High-dimensional data and problems in neuroscience

Cartoon by Jordan Collver

Ian Stevenson

stevenson.lab.uconn.edu
@ihstevenson

Overview

Tour of the backyard
what questions? what data?

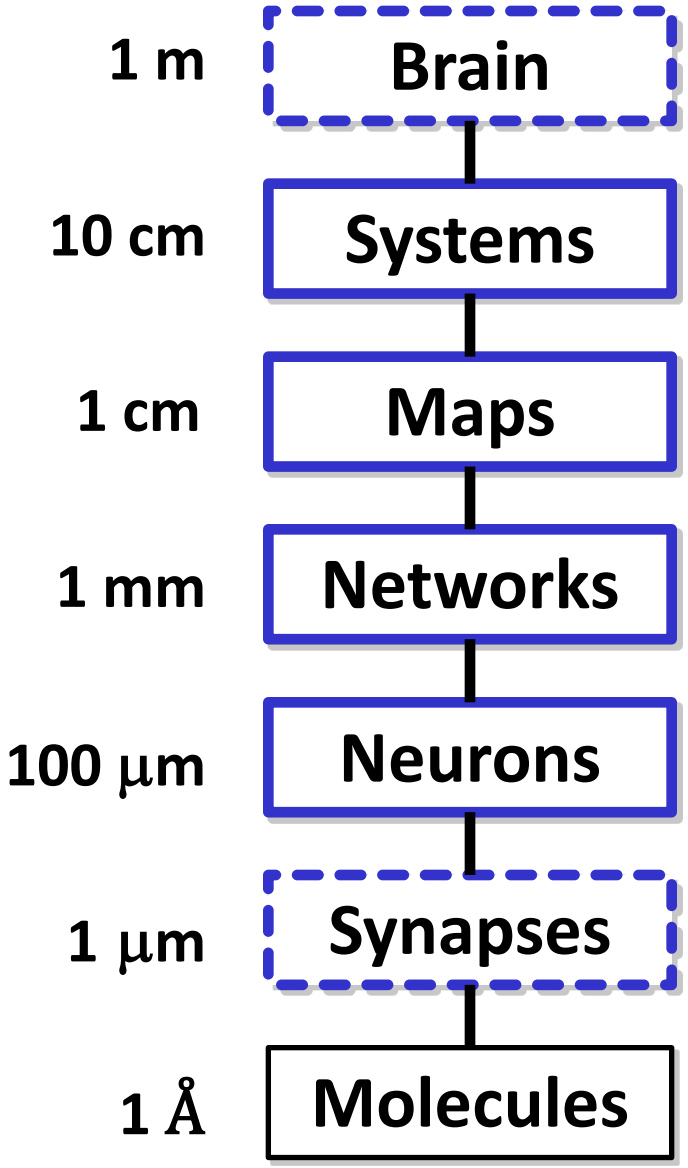
High dimensional regression with natural stimuli
a stereotypical(?) curse of dimensionality situation

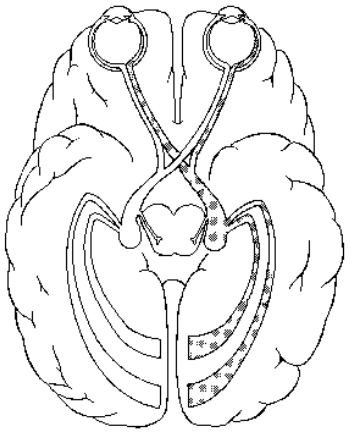
Dimensionality reduction for neural populations
non-Gaussian observations and dynamics

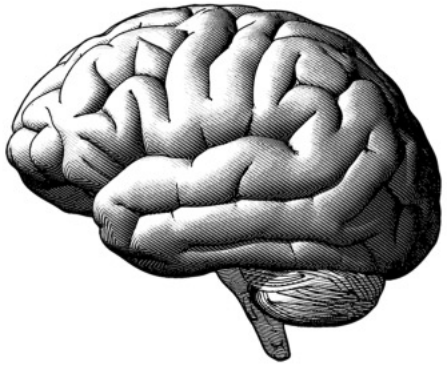
How does the brain handle high dimensional inputs?
random projections for odor perception

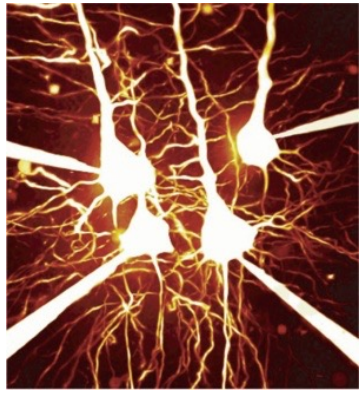
“The human brain consists of **100 billion neurons**, 100 trillion synapses, connected by hundreds of thousands of miles of axons and dendrites …and we in neuroscience call that job security.”

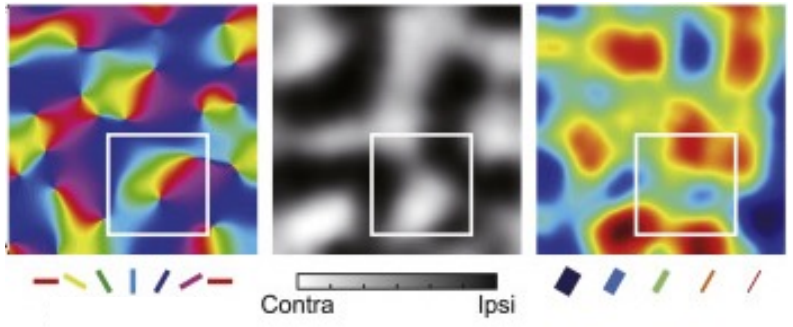
- Chuck Stevens

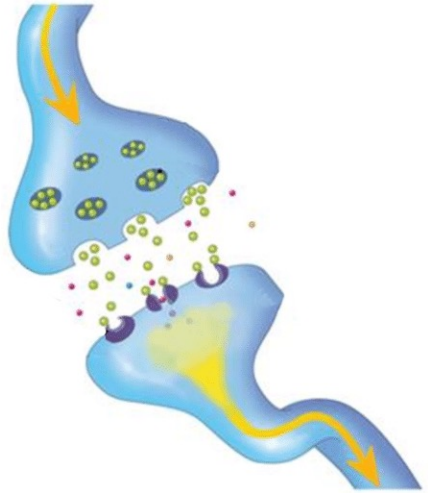

1 m
Brain
10 cm
Systems
1 cm
Maps
1 mm
Networks
100 μm
Neurons
1 μm
Synapses
1 Å
Molecules

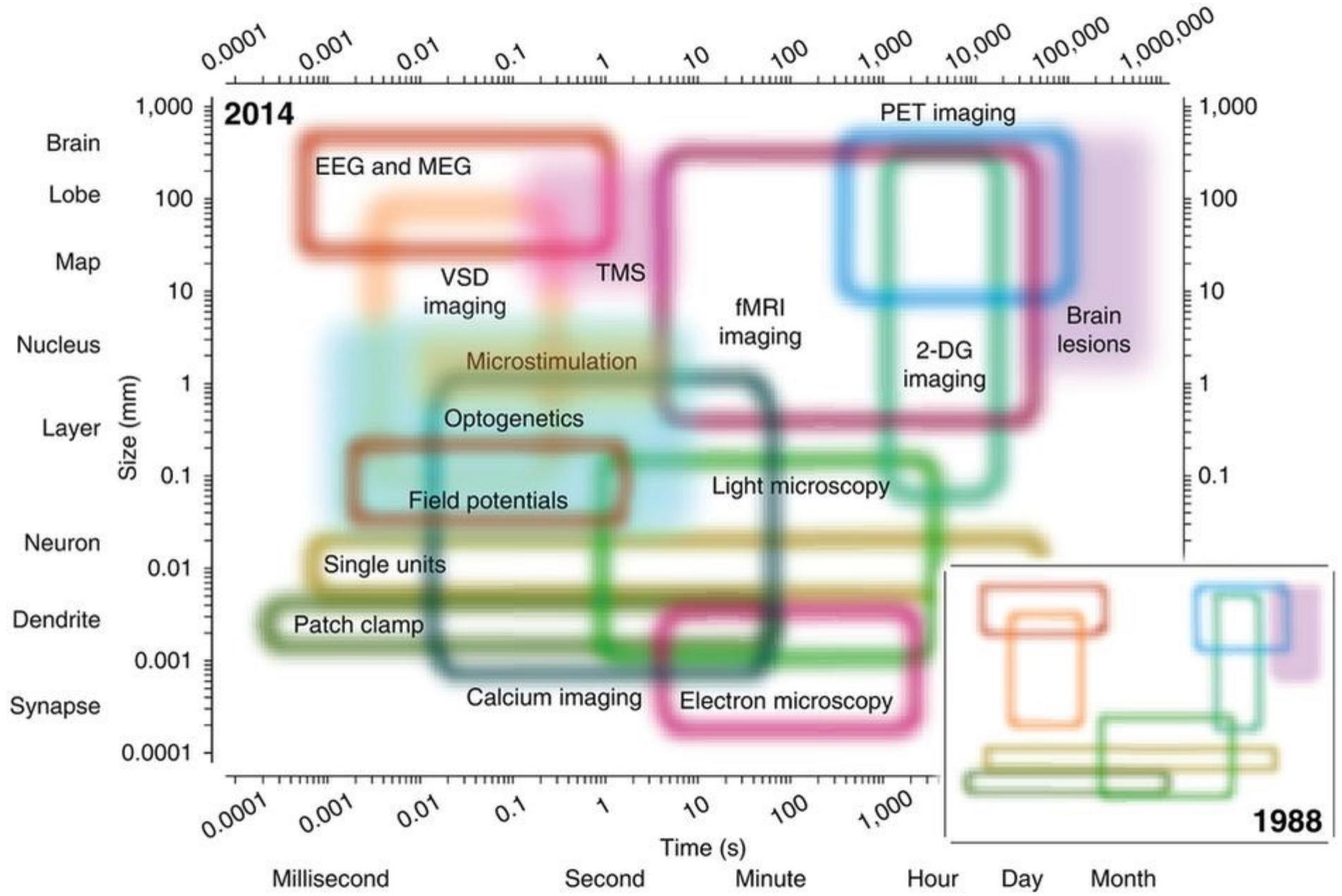




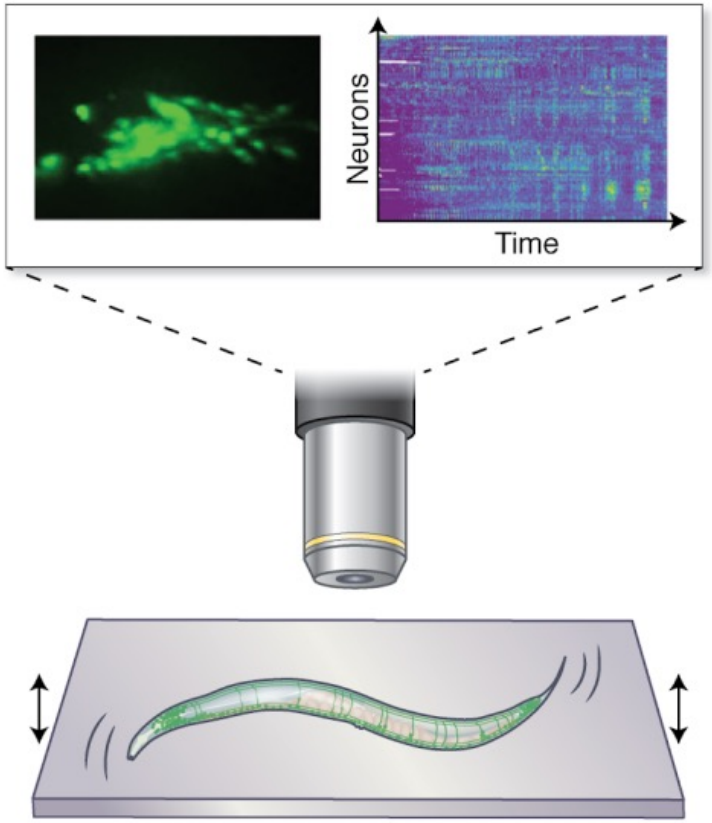


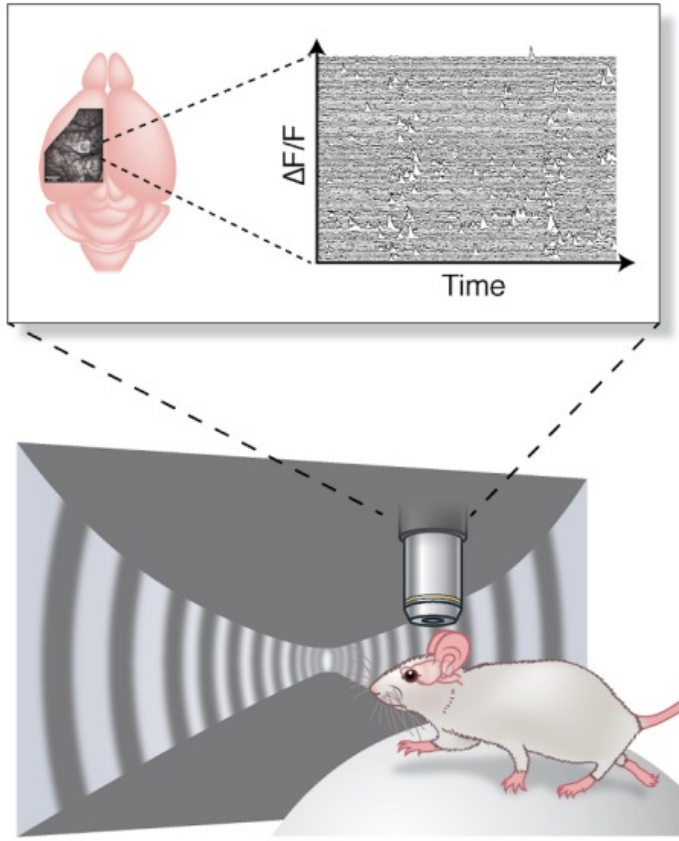

Contra
Ipsi

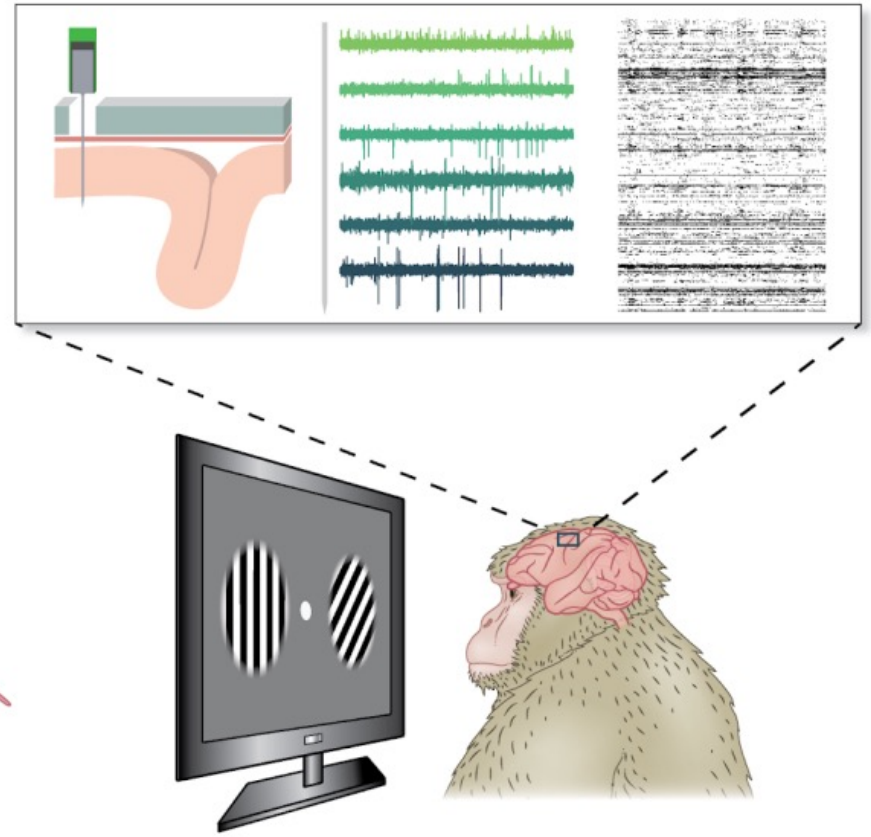


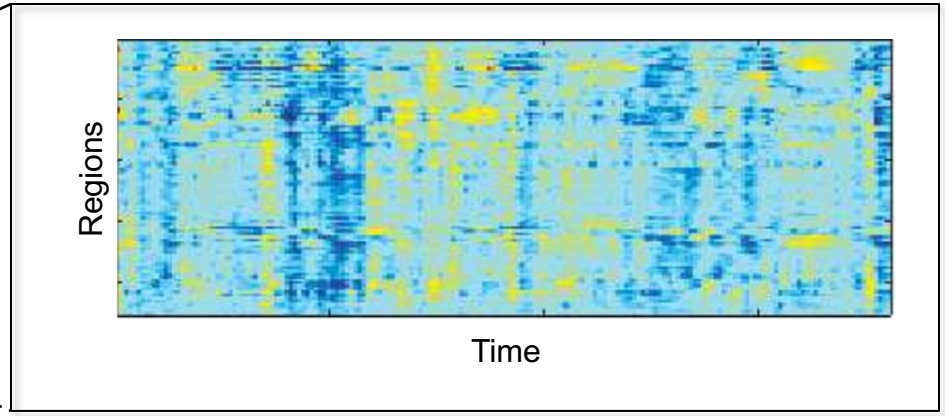


Sejnowski et al. (2014)



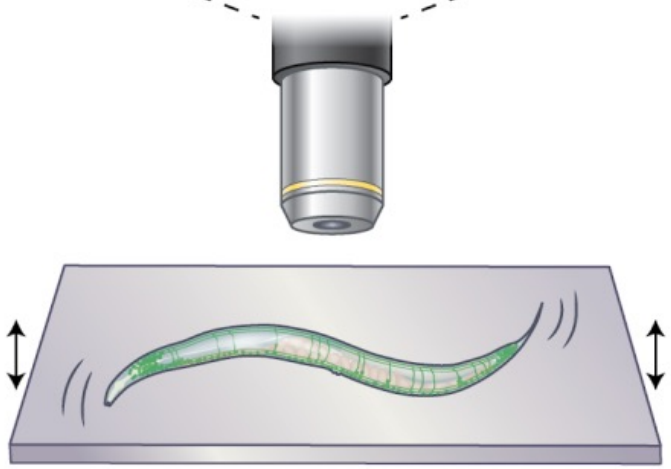




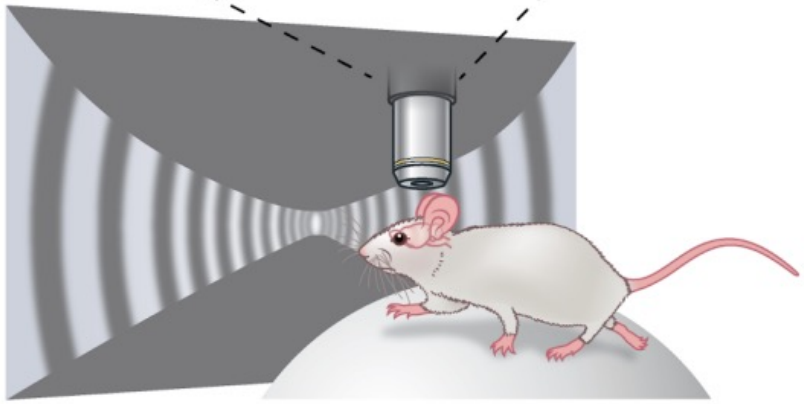


Based on Urai et al. 2022

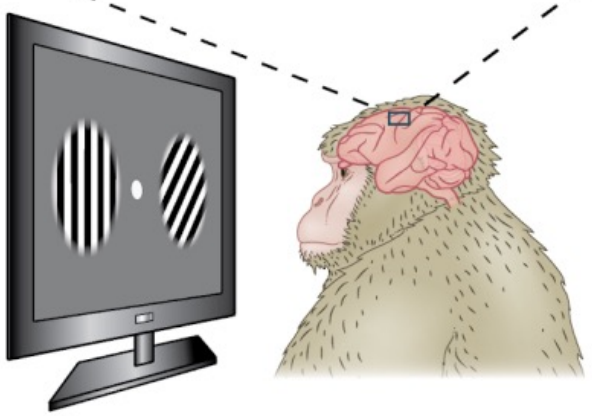
How do motor neurons coordinate for efficient locomotion?



How do neurons represent the mouse's location in space?



How do neurons in frontal areas alter neurons in sensory areas to control attention?

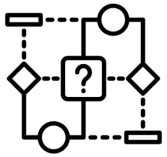


What parts of the brain are involved in learning to read?

Marr's Levels of Analysis

a framework for neuroscientific questions

Computation What does the system do?

