

Models of hippocampal computations

A brief introduction to how the
hippocampus encodes memories and
supports spatial navigation

Learning Objectives

Discuss a model of the whole hippocampus and understand how the dentate gyrus and CA3 might be involved in pattern separation and pattern completion respectively

Discuss the continuous attractor network model of place cell firing

To understand the concept of path integration and how it might contribute to spatial navigation and place cell firing

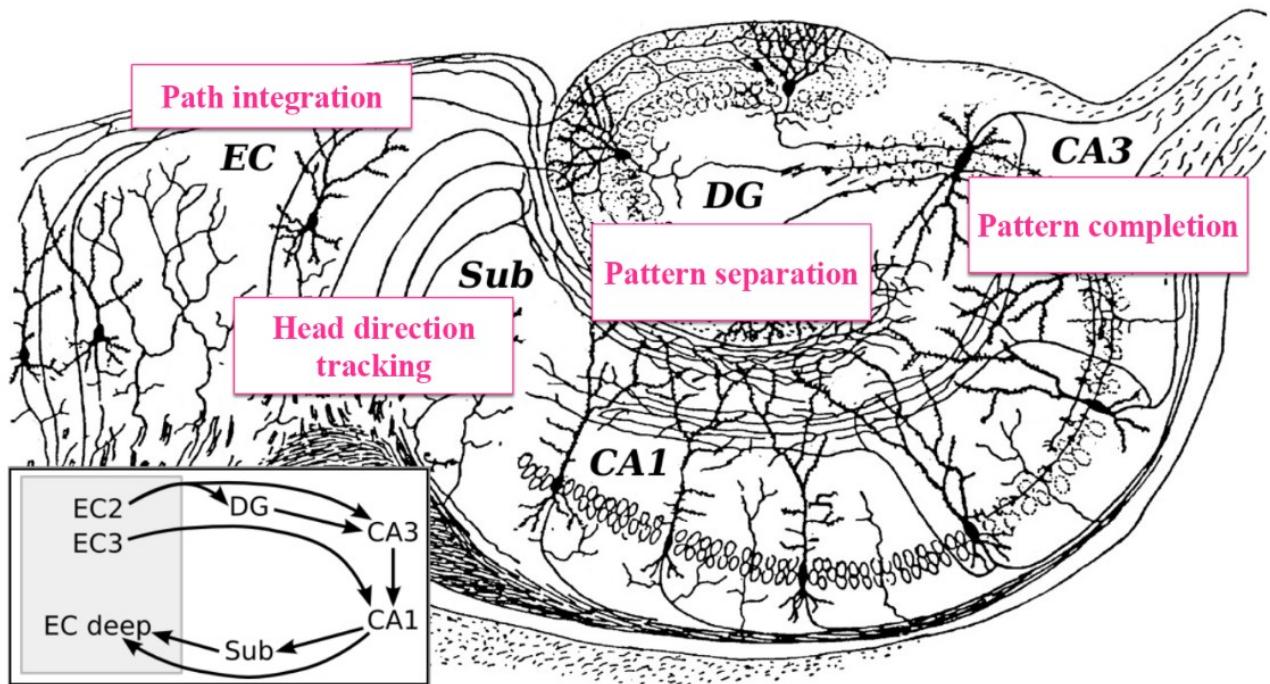
Describe the firing pattern of grid cells in entorhinal cortex and why they might be suited to provide the path integration input to place cells

Attractor networks in the hippocampus

Attractor networks in the hippocampus

- 1) Point attractors for memory recall (CA3 and DG)
- 2) Ring continuous attractors for tracking head-direction (HDCs in subiculum) and location (place cells in CA3)
- 3) Torus continuous attractors for path integration (EC)

Computations of the hippocampus



Long term memory

Long term memory tends to refer to information that is held indefinitely and includes episodic, semantic and procedural memory.

There are 3 stages within long term memory

- 1) Encoding
- 2) Storage
- 3) Recall



Connectivity within the hippocampus

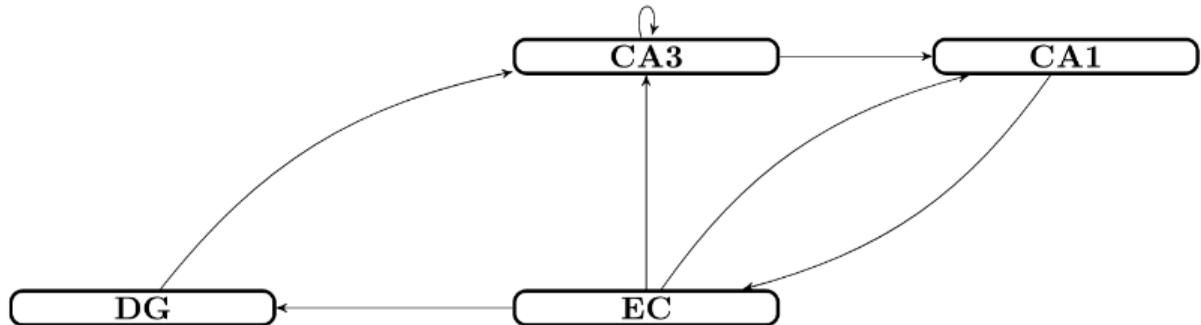


Figure 2: Connectivity of the hippocampus. A rough diagram showing the major connections between the areas of the connectivity. The set of axons running from EC to DG, CA3 and CA1 is called the perforant pathway, the mossy fibres run from DG to CA3 and the Schaffer collateral fibers go from CA3 to CA1. The loop on CA3 is supposed to represent the high level of recurrent connections in that region.

