

# Advanced Data Visualization

CS 6965

Fall 2019

Prof. Bei Wang Phillips

University of Utah



Lecture 25

# Structural Inference of High-dim Data

TOPO

# Topology Tool Kit

## TTK

<https://topology-tool-kit.github.io/>



# Installation Demo

TTK installation tutorial

```
ttk@ttk-VirtualBox: ~/code/ttk-0.9/build
-- Installing: /usr/local/lib/paraview-5.2/plugins/libUncertainDataEstimator.so
-- Set runtime path of "/usr/local/lib/paraview-5.2/plugins/libUncertainDataEstimator.so" to "/usr/local/bin/.."
ttk@ttk-VirtualBox:~/code/ttk-0.9/build$ ls /usr/local/bin/*Cnd
/usr/local/bin/blankCnd
/usr/local/bin/continuousScatterplotCnd
/usr/local/bin/contourForestsCnd
/usr/local/bin/geometrySmootherCnd
/usr/local/bin/jacobiSetCnd
/usr/local/bin/landmarkCnd
```

TTK v0.9 (c) 2017

Getting Started Guide Example Visualizations

9:44 / 10:40

CC HD YouTube

<https://youtu.be/etAe13KEWsk>







# Morse Persistence Demo (MSC)

TTK usage tutorial - Morse persistence demo

MORE VIDEOS

6:41 / 16:03

CC HD YouTube



# A few tips

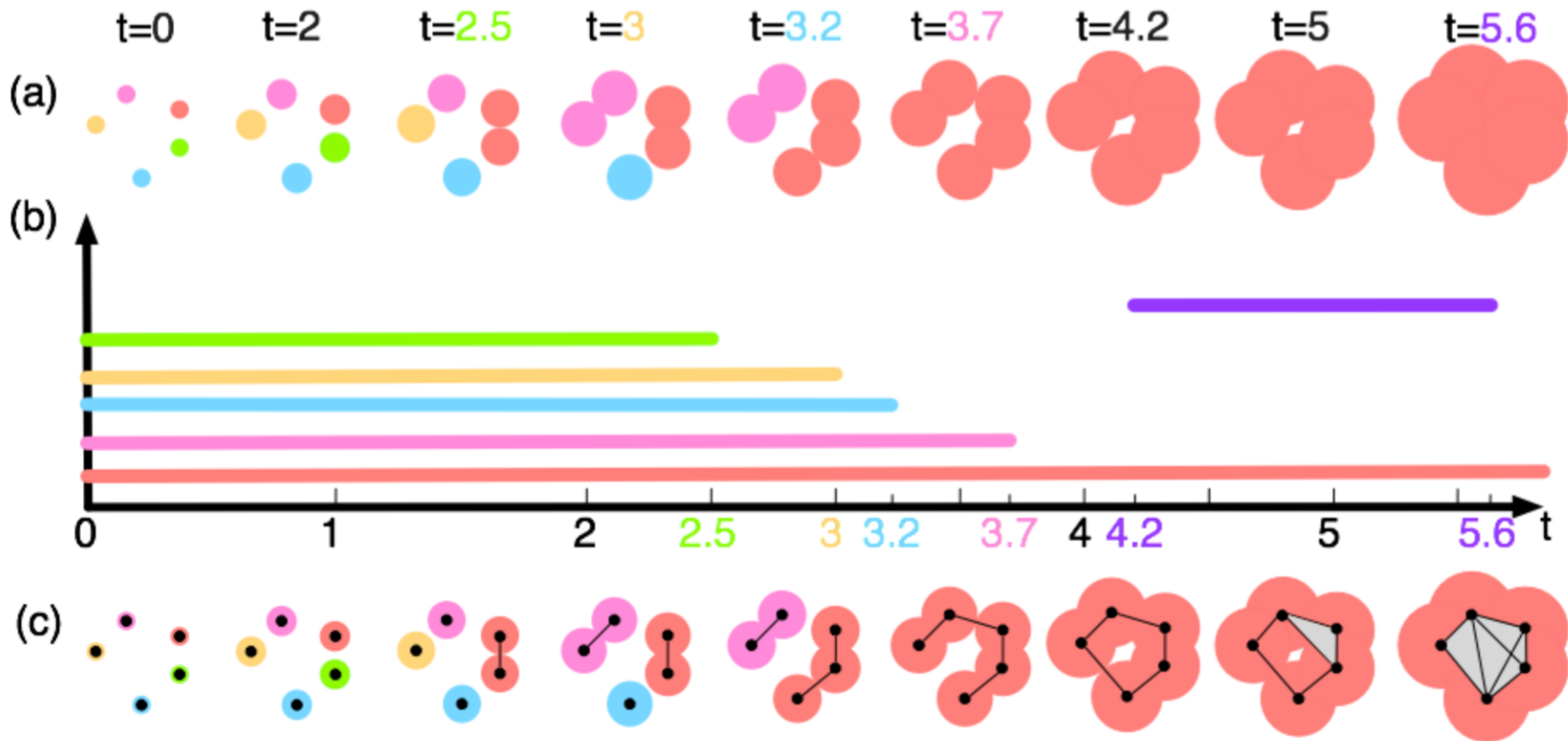
- The installation is going to take a while (4+ hours).
- Follow the demo closely, however pay attention to some differences in different versions of TTK

# Structural Inference of High-dimensional Data

A topological view



# Review: Persistent Homology





# Stability of Persistence Diagrams

Homological features encoded as barcodes or persistent diagrams

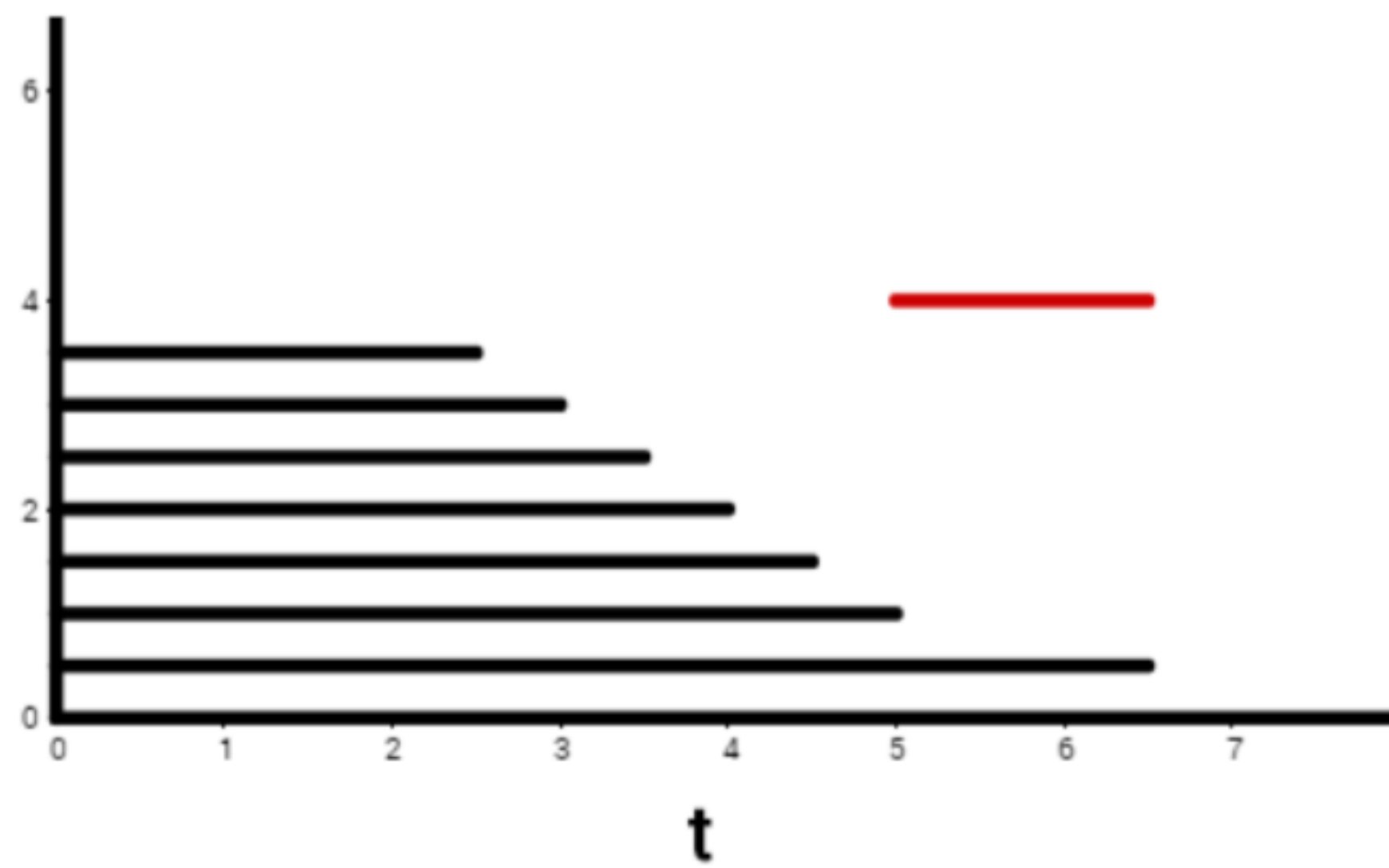


Figure: Barcode

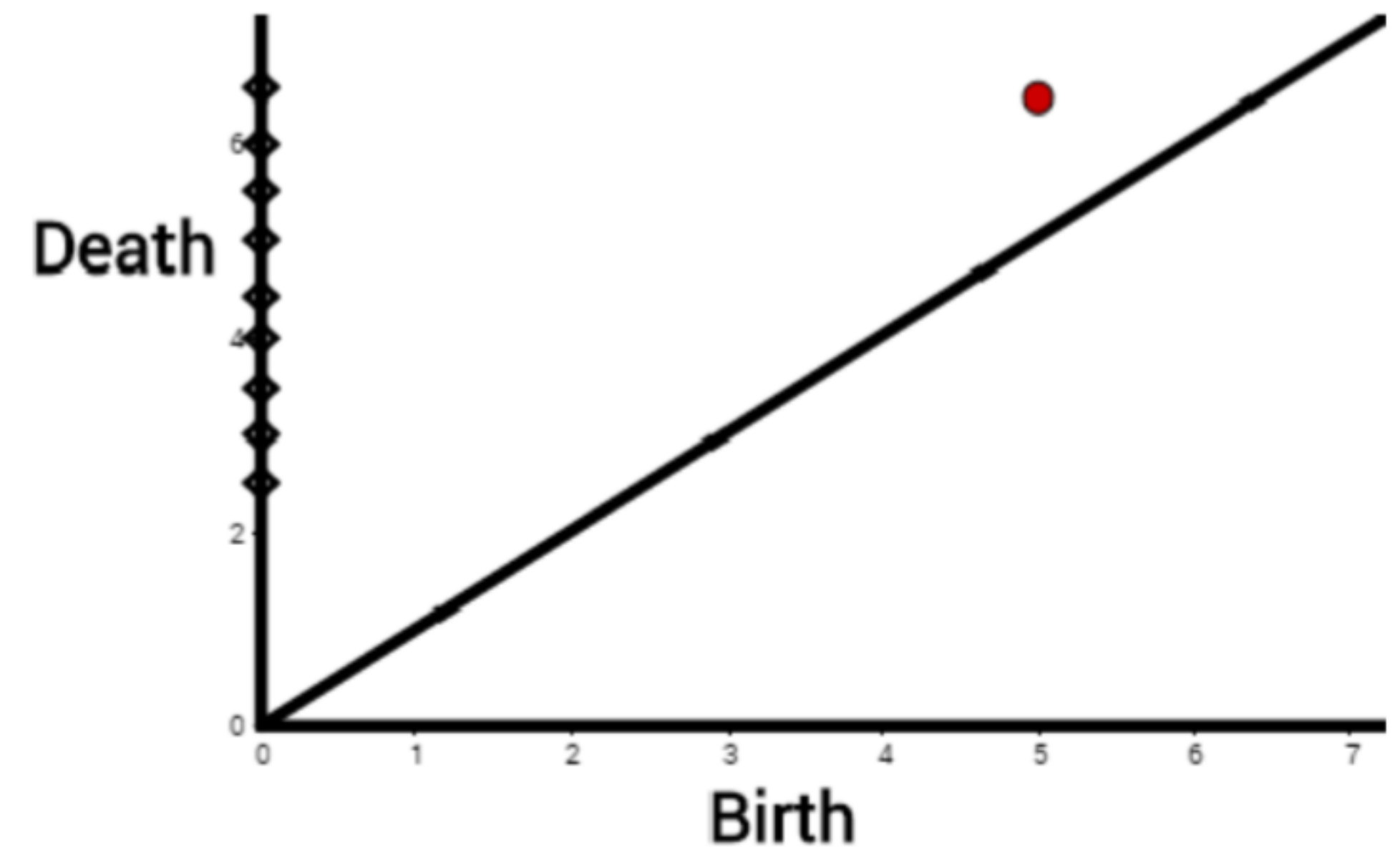
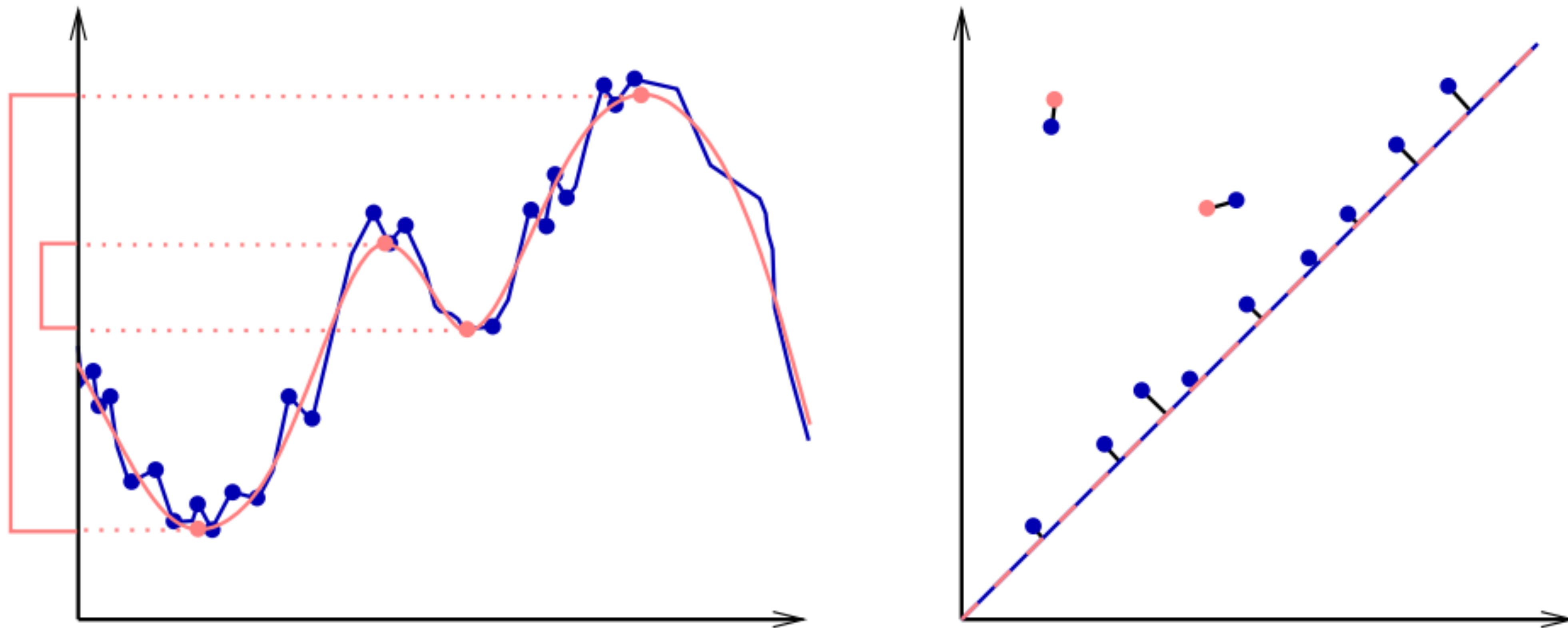


Figure: Persistence Diagram





**Figure 2: Left: two close functions, one with many and the other with just four critical values. Right: the persistence diagrams of the two functions, and the bijection between them.**

# Bottleneck Distance

We need some definitions. For points  $p = (p_1, p_2)$  and  $q = (q_1, q_2)$  in  $\mathbb{R}^2$ , let  $\|p - q\|_\infty$  be the maximum of  $|p_1 - q_1|$  and  $|p_2 - q_2|$ . Similarly for functions  $f$  and  $g$ , let  $\|f - g\|_\infty = \sup_x |f(x) - g(x)|$ . Let  $X$  and  $Y$  be multisets of points.

