

Biologically inspired neural networks for modeling brain functions and neural dynamics

Z. Sage Chen, PhD

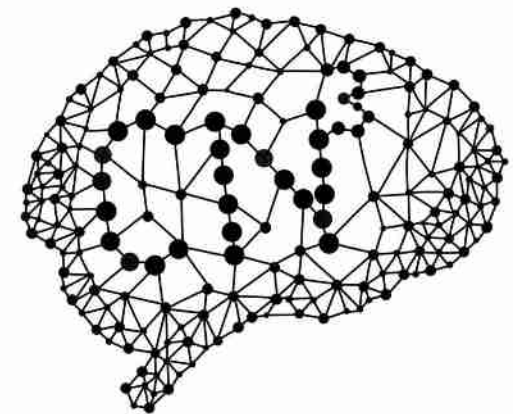
NYU Neuroscience Institute

Department of Psychiatry

Department of Neuroscience & Physiology

Department of Biomedical Engineering

Email: zhe.chen@nyulangone.org



NeuroAI

- Bridge neuroscience and AI to understand brain function & develop AI systems inspired by the brain.

The BRAIN Initiative[®]
BRAIN NeuroAI Workshop

November 12 & 13, 2024

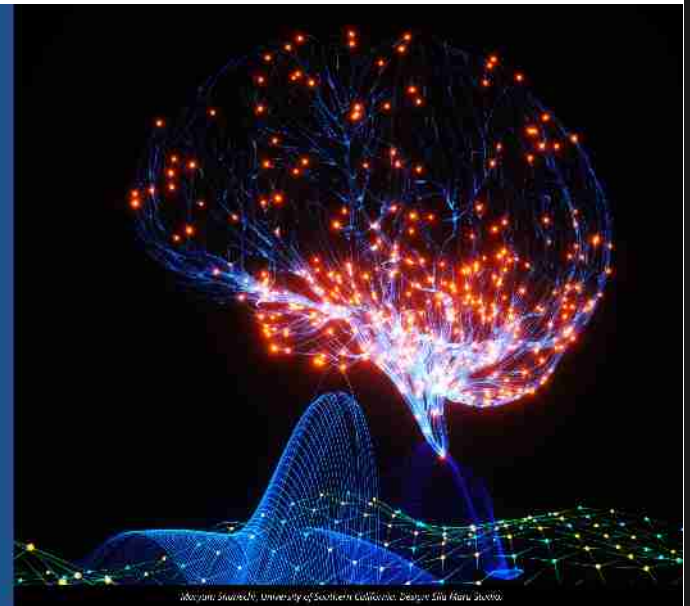
Register Now for Virtual Attendance
<https://n4solutionsllc.com/BRAINNeuroAI>

Questions? Contact us!
BRAINNeuroAIWorkshop@ninds.nih.gov

Learn More & Register



NIH National Institutes of Health
The BRAIN Initiative[®]



Examples:

BMIs to restore/enhance brain functions

Building NN models to understand brain functions

Using DNN/transformers to analyze neural data

Using generative AI to generate neural data

Developing neuro-inspired AI models & algorithms



CNS2025 Workshop: BRAIN 2.0: Emerging Research Topics in NeuroAI

Workshop organizers:
Zhe Sage Chen (New York University)
Michael Halassa (Tufts University)

Speakers
Mehdi Azabou (Columbia University)
Zhe Sage Chen (New York University)
Eva Dyer (Georgia Tech)
Tatiana Engel (Princeton University)
Michael Halassa (Tufts University)
Dmitry Krotov (MIT-IBM Watson AI Lab)

Location: Palazzo del Congresso
Whitaker, Florence, Italy
When: July 9, 2025



More details are available at



What is covered in lectures & what is not?

- High-level concepts
- Rationale (vs. alternatives)
- Some walk-through examples
- Detailed math derivation
- Detailed discussions

Two tales of computational neuroscience

- Modeling (microcircuit, behavior) ---Lecture 1
- Neural data analysis (large-scale, multi-modal) --- Lecture 2

Common goal: **linking neural representations to behaviors**

Roles of computational modeling

- **Uncovering computational principles and mechanisms --- complementary to circuit mechanisms**

neural representations at single neuron and population levels

spike timing or neural oscillatory dynamics

advantages: unlimited trials, a very large network size

- **Making experimental predictions or experimentally testable hypotheses**

computational vs. optogenetic simulation

computational vs. experimental lesion



Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in
Neurobiology

Recurrent neural networks as versatile tools of neuroscience research

Omri Barak



J Neurophysiol 122: 2504–2521, 2019.
First published October 16, 2019; doi:10.1152/jn.00467.2018.

RESEARCH ARTICLE | *Neural Circuits*

Considerations in using recurrent neural networks to probe neural dynamics



Jonathan C. Kao^{1,2}

¹Department of Electrical and Computer Engineering, University of California, Los Angeles, California; and ²Neurosciences Program, University of California, Los Angeles, California

Submitted 9 July 2018; accepted in final form 7 October 2019

Primer

Artificial Neural Networks for Neuroscientists: A Primer

Guangyu Robert Yang^{1,*} and Xiao-Jing Wang^{2,*}

¹Center for Theoretical Neuroscience, Columbia University, New York, NY, USA

²Center for Neural Science, New York University, New York, NY, USA

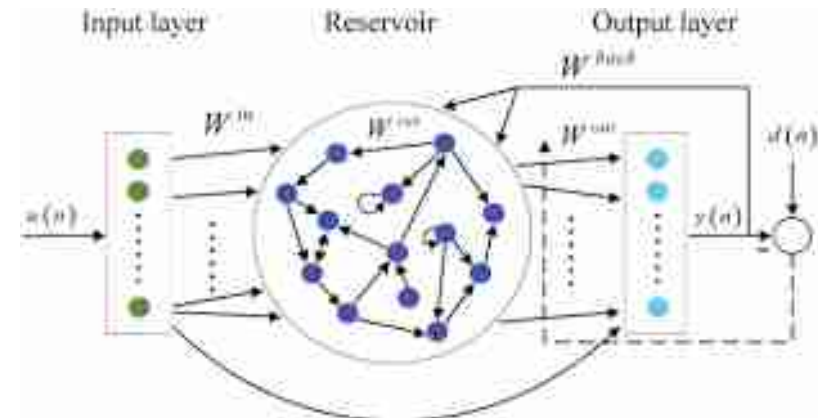
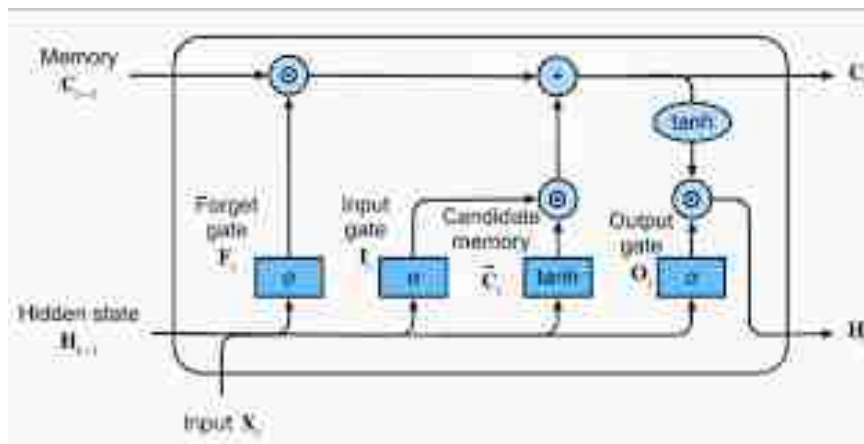
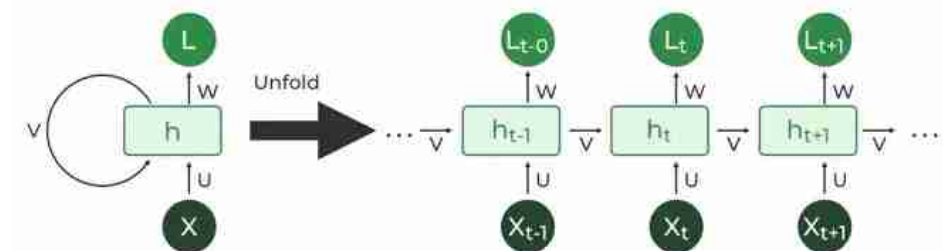
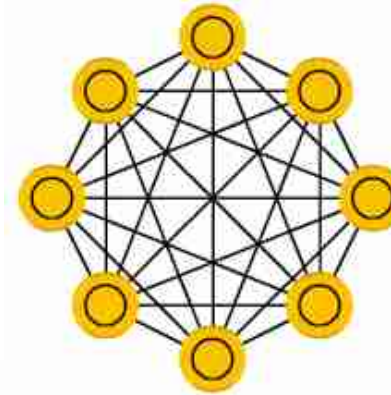
*Correspondence: robert.yang@columbia.edu (G.R.Y.), xjwang@nyu.edu (X.-J.W.)

<https://doi.org/10.1016/j.neuron.2020.09.005>

SUMMARY

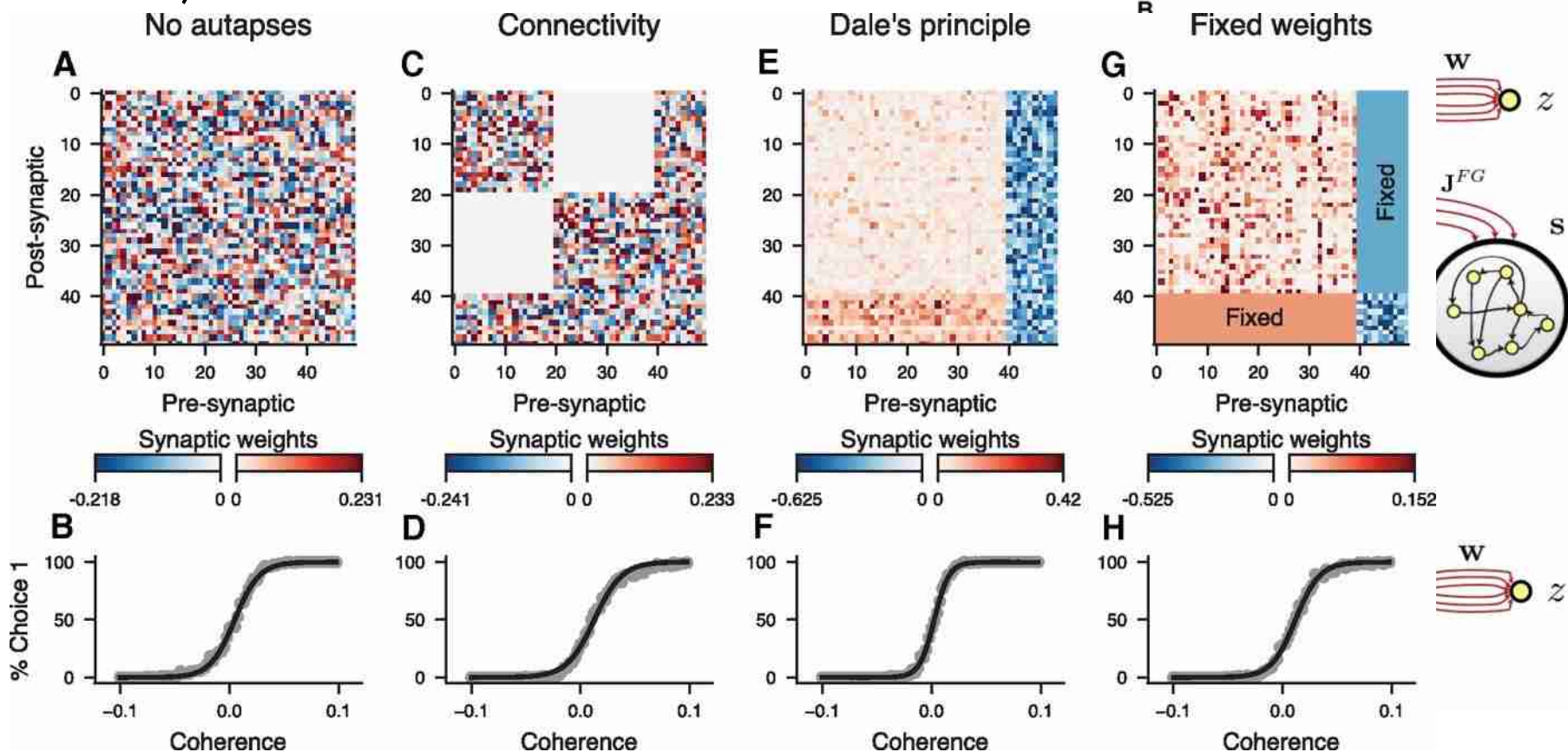
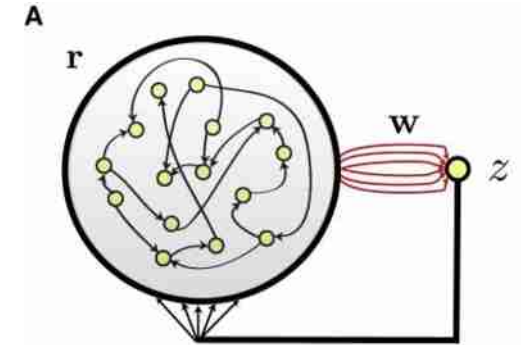
Discrete time RNNs

- Naïve RNN (hidden-to-hidden, output-to-hidden connections, etc)
- LSTM (Hochreiter & Schmidhuber, 1997)
- Echo state network (Jaeger, 2002)

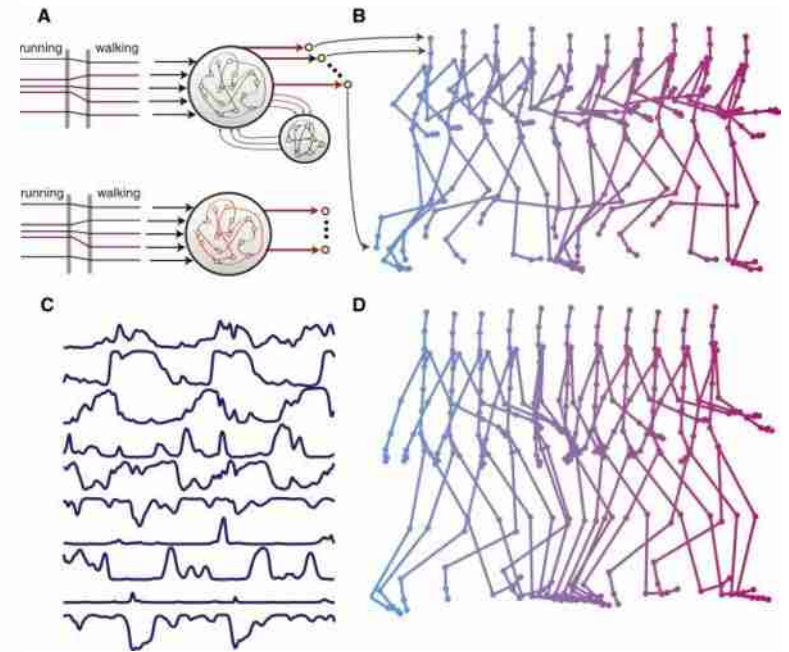
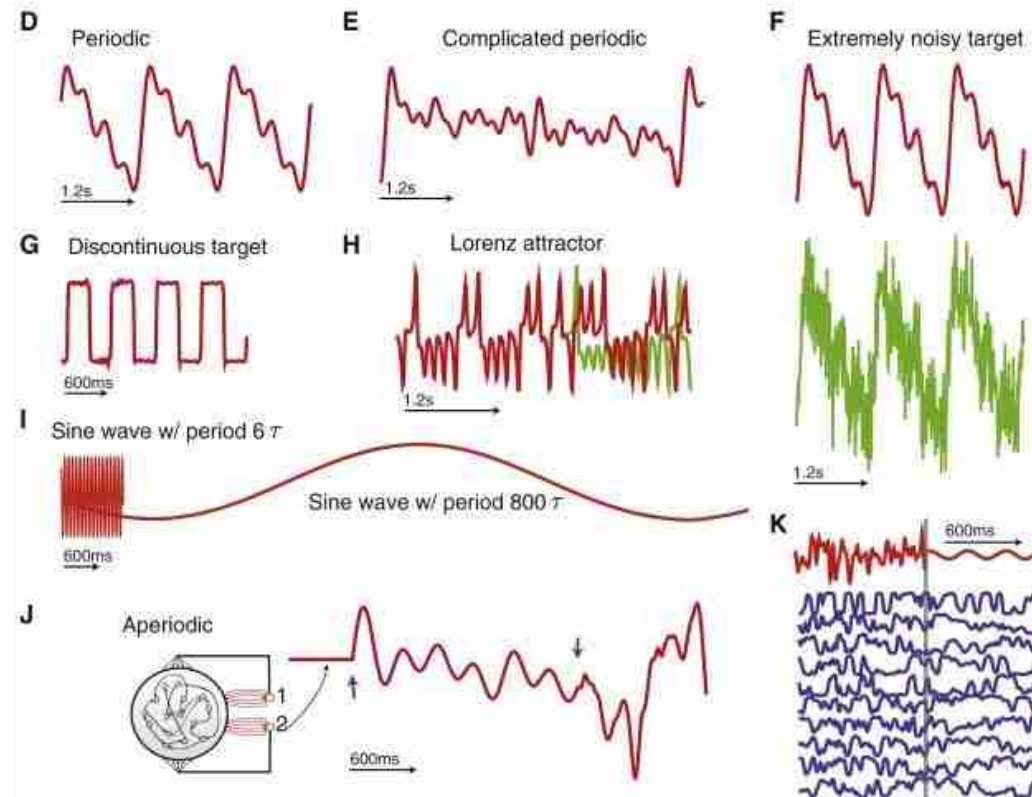
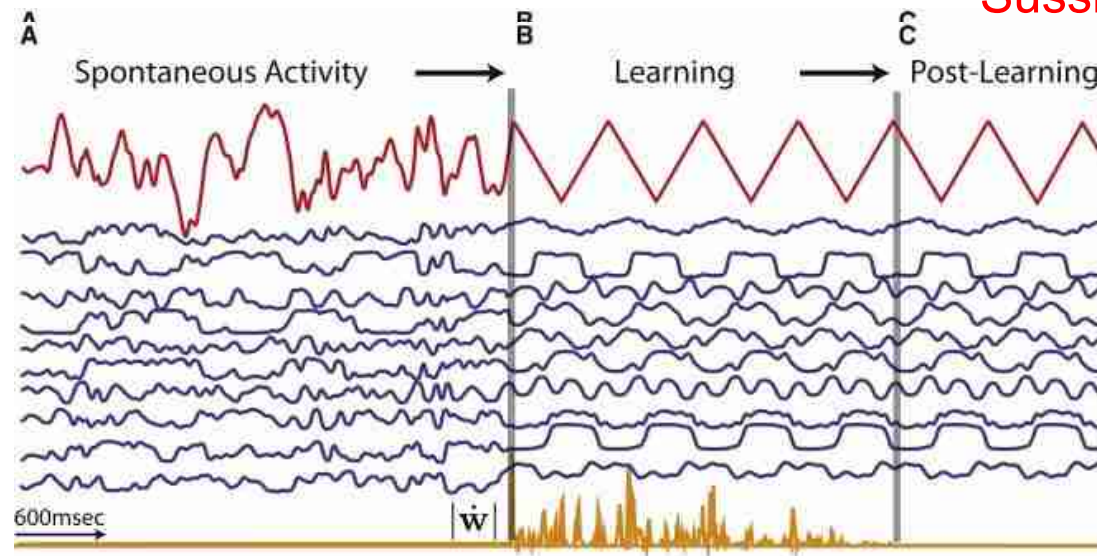


Variants of continuous-time RNNs

- Liquid state machine (Wolfgang et al. 2002)
- FORCE learning (Sussillo & Abbott, 2009)
- E/I-constrained RNN (Song et al., 2017)
- Low-rank RNN (Mastrogiuseppe & Ostojic, 2018)



Sussillo & Abbott (2009) Neuron

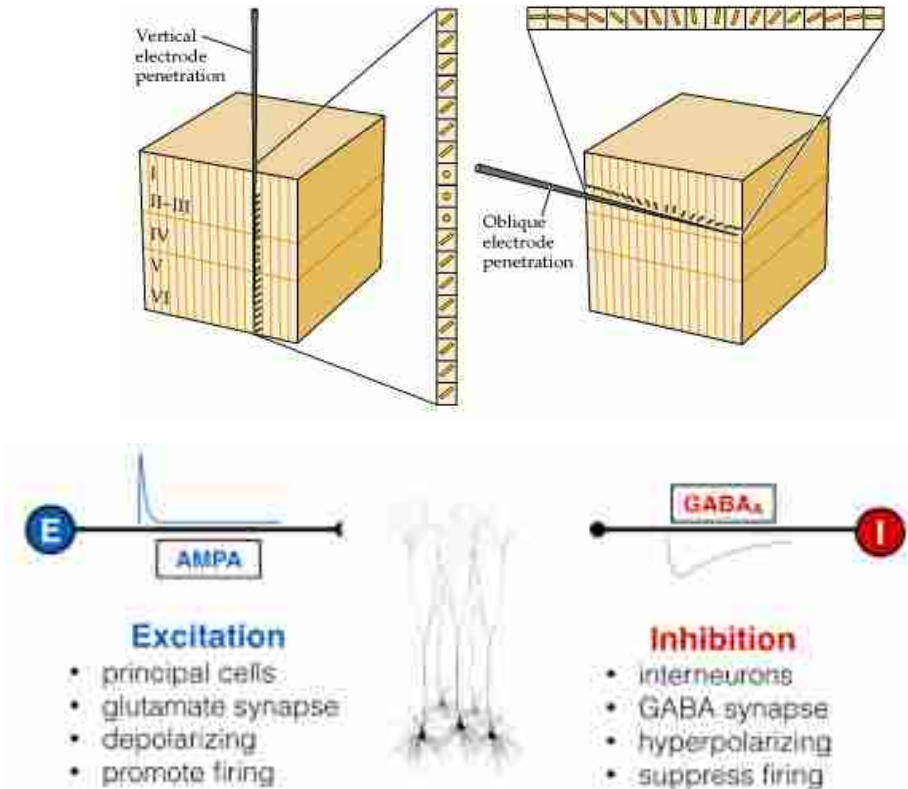


RNN Learning & neural plasticity

- Supervised, unsupervised, and reinforcement learning
- Continual learning, end-to-end seq2seq learning, ...
- Input: sensory input sequence
- Output: decision-making or behavioral readout OR
neural signals (e.g., sequence)
neural + behavioral signals

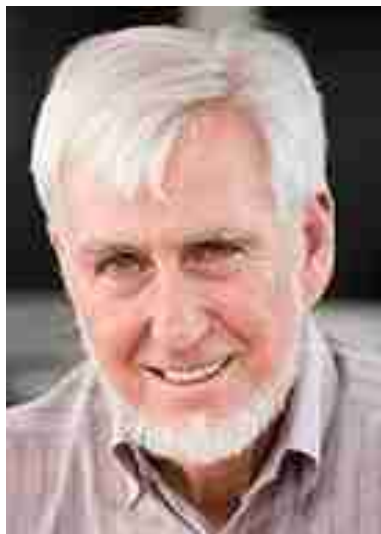
Why biological constraints matter?

- Neuronal networks have structured connectivity
- Structure often determine function
- Neuronal network have excitatory and inhibitory neurons (Dale's principle), and different inhibitory neuron subtypes
- Neuronal networks have cell-type specific or pathway-specific projections
- Synaptic-neuronal coupling: dynamic synapses



Brain GPS system: hippocampal formation (hippocampus-entorhinal cortex)

2014 Nobel Prize winners in Physiology or Medicine



J. O'Keefe



M-B. Moser

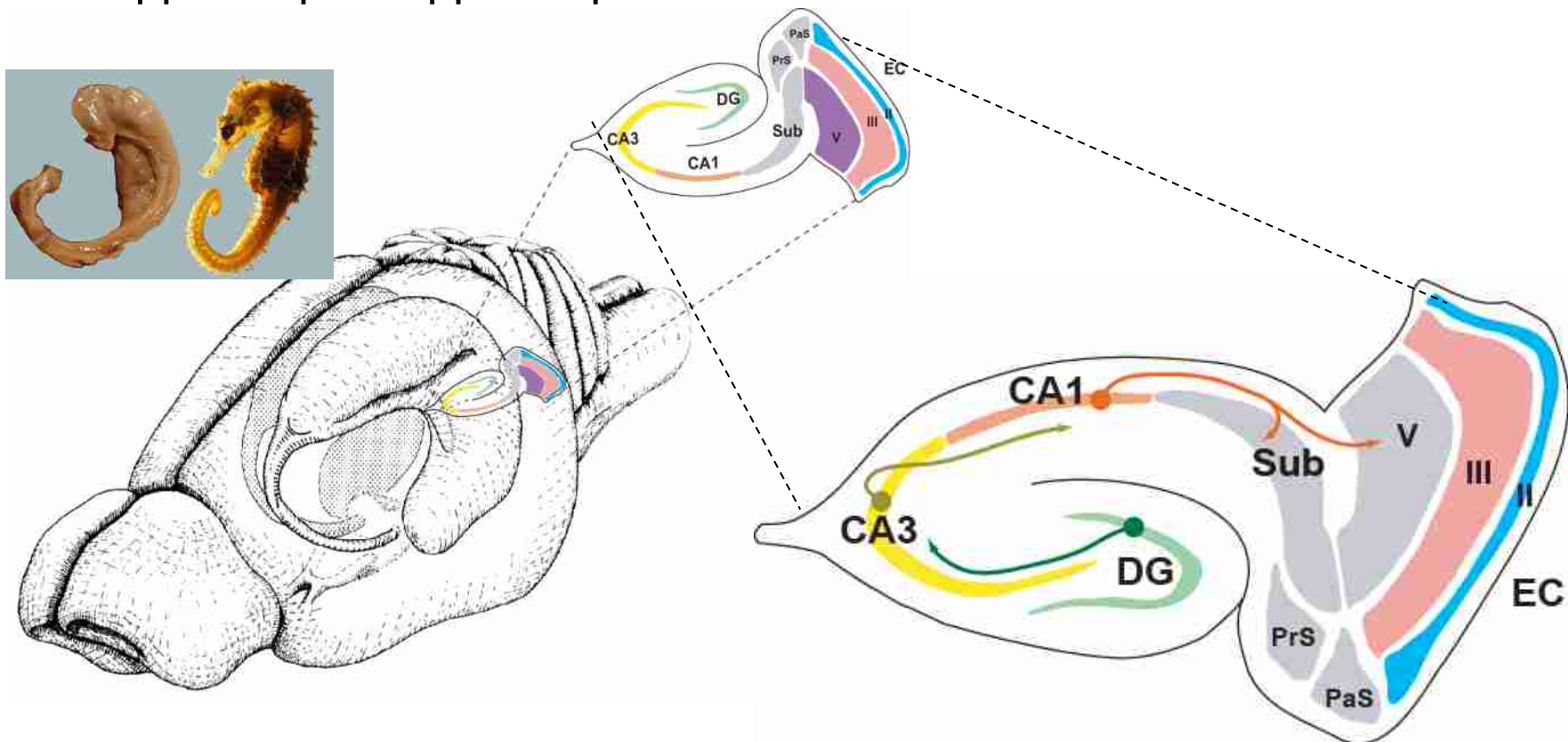
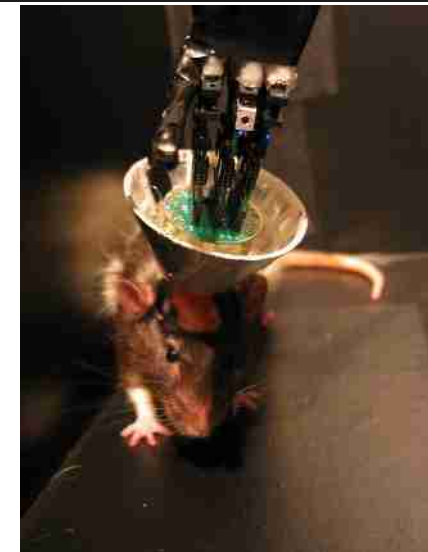


E. Moser

"for their discoveries of cells that constitute a positioning system in the brain".

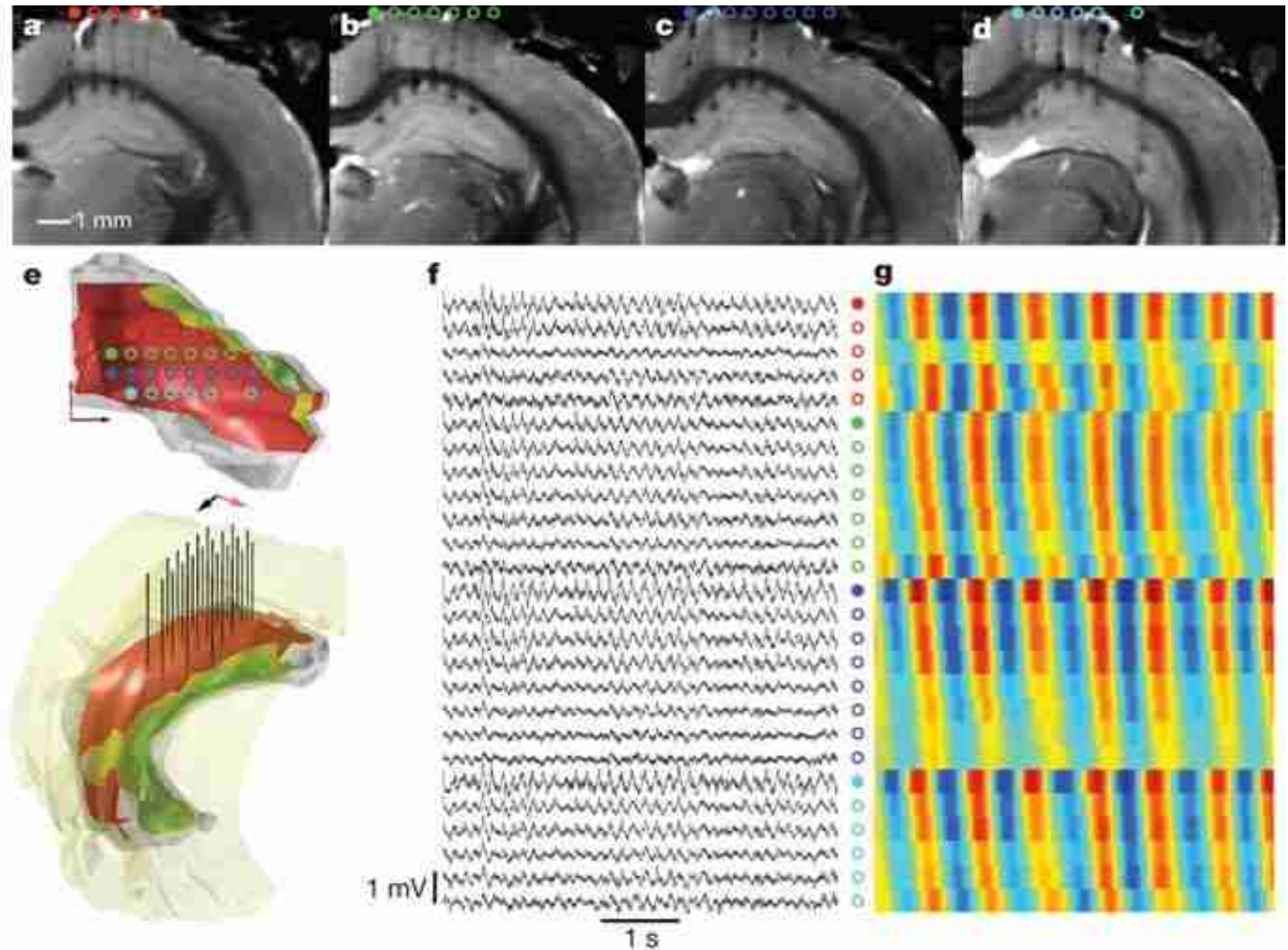
Hippocampus (HPC)

- A model system for studying brain functions of **spatial & episodic memory**
- Tetrode or silicon probe recording of rat dorsal hippocampus/hippocampi



Tetrode 4x8 grid

HPC traveling waves



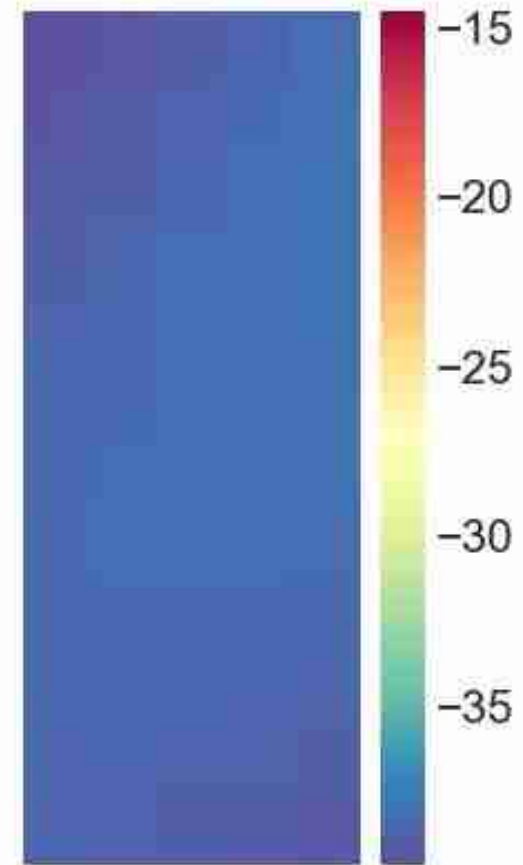
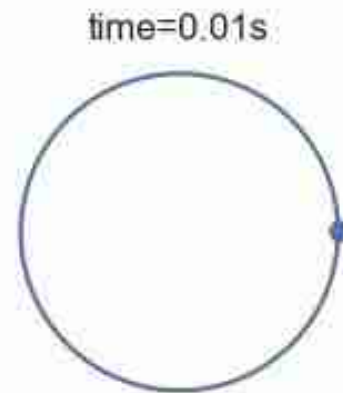
- Lubenov & Siapas. Nature (2009).
- Patel et al. Neuron (2012).
- Patel et al. J. Neurosci. (2013).
- Zhang & Jacobs. J. Neurosci. (2015).

