NEPR208 - Optimality and adaptation

Stephen Baccus
Department of Neurobiology
baccus@stanford.edu

How?

What are the mechanisms of a function?

Why?

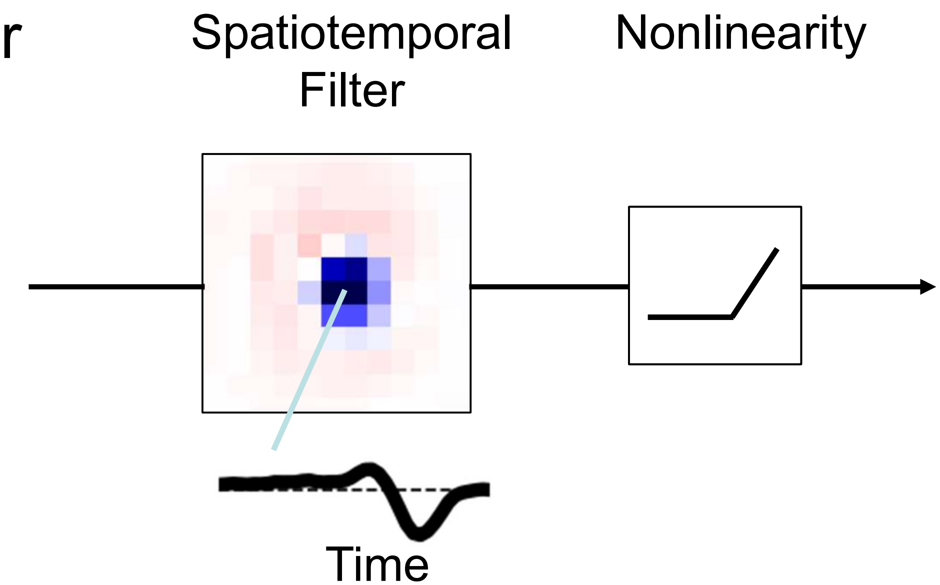
What are the functional benefits …

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?

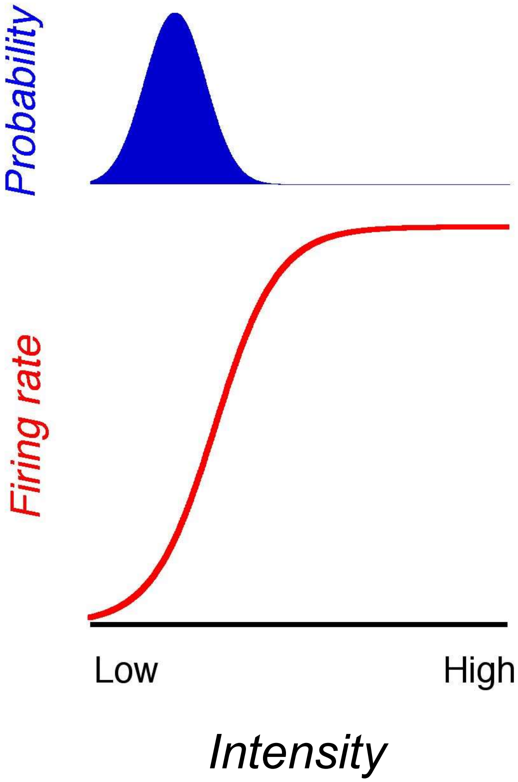
Why do cells have a certain duration filter?

Why do they have a certain shape filter?

Why does the filter change?

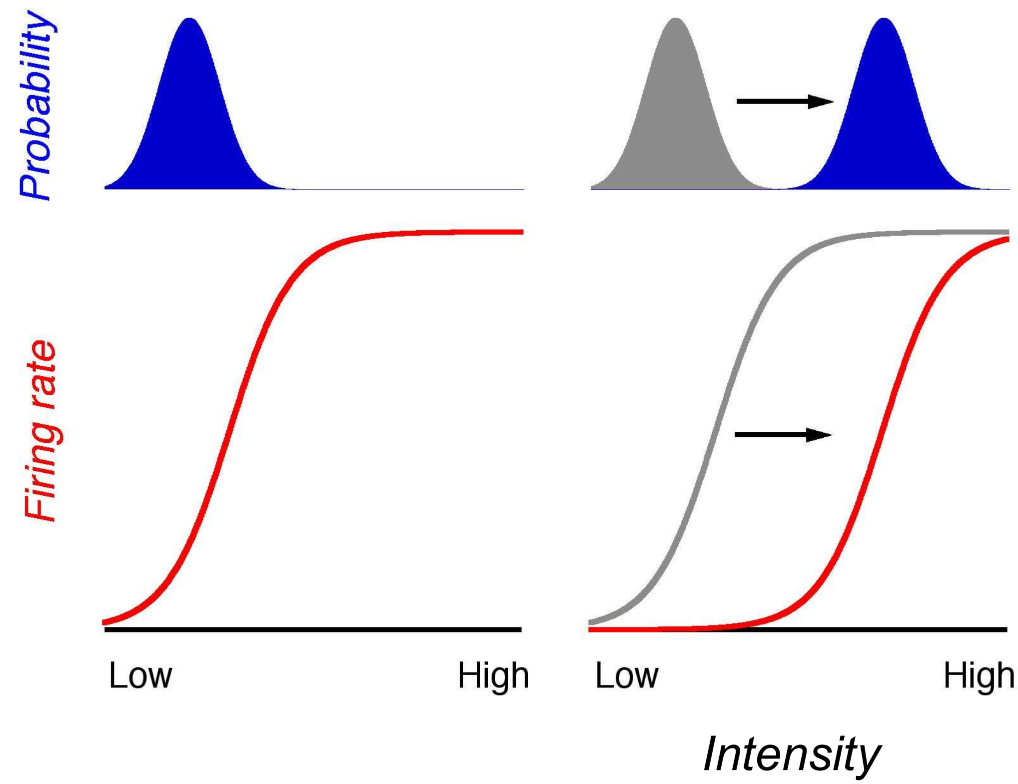
How can the nonlinearity and filter change?

