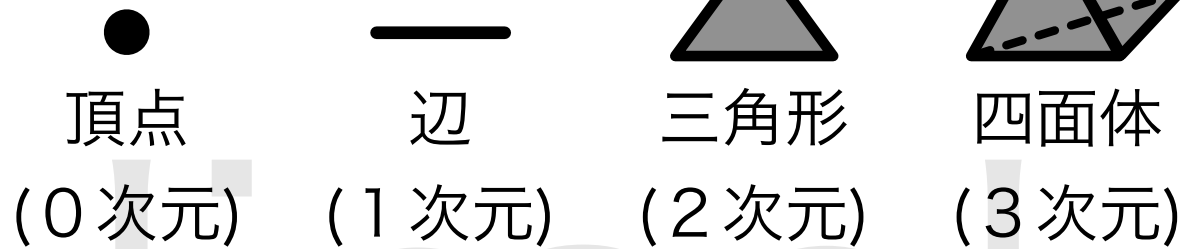
「連結成分, リング, 空洞」を表すホモロジーを  
代数的に定義してみる

### ★ ホモロジーの導入の流れ

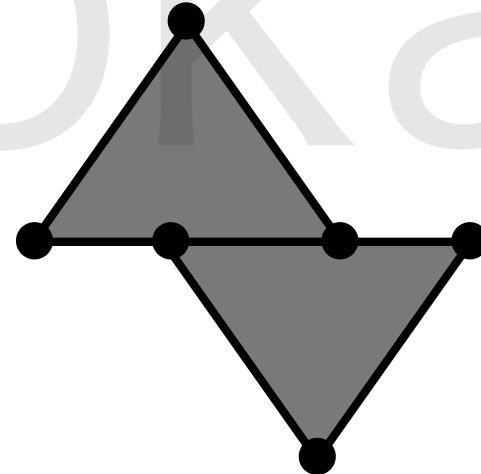


# 多面体 (単体複体)

多面体：頂点，辺，三角形，…，高次元の三角形を  
(ズレが無いように) 張り合わせた図形



多面体の例



多面体ではない例

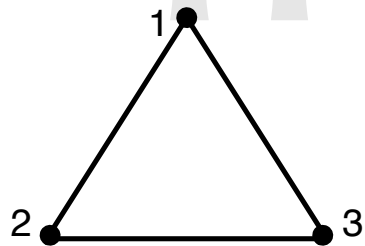
## 単体複体 (多面体)

有限集合  $V$  と  $V$  の部分集合の集まり  $\Sigma$  が

- $v \in V \implies \{v\} \in \Sigma$
- $\tau \in \Sigma, \sigma \subset \tau \implies \sigma \in \Sigma$

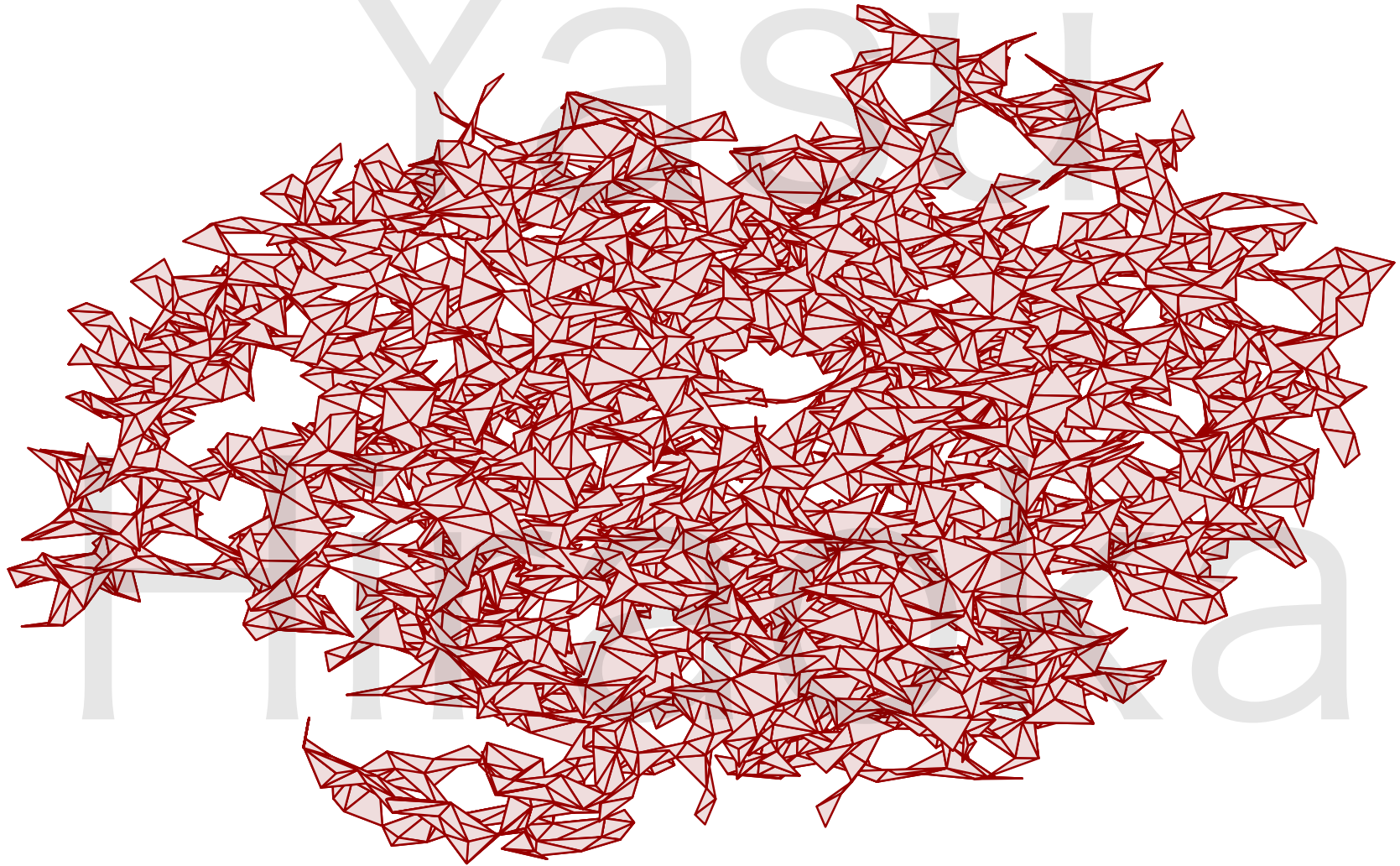
をみたすとき  $X = (V, \Sigma)$  を (抽象) 単体複体とよぶ

単体複体の例)



$$X = \{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{1\}, \{2\}, \{3\}\}$$

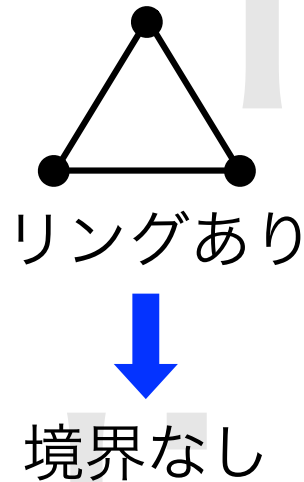
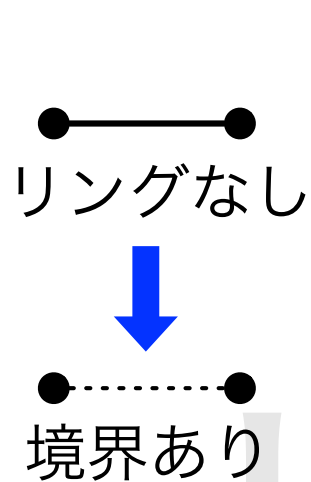
## 多面体の例



ヘモグロビンの多面体モデル

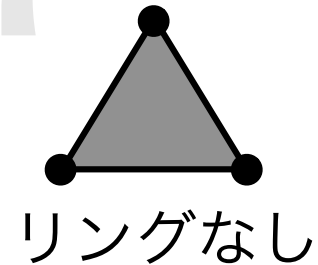
# 穴って何だ？



## ★ 1次元の穴：リング



1次元図形は境界がなければ「リング」？

そうでもない



1次元図形  には境界がないが  
2次元図形  の境界になっているので  
リングが消失！

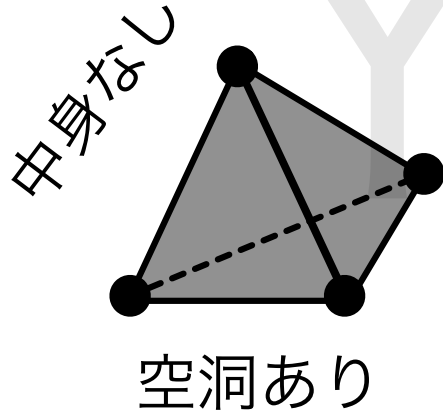
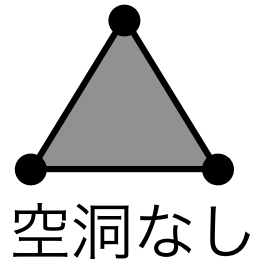
## 1次元の穴（リング）の特徴付け



境界のない1次元図形で2次元図形の境界になっているもの

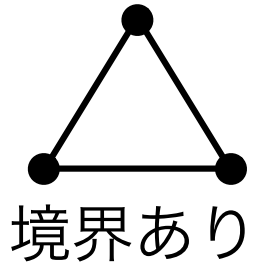
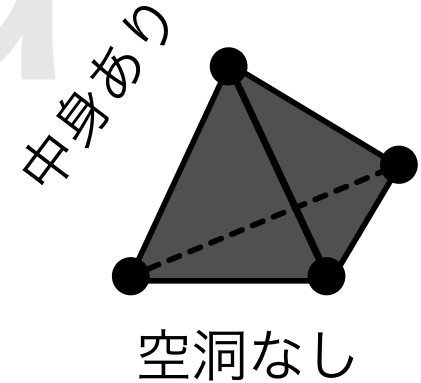
# 穴って何だ？

## ★ 2次元の穴：空洞



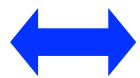
2次元図形は境界がなければ「空洞」？

そうでもない



2次元図形には**境界**がないが  
3次元図形の**境界**になっているので  
空洞が消失！

## 2次元の穴（空洞）の特徴付け



**境界**のない2次元図形で3次元図形の**境界**になっていないもの

# 穴と境界

## 1次元の穴（リング）の特徴付け



**境界**のない1次元図形で2次元図形の**境界**になっていないもの

## 2次元の穴（空洞）の特徴付け



**境界**のない2次元図形で3次元図形の**境界**になっていないもの

**KEYWORD : 境界**

ホモロジーのこころ

境界に着目した定式化を行うことで穴を代数化する

# 境界を代数的に取り出そう！

$K$ : 多面体,  $Z_2 := \{0, 1\}$  ( $0 + 0 = 1 + 1 = 0$ ,  $0 + 1 = 1 + 0 = 1$ )

