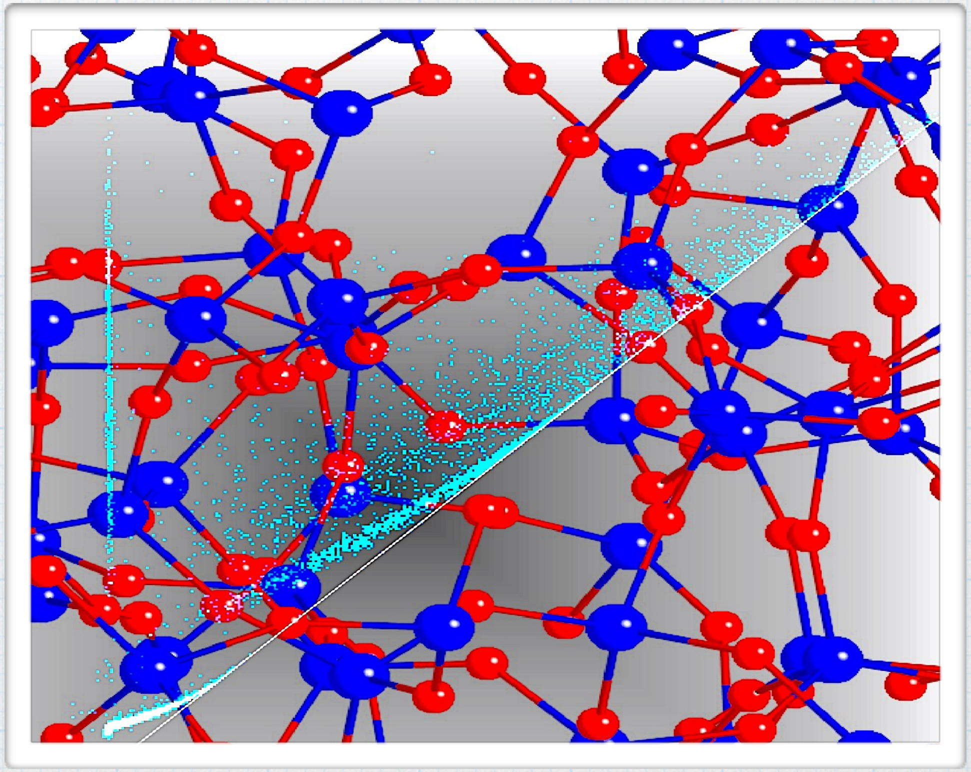
# パーシステントホモロジーと機械学習


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^2I

# 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データ解析手法
Data Has Shape, Shape Has Meaning, Meaning Drives Value

Gunnar Carlsson's Gr. (math. Stanford, AYASDI)

- （ビッグ）データ解析, ソーシャルネットワーク, 医療, 金融 etc

Robert Ghrist's Gr. (math. UPenn)

- 情報ネットワーク, センサーネットワーク

Konstantin Mischaikow's Gr. (math. Rutgers)

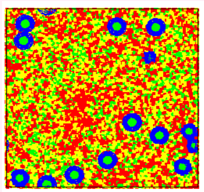
- 流体解析, データ時系列解析

東北大AIMR, CREST, SIP, MI^2I

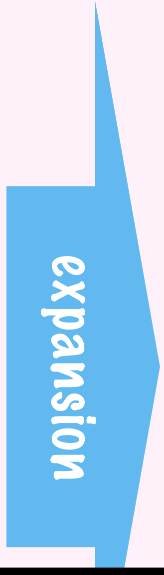
- 材料科学（ガラス, 粉体, 高分子, 金属, 蓄電池 etc）

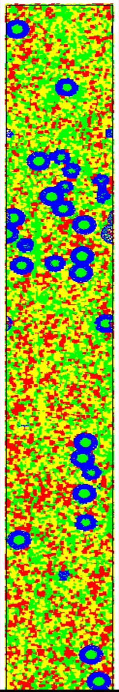
# 位相的データ解析の材料科学への応用

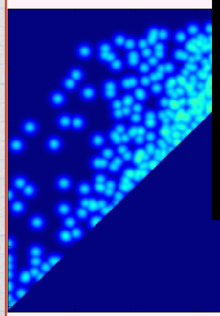
## 高分子



AFM
imag

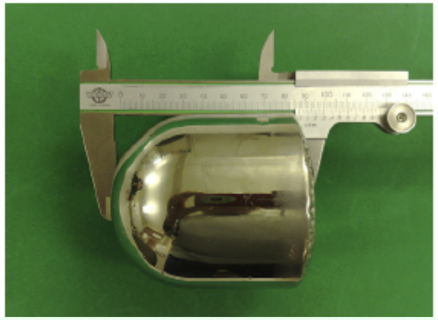


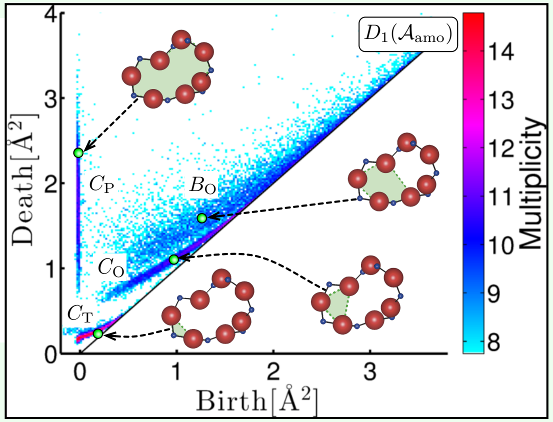


## アモルファス

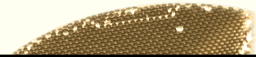
silica

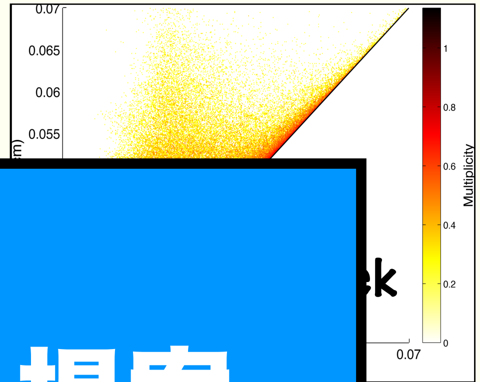


metallic glass



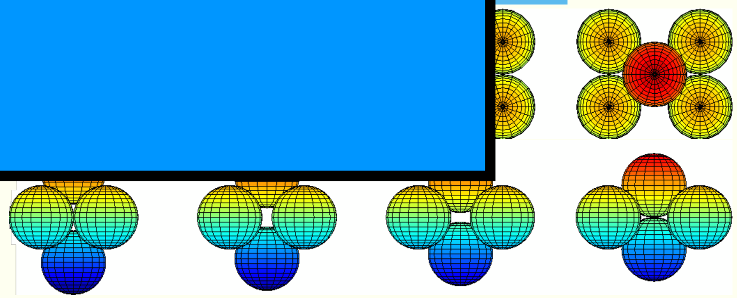

## 粉体






k

ta.