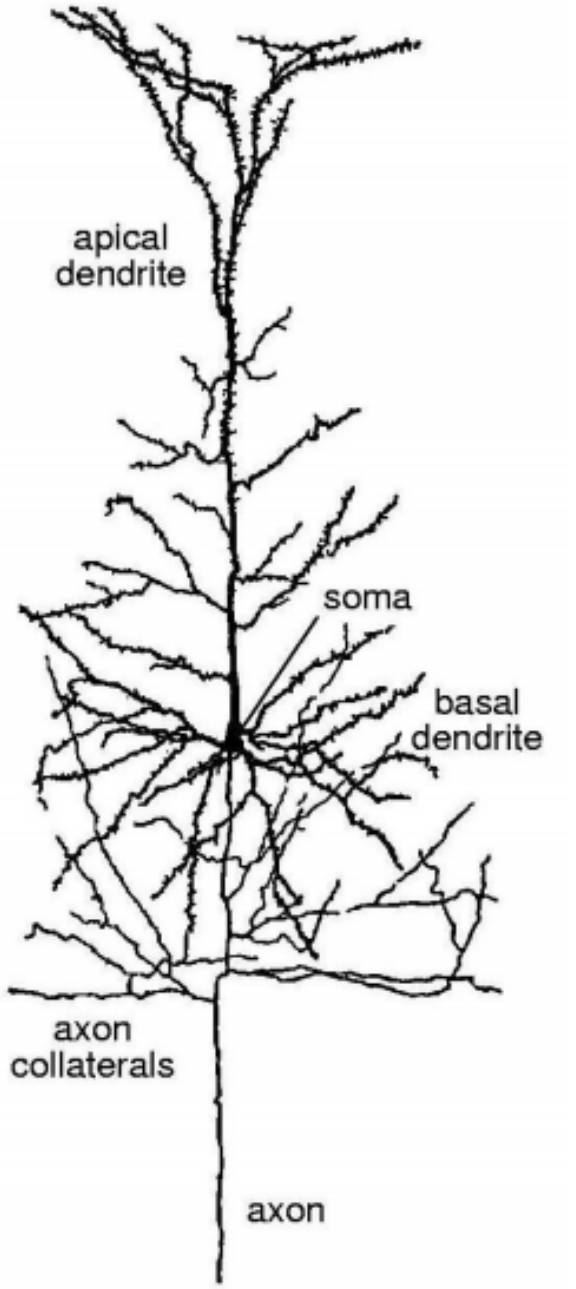


Algorithm How does the system do it?

Implementation How is the system realized?

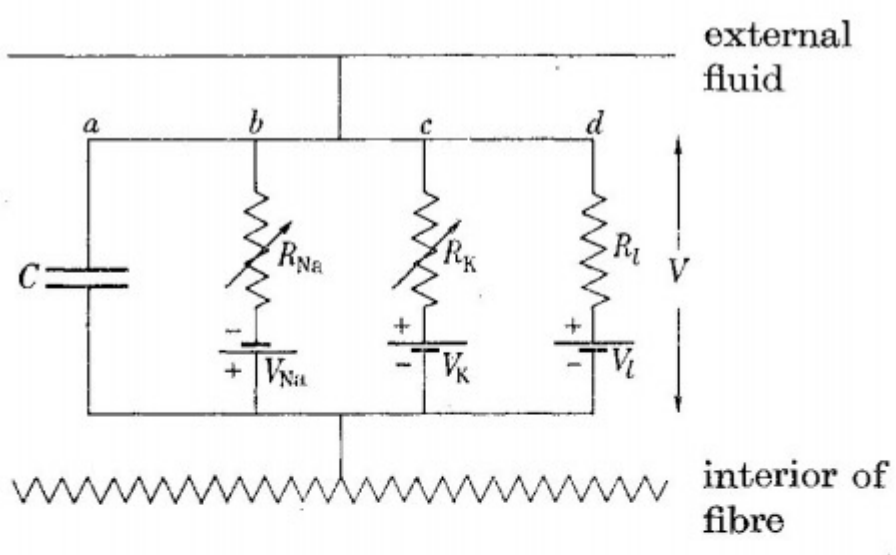
Follow-up Questions…

How do things differ in drug/disease states and can we fix it? (**medicine**)
How can we build systems that interact with the brain? (**neural engineering**)
How can we use this info for artificial systems? (**neuroAI**)

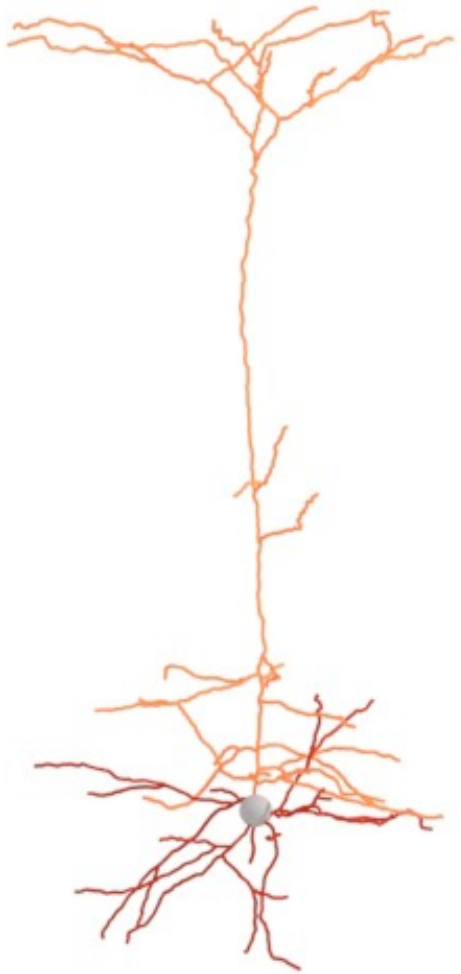


How do single neurons work?

main internal signal is membrane potential



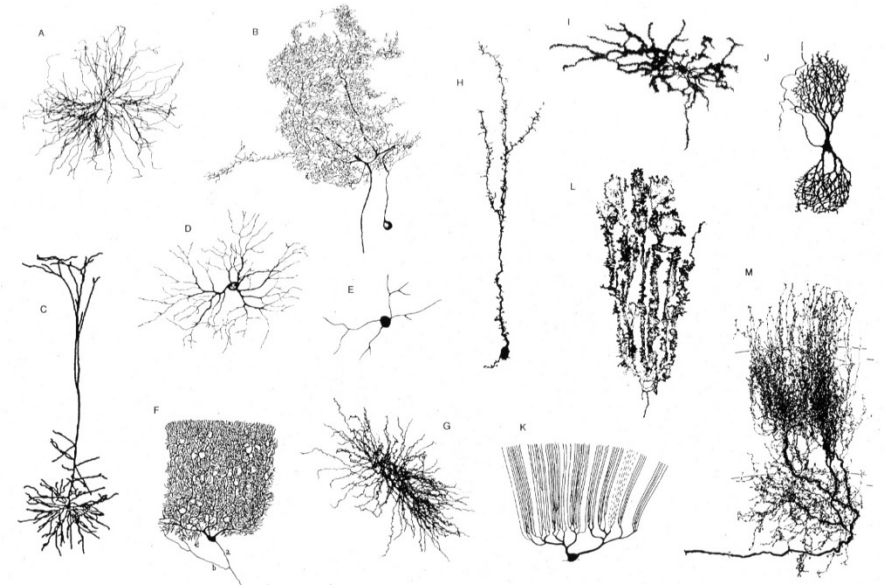
$$C\frac{dV}{dt} = -g_L(V - V_L) - I_{Na}(V) - I_K(V) - g_E(V - V_E) - g_I(V - V_I) + I_{stim}$$

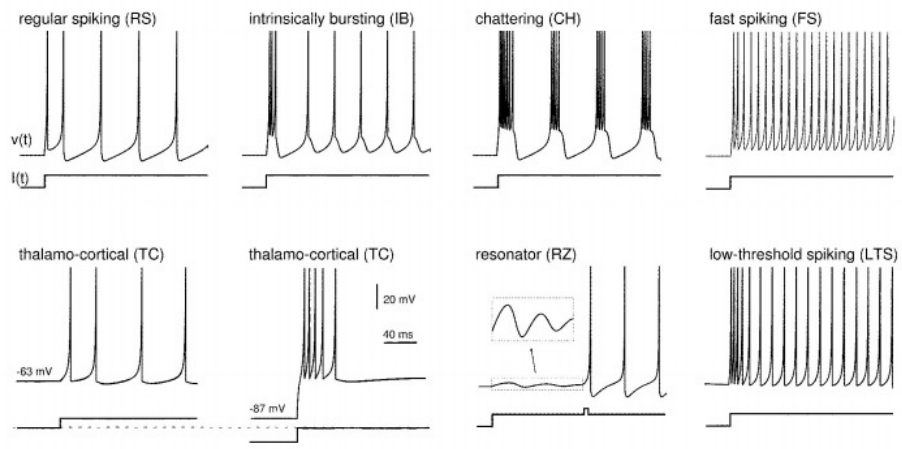


stimulus

response

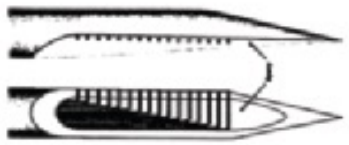
model





Allen Institute Cell Types Project

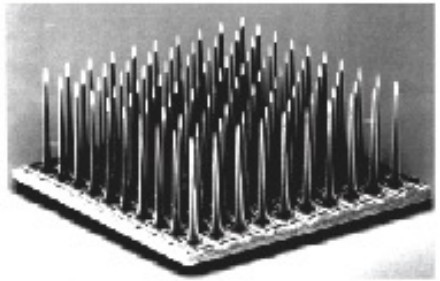
Single-Electrode



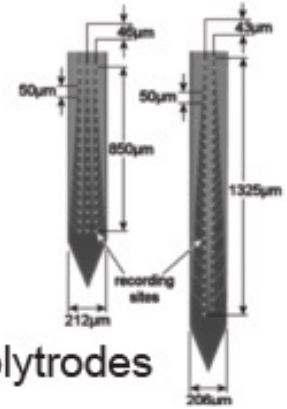
Multi-Electrode

1960 1970 1980

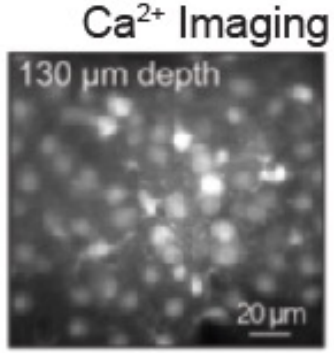
1990 2000 2010 2020

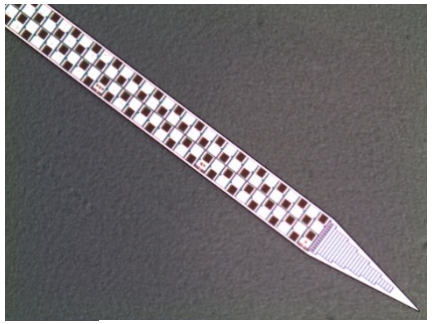


Utah Array

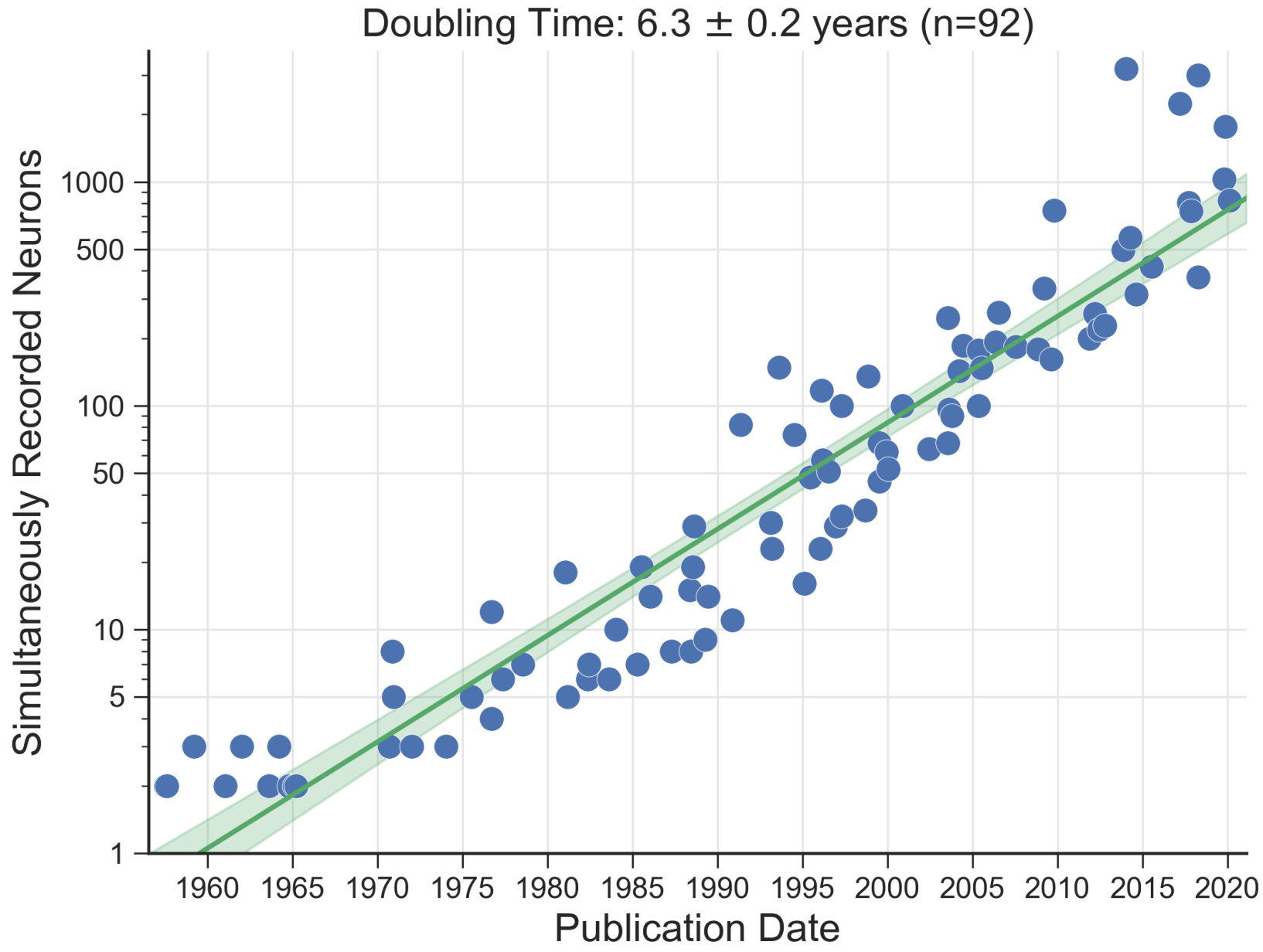


Polytrodes





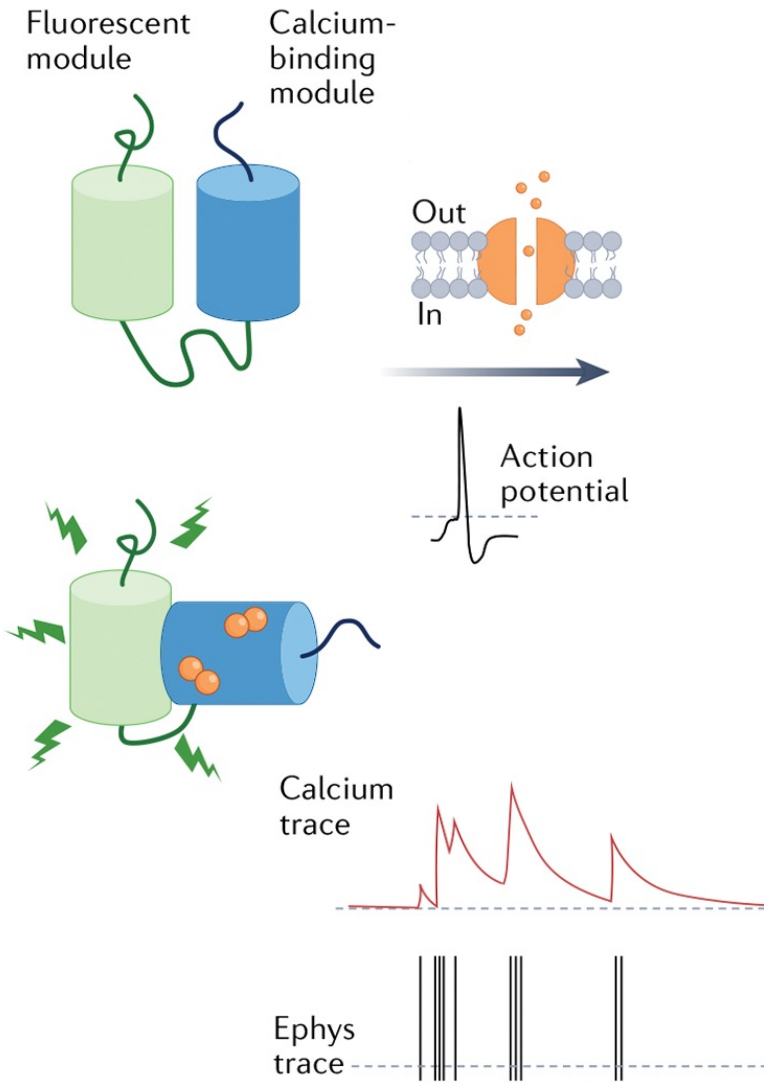
Neuropixels

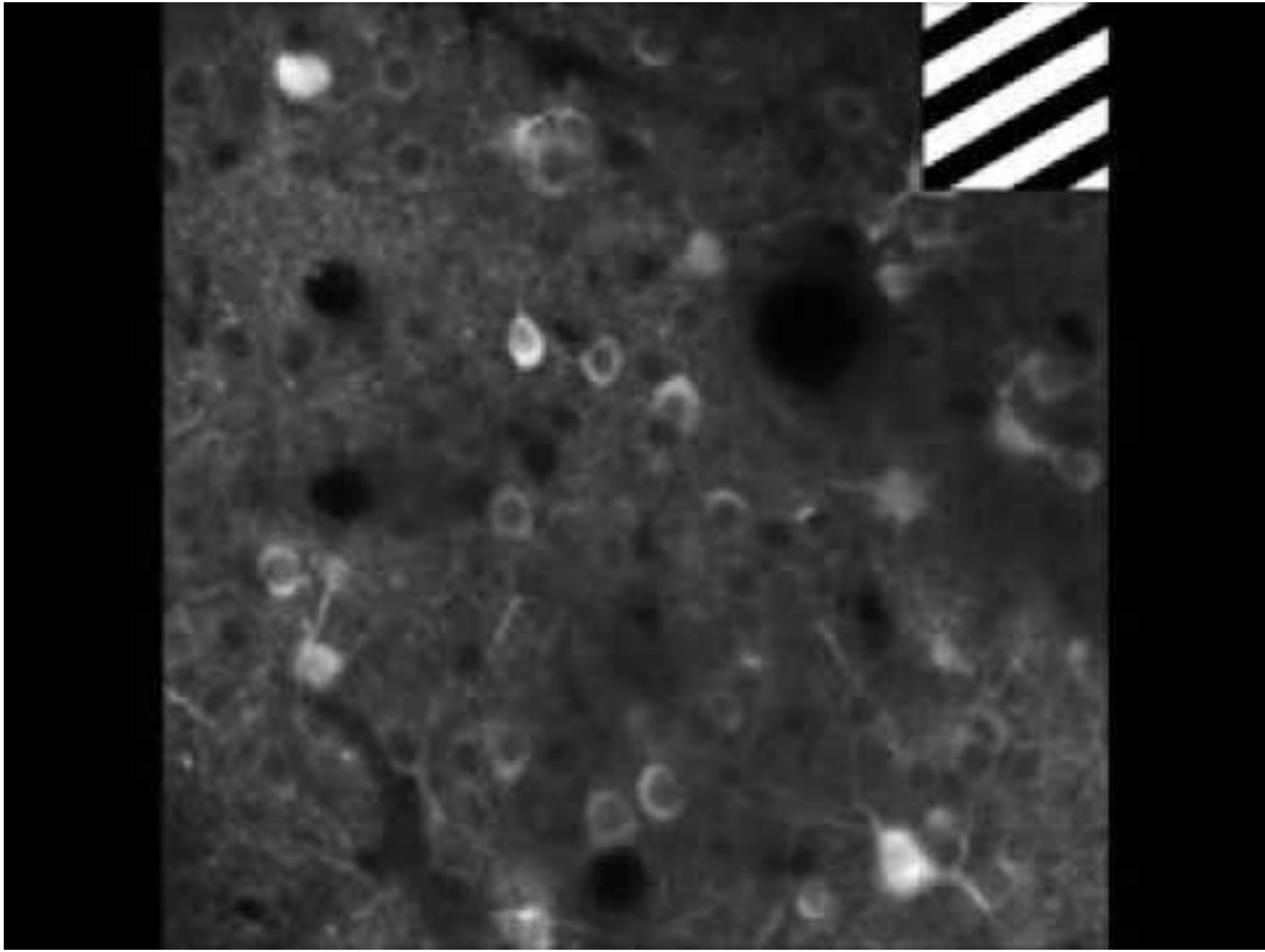


Stevenson and Kording, *Nature Neuroscience* (2011)

