# Lecture 10: Discrete classification

- Midterm is in class on Monday, May 2, 2022. Covers material up to and including Poisson Processes. 2 cheat sheets (8.5” x 11” paper) allowed.
- Past years’ midterms are on Bruin Learn. 2017 was too long.
- TA review session: Franz 1178, April 28, 2022, 6-8pm. It will be recorded.
- HW #3 is due Wednesday, May 4, 2022, uploaded to Gradescope.
- HW regrade requests should be made via Gradescope. This will route the request to the grader who graded the question.
- You're welcome to bring a calculator for the MT but we wrote the exam so you don't need one.

# Lectures 9 + 10: Discrete Classification

- Reading assignment:

  - Required (uploaded to CCLE):

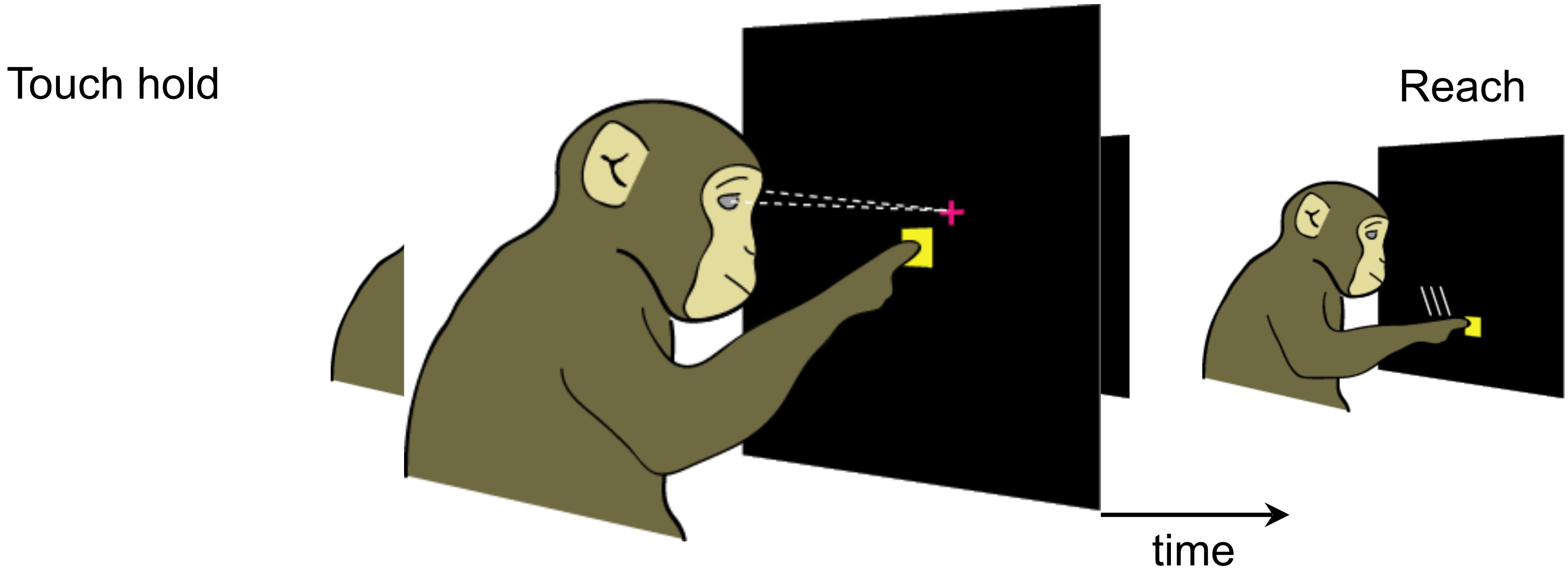
    - PRML p. 38–46: Decision Theory
    - PRML p. 179–184, 196–203: Linear Models for Classification
    - Santhanam G*, Ryu SI*, Yu BM, Afshar A, Shenoy KV (2006) A high-performance brain-computer interface. *Nature*. 442:195-198.

  - Optional (uploaded to CCLE):

    - Achtman N*, Afshar A*, Santhanam G, Yu BM, Ryu SI, Shenoy KV (2007) Free-paced high-performance brain-computer interfaces. *Journal of Neural Engineering*. 4:336-347.

    - Santhanam G, Yu BM, Gilja V, Afshar A, Ryu SI, Sahani M, Shenoy KV (2009) Factor-analysis methods for higher-performance neural prostheses. *Journal of Neurophysiology*. 102:1315-1330.

