What is an attractor?

"A **dynamical system** is a set of variables together with all the rules that determine their time-evolution."

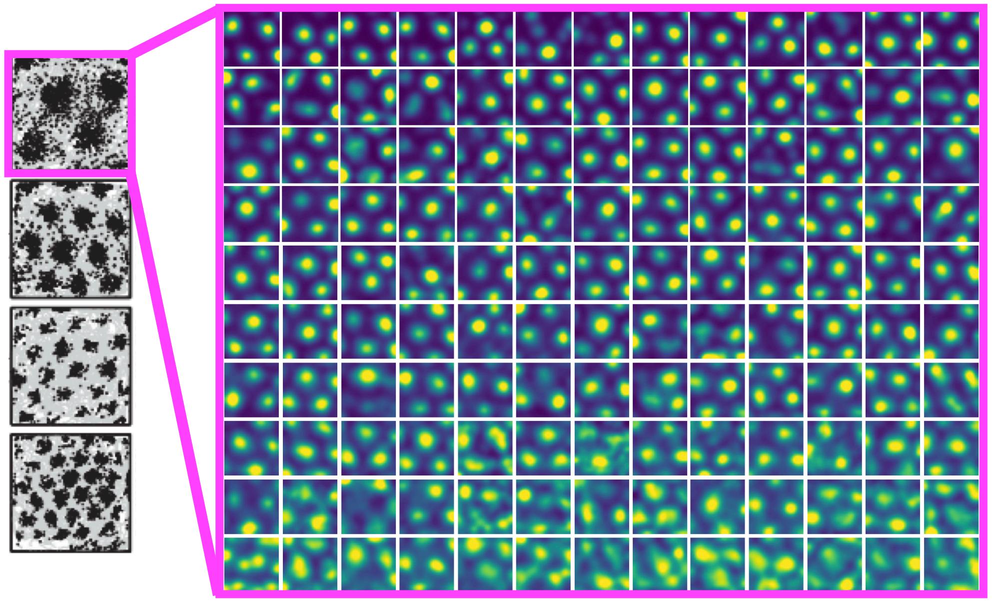
"The instantaneous value of these variables is called the **state** of the system at that moment. The state is a point (vector) in the state space of the dynamical system."

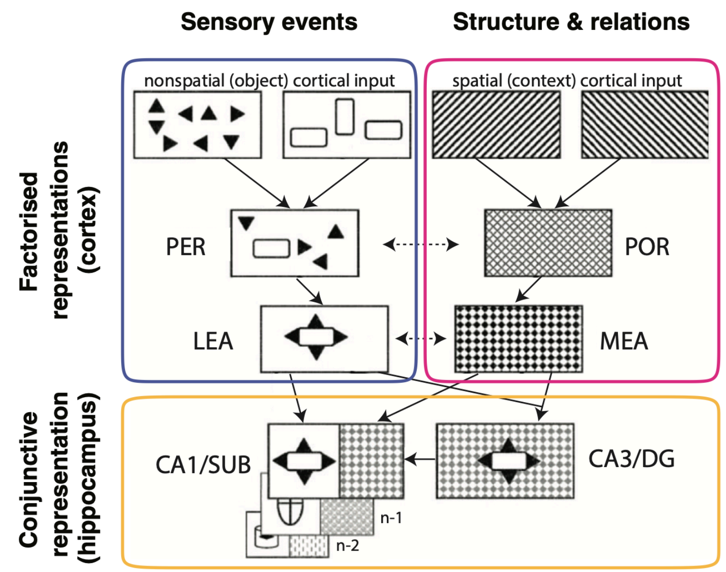
"An **attractor** is a state within a state space, to which all nearby states eventually flow."

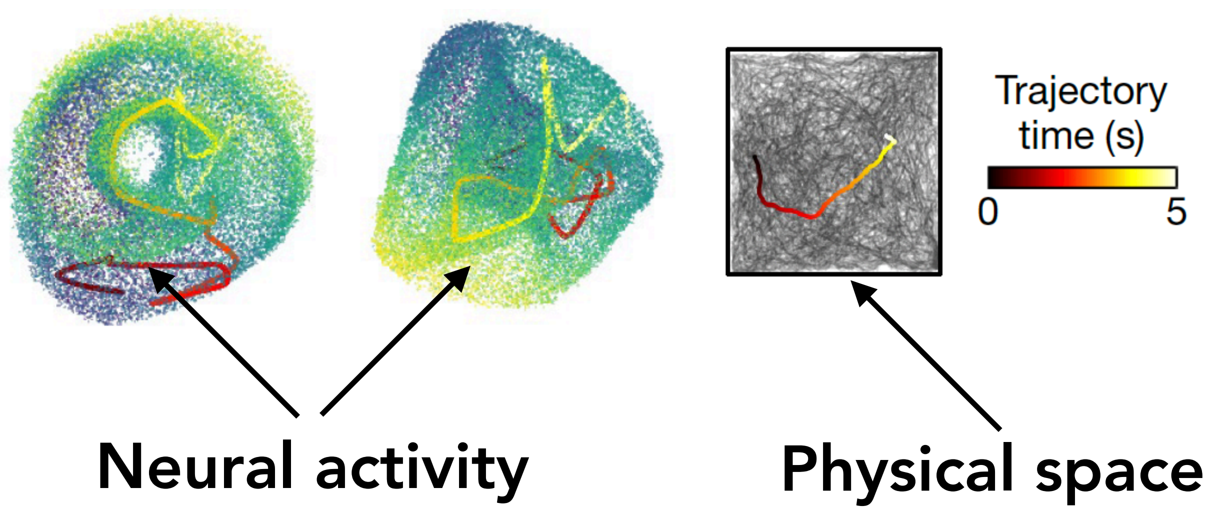
Khona, M., & Fiete, I. R. (2021). Attractor and integrator networks in the brain. *arXiv preprint arXiv:2112.03978*.

Why care about attractor neural networks?

i. understanding memory and spatial representation

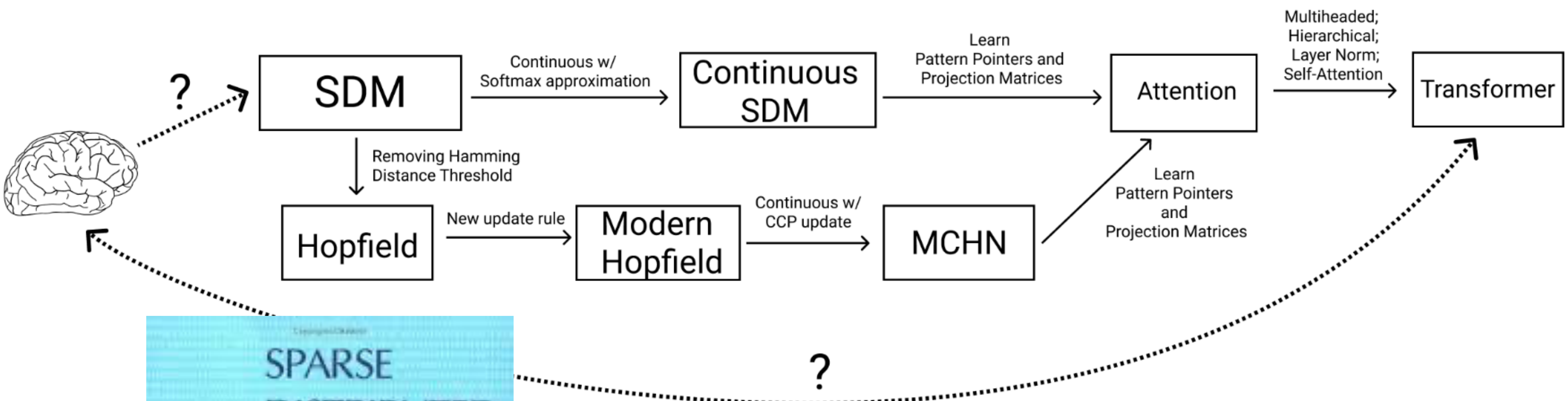


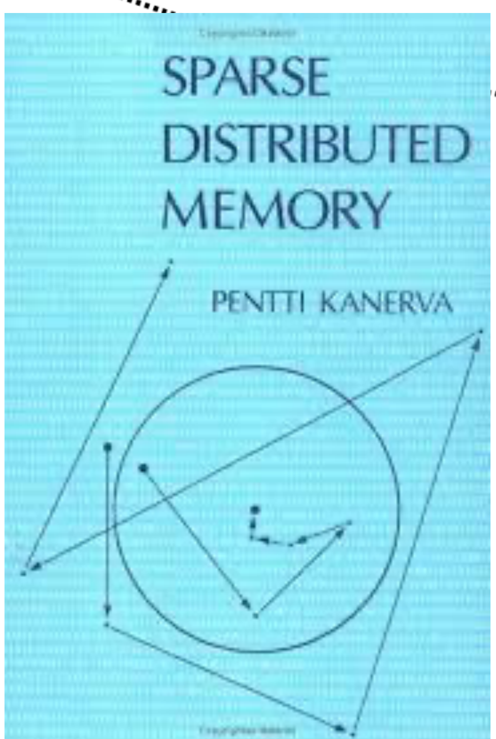




Why care about attractor neural networks?

ii. intriguing connections to AI?





