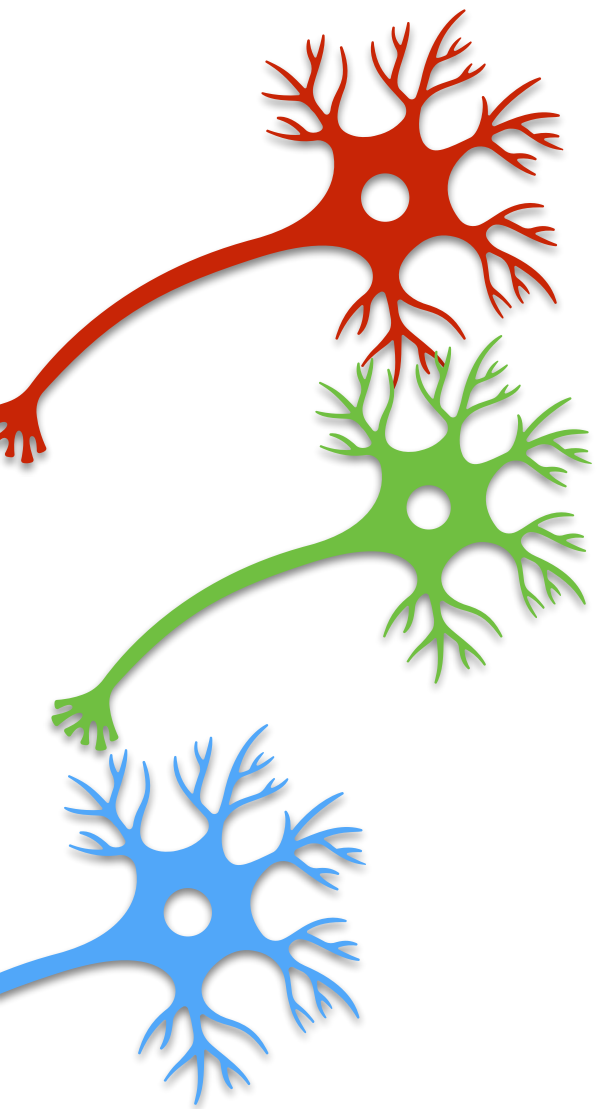
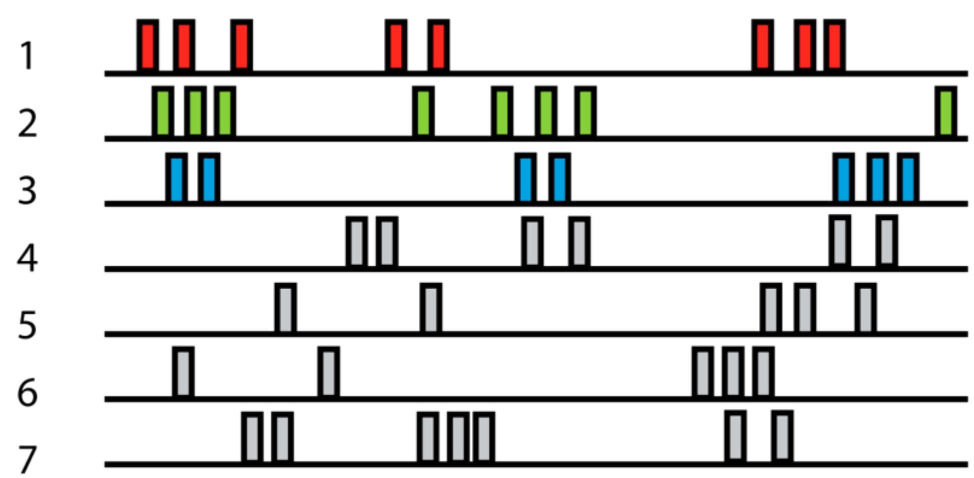
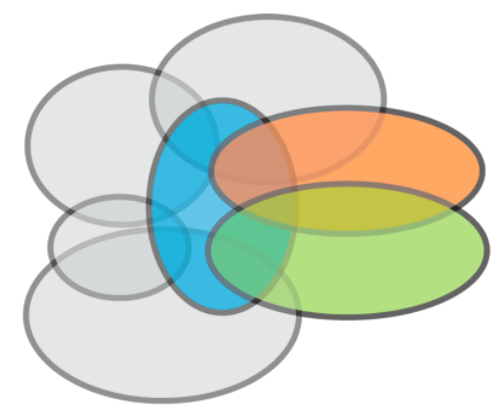
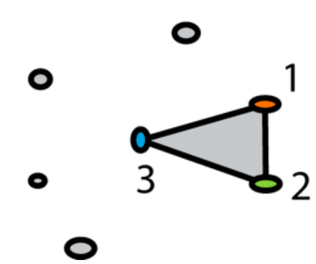
TDA+Neuro

Topological Data Analysis and Neuroscience
CSE 5339









Facundo Mémoli

Organization

- Lecturers & TAs.
  - **Facundo Mémoli.** Main lecturer.
  - **Alex McCleary.** Alex will coordinate all intro lectures.
  - **Brantley Vose.** Brantley will collect attendance and will be the point of contact for providing feedback about your presentation materials (in advance to your respective talks).
  - **Nate Clause**. Nate will be the webmaster, and will take care of all timetabling/coordination, collecting HW, collecting presentation materials, etc.
  - **Ling Zhou**. Internal coordination with group meetings.
- Webpage: **https://github.com/ndag/TDA-and-Neuro/**
- Google group: TBA

The webpage/resources

https://github.com/ndag/TDA-and-Neuro/

README.md

Topological Data Analysis in Neuroscience (TDA+Neuro)

Description/contents:

In recent years, ideas from TDA have been increasingly adopted in order to analyze neuroscie
particular interest are TDA techniques which can summarize the information contained in the j
patterns of ensmebles of neurons. This topic course will overview the main ideas in this field.

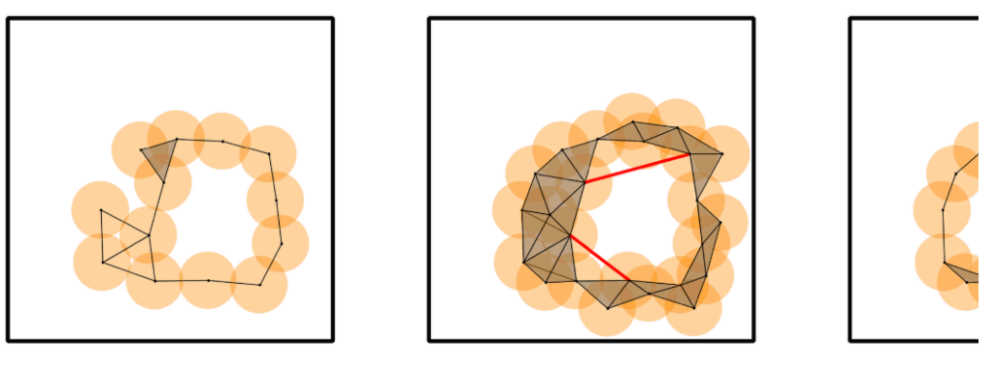


Figure 1. A filtration on a data set.

Prerequisites:

Prerequisites:

Basic familiarity with data analysis principles and with linear algebra. Familiarity with topological data analysis will be beneficial.

Some Resources:

Here are some useful papers for reference:

(1) Cell groups reveal structure of stimulus space. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000205

(2) A Topological Paradigm for Hippocampal Spatial Map Formation Using Persistent Homology. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002581

(3) The importance of forgetting: Limiting memory improves recovery of topological characteristics from neural data. https://journals.plos.org/plosone/article/comments?id=10.1371/journal.pone.0202561

(4) Evaluating State Space Discovery by Persistent Cohomology in the Spatial Representation System. https://www.frontiersin.org/articles/10.3389/fncom.2021.616748/full

(5) Some more papers: A bibliography by Chad Giusti, http://www.chadgiusti.com/bib.html .

Schedule

Class 1. 1/11/2022. Brief discussion and introduction. Lecture by Facundo Mémoli.

Class 2. 1/18/2022. Introduction to Persistence. Lecture by Facundo Mémoli.

Class 3. 1/25/2022. Multiparameter persistence. Lecture by Alex McCleary.

Class 4. 2/1/2022. Introduction to RIVET and ZigZag Persistence. Lecture by Nate Clause.

Format

- There will be 5—6 intro lectures:
  - **W1**. Facundo. Intro.
  - **W2**. Facundo. Basics of TDA. One-parameter persistent homology.
  - **W3**. Alex. Multiparameter and zigzag persistence. Overview.
  - **W4**. Nate. Multiparameter and zigzag persistence. Computational aspects.
  - **W5.** Brantley. The importance of forgetting!
  - **W6**. Ling. TBA.

Format

- After these intro lectures **all of you** enrolled in this course will be giving lectures about papers/chapters.
- **Today** we will collect your email addresses and other information such as Major, research interests, etc.
- By end of **W3** all of you will be assigned papers that you should start reading for an eventual presentation. We will let you know the dates starting **W6**
- An (evolving) list of papers is posted on the webpage.

Other stuff

- **2 credit course**: Besides class attendance, expect to spend a significant amount of time, at least 4x15 = **60 hours,** on the course: reading papers, preparing your presentation, solving exercises, etc.
- Grades (A-F) will be decided based on:
  - **Presentations** (55%). We want high quality presentations: more to be discussed in coming weeks. You'll have to provide us with your draft presentation materials a week beforehand. Brantley and Nate will have a meeting with you to provide feedback which you'll need to incorporate.
  - **HW** (15%). There will be a few HW exercises at the beginning (W1 to W6).
  - **Class participation** (15%) and **class attendance** (15%).
    - Note: mandatory class attendance.
    - If you have medical condition: notify me and Brantley immediately. We'll let you know make up assignments.

Introductions

Background

What is Neuroscience?

Wikipedia
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute

Help
Learn to edit
Community portal
Recent changes
Upload file

Tools

What links here
Related changes
Special pages
Permanent link
Page information
Cite this page

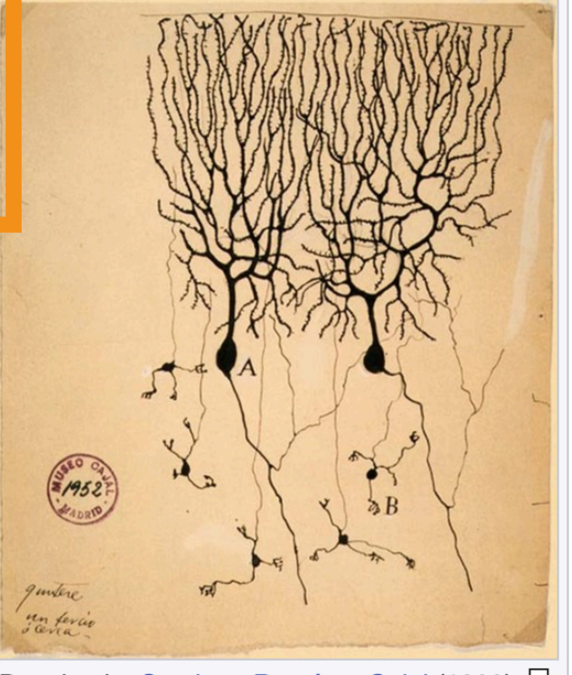
Neuroscience

From Wikipedia, the free encyclopedia

For the journal, see Neuroscience (journal).
"Brain science" redirects here. For other aspects of brain science, see cognitive science, cognitive psychology, neurology, and neuropsychology.

Neuroscience is the scientific study of the nervous system.[1] It is a multidisciplinary science that combines physiology, anatomy, molecular biology, developmental biology, cytology, computer science and mathematical modeling to understand the fundamental and emergent properties of neurons, glia and neural circuits.[2][3][4][5][6] The understanding of the biological basis of learning, memory, behavior, perception, and consciousness has been described by Eric Kandel as the "epic challenge" of the biological sciences.[7]

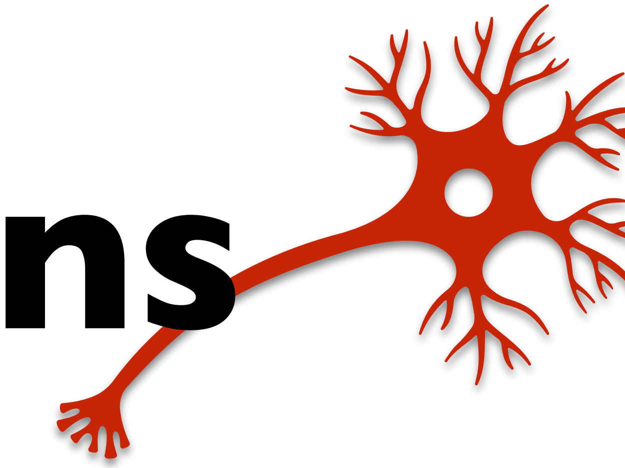
The scope of neuroscience has broadened over time to include different approaches used to study the nervous system at different scales. The techniques used by neuroscientists have expanded enormously, from molecular and cellular studies of individual neurons to imaging of sensory, motor and cognitive tasks in the brain.



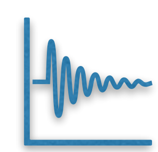
Drawing by Santiago Ramón y Cajal (1899) of neurons in the pigeon cerebellum

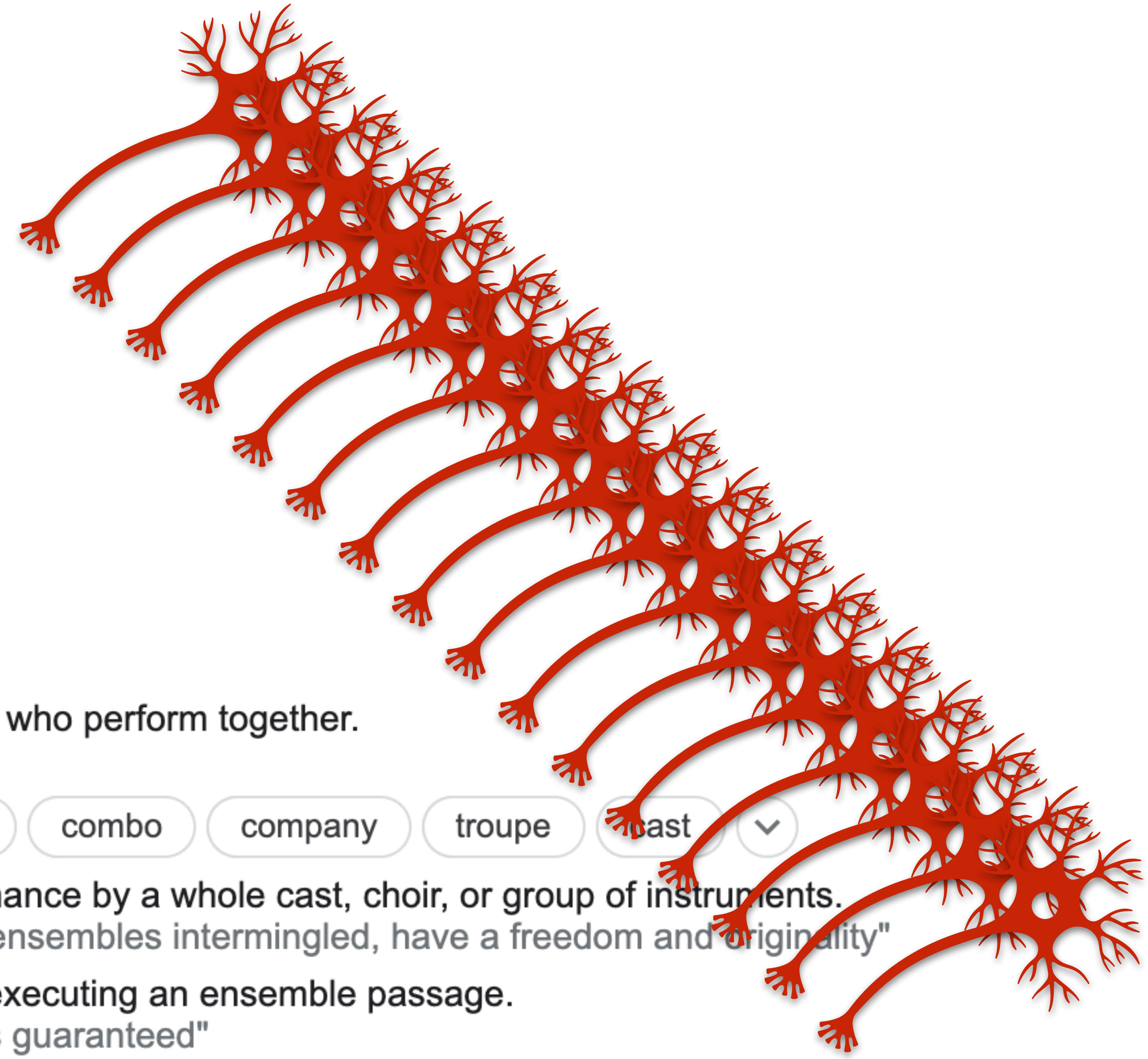
Contents [hide]

Basic unit: **neurons**



"A **neuron** or **nerve cell** is a electrically excitable cell that communicates with other cells via specialized connections called synapses."

- Our simplistic model will be:
  - Neurons are cells which 'fire', i.e. they generate electrical signals. 
  - Neurons are interconnected —> systems/**ensembles** of neurons



en·sem·ble

/änˈsämbəl/

See definitions in:

All | Music | Physics

noun

noun: **ensemble**; plural noun: **ensembles**

1. a group of musicians, actors, or dancers who perform together.
   "a Bulgarian folk ensemble"

   Similar: group · band · orchestra · combo · company · troupe · cast

   - a scene or passage written for performance by a whole cast, choir, or group of instruments.
     "Cherubini's numbers, with solos and ensembles intermingled, have a freedom and originality"
   - the coordination between performers executing an ensemble passage.
     "a high level of tuning and ensemble is guaranteed"

2. a group of items viewed as a whole rather than individually.
   "the buildings in the square present a charming provincial ensemble"

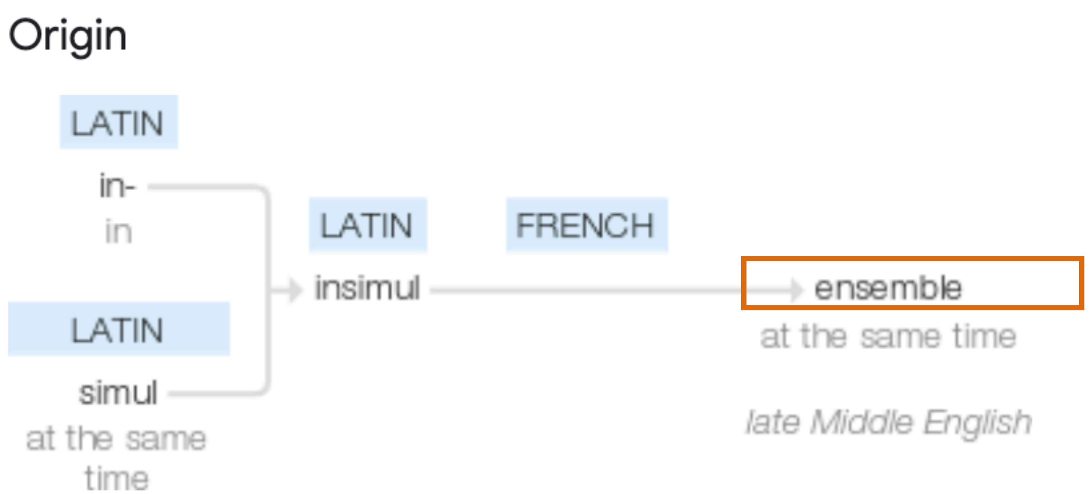
   Similar: whole · whole thing · entity · unit · unity · body · piece · object

   - a set of clothes chosen to harmonize when worn together.

