

# Dimensionality Reduction of Multi-Trial Neural Data by Canonical Polyadic Tensor Decomposition

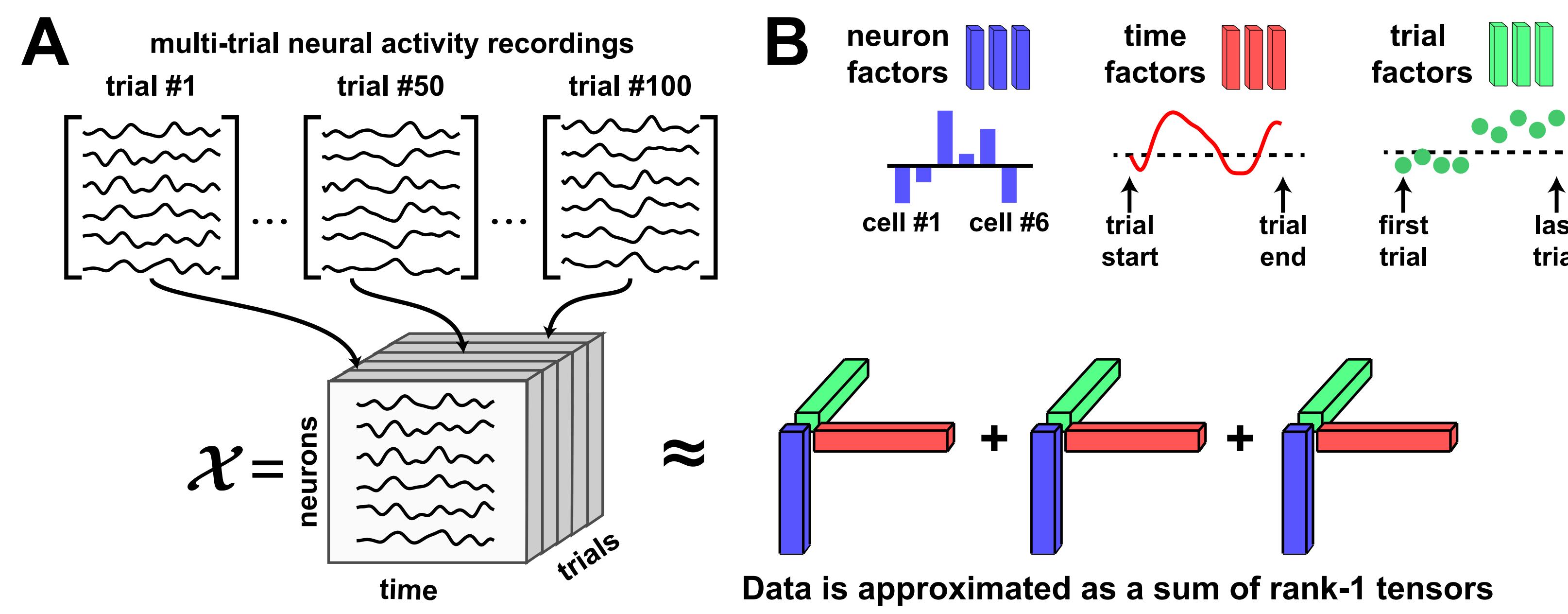


Alex H. Williams<sup>1</sup>, Hyun Kim<sup>2</sup>, Fori Wang<sup>1</sup>, Saurabh Vyas<sup>2</sup>, Krishna V. Shenoy<sup>2</sup>, Mark J. Schnitzer<sup>3</sup>, Tamara G. Kolda<sup>4</sup>, Surya Ganguli<sup>3</sup>