Adaptation to the average input

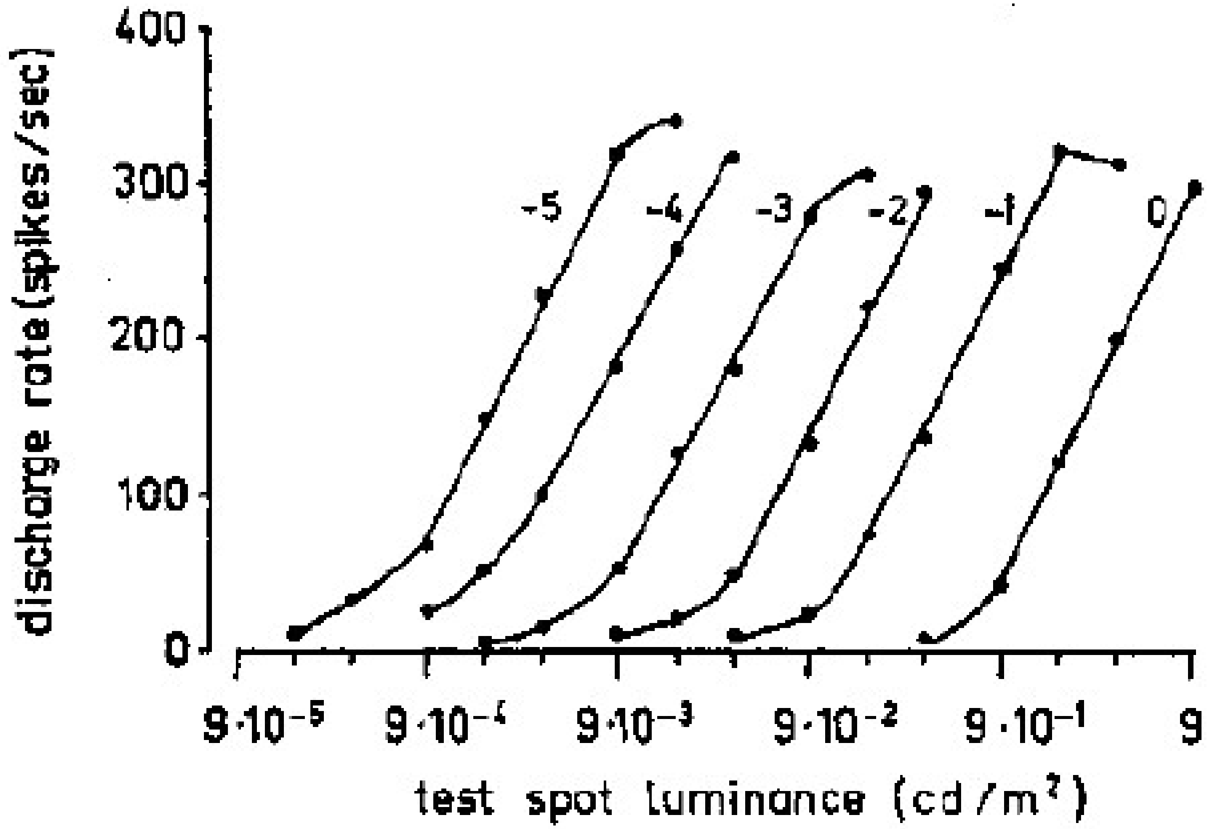


Adaptation to the average input

Light adaptation

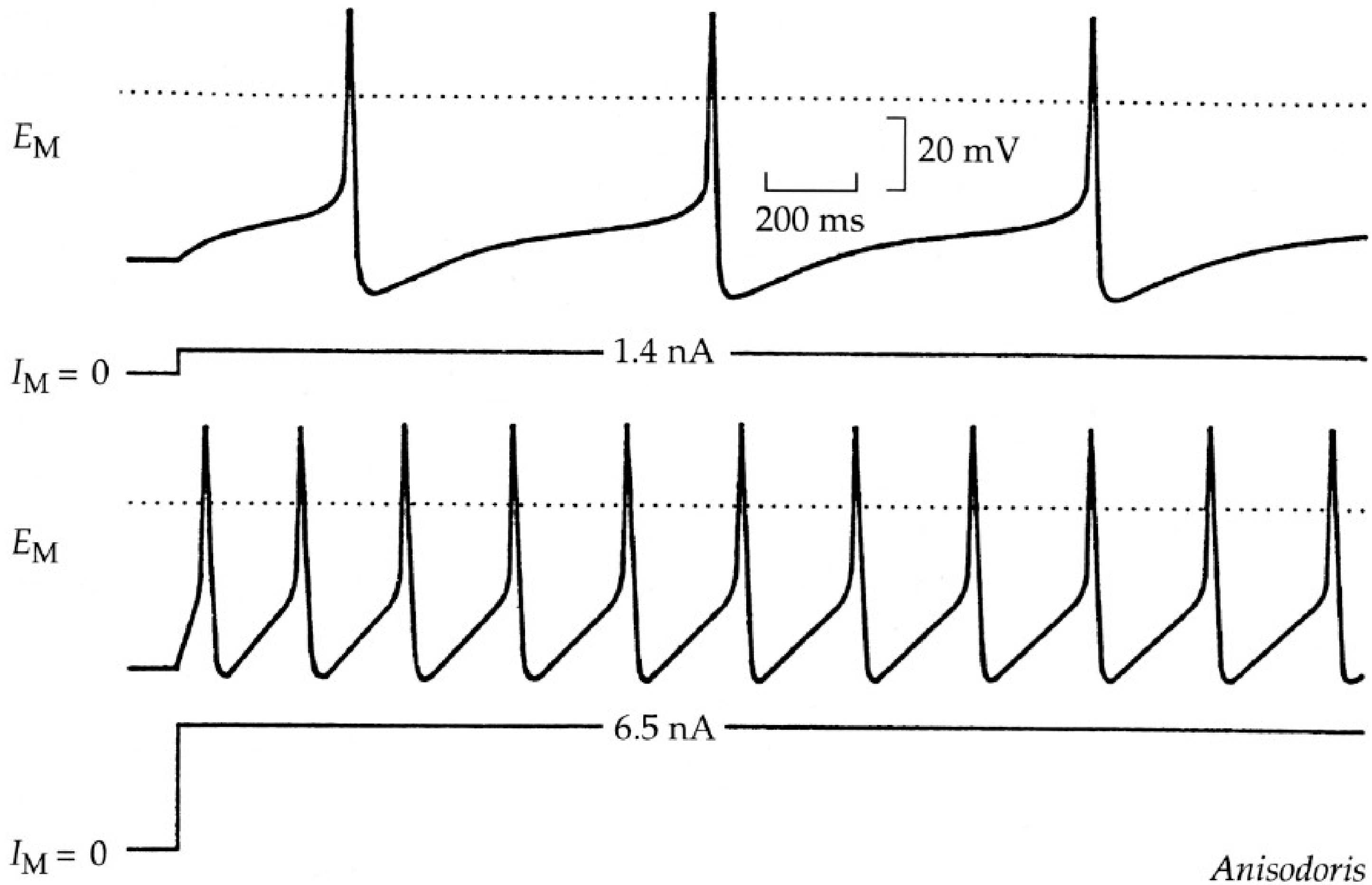


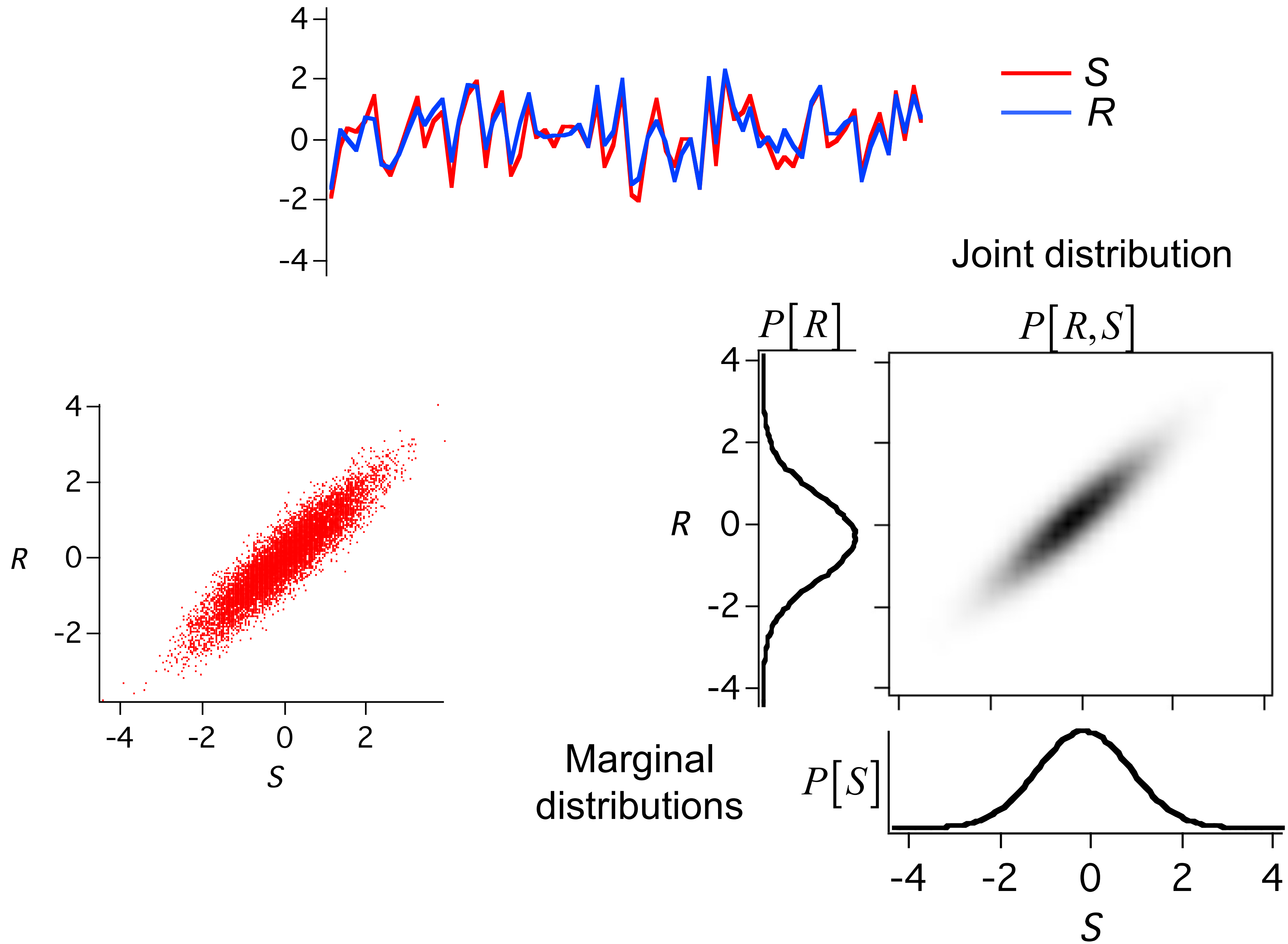
Ganglion cell response curves shift to the mean light intensity



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

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





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

A Mathematical Theory of Communication

Claude Shannon (1948)

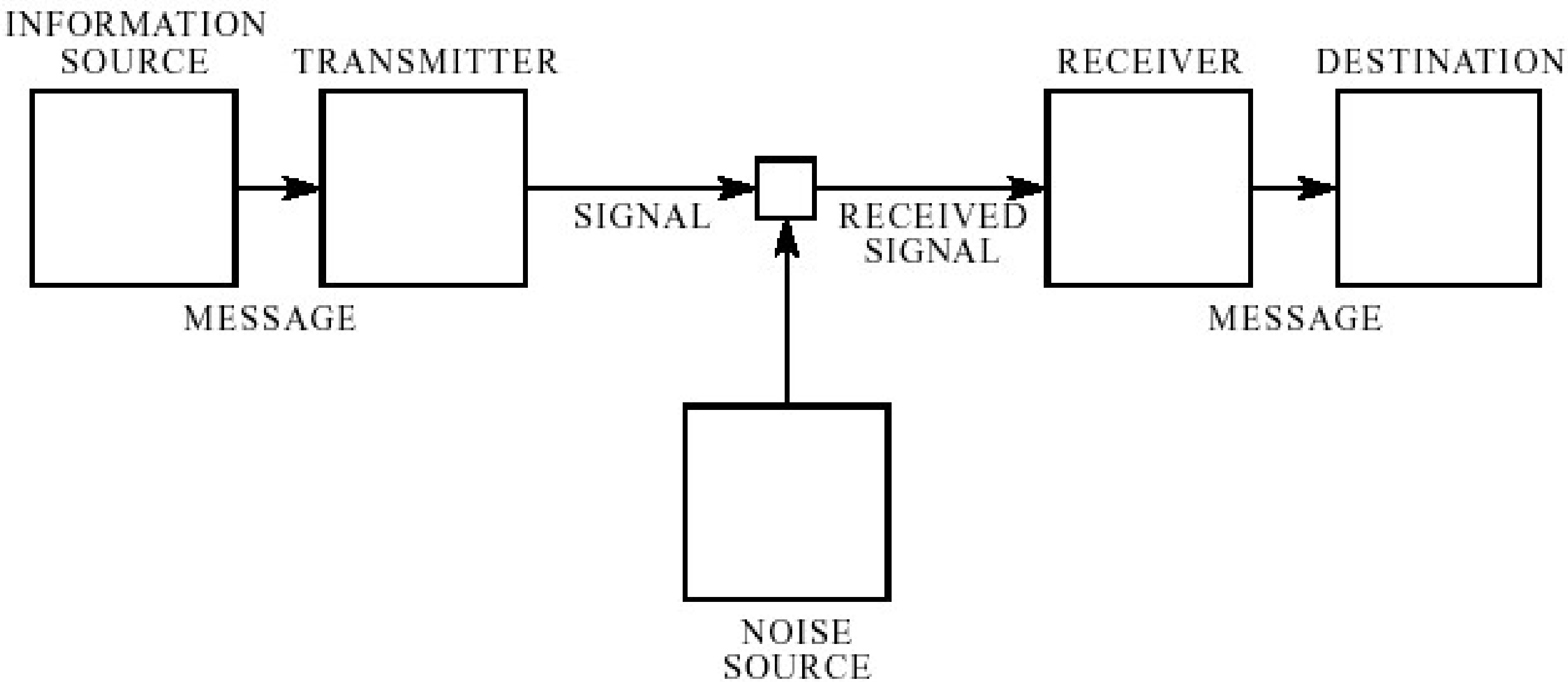


Fig. 1 — Schematic diagram of a general communication system.

Bell System Technical Journal 27, 379-423

A Mathematical Theory of Communication

Claude Shannon (1948)

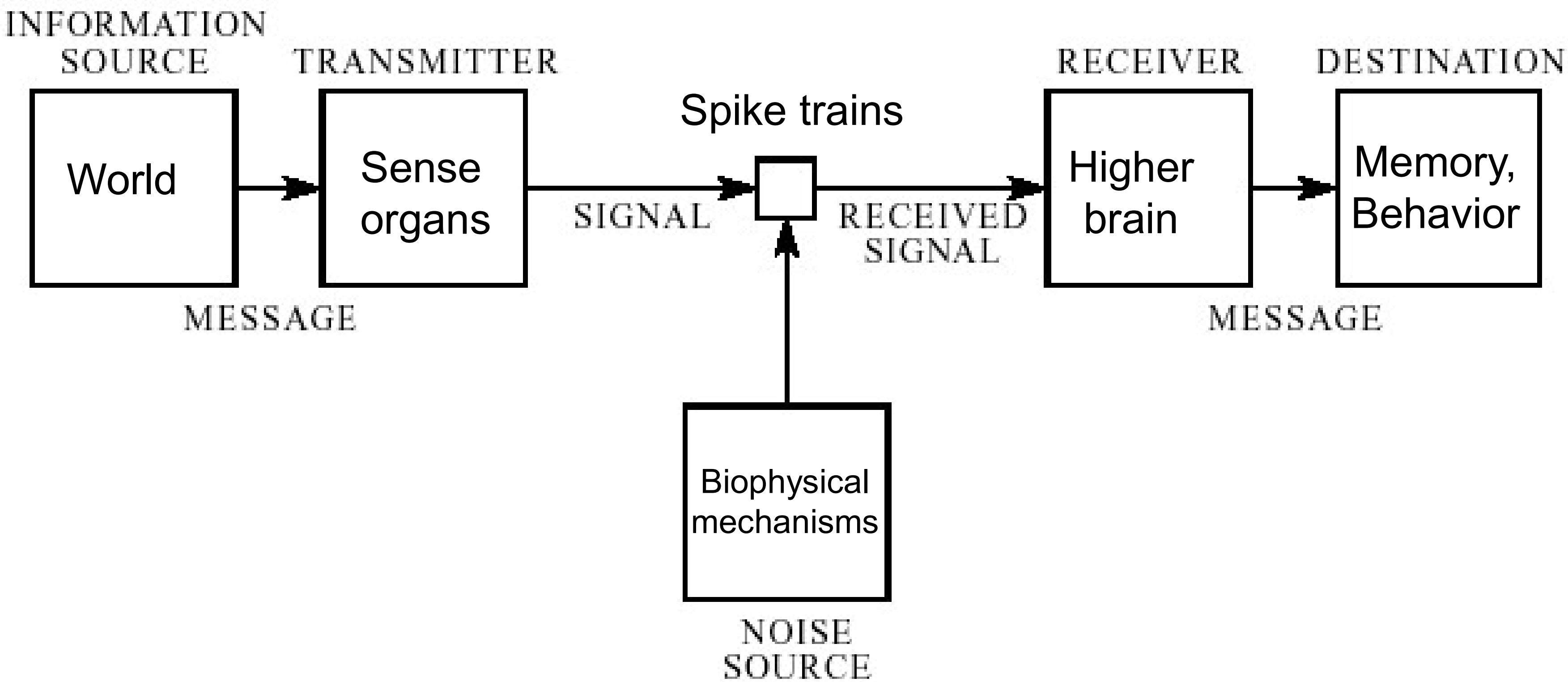


Fig. 1 — Schematic diagram of a general communication system.

