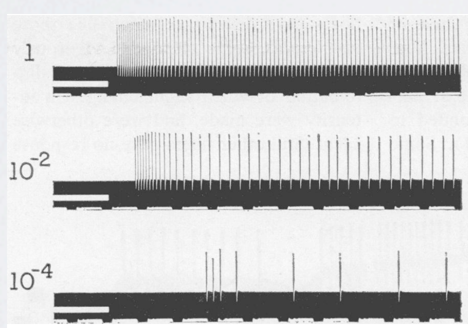
# STATISTICAL THINKING IN SPIKE TRAIN ANALYSIS (LECTURE 3)

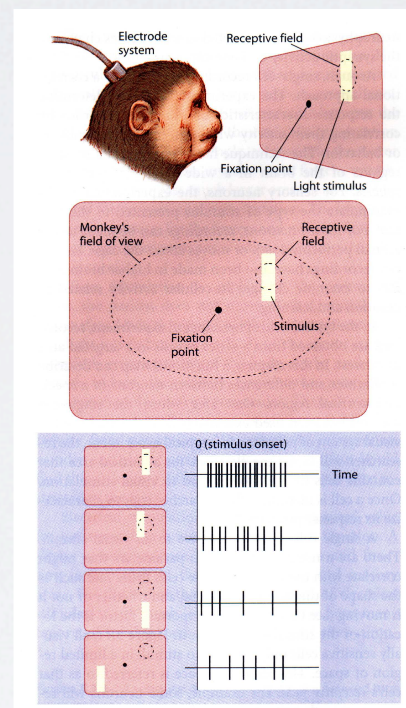

**Rob Kass**

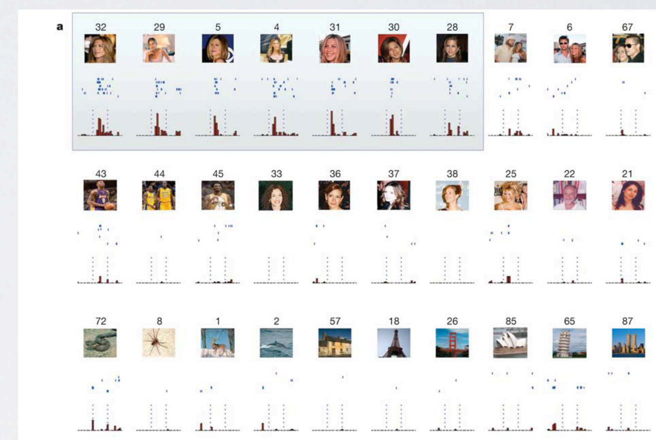
Department of Statistics
Machine Learning Department
Center for the Neural Basis of Cognition
Carnegie Mellon University

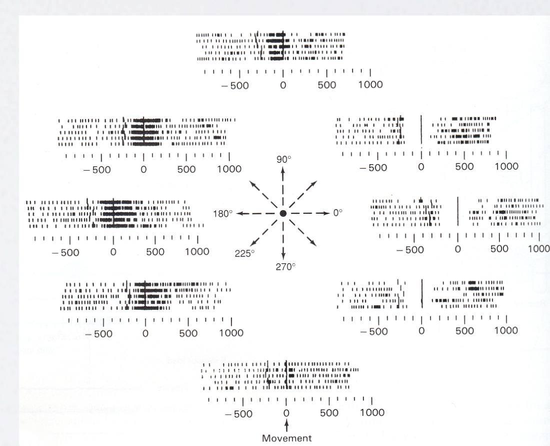
# Lecture 3. First, recap of lecture 1:

- problem of neural coding: to elucidate “the representation and transmission of information in the nervous system”
- firing rate:





# Lecture 2. First, recap:

- problem of neural coding: to elucidate "the representation and transmission of information in the nervous system"
- firing rate:

$$\frac{\text{number of spikes}}{\Delta t}$$

# Lecture 2. First, recap:

• problem of neural coding: to elucidate "the representation and transmission of information in the nervous system"

• firing rate:

$$\frac{\text{number of spikes}}{\Delta t}$$

$$\textit{Theoretical} \text{ instantaneous firing rate } \lambda(t) \text{ (or } \lambda(t|H_t))$$

## homogeneous Poisson:

$$P(\text{event in } (t, t+dt]) = \lambda dt.$$

## inhomogeneous Poisson:

$$P(\text{event in } (t, t+dt]) = \lambda(t) dt.$$

## general:

$$P(\text{event in } (t, t+dt] | H_t) = \lambda(t|H_t) dt,$$

$$(\text{for Poisson: } \lambda(t|H_t) = \lambda(t))$$

# Jacobs et al. experiment (2009, *PNAS*)

Neural coding question: Are spike counts enough to decode stimulus? Is marginal firing rate (Poisson process) enough? Or is relative spike timing needed (non-Poisson process)?

Idea: compare retinal output *in vitro* to animal behavior

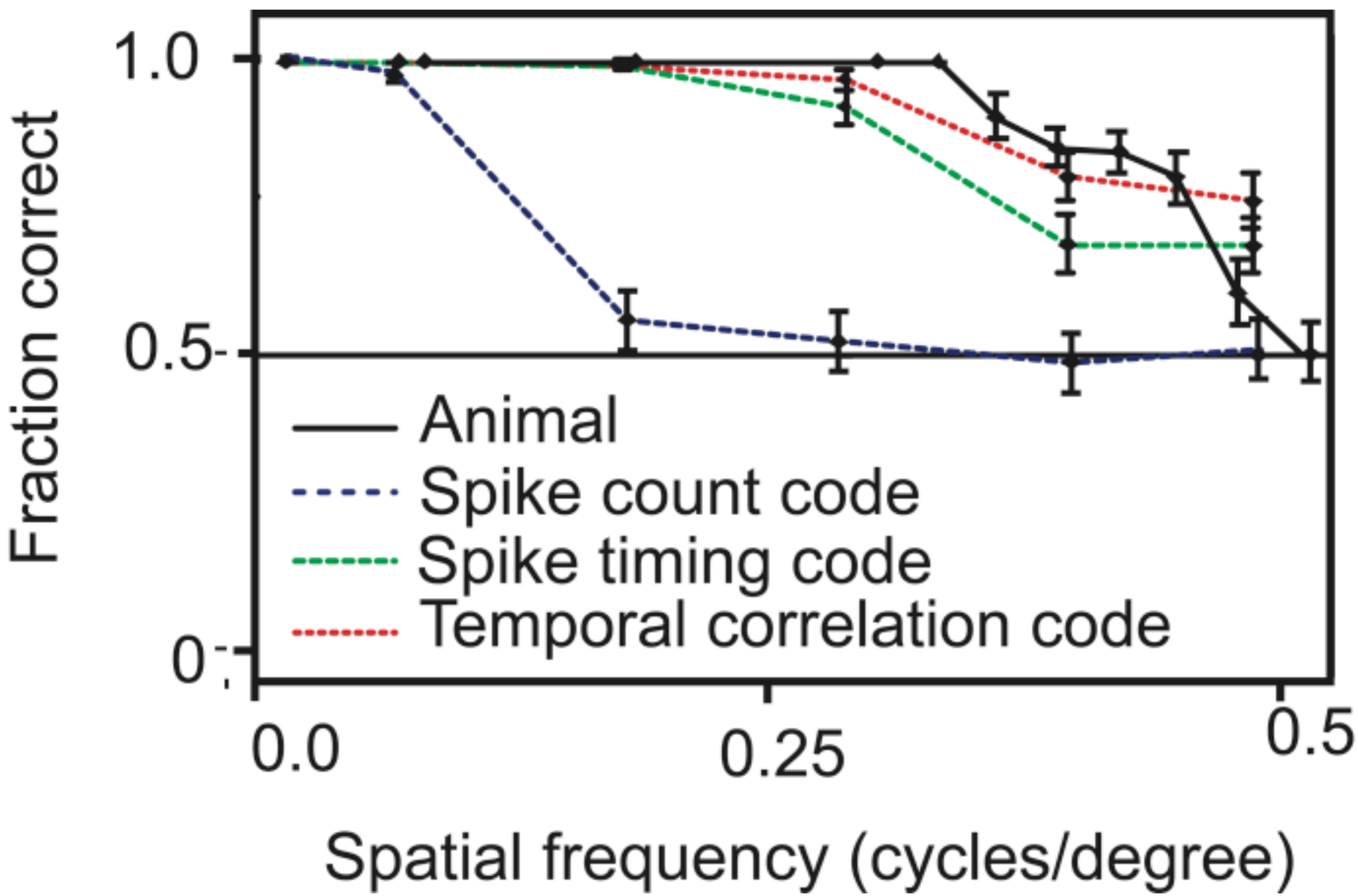
--> examine stimulus-related information contained in retinal spike trains, see how it accounts for behavior

Logic:

1. If all relevant retinal output is recorded, then all stimulus-related information must be contained in the retinal spike trains.

2. If spike counts suffice then behavioral response should be recoverable by a Bayes classifier based on spike counts.

***--> the nervous system could be less efficient than a Bayes classifier but it can't possibly be more efficient***


1.0
0.5
0
Fraction correct
0.0
0.25
0.5
Spatial frequency (cycles/degree)
Animal
Spike count code
Spike timing code
Temporal correlation code

# This time:

- A little more about statistics.
- A few research results.

# This time:

- A little more about statistics.
- A few research results.

- Which features of spike trains are “signal” and which are “noise”?
- Does the PSTH from a single neuron represent well the signal from a population of similar neurons?
- In what ways are population signals carried that are not apparent from responses of individual neurons?
- What time scales are relevant to neural coding?

# This time:

- A little more about statistics.
- A few research results.

    - Which features of spike trains are “signal” and which are “noise”?
    - Does the PSTH from a single neuron represent well the signal from a population of similar neurons?
    - In what ways are population signals carried that are not apparent from responses of individual neurons?
    - What time scales are relevant to neural coding?

- A few high-level (advanced) conceptualizations

$$\text{pdf for random variable } Y: \quad p(y|\theta)$$

$$\text{likelihood function:} \quad L(\theta) = p(y|\theta)$$

***statistical model*** assumes observation arises from some combination of signal and noise

pdf for random variable $Y$: $\quad p(y|\theta)$

likelihood function: $\quad L(\theta) = p(y|\theta)$

***statistical model*** assumes observation arises from some combination of signal and noise

"observation = signal + noise"

$$\text{pdf for random variable } Y: \quad p(y|\theta)$$

$$\text{likelihood function:} \quad L(\theta) = p(y|\theta)$$

***statistical model*** assumes observation arises from some combination of signal and noise

"observation = signal + noise"

get likelihood on parameter vector that characterizes signal and noise (at least in finite-dimensional case)

pdf for random variable $Y$: $\quad p(y|\theta)$

likelihood function: $\quad L(\theta) = p(y|\theta)$

Maximum Likelihood Estimate (MLE):

$$\hat{\theta} = \ \text{argmax } L(\theta)$$

pdf for random variable $Y$: $\quad p(y|\theta)$

likelihood function: $\quad L(\theta) = p(y|\theta)$

Maximum Likelihood Estimate (MLE):

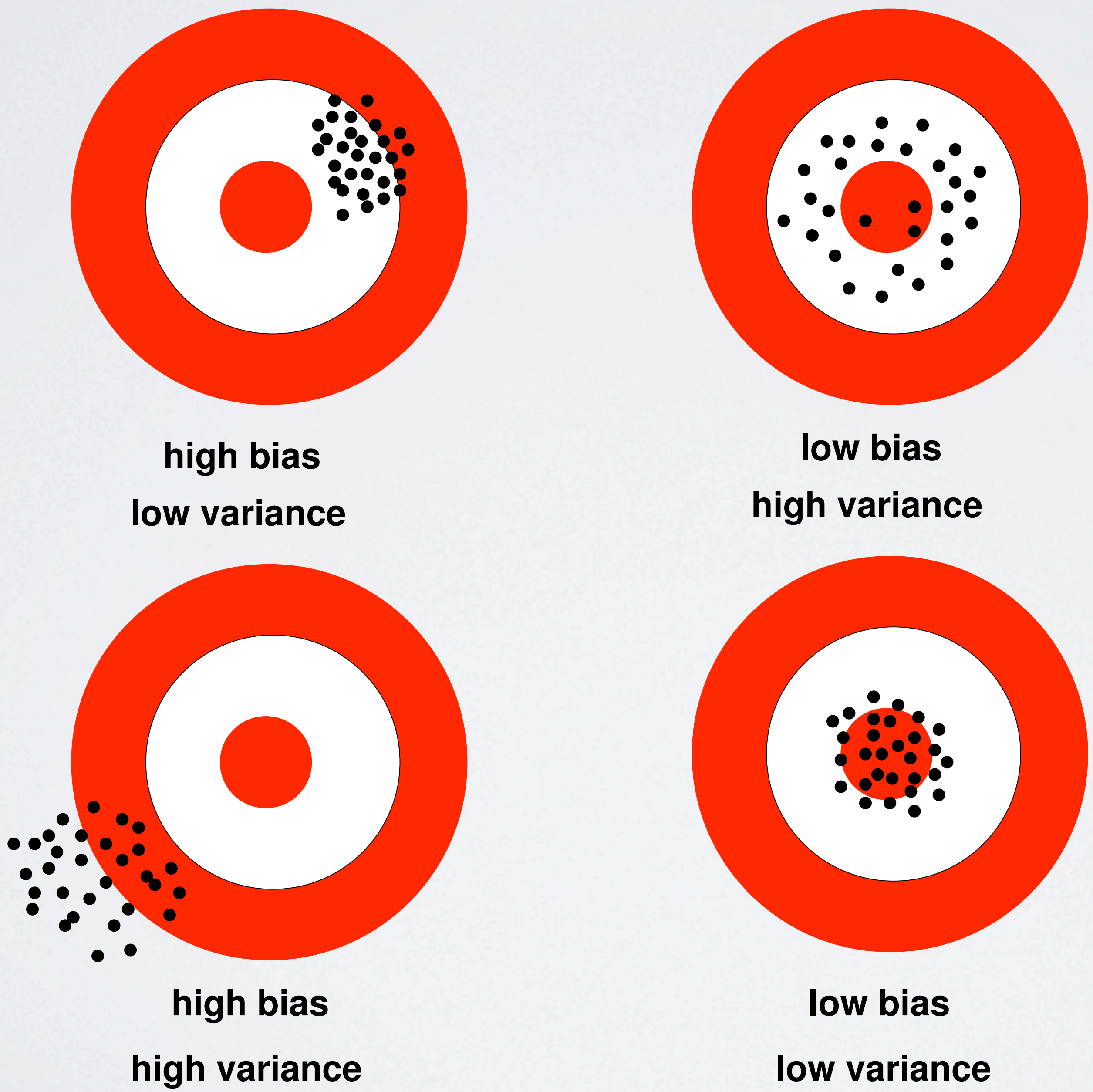
$$\hat{\theta} = \ \text{argmax} \ L(\theta)$$

but how good is this method of estimation?

Properties of ML estimation:

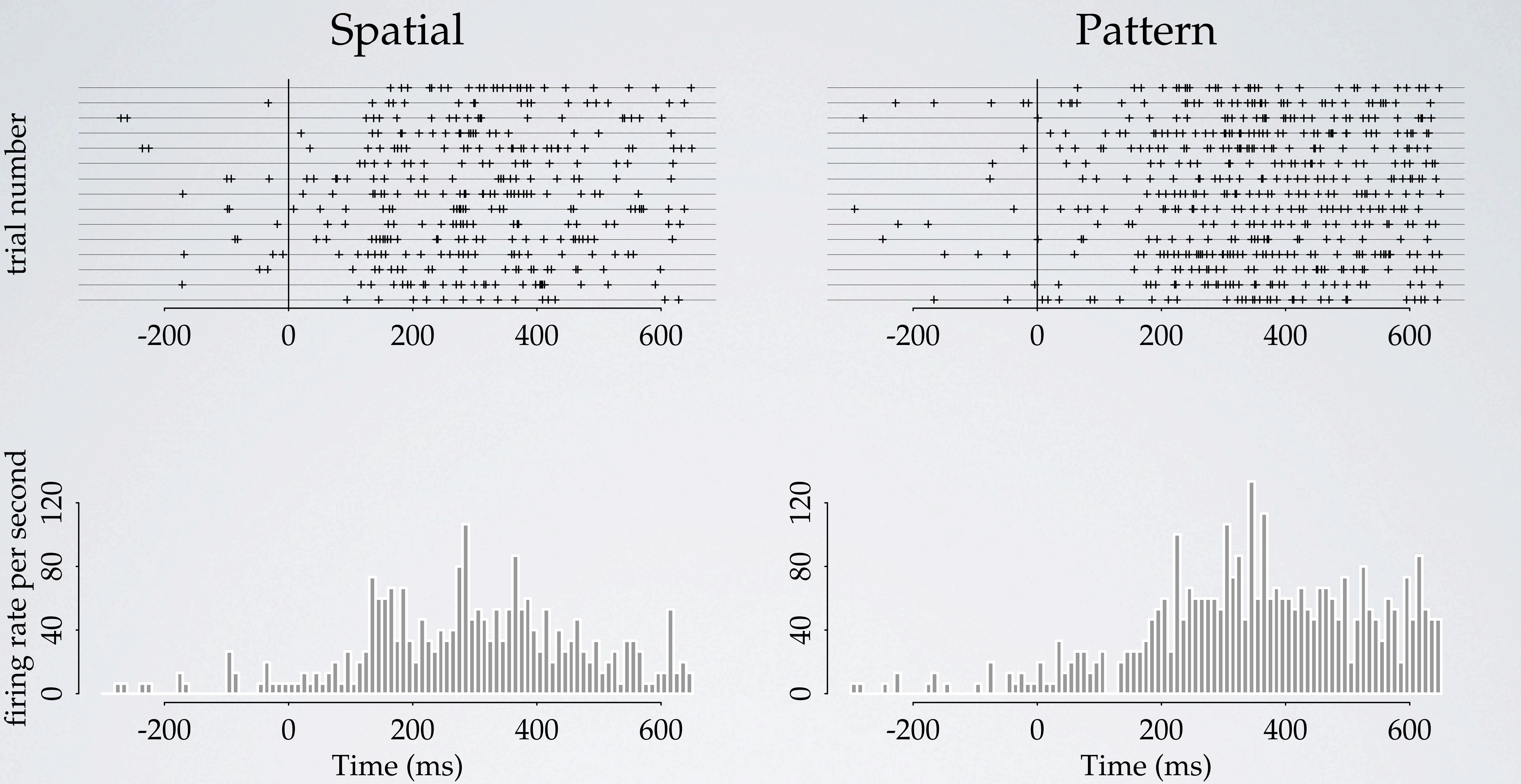
1. In large samples, ML estimation is optimal.
2. In large samples, ML estimation is approximately Bayesian.
3. In large samples, the MLE is approximately normal, and standard errors are generally easy to compute.

