Cold Spring Harbor Laboratory

bioRxiv

THE PREPRINT SERVER FOR BIOLOGY

Direct neural perturbations reveal a dynamical mechanism for robust computation

Daniel J. O'Shea, Lea Duncker, Werapong Goo, Xulu Sun, Saurabh Vyas, Eric M. Trautmann, Ilka Diester, Charu Ramakrishnan, Karl Deisseroth, Maneesh Sahani, Krishna V. Shenoy

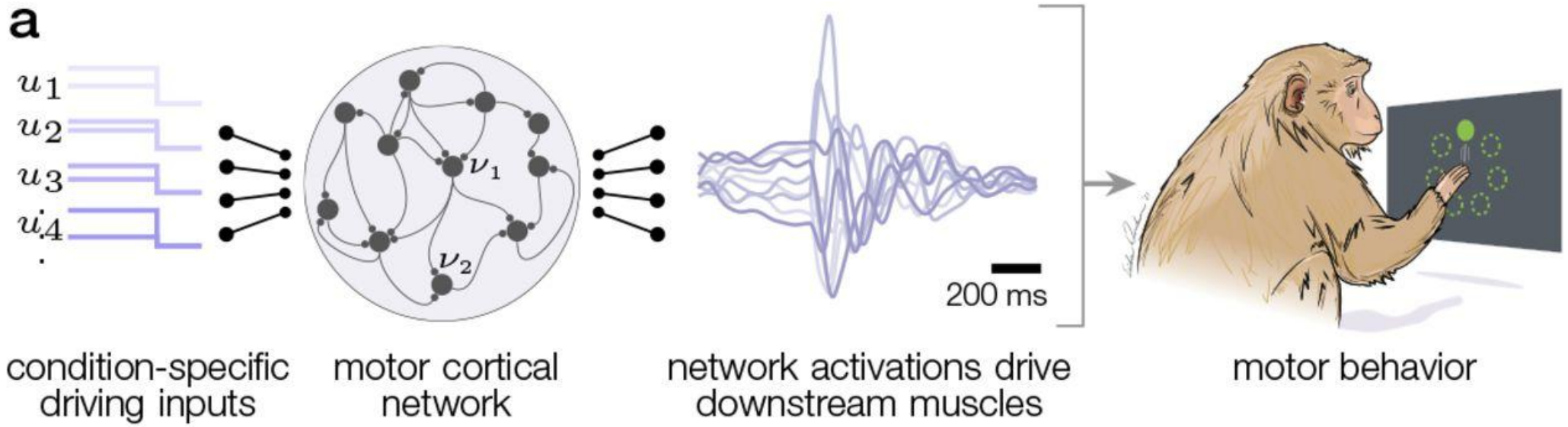
doi: https://doi.org/10.1101/2022.12.16.520768

3 0 4 0

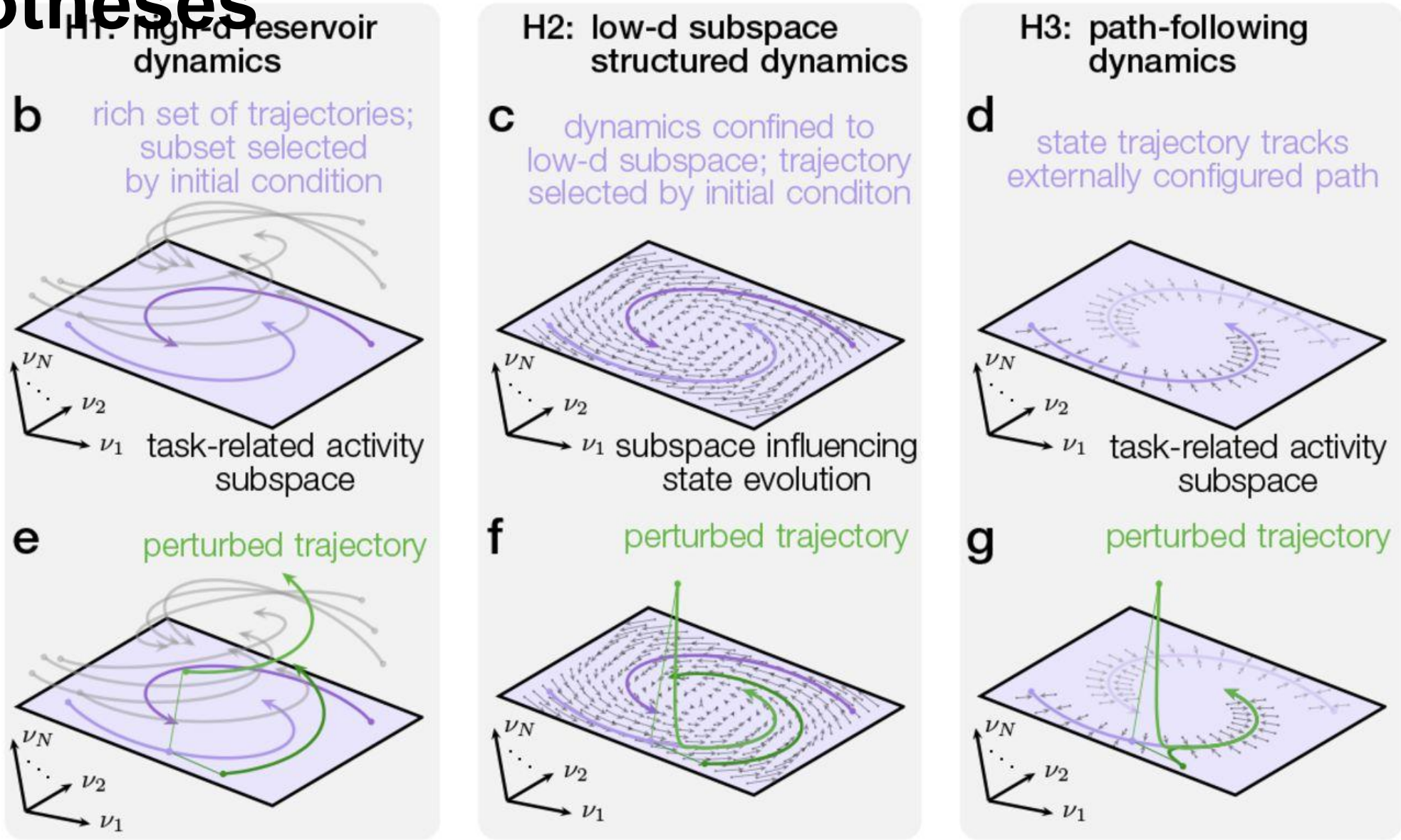
Abuzar
Mahmood
Katz Lab
CSNIC 2/15/24

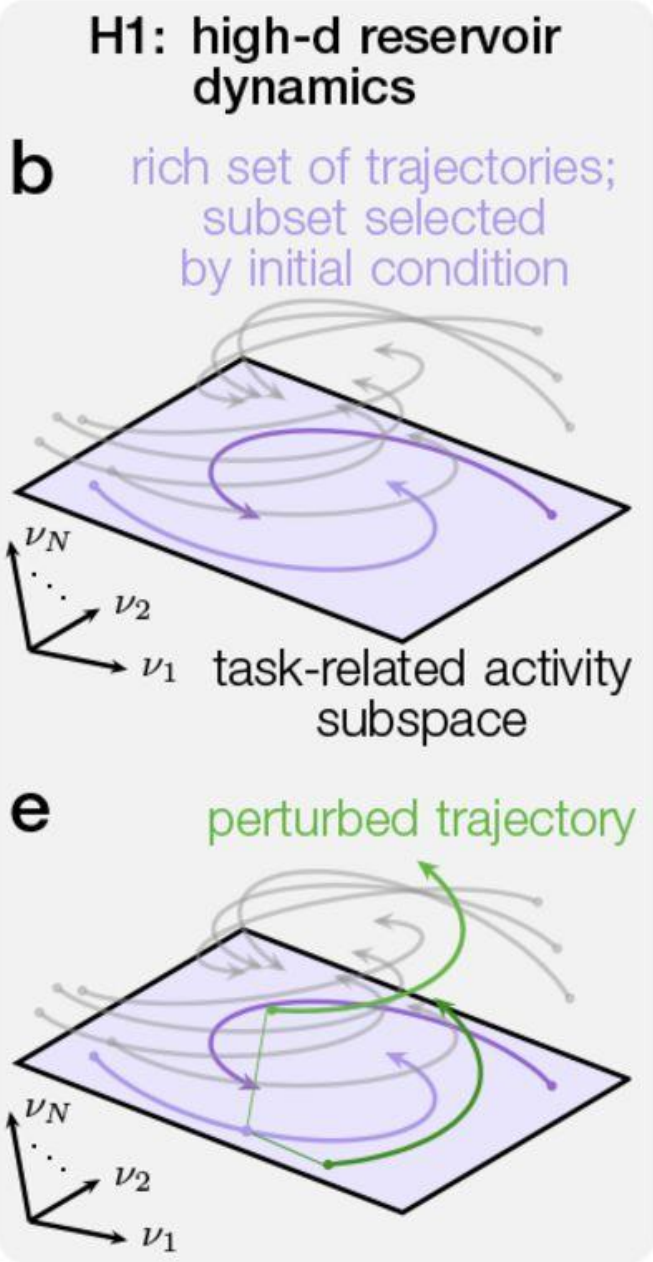
- Intro + Background
  - Hypotheses (total of 3)
  - Experimental Setup
- Results (knock out 1 hypothesis with each section):
  - Optogenetics
  - Intra-Cranial Microstimulation

Introduction



Introduction : Hypotheses





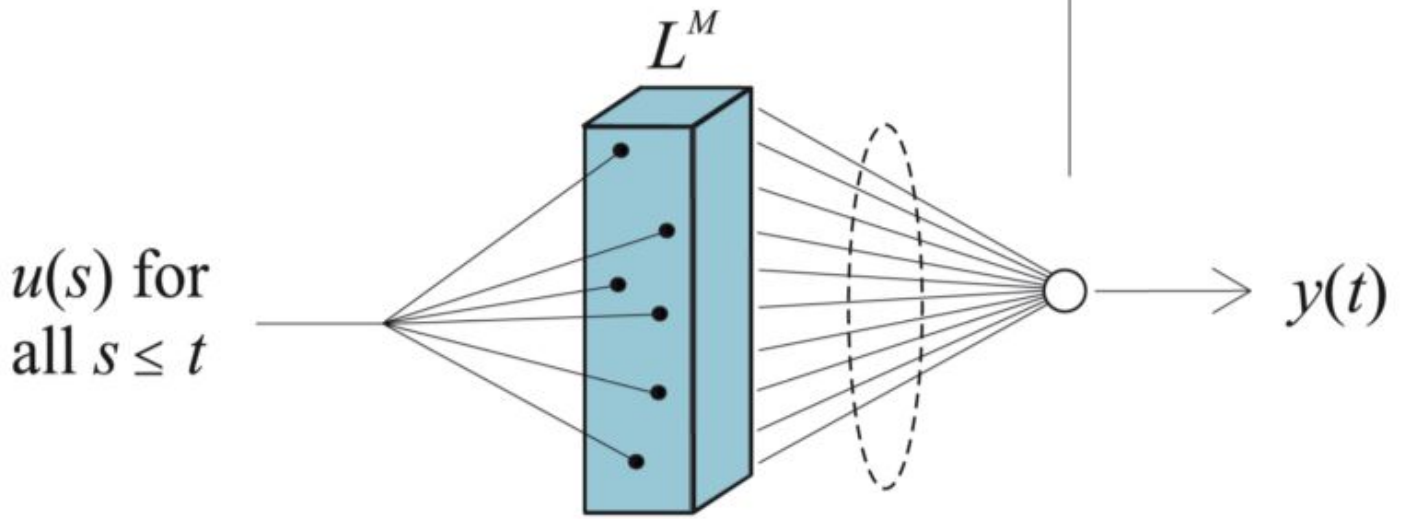
Hypothesis 1

- In reservoir networks, which retain the concurrent potential to express basis patterns in many different dimensions, perturbations of neural state should engage new dynamical modes and thus evoke complex, long-lasting transients (**Fig. 1e**)

Background : Reservoir Computing

a bank of basis filters
(or some more general dynamical system)

memoryless readout,
trained for a specific task

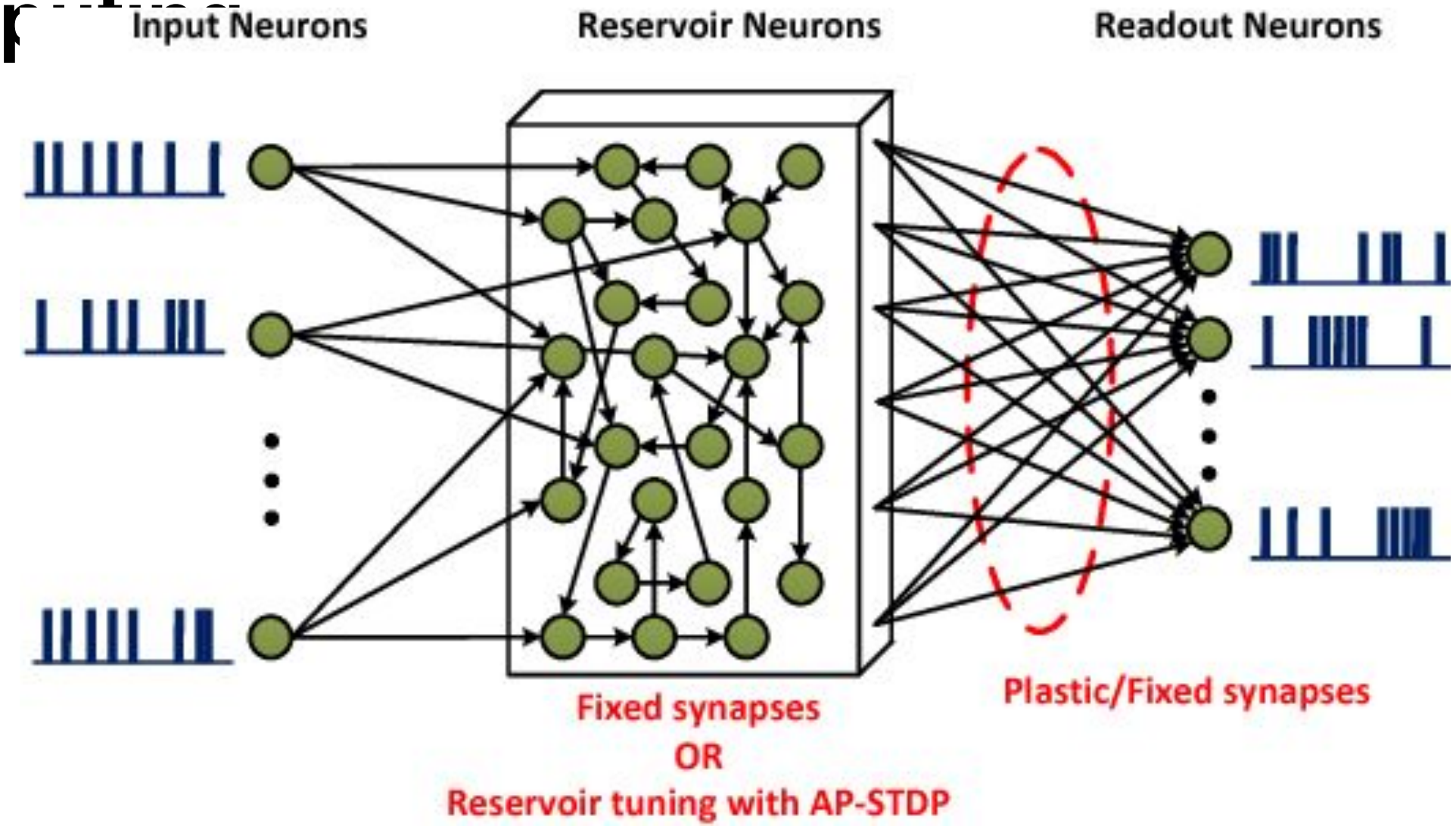


= liquid state of the
Liquid State Machine

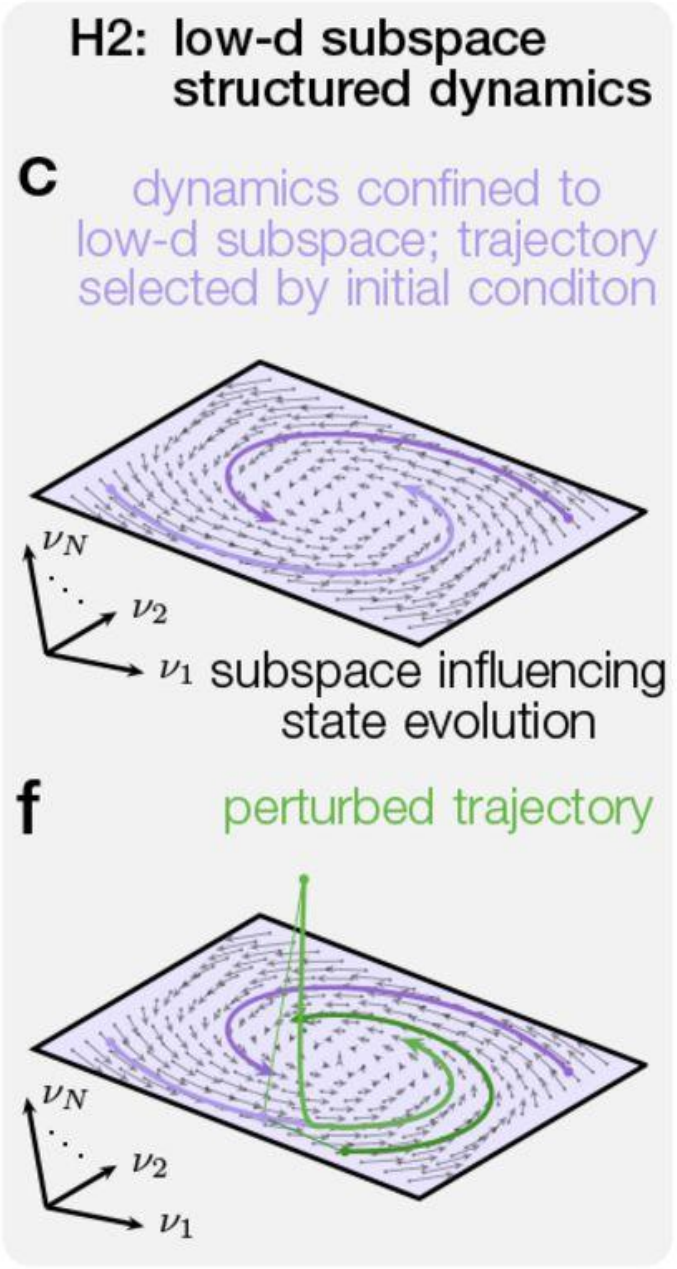
$$x^M(t) = (L^M u)(t)$$
$$y(t) = f^M(x^M(t))$$

1. Maass W. Liquid Computing. In: Cooper SB, Löwe B, Sorbi A, editors. Computation and Logic in the Real World [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007 [cited 2024 Feb 13]. p. 507–16. (Lecture Notes in Computer Science; vol. 4497). Available from: http://link.springer.com/10.1007/978-3-540-73001-9_53

Background : Reservoir Computing

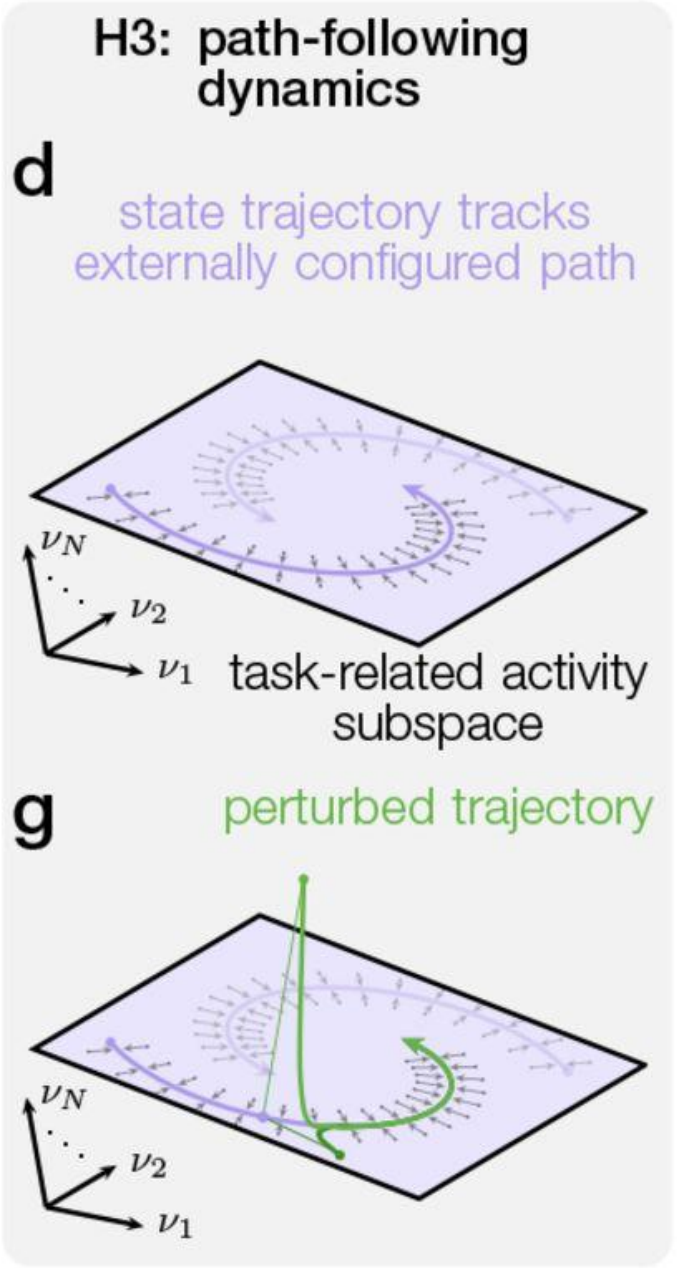


Wang, Qian and Peng Li. “D-LSM: Deep Liquid State Machine with unsupervised recurrent reservoir tuning.” *2016 23rd International Conference on Pattern Recognition (ICPR)* (2016): 2652-2657.



Hypothesis 2

- In subspace structured networks, where activity is driven by the state within a low-dimensional subspace, only perturbations that affect this subspace should elicit complex, long-lasting effects, while perturbations along all other dimensions should fail to engage with the circuit dynamics (**Fig. 1f**).