Modeling HPC traveling waves

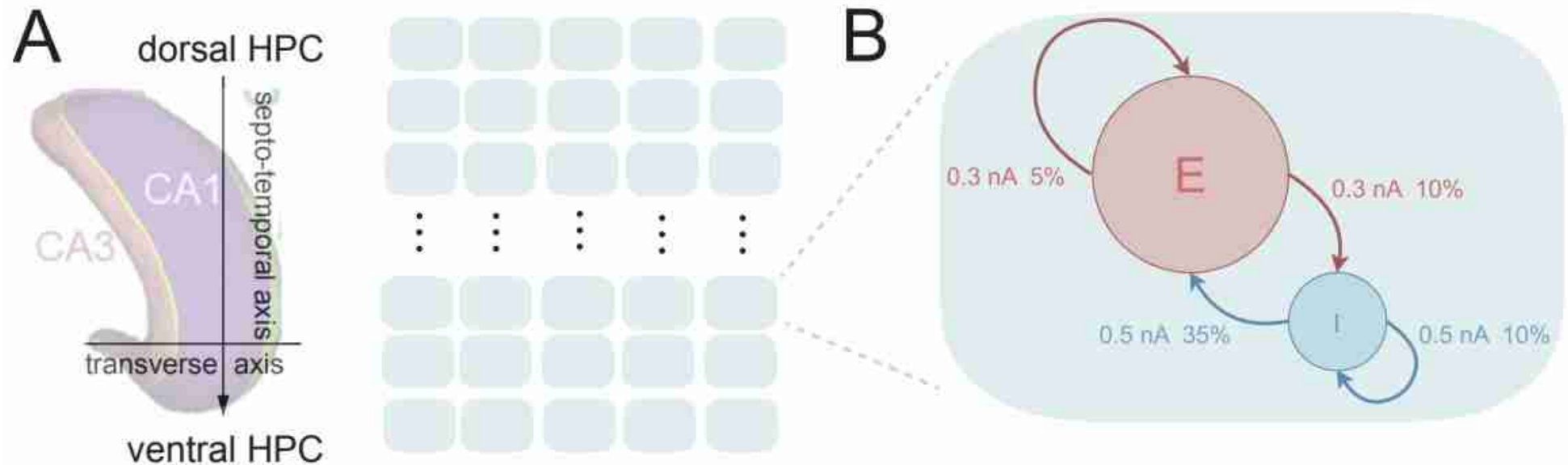


Simulated behavior

- circular track (2 m)
- default speed 0.2 m/s



Model overview



- Septotemporal/ dorsoventral/ longitudinal axis: dorsal (septal), intermediate, and ventral (temporal) segments
- Proximodistal/ transverse axis
- **12×5 blocks: each consisting of 45 leaky integrate-and-fire (LIF) excitatory neurons and 15 LIF inhibitory neurons**

Differential equations for excitatory and inhibitory neurons

- Leaky integrate-and-fire (LIF) neurons

$$\dot{V} = \frac{E_l - V}{\tau_m} + \frac{I_{exc} - I_{inh} + I_{st}}{C_m}$$
$$\dot{I}_{exc} = \frac{-I_{exc}}{\tau_{exc}}, \quad \dot{I}_{inh} = \frac{-I_{inh}}{\tau_{inh}}$$

Excitatory place cells

$$I_{st_e} = I_{0_e} \cdot \left(1 + gain \cdot \exp \left(- \frac{(m_{RF} - position(t))^2}{2l_{RF}^2} \right) \right)$$

Inhibitory theta cells

$$I_{st_i} = I_{0_i} \cdot (1 + gain \cdot \sin(2\pi f t))$$

Run speed modulation

$$gain = gain_0 + \gamma \cdot (v - v_0) e^{-\frac{n}{18}}$$

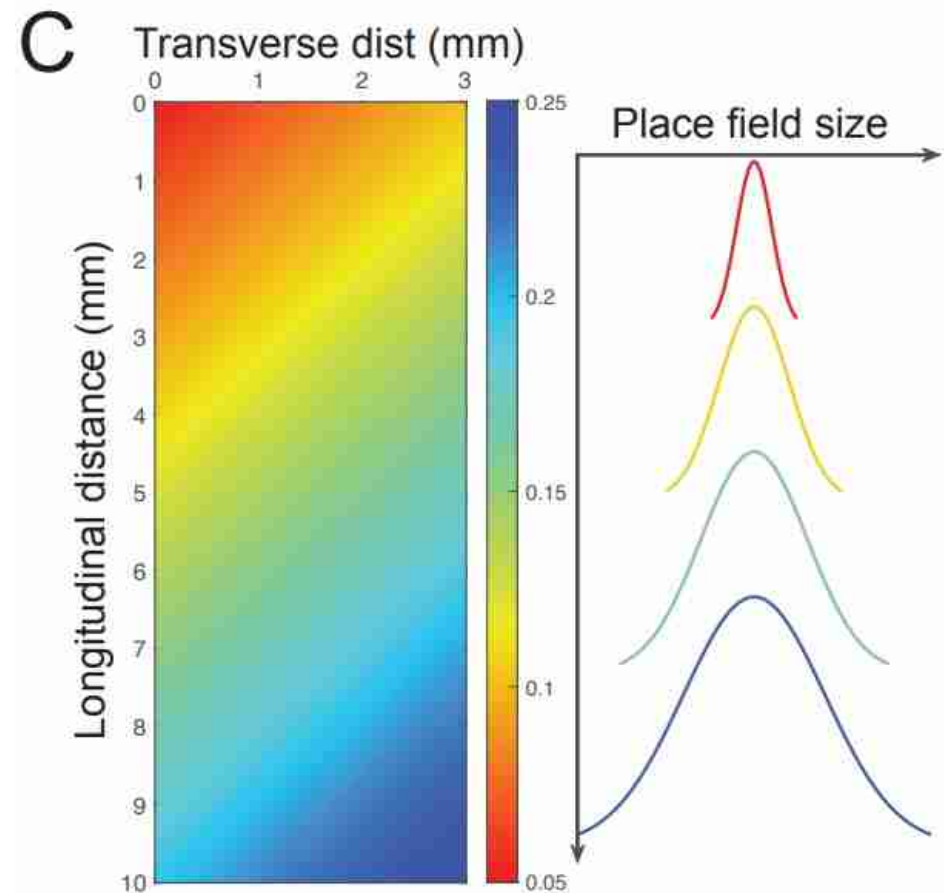
Model assumption

Along the dorsoventral axis:

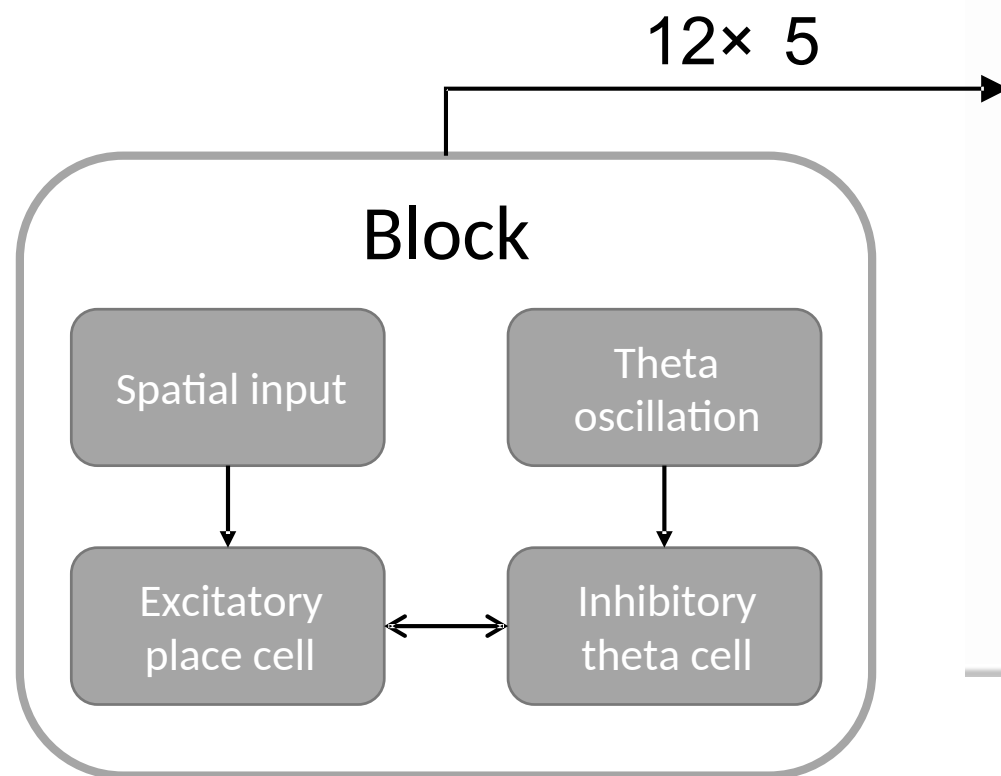
- place field sizes increase
- the magnitude of the excitatory spatial input and inhibitory theta oscillation drive decrease

Between neighboring blocks:

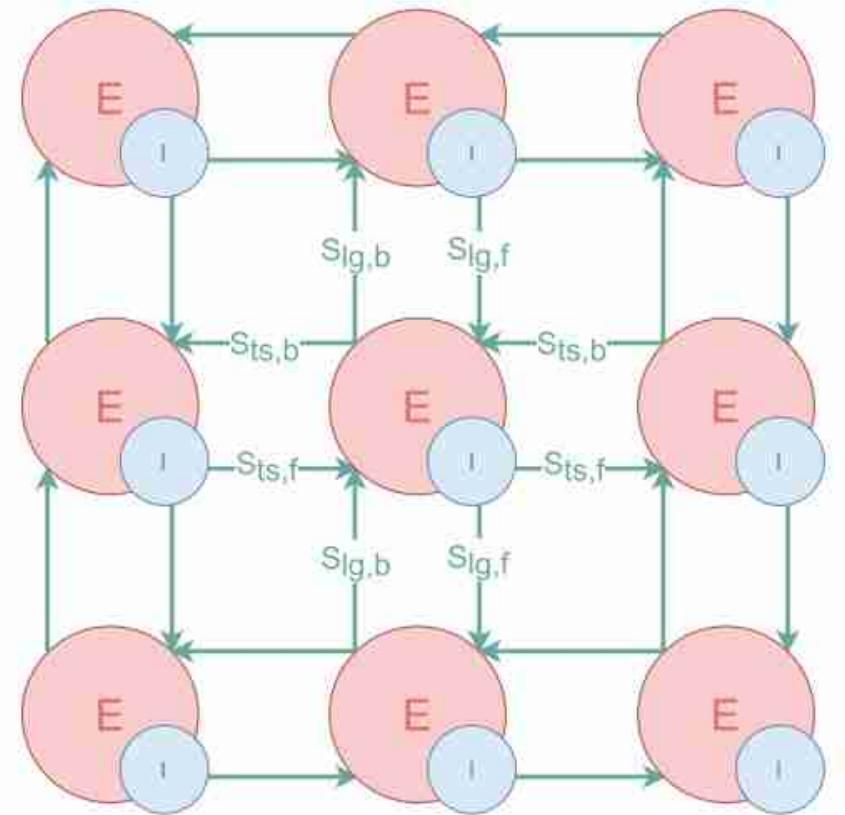
- long-range reciprocal yet asymmetric (**excitatory or inhibitory**) connectivity



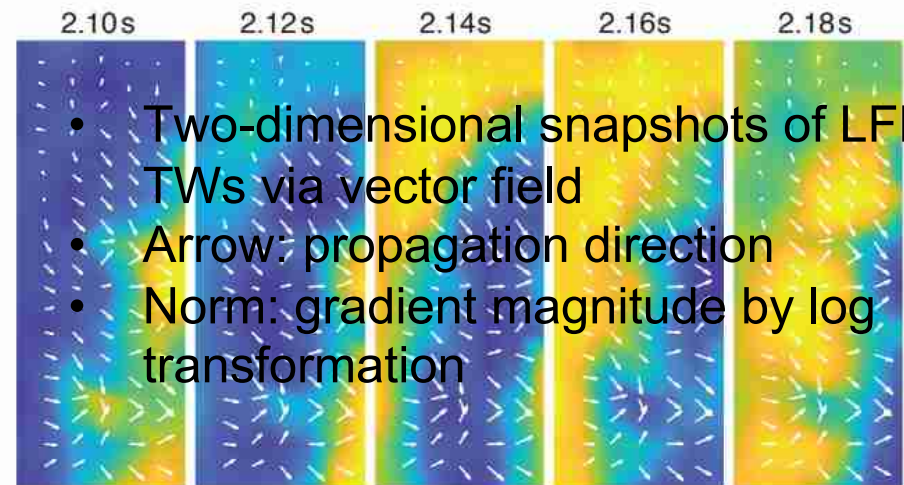
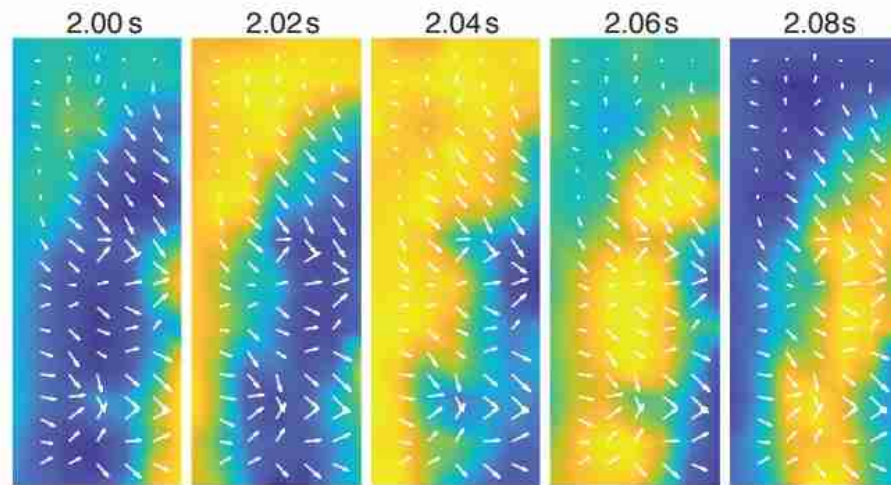
Sum-up



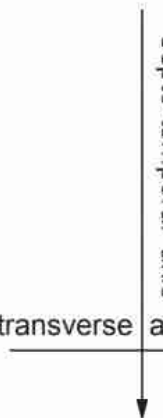
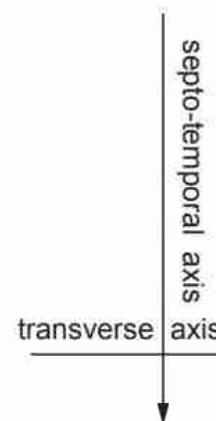
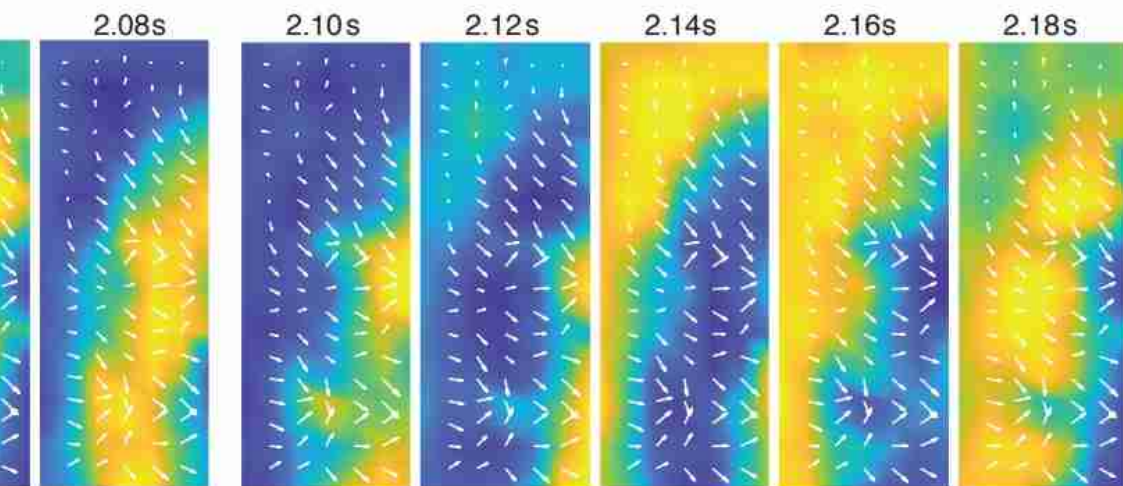
D



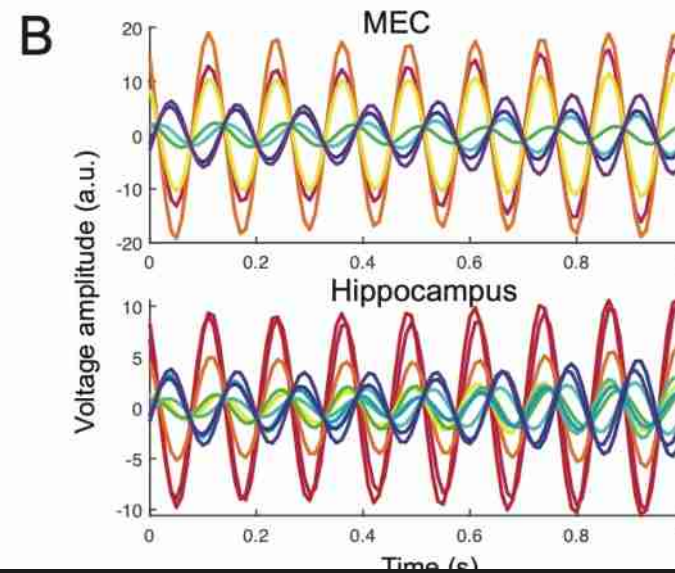
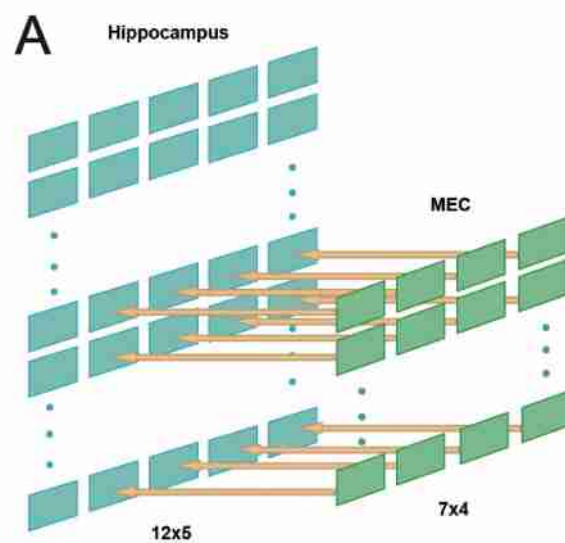
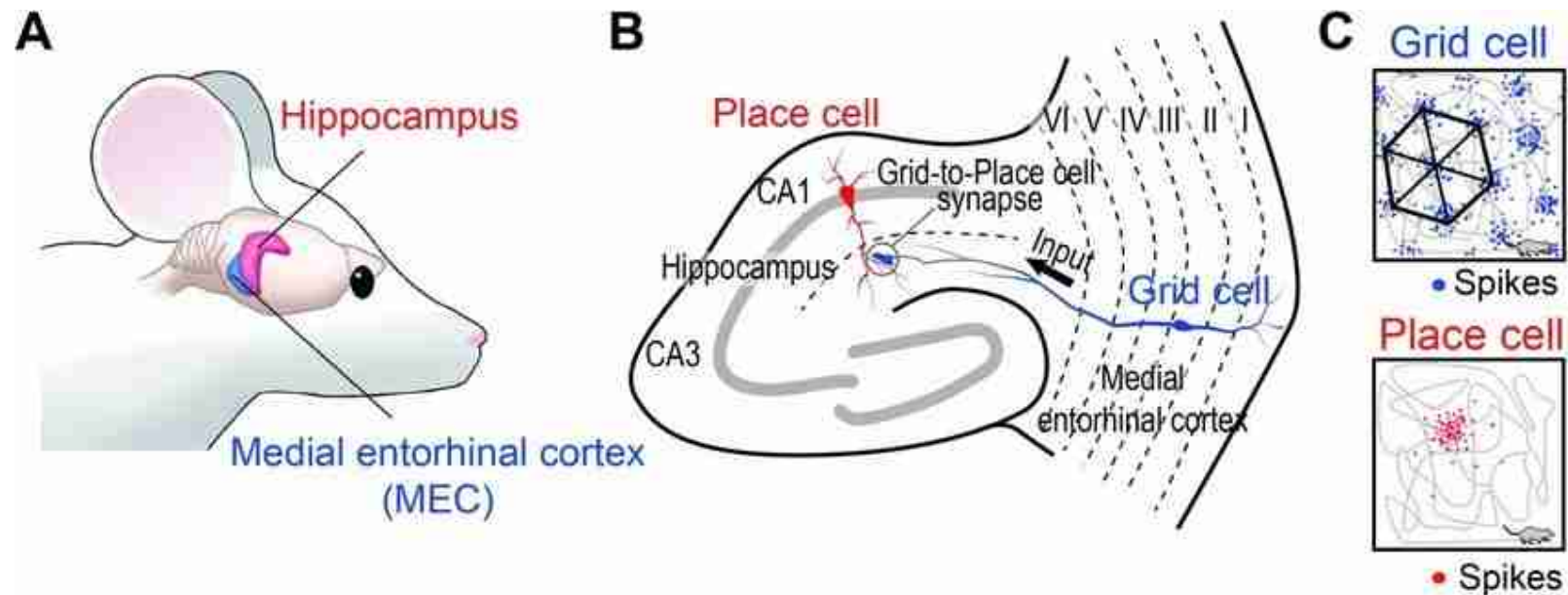
D



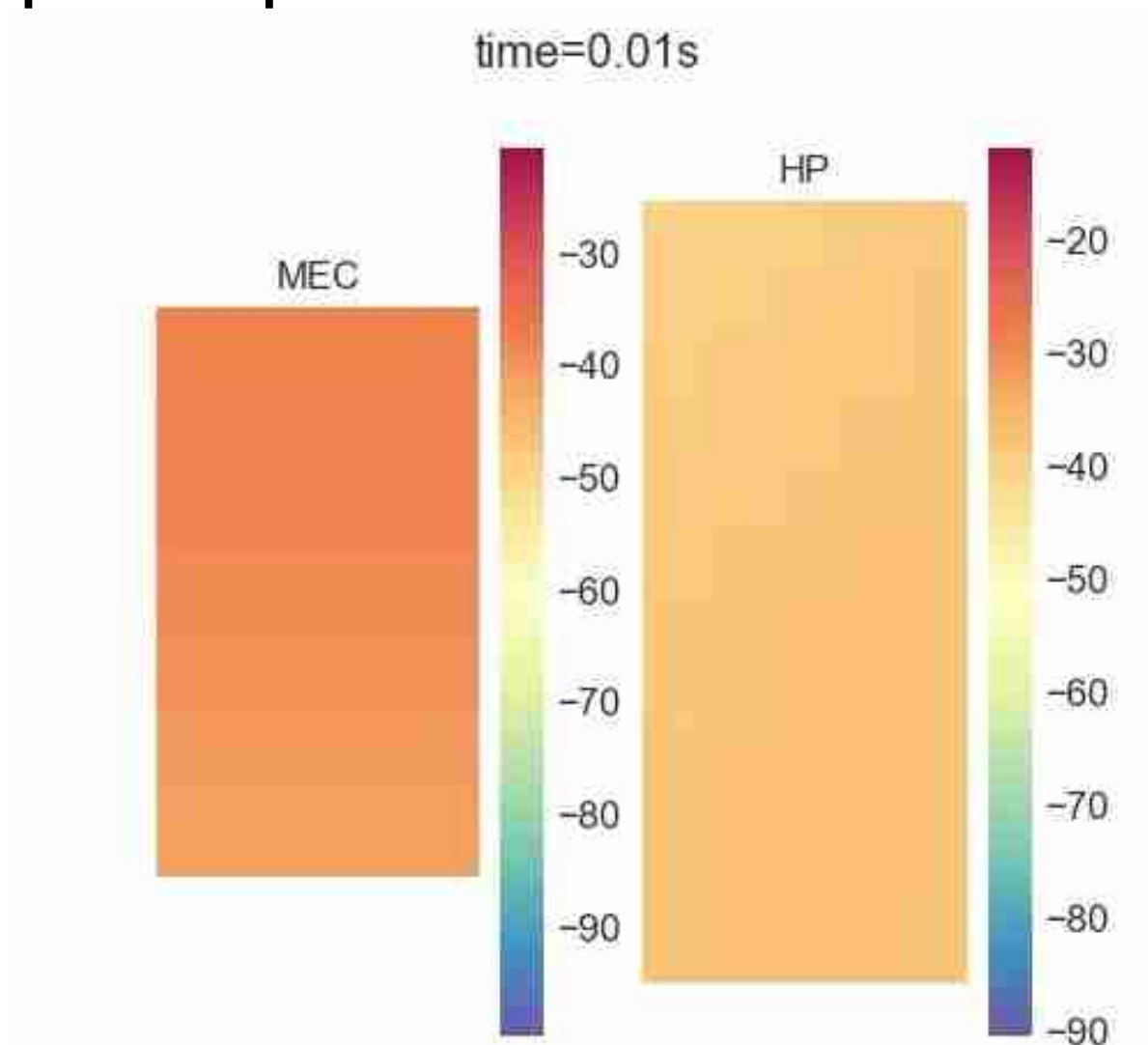
- Two-dimensional snapshots of LFP TWs via vector field
- Arrow: propagation direction
- Norm: gradient magnitude by log transformation



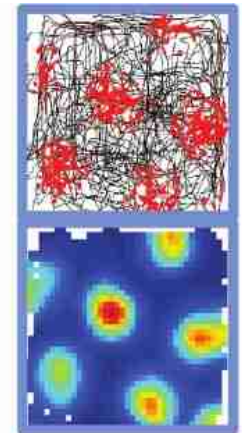
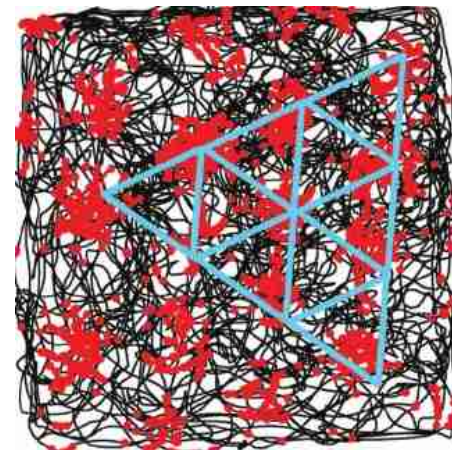
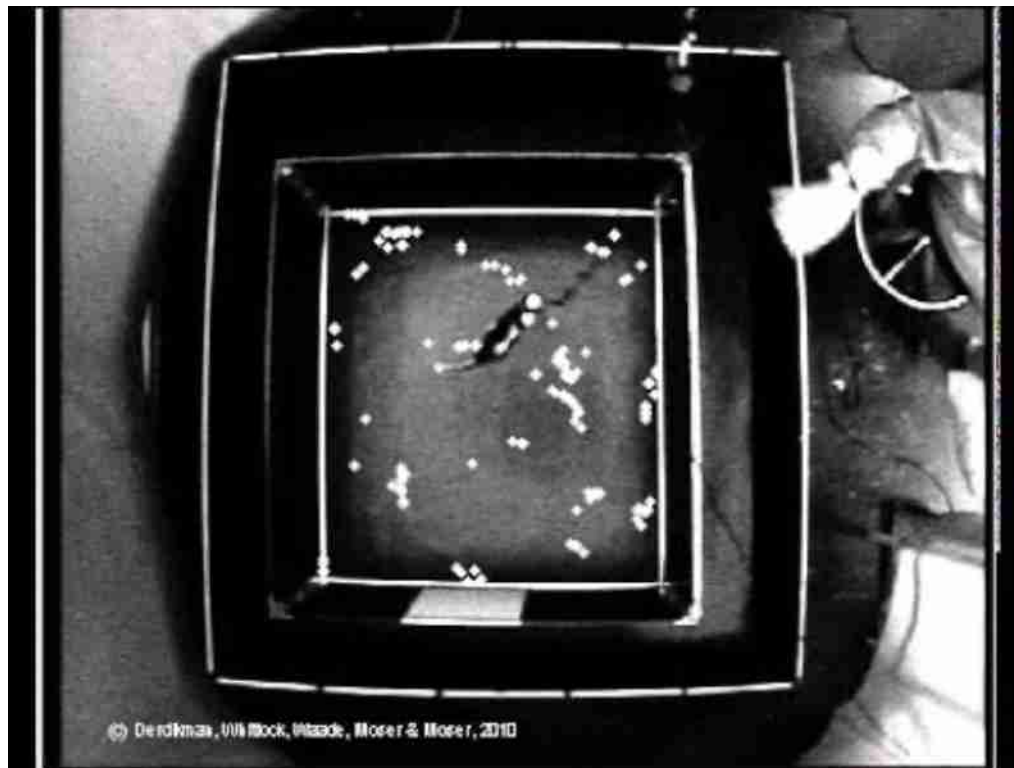
Extension to MEC traveling waves



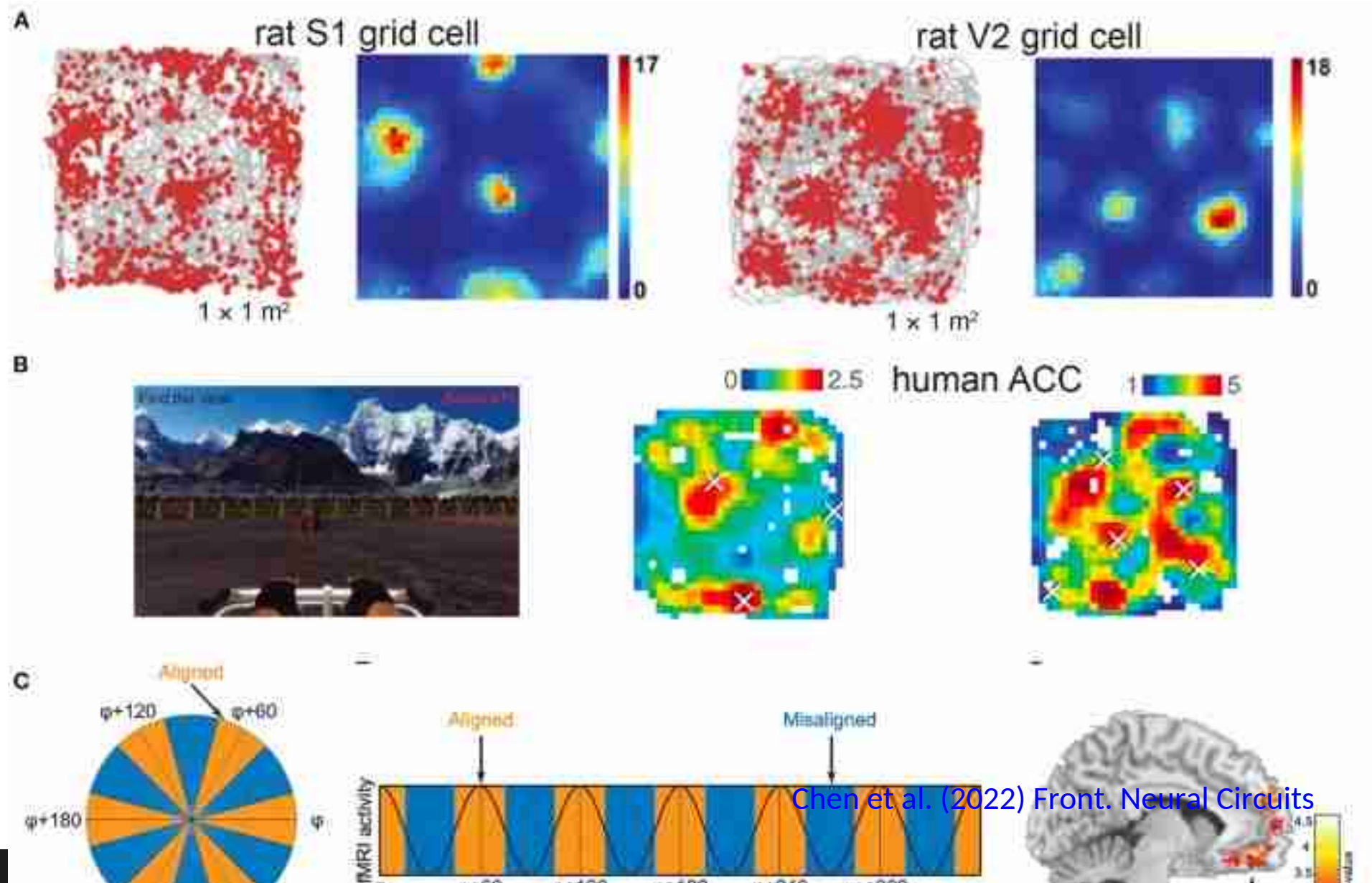
Concurrent MEC-HPC traveling waves in hippocampal formation



Example of grid cell



Grid cells outside MEC



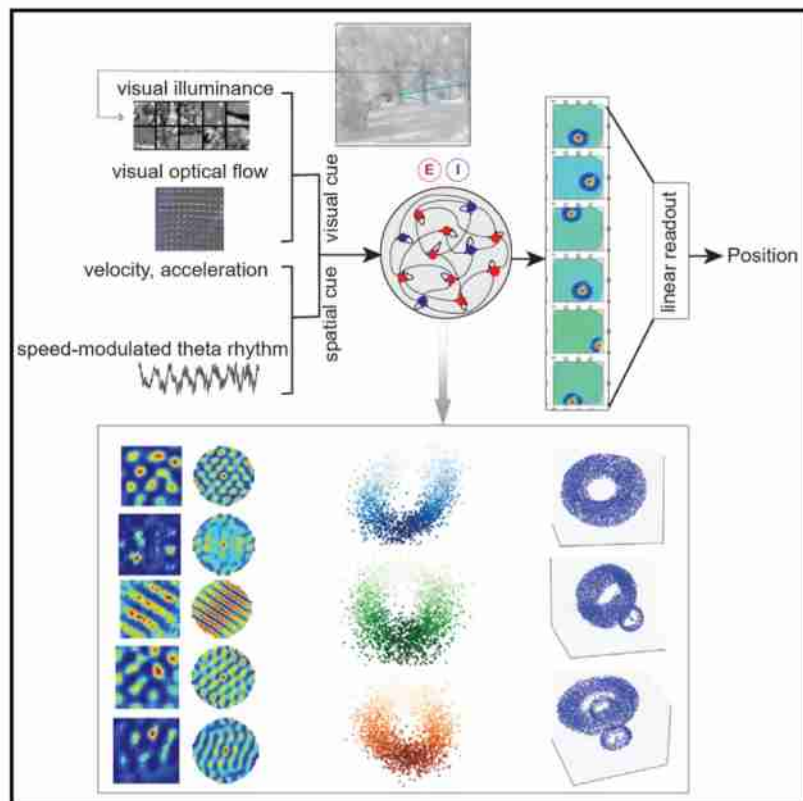
Modeling V2 cortical grid cells



Cell Reports

Excitatory-inhibitory recurrent dynamics produce robust visual grids and stable attractors

Graphical abstract



Authors

Xiaohan Zhang, Xiaoyang Long,
Sheng-Jia Zhang, Zhe Sage Chen

Correspondence

zhe.chen@nyulangone.org

In brief

Zhang et al. train biologically realistic neural network models to perform a spatial navigation task with multisensory input and discover emergent grid-like responses and a low-dimensional, multistable torus-like attractor with imposed network connectivity and sparsity constraints.

Model assumption

- RNN receives multimodal sensory input to produce behavior (animal's location)

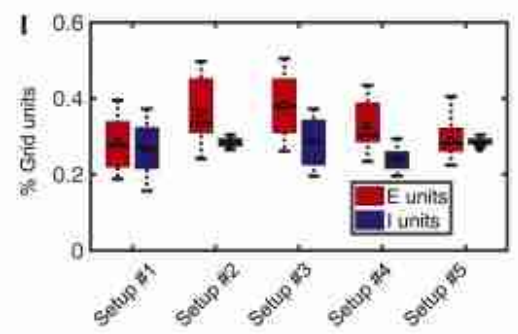
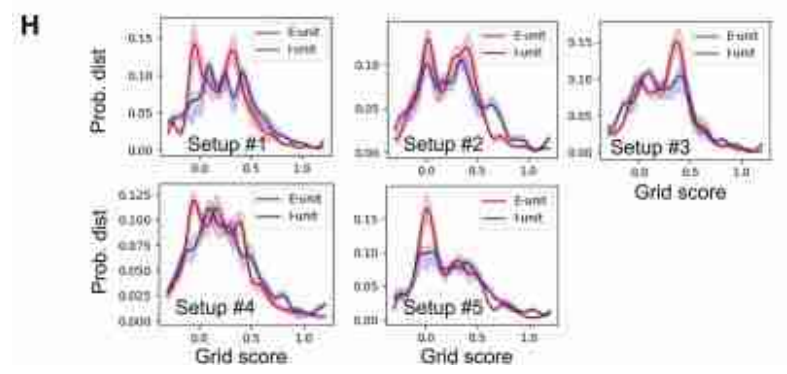
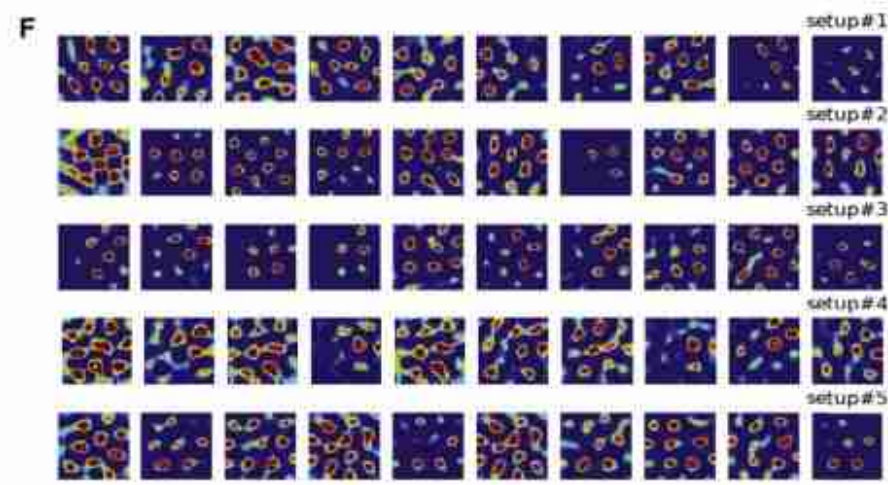
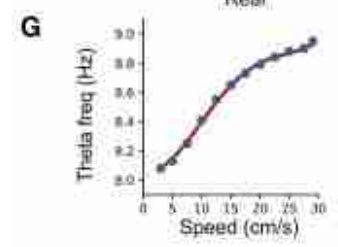
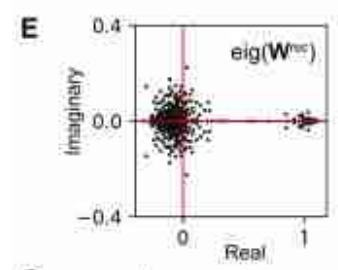
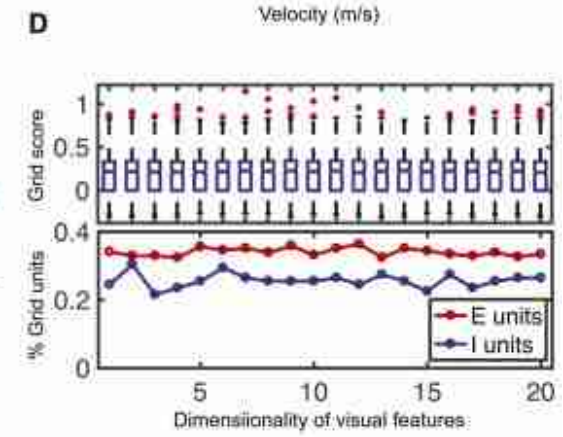
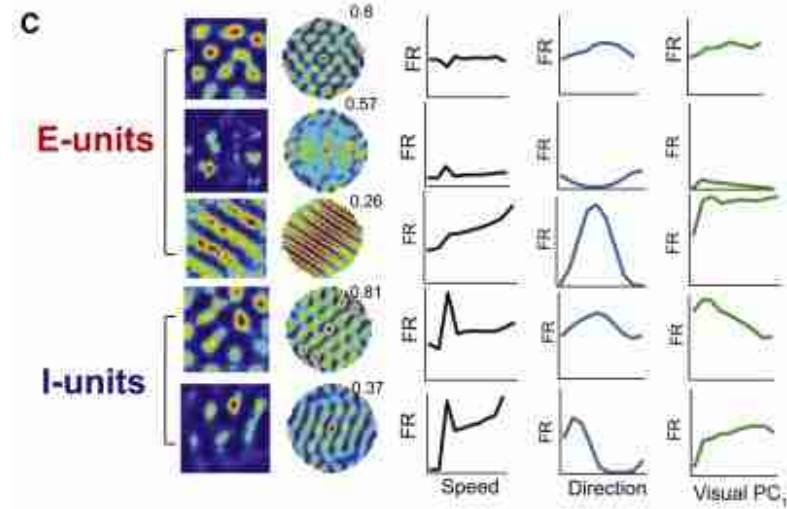
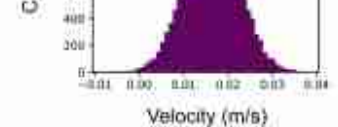
Table 1: Testing various input configurations in the E/I-RNN.

