EPFL

Week 6

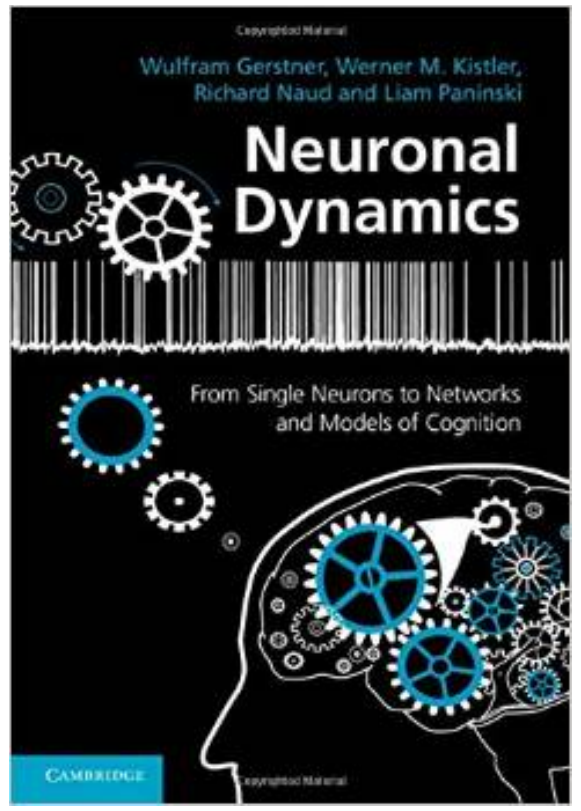
Attractor Networks and Generalizations of the Hopfield model

Wulfram Gerstner

EPFL, Lausanne, Switzerland

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

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

Lecture 11 of video series

https://lcnwww.epfl.ch/gerstner/NeuronalDynamics-MOOCall.html

1. Review of last week: overlap / correlation

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

Overlap: similarity between state *S(t)* and pattern

1. Review of last week: overlap / correlation

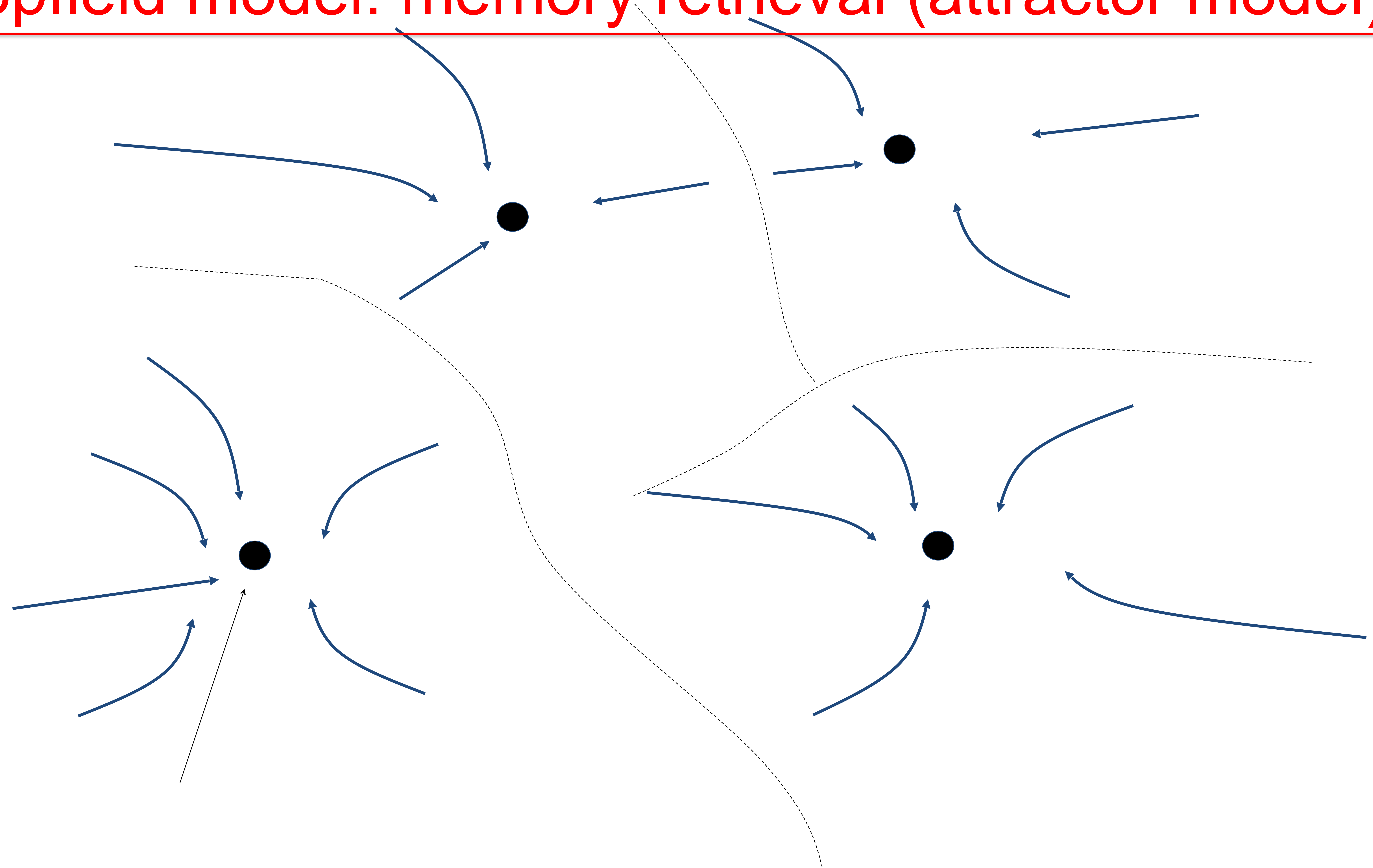
Image: Neuronal Dynamics, Gerstner et al., Cambridge Univ. Press (2014),

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

Overlap: similarity between state *S(t)* and pattern

Orthogonal patterns

1. Hopfield model: memory retrieval (attractor model)



Quiz 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.

Quiz 1: overlap and attractor dynamics

[x] [ ] The overlap is maximal
if the network state matches one of the patterns.

[x] [ ] 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.

Quiz 1: overlap and attractor dynamics

[x] [ ] The overlap is maximal
if the network state matches one of the patterns.

[x] [ ] The overlap increases during memory retrieval.

[ ] [ ] The mutual overlap of orthogonal patterns is one.

[x] [ ] In an attractor memory, the dynamics converges to a stable
fixed point.

[x] [ ] 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.

Quiz 1: overlap and attractor dynamics

[x] [ ] The overlap is maximal
if the network state matches one of the patterns.

[x] [ ] The overlap increases during memory retrieval.

[ ] [ ] The mutual overlap of orthogonal patterns is one.

[x] [ ] In an attractor memory, the dynamics converges to a stable
fixed point.

[x] [ ] 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.

6.1 Summary:

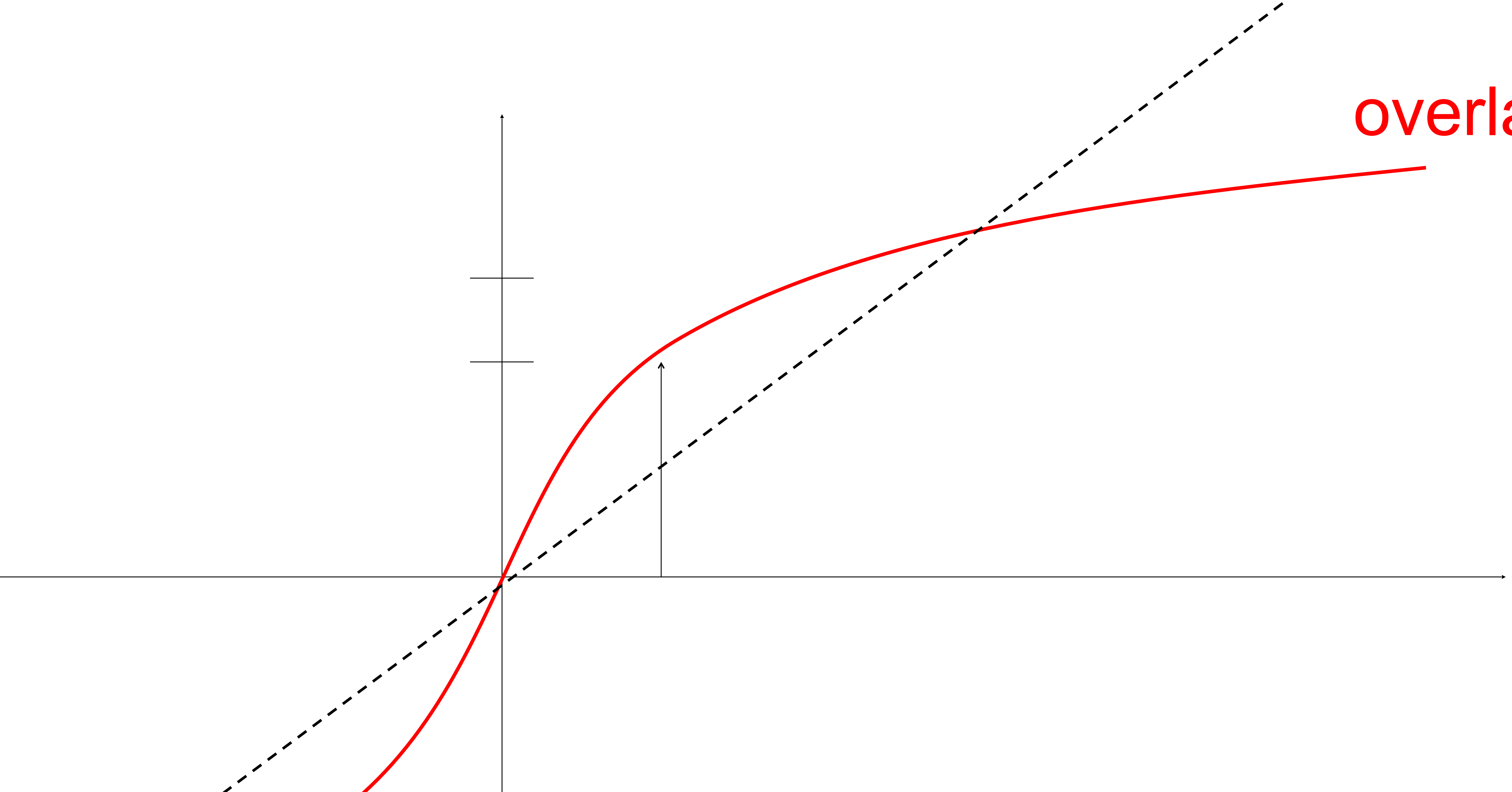
- Attractor dynamics implies that the dynamics of a network are attracted towards certain preferred states.
- These states correspond to the previously stored patterns (also called stored memories, or ‘engrams’
- The movement of the network dynamics towards these special states is called memory retrieval. Once the network is in one of these special states we say that the corresponding memory has been retrieved.
- Mathematically, the attractor states are the fixed points of the network dynamics
- Mathematically, the overlap with pattern mu measures the similarity between the momentary network state (activity of all neurons) and the stored memory mu.
- In the attractor state, and number of patterns small compared to N the overlap is (close to) 1
- Low-activity patterns are patterns where the ratio of black pixels to white pixels is not 50:50.   The mean fraction of black pixels (neurons that should be ‘ON’) is 2a – 1   where a is the expectation value of p_i^μ

