NX-414: Brain-like computation and intelligence

Alexander Mathis
alexander.mathis@epfl.ch

Co-teacher: Martin Schrimpf (martin.schrimpf@epfl.ch)

Lecture 1, February 19

Who am I?

- Assistant Professor at EPFL since 8/2020
- Working on methods for behavioral analysis, modeling sensorimotor function & skill learning

My office hours:
Wednesdays, 3 - 4 pm in SV 2811
(*in the weeks when I teach*)

Teaching team

Martin Schrimpf (Assistant Professor)
Abdulkadir Gokce (PhD student)
Merkourios Simos (PhD Student)
Hossein Mirzaei (PhD Student)
Michael Hauri (NX student)

Exercises: 1 - 3 pm on Wednesdays
In the same room

What is this class about?

Learning Outcomes

- Formulate models of brain function, implement those models in Python
- Hypothesize about potential mechanisms that give rise to behavior
- Design models of brain function
- Characterize current models of brain function

Transversal skills

- Set objectives and design an action plan to reach those objectives.
- Demonstrate the capacity for critical thinking
- Write a scientific or technical report.
- Summarize an article or a technical report.

Logistics of the class

Learning Prerequisites

- RECOMMENDED COURSES: CS-433
- IMPORTANT CONCEPTS TO START THE COURSE: Programming in Python, good mathematical background

Teaching methods

- Lectures and exercises to discuss and work on problem sets (both numerical and analytical).

Grading

Expected student activities

- Attend lectures and take notes during lectures, participate in quizzes and read scientific articles. Solve the problem sets and take the final exam.

Assessment methods

- The final mark is a combination of three evaluations:
  - problem sets/mini project (25%)
  - quizzes (25%)
  - final exam (50%)

Grading details

- We'll have analytical problem sets (*not* graded); we'll share the problem sets and solutions after the exercises
- We'll have two journal clubs (with one quiz each, overall graded with 25%). You *need* to be present in the exercises (*week 8 & 14*).
- The exam comprises analytical problems & questions (50%)
- There will be one computational project on Google Cloud (4 weeks). This part will be graded (25%).

Google Cloud

TBD: Thanks for sponsoring!

Let's start...

Did you take Neuroscience classes?

Course's title: Brain-like computation and intelligence

Brain-like computation?

We will *discuss how the brain computes*. We will focus on distributed processing, brain-like connectivity, spiking, plasticity/learning, reward-based learning…

Intelligence?

Concretely, we will try to answer *how the brain achieves specific, complex abilities* such:

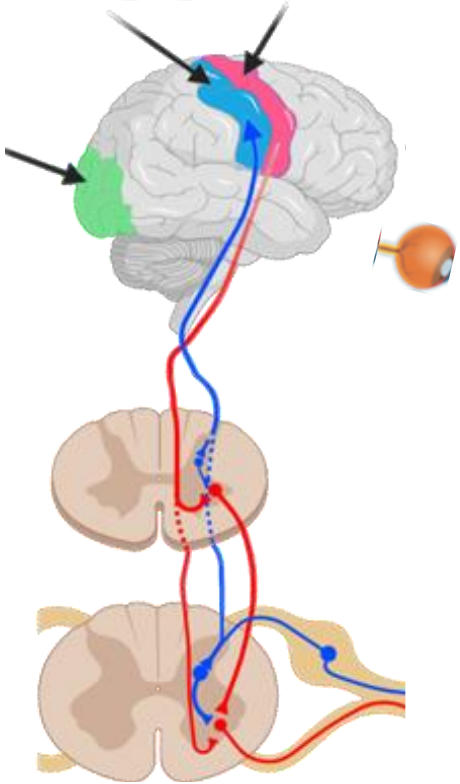
- Visually recognizing objects
- Navigating
- Language and cognition
- Reaching for objects
- Learning motor skills …

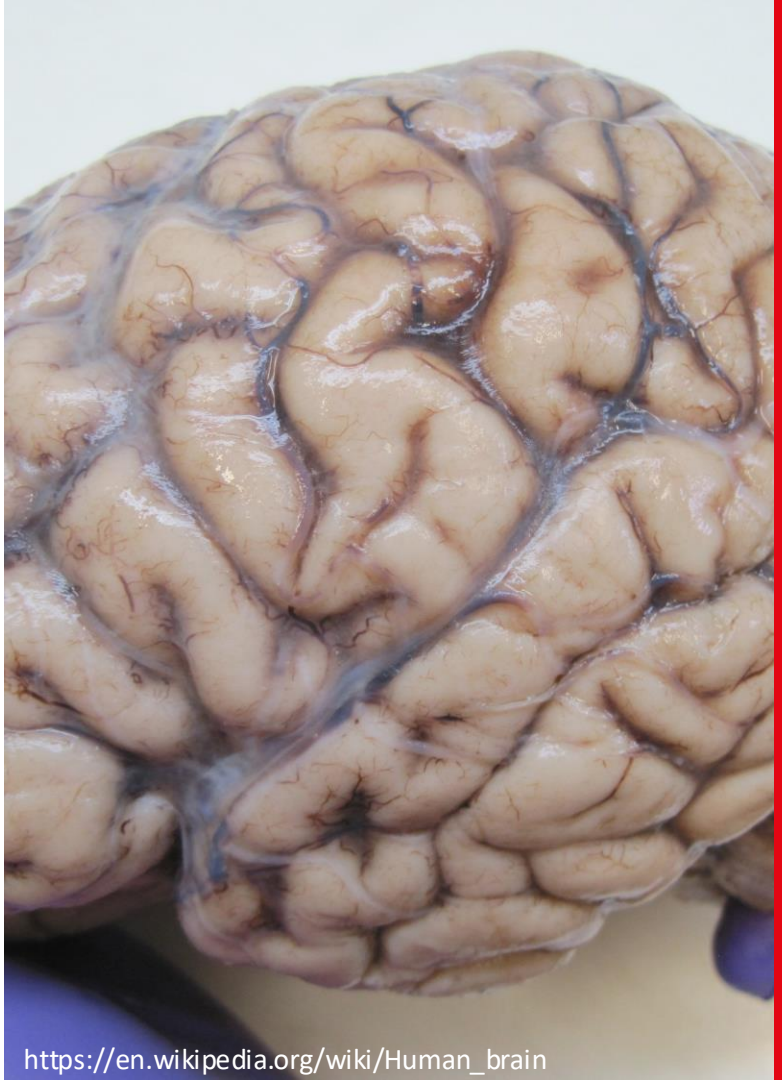
We will also discuss intelligence more broadly (later).

Class	Date	Topic
1	19/02/2025	Introduction & neural code
2	26/02/2025	Normative models
3	05/03/2025	Bayes and Brain-like circuits
4	12/03/2025	Task-driven models (Path integration)
5	19/03/2025	Task-driven models (Vison)
6	26/03/2025	Task-driven (Unsupervised, Audition, metamers, optimal stimuli)
7	02/04/2025	Task-driven (Proprioception) and Motor Control
8	09/04/2025	Language modeling in the brain I
9	16/04/2025	Language modeling in the brain II
10	23/04/2025	Easter Break
11	30/04/2025	Language modeling in the brain III (language in the service of cognition)
12	07/05/2025	Learning to control
13	14/05/2025	Brain-inspired reinforcement learning
14	21/05/2025	Skill learning
15	28/05/2025	Review

What is this course about?

Biological Intelligence

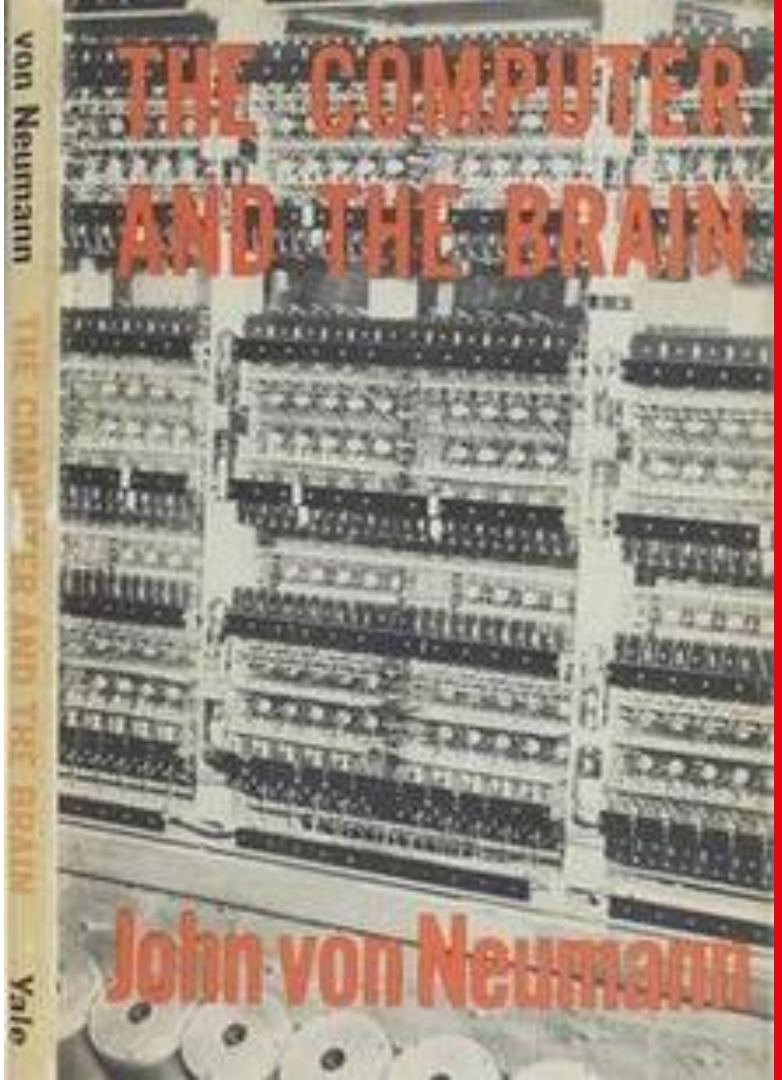




https://en.wikipedia.org/wiki/Human_brain

How does the brain compute?

vs. Turing Machines, von Neumann machines, …



How does the brain compute?