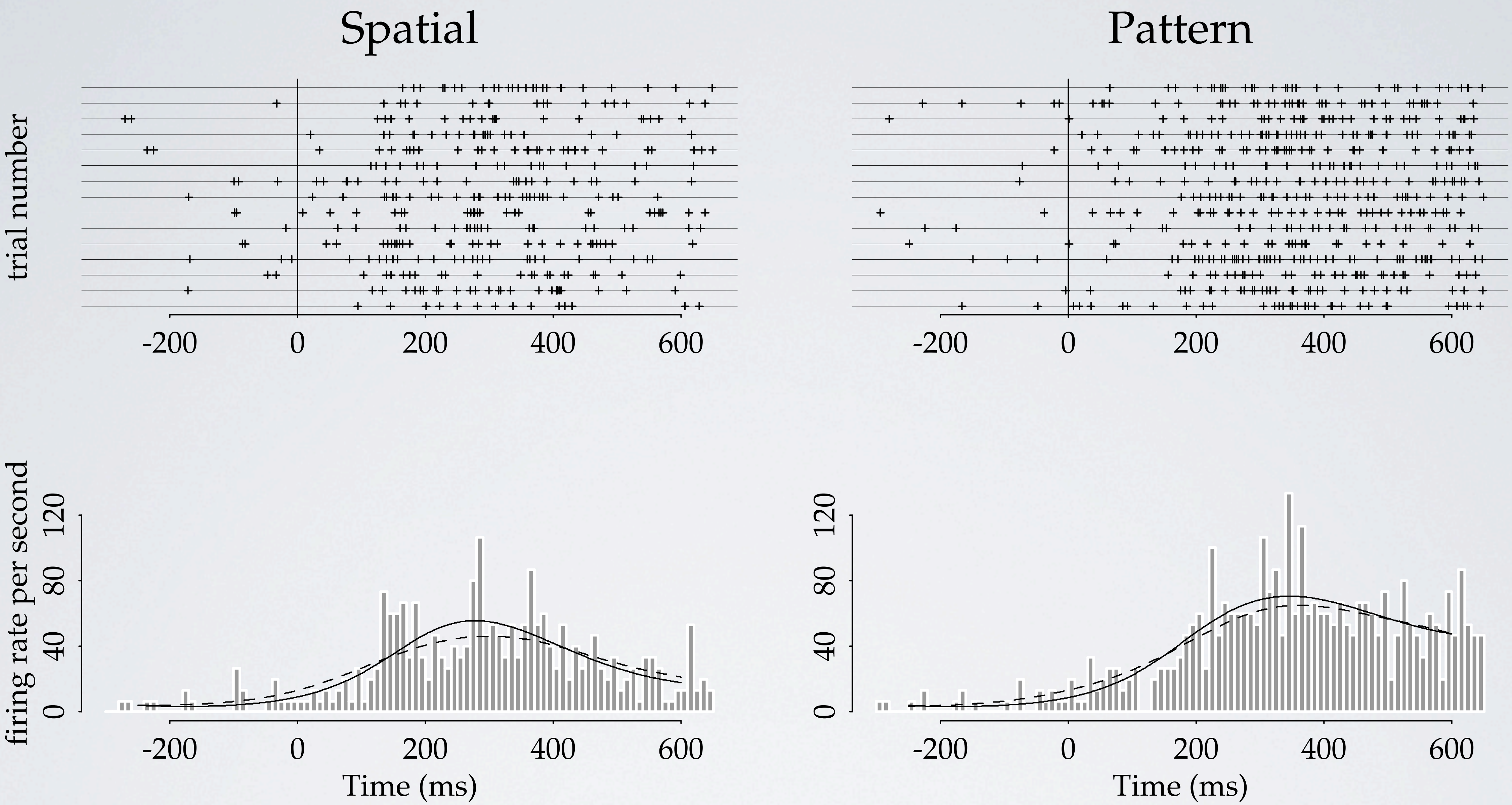
$$MSE(T) = \mathrm{Bias(T)}^2 + \mathrm{Variance(T)}.$$



typically MSE decreases as 1/n
(n is sample size)

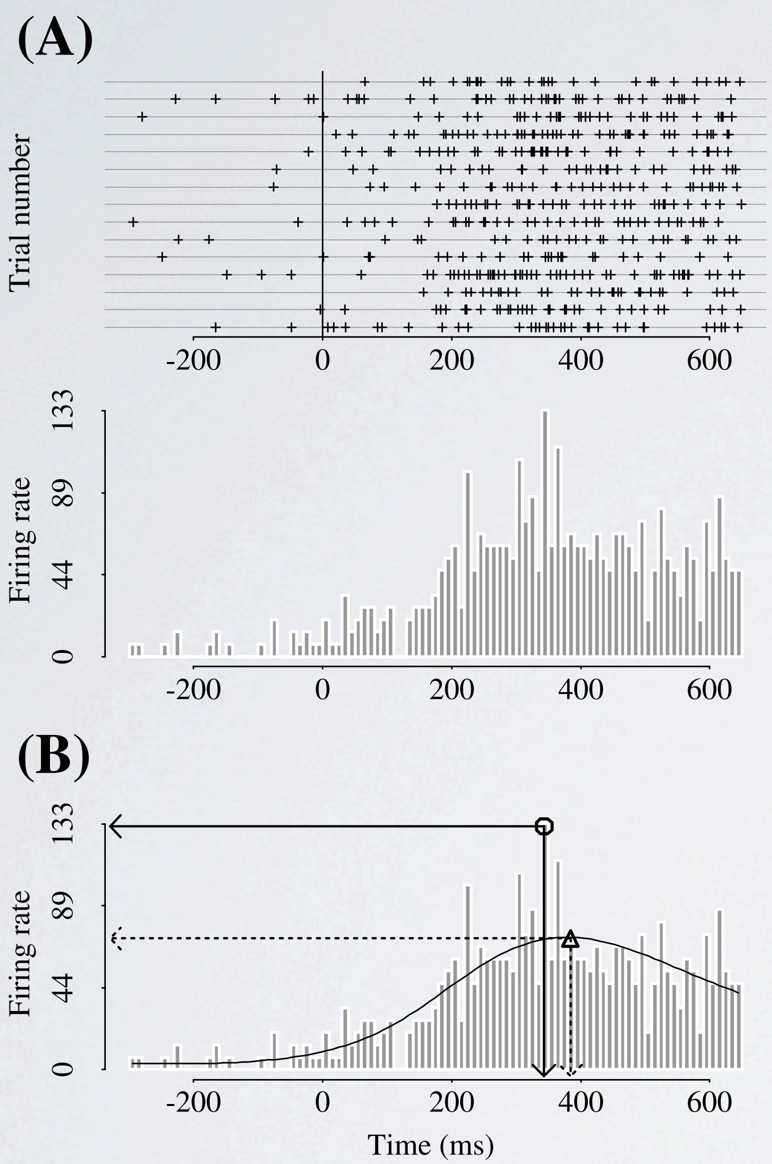
so cutting MSE in half has roughly the same effect as doubling the sample size

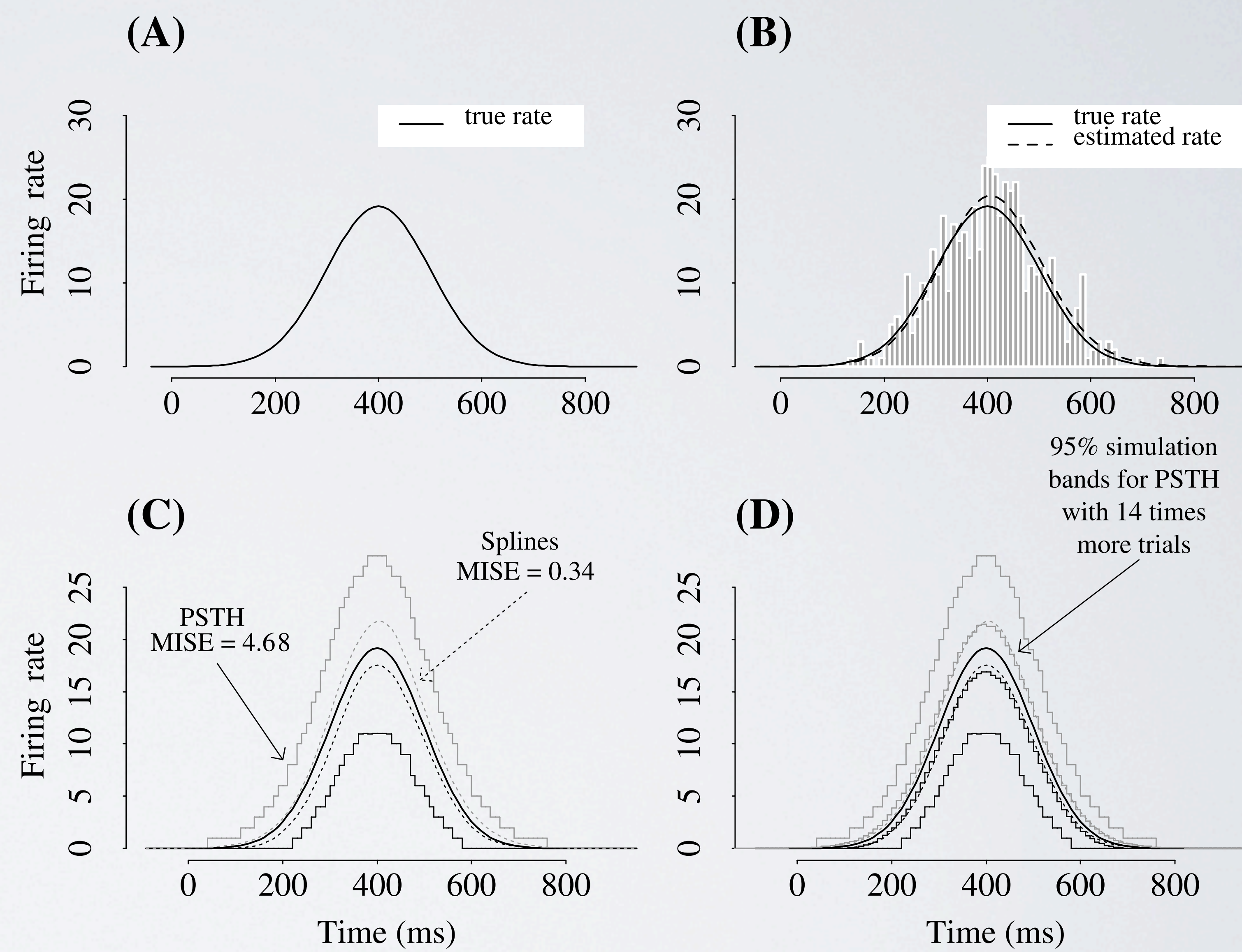

Spatial
Pattern
trial number
-200
0
200
400
600
firing rate per second
0
40
80
120
Time (ms)

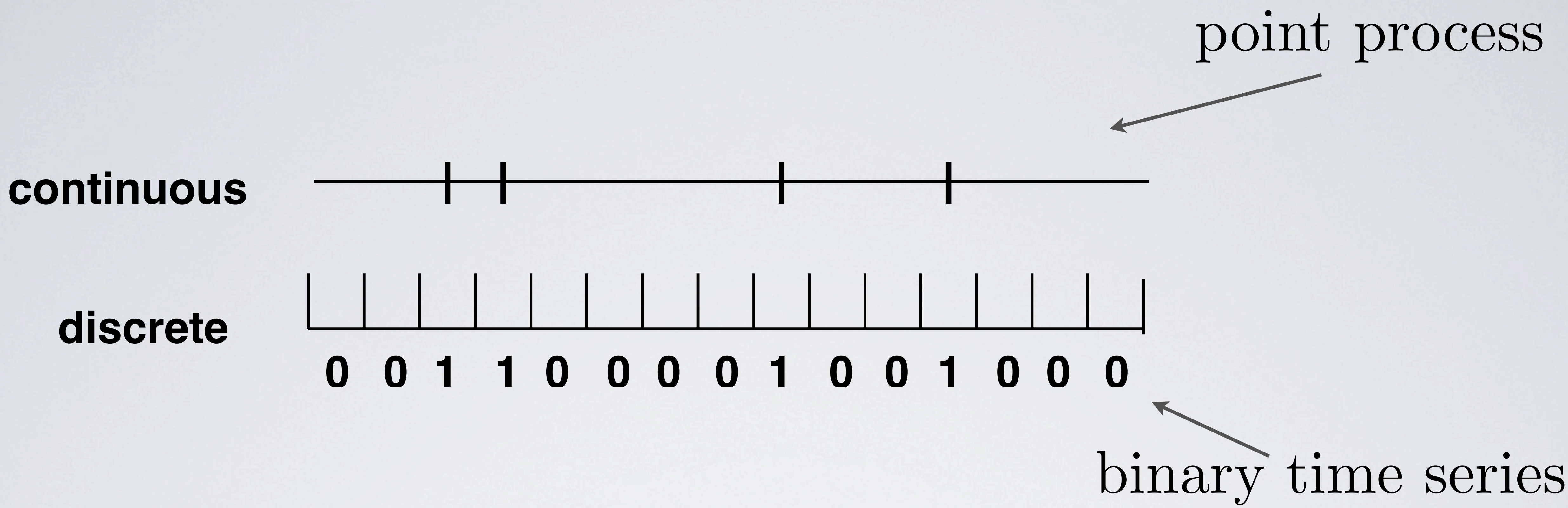


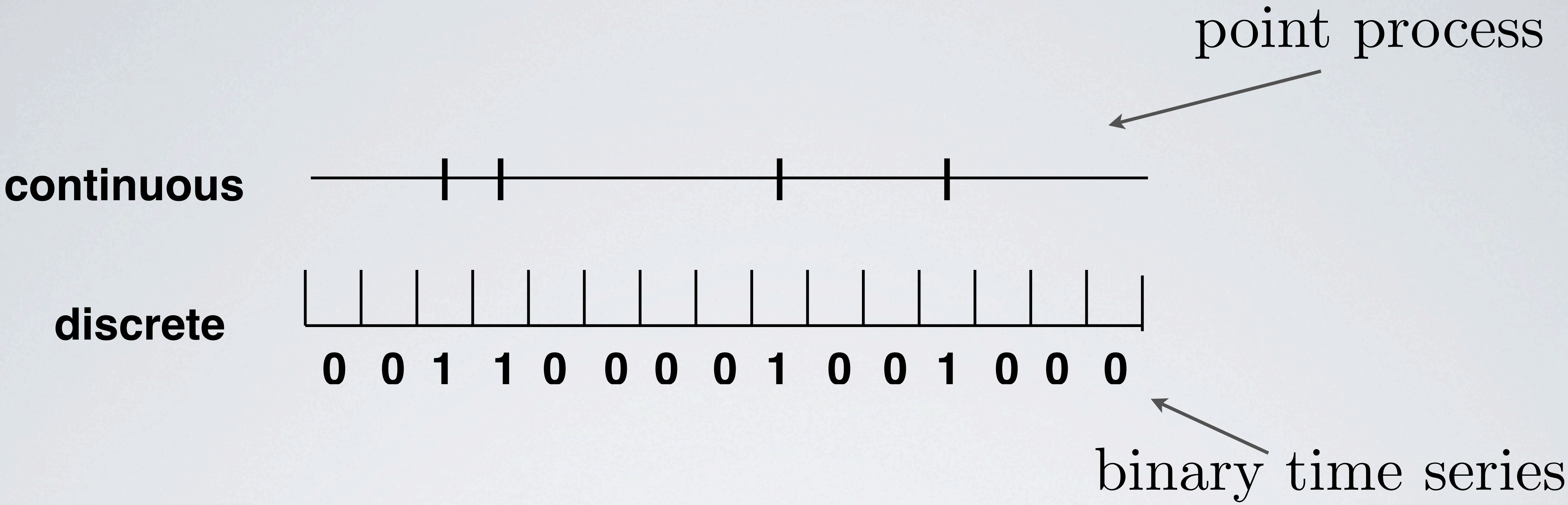
Spatial
Pattern
trial number
firing rate per second
0
40
80
120
-200
0
200
400
600
Time (ms)

# note possible huge efficiency gain from smoothing

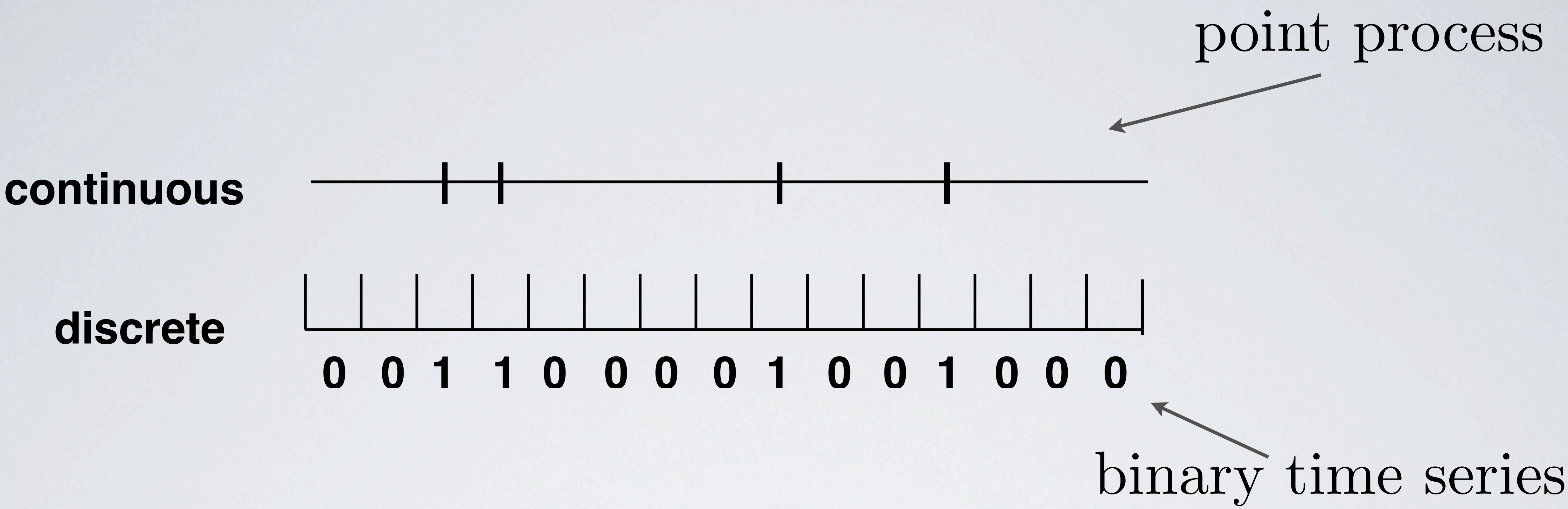



point process
continuous
discrete
0 0 1 1 0 0 0 0 1 0 0 1 0 0 0
binary time series

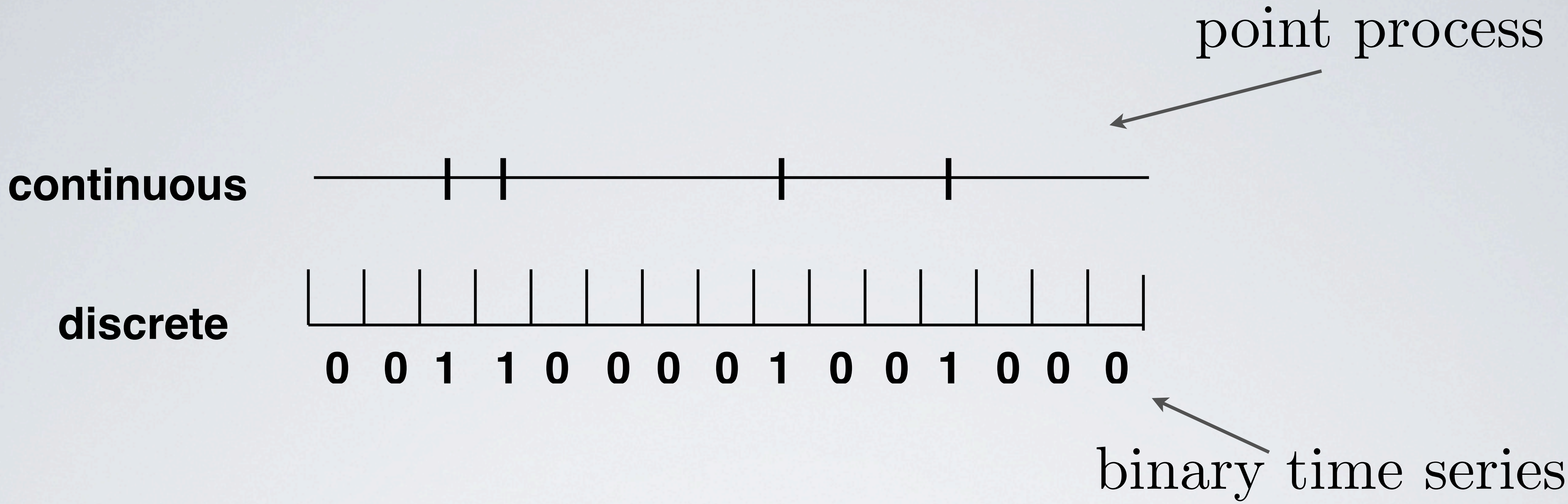



and it is not hard to show that




and it is not hard to show that

$$
\begin{aligned}
\text{point process } L(\theta) &\approx \text{binary time series } L(\theta) \\
&\approx \text{Poisson time series } L(\theta)
\end{aligned}
$$




and it is not hard to show that

$$
\begin{aligned}
\text{point process } L(\theta) &\approx \text{binary time series } L(\theta) \\
&\approx \text{Poisson time series } L(\theta)
\end{aligned}
$$

Holds for both Poisson and non-Poisson processes

# Attenuation of correlation

Selectivity for serial order among neurons from the supplementary eye field (SEF) in a 3-object eye-movement task.

# Attenuation of correlation

Selectivity for serial order among neurons from the supplementary eye field (SEF) in a 3-object eye-movement task.

A monkey moves his eyes to 3 objects in sequence.

Some neurons fire more rapidly for the third object than for the first.

# Attenuation of correlation

Selectivity for serial order among neurons from the supplementary eye field (SEF) in a 3-object eye-movement task.

A monkey moves his eyes to 3 objects in sequence.

Some neurons fire more rapidly for the third object than for the first.

Does this mean the SEF keeps track of sequencing?

Or could it be due to anticipation of reward?

Serial order selectivity index of SEF neurons in 3-object serial visual saccade task: ($FR =$ firing rate)

$$X = \frac{FR_3 - FR_1}{FR_3 + FR_1}$$

Reward expectancy index:

$$Y = \frac{FR_{big} - FR_{small}}{FR_{big} + FR_{small}}$$

Serial order selectivity index of SEF neurons in 3-object serial visual saccade task: ($FR =$ firing rate)

$$X = \frac{FR_3 - FR_1}{FR_3 + FR_1}$$

Reward expectancy index:

$$Y = \frac{FR_{big} - FR_{small}}{FR_{big} + FR_{small}}$$

Observed correlation among 54 neurons $r = .49$.
How much attenuation due to noise?

