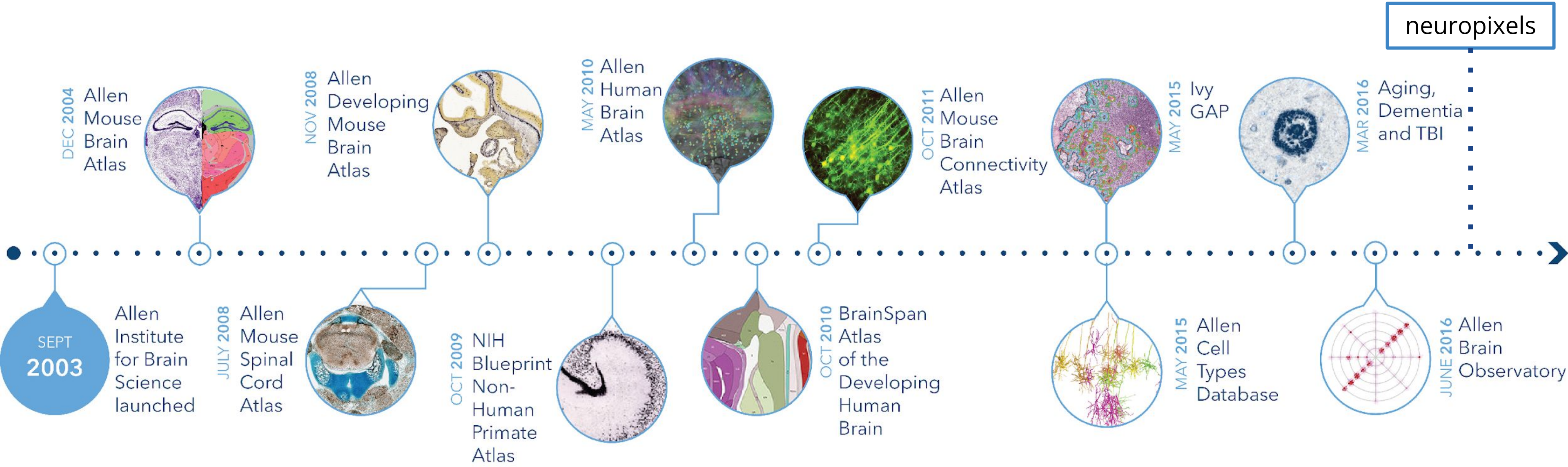
Data Structures
& Signal Processing

BIPN 162

Objectives for today

- Identify the three common features of datasets: **metadata**, **raw data**, and **processed** data
- Define **metadata** and explore the metadata associated with an NWB file
- Describe features of **time series** and common ways to analyze them
- Load & explore an NWB time series

Neuroscience datasets are increasingly being shared online



NEURODATA WITHOUT BORDERS

DANDI
dandiarchive.org

Neuroscience Datasets

Database Name	URL	Data Summary	Student Objective	Accessing API/SDK	Using Pandas	Text Mining
NeuroSynth	https://neurosynth.org/	Summarizes fMRI data from many studies	Perform meta analyses of fMRI data			X
LISC	https://lisc-tools.github.io/l	Text analysis of papers	Search publications for terms to identify interesting intersections.			X
Allen Cell Types Atlas	http://celltypes.brain-map.(	Whole-cell electrophysiology in genetically-identified cell types in mice and humans	Compare electrophysiology metrics for different cell types in mice and humans.	X	X	
Allen Brain Observatory	http://observatory.brain-m	*In vivo* 2p imaging in genetically-identified cell types in mice	Compare visual responses of cells recorded via two-photon imaging and analyze correlations.	X	X	
Allen Neuropixels	https://allensdk.readthedo	Extracellular recording & behavior in mice performing a task	Compare visual & behavioral responses of cells recorded with high density extracellular recording arrays. Also includes LFP data.	X	X	
Allen RNAseq	https://portal.brain-map.or	RNAseq in mice & humans	Compare gene expression in mice and humans in different cells & brain regions.		X	
Allen Connectivity	http://connectivity.brain-m					
Allen Synaptic Physiology	https://portal.brain-map.or					
Neuropixels Spike Sorting	http://repository.cshl.edu/i	Lots of spikes	PCA analysis to isolate units in extracellular recording			

Open Neuroscience Datasets https://bit.ly/openneurodatasets

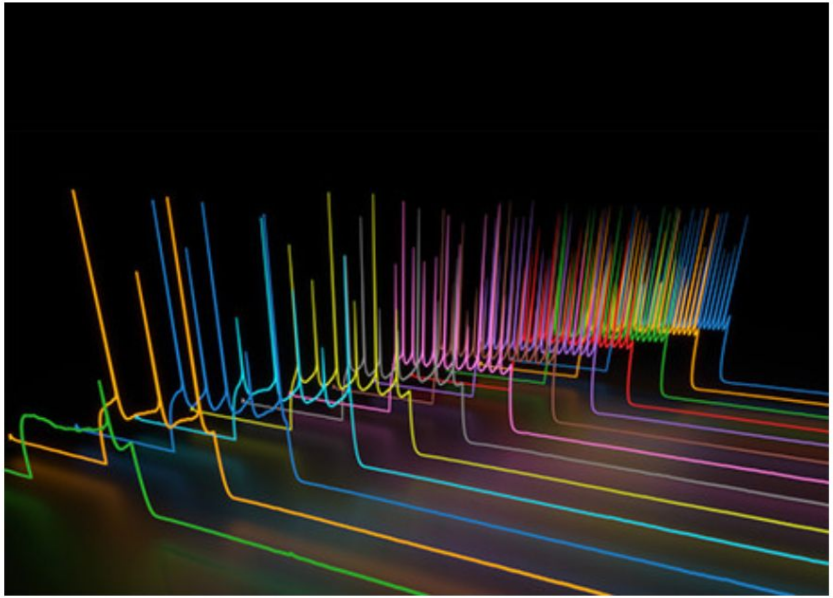
NEURODATA WITHOUT BORDERS

ABOUT COMMUNITY

Neurodata Without Borders

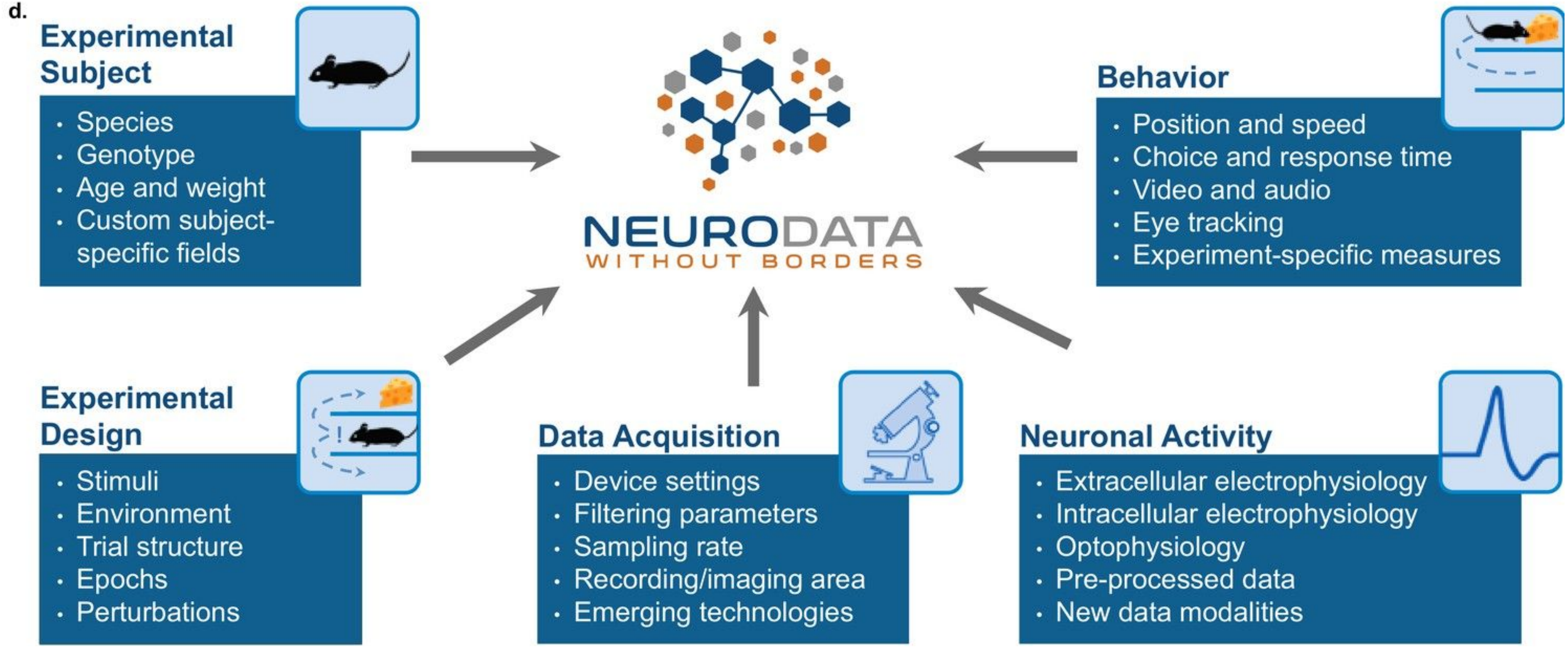
Neurodata Without Borders (NWB) is a data standard for neurophysiology, providing neuroscientists with a common standard to share, archive, use, and build analysis tools for neurophysiology data. NWB is designed to store a variety of neurophysiology data, including data from intracellular and extracellular electrophysiology experiments, data from optical physiology experiments, and tracking and stimulus data.