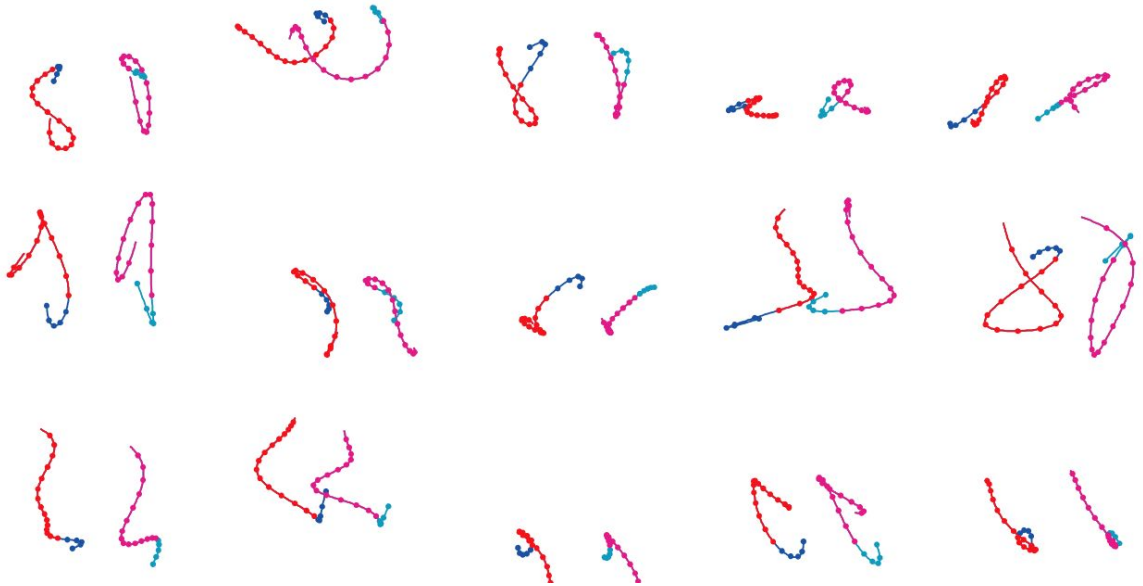
Hypothesis 3

- Lastly, in a path-following network, all perturbations away from the externally-configured trajectory will decay back rapidly (**Fig. 1g**).

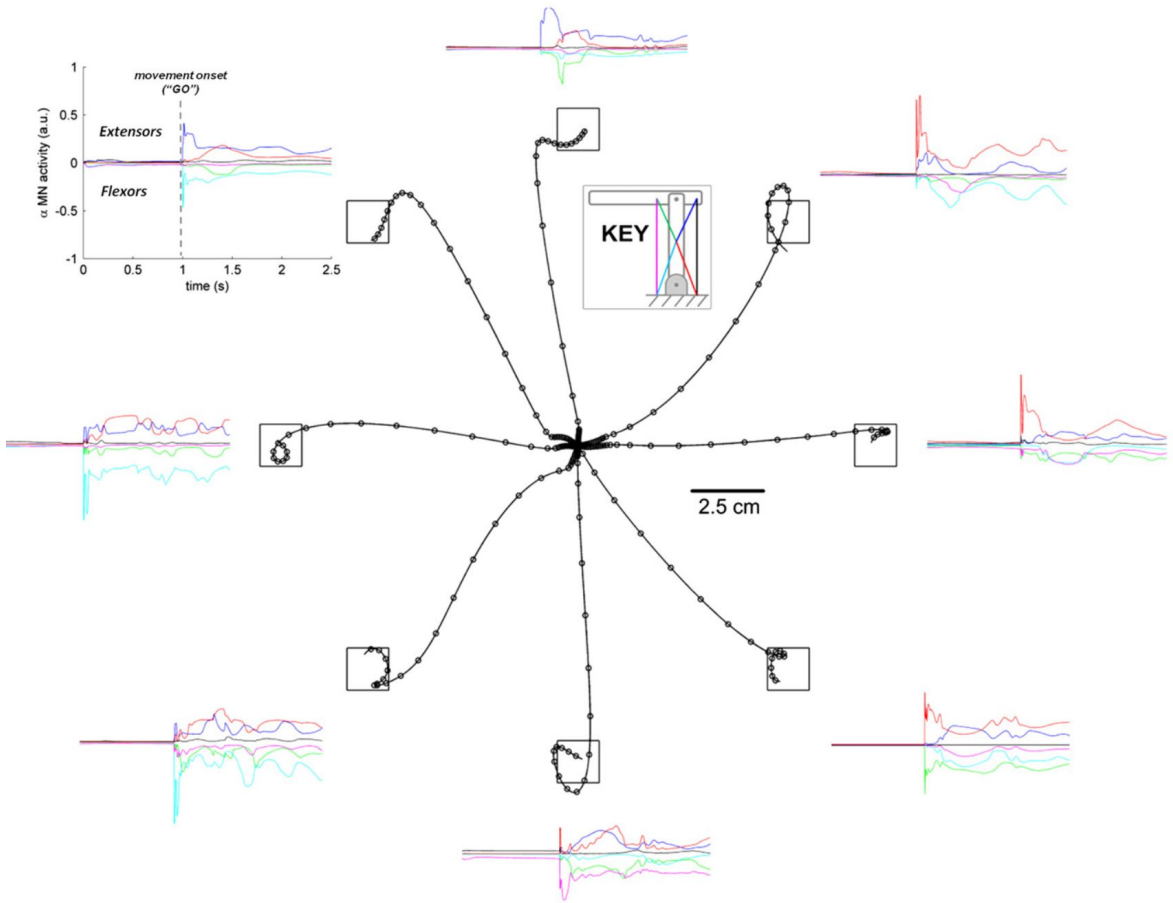
Background : Path-following Dynamics

- *path-following dynamics* (**H3**),
  - in which the neural state is constrained to move along an externally configured path. In the context of motor control, previous work has proposed that the motor cortex might serve to activate specific motor programs implemented by recurrent circuitry in the spinal cord.

A



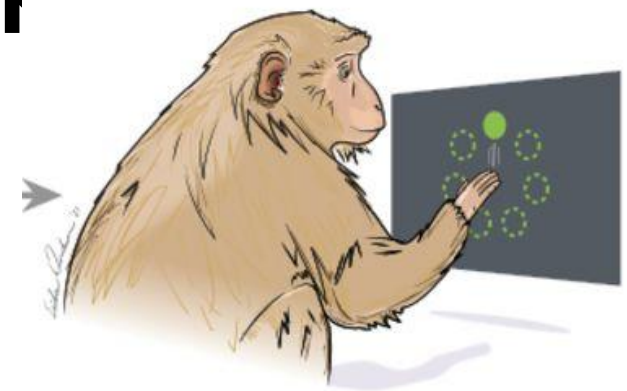
Hatsopoulos NG, Xu Q, Amit Y. Encoding of Movement Fragments in the Motor Cortex. J Neurosci. 2007 May 9;27(19):5105–14.



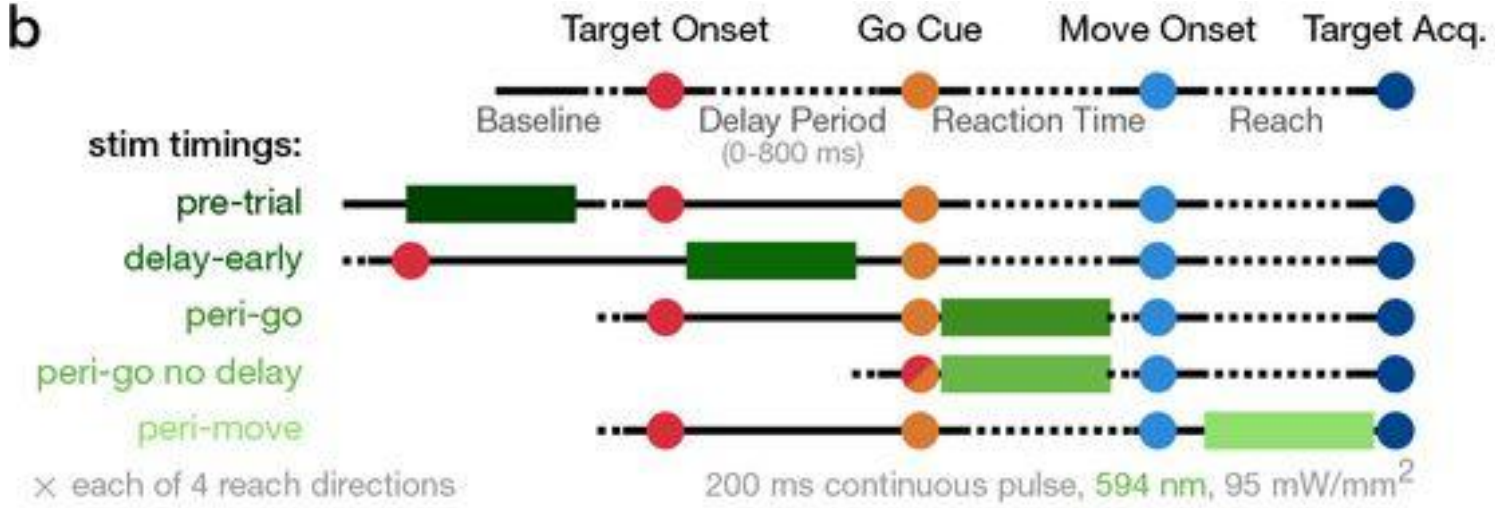
Tsianos GA, Goodner J, Loeb GE. Useful properties of spinal circuits for learning and performing planar reaches. J Neural Eng. 2014 Oct 1;11(5):056006.

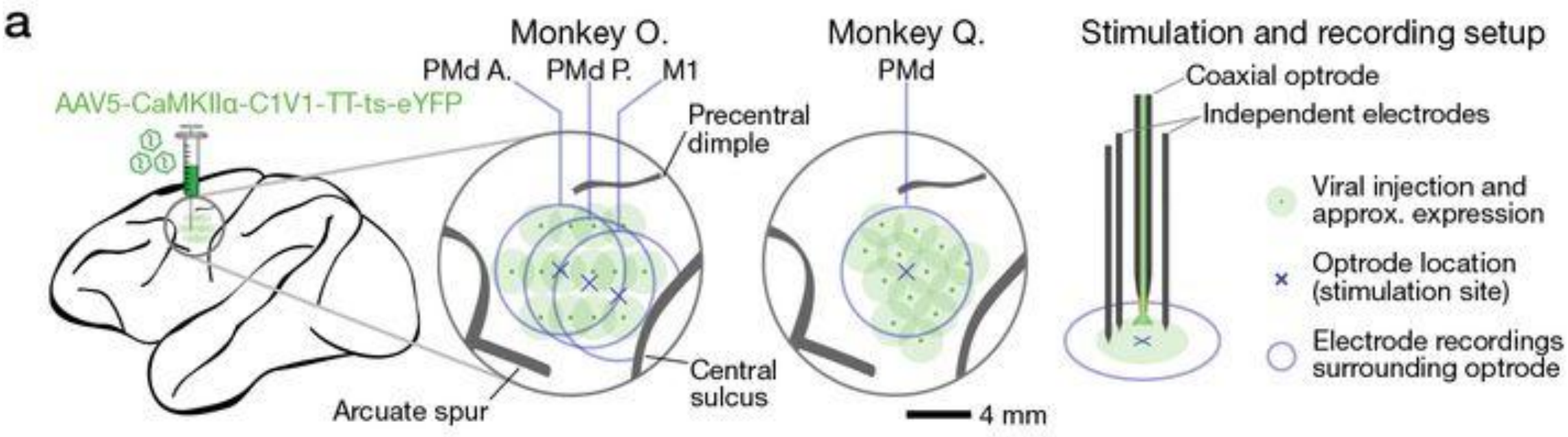
- Full extent of hypotheses justification :/
  - No references on expected effects of perturbations
  - But…model for H1 perturbation later on.

Experimental Setup (op



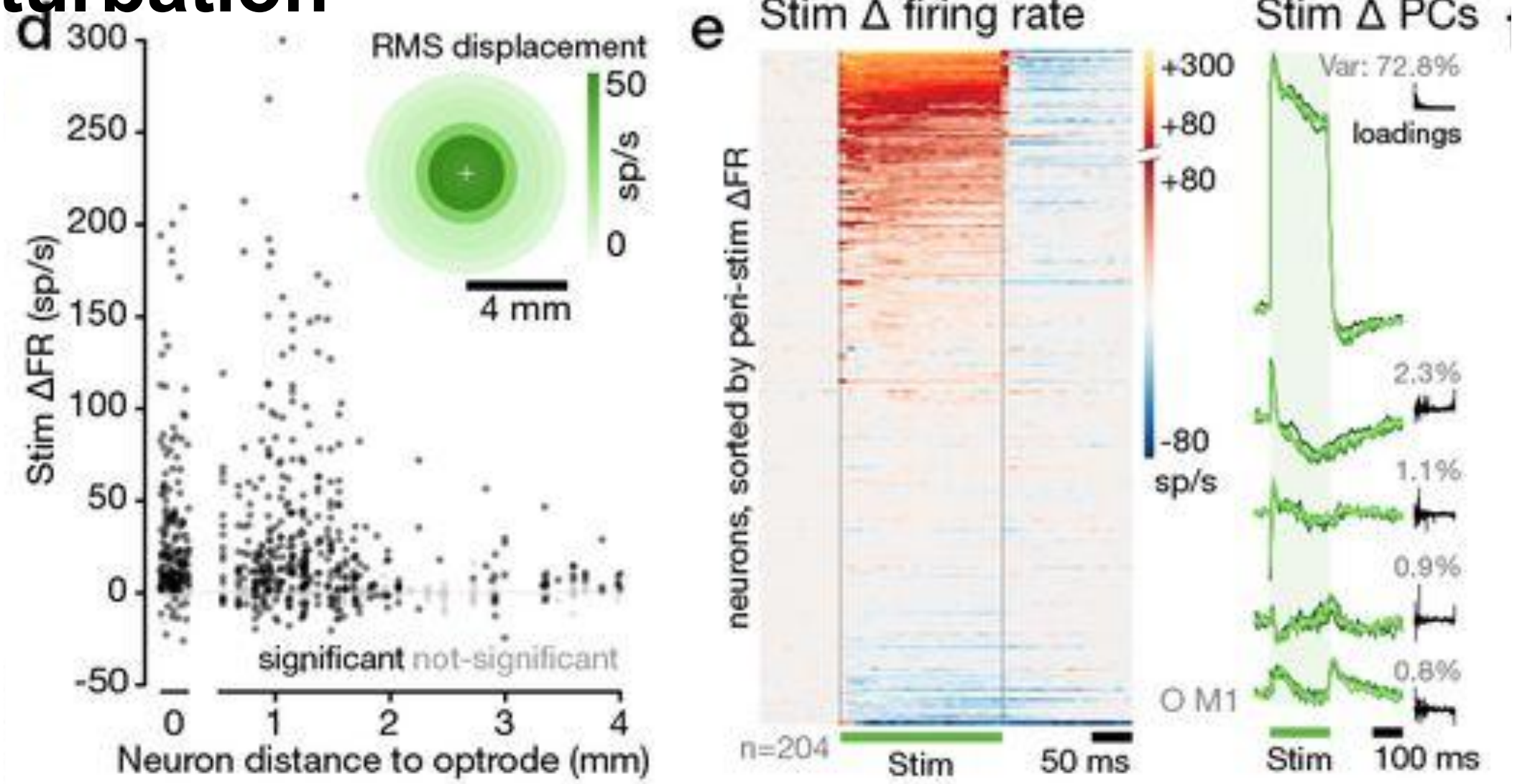
motor behavior





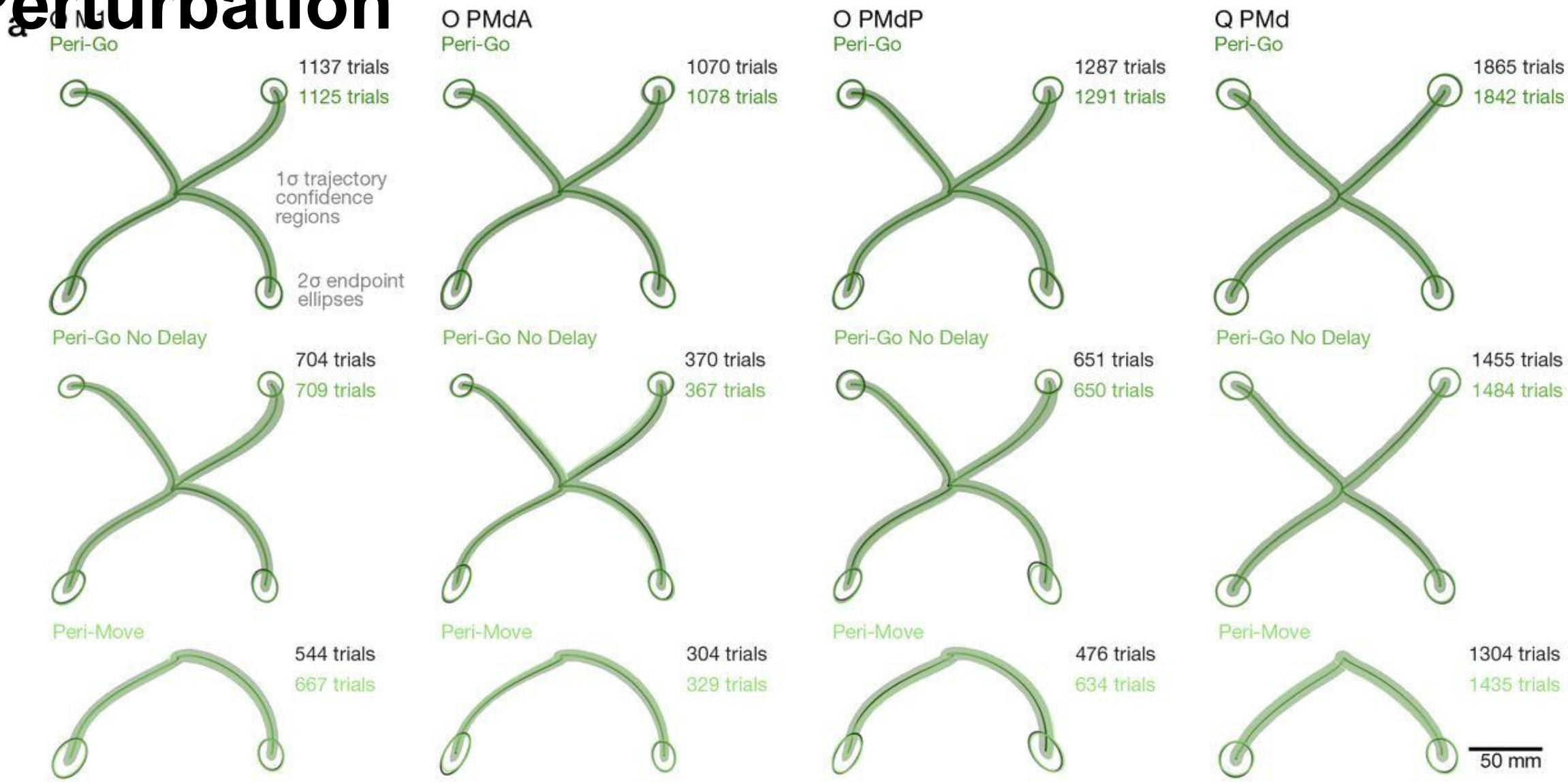
- Intro + Background
  - Hypotheses (total of 3)
  - Experimental Setup
- Results (knock out 1 hypothesis with each section):
  - Optogenetics
  - Intra-Cranial Microstimulation

Neural Effects Of Optogenetic Perturbation

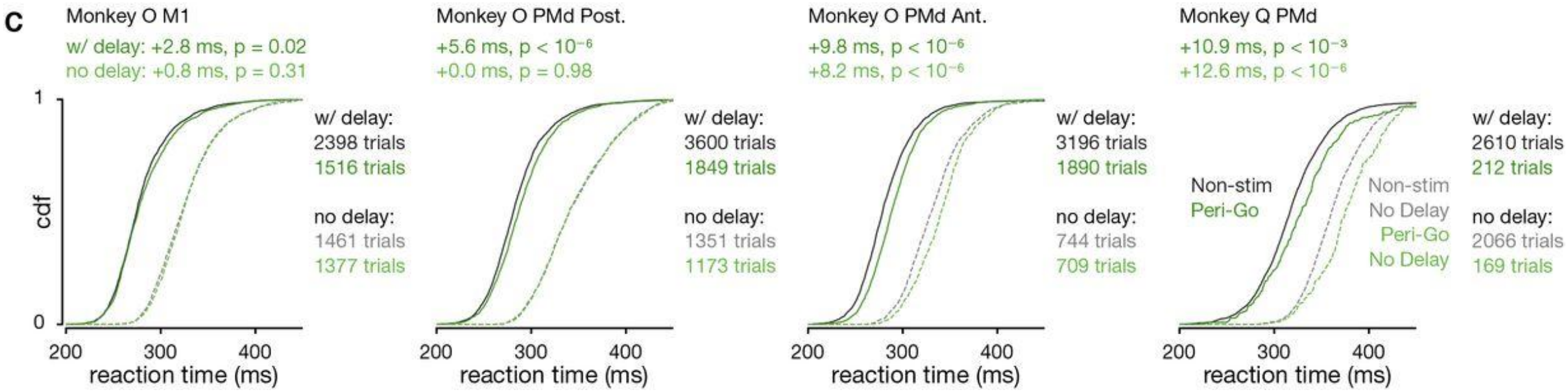


Neural Effects Of Optogenetic Perturbation

No Behavioral Effects Of Optogenetic Perturbation

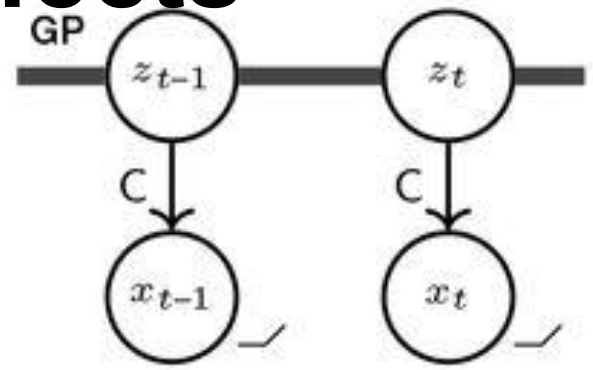


~~No~~ Behavioral Effects Of Optogenetic Perturbation



Inferring Perturbation Effects

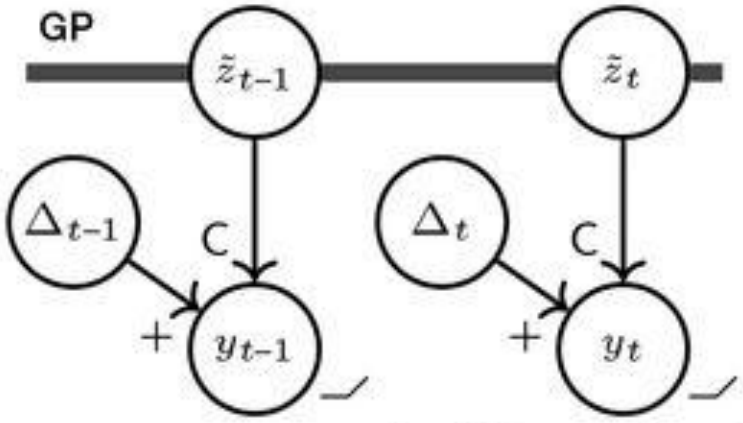
a Non-stimulation neural data:



non-stim latents
$z_t \in \mathbb{R}^K$, with GP prior

non-stim rates
$x_t = g(Cz_t + d) \in \mathbb{R}^D$

Stimulation neural data:



stim latents
$\tilde{z}_t \in \mathbb{R}^K$, with GP prior

additive component
$\Delta \in \mathbb{R}^{D \times T}$, low-rank

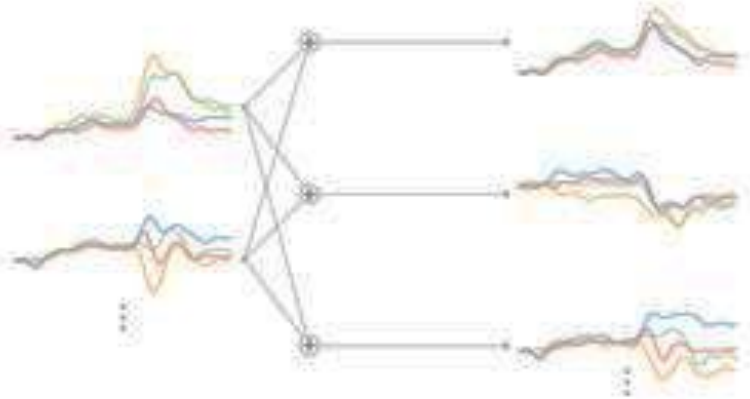
stim rates
$y_t = g\left(C\tilde{z}_t + \Delta_t + d\right)$

