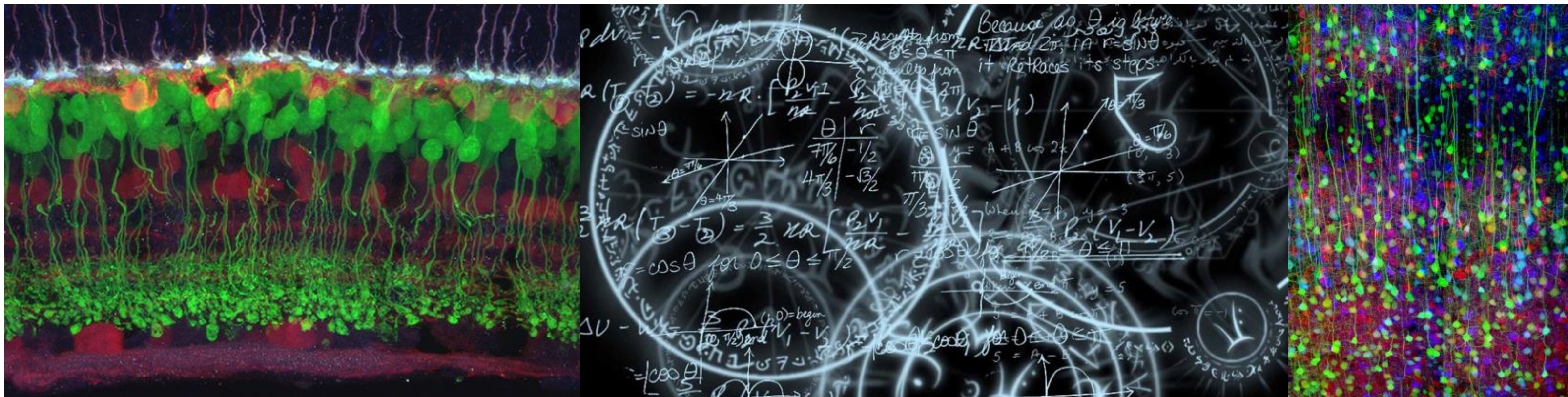
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

MEDIZINISCHE FAKULTÄT

Universitätsklinikum Tübingen



Neural Data Science

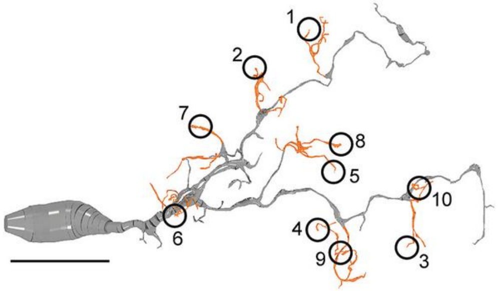
Spike detection and feature extraction

Prof. Dr. Philipp Berens
Hertie Institute for AI in Brain Health, Tübingen

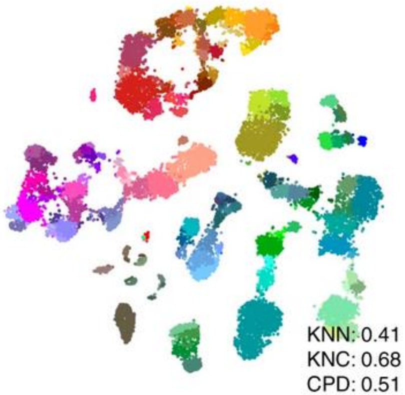
My lab

Topics & interests of the lab

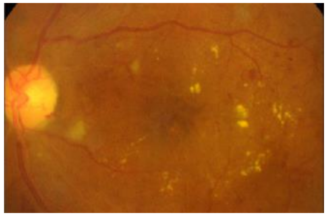
- ML for models of circuits & computations in the retina

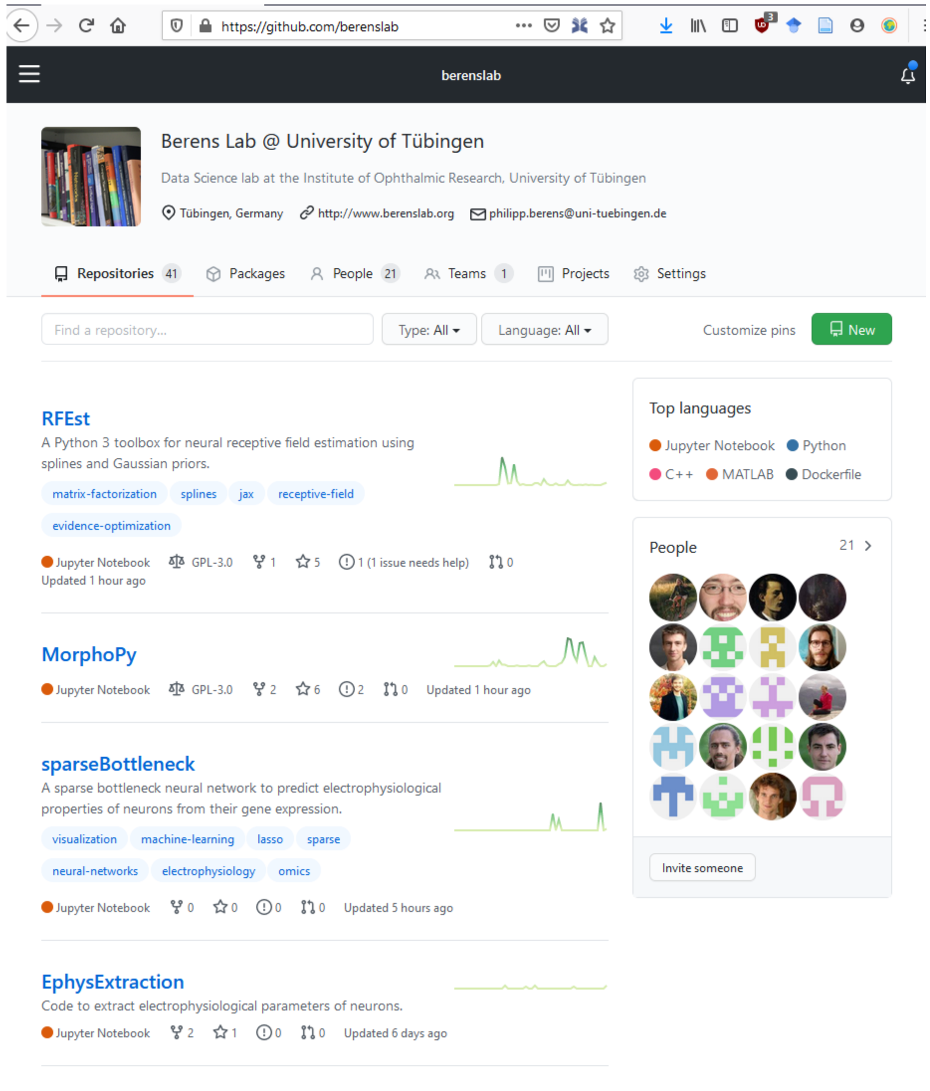


- ML for single cell transcriptomics in neuroscience



- AI-based medical diagnostics in ophthalmology



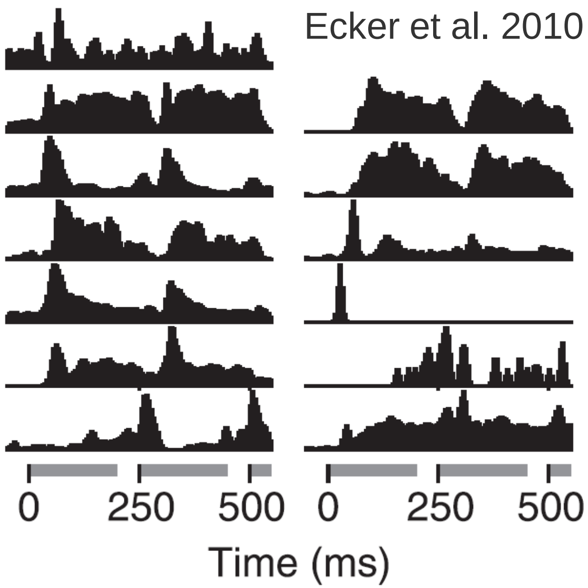

https://github.com/berenslab
berenslab
Berens Lab @ University of Tübingen
Data Science lab at the Institute of Ophthalmic Research, University of Tübingen
Tübingen, Germany
http://www.berenslab.org
philipp.berens@uni-tuebingen.de
Repositories 41
Packages
People 21
Teams 1
Projects
Settings
Find a repository...
Type: All
Language: All
Customize pins
New
RFEst
A Python 3 toolbox for neural receptive field estimation using splines and Gaussian priors.
matrix-factorization
splines
jax
receptive-field
evidence-optimization
Jupyter Notebook GPL-3.0 1 5 1 (1 issue needs help) 0
Updated 1 hour ago
MorphoPy
Jupyter Notebook GPL-3.0 2 6 2 0 Updated 1 hour ago
sparseBottleneck
A sparse bottleneck neural network to predict electrophysiological properties of neurons from their gene expression.
visualization
machine-learning
lasso
sparse
neural-networks
electrophysiology
omics
Jupyter Notebook 0 0 0 0 Updated 5 hours ago
EphysExtraction
Code to extract electrophysiological parameters of neurons.
Jupyter Notebook 2 1 0 0 Updated 6 days ago
Top languages
Jupyter Notebook
Python
C++
MATLAB
Dockerfile
People 21
Invite someone

Why is there this course?

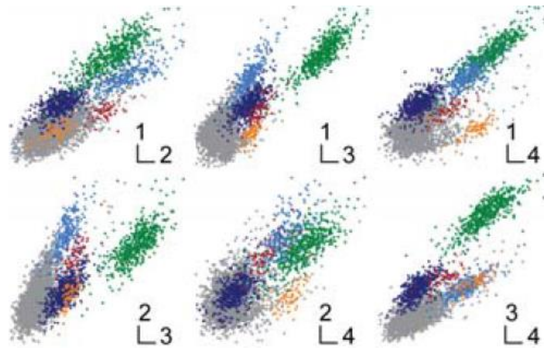
- You know about
  - Machine Learning
  - Signal Processing
  - Neural Coding
  - Neural Dynamics
  - Bioinformatics

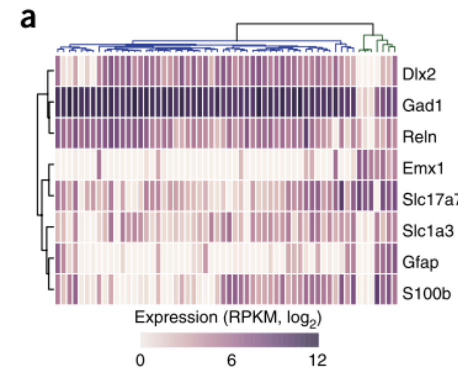
- So why care?
  - Different types of data (time series, images, graphs…)
  - Practical approach to data science (independent of topic)

