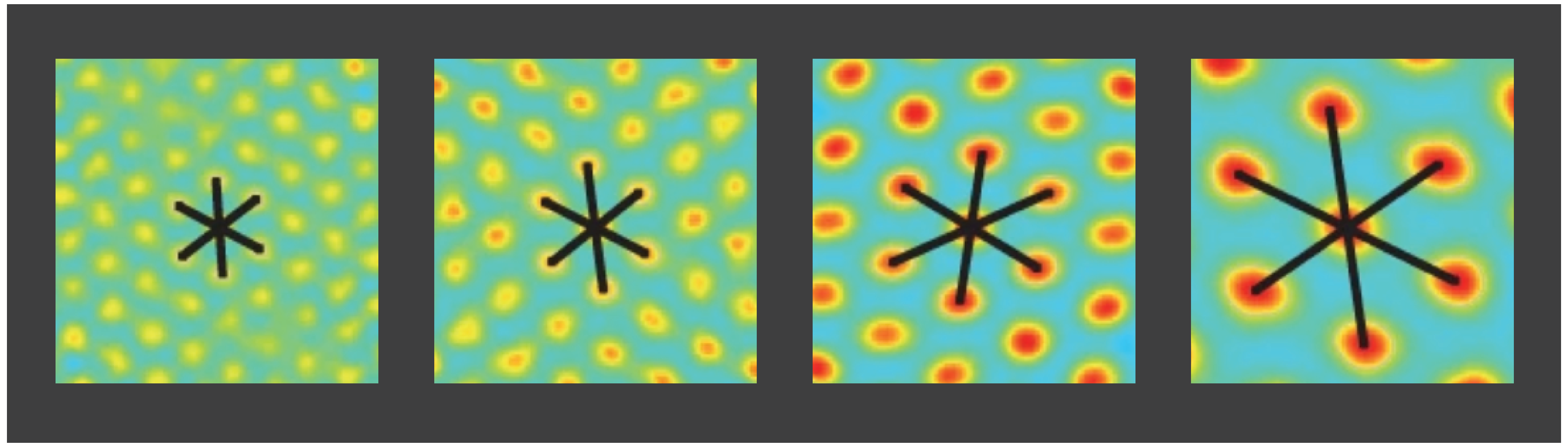


A self-supervised learning objective explains the ***modular*** organization of grid cells

Mikail Khona @KhonaMikail
(mikail@mit.edu)
Rylan Schaeffer @RylanSchaeffer
(rylanschaeffer@gmail.com)

Stanford | ENGINEERING
Computer Science

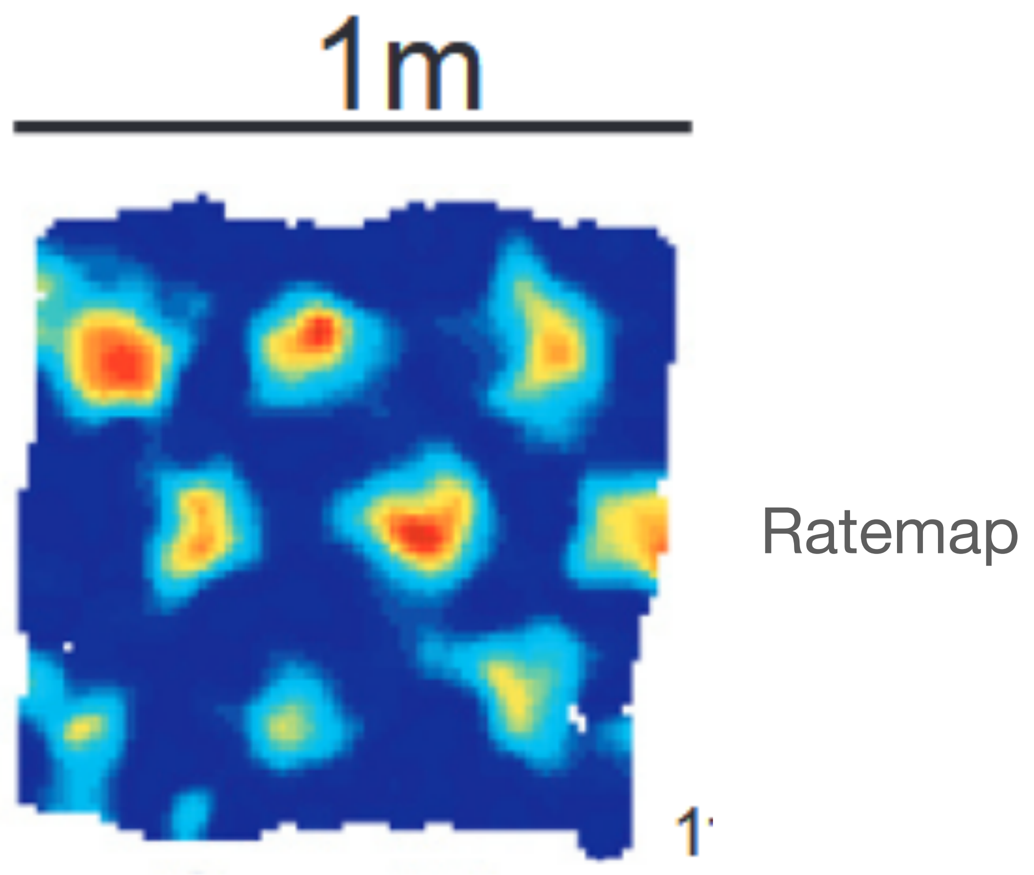
McGOVERN INSTITUTE | K. LISA YANG ICoN CENTER

MIT | brain+cognitive sciences

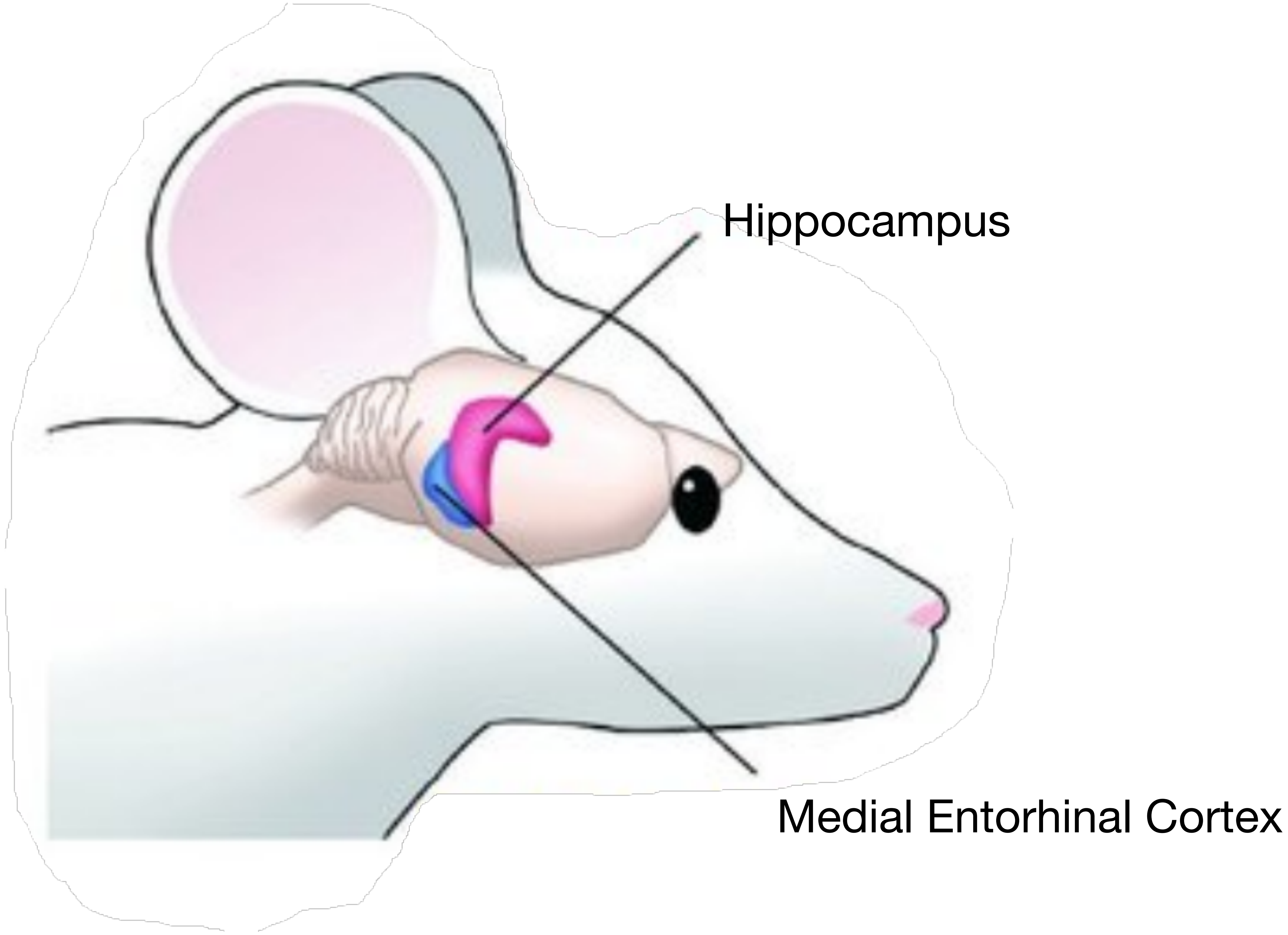
MIT Physics

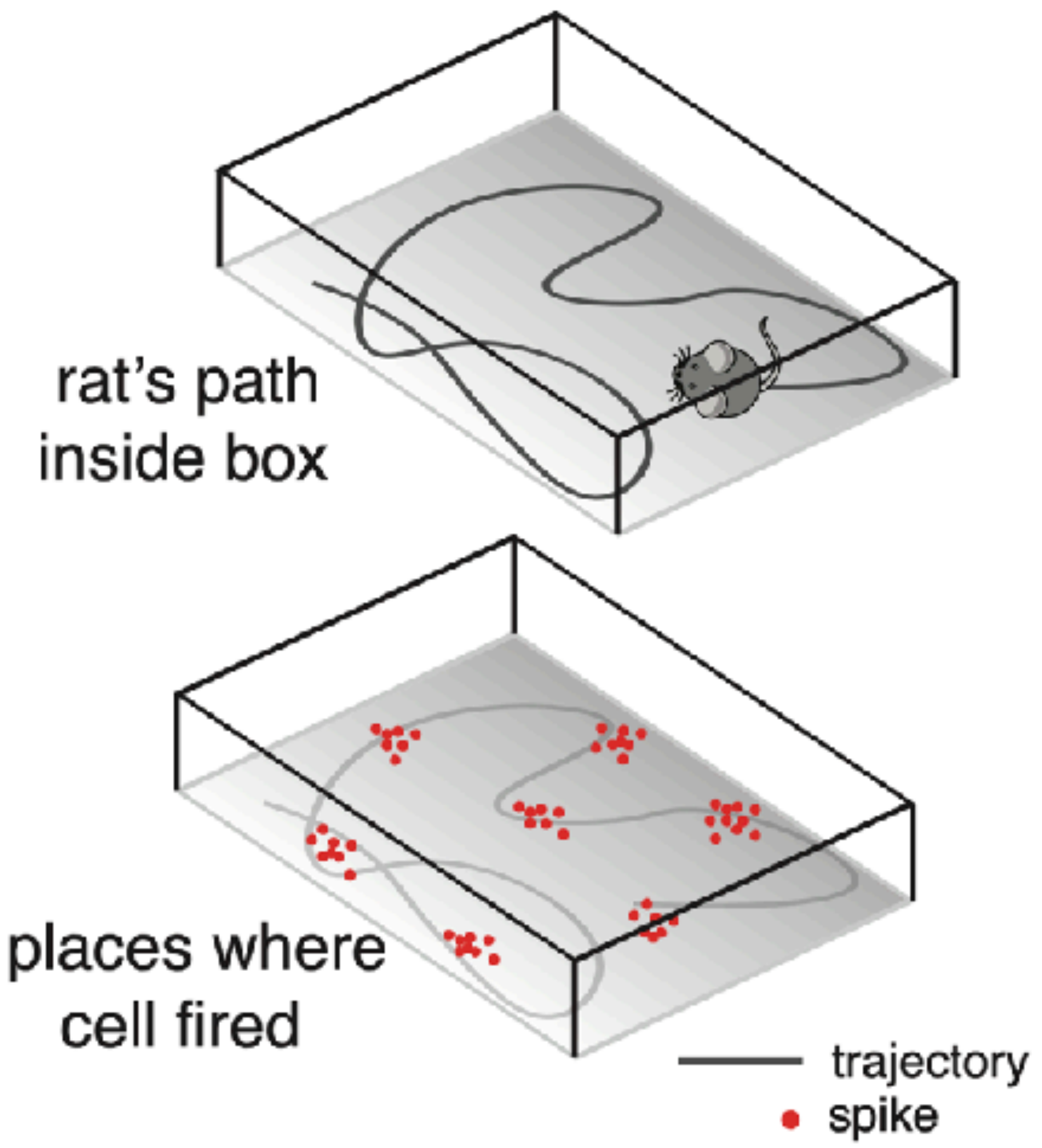
The grid cell system in the MEC

- Grid cells in the medial Entorhinal Cortex (mEC) keep track of **allocentric location** modulo a hexagonal lattice.



