Simple Measures of Visual Encoding

vs.

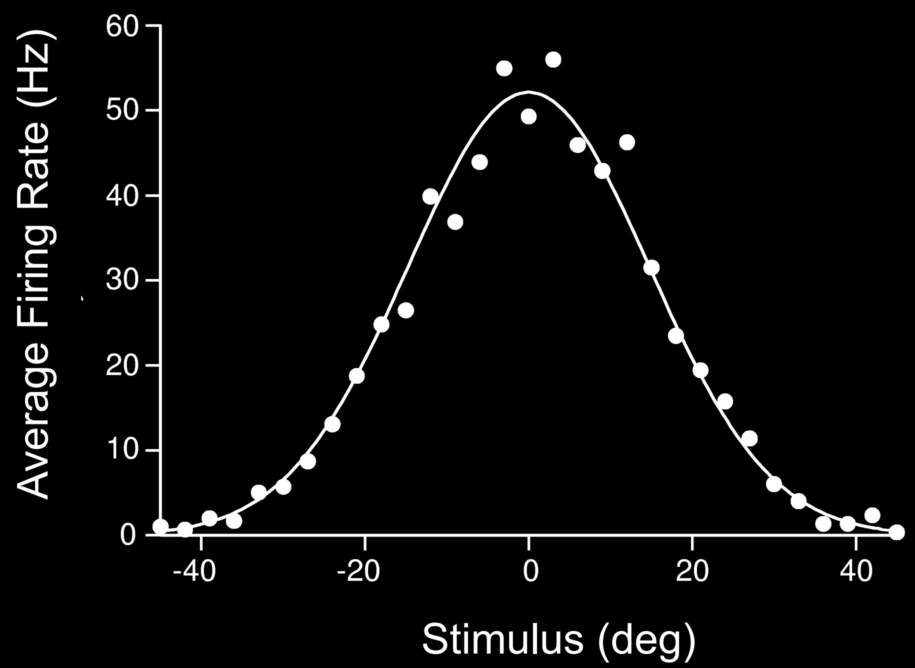
Information Theory

Simple Measures of Visual Encoding

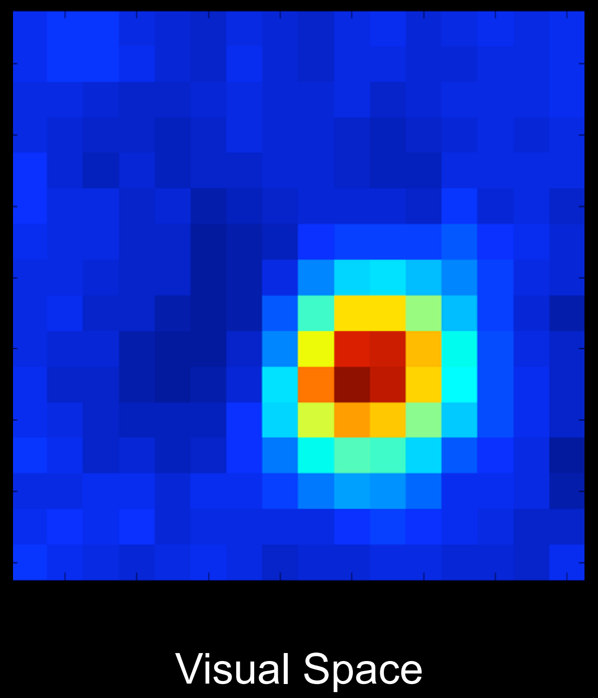
STIMULUS → □ → RESPONSE

What does a [visual] neuron do?

Tuning Curves

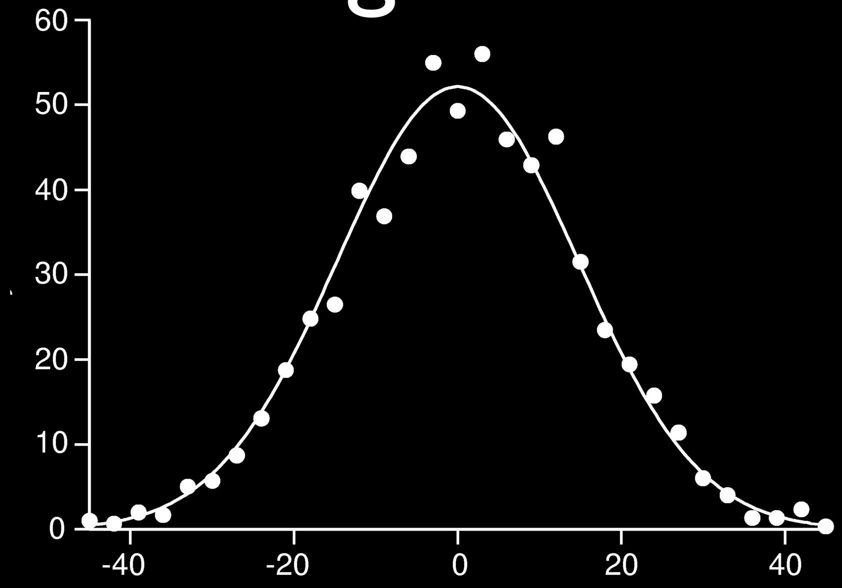


Receptive Fields



Neurons as information encoders

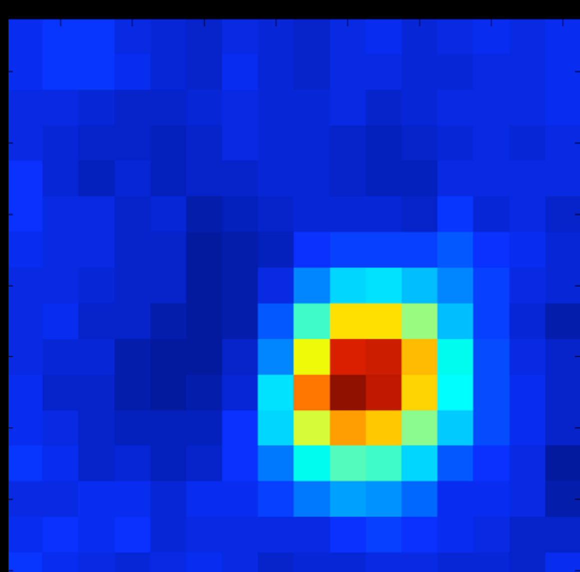
Tuning Curves



What stimuli are best-encoded by the neuron?

Stimuli a neuron respond to may not be best-encoded (Fisher Info)

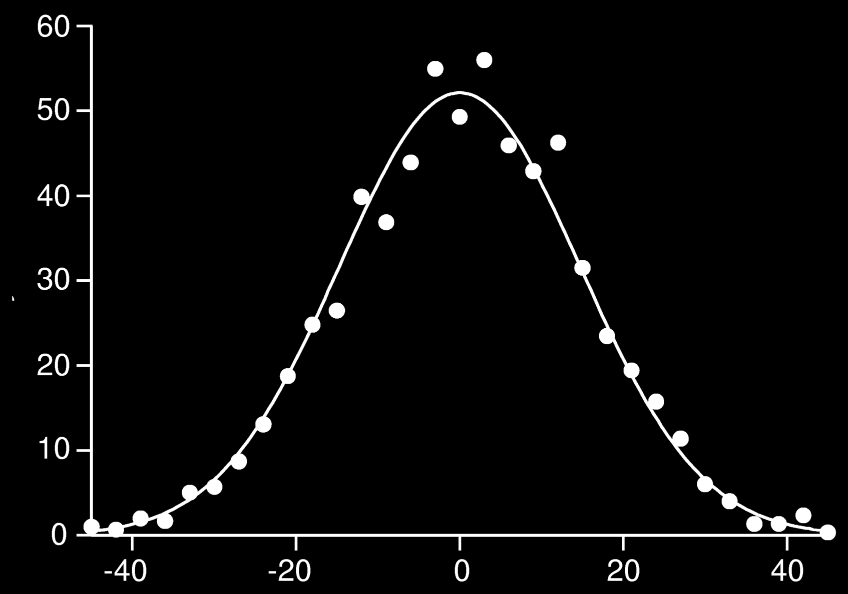
Receptive Fields

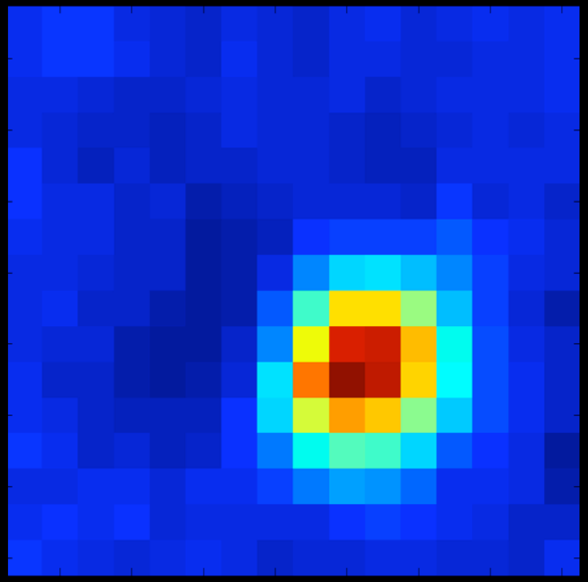


How are STRFs related to info in spike trains?

Spiking precision of neuron (and Shannon info) not predicted by RF alone

GOAL: Link information measures to neural function





Mark Goldman
(Wellesley College)

Garrett Stanley
(Harvard University)

Also thanks to KITP crew for explaining Fisher info and many other things

What does a neuron respond to?

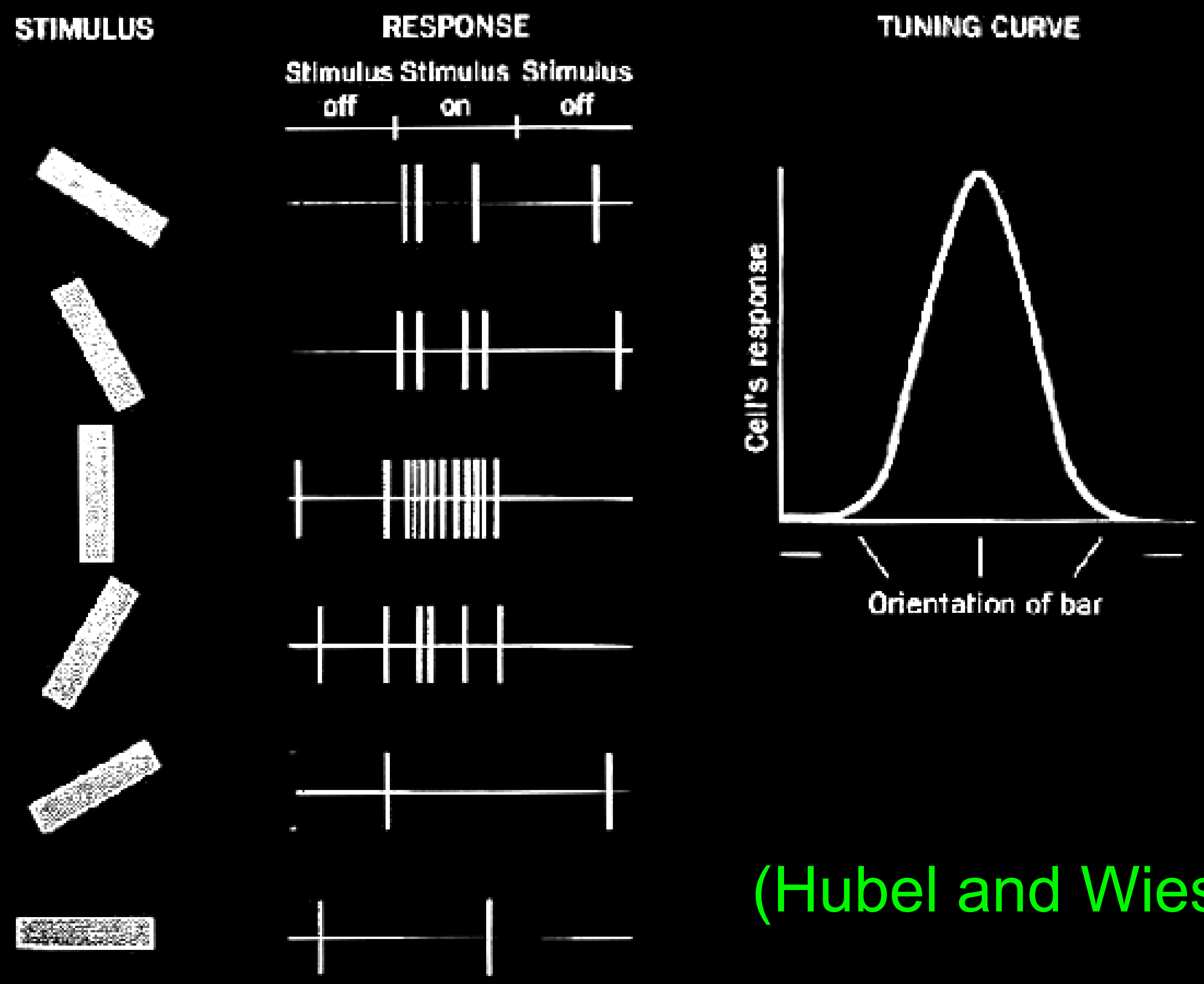
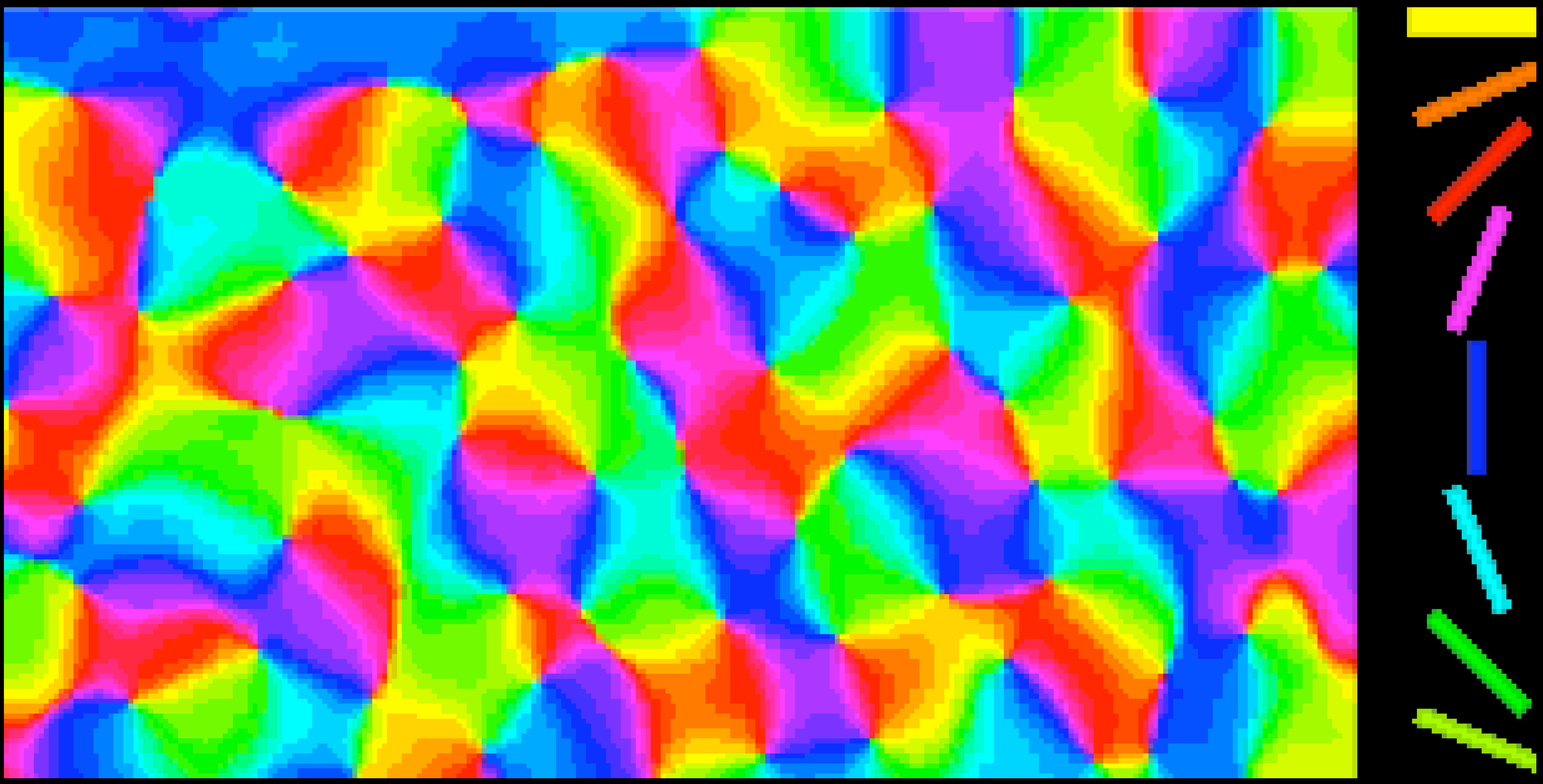