w/ rectifying non-linearity to correctly handle silencing

g Latent Variable Model

non-stim latents
latent dimensions

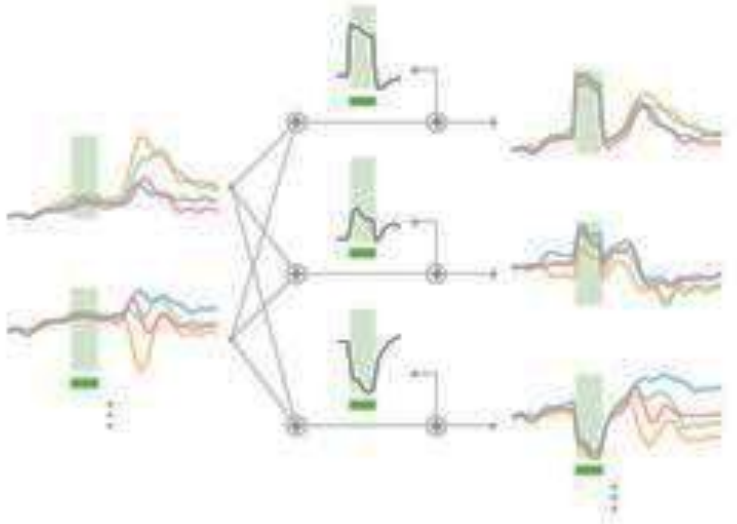
non-stim firing rates
neurons (PSTHs)



stim latents
independent of non-stim

stim deltas
condition-independent, low-rank neurons × time

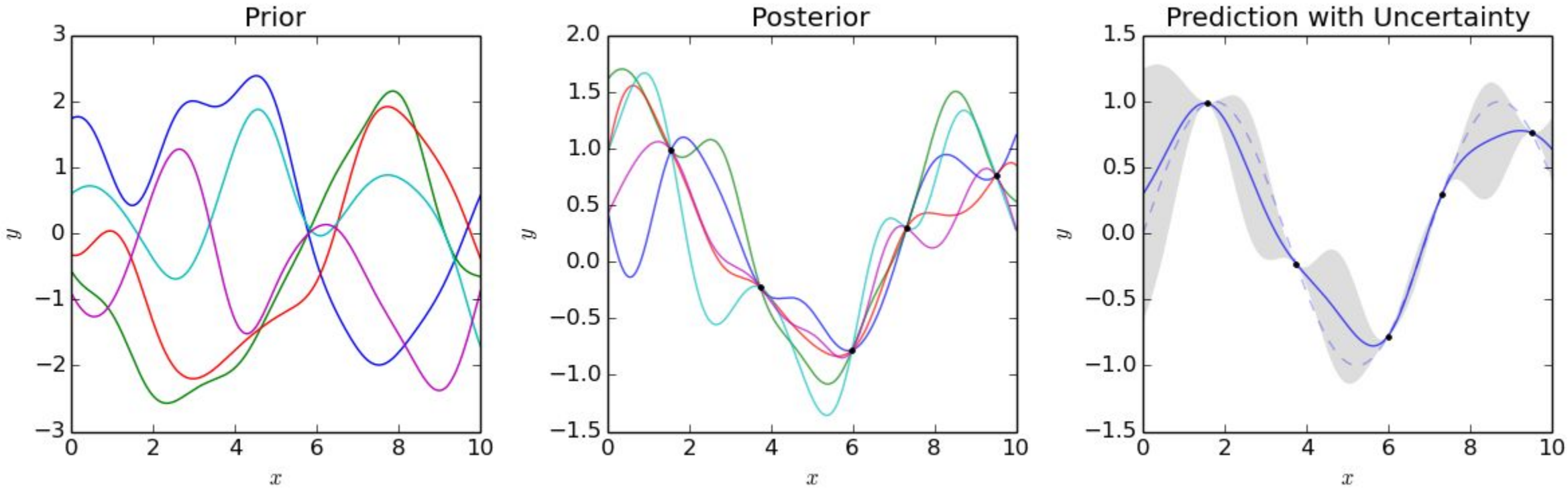
stim rates



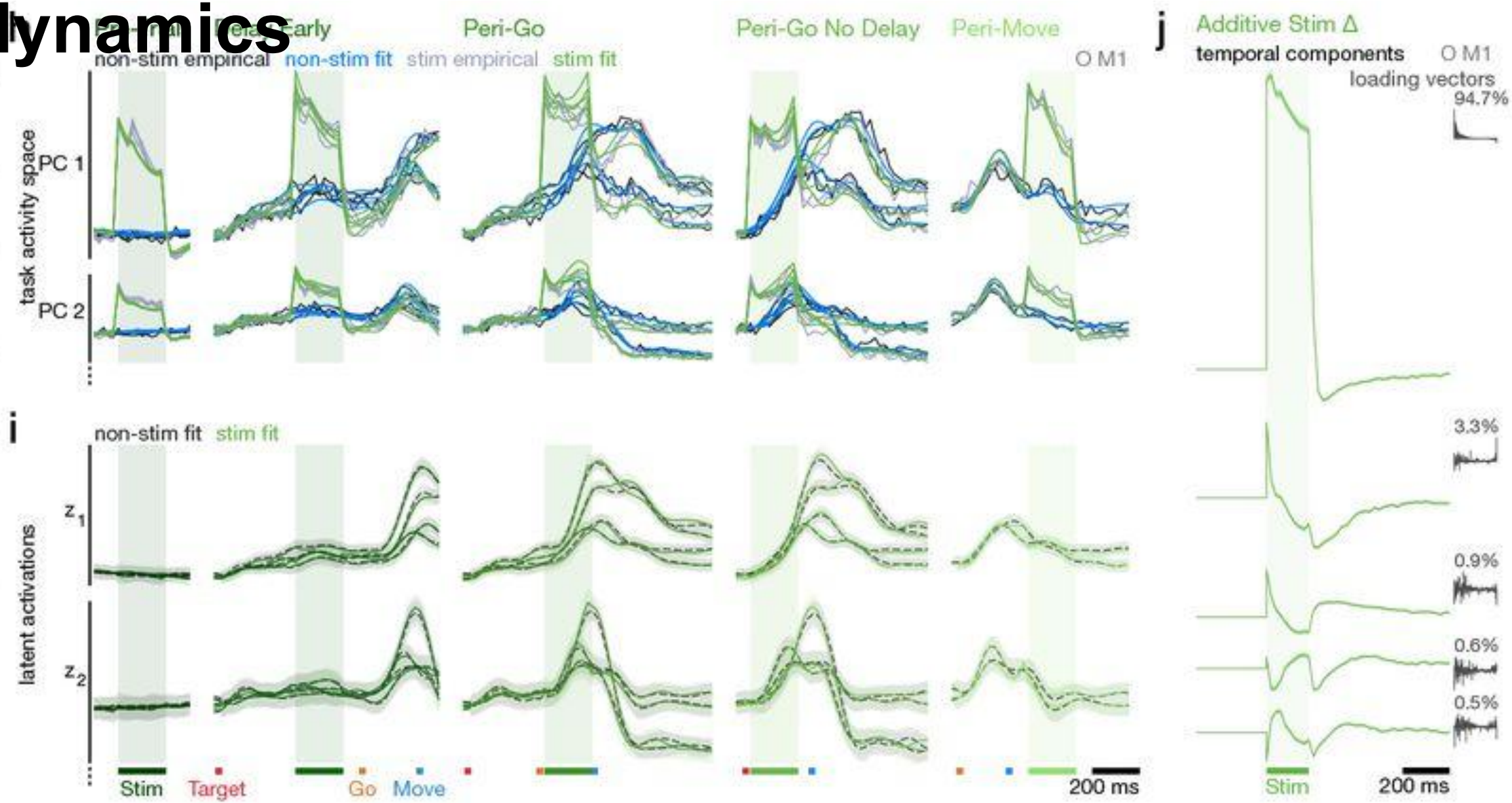
Gaussian process autocorrelation timescale = 100ms

Additive component rank = 5

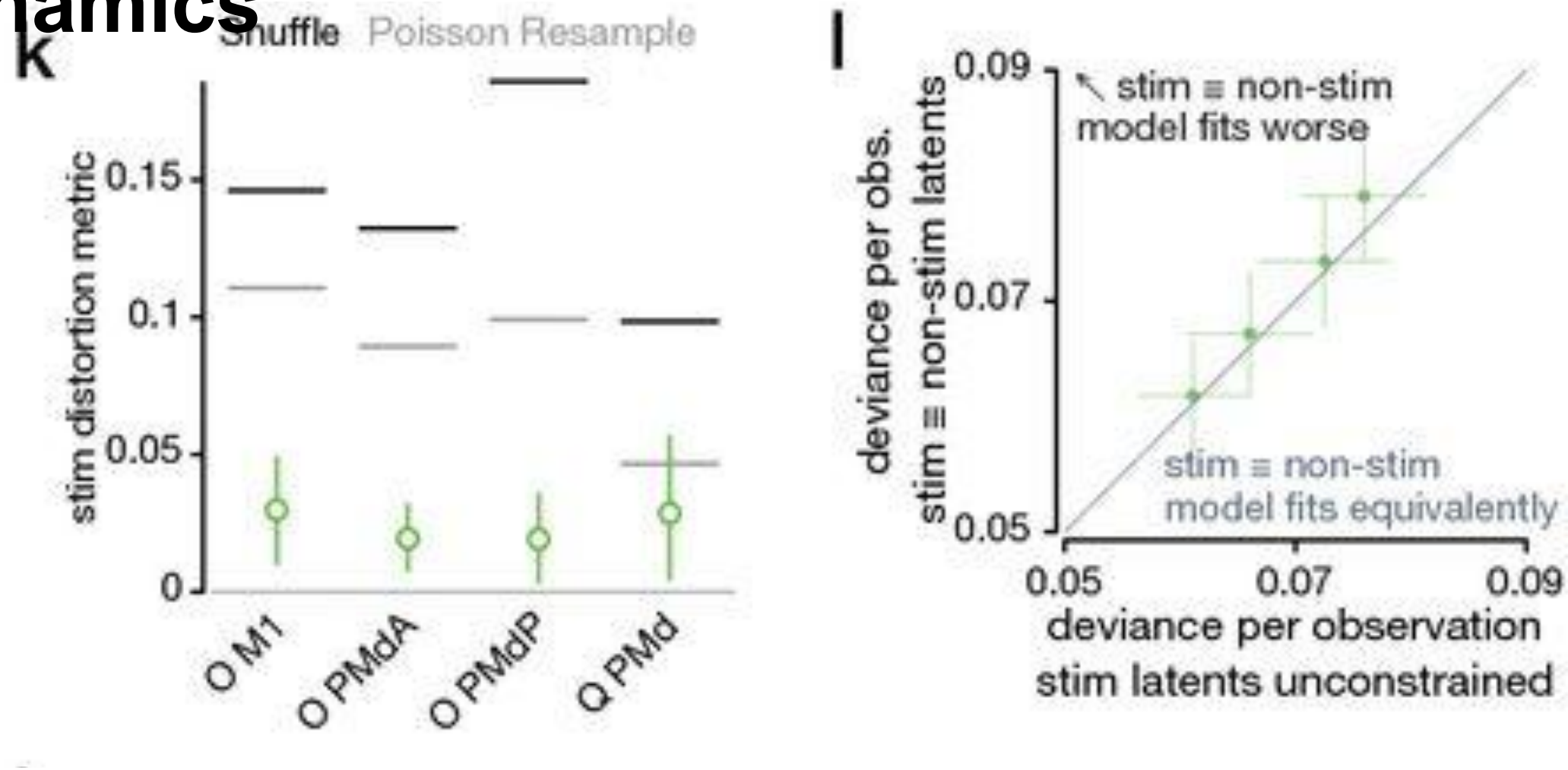
Gaussian what-now?



Perturbation doesn't affect underlying dynamics

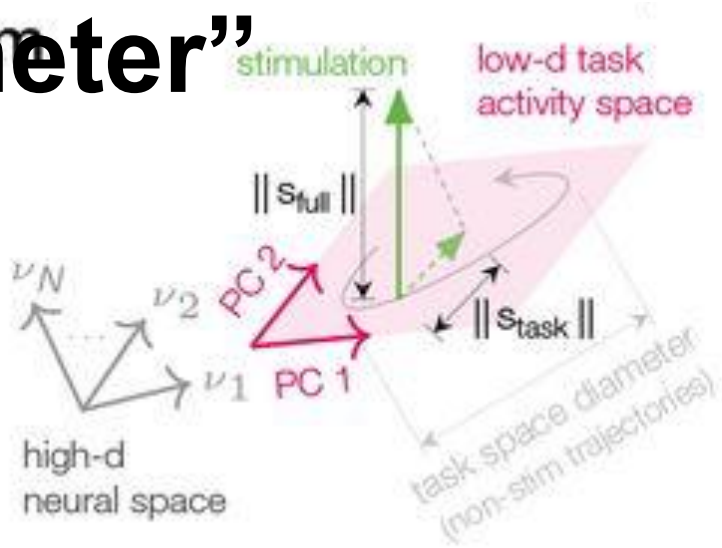


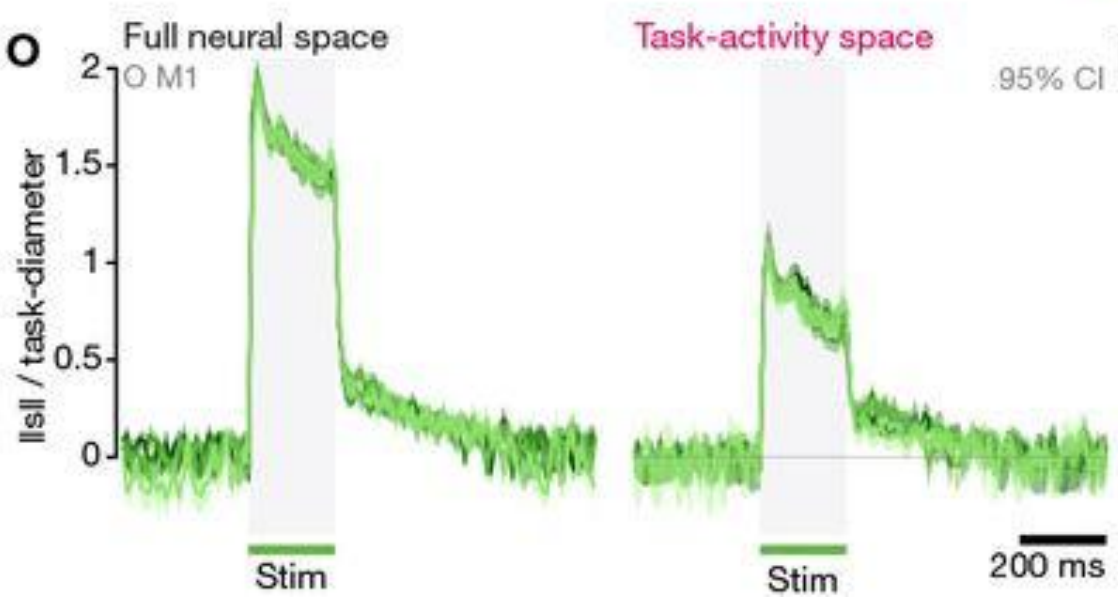
Perturbation doesn't affect underlying dynamics

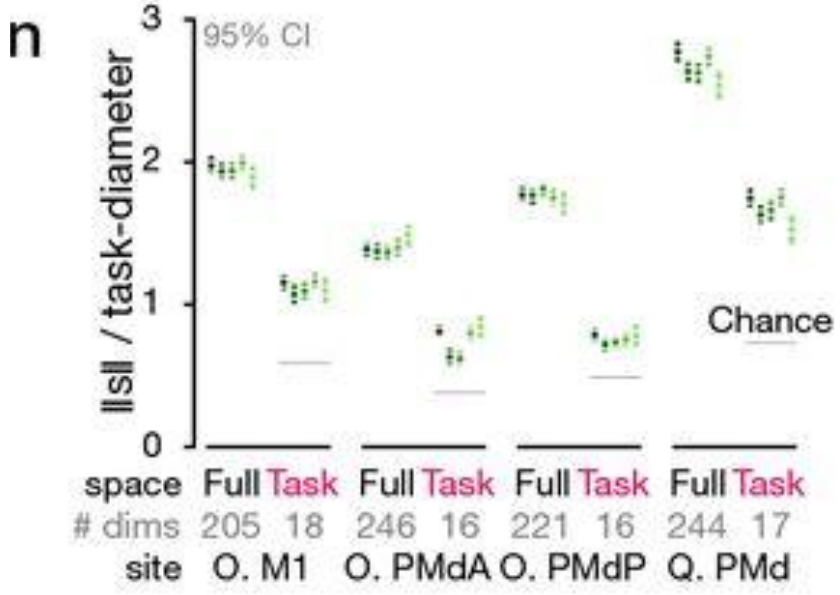


Stim-distortion: Euclidean distance after alignment using rotation, reflection, translation (no scaling)

Introducing "Task-Space Diameter"







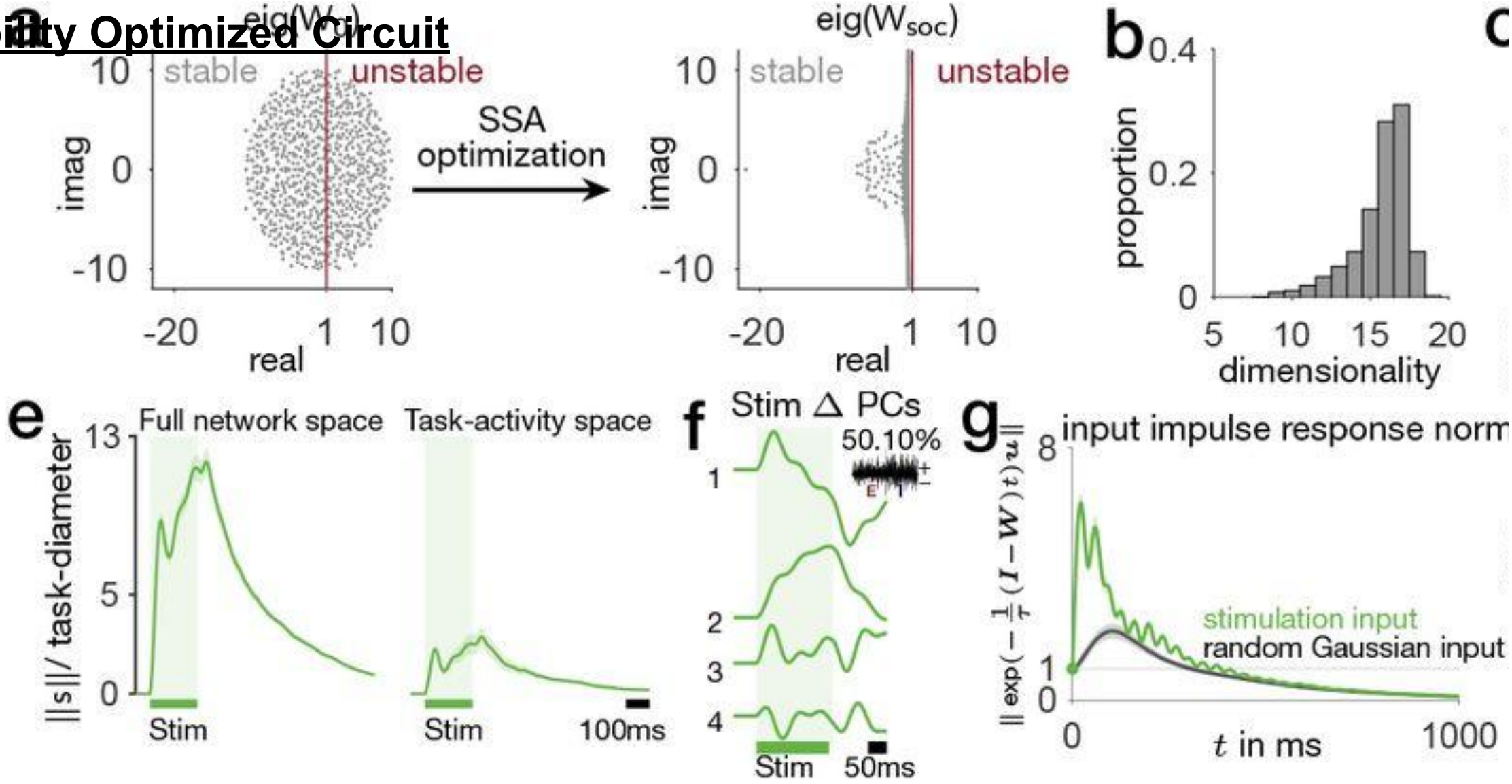
Magnitude of perturbation vector projected in "task subspace"

Task Activity Space: Dimensions with 95% of variance using principal components

Moving on to models...

Slow perturbation decay in a type of Reservoir Network

Stability Optimized Circuit

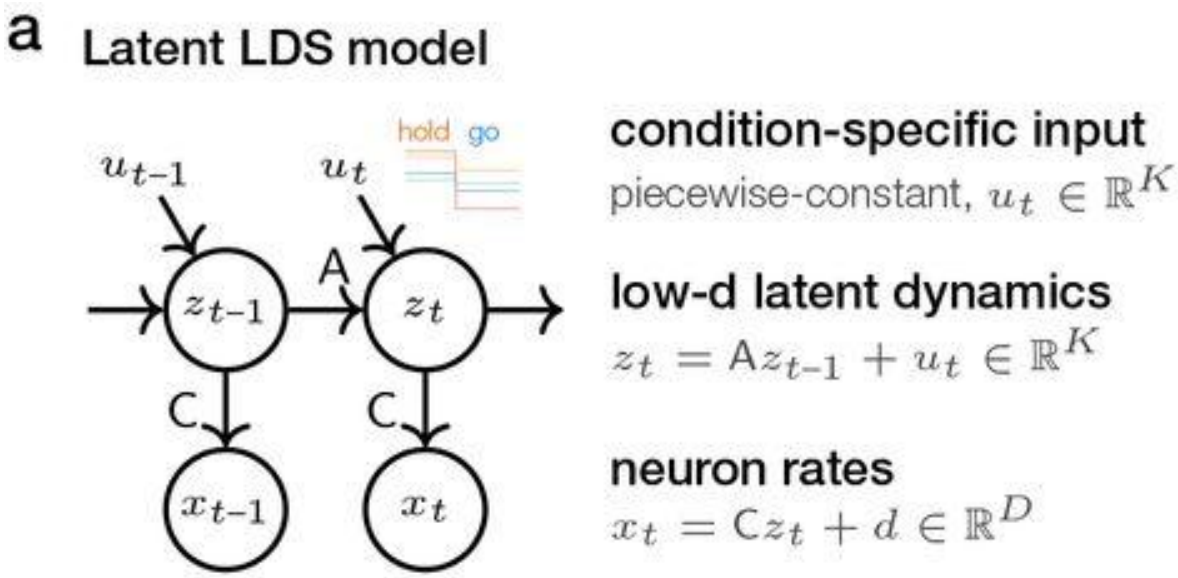


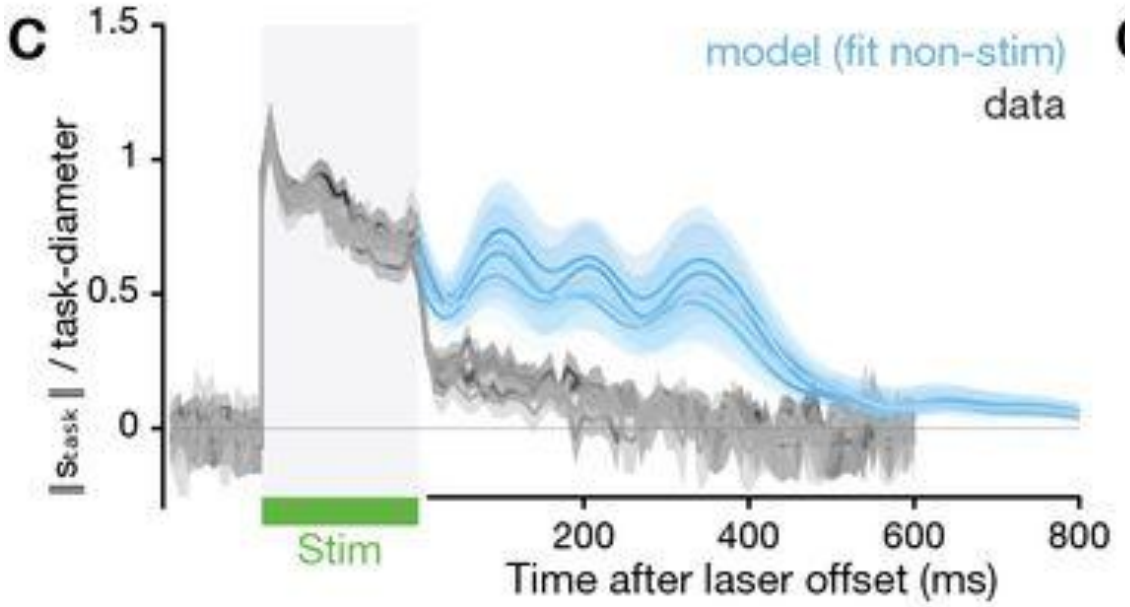
Done with H1?

