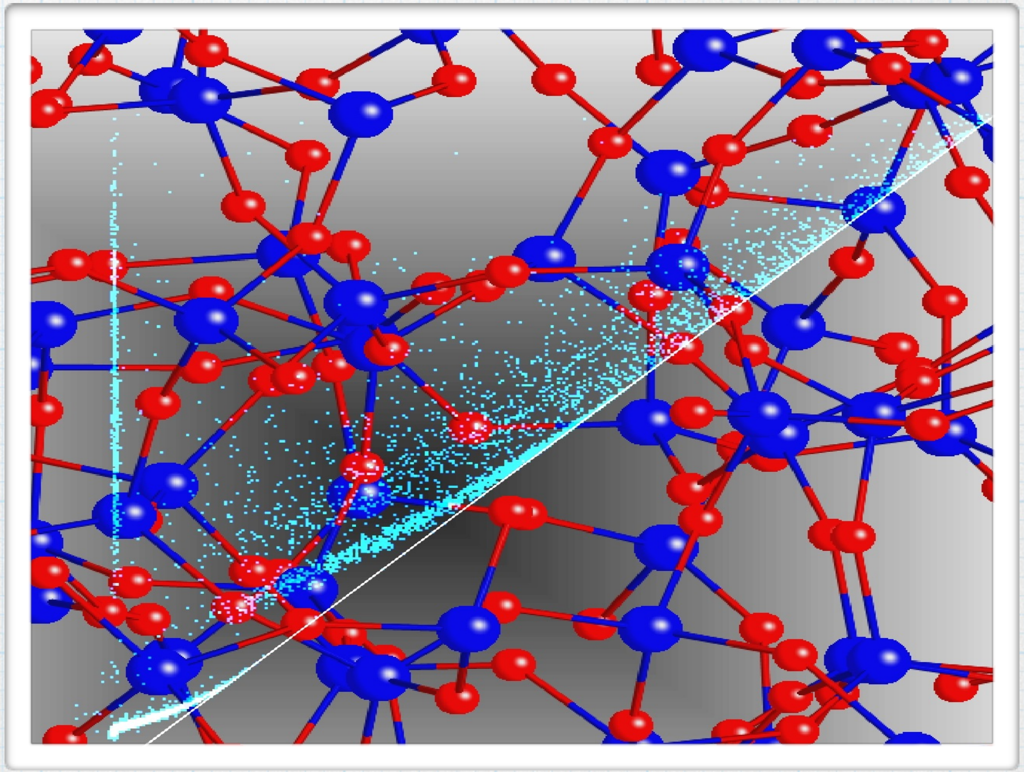
Topological Data Analysis on Materials Science

材料科学への位相的データ解析

平岡 裕章

Yasuaki Hiraoka

WPI-AIMR, Tohoku



JST CREST 「ソフトマター記述言語の創造に向けた位相的データ解析理論の構築」

Content

1. Overview of TDA

2. Persistent homology, persistence diagram

3. Materials TDA: glass analysis, materials info, proteins

4. Building mathematics from materials science:
   probability, representation, statistics, inverse problem

Collaborators

random topology

Tomoyuki Shirai

topological statistics

Kenji Fukumizu
Genki Kusano

representation

Hideto Asashiba

inverse problem

Ippei Obayashi

materials TDA

Hirata, Nakamura, Matsue, Saadatfar, Ichinomiya, Nakajima
Kotani, Nishiura, Kramar, Kohara, etc

Topological Data Analysis

TDA starts from 21st century for analyzing complicated data geometrically and topologically

Data Has Shape, Shape Has Meaning, Meaning Drives Value

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

- bigdata, social network, medical science, finance, etc

Robert Ghrist's Gr. (math. UPenn)

- network flow, sensor network

Konstantin Mischaikow's Gr. (math. Rutgers)

