Machine Learning Methods for Neural Data Analysis

Lecture 1: Course Overview

Scott Linderman

STATS 220/320 (NBIO220, CS339N)

Welcome!

Please take this short survey:

https://tinyurl.com/stats320survey

Introductions

- About me:
  - Asst. Prof. of Statistics and Computer Science (by courtesy)
  - Institute Scholar, Wu Tsai Neurosciences Institute
  - Fun fact: I can recite the alphabet backward in under 3 seconds
- TA’s:
  - Noah Cowan (Stats PhD student)
  - TBD (Most likely a stats PhD student :)

What drew you to this course?

Are you an experimental neuroscientist seeking a deeper understanding of the analysis methods you're using?

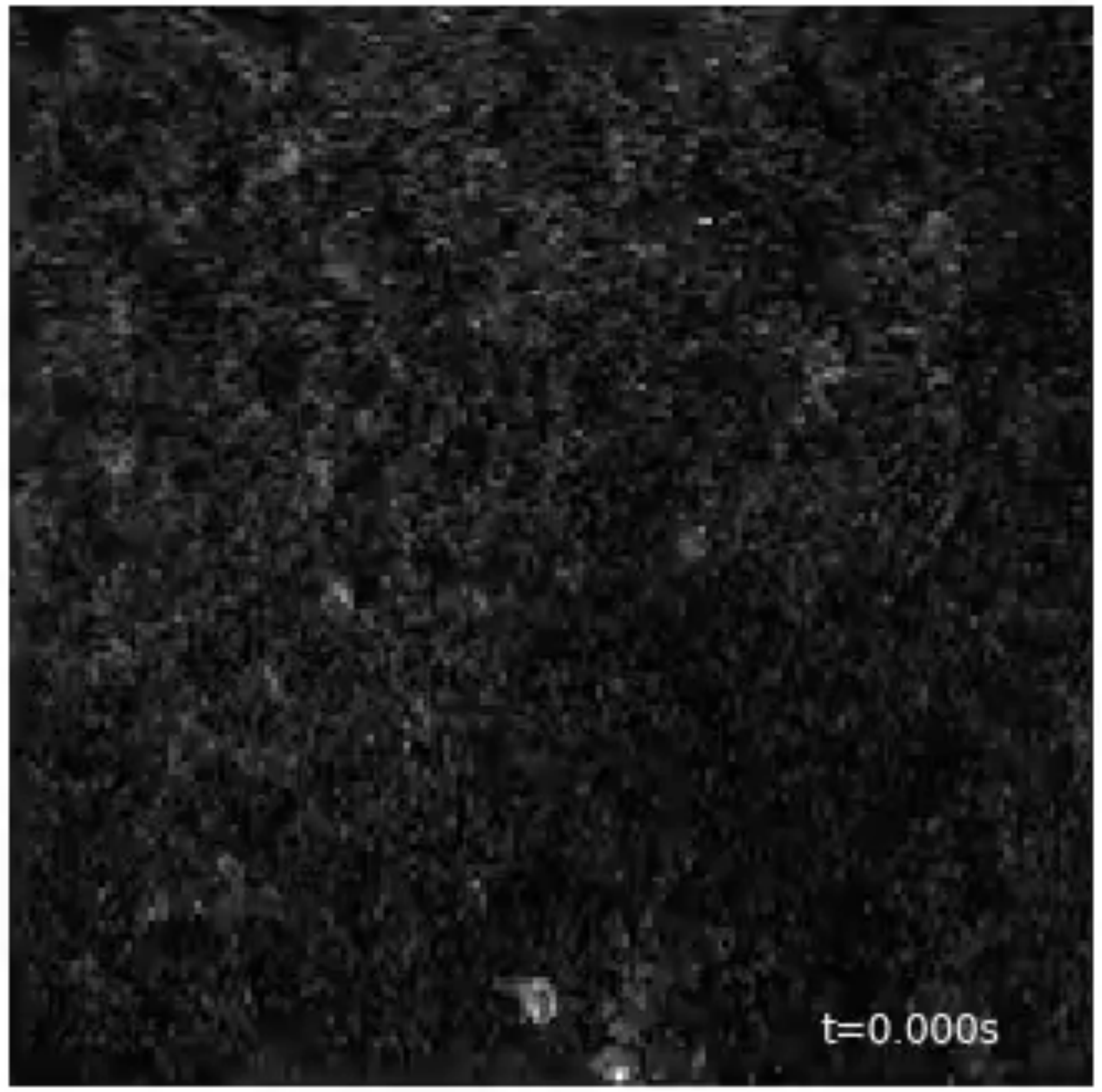
Are you interested in machine learning and want to learn more about scientific applications?

Are you a methods developer looking for challenging problems?

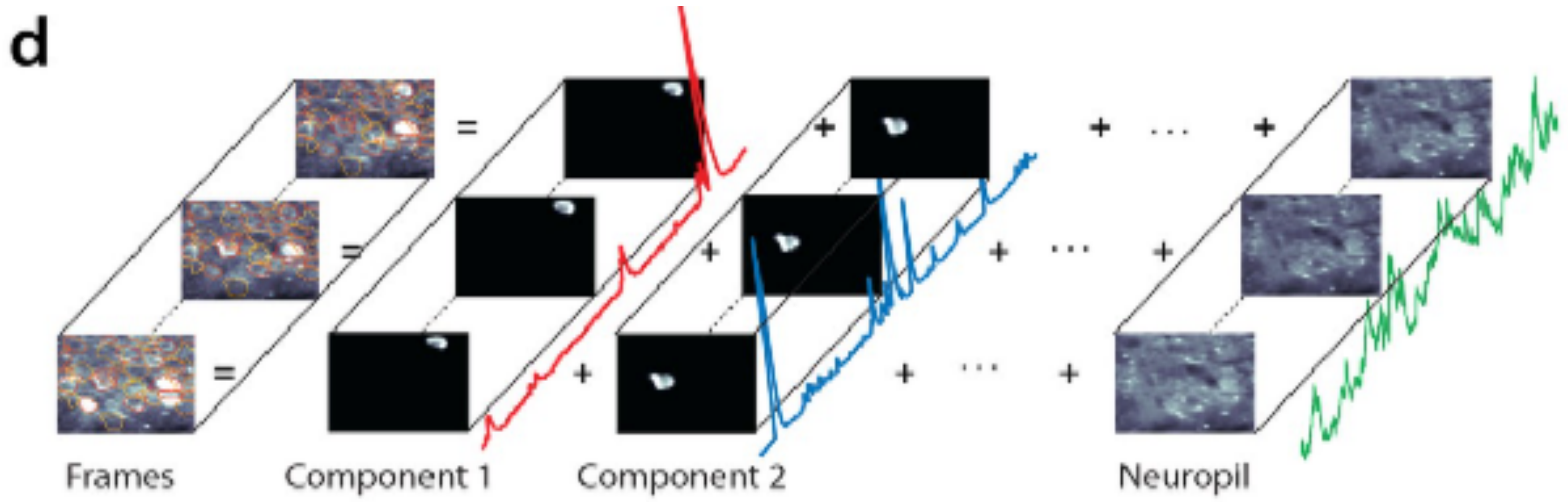
Are you interested in data science, coding, and working with real data?

What is “neural data”? What kinds of methods?

- **Setup:** suppose a neuroscientist brings you this **calcium imaging data.**
- The white blobs are **neurons**, and the flashes show when the neurons are **spiking**.
- However, there’s lots of **background noise** in the video too.
- **Question:** How would you find the neurons and extract their activity traces?