- These predicted responses were qualitatively different from the observed neural responses arguing against the presence of high-dimensional reservoir dynamics in the motor cortex (**H1**).
- BUT…
- Following [14], we set τ = 200 to match the time-scales observed in motor cortical activity patterns.

$$\tau \frac{d\boldsymbol{x}}{dt} = -\boldsymbol{x}(t) + \mathbf{W}\boldsymbol{x}(t) + \boldsymbol{I}(t)$$

Modelling perturbation of low-D dynamics





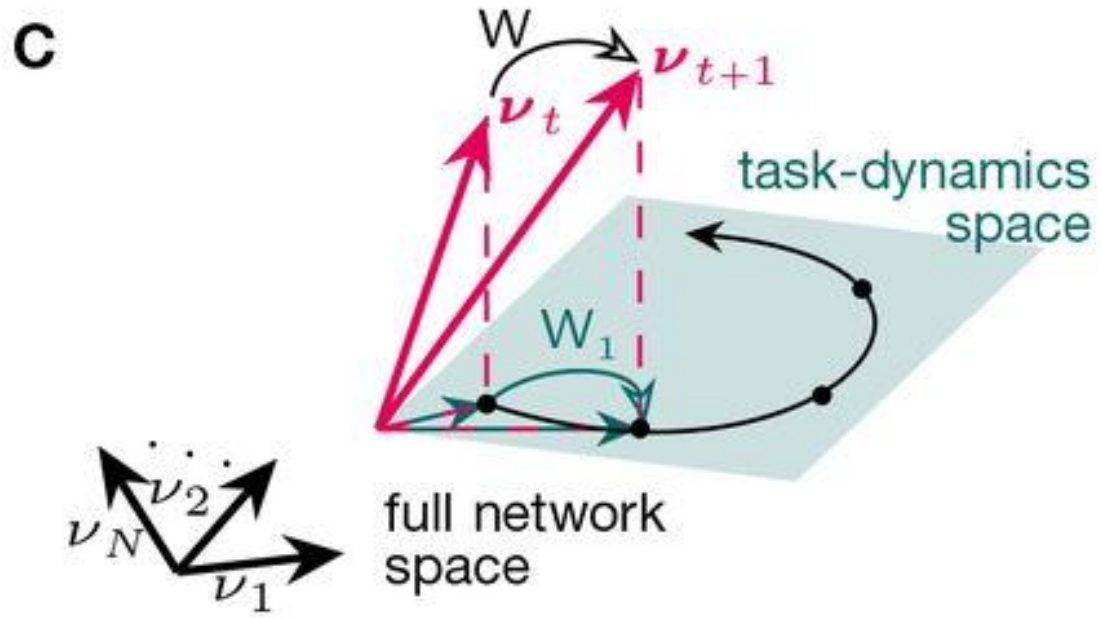
Rationalizing LDS perturbation results

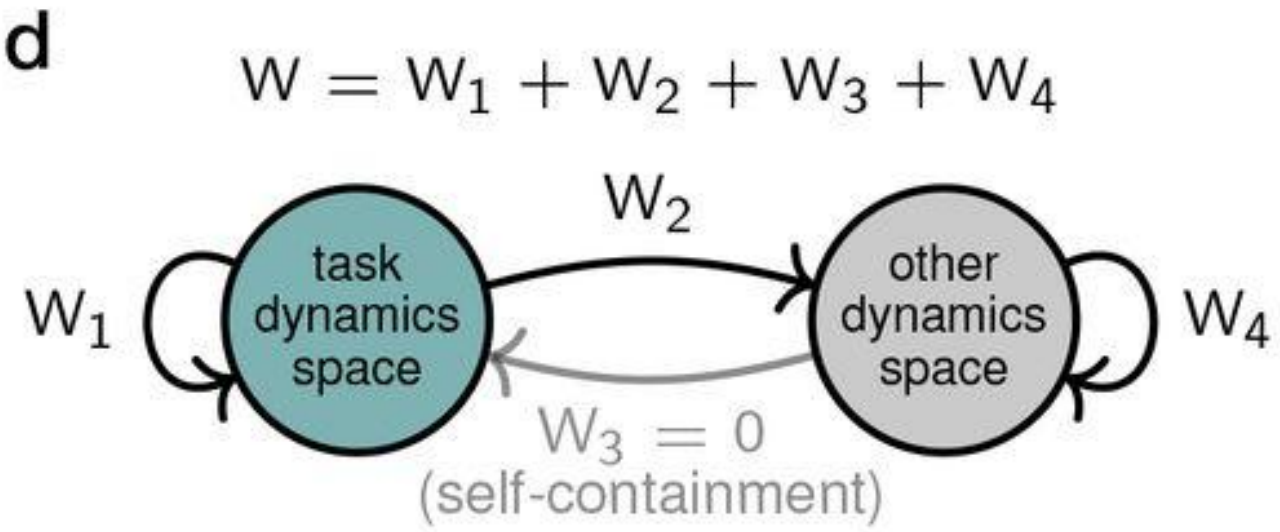
- Task activity space = Task dynamics space?
- Might not be perturbing the “task dynamics space”
- Can’t selectively perturb LDS, therefore…
- Fit model where we know “task dynamics space” and perturb that model.

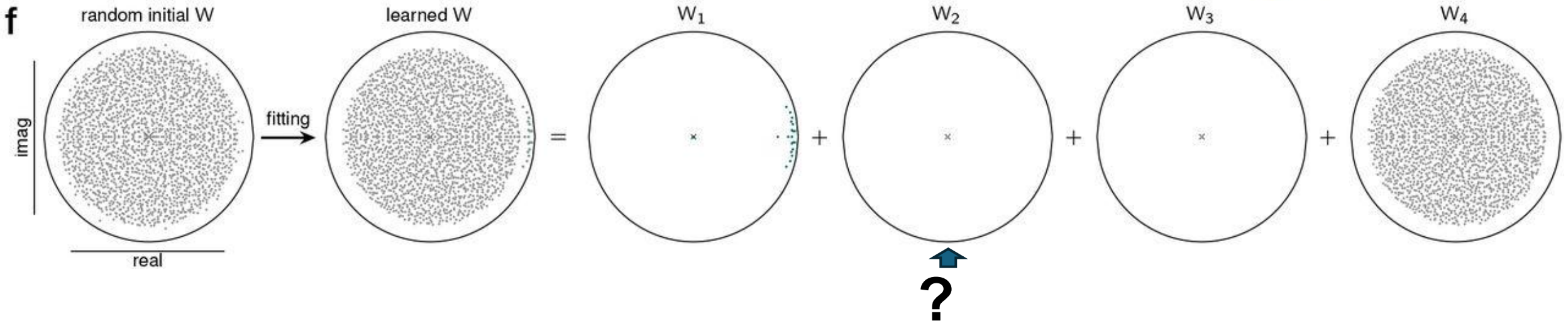
Mechanistic Explanation of perturbation results
More realistic model of subspace-structured dynamics

- Connectivity satisfies: Dale's law[53–55] and E/I balance.[53–55]
- high-dimensional (N=1500–2500 units), linear network
- low-dimensional subspace of size K=N/100 driving response
- We then optimized the E/I network connectivity matrix to maximize the likelihood that a linear readout from this task dynamics space matched the measured neural responses.
- **The E/I model was fit exclusively to data from trials without stimulation.**

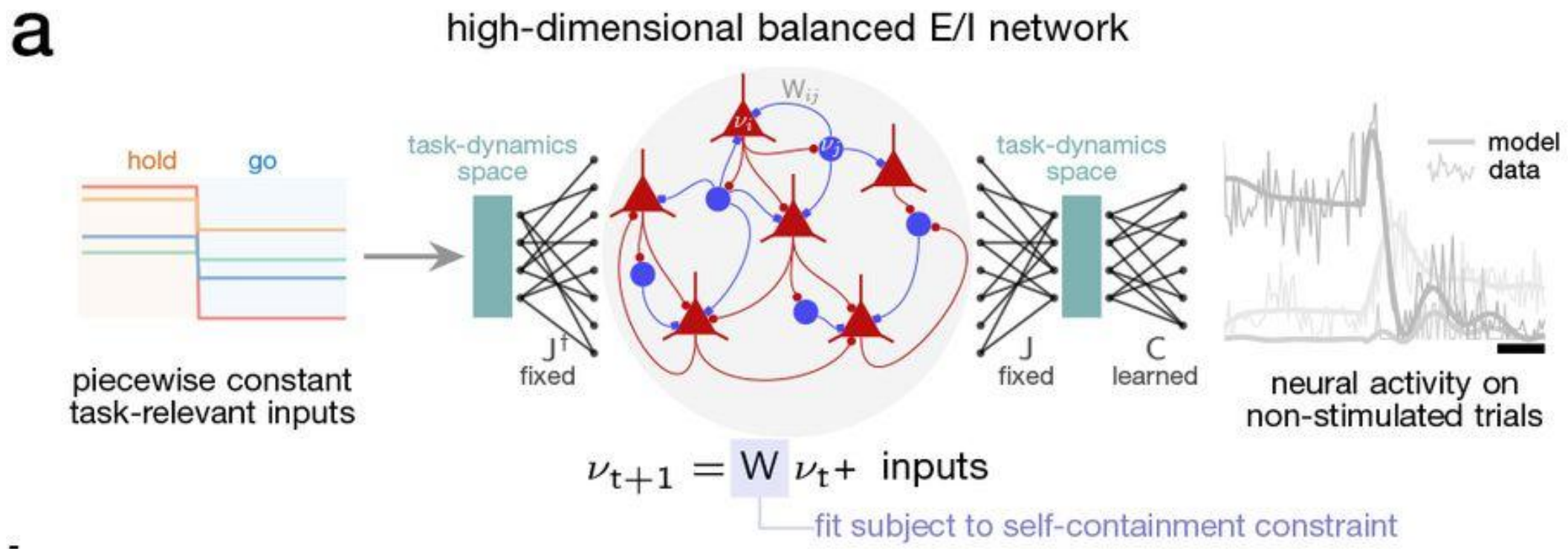
More realistic model of subspace-structured dynamics

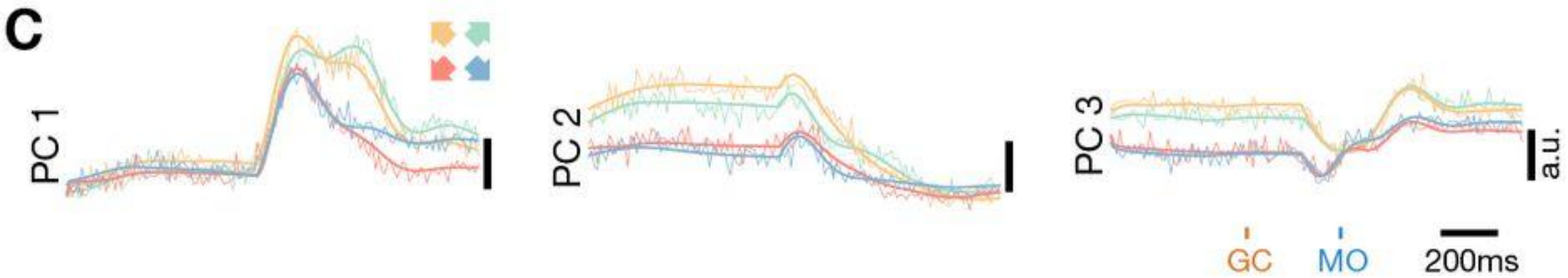




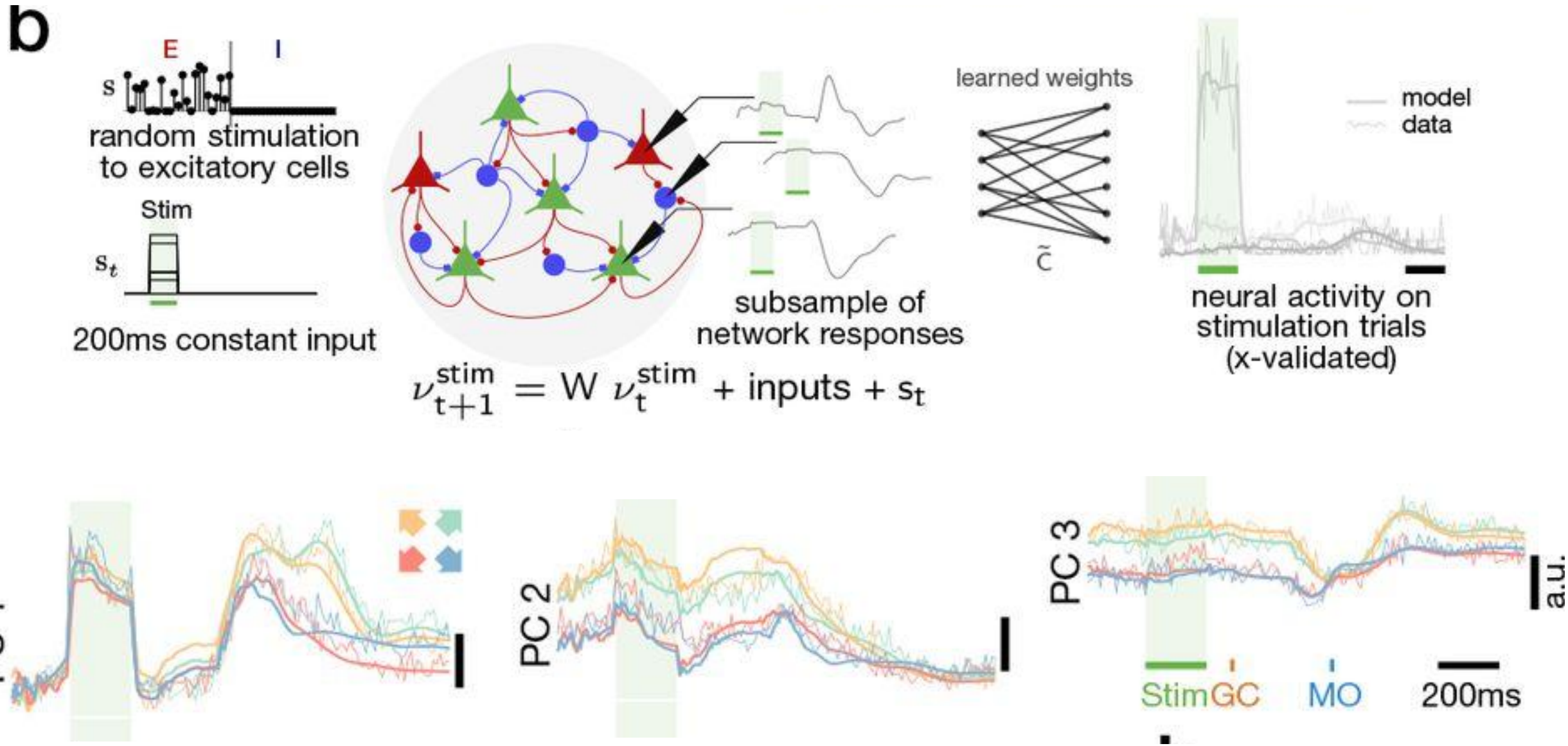


More realistic model of subspace-structured dynamics



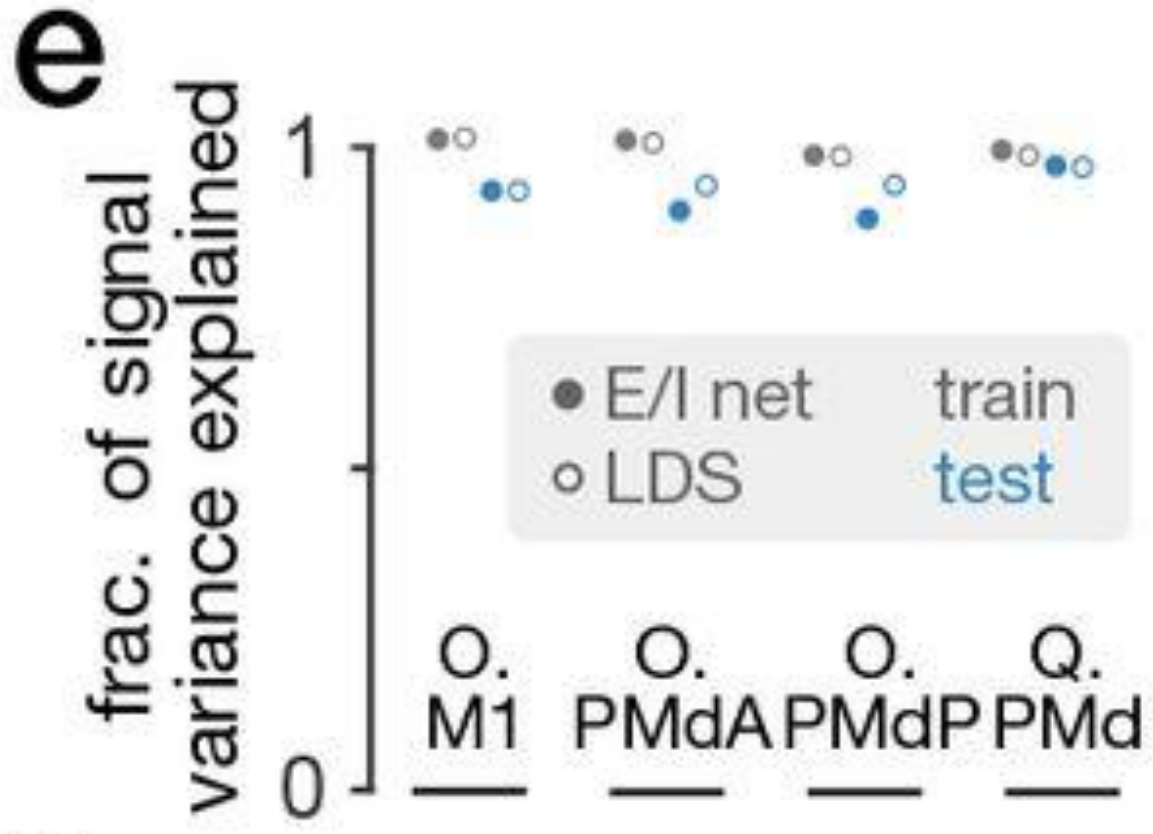


More realistic model of subspace-structured dynamics



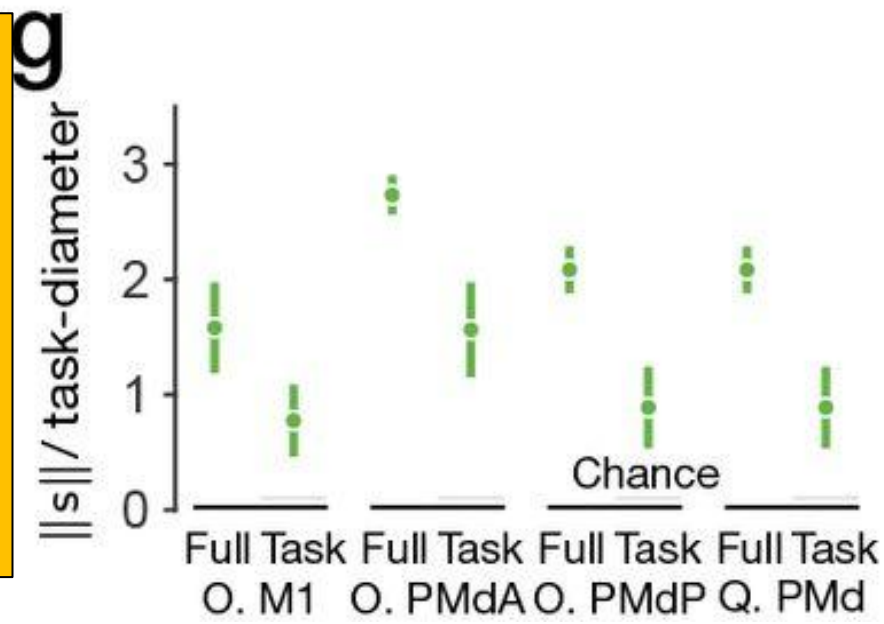
optogenetic stimulation was modeled with additive noisy, positive input patterns targeting random subsets of excitatory cells in the E/I network

