Biological Modeling of Neural Networks

Week 6

Attractor Networks and

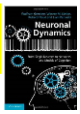
Generalizations of the Hopfield model

Wulfram Gerstner

EPFL, Lausanne, Switzerland

Reading for week 6:
NEURONAL DYNAMICS
- Ch. 17.2.5 – 17.4

Cambridge Univ. Press



6.1. Attractor networks

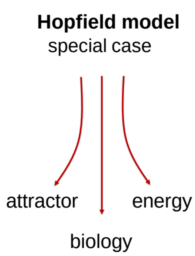
6.2. Stochastic Hopfield model

6.3. Energy landscape

6.4. Towards biology (1)
- low-activity patterns

6.5 Towards biology (2)
- spiking neurons

6.1. Review and next steps



6.1. Attractor networks

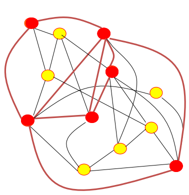
6.2. Stochastic Hopfield model

6.3. Energy landscape

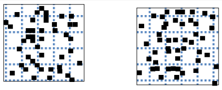
6.4. Towards biology (1)
- low-activity patterns

6.5 Towards biology (2)
- spiking neurons

6.1. Review of week 5: Memory and Hebbian assemblies



6.1. Review of week 5: Deterministic Hopfield model



Prototype $\vec{p}^1$ Prototype $\vec{p}^2$

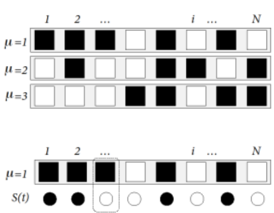
interactions $w_{ij} = \frac{1}{N} \sum_{\mu} p_i^{\mu} p_j^{\mu}$

Sum over all prototypes

- each prototype has black pixels with probability 0.5
- prototypes are random patterns, chosen once at the beginning

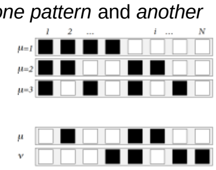
6.1. Review of week 5: overlap / correlation

Image: Neuronal Dynamics, Gerstner et al., Cambridge Univ. Press (2014),



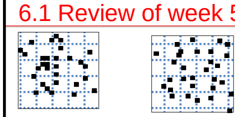
Correlation: overlap between *one pattern* and *another*

Overlap: similarity between state *S(t)* and pattern $m^{\mu} = \frac{1}{N} \sum_j p_j^{\mu} S_j$



Orthogonal patterns

6.1 Review of week 5: Deterministic Hopfield model



Prototype $\vec{p}^1$ Prototype $\vec{p}^2$

interactions $w_{ij} = \frac{1}{N} \sum_{\mu} p_i^{\mu} p_j^{\mu}$

Sum over all prototypes

Input potential $h_i = \sum_j w_{ij} S_j$

Sum over all inputs to neuron *i*
prototypes

Deterministic dynamics

dynamics $S_i(t+1) = \mathrm{sgn}[h_i(t)] = \mathrm{sgn}[\sum_j w_{ij} S_j(t)]$

Similarity measure: Overlap w. pattern 17: $m^{17}(t) = \frac{1}{N} \sum_j p_j^{17} S_j(t)$

6.1 Hopfield model: memory retrieval (with overlaps)

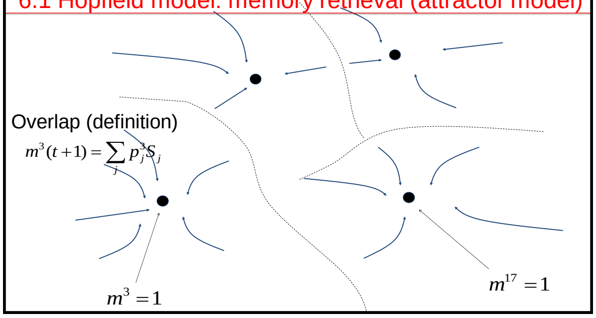
$$S_i(t+1) = \mathrm{sgn}[h_i(t)] = \mathrm{sgn}[\sum_j w_{ij} S_j(t)]$$