FIGURE 4.8 Response of a single cortical cell to bars presented at various orientations.

(Hubel and Wiesel, 1959)

Jumping to conclusions?

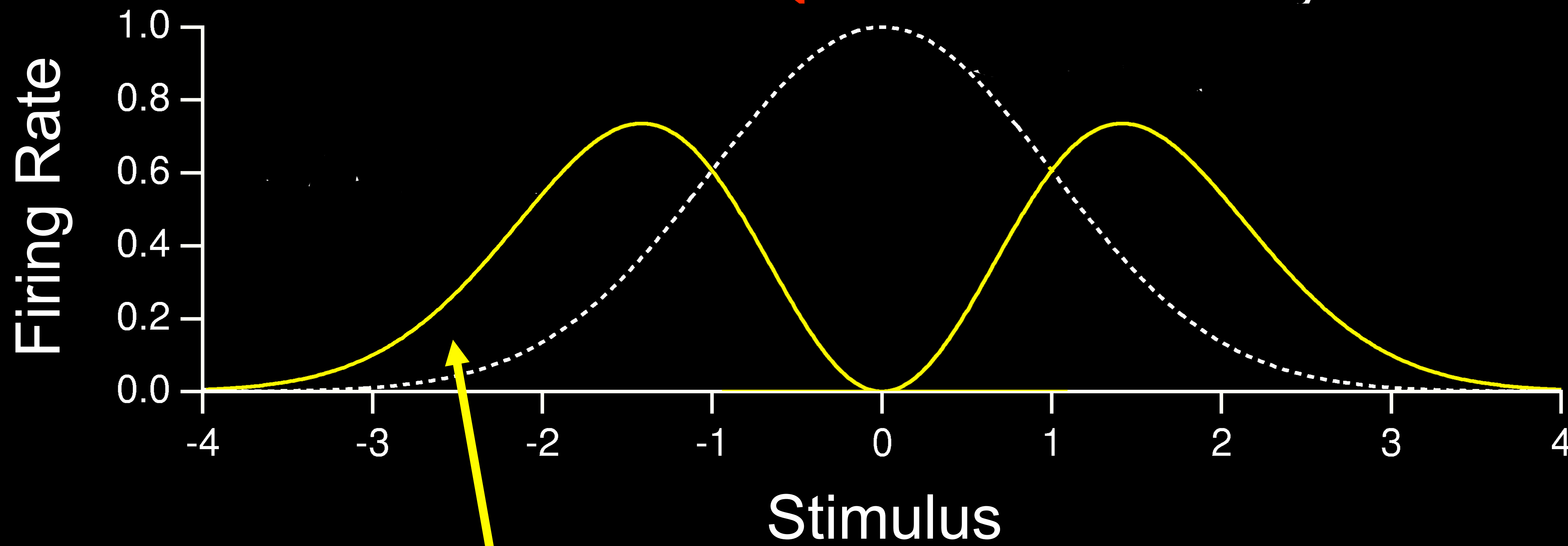
"Functional Map" of the V1 (Visual Cortex)



Discriminability and Fisher Info

Slope is zero at the peak firing rate

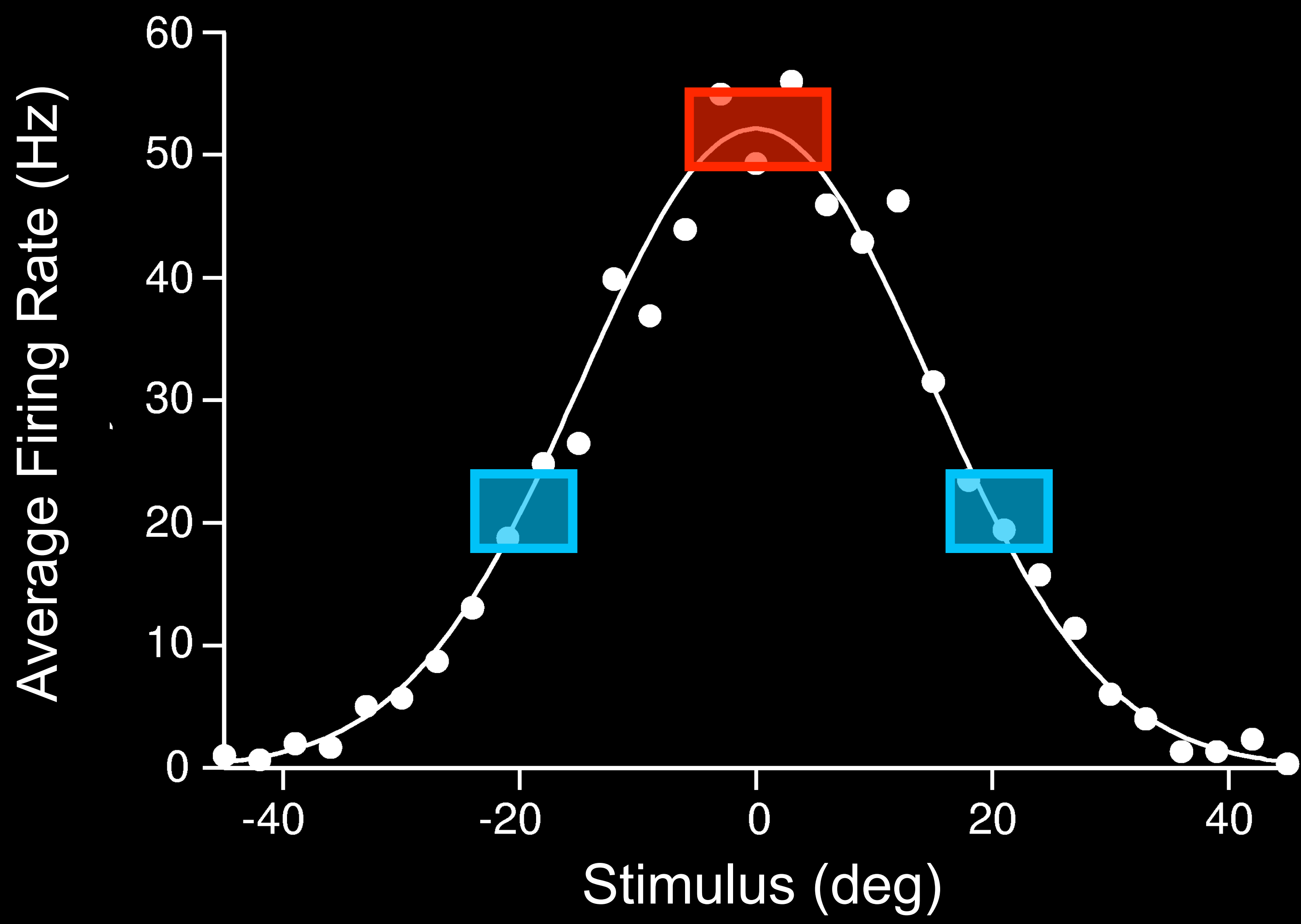
from Dayan & Abbott



Fisher Information

$$J[\theta] = -\left\langle \left[\frac{\partial}{\partial \theta} \log p(r|\theta) \right]^2 \right\rangle_r = \frac{1}{\sigma^2} [f'(\theta)]^2$$

What is a well-encoded stimulus?



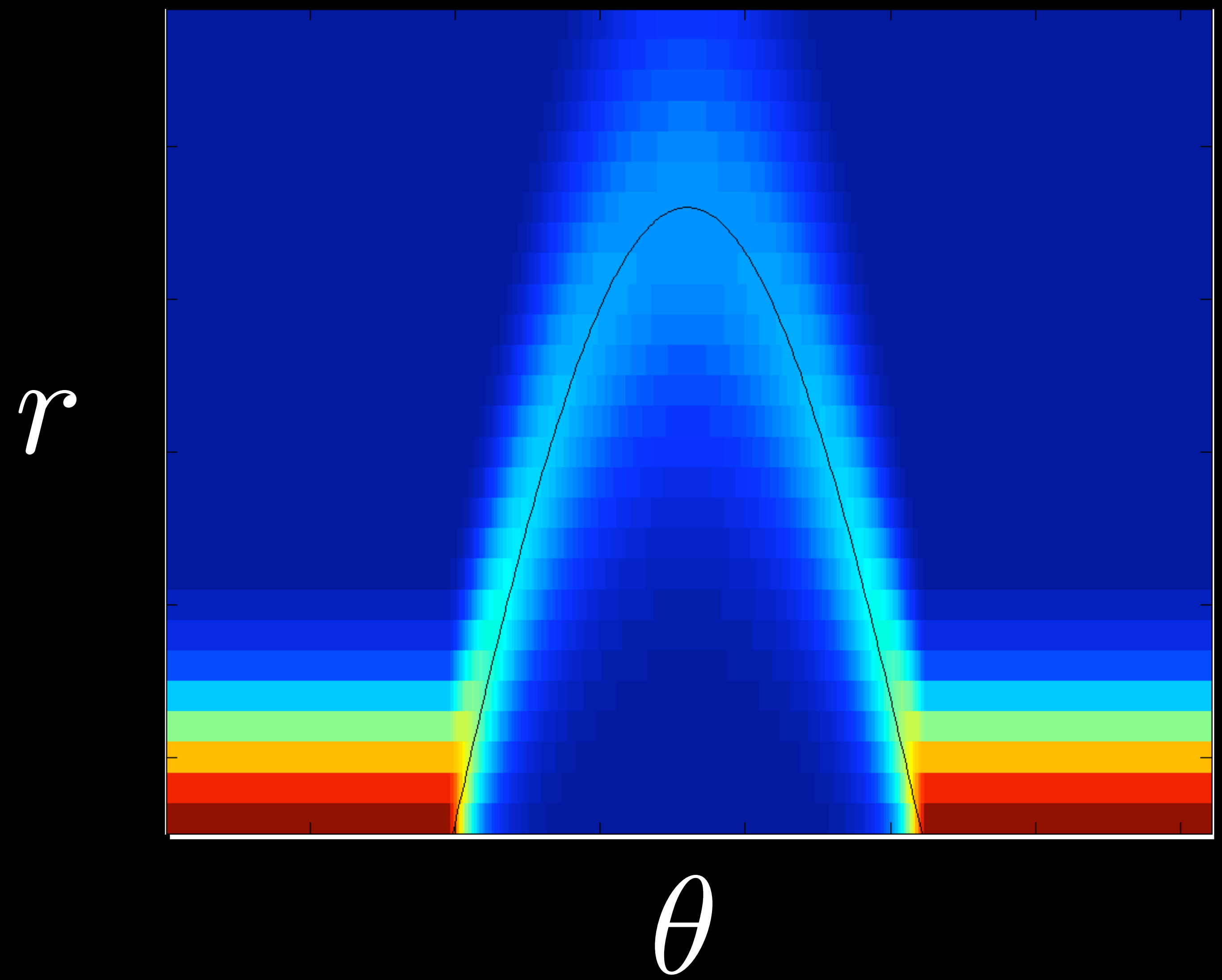
→ Stimuli that make the neuron fire the most?

→ Stimuli to which a neuron's firing rate is most sensitive?

Spikes convey info

Other spike train properties (?)

“Model independence”



How to simplify the JPD?

Mutual information

Receptive fields

....

Decomposing the mutual information

$$I[\Theta, R] = \sum_{\theta} \sum_{r} p(\theta, r) \log_2 \left[\frac{p(\theta, r)}{p(\theta) p(r)} \right]$$

Just a number

Decompose *I*[*R*,*S*] into stimulus- or response-specific quantities:

$$I[R,S] = \sum_{s \in S} p(s) i(s) \quad = \sum_{r \in R} p(r) i(r)$$

...that represent the contribution of particular symbols to *I*[*R*,*S*].

Many ways to decompose *I*[*R*,*S*] ⇒ Choice of 'i' depends on what you want to mean.