vs. Turing Machines, von Neumann machines, …

https://en.wikipedia.org/wiki/The_Computer_and_the_Brain

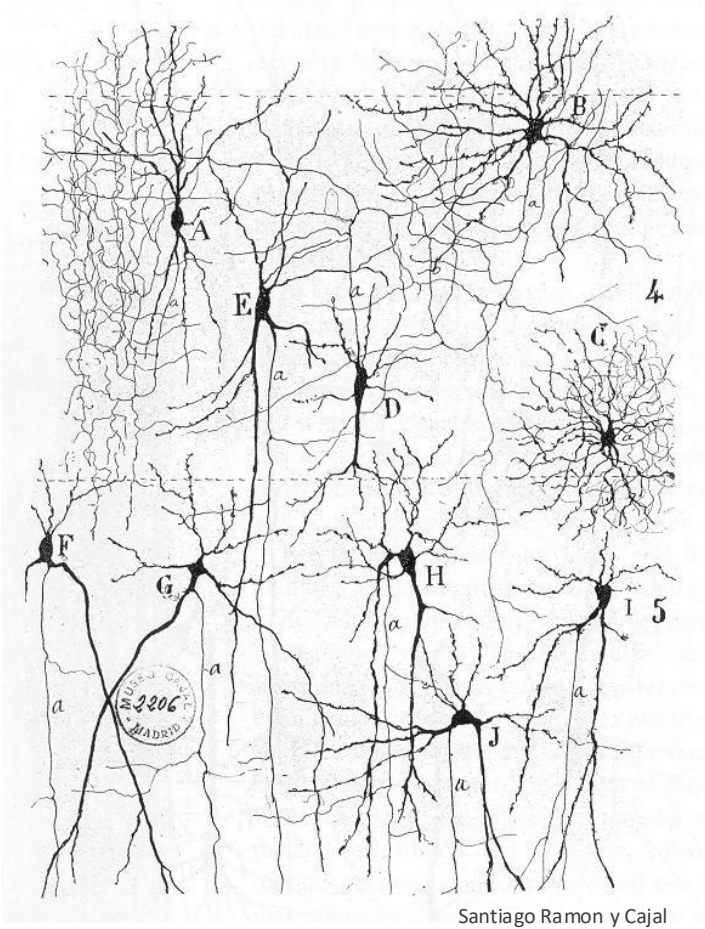
Brains are complex

Human brain:

- 86 billion (10^{10}) neurons
- ~1.5×10^{14} synapses

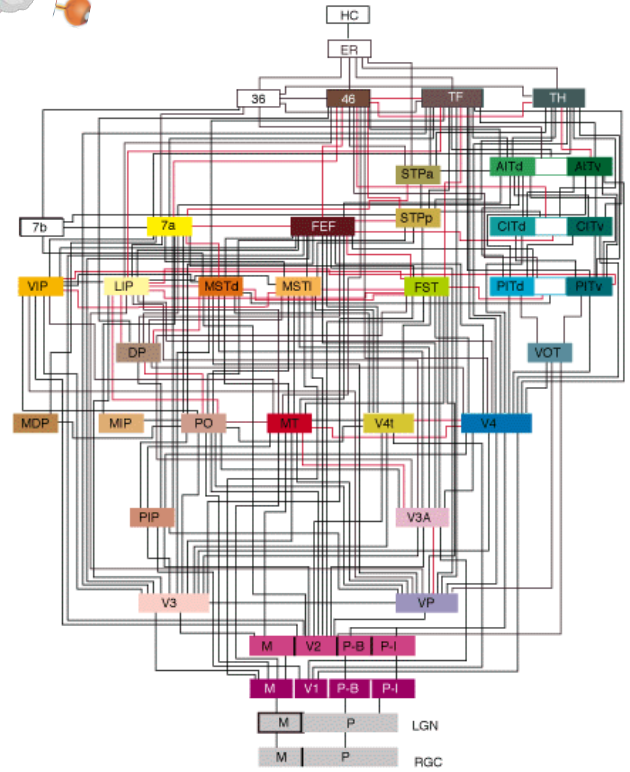
Mouse brain:

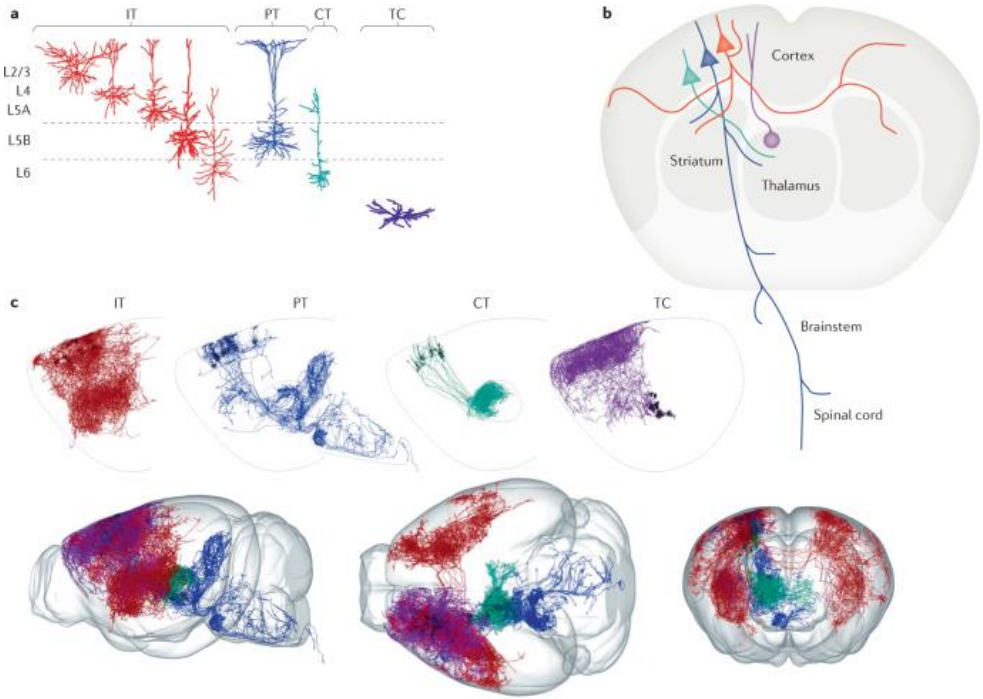
- 71 million neurons
- 10^{12} synapses



Santiago Ramon y Cajal

Neural circuits are staggeringly complex



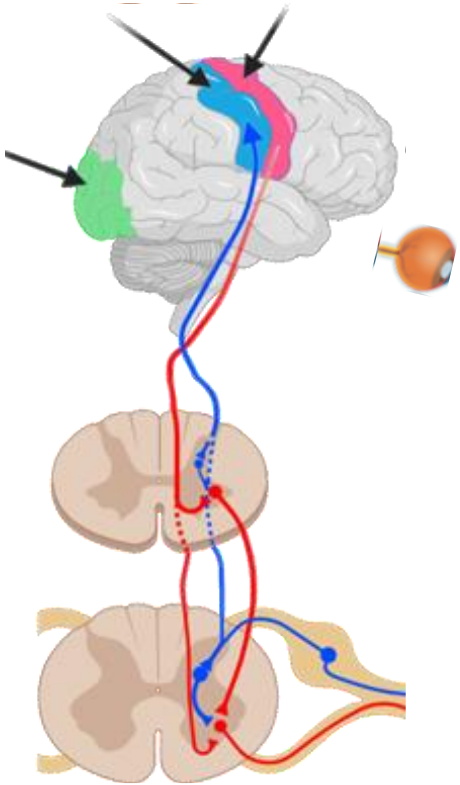


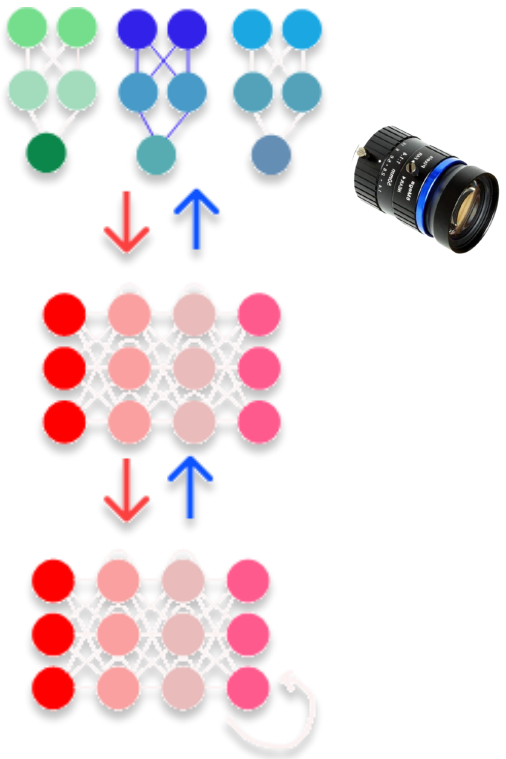
Felleman and Van Essen *Cerebral Cortex* 1991

Shepherd and Yamawaki *Nature review neuroscience* 2021

Biological Intelligence ⟷ Artificial Intelligence

Feynman: What I cannot create, I do not understand.