Neural data

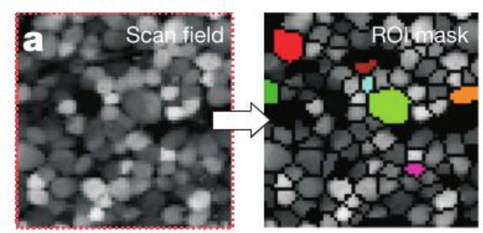


Ecker et al. 2010

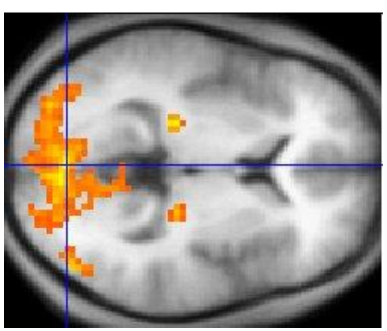




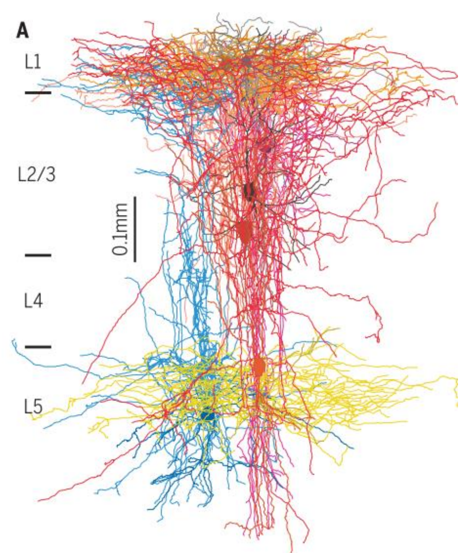
Cadwell et al. 2016



Baden et al. 2016

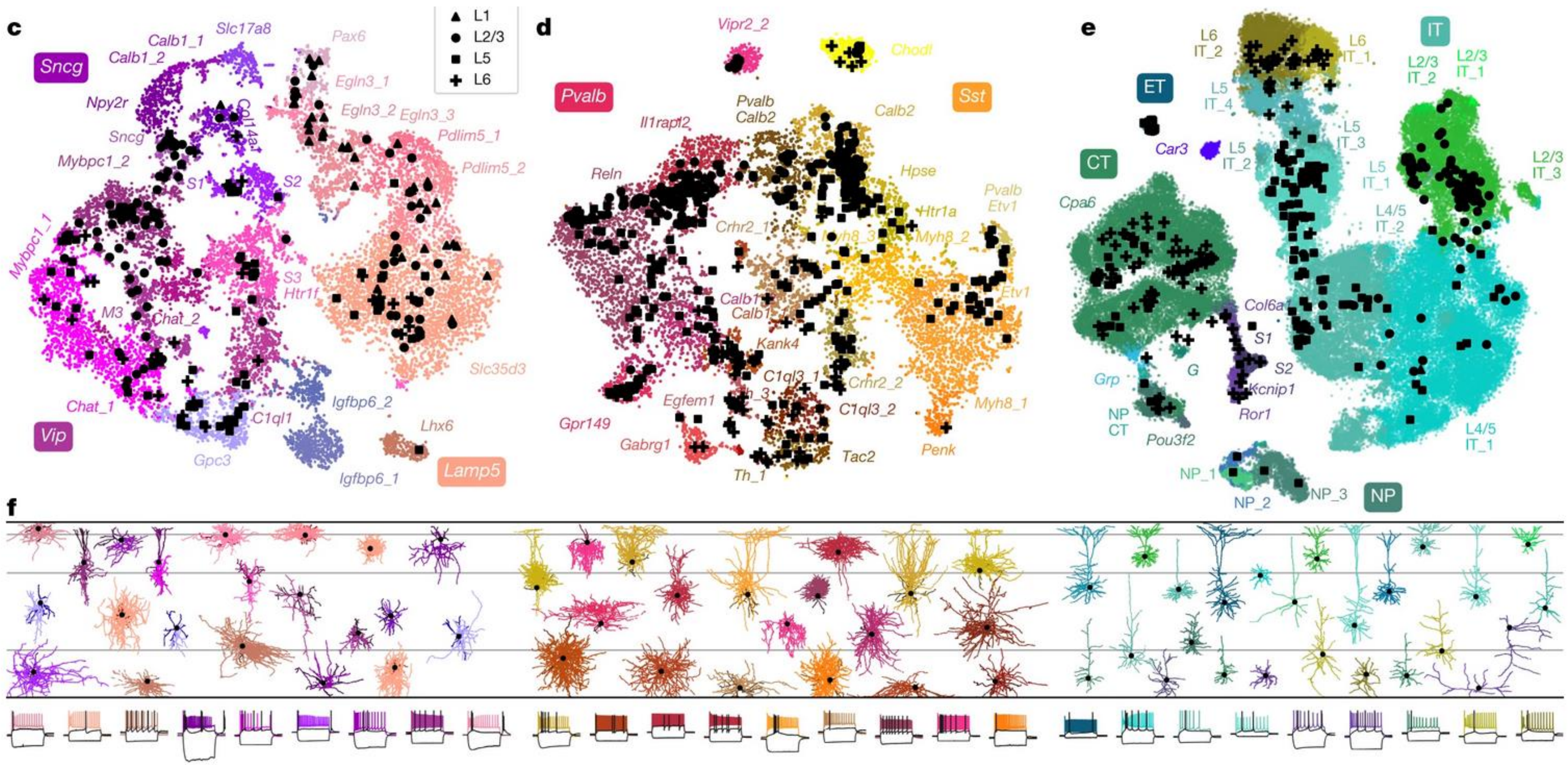


Wikipedia



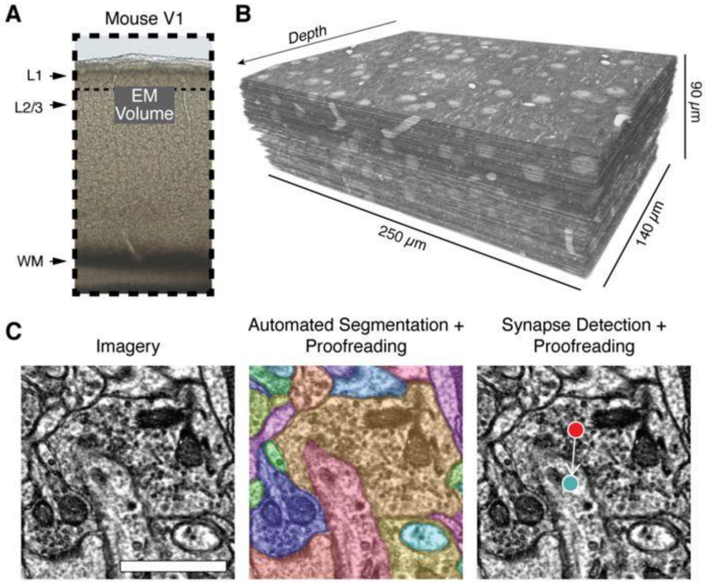
Jiang et al. 2015

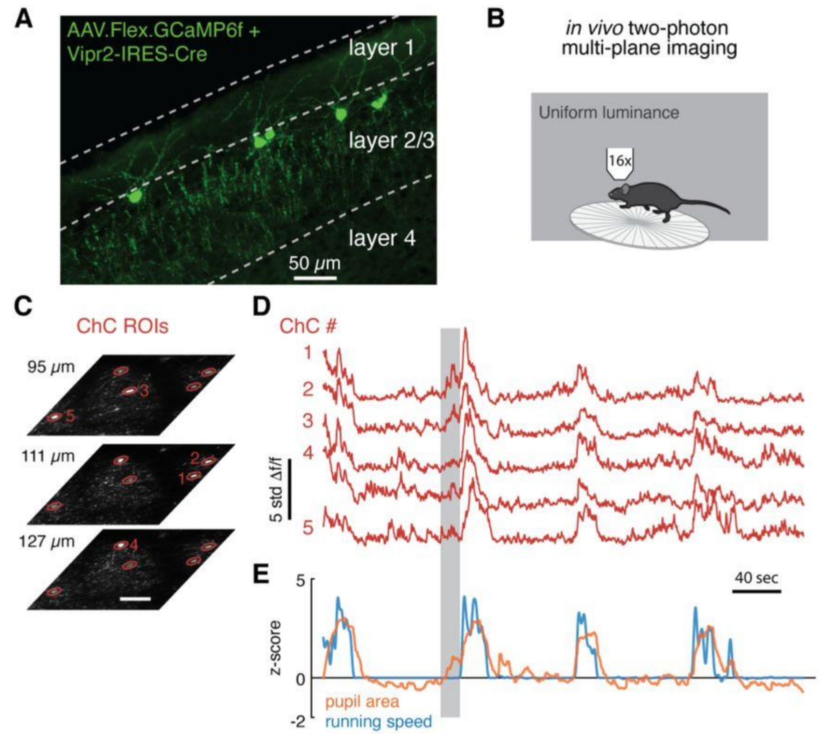
Multimodal data

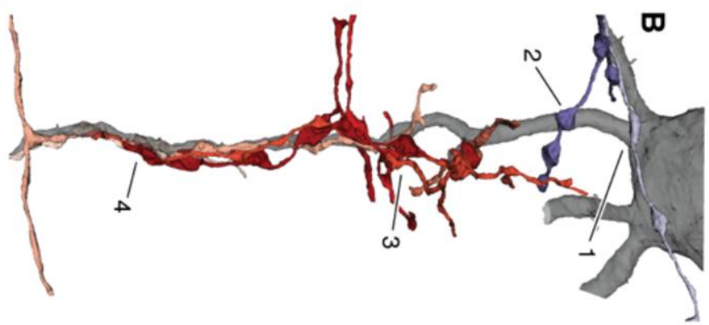


Scala, Kobak et al. Nature 2020

Multimodal data





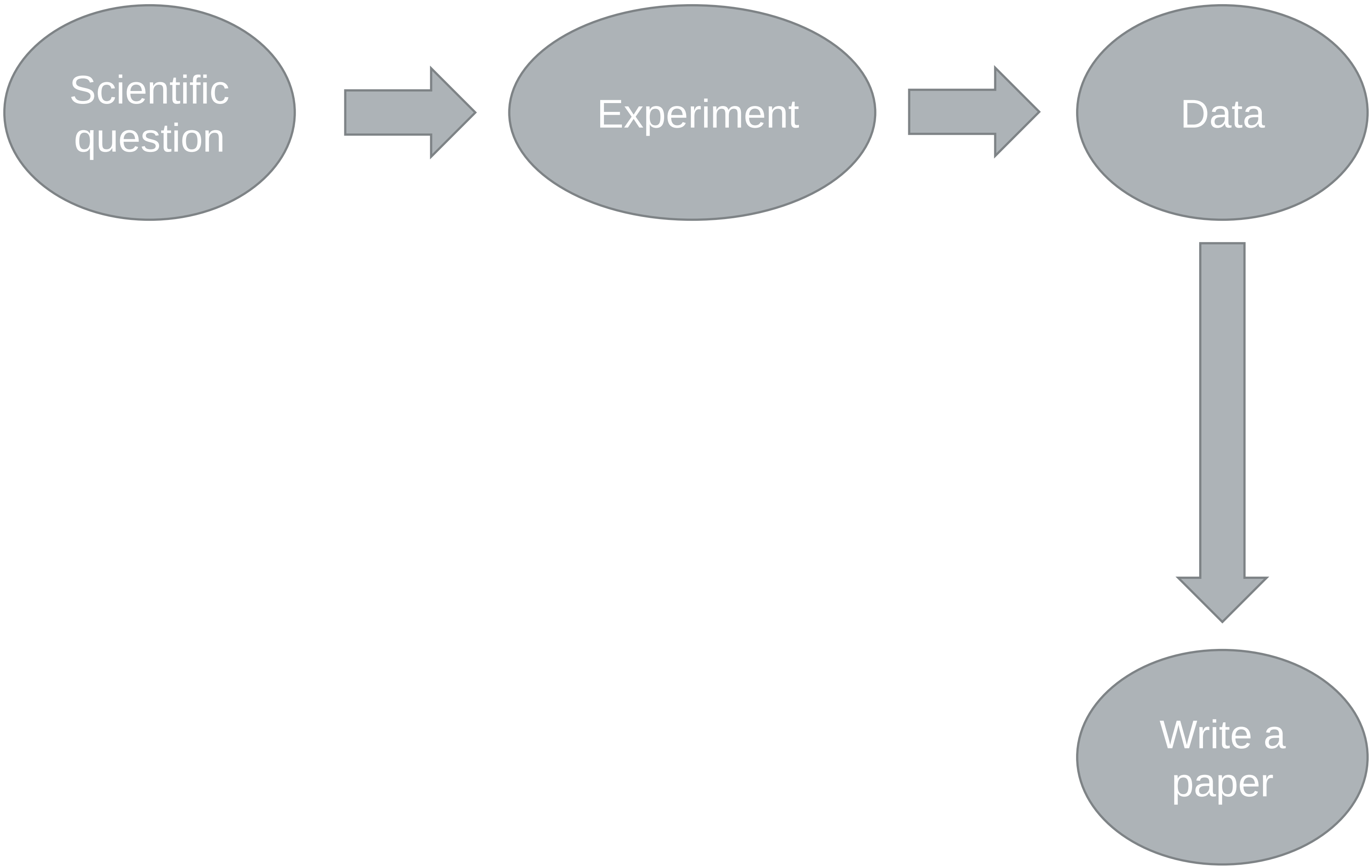


https://www.biorxiv.org/content/10.1101/2020.03.31.018952v1

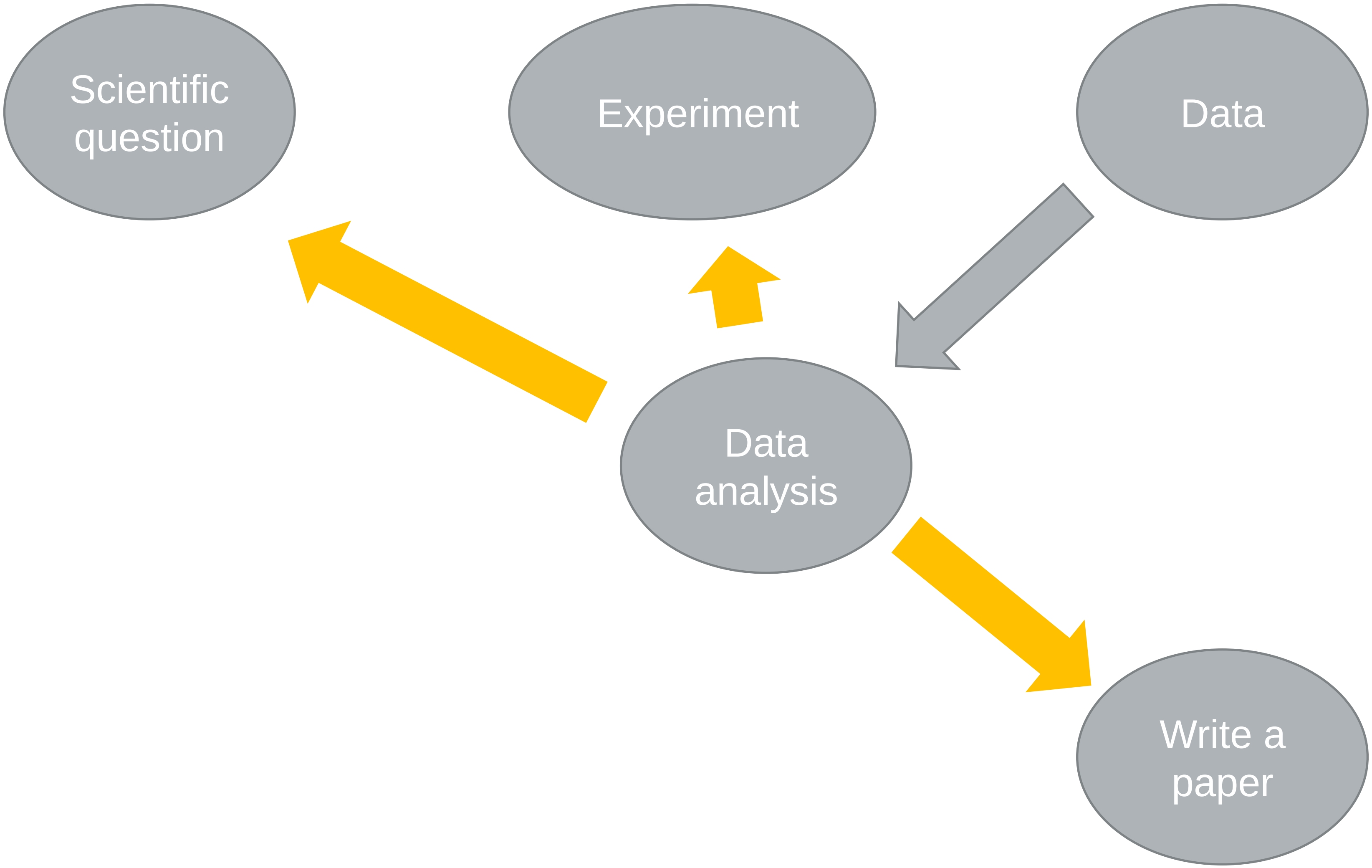
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Universitätsklinikum Tübingen

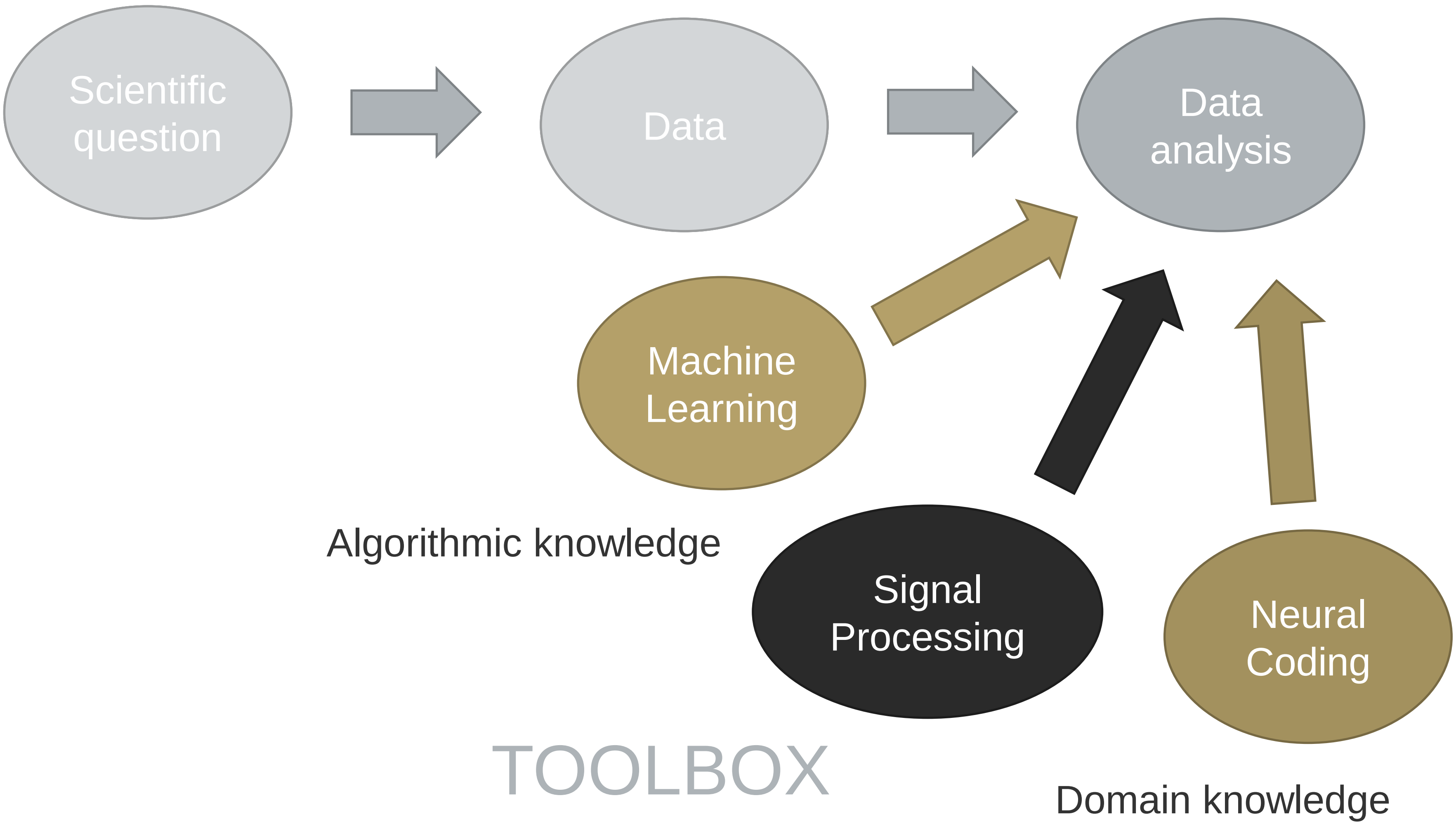
An ideal world



An alternative view



How Does Data Analysis Work?



Learning targets

- Learn about common problems in neural data science
- Analyse different data types
- Implement and apply basic algorithms to solve them
- Learn about advanced solutions
- Learn how to analyse complex real world datasets

- Practise a data science philosophy
  - Plot raw data
  - Look at examples
  - Test algorithms on artificial data
  - Be critical with your results
  - Let the problem determine the technique

What knowledge is helpful for the course?

- Python programming
- Machine learning knowledge
  - Clustering
  - Regression
  - Classification
- Some neuroscience context
  - Ask if you don't understand the purpose of something!

Ressources

- Anaconda Python distribution [link]
- Python tutorial on collab
- Youtube lecture Introduction to Machine Learning
- Lab github
- Lab website

How does this course work? (1/2)

- Lecture (Wednesday, 10-12 am)
  - Problem statement & possible solutions
  - Algorithmic details and application issues
  - Goal: working knowledge of important problems & techniques, literature overview

- Tutorial (Friday, 12am-2pm)
  - Tutors: Jonas Beck, Jan Lause & Ziwei Huang
  - 10 min presentation by PhD student
  - Discuss problem statement and open questions
  - Live coding & best practice feedback

How does this course work? (2/2)

- Programming exercises (60% of grade)
  - Work on real-world problems and **do** analysis
  - Implement algorithms discussed in lecture as model solution
  - Start programming on Fridays in tutorial
  - Work in groups of 2-3 students

- Final project (40% of grade)
  - Work on final project after end of June
  - Describe problem statement, solutions, findings, well-prepared figures, brief discussion
  - In Jupyter Notebook

- No exam

Policies

- No late submissions.

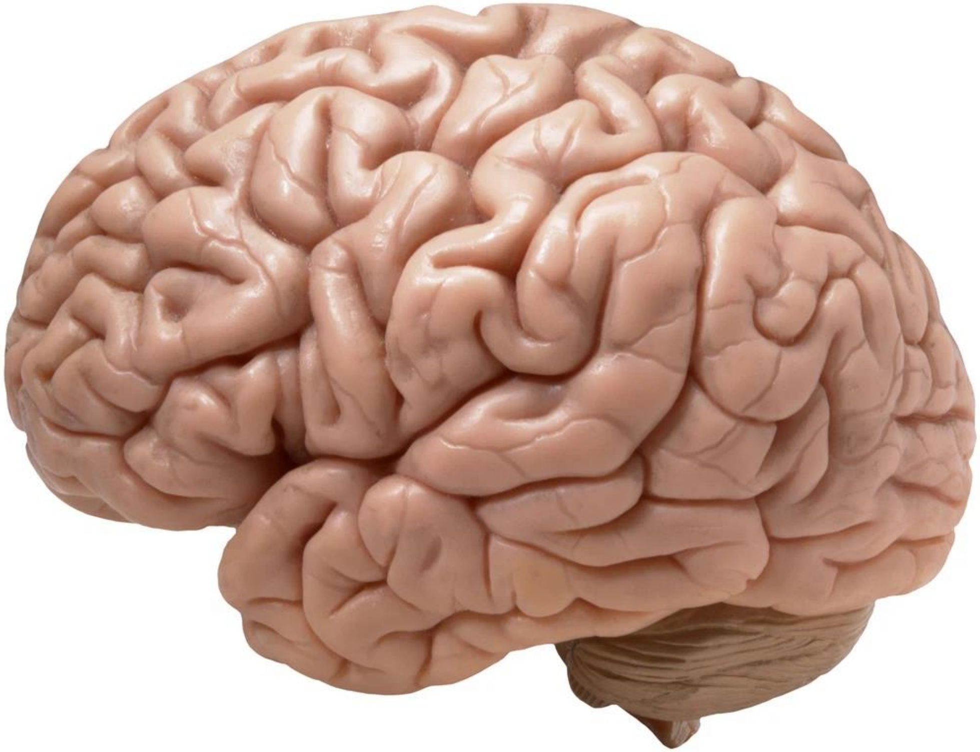
- Use any available material, including Python libraries (unless we specifically ask you not to do this).

- Talk to others about the problems, but
  - credit any source in the code
  - write your own code.

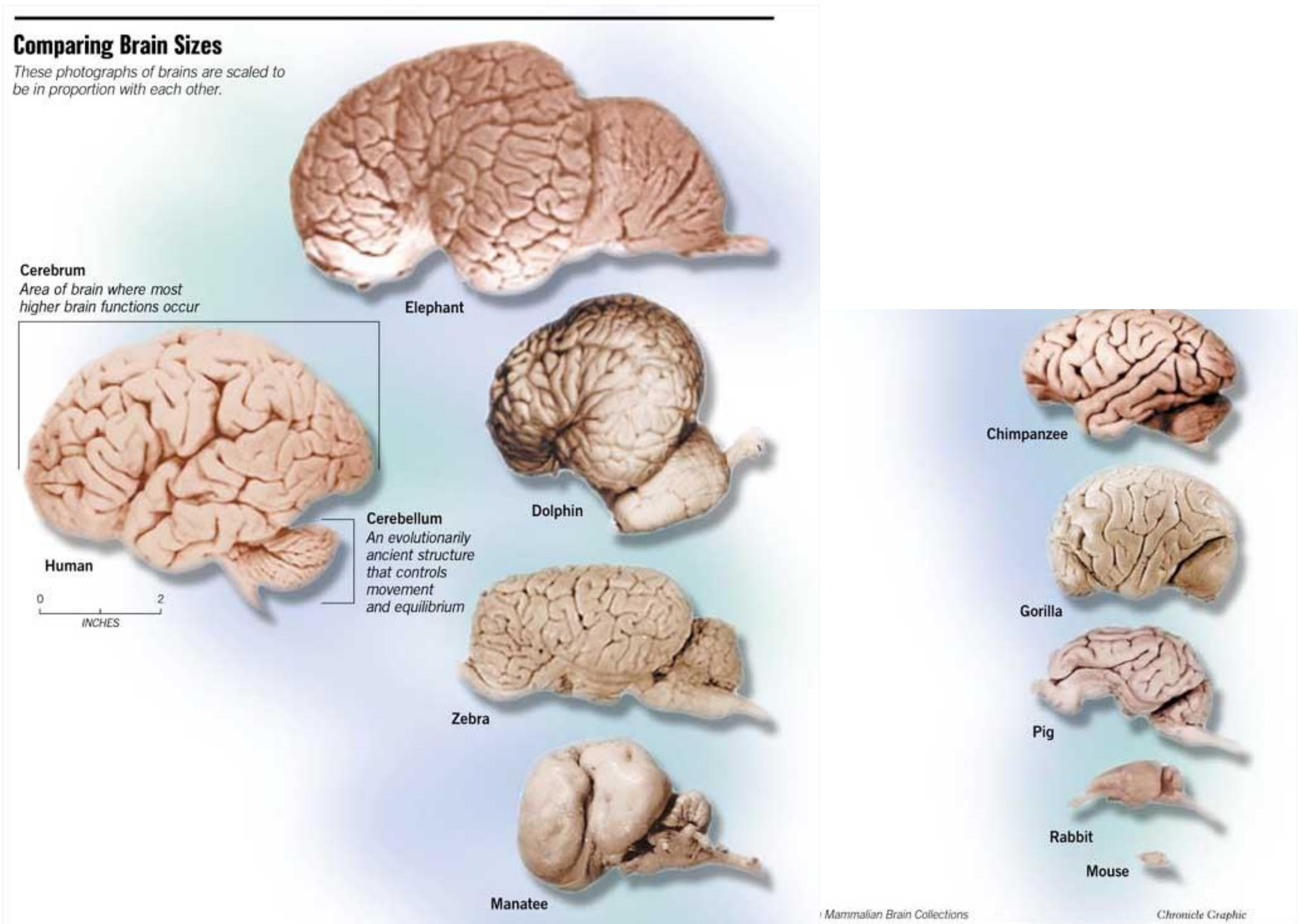
- Work as a team.

- If you have feedback how to improve the course, talk to us.