---

定義：鎖群 (さぐん)

$$C_0(K) = \left\{ \sum \alpha_i \underline{|v_i|} ; \alpha_i \in \mathbf{Z}_2 \right\}$$

**Kの頂点**

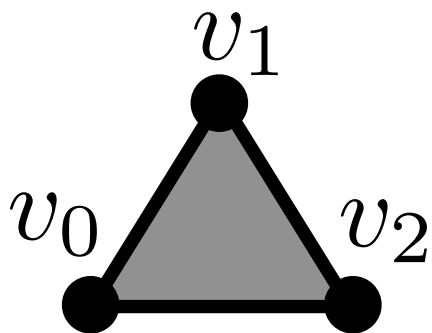
$$C_1(K) = \left\{ \sum \alpha_{ij} \underline{|v_i v_j|} ; \alpha_{ij} \in \mathbf{Z}_2 \right\}$$

**Kの辺**

$$C_2(K) = \left\{ \sum \alpha_{ijk} \underline{|v_i v_j v_k|} ; \alpha_{ijk} \in \mathbf{Z}_2 \right\}$$

**Kの三角形**

---



$$C_0(K) = \{a|v_0| + b|v_1| + c|v_2|\}$$

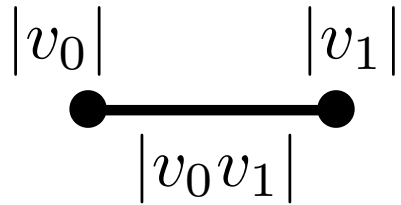
$$C_1(K) = \{a|v_0 v_1| + b|v_1 v_2| + c|v_0 v_2|\}$$

$$C_2(K) = \{a|v_0 v_1 v_2|\}$$



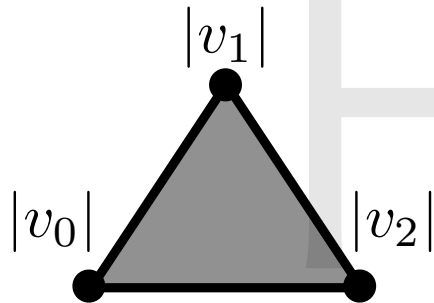
# 境界作用素

辺の境界：  $|v_0v_1|$  の境界は  $|v_0|, |v_1|$



$\partial_1$  を境界を取り出す操作（線形写像）として  
 $\partial_1|v_0v_1| := |v_0| + |v_1|$  と書く

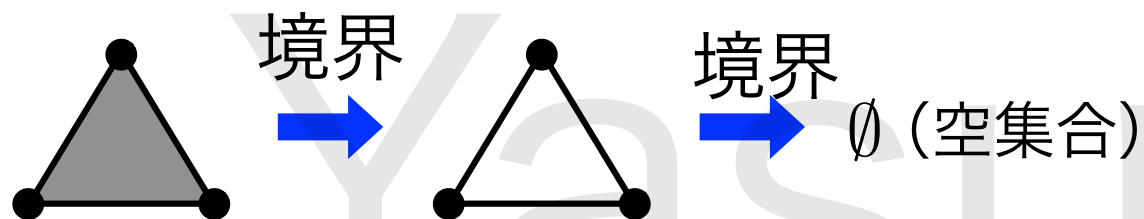
三角形の境界：  $|v_0v_1v_2|$  の境界は  $|v_0v_1|, |v_1v_2|, |v_0v_2|$



同様に境界を取り出す操作（線形写像）を  
 $\partial_2|v_0v_1v_2| := |v_0v_1| + |v_1v_2| + |v_0v_2|$   
と書く

- $\partial_1, \partial_2$  を境界作用素と呼ぶ
- $0 \longrightarrow C_2(K) \xrightarrow{\partial_2} C_1(K) \xrightarrow{\partial_1} C_0(K) \longrightarrow 0$  を鎖複体と呼ぶ

## 境界の特徴の1つ：境界の境界は無い（空集合）



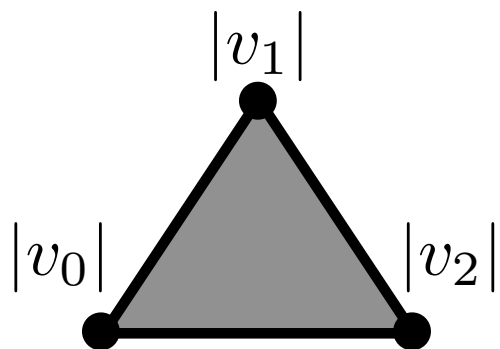
境界作用素は同様の性質を保っている？

命題： $\partial_1 \circ \partial_2 = 0$  (続けて境界をとる作用をすれば零になる)

$$C_2(K) \xrightarrow{\partial_2} C_1(K) \xrightarrow{\partial_1} C_0(K)$$

注意) 代数的に「境界の境界は無い」を表現している

ざっと確認： $\partial_1(\partial_2|v_0v_1v_2|) = \partial_1(|v_0v_1| + |v_1v_2| + |v_0v_2|)$



$$\begin{aligned} &= \partial_1|v_0v_1| + \partial_1|v_1v_2| + \partial_1|v_0v_2| \\ &= |v_0| + |v_1| + |v_1| + |v_2| + |v_0| + |v_2| \\ &= 0 (\because 1 + 1 = 0) \blacksquare \end{aligned}$$

# 穴を代数的にとりだしてみよう

思い出そう！

1次元の「穴」 (リング)



境界のない1次元図形で2次元図形の境界に  
なっていないもの ① ②

①を代数的に表すと

$$Z_1(K) := \{\partial_1 c = 0 \text{ となる } c \in C_1(K)\} = \text{Ker} \partial_1$$

②を代数的に表すと

$$B_1(K) := \{c = \partial_2 c' \text{ と書ける } c \in C_1(K)\} = \text{Im} \partial_2$$

ここで  $c' \in C_2(K)$

ちなみに命題  $\partial_1 \circ \partial_2 = 0$  より

$$B_1(K) \subset Z_1(K) \subset C_1(K)$$

## 1次元の「穴」 (リング)



境界のない1次元図形で2次元図形の境界になっ  
ていないもの ① ② ③

よって1次元の「穴」を代数的に表現すると

「 $Z_1(K)$  に対して  $B_1(K)$  を零にする操作で生き残るもの」  
① ② (操作を  $Q$  と表す) ③

すると  $z = z' + b, z, z' \in Z_1(K), b \in B_1(K)$  に対して

$$Q(z') = Q(z) + Q(b) = Q(z)$$

$Z_1(K)$  の中で違いが  $B_1(K)$  である要素ごとにグループ分けした各グループが1次元の「穴」を表現することになる

## 定義：ホモロジー

$$\begin{aligned} H_1(K) &:= Q(Z_1(K)) = Z_1(K)/B_1(K) \\ &= \{[z] \mid z \in Z_1(K)\} \end{aligned}$$

ここで  $[z]$  は  $z$  が所属するグループを表す  
を **1次元ホモロジー** とよぶ

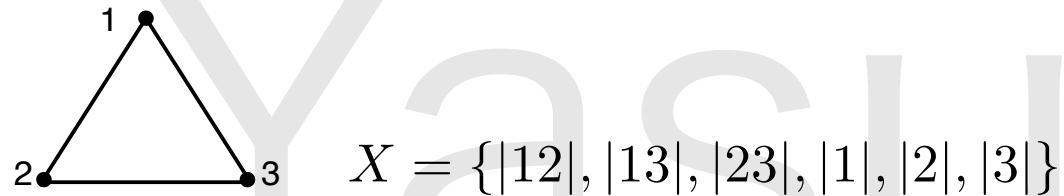
宿題) 具体的な図形に対して自分でホモロジー群を  
求めてみましょう

宿題) 2次以上の定義  $H_k(X)$  ( $k \geq 2$ ) も考えてみましょう  
(同様です)

宿題) 穴の柔らかさは実は  $B_q(K)$  が表します。なぜだか  
考えてみましょう

# ホモロジーの計算例 (1)

単体複体



鎖複体

$$0 \longrightarrow C_2(X) \xrightarrow{\partial_2} C_1(X) \xrightarrow{\partial_1} C_0(X) \longrightarrow 0$$

$$C_2(X) = \{0\} \quad C_1(X) = \{\alpha_1|12| + \alpha_2|13| + \alpha_3|23| : \alpha_i \in \mathbf{Z}_2\}$$

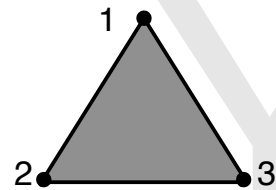
$$C_0(X) = \{\alpha_1|1| + \alpha_2|2| + \alpha_3|3| : \alpha_i \in \mathbf{Z}_2\}$$

$$\partial_2 = 0 \quad \partial_1 = \begin{array}{c} \begin{array}{ccc|c} |12| & |13| & |23| & \\ \hline 1 & 1 & 0 & |1| \\ 1 & 0 & 1 & |2| \\ 0 & 1 & 1 & |3| \end{array} \end{array} \quad \begin{array}{l} Z_1(X) = \text{Ker } \partial_1 = \text{Span} \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\} \simeq \mathbf{Z}_2 \\ B_1(X) = \text{Im } \partial_2 = 0 \end{array}$$

1次ホモロジー  $H_1(X) = Z_1(X)/B_1(X) \simeq \mathbf{Z}_2$       リング  $|12| + |13| + |23|$  に対応