Bell System Technical Journal 27, 379-423

A Mathematical Theory of Communication

Claude Shannon (1948)

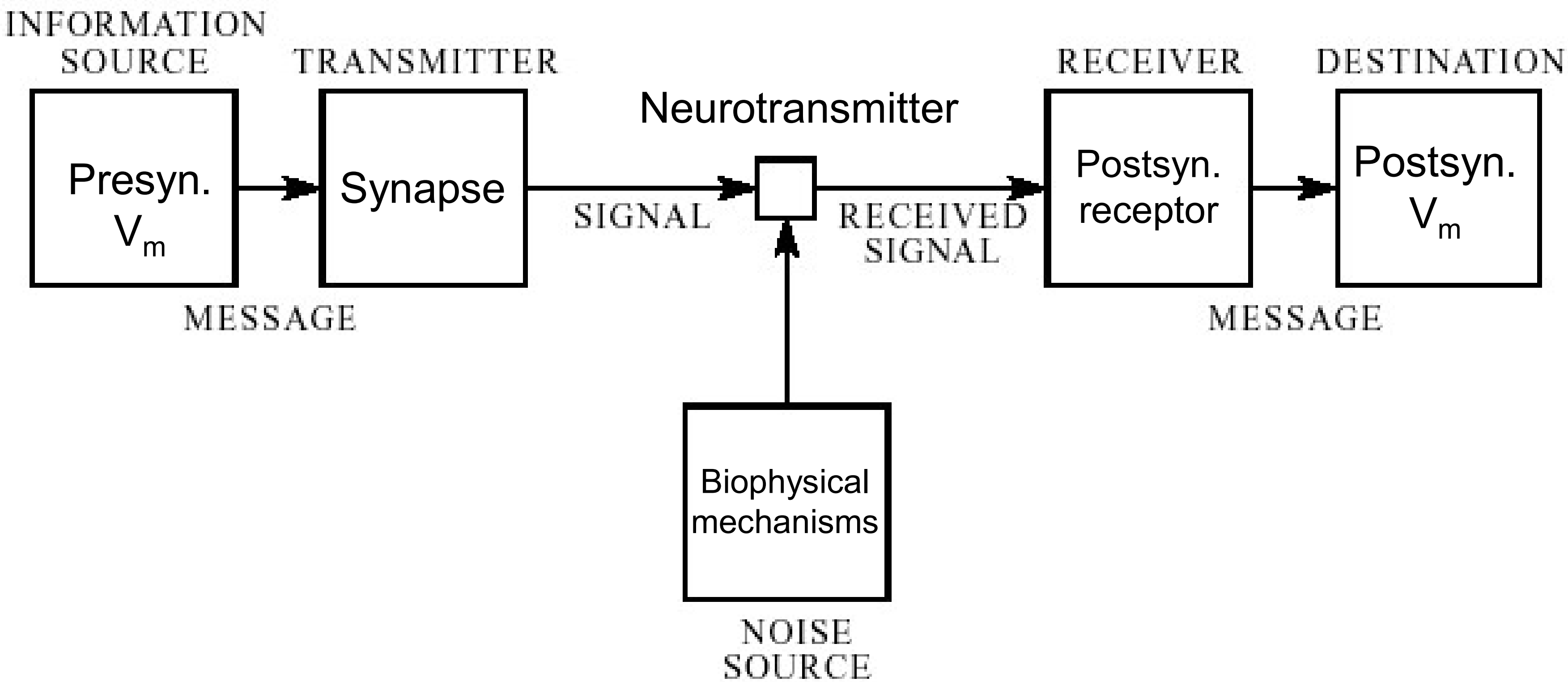


Fig. 1 — Schematic diagram of a general communication system.

Bell System Technical Journal 27, 379-423

A Mathematical Theory of Communication

Claude Shannon (1948)

What is information?

Entropy*

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

$$H(X) = -\sum_i P[x_i] \log P[x_i]$$

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

0 log(0)=0

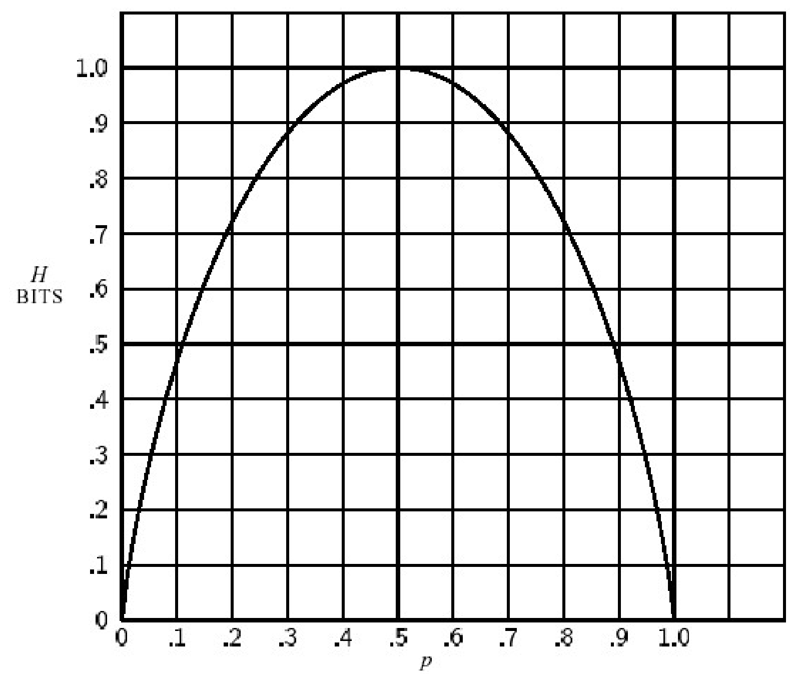


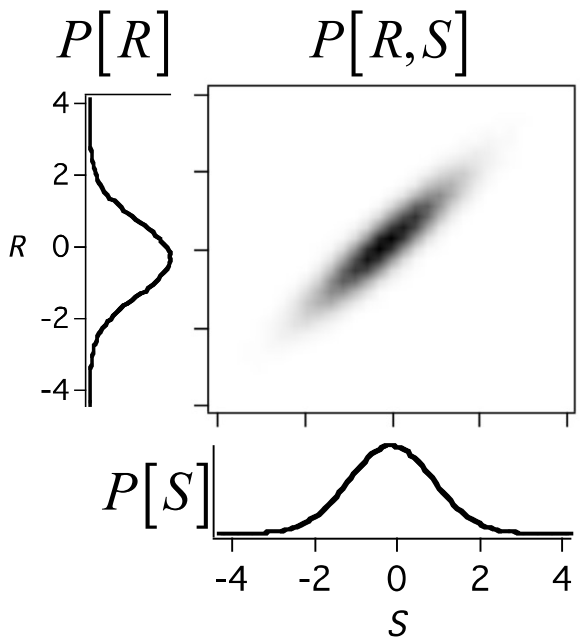
Fig. 7—Entropy in the case of two possibilities with probabilities p and $(1-p)$.

*By analogy to entropy in statistical mechanics,

$$S = k \log W$$

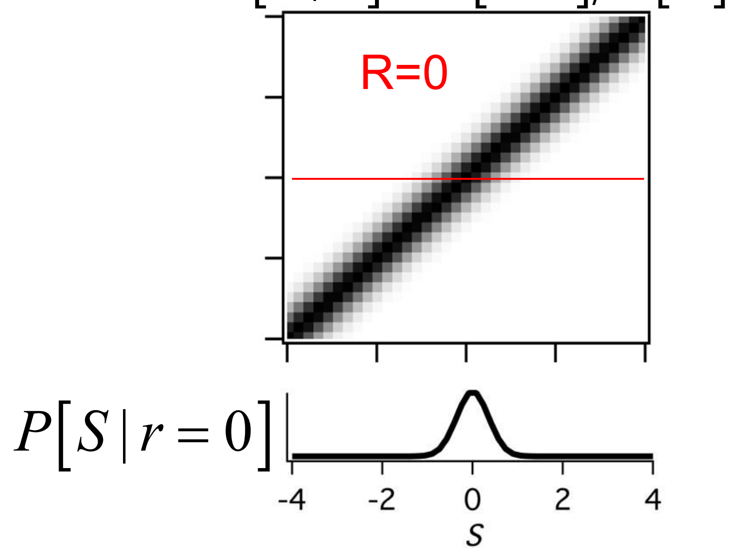
k: Boltzmann constant W: Number of possible microscopic states

Information is a reduction in entropy



Conditional distribution

$$P[S|R] = P[R,S]/P[R]$$



Conditional entropy

$$H(S|R) = -\sum_s \sum_r P(r,s)\log\left(P(s|r)\right)$$

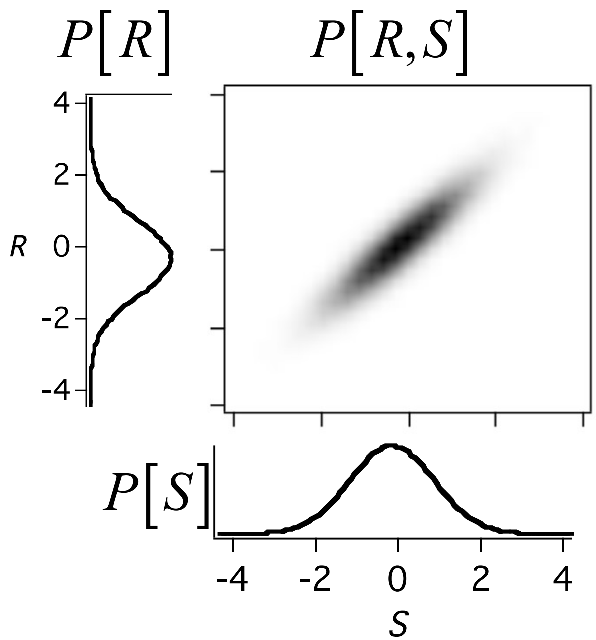
Mutual information

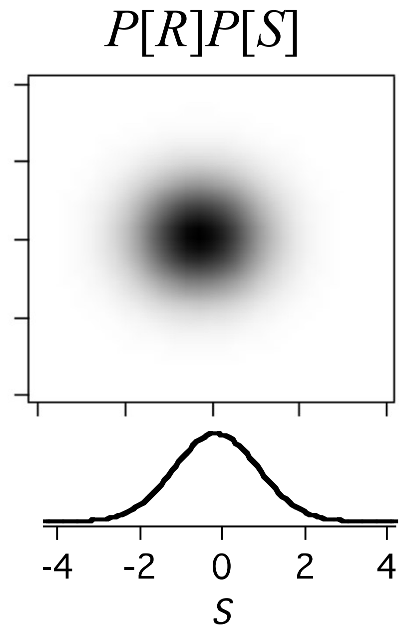
A measure, in bits, of how much information is conveyed by one random variable about another random variable. It is equal to the total entropy minus the conditional entropy.

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

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

Mutual information as the 'distance' between two probability distributions





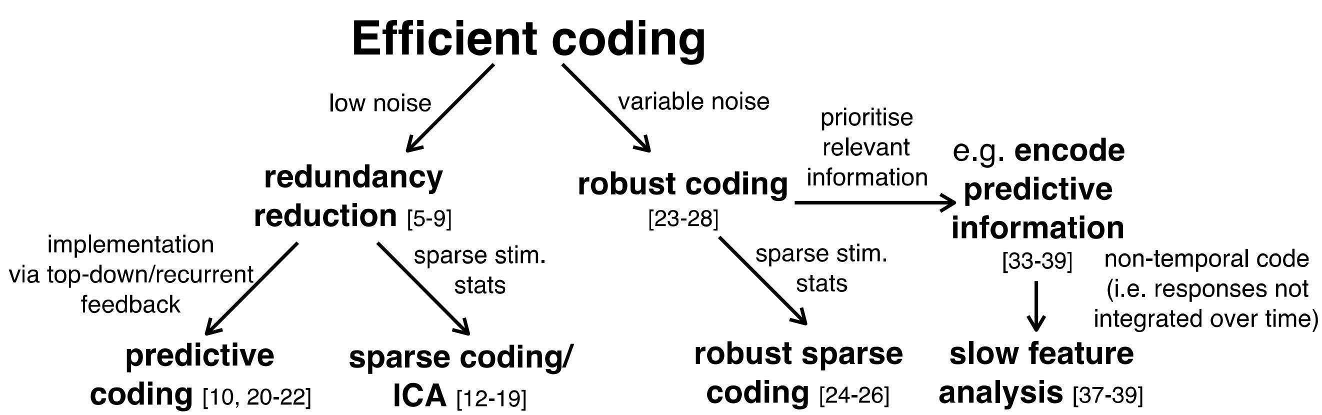
$$I(R;S)=\sum_i \sum_j P\left[R_i,S_j\right]\log\left(\frac{P\left[R_i,S_j\right]}{P\left[R_i\right]P\left[S_j\right]}\right)$$

Does the early visual system maximize information transmission?