- complex fluid, time series data

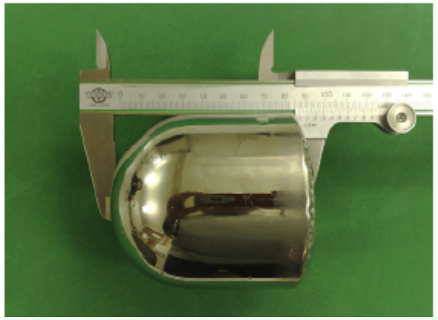
Our group in AIMR

- materials science

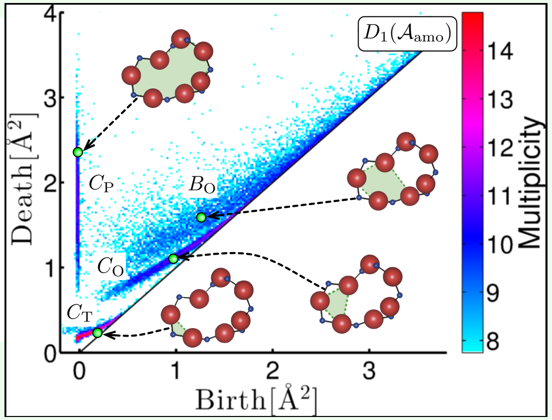
Materials TDA in AIMR

Glass

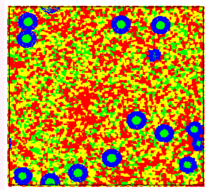
silica



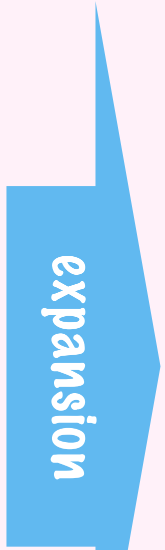
metallic glass

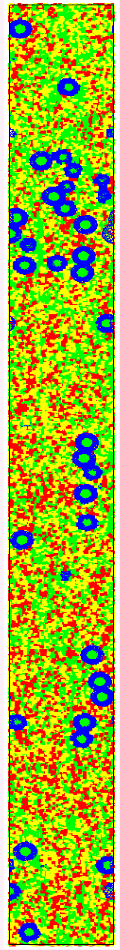


Polymer

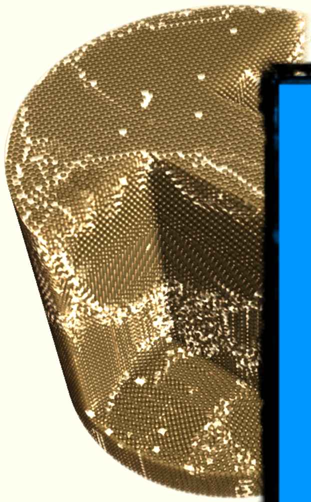


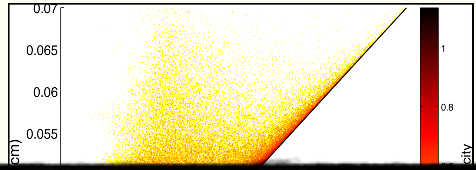
AFM



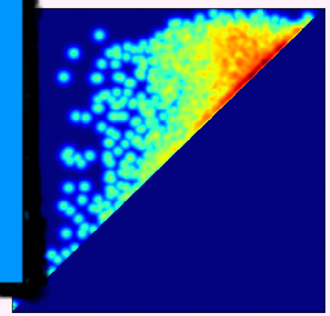


Granular system





Joint with
(M.Saadatf

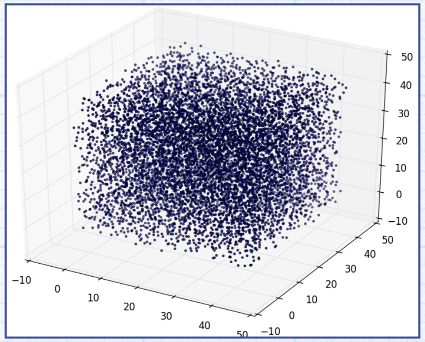


Need descriptors for order in disorder systems

Shape of Data

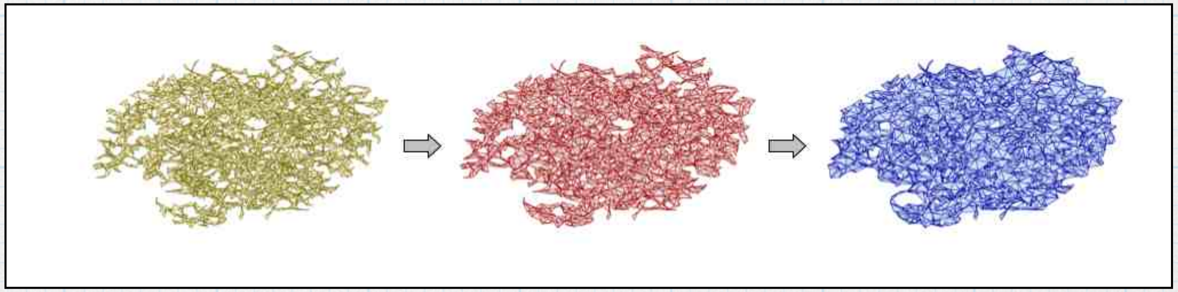
Data

point cloud, atomic configuration, digital image, sensor network



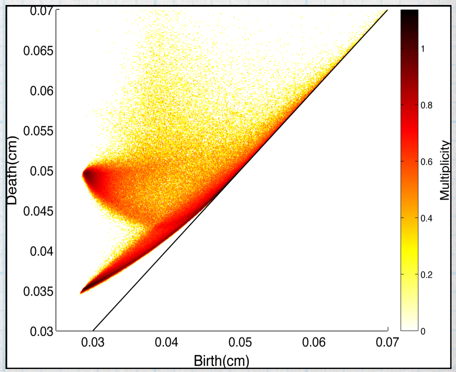
Geometric Model

simplicial complex, cell complex, quiver



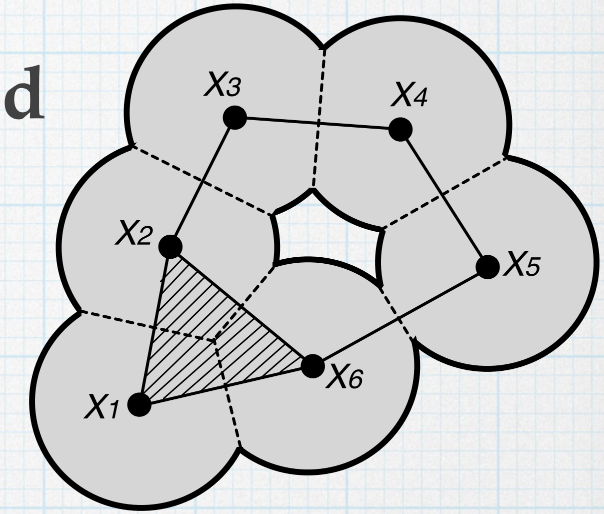
Algebraic Description

Persistent homology, sheaf cohomology, category



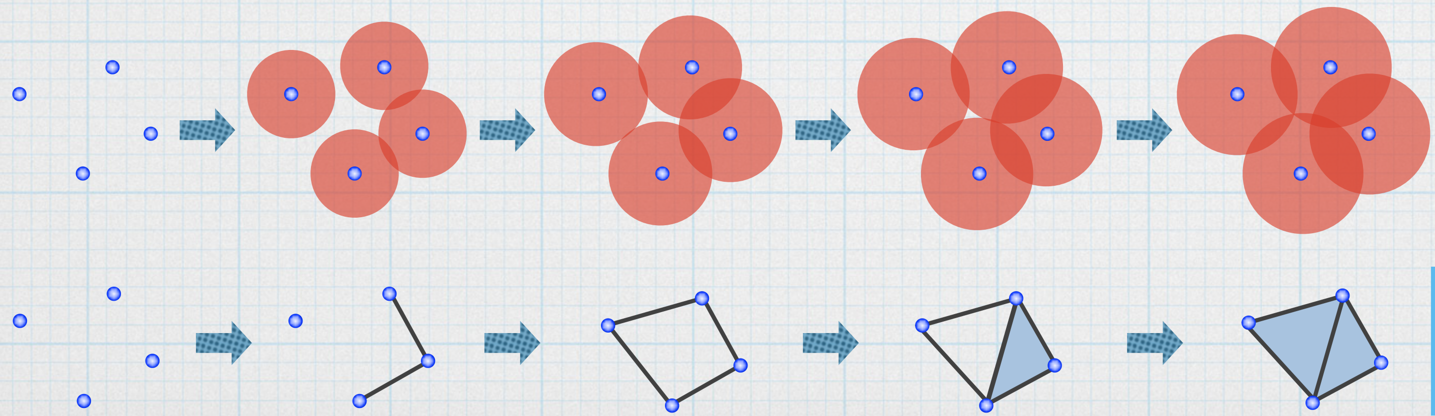
Edelsbrunner & Mücke '94

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

easier to analyze by computers



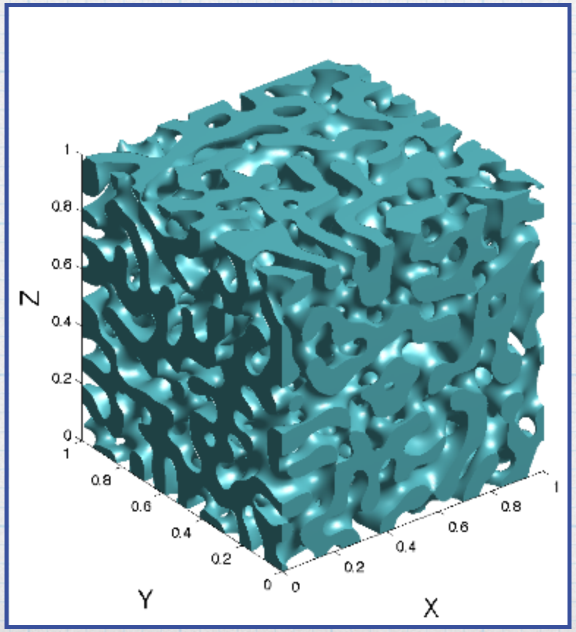
filtration:
changing resolution

Homology

geometric object $\xrightarrow{\text{input}}$ H_ℓ $\xrightarrow{\text{output}}$ number of ℓ -dim *"holes"* (ring, cavity, etc.)

homology

input



output

$H_0 = K$: 1 component

$H_1 = K^{847}$: 847 rings

$H_2 = 0$: 0 cavity

note1: K is a field ($K = \mathbf{R}, \mathbf{Q}, \mathbf{C}$ etc)

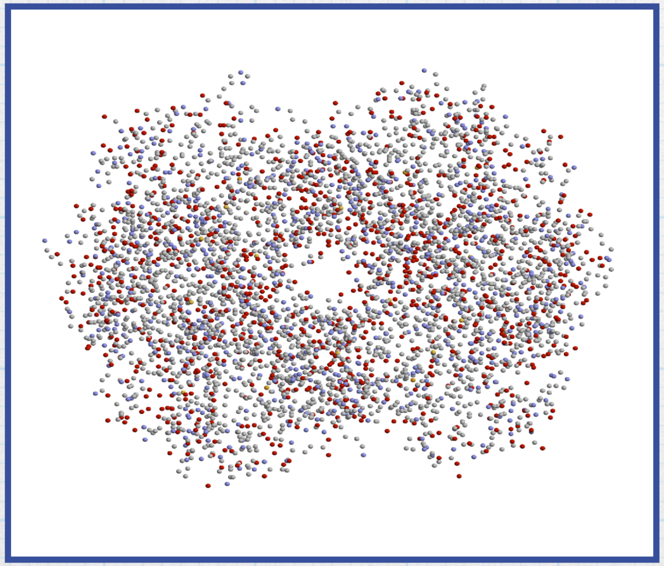
note2: CHomP/PHAT for computations

Persistent Homology

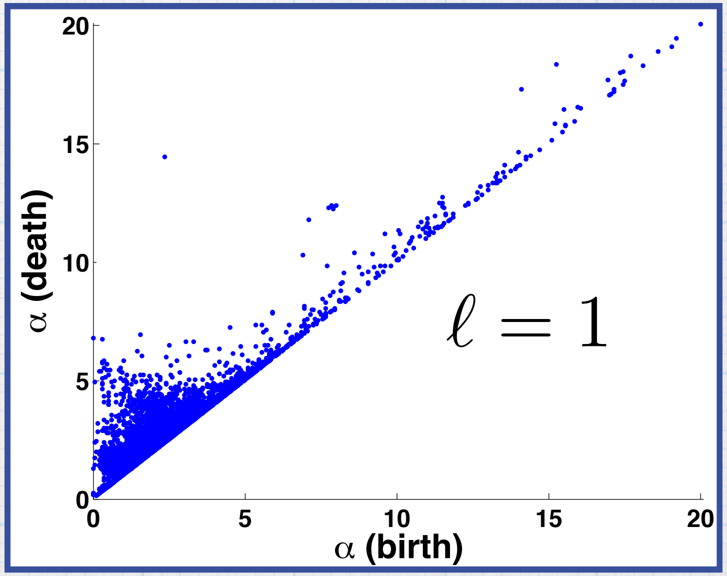
(filtered) geometric object **input** → PH_ℓ **output** → **metric info.** of **ℓ-dim holes**