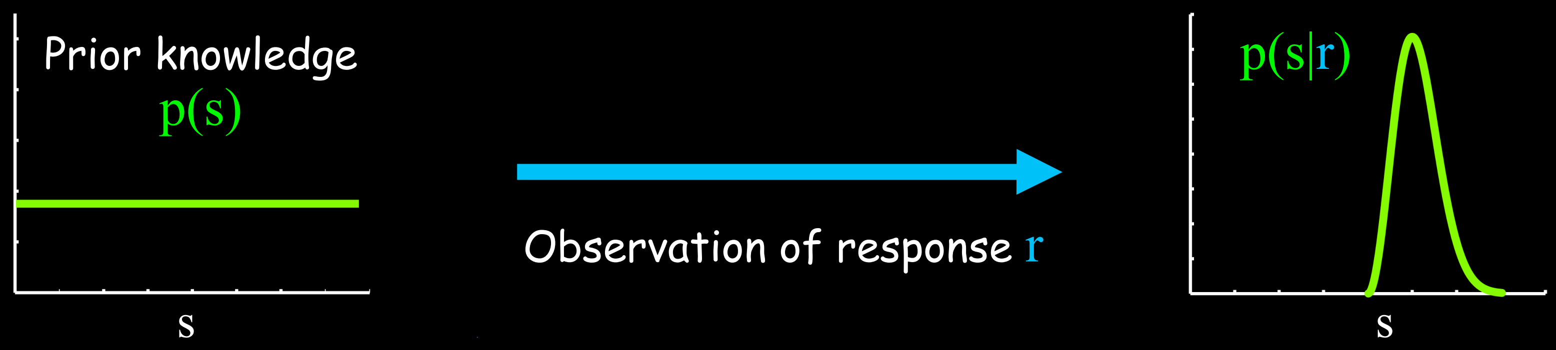
Investigated in DeWeese and Meister (1999)

Specific Information

~~Specific Surprise~~

Information measure of a response

$I[S,R]$ <
- S - ensemble of possible stimuli $p(s)$
- R - ensemble of possible responses



Uncertainty in the stimulus ensemble → ENTROPY $H[S]$

$$H[S] = -\sum_{s \in S} p(s) \log_2 p(s) \qquad H[S|r] = -\sum_{s \in S} p(s|r) \log_2 p(s|r)$$

$$H[S] - H[S|r]$$

reduction in uncertainty in the stimulus ensemble

Specific information is the appropriate decomposition for responses

$$i_{sp}(r) = H[S] - H[S|r]$$

Specific information is the only additive decomposition.

$$i_{sp}(r_1, r_2) = i_{sp}(r_1) + i_{sp}(r_2|r_1)$$

$$I[R,S] = \sum_{r \in R} p(r) i_{sp}(r)$$

Mutual information is the average reduction in uncertainty from any response.

Is specific information the "best" decomposition for stimuli?

$$i_{sp}(s) = H[R] - H[R|s]$$

Which is the best-encoded stimulus?

Joint Probability Distribution
$p(r,s)$

	r_1	r_2
s_2	1/4	0
s_1	1/4	1/2

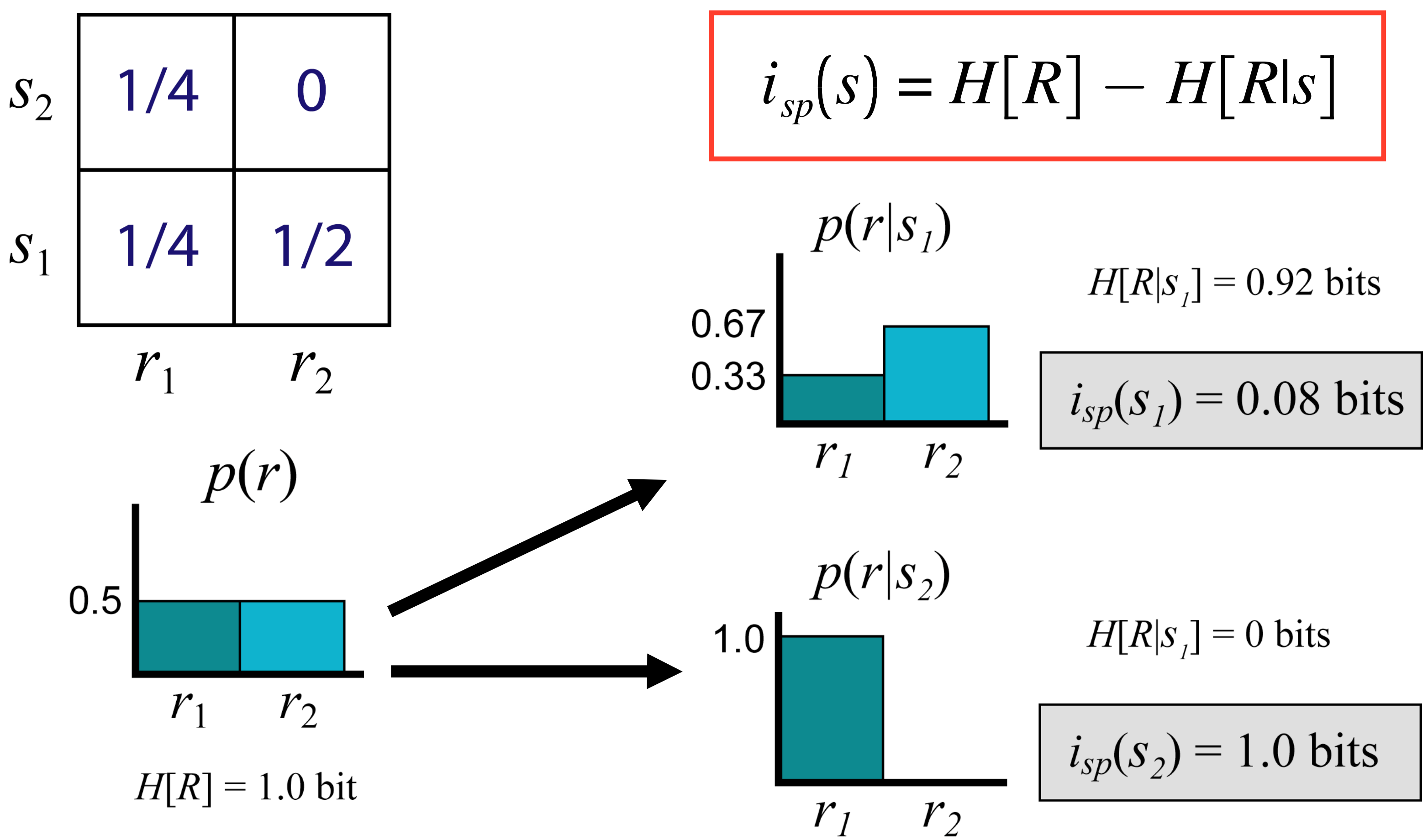
1. Equal probability (1/2) of measuring each response

2. If r_1 is measured:
Equal chance that the stimulus is s_1 or s_2

3. If r_2 is measured:
The stimulus must be s_1

→ s_1 is better encoded than s_2

Specific information selects stimuli associated with few responses



Causality breaks the symmetry of specific measures

$i_{sp}(r) = H[S] - H[S|r]$

Response ***r*** is observed.

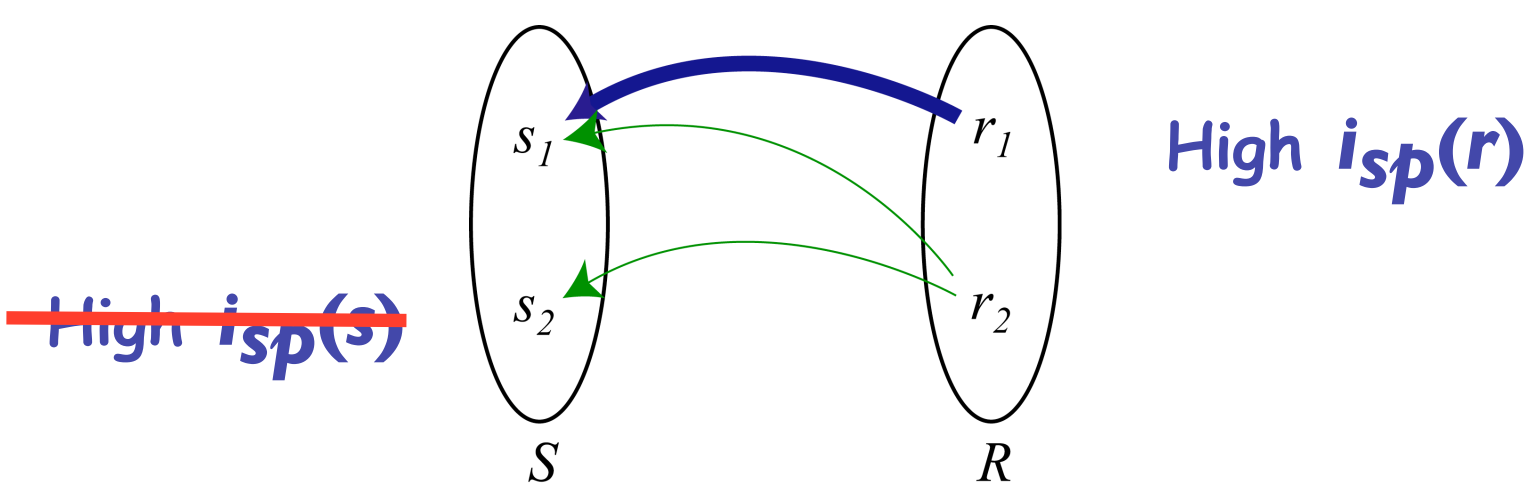
***$i_{sp}(r)$* tells you how well you know *S*.**

~~$i_{sp}(s) = H[R] - H[R|s]$~~

~~Stimulus *s* occurs.~~

~~***$i_{sp}(s)$* tells you how well you can predict *R*.**~~

Responses convey information about stimuli



A well-encoded stimulus is…

… a stimulus that is associated with informative responses.

Information of a response:

$$i_{sp}(r) = H[S] - H[S|r]$$

$i_{\mathrm{sp}}(r)$ is the reduction in uncertainty about the stimulus ensemble given a particular response *r*.

Information of a stimulus:

$$i_{SSI}(s) = \sum_{r} p(r|s) i_{sp}(r)$$

$i_{\mathrm{SSI}}(s)$ is the average reduction in uncertainty gained by a response given the presence of a particular stimulus *s*.

SSI works in the toy example…

s_2	1/4	0
s_1	1/4	1/2
	r_1	r_2

Ambiguous response!
$i_{sp}(r_1)$ = -0.19 bits

Clear response!
$i_{sp}(r_2)$ = 0.81 bits

$$i_{SSI}(s) = \sum_{r \in R} p(r|s) i_{sp}(r)$$

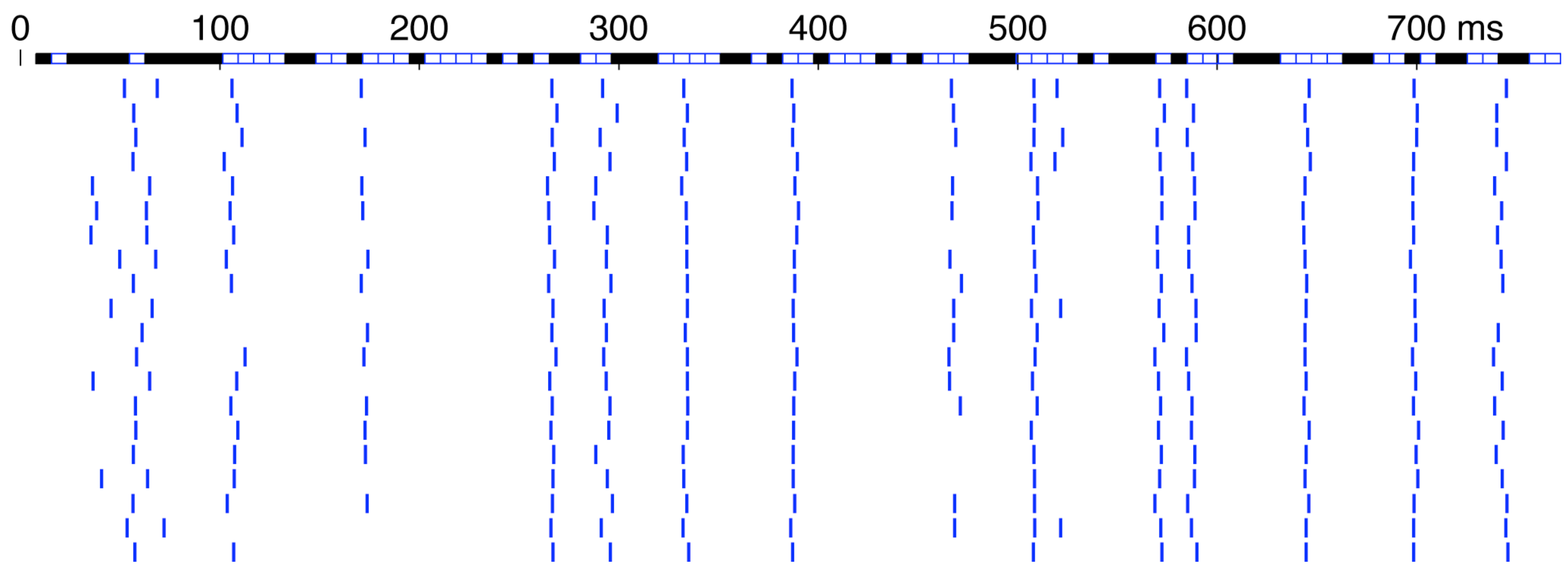
$i_{SSI}(s_2)$ = 1 (-0.19 b) + 0 (0.81 b)
= -0.19 bits

$i_{SSI}(s_1)$ = 1/3 (-0.19 b) + 2/3 (0.81 b)
= 0.48 bits

The best-encoded stimuli are those that lead to the least ambiguous responses.