Data from Sue Ann Koay and David Tank



Giovanucci et al (eLife, 2019)

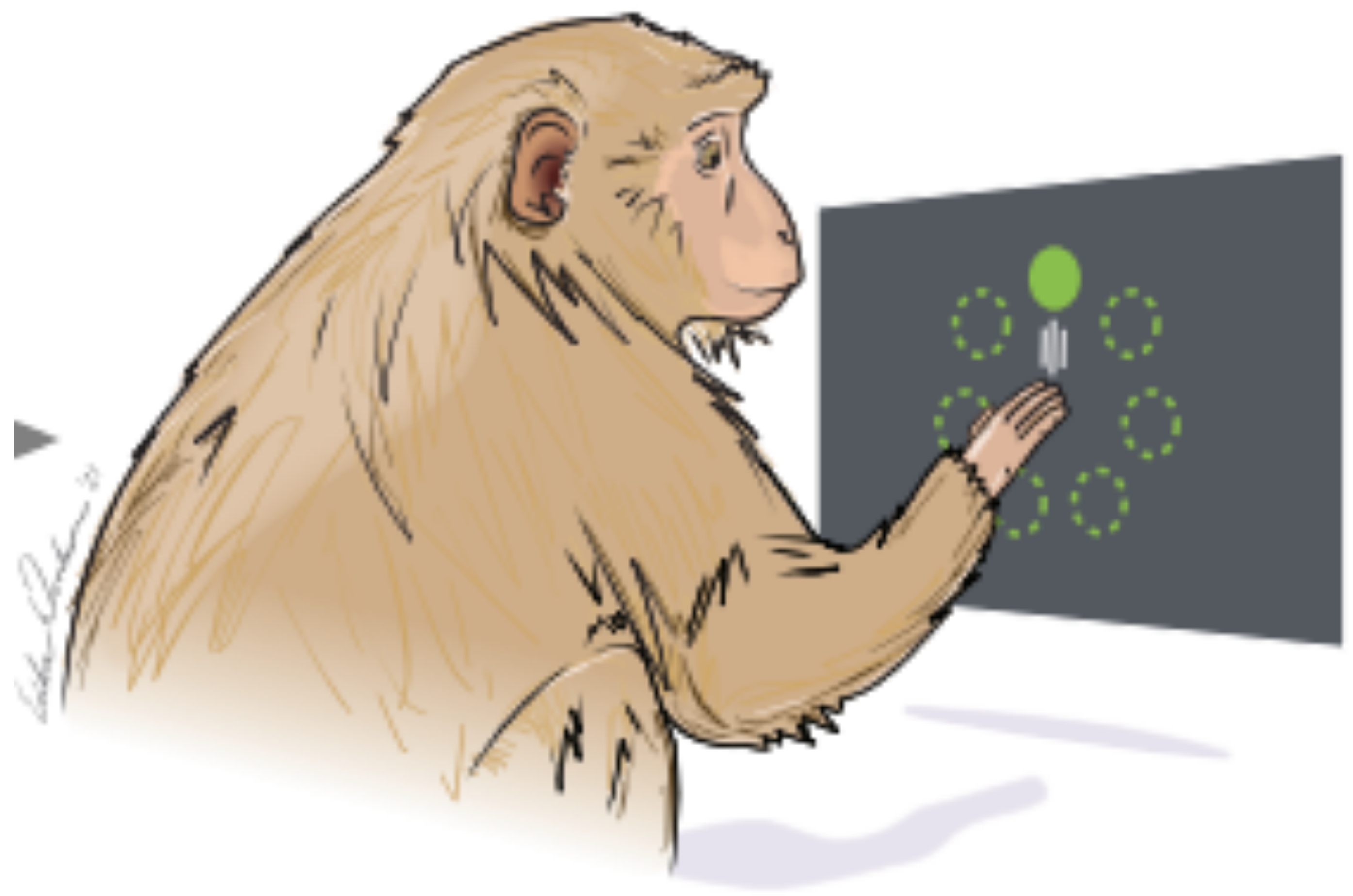
What is "neural data"? What kinds of methods?

- **Setup:** suppose you have **multi-electrode recordings** of hundreds of neurons from the brain of a monkey as it performs a **reaching task.**

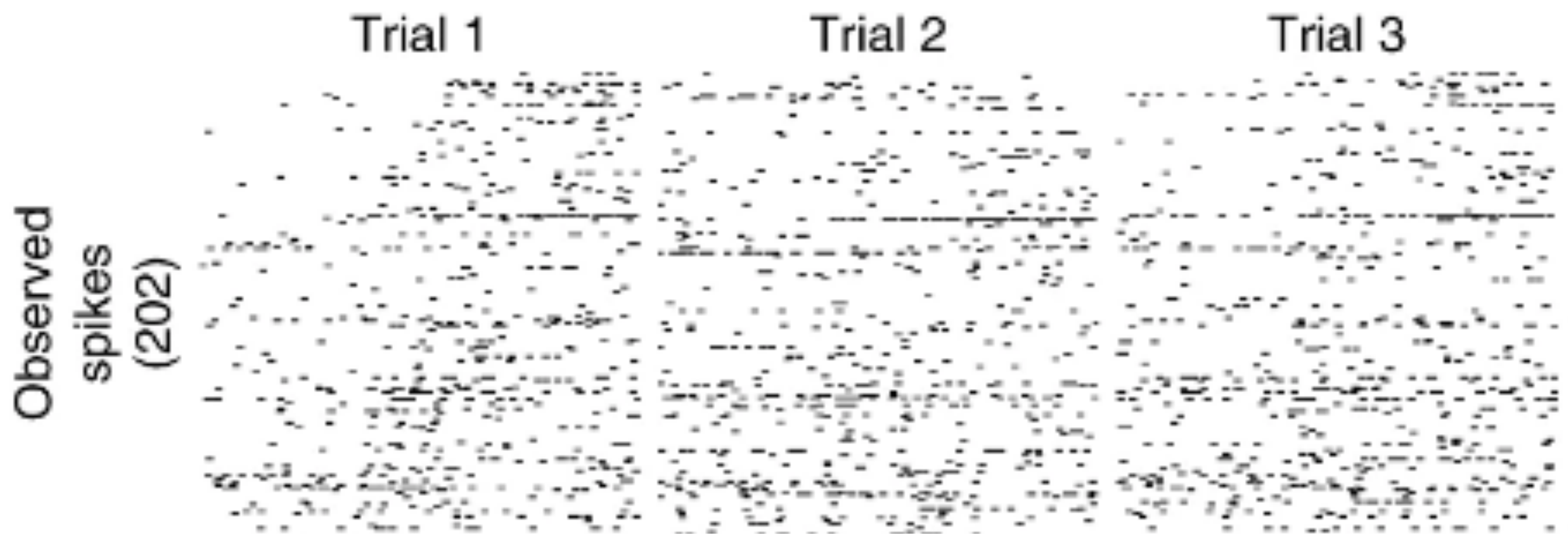
- On each trial, the monkey reaches from the center of the screen to one of the targets.

- Simultaneously, you record the **spike raster**. Each dot indicates when a neuron spiked during the trial.

- **Question:** How could you **decode** (i.e. predict) the location of the monkey's hand given the spikes?