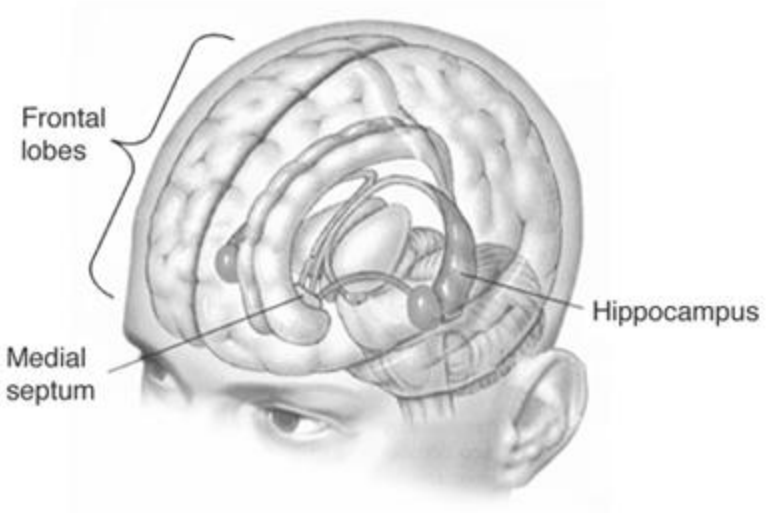




Hausmann*, Marin-Vargas*, Mathis**, Mathis**, Curr op. in Neuro., 2021

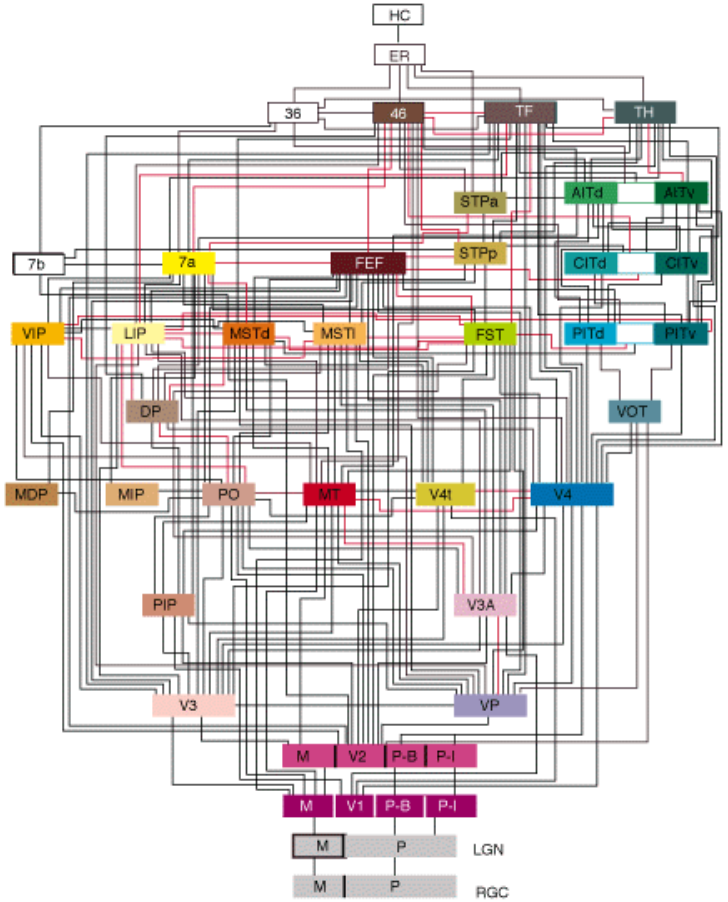
How do neuroscientists localize brain function?

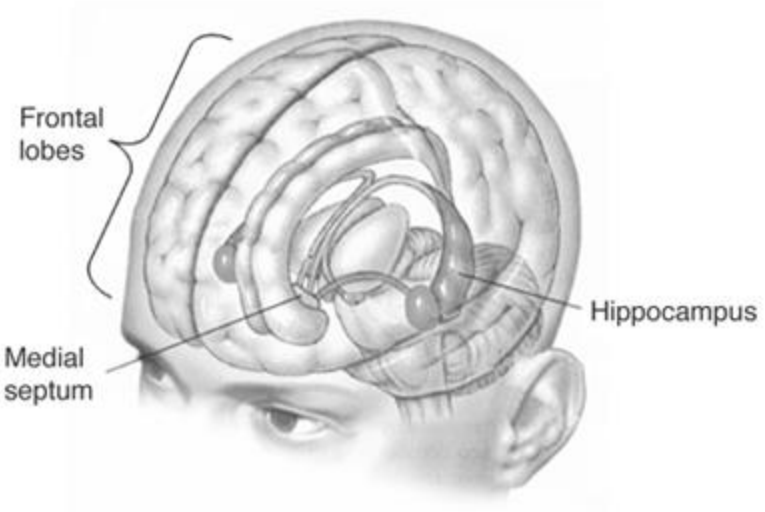
Patient H.M.



https://en.wikipedia.org/wiki/Henry_Molaison

Patient H.M.

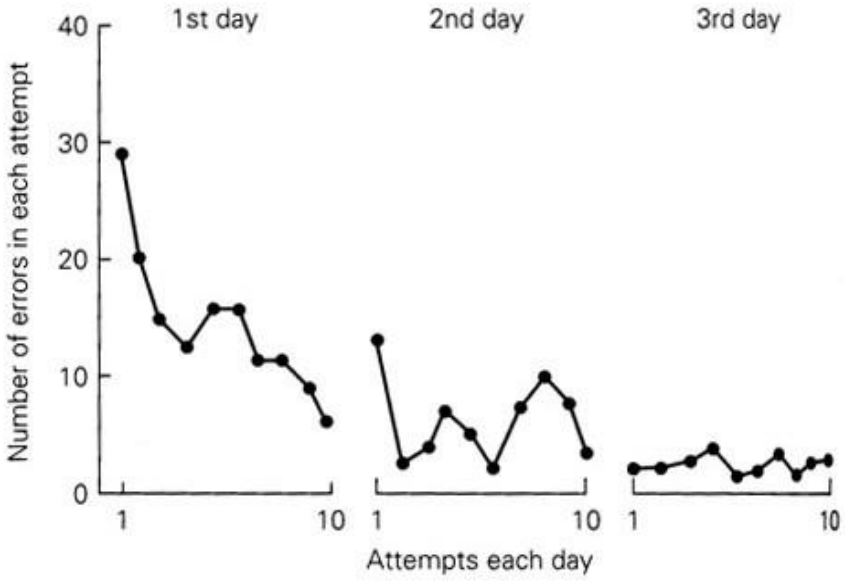




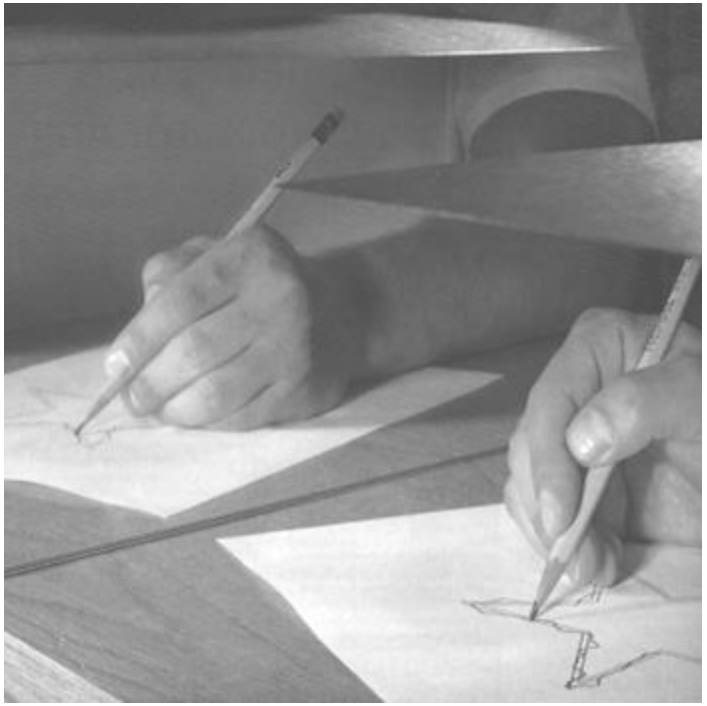
Felleman and Van Essen *Cerebral Cortex* 1991

https://en.wikipedia.org/wiki/Henry_Molaison

Milner 1962: mirror drawing could be learned over a period of days by the severely amnesic patient H.M. in the absence of any conscious memory of having practiced the task before



Milner (1962) Physiologie de l'hippocampe.

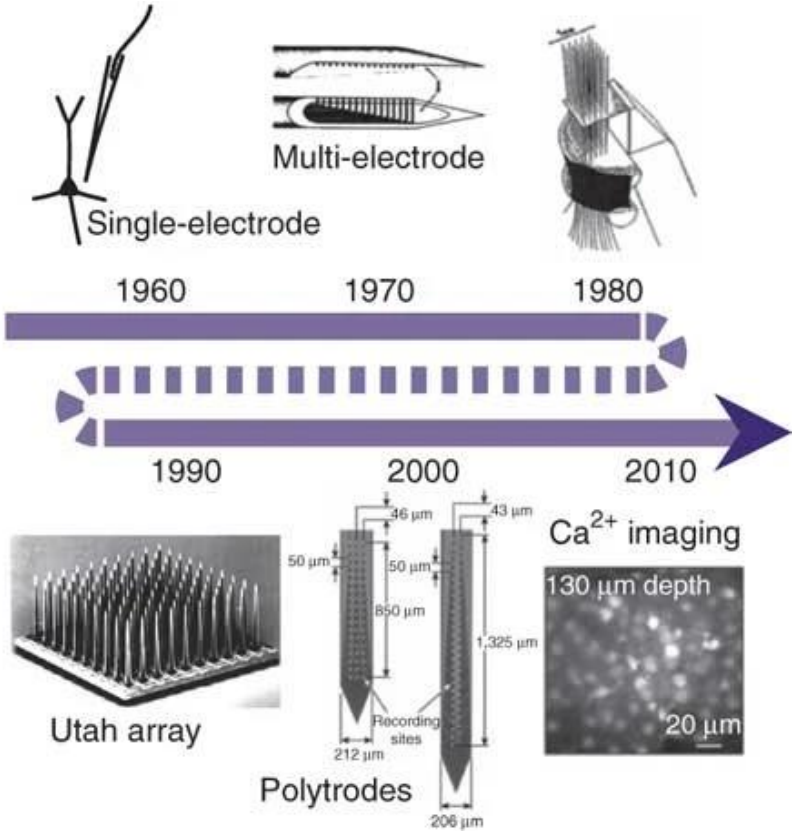


Kandel et al. Principles of Neural Science 2000 (62-2)

- HM made us realize that there are different forms of memory and that different parts of the brain support those!

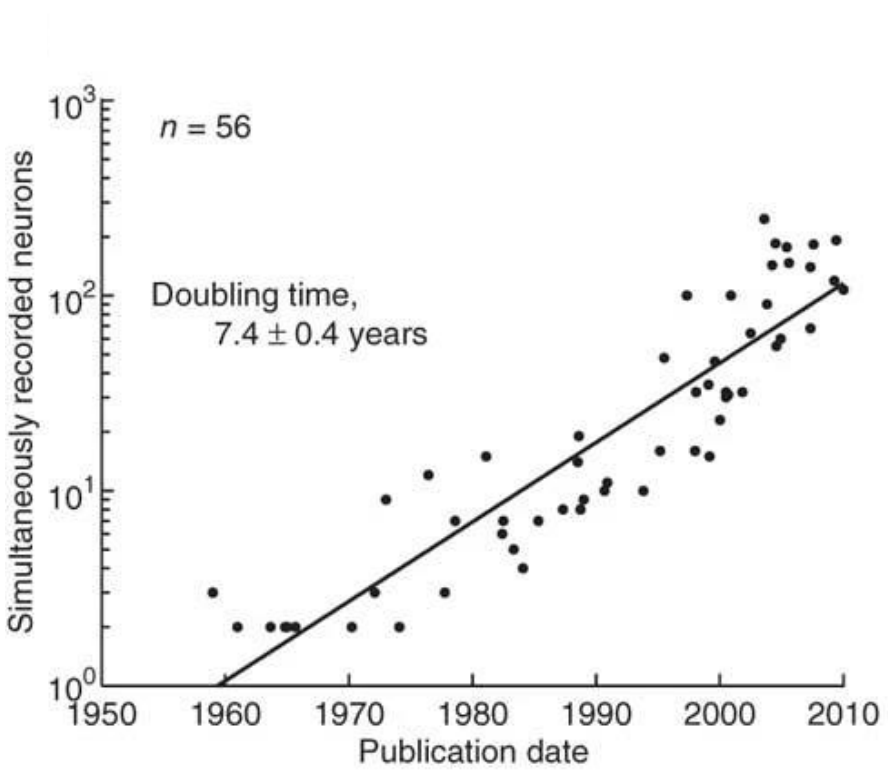
How do neuroscientists study neural function?