COGNITIVE SCIENCE 12, 299–329 (1988)

Comparison Between Kanerva's SDM and Hopfield-type Neural Networks

JAMES D. KEELER
University of California at San Diego

https://www.trentonbricken.com/SDM-And-Hopfield-Networks/

Recap/basics about Hopfield networks

- Energy function
- Synchronous vs. asynchronous update rule
- Capacity of Hopfield networks
- Hopfield ’82 (discrete-time) vs. Hopfield ’84 model (continuous time)

Energy function of Hopfield '82 model

Dynamics: **s** - N-dimensional state vector, **W** - $N \times N$ weight matrix

$$s_i = \text{sign}(\sum_j w_{ij} s_j)$$

Energy function (when $w_{ij} = w_{ji}$):

$$H = -\frac{1}{2}\sum_{ij} s_i w_{ij} s_j = -\frac{1}{2}\mathbf{s}^T \mathbf{W} \mathbf{s}$$

Derivation: See course website

This energy function:

1. Never increases at any time step

2. Is bounded from below

Synchronous vs. asynchronous dynamics

Synchronous: update all s_i at the same same time (Little 1974)

Asynchronous: each s_i is updated one at a time

Consider: $W = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and the four binary states.

What are the stable states for each update rule?

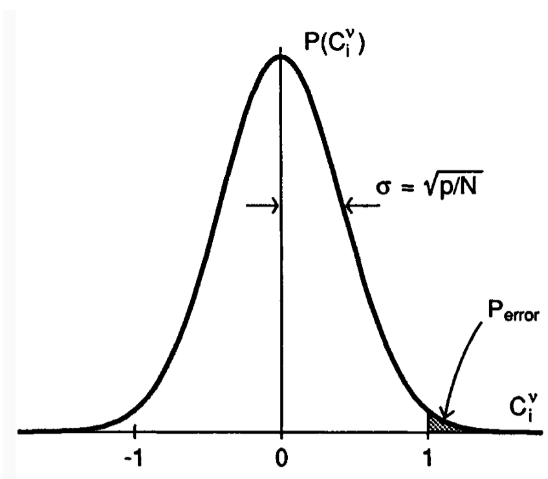
Capacity of a Hopfield network

Hebbian learning rule: $w_{ij} = \frac{1}{N}\sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu$

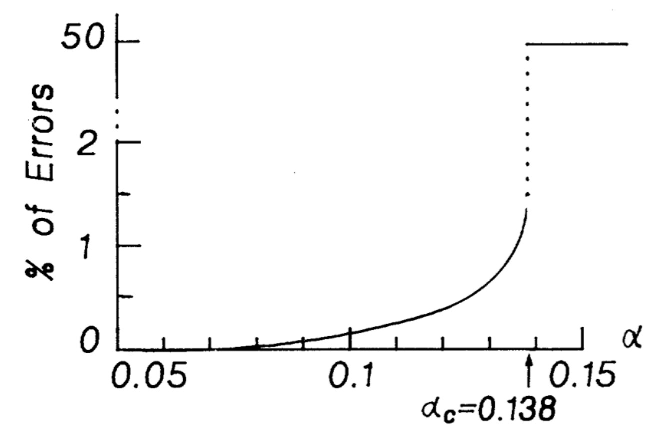
A stored pattern is **stable** for pattern ξ^ν if, for all *i*,

$$\xi_i^\nu = \text{sign}\left(\sum_j w_{ij}\xi_j^\nu\right) = \text{sign}\left(\xi_i^\nu + \frac{1}{N}\sum_{\mu\neq\nu} \xi_i^\mu \xi_j^\mu\right)$$

Chance of errors depends on N and p:



Approximate crosstalk between random patterns



Statistical mechanics derivation of capacity (Amit, Gutfreund, Sompolinsky 1990)

Hopfield (1984) model

Dynamics:

u - N-dimensional state vector, **W** - $N \times N$ weight matrix

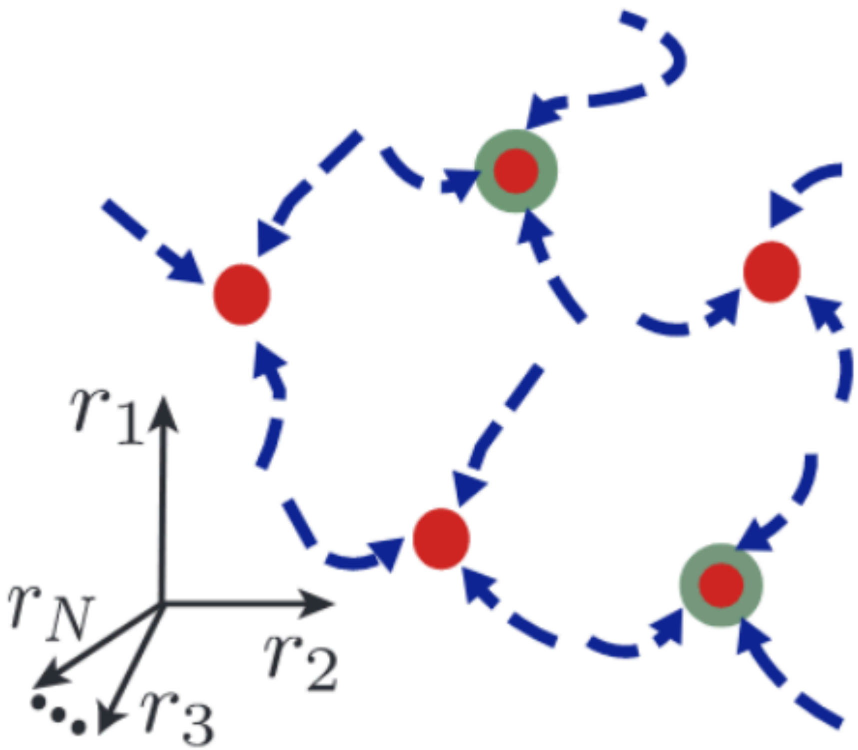
$$\tau_i \frac{du_i}{dt} = -u_i + \sum_j w_{ij} V_j$$

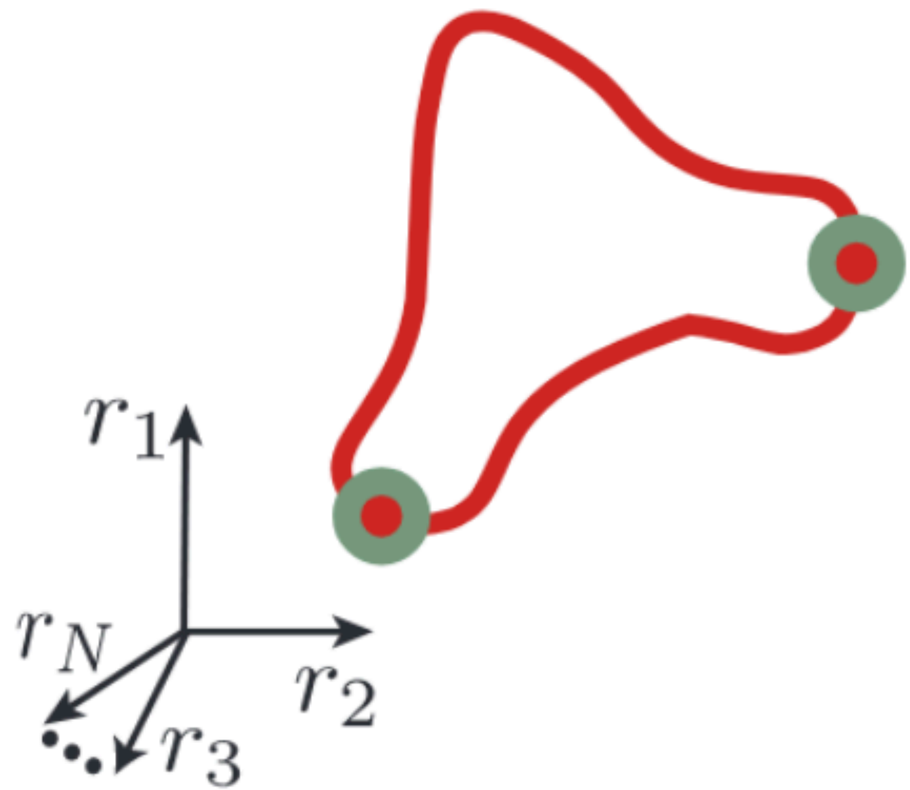
$$V_j = g(u_j); g(x) = \tanh(\beta x)$$

Energy function (when $w_{ij} = w_{ji}$ and $g(\cdot)$ is bounded):