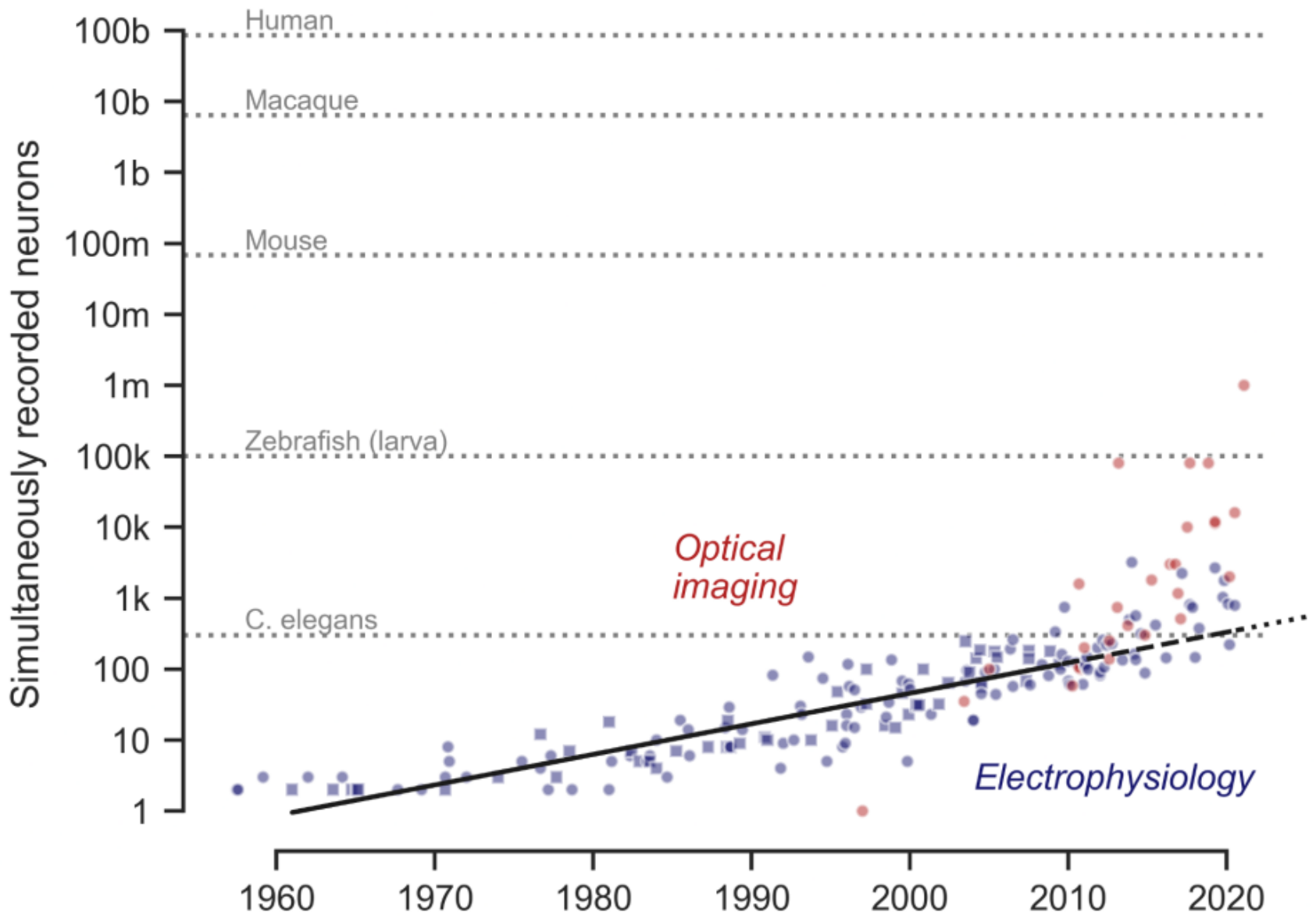
data & figure at stevenson.lab.uconn.edu

Calcium Imaging





Grienberger et al. 2022, Sur Lab



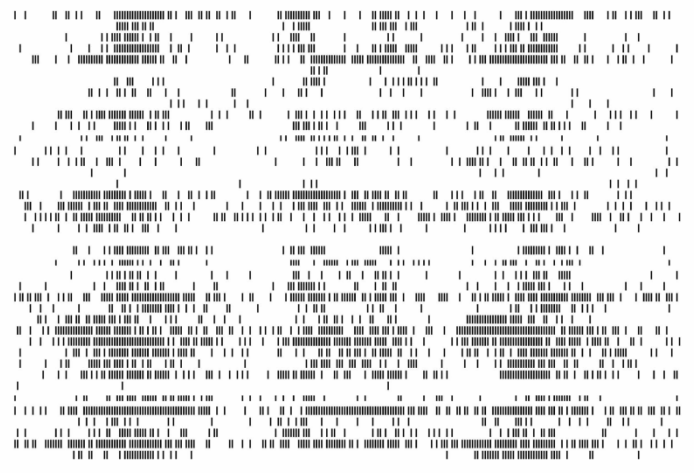
Urai et al. 2022

The Neural Coding Problem

Encoding

What features of the world explain neural activity? And how?

Stimuli and **Movement**

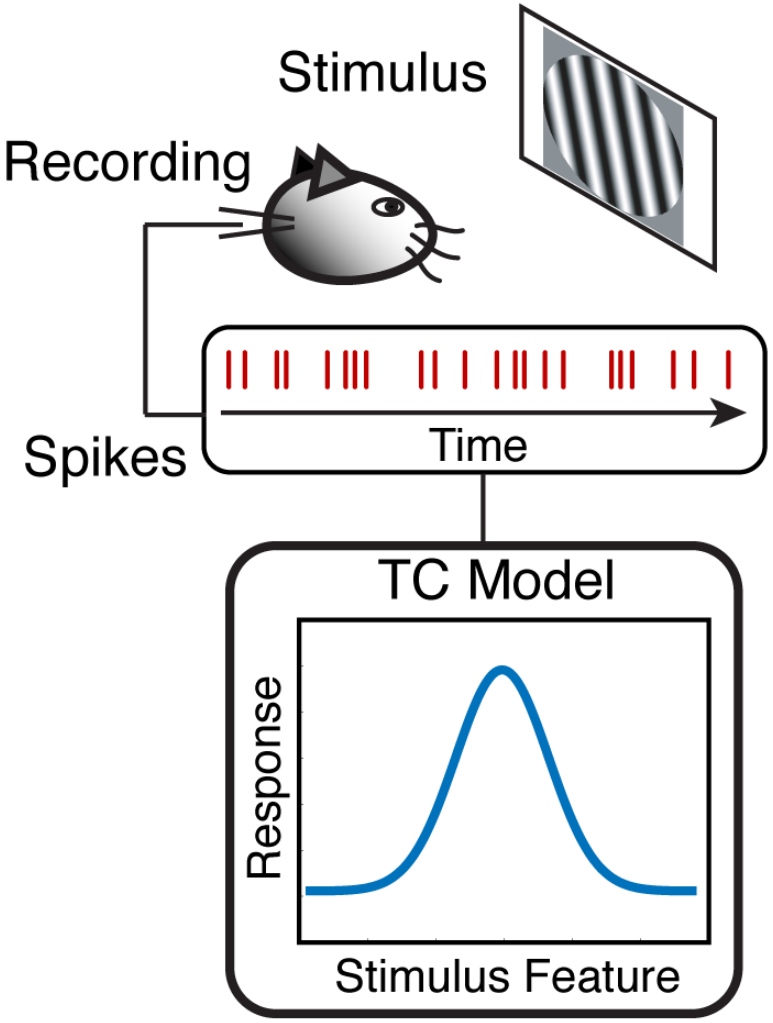


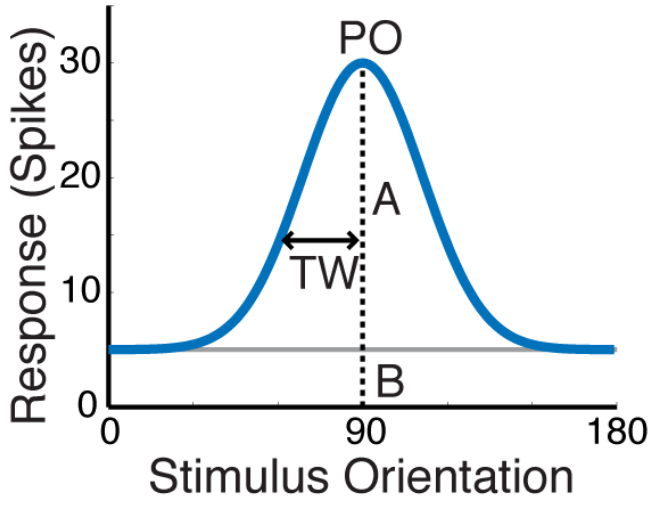
Neural Signals

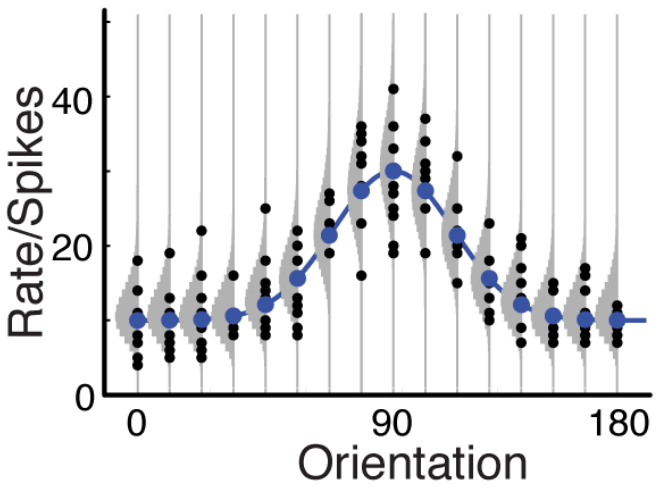
Decoding

How well can neural signals predict variables in the world?

Single neuron responses to low dimensional **stimuli**







stimulus

$$\lambda_i = f_\theta(X_i)$$

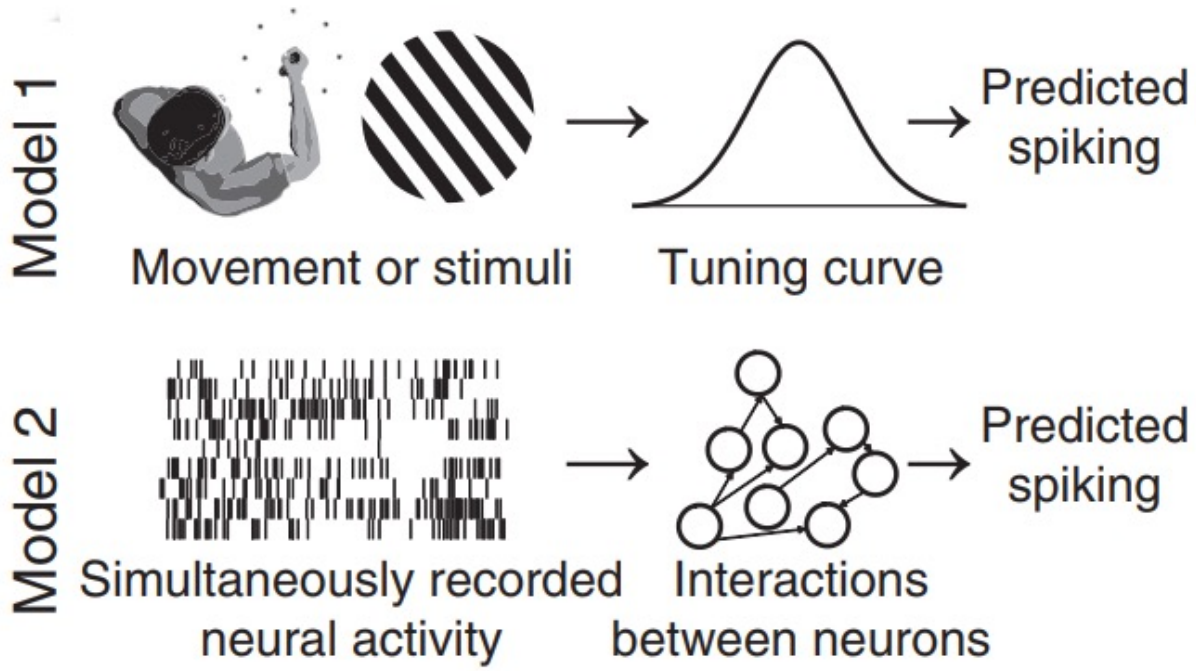
$$y_i \sim Poisson(\lambda_i)$$

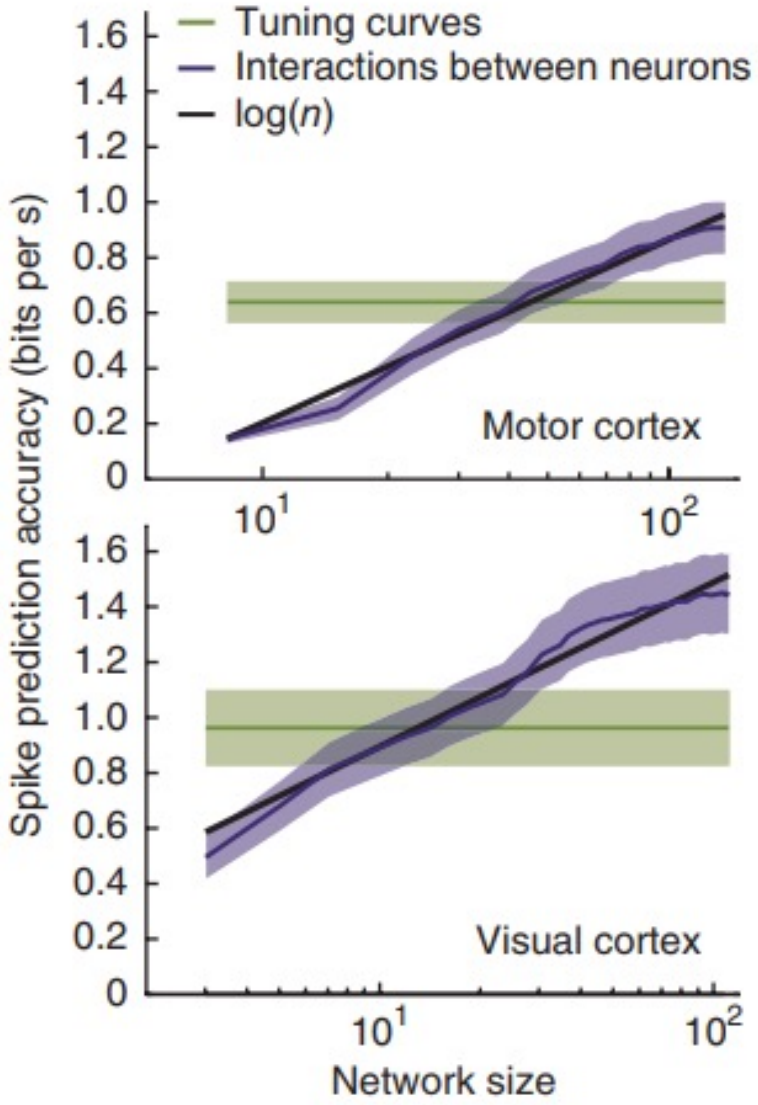
response

Cronin*, Stevenson*, Sur, and Kording. *J Neurophys* (2010)

Scaling in *d*…

some kind of blessing?

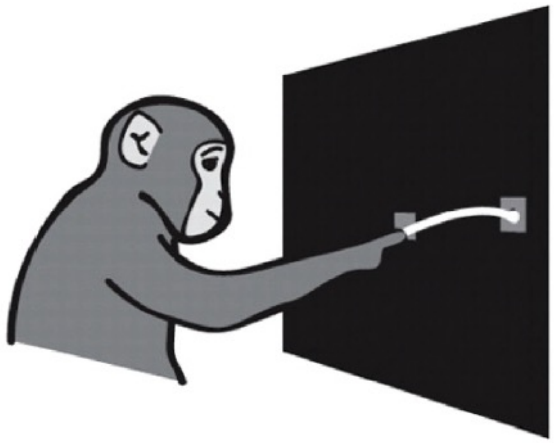


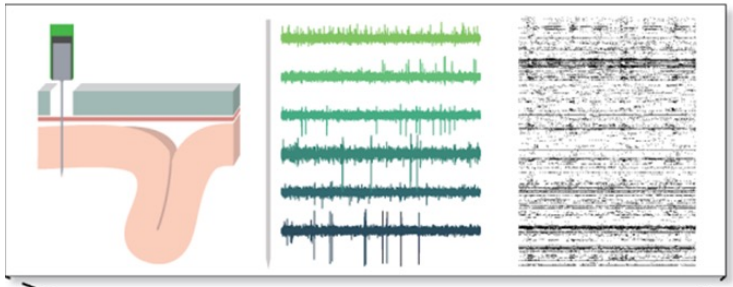


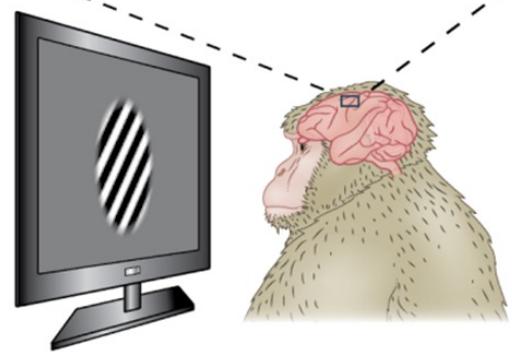
Stevenson and Kording, Nature Neuroscience. 2011

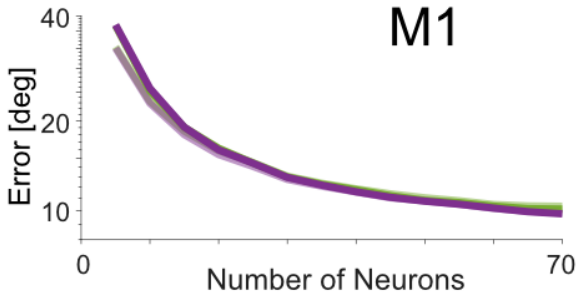
Scaling in *d*…

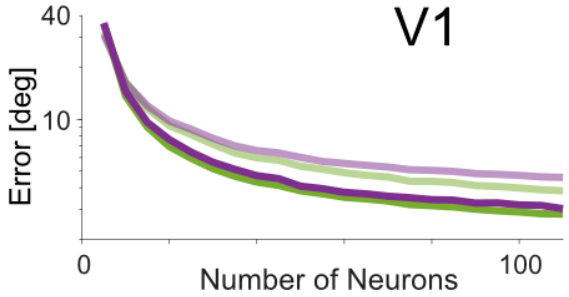
some kind of blessing? curse?

