STRUCTURES CLUSTER OF EXCELLENCE

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

ATA HSE

International Laboratory of Algebraic Topology and Its Applications

HSE University Faculty of Computer Science

Topologically Autoencoding Cognitive Maps

Maxim Beketov, HSE University

4th Workshop on Topological Methods in Data Analysis
21 September 2023, Heidelberg University

Our Team

Konstantin Sorokin,
PhD student

Maxim Beketov,
PhD student

Konstantin Anokhin,
Prof., Academician of RAS

Vladimir Sotskov,
PhD candidate

Anton Ayzenberg,
Prof.

Michael Subbotin,
MSc student

Place Cells

2014 Nobel Prize
in Physiology or Medicine

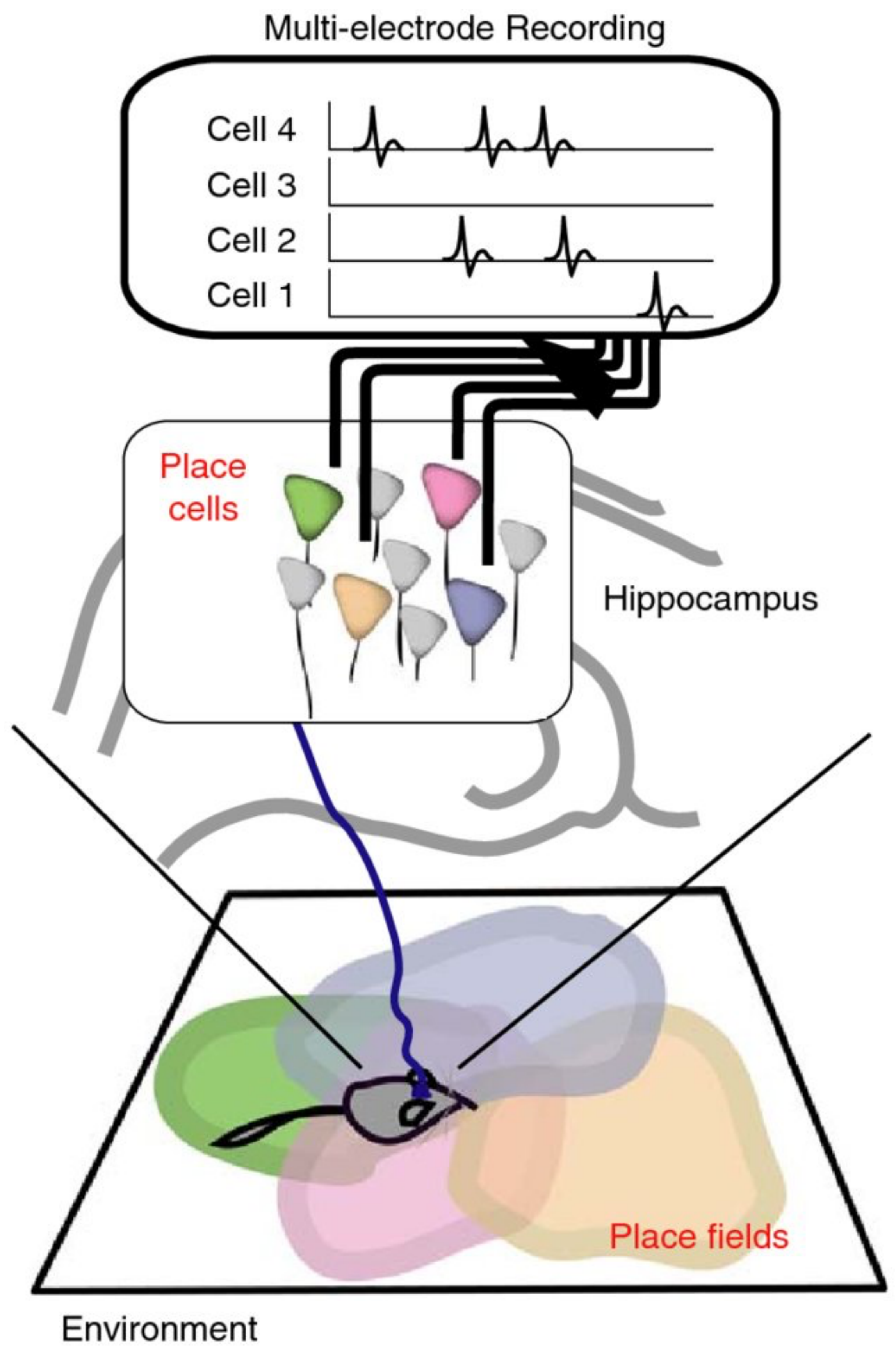
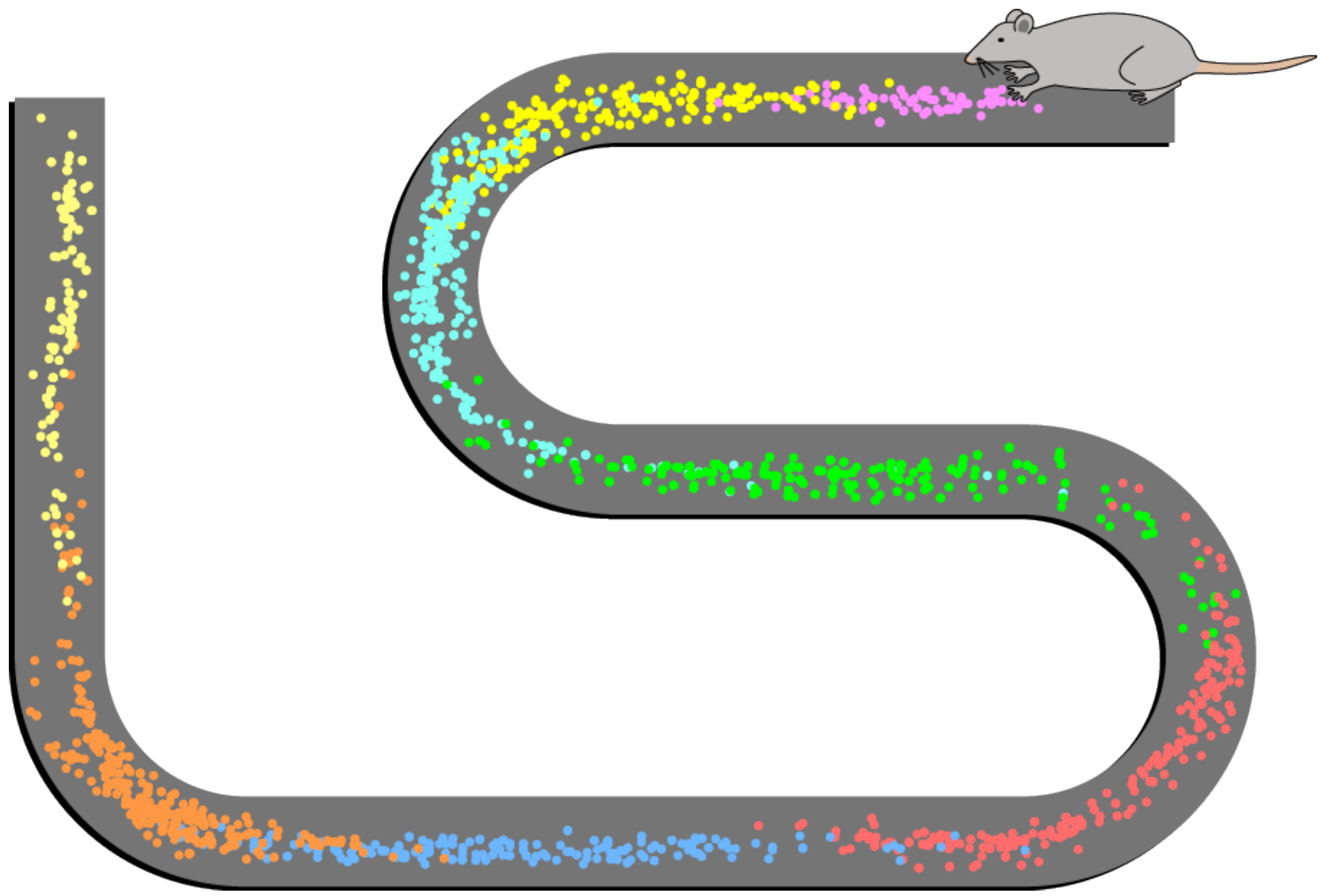
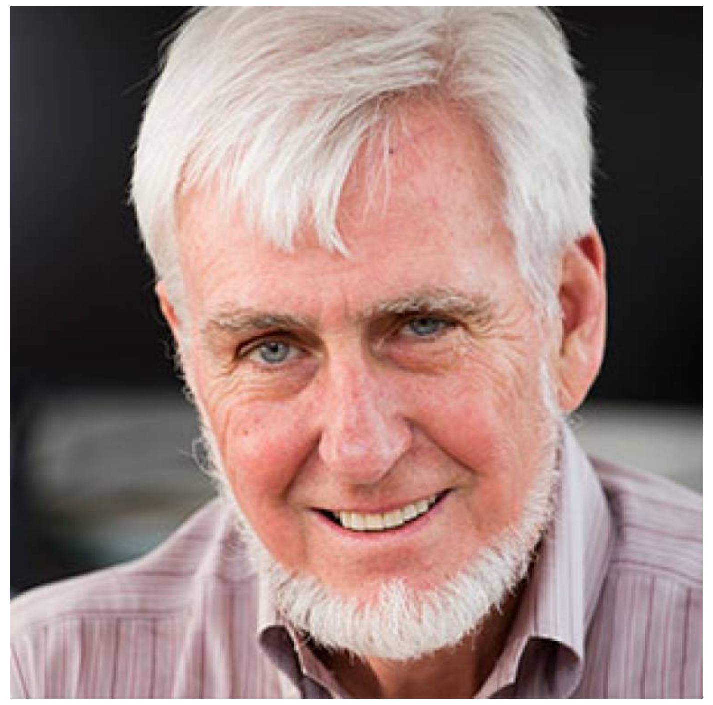


Image from: L Wagatsuma, Hiroaki, and Yoko Yamaguchi. "Neural dynamics of the cognitive map in the hippocampus." Cognitive Neurodynamics 1 (2007): 119-141:



↑ **place fields**

Image from Wiki



John O'Keefe

May-Britt Moser, Edvard I. Moser

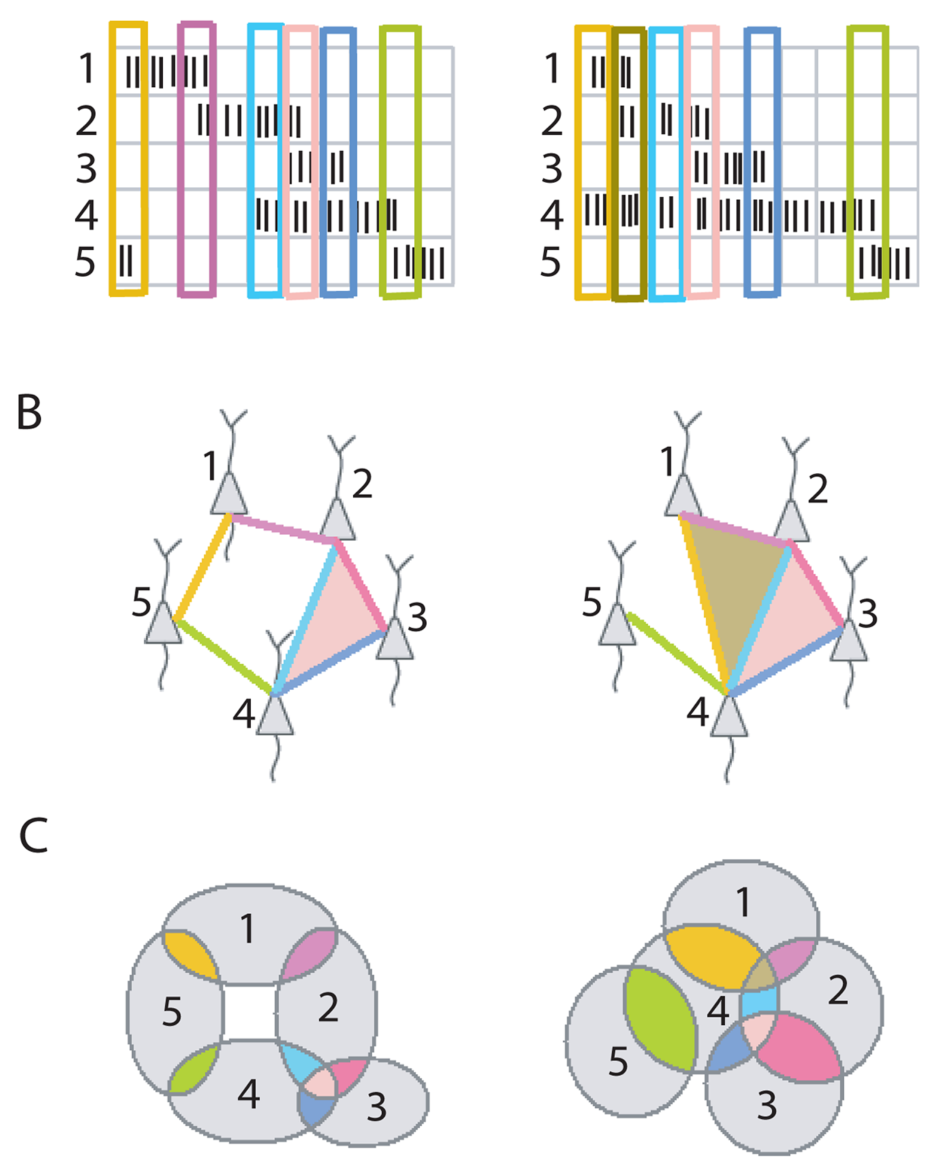
O'Keefe, John, and Jonathan Dostrovsky. "The hippocampus as a **spatial map**: preliminary evidence from unit activity in the freely-moving rat." *Brain research* (1971). *(~70k citations)*

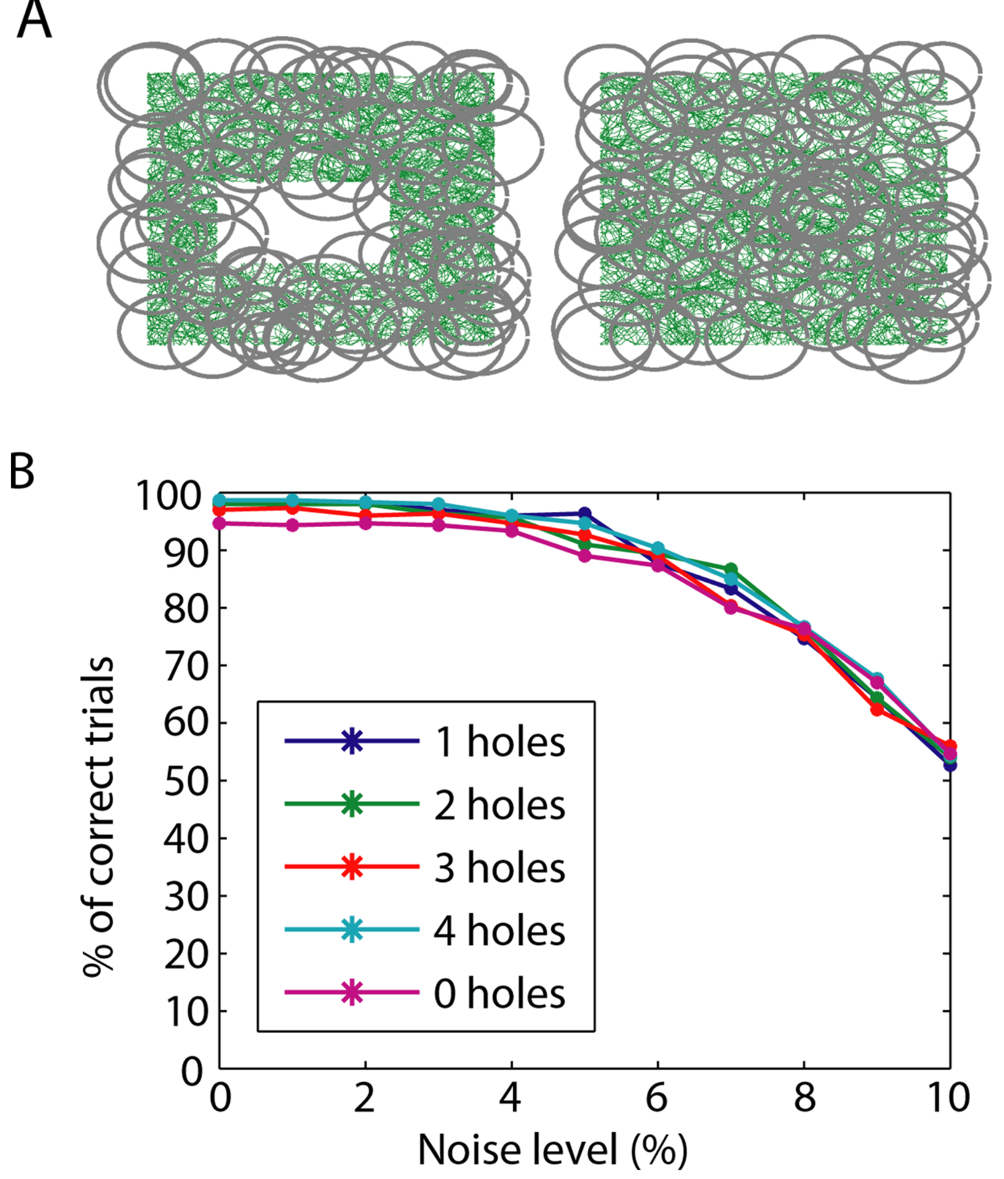
The work of Curto, Itskov, Dabaghian et al.