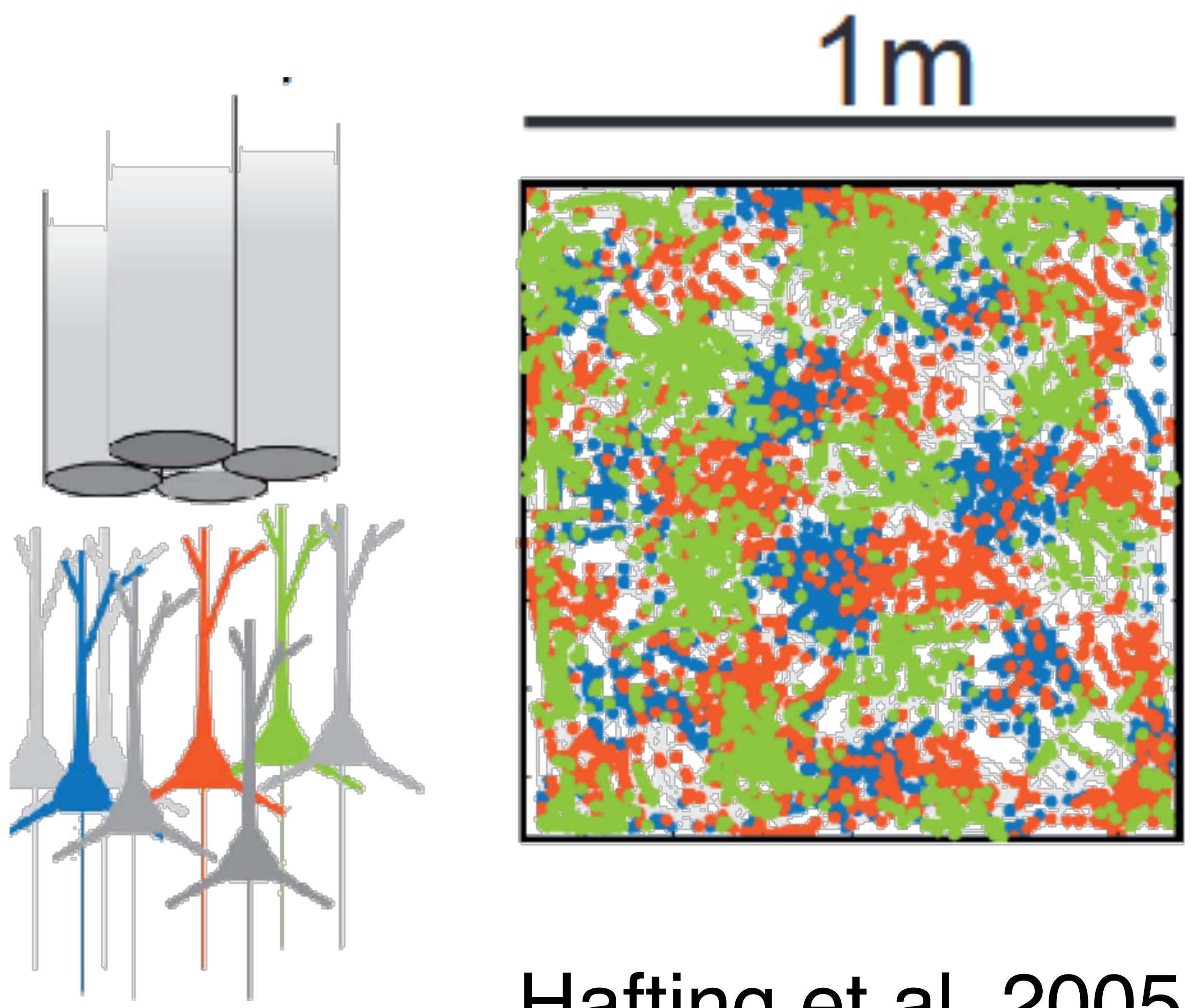
Hafting et al. 2005





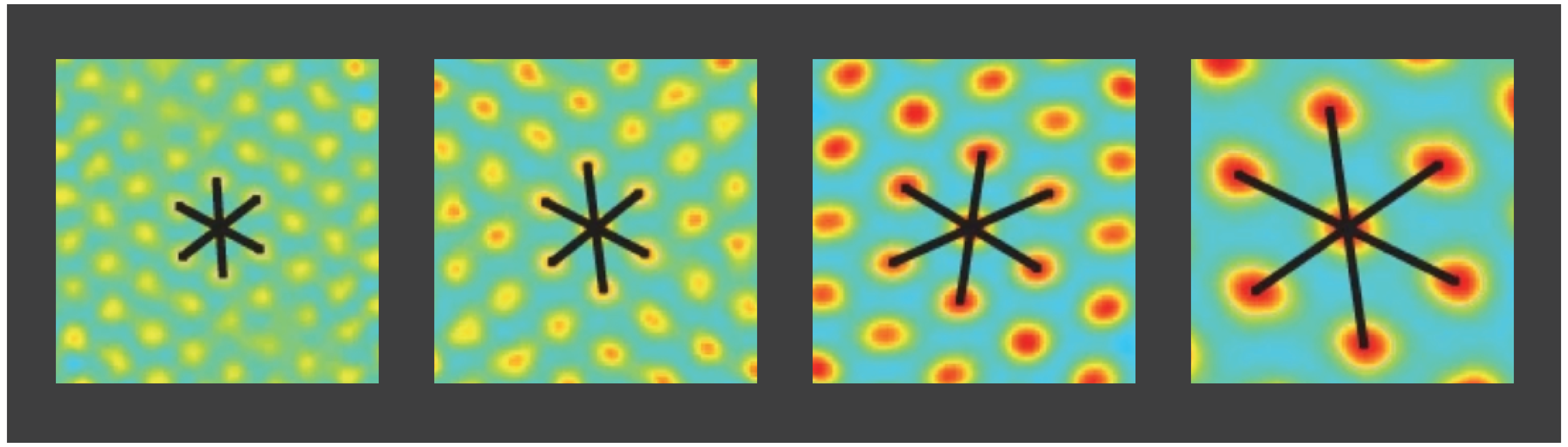
The grid cell system in the MEC

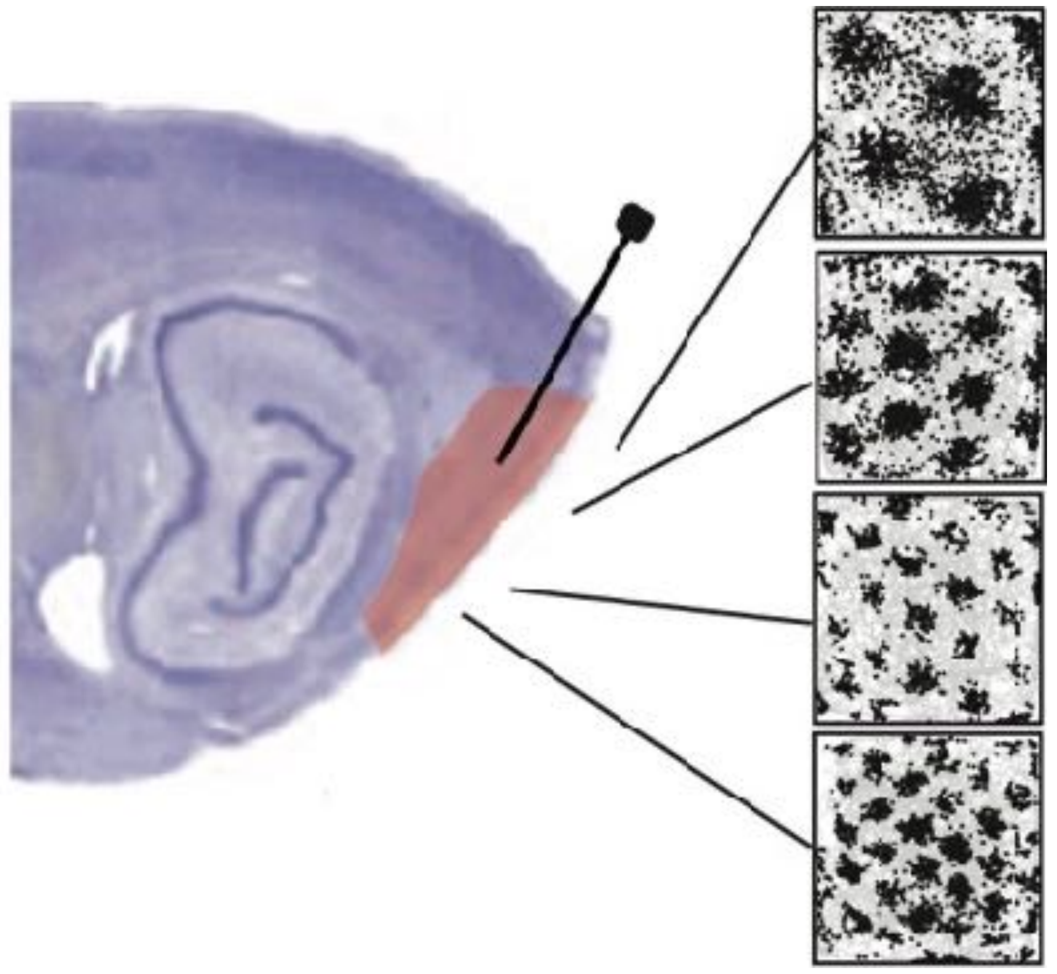
- Grid cells in the medial Entorhinal Cortex (mEC) keep track of **allocentric location** modulo a hexagonal lattice.
- Different grid cells keep track of this information with respect to lattices of different phases and lattice spacings.



Hafting et al. 2005

- Periodicity is arranged along dorso-ventral (DV) axis of mEC



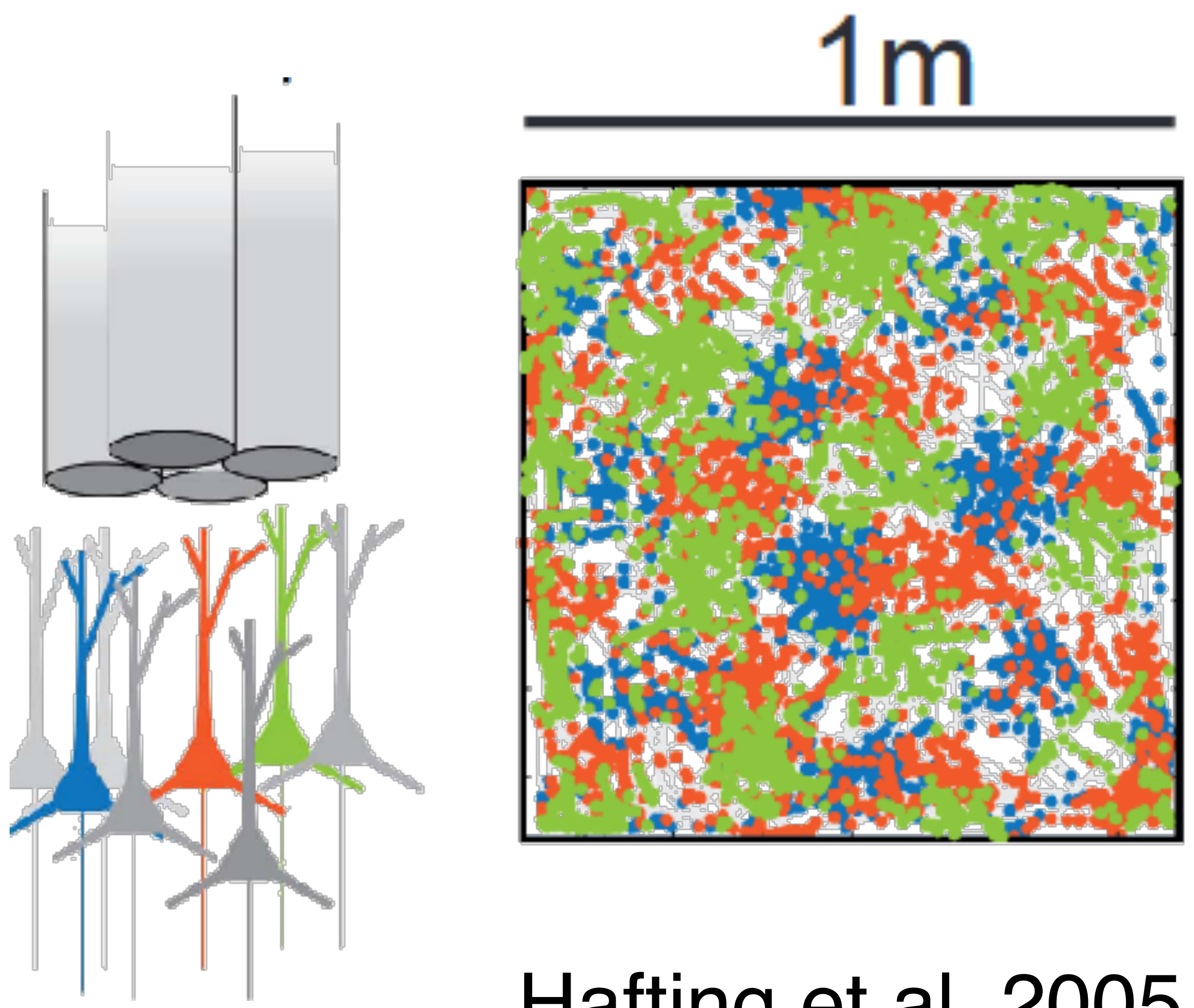


Stensola et al, The entorhinal grid map is discretized. Nature (2012)

The grid cell system in the MEC

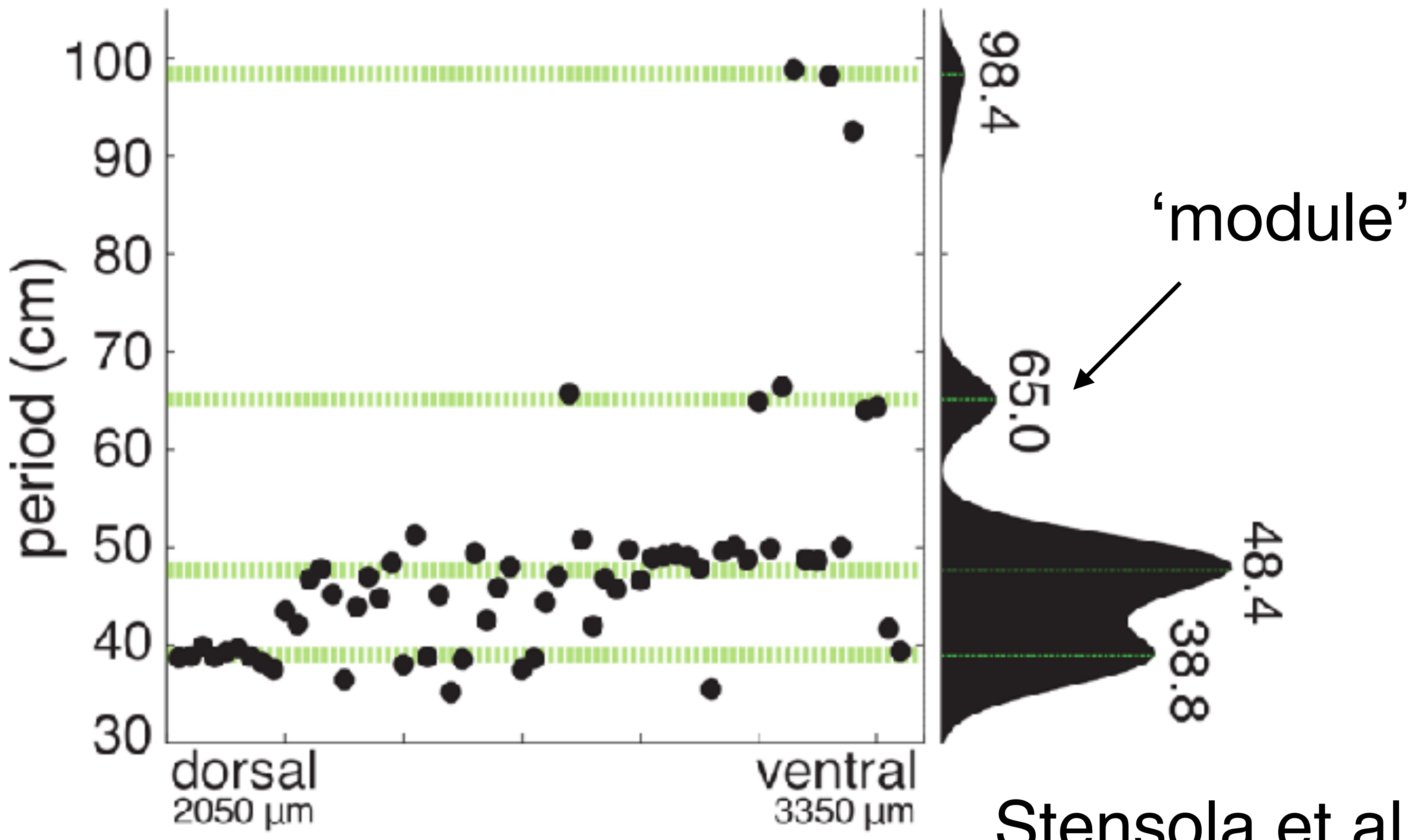
- Grid cells in the medial Entorhinal Cortex (mEC) keep track of **allocentric location** modulo a hexagonal lattice.

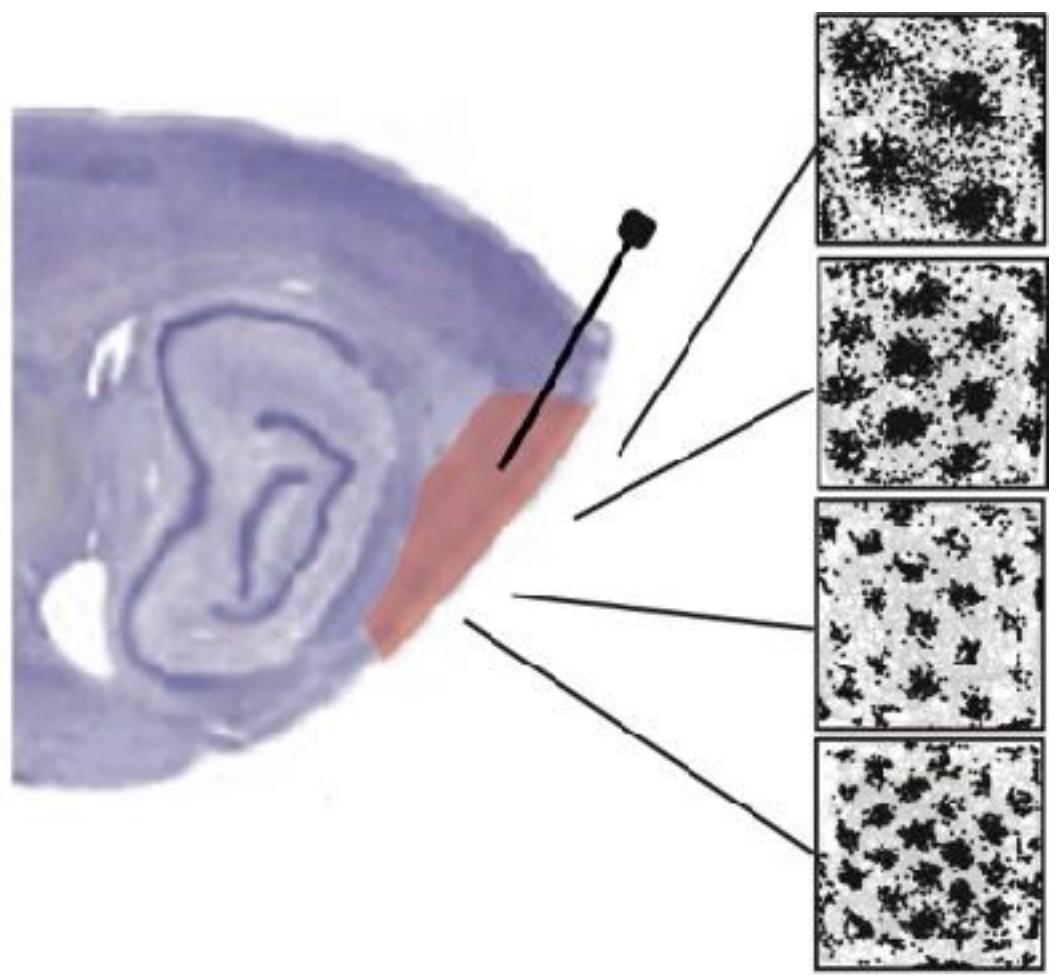
- Different grid cells keep track of this information with respect to lattices of different phases and lattice spacings.