Auto-associative memory in the hippocampus

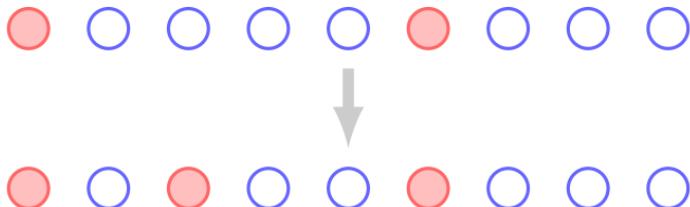
The standard paradigm for memory in the hippocampus is **auto-associative memory**. The hippocampus implements a network for this.

As we discussed auto-associative memories are patterns that represent memories along with some dynamics that complete partial patterns.

Imagine a sequence of on-off neurons where the filled circles correspond to on.



Recall occurs when the network is presented with a partial pattern and evolves into the complete pattern.



Computational models of hippocampal memory encoding recall

The most plausible site for this auto-associative network is CA3

- this area is the only one with **recurrent connections** between the principal cells (that is the most common cell type whose axons extend to other brain regions)
- In CA3, pyramidal cells are typically connected to other pyramidal cells

A leading theory is that distinct episodic memories are stored as attractors in CA3 (O'Reilly and McClelland, 1994)

The attractors in memory circuits are typically described as point attractors:

- the dynamics of the system evolve to one stable point.

During the encoding of a new memory we want to create a new attractor.

Computational models of hippocampal memory encoding and recall

During the encoding of a new memory we want to create a new attractor (set of neurons that are active for that memory). This allocation is done at random, via pattern separation mechanisms implemented by the dentate gyrus, DG.

The DG acts as a kind of teacher, activating some random set of neurons in CA3.

Encoding then involves **Hebbian plasticity** both in the EC to CA3 synapses and in the recurrent CA3-CA3 synapses.

During memory recall the external cue signal is routed via EC, completed by CA3 and then read out by CA1 (**synaptic plasticity is not involved in recall**)

Note that this implies content-addressable memory (you can recall an item based on partial information of the item itself, rather than by knowing a location in memory as in human made computers)

A highly simplified model of CA3

In this model, CA3 is made up of McCulloch-Pitts neurons and connectivity is all-to-all, as in a Hopfield network.

As before, let N be the number of neurons, x_i the activity of neuron i and w_{ij} the strength of the connection from i to j .

The sparseness (the average proportion of neurons active at any one time) is α

- α is believed to be very small in actual neurons.

For this simple model $w_{ij} = w_{ji}$ (the weights are symmetric)

During learning the patterns are activated and plastic changes are made to the synapse strengths according to a simple correlation based Hebbian plasticity rule.

Hebbian plasticity learning rule

$$\delta w_{ij} = \eta x_i x_j$$

where w_{ij} is the strength of the synapse from neuron i to neuron j, x_i and x_j are the states of the two neurons and η is a learning rate. Another version is:

$$\delta w_{ij} = \eta(x_i - a)(x_j - a)$$

where having a allows for different cut-off points between the behaviour that causes potentiation or depression. This is much more realistic.

By choosing a and rescaling η we get a correlation based Hebbian plasticity rule:

$$\Delta w_{ij} = \frac{\eta}{4}(x_i + 1 - 2\alpha)(x_j + 1 - 2\alpha)$$

Hebbian plasticity learning rule in CA3

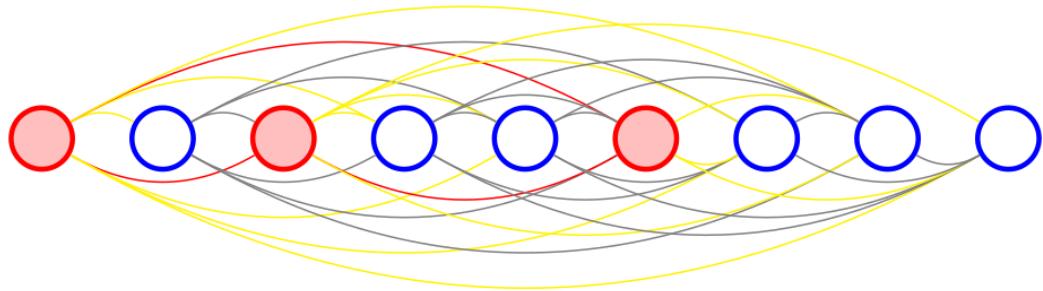
$$\Delta w_{ij} = \frac{\eta}{4}(x_i + 1 - 2\alpha)(x_j + 1 - 2\alpha)$$

where $\eta/4$ is the learning rate: this is often a small number (the four is just for notational convenience)

- but, in hippocampus, were memories need to be learned very quickly and sometimes after only a single presentation, η is large.