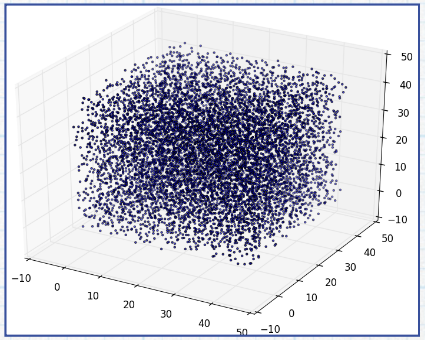
Joint with ANU
(M.Saadatfar etc)

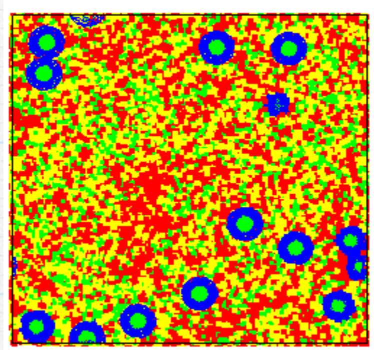
**無秩序系の記述言語を提案**

# Shape of Data

## Data

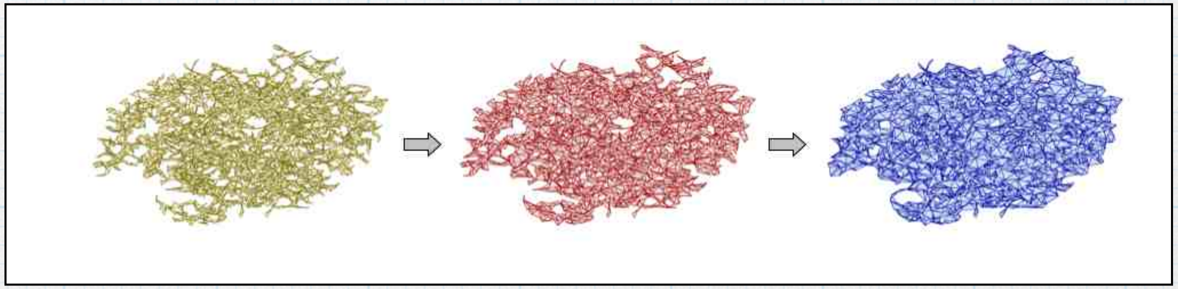
point cloud, atomic configuration, digital image, sensor network





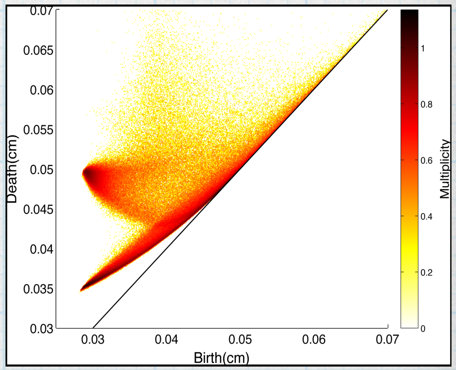
## Geometric Model

Polyhedron (Alpha, Čech, cell complex)



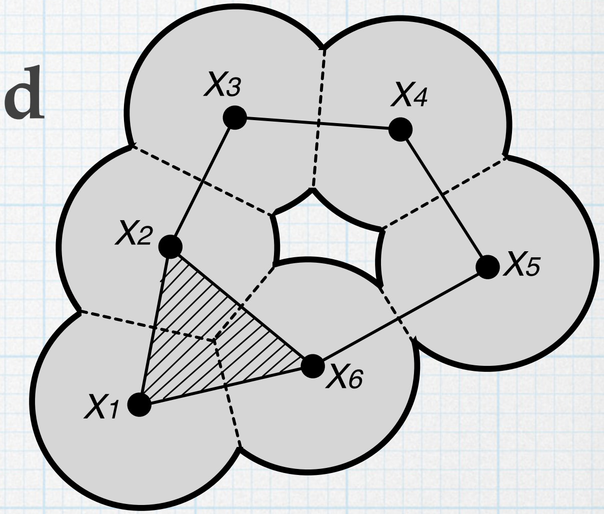
## Algebraic Description

Persistent homology, sheaf cohomology, category



Edelsbrunner & Mücke '94

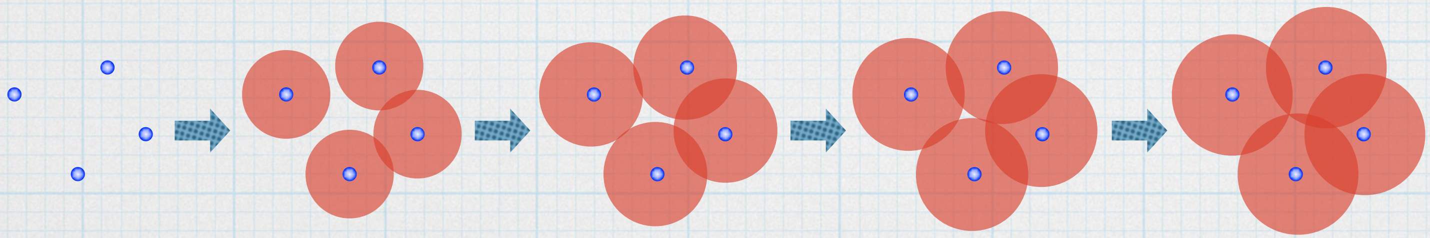
# Alpha shape

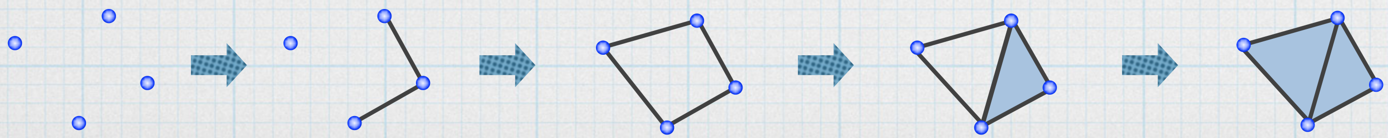



- $X = \{x_i \in \mathbf{R}^m \mid i = 1, \ldots, n\}$ : point cloud
- $\mathbf{R}^m = \cup_i V_i$ : Voronoi decomp.
- $\cup_i B_i(r) = \cup_i (B_i(r) \cap V_i)$
- Alpha shape $\mathcal{A}(X, r)$ : dual of $\{B_i(r) \cap V_i \mid i = 1, \ldots, n\}$ (simplicial complex)
- Nerve theorem: $\cup_i B_i(r) \simeq \mathcal{A}(X, r)$

