Biological Modeling of Neural Networks

Week 6

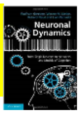
Attractor Networks and Generalizations of the Hopfield model

Wulfram Gerstner

EPFL, Lausanne, Switzerland

Reading for week 6:
NEURONAL DYNAMICS
- Ch. 17.2.5 – 17.4
- Ch. 19.1-19.2;

Cambridge Univ. Press



6.1. Attractor networks

6.2. Stochastic Hopfield model

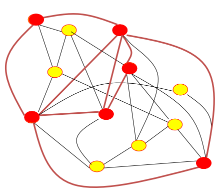
6.3. Energy landscape

6.4. Towards biology
- low-activity patterns
- spiking neurons

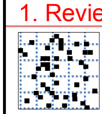
6.4 Models of synaptic plasticity
- Hebbian learning rules
- Bienenstock-Cooper-Munro rule

6.5 Online learning of memories

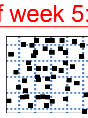
1. Review of week 5: Memory and Hebbian assemblies



1. Review of week 5: Deterministic Hopfield model



Prototype p^1



Prototype p^2

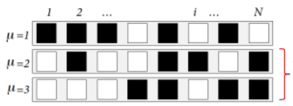
interactions $w_{ij} = \frac{1}{N}\sum_{\mu} p_i^\mu p_j^\mu$

Sum over all prototypes

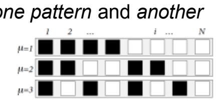
- each prototype has black pixels with probability 0.5
- prototypes are random patterns, chosen once at the beginning

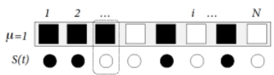
1. Review from week 5: overlap / correlation

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



Correlation: overlap between *one pattern and another*

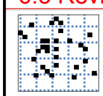




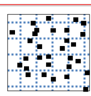
Overlap: similarity between state *S(t)* and pattern $m^\mu = \frac{1}{N}\sum_j p_j^\mu S_j$

Orthogonal patterns

6.3 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: $m^{17}(t+1) = \sum_j p_j^{17} S_j$

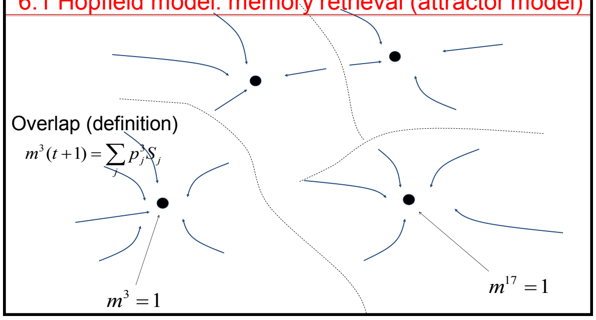
6.1 Hopfield model: memory retrieval (with overlaps)

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

blackboard

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

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)

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

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

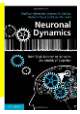
EPFL

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
- Ch. 19.1-19.2;

Cambridge Univ. Press



6.1. Attractor networks

6.2. Stochastic Hopfield model

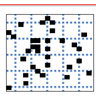
6.3. Energy landscape

6.4. Towards biology
- low-activity patterns
- spiking neurons

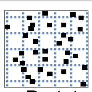
6.4 Models of synaptic plasticity
- Hebbian learning rules
- Bienenstock-Cooper-Munro rule

6.5 Online learning of memories

6.2 Stochastic Hopfield model



Prototype $\vec{p}^1$



Prototype $\vec{p}^2$

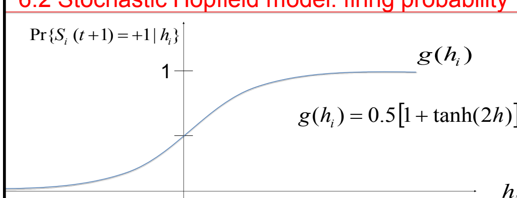
Random patterns

Interactions (1) $w_{ij} = \sum_{\mu} p_i^{\mu} p_j^{\mu}$

Dynamics (2)

$$\Pr\{S_i(t+1) = +1 \mid h_i\} = g[h_i] = g\left[\sum_j w_{ij} S_j(t)\right]$$

6.2 Stochastic Hopfield model: firing probability



$$g(h_i) = 0.5\left[1 + \tanh(2h)\right]$$

blackboard

$$\Pr\{S_i(t+1) = +1 \mid h_i\} = g[h_i] = g\left[\sum_j w_{ij} S_j(t)\right] = g\left[\sum_{\mu} p_i^{\mu} m^{\mu}(t)\right]$$

