COGS138: Neural Data Science

Lecture 14

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
- Field potential data
- Spike sorting
- Neuropixels datasets and similar
- PCA

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

Preparing for the grade check-in

- Have not posted the google form yet
- Check that you have:
  - Quiz 1, 2, 3
  - Assignments 0,1,2,3 (in process)
  - Participations Pre-course survey, Lecture questions, Lecture 11 pre-post surveys, group survey

Discussing the mid-quarter checkin

- First of all, we have a month to go, and a lot to cover. We are also going to be pulling it all together and you'll be applying the concepts covered, you haven't gotten into that but that does not mean you are done

- This is not 118E

- You should recognize where you stand on your knowledge - each course is a bare introduction to the topic - 118A/B for example

- This has not just been a review of 118a/b and 108, there is a necessary overlap and integration of topics, but we have presented a lot that is not covered in the prior courses

- The concept of this course includes an overview but is more about how to incorporate the concepts you learned into data science for neuroscientific advancement and application

- The curriculum that is approved by the university is so broad you could never go through a 'rigorous' build up of topics. It was not advertised as a 'rigorous' mathematical course, it's a data science course

Putting the time in

- You should think critically if all the topics we covered were covered in 118 and 108.
- If you think so, you need to pay better attention and do the reading I asked, and as I said, use the lectures as a beginning and not an end.
- If you don't feel like you are getting much out of the course, you need to work harder.
- You need to ask questions, of those who responded ~20% have not
- You need to read broadly on the topics,
- Think about the insights I am giving you,
- Practice with the tools I introduce.
- The final project will be the ultimate application of these topics, and your technical course foundation, along with the structure we have laid is a framework to build your project around

Course objectives

Learn how to:

- think from a “data first” perspective: what data *would* you need to answer your scientific
questions of interest?
- develop hypotheses specific to big data environments in neuroscience.
- work with many different neuroscience data types that might include data on behavior, brain structure and connectivity, single-unit spiking, field potential, gene expression, and even text-mining of the peer-reviewed neuroscientific literature.
- read and analyze data stored in standard formats (*e.g.*, Neurodata Without Borders and Brain Imaging Data Structure).
- integrate multiple heterogeneous datasets in scientifically meaningful ways.
- choose statistical model(s) informed by the underlying data.
- design a big data experiment and integrate data from multiple open data sources.
- consider alternative hypotheses and assess for spurious correlations and results.

What we told you to expect in this class

- What is my (our) role?
  - I am (we are) here to help you to learn and succeed, to open the door
  - NOT here to weed anybody out
  - NOT here to compete with you
  - Mutual respect

What we told you to expect in this class II

- What is your role?
  - Learn! Open your mind
  - Put in the effort - you must walk through the door
    - **Watch/attend lectures, do the readings, complete assignments and tests, and think about it all**
  - Treat each other well, help each other to succeed (but do your own work of course)

Course logistics we have been using

- Neural Data Science will be a lecture and project-based course.
- It will consist of three hours of lecture
  - one that is typically designated for lecture,
  - one typically designated for in-class project work and discussion.
- Homework will consist of problem sets designed to support students with their progression through the larger class projects.
- Students will be evaluated on their
  - class participation,
  - assignments,
  - readings, and
  - their large Final Project.

How to get the most out of the material

- You have taken courses on many of these topics, sometimes more than one - great!
- Each course is designed to paint a picture of a topic, but it’s like a very zoomed out low resolution photograph of a complex scene like a city - you get a sense of the colors and space at first, then choose sub-regions and zoom, revealing more pixels, and more and more
- Also like a fractal

Fractal complexity

- https://www.youtube.com/watch?v=fnuSrhGWqu4
- https://www.youtube.com/watch?v=LhOSM6uCWxk
- https://www.youtube.com/watch?v=rGwwydEWLiI
- https://www.youtube.com/watch?v=8YIZEp4lhRk

$$z_{n+1} = z_n^2 + c$$

[Fractal by Simpkins 2008]

How much time you should spend outside class

- Each UCSD class recommends students set aside **at minimum 4-6 hours** outside class for work beyond lectures
- We expect this, so **if you are doing less than this you are not putting enough time in**
- You should be doing this with all your classes, even if you could get A's doing less - if you can do the course material included faster, you should read more, study and practice more
- Unique time in life to do this, don't waste a second

What should you do if you completed the assigned reading, quiz, assignments?

- Read all references mentioned thoroughly, read further anything you don't understand within those references
- Download and spend time exploring each tool we discuss and present yourself.
  - Thoroughly go through every example,
  - Perform actions using those tools,
  - Get them set up in data hub and on your machine, you will use them during your project
- Spend time thinking about the topics, what are the implications of what we discuss? How does the background you have up to this point apply to these topics? Just because we don't test on

Where we are in the course plan…

- Covered
  - Intro to neural data science, questions, jupyter, python, pip
  - Data sources, issues, advantages, caveats and modalities - EEG/MEG, Eye tracking, Behavior, Motion Capture (Phasespace, VICON, IMU, kinematics/dynamics), Text/Speech, gene expression, sequencing, animal models, animal model limitations, alternatives
  - Tools for data science NLTK, MNE, Sentiment analysis, PyMo, mocaplib
  - FAIR data, NWB, BIDS, DANDI

Where we are in the course plan II…

- Covered continued…
  - Version control in data science and applications to neural data science, big data and reproducibility
  - Data structures, intuition and data cleaning tools/approaches
  - Scientific visualization, color theory, perceptually aware visualization, plots/ python review and choosing data representation/plots

Where we are in the course plan III…

- Covered continued…
  - Time series analysis, filtering, transformations, fourier analysis, continuous vs. discrete, analog vs. digital, frequency and other domains, stationarity, linearity vs. nonlinearity, implications, sampling, aliasing, filtering (low pass, high pass, band pass, band stop, notch, recursive, averaging, causal vs. non causal filters)
  - Statistical data analysis for neuroscience data, central tendency, distributions, independence, descriptive/inferential statistics, outliers, missing data, population and samples and implications for neural recordings, variability, mean, median, mode, variance, covariance, z scores, standard deviation, covariance, correlation in neural data science

Where we going in the course plan…

- Logistics:
    - 2-3 more assignments
    - 3-4 more lecture quizzes
    - 2 more readings and reading quizzes
    - more class participation
    - Paper - NEW
    - Project - proposal, cp1, cp2, final
- Topics
    - syllabus

New assignment

- 1) We are going over the mid-quarter checkin responses that have been received, we've only gotten about 1/4-1/3 so the sample is not fully representative
- 2) I'm assigning a new writing assignment. If there is any likelihood at all you used a chatbot or AI generated text you get a zero and a warning. Write it yourself or risk consequences, even if we suspect it and it's not certain, so you better not plagiarize. But moreover the point of this is for you to learn something, not use a tool to generate your thoughts for you. You, your thoughts and insights are not replaceable.