persistent homology

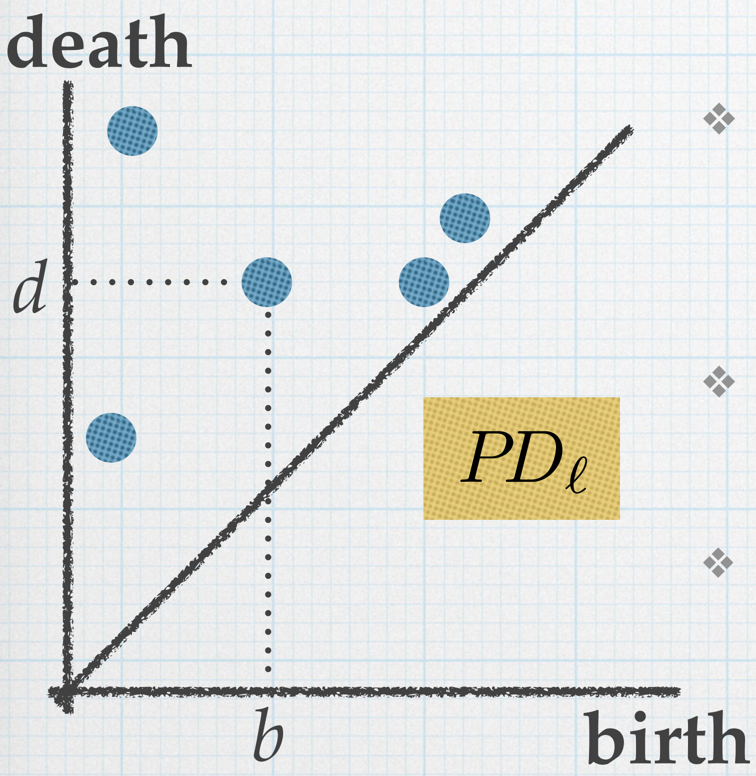
input: hemoglobin



output: persistence diagram

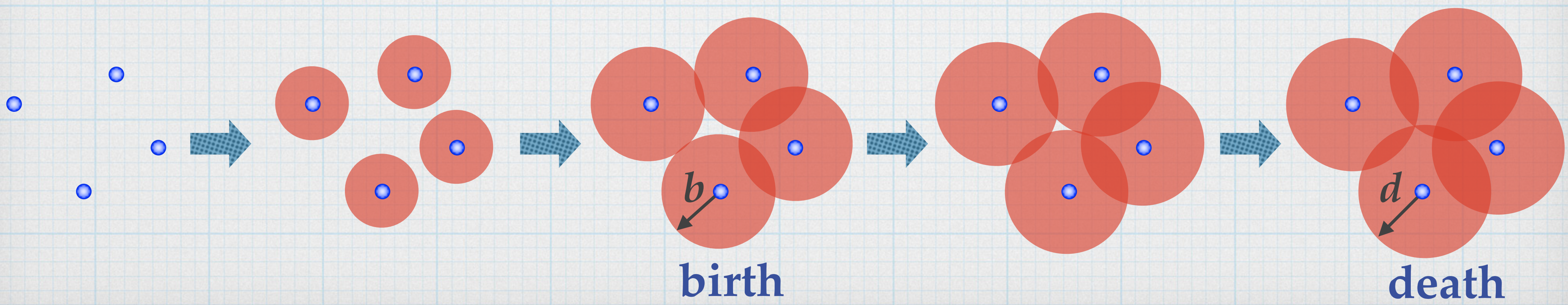


Persistence Diagram



- It consists of finite points in $\mathbf{R}^2$ above the diagonal
- Each point ● expresses an ℓ-dim hole
- b & d express birth and death scales

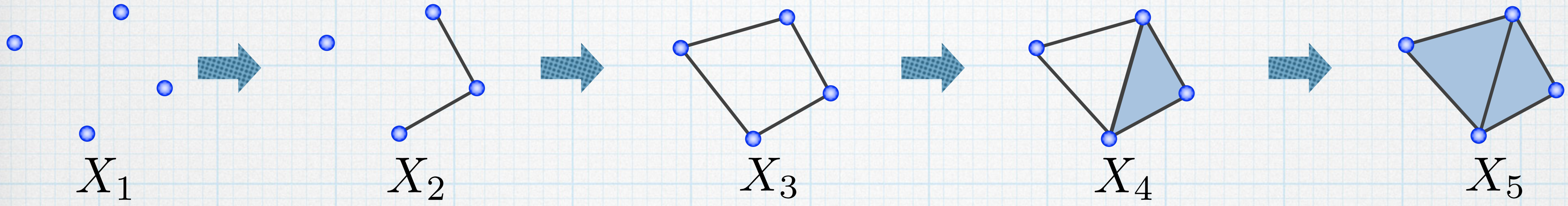
radius



Edelsbrunner, Letscher, Zomorodian, Carlsson, de Silva

Persistent Homology

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

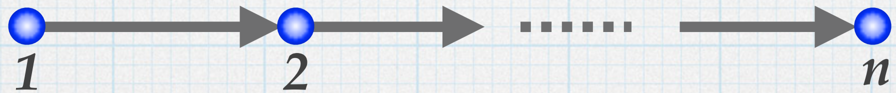


- filtration $\mathcal{X} : X_1 \subset X_2 \subset \cdots \subset X_n$

representations on A_n

- persistent homology

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



- interval decomposition (Gabriel's Theorem)

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

generator

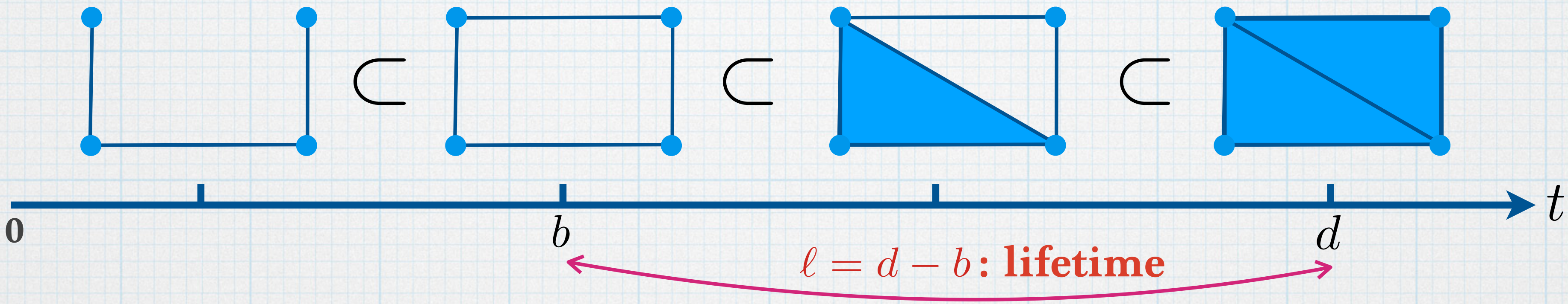
$$I[b,d] : 0 \to \cdots \to 0 \to \underset{\text{at } X_b}{K} \to \cdots \to \underset{\text{at } X_d}{K} \to 0 \to \cdots \to 0$$

Persistence Diagram

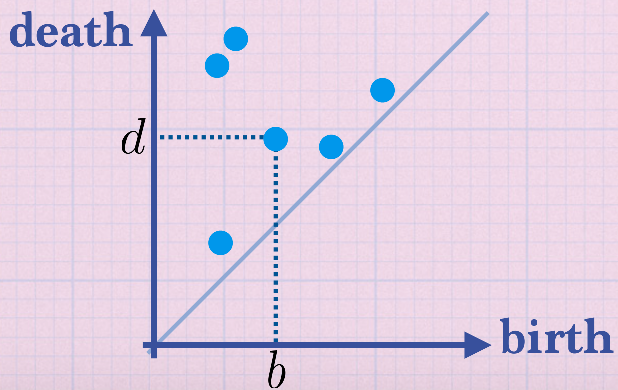
Interval decomp: $H_k(\mathcal{X}) \simeq \bigoplus_{i=1}^{p} I[b_i, d_i]$

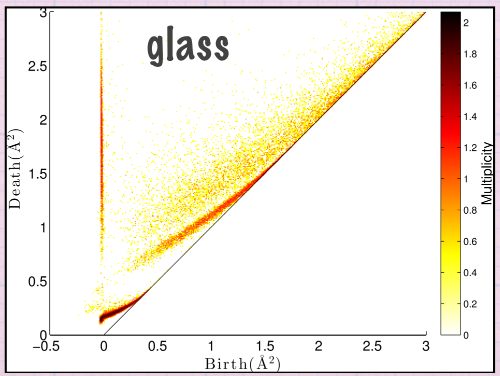
- $I[b, d]$ **represents appearance and disappearance of a topological feature at** $t = b, d$ **in** $\mathcal{X} = \{X(t)\}_t$