     Similar: outfit · costume · suit · coordinates · matching separates

   - **PHYSICS**
     a group of similar systems, or different states of the same system, often considered statistically.
     "we would have to adopt a picture in which there is an ensemble of all possible universes with some probability distribution"

etymology



late Middle English (as an adverb (long rare) meaning ‘at the same time’): from French, based on Latin *insimul*, from *in-* ‘in’ + *simul* ‘at the same time’. The noun dates from the mid 18th century.

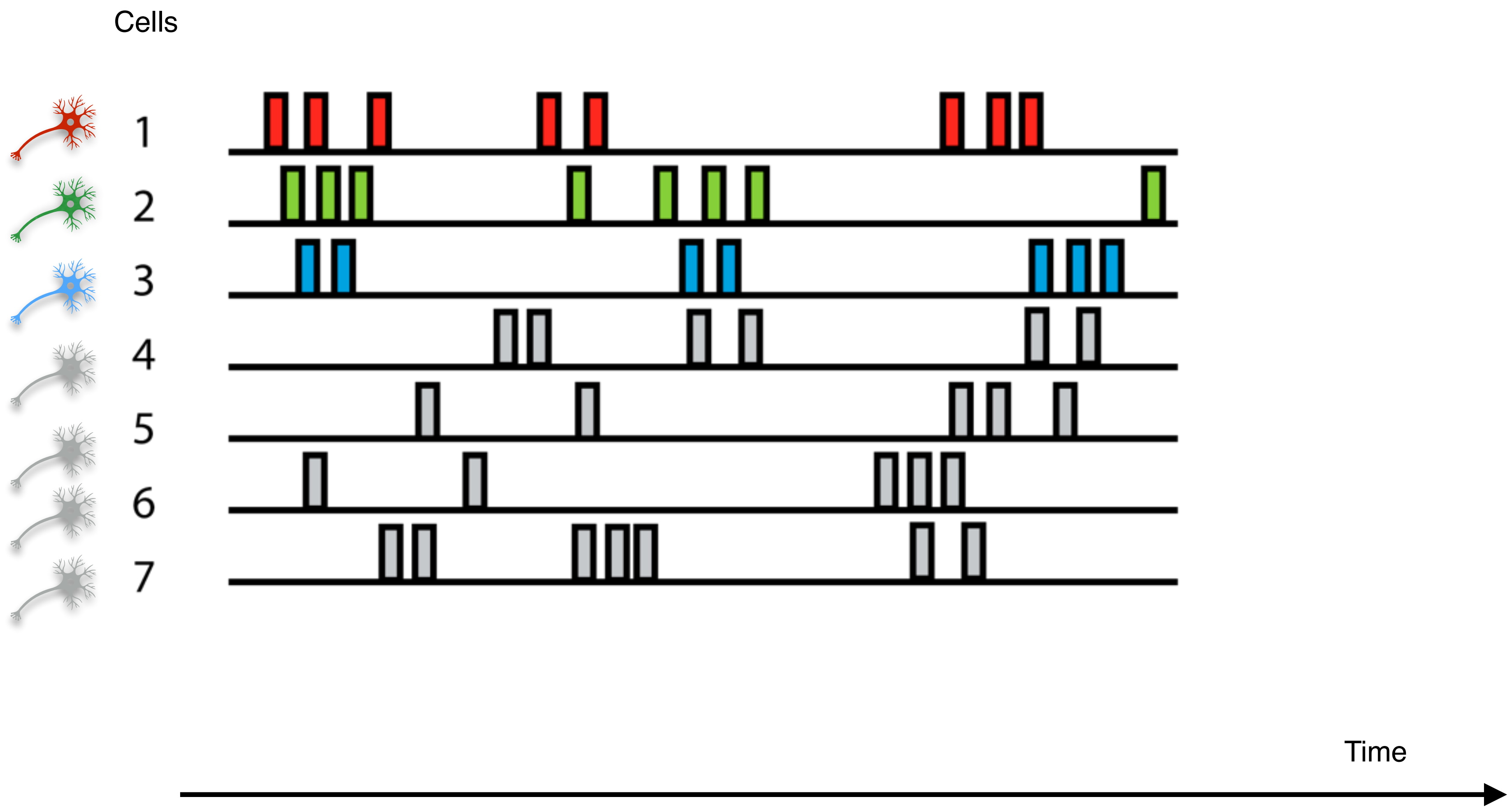
Related words: SIMULtanenous, SIMULate, Assemble, etc

In french: ‘ensemble’ means ‘together’

Neuronal ensembles

- We want to understand the information encoded by collections/ensembles of neurons.
- Neurons are observed through their (electrical) activity —> **spike trains**
- We assume a signal **s: T —> {0,1}** represents the activity/ inactivity of a given neuron. Here T is the time interval of an experiment.
    - Explore more: MIT online lecture https://www.youtube.com/watch?v=osYGG7TKcz8
- For an ensemble we have: **spike trains**

Spike train



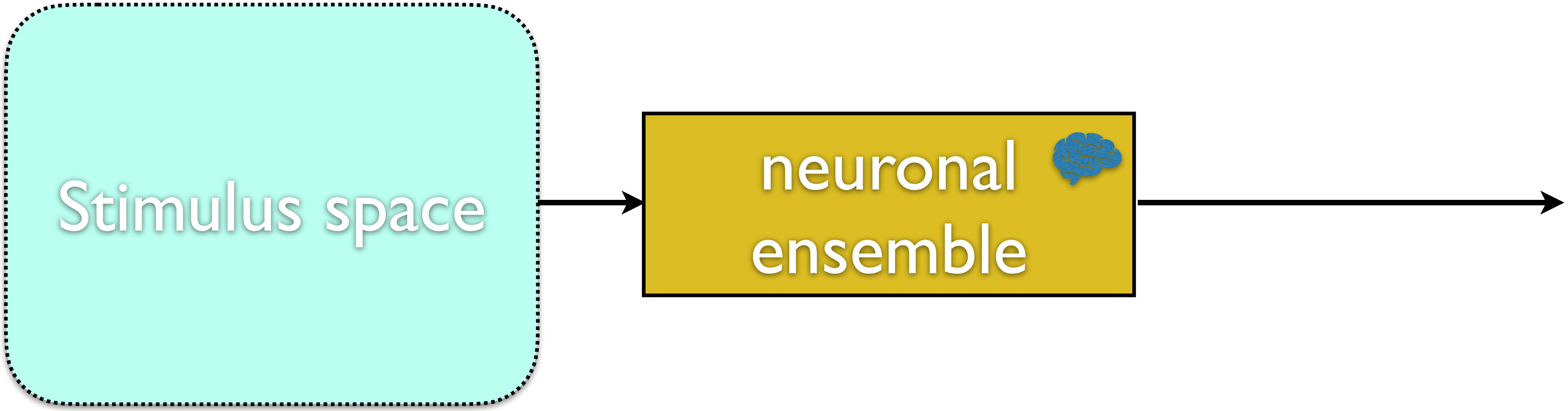
One basic setup

Stimulus could be visual, auditive etc

Neuronal ensemble could be in visual processing area (V1), or it could be in hippocampus (spatial memory), etc

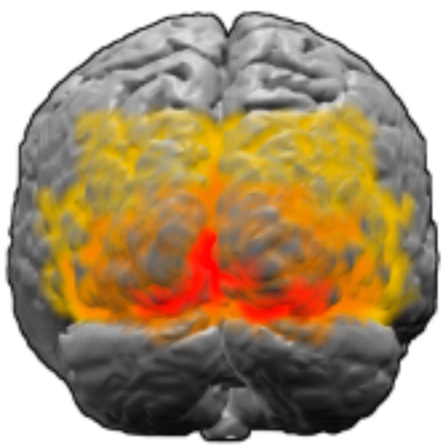
A general question: Is the neuronal ensemble able to encode for the information contained in the <u>stimulus space</u>?

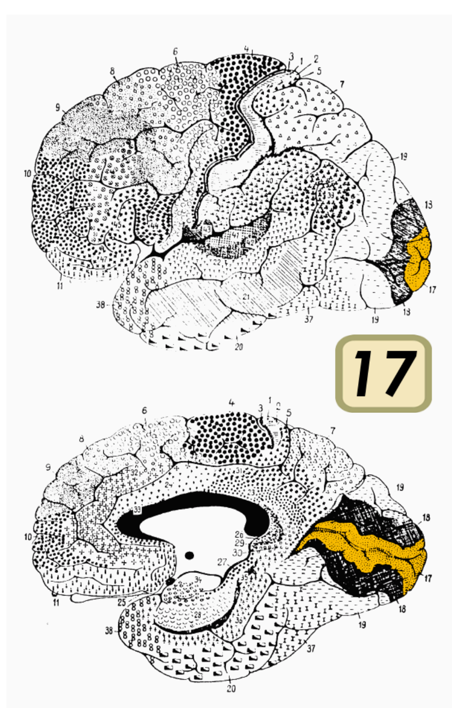
Stimulus space



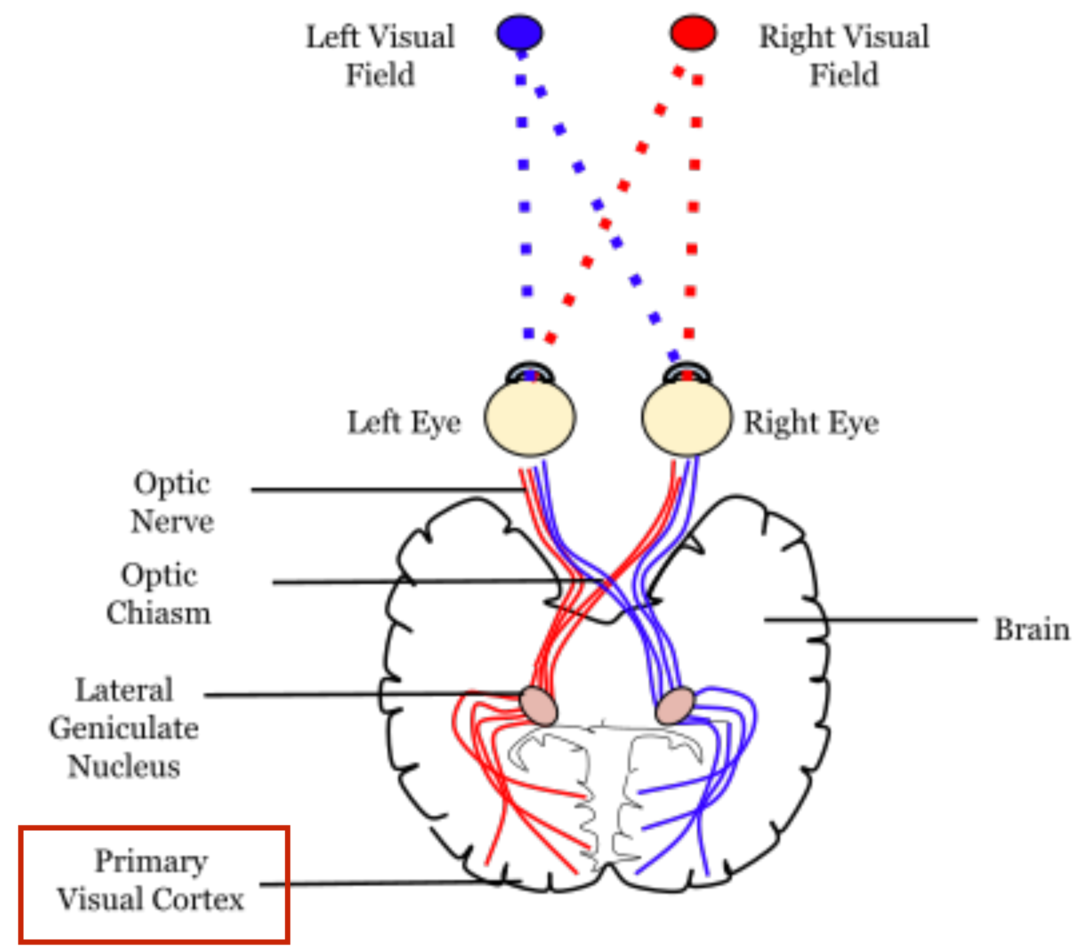
Example: V1

primary visual cortex



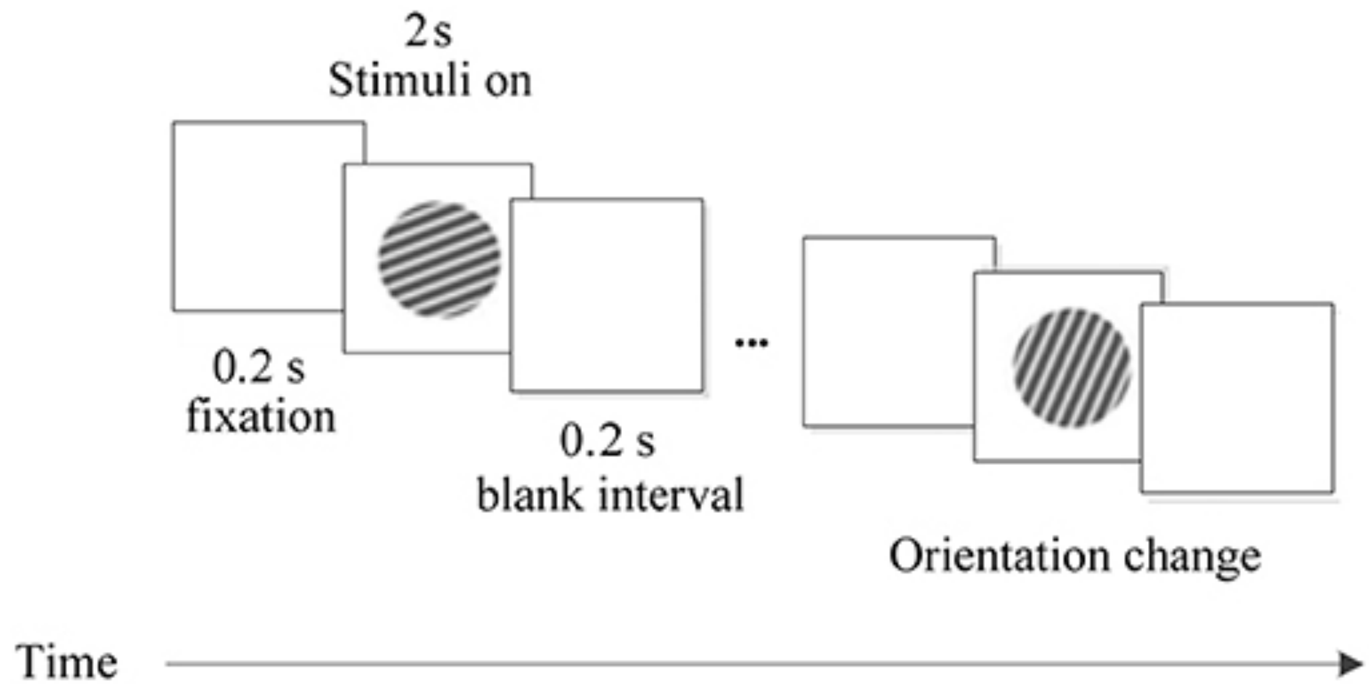


https://en.wikipedia.org/wiki/Visual_cortex

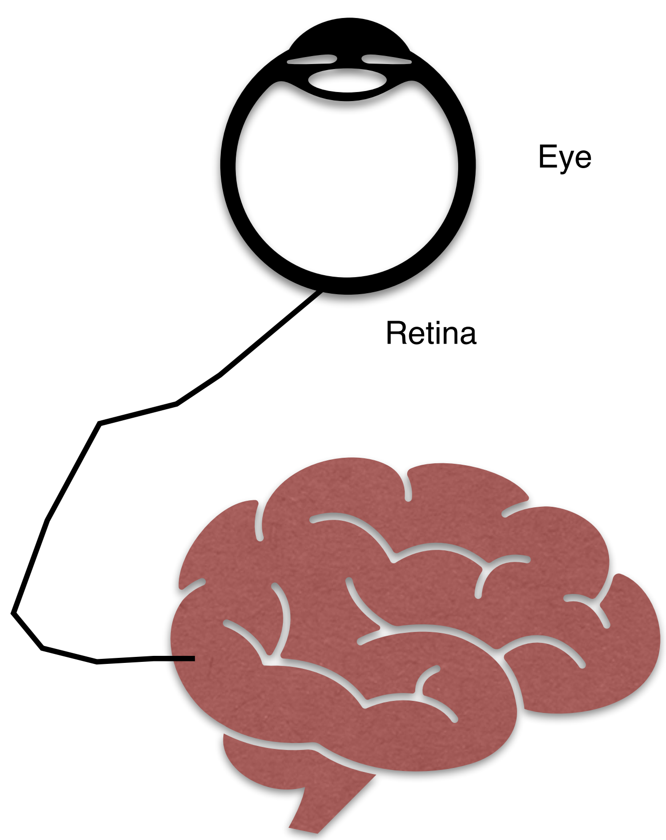


V1

The tuning properties of V1 neurons (what the neurons respond to) differ greatly over time. Early in time (40 ms and further) individual V1 neurons have strong tuning to a small set of stimuli. That is, the neuronal responses can discriminate small changes in visual orientations, spatial frequencies and colors (as in the optical system of a camera obscura, but projected onto retinal cells of the eye, which are clustered in density and fineness).[13] Each V1 neuron propagates a signal from a retinal cell, in continuation. Furthermore, individual V1 neurons in humans and animals with binocular vision have ocular dominance, namely tuning to one of the two eyes. In V1, and primary sensory cortex in general, neurons with similar tuning properties tend to cluster together as cortical columns. David Hubel and Torsten Wiesel proposed the classic ice-cube organization model of cortical columns for two tuning properties: ocular dominance and orientation. However, this model cannot accommodate the color, spatial frequency and many other features to which neurons are tuned[*citation needed*]. The exact organization of all these cortical columns within V1 remains a hot topic of current research. The mathematical modeling of this function has been compared to Gabor transforms.

