

Statistical Physics of Collective Processes in Biology

Rémi Monasson

*Laboratory of Theoretical Physics,
CNRS & Ecole Normale Supérieure, Paris*

Spin Glasses 2014, Cargèse, August 2014

Statistical Physics of Collective Processes in Biology

\cap Spin Glasses ...

= *The Hopfield model revisited*
+ ...

Why?

- Historical and scientific importance (good to look back at the model after 3 decades ...)
- Experimental context has changed (enormously!)
- Theoretical extensions
- Interest well beyond neuroscience

The paper (1982)

Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 2554–2558, April 1982
Biophysics

Neural networks and physical systems with emergent collective computational abilities

(associative memory/parallel processing/categorization/content-addressable memory/fail-soft devices)

J. J. HOPFIELD

Division of Chemistry and Biology, California Institute of Technology, Pasadena, California 91125; and Bell Laboratories, Murray Hill, New Jersey 07974

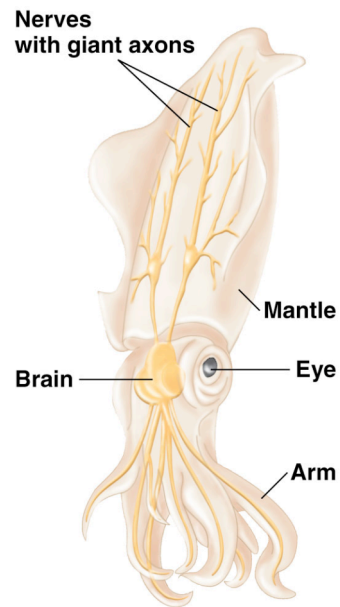
Contributed by John J. Hopfield, January 15, 1982

ABSTRACT Computational properties of use to biological organisms or to the construction of computers can emerge as collective properties of systems having a large number of simple equivalent components (or neurons). The physical meaning of content-addressable memory is described by an appropriate phase space flow of the state of a system. A model of such a system is given, based on aspects of neurobiology but readily adapted to integrated circuits. The collective properties of this model produce a content-addressable memory which correctly yields an entire memory from any subpart of sufficient size. The algorithm for the time evolution of the state of the system is based on asynchronous parallel processing. Additional emergent collective properties include some capacity for generalization, familiarity recognition, categorization, error correction, and time sequence retention. The collective properties are only weakly sensitive to details of the modeling or the failure of individual devices.

- Autoassociative memory
- Simple dynamics of components
(no clock)
- Generalization, error correction, time sequence storage, ...
- Robustness to failure of individual components

Ingredients of the model (1)

- Neuron = Ising spin $s_i(t) = 0, 1$ or $-1, +1$



© 2012 Pearson Education, Inc.

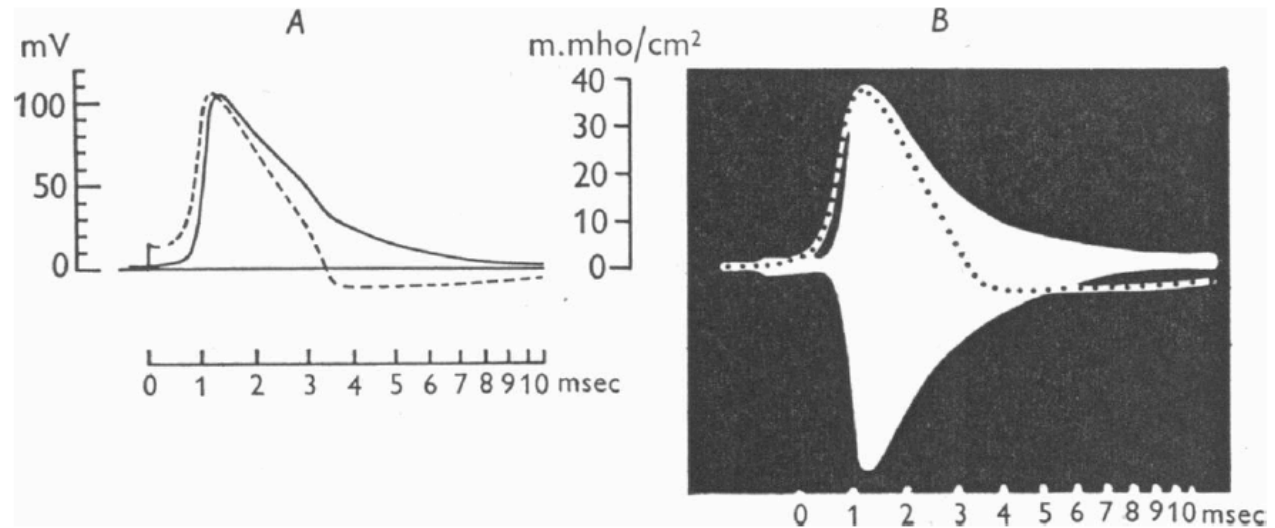


Fig. 16. *A*, solution of eqn. (26) for initial depolarization of 15 mV at a temperature of 6° C. The broken curve shows the membrane action potential in mV; the continuous curve shows the total membrane conductance ($g_{Na} + g_K + \bar{g}_l$) as a function of time. *B*, records of propagated action potential (dotted curve) and conductance change reproduced from Cole & Curtis (1939). The time scales are the same in *A* and *B*.

Hodgkin & Huxley (1952)

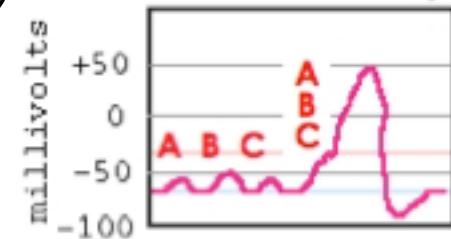
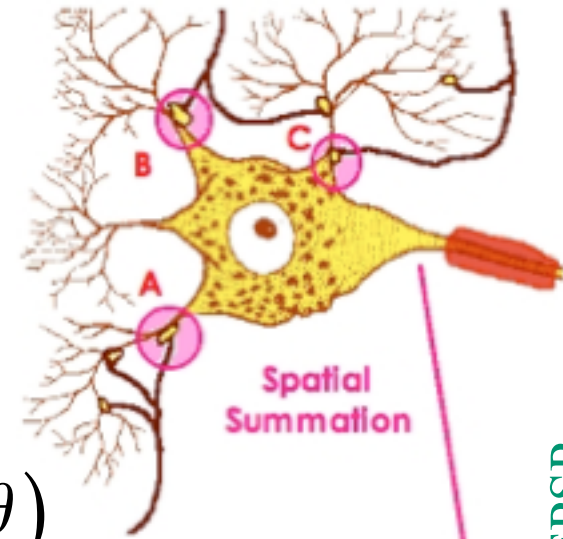
Ingredients of the model (2)

State of neuron determined by
linear summation + threshold rule:

$$S(t+1) = \text{sign}(J_A s_A(t) + J_B s_B(t) + J_C s_C(t) + \dots - \theta)$$

J = synaptic interaction

Sequential updating
(time scale about 10 ms)



A, B, C = EPSP
Excitatory Post-Synaptic Potential

A, B, C Each of these firings alone causes a partial depolarization but not enough for an action potential.

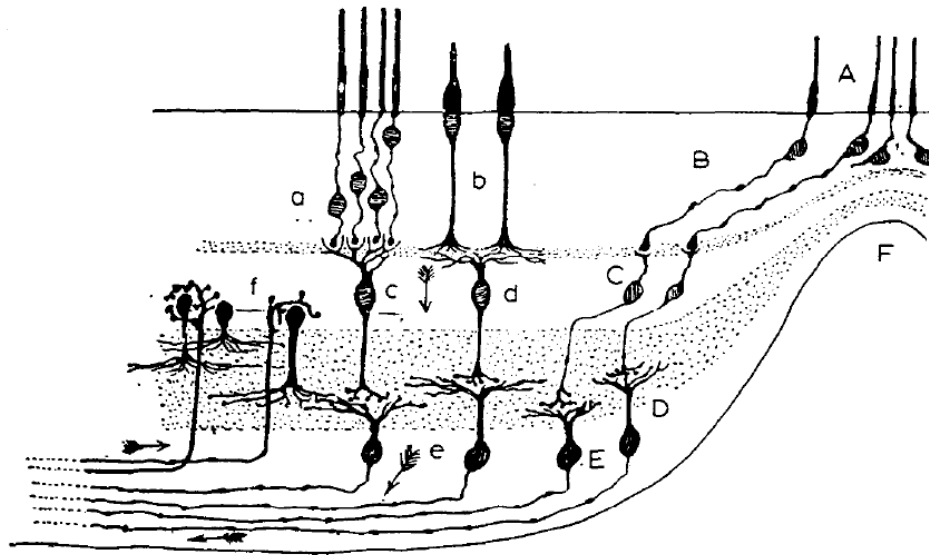
But, if A,B,C fire simultaneously their combined effects will cause an action potential

Ingredients of the model (3)

- Synapse = interaction J_{ij} obeying Hebb's rule

When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased

D. Hebb, *The organization of behavior* (1949)

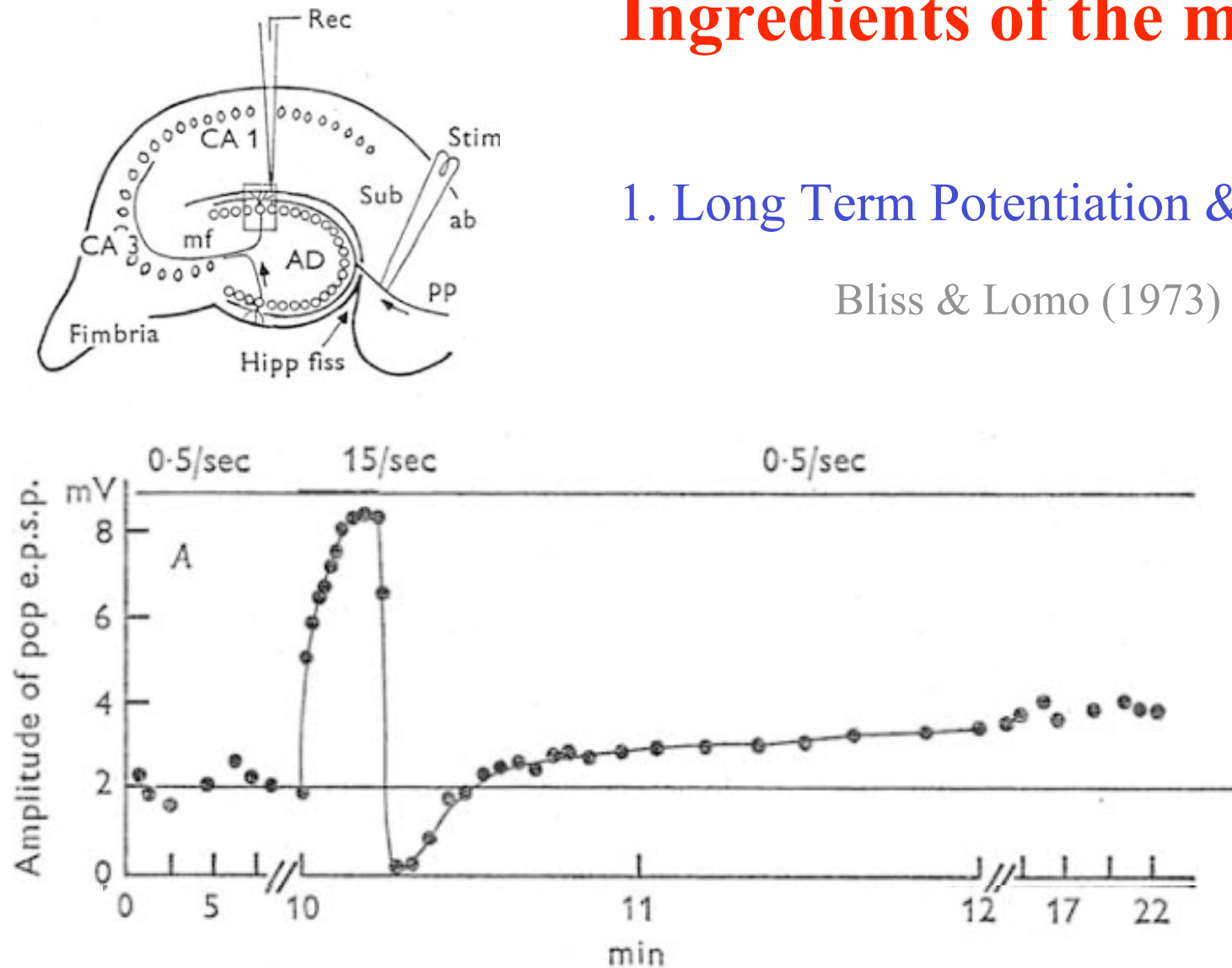


Connectionnism postulated by Ramon Y Cajal as soon as 1894: memory comes from changes in connections rather than appearance of new neurons