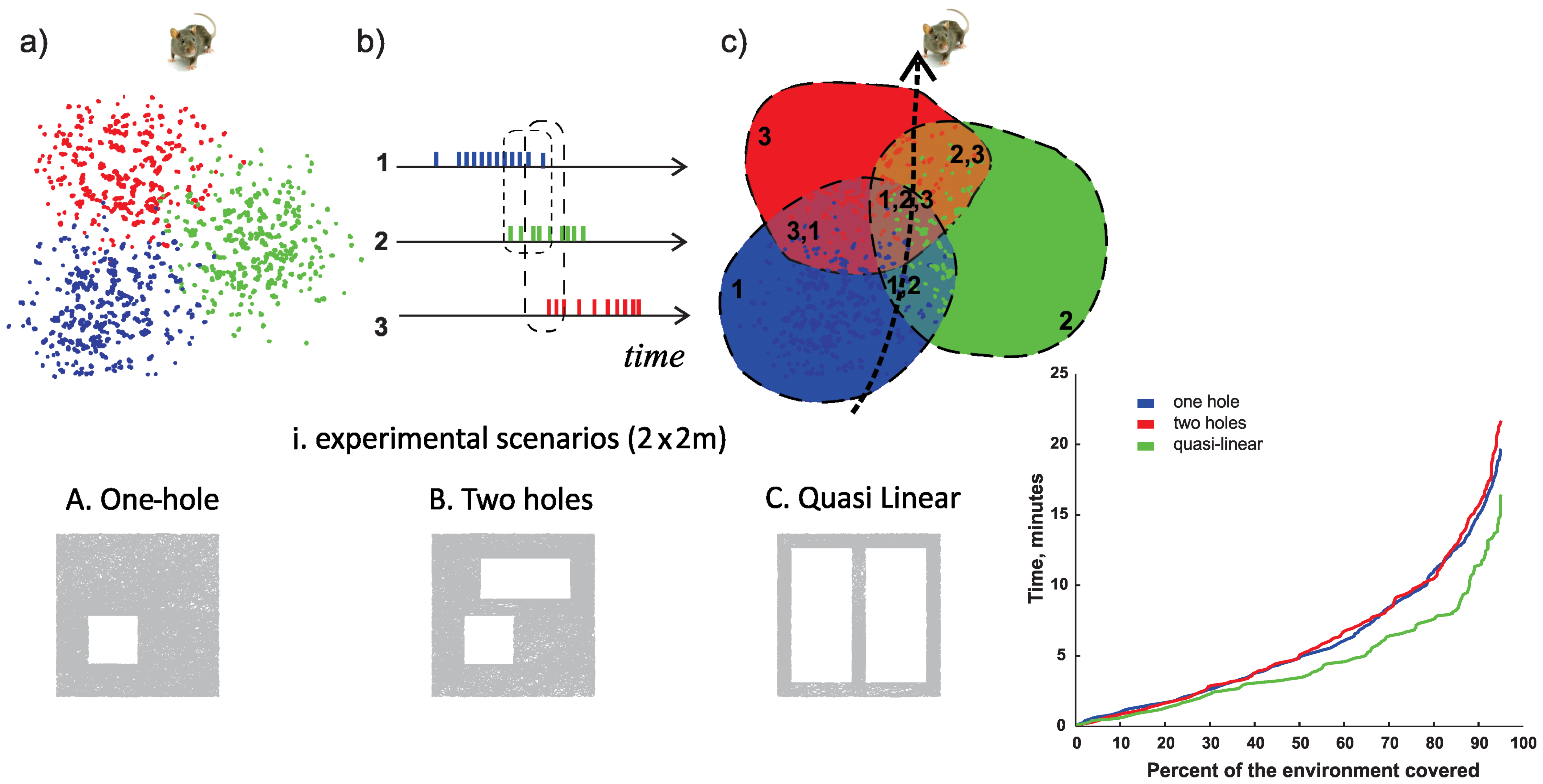
Carina Curto

Vladimir Itskov

Yury Dabaghian



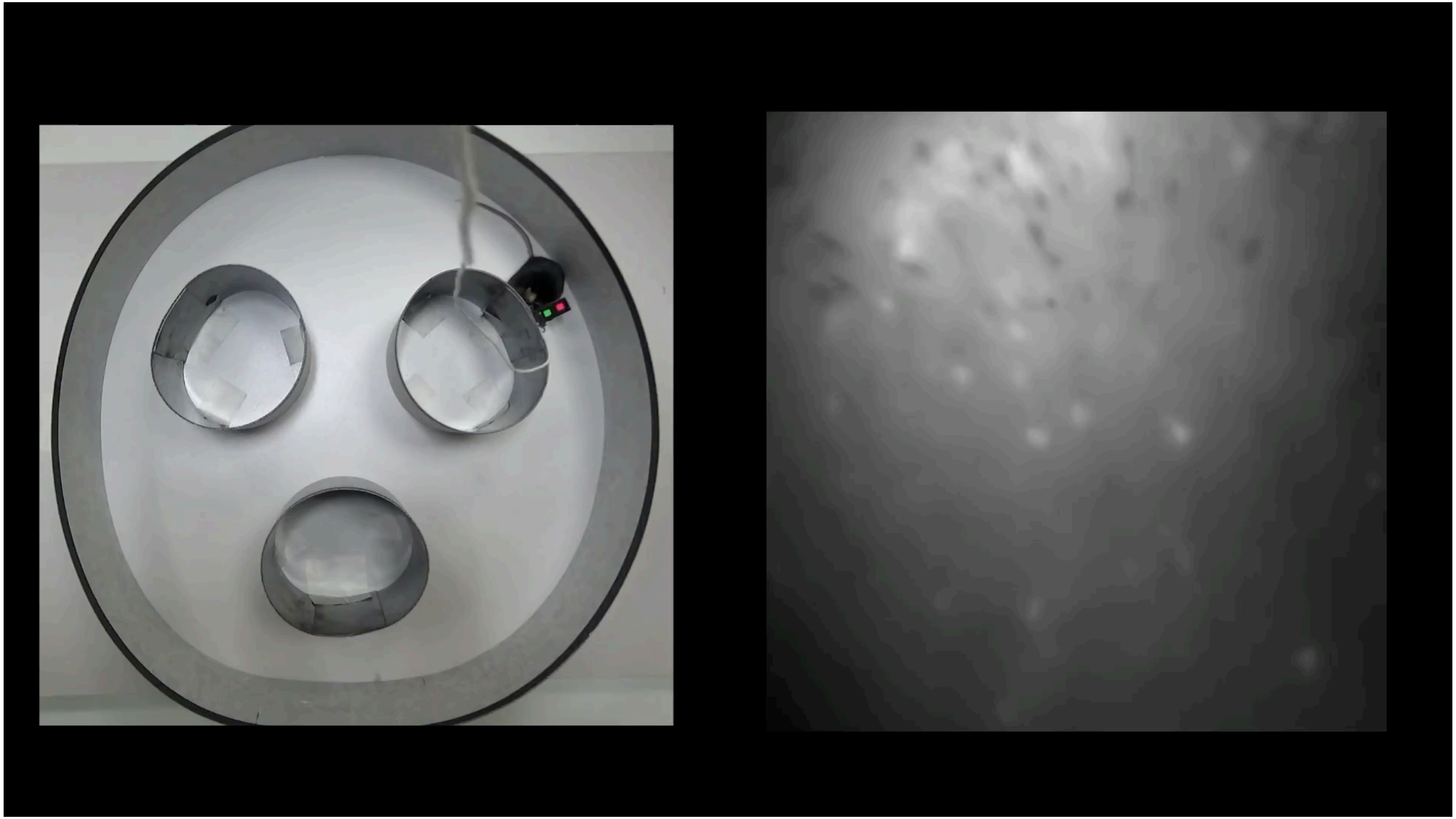




Curto, Carina, and Vladimir Itskov.
"Cell groups reveal structure of stimulus space."
PLoS computational biology 4.10 (2008): e1000205.

Dabaghian, Y., F. Memoli, L. Frank, and G. Carlsson.
"A topological paradigm for hippocampal spatial map formation
using persistent homology." *PLoS Computational Biology* 8, no. 8 (2012).

Our experiments (conducted by V. Sotskov)



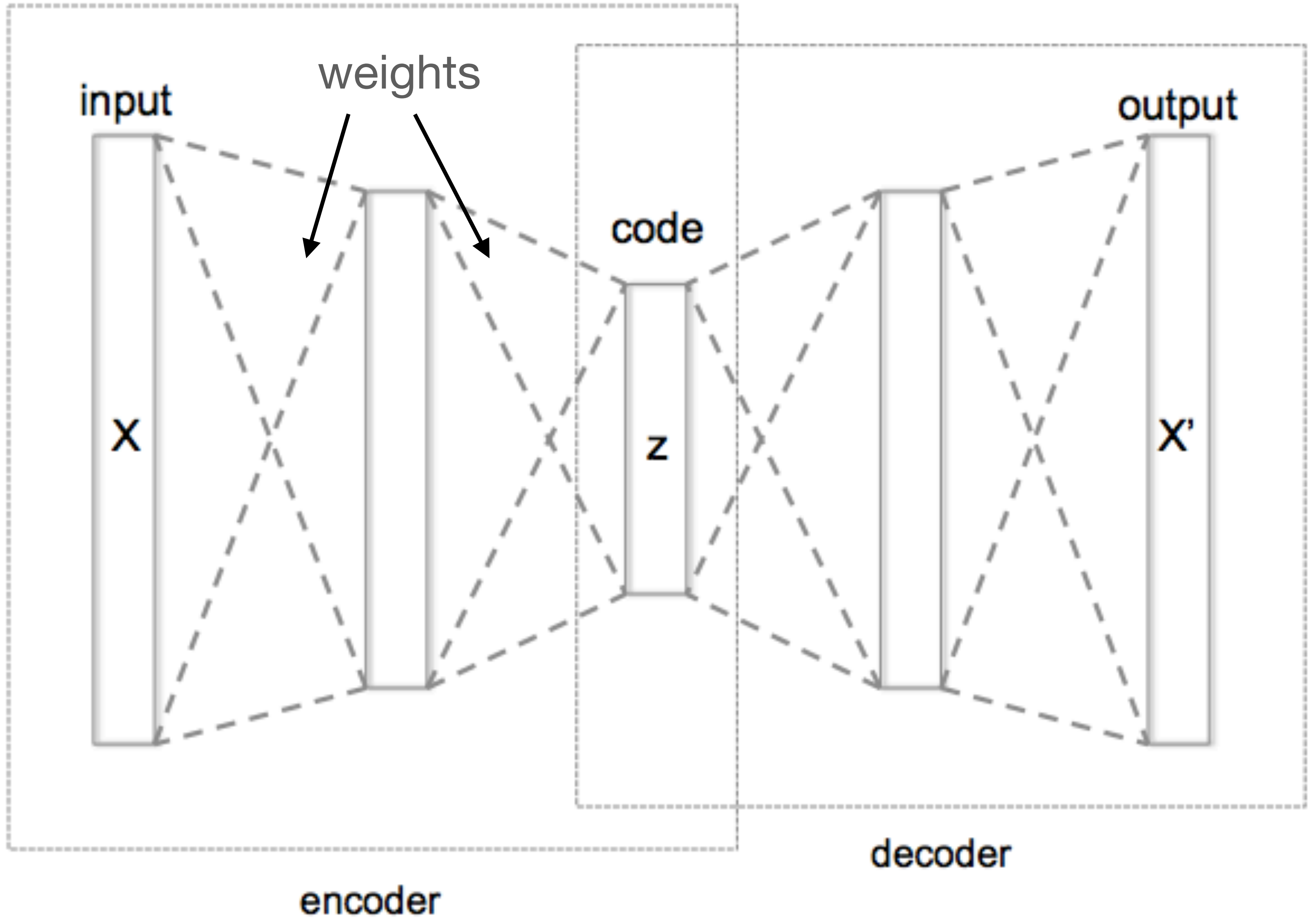
We’ve had: several mice; arenas with **1, 2, 3 holes**; ~100-300 visible neurons (calcium imaging)

Autoencoders

Classic Autoencoder is a neural network that learns to map X to Z (lower-dimensional **latent** space) and then Z to X’, minimizing the **reconstruction** loss

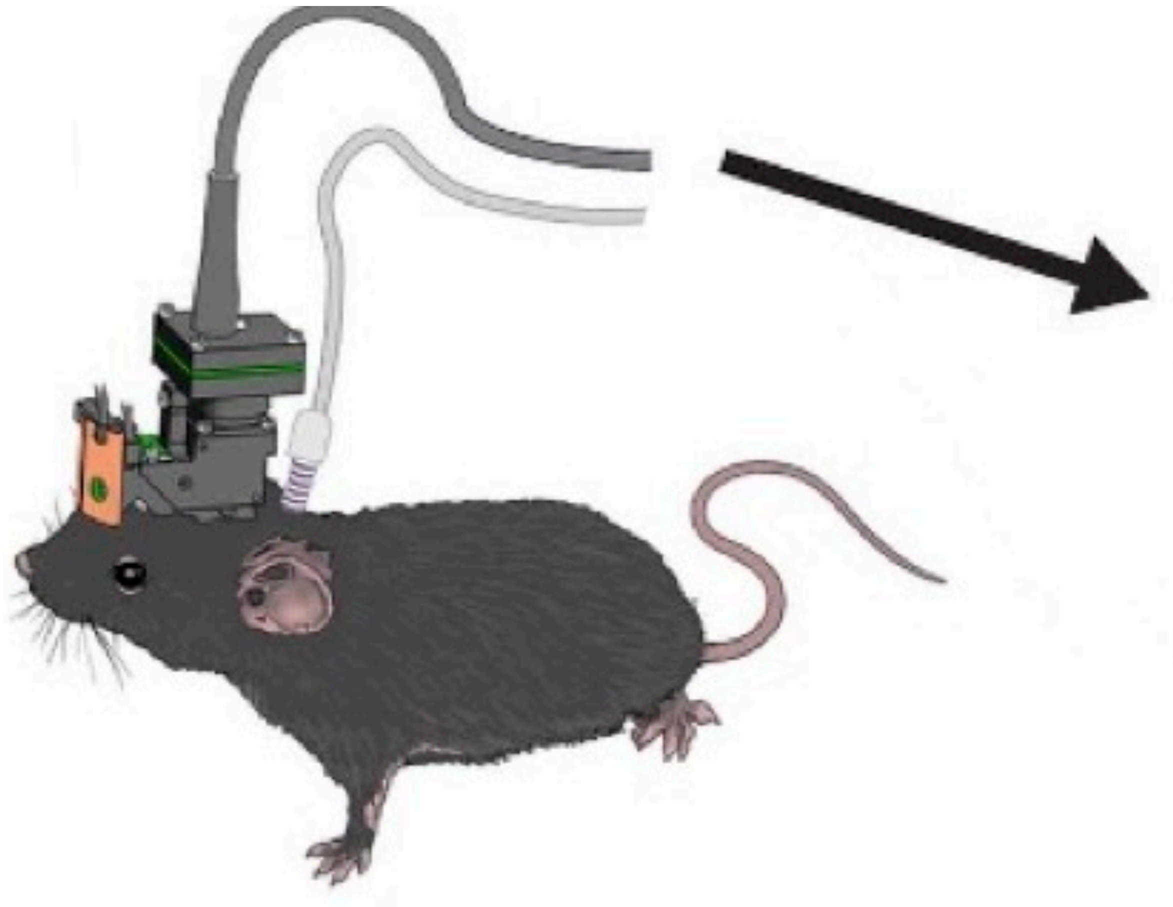
$$\underset{(weights)}{\hat{w}} = \underset{w}{\text{argmin}} ||X - \underbrace{\text{dec}(\text{enc}(X))}_{\equiv X'}||^2$$

