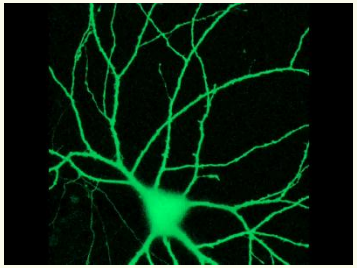
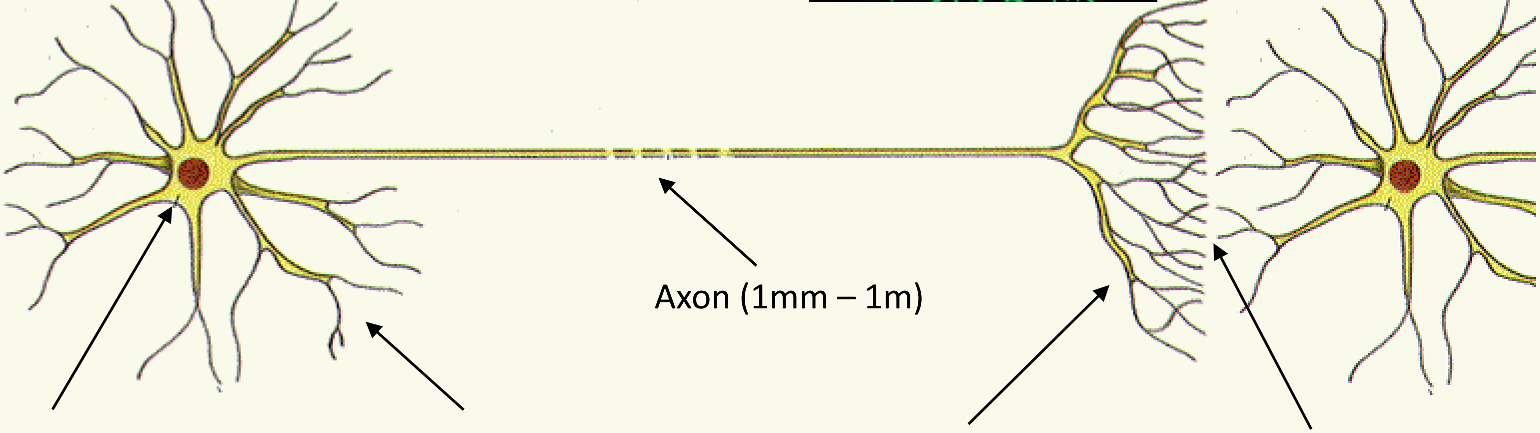
Lecture III

Source coding, channel capacity and neuronal systems

Nerve Cells





Axon (1mm – 1m)

Cell body (soma)

Dendrites

Terminal branches
of the Axon

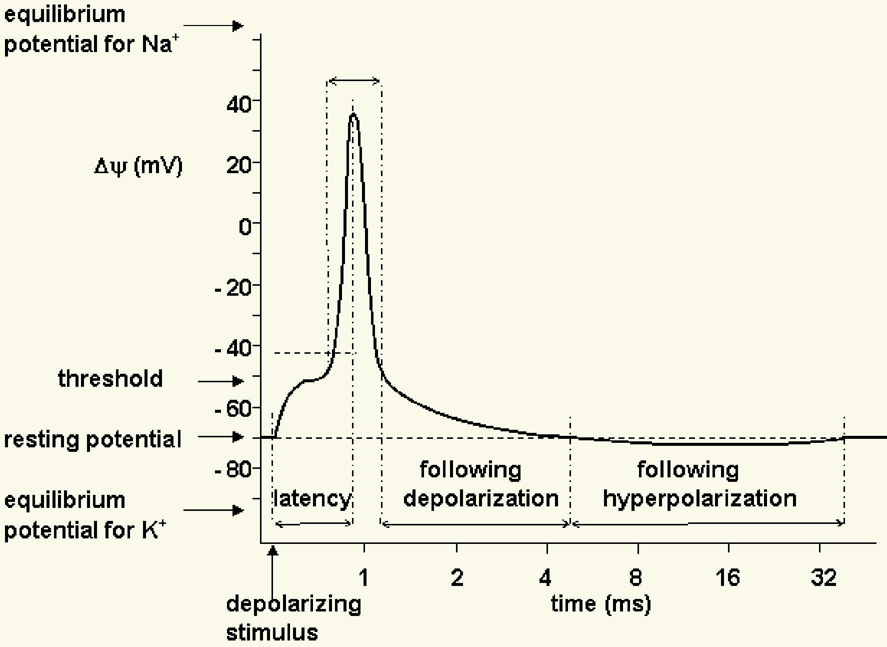
Synaptic connection

Integrate (over inputs)
and fire (action potential\spike)

The action potential of neurons

- **Action potential** = large sharp fluctuation of membrane potential propagating along the axon and which carries information.
- Generation and propagation enabled by ion channels.

- **Threshold** for generating action potential.
- **All-or-none law** ensures full size .
- **frequency coding** of strength and latency of initial stimulus. Amplitude almost independent of stimulus
- Impossible to fire just after previous firing.

