2 Books

Theoretical Neuroscience: How questions?

- How to generate spikes as a single neuron?
- How to build SNN?

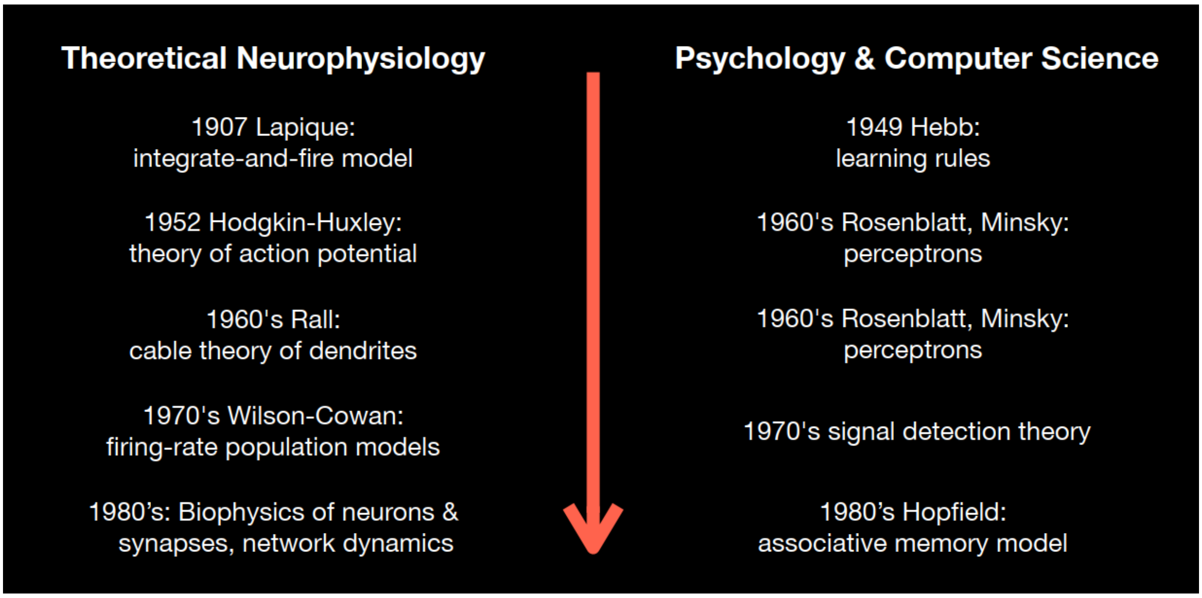
Principles of Neural Design: Why questions?

- Why do we have a brain?
- Why do we have a bigger brain?
- Why do we have different neuron types?

History Before 1988

Sejnowski et al., 1988: a famous review paper by 3 notable scientists.

In 2022 Fall, we covered 8 topics of these 10: except for cable theory and Wilson-Cowan model. You can read these on *Theoretical Neuroscience*



(figure from class slides)

There has been a lot progress after 1988.

In exactly 1988, Hinton and colleagues developed error-back-propagation.

Principles/Axioms of Neuroscience?

In physics, we always have some principles, like Newton-3-laws, thermodynamics-3-laws, Maxwell equations, Schrodinger equation, equal probability principle.

But in biology, we don't have such things. Maybe the only thing can be called principle in biology is Darwin's **Nature Selection**.

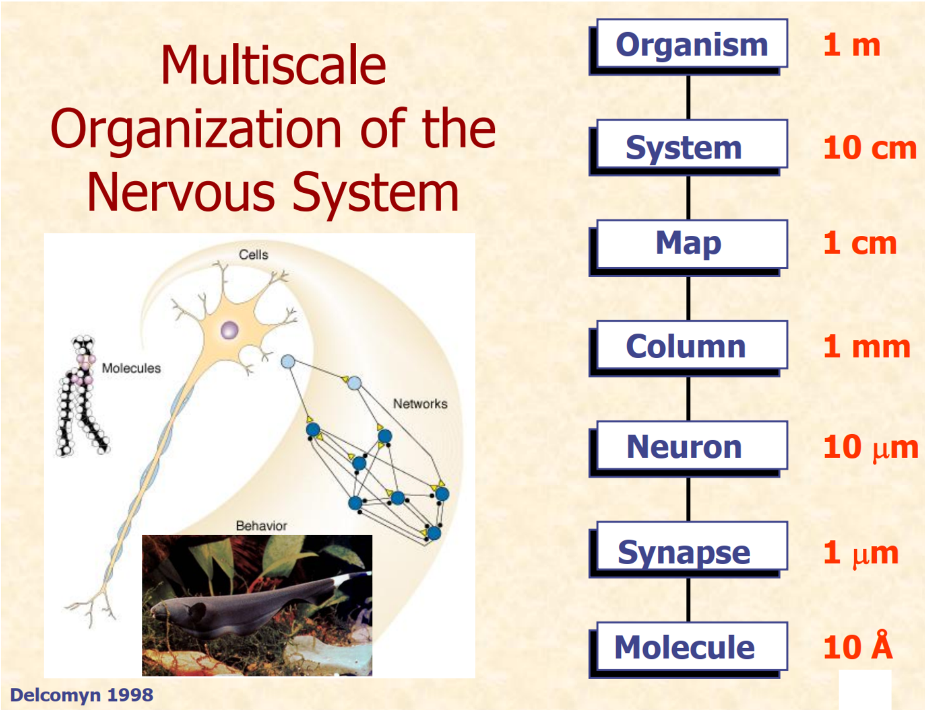
Goals of Neuroscience?

1. Understand how the brain works

More Is Different

[More Is Different: Broken symmetry and the nature of the hierarchical structure of science](): a famous paper by Philip Anderson in 1972

This "More Is Different" thing also plays important role in Neuroscience.