When noise is added to a pair of variables, correlation is attenuated:

$$Cor(X + \text{ noise}_x, Y + \text{ noise}_y) < Cor(X, Y)$$

When noise is added to a pair of variables, correlation is attenuated:

$$Cor(X + \text{ noise}_x, Y + \text{ noise}_y) < Cor(X, Y)$$

Spearman (1904): correction using noise standard deviations

When noise is added to a pair of variables, correlation is attenuated:

$$Cor(X + \text{ noise}_x, Y + \text{ noise}_y) < Cor(X, Y)$$

Spearman (1904): correction using noise standard deviations

We used hierarchical (random-effects) model and
Bayesian estimation of correlation

Simulation results, MSE:

| | $r$ | corrected $r$ | Bayesian |
|---|---|---|---|
| $n = 15$ | 0.091 | 0.034 | 0.0072 |
| $n = 60$ | 0.014 | 0.0052 | 0.00086 |

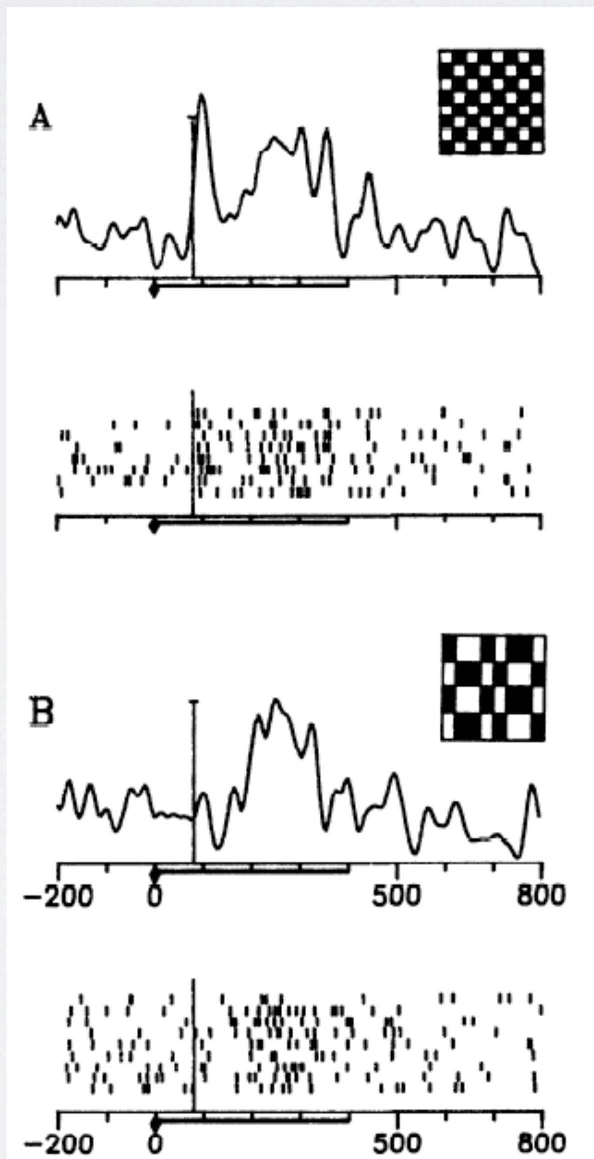
95% confidence interval coverage probability:

| | $z$ | corrected $z$ | Bayesian |
|---|---|---|---|
| $n = 15$ | 0.092 | 0.85 | 0.94 |
| $n = 60$ | 0.61 | 0.96 | 0.95 |

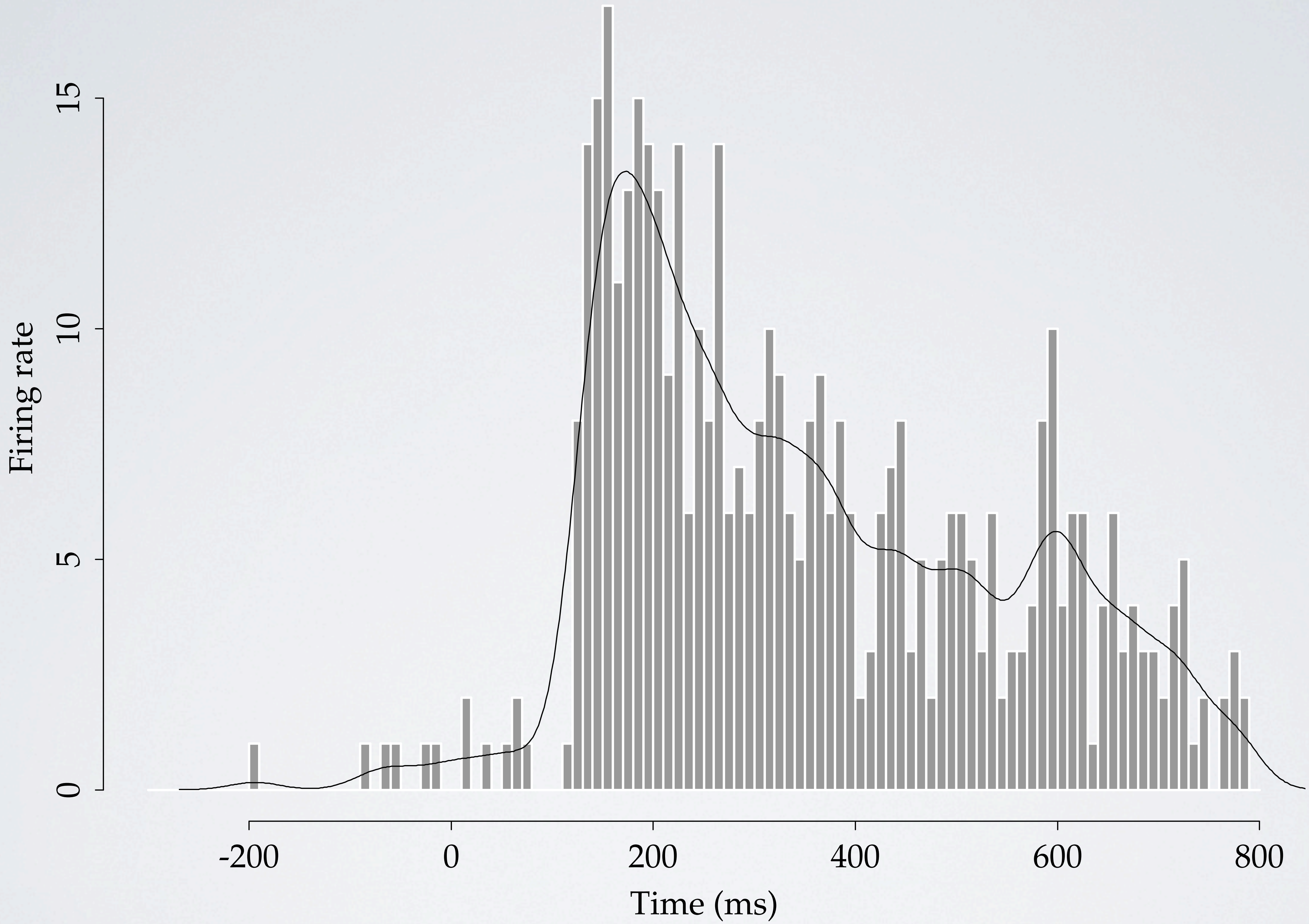
Correlation between serial order selectivity index and reward expectancy index

$$
\begin{aligned}
r &= .49 \\
\hat{\rho} &= .82 \\
& \quad (.77, .88)
\end{aligned}
$$

Optican and Richmond (1987) recorded from IT during presentation of Walsh functions and considered temporal structure of spike trains, concluding that “neurons in [IT] convey messages by temporal modulation of firing rate”



# response of neuron in inferotemporal cortex



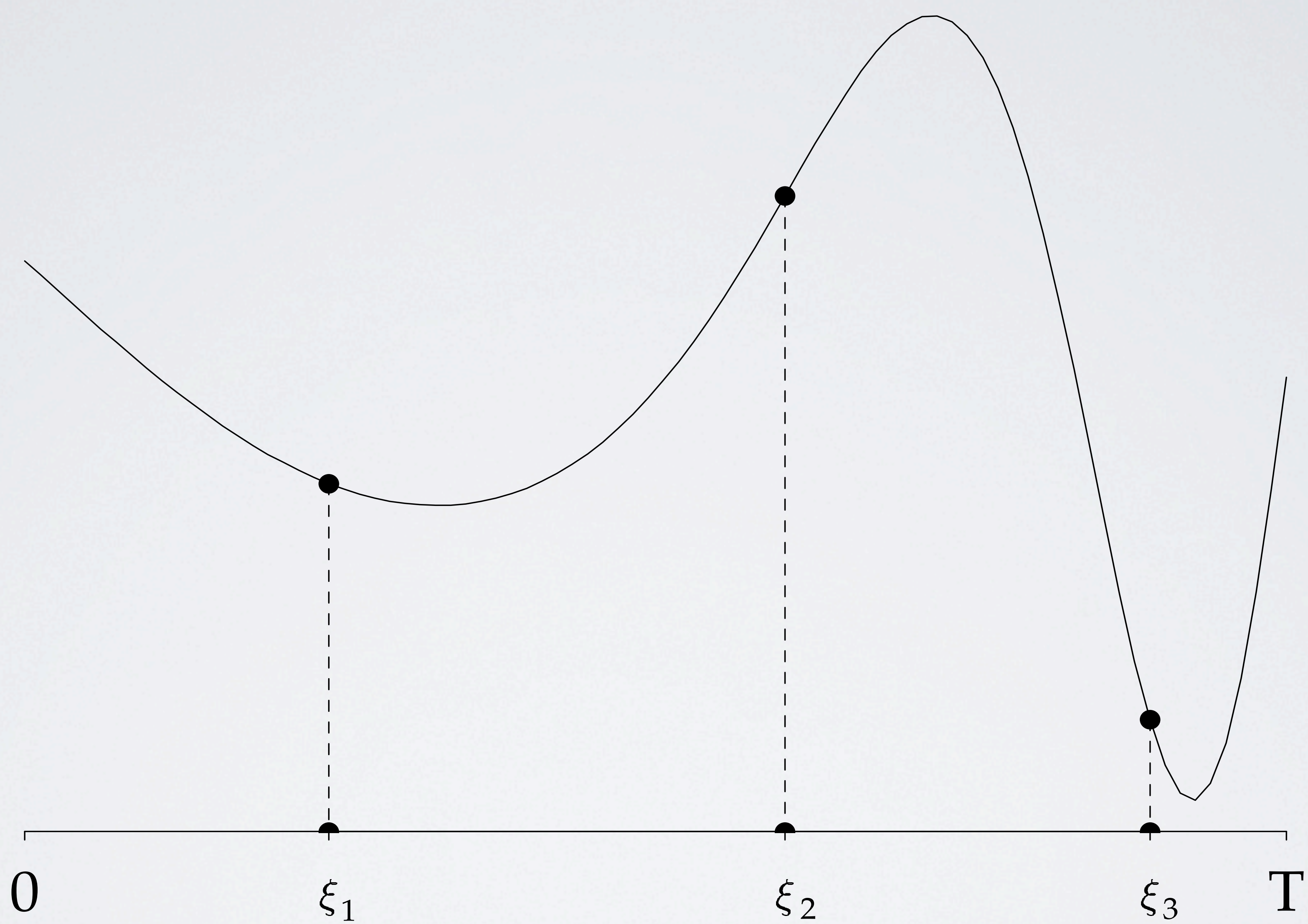
## Bayesian Adaptive Regression Splines (BARS)

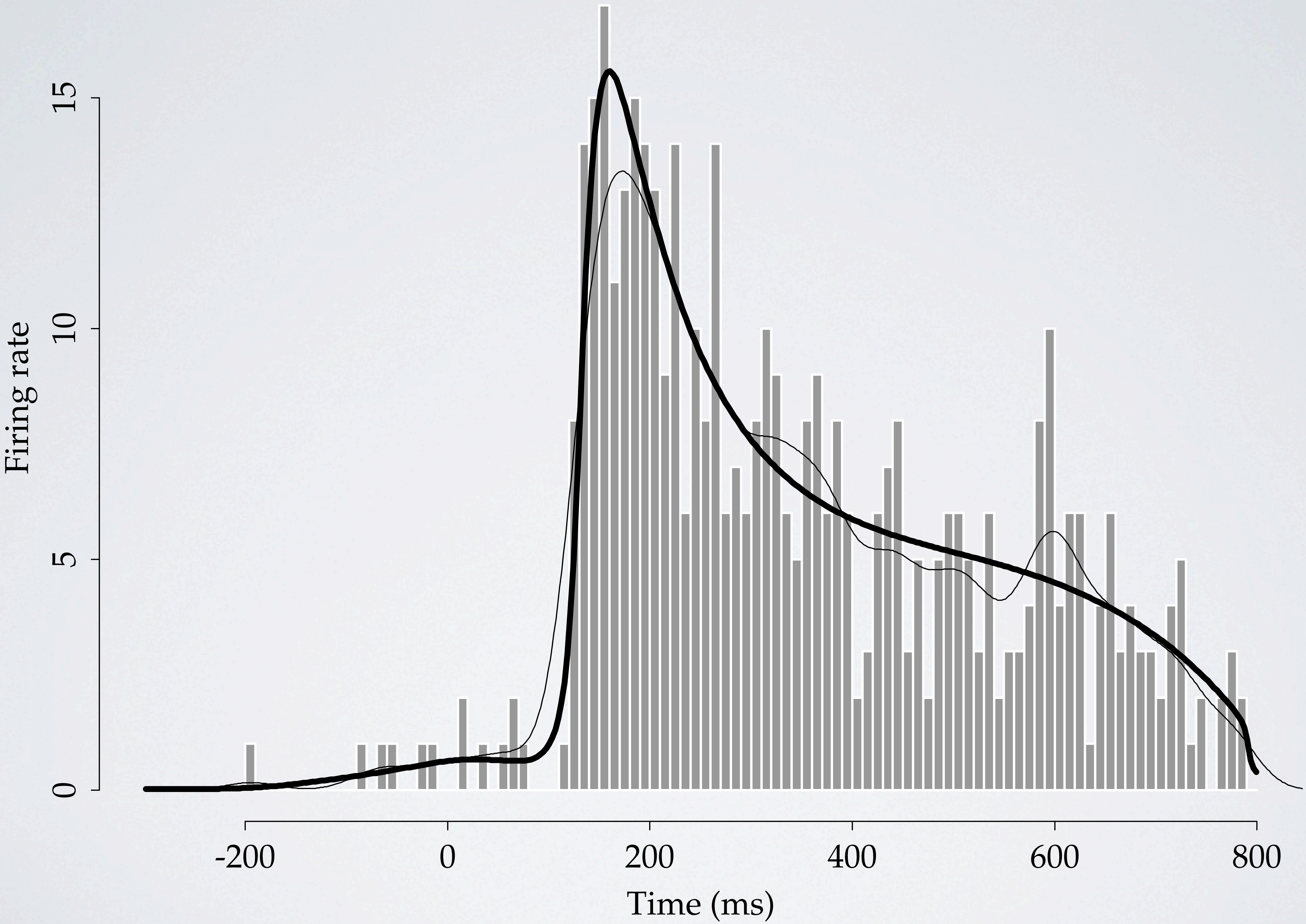
- Uses reversible-jump MCMC to find good knot sets.

Bayesian Adaptive Regression Splines (BARS)

- Uses reversible-jump MCMC to find good knot sets.
- Posterior based on BIC.
- Careful implementation.
- Produces smaller MSE than other methods.

# cubic spline with 3 knots




Firing rate
0
5
10
15
-200
0
200
400
600
800
Time (ms)

Another statistical question:

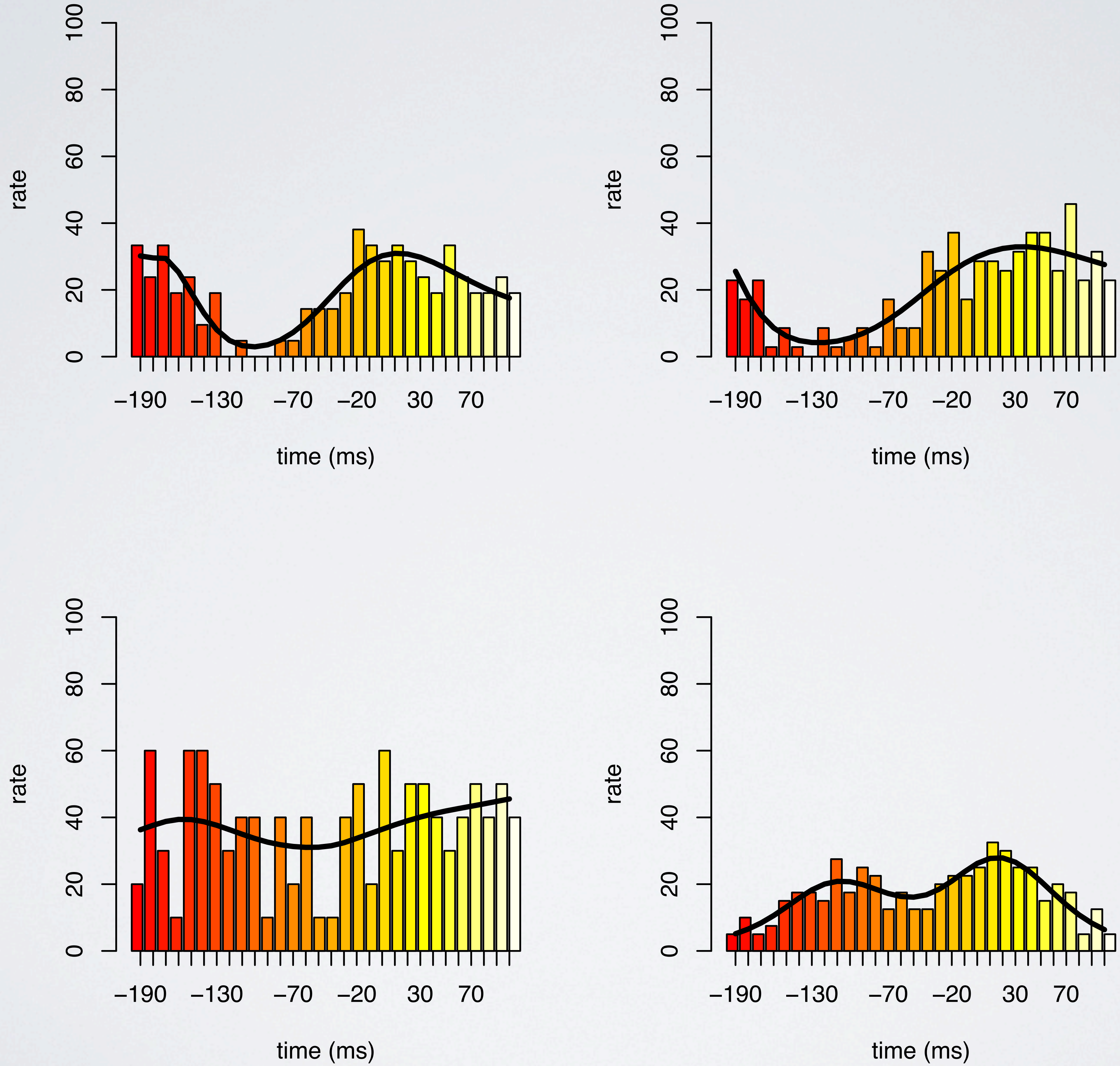
In comparing firing rates across conditions, which time interval should be used?

Another statistical question:

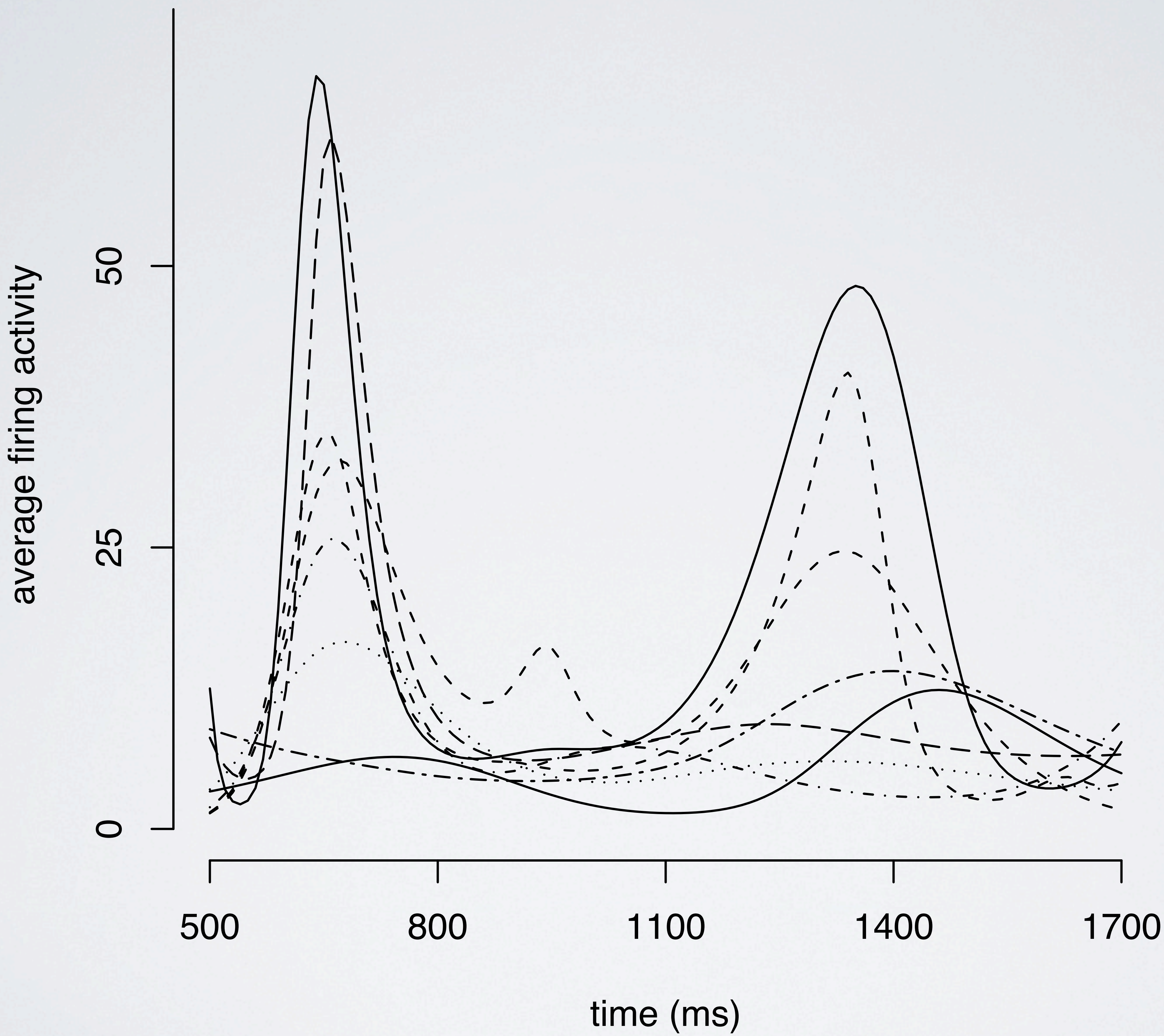
In comparing firing rates across conditions, which time interval should be used?

Our answer: don't need to pick an interval; use entire firing-rate function!