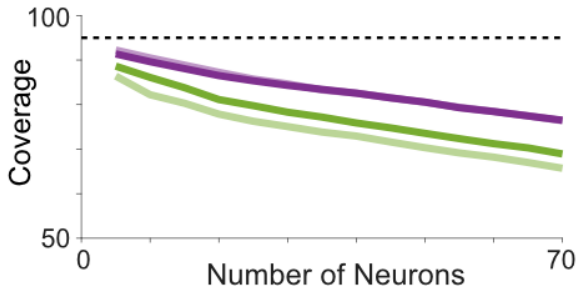


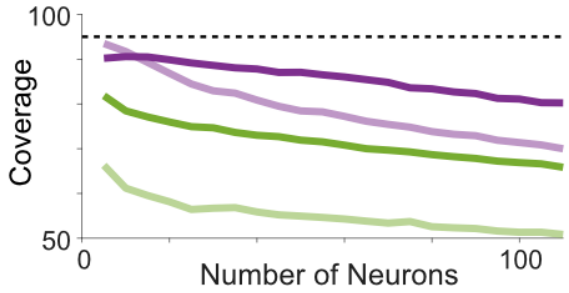








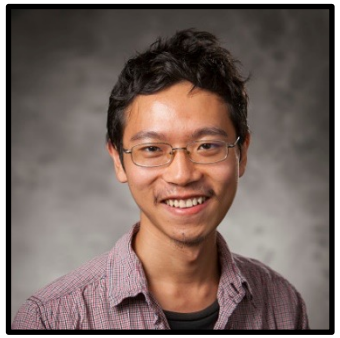




Ganchao's Model

(NBGLLVM)

Old Model

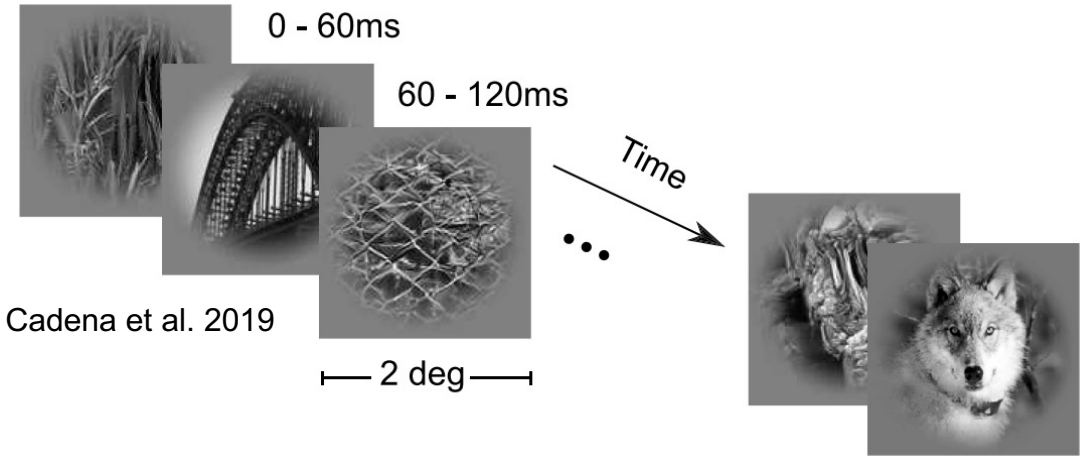


Ganchao Wei, PhD

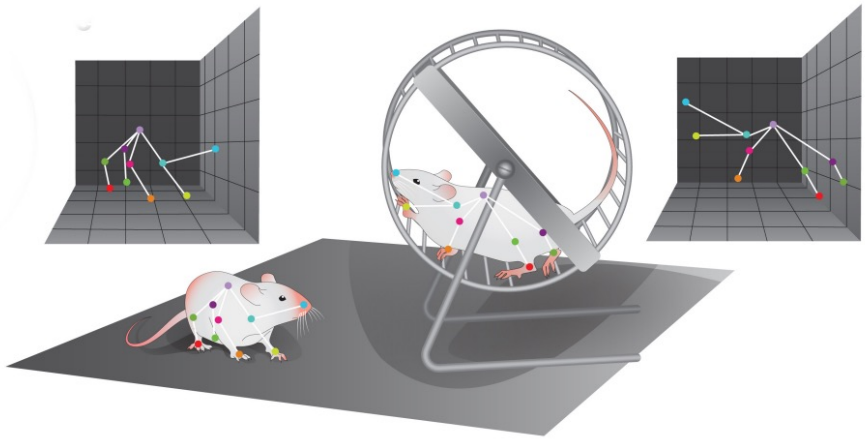
UConn Stats→ Duke

Wei et al. J Neurosci. 2024

Single neuron responses to the natural world

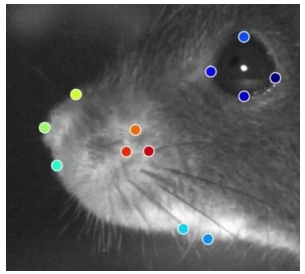


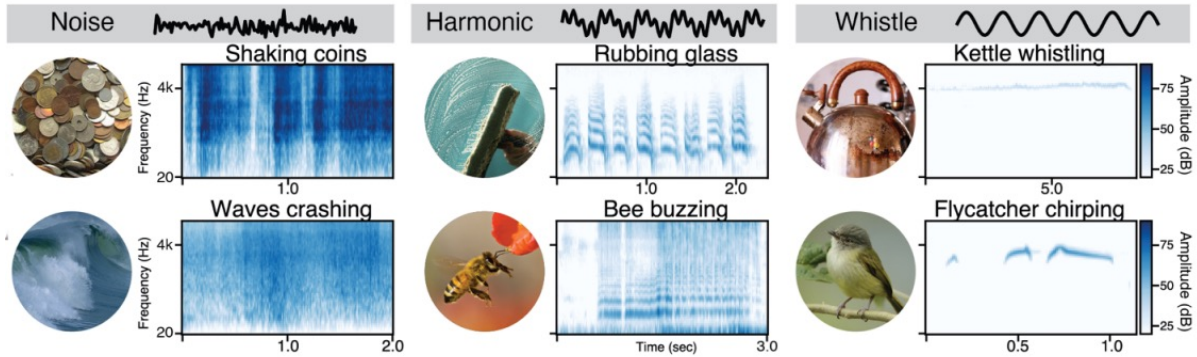
Cadena et al. 2019



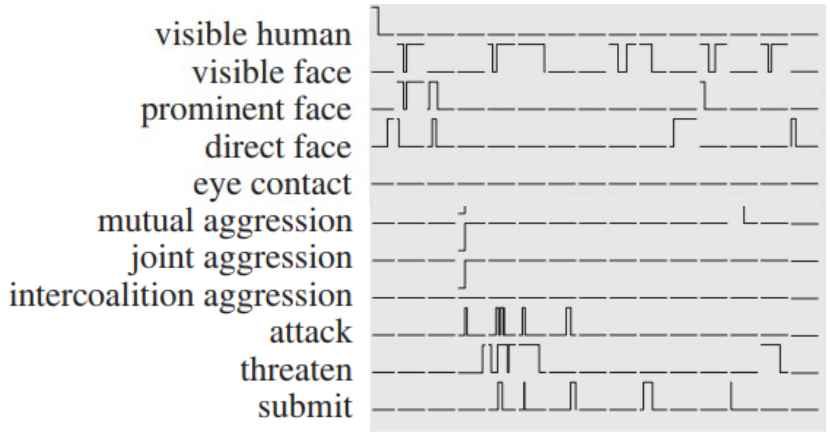
Mathis and Mathis, 2019

Syeda et al. 2024

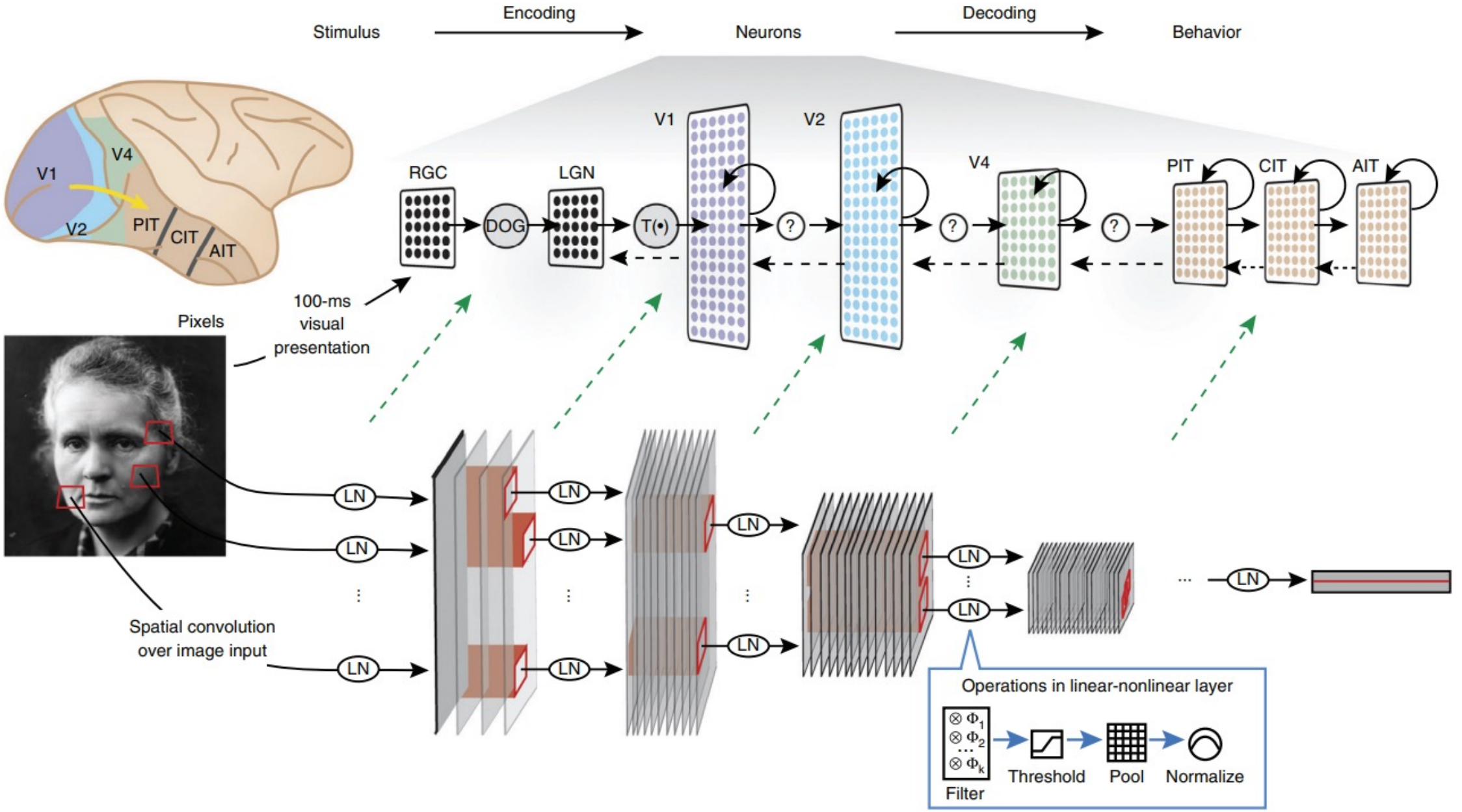




Cusimano et al. 2023

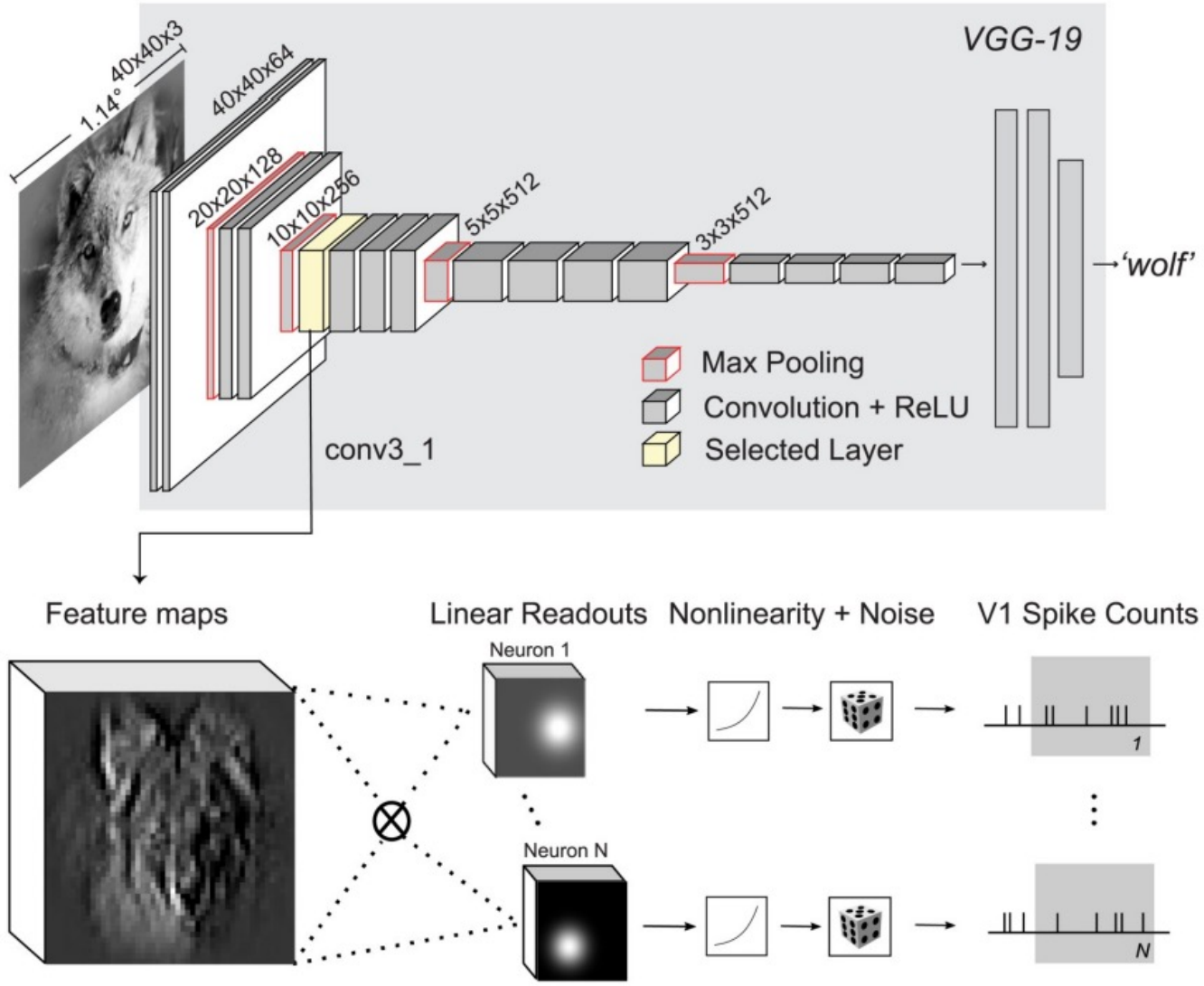


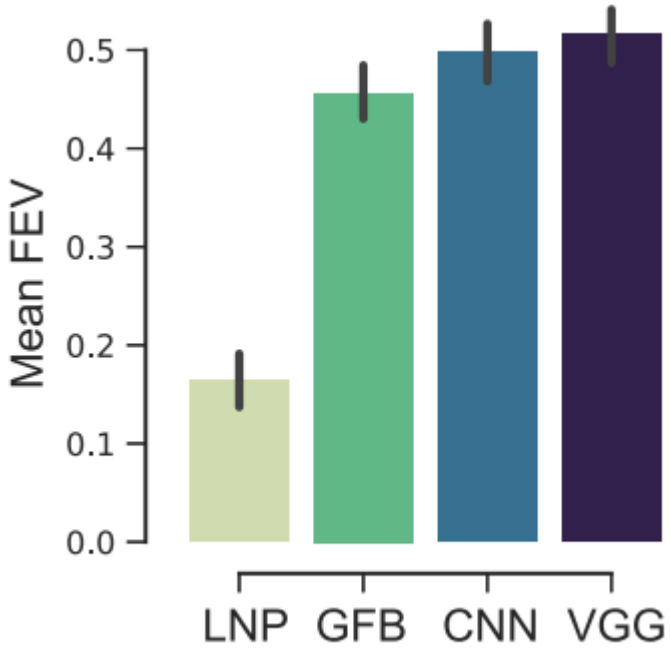
Adams et al. 2020



Yamins et al. 2014

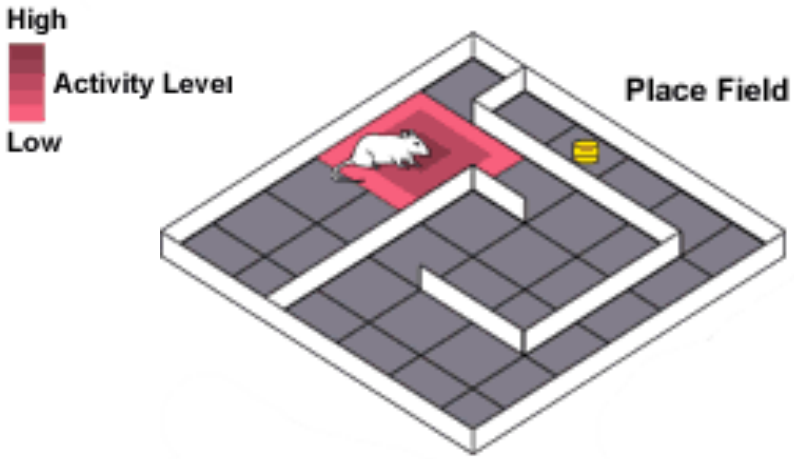
What image features are V1 neurons sensitive to?