(figure from class slides)

David Marr vs Henry Markram

Marr

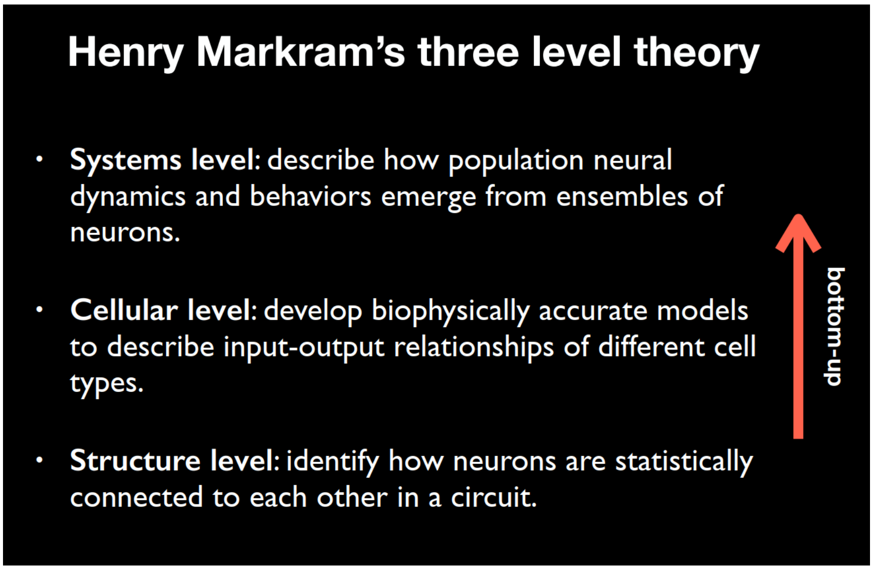
Computational theory	Representation and algorithm	Hardware implementation
What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?	How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?	How can the representation and algorithm be realized physically?

Figure 1–4. The three levels at which any machine carrying out an information-processing task must be understood.

(figure from Vision)

In the past, algorithm needs a lot math. But now, just 炼丹。

Markram



This approach **succeeds** in physics (mechanics, thermodynamics, electrodynamics), but it **seems** to fail in neuroscience.

> Think about the AlexNet, the ResNet or the transformer. You know every connection between each pair of artificial neurons. But can you tell me how the transformer works?

PS: In 2009, Henry Markram claimed that he could build human brain in ten years. But he failed, along with the blue brain project. Blue Brain Project and Human Brain Project were highly related to each other and both got a lot of criticism. See this paper for the controversy.

Think

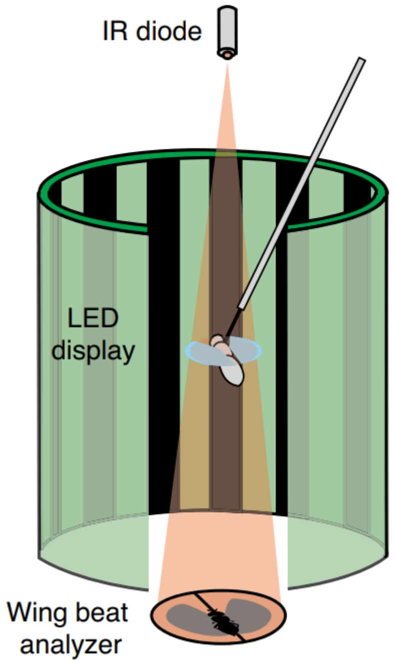
A friend of mine said that: both Marr's and Markram's theory has the biggest difficulty from the 2nd level to the 3rd level (from algorithm to hardware and from cell to system), do you agree with him?

Example: Vision of Fruit-fly

An experiment

Original paper in 2007.

PS: we want to use dragon-fly one day!

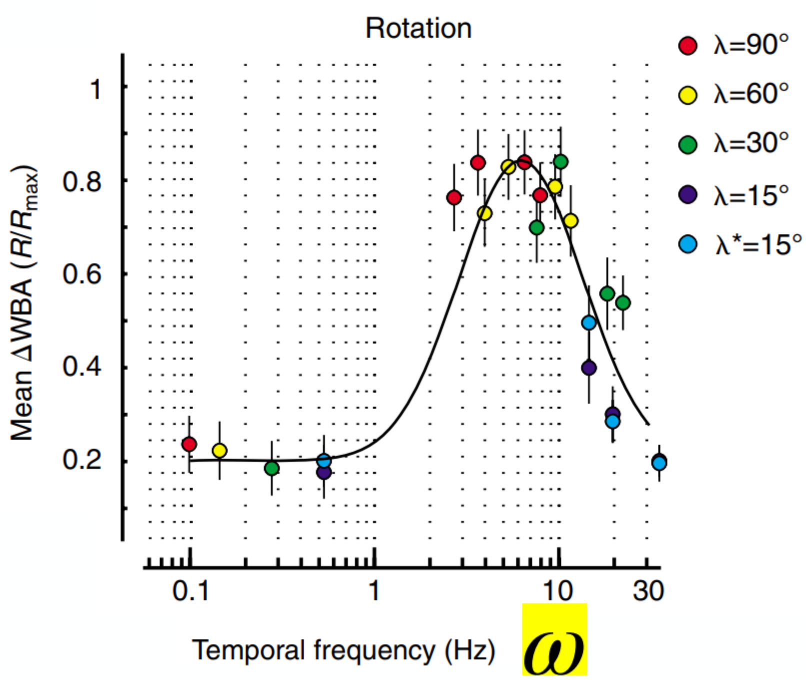


The stimulus is:

$$s(t) = sin(wt - kx)$$

The response is wing beat.

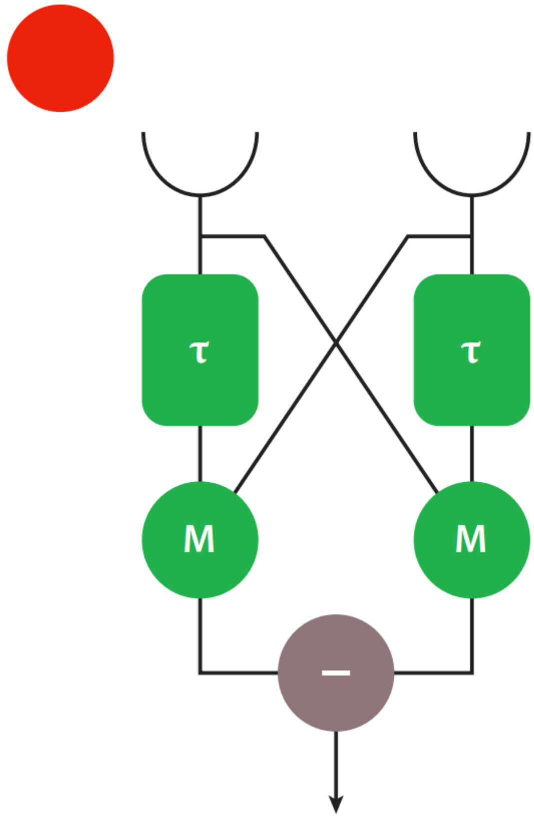
The response-stimulus plot is:



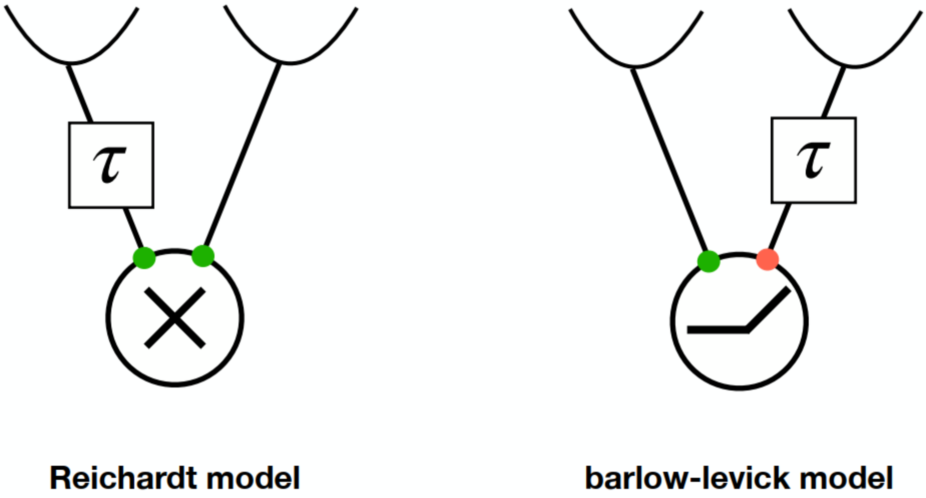
A theory

[Original paper](#) in 1956.

Attention: this model is for 2 near photo receptors in the retina, not for left eye and right eye!!!

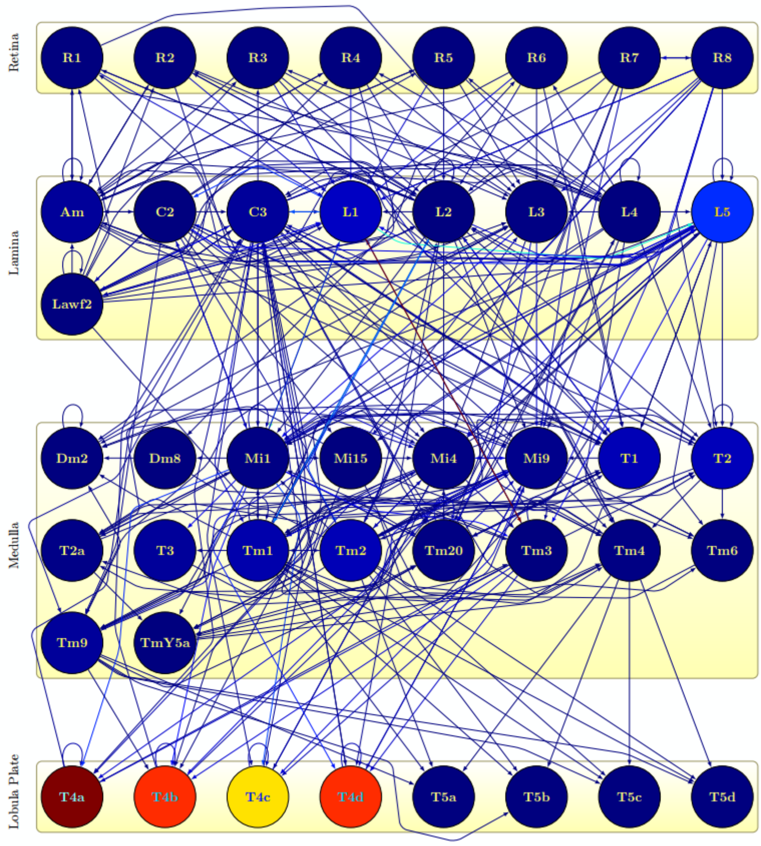


Use ReLU instead of multiplication.



Open the black box

Try to open the black box in Drosophila.



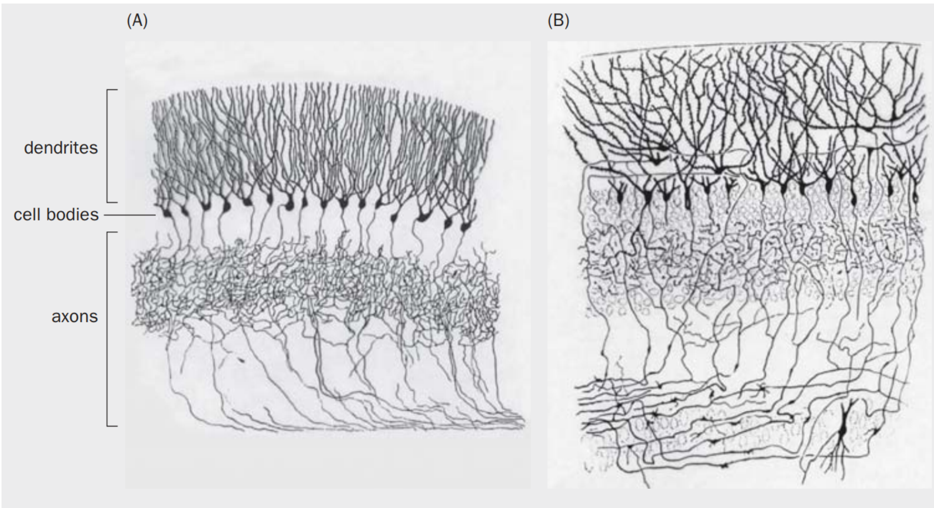
Think

Does this example have David Marr's 3 levels?