DEFINITION. The *Hausdorff distance* and the *bottleneck distance* between  $X$  and  $Y$  are

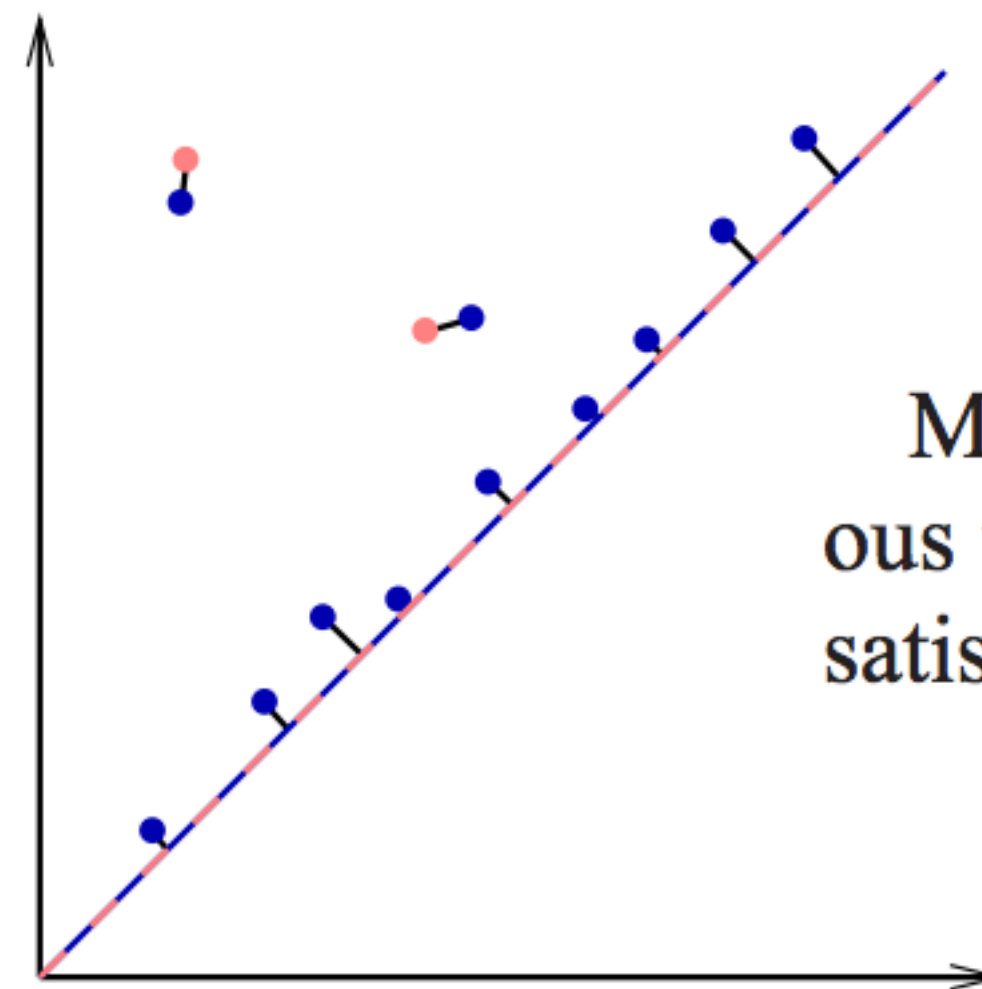
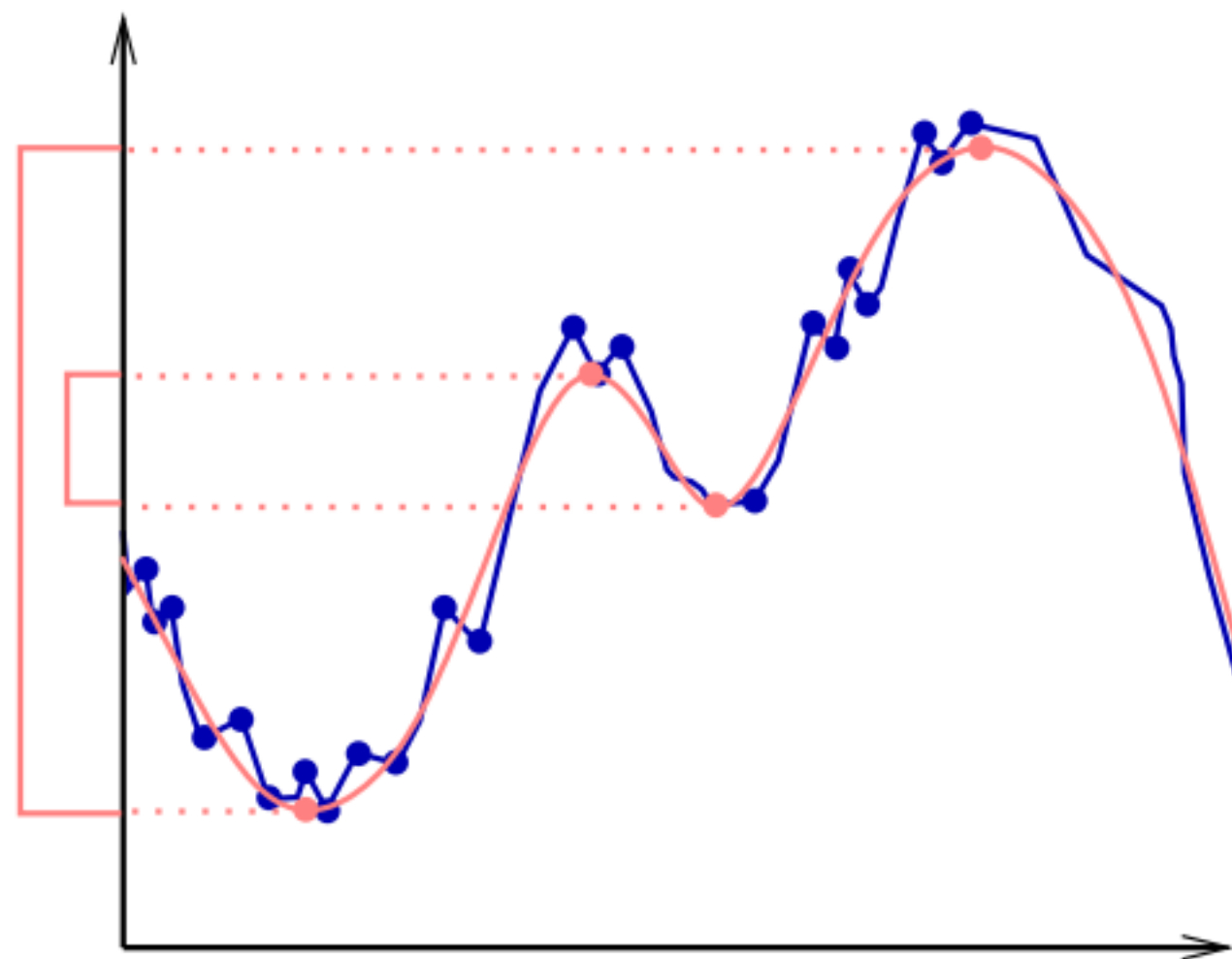
$$d_H(X, Y) = \max\left\{\sup_x \inf_y \|x - y\|_\infty, \sup_y \inf_x \|y - x\|_\infty\right\}$$

$$d_B(X, Y) = \inf_\gamma \sup_x \|x - \gamma(x)\|_\infty,$$

where  $x \in X$  and  $y \in Y$  range over all points and  $\gamma$  ranges over all bijections from  $X$  to  $Y$ . Here we interpret each point with multiplicity  $k$  as  $k$  individual points and the bijection is between the resulting sets.

# Bottleneck Distance Stability

$$d_B(X, Y) = \inf_{\gamma} \sup_x \|x - \gamma(x)\|_{\infty}$$

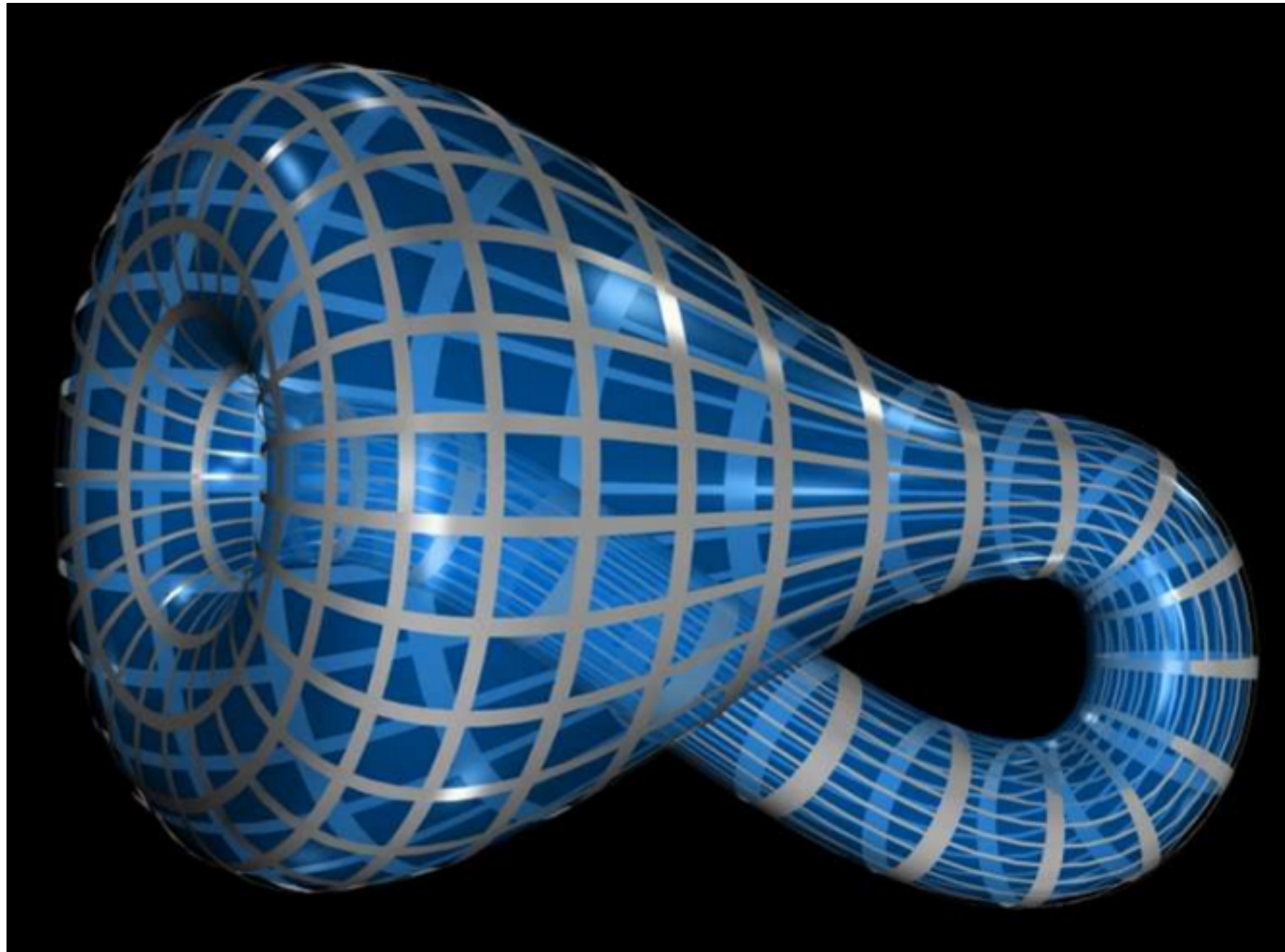


MAIN THEOREM. Let  $\mathbb{X}$  be a triangulable space with continuous tame functions  $f, g : \mathbb{X} \rightarrow \mathbb{R}$ . Then the persistence diagrams satisfy  $d_B(D(f), D(g)) \leq \|f - g\|_{\infty}$ .

$$d_B(D(f), D(g)) \leq \|f - g\|_{\infty}$$

Cohen-SteinerEdelsbrunnerHarer2005





# Space of Natural Images

[CarlssonIshkhanovSilva2008]

<http://webee.technion.ac.il/Sites/People/adler/Boris-topology.pdf>

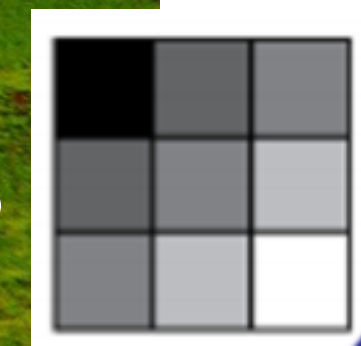








Taking 3 by 3 patches





# Highlight

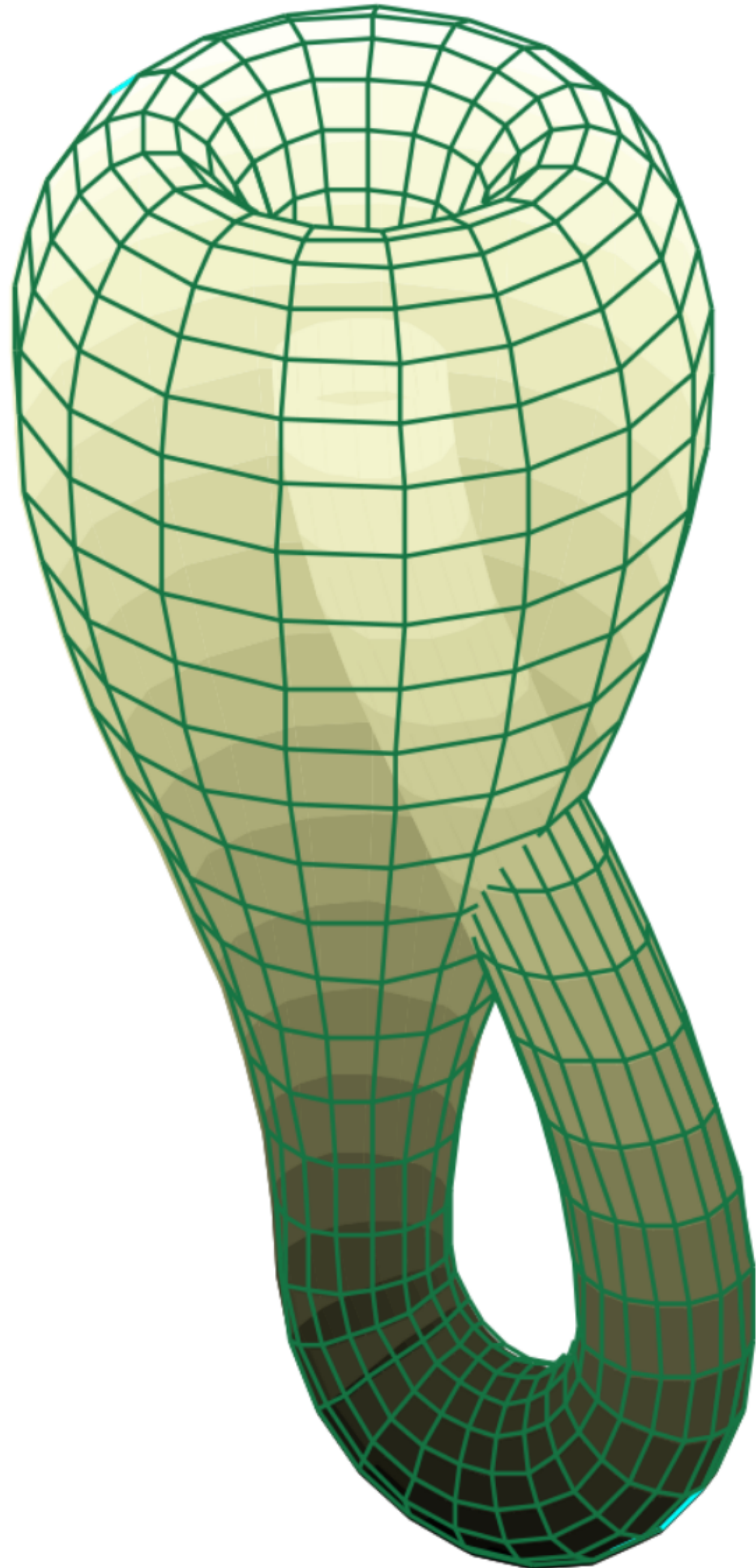
In this paper we continued the qualitative (topological) approach to the study of the space of 3 by 3 patches coming from natural images initiated in [8]. Perhaps the key advantage of this approach is that it allows one to find highly non-linear yet extremely important subsets within the data which otherwise would be very hard to discover using more common statistical techniques.