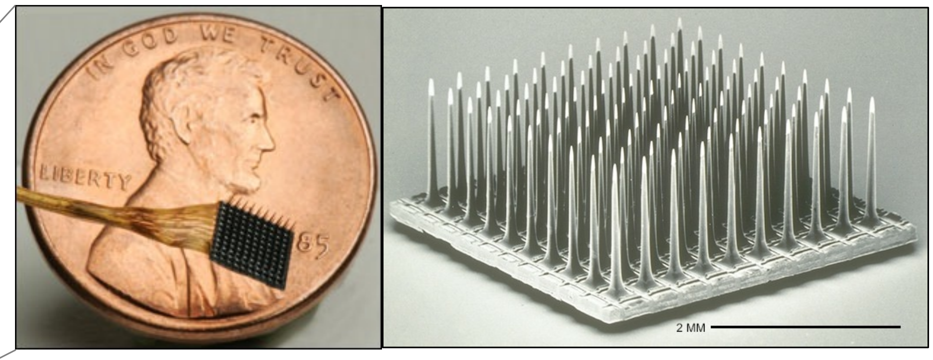
# Delayed reach task



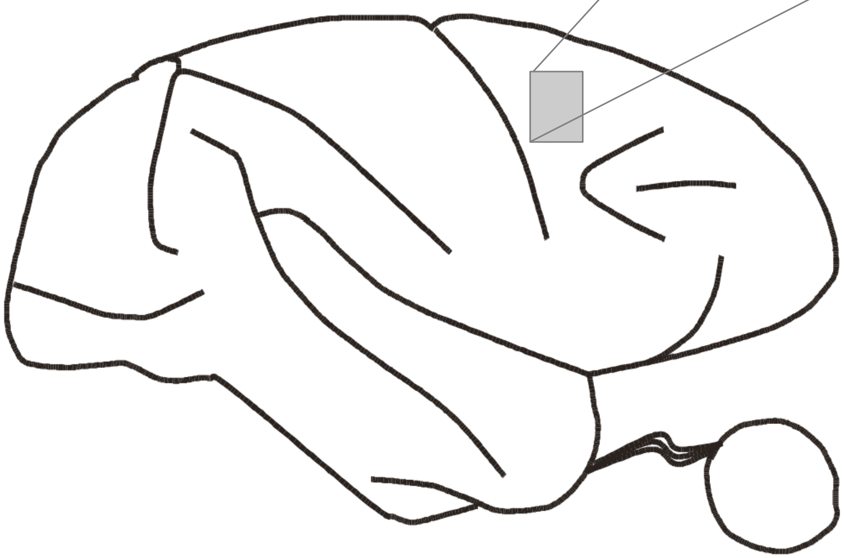
# Delayed reach task

# Recordings in premotor cortex

plan activity

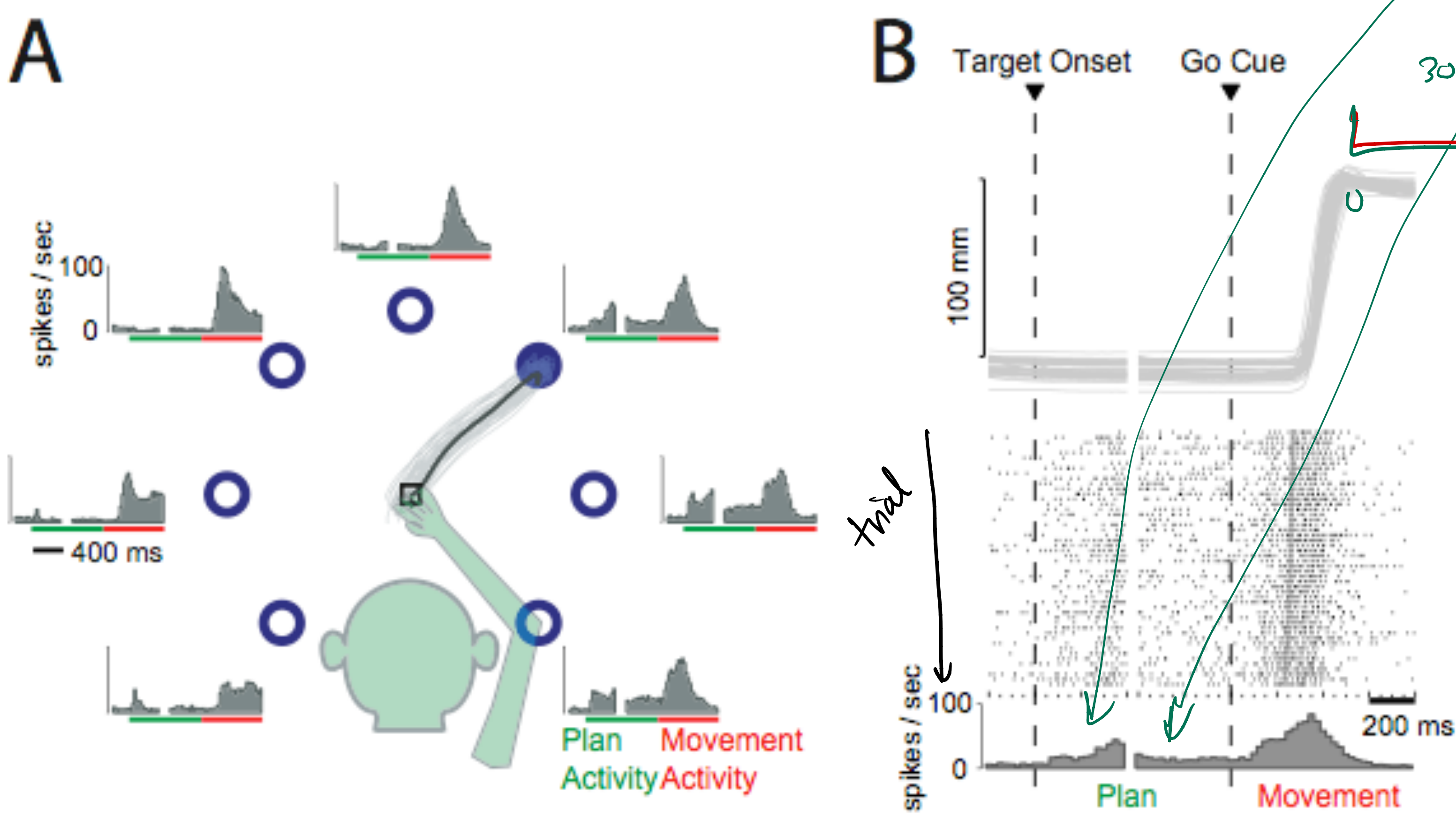



Utah 96-electrode array



# Delayed Reach Task, Plan & Movement Activity

- Firing rate of a