Ingredients of the model (4)

1. Long Term Potentiation & Depression

Bliss & Lomo (1973)



+ other
scheme
for LTD

2. Spike-timing dependent plasticity (from 10 ms to few minutes)

Ingredients of the model (5)

LETTER

Long Term Depression/Potentiation

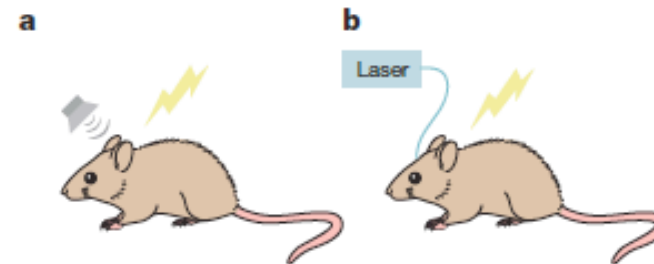
doi:10

Engineering a memory with LTD and LTP

Sadegh Nabavi^{1*}, Rocky Fox^{1*}, Christophe D. Proulx¹, John Y. Lin², Roger Y. Tsien^{2,3} & Roberto Malinow¹

It has been proposed that memories are encoded by modification of synaptic strengths through cellular mechanisms such as long-term potentiation (LTP) and long-term depression (LTD)¹. However, the causal link between these synaptic processes and memory has been difficult to demonstrate². Here we show that fear conditioning^{3–8}, a type of associative memory, can be inactivated and reactivated by LTD and LTP, respectively. We began by conditioning an animal to associate a foot shock with optogenetic stimulation of auditory inputs targeting the amygdala, a brain region known to be essential for fear conditioning^{3–8}. Subsequent optogenetic delivery of LTD conditioning to the auditory input inactivates memory of the shock. Then subsequent optogenetic delivery of LTP conditioning to the auditory input reactivates memory of the shock. Thus, we have engineered inactiva-

the memory of the shock by LTD (Fig. 2a, b, f). No memories can be reactivated. To these animals we c LTP protocol (see Methods). One day later, anim



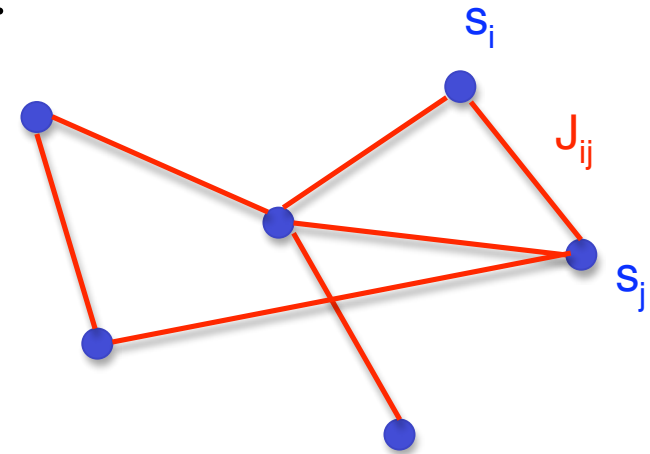
The model

- sets of uncorrelated patterns to be ‘stored’:

$$\xi_i^\mu = \pm 1$$

Index of pattern = 1, ..., P

Index of neuron = 1, ..., N

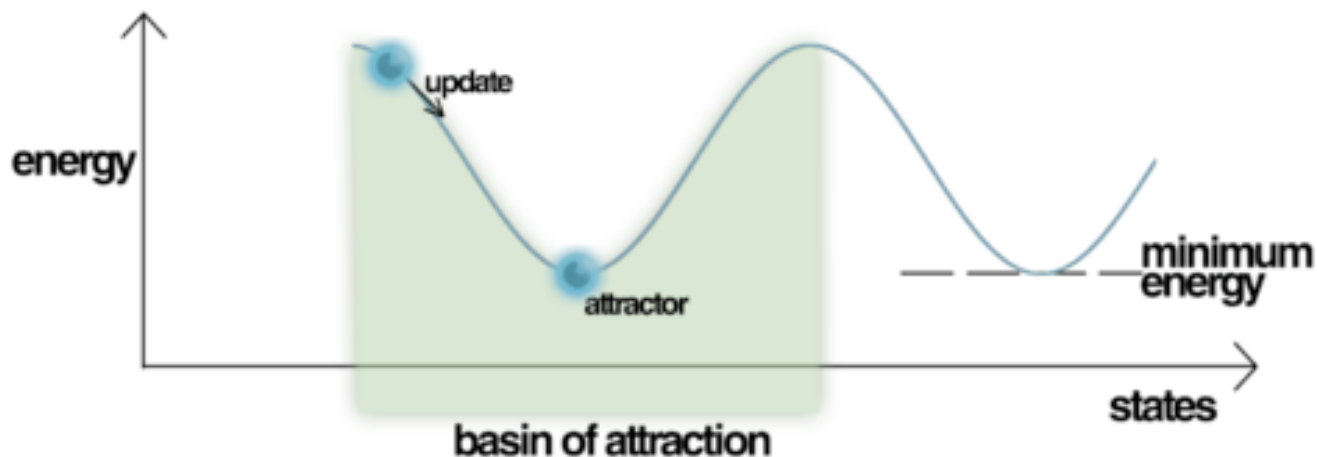


- synaptic interactions:
$$J_{ij} = \frac{1}{N} \sum_{\mu} \xi_i^{\mu} \xi_j^{\mu}$$
- updating rule:
$$s_i(t+1) = \text{sign} \left(\sum_j J_{ij} s_j(t) - h_i \right)$$

Q: Is the final state close to one of the patterns?

A: Yes, if number P of patterns small enough ...

$$E[s_1, s_2, \dots, s_N] = -\frac{1}{2} \sum_{i,j} J_{ij} s_i s_j + \sum_i h_i s_i$$



Fixed point condition (zero threshold):

$$\xi_i^\mu = \text{sign} \left(\sum_j J_{ij} \xi_j^\mu \right) = \text{sign} \left(1 + \sum_{v(\neq \mu)} \xi_i^v \left[\frac{1}{N} \sum_j \xi_j^v \xi_j^\mu \right] \right) \quad ; \quad \text{noise} = O\left(\sqrt{P/N}\right)$$

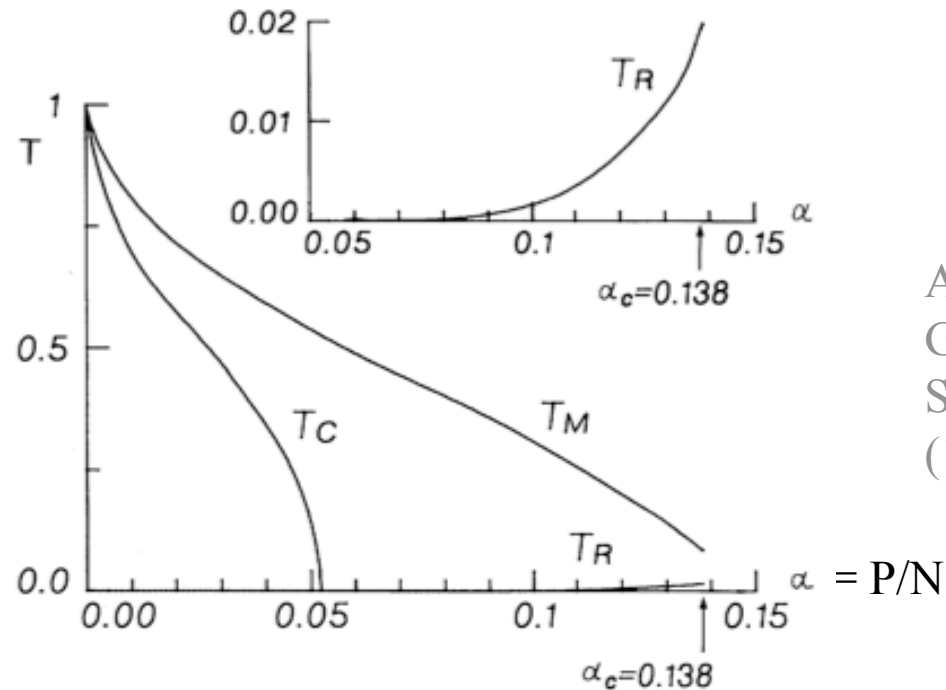
Analytical solution and phase diagram

Order parameters:

$$m^\mu = \frac{1}{N} \sum_i \overline{\langle s_i \rangle \xi_j^\mu}$$

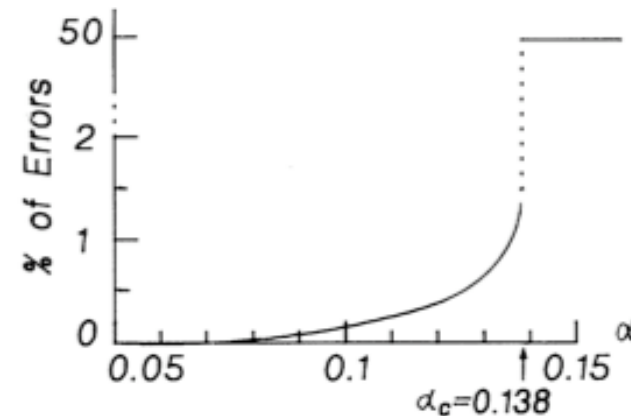
$$s = \frac{1}{P} \sum_{\mu > 1} (m^\mu)^2$$

$$q = \frac{1}{N} \sum_i \overline{\langle s_i \rangle^2}$$



Amit,
Gutfreund,
Sompolinsky
(1984)

Error at $T=0$:



Criticisms against the Hopfield model

- Symmetric J
- Oversimplified dynamics (Spikes? No adaptation?)
- Spurious states
- Attractors = specific activity configurations
- Experimental evidence?

Why is it interesting/important?

- One of the (very few) paradigms in theoretical neuroscience (Hebb's rule & attractors)
- Model enjoys many interesting properties
- Conceptually important (beyond neuroscience):
unique **J (coupling matrix)** allows for vastly different activity states \mathbf{s}
- Analytical solution two years later by AGS with spin glass techniques
- Opened the way to applications of spin glasses beyond physics (optimization problems, error-correcting codes, ...)
- Deep 'sociological' impact in the physics community towards theoretical biology

The Hopfield model today

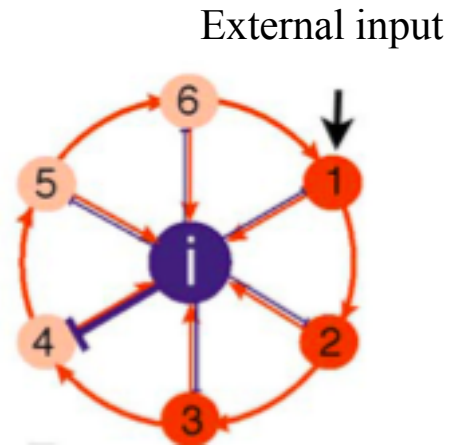
- Attractors in neuroscience: recent experimental « evidence »
+ theoretical developments

with special attention to the representation of space in the brain

- Statistical mechanics development & Inverse problems in biology:
 - Ising-based models and the inverse Hopfield problem
 - Applications to proteomics, gene networks, neuroscience, ...
(mostly this afternoon!)