Discussion

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

Physicalist perspective

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

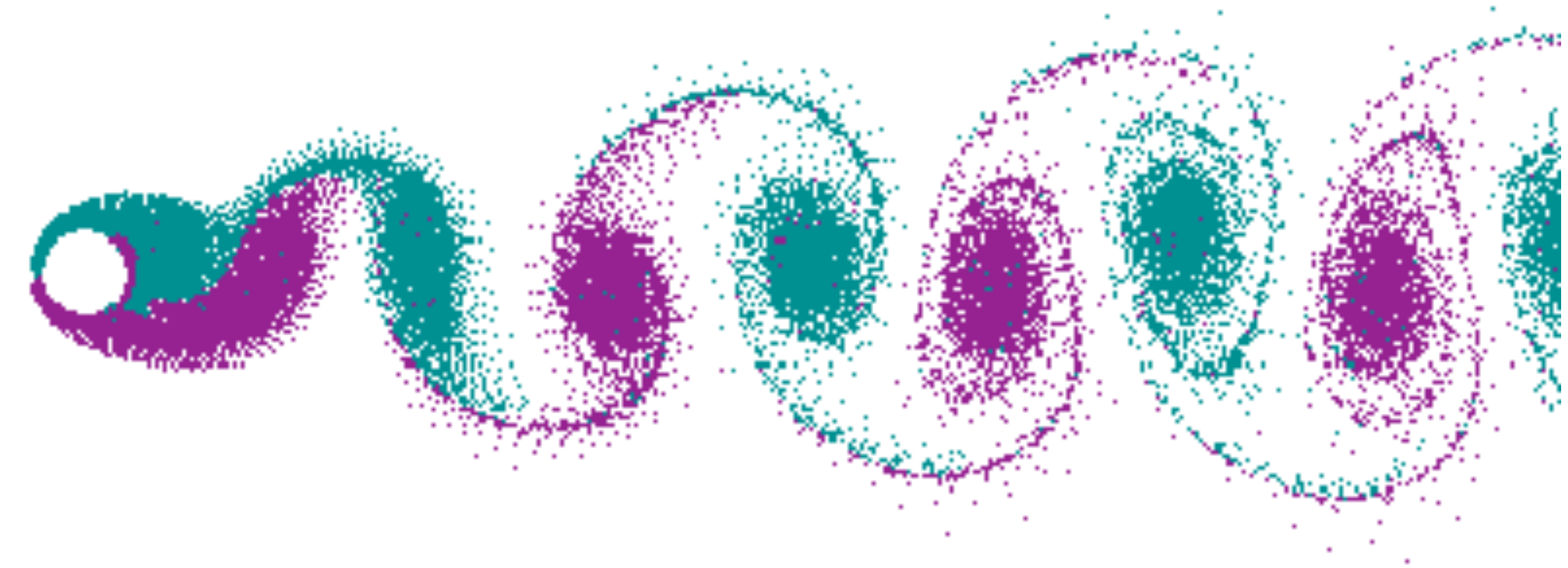
Underlying dynamics need to be exposed

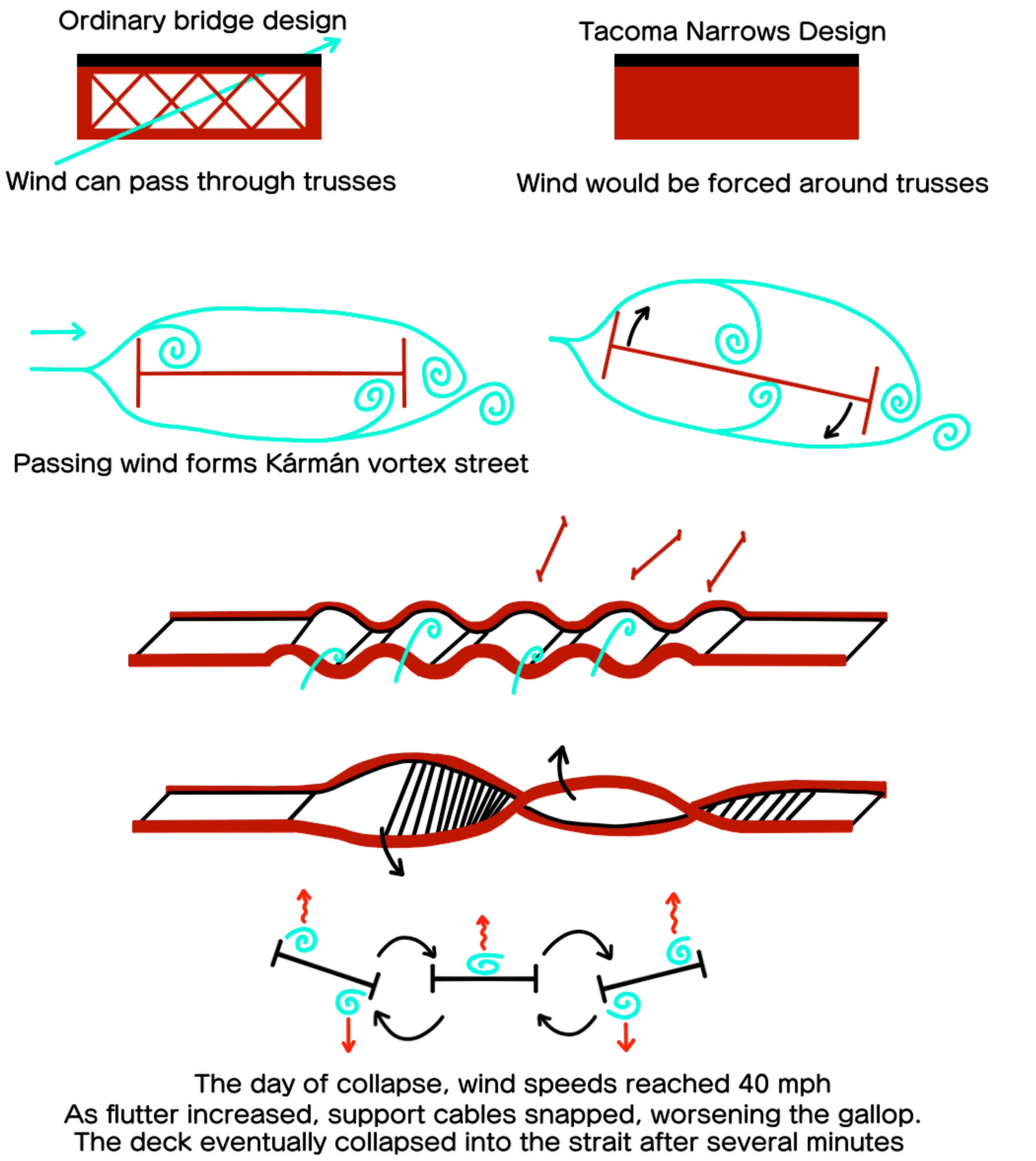
- Tacoma narrows bridge disaster
  - 1st order vs. higher order
  - https://en.wikipedia.org/wiki/Tacoma_Narrows_Bridge_(1940)