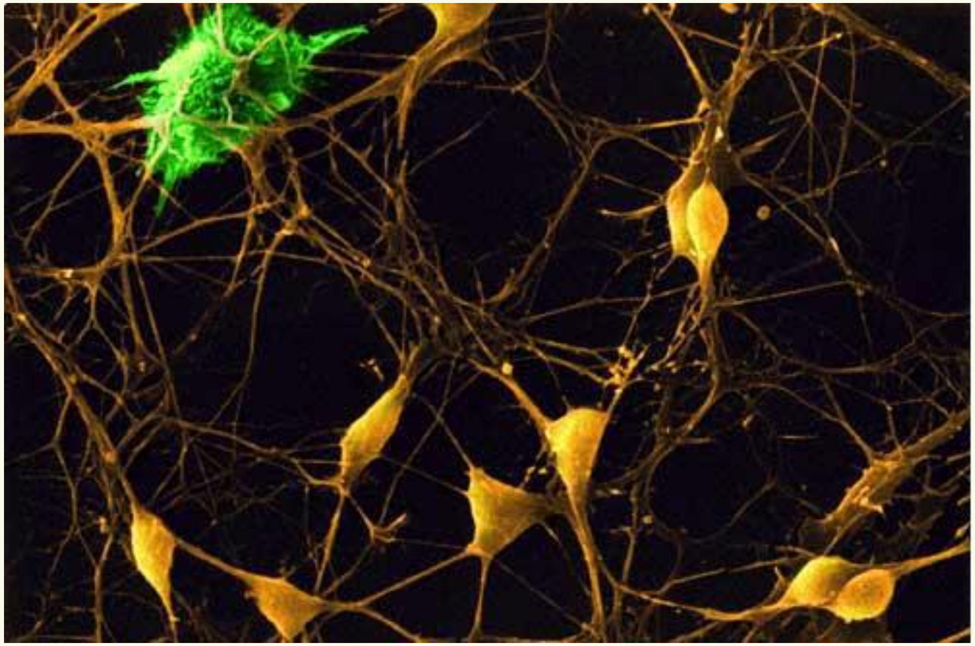


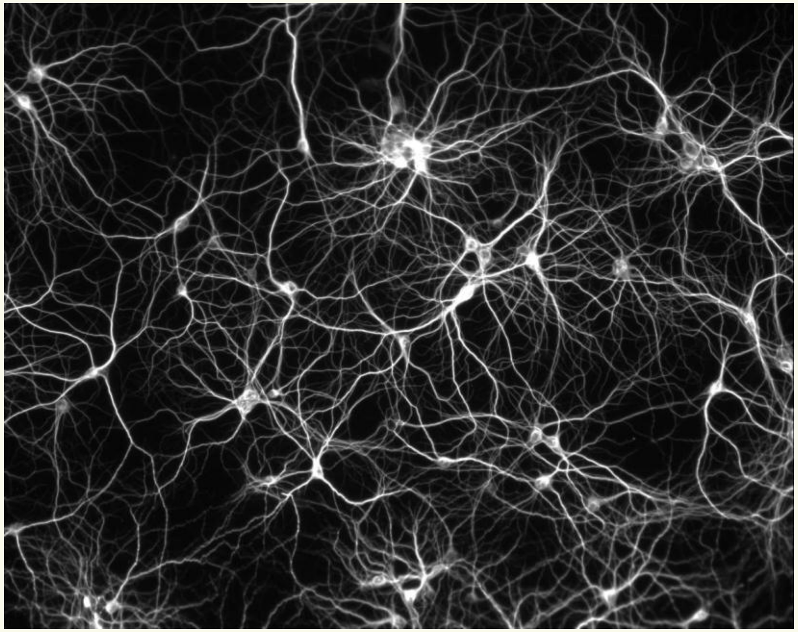
The problem

- Human brain: 10^{11} neurons (~Milky Way), >10^{14} connections.

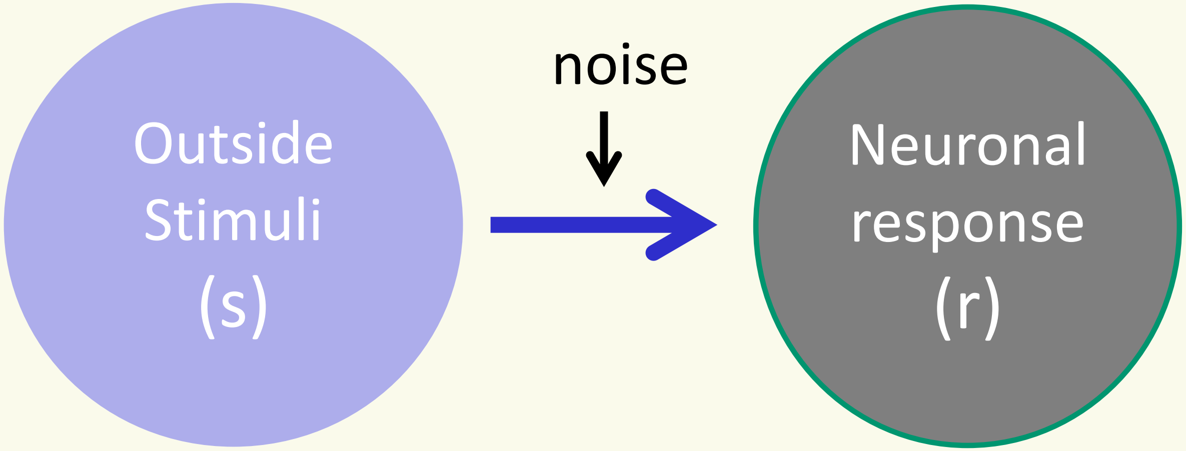
What should we aim to measure? How many neurons to record from?
How should we look at the data?

One possible point of view: The brain is an information gathering and processing device.
To understand its function we should measure not heat, or spike velocities but…
information gathering is measurable and the currency is bits.
(info processing may be viewed as feature extraction also measurable in bits).





The “archetype” neuronal information channel



p(s,r):

From the “researcher’s point of view”:
Given an outside stimuli, what neuronal responses could it elicit? p(r|s)*p(s)

Conversely, taking the “man\animal point of view”:
Given a neuronal response, what can we tell about outside world? p(s|r)*p(r)

maximizing information between source and response

The point of view we take:

We live an environment where p(s) is given

(for example: a monkey raised in an environment with horizontal stripes only will grow to be blind to perpendicular stripes).

- What could be a optimal design for a neuron?
- How would it convey a maximal amount of information regarding this given environment?

We want to choose a function *r = f (s)* – What would be a good choice?

Assumption (for this 1st example): Our channel is *not noisy.*

Histogram flattening

- We would like to maximize the mutual information between the environment and the response of the Neuron:

$$I(r,s) = H(r) - H(r \mid s)$$

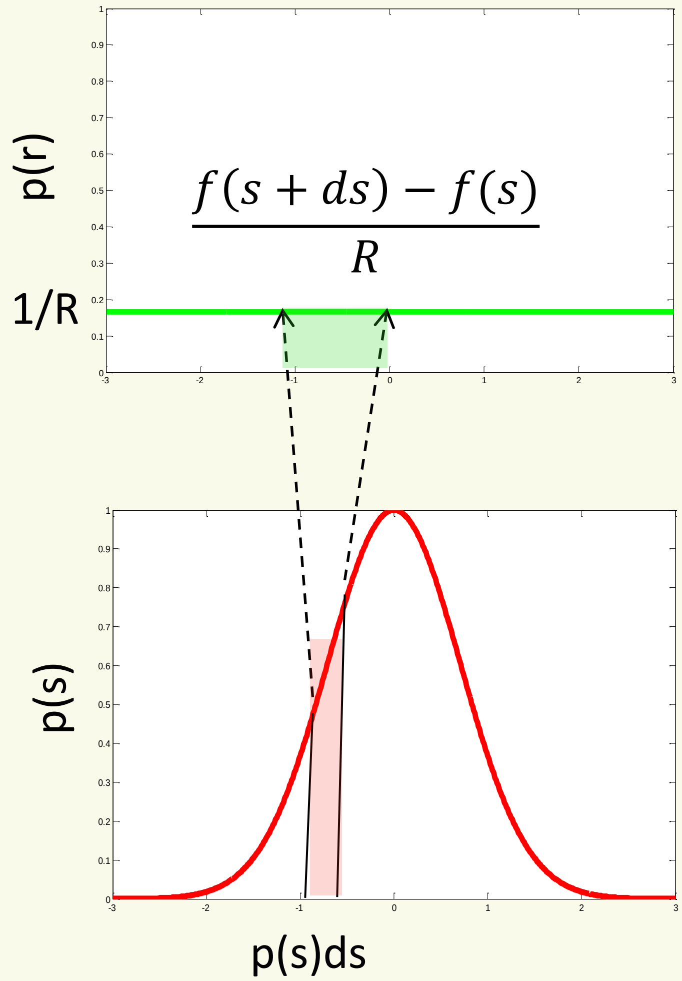
- For a noiseless channel ($H(r|s) = 0$) this is just:

$$I(r,s) = H(r) = -\int_0^R p(r)\log p(r)dr$$

- Which is maximized for constant probability

$$p(r) = \frac{1}{R}.$$

Histogram flattening



$$\frac{f(s+ds)-f(s)}{R}$$

Without loss of generality:
f(s) is continuous
and monotone increasing.

