Computational Neuroscience: Neuronal Dynamics of Cognition

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

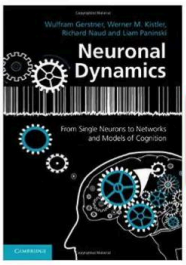
Attractor Networks and Generalizations of the Hopfield model

Wulfram Gerstner

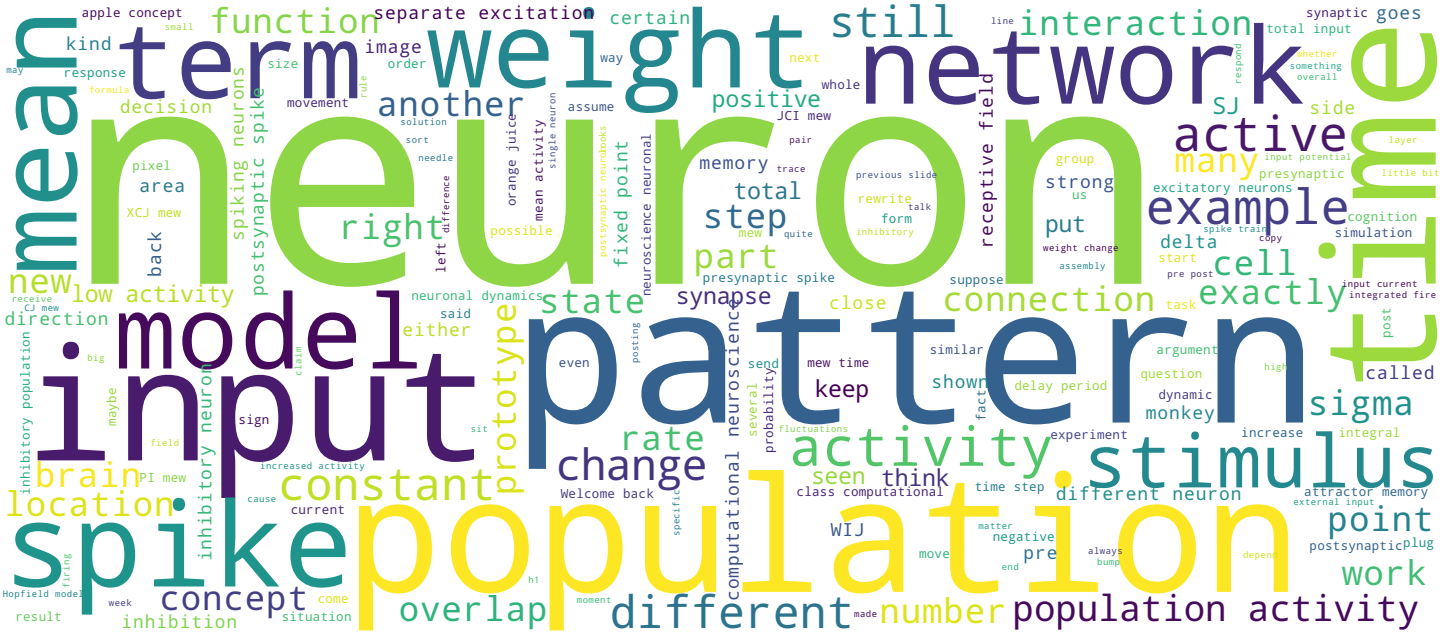
EPFL, Lausanne, Switzerland

Reading:
NEURONAL DYNAMICS
- Ch. 17.2.5 – 17.4

Cambridge Univ. Press



1. Attractor networks
2. Stochastic Hopfield model
3. Energy landscape
4. Towards biology (1)
   - low-activity patterns
5. Towards biology (2)
   - spiking neurons

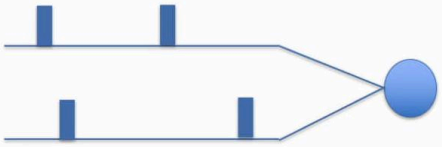


Search MOOC

Video

EPFL

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)

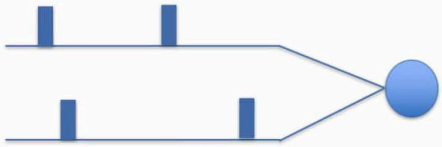
Welcome back to the class computational neuroscience neuronal dynamics of cognition. We have been looking at attractor memories. But on the way towards more biological realism, there's still a few steps to be taken. In particular, you still have the problem that neurons are binary plus minus one. So if you think of the input to neuron and I, the total input, and if you take SJ as plus or minus one. And this is like you have positive pulses coming in and negative pulses coming on. Coming into a posting up at neurons, and that's not very, very realistic. In fact, real neurons have binary variables. They have spikes short pulses that are on or absent. So a better description would be binary input in the form of zeros and ones.