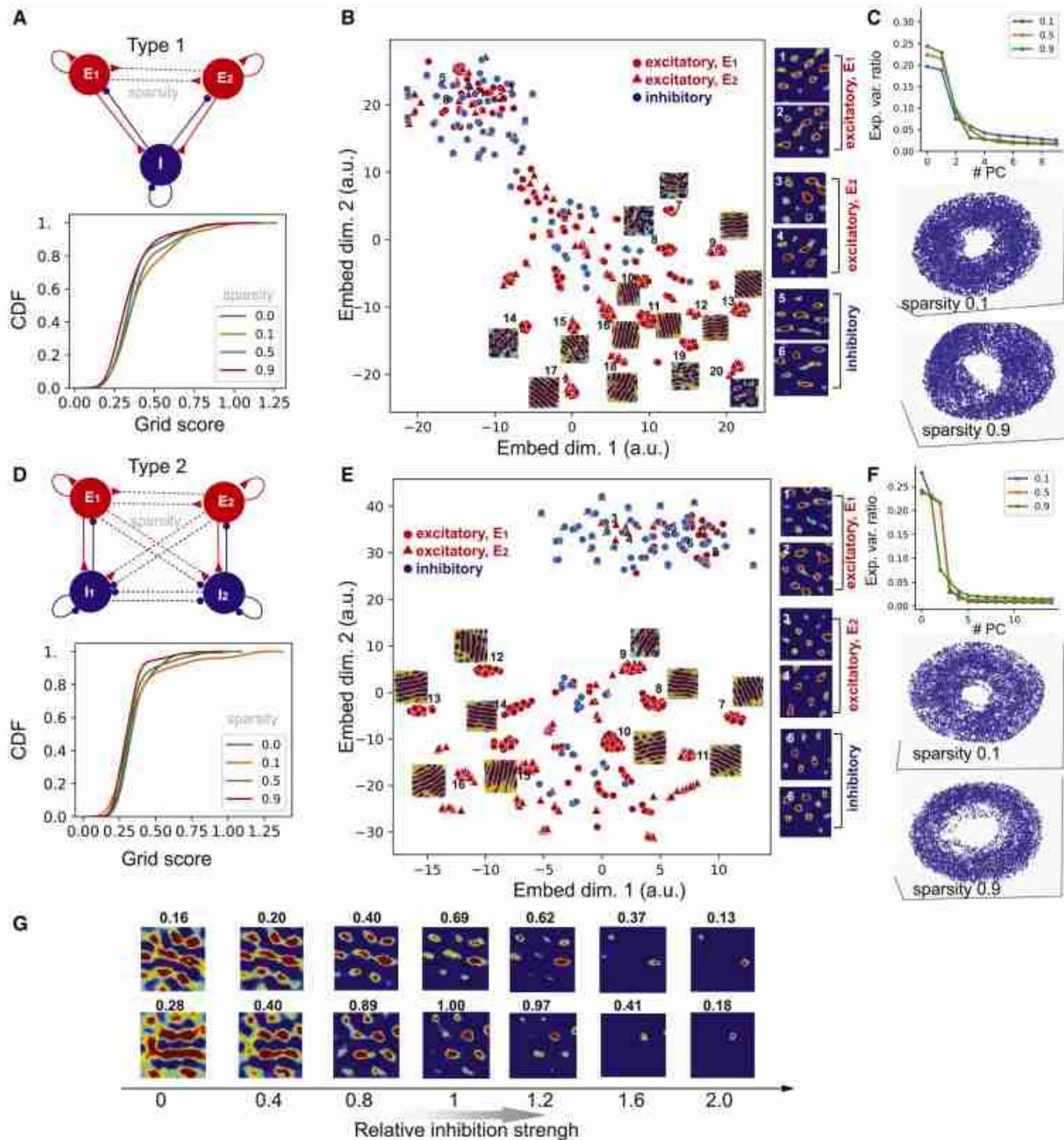
Setup	RNN Input	Dimensionality N_{in}
#1	Velocity	2
#2	Velocity + Visual luminance	$2+(1-20)$
#3	Acceleration + Visual luminance	$2+20$
#4	Theta freq + direction + Visual luminance	$1+1+20$
#5	Visual optical flow	512

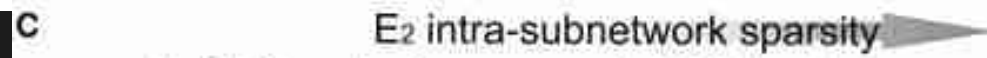
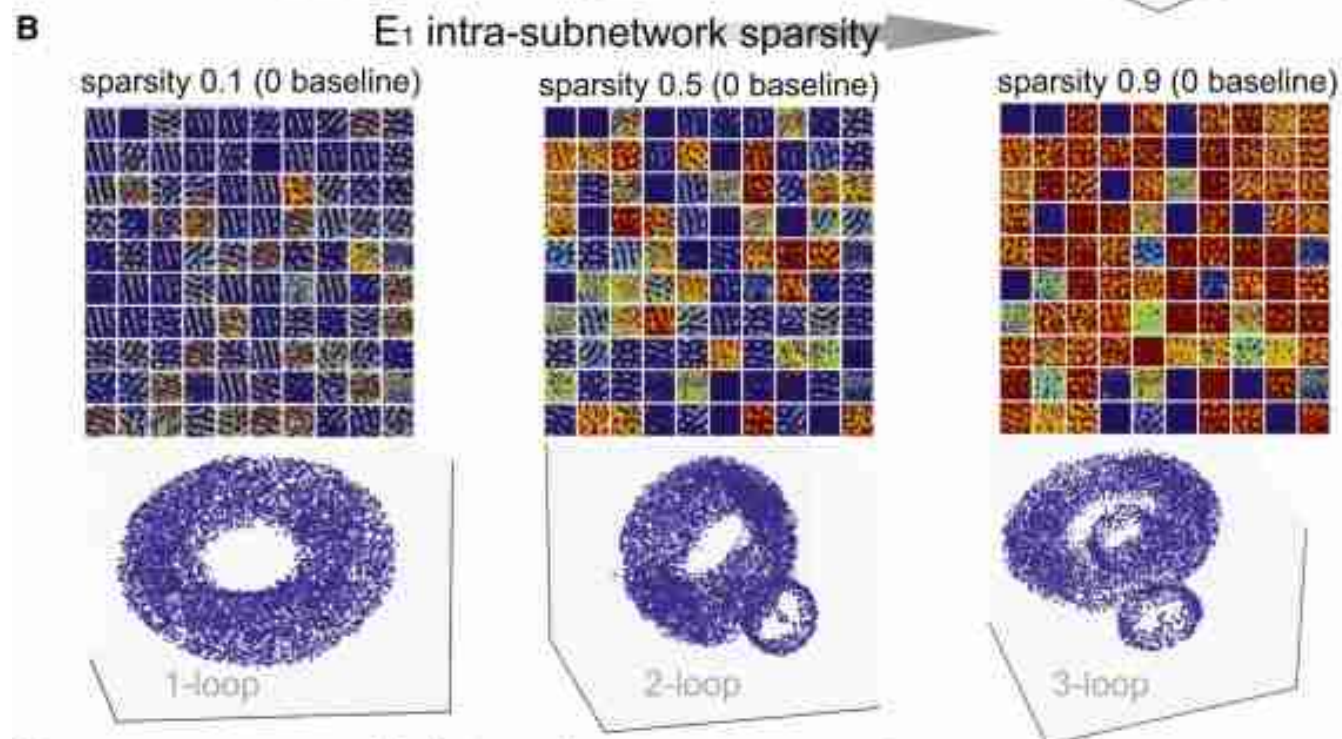
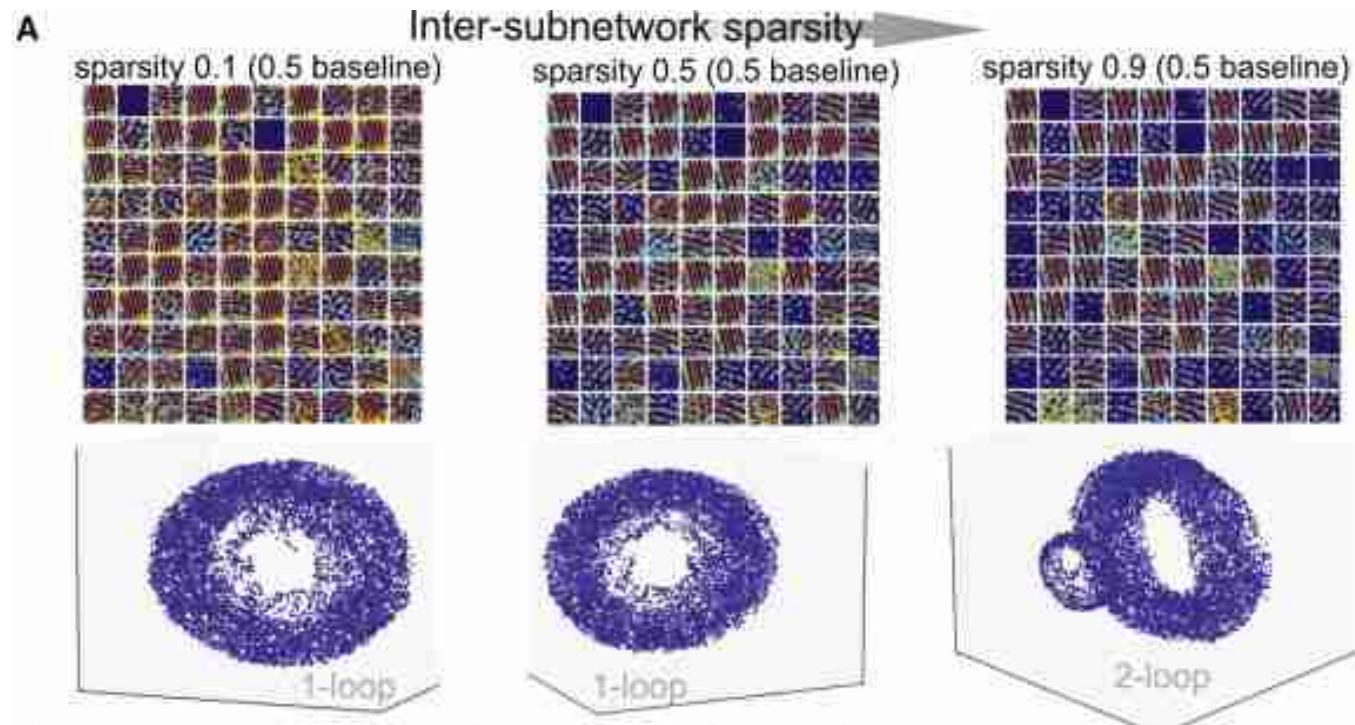
speed-modulated theta rhythm



80%







Spiking RNN for modeling PFC working memory and divided attention in decision making

- Can we use biologically constrained SRNN to produce experimental findings in task-performing mice?



Cognitive Computation (2023) 15:1167–1189
<https://doi.org/10.1007/s12559-022-09994-2>

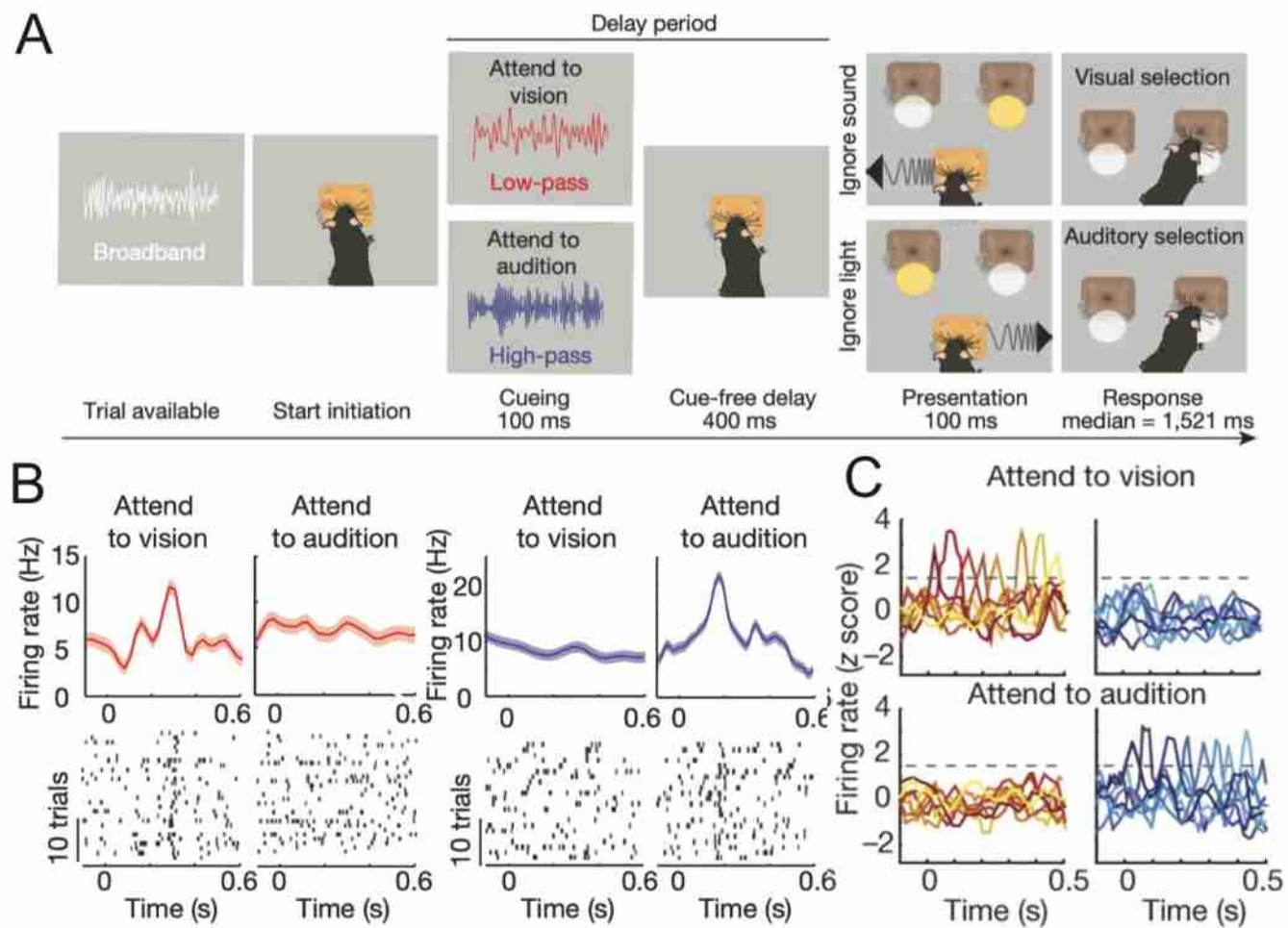


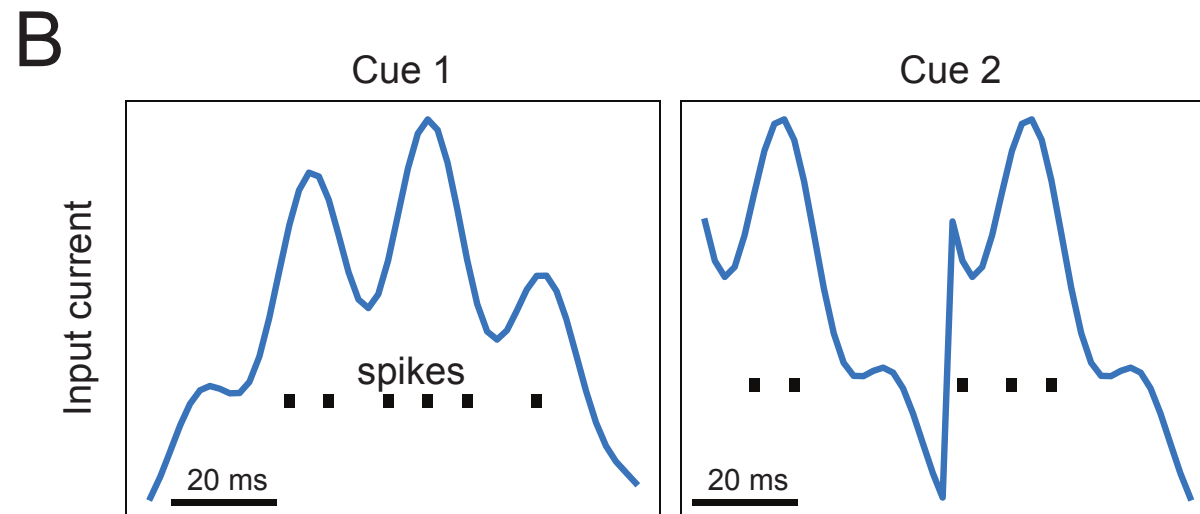
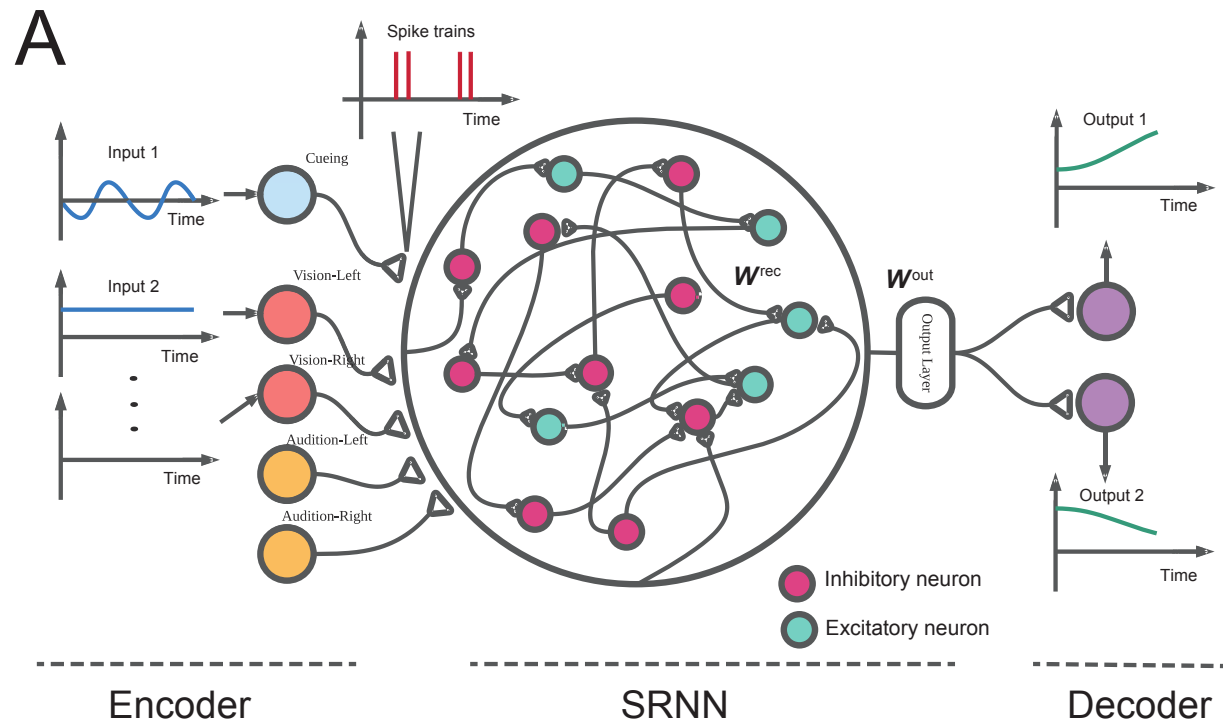
Spiking Recurrent Neural Networks Represent Task-Relevant Neural Sequences in Rule-Dependent Computation

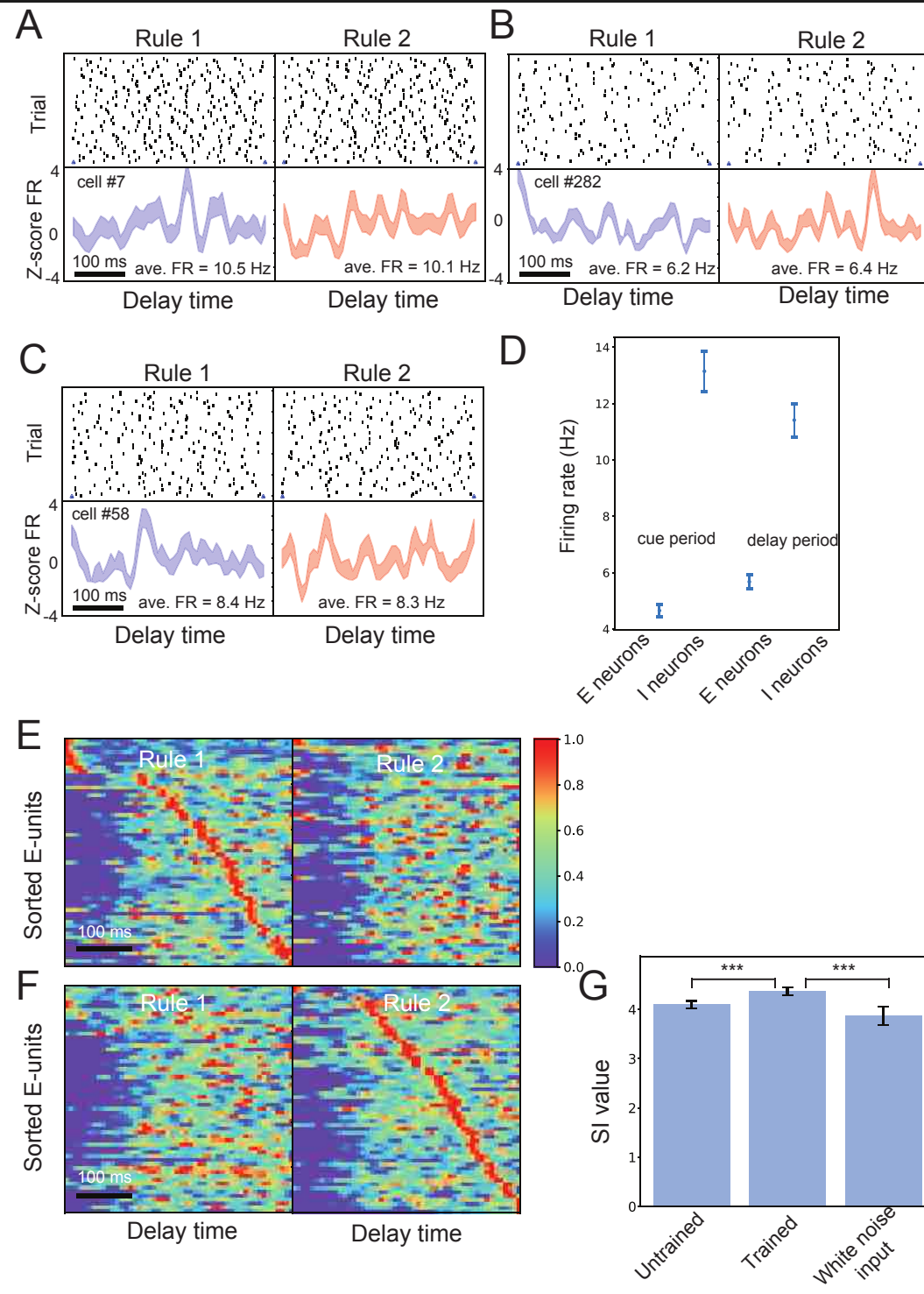
Xiaohe Xue¹ · Ralf D. Wimmer² · Michael M. Halassa² · Zhe Sage Chen^{3,4,5}

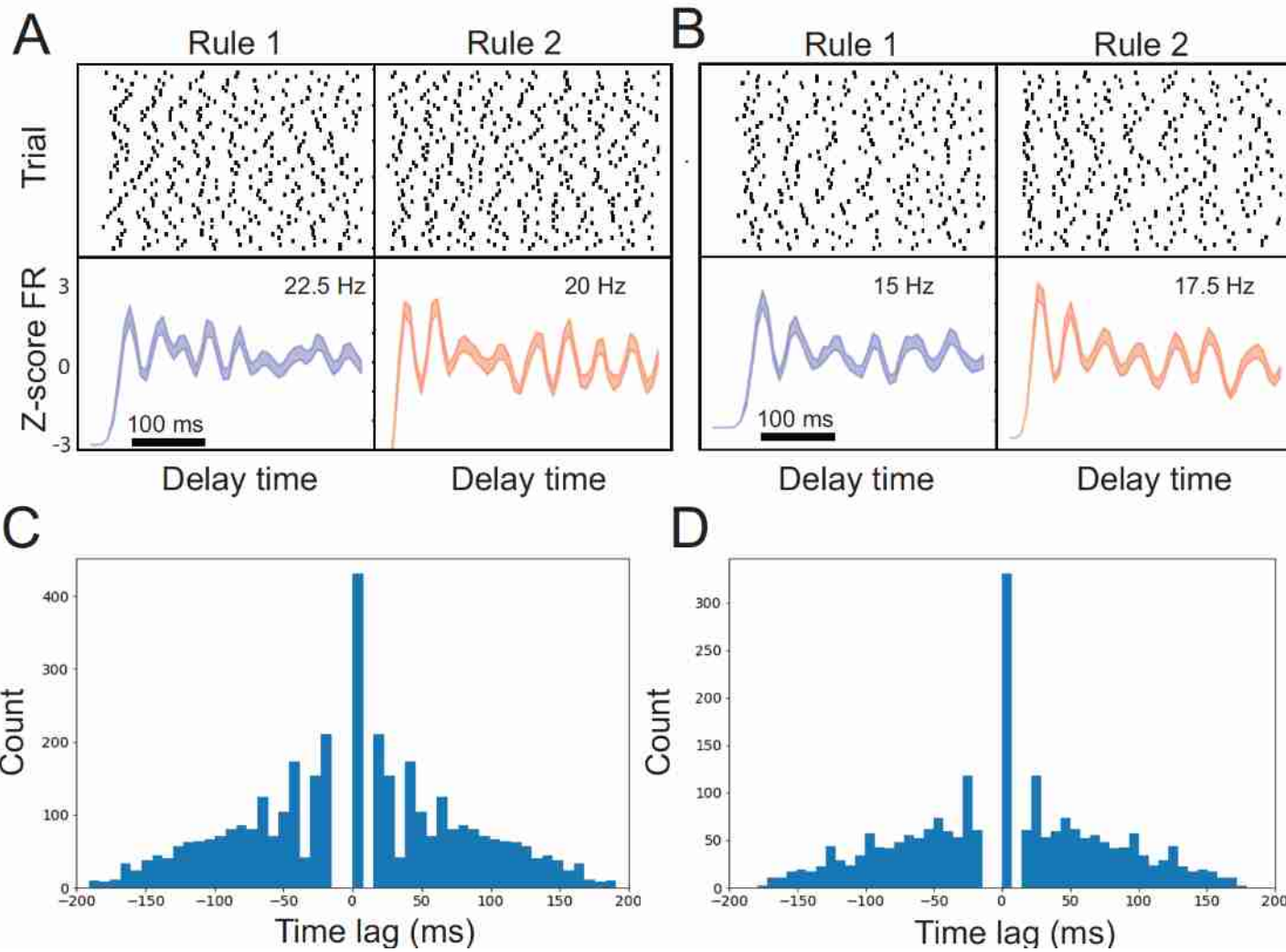
Received: 16 August 2021 / Accepted: 13 January 2022 / Published online: 5 February 2022

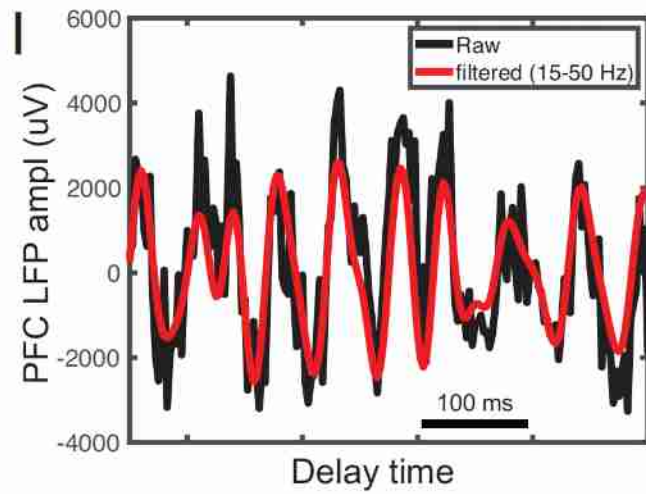
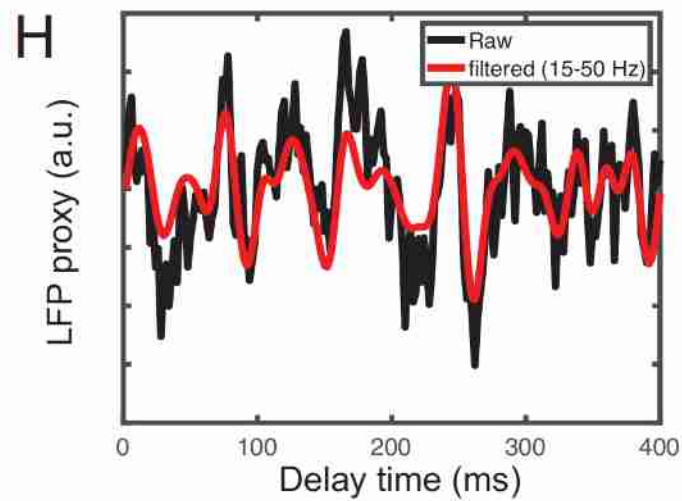
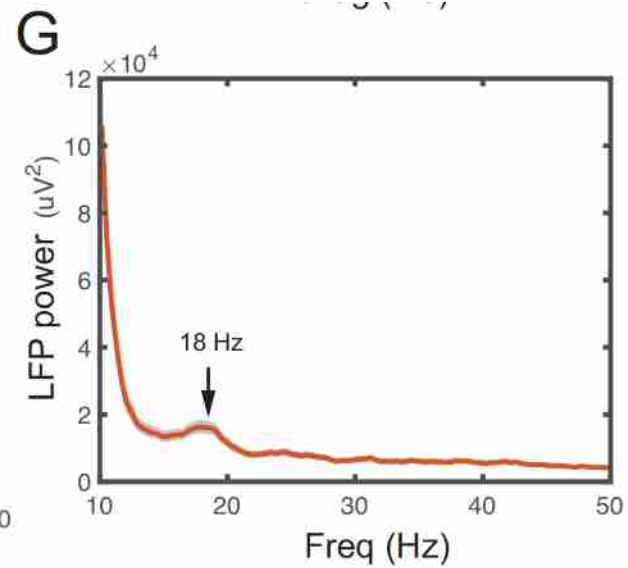
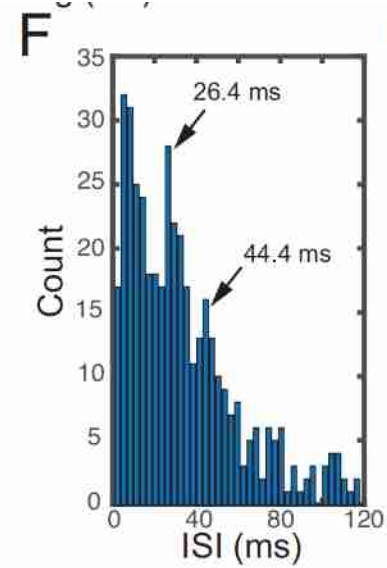
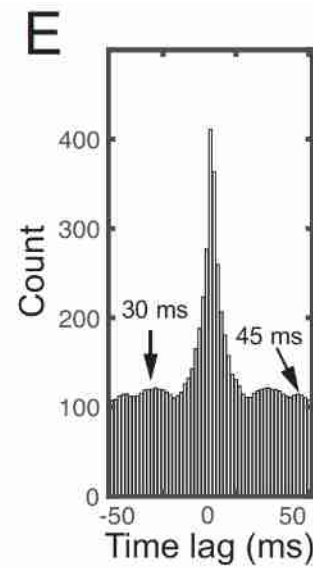
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022













Mediodorsal thalamus regulates task uncertainty to enable cognitive flexibility

Received: 23 October 2023

Accepted: 7 March 2025

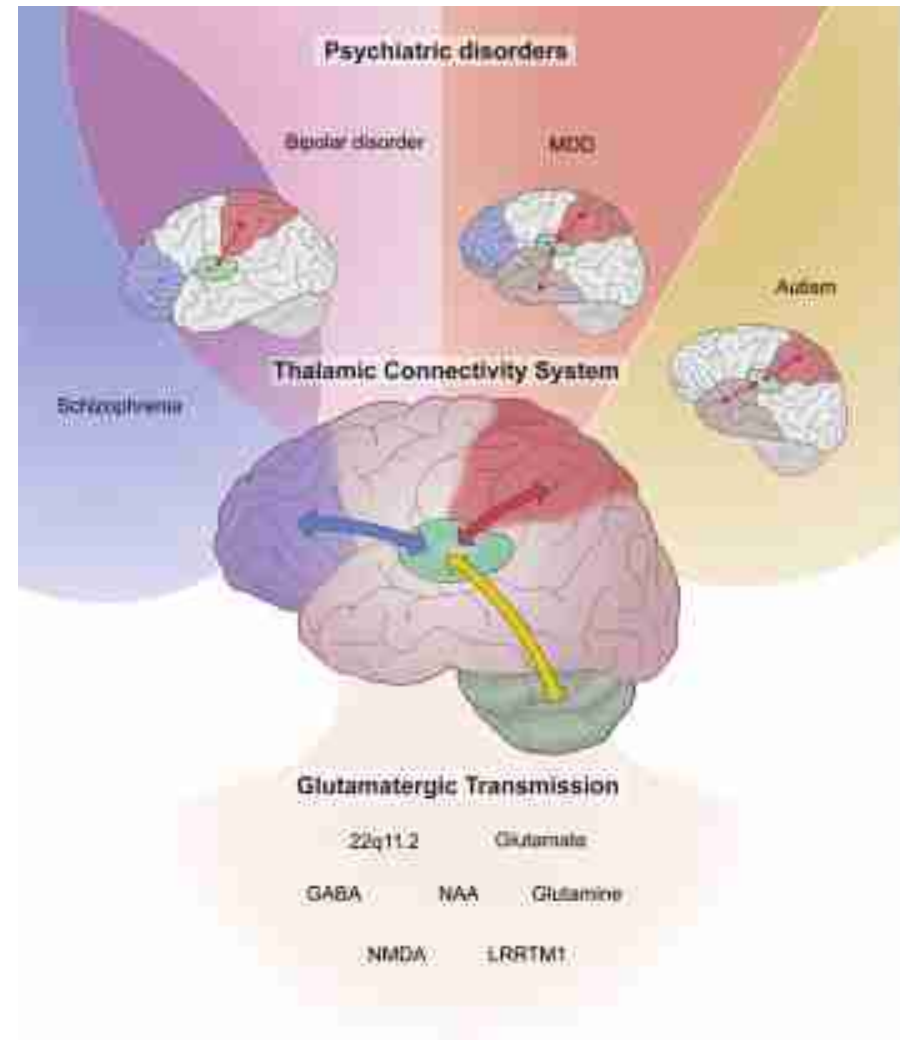
Xiaohan Zhang¹, Arghya Mukherjee², Michael M. Halassa² &
Zhe Sage Chen^{1,3,4,5} ✉



Rikhye RV, Gilra A, & Halassa MM (2018). Nat. Neurosci. 21, 1753-1763 (2018).
Mukerjee A, Lam NH, Wimmer RD, & Halassa MM. Nature, 600, 100-104 (2021).
Uddin LQ. Nat. Rev. Neurosci. (2021)
Waltz JA. Neuroscience (2017)
Standke et al. Neuroimage: Clinical (2021)

Mediodorsal (MD) thalamus

- A critical partner with prefrontal cortex (PFC) in cognitive and executive control
- Dysfunctional MD connectivity is implicated in many mental disorders
- **Question:** How to dissect circuit mechanisms (e.g., cell type specificity, thalamocortical connectivity) that link to cognitive deficits?