Ring & continuous attractors (1)

$N = 6$ neurons



- balance of local excitation + global inhibition
= localization of activity (bump or cell assembly)
- if inhibition not global, more than one bumps, see later ...

A continuous-attractor model in statistical physics (1)

Translation-invariant + long-range interactions:
exactly solvable lattice-gas model for the
liquid/gas transition

Lebowitz and Penrose (1966)

$$E_J[\sigma] = - \sum_{i < j} J_{ij} \sigma_i \sigma_j$$

$$J_{ij} = J_w(\|i - j\|)$$

Order parameter =
Coarse-grained activity:

$$\rho(x) \equiv \lim_{\epsilon \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{\epsilon N} \sum_{(x - \frac{\epsilon}{2})N \leq i < (x + \frac{\epsilon}{2})N} \langle \sigma_i \rangle_J$$

Single spin
self-consistent
equations:

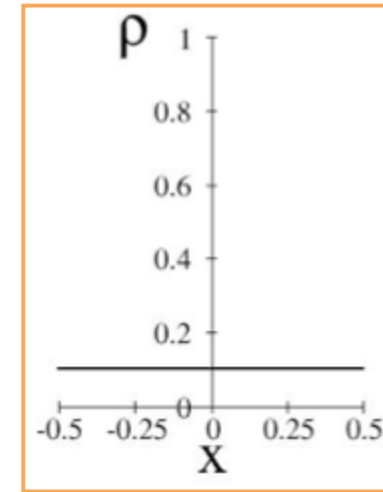
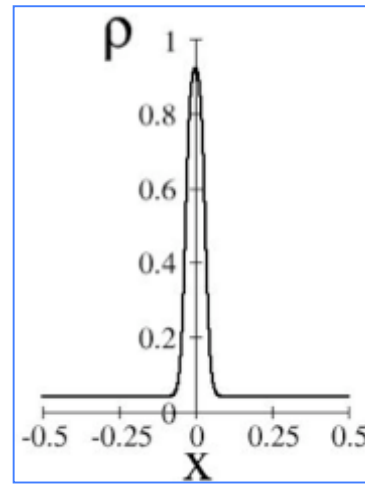
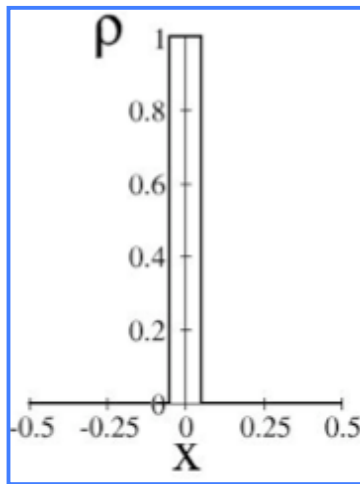
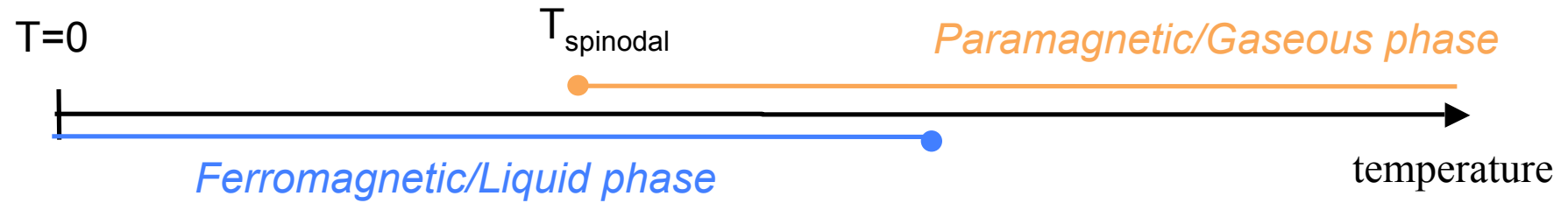
$$\rho(x) = \frac{1}{1 + e^{-\mu(x)/T}} ,$$

$$\mu(x) = \int dy J_w(x - y) \rho(y) + \lambda$$

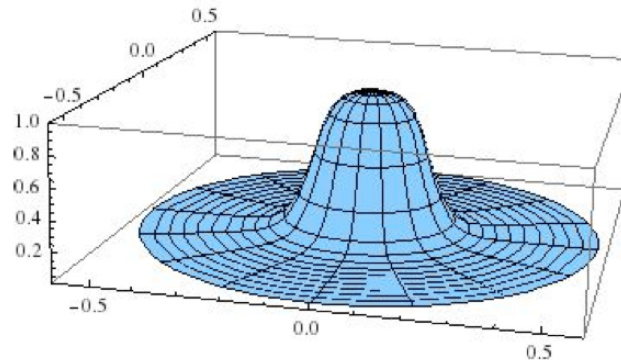

(imposes global activity)

(Similar to rate model for neurons ...)

A continuous-attractor model in statistical physics (2)

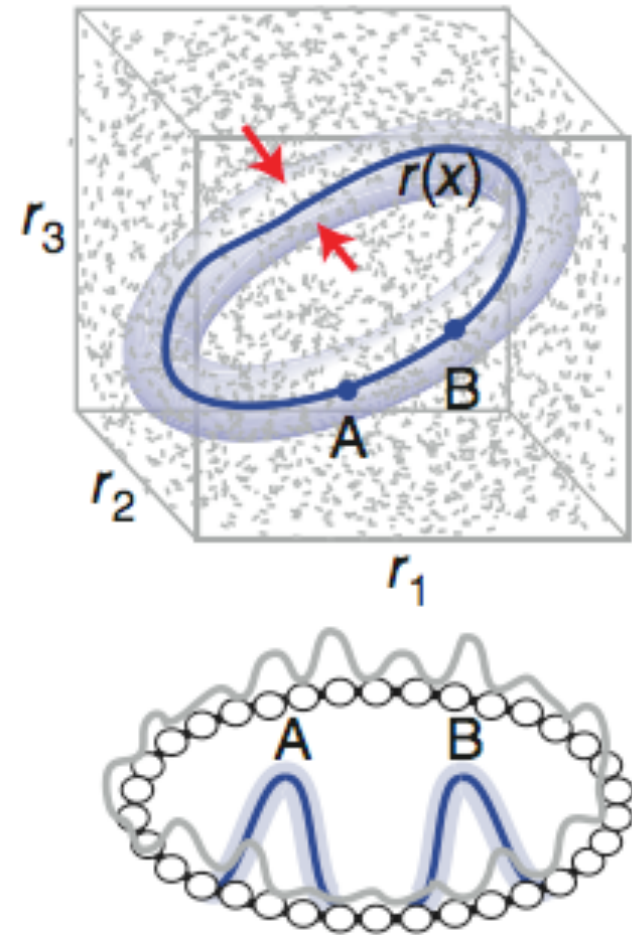


True also in
 $D=2$:

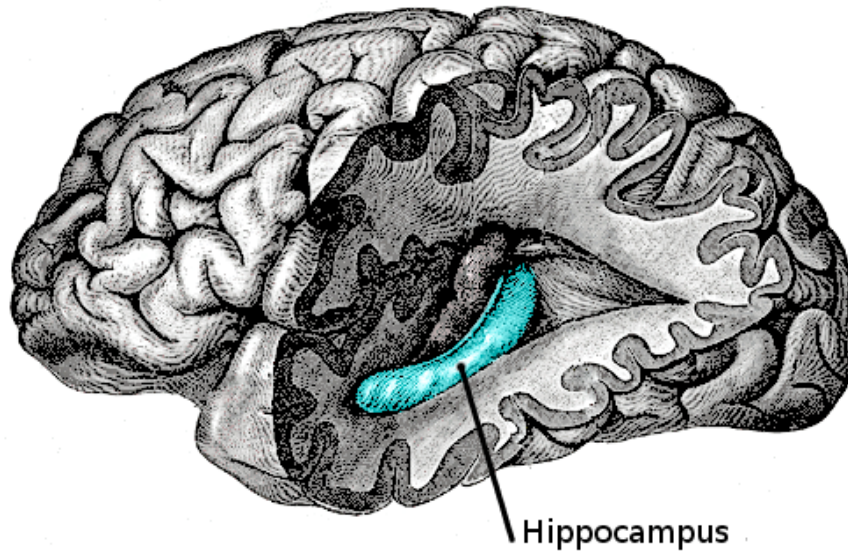


Ring & continuous attractors (2)

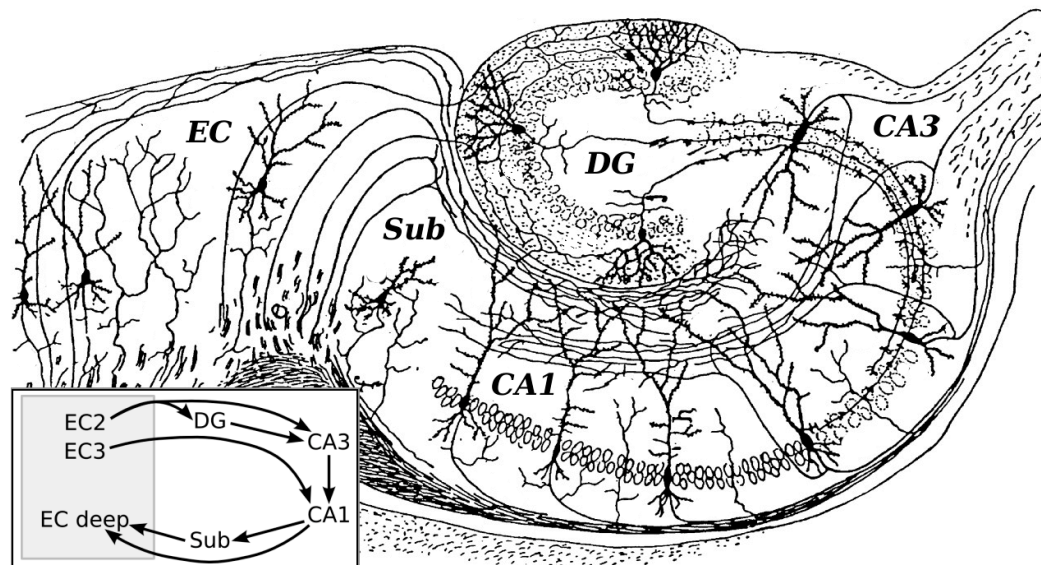
- active bump = collective coordinate for the neural activity (robust encoding)
- successive firing of neurons along the ring in 1D or higher D = continuous attractor
- bump driven by external input (stimulus or other neural activity) or diffusion
- presence of heterogeneities in the interaction?
- evidence for continuous attractors in experimental recordings