Specific Application: Neurons in the Early Visual System

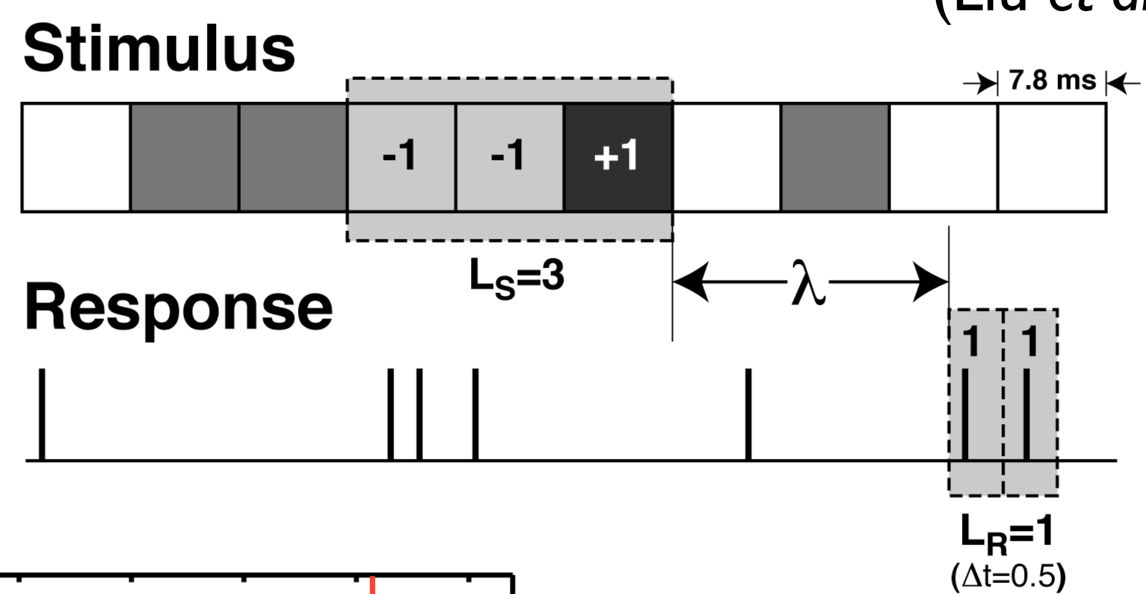
Keat, Reinagel, Reid, and Meister (2001)

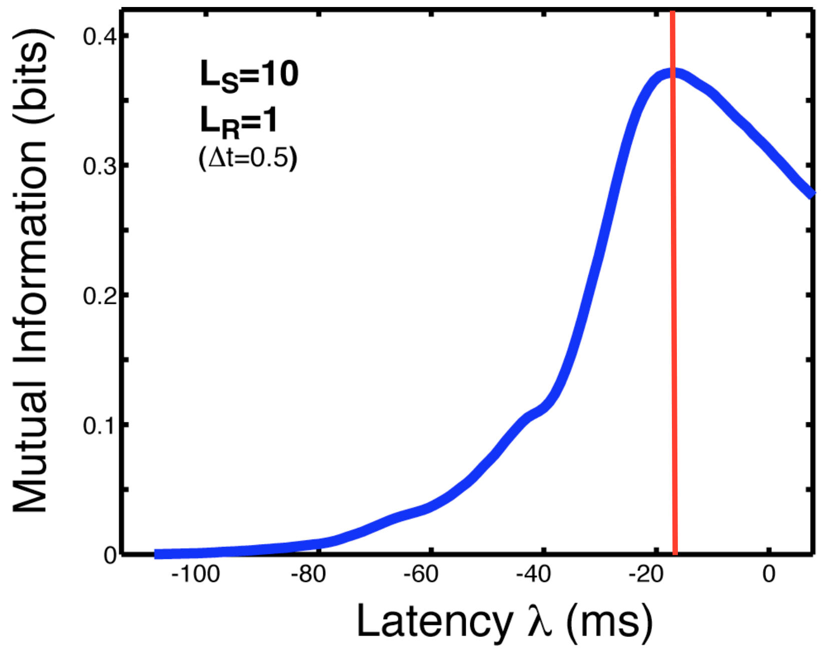


- Visual neuron responding to full-field flicker stimulus
- Realistic "data" (including trial-to-trial variability)

Mutual Information Calculation

(Liu *et al.*, 2001)

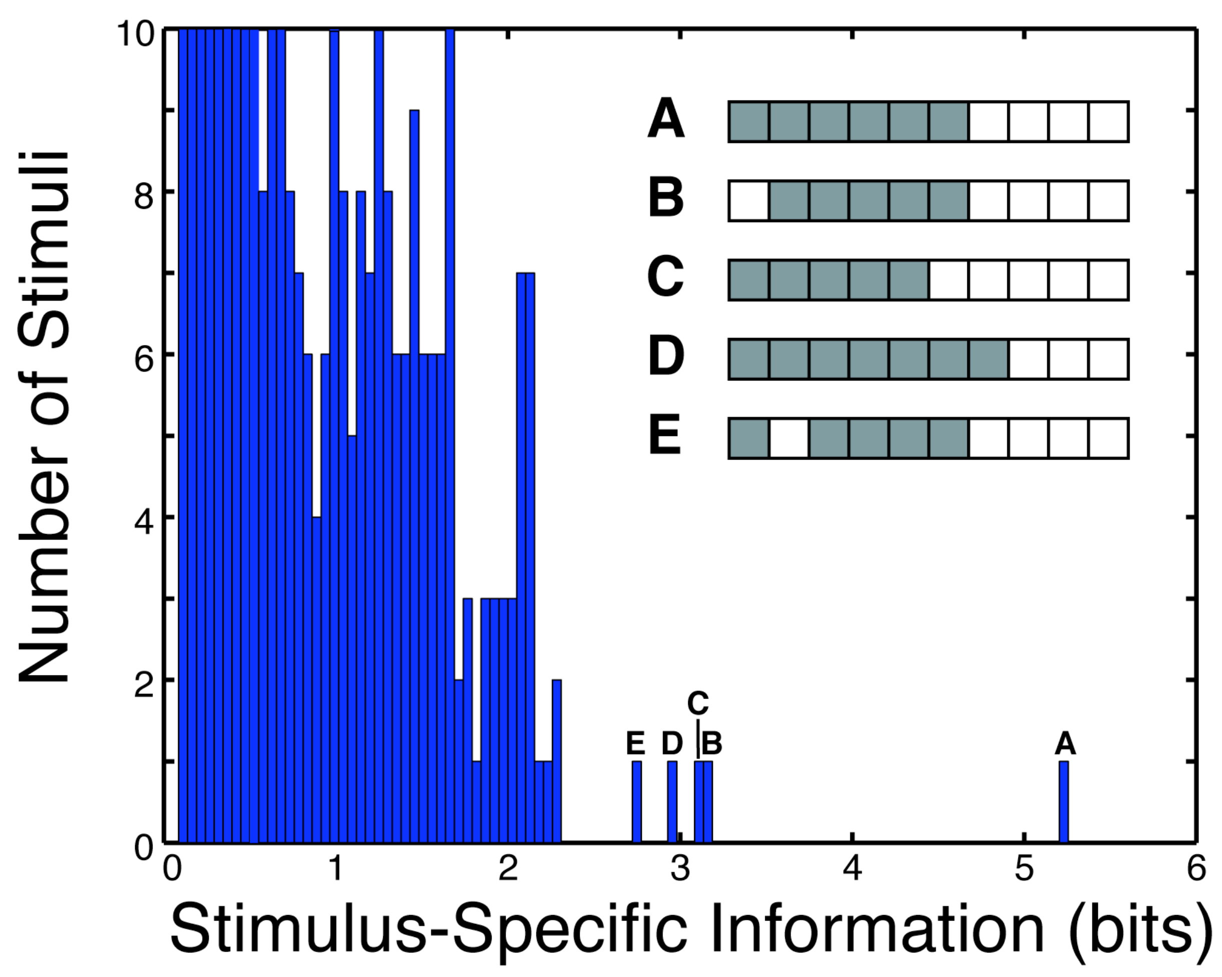




Mutual information depends on L_S, L_R, and latency.

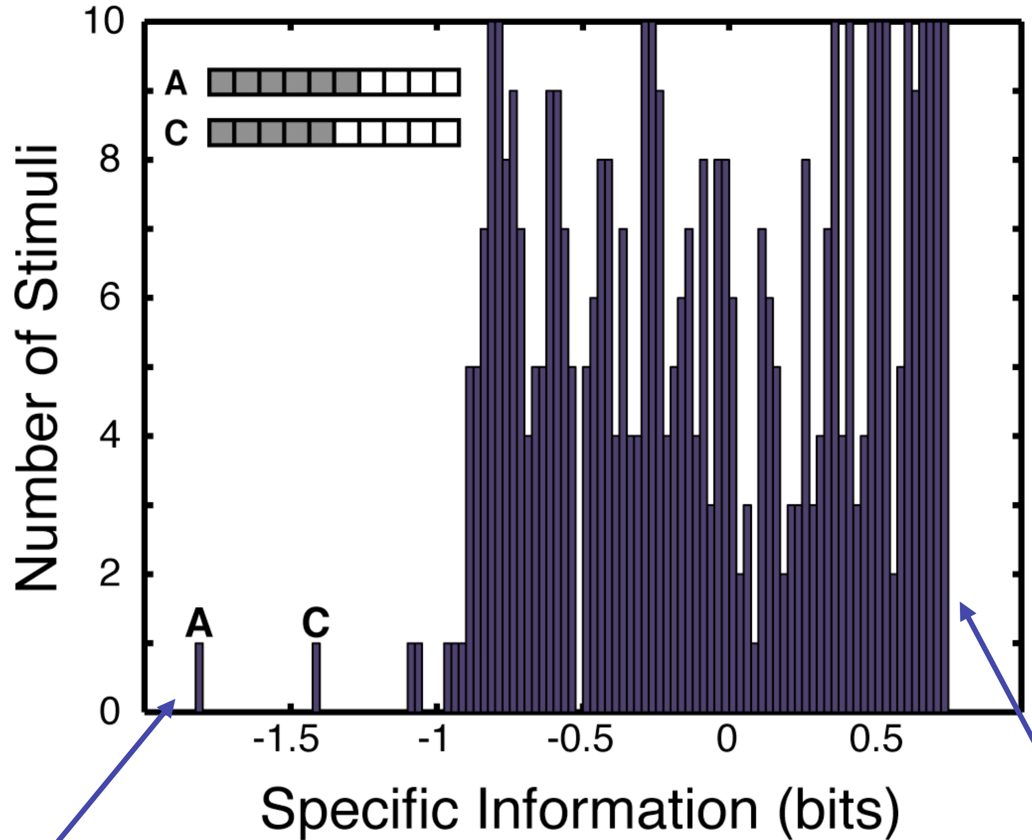
SSI of visual stimuli

$$i_{SSI}(s) = \sum_{r \in R} p(r|s)\, i_{sp}(r)$$



Specific information of visual stimuli

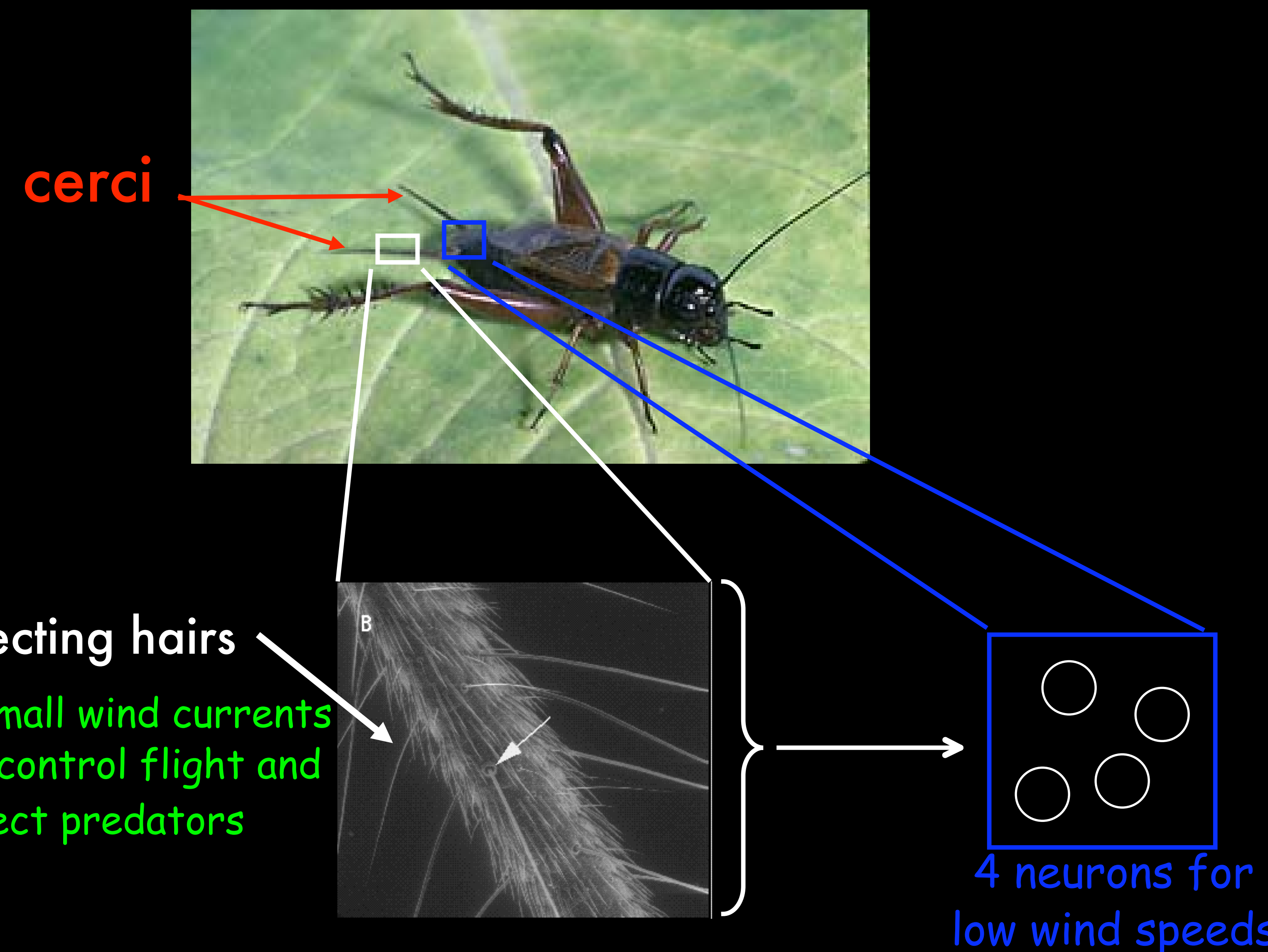
$$i_{sp}(s) = H[R] - H[R|s]$$



Stimuli that often cause spikes.