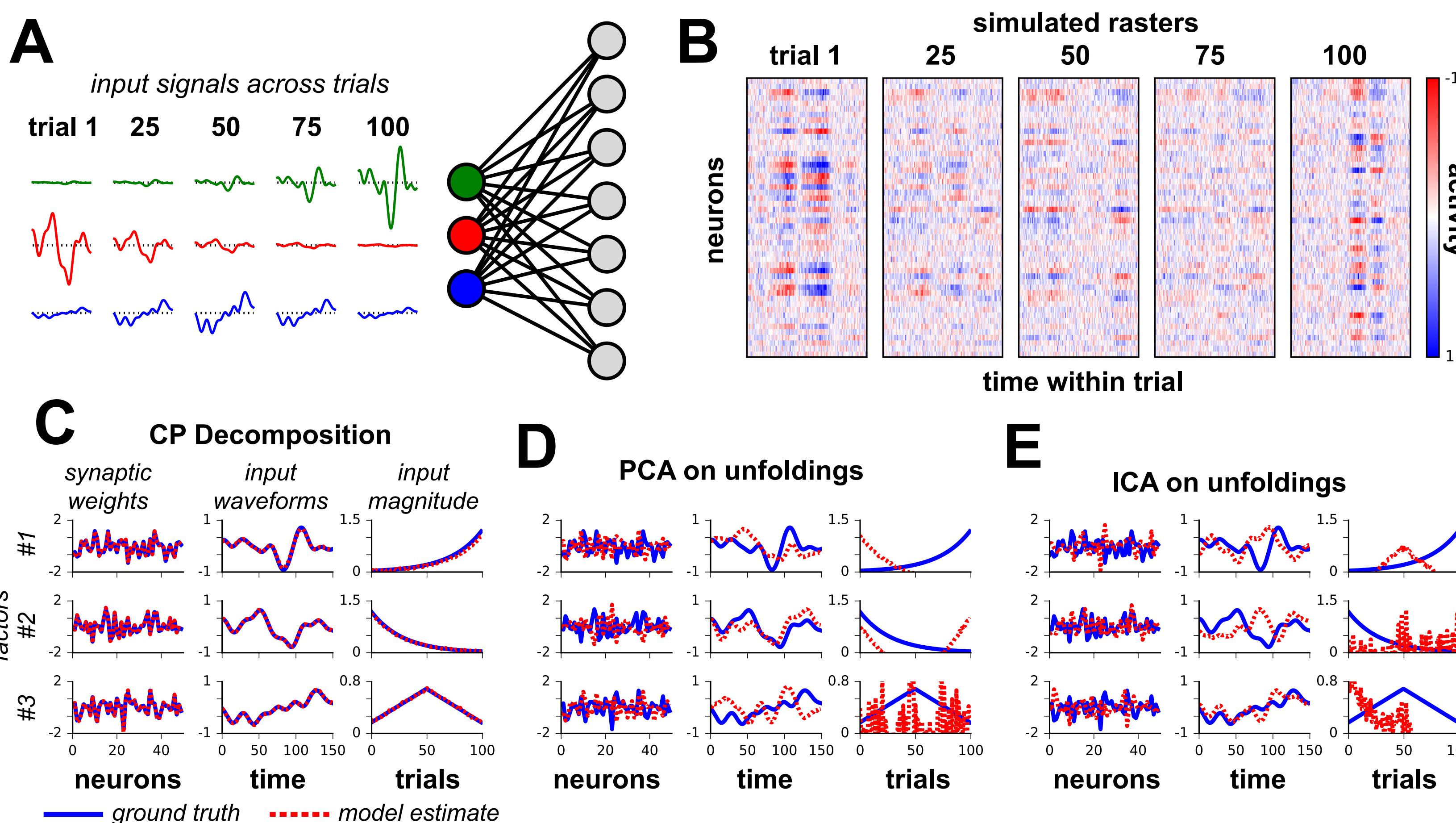
## Summary

Modern technologies enable neuroscientists to record from many hundreds of neurons for weeks or even months. Commonly used methods for dimensionality reduction identify low-dimensional features of within-trial neural dynamics, but do not model changes in neural activity across trials. We represent multi-trial data as a three-dimensional data array (a third-order tensor) with each entry indicating the activity of a particular neuron at a particular time on a particular trial, and use canonical polyadic (CP) tensor decomposition to find low-dimensional signatures of within and across-trial changes in neural dynamics. This approach produces more informative descriptions of experimental and synthetic multi-trial neural data than classic techniques such as principal and independent components analysis (PCA & ICA).