Notes

Summary

0m 05s

5. attractor memory with spiking neurons



Total input to neuron i

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

$$= \sum_j \omega_{ij} \cdot (2\sigma_j - 1)$$

$$= \sum_j \tilde{\omega}_{ij} \sigma_j - c_i$$

$2\omega_{ij}$

$$\sum_j \omega_{ij} = \sum_j \sum_\mu \frac{1}{N} p_i^\mu \cdot p_j^\mu$$

$$= \sum_\mu p_i^\mu \left(\frac{1}{N} \sum_{j=1}^{N} p_j^\mu \right)$$

- rewrite binary state variable:

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

$$S_i = 2\sigma_i - 1$$

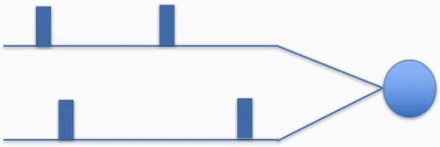
So let us try to rewrite the equation. I just replace s by sigma. Let's propose a combination that si is (two sigma I), minus one. Let's check this. If I plug in sigma equals one. I can now put 1. If plug in sigma equals zero, I get an output of minus one. So this is just a rewrite, which works in both directions. But let's plug this in, so I replace the SJ by two sigma j minus one. And I keep the rest. So I have sum over J WI J, two sigma one, two sigma j minus one. I can split this up, I can say this is sum over J WI J Tilde sigma J minus some constant CI. What is this WIJ Tilde is just to WIJ. What is this neuron specifically constant CI. Well, ci is some over JWIJ. Now let's suppose we have the standard Hopfield weights for patterns that have mean activity 50%. So the wij is PI-mew PJ-mew with an appropriate constant, say one over n. And for each, the weights have to sum up out some of the mew, and they're still have sum over J because I consider this some JWJ, which I can rewrite as sum over J one over NPJ mew times. PI-mew, sum over mew. If the mean activity, if a sum of all neurons in the network. If the mean activity is exactly zero, not just expected to be zero, but exactly zero, then this constant would disappear.

Notes

Summary

0m 56s

5. attractor memory with spiking neurons



Total input to neuron i

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

$\downarrow$

σ_j

Separation of excitation/ inhibition

- rewrite weights:

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

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

$$b = 0$$

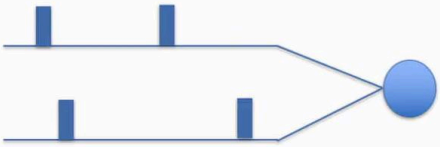
Otherwise I would have to keep this constant. Now the same argument can also be applied for low activity patterns. CI is some of the JWIJ for low activity patterns, I will take ci mew minus BCJ-mew minus A. Sum over mew, sum over J, with some constant in front. Same argument as before. I take sum over J. I take XCJ mew minus a. And I have a term. Some over JCJ mew minus a. The activity of these patterns, is exactly equal to A. The exactly equal to 10%. These term goes. Now with these new variables, zero and one. It's much easier to interpret input as spikes, so I will have many different time steps, and the each time step, there's a value of one, or else there is zero. But this was the first step, we went from s, to binary variables, sigma. Now the second step is you would like to separate excitation and inhibition. Note here that weights can be positive or negative. So if I have XCJ-mew equal to one, positive term here. If I have XCJ-mew equal to zero. I have a negative term. And this is true even if I said the constant be equal to zero. I would still have positive and negative weights, which is biologically not plausible because in biology, neurons that send out positive weights are called excitatory neurons, and these are different neurons than those that send out negative weights, which are called inhibitory neurons.

Notes

Summary

3m 14s

5. attractor memory with spiking neurons



Total input to neuron i

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

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

Separation of excitation/ inhibition
- rewrite weights:

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

$$b = 0$$

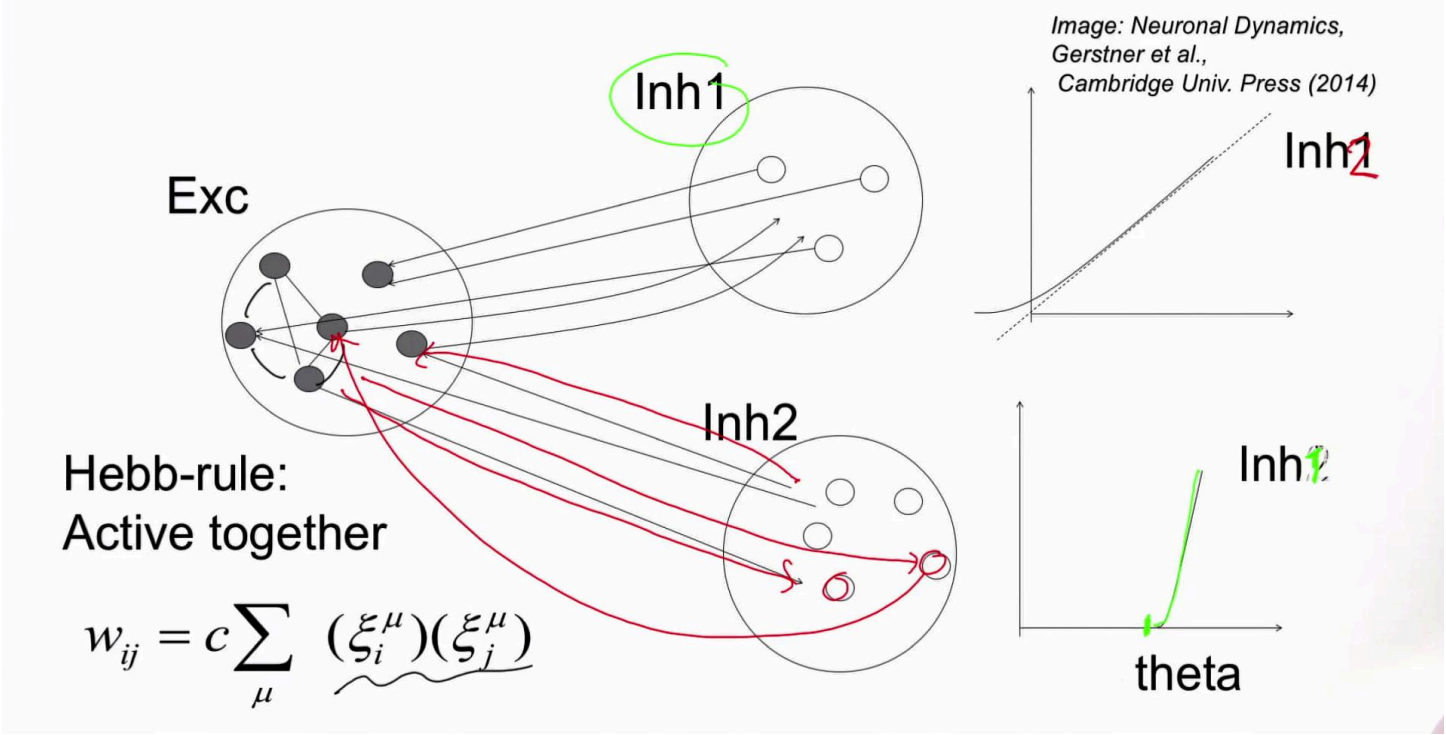
Now, if you do a little calculation, HI of T, and take this be equal zero situation, c times sum over JCI-mew times, CJ-mew, minus A, and here for simplicity, I've dropped the sum over mew times sigma J i made the replacement under the conditions discussed on the previous slide, you see that we can separate this, I can pull out the term with a. And then I have some over JCI-mew time's sigma J of T, with a minor sign. And I would keep the other term, some over JCI-mew CJ-mew sigma JFT. Now here, this guy is 01. This guy 01 for these rates are either 01, which means they are never negative, they're always positive or zero. Same thing here. If a minus side. A is a positive constant say 10% 0.1 CI-mew is either 0 or 1, and minus side. So this is like separate inhibition. These are neurons that just send out inhibitory spikes. However, it's still the same neuron. So we still have not succeeded to separate excitation and inhibition.