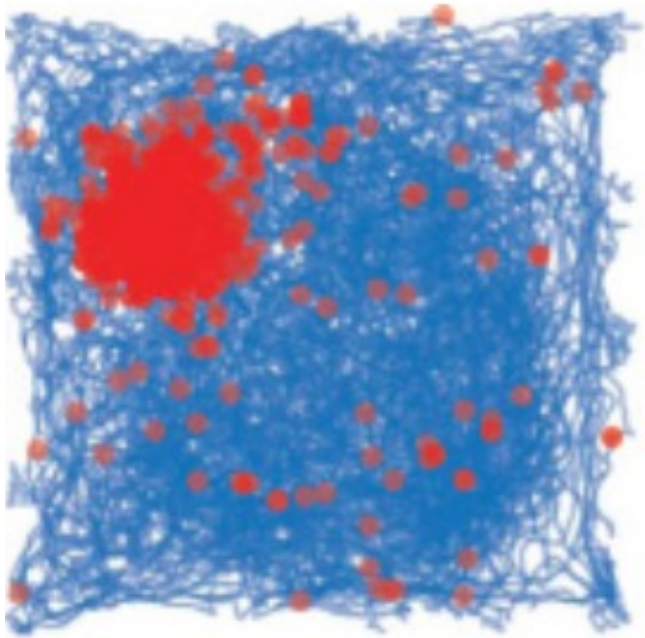


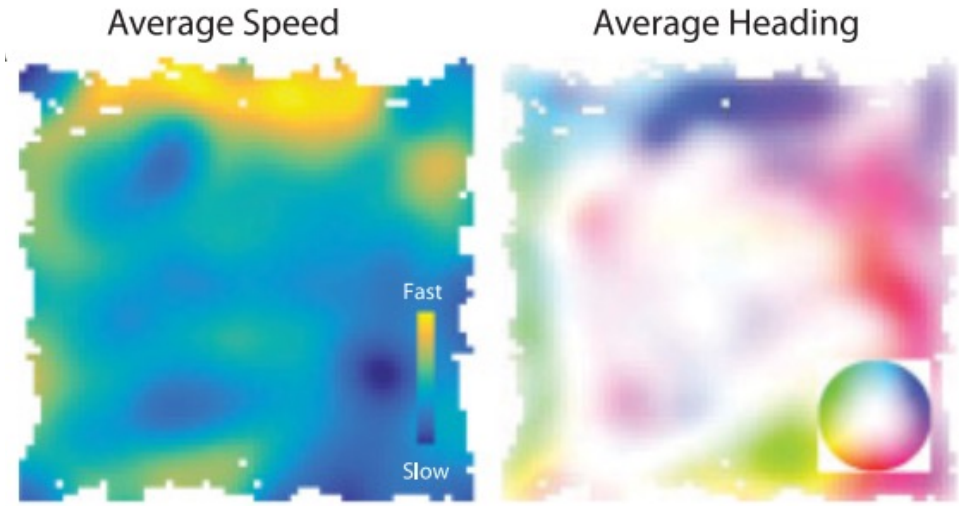


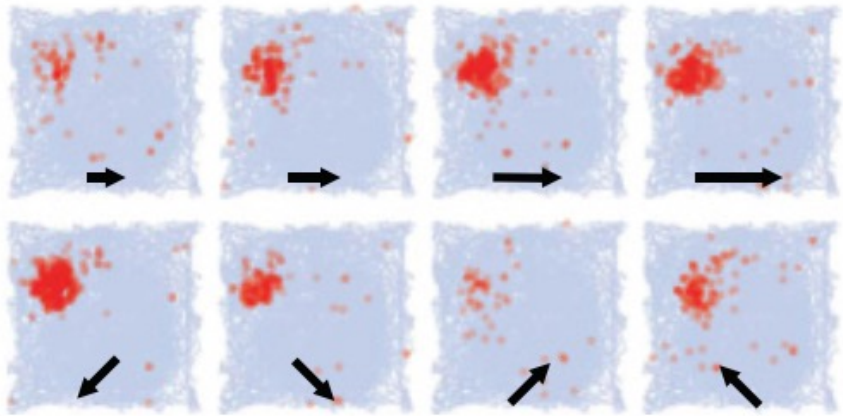
Cadena et al. 2019

Omitted variable bias is potentially a major problem



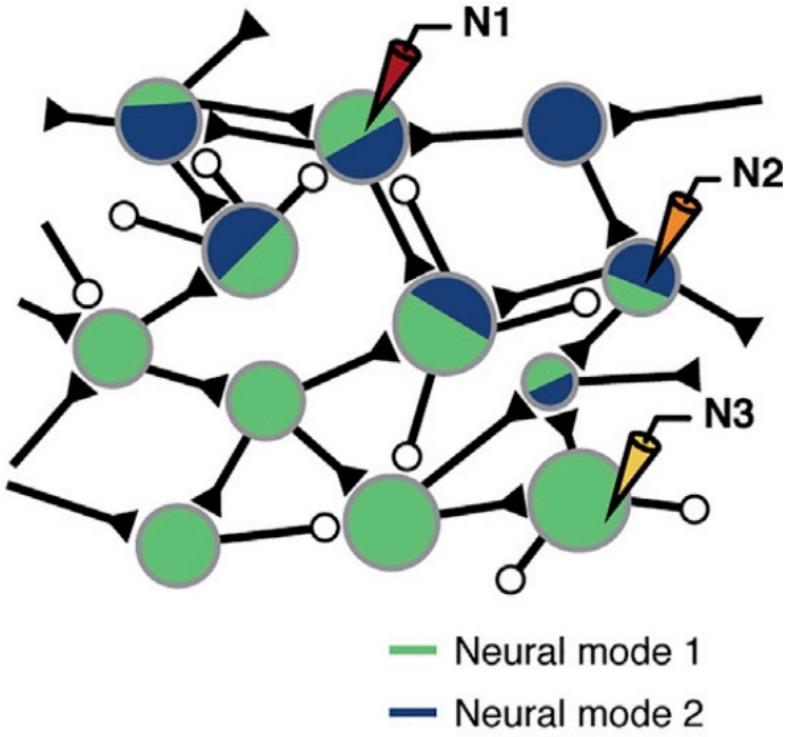


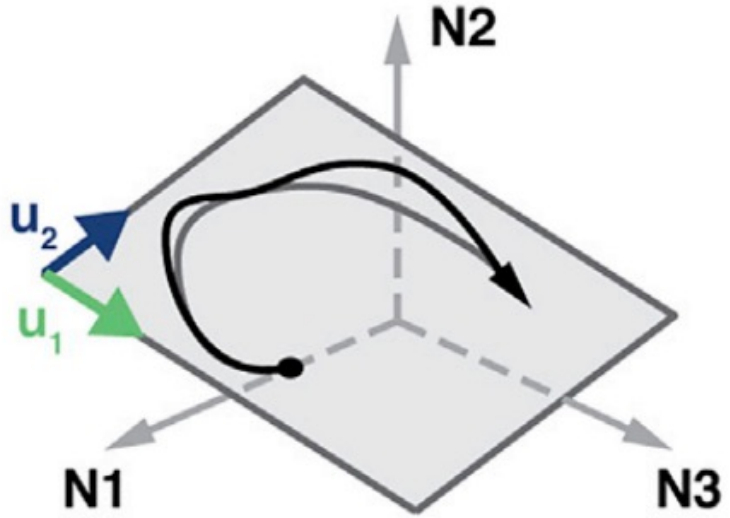


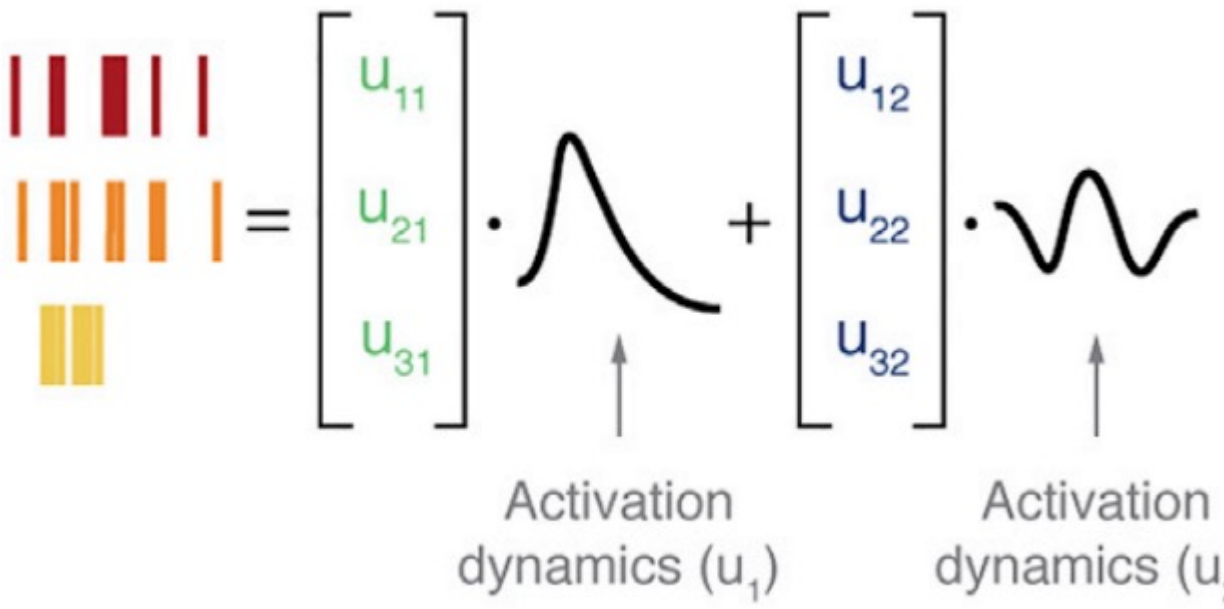


Stevenson, Neural Computation. 2018

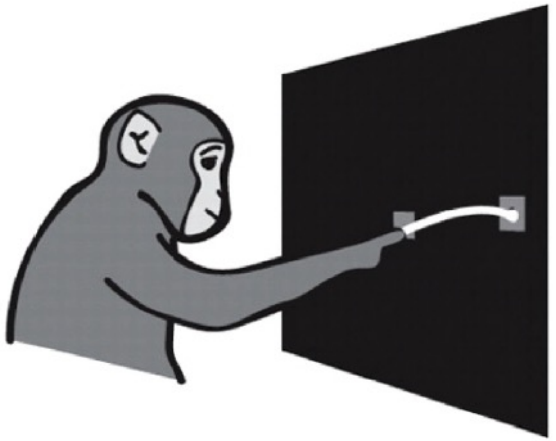
Dimensionality Reduction for Populations of Neurons

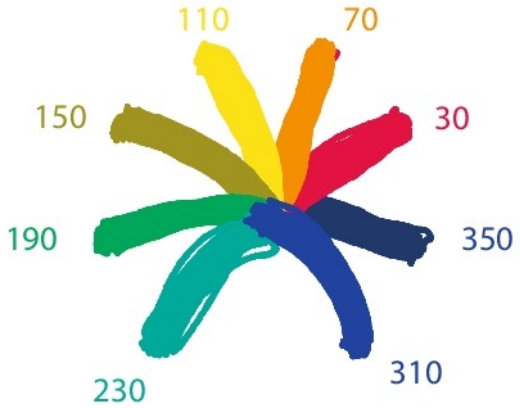


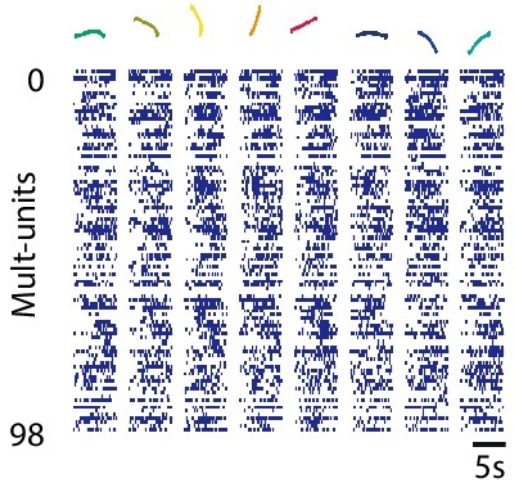


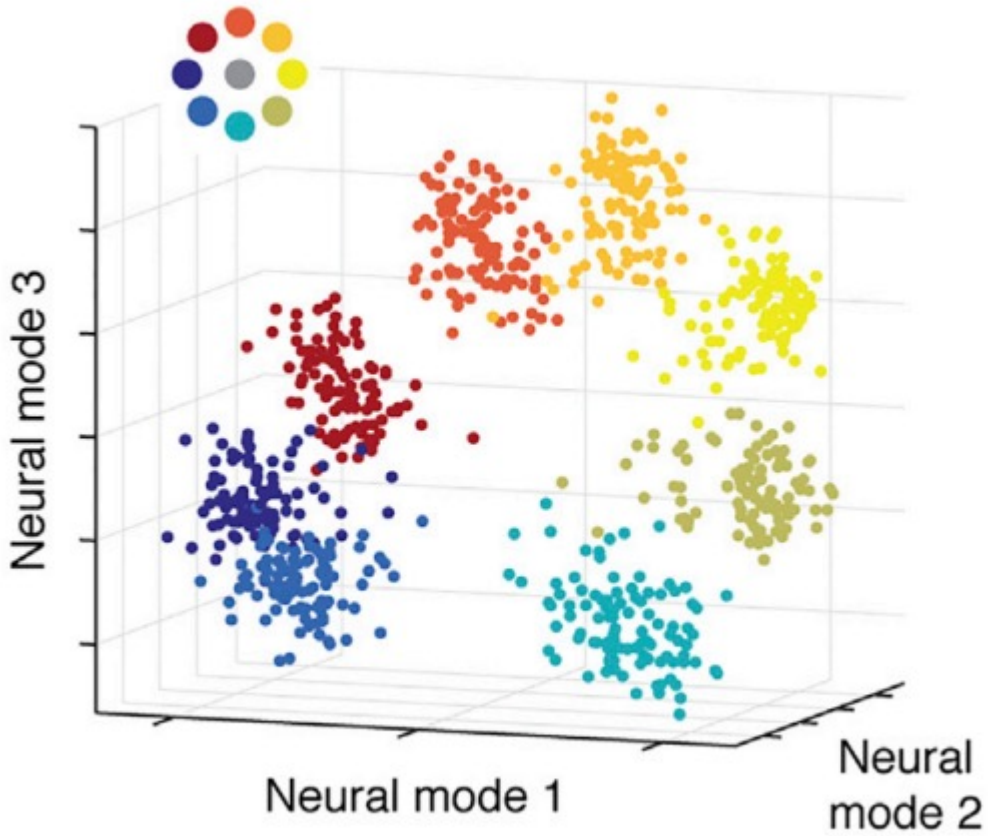


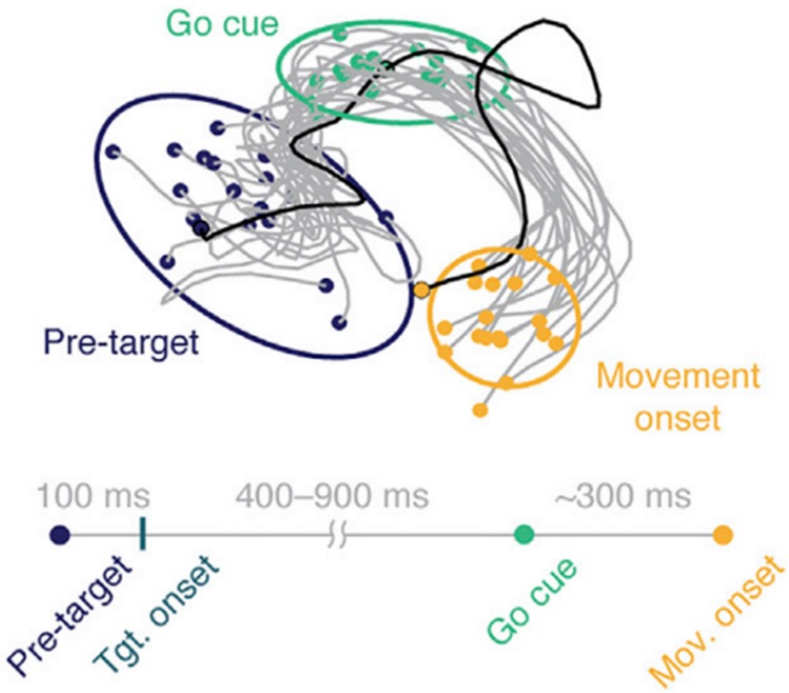
Gallego et al. 2017





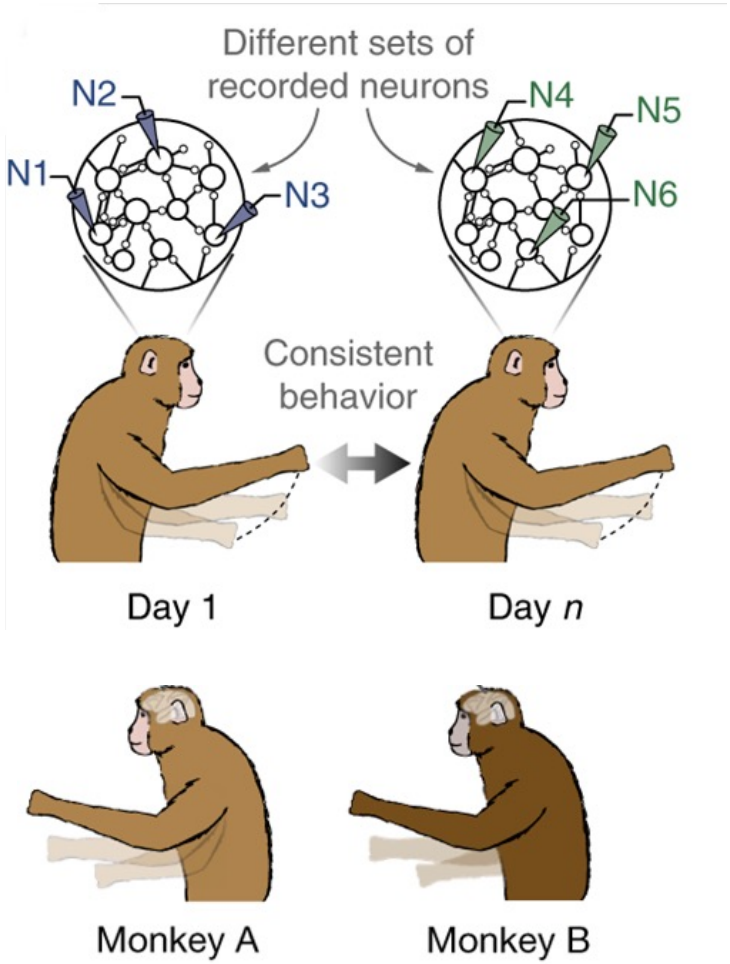


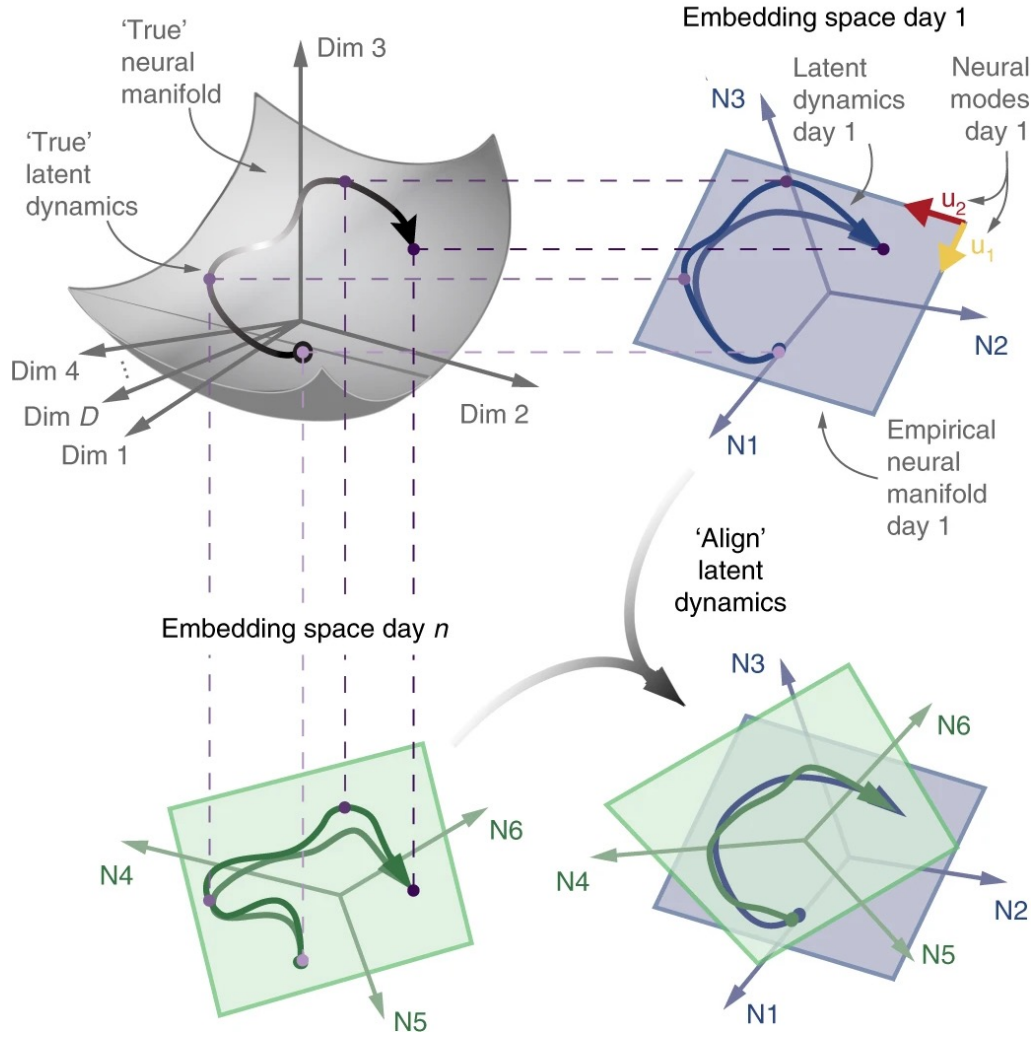


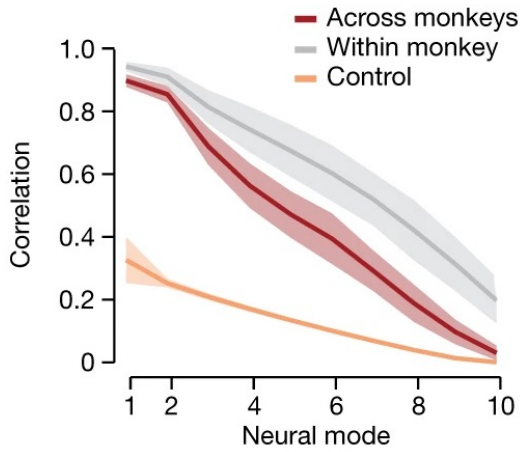


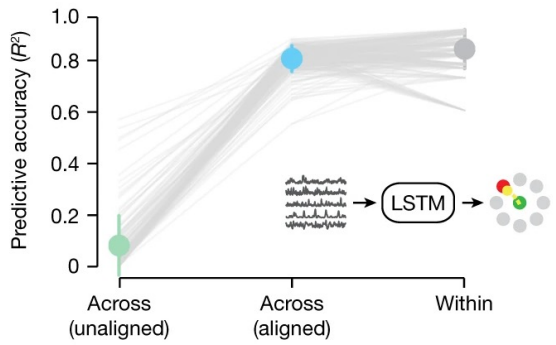
Based on Mitchell-Heggs et al. 2023, Gallego et al. 2017

Latent dynamics can be aligned over time/between animals









Gallego et al. 2020 and Safaie et al. 2023

Partial List of **Dimensionality Reduction Models for Neuro**

	Dynamics	Mapping	Observation	Additional Focus	
PCA	Static	Linear	n/a		Mazor (2005), Ahrens (2012)
FA	Static	Linear	Gaussian		Santhanam (2009), Churchland (2010)
ICA	Static	Linear	n/a		Laubach (1999), Lopes-dos-Santos (2013)
NMF	Static	Linear	Gaussian		Kim (2005), Xie (2017)
Isomap	Static	Nonlinear	n/a		Stopfer (2003), Carrillo-Reid (2008)
dPCA	Static	Linear	Gaussian	Task Variables	Kobak (2016)
TCA	Static	Linear	Gaussian	Trial Structure	Williams (2018)
NBFA	Static	Linear	NB	Noise	Pillow (2012)
jPCA	LDS	Linear	Gaussian	Rotations	Churchland (2012)
RLVM	LDS	NN	Gaussian	Rectification	Whiteway (2017)
PLDS	LDS	Linear	Poisson		Smith (2003), Macke (2012)
PODS	LDS	Linear/NN	Poisson/GC		Gao (2016)
SLDS	Switching LDS	Linear	Gaussian		Petreska (2011)
RSLDS	Recurrent Switch	Linear	Gaussian		Linderman (2017)
LFADS	RNN	Linear	Poisson		Pandarinath (2018)
GP-RNN	RNN	GP	Poisson		She (2020)
GPFA	GP	Linear	Gaussian		Yu (2009)
vLGP	GP	Linear	Poisson		Zhao (2017)
P-GPLVM	GP	GP	Poisson		Wu (2017)
mixDPFA	LDS	Mixture	Poisson/NB		Wei (2022)

Based on Hurwitz et al. 2021

Neural systems *handle* high-dimensional input

Major theoretical view is efficient coding and redundancy reduction
Some systems appear to have ~random projections