The Debate between Cajal and Golgi: the Start of Neuroscience

Golgi: axons form a big net. (A)

Cajal: axons belongs to each neuron. (B)



(figure from class slides)

PS: Golgi didn't mean neurons form a big net, he meant axons form a big net.

PPS: in recent years, jellyfish's axons are found to be like Golgi's theory.

Outline

Here, you already know the basic structure of a neuron. Below we will talk about the brain through a top-down approach, which is adapted from CH1-CH3 of *Principles of Neural Design*.

- Why do we need a brain?
- Why do we need columns (similar neurons located together)?
- Why do we need axons and dendrites?

Why Do We Need a Brain?

E.coli

E.coli has no brain, but it can do biased random walk and it has memory. (This is real memory! not like "memory" of Hopfield Network)

But no brain has some drawbacks:

- only very short memory
- can't sense more chemicals
- can't control motor more precisely

C.elegans

motor

100 muscles, distributed evenly, generating oscillations.

This oscillation needs a brain.

This oscillation is preserve in many animals. Walk, run is also generated by oscillations.

sensory

They can sense much more things than E.coli or other no-brain-creature, including T, light, chemistry compounds.

PS: some design aspects of their brains

- each single neuron is important
- chemistry > electrical
- favor analogue over digital: digital signals need more energy.
- use stereotype things: they don't want to reinvent wheels too.

Why Do We Need a Bigger Brain?

see CH3 of *Principles of Neural Design*

Why Do We Need Columns?

Minimize Wire

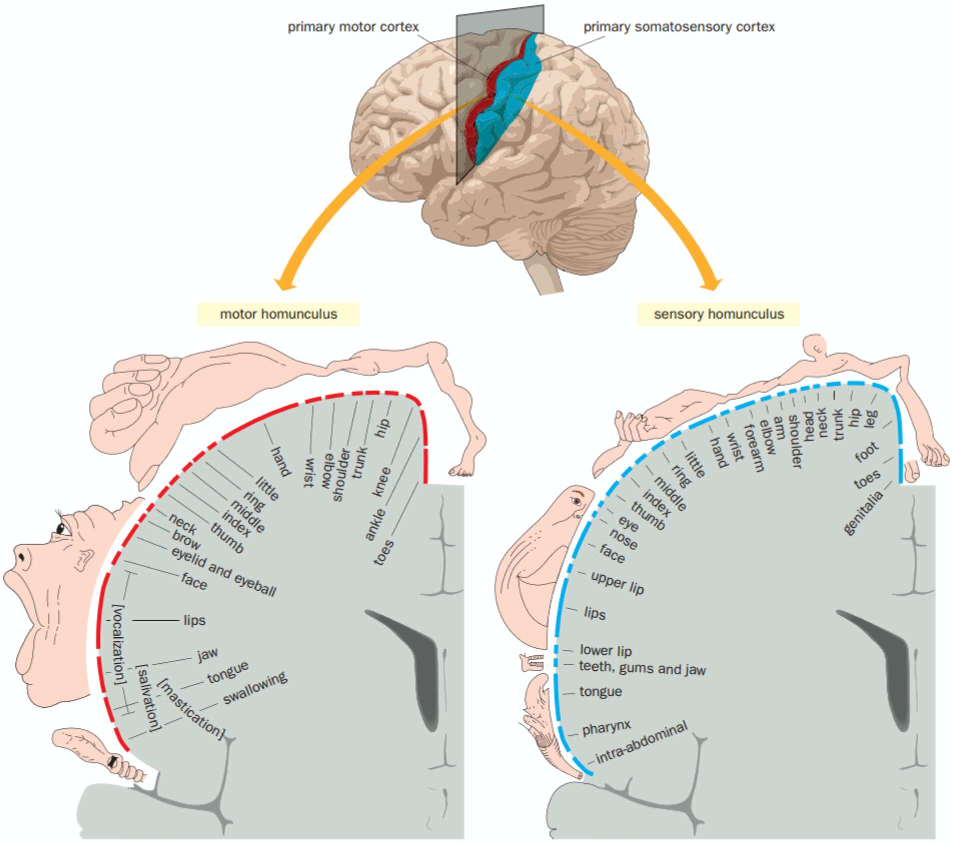
compute locally if you can?

neurons that fire together should locate together?

Some scientists believe these two statements.

Personally, I am not sure, although we have some evidence:

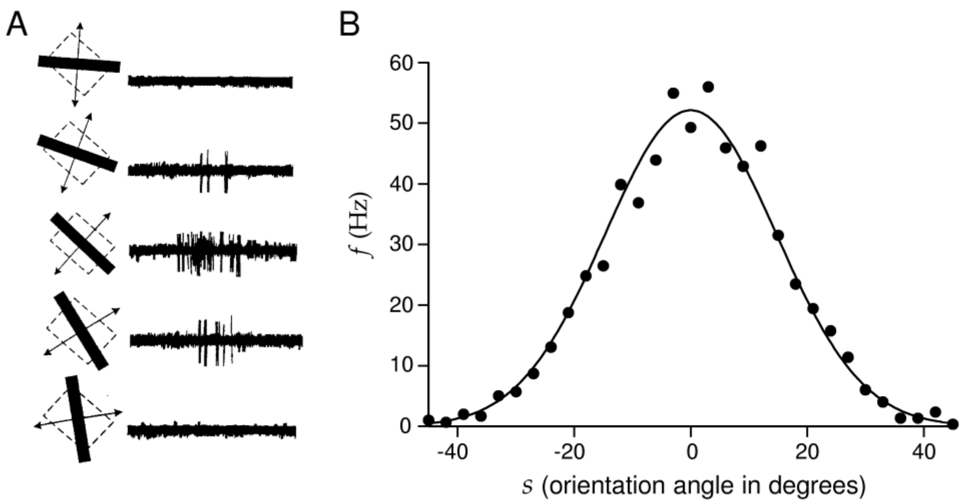
eg1: Map of Penfield (human cortex)



(figure from class slides)