## Canonical Polyadic (CP) Tensor Decomposition



## Motivating Example (Linear 1-layer network)



CP decomposition can be interpreted as a 1-layer linear network with input signals that change in magnitude across trials (**A**). Simulated data with three inputs, 50 neurons, 100 trials, and additive Gaussian noise, is shown in (**B**). These data are approximately rank-3 since the dynamics of each neuron are a linear combination of the three input signals. CP decomposition (**C**) identified the synaptic weights (left column), input signal waveforms (middle column), and how the inputs changed in magnitude across trials (right column), up to a re-scaling and permutation of the factors (see panel on Properties and Intuition). Applying PCA (**D**) or ICA (**E**) to the three tensor unfoldings (matricizations) of the data did not recover this structure.

## Model Properties and Intuition

Denote the activity of neuron  $n$  at time  $t$  on trial  $k$  as  $x_{ntk}$ . The full dataset is a third-order tensor  $\mathcal{X}$  and the population activity on a given trial is a matrix  $\mathbf{X}_{::k}$ . CP decomposition approximates the data as a sum of rank-one tensors, which is a direct generalization of PCA to higher-order data arrays.

**Def:** a tensor  $\mathcal{Y}$  is *rank-one* if it is expressible as a vector outer product:

$$\mathcal{Y} = \mathbf{u} \circ \mathbf{v} \circ \mathbf{w} \iff y_{ijk} = u_i v_j w_k$$

The boxes below show two equivalent formulations of PCA and CP decomposition to highlight their similarities.

### PCA on a single trial

$$\begin{aligned} \underset{\mathbf{A}, \mathbf{B}}{\text{minimize}} \quad & \|\mathbf{X}_{::k} - \hat{\mathbf{X}}\|_F^2 \\ \text{subject to} \quad & \hat{\mathbf{X}} = \sum_{r=1}^R \mathbf{a}_{:r} \circ \mathbf{b}_{:r} \\ \underset{\mathbf{A}, \mathbf{B}}{\text{minimize}} \quad & \|\mathbf{X}_{::k} - \mathbf{AB}^T\|_F^2 \end{aligned}$$

### CP decomp. on all trials

$$\begin{aligned} \underset{\mathbf{A}, \mathbf{B}}{\text{minimize}} \quad & \|\mathcal{X} - \hat{\mathcal{X}}\|_F^2 \\ \text{subject to} \quad & \hat{\mathcal{X}} = \sum_{r=1}^R \mathbf{a}_{:r} \circ \mathbf{b}_{:r} \circ \mathbf{c}_{:r} \\ \underset{\mathbf{A}, \mathbf{B}, \mathbf{C}}{\text{minimize}} \quad & \sum_{k=1}^K \|\mathbf{X}_{::k} - \mathbf{ADiag}(\mathbf{c}_{:k})\mathbf{B}^T\|_F^2 \end{aligned}$$

We respectively refer to  $\mathbf{a}_{:r}$ ,  $\mathbf{b}_{:r}$ , and  $\mathbf{c}_{:r}$  as the *neuron factors*, *time factors*, and *trial factors*, of the CP decomposition. These vectors are collected into the columns of the matrices  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$ , which we call the *factor matrices*. A crucial disadvantage of PCA is that its objective function is invariant to any invertible linear transformation (since  $\mathbf{AB}^T = \mathbf{AF}^{-1}\mathbf{FB}^T = \mathbf{A}'\mathbf{B}'^T = \hat{\mathbf{X}}$ ), meaning that the latent factors can only be recovered up to a linear subspace. In contrast, due to a seminal result of Kruskal [1977], the factors of higher-order CP decomposition are essentially recoverable so long as they are linearly independent.

We fit CP decompositions by alternating least-squares (ALS) as reviewed in [Bader & Kolda, 2009, SIAM Review]. This procedure may converge to local minima, but similar factors (after alignment by permutation and re-scaling) were consistently obtained from different random initializations in all applications.

## Acknowledgements and Funding:

This work was supported by grants from Howard Hughes Medical Institute, the McNight Foundation, the Burroughs-Wellcome Fund, and the Department of Energy Computational Science Graduate Fellowship. We thank Jeffrey Seely (Columbia University) and Madeleine Udell (Cornell University) for helpful conversations, and Chris Stock (Stanford University) for providing simulated data from the Mante-Susillo model network.