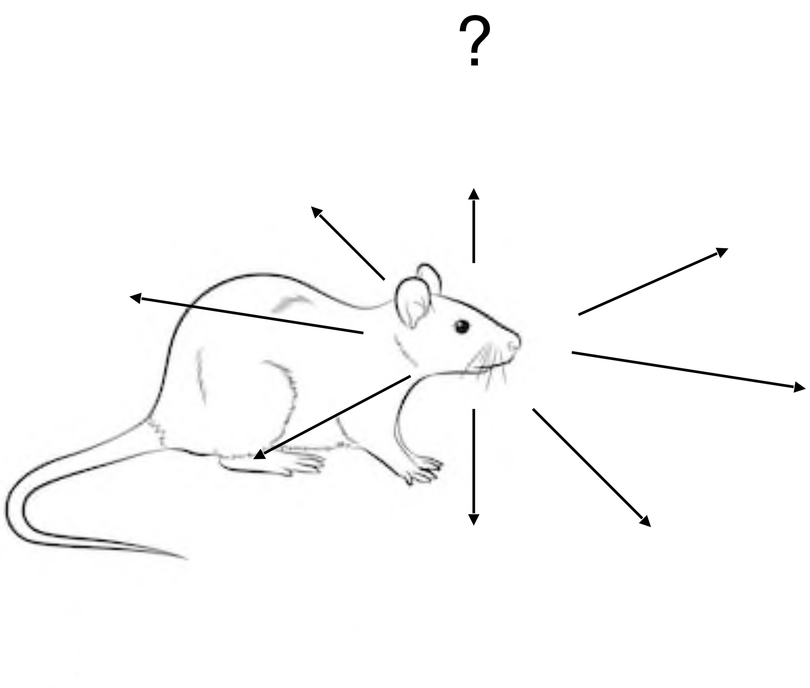
$$H = -\frac{1}{2} \sum_{ij} w_{ij} V_i V_j + \sum_i \int_0^{V_i} g^{-1}(V) dV$$