Notes

Summary

5m 16s

5. Separation of excitation and inhibition



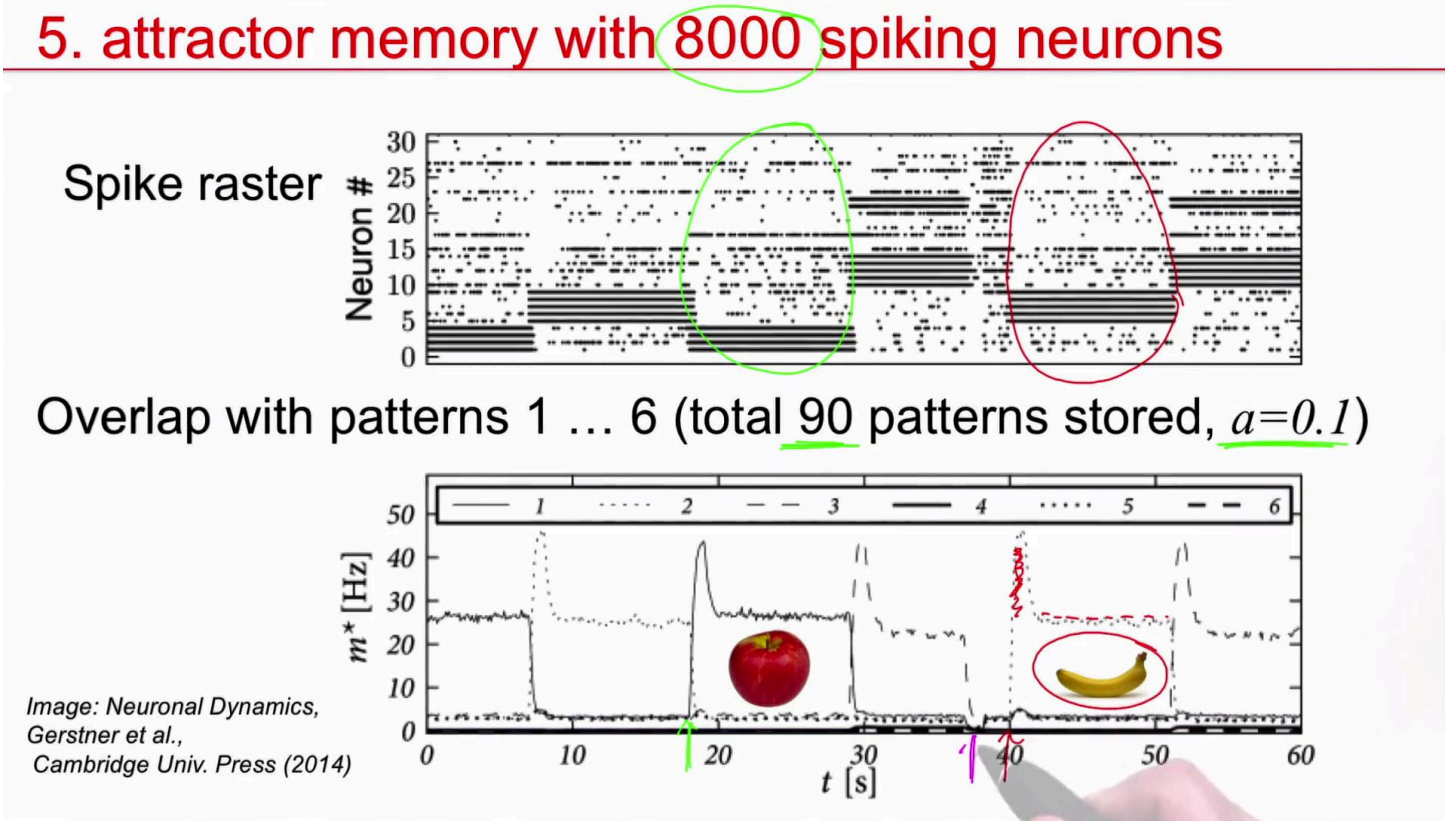
However, let's just assume that we have two populations of neurons. Those excitatory and inhibitory neuron whereby excitatory neurons have weights, which are just given by this formula here, which means each term here is either 01. So overall, I'm excited to weights, so I have positive weights and the term that appeared with a negative sign is embedded here as interaction with inhibitory neurons, which, in turn, send information back to the excitatory neurons. So this way, the two terms we had on a previous slide, are separated out into two different populations. Now there are two of these inhibitory populations on the previous calculation we needed one, which is nearly linear or main linear. So, this would be inhibitory to in our situation. And then I would have another one inhibitory population one, which controls the overall activity of this network so that never more than say 10% of the neurons can be active at the same time to need a very steep input output curve, which has a sharp threshold theta. Okay, so what we have seen, we have seen that I can go from plus minus one neurons to spiking neurons. I can go to very little activity, and I can separate excitation, and inhibition.

Notes

Summary

6m 58s

5. attractor memory with 8000 spiking neurons



And here's known application. So we have loaded, several different patterns. In total, 90 patterns. Each of these patterns is low activity. So if I have 8000 neurons total that means, each pattern involves roughly 800 neurons, and you see here, just 30 neurons out of the thousand. And you see that during this pattern, the first four neurons are active, but there are also many other neurons that are active, some neurons that are not active and there are some new ones that are active at low rates. So here we give a partial input of this pattern thing of this concept. A concept, could be the apple concept. And then the input is removed. The network is in the free recall, and you see that it thinks about the apple the apple concept is rich retrieved. The apple concept is represented during all this time. At a later time, we give a different input, say it's a banana input. It's a different subset of neurons, that's active, the input is short, and it's only during this short time period. But the activity persists for a very long time. So, the network, loads a memory, and that it stays in the network, in the form of working memory. At this location here.