Hafting et al. 2005

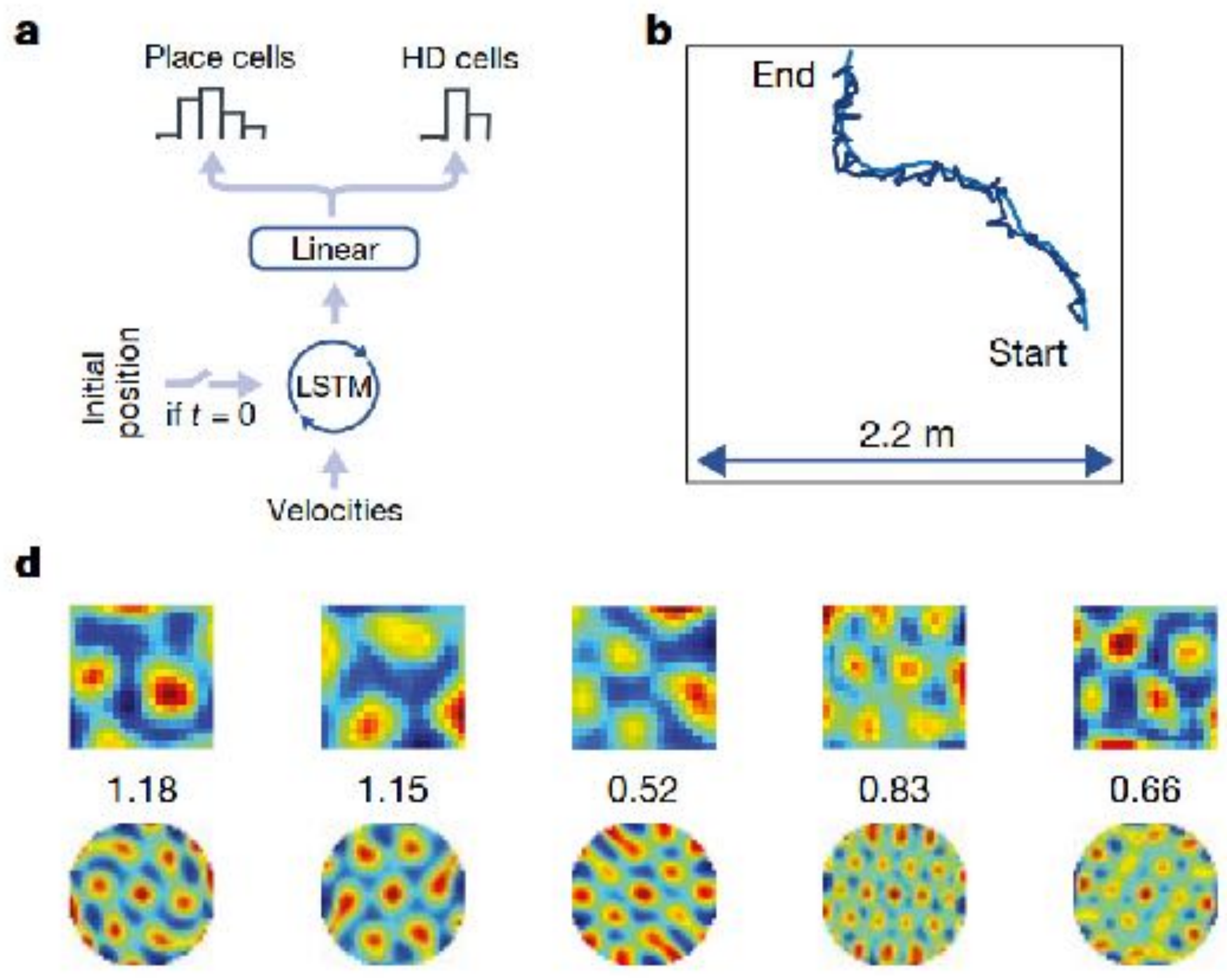
- Periodicity is arranged along dorso-ventral (DV) axis of mEC



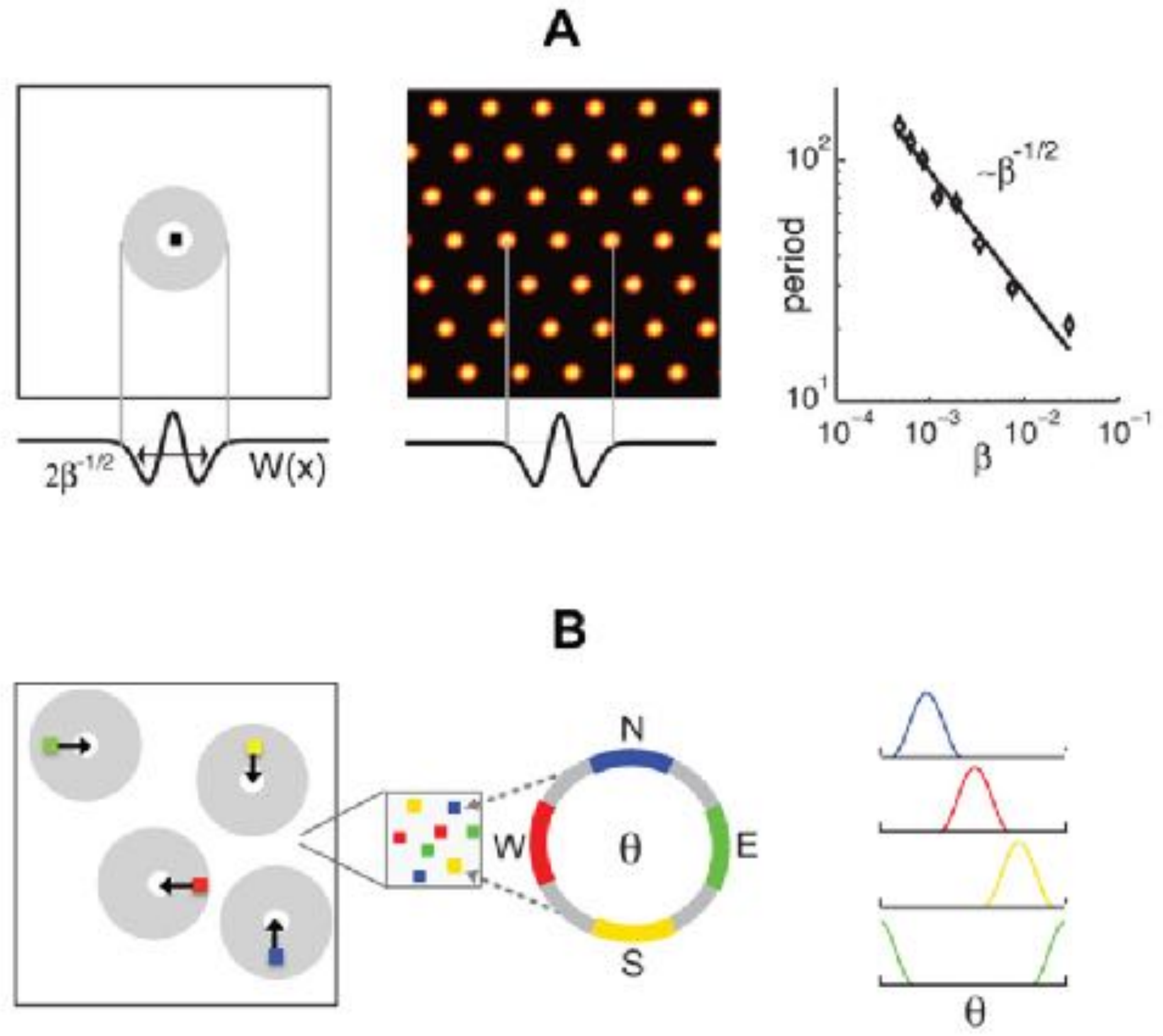


Stensola et al, The entorhinal grid map is discretized. Nature (2012)