The NWB team consists of neuroscientists and software developers who recognize that adoption of a unified data format is an important step toward breaking down the barriers to data sharing in neuroscience.

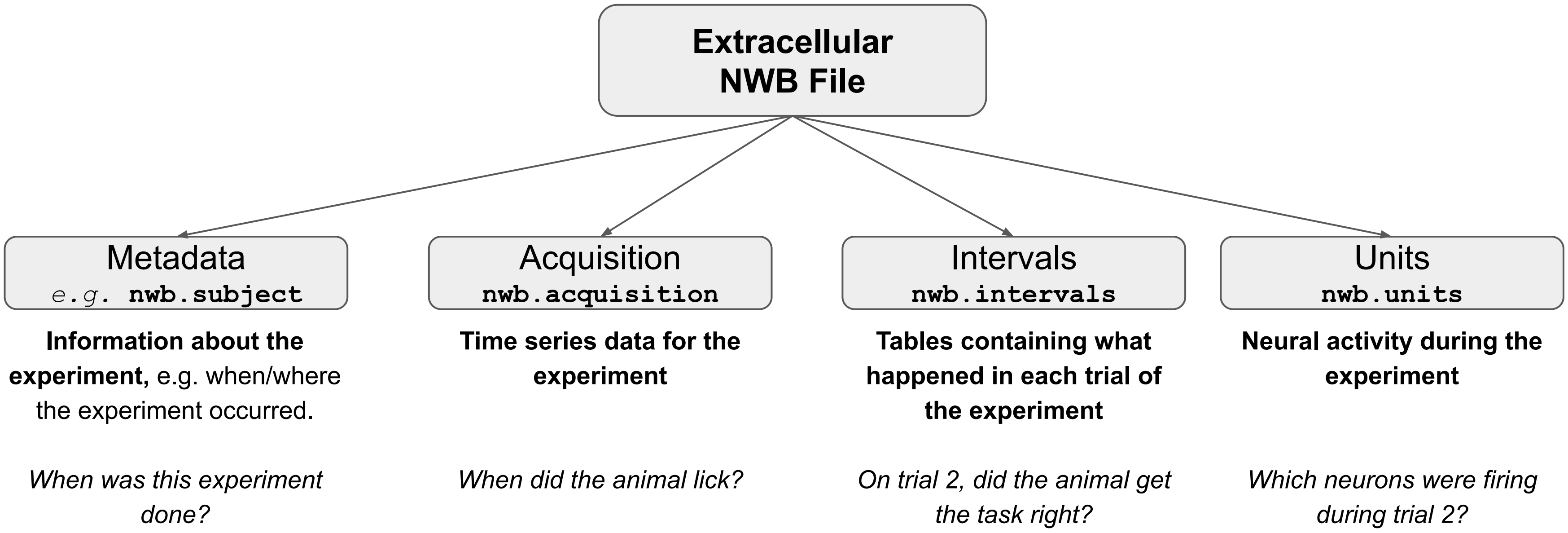


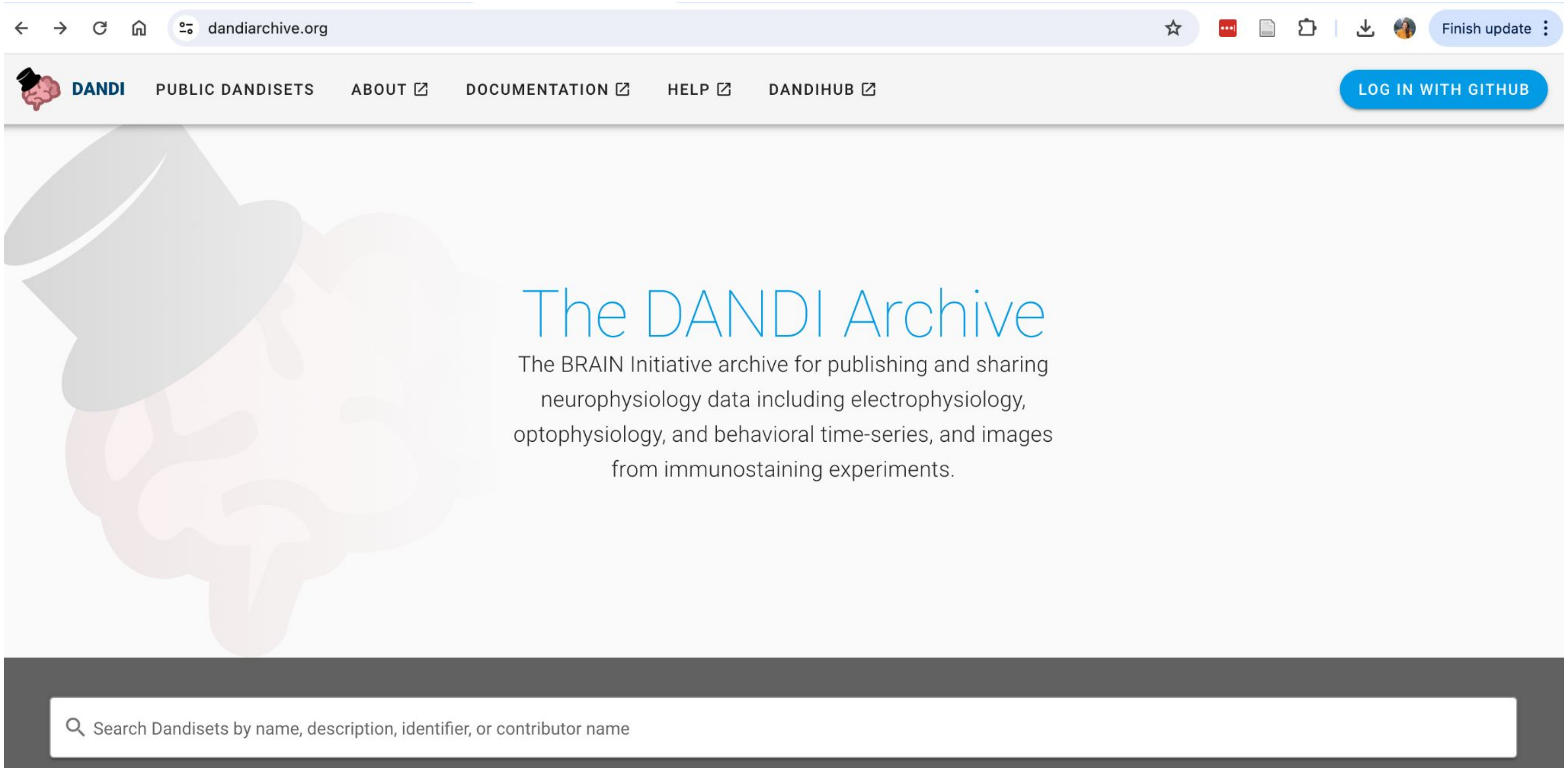
(Courtesy of the Allen Institute for Brain Science)