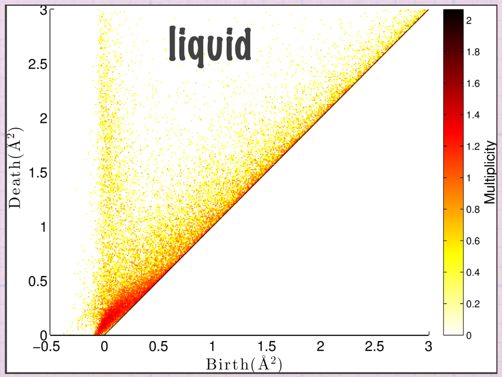
birth time **death time**



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

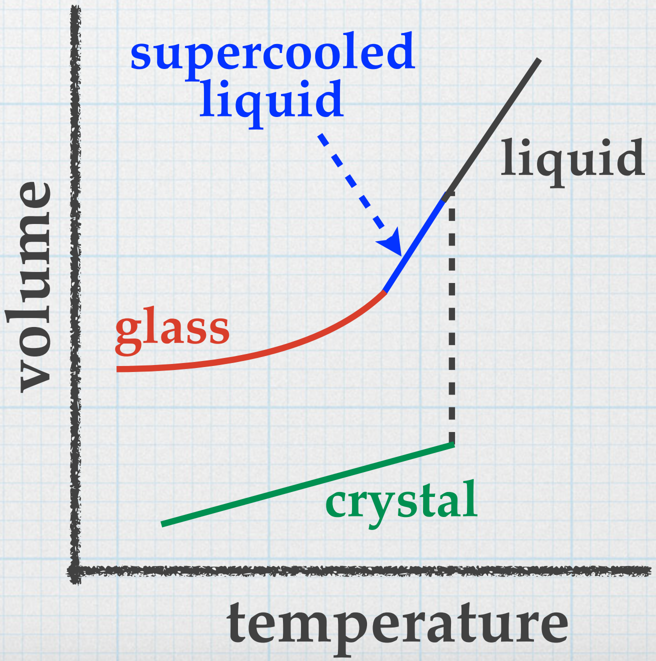




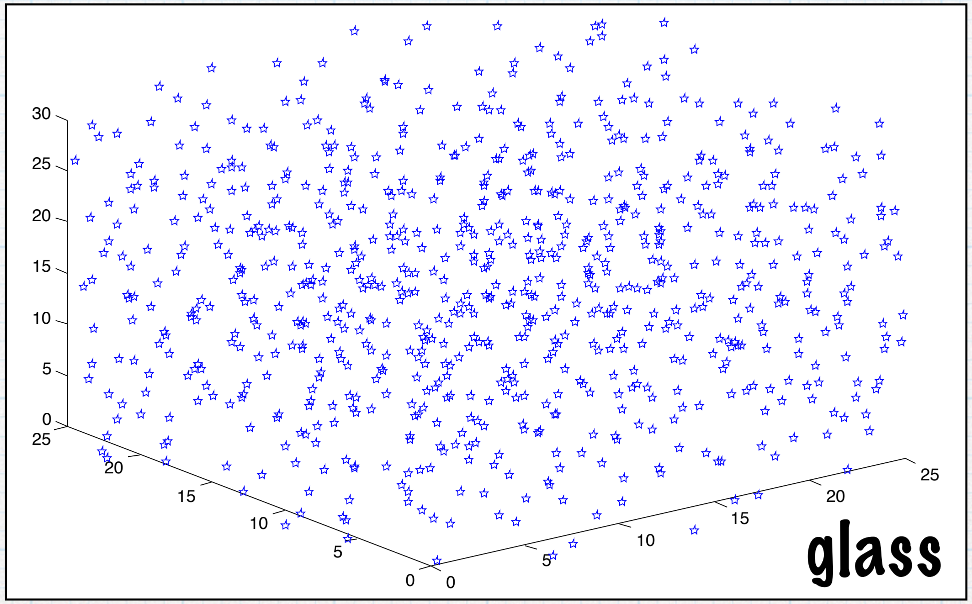


What is glass?

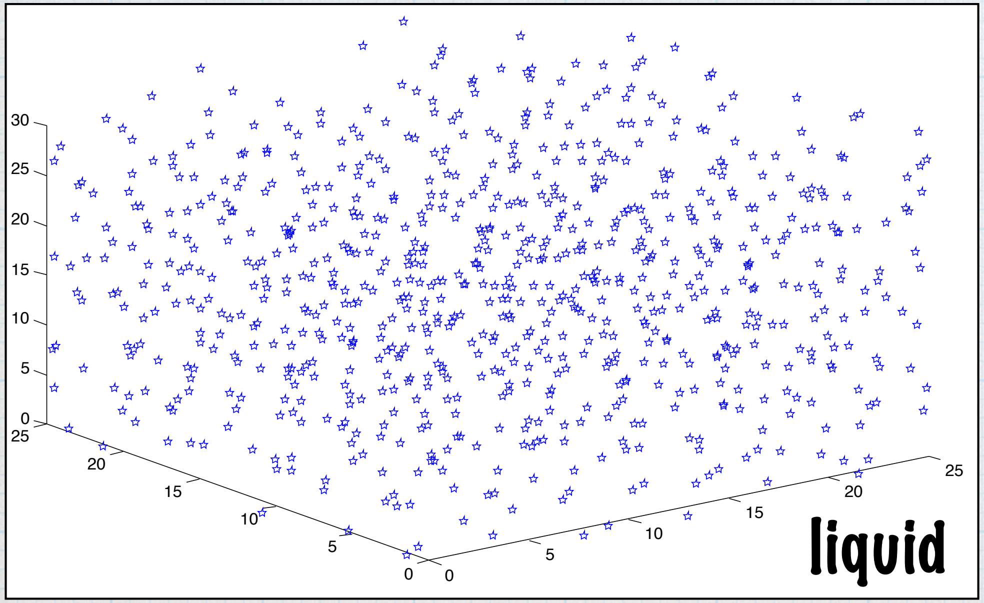
- **Not yet fully answered to "what is glass?"**
- **Not liquid, not solid, but something in-between**
- **Molecules have disordered configurations, but sufficient cohesion to maintain rigidity**
- **Further geometric understandings of atomic configurations are required**
- **Solar Energy Glass, DVD,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

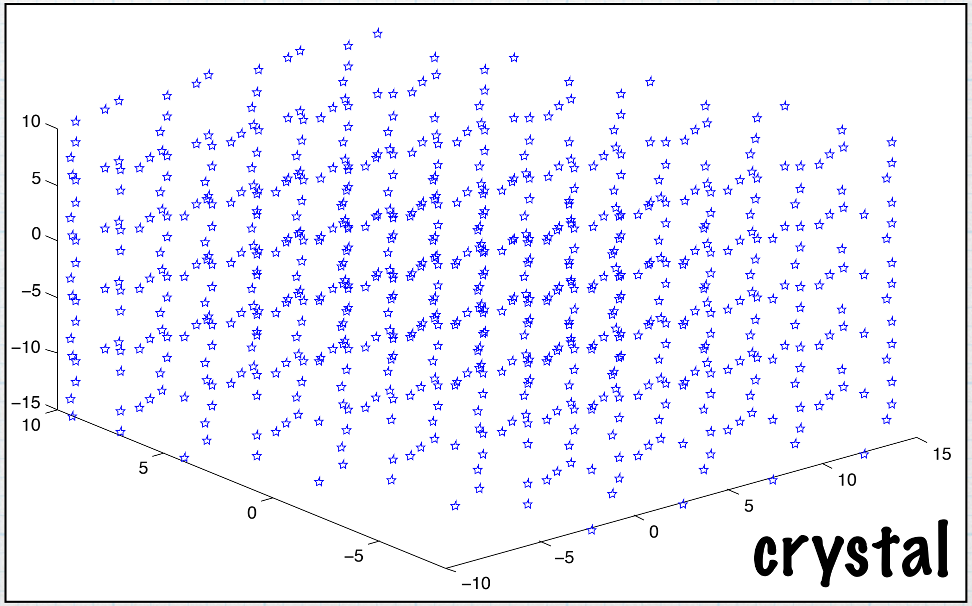


Atomic configurations of silica (SiO_2)



obtained by molecular dynamics (MD) simulation

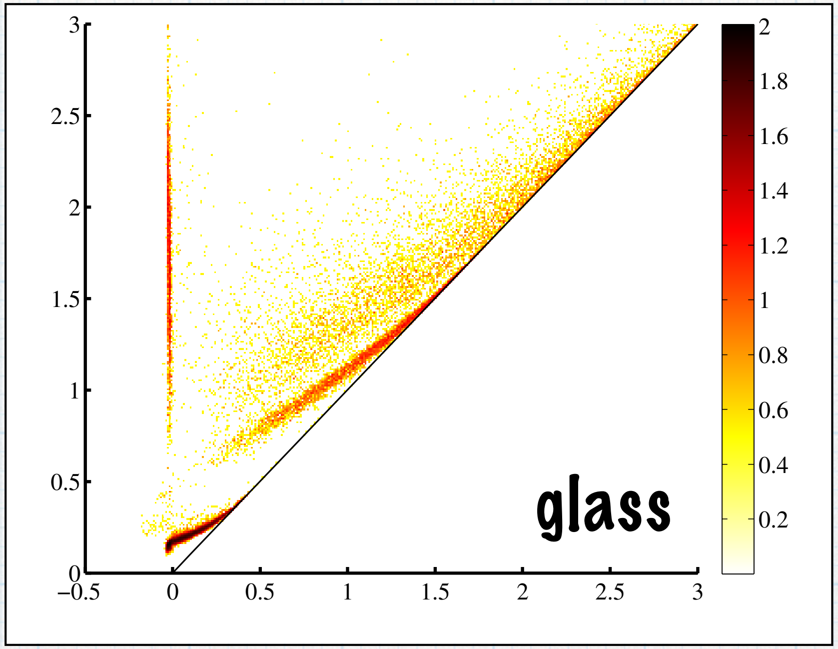




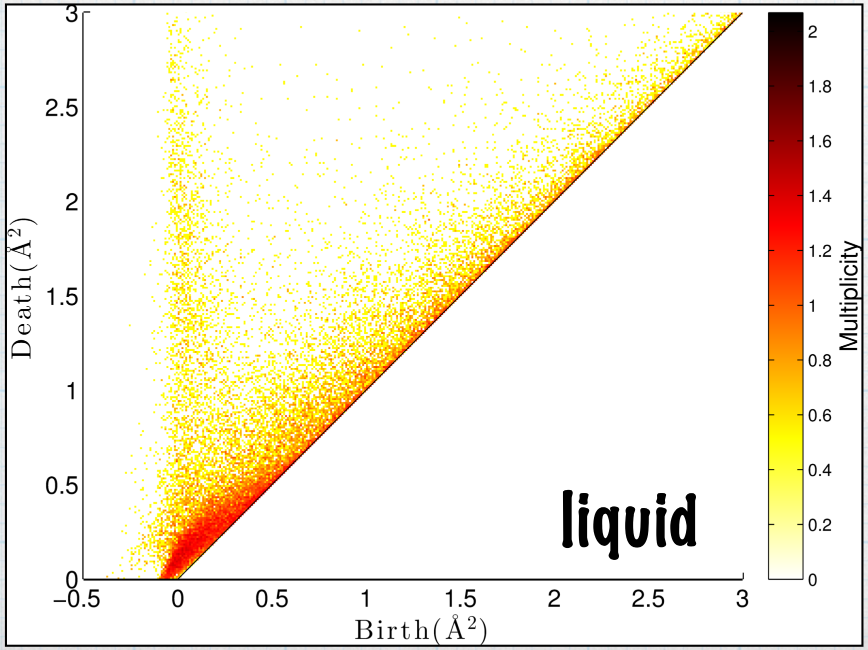
Nakamura, H. , Nishiura et al. Nanotechnology 26 (2015)