Blackboard-1

$$S_i(t+1) = \mathrm{sgn}[\sum_\mu p_i^\mu m^\mu(t)]$$

$$m^\mu(t+1) \leftarrow \quad m^\mu(t)$$

6.1 Hopfield model: memory retrieval (attractor model)



Overlap (definition)

$$m^3(t+1) = \sum_j p_j^3 S_j$$

6.1 Hopfield model: memory retrieval (attractor model)

Hopfield model
special case

attractor   energy

biology

Attractor networks:
dynamics moves network state to a fixed point

Hopfield model:
for a small number of patterns, states with overlap 1 are fixed points

Aim for today:
generalize!

Quiz 6.1: overlap and attractor dynamics

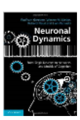
[ ] The overlap is maximal
if the network state matches one of the patterns.
[ ] The overlap increases during memory retrieval.
[ ] The mutual overlap of orthogonal patterns is one.
[ ] In an attractor memory, the dynamics converges to a stable fixed point.
[ ] In a perfect attractor memory network, the network state moves towards one of the patterns.
[ ] In a Hopfield model with N random patterns stored in a network N neurons, the patterns are attractors.
[ ] In a Hopfield model with 200 random patterns stored in a network 1000 neurons, all fixed points have overlap one.

Biological Modeling of Neural Networks

Week 6
Attractor Networks and Generalizations of the Hopfield model

Wulfram Gerstner
EPFL, Lausanne, Switzerland

Reading for week 6:
NEURONAL DYNAMICS
- Ch. 17.2.5 – 17.4



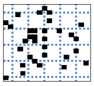
Cambridge Univ. Press

6.2 Stochastic Hopfield model

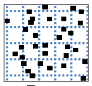
Neurons may be noisy:

What does this mean for attractor dynamics?

6.2 Stochastic Hopfield model



Prototype $\vec{p}^1$



Prototype $\vec{p}^2$

Random patterns

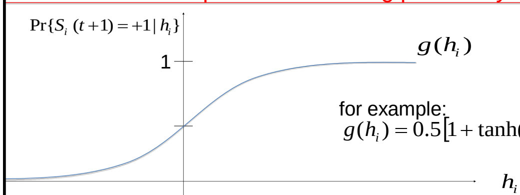
Interactions (1)

$$w_{ij} = \frac{1}{N} \sum_{\mu} p_i^{\mu} p_j^{\mu}$$

Dynamics (2)

$$\Pr\{S_i\,(t+1) = +1 \,|\, h_i\} = g[h_i] = g\left[\sum\nolimits_j w_{ij} S_j\,(t)\right]$$

6.2 Stochastic Hopfield model: firing probability



for example:
$g(h_i) = 0.5[1 + \tanh(2h)]$

Blackboard-2

$$\Pr\{S_i\,(t+1) = +1 \,|\, h_i\} = g[h_i] = g\left[\sum\nolimits_j w_{ij} S_j\,(t)\right] = g\left[\sum\nolimits_{\mu} p_i^{\mu} m^{\mu}\,(t)\right]$$

6.2 Stochastic Hopfield model

Dynamics (2)

$$\Pr\{S_i\,(t+1) = +1 \,|\, h_i\} = g[h_i] = g\left[\sum\nolimits_j w_{ij} S_j\,(t)\right]$$

$$\Pr\{S_i\,(t+1) = +1 \,|\, h_i\} = g\left[\sum\nolimits_{\mu} p_i^{\mu} m^{\mu}\,(t)\right]$$

Assume that there is only overlap with pattern 17:
two groups of neurons: those that should be 'on' and 'off'

$$\Pr\{S_i\,(t+1) = +1 \,|\, h_i = h^+\} = g\left[m^{17}\,(t)\right] \quad \text{for } i \text{ with } p_i^{17} = +1$$

$$\Pr\{S_i\,(t+1) = +1 \,|\, h_i = h^-\} = g\left[-m^{17}\,(t)\right] \quad \text{for } i \text{ with } p_i^{17} = -1$$

