COGS138: Neural Data Science

Lecture 13

C. Alex Simpkins, PhD
UCSD Dept. of Cognitive Science, Spring 2023
RDPRobotics, LLC
http://casimpkinsjr.radiantdolphinpress.com/pages/cogs138_sp23
rdprobotics@gmail.com | csimpkinsjr@ucsd.edu

Plan for today

- Announcements
- Review - Last time
- Connectionist views and assumptions discussion
- Single cell recording
- High density recording
- Spike sorting
- Neuropixels datasets and similar

Chat/AI/etc responses not acceptable

Announcements

- **Mid-quarter check-in survey assignment (required)**
- **Grade check-in - survey upcoming, to address any issues now and any concerns**
  - Check for missing quizzes (canvas)
  - Check for missing class participation (look over assignments page for completion of everything assigned, will import to canvas over the weekend if possible)
  - Check for data hub assignments (imported to canvas by weekend, check on data hub for now)
- **Assignment 3 - quick viewing**
- **Reading 3 - posted, due next Tuesday**
- **Lecture Quiz - posted, due Tuesday**
- **Previous project review:** https://forms.gle/f43RLeJyUgHX22HG7

Last time

Course links

Website	http://casimpkinsjr.radiantdolphinpress.com/pages/cogs138_sp23	Main face of the course and everything will be linked from here. Lectures, Readings, Handouts, Files, links
GitHub	https://github.com/drsimpkins-teaching	files/data, additional materials & final projects
datahub	https://datahub.ucsd.edu	assignment submission
Piazza	https://piazza.com/ucsd/spring2023/cogs138_sp23_a00/home (course code on canvas home page)	questions, discussion, and regrade requests
Canvas	https://canvas.ucsd.edu/courses/44897	grades, lecture videos
Feedback	https://forms.gle/gSEWi1ZKS4BLDPBH7	If I ever offend you, use an example you are uncomfortable with, or to provide general feedback. Please remain constructive and polite

About the final projects

Finding the project files

- Blank starter document for report, handouts and info to start with (draft): https://github.com/drsimpkins-teaching/cogs138/tree/main/main_project

- Links to old projects: https://github.com/drsimpkins-teaching/cogs138/tree/main/main_project/Wi2021

Distributions

- http://localhost:8888/notebooks/Documents/teaching/cogs138/old/Tutorials-master/Distributions.ipynb
- http://localhost:8888/notebooks/Documents/teaching/cogs138/old/Tutorials-master/Central%20Limit%20Theorem.ipynb
- http://localhost:8888/notebooks/Documents/teaching/cogs138/old/Tutorials-master/Correlation%20resampling.ipynb

Math and symbol review

- http://casimpkinsjr.radiantdolphinpress.com/pages/cogs138_sp23/handouts/greek_letters_review.pdf
- http://casimpkinsjr.radiantdolphinpress.com/pages/cogs138_sp23/handouts/math_review.pdf
- Handouts page on website:
    - http://casimpkinsjr.radiantdolphinpress.com/pages/cogs138_sp23/handouts.html

Demo in python…

Central tendency for neural data

Python docs on statistics

- Individual stats:
  - https://docs.python.org/3/library/statistics.html
- Comparisons:
  - https://github.com/drsimpkins-teaching/cogs138/blob/main/Tutorials-master/12-StatisticalComparisons.ipynb

Correlations and pitfalls in neural data science

Perspective taking, painting a picture with data, pitfalls on both sides

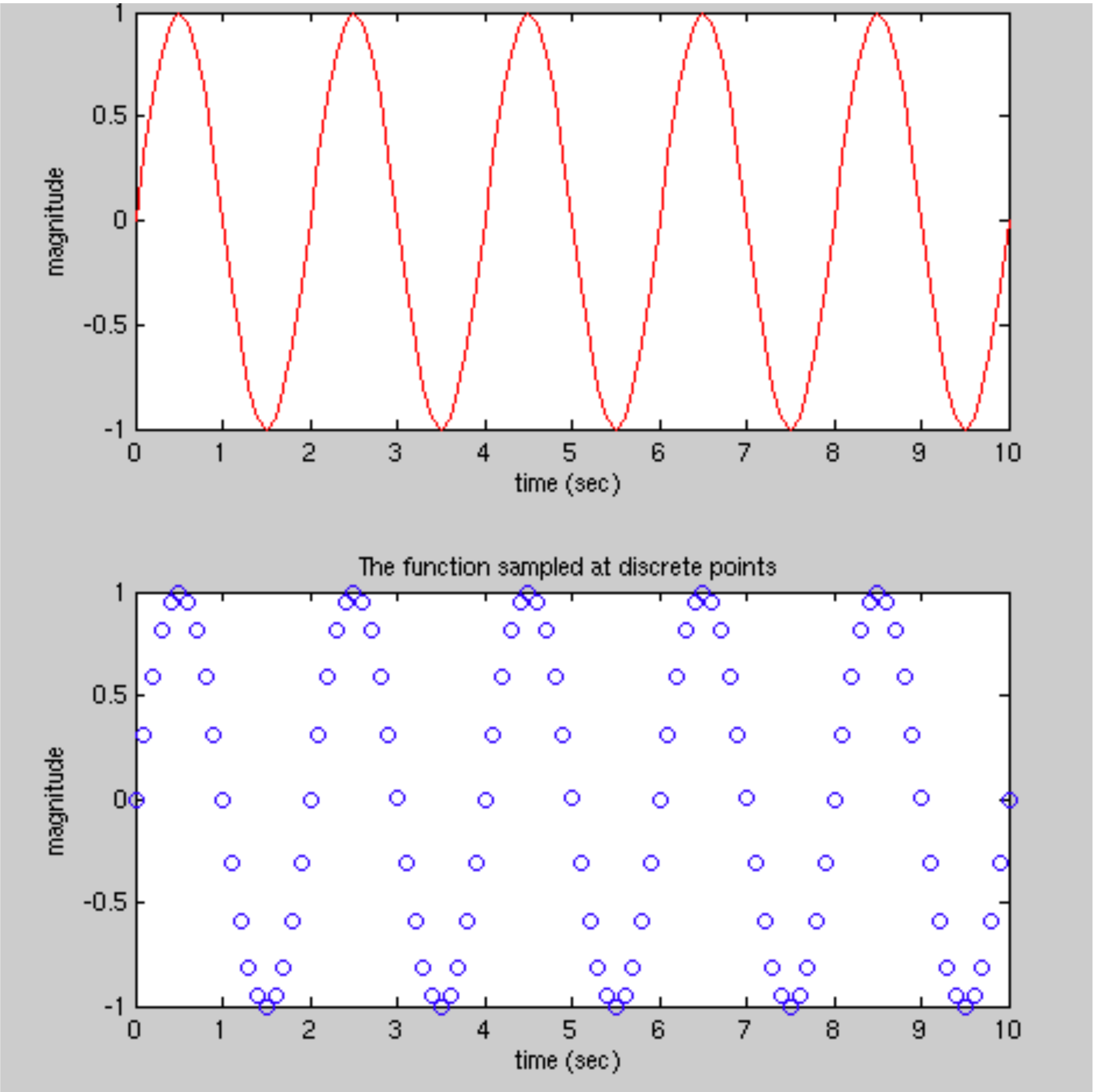
- Churchland's paper(s)
  - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3393826/
  - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6908571/
- Criticisms of Churchland's paper
  - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6908571/

More on sampling, discretization, filtering

- Review of continuous vs. discrete quantities
- Analog vs. Digital
- Discretization, sampling, aliasing
- Filter theory, frequency response, filter types
- Linearity

Continuous vs. Discrete quantities

- Information storage
  - **Continuous** signals have information at every point in time
  - **Discrete** signals have info only at specified intervals (fixed or variable)



Examples of continuous and discrete systems

- Continuous or discrete?
  - # of people in this class
  - # of Time zones
  - Time
  - Answers on multiple choice tests
  - A Sound
  - Body temperature

- Discrete
- Discrete
- Continuous
- Discrete
- Continuous
- Continuous

Analog vs. Digital quantities

- Information storage
  - **Analog** contains infinite information
  - **Digital** contains limited information, depending on the number of bits of information the digital value can store
    - 0 or 1 in each bit means each bit multiplies the possible combinations of numbers by 2
    - 2^4 = 0-15 (a 4-bit number, 16 different values)
    - 2^8 = 0-255 (an 8-bit number, 256 different values)
    - 2^16 = 0-65535 (a 16-bit number, 65536 different values)

More on digital quantities

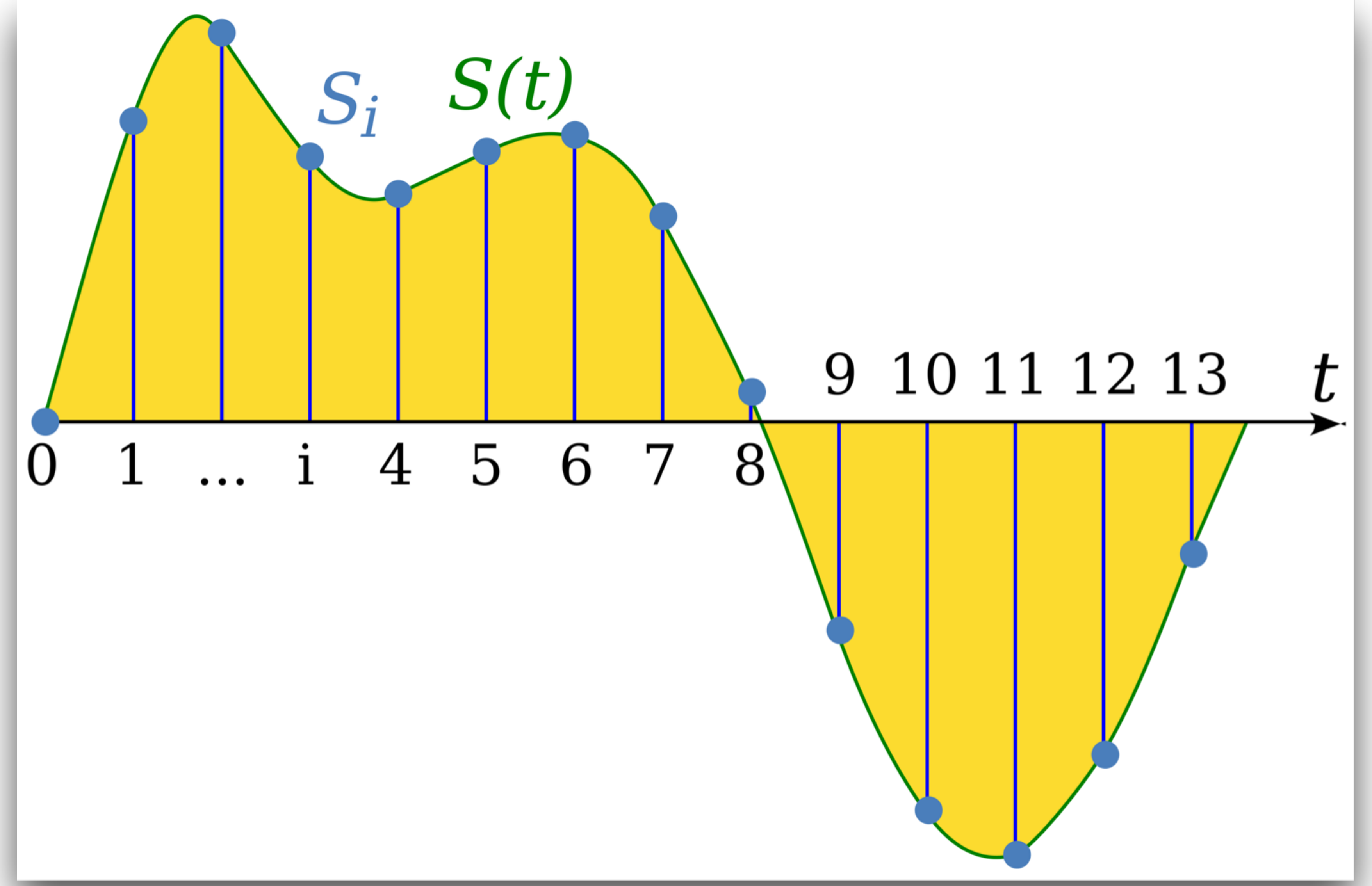
- Measuring an EEG boils down to recording a sequence of numbers into computer memory, stored in values of a specific size, such as 8 bit numbers.
  - i.e. signal is 0-5V, digitized with 8 bit ***precision*** would yield a ***resolution*** of 5V/256 = 0.020V, or 20mV (mV = ‘milli-Volts’)
  - **Resolution -** defined as the smallest quantity which can be reliably measured
  - **Digital Precision -** The number of bits of information contained in a digital quantity
- Also important for computations
  - Round off errors can accumulate
    - Example
      - 2.245+3.432+1.234 = 6.911
      - 2+3+1 = 6, and that’s only 3 samples! Imagine 1000/sec (1kHz) !
  - More on this later

Discretization

- Measuring a continuous (analog) signal means capturing information at specified (fixed or variable) intervals
  - **Sampling frequency -** the frequency at which data is recorded from a signal (Typically in Hz, ie 5kHz)
- When capturing data, or when manipulating data which has been discretized, there are several issues to consider
  - Aliasing (not the TV show:)
  - Sampling rates
  - Post-processing – filtering data to remove unwanted information while retaining desired information