The representation of space in the brain

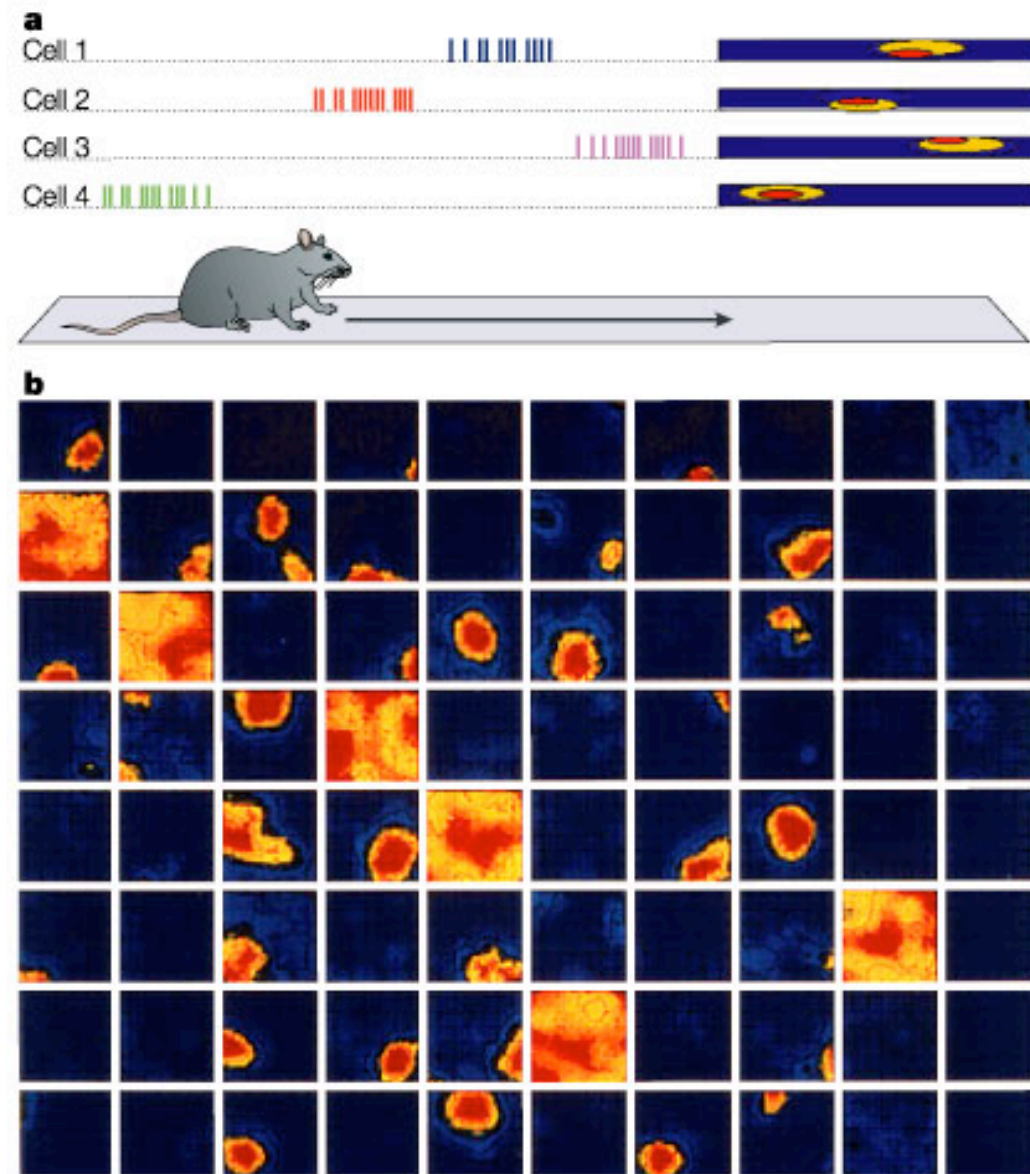


- necessary to form and retain new memories
- deeply intra-connected and connected to neighboring cortical regions, e.g. EC
- Hippocampus and EC fundamentally involved in the representation of space



O'Keefe, Dostrovsky (1971)

Place cells in the hippocampus (1)

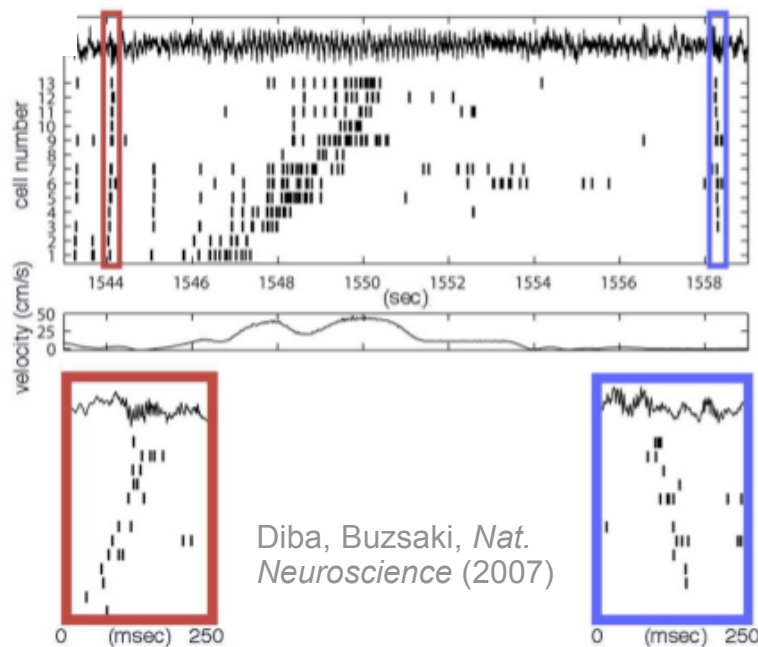
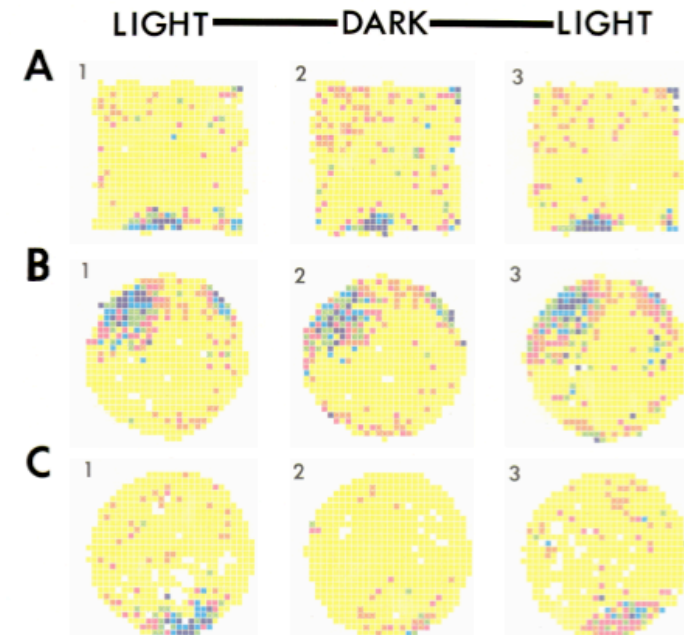


Nakazawa, McHugh, Wilson & Tonegawa
Nature Reviews Neuroscience **5**, 361-372 (2004)

Place cells in the hippocampus (2)

1. Place cell activity is largely maintained in the absence of visual cue

Quirk, Muller, Kubie,
J. Neuroscience (1990)



Diba, Buzsaki, *Nat. Neuroscience* (2007)

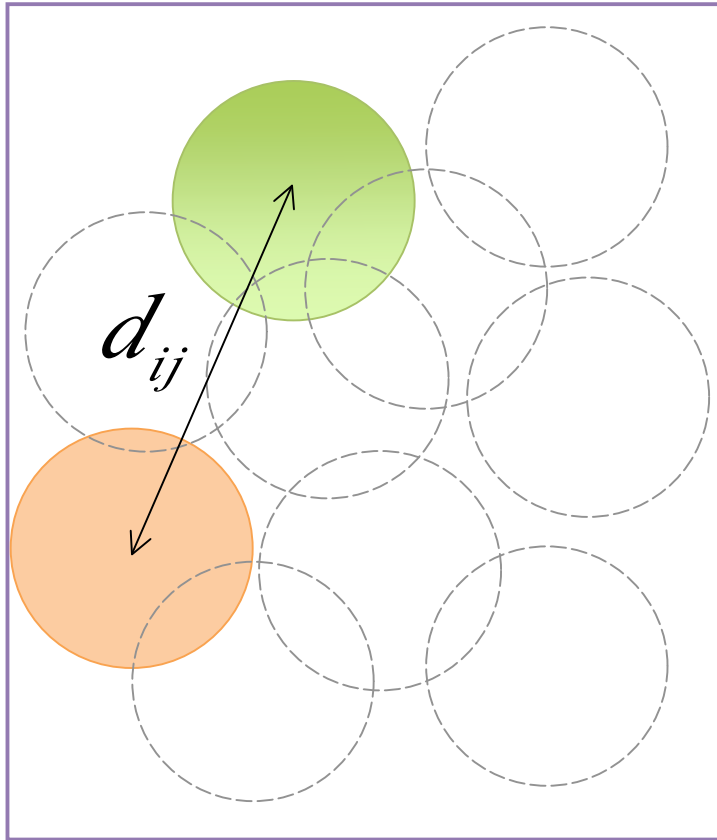
2. Place cell activity can take place on compressed time scales

⇒ Internally generated
(input independent)
network activity

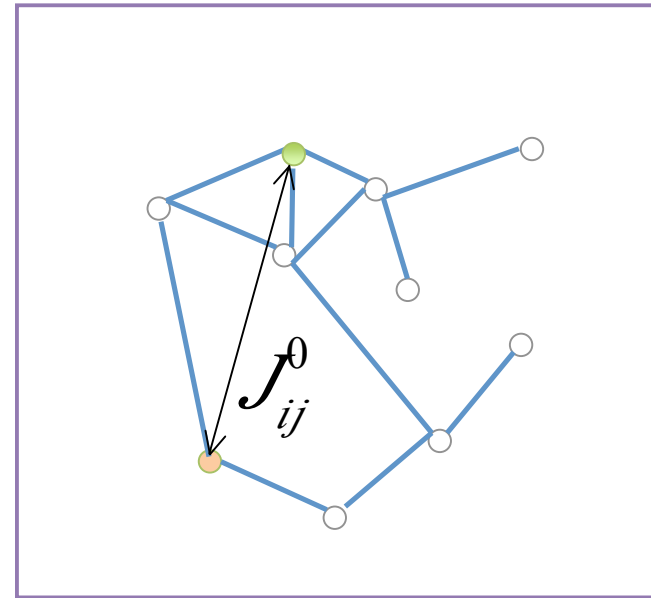
Attractor model for one environment

Neuron = binary state, silent or active : $\sigma_i = 0, 1$

Physical space



Neural network



$$J_{ij}^0 = \begin{cases} \frac{1}{N} & \text{if } d_{ij} \leq d_c, \\ 0 & \text{if } d_{ij} > d_c \end{cases}$$

Identical to Lebowitz-Penrose lattice-gas model (if perfect learning)!

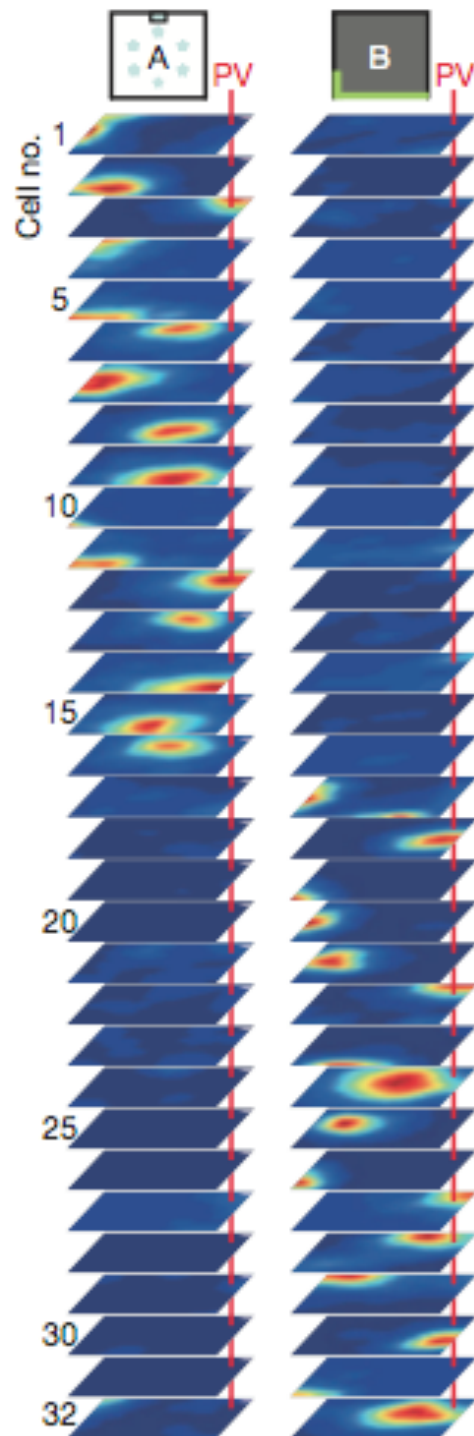
Place cells in the hippocampus (3)

1. Fundamental property: remapping

- *global remapping observed in new environment; statistically compatible with random reallocation of place fields*
- *only a fraction of cells active in one environment (orthogonalization of space representations)*
- *rate remapping follows slight changes (e.g. of cues)*

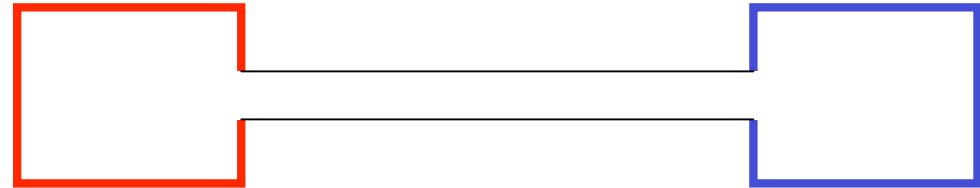
2. Context dependence: there is more to place cells than « place » ...