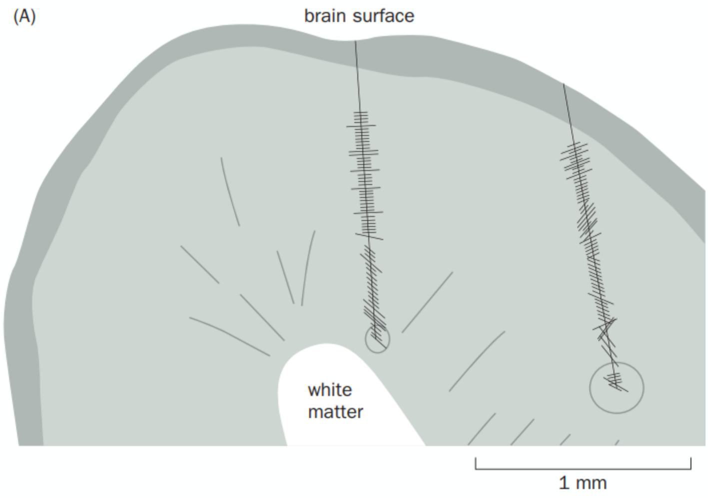
This map is first discovered by Wilder Penfield.

eg2: Map of Hubel and Wiesel (cat V1)



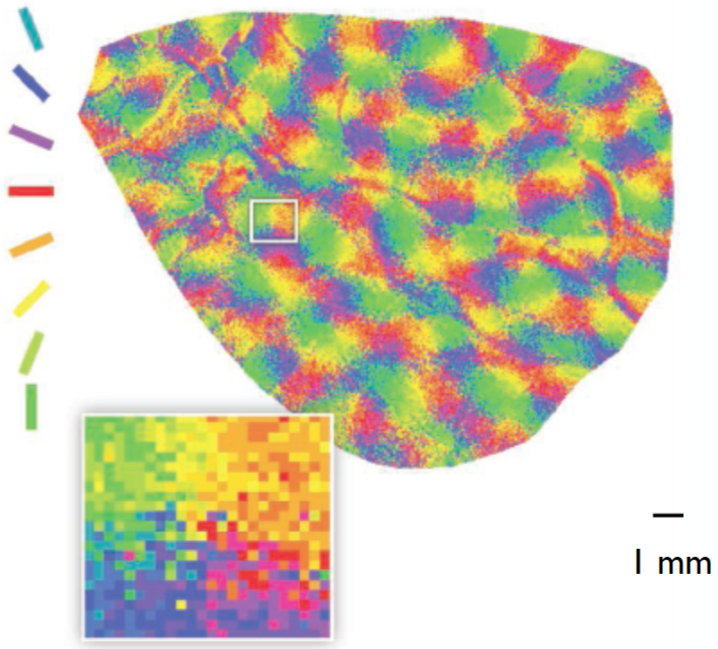
(figure from class slides)

eg3: Map of Hubel and Wiesel (Column)



(figure from class slides)

eg4: uniform distribution vs Gaussian distribution



(figure from class slides)

It will to a homework to discuss that which is better, uniform distribution (salt-and-pepper) or Gaussian distribution (pin-wheel).

Why Do We Need Axons and Dendrites?

Cajal thought that the neurons, the brain can save time, space, and material.

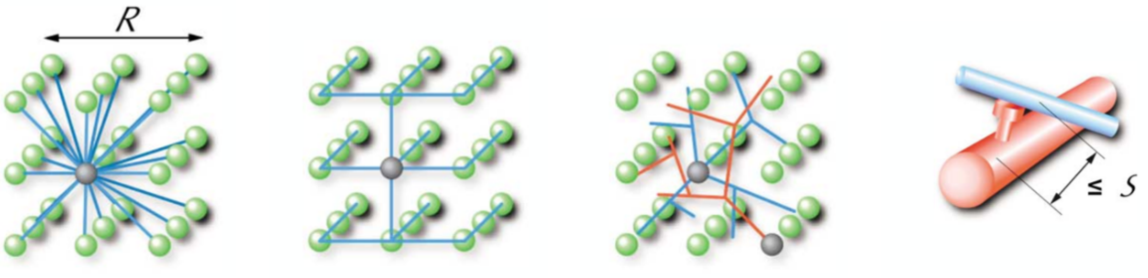
Alan Turing also had similar opinions.

We make 2 assumption

- every 2 neuron must have connection
- minimize the wire

We will see, under the 2 assumptions, we can get results that are in line with reality, so these 2 assumptions are reasonable. If not, we must change our assumptions. (This approach is similar to Newton's, Maxwell's and Schrodinger's).

Under these 2 assumptions, we can make 4 designs



Design I **Design II** **Design III** **Design IV**

(figure adapted from [Chklovskii, 2004])

Think

Is it right to think brain as being designed?

Personally, I think it is not right, although it can give us some ideas and insights for designing an artificial neural network.

The brain is not designed, it just evolved.

This is like "能存活的生物不是最聪明的，也不是最强壮的，也不是最节能的，而是最适合环境的。"

It is true that more clever, strong, efficient creatures are more likely to survive, but not always the case.

Quote from Francis Crick

> Biologists must constantly keep in mind that what they see was not designed, but rather evolved. It might be thought, therefore, that evolutionary arguments would play a large part in guiding biological research, but this is far from the case. It is difficult enough to study what is happening now. To try to figure out exactly what happened in evolution is even more difficult. Thus evolutionary arguments can usefully be used as hints to suggest possible lines of research, but it is highly dangerous to trust them too much. It is all too easy to make mistaken inferences unless the process involved is already very well understood.

Is it right to use Occam’s Razor in biology?

Personally, I think it is not right.

Unlike the above question, we have some experimental evidences indicating that the brain has a lot redundancy, which means that Occam's Razor is wrong in biology.

Quote from Francis Crick

> While Occam's razor is a useful tool in the physical sciences, it can be a very dangerous implement in biology. It is thus very rash to use simplicity and elegance as a guide in biological research. While DNA could be claimed to be both simple and elegant, it must be remembered that DNA almost certainly originated fairly close to the origin of life when things were necessarily simple or they could not have got going.