6.2 Stochastic Hopfield model

Dynamics (2)

$$\Pr\{S_i(t+1)=+1 \mid h_i\} = g[h_i] = g\left[\sum_j w_{ij} S_j(t)\right]$$

$$\Pr\{S_i(t+1)=+1 \mid h_i\} = g\left[\sum_\mu p_i^\mu m^\mu(t)\right]$$

blackboard

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

$$\Pr\{S_i(t+1)=+1 \mid h_i = h^+\} = g\left[m^{17}(t)\right]$$

$$\Pr\{S_i(t+1)=+1 \mid h_i = h^-\} = g\left[-m^{17}(t)\right]$$

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

Exercise 3 now: Stochastic Hopfield

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

15 minutes,
Try to get
As far as possible

Suppose initial overlap with pattern 17 is 0.4;
Find equation for overlap at time *(t+1)*,
given overlap at time *(t)*

Next lecture
9:50

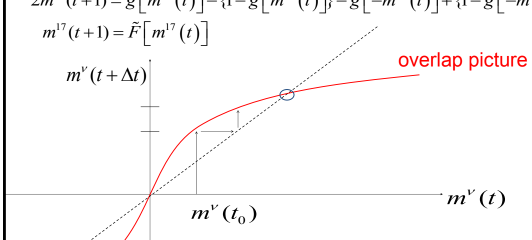
Hint: Use result from blackboard and consider 4 groups of neurons
- Those that should be ON and are ON
- Those that should be ON and are OFF
- Those that should be OFF and are ON
- Those that should be OFF and are OFF

6.2 Stochastic Hopfield model: memory retrieval

Overlap: Neurons that should be 'on' Neurons that should be 'off'

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

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



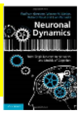
Biological Modeling of Neural Networks

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
- Ch. 19.1-19.2;

Cambridge Univ. Press



6.1. Attractor networks

6.2. Stochastic Hopfield model

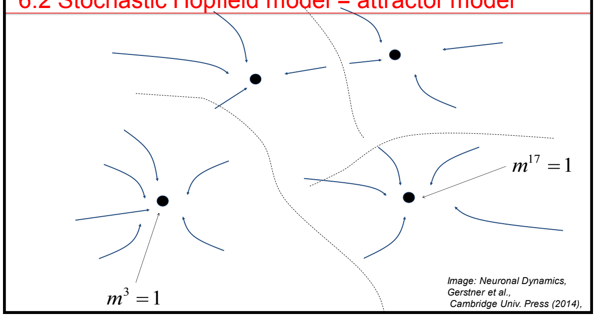
6.3. Energy landscape

6.4. Towards biology
- low-activity patterns
- spiking neurons

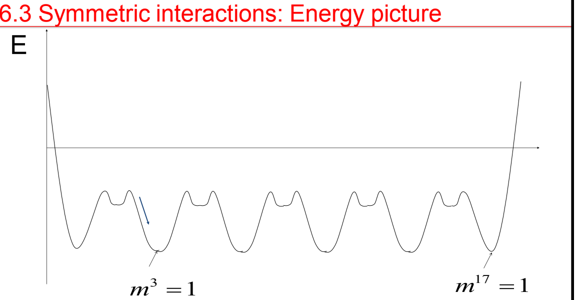
6.4 Models of synaptic plasticity
- Hebbian learning rules
- Bienenstock-Cooper-Munro rule