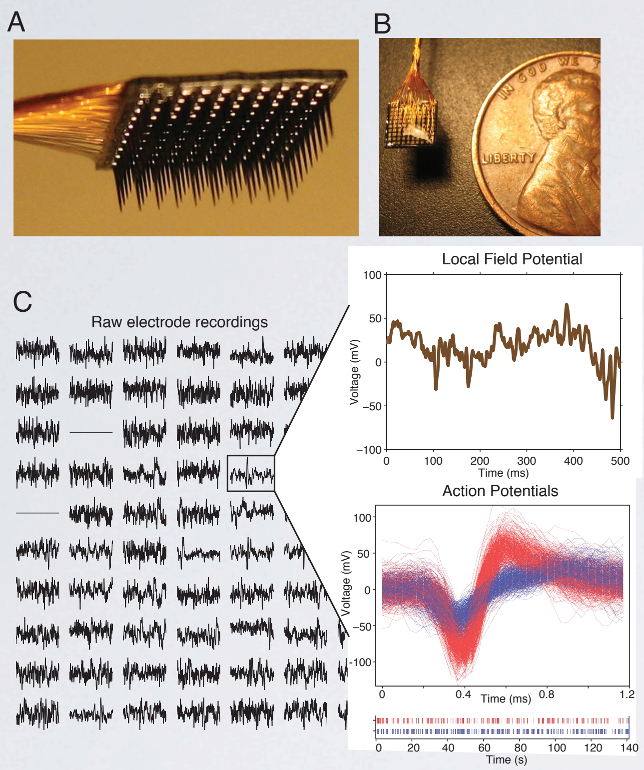
$$H_0 : \lambda_1(t) = \lambda_2(t)$$

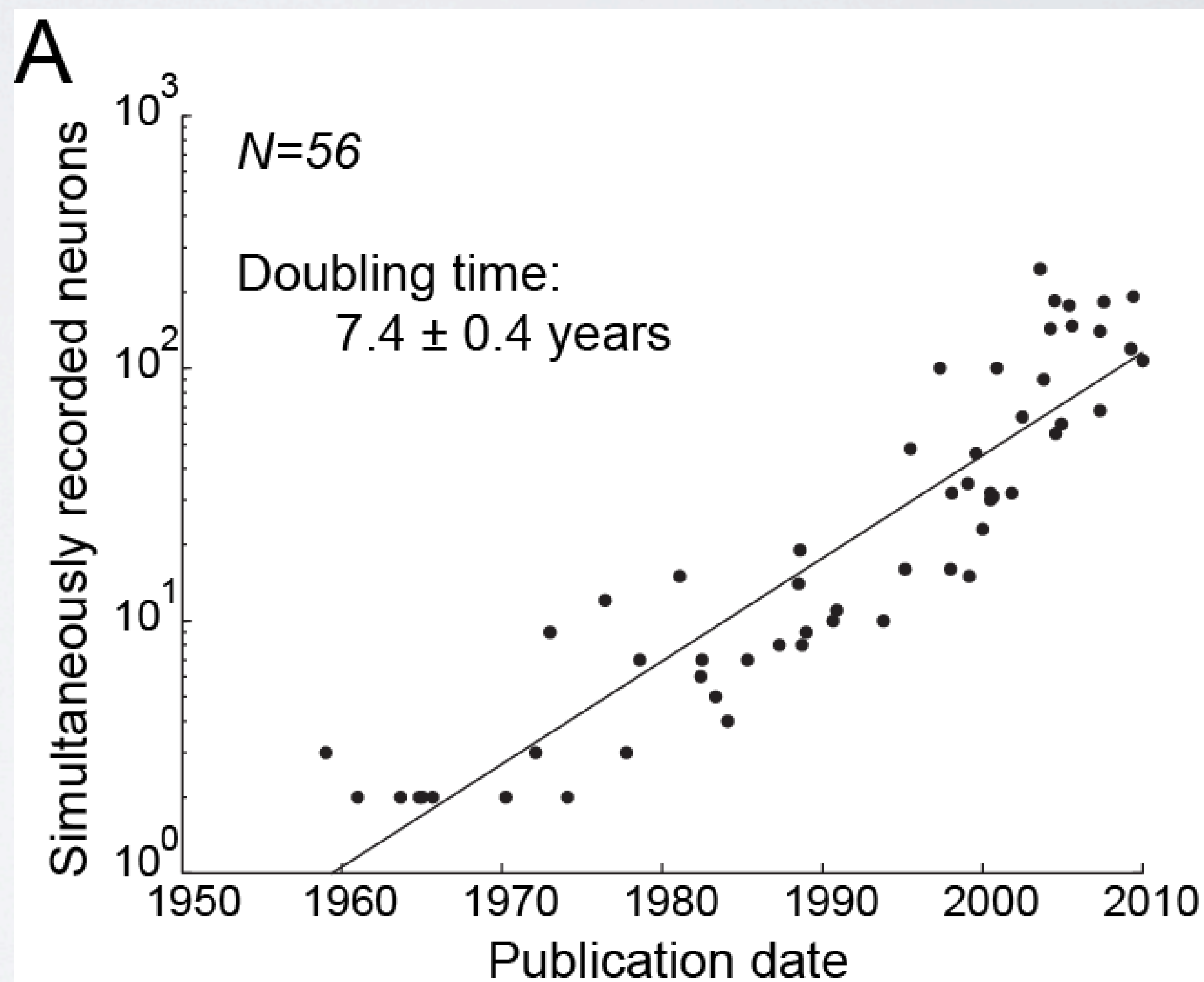


$$H_0 : \lambda_1(t) = \lambda_2(t) = \cdots = \lambda_p(t)$$



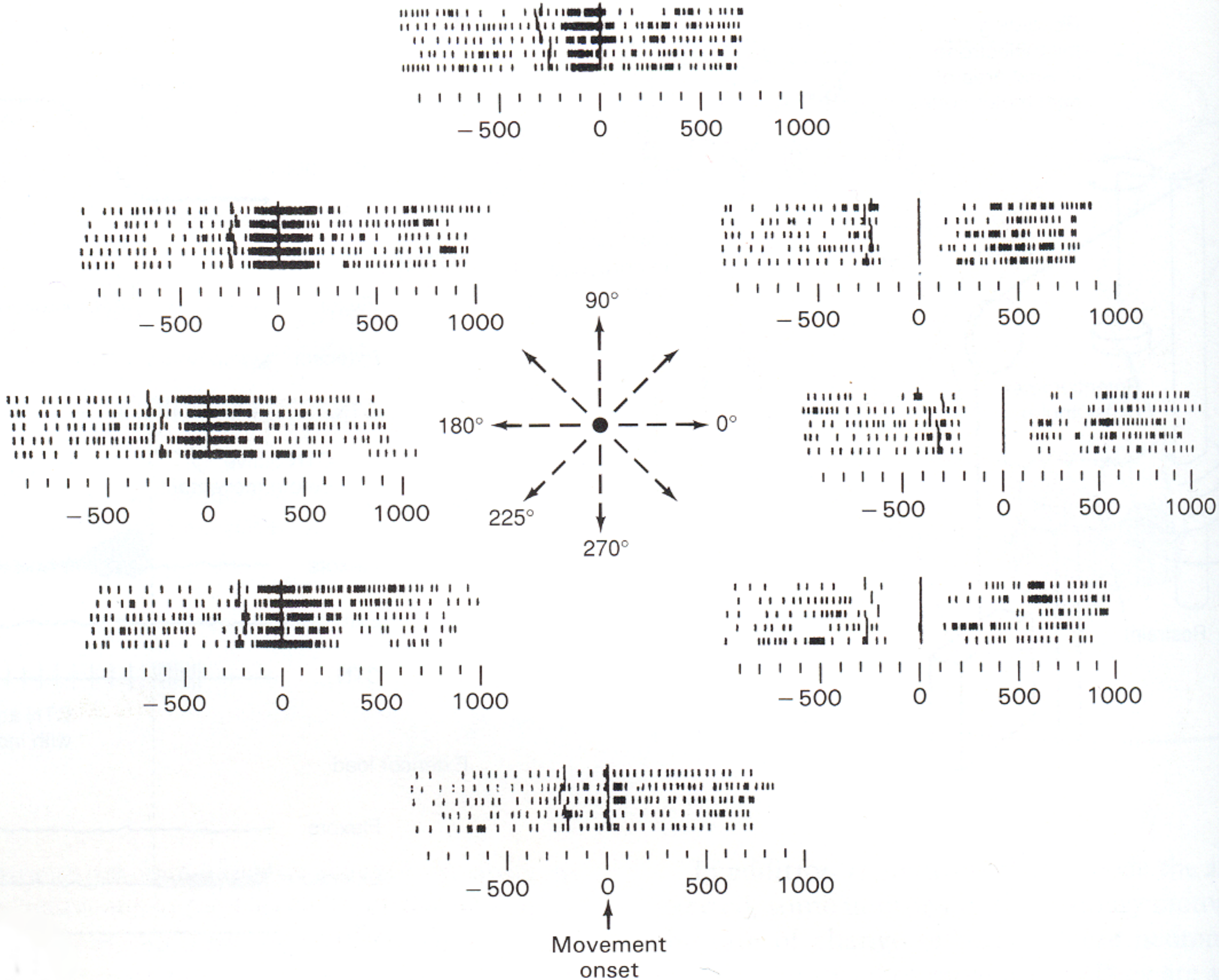
# Simultaneously-recorded multiple spike trains



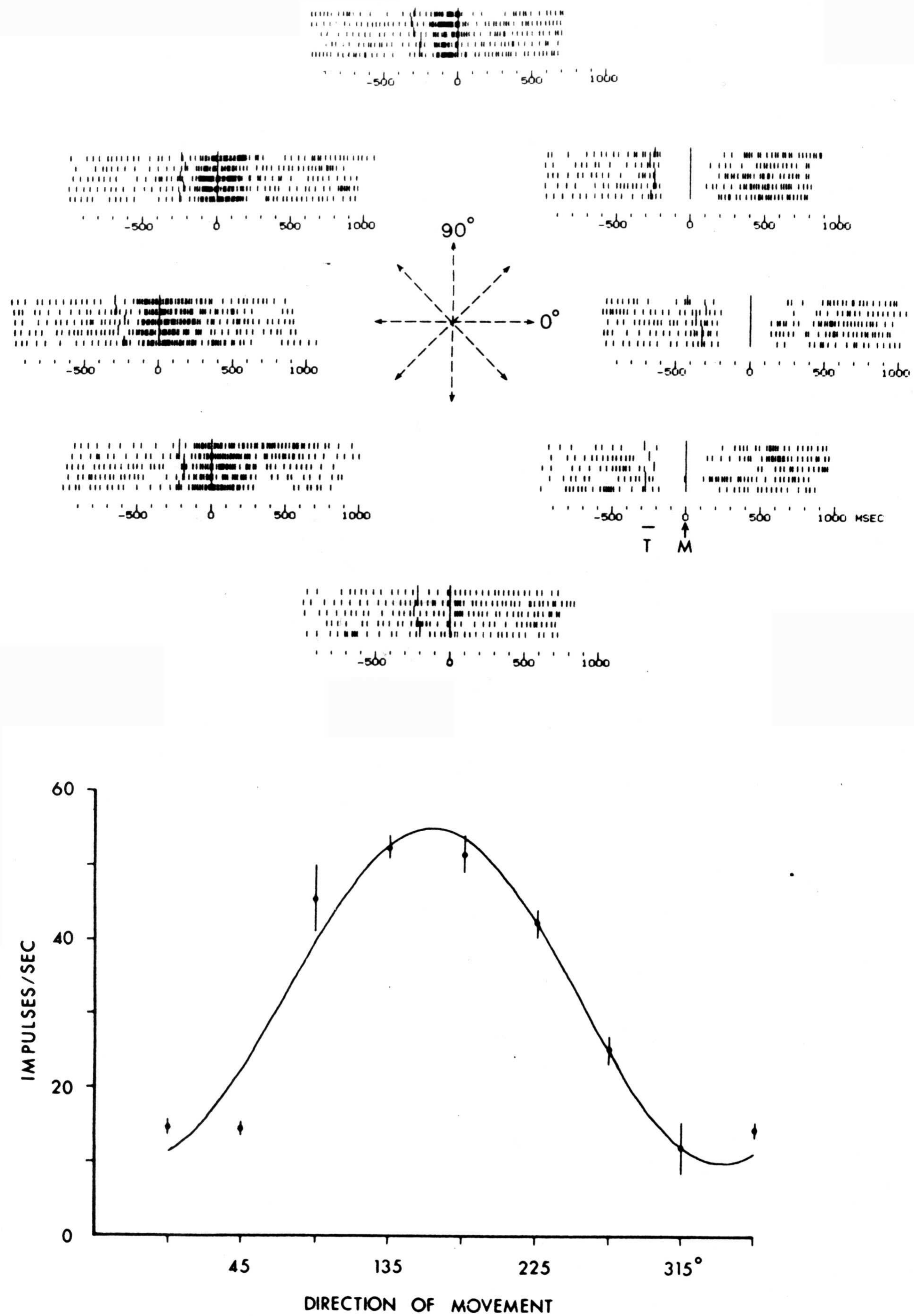




Stevenson and Koerding (2010, *Nature Neuroscience*)

# Motor control and neural prosthetics



90°
180°
0°
225°
270°
−500
0
500
1000
Movement
onset



-500
0
500
1000
90°
0°
1000 MSEC
T
M
60
40
20
0
IMPULSES/SEC
45
135
225
315°
DIRECTION OF MOVEMENT

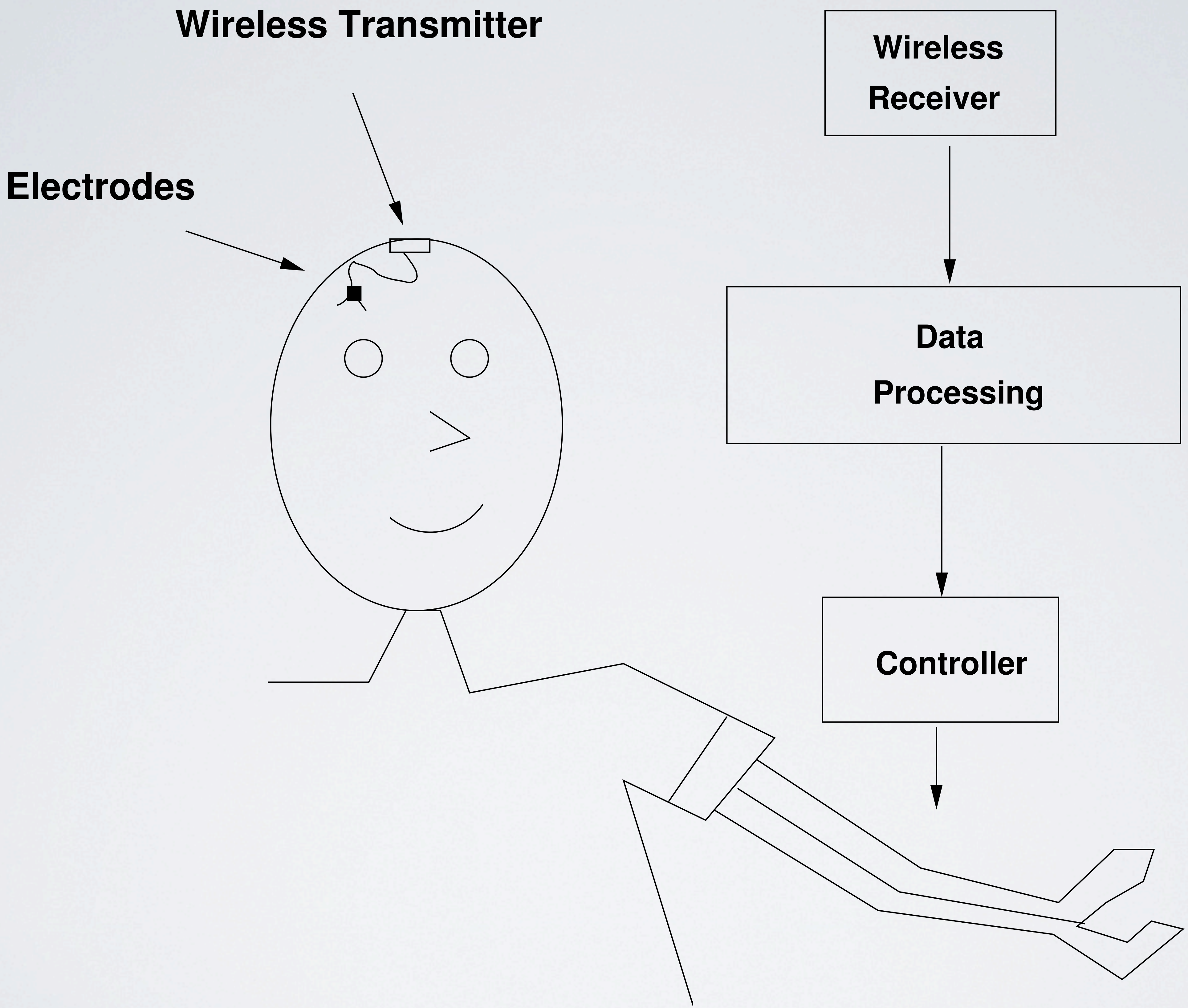
# DIRECTIONAL TUNING MAY BE CAPTURED TO CREATE A PROSTHETIC DEVICE

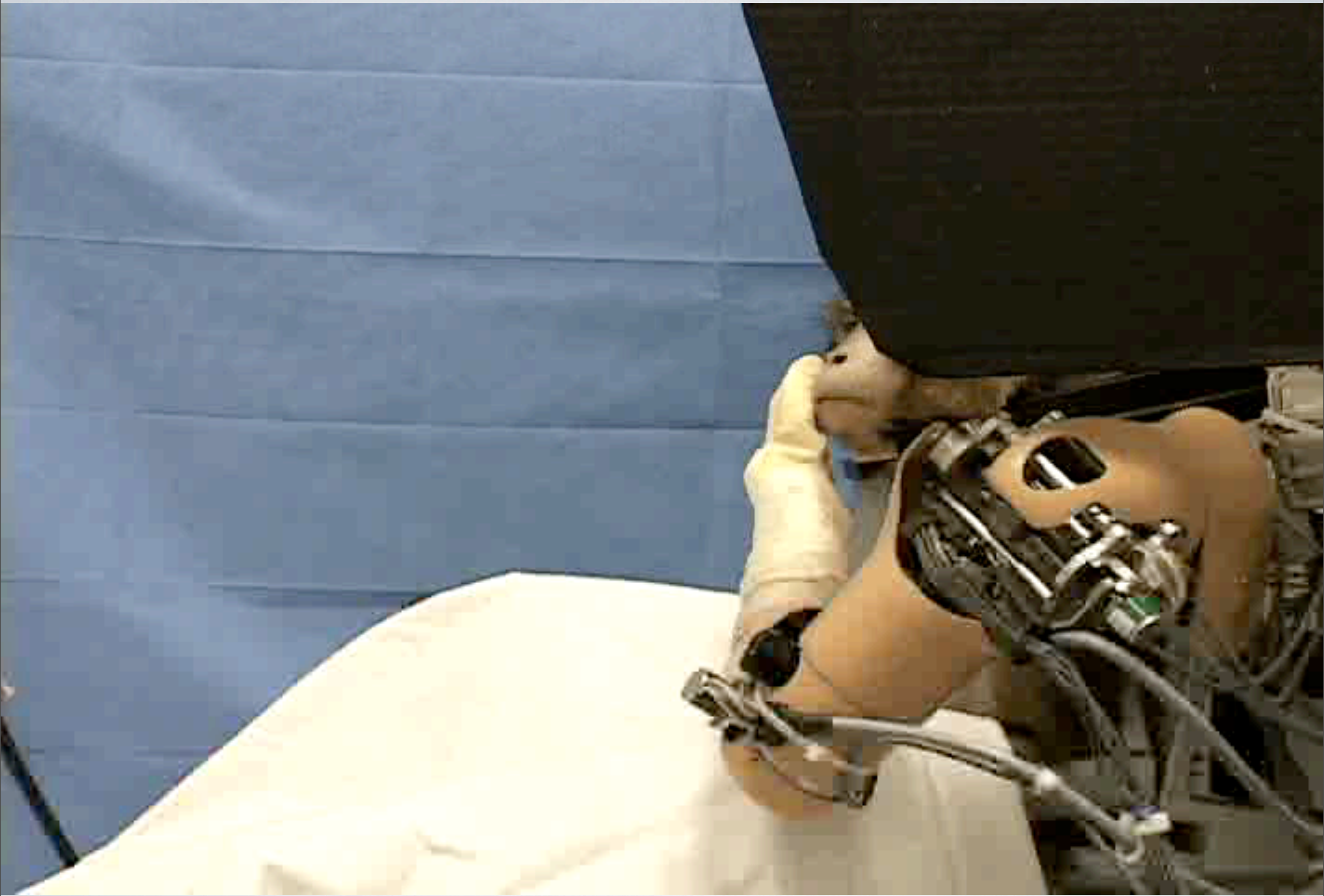
# DIRECTIONAL TUNING MAY BE CAPTURED TO CREATE A PROSTHETIC DEVICE

- Motor cortical neurons are broadly velocity-tuned

# DIRECTIONAL TUNING MAY BE CAPTURED TO CREATE A PROSTHETIC DEVICE

- Motor cortical neurons are broadly velocity-tuned
- By combining activity of dozens of neurons, hand movement can be predicted