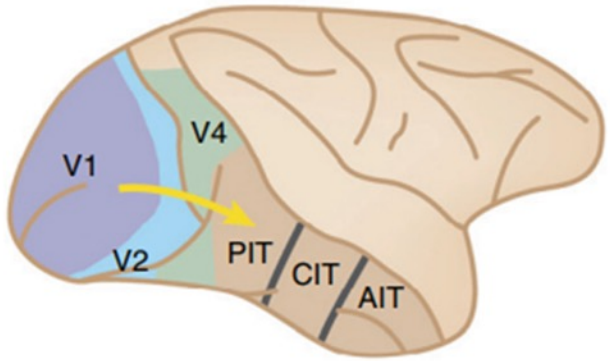
…biological systems have special constraints

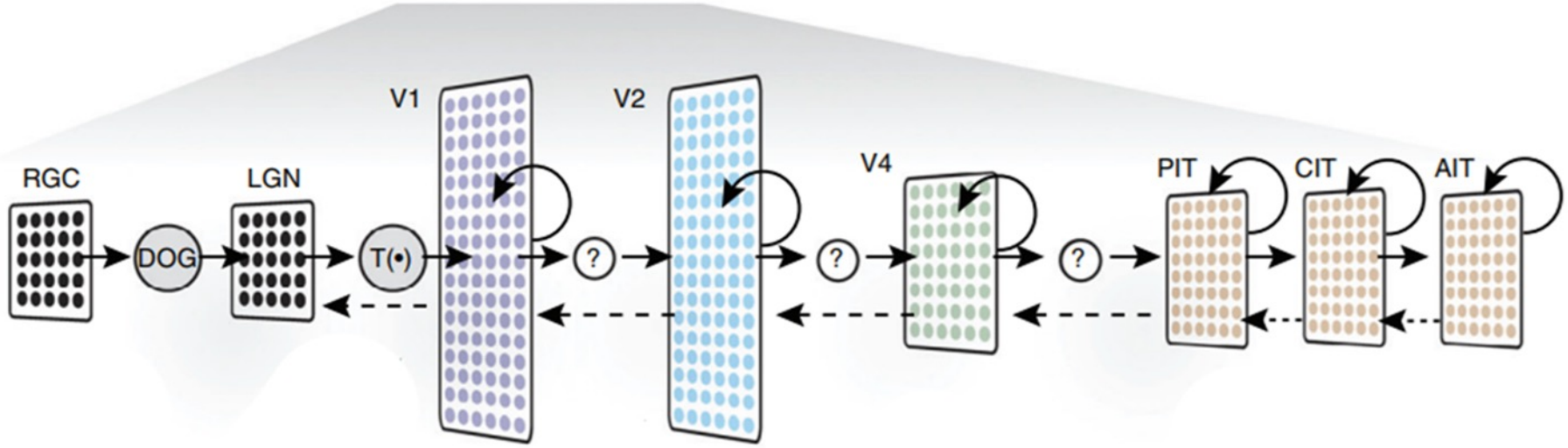
Changing input distribution (e.g. vision during the day vs night)
Bio/physics introduces noise, limits dynamic range
Constraints on energy, wiring
Representations support multiple objectives (e.g. tracking, few-shot learning, anomaly detection)

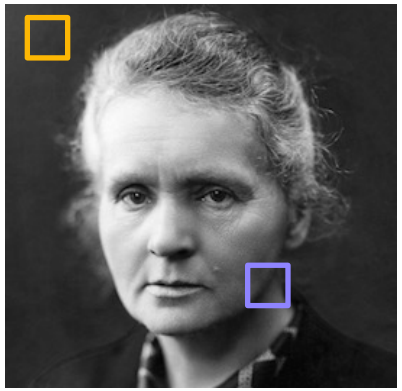
Dimensionality expansion is a theme in sensory systems

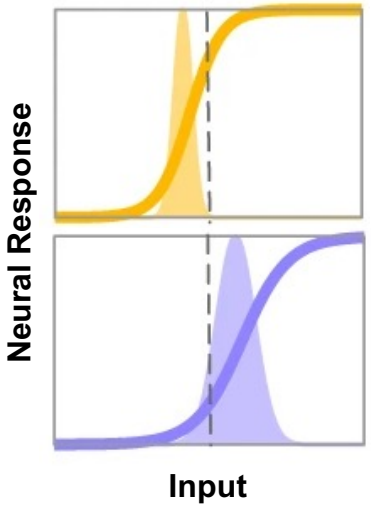
Human visual system… 100M rods and cones – 1M RGCs – 1M LGN – 250M V1
Human auditory system… 15K hair cells – 30K cochlear nerve fibers – 100M A1

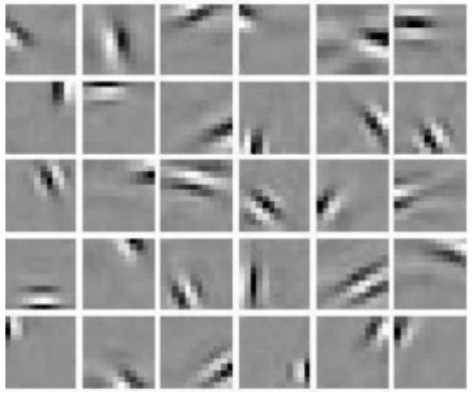
Efficient Coding and Redundancy Reduction

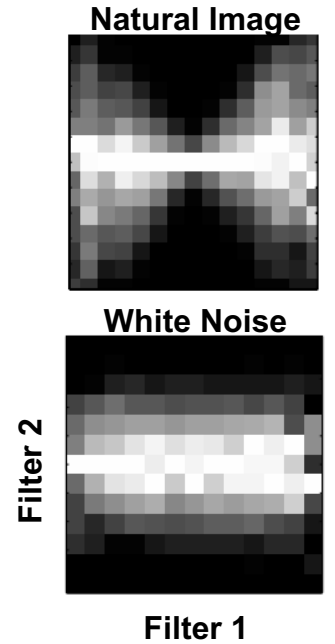






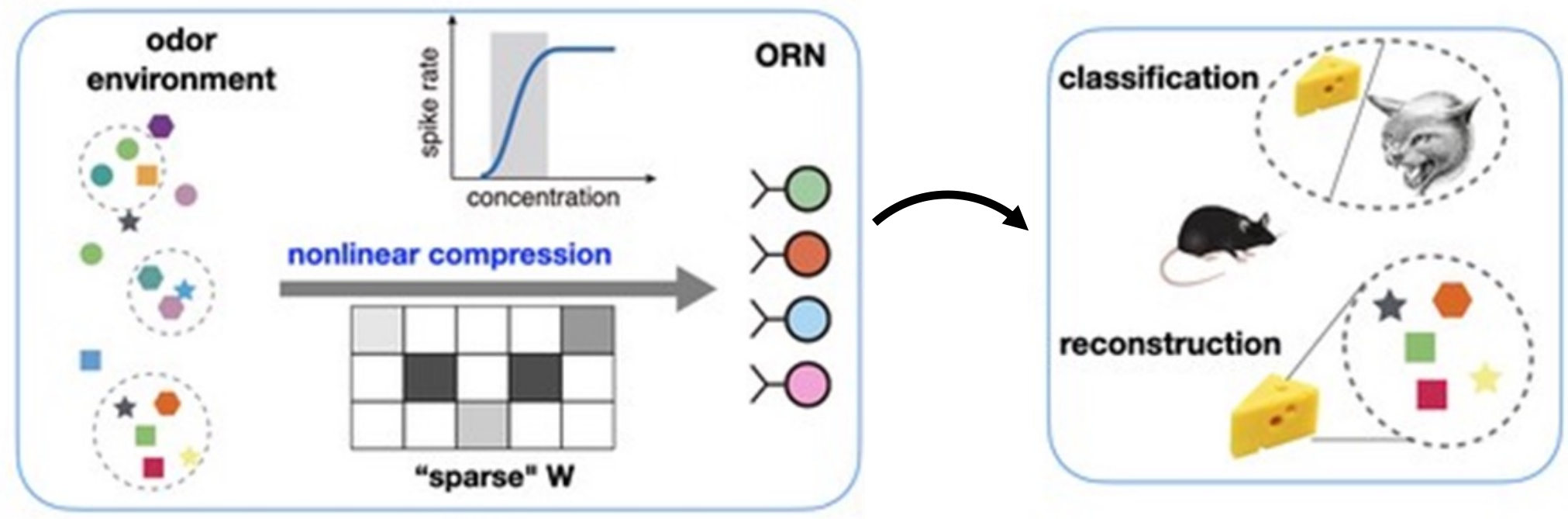




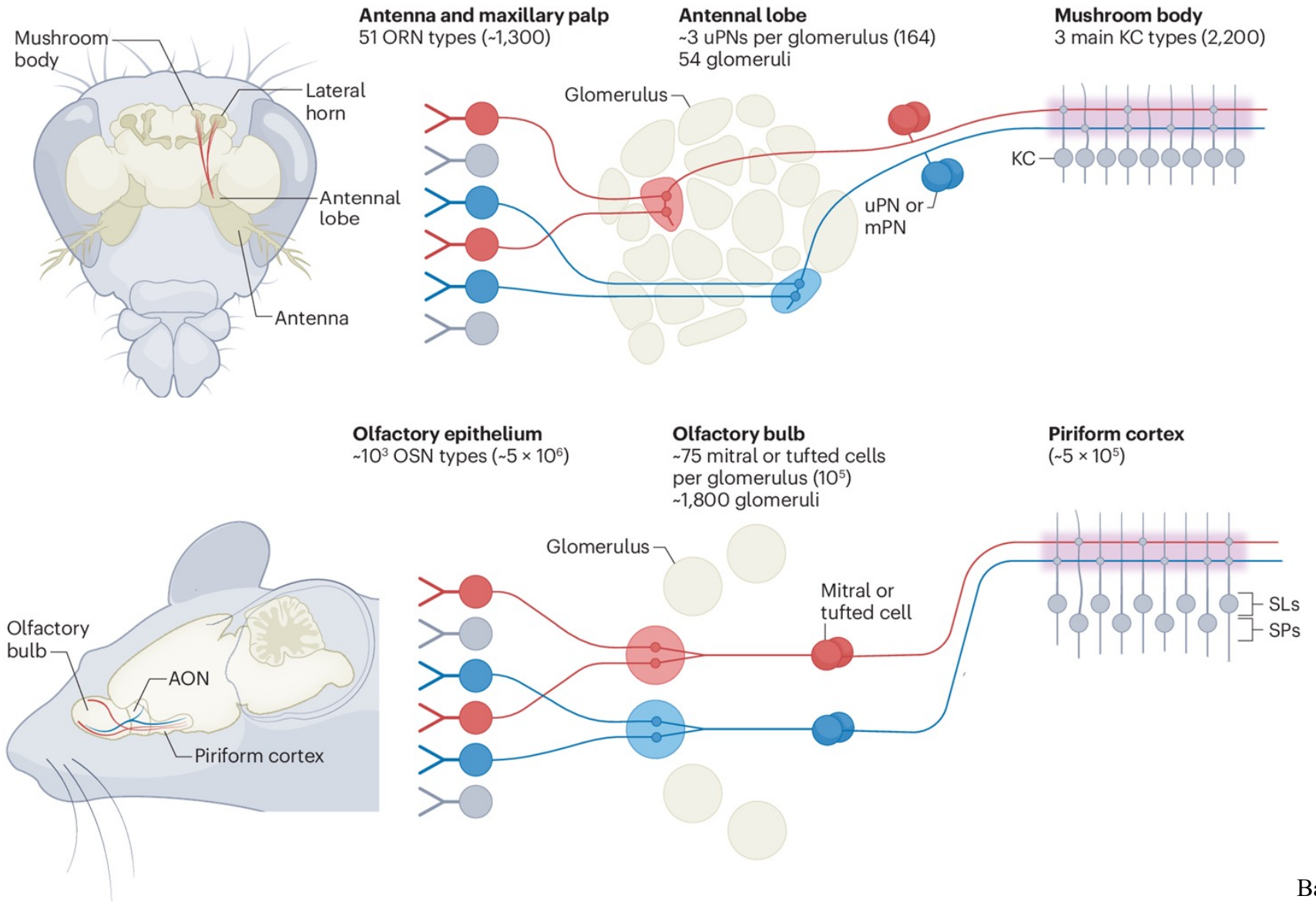


Yamins et al. 2014, Młynarski et al. 2021, Olshausen and Field 1996, Schwartz and Simoncelli 2001

Odor Perception and the Johnson-Lindenstrauss Lemma?



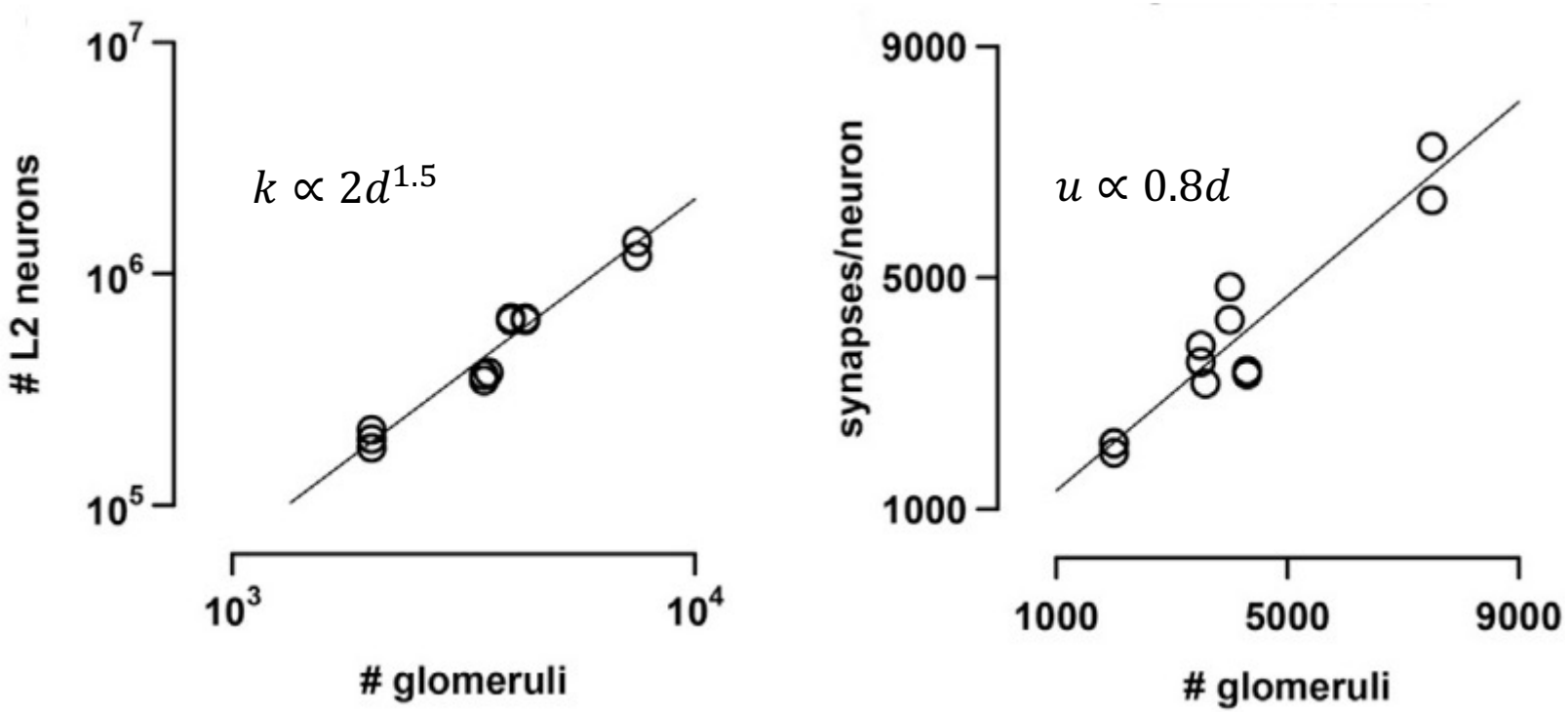
Odor Perception and the Johnson-Lindenstrauss Lemma?

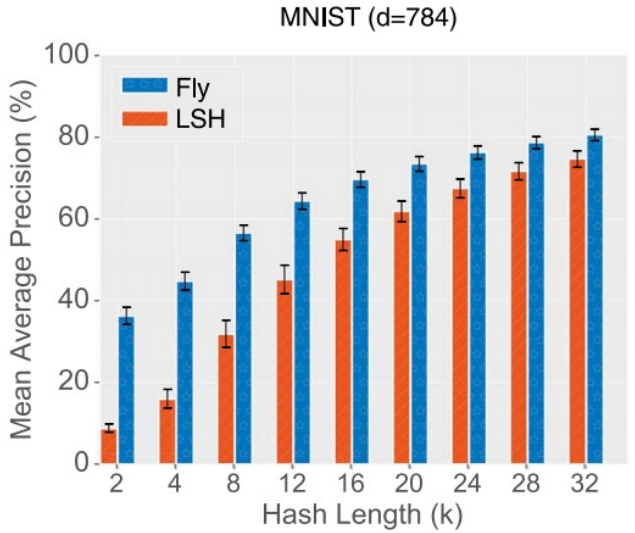


Based on Fulton et al. NRN 2024

Scaling Rules for Odor Perception

For dense random projection $f: R^d \rightarrow R^k$ JL is based on $k \geq \frac{3}{c\epsilon^2} \ln n$ but …





Srinivasan and Stevens, 2019

Dasgupta et al. 2017

Takeaways

Data in neuroscience is growing rapidly

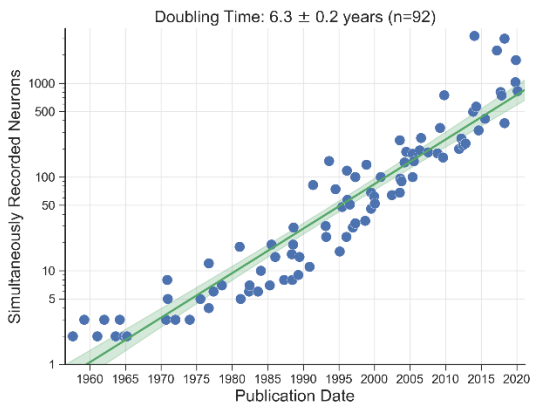
High dimensional data analysis problems abound…

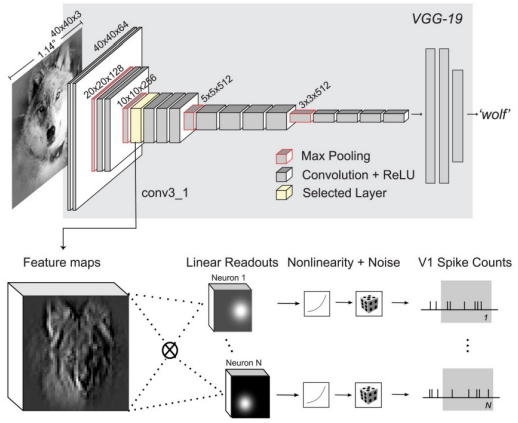
what (high-d) natural variables are neurons sensitive to?

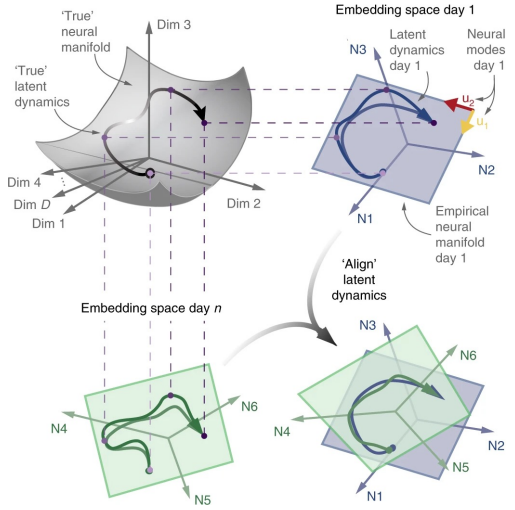
how can we understand (high-d) populations of neurons?