Since α is very small for real networks there will be a large increase, $\tilde{\eta}$ for the connection between two neurons that are active at the same time ($x_i = 1$ and $x_j = 1$), a tiny increase $\tilde{\eta}\alpha^2$ for pairs of neurons that are inactive at the same time ($x_i = -1$ and $x_j = -1$) and a medium size decrease $-\eta\alpha$ for pairs of neurons where one is active ($x_i = 1$) and one inactive ($x_j = -1$).

Hebbian plasticity in CA3



Learning in the associate network. The pattern has been imposed and connection strengths are changed.

$$\delta w_{ij} = \eta(x_i - a)(x_j - a)$$

The red links increase by $\eta(1 - a)^2$ and the gray by $\eta(-a)^2$, the yellow links decrease by $\eta a(1 - a)$.

Hebbian plasticity in CA3

During recall some of the neurons are held in the active state and the rest of the network evolves according to a threshold input rule. That means each neuron has an input given by:

$$h_i = \sum w_{ij} x_j$$

and is set in the active state if $h_i > \theta$ where θ is a threshold which is set to different values for different networks.

The idea is that after learning the pattern $\{0, 2, 5\}$



the connections between these nodes will be strong, so if the network has nodes $\{0, 5\}$ activated



the value $h_2 = w_{12} + w_{52}$ will be larger than threshold and the subsequent dynamics will switch neuron 2 on.

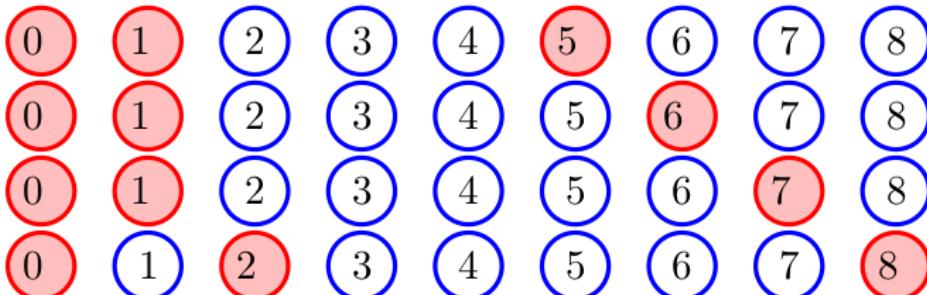
Capacity and correlated patterns

The estimates of capacity assume that the patterns are all independent.

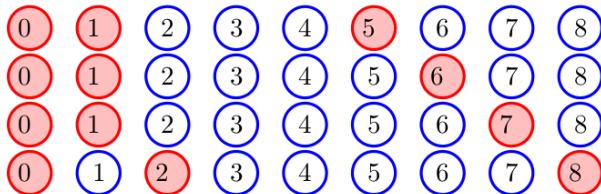
If they aren't, the capacity is reduced. If patterns share some fragments or subpatterns then the connections in these subpatterns become very strong, perhaps dominating other elements in the patterns

- these may be the elements that make them different!

Consider the four patterns:

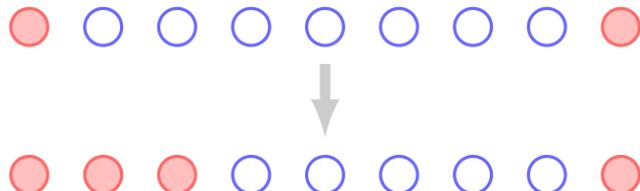


Correlated patterns



The connection between neurons 0 and 1 will become very strong because this connection is present in three patterns out of four.

It is likely that this will result in this erroneous completion:



Or:



Implications of correlated patterns

This means that auto-associative networks are not able to effectively store anything except random patterns!

- This is why they have never proved useful for machine learning

In cortical memory, since there are multiple presentations supported by a hippocampal representation, the memory system has the opportunity to learn different, similar, memories.

The hippocampal circuitry is critical in our ability to learn correlated patterns.

The goal here, however, is to learn the memory quickly after a small number of presentations!

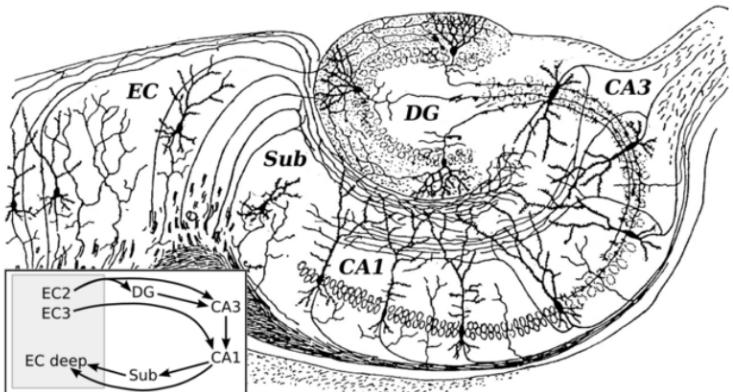
Hippocampal circuitry that enables memories for correlated patterns

In the case of hippocampus it has been proposed, that this problem is solved through the EC-DG-CA3 pathway

- one role of the dentate gyrus may be to **randomize** the connectivity between EC and CA3.