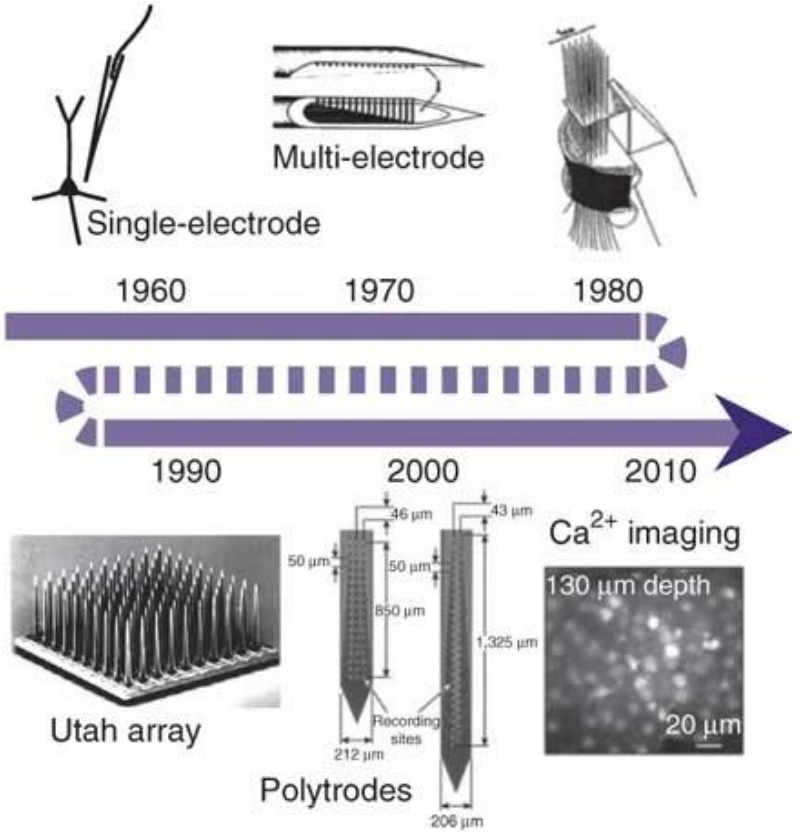
Single unit recoding: “Moore’s law”



How advances in neural recording affect data analysis, Stevenson & Kording Nature Neuroscience 2011

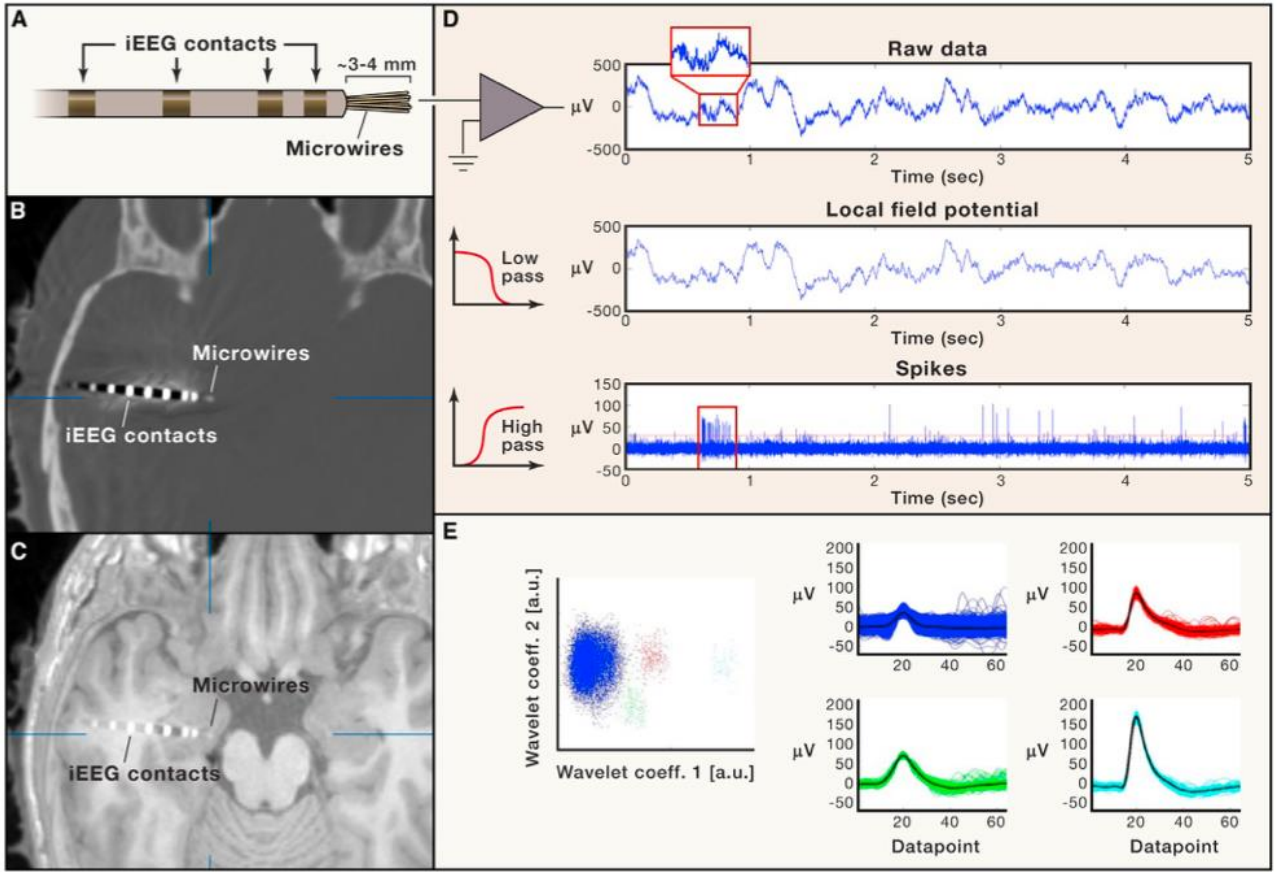
Single unit recoding: "Moore's law"





How advances in neural recording affect data analysis, Stevenson & Kording Nature Neuroscience 2011

Extracellular recordings

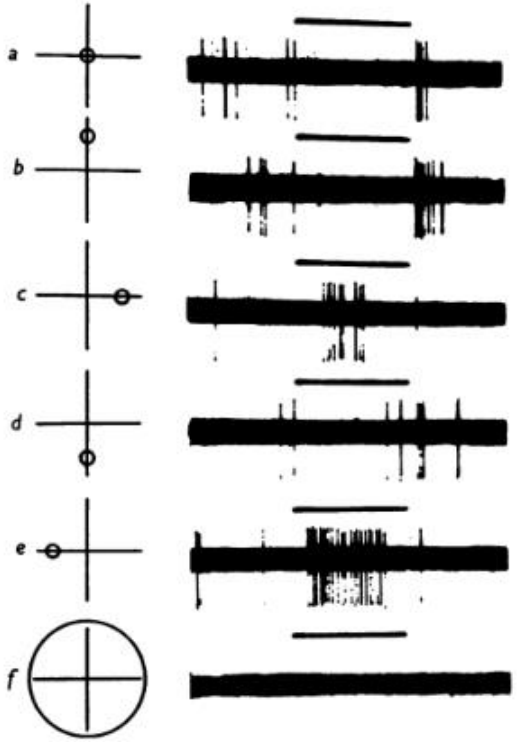


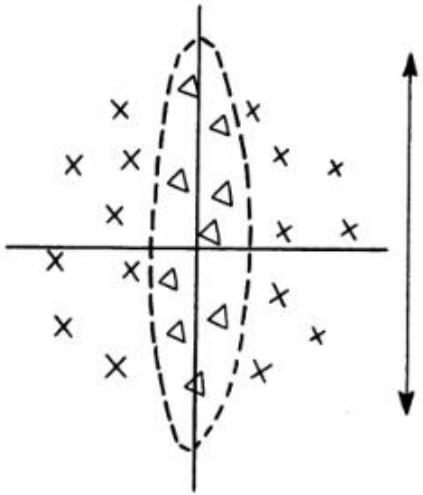
Quiroga, Cell 2019

What do neurons encode?

What do neurons care about?

1 degree spot of light





5 degree spot of light

J. Physiol. (1959) 148, 574–591

RECEPTIVE FIELDS OF SINGLE NEURONES IN THE CAT'S STRIATE CORTEX

By D. H. HUBEL* and T. N. WIESEL*

From the Wilmer Institute, The Johns Hopkins Hospital and University, Baltimore, Maryland, U.S.A.

(*Received* 22 *April* 1959)

Orientation selectivity

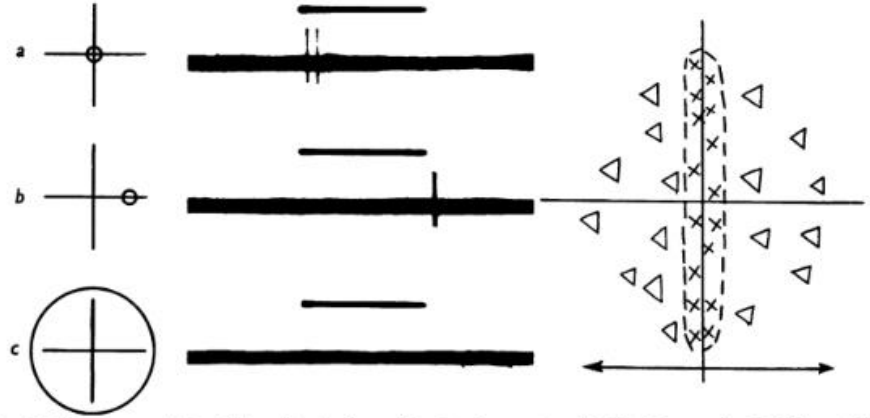


Fig. 2. Responses of a unit to stimulation with circular spots of light. Receptive field located in area centralis of contralateral eye. (This unit could also be activated by the ipsilateral eye.) *a*, 1° spot in the centre region; *b*, same spot displaced 3° to the right; *c*, 8° spot covering entire receptive field. Stimulus and background intensities and conventions as in Fig. 1. Scale, 6°.