typical PMd neuron in a delayed-reach behavioral task




400ms

300

900 ms

800

400

800ms

One neuron

trial

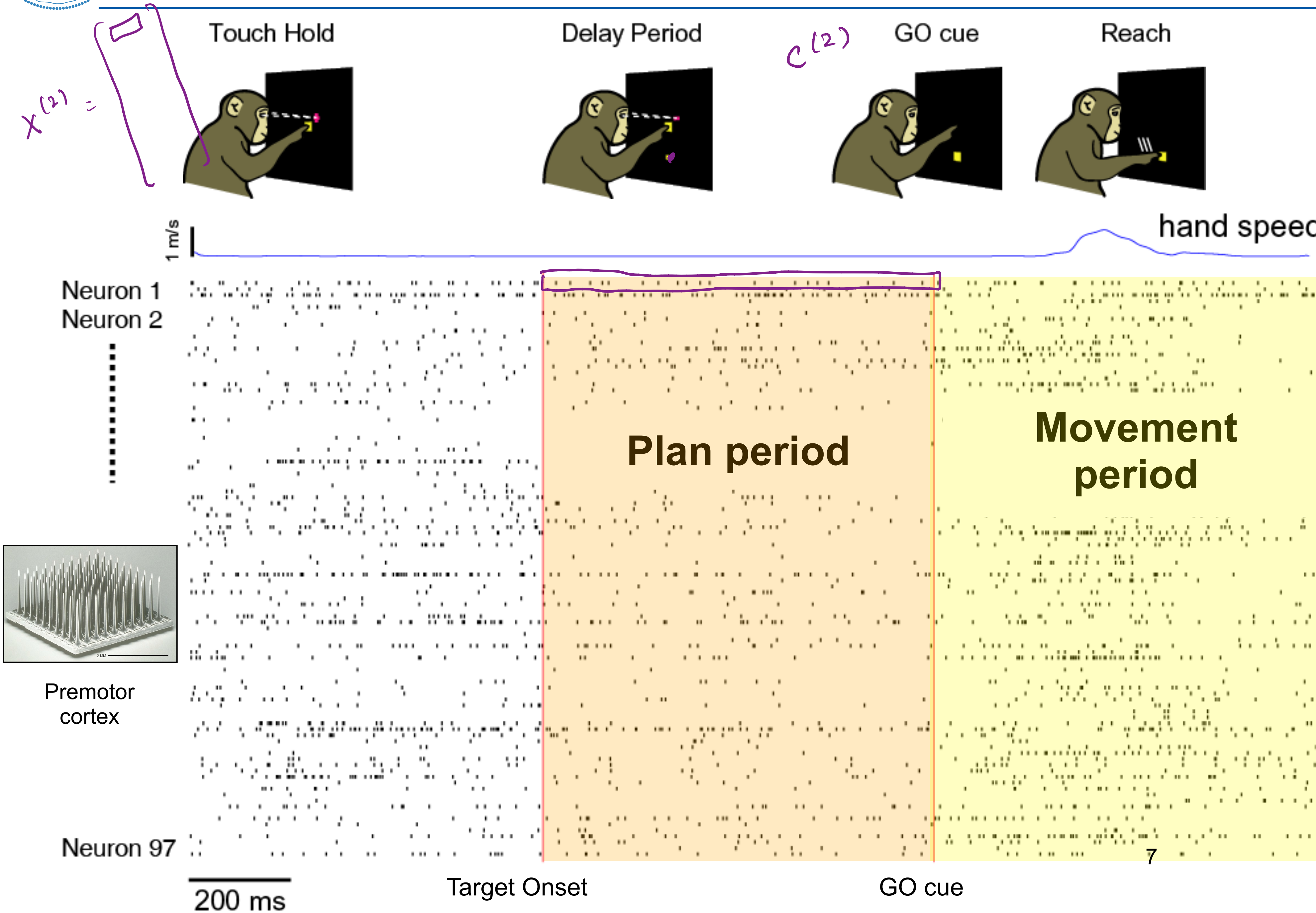
400ms

500ms

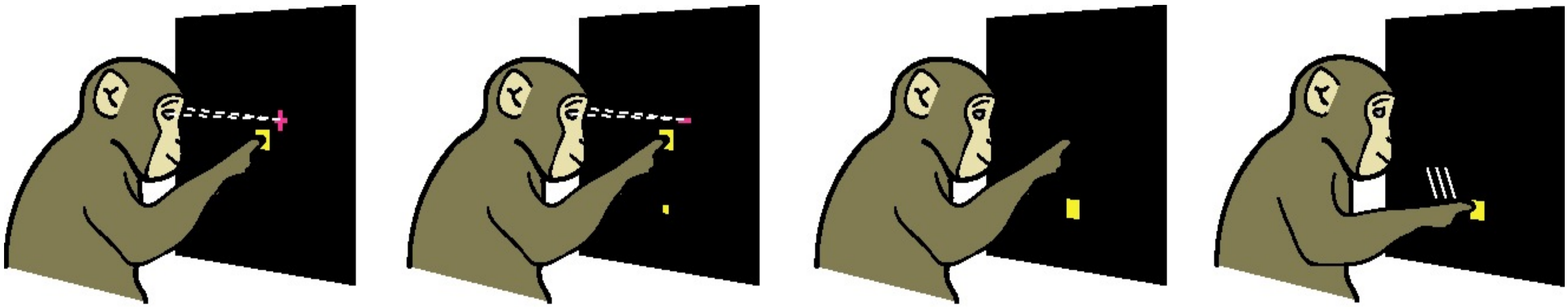
Delay period

time

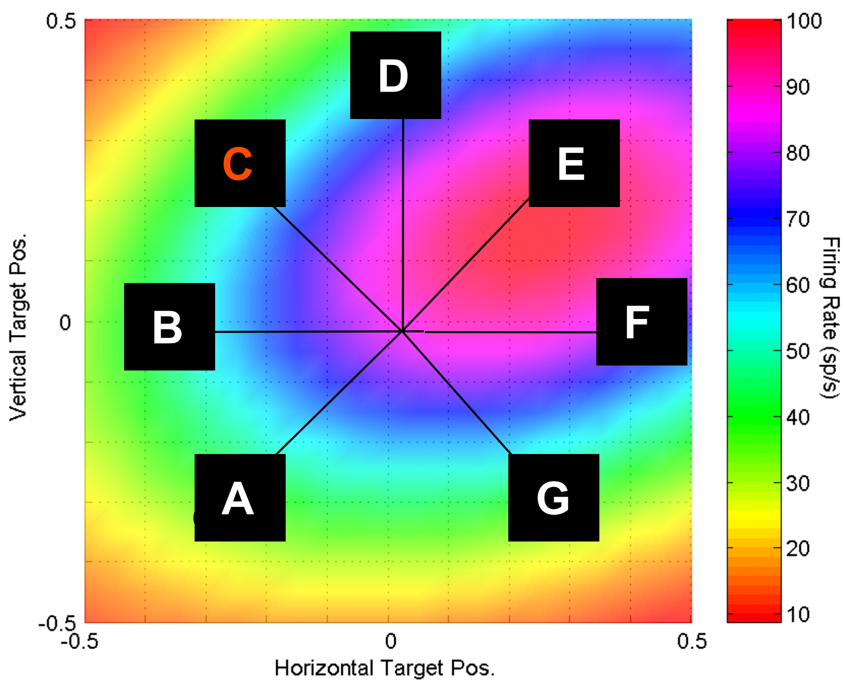
# Delayed Reach Task, Plan & Movement Activity