Sampling

- **Sample -** We record data at specific points in time
- **Period -** The time between samples, ***T* [sec]**
- **Sample frequency -** The frequency of sampling, *f* [Hz]



$$f = \frac{1}{\Delta T}$$

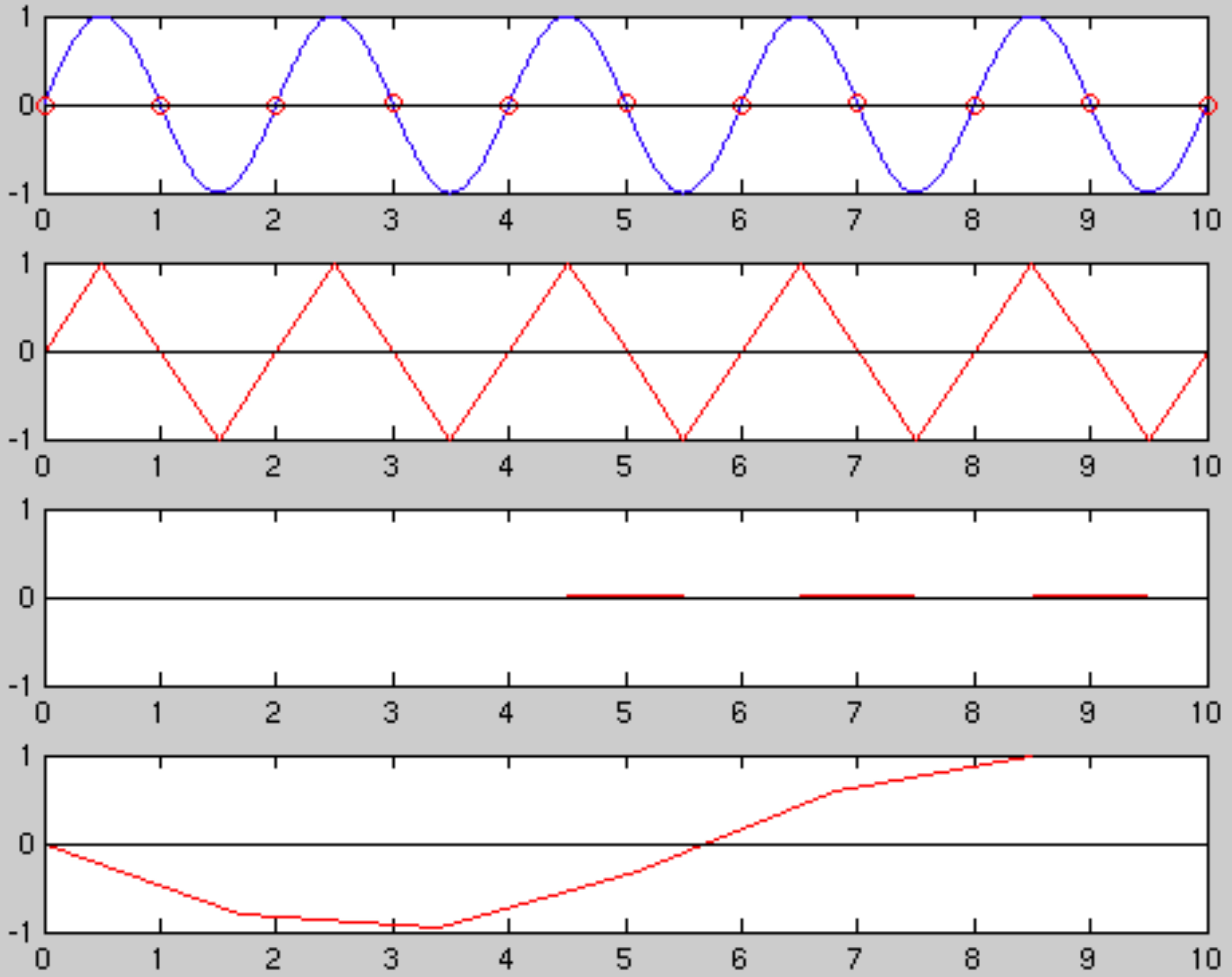
https://en.wikipedia.org/wiki/Sampling_%28signal_processing%29

Nyquist and Sampling

- Stories
  - Running in the dark with periodic lights on the ground, with sharp turns
  - Ping pong (no sound, periodic view of the system)
- As a rule of thumb, you must sample AT LEAST twice as fast as the highest frequency you want to measure
  - **Nyquist frequency -** max freq. that can be measured [Hz]
  - **Nyquist rate -** sampling frequency (which is 2x the nyquist frequency) required to sample at the nyquist frequency
  - 20 times as fast is better
  - Filter out higher frequency components

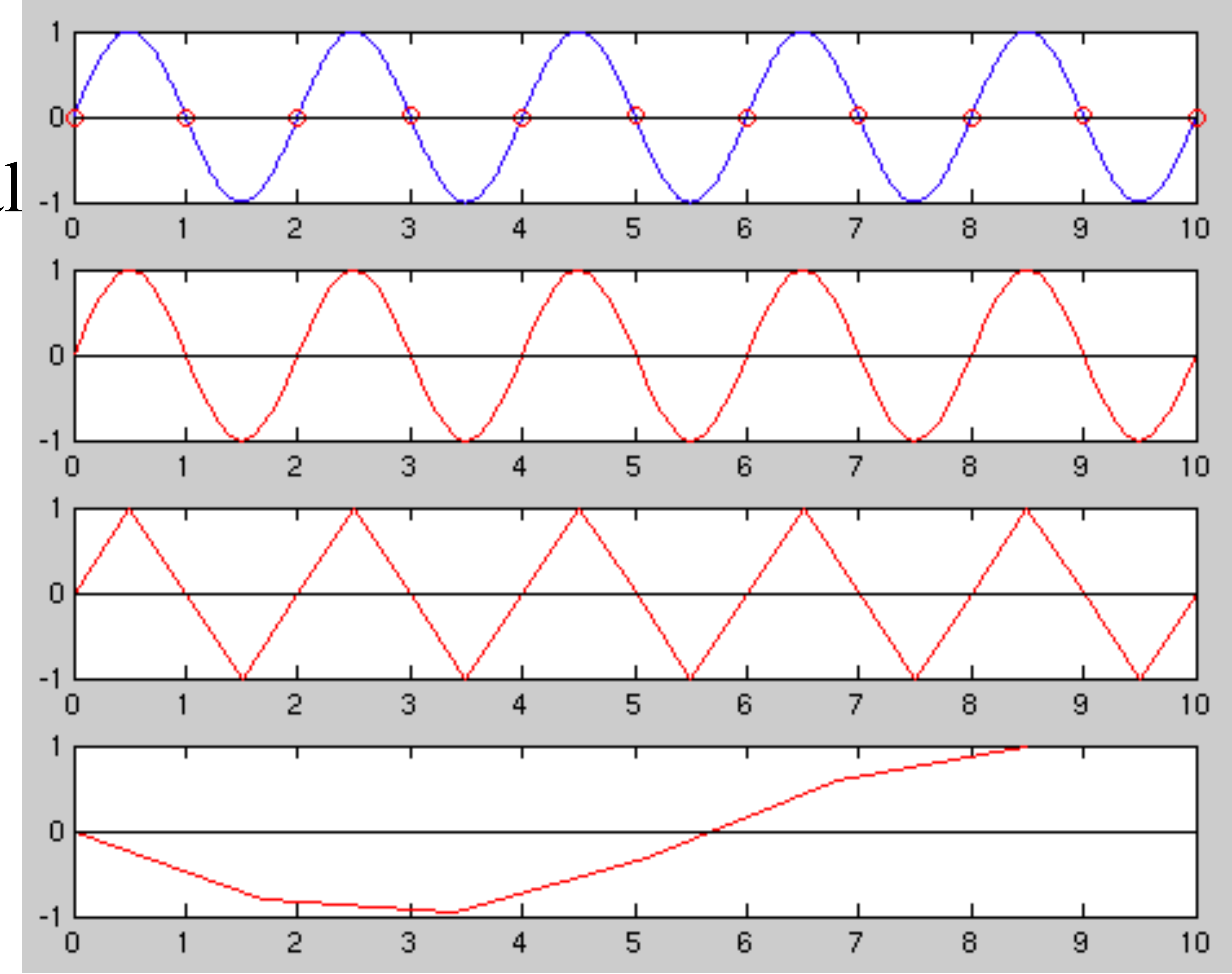
Nyquist frequency

$$\boxed{f_n = \frac{1}{2\Delta T}}$$


1
0
-1
0 1 2 3 4 5 6 7 8 9 10

What do we see in this picture?

- **Aliasing -** the corrupting of a signal by components of higher frequencies overlapping into the lower frequency



How do we solve this?

- Filter out the frequencies we don’t want
  - Low pass filter
  - High pass filter

Examples: Visual discretization

- Color shading

- Color and visual boundaries:

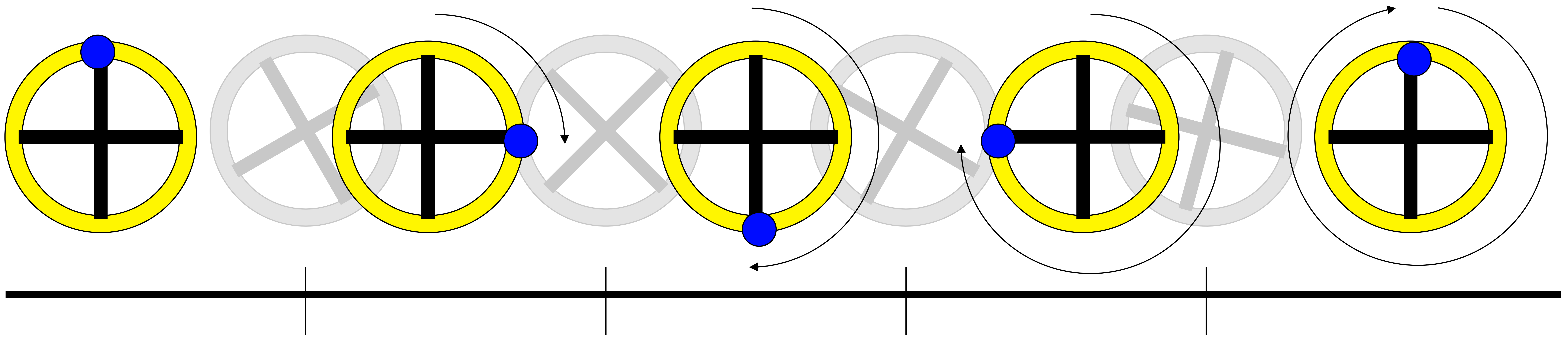
Few colors and low spatial resolution

Low spatial resolution only

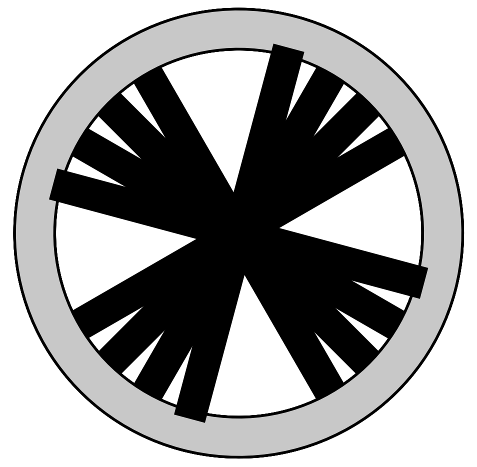
High spatial resolution and colors

Example: Sampling and Aliasing

- The wheel spokes example…<Live demo>



- We're sampling at too slow a rate to accurately see the spokes rotate, and at a particular rotational velocity of the wheel, we see an 'aliased' reverse rotation!

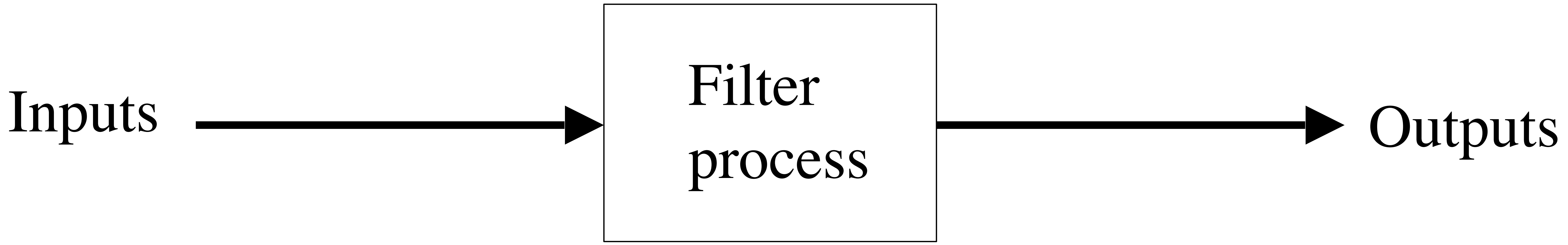


Obviously aliasing can be bad…

- Aliasing can lead to improper interpretations of data
  - **So what do we do about it?**
    - We must first sample at twice the rate of the fastest signal we care about
    - Filter our data (humans do this, and so do cognitive scientists!)