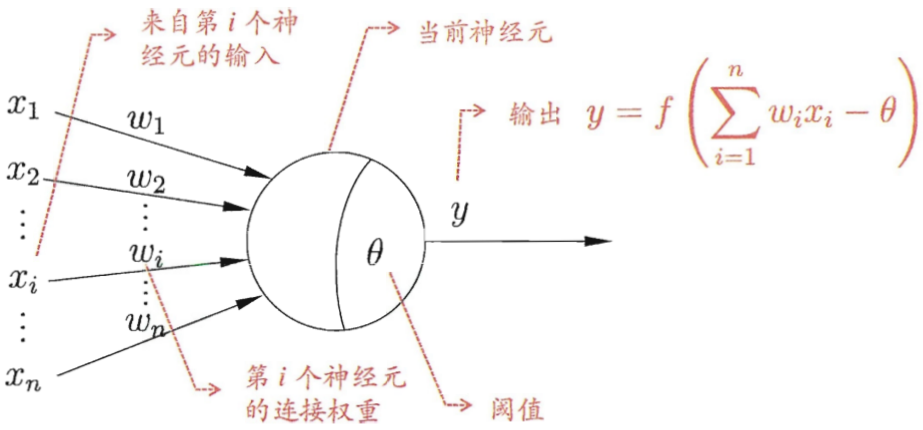
Model of Single Neuron

Below, I will introduce 3 types of single neuron models, from the simple to the complex.

McCulloch-Pitts Model will be used in the Part 2 and Part 3 of the class, Integrate-and-Fire Model will be used in Part 1 and Part 3 and Hodgkin-Huxley Model will only be use in Part 1.

All in all, you must **master** (not just know) these 3 models during this course.

McCulloch-Pitts Model



(figure from 《机器学习》周志华)

$$y = f(\sum_{i=1}^{n} w_i x_i - \theta)$$

$$where\ f(x)\ can\ be\ Heaviside\ or\ Sigmoid\ or\ ReLU$$

M-P model is usually used everywhere in ML and DL, only seldom will they use Integrate-and-Fire Model. When they say they are going to simulate an SNN (Spiking Neural Network), mostly they use IF Model or its variant (see them in 吴思's BrainPy book). They almost never use HH Model because its computational expansivity.

M-P model is also used in Comp Neuro, like the head direction model by Haim Sompolinsky and Kechen Zhang. (Ben-Yishai, 1995) (Kechen Zhang, 1996)

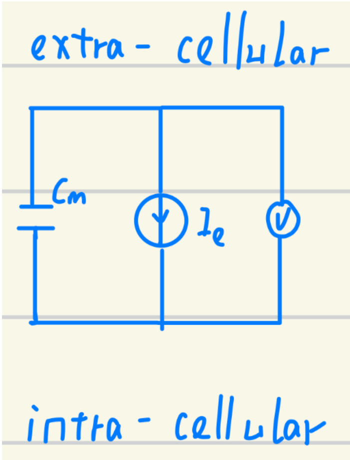
Lapique-Lapique Model (Integrate-and-Fire Model)

Integrate-and-Fire Model was proposed by Louis Lapique and his wife (as he insisted), this is why I call it Lapique-Lapique Model.

All IF Models has 2 imposed equation:

$$\begin{cases} V(t_{spike}^{-}) = V_{threshold} \\ V(t_{spike}^{+}) = V_{resting} \end{cases}$$

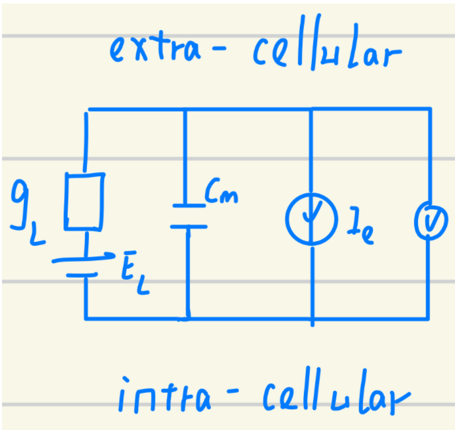
Perfect IF Model



Looking at its circuit diagram, you can easily write an ODE:

$$C\frac{dV}{dt} = I(t)$$

Leaky IF Model



Looking at its circuit diagram, you can easily write an ODE:

$$C\frac{dV}{dt} = I(t) + (-g_L(V - E_L))$$

Variant

In this class, we will only use perfect IF and leaky IF.

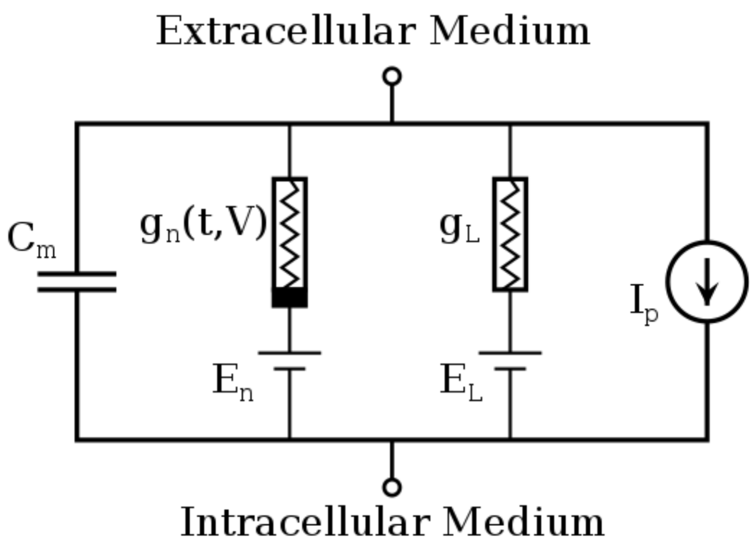
But why always within the class? Why always wait the teacher to teach you? See adaptive, fractional derivative, exponential IF Model here.

Hodgkin-Huxley Model

I will call HH model the most beautiful model in neuroscience.

HH model is a biological plausible model, so, you need to know some biology knowledge about it.