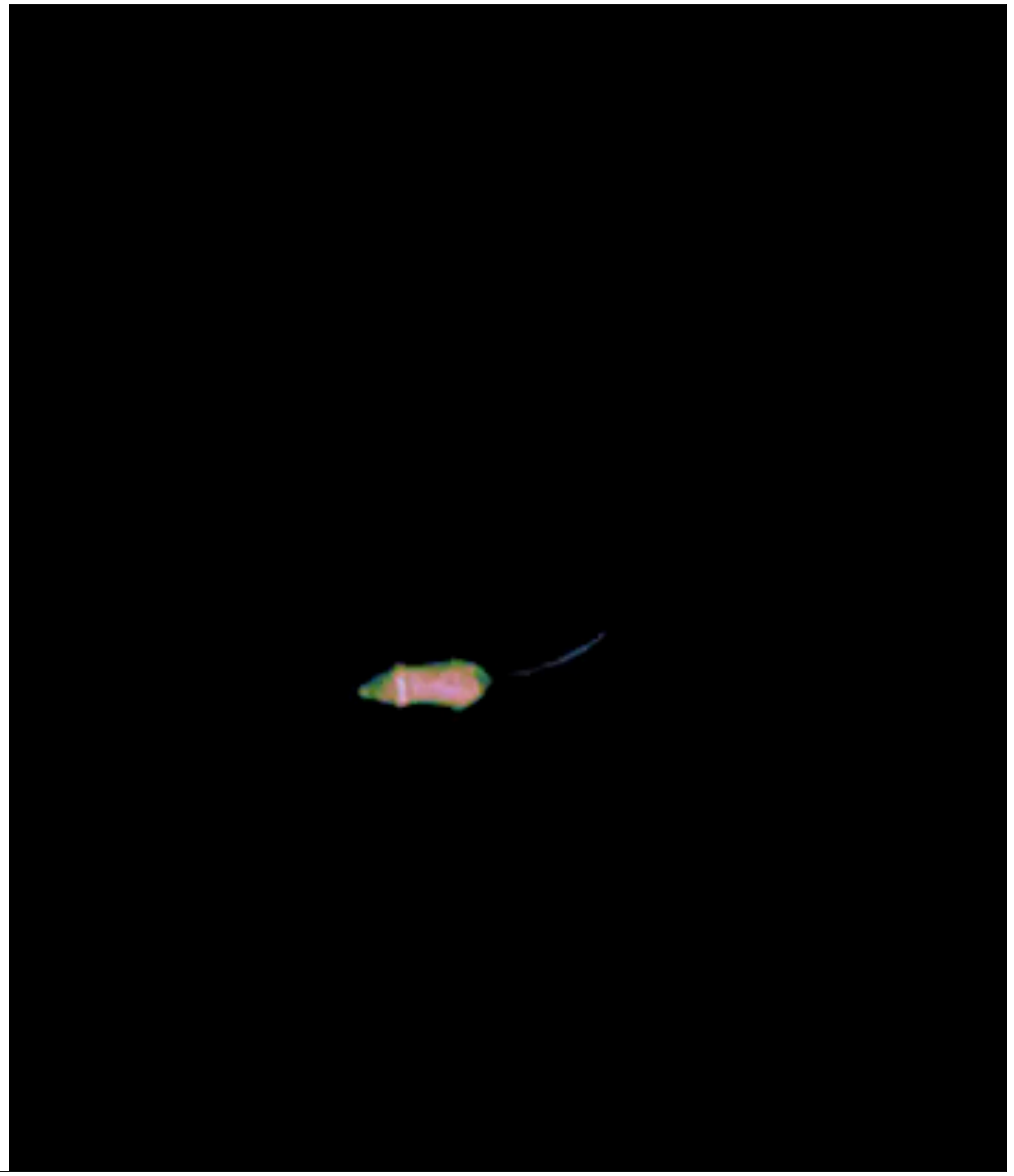
O'Shea, Duncker et al., 2022

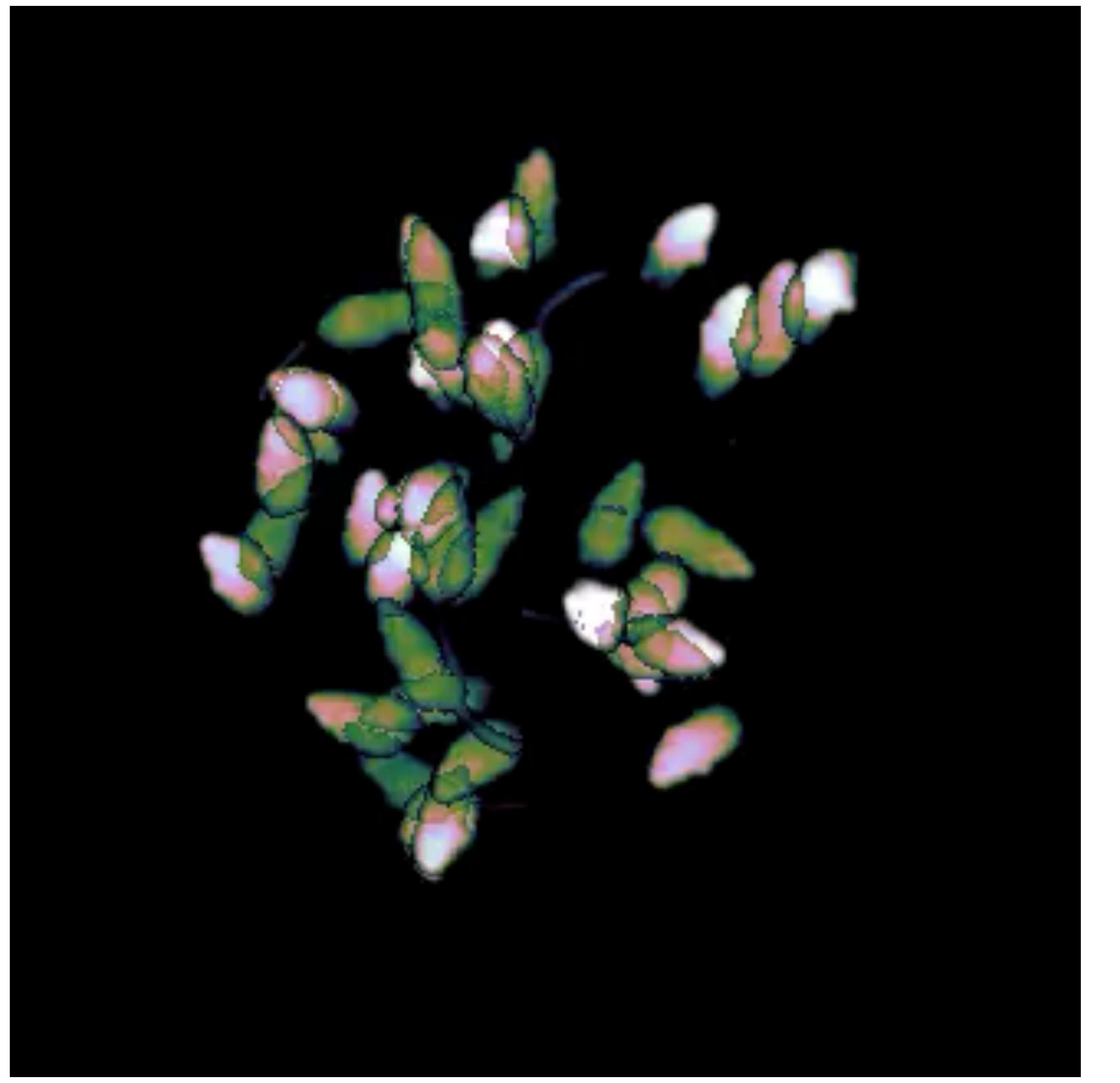


Pandarinath et al., 2019

What is “neural data”? What kinds of methods?

- **Setup:** suppose you have **depth video** of a freely moving mouse, and you want to identify **stereotyped movements**.
- If you could identify such movements, you could then study their **neural correlates**.
- Ideally, you’d like to find these stereotyped movements in an **unsupervised manner**.
- **Question:** How could you segment this video into motifs like these?