Stimuli the neuron does not respond to

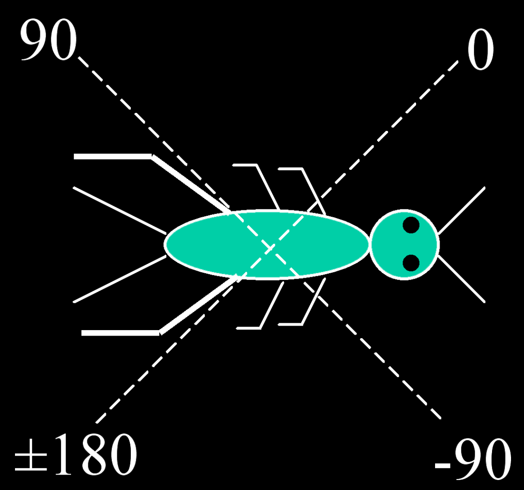
The Cricket Cercal System



Wind-detecting hairs

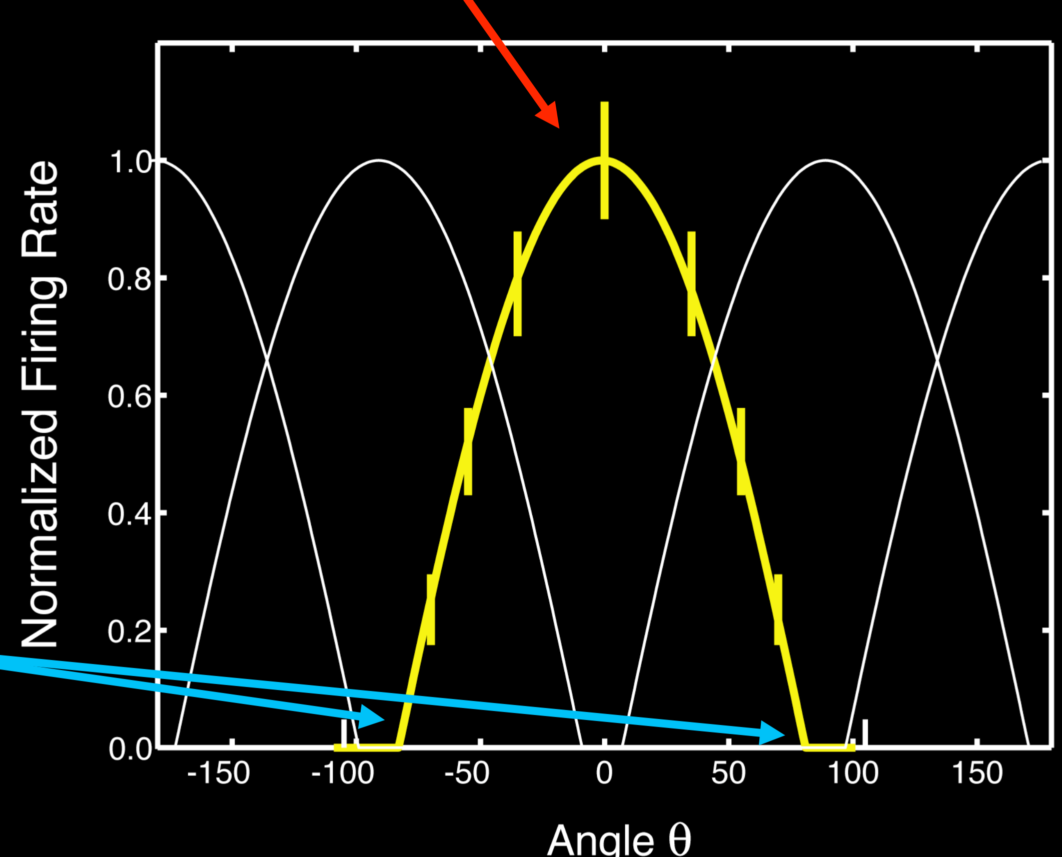
- detect small wind currents
- used to control flight and detect predators

Tuning Curves of Wind-Sensitive Neurons



Adapted from
Miller *et al.*, 1991

Do these neurons best encode stimuli at their peak firing rates?

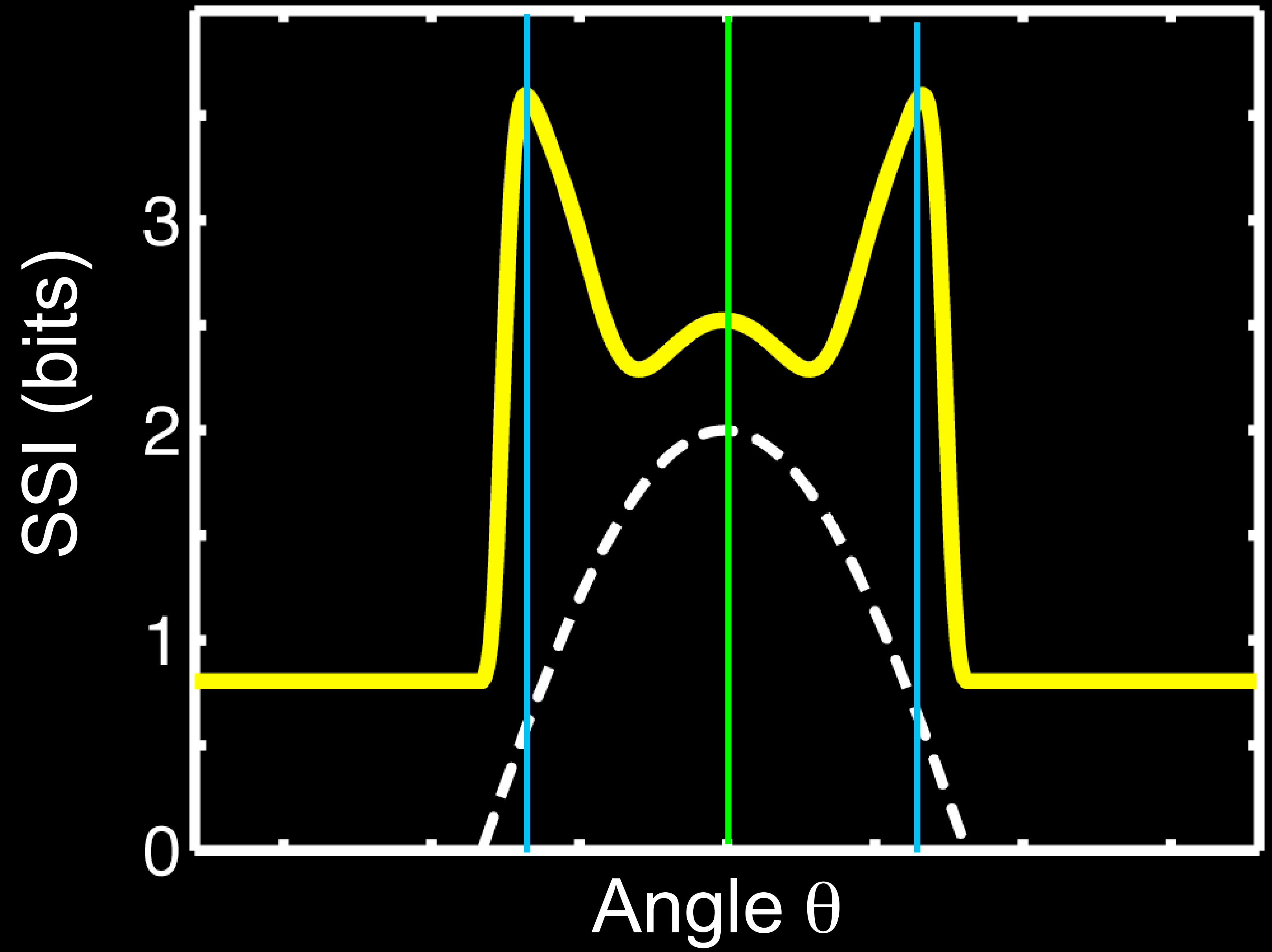


...or stimuli at the maximum slopes?

SSI for a single isolated neuron

SSI calculated directly from p(r,s) given by Miller *et al.*, 1991

Peaks are near maximum slope



1) but peak is not exactly at maximum slope

2) local maximum at peak firing rate

Which responses are informative?

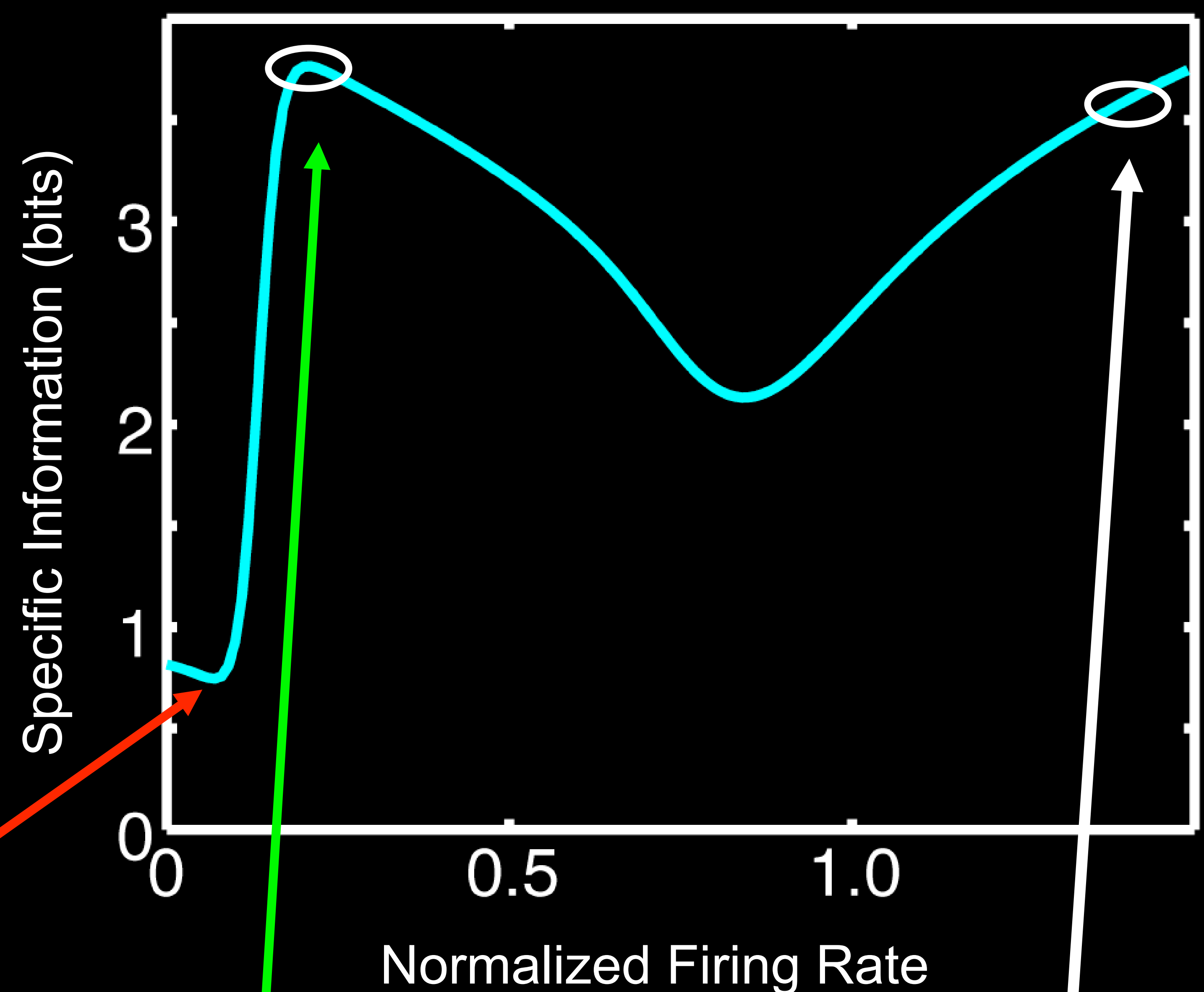
Large specific information

↓

Large reduction in uncertainty in stimulus

↓

Occurs where few stimuli correspond to given response



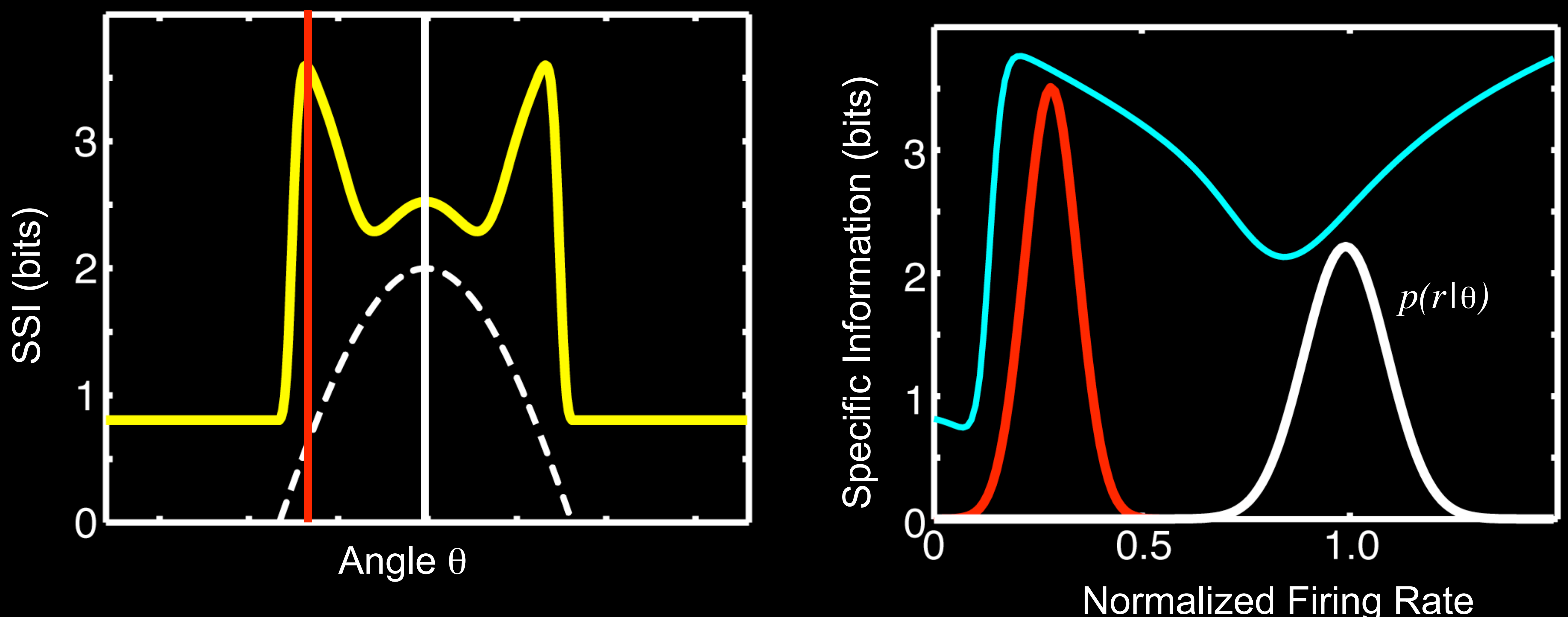
Small because many stimuli can lead to low rates (due to noise)