During learning, neurons in EC and CA3 are matched via DG and connections from EC to DG and from DG to CA3 are essentially random.

In this way the DG is thought to be critical for **pattern separation**.



Hippocampal circuitry that enables memories for correlated patterns

This randomisation of connections reduces overlap through a k-winner takes all mechanism.

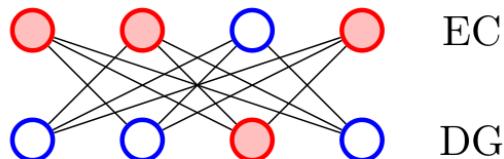
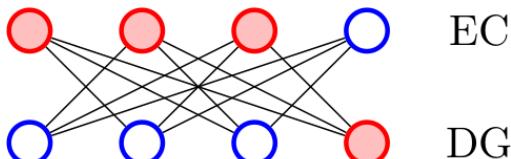
winner-take-all: neurons in a layer compete for activation

Roughly, it is supposed that local inhibition ensures that the k most active neurons in the DG layer ‘win’.

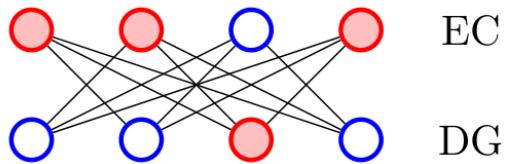
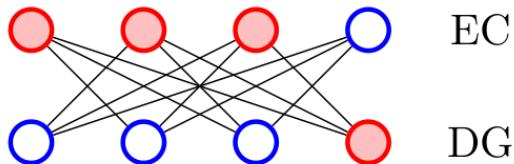
local inhibition: neurons suppress the activity of neurons in the same layer

This might reduce overlap between pattern representations in the brain as shown below where $k = 1$ and two similar patterns result in a different neuron being active:

VS.



Hippocampal circuitry that enables memories for correlated patterns



This randomization might be repeated in the subsequent connection between DG and CA3.

This mechanism may explain why neurons in DG are being born all the time (neurogenesis: new neurons even in adulthood), perhaps their role is to create these random connections.

Models of the whole hippocampus

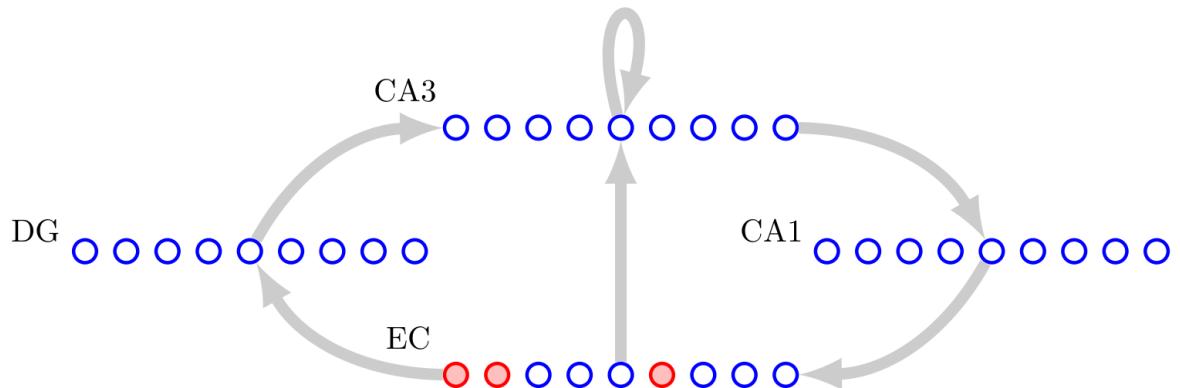
We are now almost in a position to consider models of the hippocampus as a whole:

- we know that CA3 may act as an auto-associative memory network performing pattern completion
- we know that the DG may act to randomise inputs from EC to CA3 to perform pattern separation
- we know that the EC provides the hippocampus with the patterns to be remembered

... but, so far CA1 hasn't been mentioned.

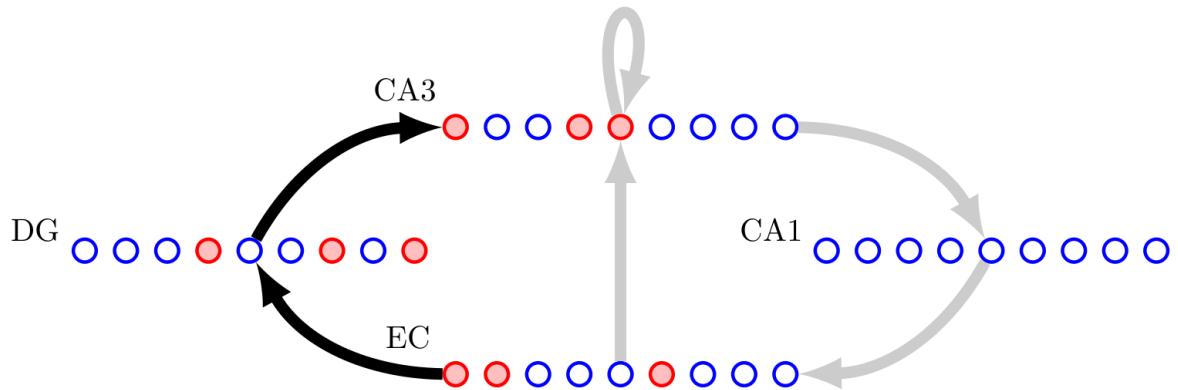
It has been suggested that the role of CA1 is to relay patterns back to the EC.