Wireless Transmitter
Electrodes
Wireless
Receiver
Data
Processing
Controller



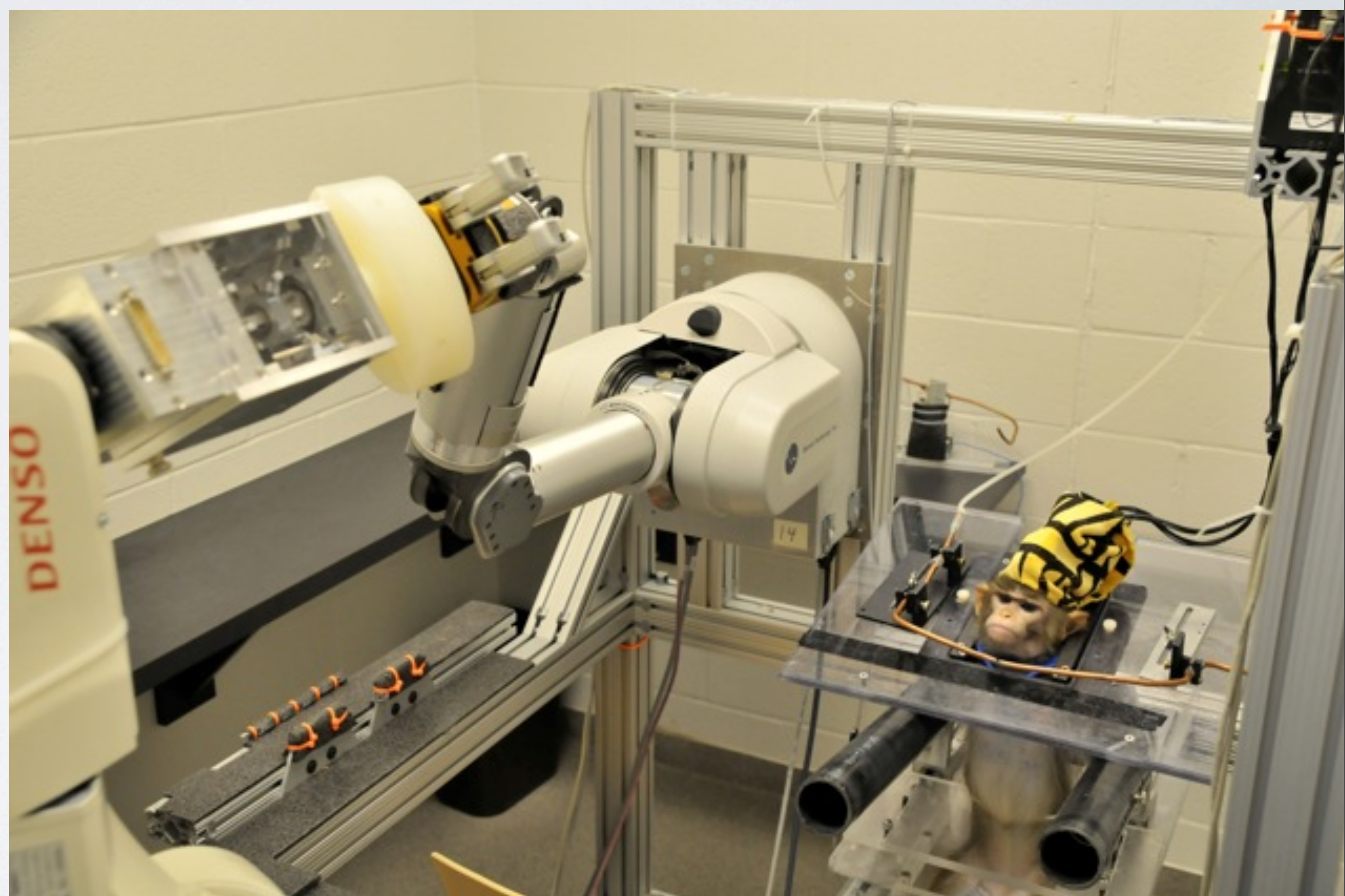
# State-Space Model (Bayesian Decoding)

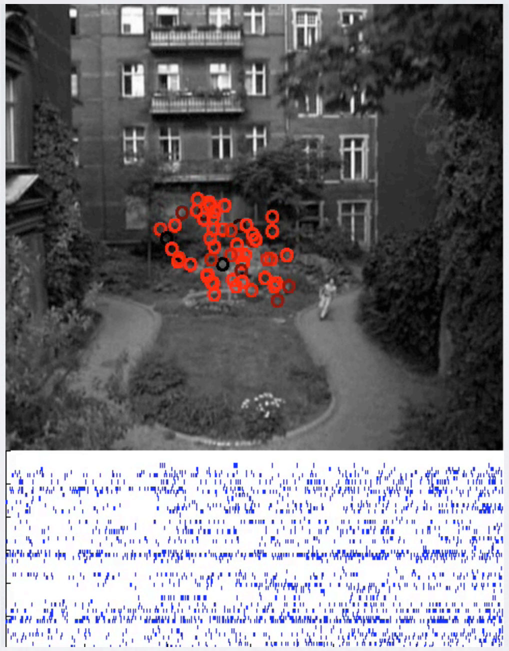
1. Observation model: Prob(spike) = f(state;noise)

2. State model: state = $\text{state}_t$ (position, velocity, etc.)
evolves over time

3. Bayes' theorem gives prediction of $\text{state}_t$
using data together with previous predicted states

# Current work: finger movement and gripper control





# Spike train modeling within trials

conditional intensity

$$FR = \lambda(t|H_t)$$

Modeling starting point: $\log \lambda(t|H_t) = f(t, H_t, x)$
Right-hand side typically involves several additive terms

$$\begin{aligned} \log FR &= \text{stimulus effects} + \text{coupling effects} \\ &+ \text{history effects} + \text{global network effects} \end{aligned}$$

# Spike train modeling within trials

conditional intensity

$$FR = \lambda(t|H_t)$$

Modeling starting point: $\log \lambda(t|H_t) = f(t, H_t, x)$
Right-hand side typically involves several additive terms

$$\begin{aligned} \log FR &= \text{stimulus effects} + \text{coupling effects} \\ &+ \text{history effects} + \text{global network effects} \end{aligned}$$

Solution: L1 regularized Poisson regression

# Spike train modeling within trials

conditional intensity

$$FR = \lambda(t|H_t)$$

Modeling starting point: $\log \lambda(t|H_t) = f(t, H_t, x)$
Right-hand side typically involves several additive terms

$$\begin{aligned} \log FR &= \text{stimulus effects} + \text{coupling effects} \\ &+ \text{history effects} + \text{global network effects} \end{aligned}$$

Solution: L1 regularized Poisson regression

Kelly, Kass, Smith, Lee (2010, *NIPS*)
Pillow, et al. (2008, *Nature*)
Truccolo, et al. (2010, *Nature Neuroscience*)

# Synchronous firing among neurons

two or more neurons fire nearly at the same time

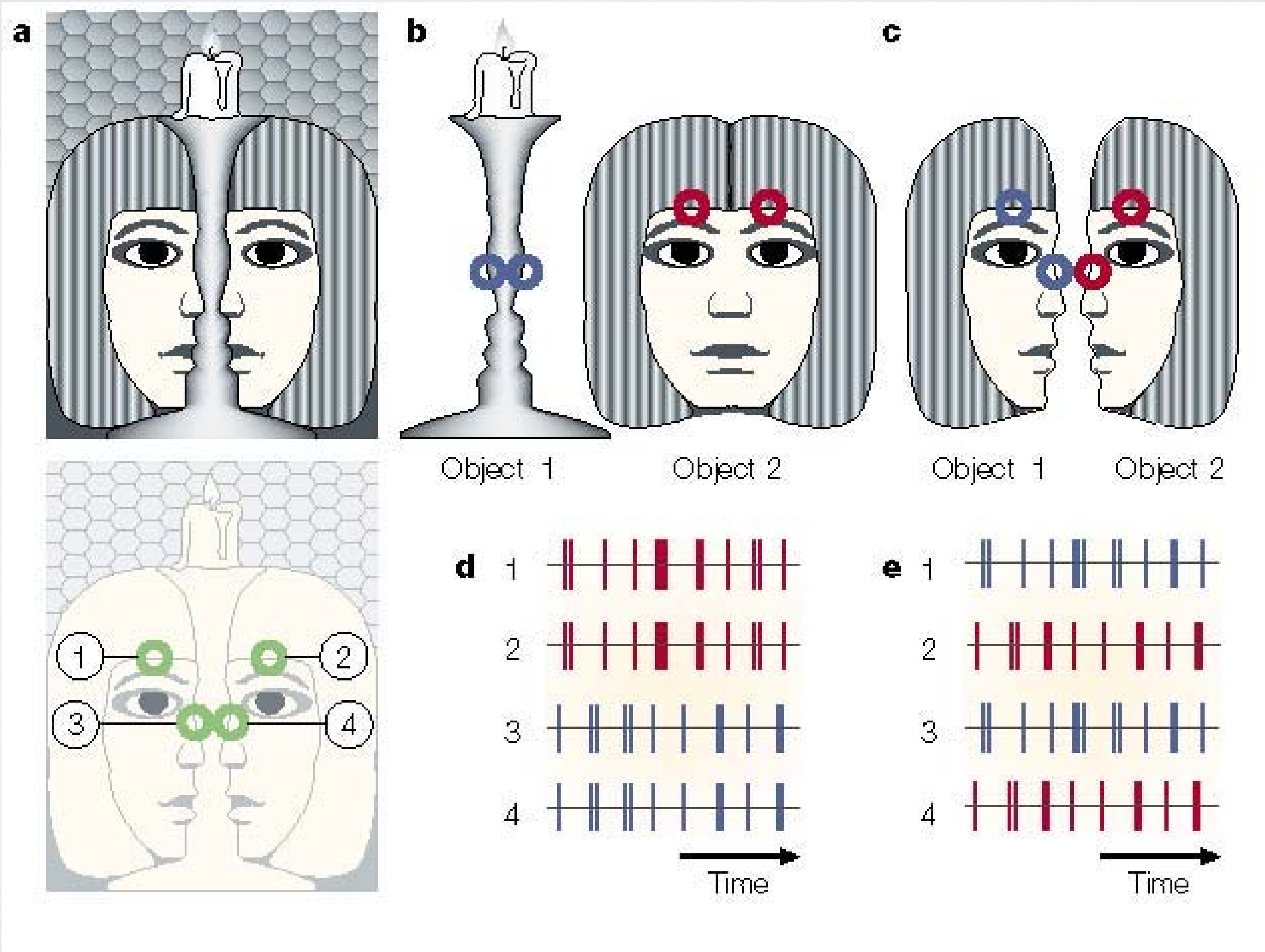
i.e. within a bin of width $\delta$ (here, $\delta = 5$ ms)

# Synchronous firing among neurons

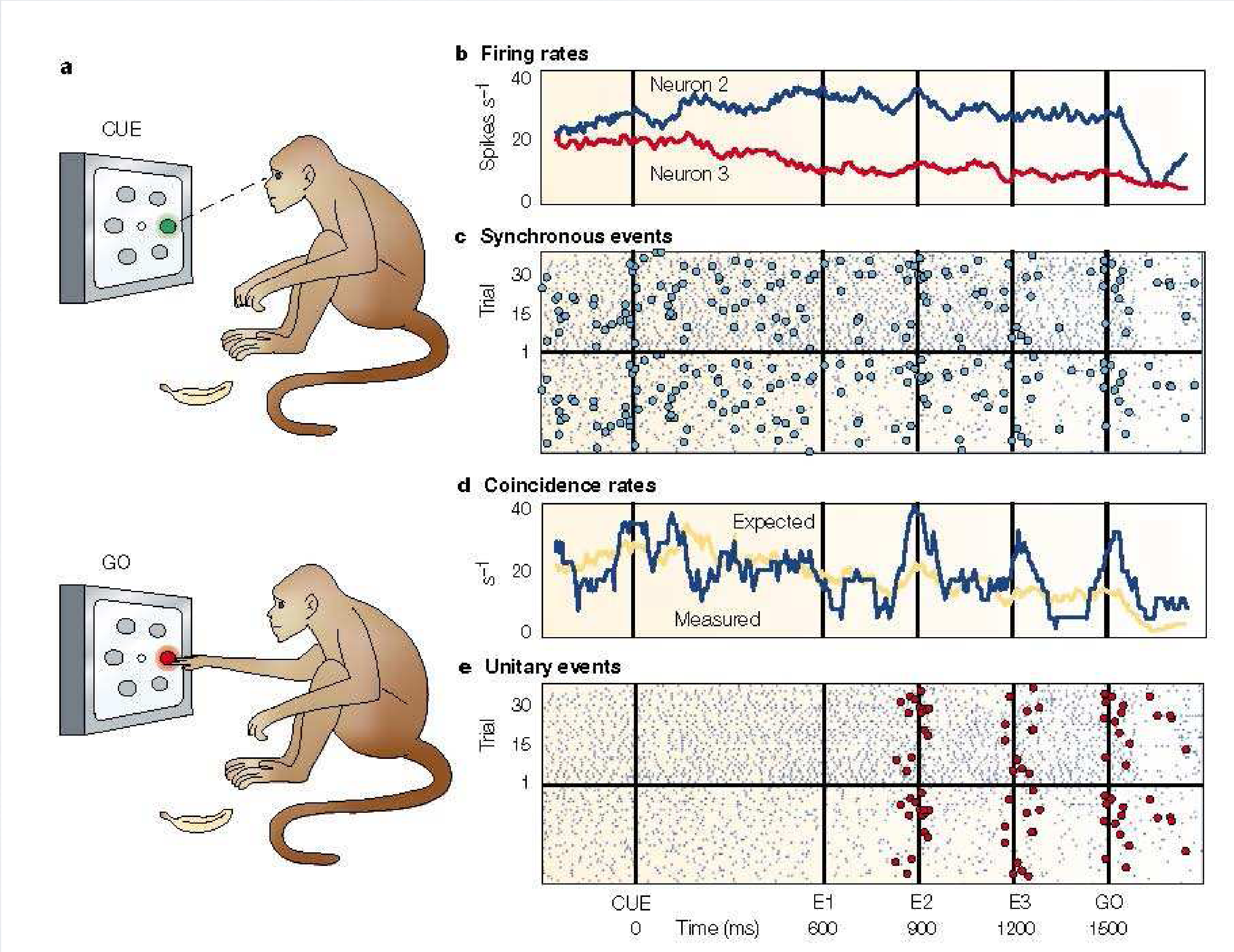
two or more neurons fire nearly at the same time

i.e. within a bin of width $\delta$

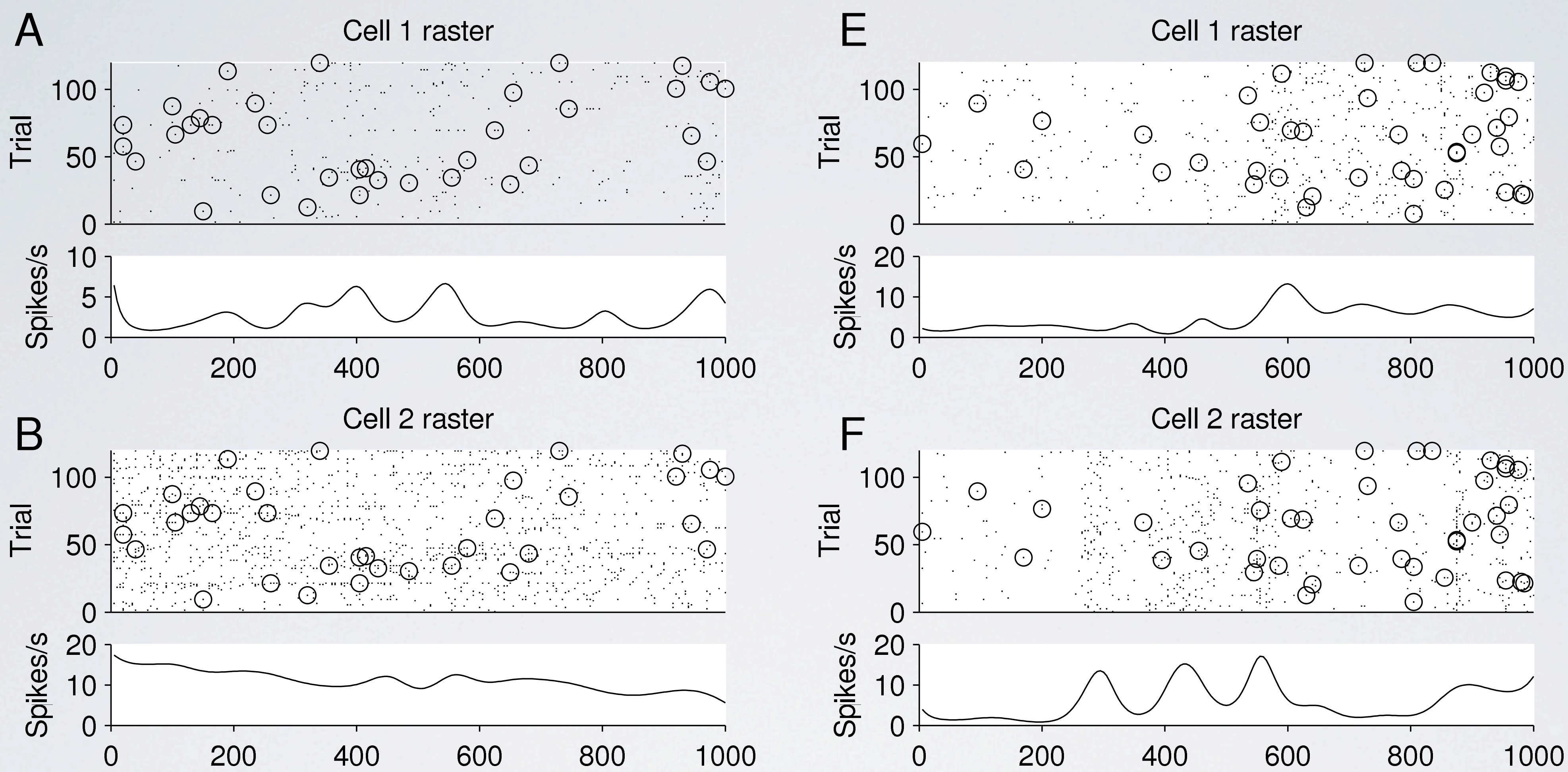
Engel, Fries, Singer (2001)



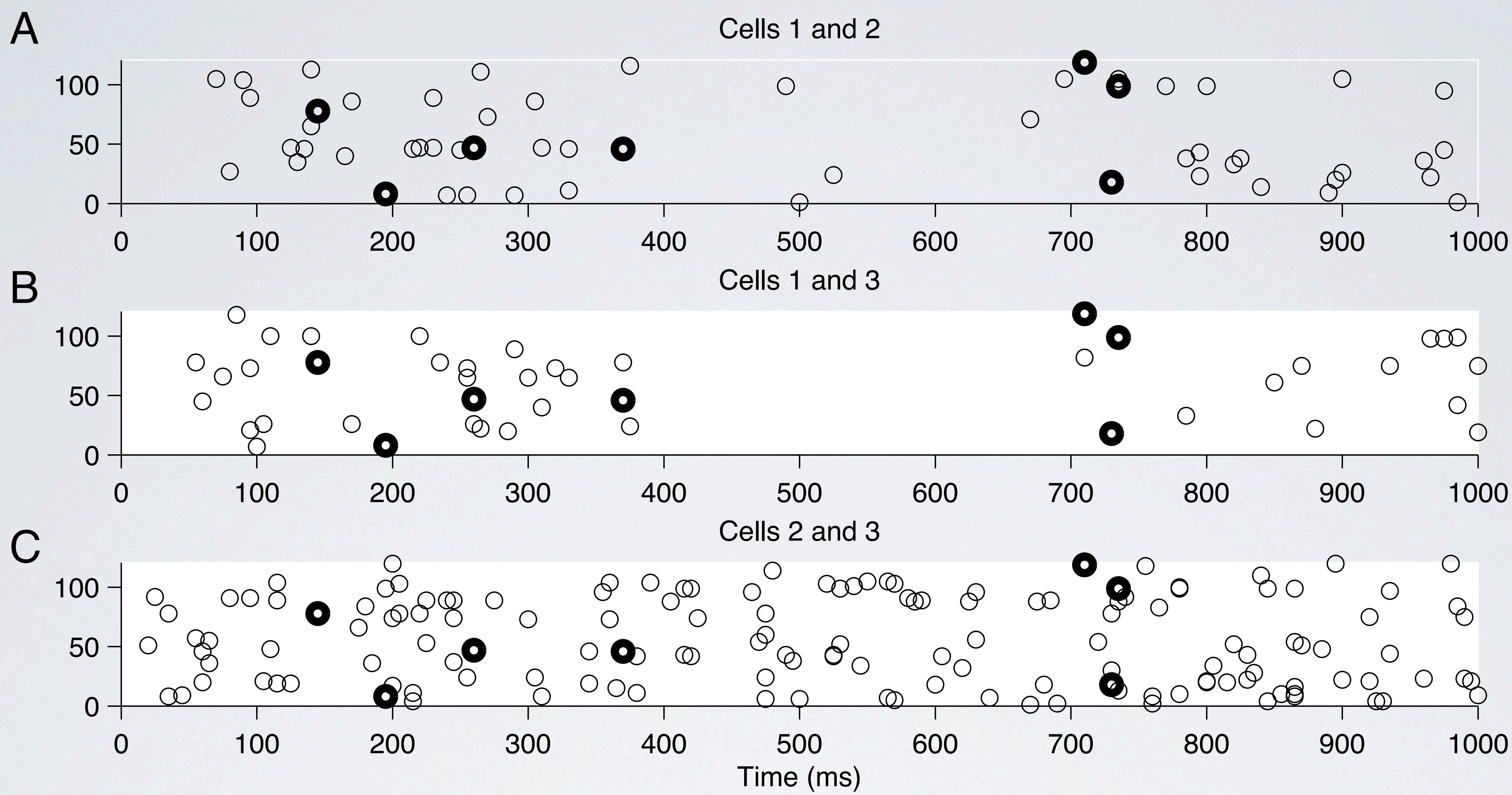
Riehle, Grün, Diesmann, and Aertsen (1997) examined neurons in motor cortex during a delayed-response hand-reaching task and found occasional synchronous firing, within 5 ms window, at times of anticipation of the signal to move.