Large because of the tuning curve slope

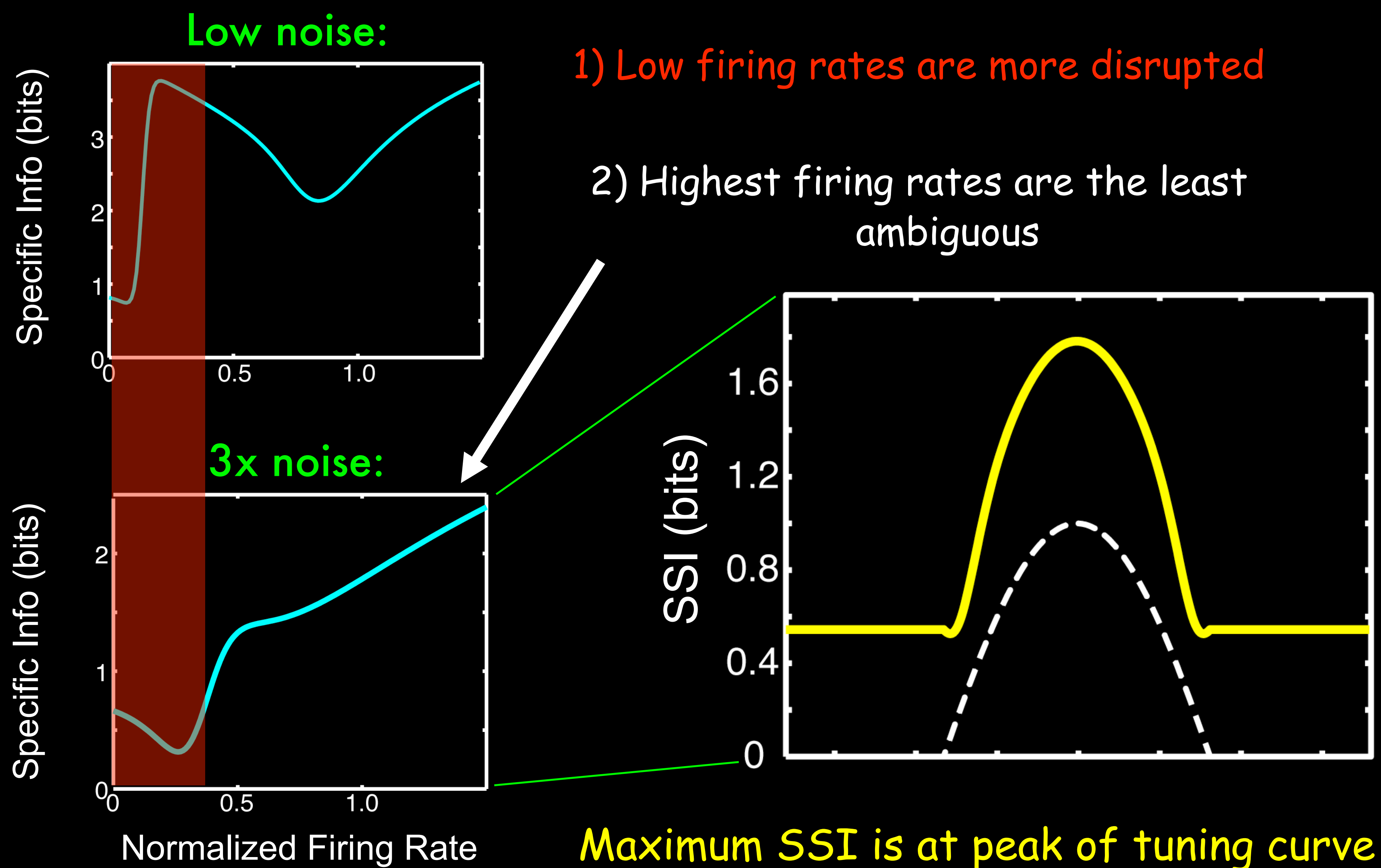
Large because few stimuli can lead to very high rates

Explanation for SSI curve shape

$$i_{SSI}(s) = \sum_{r \in R} p(r|s)\, i_{sp}(r)$$ = average specific information for a given stimulus



Increasing the noise leads to a transition

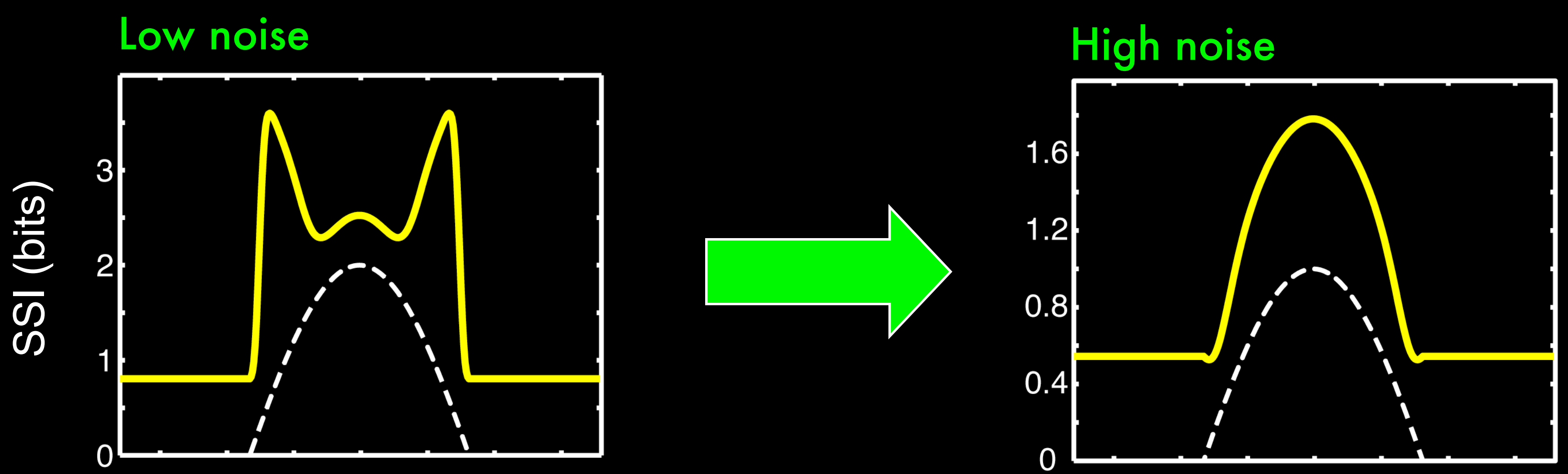


1) Low firing rates are more disrupted

2) Highest firing rates are the least ambiguous

Maximum SSI is at peak of tuning curve

Neuron Variability Causes a Transition



The meaning of a tuning curve can be dramatically different for single neurons.

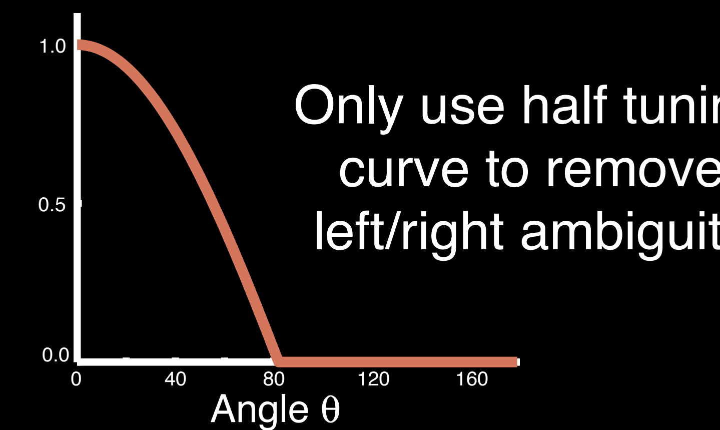
1. Why does this disagree with Fisher information?

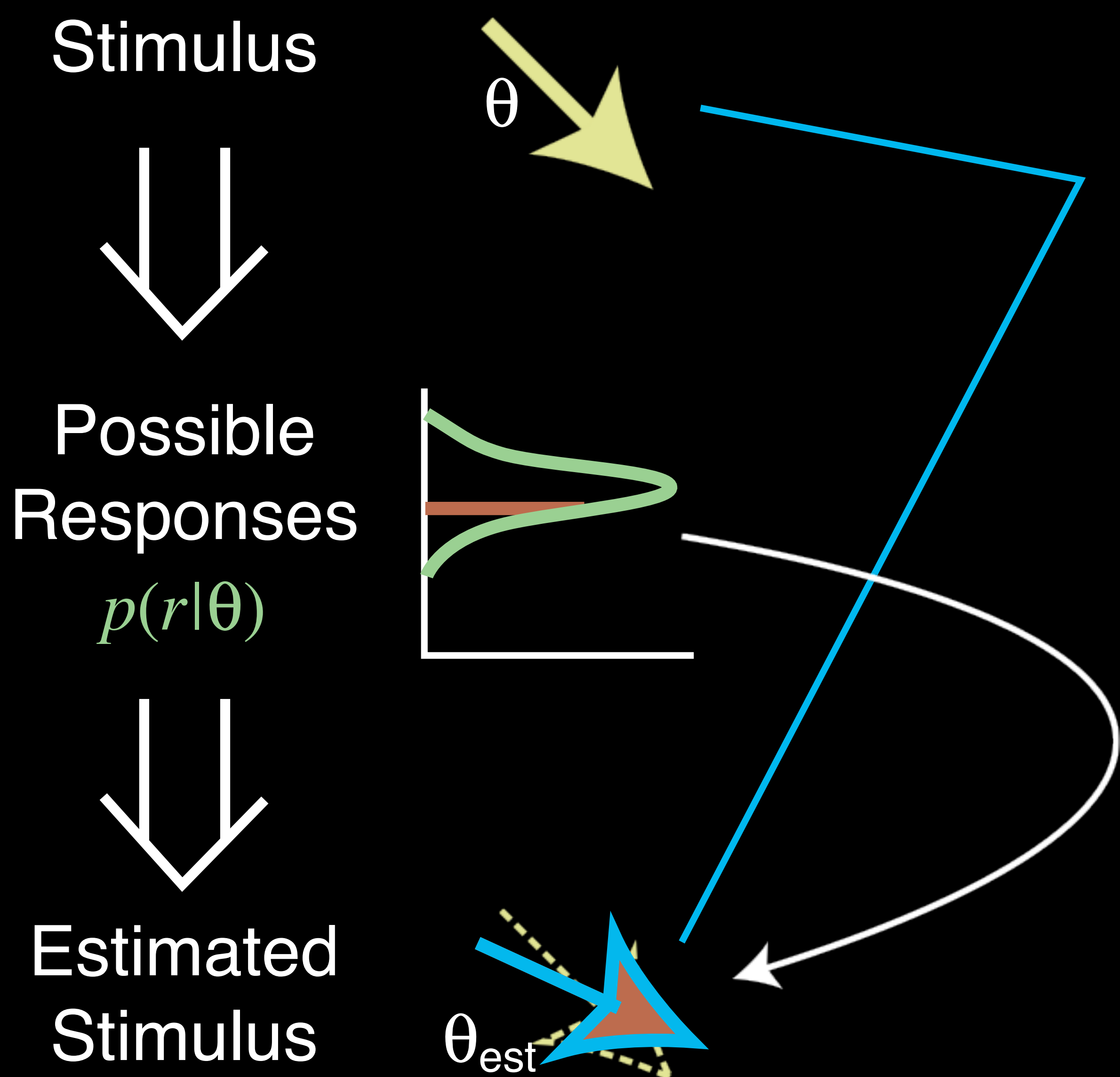
2. What happens in the context of a population?

Do effects of cooperativity change the relative importance of stimuli to a neuron?