See also Rübel et al. (2022): https://elifesciences.org/articles/78362



Rübel et al. (2022): https://elifesciences.org/articles/78362


Extracellular NWB File
Metadata
e.g. nwb.subject
Acquisition
nwb.acquisition
Intervals
nwb.intervals
Units
nwb.units
Information about the experiment, e.g. when/where the experiment occurred.
Time series data for the experiment
Tables containing what happened in each trial of the experiment
Neural activity during the experiment
When was this experiment done?
When did the animal lick?
On trial 2, did the animal get the task right?
Which neurons were firing during trial 2?



http://dandiarchive.org

Signal processing

Anything recorded continuously over time is a **time series**

(a set of data points generated from successive measurements over time)

Commonly encountered time series data in biology

- Gene expression data over time
- Neurophysiology recordings (e.g. electrophysiology, imaging)
- Circadian rhythm data
- Medical observations over time
- Animal movement
- Physiology data (e.g. heart rate/ECG, pulse rate, respiration, etc.)
- Molecules/proteins/cells moving

White blood cell tracking bacteria
Image info

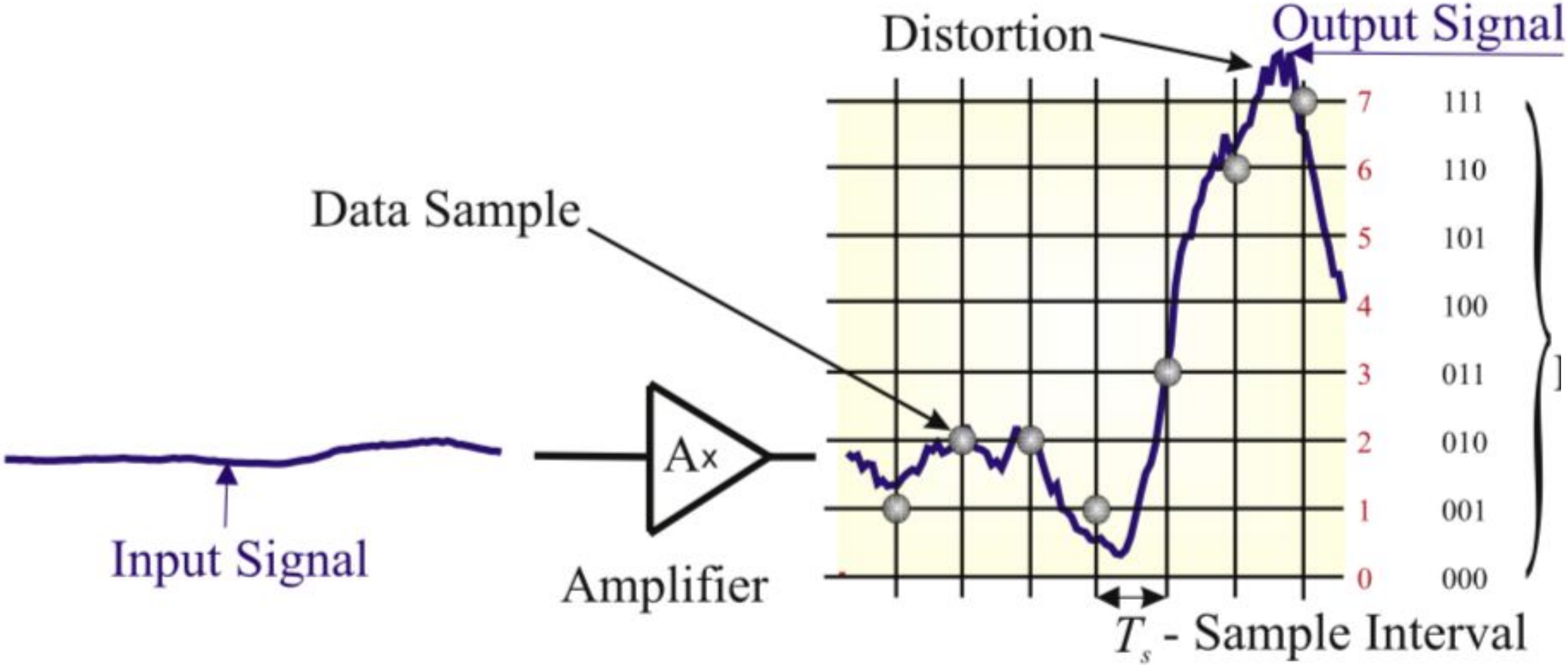
Common signal processing approaches

- Preprocessing & data cleaning
  - Removing outliers and/or noise
- Filtering
  - Using convolution
  - Using frequency
- **Looking for correlations in time**
- Clustering & classification
- Dimensionality reduction or segmentation
- Prediction
- Anomaly or peak detection — ***spike sorting***

Real world signals (e.g., brain and sound waves) are continuous in time. In other words, they're **analog**.

However, our computers are **digital** — based on **binary** representations of information.

When we work with continuous signals, then, we're always **sampling** and **digitizing**) them.