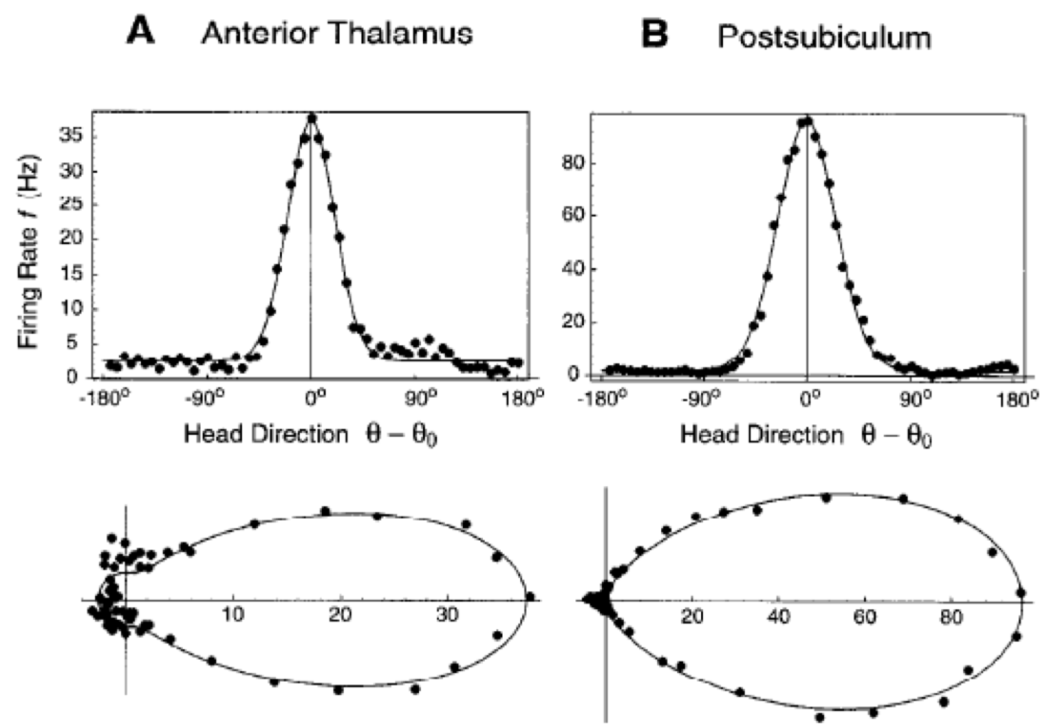
Point vs. ring attractors





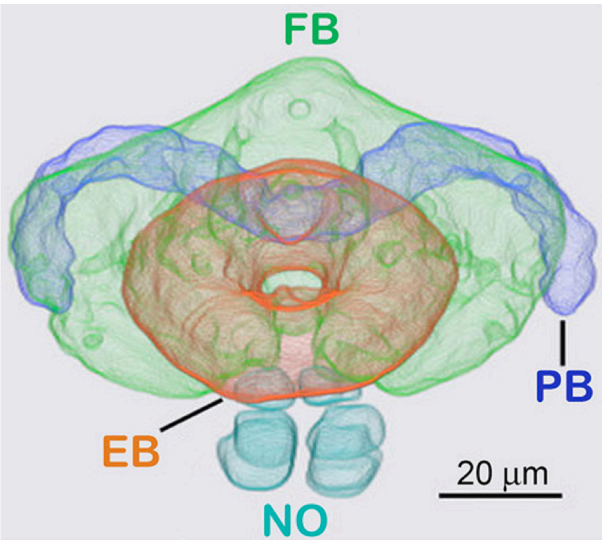
Head-direction cells

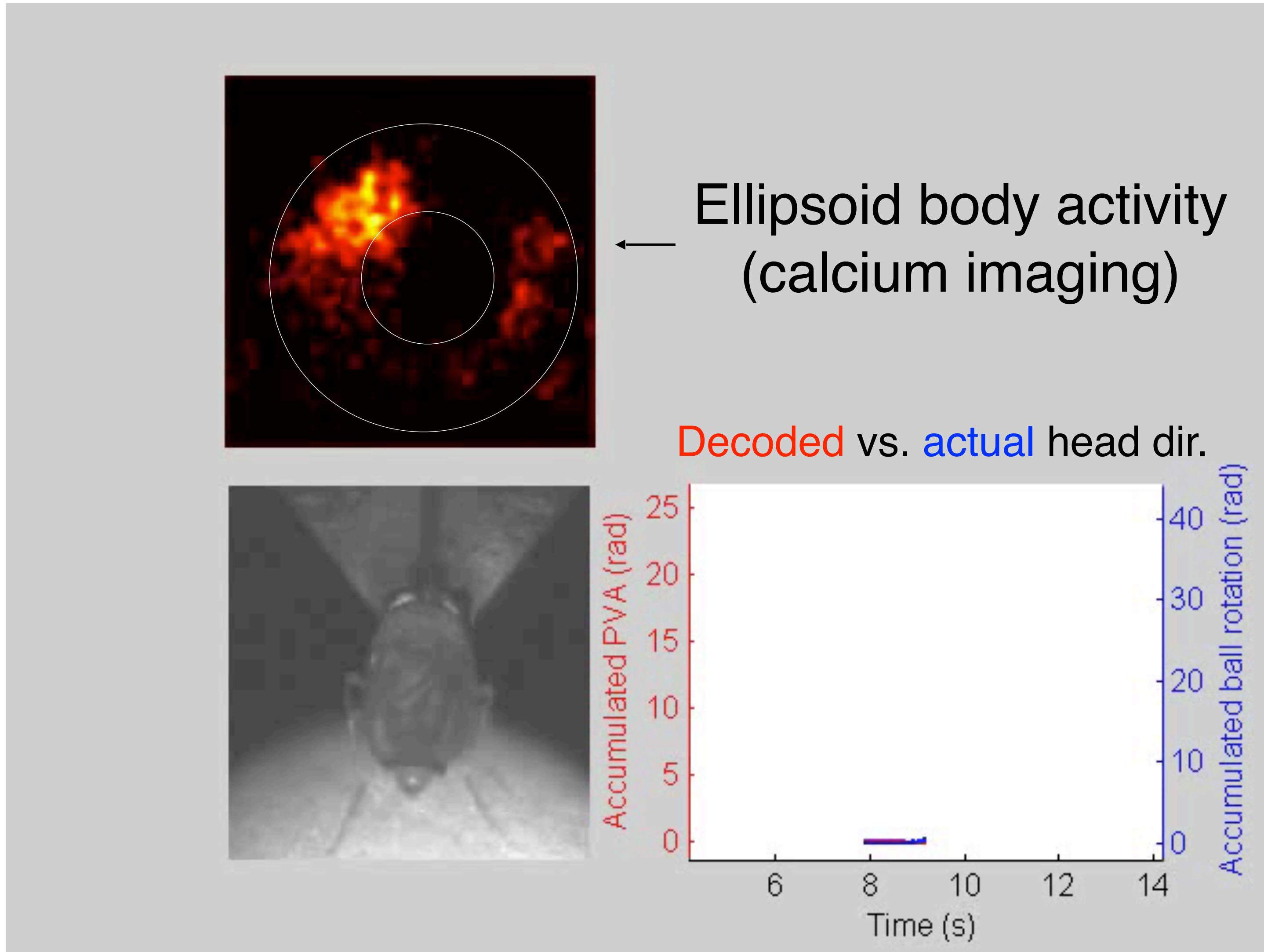




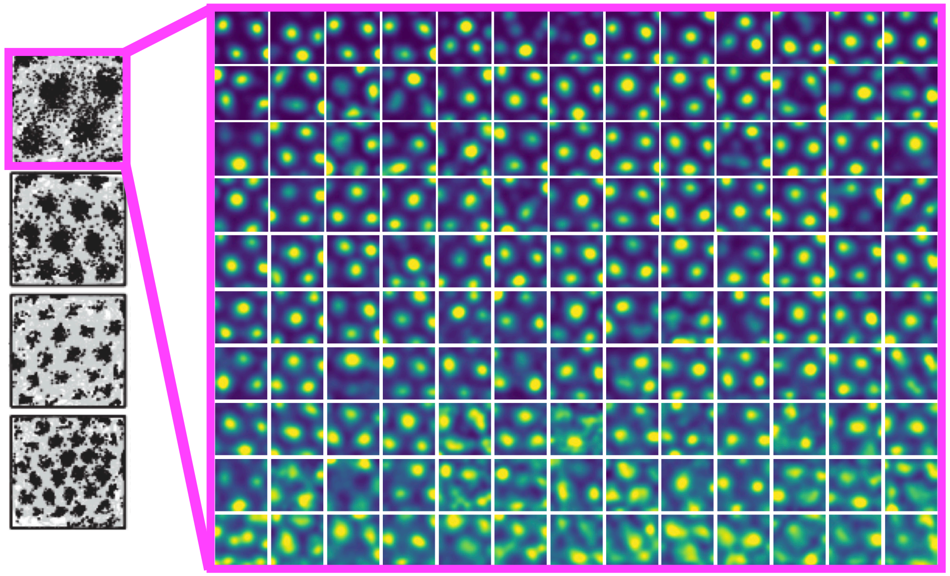
Head-direction cells in ellipsoid body of *Drosophila*

(Seelig & Jayaraman 2015)



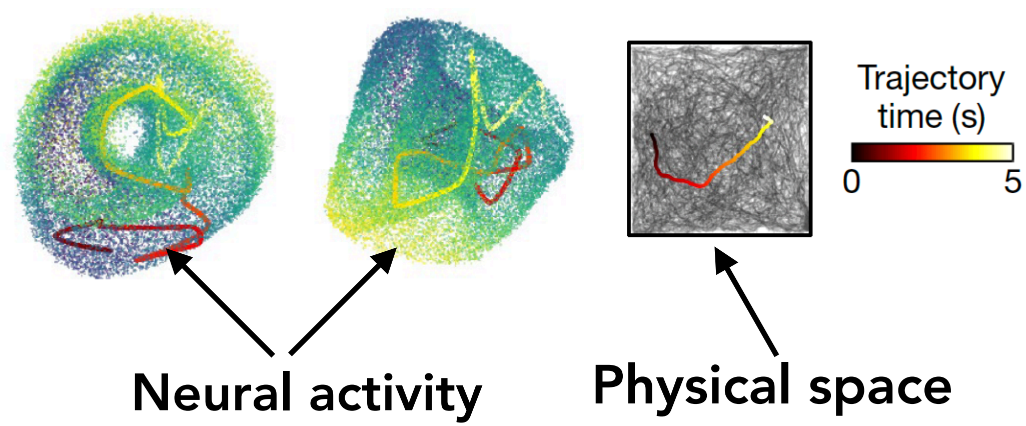


Describing network activity of grid cells



Grid cells are organized into modules. Each has many grid cells of (approximately) the same spatial scale.

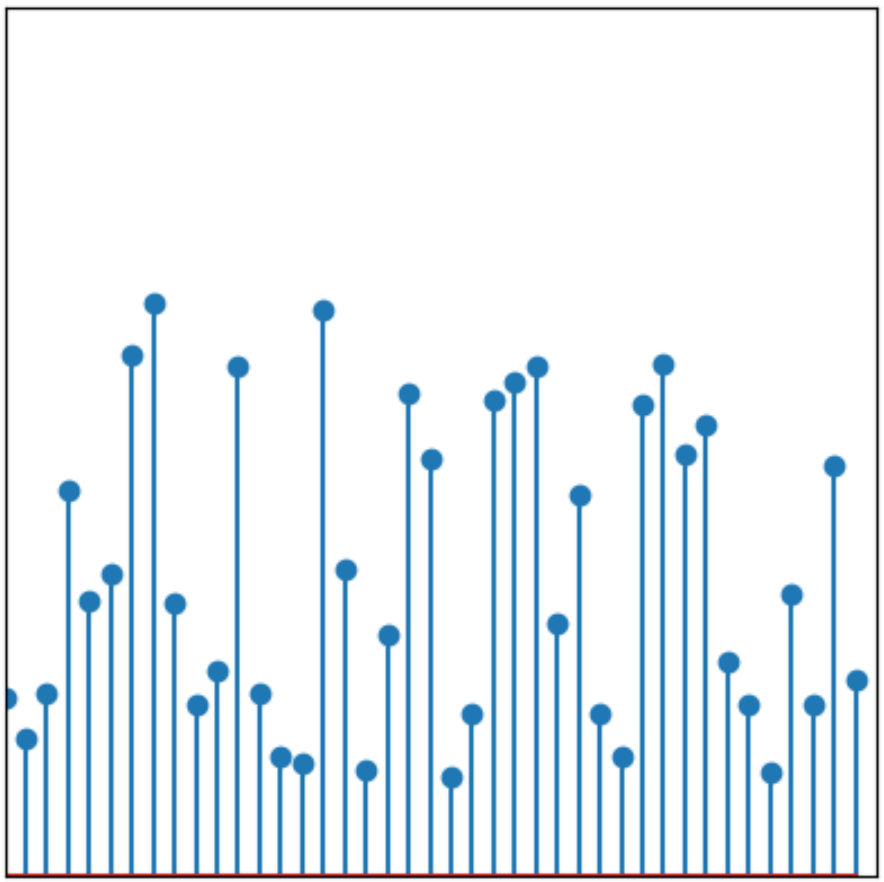
Continuous attractor network hypothesis (for grid cells): Neural weights and activity patterns organized as a **torus**