6.5 Online learning of memories

6.2 Stochastic Hopfield model = attractor model



6.3 Symmetric interactions: Energy picture

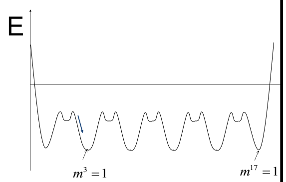


6.3 Symmetric interactions: Energy picture

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

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

blackboard



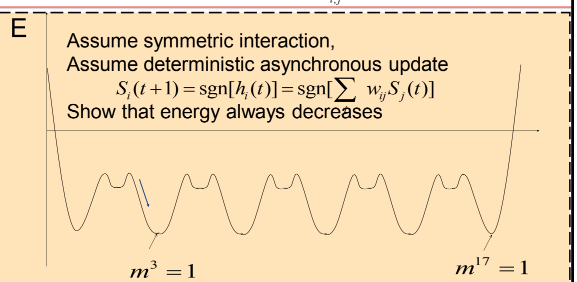
Exercise 4 now: energy

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

Assume symmetric interaction,
Assume deterministic asynchronous update

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

Show that energy always decreases



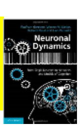
Biological Modeling of Neural Networks

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
- Ch. 19.1-19.2;

Cambridge Univ. Press



6.1. Attractor networks

6.2. Stochastic Hopfield model

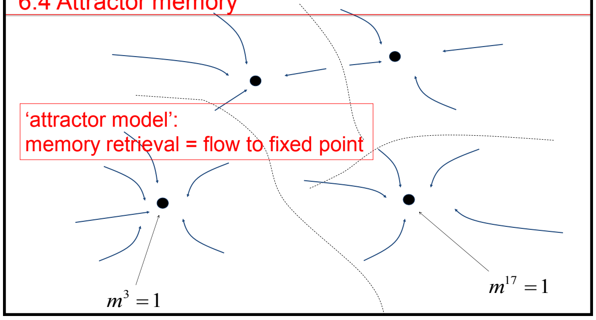
6.3. Energy landscape

6.4. Towards biology
- low-activity patterns
- spiking neurons

6.4 Models of synaptic plasticity
- Hebbian learning rules
- Bienenstock-Cooper-Munro rule

6.5 Online learning of memories

6.4 Attractor memory



6.4 attractor memory with spiking neurons

Memory with spiking neurons

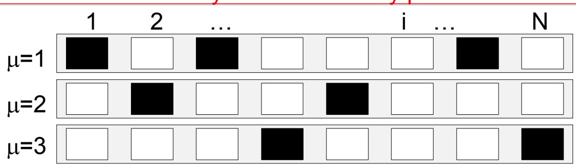
-Mean activity of patterns?
-Better neuron model?
-Separation of excitation and inhibition?
-Modeling with integrate-and-fire model?
-Neural data?

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 (mean =a<0)** →
e.g., 10 percent of neurons should be active in each pattern

$$w_{ij} = \tfrac{1}{N}\sum_{\mu}(p_i^{\mu} - b)(p_j^{\mu} - a)$$

Some constant   Mean activity of pattern

6.4 attractor memory with 'low' activity patterns

Random patterns +/-1 with **low activity (mean =a<0)** →
e.g.10 percent of neurons should be active in each pattern

$$w_{ij} = \tfrac{1}{N}\sum_{\mu}(p_i^{\mu} - b)(p_j^{\mu} - a)$$

blackboard

Introduce overlap

Introduce dynamics

Exercise 5 NOW- from Hopfield to spikes

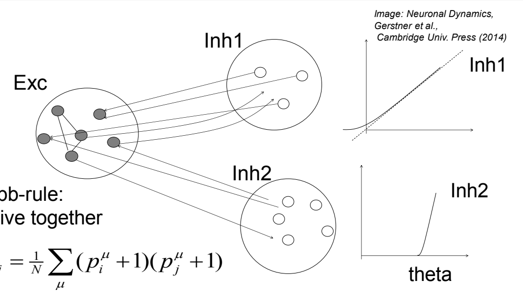
In the Hopfield model, neurons are characterized by a binary variable S_i = +/-1. For an interpretation in terms of spikes it is, however, more appealing to work with a binary variable x_i which is zero or 1.
(i) Write $S_i = 2x_i - 1$ and rewrite the Hopfield model in terms of x_i.
What are the conditions so that the input potential is

$$h_i = [\sum_j w_{ij} x_j(t)]$$

(ii) Interpretation: can you also restric the weights to excitation only?