digitizing

sampling

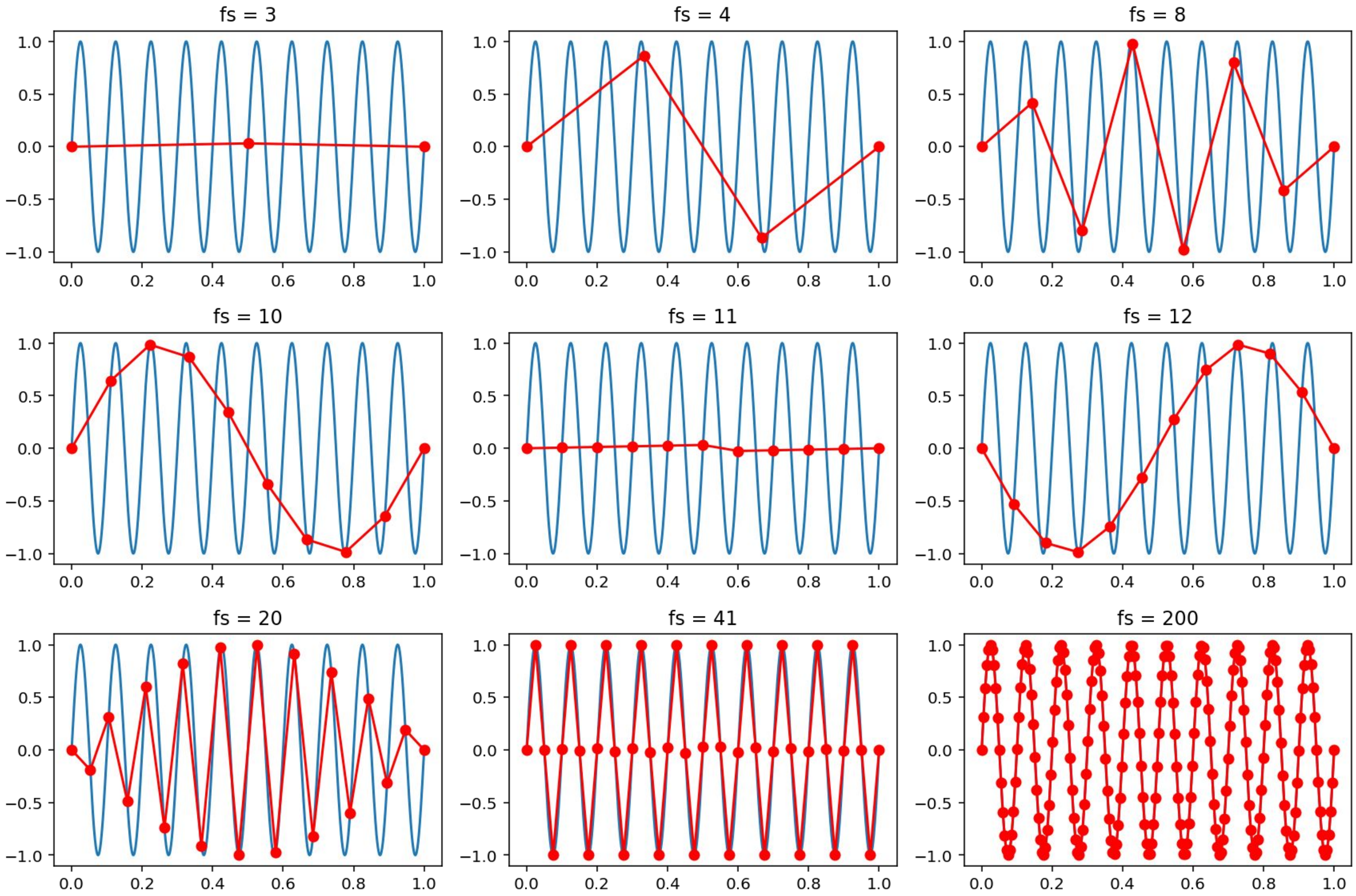
Important signal processing terms

Sampling frequency (or rate): how often was a sample collected?

- Lower for EEG (typically around 100 Hz), very high for spiking data — why?
- Trade off: higher sampling rate means bigger data!

Nyquist rate: to accurately represent a periodic signal with a frequency, you have to sample ***at least*** twice as fast, otherwise you will get aliasing

- Relatedly, we can only detect frequencies at half the sampling frequency: the **Nyquist frequency**

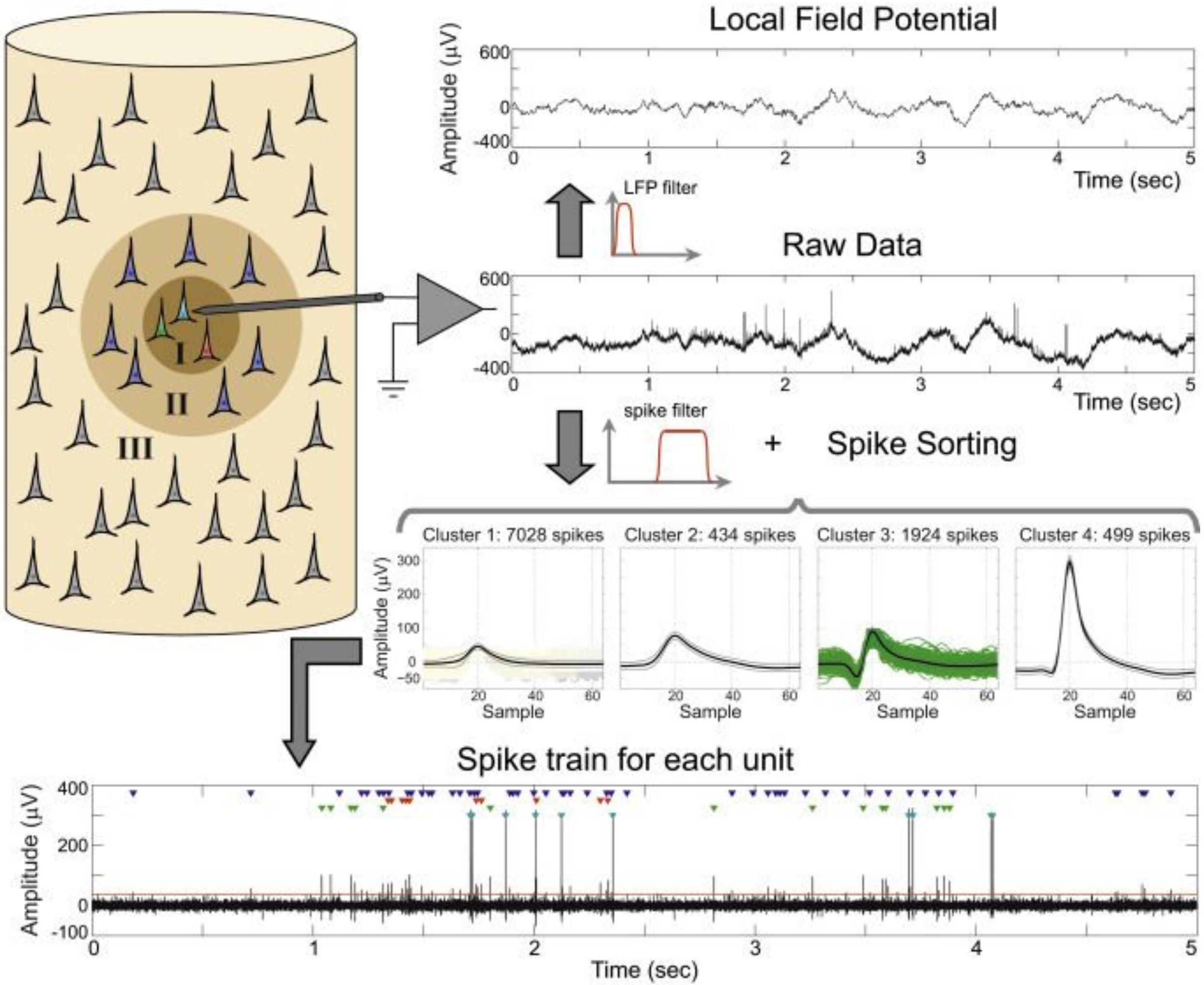


Blue = **the signal we're trying to measure**

Red = **our sampling of that signal**

fs = sampling frequency

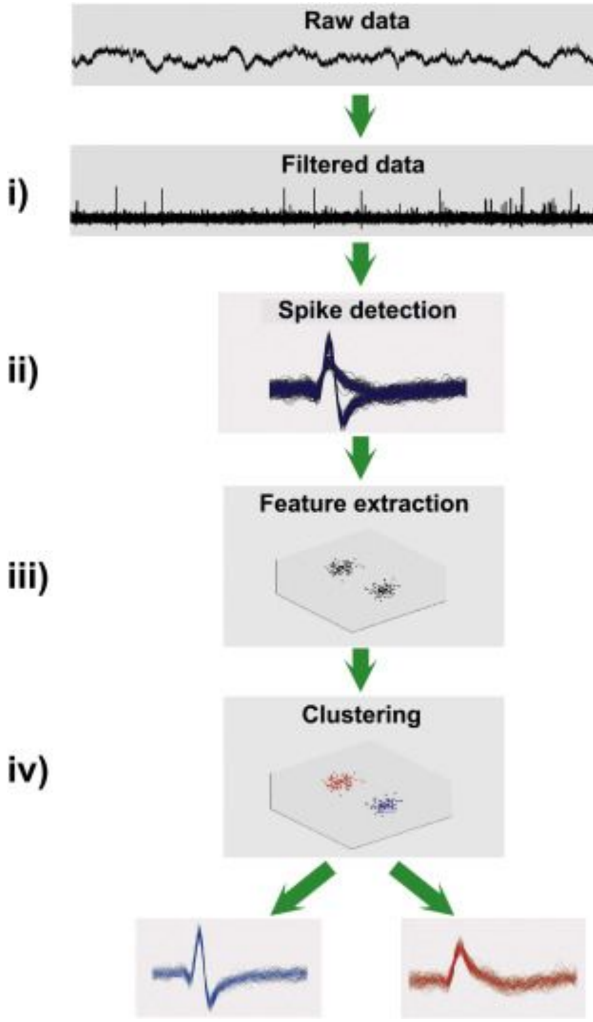
Extracellular recordings require spike sorting