$$dp = p(s)ds = p(r)dr = \frac{1}{R}\frac{df}{ds}ds$$

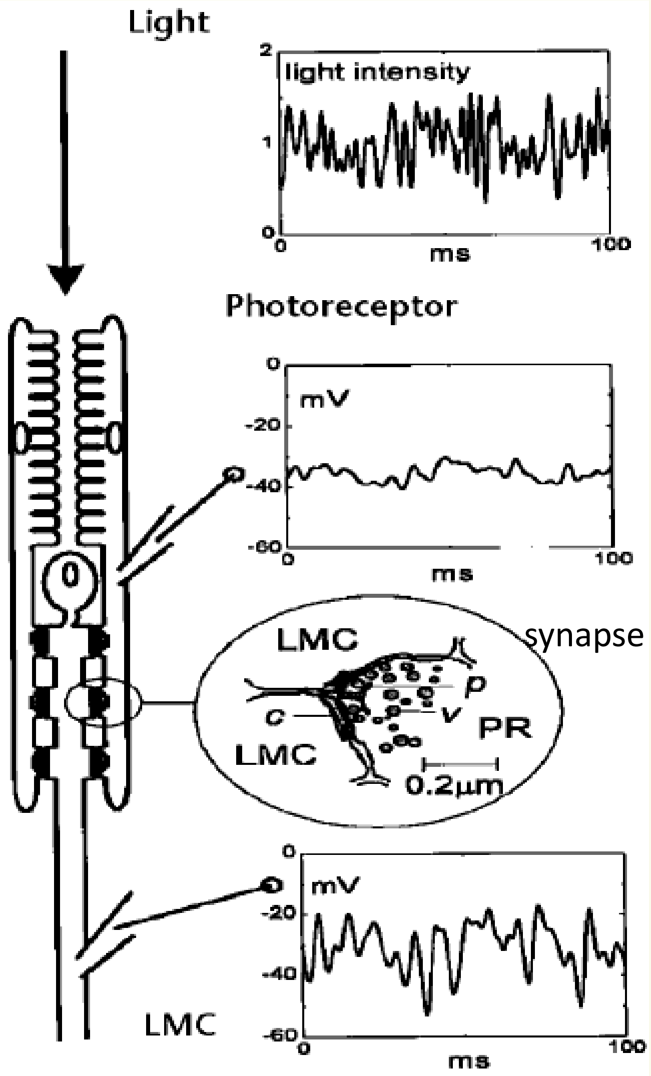
$$\Downarrow$$

$$f(s) = R\int_{-\infty}^{s} p(s')ds'$$

We would choose *f*(*s*) to be the cumulative distribution of *p*(*s*).

EXPERIMENT

Example: Fly LMC – large monopolar cell



- Photoreceptors and an LMC of the blowfly retina code light level in a single pixel of the compound eye.
- Six photoreceptors (two shown) carrying the same signal converge on a single LMC and drive it via multiple parallel synapses.
- The signals are intracellular recordings of the graded changes **(not spikes!)** of membrane potential induced by fast changes in contrast.

Simon B. Laughlin, Rob R. de Ruyter van Steveninck & John C. Anderson, 1998

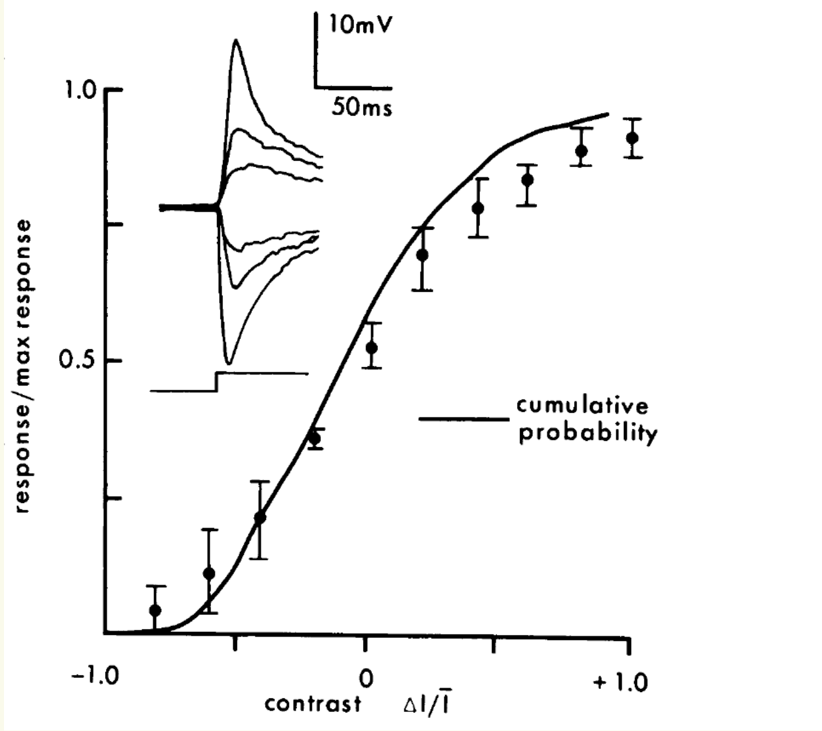
EXPERIMENT

Example: Fly LMC

- Simon Laughlin (1981) measured the cumulative contrast distribution in the flies' natural environment (lakeside vegetation): *p(s)*

- He then measured the response of the cell to contrast steps (inset).

- And averaged over trials (data points): this is *f(s)*.

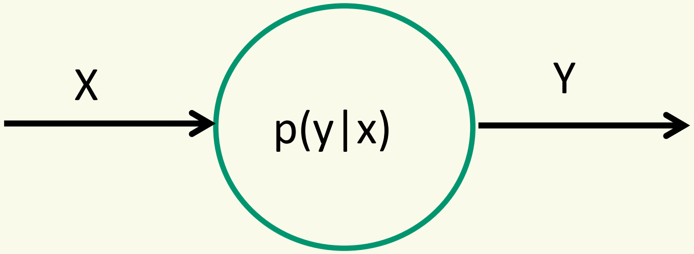


He found that:

f(s) is very close to the cumulative distribution associated with p(s)!