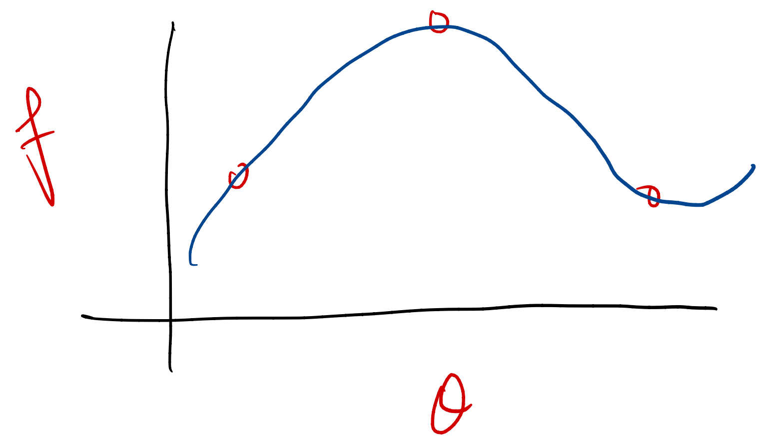


# Plan Activity Reflects Movement Endpoint

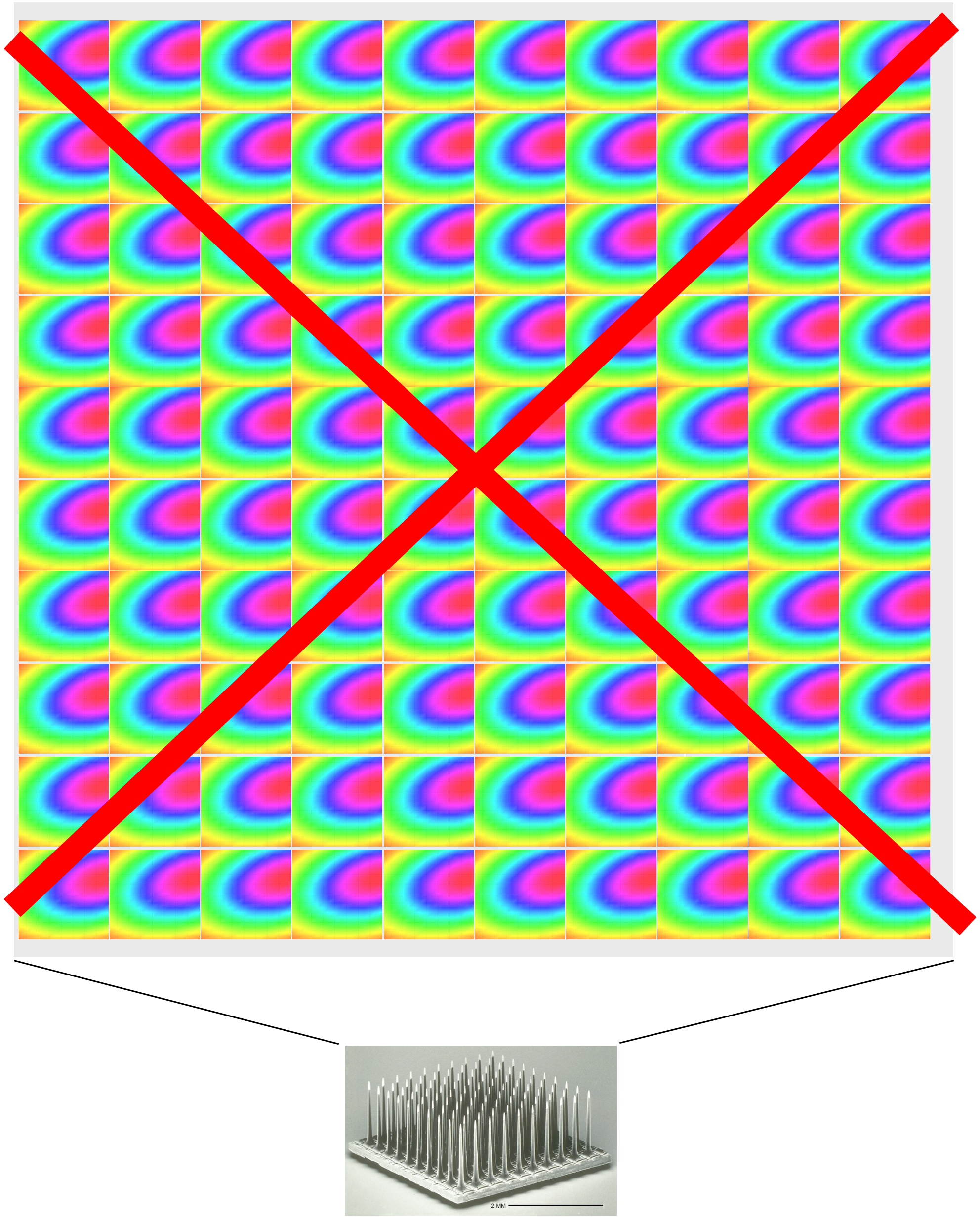


Typical spike rate from 1 neuron

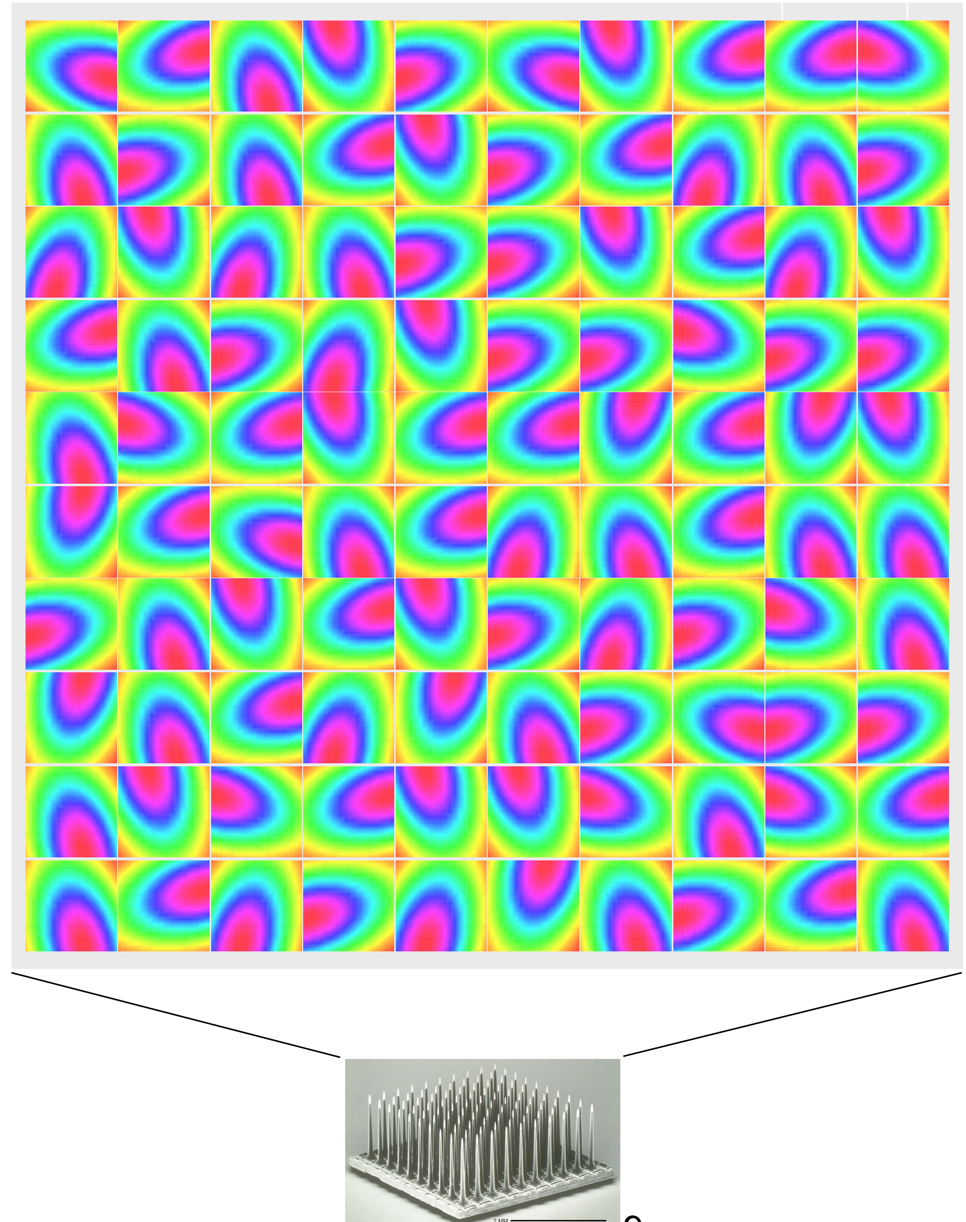


# Each neuron is “tuned” differently

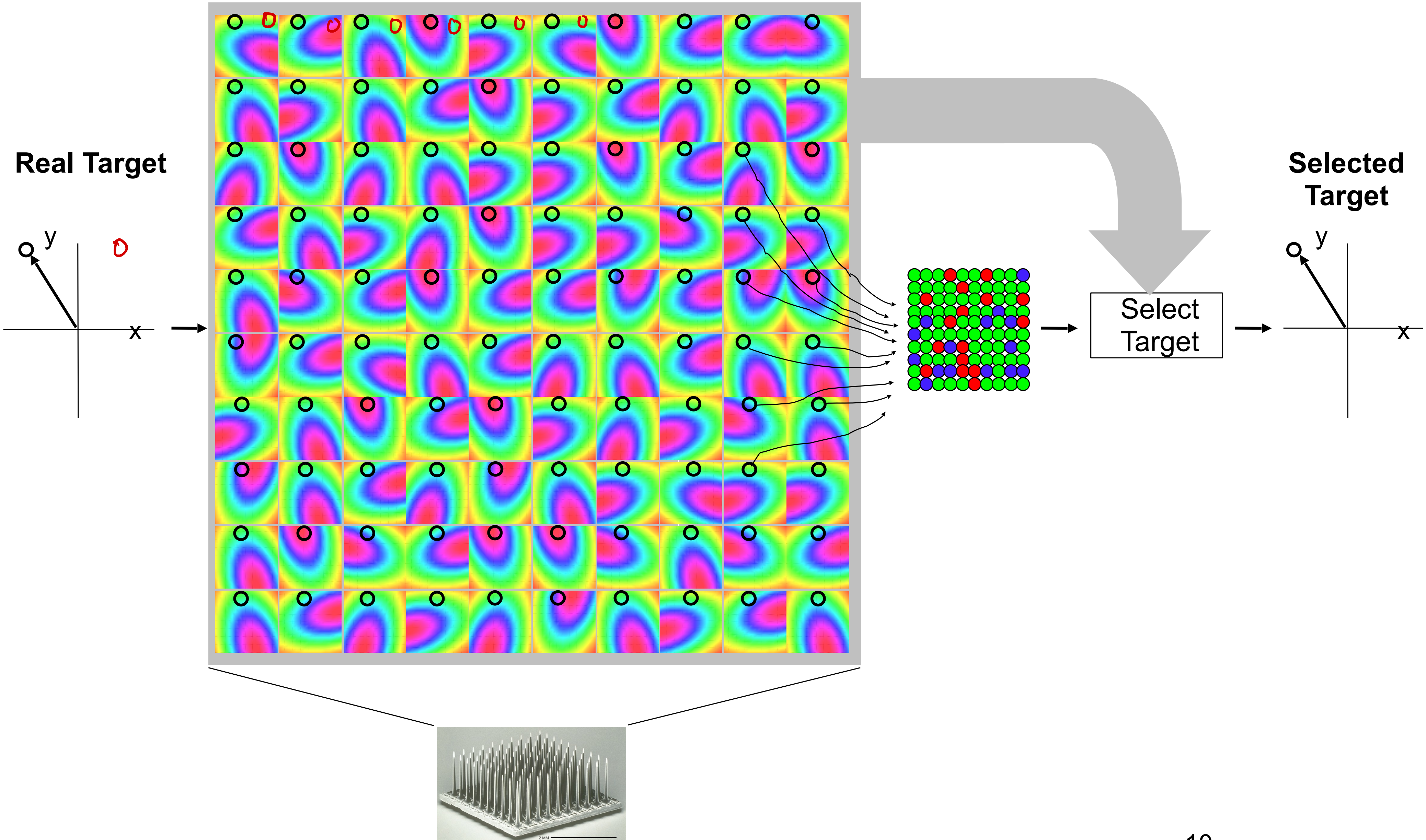


100 electrode array



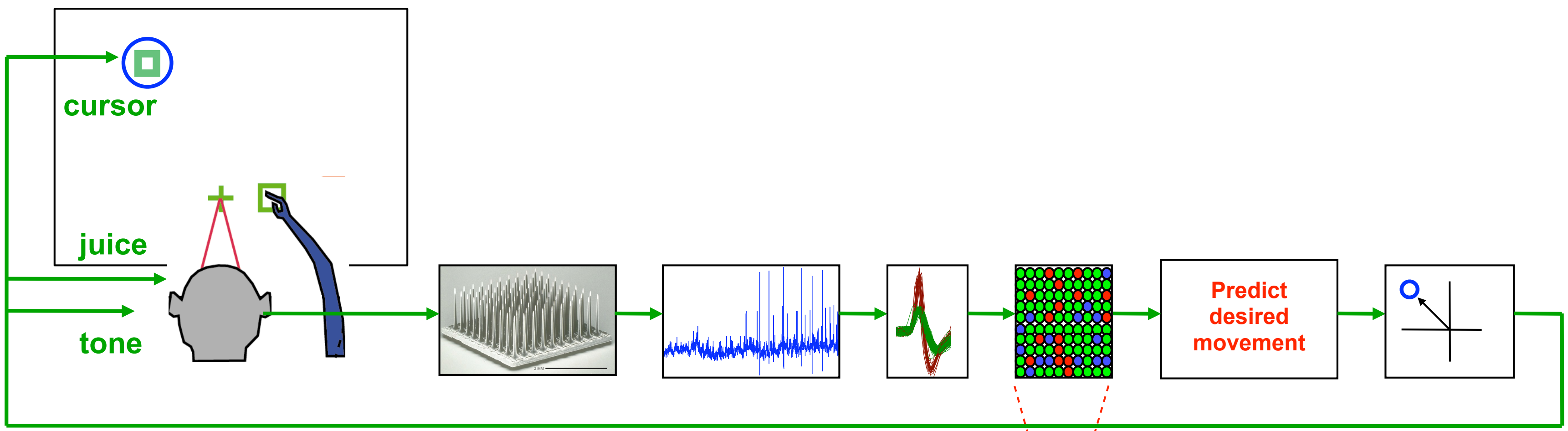
100 electrode array

# How to predict the desired target




100 electrode array