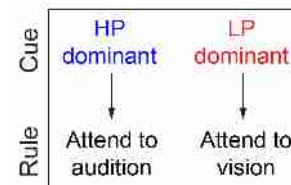
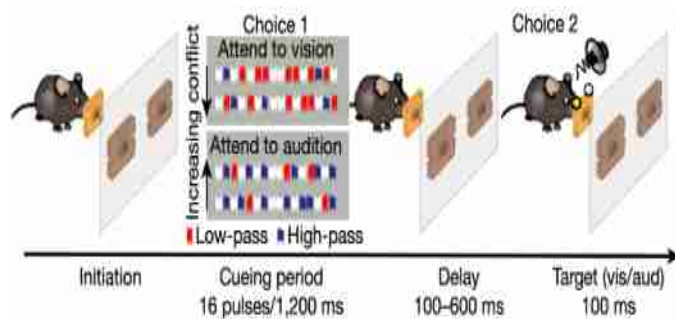
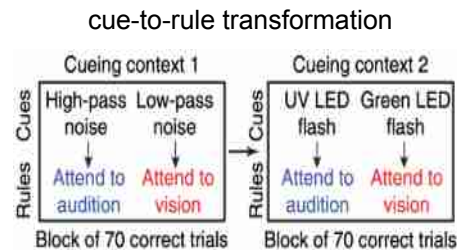
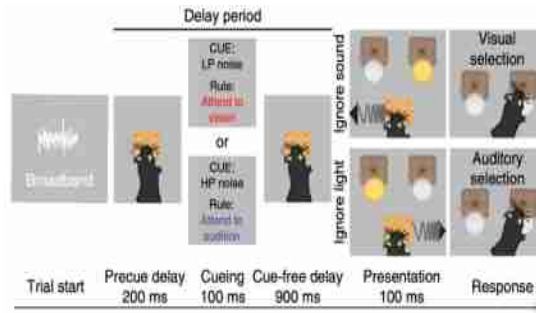


Flexible decision making and task un



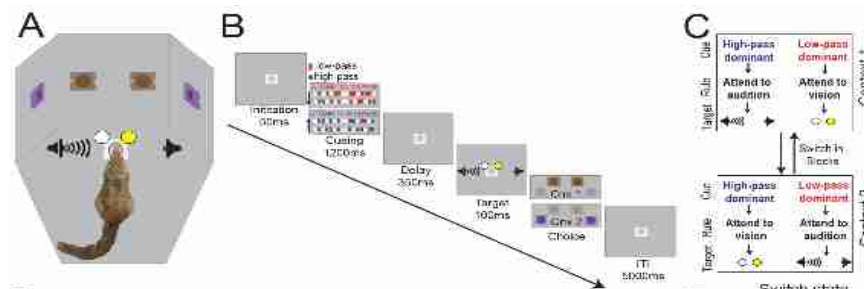
Mapping uncertainty

Rikhye et al. (2018) Nat. Neurosci.



Cueing uncertainty

Mukerjee et al. (2021) Nature

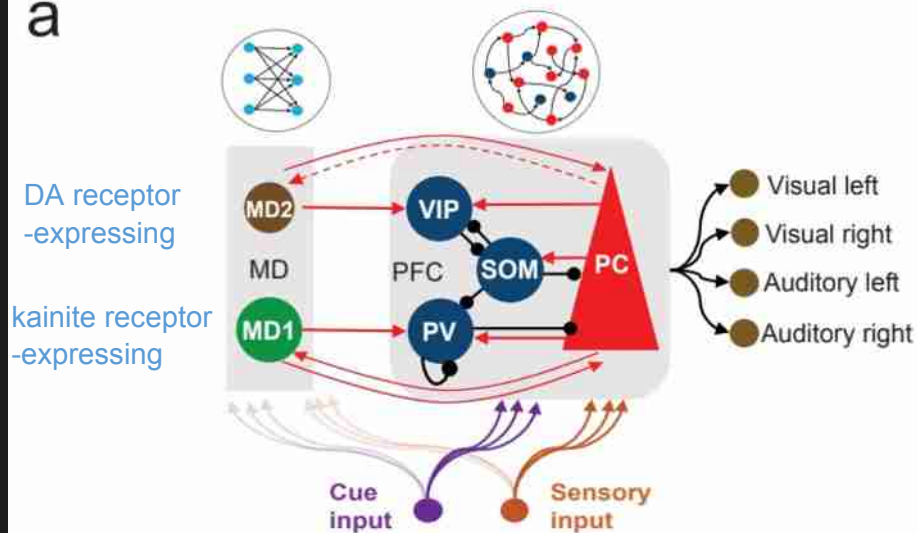


Cueing + Mapping uncertainties

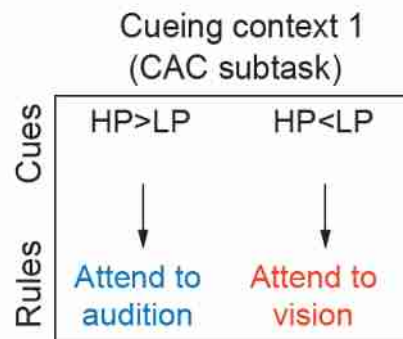
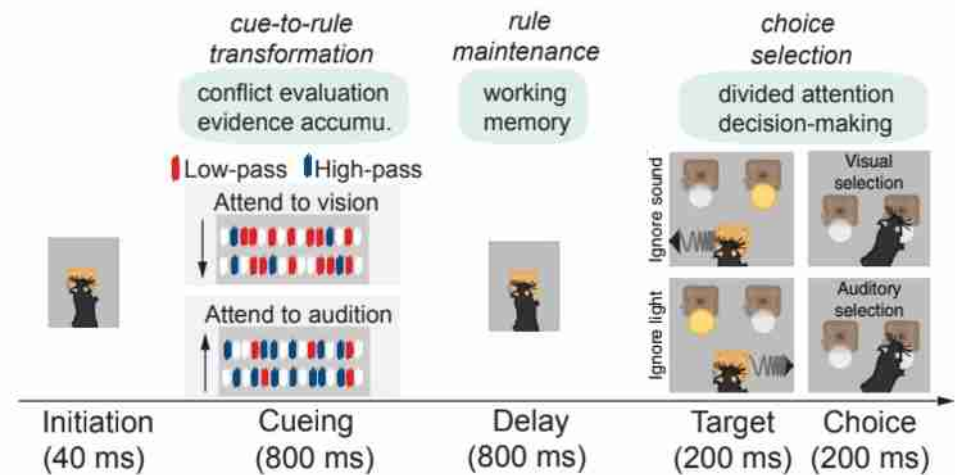
Lam et al. (2025) Nature

Cell-type specific circuit modeling

a



b



Congruence: 0.57

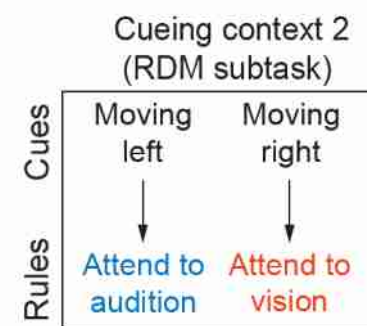
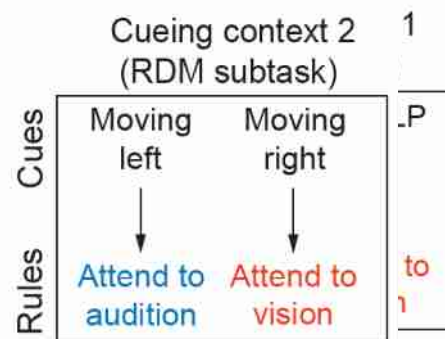
Signal density: 7

Congruence: 0.85

Signal density: 7

Congruence: 0.57

Signal density: 14



RDM subtask

Coherence: +0.6

Signal density: 10

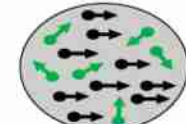
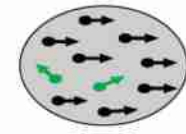
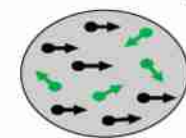
Coherence: +0.8

Signal density: 10

Coherence: +0.6

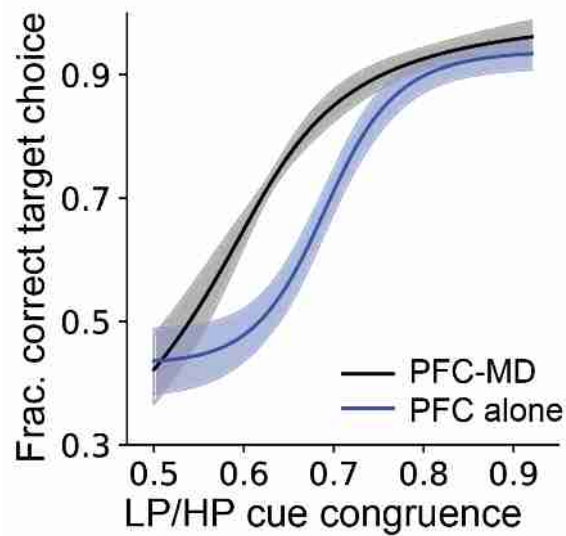
Signal density: 15

Signal-driven uncertainty



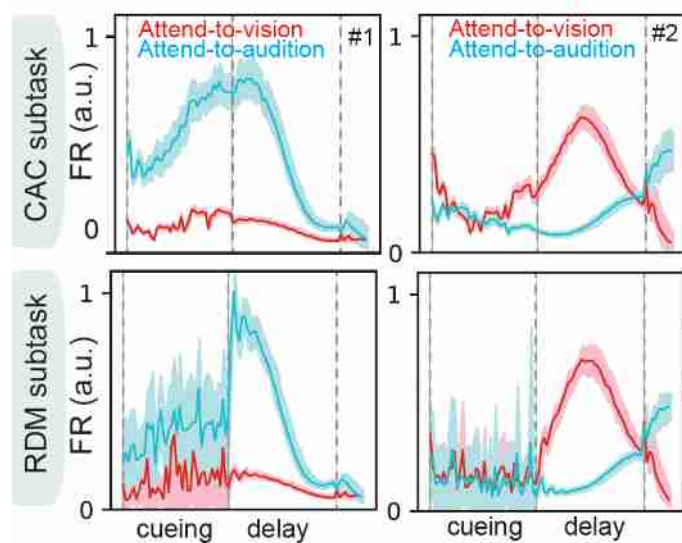
Psychometric curves: PFC-MD vs PFC-alone models

- PFC-MD model outperforms PFC-alone under medium-high cue uncertainty
- MD improves evidence accumulation under noisy cues.

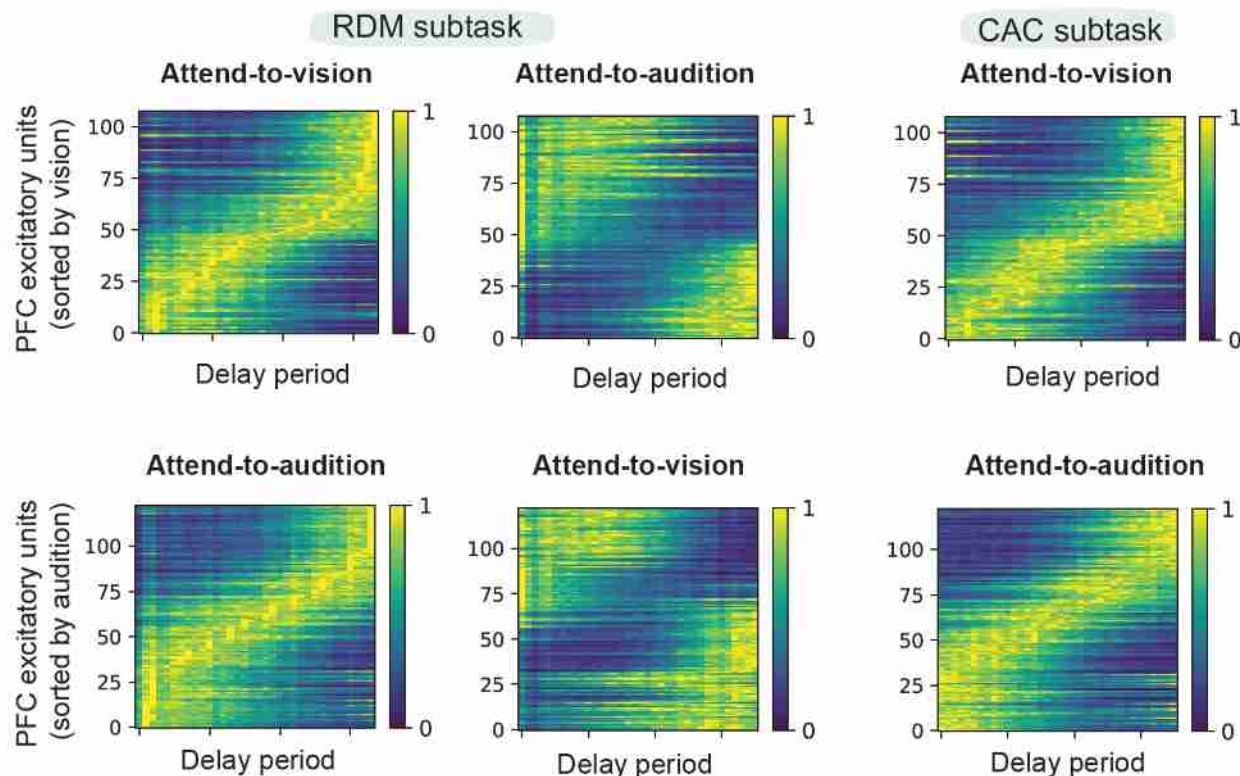


PFC sequences show rule tuning specificity and context invariance

Single PFC neuronal rule tuning



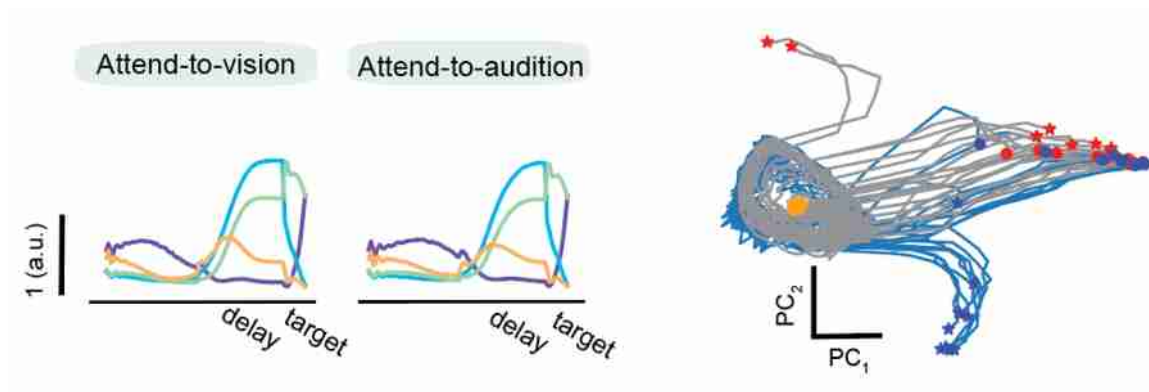
PFC population encoding rule during WM delay



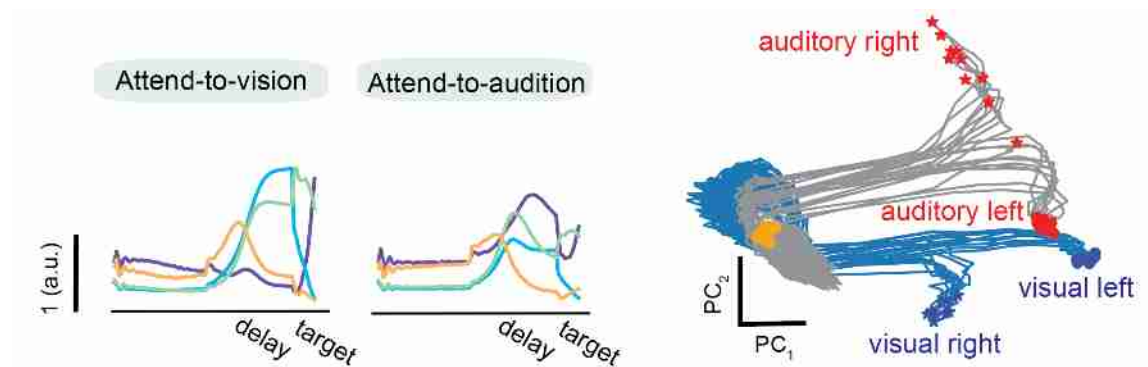
E/I imbalance affects PFC unit tuning and neural trajectory

- MD activation recovers normal PFC rule tuning and neural trajectory

E/I imbalance

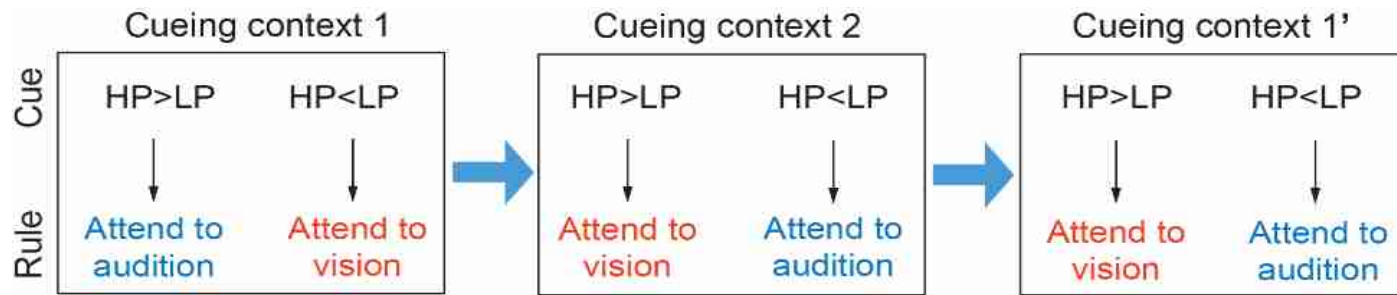


MD activation

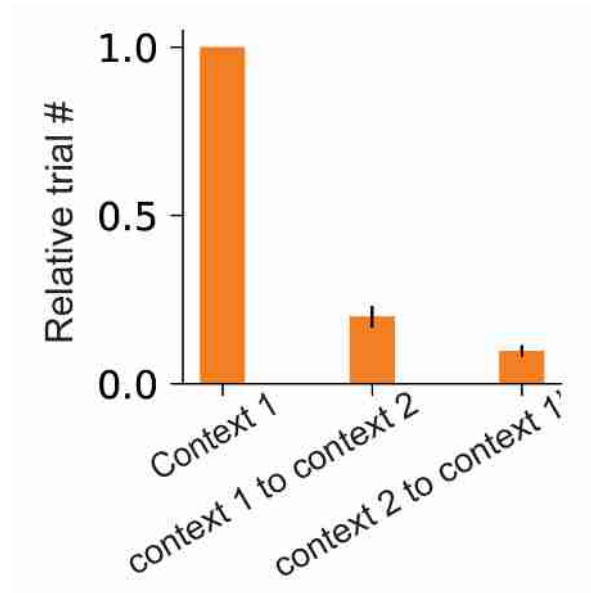


Context switching

Cue-to-rule remapping

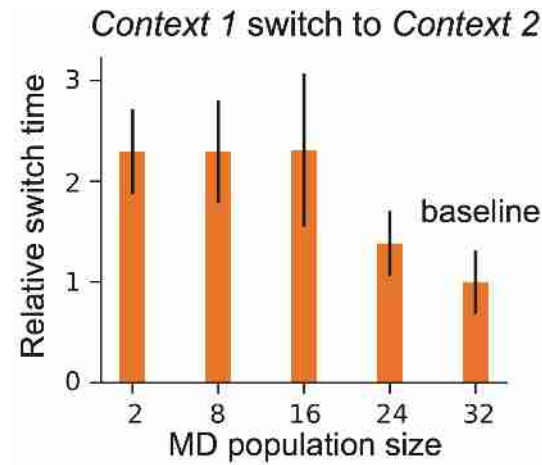


Learning speed in context switching



The size of MD impacts switching speed

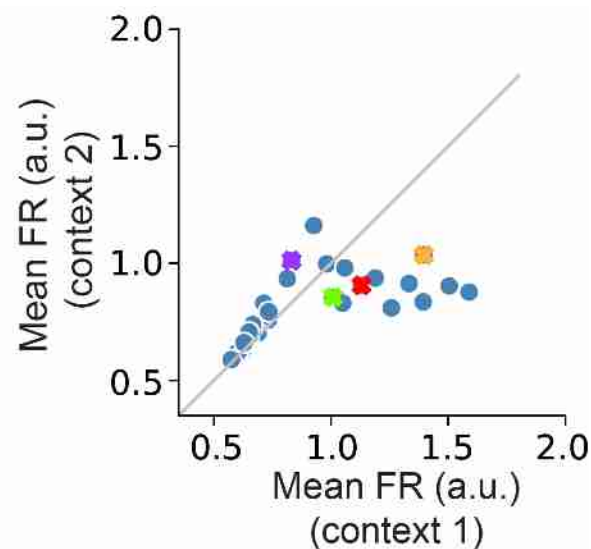
Decreasing the size of MD units increases the switch time



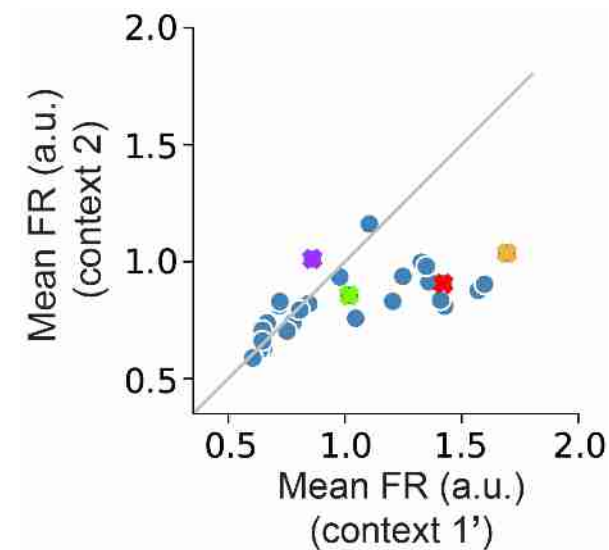
MD representation

MD units show context-invariant tuning between Context 1 and Context 1'

context 1 switch to context 2



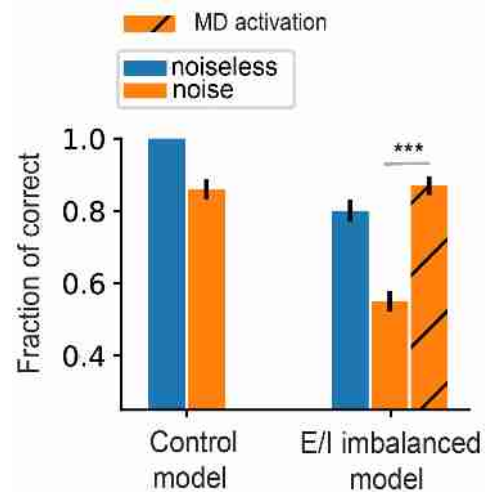
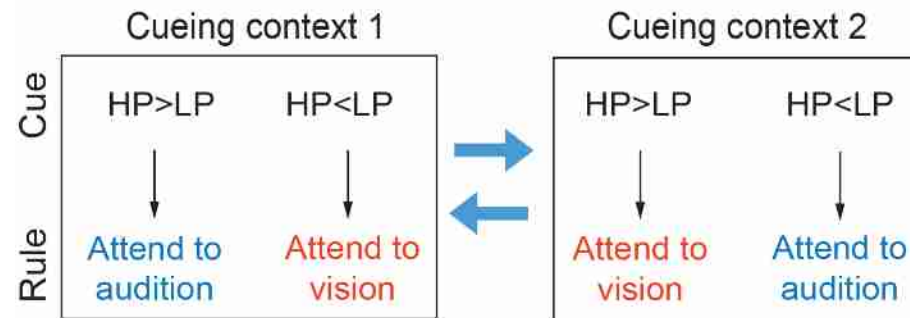
context 2 switch to context 1'



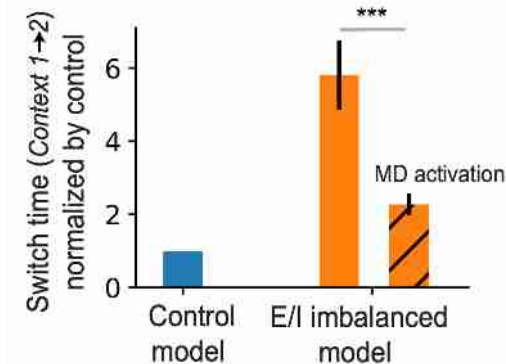
MD activation rescues cognitive deficits

noise interference
in delay period

reduced cortical
PV inhibition: E/I imbalanced

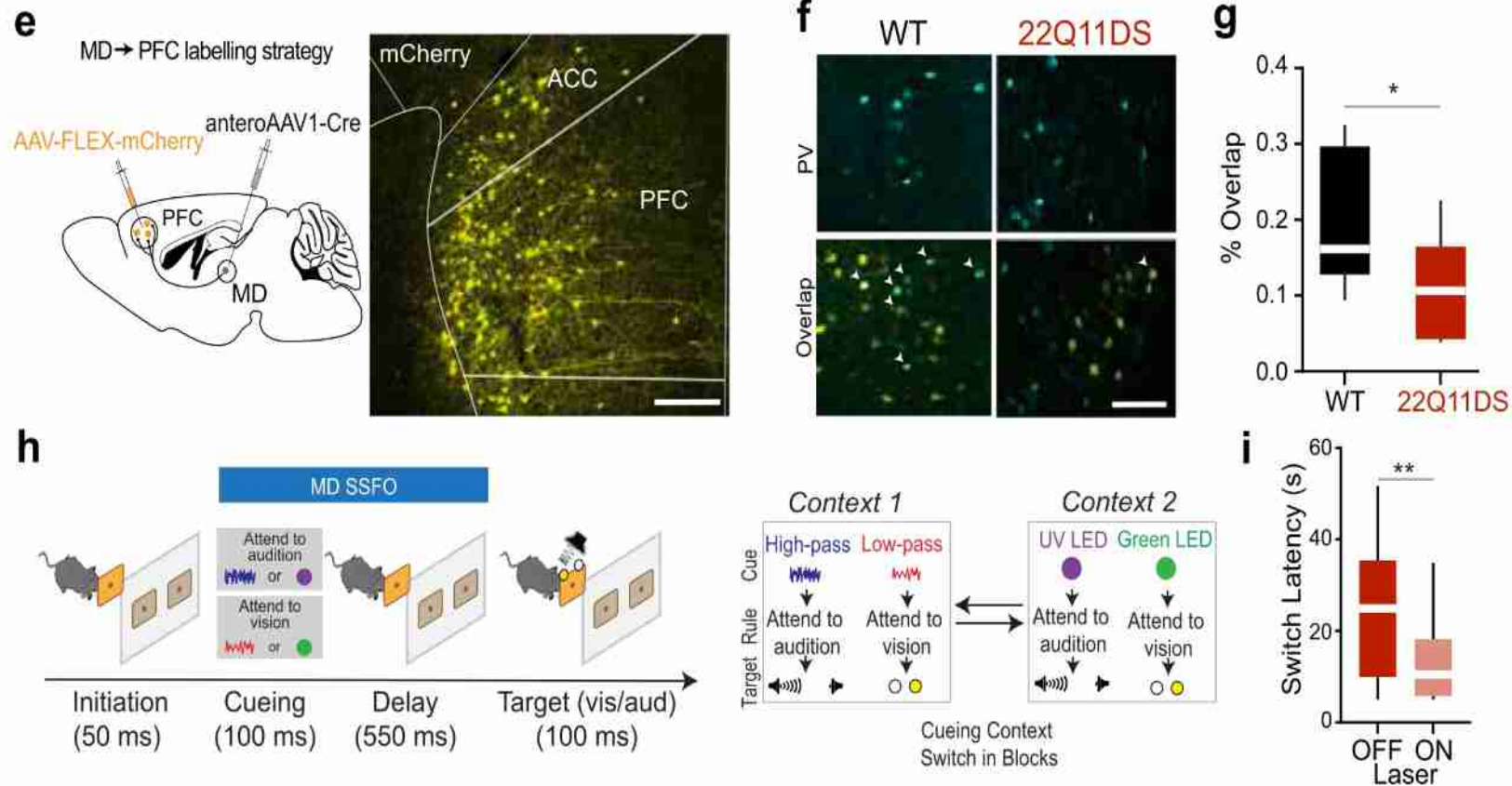


MD activation improved the task performance in the presence of **noise** and **E/I imbalance**



MD activation rescued switching deficit.

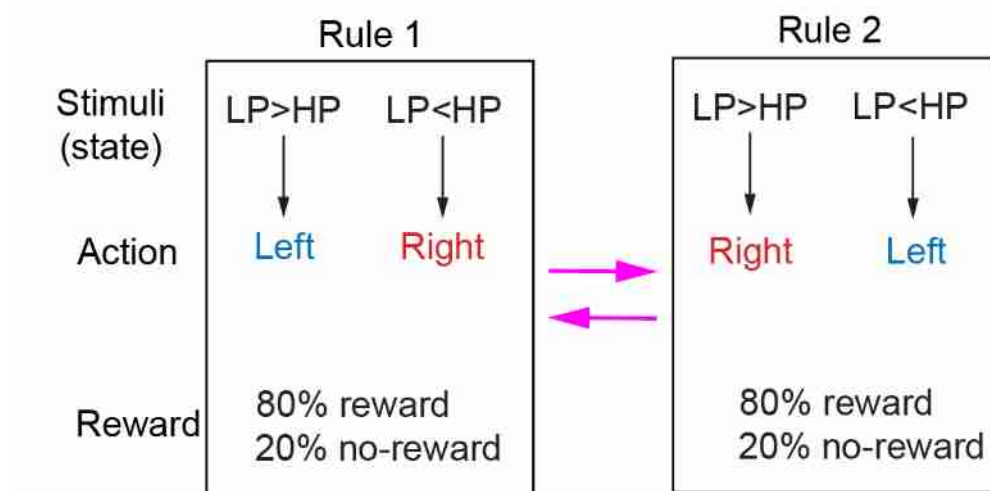
Experimental validation in a schizophrenia mouse model (22Q11)



activating MD can
rescue switching latency

Generalization to outcome or reward uncertainty

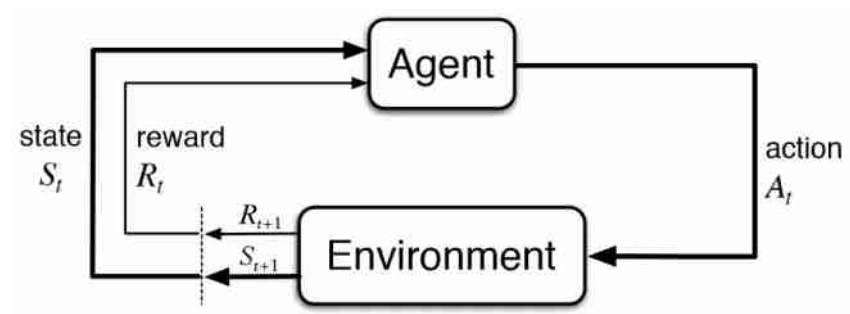
Decision-making task with all **cueing, mapping & outcome** uncertainties



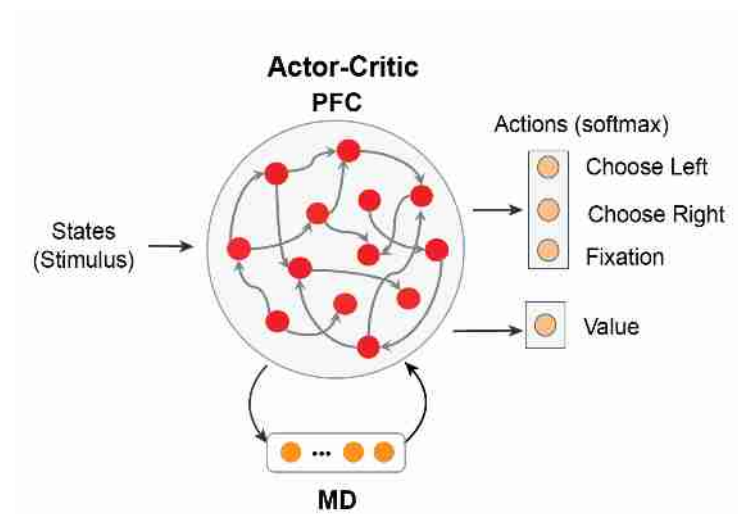
Limitation of supervised learning: the current trial outcome will not impact the decision of the next trial

Reinforcement learning (RL)

- Temporal credit assignment: learning policy to maximize the expected future reward
- Policy gradient optimization

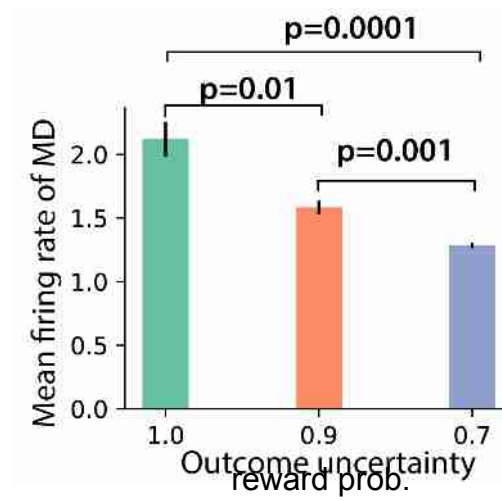


{state, action, reward}

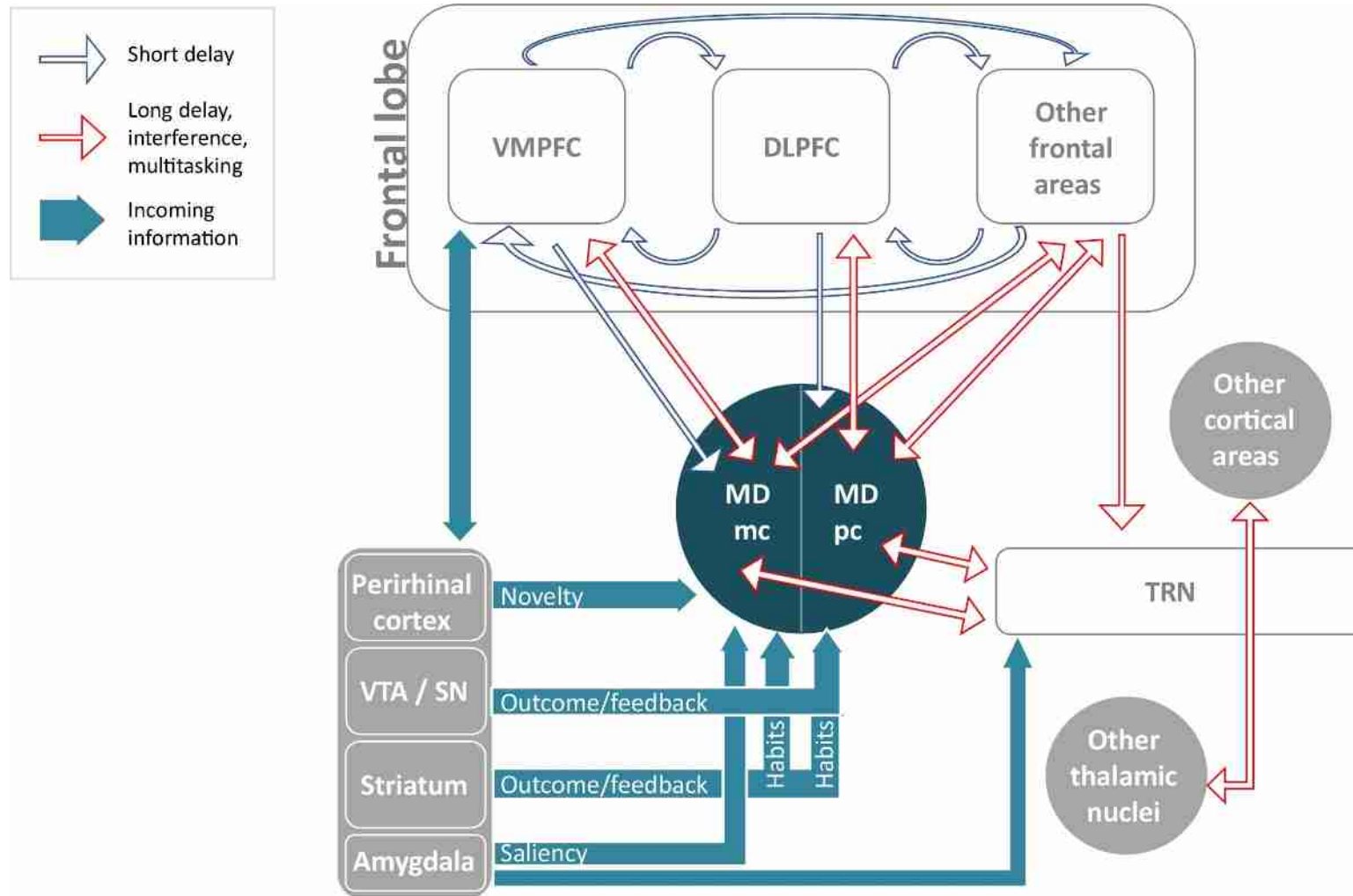


MD encodes outcome uncertainty

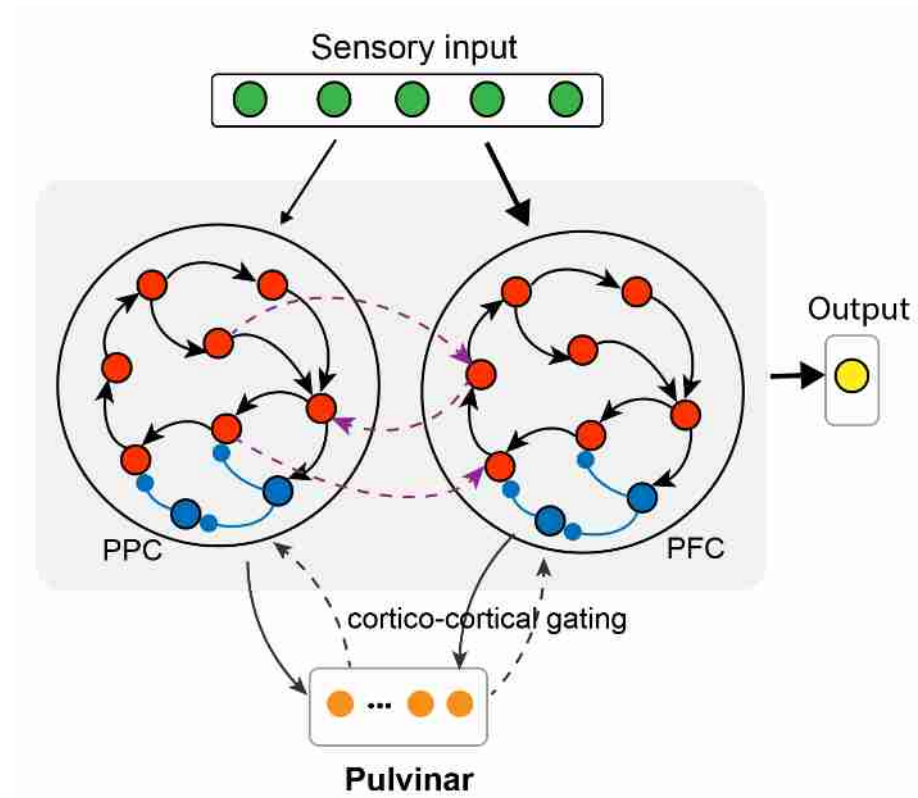
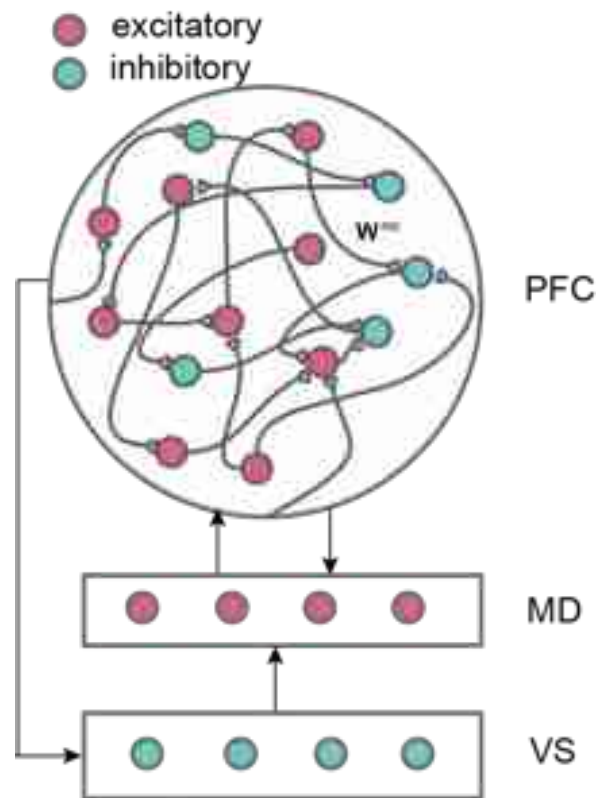
- MD activity modulates with outcome uncertainty