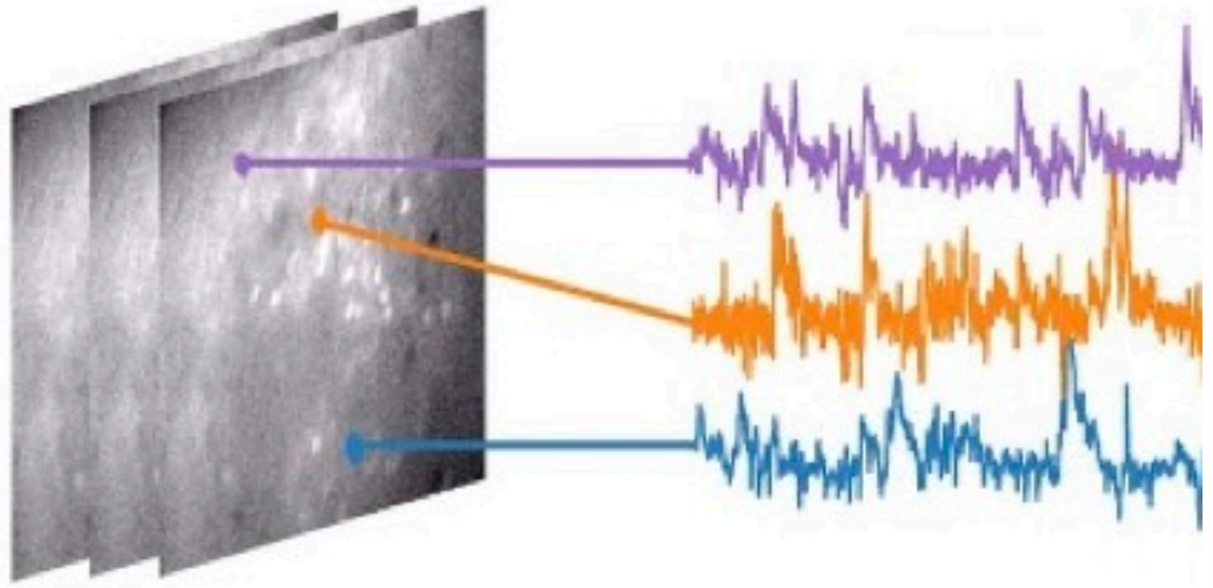
(Some regularization term on Z should be added, of course)



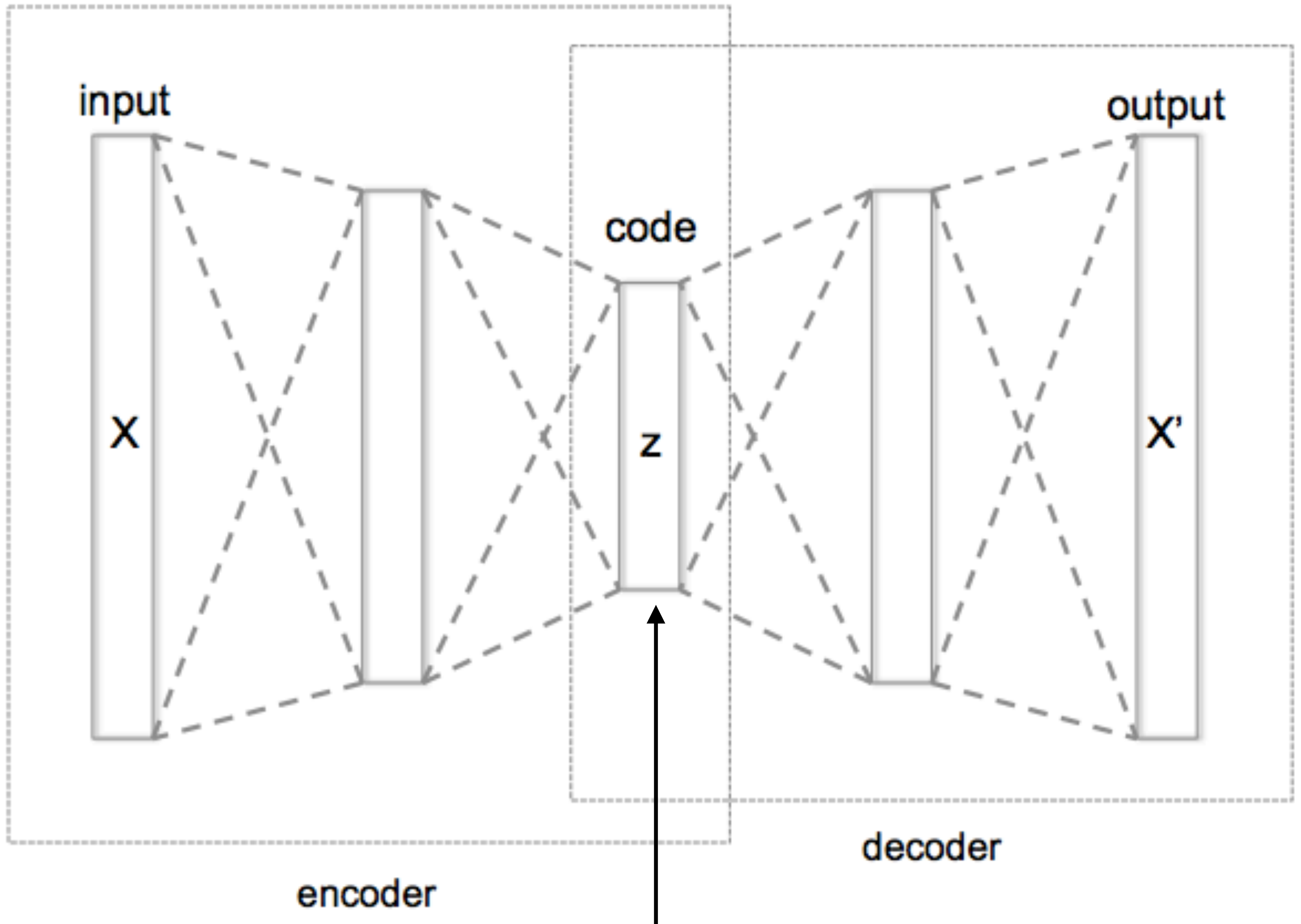
$$\text{enc}_w : X \to Z \qquad \text{dec}_w : Z \to X'$$