Channel Capacity

Discrete memory-less channel



Definition 1:

$$C = \max_{p(x)} I(X,Y)$$

Definition 2:

Maximum number of bits transferred per use of channel

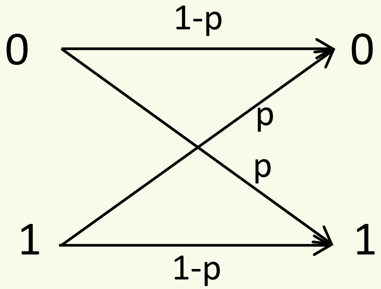
Properties of capacity

- $C \geq 0$ *(reasonable, if we think of capacity in the water-pipe analog)*
- $C \leq H(X), H(Y) \leq \log|X|, \log|Y|$

(channel capacity is limited by size of alphabet it can handle)

Shannon proved equivalence of these two definitions (sketched below).

Example: Binary symmetric Channel



$$
\begin{aligned}
I(X,Y) &= H(Y) - H(Y|X) \\
&= H(Y) - \sum p(x) H(Y|X=x) \\
&= H(Y) - \sum p(x) H(p) \\
&= H(Y) - H(p) \\
&\leq 1 - H(p)
\end{aligned}
$$

$$
\begin{aligned}
H(Y|X=x) &= -p \log p - (1-p) \log(1-p) \\
&= H(p)
\end{aligned}
$$

Equality is obtained by choosing p(x)=(½ , ½) which makes p(y)=(½ , ½) as well.

Hence:

C = 1-H(p) (Note, *p=½* gives *C=0*)

Simple example: noisy typewriter

The channel:

A->{A,B} ; B->{B,C} ….. Z->{Z,A} (all errors are probability ½).

$$C = \max\ I(X,Y) = \max H(Y)-H(Y|X)$$

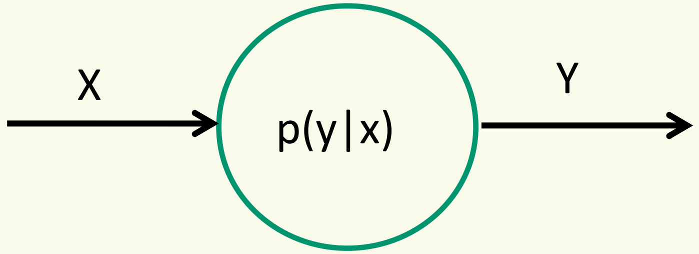
$$= \log 26 - \log 2 = \log 13$$

On the other hand if we use a code of block length n=1 consisting of every other letter we get an error free channel with 13 possible messages.

Choosing the messages equi-probably

We can achieve transmission of log 13 per transmission.

Channel coding theorem (intuitive explanation)



A communication channel is defined by *p(y|x)*.

We can choose *p(x)* as we wish to try and achieve optimal information transfer.

Once we choose *p(x)* this also sets *p(y)=p(y|x)*p(x)*

Channel coding theorem
(intuitive explanation cont.)

We use code words of length n: $\vec{X}_i \quad i = 1 \dots k.$
Each transmission uses the channel n times.
How much information can we pass per use?

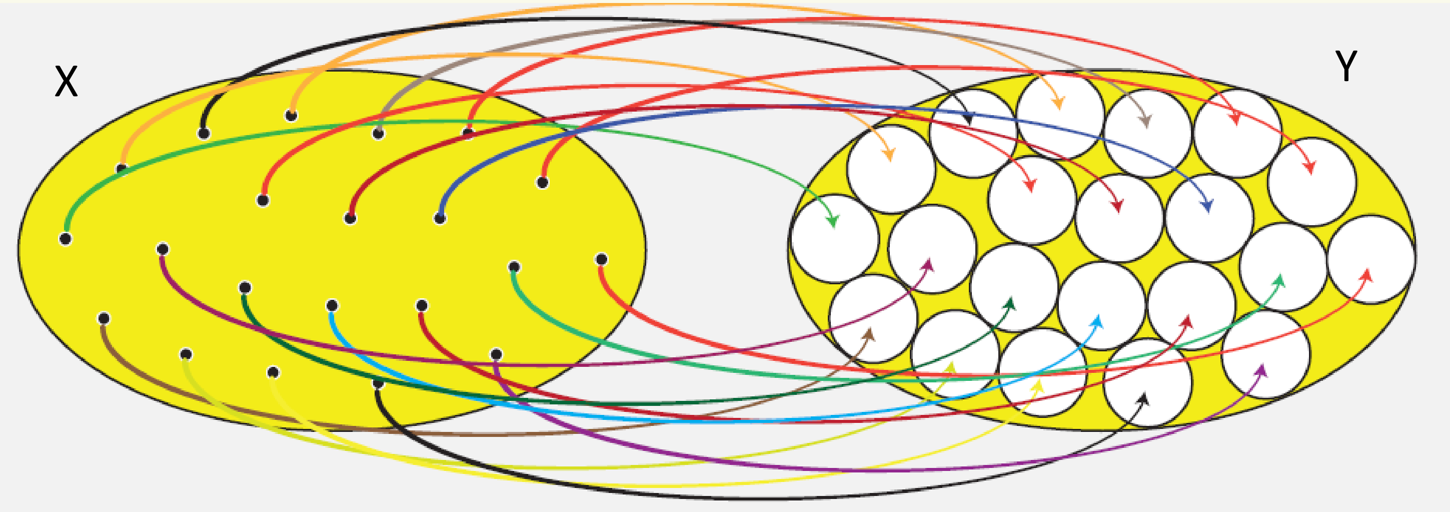
Due to the noise in the channel
each time $\vec{X}_i$ is transmitted a different message will be received.
Call these groups of messages $\{\vec{Y}_j\}_i$

The idea is:
If we choose good $\vec{X}_i$'s + take n to be very large all the $\{\vec{Y}_j\}_i$ i=1…k will be disjoint.
This means that all messages $\vec{X}_i$ can be passed reliably (no channel noise).
The maximal number of bits we can pass using **n** transmissions is ***log (k)***
Per transmission the rate is ***R = log(k)/n*** *bits/channel use*.

(achieved by using
codewords equiprobably).

The question becomes:
How many disjoint sets $\{\vec{Y}_j\}_i$ can we fit in the space Y^n?

What is the maximal number of disjoint output messages ?