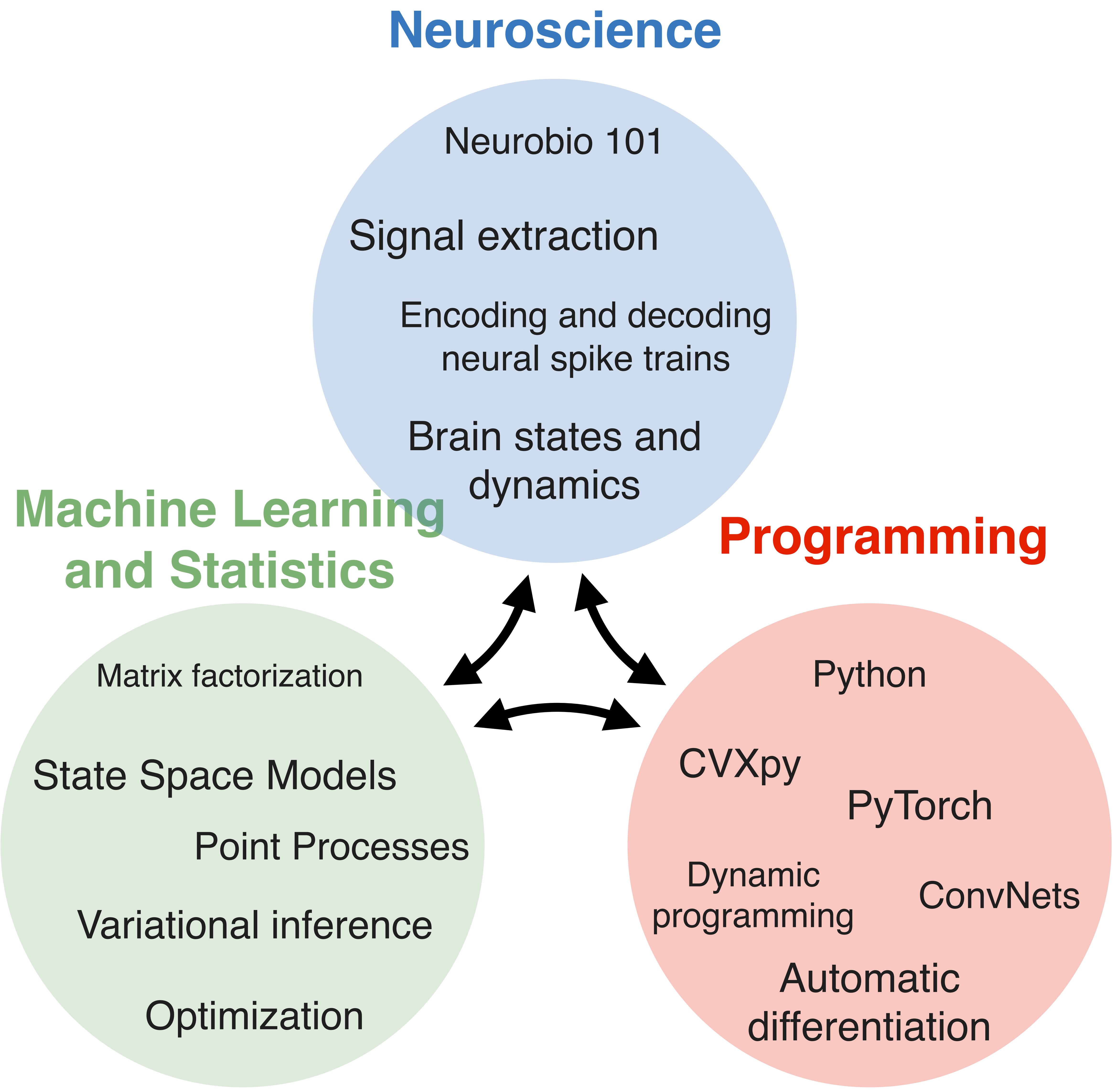


What is this course about?

- Learning about modern neural and behavioral recording methods.
- Understanding key scientific questions and data analysis challenges.
- Developing probabilistic models to tackle these challenges.
- Deriving algorithms for inference and estimation with these models.
- Implementing these algorithms in Python and applying them to data.

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.



Course Outline

- Unit I: Extracting signals of interest from raw data
- Unit II: Encoding and decoding neural spike trains
- Unit III: Latent variable models of neural and behavioral data
- Unit IV: Current research topics

- See the syllabus on the course website:

https://slinderman.github.io/stats320

Online Textbook

- I have been writing an online textbook to accompany this course:

  https://slinderman.github.io/ml4nd

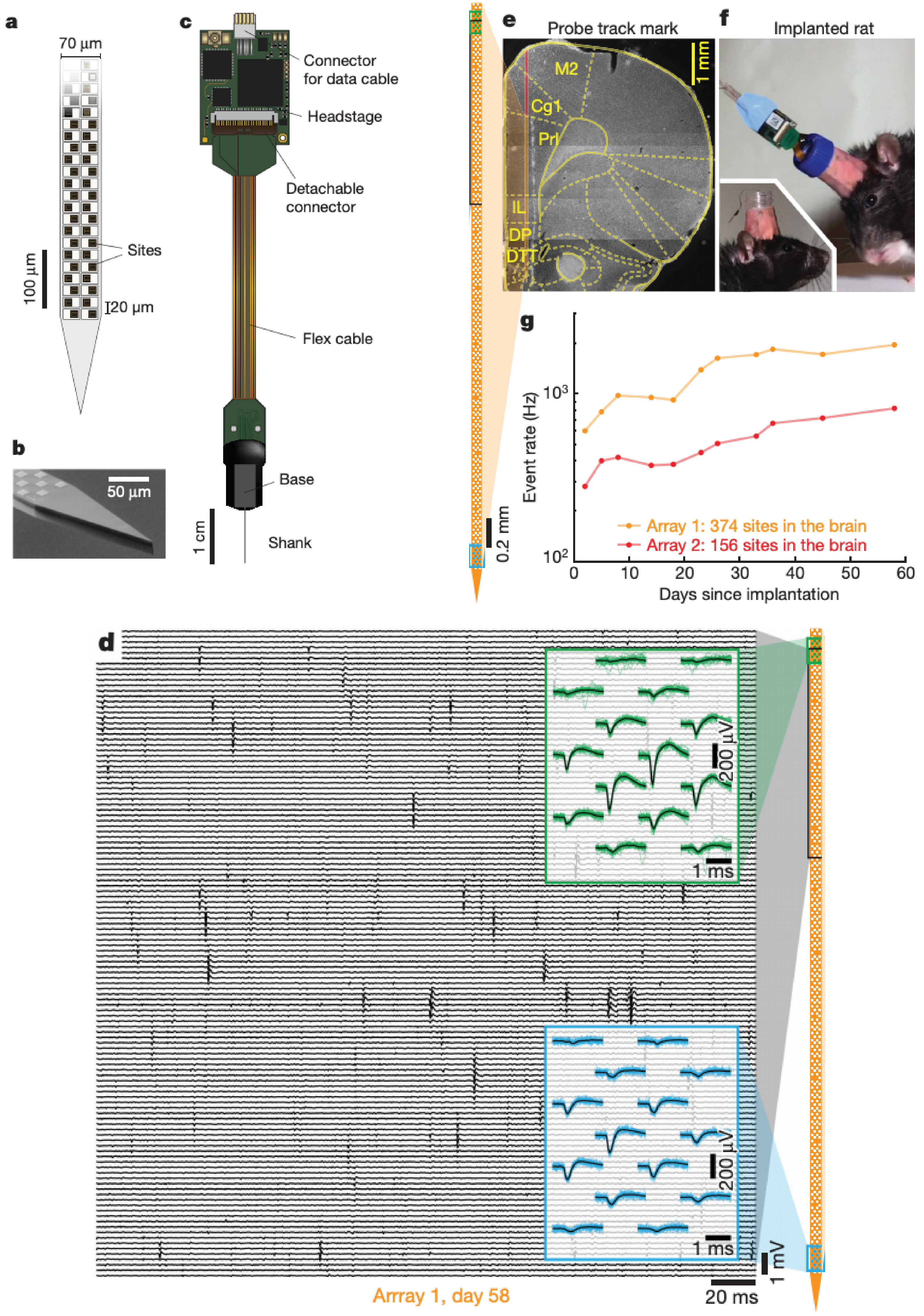
- The book contains several chapters to accompany the lectures.

- It also contains the labs, which will be your weekly assignments (more on this shortly!)

Unit 1: Signal Extraction

Spike Sorting

- Modern recording probes like **Neuropixels** measure the electrical activity of **hundreds of cells** across **multiple brain regions** simultaneously.
- The raw data is a **multidimensional time series** of **voltage measurements**, one for each recording site on the probe.
- 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.
- <u>Methods:</u> **mixture models, dimensionality reduction**