# End-to-End System Performance Measurements

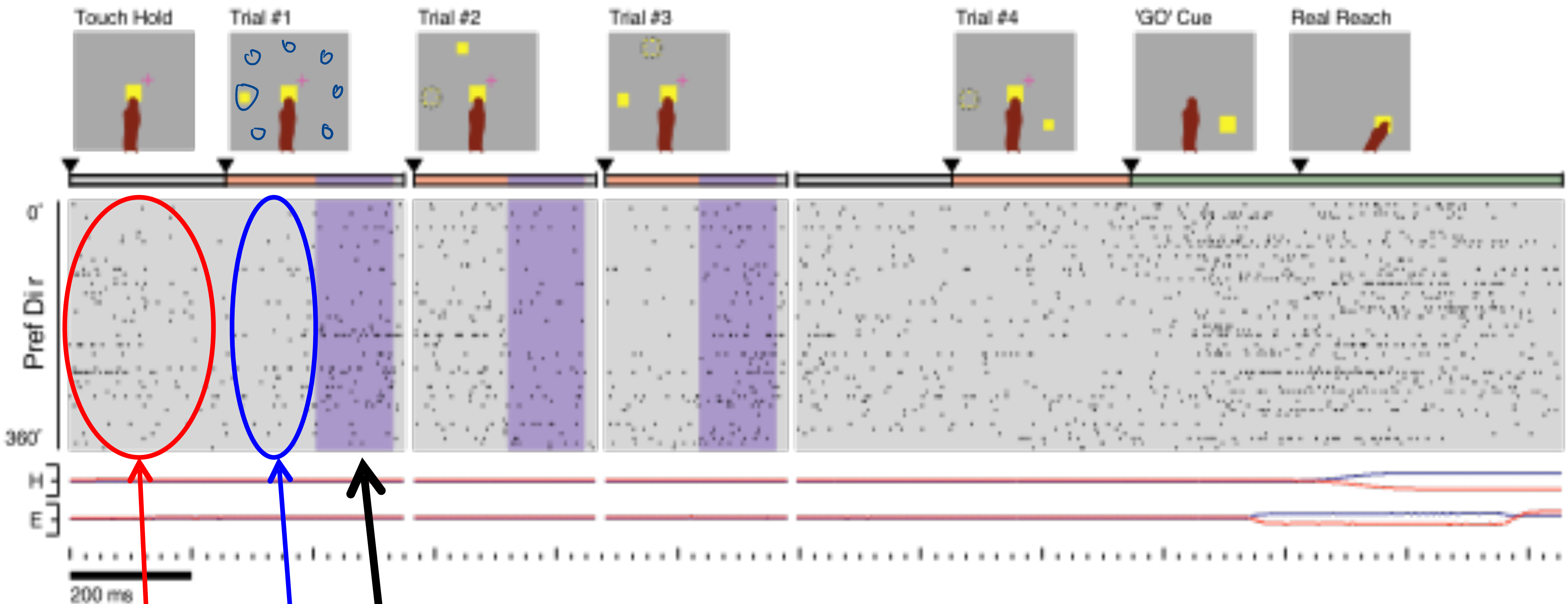




*If there is "magic", here it is...*


Santhanam et al. *J Neurophysiol* 2009
Cunningham et al., *J Neurophysiol* 2008
Kemere et al., *J Neurophysiol* 2008
Batista... Shenoy, *IEEE TNSRE* 2007
Yu et al., *J Neurophysiol* 2007
Achtman et al., *J Neural Engineering* 2007
Santhanam et al. *Nature* 2006
Kemere et al. *IEEE TBME* 2004
Shenoy et al. *NeuroReport* 2003