Autoencoding cognitive maps

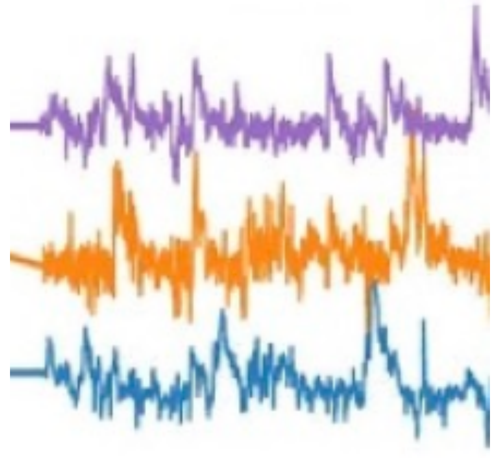


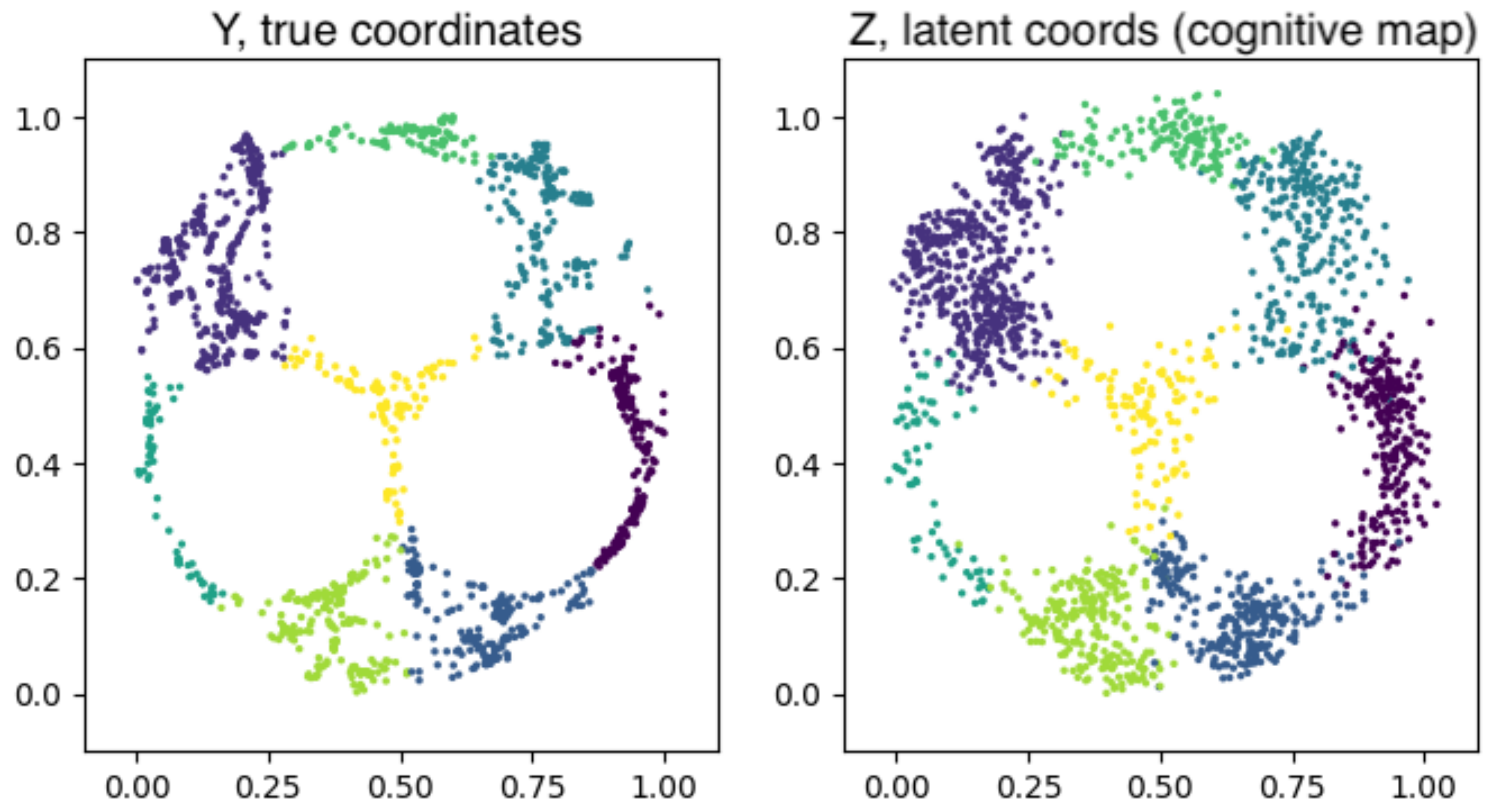


~30 “most informative” neurons selected



reconstruction

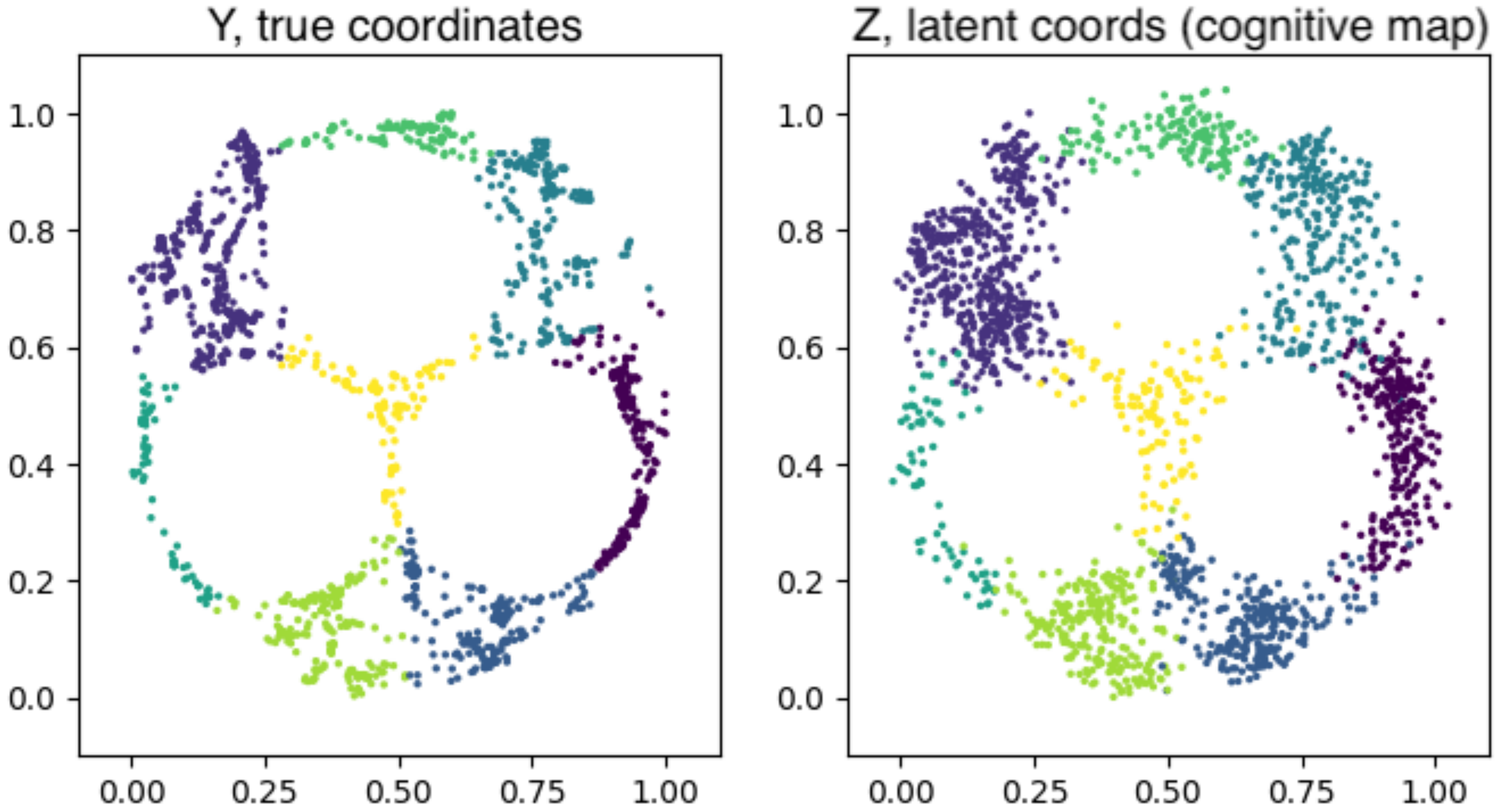


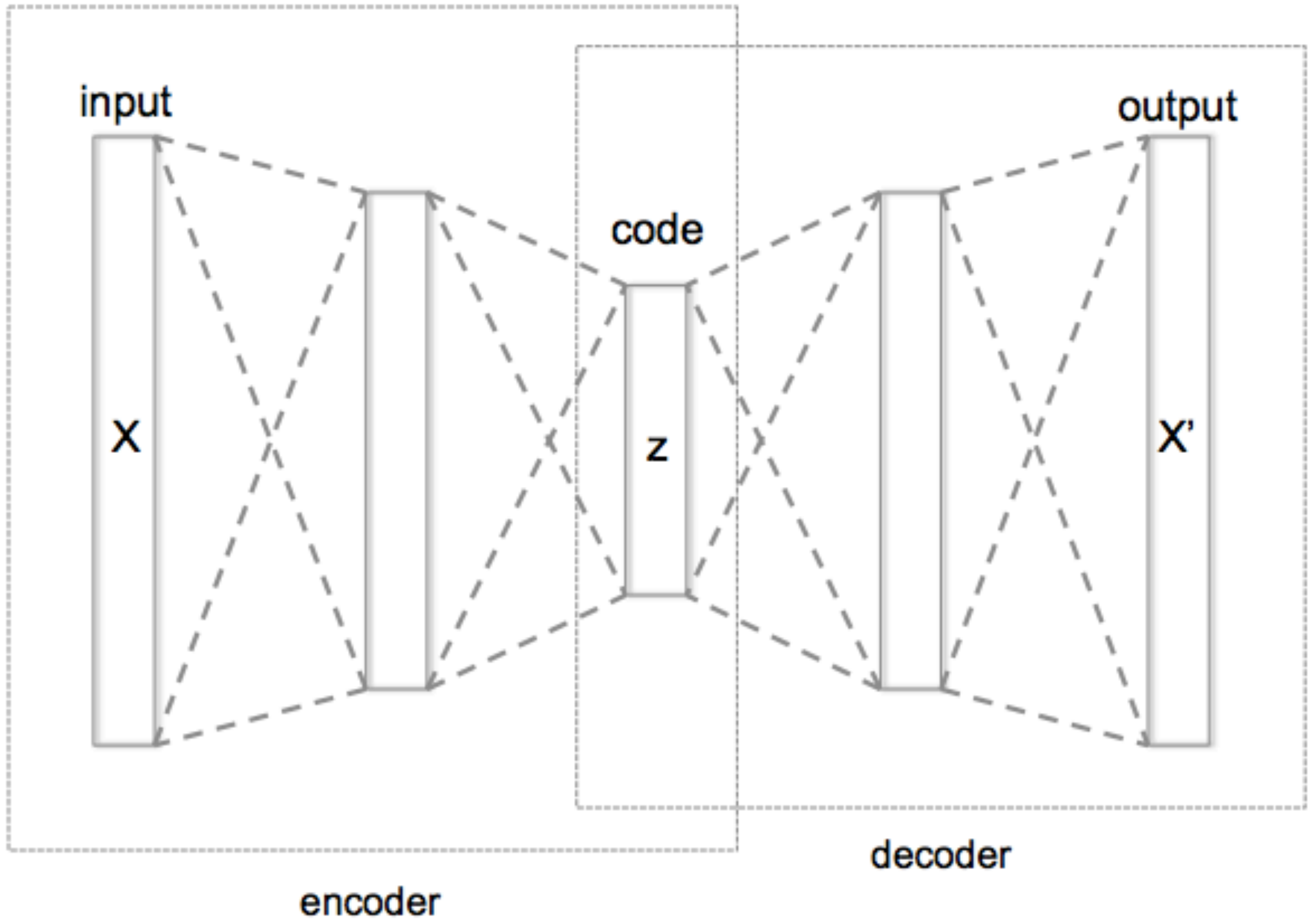


2D **cognitive map**

Image from:
Trevathan, James K., et al. "Calcium imaging in freely moving mice during electrical stimulation of deep brain structures." Journal of neural engineering 18.2 (2021): 026008.

