NEPR208 - Optimality and adaptation

Stephen Baccus
Department of Neurobiology
baccus@stanford.edu

What?

What are the important biological functions ?

How?

How is a function performed by mechanisms?

Why?

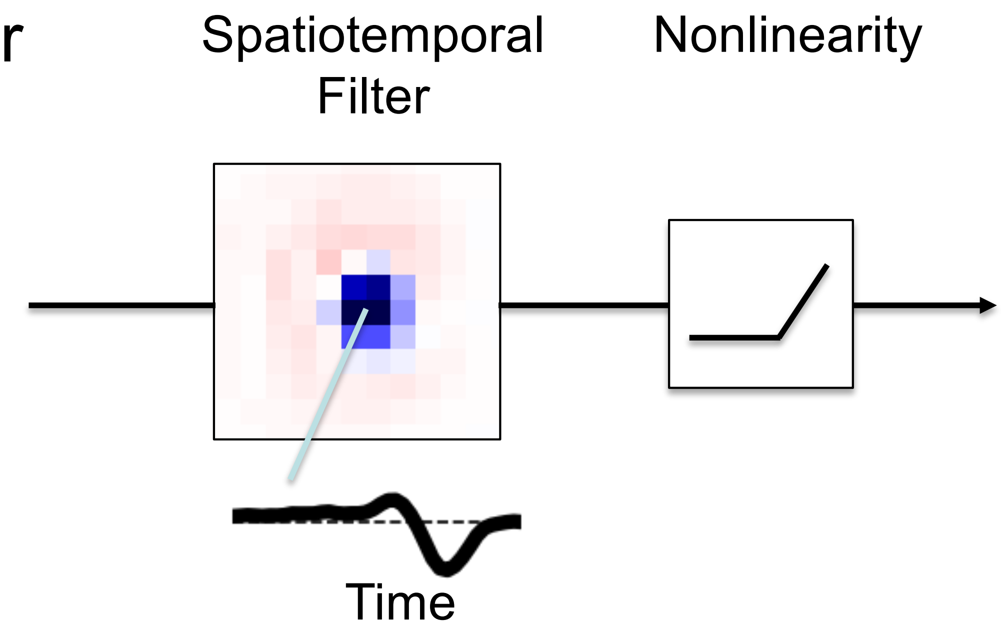
What are the functional benefits …

of a particular neural code?

of specific mechanisms?

Why this neural code?

Linear-Nonlinear (LN) Model



Functional advantages of response properties and changes in those properties

Why do cells have a particular nonlinear response function?

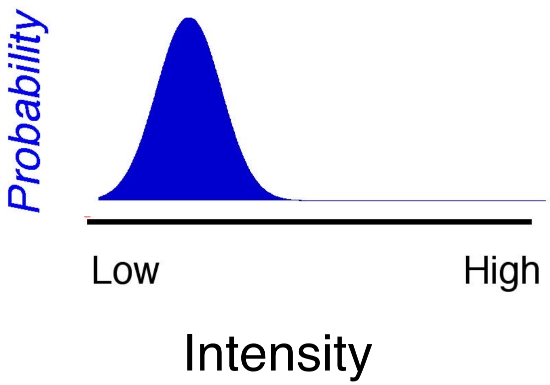
Why does the nonlinearity change?

Why do cells have a certain duration filter?

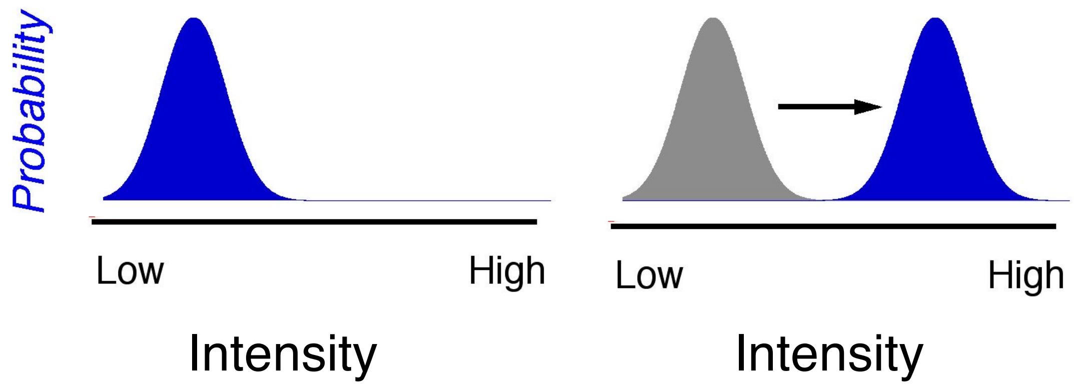
Why do they have a certain shape filter?

Why does the filter change?

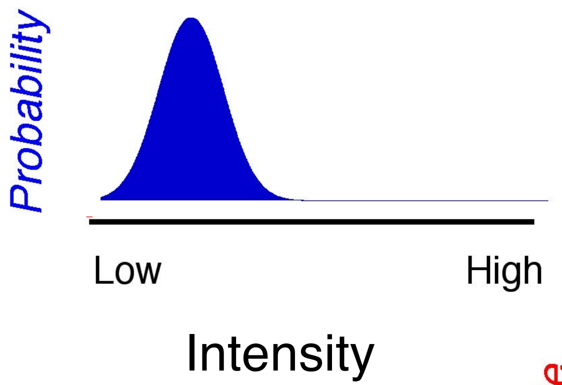
What are the statistics of the stimulus?

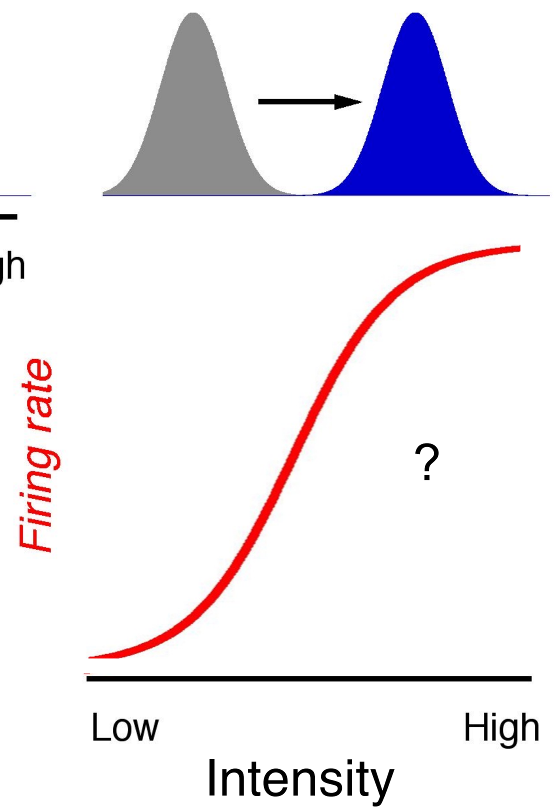


Stimulus statistics change over long timescales

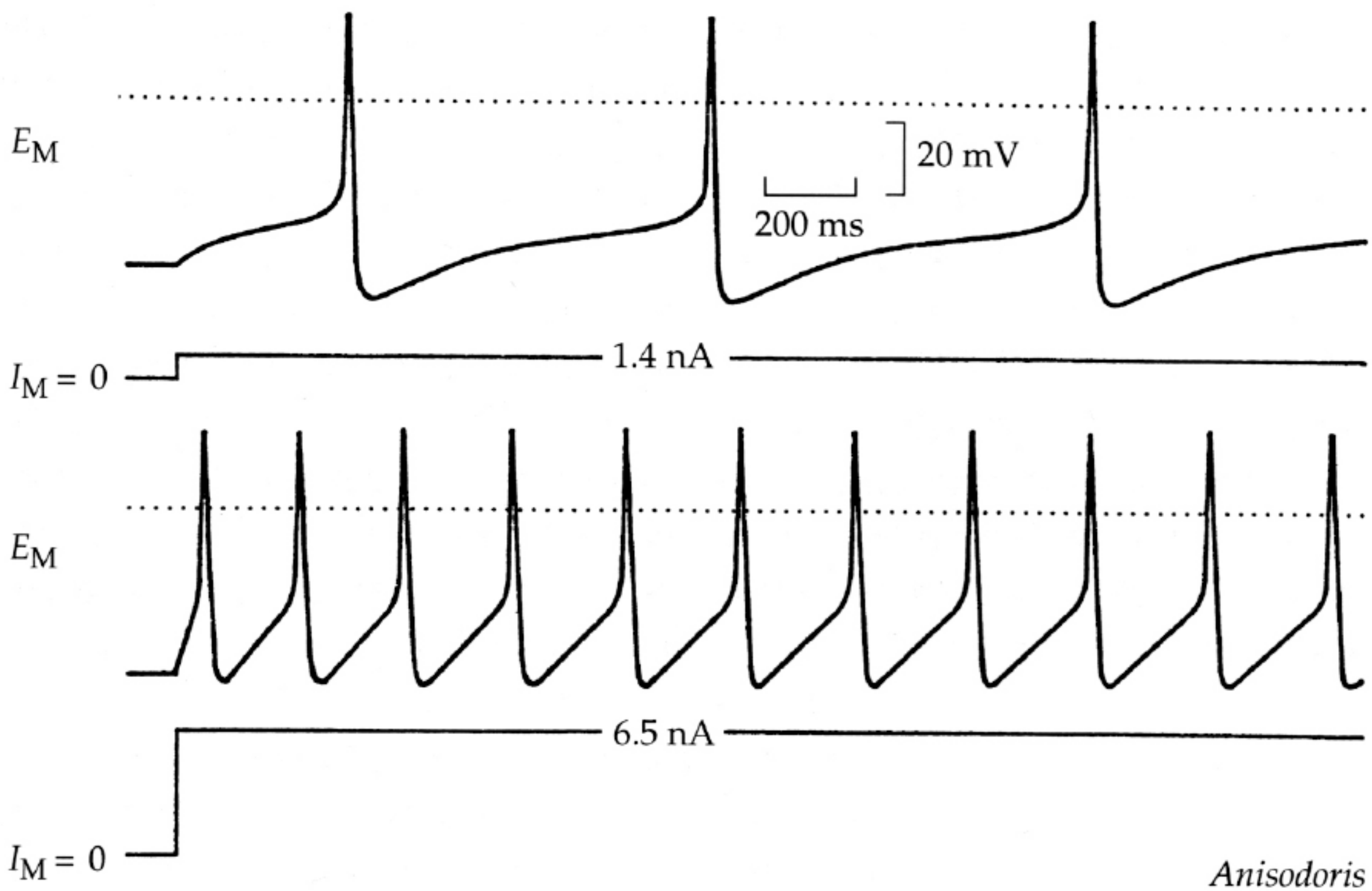


What is the best way to encode stimuli with changing statistics?