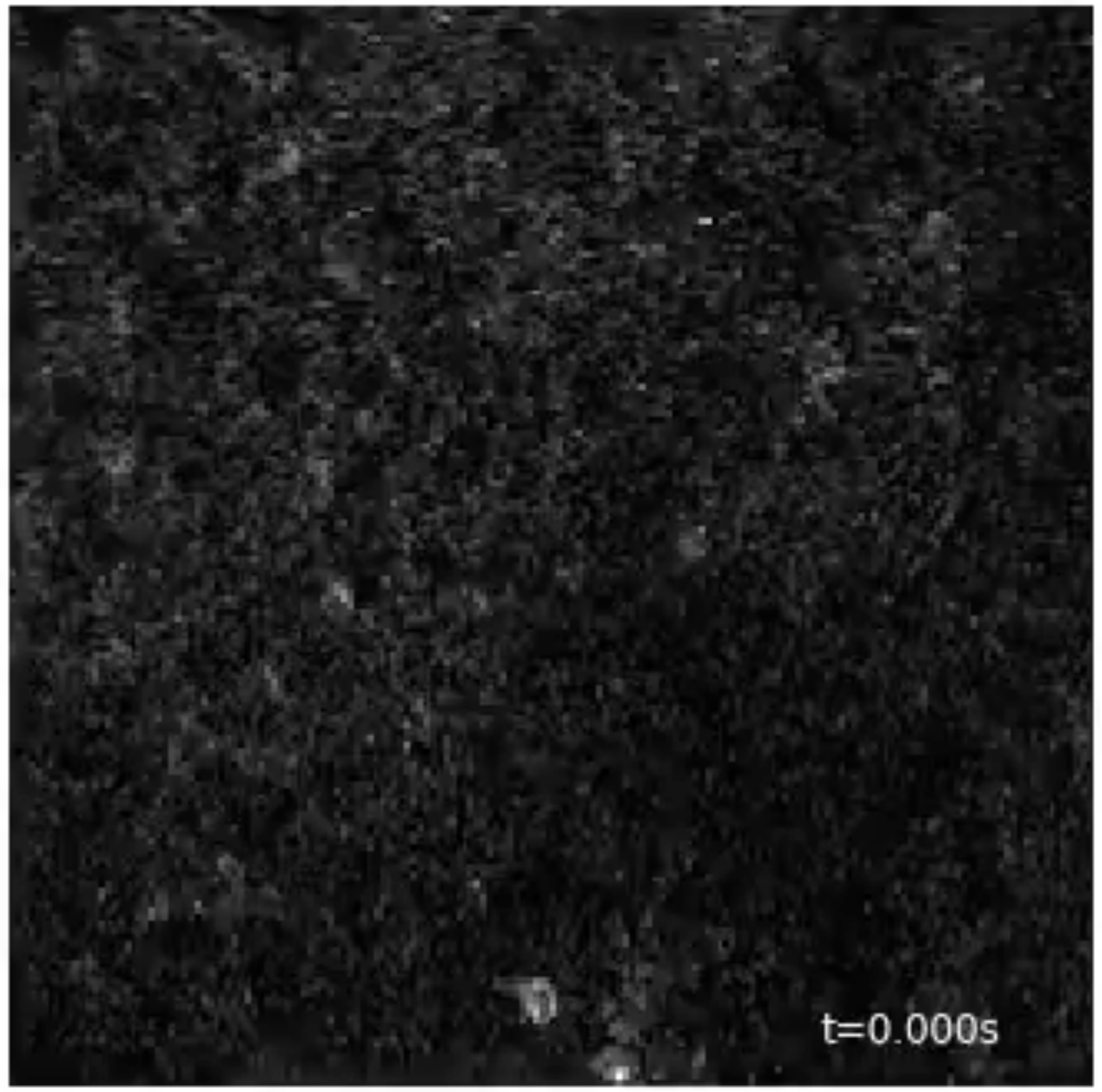


Jun et al, 2017.

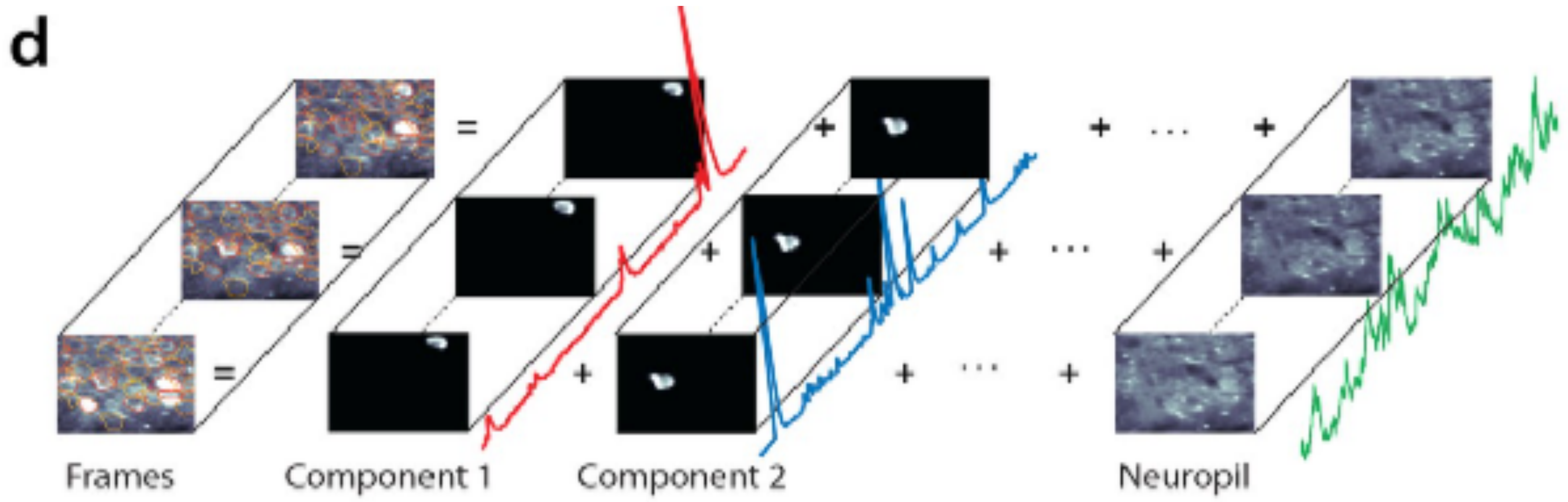
Unit 1: Signal Extraction

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.
- Methods: **matrix factorization, convex optimization**



Data from Sue Ann Koay and David Tank

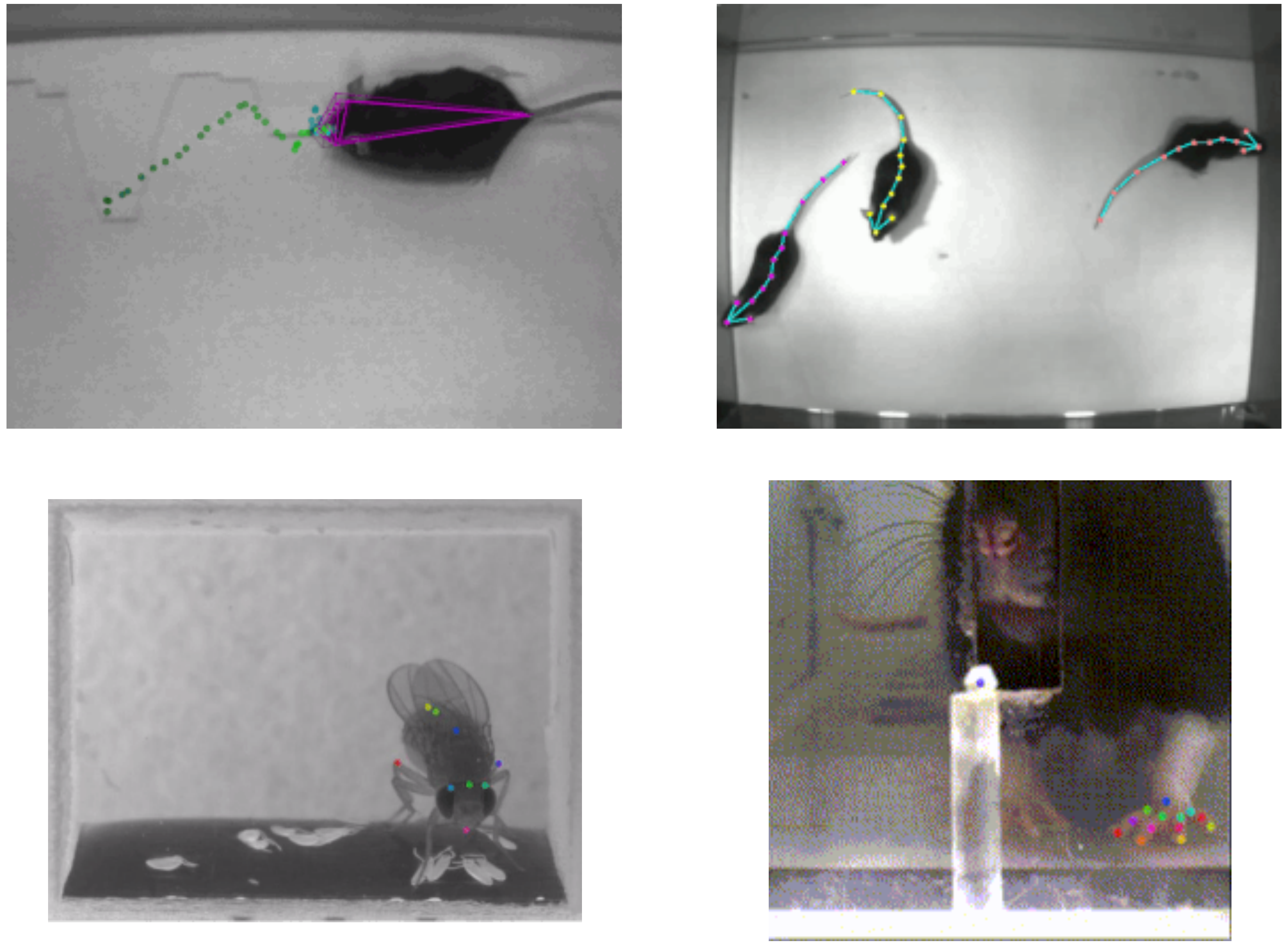


Giovanucci et al (eLife, 2019)