Thus we filter our data…

- **Filter** - an operation or process which alters input data according to some mathematical relationship or heuristic rule to produce output data which is more desirable



Computational filtering

- *Noisy auditory data can be filtered to remove undesired signals*
- *EEG signals can be filtered to remove 60Hz noise from AC lines nearby*
- *Other sensor signals can be filtered to improve results*

Frequency Response

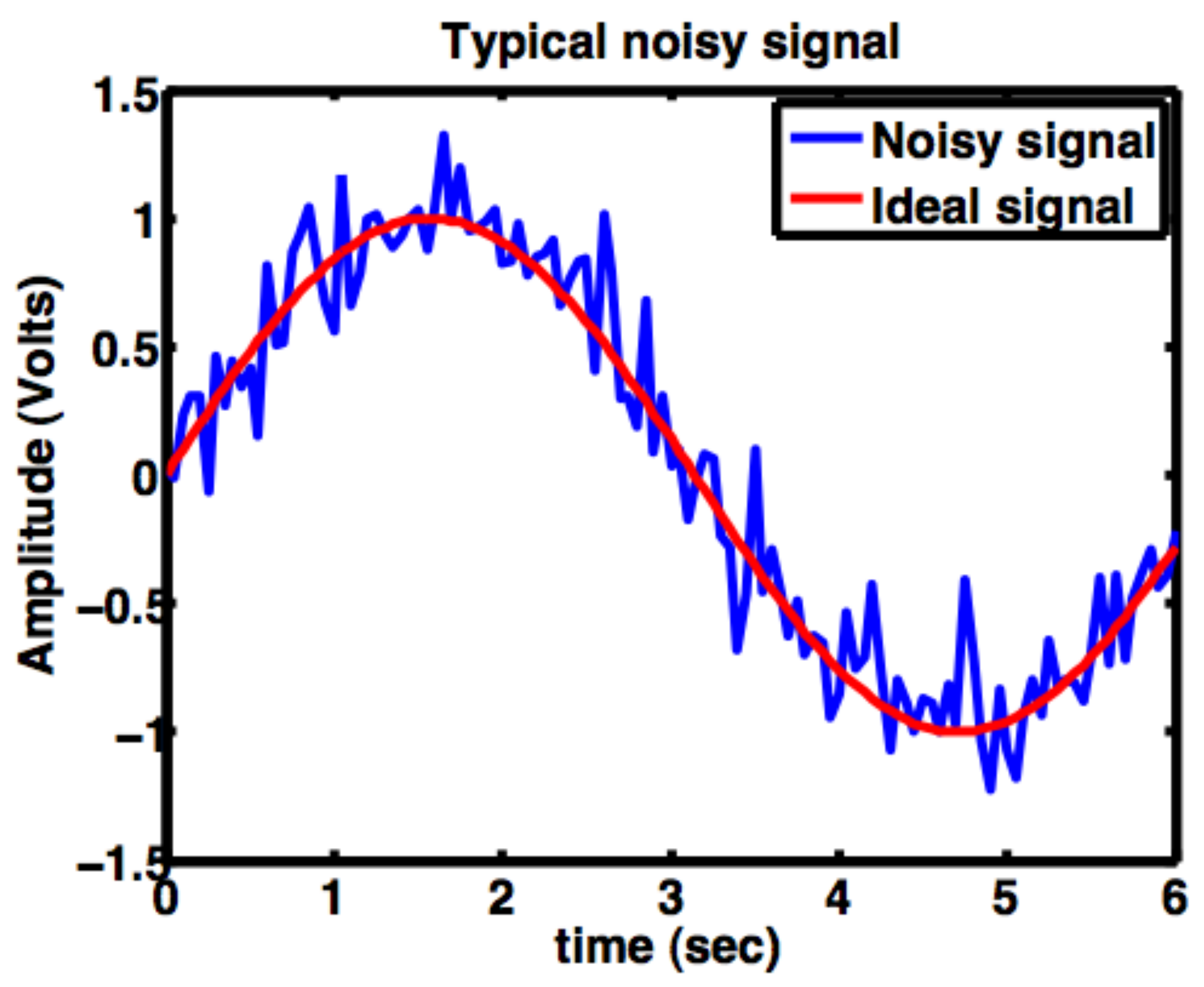
- Linearity of systems vs. nonlinearity
- The response of a linear system to a sinusoidal input is a sinusoidal output with the amplitude and phase shifted in some way
- This is useful for characterizing the behavior of some signal over a range of possible input frequencies
- Example with the chalk

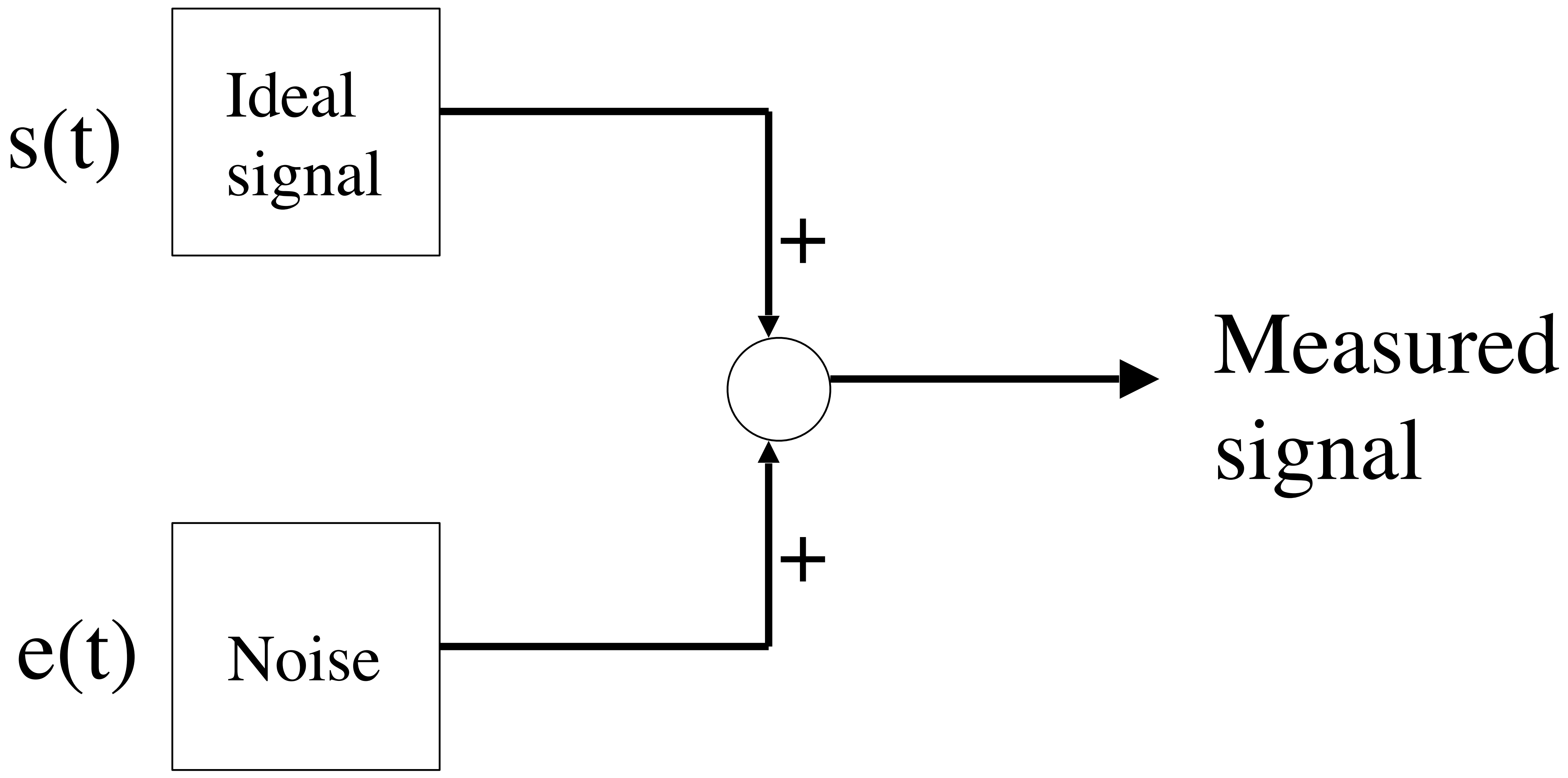
Common filter types in signal processing

- **Low-pass filter** - (ideal) attenuates high frequency data, while allowing low frequency data to pass unchanged
- **High-pass filter** - (ideal) attenuates low frequency data, while allowing high frequency data to pass unchanged
- **Band-pass filter** - (ideal) attenuates all frequencies except a particular frequency band (or bands)
- **Band-stop filter** - (ideal) attenuates one or a selection of frequency ranges of data, allowing all the rest to pass unchanged
- Actual filters are not exactly ideal…which we will discuss

Signals and noise…

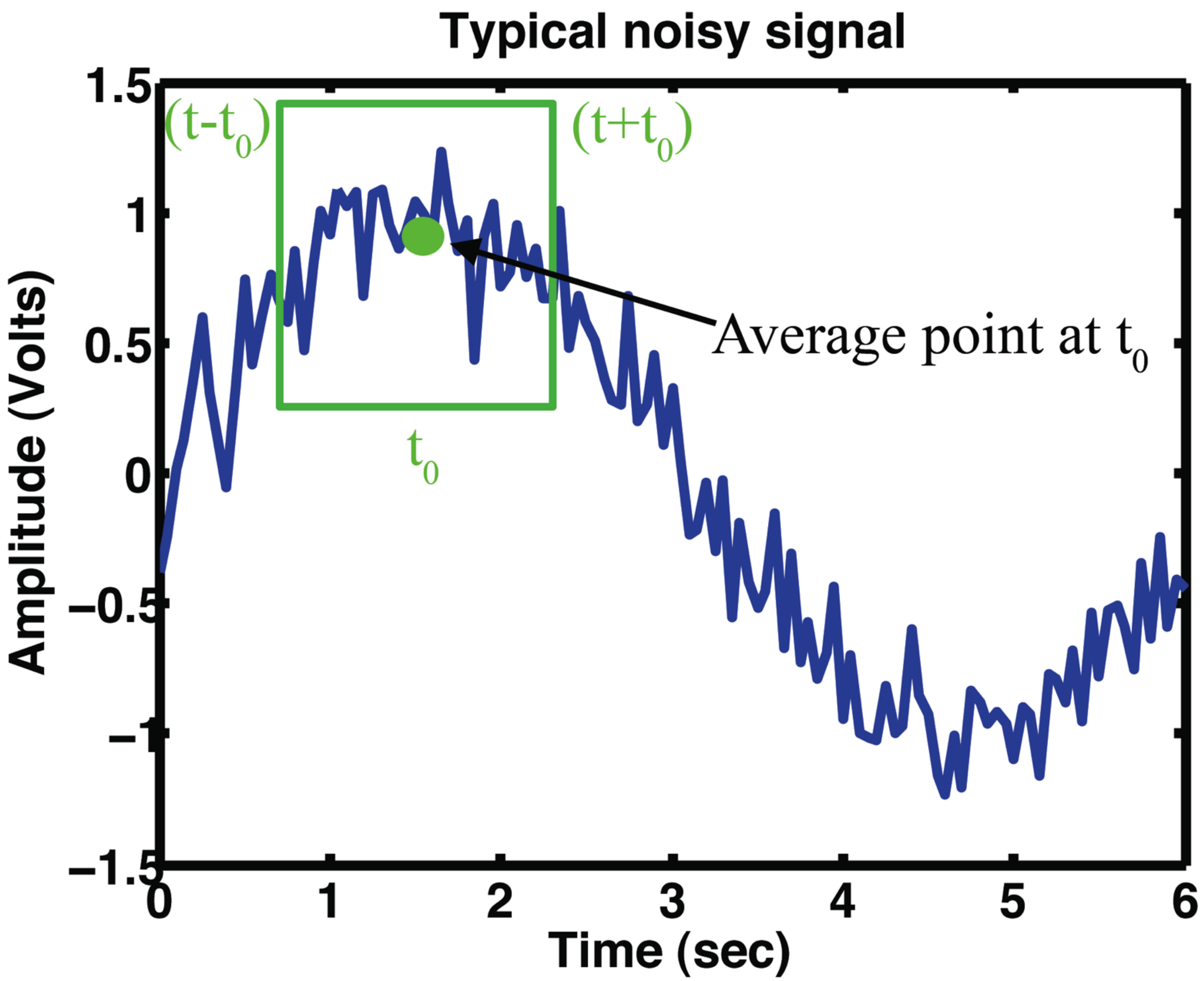
- By making assumptions about the properties of the unwanted 'noise' *e(t)*, we can reconstruct an appropriate *estimate* of the original signal *s(t)*
  - **Noise - any unwanted portion of a signal, lumped together. It may come from multiple sources but tends toward some statistically predictable properties**





Low-pass filtering

- So the effect is this



More on linearity vs. nonlinearity

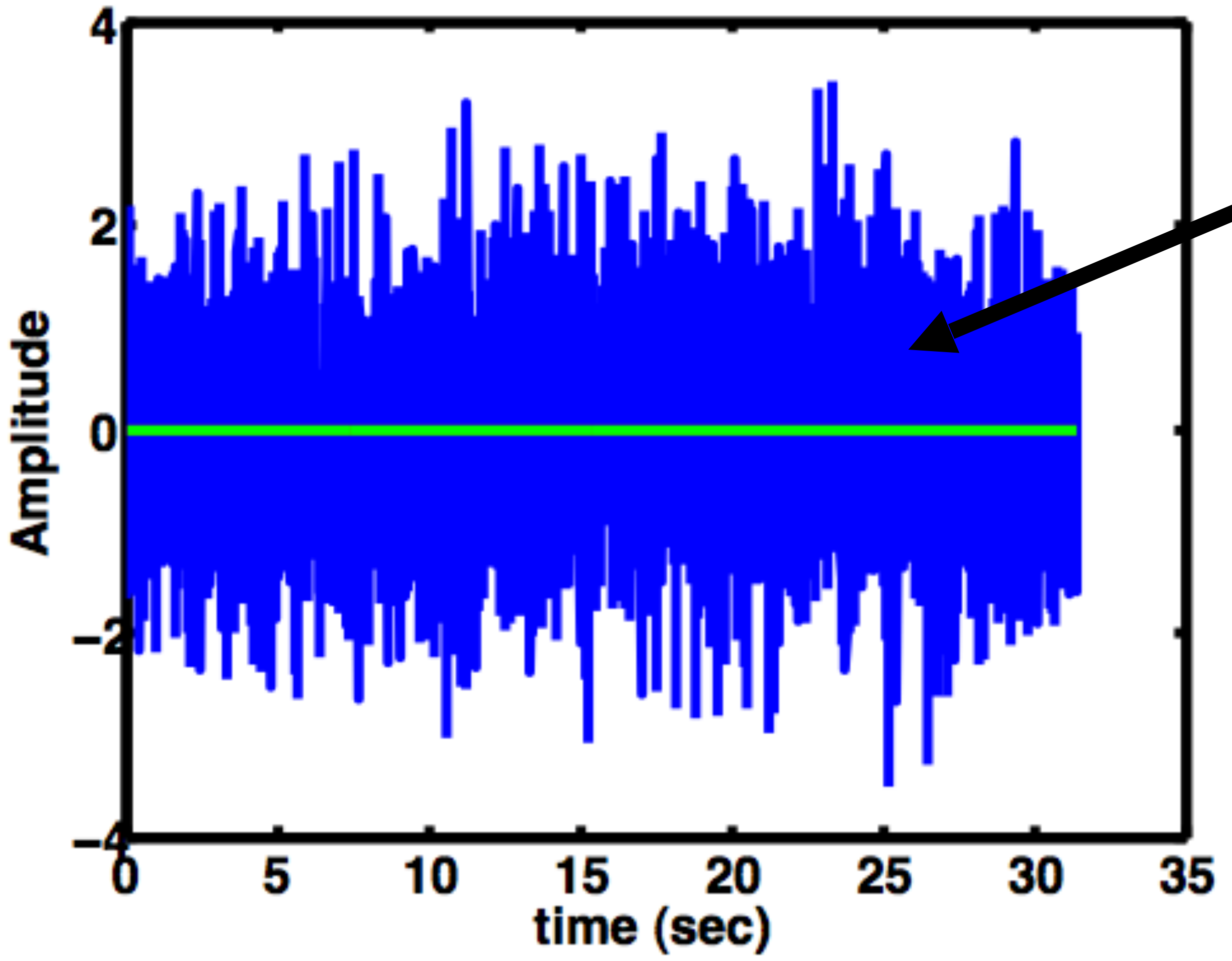
- Power
  - **A linear system is a system whose dependent variables are related to its independent variables by a power of one**
- Linear systems have these particular properties (and they are very favorable)
  - **Additive** $T[x_1(n) + x_2(n)] = T[x_1(n)] + T[x_2(n)]$
  - **Homogeneous** $T[cx(n)] = cT[x(n)]$
- Linear differential equations are more well-understood than nonlinear differential equations