10 minutes,
Try to get
As far as possible

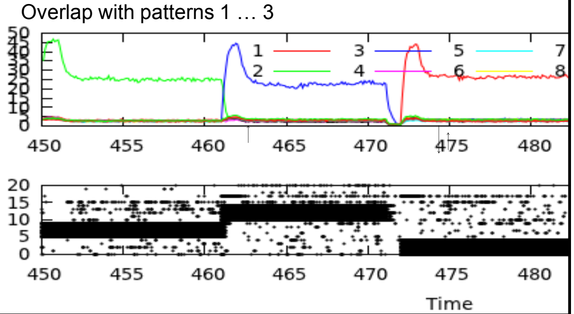
6.4 Separation of excitation and inhibition



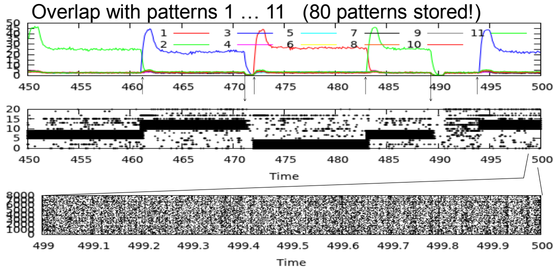
Hebb-rule:
Active together

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

Overlap with patterns 1 … 3



Overlap with patterns 1 … 11 (80 patterns stored!)



Memory with spiking neurons

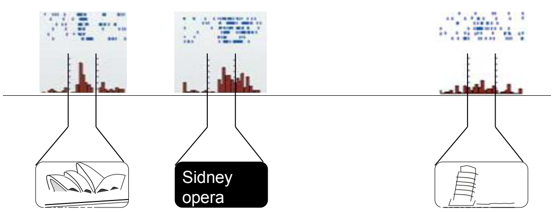
-Low activity of patterns?
-Separation of excitation and inhibition?
-Modeling with integrate-and-fire?

All possible

-Neural data?

6.4 memory data (review from week 5)

Human Hippocampus

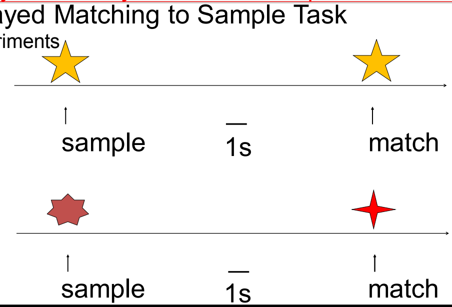


Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature, 435:1102-1107.

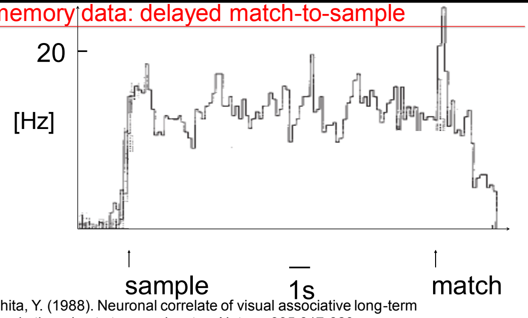
6.4 memory data: delayed match to sample

Delayed Matching to Sample Task

Animal experiments

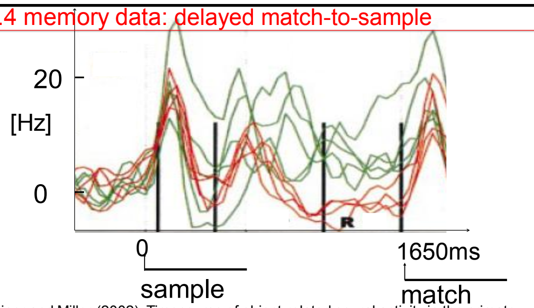


6.4 memory data: delayed match-to-sample



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.