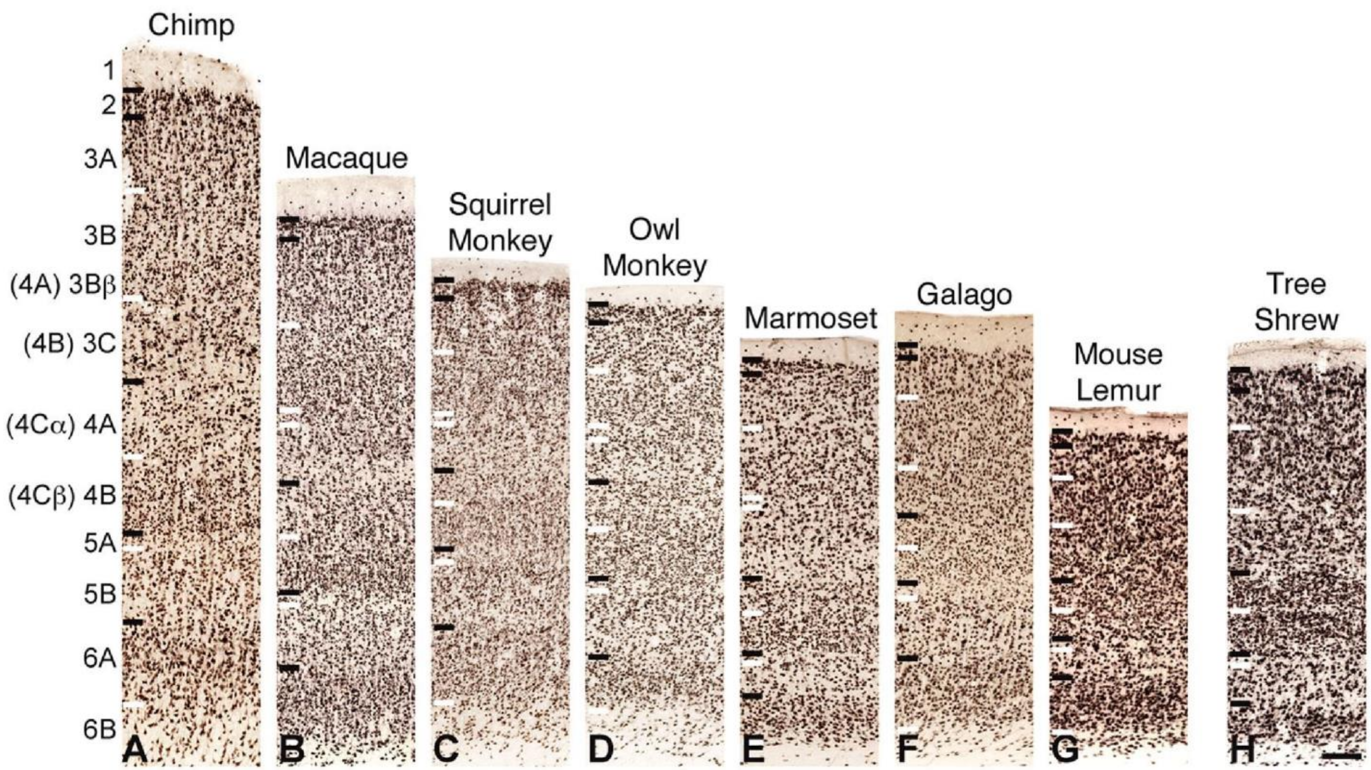
The human brain



Other brains

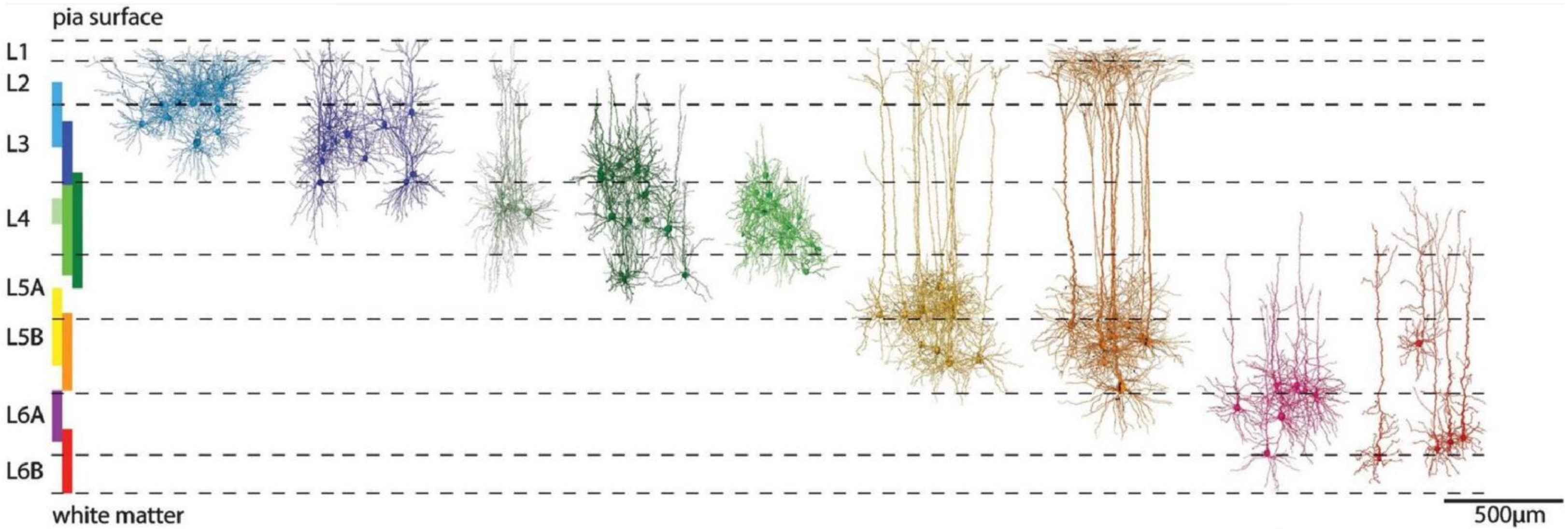


Layered cortex



Balaram & Kaas 2014 Front Neuroanat

Different cells

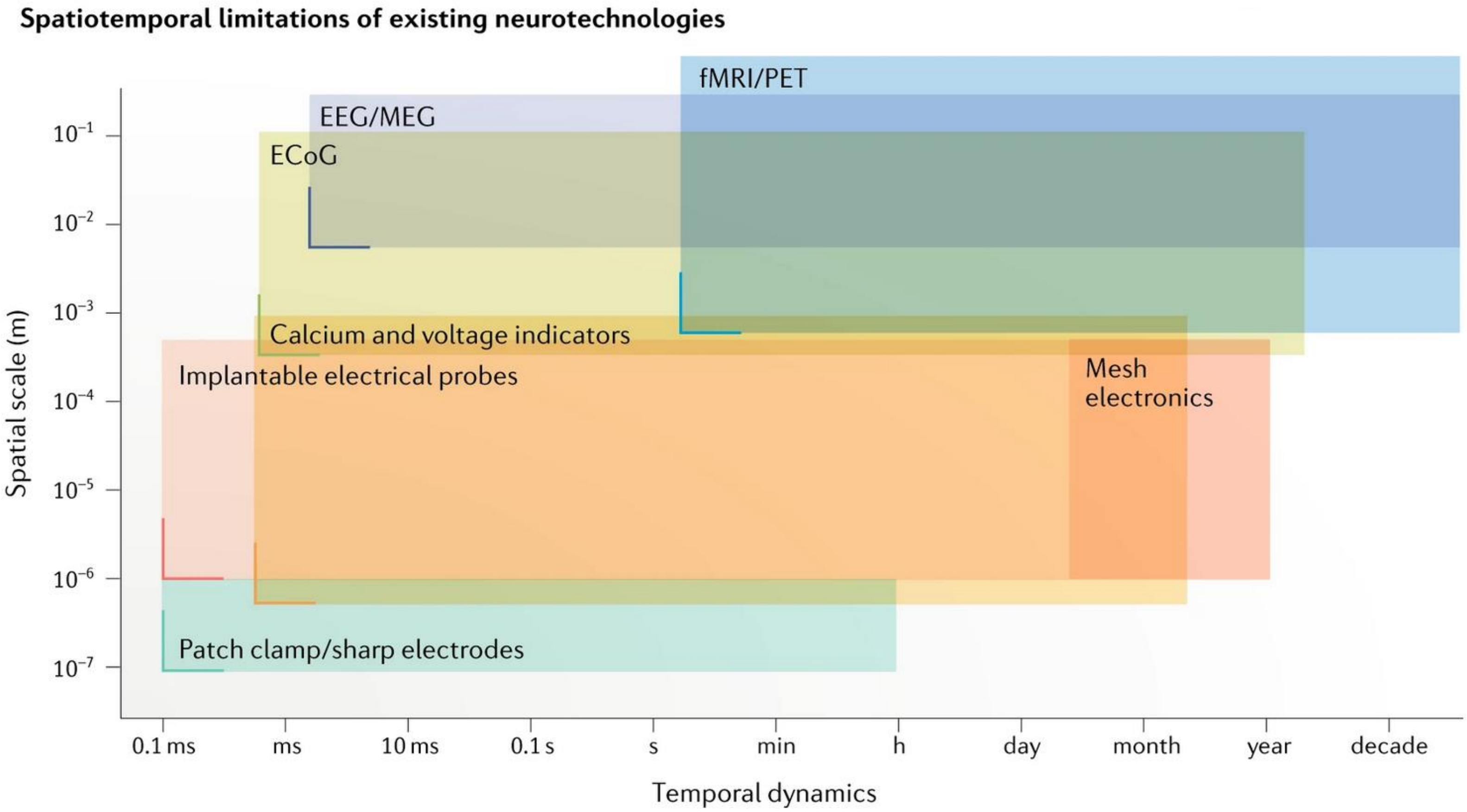


Oberlaender et al. 2012

Eberhard Karls Universität Tübingen

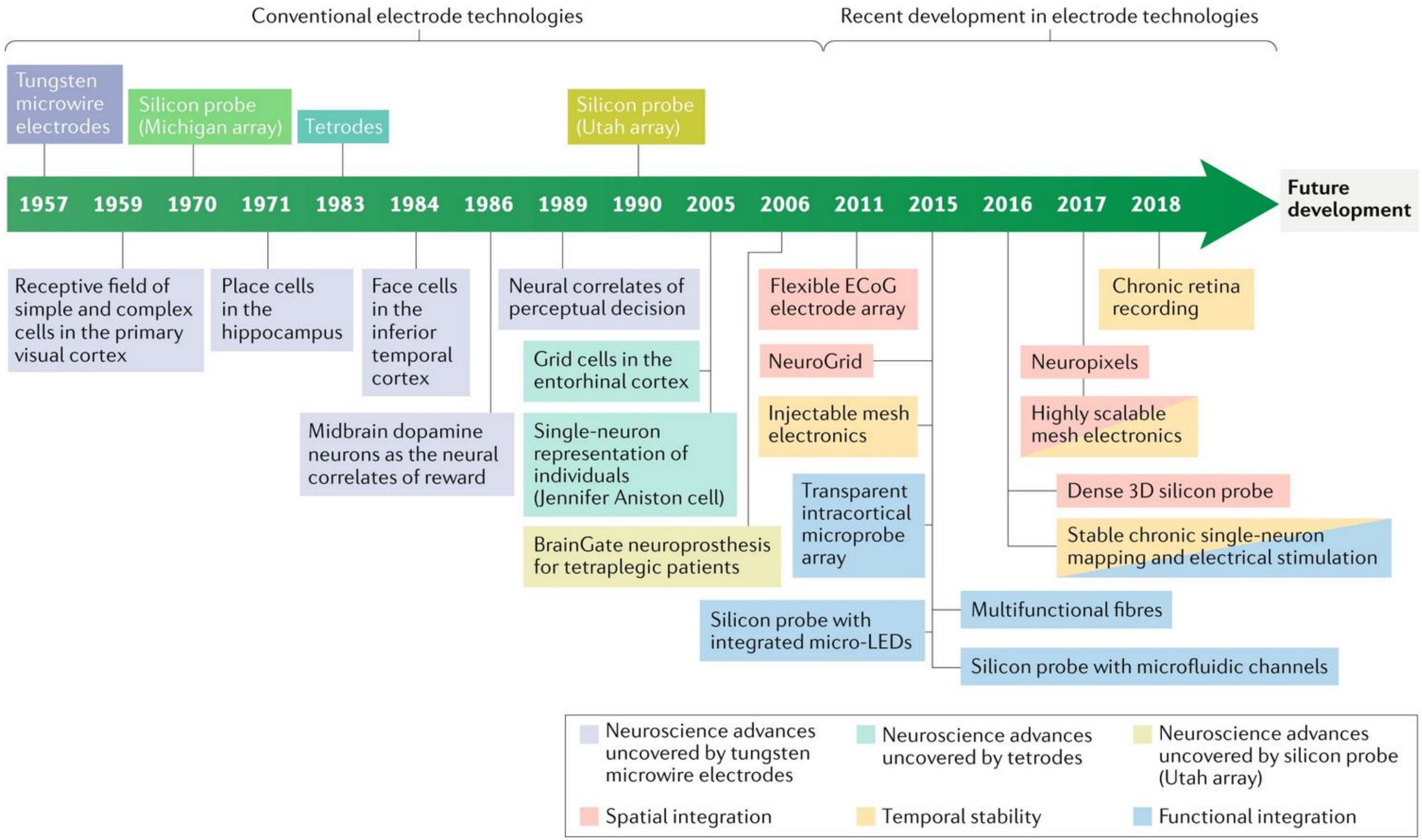
Universitätsklinikum Tübingen

Methods for recording brain activity



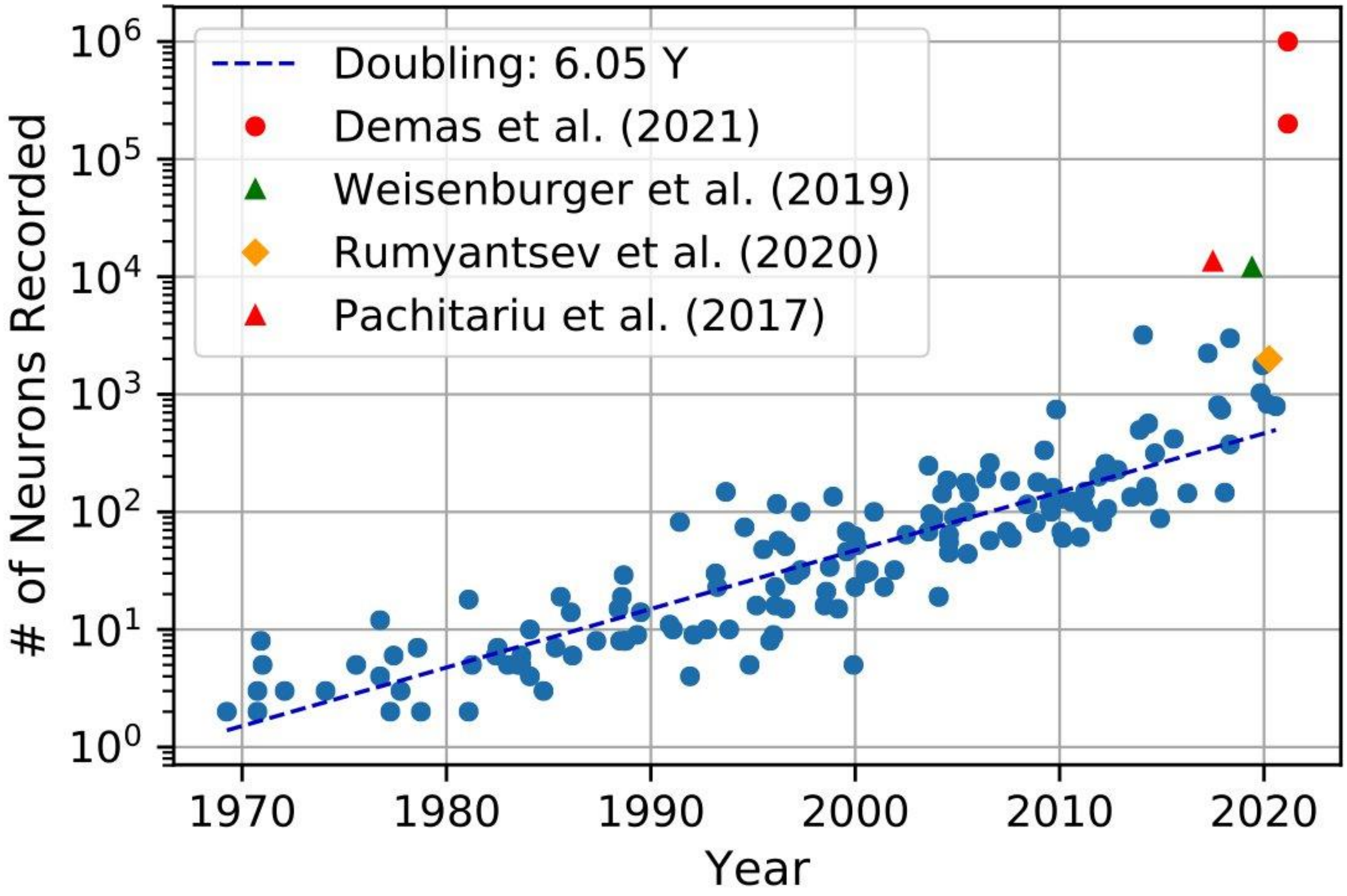
Hong & Lieber, 2019

Methods for recording brain activity

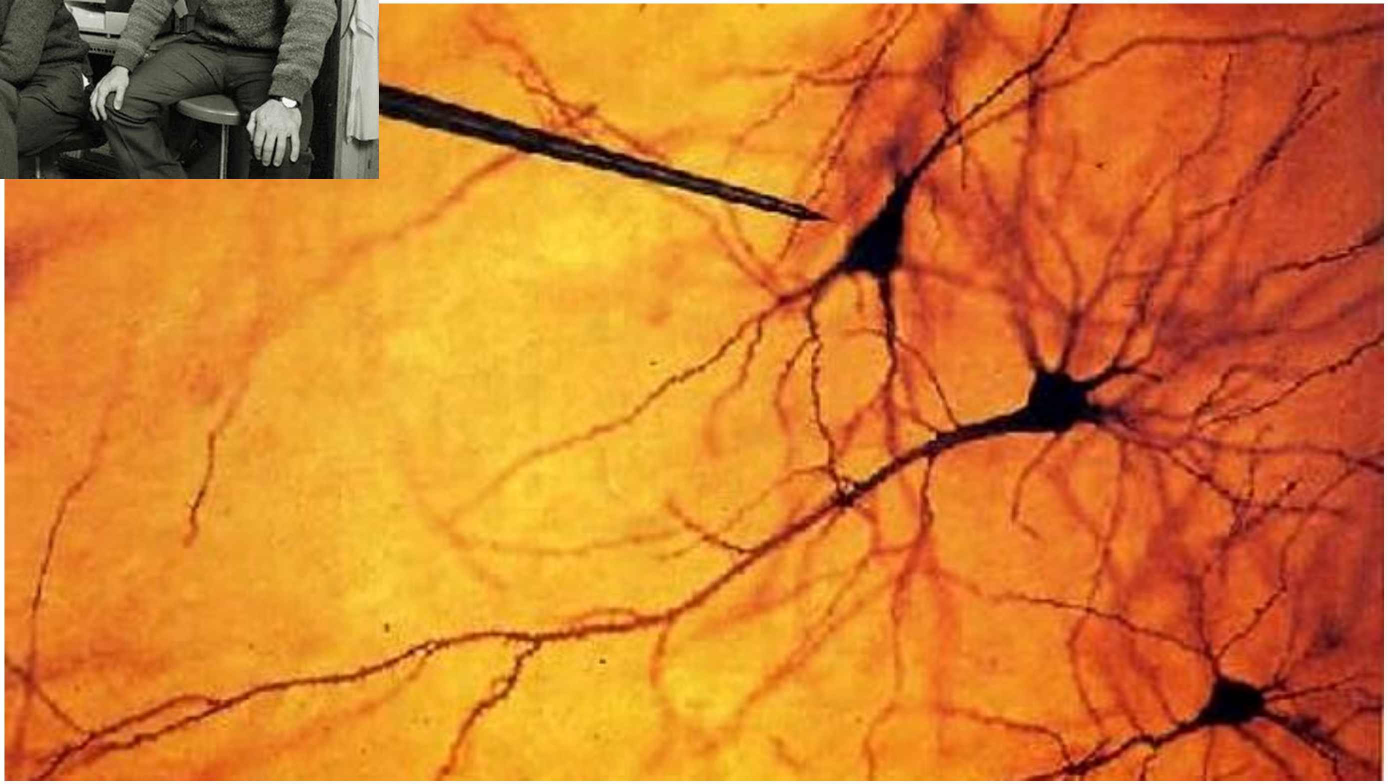


Hong & Lieber, 2019

New techniques allow more neurons

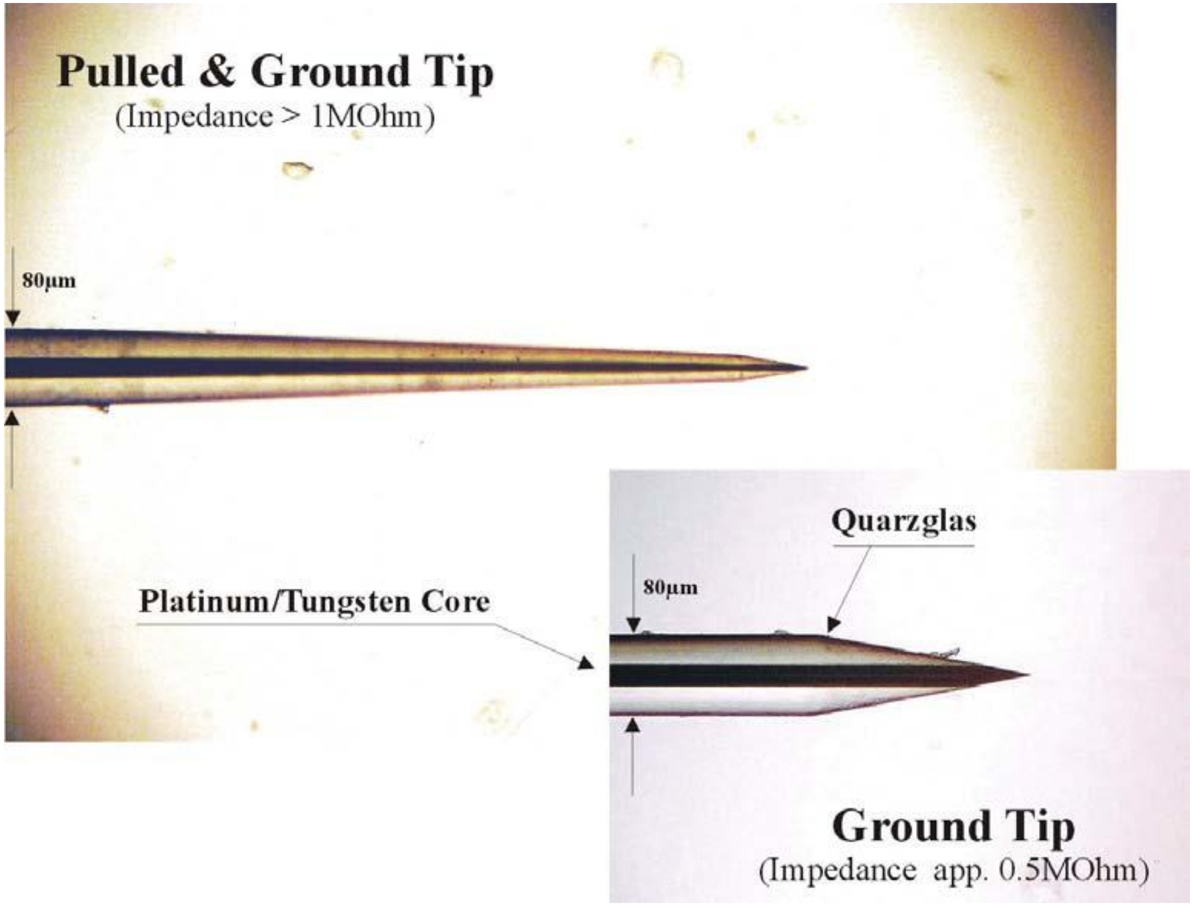


David Hubel & Thorsten Wiesel

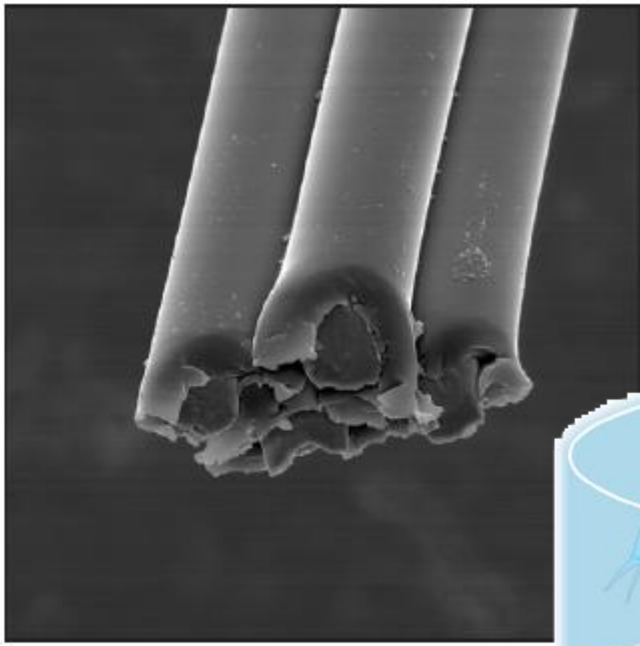


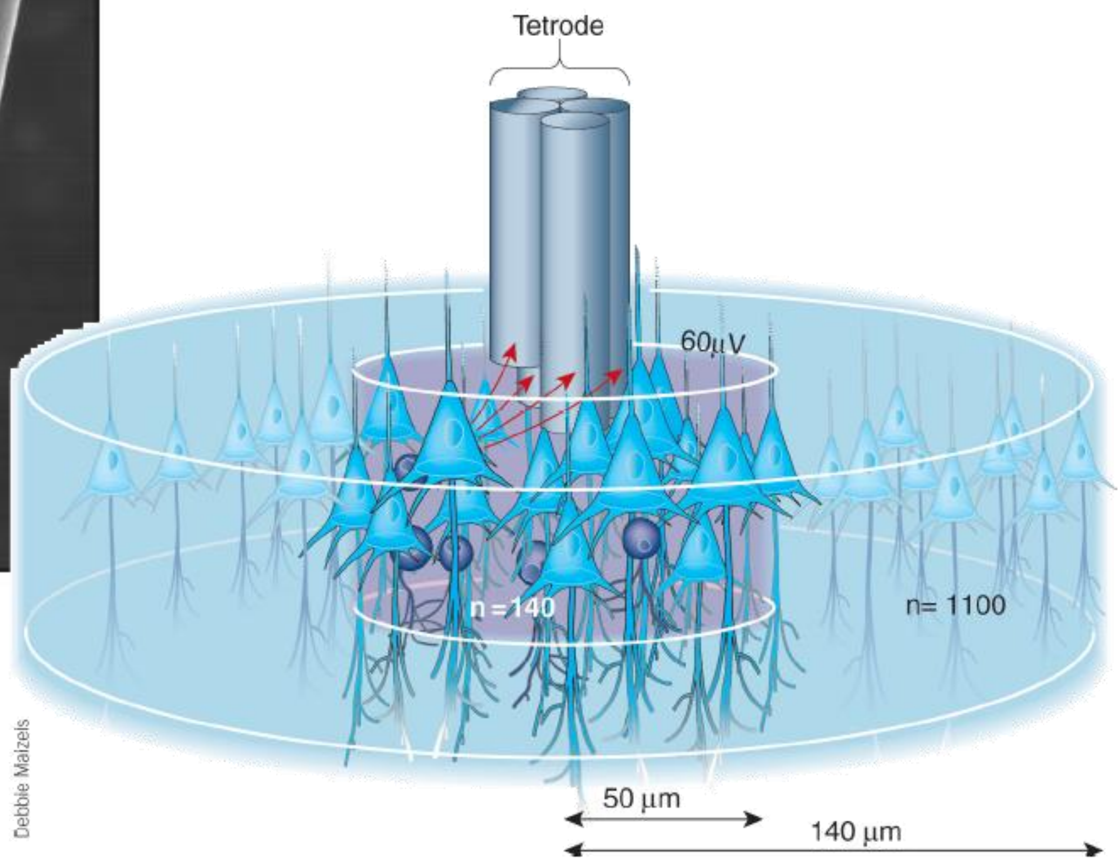
harvard.edu

Tungsten microelectrodes

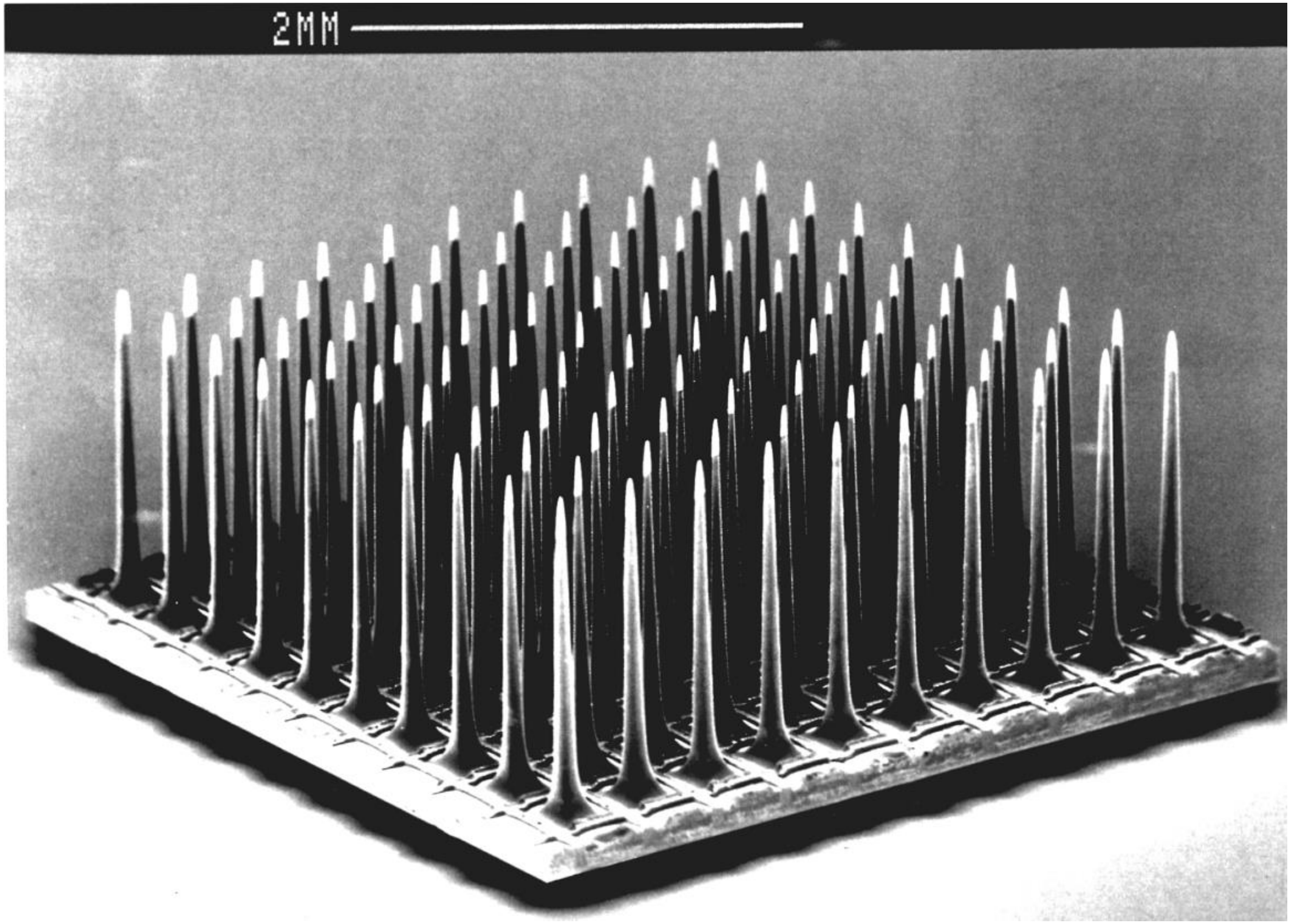


Tetrodes



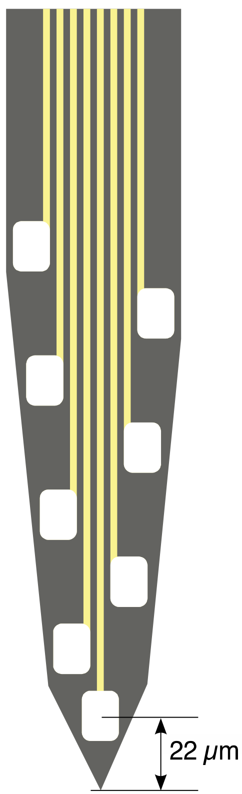


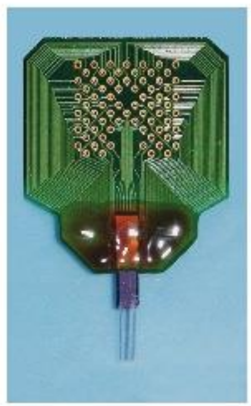