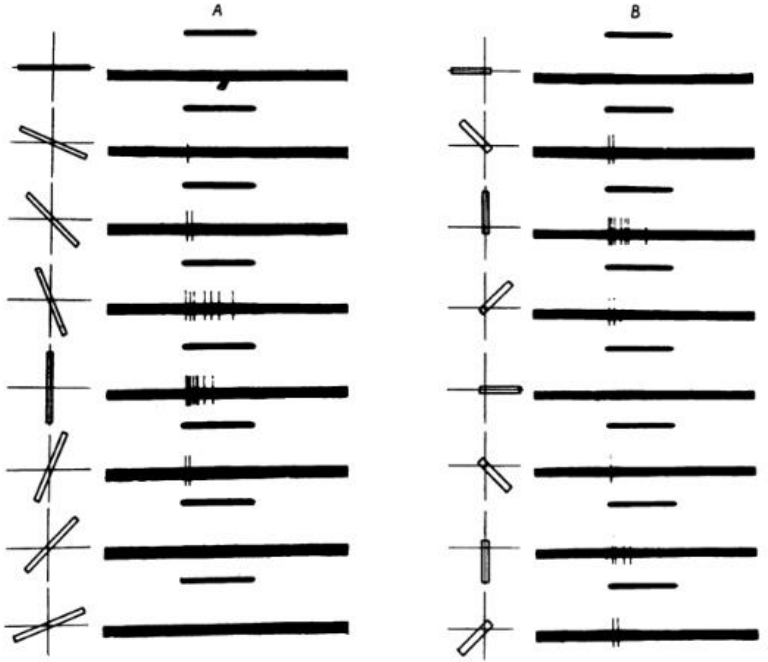


Fig. 3. Same unit as in Fig. 2. *A*, responses to shining a rectangular light spot, 1° × 8°; centre of slit superimposed on centre of receptive field; successive stimuli rotated clockwise, as shown to left of figure. *B*, responses to a 1° × 5° slit oriented in various directions, with one end always covering the centre of the receptive field: note that this central region evoked responses when stimulated alone (Fig. 2*a*). Stimulus and background intensities as in Fig. 1; stimulus duration 1 sec.

Hubel &Wiesel 1959

Reaching task

ON THE RELATIONS BETWEEN THE DIRECTION OF TWO-DIMENSIONAL ARM MOVEMENTS AND CELL DISCHARGE IN PRIMATE MOTOR CORTEX[1]

APOSTOLOS P. GEORGOPOULOS,[2] JOHN F. KALASKA,[3] ROBERTO CAMINITI,[4] AND JOE T. MASSEY[5]

Departments of Physiology and Neuroscience, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

Received October 30, 1981; Revised April 30, 1982; Accepted May 21, 1982

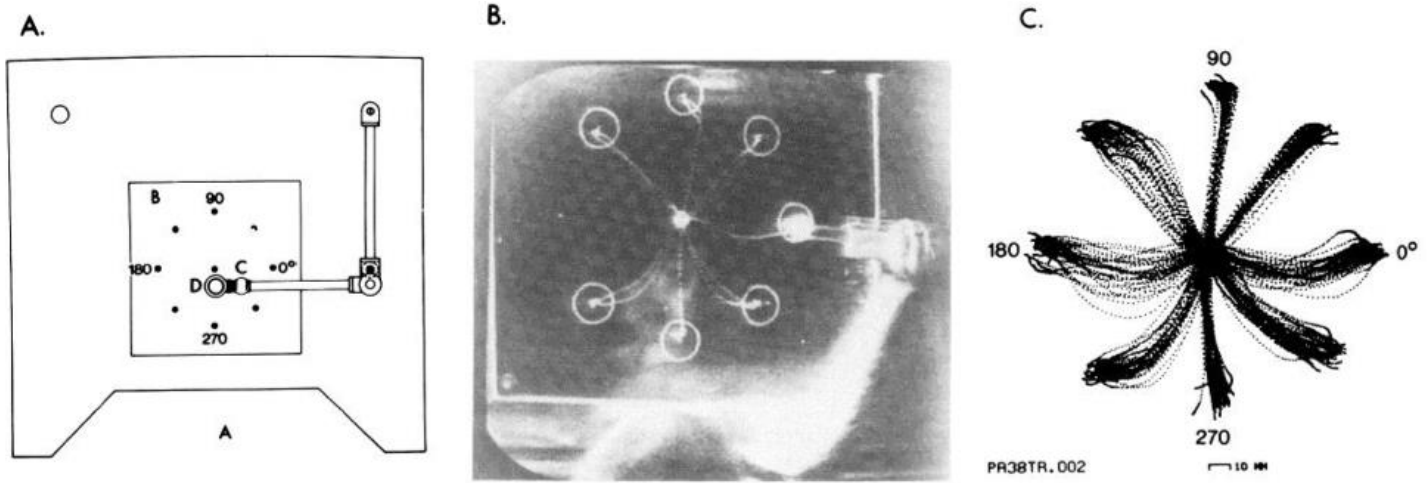


Figure 1. A, Diagram of the behavioral apparatus. The monkey sits at *position A*, facing a 25-cm square working surface (*B*) on which there are nine light-emitting diodes (LEDs). One LED is at the center of the working surface and eight are on a circle with an 8-cm radius; they are numbered from 0 to 315° counterclockwise. The monkey grasps an articulated manipulandum at its end (*C*) and moves it across the *x-y* surface of the plane to capture within a clear plastic circle (*D*) whichever LED is illuminated. The plane is tilted 15° from the horizontal toward the animal. *B*, Overhead view of a monkey performing the task displayed on a television monitor. The monkey has moved the manipulandum from the center to the target LED (in this case, the movement direction is 0°) to complete a trial. The trajectories of movement for this trial and for a few previous trials are superimposed on the television image as *light lines*. The *small circles* are the 25-mm-diameter target windows around each target LED. The starting (*center*) window had a diameter of 15 mm. *C*, Trajectories of 30 movements to each target made by a well trained monkey. Each *dot* is the position of the center of the target-capturing circle on the end of the manipulandum taken at 10-msec intervals.

Variability of responses

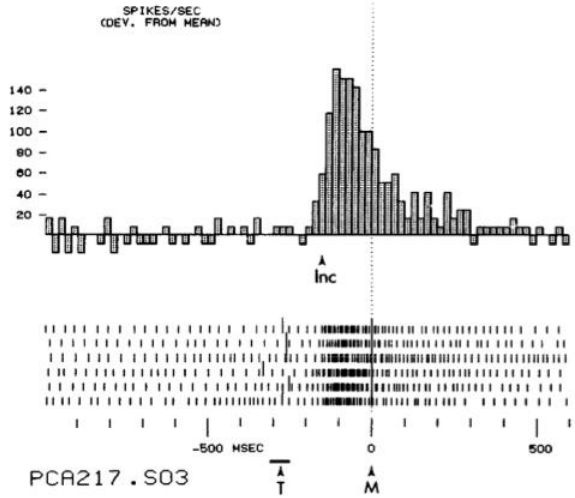


Figure 2. Example of the determination of the timing of the first change in neuronal activity using the method described in the text. Impulse activity was recorded from a single neuron during 6 movements toward the same target and is displayed as a raster (*bottom*) and as a perievent histogram (*top*). All trials and the histogram are oriented to the onset of movement. The time of appearance of the target for each trial is indicated by a *longer vertical line* in the spike train to the *left* of the orientation point. The histogram is plotted as the impulses per sec deviation from the mean control rate of discharge observed before the appearance of the target LED. The bin width of the histogram is 20 msec. *Inc*, Onset of the initial increase of activity as determined by method described in the text; *M*, onset of movement; *T*, mean time ± 1 SD of the appearance of the target.

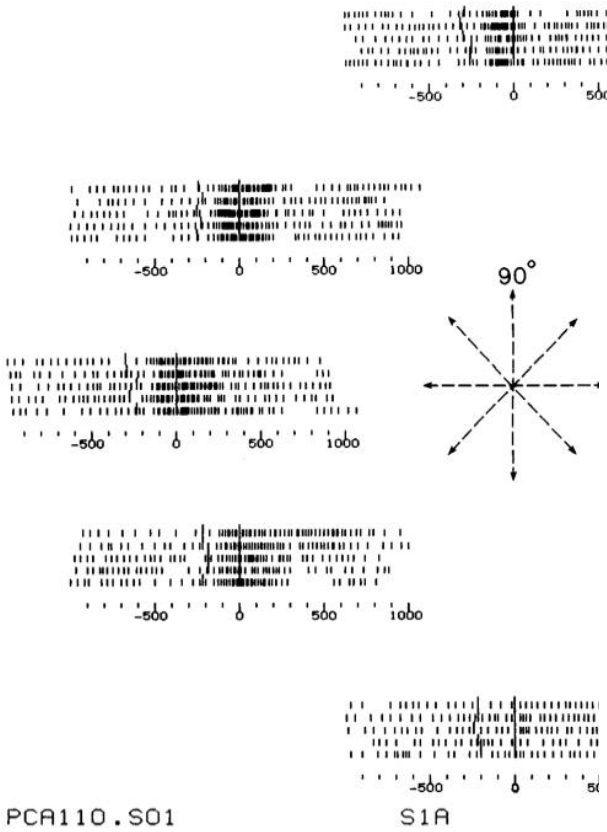
ON THE RELATIONS BETWEEN THE DIRECTION OF TWO-DIMENSIONAL ARM MOVEMENTS AND CELL DISCHARGE IN PRIMATE MOTOR CORTEX[1]

APOSTOLOS P. GEORGOPOULOS,[2] JOHN F. KALASKA,[3] ROBERTO CAMINITI,[4] AND JOE T. MASSEY[5]

Departments of Physiology and Neuroscience, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