2. Stochastic Hopfield model: memory retrieval

Overlap:

Neurons that should be 'on'

Neurons that should be 'off'



2. Stochastic Hopfield model = attractor model

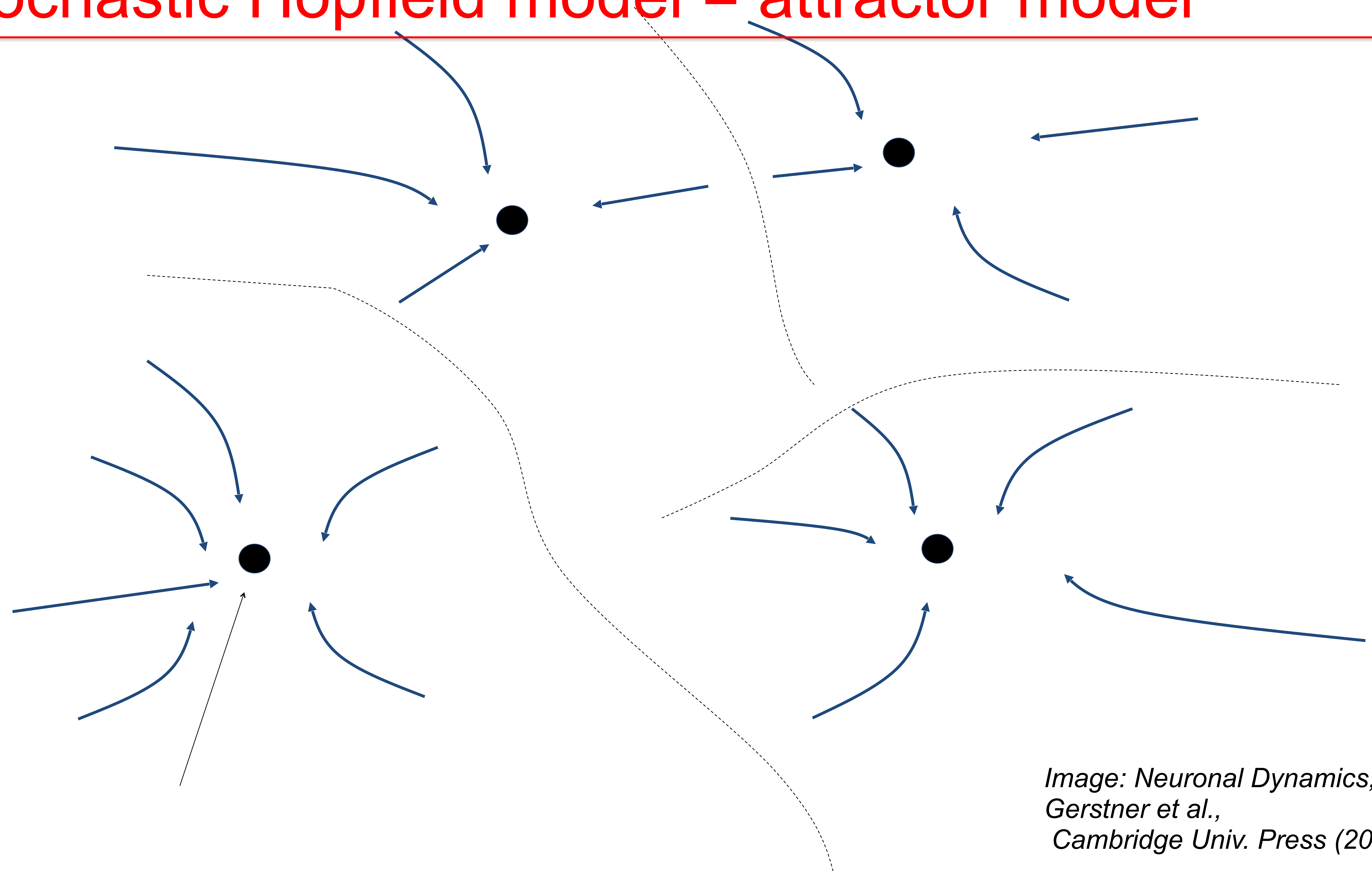


Image: Neuronal Dynamics, Gerstner et al., Cambridge Univ. Press (2014),

Quiz 2: Stochastic networks and overlap equations

[ ] The update of the overlap leads always to a fixed point with overlap m=1

[ ] The update equation as derived here implicitly assumed **orthogonal** patterns because otherwise we would have to analyze overlaps with several patterns in **parallel**

[ ] The update equation as derived here requires a function

Quiz 2: Stochastic networks and overlap equations

[ ] [ ] The update of the overlap leads always to a fixed point with overlap m=1

[x] [ ] The update equation as derived here implicitly assumed **orthogonal** patterns because otherwise we would have to analyze overlaps with several patterns in **parallel**

[ ] [ ] The update equation as derived here requires a function

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

6.2 Summary:

- Instead of deterministic dynamics, we can also study stochastic dynamics.

- Memory retrieval is possible with stochastic dynamics, if the number of patterns is small (in some sense) to the total number of neurons. In this case, the similarity variables (also called overlap variable) have negligible noise, because they involve averaging over a large number of neurons.

- Similar to deterministic dymanics we find fixed points with a large overlap.

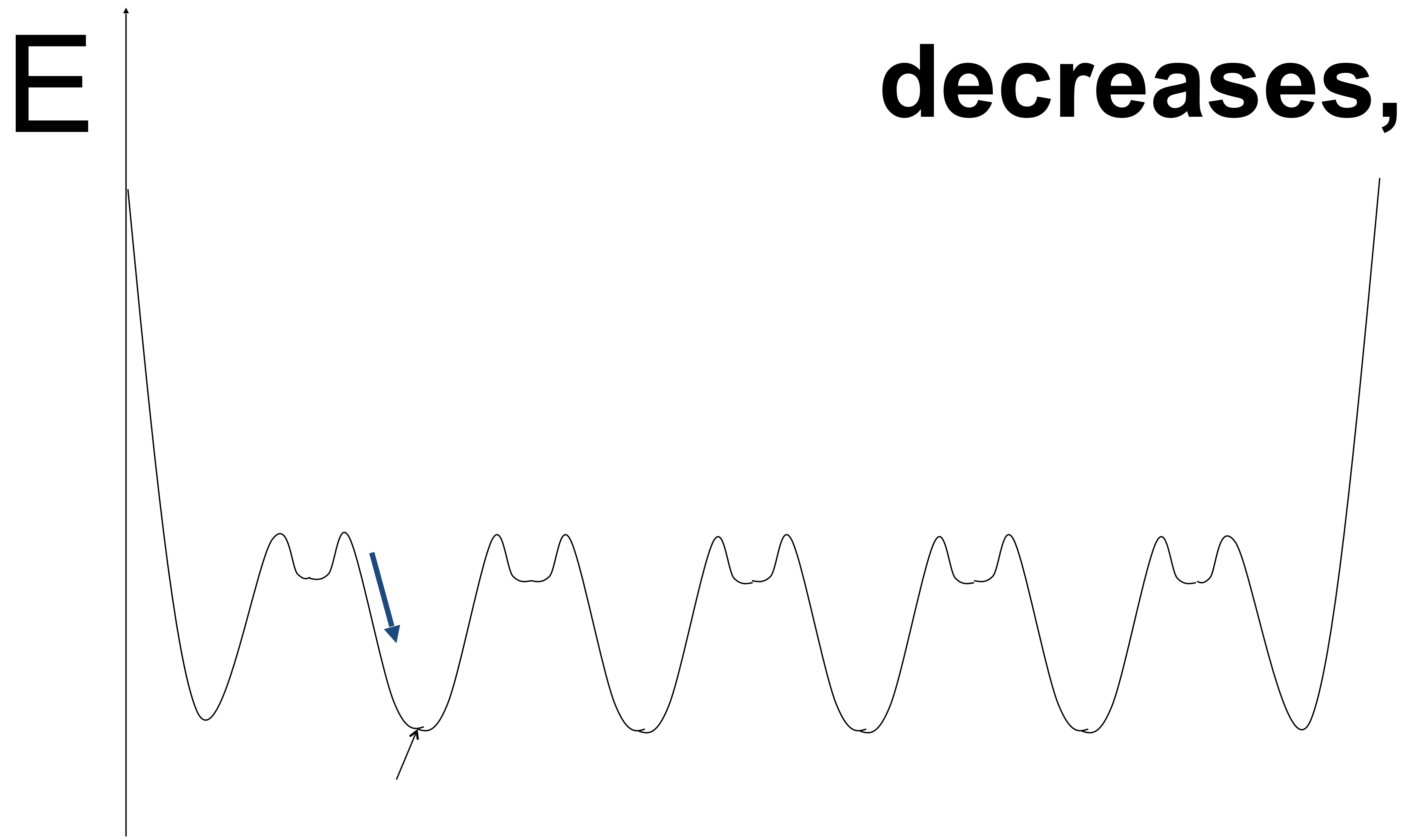
- Since random patterns are nearly orthogonal, a large overlap with one of the patterns implies that the overlap with other patterns in this network state is small.