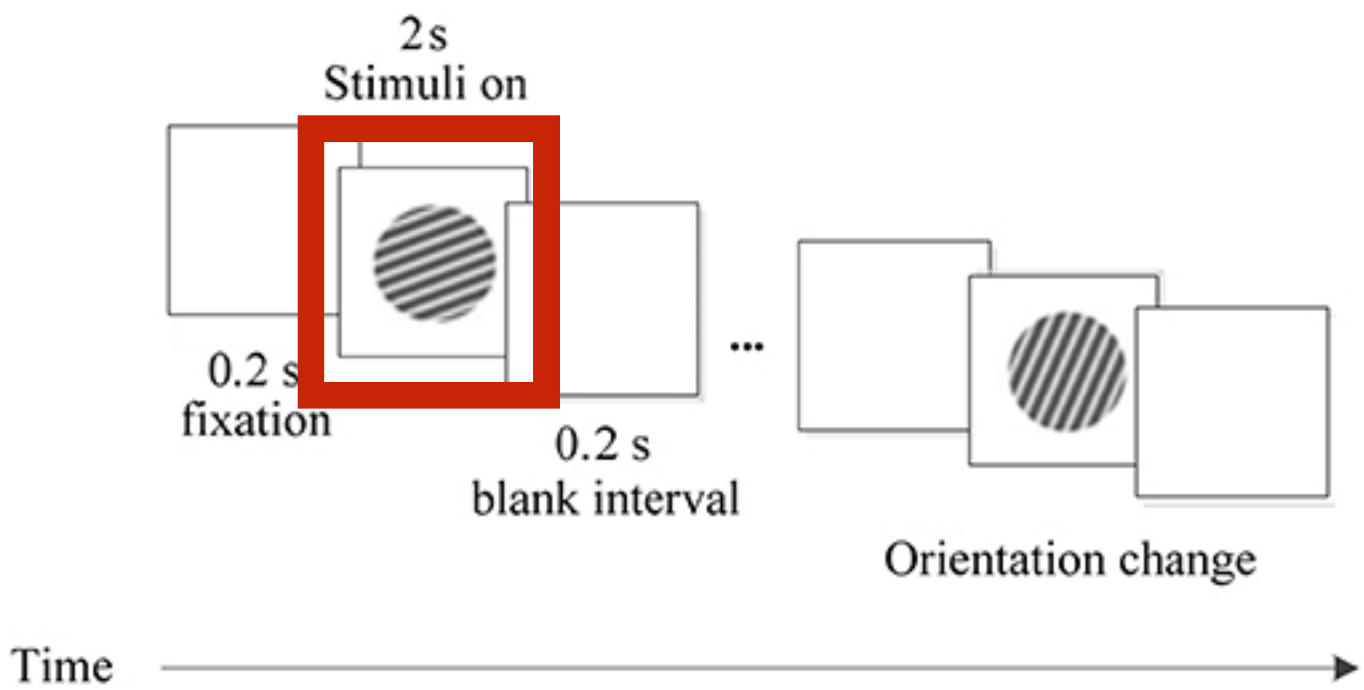


https://www.frontiersin.org/articles/10.3389/fncom.2019.00047/full

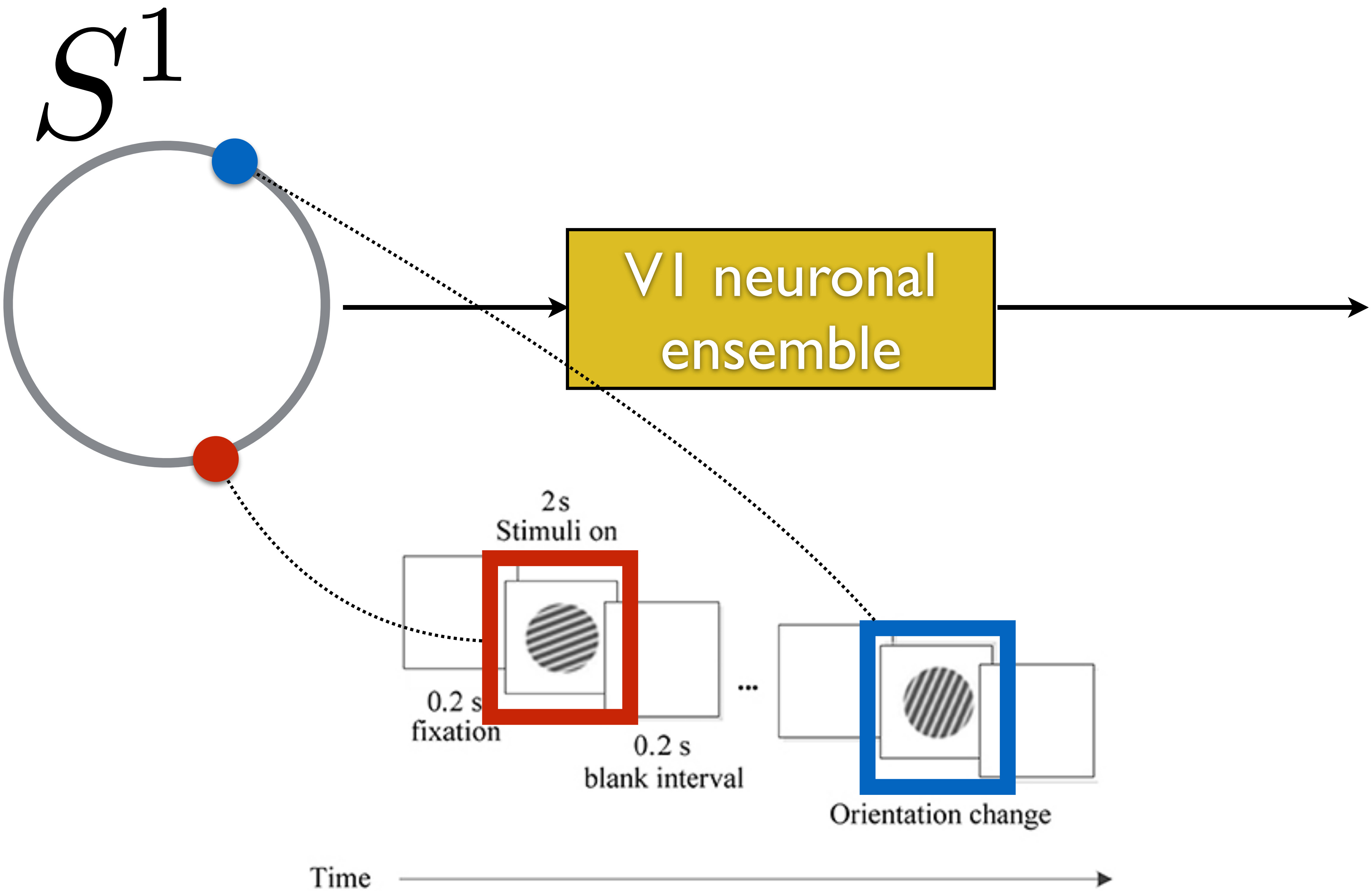


A question

- What is the ‘stimulus space’ corresponding to all possible images like:



Stimulus space



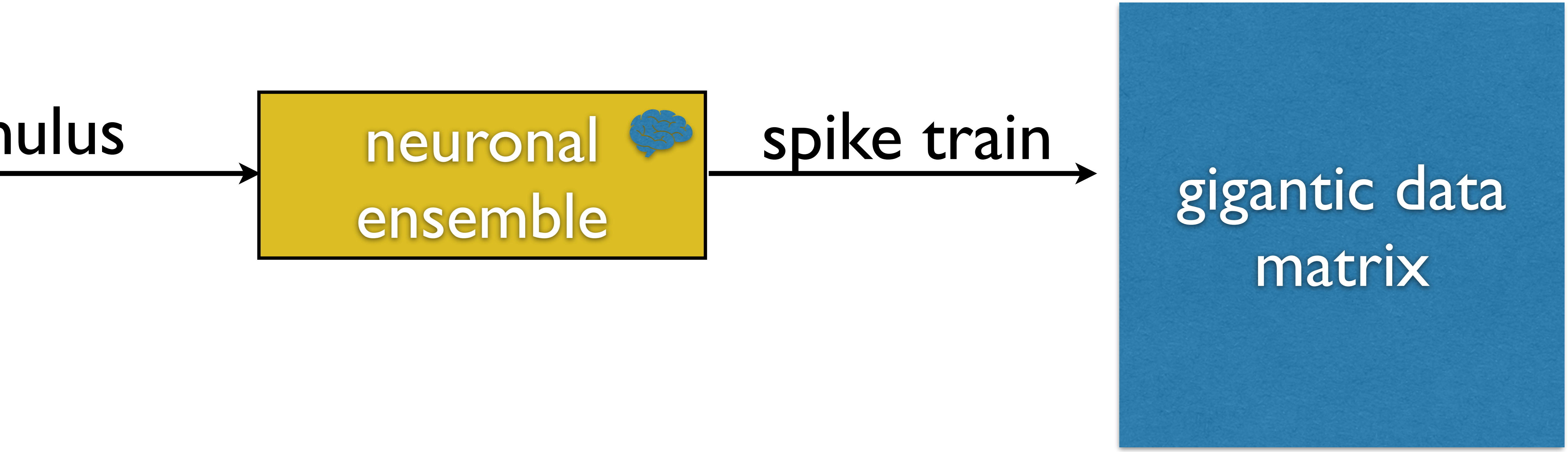
One basic setup

Stimulus could be visual, auditive etc

Neuronal ensemble could be in visual processing area (V1), or it could be in hippocampus (spatial memory), etc

A general question: Is the neuronal ensemble able to encode for the information contained in the <u>stimulus space</u>?

<u>In the case of V1 we'd want to check whether we can *"detect the presence of S^1"* in the spike train!</u>



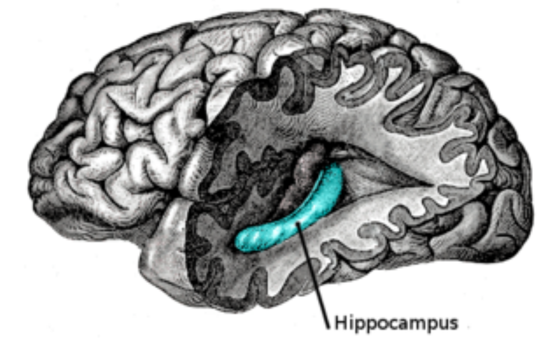
time X number of neurons

What is TDA?

- **T**opological **D**ata **A**nalysis is a collection of tools/concepts which permit extracting information from datasets. These tools often use ideas from topology, a branch of mathematics dealing with "the characterization of shape".

Terminology/Basics

- simplicial complexes
- Betti numbers
- persistence barcodes
- Covers & the nerve theorem
- …



An introduction to Persistent Homology via an example from neuroscience

Facundo Mémoli
memoli@math.osu.edu

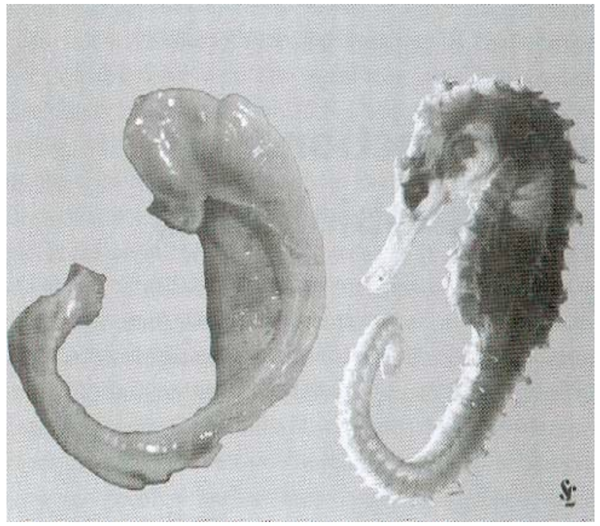
Some references

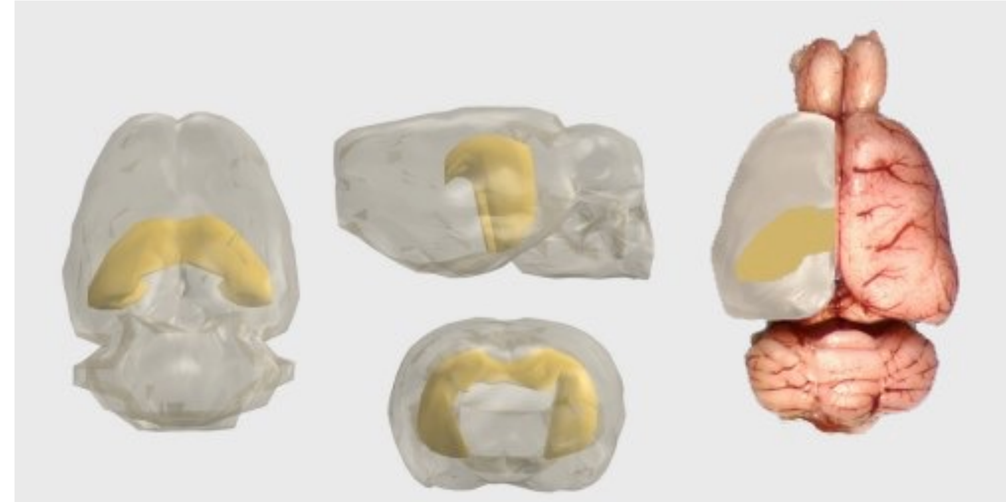
- Curto C, Itskov V (2008) **Cell Groups Reveal Structure of Stimulus Space.** PLoS Comput Biol 4(10): e1000205. doi:10.1371/journal.pcbi.1000205

- Dabaghian Y, Mémoli F, Frank L, Carlsson G (2012) **A Topological Paradigm for Hippocampal Spatial Map Formation Using Persistent Homology.** PLoS Comput Biol 8(8): e1002581. doi:10.1371/journal.pcbi.1002581

- Carina Curto, Pennsylvania State University: **What can topology tell us about the neural code?** Arxiv. AMS lecture, 2015.

- Chad Giusti, Robert Ghrist, Danielle S. Bassett. **Two's company, three (or more) is a simplex. Algebraic-topological tools for understanding higher-order structure in neural data.** Journal of Computational Neuroscience, 2016.





- **Hippocampus** is an area of the brain that has been associated with **spatial memories**.

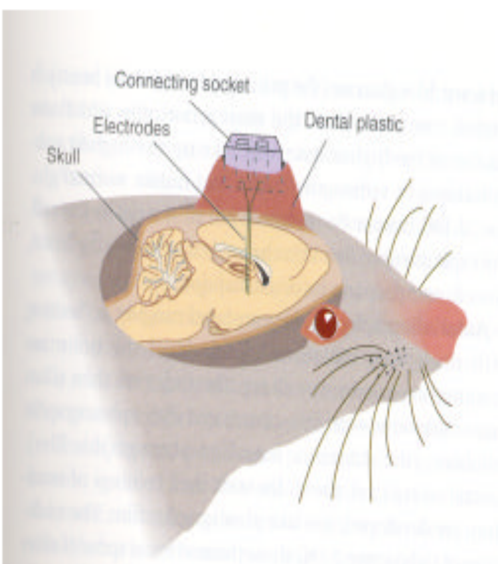
- It has a shape that reminded early anatomists of a **seahorse** and hence the name. (In greek: *hippos* = “horse”, *kampos* = “sea monster”)

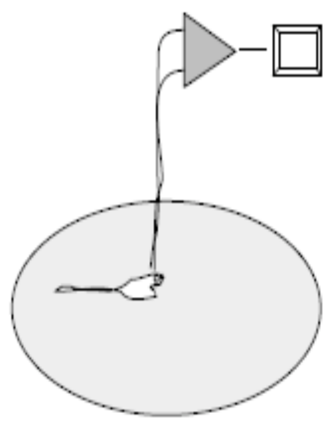
- hippocampus of **rats** is the one most studied.

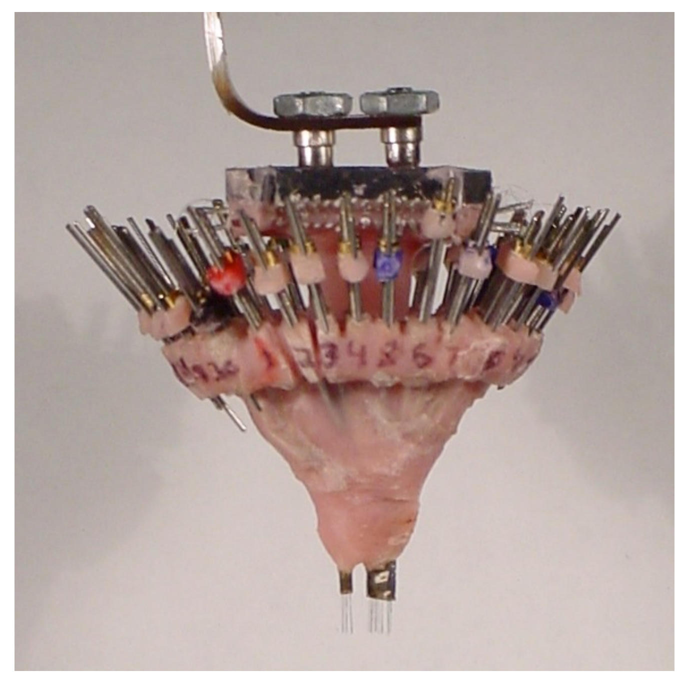
- If hippocampus is **damaged**, it has been observed that animal loses full ability to solve **spatial navigation** tasks.

- It has been established that **firing activity** of rat hippocampus cells has **correlation** with **current spatial location** (O'Keefe and Dostrovsky (1971)).