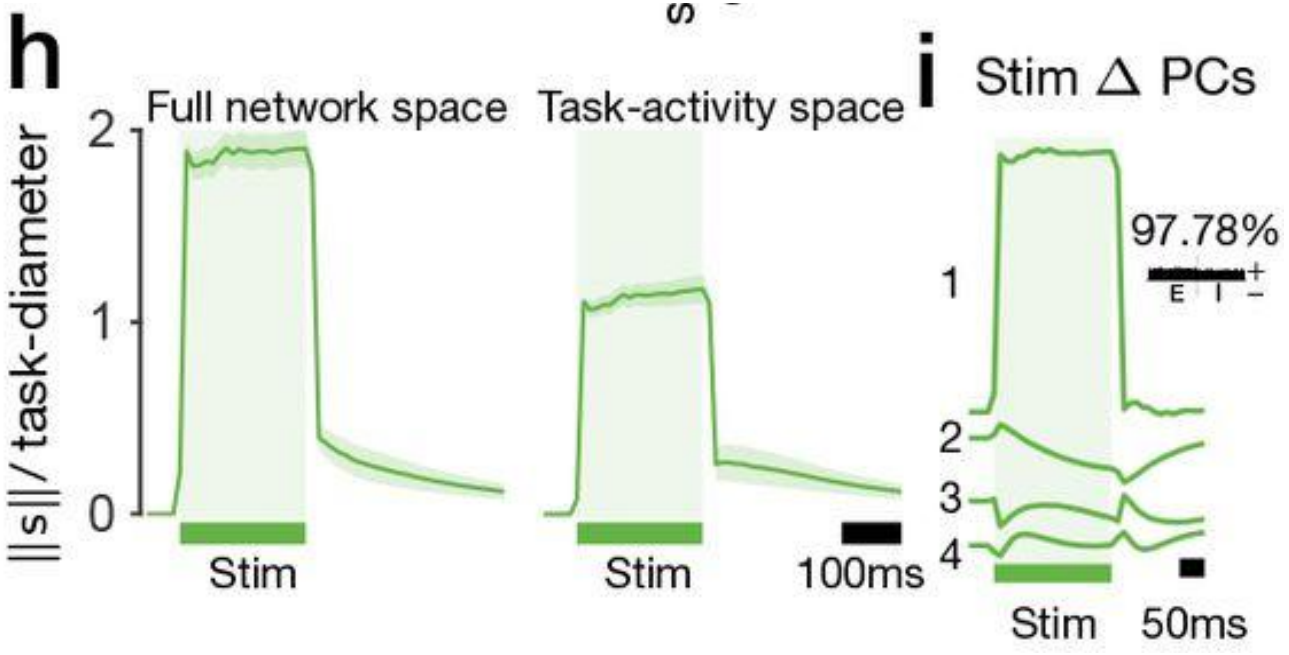
More realistic model of subspace-structured dynamics



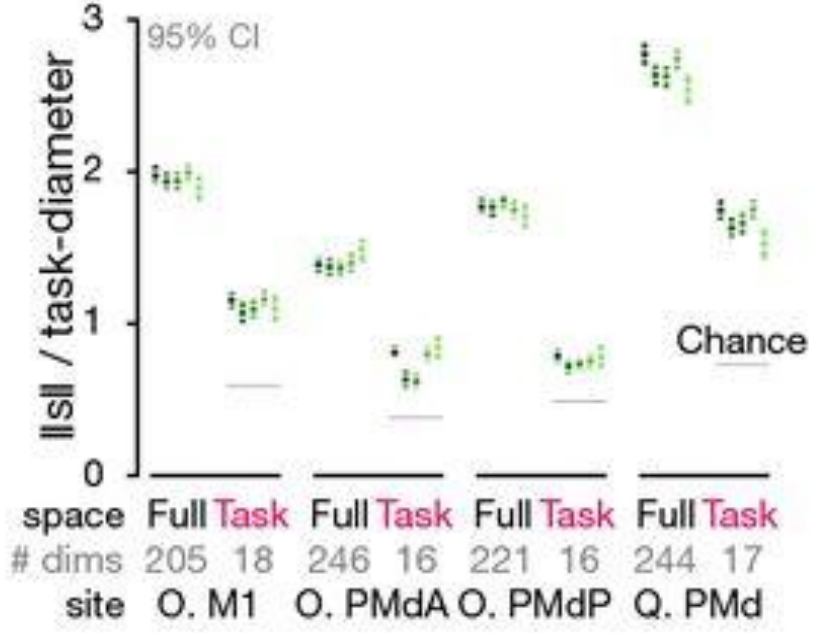
More realistic model of subspace-structured dynamics

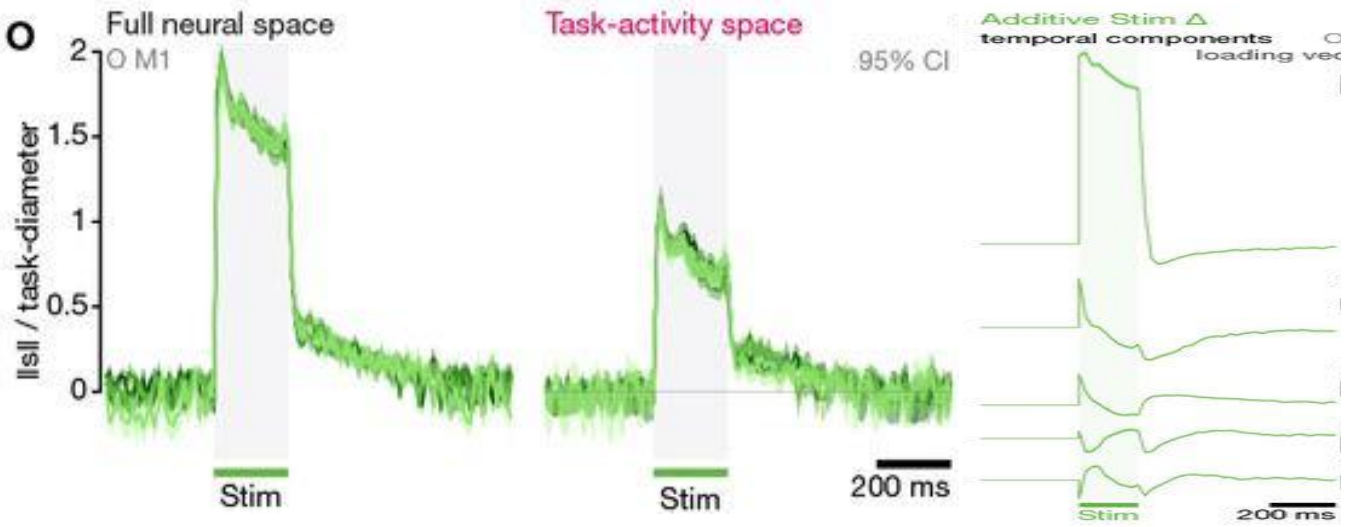
Model



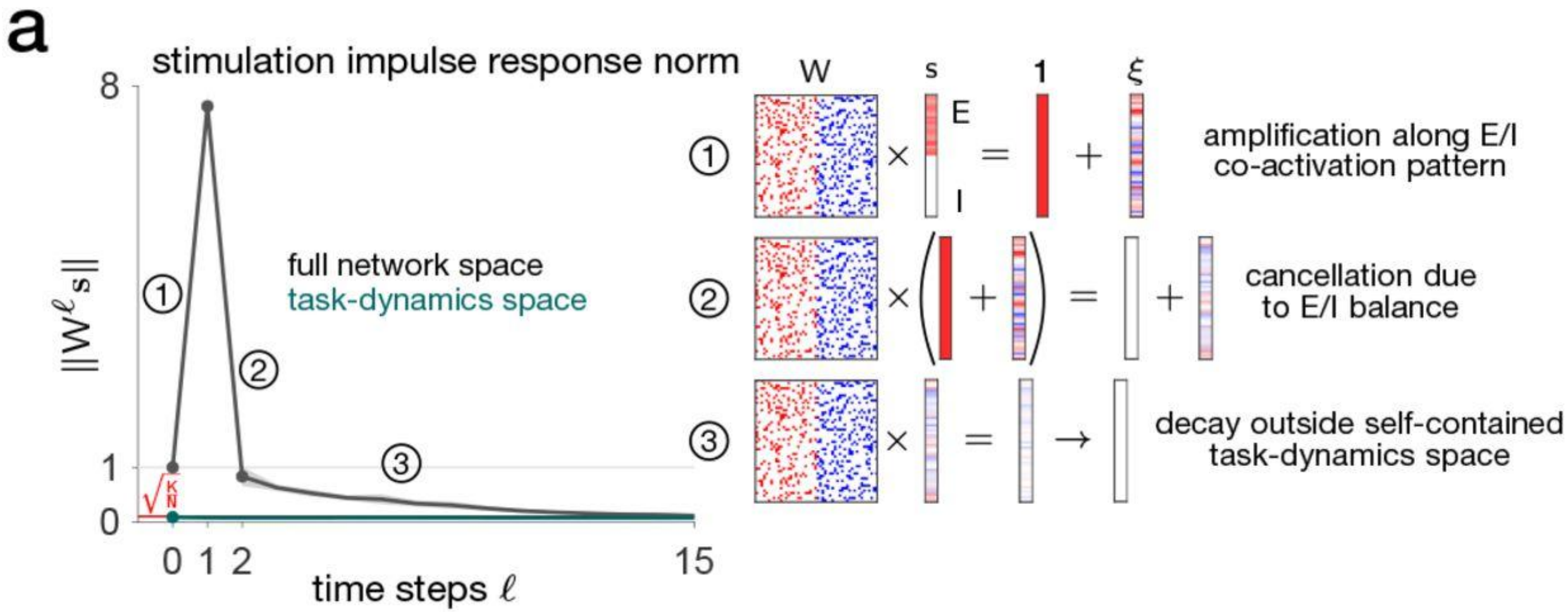


Data

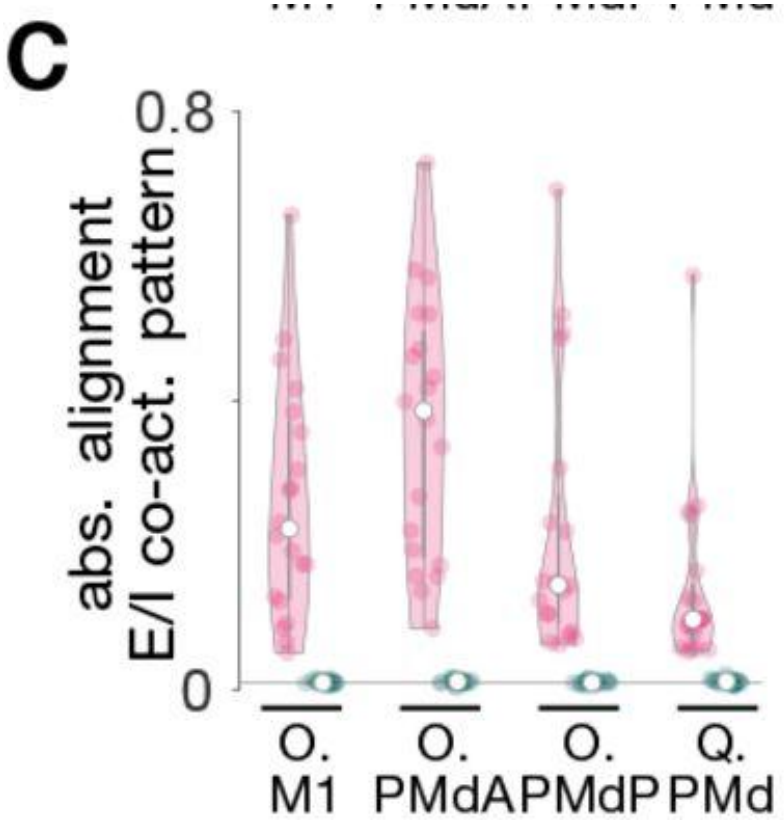


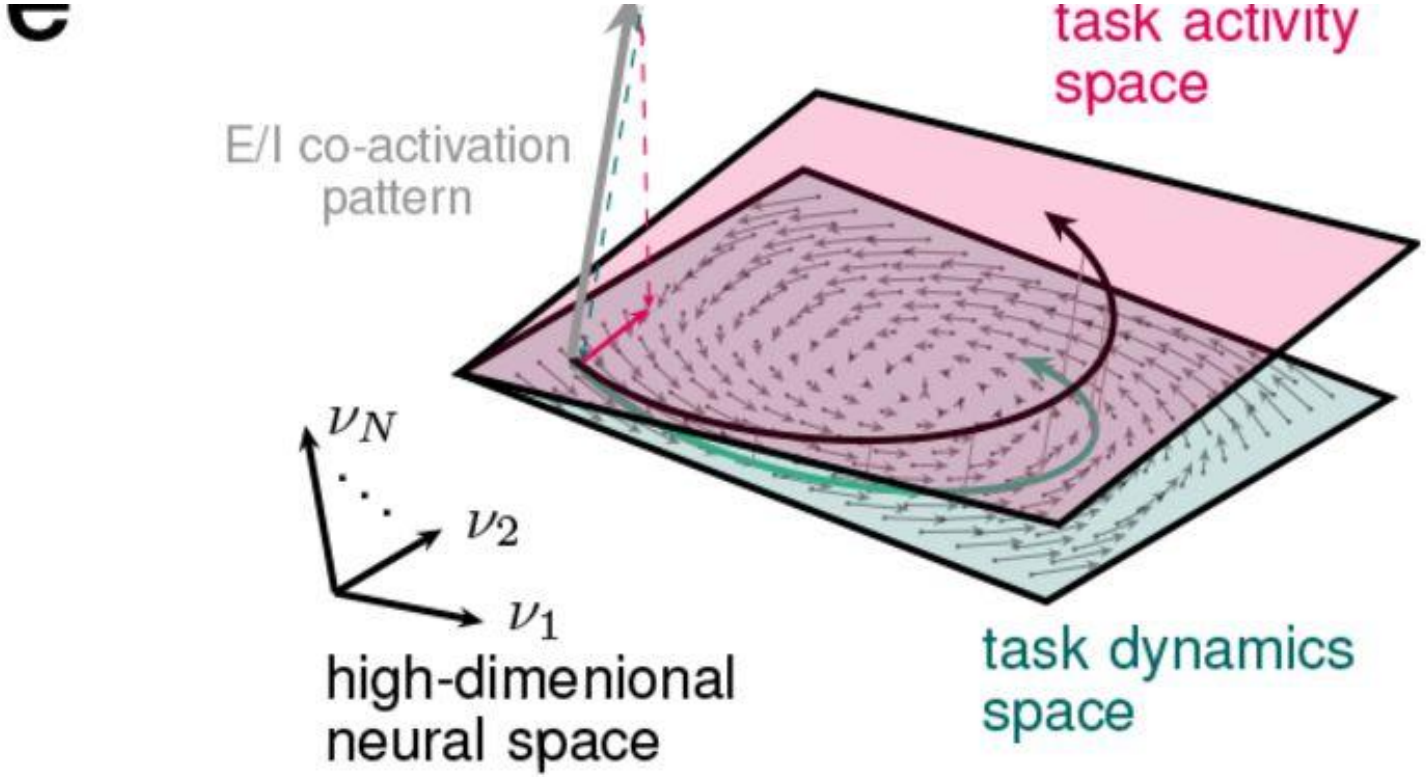


Mechanistic explanation for perturbation results



Mechanistic explanation for perturbation results





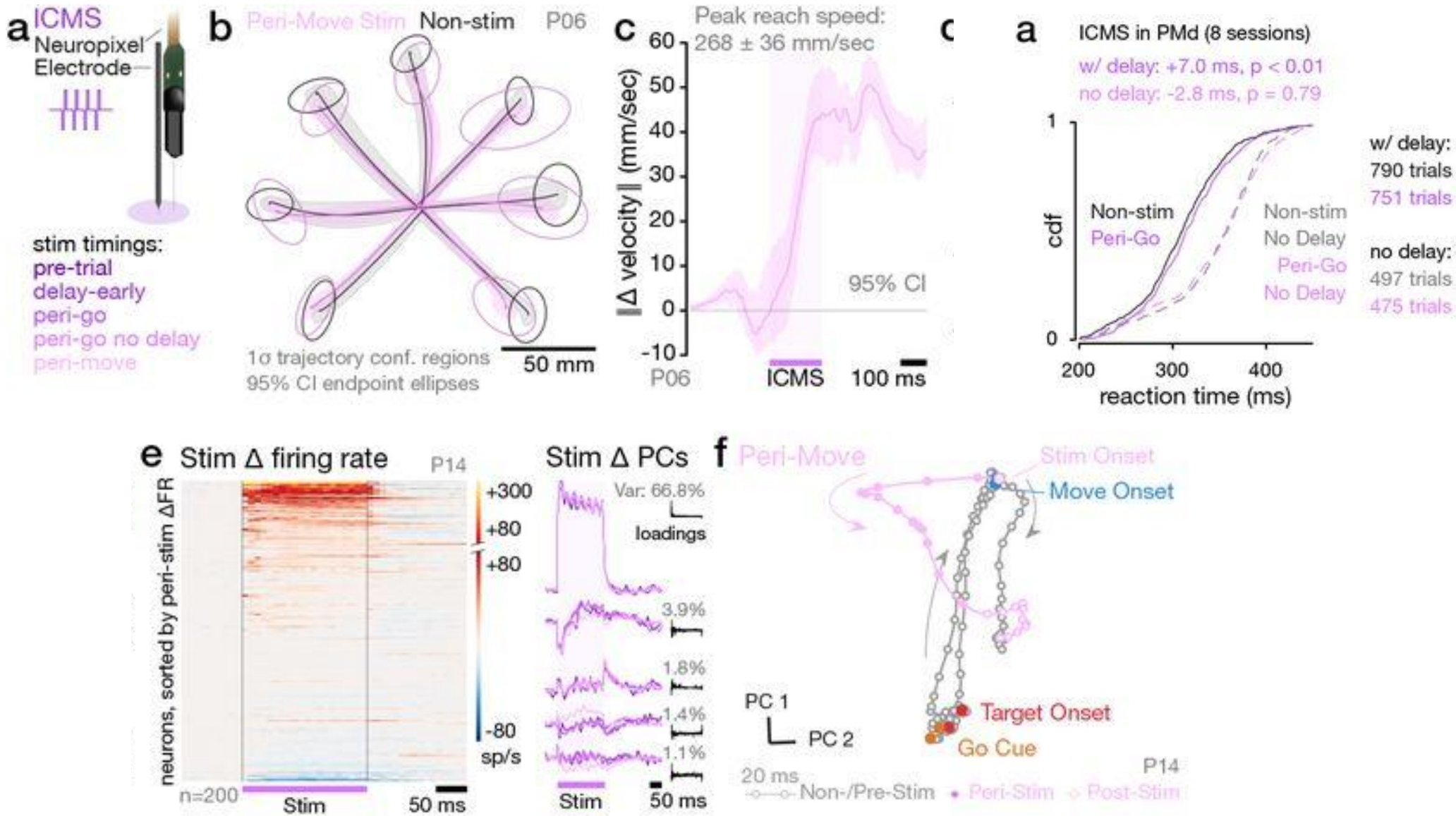
Checkpoint

- Path-following dynamics (H3): all local perturbations should decay rapidly back towards the externally imposed trajectory
- Subspace-structured dynamics (H2) : If perturbation doesn't reach task-dynamics subspace…perturbations should also decay rapidly.
- Therefore, **H2** and **H3** cannot be distinguished based on the results from these optogenetic perturbations alone.

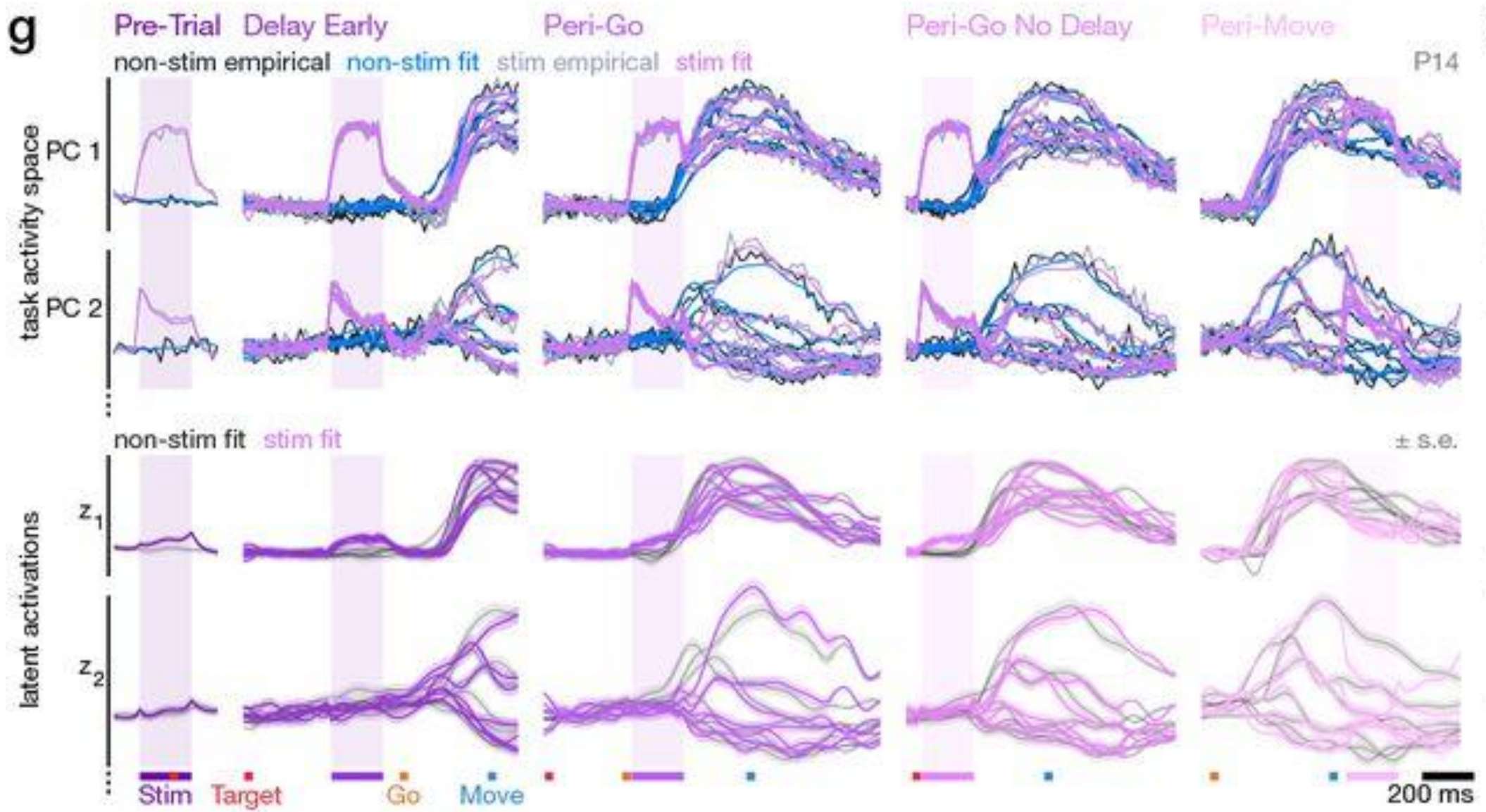
- Intro + Background
  - Hypotheses (total of 3)
  - Experimental Setup
- Results (knock out 1 hypothesis with each section):
  - Optogenetics
  - Intra-Cranial Microstimulation

- We reasoned that the ability to influence task-relevant behavioral outcomes might indicate that an experimental intervention is able to induce perturbations within the task dynamics space.
- Intracortical electrical microstimulation (ICMS), which is known to disrupt motor preparation[60] and to evoke movements readily when delivered to motor cortex.