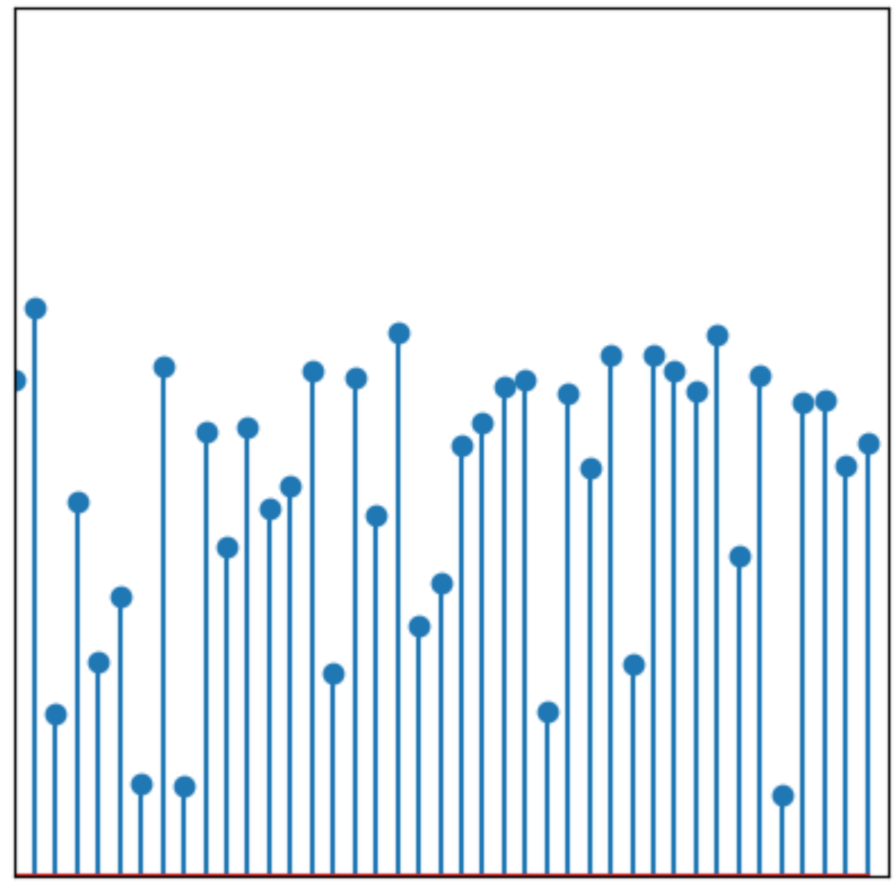
Gardner, R. J., Hermansen, E., Pachitariu, M., Burak, Y., Baas, N. A., Dunn, B. A., ... & Moser, E. I. (2022). Toroidal topology of population activity in grid cells. Nature, 602(7895), 123-128.

Zhang's (1996) ring attractor model

Bump formation

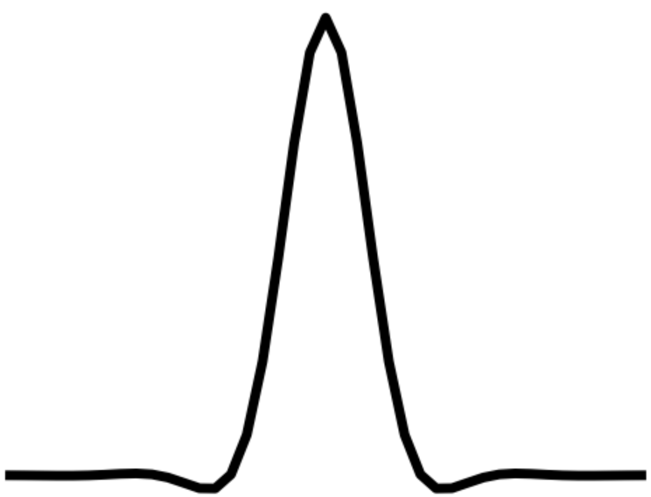


Shifting the bump

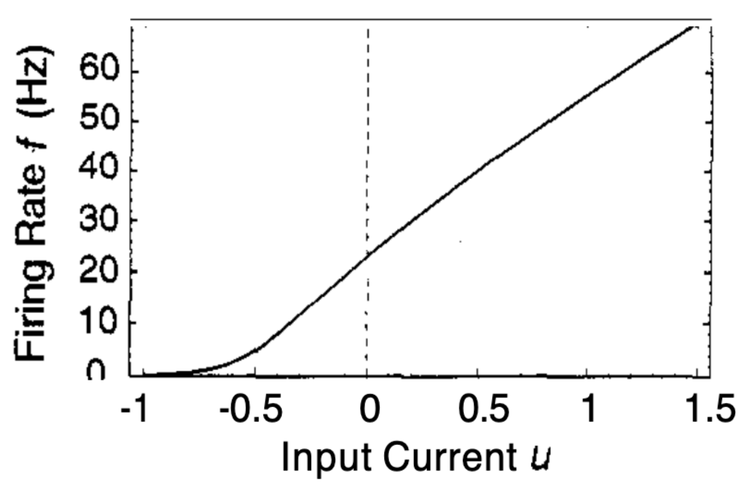


Zhang (1996) dynamics

$$\tau \frac{du}{dt} = -u + w \circledast \sigma(u)$$

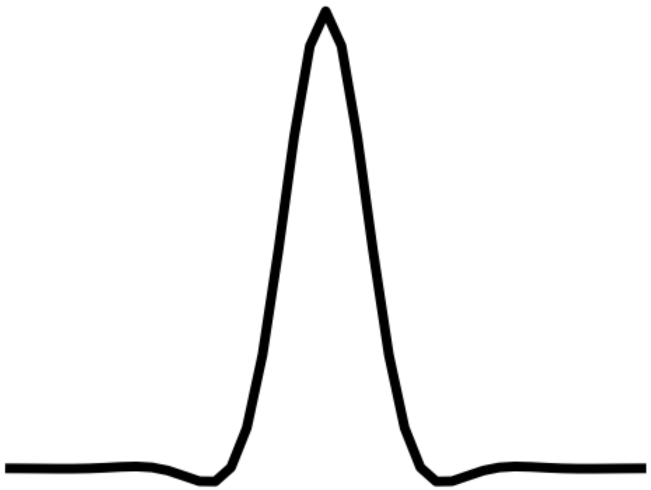


w

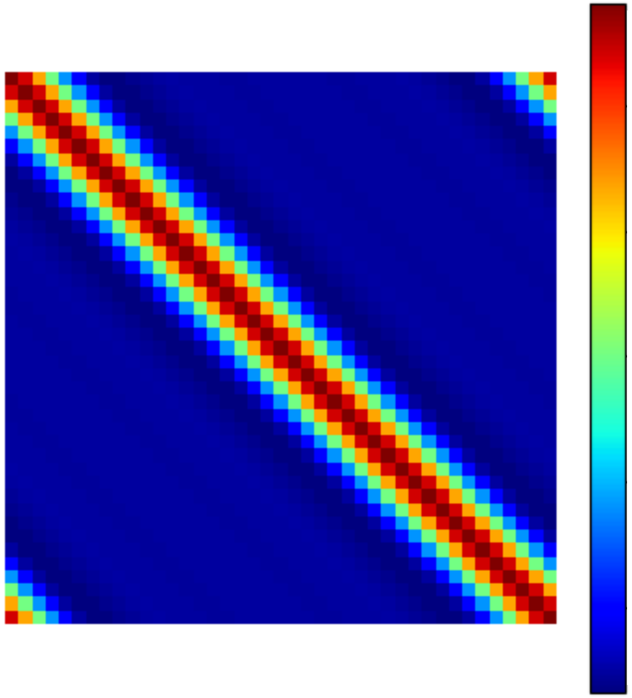


$\sigma(u)$

Designing a ring attractor



Weight profile



Weight matrix

Toeplitz!