Remember, for a particular choice of p(x), p(y) is completely determined:
The total number of typical vectors in Y^n is: $2^{nH(Y)}$

(the probability to get one of these vectors after a transmission is ~1)

Using the same argument, the size of each set of vectors $\{Y_j\}_i$: $2^{nH(Y|X=\overrightarrow{Xi})}$

So the average size of these output sets is just : $2^{nH(Y|X=Xi)}$

We divide these total available “area” by mean message area.
to see how many disjoint sets fit in the entire space: $k = 2^{n(H(Y)-H(Y|X))} = 2^{nI(X,Y)}$

Channel coding theorem
(intuitive explanation cont.)

Therefore given p(x) we can hope to find (upper bound) k messages, $\vec{X}_i$, with disjoint outputs.

$$k = 2^{n(H(Y)-H(Y|X))} = 2^{nI(X,Y)}$$

Disjoint outputs means we pass these messages without error to achieve a rate of:

$$R = 1/n * \log k = 1/n * \log 2^{nI(X,Y)} = I(X,Y) \quad \text{bits/channel use}$$

To get the most bits across we choose p(x) that maximizes I(X,Y), but this is the original definition of capacity:

$$R = \max_{p(x)} I(X,Y) = C \qquad \square$$

If we want to transmit messages at a rate $R \leq C$ we can do this with arbitrarily small errors!

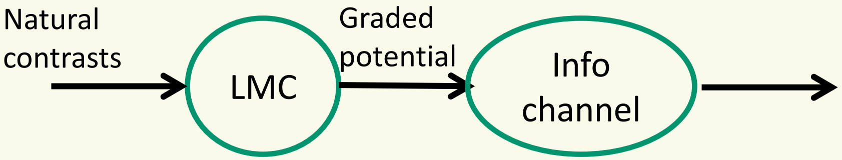
Back to neuroscience

A small comment for intuition:

We saw that for some information channels

capacity is reached by using the input alphabet equi-probably.

Going back to the fly LMC:



The graded potential is:

1. in volts (a message that can be transmitted between neurons).
2. idealy, no info loss as the message passes the LMC.
3. distributed equiprobably.

1,2 ➔ We can think of the LMC as a source encoder.

3 ➔ it may be a good encoder since
it may help to achieve high information rates in the downstream communication channel.

EXPERIMENT

Capacity of neuronal link

D.M. MacKay and W.S. McCulloh (1952)

Looked at spiking neurons and wanted to compare estimates for channel capacity between two different scenarios:

1. Information is encoded by spike sequences {1000100110…}

2. Information is encoded by exact time until next spike. $\{\tau_1, \tau_2, \tau_2, \ldots\}$

EXPERIMENT

1. Spike sequences

$$C = \max_{p(x)} [H(X)-H(X|Y)] \leq \max_{p(x)} H(X)$$

What is X depends on the model.
Here it all possible spike sequences.

After each spike there is a refractory period of T_R.

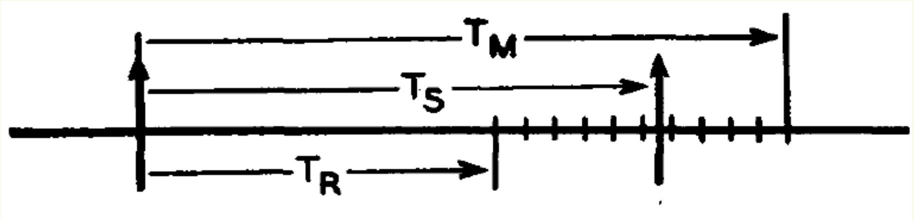
To obtain their estimate they divide time into bins of duration T_R , each bin can have at most one spike.

The maximal rate of information is when the probability for a spike is ½.

The number of bits per msec is then *1/ T_R bits/msec.*

For spike frequencies around 250Hz. *T_R=4 msec* and the bit rate is *C~1/4 bit/msec*.

2. Spike timings



The time until next spike, T_S is somewhere between the refractory time T_R and some maximal time T_M (to be determined below). Neurons can measure spikes to within a window of dT.

➔ Number of possible messages $\frac{T_M - TR}{dT}$. They estimate *dT* at 0.05msec.

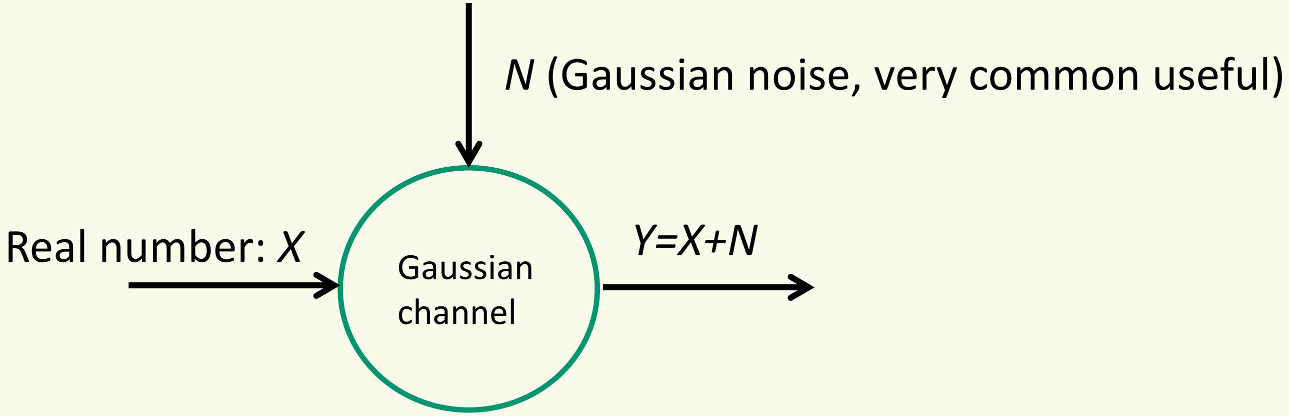
Average number of messages per msec: $\frac{2}{T_M + T_R}$.

So that the bit rate is: $\frac{2}{T_M + TR} * \log \frac{T_M - TR}{dT}$ where p(x) is the uniform probability over all possible timings

For T_R*=4 msec* we get a maximal value for T_M*=6.7 msec.*
The rate of information is now: C ~ *1.1 bit/msec*

(about 4 times higher than previous coding procedure).

And back to theory…
Channel capacity of the Gaussian information channel



Start by calculating the mutual information between input and output.

$I(X,Y) = H(Y)+H(X)-H(Y,X)$ definition of mutual information

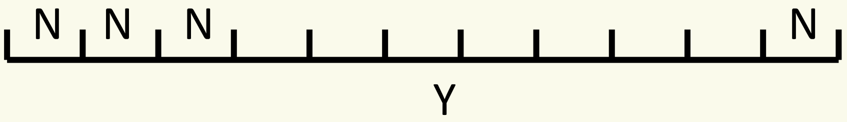
$=H(Y)+H(X)-H(X,X+N)$ Gaussian channel definition

$= H(Y)+H(X)-H(X,N)$ $p(X,X+N)=p(X,N)$

$= H(Y)+H(X)-H(X)-H(N)$ X and N are independent

$= H(Y)-H(N)$

The answer is intuitive:
Not all output states are distinguishable , this is due to the noise.