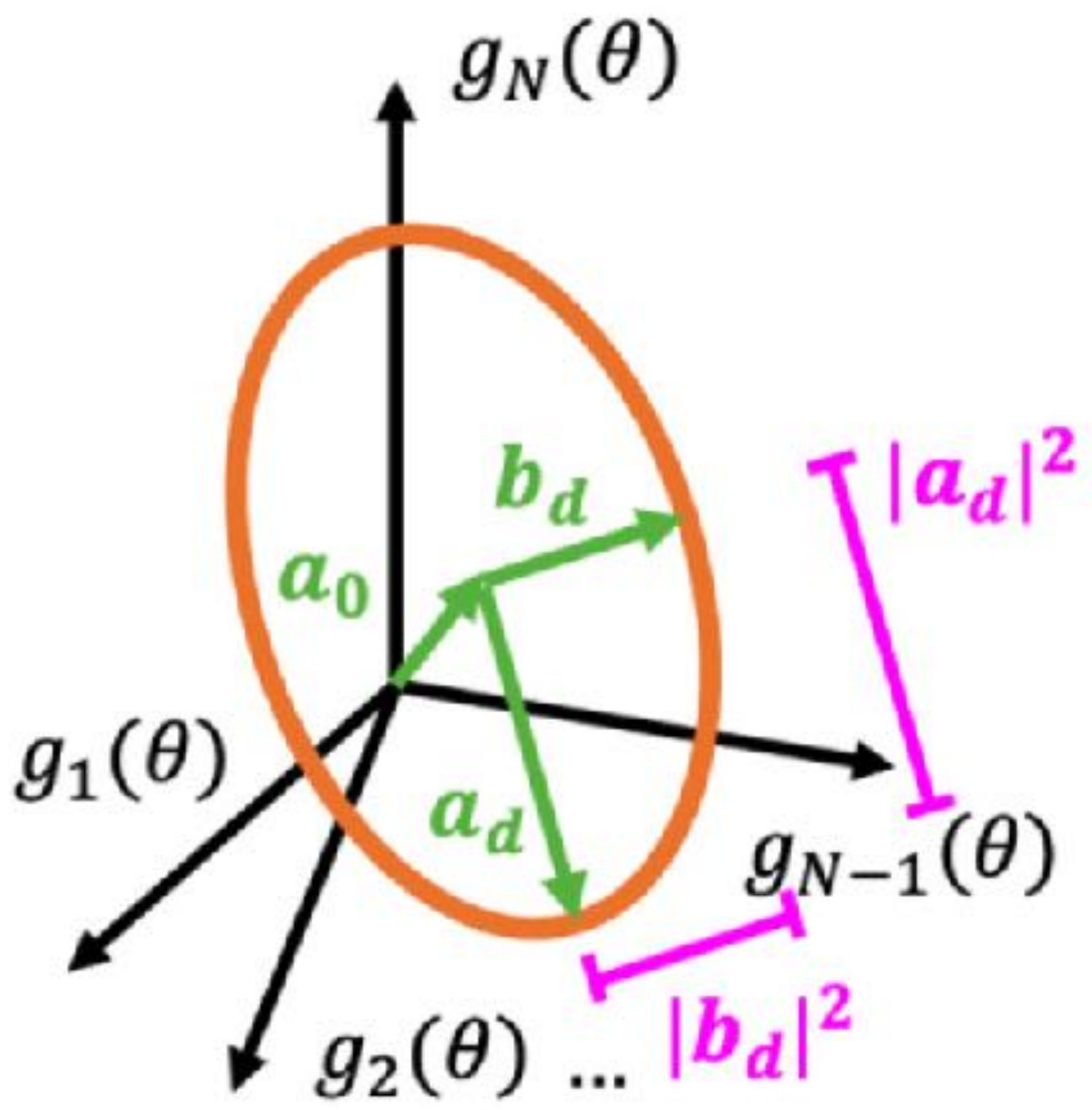
Supervised learning



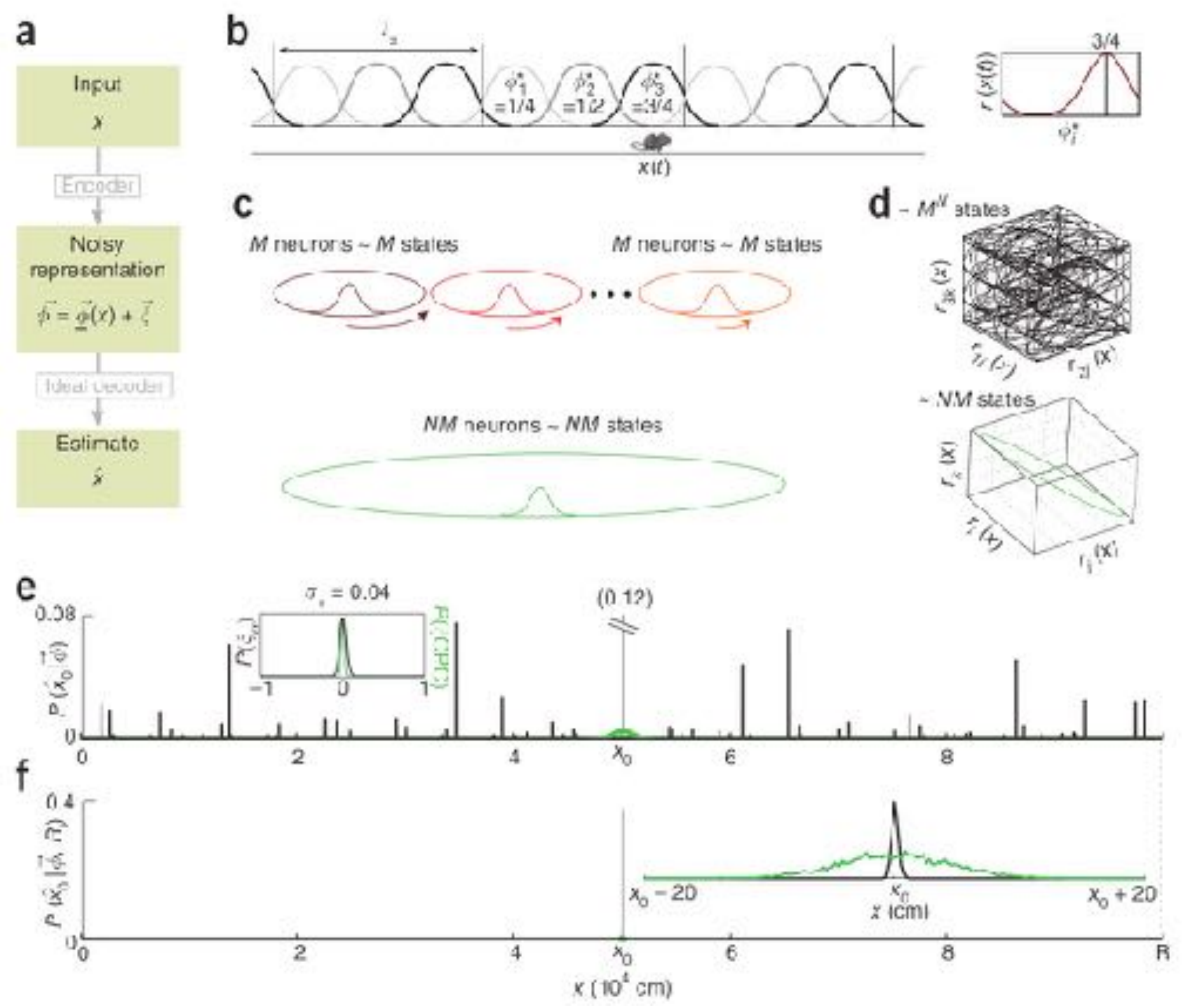
Continuous attractors



Basis function optimization

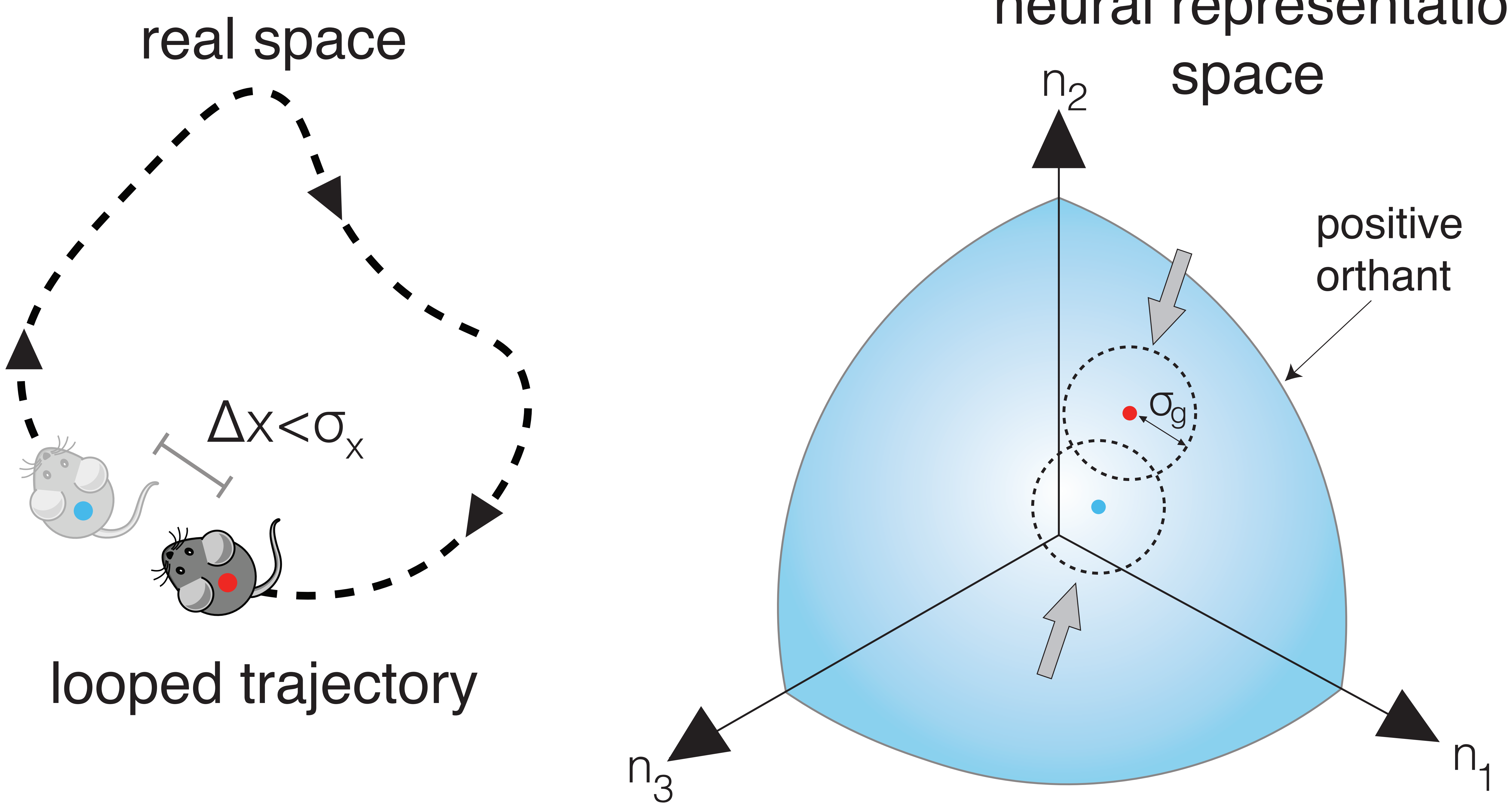


Coding theory



We can use these insights to formulate a self-supervised learning SSL problem

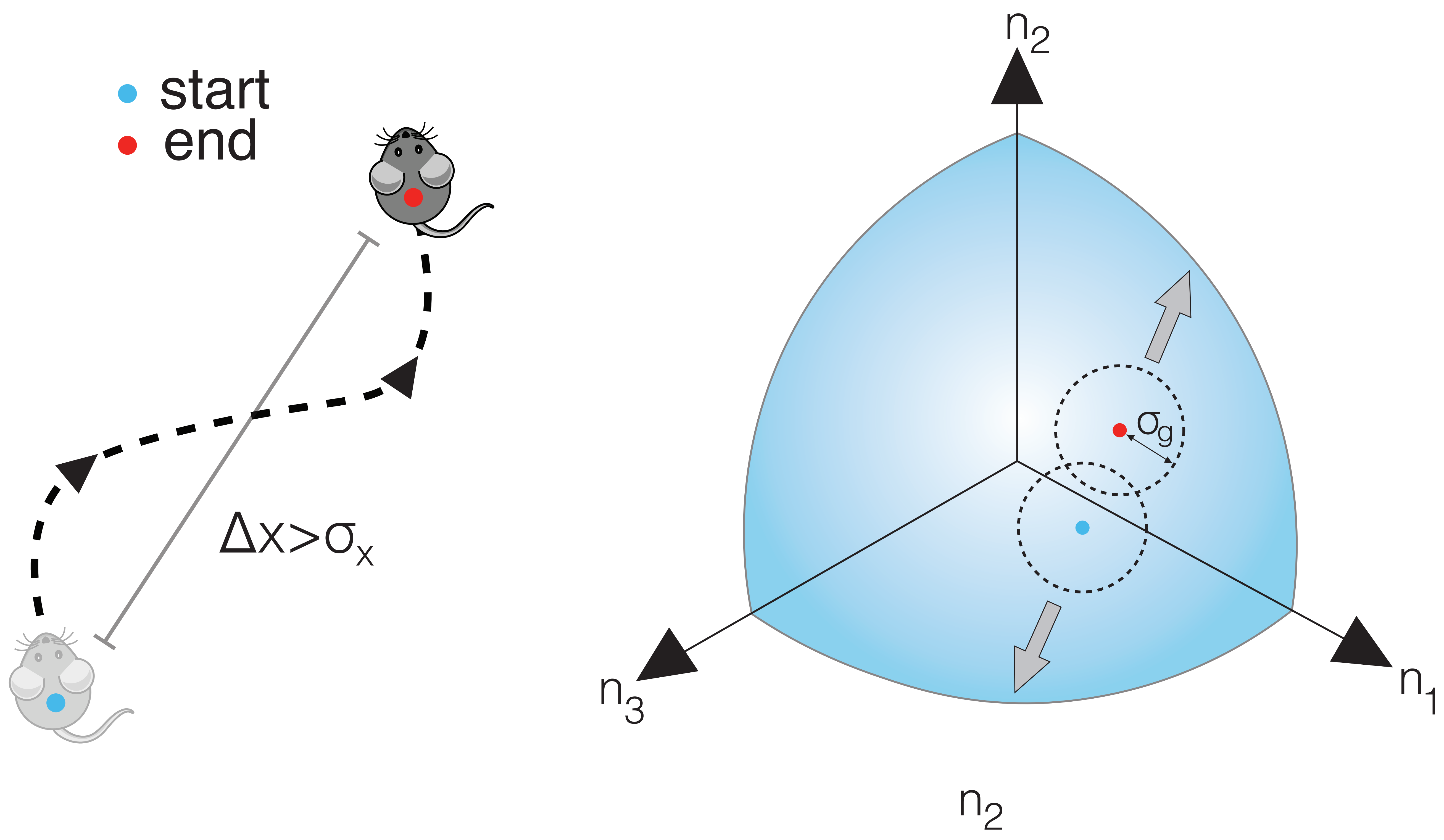
Path Invariance



This ‘loop closure’ property is needed for path integration

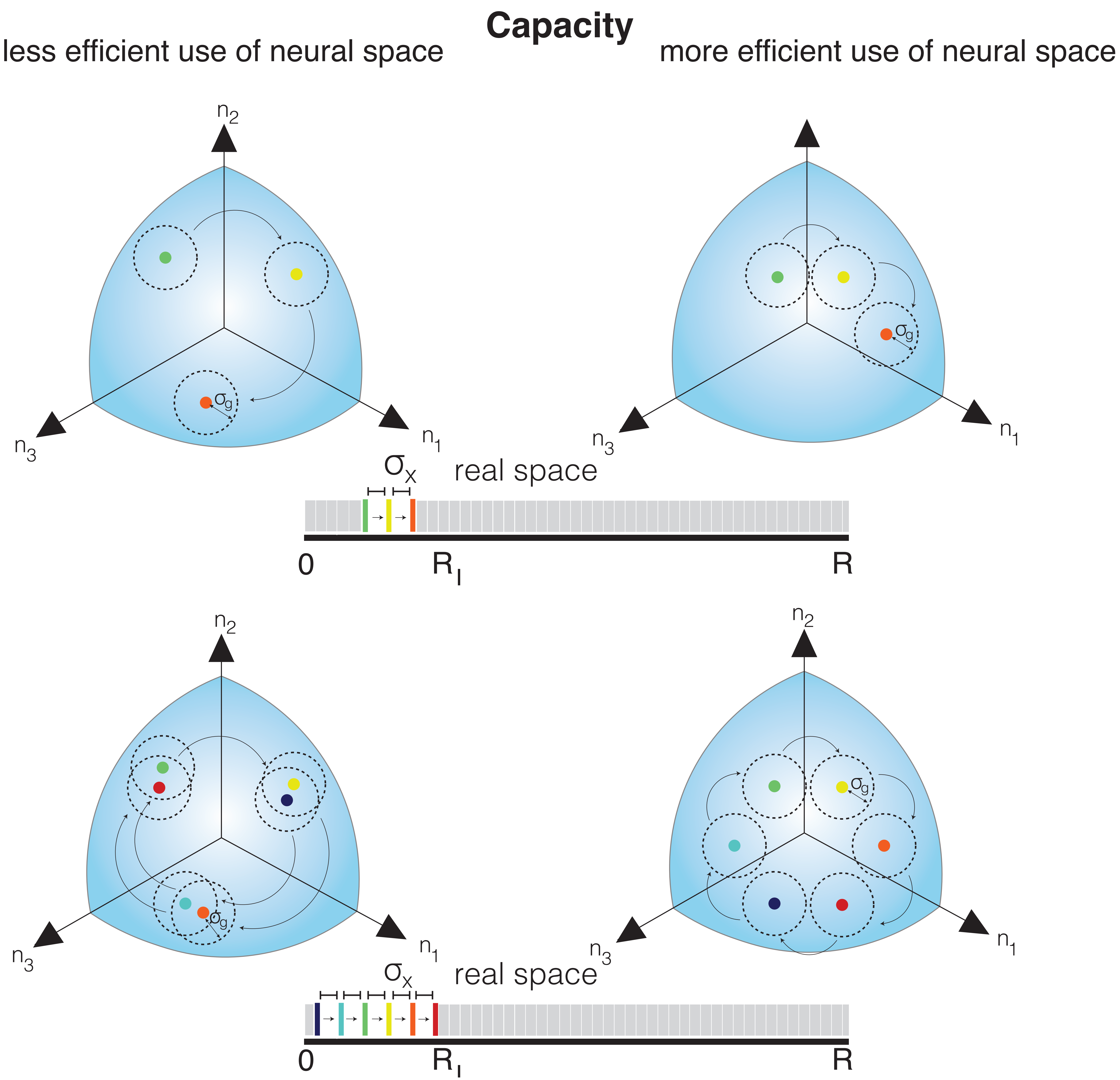
We can use these insights to formulate a self-supervised learning SSL problem

Separation



See also: Dorrell et al (2023)

We can use these insights to formulate a self-supervised learning SSL problem

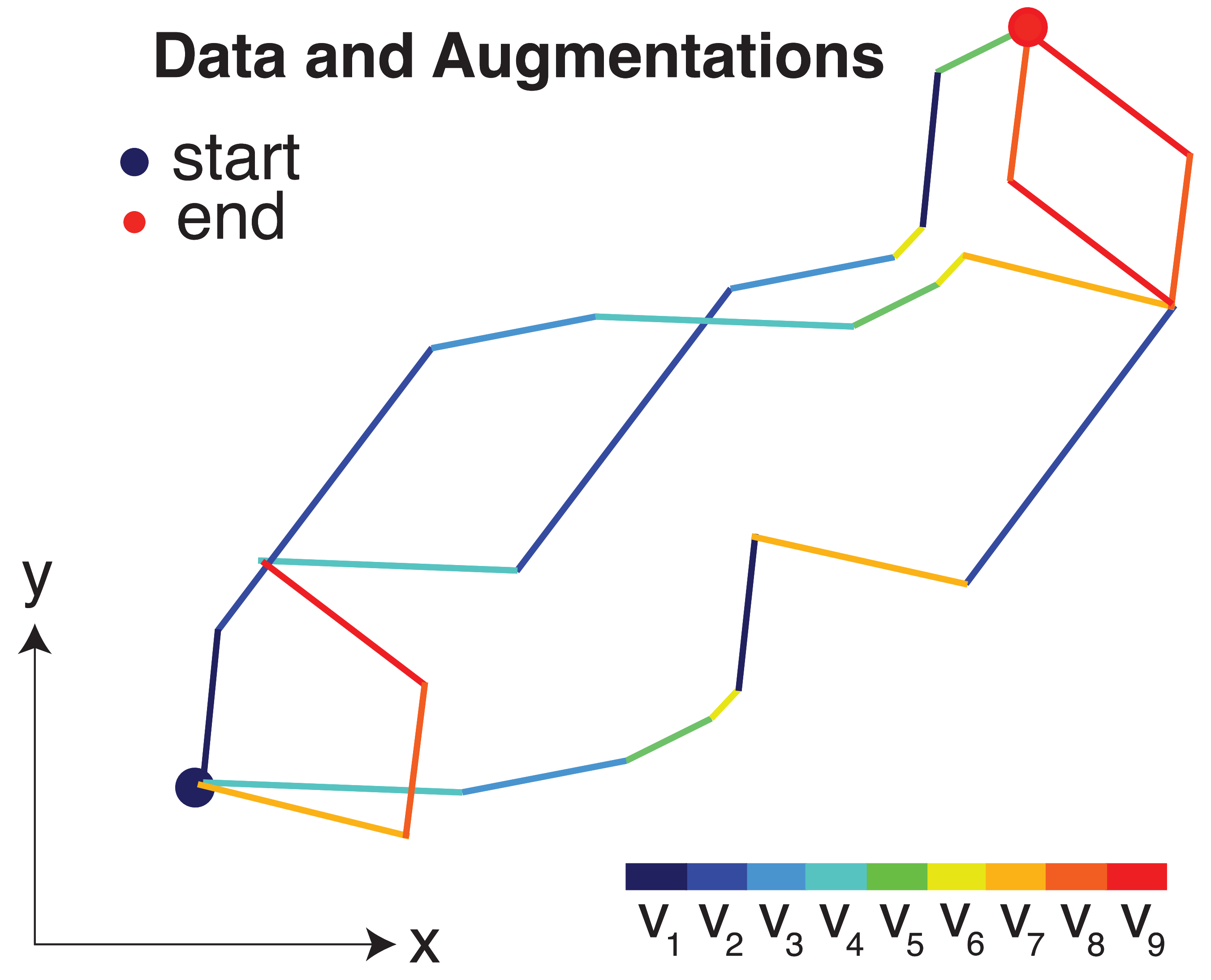


Extending the self-supervised learning SSL problem to spatial navigation

To create a trajectory, we sample T velocities $(\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_T)$, with $\mathbf{v}_t \sim_{i.i.d.} p(\mathbf{v})$

To create a batch, we create B random permutations: $\{\pi_b\}_{b=1}^{B}$

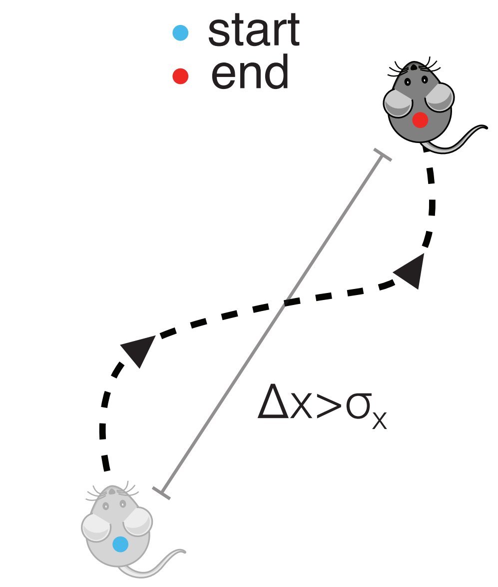
$$\pi_b : [T] \rightarrow [T]$$



Formulating a self-supervised learning SSL problem: loss functions

$$\mathcal{D}_{\text{gradient step}} = \left\{ \left(\mathbf{v}_{\pi_b(1)}, \mathbf{v}_{\pi_b(2)}, \ldots, \mathbf{v}_{\pi_b(T)} \right), \left(\mathbf{g}_{\pi_b(1)}, \mathbf{g}_{\pi_b(2)}, \ldots, \mathbf{g}_{\pi_b(T)} \right) \right\}_{b=1}^{B}$$

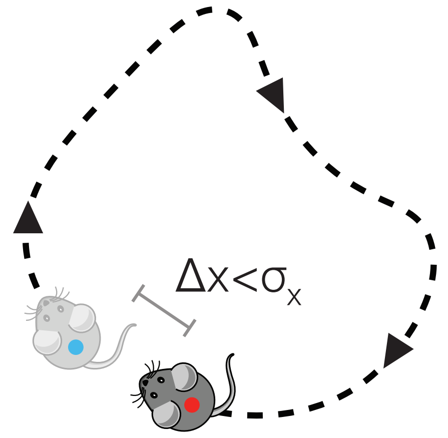
with shared initial state $\mathbf{g}_0$



real space

$$\mathcal{L}_{Sep} = \sum_{\substack{\forall \pi_b, \pi_{b'}, t, t': \\ ||\mathbf{x}_{\pi_{b'}(t)} - \mathbf{x}_{\pi_b(t')}||_2 > \sigma_x}} \exp\left(- \frac{||\mathbf{g}_{\pi_{b'}(t)} - \mathbf{g}_{\pi_b(t')}||_2^2}{2\sigma_g^2} \right)$$

1 coarse grained bit of information about *relative*, and not absolute, spatial location



looped trajectory

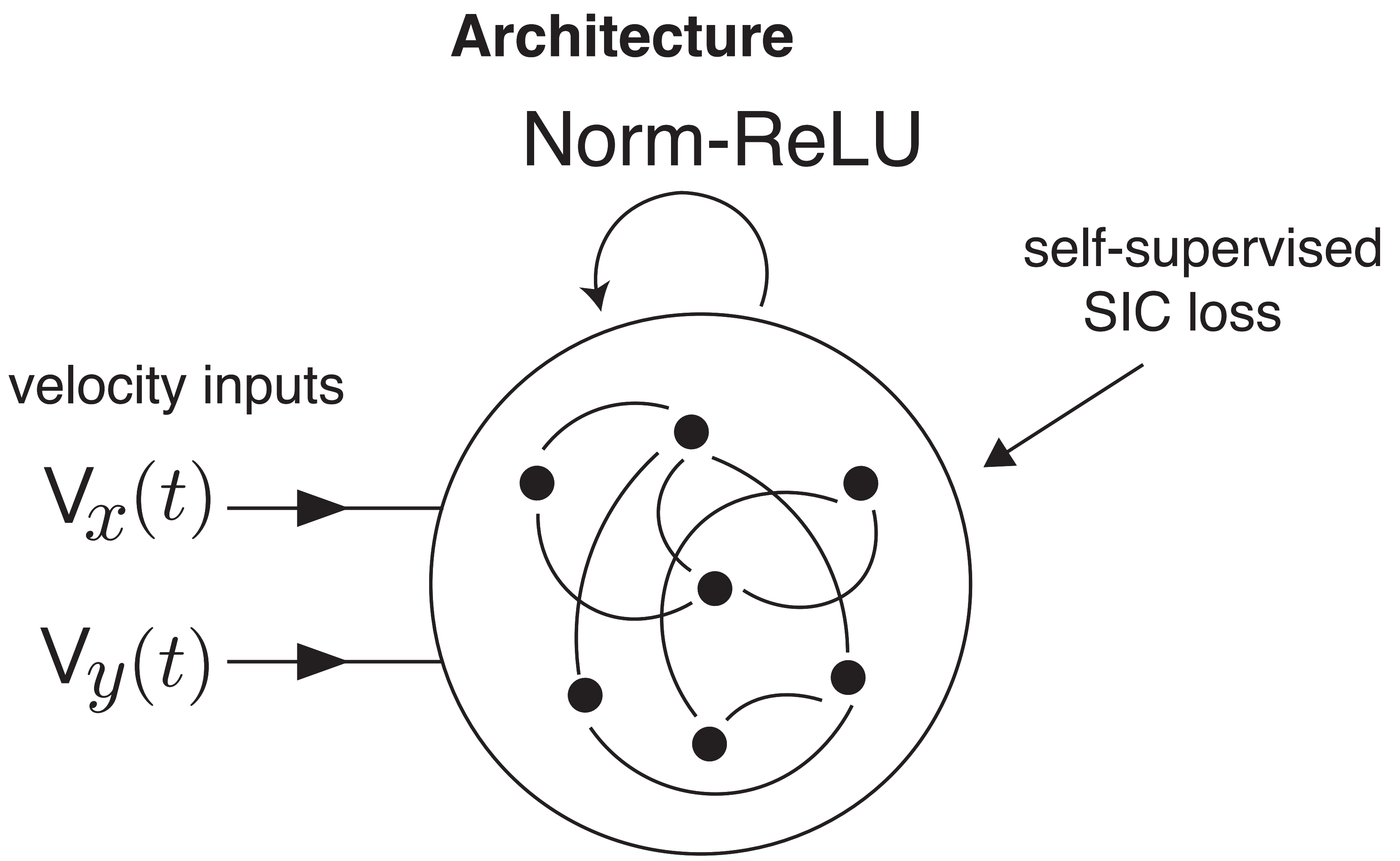
$$\mathcal{L}_{Inv} = \sum_{\substack{\forall \pi_b, \pi_{b'}, t, t': \\ ||\mathbf{x}_{\pi_{b'}(t)} - \mathbf{x}_{\pi_b(t')}||_2 < \sigma_x}} \left|\left| \mathbf{g}_{\pi_b(t)} - \mathbf{g}_{\pi_{b'}(t')} \right|\right|_2^2$$

Contrast with supervised approaches that provide *absolute* spatial information at *all* times