# ホモロジーの計算例 (2)

単体複体



$$X = \{|123|, |12|, |13|, |23|, |1|, |2|, |3|\}$$

鎖複体

$$0 \longrightarrow C_2(X) \xrightarrow{\partial_2} C_1(X) \xrightarrow{\partial_1} C_0(X) \longrightarrow 0$$

$$C_2(X) = \{\alpha|123| : \alpha \in \mathbf{Z}_2\}$$

$$C_1(X) = \{\alpha_1|12| + \alpha_2|13| + \alpha_3|23| : \alpha_i \in \mathbf{Z}_2\}$$

$$C_0(X) = \{\alpha_1|1| + \alpha_2|2| + \alpha_3|3| : \alpha_i \in \mathbf{Z}_2\}$$

$$\partial_2 = \begin{array}{c} |123| \\ \left[ \begin{array}{c} 1 \\ 1 \\ 1 \end{array} \right] \begin{array}{l} |12| \\ |13| \\ |23| \end{array} \end{array} \quad \partial_1 = \begin{array}{c} |12| \quad |13| \quad |23| \\ \left[ \begin{array}{ccc} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{array} \right] \begin{array}{l} |1| \\ |2| \\ |3| \end{array} \end{array} \quad Z_1(X) = \text{Ker} \partial_1 = \text{Span} \left\{ \left[ \begin{array}{c} 1 \\ 1 \\ 1 \end{array} \right] \right\} \simeq \mathbf{Z}_2$$

$$B_1(X) = \text{Im} \partial_2 = Z_1(X)$$

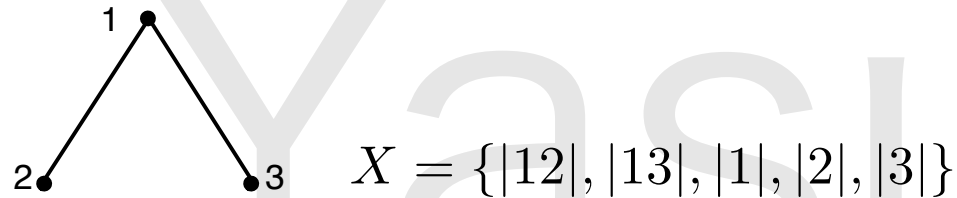
1次ホモロジー

$$H_1(X) = Z_1(X)/B_1(X) = 0$$

リングなしに対応

# ホモロジーの計算例 (3)

単体複体



鎖複体

$$0 \longrightarrow C_2(X) \xrightarrow{\partial_2} C_1(X) \xrightarrow{\partial_1} C_0(X) \longrightarrow 0$$

$$C_2(X) = \{0\} \quad C_1(X) = \{\alpha_1|12| + \alpha_2|13| : \alpha_i \in \mathbf{Z}_2\}$$

$$C_0(X) = \{\alpha_1|1| + \alpha_2|2| + \alpha_3|3| : \alpha_i \in \mathbf{Z}_2\}$$

$$\partial_2 = 0 \quad \partial_1 = \begin{bmatrix} |12| & |13| \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{matrix} |1| \\ |2| \\ |3| \end{matrix} \quad \begin{matrix} Z_1(X) = \text{Ker} \partial_1 = 0 \\ B_1(X) = \text{Im} \partial_2 = 0 \end{matrix}$$

1次ホモロジー  $H_1(X) = Z_1(X)/B_1(X) = 0$       リングなしに対応



# ホモロジー線形写像

包含関係  $X \subset Y$   $\rightarrow$  線形写像  $C_k(X) \hookrightarrow C_k(Y)$

$$Z_k(X) \hookrightarrow Z_k(Y)$$

$$B_k(X) \hookrightarrow B_k(Y)$$

$$c \mapsto c \quad \text{単射}$$

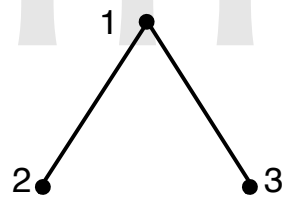
## $\rightarrow$ ホモロジー線形写像

$$H_k(X) = \frac{Z_k(X)}{B_k(X)} \longrightarrow H_k(Y) = \frac{Z_k(Y)}{B_k(Y)}$$

$$[c] \mapsto [c]$$

注) 一般に単射とは限らない  
( [ ] は各  $X, Y$  で! )

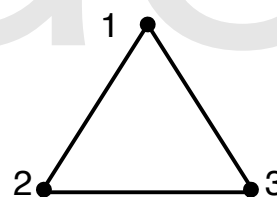
例)



$H_1$

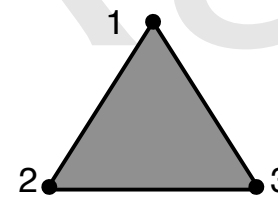
0

$\longrightarrow$



$Z_2$

$\longrightarrow$

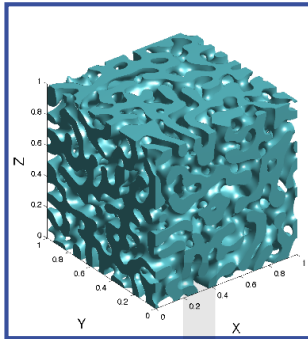


0

ホモロジー線形写像はリングの発生と消滅を表現  $\rightarrow$  パーシステントホモロジー

## 2. パーシステントホモロジー入門

### ホモロジー



ホモロジー

$$\begin{aligned} H_0(\text{cube}) &= \mathbb{R} && : 1\text{個の連結成分} \\ H_1(\text{cube}) &= \mathbb{R}^{847} && : 847\text{個のわっか} \\ H_2(\text{cube}) &= 0 && : 0\text{個の空洞} \end{aligned}$$

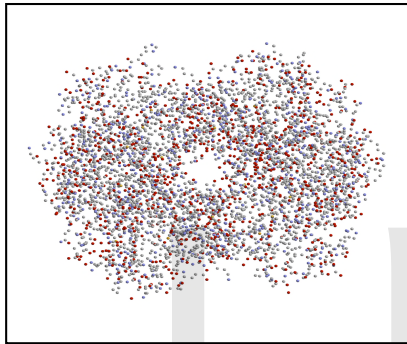
- ポアンカレによって約100年前に考案
- その後、数学者によって数学研究にのみ使われる
- 計算機の発達に伴い計算機を用いてホモロジーを計算する方法が開発される (計算ホモロジー)
- 生命・材料・情報などへ応用されはじめる (TDAの源流)

# 2. パーシステントホモロジー入門

## パーシステントホモロジー

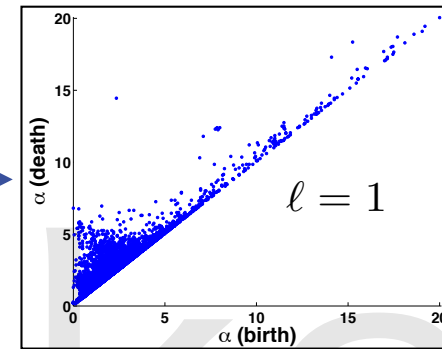


ヘモグロビン



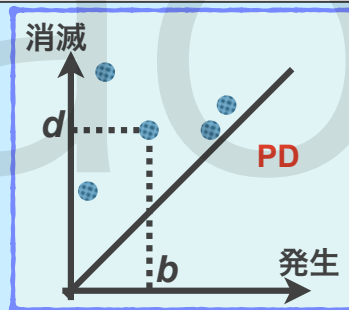
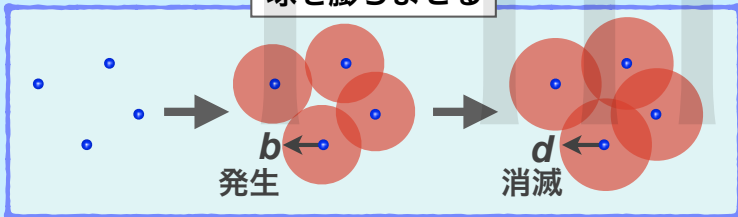
パーシステントホモロジー

パーシステント図(PD)



### PD構成法

球を膨らませる



- PDの各点はデータ内の穴
- 発生軸は穴の発生パラメータ
- 消滅軸は穴の消滅パラメータ
- 対角線付近の点はノイズ
- 対角線から離れた点はロバスト

- 数学者Edelsbrunner, Carlsson等によって開発 (2003, 2005年)
- 穴のサイズ, 形状, 階層構造などを扱える
- 数学的進化と同時に, 現在急速に諸科学への応用が進められている

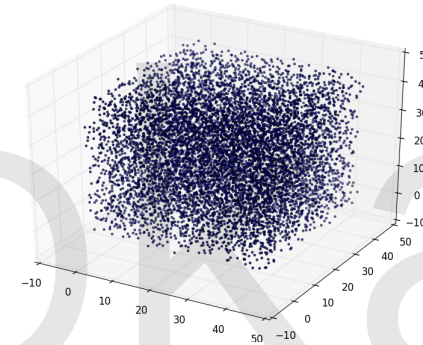
# ホモロジーとパーシステントホモロジーの入力の違い

- ホモロジー：（単一の）図形
- パーシステントホモロジー：図形のフィルトレーション  
→ トポロジーにスケール情報を付与できる

## 代表的な入力の構成法

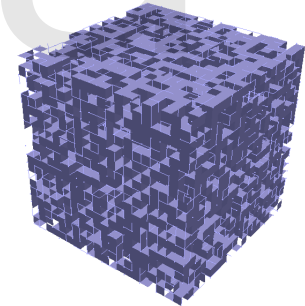
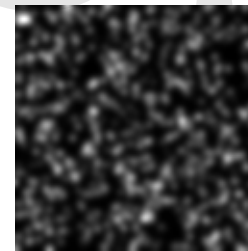
### 空間点列

- チェック複体・アルファ複体
- リップス複体



### (高次元) 画像データ

- 方体複体

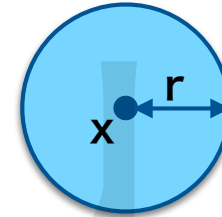


まずはこれらの幾何モデルの解説を行う

# チェック複体

• 入力点列  $X = \{x_i \in \mathbf{R}^m \mid i = 1, \dots, n\}$

• 球体  $B_r(x) = \{y \in \mathbf{R}^m \mid \|y - x\| \leq r\}$

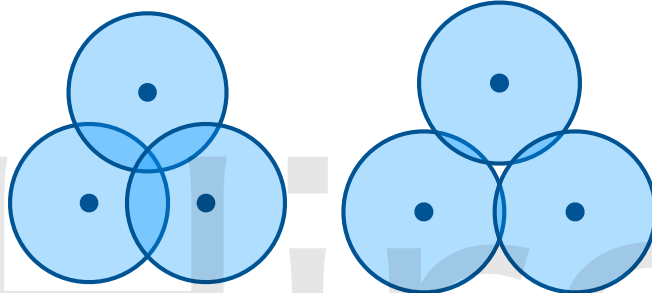


• チェック複体  $\mathcal{C}(X, r)$

$\{|x_{i_0} \cdots x_{i_k} \mid \bigcap_{j=0}^k B_r(x_{i_j}) \neq \emptyset\}$  で定まる単体複体

例)