(overdispersion, dependence on other e.g. olfactory stimuli, on task, ...)

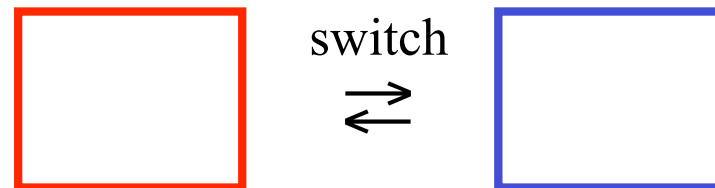


Teleportation (1)

- Rat in two different environments



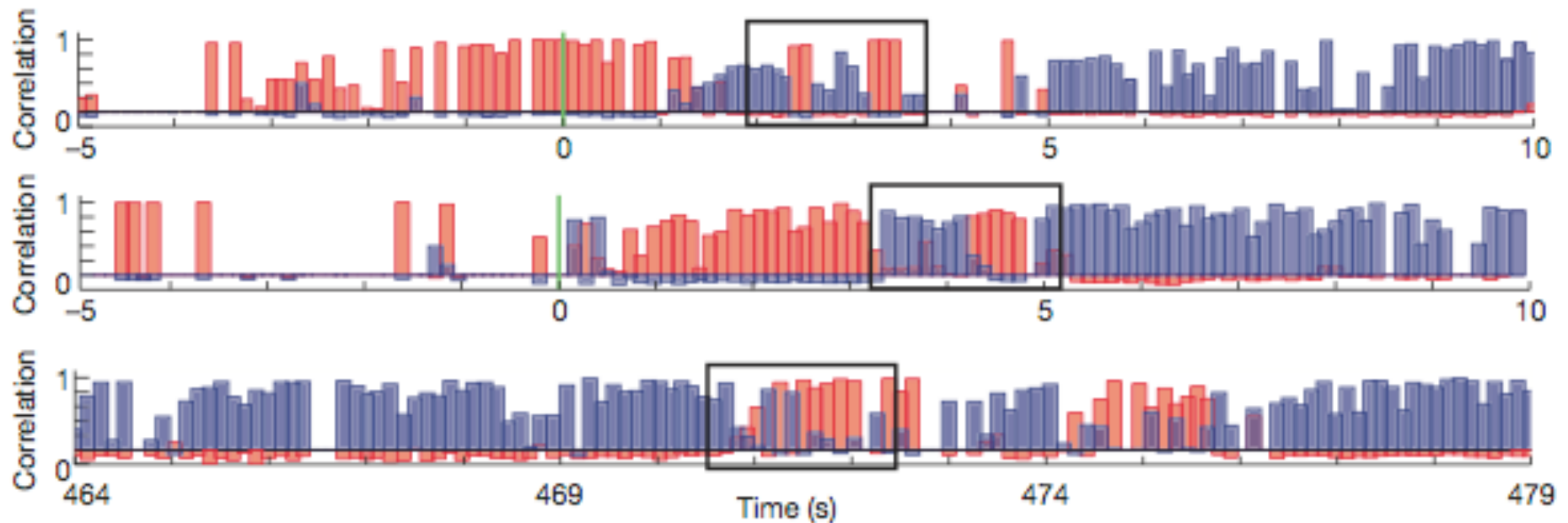
- Place fields are specific to each environment
- Population vectors (average activity) specific to each environment
- Sudden change of environment?



Jezek, Henriksen, Treves, Moser & Moser,
Nature 478, 246 (2011)

Teleportation (2)

$$m^\mu = \text{overlap}(\vec{r}(t), \vec{r}^\mu) \quad \text{with } \mu = 1, 2$$

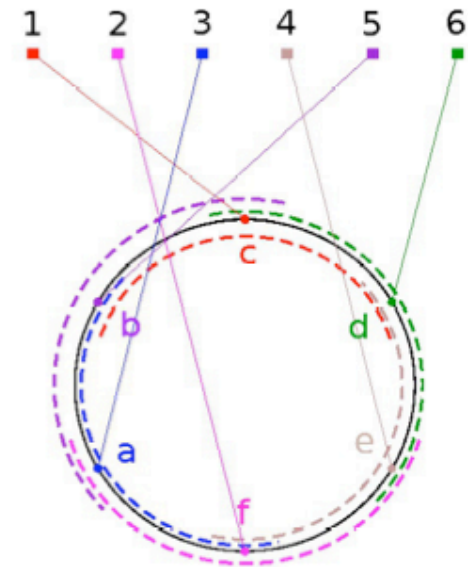
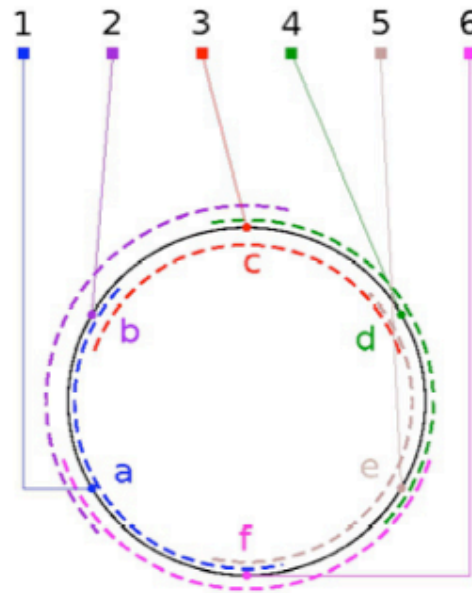


How are different environments 'stored' in the hippocampus?
What is the dynamics of the neural activity within one environment?
In between two environments?

Model: random remappings

Hypothesis: place fields are randomly remapped onto neurons

Example in
dimension $D=1$:



New environment = random permutation π
(no dilution here, easy to incorporate in model ...)

Battaglia, Treves (1998)
Tsodyks (1999)
Hopfield (2010)
R.M., Rosay (2014)

Model: statistical mechanics formulation

Interaction matrix for
L+1 environments:

$$J_{ij} = \sum_{\ell=0}^L J_{ij}^{\ell} = J_{ij}^0 + \sum_{\ell=1}^L J_{\pi^{\ell}(i)\pi^{\ell}(j)}^0$$

Probability of activity
configuration:

$$P_J(\boldsymbol{\sigma}) = \frac{1}{Z_J(T)} \exp \left(- E_J[\boldsymbol{\sigma}]/T \right)$$

‘Energy’ :
(-log likelihood)

$$E_J[\boldsymbol{\sigma}] = - \sum_{i < j} J_{ij} \sigma_i \sigma_j$$

Partition
function:

$$Z_J(T) = \sum_{\boldsymbol{\sigma} \text{ with constraint } \sum_{i=1}^N \sigma_i = f N} \exp \left(- E_J[\boldsymbol{\sigma}]/T \right)$$

← (inhibition)

Multi-environment case: order parameters

Hypothesis: Look for activity localized in one environment, delocalized in the others

Local density of activity
averaged over
environments:

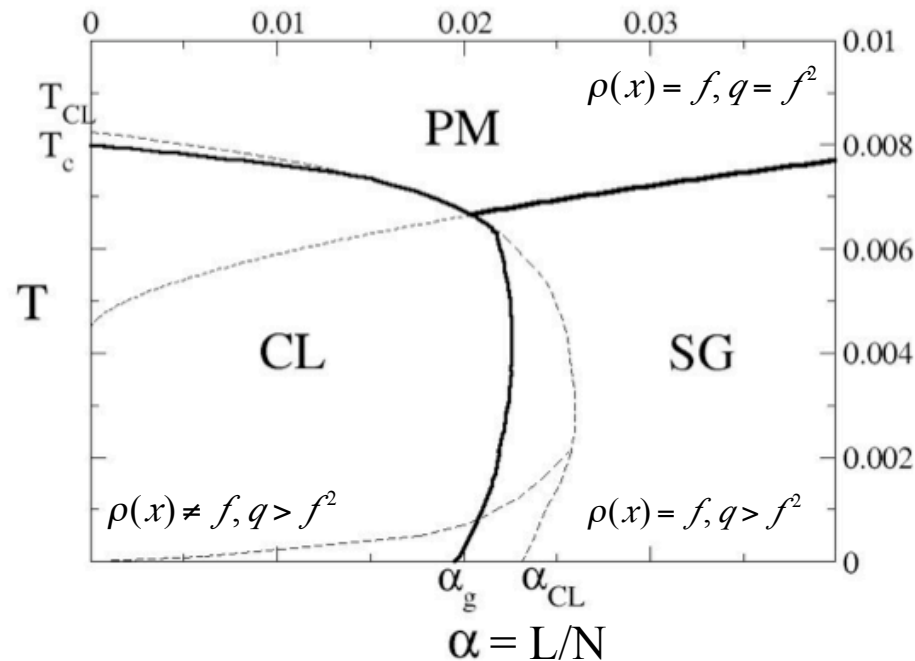
$$\rho(x) \equiv \lim_{\epsilon \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{\epsilon N} \sum_{(x - \frac{\epsilon}{2})N \leq i < (x + \frac{\epsilon}{2})N} \overline{\langle \sigma_i \rangle_J}$$

Edwards-Anderson overlap
(measures spatial heterogeneities
in the activity):

$$q \equiv \frac{1}{N} \sum_{i=1}^N \overline{\langle \sigma_i \rangle_J^2}$$

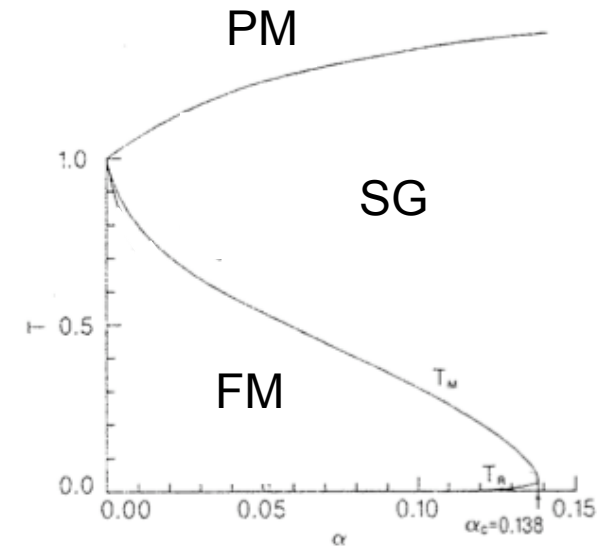
Phase diagram(s)

$$H = - \sum_{i < j} \sum_l J_{\pi^l(i), \pi^l(j)}^0 \sigma_i \sigma_j$$



vs.

$$H = - \sum_{i < j} \left(\sum_{\mu} \xi_i^{\mu} \xi_j^{\mu} \right) \sigma_i \sigma_j$$



Amit, Gutfreund, Sompolinsky (1984)

Storage of an extensive number of spatial charts in an attractor neural network...