Fourier transforms

- Frequency domain example : Musical note vs. the sound
  - **More parsimonious to describe a song in terms of its notes than time domain signal (when creating a ‘model’ for a song which can be communicated)**

We return to noisy data which we want to ‘clean up’

- We do this by removing undesired components of the signal
- One way to do this is *averaging* out the noise
- If it’s Gaussian and additive…

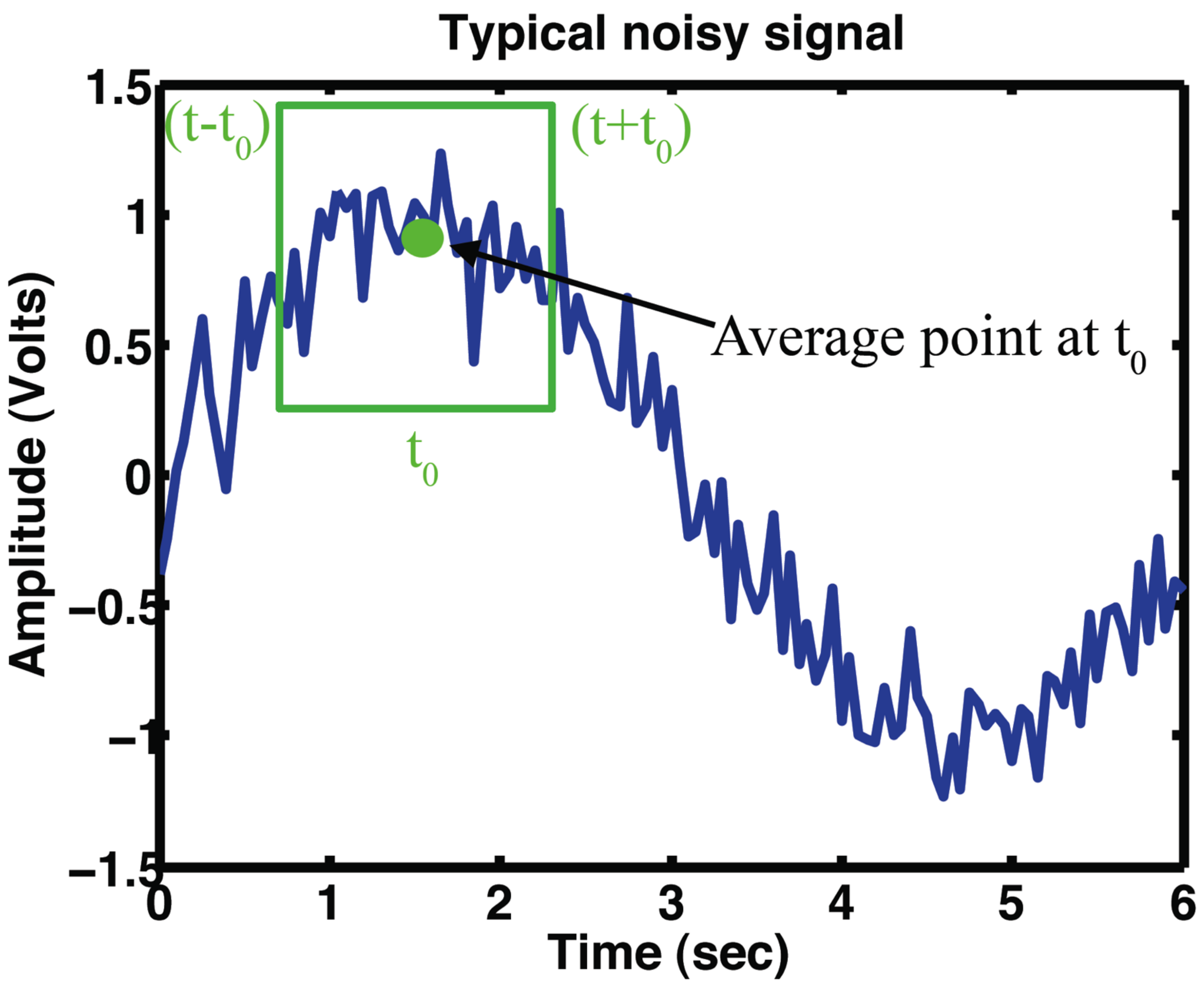


This is gaussian noise, and the average of this is approximately the green line, 0

$-5 + 5 = 0$

How to do it

- **Decide on a 'window' of data to average over, which is narrower than the fastest component to your changing signal**
- **Sum up over that window of points and divide by the number of points (average)**



Continuous form

$$x_f(t) = \int_{t-t_0}^{t+t_0} x(\tau)d\tau$$

Discrete form

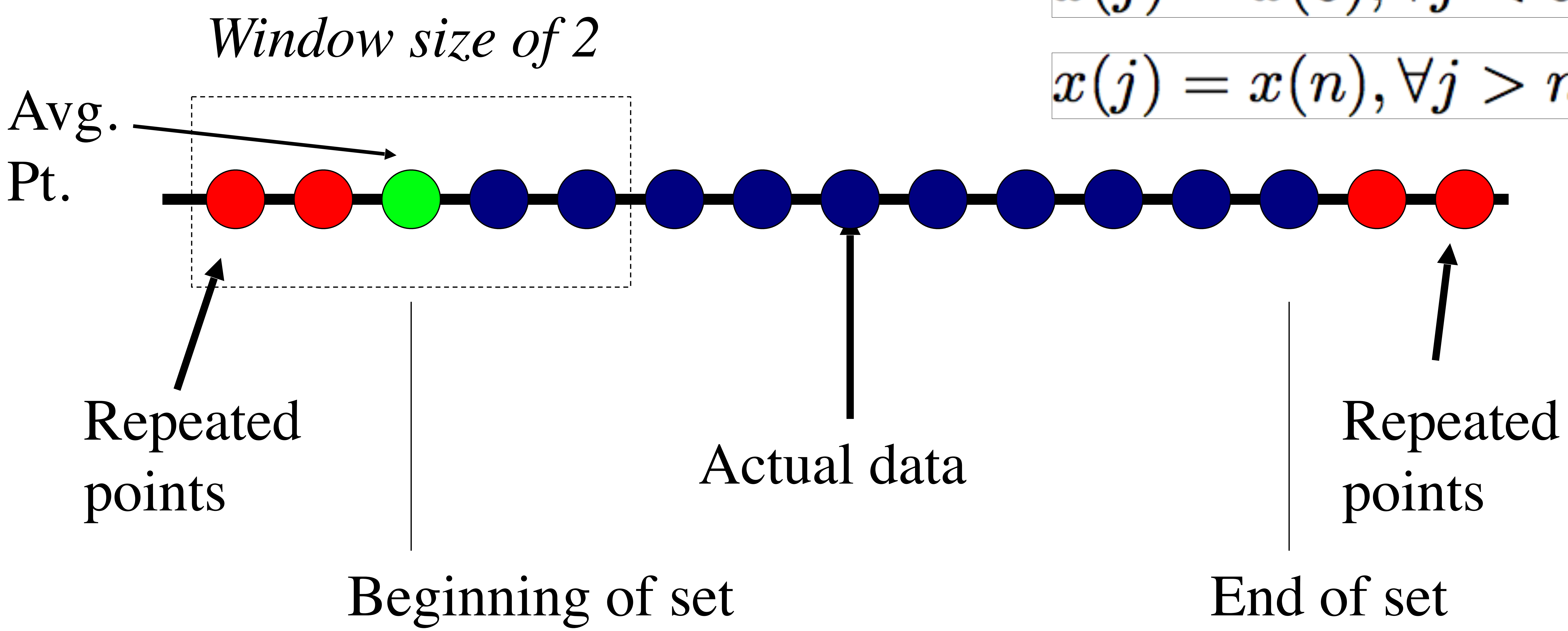
$$x_f(i) = \frac{1}{2k+1}\sum_{j=i-k}^{i+k} x(j)$$

A few details

- What about at the ends of the data where we don't have information before (at the beginning of the data set) or after (at the end of the data set)?
  - **Copy the first or last point and repeat as necessary**

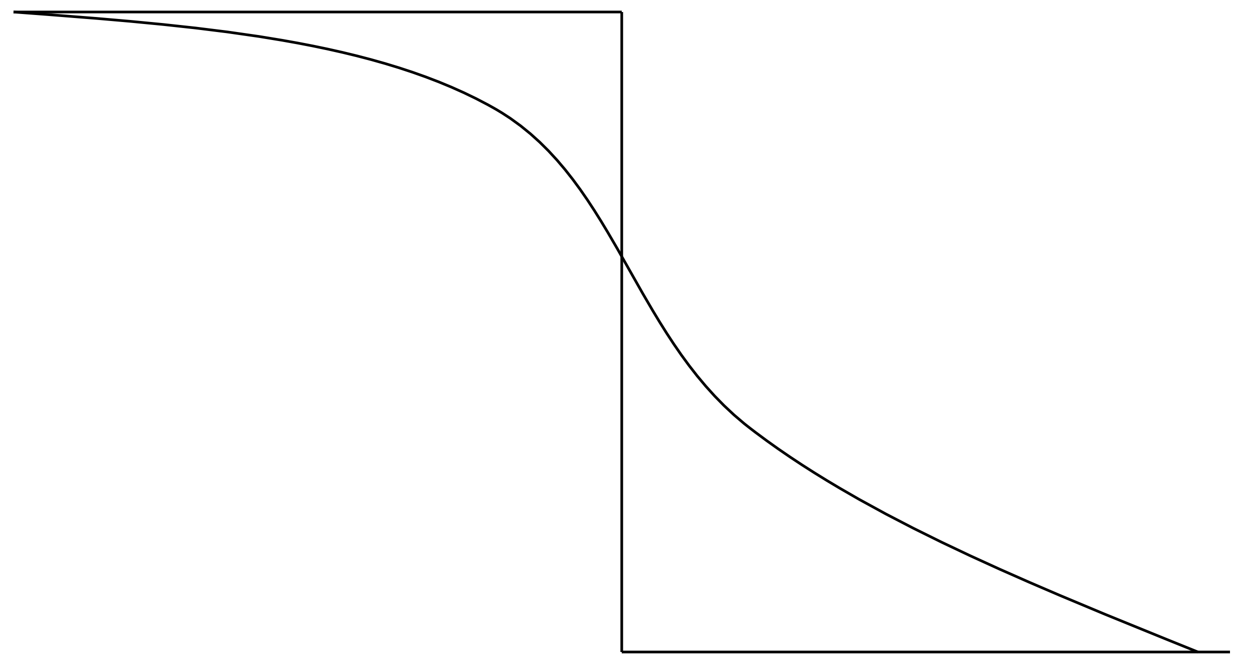
$$x(j) = x(0), \forall j < 0$$

$$x(j) = x(n), \forall j > n$$



Disadvantages…

- Need to have all data in memory already, so it isn't an 'online' filter
- Causality
  - **If we care about an exact event timing, this is a poor filter to use:**



Signal anticipates changes!

Solution

- Recursive filter
    - Solves causality issue
    - Easy to implement as we saw last time

On to today…

Physicalist perspective

- Neuroscience perspective
- Other perspectives
- Which is ‘true?’
- Does it matter?
- Animal model assumption of mapping

Underlying dynamics need to be exposed

- Tacoma narrows bridge disaster
  - 1st order vs. higher order
  - https://en.wikipedia.org/wiki/Tacoma_Narrows_Bridge_(1940)

https://en.wikipedia.org/wiki/Tacoma_Narrows_Bridge_(1940)#/media/File:Opening_day_of_the_Tacoma_Narrows_Bridge,_Tacoma,_Washington.jpg