**a collection of  $4 \cdot 10^6$  3 by 3 patches**



# High-level techniques

- Persistent homology
- The space of patches: preprocessing
- Klein bottle

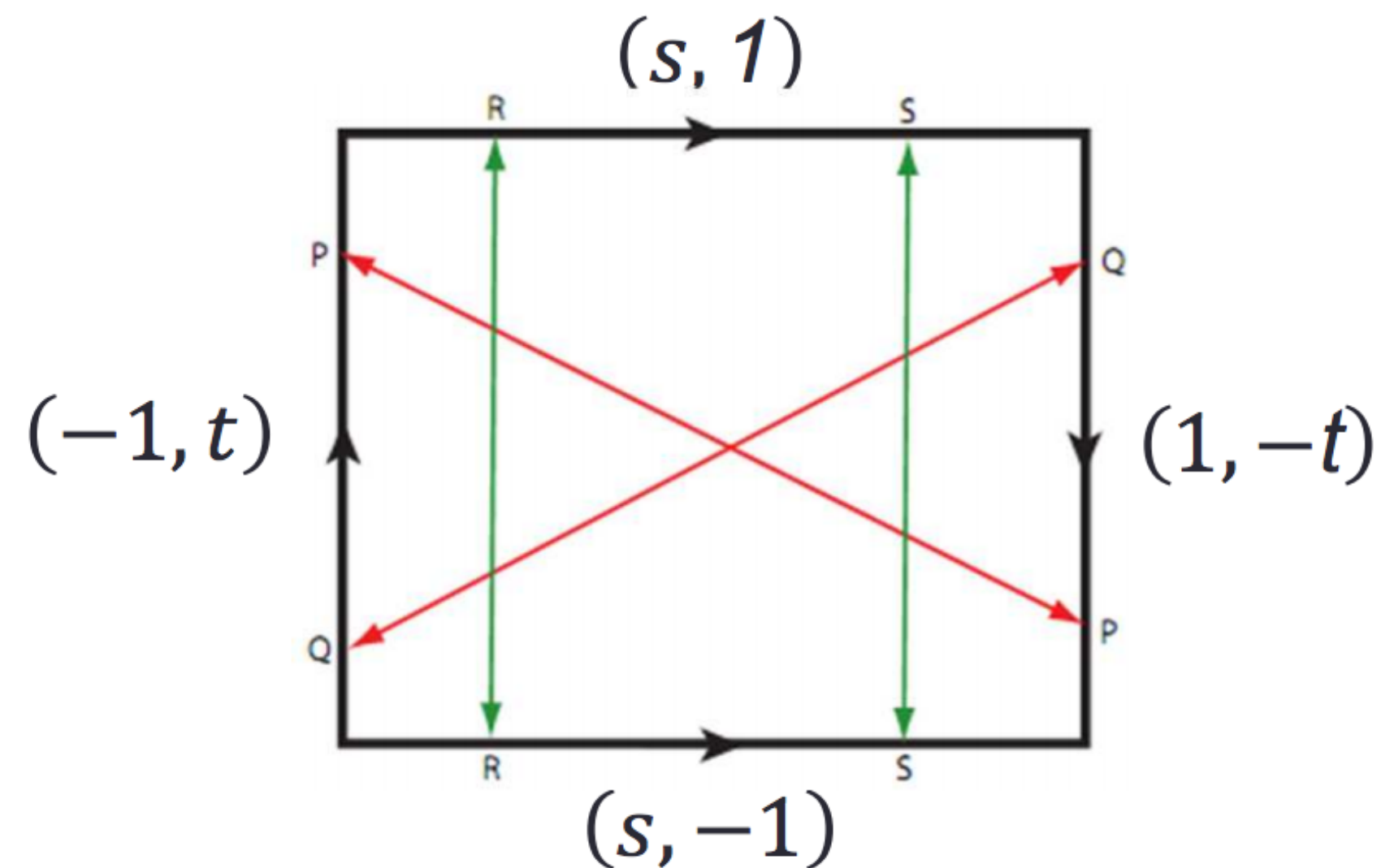


# Klein Bottle

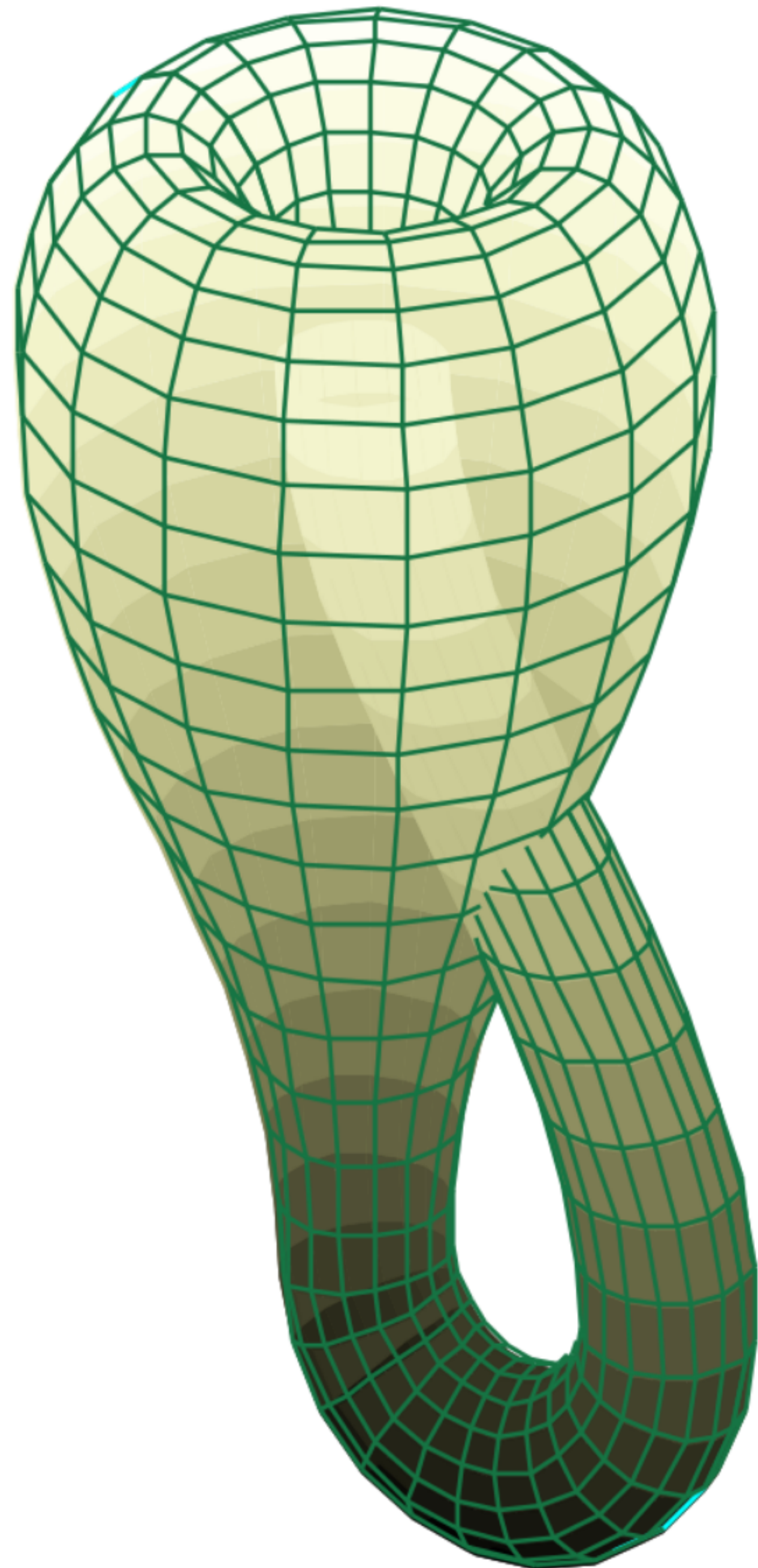
- Created by identifying points of a unit square:

$$(-1, t) \Leftrightarrow (1, -t)$$

$$(s, -1) \Leftrightarrow (s, 1)$$







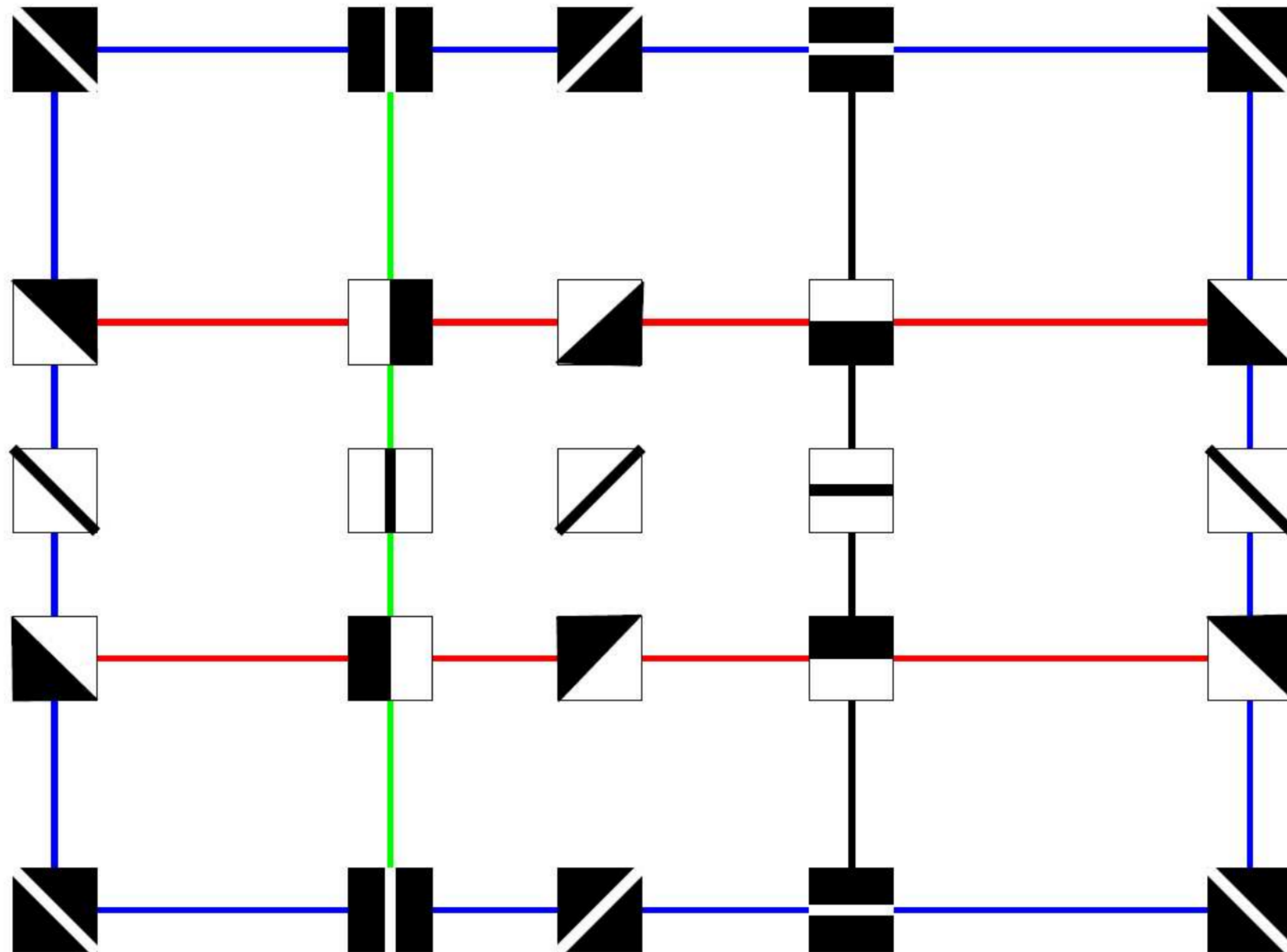
# Klein Bottle



[https://en.wikipedia.org/wiki/Klein\\_bottle](https://en.wikipedia.org/wiki/Klein_bottle)

<https://www.youtube.com/watch?v=AAAsICMPwGPY>





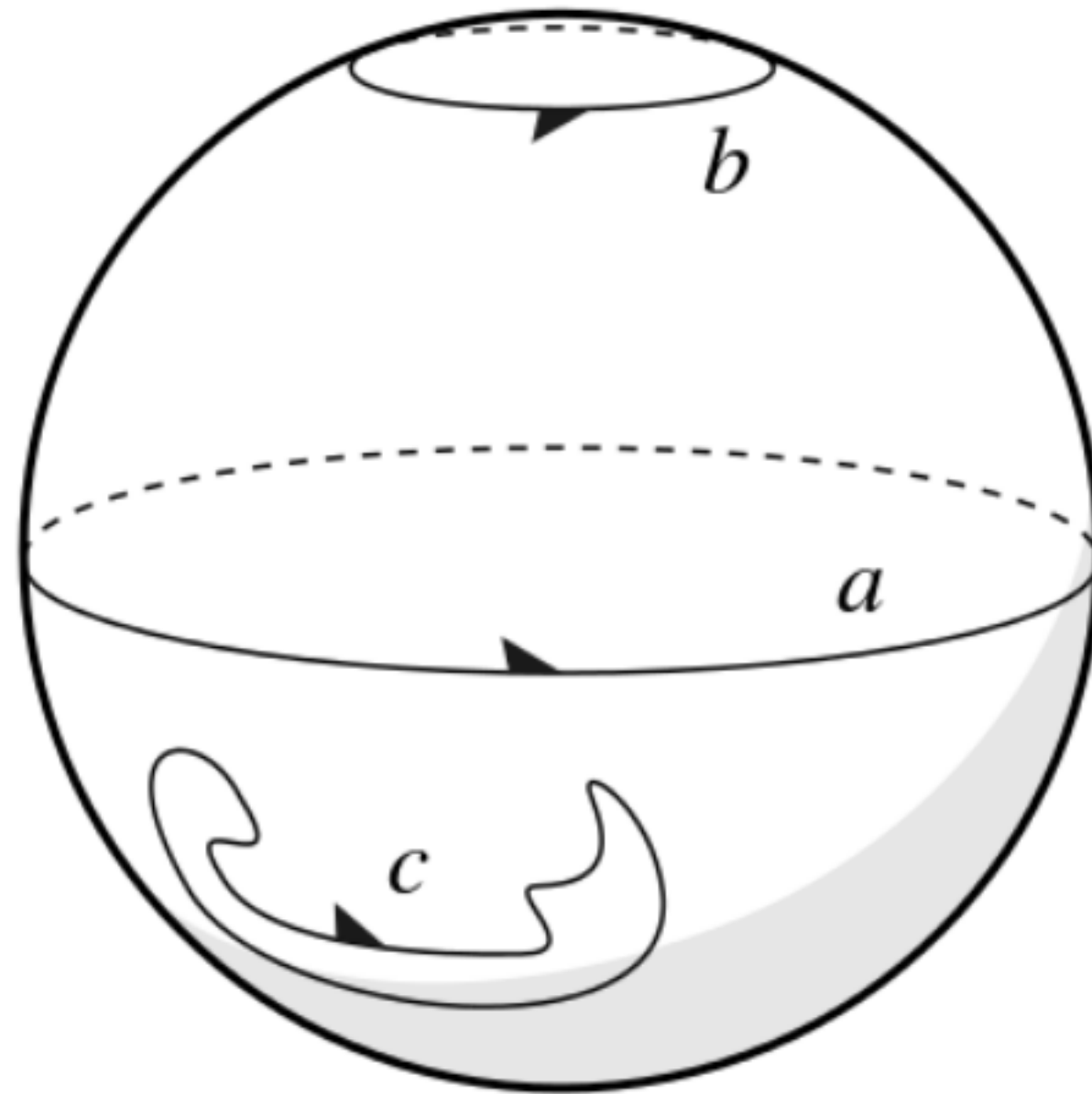
3 by 3 patches parametrized by the Klein bottle



# How do we infer the Klein bottle?

- Via persistent homology...
- The (persistent) homology of the space of (many) images matches that of a Klein bottle under  $\mathbb{Z}_2$  homology.