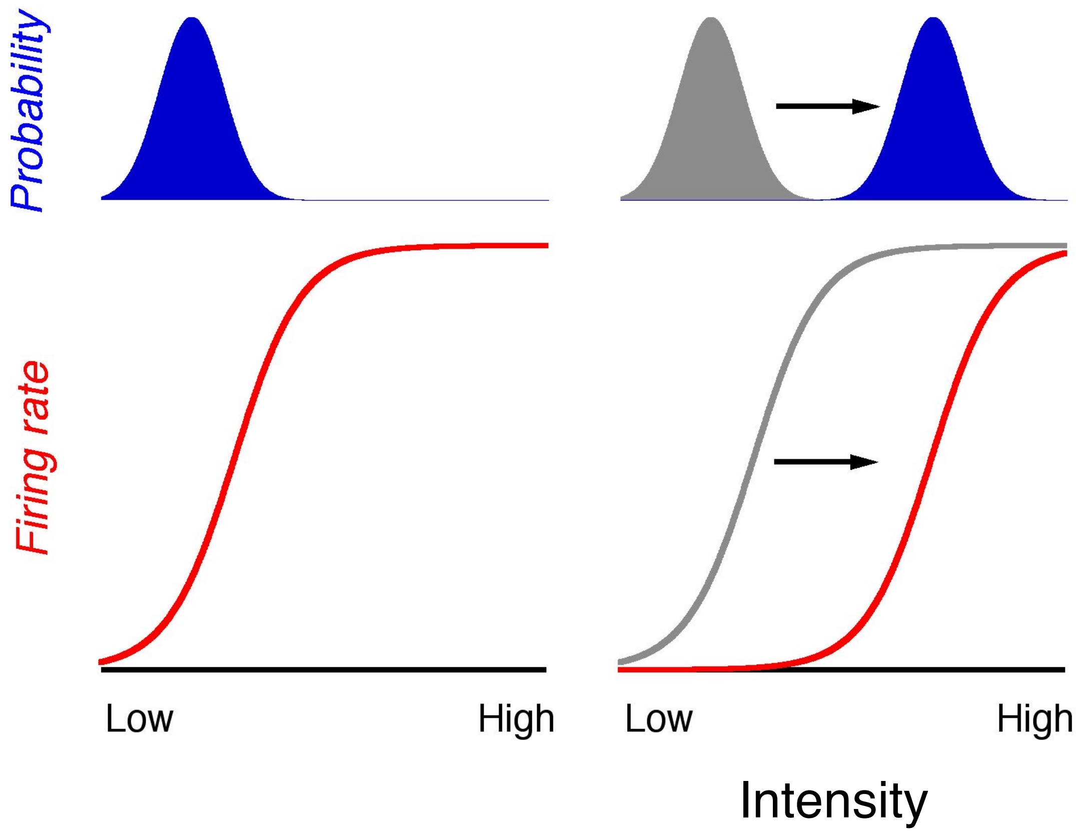


Neurons have a limited dynamic range set by maximum and minimum output levels, and by noise

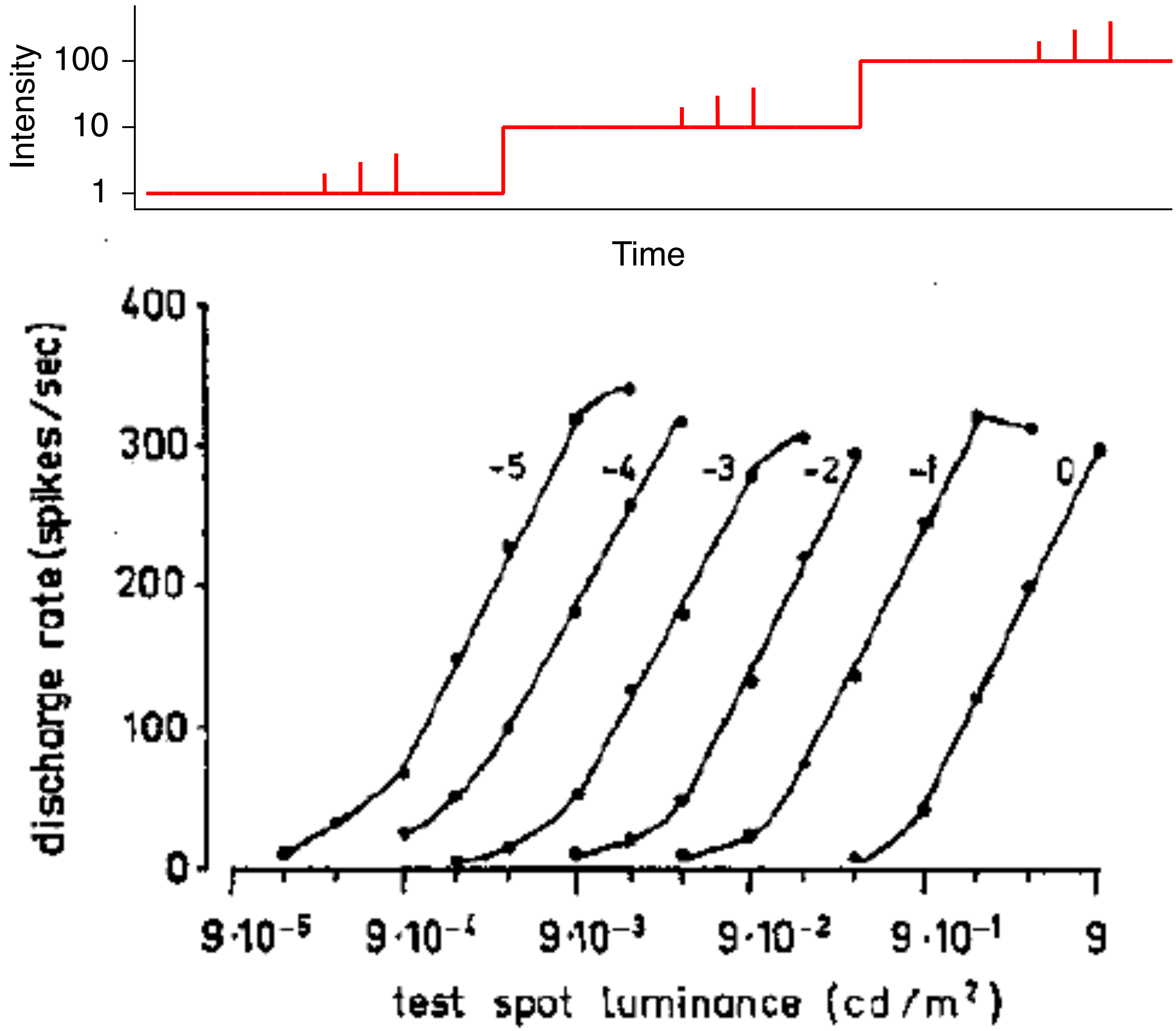


Adaptation to the average input

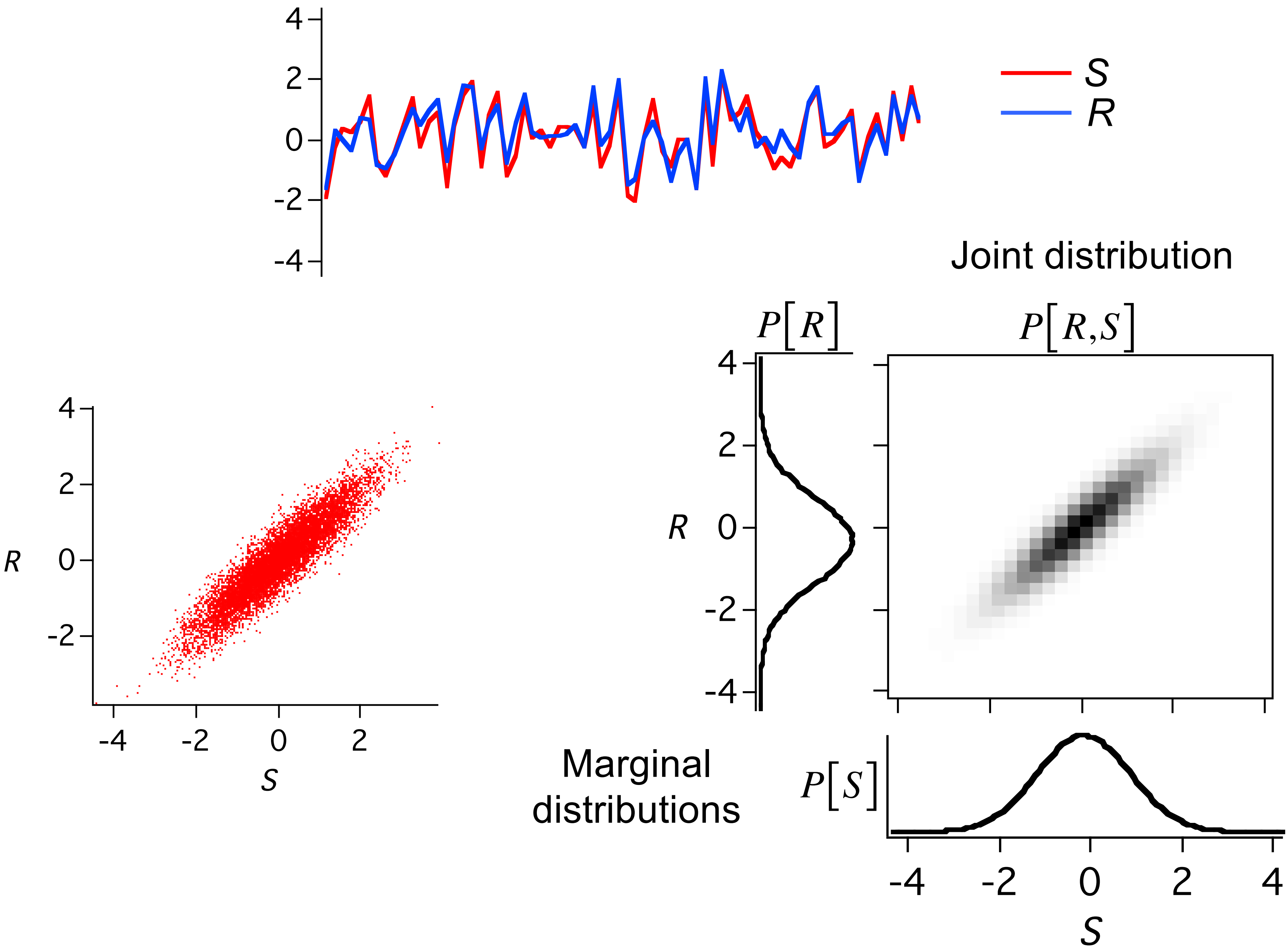
Light adaptation



Ganglion cell response curves shift to the mean light intensity



Sakmann and Creuzfeldt, Scotopic and mesopic light adaptation in the cat's retina (1969)


S
R
Joint distribution
P[R]
P[R,S]
R
S
Marginal distributions
P[S]
S

A Mathematical Theory of Communication
Claude Shannon (1948)

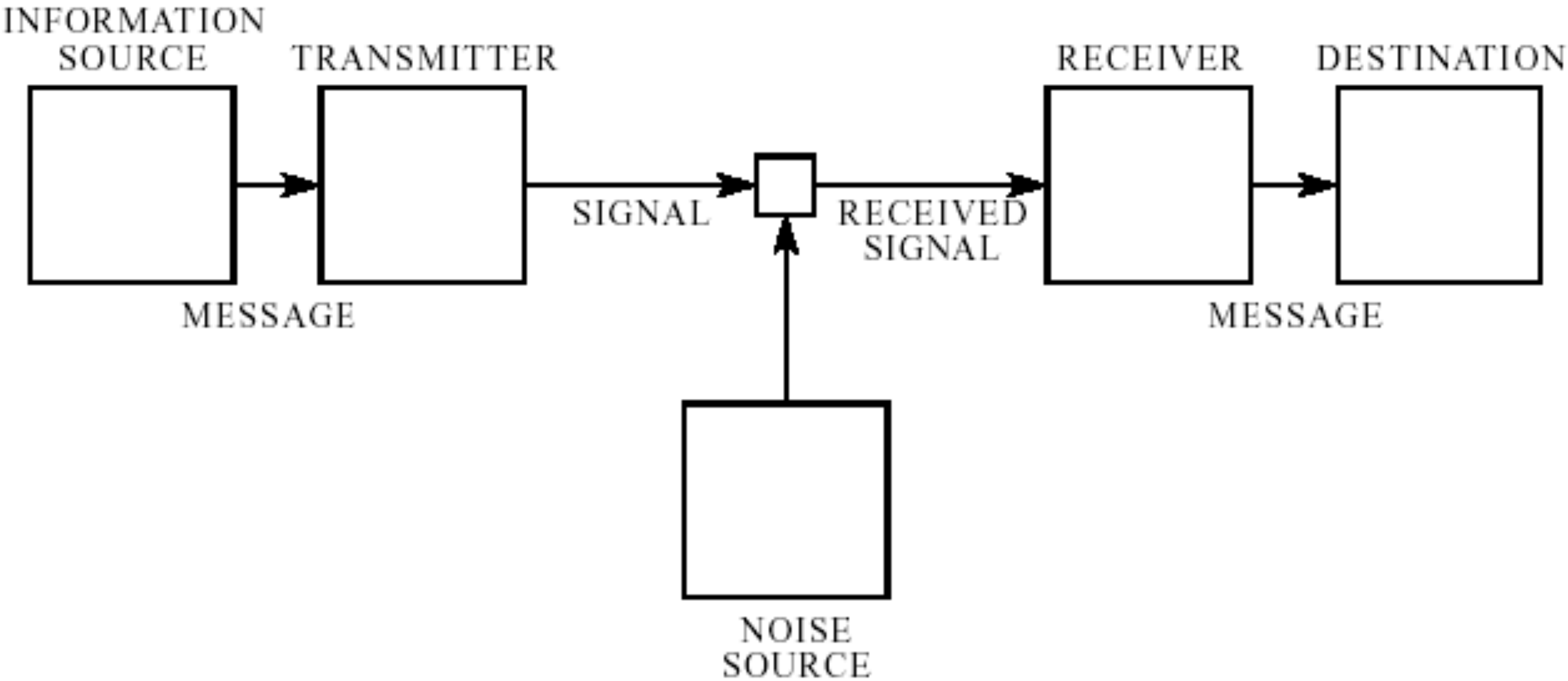


Fig. 1—Schematic diagram of a general communication system.