Model oversimplification vs. real brain structures



Increase model complexity: three brain structures



Summary

- Levels of model assumptions and abstraction depends on the purpose of research
- Derive guiding computational principles underlying complex memory/cognition/motor functions
- Biologically constrained network structure and connectivity produce emergent properties of neurons in task-performing RNNs

Open-source software resource

- <https://github.com/murraylab/PsychRNN>

Homework: Training an E/I-RNN for decision-making tasks

- Train and simulate E/I-RNNs using open-source software; Analyze the network connectivity and eigenvalue spectrum; Visualize the activity of single neurons and population activity; Change network connectivity and evaluate the impact on neural dynamics
- <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004792>
- <https://www.nature.com/articles/nature12742>
- <https://www.sciencedirect.com/science/article/pii/S2589004221008877?via%3Dihub>

Latent variable models for neural data analysis

Z. Sage Chen, PhD

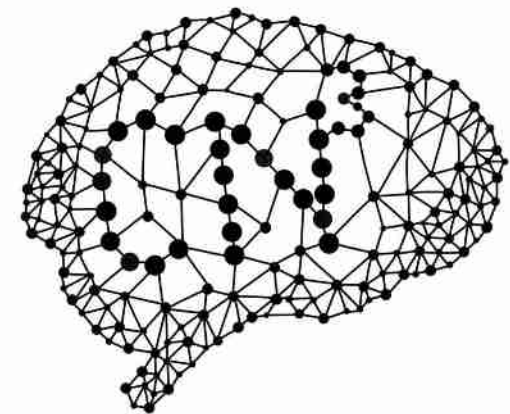
NYU Neuroscience Institute

Department of Psychiatry

Department of Neuroscience & Physiology

Department of Biomedical Engineering

Email: zhe.chen@nyulangone.org

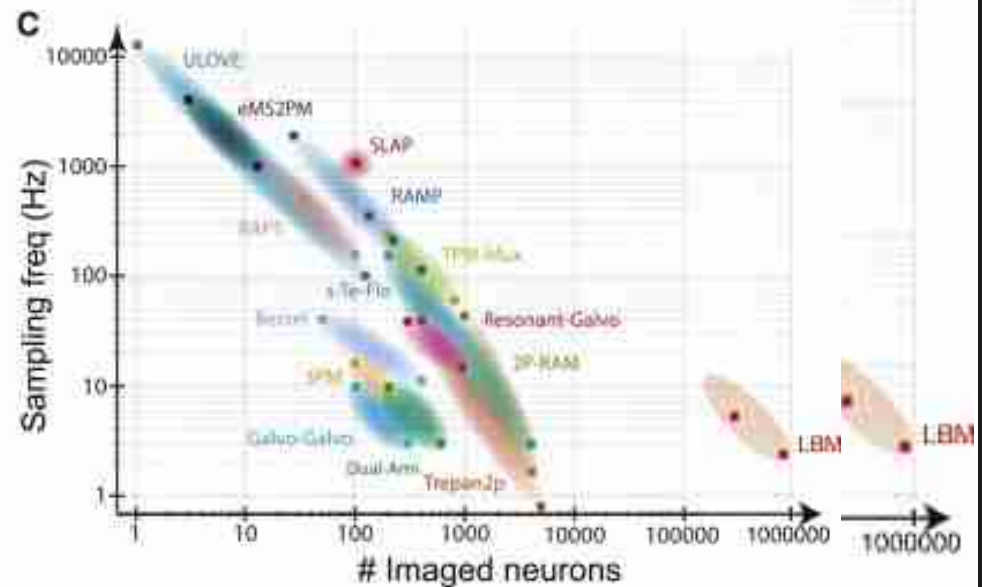
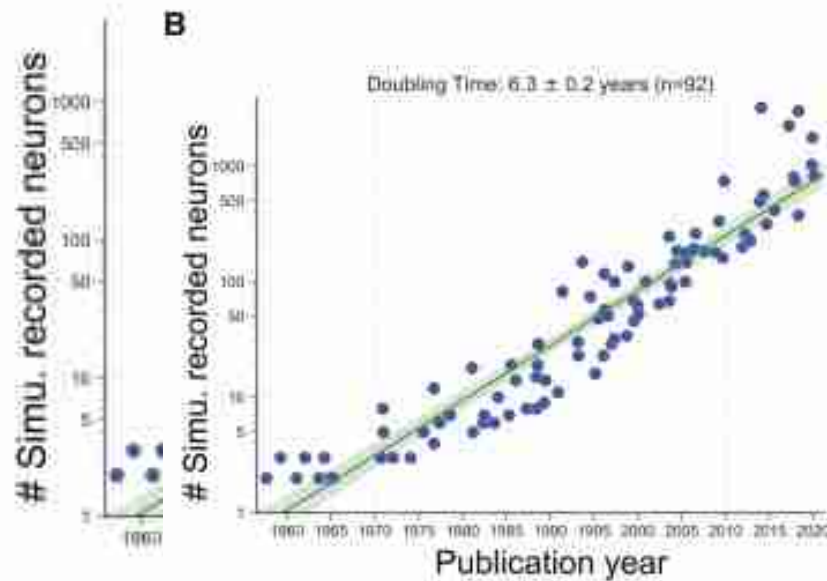


A

Large-scale neural recordings in animals

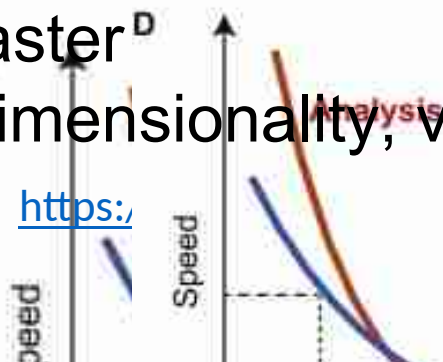
One million dollar question: can one concurrently record spikes from every neuron?
(Kleinfeld et al. 2019 Neuron)

B



(faster
(dimensionality, volume and size)

<https://>



E



Lecoq et al. (2019) J. Neurosci.
Chen and Pesaran (2021) Neuron

Opportunities

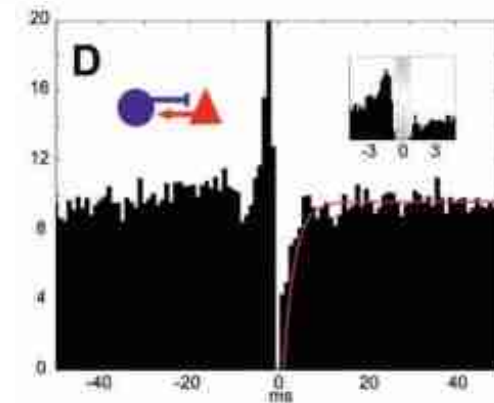
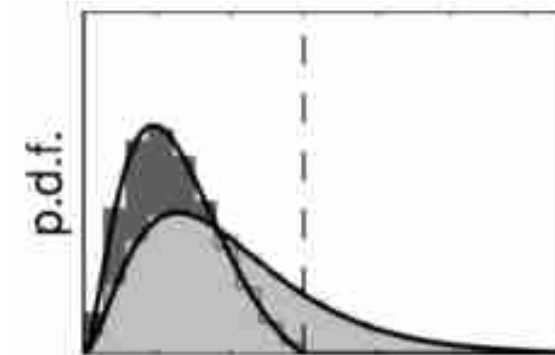
- Ability to examine coordinated brain activities in complex tasks
- Ability to address the “dimensionality of neural codes” for specific tasks
- Ability to subsample / resample the neuronal space to assess the complexity or heterogeneity of response

Outline

- Starting point: Linear and generalized linear models
- Inference as optimization: Likelihood and Bayesian Inference
- A few neuroscience examples
- Nonlinearity: Generative AI applications

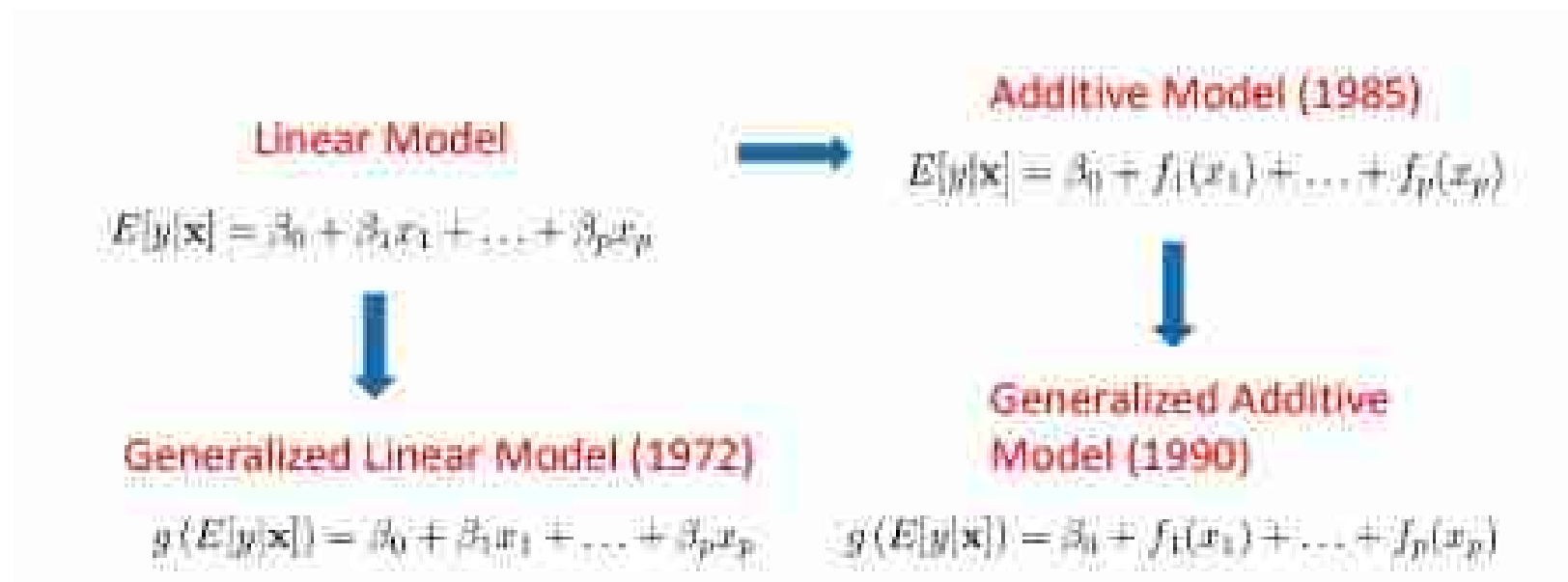
Characterizing neuronal responses

- Modeling spike count or ISI (**rate code**)
- Modeling spike trains (**timing code**)
- Modeling spike train correlation (**correlation code**)



Beyond the linear model

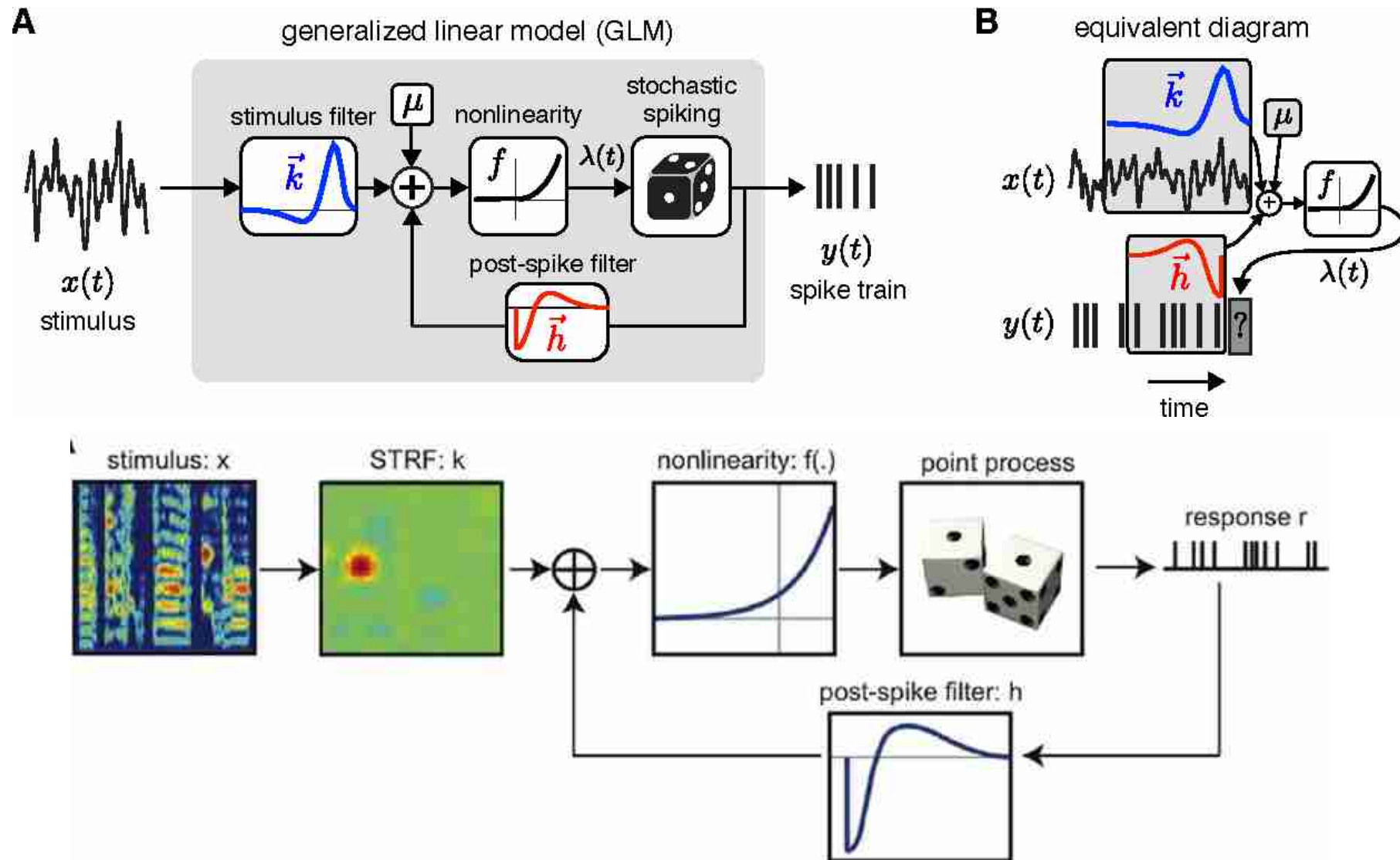
- Covariate (input $\mathbf{x} = [x_1, x_2, \dots, x_p]$) $\rightarrow$ Response (output y)
- $y = \boldsymbol{\beta}\mathbf{x}$, where $\boldsymbol{\beta}$ denotes matched filter (receptive field)
- Mapping the **stochastic relationship** between input and output



Generalized linear model (GLM)

Distribution	Support of distribution	Typical uses	Link name	Link function
Normal	real: $(-\infty, +\infty)$	Linear-response data	Identity	$\mathbf{X}\boldsymbol{\beta} = \mu$
Exponential	real: $(0, +\infty)$	Exponential-response data, scale parameters	Inverse	$\mathbf{X}\boldsymbol{\beta} = -\mu^{-1}$
Gamma				
Inverse Gaussian	real: $(0, +\infty)$		Inverse squared	$\mathbf{X}\boldsymbol{\beta} = -\mu^{-2}$
Poisson	integer: $[0, +\infty)$	count of occurrences in fixed amount of time/space	Log	$\mathbf{X}\boldsymbol{\beta} = \ln(\mu)$
Bernoulli	integer: $[0, 1]$	outcome of single yes/no occurrence	Logit	$\mathbf{X}\boldsymbol{\beta} = \ln\left(\frac{\mu}{1-\mu}\right)$
Binomial	integer: $[0, N]$	count of # of "yes" occurrences out of N yes/no occurrences		
Categorical	integer: $[0, K)$	outcome of single K-way occurrence		
	K-vector of integer: $[0, 1]$, where exactly one element in the vector has the value 1			
Multinomial	K-vector of integer: $[0, N]$	count of occurrences of different types (1 .. K) out of N total K-way occurrences		

Example: modeling neuronal spiking activity



Beyond Poisson

Poisson: Mean = variance

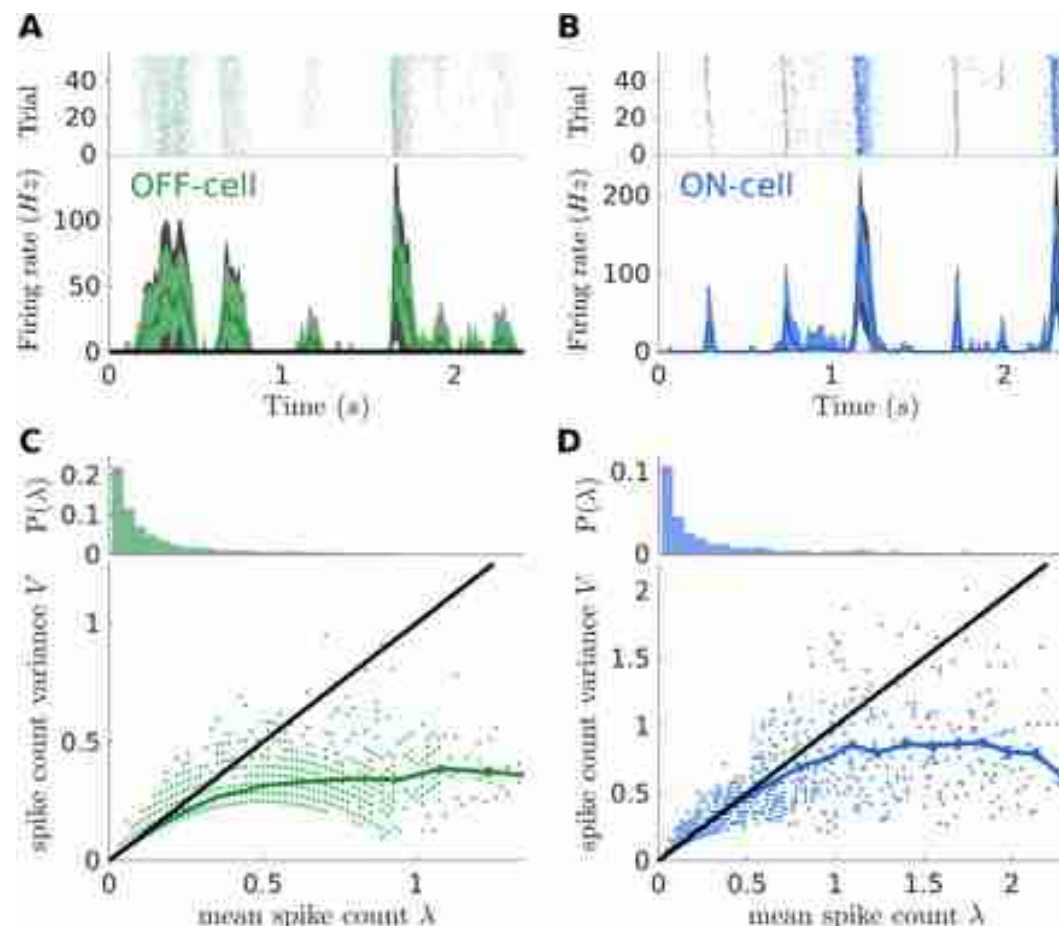
Mean < variance: underdispersed

Mean > variance: overdispersed

QUESTIONS

How to model overdispersed firing?

How to calibrate it?



Distribution	Range of random variables	Degree of freedom
Gaussian	Real-valued	Mean μ , variance σ^2
Binomial	Binary	$0 < p < 1$
Poisson	Nonnegative integer	Rate $\lambda > 0$
Negative binomial	Nonnegative integer	$r > 0, 0 < p < 1$
Conway–Maxwell–Poisson (Stevenson 2016)	Nonnegative integer	$\lambda > 0, \nu \geq 0$
Generalized Gaussian (Gao et al. 2016)	Nonnegative integer	$\lambda, g(\cdot)$

Computational goals

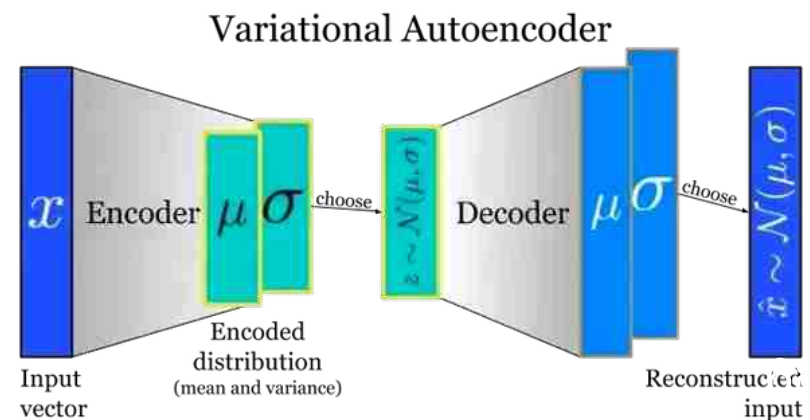
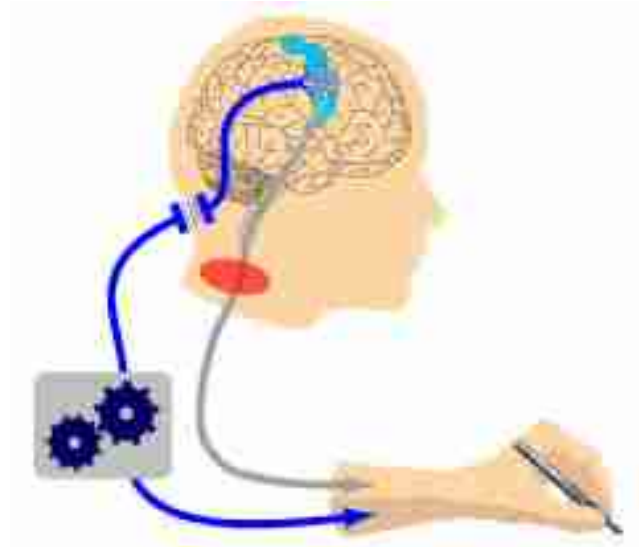
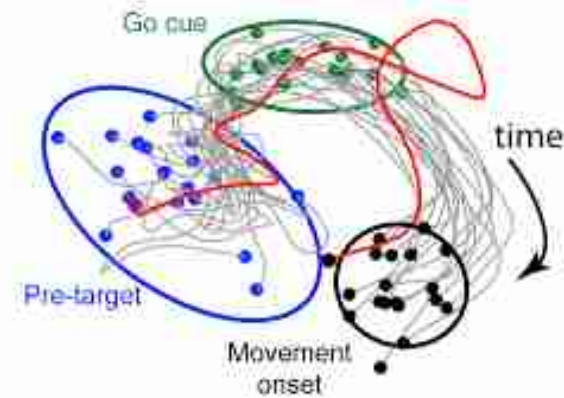
- 1) Uncovering latent structures or neural representations in task behaviors, and establish their connections to behavioral measures
- 2) Explain the neuronal response variability (e.g., Poisson to doubly stochastic)
- 3) Manifold learning and dimensionality reduction (task-relevant information)

Latent variables

- **Purpose:** characterize unobserved random, or deterministic but unknown, or missing variable that represent either quantitative or categorical information (e.g., membership)
- Continuous or discrete or mixed
- Degree of unknown is characterized by the uncertainty (confidence intervals, probability distribution)
- Time is the essence: embedded in stochastic processes (Markovian, Gaussian, Poisson, Cox, ...)

Neuroscience applications

- Dimensionality reduction and visualization
- Inverse problems: decoding, deconvolution, denoising
- Detection of change points or seizures, prosthetic control
- Data compression, augmentation
- Account for response variability: common input, top-down attention, modulatory input



PPCA and factor analysis

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})$$

$$p(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x}|\mathbf{W}\mathbf{z} + \boldsymbol{\mu}, \sigma^2\mathbf{I})$$

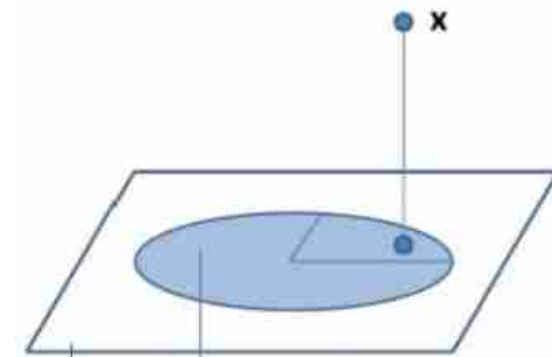
$$\begin{aligned} p(\mathbf{x}) &= \int p(\mathbf{x}|\mathbf{z})p(\mathbf{z}) d\mathbf{z} \\ &= \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \sigma^2\mathbf{C}) \end{aligned}$$

$$\mathbf{C} = \mathbf{W}\mathbf{W}^\top + \sigma^2\mathbf{I}$$

Factor Analysis

W: non-orthogonal

Noise variance **Λ** : non-isotropic

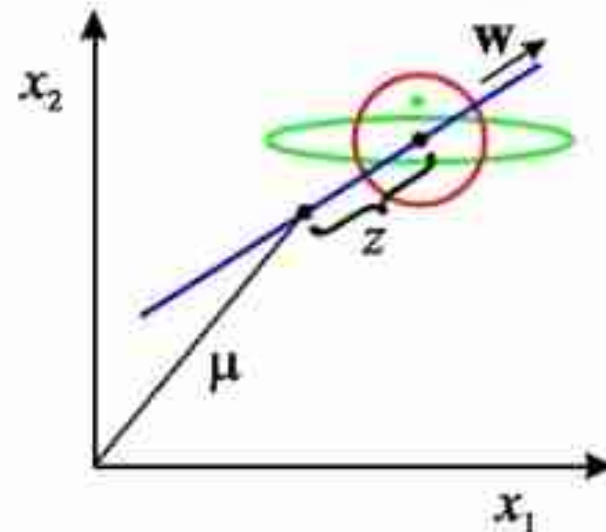


PCA: uniform prior on the subspace

PPCA: Gaussian dist.

PCA

factor analysis



$$\mathbf{x} = \mathbf{W}\mathbf{z} + \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

State space model (SSM)

- a) dynamic latent continuous “state”
- b) statistical characterization of spatiotemporal neural observations

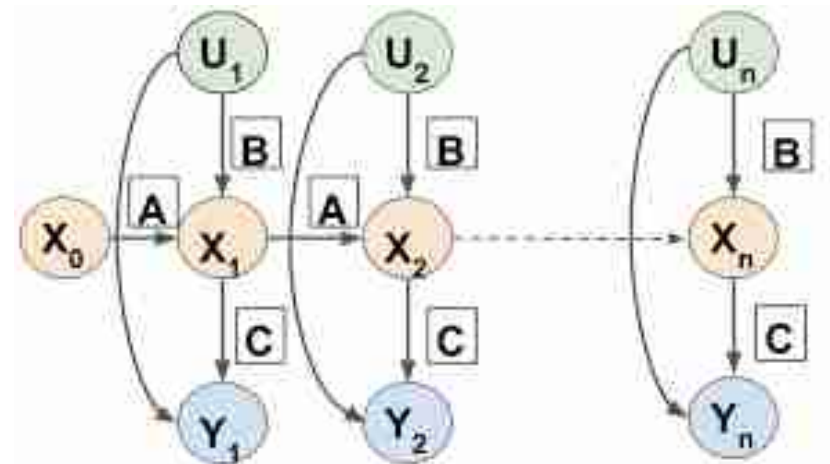
$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{w}_k) \quad \text{state equation for } P(\mathbf{x}_k | \mathbf{x}_{k-1})$$

$$\mathbf{y}_k = g(\mathbf{x}_k, \mathbf{v}_k) \quad \text{observation equation for } P(\mathbf{y}_k | \mathbf{x}_k)$$



Simple example: linear dynamical systems (LDS) or linear Gaussian dynamics (Kalman filter)

$$\begin{aligned}\mathbf{x}_{t+1} &= \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_{t+1} + \mathbf{n}_{t+1} \\ \mathbf{y}_t &= \mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t + \mathbf{v}_t\end{aligned}$$



Recursive Bayesian estimation

Posterior estimation

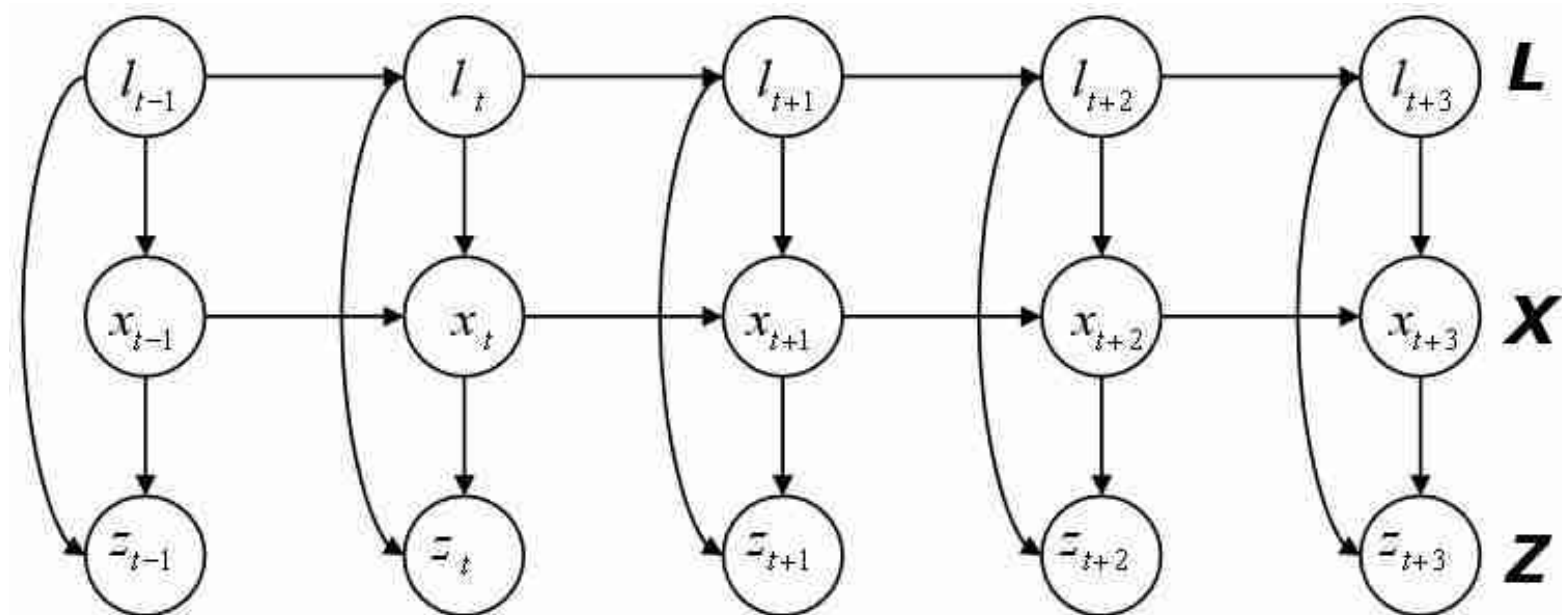
$$\begin{aligned} p(\mathbf{x}_t | \mathbf{y}_{0:t}) &= \frac{p(\mathbf{x}_t) p(\mathbf{y}_{0:t} | \mathbf{x}_t)}{p(\mathbf{y}_{0:t})} = \frac{p(\mathbf{x}_t) p(\mathbf{y}_t, \mathbf{y}_{0:t-1} | \mathbf{x}_t)}{p(\mathbf{y}_t, \mathbf{y}_{0:t-1})} \\ &= \frac{p(\mathbf{x}_t) p(\mathbf{y}_t | \mathbf{x}_t, \mathbf{y}_{0:t-1}) p(\mathbf{y}_{0:t-1} | \mathbf{x}_t)}{p(\mathbf{y}_t | \mathbf{y}_{0:t-1}) p(\mathbf{y}_{0:t-1})} \\ &= \frac{p(\mathbf{x}_t) p(\mathbf{y}_t | \mathbf{x}_t, \mathbf{y}_{0:t-1}) p(\mathbf{x}_t | \mathbf{y}_{0:t-1}) p(\mathbf{y}_{0:t-1})}{p(\mathbf{y}_t | \mathbf{y}_{0:t-1}) p(\mathbf{y}_{0:t-1}) p(\mathbf{x}_t)} \\ &= \frac{p(\mathbf{y}_t | \mathbf{x}_t, \mathbf{y}_{0:t-1}) p(\mathbf{x}_t | \mathbf{y}_{0:t-1})}{p(\mathbf{y}_t | \mathbf{y}_{0:t-1})} = \frac{p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}_{0:t-1})}{p(\mathbf{y}_t | \mathbf{y}_{0:t-1})}, \end{aligned}$$

One-step state prediction (Chapman-Komogorov equation)