Recording spikes from hippocampus cells

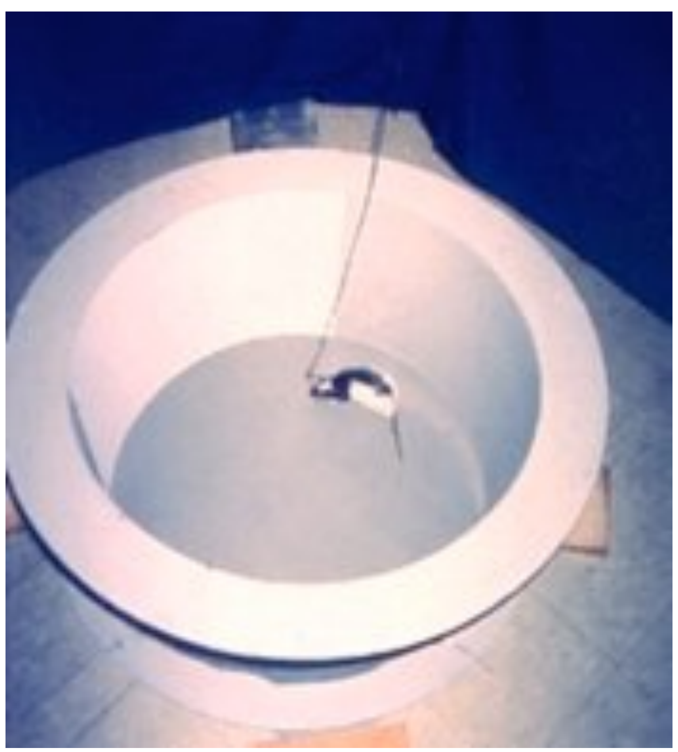




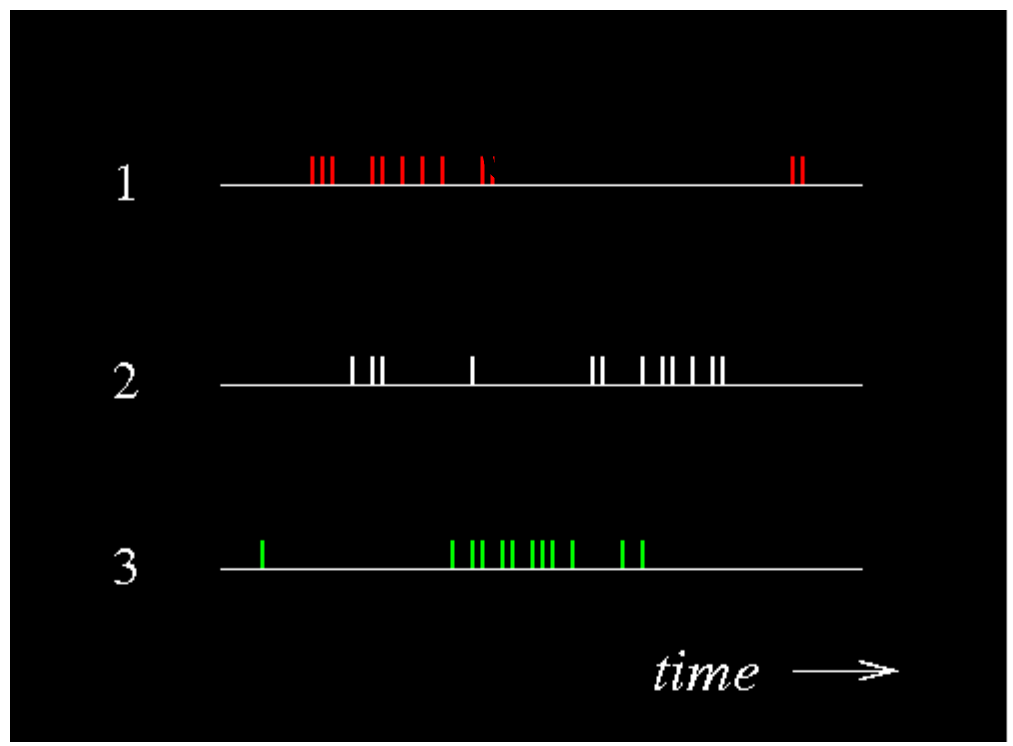


electrode 'hat'

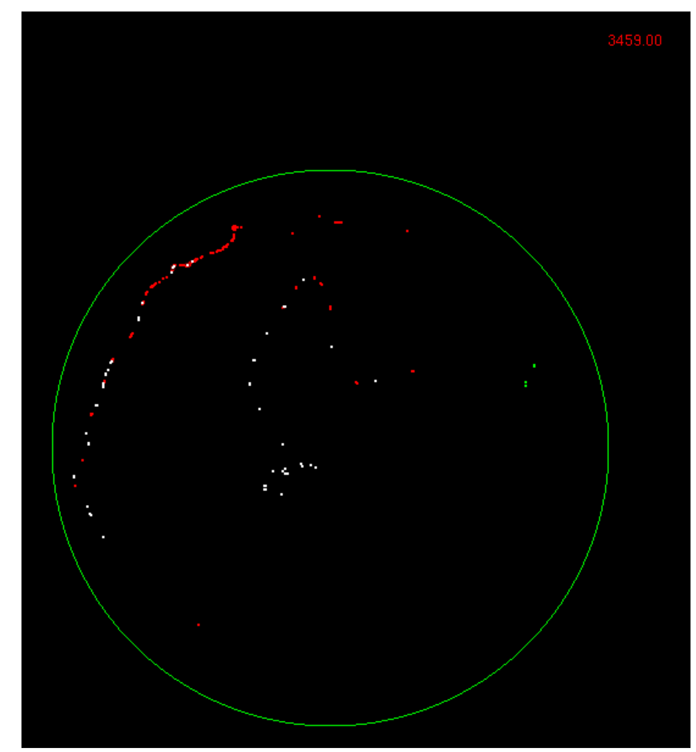
Circular **Arena**



arena



spike trains

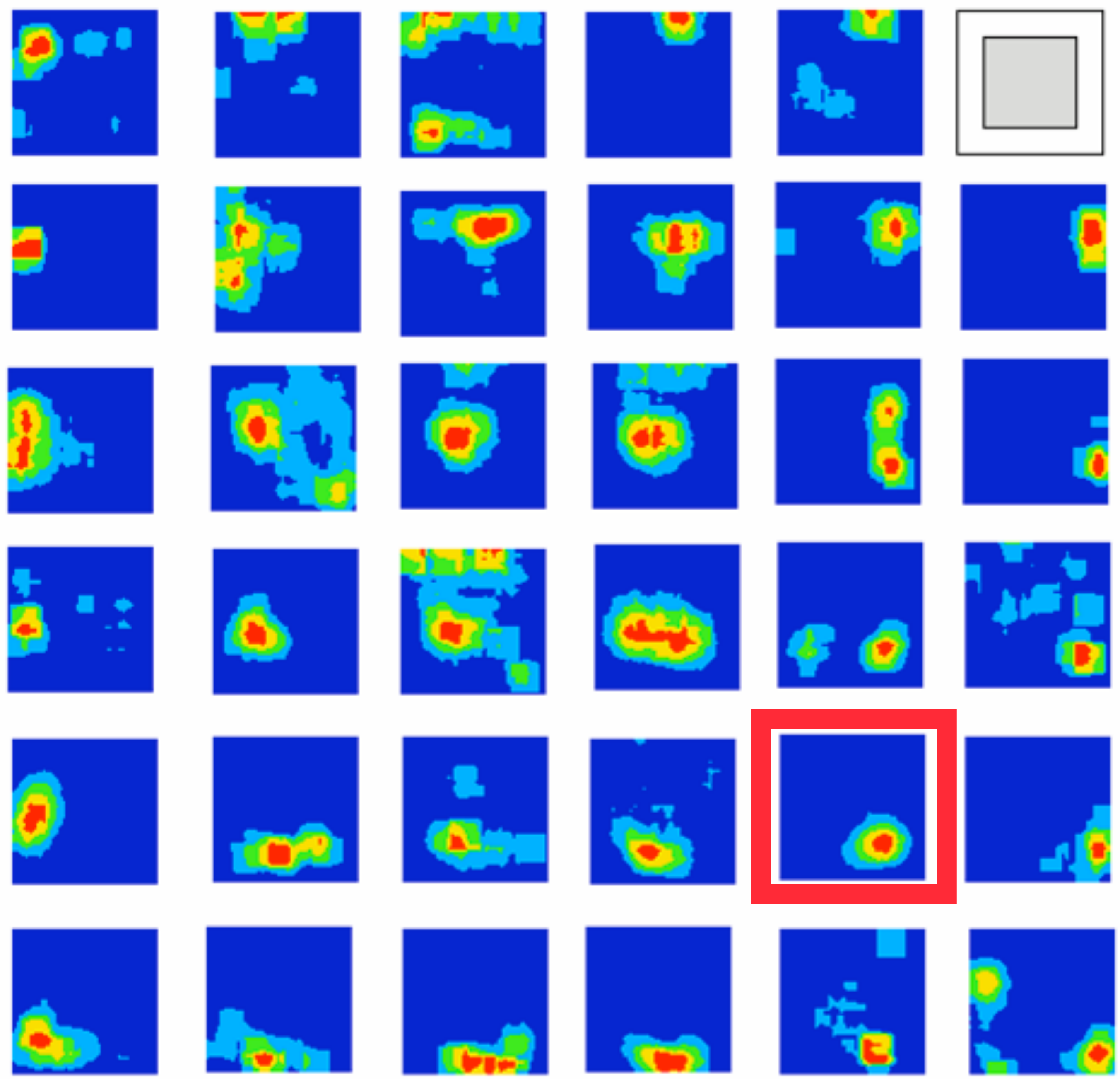


- Specifically, a given **cell** becomes active **only** when rat is located in a relatively **small area** of environment that it considers to be familiar.
  - Hence these cells are referred to as **Place Cells (PC)**
  - and the areas where they are active are called **Place Fields** (areas highlighted by these neurons) **(PF)**

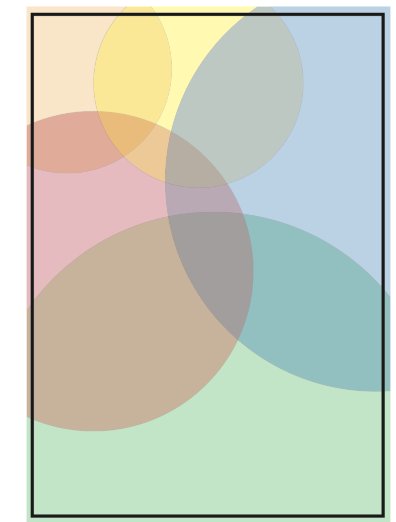
Some data about PFs

- When rat first enters a **new arena**, it takes **4-8 mins** for the PFs to be **created/stabilized.**
- PFs tend to be highly stable over **repeated** exposures to the **same** environment.
- PFs can have very different shapes. Typical computational model is non-isotropic Gaussian. Typical size could be from **10-70 cm**, but this varies. Shape of PFs core is typically **convex**, and **elliptical**.
- Typical **peak firing rate** (height of PF profile) **20 hz.**
- PCs can be **directional**!
- Nearby PCs can have totally different PFs.

35 SIMULTANEOUSLY RECORDED PLACE CELLS



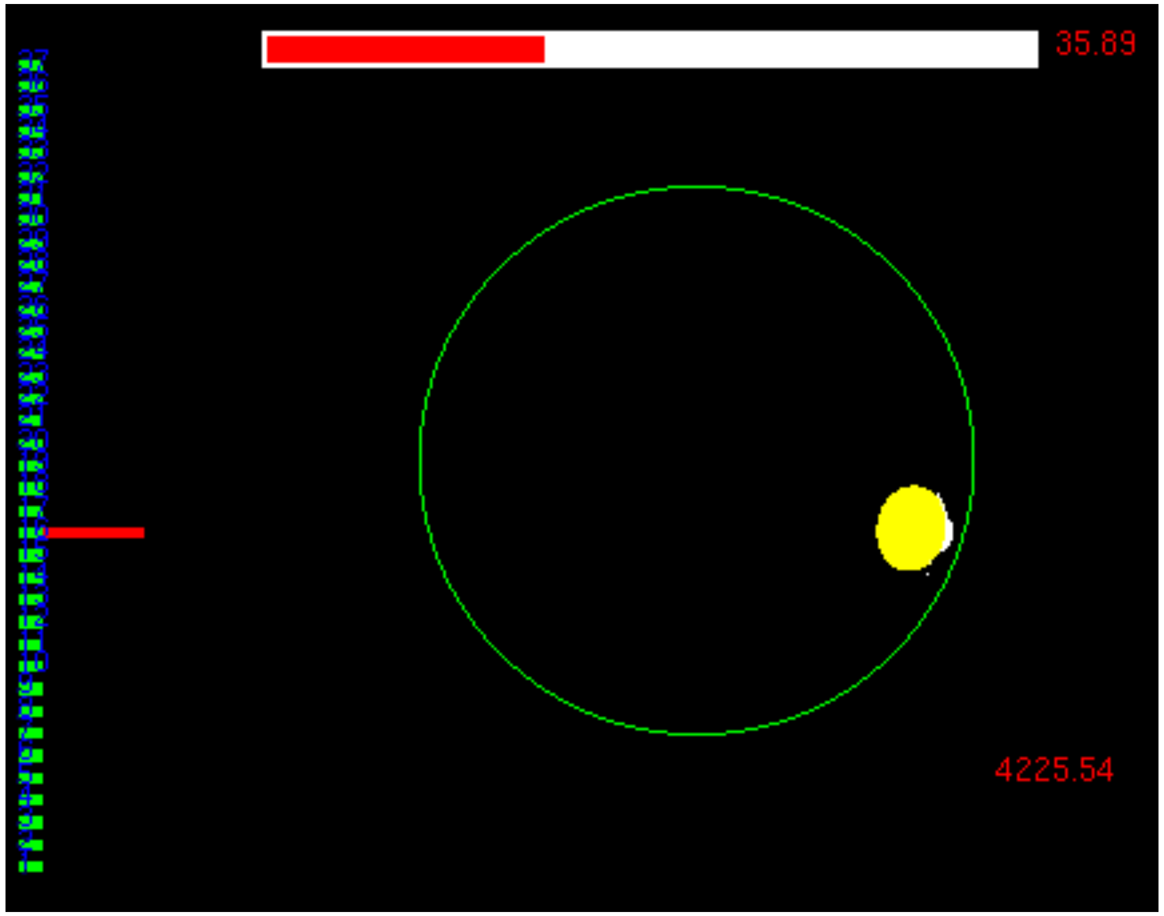
- Experimental evidence suggests that the collection of **PFs covers the whole environment that the rat considers familiar.** In this example: PF1, PF2, PF3, PF4, PF5



- It has been shown that if one assumes the knowledge of the position and size of 70--80 **PFs** in a 1m x 1m arena, then one can predict (with high accuracy) the **current location** of the rat at any time based on the **firing pattern** of the **PC ensemble.** (Wilson and McNaughton (1993) , Brown et al,

Space and trajectory reconstruction

- **Trajectory reconstruction:**
  - Estimate location and size of place fields.
  - Decode animal's position from new spike trains based on these estimates (Brown et. al 1998)
- The analysis is based on **external** geometric information
- What if the external "place" tags are removed?



Space Reconstruction Experiment (SRE)

- Assume rat is running around freely in a certain arena, away from experimenter/observer.
- The experimenter receives real time signals from electrodes that are implanted into rat's hippocampus and is free to analyze the information in any way in order to extract as much information as possible about geometry of the arena.
- **QUESTION**: how much will the experimenter be able to de deduce about the **geometrical/spatial** properties of the environment?
  - more optimistically: can this reveal how the brain may be representing space?

- From a **biological** point of view it seems that the question of what **information** one can gather **about the environment** should be approached in a **parsimonious** way in the sense that **first** one should

    - identify what is the **most robust** type of information that one could hope to capture.

    - check **stability** of this information to variations present in the biological system (**randomness** for example).

Space reconstruction experiment (SRE)

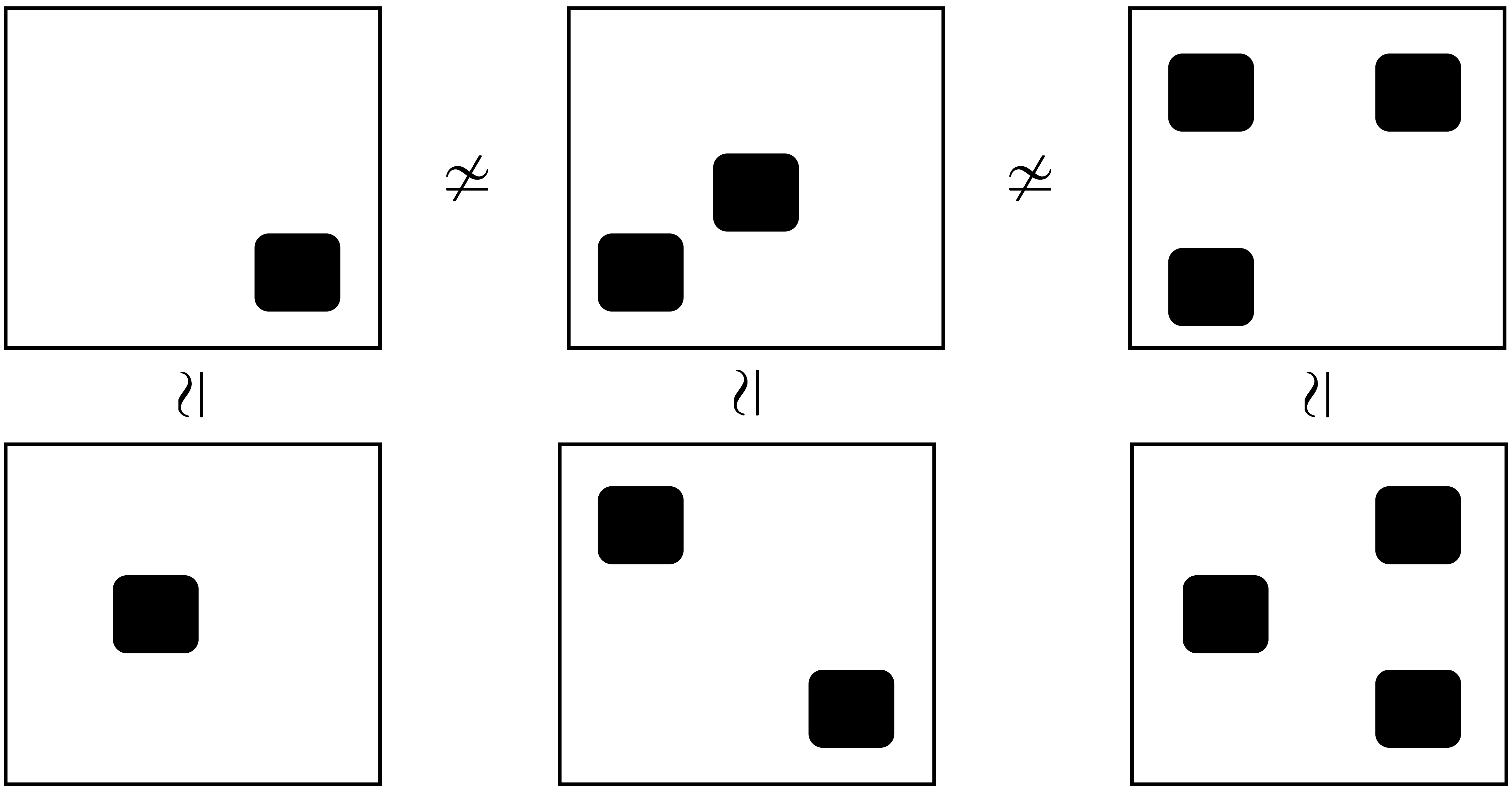
what can you say about **stimulus** by just looking at **ouput** (activity of ensembles of PCs)?

So from now on: we assume we only have access to **spike trains**

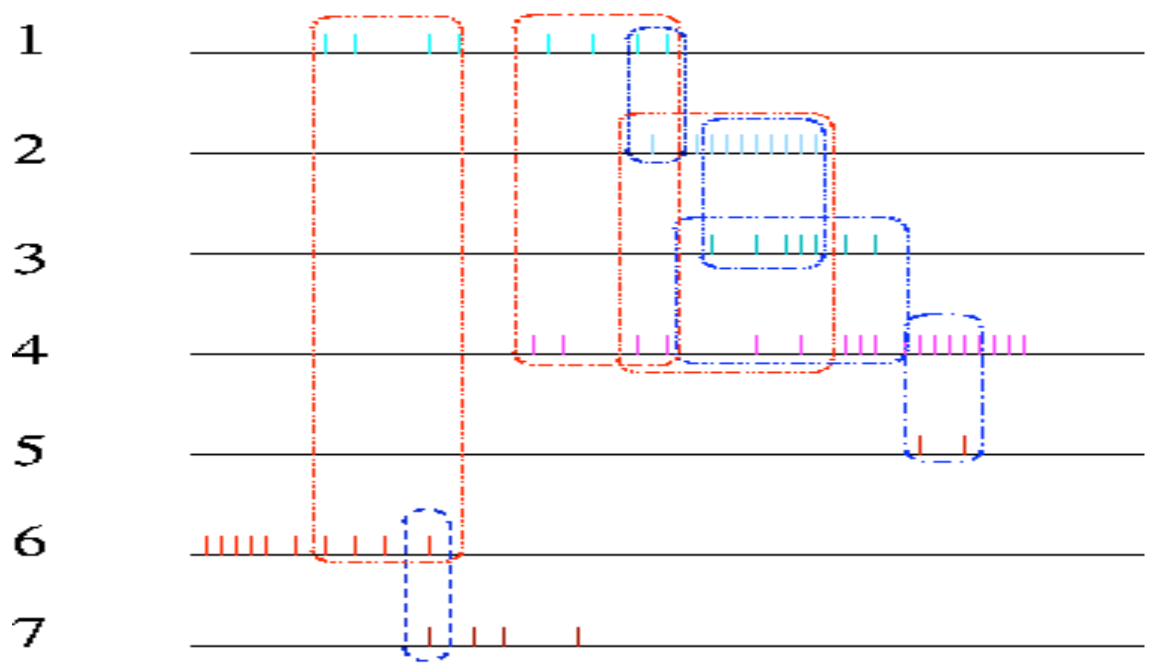
Moving on..