Physical Perspective



(figure from Internet)

In HH model, we use 3 additional ODEs to replace the 2 imposed eqs of IF model. And its first ODE is also gotten from its circuit diagram.

$$I = C_m \frac{\mathrm{d}V_m}{\mathrm{d}t} + \bar{g}_{\mathrm{K}} n^4 (V_m - V_K) + \bar{g}_{\mathrm{Na}} m^3 h (V_m - V_{Na}) + \bar{g}_l (V_m - V_l)$$

$$\frac{dn}{dt} = \alpha_n(V_m)(1-n) - \beta_n(V_m)n$$

$$\frac{dm}{dt} = \alpha_m(V_m)(1-m) - \beta_m(V_m)m$$

$$\frac{dh}{dt} = \alpha_h(V_m)(1-h) - \beta_h(V_m)h$$

(figure from Internet)

There are a few things you need to think:

- What is the meaning of α and β ?
- Why the exponent is $4, 3, 1$ for n, m, h in the first ODE?

Biological Perspective

There are a few things you need to know:

- How to calculate the V_K and V_{Na}? (Sometimes they are called Nernst Equilibrium Potential)
- What are the concentration of Na^+, K^+, Cl^-, Ca^{2+} inside the neuron and outside the neuron? How will they change during the action potential? How much will they change during the action potential?
- What does the $Na^+ - K^+\ pump$ do in the action potential? Is it the same thing as the ion channel?

THINK

Isaac Newton used mass-point-model when he studied the movement of the sun and the earth.

Charles-Augustin de Coulomb used charge-point-model when he studied the force of 2 small balls.

Mr. & Mrs. Lapique, Alan Hodgkin and Andrew Huxley used circuit diagram to model the neuron.

Do you think these three methods are from the same vein?

Meaning of n, m, h

$$g_{K+} = P\bar{g_{K+}}$$

Here, P is probability.

In HH model, n, m, h are all probability, their range is $[0, 1]$.

ODE of n, m, h

$$\frac{dn}{dt} = \alpha(V)(1-n) - \beta(V)n$$

See open and close as 2 states in a Markov Chain, and $\alpha(V), \beta(V)$ are transfer probability.

Resting State

When $\frac{dn}{dt} = 0$, $n = \frac{\alpha_n}{\alpha_n+\beta_n}$, denote it as n_∞

Denote $\tau_n = \frac{1}{\alpha_n+\beta_n}$, we'll get

$$\tau_n \frac{dn}{dt} = -n + n_\infty(V)$$

How to Get α and β ?

Go back to experiments, HH found that the plot of $n_\infty \sim V$ was very like sigmoid.

So they use this function to model (This is straightforward that the don't use other sigmoid function, think of Boltzmann distribution!)

$$n_\infty = \frac{1}{1+e^{F_0-\beta V}}$$

From n_∞, you can get α and β.

Exponent of n, m, h

The plot of $n_\infty \sim V$ is more up-and-left in experiments.

They use exponent to model (This is not so straightforward like above).

They choose 4 for n, 3 for m, 1 for h.

In 1998, structure biologists found that each potassium channel is consist of 4 sub-modules and it will open if and only if all 4 sub-modules are open~

Ion channels

In HH's paper, they only consider sodium and potassium.

Nowadays, some researchers consider more than 10 type of ions, including calcium and chloride.

By the way, C.elegans have potassium, calcium and chloride, but no sodium!

Sensory Neuron in Visual System

We will use visual system to talk about encoding.

We want to investigate the sensory neuron, like Hubel and Wiesel in 1970s.

We will present some stimulus and see the response.

One of the most simple stimulus is white noise. (PS: white noise can be any distribution, i.e., can be Gaussian or not. I will talk about it from the point of stochastic process in the TA class).

We ask 2 questions:

- Q1: What does the stimulus before the spike of a sensory neuron look like?
- Q2: Can we predict the response of a sensory neuron?

Answer Q1: The Stimulus Before the Spike

See Peter's book.

Answer Q2

Functional Derivative

Why using functional derivative? We just want to investigate sensory neurons.

If $r[s(t)] = s(t)$, then...

Calculate Derivative

Just like what we do in physics and ML.

Finally

If we present white noise $s(t)$ with variance of σ^2, we just need to measure $r(t)$ and calculate $\frac{Q_{rs}(-\tau)}{\sigma^2}$, it will be the optimal kernel.

After that, when we present a new stimulus $s_{new}(t)$, we can calculate $r_{new-predicted}(t)$ for it.

D in visual system

- Earlier than Yann LeCun?
  - yes
- Do we have translational invariant D of biological neurons?

- Do we have translational invariant D of biological neurons?
    - yes
- D of retinal ganglion cell
    - use the exp data to fit the parameters.
    - you can also use other function besides the Mexican hat.
- D of simple cell in V1
    - $exp \cdot cos$
    - first layer of CNN like this, black-white-black..., oriented (still a puzzle)

Use constant stimulus

integral of D should be 0.

THINK

- Do we have figs of $r(t)$ vs $r_{est}(t)$
- Seems this D only works for white noise. i.e. we can only use it to predict r for white noise s

From Linear to Non-linear

Method 1: add an activation function. (the same vain as in DL)

Method 2: use more terms in Wiener Series.

Introduction

In the last two lectures, we discussed encoding. Today, we will talk about decoding. Here, encoding refers to how animals use neurons to represent external information (including vision, hearing, smell, taste, touch, and pain, although we only covered vision in our lectures). Decoding refers to how information from neurons is reconstructed into external information.

First, it's important to note: scientists need to decode, Elon Musk's Neuralink needs to decode, but animals may not necessarily need to decode. At least, there is no conclusive evidence yet that animals must decode.

In one experimental paradigm, scientists show monkeys some dots.

Today's lecture will only cover this paradigm.

What we discuss does not go beyond Chapter 3 of Peter Dayan's book.