'Efficient Coding' - Horace Barlow

Mutual information	Total entropy	Conditional ("noise") entropy

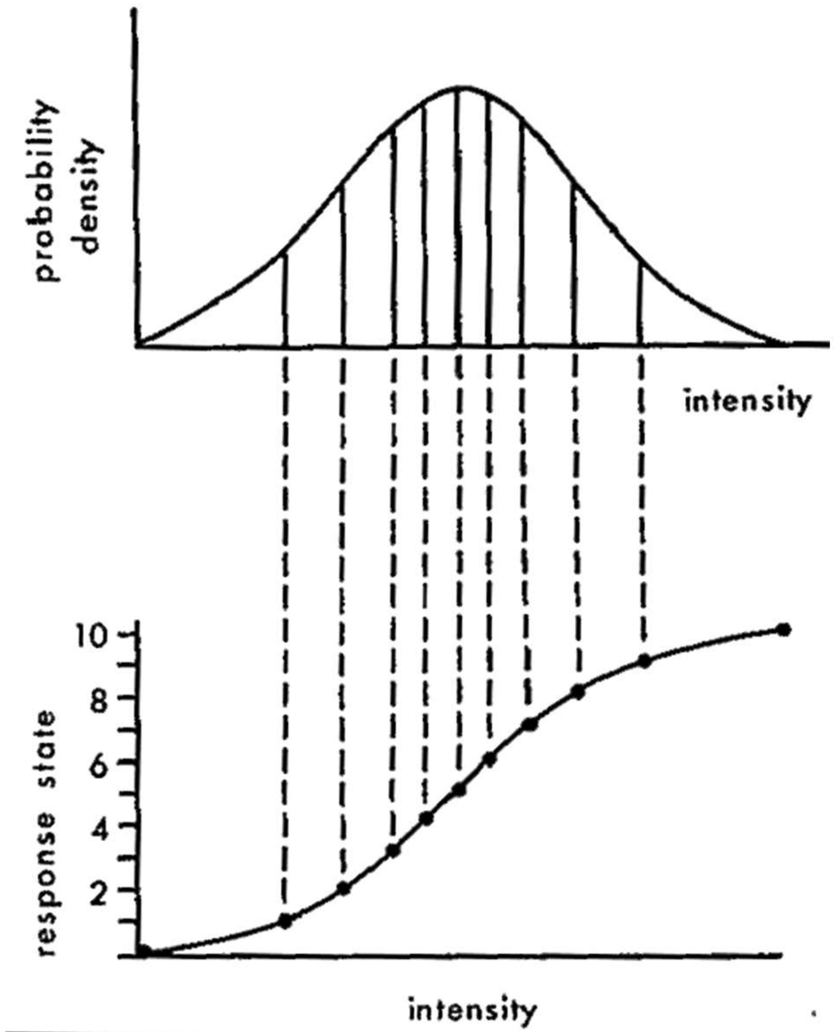
$$I(S;R) = H(S) - H(S|R)$$

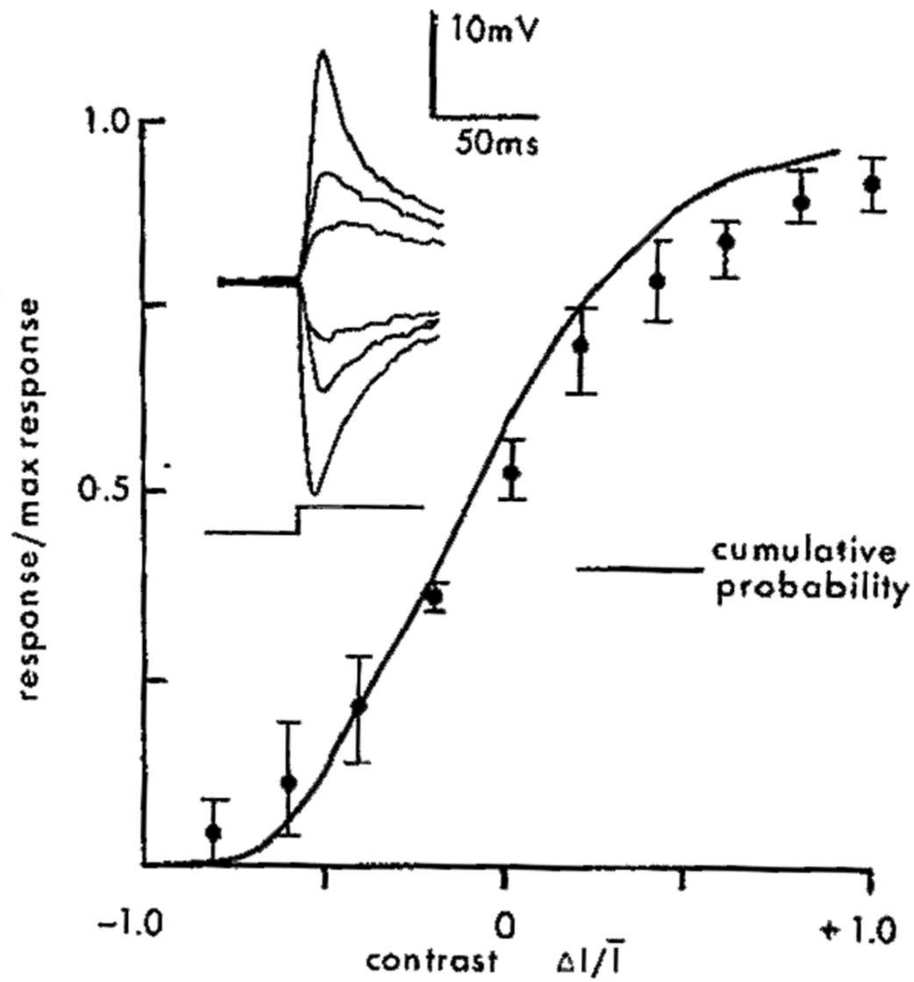


References in notes section

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.

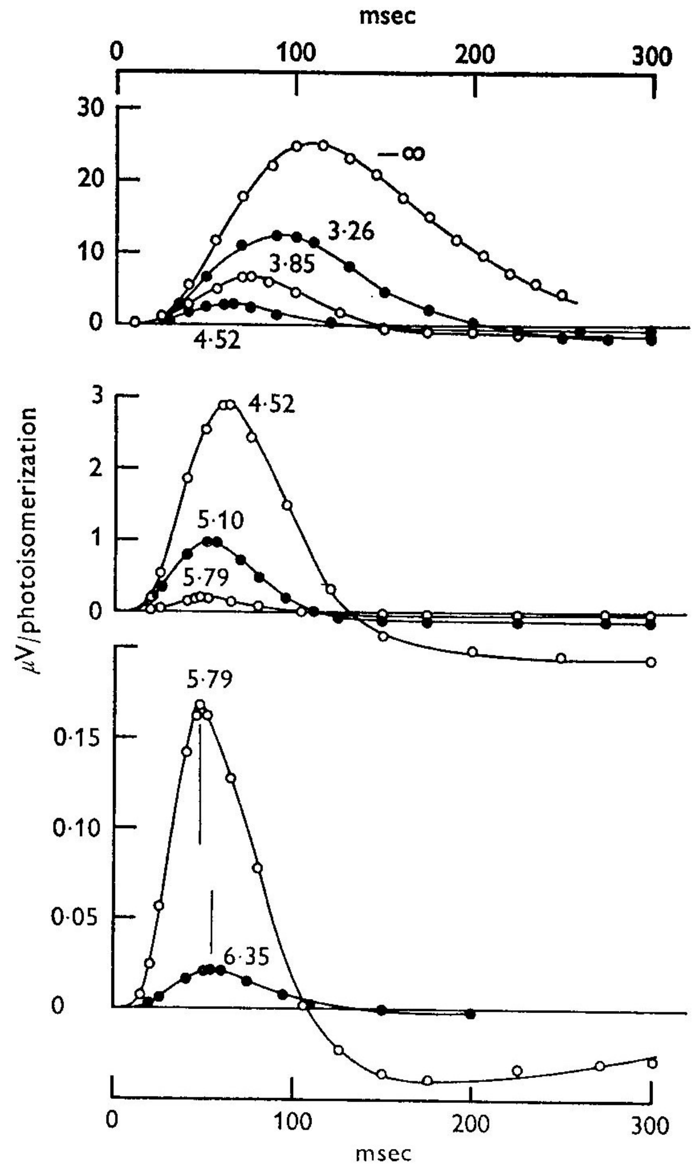
Maximizing information by a nonlinearity that maximizes total entropy