Biological Modeling of Neural Networks

Week 6

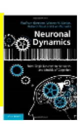
Attractor Networks and Generalizations of the Hopfield model

Wulfram Gerstner

EPFL, Lausanne, Switzerland

Reading for week 6:
NEURONAL DYNAMICS
- Ch. 17.2.5 – 17.4
- Ch. 19.1-19.2;

Cambridge Univ. Press



6.1. Attractor networks

6.2. Stochastic Hopfield model

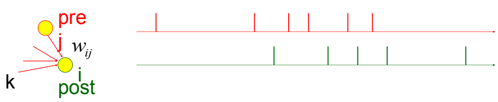
6.3. Energy landscape

6.4. Towards biology
- low-activity patterns
- spiking neurons

6.4 Models of synaptic plasticity
- Hebbian learning rules
- Bienenstock-Cooper-Munro rule

6.5 Online learning of memories

6.2 Hebbian Learning (rate models)



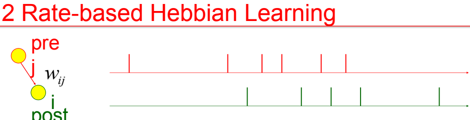
When an axon of cell **j** repeatedly or persistently takes part in firing cell **i**, then j's efficiency as one of the cells firing i is increased

Hebb, 1949

- local rule
- simultaneously active (correlations)

Rate model:
active = high rate = many spikes per second

6.2 Rate-based Hebbian Learning

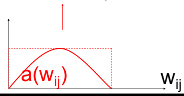


Blackboard

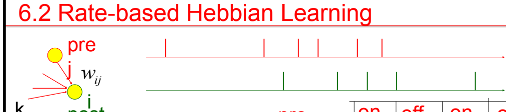
$$\frac{d}{dt} w_{ij} = F(w_{ij}; \nu_j^{pre}, \nu_i^{post})$$

$$\frac{d}{dt} w_{ij} = a_0 + a_1^{pre}\nu_j^{pre} + a_1^{post}\nu_i^{post} + a_2^{corr}\nu_j^{pre}\nu_i^{post} + \dots$$

a = a(w_{ij})



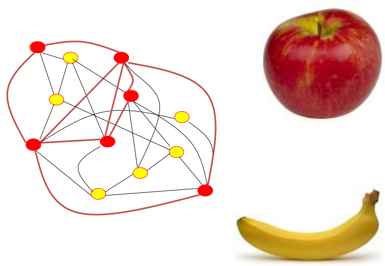
6.2 Rate-based Hebbian Learning



$$\frac{d}{dt} w_{ij} = a_2^{corr}\nu_j^{pre}\nu_i^{post}$$

pre	on	off	on	off
post	on	on	off	off
	+	0	0	0

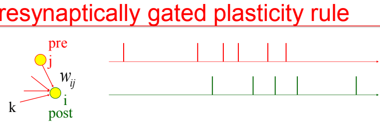
Review from week 5: Hebbian Learning



6.2 Rate-based Hebbian Learning

pre / post	on / on	off / on	on / off	off / off
$\frac{d}{dt} w_{ij} = a_2^{corr} \nu_j^{pre} \nu_i^{post}$	+	0	0	0
$\frac{d}{dt} w_{ij} = a_2^{corr} \nu_j^{pre} \nu_i^{post} - c$	+	-	-	-
$\frac{d}{dt} w_{ij} = a_2^{corr} \nu_j^{pre} (\nu_i^{post} - \vartheta)$	+	0	-	0
$\frac{d}{dt} w_{ij} = a_2^{corr} (\nu_j^{pre} - \vartheta)(\nu_i^{post} - \vartheta)$	+	-	-	+

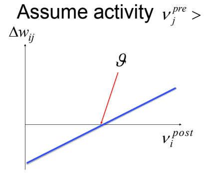
6.2 Presynaptically gated plasticity rule



presynaptically gated

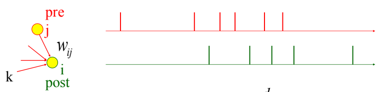
$$\frac{d}{dt} w_{ij} = a_2^{corr} (\nu_i^{post} - \vartheta) \nu_j^{pre}$$

Assume activity $\nu_j^{pre} > 0$



6.2 Bienenstock-Cooper-Munro rule

Bienenstock, Cooper Munro, 1982



presynaptically gated

$$\frac{d}{dt} w_{ij} = a_2^{corr} (\nu_i^{post} - \vartheta) \nu_j^{pre}$$

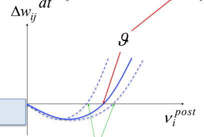
BCM: 3rd order ('triplet')

$$\frac{d}{dt} w_{ij} = b\, \nu_i^{post} (\nu_i^{post} - \vartheta) \nu_j^{pre}$$