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



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



観察)  $\bigcup_{x \in X} B_r(x)$  に穴がある



$\mathcal{C}(X, r)$  に穴がある

脈体定理

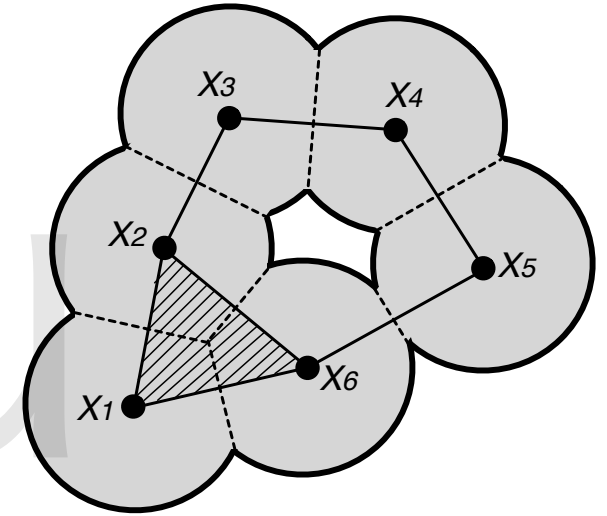
(Nerve Theorem)

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

- 切り張りしない連続変形でうつりあえる (ホモトピー同値)
- 穴の情報は保持
- 左辺：原子配置、センサー配置、データ点列  
右辺：計算機で扱いやすい  
左辺と右辺でトポロジカルな情報の損失がない

# アルファ複体

- 入力点列  $X = \{x_i \in \mathbf{R}^m \mid i = 1, \dots, n\}$
- $\mathbf{R}^m = \cup_i V_i$  : ボロノイ分割
- $\cup_i B_i(r) = \cup_i (B_i(r) \cap V_i)$
- アルファ複体  $\mathcal{A}(X, r)$



$\{B_i(r) \cap V_i \mid i = 1, \dots, n\}$  の双対単体複体

(チェック複体と同様に定義)

- 脈体定理は成立  $\cup_{x \in X} B_r(x) \simeq \mathcal{A}(X, r)$
- 単体の個数を削減できる (計算量の削減)

➡ 実用的にはチェック複体ではなくアルファ複体の方がよく用いられる。ただし  $m=3$  まで

アルファ複体構成ソフトウェア : CGAL <http://www.cgal.org>

# リップス複体

## チェック複体とアルファ複体の問題点

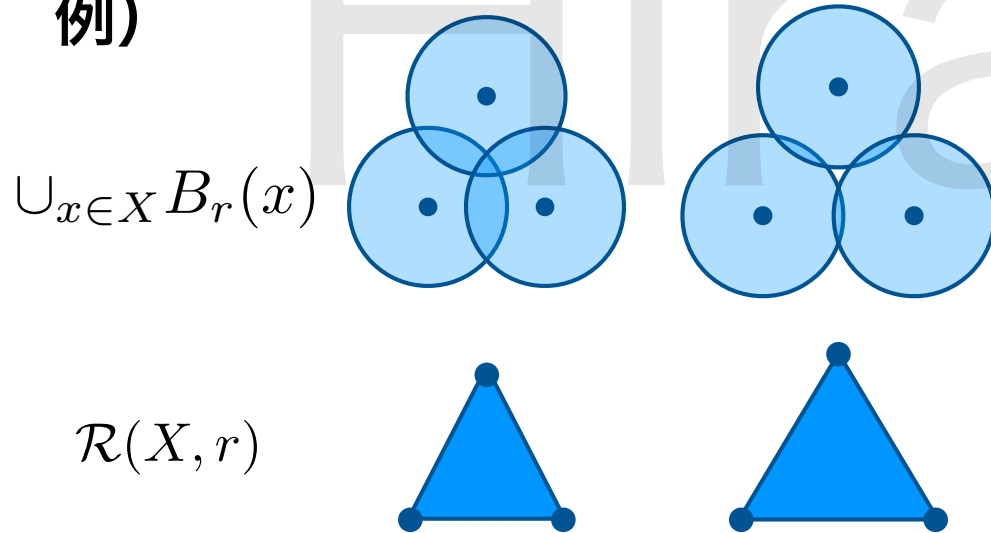
…… 高次元空間データの計算が困難

- 入力点列  $X = \{x_i \in \mathbf{R}^m \mid i = 1, \dots, n\}$
- リップス複体  $\mathcal{R}(X, r)$

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

で定まる単体複体 (各ペアで交わりを調べれば良い)

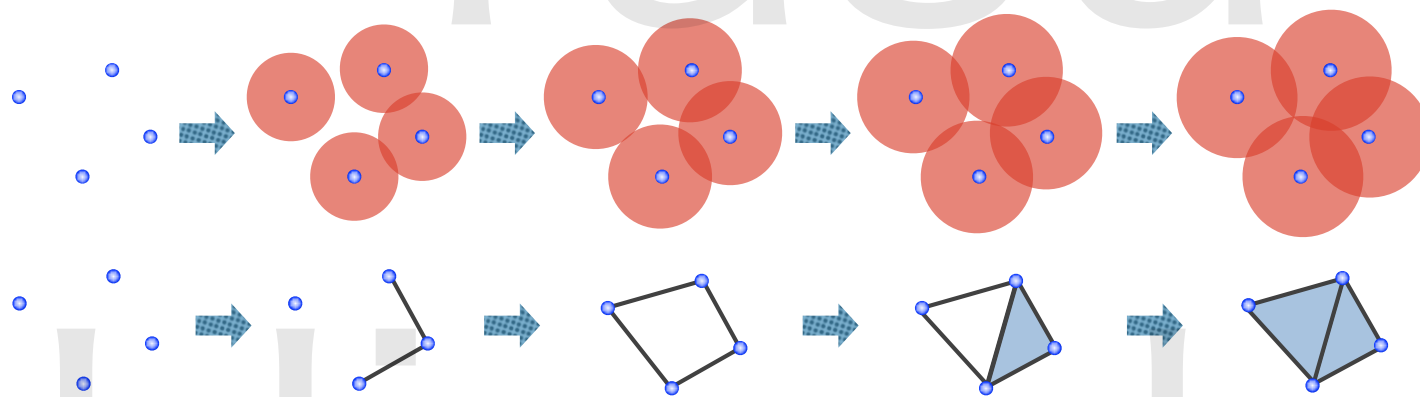
例)



- 穴の情報は保たれない
- 高次元空間のデータでも計算可能 (2点間の距離がわかれば構成できる)

# フィルトレーションの構成法

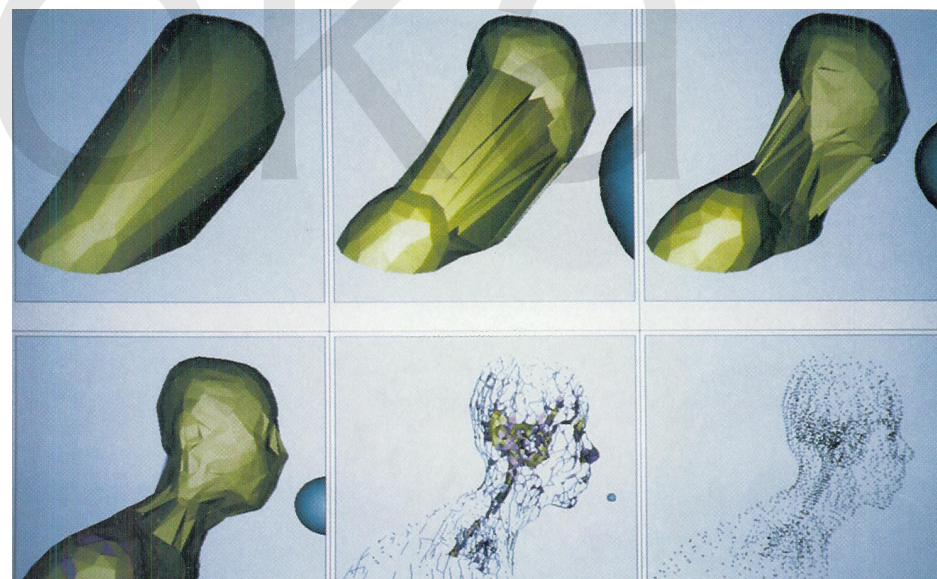
- $K(X, r) = \mathcal{C}(X, r), \mathcal{R}(X, r), \text{ or } \mathcal{A}(X, r)$  とする
- $K(X, r) \subset K(X, s)$  for  $r \leq s$  となっている  
(各パラメータで交わりを調べているので)



- $\{K(X, r)\}_{r \geq 0}$

フィルトレーションという  
(図形の増大列のこと)

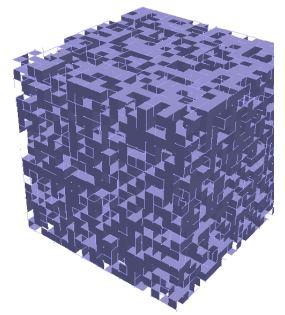
- パラメータ  $r$  はデータの  
解像度を調整





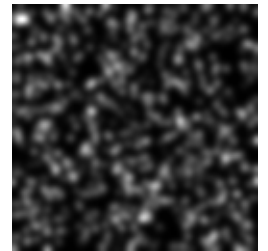
# 方体複体

(ピクセルやボクセルの集まりからなる図形)



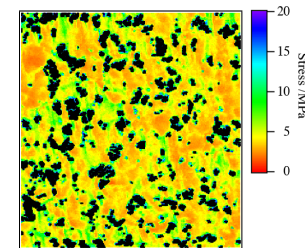
## 方体複体のフィルトレーションの作り方

- 2値化されている場合：黒領域を徐々に太らせていく
- 2値化されていない場合（例：グレースケール画像）



方法1：適当に2値化して太らせる

方法2：レベルセット法



- 関数  $f : X \rightarrow \mathbb{R}$

(画像  $X$  上のグレースケール値とみなす)