Bell System Technical Journal 27, 379-423

A Mathematical Theory of Communication

Claude Shannon (1948)

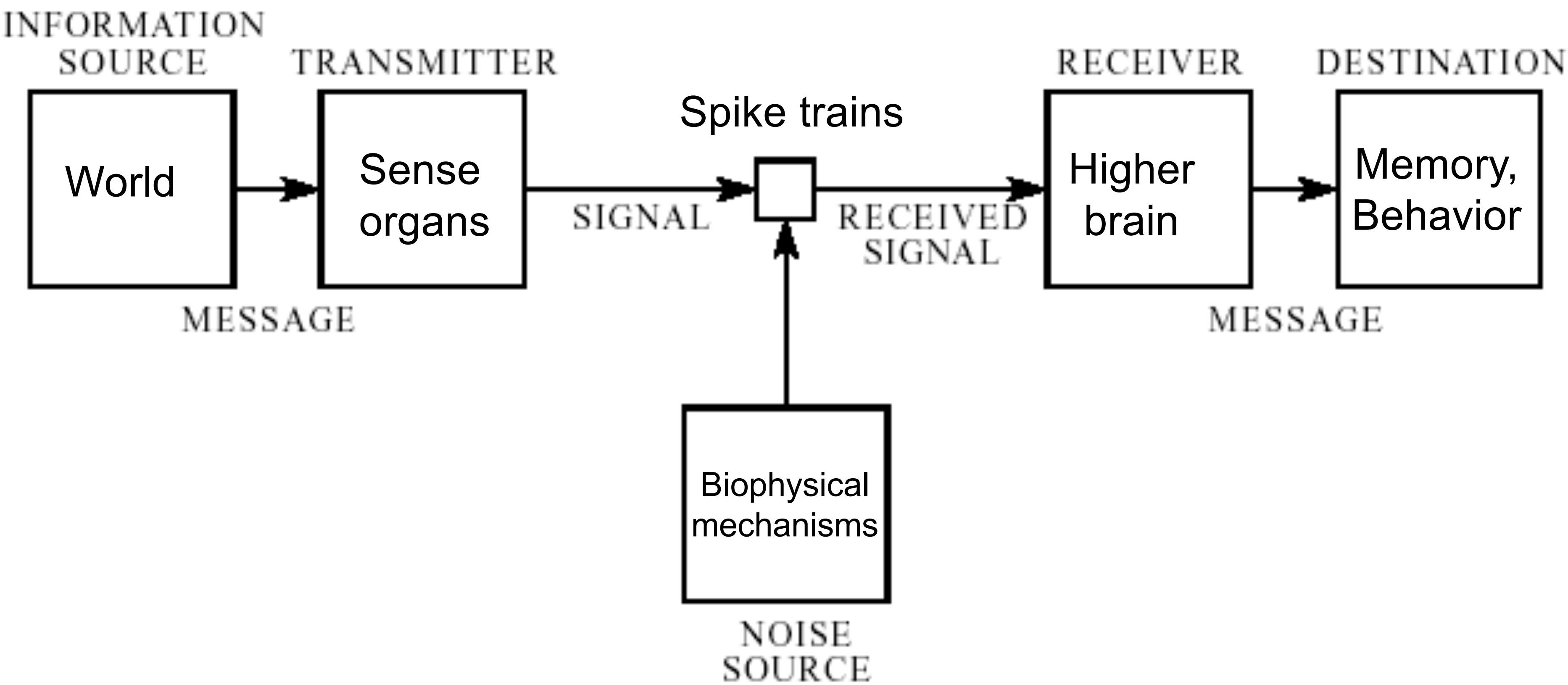


Fig. 1—Schematic diagram of a general communication system.

Bell System Technical Journal 27, 379-423

A Mathematical Theory of Communication

Claude Shannon (1948)

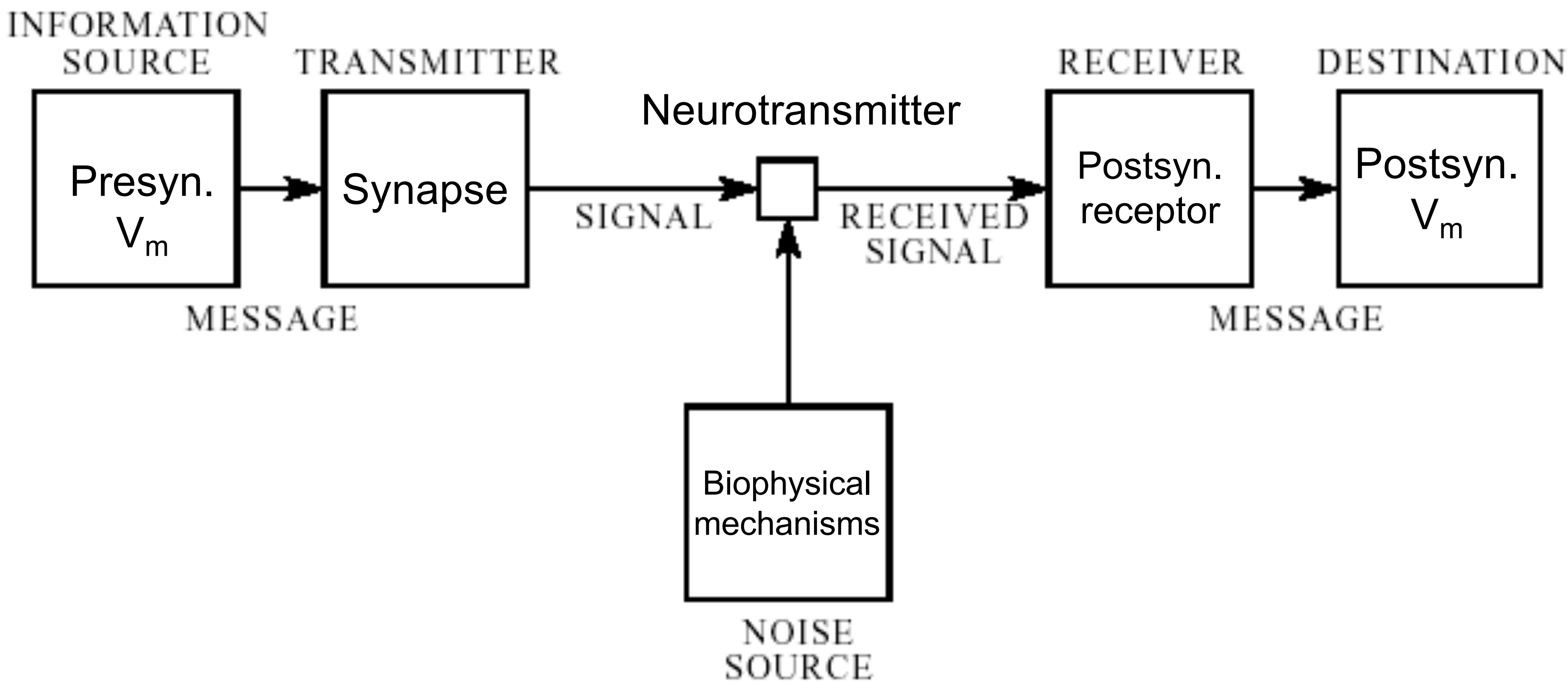


Fig. 1—Schematic diagram of a general communication system.

Bell System Technical Journal 27, 379-423

A Mathematical Theory of Communication

Claude Shannon (1948)

What is information?

A measure of uncertainty of a random variable. The maximum possible amount of information there is to be learned from a variable.

$$S = k \log W$$

S: Entropy

k: Boltzmann constant

W: Number of possible microscopic states

<u>A Mathematical Theory of Communication</u>

Claude Shannon (1948)

What is information?

<u>Entropy</u>

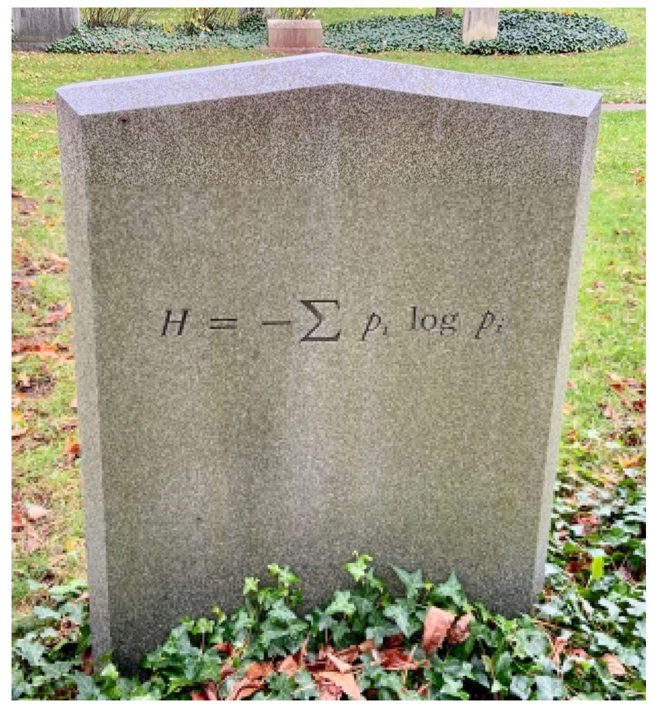
A measure of uncertainty of a random variable in bits. The maximum possible amount of information there is to be learned from a variable.

$$H(X) = -\mathrm{E}(\log P[x_i]) = -\sum P[x_i] \log P[x_i]$$

Entropy of a fair coin =

- 1/2 log(1/2) - 1/2 log(1/2) = 1 bit

0 log(0)=0



Entropy is maximal when all possibilities are equally likely

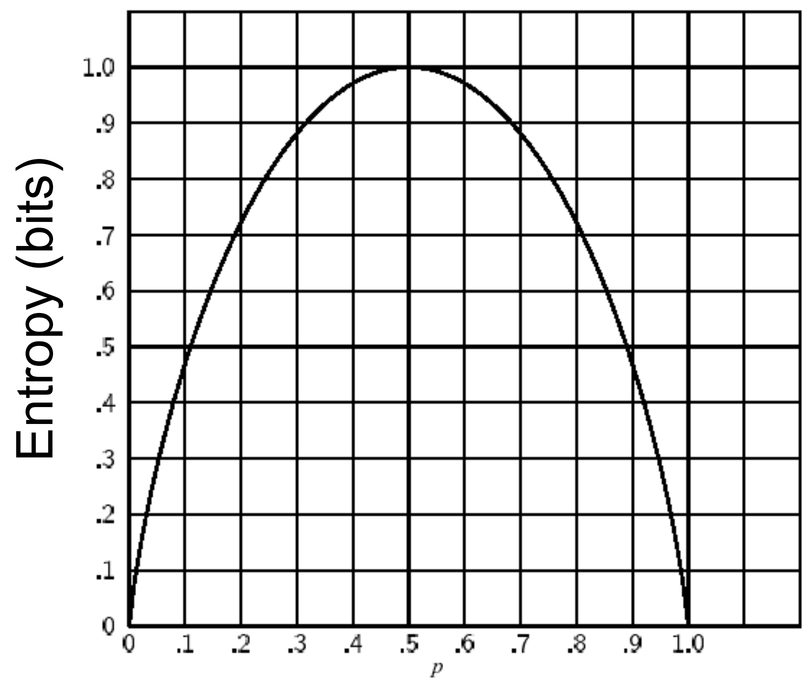
$$H(X) = -\mathrm{E}(\log P[x_i]) = -\sum P[x_i] \log P[x_i]$$