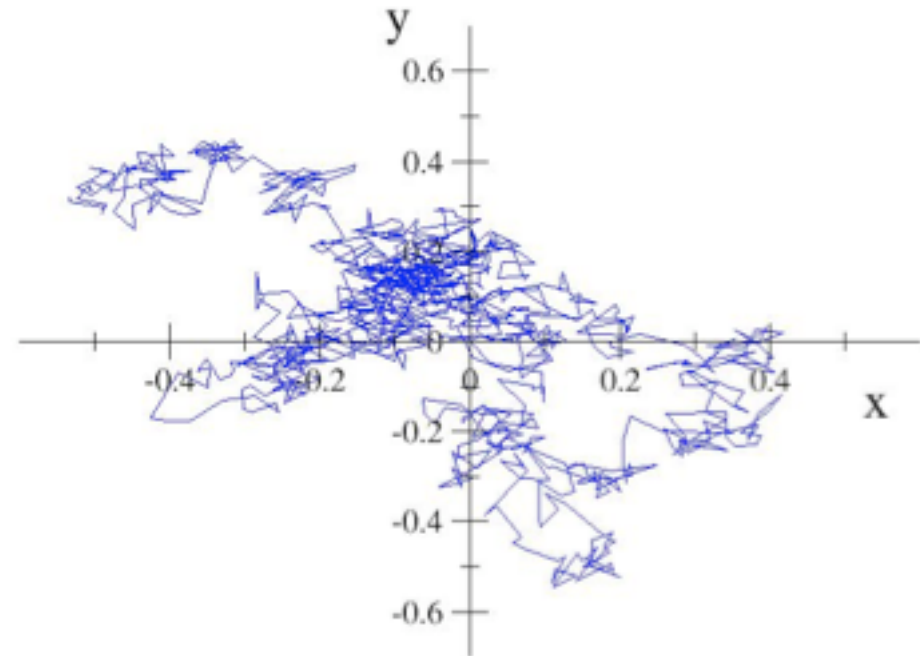
... very robust to neural noise (temperature) !

Dynamics within one environment (LP model)

Microscopic neural dynamics: emergence of a quasi-particle

relaxation towards equilibrium density for all modes, with thermalization at 'temperature' of the order of $1/N$ except for zero mode (translation of bump), which diffuses with $D=O(1/N)$

Einstein relation = ok!



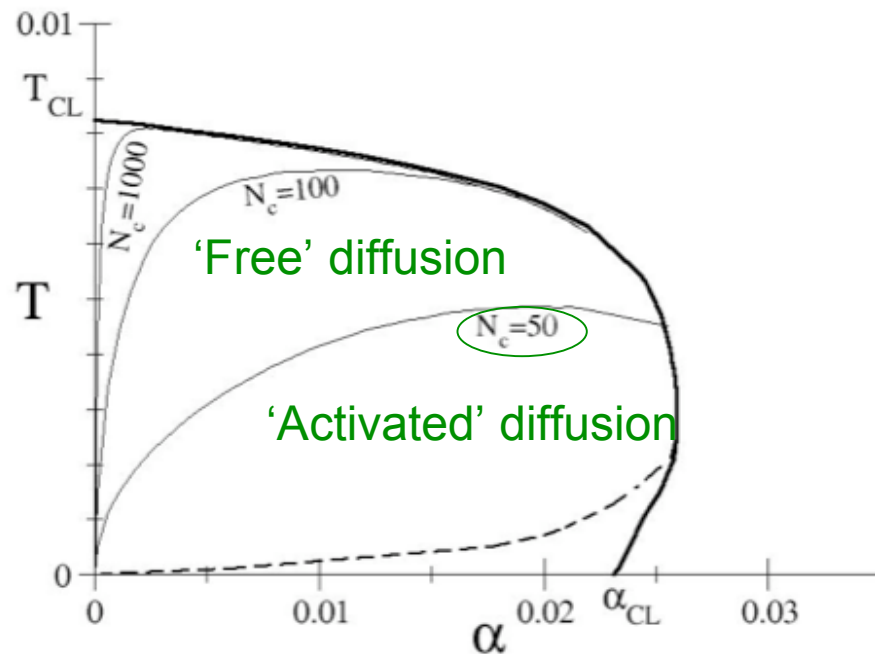
Trajectory of clump center in $D=2$
($N=45 \times 45$ spins, $\alpha=0.001$, $T=0.004$)

Dynamics within one environment ($\alpha > 0$)

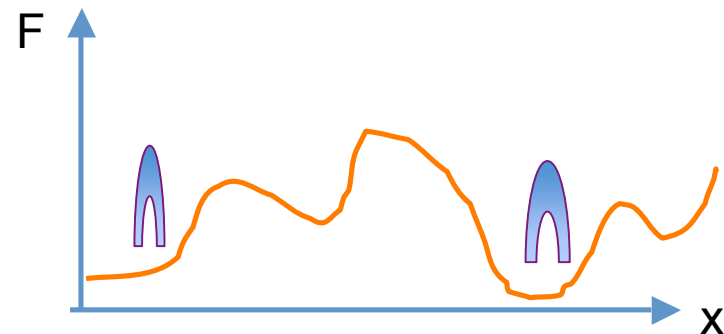
$$\overline{Z_J(T)}^n = \int df \mu_N(f) e^{N(-n\beta f)} = e^{N(-n\beta f_{av} + \frac{n^2}{2}\Gamma + O(n^3)) + o(N)}$$

Fluctuations

$$(\overline{\Delta F^2})^{1/2} = T\sqrt{\Gamma}\sqrt{N}$$

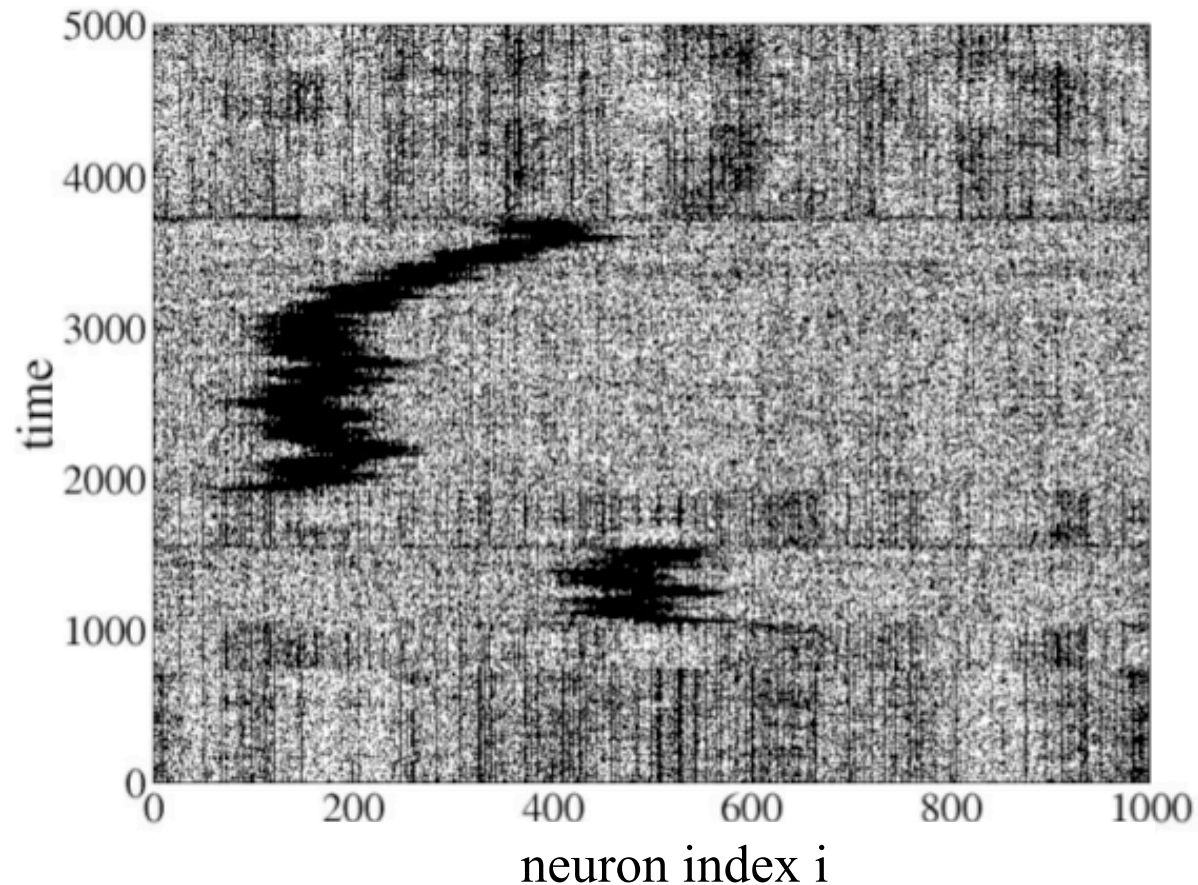


We can go further ...



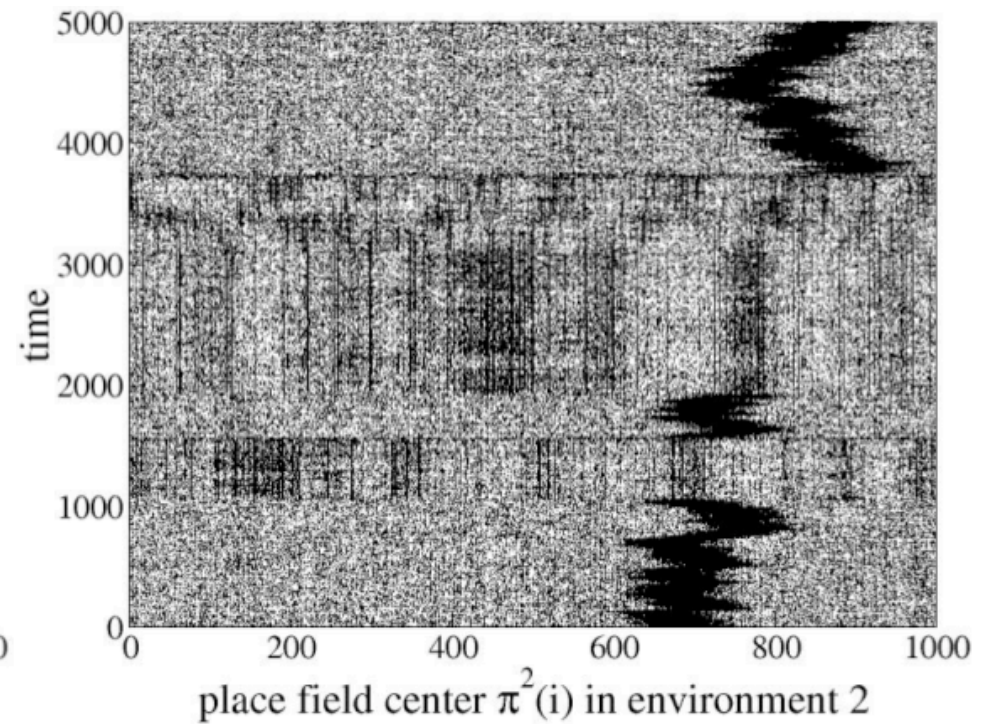
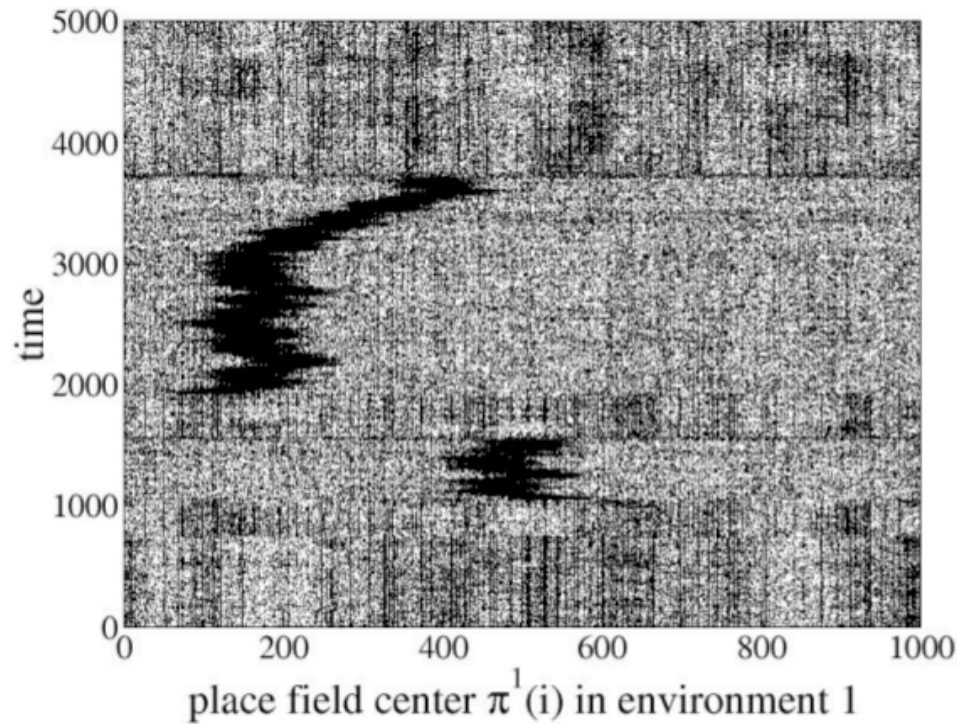
Compute $\overline{(\Delta F(x)\Delta F(y))^{1/2}} = T\sqrt{\Gamma(x-y)}\sqrt{N}$
 Technically: two sets of $n/2$ replicas, $n \rightarrow 0$