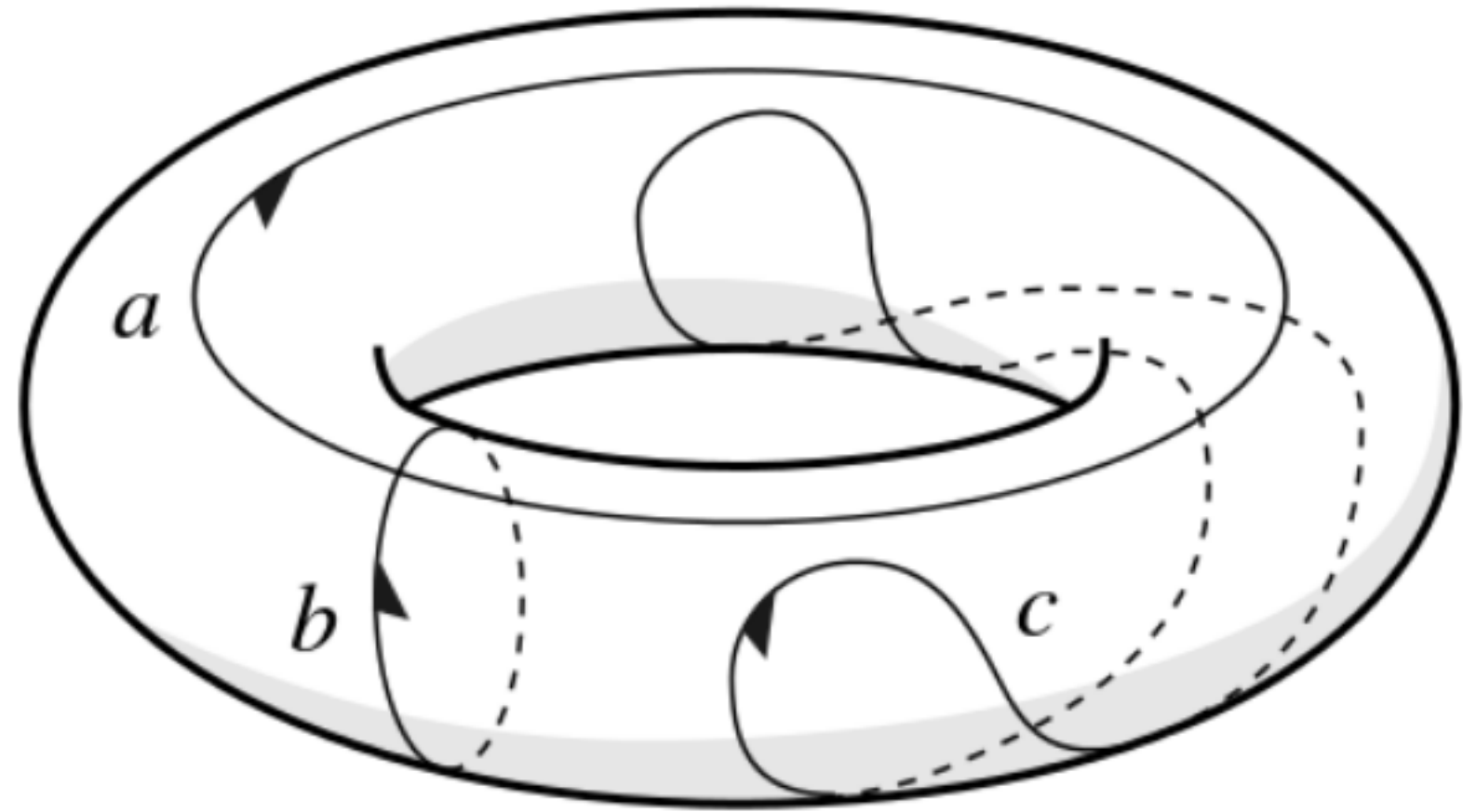
# Back to homology



- Any cycle on the sphere can shrink to a point

$b_0$ ◆	$b_1$ ▼	$b_2$ ◆
1	0	1

# Back to homology

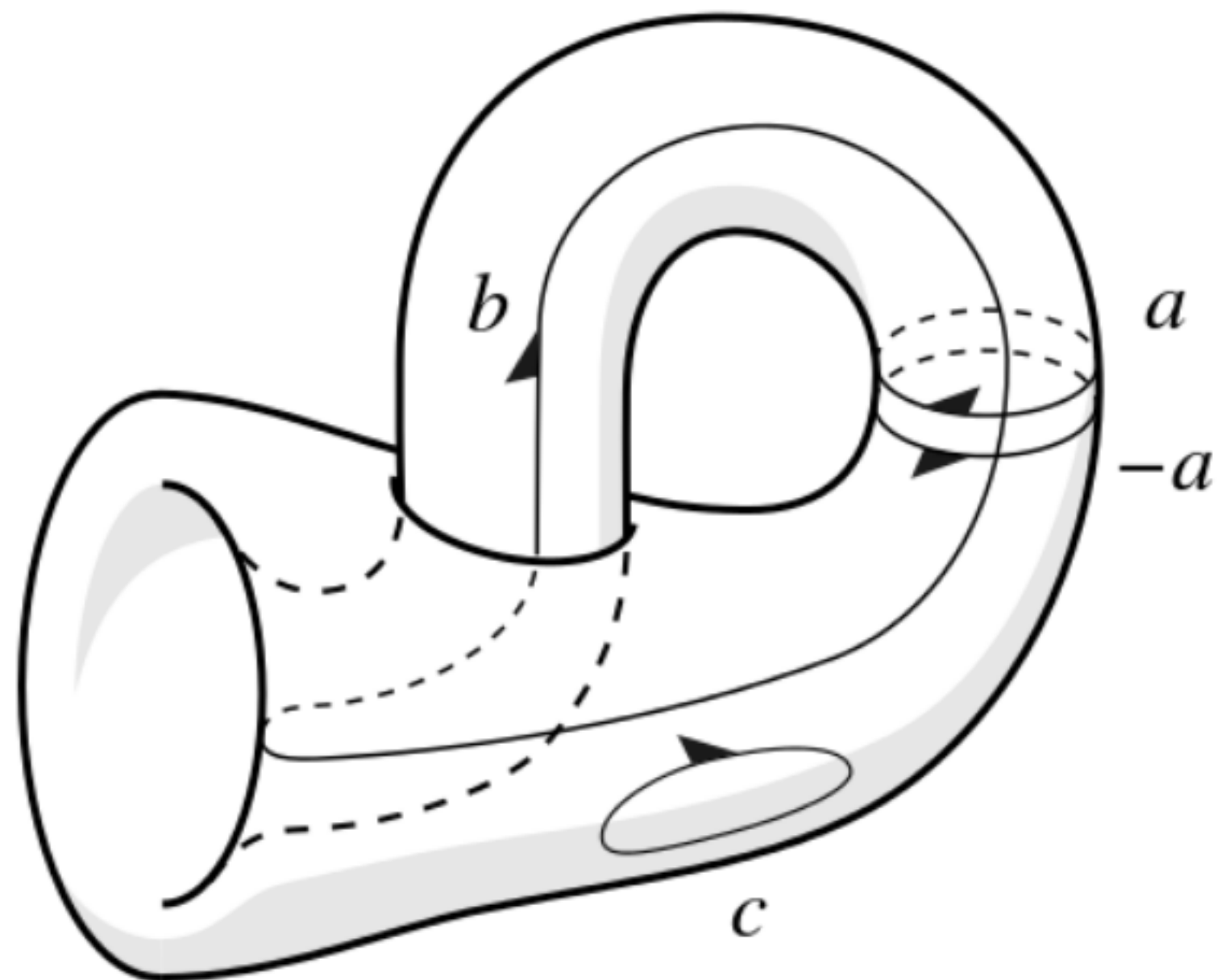


- Not all cycles on the torus can shrink to a point

$b_0$	◆	$b_1$	▼	$b_2$	◆
1		2		1	



# Back to homology



- Not all cycles on the Klein bottle can shrink to a point

$b_0$	◆	$b_1$	▼	$b_2$	◆
1		1		0	

- Under  $\mathbb{Z}_2$  homology, the rank is identical to that of a torus

# Persistent homology

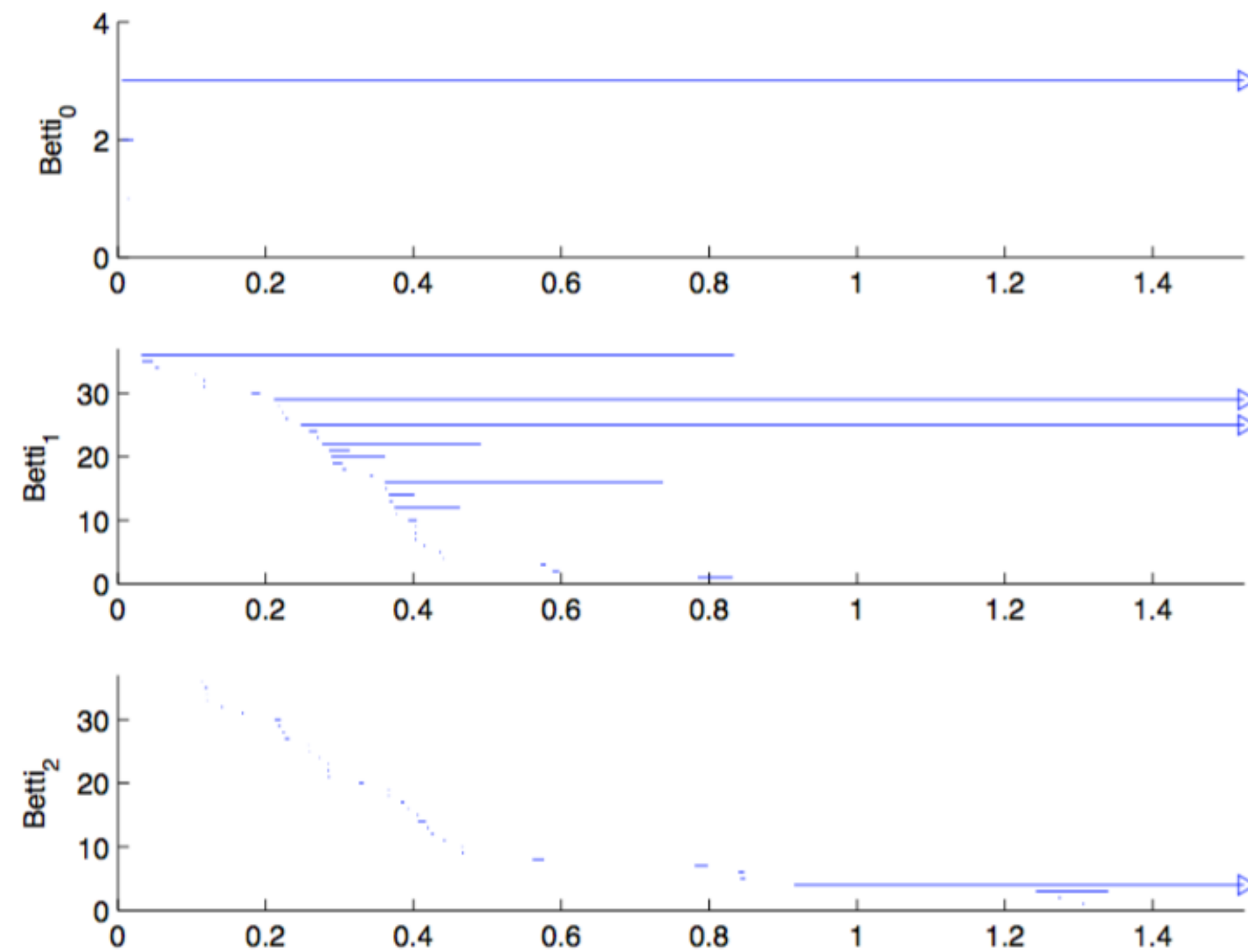


Figure 9: *PLEX* results for  $X(100, 10)$  in  $\mathcal{M}$

# Application

- Image compression: replace the high-contrast patches of the given image by the points on the surface of the embedded Klein bottle that best approximate them.



# Application: Texture Representation

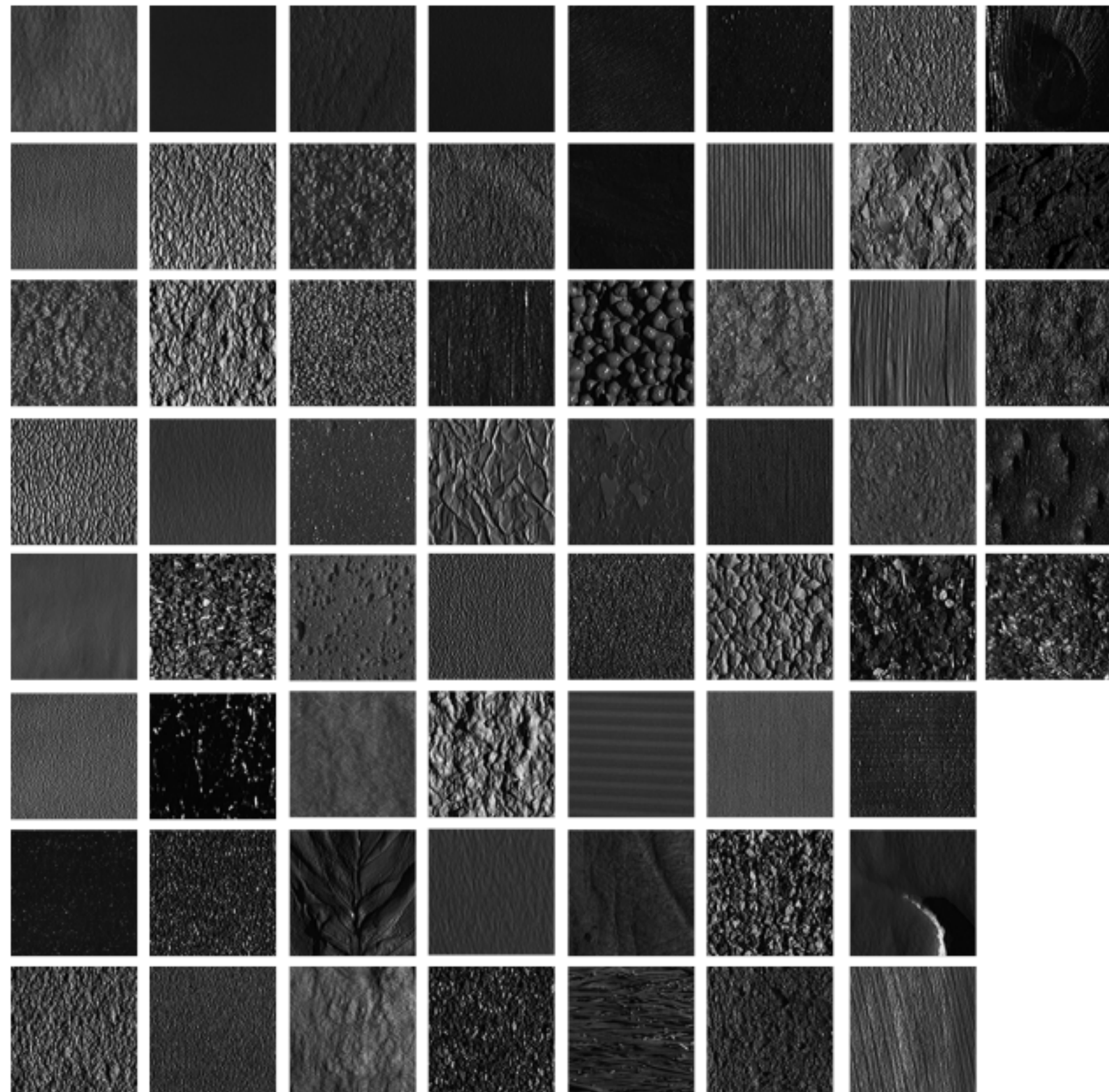


Figure 12: The 61 materials in the CURET data set

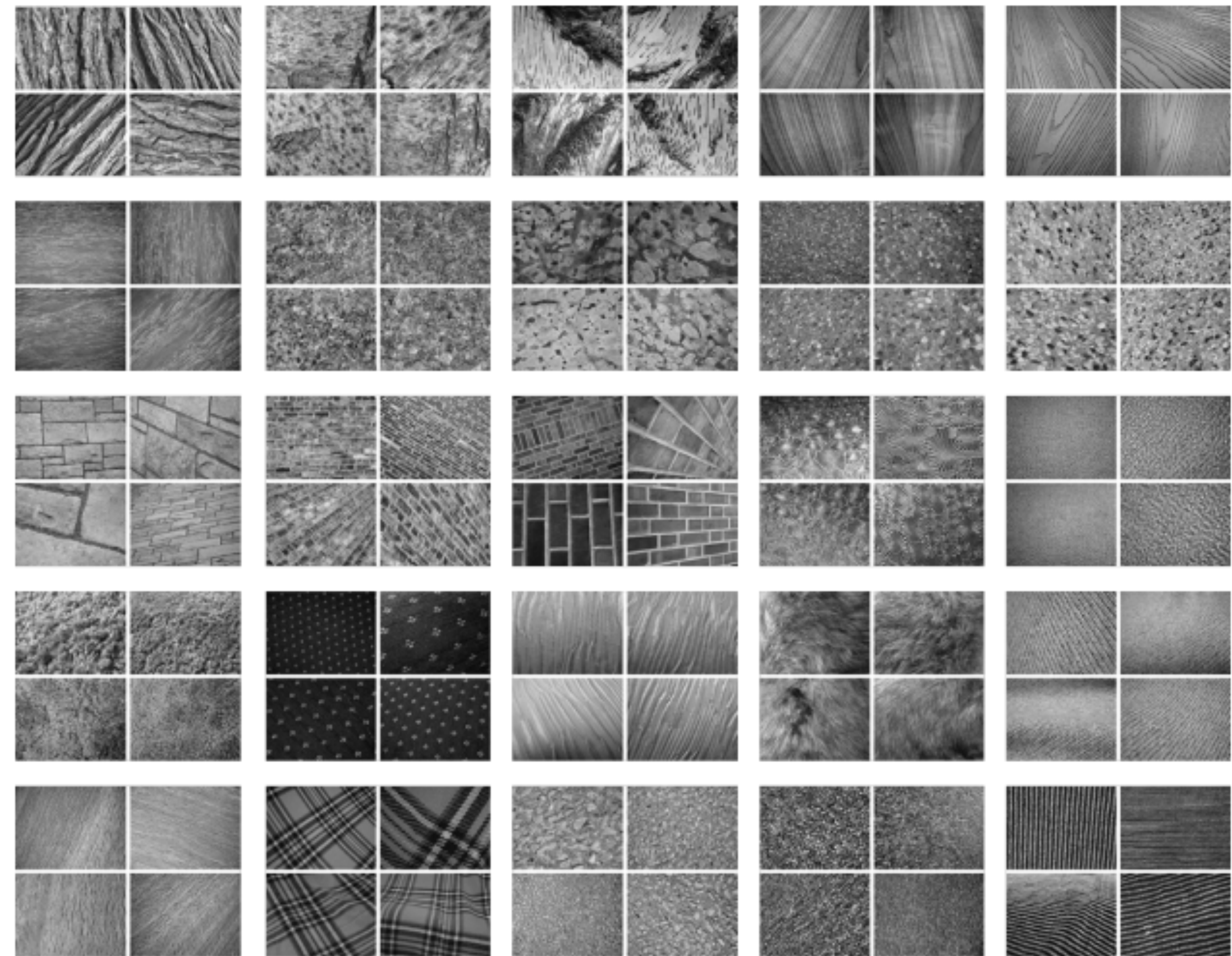
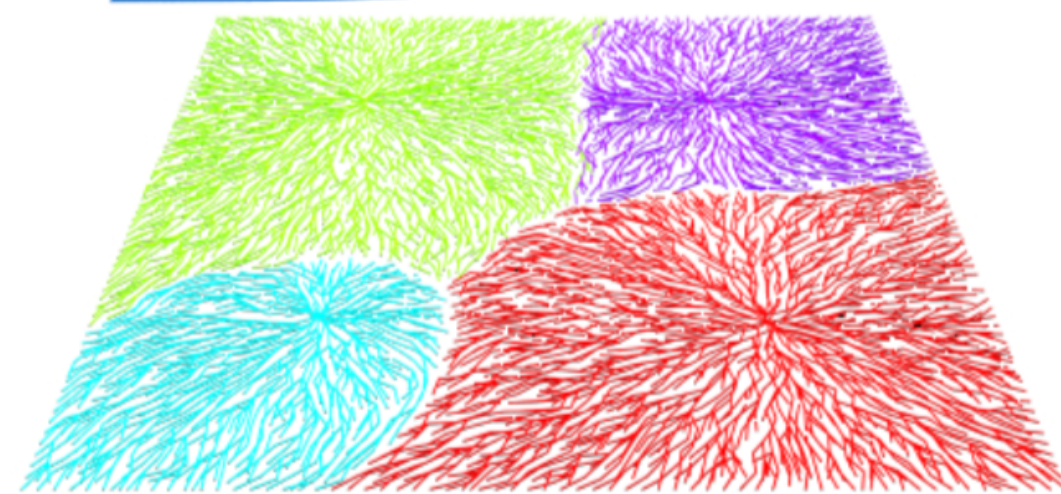
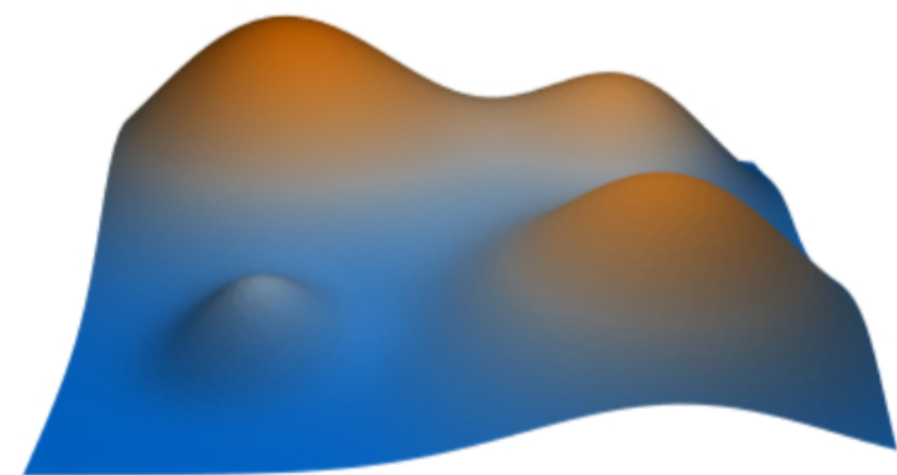
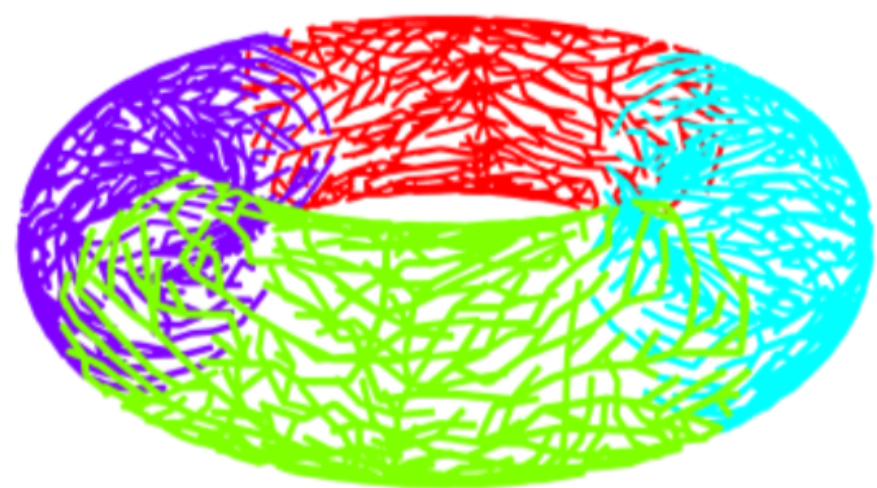
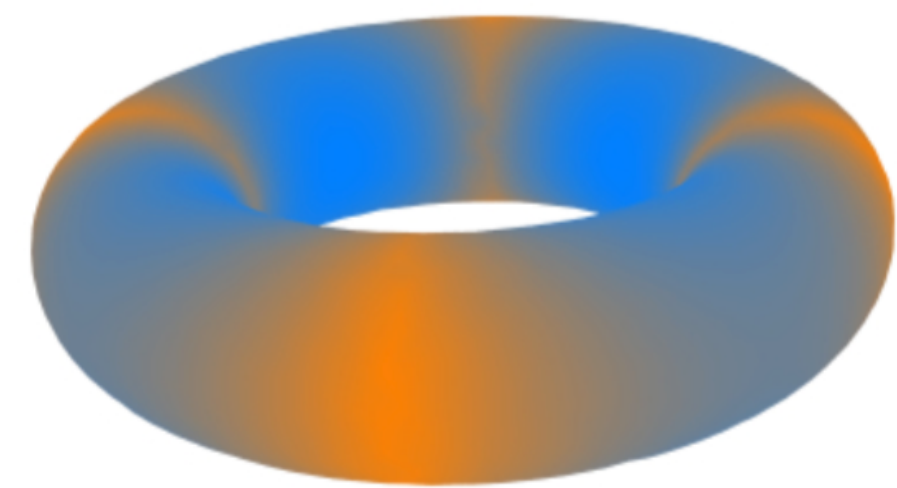