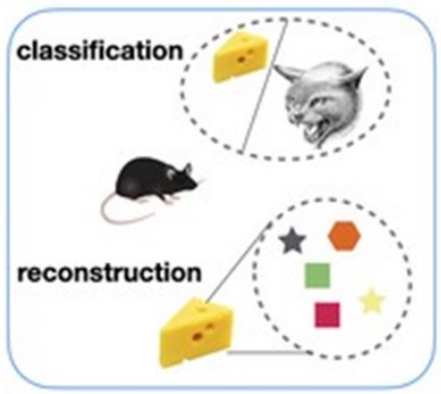
Neural systems solve high dimensional problems too

some solutions (e.g. odor perception) may be related





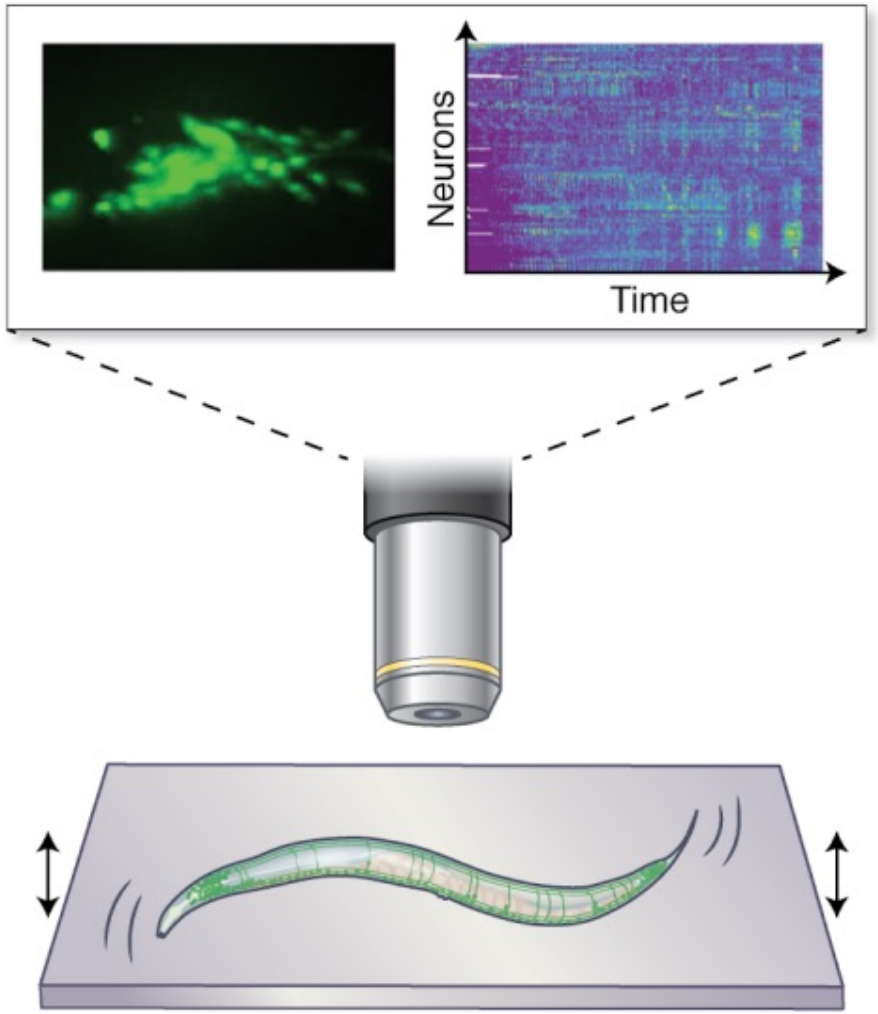




Ian's Discussion…

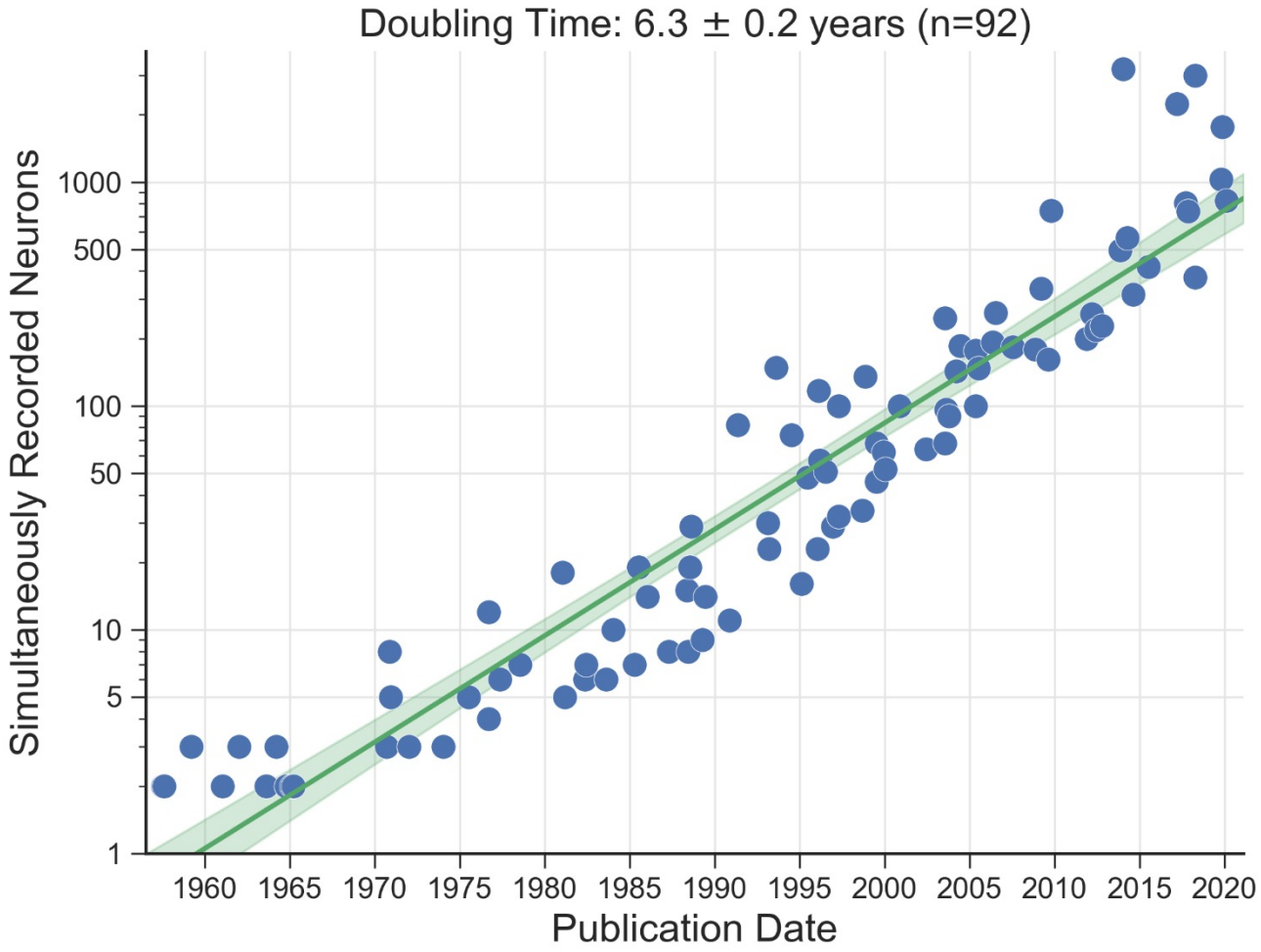
Advertising

Open Questions
from a Neural Data Scientist



Theory is great, how can I apply it?

“100B neurons = job security” – Chuck Stevens
“Data analysis is the rate-limiting step in many neuroscience PhDs, and it’s getting worse…” - me

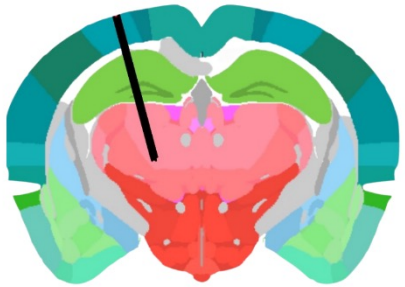


Stevenson and Kording, *Nature Neuroscience* (2011)

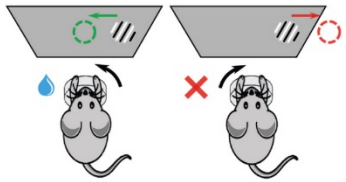
data & figure at stevenson.lab.uconn.edu

In the pipeline effort #1

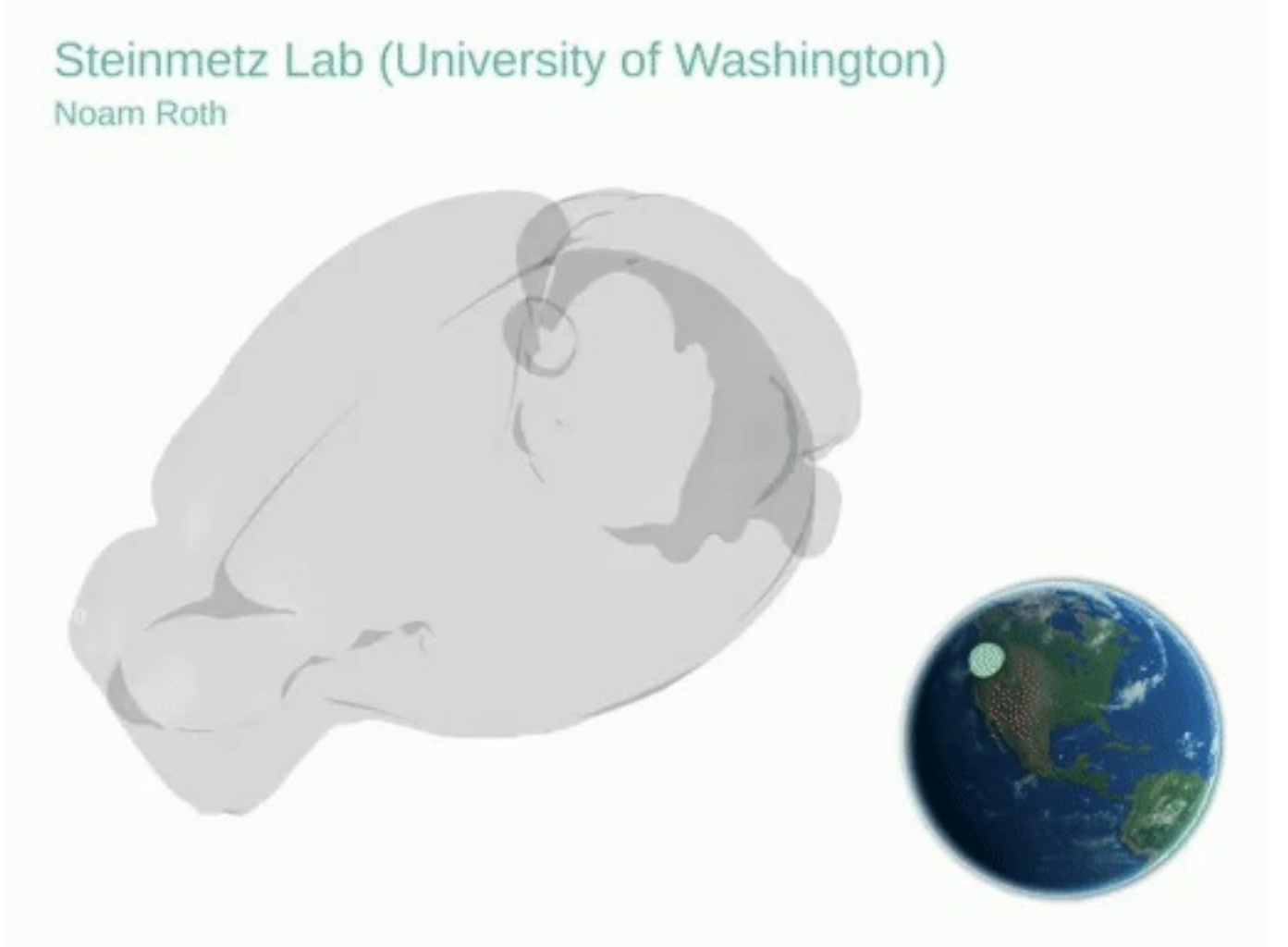
INTERNATIONAL
BRAIN
LABORATORY



Reproducible Electrophysiology



Standardized Behavior

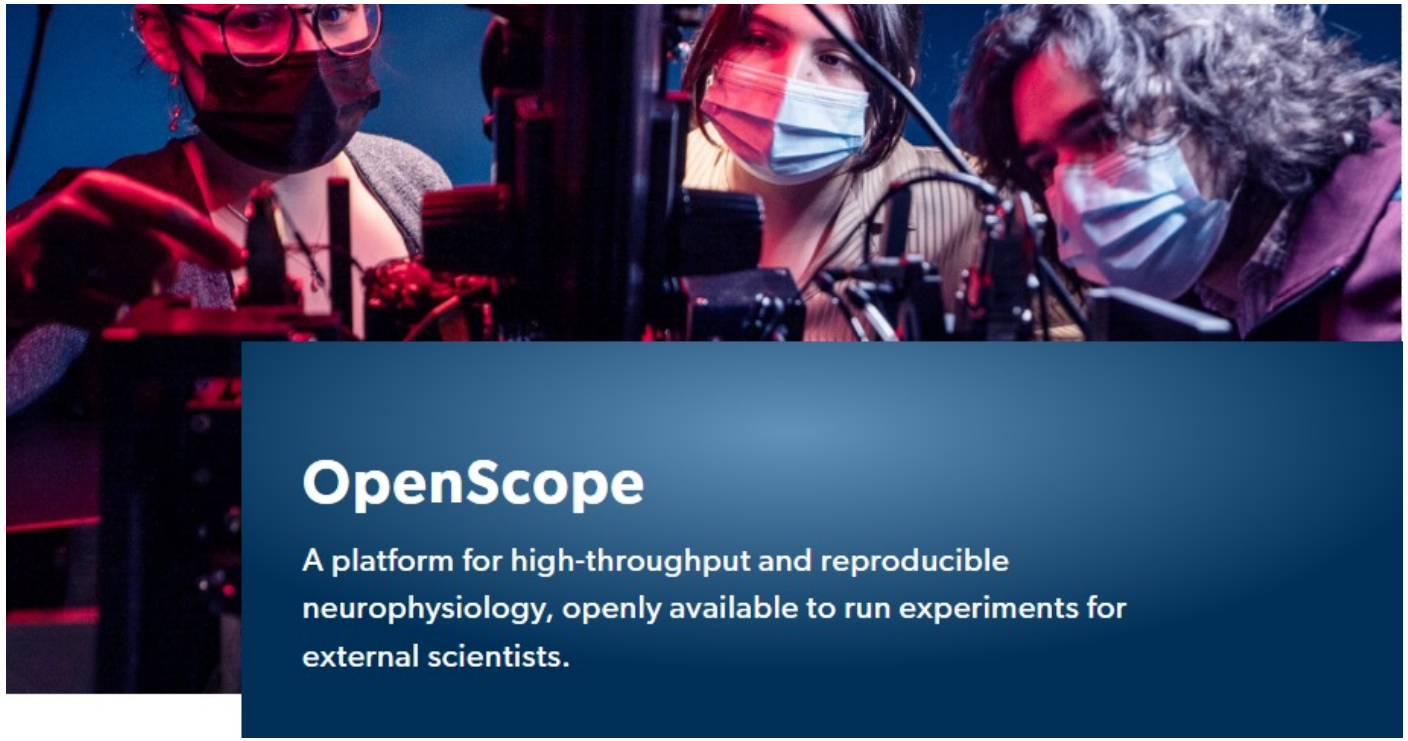


internationalbrainlab.com

In the pipeline effort #2

OpenScope

ALLEN INSTITUTE

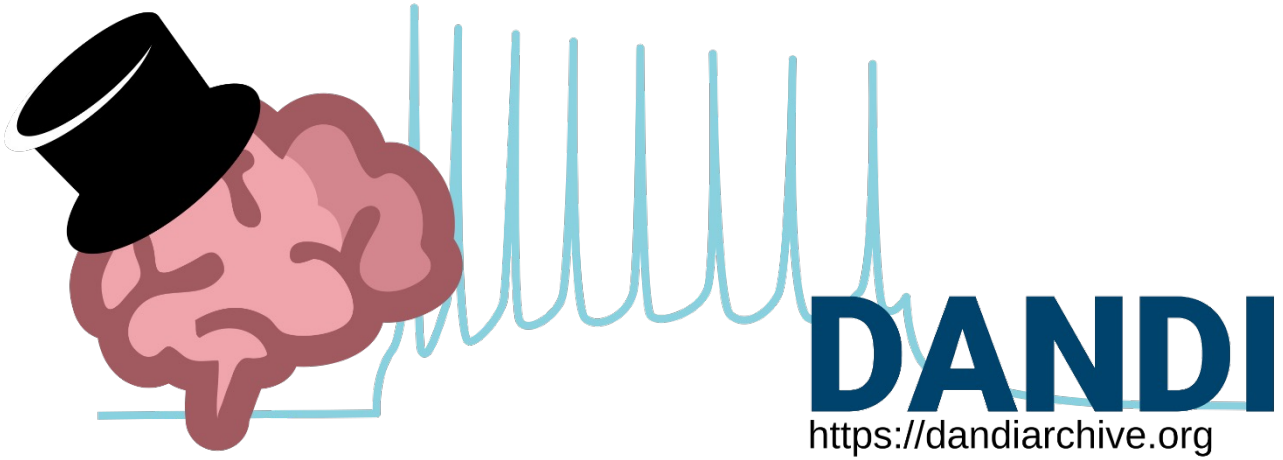


OpenScope is the first astronomical observatory in Neuroscience. Scientists across the world can propose new experiments that will be executed freely on the Allen Brain Observatory pipeline.

- Record from thousands of neurons across the mouse brain using either multi-probes Neuropixels recordings or multi-area two photon calcium imaging.
- An end-to-end standardized experimental platform including brain surgery, animal behavioral training, neuronal recordings and brain reconstruction.

alleninstitute.org/division/neural-dynamics/openscope/

Neural Data Repositories, Standards, and Competitions

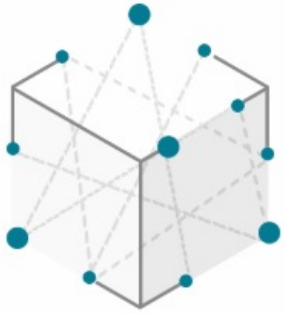


DANDI

https://dandiarchive.org

Neural Latents Benchmark

A Benchmark for Models of Neural Data



OpenNEURO

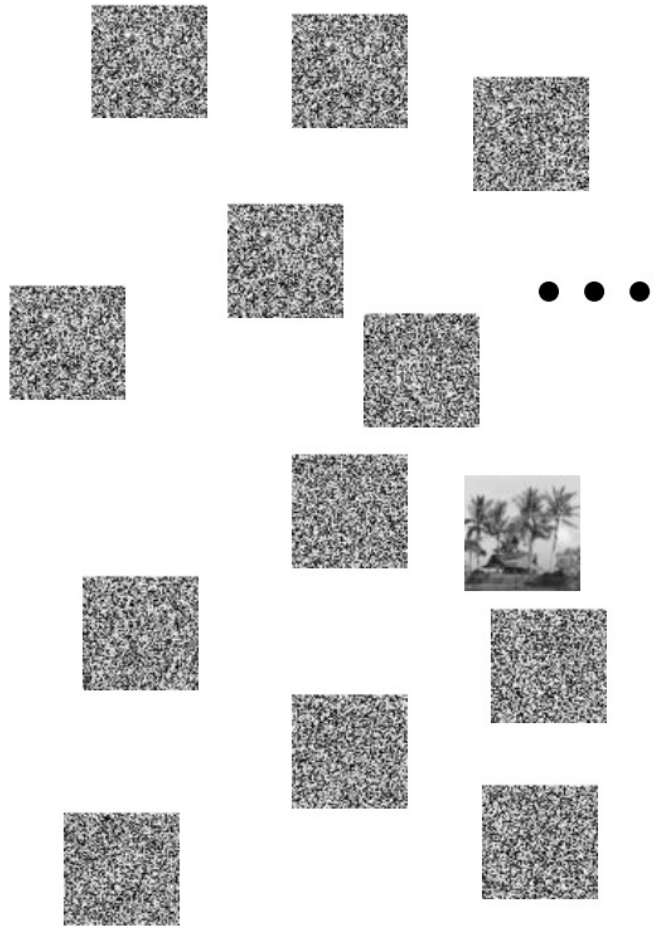
NEURODATA WITHOUT BORDERS

Collaborative Research in Computational Neuroscience

crcns.org

Questions from a Neural Data Scientist…

"The distribution of natural images is complicated. Perhaps it is something like **beer foam**, which is mostly empty but contains a thin mesh-work of fluid which fills the space and occupies almost no volume. The fluid region represents those images which are natural in character." [Ruderman 1996]



Which results apply to natural distributions?