## Author Affiliations and Contact Information:

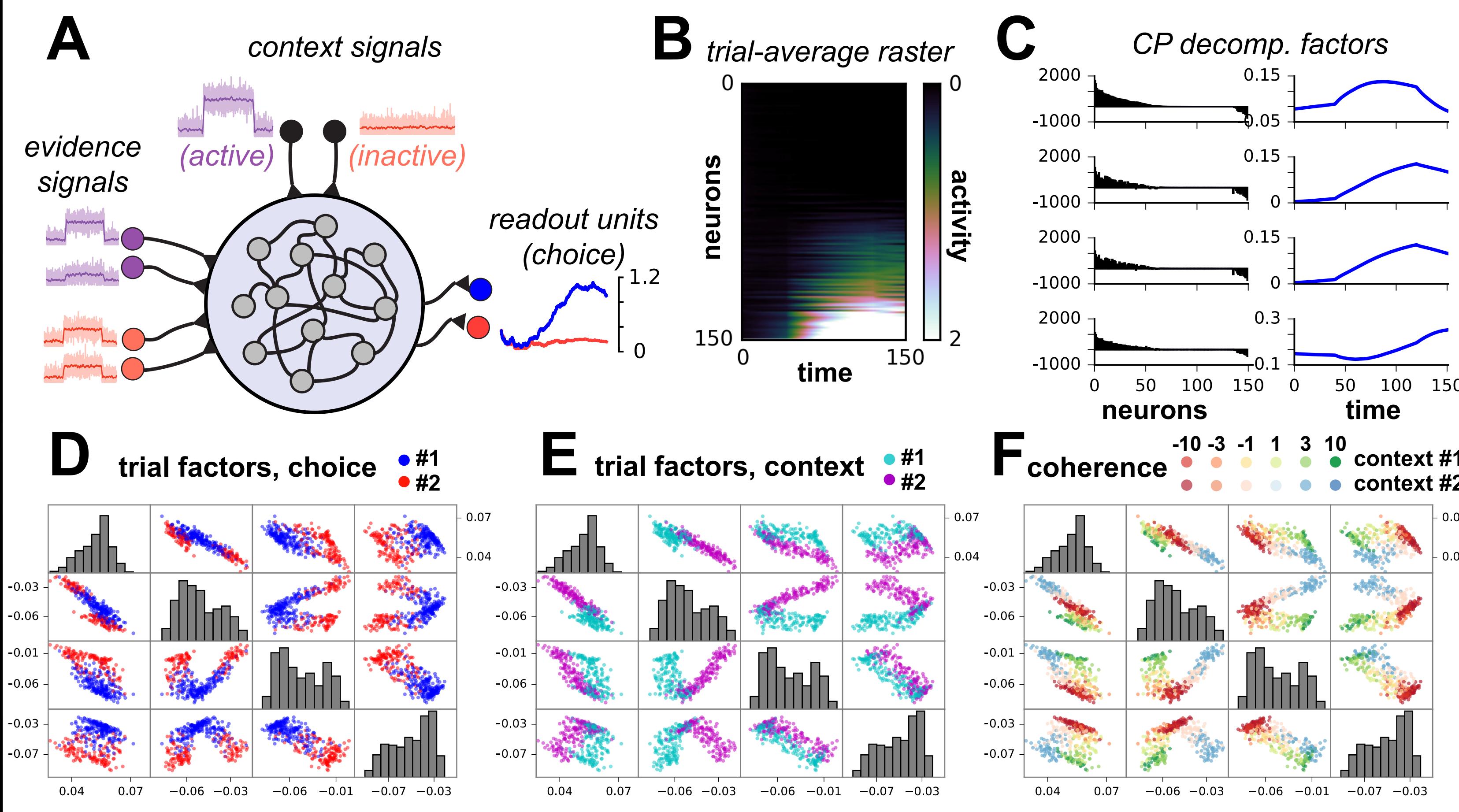
Departments of Neurobiology<sup>1</sup>, Electrical Engineering<sup>2</sup>, & Applied Physics<sup>3</sup>, Stanford University, Stanford, CA. Data Sciences & Cyber Analytics Department<sup>4</sup>, Sandia National Labs, Livermore, CA. \*Send correspondence to [ahwillia@stanford.edu](mailto:ahwillia@stanford.edu) and [srganguli@stanford.edu](mailto:srganguli@stanford.edu)