Overlap (next time step) $m^{17}(t+1) = \frac{1}{N} \sum_j p_j^{17} S_j(t+1) = ???$

Exercise 1 now: Stochastic Hopfield

9 minutes,
Try to get
As far as possible

Overlap $m^{17}(t+1) = \frac{1}{N}\sum_j p_j^{17} S_j(t+1)$

Suppose initial overlap with pattern 17 is 0.4;
Find equation for overlap at time *(t+1)*,
given overlap at time *(t)*.
Assume overlap with other patterns stays zero.

Next lecture
9:55

Hint: Use result from blackboard and consider 4 groups of neurons
- Those that should be ON and are ON
- Those that should be ON and are OFF
- Those that should be OFF and are ON
- Those that should be OFF and are OFF

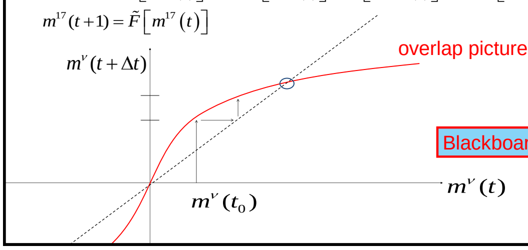
6.2 Stochastic Hopfield model: memory retrieval

Overlap:

Neurons that should be 'on' — Neurons that should be 'off'

$$2m^{17}(t+1) = g\left[m^{17}(t)\right] - \{1 - g\left[m^{17}(t)\right]\} - g\left[-m^{17}(t)\right] + \{1 - g\left[-m^{17}(t)\right]\}$$

$$m^{17}(t+1) = \tilde{F}\left[m^{17}(t)\right]$$



Blackboard-3

6.2 Stochastic Hopfield model: memory retrieval

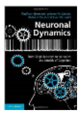
- Memory retrieval possible with stochastic dynamics
- Fixed point at value with large overlap (e.g., 0.95)
- Need to check that overlap of other patterns remains small
- Random patterns: nearly orthogonal but 'noise' term

Biological Modeling of Neural Networks

Week 6
Attractor Networks and
Generalizations of the Hopfield model

Wulfram Gerstner
EPFL, Lausanne, Switzerland

Reading for week 6:
NEURONAL DYNAMICS
- Ch. 17.2.5 – 17.4



Cambridge Univ. Press

6.2 Stochastic Hopfield model = attractor model

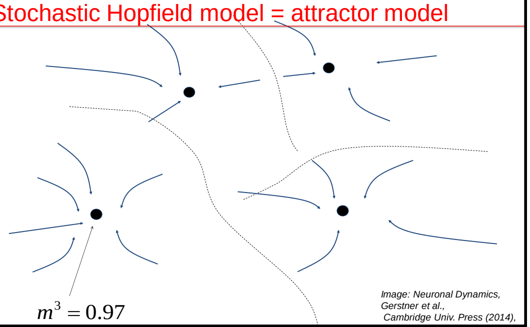
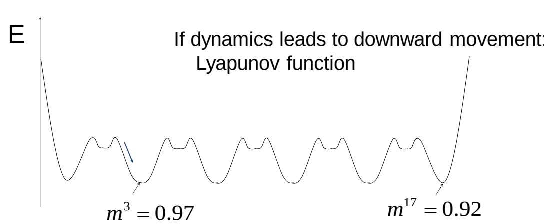


Image: Neuronal Dynamics, Gerstner et al., Cambridge Univ. Press (2014),

6.3 Symmetric interactions: Energy picture

If dynamics leads to downward movement: Lyapunov function

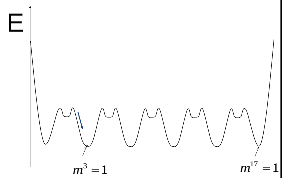


6.3 Symmetric interactions: Energy picture

$$E = -\frac{1}{2}\sum_{i,j} w_{ij} S_i S_j$$

- Rewrite in terms of overlaps
- Random patterns vs. orthogonal patterns
- Random state vs. overlap state