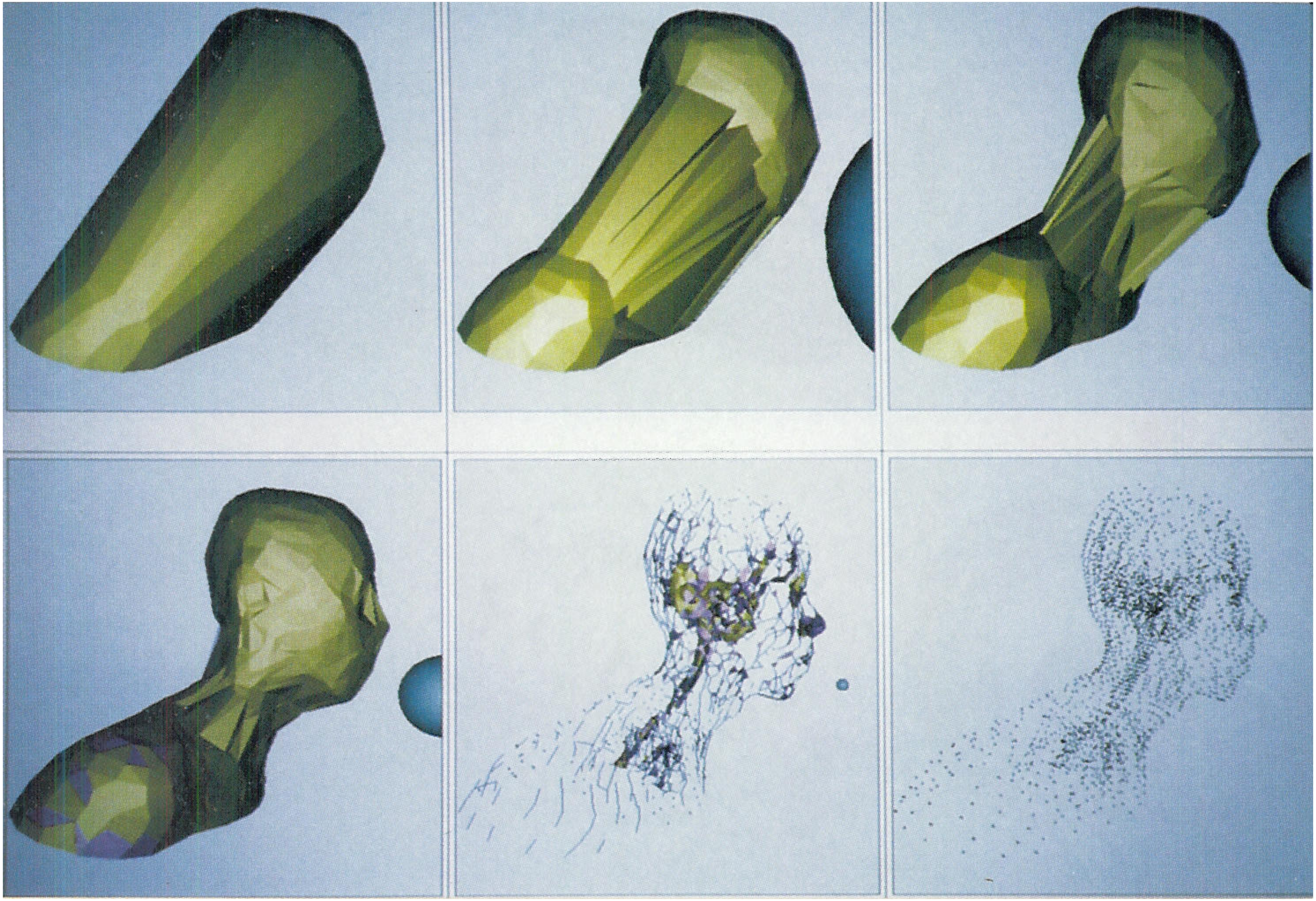
easier to analyze by computers

- $\mathcal{A}(X, r) \subset \mathcal{A}(X, s)$ for $r < s$





filtration:
changing resolution

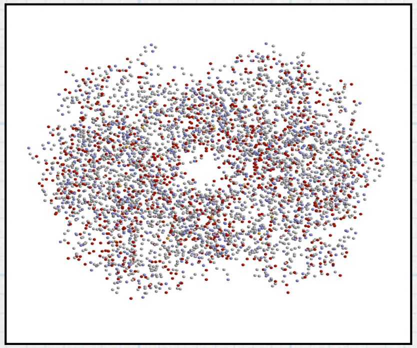


(ref. Edelsbrunner, Mucke)

# パーシステントホモロジー

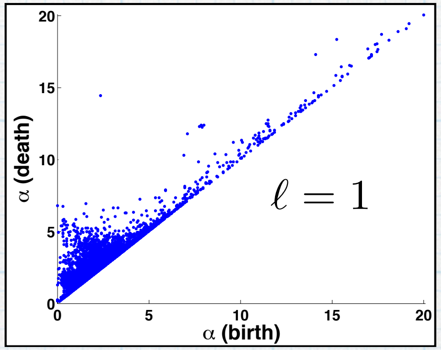
データ —入力→ $PH_\ell$ —出力→ $\ell$ 次元の穴のサイズや形状の情報

ヘモグロビン



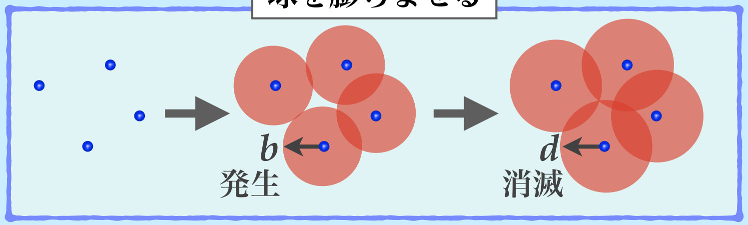
パーシステントホモロジー →

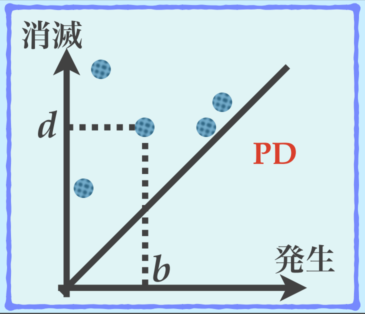
パーシステント図(PD)




**PD構成法**

球を膨らませる



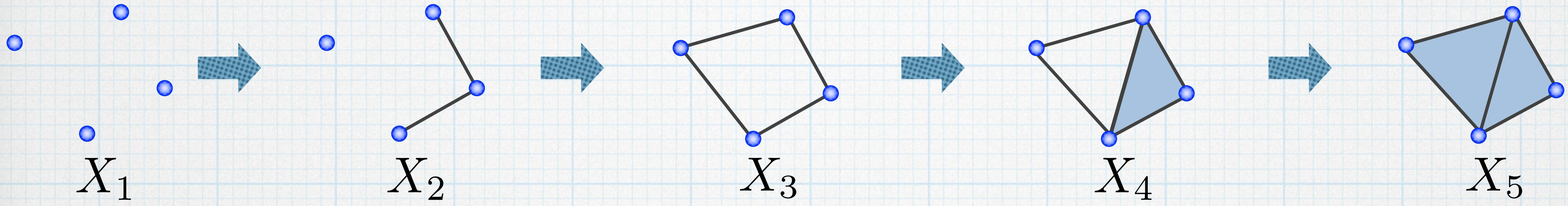

- PDの各点はデータ内の穴
- 発生軸は穴の発生パラメータ
- 消滅軸は穴の消滅パラメータ
- 対角線付近の点はノイズ
- 対角線から離れた点はロバスト

- 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

Edelsbrunner, Letscher, Zomorodian, Carlsson, de Silva

# Persistent Homology

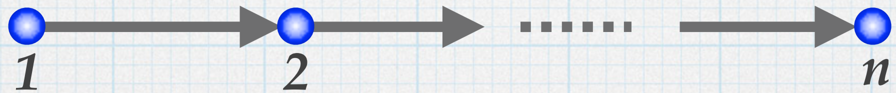
$PH_1(\mathcal{X}) \simeq I[3,4]$




- **filtration** $\mathcal{X} : X_1 \subset X_2 \subset \cdots \subset X_n$
- **persistent homology**

