Machine Learning Methods for Neural Data Analysis

Looking back and looking forward

Scott Linderman

STATS 220/320 (NBIO220, CS339N).

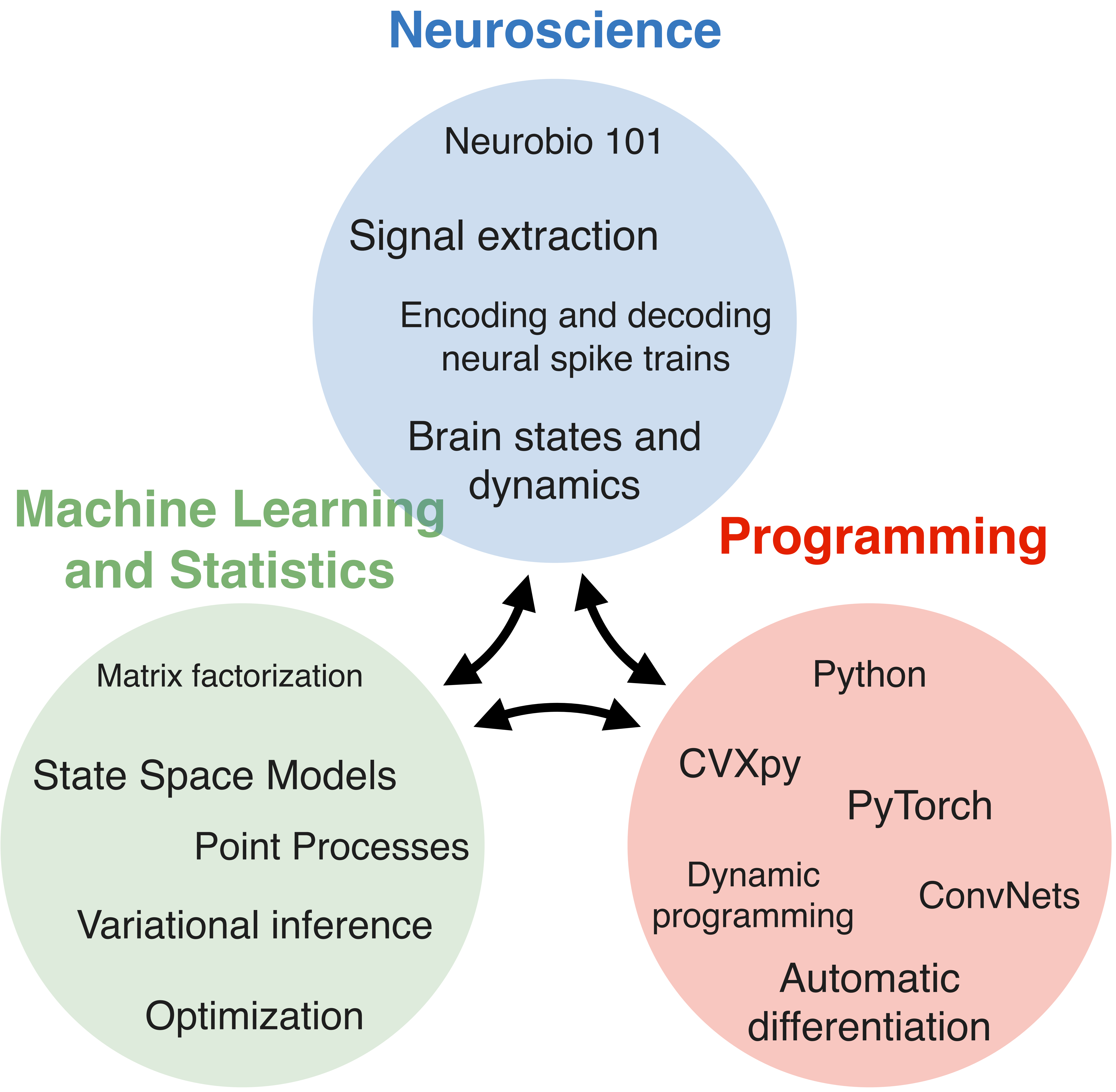
Agenda

- Looking back: What have we covered
- Unit V?

Looking back

Learning Objectives

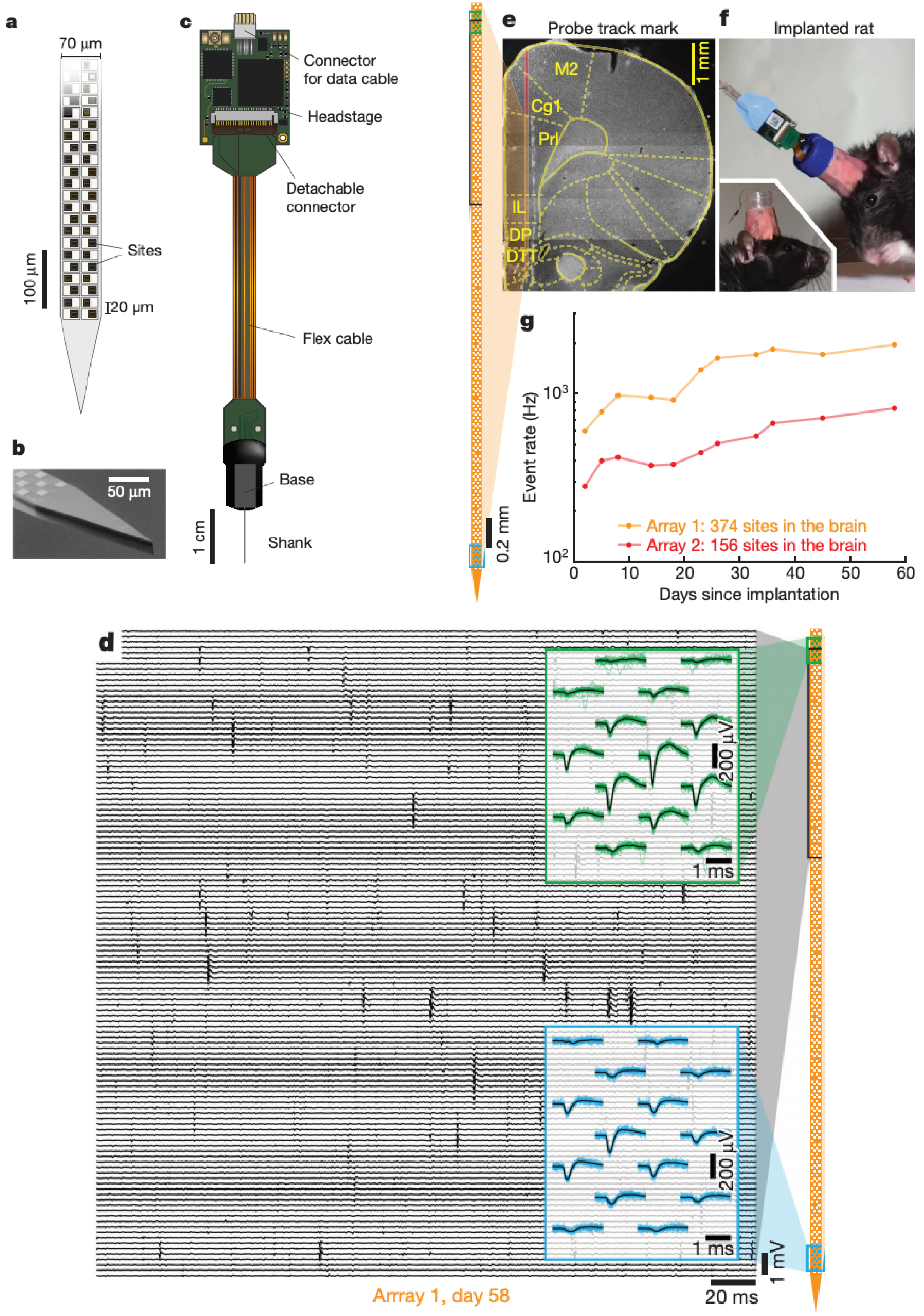
- **Understand** where neural data comes from, what the key analysis problems are, and how state-of-the-art methods work.
- **Develop** probabilistic models for neural data analysis and algorithms to fit those models.
- **Implement** models and algorithms in Python/PyTorch and apply them to real data.
- **Generalize** to new problems and datasets in a course project.



Unit I: Signal Extraction

Spike Sorting

- Modern recording probes like **Neuropixels** measure the electrical activity of **hundreds of cells** across **multiple brain regions** simultaneously.
- When neurons near the probe fire an **action potential**, it registers a **spike in the voltage** on nearby channels.
- Our goal is to **find the spikes** in this time series and **assign a neuron label** based on its waveform.
- **What we learned:** mixture models, matrix factorization, MAP inference, coordinate ascent.

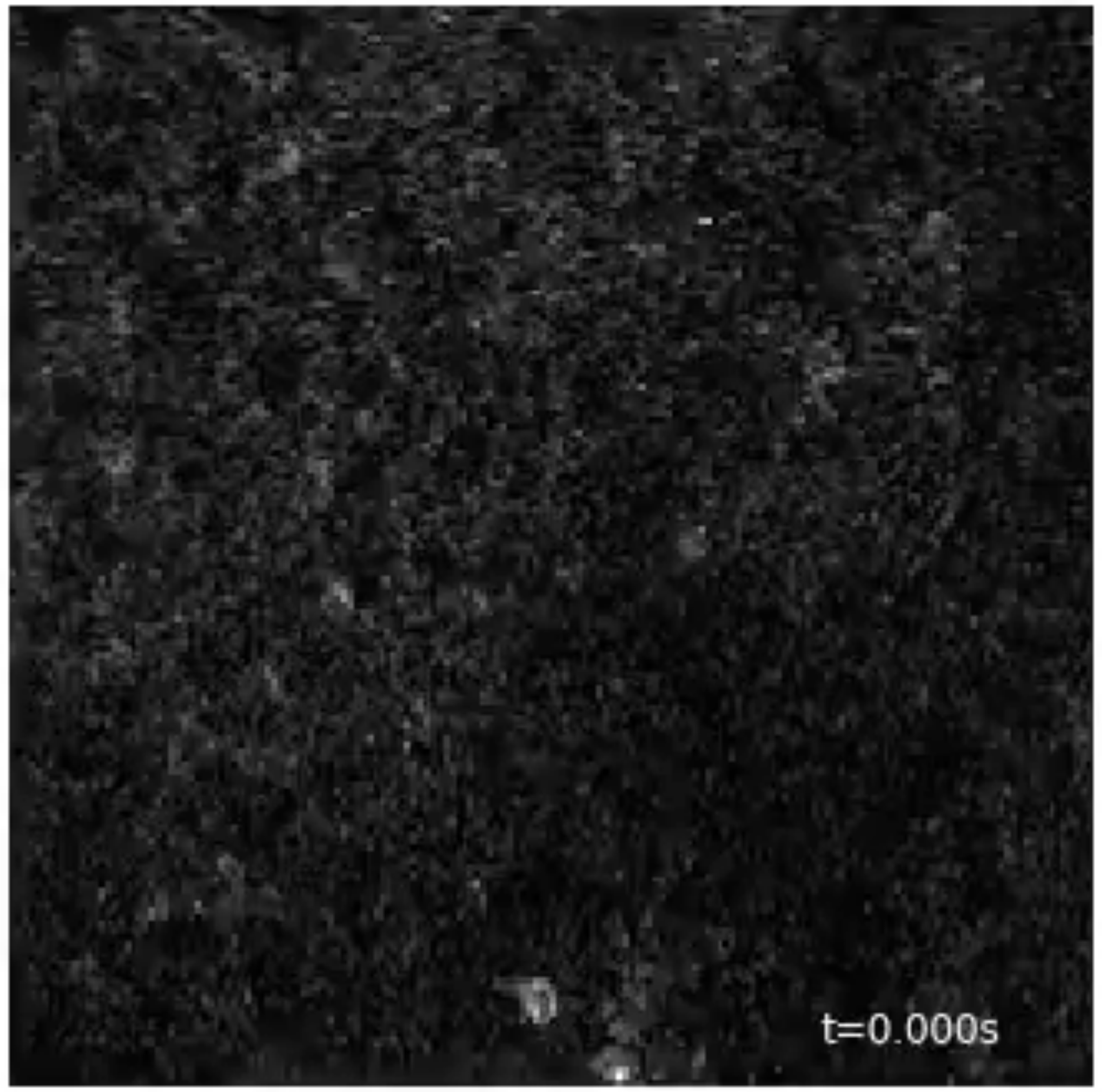


Jun et al, 2017.