Unit 1: 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.
- <u>Methods</u>: **CNNs**, **transfer learning**

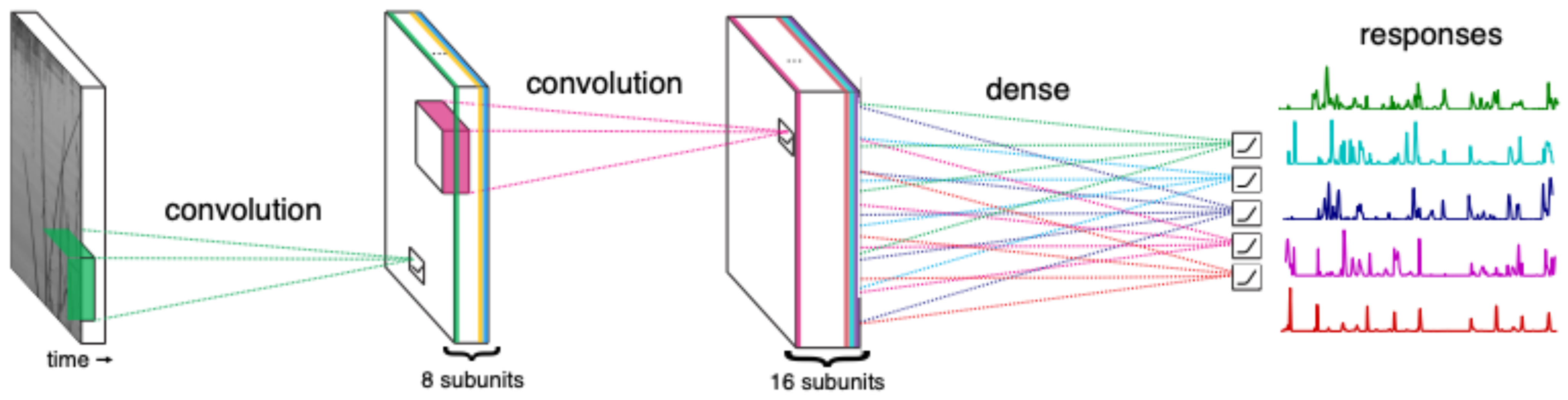


Mathis et al (Nat. Neuro., 2018)
https://github.com/DeepLabCut/DeepLabCut

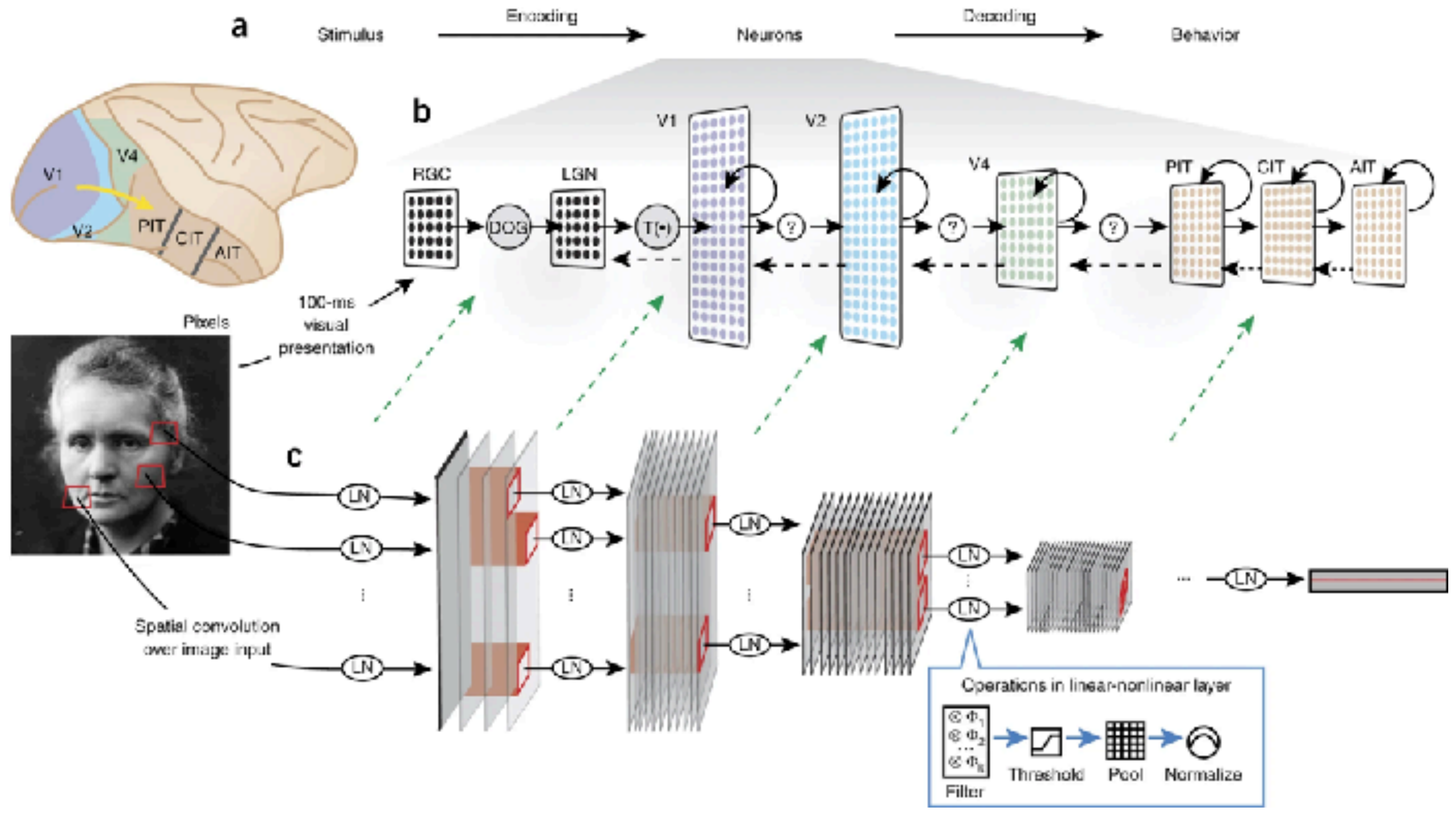
Unit 2: Encoding and Decoding Neural Spike Trains

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.
- <u>Methods</u>: **GLMs, Poisson processes**



McIntosh et al (NeurIPS 2016)