- What do **PC ensembles** encode for?
    - topology? metric??
  - type of information about the arena that we will try to detect: **topological**.
  - How can we do this? There is a natural construction..

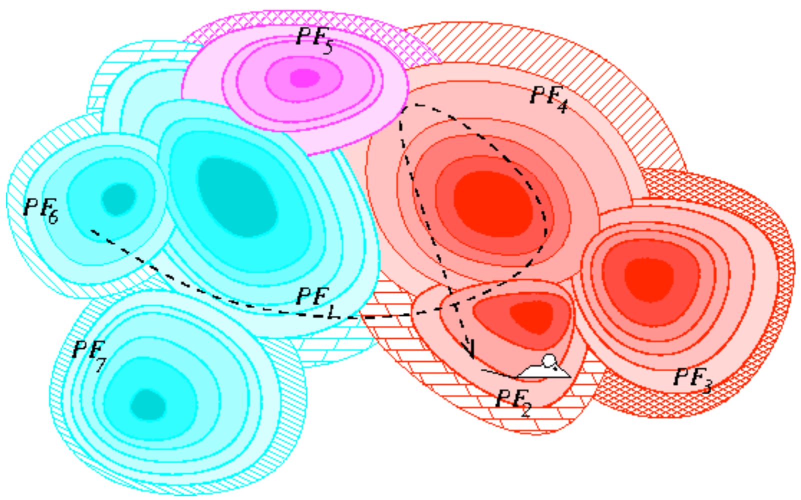
Different arenas



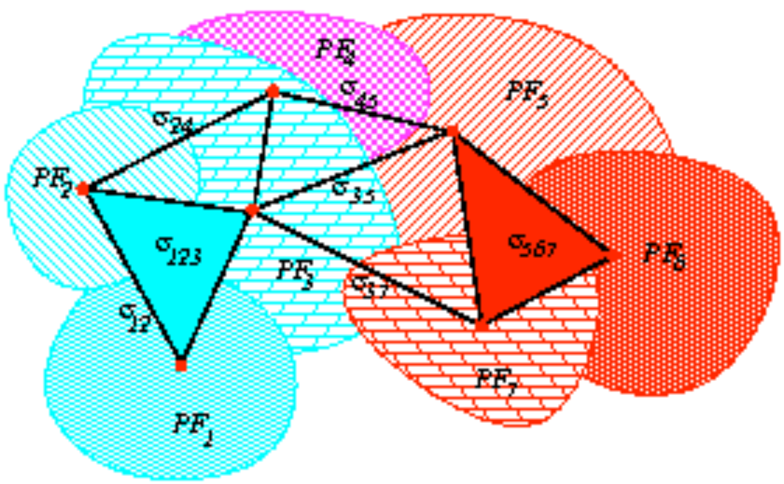
different number of obstacles…



Spike trains, temporal overlap

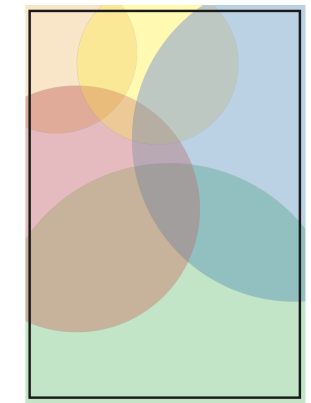


Place fields,
Spatial overlap

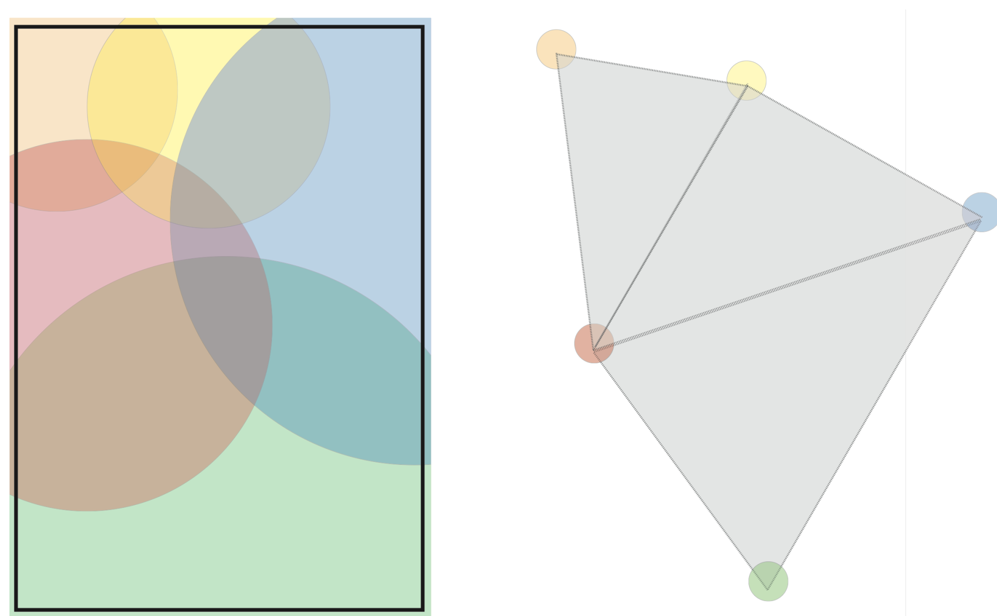


If we had access to patterns of spatial overlap, we could invoke (or dream of invoking) the so called **Nerve Theorem**: the "topology" of the nerve complex associated to the PFs is the same as that of the arena. But we do not have access to PFs directly

Remember: PFs cover the arena —> Nerve Theorem



Theorem. *If $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ is an open cover of a compact space X such that every non-empty intersection of finitely many sets in $\mathcal{U}$ is contractible, then X and (the geometric realization of) $\mathcal{N}(\mathcal{U})$ are homotopy equivalent.*

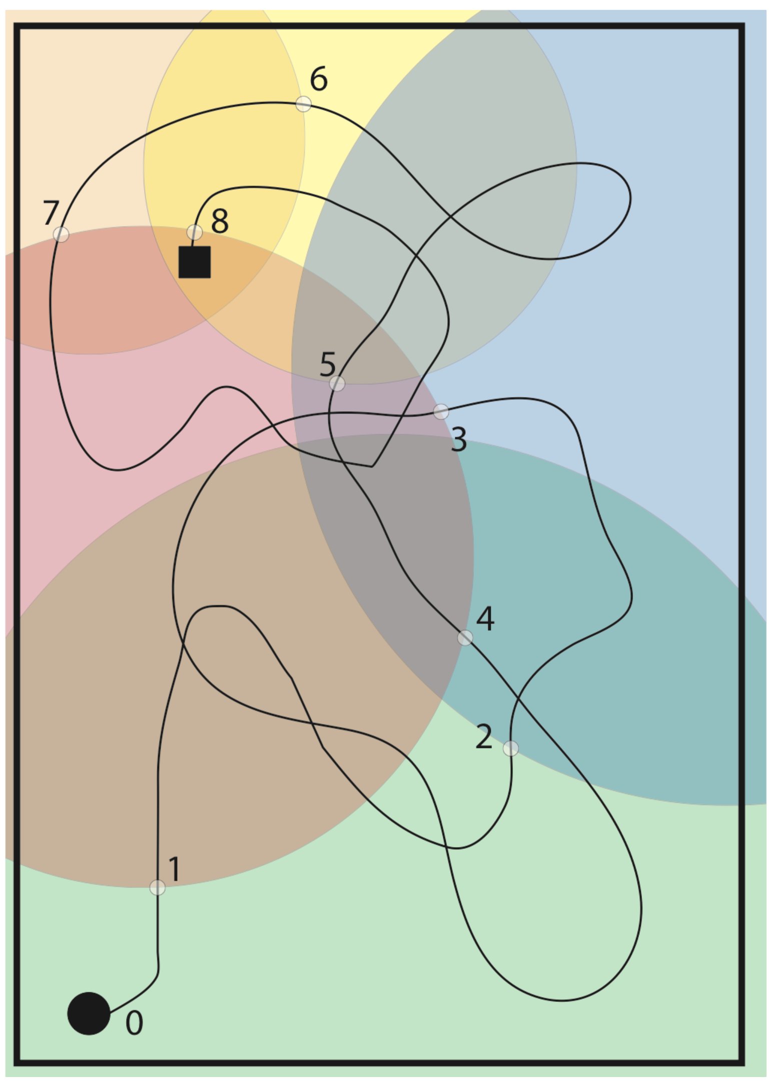


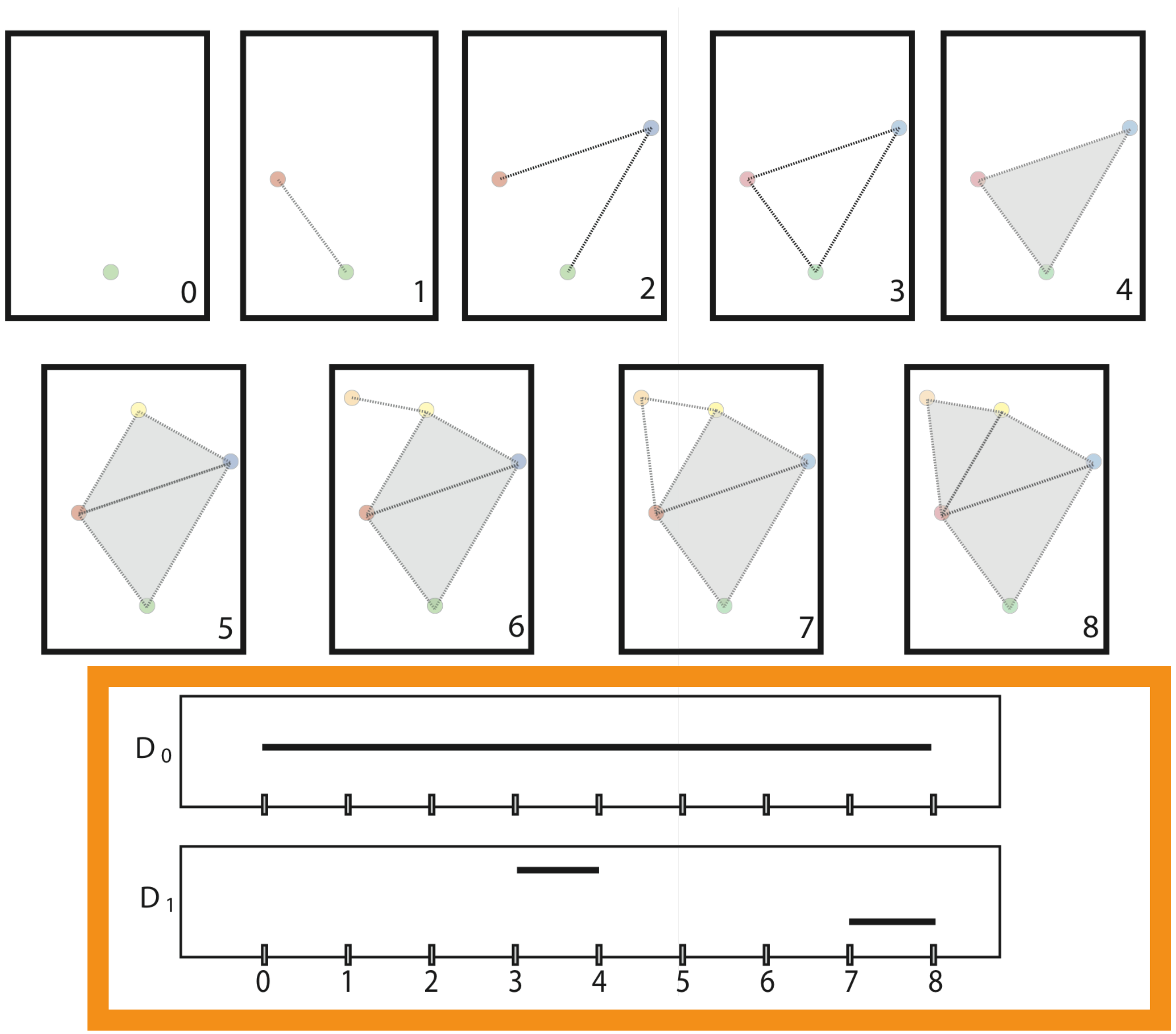
X and $\mathcal{U}$ $\qquad$ $\mathcal{N}(\mathcal{U})$ $\qquad$ $\mathcal{U} = \{$PF1, PF2, PF3, PF4, PF5$\}$

We do not have access to the cover itself ... we only know the spike trains!

So, from the spike trains, we can infer the structure of the Nerve complex.

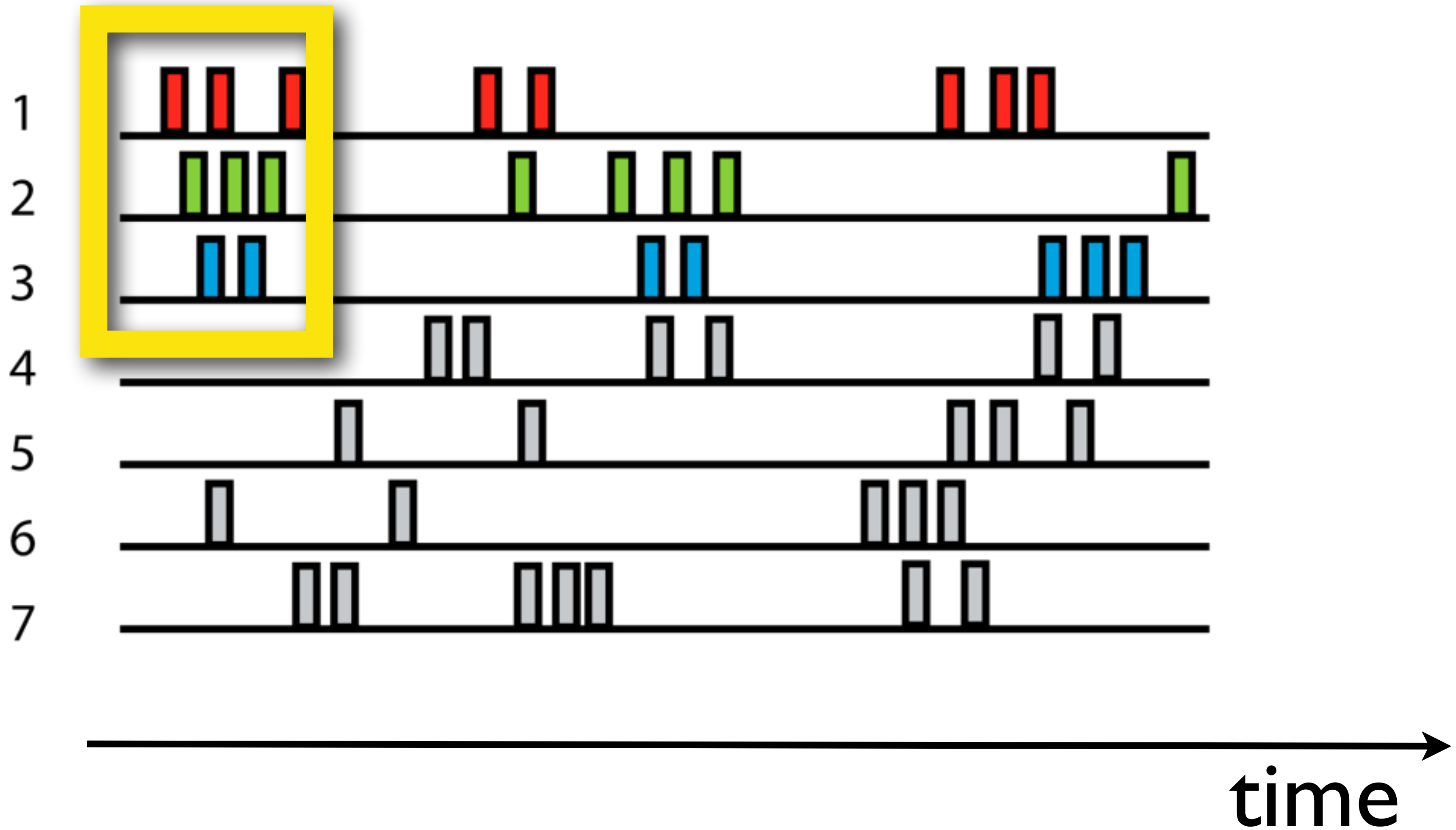
But we can do this in a time dependent way —> Learning Process