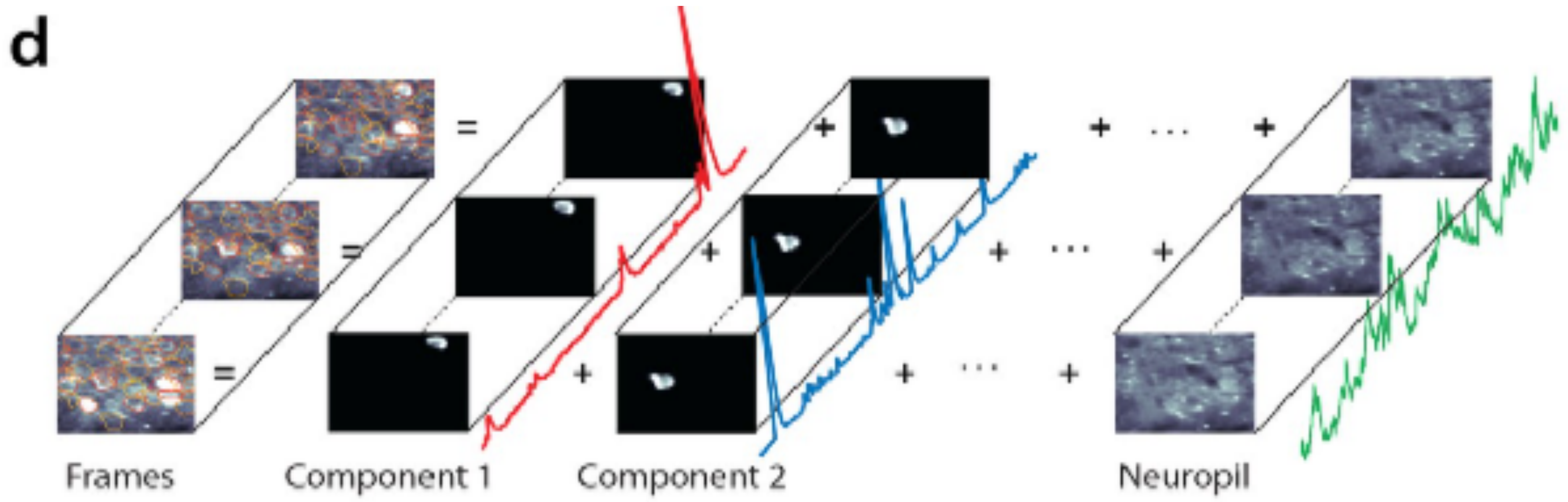
Unit I: Signal Extraction

Demixing calcium imaging data

- When neurons spike, there's a large influx of **calcium ions (Ca^{2+})** into the cell.
- **Genetically encoded calcium indicators** (GECIs) bind to calcium ions, and when light is shone on them they fluoresce.
- Using these indicators, neuroscientists can **optically record** calcium concentrations, a good proxy for neural spiking.
- **Demixing videos to identify cells** and **deconvolving traces to identify spikes** is an area of active research.
- **What we learned:** convolutional matrix factorization, convex optimization, CVXpy



Data from Sue Ann Koay and David Tank

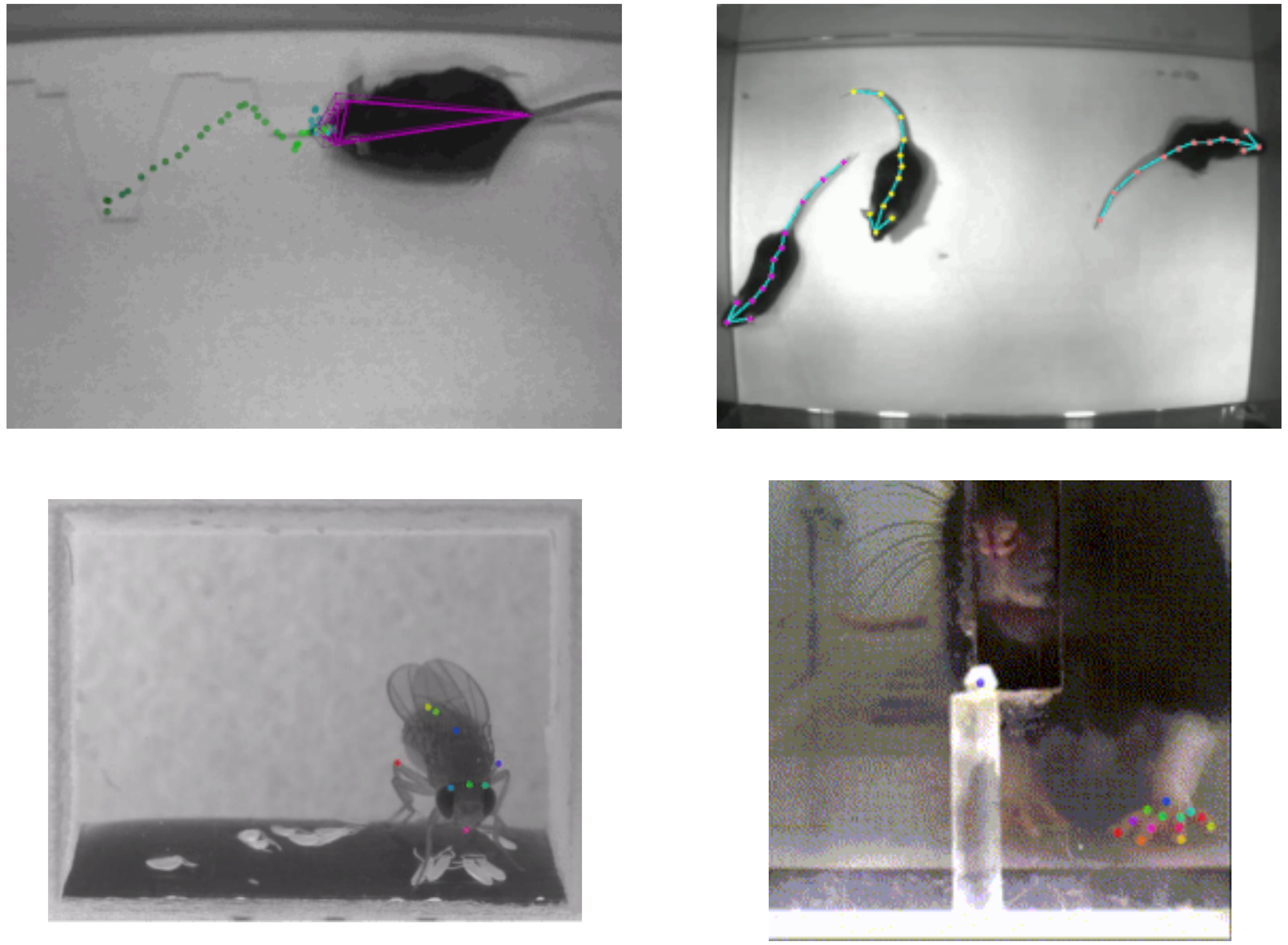


Giovanucci et al (eLife, 2019)

Unit I: Signal Extraction

Markerless pose tracking

- We want to understand how neural activity produces behavior.
- First, we need to **quantify motor outputs**, ideally in unconstrained animals.
- State of the art methods for **markerless pose tracking** use **deep convolutional neural networks (CNNs)** to find keypoints in videos.
- **What we learned:** logistic regression, convolutional neural networks, transfer learning, DataLoaders, torchvision.models.

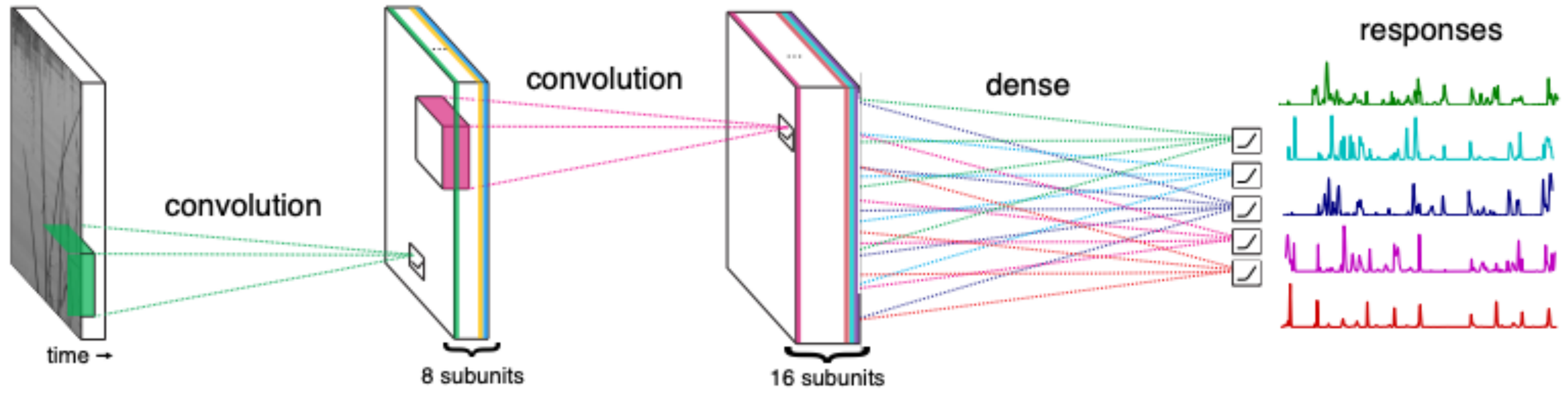


Mathis et al (Nat. Neuro., 2018)
https://github.com/DeepLabCut/DeepLabCut

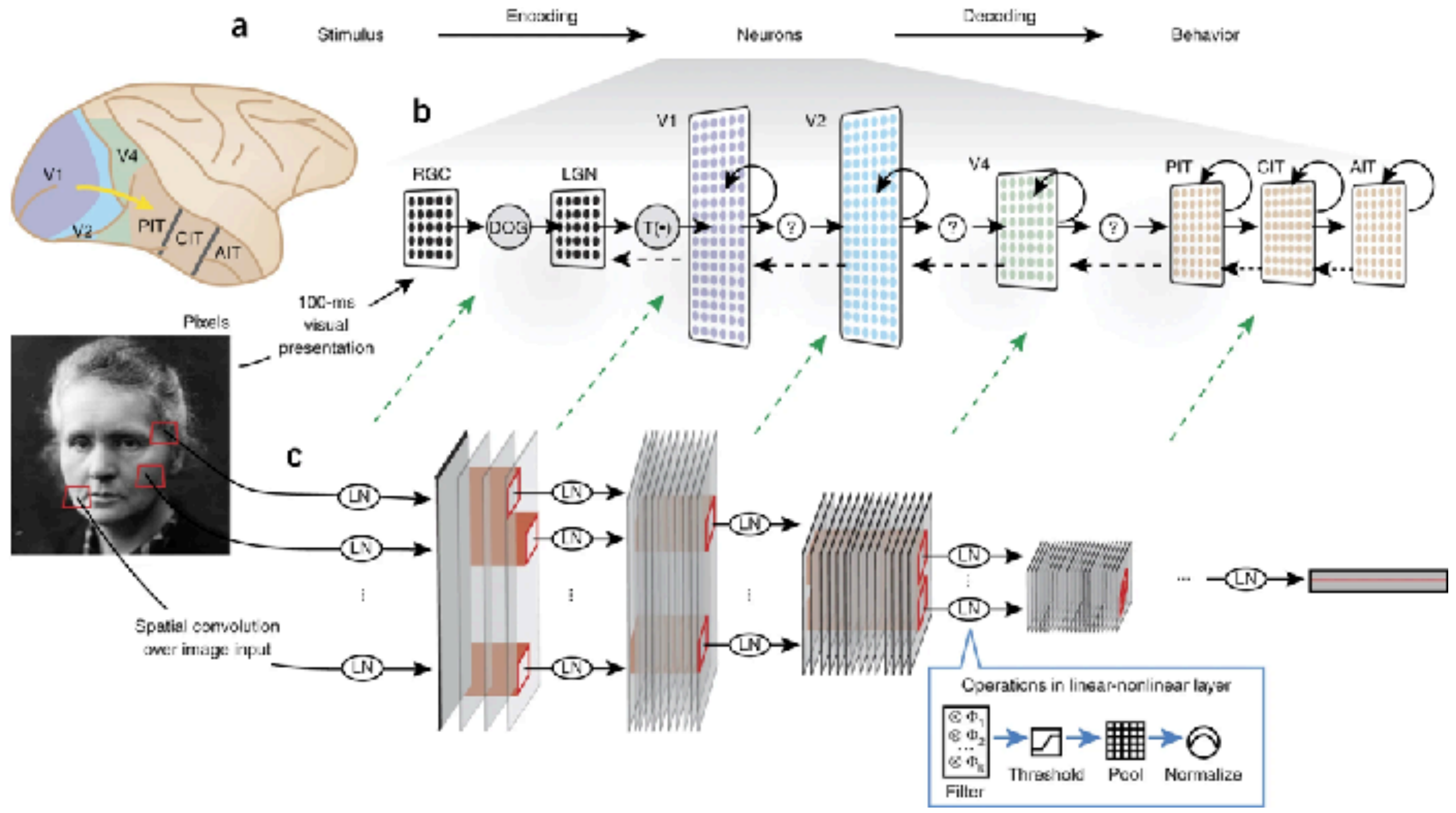
Unit II: Encoding and Decoding Neural Spike Trains

Predicting neural responses to images

- CNNs aren't just useful for signal extraction, they're also our best models for how the **visual system encodes sensory inputs.**
- Of course, we see a constantly changing visual scene. We'll build models that **take in movies** and **output neural firing rates.**
- Neural spikes are modeled as draws from a **Poisson process** with these firing rates.
- **What we learned:** generalized linear models, Poisson processes, random graph models, more CNNs.