A model of the whole hippocampus



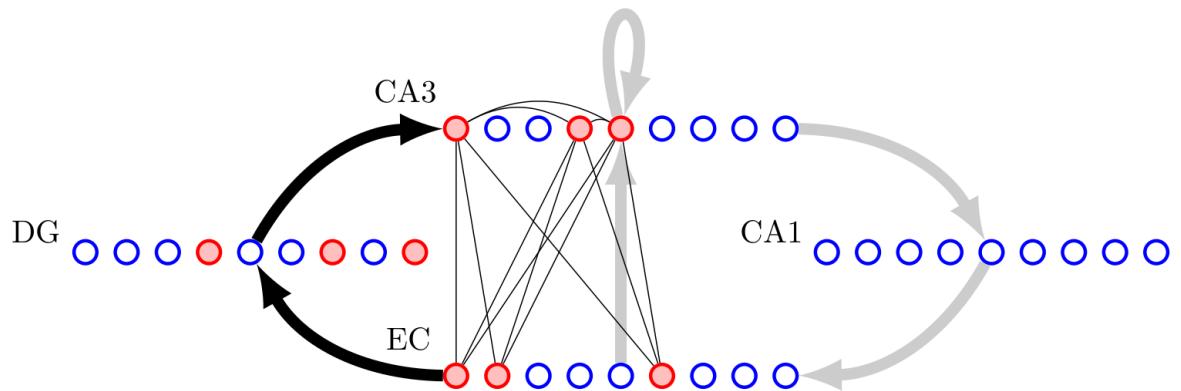
Connections, possibly random, between EC and DG and between DG and CA3, along with the 'k-winner takes all' mechanism, causes activity in CA3.

A model of the whole hippocampus



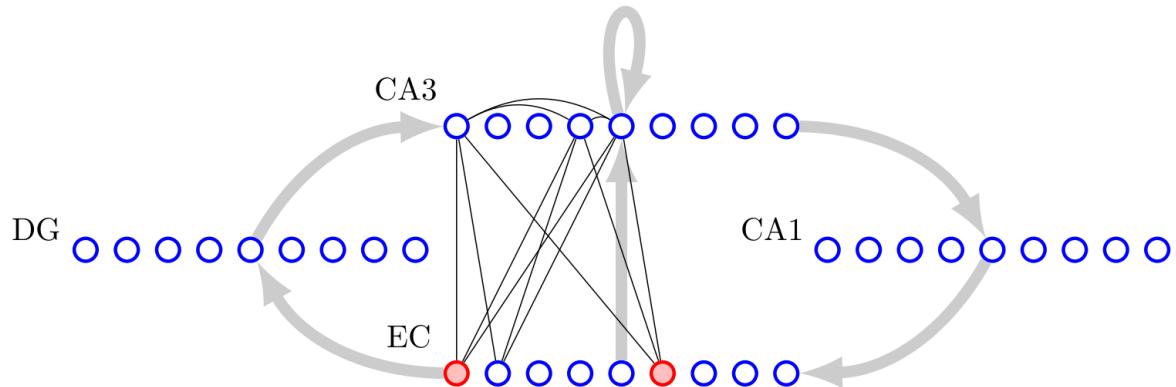
This representation is then learned by Hebbian plasticity between EC and CA3.

A model of the whole hippocampus

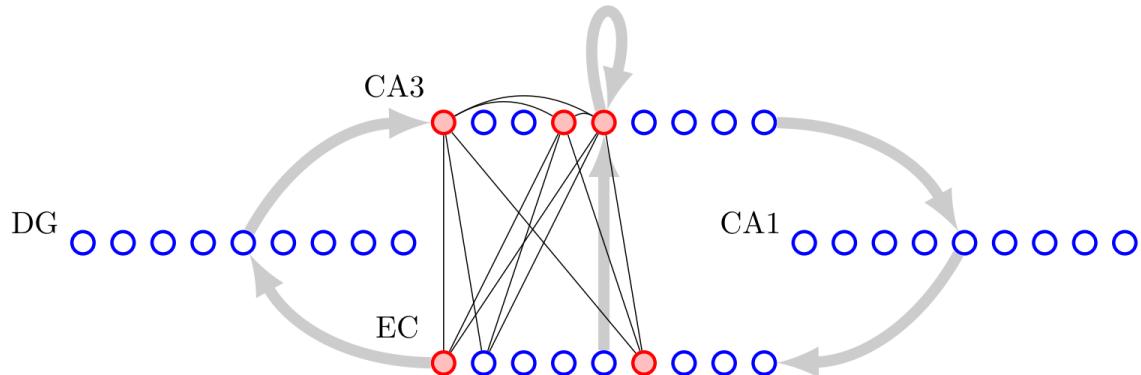


Hebbian learning also strengthens links to map the pattern to CA1 and link that to EC.