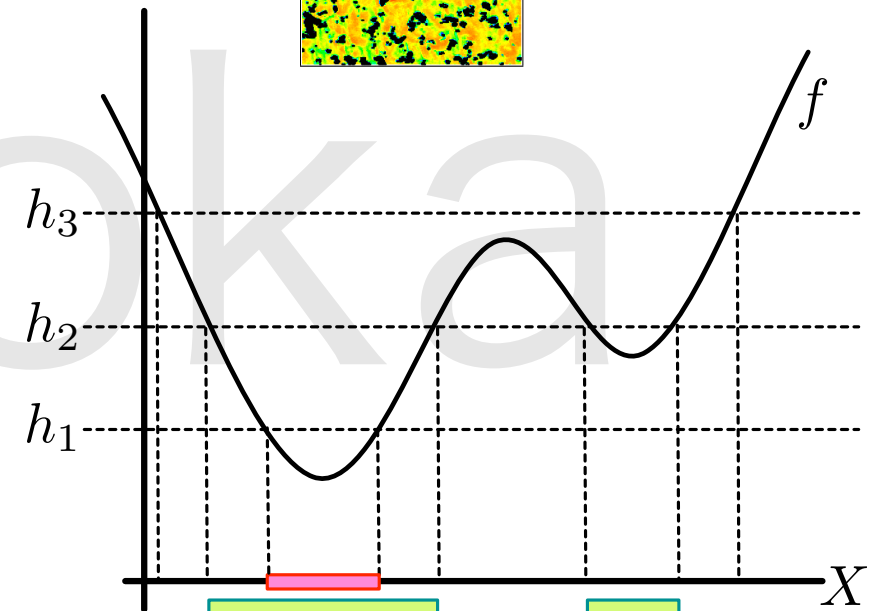
- レベルセット

$$X_h := \{x \in X \mid f(x) \leq h\}$$

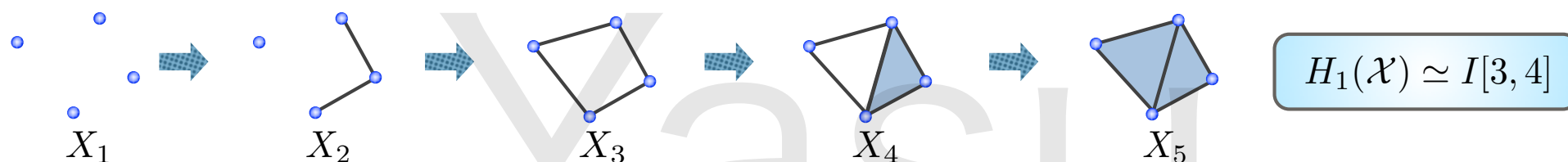
- フィルトレーション

$$X_{h_1} \subset X_{h_2} \subset \cdots \subset X_{h_T}$$

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

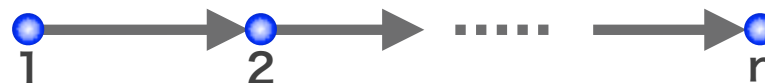


# パーシステントホモロジー



- フィルトレーション  $\mathcal{X} : X_1 \subset X_2 \subset \dots \subset X_n$  ホモロジー線形写像でつなげる
- パーシステントホモロジー  $H_\ell(\mathcal{X}) : H_\ell(X_1) \rightarrow H_\ell(X_2) \rightarrow \dots \rightarrow H_\ell(X_n)$

representations on  $A_n$



- パーシステントホモロジーの一意分解定理

$$H_\ell(\mathcal{X}) \simeq \bigoplus_{i=1}^s I[b_i, d_i] \quad I[b, d] : 0 \rightarrow \dots \rightarrow 0 \rightarrow K \rightarrow \dots \rightarrow K \rightarrow 0 \rightarrow \dots \rightarrow 0$$

at  $X_b$ 
at  $X_d$

b: 発生時刻 (birth) , d: 消滅時刻 (death)

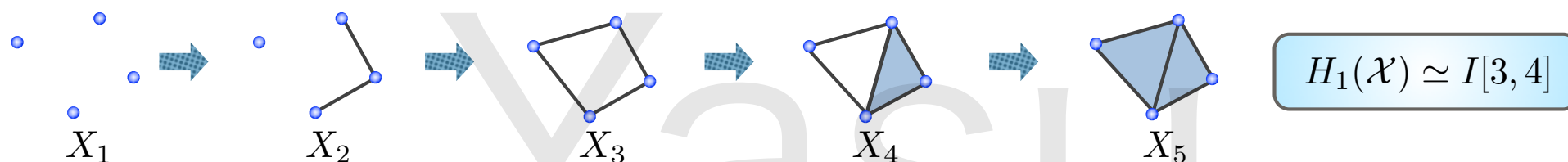
(ここで  $K = \mathbb{Z}_2$ )

- 区間  $I[b, d]$  を生成元ともよぶ。穴の発生, 消滅, 存続を表現
- $d-b$  を生存時間とよぶ
- 生存時間が短い穴は入力データの摂動で消滅しうる (ノイズ)
- 生存時間が長い穴は入力データの摂動に強い (ロバスト)



定量的に穴の違いを識別可能 (古典的なトポロジーから進化)

## パーシステント図



### パーシステントホモロジーの一意分解定理

(ここで  $K = \mathbb{Z}_2$ )

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

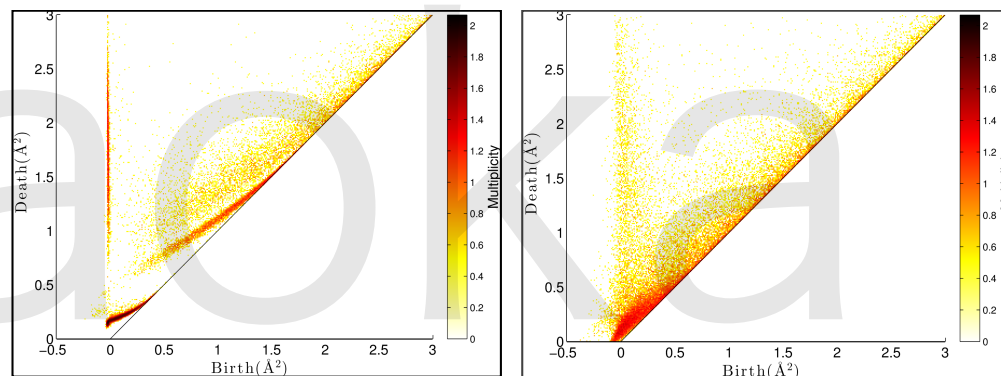
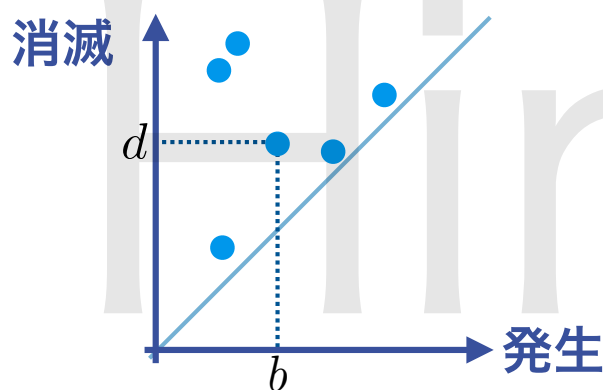
$$I[b, d] : 0 \rightarrow \dots \rightarrow 0 \rightarrow K \rightarrow \dots \rightarrow K \rightarrow 0 \rightarrow \dots \rightarrow 0$$

at  $X_b$                       at  $X_d$

b: 発生時刻 (birth) , d: 消滅時刻 (death)

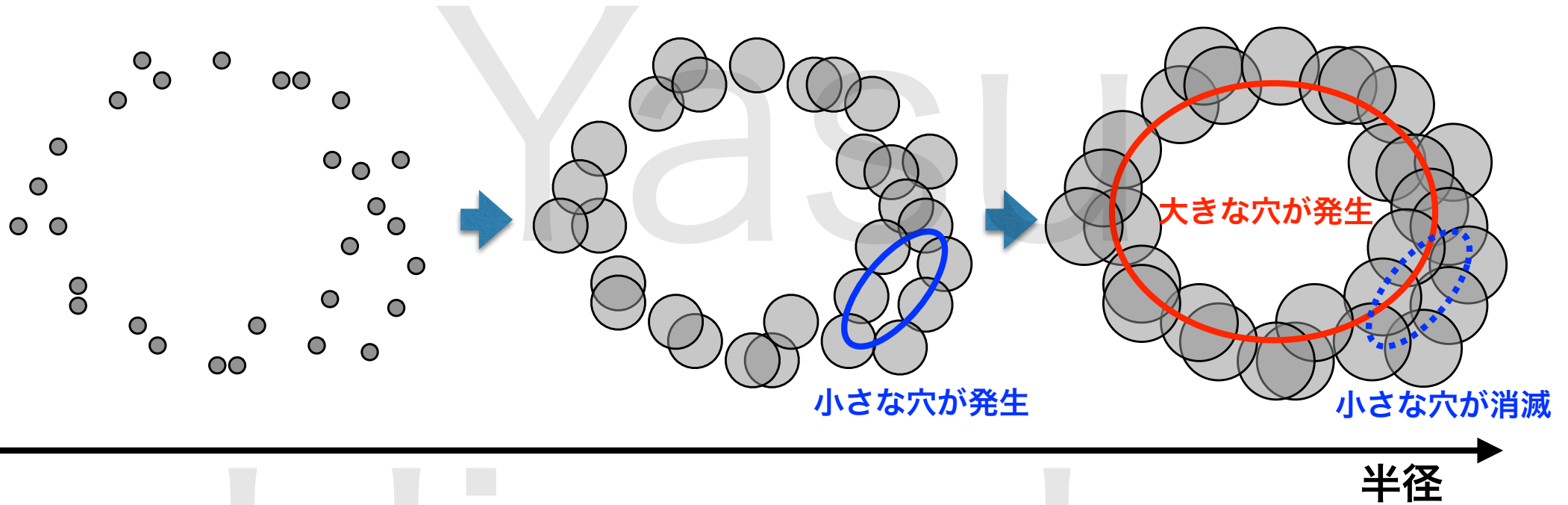
### パーシステント図 (PD)

$$D_k(\mathcal{X}) = \{(b_i, d_i) \in \mathbb{R}_{\geq 0}^2 : i = 1, \dots, p\}$$

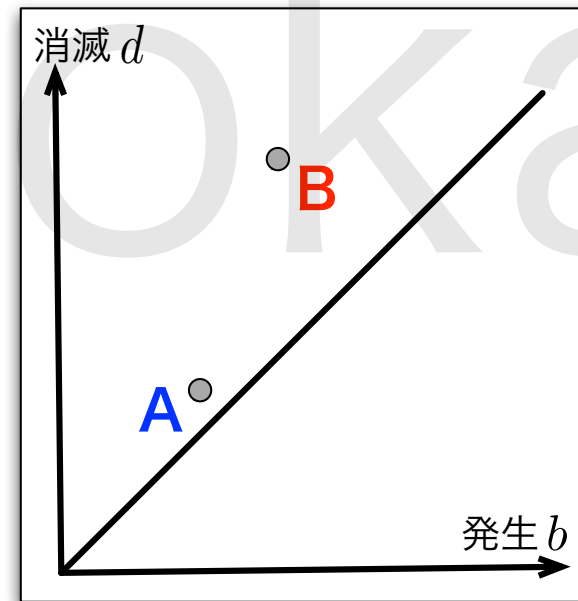


- PD内の点は一意分解定理に現れる区間に対応
- 対角線付近の点は生存時間が短く、対角線から離れた点は生存時間が長い
- PDを比較することでデータの穴情報をマルチスケールで比較することが可能 (一意分解定理があるので比較は意味を持つ)