Received October 30, 1981; Revised April 30, 1982; Accepted May 21, 1982

Coding for the direction of movement



Tuning curve:

$$f(\phi) = f_1 + f_2 \cos(\phi - \phi_0)$$

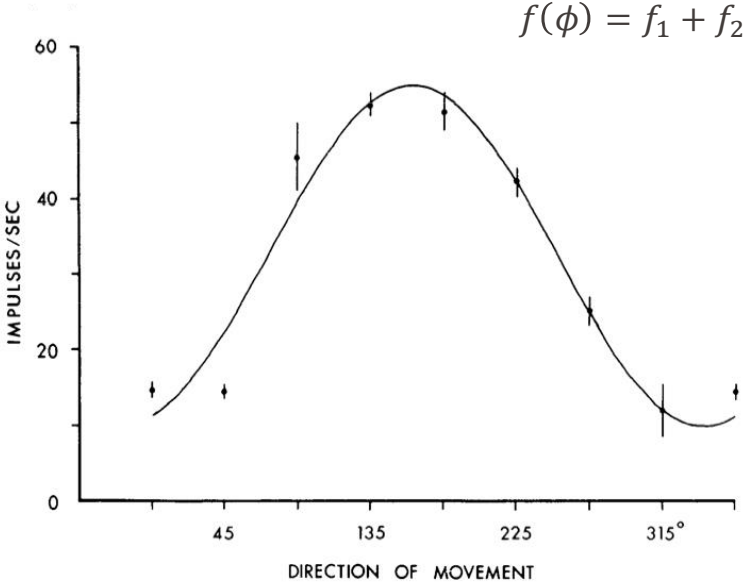


Figure 4. Orderly variation in the frequency of discharge of a motor cortical cell with the direction of movement. *Upper half*, Rasters are oriented to the movement onset, *M*, and show impulse activity during five repetitions of movements made in each of the eight directions indicated by the *center diagram*. Notice the orderly variation in cell's activity during the RT, MT, and TET. *Lower half*, Directional tuning curve of the same cell. The discharge frequency is for TET. The data points are the mean ± SEM. The regression equation for the fitted sinusoidal curve is $D = 32.37 + 7.281 \sin \theta - 21.343 \cos \theta$, where D is the frequency of discharge and θ is the direction of movement or, equivalently, $D = 32.37 + 22.5 \cos (\theta - \theta_0)$, where θ_0 is the preferred direction ($\theta_0 = 161°$).

Place cells, grid cells and the brain's spatial representation system

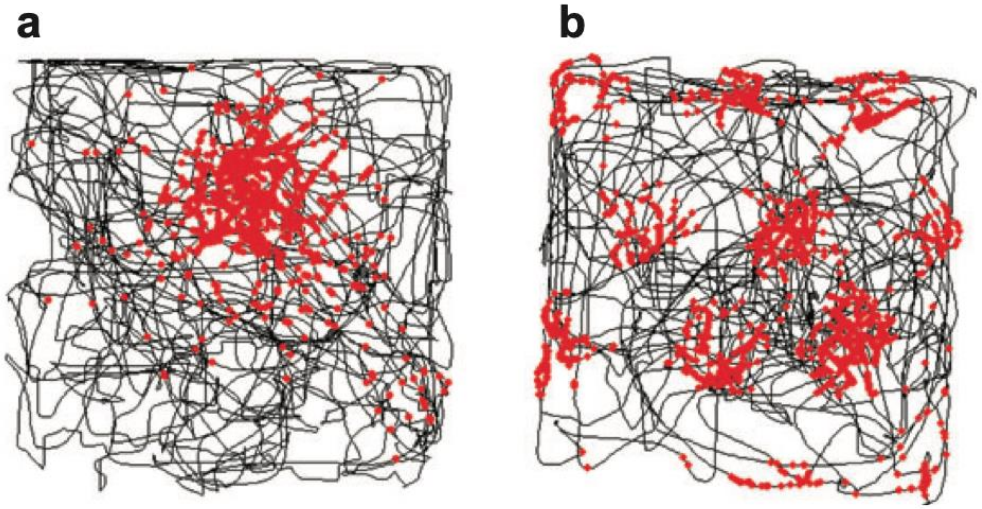


Figure 1

Place cell in the hippocampus (*a*) and grid cell in the medial entorhinal cortex (MEC) (*b*). Spike locations (*red*) are superimposed on the animal's trajectory in the recording enclosure (*black*). Whereas most place cells have a single firing location, the firing fields of a grid cell form a periodic triangular matrix tiling the entire environment available to the animal.

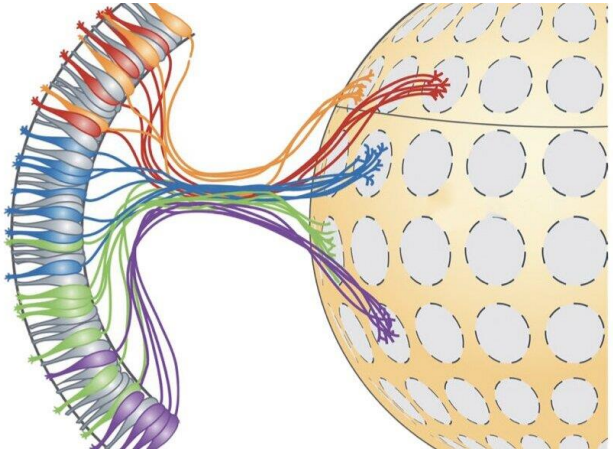
Moser, Kropff, Moser, Annual Review Neuro 2008

Further watching…

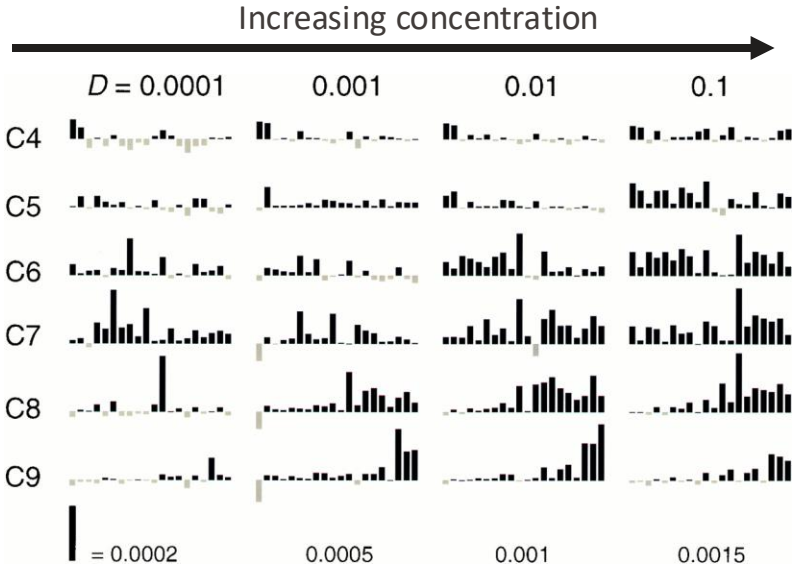
Check out

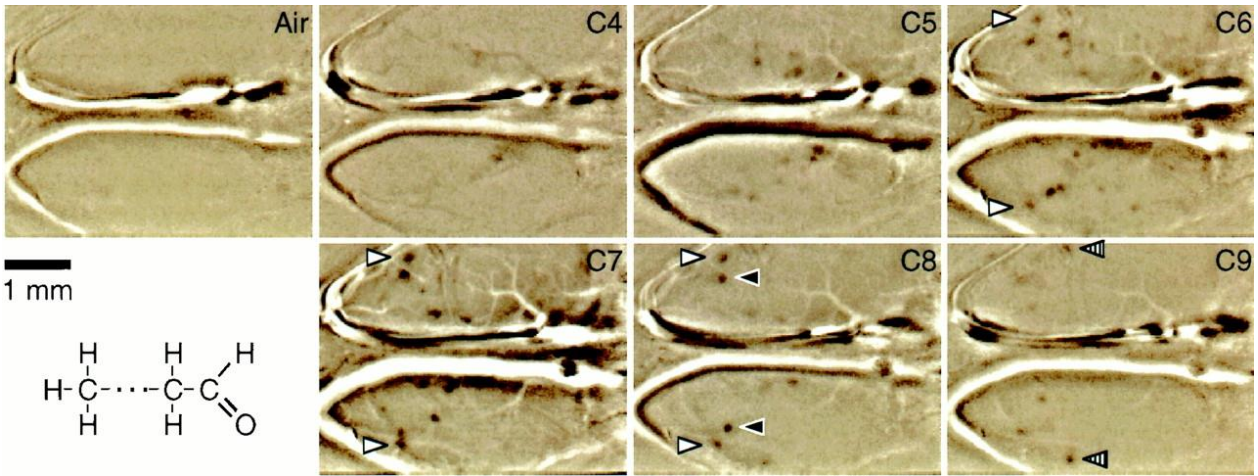
- David H. Hubel's Nobel Prize Lecture from 1981: https://www.youtube.com/watch?v=k2Zz2Re5BCc
- Torsten Wiesel's Nobel Prize Lecture from 1981: https://www.youtube.com/watch?v=zVRvzoATHmA
- May-Britt Moser's, Edvard Moser's and John O'Keefe's Nobel Prize Lecture from 2014: https://www.youtube.com/watch?v=P0tXhEbvjjg