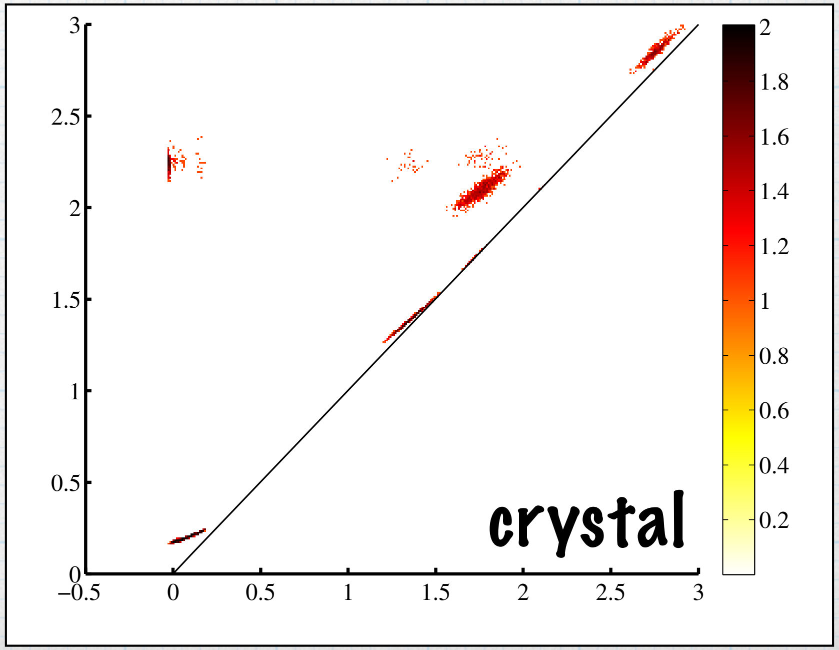
1st-Persistence diagrams of silica

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



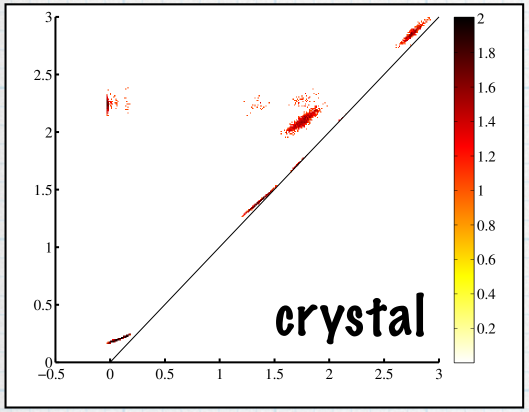
PD1 of weighted alpha filtrations

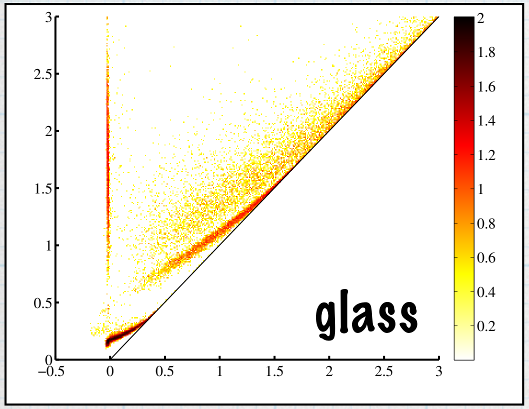


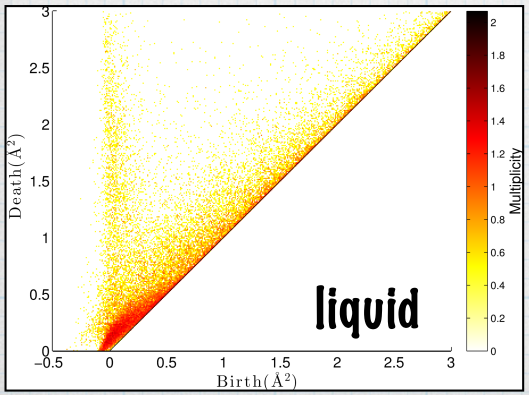


Nakamura, H., Nishiura et al. Nanotechnology 26 (2015)

Support dim and order parameter







0-dim support results from periodic atomic locations of crystals

transition

1-dim support (curves!) appears

transition

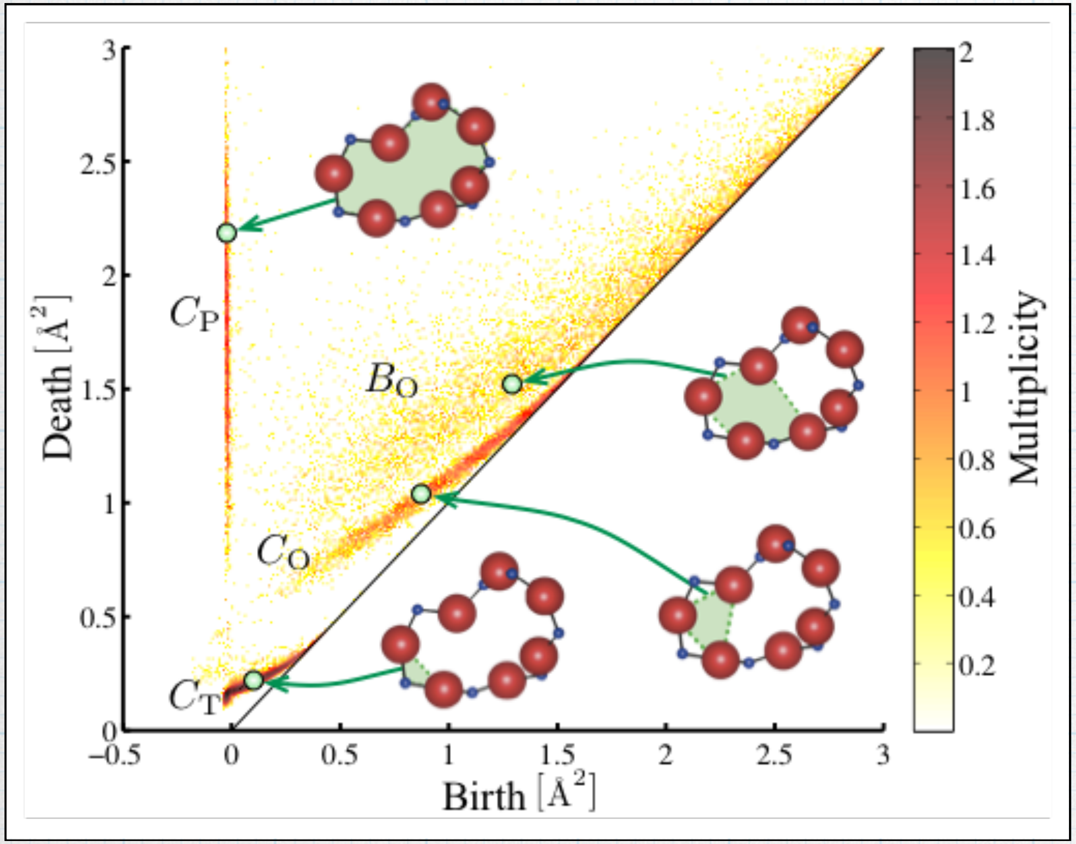
2-dim support results from random atomic locations of liquids

Remark

1. topological approach for glass transition ?
2. I'll talk about this later from statistical viewpoint

Nakamura, H. , Nishiura et al. Nanotechnology 26 (2015)

Geometric origins of curves

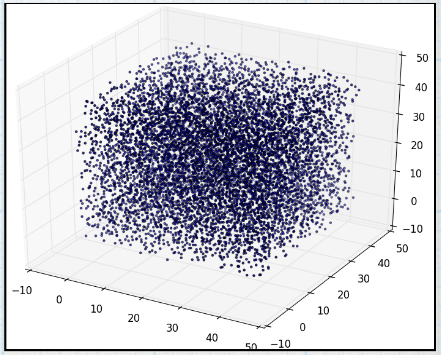


what characterize glass structures?

what are the geometric origins of the curves?

inverse problem!