Regularization $$\mathcal{L}_{Cap} = - \left|\left| \frac{1}{BT} \sum_{\pi_b, t} \mathbf{g}_{\pi_b(t)} \right|\right|_2^2$$

Regularization $$\mathcal{L}_{ConIso} \stackrel{\text{def}}{=} \mathbb{V}\left[\left\{ \frac{||g_t - g_{t-1}||}{||v_t||} \right\}_{t\,:\,0<||v_t||<\sigma_x} \right]$$

Formulating a self-supervised learning SSL problem: architecture



$$W(\mathbf{v}_t) = MLP(\mathbf{v}_t)$$

$$\mathbf{g}_t = \sigma(W(\mathbf{v}_t)\,\mathbf{g}_{t-1})$$

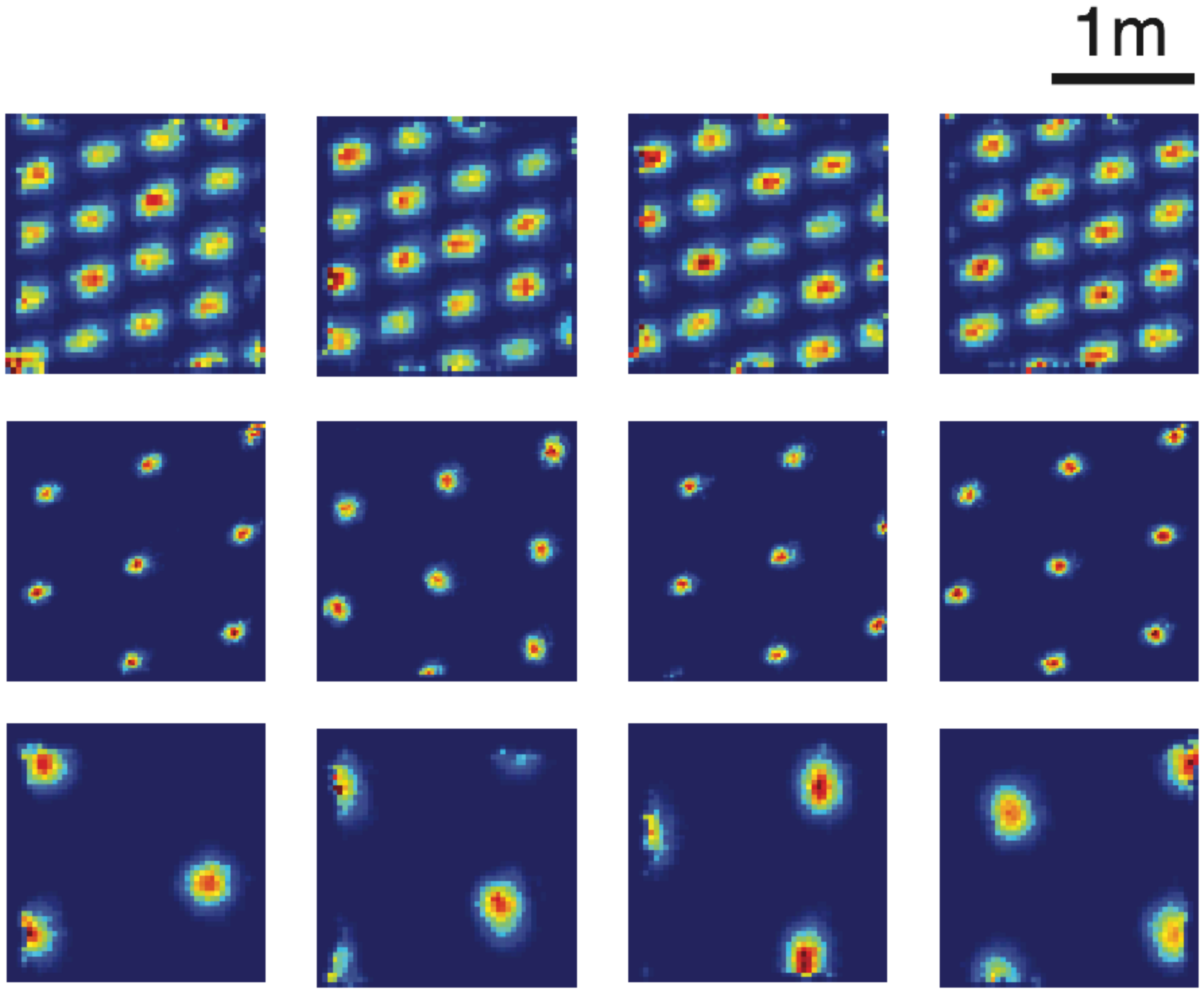
$$\sigma(\cdot) = Norm(ReLU(\cdot)) = ReLU(\cdot)\,/\,||ReLU(\cdot)||$$

Normalization of neural population activity: prevent trivial solutions often found by contrastive SSL

Understanding Dimensional Collapse in Contrastive Self-supervised Learning. Jing et al (2021)

What shapes the loss landscape of self-supervised learning? Liu et al (2023)

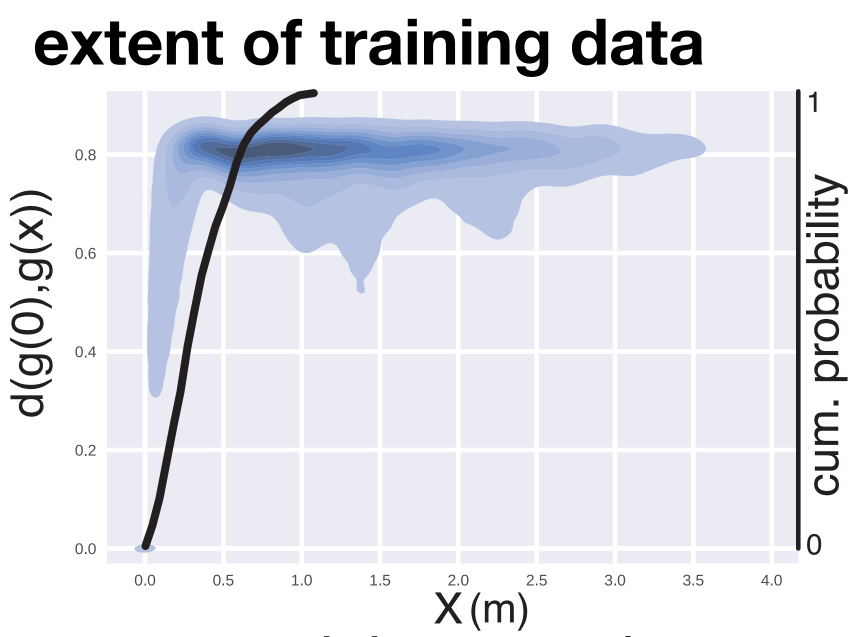
Result: It is *possible* to get multi-periodic grid-like solutions!

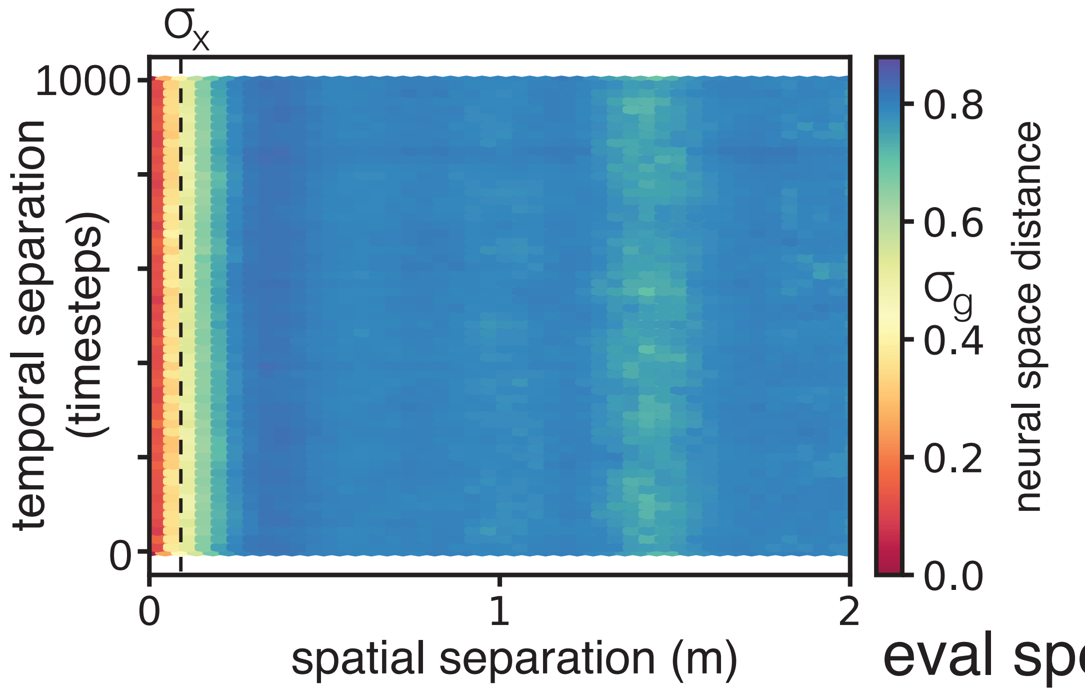


12 cells from a 128 neuron RNN

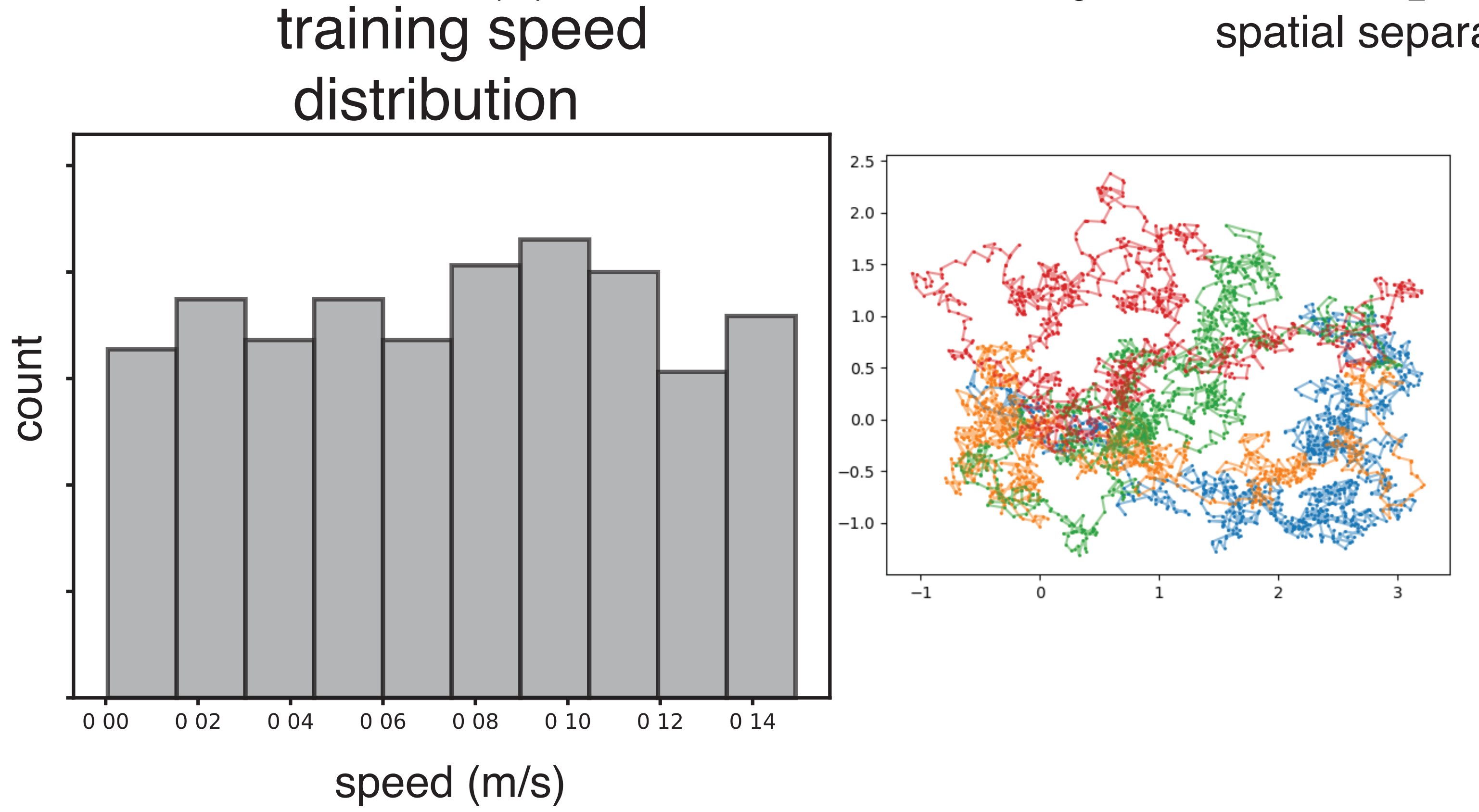
Solutions generalize to larger environments and distinct input statistics without any additional training