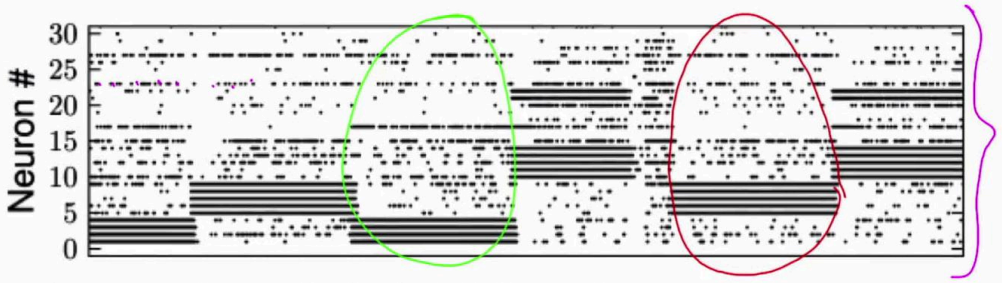
Notes

Summary

8m 27s

5. attractor memory with 8000 spiking neurons

Spike raster



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

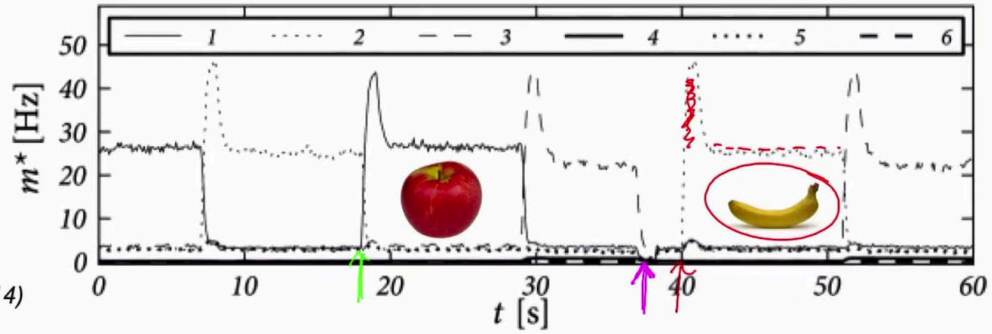


Image: Neuronal Dynamics, Gerstner et al., Cambridge Univ. Press (2014)

We gave a strong input to the inhibitory neurons, and the net result is that the network falls back in a spontaneous state where none of the learn patterns is active. Some neurons are active and the others are not active. So, this network of 8000 neurons total can store, many patterns. Each neuron is involved in the retrieval of many of these patterns, so the concepts are distributed across different neurons, and each neuron participates in several of these concepts. The whole thing works with spiking neurons that emit spikes, occasionally, from time to time.

Notes

Summary

9m 49s

5. attractor memory with spiking neurons

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?

So what you see here is, yes it's possible to have memory with spiking neurons. We can do it with low activity of patterns, yes, we can separate excitation In addition, and this is a model with spiking neurons of the integrated fire type. We have asymmetric weights. We have flow connection probability. All this is possible. The question now is, how does this relates to neural data?

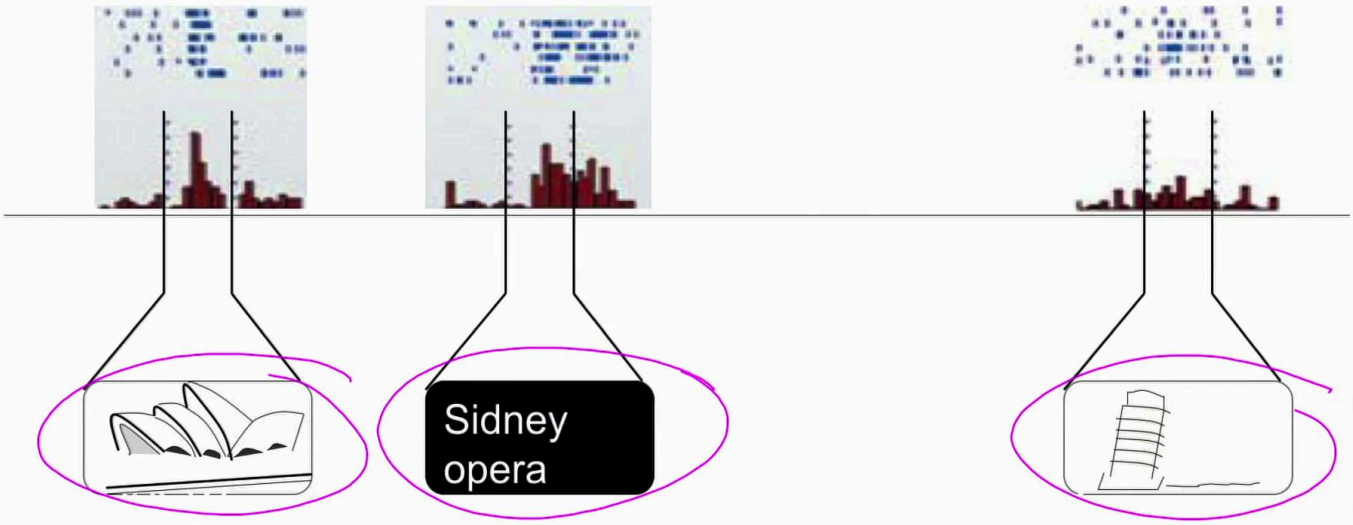
Notes

Summary

10m 32s

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.