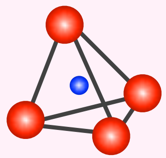
(e.g., find optimal cycles)



hierarchical ring structure

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

C_T: triangles on tetrahedra

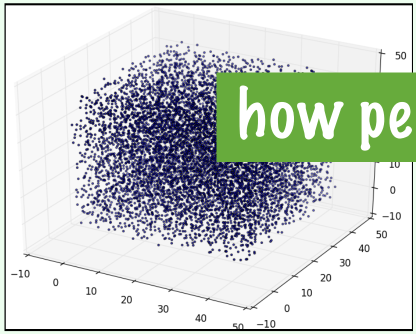


B_O: oxygen rings ($\geq$ four)

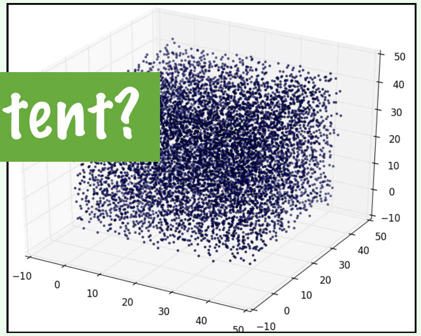
From materials science to math

representation theory

glass → compressed

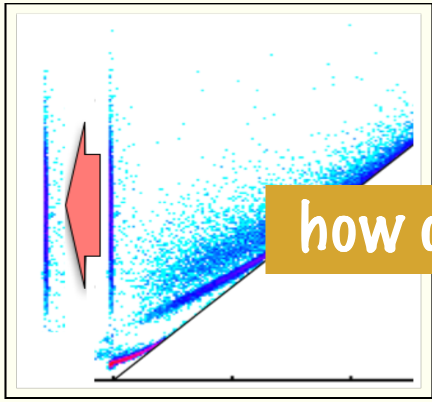


how persistent?

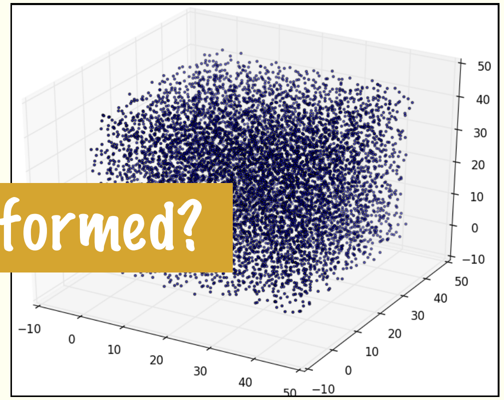


study persistence of both spatial and compression senses in pressurization

inverse problem

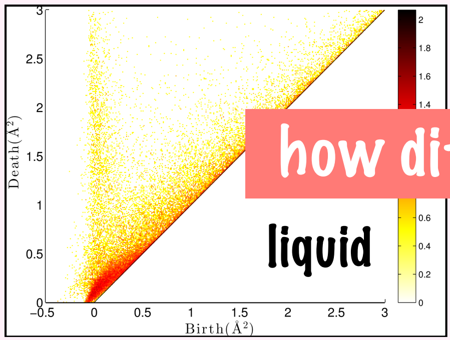


how deformed?

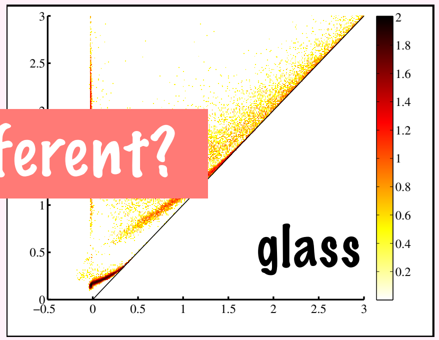


develop methods to study inverse problems (applications: design, rigidity, etc)

random topology

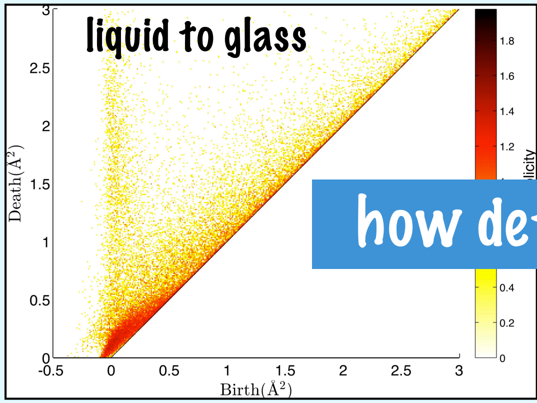


how different?

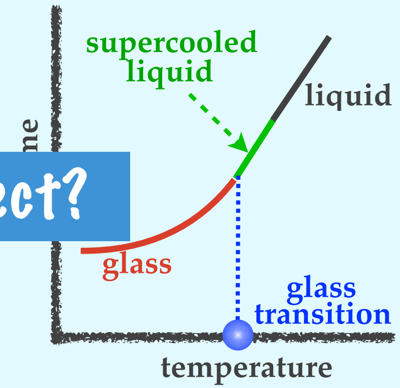


understand random topological features to characterize ordered structures

topological statistics



how detect?



detect glass transition and capture principal behaviors in statistical methods

Representation and Generalized PD

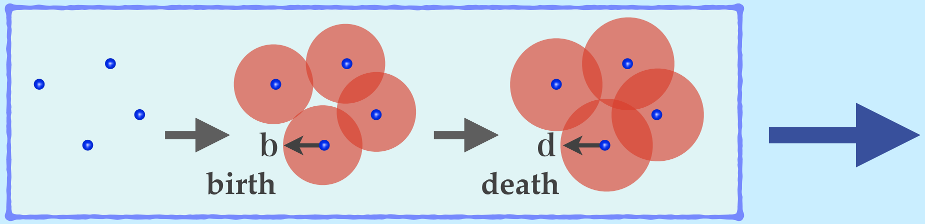
- filtration $\mathcal{X} : X_1 \subset \cdots \subset X_n$
- persistent homology $PH_\ell(\mathcal{X}) : H_\ell(X_1) \to \cdots \to H_\ell(X_n) \simeq \bigoplus I[b,d]$ ★

key property
interval decomposition

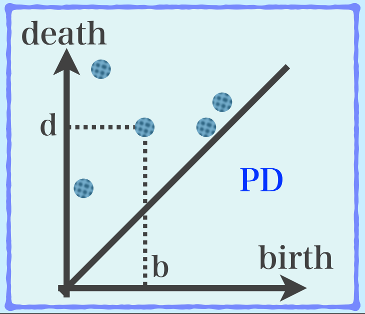
$$I[b,d] : 0 \to \cdots \to 0 \to \underset{\text{at } X_b}{K} \to \cdots \to \underset{\text{at } X_d}{K} \to 0 \to \cdots \to 0$$

K: field

- persistence diagram (PD) $D_\ell(\mathcal{X}) = \{(b,d) \in \mathbf{R}^2 \text{ appearing in} ★\}$



blow up balls



- each point in PD is a hole in data
- points close to diagonal are noisy
- points away from diagonal are robust

- A_n persistent module $H_\ell(X_1) \leftrightarrow H_\ell(X_2) \leftarrow \cdots \to H_\ell(X_{n-1}) \leftrightarrow H_\ell(X_n) \simeq \bigoplus I[b,d]$

representation on A_n quiver

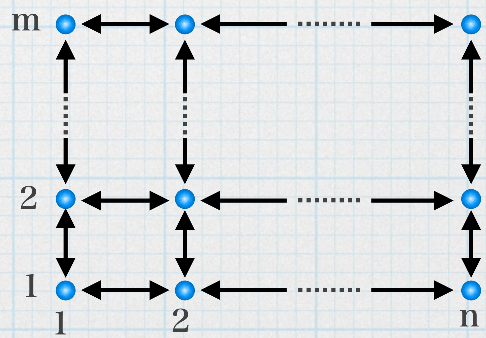
interval decomposition
is available

Applications: time series analysis, protein folding, etc

- $A_m \otimes A_n$ persistent module (Escolar & H. Discrete Comput. Geom.)

indecomposable decomposition are not trivial!

research topics: matrix methods
and bocs for $A_m \otimes A_n$
(Asashiba, Escolar, Takeuchi, H)



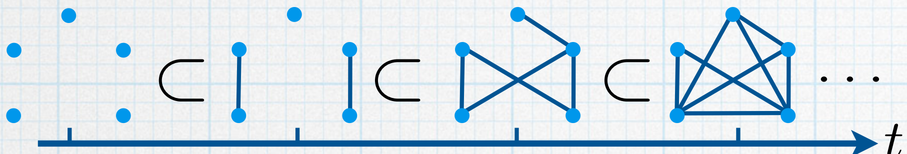
Applications: spatio-temporal TDA

see poster for details!