McIntosh et al (NeurIPS 2016)

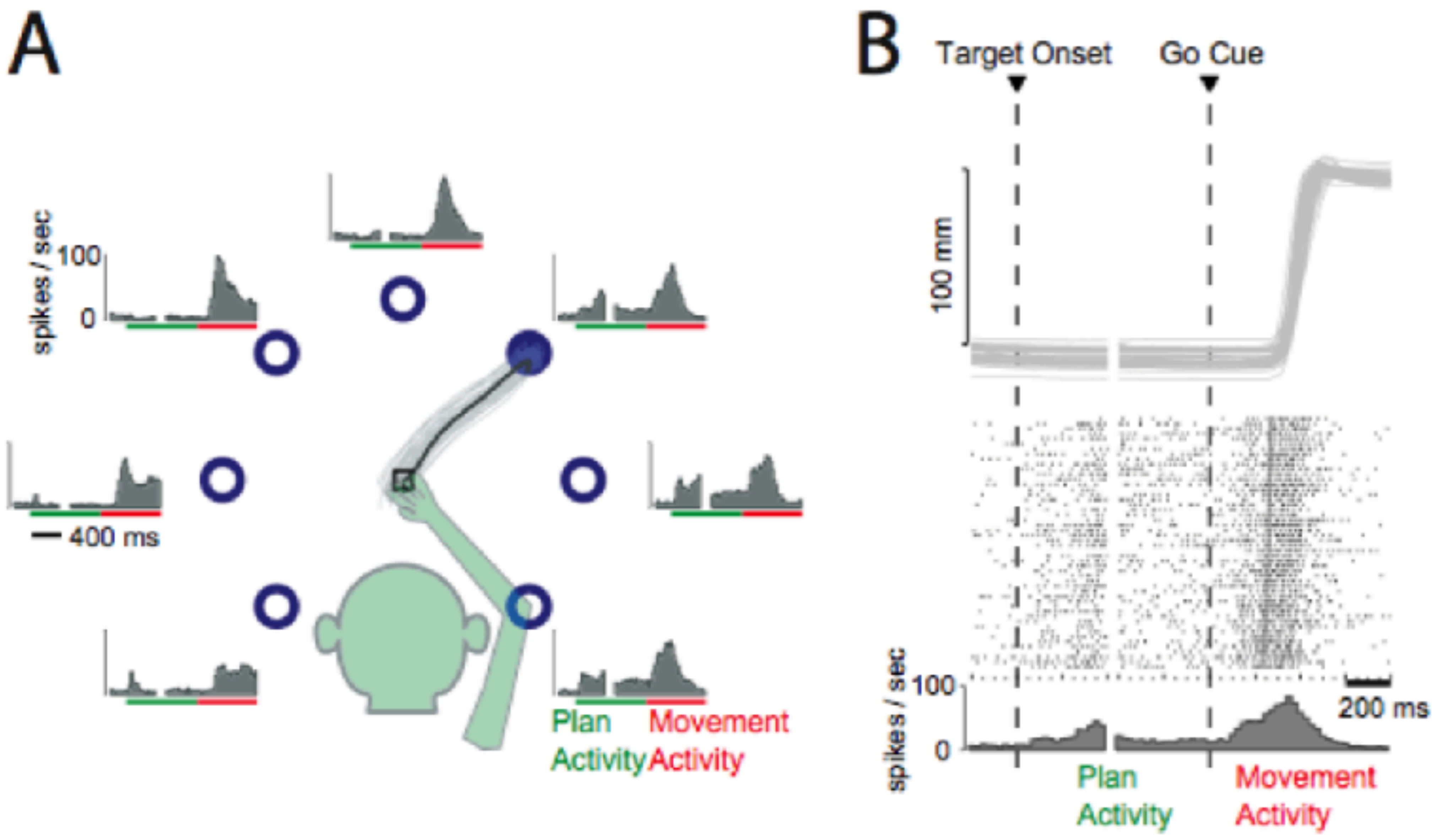


Yamins and DiCarlo (Nat. Neuro. 2016)

Unit II: Encoding and Decoding Neural Spike Trains

Decoding arm movements from neural data

- We also want to understand how to **decode motor outputs from neural activity**.
- This is a central challenge in **building neural prostheses**.
- Neurons in motor cortex, in particular, fire at different rates for different movements.
- We can leverage these differences to **infer movements** from **neural data.**
- **What we learned:** Bayesian decoders, linear dynamical systems, natural and mean parameters of the Gaussian distribution.

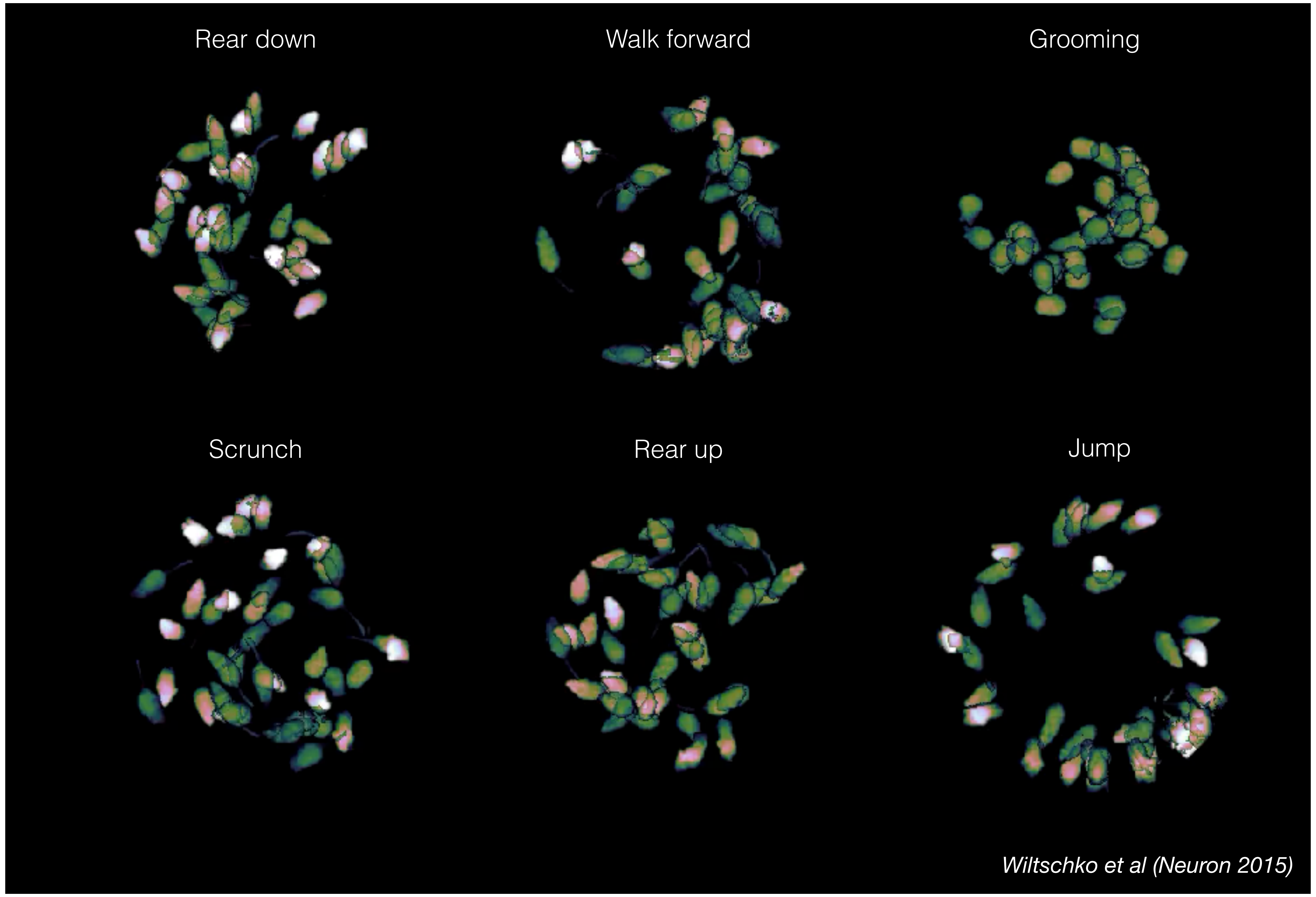


Prof. Krishna Shenoy, EE124

Unit III: Latent variable models of neural and behavioral data

Summarizing behavior with movement “syllables”

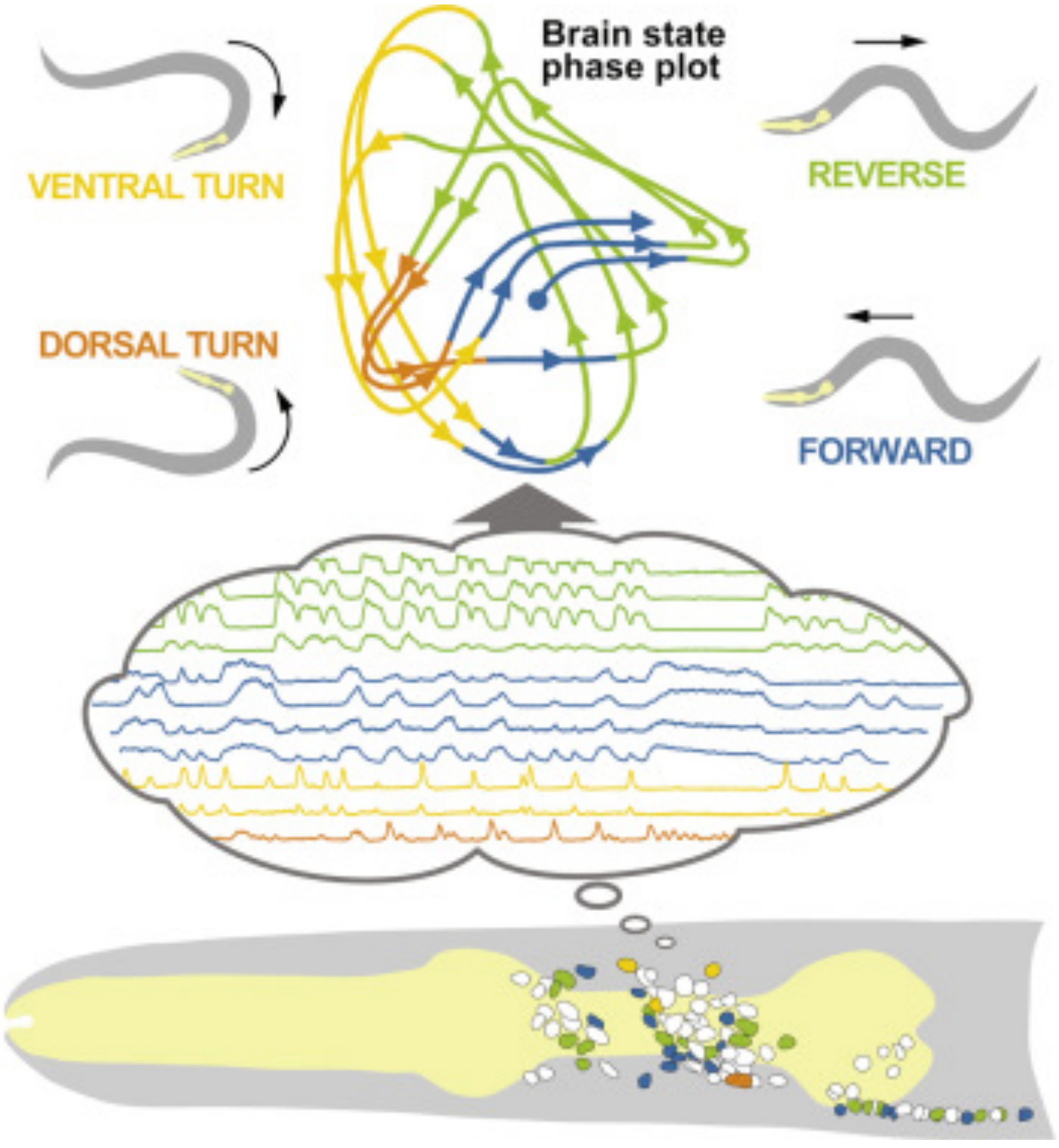
- We can learn a lot about the brain by understanding the **structure of its outputs**.
- Recently, there’s been a “**call to action**” to better characterize animal behavior. Krakauer et al (Neuron, 2017); Datta et al. (Neuron, 2019)
- **Latent variable models** offer a compelling means of **summarizing behavior** in terms of **hidden states**, or “syllables,” of movement.
- We’ll build **autoregressive hidden Markov models** to extract such syllables from video data.
- **What we learned:** expectation-maximization (EM), hidden Markov models, sufficient statistics



Unit III: Latent variable models of neural and behavioral data

Discovering dynamical states in whole-brain recordings

- A remarkable property of brain activity is that it is often **lower dimensional** than the sheer number of neurons.
- Moreover, the **dynamics** within this low dimensional space are often **indicative of** the animal's **behavior**.
- We will study **state space models** for characterizing these low dimensional dynamics.
- **What we learned:** factor analysis, switching LDS, variational EM, SSM



Kato et al (Cell, 2015)

Unit IV: Current research topics

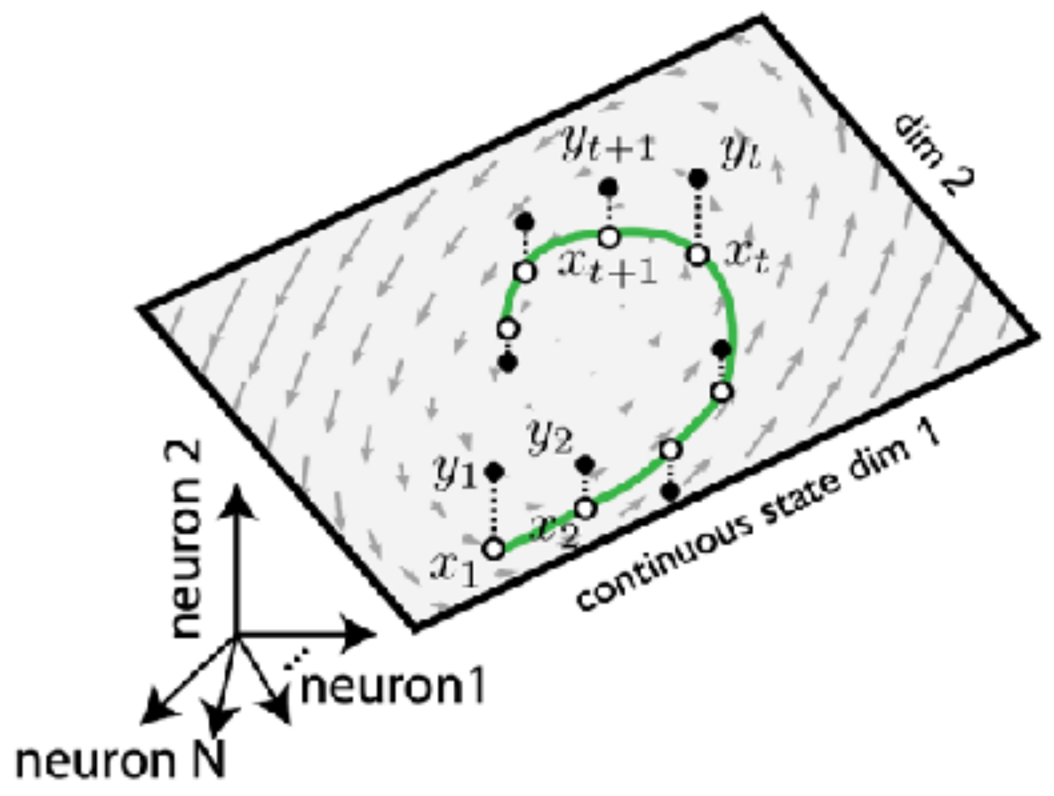
State space models and inference algorithms