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



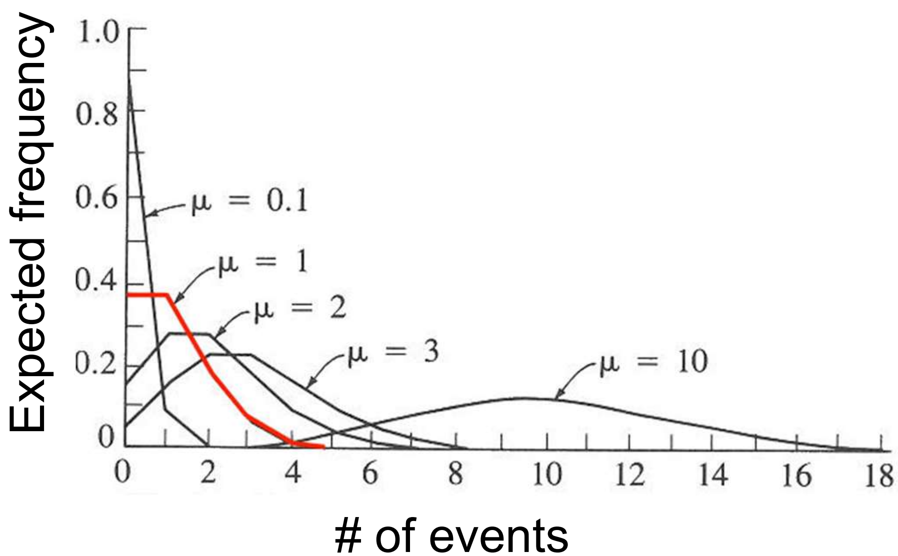
Baylor & Hodgkin 1974

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

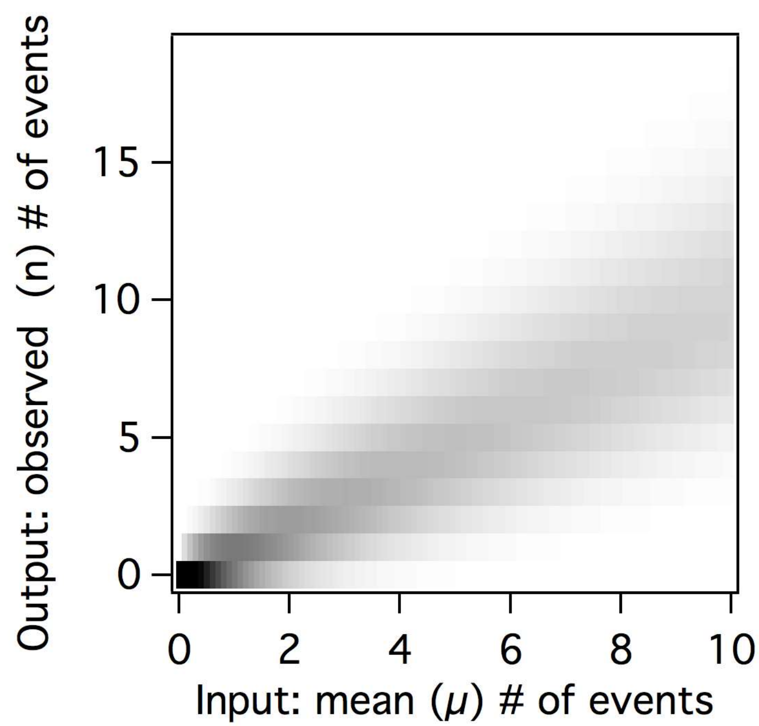
$$\frac{e^{-\mu}\mu^{n}}{n!}$$

μ = mean # of events in a time interval

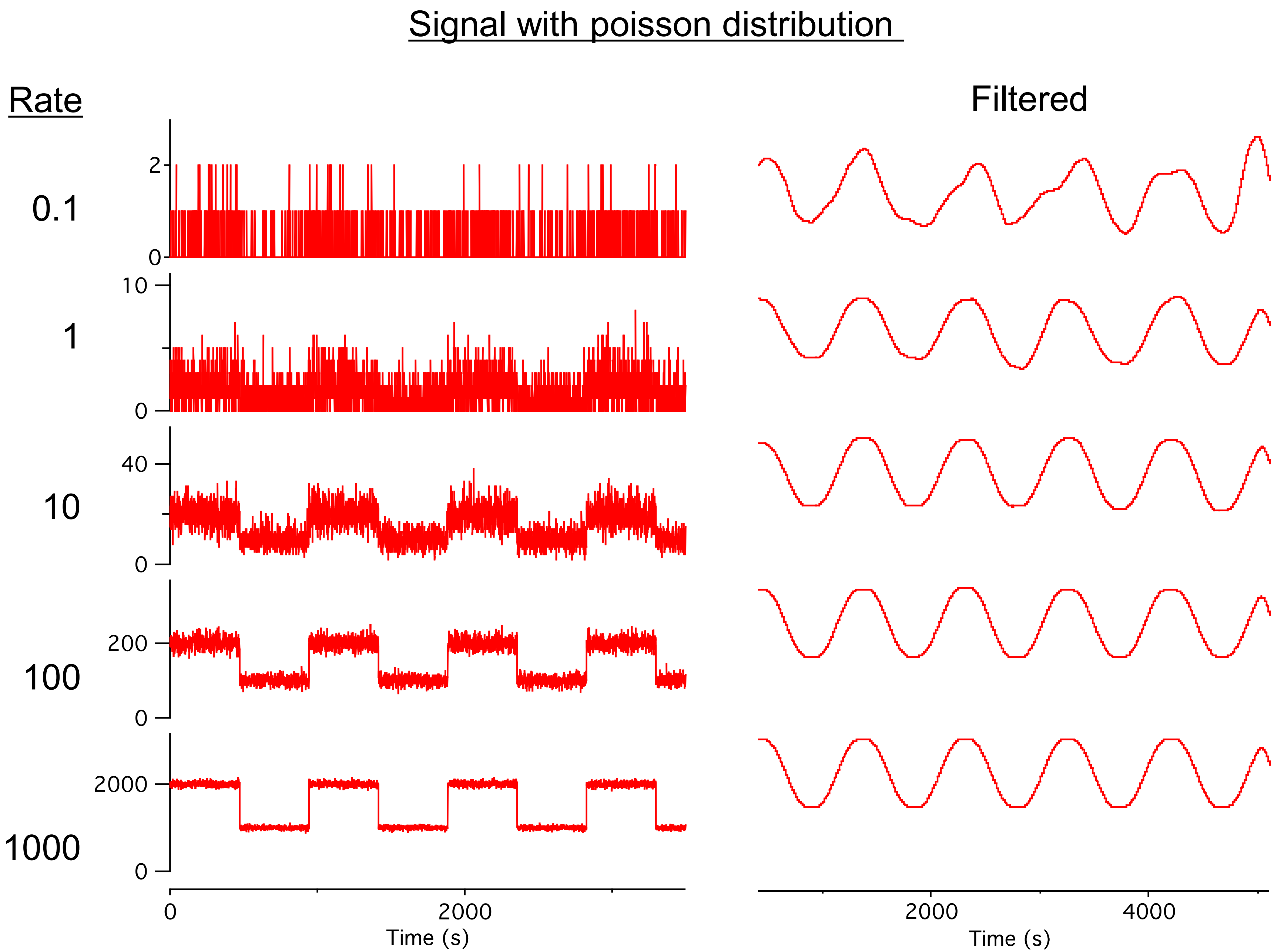
n = events in a time interval



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

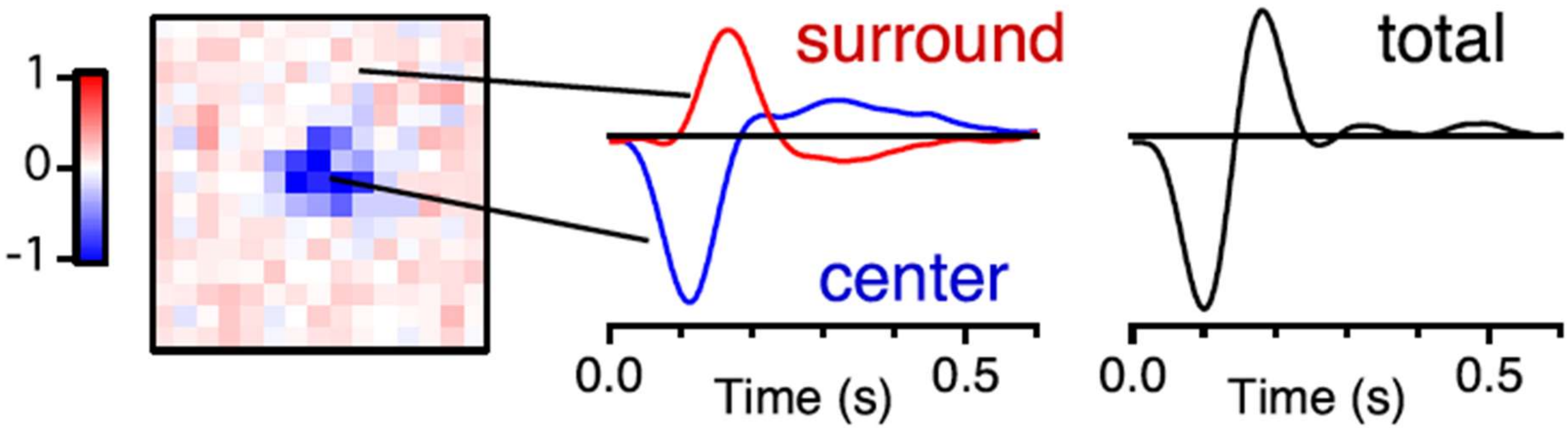


variance=mean=μ


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?

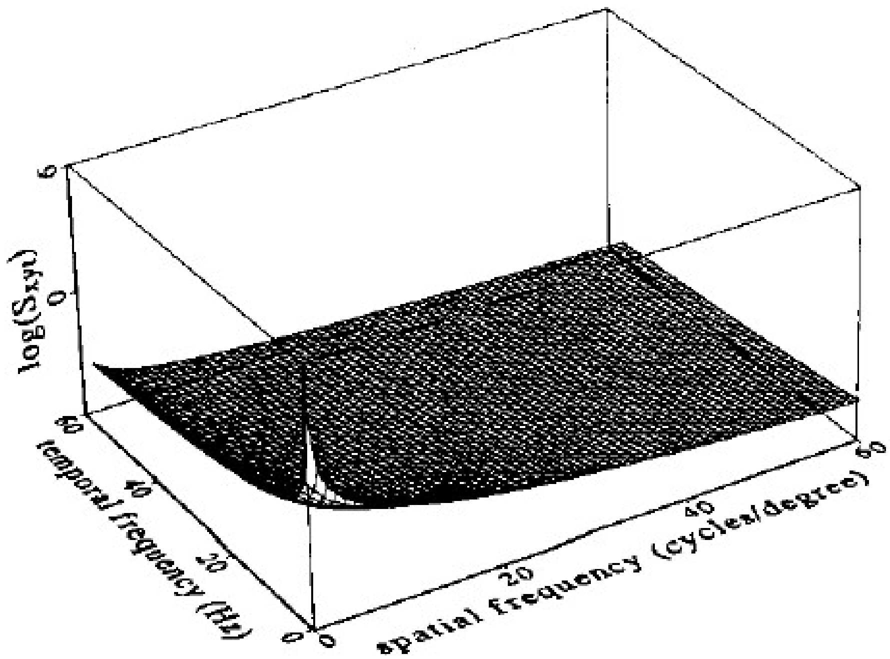
Retinal bipolar cell receptive field

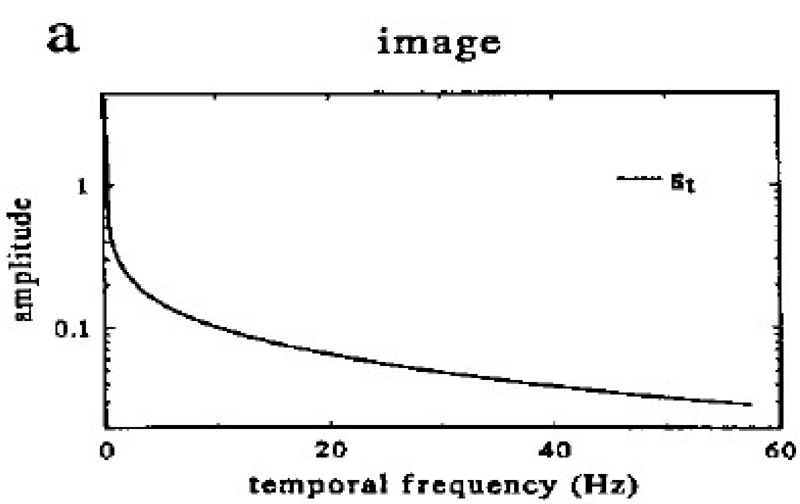


Baccus, Olveczky, Manu & Meister, 2008

Theory of maximizing information in a noisy neural system

Natural visual scenes are dominated by low spatial and temporal frequencies



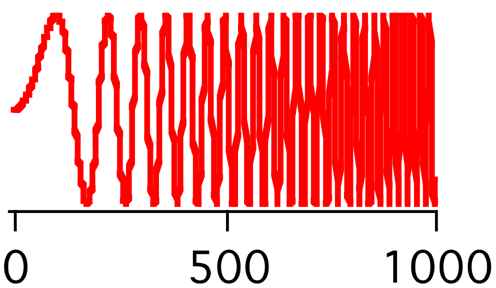


J.H. van Hateren. Real and optimal neural images in early vision. *Nature* 360:68-70 (1992)

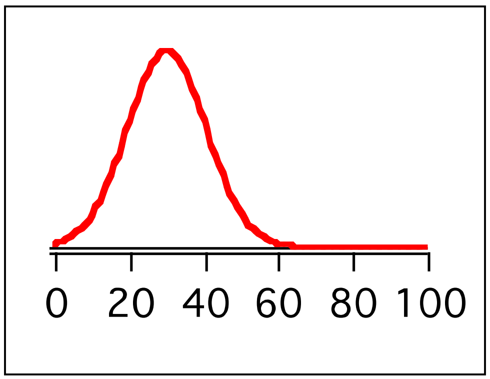
J.H. van Hateren, Spatiotemporal contrast sensitivity of early vision. *Vision Res.*, 33:257-67 (1993)

Linear filter and frequency response

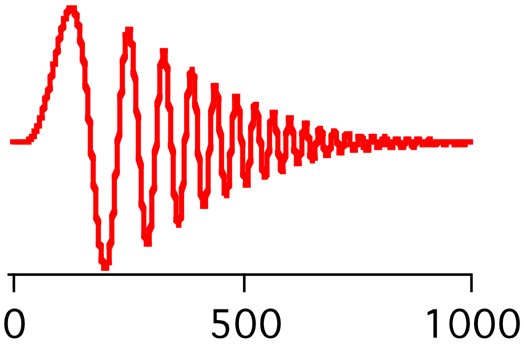
Stimulus



Filter



Response



Convolution theorem

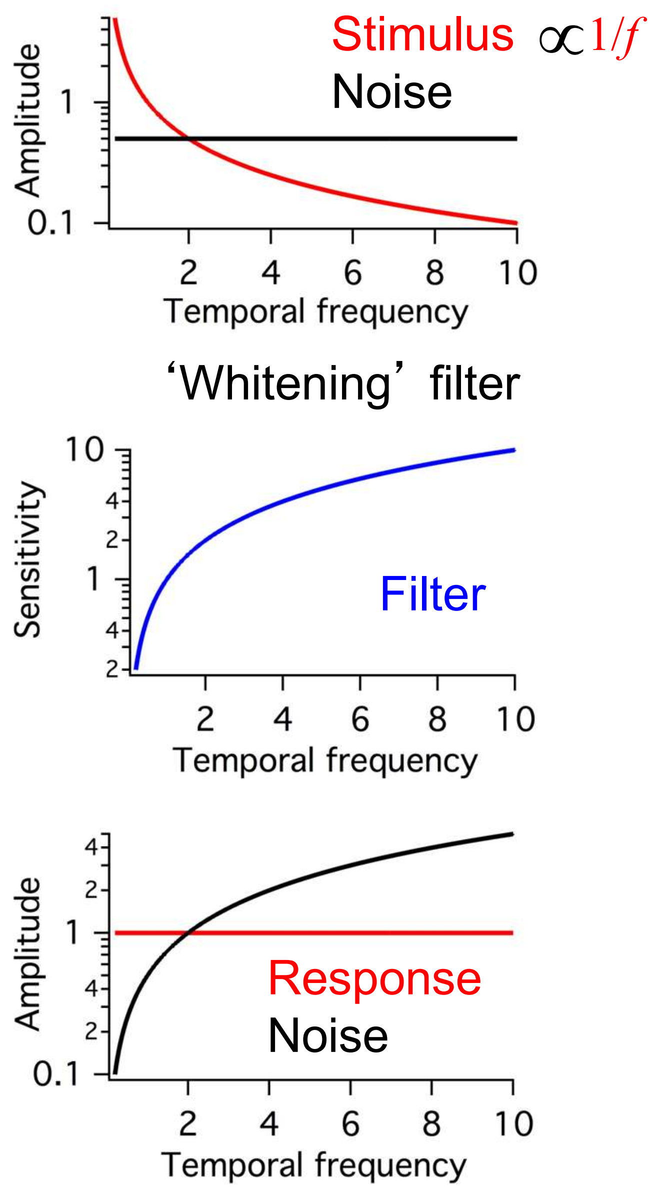
$$h(t) = f(t) * g(t) \quad \Leftrightarrow \quad \tilde{h}(\omega) = \tilde{f}(\omega)\tilde{g}(\omega)$$

a convolution in the time domain

is a simple product in the frequency domain

What filter maximizes information?