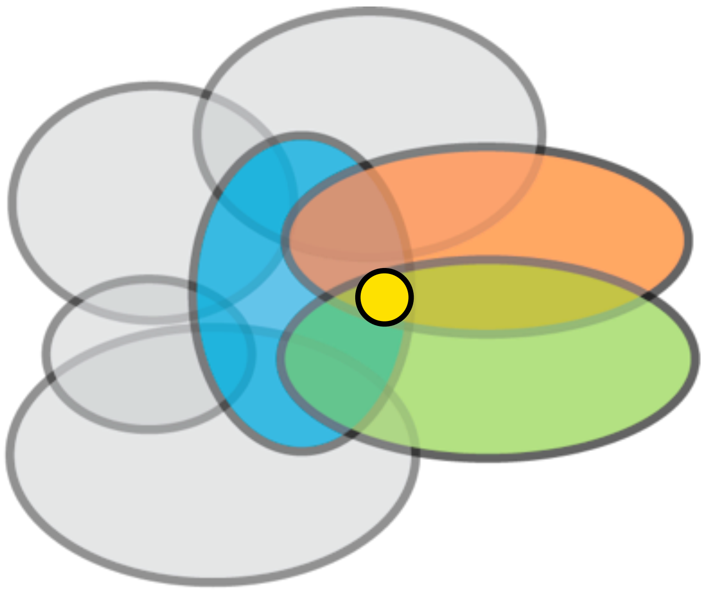




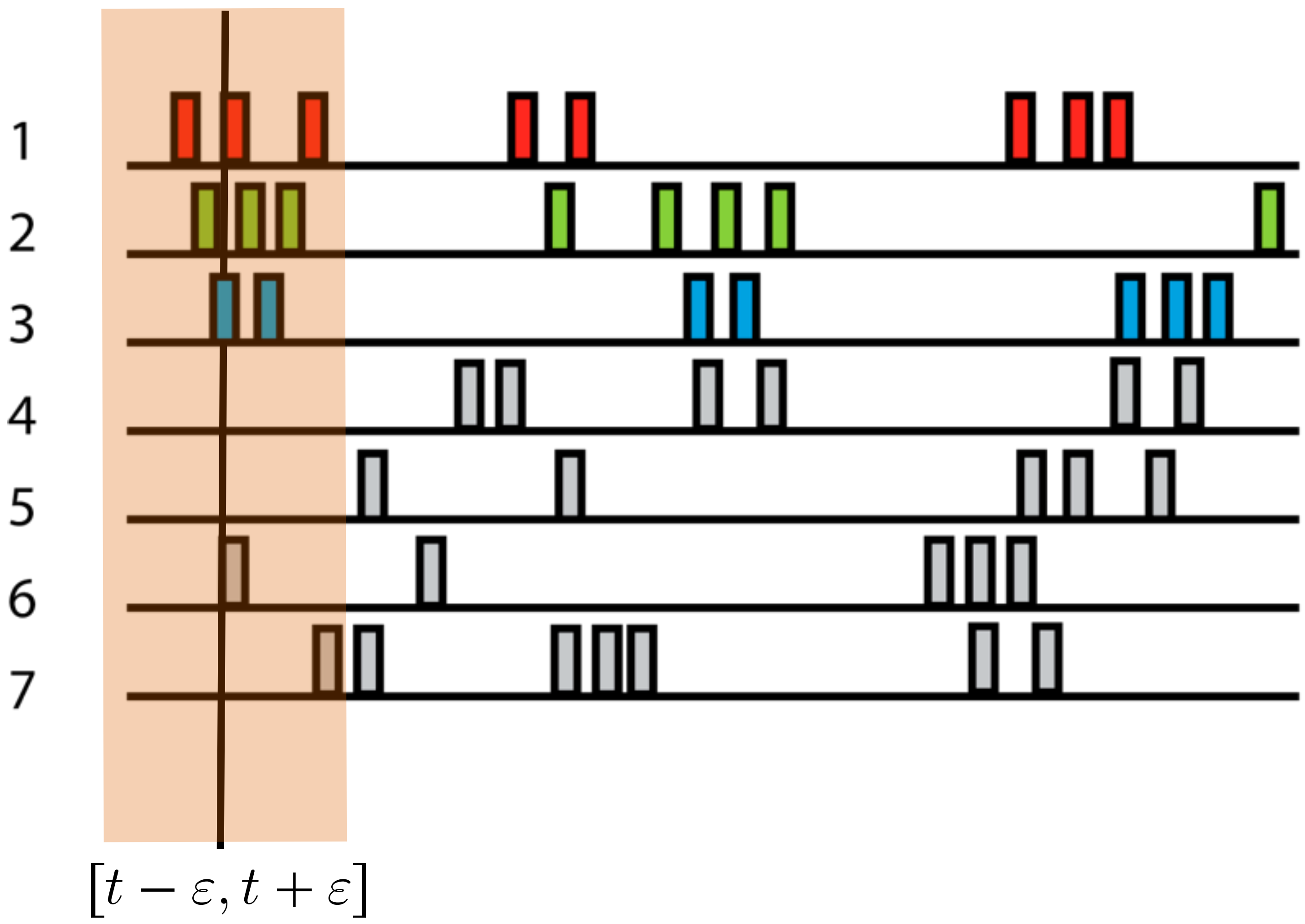
These are Betti Barcodes —> In general compute them with Persistent Homology techniques

Using spiking info





- We form a **vertex set** $V = \{v_1, v_2, \ldots, v_n\}$, with exactly one vertex for each **PF** we have.
- A **simplex** σ is a collection of different vertices $[v_0, v_1, \ldots, v_k]$. We say that σ is a k-simplex and that k is the **dimension** of σ.
- Let $K_d(V)$ denote the collection of all simplices that have dimension up to d.



Fix $\epsilon \geqslant 0$ and $n_0 \in \mathbb{N}$.

- One associates an entry time or **filtration time** $F(\sigma)$ to each simplex $\sigma = [v_0, v_1, \ldots, v_k]$:

  $$F([v_0, \ldots, v_k]) := \min \left\{ t \in [0, T] \,\text{s.t.}\, \left| s(v_i) \cap [t-\varepsilon, t+\varepsilon] \right| \geqslant n_0 \text{ for all } i = 0, 1, \ldots, k \right\}$$

  for some fixed $\varepsilon > 0$ and $n_0 \in \mathbb{N}$. Here $s(v_i)$ denotes the spike train observed at the i-th cell.

- One obtains a pair $\big(K_d(V), F\big)$, which is called a **filtration**.

- One could only consider the simplicial complex composed by those simplices σ of $K_d(V)$ for which $F(\sigma) < \infty$.

$$\big(K_d(V), F\big)$$

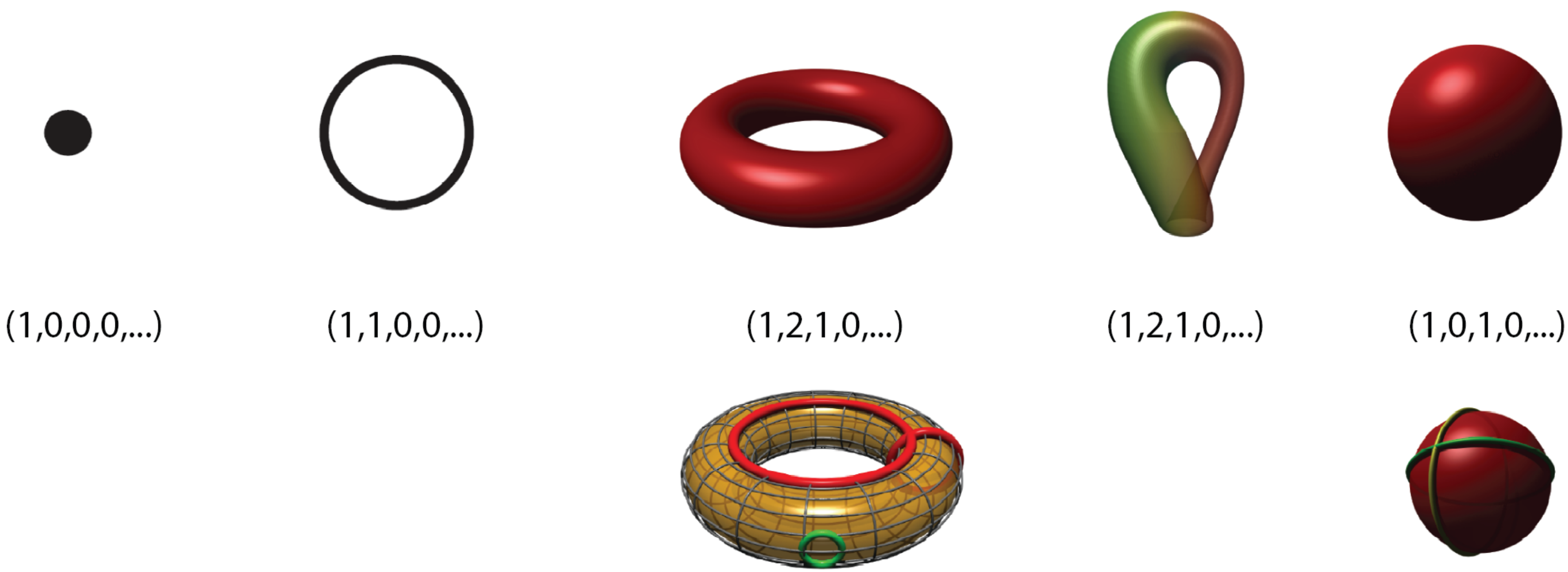
$$K_1 \subset K_2 \subset K_3 \subset \ldots \subset K_\ell$$

these are the objects of study of **persistent topology**

(Persistent) Topological Invariants

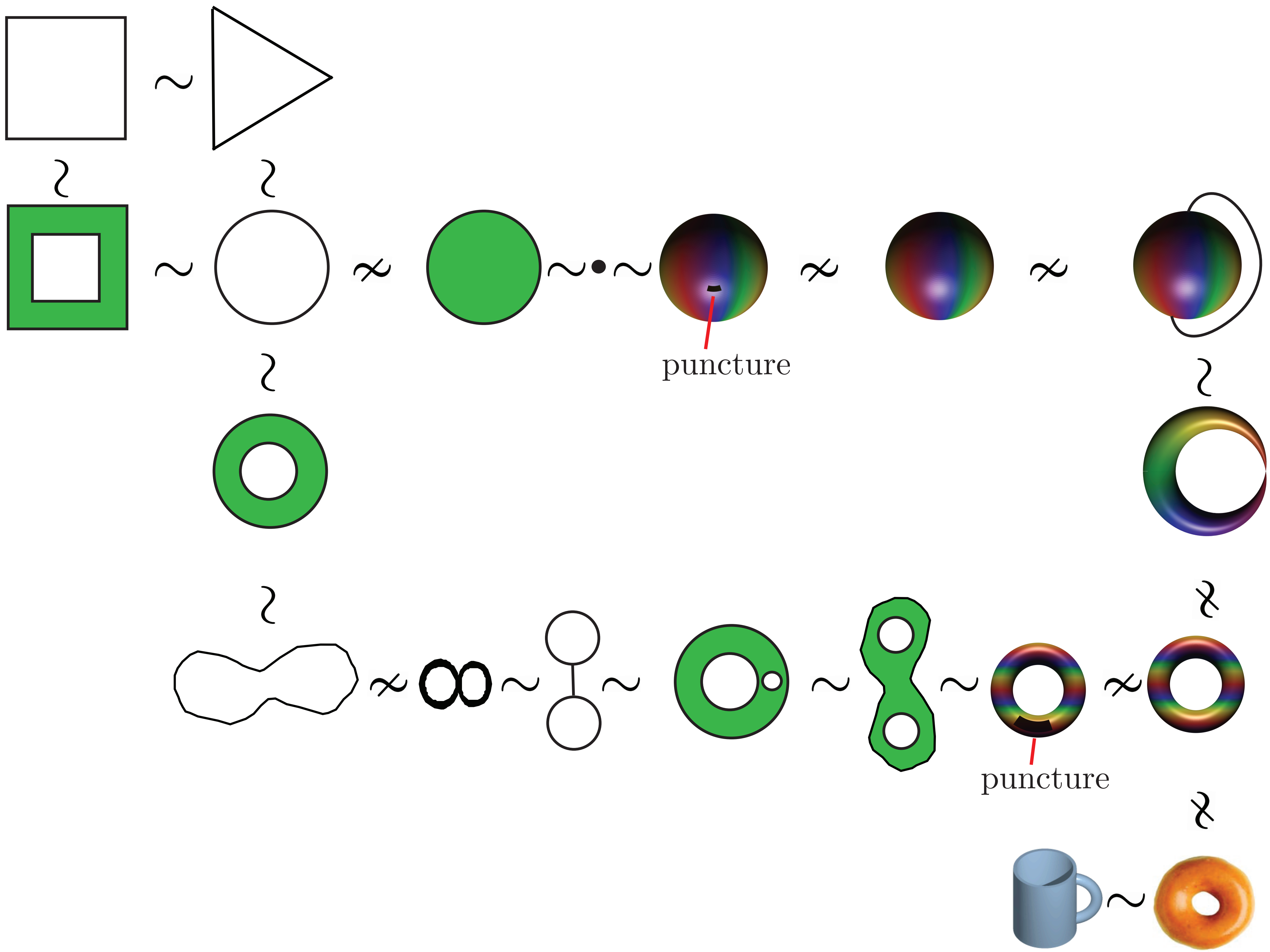
Topological invariants: Betti numbers

For a topological space X, the k-th **Betti number** $b_k(X)$ is a non-negative integer which measures the k-dimensional connectivity of a space X.



Betti numbers are **insensitive** to fairly general deformations of the underlying space!

Rubberband geometry: homotopy equivalence



Important concept:
every ‘shape’ has a **signature** in terms of Betti numbers.

$$\left(K_d(V), F\right)$$

$$K_1 \subset K_2 \subset K_3 \subset \ldots \subset K_\ell$$

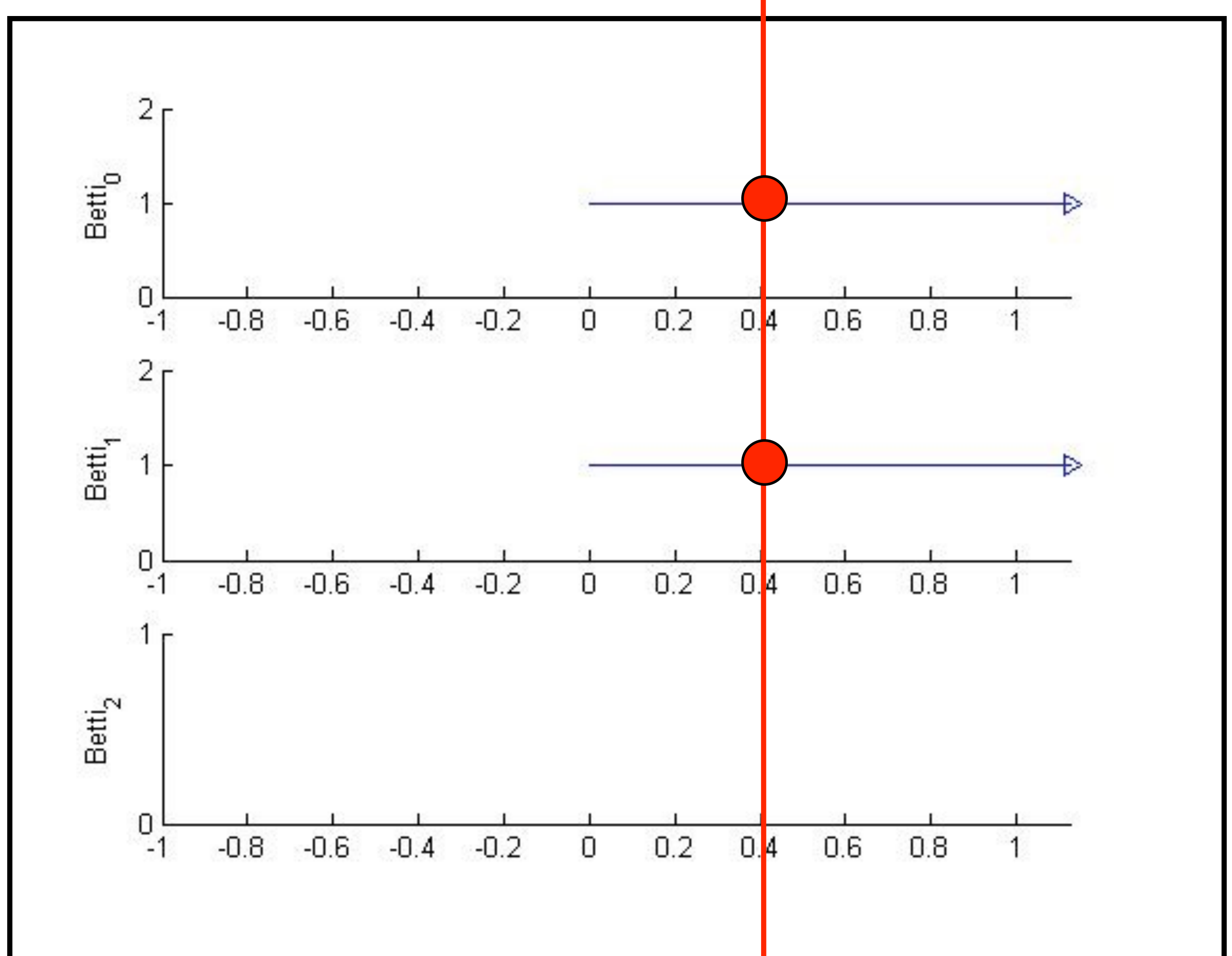
these are the objects of study of **persistent topology**

We don’t have a single topological space, but a family of them, indexed by “scale”

Scale Dependent b_k

- Actually, for each R in [0,1] one computes $b_k(X,R)$, that is, a <u>**scale dependent**</u> $b_k(X)$: look at filtrations and **persistence!**
- This leads to BARCODES:

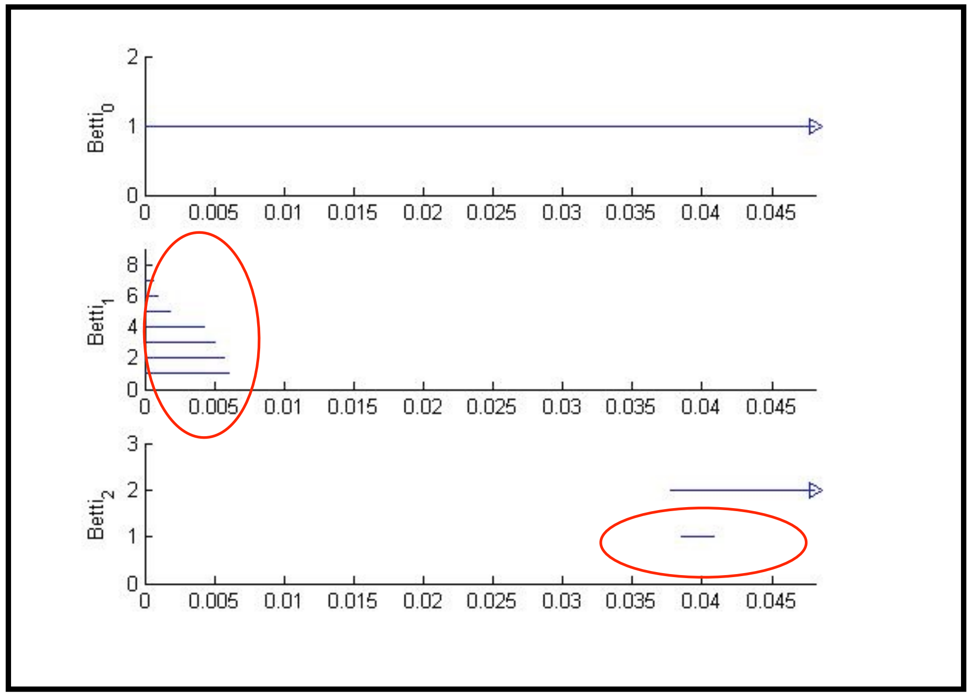
$$\left\{ \begin{array}{l} b_0(0.4) = 1, \\ b_1(0.4) = 1, \\ b_2(0.4) = 0 \end{array} \right.$$



Barcode for S^1

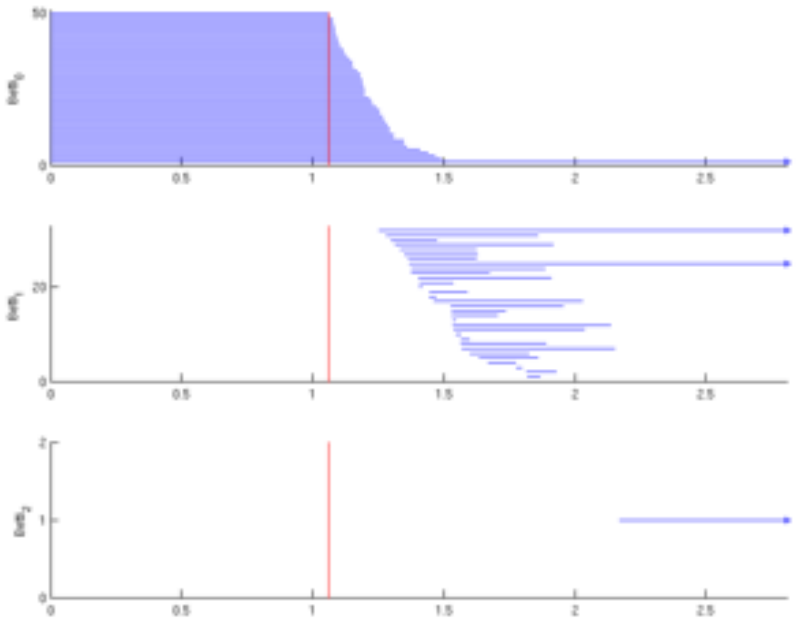
- Long lines are associated to ‘<u>persistent</u>’ features in the dataset– **dominant** features.