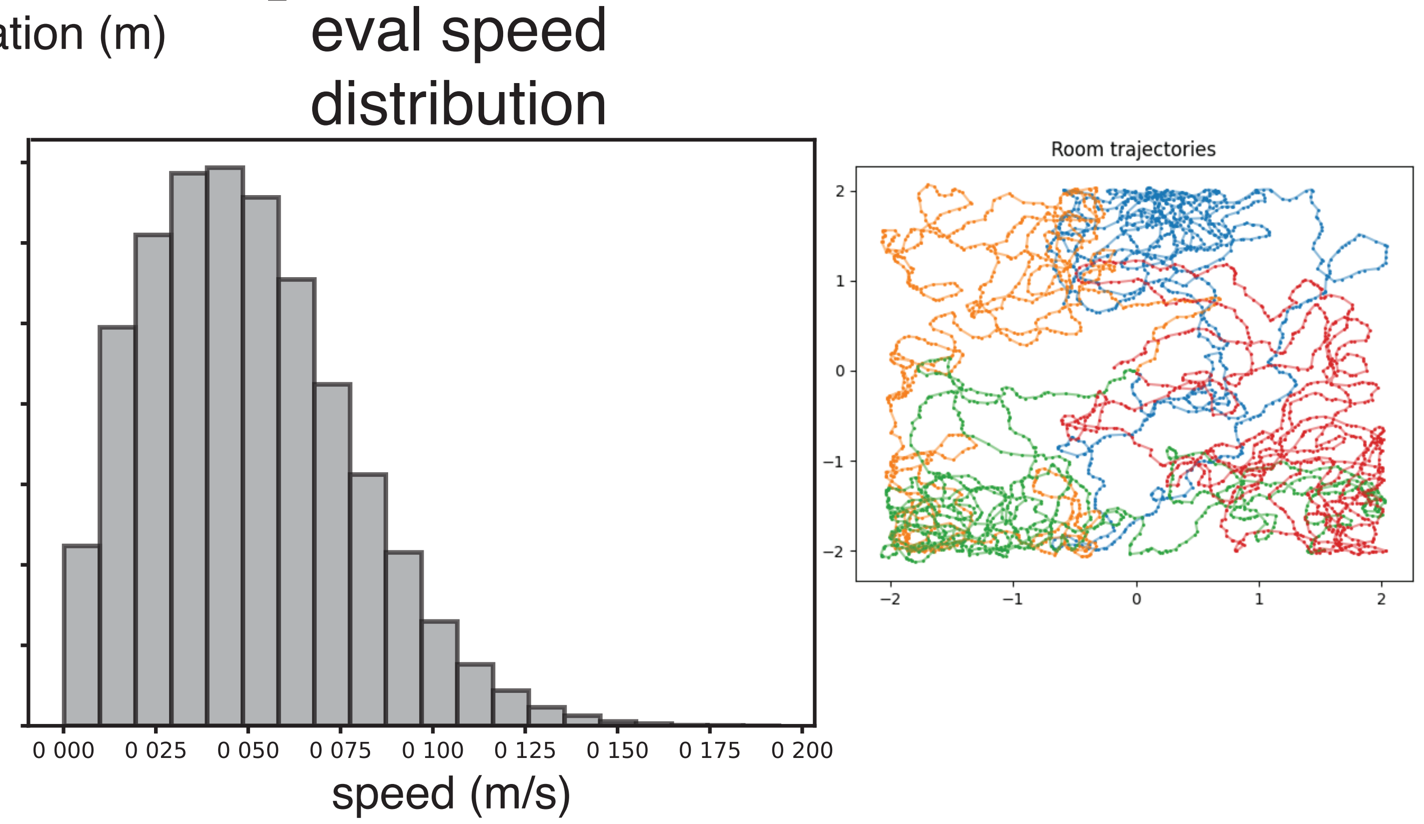
Out of distribution generalization

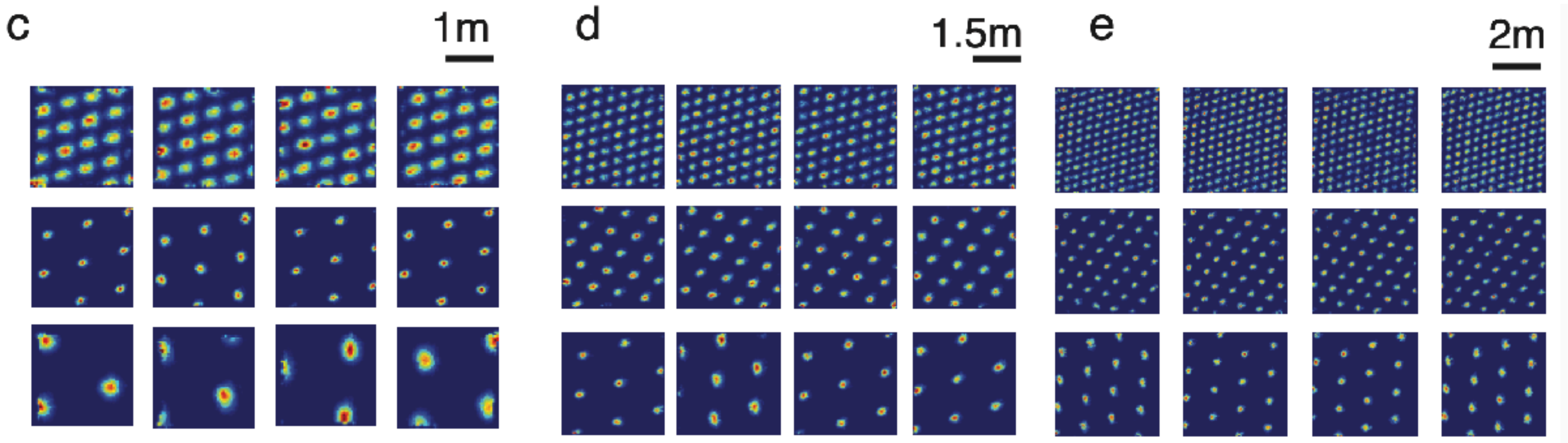




Perfect path integration

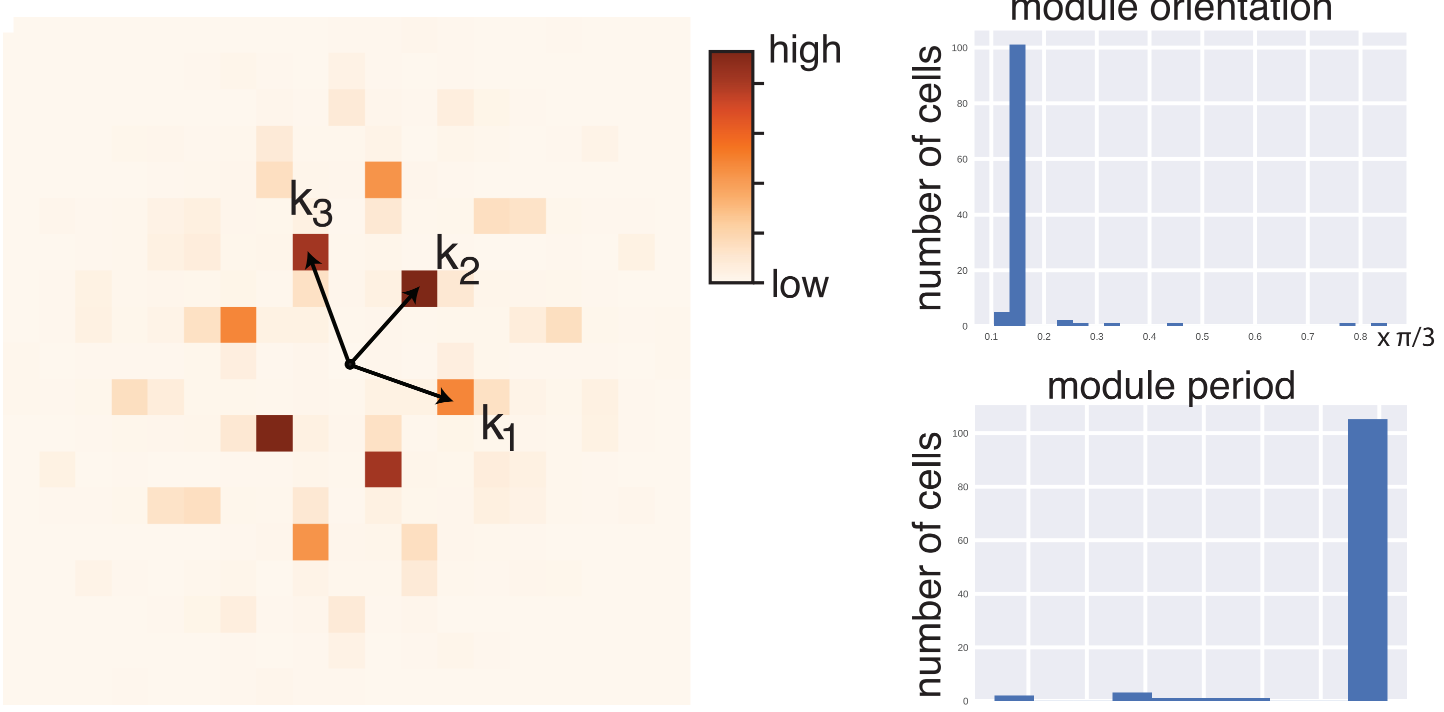




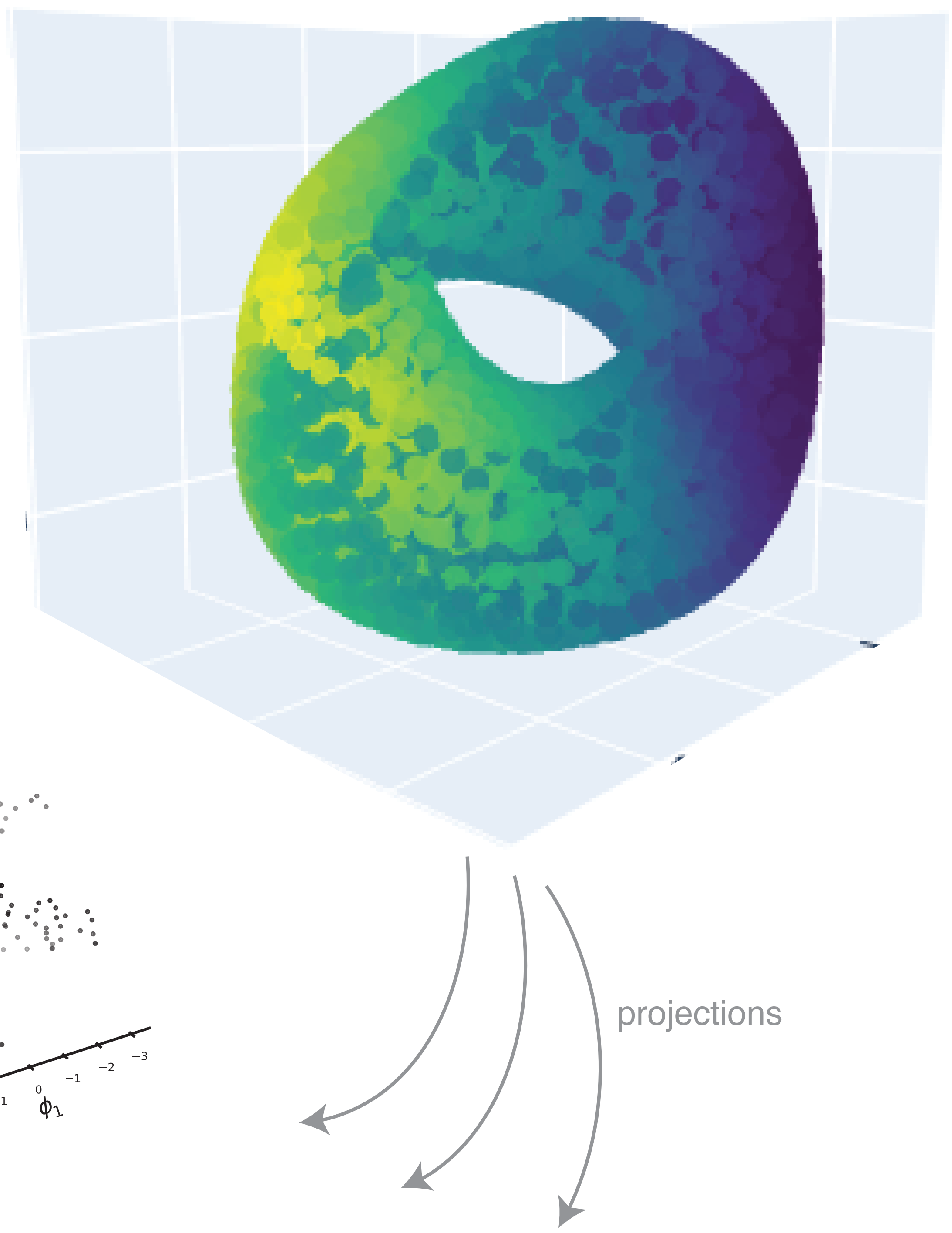


Dissecting a single module shows key properties of grid cells

mean fourier power

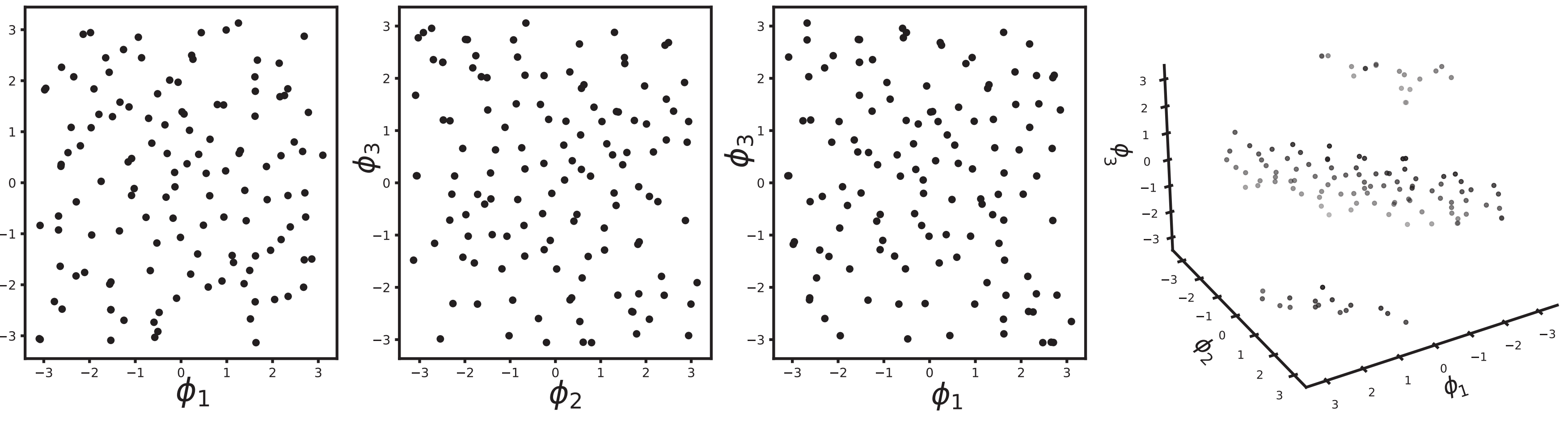


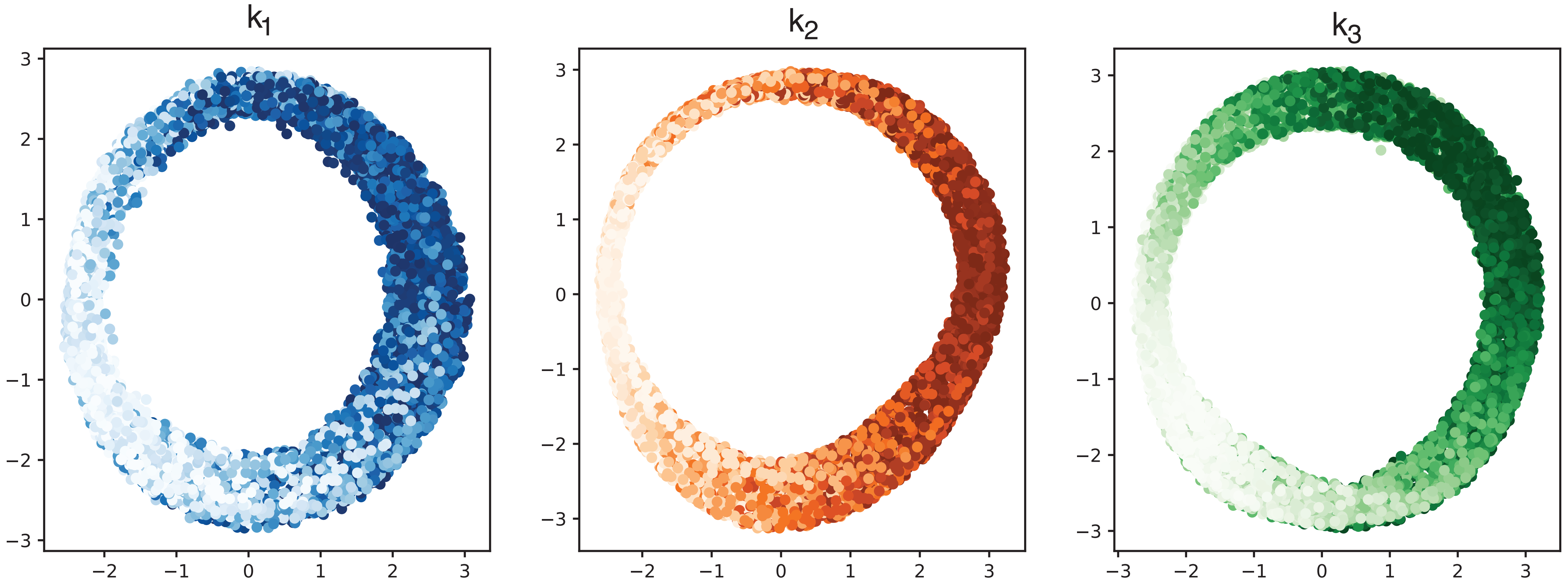
toroidal state space



Gardner et al Nature (2022)

phase coverage





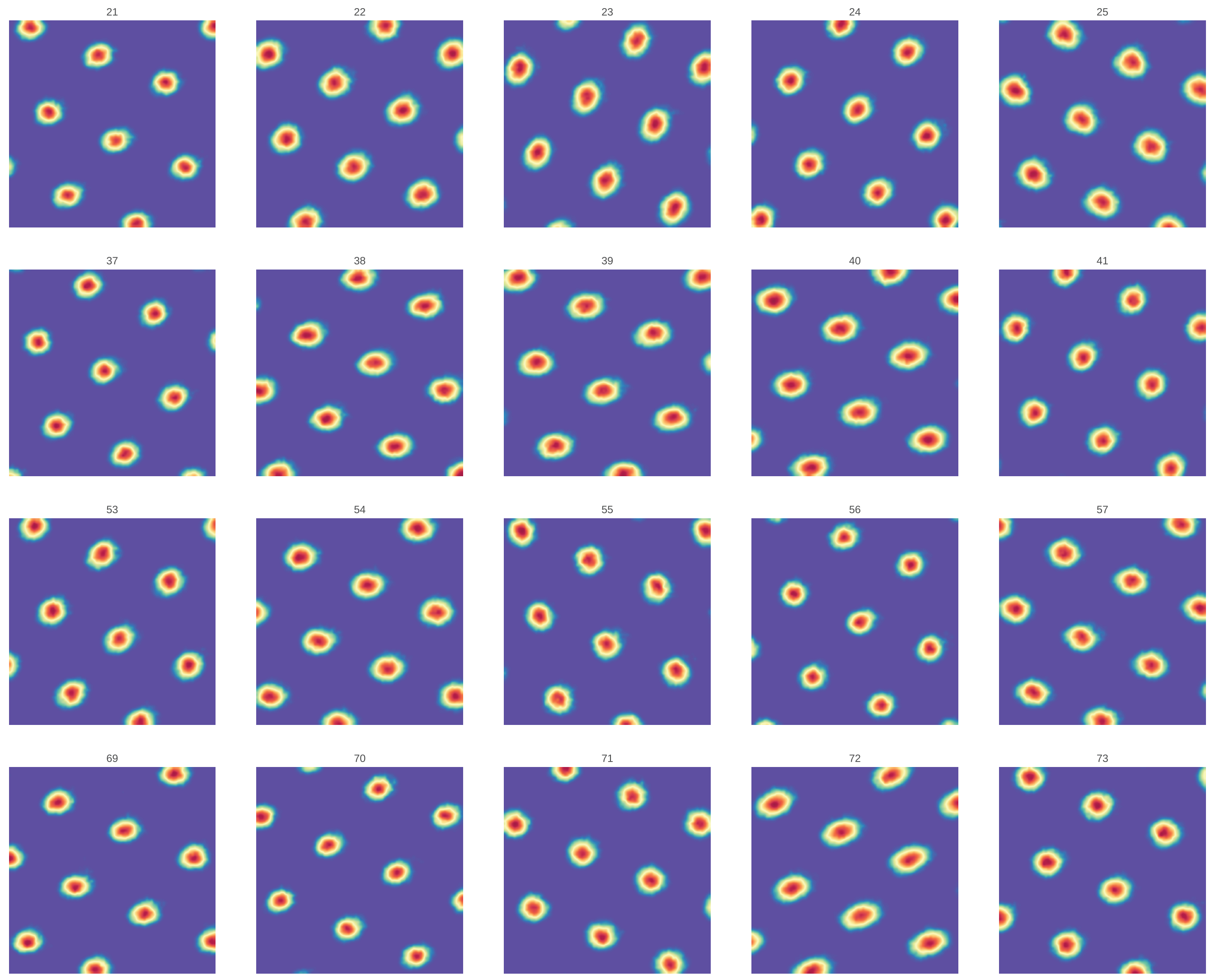
Signature of high dimensional 'twisted torus' topology!