Entropy of a fair coin =

- 1/2 log(1/2) - 1/2 log(1/2) = 1 bit

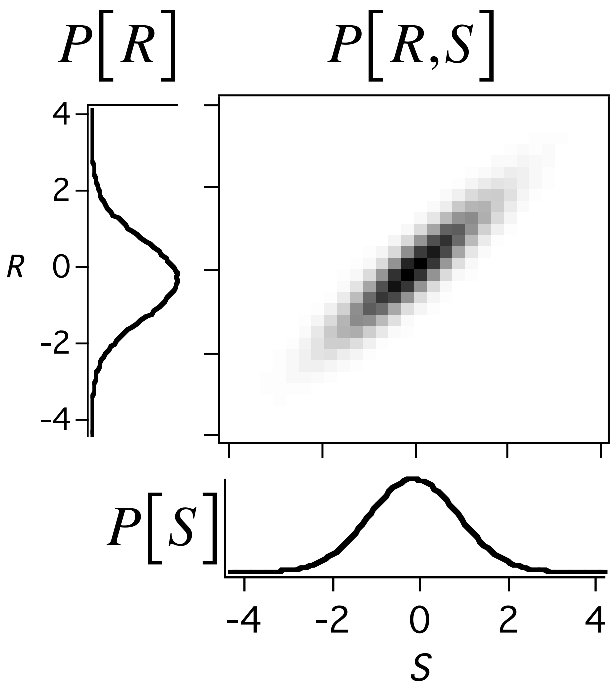
of an unfair coin =

- 3/4 log(3/4) - 1/4 log(1/4) = ~0.8 bits



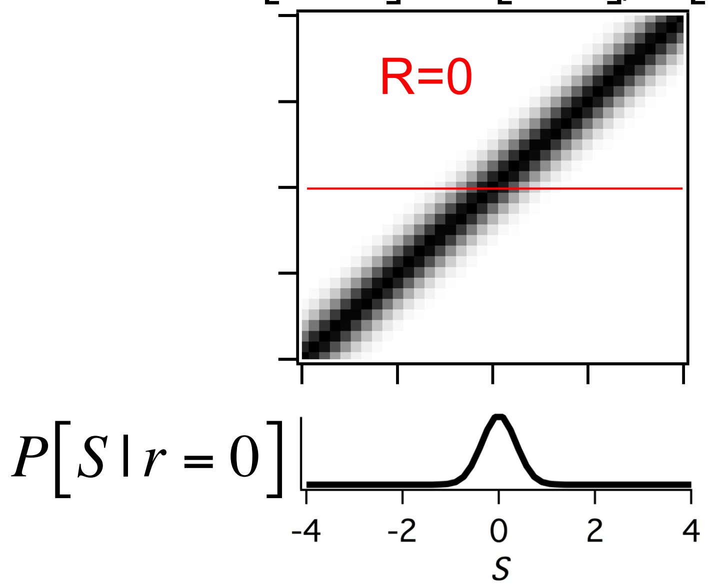
Entropy with two possibilities of probability p and 1 - p

Information is a reduction in entropy



Conditional distribution

$$P[S \mid R] = P[R,S] / P[R]$$



Conditional entropy

$$H(S \mid R) = -\mathrm{E}_{P[S,R]}(\log P[S \mid R])$$

$$= -\sum_{S}\sum_{R} P[S,R] \log P[S \mid R]$$

Mutual information

A measure, in bits, of how much information is conveyed by one random variable about another random variable. It is equal to the total entropy minus the conditional entropy.

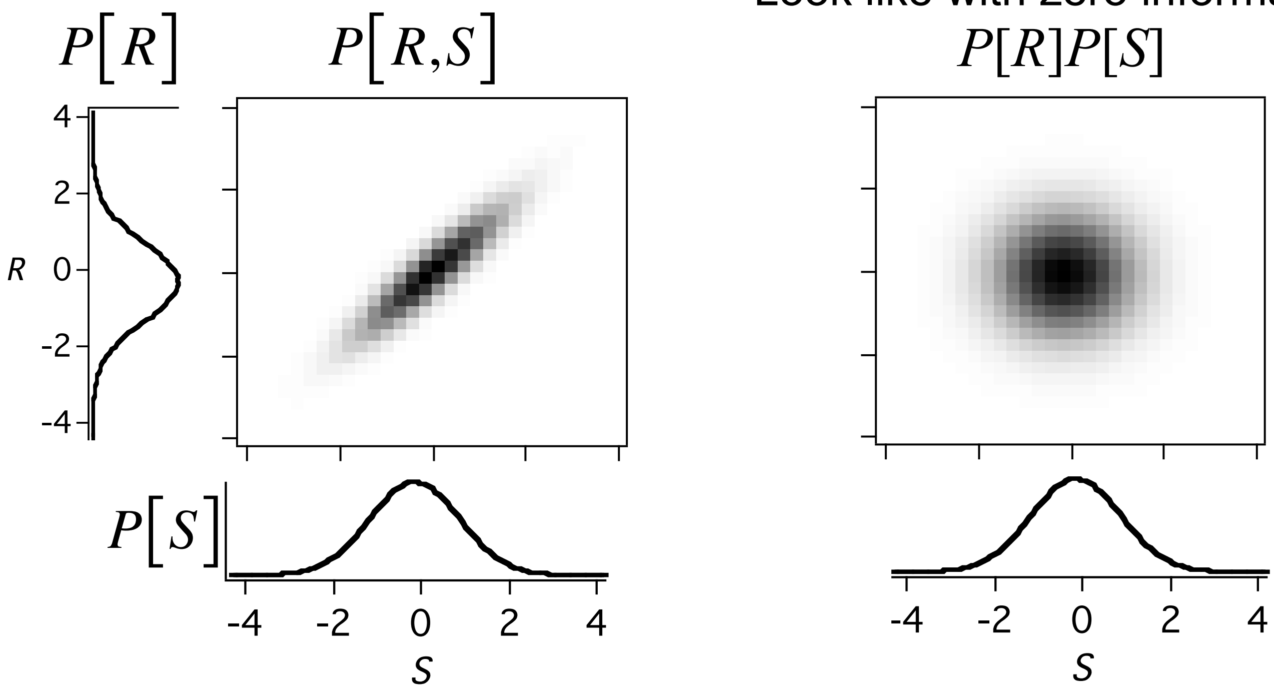
$$I(S;R) = H(S) - H(S|R)$$

$$I(R;S) = H(R) - H(R|S)$$

$$I(R;S) = I(S;R)$$

Mutual information as the ‘distance’ between two probability distributions

Product distribution – what things would Look like with zero information



$$I(S;R) = E_{P(S,R)}\left(\log\left(\frac{P(S,R)}{P(S)P(R)}\right)\right)$$

$$= \sum_{S}\sum_{R} P(S,R)\log\left(\frac{P(S,R)}{P(S)P(R)}\right)$$

$$D_{KL}(P(S,R) \parallel P(S)P(R))$$

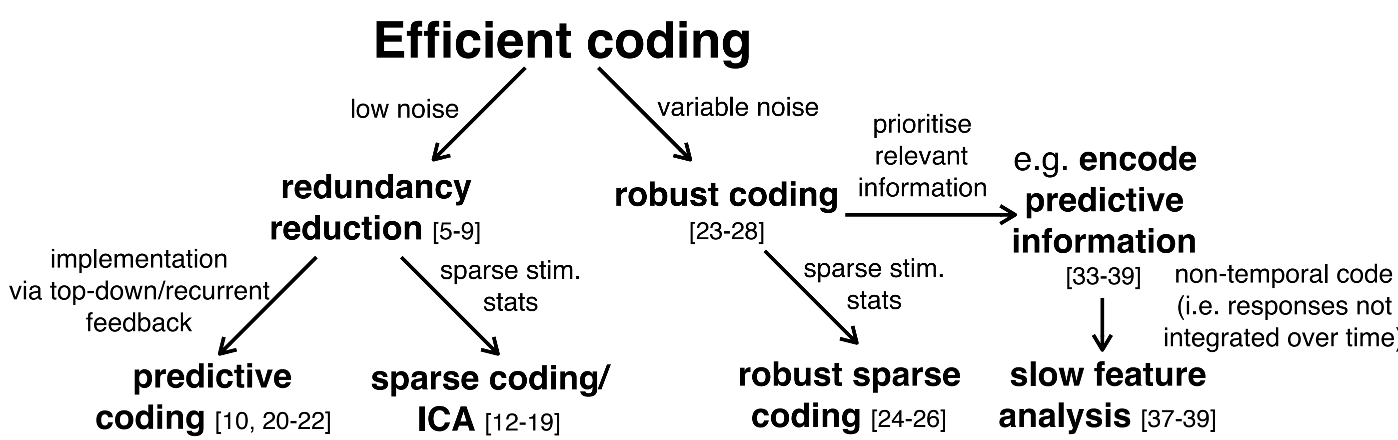
Kullback-Leibler divergence

Does the early visual system maximize information transmission?