And back to theory…
Channel capacity of the Gaussian information channel

We can now calculate the channel capacity, of course if we choose very far apart X values the noise can be made negligible.
But what if we have limited average power at the input how much can we send?

Mean power=E(X^2) =const,
This is the same as limiting the input variance since: var(X)=E(X^2)-E(X)^2

And this is equivalent to limiting output variance since var(Y)=var(X)+var(N). So….

$$C=\max I(X,Y) = \max (H(Y))-H(N)$$

The maximum is taken only over p(Y) with set variance.
As we have seen the answer is that p(Y) is a Gaussian
since noise is Gaussian this happens when X is also Gaussian.

The entropy of a Gaussian is $H(X) = 1/2\log(2\pi e\,(var(X)))$

$$\boxed{\begin{array}{c} C\ = 1/2\log 2\pi e\,(var(X)+var(N))-1/2\log 2\pi e\,(var(N)) \\ = 1/2\log\,(1+SNR) \end{array}}$$

The higher the SNR the more info that can be passed.

Time dependent signals

Follow “The rate of info transfer at graded synapses” Van Steveninck, Laughlin 1996.

We aim to take another look at the LMC neuron, when the input is a continuous signal. We describe f(t) in Fourier space.

f(t) = $\sum_{n=-\infty}^{\infty} f_n \exp(-iw_n t)$

Since f(t) is real f_n=$(f_{-n})^*$

IN previous cases our channel inputs were random numbers *(i)* taken from some distribution *prob(i)*.

In the present case, we want to do the same thing for functions. We want to define a distribution: *p(f(t)).*

Gaussians are prevalent in nature we first focus on Gaussian random functions.

Gaussian random time dependent functions.

$$f(t) = \sum_{n=-\infty}^{\infty} f_n \exp(-i\omega_n t)$$

The f_n coefficients are chose from Normal distributions. Such that:

$$\begin{cases} \langle f_n f_{-m} \rangle = 0 & n \neq m \\ \langle f_n f_{-n} \rangle = \sigma^2(\omega_n) & \end{cases}$$

The variance of f can now be easily computed:

$\langle f(t)^2 \rangle = \sum < fnf\ \_\ _n > \exp(-i(\omega_n + \omega\ \_\ _n)t)$

$= \sum \sigma^2(\omega_n)$

By construction,
each frequency is associated with its own variance.
➔ We can look at each frequency as carrying a separate piece of information through a separate channel with its own specified noise.

How much information each independent variable

We have moved back to what we know (scalar channels):

Transmitting the function *f(t)* is equivalent to transmitting the list of independent, random Gaussian variables f_n!

If we assume Gaussian random noise that is frequency dependent: N(ω_n).
So each channel passes:

$$\text{I }(\omega_n)=1/2 \text{ log } (1 + \frac{\sigma^2(\omega_n)}{N^2(\omega_n)})$$

Total mutual information ➔ $\text{I} = \sum I(\omega_n) = \frac{1}{2}\sum \log(1 + \frac{\sigma^2(\omega_n)}{N^2(\omega_n)})$

Taking the continuous limit

$<f(t)^2> = \sum \sigma^2(\omega_n)$

$= \sum \frac{\omega_{n+1}-\omega_n}{\omega_{n+1}-\omega_n} \sigma^2(\omega_n) = \sum \frac{\Delta\omega}{2\pi} T\sigma^2(\omega_n) \overset{T->\infty}{->} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} T\,\sigma^2(\omega)$

$= \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} S(\omega)$, $S(\omega) = \lim_{T->\infty} T\,\sigma^2(\omega)$

Power spectrum
units: [variance/Hz]

Total information per period T: $I \to \frac{T}{2} \int \frac{dw}{2\pi} \log(1 + \frac{S^2(\omega)}{N^2(\omega)})$ bits

Rate of information transfer: $R = \frac{1}{2} \int \frac{dw}{2\pi} \log(1 + \frac{S^2(\omega)}{N^2(\omega)})$ bits/second

EXPERIMENT

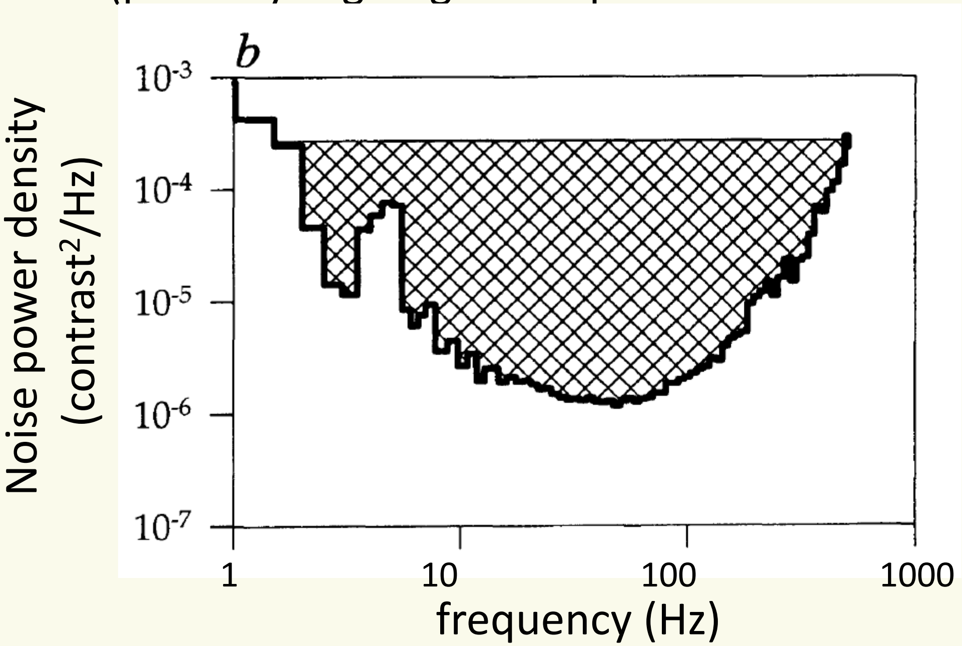
How to maximize information transfer using f(t)

In the discrete Gaussian channel we reach channel capacity by choosing the inputs from a Gaussian distribution. What should we do now?

Q: Given limited power, how should we distribute it among the different frequencies?

Remember we have frequency dependent Gaussian noise N(ω).