Any $n \times n$ matrix A of the form

$$A = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & \cdots & a_{-(n-1)} \\ a_1 & a_0 & a_{-1} & \ddots & & \vdots \\ a_2 & a_1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & a_{-1} & a_{-2} \\ \vdots & & \ddots & a_1 & a_0 & a_{-1} \\ a_{n-1} & \cdots & \cdots & a_2 & a_1 & a_0 \end{bmatrix}$$

is a **Toeplitz matrix**. If the i, j element of A is denoted $A_{i,j}$ then we have

$$A_{i,j} = A_{i+1,j+1} = a_{i-j}.$$

$$\tau \frac{du}{dt} = -u + W\sigma(u)$$

$$\rightarrow \tau \frac{du}{dt} = -u + w \circledast \sigma(u)$$

Terry Sejnowski (Salk Institute for Biological Studies)
Brains and AI

NICE 2024, https://niceworkshop.org, 23 April 2024

https://flagship.kip.uni-heidelberg.de/video/meeting_257_video_9437.mp4

Designing the weight matrix

Zhang '96 dynamics:

$$\tau \frac{du}{dt} = -u + w \circledast \sigma(u)$$

Steady state:

$$u = w \circledast \sigma(u)$$

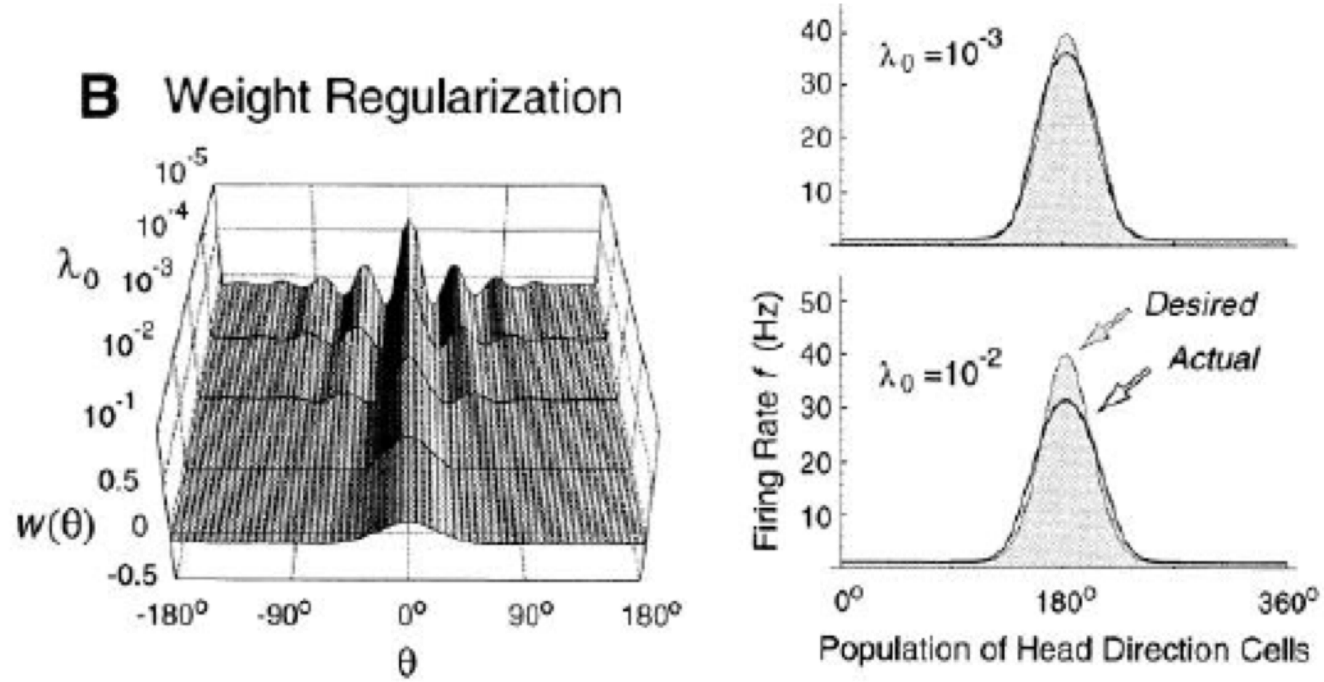
$$\iff \mathcal{F}(u) = \mathcal{F}(w) \odot \mathcal{F}(\sigma(u))$$

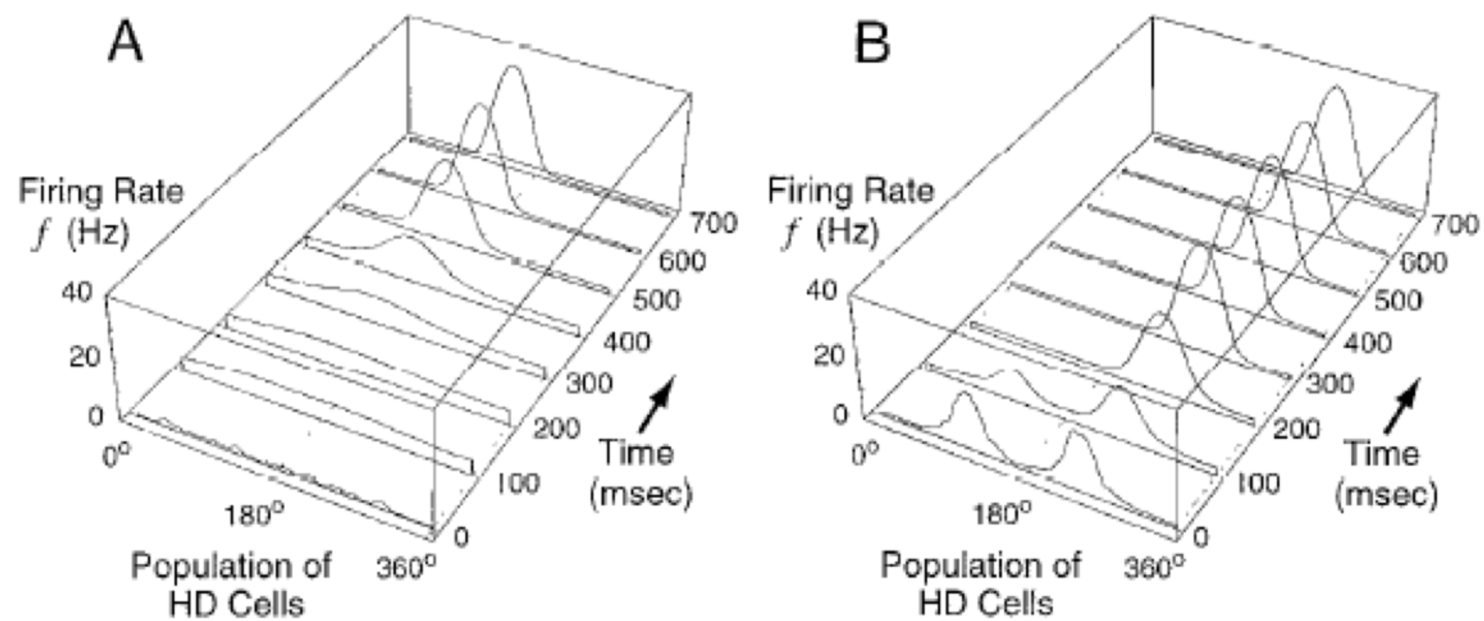
Solving for weight matrix:

$$\mathcal{F}(w^*) = \frac{\mathcal{F}(u)}{\mathcal{F}(\sigma(u))}$$

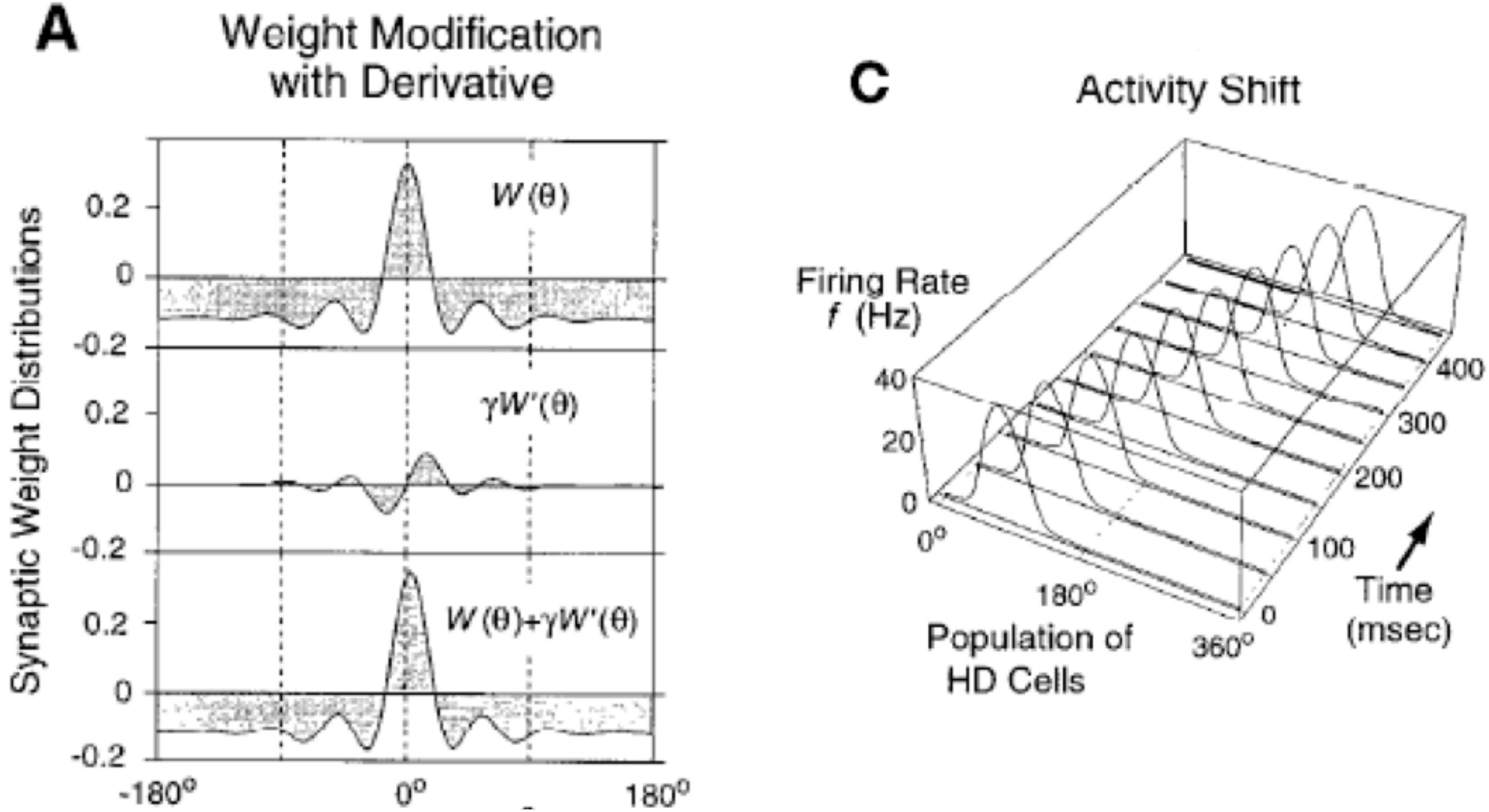
Regularize to ensure unique solution:

$$\mathcal{F}(w^*) = \frac{\mathcal{F}(u) \odot \mathcal{F}(\sigma(u))}{|\mathcal{F}(\sigma(u))|^2 + \lambda}$$

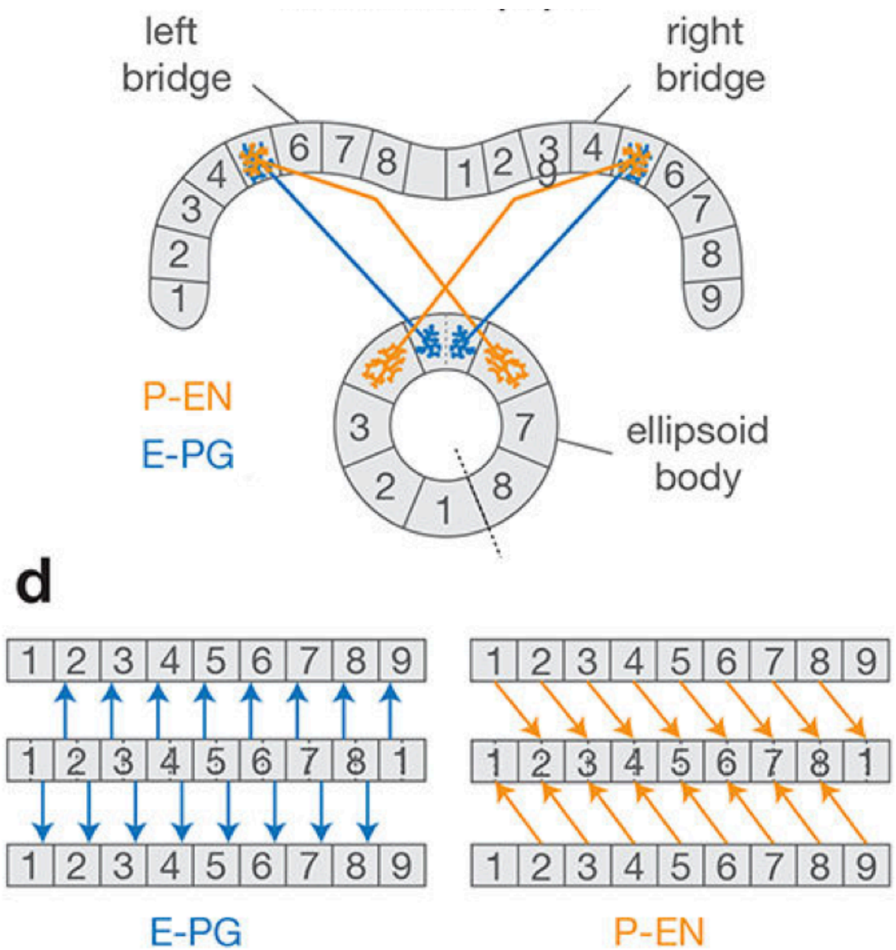

B Weight Regularization
λ0
10-5
10-4
10-3
10-2
10-1
0.5
0
-0.5
w(θ)
-180°
-90°
0°
90°
180°
θ
λ0 =10-3
λ0 =10-2
Desired
Actual
Firing Rate f (Hz)
40
30
20
10
50
0°
180°
360°
Population of Head Direction Cells

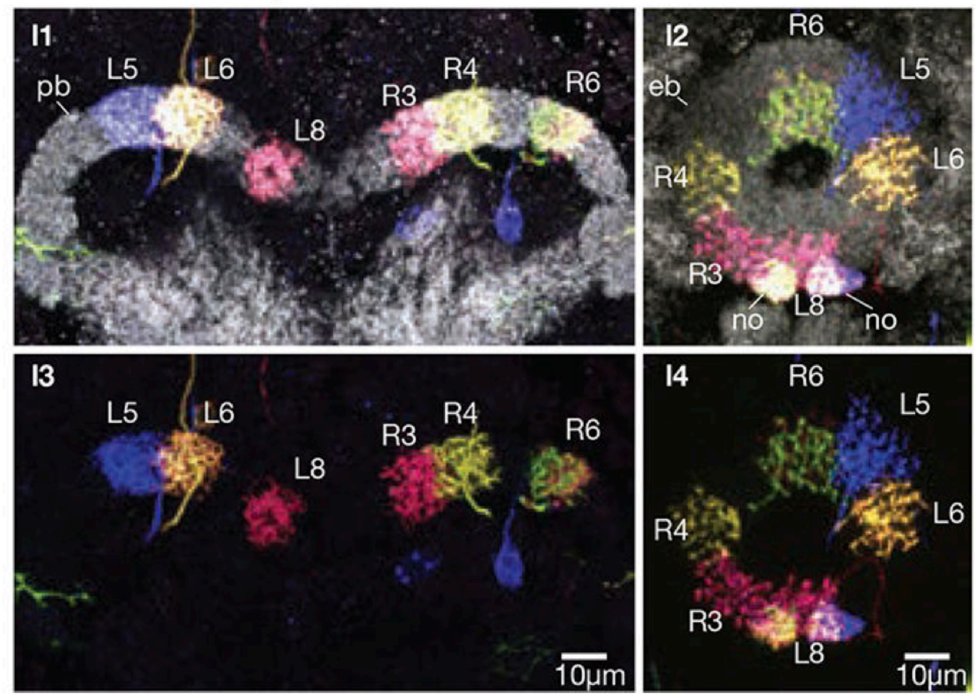

A
B
Firing Rate
f (Hz)
40
20
0
0°
180°
360°
Population of
HD Cells
Time
(msec)
0
100
200
300
400
500
600
700