# pair 1 pair 2



A
Cell 1 raster
Trial
Spikes/s
B
Cell 2 raster
E
Cell 1 raster
F
Cell 2 raster
0 200 400 600 800 1000


A
Cells 1 and 2
B
Cells 1 and 3
C
Cells 2 and 3
0
50
100
0
100
200
300
400
500
600
700
800
900
1000
Time (ms)

Vol 440|20 April 2006|**doi:10.1038/nature04701** nature

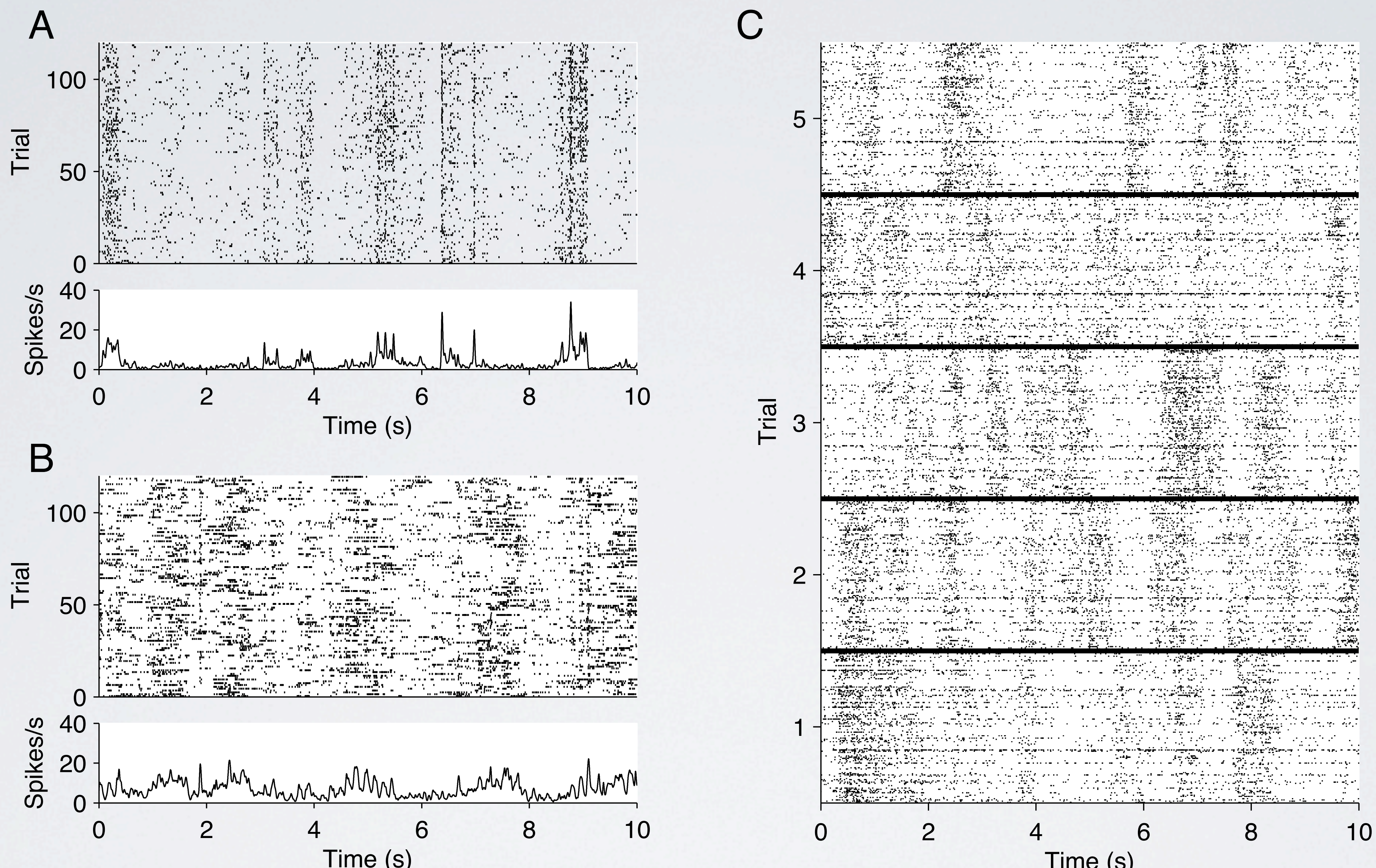
ARTICLES

# Weak pairwise correlations imply strongly correlated network states in a neural population

Elad Schneidman[1,2,3], Michael J. Berry II[2], Ronen Segev[2] & William Bialek[1,3]

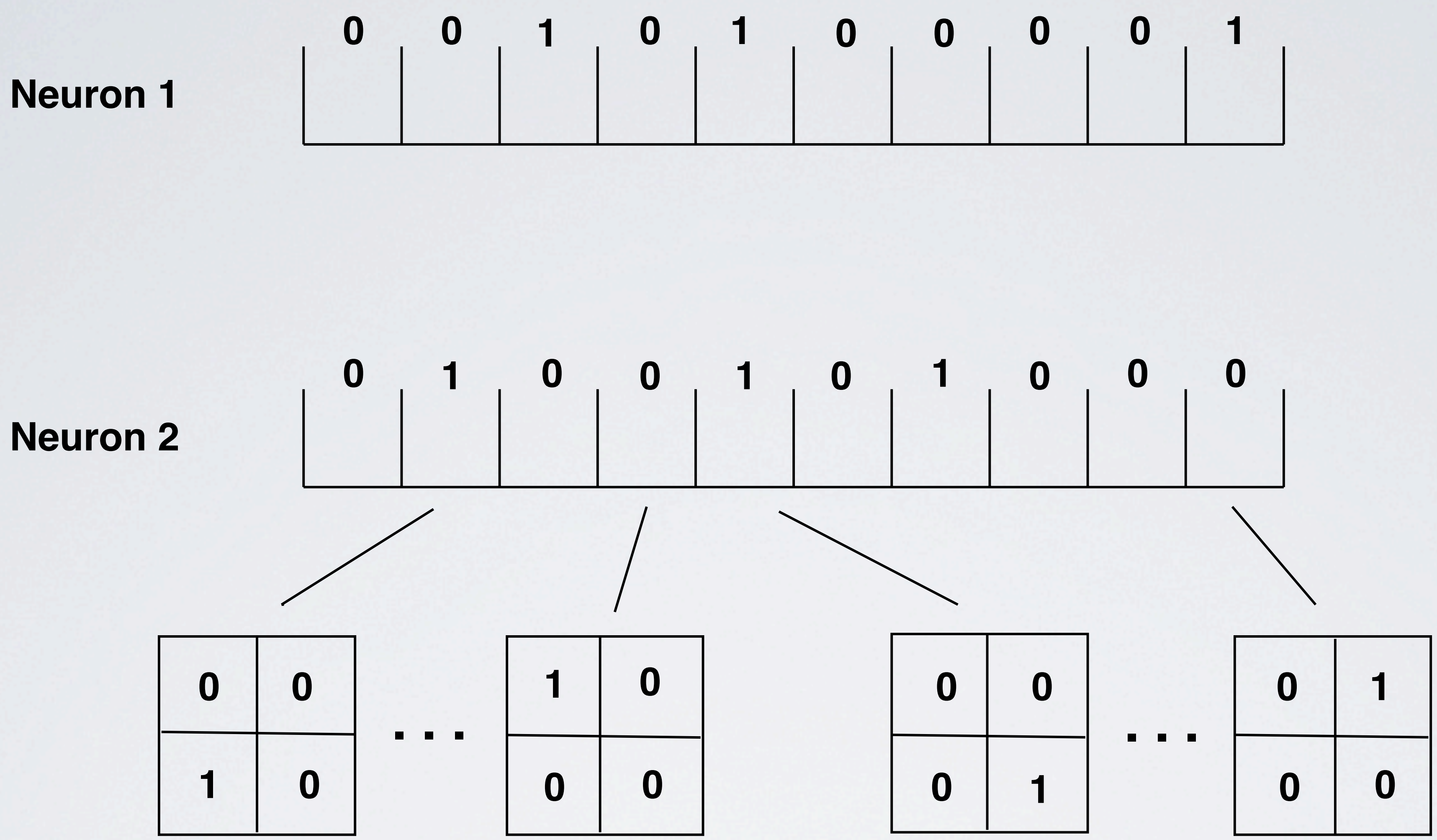
one neuron

all 128 neurons

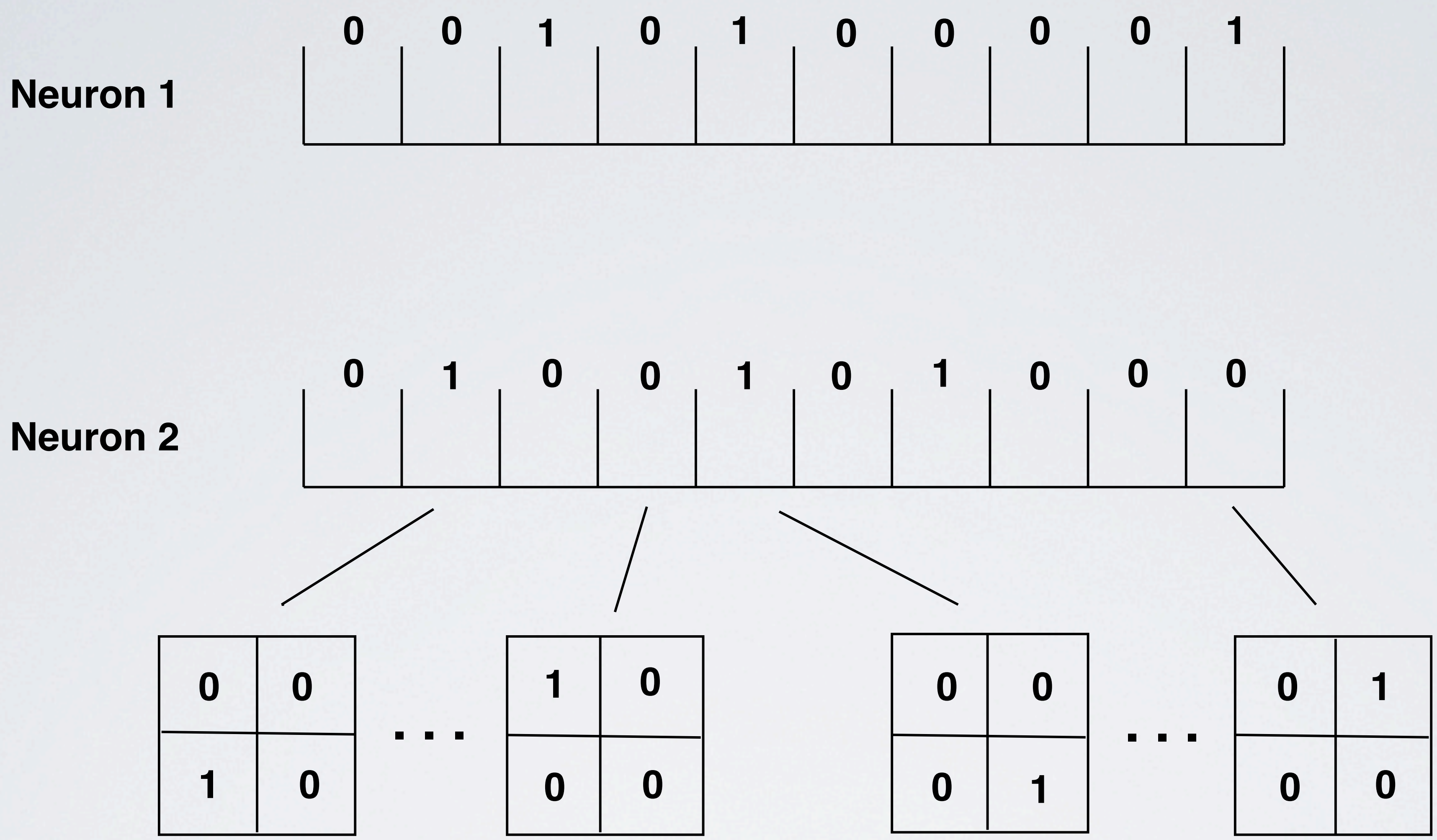



another neuron

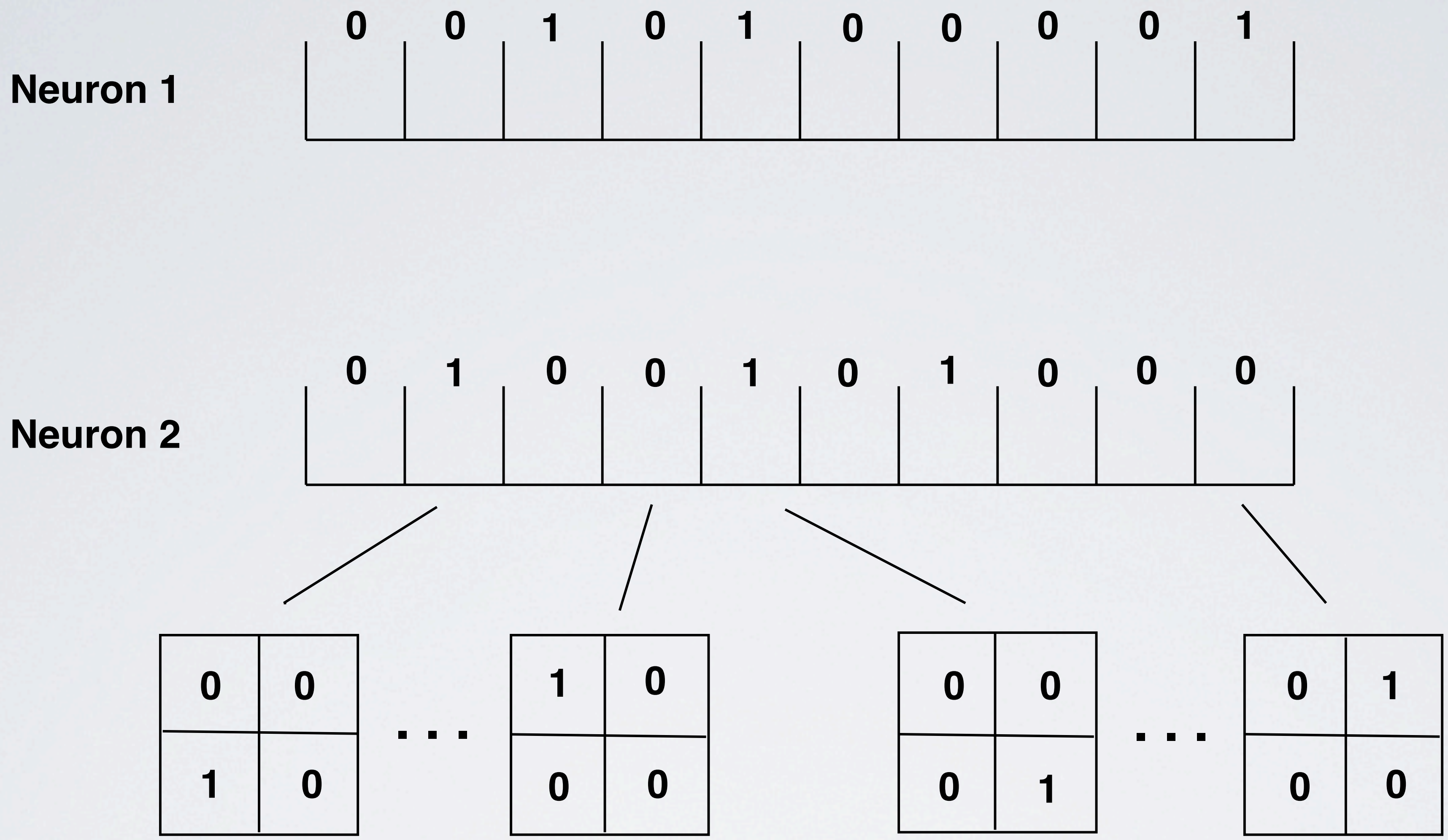
| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Neuron 1** | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| **Neuron 2** | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 |



Neuron 1
0 0 1 0 1 0 0 0 0 1
Neuron 2
0 1 0 0 1 0 1 0 0 0
0 0
1 0
. . .
1 0
0 0
0 0
0 1
. . .
0 1
0 0




time series of 2x2 tables




time series of 2x2 tables

for n neurons, time series of 2^n tables

Basic idea:

modify single-neuron regression methods

binary data --> polytomous data

in order to treat time series of 2^n tables

Complications:

1. nonstationary stimulus effects
2. history effects

must avoid huge numbers of parameters

math problem: how can synchrony be accommodated in point process framework?

Kass, Kelly, Loh (2011, *Annals of Applied Statistics*)

# Under independence: for neurons i and j

$$P(\text{joint spikes}_{ij} \text{ in } (t, t+\delta)) = P(\text{spike}_i \text{ in } (t, t+\delta)) \cdot P(\text{spike}_j \text{ in } (t, t+\delta))$$

# Under independence: for neurons i and j

$$P(\text{joint spikes}_{ij} \text{ in } (t, t+\delta)) = P(\text{spike}_i \text{ in } (t, t+\delta)) \cdot P(\text{spike}_j \text{ in } (t, t+\delta))$$

$$P_{11} = P_{1+} \cdot P_{+1}$$

independence: $P_{11} = P_{1+} \cdot P_{+1}$

general: $P_{11} = P_{1+} \cdot P_{+1} \cdot \zeta$

$\zeta - 1$ is excess proportion of joint spiking

$$P_{11}(t) = P_{1+}(t) \cdot P_{+1}(t) \cdot \zeta(t)$$

$$P_{11}(t) = P_{1+}(t) \cdot P_{+1}(t) \cdot \zeta(t)$$

$$\log P_{11}(t) = \log P_{1+}(t) + \log P_{+1}(t) + \log \zeta(t)$$

$$P_{11}(t) = P_{1+}(t) \cdot P_{+1}(t) \cdot \zeta(t)$$

$$\log P_{11}(t) = \log P_{1+}(t) + \log P_{+1}(t) + \log \zeta(t)$$

$$\log \lambda^{ij}(t|H_t) = \log \lambda^{i}(t|H_t) + \log \lambda^{j}(t|H_t) + \log \zeta(t)$$

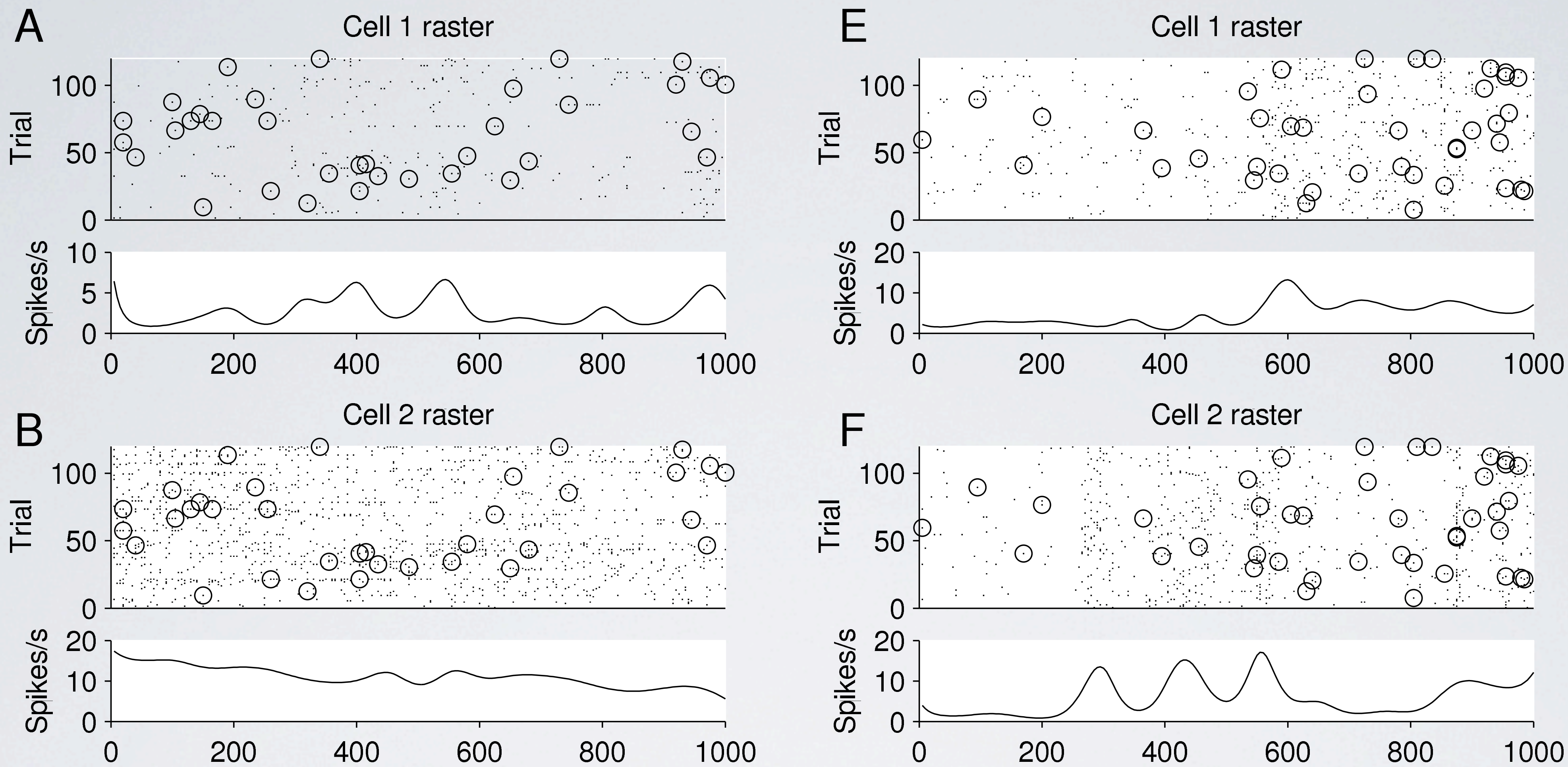
# continuous-time loglinear model of firing-rate (intensity)

$$\log \lambda^{ij}(t|H_t) = \log \lambda^i(t|H_t^i) + \log \lambda^j(t|H_t^j) + \log \zeta$$

or

$$\begin{aligned} \log FR_{ij}(t) &= stimulus_i(t) + history_i(t) \\ &+ stimulus_j(t) + history_j(t) \\ &+ excess \end{aligned}$$