Map to
Taylor expansion?

triplet pair

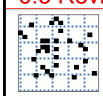
$$\frac{d}{dt} w_{ij} = b\, \nu_i^{post} (\nu_i^{post} - \vartheta) \nu_j^{pre}$$



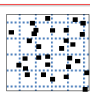
assume $\nu_j^{pre} > 0$

Homeostasis $\vartheta = f(\bar{\nu}_i^{post})$

6.3 Review of week 5: Deterministic Hopfield model



Prototype $\vec{p}^1$



Prototype $\vec{p}^2$

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

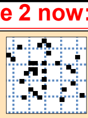
Sum over all prototypes

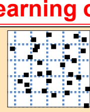
Input potential $h_i = \sum_j w_{ij} S_j$

Sum over all inputs to neuron *i* prototypes

Exercise 2 now: learning of prototypes

Prototype $\vec{p}^1$





Prototype $\vec{p}^2$

interactions

(1) $w_{ij} = \sum_{\mu} p_i^{\mu} p_j^{\mu}$

Sum over all prototypes

a) Show that (1) corresponds to a rate learning rule

(2) $\frac{d}{dt} w_{ij} = a_2^{corr} (\nu_j^{pre} - \vartheta)(\nu_i^{post} - \vartheta)$

Take 8 minutes,
start the exercise
Next lecture at 11:35

Assume that weights are zero at the beginning;

Each pattern is presented (enforced) during 0.5 sec (One after the other).

note that $p_j^{\mu} = \pm 1$ but $\nu_j \geq 0$

b) Compare with: $\frac{d}{dt} w_{ij} = a_0 + a_1^{pre} \nu_j^{pre} + a_1^{post} \nu_i^{post} + a_2^{corr} \nu_j^{pre} \nu_i^{post} + \ldots$

c) Is this unsupervised learning?

Biological Modeling of Neural Networks

Week 6
Hebbian LEARNING and
ASSOCIATIVE MEMORY

Wulfram Gerstner
EPFL, Lausanne, Switzerland

6.1 Synaptic Plasticity
- Hebbian Learning
- Short-term Plasticity
- Long-term Plasticity
- Reinforcement Learning

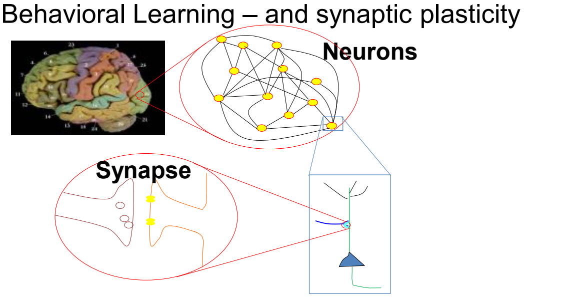
6.2 Models of synaptic plasticity
- Hebbian learning rules
- Bienenstock-Cooper-Munro rule

6.3 Hopfield Model
- probabilistic
- energy landscape

6.4 Attractor memories

6.5 Online learning of memories

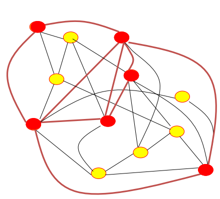
Behavioral Learning – and synaptic plasticity



Synaptic Plasticity = Change in Connection Strength

6.5 Review: Hebbian Learning/Assemblies

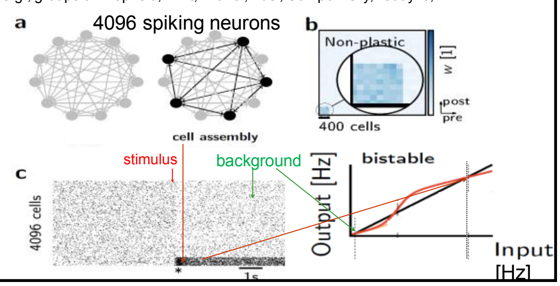
Recall:
Partial info



item recalled

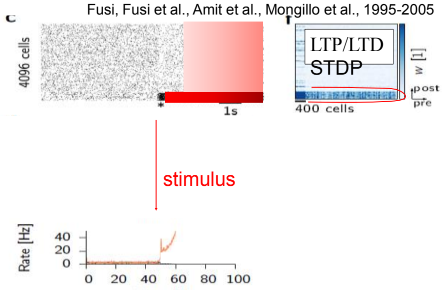
6.5 Preconfigured memory: bistable network