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



Adapting to different levels of noise

High SNR – increase total entropy

Filter to whiten in the presence of noise

Low SNR – reduce noise entropy

Low pass filter reduces noise

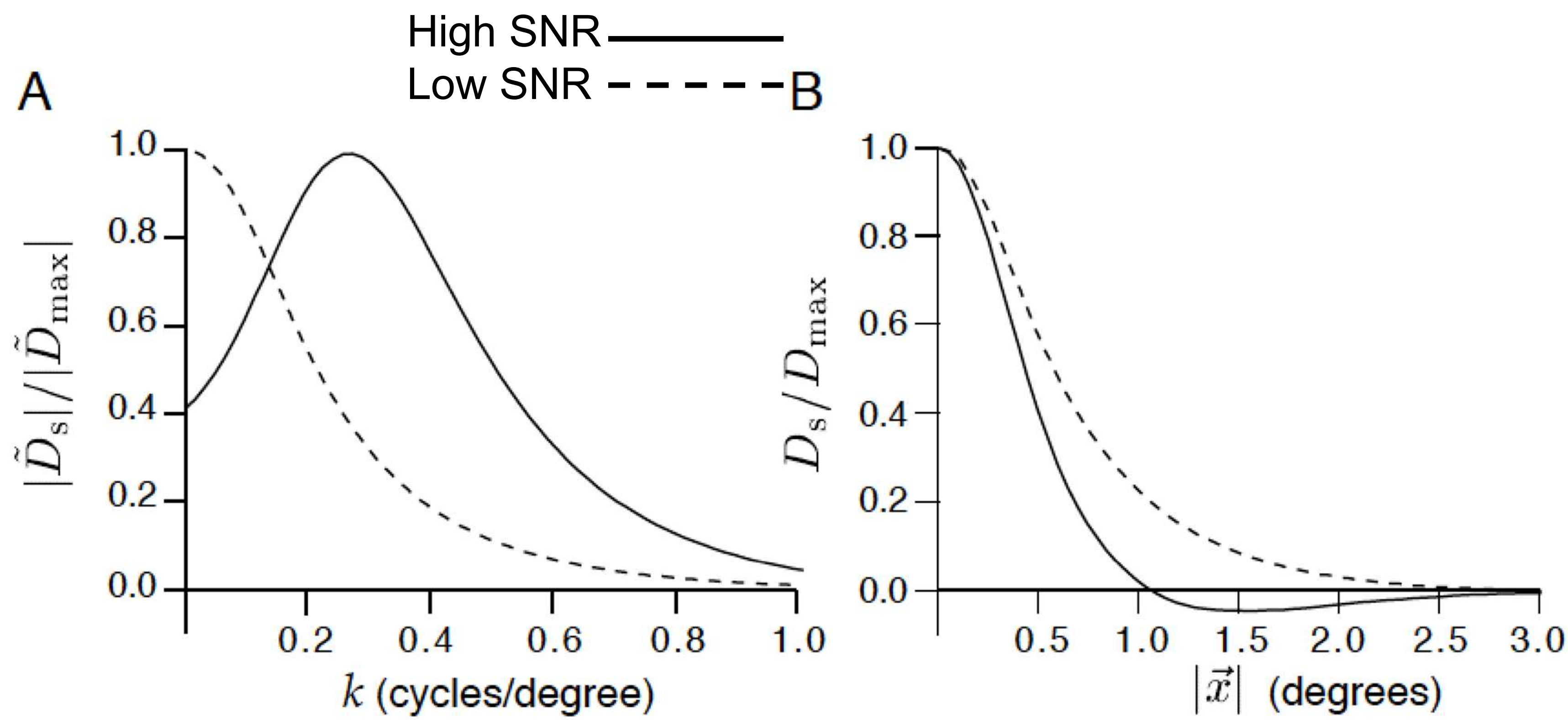
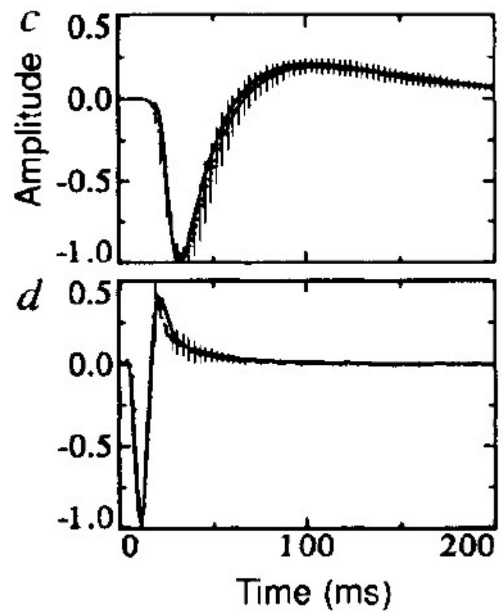


Figure 4.3: Receptive field properties predicted by entropy maximization

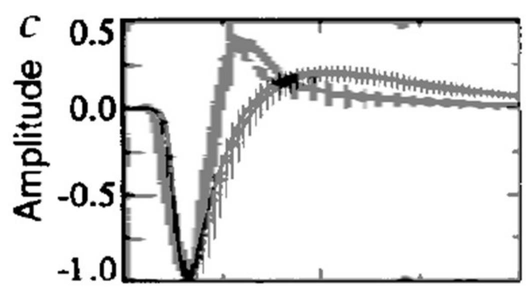
Theory of maximizing information in a noisy neural system

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



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

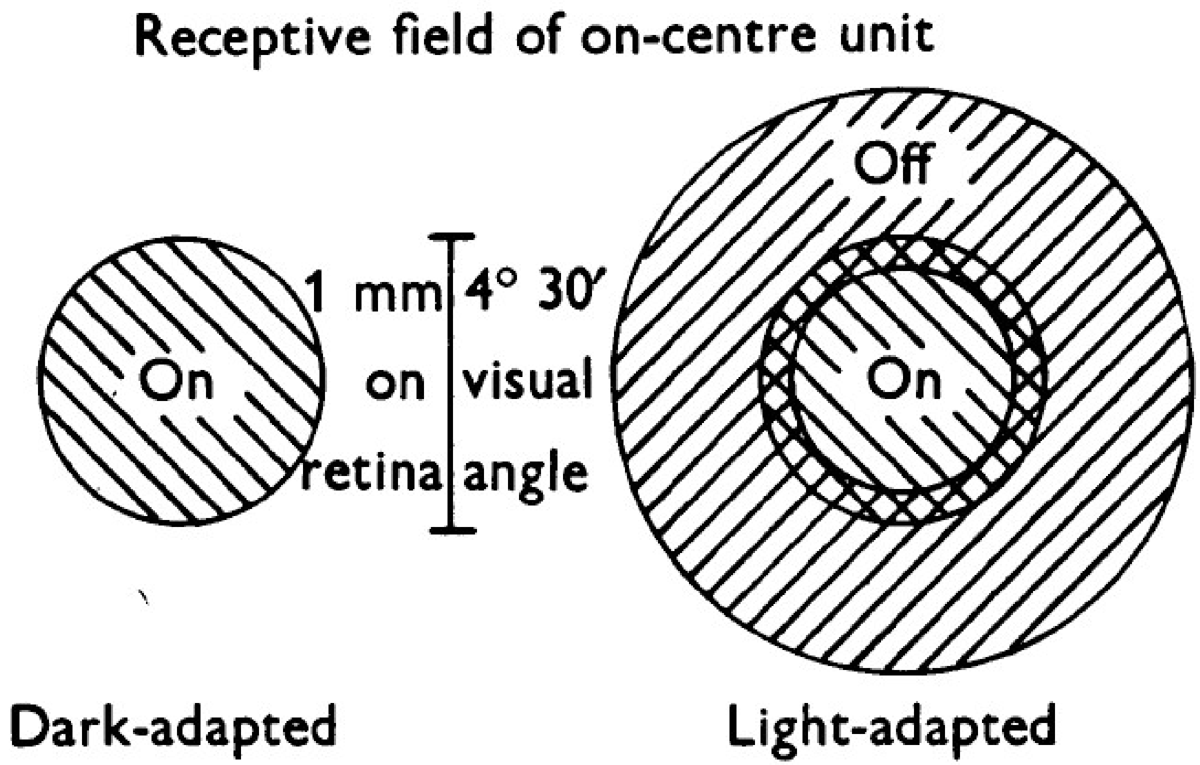
High background intensity
Emphasizes change, is more differentiating
(real and theoretical optimum)



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

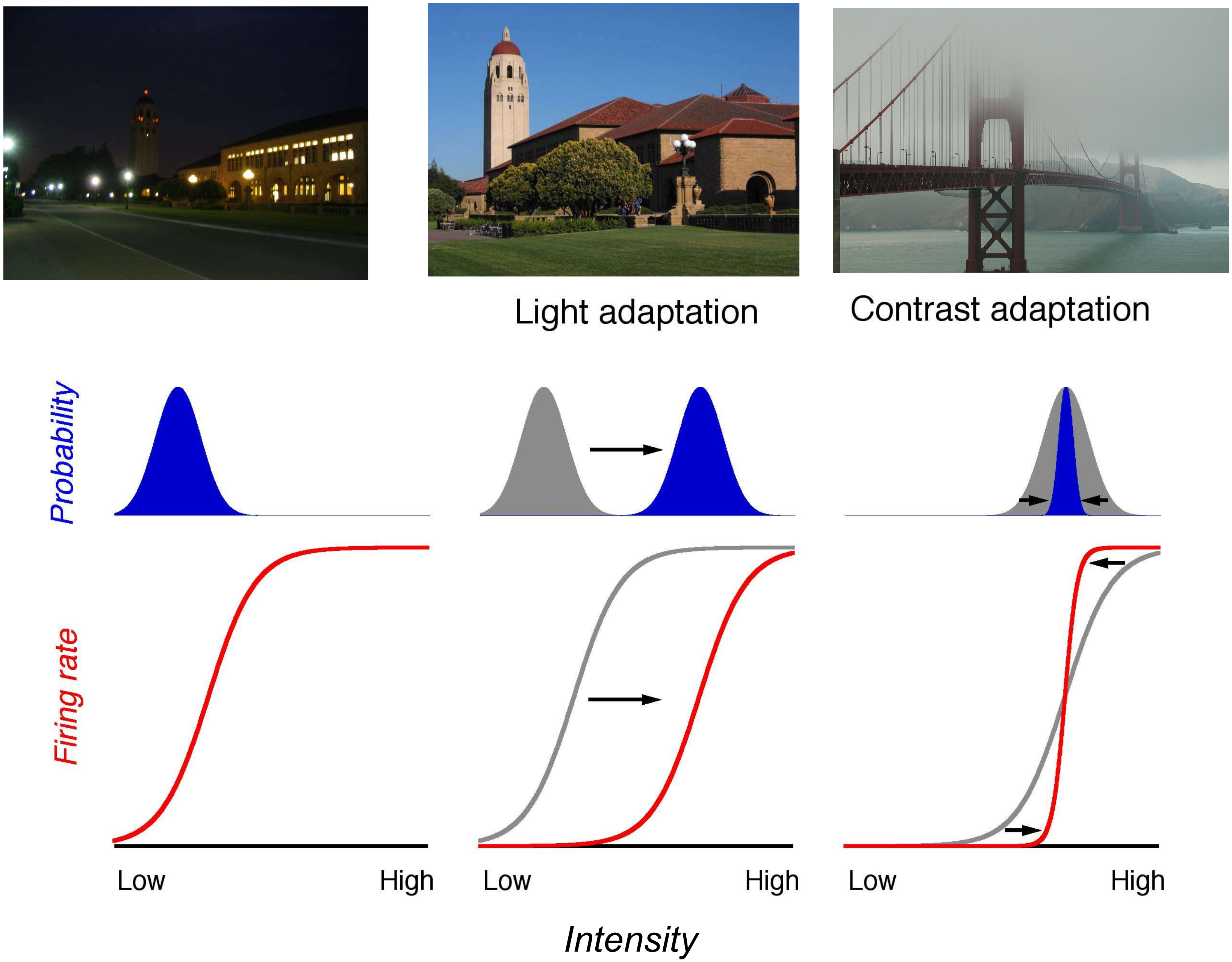
Low frequencies dominate in natural scenes