Blackboard-4



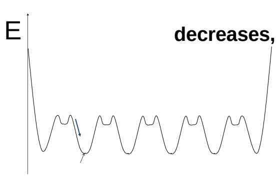
6.3. Symmetric interactions: Energy/Lyapunov function

Assume symmetric interaction,
Assume deterministic asynchronous update

$$S_i(t+1) = \text{sgn}[h_i(t)] = \text{sgn}[\sum_j w_{ij} S_j(t)]$$

Claim: the energy $E = -\frac{1}{2}\sum_{i,j} w_{ij} S_i S_j$

decreases, if neuron *k* changes



J.J. Hopfield (1982) Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA 79, pp. 2554–2558

Exercise 2 now: energy $E = -\frac{1}{2}\sum_{i,j} w_{ij} S_i S_j$

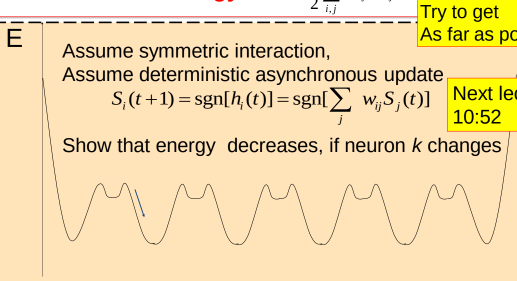
9 minutes,
Try to get
As far as possible

Assume symmetric interaction,
Assume deterministic asynchronous update

$$S_i(t+1) = \text{sgn}[h_i(t)] = \text{sgn}[\sum_j w_{ij} S_j(t)]$$

Show that energy decreases, if neuron *k* changes

Next lecture
10:52



6.3 Symmetric interactions: Energy picture

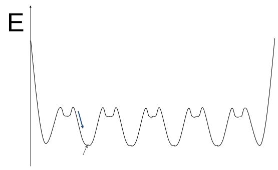
$$E = -\frac{1}{2}\sum_{i,j} w_{ij} S_i S_j$$

Assume symmetric interaction,
Assume deterministic asynchronous update

$$S_i(t+1) = \mathrm{sgn}[h_i(t)] = \mathrm{sgn}[\sum_j w_{ij} S_j(t)]$$

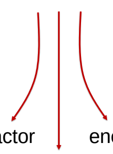
Claim: energy decreases, if neuron *k* changes

Proof: blackboard-5



6.3. Energy picture

Hopfield model
special case



energy picture is rather general
- has been very influential

energy picture is a side-track:
- it needs symmetric interactions

energy picture is useful
- it shows that it should be possible to learn other patterns than mean-zero random patterns

Quiz 6.2: Energy picture and Lyapunov function

Let $E = -\frac{1}{2}\sum_{i,j} w_{ij} S_i S_j$ be the energy of the Hopfield model

and $S_i(t+1) = \mathrm{sgn}[h_i(t)] = \mathrm{sgn}[\sum_j w_{ij} S_j(t)]$ the dynamics.

[ ] The energy picture requires random patterns with prob = 0.5
[ ] The energy picture requires symmetric weights
[ ] It follows from the energy picture of the Hopfield model that the only fixed points are those where the overlap is exactly one
[ ] In each step, the value of a Lyapunov function decreases or stays constant
[ ] Under deterministic dynamics the above energy is a Lyapunov function