Olfactory coding



Wang et al. Nature Neuro 2022

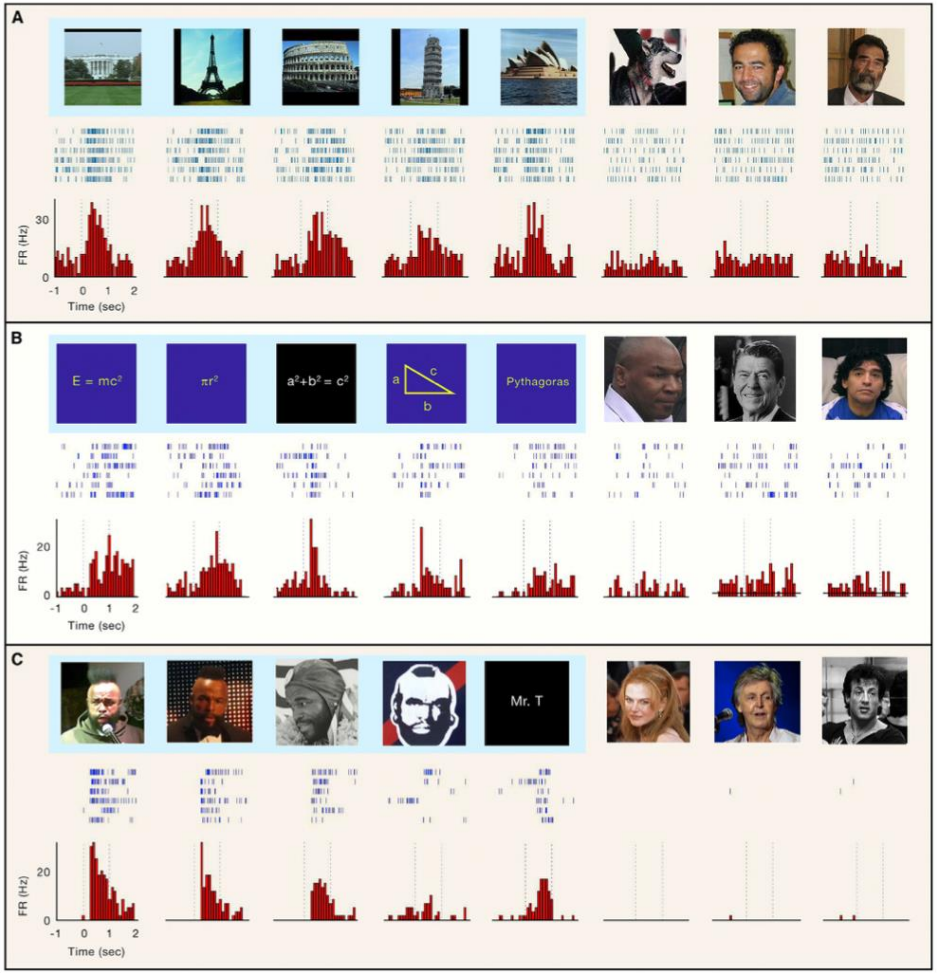




Meister & Bonhoeffer, J Neuro 2001

Single units in the human temporal lobe

Neurons can correspond to complex concepts shown in *different ways* in a selective fashion, e.g. “buildings” or “math”

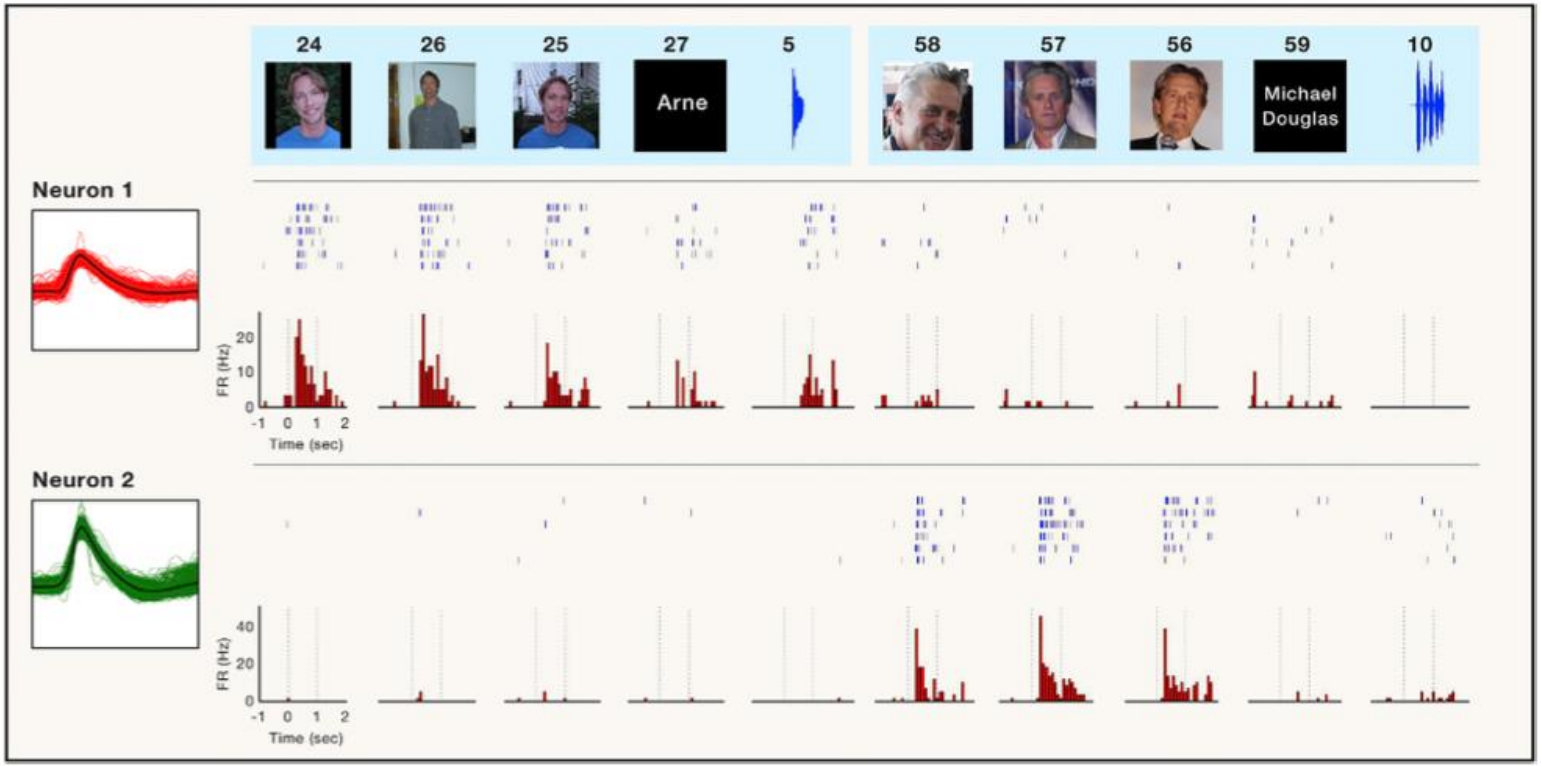


Raster plots and peristimulus histograms for 3 neurons with stimuli

Quiroga, Cell 2019

Multimodal invariance (visual, audio and text)

Concept cells

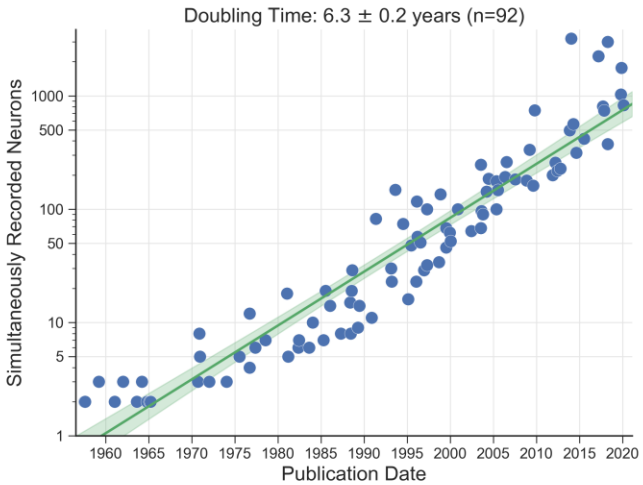


Quiroga, Cell 2019

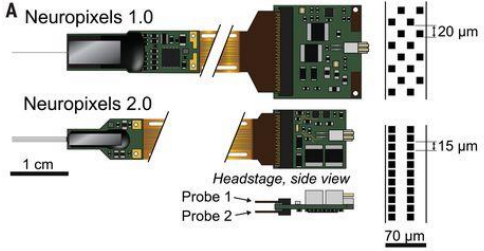
Neuropixel 2.0: high-density probe for stable, long-term recordings

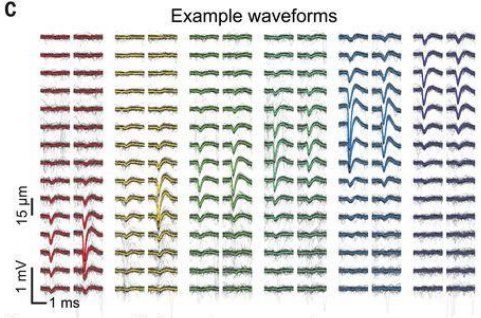
Steinmetz et al. 2020, Science

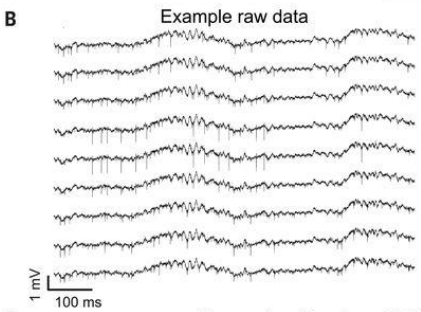
- 10,240 recording sites in one implant with ~1.1g

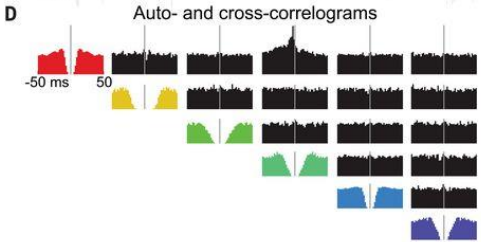


https://stevenson.lab.uconn.edu/scaling/









Encoding models

We have seen that

- neural responses to a stimulus are variable
- thus, neurons are typically described by stochastic models
- the (mean) firing rate is often described by “tuning curves”

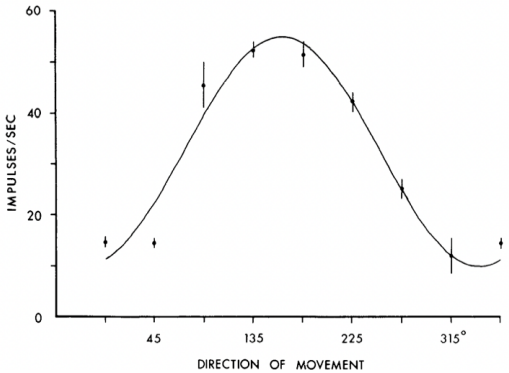
Example tuning curves:

- Gaussian tuning curve with preferred direction (location) μ and width σ:

$$f(\theta) = f_{max}\exp\left(-\frac{(\theta-\mu)^2}{2\sigma^2}\right)$$

- Cosine tuning curve with preferred phase ϕ_0 and rates f_1, f_2:

$$f(\phi) = f_1 + f_2\cos(\phi-\phi_0)$$



Encoding models

Imagine a population of neurons encoding a continuous variable $x \in \mathbb{R}^{\mathrm{D}}$ with response $k = (k_1, \ldots k_{\mathrm{N}})$.