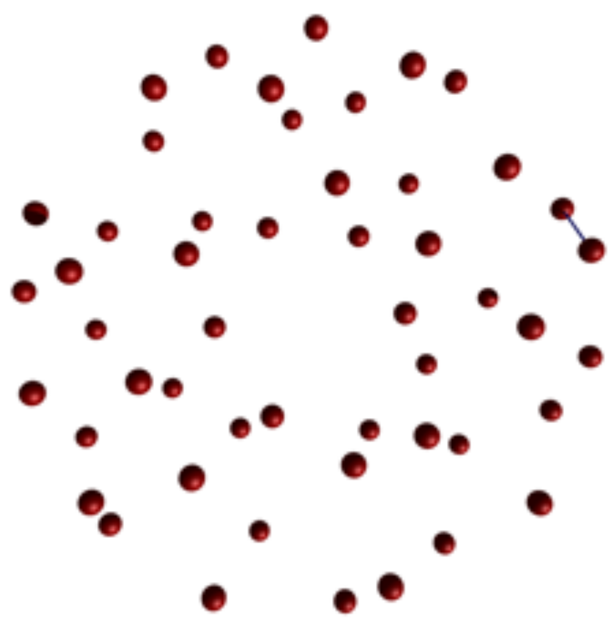
- Short lines indicate small features/**noise**.

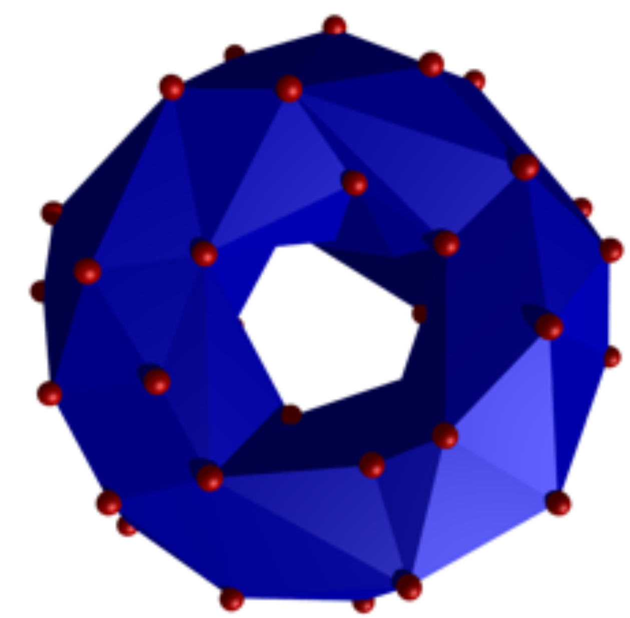


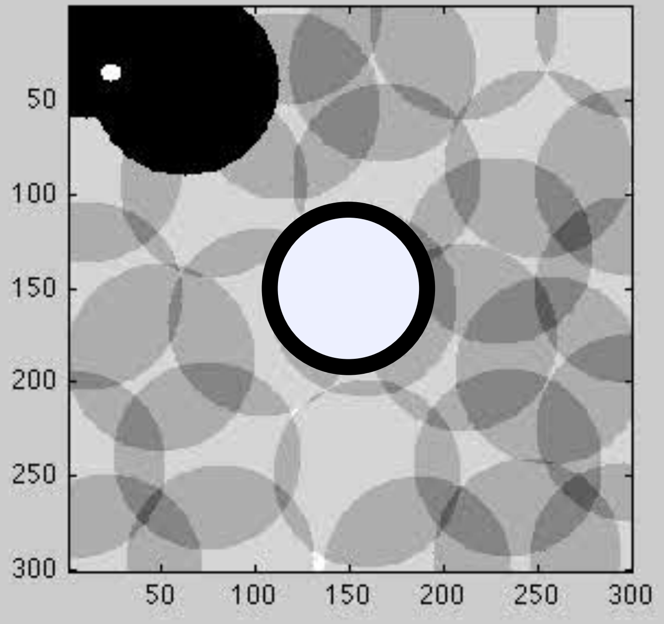
<u>Example</u>: noisy sphere

Example: Rips complex

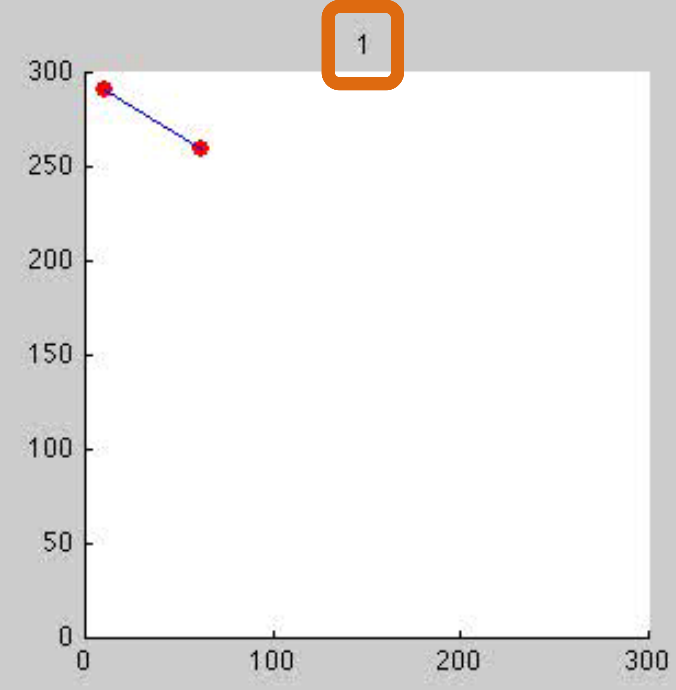


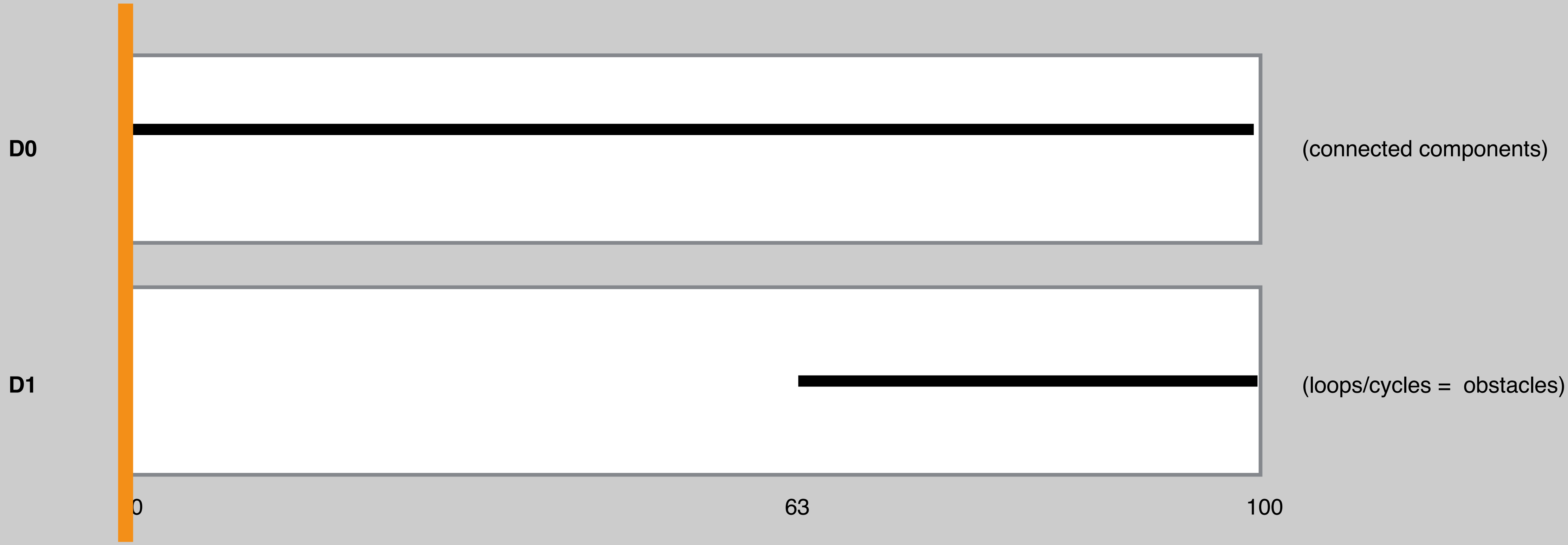






Arena with 1 obstacle





In this case, the **learning time** is **63**: **first time** after which the **hippocampal network** correctly encoded the topology of the arena

Implementation: **java-plex**

- Barcodes arise from filtrations.

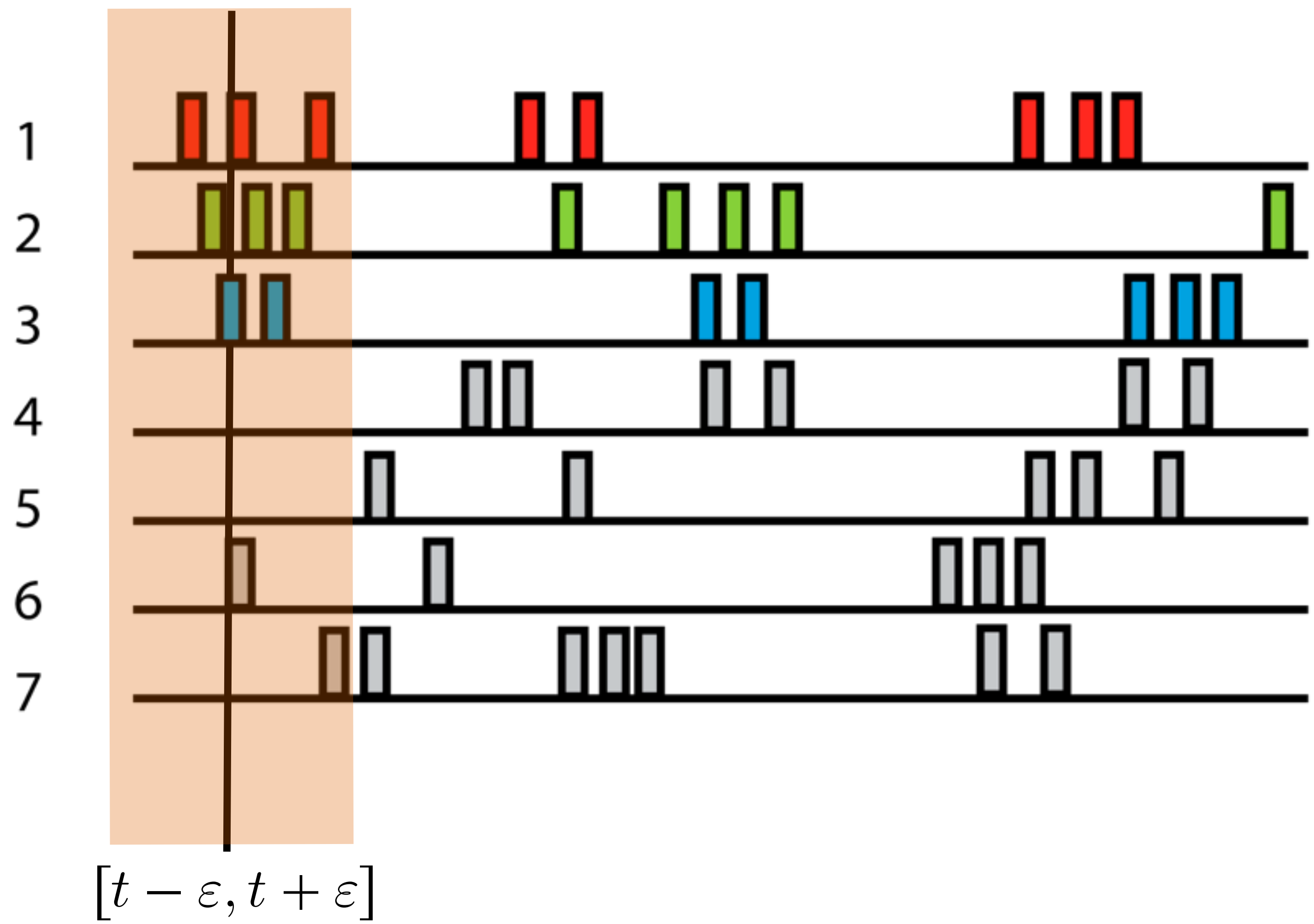
- Use software javaPlex to compute them (polynomial cost):

http://appliedtopology.github.io/javaplex/

H. Edelsbrunner, D. Letscher and A. Zomorodian, Topological Persistence and Simplification, FOCS 2000.

Experiments:

- working with simulated data (Poisson model)
- we generate arenas with different number of '**features**' (holes) and generate a distribution of PF centers and a **trajectory that remains fixed**.
- vary size **σ** and peak firing rate **A** of PF profiles. (chosen according to some probability distribution.)
- For σ and A fixed we run simulation many times. Each time we obtain a multi-electrode spike train (MEST).
- For each MEST, compute barcode diagrams.
- compute statistics on results.

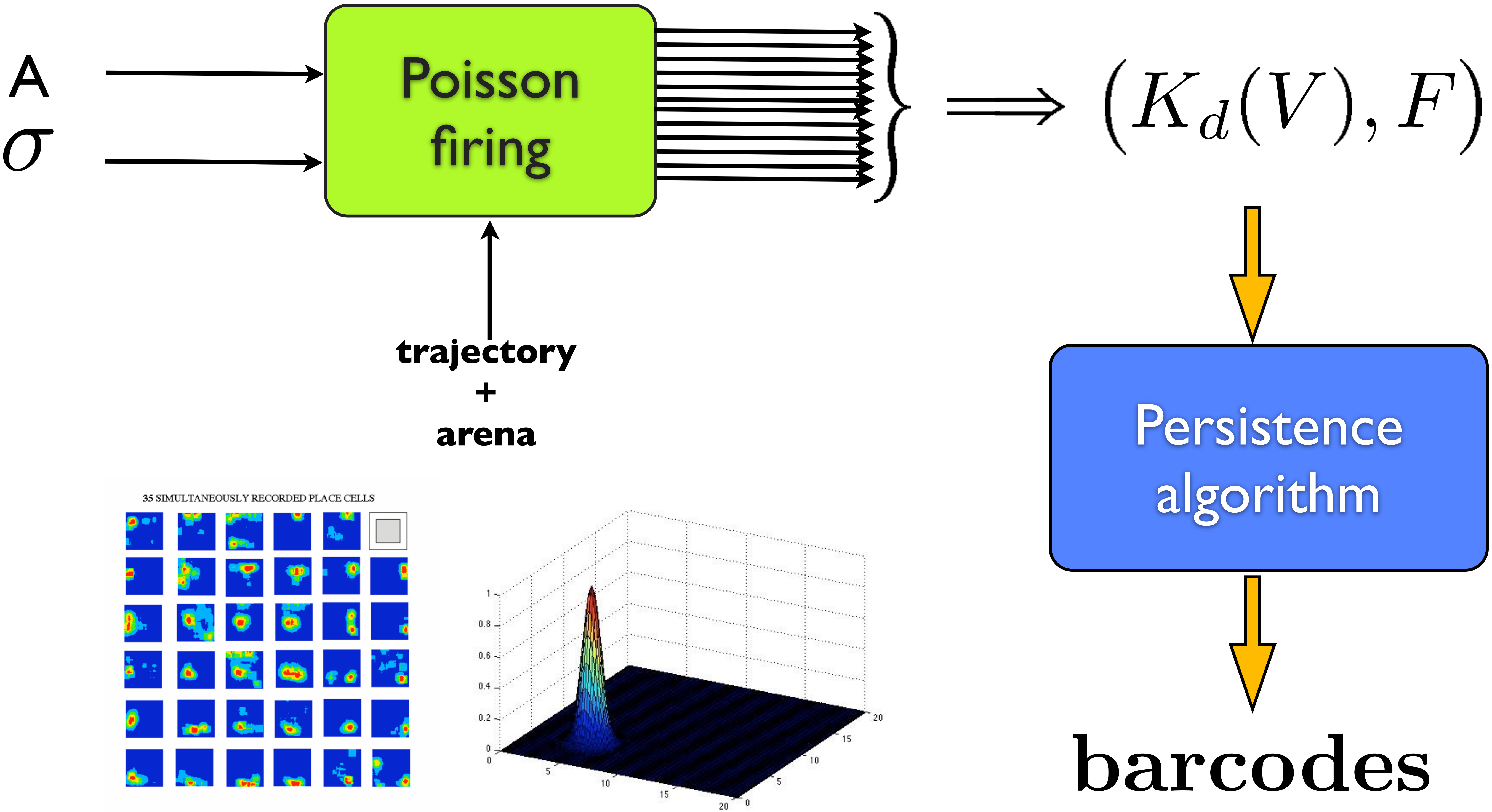


- **filtration time** $F(\sigma)$ of each simplex $\sigma = [v_0, v_1, \ldots, v_k]$:

  $$F([v_0, \ldots, v_k]) := \min \left\{ t \geqslant 0 \,\text{s.t.}\, \left| s(v_i) \cap [t-\varepsilon, t+\varepsilon] \right| \geqslant n_0 \text{ for all } i = 0, 1, \ldots, k \right\}$$

  for some fixed $\varepsilon > 0$ and $n_0 \in \mathbb{N}$. Here $s(v_i)$ denotes the spike train observed at the i-th cell.