3. Symmetric interactions: Energy/Lyapunov function

Assume symmetric interaction,
Assume deterministic asynchronous update

Claim: the energy

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

3. Energy picture

energy picture historically important:

- capacity calculations

Amit-Gutfreund-Sompolinsky (1985) Spin-glass models of neural network Phys. Rev. A. 32: 1007-1018

energy picture is very general:

- also possible for patterns with non-zero mean.

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

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.

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

Peretto (1984) Collective Properties of Neural Networks: A statistical Mechanics Approach Biol. Cybern. 50: 51-62 (received November 1983).

Amit-Gutfreund-Sompolinsky (1985) Spin-glass models of neural network Phys. Rev. A. 32: 1007-1018

Energy picture with rate neurons:

Cohen and Grossberg (1983) Absolute Stability of Global Pattern Formation and Parallel Memory Storage by Competitive Neural Network in IEEE Trans. Systems, Man, Cybernetics

Hopfield (1984), Neurons with graded response have collective computational properties like those of two-state neurons. PNAS 81: 3088-3092

3. Energy picture

energy picture historically important:

- capacity calculations

Amit-Gutfreund-Sompolinsky (1985) Spin-glass models of neural network Phys. Rev. A. 32: 1007-1018

energy picture is very general:

- also possible for patterns with non-zero mean.

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

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.

energy picture is a side-track:

- it needs symmetric interactions

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

Peretto (1984) Collective Properties of Neural Networks: A statistical Mechanics Approach Biol. Cybern. 50: 51-62 (received November 1983).

Amit-Gutfreund-Sompolinsky (1985) Spin-glass models of neural network Phys. Rev. A. 32: 1007-1018

Energy picture with rate neurons:

Cohen and Grossberg (1983) Absolute Stability of Global Pattern Formation and Parallel Memory Storage by Competitive Neural Network in IEEE Trans. Systems, Man, Cybernetics

Hopfield (1984), Neurons with graded response have collective computational properties like those of two-state neurons. PNAS 81: 3088-3092

Quiz 3: Energy picture and Lyapunov function

Let be the energy of the Hopfield model

and 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

Quiz 3: Energy picture and Lyapunov function

Let be the energy of the Hopfield model

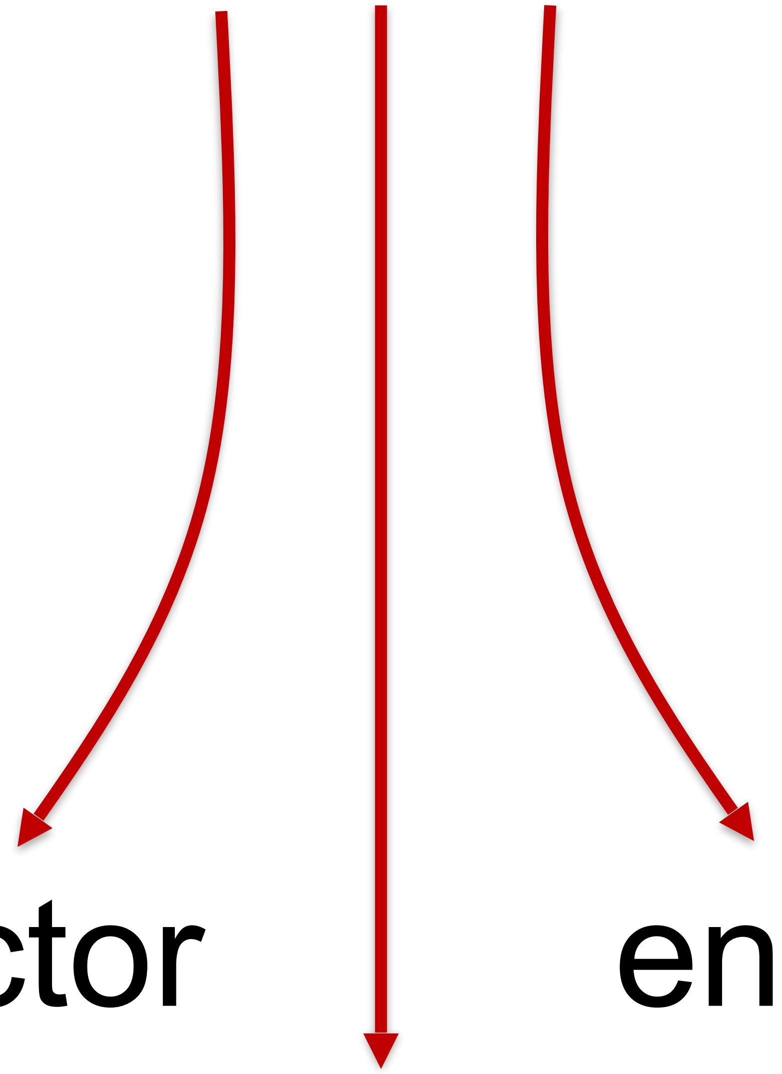
and the dynamics.

[ ] [ ] The energy picture requires random patterns with prob = 0.5
[x] [ ] 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
[x] [ ] In each step, the value of a Lyapunov function decreases or stays constant
[x] [ ] Under deterministic dynamics the above energy is a Lyapunov function

3. Energy picture

Hopfield model

special case



attractor

energy

biology

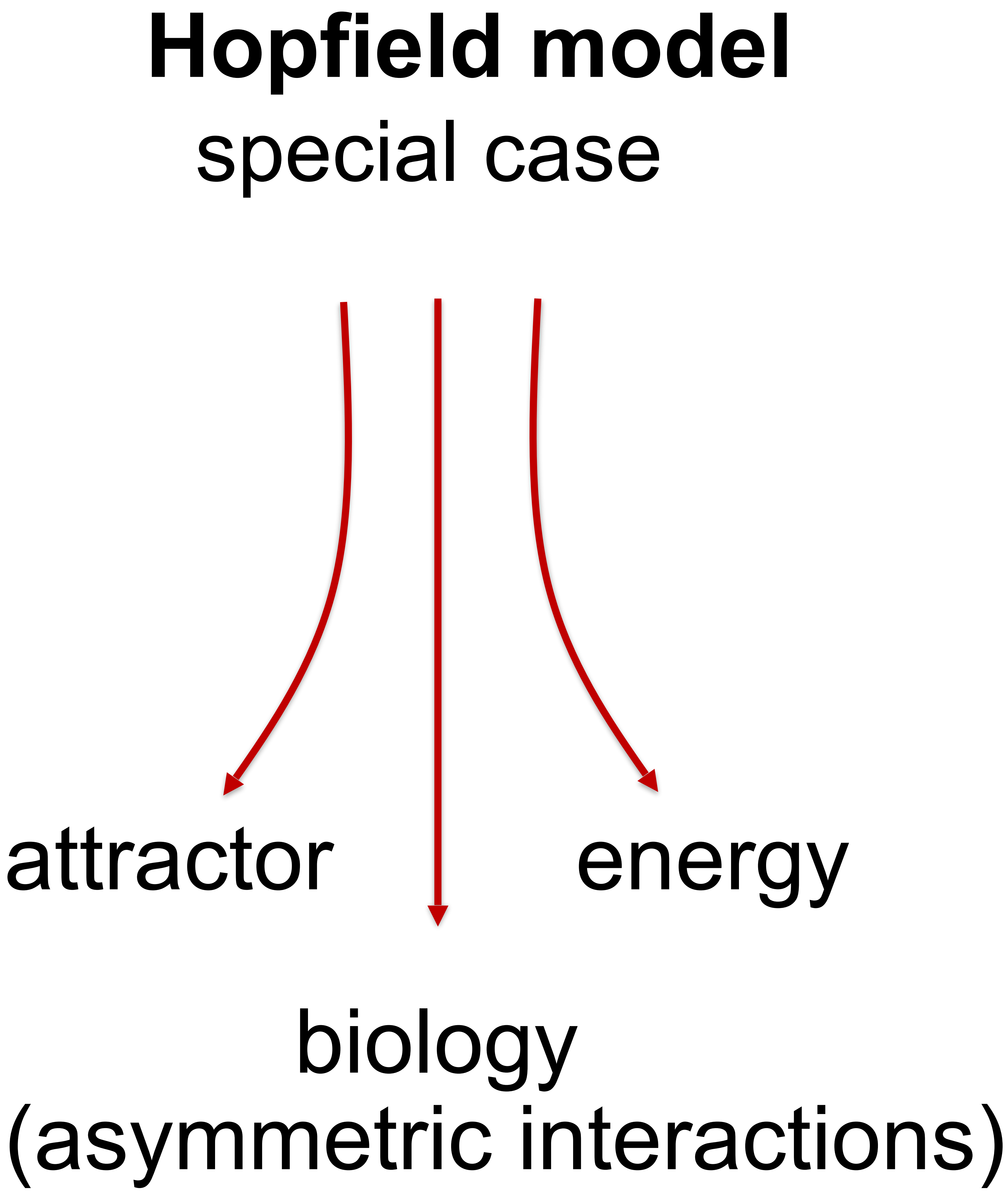
(asymmetric interactions)

3. Energy picture

We will **not use** the energy picture – because the assumption of symmetric interactions is too restrictive.