representations on $A_n$

$$H_\ell(\mathcal{X}) : H_\ell(X_1) \to H_\ell(X_2) \to \cdots \to H_\ell(X_n)$$




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

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

generator

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

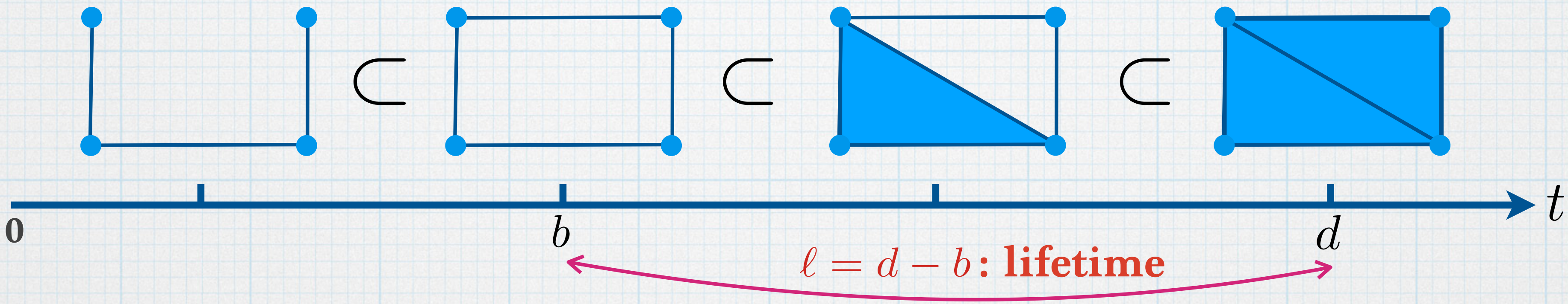
at $X_b$ at $X_d$

# Persistence Diagram

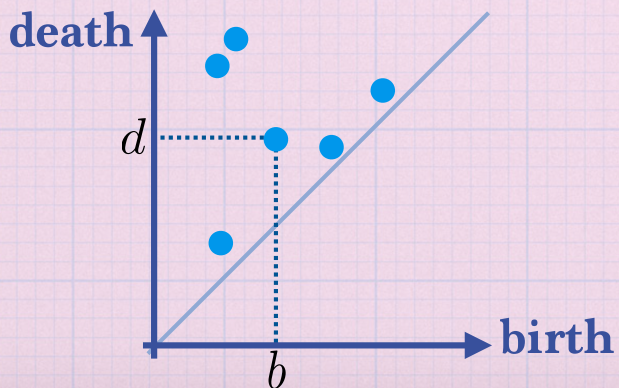
**Interval decomp:** $H_k(\mathcal{X}) \simeq \bigoplus_{i=1}^{p} I[b_i, d_i]$

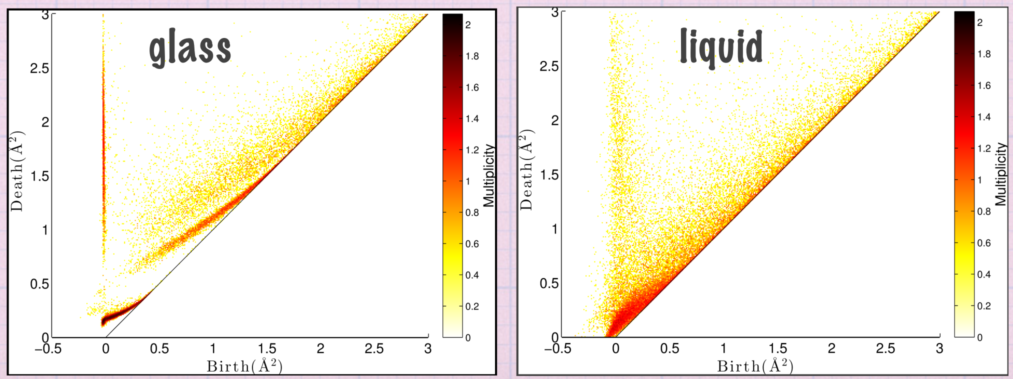
- $I[b,d]$ **represents appearance and disappearance of a topological feature at** $t = b, d$ **in** $\mathcal{X} = \{X(t)\}_t$

**birth time** **death time**



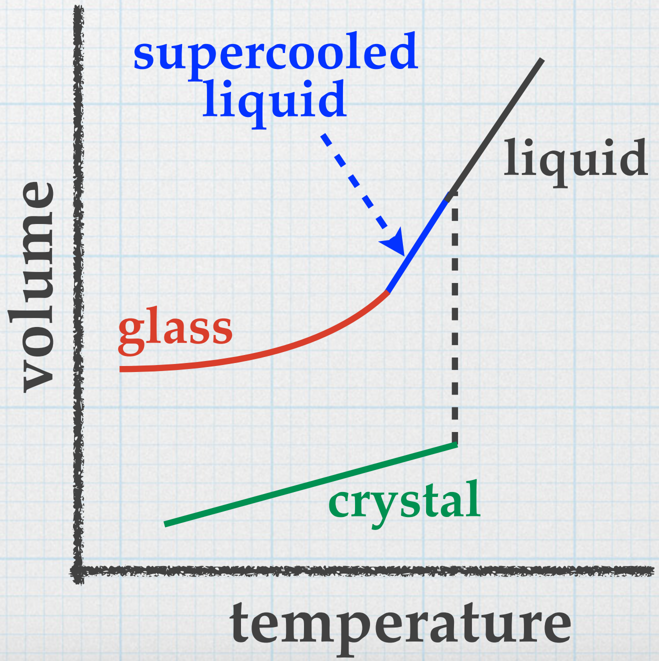

- $D_k(\mathcal{X}) = \{(b_i, d_i) \in \mathbb{R}^2_{\geq 0} : i = 1, \ldots, p\}$: **persistence diagram**