Random Topology

Characterize stochastic properties of data

Random simplicial complexes and its persistence diagrams

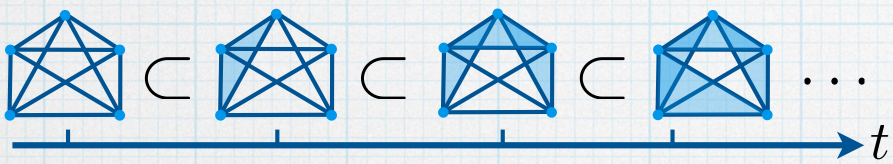


Erdös–Rényi random graph (process)

- percolation, Erdös–Rényi thm
- Frieze's $\zeta\,(3)$ theorem
- asymptotic behaviors, etc

higher dim. generalization

higher dim. generalization



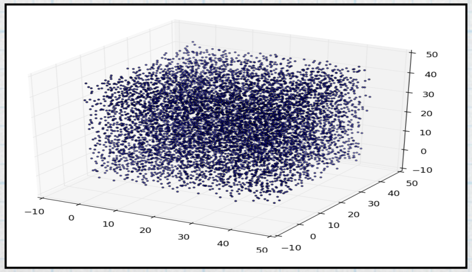
random simplicial complex (process)

- Generalized Frieze's $\zeta\,(3)$ theorem
  (H & Shirai. arXiv:1503.05669)

Research Topics

- limiting theorem of PD
- high dimentional Wilson algorithm
- random walk on simplicial complex
- birth-death Markov chain and generalized persistence

Random point process on $\mathbf{R}^n$ and its persistence diagrams



Research Topics

- PD for (quasi) periodic configuration and its perturbation
- characterizing PDs for softmatter
- Point process (Poisson, Cox, etc.) and their PDs

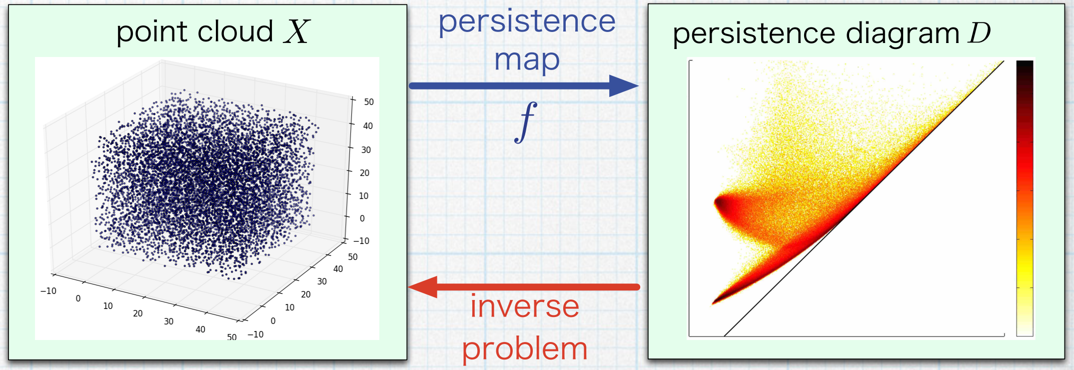
Persistent Inverse Problem

Continuation of point clouds via PDs

Gameiro, H, Obayashi. Physica D (2016)

- Given initial correspondence $D = f(X)$ and a target PD $D' \sim D$, solve

$$X' \text{ s.t. } D' = f(X')$$



- persistence map f is differentiable
- apply the continuation method in dynamical systems (with pseudo-inverse operator)
- existence and uniqueness, bifurcation, singularity
- Applications: material design, prediction of geometric structure

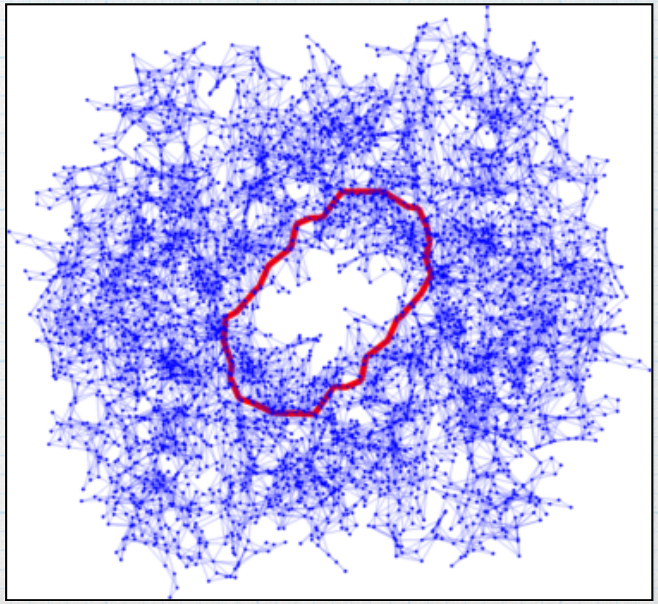
Compressive sensing for persistent homology

$$\begin{array}{ccccc} X & \longmapsto & C_*(X) & \longmapsto & D_*(X) \\ \text{point cloud} & & \text{chain cplx} & & \text{PD} \end{array}$$

- find minimum representative z_* in a homology class
- sparse optimization problem: $\dim C_*(X) \gg ||z_*||_0$

$\Rightarrow$ compressive sensing for z_*

- possible to extract detailed geometric features



Topological Statistics

Statistical reliability of PDs

background

Lack of sample data for statistics

(ex: soft-matter experiments in limited situation)

Reliability of PDs?

Research topics

bootstrap method for PDs using $A_2 \otimes A_n$

$$
\begin{tikzcd}
H_*(Y_1) \arrow[r] & H_*(Y_2) \arrow[r] & H_*(Y_3) \arrow[r] & \dots \arrow[r] & H_*(Y_n) \\
H_*(X_1) \arrow[r] \arrow[u] & H_*(X_2) \arrow[r] \arrow[u] & H_*(X_3) \arrow[r] \arrow[u] & \dots \arrow[r] & H_*(X_n) \arrow[u]
\end{tikzcd}
$$

original PH

resampling PH

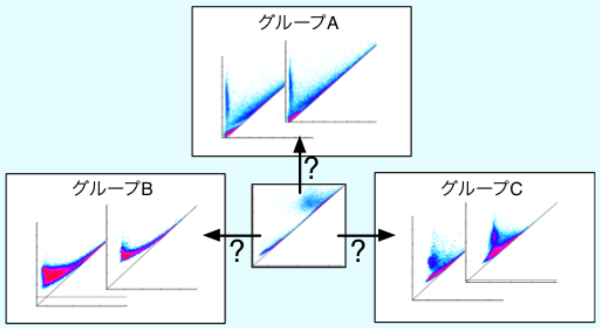
indecomposable decom. provides reliability of PD!

Kernel method: Statistics for PDs

background

Statistical methods for PDs

(SVM, PCA, change point detects, regression, etc)



classification???

research topics

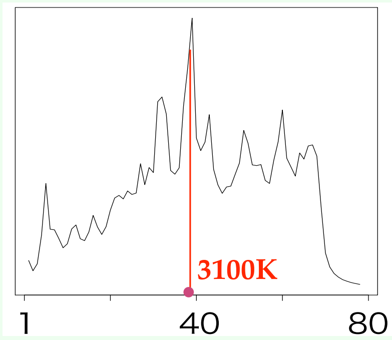
Develop kernel methods for PDs

(Kusano, Fukumizu, H)

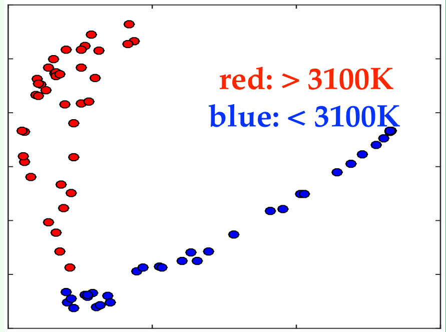
- kernel method for PDs: persistence weighted Gaussian kernel (PWGK)
- stability theorem
- vectorization of PD

see poster for details!

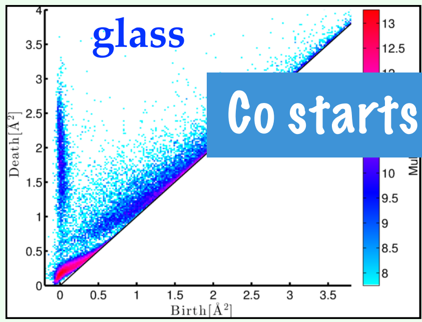
Application

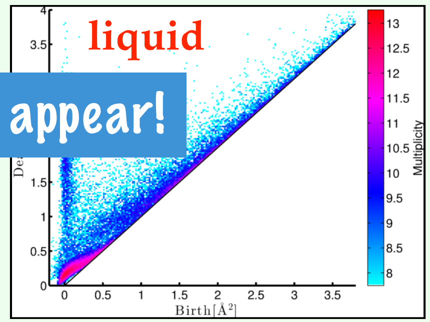


kernel Fisher discriminant



kernel PCA



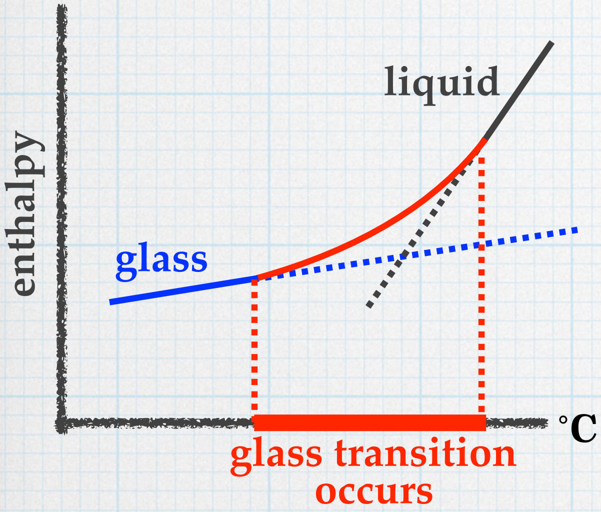


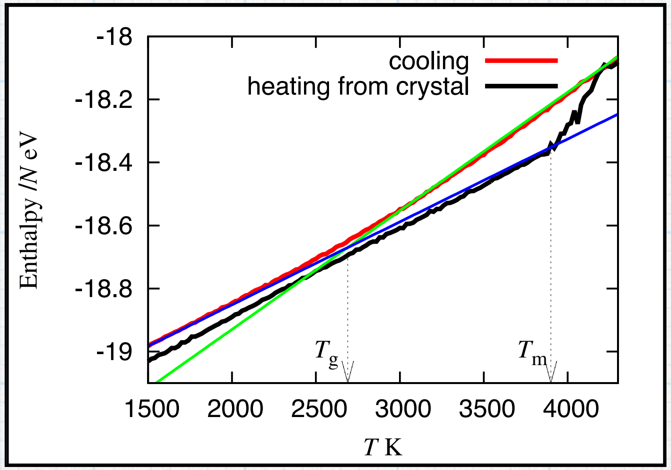
Co starts to appear!