Sorscher*, Mel* et al (2023)

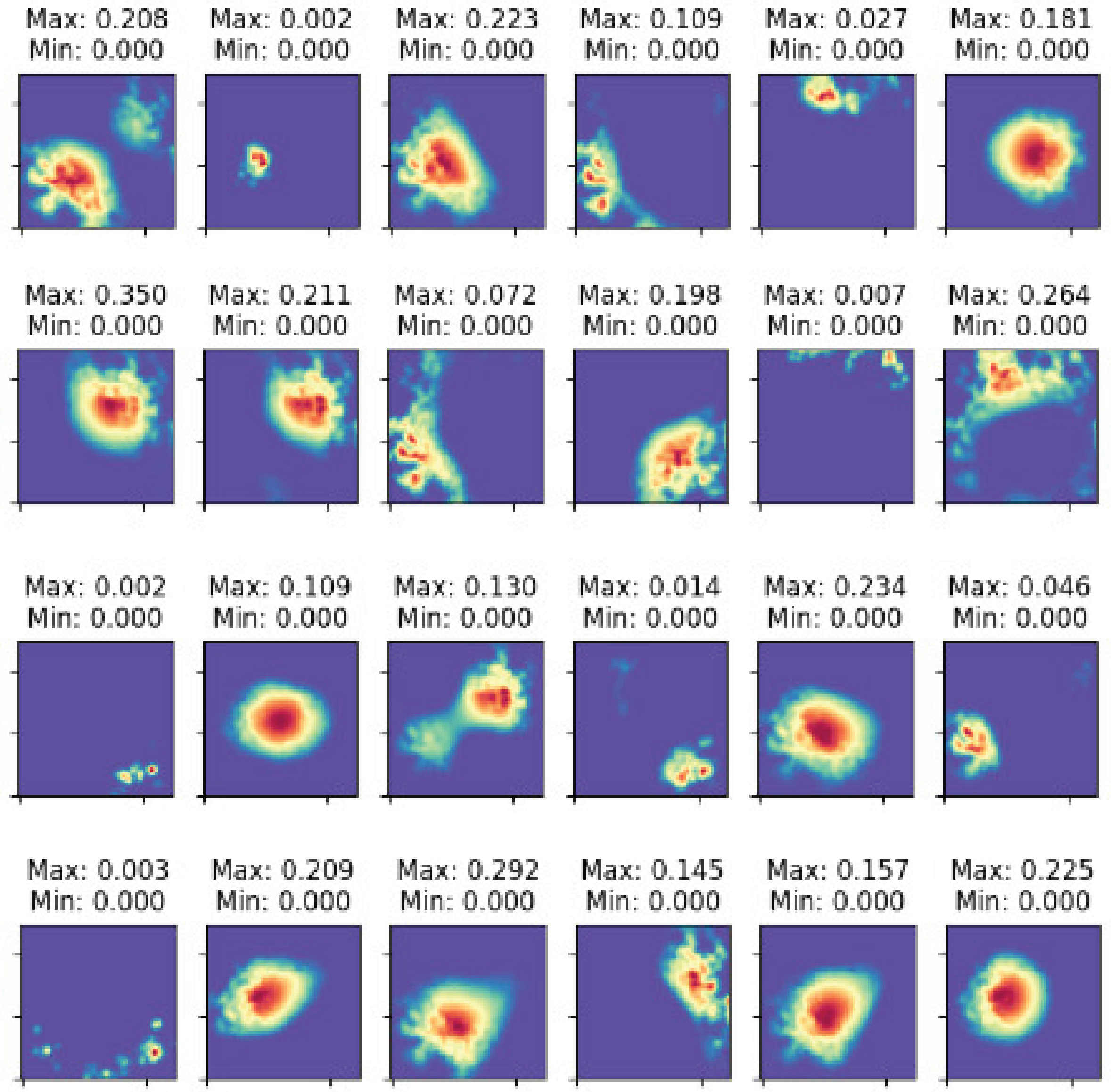
When does a multi-periodic solution NOT appear?

Still see 1 perfect module of grids!