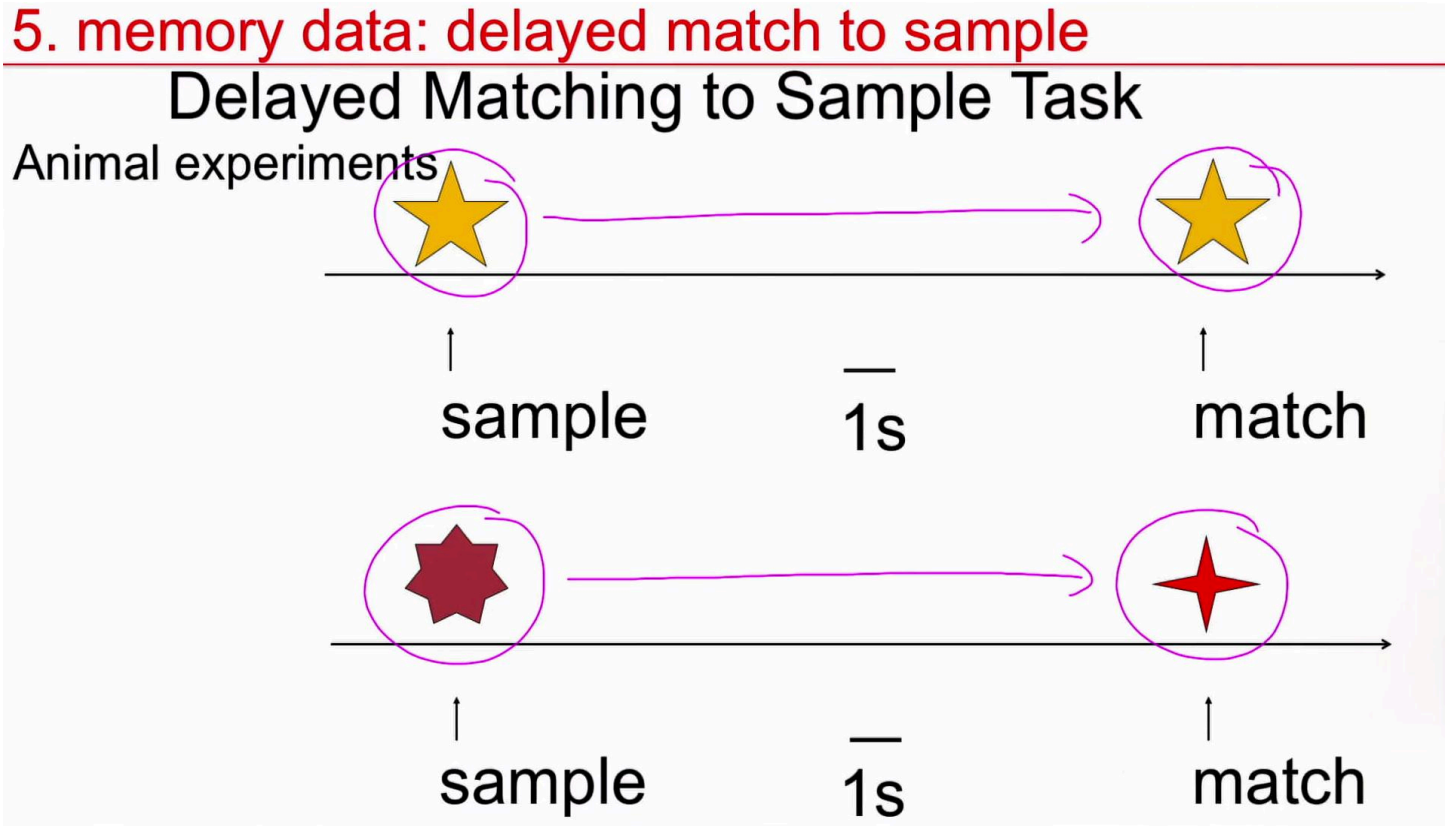
Let me come back to what we saw last week. We said that there are neurons in a certain region of the human brain in the hippocampus or nearby. And these neurons, the same neuron response, whether the image is dead of the opera, or whether the Word says, Sydney Opera. Here it's not several neurons, but it's the same neuron over several repetitions. So in total. This image was shown six times. This image was shown this six times. Now if you show a different image, then these neurons, do not respond as strongly. Note that the stimulus, the information is only given at the beginning and the activity continues after the stimulus has been removed. So this is an indication, potentially of happy and assembly in the human brain.