Figure 14: Four exemplars from each of the twenty five classes in the UIUC Texture data set.



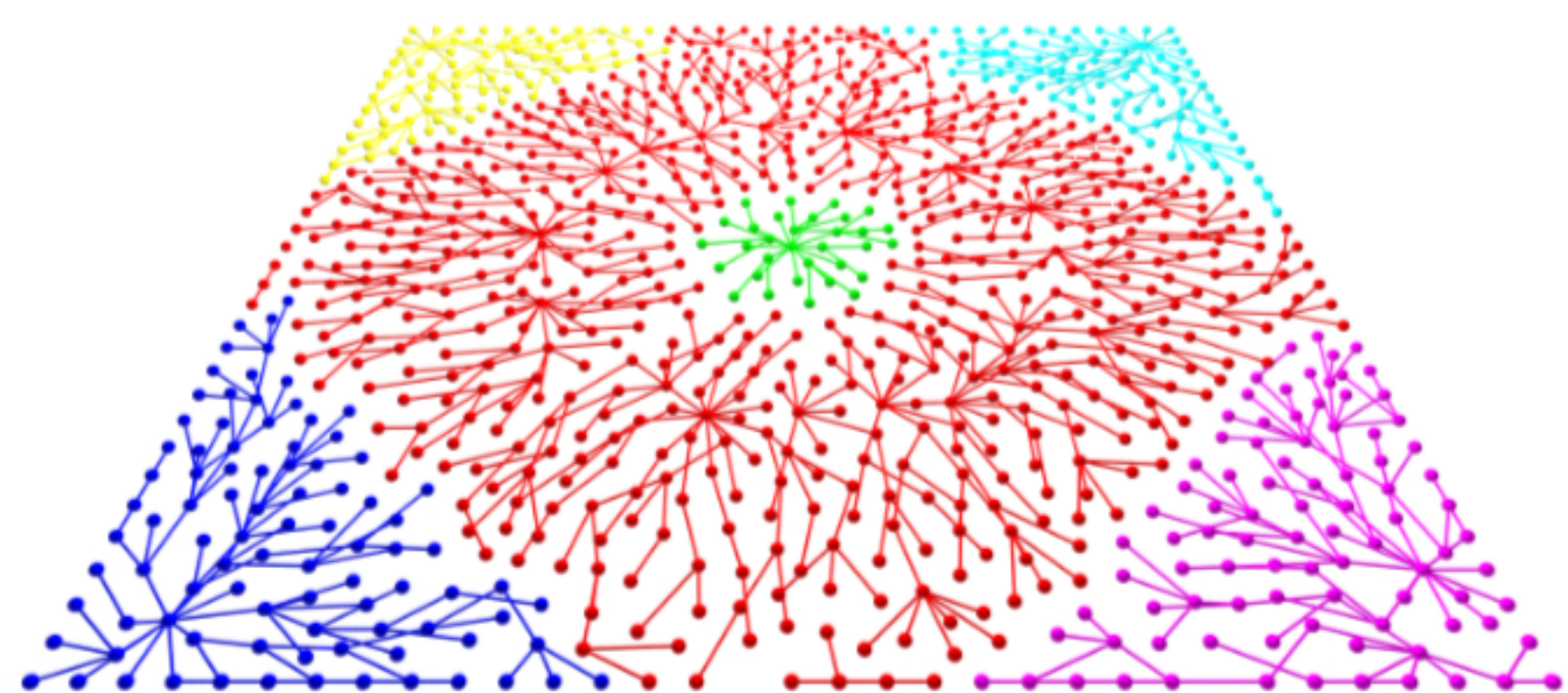
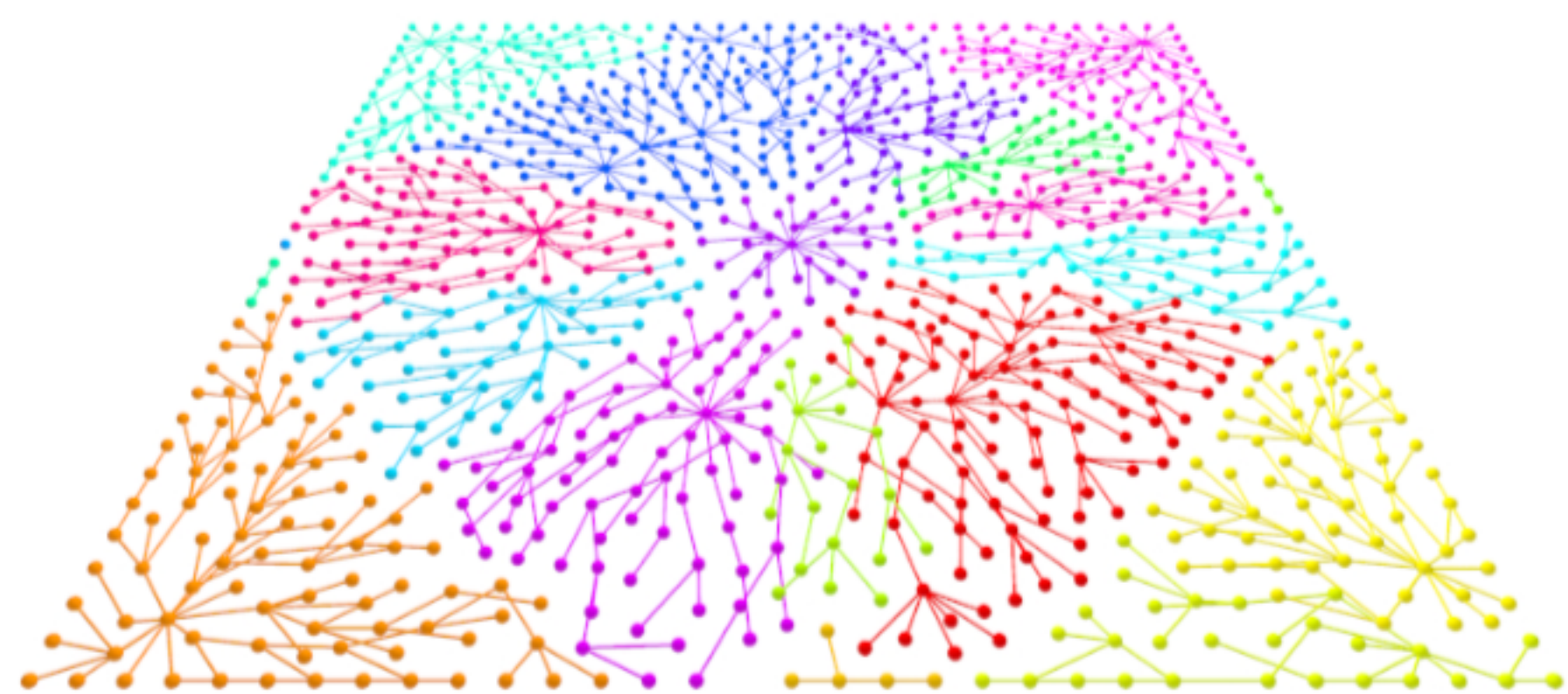
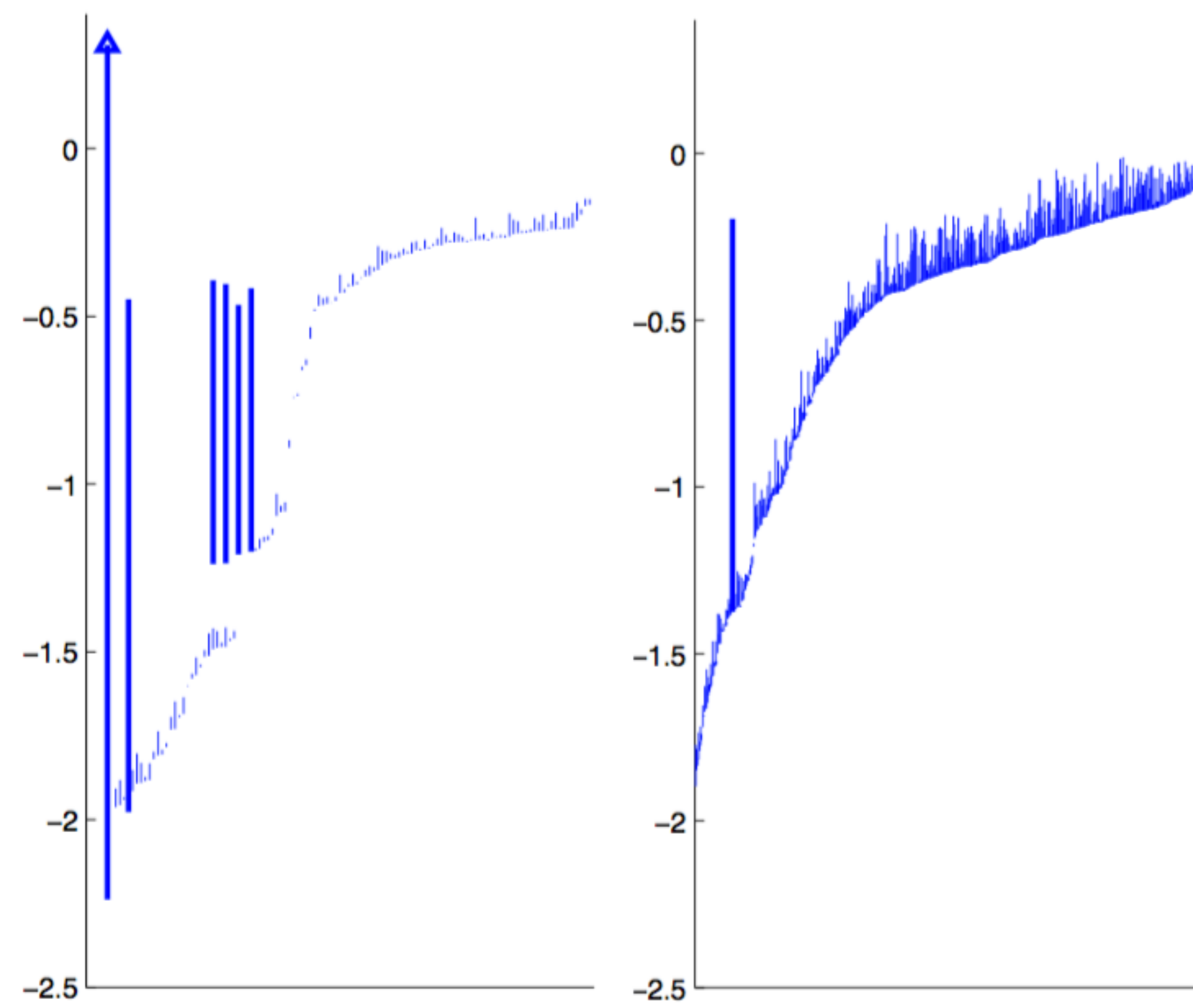
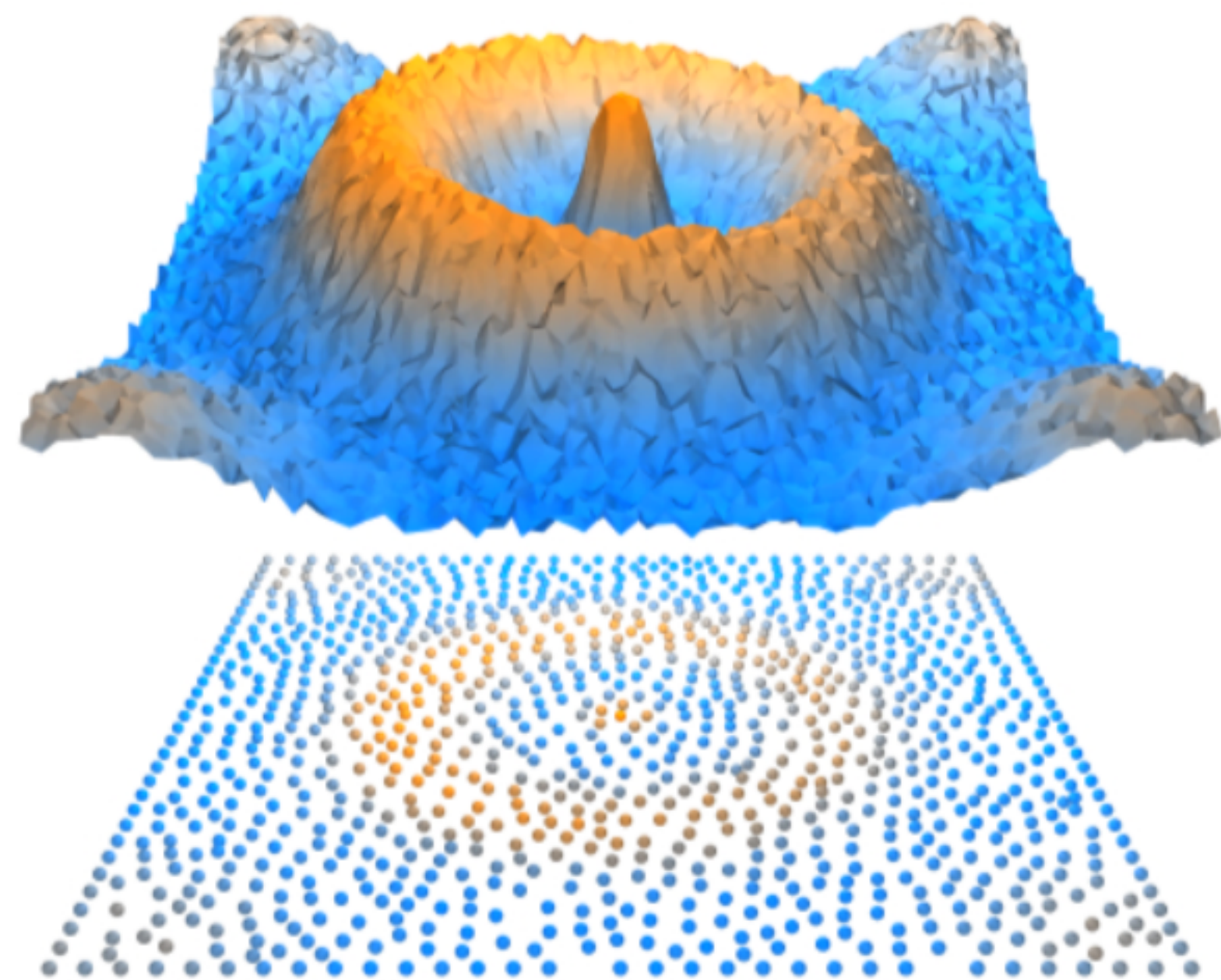