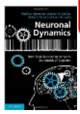
Biological Modeling of Neural Networks

Week 6
Attractor Networks and
Generalizations of the Hopfield model

Wulfram Gerstner
EPFL, Lausanne, Switzerland

Reading for week 6:
NEURONAL DYNAMICS
- Ch. 17.2.5 – 17.4

Cambridge Univ. Press



6.1. Attractor networks

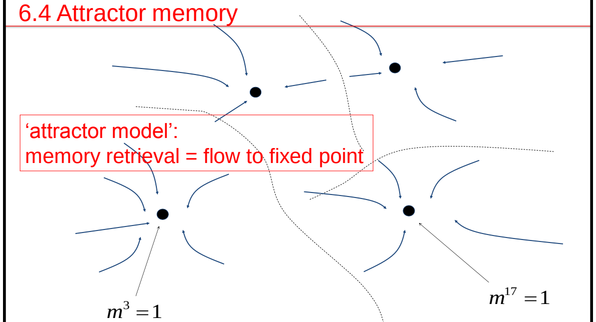
6.2. Stochastic Hopfield model

6.3. Energy landscape

6.4. Towards biology (1)
- low-activity patterns

6.5 Towards biology (2)
- spiking neurons

6.4 Attractor memory

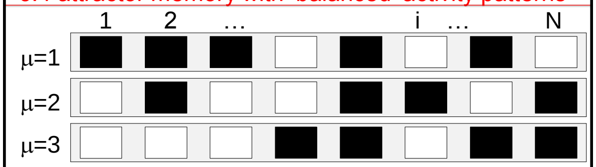


6.4 attractor memory in realistic networks

Memory in realistic networks

-Mean activity of patterns?
-Asymmetric connections?
-Separation of excitation and inhibition?
-Better neuron model?
-Modeling with integrate-and-fire model?
-Low probability of connections?
-Neural data?

6.4 attractor memory with 'balanced' activity patterns



Random patterns +/-1 with zero mean →
50 percent of neurons should be active in each pattern

$$w_{ij} = \frac{1}{N}\sum_{\mu} p_i^{\mu} p_j^{\mu}$$

6.4 attractor memory with 'low' activity patterns

Random patterns +/-1 with **low activity (prob{black}=a<0.5)** →
e.g., 10 percent of neurons should be active in each pattern

$$w_{ij} = c\sum_{\mu} (\xi_i^{\mu} - b)(\xi_j^{\mu} - a)$$

(so far: b=a=0.5)

$\xi_i^{\mu} \in \{0,1\}$

Some constant: *b=0* or *b=a*

Mean activity of pattern

6.4 attractor memory with 'low' activity patterns

Random patterns +/-1 with **low activity (mean a<0.5)** →
e.g.10 percent of neurons should be active in each pattern

$$w_{ij} = c\sum_{\mu} (\xi_i^{\mu} - b)(\xi_j^{\mu} - a) \qquad \xi_i^{\mu} \in \{0,1\}$$

Blackboard-6

Introduce overlap $m^{\mu}(t) = c\sum_{j} (\xi_j^{\mu} - a) S_j(t)$

Introduce dynamics

b=0 or *b=1*

6.4 attractor memory with 'low' activity patterns

- attractor dynamics possible:

$$m^{\mu}(t+1) = \hat{F}[m^{\mu}(t)]$$

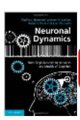
- no need for symmetric weights
- capacity calculations possible (analogous to last week)

Biological Modeling of Neural Networks

Week 6
Attractor Networks and Generalizations of the Hopfield model

Wulfram Gerstner
EPFL, Lausanne, Switzerland

Reading for week 6:
NEURONAL DYNAMICS
- Ch. 17.2.5 – 17.4



Cambridge Univ. Press

6.1. Attractor networks

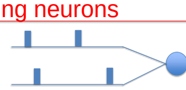
6.2. Stochastic Hopfield model

6.3. Energy landscape

6.4. Towards biology (1)
- low-activity patterns

6.5 Towards biology (2)
- spiking neurons

6.5 attractor memory with spiking neurons



Total input to neuron i

$$h_i(t) = \sum_j w_{ij} S_j(t)$$

- rewrite binary state variable:

$$S_i(t) = \pm 1 \quad \rightarrow \quad \sigma_i(t) \in \{0,1\}$$

- use low firing probability (in time)
- Use low activity (across neurons)

Exercise 3 NOW- from Hopfield to spikes

In the Hopfield model, neurons are characterized by a binary variable S_i = +/-1. For an interpretation in terms of spikes it is, however, more appealing to work with a binary variable x_i which is zero or 1.