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**

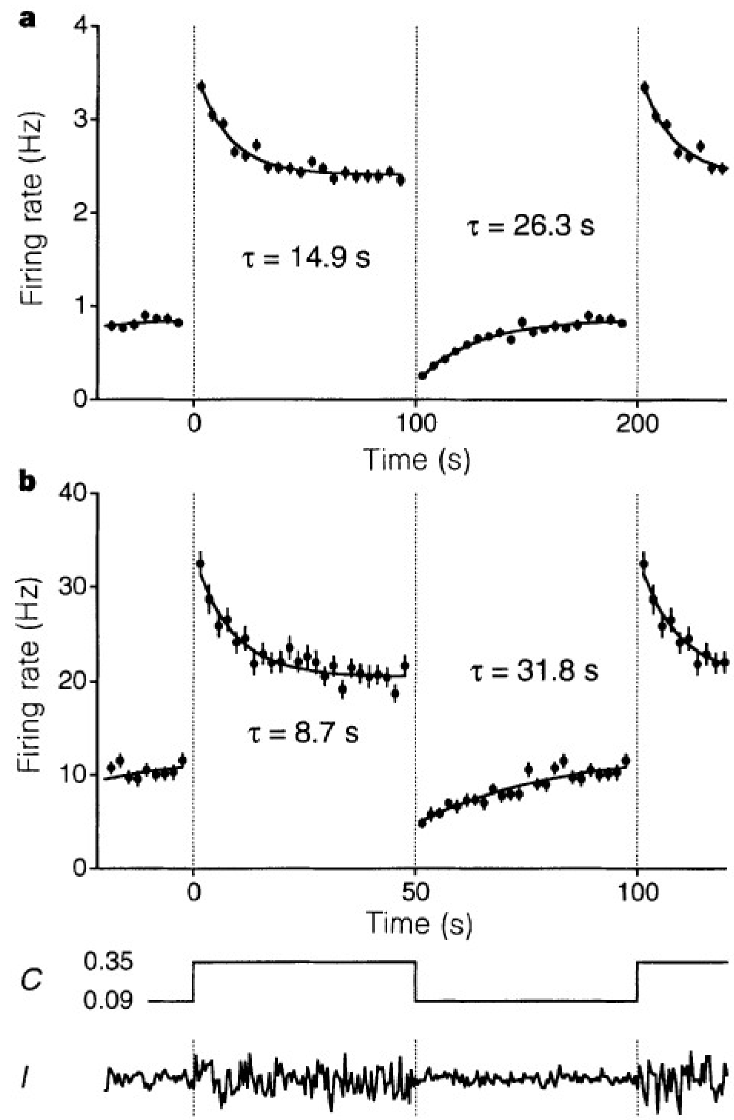
Adaptation to mean and variance



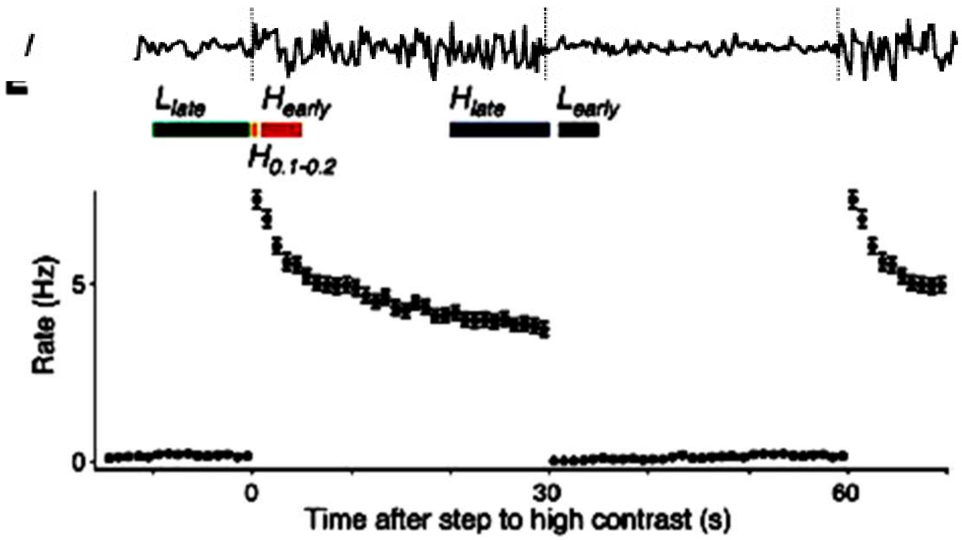
Retinal contrast adaptation

Salamander

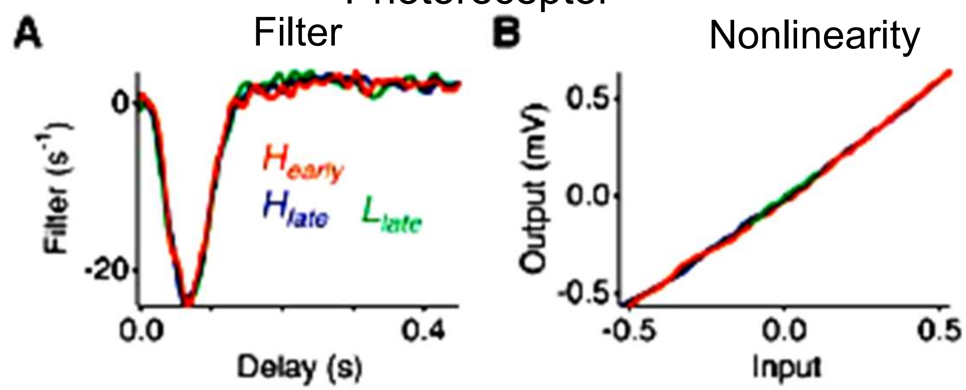
Rabbit



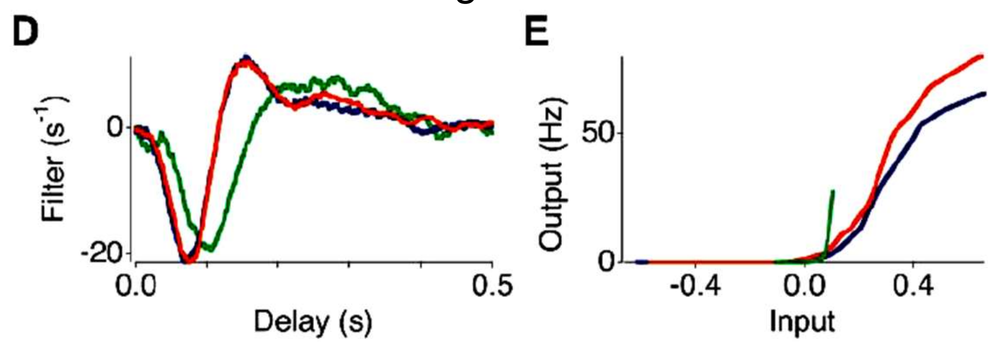
Smirnakis et al., Adaptation of retinal processing to image contrast and spatial scale. *Nature,* 386:69-73 (1997).



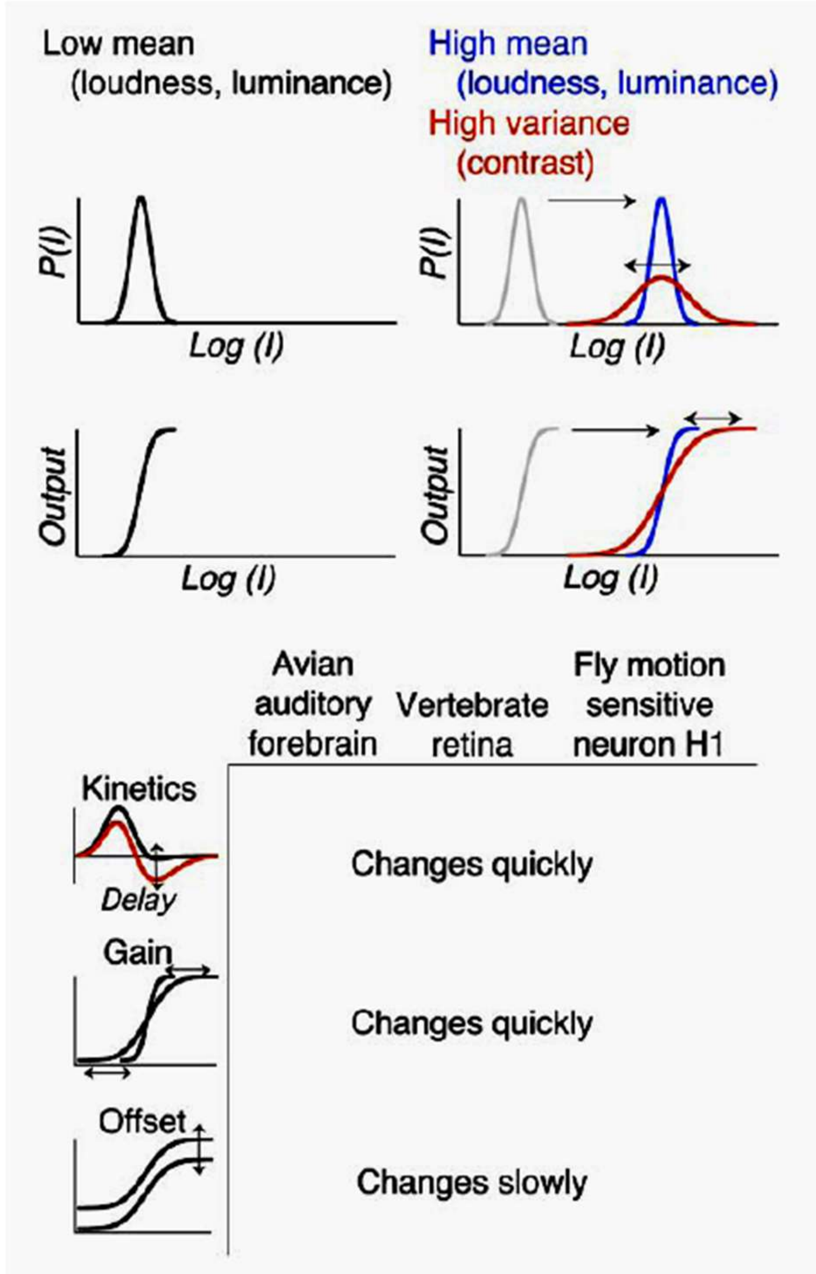
Photoreceptor



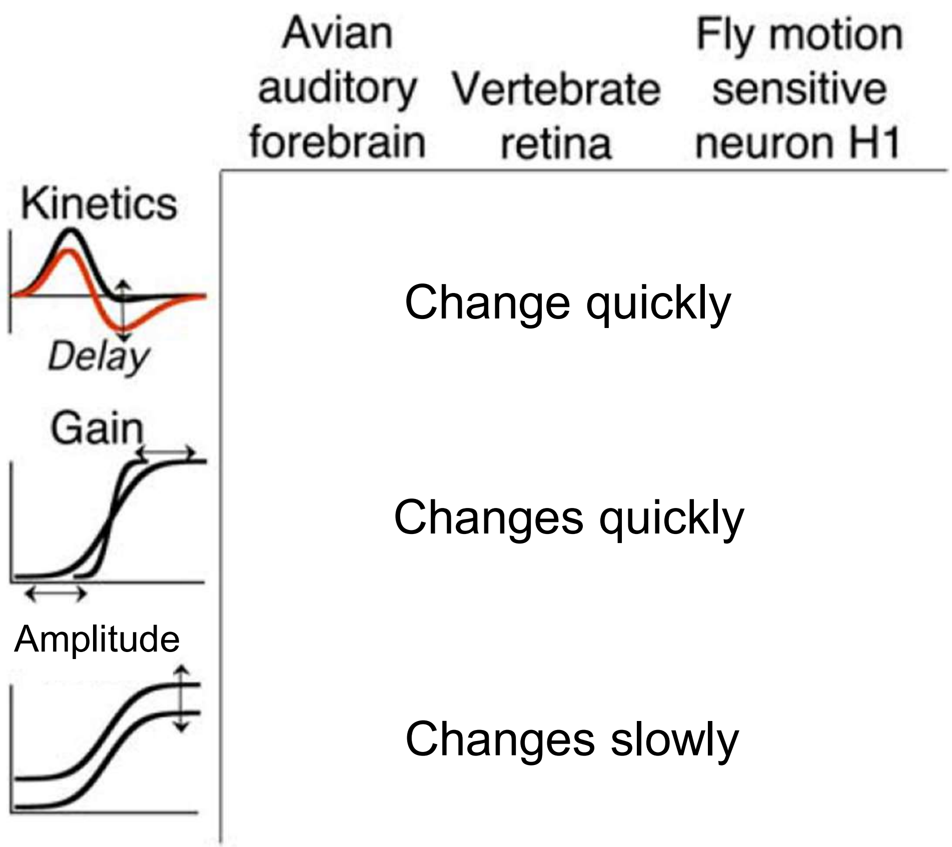
Ganglion cell



Baccus & Meister. Fast and slow contrast adaptation in retinal circuitry. (2002).