Demo Code: <https://github.com/ahwillia/tensor-demo>

MATLAB Toolbox: <http://www.sandia.gov/~tgkolda/TensorToolbox/>

## Case Study #1 (Artificial Recurrent Network)

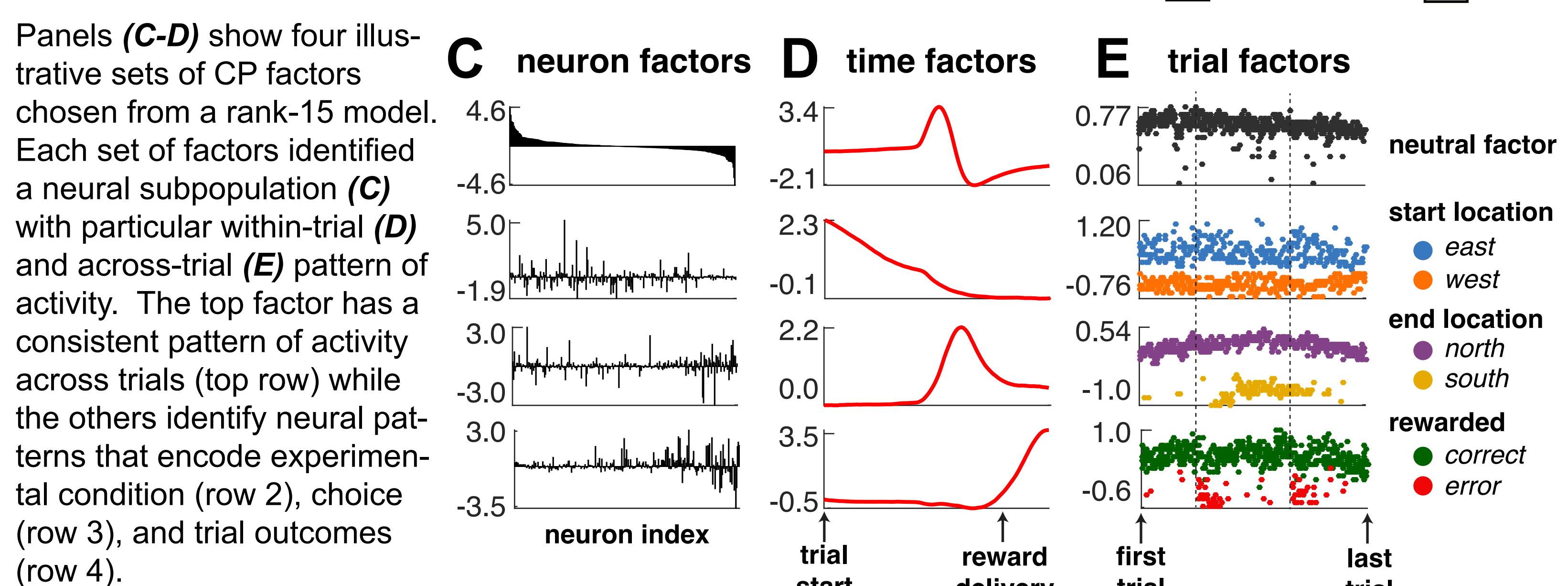
We used the Python package pycog [Song et al, 2016, PLoS Comp Bio] to train a recurrent network with rectified linear hidden units (**A**) to perform a context-dependent sensory integration task [Mante et al, 2013, Nature]. On each trial, one of two context-indicating input signals (**top, A**) was active. Each contextual cue was matched to a pair of noisy evidence signals (**left, A**), and the network used two readout units (**right, A**) to indicate which of the two evidence signals associated with the activated context was larger. We simulated trials from 3 coherence conditions across both contexts; the trial-averaged population activity is shown in (**B**). A rank-4 CP decomposition captures roughly 90% of variance in the data; the neuron and time factors shown in (**C**) indicate that a small fraction of neurons dominate and have ramping activity profiles. Each of these subpopulations is related to task-relevant variables as can be seen by plotting the pairwise relationships between trial factors in the trial factors (**D-F**). Within this reduced, 4-dimensional space, the trials clearly separate based on choice (**D**), trial context (**E**), and strength of evidence (**F**).