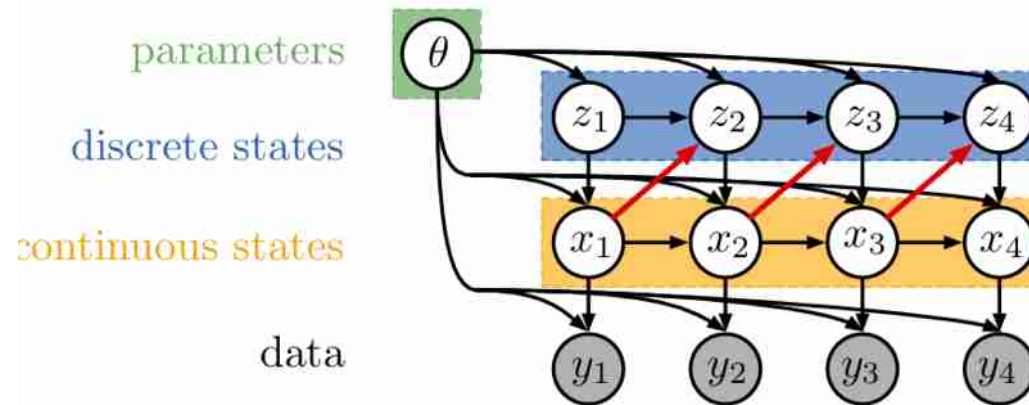
$$p(\mathbf{x}_{t+1} | \mathbf{y}_{0:t}) = \int p(\mathbf{x}_{t+1} | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}_{0:t}) d\mathbf{x}_t,$$

$$p(\mathbf{x}_{t+1} | \mathbf{y}_{0:t}, S_{0:t}) = \int \sum_{S_t} p(\mathbf{x}_{t+1} | \mathbf{x}_t, S_{t+1}) p(S_{t+1} | S_t) p(\mathbf{x}_t | \mathbf{y}_{0:t}, S_{0:t}) d\mathbf{x}_t.$$

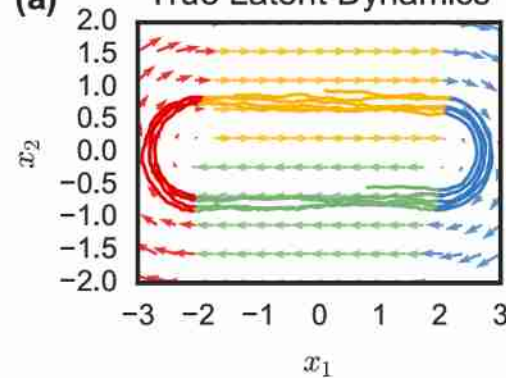
Generalization: Switching LDS (sLDS)



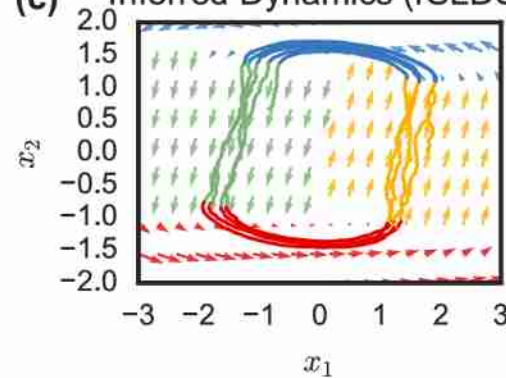
Recurrent switching LDS (rsLDS)



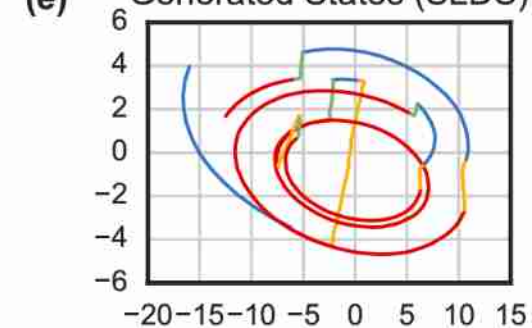
(a) True Latent Dynamics



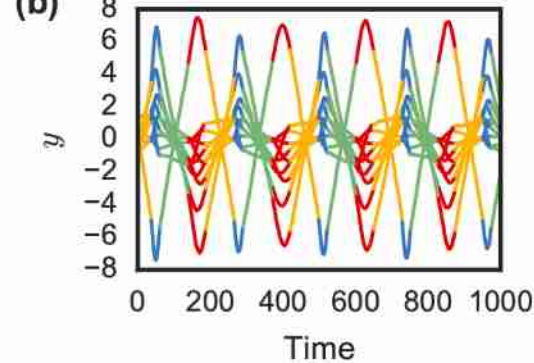
(c) Inferred Dynamics (rSLDS)



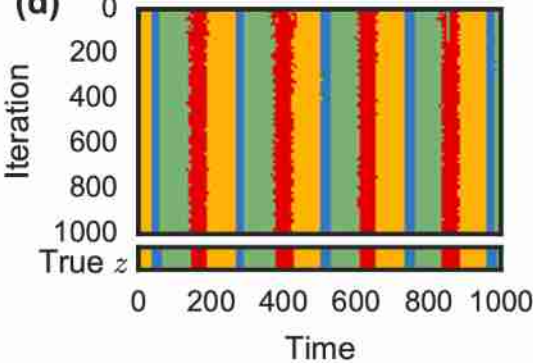
(e) Generated States (SLDS)



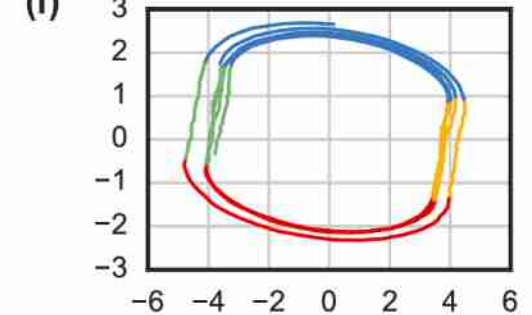
(b) Observed Data



(d) Discrete State Samples



(f) Generated States (rSLDS)



(Finite) mixture models

Purposes: density estimation, clustering, dimensionality reduction

- Mixtures of Gaussians / t-distributions

$$\mathcal{P}(\xi_t) = \sum_{i=1}^K \pi_i \mathcal{N}(\xi_t | \mu_i, \Sigma_i)$$

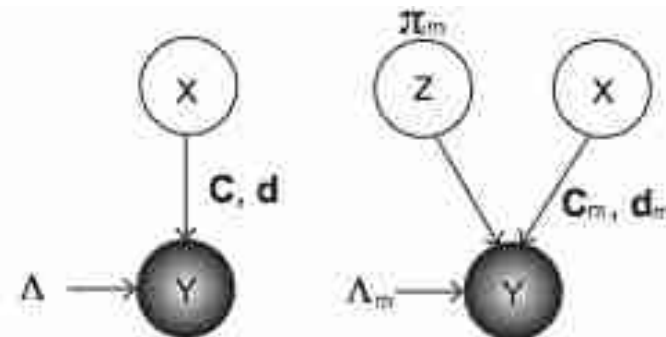
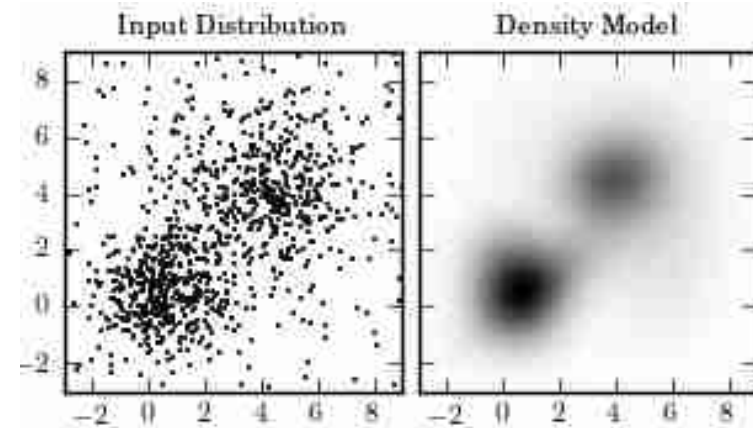
N datapoints of dimension D

$$\mathcal{N}(\xi_t | \mu_i, \Sigma_i) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} (\xi_t - \mu_i)^T \Sigma_i^{-1} (\xi_t - \mu_i) \right)$$

- Mixtures of factor analysis (MFA)

$$\text{s.t. } 0 < \pi_m < 1, \quad \sum_{m=1}^M \pi_m = 1$$

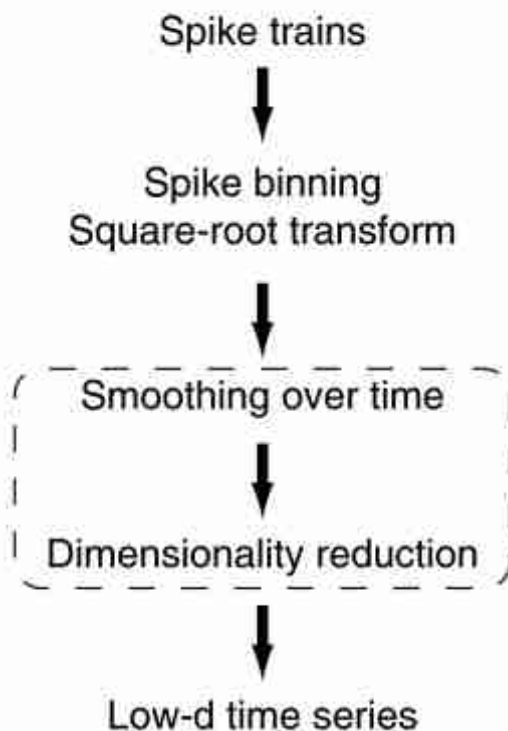
$$\mathbf{y}_t \sim \mathcal{N}(\mathbf{C}_m \mathbf{x}_t + \mathbf{d}_m, \Lambda_m)$$



GPFA (Gaussian process factor analysis)

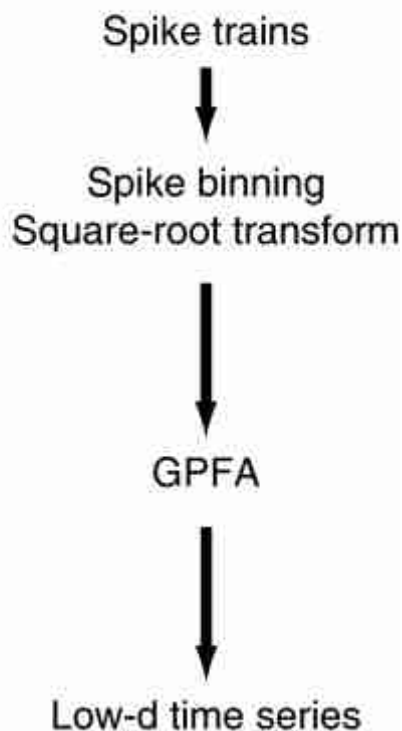
A

Two-stage methods



B

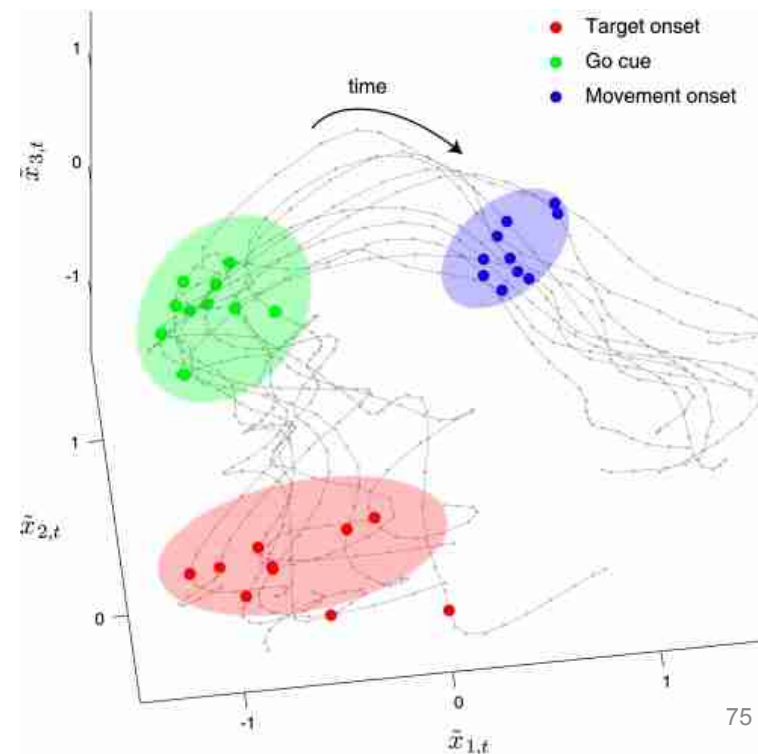
GPFA



$$\mathbf{x}_{i,:} \sim \mathcal{N}(\mathbf{0}, K_i)$$

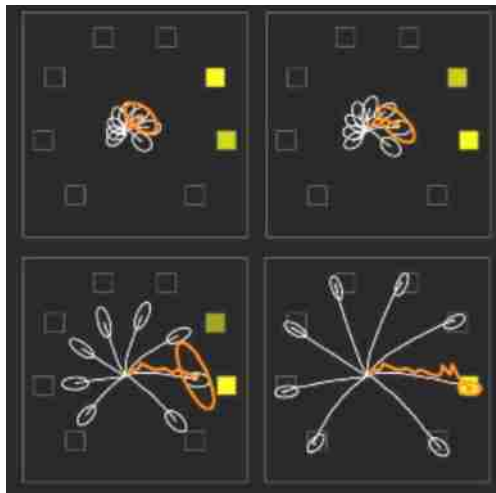
$$\mathbf{y}_{:,t} | \mathbf{x}_{:,t} \sim \mathcal{N}(C\mathbf{x}_{:,t} + \mathbf{d}, R)$$

$$K_i(t_1, t_2) = \sigma_{f,i}^2 \exp\left(-\frac{(t_1 - t_2)^2}{2\tau_i^2}\right) + \sigma_{n,i}^2 \cdot \delta_{t_1,t_2}$$

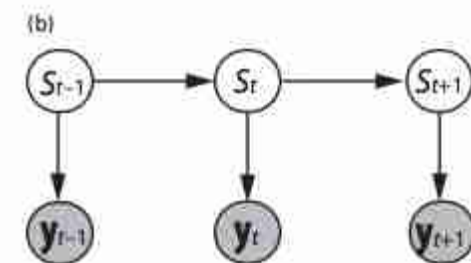
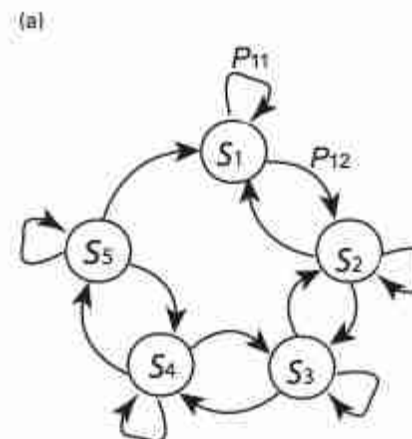


Hidden Markov model

- Mixture model $\rightarrow$ introduce latent state dynamics (Markovian)
- Stochastic observations: multinomial, Gaussian, Poisson, ...



- Examples:
 - Trajectory decoding
 - UP/DOWN state detection
 - sleep staging



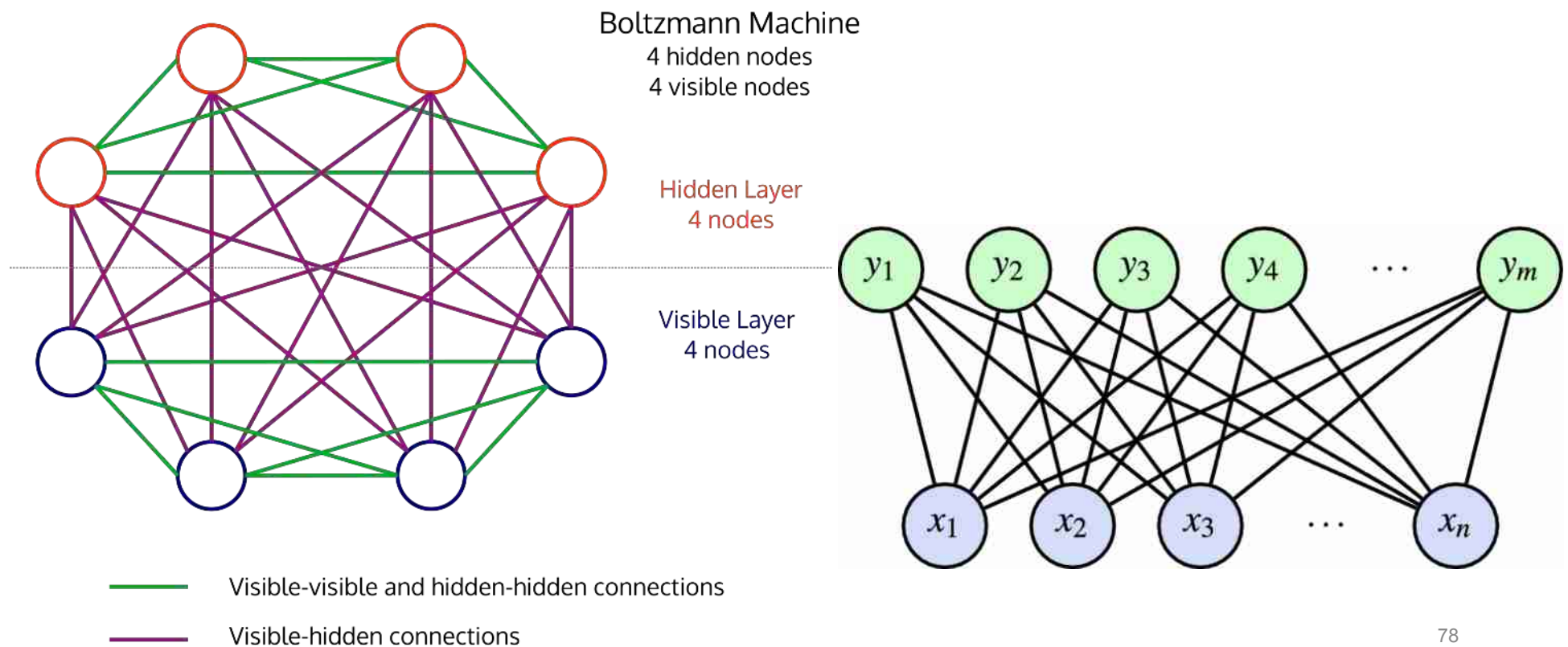
Mixed latent variable models

	Continuous	Discrete	Mixed
Static	PPCA, FA	Mixture models (GMM)	Mixtures of FA
Dynamic	LDS, GPFA, LFADS	HMM, HSMM	Switching LDS Mixture of GPFA HMM-GPFA

Energy-based latent variable models

- Exponential family Restricted Boltzmann machine (RBM): static data

Welling, Rosen-zvi, Hinton, NIPS2005



Likelihood inference

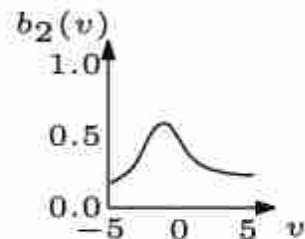
- Likelihood: Probability of data as a function of the parameter is the likelihood of the parameter θ
- Observed data likelihood: $p(Y | X, \theta)$ ---- ML inference
- Complete data likelihood: $p(X, Y | \theta)$
- Expected complete data log-likelihood: $E_x[\log p(X, Y | \theta)]$ --- EM algorithm
- Marginal likelihood: $p(Y)$ --- VB inference

Homework: EM algorithm for HMM

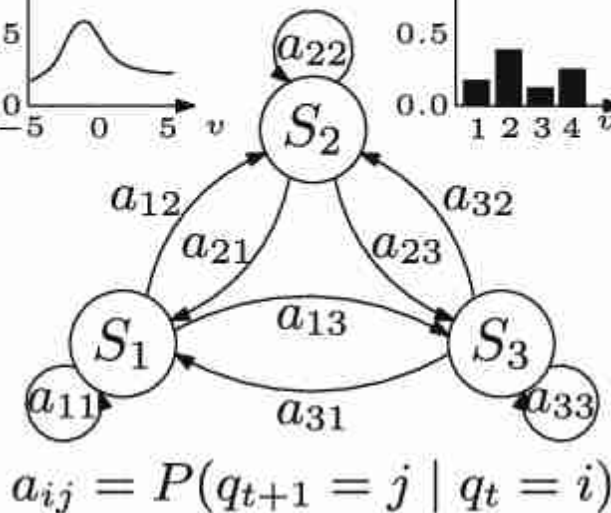
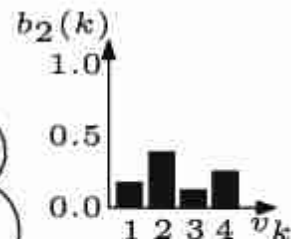
- Baum-Welch algorithm for model estimation
- Viterbi algorithm for inference (dynamic programming)

Continuous
output

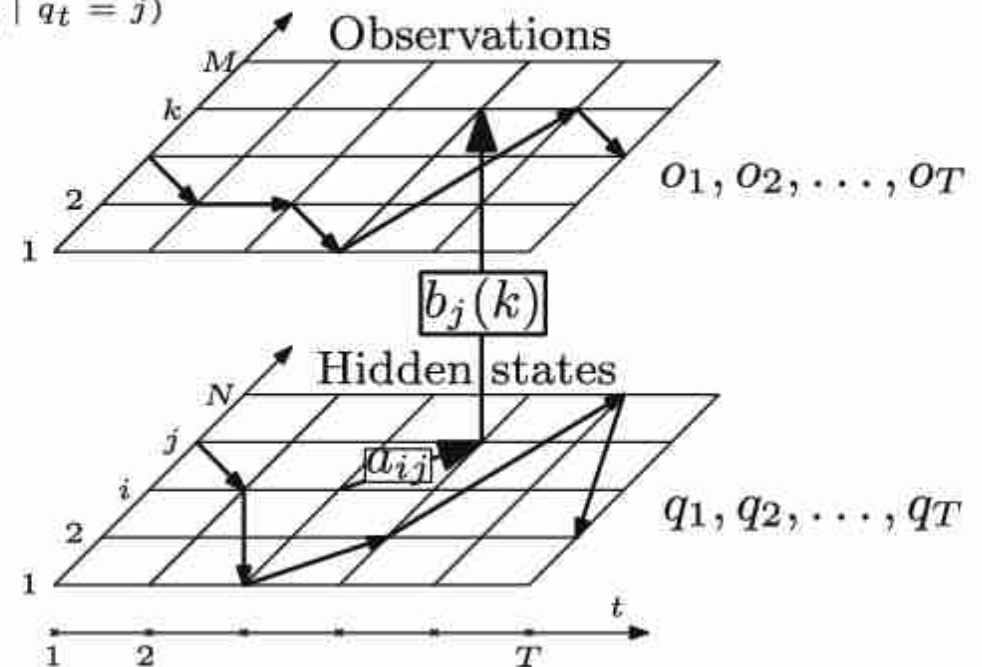
$$b_j(k) = f(o_t = v \mid q_t = j)$$



$$b_j(k) = P(o_t = v_k \mid q_t = j)$$



Discrete
output



Bayesian inference



- Mean field approximation

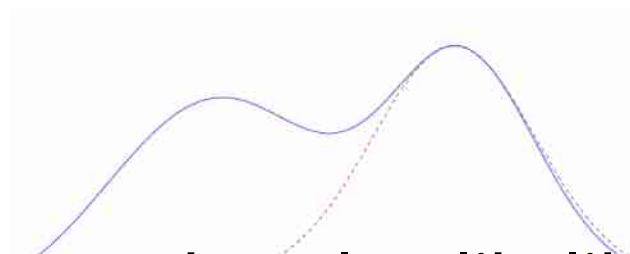
$$\begin{aligned} p(Z, \theta | Y) &\approx p(Z | Y) p(\theta | Y) \\ &= \frac{p(Y | Z, \theta) p(Z) p(\theta)}{p(Y)} \\ &= \frac{p(Y | Z, \theta) p(Z) p(\theta)}{\int \int p(Y | Z, \theta) p(Z) p(\theta) dZ d\theta} \end{aligned}$$

- Marginal likelihood (“evidence”)

$$p(Y) = \int \int \int p(Y | X, Z, \theta) p(X) p(Z) p(\theta) dX dZ d\theta$$

Gaussian (Laplace) approximation

- Saddle-point method
- Taylor expansion around the mode at log-likelihood



$$\ln f(x) = \ln f(x_0) + \underbrace{\frac{\partial \ln f(x)}{\partial x} \Big|_{x=x_0}}_{\text{second term}} \cdot (x - x_0) + \frac{1}{2} \frac{\partial^2 \ln f(x)}{\partial x^2} \Big|_{x=x_0} \cdot (x - x_0)^2 + h.o.t...$$

$$\ln f(x) = \ln f(x_{\max}) + \frac{1}{2} \frac{\partial^2 \ln f(x)}{\partial x^2} \Big|_{x=x_{\max}} \cdot (x - x_{\max})^2$$

$$\int e^{\ln f(x)} dx = \underbrace{e^{\ln f(x_{\max})}}_{\text{constant}} \int \exp \left[\underbrace{\frac{1}{2} \frac{\partial^2 \ln f(x)}{\partial x^2} \Big|_{x=x_{\max}}}_{\text{constant}} \cdot (x - x_{\max})^2 \right] dx$$

$$\approx e^{L(x_{\max})} \int \exp \left[-\frac{(x - x_{\max})^2}{2\sigma^2} \right] dx$$

First-order Laplace method

- Procedure

(1) find a local maximum $x_{\max}$ of the given *pdf* $f(x)$

(2) calculate the variance $\sigma^2 = -\frac{1}{f''(x_{\max})}$

(3) approximate the *pdf* with $p(x) \approx \mathcal{N}[x_{\max}, \sigma^2]$

Variational Bayes (VB) approximation

Optimization: maximize the lower bound of marginal (log) likelihood

$$\begin{aligned}\log p(Y) &= \log \int d\theta \int dX p(\theta) p(Z, Y|\theta) \\ &= \log \int d\theta \int dZ q(Z, \theta) \frac{p(\theta) p(Z, Y|\theta)}{q(Z, \theta)} \\ &\geq \int d\theta \int dZ q(Z, \theta) \log \frac{p(\theta) p(Z, Y|\theta)}{q(Z, \theta)} \\ &= \left\langle \log p(Z, Y, \theta) \right\rangle_q + \mathcal{H}_q(Z, \theta) \equiv \mathcal{F}(q(Z, \theta))\end{aligned}$$

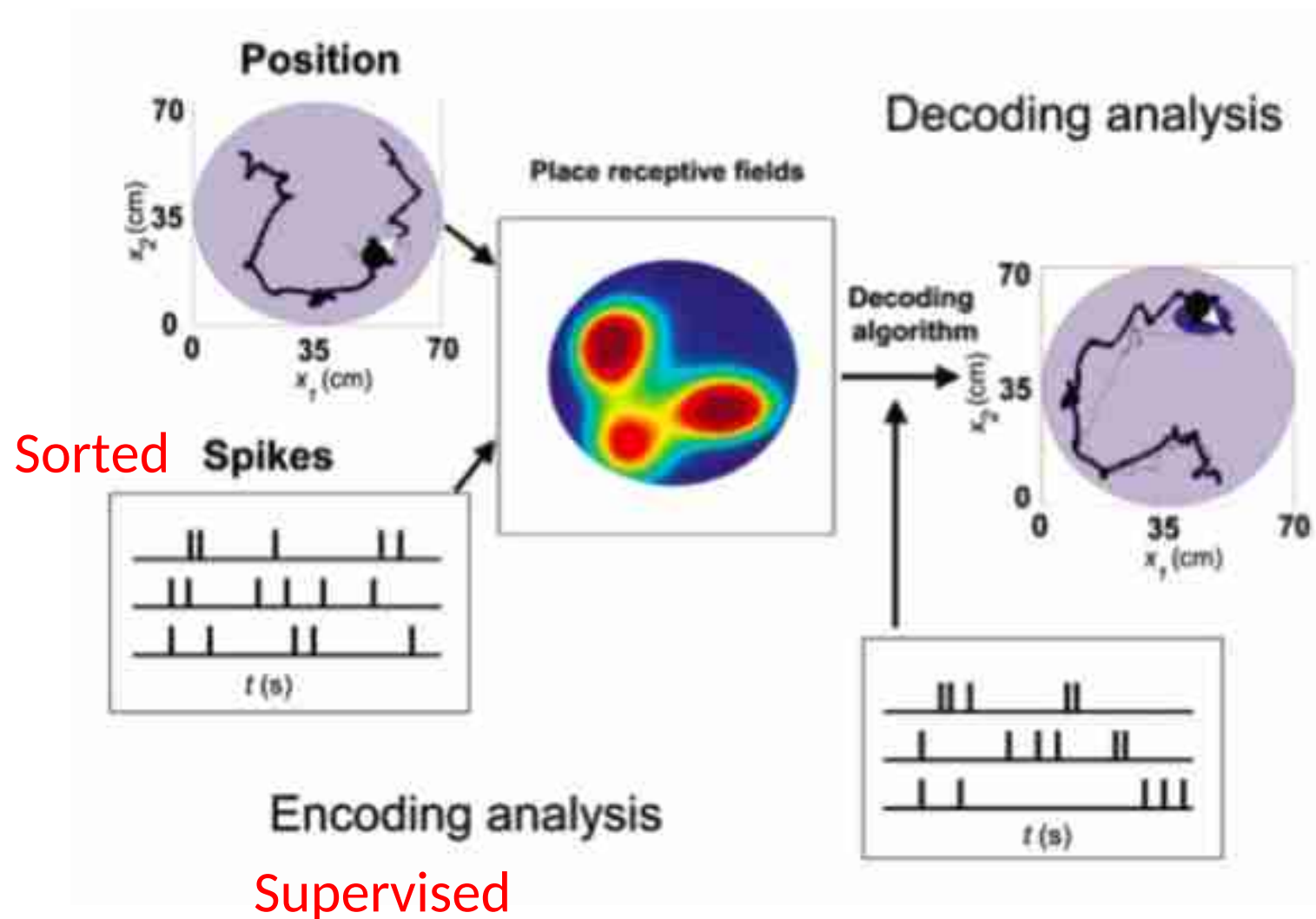
$p(\theta)$: prior distribution
 $p(Z, Y|\theta)$: complete data likelihood
 $q(Z, \theta)$ variational posterior dist
 $\mathcal{F}(q(Z, \theta))$ free energy

VB-EM algorithm:

VB-E step:

VB-M step:

State-space model for Bayesian decoding analysis



Brown et al. (1998) J Neurosci
Zhang et al. (1998) J Neurophysiol

State-space neural encoding & decoding analyses

$$W_x(\Delta_k) = \begin{bmatrix} \sigma_{x_1}^2 & \rho\sigma_{x_1}\sigma_{x_2} \\ \rho\sigma_{x_1}\sigma_{x_2} & \sigma_{x_2}^2 \end{bmatrix} \Delta_k,$$

Brown et al. (1998) J Neurosci

State equation:

$$x(t_k) - x(t_{k-1}) \sim N(0, W_x(\Delta_k)); \quad (8)$$

Observation equation:

$$f(\mathbf{I}(t_k)|x(t_k), t_{k-1}) = \prod_{c=1}^C \{[\lambda^c[x(t_k)]\Lambda^c[\theta(\Delta_k)]]^{I_c(t_k)} \exp\{-\lambda^c[x(t_k)]\Lambda^c[\theta(\Delta_k)]\}\}; \quad (9)$$

One-step prediction equation:

$$\hat{x}(t_k|t_{k-1}) = \hat{x}(t_{k-1}|t_{k-1}); \quad (10)$$

One-step prediction variance:

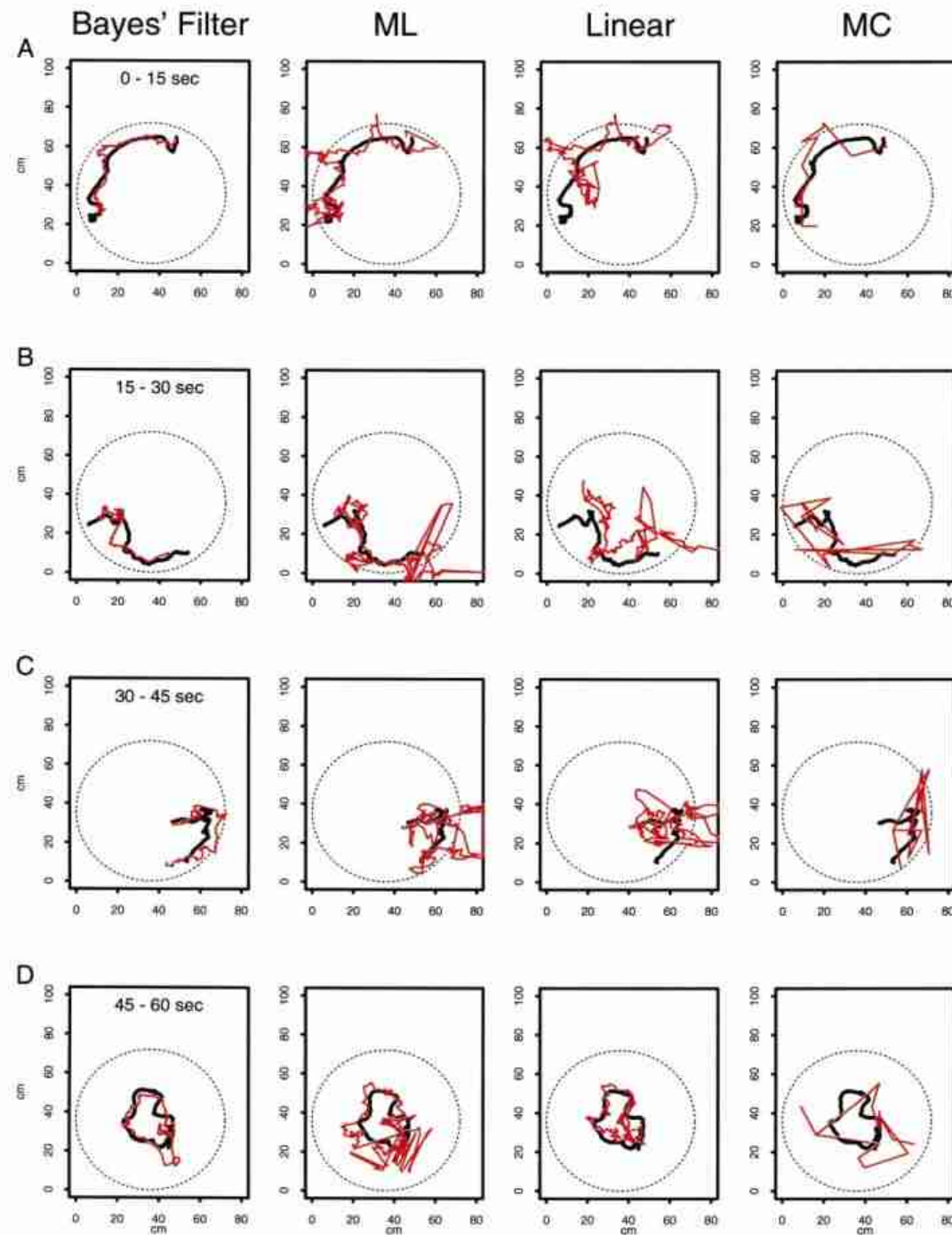
$$W(t_k|t_{k-1}) = W_x(\Delta_k) + W(t_{k-1}|t_{k-1}); \quad (11)$$

Posterior mode

$$\begin{aligned} \hat{x}(t_k|t_k) &= [W(t_k|t_{k-1})^{-1} + \sum_{c=1}^C A_c[\hat{x}(t_k|t_k), \theta(\Delta_k)]W_c^{-1}]^{-1} \\ &\quad \times [W(t_k|t_{k-1})^{-1}\hat{x}(t_k|t_{k-1}) + \sum_{c=1}^C A_c[\hat{x}(t_k|t_k), \theta(\Delta_k)]W_c^{-1}\mu_c]; \end{aligned} \quad (12)$$

Posterior variance:

$$\begin{aligned} W(t_k|t_k) &= [W(t_k|t_{k-1})^{-1} + \sum_{c=1}^C A_c[\hat{x}(t_k|t_k), \theta(\Delta_k)]W_c^{-1} \\ &\quad + \sum_{c=1}^C \lambda^c[\hat{x}(t_k|t_k)]\Lambda^c[\theta(\Delta_k)]W_c^{-1}(\hat{x}(t_k|t_k) - \mu_c)(\hat{x}(t_k|t_k) - \mu_c)'W_c^{-1}]^{-1}, \end{aligned} \quad (13)$$