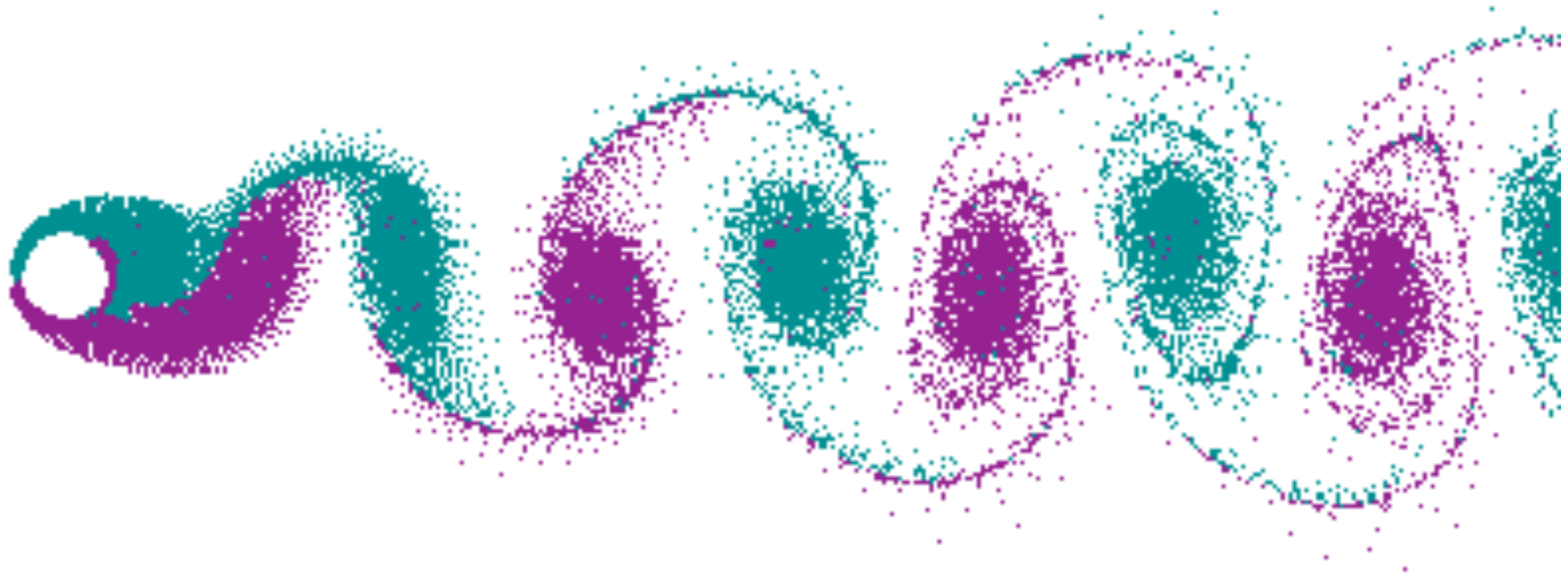
Underlying dynamics need to be exposed

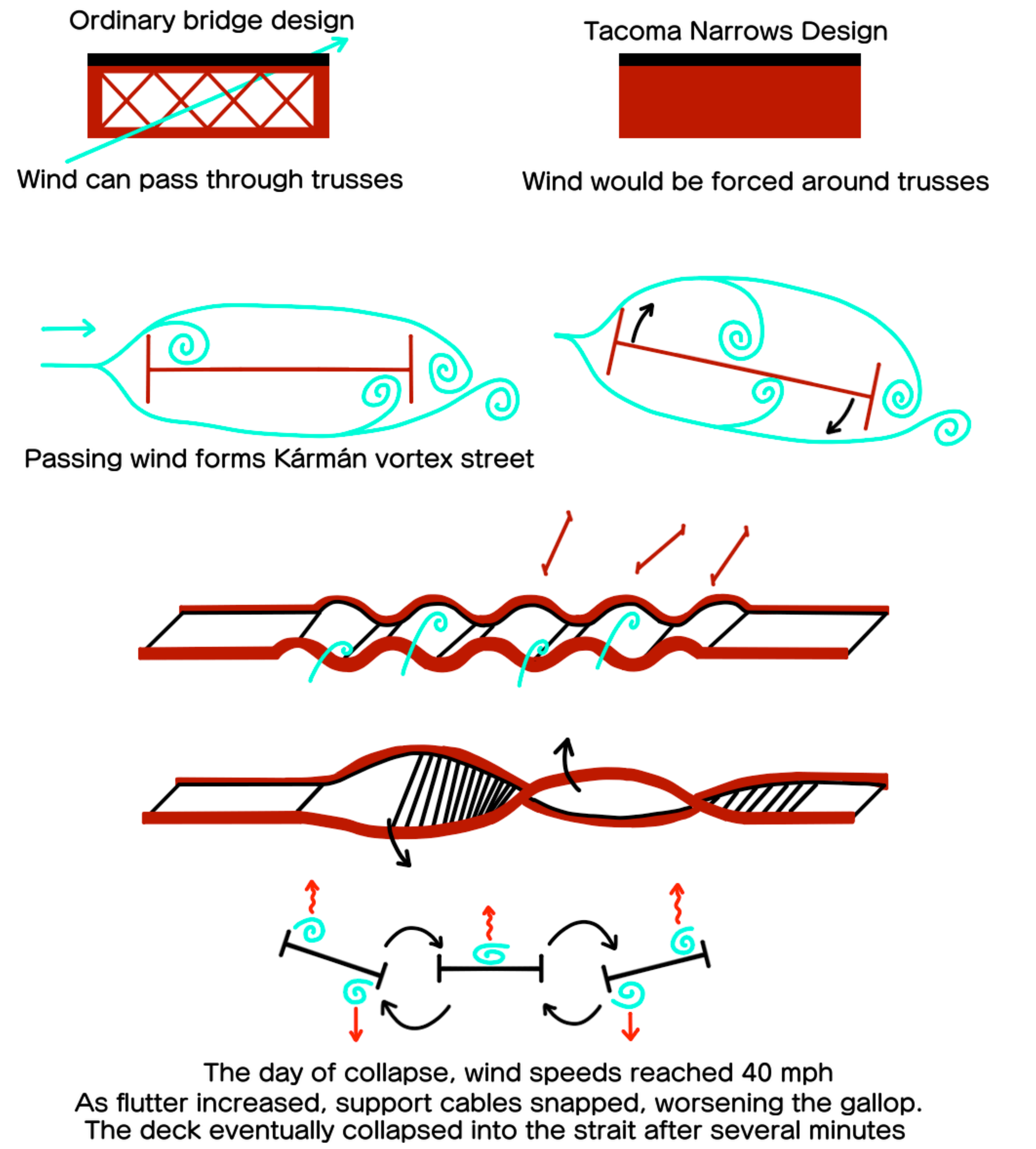
- Designed to withstand first order forces but vibration was not considered

- Resonance

- Vortex shedding?

    - Forced oscillation close to resonance fre





https://en.wikipedia.org/wiki/Tacoma_Narrows_Bridge_(1940)

How do we then learn about unknown dynamics?

- Learn by **experience**, **experimentation**, **hypothesis generation**, **data science!**
    - “The Tacoma Narrows bridge failure has given us invaluable information ... It has shown [that] every new structure [that] projects into new fields of magnitude involves new problems for the solution of which neither theory nor practical experience furnish an adequate guide. It is then that we must rely largely on judgment and if, as a result, errors, or failures occur, we must accept them as a price for human progress.” [Othmar Ammann]
    - Following the incident, engineers took extra caution to incorporate aerodynamics into their designs, and wind tunnel testing of designs was eventually made mandatory.

Motivation and warnings for the use of neural data science to answer big questions

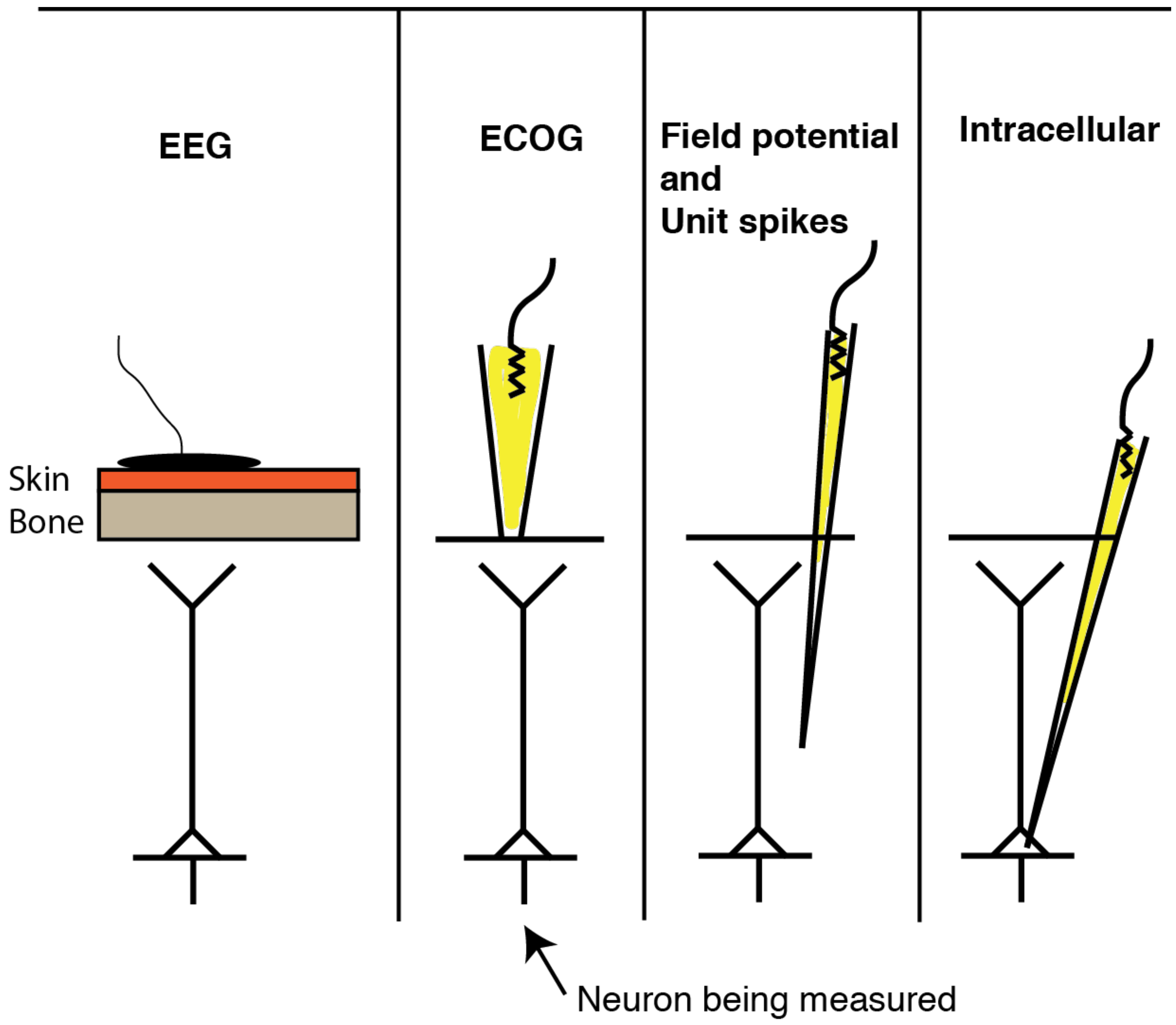
- **Power**- It’s powerful
- **Scope is bigger** - We are making models and theories that are useful, of increasing complexity and scope - drawing connections from broad sources
- **Models are finite, world is infinite** - The models are only models, and thus finite, and do not capture the infinite dynamics of the system
- **All models have assumptions** - All models and fields are based upon assumptions
  - Recognize that and seek to make more and more useful studies, models, data science tools for neuroscience and beyond
  - Treat all who say they are displaying the underlying mechanisms with skepticism (take the best of their theory and apply it but recognize for what it is)

From statistics to recordings…

- We have discussed parametric and nonparametric models
- NWB, DANDI, BIDS
- EEG/MEG, MOCAP, LISC
- Data science and neuroscience perspective on it
- Modeling concepts
- Now let's consider maps between neurons and systems

Neuronal recording approaches

- Electroencephalography (EEG)
- Electrocorticogram (ECOG)
- Local Field Potential (LFP)
- Unit Spikes (US)
- Intracellular (IC)
- Number of units recorded by size:
  - EEG>ECOG>FP>>US>IC

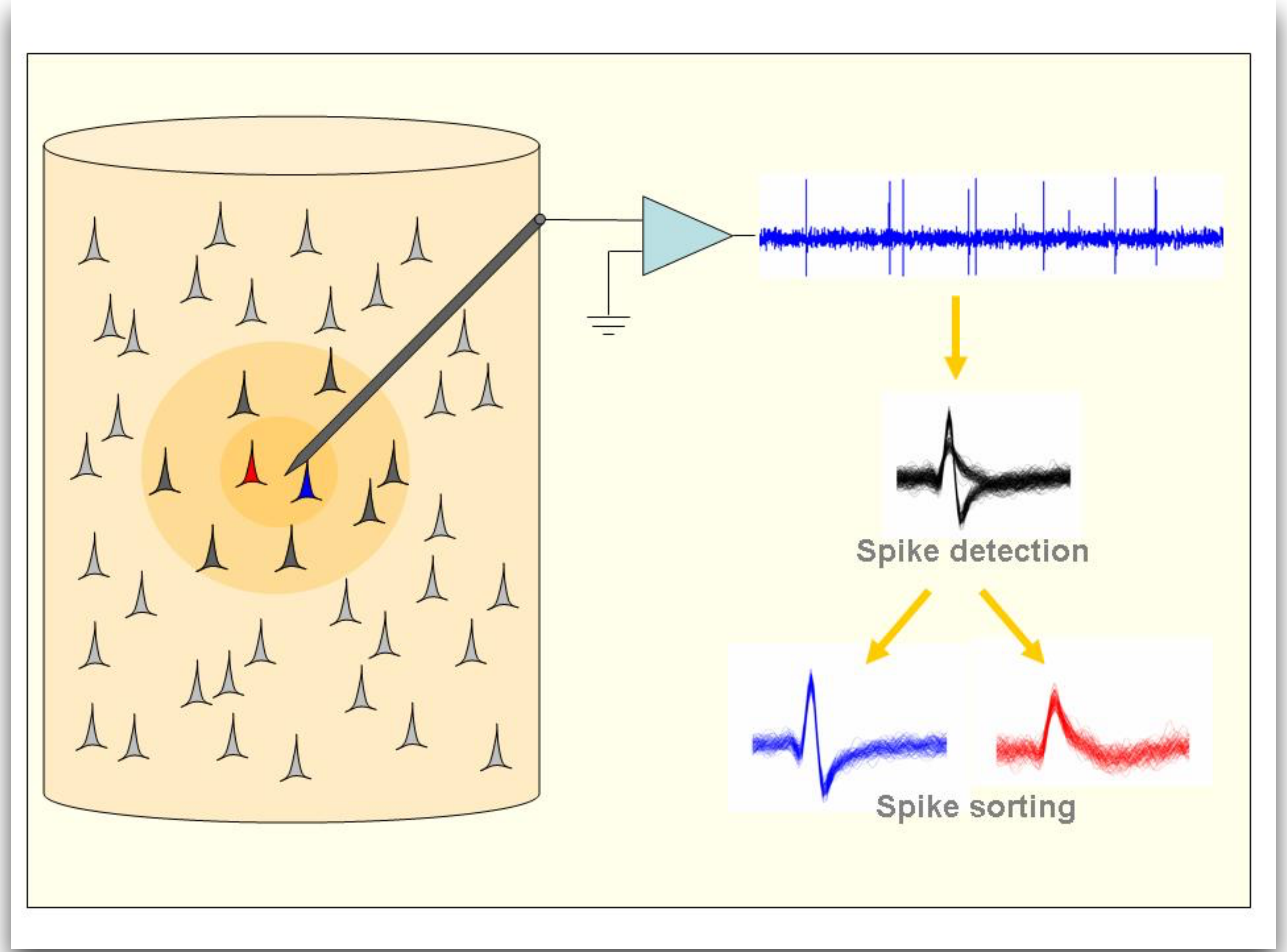


Problems with field level recording

- Like recording the crowd in a statium's voices as opposed to a single voice in the crowd (single neuron)
- Cannot differentiate sub-regions of the area measured, only aggregate
- Some neuronal activity is highly spatially independent
- Often we are interested in understanding the specific detailed structure of the connections, cannot determine here
- Localization not unique
- Sparse neuron patterns are hidden completely

What is “Single unit recording?”