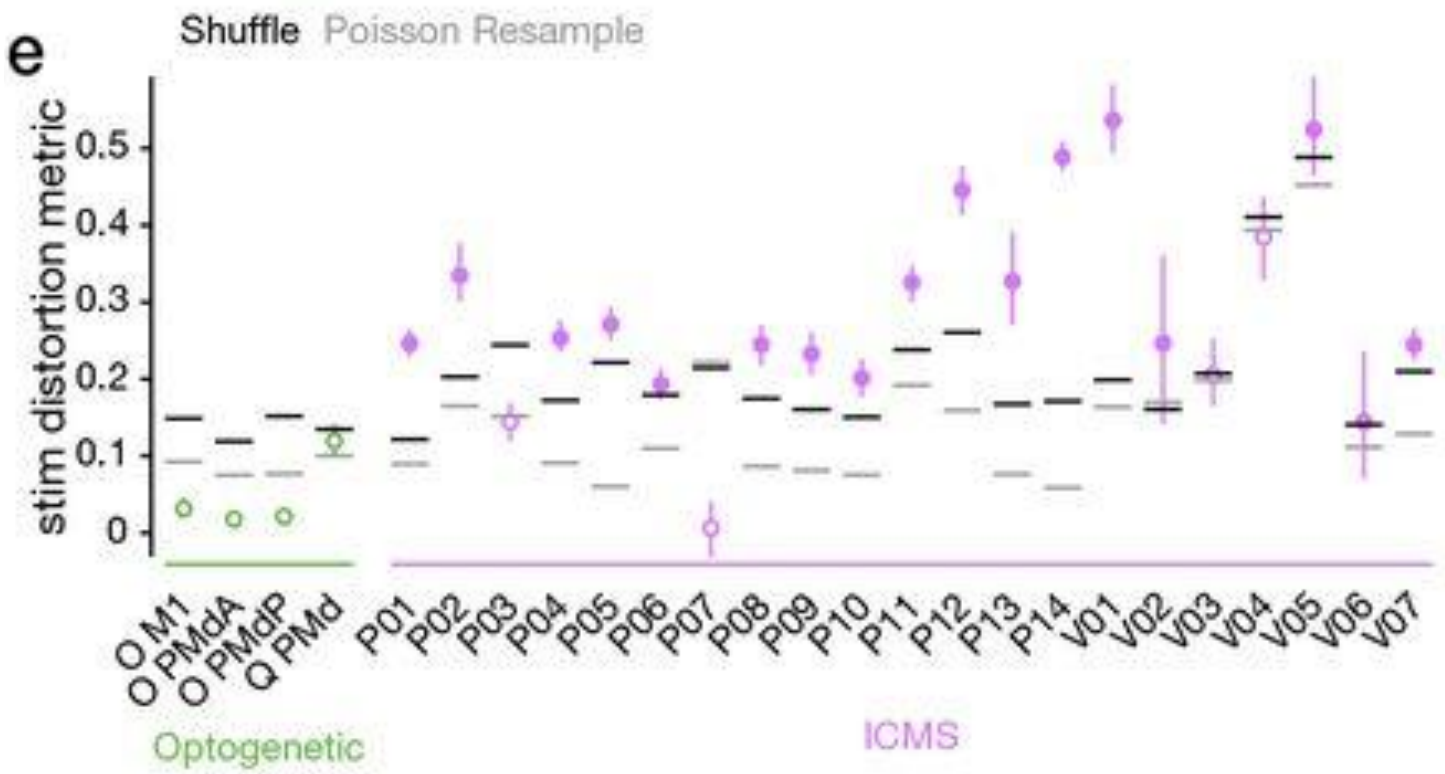
ICMS Effects

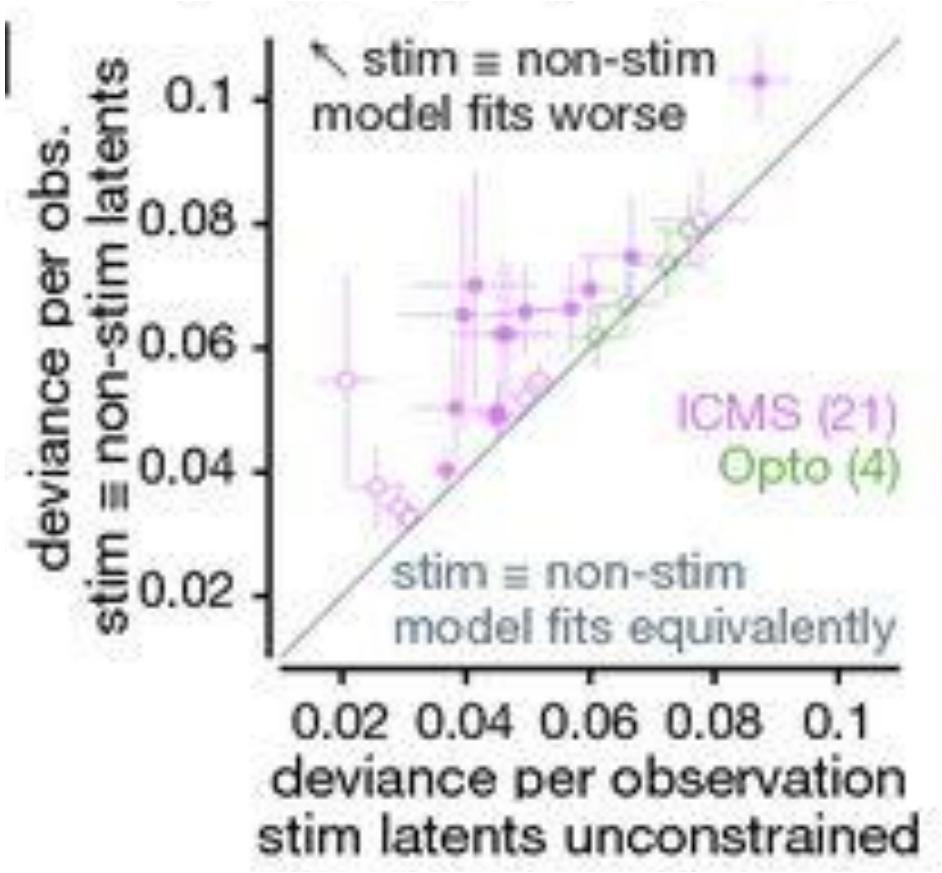


ICMS Effects

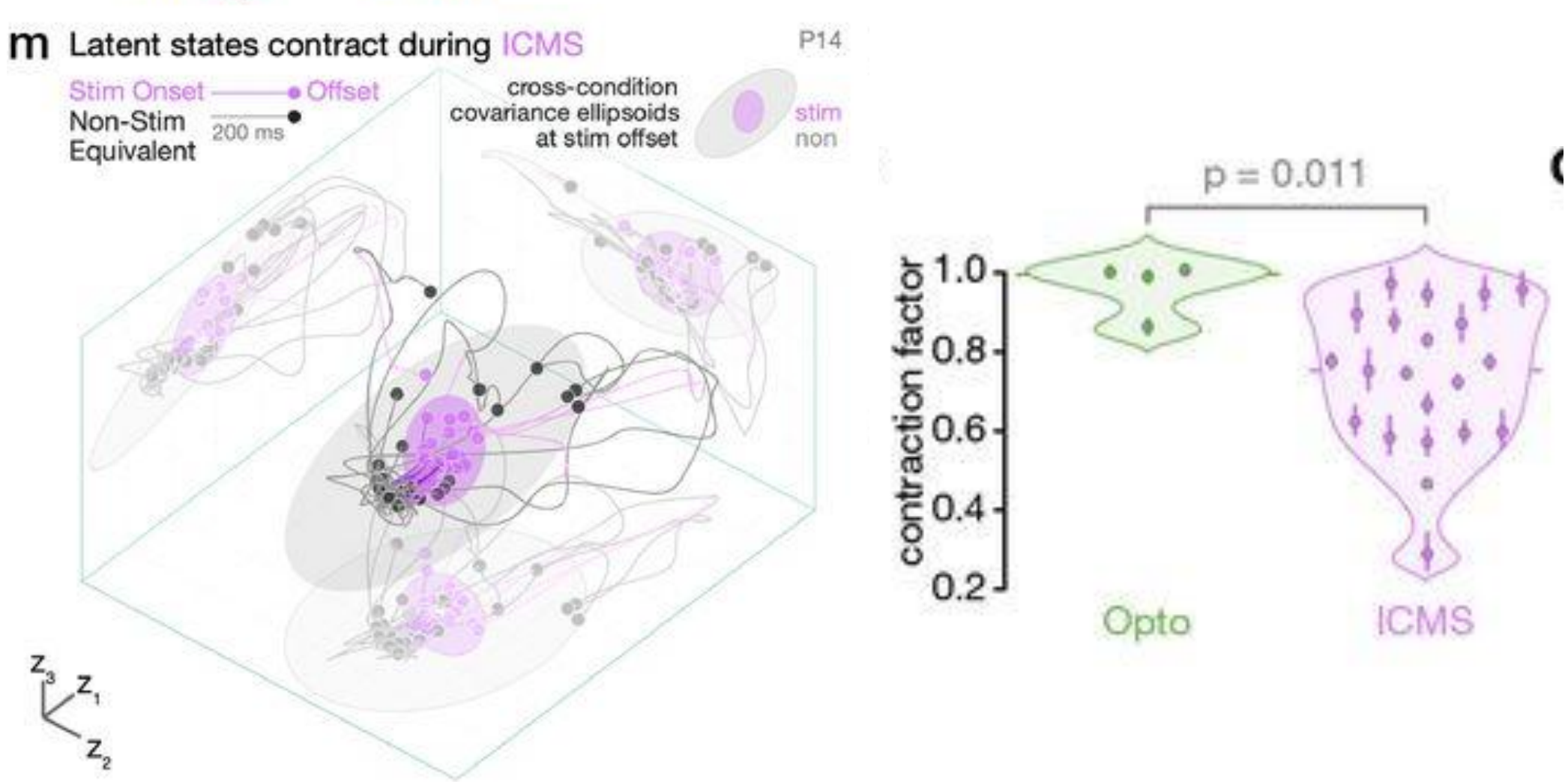


ICMS has stronger (actually significant) effect on neural dynamics





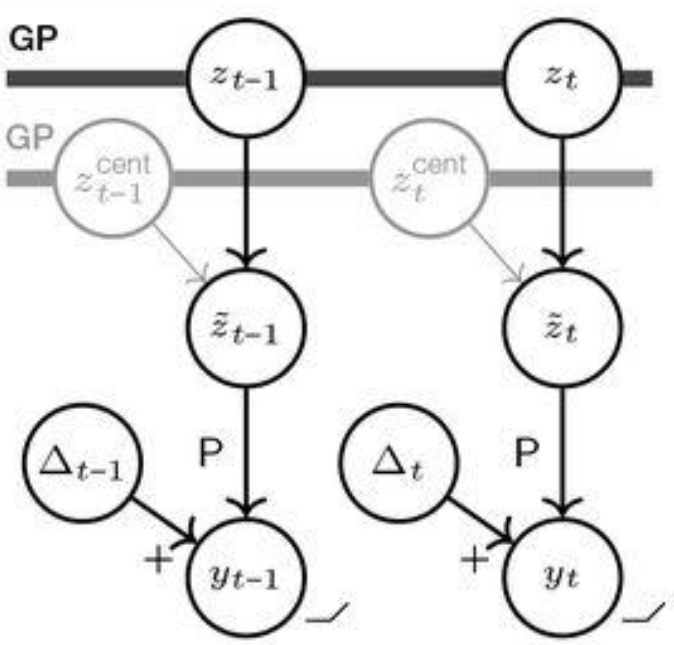
ICMS causes “contraction” of neural trajectory



For each session, we computed a contraction factor, defined as the ratio of variance in the stimulated vs. non-stimulated neural states, with one corresponding to no contraction, and zero indicating complete contraction to a point.

Modelling ICMS-caused contraction

k Stimulation neural data, contraction-only model:



non-stim latents
shared between non-stim and stim

centroid of contraction
shared across conditions

stim latents (contracted)
R rotation, B_t diag $\in \mathbb{R}^{K \times K}$

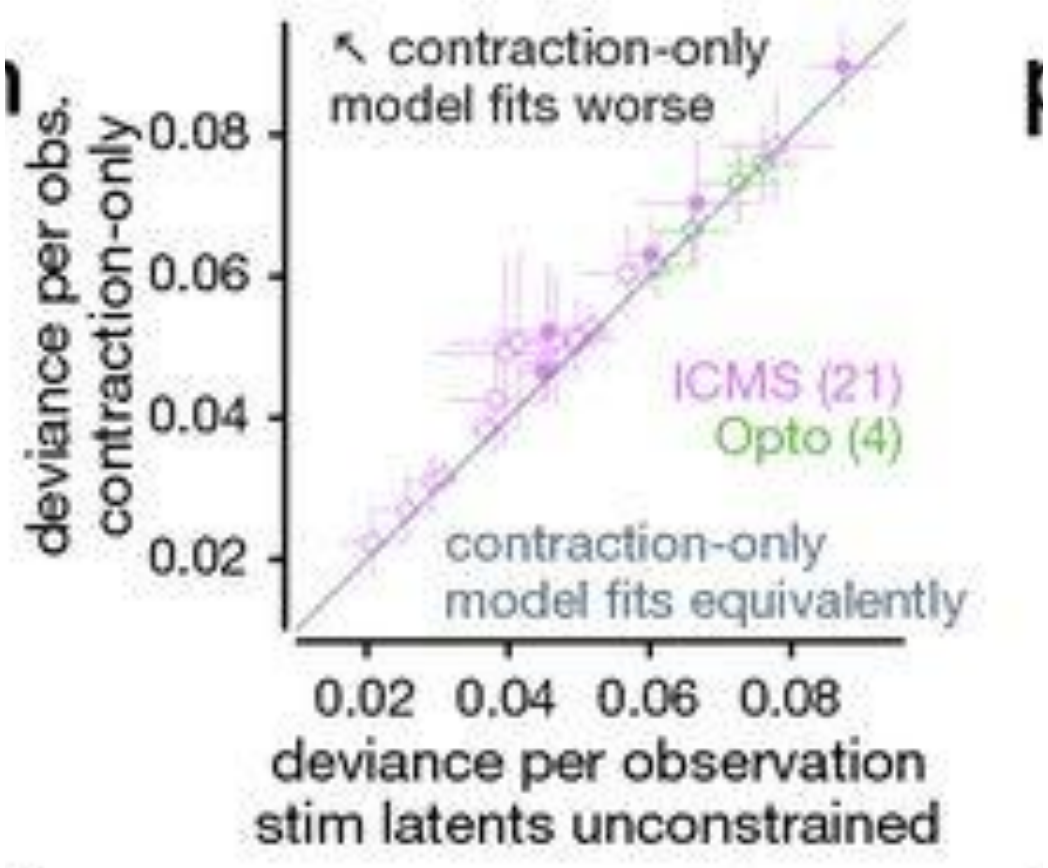
$$\tilde{z}_t = \left(z_t - z_t^{cent}\right) RB_tR^T + z_t^{cen}$$

additive component
$\Delta \in \mathbb{R}^{D \times T}$, low-rank

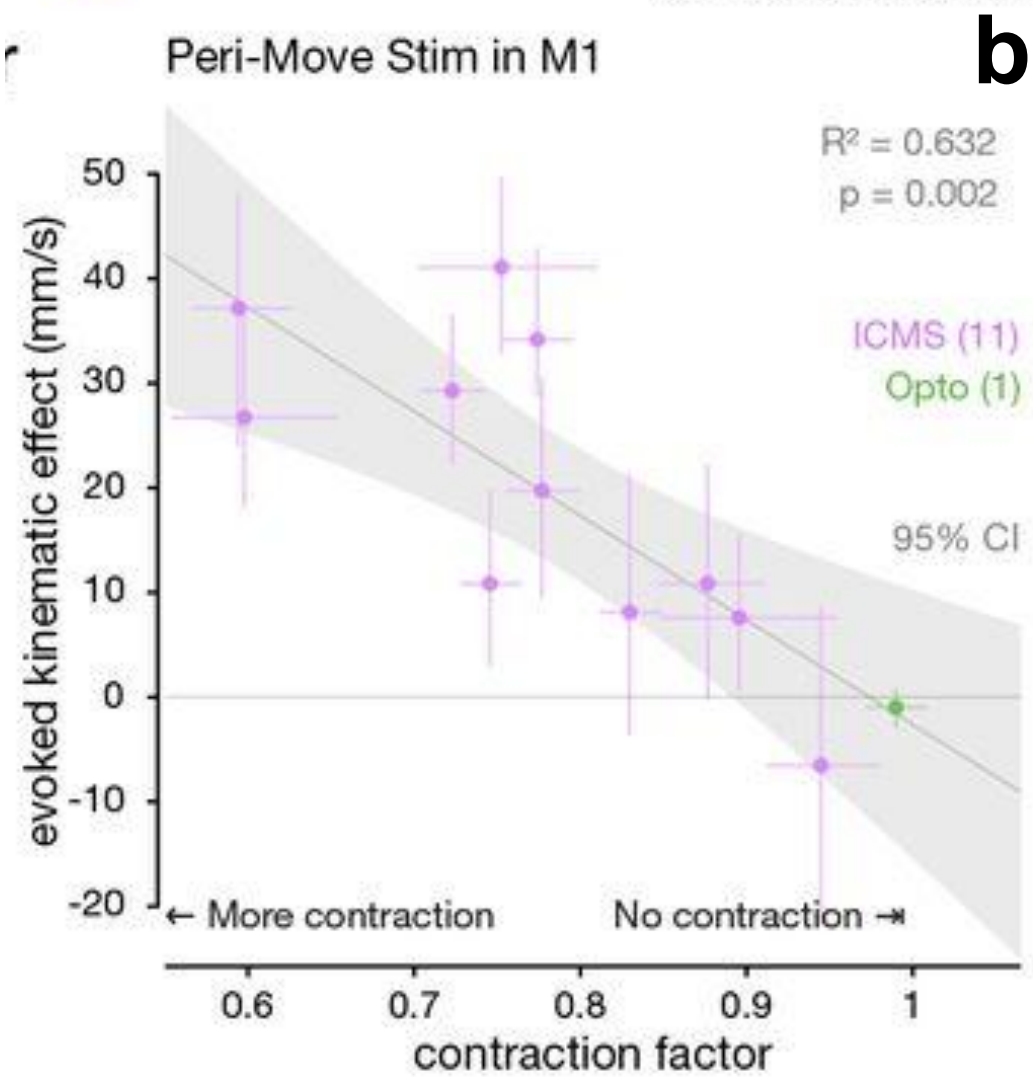
stim rates

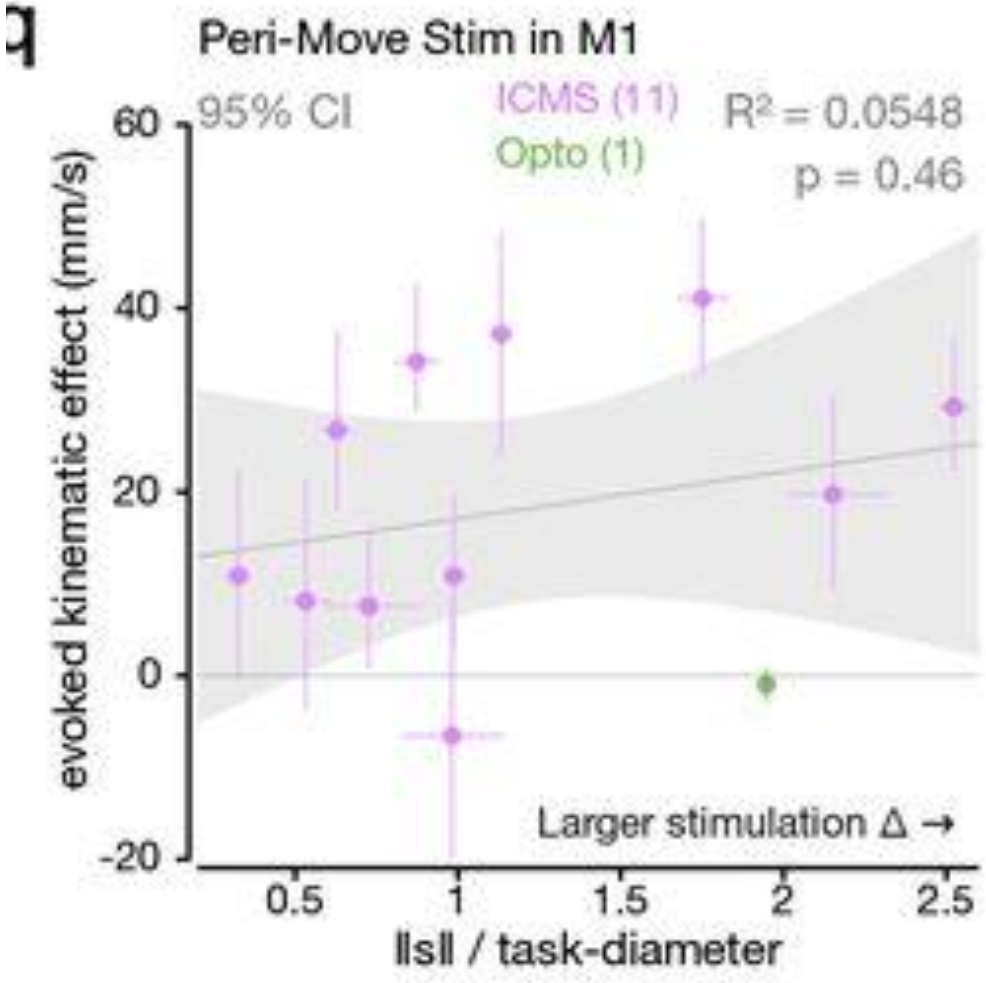
$$y_t = g\left(\tilde{z}_t P + \Delta_t + \mu\right)$$

w/ rectifying non-linearity to correctly handle silencing



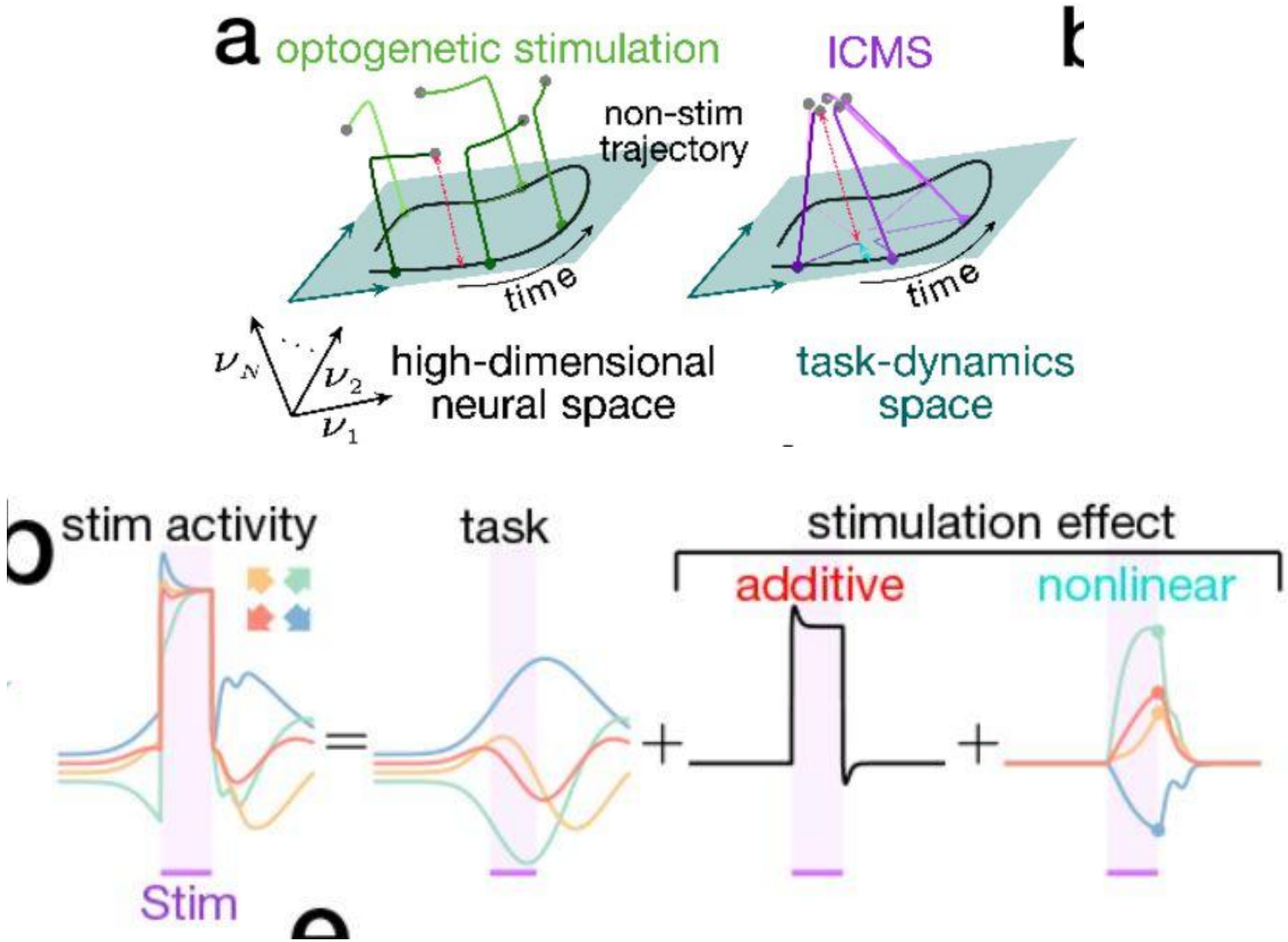
ICMS-caused dynamics contraction is related to behavior