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

Glass transition temperature

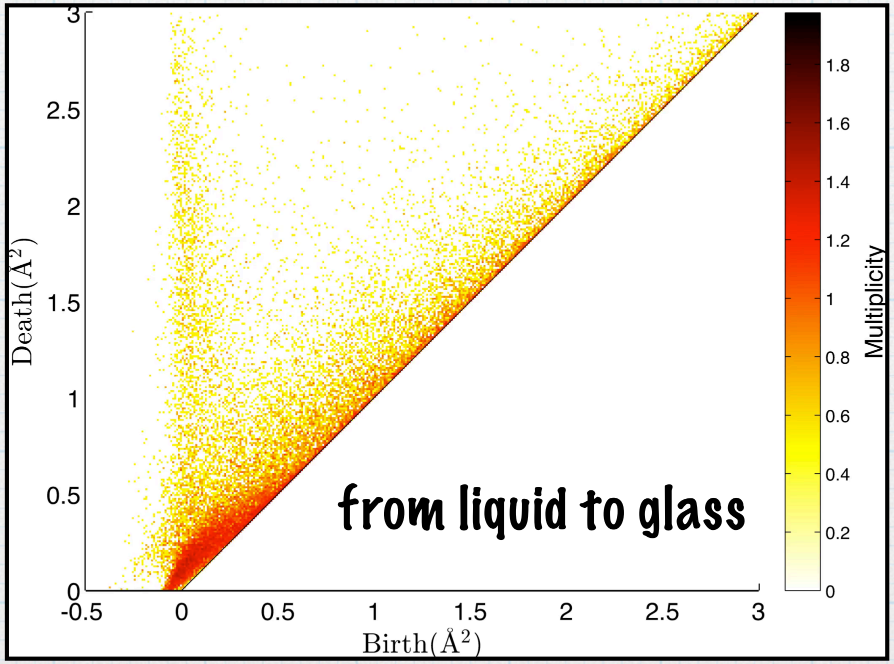
Convention to determine GTT in physics:





Since two slopes are close, GTT interval becomes large.

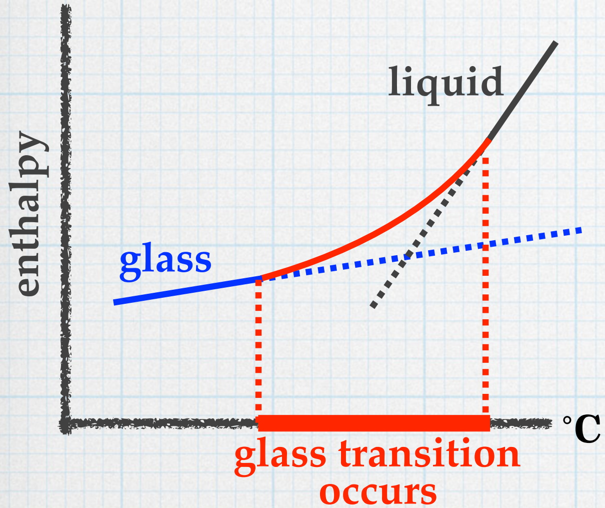
kernel change point detection and kernel PCA based on PWGK

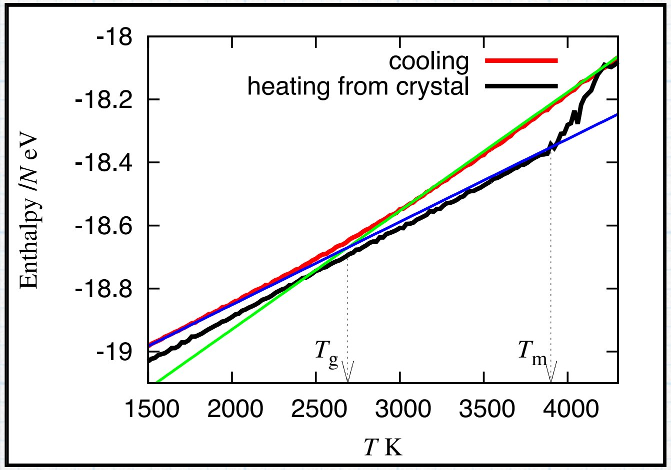


G. Kusano, K. Fukumizu, Y.H. ICML (2016)

Glass transition temperature

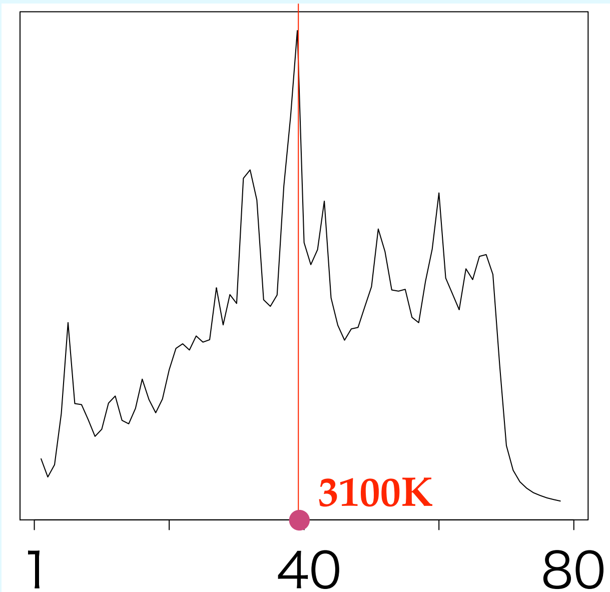
Convention to determine GTT in physics:



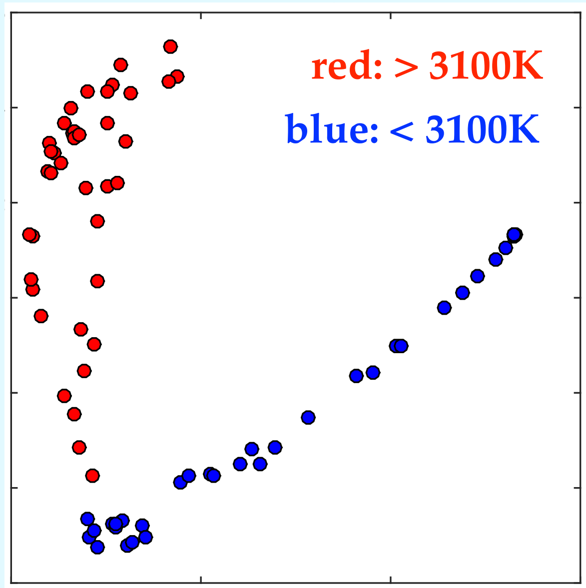


Since two slopes are close, GTT interval becomes large.

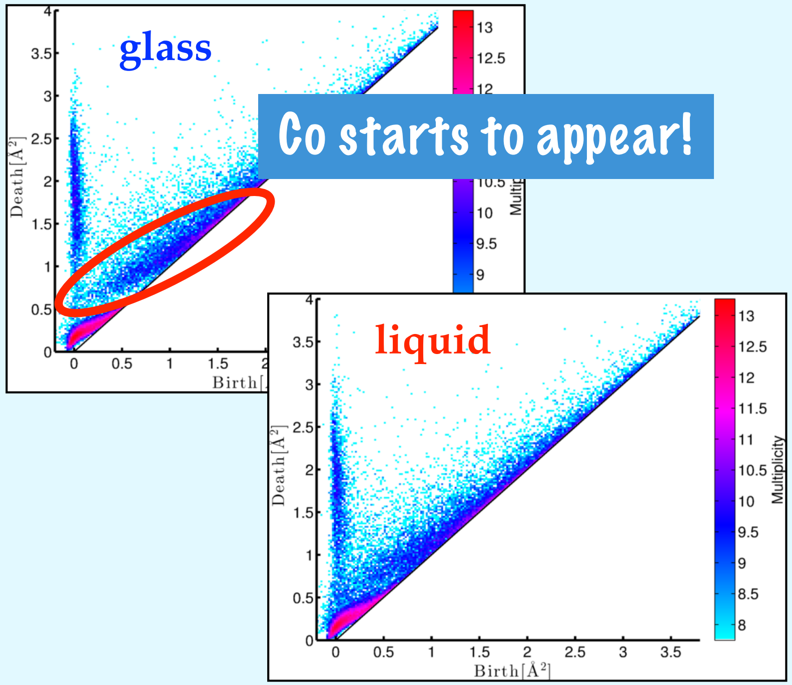
kernel change point detection and kernel PCA based on PWGK



kernel Fisher discriminant



kernel PCA



Collaborators

random topology

Tomoyuki Shirai

topological statistics

Kenji Fukumizu
Genki Kusano

representation

Hideto Asashiba

inverse problem

Ippei Obayashi

materials TDA

Hirata, Nakamura, Matsue, Saadatfar, Ichinomiya, Nakajima
Kotani, Nishiura, Kramar, Kohara, etc