Autoencoding with a metric reconstruction penalty





3 layers were enough for this
(This in on test set, unseen on training ofc)

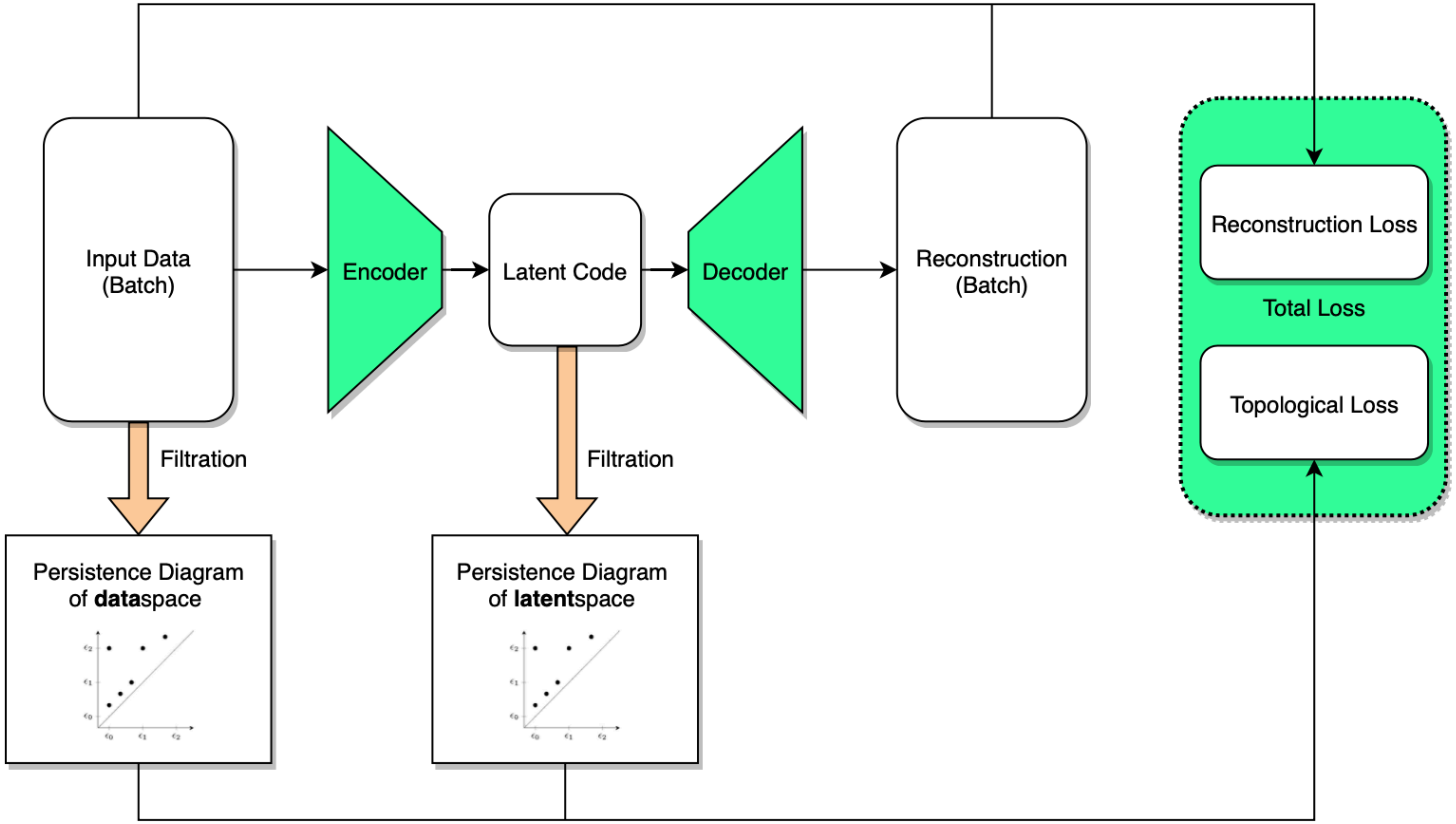
We forced the AE to learn the true coordinates with an additional metric dissimilarity loss:

$$\text{Total loss} = ||X - \text{dec}(\text{enc}(X))||^2 + \lambda ||d_{ij}(Y) - d_{ij}(Z)||$$

Reconstruction loss

Topological Autoencoders

Moor, Michael, Max Horn, Bastian Rieck, and Karsten Borgwardt "Topological autoencoders"
In *International conference on machine learning*, pp. 7045-7054. PMLR, 2020.https://arxiv.org/abs/1906.00722



(Image from M. Moor's blogpost on the topic – https://michaelmoor.ml/blog/topoae/main/)

Topological Autoencoders: details

PH is stored as a collection of $\{D_0, D_1, D_2, \dots D_d, \dots\}$

diagrams – pairs of *(birth_scale, death_scale)*

and **pairings** $\{\pi_0, \pi_1, \dots, \pi_d, \dots\}$

pairs of (s_i, s_j) where each d-dimensional feature is born in i-th simplex s_i and dead as j-th simplex s_j appears

In present work, authors **only track 0-homology** – that is, AE tries to preserve the number (and structure) of **connected components**! (tracking 1-homology was also tried)

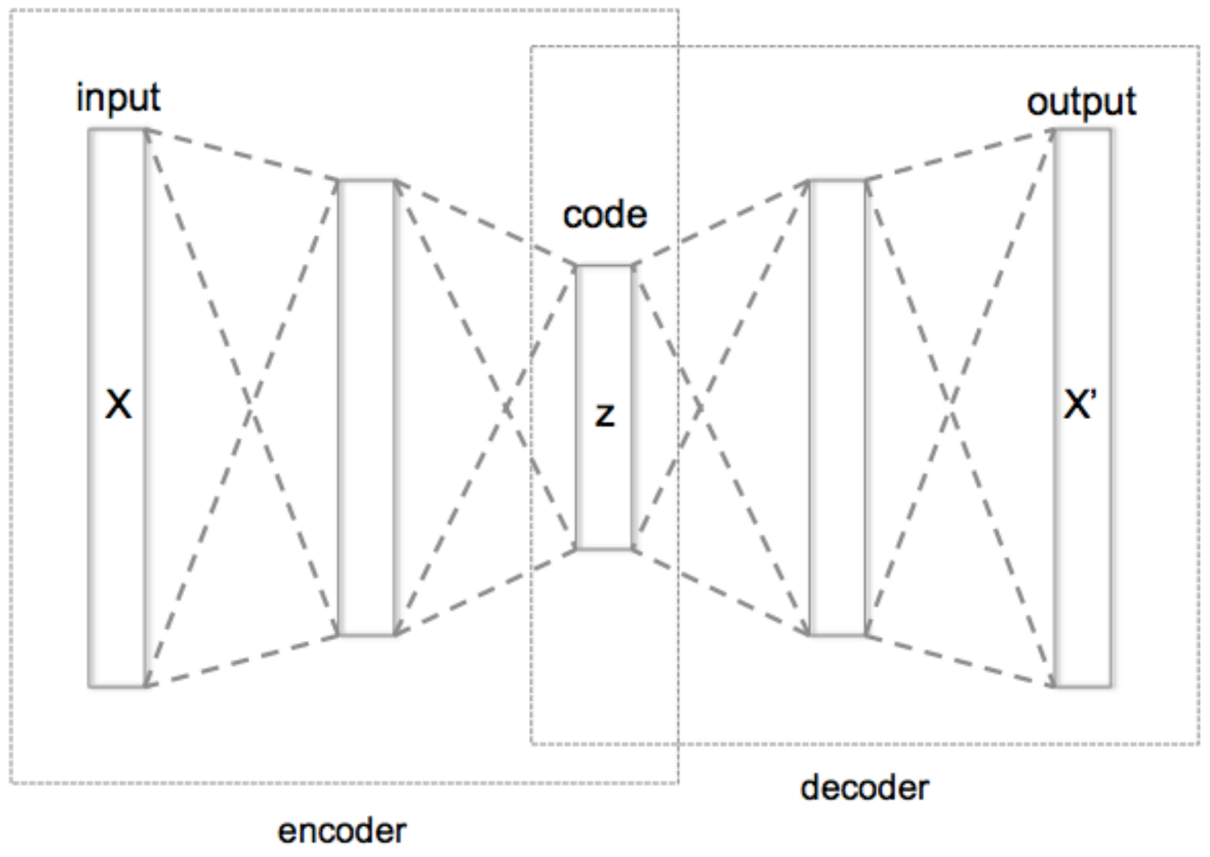
To do that, they only need $\mathbf{A}^S$ – the **distance matrix** of the point cloud, and the

0-pairings – that is, edges

Under the hood, they compute the minimum spanning tree, which is $O(n^2\alpha(n))$ complexity

Topological Autoencoders: details

The topological loss is "two-sided" in X and Z



$$\mathscr{L}_{topo} = \mathscr{L}_{X\to Z} + \mathscr{L}_{Z\to X}$$

$$\mathscr{L}_{X\to Z} = \frac{1}{2}\left\| \mathbf{A}^X\left[\pi^X\right] - \mathbf{A}^Z\left[\pi^X\right] \right\|^2$$ and $$\mathscr{L}_{Z\to X} = \frac{1}{2}\left\| \mathbf{A}^Z\left[\pi^Z\right] - \mathbf{A}^X\left[\pi^Z\right] \right\|^2$$

Now, for each batch, the AE parameters are encoded to minimize the loss, the gradient is quite simple:

$$\frac{\partial}{\partial \boldsymbol{\theta}}\mathscr{L}_{X\to Z} = \frac{\partial}{\partial \boldsymbol{\theta}}\left(\frac{1}{2}\left\| \mathbf{A}^X\left[\pi^X\right] - \mathbf{A}^Z\left[\pi^X\right] \right\|^2\right) = -\left(\mathbf{A}^X\left[\pi^X\right] - \mathbf{A}^Z\left[\pi^X\right]\right)^\top \left(\frac{\partial \mathbf{A}^Z\left[\pi^X\right]}{\partial \boldsymbol{\theta}}\right)$$

$$= -\left(\mathbf{A}^X\left[\pi^X\right] - \mathbf{A}^Z\left[\pi^X\right]\right)^\top \left(\sum_{i=1}^{\left|\pi^X\right|} \frac{\partial \mathbf{A}^Z\left[\pi^X\right]_i}{\partial \boldsymbol{\theta}}\right)$$