(i) Write $S_i = 2\sigma_i - 1$ and rewrite the Hopfield model in terms of σ_i

(ii) What are the conditions so that the input potential is

$$h_i(t) = \sum_j w_{ij}\sigma_j(t)$$

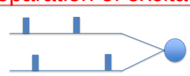
(iii) Interpretation: can you also restrict the weights to excitation only? Assume low-activity patterns

$$w_{ij} = c\sum_\mu (\xi_i^\mu - b)(\xi_j^\mu - a)$$

and pick *b=0*

10 minutes,
Try to get
As far as possible
Lecture: 10:35

6.5. Separation of excitation and inhibition



Blackboard-7

Total input to neuron *i*

$$h_i(t) = \sum_j w_{ij} S_j(t)$$

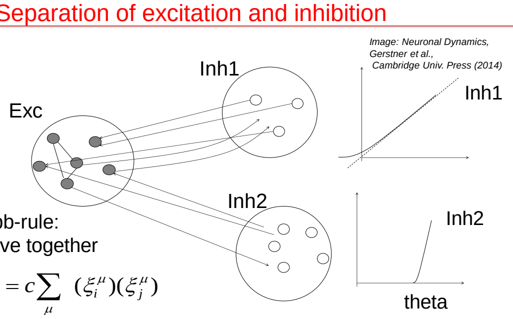
Separation of excitation/ inhibition
- rewrite weights:

$$w_{ij} = c\sum_\mu (\xi_i^\mu - b)(\xi_j^\mu - a)$$

$$\xi_i^\mu \in \{0,1\}$$

$$b = 0$$

6.5 Separation of excitation and inhibition

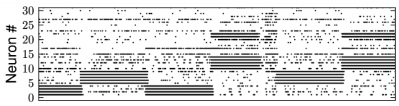


Hebb-rule:
Active together

$$w_{ij} = c\sum_\mu (\xi_i^\mu)(\xi_j^\mu)$$

6.5 attractor memory with 8000 spiking neurons

Spike raster



Overlap with patterns 1 … 6 (total 90 patterns stored, $a=0.1$)

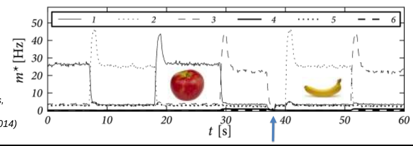


Image: Neuronal Dynamics, Gerstner et al., Cambridge Univ. Press (2014)

6.5 attractor memory with spiking neurons

Memory with spiking neurons

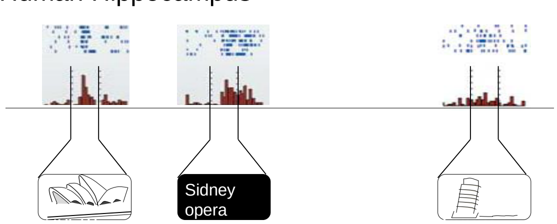
-Low activity of patterns?
-Separation of excitation and inhibition?
-Asymmetric connections
-Modeling with integrate-and-fire?
-Low connection probability

All possible

-Neural data?

6.4 memory data (review from week 5)

Human Hippocampus

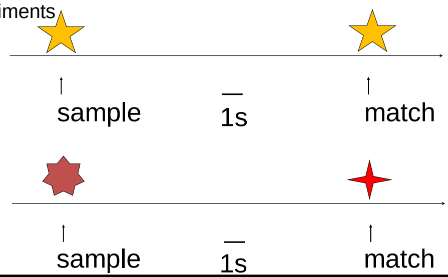


Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature, 435:1102-1107.