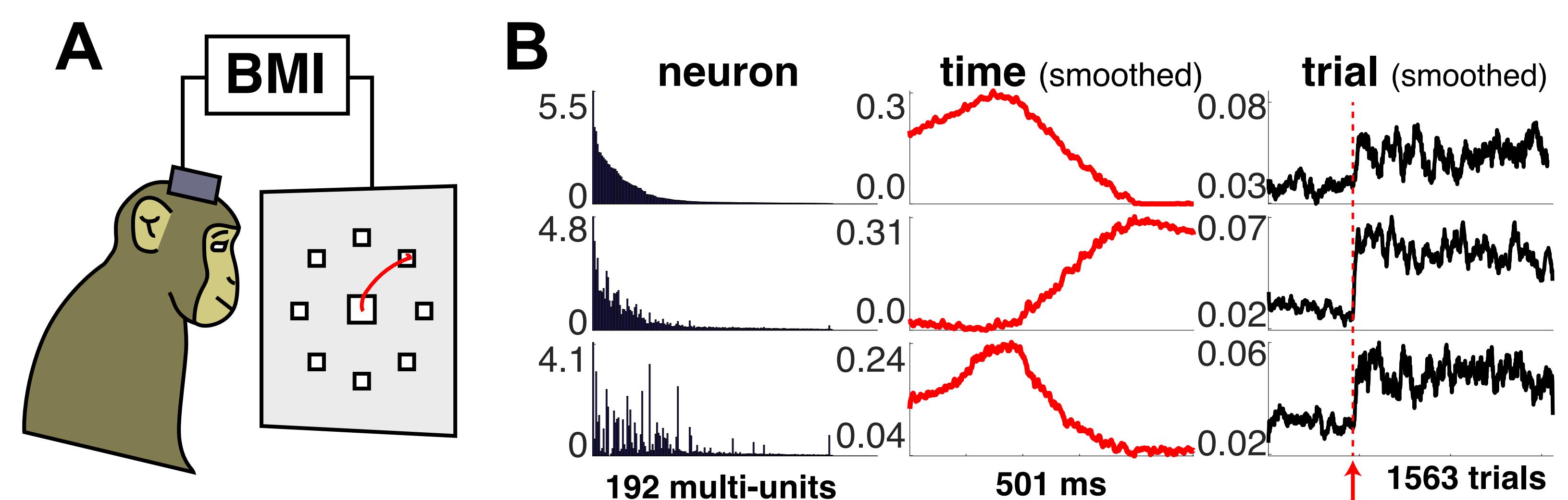
## Case Study #2 (Mouse Prefrontal Cortex)

Calcium dynamics of ~500 neurons in the medial prefrontal cortex were imaged with a miniature fluorescence microscope in mice performing a spatial navigation task. Mice began each trial in either the east or west arm of a four-armed maze, with the opposing arm blocked. Mice were trained to follow either an egocentric (**A**) or allocentric (**B**) navigation strategy for a water reward. The rewarded strategy was periodically switched, prompting the mouse to adapt strategies.

**CP decomposition identifies neural populations that are sensitive to mouse position, behavioral strategies, and rewards**



## Case Study #3 (Primate Motor/Premotor Cortex)



A Rhesus monkey was trained to make point-to-point reaches to visual targets in a 2D plane with a virtual cursor controlled by their contralateral arm (**A**). We recorded approximately 200 units using multi-electrode arrays (Utah arrays, Blackrock Microsystems) implanted in the pre-motor and motor cortex. We first fit a velocity Kalman filter to decode cursor velocities from firing neural rates, and then perturbed this decoder by rotating the output cursor velocities (a visuomotor rotation) by 30 degrees. Panel (**B**) shows the results of a rank-3 CP decomposition applied to this dataset. The factors identify neural populations which are preferentially activated by the BMI perturbation.