Shifting the bump

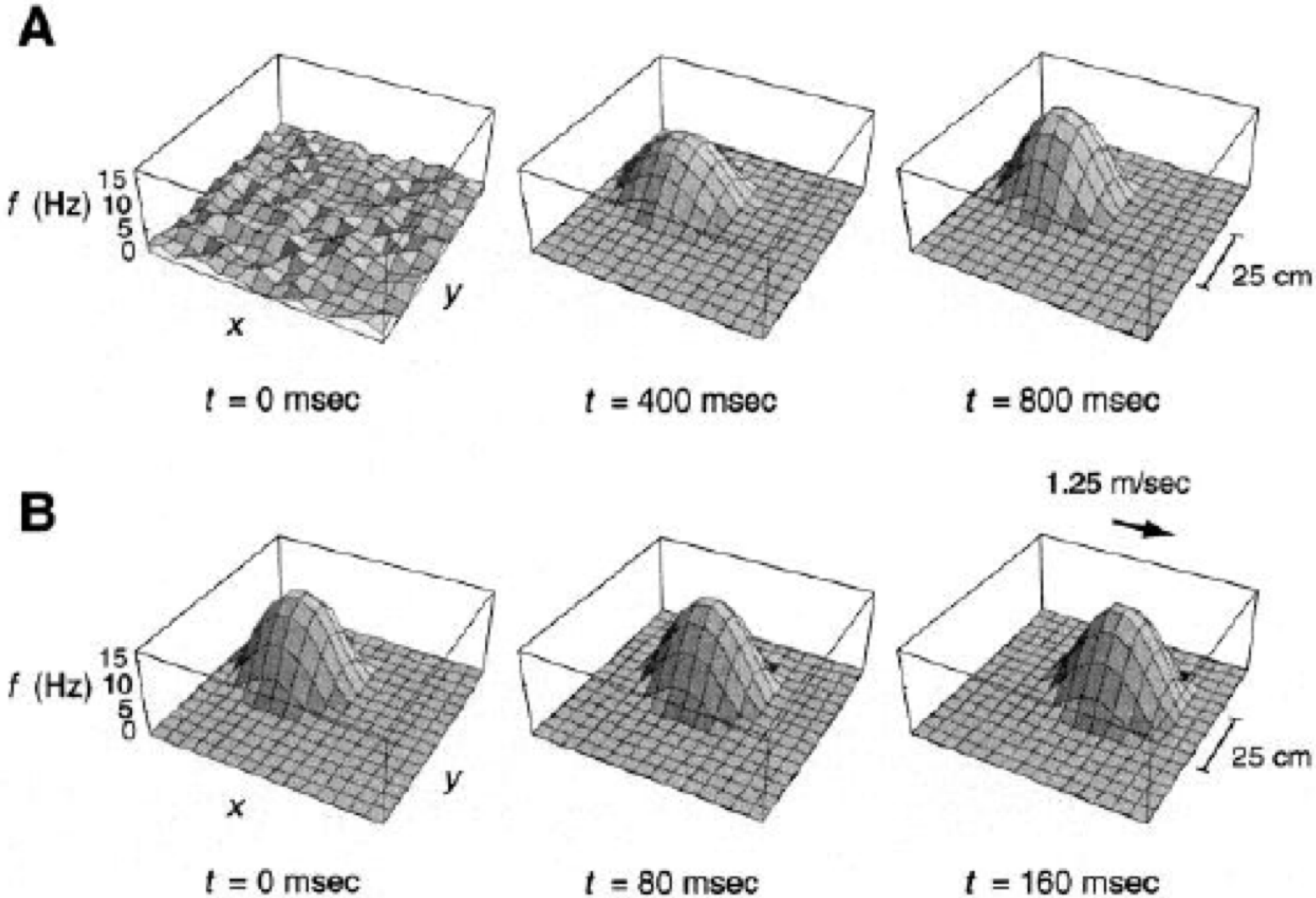


A shift mechanism in the fly

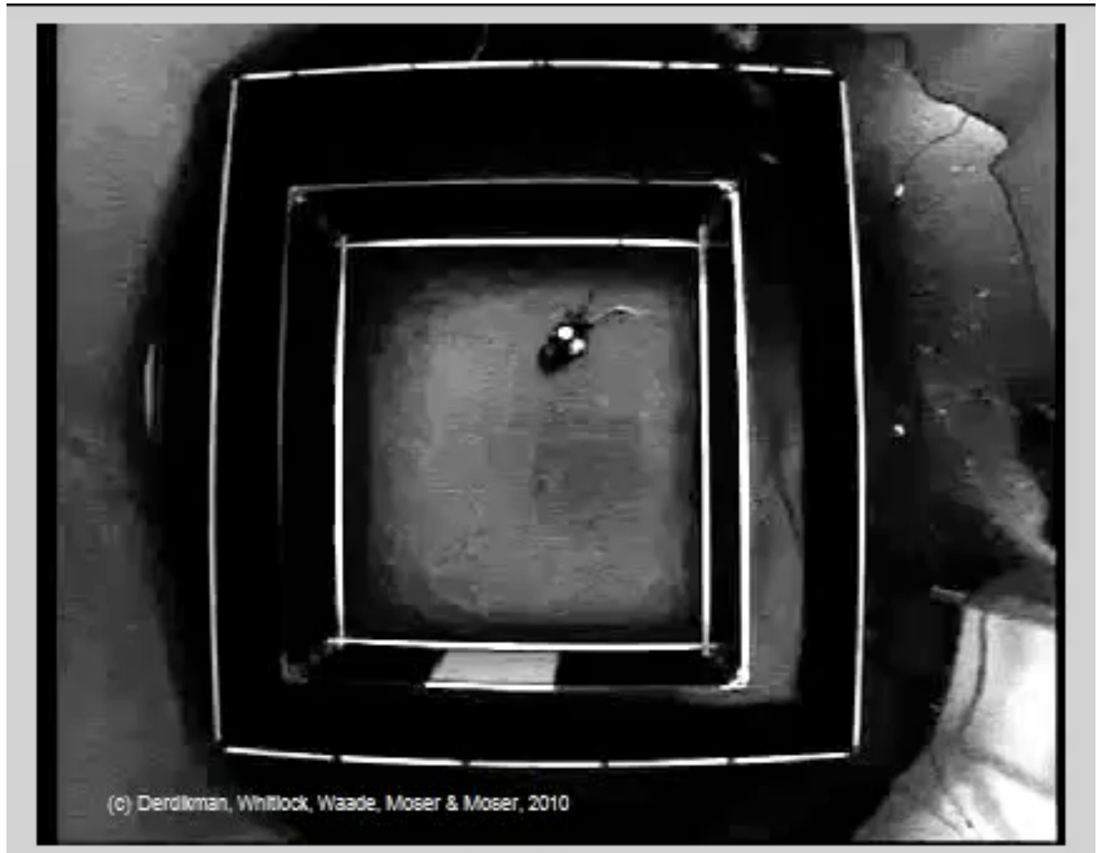


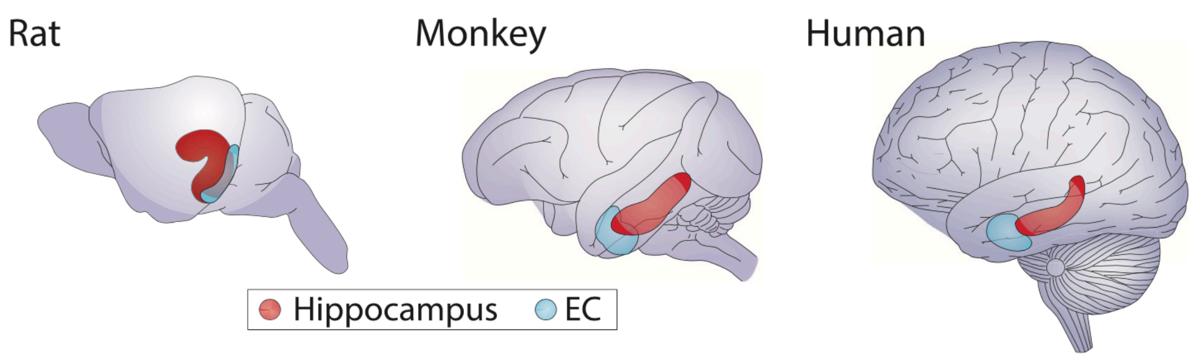


2D continuous attractors (torus)



Grid cells in medial entorhinal cortex





Accurate Path Integration in Continuous Attractor Network Models of Grid Cells

Yoram Burak[1,2]*, Ila R. Fiete[2,3]