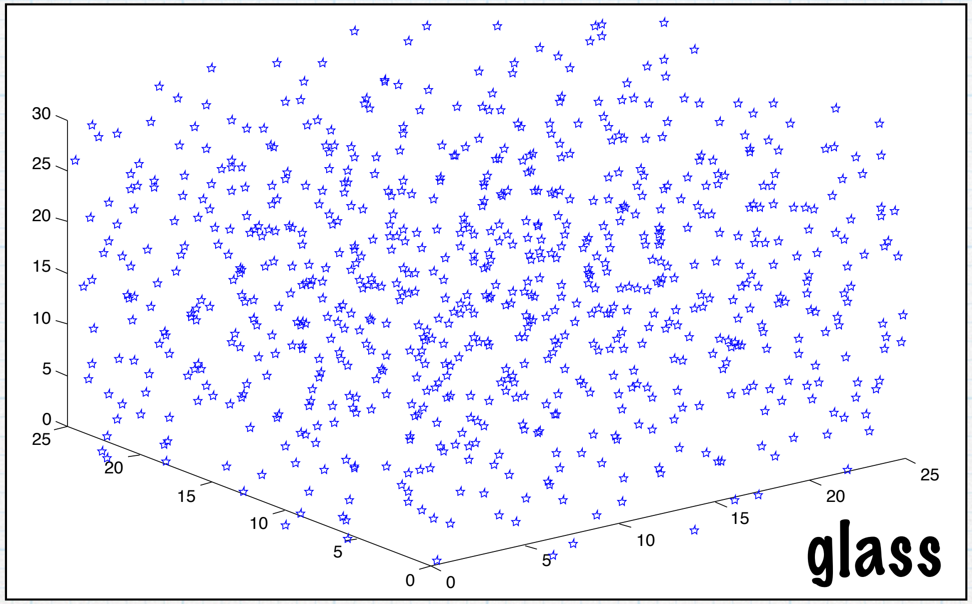



# What is glass?

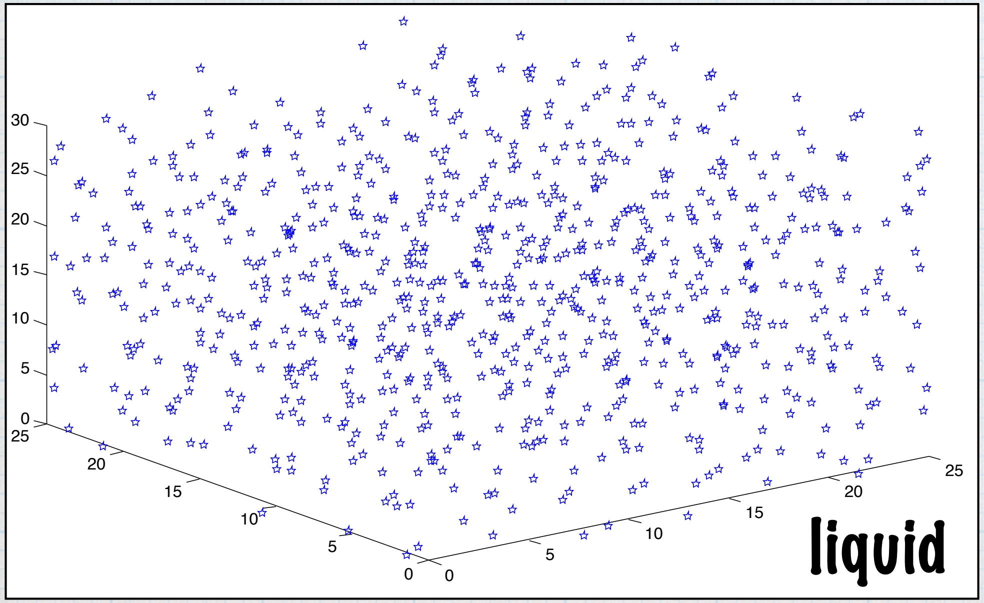
* Not yet fully answered to "what is glass?"
* Not liquid, not solid, but something in-between
* Atomic configuration looks random, but sufficient cohesion to maintain rigidity
* Further geometric understandings of atomic configurations are required
* Solar Energy Glass, DVD, BD, etc.

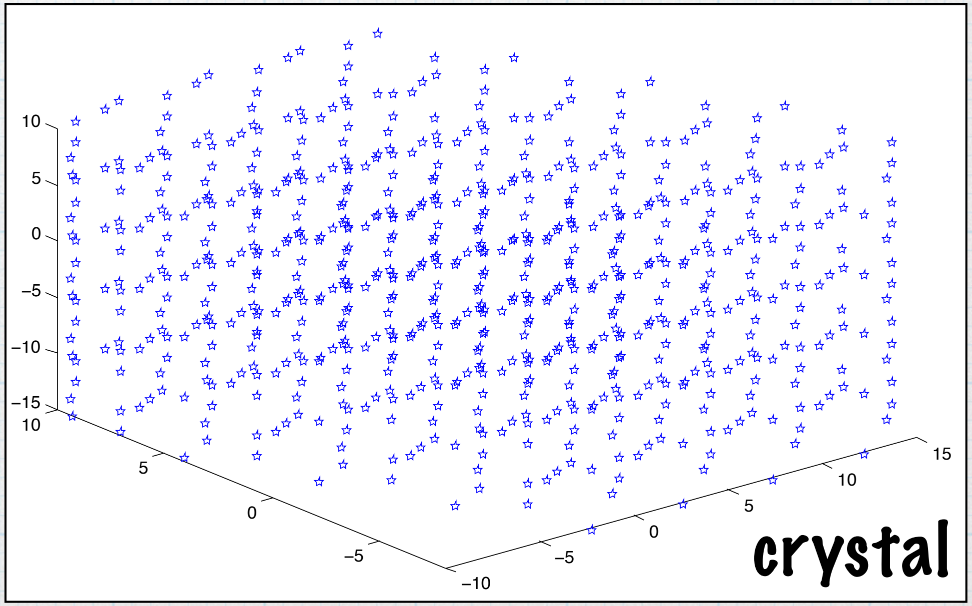


# Atomic configurations of silica ($SiO_2$)




obtained by molecular dynamics (MD) simulation

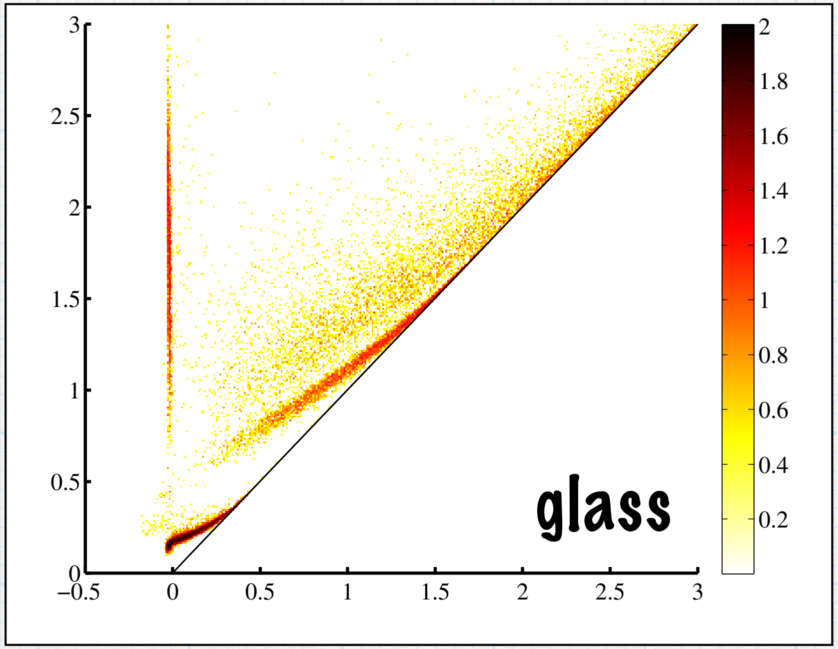


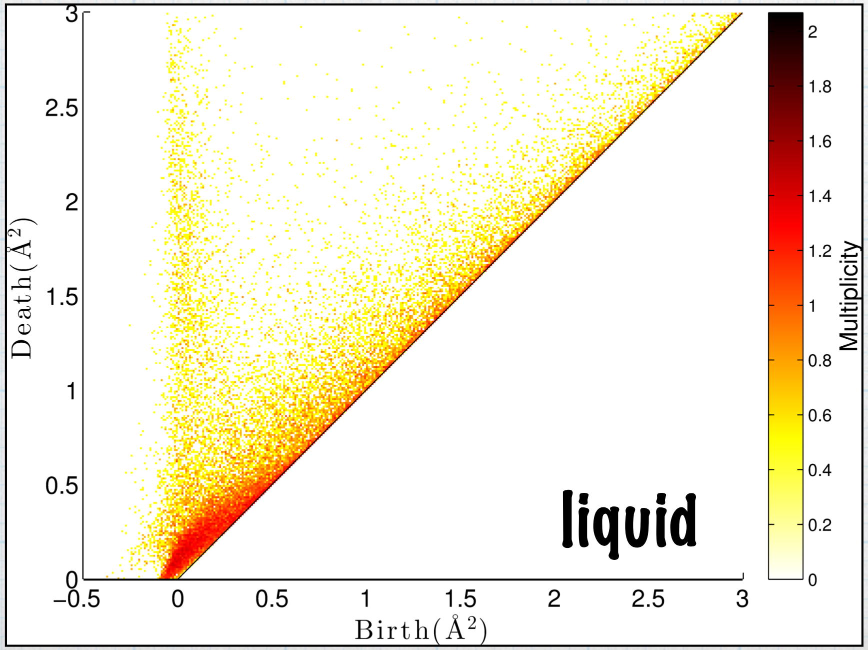
Y.H., et al. PNAS (2016)