# Task timeline



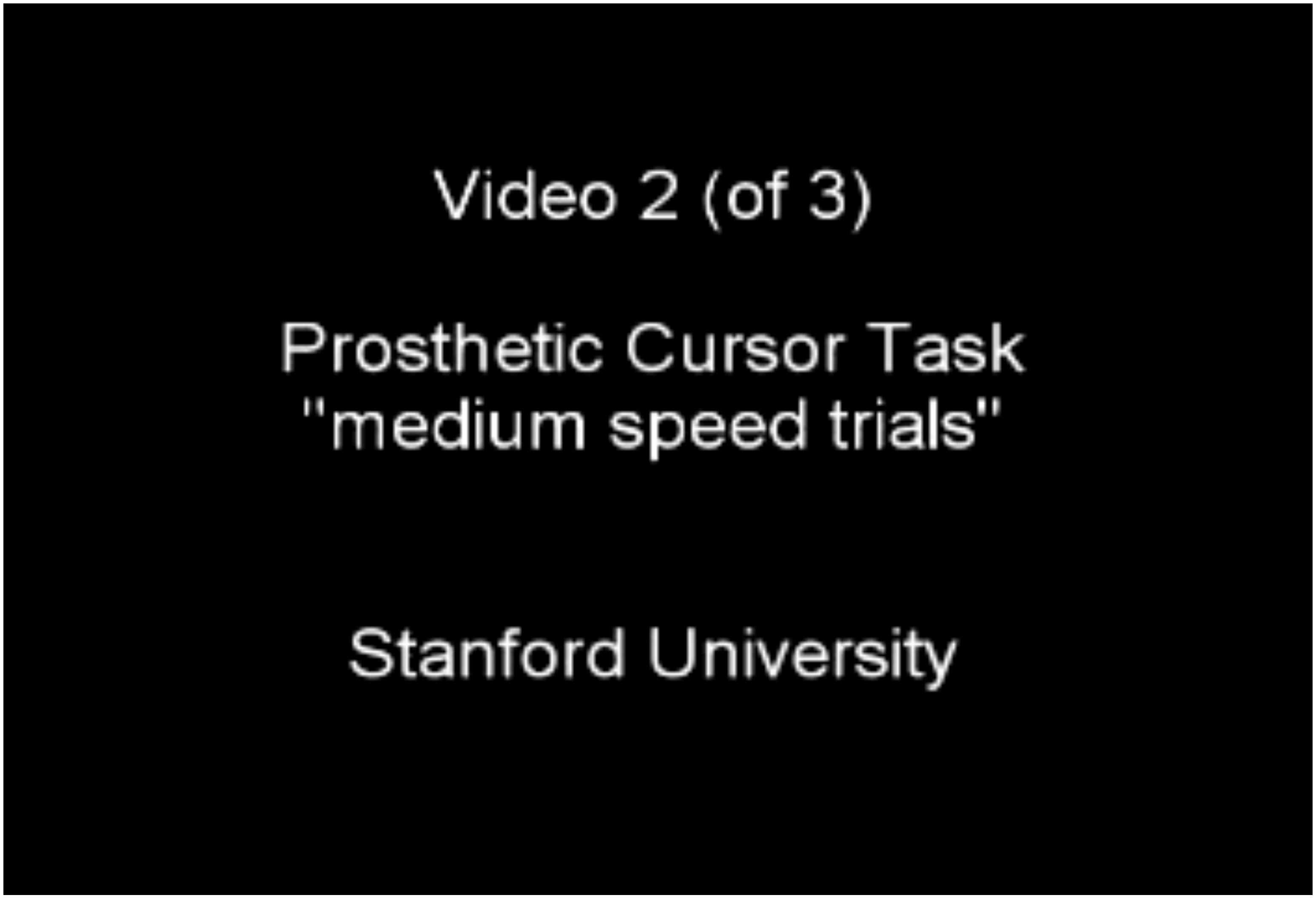

Classify this **plan** activity to one of 8 targets

Time required for target information to arrive in premotor cortex (due to action potential propagation, visual processing)