A: (proof by Lagrange multipliers - but not now):

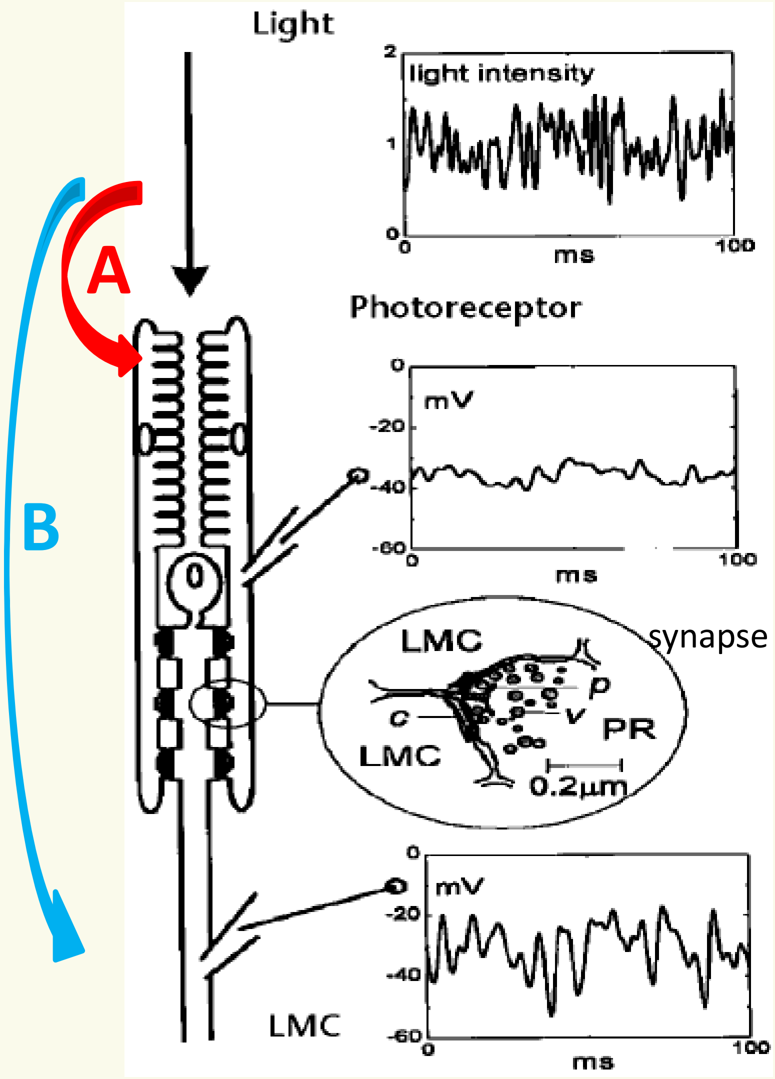


Noise whitening or water filling.

Similar to our first example on LMC:
To transfer much information
we want to make the signal (+noise)
look as random (surprising) as possible!

EXPERIMENT

Channel capacity of fly neurons for continuous signals



Question:
What is the channel capacity of channels A and B?

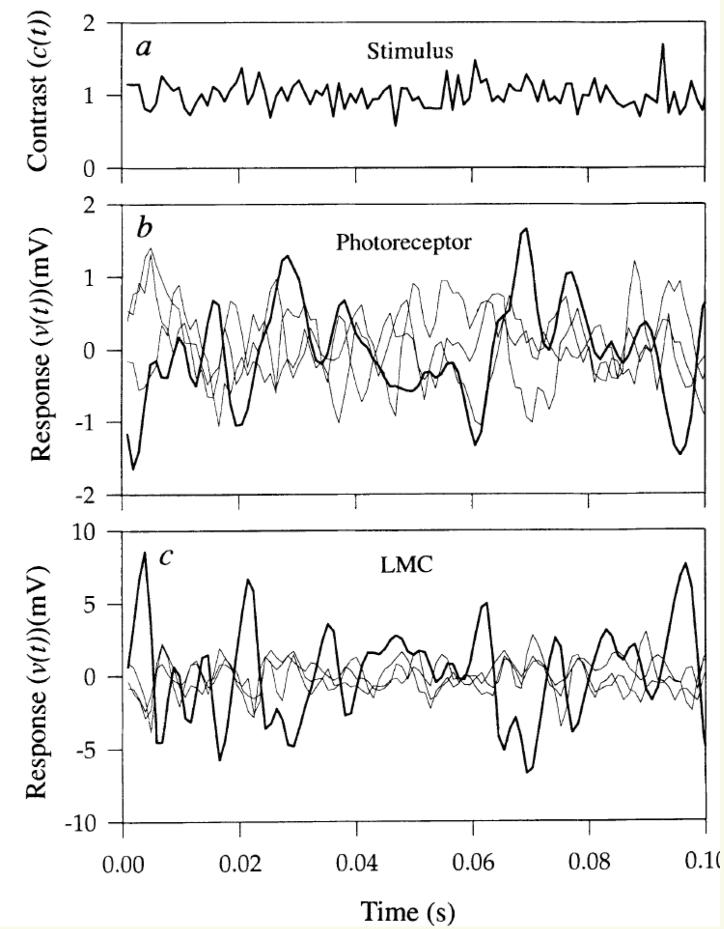
Answer:
In the next we follow
Ruyter van Steveninck and Laughlin (1996)

Simon B. Laughlin, Rob R. de Ruyter van Steveninck & John C. Anderson, 1998

EXPERIMENT

Channel capacity of fly photoreceptor and LMC

Step 1: measure the noise:



- Present stimulus c(t) multiple times.

- For each time measure cells' response v(t).
- Calculate < v(t)> (mean over trials).
- For each trial the noise is defined by:

  n(t)= v(t) - < v(t)>

To calculate capacity we need,
SNR(ω)=S(ω)/N(ω) so we should:

- Fourier transform signals and noises
- Make sure S and N have the same units

  (they don't now!)

EXPERIMENT

Channel capacity of fly photoreceptor and LMC

Step 2: calculate the signal to noise ratio:

To calculate capacity we need,
SNR(ω)=S(ω)/N(ω) so we should:

- Fourier transform signals and noises
- Make sure S and N have the same units (they don't now!)

$$f(\omega)=\frac{1}{\sqrt{2\pi}}\int f(t)\exp(-i\omega t)dt$$

$$S(\omega) = \frac{f(\omega)^2}{2\pi}$$

We have multiple measurements of the noise n(t):

- Fourier transform each one.
- N(ω) The variance of the noise at frequency ω is taken over all f(ω) measurements.