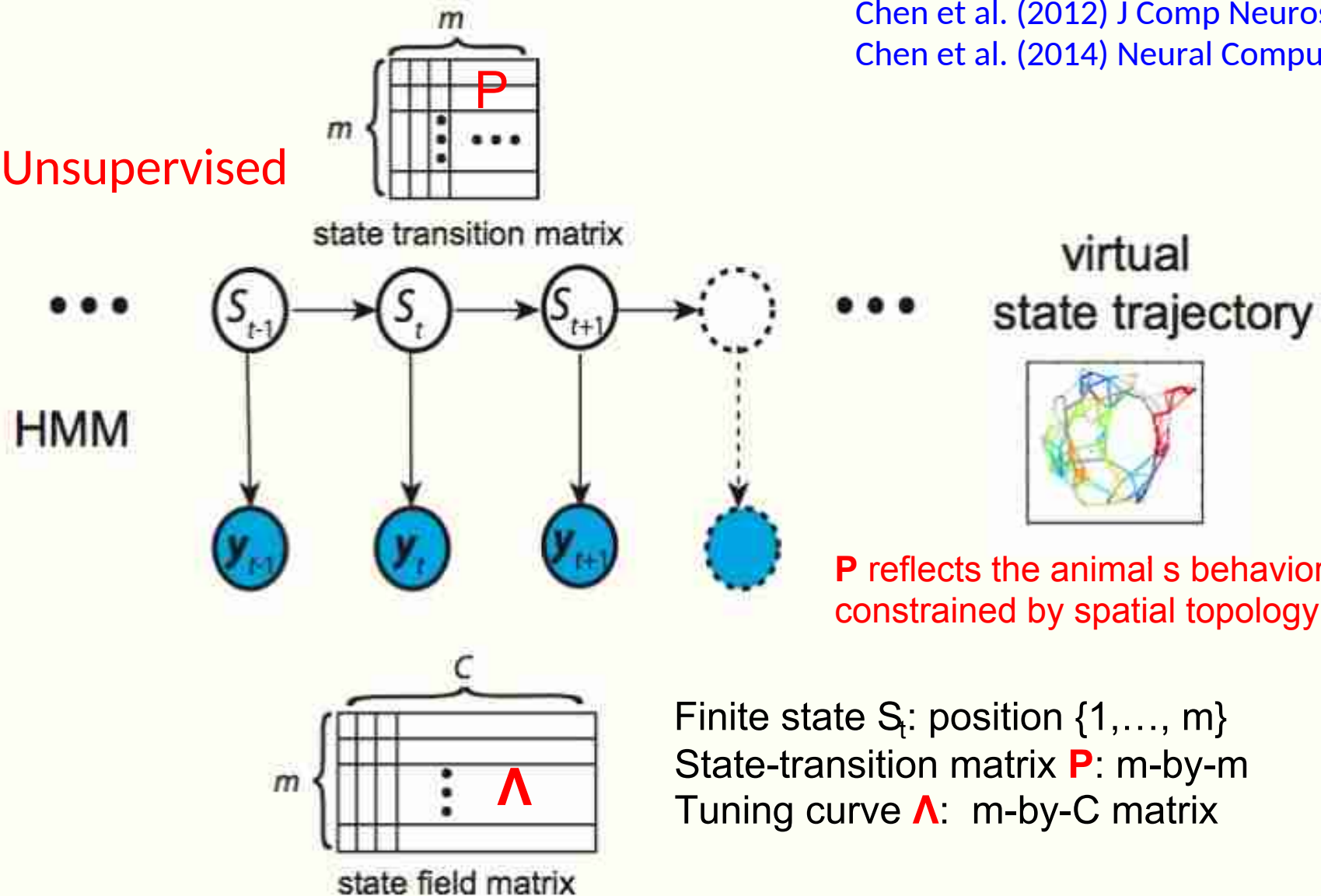


Brown et al. (1998) J Neurosci

HMM decoding hippocampal population codes

Chen et al. (2012) J Comp Neurosci
Chen et al. (2014) Neural Comput

Unsupervised



Finite state S_t : position $\{1, \dots, m\}$
State-transition matrix P : m -by- m
Tuning curve Λ : m -by- C matrix

$$y_{c,t} \sim \text{Poisson}(\lambda_{ci}; S_t=i)$$

Bayesian Inference

- Likelihood inference is inefficient for small data sample size; need to impose structured priors onto the constrained solution space ($m^2 + mC + m$)
- Bayesian inference: assume a **factorial** posterior and **conjugate** priors

$$\begin{aligned} q(\pi, P, \Lambda, S_{1:T}) &= q(\pi, P, \Lambda) q(S_{1:T}) \\ &\approx p(\pi, P, \Lambda, S_{1:T} | y_{1:T}) \end{aligned}$$

$$\begin{aligned} p(S_1 | \pi) &= \text{Multinomial}(S_1 | \pi), \\ p(S_t | S_{t-1}, P) &= \text{Multinomial}(S_t | P_{S_{t-1}, :}), \\ p(\mathbf{y}_t | S_t, \Lambda) &= \prod_{c=1}^C \text{Poisson}(y_{c,t} | \lambda_{c,S_t}). \end{aligned}$$

Bayesian nonparametrics

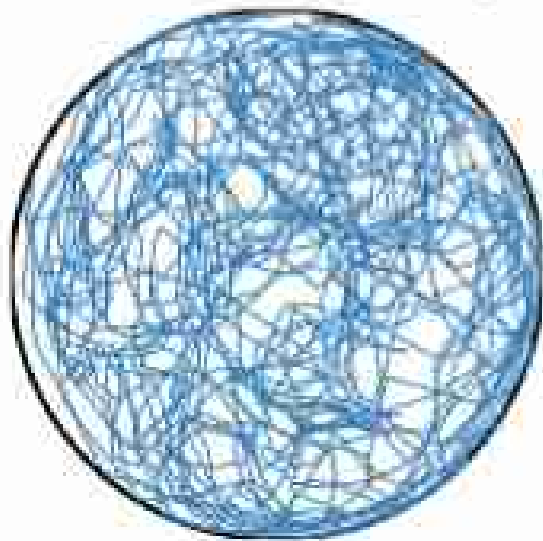
- Automatic model selection
- Nonparametric Bayesian: infinite HMM ($m \rightarrow \infty$; hierarchical Dirichlet process) and HDP-HMM

Linderman, Johnson, Wilson, Chen (2016) J. Neurosci. Methods.
Chen, Linderman, Wilson (2016). Proc. IEEE MLSP

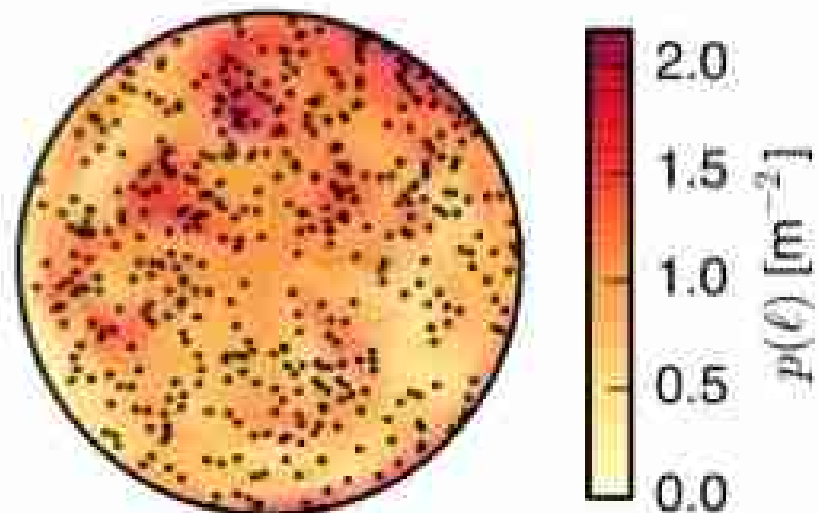
Rat open field data

47 hippocampal neurons, 9.8 min recording, 1 m diameter

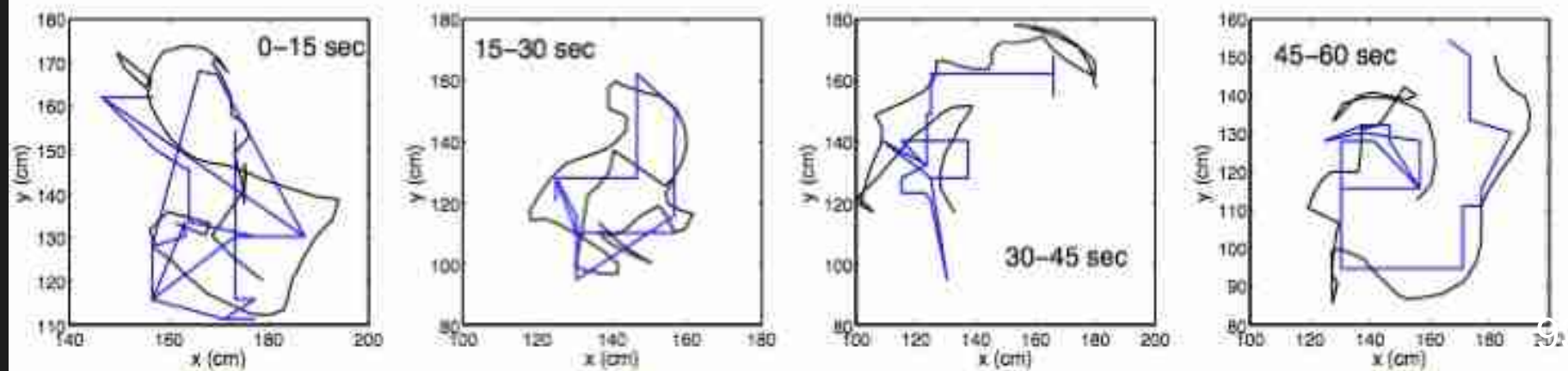
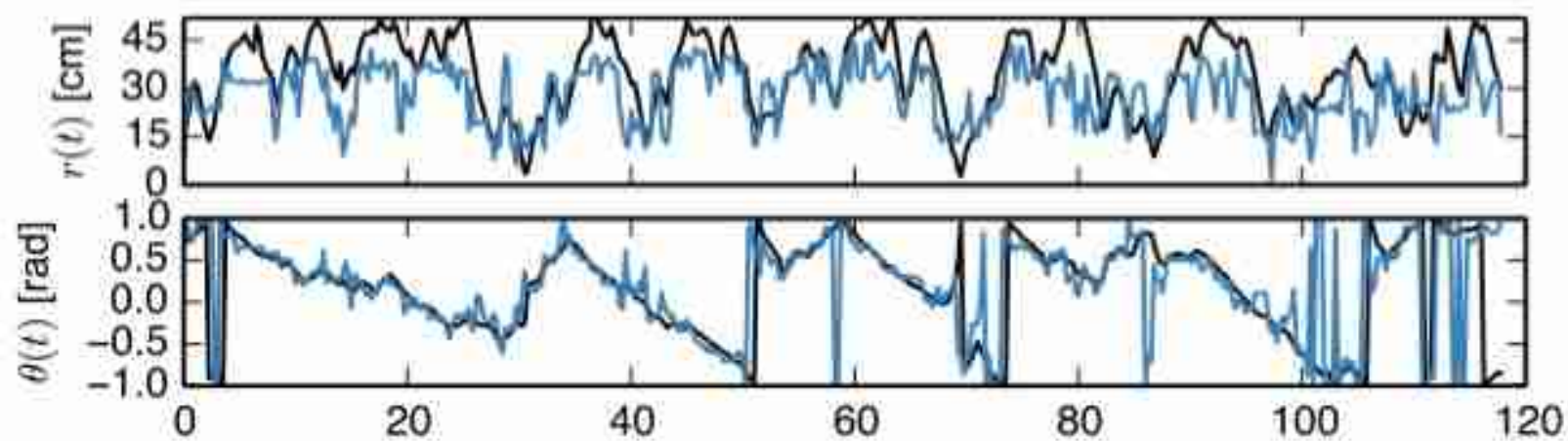
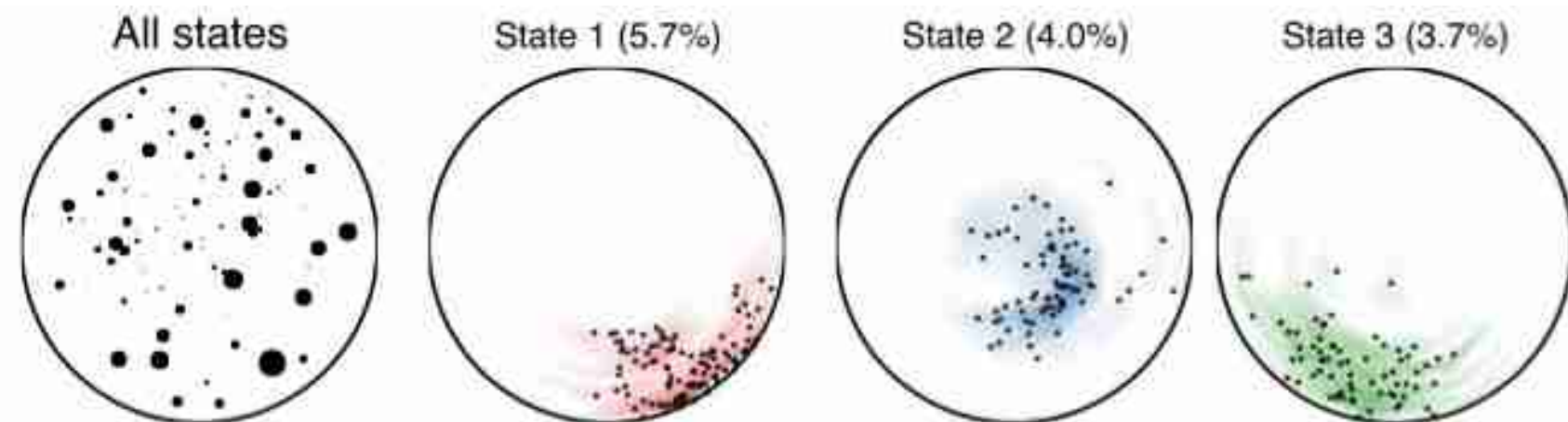
Location Trajectory



Empirical Location Distribution



Data: Loren Frank & Matt Wilson

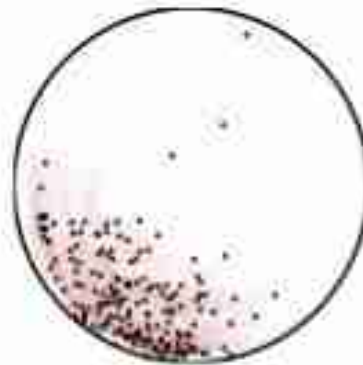


Place fields vs. state fields

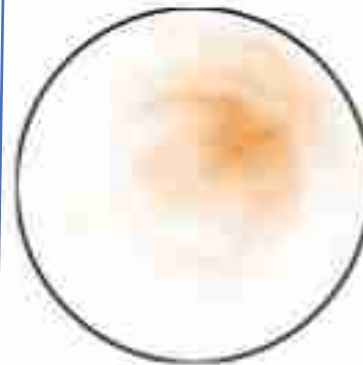
Inferred Place Field 1



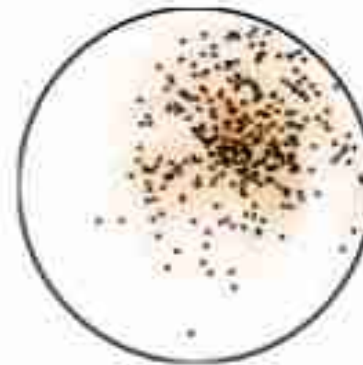
True Place Field 1



Inferred Place Field 14



True Place Field 14



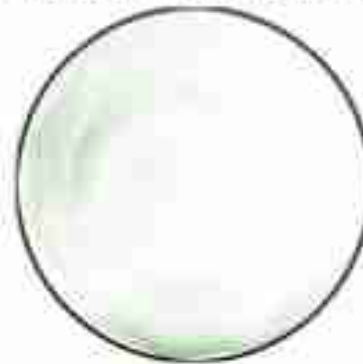
Inferred Place Field 29



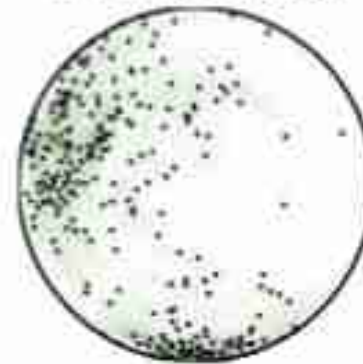
True Place Field 29



Inferred Place Field 39



True Place Field 39



Motor control via Kalman filtering

Latent variable model: LDS

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{w}_k$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{v}_k$$

Inference

$$\hat{\mathbf{x}}_{k|k} = \mathbf{A}\hat{\mathbf{x}}_{k-1|k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{C}\mathbf{A}\hat{\mathbf{x}}_{k-1|k-1})$$

$$\mathbf{P}_{k|k-1} = \mathbf{A}\mathbf{P}_{k-1|k-1}\mathbf{A}^T + \mathbf{Q}$$

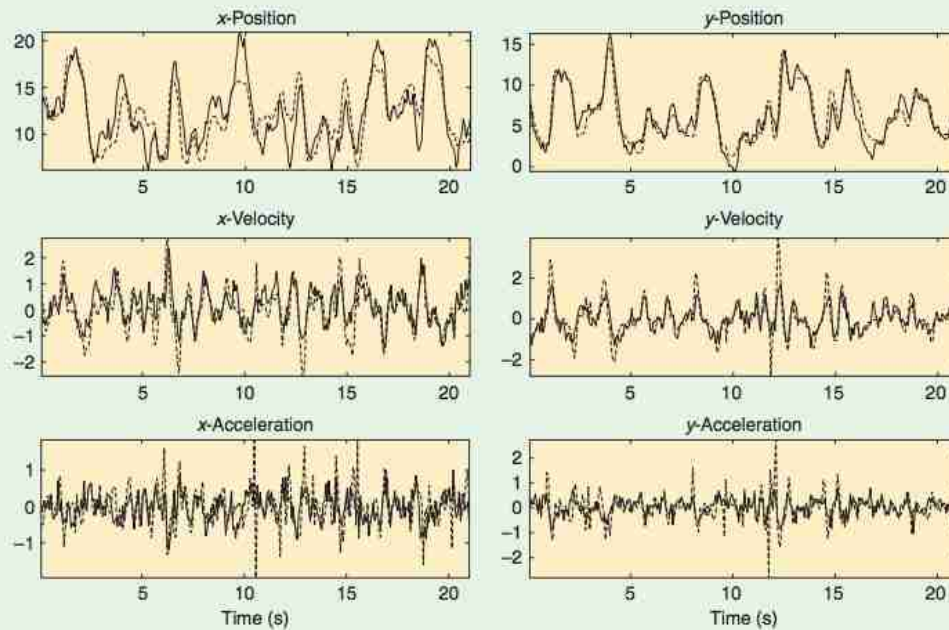
$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{C}^T(\mathbf{C}\mathbf{P}_{k|k-1}\mathbf{C}^T + \mathbf{R})^{-1}$$

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{C})\mathbf{P}_{k|k-1}$$

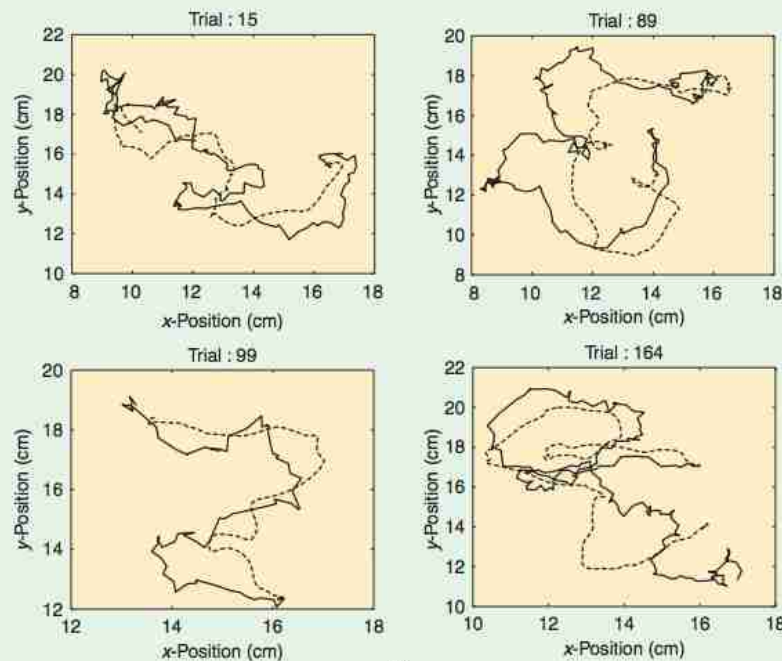
Augmented latent variable model

$$\begin{pmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{pmatrix} = \mathbf{A} \begin{pmatrix} \mathbf{x}_{k-1} \\ \mathbf{z}_{k-1} \end{pmatrix} + \mathbf{w}_k$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{z}_k + \mathbf{v}_k$$



(a)



(b)

Figure 6. Decoding M1 population activity. (A) Pinball task. Reconstruction of 6-dimensional kinematic state: true (dashed) and reconstructed (solid) kinematic variables. Only 21 s of test sequence are shown. (B) Pursuit tracking task. Reconstruction of hand position on a few test trials: true (dashed) and reconstructed (solid) hand trajectories. (Adapted from [38] with permission, © MIT Press, 2005). (Figure courtesy of Dr. Wei Wu at Florida State University).

Likelihood: Gaussian $\rightarrow$ Poisson

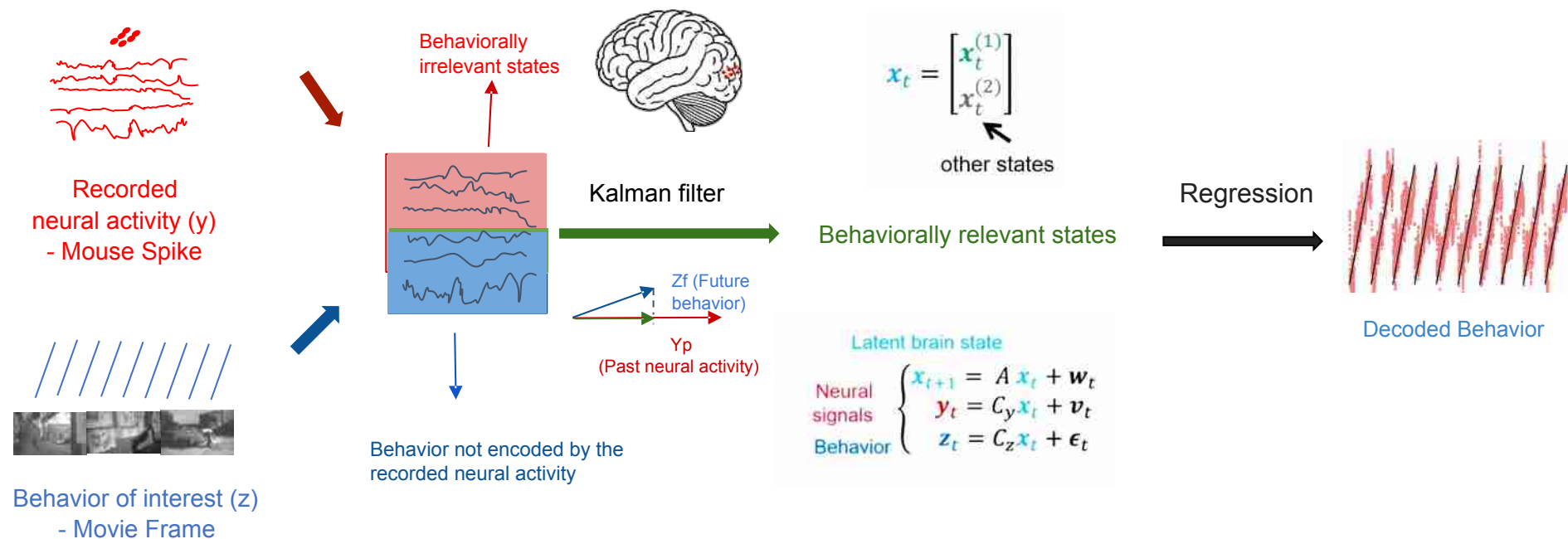
- $\mathbf{x}_t = \{M_t, V_t\}$ denotes 3D hand position and velocity (6-dimensional)
- Generalized linear model (GLM) in the likelihood function
- Learning and Inference: Gaussian approximation, particle filter

$$\mathbf{x}_t = \begin{bmatrix} M_t \\ V_t \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{3 \times 3} & \Delta t \mathbf{I}_{3 \times 3} \\ \mathbf{0} & 0.98 \mathbf{I}_{3 \times 3} \end{bmatrix} \mathbf{x}_{t-1} + \begin{bmatrix} \mathbf{0} \\ \mathbf{e}_t \end{bmatrix},$$
$$y_{t,c} \sim \text{Poisson}(\lambda_c(M_{t-\text{lag}}, V_{t-\text{lag}}, \sigma_c \mathbf{e}_{t,c})),$$

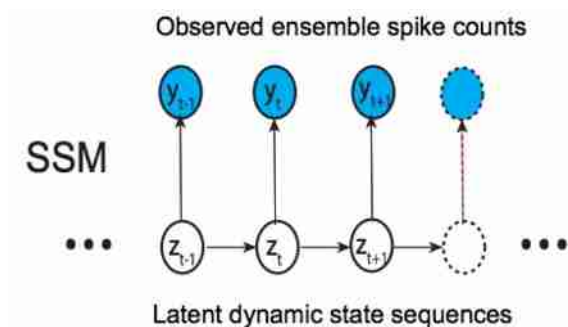
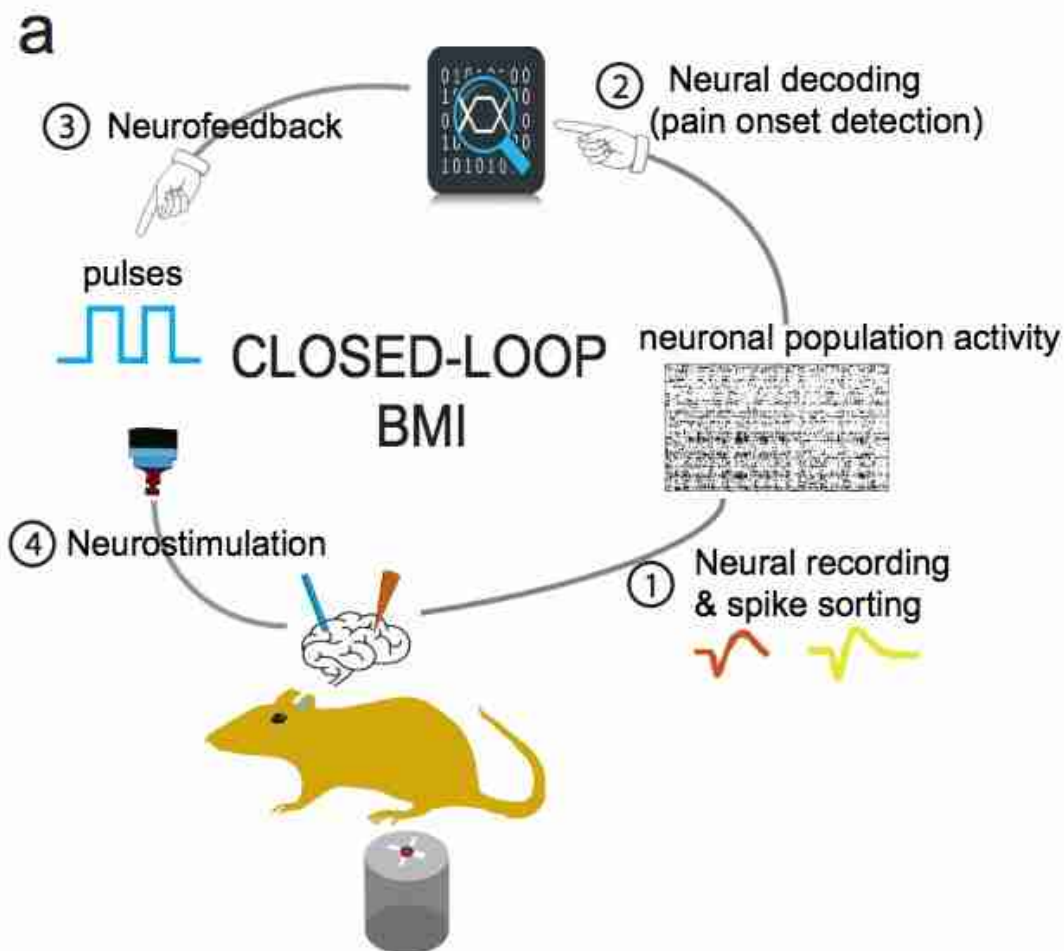


Modeling behaviorally relevant neural dynamics enabled by preferential subspace identification

Omid G. Sani¹, Hamidreza Abbaspourazad¹, Yan T. Wong^{2,5}, Bijan Pesaran² and Maryam M. Shanechi^{1,3,4} ✉



Detection of acute pain onset



Latent variable model: PLDS

$$\mathbf{z}_k = \mathbf{a}\mathbf{z}_{k-1} + \boldsymbol{\epsilon}_k$$

$$\boldsymbol{\eta}_k = \mathbf{c}\mathbf{z}_k + \mathbf{d}$$

$$\mathbf{y}_k \sim \text{Poisson}(\exp(\boldsymbol{\eta}_k)\Delta)$$

Inference

$$\hat{\mathbf{z}}_{k|k-1} = \mathbf{a}\hat{\mathbf{z}}_{k-1|k-1}$$

$$\mathbf{Q}_{k|k-1} = \mathbf{a}^2\mathbf{Q}_{k-1|k-1} + \boldsymbol{\sigma}^2$$

$$\hat{\mathbf{y}}_{k|k-1} = \exp(\mathbf{c}\hat{\mathbf{z}}_{k|k-1} + \mathbf{d})\Delta$$

$$\mathbf{Q}_{k|k}^{-1} = \mathbf{Q}_{k|k-1}^{-1} + \mathbf{c}^T \text{diag}(\hat{\mathbf{y}}_{k|k-1})\mathbf{c}$$

$$\hat{\mathbf{z}}_{k|k} = \hat{\mathbf{z}}_{k|k-1} + \mathbf{Q}_{k|k}\mathbf{c}^T(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1})$$

Pain detection

$$P(\text{Z-score} > \hat{z}_{k|T}) = 1 - P(\text{Z-score} \leq \hat{z}_{k|T})$$

$$= 1 - \int_{-\infty}^{\hat{z}_{k|T}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

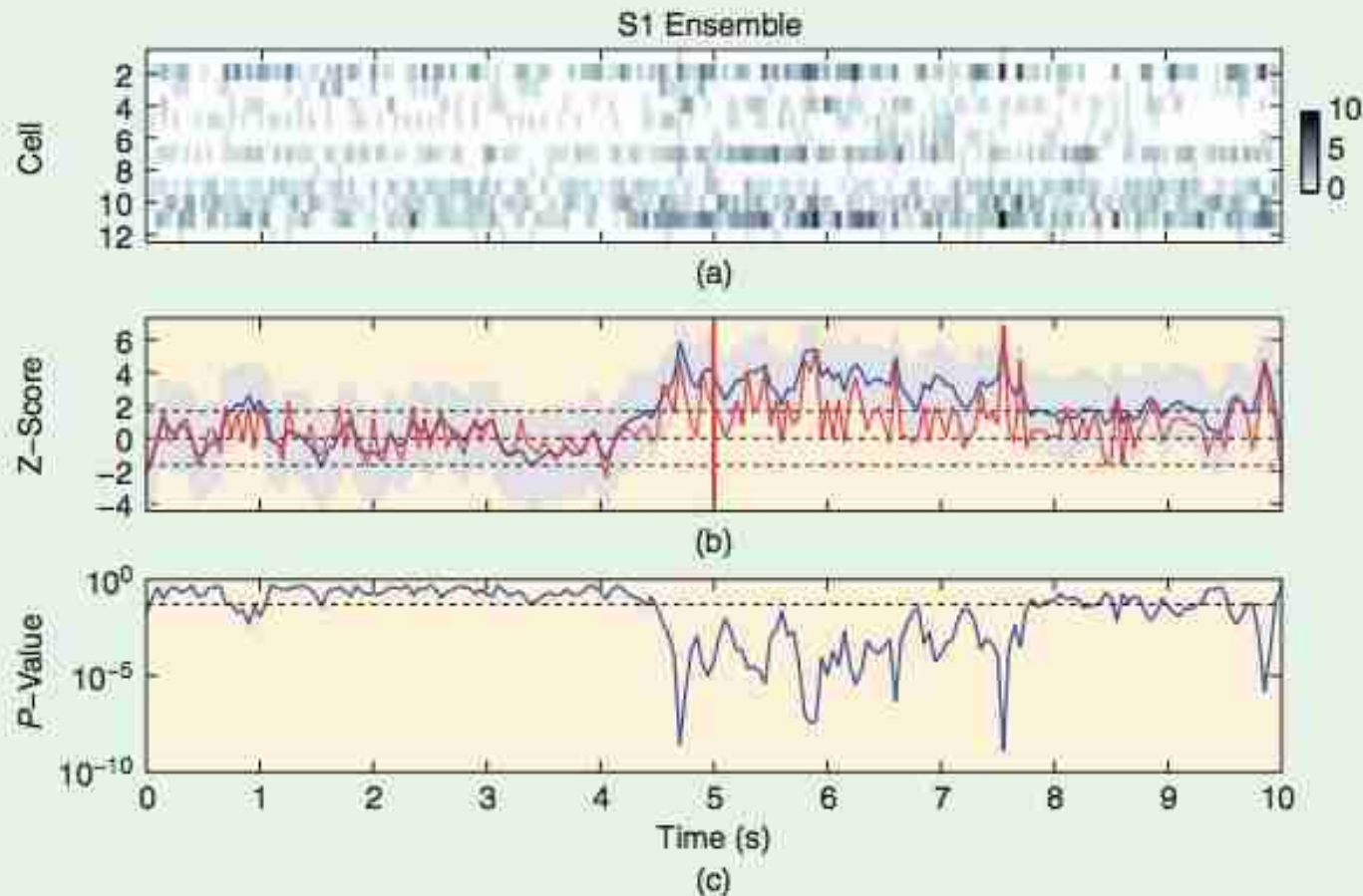
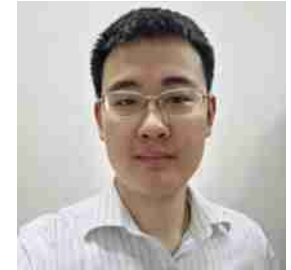
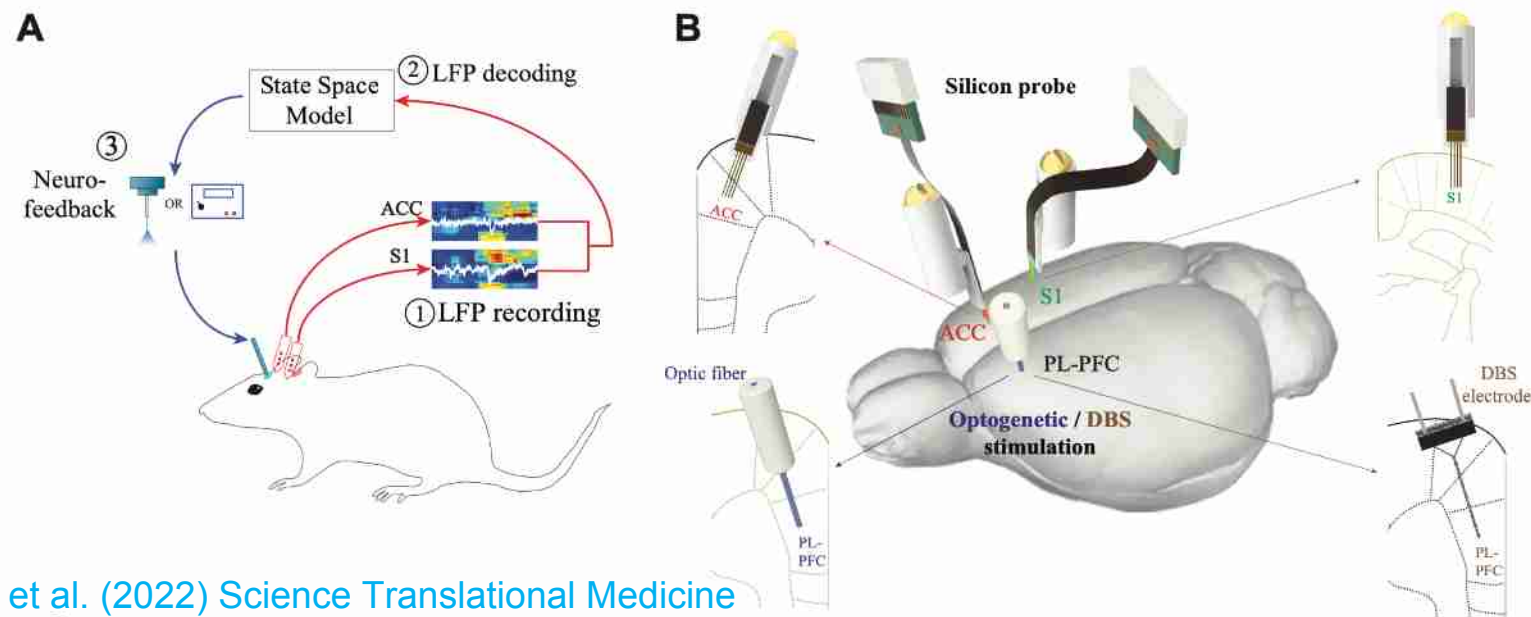
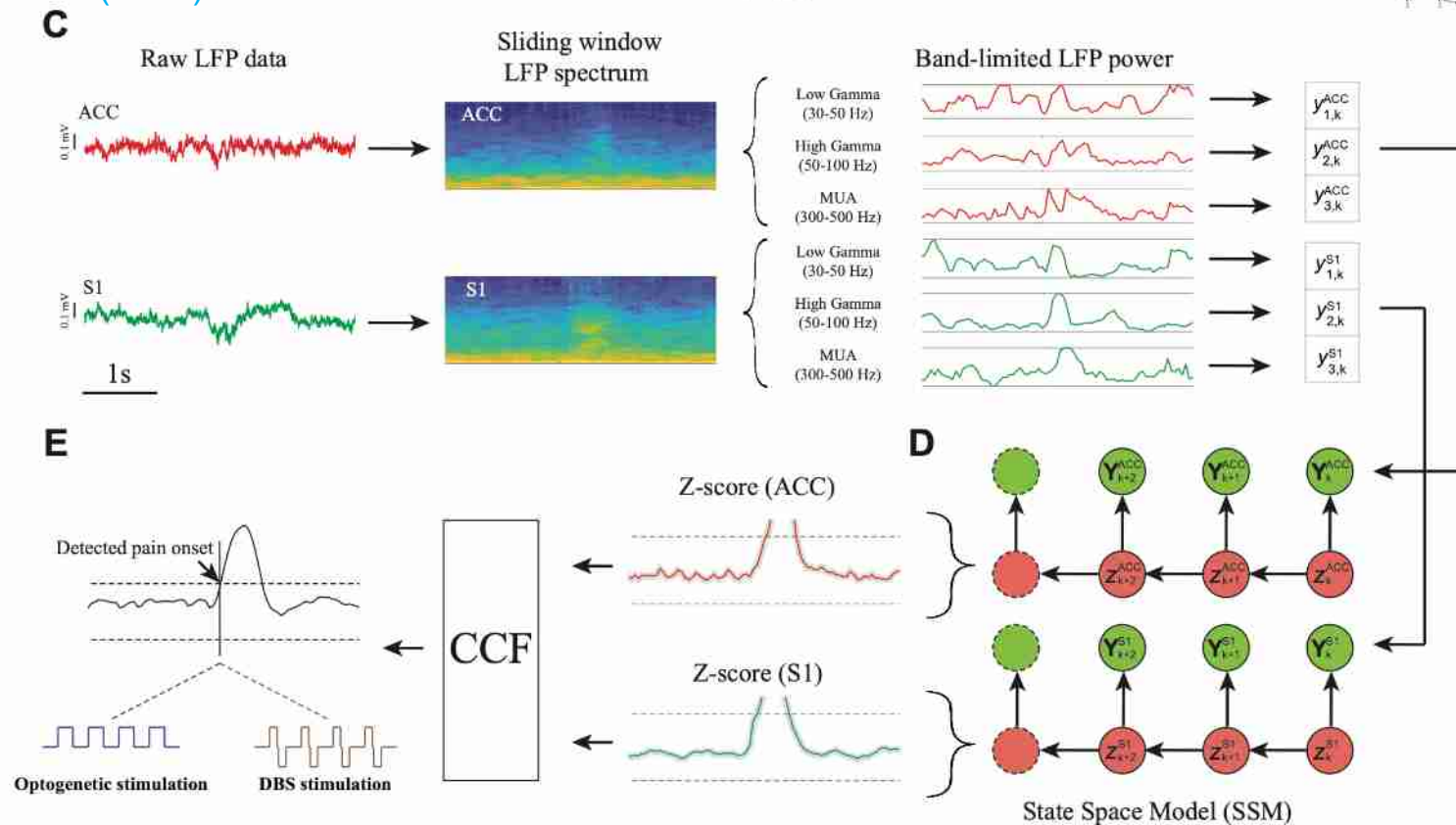