Here:

- x is the stimulus
- $k = (k_1, \ldots k_{\mathrm{N}})$ denotes the spike count firing rate (for N neurons)

Then several probabilities are relevant:

- $P(x)$ the probability of the stimulus x being presented – often called prior probability
- $P(k)$ the probability of the response k being recorded
- $P(k|x)$ the conditional probability of evoking spike rate k given that stimulus x was presented (called likelihood for observed k)
- $P(x|k)$ the conditional probability of observing stimulus x given k was recorded

Example encoding models

For a single neuron with Poisson emission and tuning curve $\lambda(x)$ we get:

$$P(k|\mathbf{x}) = \frac{\lambda(\mathbf{x})^k}{k!}\exp\big(-\lambda(\mathbf{x})\big)$$

For N neurons that fire (statistically) independently:

$$P(k|\mathbf{x}) = \prod_{i=1}^{N} \frac{\lambda_i(\mathbf{x})^{k_i}}{k_i!}\exp\big(-\lambda_i(\mathbf{x})\big)$$

This *encoding model* describes the response properties of a group of neurons to stimulus x.

How can we decode the stimulus?

Let's assume we observed the spike count vector k, what can we say about the stimulus (location x)?

Bayes theorem:

$$P(x|k) = \frac{P(x,k)}{P(k)} = \frac{P(x,k)\cdot P(x)}{P(k)\cdot P(x)} = P(k|x)\cdot\frac{P(x)}{P(k)}$$

Thus, we find that

$$P(x|k) \propto P(k|x)P(x)$$

Well-known estimators/decoders

Maximum a posteriori (MAP) estimator:

$$x_{MAP}(k) = argmax_x P(x|k) = argmax_x P(k|x)P(x)$$

Maximum likelihood estimator (ML):

$$x_{ML}(k) = argmax_x P(k|x)$$

Assessing performance with bias and variance

Consider an estimator for the random variable X as x_{est}.

Then the bias is given as

$$b_{est}(x) = \langle x_{est} \rangle - x$$

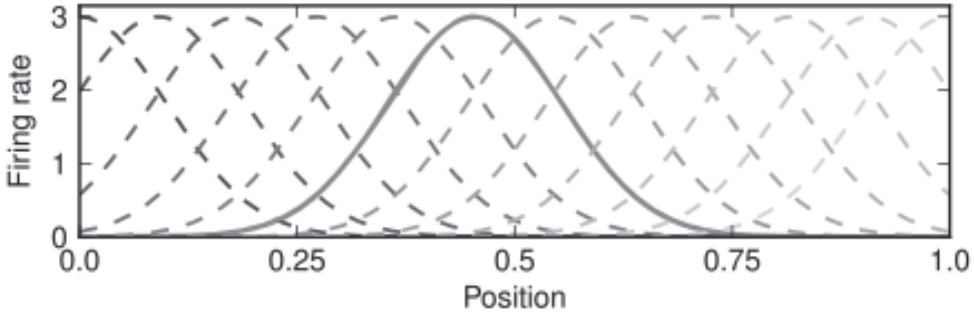
and the variance as

$$\sigma^2_{est}(x) = \langle (x_{est} - \langle x_{est} \rangle)^2 \rangle$$

Note, here $\langle . \rangle$ is given by averaging over P(k|x), i.e.

$$\langle x_{est} \rangle = \mathbb{E}_{\mathrm{K}}(x_{est}) = \int_k x_{est}\,(\mathrm{k}) P(k|x) dk$$

How well does a population of neurons encode a stimulus?



- Decoding can be used to understand the limit of information in a population
- However, we need *optimal decoders*, otherwise our conclusions might reflect limitations of our decoders, rather than bounds on the neural system being studied
- How can we get bounds on optimal decoders?

Cramer-Rao inequality and Fisher information

For any biased estimator it holds that

$$\sigma_{est}^2(x) \geq \frac{(1 + b'_{est}(x))^2}{I(x)}$$

Note that for unbiased estimators, we have

$$\sigma_{est}^2(x) \geq \frac{1}{I(x)}$$

With “Fisher information” defined as

$$I(x) = \int p(k|x) \left(-\frac{\partial^2 lnp(k|x)}{\partial x^2} \right) dk$$

Equivalently one can define

$$I(x) = \int p(k|x) \left(\frac{\partial lnp(k|x)}{\partial x} \right)^2 dk$$

Cauchy-Schwartz inequality

For any pair of vectors in an inner product space it holds that:

$$\langle u, v\rangle^2 \leq \langle u, u\rangle\langle v, v\rangle$$

Proof: Due to the semi-positiveness of the inner product, the following term is positive, as it is a square:

$$\langle(\langle u, u\rangle v - \langle u, v\rangle u)^2\rangle \geq 0$$

Computing the square gives

$$\langle u, u\rangle^2\langle v, v\rangle - \langle u, v\rangle^2\langle u, u\rangle \geq 0$$

which proves the CS. Note we write: $\langle \mathrm{w}^2\rangle = \langle w, w\rangle$ as a shorthand.

Corollary: For random variables u, v with inner product:

$$\langle u, v\rangle := \mathbb{E}(uv)$$

- CS yields: $|\mathbb{E}(uv)|^2 \leq \mathbb{E}(u^2)\mathbb{E}(v^2)$

Hence CS states: $|\mathbb{E}(uv)|^2 \leq \mathbb{E}(u^2)\mathbb{E}(v^2)$

To prove the Cramer-Rao bound let's set: $u := \partial \ln p / \partial x$ and $v := x_{est} - \langle x_{est} \rangle$.

By design we find

$$\langle u, u \rangle = I(x)$$

and

$$\langle v, v \rangle = \mathbb{E}\left((x_{est} - \langle x_{est} \rangle)^2\right) = \sigma_{est}^2$$

Thus, the CS gives us

$$\sigma_{est}^2 I(x) \geq \langle \frac{\partial \ln p}{\partial x} (x_{est} - \langle x_{est} \rangle) \rangle^2$$

Let's just copy the last expression: $\sigma_{est}^2 I(x) \geq \langle \frac{\partial \ln p}{\partial x} (x_{est} - \langle x_{est} \rangle) \rangle^2$

We note that the bias is given by

$$x + b_{est} = \langle x_{est} \rangle = \int x_{est} p(k|x) d\mathrm{k}$$

Let's take the derivative with respect to the stimulus x, we note:

$$1 + b'_{est}(x) = \int x_{est}(k) \frac{\partial p(k|x)}{\partial x} dk = \int x_{est} p(k|x) \frac{\partial lnp(k|x)}{\partial x} dk$$

Note that

$$\int \langle x_{est} \rangle p(k|x) \frac{\partial lnp(k|x)}{\partial x} dk = \langle x_{est} \rangle \int p(k|x) \frac{\partial lnp(k|x)}{\partial x} dk = \langle x_{est} \rangle \int \frac{\partial p(k|x)}{\partial x} dk = 0$$

The last step holds as $\int p(k|x) dk = 1$ .Thus, we find that:

$$1 + b'_{est}(x) = \int (x_{est} - \langle x_{est} \rangle) p(k|x) \frac{\partial lnp(k|x)}{\partial x} dk$$

Just plugging this in the RHS of the CS finishes the proof.

Cramer-Rao inequality and Fisher information

For any biased estimator it holds that

$$\sigma_{est}^2(x) \geq \frac{(1 + b'_{est}(x))^2}{I(x)}$$

Note that for unbiased estimators, we have

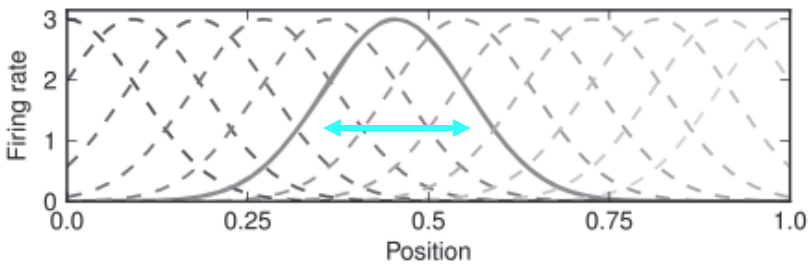
$$\sigma_{est}^2(x) \geq \frac{1}{I(x)}$$

With Fisher information defined as

$$I(x) = \int p(k|x) \left(-\frac{\partial^2 lnp(k|x)}{\partial x^2} \right) dk$$

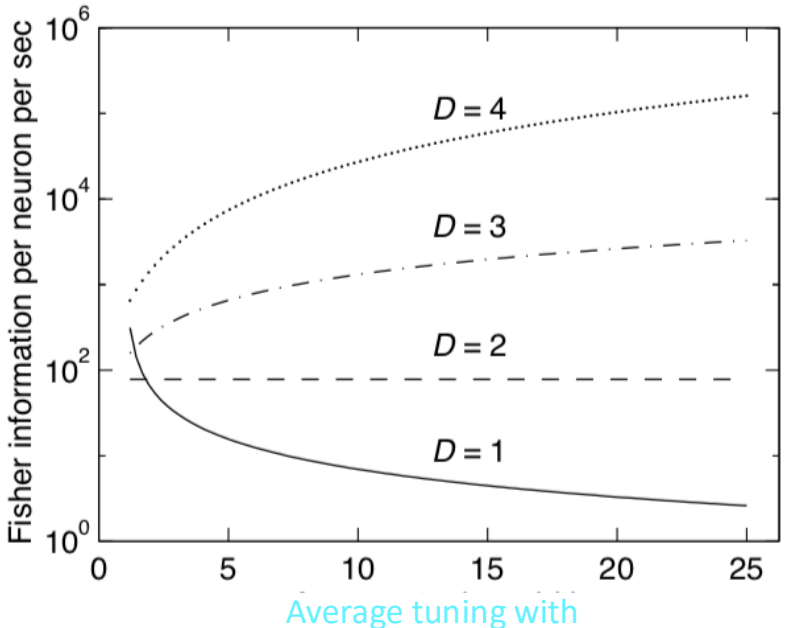
How much information do populations of neurons contain?

tuning with



You'll compute this in the exercises!!

D dimension of the stimulus space



Zhang & Sejnowksi, Neural Computation 1999

Take-home messages

- Neurons are tuned to many different external variables (location, movement direction, odors, concepts, …)
- The nervous system is (supposedly) interested in inferring what is happening in the world with neural patterns
- External variables are encoded in a noisy, distributed way
- The Fisher information allows us to study how well a population of neurons encode a stimulus/movement plan/etc.