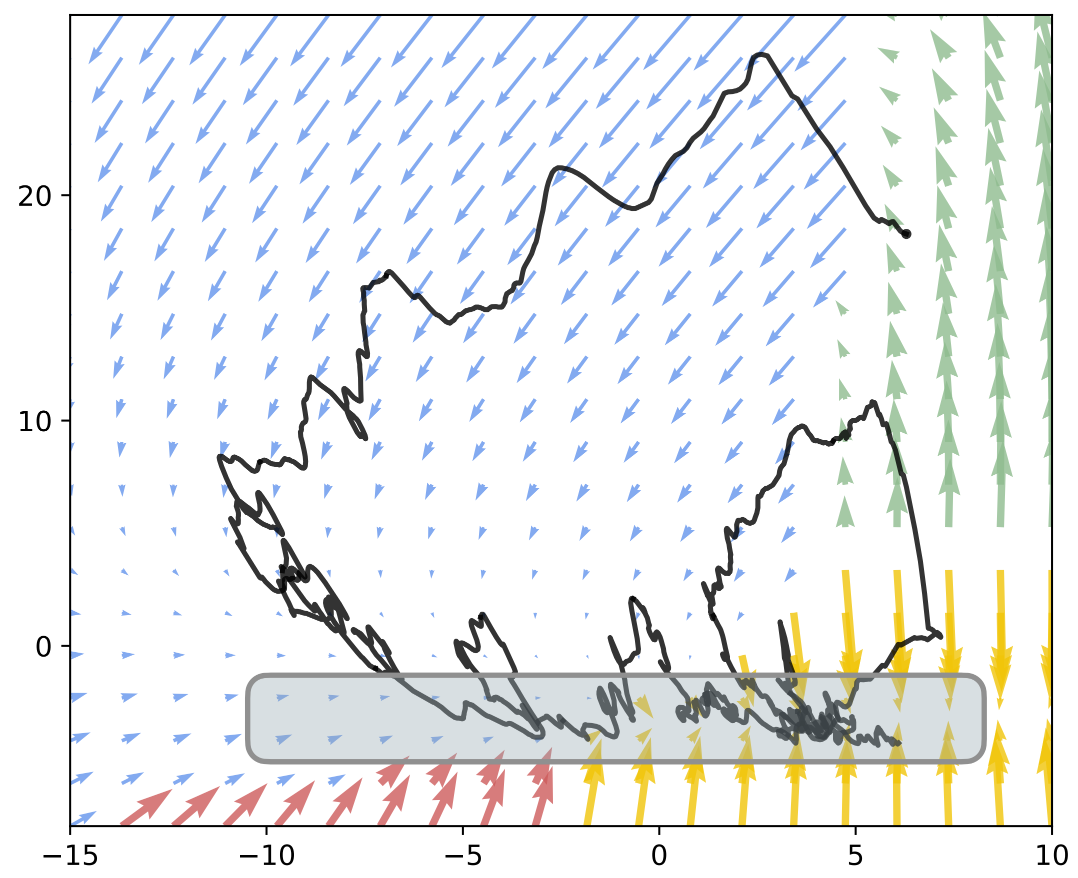
- Designing more flexible and interpretable models for neural and behavioral time series remains an important area of research.

- In parallel, we are working on improved algorithms for inferring states and estimating parameters of these models too.

- As datasets grow, our models must as well. At some point, these so-called foundation models could unlock new possibilities for understanding neural computation.

- **What we learned:** Gaussian processes, SDEs, VAEs, foundation models.

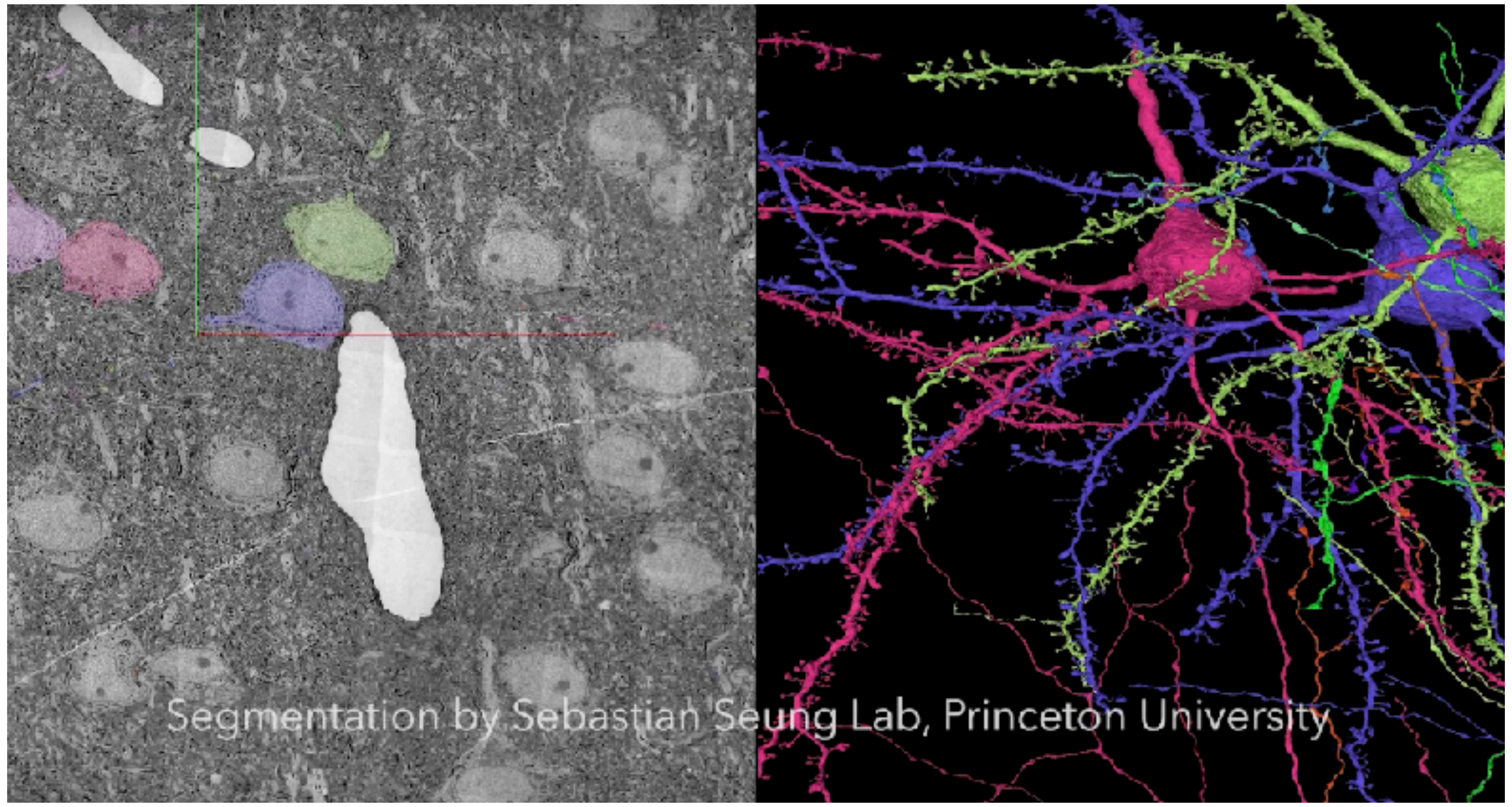


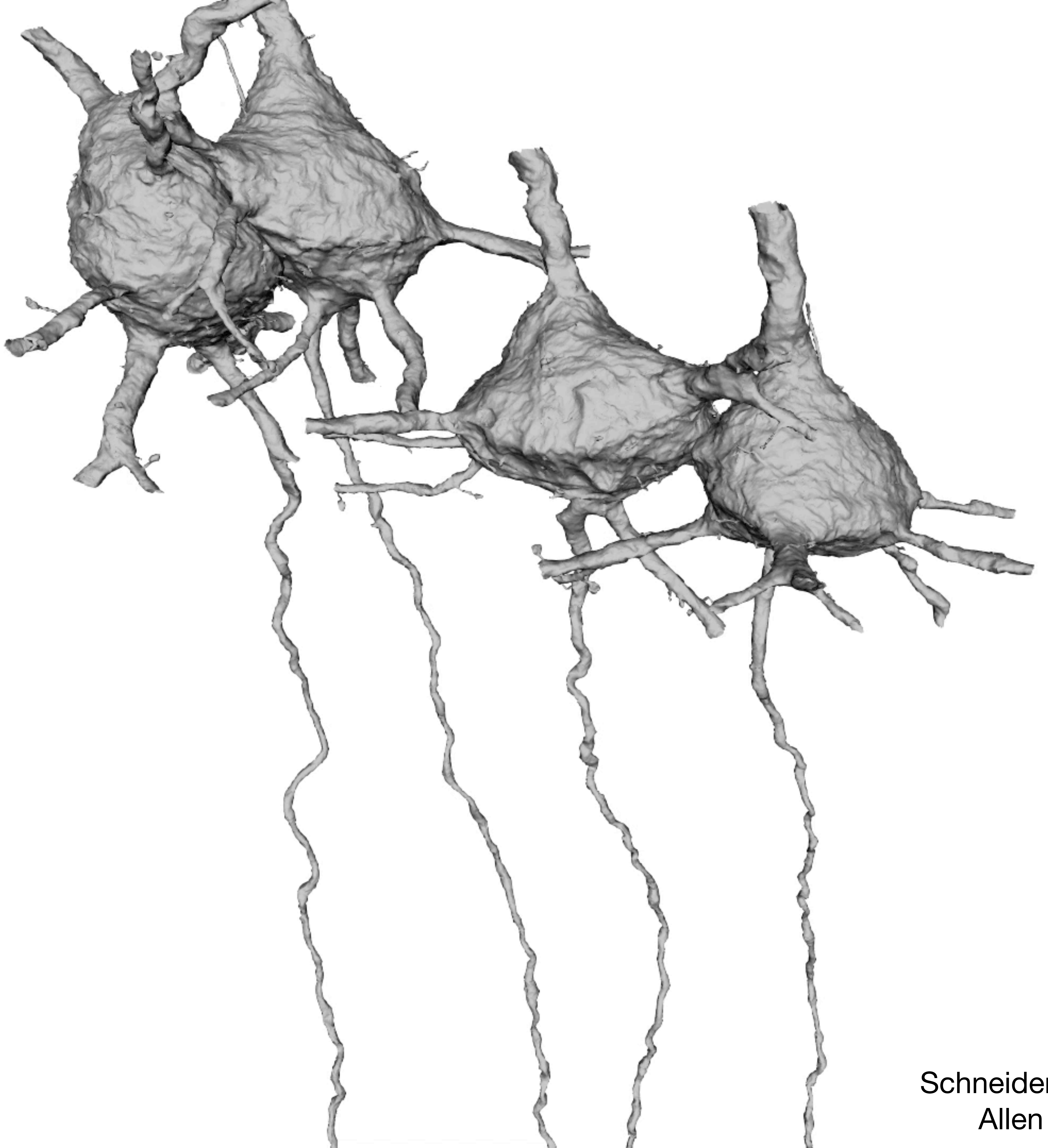


Unit V: More data, more problems

Connectomics

- Finding the “wiring diagram” of the brain by reconstructing cells from electron microscopy image stacks.
- The connectome of *C. elegans* (c.f. Lab 8) was published by White et al, **1986**, with 302 neurons and ~7k synapses.
  - It has recently been refined (Cook et al, 2019) for both sexes and over development (Witvliet et al, 2020).
- Large scale efforts are underway to map the connectome of **other model organisms**: Drosophila (Xu et al, 2020), larval zebrafish (Kunst et al, 2019), mouse (Oh et al, 2014; Schneider-Mizell et al, 2020)
- **Statistical challenges:** image segmentation, 3D reconstruction, shape analysis, network analysis, …





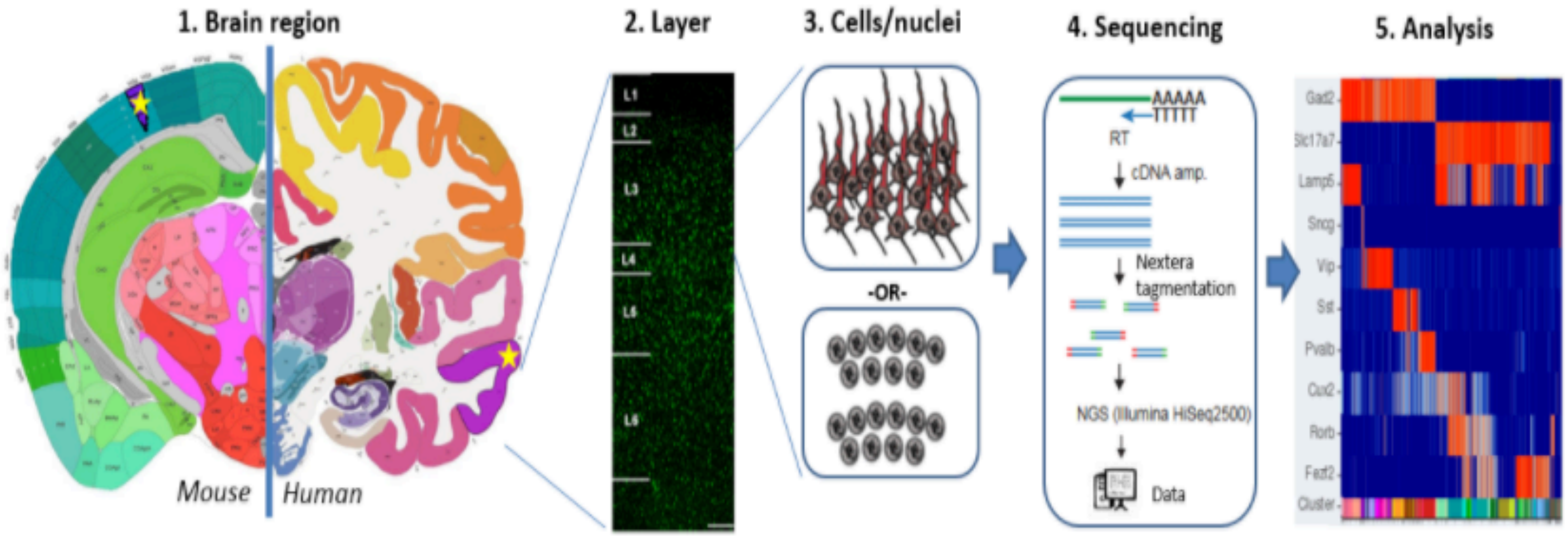
Schneider-Mizell et al. bioRxiv 2020
Allen Institute for Brain Science