At low contrast these cells have a linear response function.

$$v(t) = \int dt' m(t')c(t-t')$$

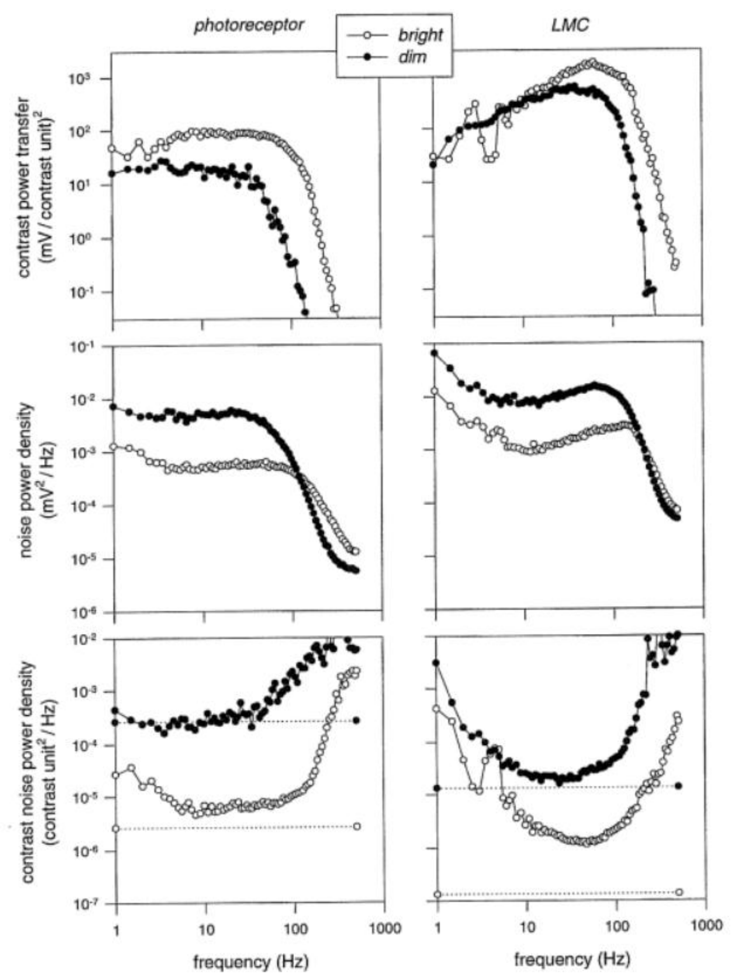
Fourier transforming this we get: V(ω)=M(ω)*C(ω).   M is called the transfer function

In our case: <V(ω)>=M(ω)*C(ω).   From this we find M(ω) and can now move noise to correct units:

$$N_{eff}(\omega) = \frac{N(\omega)}{|M(\omega)|^2}$$

Channel capacity of fly photoreceptor and LMC

Step 3: experimental measurements of the signal to noise ratio: (two cell types, two different intensities)



$C(\omega)$

$N(\omega)$: note the noise depends on light intensity (due to physical limits)
➔ SNR grows nonlinearly with intensity.

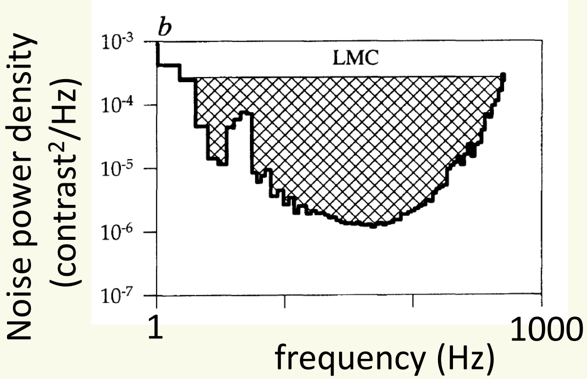
$N_{eff}(\omega)$

➔ $SNR(\omega)= S(\omega)/N_{eff}(\omega)$

Channel capacity of fly photoreceptor and LMC

Step 3: compute channel capacity

- Assume some total signal power (here, 0.1 is natural variance of contrast in natural scenes).
- For a given cell and light intensity we calculate $N_{eff}(\omega)$.
- Use the water filling analogy to calculate the input signal $C_{max}(\omega)$ the maximizes mutual information in a channel with this given noise.
- Calculate the power spectrum of the optimal signal: $S_{max}(\omega)$ and the resulting $SNR_{max}(\omega)$



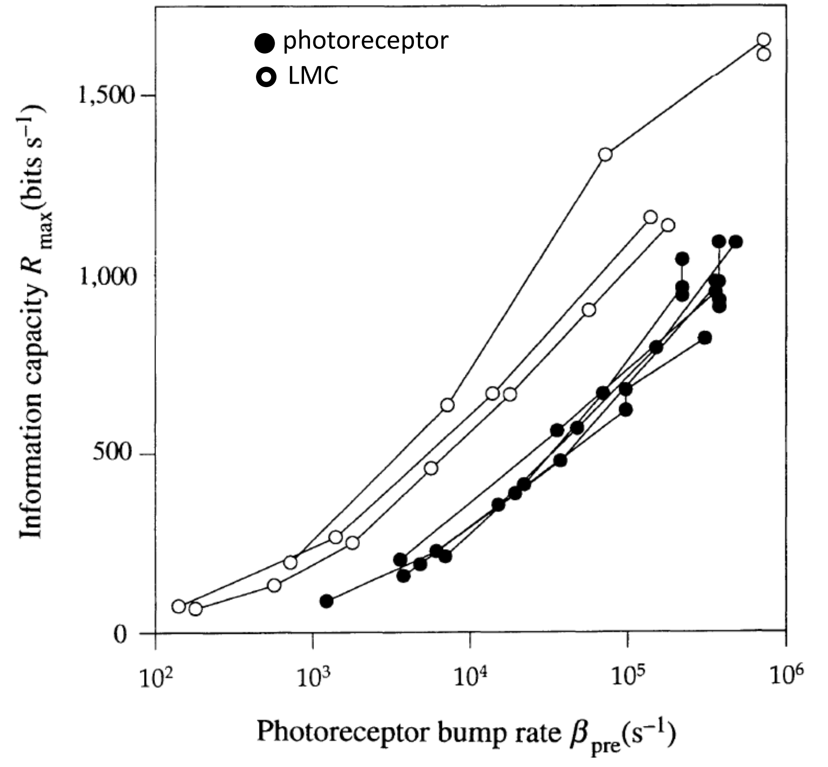
- Use the formula to calculate the maximal mutual information:

$$R_{max} = \frac{1}{2}\int \frac{d\omega}{2\pi} \log\left(1 + SNRmax(\omega)\right) bits/sec$$

EXPERIMENT

Channel capacity of fly photoreceptor and LMC

information transfer rate in the channel
environment →photoreceptor/LMC.



light intensity

The higher the light intensity the less the noise plays a role and the higher the channel capacity.

LMC has higher capacity
Reasonable since it has 6 photoreceptors as inputs.

Goes up to 1,500 bits/sec!

(They measured that with
natural (suboptimal) input signals
the information rate
is not much lower than this capacity.)