## 入力データ

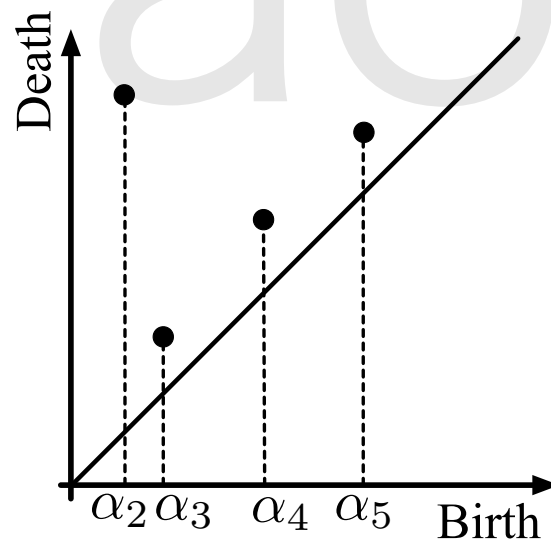
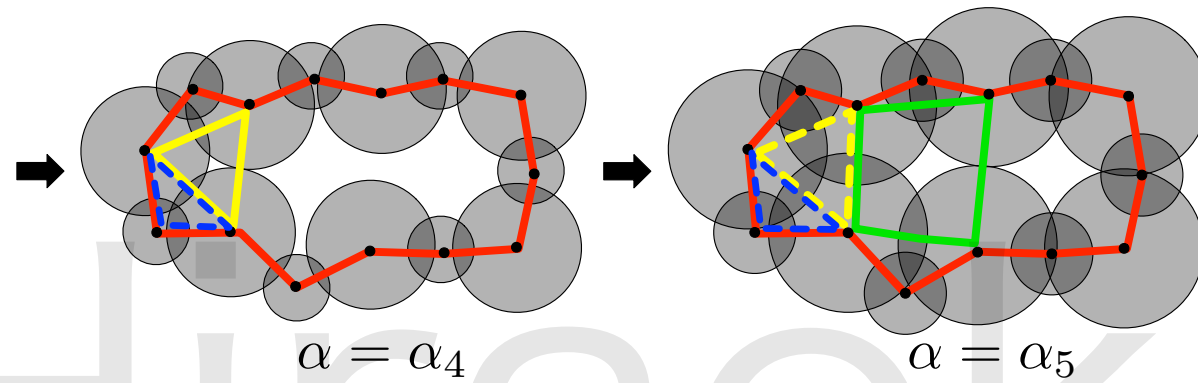
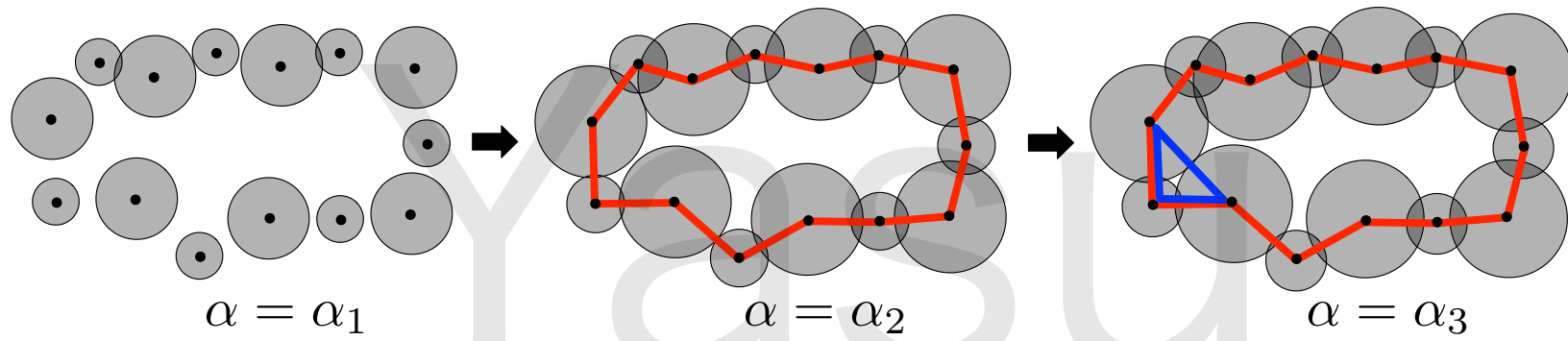


- 点Aが青の小さな穴に対応
- 点Bが赤の大きな穴に対応
- PDから穴の発生・消滅時刻、大きさなどがわかる



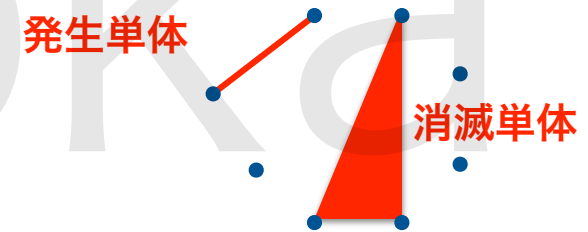
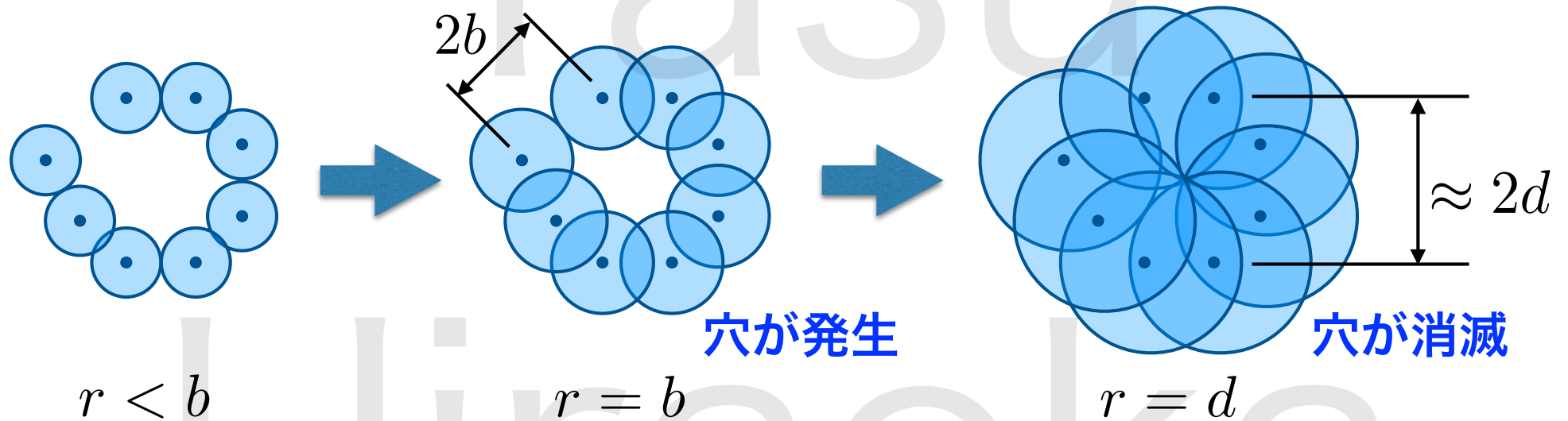
PD1

# PDの例



# PDからの幾何情報の復元 (空間点データの場合のみ説明)

発生時刻  $b$  : 穴を構成する点たちの隣接距離の半分  
消滅時刻  $d$  : (だいたい) 穴の半径



上の例での発生・消滅単体

発生単体 : 発生時刻を決める単体  
消滅単体 : 消滅時刻を決める単体

- 発生単体と消滅単体は計算の過程で求めている
- 描かせることで穴の位置や形状がわかる (逆問題の一種)

# パーシステントホモロジー周辺の最新の話題 (数学編)

**確率論**：ランダムグラフの高次元化にあたるランダムトポロジー研究が急速に発展中。ベッチ数やパーシステント図の極限定理が近年示されつつある。

**表現論**：クイバー (or 多元環) の表現としての一般化。Auslander-Reiten理論やグレブナー基底を用いたパーシステントホモロジーの高次元化が進められている。

**統計**：パーシステント図をデータとして統計解析を行う位相的統計理論の整備が進められる。パーシステント図に対する各種ベクトル化を用いた統計解析法が開発されている (詳しくは福水さんの講演)

**逆問題**：パーシステント図で検出された特徴を元のデータ空間に引き戻す逆問題は現実の問題において極めて重要。スパース最適化や追跡法を用いた手法が開発されている (詳しくは大林さんの講演)

これらの課題は「CREST TDA」で重点的に研究開発に取りかかる

平岡裕章 (代表：東北大)、浅芝秀人 (静岡大)、白井朋之 (九大)、福水健次 (統数研)、一宮尚志 (岐阜大)

# パーシステントホモロジー周辺の最新の話 (応用編)

材料科学：ソフトマターやガラスのMD計算や実験画像へのPD構造解析  
後で説明 (AIMR, CREST TDA, SIP, MI<sup>2</sup>I)

脳科学：ニューロンの相関関係に対するリップス複体を用いたPD解析  
Giusti et al. PNAS (2015)

生命科学：タンパク質構造のPD解析 Gameiro et al. (2015), Cang et al (2015)

情報通信：センサーネットワークに対するリップス複体を用いたPD解析  
Robert Ghrist (UPenn)