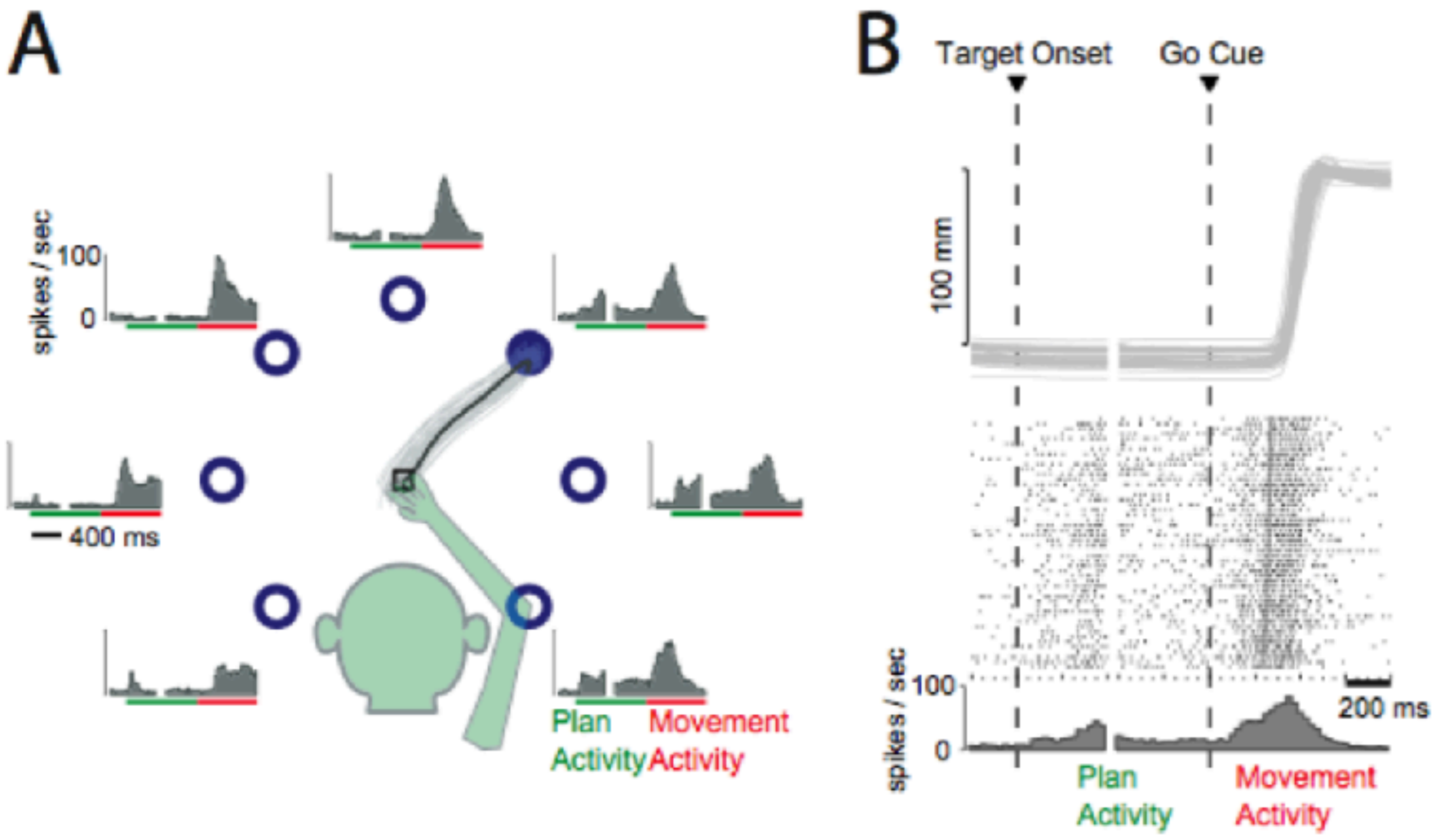


Yamins and DiCarlo (Nat. Neuro. 2016)

Unit 2: 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.**
- <u>Methods</u>: **linear Gaussian models, Bayesian decoders**

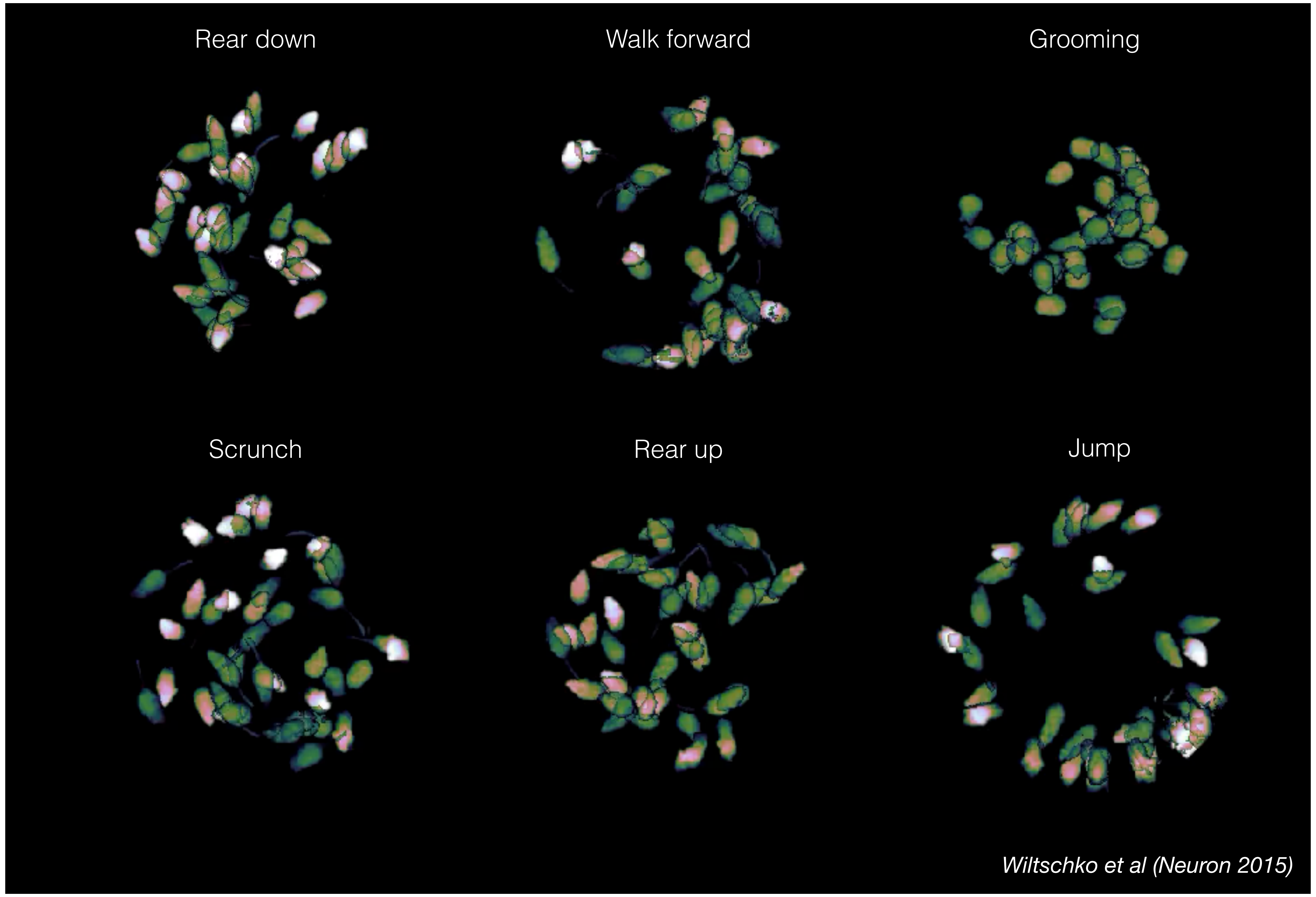


Prof. Krishna Shenoy, EE124

Unit 3: Latent variable models of neural and behavioral data

Summarizing behavior with movement “syllables”