My basic intuition…

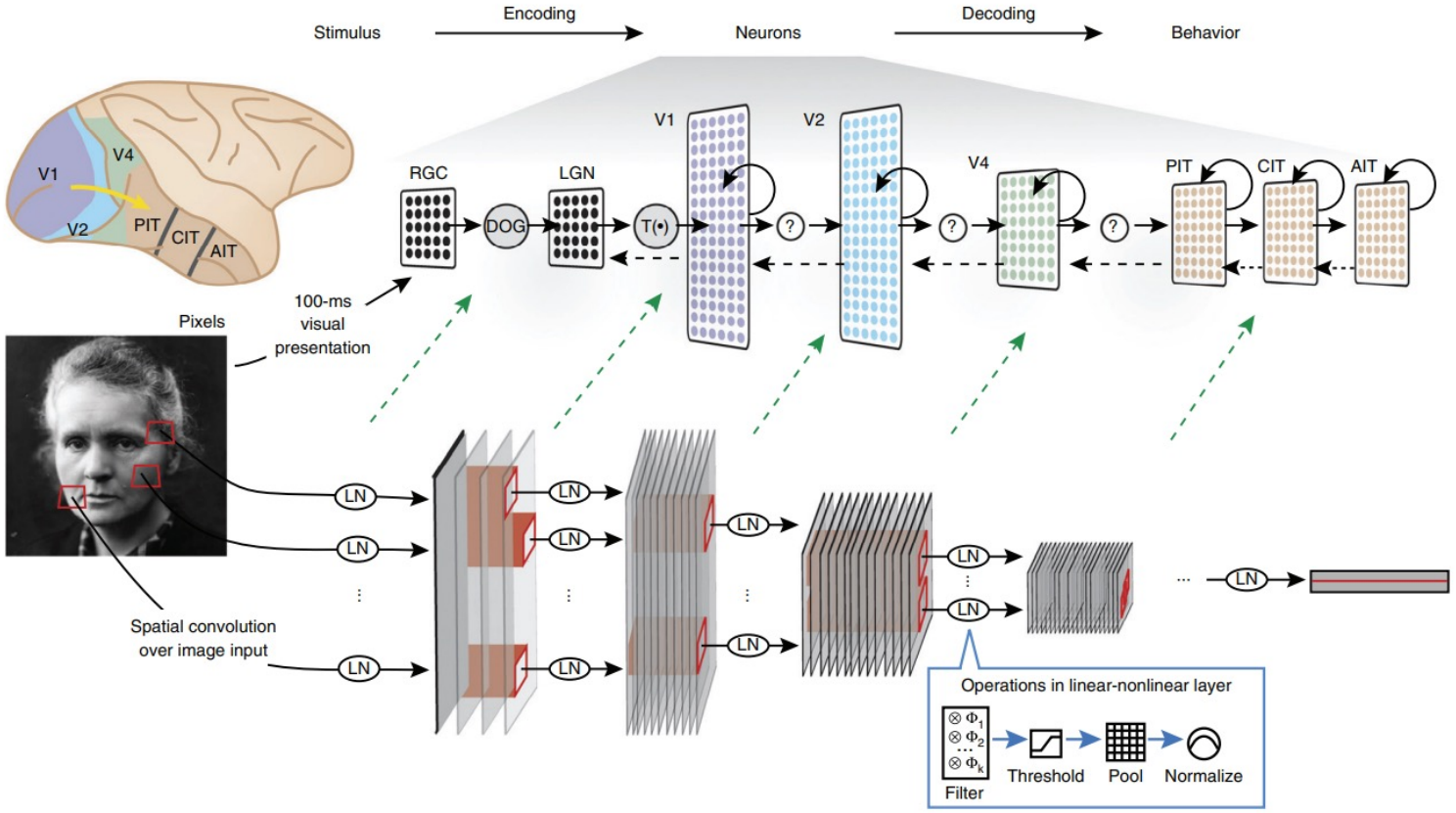
- volume of natural images << volume of a Gaussian
- not everything is on the convex hull?
- many non-linearities, continuous trajectories
- still ~orthogonal
- distances are really interesting
- many transformations that don't preserve distance

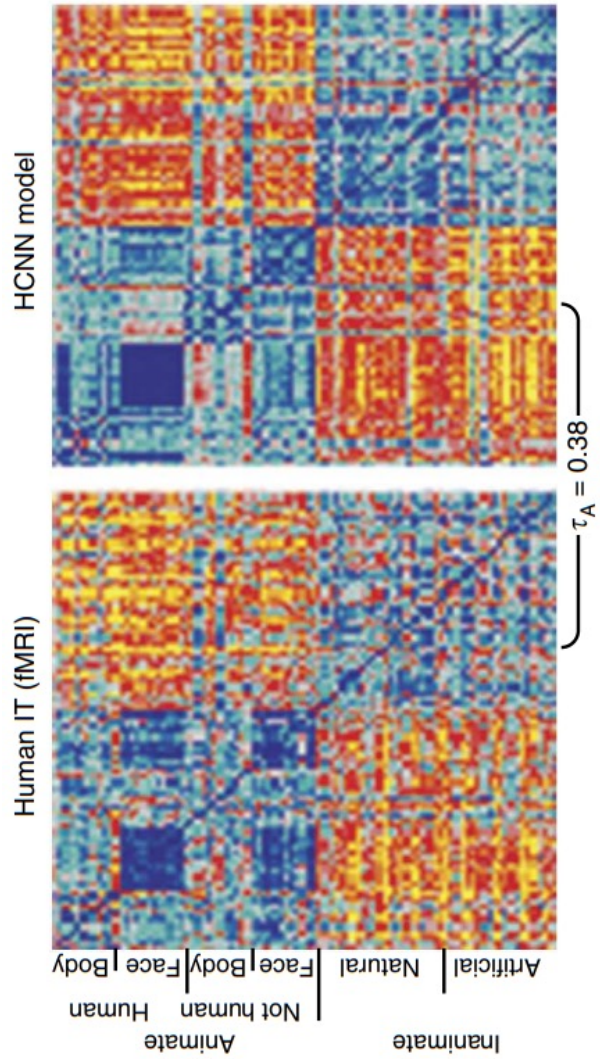
Many of these things can just be measured!

via Siwei Lyu

Questions from a Neural Data Scientist…

Comparing high-d representations?
”representational similarity analysis”

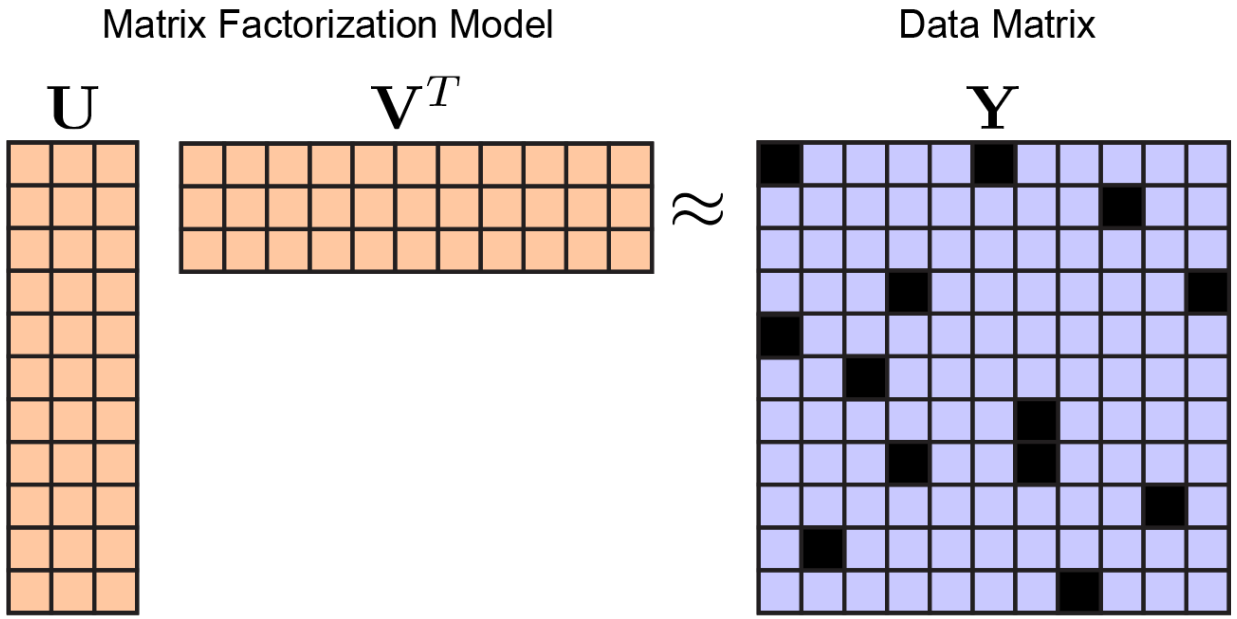


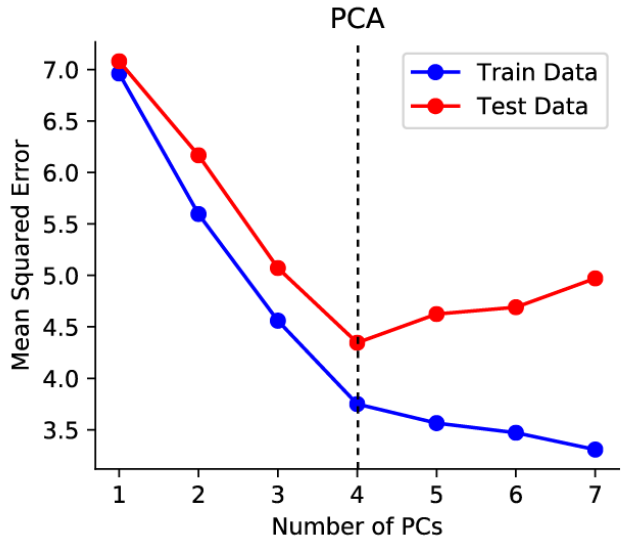


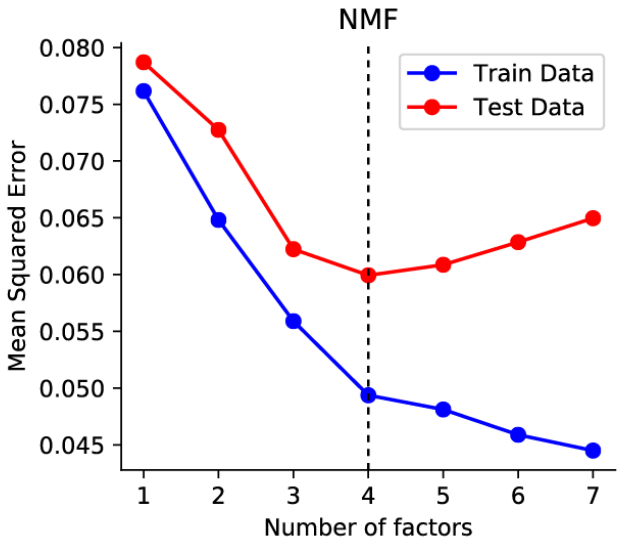
Yamins et al. 2014

Questions from a Neural Data Scientist…

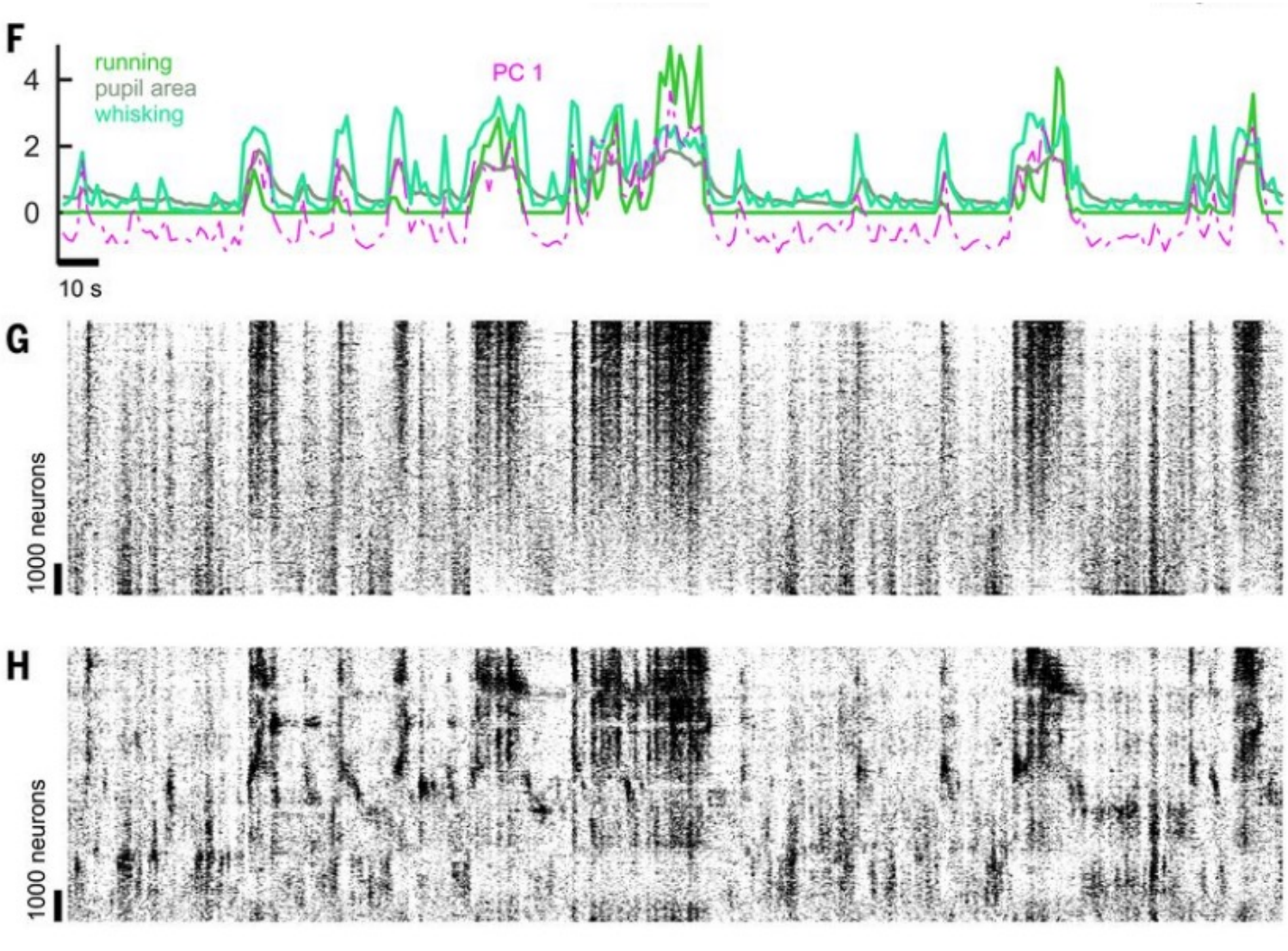
Selecting low-d representations?
noise vs model misspecification

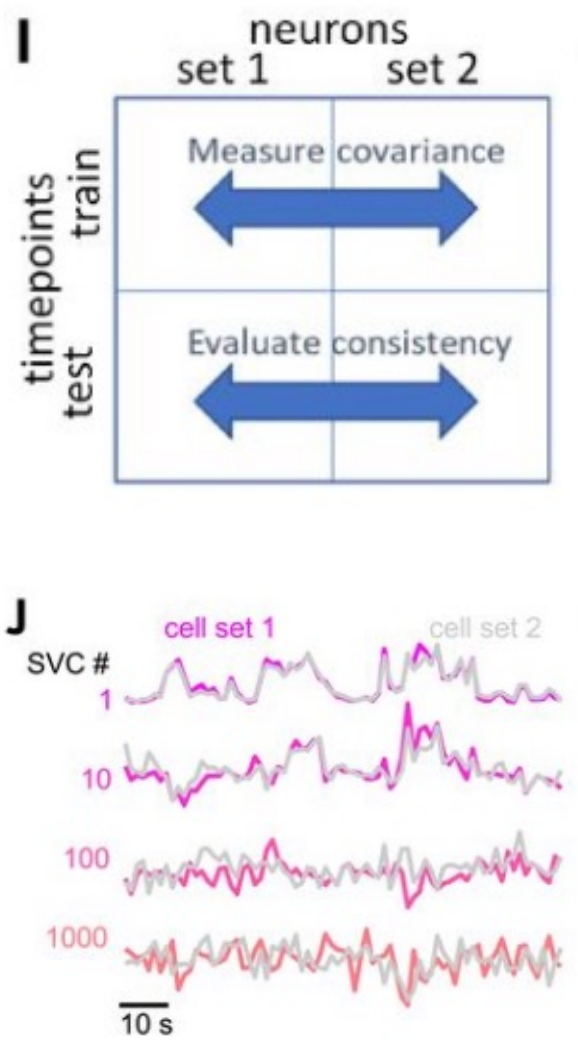






Via Alex Williams





Stringer et al. 2019

Neuron

Simultaneous, cortex-wide dynamics of up to 1 million neurons reveal unbounded scaling of dimensionality with neuron number

Highlights

- Cellular-resolution, cortex-wide volumetric imaging of neuronal population dynamics
- Dimensionality of cortical dynamics exhibits power law scaling with number of neurons
- Half of neural variance contains high-dimensional dynamics without behavioral correlate
- Neural ensembles are highly distributed, fine grained, and cover a continuum of timescales

Authors

Jason Manley, Sihao Lu, Kevin Barber, ..., David Meyer, Francisca Martínez Traub, Alipasha Vaziri

Correspondence

vaziri@rockefeller.edu

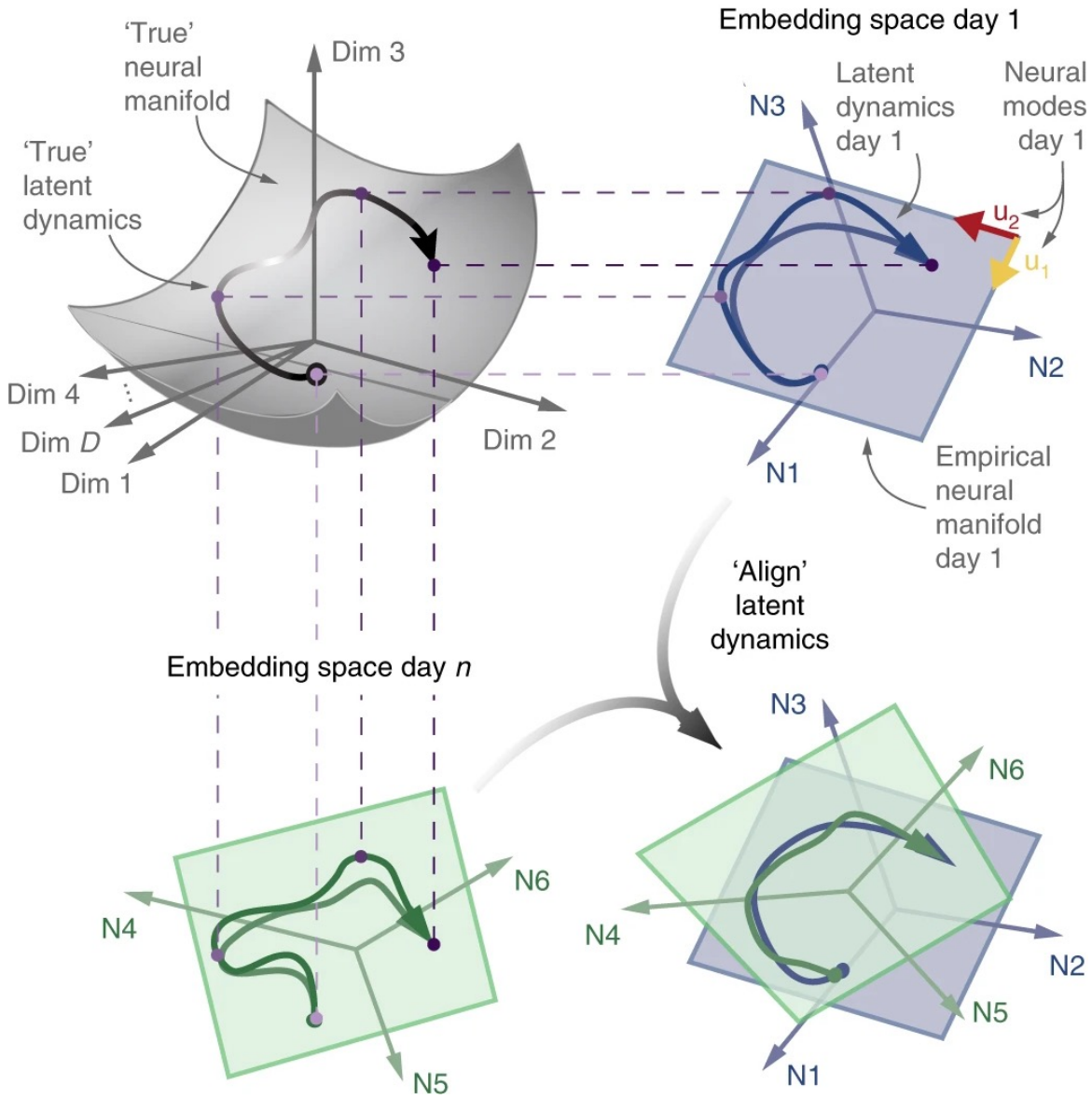
In brief

Neuronal populations have often been argued to exhibit low-dimensional

Questions from a Neural Data Scientist…

Aligning low-d representations?
CCA on aligned data
distribution alignment
deep nets (Eva Dyer)

can we do better? Or some guarantees?



Gallego et al. 2020 and Safaie et al. 2023