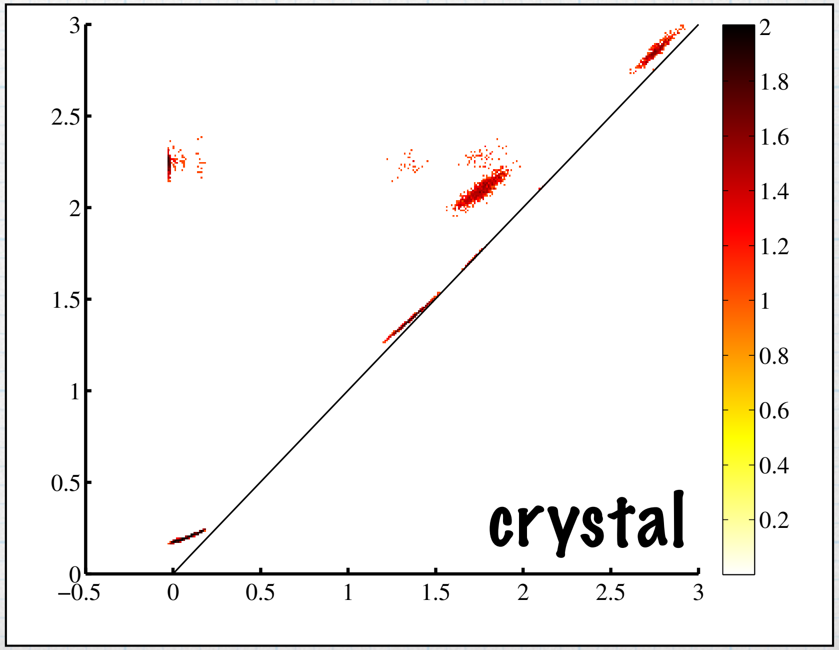
# 1 dim persistence diagrams of silica

All PD computations are performed by HomCloud



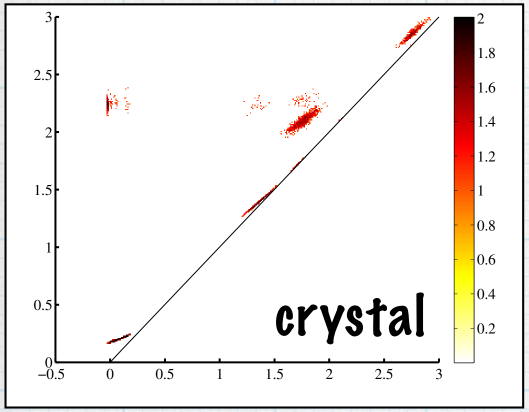

PD1 of alpha filtrations



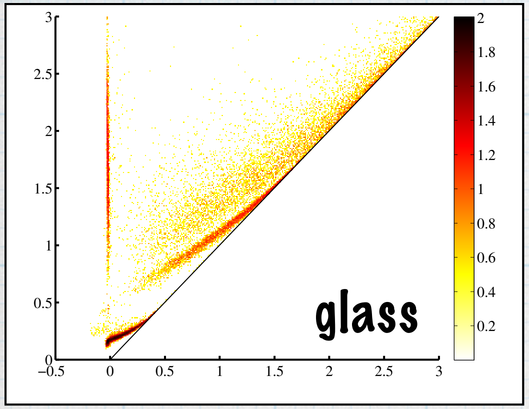
Y.H., et al. PNAS (2016)

# Support dim and order parameter



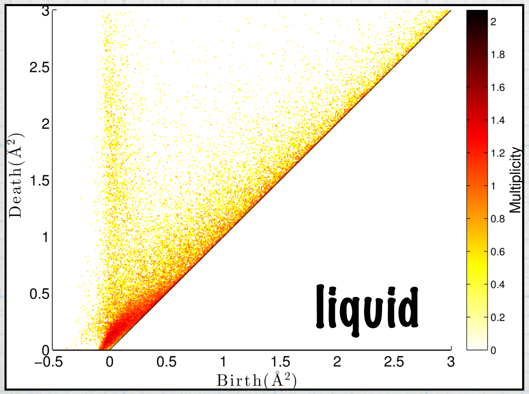

**0-dim** support results from **periodic** atomic locations of crystals

transition




**1-dim** support (curves!) appears

transition

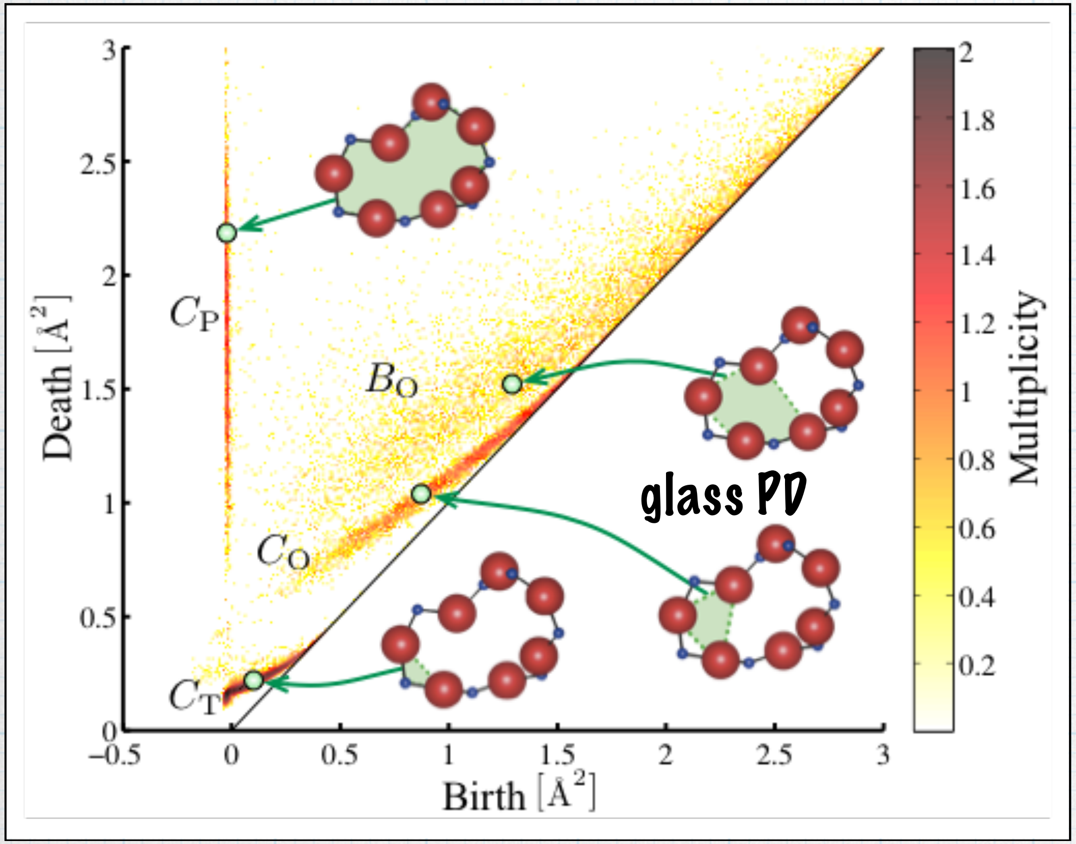



**2-dim** support results from **random** atomic locations of liquids

Remark : This suggests a topological approach for glass transition.