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

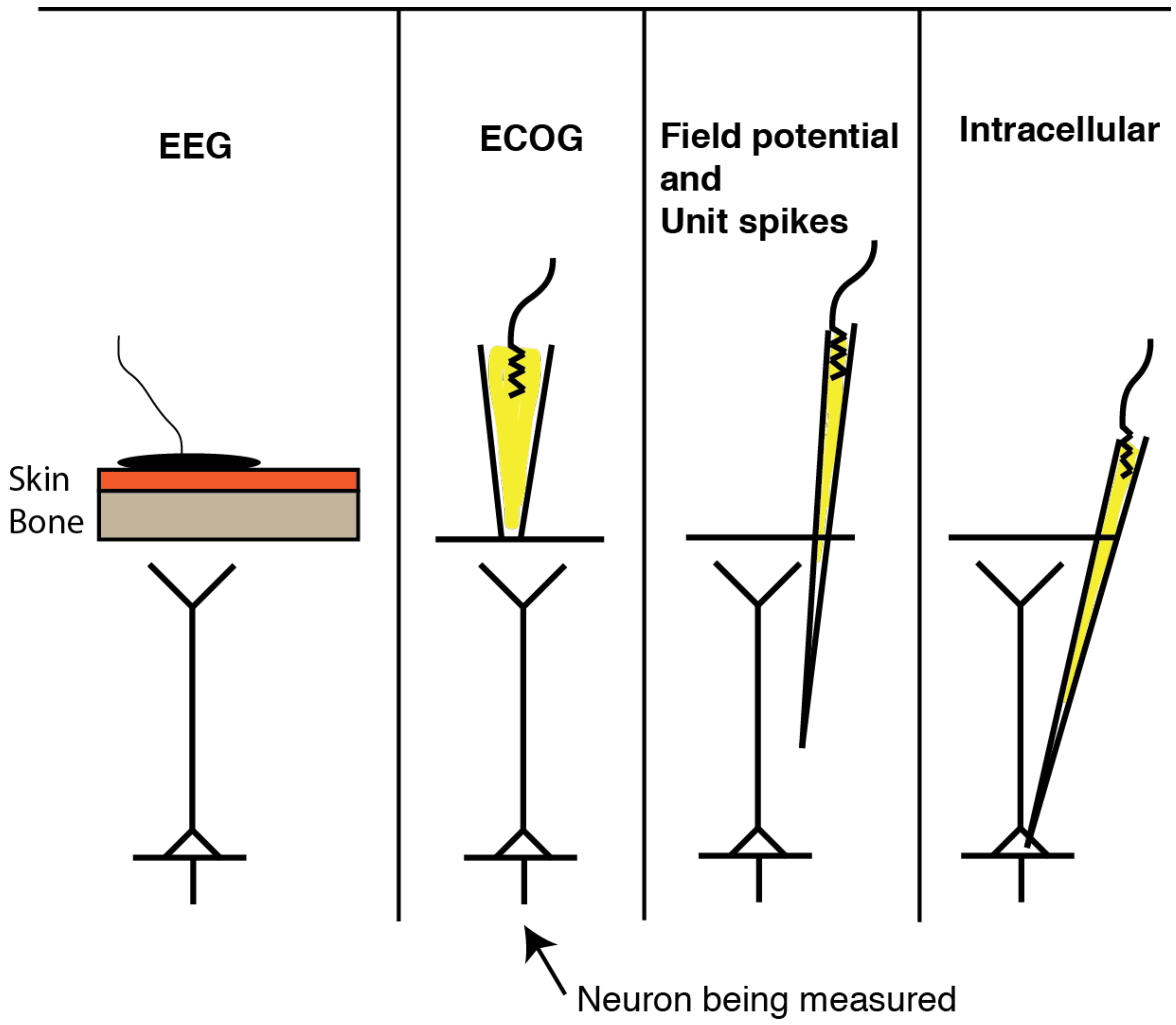
- **Power**- It's powerful
- **Scope is bigger** - We are making models and theories that are useful, of increasing complexity and scope - drawing connections from broad sources
- **Models are finite, world is infinite** - The models are only models, and thus finite, and do not capture the infinite dynamics 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

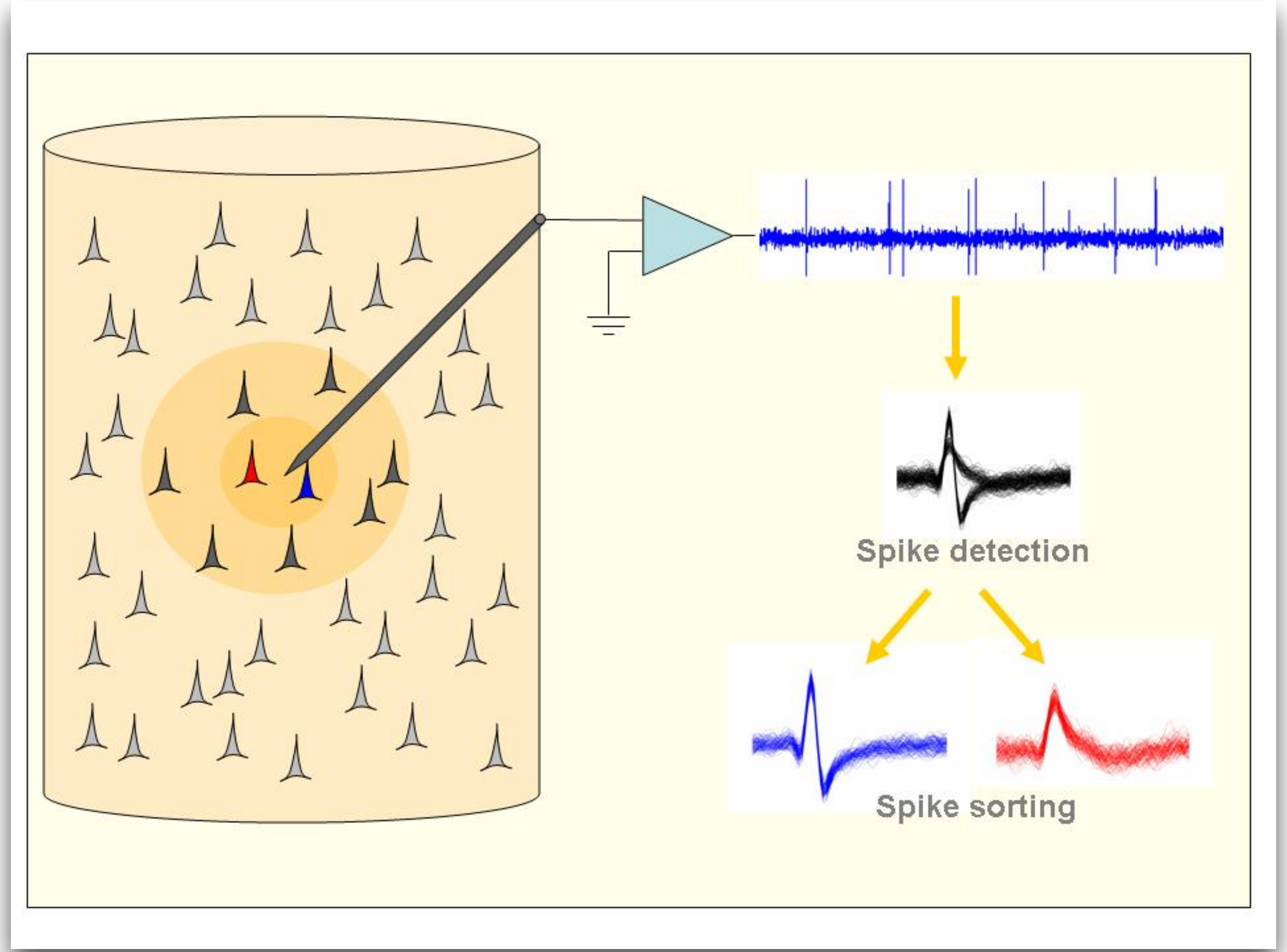


Problems with field level recording

- 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?”