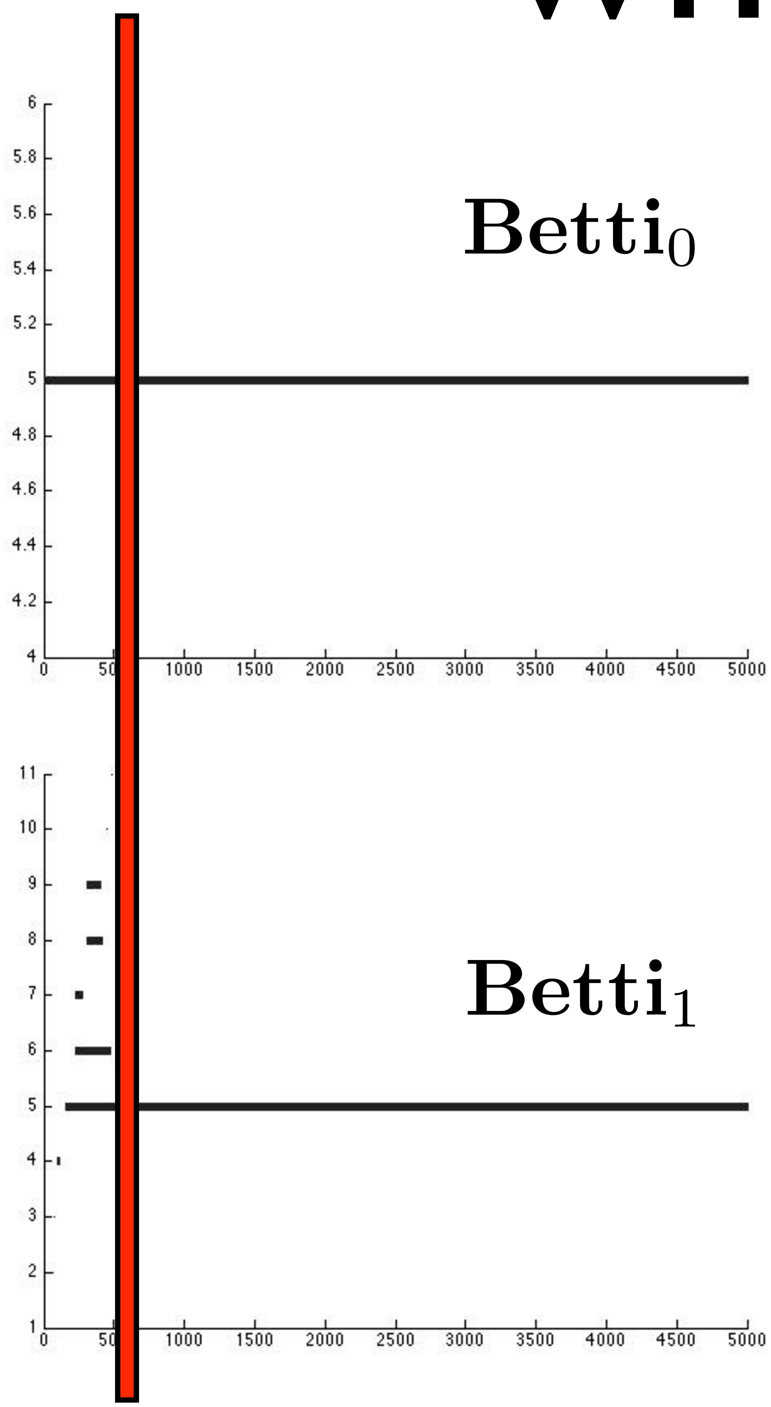
- **Filtration**: $\big(K_d(V), F\big) \implies$ **use persistence algorithm** and get **barcodes**.



The standard model for PF profiles is that of a Gaussian kernel with deviation σ, peak height A and center $\vec{x}_0$:

$$\rho(\vec{x}) = A \cdot e^{-\|\vec{x}-\vec{x}_0\|^2/2\sigma^2}$$

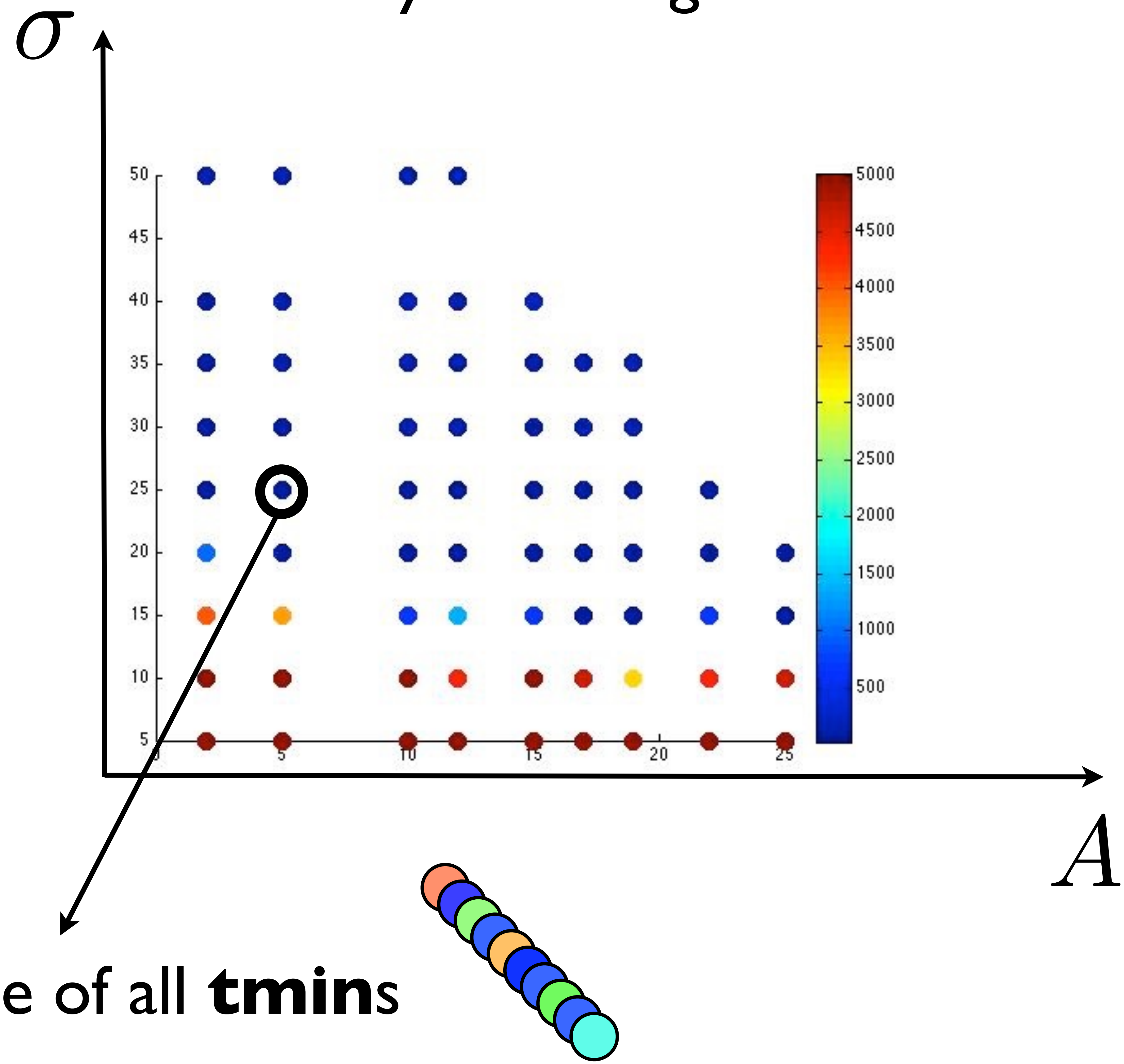
what we retain



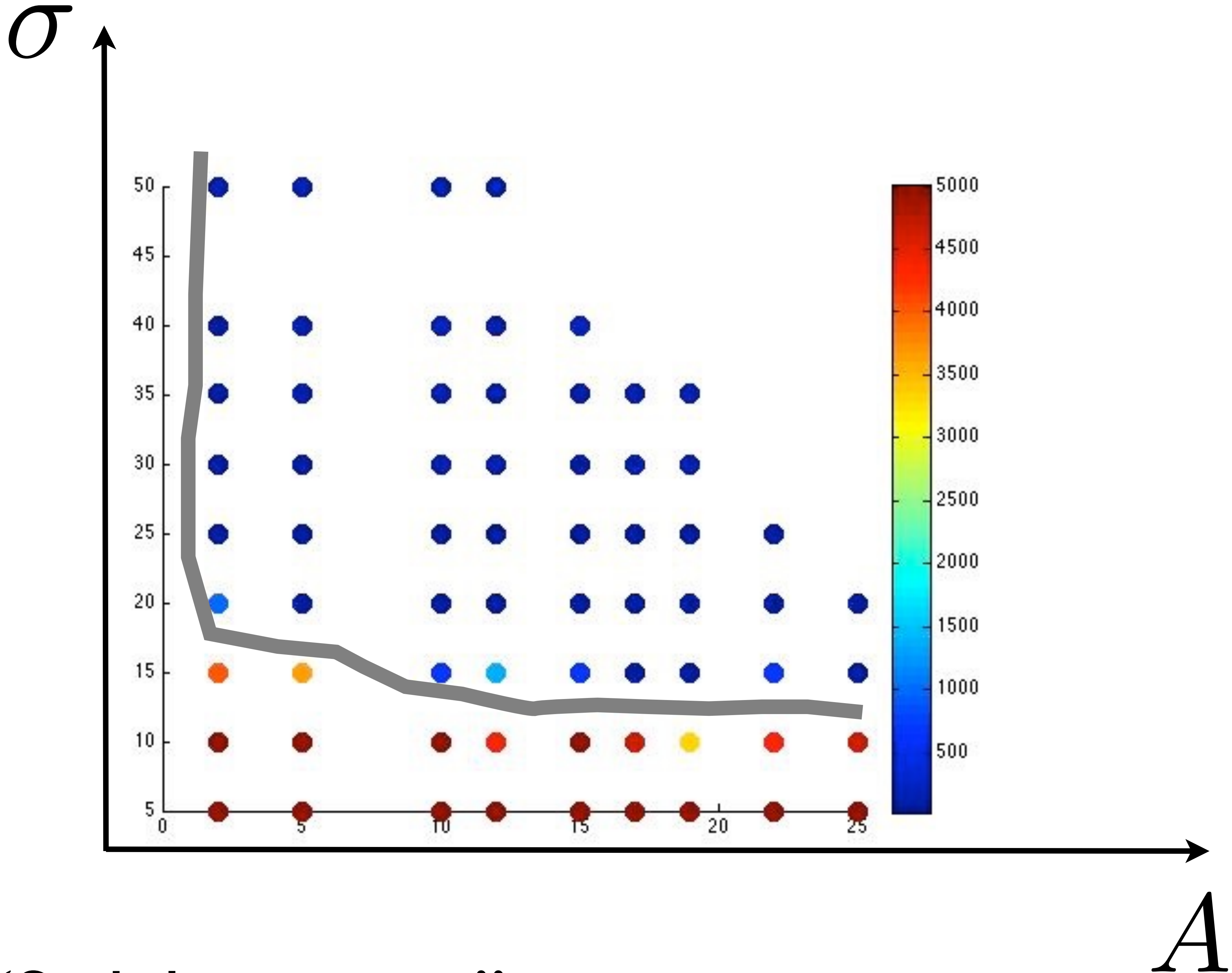
- For each run we record the first time **tmin** when the Betti signature is correct.

- tmin measures stability of the map (the smaller **tmin**, the more reliable the topological map)

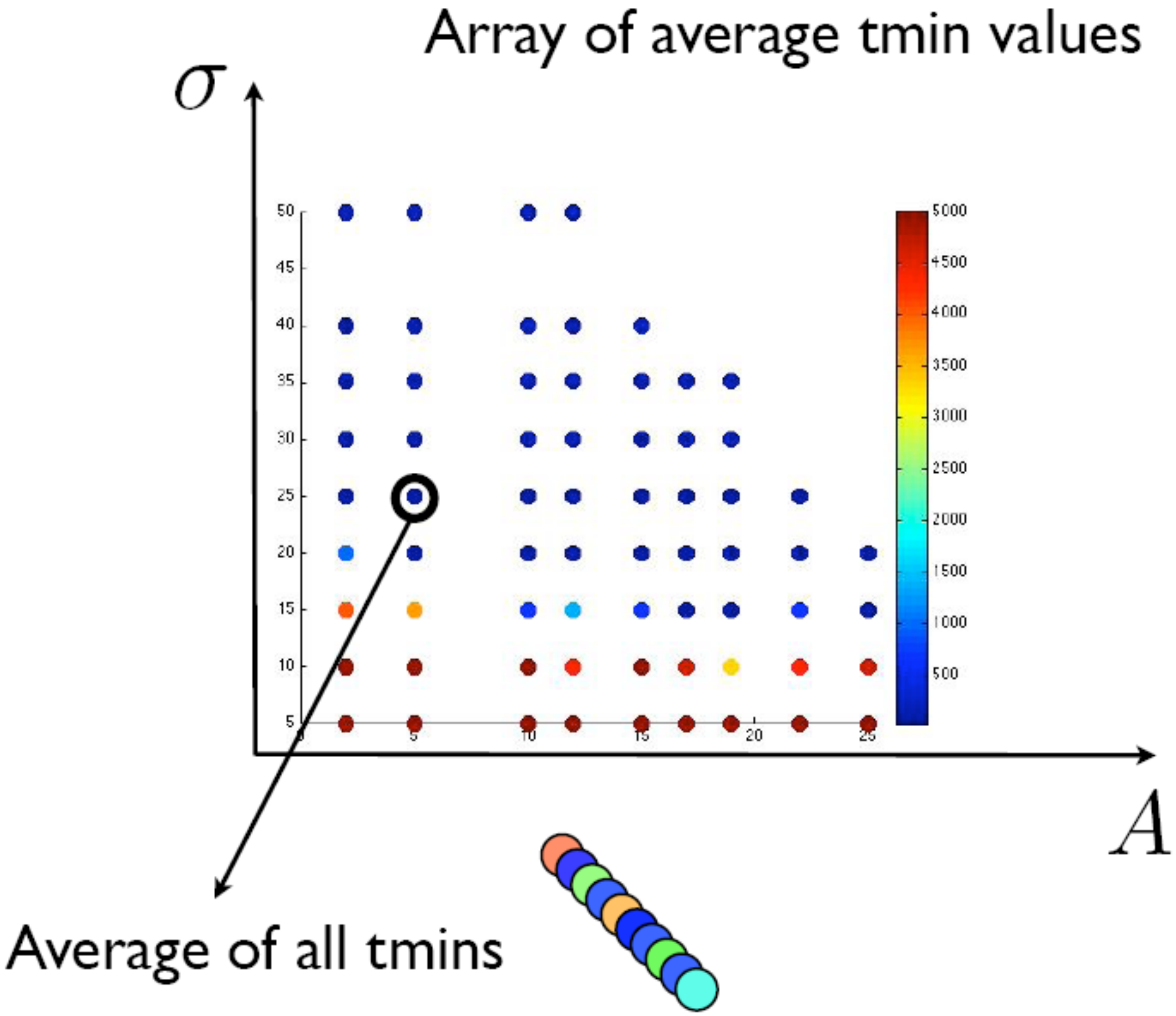
Array of average **tmin** values

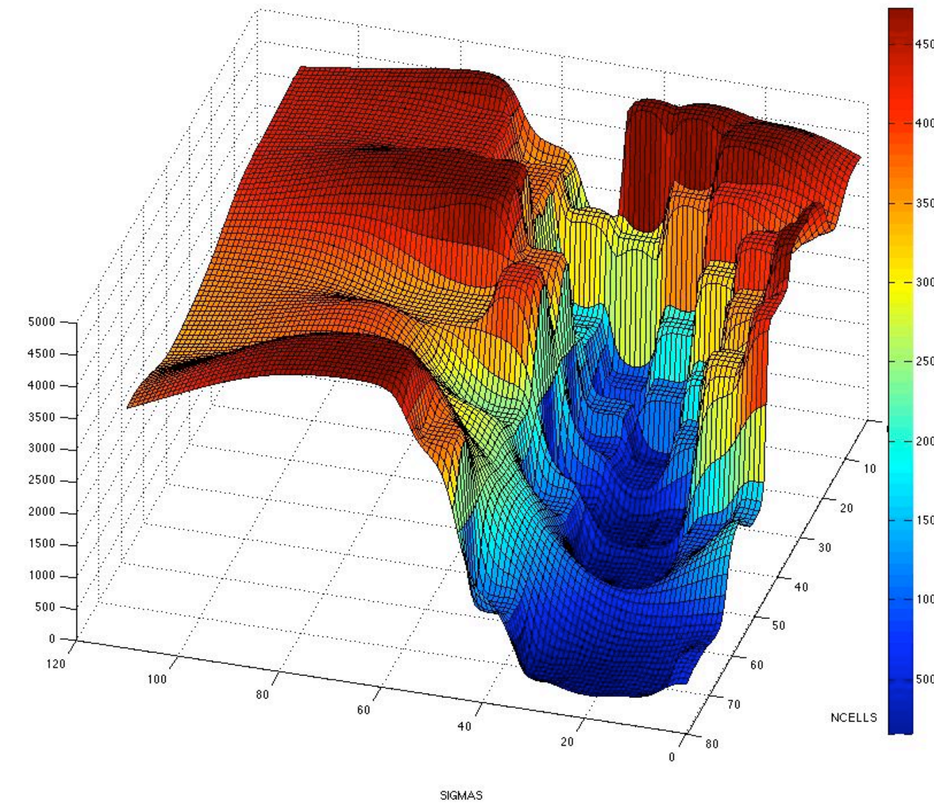


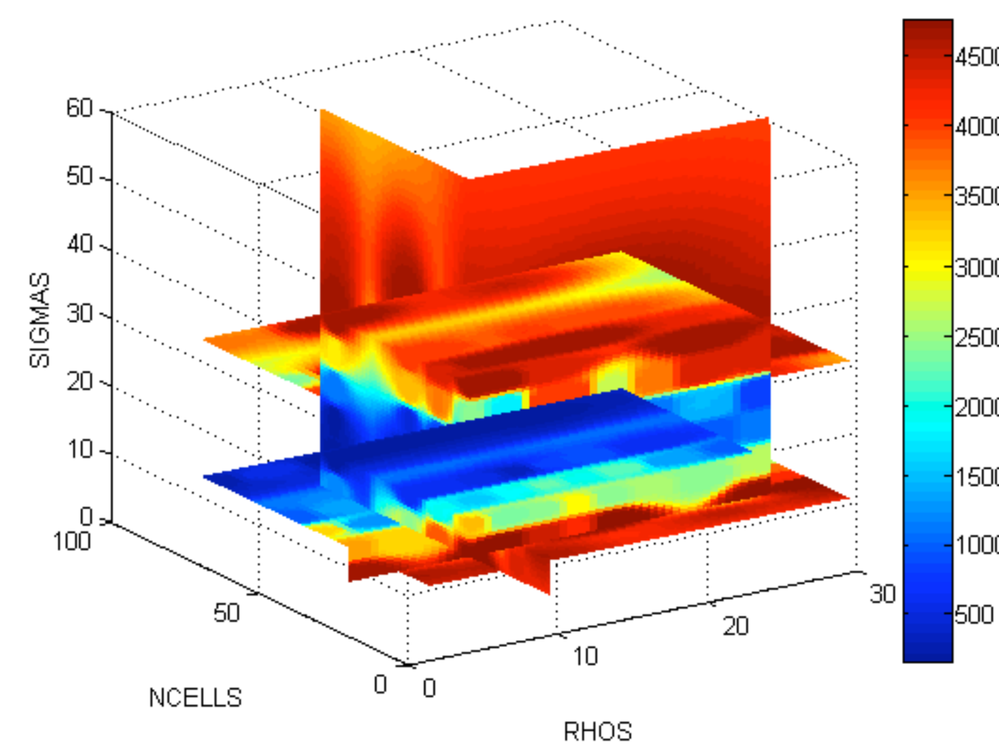
Average of all **tmin**s



"Stability region"







- There is a clearly defined region of effective map representation.

- For the values of place cell and place field parameters outside of the stability region the correct map can not be learned reliably – the signature may never converge to the correct answer.

Results

- The **topological stability** of the place field map is achieved only for a **specific range** of parameters of the place cell firing activity.

Discussion

- Topology is **parsimonious** info: there are many metrics that give the same topology.
- Use Betti numbers to characterize the arena.
- Use **persistence topology** tools to compute scale dependent Betti numbers and thus construct a measure of **robustness**.
- Study **stability** of this information.
- —> better analysis should include **forgetting** connections/associations between neurons. Leads to ZigZag persistence!