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

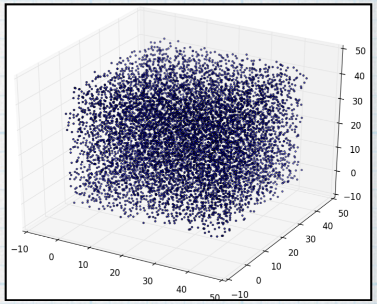
# Geometric origins of curves: inverse problem




what characterizes glass structures?

what is the geometric origin of the curves?

inverse problem!

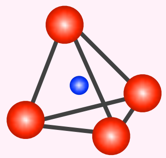


- optimal cycle
  Dey, et.al, 2011, Escolar and H, 2015
- continuation
  Gameiro, Obayashi, H. Physica D, 2015

## 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)**

Kusano, Fukumizu, Y.H. ICML (2016)

# PD and kernel method

- vectorization of PDs for statistics (persistence landscape, etc)
- Reininghaus et al. proposes the persistence scale-space kernel (PSSK) on PDs based on a heat diffusion

- **Our method: persistence weighted Gaussian kernel (PWGK)**
- **PWGK can explicitly control the effect of persistence**

**Let $\Omega$ be a LCH and $k : \Omega \times \Omega \to \mathrm{R}$ be a positive def. kernel**

$\mathcal{H}_k$ **: reproducing kernel Hilbert space (RKHS) s.t.** $f(x) = \langle f, k(\cdot, x) \rangle_{\mathcal{H}_k}$ $\forall x \in \Omega, \forall f \in \mathcal{H}_k$

$M_b(\Omega)$ **: space of all finite signed Radon measures on** $\Omega$

$k$ **is called** $C_0$**-universal if**

- $k(x,x)$ **is a** $C_0$**-function (cont. & vanishing at** $\infty$ **)**
- $\mathcal{H}_k$ **is dense in** $C_0(\Omega)$

**Thm: Kernel embedding of distributions (Fukumizu et al, 2011)**

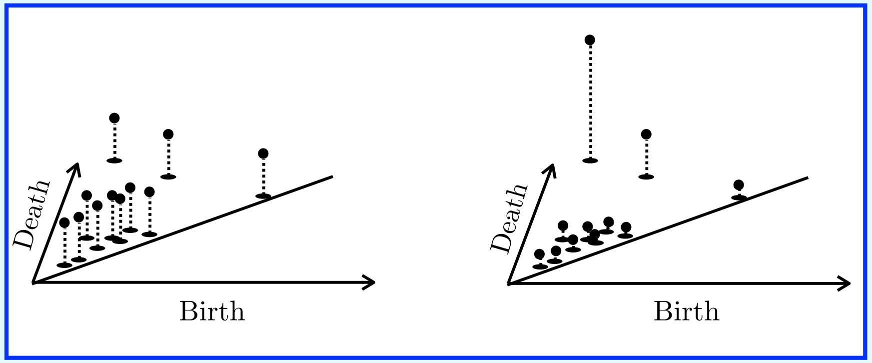
- $\Phi_k : M_b(\Omega) \to \mathcal{H}_k, \ \mu \mapsto \int k(\cdot, x) d\mu(x)$ **is injective.**
- $d_k(\mu, \nu) = \|\Phi_k(\mu) - \Phi_k(\nu)\|_{\mathcal{H}_k}$ **defines a distance on** $M_b(\Omega)$**.**

Kusano, Fukumizu, Y.H. ICML (2016)

# Persistence weighted Gaussian kernel (PWGK)

- **weighted persistence diagram**

$$\mu_D^w = \sum_{x \in D} w(x)\delta_x$$




**ex)** $w_{\mathrm{arc}}(x) = \tan^{-1}(C\mathrm{pers}(x)^p)$

$C$ **and** $p$ **explicitly control the effect of persistence in statistics.**

- **PWGK**

**The Gaussian kernel** $k_G(x,y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$ **is C0-universal**

➡ $\Phi_{k_G} : M_b(\Delta) \to \mathcal{H}_{k_G},\ \mu_D^w \mapsto \int k_G(\cdot, x) d\mu_D^w(x)$ **becomes injective**

$d_{k_G}^w(D,E) = \|\Phi_{k_G}(\mu_D^w) - \Phi_{k_G}(\mu_E^w)\|_{\mathcal{H}_{k_G}}$ **defines a distance**

- **Stability Theorem**

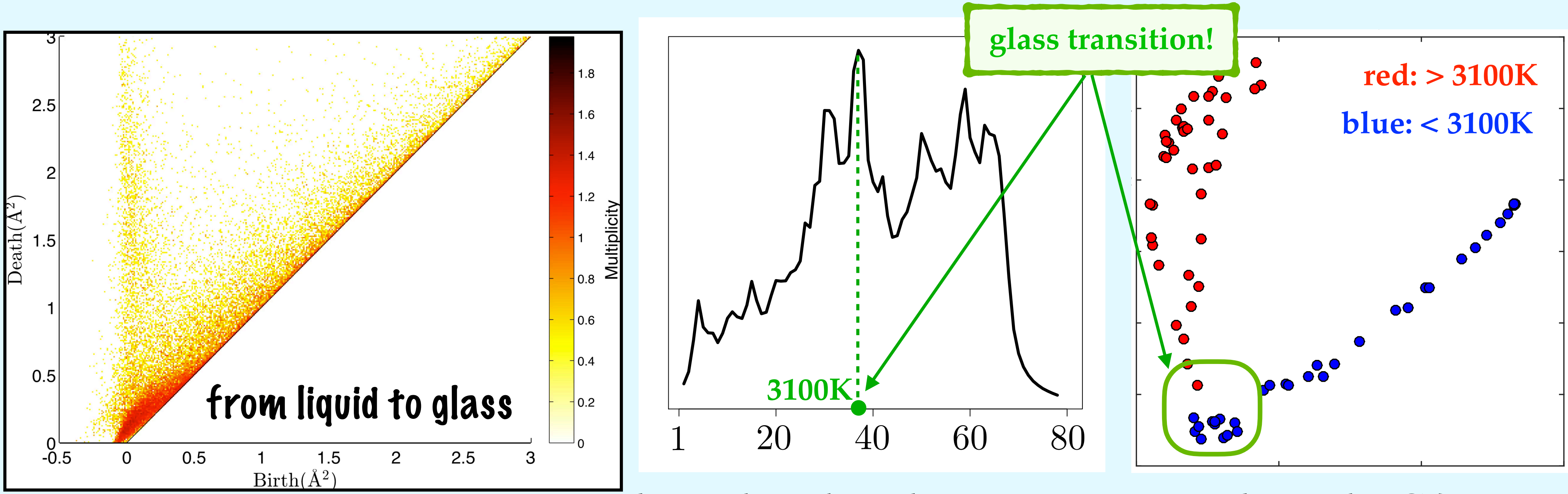
**Let** $X, Y \subset \mathrm{R}^d$ **be finite subsets and** $p > d+1$. **Then,**

$$d_{k_G}^{w_{\mathrm{arc}}}(D(X), D(Y)) \leq L d_H(X,Y)$$

Kusano, Fukumizu, Y.H. ICML (2016)