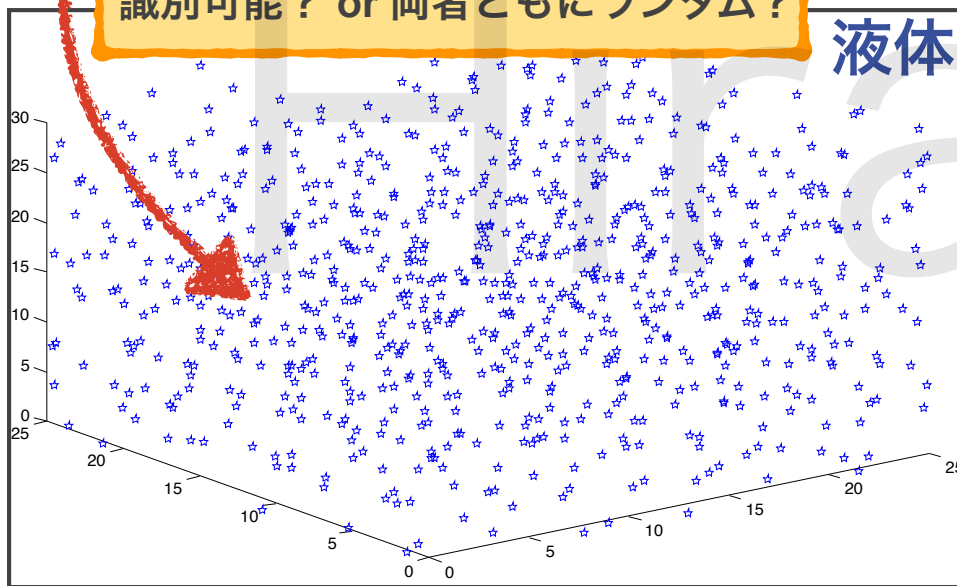
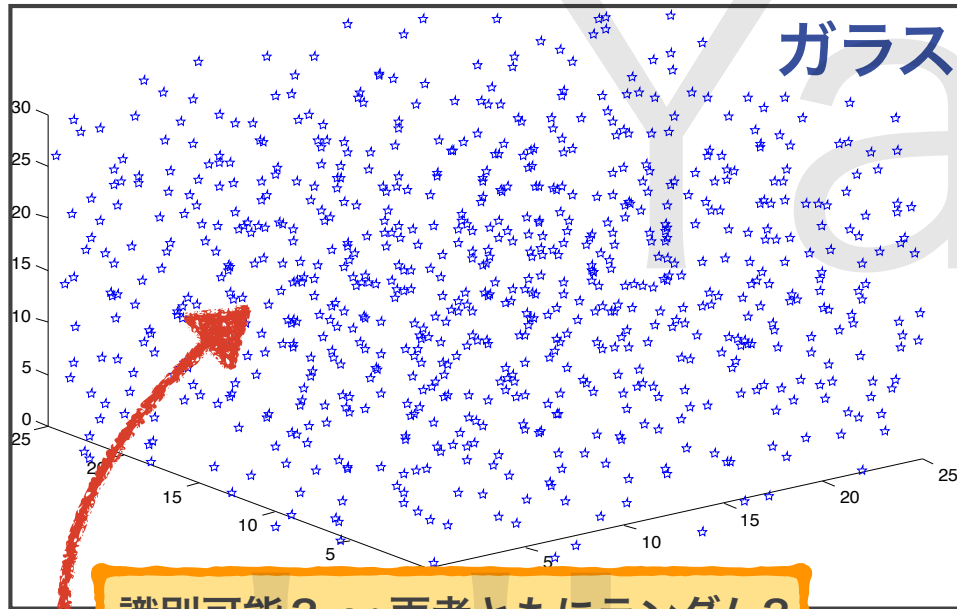
ビッグデータ解析：ソーシャルネットワーク, 医療, 創薬, 金融 etc  
AYASDI, Inc.

宇宙論：何かのパーシステントホモロジーを計算してみたい

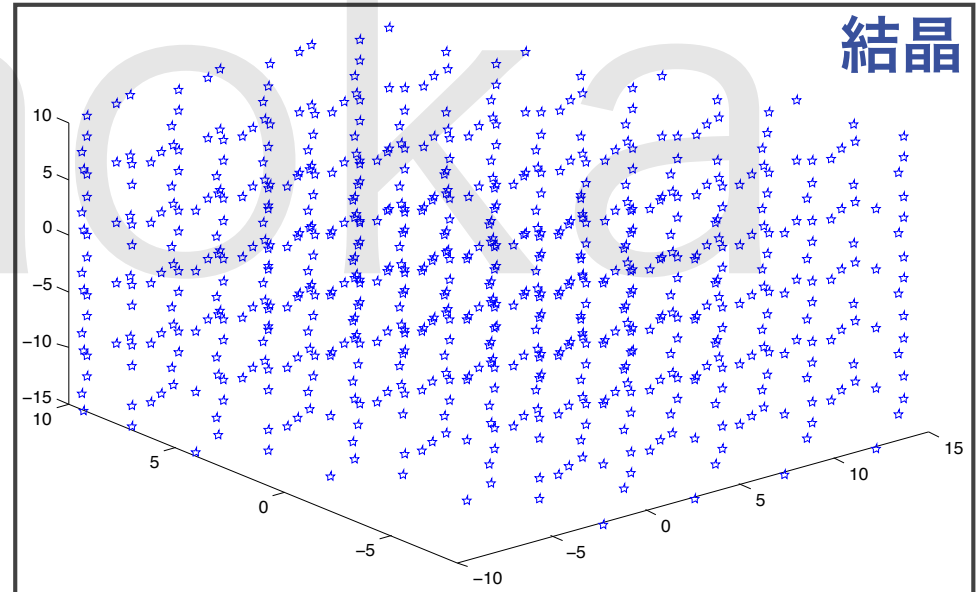
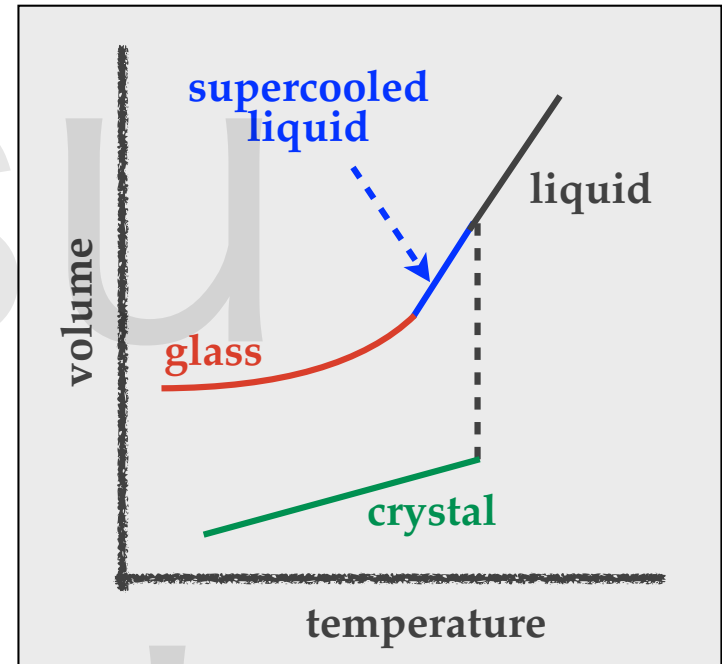
The Topology of the Cosmic Web in Terms of Persistent Betti Numbers.  
Edelsbrunnerが絡んでいるよう



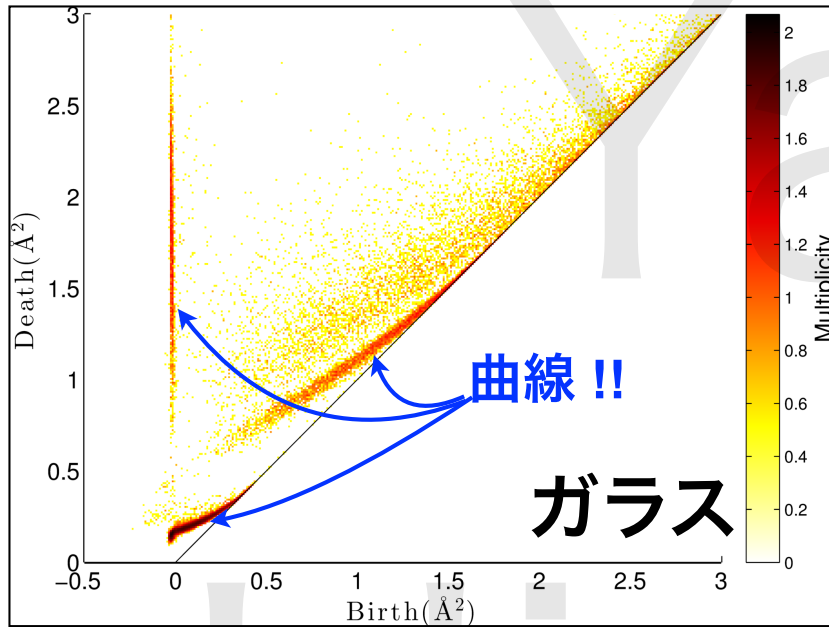
# シリカの原子配置



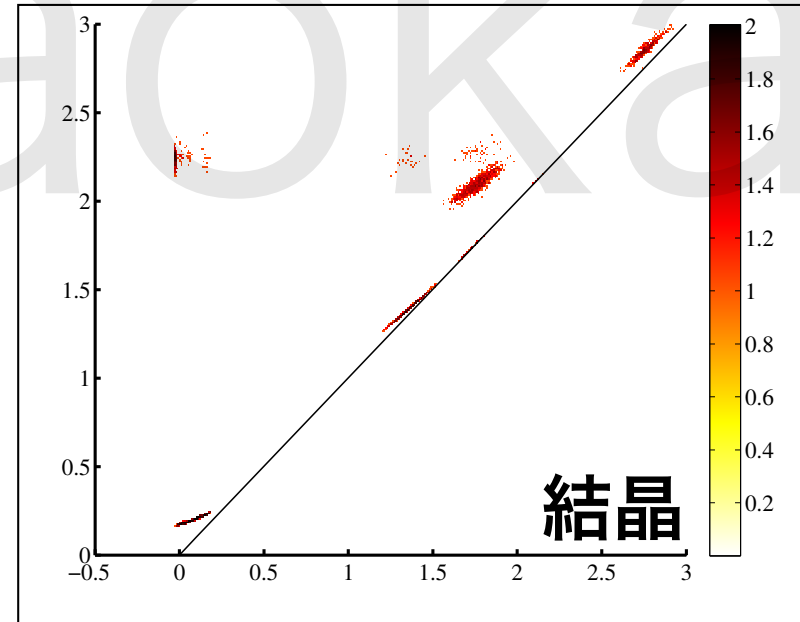
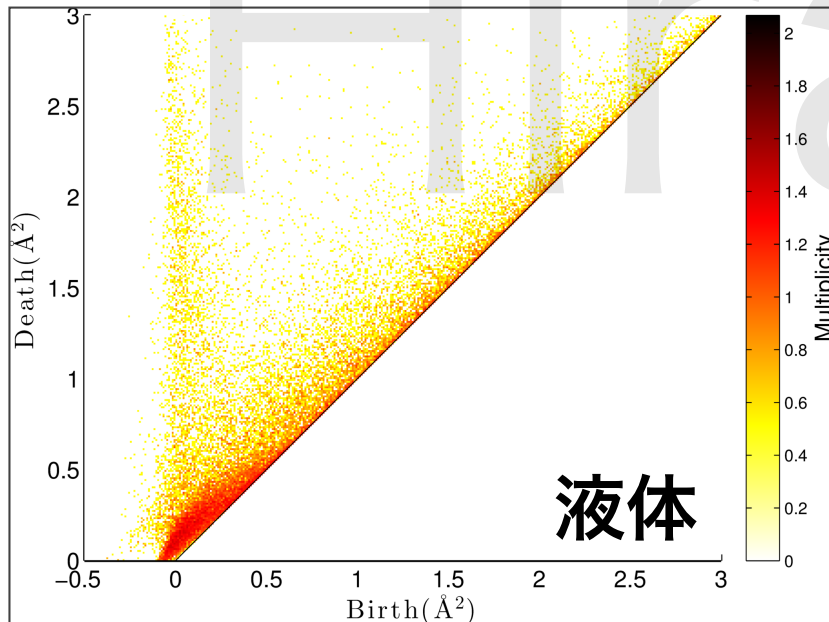
識別可能？ or 両者ともにランダム？