Genetic Sequencing

Characterizing cell types

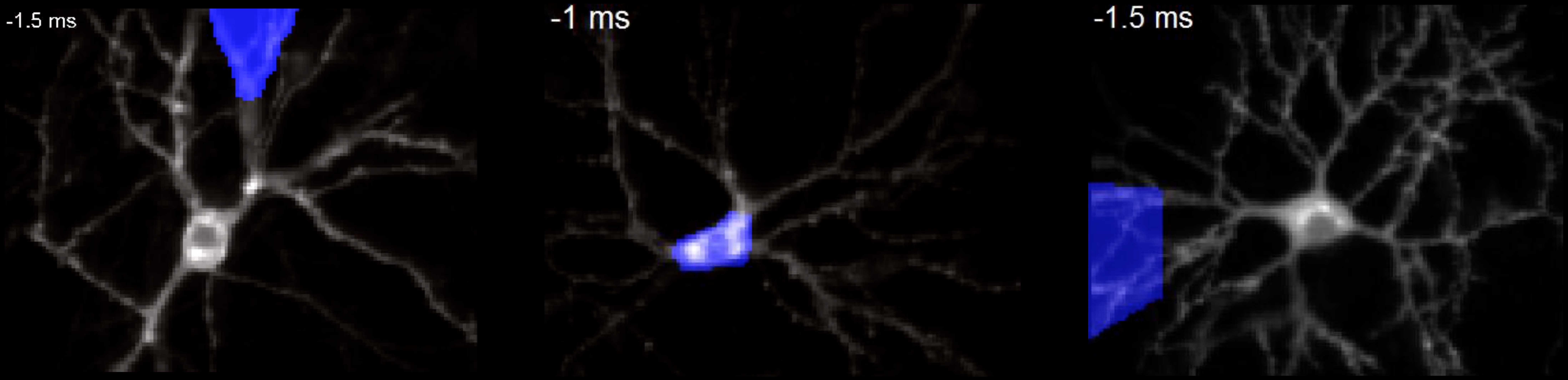
Background: Profiling Cellular Diversity in the Brain

The mammalian brain is composed of many cell populations that differ based on their molecular, morphological, electrophysiological and functional characteristics. Classifying these cells into types is one of the essential approaches to defining the diversity of the brain's building blocks. This project seeks to characterize cortical diversity at the cellular level for several neuroanatomical areas in both mouse and human. Additional data will be released regularly, building toward a complete picture of cellular diversity of the brain and how this diversity is conserved across species.



https://portal.brain-map.org/atlases-and-data/rnaseq

Voltage Imaging



Hochbaum et al (2014)

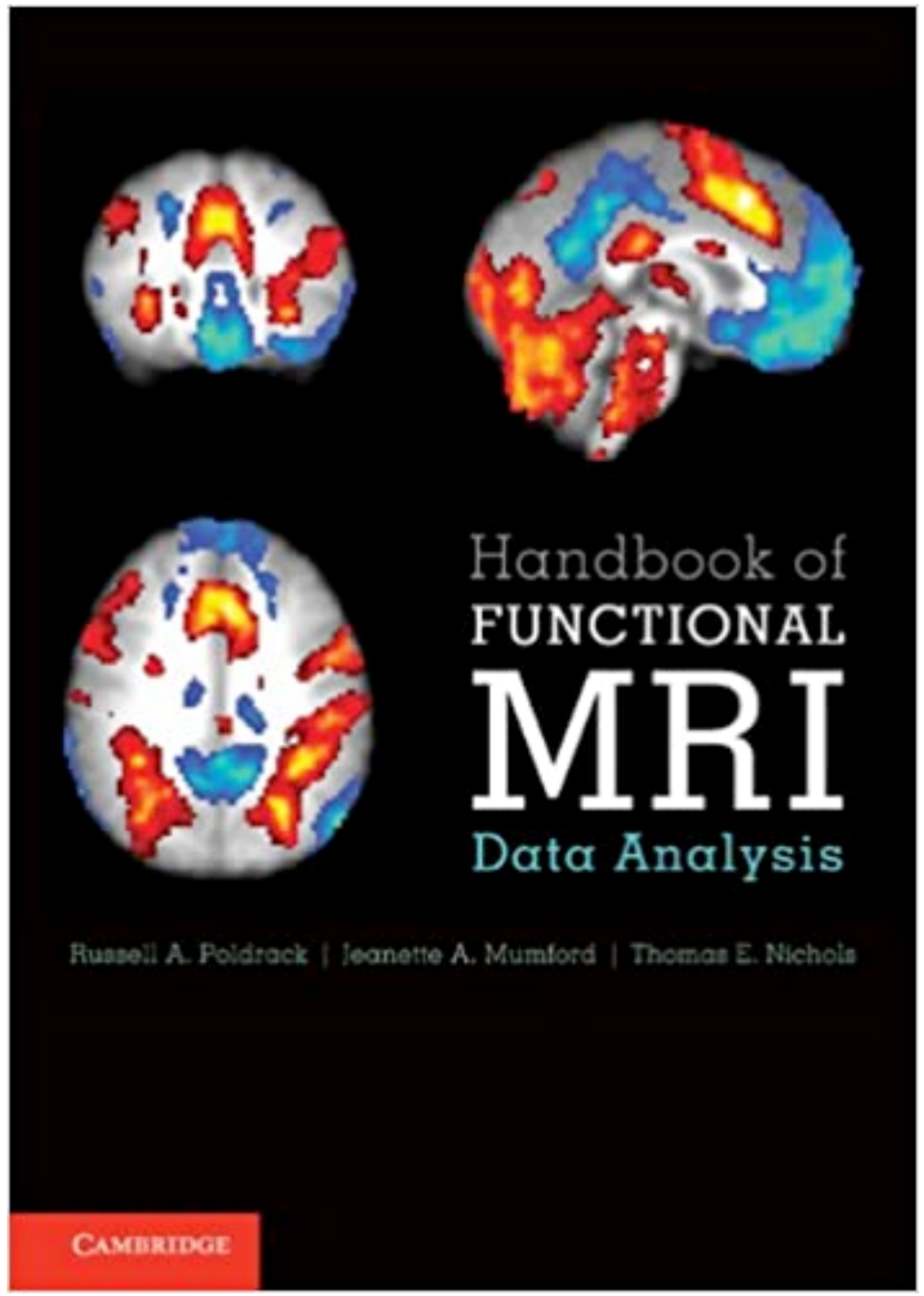
Goal: develop a model that captures spatiotemporal voltage dynamics and use it to smooth noisy imaging data with low temporal resolution.

Challenge: we don't know the precise ion channel kinematics.

Functional Magnetic Resonance Imaging (fMRI)

Measuring neural activity in the human brain

- fMRI measures the **blood oxygenation level dependent (BOLD) contrast** to measure blood flow, which is correlated with neural activity.
- “Resting state” fMRI has been used to characterize the **default mode network** of correlated brain regions and apparent brain states.
- fMRI is one of our best tools for measuring human brain activity in healthy subjects.
- **Statistical challenges**: Multivariate analysis, functional connectivity, hypothesis testing.



Electroencephalography (EEG)

- EEG measures electrical activity in the brain via electrodes positioned along the scalp.
- It is (relatively) non-invasive, but electrical signals are filtered and attenuated by the skull.
- Commonly used to diagnose epilepsy, sleep disorders, etc.
- Recent work by **Prof. Emery Brown** (MIT) uses EEG to monitor patients during anesthesia.
- **Statistical challenges:** signal processing, spectral analysis, state space modeling.

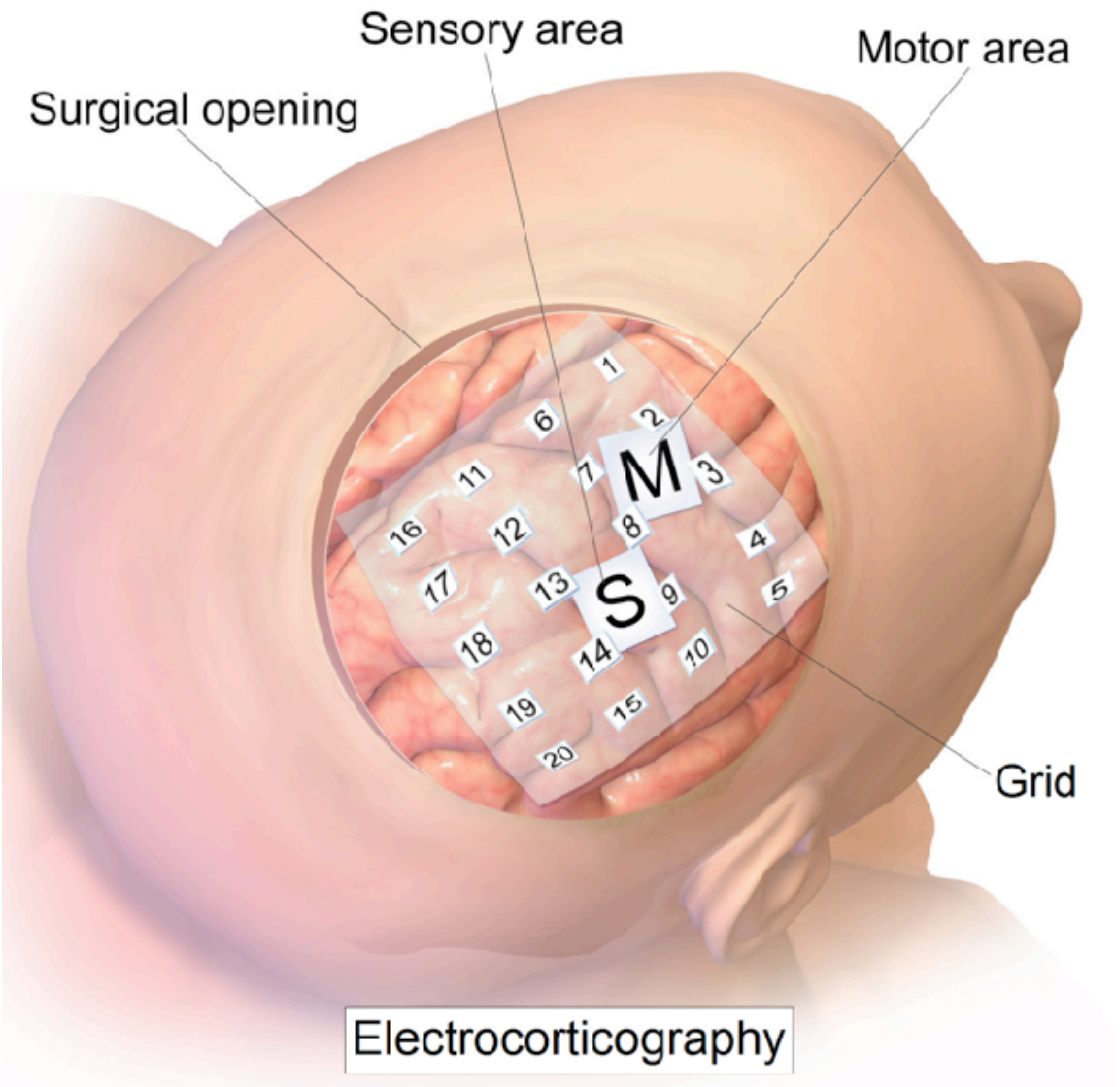
(not me)

https://arstechnica.com/information-technology/2018/04/hacking-your-brain-researchers-discover-security-bugs-in-eeg-systems/

Electrocorticography (ECoG)

- Sometimes called "intracranial EEG," ECoG measures electrical activity in the brain via electrodes placed on the exposed surface of the brain.
- As such, it is only used for patients who need neurosurgery, e.g. due to medically-intractable epilepsy.
- Surface and depth electrodes record **local field potentials (LFP)** and single neurons.
- ECoG is a promising technique for developing brain-computer interfaces. C.f. work from the **Henderson Lab** at Stanford and Eddie Chang's group at UCSF.
- **Statistical challenges:** dynamical systems modeling, neural decoding, transfer learning (between patients with different grid placements).