We are interested in biology:
while brain-areas are typically reciprocally connected, it is unlikely that we would find perfect symmetry of interactions on a **neuron-to-neuron basis.**

Nearly all calculations can be done without the energy picture



6.3 Summary:

- If the weight matrix is symmetric, then network dynamics can be described as down-hill movement in an energy landscape.

- The Energy is also called a Lyapunov function of the system.

- In the Hopfield model, states of lowest energy correspond to the attractors (which in term correspond to states with overlap 1, and hence to the stored memories)

- The energy picture is much more restrictive than the general attractor dynamics.

4. attractor memory with ‘low’ activity patterns

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

with overlap

yield dynamics

$b=0$ or $b=1$

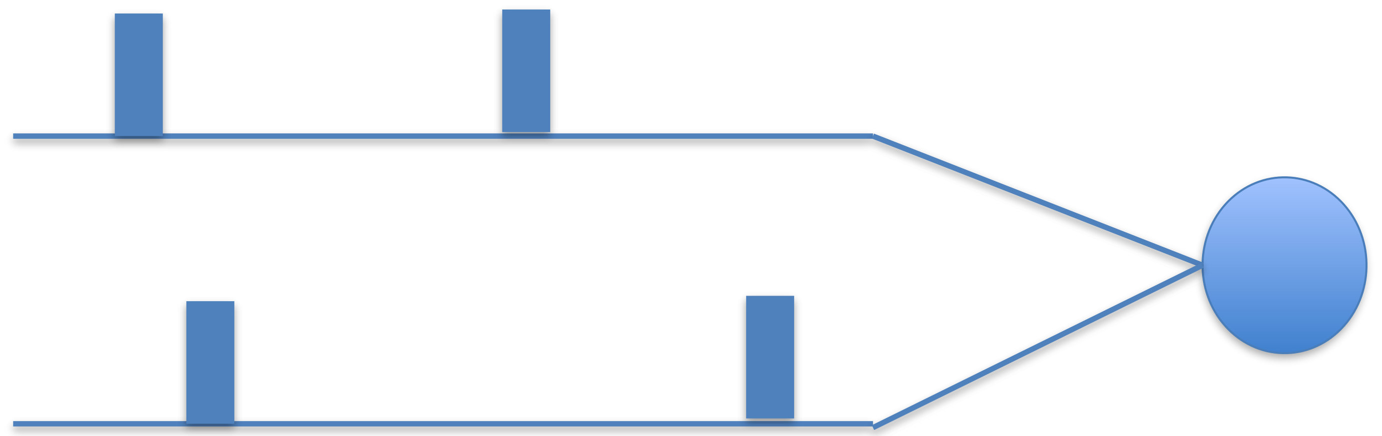
6.4 Summary:

- In the standard-Hopfield model, the patterns that are stored have 50 percent active and 50 percent inactive neurons. Low activity-patterns means that in each pattern only a small number of neurons is active.

- With an appropriate choice of the connection weights, low-activity patterns can become stable fixed points of the attractor dynamics

- Note that there is no need for the weights to be symmetric!

5. attractor memory with spiking neurons

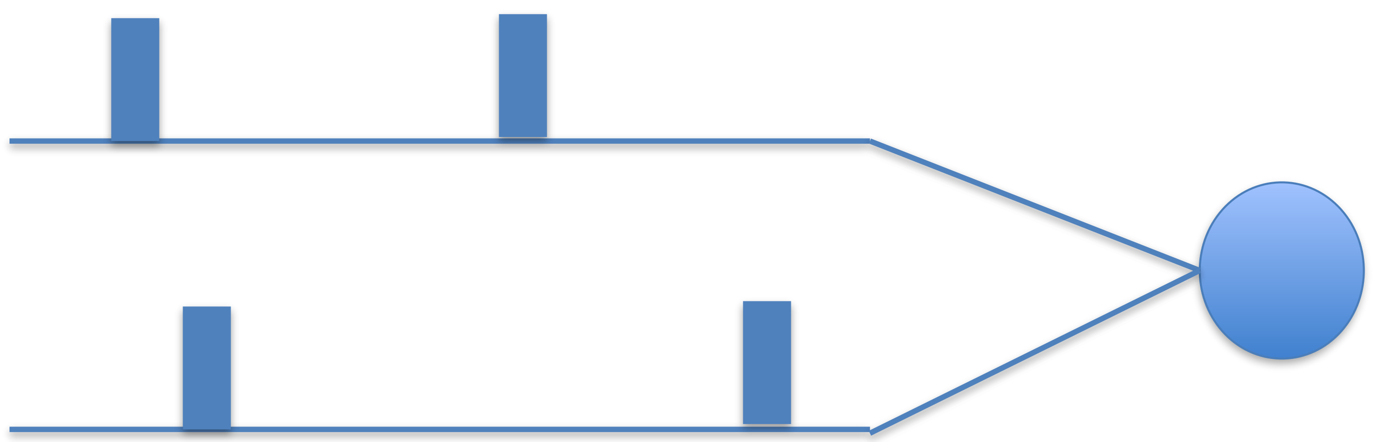


Total input to neuron i

- rewrite binary state variable:

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

5. attractor memory with spiking neurons

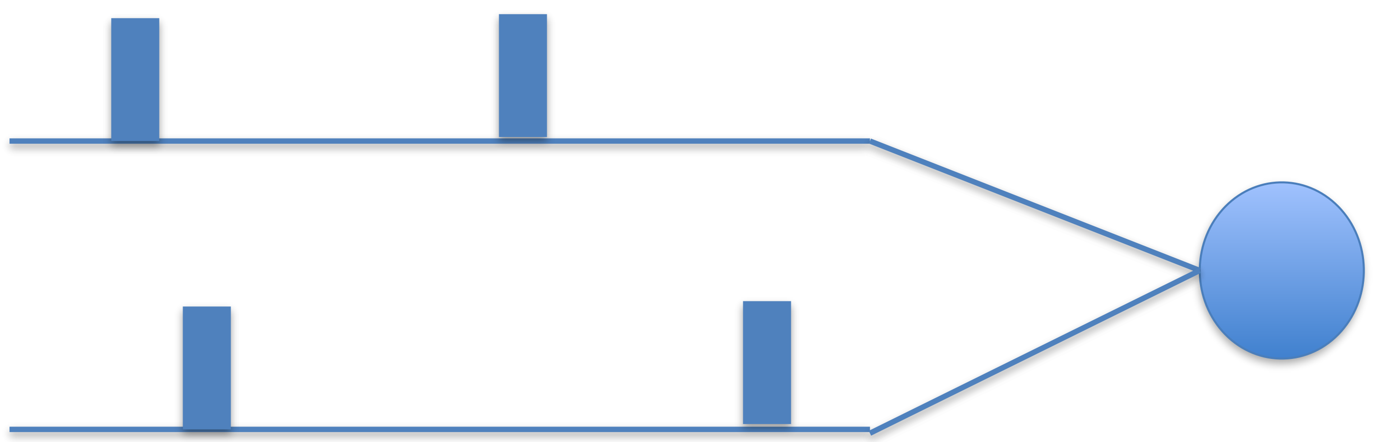


Total input to neuron i

- rewrite binary state variable:

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

5. attractor memory with spiking neurons

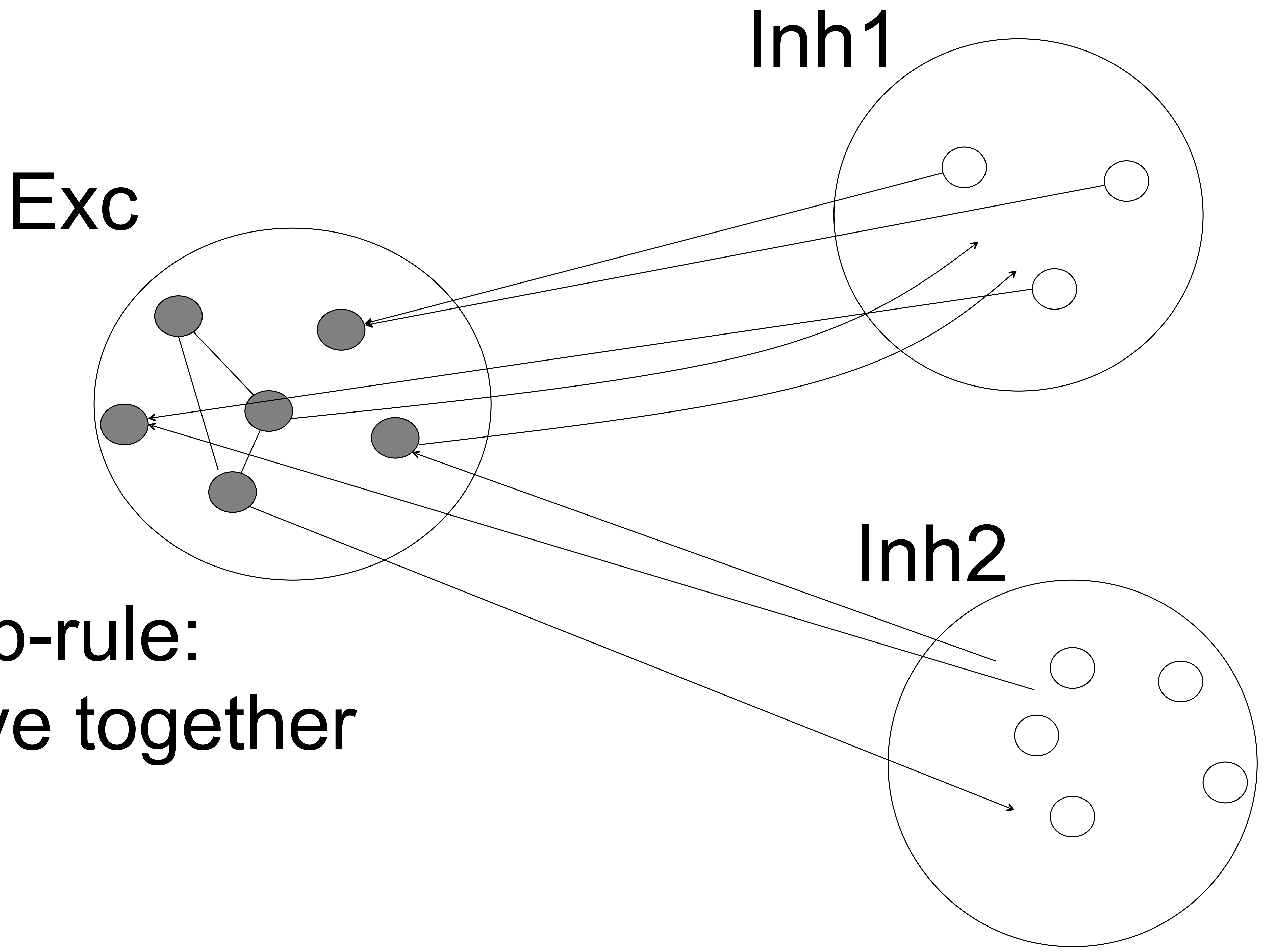


Total input to neuron i

Separation of excitation/ inhibition

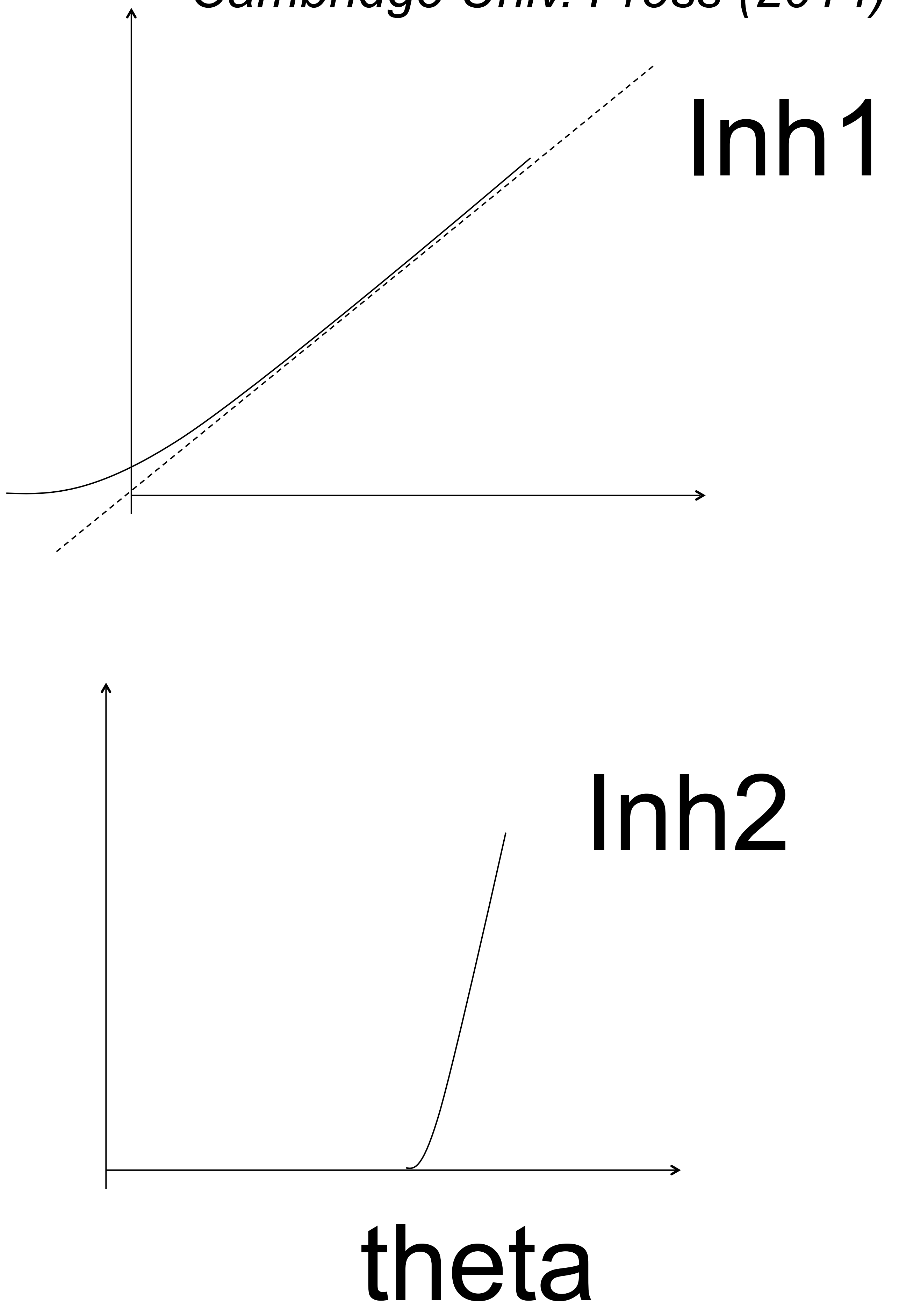
- rewrite weights:

5. Separation of excitation and inhibition



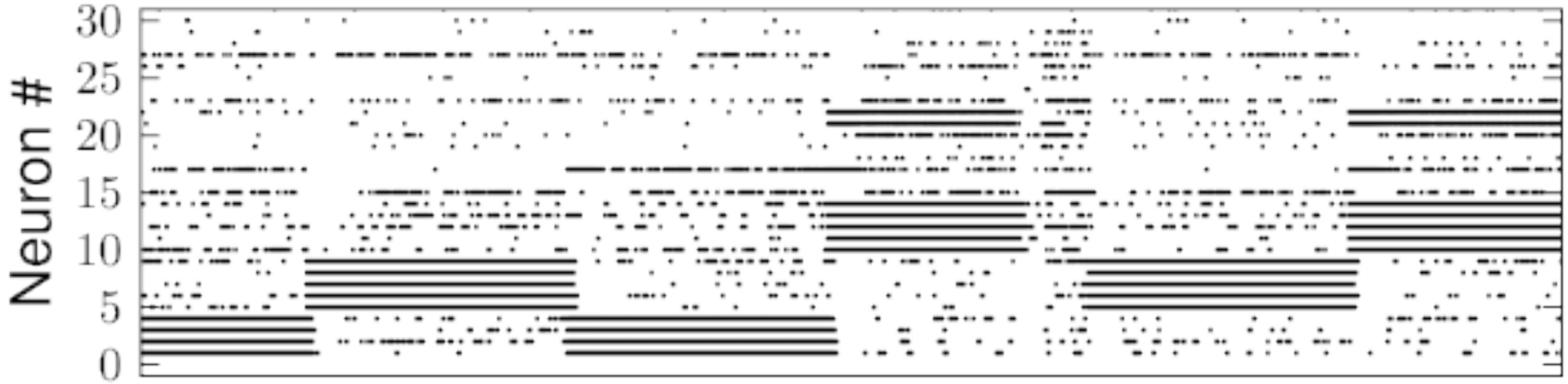
Hebb-rule:
Active together

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



5. attractor memory with 8000 spiking neurons

Spike raster



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

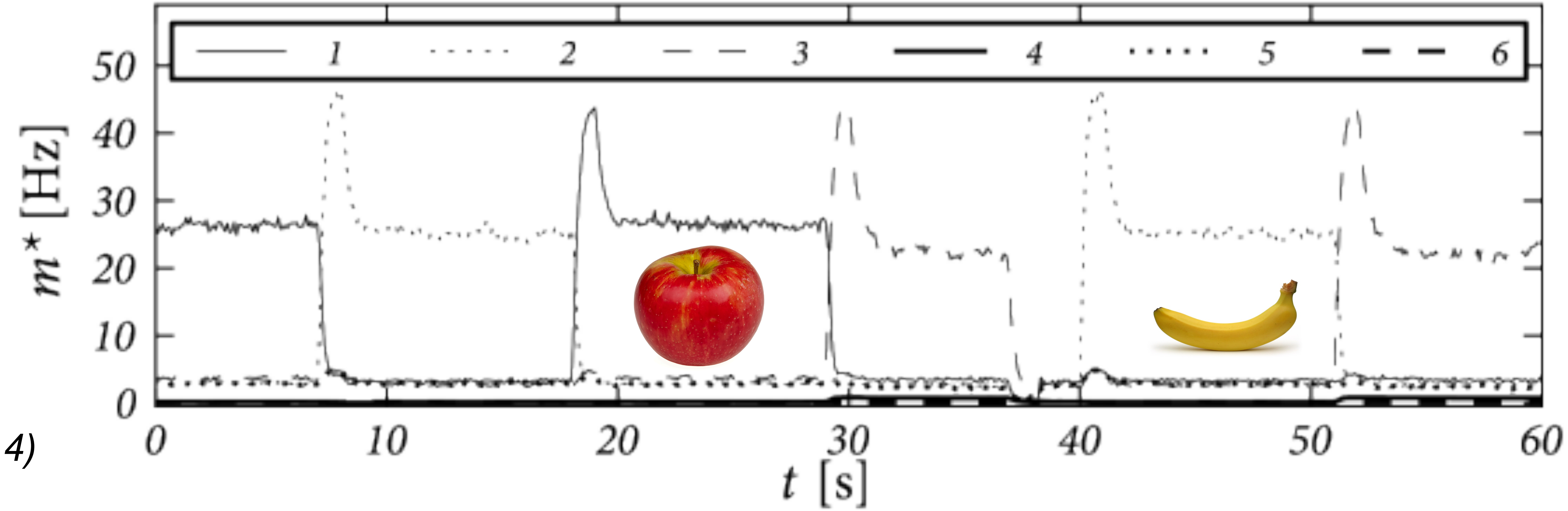


Image: Neuronal Dynamics, Gerstner et al., Cambridge Univ. Press (2014)

Memory with spiking neurons

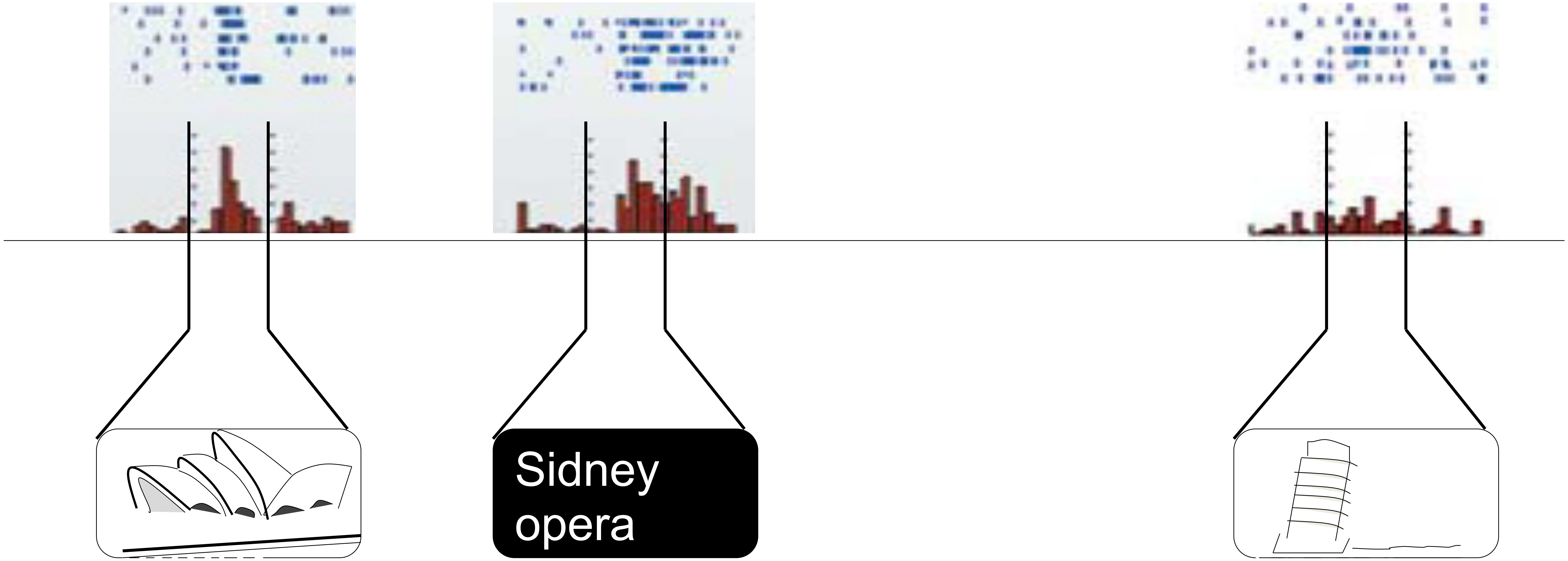
-Low activity of patterns?
-Separation of excitation and inhibition?
-Modeling with integrate-and-fire?
-Asymmetric weights
-Low connection probability

All possible

-Neural data?

5. memory data (review from week 5)

Human Hippocampus

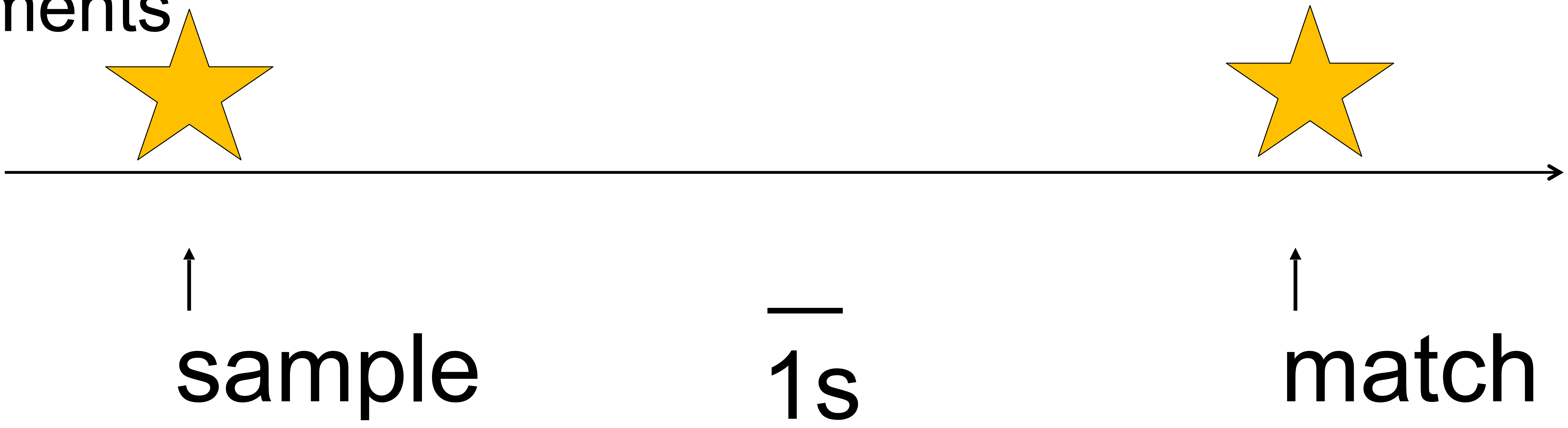


Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature, 435:1102-1107.