Mechanistic model of ICMS-caused trajectory contraction

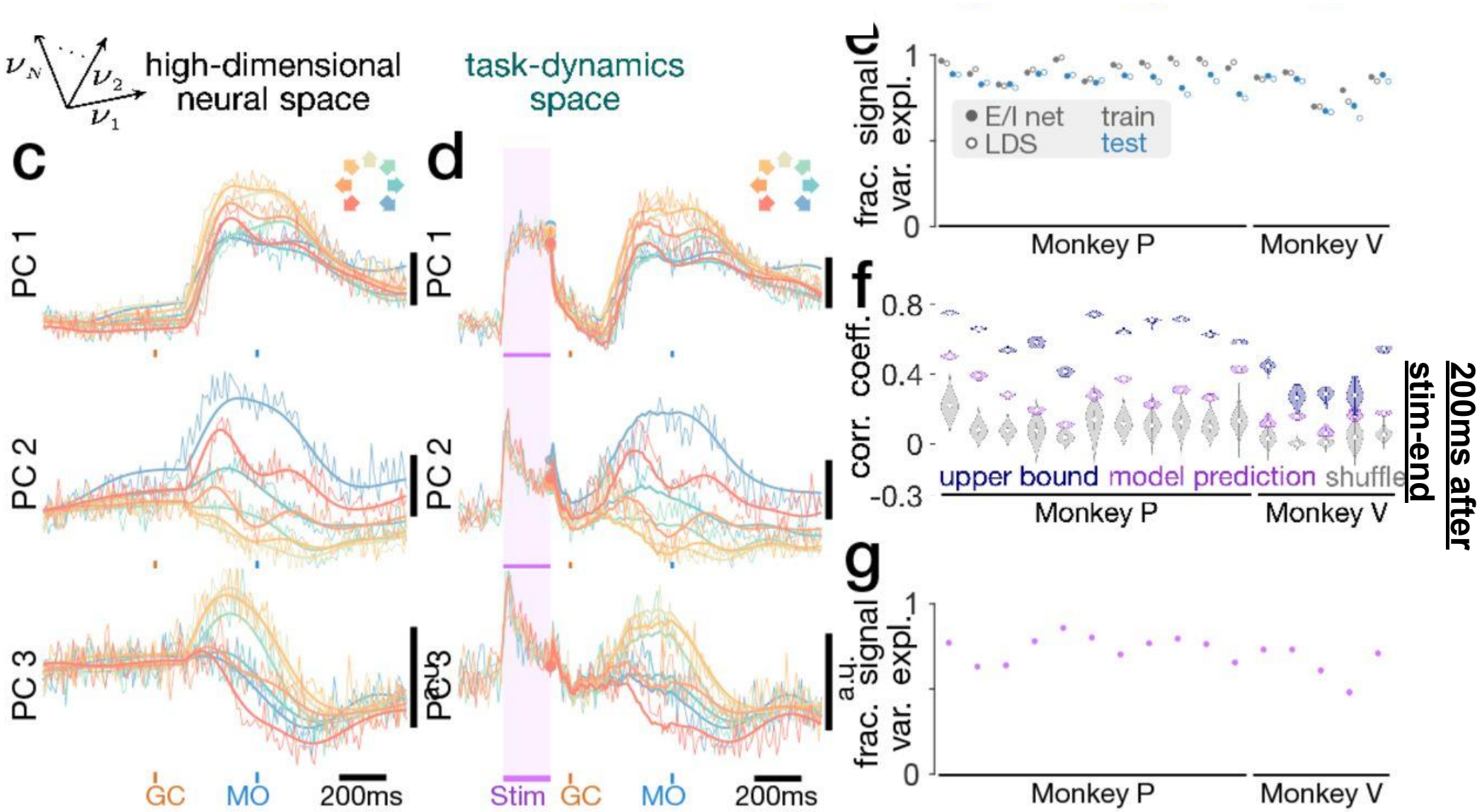
- The E/I network model was again fit exclusively to neural activity recorded in the absence of stimulation.
- The additive component of the stimulation effect was modeled as the fitted E/I network's response to a constant, additive input vector. The elements of this vector were drawn at random from a Gaussian distribution, since ICMS was not targeted to specific cell-types.
- The network responses to the additive perturbation were unable to capture the condition-dependent, nonlinear component of the stimulation responses.


a
optogenetic stimulation
ICMS
non-stim
trajectory
time
time
ν_N
ν_2
ν_1
high-dimensional
neural space
task-dynamics
space
b
stim activity
task
stimulation effect
additive
nonlinear
=
+
+
Stim

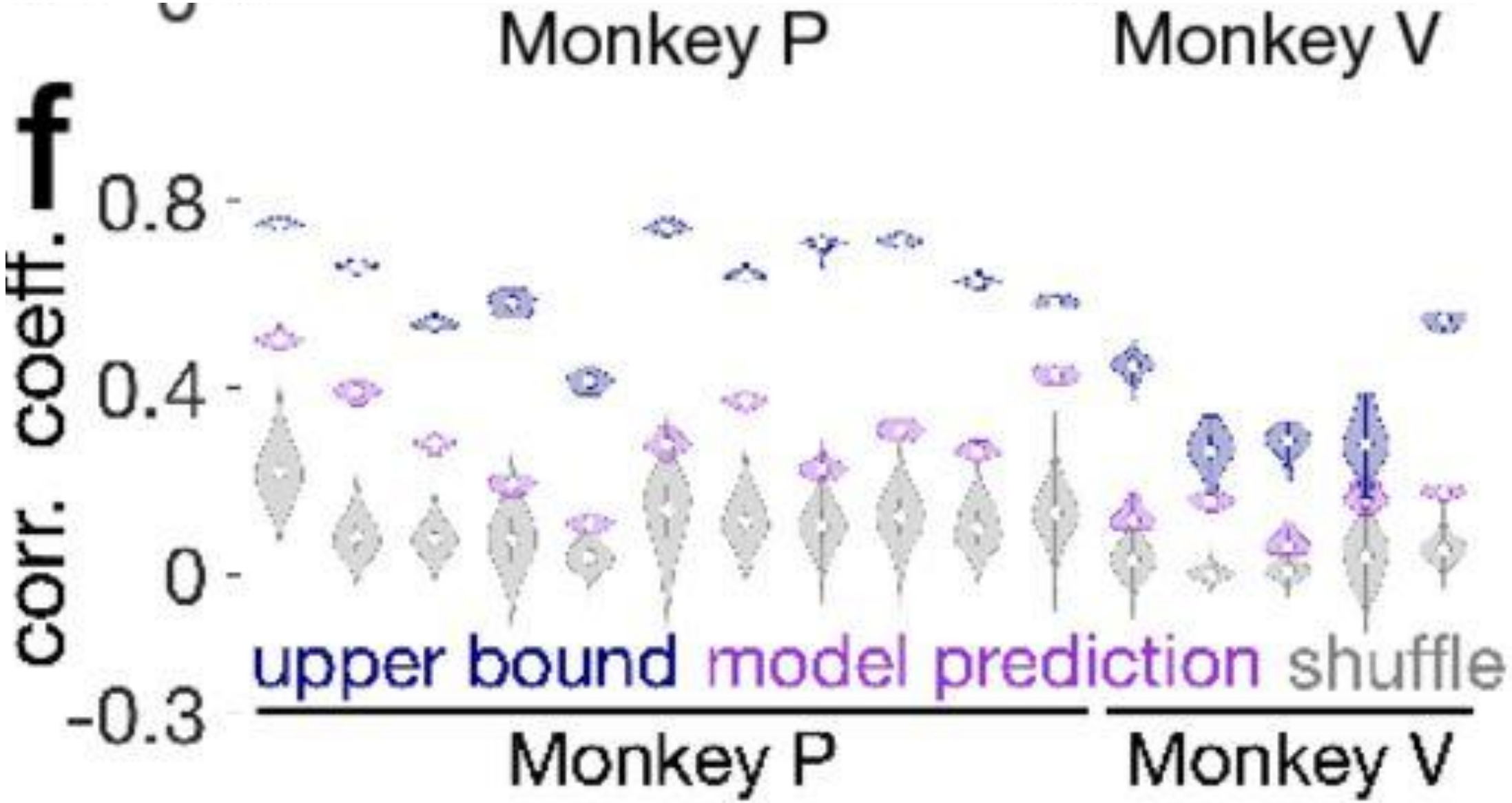
Mechanistic model of ICMS-caused trajectory contraction

- Thus, the nonlinear stimulation effects could be estimated by subtracting estimates of both task-related activity and additive effects from the recorded neural firing rates (Fig. 6b).
- We projected the state of the nonlinear stimulation effect at the last time-point of ICMS into the task dynamics space of our E/I model.
- Taking this projected state as the initial condition, we then predicted forward in time using the learned dynamics within the task dynamics space and evaluated Pearson's correlation coefficient between the model prediction and empirical nonlinear stimulation effect throughout the 200ms following the end of stimulation.

E/I model does(n't) explain non-linear effects of ICMS



- **Strikingly, the model predictions showed a strong correlation with the empirical effect in the majority of datasets (Fig. 6f).**
- **This demonstrates that the stimulation indeed engaged the task dynamics space in a way that is predictable based on normal task-related dynamical structure, without any further adjustment of model parameters.**


Monkey P
Monkey V
f
corr. coeff.
0.8
0.4
0
-0.3
upper bound
model prediction
shuffle
Monkey P
Monkey V

- The observation of long-lasting, predictable effects following ICMS constitutes strong evidence against path-following dynamics (**H3**). Instead, neural population responses to both optogenetic stimulation and ICMS were indicative of subspace structured dynamics embedded within an E/I network (**H2**).

Abu's Conclusions

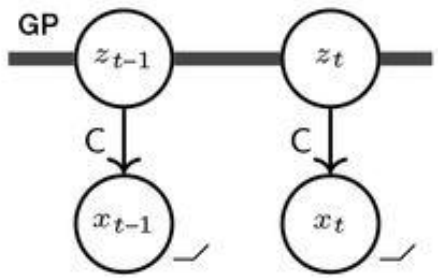
- Optogenetics:
  - Not hitting enough of motor cortex, or
  - Motor cortex is not involved in underlying dynamics
- ICMS
  - Finally hitting enough of motor cortex, or
  - Finally reaching enough circuitry are involved in underlying dynamics

Concerns

- No literature support for prediction of perturbation effects.
- Subspace-driven dynamics and Reservoir dynamics may be equivalent if only a few modes in the reservoir are driving the output?
- Their E/I model is a reservoir model?
  - We then optimized the E/I network connectivity matrix to maximize the likelihood that a linear readout from this task dynamics space matched the measured neural responses.

Supplement

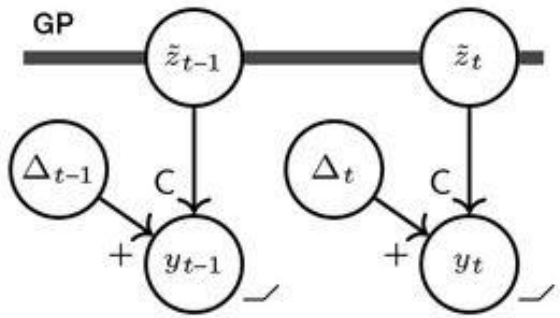
a **Non-stimulation neural data:**



non-stim latents
$z_t \in \mathbb{R}^K$, with GP prior

non-stim rates
$x_t = g(Cz_t + d) \in \mathbb{R}^D$

Stimulation neural data:



stim latents
$\tilde{z}_t \in \mathbb{R}^K$, with GP prior

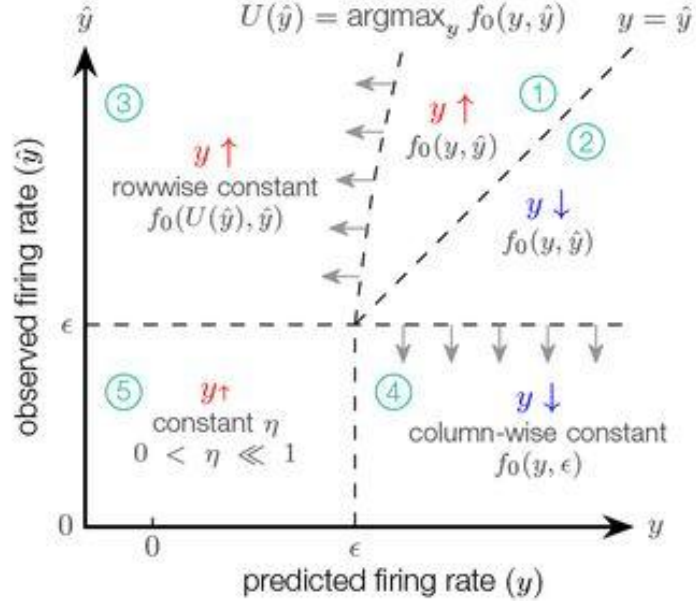
additive component
$\Delta \in \mathbb{R}^{D \times T}$, low-rank

stim rates
$y_t = g\left(C\tilde{z}_t + \Delta_t + d\right)$

w/ rectifying non-linearity to correctly handle silencing

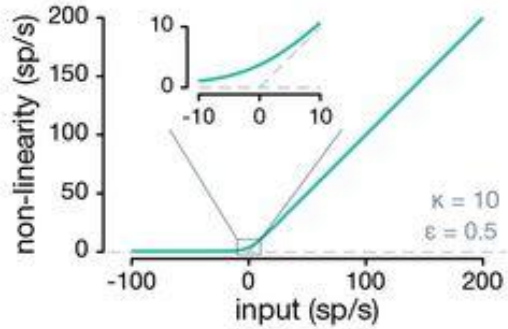
b **Soft-plus rectified Poisson observation model:**

Schematic of $f(y, \hat{y}) = \frac{\partial \ell}{\partial y}\Big|_{y,\hat{y}}$ defined piecewise via f_0 for Poisson model with soft-plus $g(\cdot)$ applied.



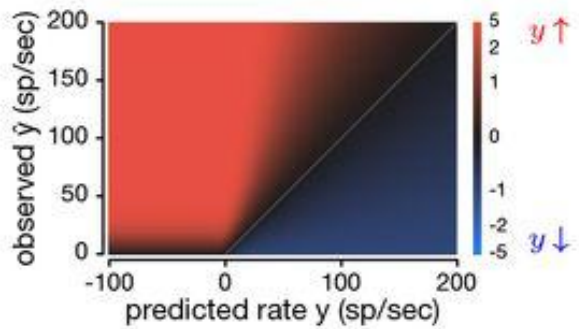
Soft-plus rectification of predicted rates

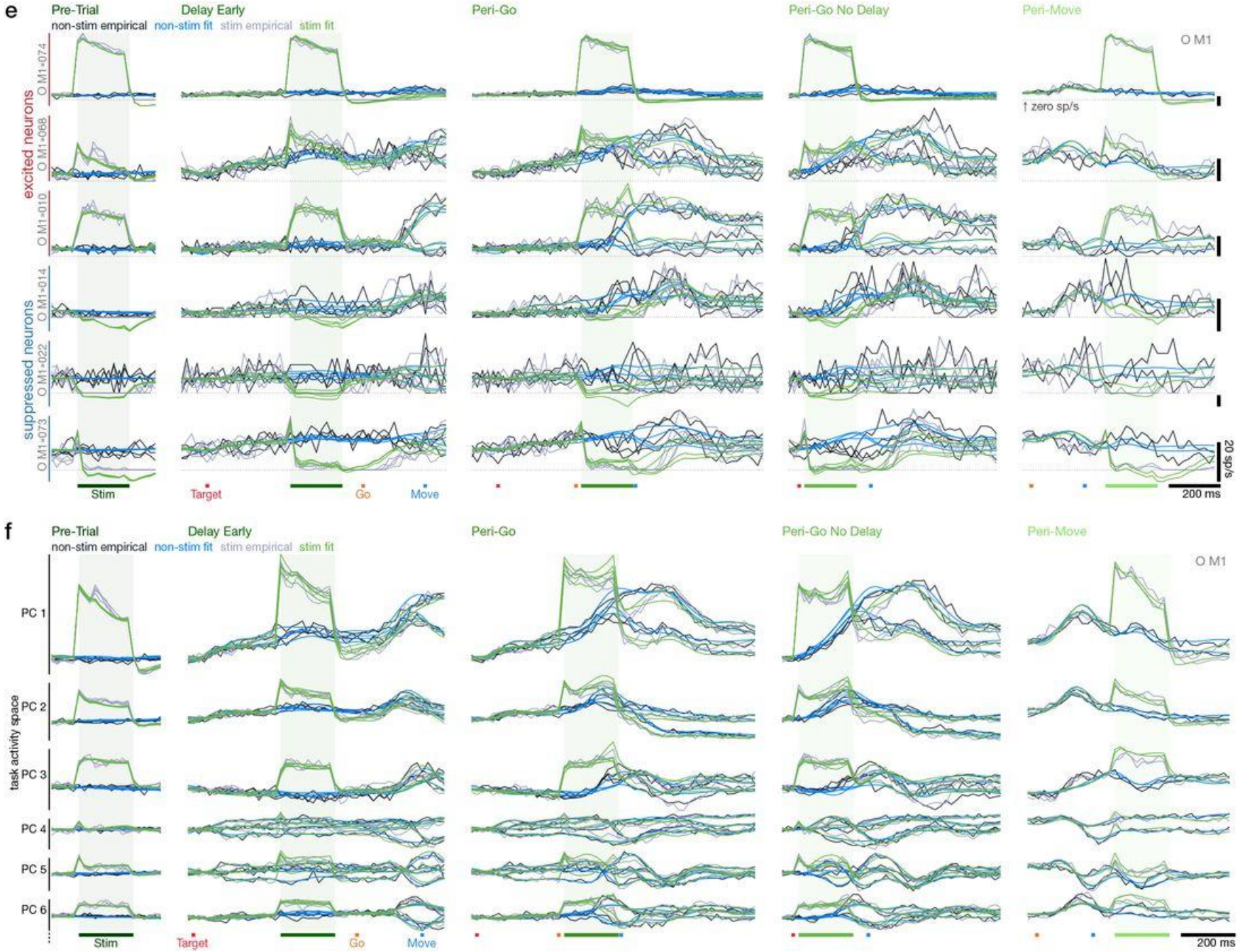
$g_{\kappa,\epsilon}(y) = \frac{1}{\kappa} \log(e^{\kappa y} + e^{\kappa \epsilon})$