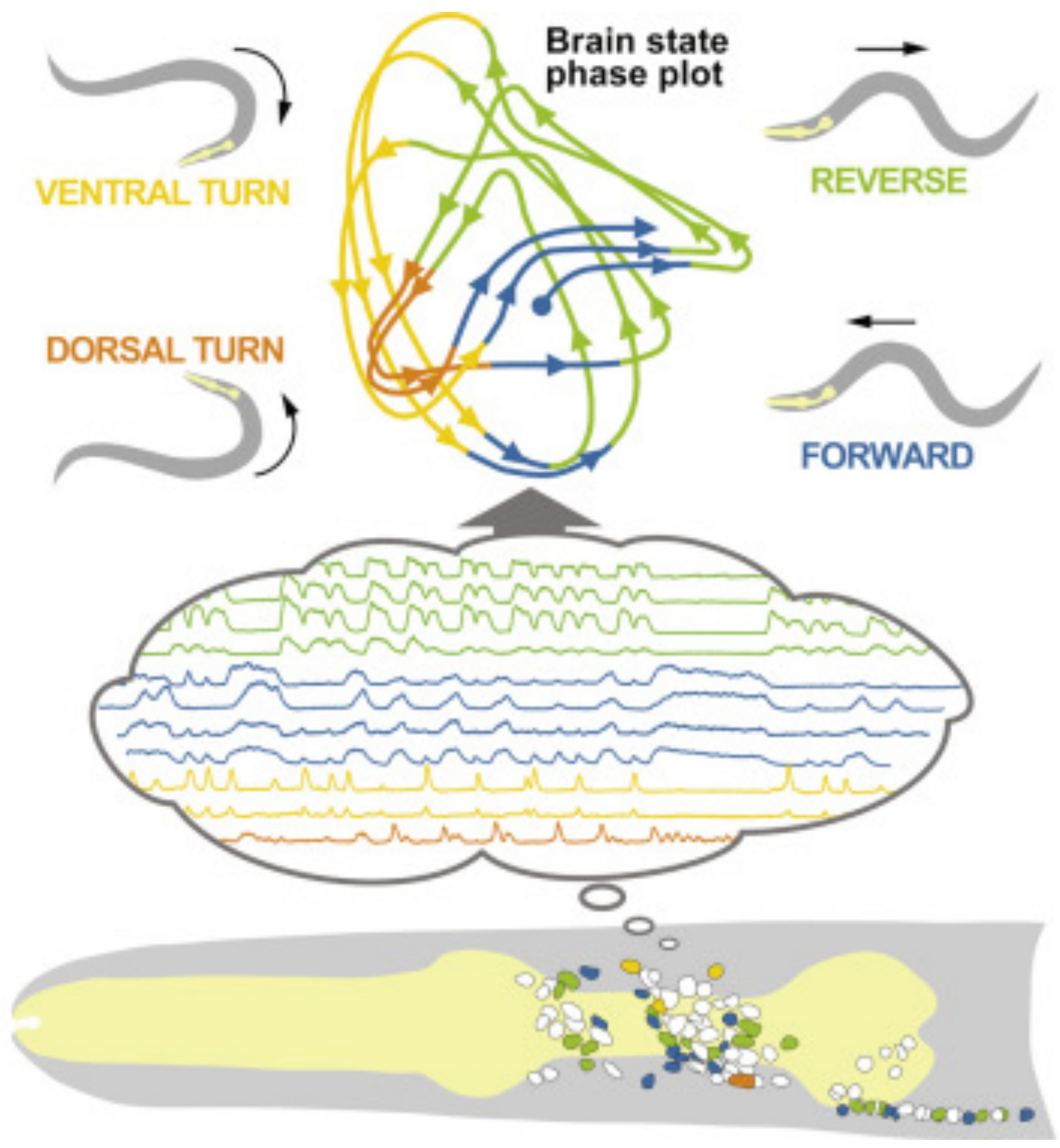
- We can learn a lot about the brain by understanding the **structure of its outputs**.
- Recently, there’s been a “**call to action**” to better characterize animal behavior.
    - Krakauer et al (Neuron, 2017); Datta et al. (Neuron, 2019)
- **Latent variable models** offer a compelling means of **summarizing behavior** in terms of **hidden states**, or “syllables,” of movement.
- We’ll build **autoregressive hidden Markov models** to extract such syllables from video data.
- <u>Methods</u>: **HMMs, forward-backward algorithm**



Unit 3: 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.
- <u>Methods</u>: **SLDS, variational inference**



Kato et al (Cell, 2015)

What *won't* we cover in this course?

- Many recording modalities: EEG, ECoG, fMRI, MEG.
- Complete details of probabilistic modeling and inference.
- Lots of neurobiology. (For that, see Prof. Luo's course in BIO/NBIO.)

Logistics

Website

- The course syllabus is on the website:

https://slinderman.github.io/stats320/

- We will use Ed for announcements and discussion.
- We will use Gradescope for submissions.

Lectures

- Lectures on **Monday/Wednesday, 1:30pm-2:20pm PT**.
- **Slides** will be posted on the website before each lecture.
- Each lecture will have **assigned reading** in the online textbook. The book contains references to research papers for you to dig deeper, if you'd like.

Labs

- You'll work in **teams of two** to **implement a method** from lecture (with lots of **starter code!**) and **apply it to data.**
- Your **weekly assignment** will be to finish the lab with your teammate.
- The catch: *you may not work with the same person twice!*
- Labs will be released on Wednesdays at midnight and due the following Wednesday at midnight.
- We will have office hours on Monday and Tuesday to help you debug your code and answer any questions.

Office hours

- **Scott**: Tuesday 10:30am - 12pm PT in Wu Tsai Neurosciences Institute, 2nd floor Theory Center.
- **Noah**: Monday 10am-12pm in CoDa B40

Final project

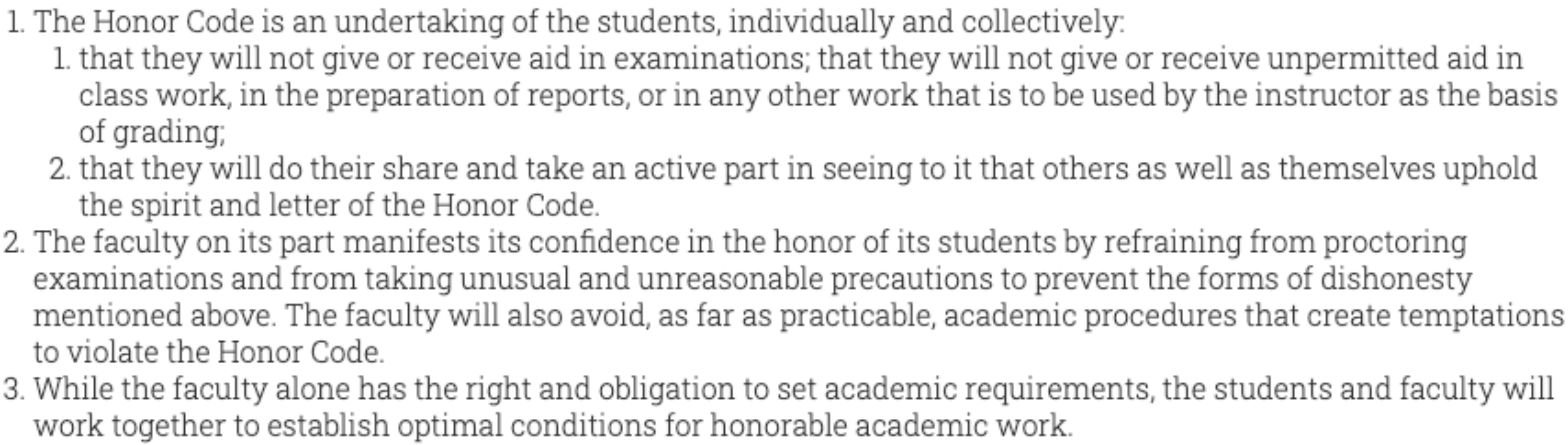
- The final project is an opportunity to **apply what you've learned** to a **problem of interest** to you.
- You'll work in teams of 2, but you can choose your teammate.
- For example, you could:
  - **Implement a method** from a recent research paper and recapitulate its results on synthetic data.
  - Apply methods developed in class to **study a dataset of interest** to you.
  - **Propose and implement an extension** to an existing method that would address some of its limitations.
- You'll submit a proposal partway through the course and a final report + code at the end.

Grading

7 Labs	7 x 10% each = 70% total
Final project	25%
Class participation	5%

- We will give more information on how final projects will be assessed later in the course.

Honor Code


1. The Honor Code is an undertaking of the students, individually and collectively:
   1. that they will not give or receive aid in examinations; that they will not give or receive unpermitted aid in class work, in the preparation of reports, or in any other work that is to be used by the instructor as the basis of grading;
   2. that they will do their share and take an active part in seeing to it that others as well as themselves uphold the spirit and letter of the Honor Code.
2. The faculty on its part manifests its confidence in the honor of its students by refraining from proctoring examinations and from taking unusual and unreasonable precautions to prevent the forms of dishonesty mentioned above. The faculty will also avoid, as far as practicable, academic procedures that create temptations to violate the Honor Code.
3. While the faculty alone has the right and obligation to set academic requirements, the students and faculty will work together to establish optimal conditions for honorable academic work.

https://communitystandards.stanford.edu/policies-and-guidance/honor-code

Questions?

Survey

Again, please take this short survey:

https://tinyurl.com/stats320survey