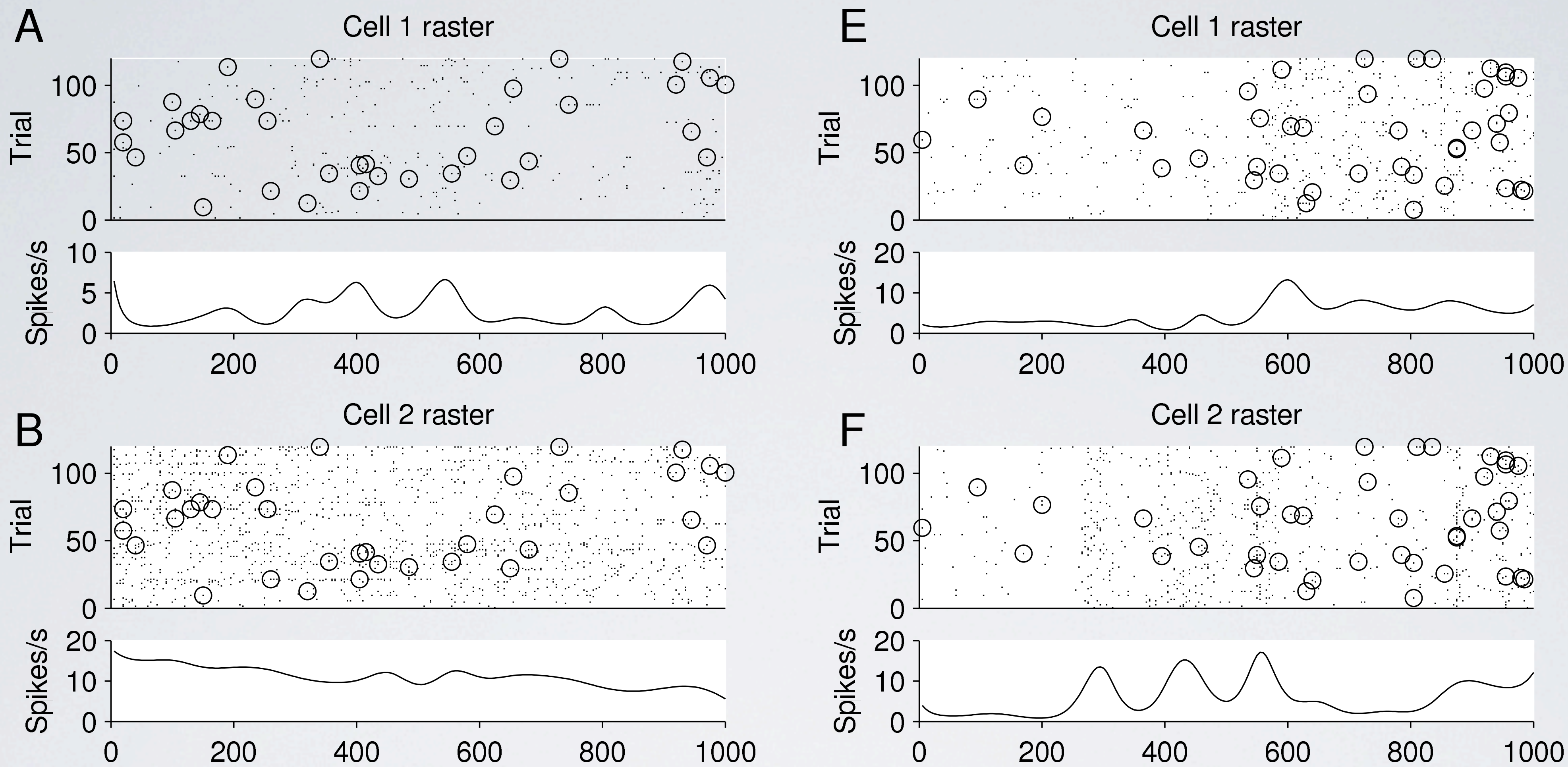
# pair 1      pair 2




$$\log \hat{\zeta} = .06 \pm .15$$

$$\log \hat{\zeta} = .82 \pm .23$$

# pair 1      pair 2

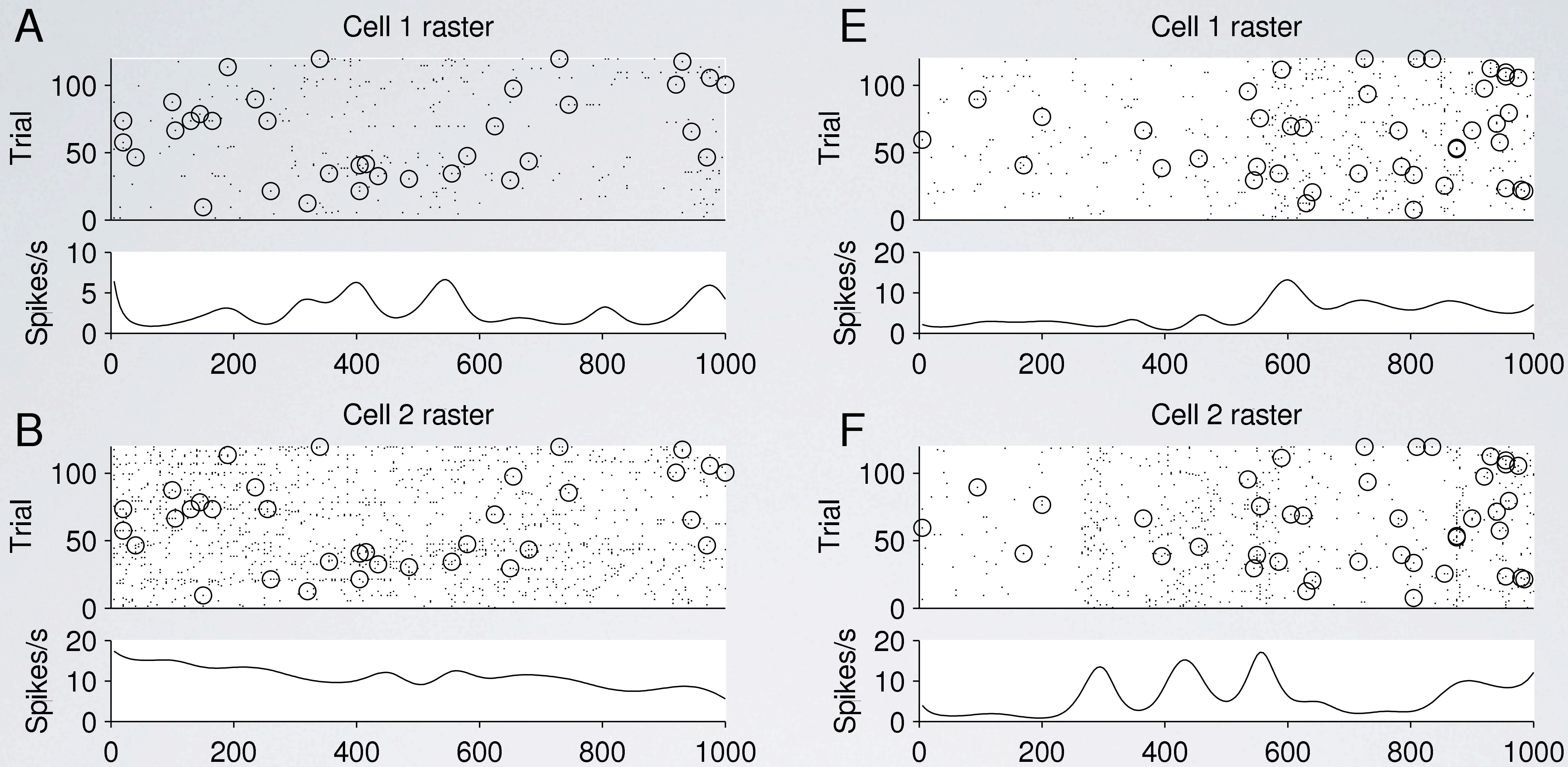



Both pairs: excess synchrony

$$\log \hat{\zeta} = .06 \pm .15$$

$$\log \hat{\zeta} = .82 \pm .23$$

# pair 1      pair 2




## Both pairs: excess synchrony

left pair: synchrony due to slow waves

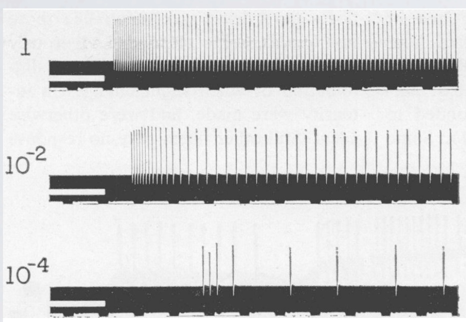
$$\log \hat{\zeta} = .06 \pm .15$$

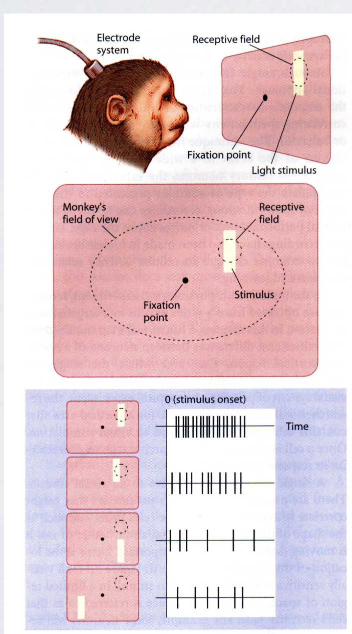
right pair: not due to slow waves

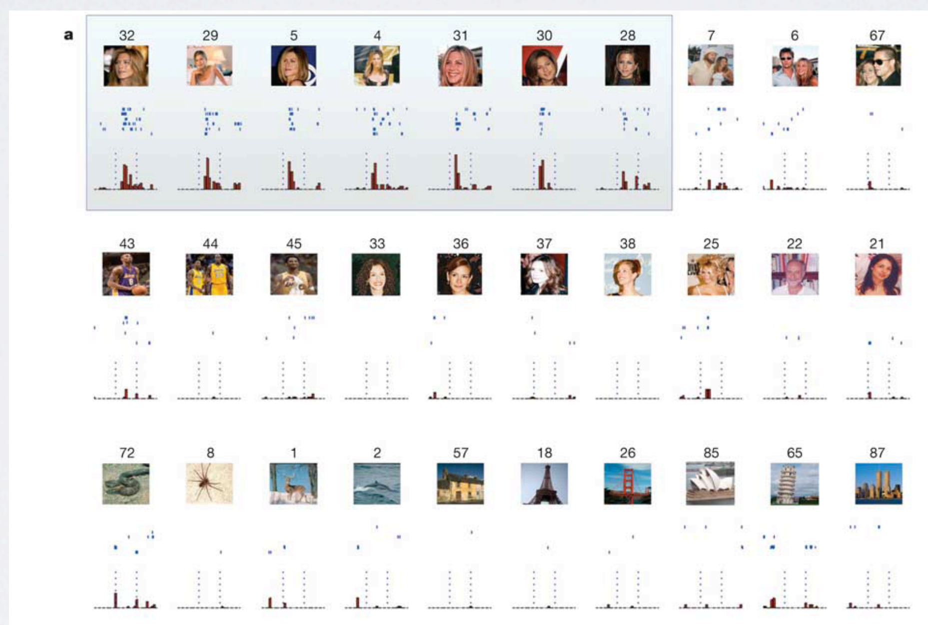
$$\log \hat{\zeta} = .82 \pm .23$$

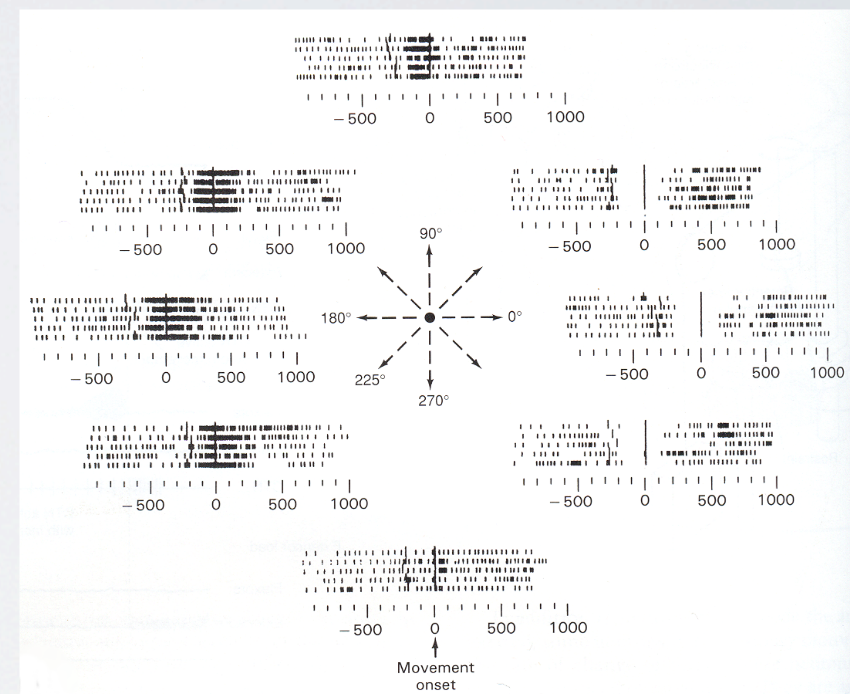
# Recap and Comments

One way neurons represent and transmit information is by increasing firing rate---represented by normalized counts

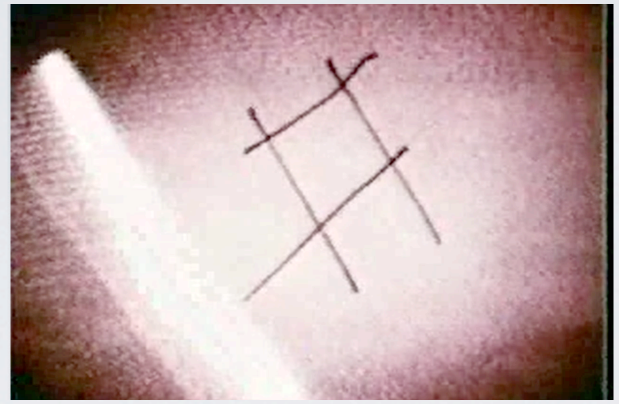


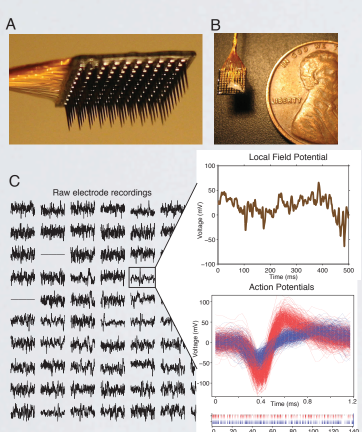



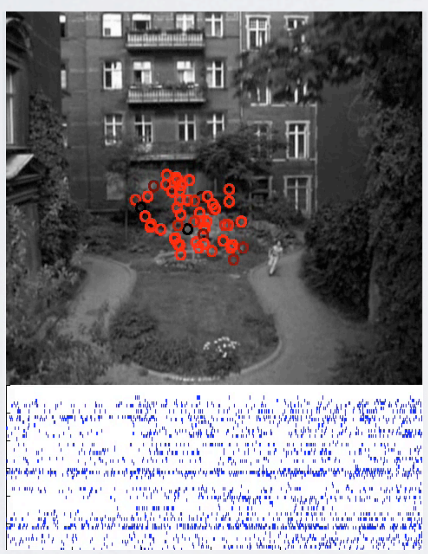



1963:



current state of the art:



$$
\begin{aligned}
\log FR &= \text{stimulus effects} + \text{coupling effects} \\
&+ \text{history effects} + \text{global network effects}
\end{aligned}
$$

(Ryan Kelly PhD thesis)

# Neural coding of firing rate may take any of several forms

$$\log FR = \text{stimulus effects} + \text{covariate effects}$$

$FR$ defined as

- $E(\text{count})/\text{time}$
- $\lambda(t)$
- $\lambda(t|H_t)$

# Neural coding of firing rate may take any of several forms

$$\log FR = \text{stimulus effects} + \text{covariate effects}$$

$FR$ defined as

- $E(\text{count})/\text{time}$
- $\lambda(t)$
- $\lambda(t|H_t)$

Kass and Ventura (2001, *Neural Computation*)
Koyama and Kass (2008, *Neural Computation*)

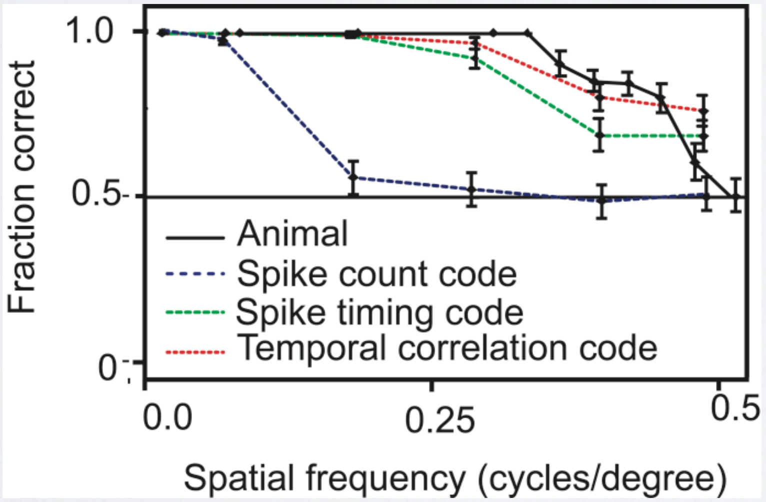
# Neural coding of firing rate may take any of several forms

$$\log FR = \text{stimulus effects} + \text{covariate effects}$$

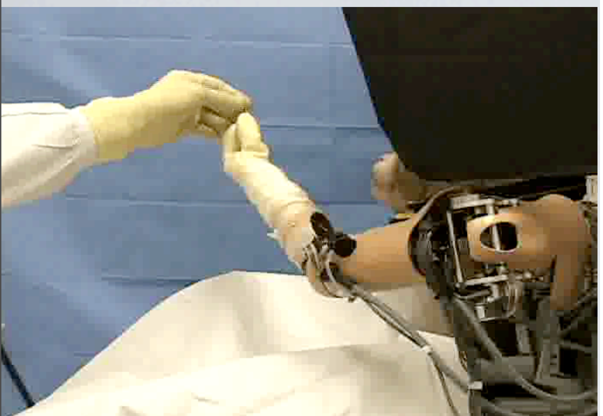
$FR$ defined as

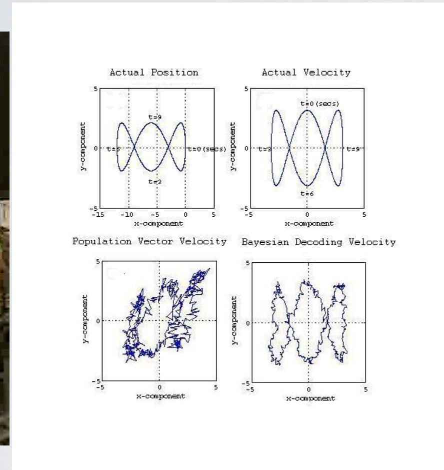
- $E(\text{count})/\text{time}$
- $\lambda(t)$
- $\lambda(t|H_t)$

Kass and Ventura (2001, *Neural Computation*)
Koyama and Kass (2008, *Neural Computation*)

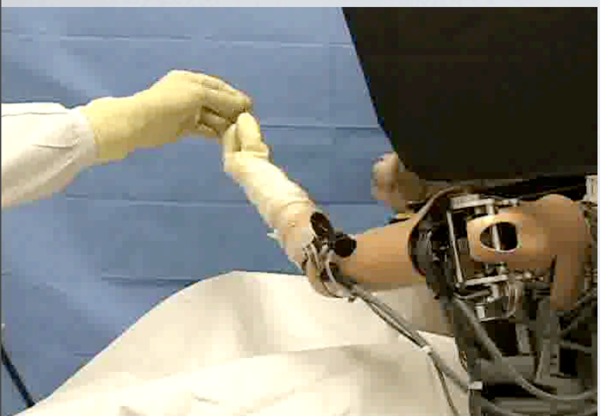


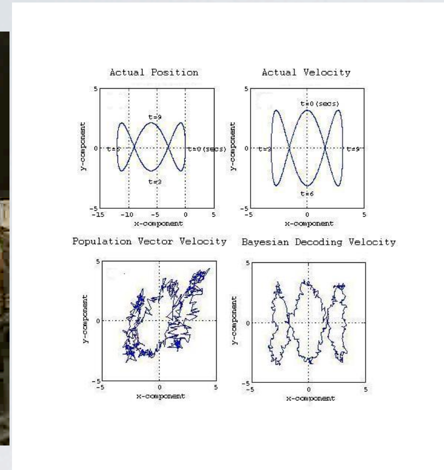
# Neural prosthetics: neural coding and decoding





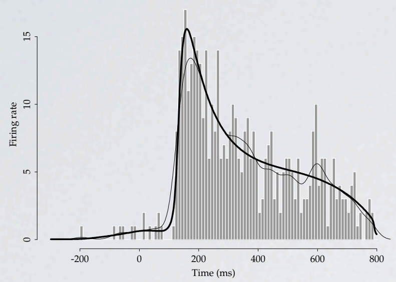
# Neural prosthetics: neural coding and decoding



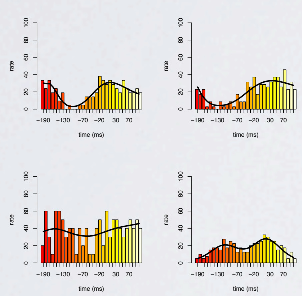


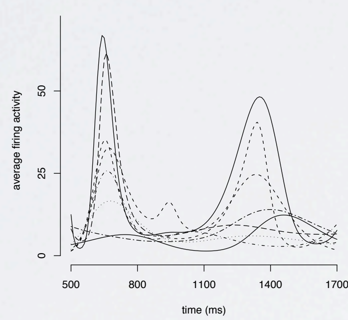

Brockwell, Rojas, Kass (2004, *J. Neurophsyiology*)
Brockwell, Kass, and Schwartz (2007, *Proc. IEEE*)
Koyama, Castellanos Perez-Bolde, Shalizi, and Kass (2010, *JASA*)

# Smoothing of firing-rate functions via BARS

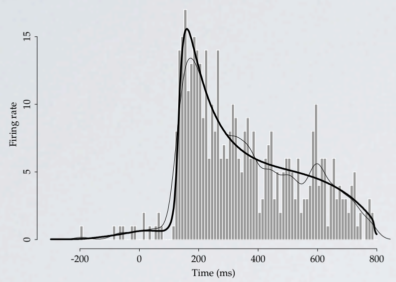



DiMatteo, Genovese, and Kass (2001, *Biometrika*)
Kass, Ventura, Cai (2003, *Network*)
Wallstrom, Liebner, and Kass (2008, *J. Comput. Software*)