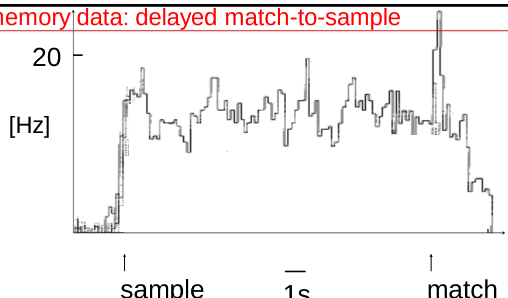
6.4 memory data: delayed match to sample

Delayed Matching to Sample Task

Animal experiments

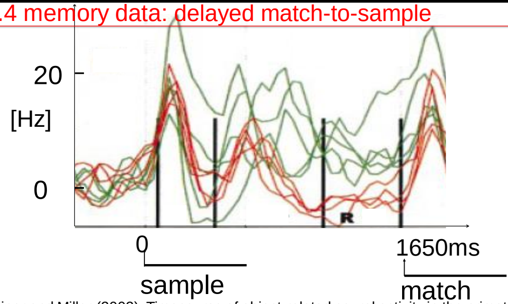


6.4 memory data: delayed match-to-sample



Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. Nature, 335:817-820.

6.4 memory data: delayed match-to-sample



Rainer and Miller (2002). Timecourse of object-related neural activity in the primate prefrontal cortex during a short-term memory task. Europ. J. Neurosci., 15:1244-1254.

6.5 Attractor Networks and Generalized Hopfield Model

Memory with spiking neurons

- Low activity of patterns!
- Separation of excitation and inhibition!
- Modeling with integrate-and-fire!
- Asymmetric connections!
- Sparse connectivity!

All possible

Attractor Memory Model

- abstract concept
- Concept stable under generalizations
- Neural data?

References: Attractor Memory Networks

Abbott, Amit, Brunel, Fusi, Gerstner, Herz, Hertz, Sompolinsky,Tsodys, Treves, van Vreeswijk, van Hemmen and many others!

Recommended textbook:

J. Hertz, A. Krogh and R. G. Palmer (1991) *Introduction to the Theory of Neural Computation.* Addison-Wesley

- L. F. Abbott and C. van Vreeswijk (1993) Asynchronous states in a network of pulse-coupled oscillators. Phys. Rev. E 48, pp. 1483–1490.

•D. J. Amit, H. Gutfreund and H. Sompolinsky (1985) Storing infinite number of patterns in a spin-glass model of neural networks. Phys. Rev. Lett. 55, pp. 1530–1533.

•D. J. Amit, H. Gutfreund and H. Sompolinsky (1987) Information storage in neural networks with low levels of activity. Phys. Rev. A 35, pp. 2293–2303..

•D. J. Amit and N. Brunel (1997) A model of spontaneous activity and local delay activity during delay periods in the cerebral cortex. Cerebral Cortex 7, pp. 237–252

-D. J. Amit and M. V. Tsodyks (1991) Quantitative study of attractor neural networks retrieving at low spike rates. i: substrate — spikes, rates, and neuronal gain.. Network 2, pp. 259–273.

-A.V. M. Herz, B. Sulzer, R. Kühn and J. L. van Hemmen (1988) The Hebb rule: representation of static and dynamic objects in neural nets.. Europhys. Lett. 7, pp. 663–669

- A. Treves (1993) Mean-field analysis of neuronal spike dynamics. Network 4, pp. 259–284.

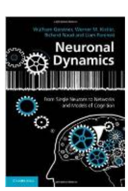
-M. Tsodyks and M.V. Feigelman (1986) The enhanced storage capacity in neural networks with low activity level. Europhys. Lett. 6, pp. 101–105.

The end

Documentation:
http://neuronaldynamics.epfl.ch/

Online html version available

Reading for this week: NEURONAL DYNAMICS - Ch. 17.2.5 - 17.4



Cambridge Univ. Press