Notes

Summary

10m 59s

5. memory data: delayed match to sample



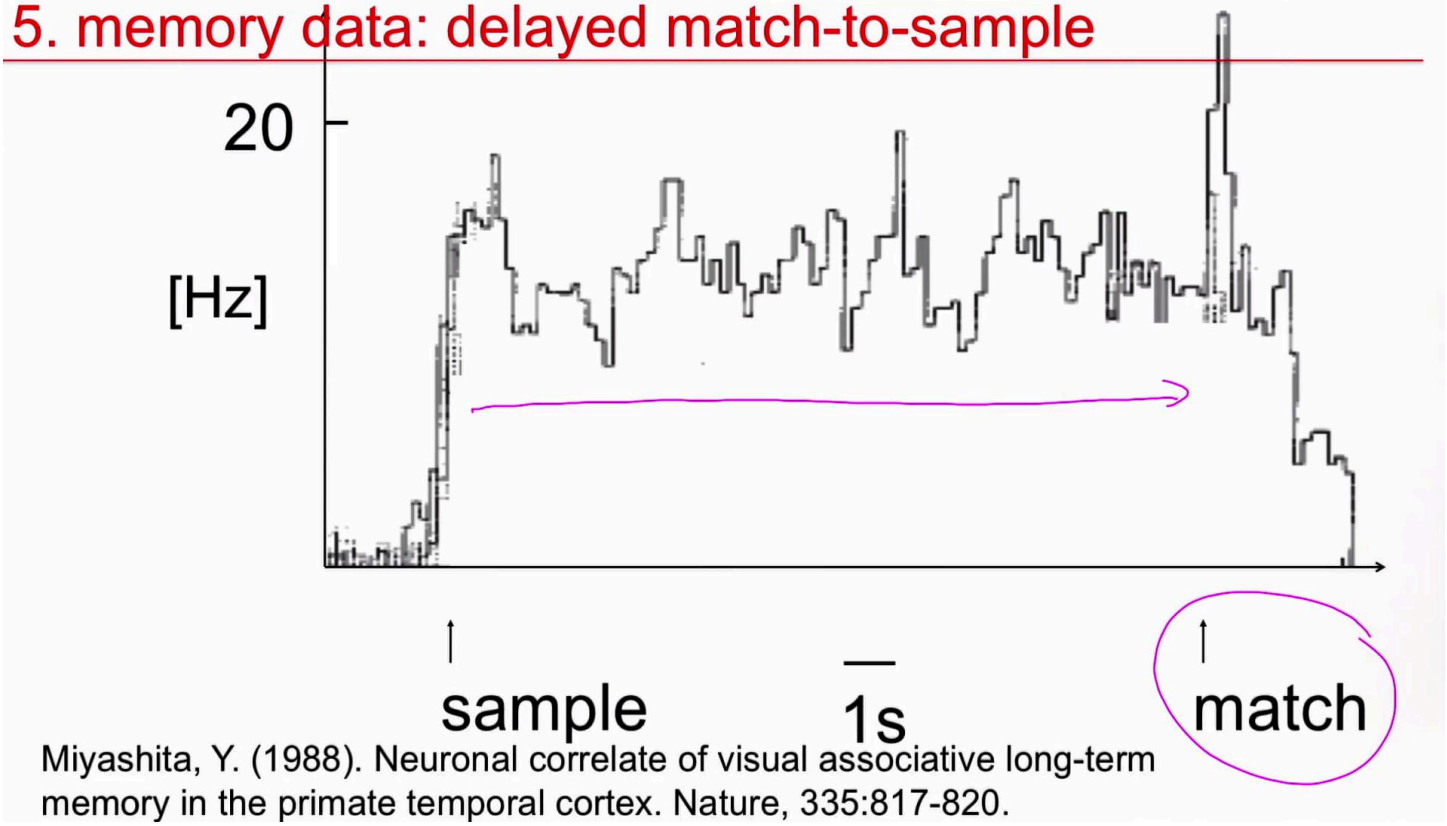
You have also looked at animal experiments with monkeys, recording the monkey brain. Now, a monkey you cannot simply ask, what do you see right now. But what you can do is you can involve the monkey in a task. For example, you can present a first stimulus. And then, one second later. Two things later five seconds later, a second stimulus is shown. And if it's the same stimulus. The monkey presses the right button and gets the orange, orange juice. And if the stimulus is not the same. It has to press the left button in order to get the orange juice. Now, in order to solve this task. The brain has to memorize the activity for the delay period, where no stimulus is given.

Notes

Summary

11m 59s

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.