Utah array

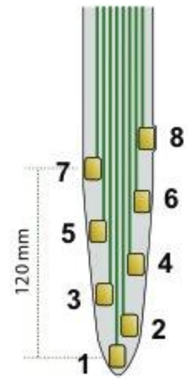


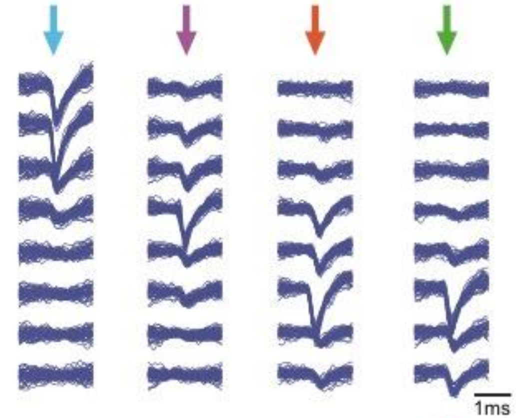
Silicon probes

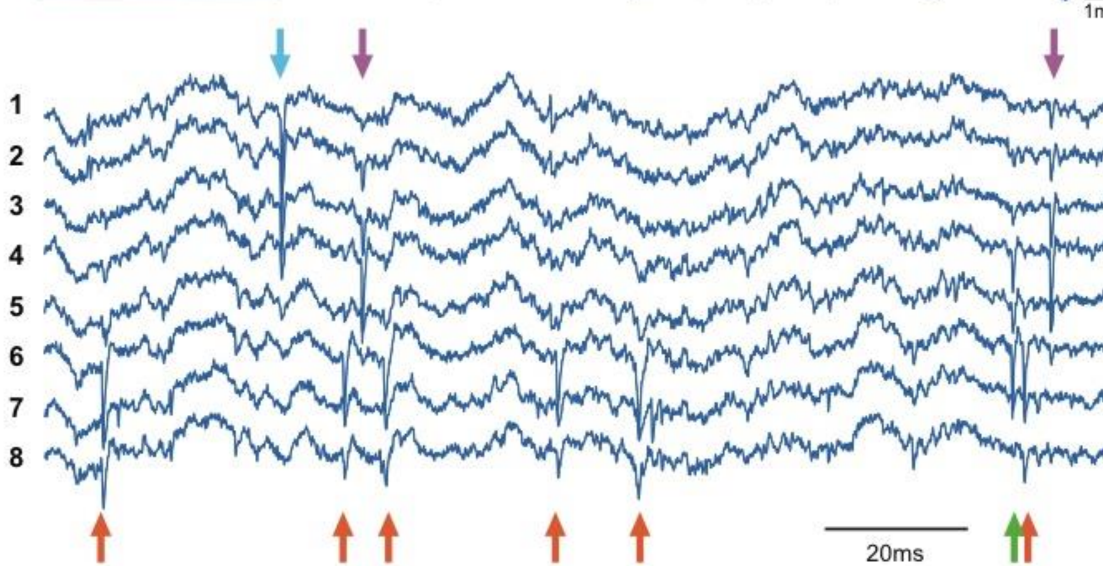
NeuroNexus

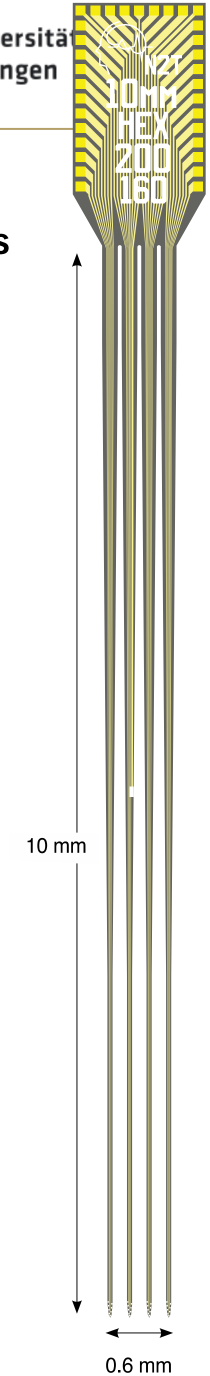




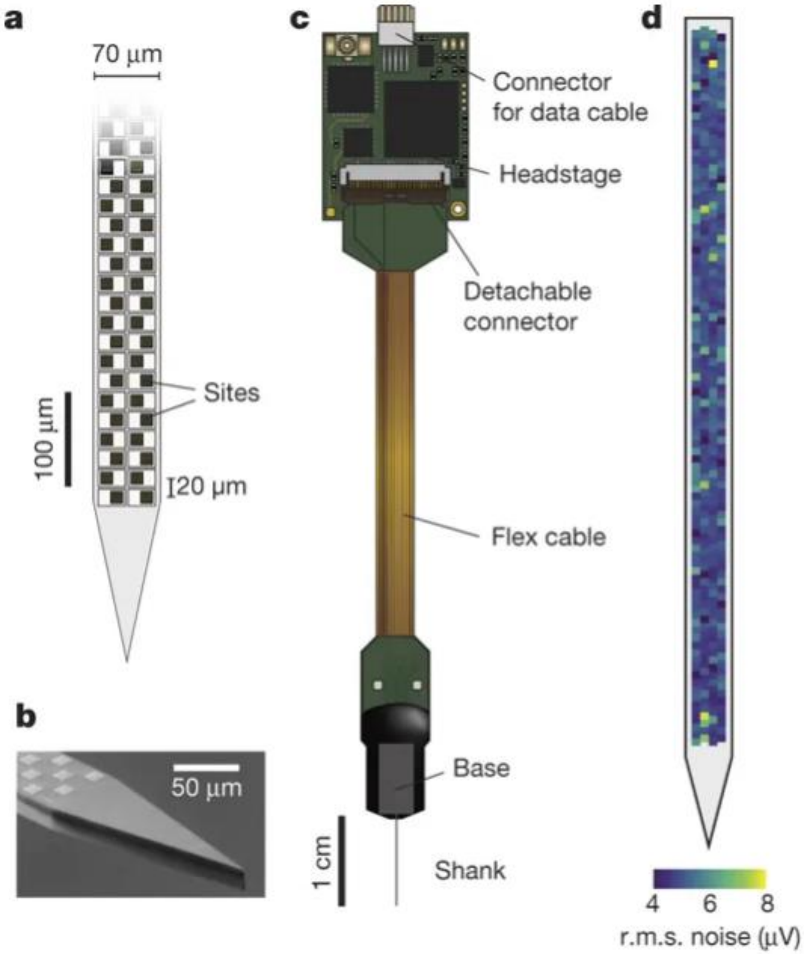


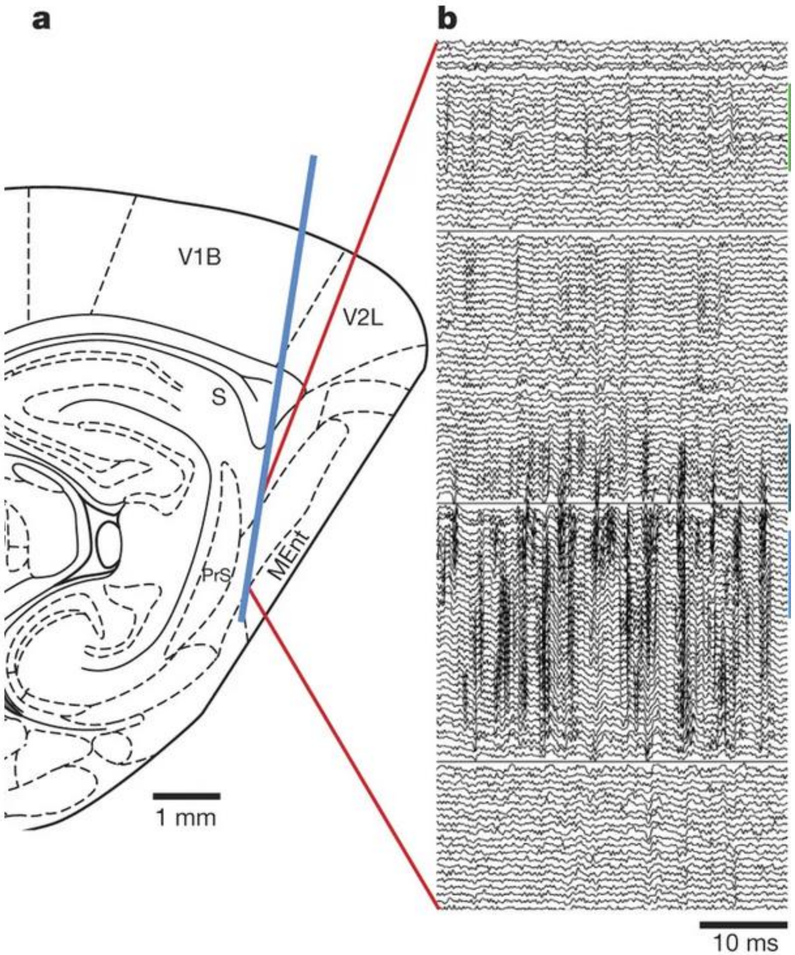






Neuropixels





960 sites

Jun et al, Nature, 2017

Neuropixels

Science

Current Issue First release papers Archive About Submit manusc

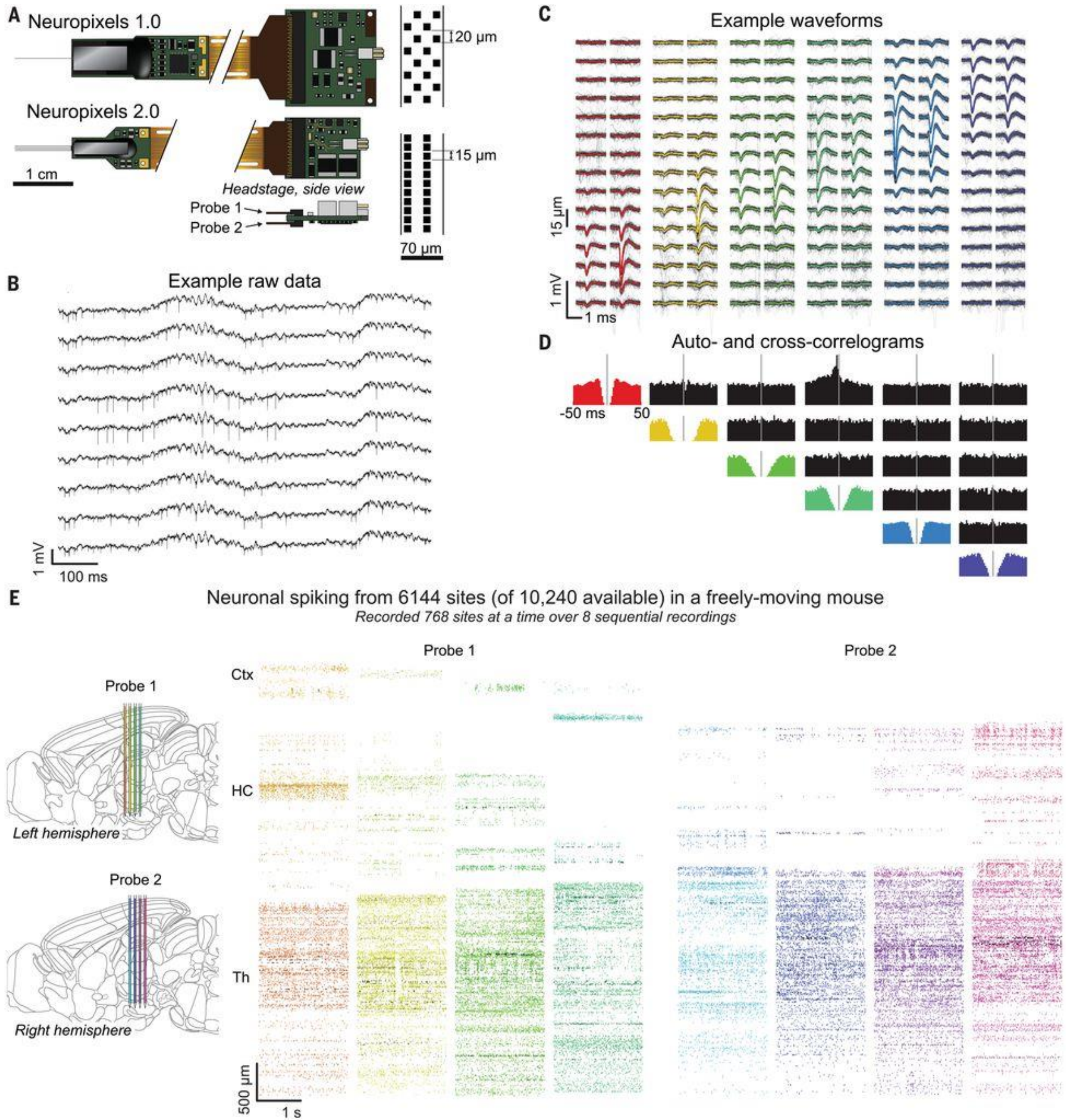
HOME > SCIENCE > VOL. 372, NO. 6539 > NEUROPIXELS 2.0: A MINIATURIZED HIGH-DENSITY PROBE FOR STABLE, LONG-TERM BRAIN RECORDINGS

RESEARCH ARTICLE

Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings

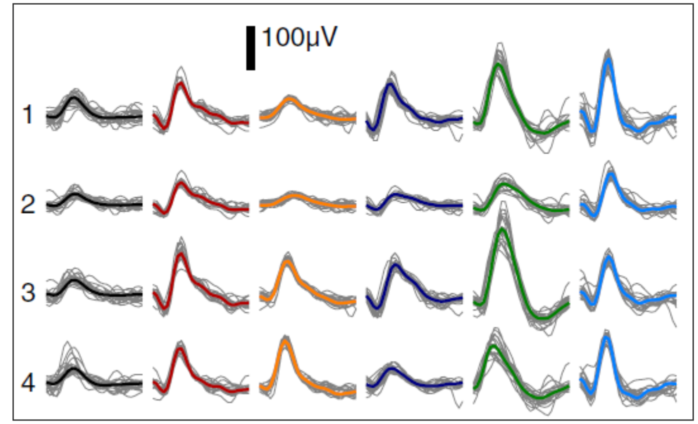
NICHOLAS A. STEINMETZ, CAGATAY AYDIN, ANNA LEBEDEVA, MICHAEL OKUN, MARIUS PACHITARIU, MARIUS BAUZA, MAXIME BEAU, JAI BHAGAT, CLAUDIA BÖHM, [...], AND TIMOTHY D. HARRIS +29 authors Authors Info & Affiliations