Dynamics: transitions between environments

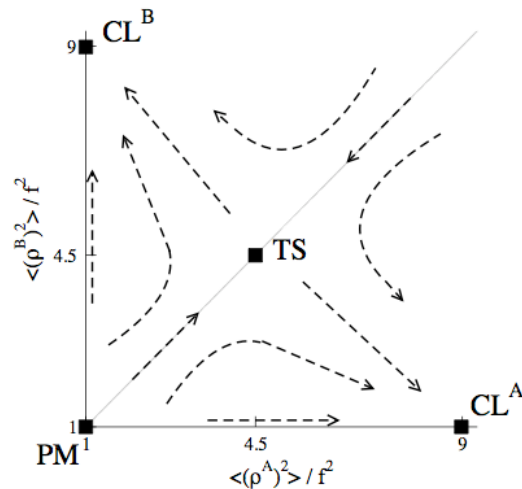


- simulations with two environments (close to top boundary of localized phase)
- ‘neurons’ ordered according to their place fields in env. 1

Dynamics: transitions between environments

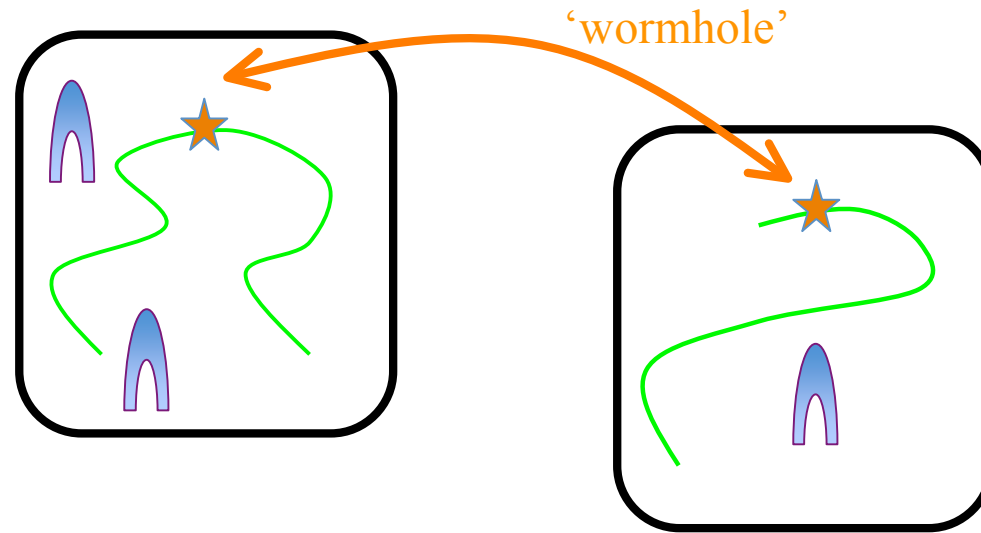


Theory:



- activity partially localized in both environments in TS
- ‘spurious state’ is important!

Schematic picture of the dynamics

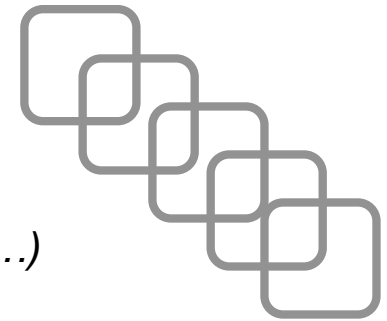


Competition between (activated) diffusion and transitions between charts...

Depends on N (or effective N)

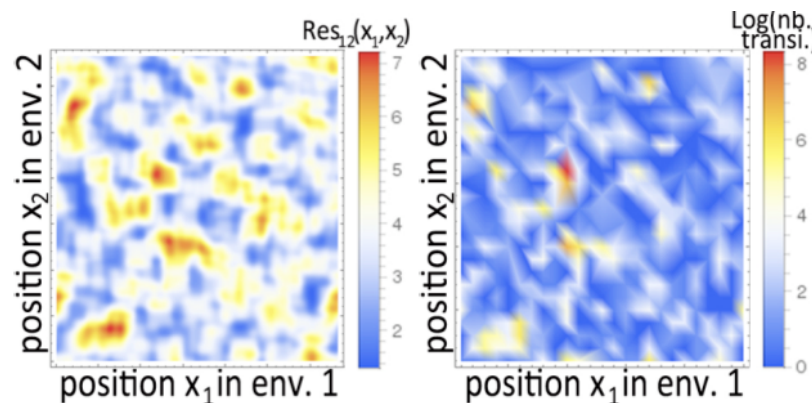
How to enhance motility in disordered landscape?

(modulation of activity, orthogonalization of maps, adaptation, ...)



Q: Where do transitions take place?

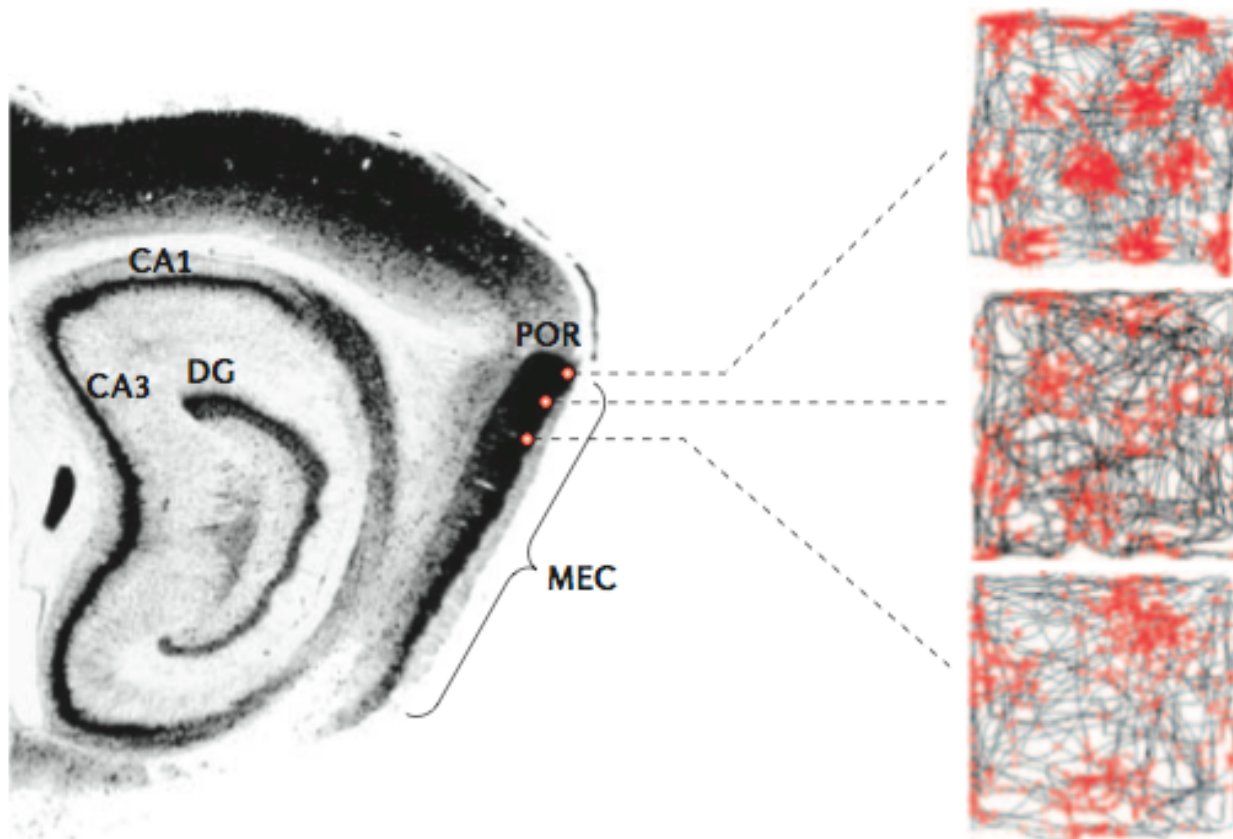
And in experiments?



$N=333,$
 $T=0.006,$
 2 env. in $D=1$

Evidence for continuous attractors in the brain (1)

Hafting, Fyhn, Molden, Moser & Moser, Nature 436, 801-806 (2005)

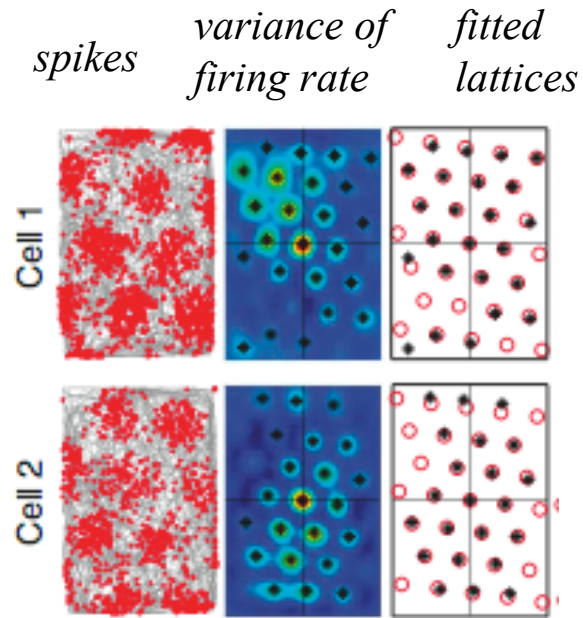


Trajectory of a rat through a square environment is shown in black. Red dots indicate locations at which a particular entorhinal grid cell fired.