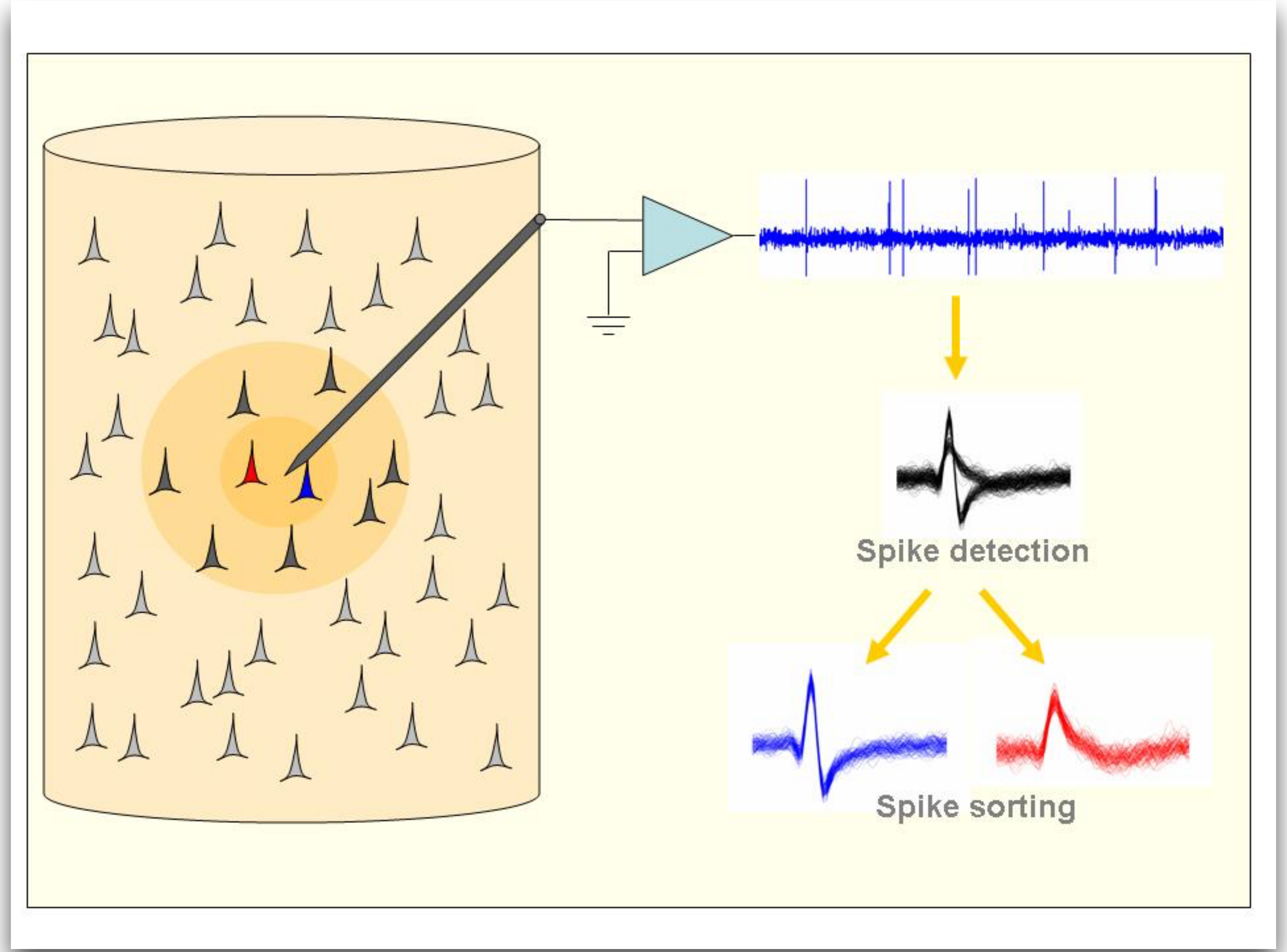
- Methods of measuring the electro-physiological response from single neurons with a micro-electrode
- An action potential generated by a neuron firing travels as a current down the excitable membrane regions through the soma and axon
  - The rate of change in voltage w.r.t. time is recorded
- Recorded extracellularly (several possible approaches)
- Micro-electrodes - High impedance, fine-tipped and conductive



http://www.scholarpedia.org/article/Spike_sorting#:~:text=Spike%20sorting%20is%20the%20grouping,activity%20of%20different%20putative%20neurons.

What is "Single unit recording?"

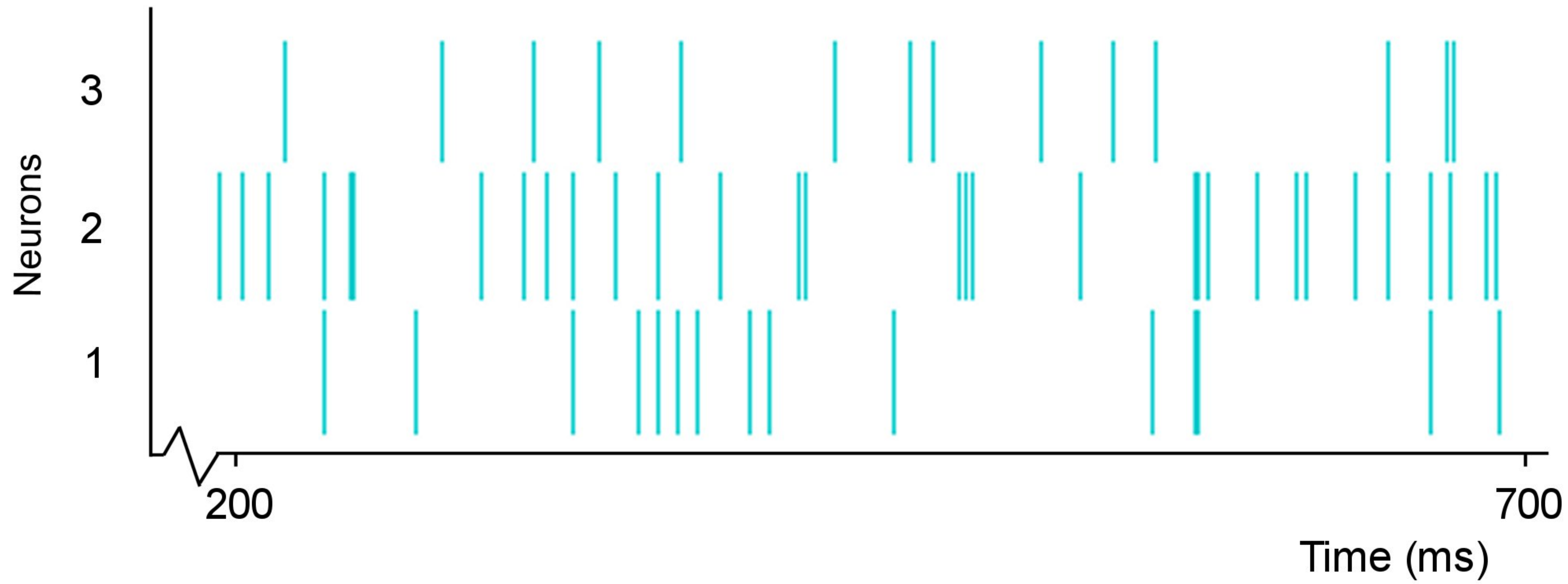
- Each neuronal response will have a unique characteristic shape
- Determined by
    - Morphology of the neuron (esp. dendrites)
    - Relative orientation/distance of measuring electrode

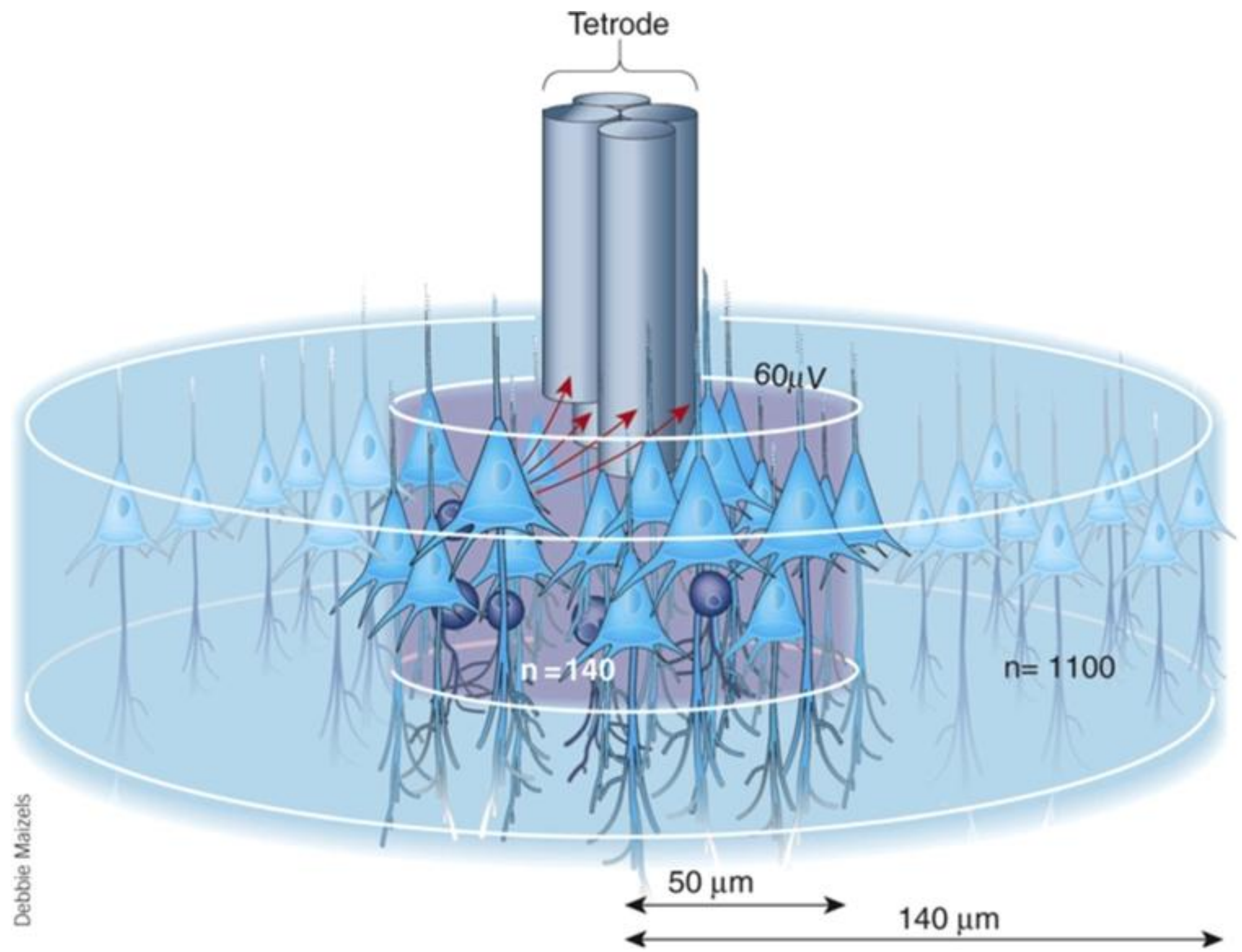


http://www.scholarpedia.org/article/Spike_sorting#:~:text=Spike%20sorting%20is%20the%20grouping,activity%20of%20different%20putative%20neurons.

Recording Action Potentials

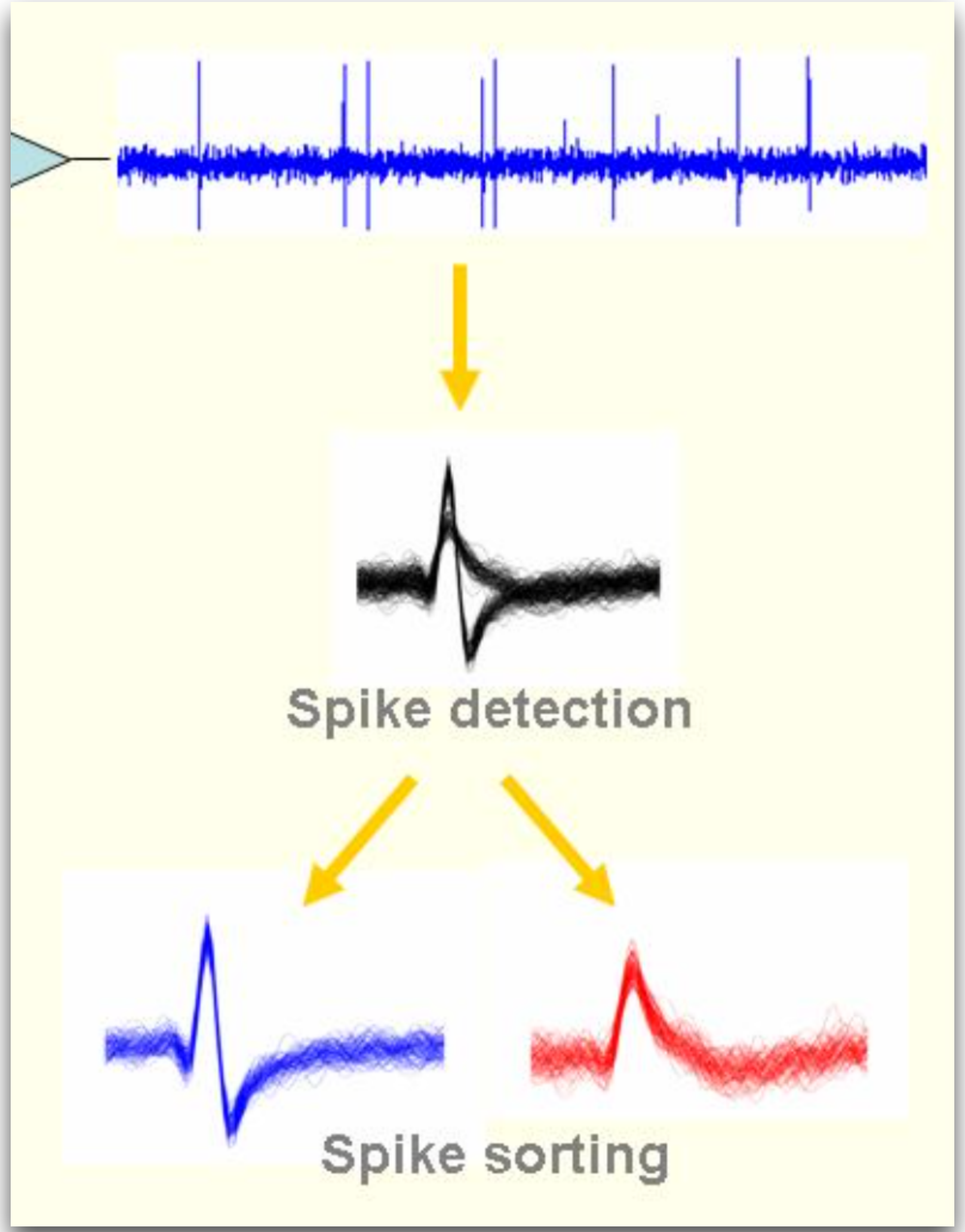
- Use tetrodes - 4 tiny pieces of tungsten or copper, twisted together, recording 4 at a time
- Implanted in brain using stereotactic device
- Must know specific coord. to record from
- When neurons fire, can use all 4 electrodes to localize which cell is firing
    - Sees differences between all 4 (ie bigger on left electrode than right, that cell is firing)