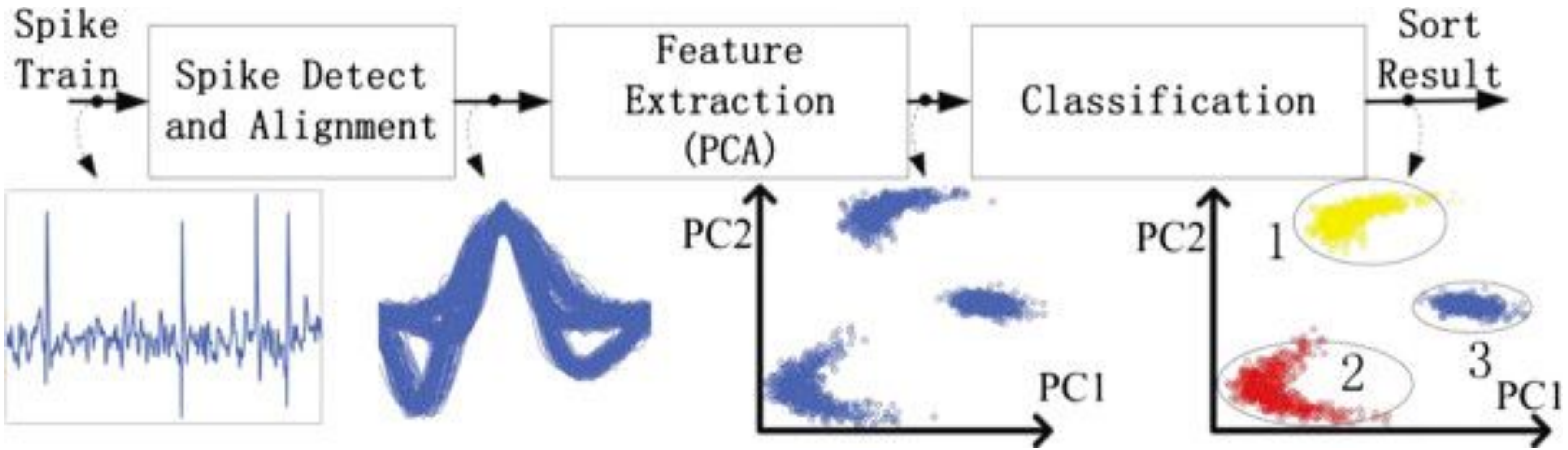


From Rey et al. (2015) "Past, present and future of spike sorting techniques"

Basic steps for spike sorting

1. a bandpass **filter** is applied, e.g., between 300 Hz and 3000 Hz, to keep the most useful part of the spectrum for spike sorting
2. **spikes** are detected, usually with an amplitude threshold applied to the filtered data
3. relevant **features** of the spike shapes are extracted via a dimensionality reduction technique (PCA)
4. those features are the input to a **clustering algorithm** that performs the classification of the waveforms and associate each cluster to a unit.





PCA is also commonly (traditionally) used for extracellular spike sorting. After projecting onto these dimensions, you can then use a clustering algorithm (or detect them manually...)

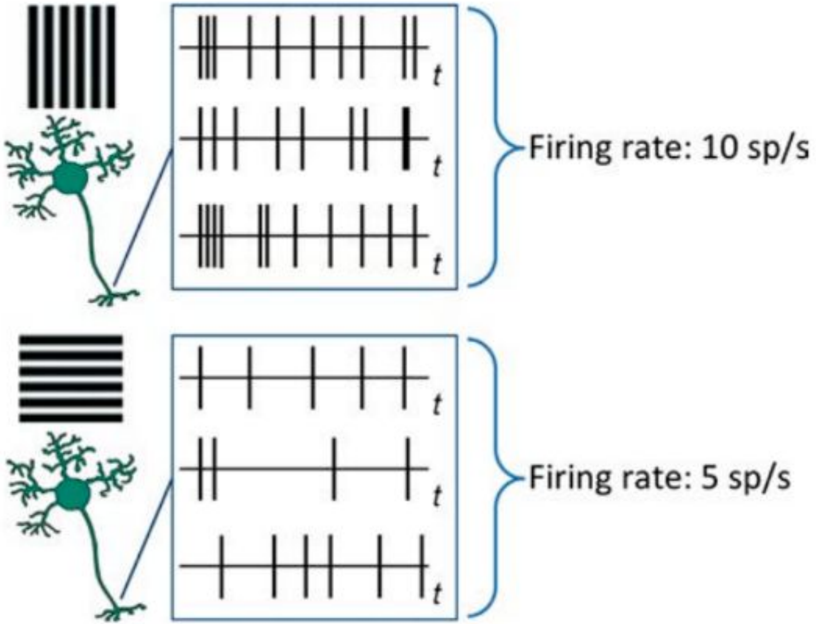
Spike Correlations

Correlations can provide information about the **functional architecture** of networks

For example, **connectivity in the retina** (Greschner et al. 2011),

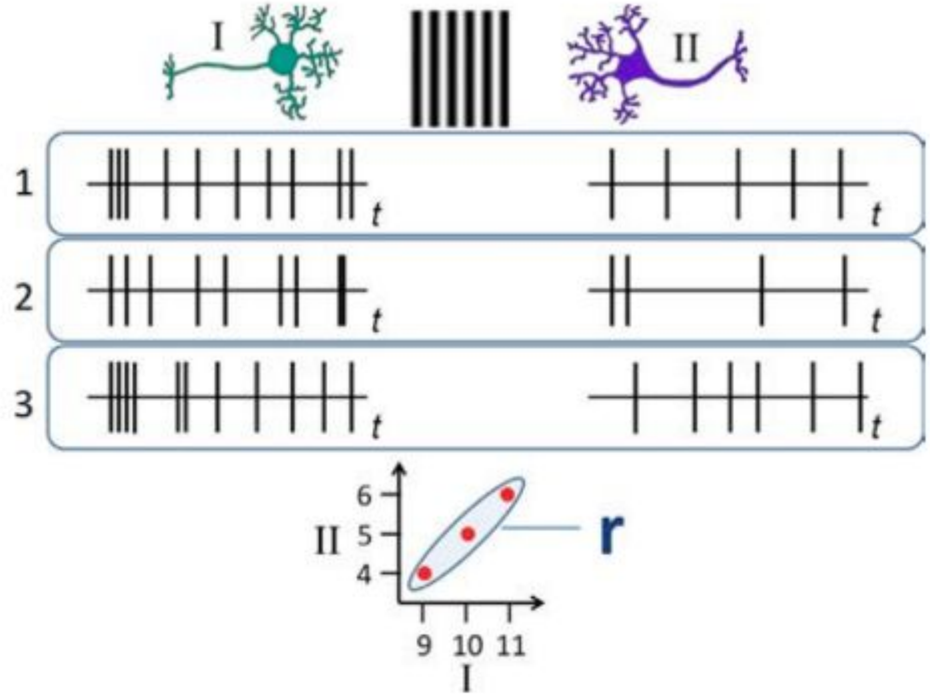
between the **visual thalamus and cortex** (Reid & Alonso, 1995),

and **between neurons in cortex** (Aertsen et al., 1989 & Alonso & Martinez, 1998)



Correlations between pairs of neurons can arise over a range of timescales

- Spikes may be precisely aligned in time, down to the millisecond
- Excitability may change on the order of seconds or longer



Neuroscientists often look for two different types of correlations: **signal** and **noise** correlations.