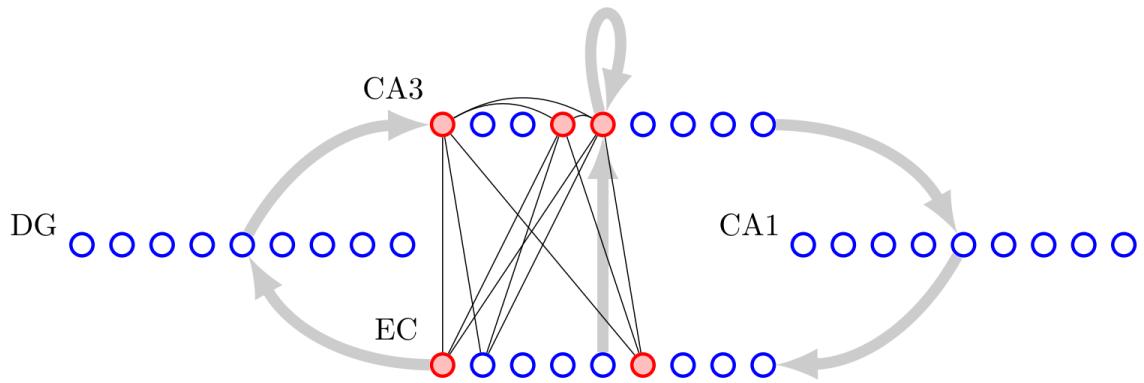
A model of the whole hippocampus



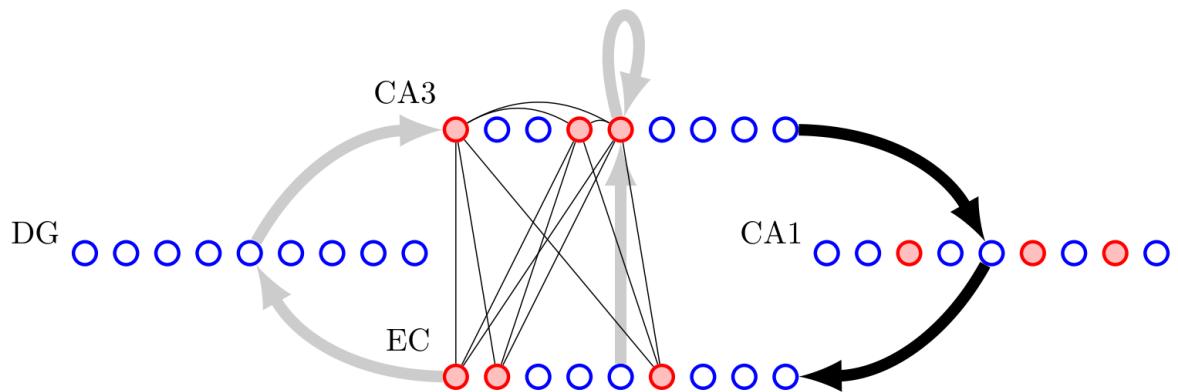
Because of the connections from EC to CA3 and the recurrent connections in CA3, this excites the pattern in CA3:



A model of the whole hippocampus



This memory is sent back to EC via CA1, and recall has occurred!



A model of the whole hippocampus

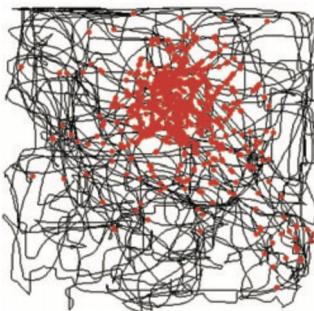
This doesn't explain how the hippocampus switches between the learning and retrieval phase.

One suggestion is just that the level of stimulation is different
- larger activity during learning excites the pathway that goes via
DG

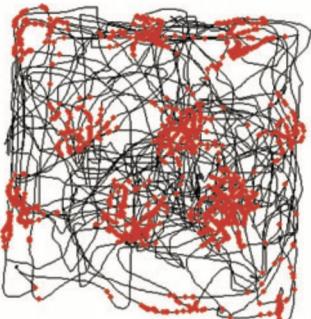
Another suggestion is that the neuromodulator acetylcholine or dopamine
is important.