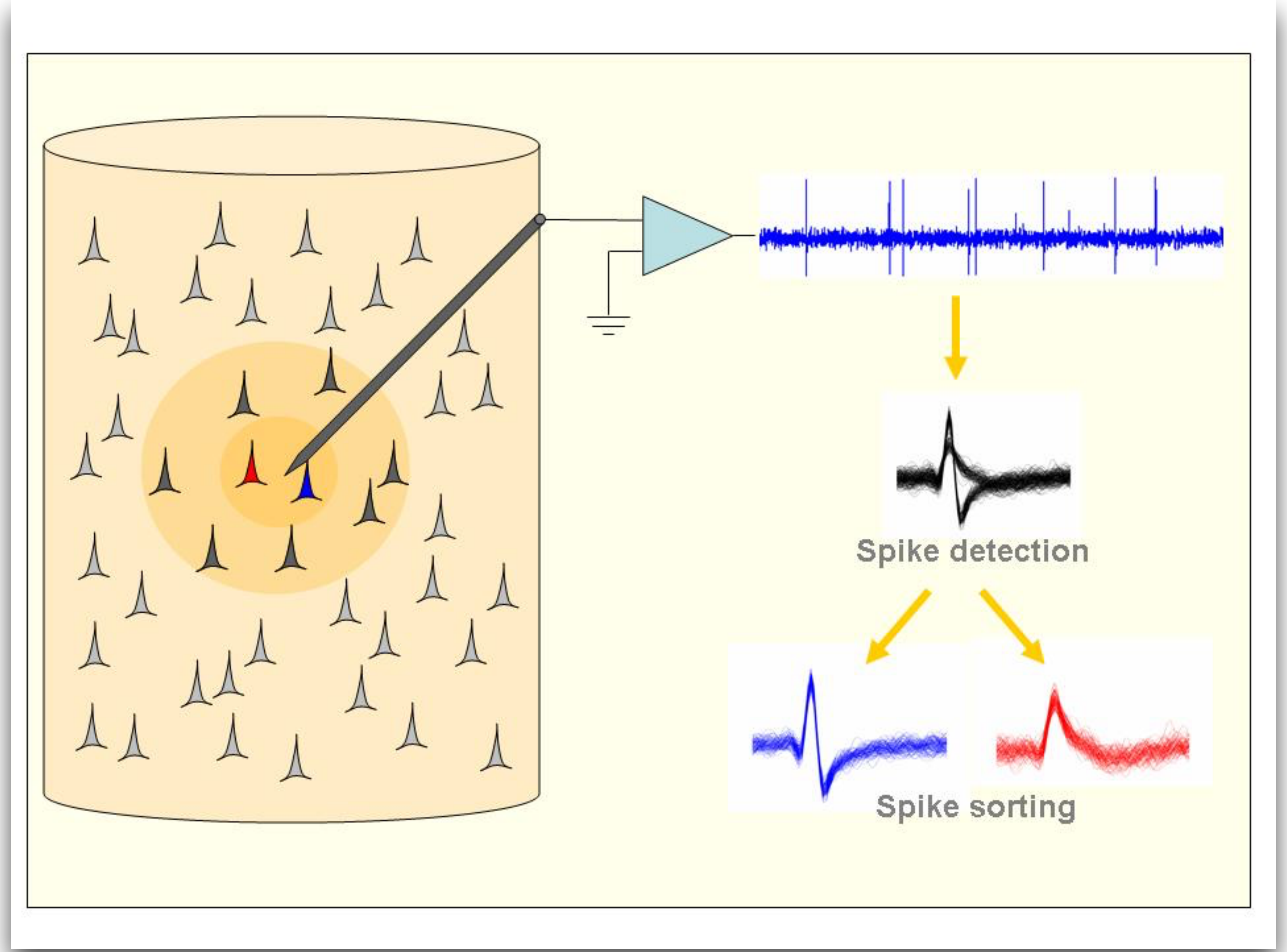
- Methods of measuring the electro-physiological response from single neurons with a micro-electrode
- An action potential generated by a neuron firing travels as a current down the excitable membrane regions through the soma and axon
    - The rate of change in voltage w.r.t. time is recorded
- Recorded 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?"

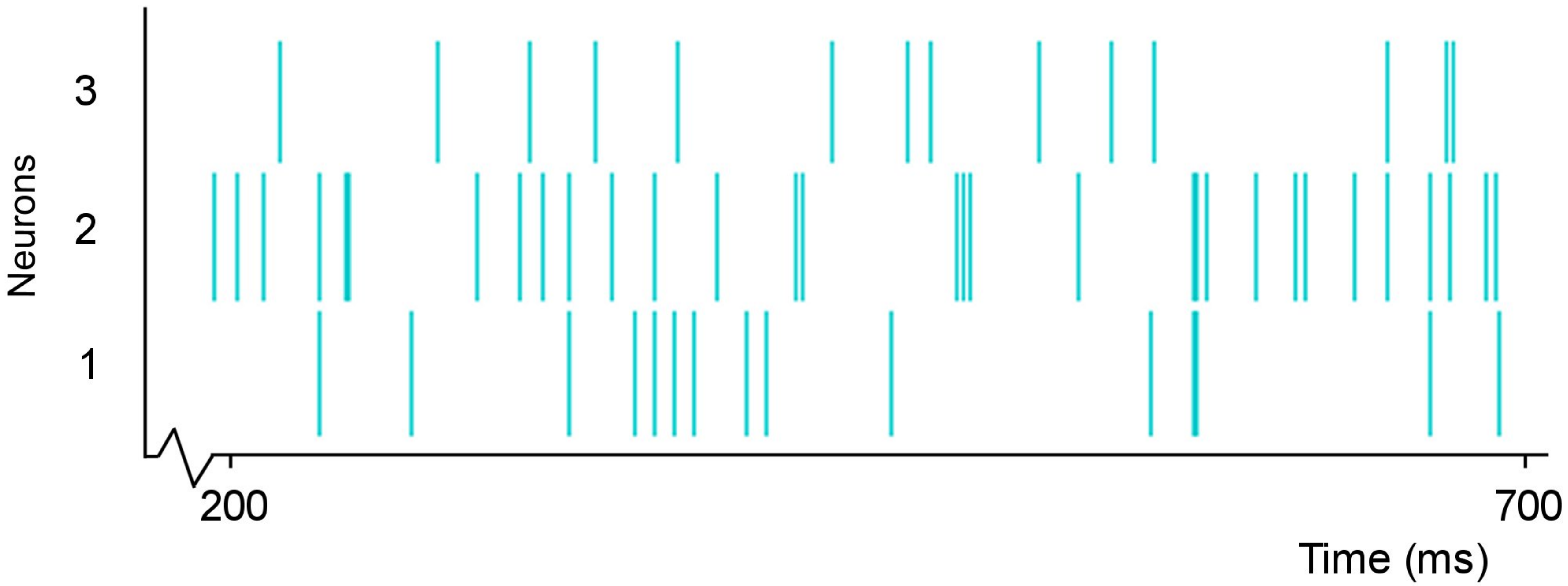
- Each neuronal response will have a unique characteristic shape
- Determined by
  - Morphology of the neuron (esp. dendrites)
  - Relative orientation/distance of measuring electrode