Reconstruction Error and Encoding

A straight-forward way of evaluating the neuron's performance in coding stimuli:





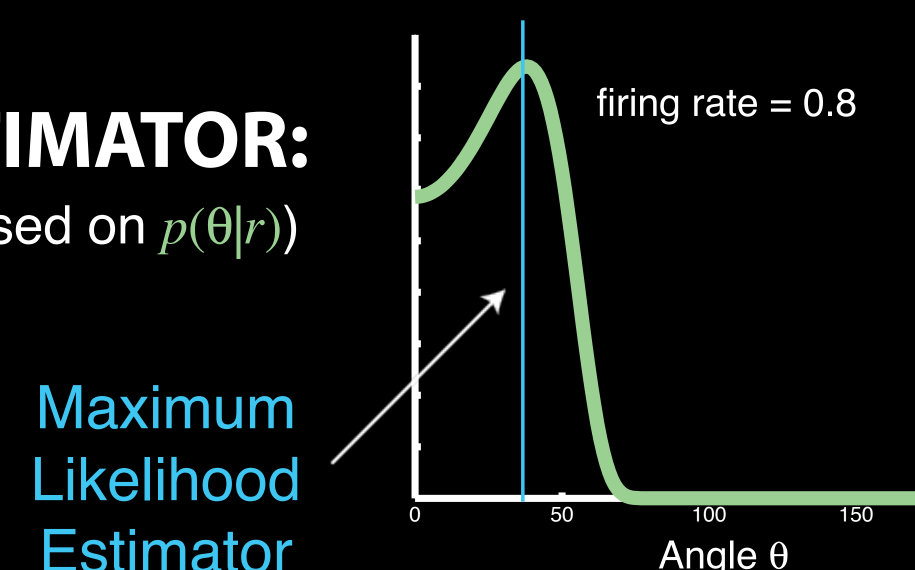
Error Term for each stimulus:

$$E[\theta] = \sum_r p(r|\theta)[\hat{\theta}(r) - \theta]^2$$

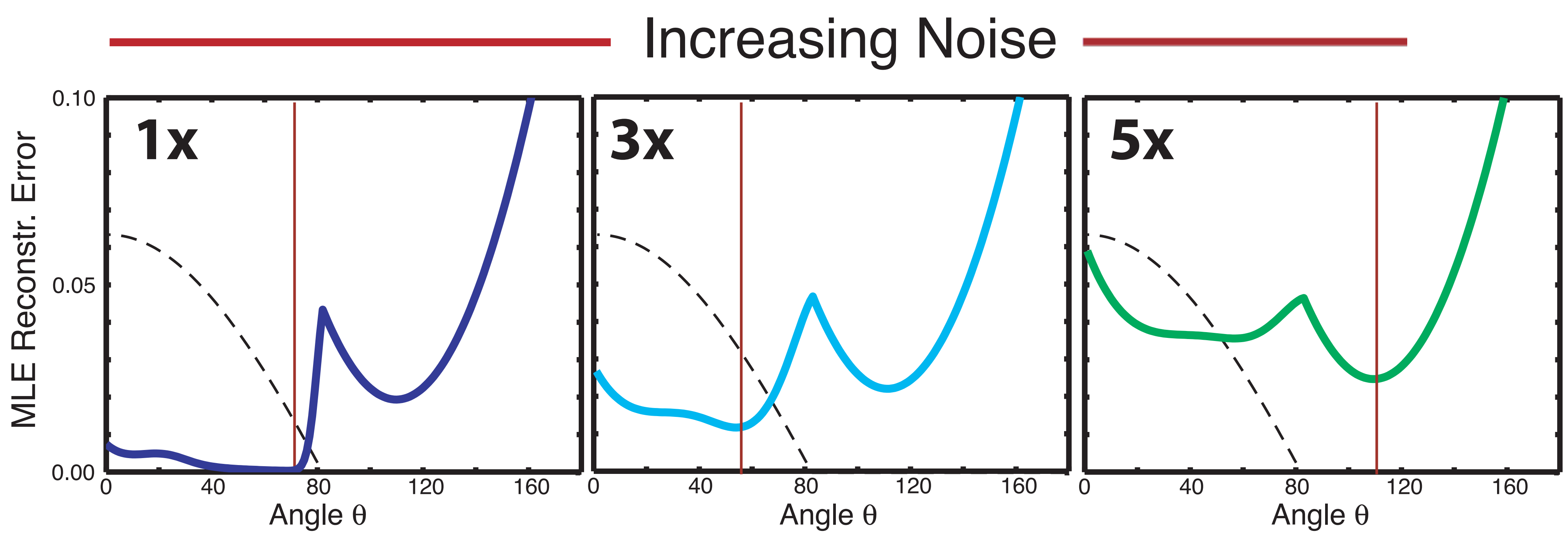
(averaged over all responses)

ESTIMATOR:

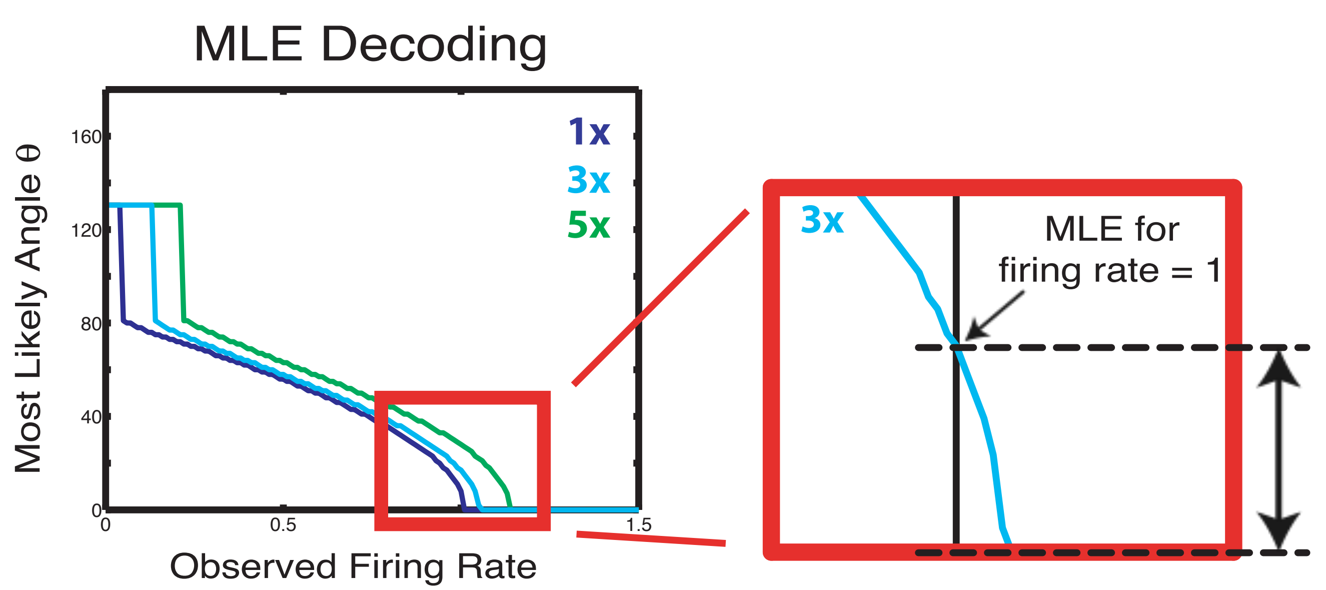
(based on $p(\theta|r)$)

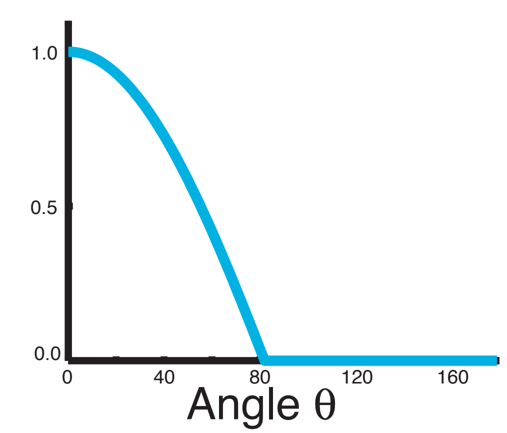


Stimulus with Minimum Reconstruction Error Has No Slope-to-Peak Transition

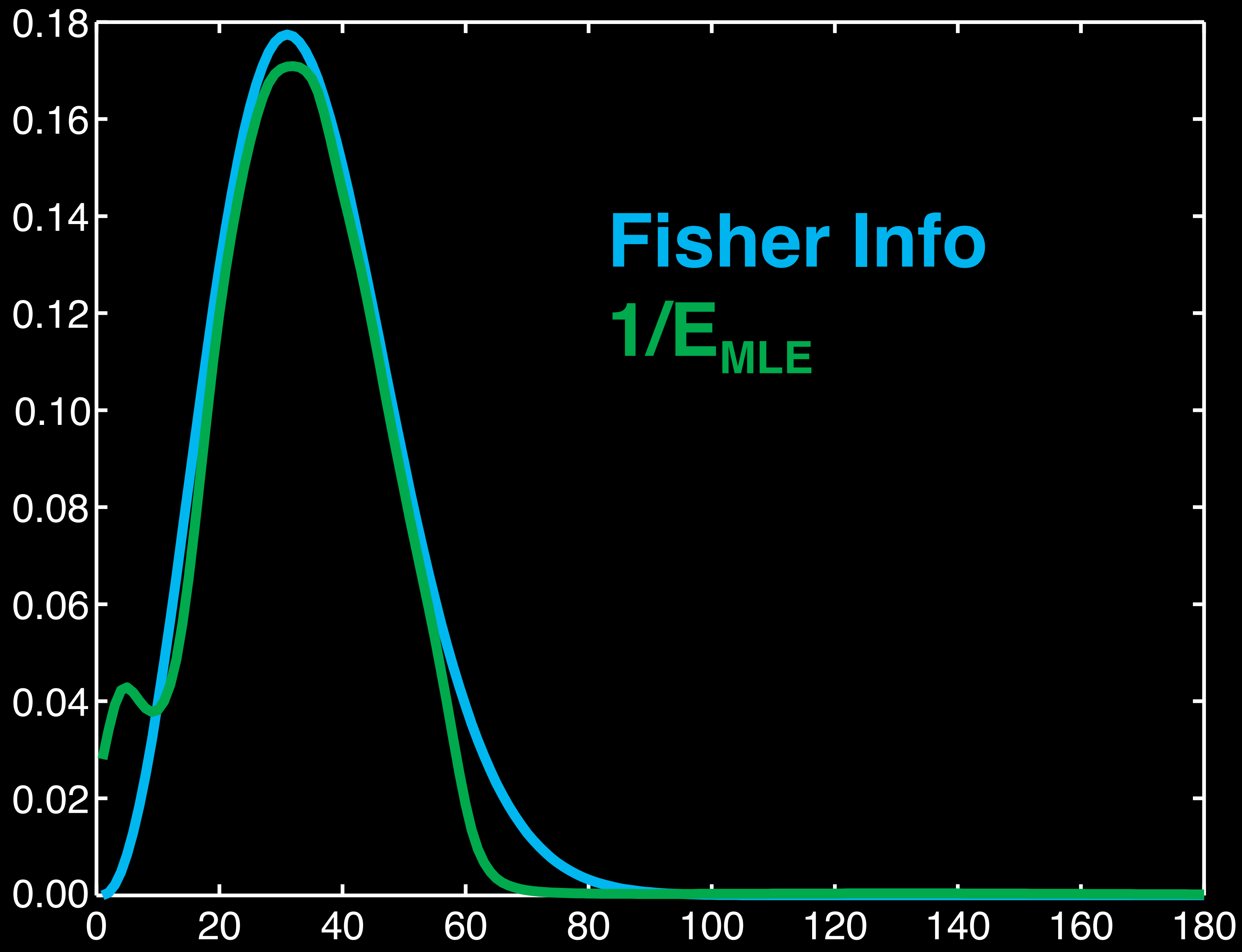


Why Not?

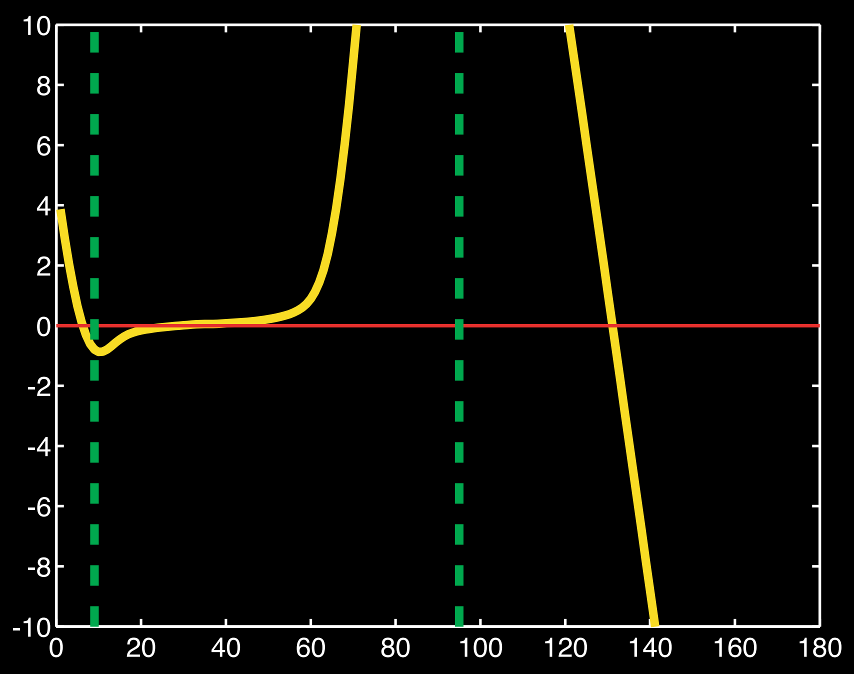




Fisher Info and Mean-Sq. Error



MLE Estimator
is BIASED

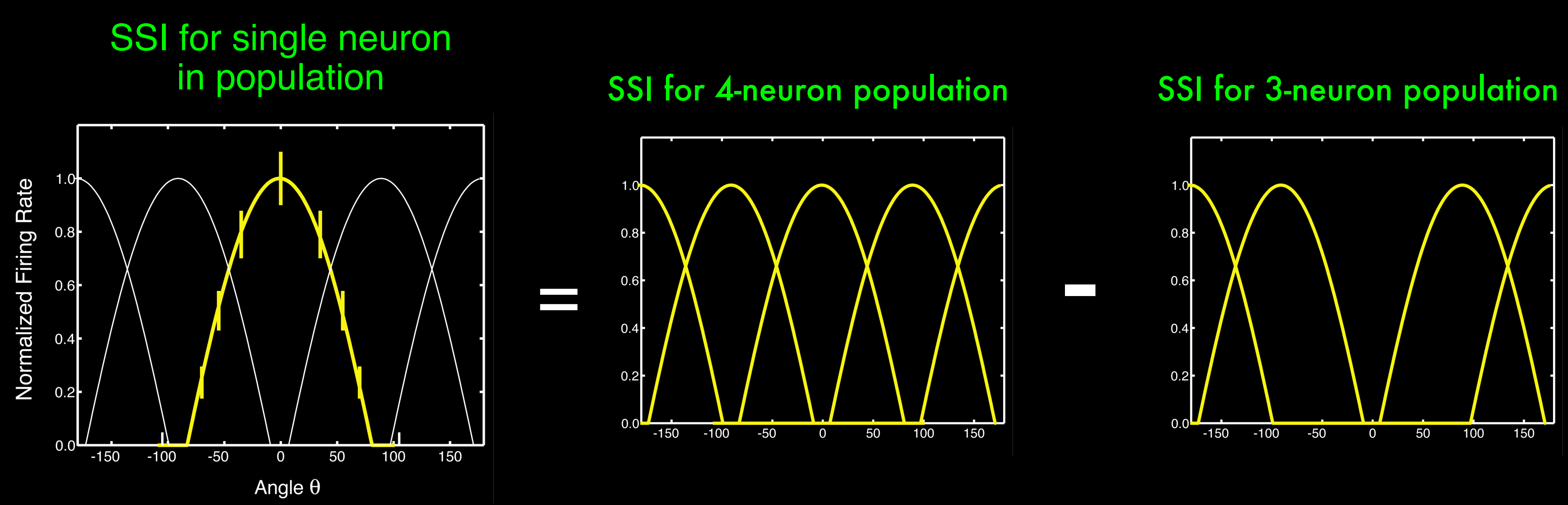


Cramer-Rao bound doesn't hold (?)