'Efficient Coding' - Horace Barlow

Mutual information	Total entropy	Conditional ("noise") entropy

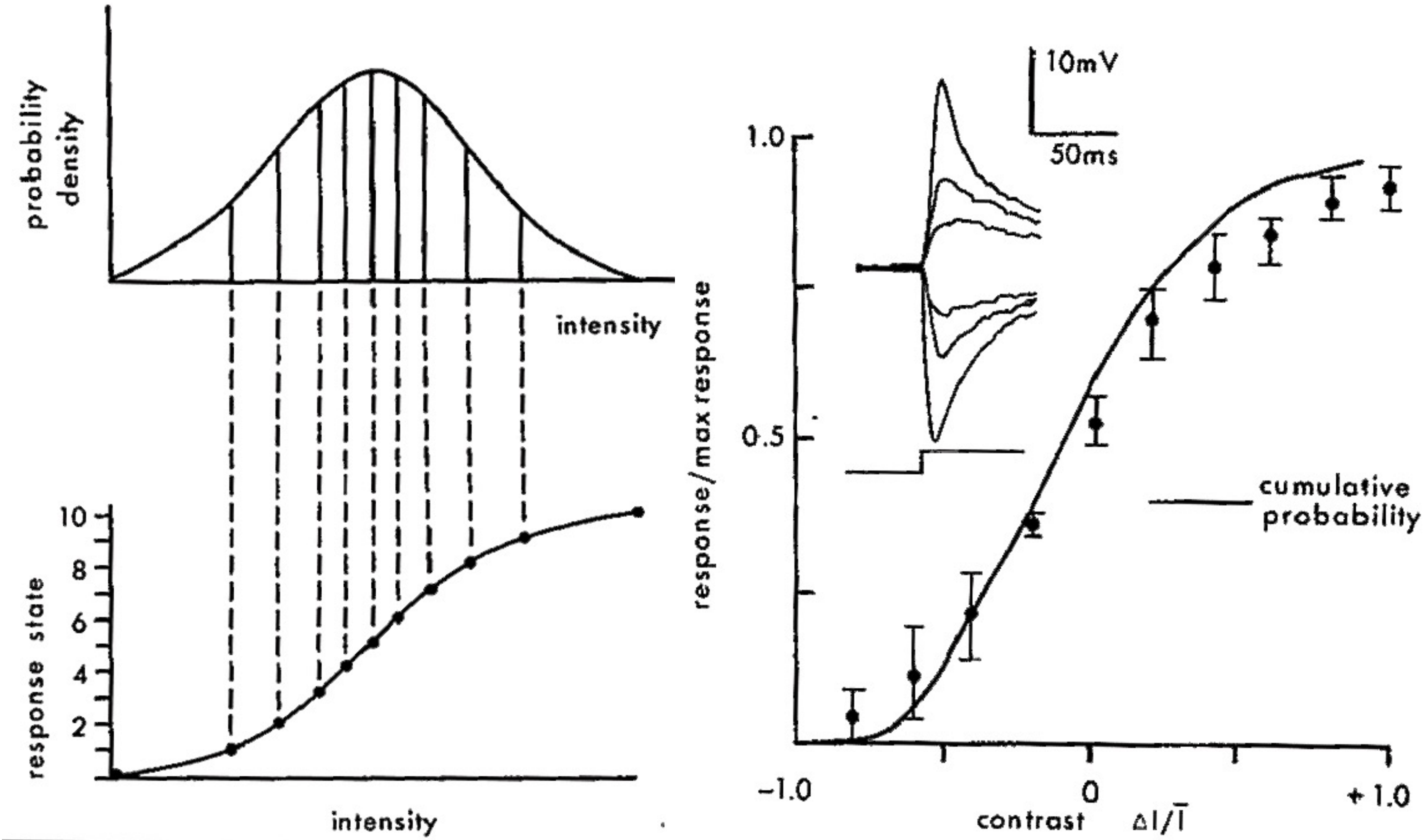
$$I(S;R) = H(S) - H(S|R)$$



References in notes section

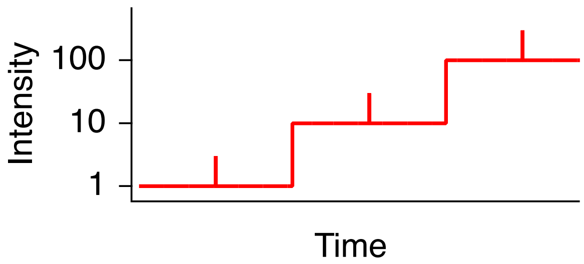
Chalk, Matthew, Olivier Marre, and Gašper Tkačik. "Toward a unified theory of efficient, predictive, and sparse coding." *Proceedings of the National Academy of Sciences* 115.1 (2018): 186-191.

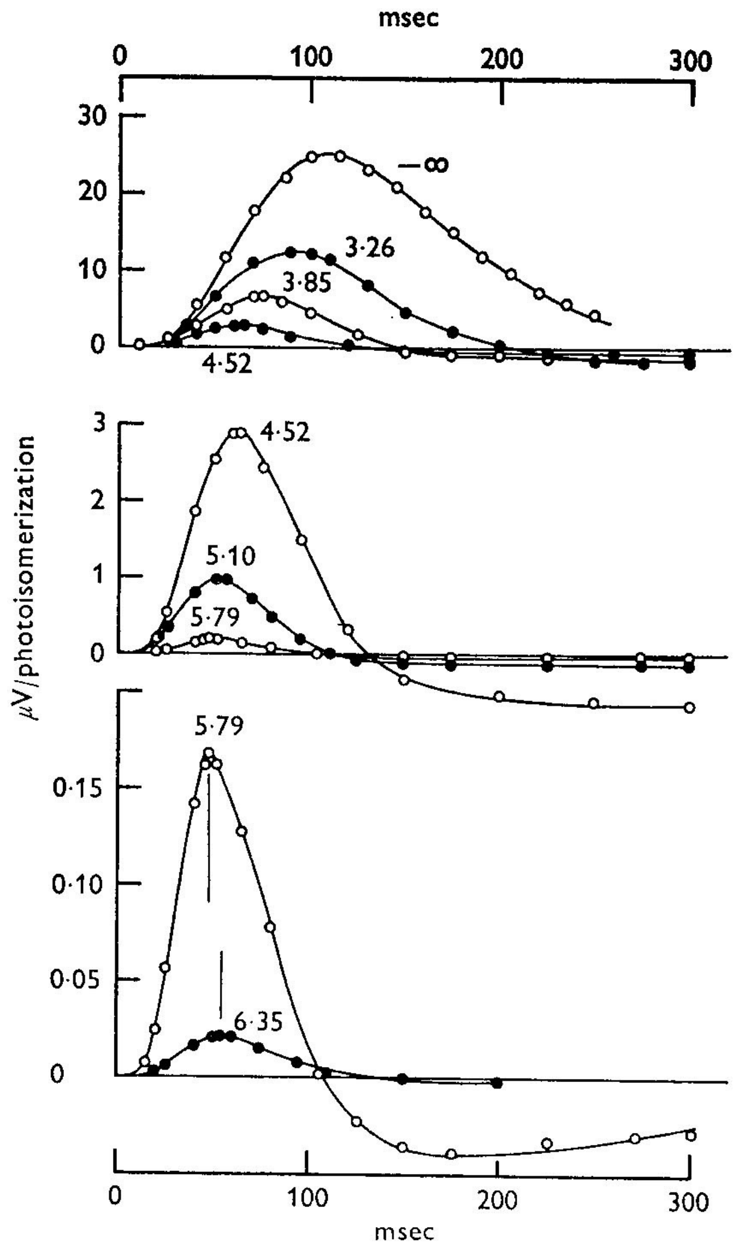
Maximizing information by a nonlinearity that maximizes total entropy



Simon Laughlin, A simple coding procedure enhances a neuron's information capacity Z. Naturforsch, 36c: 910-912 (1981)