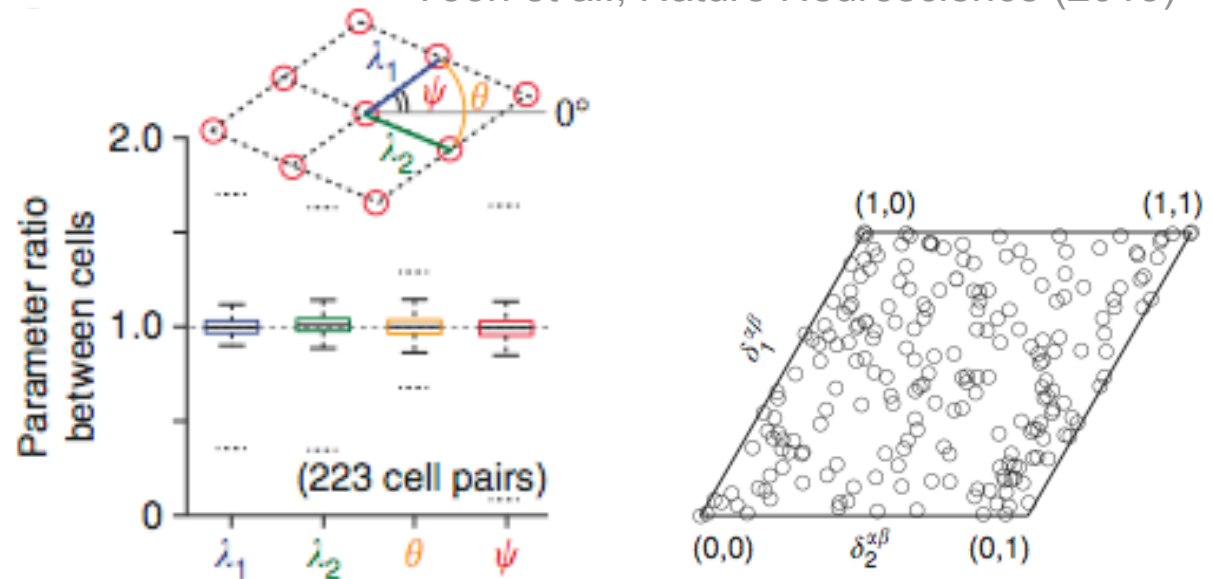
Grid cell properties:

- fire on triangular lattice
- neighbouring cells differ by translation of their grids
- ‘far away’ cells also differ by grid rotation
- mesh sizes vary with recording depth in MEC
- geometric organization of grids (5 sizes, ratio 1.4)
- establish very fast in a new environment and stabilize over days
- found in rodents, monkeys, bats
- 2D continuous attractor models (with local inhibition)

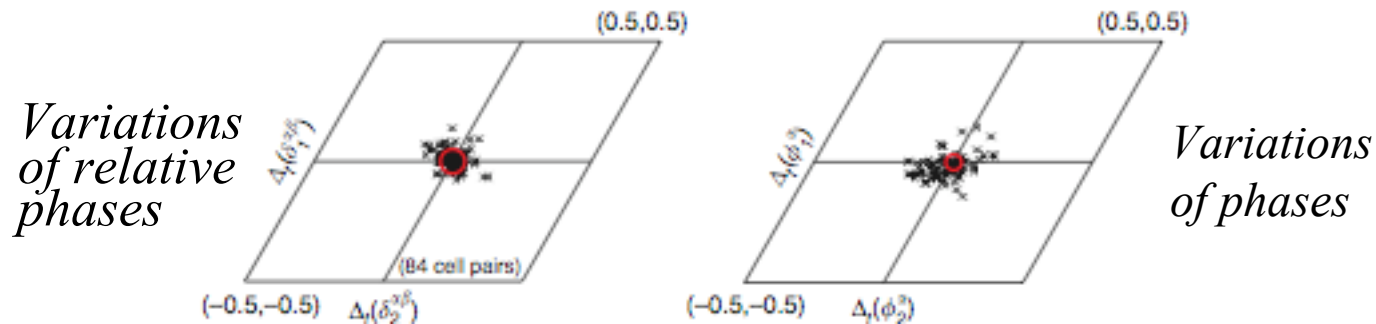
Evidence for continuous attractors in the brain (2)



Yoon et al., Nature Neuroscience (2013)



- neighbouring cells define identical 2D lattices, up to a 2D translation
- relative values of translation parameters are more stable over long periods of time than parameters themselves

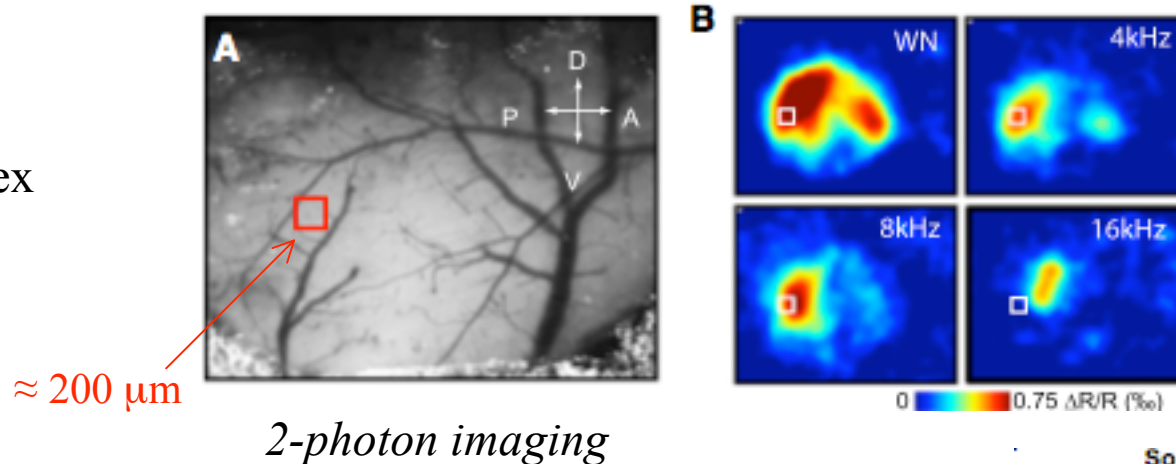


- stability against moderate perturbations e.g. environment reshaping ...

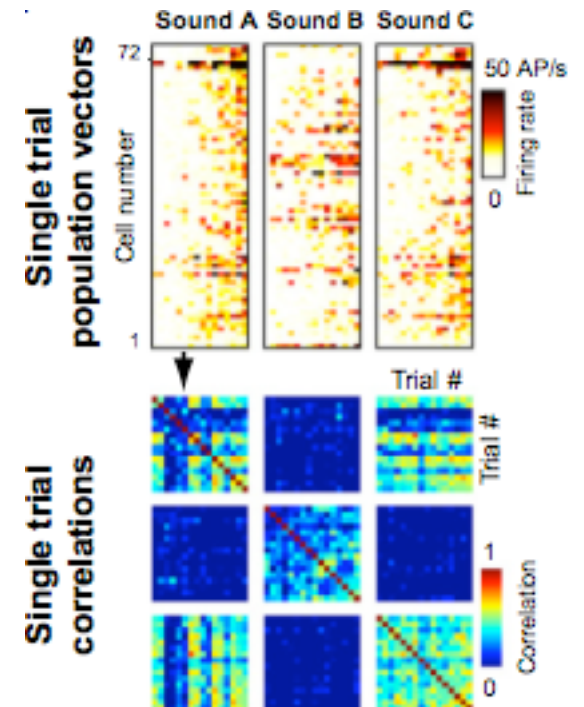
Cell assemblies in the auditory cortex (1)

Bathellier, Ushakova, Rumpel, Neuron 2012

Superficial
auditory cortex
of a mouse

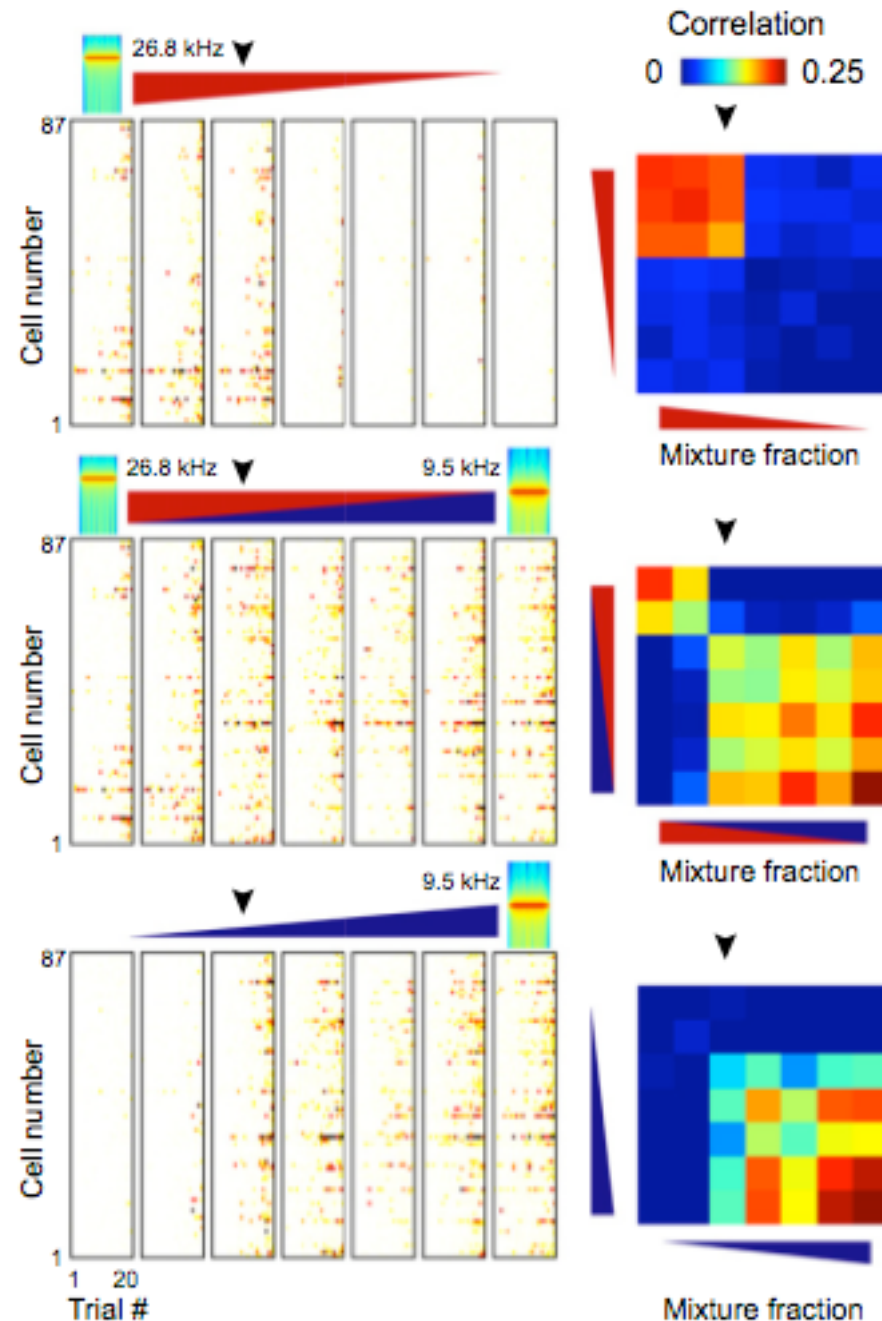


- response depends on stimulus
- neural activity is highly stochastic at the individual level
- less fluctuations at the population level
- what happens for mixed stimuli?



Cell assemblies in the auditory cortex (2)

- compatible with discrete (Hopfield-like) attractor model
- existence of a limited number of « attractors » (74 in 14 mice?)
- combinatorial coding based on those neural sub-populations



Conclusion

Situation has drastically evolved since 1982:

- Various experimental techniques (recording, imaging, reversible & specific stimulation, ...) provide detailed knowledge of the neural activity at the cell level
- Evidence for the existence of discrete and continuous attractors
- Attractors (memories) can be modified
- Dynamics ...
- Theoretical challenges: out-of-equilibrium effects,
nature of representations (what is “place”?),
biological constraints, ...
inverse approaches to interpret/analyze data

Brief & recent bibliography

Space: Grid, place, ... cells: *A Sense of Where You Are*, New York Times,
April 30th, 2013

B.L. McNaughton et al., Nature Reviews Neuroscience 7, 663 (2006)

A. Samsonovitch, B.L. McNaughton, J. Neuroscience 17, 5900 (1997)

Cell assemblies: G. Buzsaki, Neuron 68, 362 (2010)

Continuous-attractor models & transitions in the hippocampus:

F.P. Battaglia and A. Treves, Physical Review E 58, 7738 (1998)

M. Tsodyks, Hippocampus 9, 481 (1999)

J.J. Hopfield, Proceedings of the National Academy of Sciences 107, 1648 (2010)

R.M., S. Rosay, Physical Review E 87, 062813 (2013), E 89, 032803 (2014)

Assemblies in the A.C.: B. Bathellier, L. Ushakova, S. Rumpel, Neuron 76, 435 (2012)

Memory modification:

S. Ramirez et al., Science 341, 387 (2013)

S. Nabavi et al. Nature 511, 348 (2014)

Please, ask for more references ...