SCIENCE • 16 Apr 2021 • Vol 372, Issue 6539 • DOI: 10.1126/science.abf4588

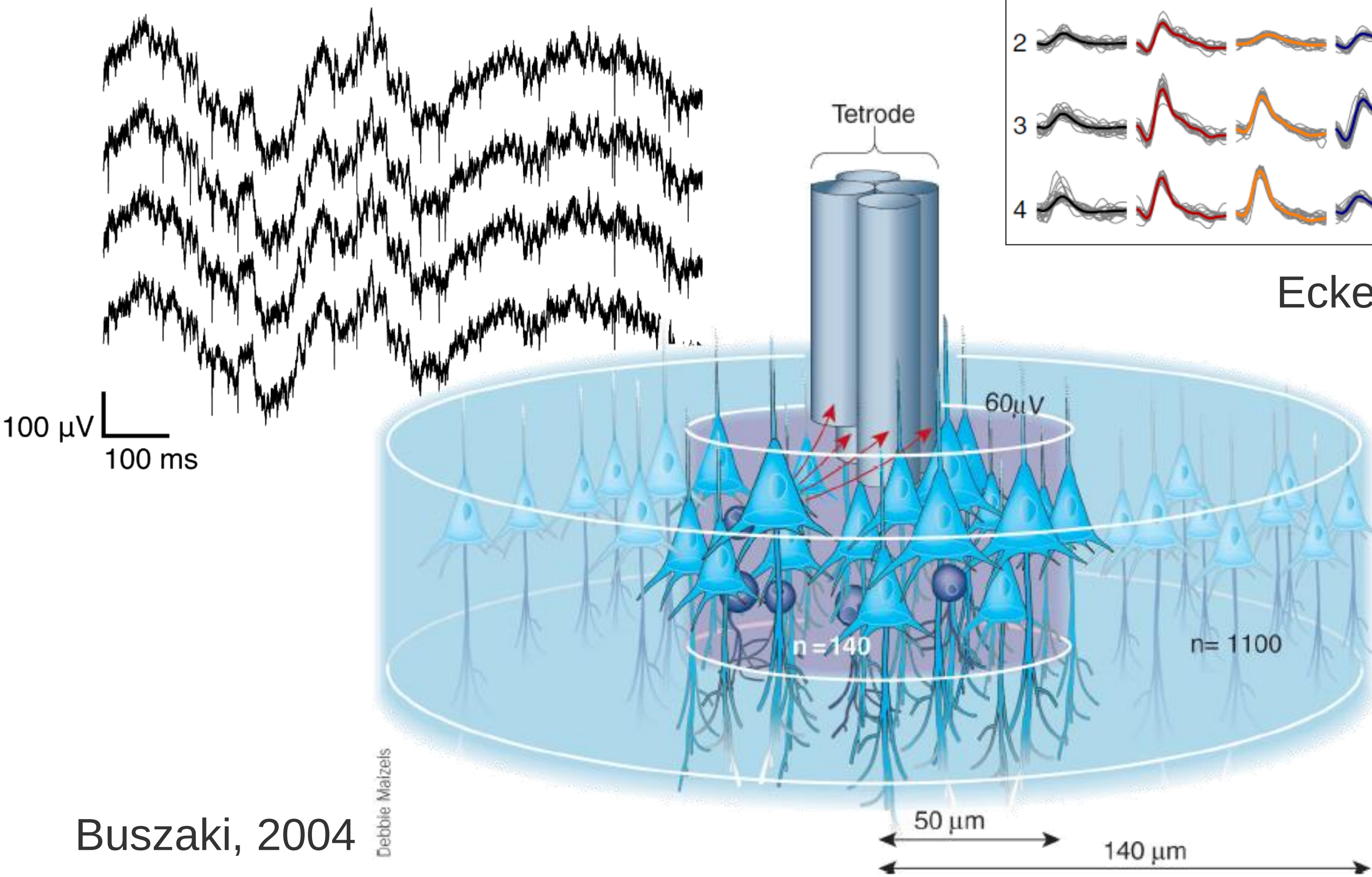


A Neuropixels 1.0
Neuropixels 2.0
1 cm
Headstage, side view
Probe 1
Probe 2
20 µm
15 µm
70 µm
B Example raw data
1 mV
100 ms
C Example waveforms
15 µm
1 mV
1 ms
D Auto- and cross-correlograms
-50 ms
50
E Neuronal spiking from 6144 sites (of 10,240 available) in a freely-moving mouse
Recorded 768 sites at a time over 8 sequential recordings
Probe 1
Probe 2
Ctx
HC
Th
Left hemisphere
Right hemisphere
500 µm
1 s

The Problem

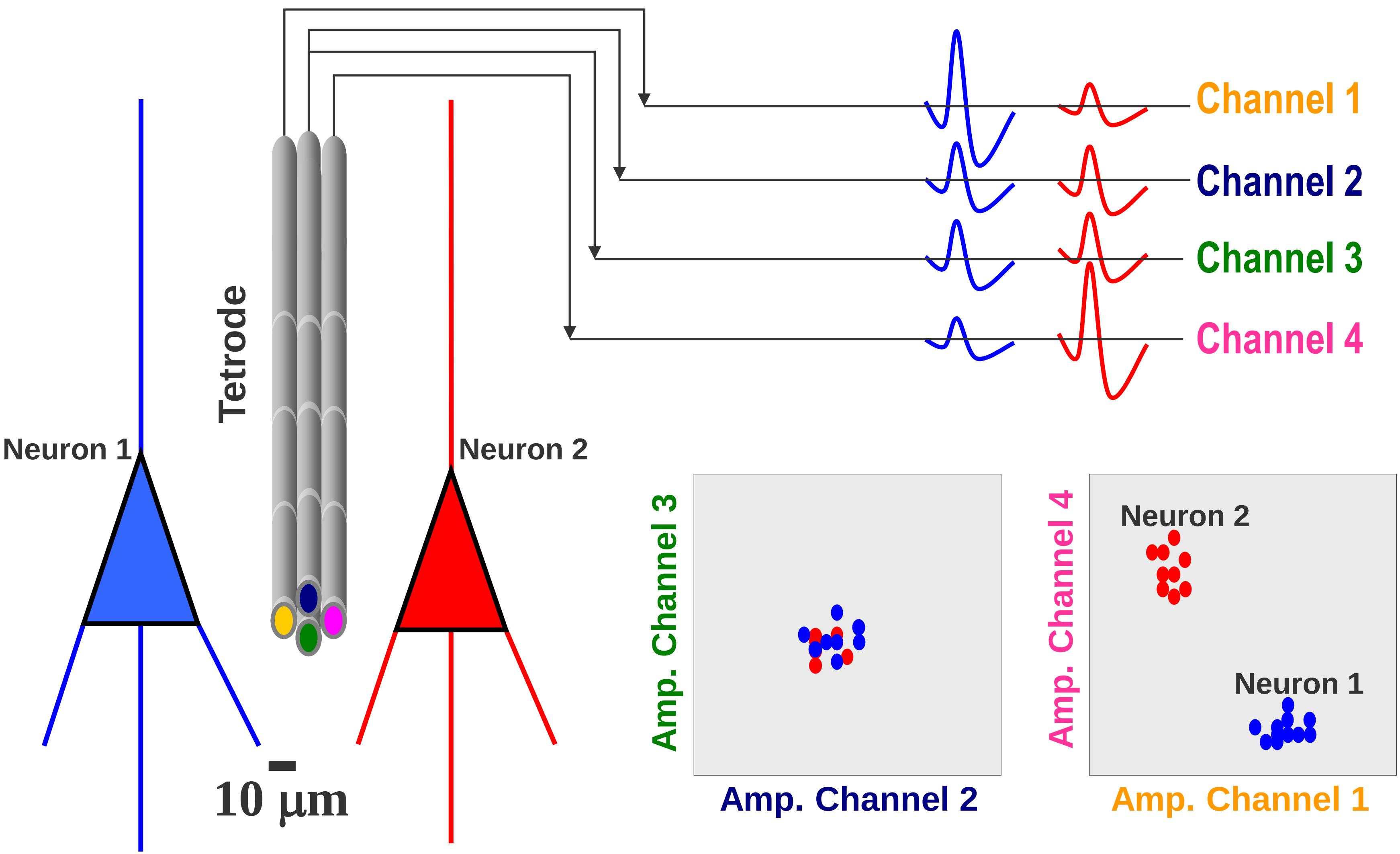


Ecker et al., 2010



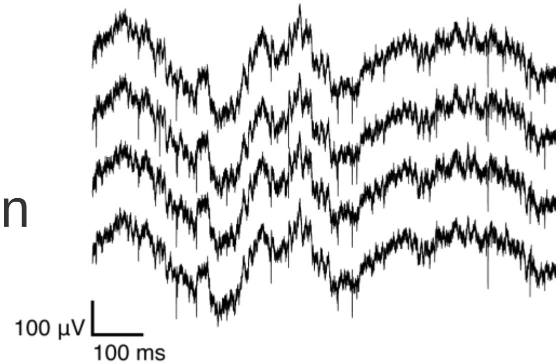
Buszaki, 2004

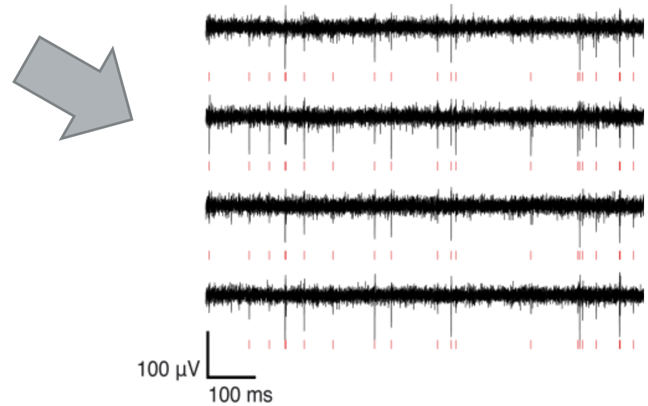
Triangulation

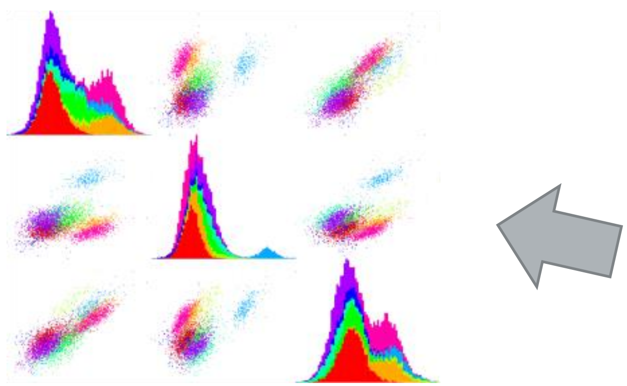


Spike sorting

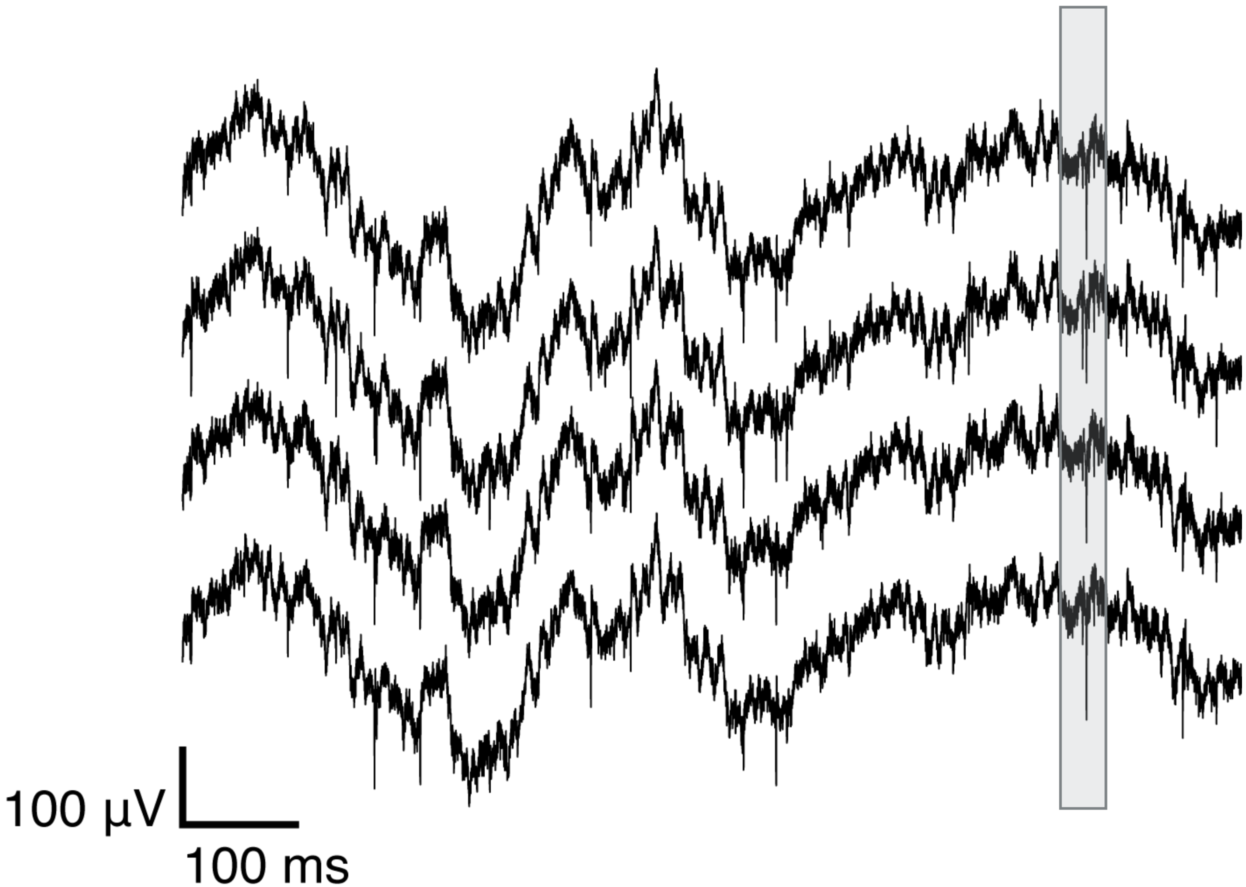
- Raw data
- Spike detection
- Feature extraction
- Clustering
- Verification