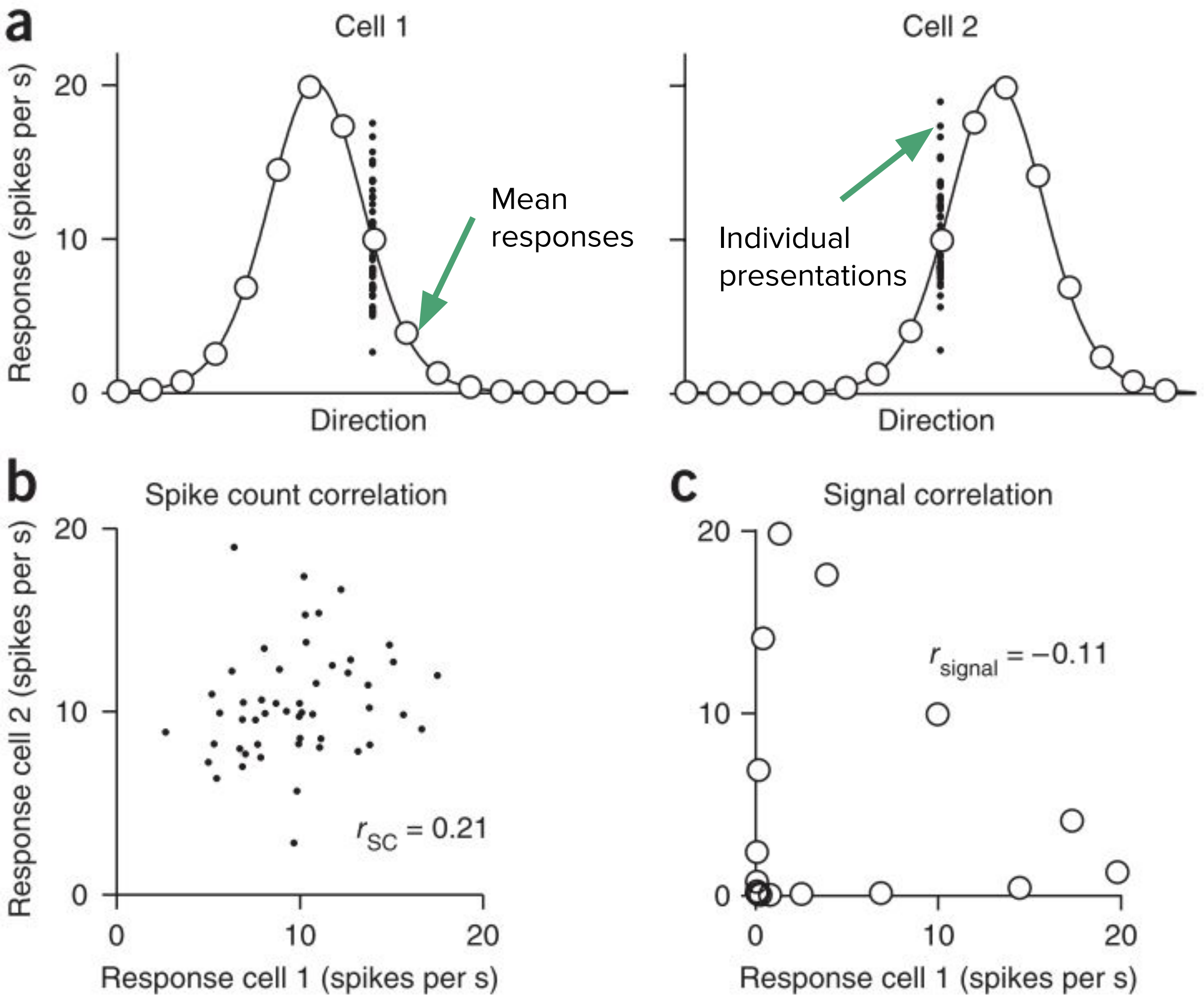
Small dots = individual presentations;
Big circles = mean responses

a. Tuning curves for two hypothetical cells

b. Spike count correlation (**noise** correlation) for those cells

c. Signal correlation

From Cohen & Kohn, 2011



Signal correlations

measures the correlation coefficient between the cells’ mean responses to different stimuli.

often used to quantify the extent to which a pair of neurons has similar tuning or other functional properties.

decreases in this type of correlation can lead to sparsening of population responses

Noise correlations

measures the correlation between responses to repeated presentations of identical stimuli, under the same behavioral conditions

“spike count correlations”