# PD detects glass transition and its geometry

**PWGK**




**kernel Fisher discriminant** **kernel PCA**

✦ **PWGK successfully detects GT**

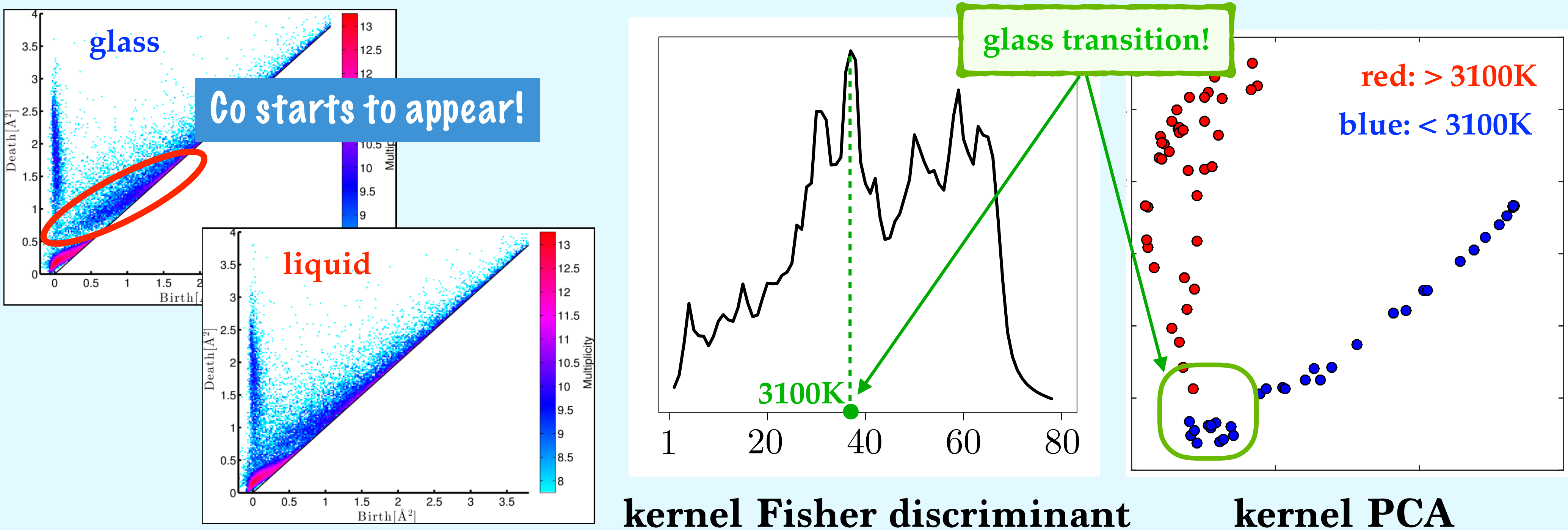
✦ **What is the geometric trigger to generate the phase transition?**

→ **Inverse problem!**

Kusano, Fukumizu, Y.H. ICML (2016)

# PD detects glass transition and its geometry

**PWGK**



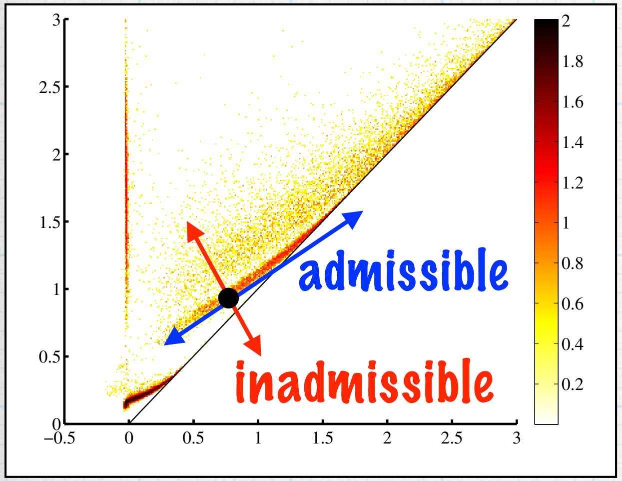

**kernel Fisher discriminant** **kernel PCA**

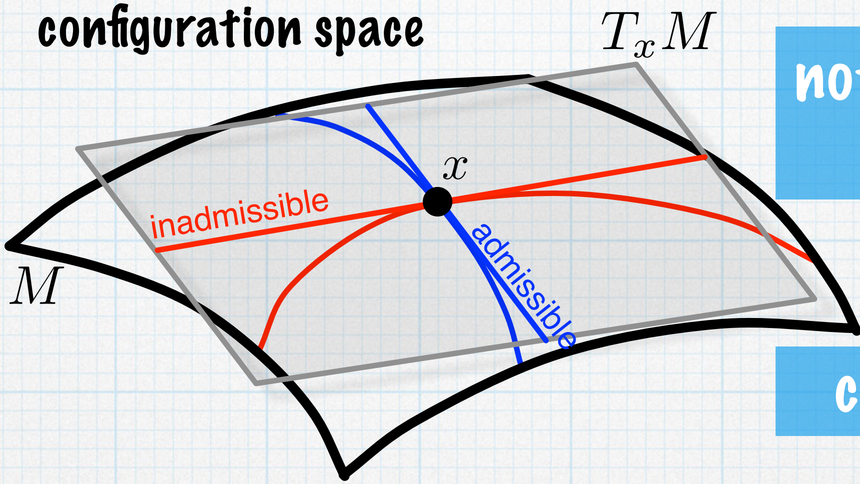
- ✦ **PWGK successfully detects GT**
- ✦ **What is the geometric trigger to generate the phase transition?**

→ **Inverse problem!**

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

# Curves and constrains



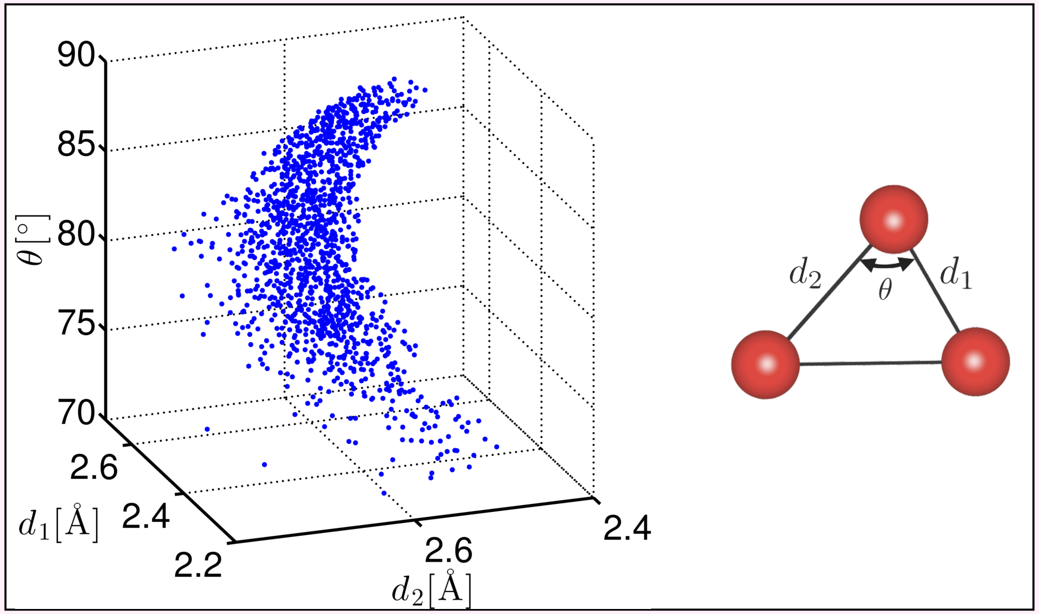




not allowed to move in the normal direction

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)