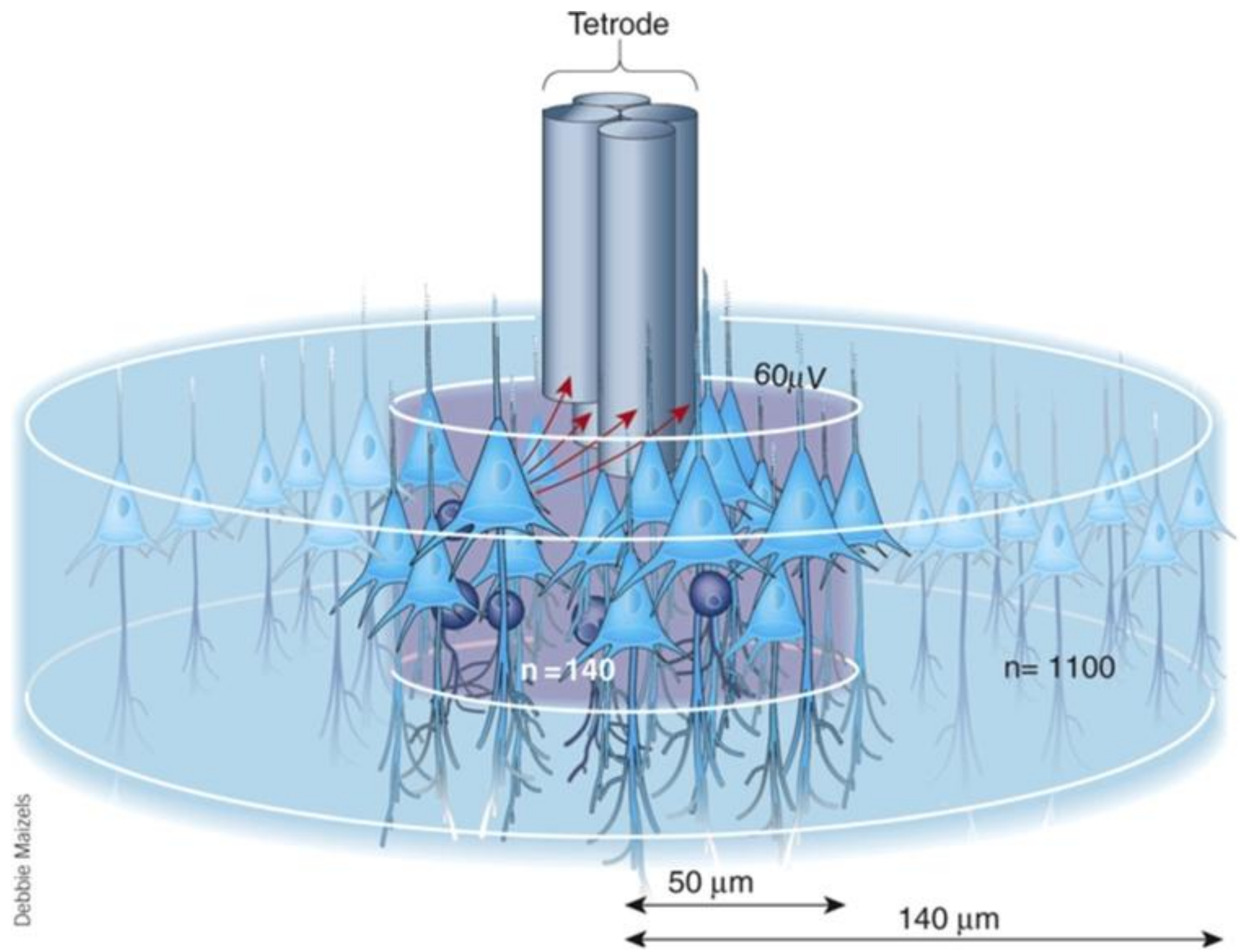


http://www.scholarpedia.org/article/Spike_sorting#:~:text=Spike%20sorting%20is%20the%20grouping,activity%20of%20different%20putative%20neurons.

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)