Noise, or **spike count correlations** (r_{sc}) tend to be small but positive

and tend to be higher for neurons that are near each other

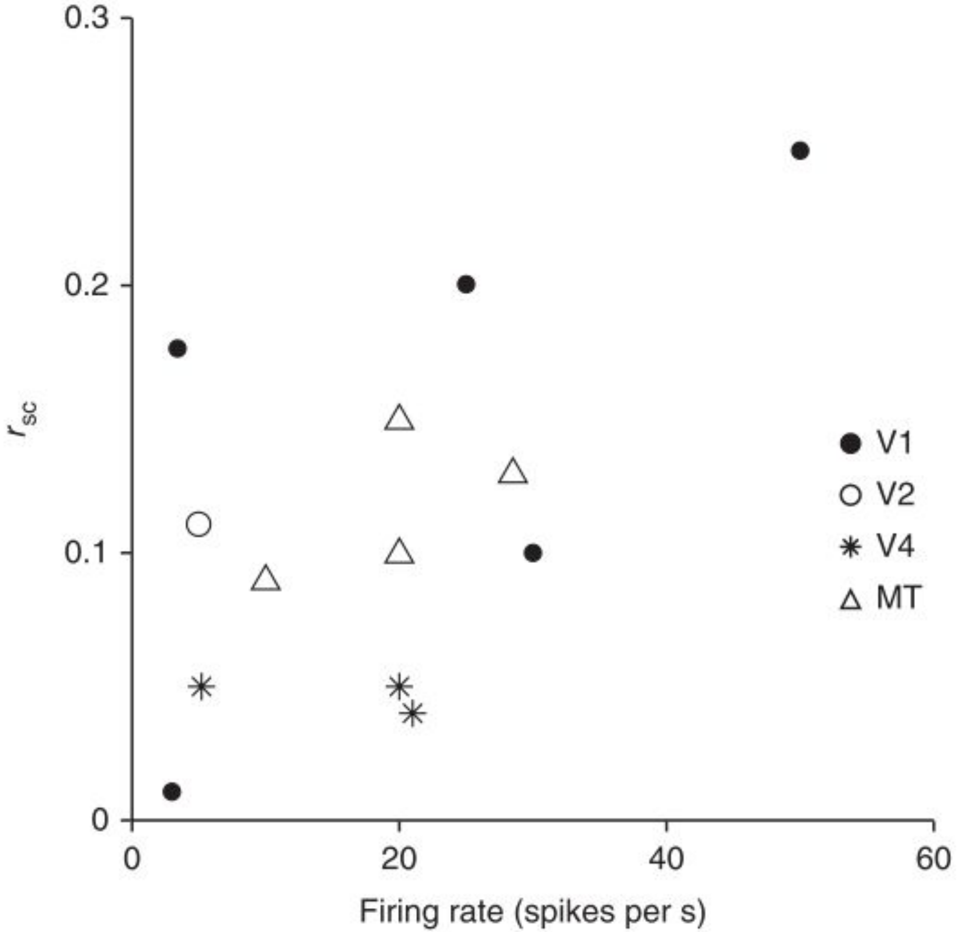
From Cohen & Kohn, 2011

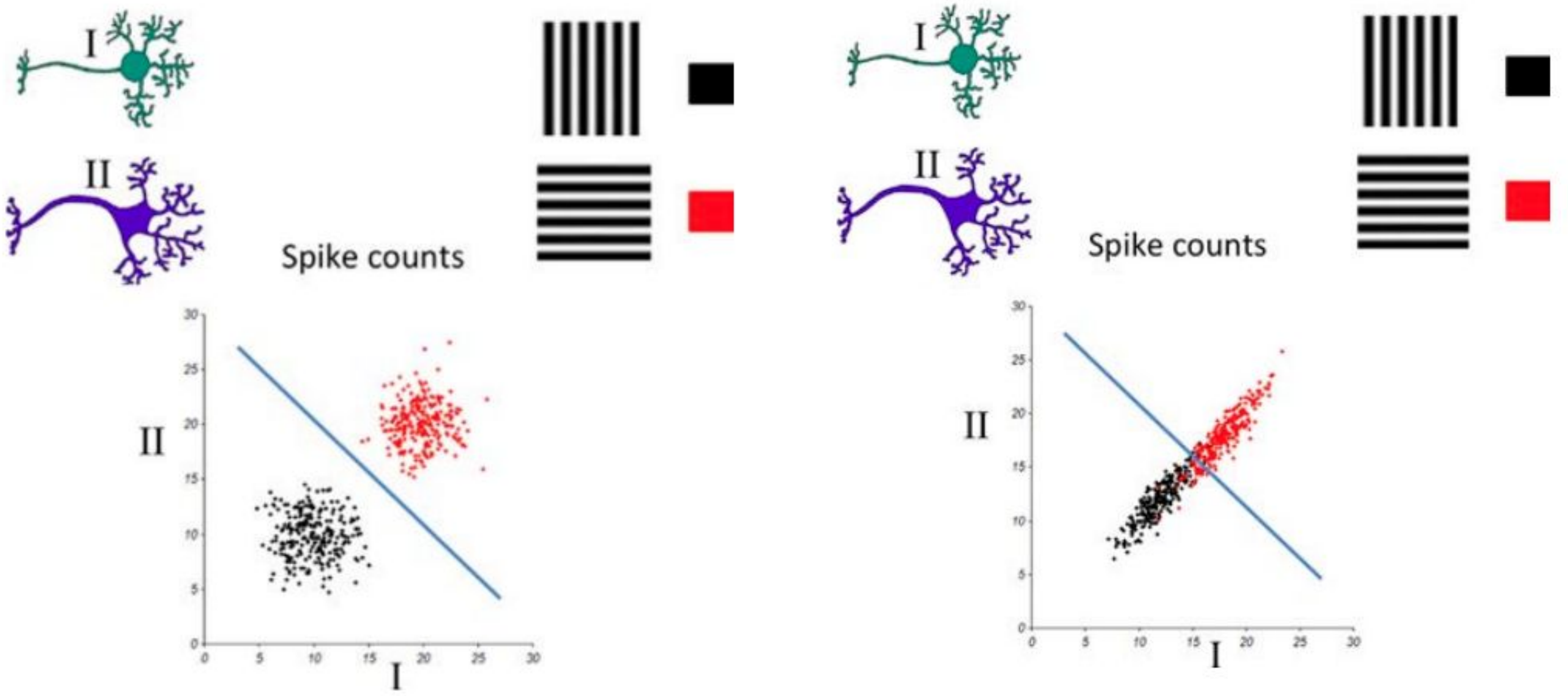
Table 1 Summary of studies measuring spike count correlations in primates

Reference number	Area	Firing rate (spikes per s)	Duration (ms)	State (task, anesthesia, etc.)	r_{SC}
12*	V1	~25	2,560	Anesthetized	0.2
26*	V1	3.4	1,280	Anesthetized	0.18
23	V1		1,894	Anesthetized	0.25
31	V1			Anesthetized	0.26
13*	V1	~50	1,860	Fixation	0.25
28*	V1	~3	500	Fixation	0.01
82	V1		400	Tracing	0.18
83*	V1	30	1,000	Discrimination	0.1
A. Zandvakili and A.K., unpublished data*	V2	5	1,000	Anesthetized	0.11
M. Smith and M. Sommer (University of Pittsburgh), personal communication*	V4	5.2	1,000	Fixation	0.05
7*	V4	21	200	Attention/detection task	0.04
8*	V4	>5, ~20	800	Attention/tracking task	0.05
A.B.G. Graf (New York University), personal communication*	MT	~10	300	Anesthetized	0.09
29*	MT	~20	500	Fixation	0.1
15*	MT	28.5	500	Discrimination	0.13
6/22*	MT	~20	1,000	Discrimination	0.15
84	Perirhinal	~12	200–500	Fixation/matching task	0.02
85	Supp motor area		66 or 200	Serial reaching	0.013
27	Supp motor area	~15	200	Reaching	0.02
86	Premotor areas	~5	400	Grasping/imagery task	0.02
87	M1	~20	600	Reaching	0.1–0.2
25	Motor/parietal; areas 2/5	~5	1,000	Reaching	0.02–0.04
88	Substantia nigra	58	500	Cue matching	0.01–0.04

Variability in r_{sc} across studies can largely be explained by differences in firing rate

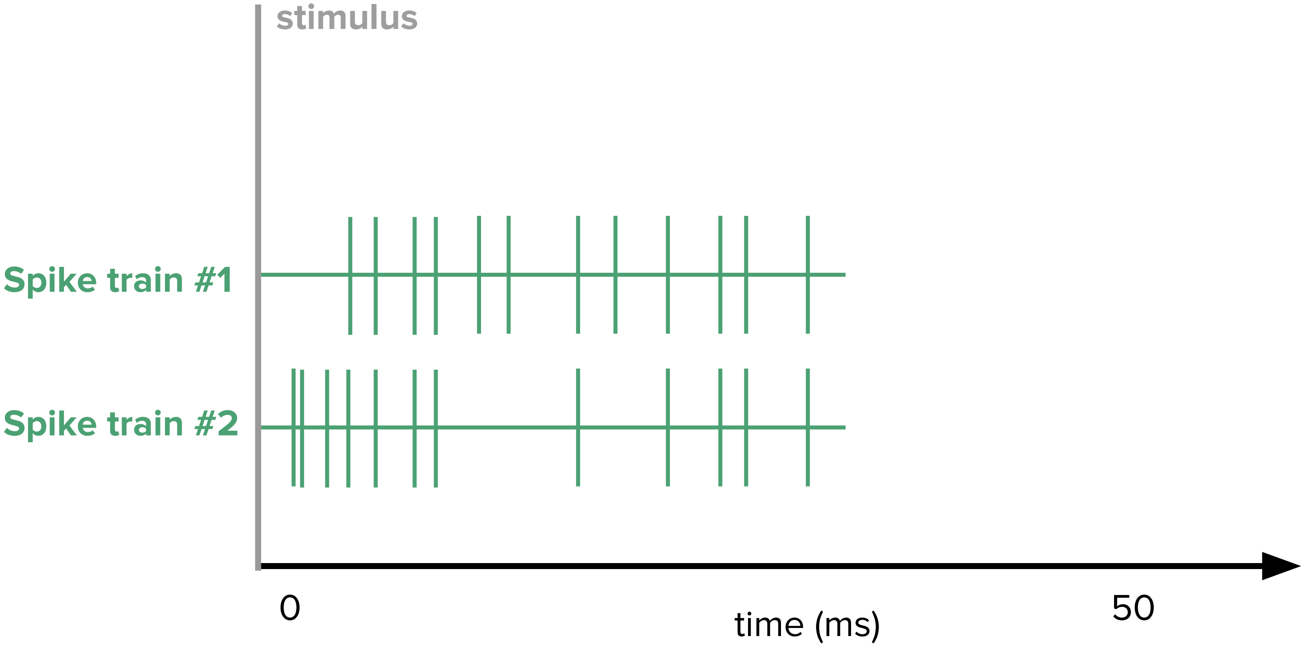
From Cohen & Kohn, 2011



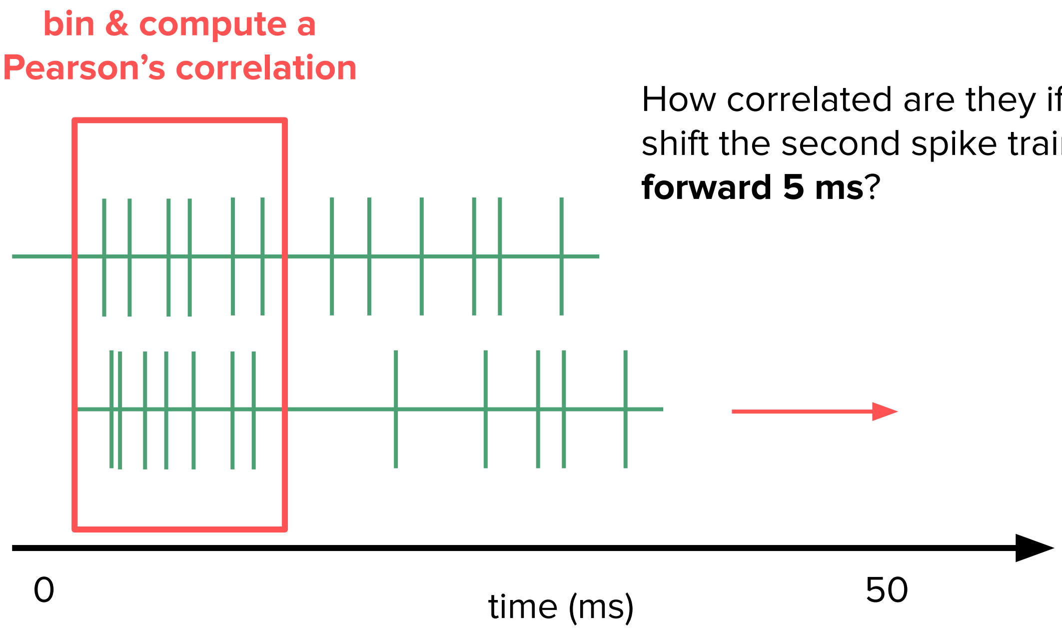


Spike count correlations change how well a population of neurons can decode a stimulus