**Baseline** neural activity

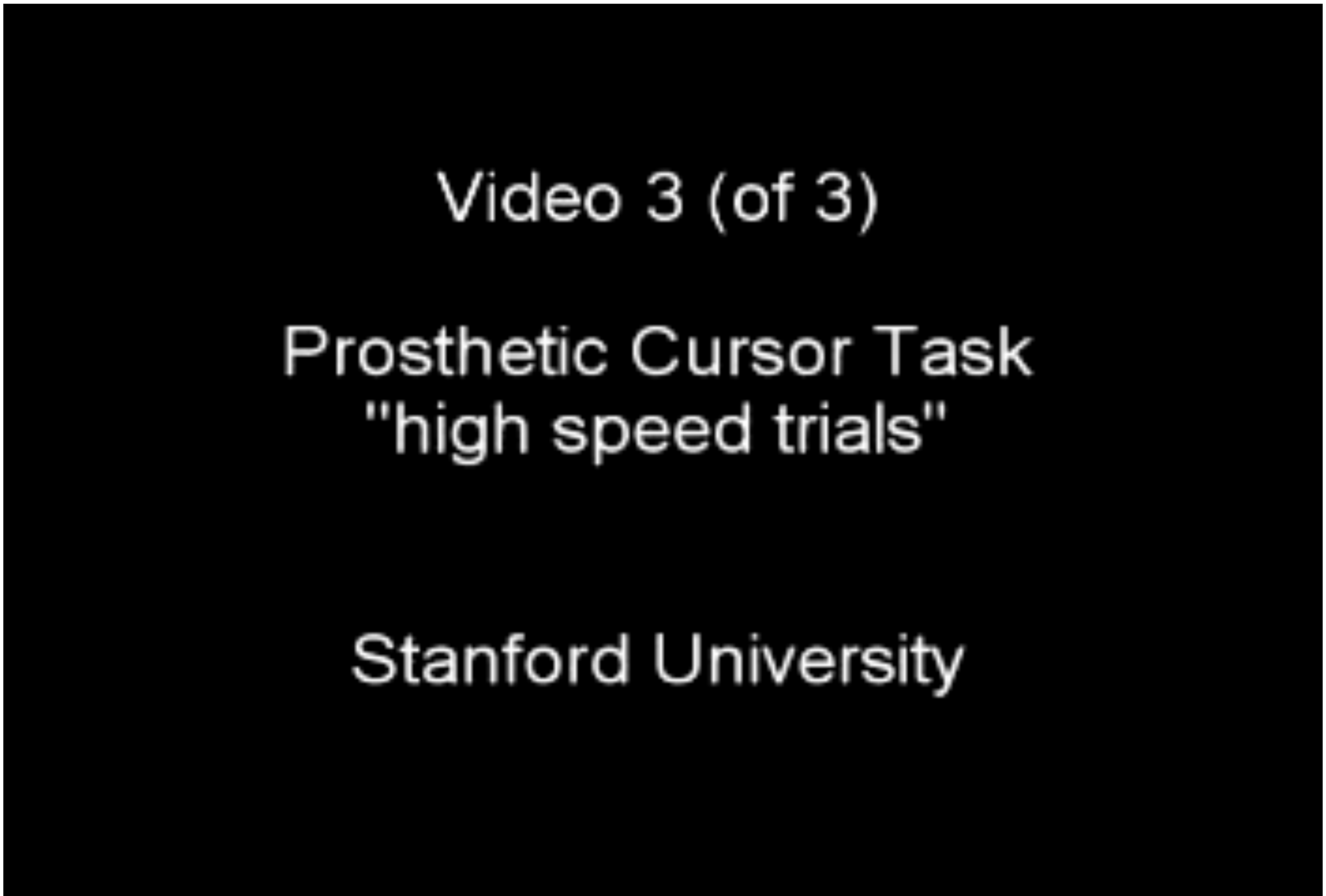
# Brain-controlled target selection

**Approximately 2.5 bps (~5 words/min equivalent)**



# High-performance brain-controlled target selection

**Approximately 5.0 bps (~10 words/min equivalent)**






Santhanam, Ryu, Yu, Afshar & Shenoy, *Nature* 2006.

# Homework 4

- You will be implementing such a decoder and applying it to real neural data (planning activity).

- The neural data were recorded from the monkey you see in the video.

# Classification algorithms

• MANY classification algorithms are available.  Examples include:

Discriminant functions
Probabilistic generative models
Probabilistic discriminative models

Neural networks
Gaussian processes
Support vector machines
Relevance vector machines

• They mainly differ in the *cost function* that is optimized to find the decision boundary.

# Classification algorithms

• It would take many weeks, if not more than half of the course to go through each of these in detail.

• Rather than give you an overview of all possible algorithms, we will go in depth into one class of algorithms (Probabilistic generative models) that is commonly used and has found great success in its application to neural data.

• The classifier used in the prosthetic system I just described is based on a probabilistic generative model.

SEE “CLASSIFICATION” HANDOUT

# Discrete classification:

$\vec{x} \in \mathbb{R}^D \qquad \vec{y} \in \mathbb{R}^D \qquad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_D \end{bmatrix}$

$\rightarrow \quad \alpha(\vec{x}) = \vec{y}^T \vec{x} = x_1 y_1 + x_2 y_2 + \dots + x_D y_D$

scalar $\in \mathbb{R}$

$$\frac{\partial \alpha(\vec{x})}{\partial \vec{x}} = \begin{bmatrix} \frac{\partial \alpha(\vec{x})}{\partial x_1} \\ \frac{\partial \alpha(\vec{x})}{\partial x_2} \\ \vdots \\ \frac{\partial \alpha(\vec{x})}{\partial x_D} \end{bmatrix} \in \mathbb{R}^D$$

$$\frac{\partial \alpha(\vec{x})}{\partial \vec{x}} = \frac{\partial (\vec{y}^T \vec{x})}{\partial \vec{x}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_D \end{bmatrix} = \vec{y}$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & & \dots & a_{mn} \end{bmatrix}$$

$A \in \mathbb{R}^{m \times n}$

$\alpha = x^T A y \quad \in \mathbb{R}$

$$\frac{\partial \alpha}{\partial A} = \begin{bmatrix} \frac{\partial \alpha}{\partial a_{11}} & \frac{\partial \alpha}{\partial a_{12}} & \dots & \frac{\partial \alpha}{\partial a_{1n}} \\ \frac{\partial \alpha}{\partial a_{21}} & & \ddots & \vdots \\ \vdots & & & \frac{\partial \alpha}{\partial a_{mn}} \\ \frac{\partial \alpha}{\partial a_{m1}} & & \dots & \end{bmatrix}$$

$$\alpha(\vec{x}) = \vec{x}^T A \vec{x} \qquad \vec{x} \in \mathbb{R}^D \qquad A \in \mathbb{R}^{D \times D}$$

$$\frac{\partial \alpha(\vec{x})}{\partial \vec{x}} \in \mathbb{R}^D$$

$$x^T A x = \sum_{i=1}^{D} \sum_{j=1}^{D} x_i a_{ij} x_j$$

$i=1,\ j=1 \qquad x_1 \cdot a_{11} \cdot x_1 = a_{11} \cdot x_1^2$

$\rightarrow i=1,\ j \neq 1, \quad x_1 a_{12} x_2 + x_1 a_{13} x_3 + \dots + x_1 a_{1D} x_D = \sum_{j=2}^{D} x_1 a_{1j} x_j$

$\rightarrow j=1,\ i \neq 1 \quad = \sum_{i=2}^{D} x_i a_{i1} x_1$

$$\frac{\partial x^T A x}{\partial x_1} = 2 a_{11} x_1 + \sum_{j=2}^{D} a_{1j} x_j + \sum_{i=2}^{D} a_{i1} x_i$$

$$= \sum_{j=1}^{D} a_{1j} x_j + \sum_{i=1}^{D} a_{i1} x_i = (Ax)_1 + (A^T x)_1$$

$$\boxed{\frac{\partial x^T A x}{\partial x_i} = (Ax)_i + (A^T x)_i}$$

$$\begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \dots \\ a_{21}x_1 + a_{22}x_2 + \dots \\ \\ \\ \end{bmatrix} = \underset{\in \mathbb{R}^D}{Ax} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1D} \\ a_{21} & & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{D1} & \dots & & a_{DD} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix}$$

$$\frac{\partial x^T A x}{\partial x} = \begin{bmatrix} (Ax)_1 + (A^T x)_1 \\ (Ax)_2 + (A^T x)_2 \\ \vdots \\ (Ax)_D + (A^T x)_D \end{bmatrix}$$

$$= Ax + A^T x$$

$$\boxed{\frac{\partial x^T A x}{\partial x} = (A + A^T)\, x}$$

$\mathbb{R}^D$

$A$: $(D \times D)$, $A^T$: $(D \times D)$, $x$: $D \times 1$

$(D \times D) \cdot (D \times 1) \Rightarrow D \times 1$

What if $D = 1$? $A$ is a scalar, $x$ is a scalar

$$\alpha(x) = x^T A x = x \cdot A \cdot x = Ax^2$$

$$\frac{\partial \alpha(x)}{\partial x} = 2Ax$$

$A, B, C, D, E$

$$Tr(A) = \sum_{i=1}^{D} a_{ii}$$

$$Tr(ABCDE) = Tr(BCDEA) = Tr(CDEAB) = Tr(DEABC) = Tr(EABCD)$$

**Goal:** $\vec{x}$ $\Longrightarrow$ classify which 1 of $\mathbb{K}$ targets (classes) the monkey is planning to reach to.