Single Neuron in a Population Context

The population SSI can be calculated.

What is the contribution of a single neuron to the population SSI?



SSI for a single neuron in a population equals information lost if this neuron were deleted.

Population transition from slope to peak at higher noise levels

LOW NOISE
(T&M noise)

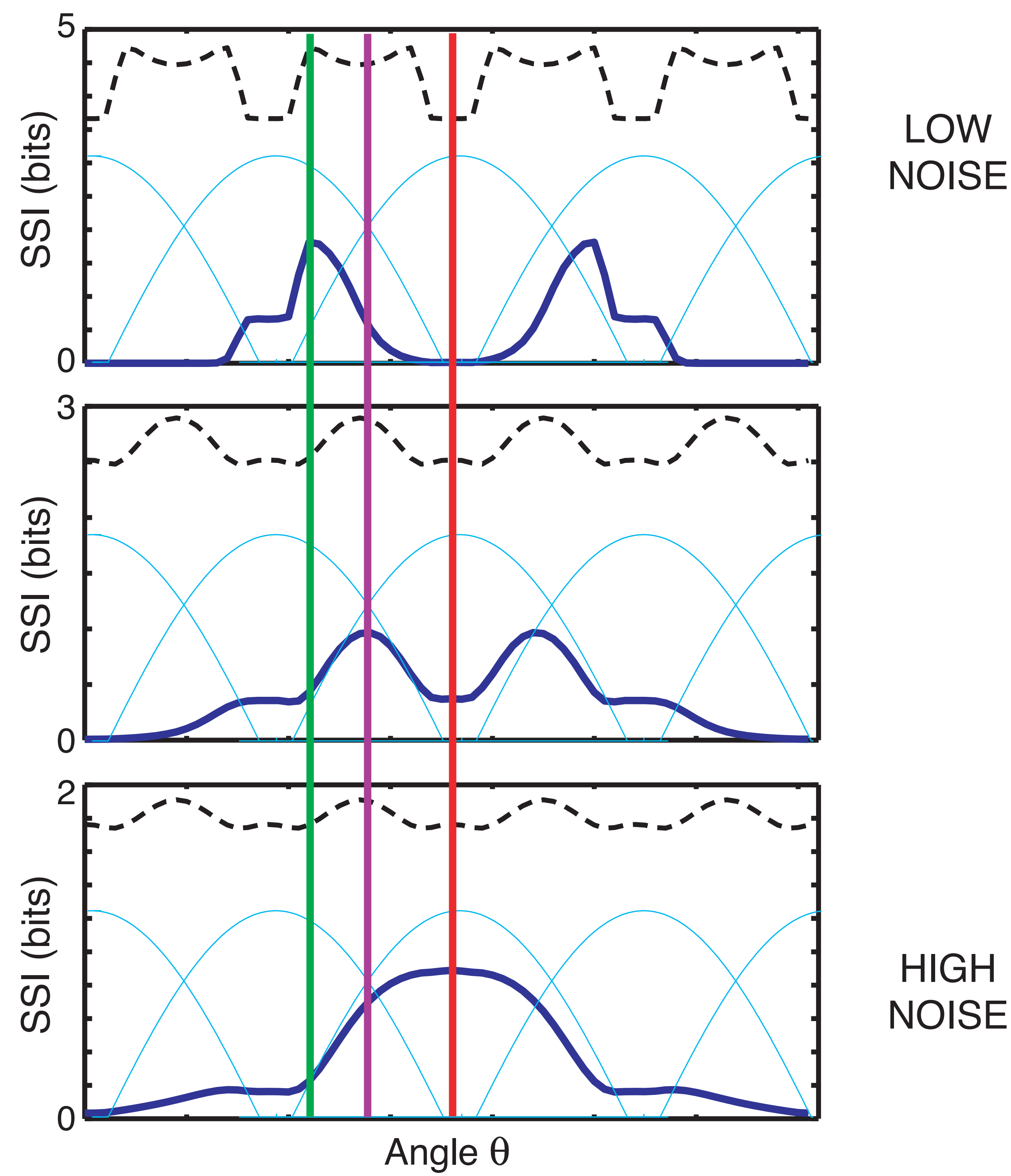
SSI peak near maximum slope

HIGH NOISE
(3x T&M noise)

SSI peak at intersections

HIGHEST NOISE
(5x T&M noise)

SSI peak at peak



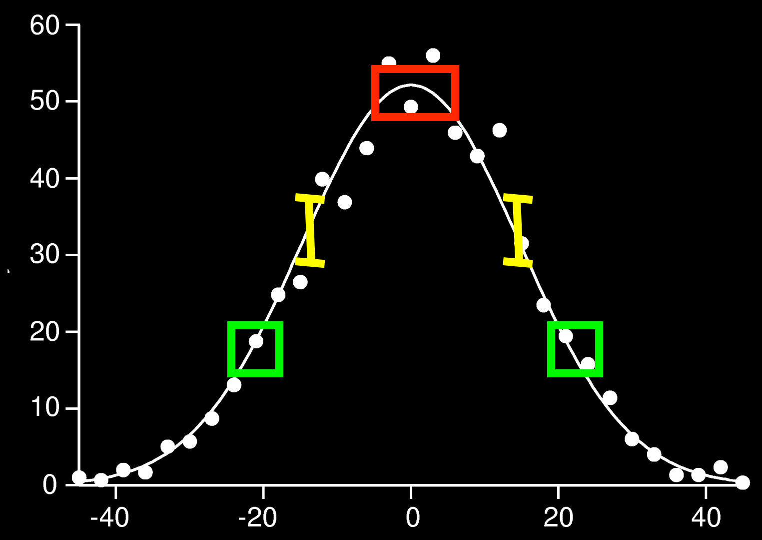
What is the function of a sensory neuron?

Old way of characterizing: Tuning curve → Which stimuli make a neuron fire the most?

New Way: Stimulus-specific information (SSI) → Which stimuli are best-encoded by a neuron?

Single neuron SSI

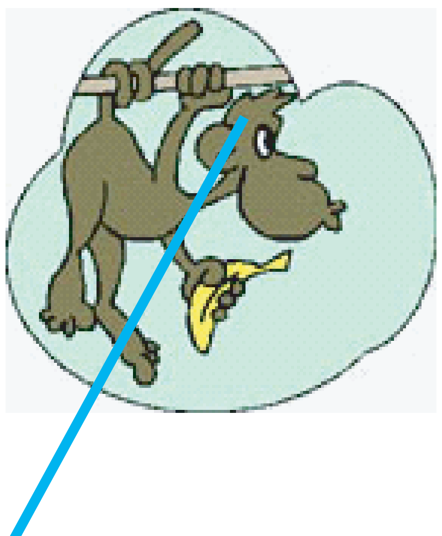
- Low noise -> large slope
- High noise -> peak

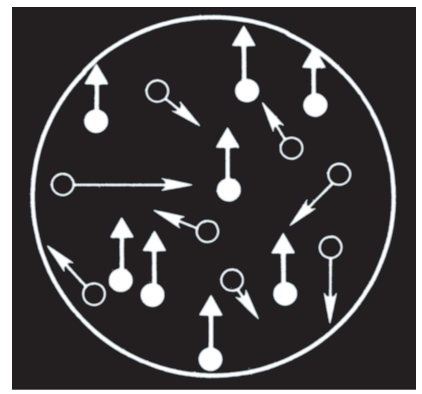


The meaning of the tuning curve depends on the amount of variability

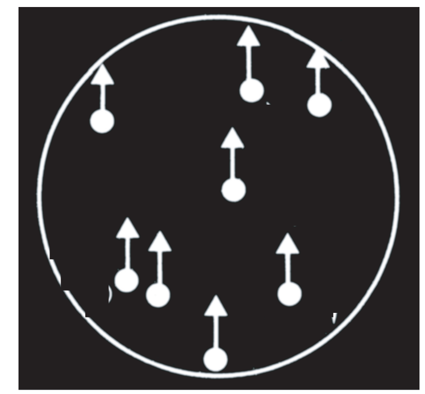
Experimental Test: Neuron-Behavior Correlation as a Function of Noise

1) Monkey points to perceived continuum direction of motion of random dots in both low noise and high noise conditions





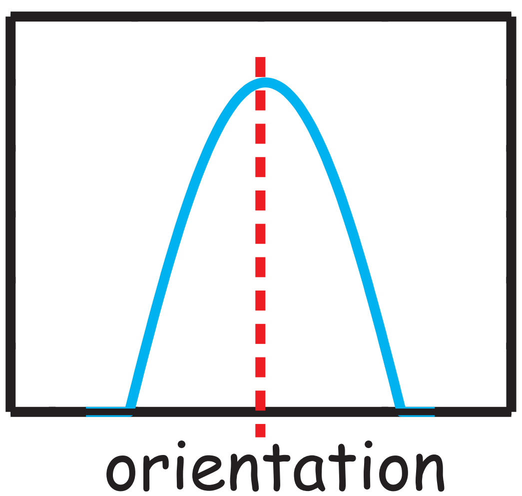
High Noise



Low Noise

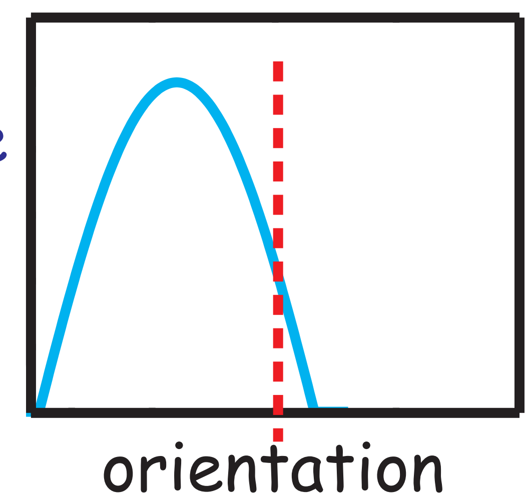
2) Record from area MT neurons chosen so that direction of motion is along:

a) peak firing rate



-OR-

b) peak slope



PREDICTION: Monkey behavior correlates best with:

a) peak firing rate neuron in high noise condition

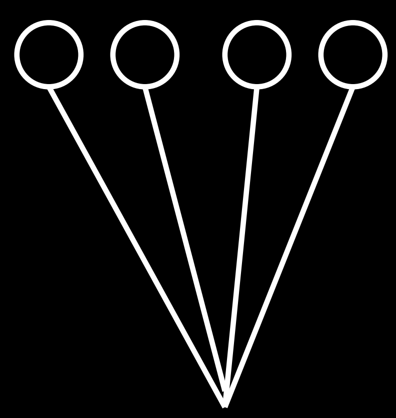
b) peak slope neuron in low noise condition

Implications of Transition for Neural Coding

Sensory neurons operate at different noise levels

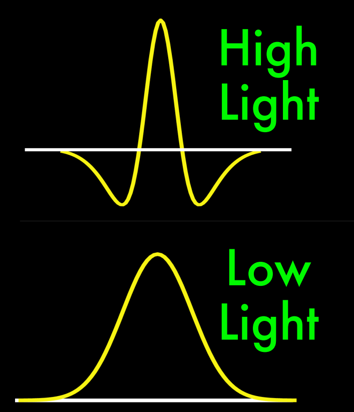
(Noise levels reported by Miller et al were for a given range of wind velocities and integration times)

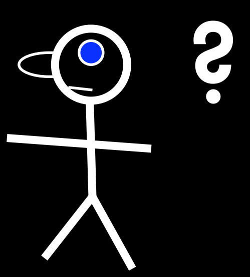
How would this be reflected?



1. Change in tuning curve shape to preserve what individual neurons encode

(e.g. retinal ganglion cells?)

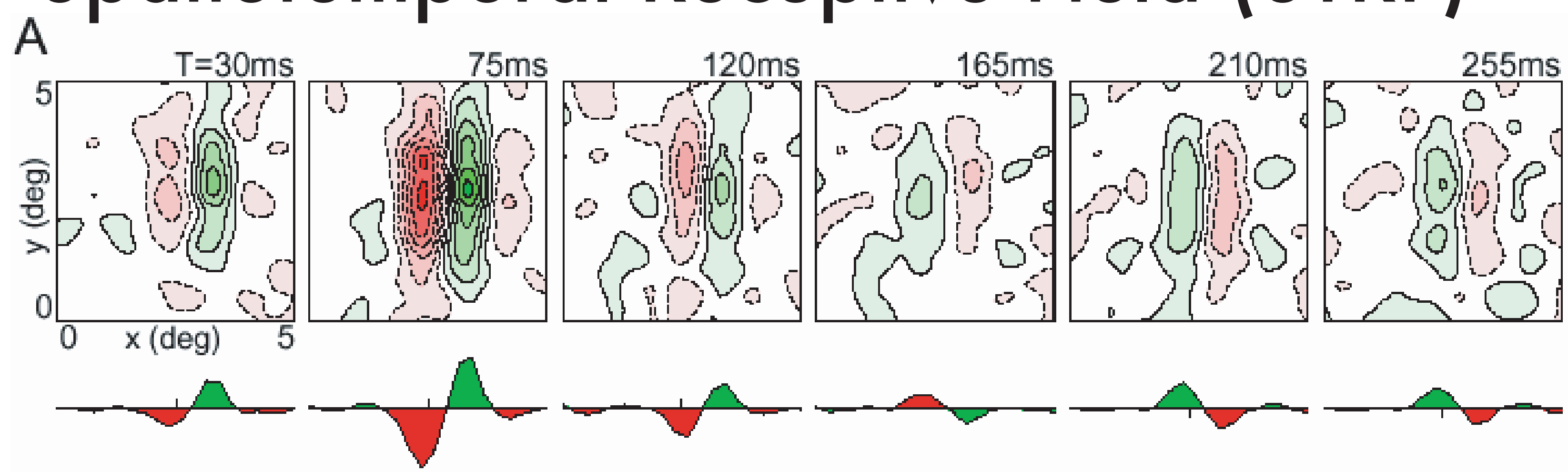




2. Change the strategy for decoding the neural information downstream?

Talk ended here (without getting to receptive field vs. information part of talk)

Spatiotemporal Receptive Field (STRF)



V1 Simple Cell (Deangelis et al., 1995)

Paper on Stimulus-specific information SSI:

Butts DA (2003) What is the information associated with a particular stimulus? *Network* **14**: 177-187.

See http://batman.med.harvard.edu/dbutts for more info.