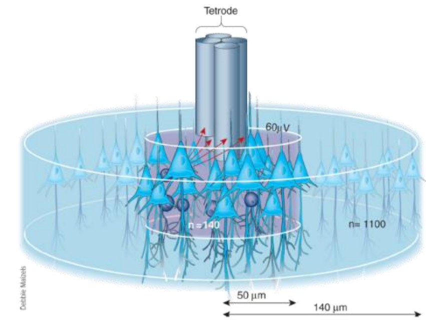






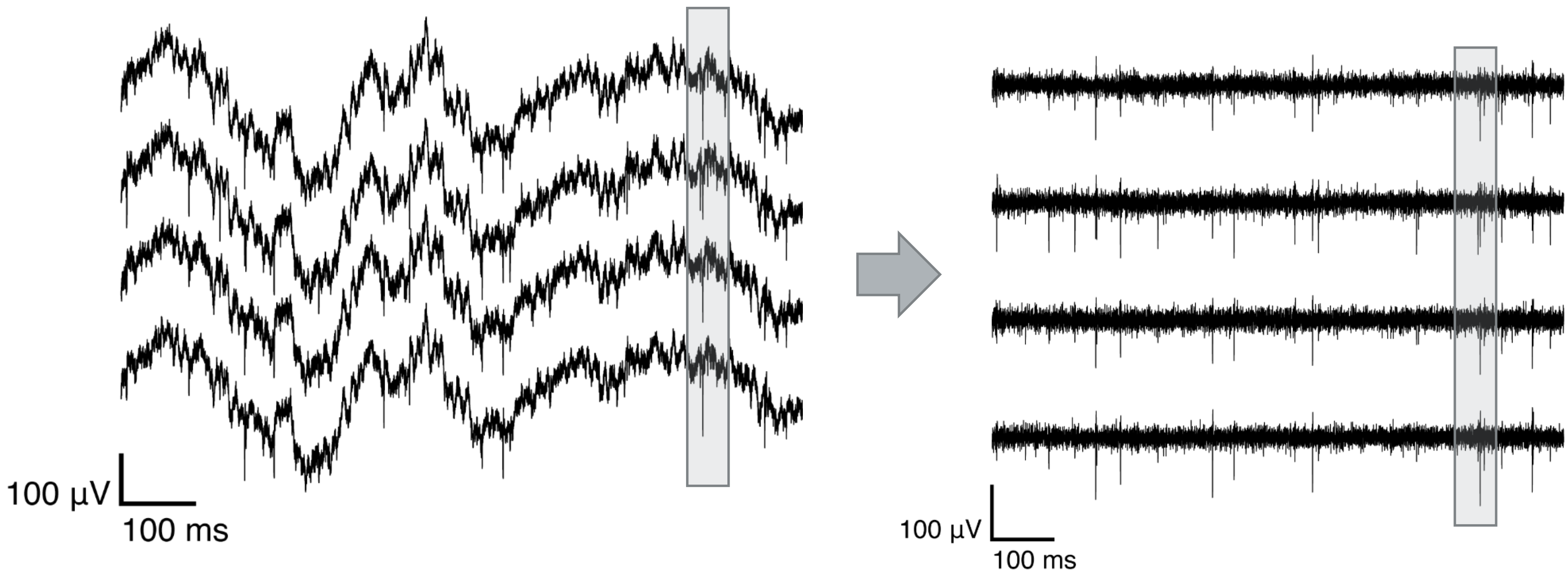
Raw signal





Tetrodes: 4 channels in parallel!

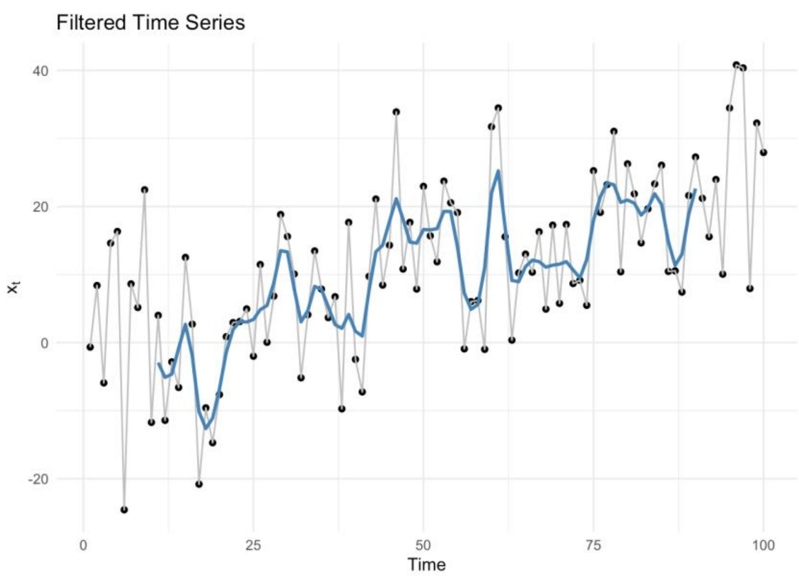
Remove LFP and high frequency noise

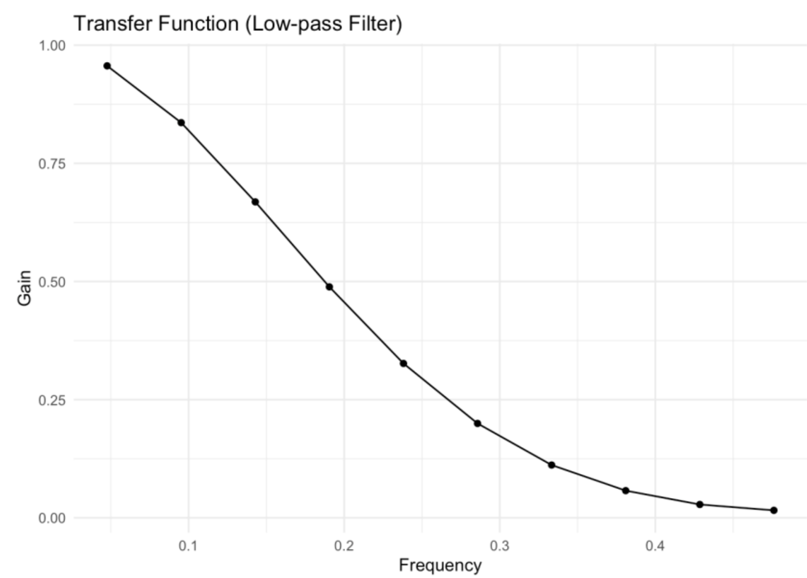


A quick detour on time series filtering

https://bookdown.org/rdpeng/timeseriesbook/filtering-time-series.html

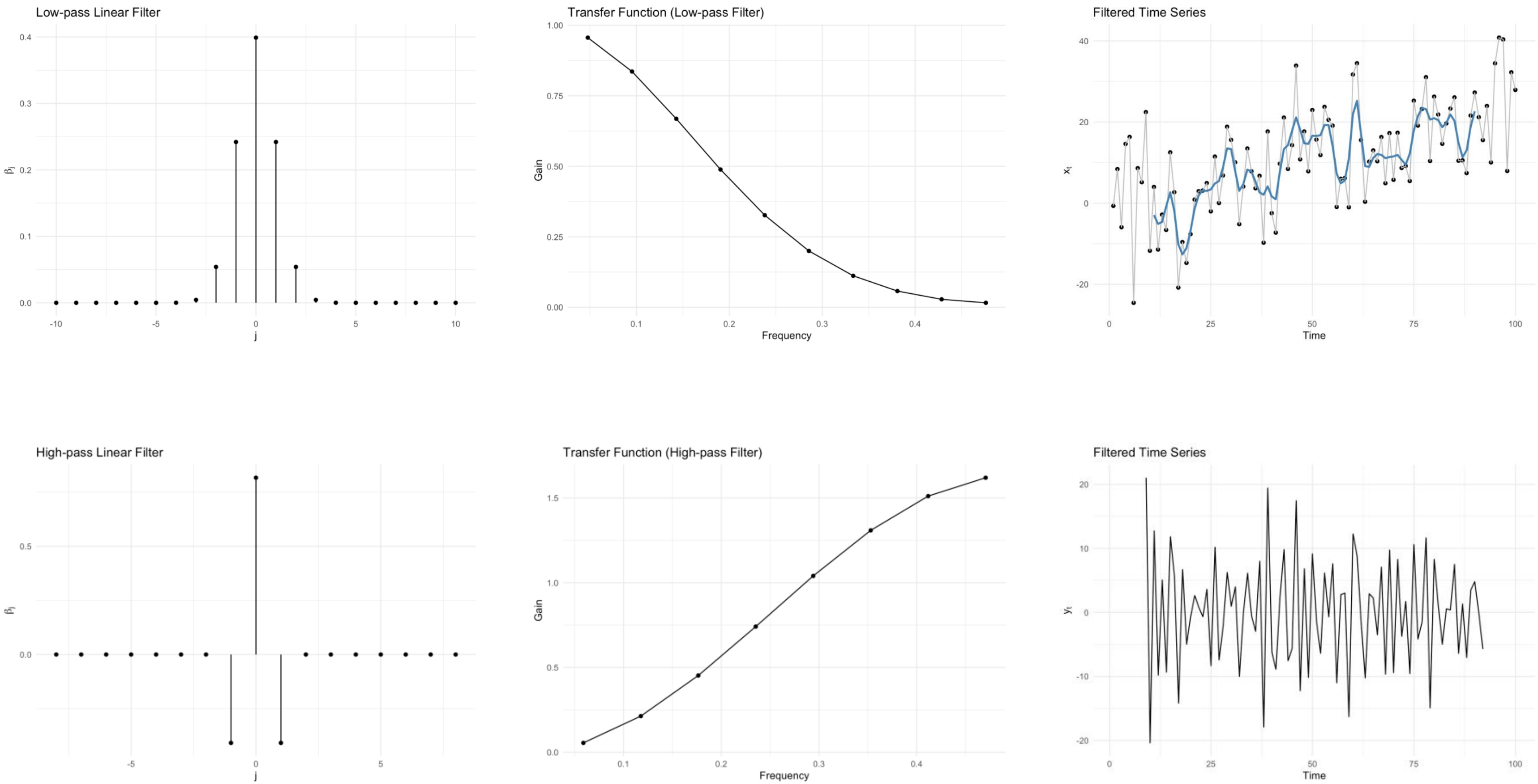
$$y_t = \sum_{j=-\infty}^{\infty} \beta_j x_{t-j}$$



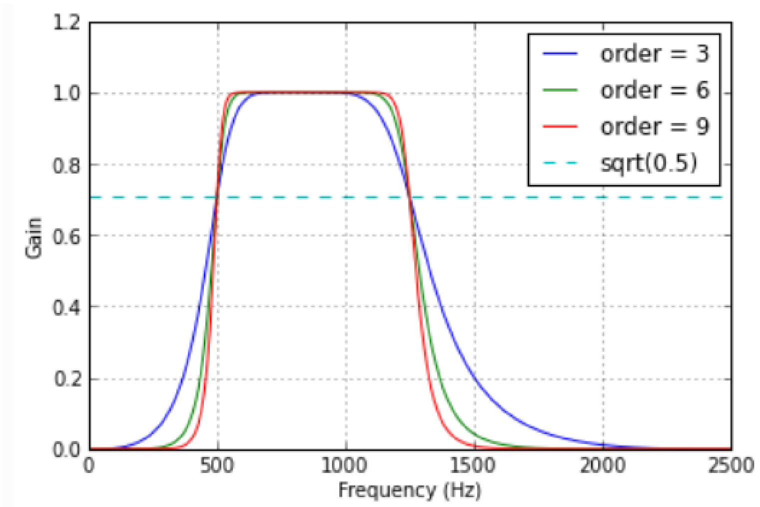


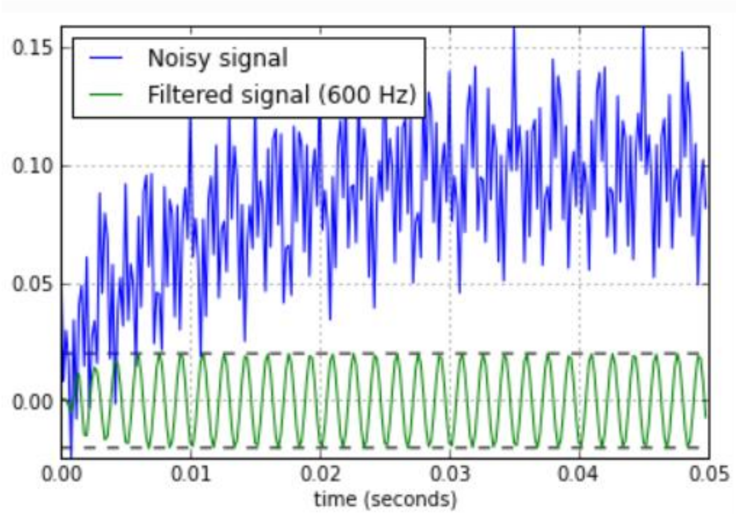
A quick detour on time series filtering

https://bookdown.org/rdpeng/timeseriesbook/filtering-time-series.html



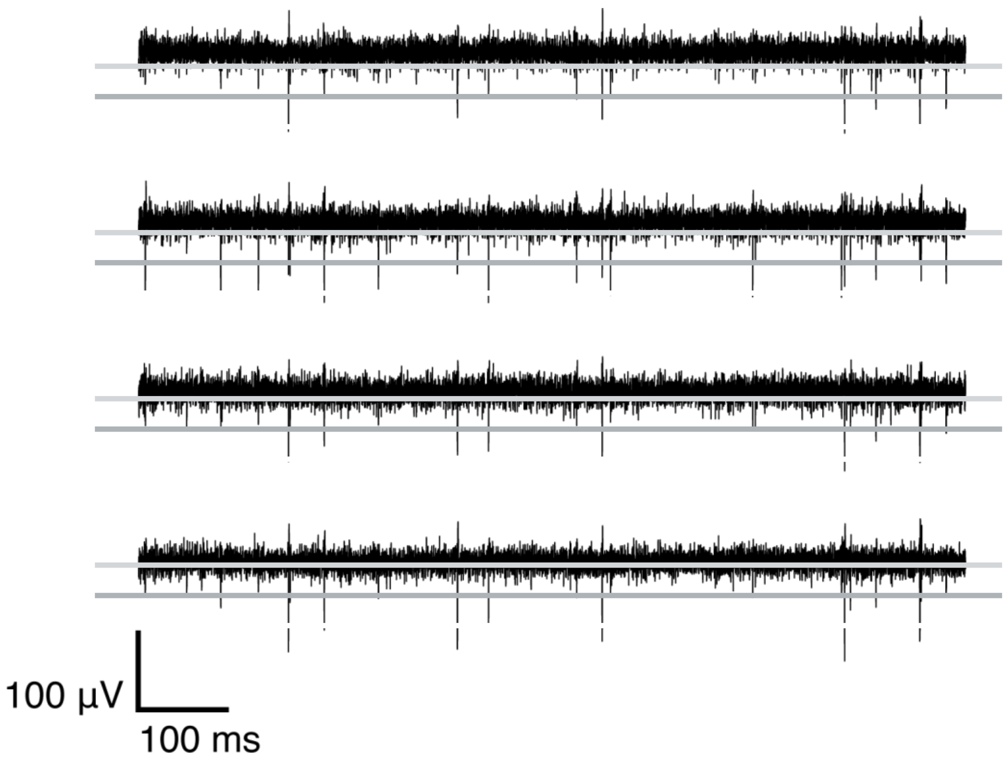
Butterworth bandpass filter





http://scipy-cookbook.readthedocs.io/items/idx_signal_processing.html

Spike detection



Threshold at $\boldsymbol{N\sigma}$

Low threshold

High threshold

Robust estimation of $\boldsymbol{\sigma}$:

$$\hat{\boldsymbol{\sigma}} = \boldsymbol{median}\left(\frac{|x - \overline{x}|}{0.6745}\right)$$

Quian-Quiroga 2004

Why is this an estimate of the standard deviation?