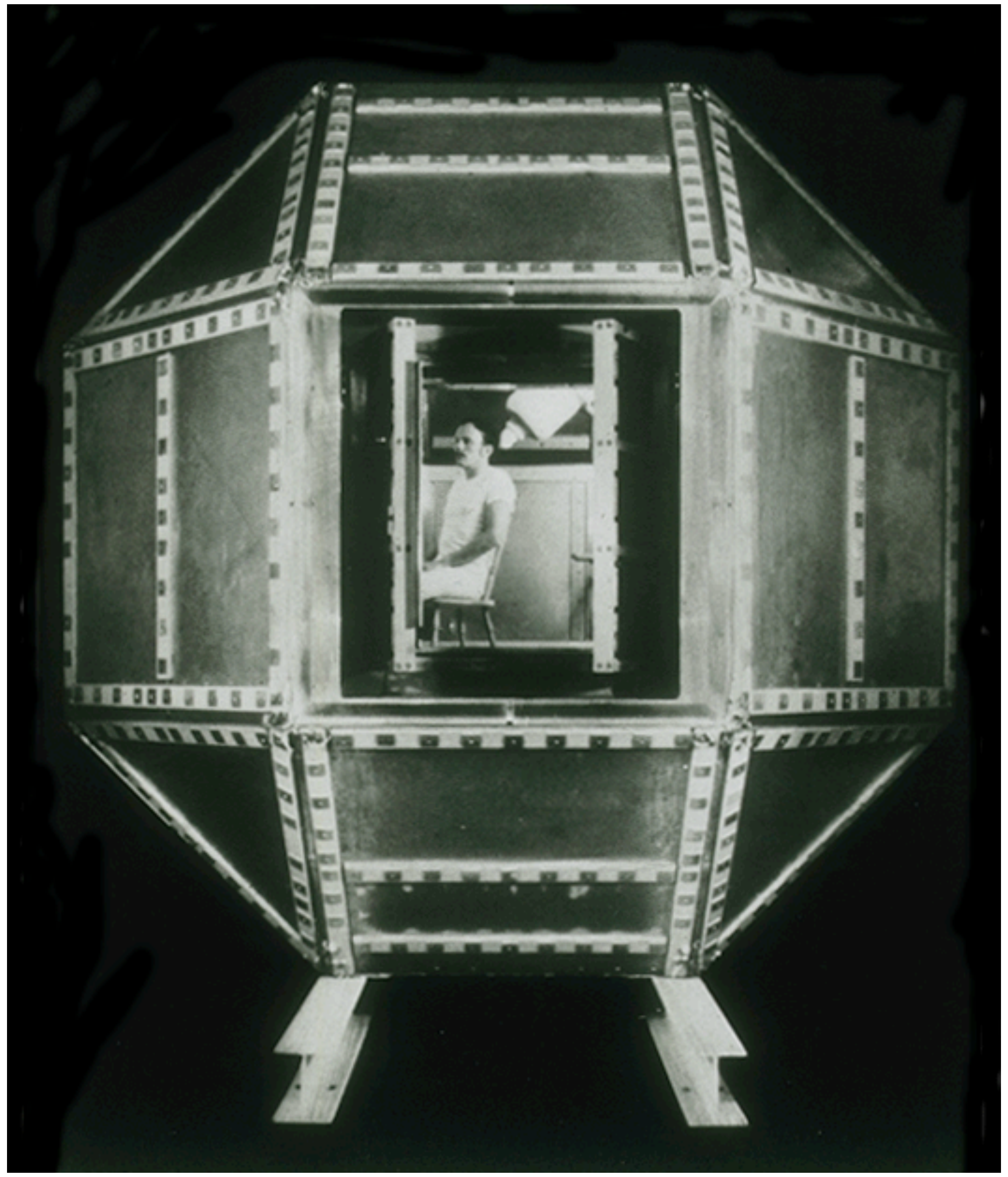
(not me)



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

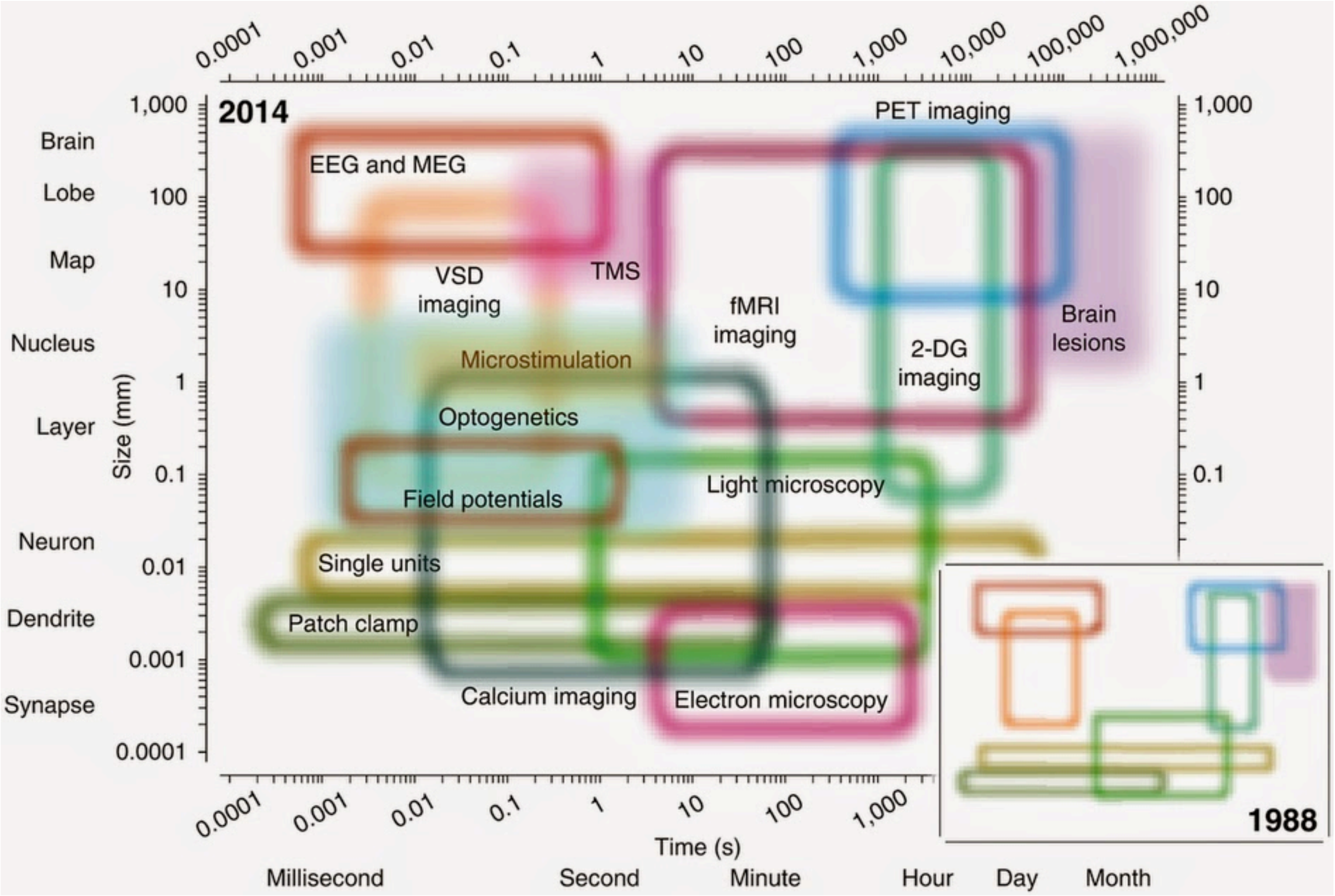
Magnetoencephalography (MEG)

- MEG measures brain activity via the magnetic fields induced by ionic currents.
- MEG has a fast temporal response (10ms), making it particularly useful for fast computations like rapid image classification and speech.
- See Laura Gwilliams' work at Stanford!
- **Statistical challenges:** source localization, "beam forming,"



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

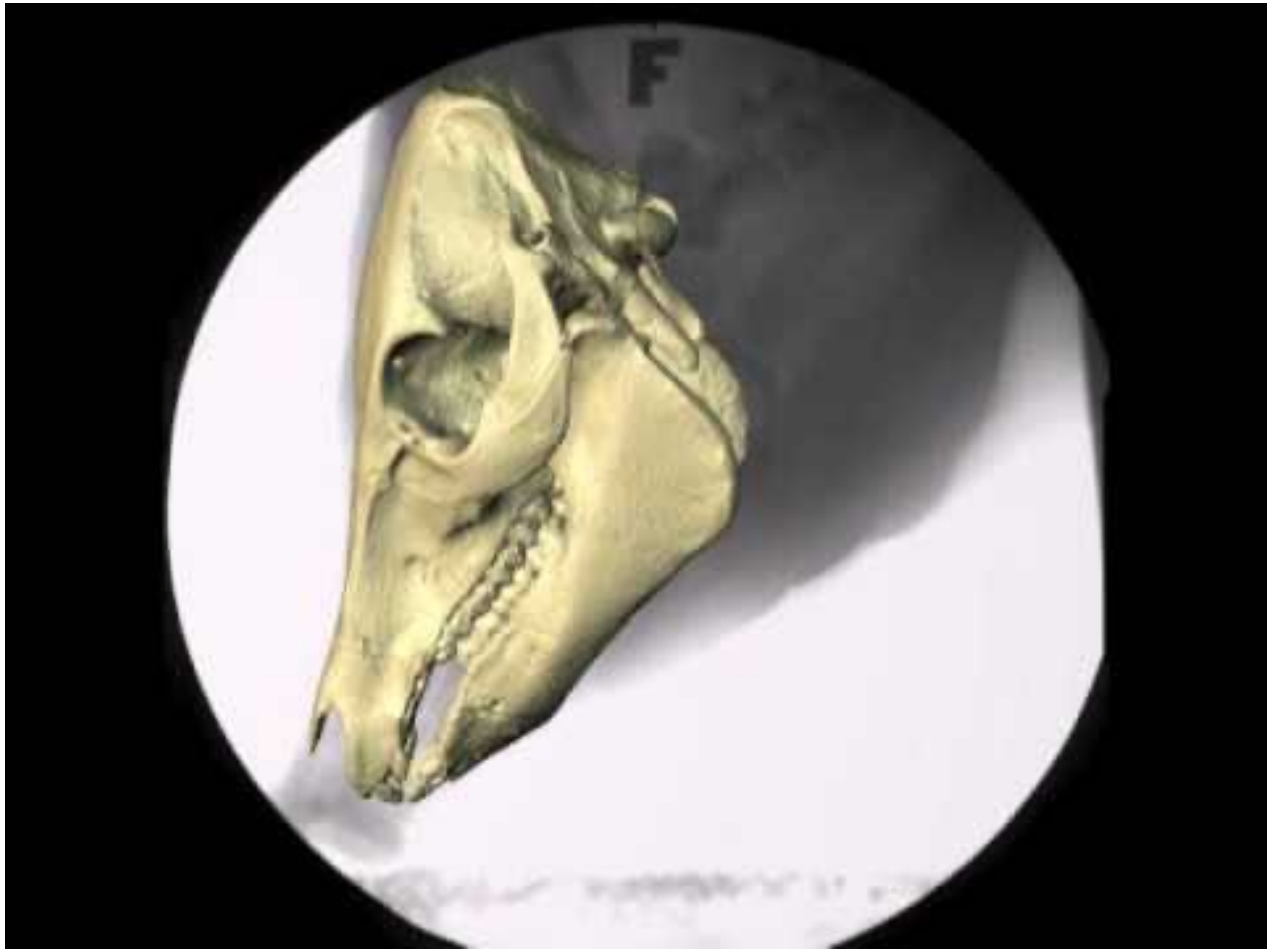
Spatial and temporal trade-offs



Sejnowski et al, Nature Neuro. 2014.

X-ray Reconstruction of Moving Morphology (XROMM)

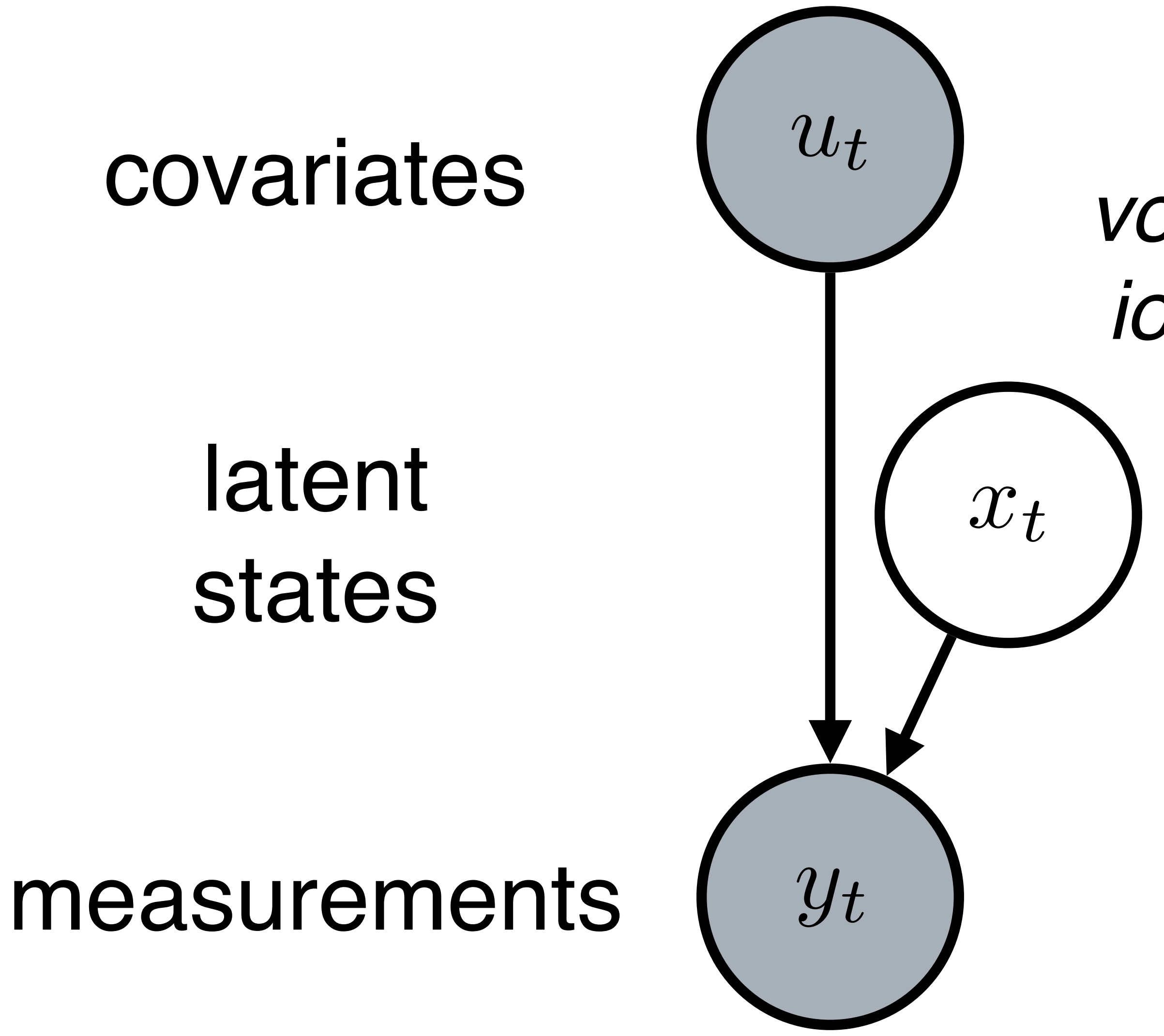
“XROMM combines **3D models of bone morphology** with movement data from **biplanar x-ray video** to create highly accurate (±0.1 mm) reanimations of the 3D bones moving in 3D space.”