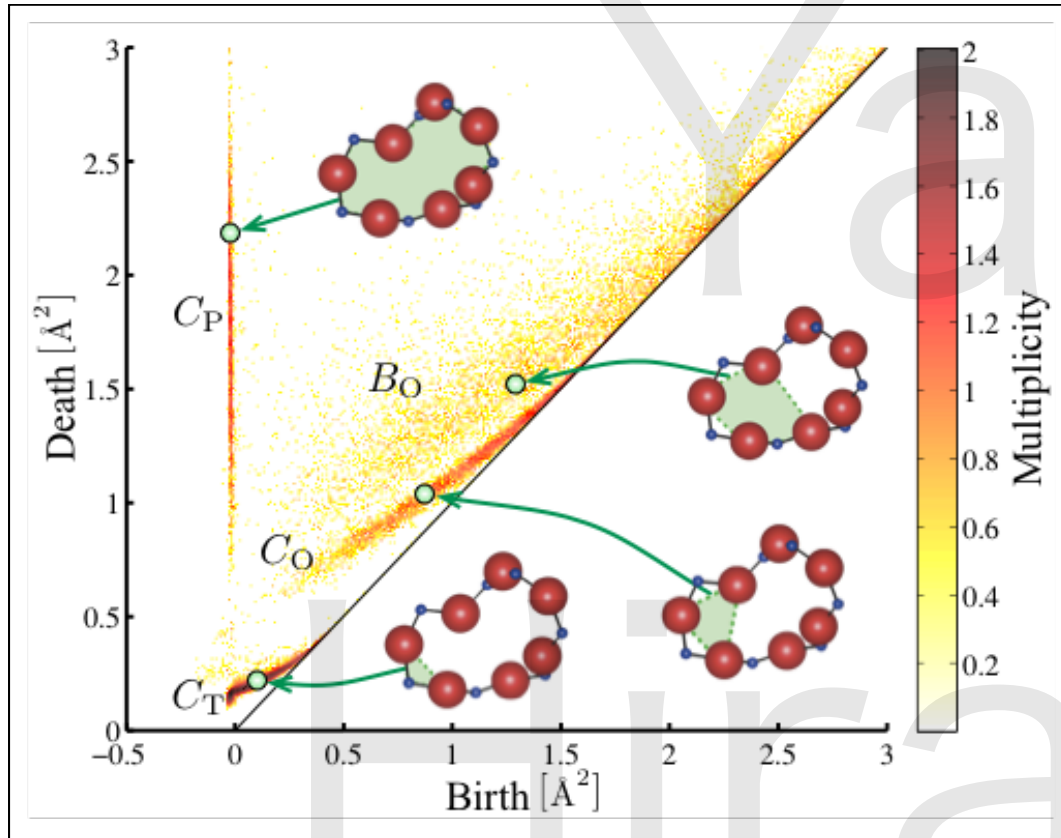
# シリカのパーシステント図



- PD1を表示 (リング構造に着目)
- 結晶の規則性は 0次元的分布
- 液体のランダム性は 2次元的分布
- ガラスは **1次元的分布 (曲線) !!**



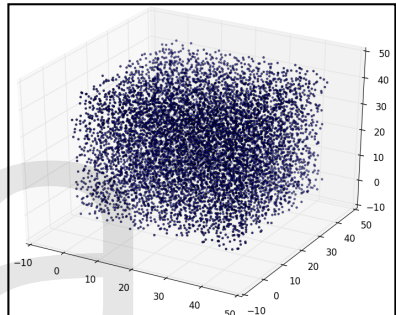
# ガラスの階層的幾何構造



ガラスの幾何構造

PD内の曲線の幾何学的な起源

逆問題



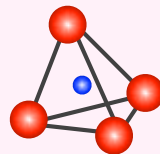
- optimal cycle  
Escobar and H. 2015.

- continuation  
Gameiro, Obayashi, H. Physica D, 2015

## 階層的リング構造

C<sub>P</sub>: primary rings generating the others → C<sub>O</sub>: three oxygen rings

C<sub>T</sub>: triangles on tetrahedra

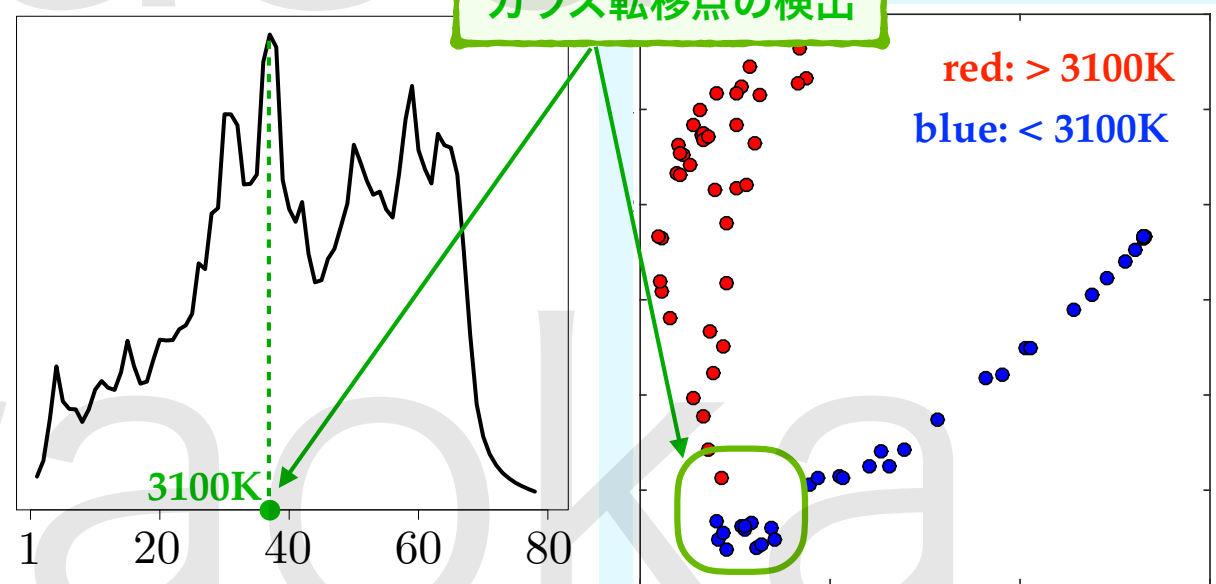
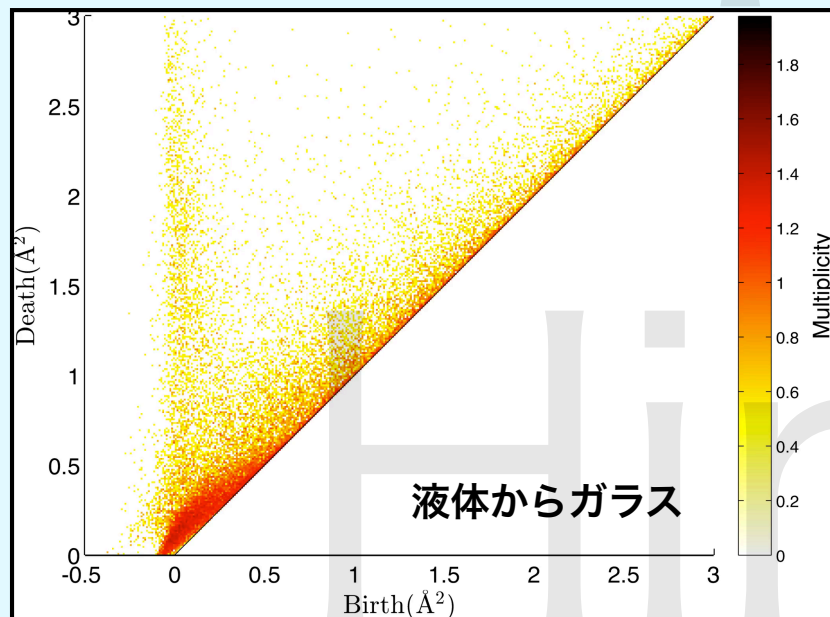


→ B<sub>O</sub>: oxygen rings (≥ four)

# パーシステント図と機械学習

## カーネル法

### Persistence Weighted Gaussian Kernel (PWGK)



カーネル変化点検出法

カーネルPCA

- ◆ PWGKを用いてガラス転移点の検出に成功
- ◆ 逆問題法を適用しガラス転移を誘発する幾何構造を解明
- ◆ マテリアルズインフォマティクスへの応用も進める

# 参考文献

シリーズ・現象を解明する数学  
Introduction to Interdisciplinary Mathematics:  
Phenomena, Modeling and Analysis

三村昌泰, 竹内康博, 森田善久: 編集

## 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

平岡裕章 著

共立出版

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)

平岡裕章, データに潜む幾何構造: パーシステントホモロジー-数理科学6月号 (2015) .

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).

東北大AIMR 平岡研究室webpage

[http://www.wpi-aimr.tohoku.ac.jp/hiraoka\\_lab/index.html](http://www.wpi-aimr.tohoku.ac.jp/hiraoka_lab/index.html)

**AYASDI webpage**

<https://www.ayasdi.com>

**Rob Ghrist webpage**

<https://www.math.upenn.edu/~ghrist/>