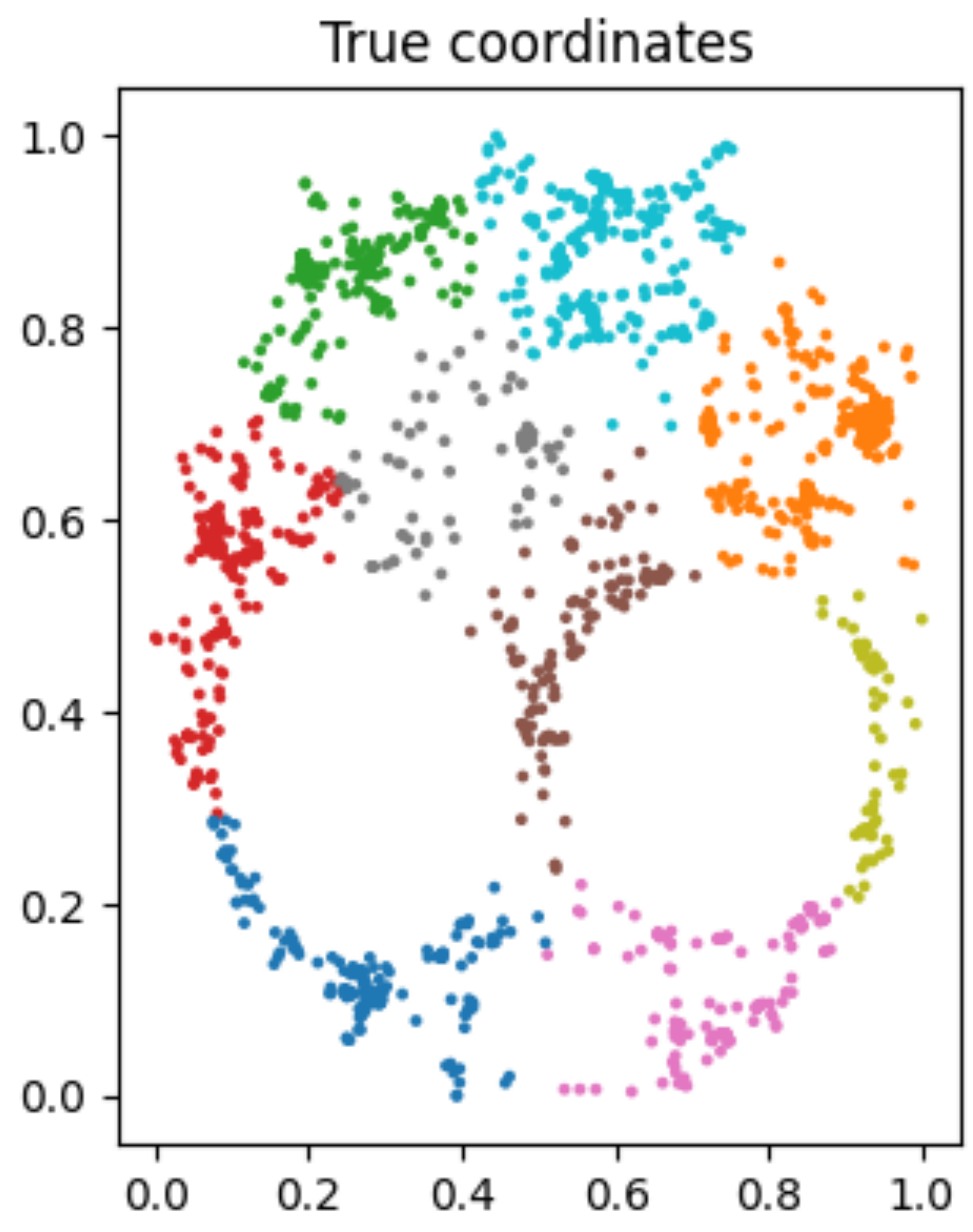
For L to be differentiable, **pairwise distances** (entries of A) should be **unique!** (Otherwise the pairing provides a discontinuity)

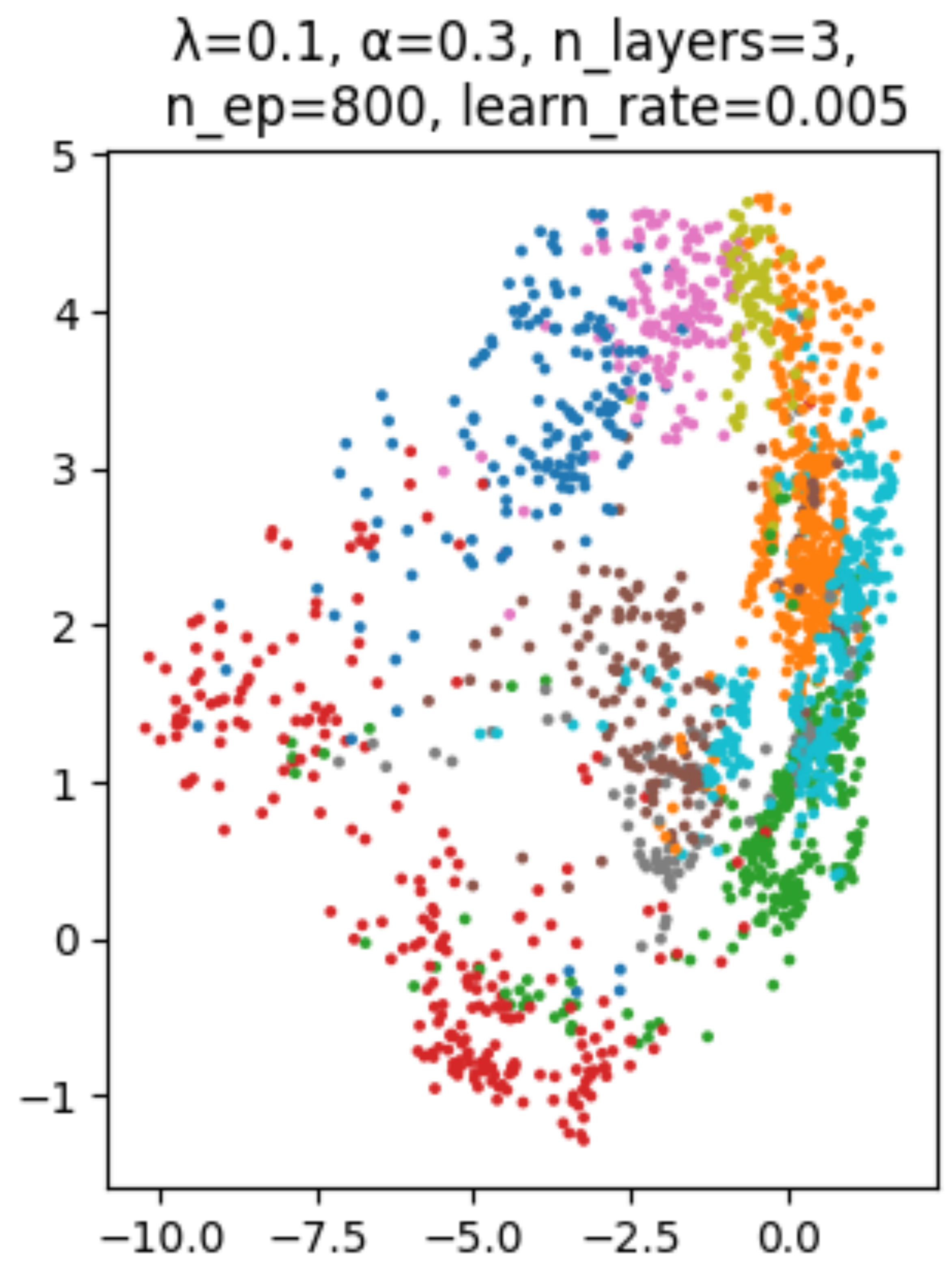
Topologically autoencoding

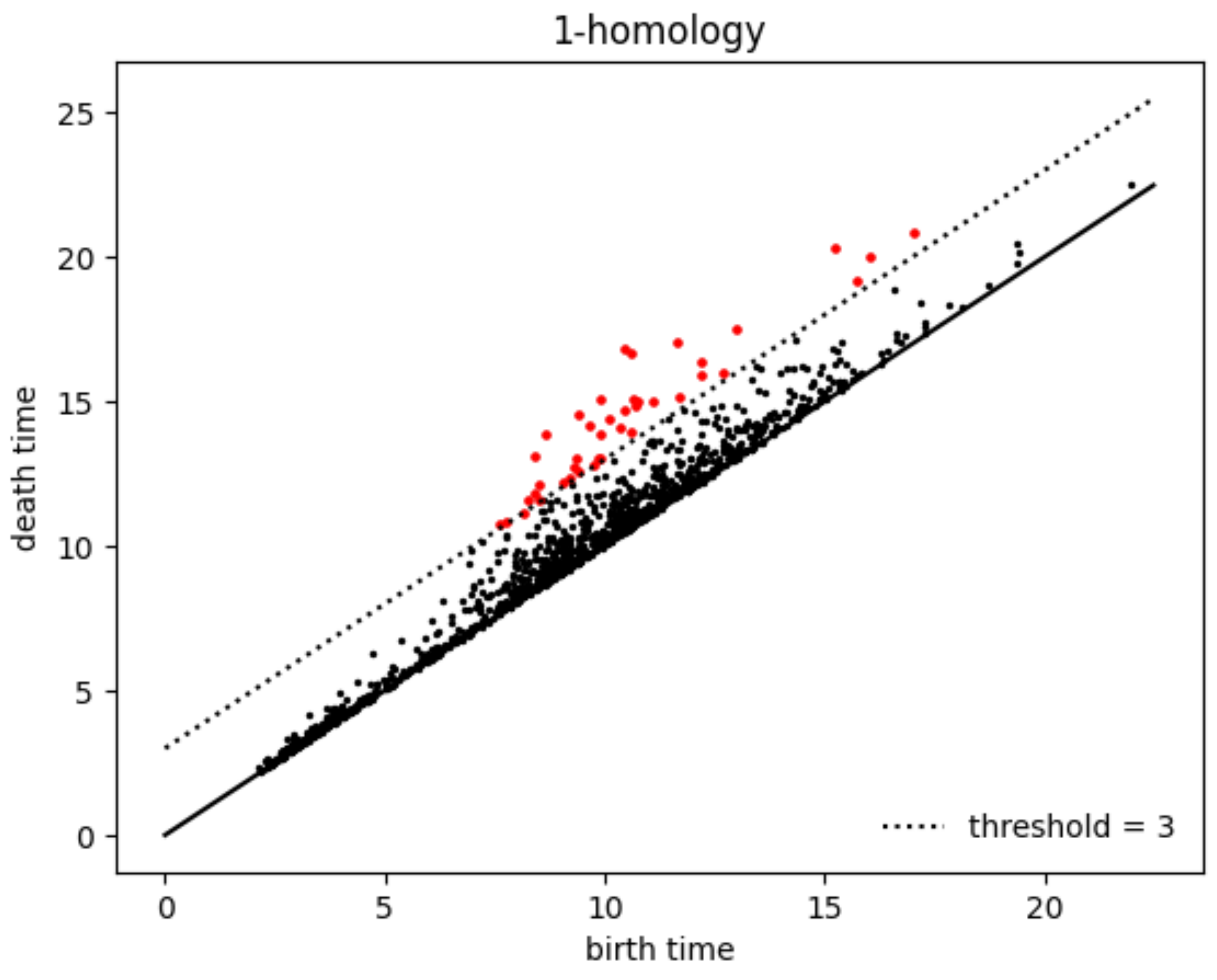
We tuned TopoAEs a bit:

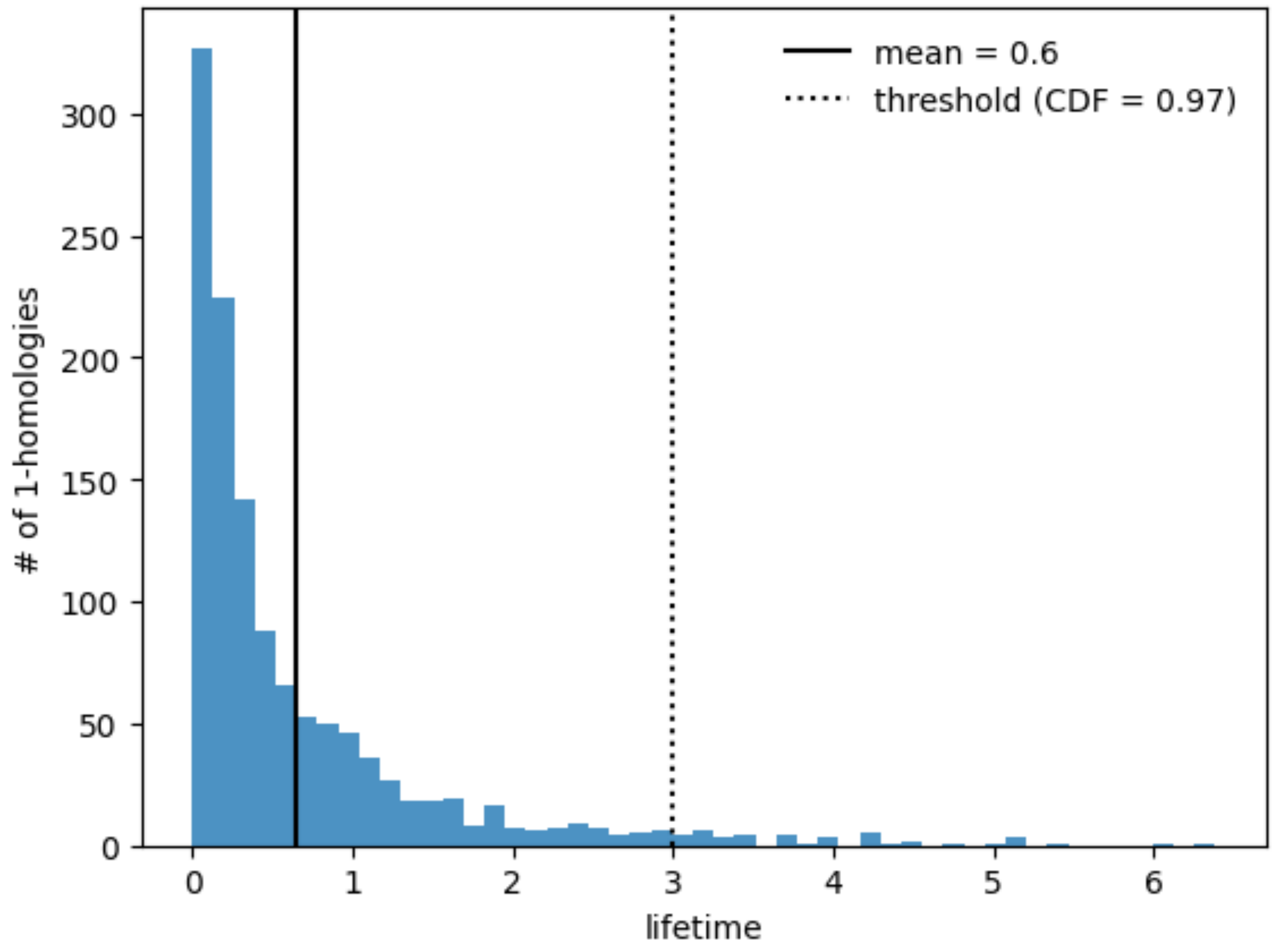
1) to only penalise for difference in **1-homologies**

2) to only preserve **long-living** (important) homologies



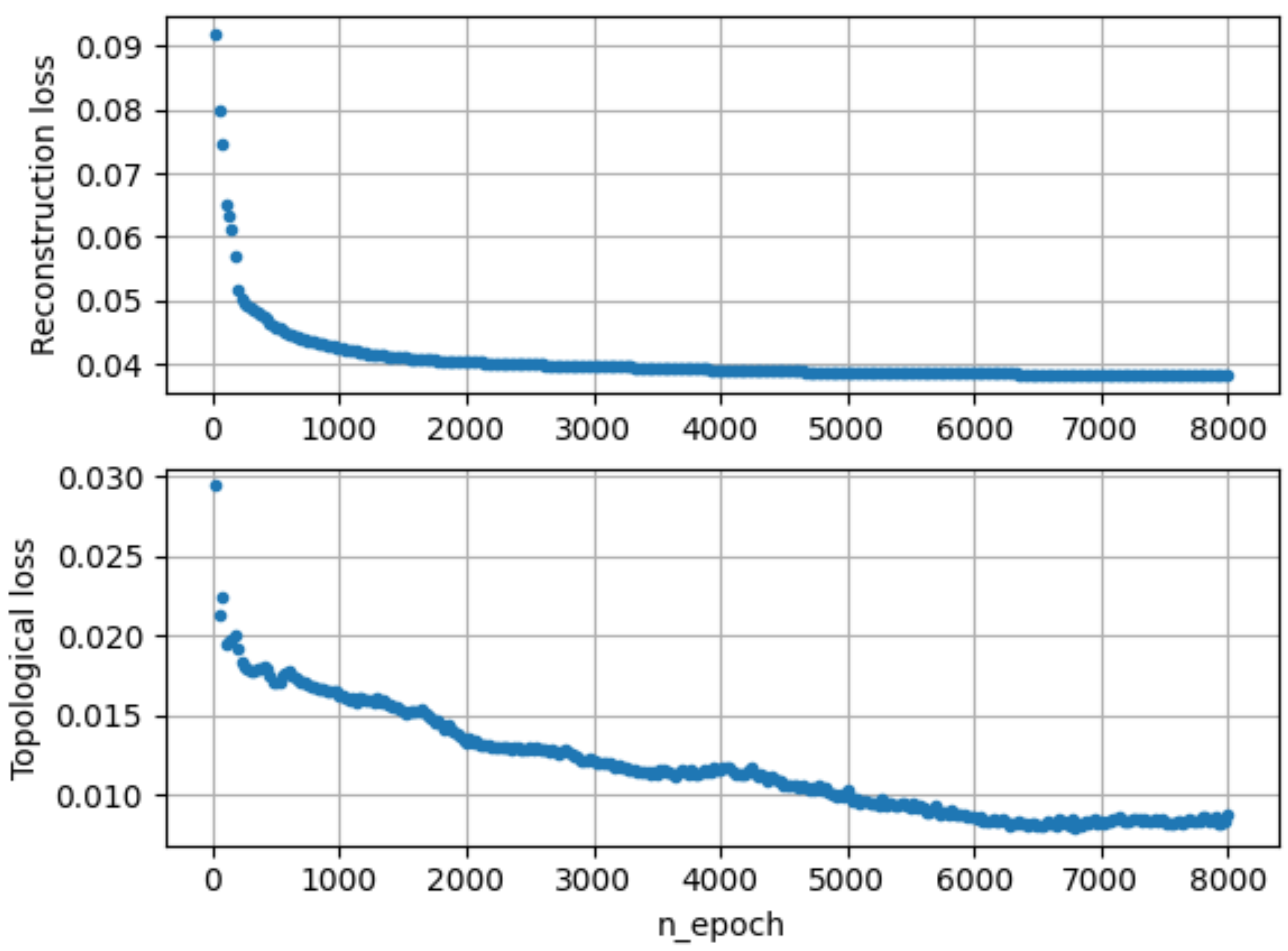


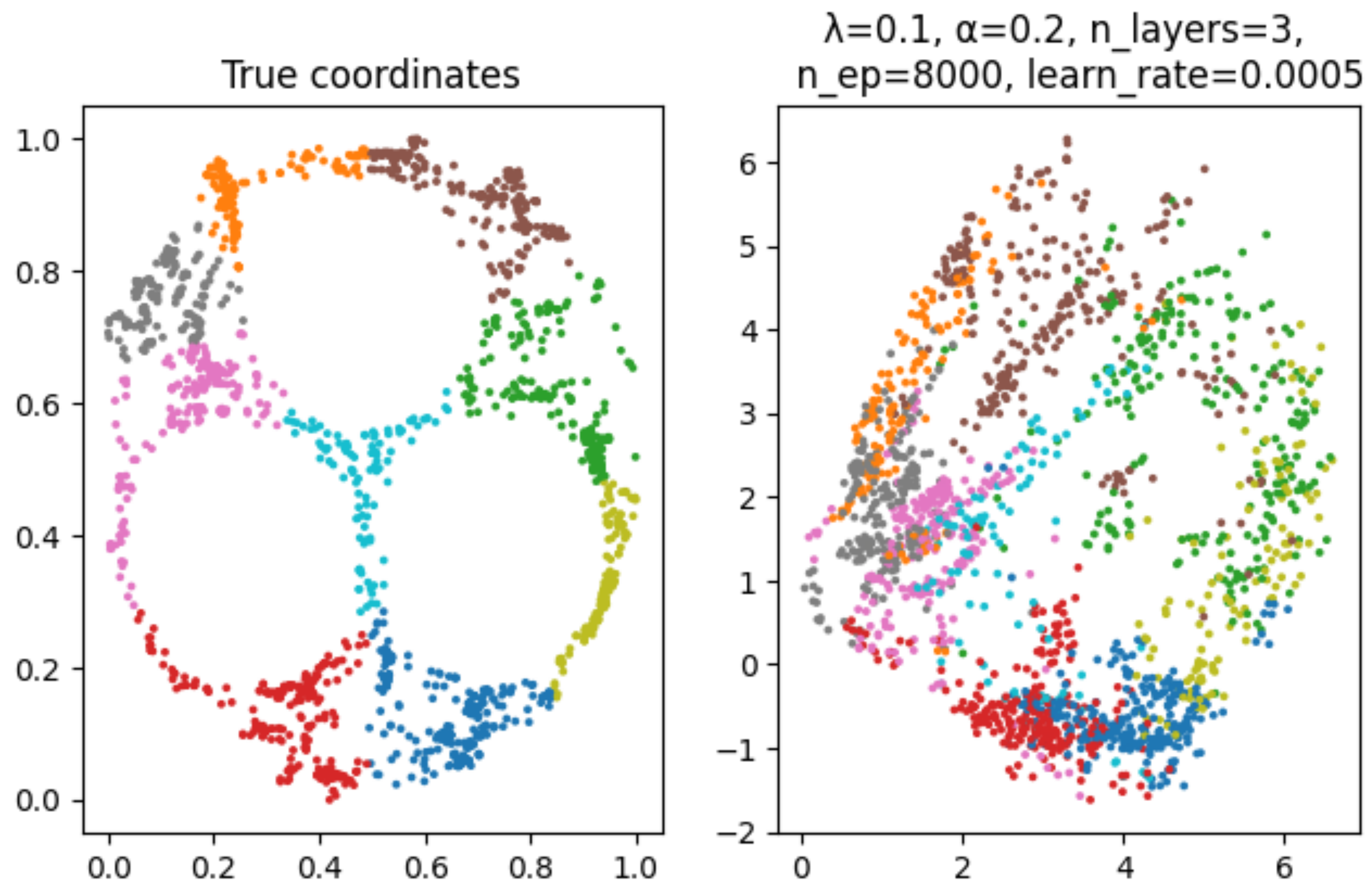




Topologically autoencoding

This is still work in progress!





Stay tuned! :)

https://cs.hse.ru/en/ata-lab/

maxbeketov@outlook.com

Thank you for your attention!