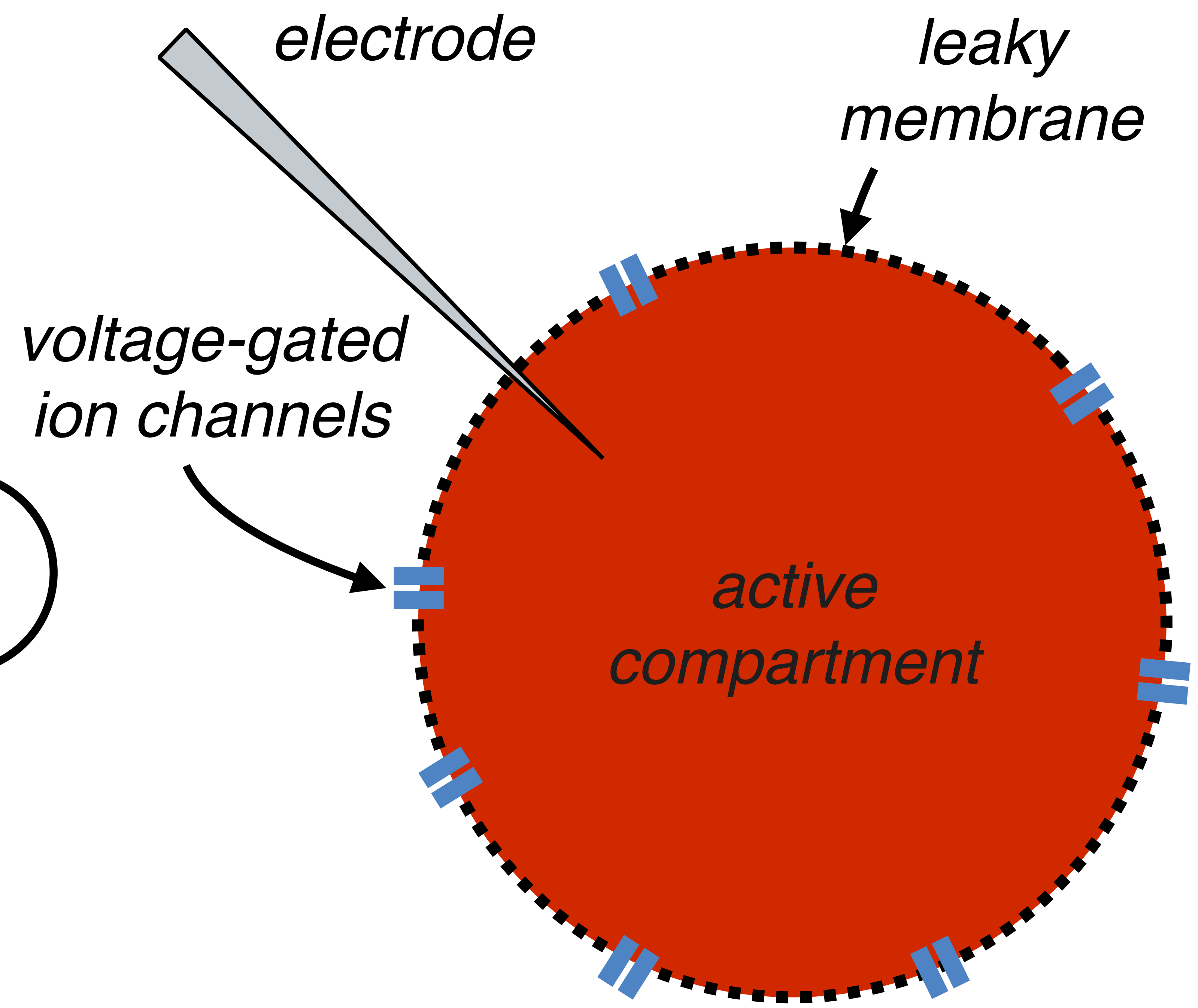


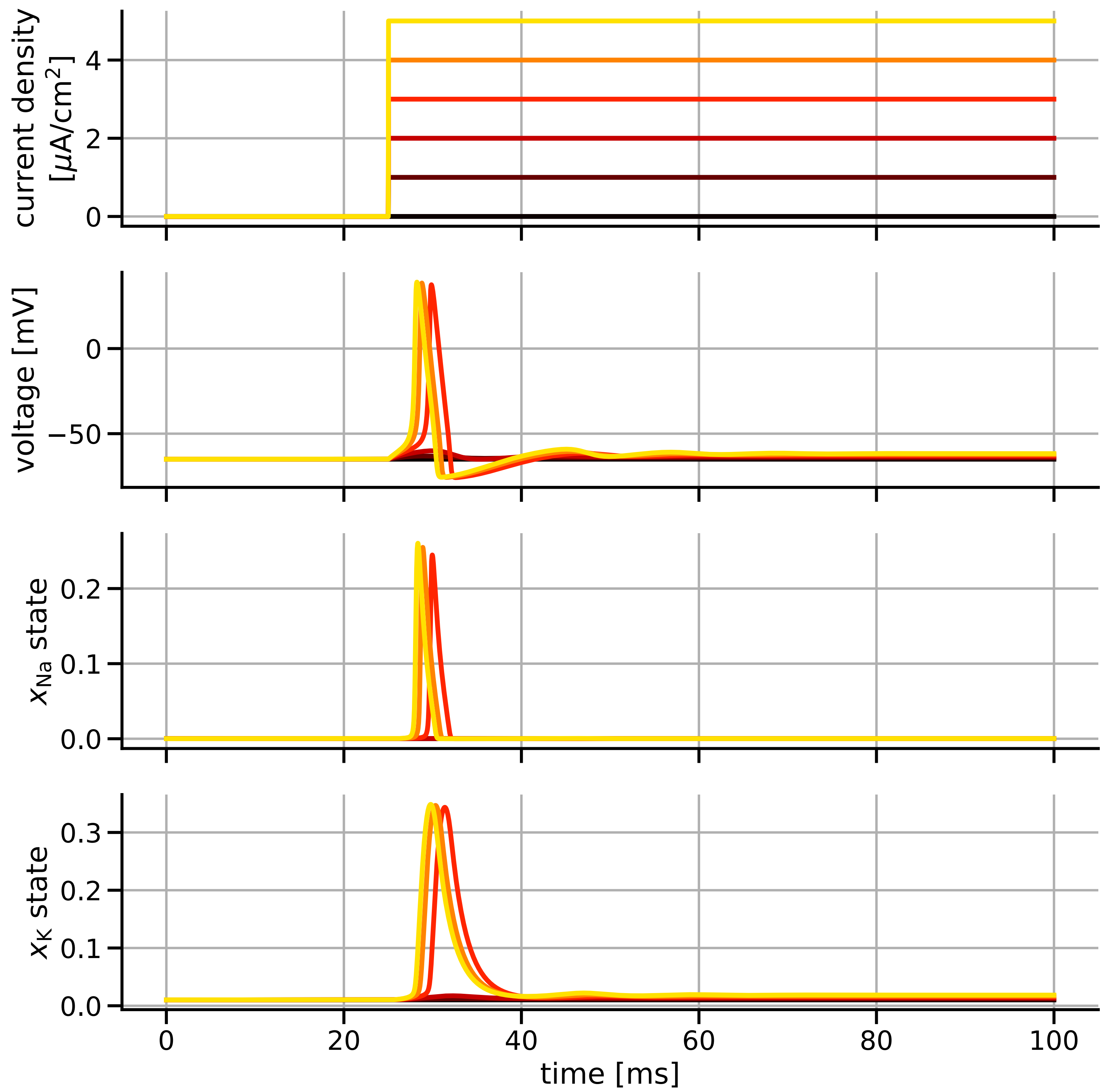
https://www.xromm.org/

Unit V: More models, more problems

Biophysical models



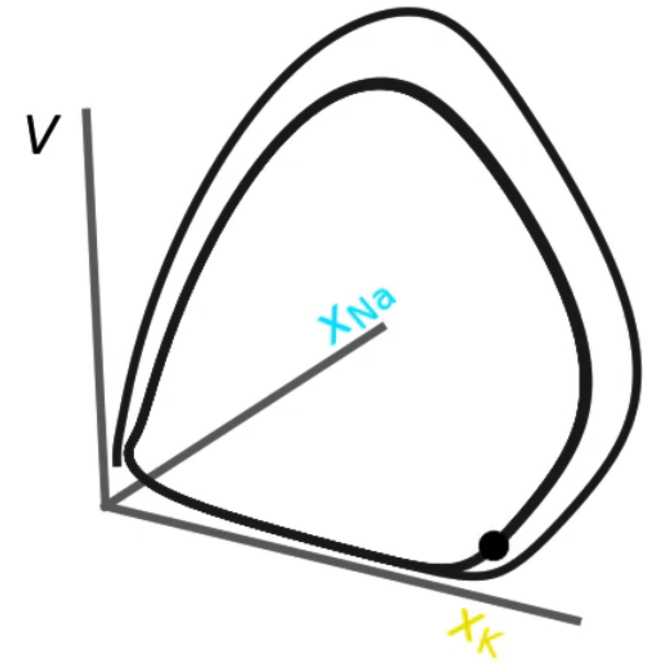




$$C\frac{\mathrm{d}V}{\mathrm{d}t} = I_{\text{in}} - g_L(V - E_L) - \sum_k g_k x_k (V - E_k)$$