Turtle Cones: Sensitivity *and Kinetics* change with mean luminance





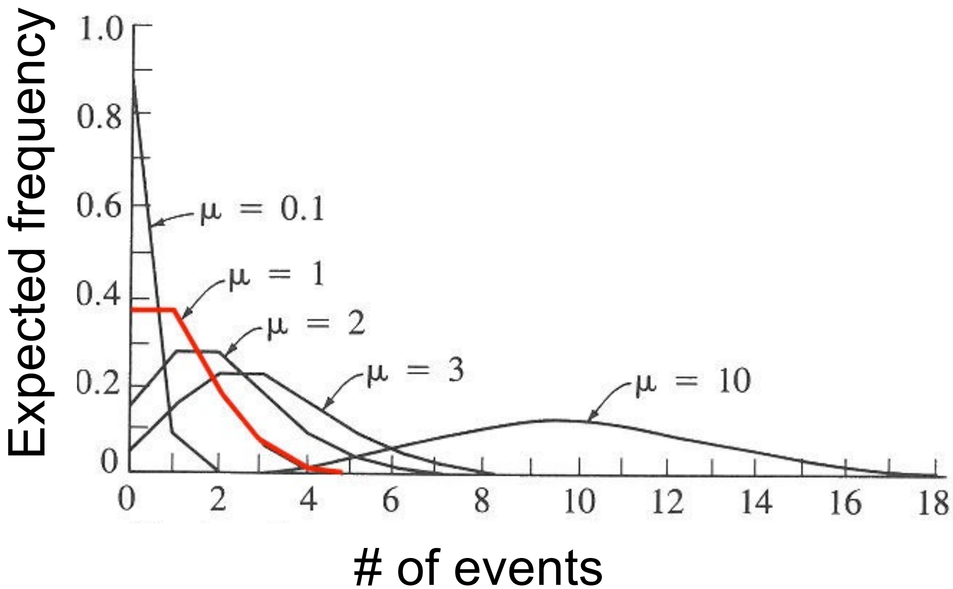
Baylor & Hodgkin 1974

Events with Poisson statistics $\mathrm{P}[n,\mu]$

$$\frac{e^{-\mu}\mu^{n}}{n!}$$

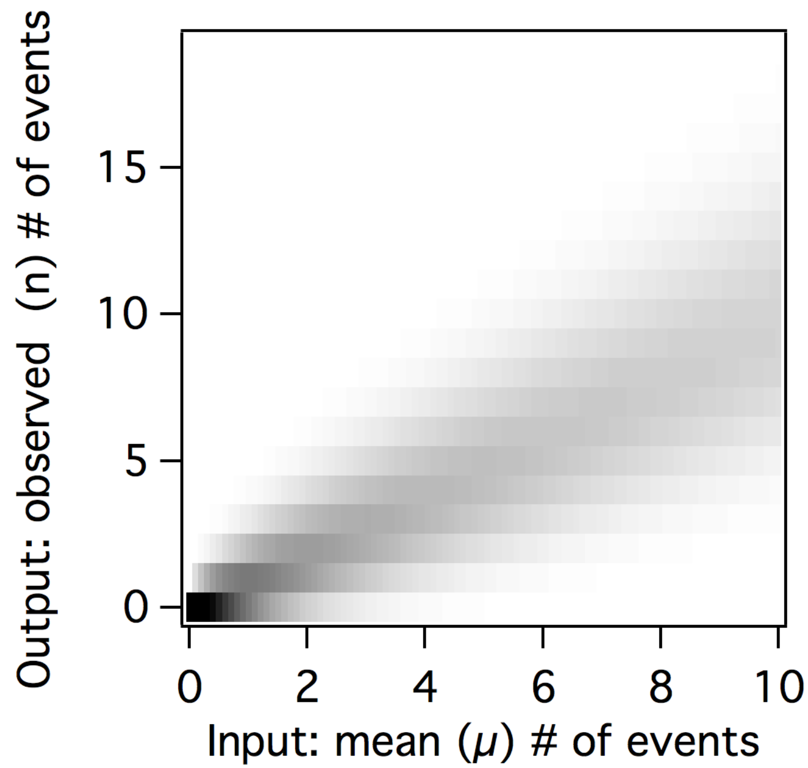
μ = mean # of events in a time interval

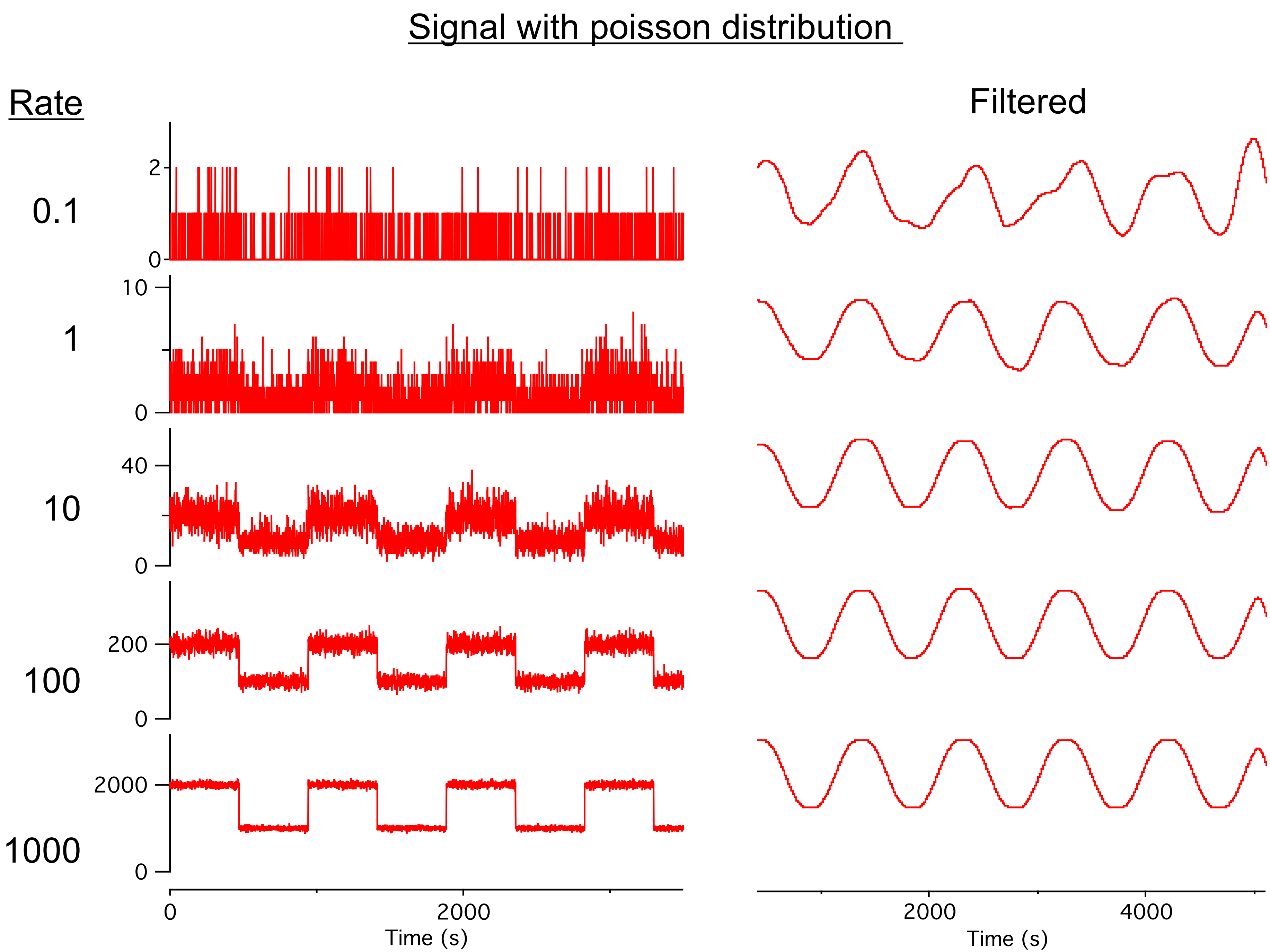
n = events in a time interval



variance=mean=μ

Joint probability distribution $\mathrm{P}[n,\mu]$

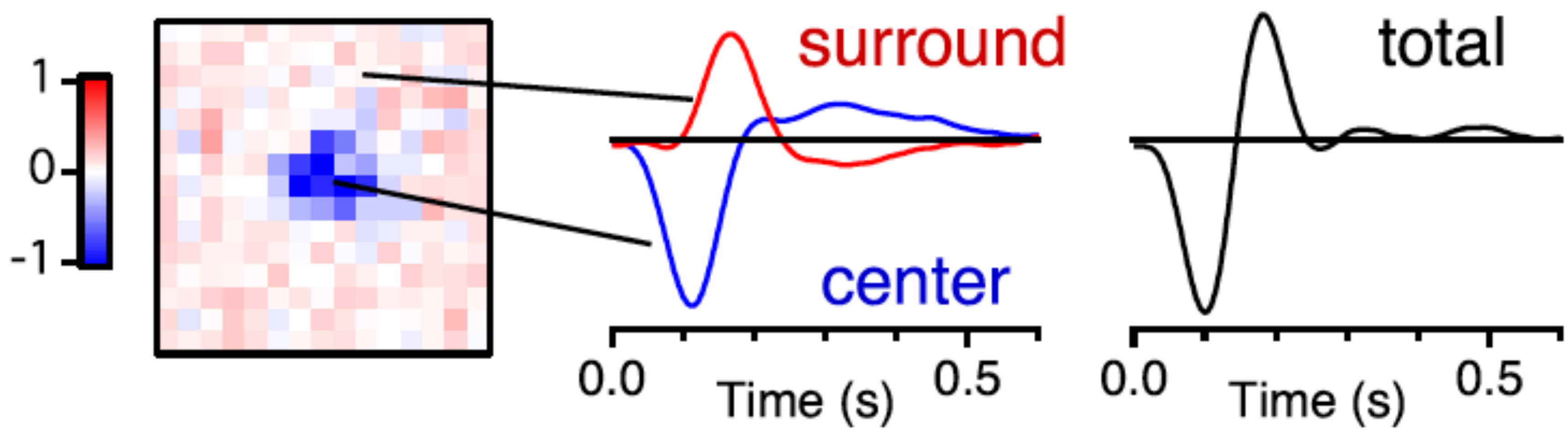




Signal with poisson distribution
Rate
Filtered
0.1
1
10
100
1000
2
0
10
0
40
0
200
0
2000
0
0
2000
Time (s)
2000
4000
Time (s)

What receptive field maximizes information transmission?

Retinal bipolar cell receptive field

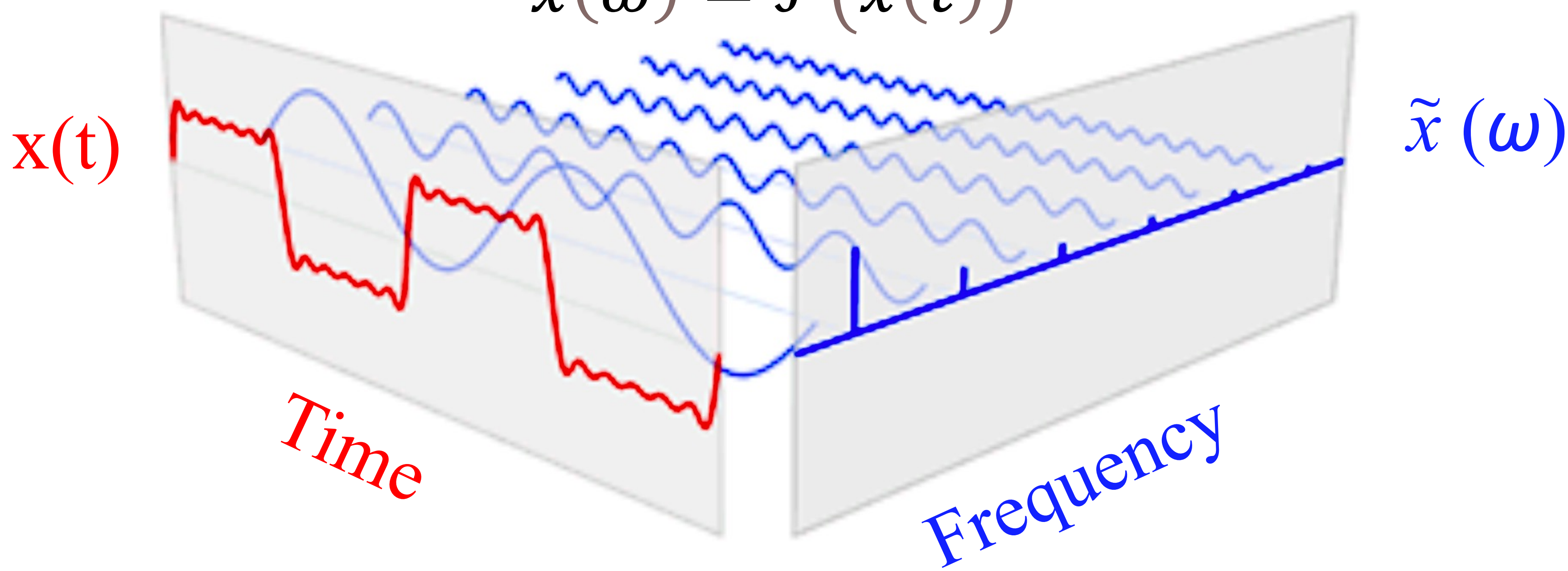


Baccus, Olveczky, Manu & Meister, 2008

Fourier transform

- Represents a function as a weighted sum of sinusoids
- Projects the function onto an orthogonal basis - sinusoids

$$\tilde{x}(\omega) = \mathcal{F}\left(x(t)\right)$$



- Fourier sine transform, cosine transform

$$F_{sin}(\omega) = \int_{-\infty}^{\infty} f(x)\sin(2\pi\omega x)dx \qquad F_{cos}(\omega) = \int_{-\infty}^{\infty} f(x)\cos(2\pi\omega x)dx$$

Fourier transform

- Represents a function as a weighted sum of sinusoids
- Projects the function onto an orthogonal basis - sinusoids

- Fourier sine transform, cosine transform

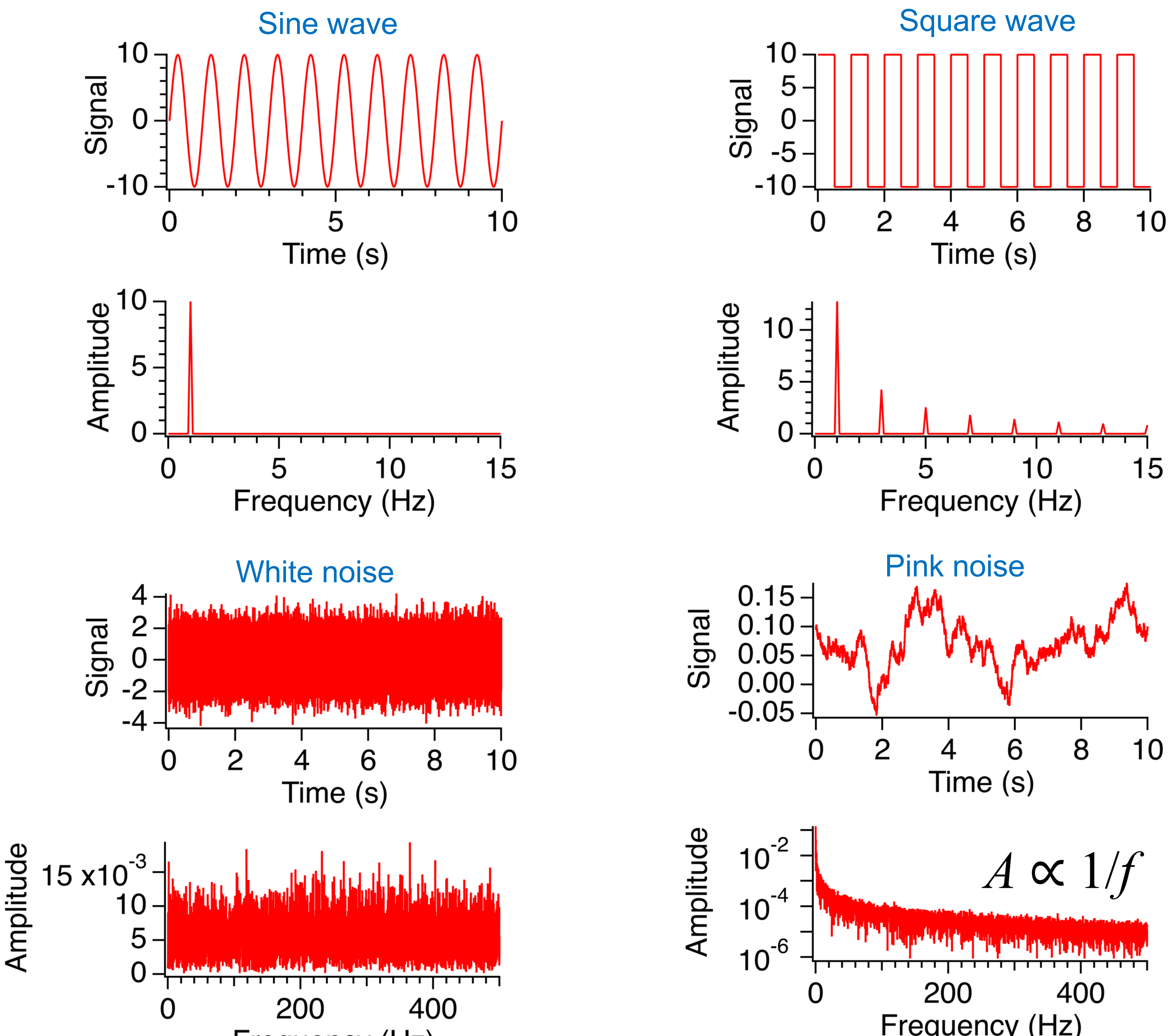
$$F_{sin}(\omega) = \int_{-\infty}^{\infty} f(x)\sin(2\pi\omega x)dx \qquad F_{cos}(\omega) = \int_{-\infty}^{\infty} f(x)\cos(2\pi\omega x)dx$$