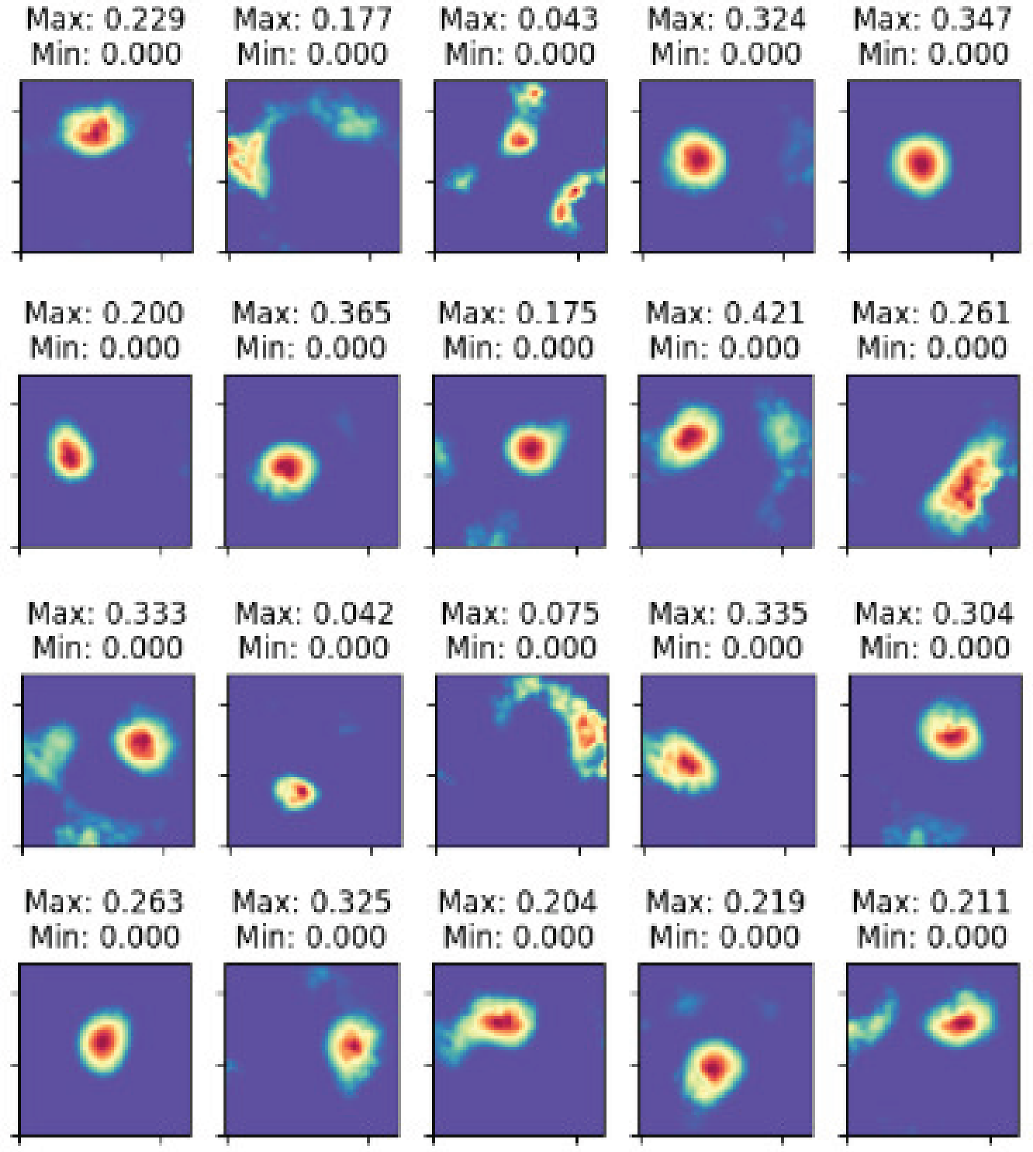
capacity loss ablation



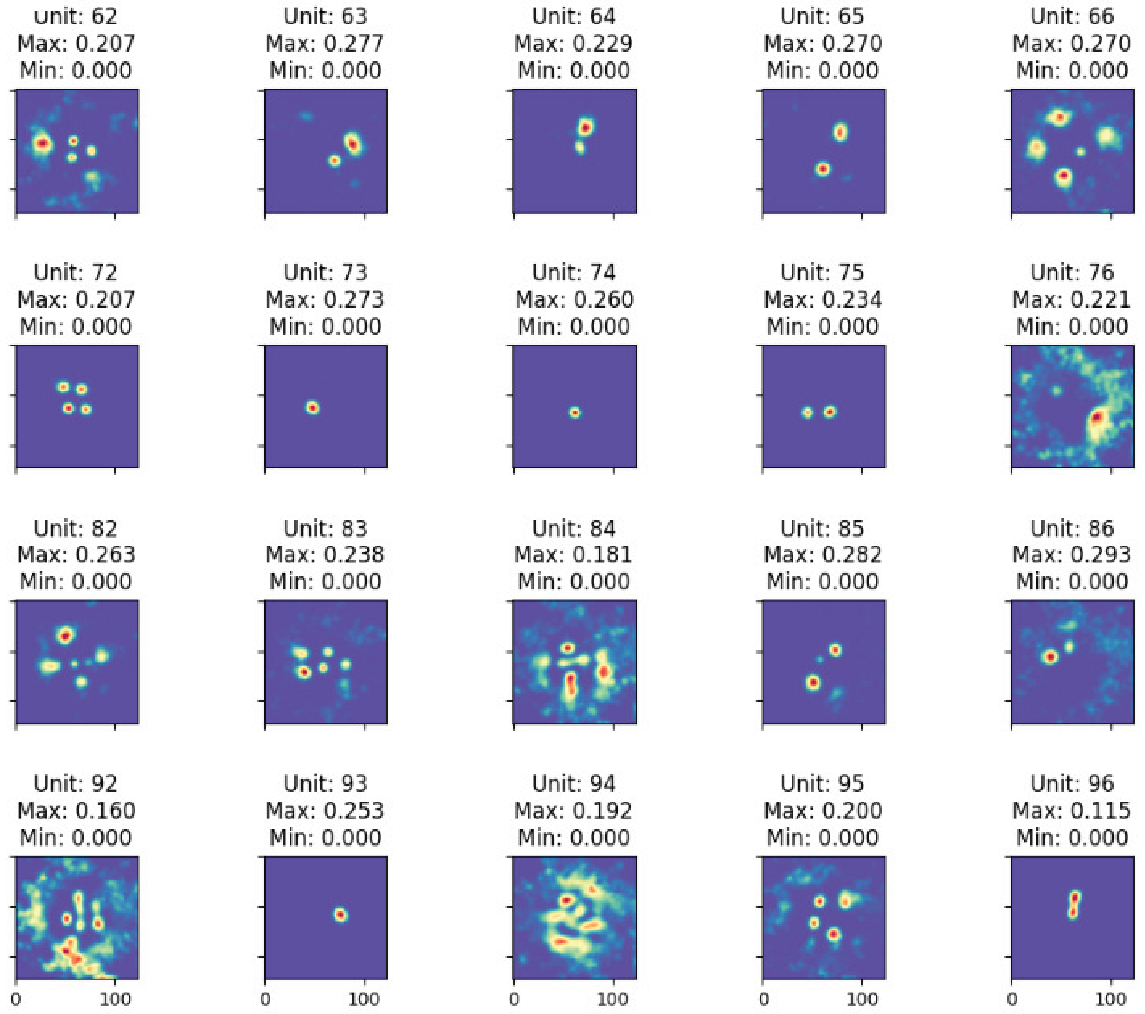
capacity loss ablation $+\sigma_g=0.1$



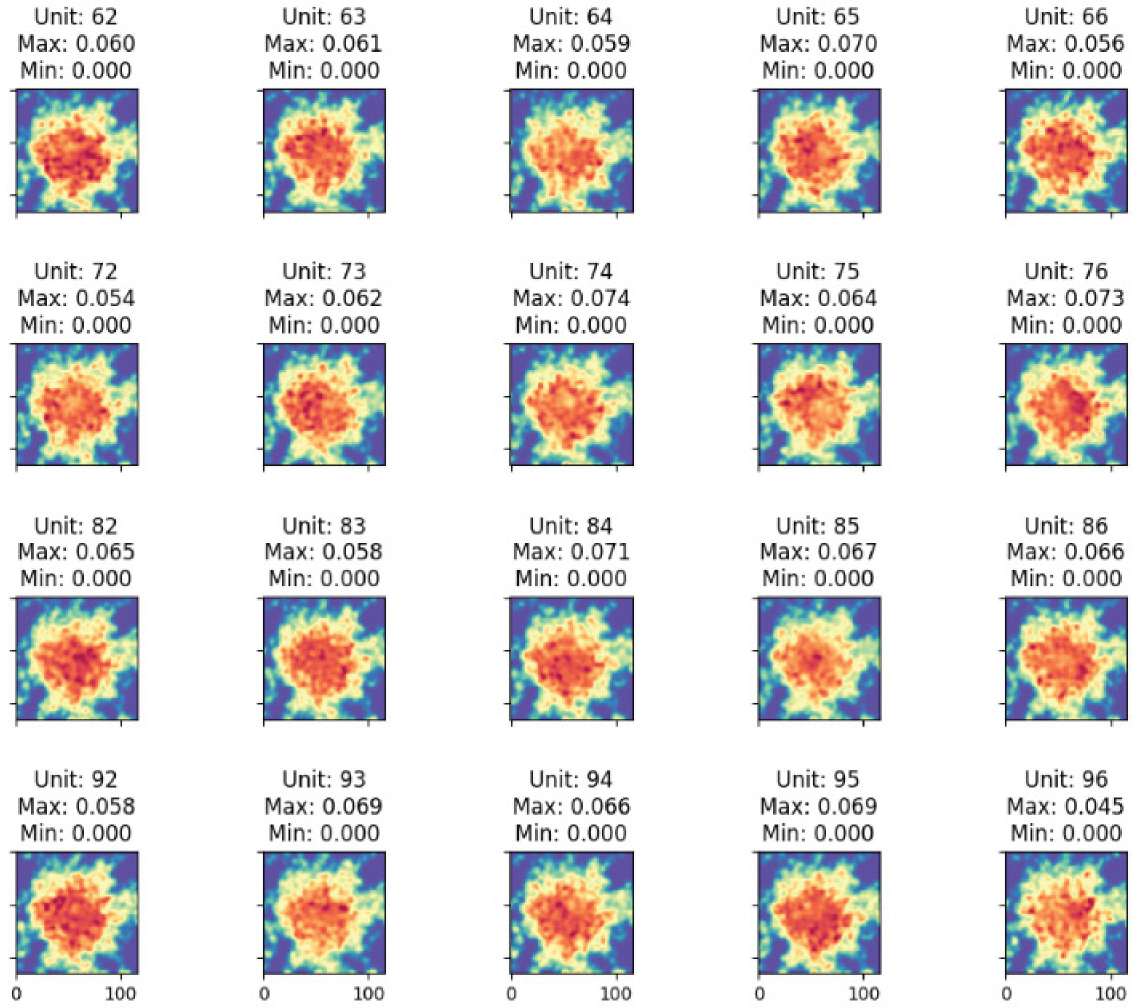
capacity loss ablation $+\sigma_g=0.2$



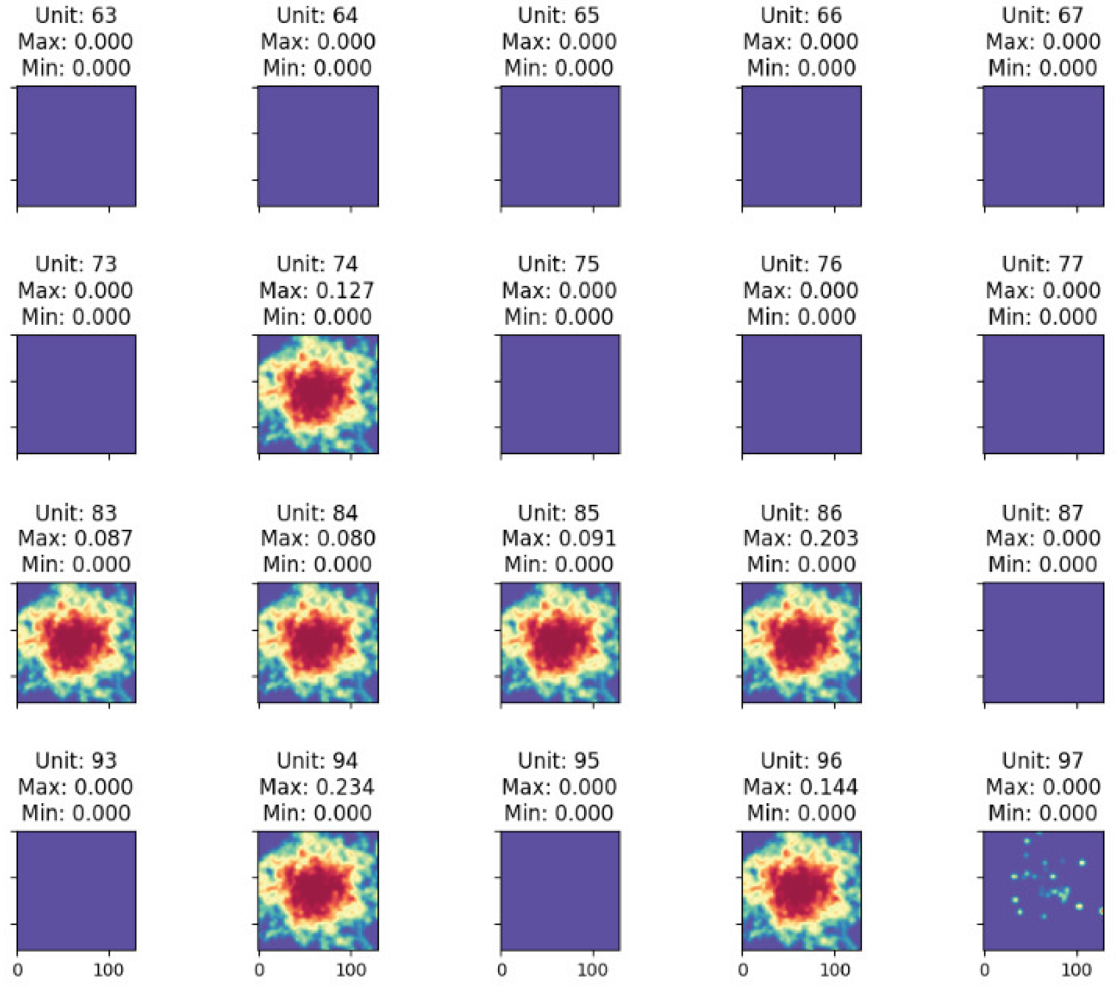
trajectory permutations ablation

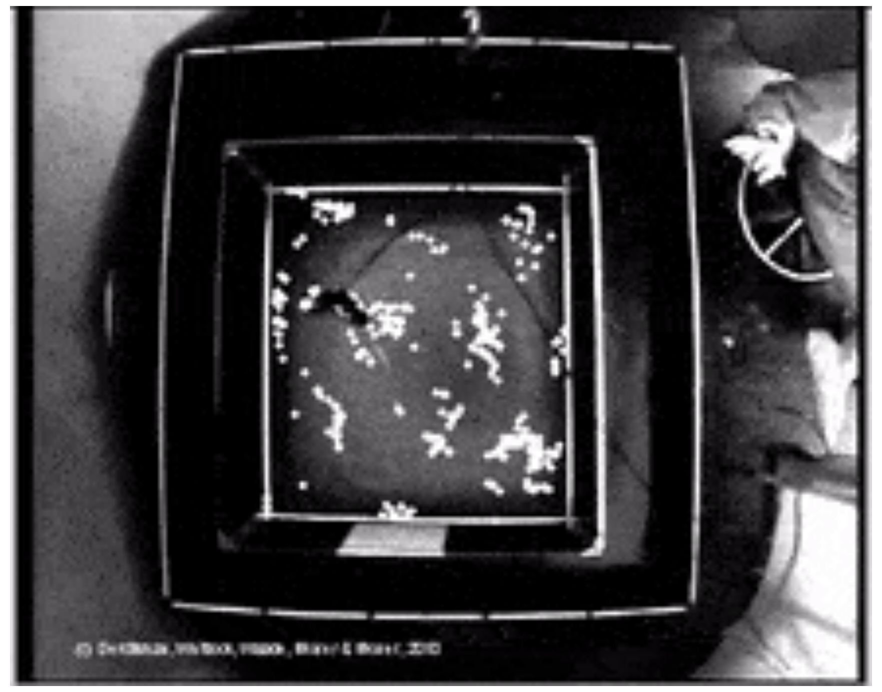


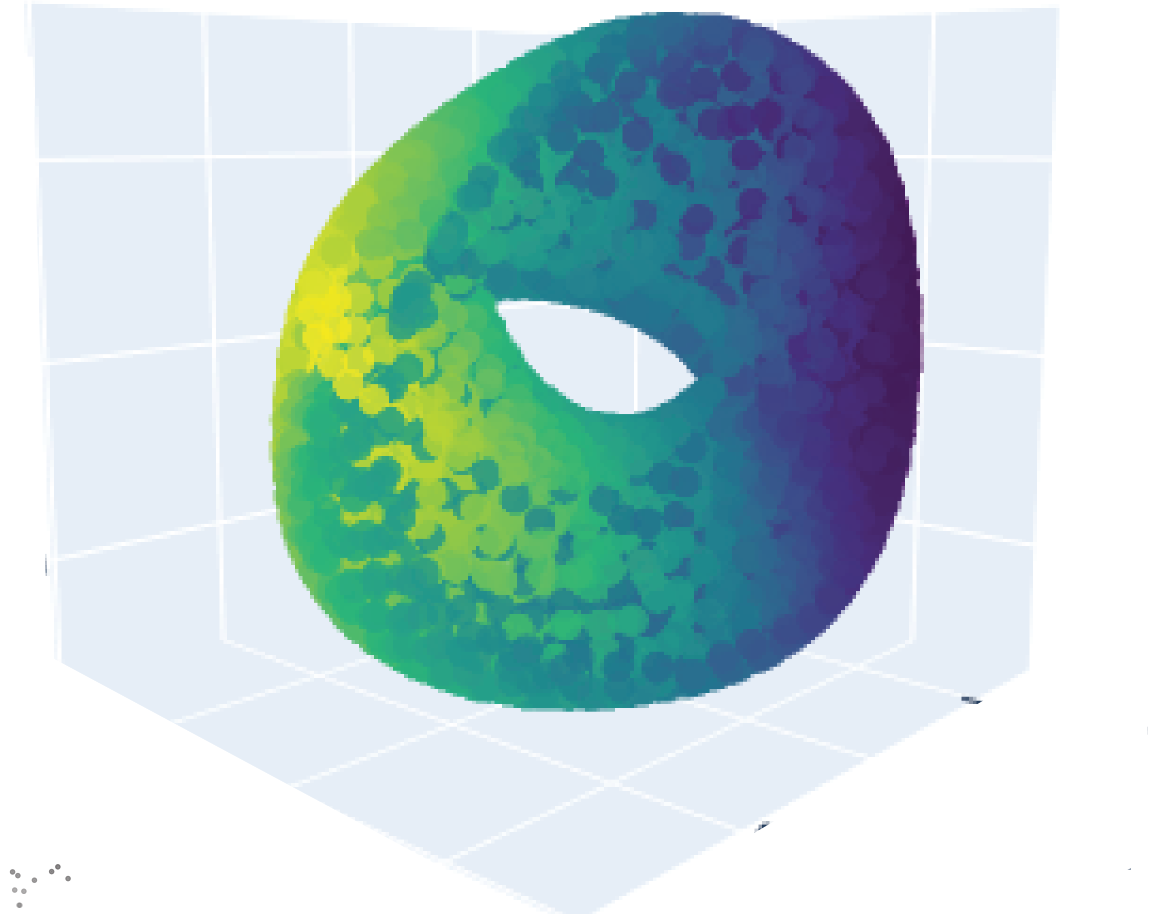
path invariance loss ablation

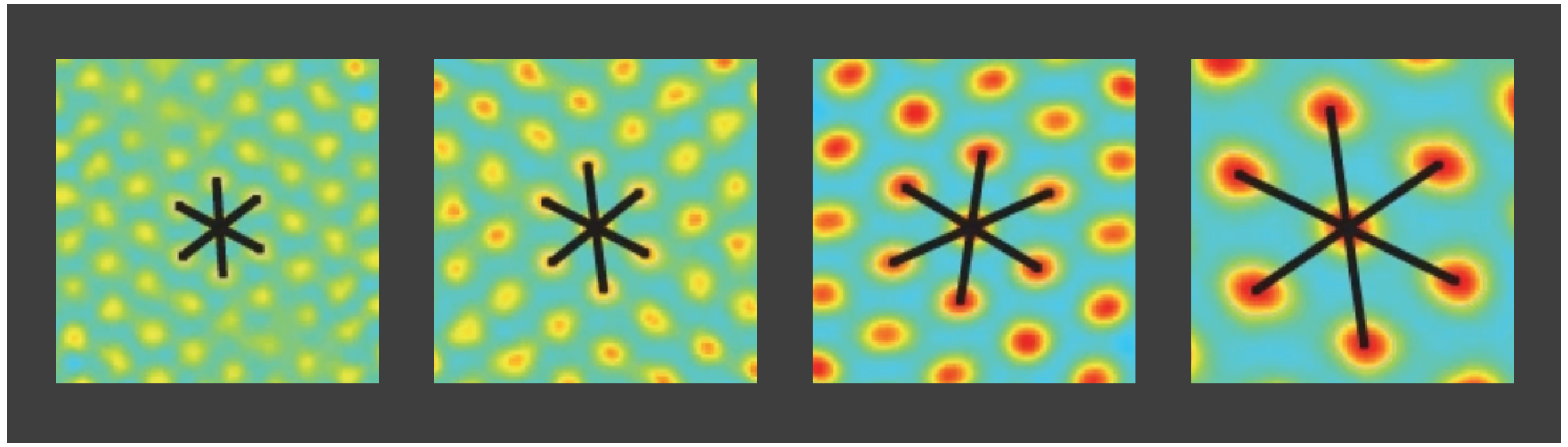


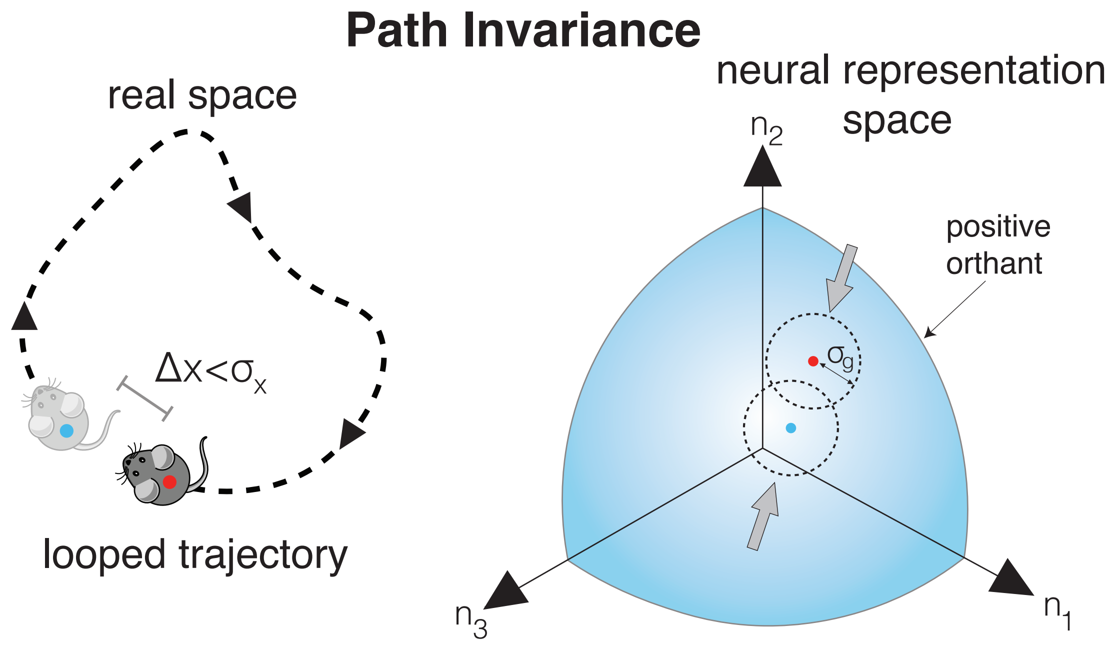
separation loss ablation

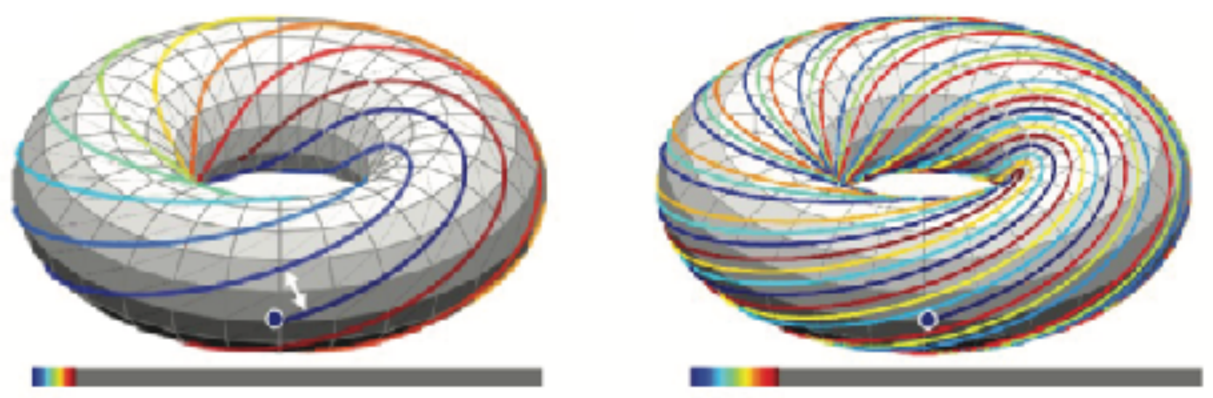


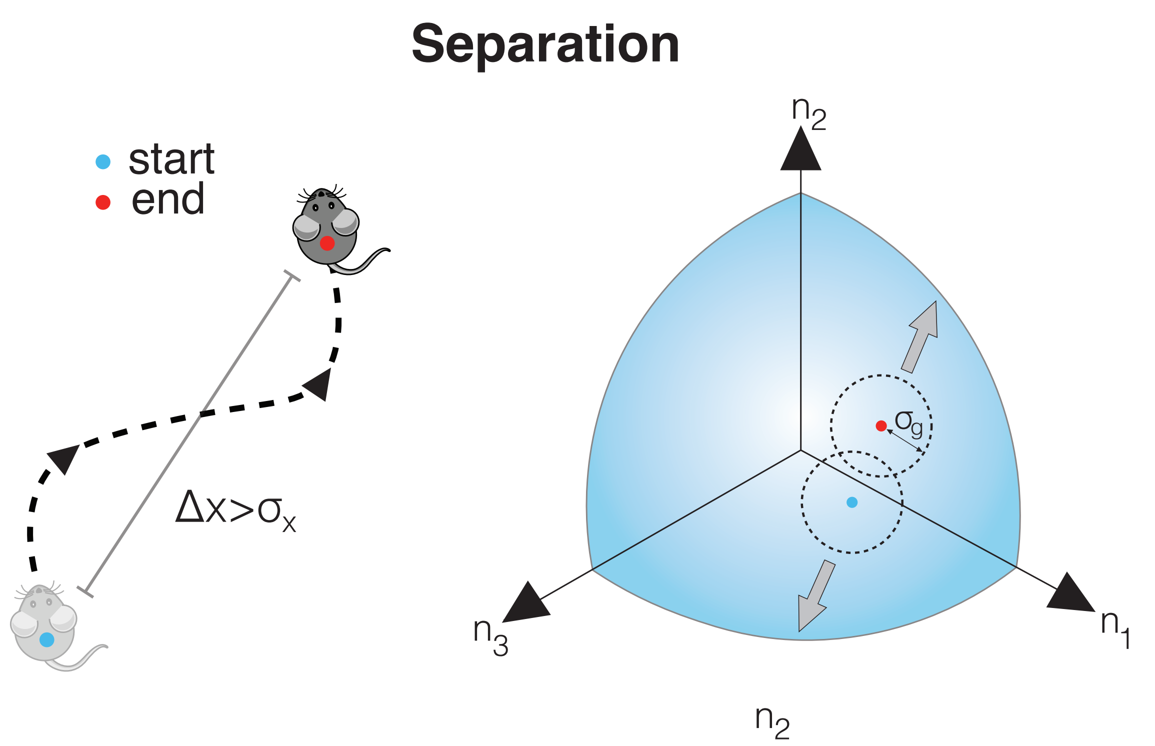


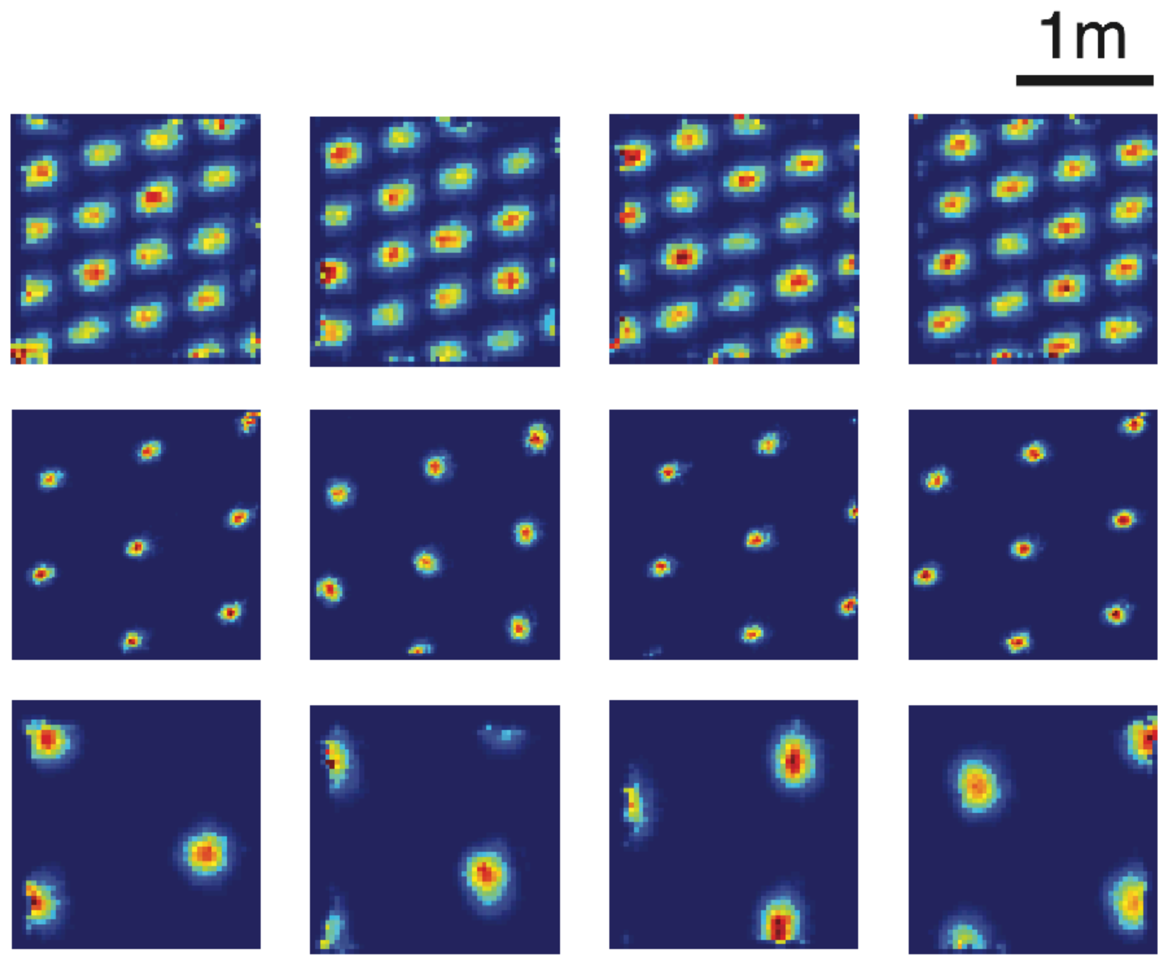












Thank you!

Mikail Khona @KhonaMikail (mikail@mit.edu)
Rylan Schaeffer @RylanSchaeffer (rylanschaeffer@gmail.com)

Self-Supervised Learning of Representations for Space Generates Multi-Modular Grid Cells, NeurIPS 2023

MIT brain+cognitive sciences | MIT Physics | McGOVERN INSTITUTE | K. LISA YANG ICoN CENTER | Stanford ENGINEERING Computer Science