$$\frac{\mathrm{d}x_k}{\mathrm{d}t} = f_k(x_k, V)$$

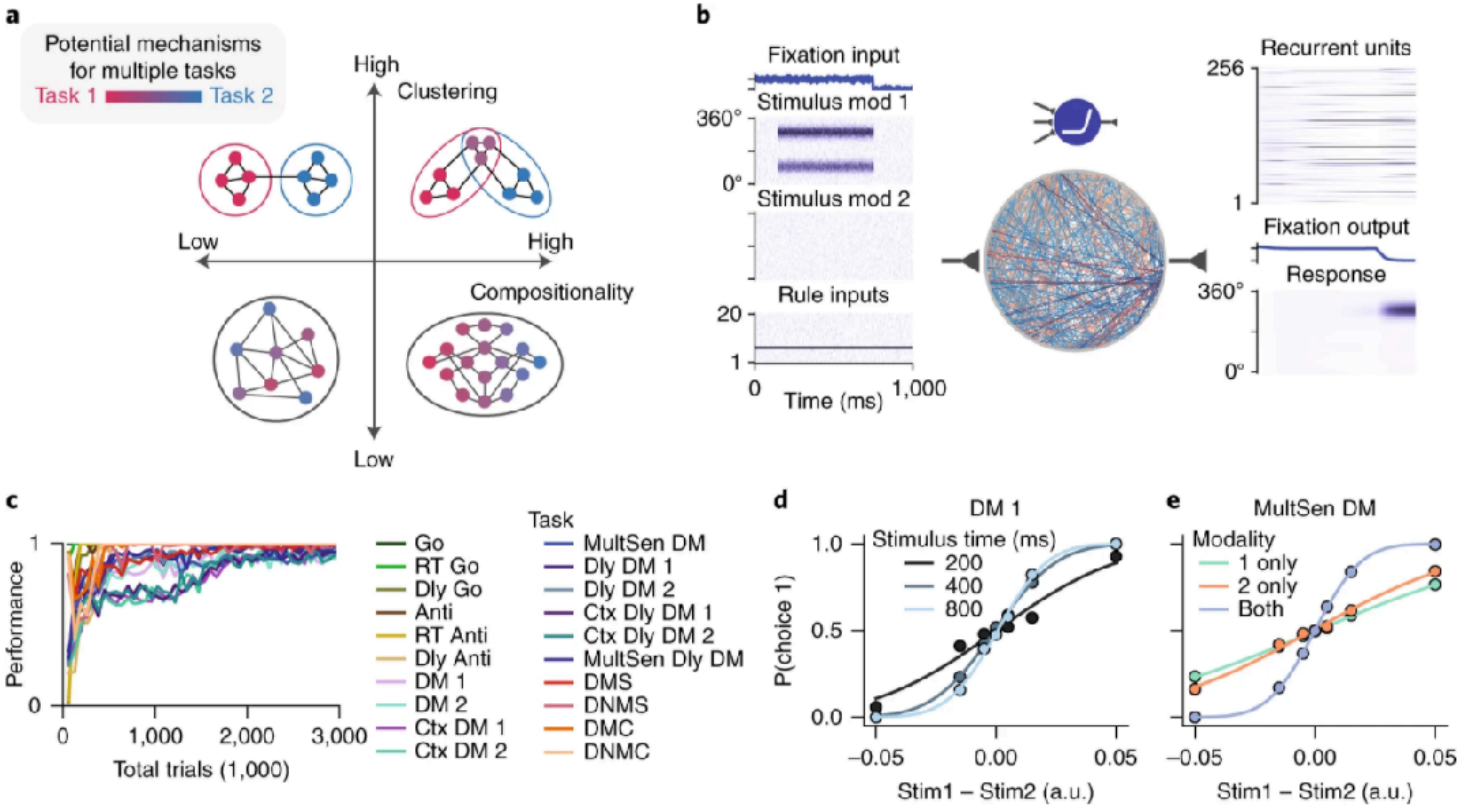
ion channel states
and their dynamics



Task-based modeling with RNNs

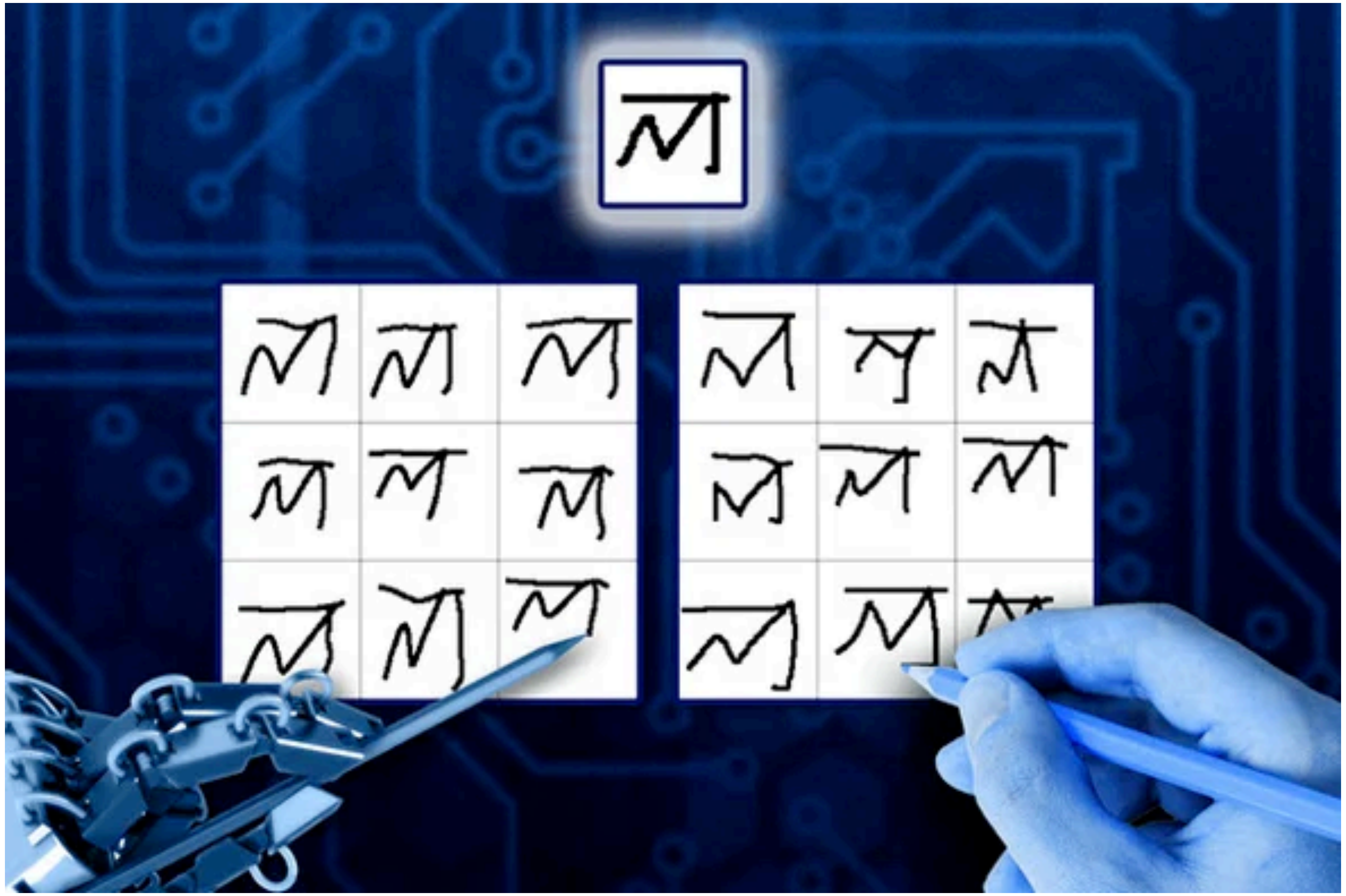
Fig. 1: A recurrent neural network model is trained to perform a large number of cognitive tasks.

From: Task representations in neural networks trained to perform many cognitive tasks



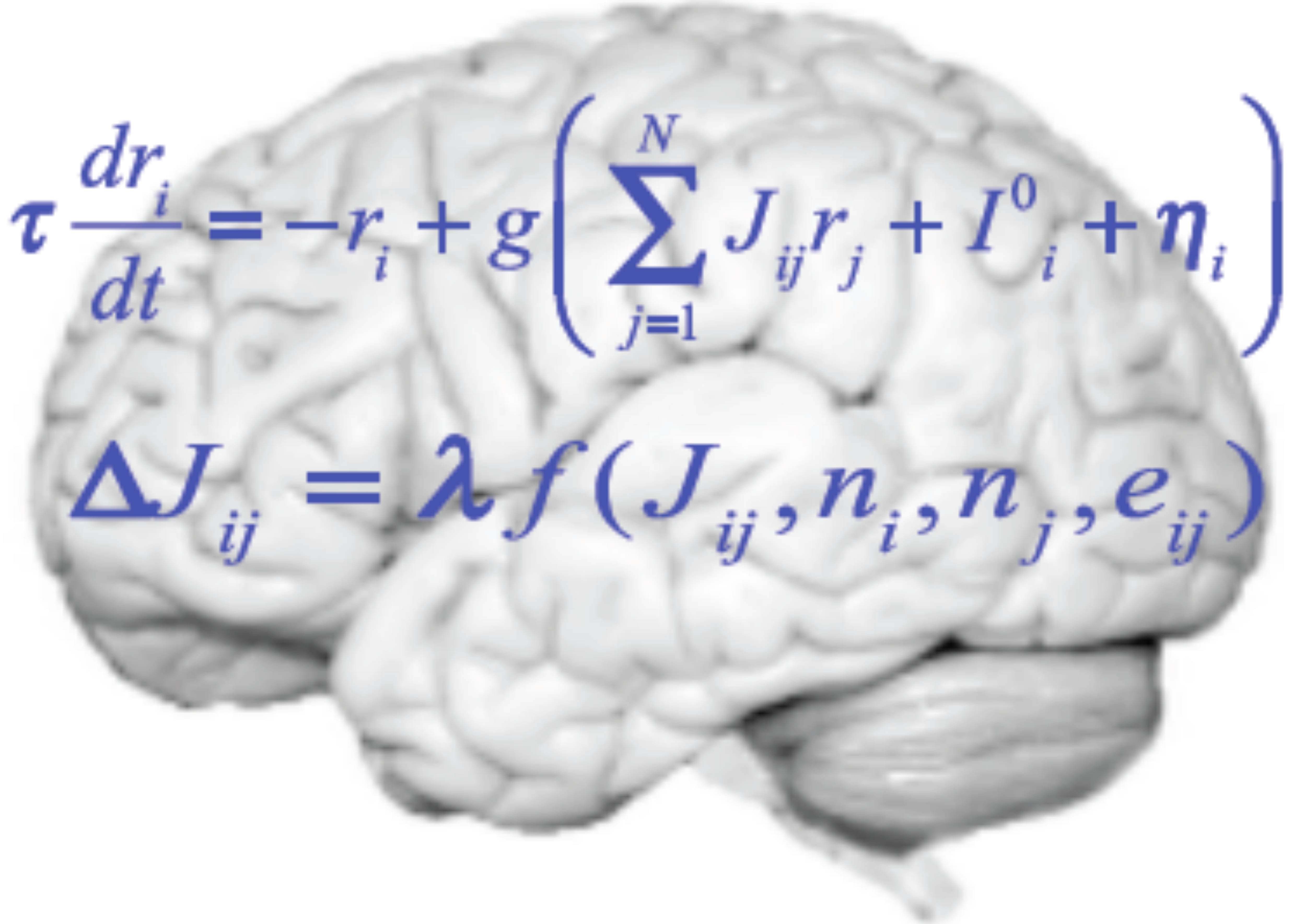
Yang et al, Nature Neuro. 2019.

The Bayesian Brain



https://blogs.scientificamerican.com/cross-check/are-brains-bayesian/
Lake et al (Science, 2015)

Theoretical Neuroscience



Related Faculty at Stanford

(I'm sure I'm missing many!)

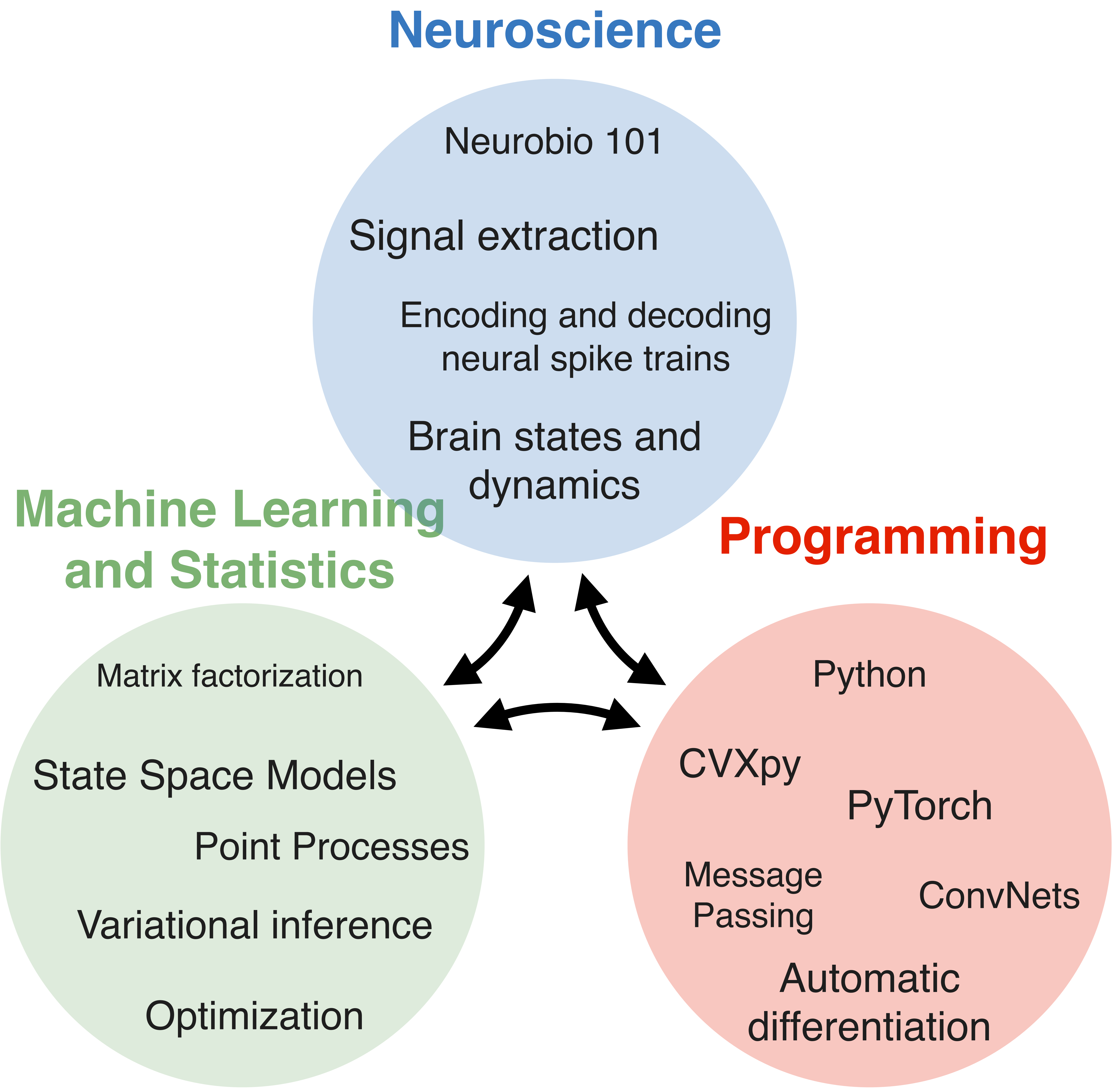
- Rosa Cao (Philosophy)
- EJ Chichilnisky (EE)
- Karl Deisseroth (Psychiatry and BioE)
- Shaul Druckmann (Neurobiology)
- Surya Ganguli (Applied Physics)
- Justin Gardner (Psychology)
- Tobias Gerstenberg (Psychology)
- Noah Goodman (Psychology and CS)
- Laura Gwilliams (Psychology, Wu Tsai, SDS)
- Liqun Luo (Biology)
- Jay McClelland (Psychology)
- Paul Nuyujukian (BioE)
- Russ Poldrack (Psychology)
- Robert Sapolsky (Biology)
- Mark Schnitzer (Applied Physics)
- Dan Yamins (Psychology and CS)

Conclusion

- It’s an exciting age for neuroscience: a confluence of technological advances is offering many new ways to measure the brain in action.
- In parallel, machine learning and statistics are experiencing a renaissance of their own, fueled by advances in deep learning, generative modeling, efficient hardware, and more.
- At the intersection, there is an array of exciting problems to tackle, ranging from using ML and statistical methods to better analyze and glean insight from neural data, to using ML to advance new hypotheses for how the brain computes.
- This course has introduced a few of the current methods for analyzing brain data. In fact, you’ve now built many of these tools from the ground up!
- We’ve just scratched the surface though, and there are tons of great courses and books to explore if you want to learn more.

Feedback

- As always, we would greatly appreciate your feedback on the course.
- Were the learning objectives (*understand, develop, implement, & generalize*) achieved?
- In particular, we dropped the in-class lab component this year. Do you think we should revive it next time?



Thank You!