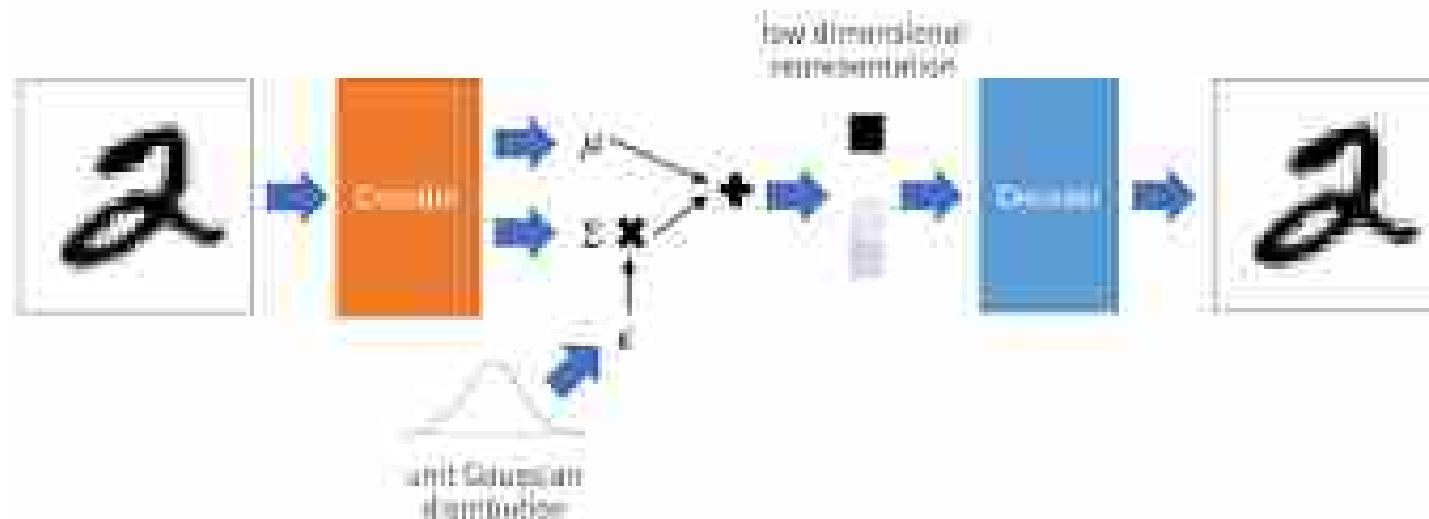
Figure 8. Detecting acute thermal pain signals. (A) neuronal ensemble spike counts (12 single units) from the rat S1 area. Bin size 50 ms. Color bar indicates spike count, with dark color representing high spike count. (B) Estimated mean Z-score (blue curve) from the smoothed state estimate $\hat{z}_{k|T}$. Vertical red line indicates the animal's paw withdrawal—an indicator of acute pain behavior. Horizontal dashed lines mark the thresholds of significant zone. Shaded area marks the confidence intervals. (C) Equivalent P-value derived from the mean Z-score (blue curve) of panel B (Adapted from [69], © IEEE, 2016).



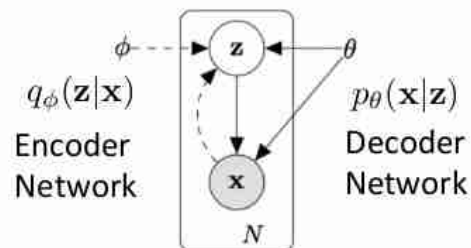
Sun et al. (2022) Science Translational Medicine



Variational autoencoder (VAE) and deep VAE

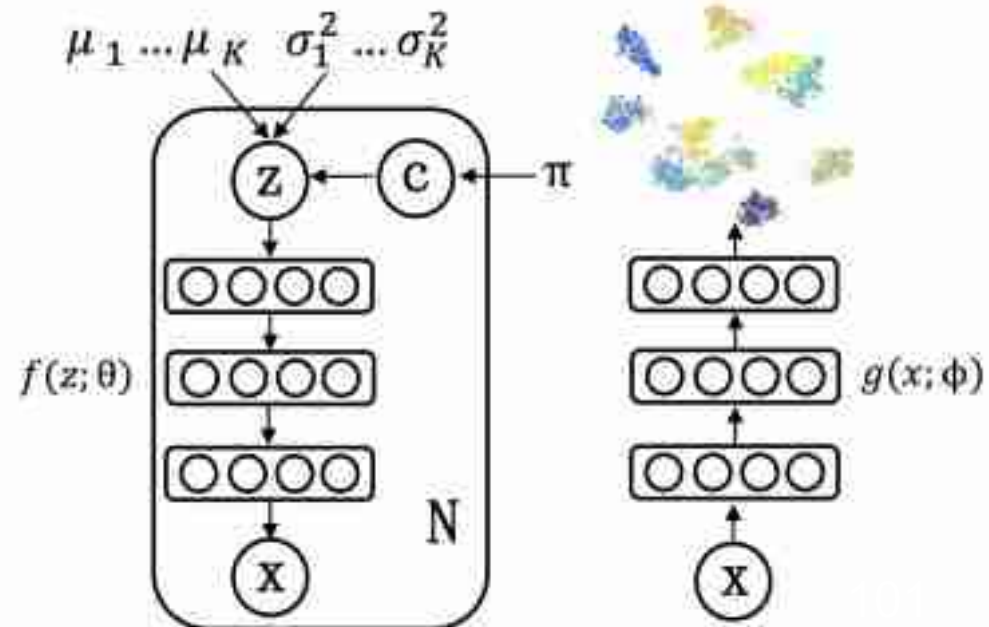


Variational Autoencoder

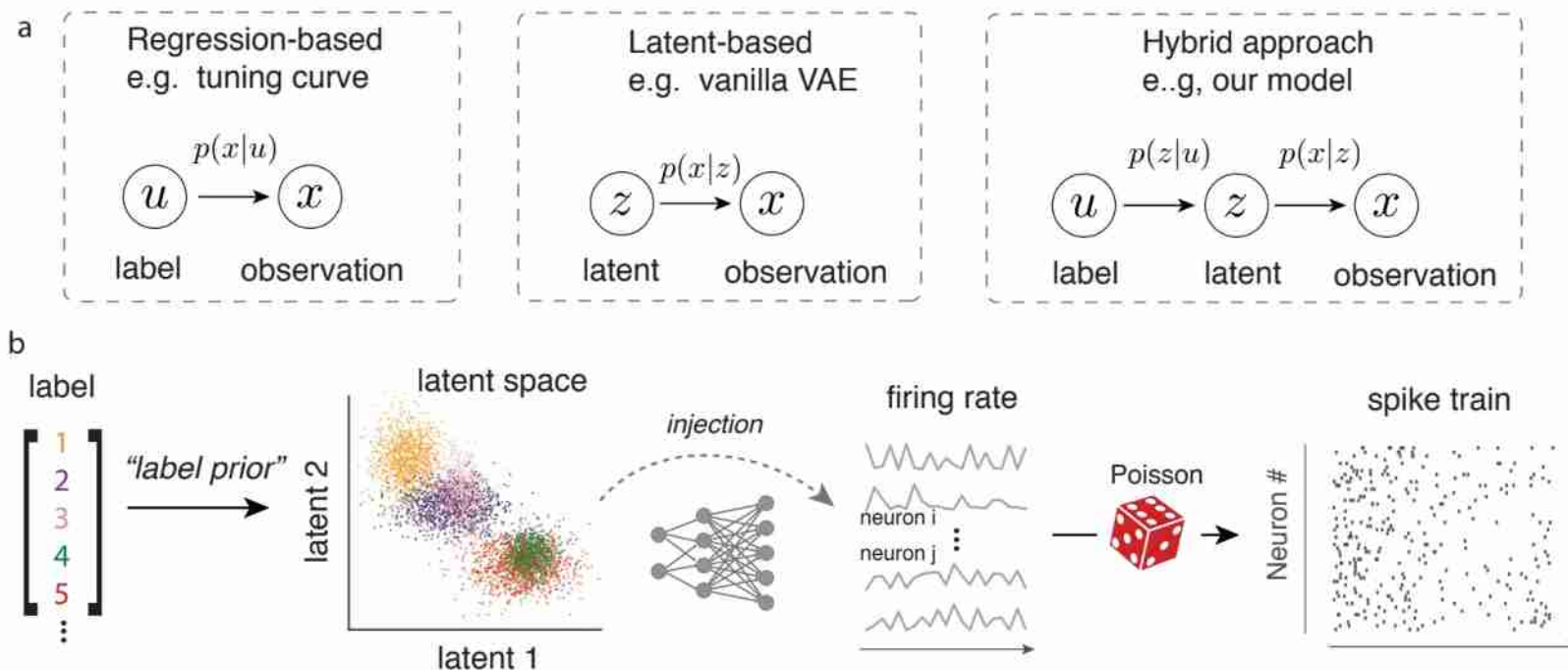


Minimize: $D_{KL}[q_\phi(z|x)||p_\theta(z|x)]$

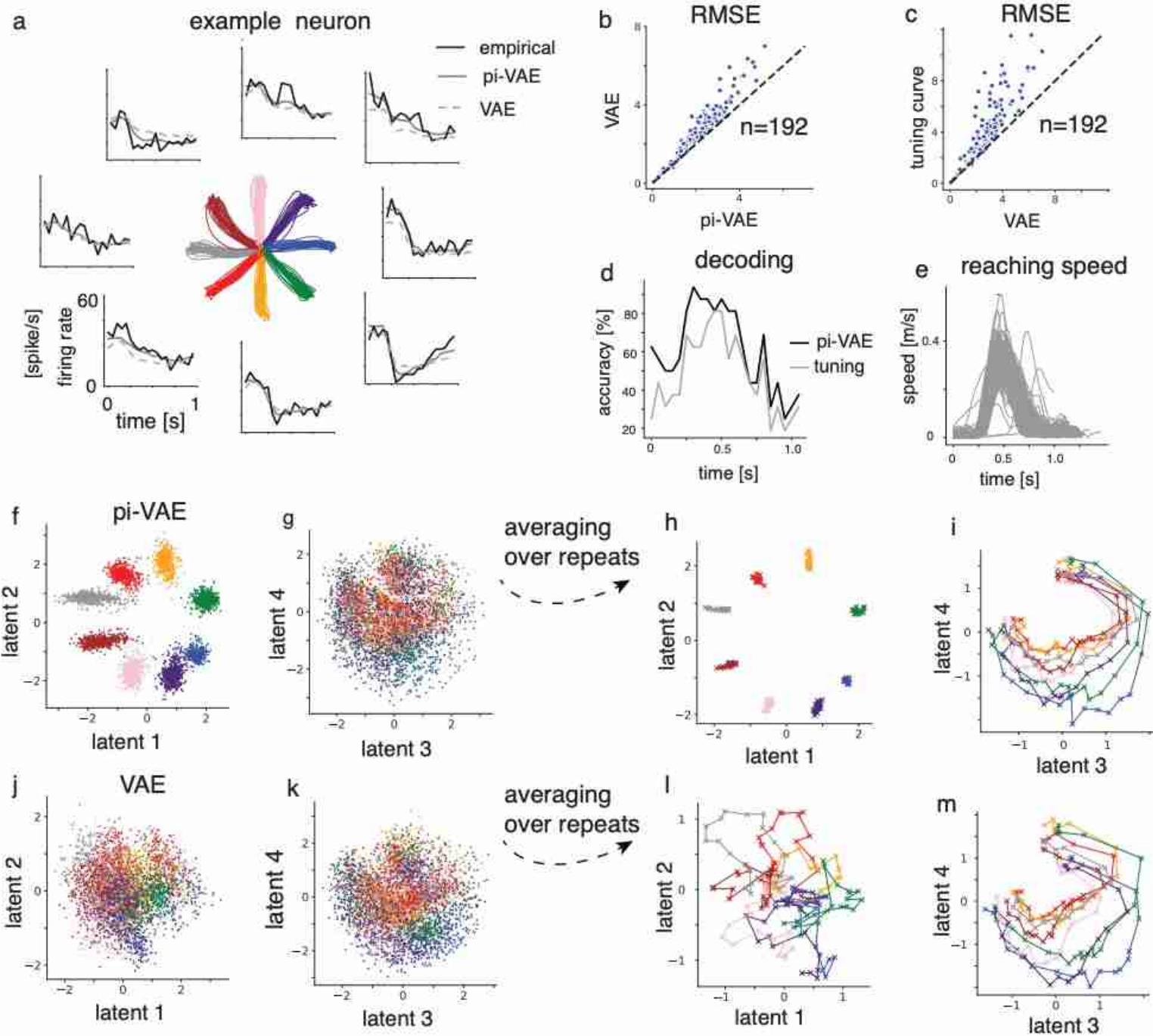
Intractable: $p_\theta(z|x) = \frac{p_\theta(x|z)p_\theta(z)}{p_\theta(x)}$



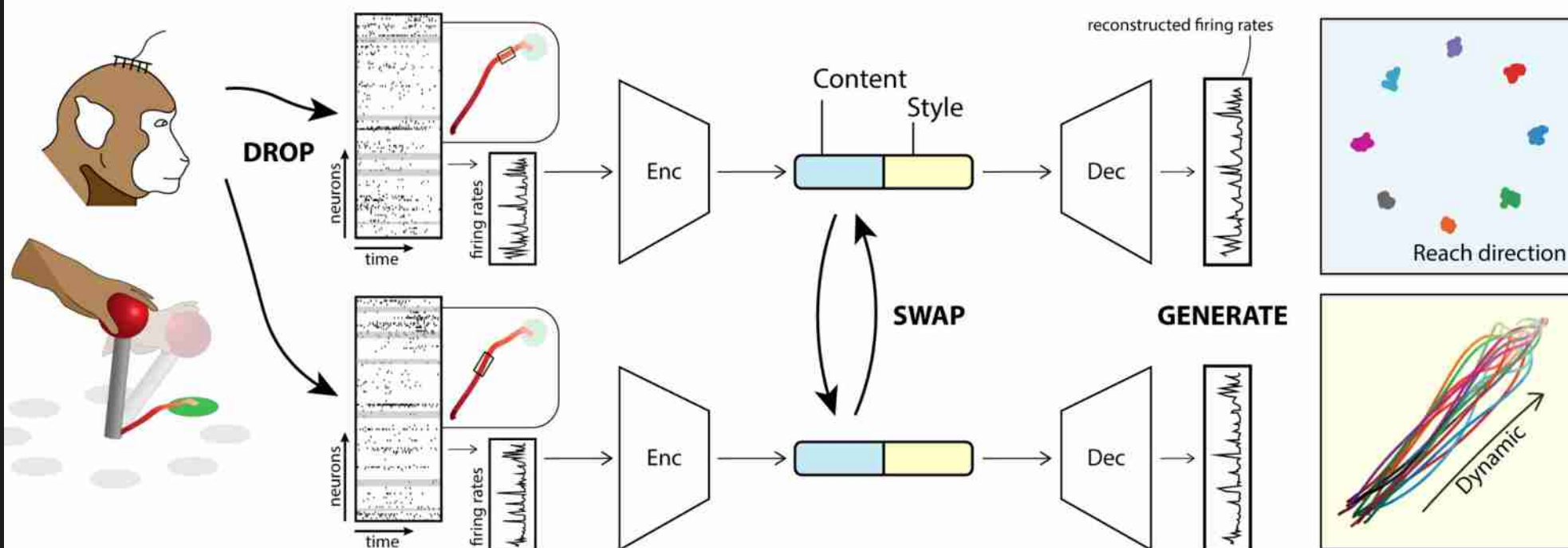
pi-VAE: latent variable model of neural activity while simultaneously modeling the relationship between latent and task-relevant variables



Zhou and Wei, NIPS 2020

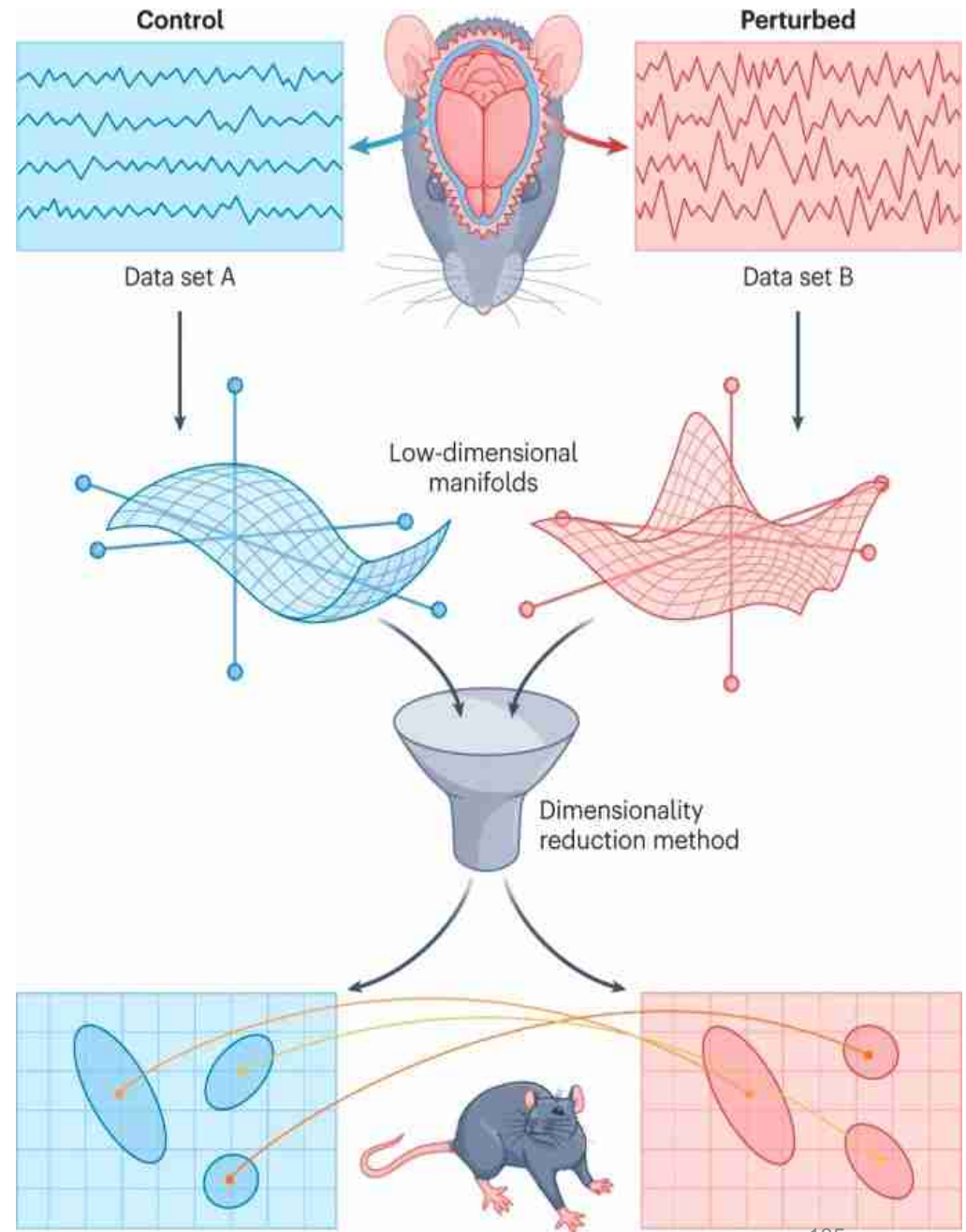


Generative AI: self-supervised learning approach for generating neural activity

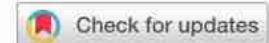


Manifold alignment

- Latent spaces cannot be directly compared
- Finding correspondence across distinct high-dimensional neural recordings

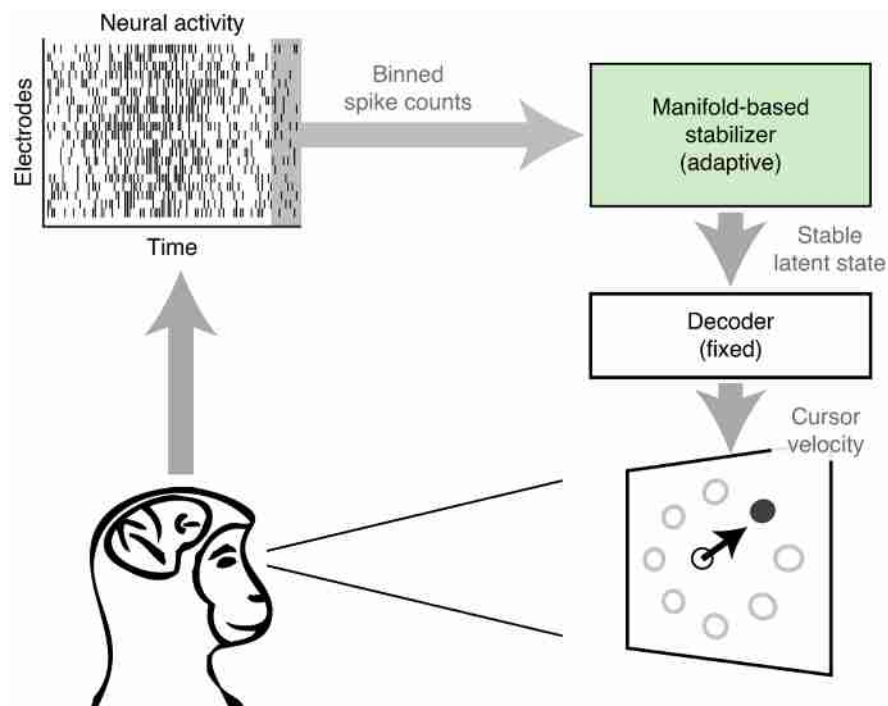


Ham, Lee, Saul, ICML'2003, UAI'05
Dabagia et al., Nat. Biomed. Eng. 2023



Stabilization of a brain–computer interface via the alignment of low-dimensional spaces of neural activity

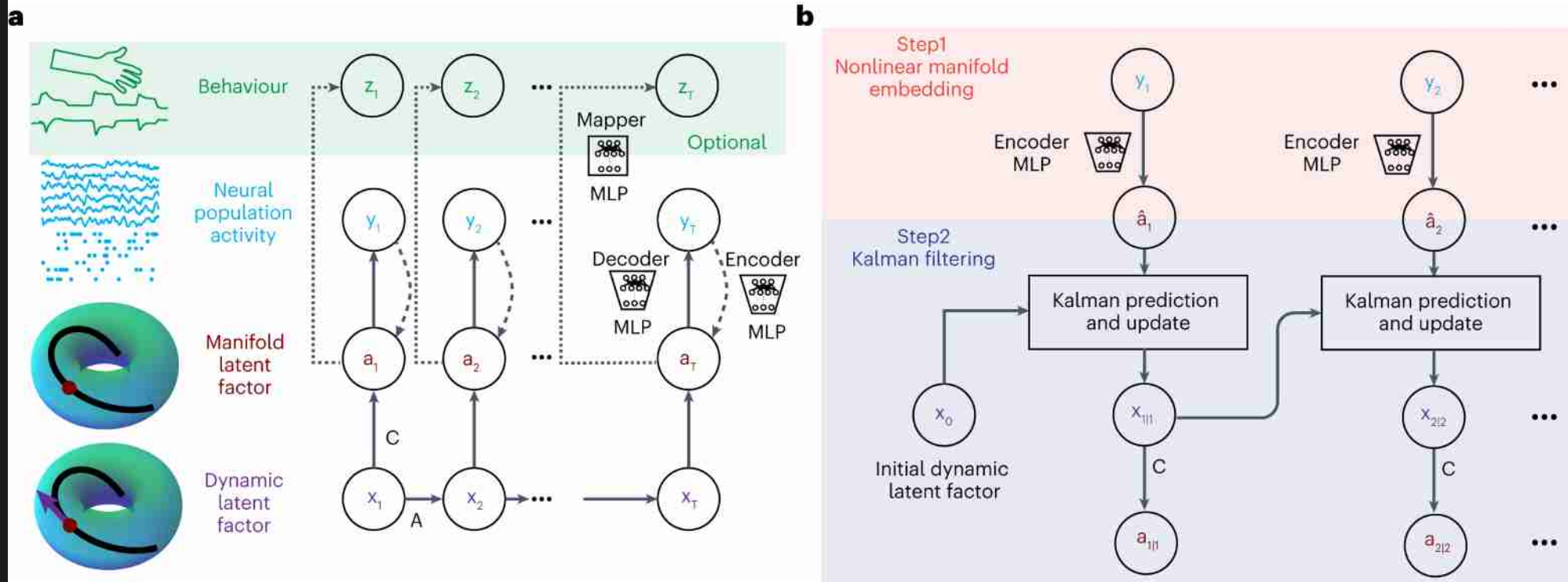
Alan D. Degenhart^{1,2,3,4,5,14}, William E. Bishop^{3,6,7,14}, Emily R. Oby^{2,3,4,5,8}, Elizabeth C. Tyler-Kabara^{2,9,10,11}, Steven M. Chase^{3,12,13,15}, Aaron P. Batista^{2,3,4,5,15} and Byron M. Yu^{1,3,12,13,15} ✉



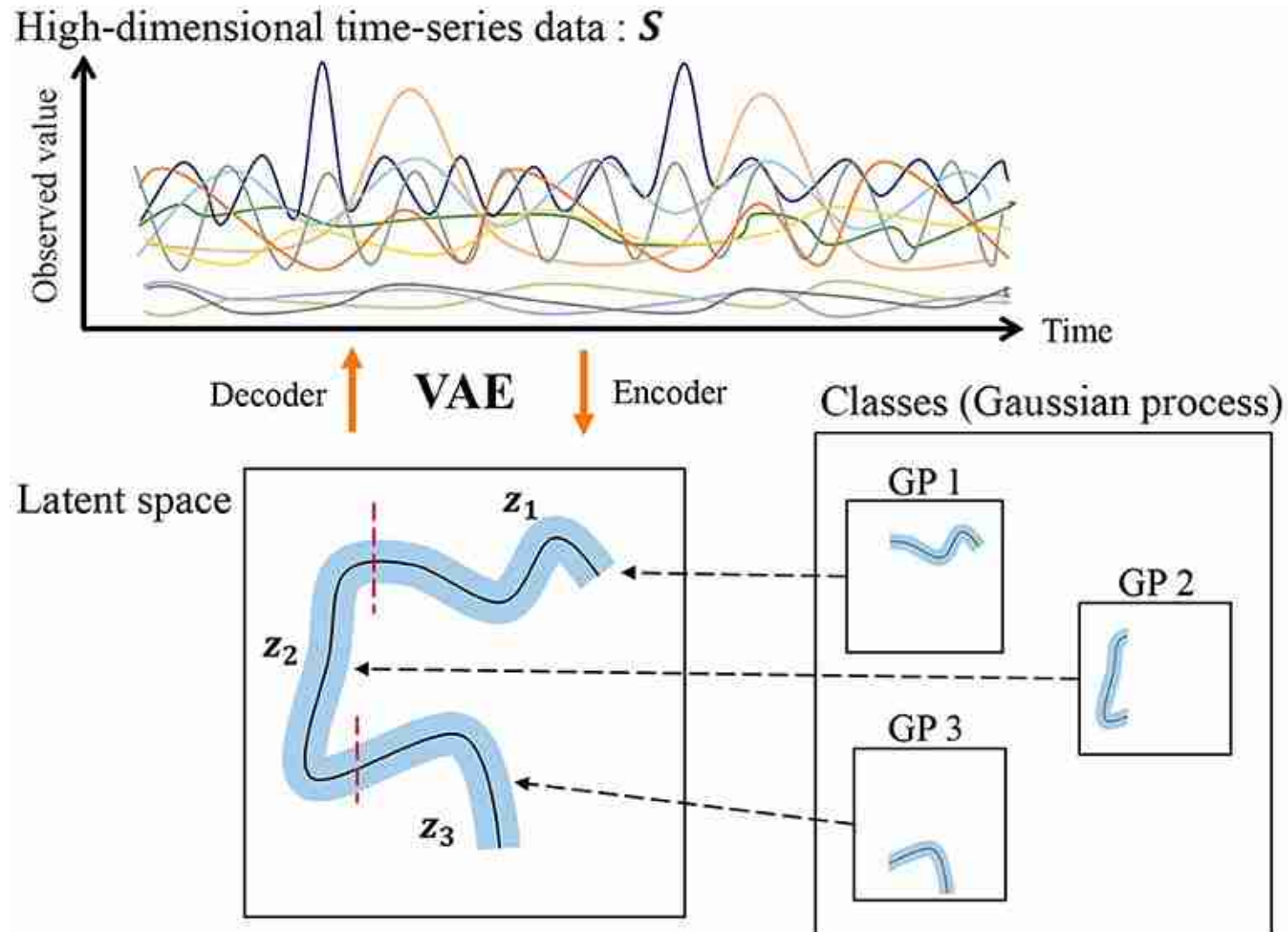
Neural recording instability
can result in poor performance in BMI

Latent model + nonlinear manifold embedding (VAE, DNN)

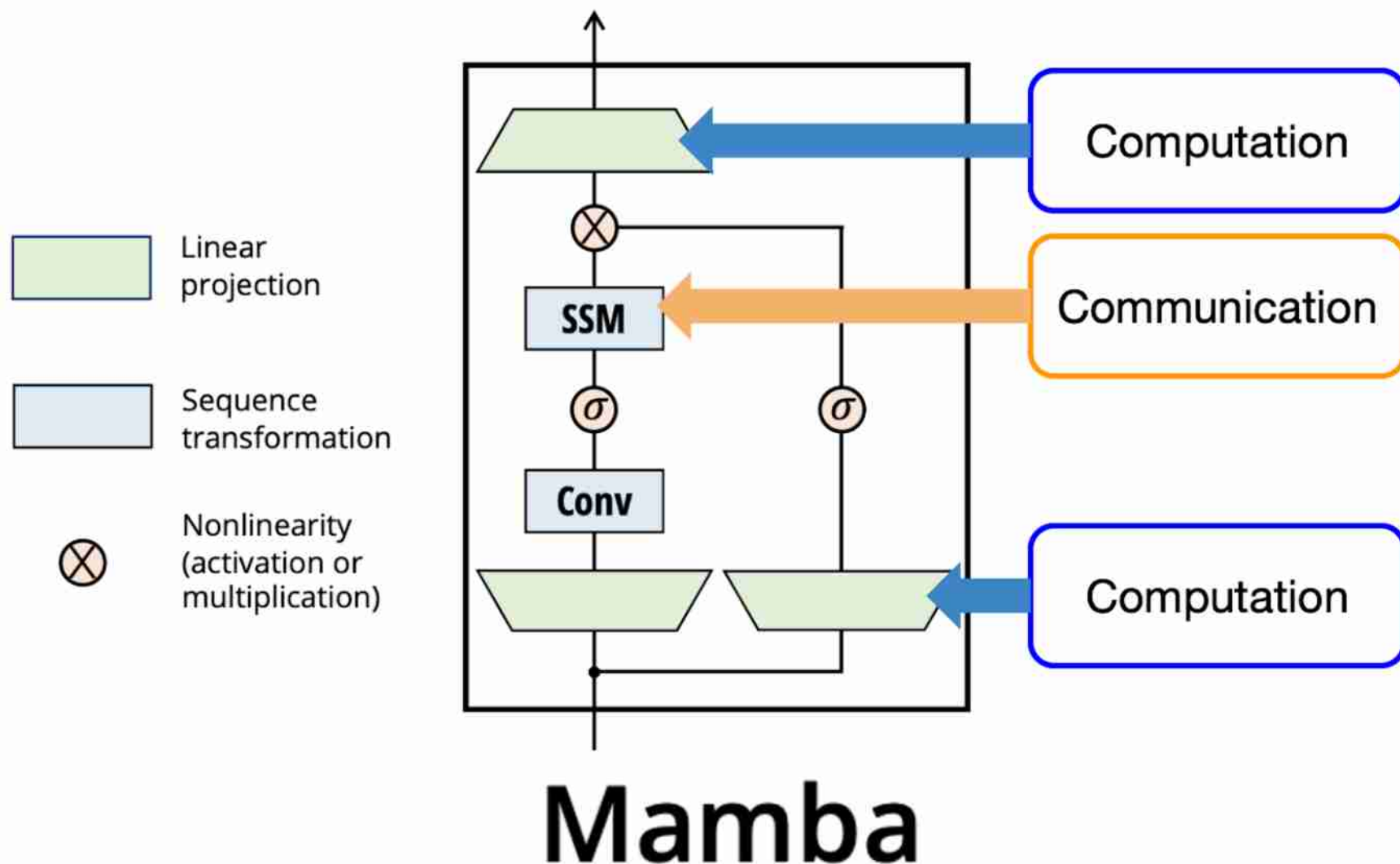
Learning: unsupervised, self-supervised (masking)



VAE-GP-HSMM



Mamba: control theory-inspired State Space Model (SSM)



Take home messages

- Low-dimensional (dynamic or static) latent variable models provide a powerful framework for analysis of high-dimensional neural data (while simultaneously modeling behavioral data): inter-trial/subject variability
- Generative vs. discriminative (or energy-based) models
- Learning as an optimization problem, whereas inference as an inverse problem: latent source estimation, source deconvolution, detection

Software source

- <https://github.com/lindermanlab/ssm>
- <https://web.stanford.edu/~swl1/>
- <https://probml.github.io/ssm-book/root.html>

Homework: Training an HMM to infer animal's position from hippocampal spiking activity

- Derive the EM or VB algorithm for latent variable models (e.g., HMM, mixtures of factor analyzers)
- <https://direct.mit.edu/neco/article/26/1/1/7932/Neural-Representation-of-Spatial-Topology-in-the>
- <https://zenodo.org/record/5527940>
- <https://s-shinomoto.com/hmm/>