Storage of spatial maps in an attractor neural network model of the hippocampus

R. Monasson, LPT-ENS, Paris
Plan

- Biological facts and motivations
- Model and statistical mechanics framework
- Storage of environments (spatial charts)
- Dynamics within one chart
- Transitions between charts
- Conclusion & Perspectives
Representation of Space in the Brain

Electrode recordings: O'Keefe & Dostrovsky (1971)
Cells in the Hippocampus respond to position in space in a specific way (called place cells)
Place cells

Kazu Nakazawa, Thomas J. McHugh, Matthew A. Wilson & Susumu Tonegawa
Where do place fields come from?

Medio-enthorinal cortex ➔ Hippocampus
Grid cells


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

Spatial autocorrelogram of the neuronal activity of the grid cell from the left figure.

- weighted sums of grid cell activities may produce localized activity (place fields)
- changes in weights results in ‘random’ remappings
Teleportation (1)

- Rat in 2 different environments
- Place fields are specific to each environment
- Population vectors (average activity) specific to each environment
- Sudden changes of environment?

How are different environments ‘stored’ in the hippocampus? What is the dynamics of the neural activity within one environment? In between two environments?
Model: one environment (1)

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

\[ J_{ij}^0 = \begin{cases} \frac{1}{N} & \text{if } d_{ij} \leq d_c \\ 0 & \text{if } d_{ij} > d_c \end{cases} \]
Model: one environment (2)

Physical space

Neural network

- we choose $d_c$ so that each neuron is connected to $wN$ other neurons ($w<<1$, but long range interactions)
- Interaction matrix invariant under translations (not necessary)
Model: random remappings

Hypothesis: place fields are randomly remapped onto neurons

Example in dimension $D=1$:

New environment = random permutation $\pi$

Battaglia, Treves (1998); Tsodyks (1999); Hopfield (2010)
Model: statistical mechanics formulation

Interaction matrix for L+1 environments:

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

Probability of activity configuration:

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

‘Energy’: (=-log likelihood)

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

Partition function:

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

(inhibition)
Case of a single environment (1)

Translation-invariant and long-range interactions: exactly solvable model (J.L. Lebowitz and O. Penrose, Journal of Mathematical Physics 7, 98 (1966))

Order parameter = Coarse-grained activity:

\[ \rho(x) \equiv \lim_{\epsilon \to 0} \lim_{N \to \infty} \frac{1}{\epsilon N} \sum_{(x-\frac{\epsilon}{2})N \leq i < (x+\frac{\epsilon}{2})N} \langle \sigma_i \rangle J \]

Single neuron 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)
Case of a single environment (2)

$T=0 \quad \Rightarrow \quad Paramagnetic\ phase$  

Localized phase \hspace{1cm} $T_{\text{spinodal}}$

True also in $D=2$:
Multi-environment case: averaging over remappings

Free energy depends on realization of random interactions $J$

$$= \text{permutations } \pi_1, \pi_2, \ldots, \pi_L$$

Hypothesis: concentration when $N \rightarrow \infty$

Replica method:

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

Average free energy

Fluctuations
Multi-environment case: averaging over remappings

\[ Z_j(T)^n = \sum_{\vec{\sigma}} \exp \left[ \beta \sum_{a=1}^{n} \sum_{i<j} \left( J_{ij}^0 + \sum_{\ell=1}^{L} J_{ij}^\ell \right) \sigma_i^a \sigma_j^a \right] \]

\[ = \sum_{\vec{\sigma}} \exp \left[ \beta \sum_{a=1}^{n} \sum_{i<j} J_{ij}^0 \sigma_i^a \sigma_j^a \right] \Xi(\vec{\sigma})^L, \quad \text{where} \]

\[ \Xi(\vec{\sigma}) = \frac{1}{N!} \sum_{\pi^\ell} \exp \left[ \beta \sum_{i<j} J_{ij}^0 \sum_{a=1}^{n} \sigma_{\pi^\ell(i)}^a \sigma_{\pi^\ell(j)}^a \right] \]

Average over permutations is not immediate ... but can be done when N\(\to\infty\) (some similarity with Itzykson-Zuber integral, but discrete group here)

\[ \log \Xi(\vec{\sigma}) = -\frac{\beta}{2} nf(1-f) + N\frac{\beta}{2} nwf^2 \]

\[ - \sum_{\lambda \neq 0} \text{Trace} \log \left[ \text{Id}_n - \beta \lambda (q - f^2 1_n) \right] \]

\[ q^{ab} \equiv \frac{1}{N} \sum_j \sigma_j^a \sigma_j^b \]
Multi-environment case: order parameters