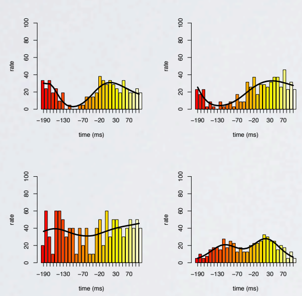


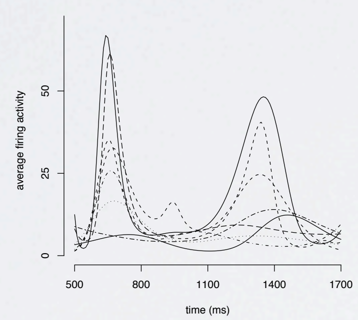


# Smoothing of firing-rate functions via BARS



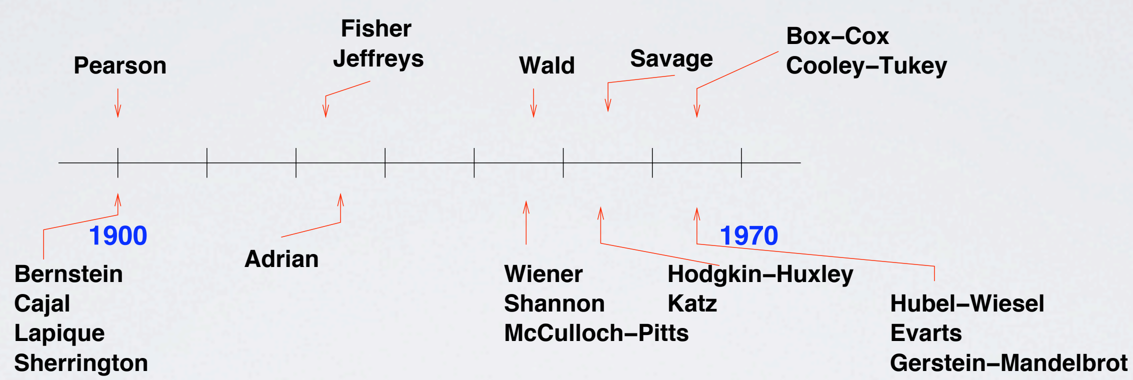

DiMatteo, Genovese, and Kass (2001, *Biometrika*)
Kass, Ventura, Cai (2003, *Network*)
Wallstrom, Liebner, and Kass (2008, *J. Comput. Software*)





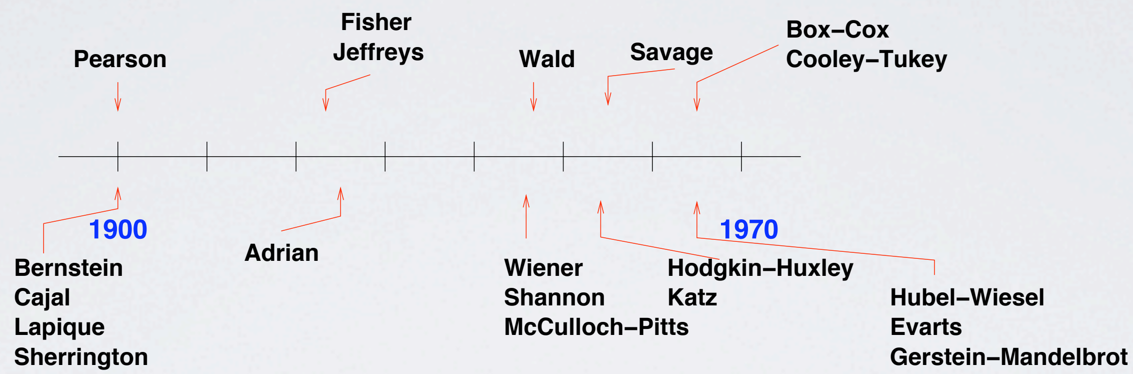

Behseta, Kass, Wallstrom (2005, *Biometrika*)
Behseta and Kass (2005, *Statistics in Medicine*)
Behseta, Moorman, Olson, and Kass (2007, *Statistics in Medicine*)

# Foundations for spike train analysis in place since ~ 1970




# Rapid progress in past ~10 years

# Foundations for spike train analysis in place since ~ 1970




# Rapid progress in past ~10 years

Brown, Kass, Mitra (2004, *Nature Neuroscience*)

Kass, Ventura, Brown (2005, *J. Neurophysiology*)

# “Statistical Thinking”

# “Statistical Thinking”

Brown and Kass, “What is Statistics?” (2009, *Amer. Statist*.)

# “Statistical Thinking”

Brown and Kass, “What is Statistics?” (2009, *Amer. Statist*.)

1. Use statistical models:
   variation ---> knowledge and uncertainty

2. Analyze procedures

# “Statistical Thinking”

## Correction for attenuation of correlation in spike counts

$$
\begin{aligned}
r &= .49 \\
\hat{\rho} &= .82 \\
&\phantom{=} (.77, .88)
\end{aligned}
$$

# “Statistical Thinking”

Correction for attenuation of correlation in spike counts

$$
\begin{aligned}
r &= .49 \\
\hat{\rho} &= .82 \\
& \quad (.77, .88)
\end{aligned}
$$

Behseta, Berdyyeva, Olson, and Kass (2009, *J. Neurophysiology*)

# “Statistical Thinking”

Brown and Kass, “What is Statistics?” (2009, *Amer. Statist*.)

# “Statistical Thinking”

Brown and Kass, “What is Statistics?” (2009, *Amer. Statist*.)

1. Use statistical models:
   variation ---> knowledge and uncertainty

2. Analyze procedures

# “Statistical Thinking”

Brown and Kass, “What is Statistics?” (2009, *Amer. Statist*.)

1. Use statistical models:
   variation ---> knowledge and uncertainty

2. Analyze procedures

*something else ...*

# Aesthetics

loglinear models:

$$\begin{aligned}\log FR &= \text{stimulus effects} + \text{coupling effects} \\ &+ \text{history effects} + \text{global network effects}\end{aligned}$$

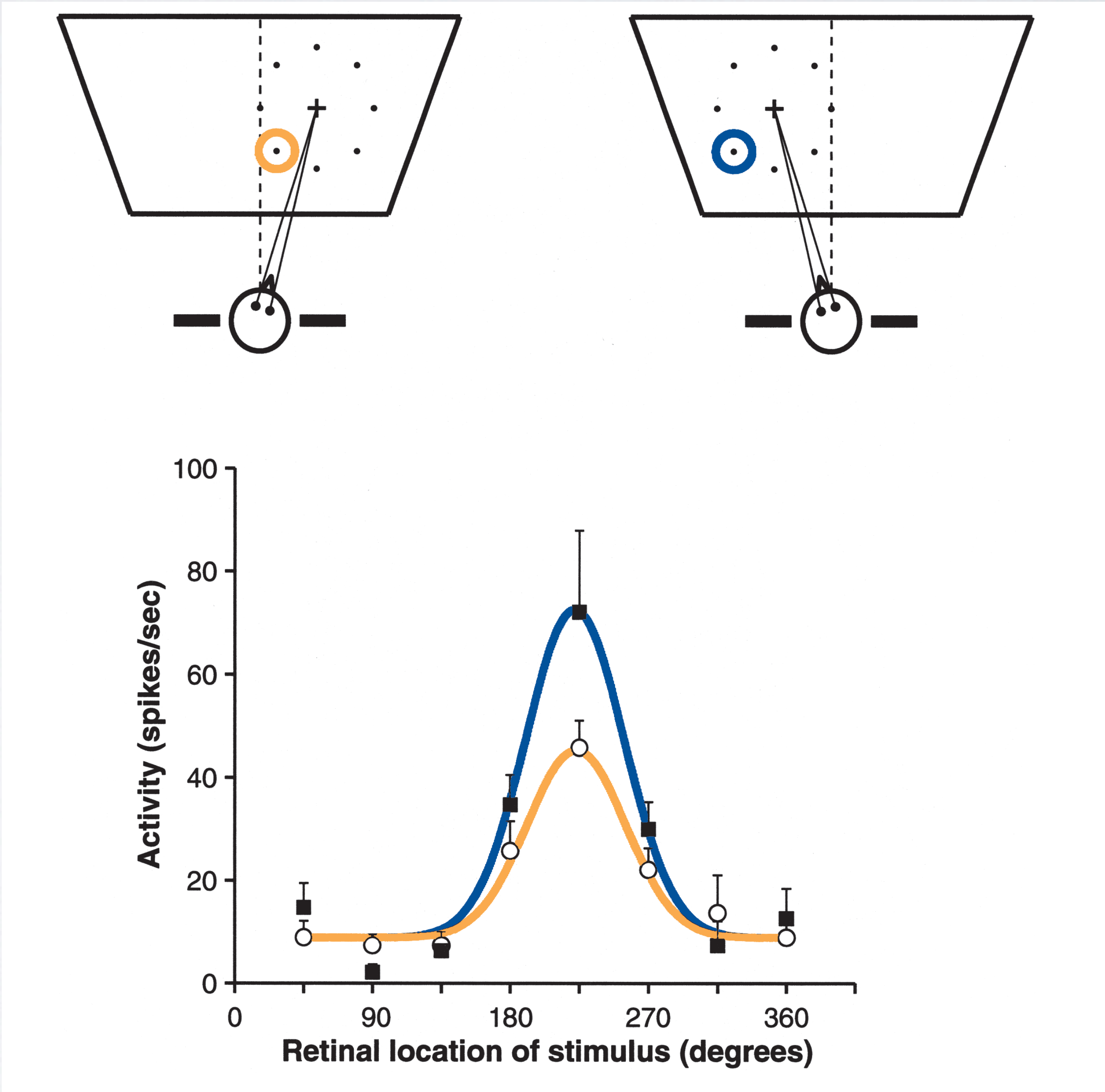
multiplicative firing-rate effects

# Aesthetics

## loglinear models:

$$\log FR = \text{stimulus effects} + \text{coupling effects} + \text{history effects} + \text{global network effects}$$

## multiplicative firing-rate effects




Salinas and Sejnowski (2001)

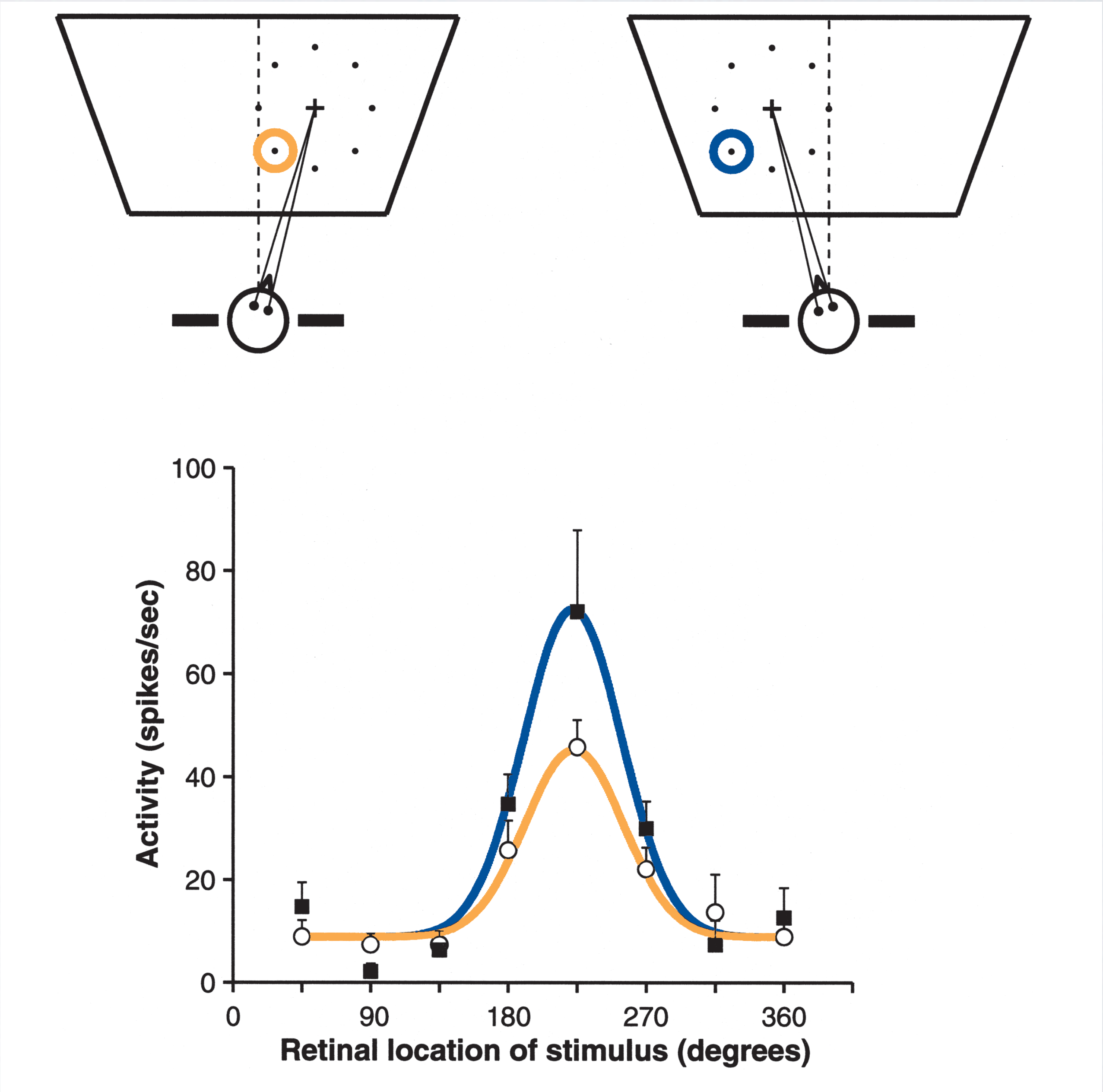
# Aesthetics

## loglinear models:

$$\begin{aligned} \log FR &= \text{stimulus effects} + \text{coupling effects} \\ &+ \text{history effects} + \text{global network effects} \end{aligned}$$

## multiplicative firing-rate effects

## gain modulation



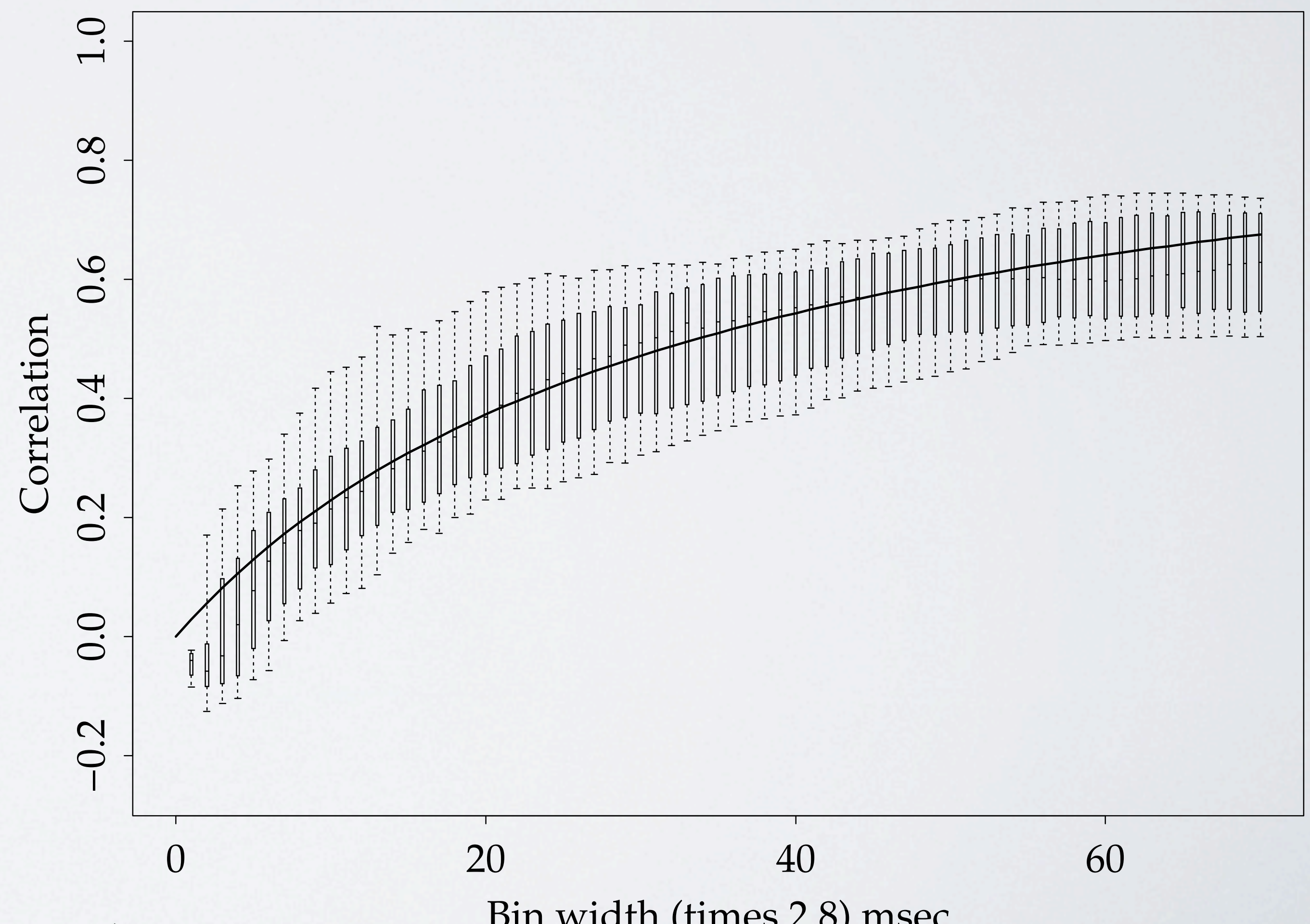

Salinas and Sejnowski (2001)

# Aesthetics

loglinear models:

multiplicative firing-rate effects

increase in correlation of spike counts with increasing bin size




Kass and Ventura (2006, *Neural Computation*)

# CLAIM:

$$\text{computational neuroscience} = \text{theoretical neuroscience} \cup \text{analysis of neural data}$$

# CLAIM:

$$\text{computational neuroscience} = \text{theoretical neuroscience} \cup \text{analysis of neural data}$$

how do statistical models fit in?

# CLAIM:

| computational neuroscience | = | theoretical neuroscience | $\cup$ | analysis of neural data |
|---|---|---|---|---|

how do statistical models fit in?

1. They are important for analysis of neural data

2. Perhaps they also fill much of the intersection

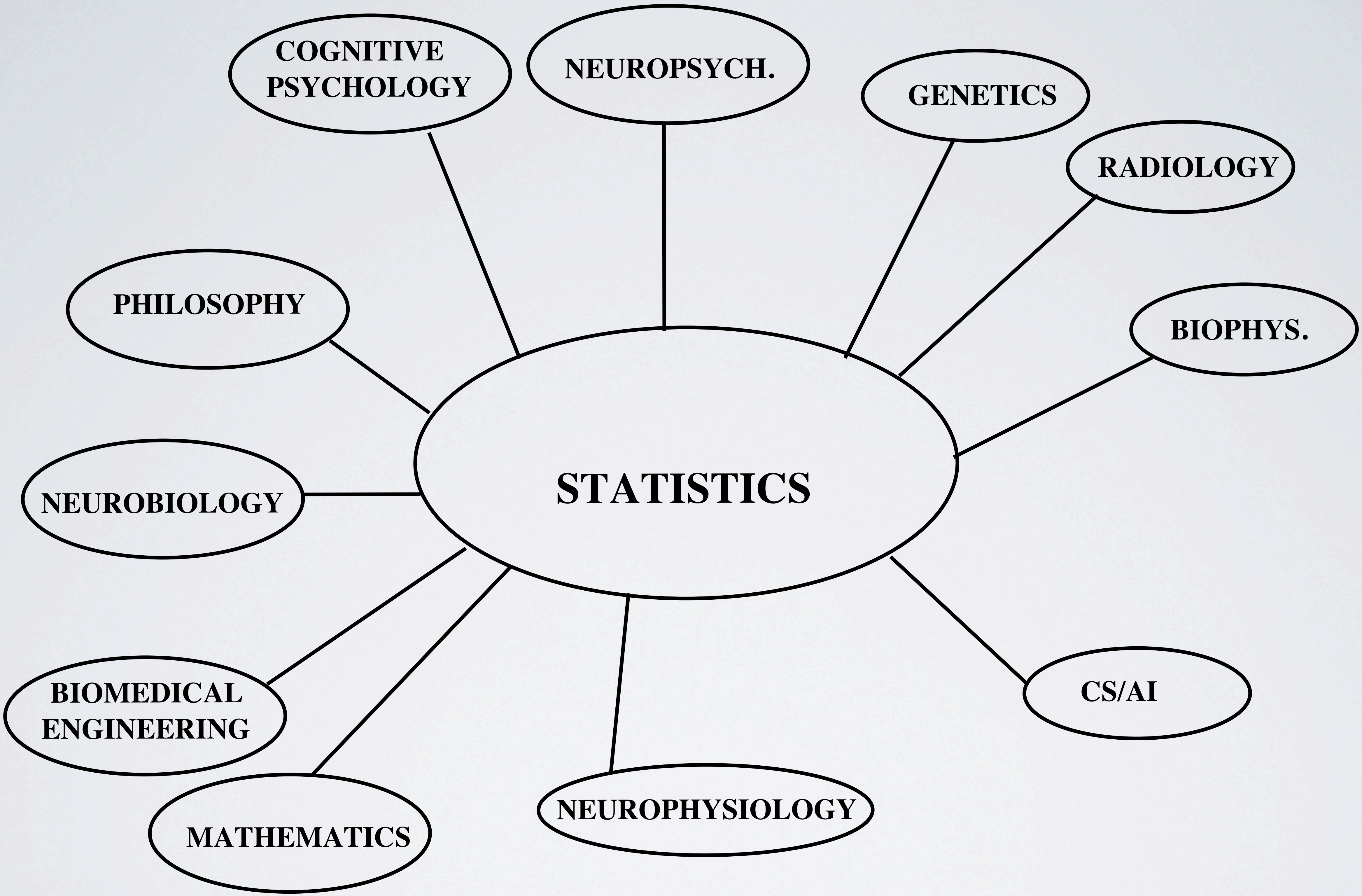
# A final picture ...

# Many diverse disciplines

biology
neurophysiology
psychology
neurology
psychiatry
radiology
mathematics
physics
bioengineering
computer science
statistics

# A Statistician's View of Cognitive Neuroscience

# A Statistician's View of Cognitive Neuroscience



## A Statistician's View of Cognitive Neuroscience