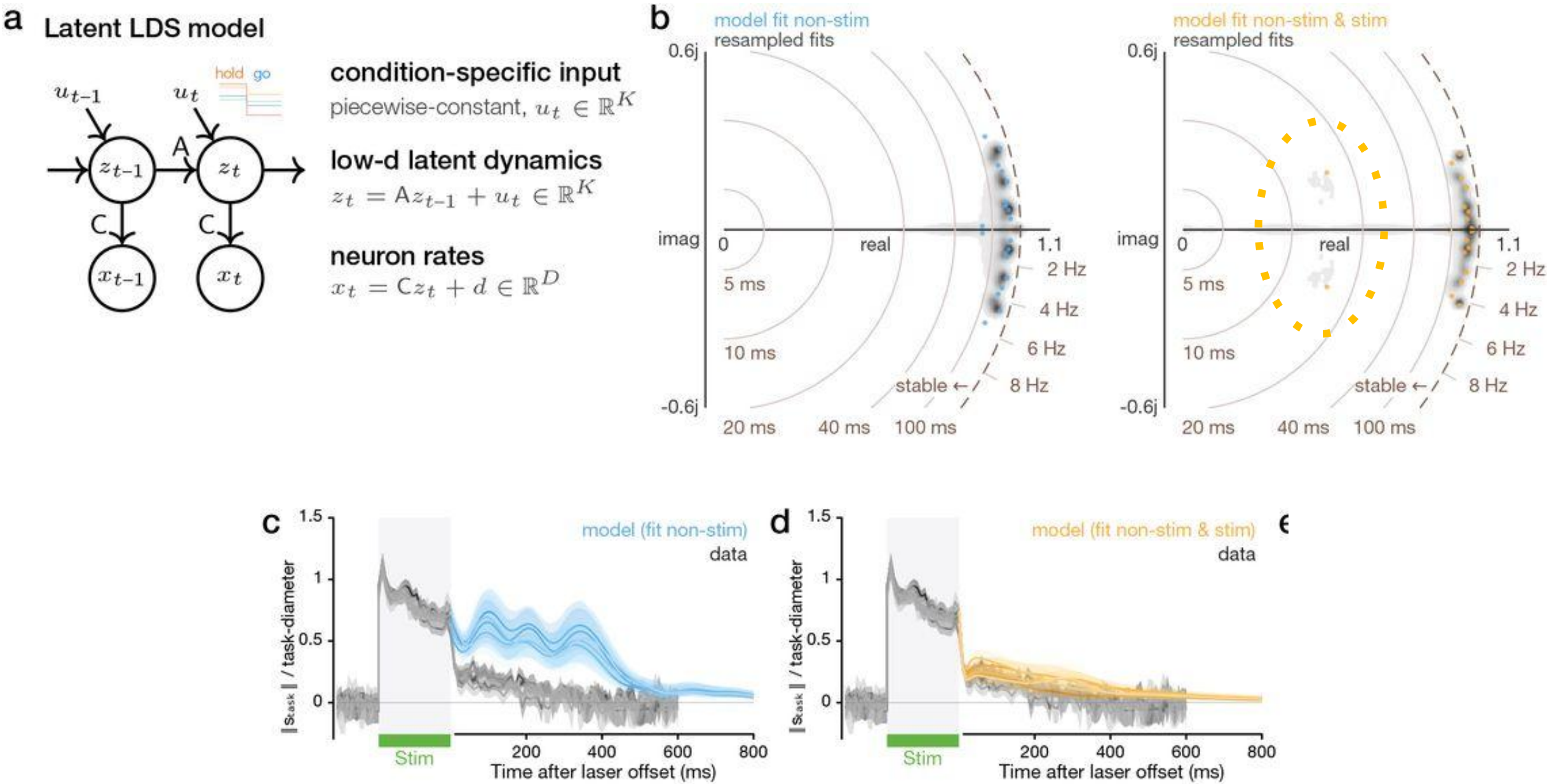


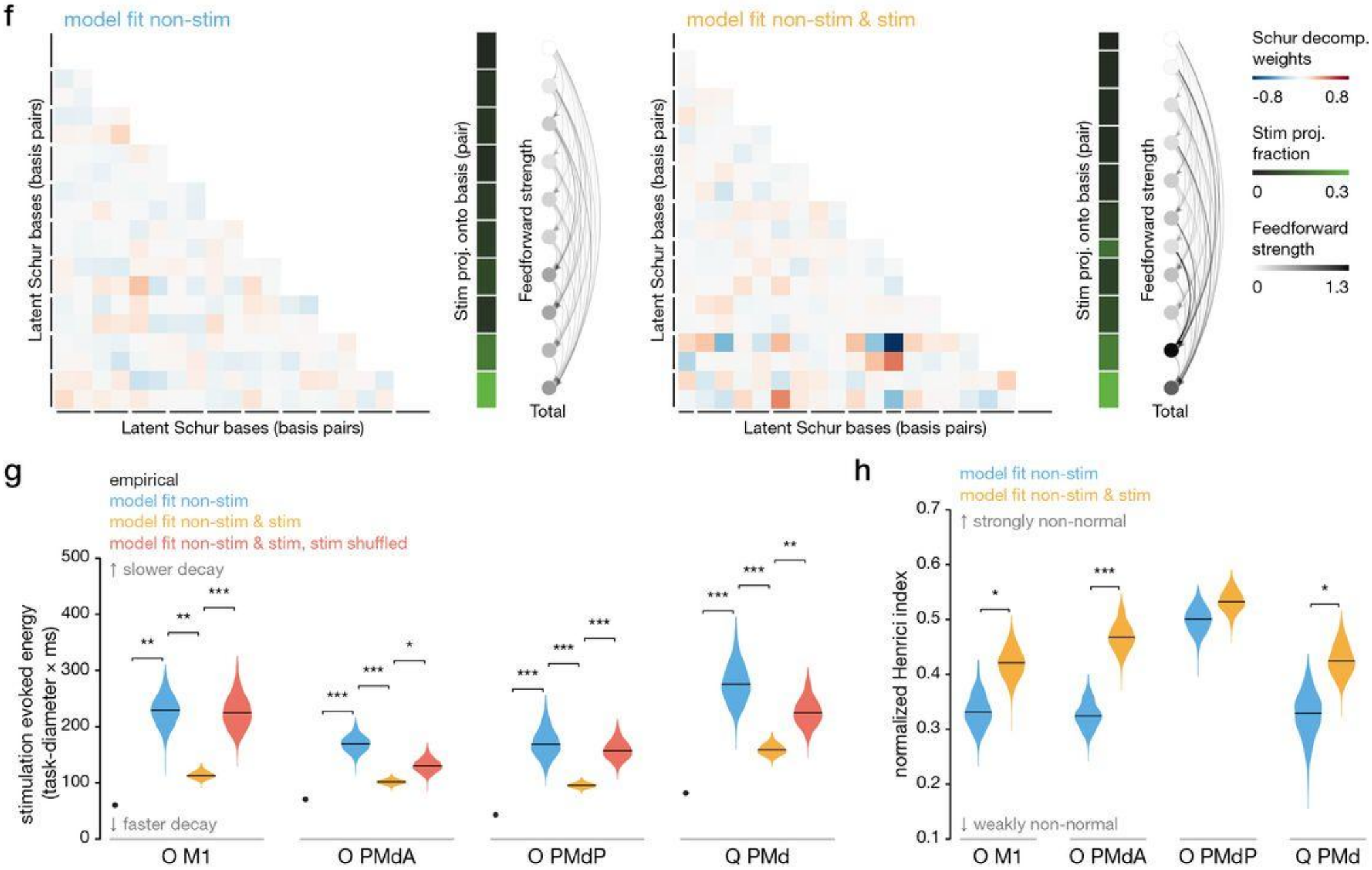
d **Gradient of predicted rate**

$f_{\kappa,\epsilon,\eta}(y, \hat{y}) = \frac{\partial \ell}{\partial y}\Big|_{y,\hat{y}}$

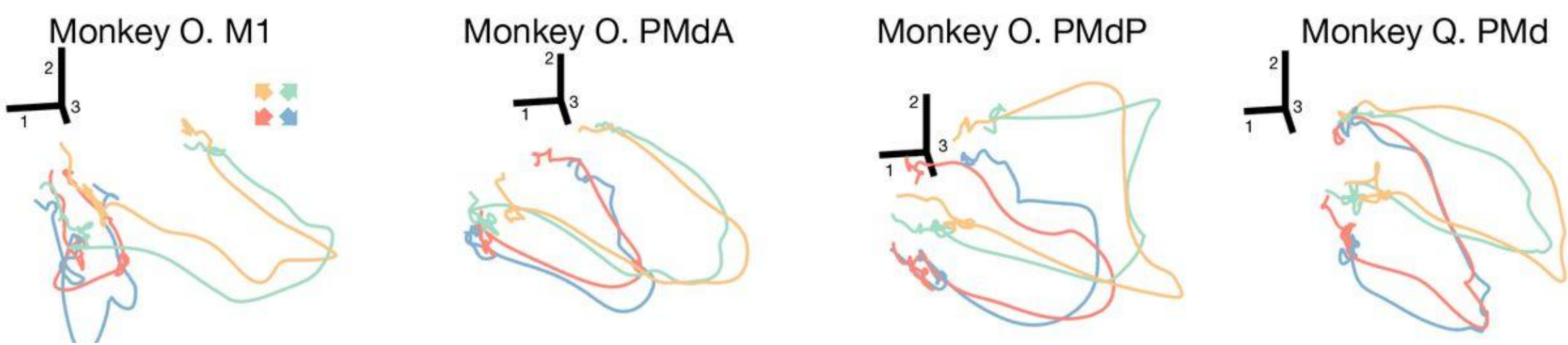



e
Pre-Trial
Delay Early
Peri-Go
Peri-Go No Delay
Peri-Move
non-stim empirical
non-stim fit
stim empirical
stim fit
O M1
excited neurons
suppressed neurons
O M1-074
O M1-068
O M1-010
O M1-014
O M1-022
O M1-073
↑ zero sp/s
20 sp/s
Stim
Target
Go
Move
200 ms
f
Pre-Trial
Delay Early
Peri-Go
Peri-Go No Delay
Peri-Move
non-stim empirical
non-stim fit
stim empirical
stim fit
O M1
task activity space
PC 1
PC 2
PC 3
PC 4
PC 5
PC 6
Stim
Target
Go
Move
200 ms


a
Latent LDS model
hold go
u_{t-1}
u_t
A
z_{t-1}
z_t
C
C
x_{t-1}
x_t
condition-specific input
piecewise-constant, $u_t \in \mathbb{R}^K$
low-d latent dynamics
$z_t = Az_{t-1} + u_t \in \mathbb{R}^K$
neuron rates
$x_t = Cz_t + d \in \mathbb{R}^D$
b
model fit non-stim
resampled fits
model fit non-stim & stim
resampled fits
0.6j
-0.6j
imag
real
0
1.1
2 Hz
4 Hz
6 Hz
8 Hz
stable ←
5 ms
10 ms
20 ms
40 ms
100 ms
c
d
e
model (fit non-stim)
model (fit non-stim & stim)
data
$\|x_{task}\|$ / task-diameter
1.5
1
0.5
0
Stim
200
400
600
800
Time after laser offset (ms)

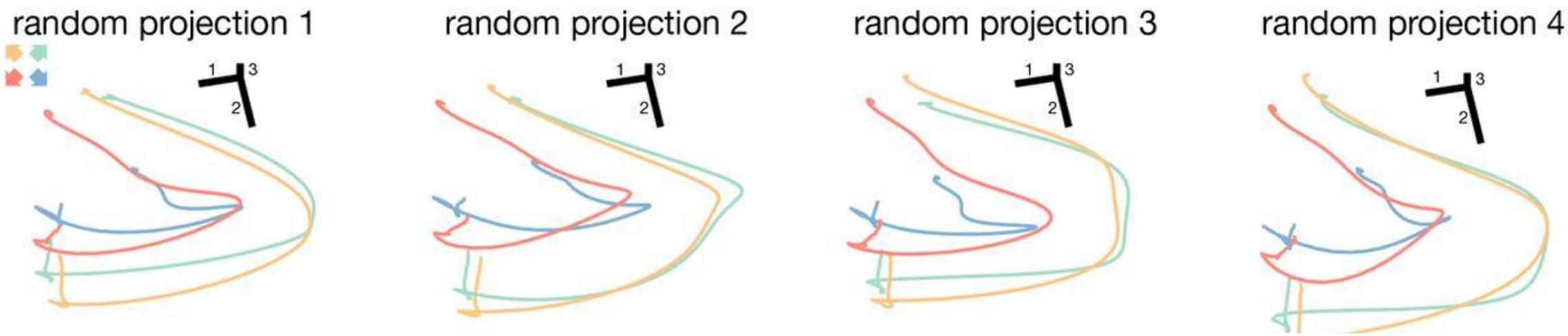

f
model fit non-stim
Latent Schur bases (basis pairs)
Latent Schur bases (basis pairs)
Stim proj. onto basis (pair)
Feedforward strength
Total
model fit non-stim & stim
Latent Schur bases (basis pairs)
Latent Schur bases (basis pairs)
Stim proj. onto basis (pair)
Feedforward strength
Total
Schur decomp. weights
-0.8
0.8
Stim proj. fraction
0
0.3
Feedforward strength
0
1.3
g
empirical
model fit non-stim
model fit non-stim & stim
model fit non-stim & stim, stim shuffled
↑ slower decay
↓ faster decay
stimulation evoked energy (task-diameter × ms)
500
400
300
200
100
0
O M1
O PMdA
O PMdP
Q PMd
h
model fit non-stim
model fit non-stim & stim
↑ strongly non-normal
↓ weakly non-normal
normalized Henrici index
0.7
0.6
0.5
0.4
0.3
0.2
0.1
O M1
O PMdA
O PMdP
Q PMd

a



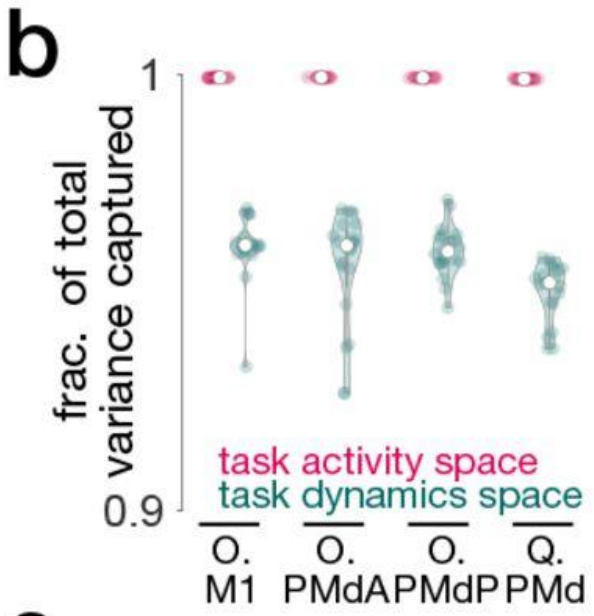
Random Projections of Activity from a trained network

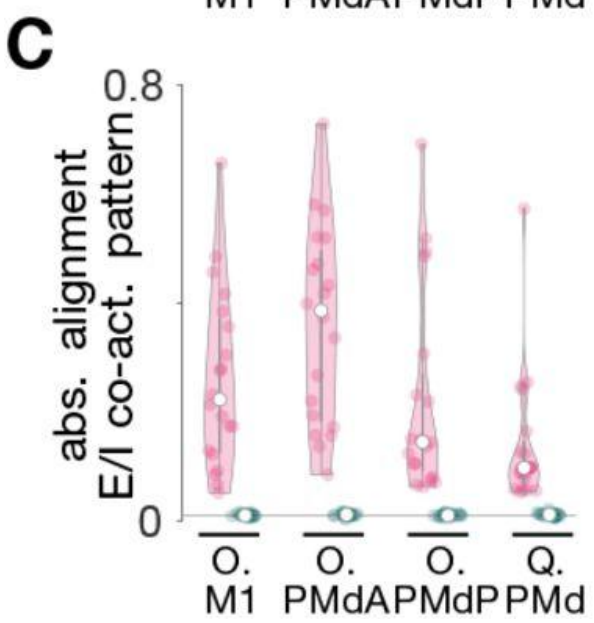
e

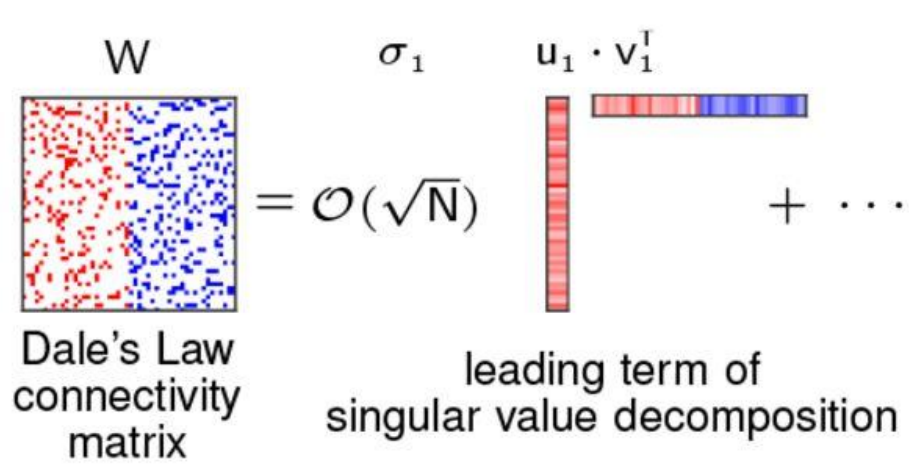


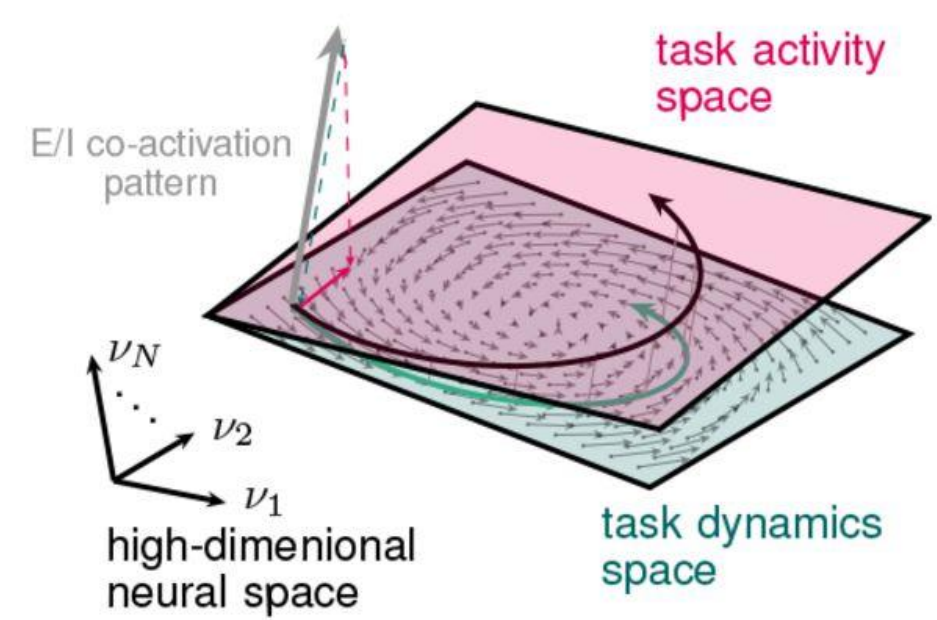
Mechanistic explanation for perturbation results

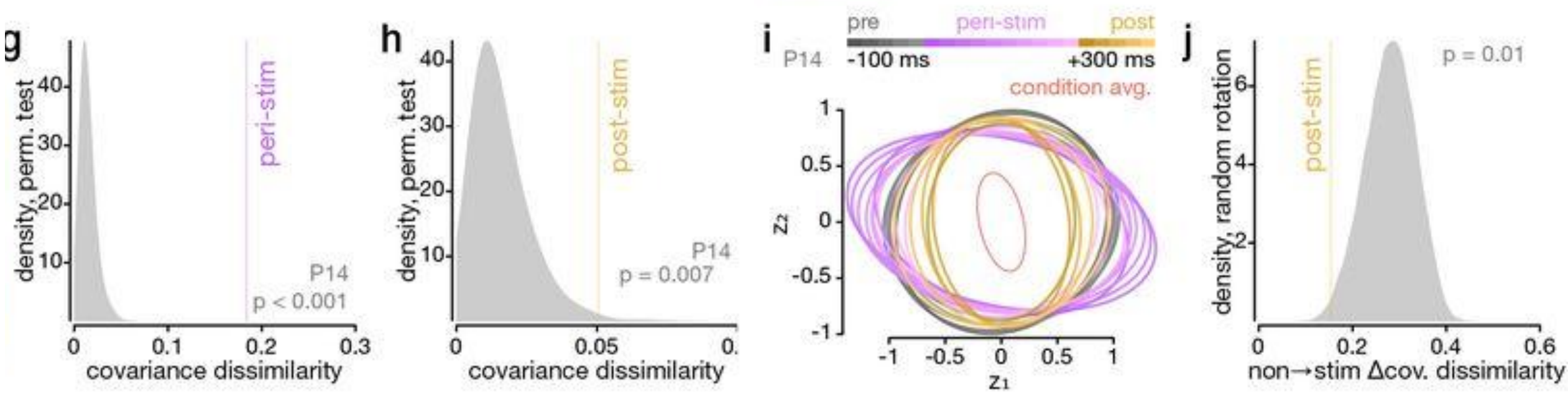
The task-activity space dimensionality is chosen to match that of the task-dynamics space for each dataset.

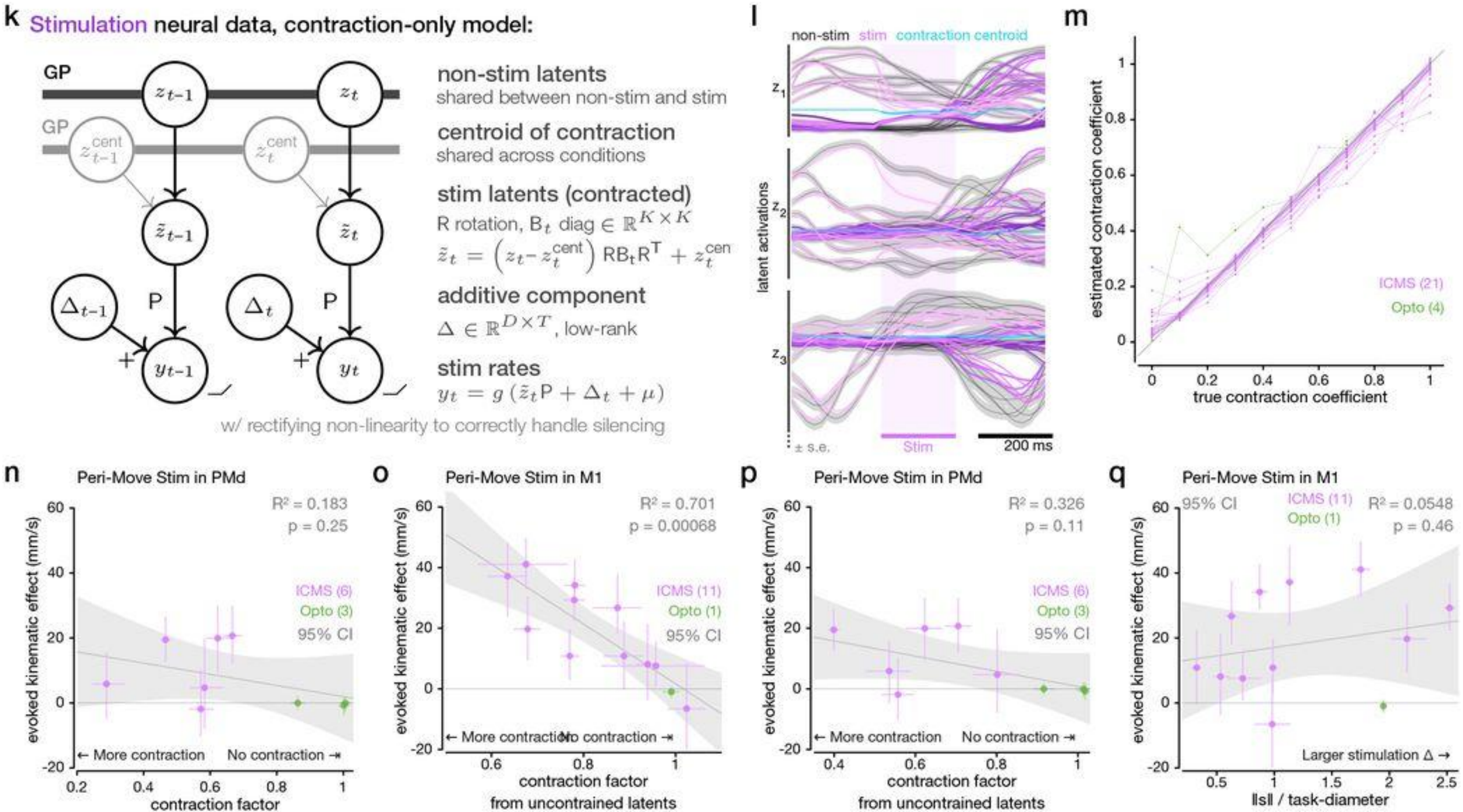









g
density, perm. test
peri-stim
P14
p < 0.001
covariance dissimilarity
h
density, perm. test
post-stim
P14
p = 0.007
covariance dissimilarity
i
pre
peri-stim
post
P14
-100 ms
+300 ms
condition avg.
z2
z1
j
density, random rotation
post-stim
p = 0.01
non→stim Δcov. dissimilarity


k
Stimulation neural data, contraction-only model:
GP
GP
non-stim latents
shared between non-stim and stim
centroid of contraction
shared across conditions
stim latents (contracted)
R rotation, B_t diag ∈ ℝ^{K×K}
z̃_t = (z_t − z_t^cent) RB_tR^T + z_t^cen
additive component
Δ ∈ ℝ^{D×T}, low-rank
stim rates
y_t = g(z̃_t P + Δ_t + μ)
w/ rectifying non-linearity to correctly handle silencing
l
non-stim
stim
contraction centroid
latent activations
± s.e.
Stim
200 ms
m
estimated contraction coefficient
true contraction coefficient
ICMS (21)
Opto (4)
n
Peri-Move Stim in PMd
R² = 0.183
p = 0.25
ICMS (6)
Opto (3)
95% CI
evoked kinematic effect (mm/s)
← More contraction
No contraction →
contraction factor
o
Peri-Move Stim in M1
R² = 0.701
p = 0.00068
ICMS (11)
Opto (1)
95% CI
evoked kinematic effect (mm/s)
← More contraction
No contraction →
contraction factor
from uncontrained latents
p
Peri-Move Stim in PMd
R² = 0.326
p = 0.11
ICMS (6)
Opto (3)
95% CI
evoked kinematic effect (mm/s)
← More contraction
No contraction →
contraction factor
from uncontrained latents
q
Peri-Move Stim in M1
95% CI
ICMS (11)
Opto (1)
R² = 0.0548
p = 0.46
evoked kinematic effect (mm/s)
Larger stimulation Δ →
||sI|| / task-diameter