Spike sorting

- Spike sorting is a process of processing neuronal data and grouping neuronal spikes into clusters based on their firing characteristics/shape
- 'Which spike corresponds to which neuron' from a cluster that is recorded



Spike sorting motivation

- Progress is reliant on

1. Simultaneous recording from a large sample of cells
2. Sorting out which cells are coordinated and how

Spike sorting motivation

- Good spike sorting algorithms can extract information from a few local neurons to near electrode
- Examples of uses:
  - Close-by neurons
  - Topographical organization
  - Sparsely firing neurons

Spike sorting methods

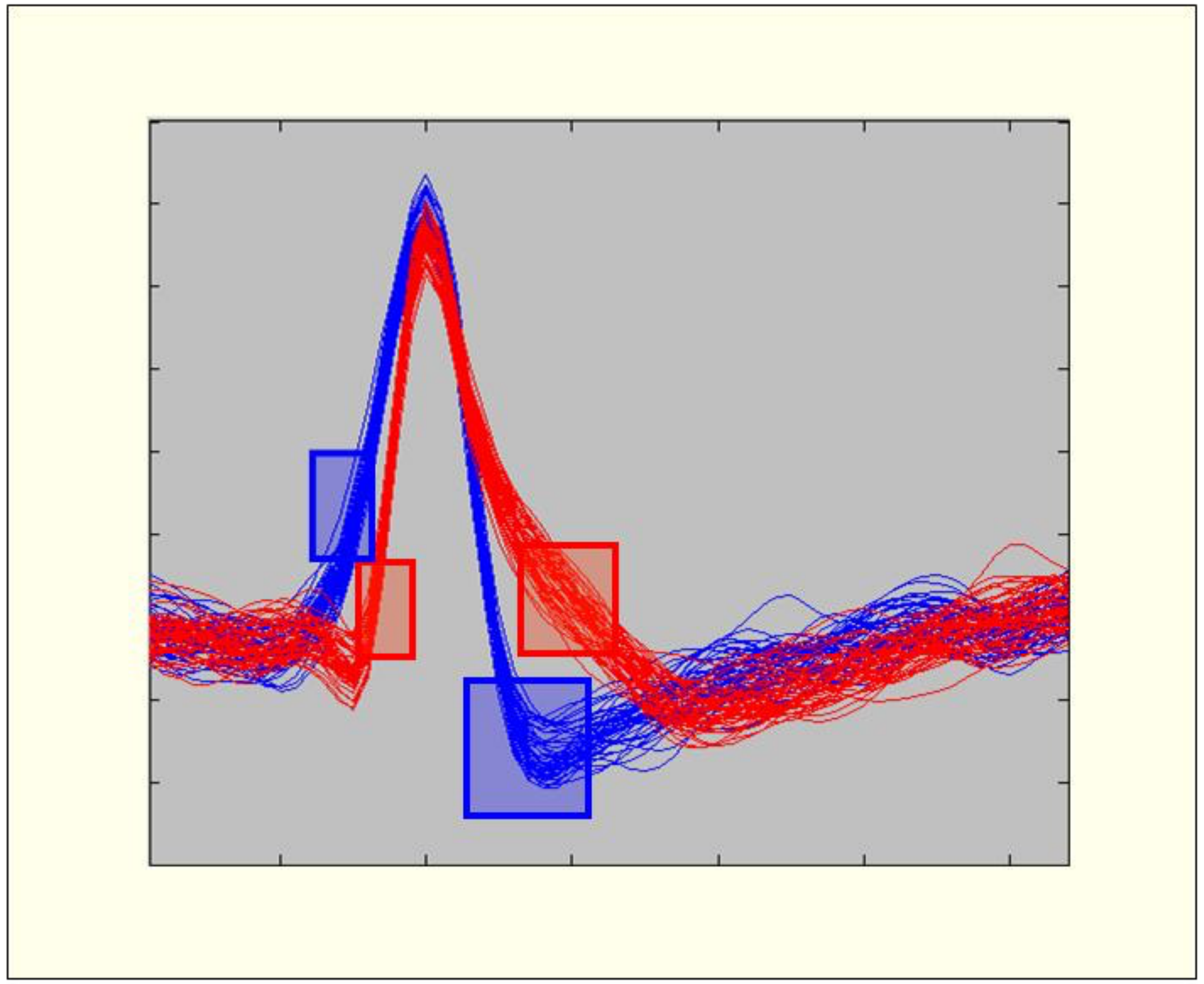
- Many exist, from simple to sophisticated (a sample here)
  - **Amplitude discriminator**
  - **Window discriminator**
  - **Characteristic shape/template matching**
  - **Supervised learning**

Spike sorting methods

- **Amplitude discriminator**
  - Advantages
    - Simplest
    - Implementable online
  - Disadvantages
    - Similar neurons can have different shapes but same amplitude

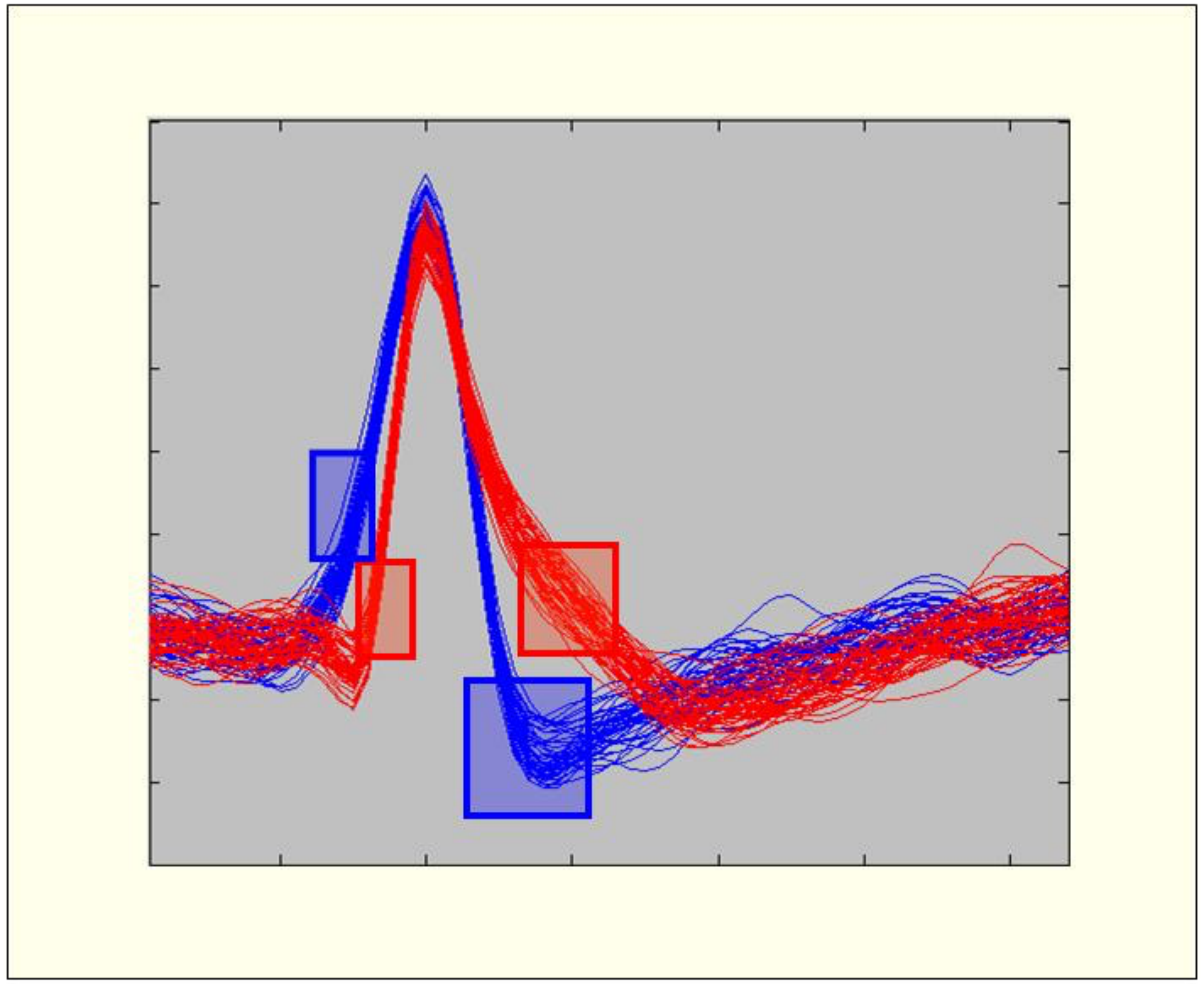
Spike sorting methods

- **Window discriminator**
  - Advantages
    - Spikes crossing multiple windows to a neuron
    - One of most popular, often in commercial systems
    - Online implementation
  - Disadvantages
    - Cannot sort more than a few channels simultaneously because needs manual tuning by user, possibly readjustment, “black art”
    - Overlapping shapes makes it hard to set up windows to separate (subjectivity)
    - Sparsely firing neurons can be missed



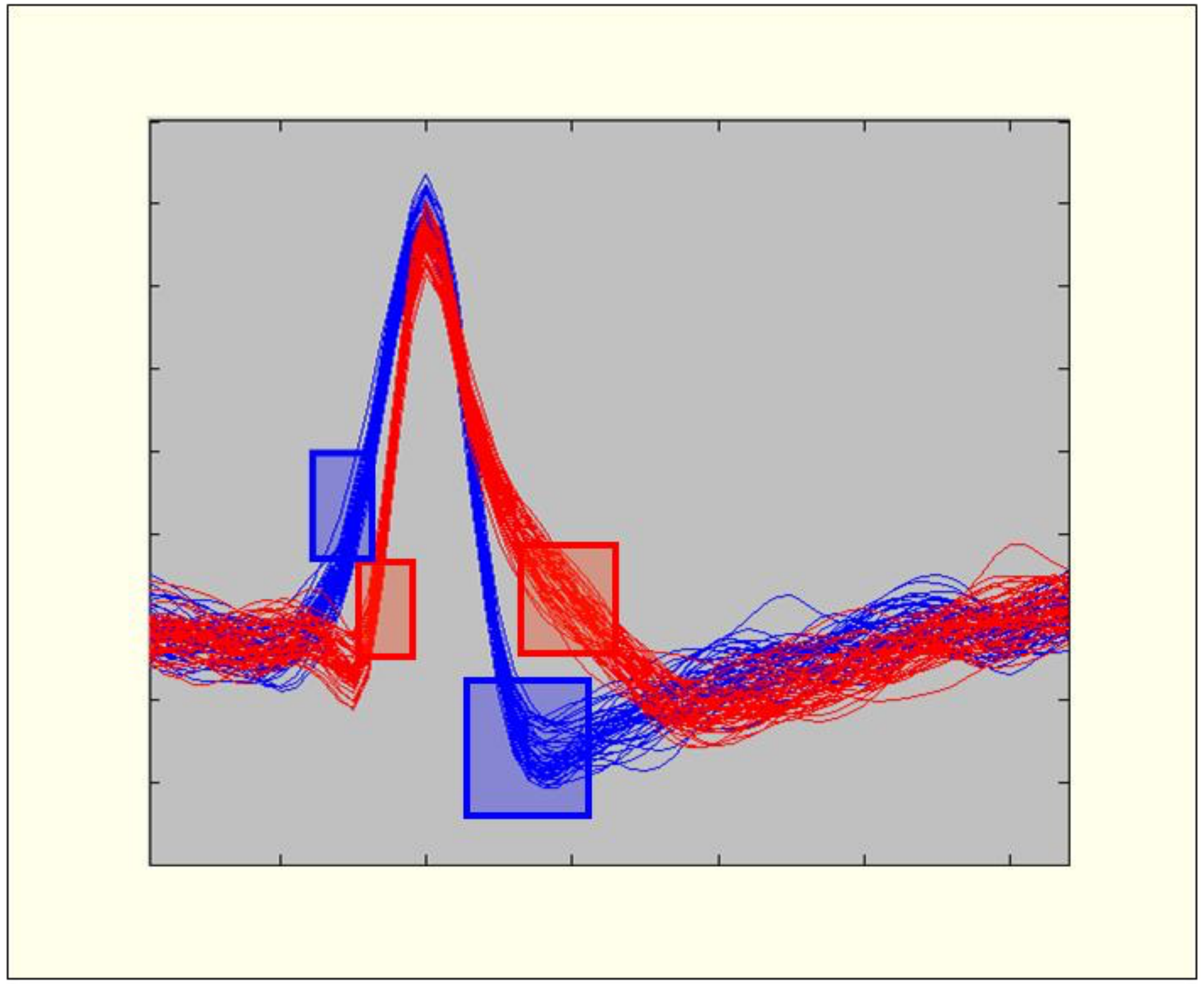
Spike sorting methods

- **Characteristic shape/template matching**
  - assign shape e.g. red/blue
  - group by shapes (often mean square distance metric)
  - Advantages
    - Online implementation possible again
  - Disadvantages
    - user intervention, not good for large n channels
    - adjustment of templates during experiment
    - may not be clear what templates or how many necessary
    - sparse neuron firing missed