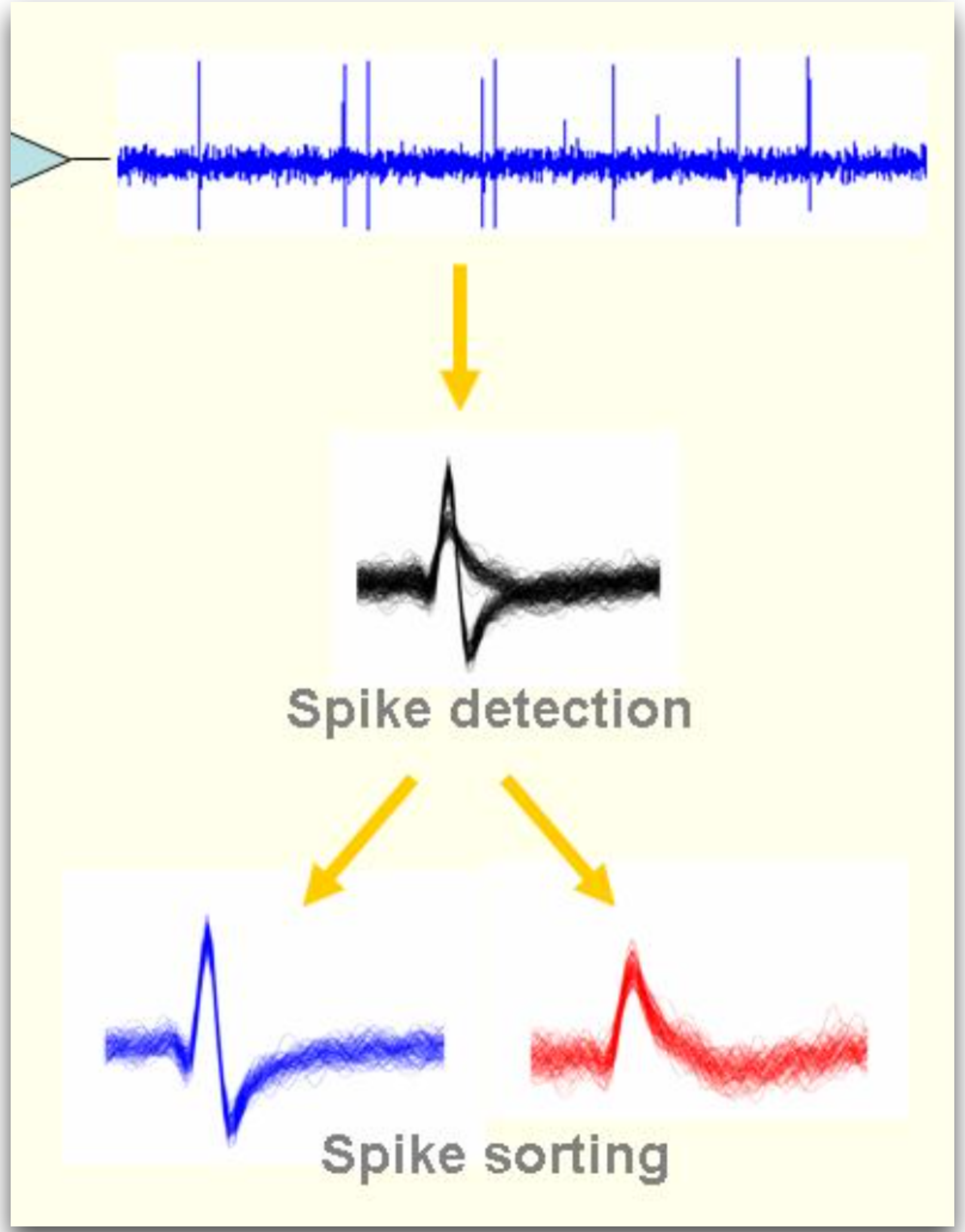




Spike sorting

- 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

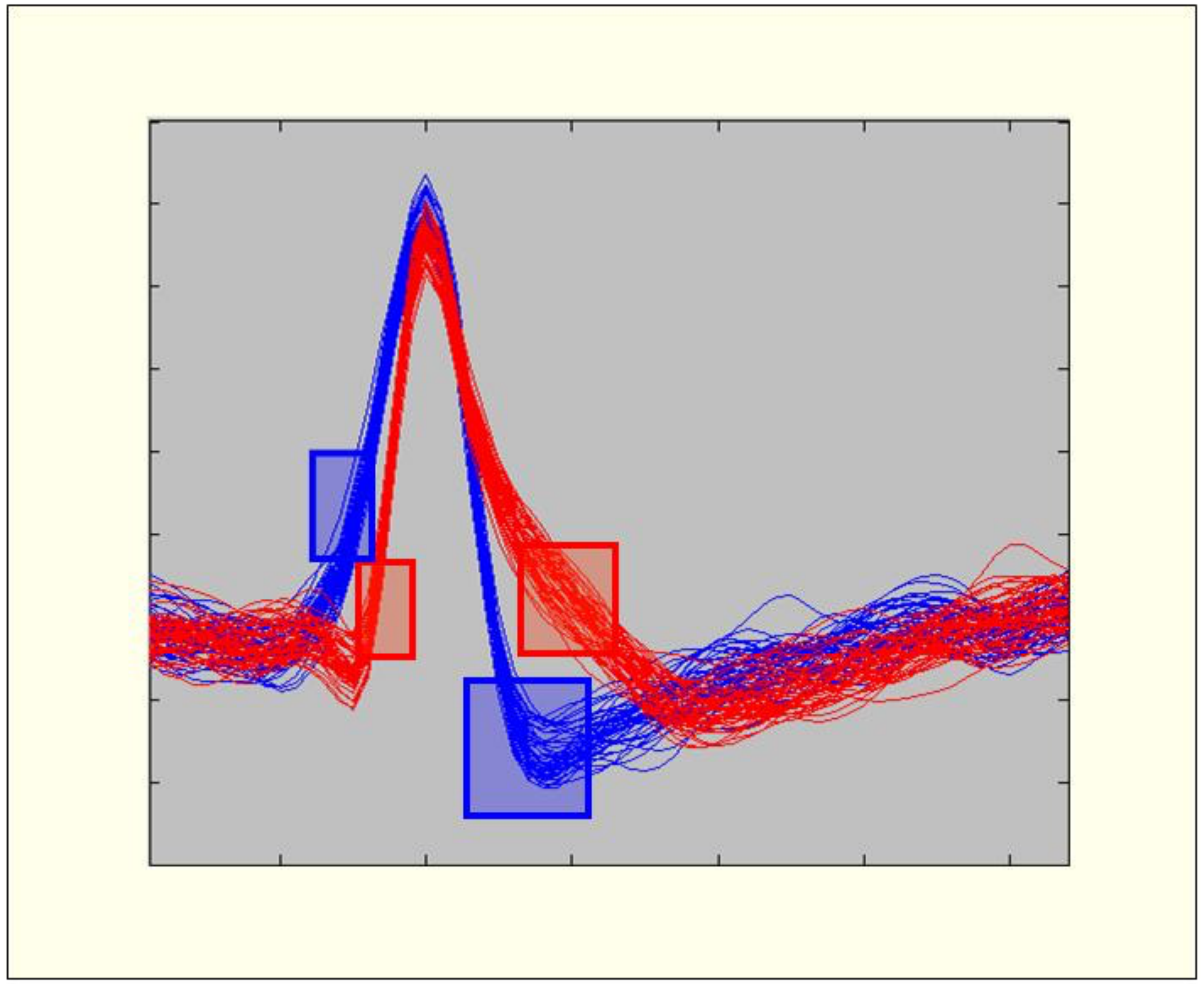
- Many exist, from simple to sophisticated (a sample here)
  - **Amplitude discriminator**
  - **Window discriminator**
  - **Characteristic shape/template matching**
  - **Supervised learning**