We can also ask how correlated spike trains are, by shifting them forward and backward in time



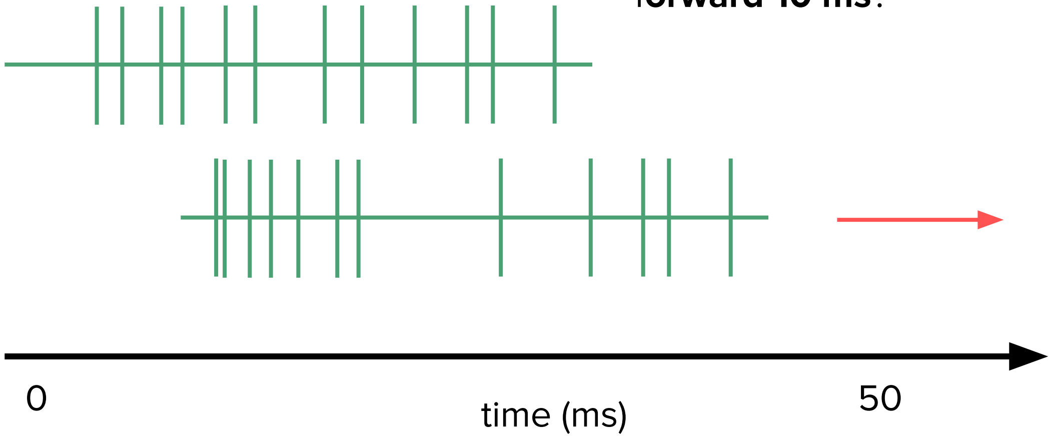
We can also ask how correlated spike trains are, by shifting them forward and backward in time



How correlated are they if we shift the second spike train **forward 5 ms**?

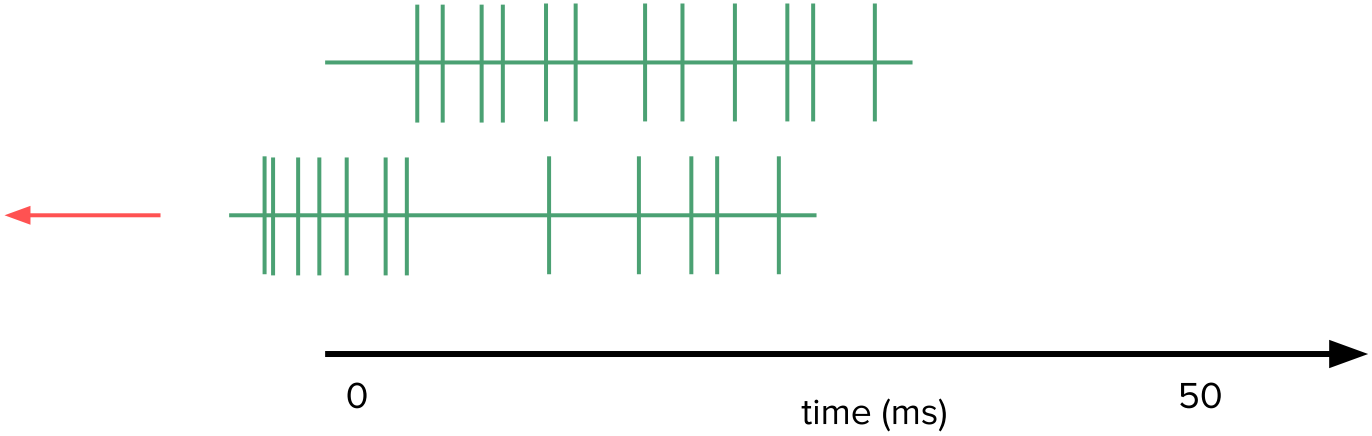
We can also ask how correlated spike trains are, by shifting them forward and backward in time

How correlated are they if we shift the second spike train **forward 10 ms**?

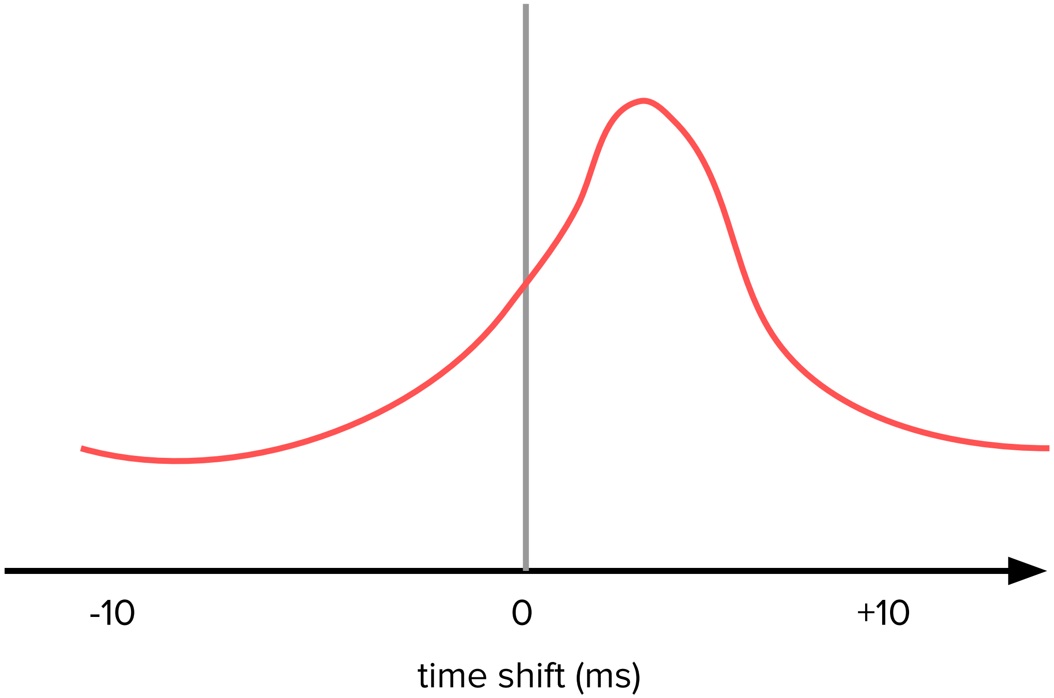


We can also ask how correlated spike trains are, by shifting them forward and backward in time

How correlated are they if we shift the second spike train **back 10 ms**?



We can also ask how correlated spike trains are, by shifting them forward and backward in time



nature communications

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature communications › articles › article

Article | Open access | Published: 28 January 2021

Mouse entorhinal cortex encodes a diverse repertoire of self-motion signals

Caitlin S. Mallory, Kiah Hardcastle, Malcolm G. Campbell, Alexander Attinger, Isabel I. C. Low, Jennifer L. Raymond & Lisa M. Giocomo

Nature Communications **12**, Article number: 671 (2021) | Cite this article

Abstract

Neural circuits generate representations of the external world from multiple information streams. The navigation system provides an exceptional lens through which we may gain insights about how such computations are implemented. Neural circuits in the medial

Caitlin Mallory (Postdoc @ UC Berkeley)

Lisa Giocomo (Professor @ Stanford)

Dataset that we'll explore today!

Additional Resources (Signal processing)

https://mark-kramer.github.io/Case-Studies-Python/03.html

https://voyteklab.com/oscillations/publications/interpreting-spectrum/

Related UCSD classes:

COGS 118C. Neural Signal Processing

DSC 120. Signal Processing for Data Analysis