$\downarrow$

Firing rate of D neurons

**TRAIN:** learn a $\boxed{\text{model}}$ that performs this classification | $f(\vec{x}) \rightarrow$ class

$$(\vec{x}^{(i)}, c^{(i)})$$

$i$ indicates the $i$th trial

$\vec{x}^{(i)} \rightarrow \mathbb{R}^{96}$

$c^{(i)} \rightarrow$ scalar

$i = 1, \ldots, N$

N total trials

$i = 11$

$x^{(i)}$ "PMd neurons' f.r.s when planning to the right target $c^{(i)}$"

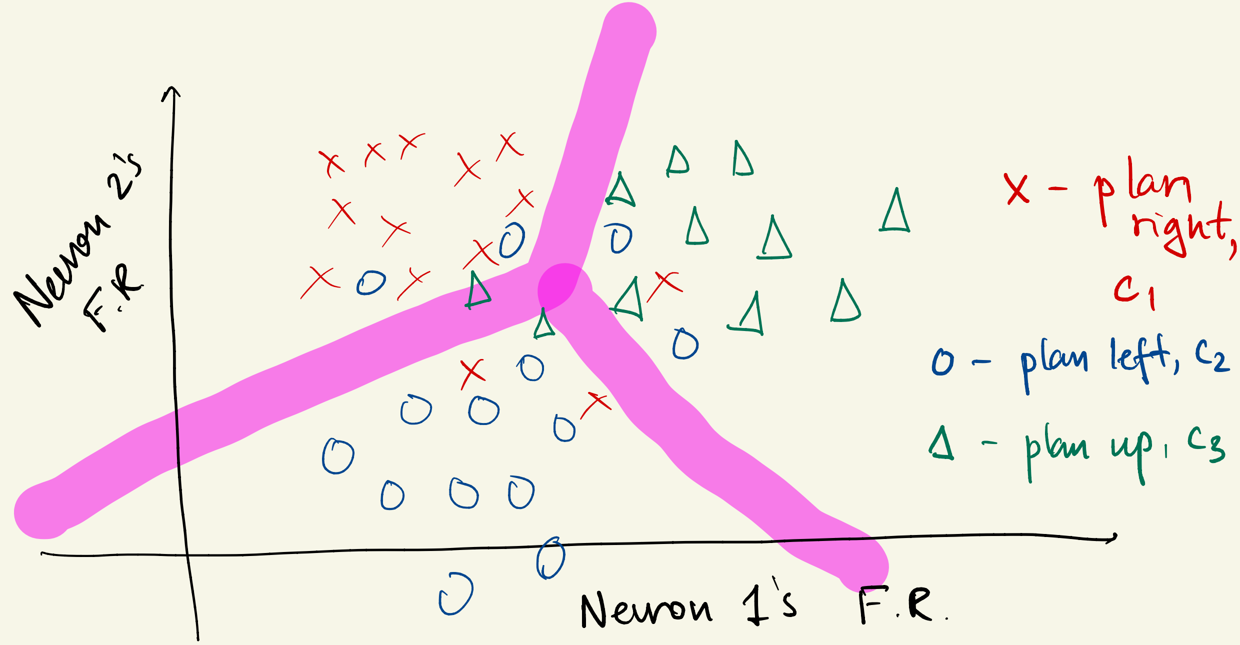
$c^{(i)}$ : right target

**TEST:** Take the model from training

$x^{(j)} \rightarrow \boxed{\text{Model from training}} \rightarrow c^{(j)}$

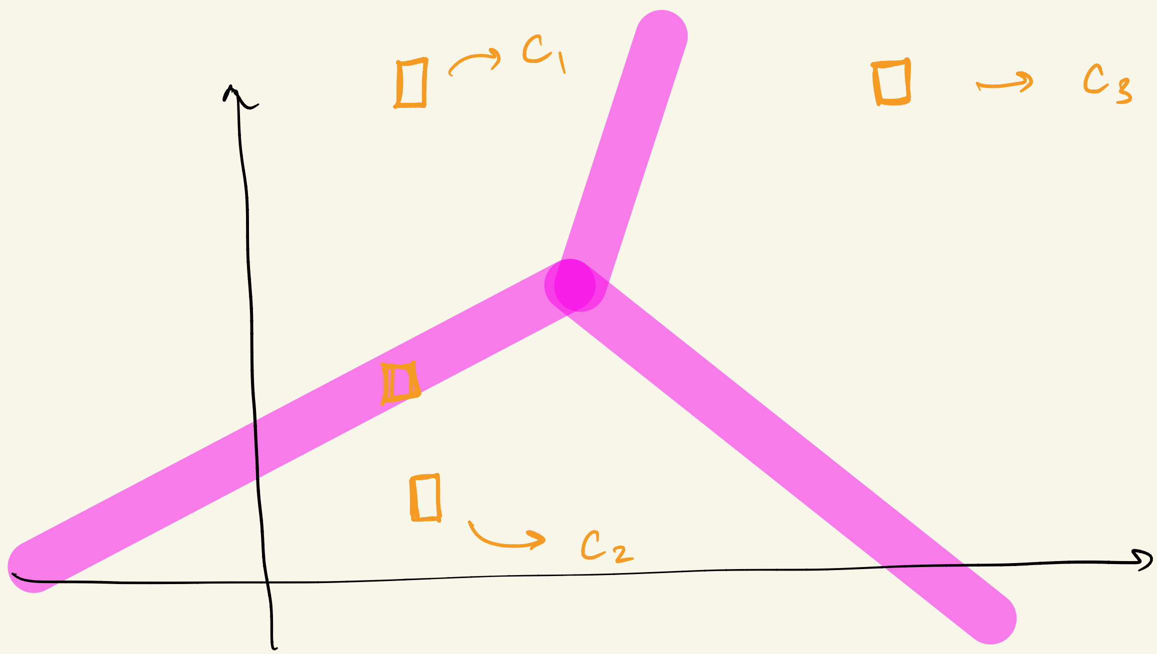
# TRAIN:

D = 2




In training, we learn the boundaries.

# TEST:



# Probabilistic Generative Models (PGMs)

$x^{(i)}$ : firing rate $\in \mathbb{R}^D$ for trial $i$

$c^{(i)}$ : $c_k$, class of trial $i$ $\quad k = \{1, 2, \dots, K\}$

**TRAIN**: $p(\vec{x} \mid c_k)$

We will have $K$ of these distributions

↳ continuous distribution.

**TEST**: Given $x^{(i)}$, we want to know which class $c_k$ it belongs to.

$$p(c_k \mid x^{(i)})$$

$$p(c_1 \mid x^{(i)}) = 0.3$$

$$p(c_2 \mid x^{(i)}) = 0.2$$

$$p(c_3 \mid x^{(i)}) = 0.5$$

**TEST :**

$$\hat{k} = \arg\max_k \; p(c_k \mid \vec{x})$$

$$= \arg\max_k \; \frac{p(c_k)\, p(\vec{x} \mid c_k)}{p(\vec{x})}$$

$$= \arg\max_k \; p(c_k)\, p(\vec{x} \mid c_k)$$

$$p(c_k) = \frac{1}{K} \qquad \text{(uniform prior)}$$

$$= \arg\max_k \; p(\vec{x} \mid c_k)$$