Spike sorting methods

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

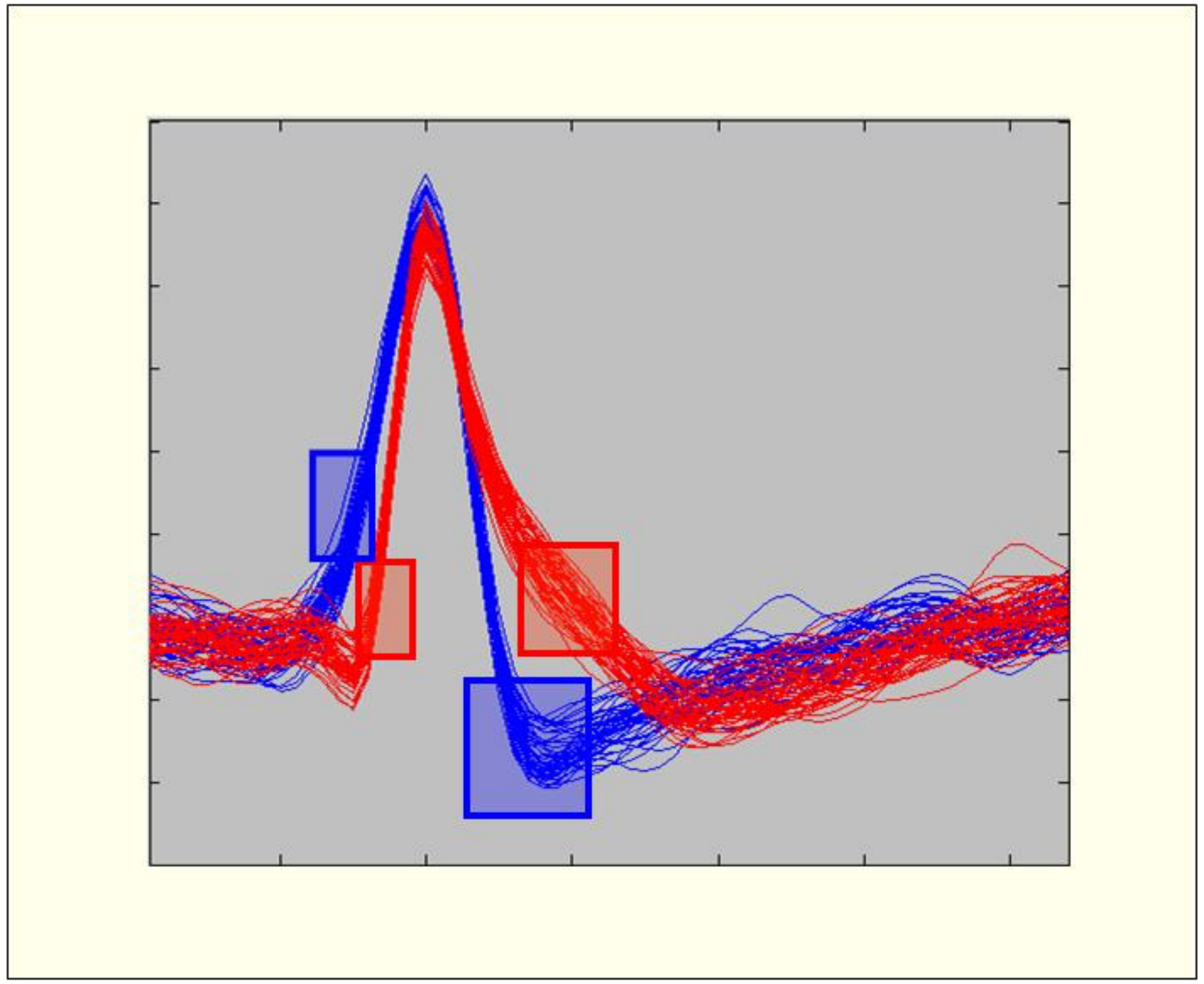
Spike sorting methods

- **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



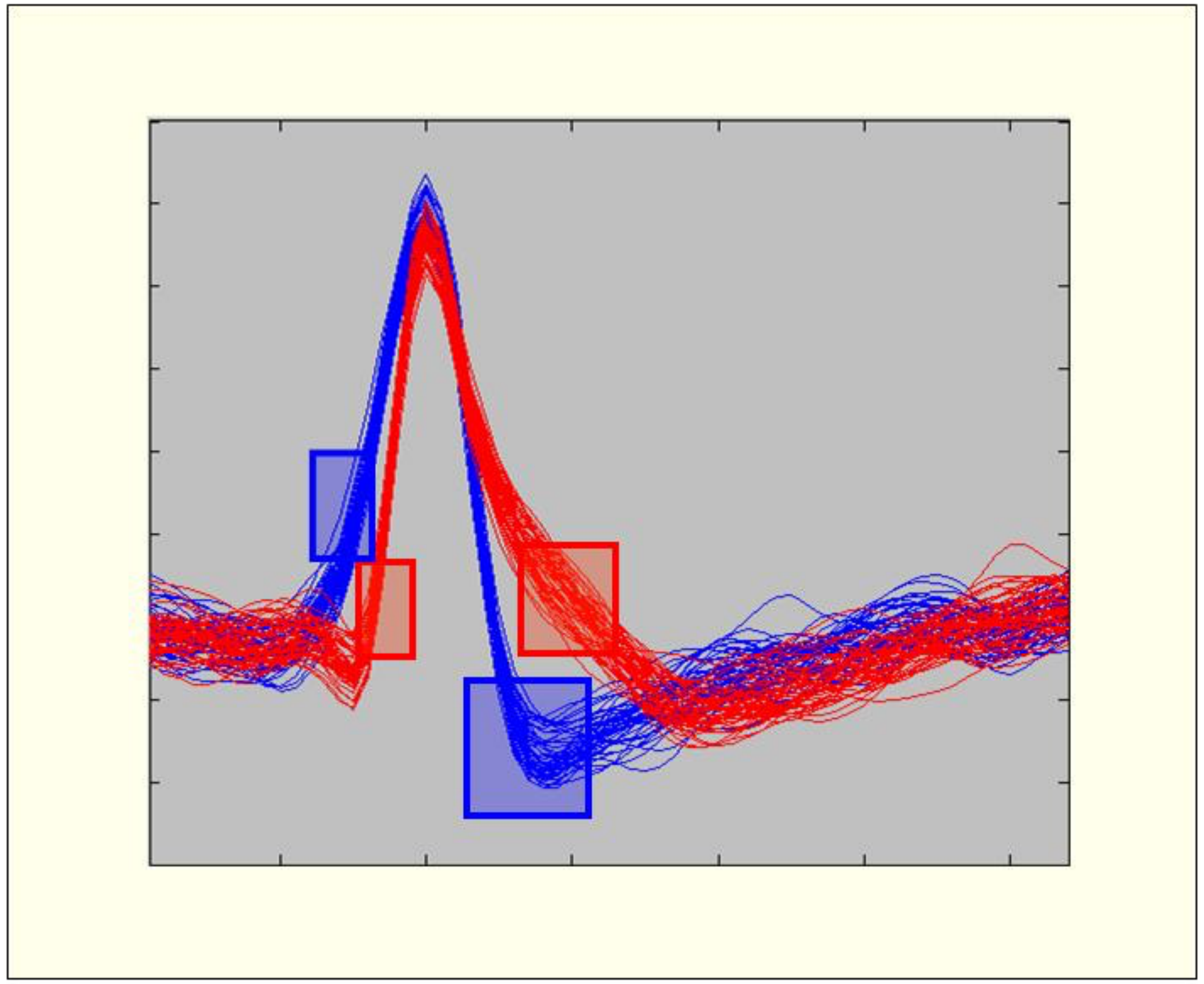
Spike sorting methods

- **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



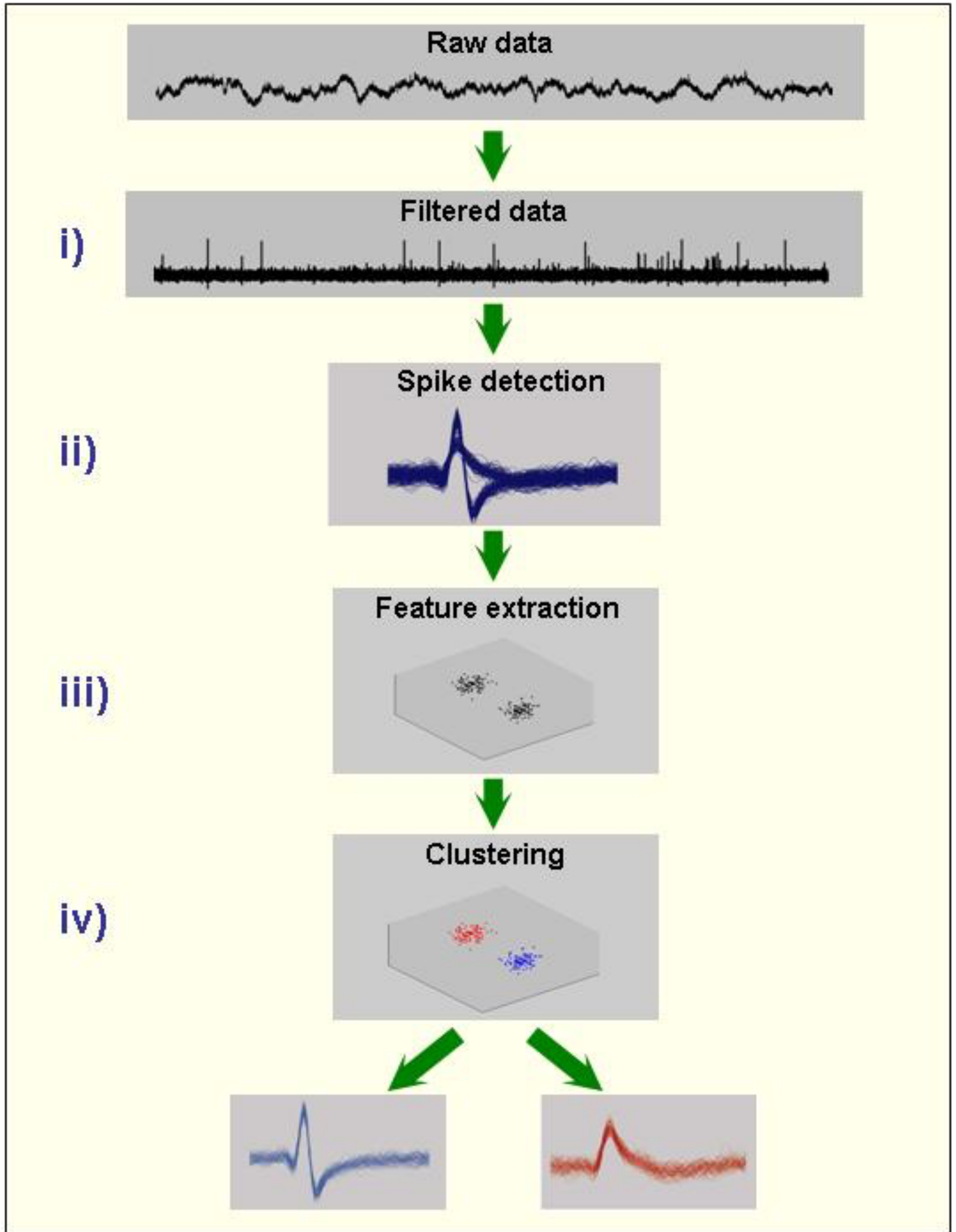
Spike sorting methods

- **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)