5. memory data: delayed match to sample

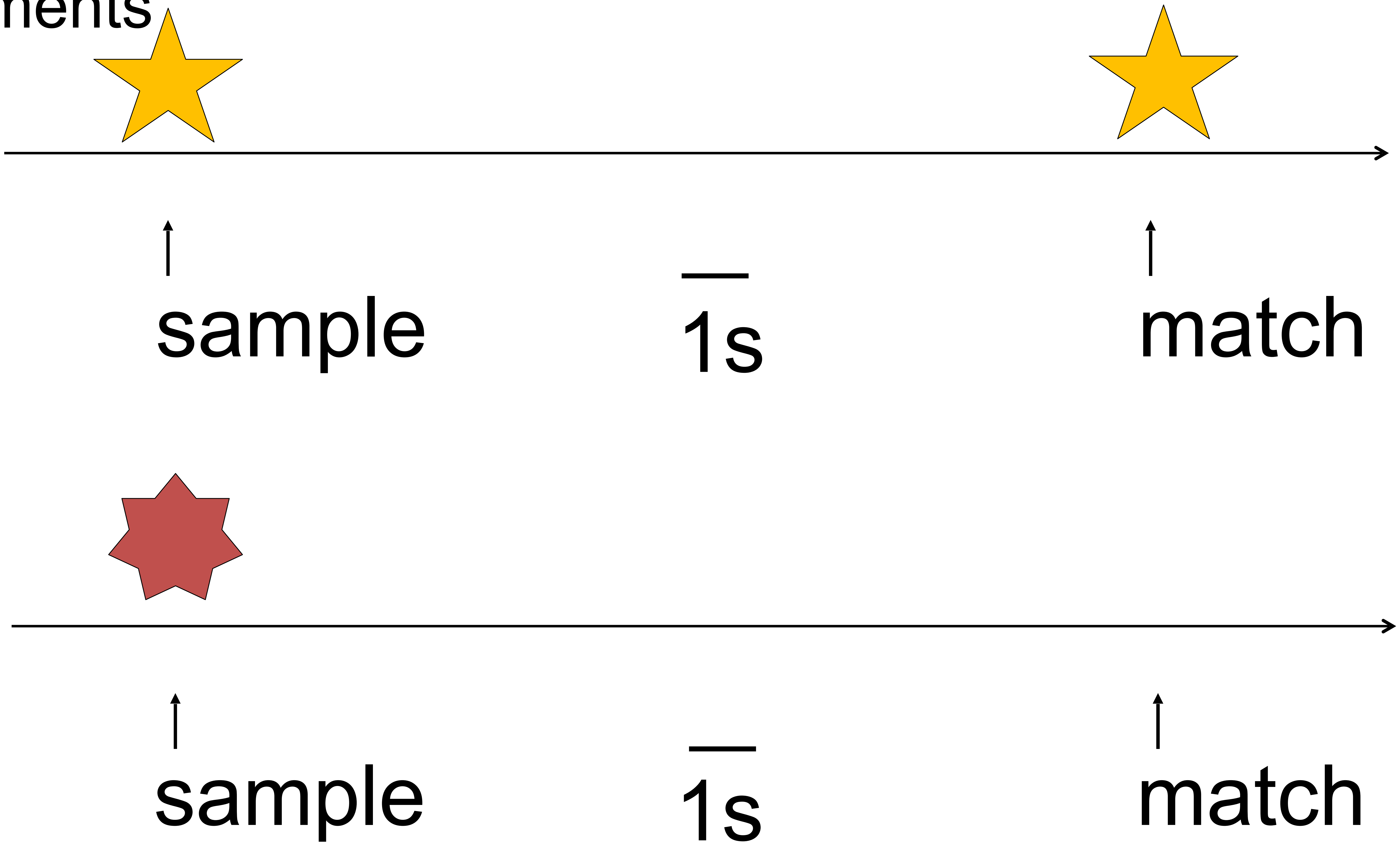
Delayed Matching to Sample Task

Animal experiments



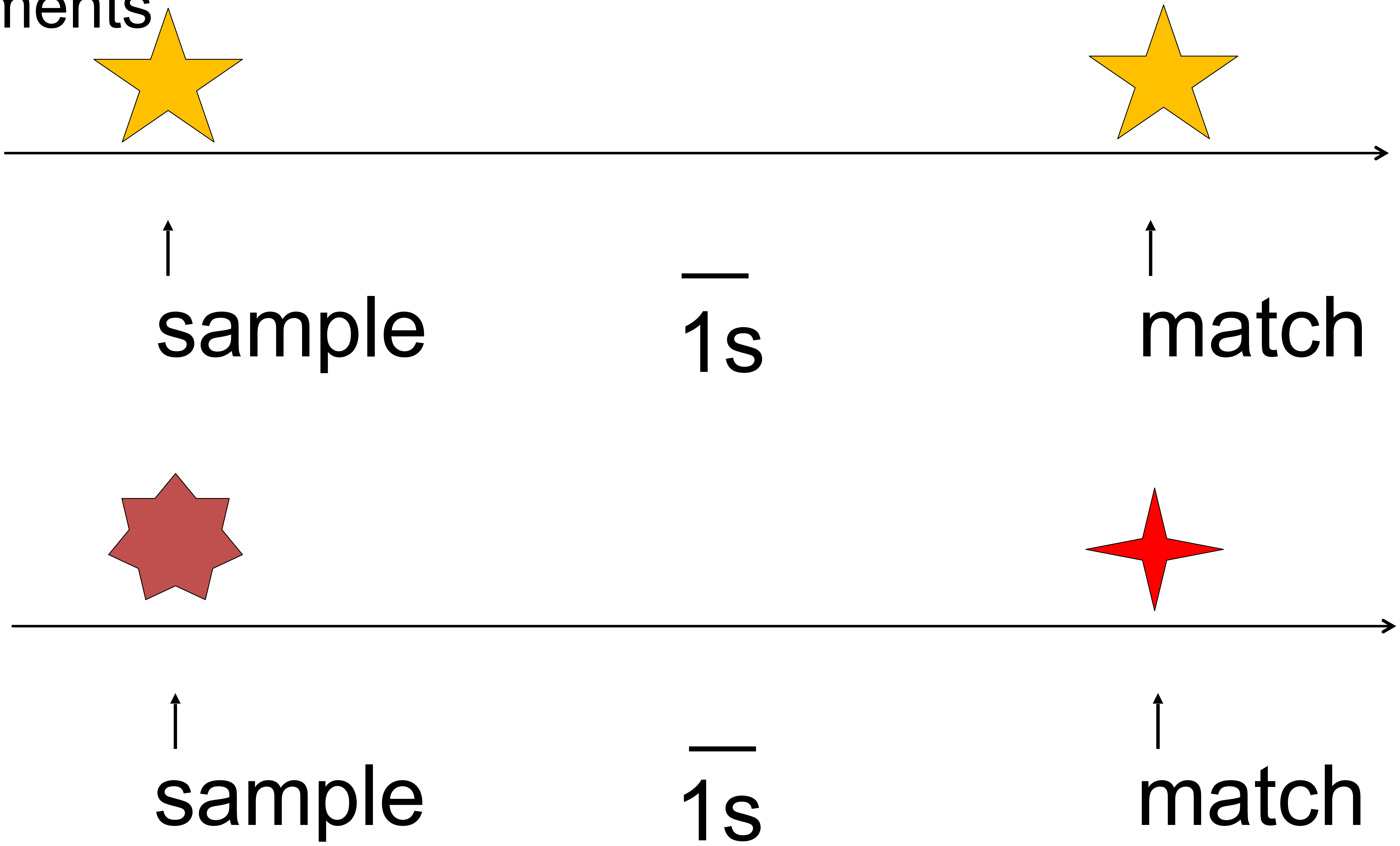
Delayed Matching to Sample Task

Animal experiments

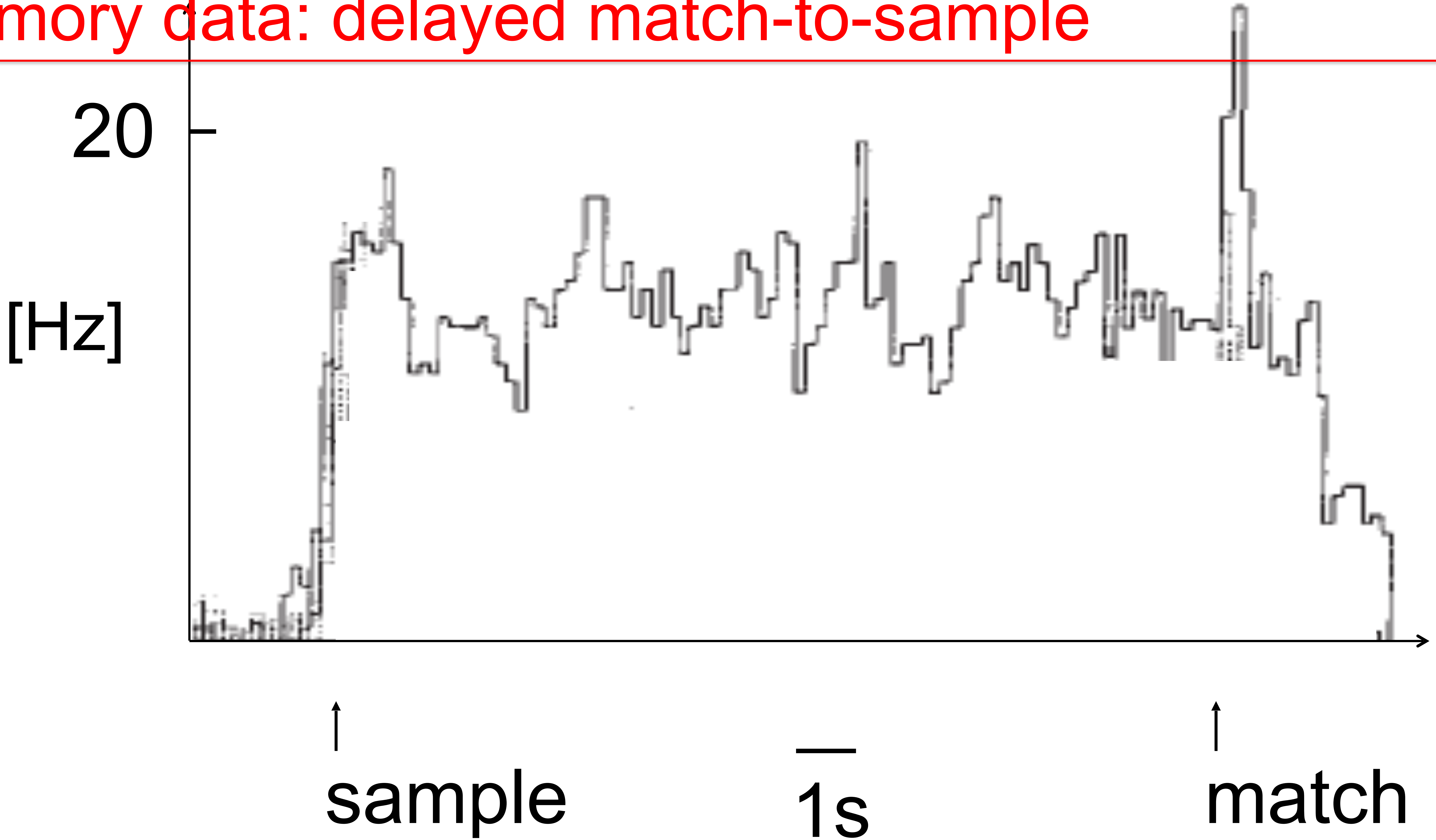


Delayed Matching to Sample Task

Animal experiments

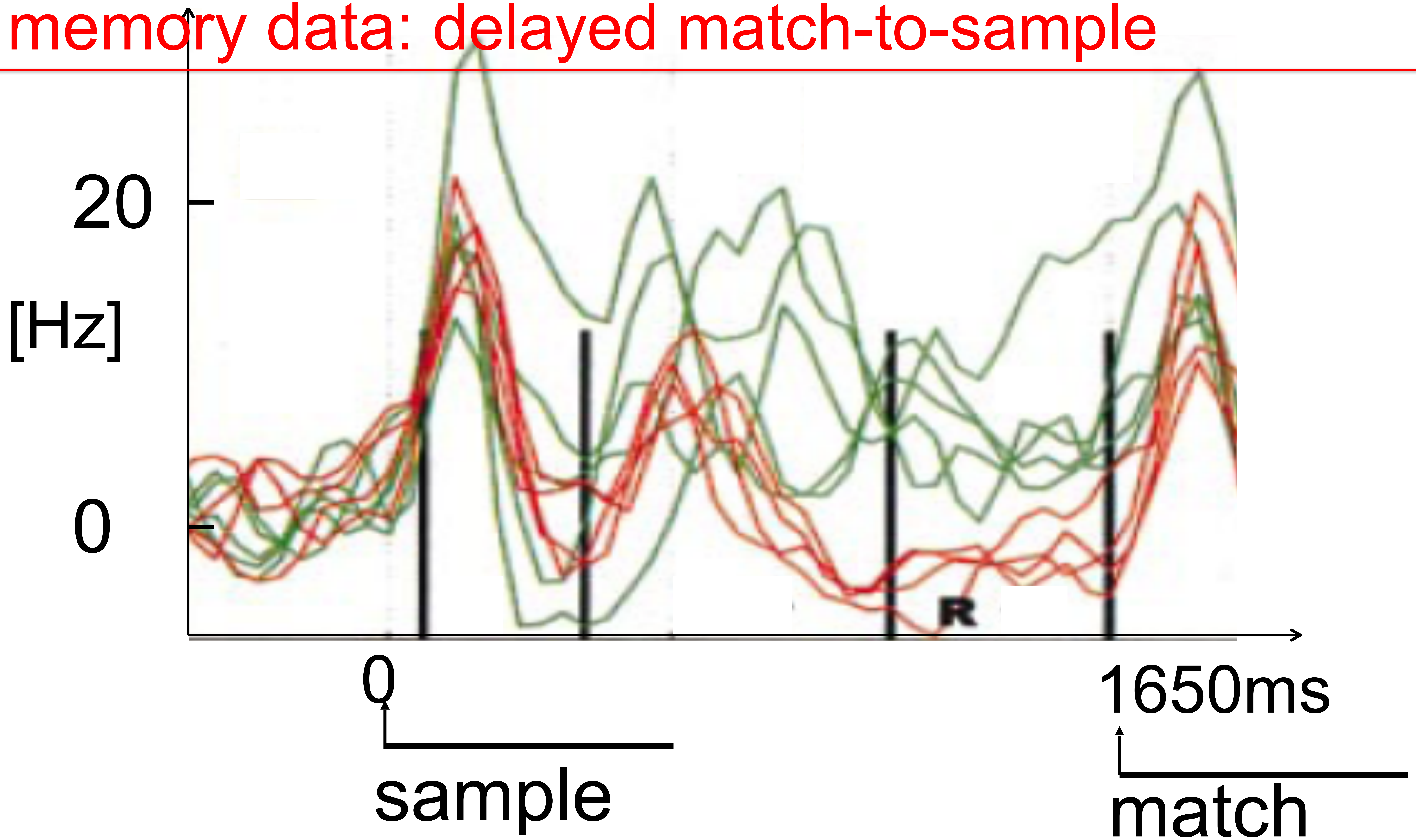


5. memory data: delayed match-to-sample



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

5. memory data: delayed match-to-sample

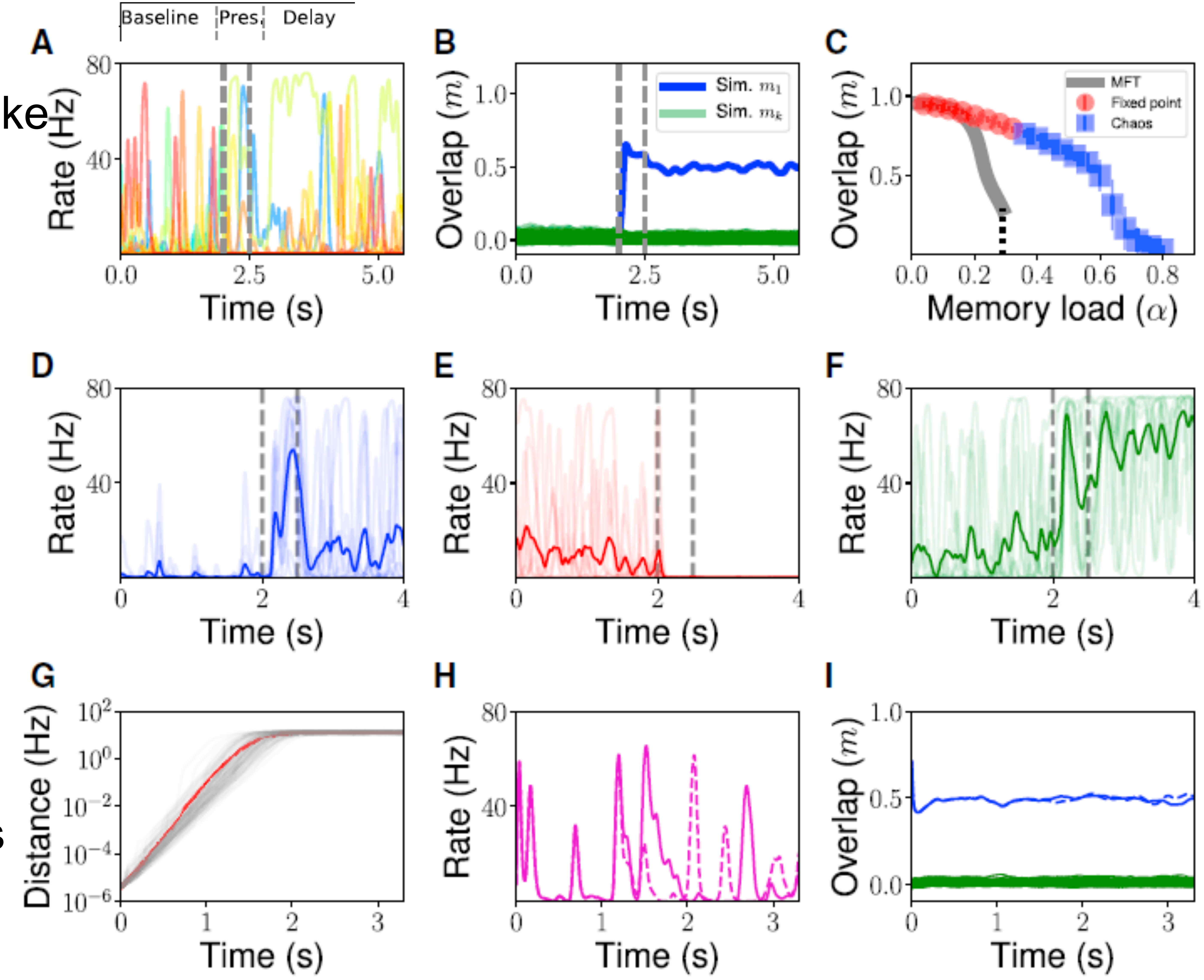


Rainer and Miller (2002). Timecourse of object-related neural activity in the primate prefrontal cortex during a short-term memory task. Europ. J. Neurosci., 15:1244-1254.

A: examples of chaotic-like Firing rates.
B,C. Overlap is stable

D-F:
Three sample neurons

G: 1-overlap (distance)
H: Same neuron, two different initial conditions
I: overlap, two different initial conditions



U. Pereira and N. Brunel, Attractor Dynamics in Networks with Lear
Inferred from In Vivo Data *Neuron 2018*

Memory in realistic networks

-Mean activity of patterns? → can be low
-Asymmetric connections? → possible
-Better neuron model? → possible
-Separation of excitation/inhibition? → possible
-Low probability of connections? → possible

Attractor Memory model

- Abstract concept!
- Influential!
- General!
- Neural data!
- consistent with spikinbg variability!

5. attractor memory in realistic networks

Memory in realistic networks

-Mean activity of patterns? → can be low
-Asymmetric connections? → possible
-Better neuron model? → possible
-Separation of excitation/inhibition? → possible
-Low probability of connections? → possible

Attractor Memory model

- Abstract concept!
- Influential!
- General!
- Neural data!
- consistent with spikinbg variability!

Outlook to final week:

- Neural manifolds
- Low-rank connectivity

6.5 Summary:

From a biological perspective, we need to consider several modifications of the Hopfield model.

- The stored patterns should be low-activity patterns.

- Neurons are not ‘spins’ with values +/-1, but send out spikes that are better described by values 0/1.

- After a spike there is a refractory time or at least a reset

- A given neuron sends out either excitatory or inhibitory synapses, but normally not both.

- Connections are not symmetric

All these modifications can be implemented in attractor networks. Moreover in delayed-match to sample tasks, activity traces suggest vaguely attractor dynamics.

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.
- U. Pereira and N. Brune (2018) I, Attractor Dynamics in Networks with Learning Rules inferred from In Vivo Data Neuron 99, 227–238