e.g., groups of Hopfield, Amit, Brunel, Fusi, Sompolinsky, Tsodyks,

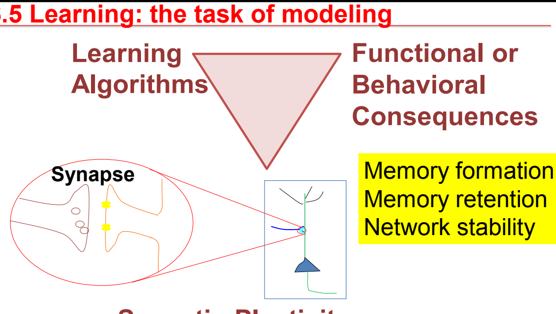


6.5 Learning the memory: very hard

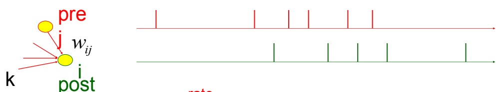
Fusi, Fusi et al., Amit et al., Mongillo et al., 1995-2005



6.5 Learning: the task of modeling



6.5 Review: Rate models of Hebbian learning



rate rate

$$\frac{dw_{ij}}{dt} = F(w_{ij};\nu_j^{pre},\nu_i^{post})$$

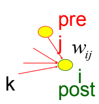
- local rule
- simultaneously active

pair

$$\frac{dw_{ij}}{dt} = a_0 + a_1^{pre}\nu_j^{pre} + a_1^{post}\nu_i^{post} + a_2^{corr}\nu_j^{pre}\nu_i^{post} +$$

depend on w_{ij}

6.5 Induction of Plasticity

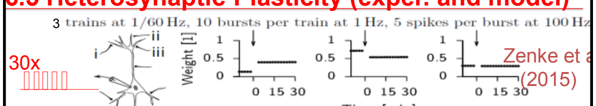


- **homosynaptic/Hebb** ('pre' and 'post')
- **heterosynaptic plasticity** (pure 'post'-term)
- **transmitter-induced** (pure 'pre'-term)

$$\frac{dw_{ij}}{dt} = a_0 + a_1^{pre}\nu_j^{pre} + a_1^{post}\nu_i^{post} + a_2^{corr}\nu_j^{pre}\nu_i^{post} +$$

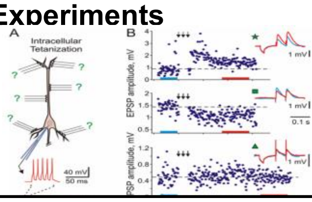
$$+a_3^{BCM}\nu_j^{pre}(\nu_i^{post})^2 \quad +a_4^{post}(w_{ij})[\nu_i^{post}]^4$$

6.5 Heterosynaptic Plasticity (exper. and model)



Zenke et al. (2015)

Experiments



Chen et al. 2013,
Chistiakova et al. 2014
See also:
Lynch et al. 1977

6.5 Induction of Plasticity (rate-based)

- nonlinear Hebb for potentiation

$$+a_3^{BCM}\nu_j^{pre}(\nu_i^{post})^2$$

- pre-post for depression

$$-a_2^{LTD}\nu_j^{pre}\nu_i^{post}$$

Bienenstock et al., 1982
Pfister and Gerstner, 2006

- heterosynaptic plasticity (pure 'post')

$$-a_4^{het}(w_{ij}-z_{ij})[\nu_i^{post}]^4$$

- transmitter-induced (pure 'pre')

$$+a_1^{pre}\nu_j^{pre}$$

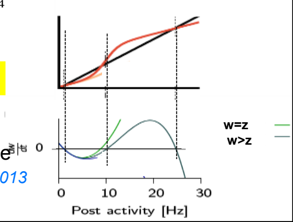
6.5 Plasticity model in network

$$\frac{dw_{ij}}{dt}=a_1^{pre}\nu_j^{pre}-a_2^{LTD}\nu_j^{pre}\nu_i^{post}+a_3^{BCM}\nu_j^{pre}(\nu_i^{post})^2$$
$$-a_4^{het}(w_{ij}-z_{ij})[\nu_i^{post}]^4$$