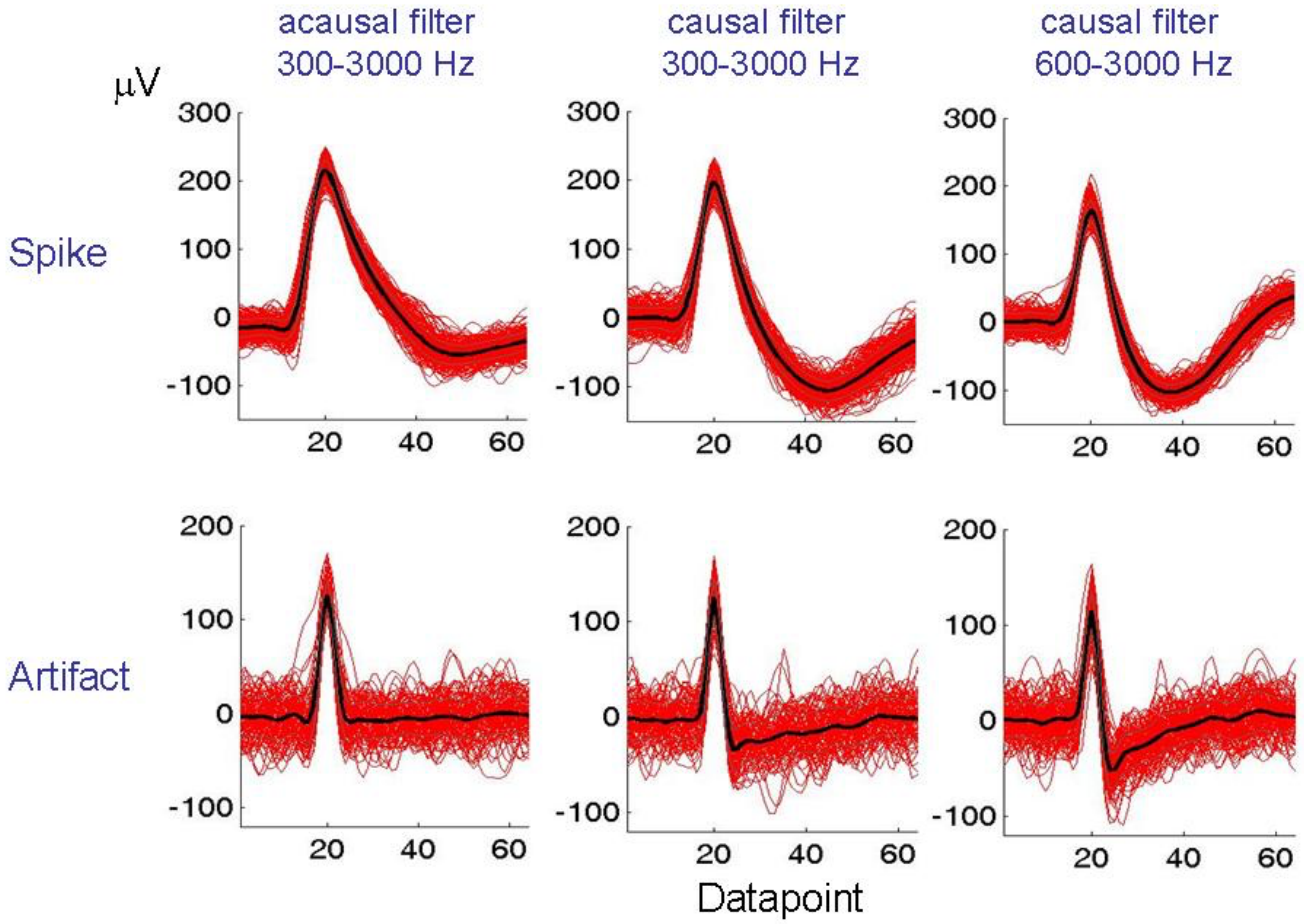
Spike sorting steps

- 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

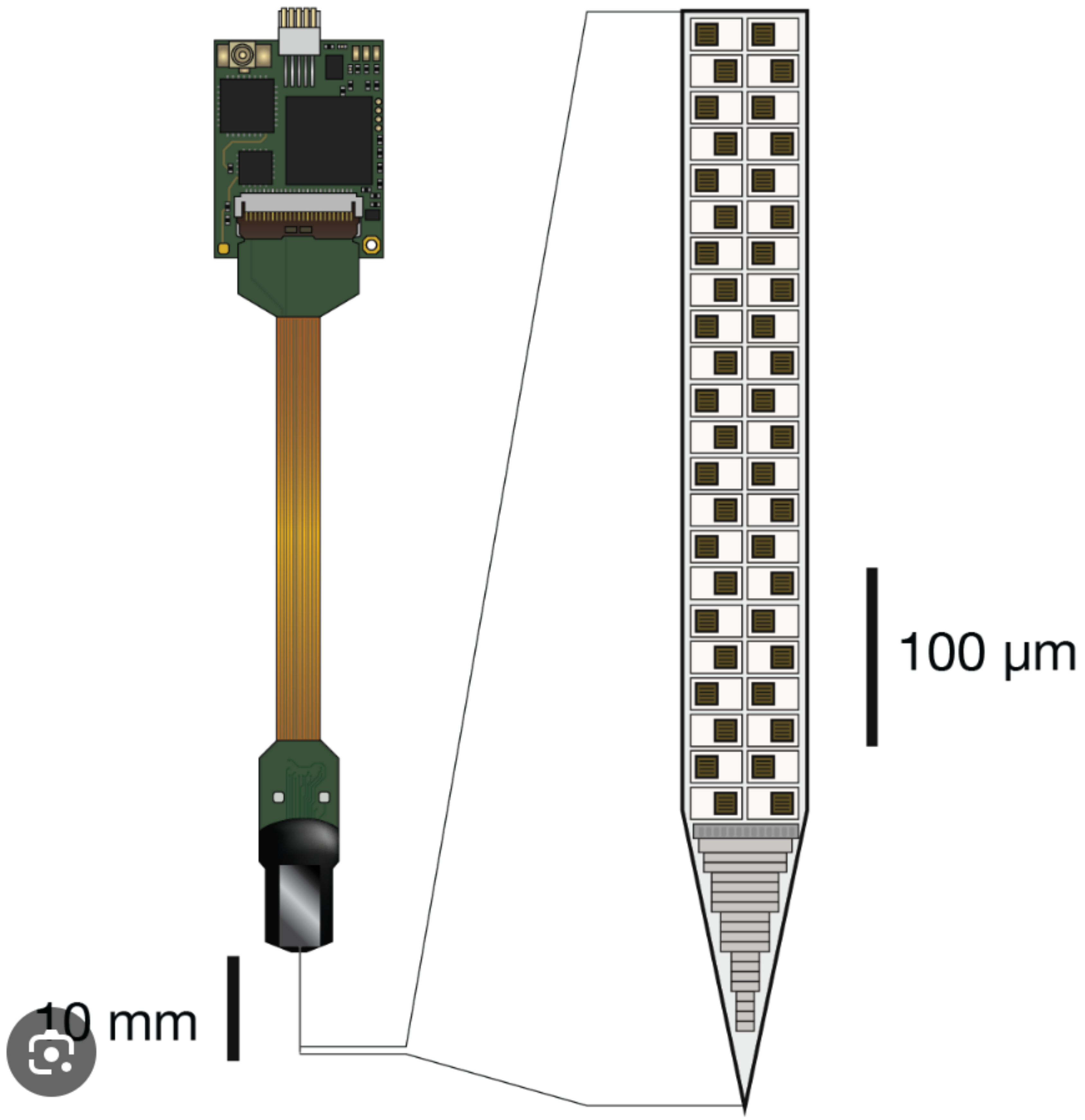
- Tetrodes
- Overlapping spikes
- Bursting cells
- Non-gaussian clusters

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

On to today…

Completing spike sorting, PCA and examples

Field potential data

- Let’s do some loading processing and filtering of field potential data
    - Install modules
    - Get data
    - Load data
    - Bandpass filtering

Filtering LFP data

PCA in neural data science