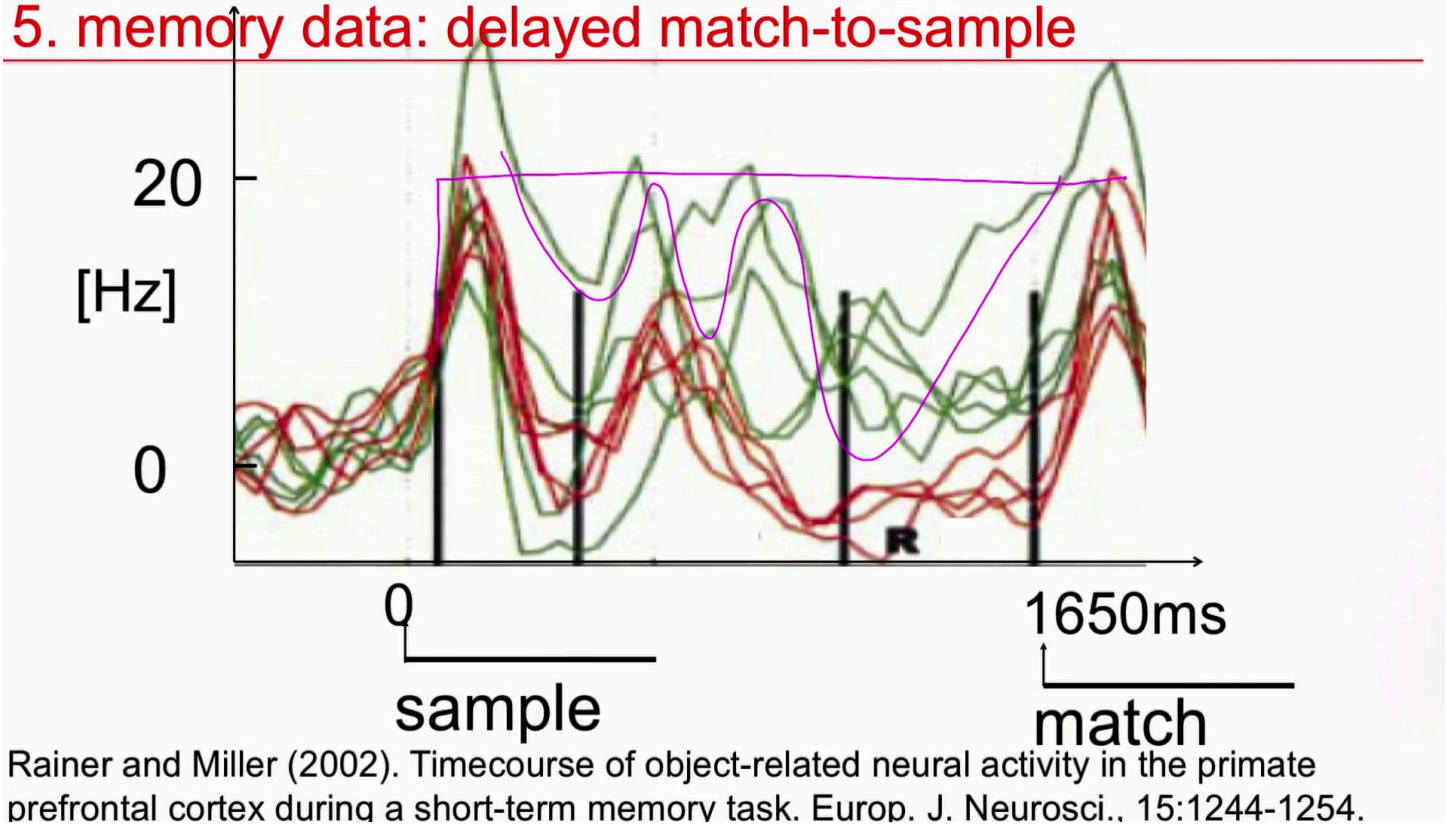
And this is what the neurons are doing during the delay period when the stimulus is shown, then some humans show increased activity, and this increased activity remains during the whole delay period until the final stimulus is shown, and the monkey receives the orange juice, or some other we work. This looks like a very convincing curve. It's in primate temporal cortex.

Notes

Summary

12m 50s

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.

People did similar experiments in other areas, for example in their prefrontal cortex. You see sort of the same kind of pattern. However, it's not the kind of thing you would expect from attractive networks. It's not sitting at a stable value. It's in fact going up and down quite a bit. So, we don't claim that this is a sign of attractor networks. However, it’s very close. It's not too bad. Somehow memory is kept in the brain and this is indicated by increased activity.

Notes

Summary

13m 19s

5. attractor memory in realistic networks

Memory in realistic networks

-Mean activity of patterns?
-Asymmetric connections?
-Better neuron model?
-Separation of excitation/inhibition?
-Low probability of connections?

Attractor Memory model
- Abstract concept!
- Influential!
- General!
- Neural data?

Simulations as well as the theoretic analysis shows that the concept of attractor memory can be translated into rather realistic networks. Realistic networks here means, we can change the mean activity of patterns to make it more plausible very low activity, only a small fraction of neurons is active at the same time. We can go from symmetric connects asymmetric connections. That's not a problem. We can use a better neuron model. For example, integrated fire model, but you could also use a Hodgkin Huxley newer and model. We can separate excitation and inhibition in different types of neurons. We can work with a low probability of connections. So, there's a big way towards biology that has been claimed. Nevertheless, abstractor memory models, remain a rather abstract concept. They have been very influential. They are general, they can be adapted in different forms. But we don't claim that neural data, uniquely says, this has to be an attractor memory for memory for memory retrieval, but it's not too bad.

Notes

Summary

13m 58s

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.

It's pretty close. So many researchers have contributed to this field of attractor memory networks. I would say right after John Hartfield published his first paper on attractor memory no models in 1982. , quite a few theoreticians took up this topic, and try to extend it towards biology. Along the lines, I just sketched today.

Notes

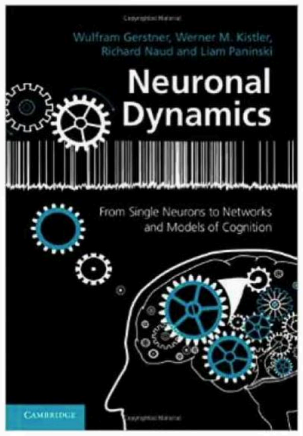
Summary

15m 11s

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

As usual, you find documentation, online under neuronaldynamics.efl.ch.

Notes

Summary

15m 38s