# Analysis of Scalar Fields over Point Cloud

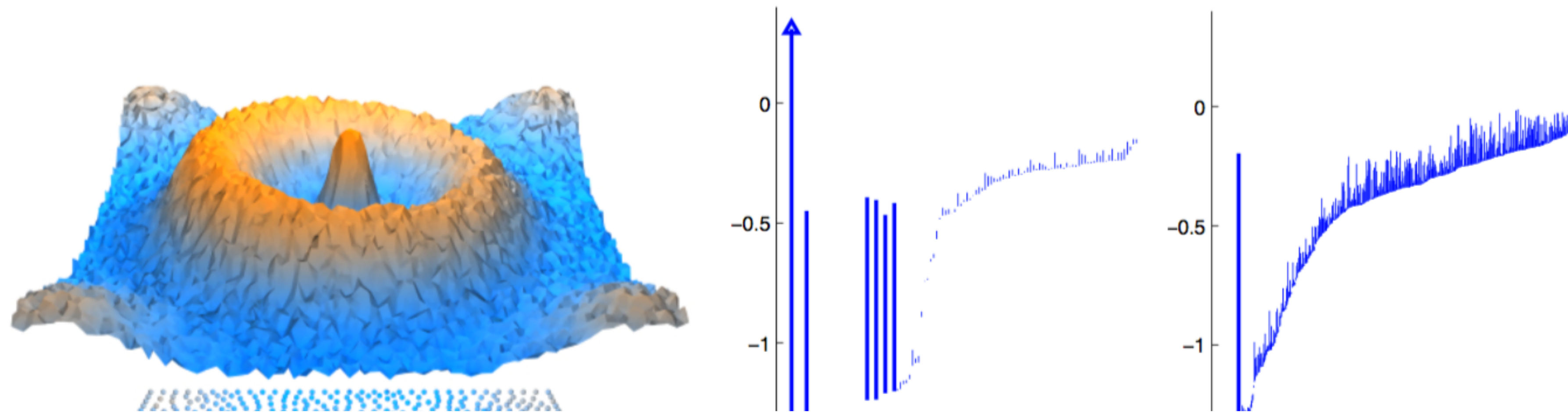
[ChazalGuibasOudot2008]

[http://ailab.ijs.si/primoz\\_skraba/papers/sf\\_SODA.pdf](http://ailab.ijs.si/primoz_skraba/papers/sf_SODA.pdf)





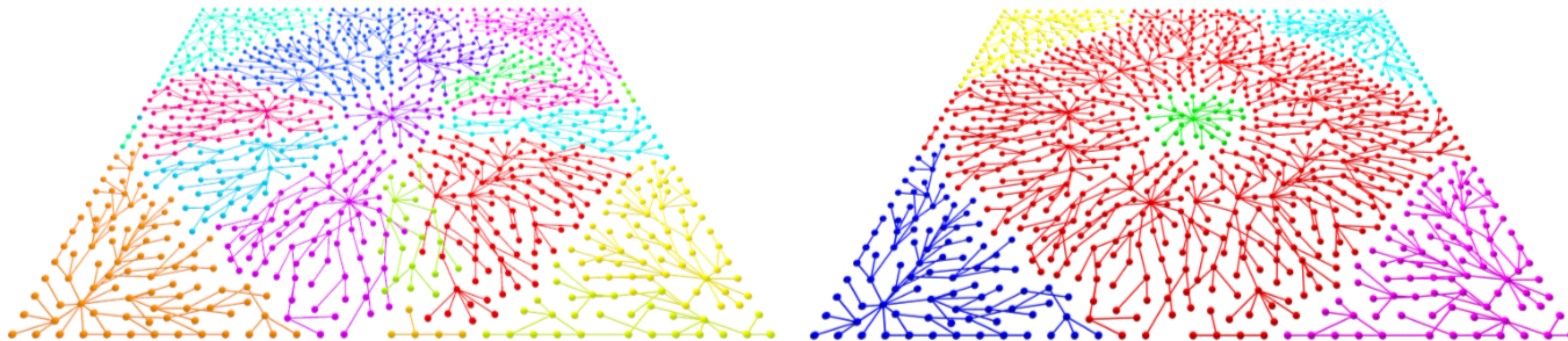




**Figure 1:** *Top row, left: a noisy scalar field  $f$  defined over a sampled planar square domain  $\mathbb{X}$ ; center and right: approximations of the 0- and 1-dimensional persistence barcodes of  $(-f)$  generated by our method from the values of  $f$  at the sample points and from their approximate pairwise geodesic distances in  $\mathbb{X}$ . The six long intervals in the 0-dimensional barcode correspond to the six prominent peaks of  $f$  (including the top of the crater), while the long interval in the 1-dimensional barcode reveals the ring shape of the basin of attraction of the top of the crater. Bottom row: approximate basins of attraction of the peaks of  $f$ , before (left) and after (right) merging non-persistent clusters, thus revealing the intuitive structure of  $f$ .*



**Figure 1:** *Top row, left: a noisy scalar field  $f$  defined over a sampled planar square domain  $\mathbb{X}$ ; center and right: approximations of the 0- and 1-dimensional persistence barcodes of  $(-f)$  generated by our method from the values of  $f$  at the sample points and from their approximate pairwise geodesic distances in  $\mathbb{X}$ . The six long intervals in the 0-dimensional barcode correspond to the six prominent peaks of  $f$  (including the top of the crater), while the long interval in the 1-dimensional barcode reveals the ring shape of the basin of attraction of the top of the crater. Bottom row: approximate basins of attraction of the peaks of  $f$ , before (left) and after (right) merging non-persistent clusters, thus revealing the intuitive structure of  $f$ .*

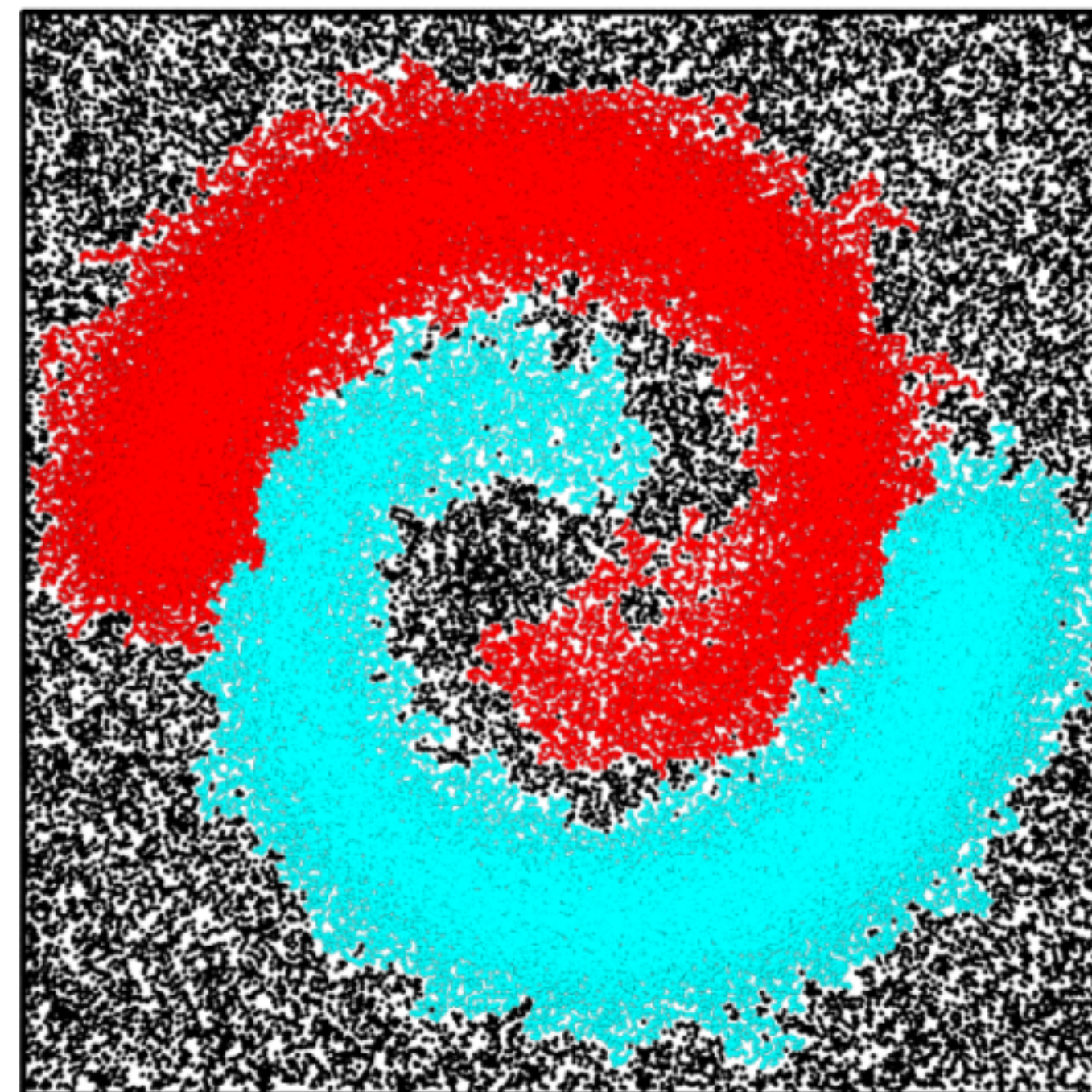
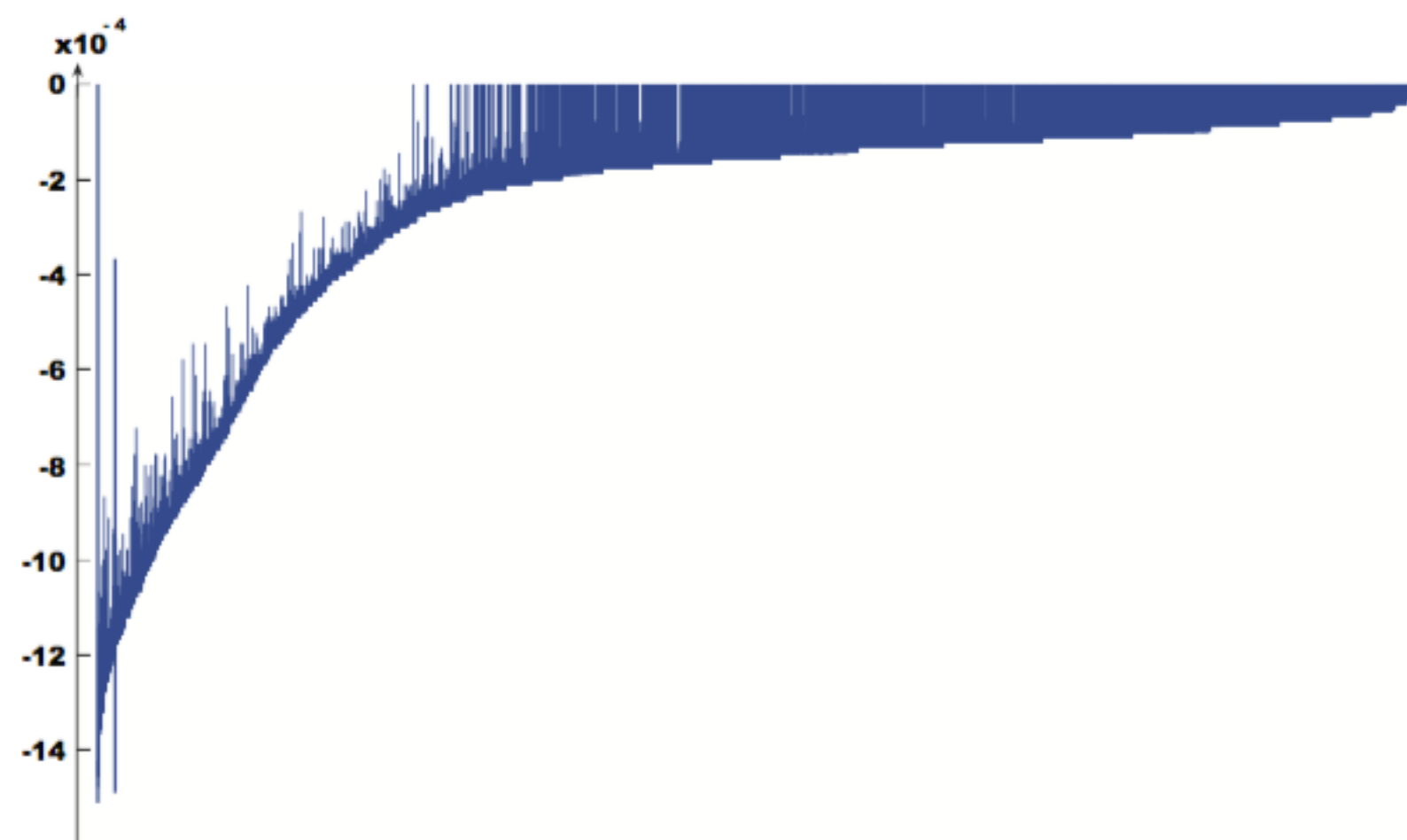
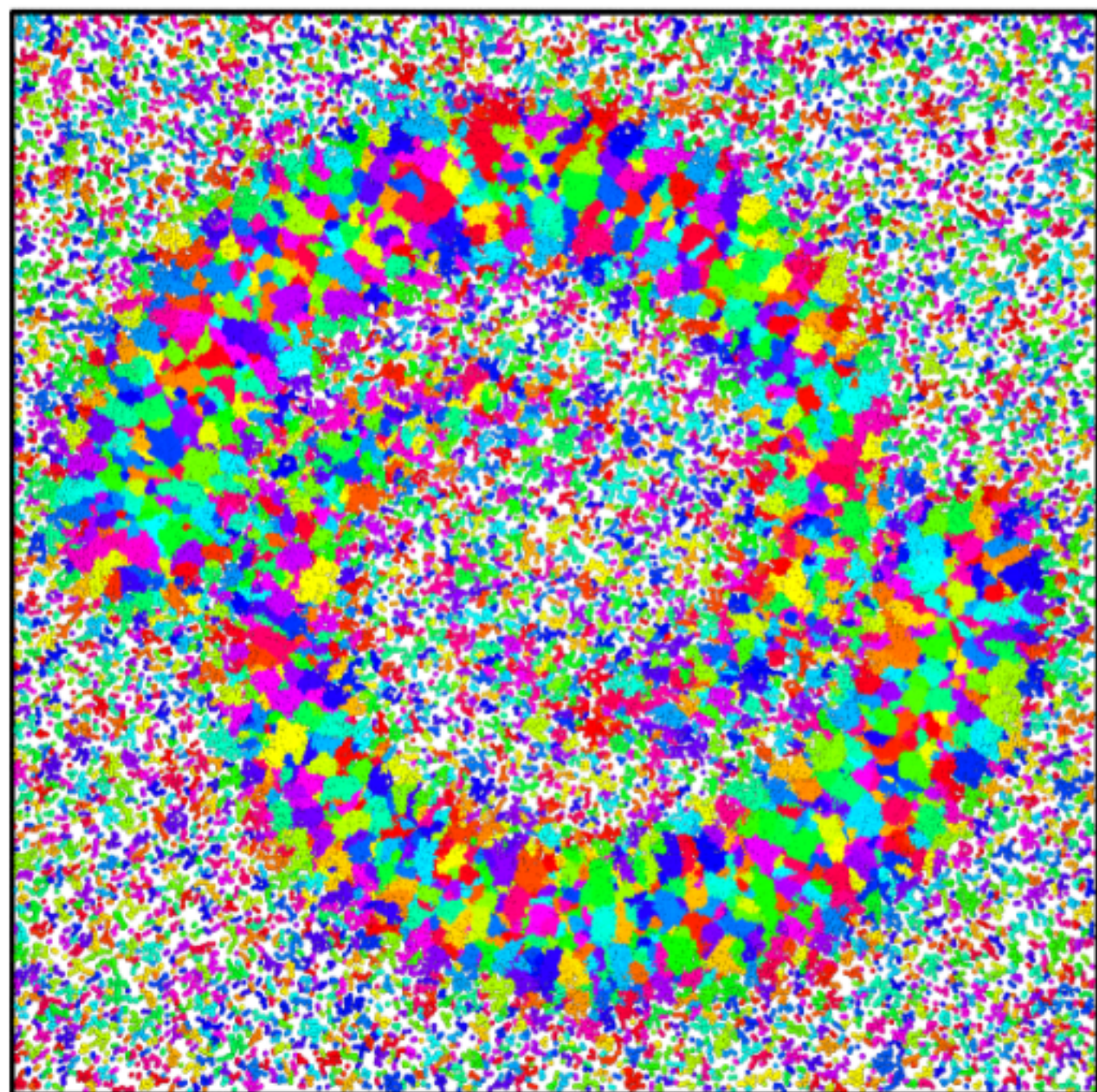
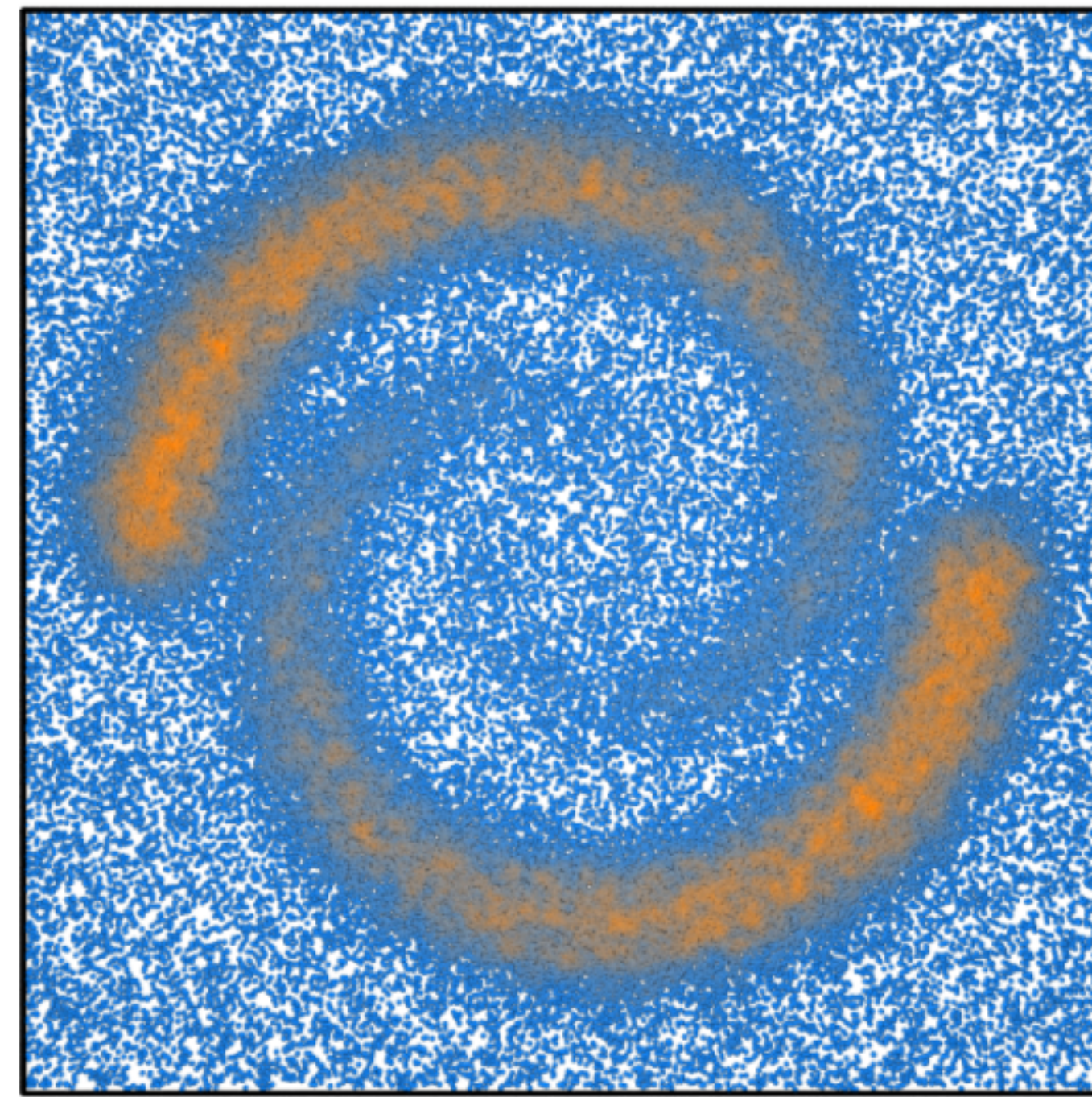
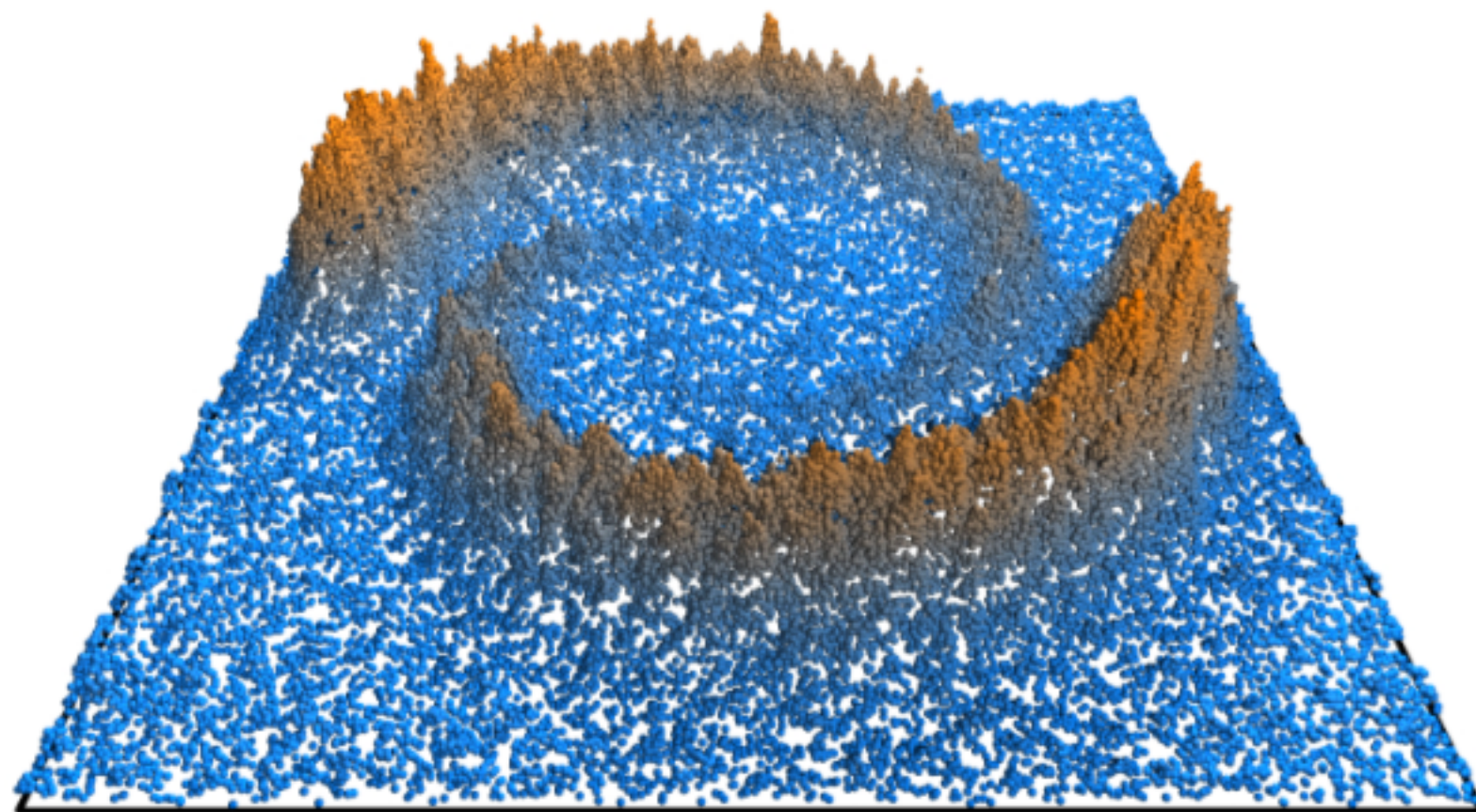
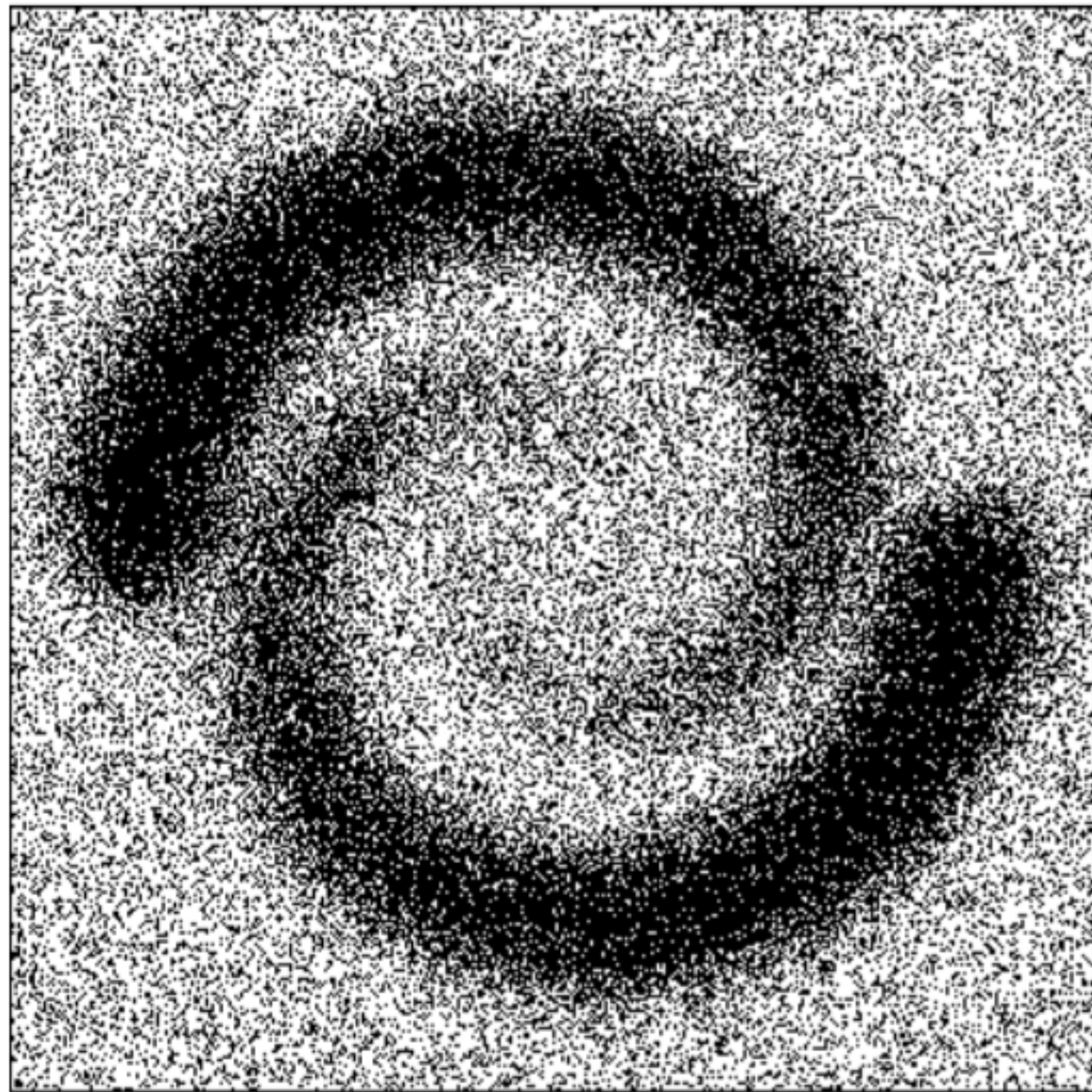