- Euler's equation $\quad e^{i\theta} = \cos(\theta) + i\sin(\theta)$

- Complex Fourier transform

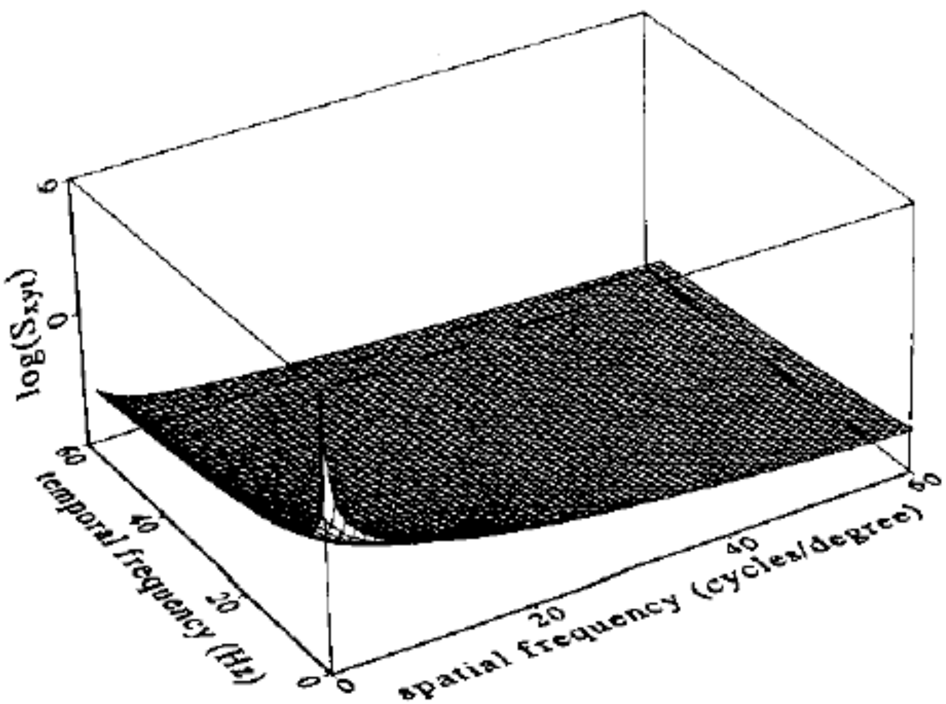
$$F(\omega) = \int_{-\infty}^{\infty} f(x)e^{-i2\pi\omega x}dx$$

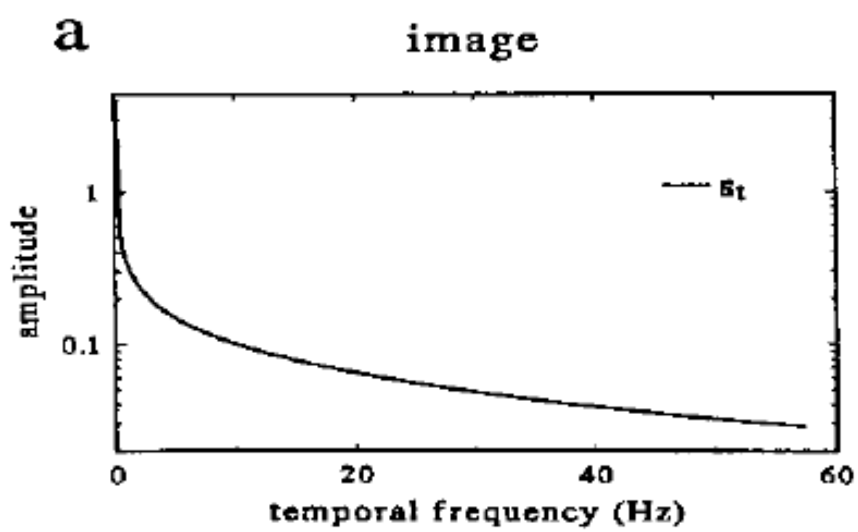
Fourier amplitude spectrum examples



<u>Theory of maximizing information in a noisy neural system</u>

Natural visual scenes are dominated by low spatial and temporal frequencies





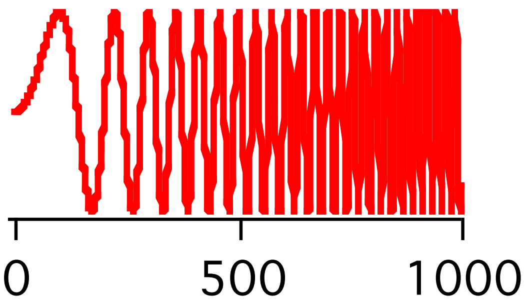
J.H. van Hateren. Real and optimal neural images in early vision. *Nature* 360:68-70 (1992)

J.H. van Hateren, Spatiotemporal contrast sensitivity of early vision. *Vision Res.*, 33:257-67 (1993)

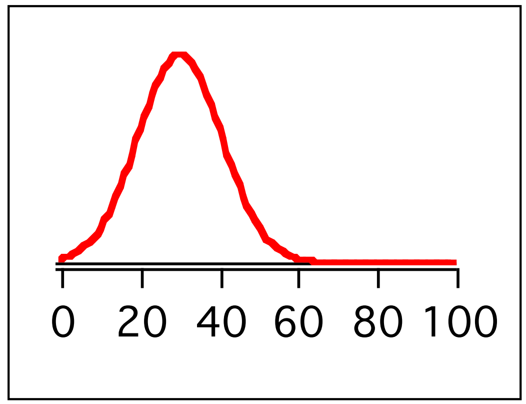
Atick & Redlich, What does the retina know about natural scenes? *Neur. comp.*, *4*: 196-210 (1992)

Linear filter and frequency response

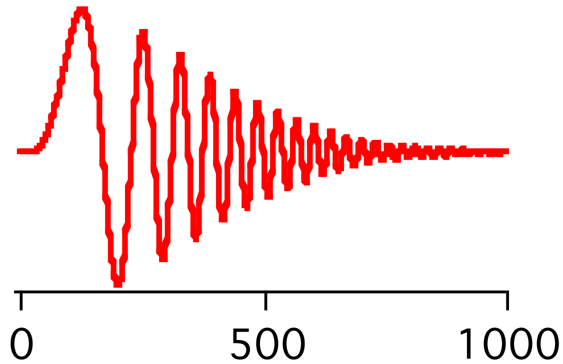
Stimulus



Filter



Response



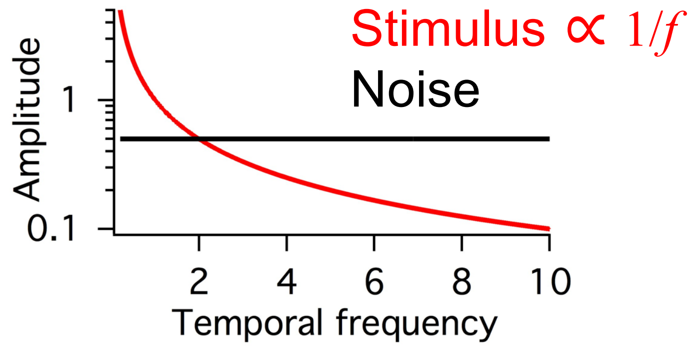
Convolution theorem

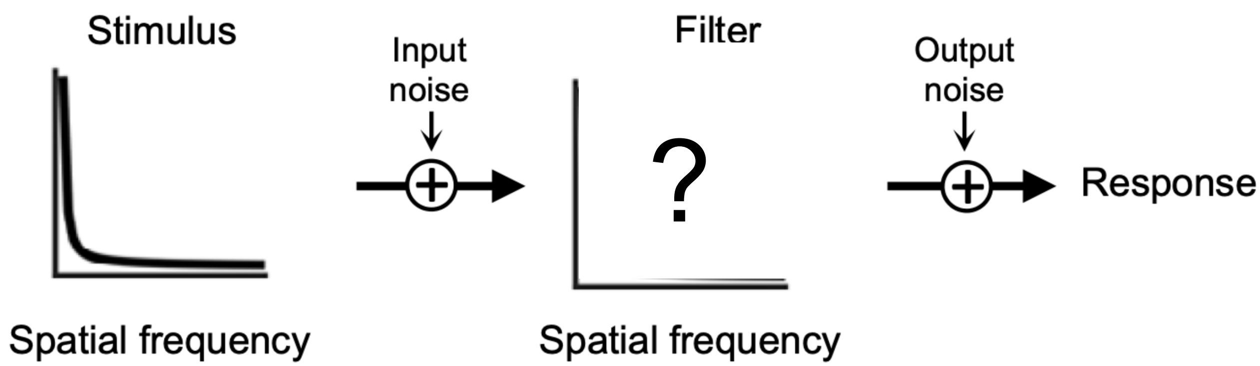
$$h(t) = f(t) * g(t) \quad \Leftrightarrow \quad \tilde{h}(\omega) = \tilde{f}(\omega)\tilde{g}(\omega)$$

a convolution in the time domain

is a simple product in the frequency domain

What filter maximizes information?





Mutual information | Total entropy | Conditional ("noise") entropy

$$I(R;S) = H(R) - H(R|S)$$

$$I(S;R) = H(F(S+N_i)+N_e) - H(F(S+N_i)+N_e|S)$$

Distributing signal power for optimal transmission

Shannon-Hartley theorem

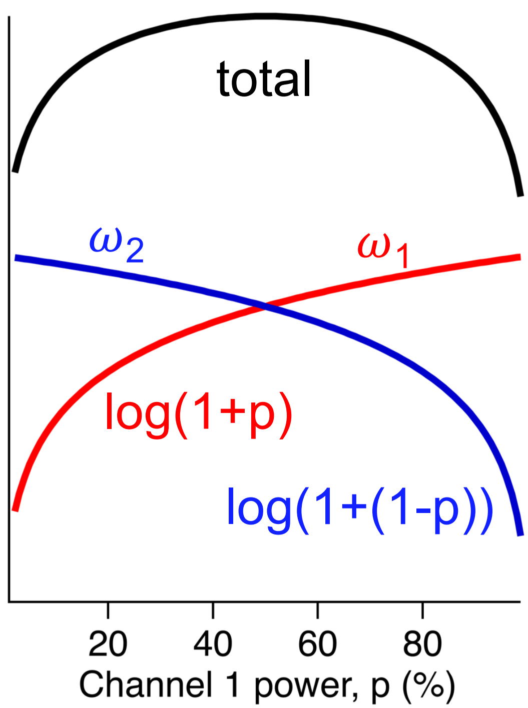
$$C = B\log_2\left(1+\frac{S}{N}\right)$$

Capacity
Bandwidth
Signal power
Noise power, additive Gaussian

For multiple frequencies, ω

$$C = \int_{-\infty}^{\infty} \log_2\left(1+\frac{S(\omega)}{N(\omega)}\right) d\omega$$