Moving Dots Experiment

Definition of Coherence

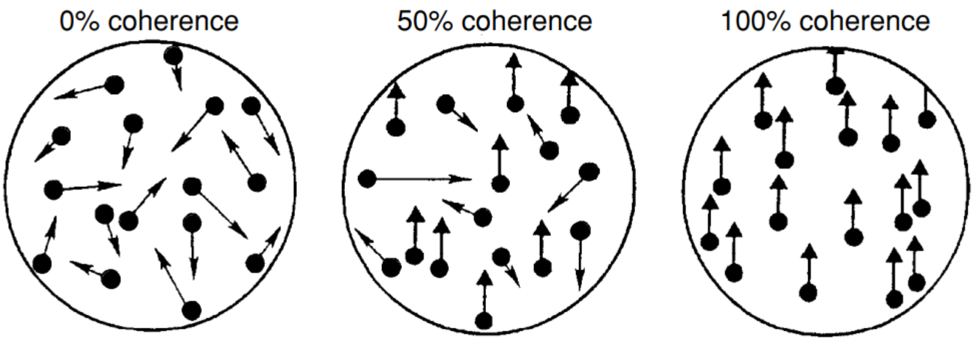


Figure 3.1 The moving random-dot stimulus for different levels of coherence. The visual image consists of randomly placed dots that jump every 45 ms according to the scheme described in the text. At 0% coherence the dots move randomly. At 50% coherence, half the dots move randomly and half move together (upward in this example). At 100% coherence all the dots move together. (Adapted from Britten et al., 1992.)

A single neuron?

You have to find the right neuron.

Professor Wen mentioned in the lecture that a single neuron can achieve the level of population coding, but this comes with a prerequisite: you must find the right single neuron. An experiment in 1992 recorded the activity of a neuron in the MT (middle temporal) area. Obviously, if you record a neuron from a brain area that has nothing to do with vision, you won't see this effect.

Two-alternative force-choice

The experiment conducted in 1992 utilized a two-alternative forced-choice (2AFC) experimental paradigm. In such cases, we can make some mathematical derivations, as seen in Sections 3.1 and 3.2 of Peter Dayan's book.

α and β

The concepts of α and β originate from hypothesis testing and are frequently used tools in computational neuroscience and machine learning.

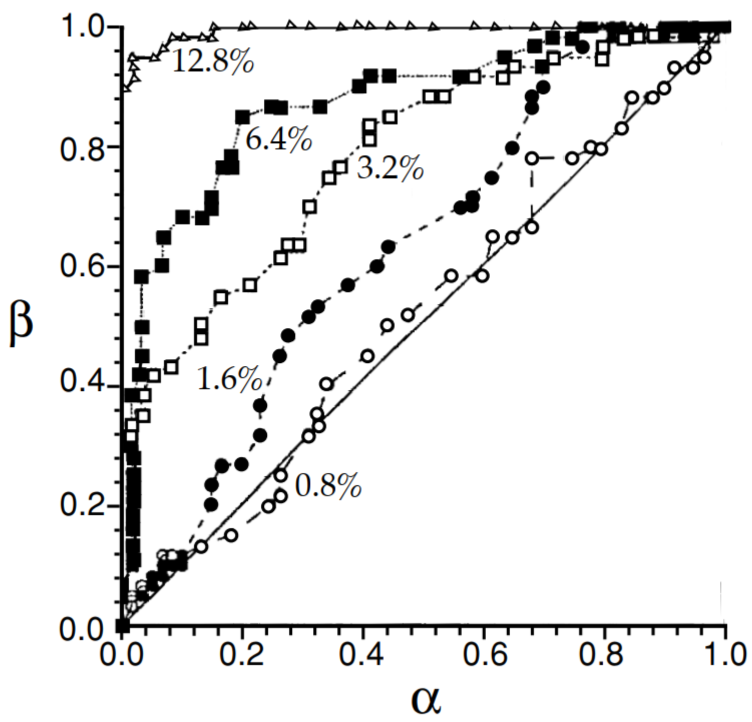
α, often referred to as the significance level, is the probability of rejecting the null hypothesis when it is true (Type I error). β, on the other hand, is the probability of failing to reject the null hypothesis when it is false (Type II error).

Remember:

$$\alpha := P(reject\ H_0|H_0) = P(current\ or\ more\ extreme\ data|H_0) = p - value$$

Receiver Operating Characteristic

ROC说白了无非就是$\beta \sim \alpha$的这张图，计算神经科学和ML也都经常使用此工具。



在理想情况下，ROC下的面积为1

在瞎选情况下，ROC下的面积为0.5

d

我们定义了$d = \frac{\langle r_+\rangle - \langle r_-\rangle}{\sigma}$来区分两个峰是否能够分开，这种思想和物理上的Rayleigh Criterion是师出同源的。

Answer some questions

Q1: Why do you think the two σ are equal?

Answer: You can calculate the standard deviation of the experimental data. If there is a significant difference, you should use different σ to model it. If the difference is small, you can use the same σ.

Q2: What is Noise Correlation?

Answer: To deeply understand, you need to manually work through the formulas in Sections 3.1 and 3.2 of Peter Dayan's book. You will find that Noise Correlation is just one term in the equation.

We want to determine the weights W between the first and second layers in a CNN, where the first layer is stimulated by white noise, and the second layer consists of the retina's bipolar cells. How can we determine this? Experimentally, we know the white noise stimuli, and we can also measure the electrophysiological data of the second layer neurons. We believe that the W that maximizes $I(x, y)$ might be the one that biology actually uses.

The W that maximizes $I(x, y)$ leads us to deduce that in the Fourier space, y components are independent. This can be verified experimentally, and it has been found to be correct.

The biological significance of this is that organisms have adopted such W:

1. W maximizes mutual information, meaning it maximizes the amount of information transferred between the two layers.
2. W allows the second layer neurons to encode different frequencies.

The method of maximizing mutual information between two layers can also be used to train artificial neural networks, but clearly, it is not as fast as backpropagation.

This video provides a popular science overview of specificity coding (which I like to call bijection coding), population coding, and sparse coding. It's very engaging, interesting, and relaxing. We can connect the concepts discussed after single neuron encoding.

The Jennifer Aniston cell is similar to specificity coding, while the work of Hubel, Wiesel, and LeCun is akin to sparse coding. The cyclic matrix mentioned previously also represents sparse coding. The attractor model of working memory represents population coding, as do Hopfield networks and the "remember faster, forget faster" toy model discussed today.

These models initially excite mathematicians and physicists, but the question remains: can they be experimentally validated? For models discussed today, experimental validation is unlikely within the next fifty years, as it is difficult to record synaptic changes.

Certain aspects of the models, such as "remember faster, forget faster," resonate with everyday experiences, but their experimental validation remains to be seen. Many models can produce the effect of "remember faster, forget faster."