# High-level techniques

- Persistent homology
- Merge low persistent clusters while maintain high persistent clusters





[ChazalGuibasOudot2008]



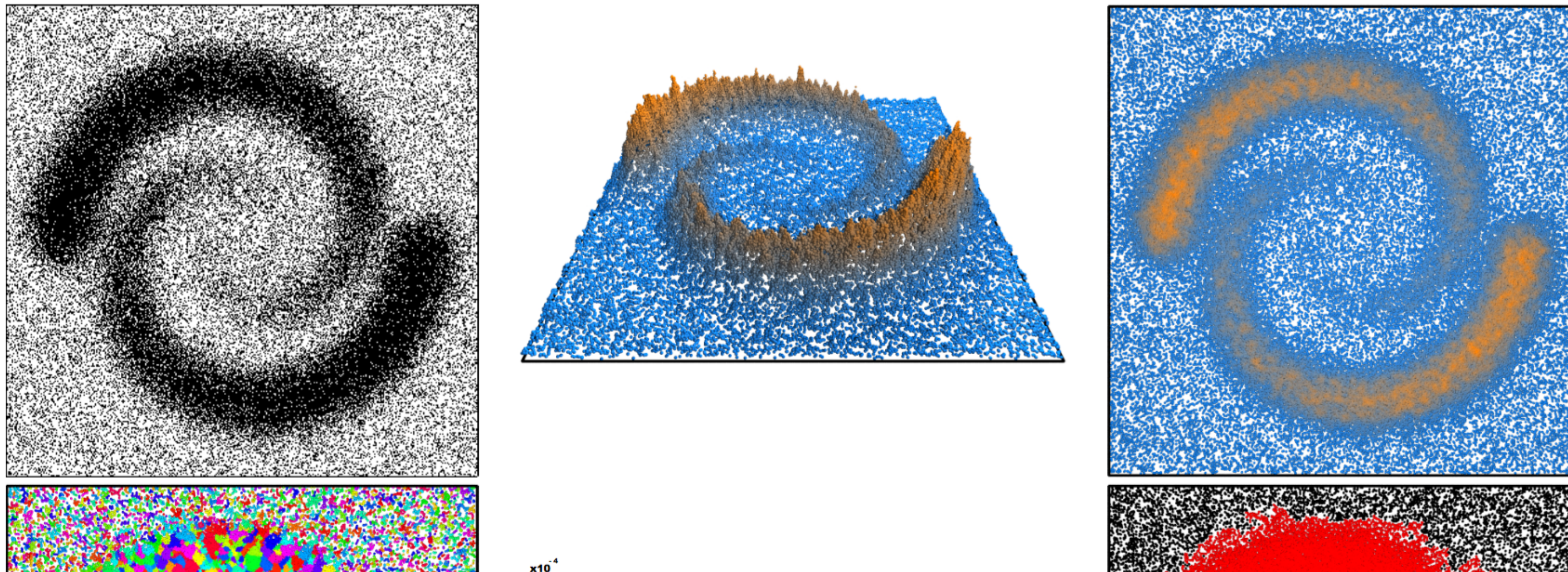
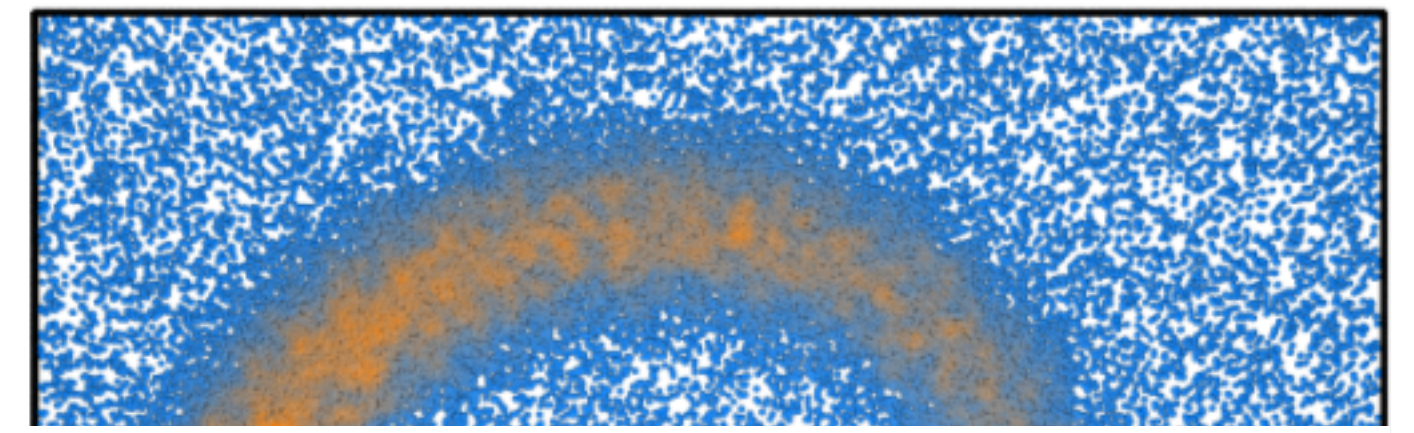
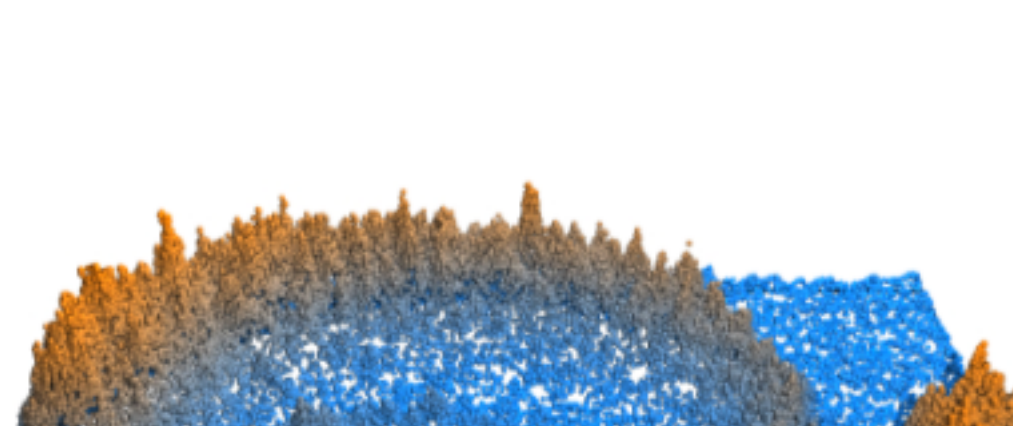
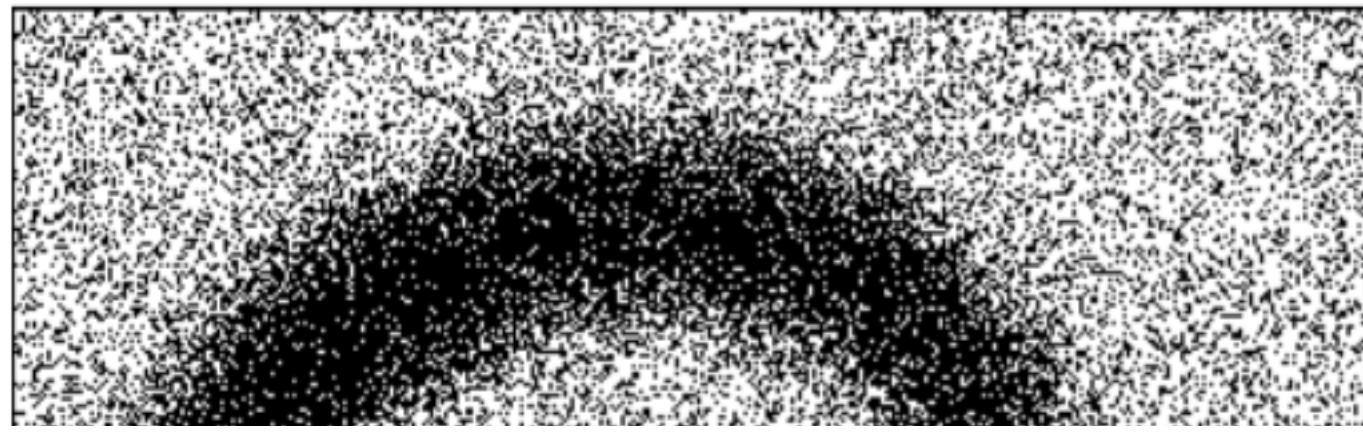