With a constraint on signal power,
$S(\omega) + N(\omega)$ should be constant



"Water-filling" procedure

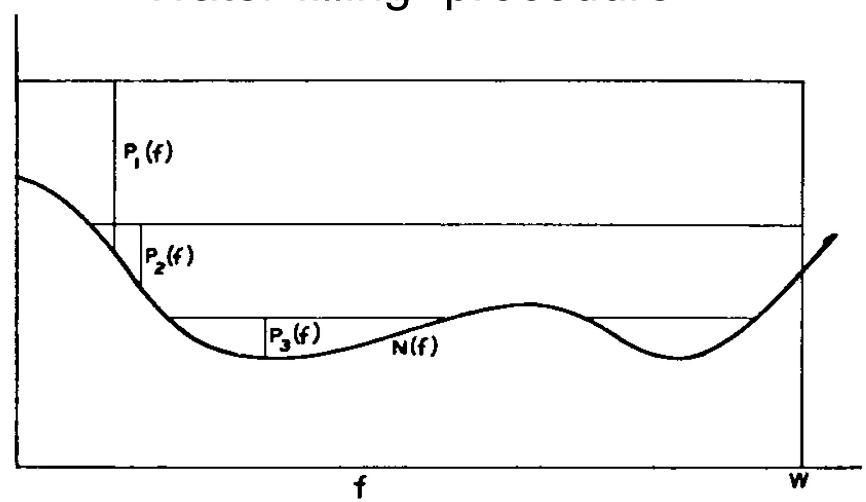


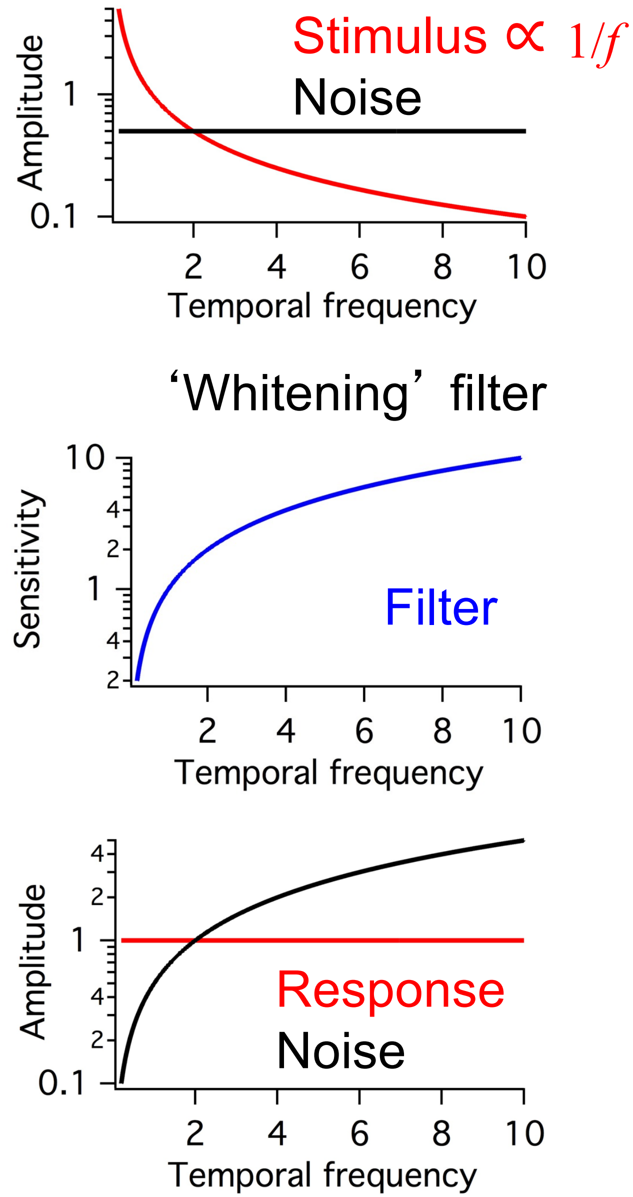
Fig. 8. Best distribution of transmitter power.

Shannon, 1949

What filter maximizes information?

increases total entropy but limits noise entropy - whitens but also cuts out noise

$$I(S;R) = H(F(S+N_i)+N_e) - H(F(S+N_i)+N_e|S)$$



Van Hateren (1993)
Atick & Redlich (1990)
Linsker (1988)

Adapting to different levels of noise

<u>High SNR – increase total entropy</u>

Filter to whiten in the presence of noise

<u>Low SNR – reduce noise entropy</u>

Low pass filter reduces noise

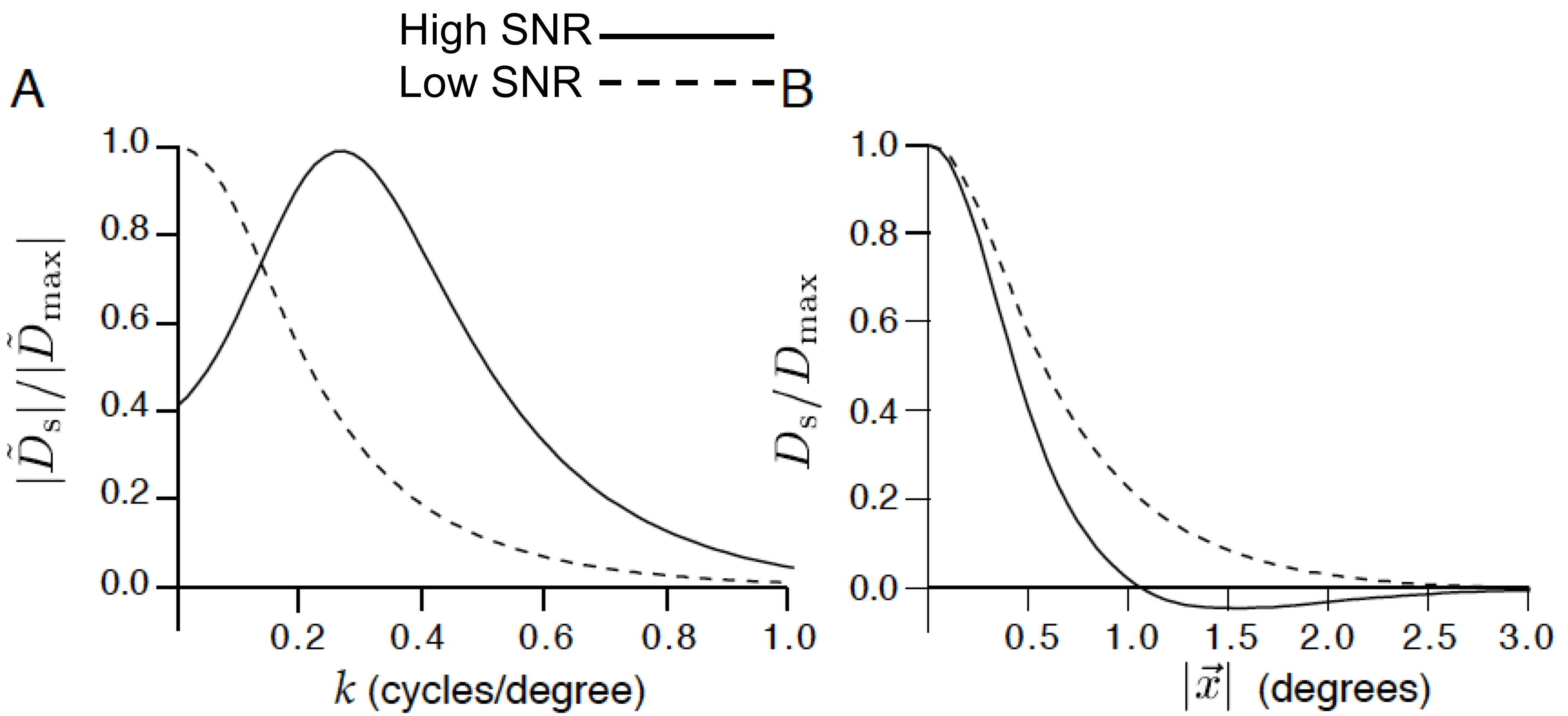
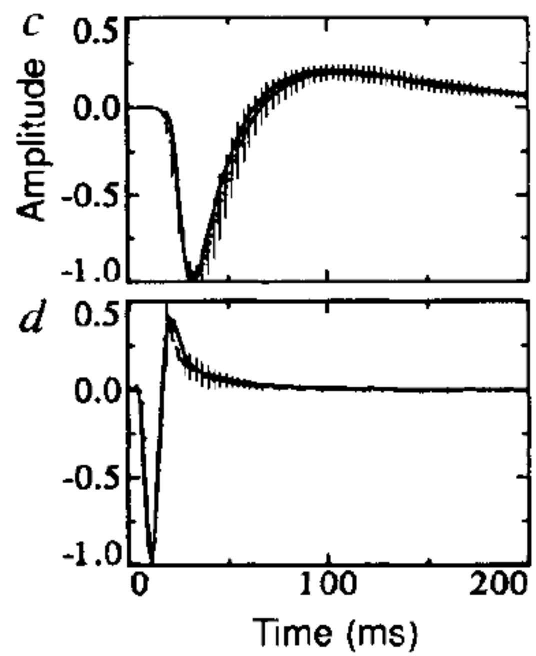


Figure 4.3: Receptive field properties predicted by entropy maximization

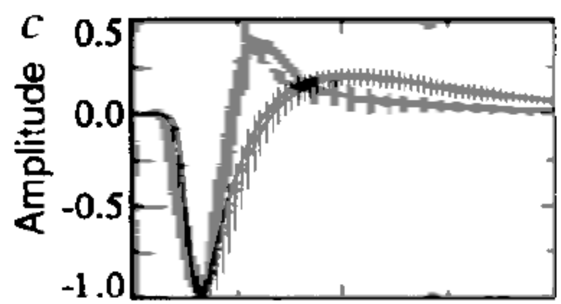
Theory of maximizing information in a noisy neural system

Filter of fly Large Monopolar Cells,
2nd order visual neuron



Low background intensity
Integrates over time
(real and theoretical optimum)

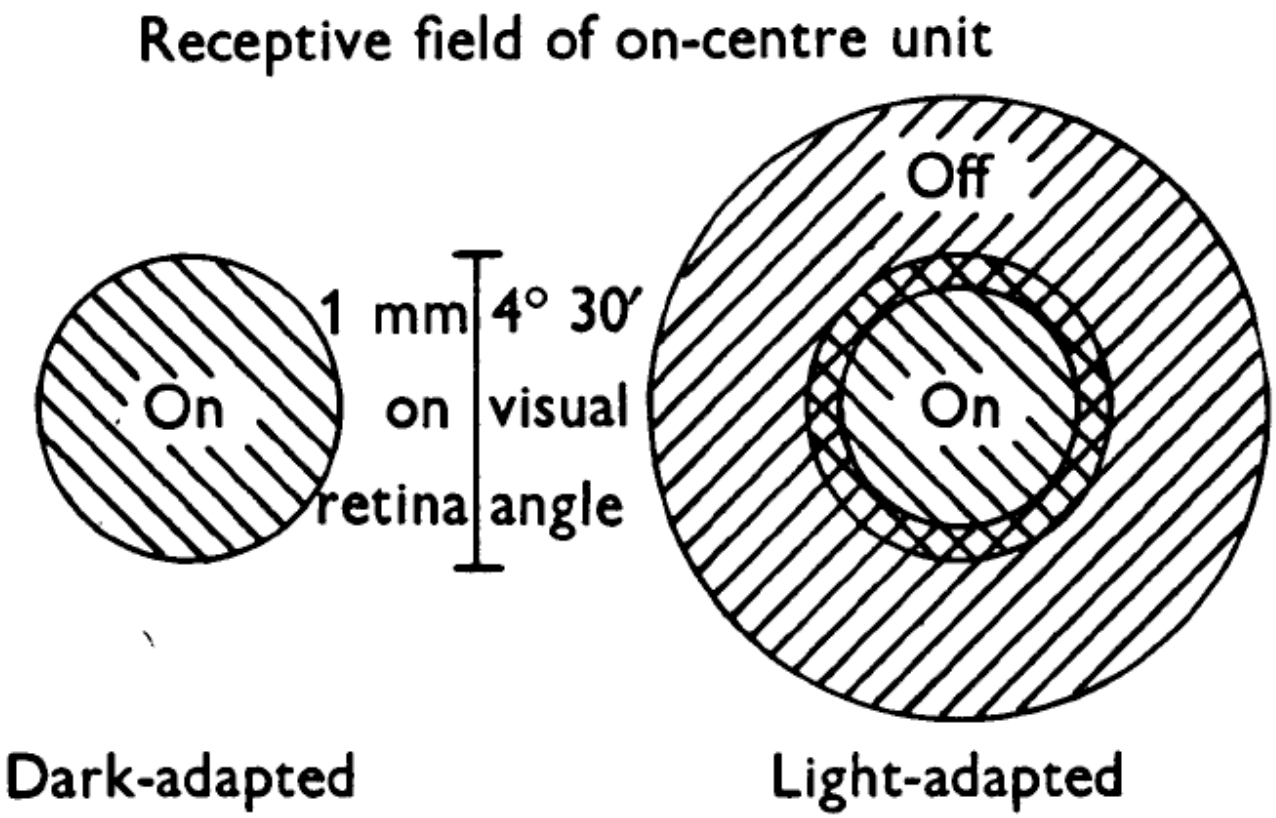
High background intensity
Emphasizes change, is more differentiating
(real and theoretical optimum)



Both, scaled in time to the first peak

J.H. van Hateren. Real and optimal neural images in early vision. *Nature* 360:68-70 (1992)

Spatial adaptation in retinal ganglion cells



Barlow, Fitzhugh & Kuffler (1957)

Theories of efficient coding:

To maximize information transmission, at high SNR when noise entropy is lower, an ideal encoder should increase total entropy and use all output values with equal probability

Low frequencies dominate in natural scenes

The highest frequencies should be rejected to reduce noise entropy as they carry little information

An efficient encoder at **high SNR** should amplify higher frequencies more than low frequencies with a **bandpass filter**

But at **low SNR** when noise entropy is high, most higher frequencies should be rejected and low frequencies should be used, shifting to a **lowpass filter**