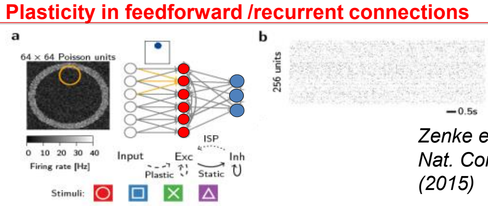
→Self-stabilizing!

Heterosynaptic plasticity must act on the same time scale

Zenke+Gerstner, PLOS Comp. B. 2013
Zenke et al., Nat. Comm., 2015

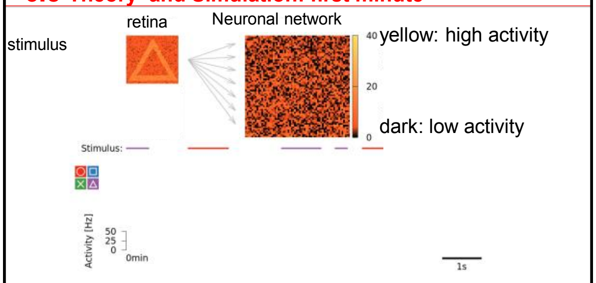


6.5. Plasticity in feedforward /recurrent connections



Zenke et al., Nat. Comm. (2015)

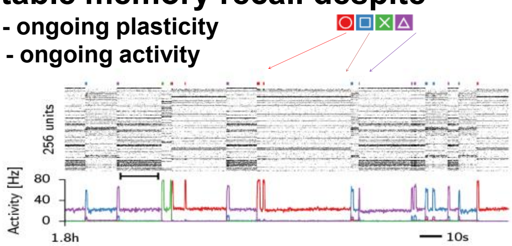
6.5 Theory and Simulation: first minute



6.5 Plasticity model in network

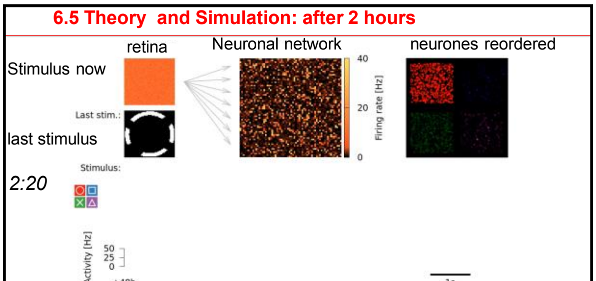
Stable memory recall despite

- **ongoing plasticity**
- **ongoing activity**



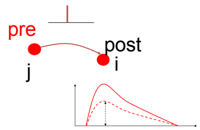
Zenke et al., Nat. Comm. (2015)

6.5 Theory and Simulation: after 2 hours



6.5 Synaptic changes – review and summary

Induction of changes
- fast (if stimulated appropriately)
- slow (homeostasis)



Persistence of changes
- long (LTP/LTD)
- short (short-term plasticity)

Functionality
- **useful for learning a new behavior/new memories**
- useful for development (wiring for receptive field development)
- **useful for activity control in network (homeostasis)**
- useful for coding

Biological Modeling of Neural Networks

Week 6
Hebbian LEARNING and ASSOCIATIVE MEMORY

Wulfram Gerstner
EPFL, Lausanne, Switzerland

6.1 Synaptic Plasticity
- Hebbian Learning
- Short-term Plasticity
- Long-term Plasticity
- Reinforcement Learning

6.2 Models of synaptic plasticity
- Hebbian learning rules
- Bienenstock-Cooper-Munro rule

6.3 Hopfield Model
- probabilistic
- energy landscape

6.4 Attractor memories

$\boxed{\textbf{6.5 Online learning of memories}}$

The end