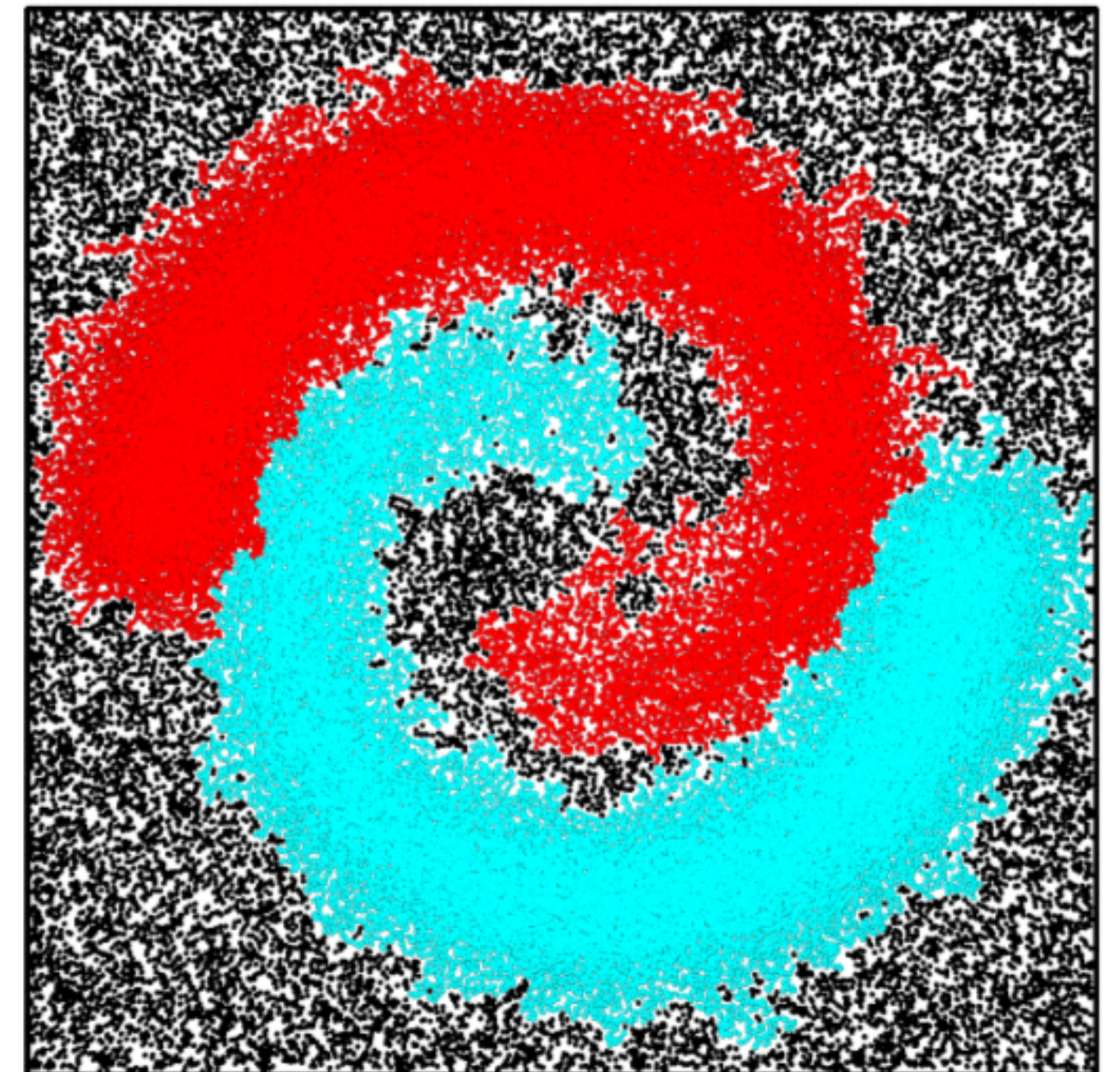
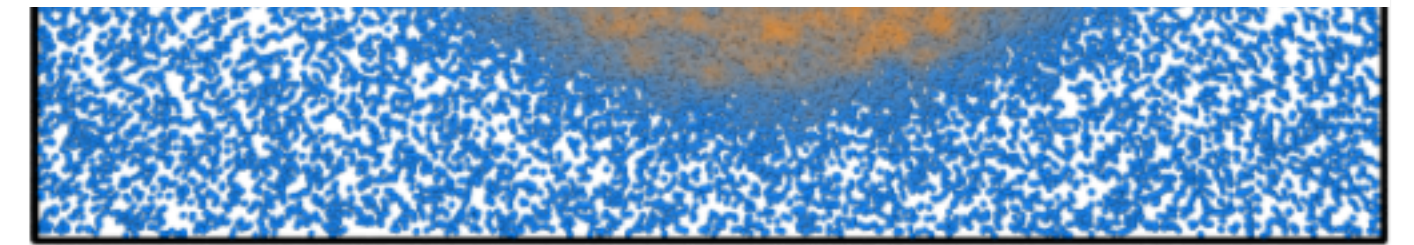
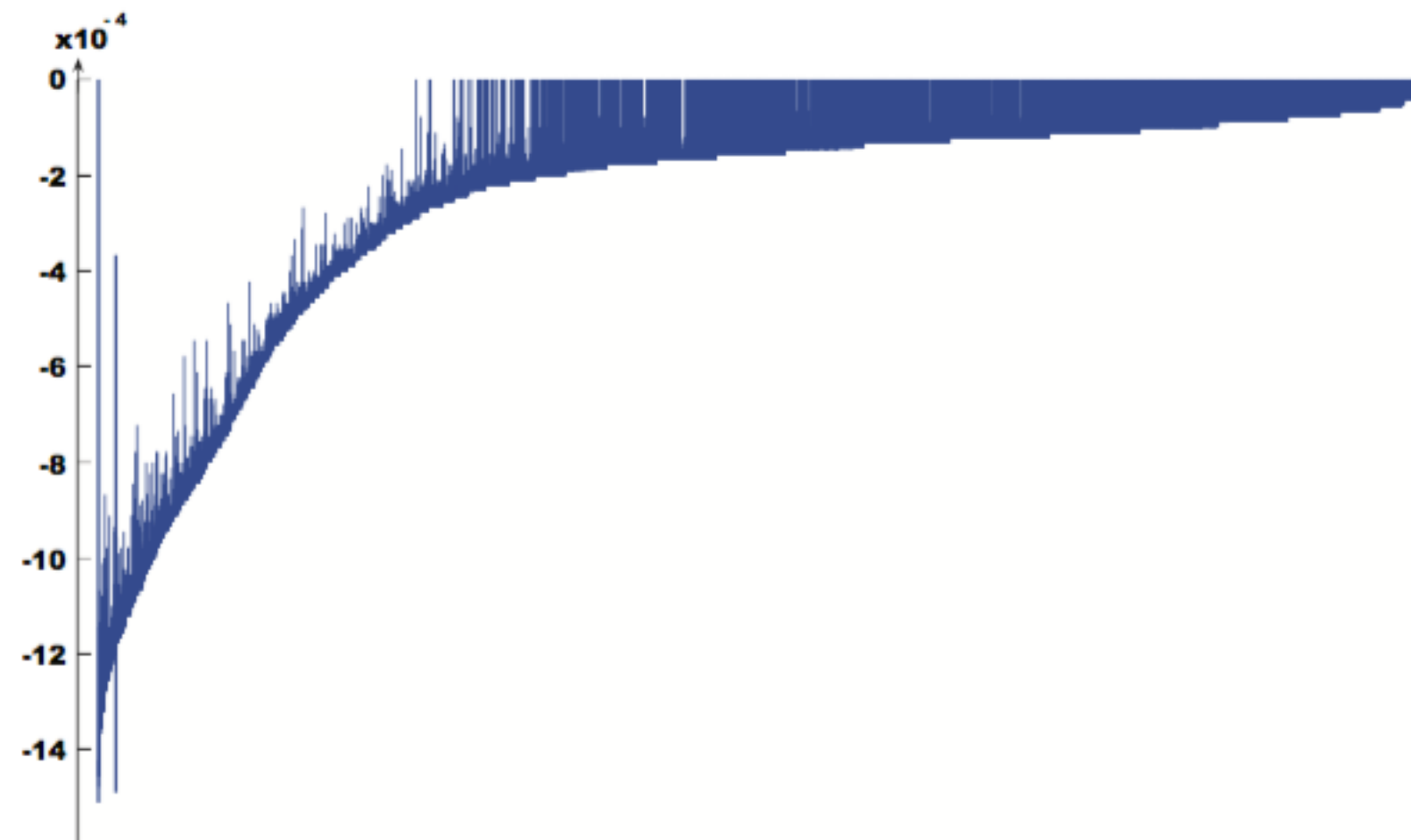
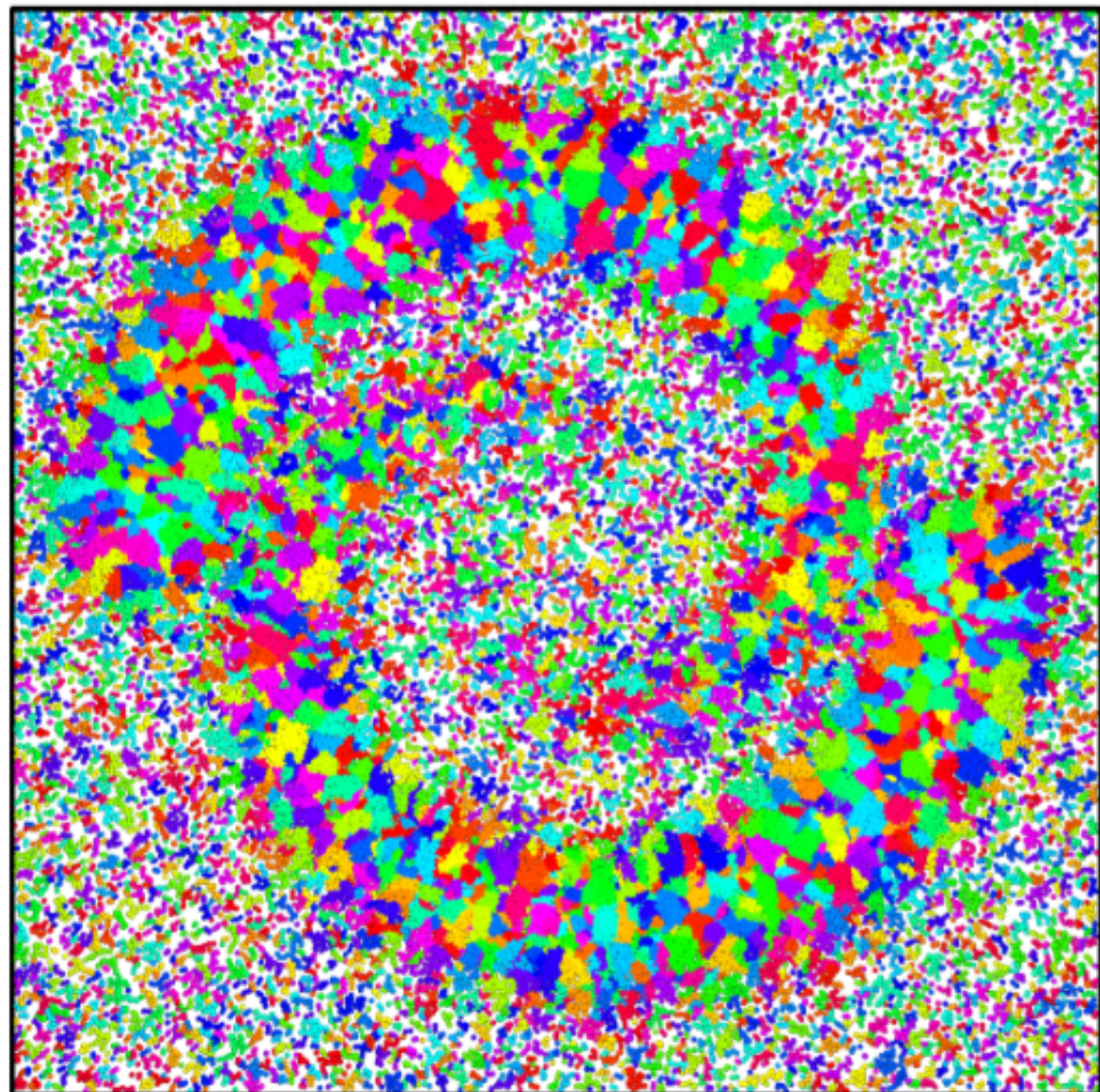
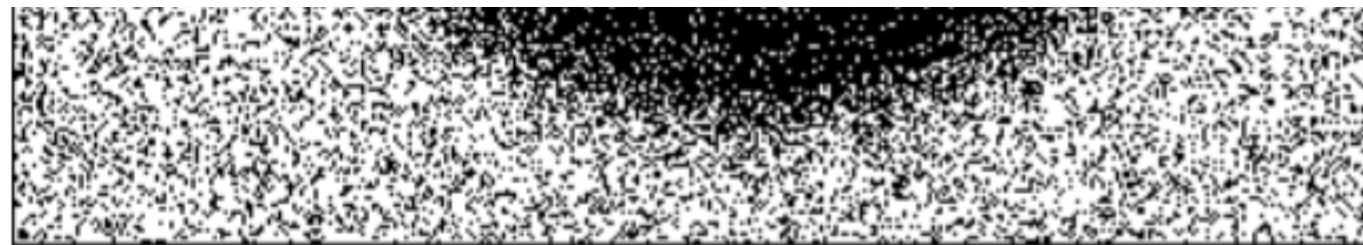
Figure 2: A result in clustering. The top row shows the input provided to the algorithm of Section 4.2: the data points (left), or rather their pairwise Euclidean distances, and the estimated density function  $f$  (center and right). The 3-d view of  $f$  illustrates how noisy this function can be in practice, thereby emphasizing the importance of our robustness result (Theorem 3.2). The bottom row shows the estimated basins of attraction of the peaks of  $f$ , before (left) and after (right) merging non-persistent clusters. The 0-dimensional persistence barcode of  $(-f)$  (center) contains two prominent intervals corresponding to the two main clusters. Since the estimated density is everywhere non-negative, the barcode has been thresholded at 0. Thus, intervals reaching 0 correspond to independent connected components in the Rips graph. Among those, the ones that appear lately are treated as noise and their basins of attraction shown in black, since their corresponding density peaks are low.







**Figure 2:** A result in clustering. The top row shows the input provided to the algorithm of Section 4.2: the data points (left), or rather their pairwise Euclidean distances, and the estimated density function  $f$  (center and right). The 3-d view of  $f$  illustrates how noisy this function can be in practice, thereby emphasizing the importance of our robustness result (Theorem 3.2). The bottom row shows the estimated basins of attraction of the peaks of  $f$ , before (left) and after (right) merging non-persistent clusters. The 0-dimensional persistence barcode of  $(-f)$  (center) contains two prominent intervals corresponding to the two main clusters. Since the estimated density is everywhere non-negative, the barcode has been thresholded at 0. Thus, intervals reaching 0 correspond to independent connected components in the Rips graph. Among those, the ones that appear lately are treated as noise and their basins of attraction shown in black, since their corresponding density peaks are low.





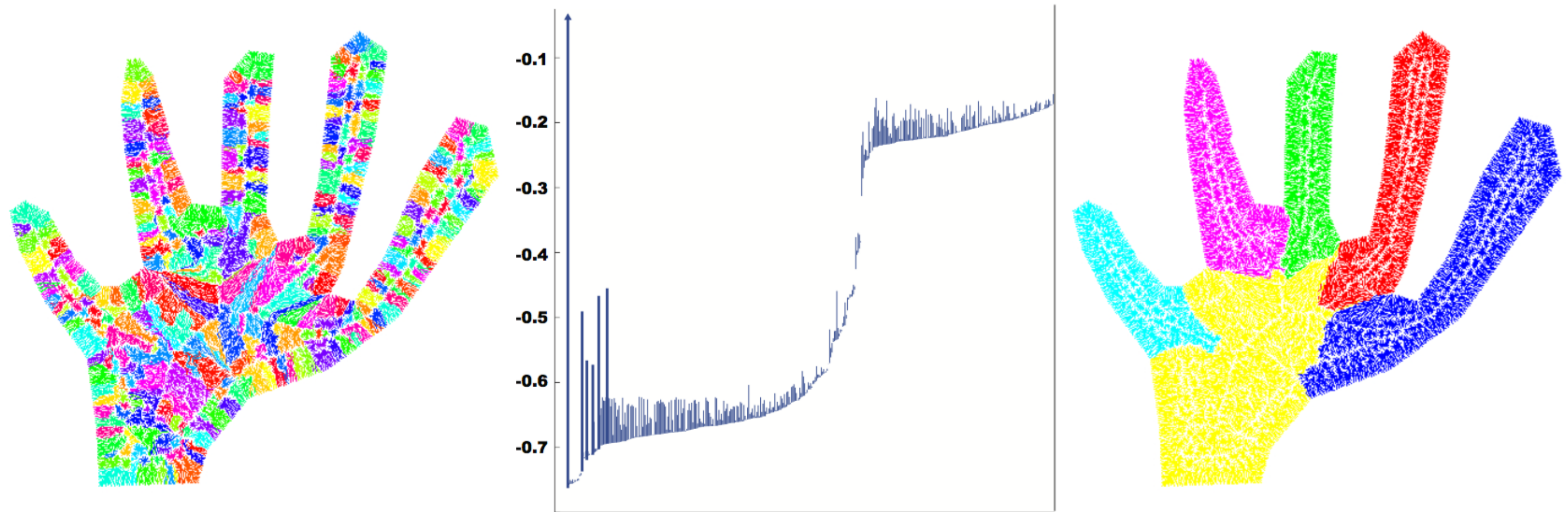


Figure 3: Segmentation result on a sampled hand-shaped 2-D domain. The segmentation function is the (normalized) diameter of the set of nearest boundary points. The barcode shows six long intervals, corresponding to the palm of the hand and to the five fingers. The results before and after merging non-persistence clusters are shown respectively to the left and to the right of the barcode.





# Thanks!

Any questions?

You can find me at: [beiwang@sci.utah.edu](mailto:beiwang@sci.utah.edu)



# CREDITS

Special thanks to all people who made and share these awesome resources for free:

- ☐ Presentation template designed by [Slidesmash](#)
- ☐ Photographs by [unsplash.com](#) and [pexels.com](#)
- ☐ Vector Icons by [Matthew Skiles](#)



# Presentation Design

This presentation uses the following typographies and colors:

## Free Fonts used:

<http://www.1001fonts.com/oswald-font.html>

<https://www.fontsquirrel.com/fonts/open-sans>

## Colors used