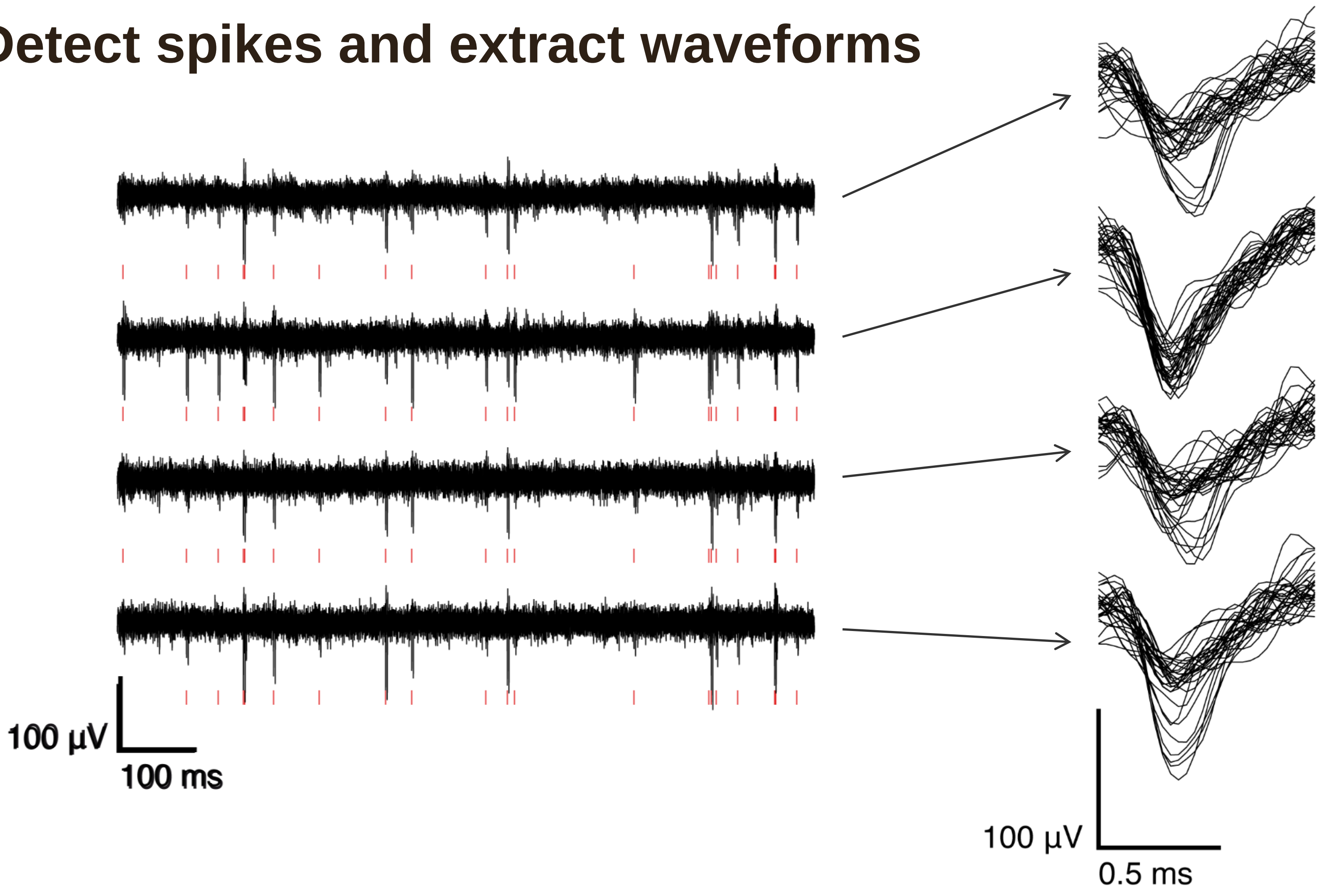
- $MAD = median(|x - \bar{x}|)$

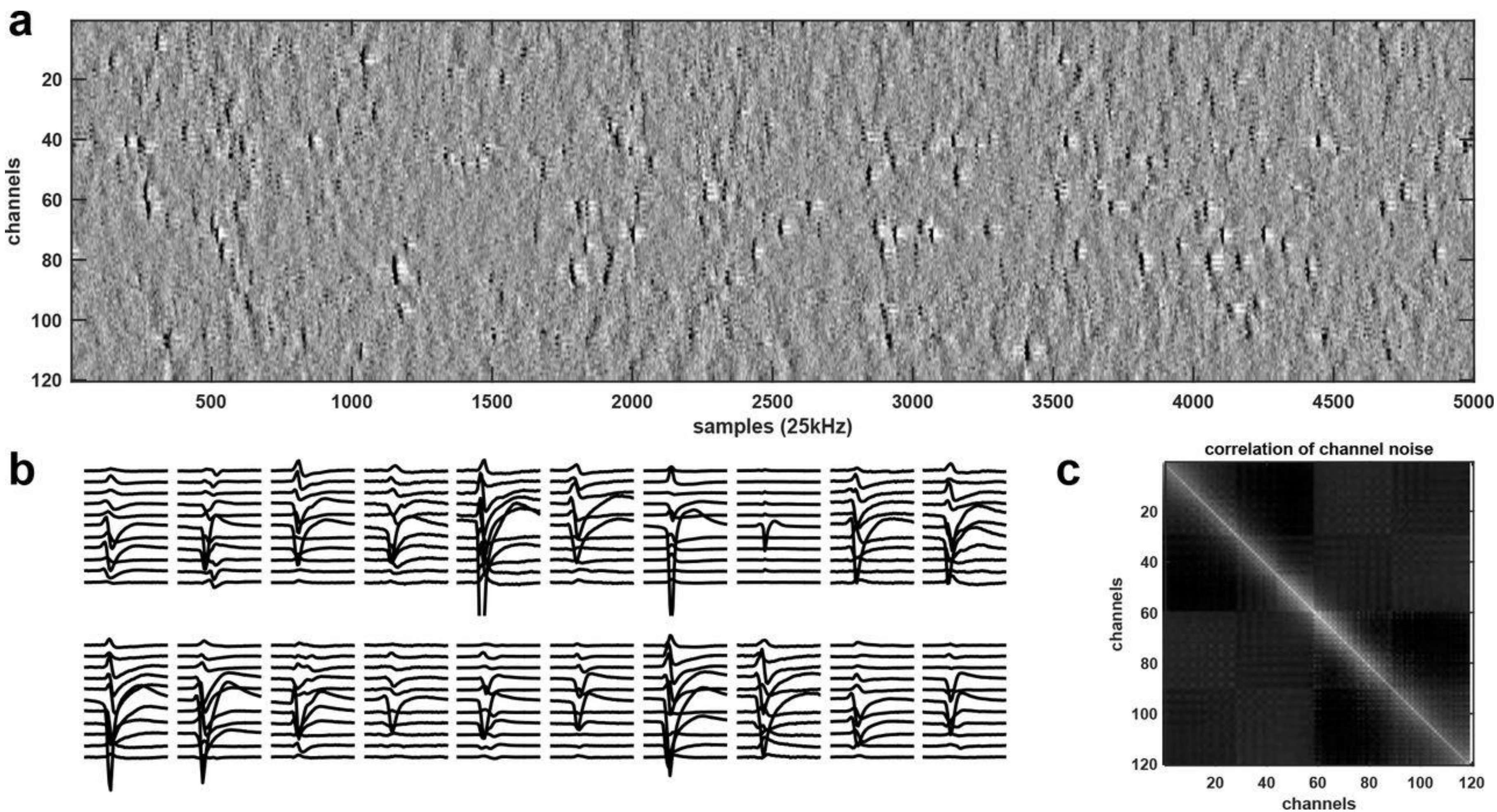
- $\frac{1}{2} = P(|X - \mu| < MAD) = P\left(\frac{|X-\mu|}{\sigma} < \frac{MAD}{\sigma}\right) = P\left(|Z| < \frac{MAD}{\sigma}\right)$

- $\Phi\left(\frac{MAD}{\sigma}\right) - \Phi\left(-\frac{MAD}{\sigma}\right) = \frac{1}{2}$
- $\Phi\left(\frac{MAD}{\sigma}\right) - 1 + \Phi\left(\frac{MAD}{\sigma}\right) = \frac{1}{2}$
- $\Phi\left(\frac{MAD}{\sigma}\right) = \frac{3}{4}$

- $\sigma = \frac{1}{\Phi^{-1}\left(\frac{3}{4}\right)} MAD = \frac{1}{1.4826} MAD = 0.675\, MAD$

Detect spikes and extract waveforms

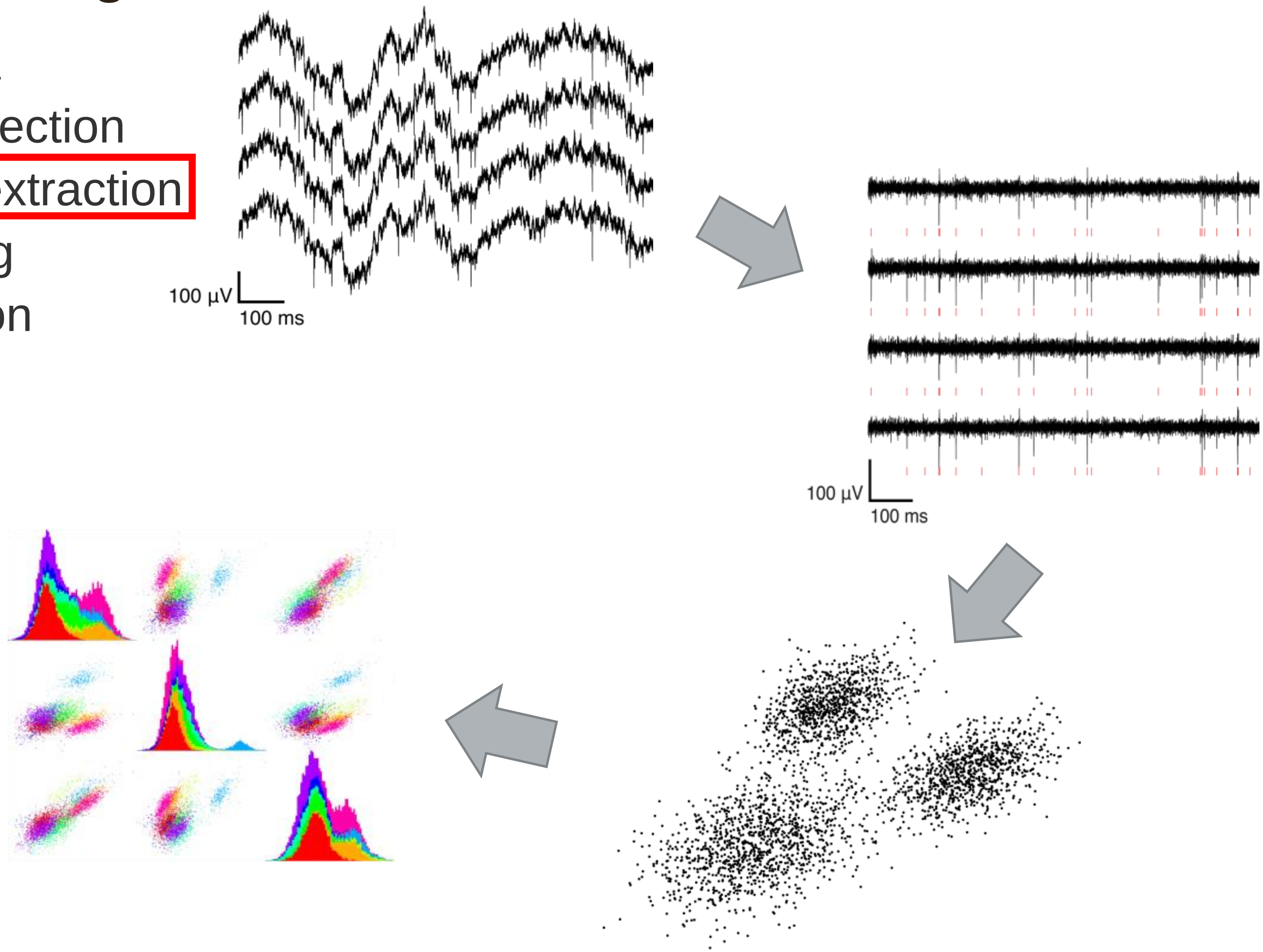




https://www.biorxiv.org/content/10.1101/061481v1.full

Spike sorting

- Raw data
- Spike detection
- Feature extraction
- Clustering
- Verification

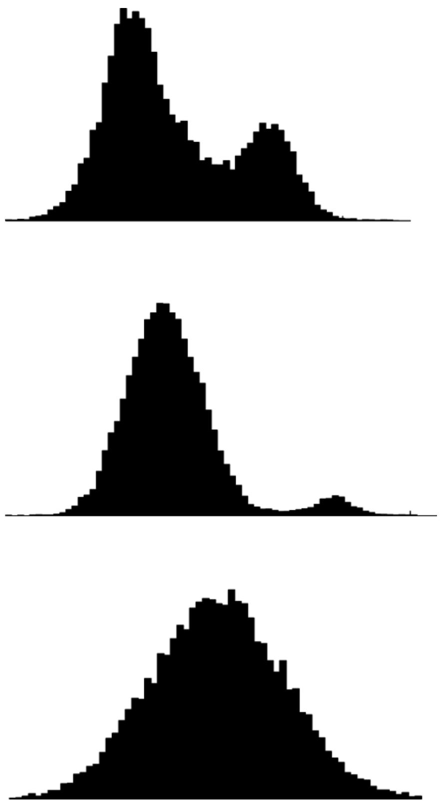


Problem: curse of dimensionality

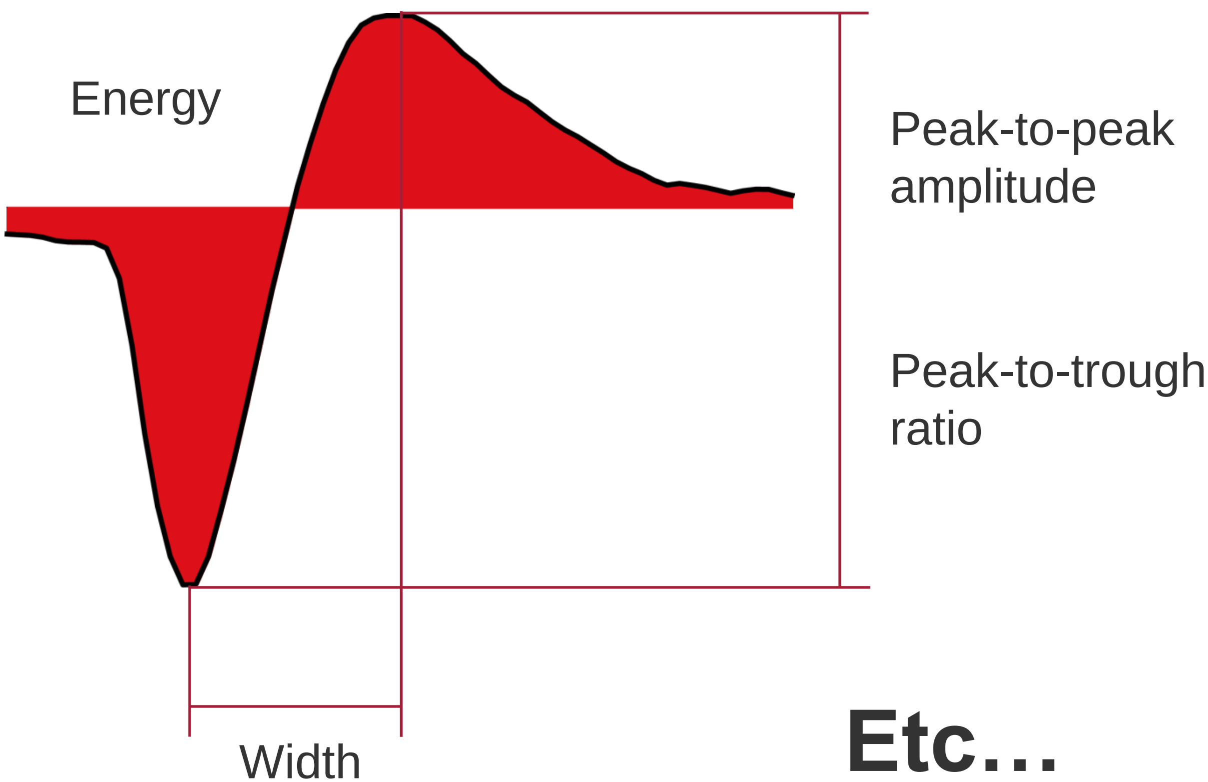
- 4 channels × 32 samples = 128 dimensions
- Problems:
  - Memory issues
  - Computing time
  - Model complexity / number of parameters
- Dimensionality reduction

Which features to choose?

- Goal: small set of features that discriminate between different neurons
- Robust against noise?
- What is good may depend on the spike sorting algorithm (shape of distribution)



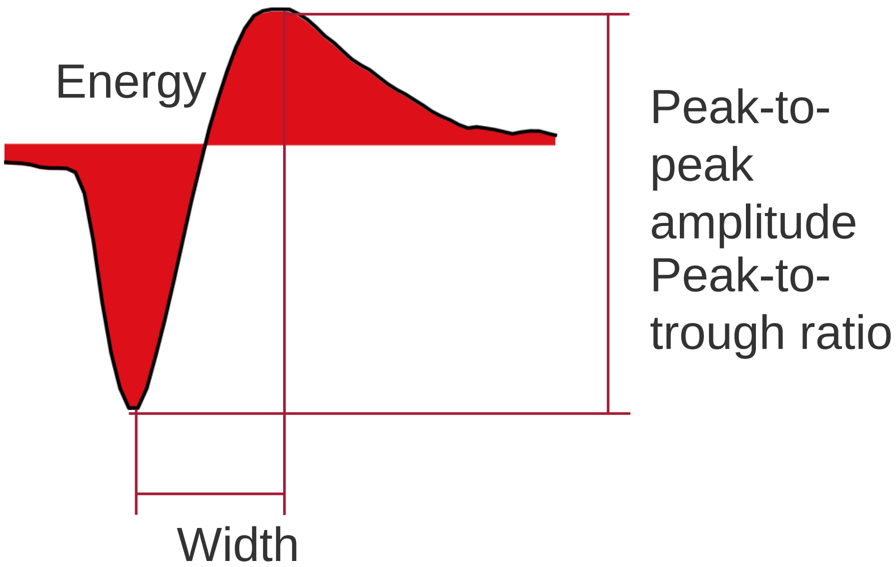
“Descriptive features”

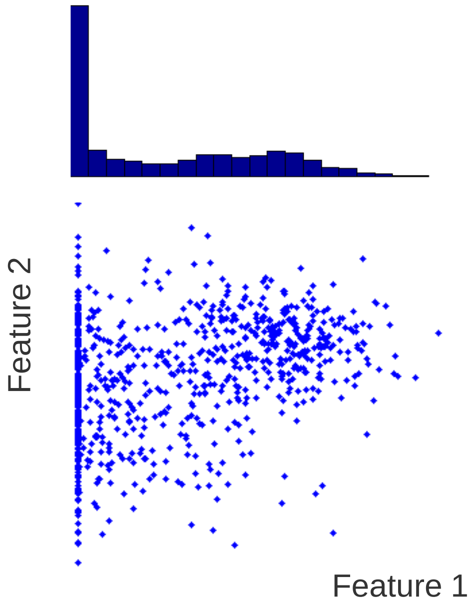


Etc…

Problems

- Sometimes undefined (no trough)
- Non gaussian distributions
- Correlated





See also Quian-Quiroga, 2004

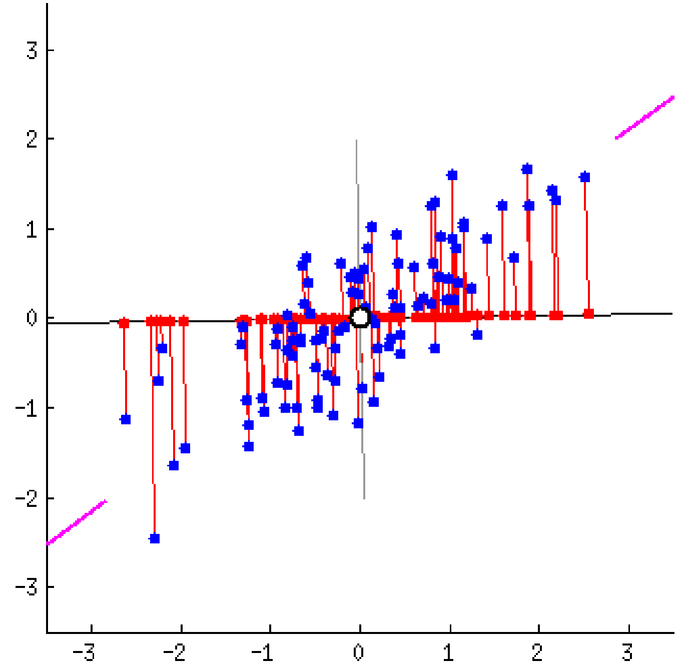
PCA

1. Find direction capturing maximum variance

$$w = \arg\max_{w} \sum_{i} \left(w^T x_i\right)^2, s.t. \left|\left|w\right|\right| = 1$$

2. Minimal reconstruction error

$$w = \arg\min_{w} \left|\left|X - ww^T X\right|\right|^2$$



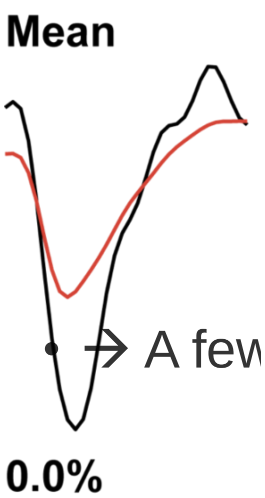
→ solve eigenvalue problem

$$w = eig(cov(X))$$

https://stats.stackexchange.com/questions/2691/making-sense-of-principal-component-analysis-eigenvectors-eigenvalues

Principal component analysis (PCA)