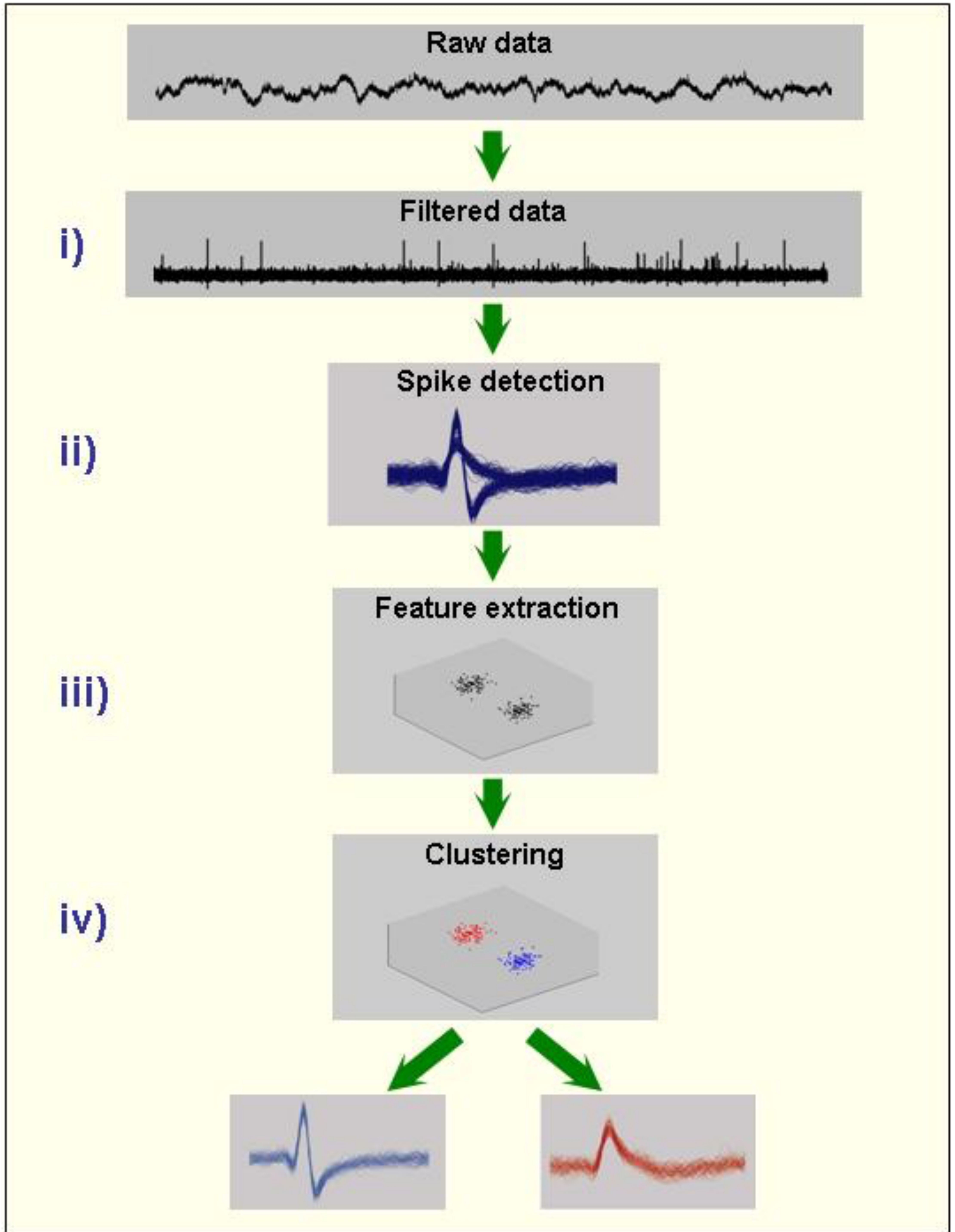
Spike sorting methods

- **Supervised learning**
  - Deals with large number of channels
  - Advantages
    - Deals with large number of channels
  - Disadvantages
    - Time consuming
    - Subjective
    - Can’t be used during experiment (with caveats)



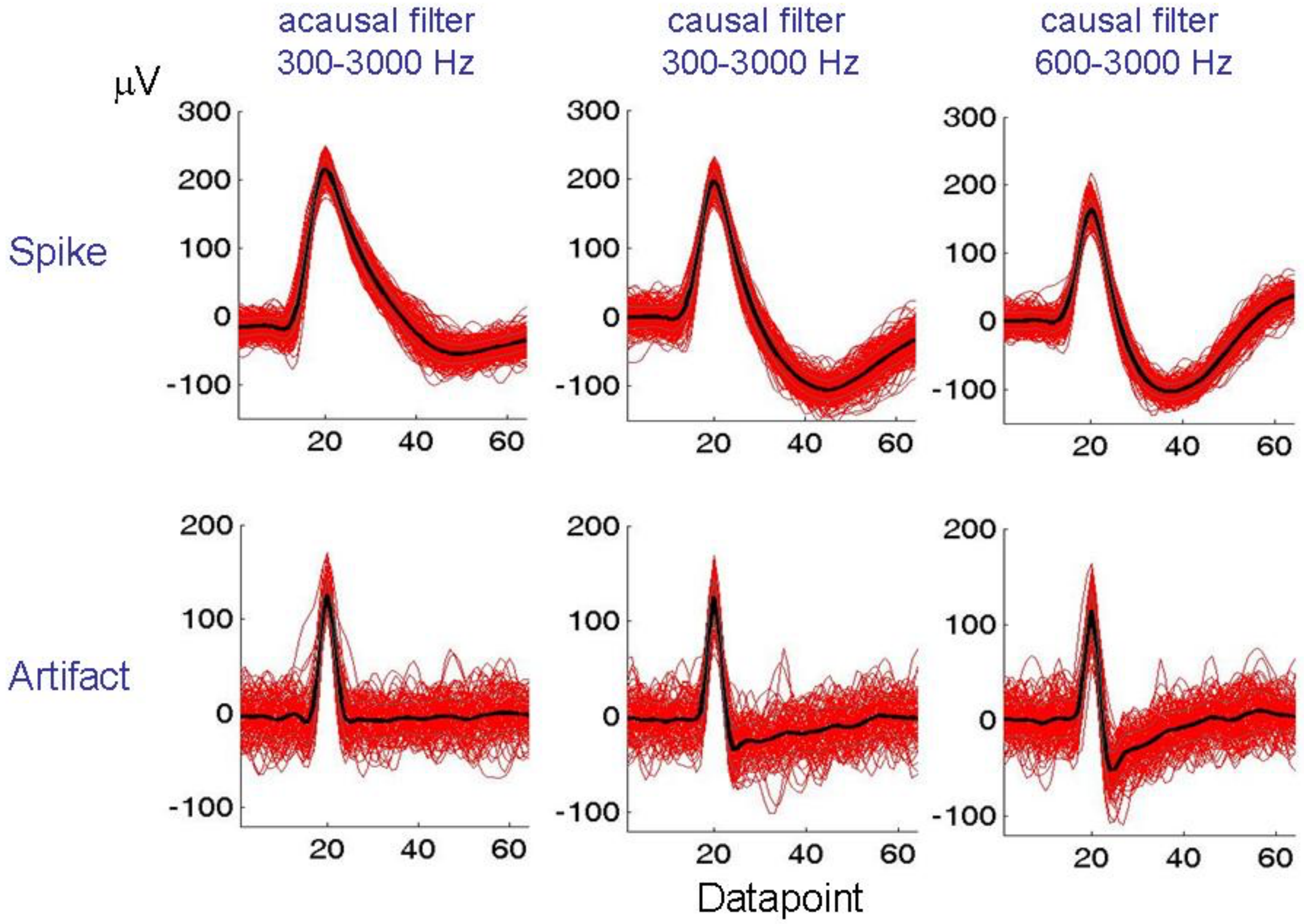
Spike sorting steps

- Filtering
- Spike detection
- Feature extraction
- Clustering



Filtering

- Bandpass (300-3k)
- non causal
- always a compromise
- causal filters would introduce phase distortion
- often commercial systems implement hardware filtering such as butterworth - receive already distorted data



Issues and challenges

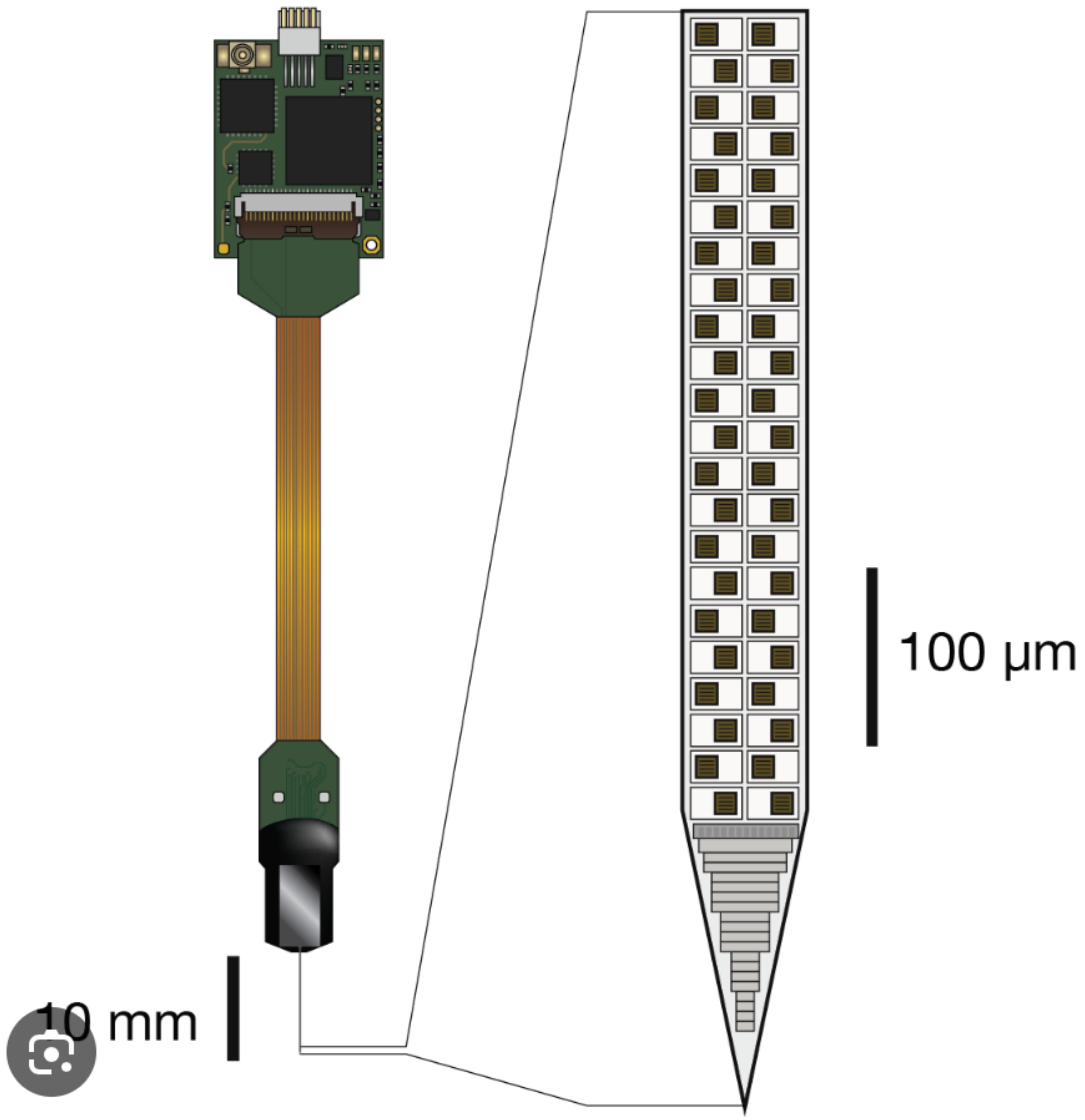
- Tetrodes
- Overlapping spikes
- Bursting cells
- Non-gaussian clusters

Introduction to python modules associated

- https://pypi.org/project/spikeinterface/
- https://elifesciences.org/articles/61834
- https://github.com/topics/spike-sorting
- https://core.ac.uk/download/pdf/52193212.pdf
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7704107/

High density single cell recording

- Multiple organizations are developing micro electrodes with multiple recording sites per electrode
- Record thousands of neurons simultaneously instead of only a few
- Better picture of entire brain areas in realtime
- Animal models and recently human



High density recording defined

- Chung, J. E., Sellers, K. K., Leonard, M. K., Gwilliams, L., Xu, D., Dougherty, M. E., Kharazia, V., Metzger, S. L., Welkenhuysen, M., Dutta, B., & Chang, E. F. (2022). High-density single-unit human cortical recordings using the Neuropixels probe. *Neuron*, *110*(15), 2409–2421.e3. https://doi.org/10.1016/j.neuron.2022.05.007

Intro to neuropixels

- https://www.neuropixels.org
- https://portal.brain-map.org/explore/circuits/visual-coding-neuropixels

References

- http://www.scholarpedia.org/article/Spike_sorting#:~:text=Spike%20sorting%20is%20the%20grouping,activity%20of%20different%20putative%20neurons.

- https://www.frontiersin.org/articles/10.3389/fninf.2022.851024/full

Additional notes

- There are several recording methods
- 2-photon
- genetic mouse models - phosphorescence

Field potential data

- Let’s do some loading processing and filtering of field potential data