Local density of activity averaged over environments:

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

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

\[
q \equiv \frac{1}{N} \sum_{i=1}^{N} \frac{\langle \sigma_i \rangle_J^2}{J}
\]
Phases and Transition Lines

\[ \rho(x) = f, q = f^2 \]

\[ \rho(x) \neq f, q > f^2 \]
Monte Carlo Dynamics:
• Pick up one silent (index i) and one active spin (index j)
• Compute variation of energy when spins are swapped
• Accept with Metropolis rule

Observe:
• Static properties ...
• Stability and fluctuations of clump (=quasi-particle?)
• Motion of clump within one environment
• Transitions between environments
Check of equilibrium properties

Local field acting on spin $i$:

$$h_i = \sum_j J_{ij}^0 \sigma_j + \sum_{l,j} J_{ij}^l \sigma_j$$

Gaussian random variable of variance $\alpha r$, ($r$ = conjugated parameter to $q$)

$(D=1, N=10000, T=0.004, \alpha=0.01)$
The clump is a quasi-particle …

Master equation for spin dynamics (case of a single chart) \(\Rightarrow\)

\[
\frac{\partial \rho}{\partial t}(x,t) = -A[{\rho}] \frac{\delta F[{\rho}]}{\delta \rho(x,t)} + z(x,t)
\]

- \( F \) = free energy functional used to determine equilibrium
- \( z \) = Gaussian random field, uncorrelated in time, correlated in space, of the order of \( N^{-1/2} \)

\(\Rightarrow\) Relaxation towards equilibrium density for all modes, with thermalization at ‘temperature’ of the order of \( N^{-1} \) **except** for zero mode (translation of clump), which diffuses with \( d=O(N^{-1}) \)
...with weak fluctuations across environment

Calculation of

$$\delta \sigma_i^2 = (\langle \sigma_i \rangle - \langle \sigma_i \rangle)^2 = q(x) - \rho^2(x)$$

(D=1, f=0.1, w=0.05, left: T=0.005, right: T=0.007)
Dynamics within one environment (1)

Trajectory of clump center in D=2
(N=45x45 spins, $\alpha=0.001$, $T=0.004$)

Position of clump center in D=1 as a function of time (MC rounds= t x 20)
(N=2000 spins, $\alpha=0.003$, $T=0.006$)
Dynamics within one environment (2)

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

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

We can go further ...

Compute \((\Delta F(x)\Delta F(y))^{1/2} = T \sqrt{\Gamma (x-y)} \sqrt{N}\)

Technically: two sets of \(n/2\) replicas, \(n \to 0\)
Dynamics within one environment (3)

We expect

\[ d \approx \frac{1}{N} e^{-\sqrt{NG(\alpha, T)}} \]

Diffusion coefficient (log. scale)

[\(D=1,\)
\(N=1000,\)
\(T=0.006,\)
\(\alpha=0.003,\)
10000 rounds]
Dynamics: transitions between environments
Dynamics: transitions between environments

• Repeated transitions allow us to determine effective barrier heights and transition times:
  \[ \tau \approx e^{Nb(1\rightarrow2)} \]

• Transitions take place at preferred locations in space(s)
Conclusion & Perspectives

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

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

\[
H = -\sum_{i<j} \sum_{l} J_{i,l,j}^0 \sigma_i \sigma_j
\]

Conclusion & Perspectives

Complete picture of the dynamics within and between charts?

Competition between (activated) diffusion and transitions ...  
Depends on N (or effective N)  
How to enhance diffusion in disordered landscape?  
(modulation of activity, orthogonalization of maps, adaptation, ...)

Conclusion & Perspectives

Complete picture of the dynamics within and between charts?
Conclusion & Perspectives

Comparison/relationship with experiments:

• quantitative estimates of parameters in the model?
• specific locations for transitions?
• differences between $D=1,2,3$?
• dynamical behavior of clump under external forcing (visual clues)?
• extension to other spaces, context-dependent place fields?
To go further

Some references:


and Synopsis in Physics, Knowing your Place, by D. Voss (http://physics.aps.org/)