Model of the ‘cognitive map’

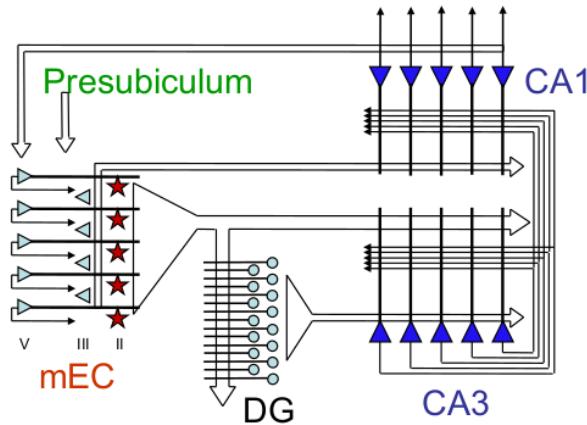
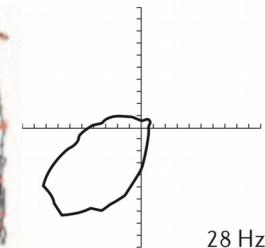
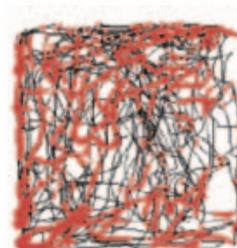
Place cell



Grid cell



Head direction cell



As we have seen
the hippocampus is
also involved in
spatial navigation

Continuous attractor network models of place cell firing

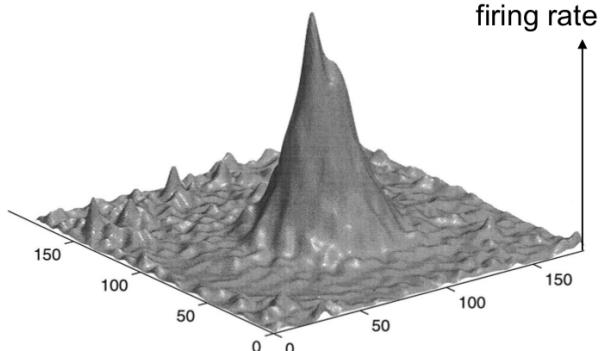
An attractor network is a network of connected nodes whose time **dynamics settle to a stable pattern**, e.g. in the Hopfield model.

Attractor representations are patterns of neural activity to which **similar patterns of activity evolve** under the dynamics of the system

A ‘continuous attractor’ is a set of patterns to which other patterns are attracted, but within which the pattern of activity can change smoothly from one to another.

Zhang (1996) and McNaughton et al. (1996) proposed that the patterns of **activity shown by place cells formed a continuous attractor**

- all of them can be thought of as having the same ‘bump’ shape:



Continuous attractor network models of place cell firing

Imagine the place cells arranged in a sheet so that each cell's location reflects the location of its firing field in the environment

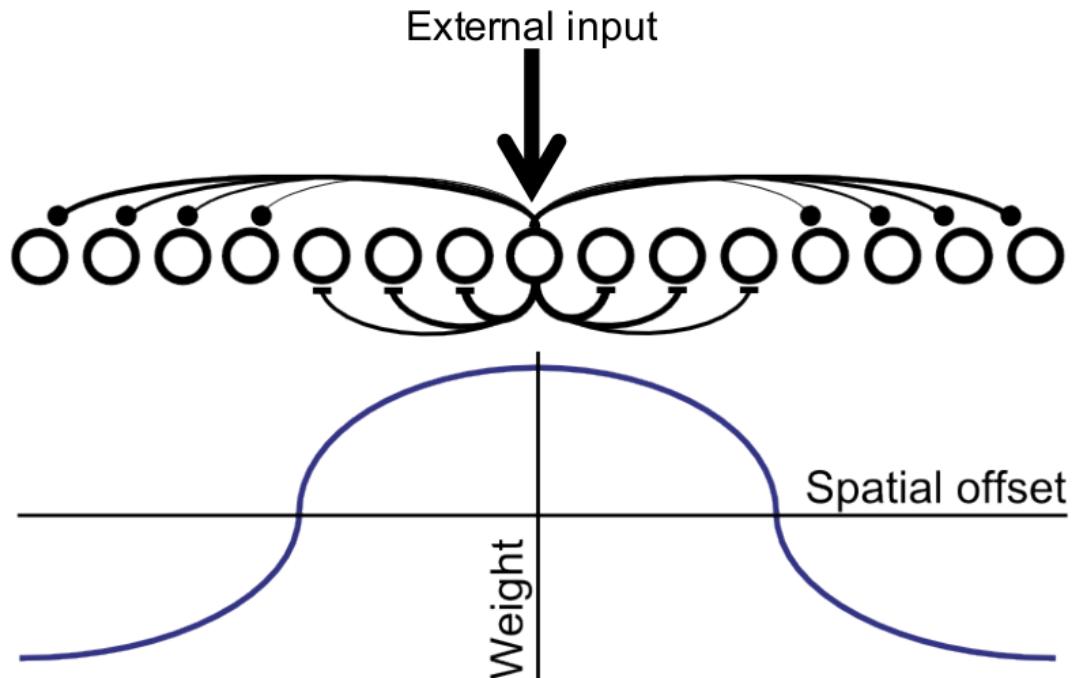
- The bump indicates the rat's location and moves as the rat moves.

The simplest example of a continuous attractor network is on a linear track.