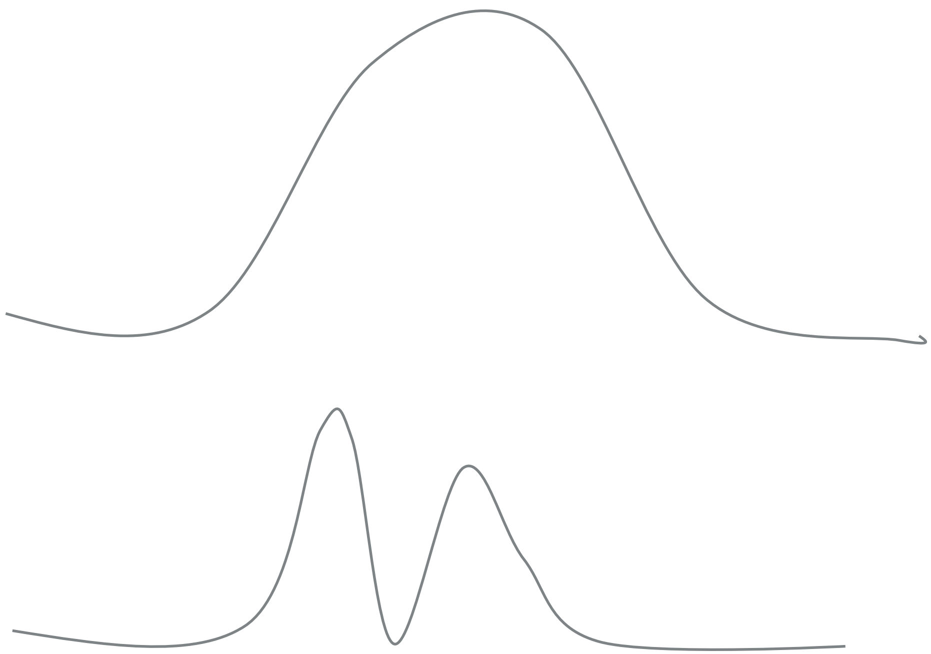
- Finds an orthogonal basis for the data
- First PC is the direction of largest variance



- → A few

Problem with PCA

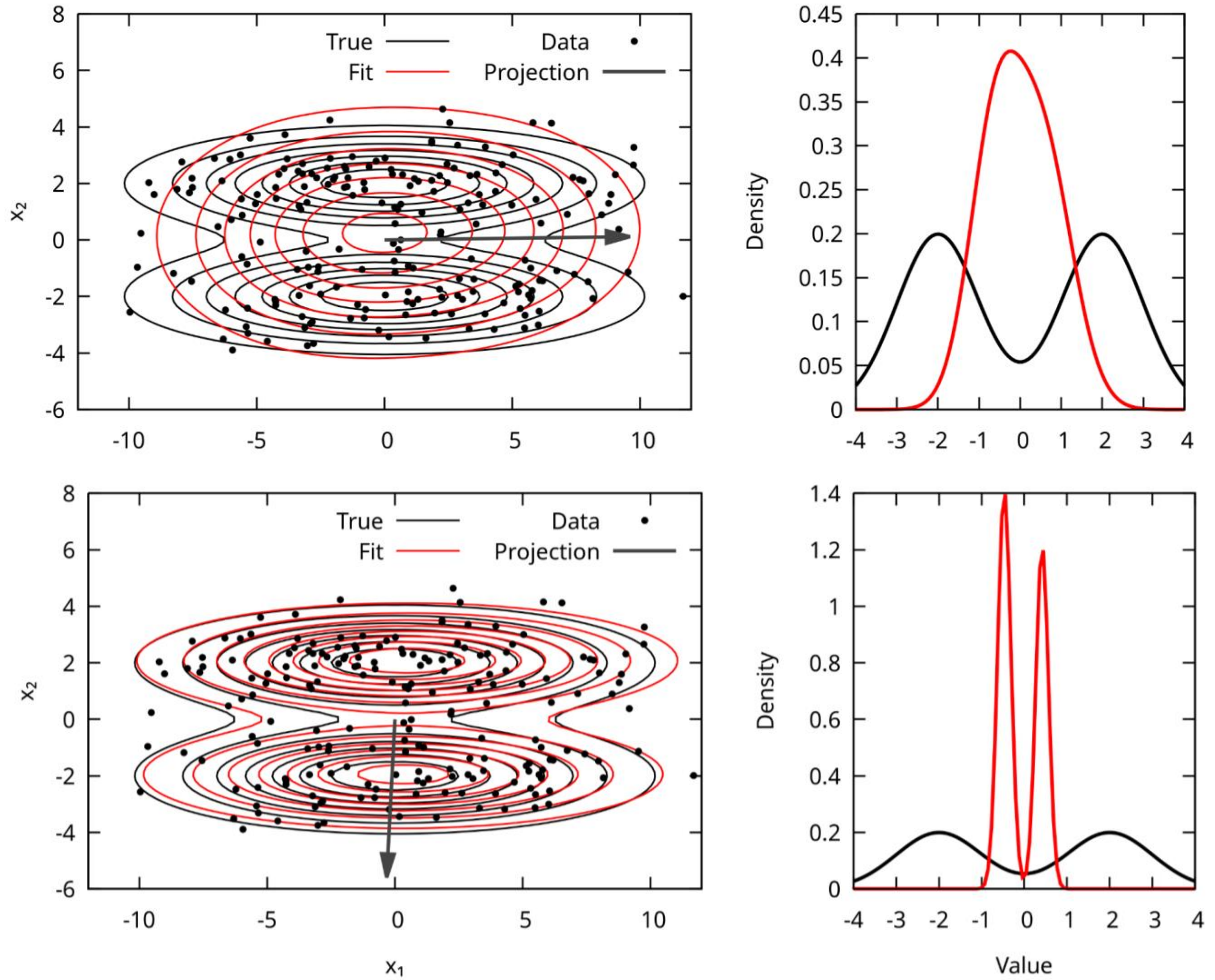
Largest variance – multiple peaks



Larger variance

Better for clustering

Hierarchical mixture of Gaussians

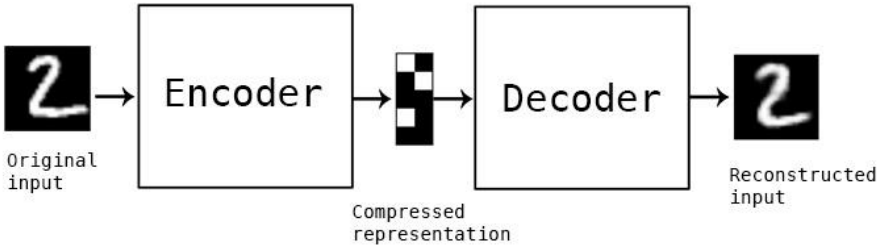


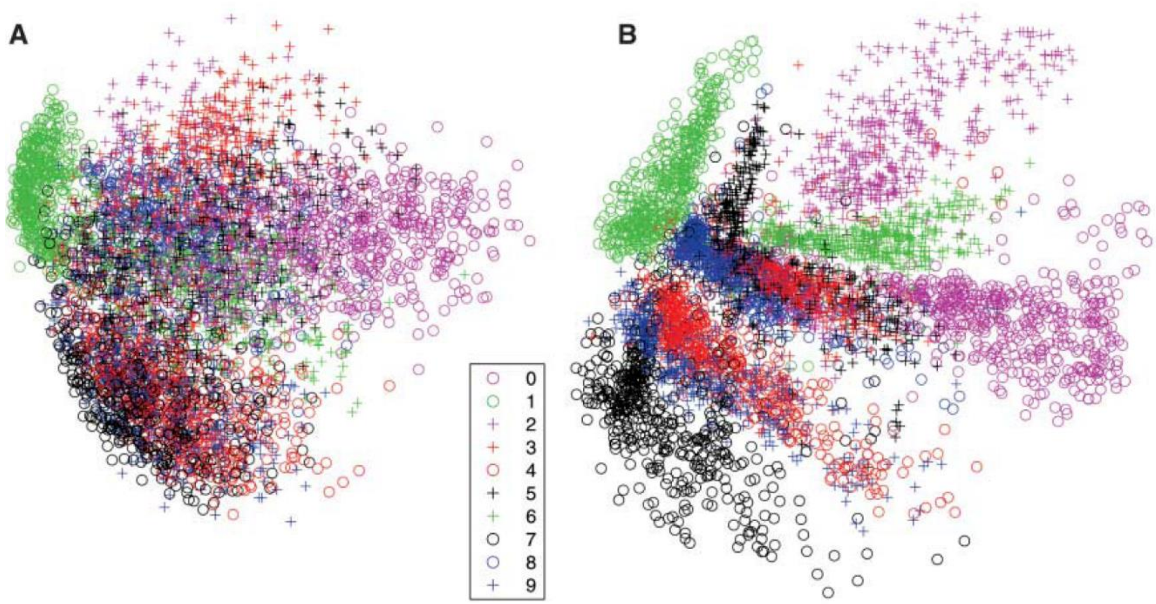
Two-stage PCA + MoG

Joint PCA + MoG

Sokoloski & Berens, arxiv

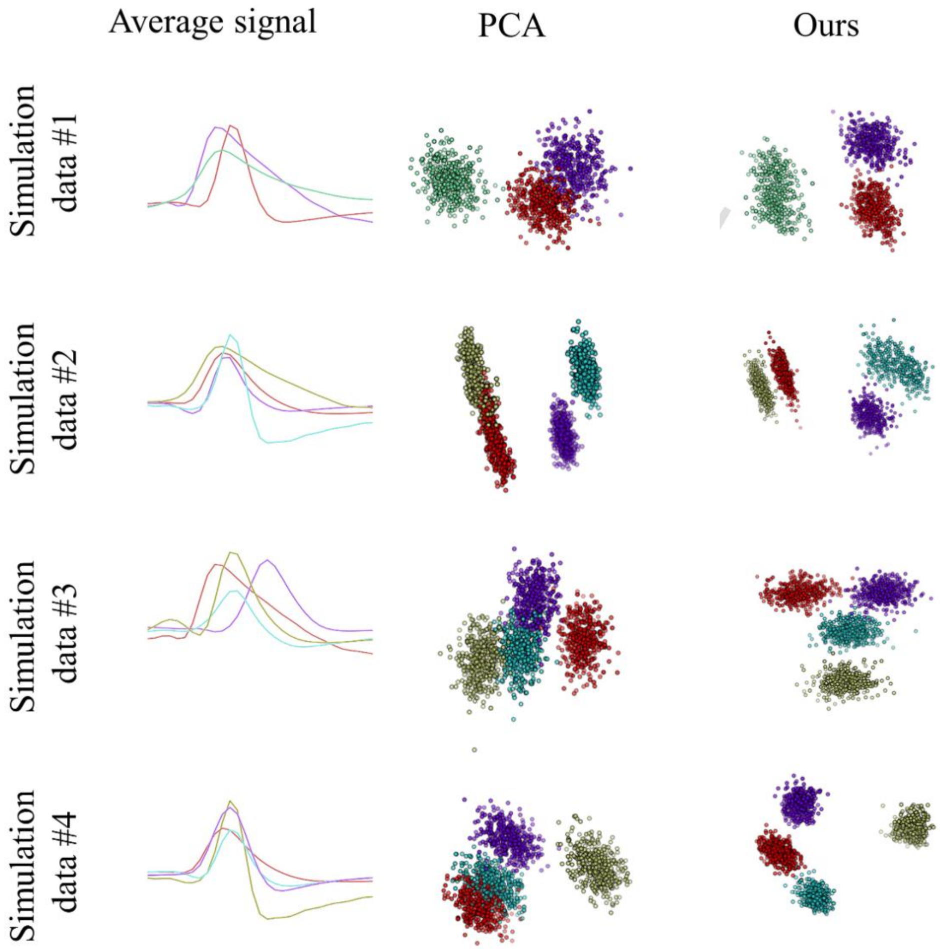
Autoencoder networks

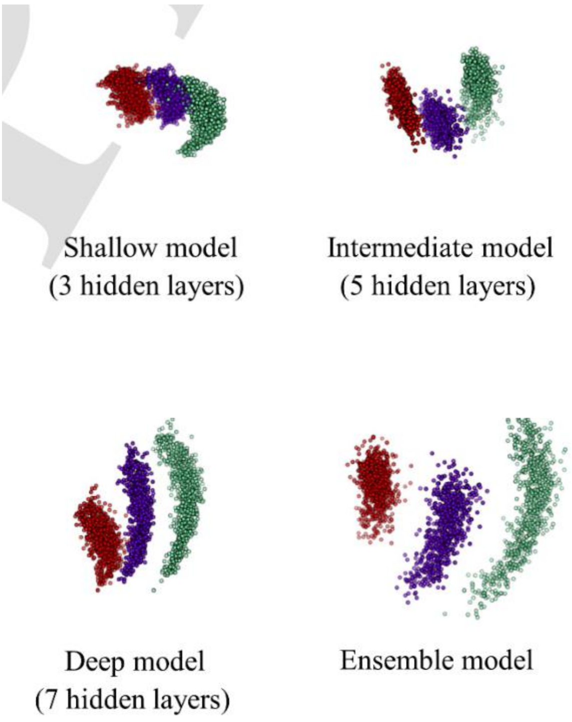




Hinton & Salakhutdinov, 2006

Example

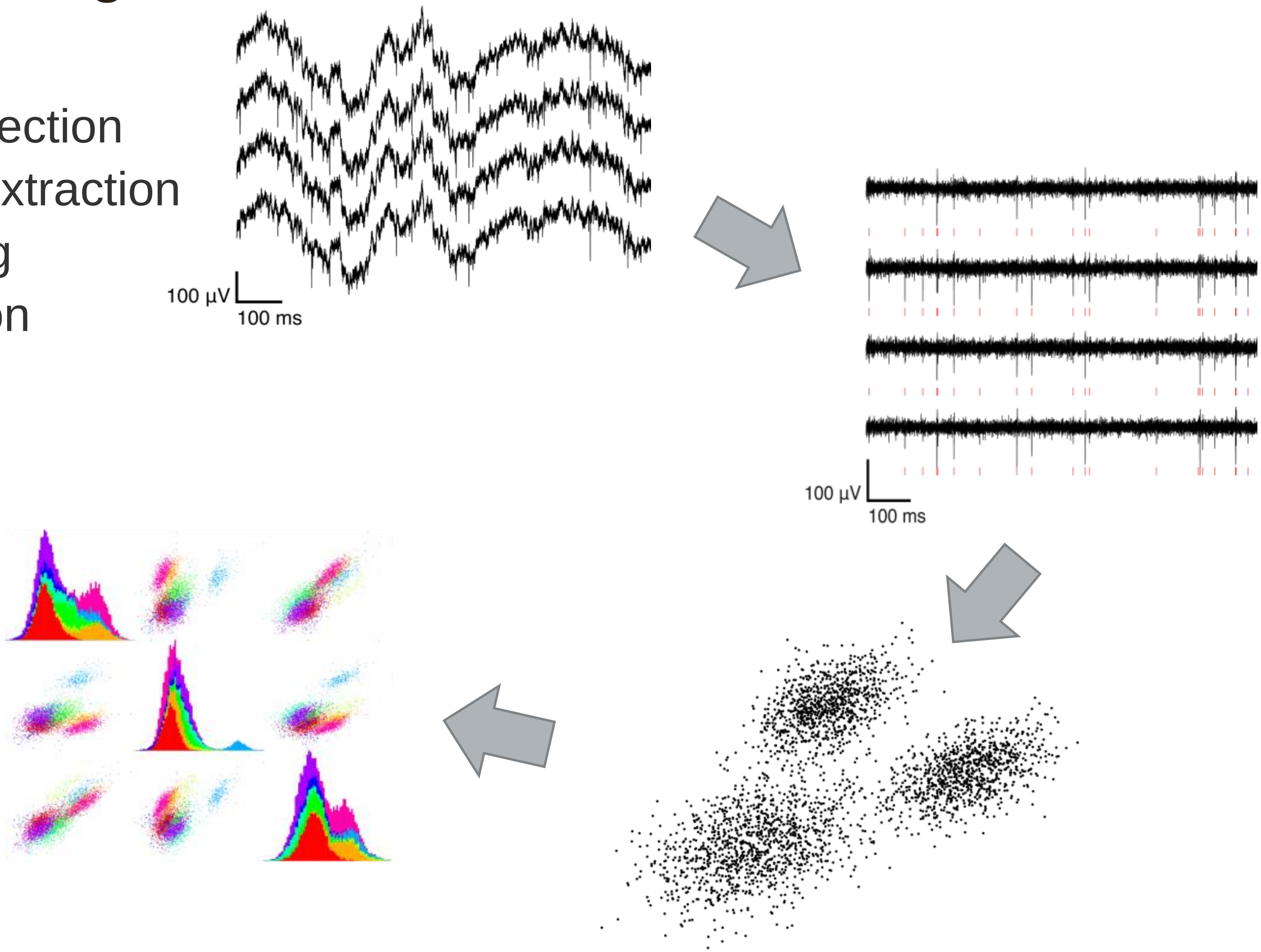




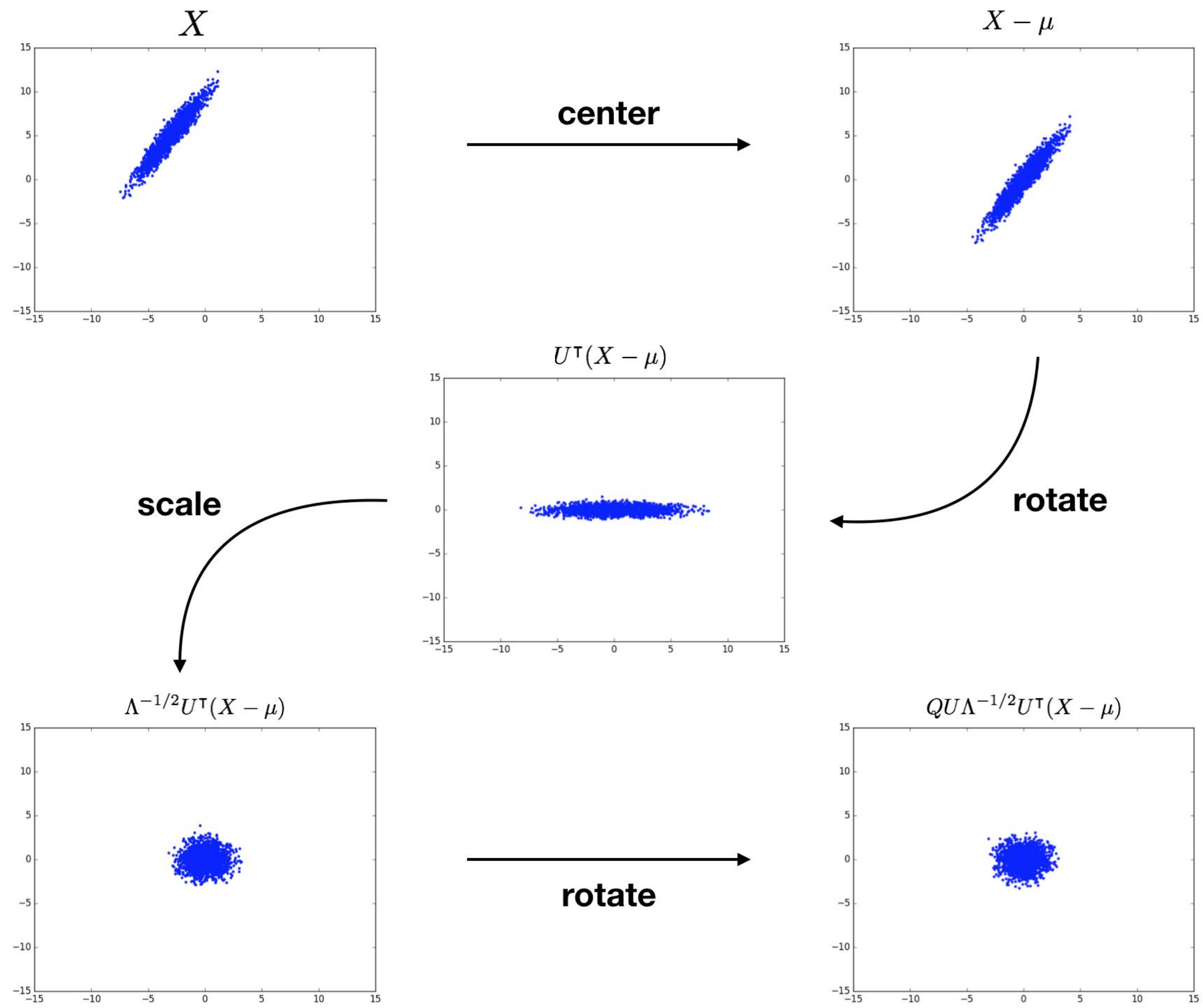
Eom et al. 2020

Spike sorting

- Raw data
- Spike detection
- Feature extraction
- Clustering
- Verification



Whitening



https://joelouismarino.github.io/posts/2017/08/statistical_whitening/