Low mean
(loudness, luminance)
High mean
(loudness, luminance)
High variance
(contrast)
P(I)
Log (I)
P(I)
Log (I)
Output
Log (I)
Output
Log (I)
Avian auditory forebrain
Vertebrate retina
Fly motion sensitive neuron H1
Kinetics
Delay
Changes quickly
Gain
Changes quickly
Offset
Changes slowly

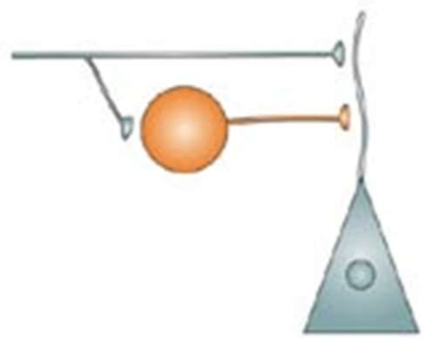
Common properties of contrast adaptation



Nagel & Doupe, 2006
Fairhall et al., 2001

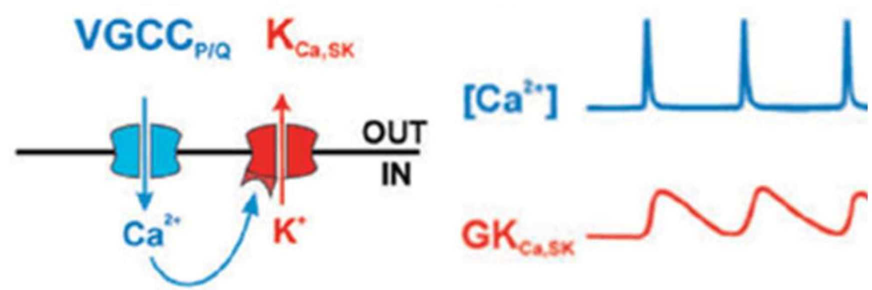
Change in sensitivity by *modulation*

Feedforward inhibition



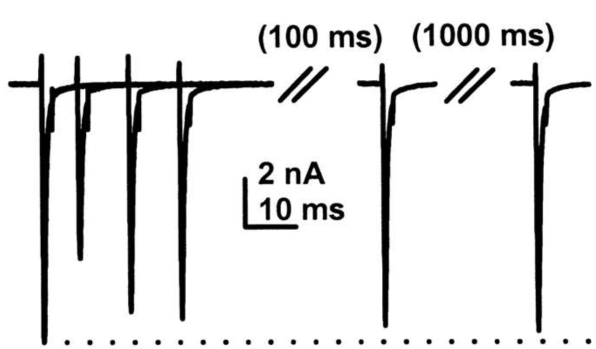
Feedback inhibition

Spike dependent conductances

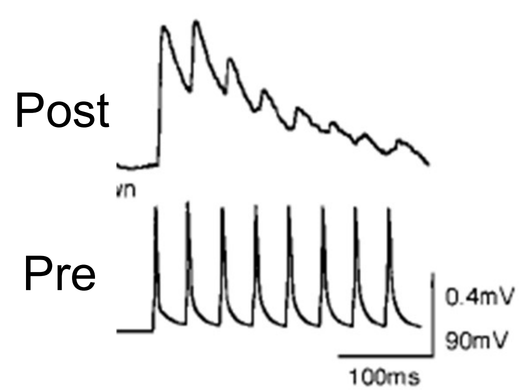


Change in sensitivity by *depletion*

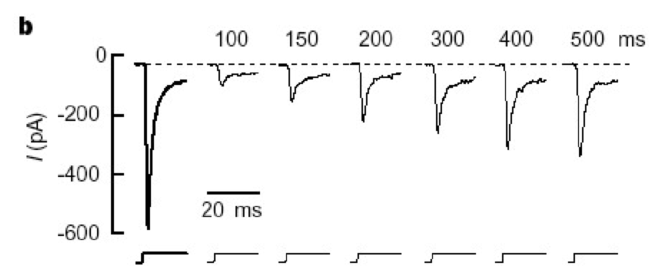
Ion channel inactivation



Short-term synaptic plasticity synaptic depression



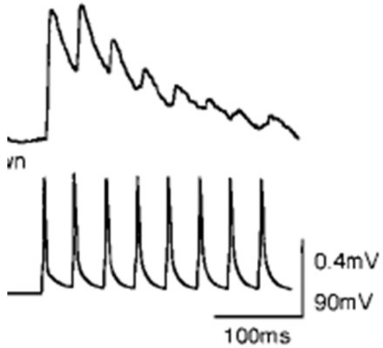
Receptor desensitization

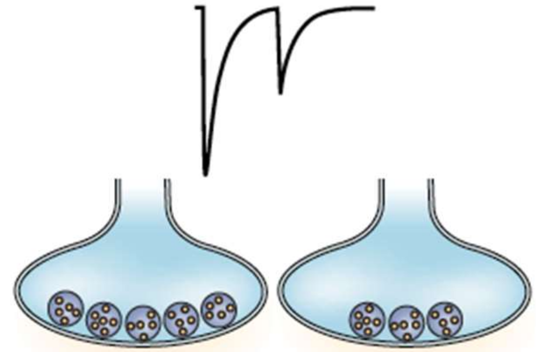


Short-term synaptic plasticity – synaptic depression

Post

Pre





Depletion of available vesicles as a mechanism for depression

n: Number of vesicles

p: Probability of vesicle release

Release = n x p

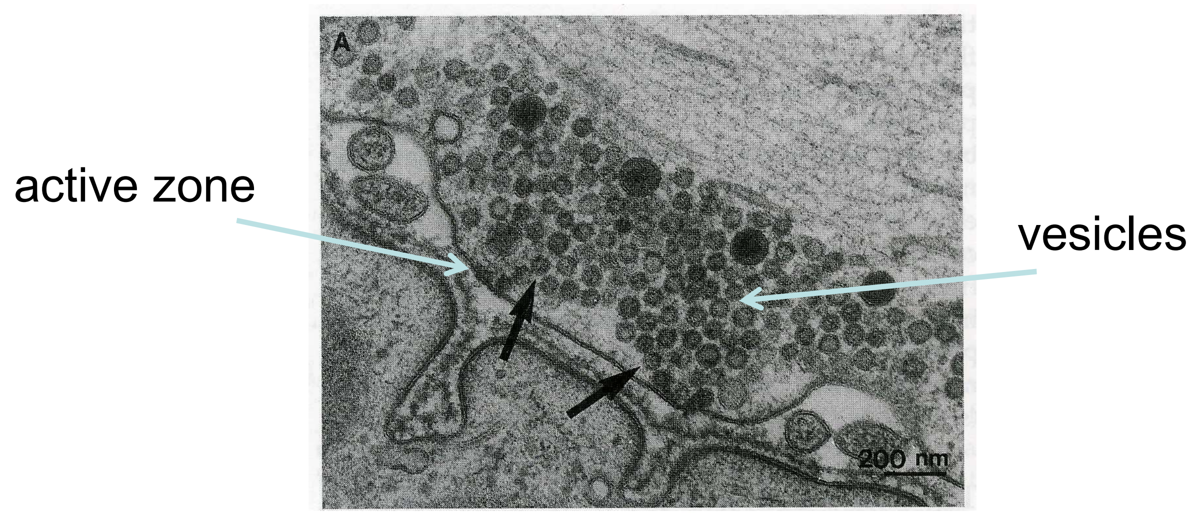
$$\frac{dn(t)}{dt} = \frac{1 - n(t)}{\tau_f} - \sum_j \delta\left(t - t_j\right) \cdot p \cdot n(t)$$

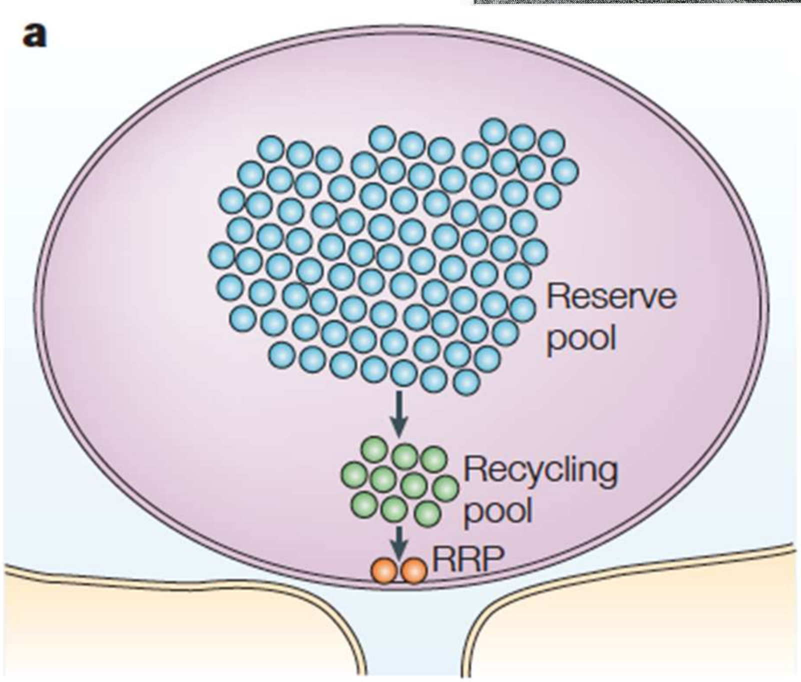
At a given time, t, the rate of change of the number of vesicles is the difference between the current and baseline number of vesicles per time interval t. At the time of each spike, release vesicles equal to the the number of vesicle times the probability of vesicle release.

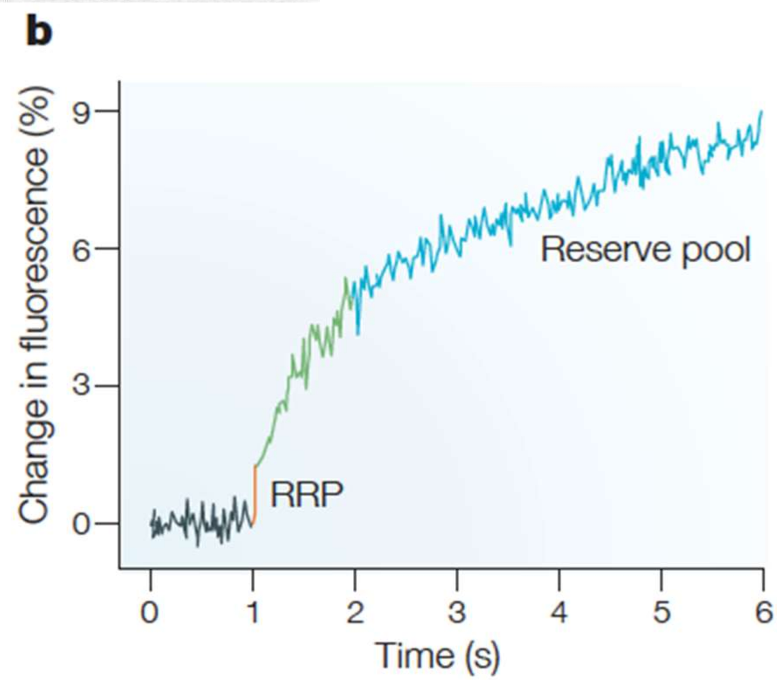
Hennig, 2013. Theoretical models of synaptic short term plasticity

colorized math, betterexplained.com

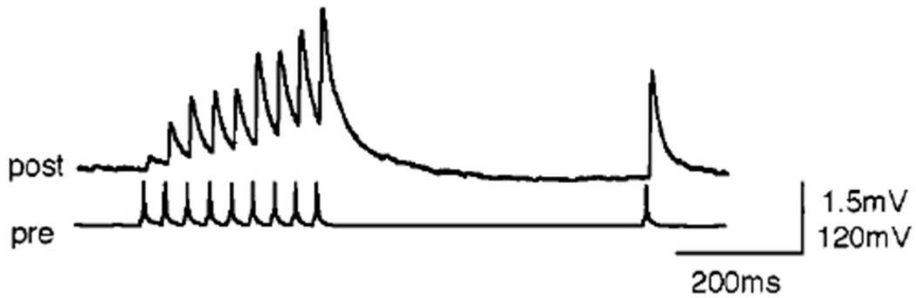
Vesicle release has dynamics over multiple timescales

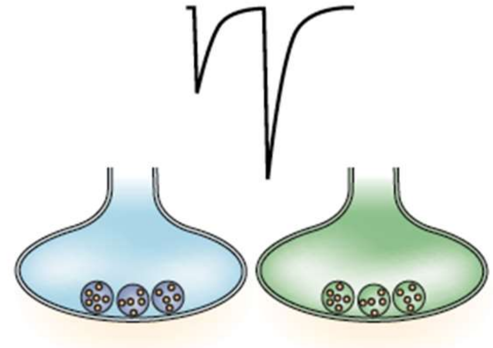






Short-term synaptic plasticity – synaptic facilitation





Residual calcium as a mechanism for increased release

$$\frac{dp(t)}{dt} = \frac{p_0 - p(t)}{\tau_f} + \sum_j \delta(t - t_j) \cdot a_f \cdot (1 - p(t))$$

Release probability exponentially decays to a baseline probability, p_0 with a time constant of τ. At the time of each spike, release probability increases by a facilitating amount that decreases to zero when release probability is one.

Does synaptic facilitation play role in working memory?

Mongillo, Barak & Tsodyks, Synaptic Theory of Working Memory (2008)

Functional advantages of response properties and changes in those properties

Why do cells have a particular nonlinear response function and why does it change?

To use the limited dynamic range of the system efficiently given the distribution of natural inputs.

Why do cells have a certain duration and shape filter?

To maximize or increase information transmission given the natural statistics of the encoded signals.

Why does the filter change?

To increase information transmission when the signal to noise ratio changes.

How can the nonlinearity and filter change?

Adaptation from multiple mechanisms, biochemical cascades, synaptic release, receptors, ion channels and network properties.