The requisite synaptic connectivity is often referred to as 'Mexican hat'

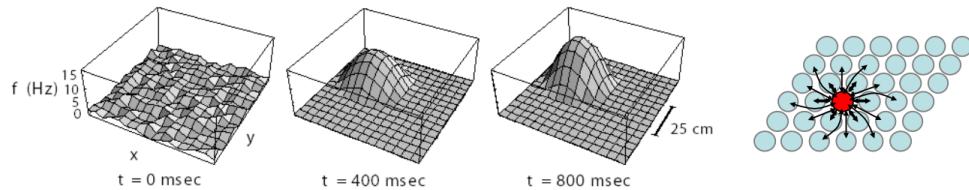
It is important to remember that the physical arrangement of these cells in the brain does NOT reflect the arrangement of their firing fields in the environment!

Continuous attractor network models of place cell firing

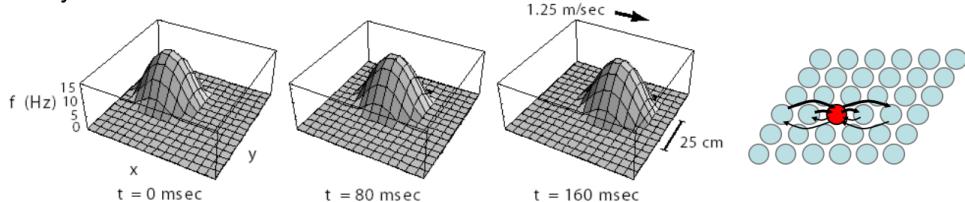


Continuous attractor network models of place cell firing

We can then generalise this model to two dimensions to form a continuous attractor, such that the ‘activity bump’ can move smoothly in any direction:



Adding asymmetric connections in a particular direction makes the activity bump shift in that direction, with a speed proportional to the strength of the asymmetric connections



Continuous attractor network models of place cell firing

This is a 2-dimensional model that relies on speed and direction information alone.

- Head direction and running speed gets combined

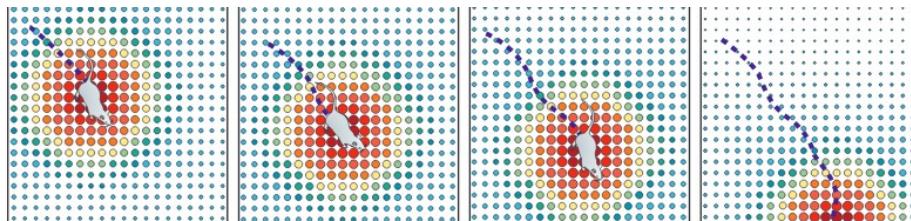
If the asymmetric connections between place cells along a given direction have weights proportional to the speed of movement in that direction, then the activity bump will automatically track the location of the rat using only these motion signals: i.e. the network performs ‘path integration’.

But: what about boundaries?

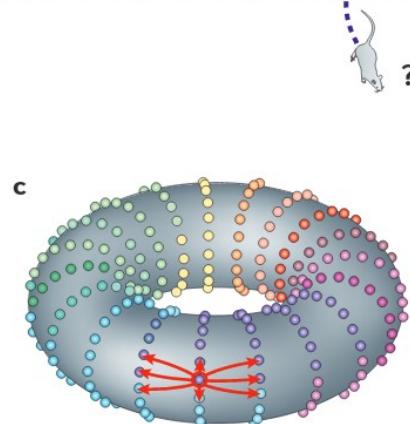
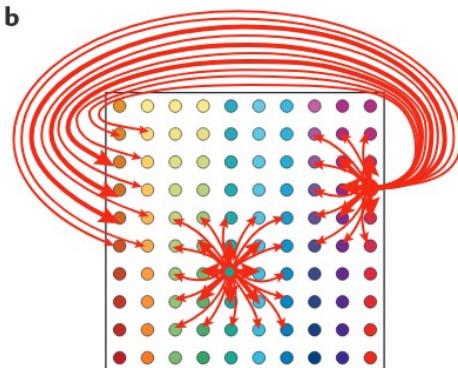
- what happens when the rat runs outside the area that the cells represent? – the problem of edge effects

Torus continuous attractor network models with grid cell firing

Periodic fields are necessary to overcome the boundary problem.



Grid cells form a grid-like structure of place fields repeating at regular intervals over the entire environment.



As an animal moves through its environment, the location-specific activity in the grid cell network is (probably!) updated principally by a path integration based mechanism.

Summary of network models performing path integration

Path integration is the process by which an animal updates its location (relative to some starting position) by keeping a cumulative record of internal movement information.

Path integration relies on place cells, HDCs and grid cells

A small set of grid cells could help perform path integration - updating their relative firing rates to track movement (distance!) of the animal

Connections between grid cells could be learned at the multiple locations within an environment at which they fire

The grid cells could thus provide the path integration input to place cells: place cells fire when all of the grid cell inputs overlap

Conclusions

The CA3 region of the hippocampus supports auto-associative memory

The CA3 is provided with patterns from EC, which are randomised by DG

The circuitry of the hippocampus is critical for processing episodic memories: both encoding and recall

Place cells form a continuous attractor that helps to compute path integration information from the grid cells, which themselves can be modelled using a toroidal continuous attractor

Reference: McNaughton, B., Battaglia, F., Jensen, O. et al. Path integration and the neural basis of the 'cognitive map'. *Nat Rev Neurosci* 7, 663–678 (2006).
<https://doi.org/10.1038/nrn1932>