Two lectures on Vision

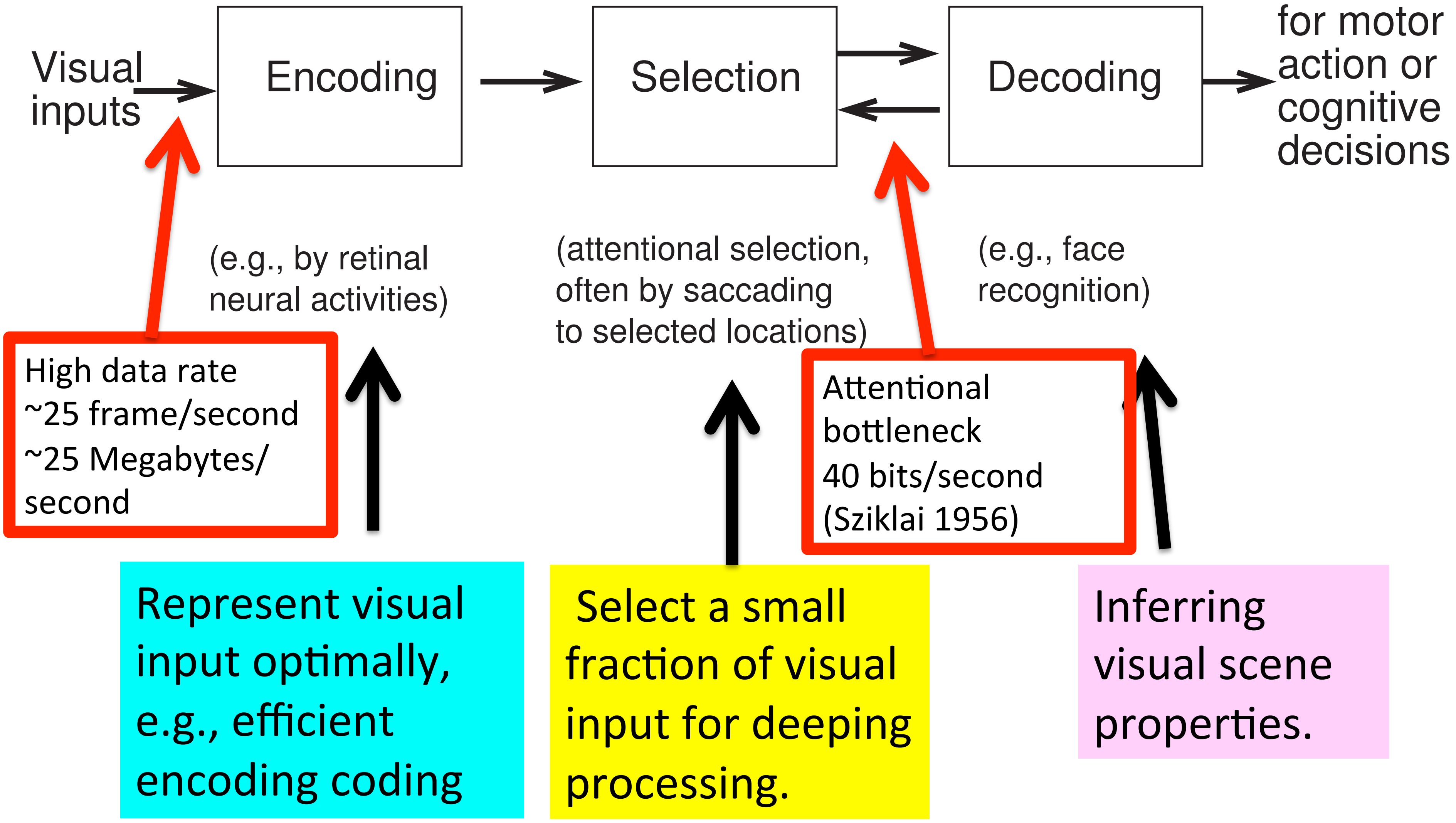
(1) Today: Visual encoding

(2) Tomorrow: Visual selection

For supplementary reading material: see my book “Understanding Vision: theory, models, data”, Oxford University Press, 2014

Li Zhaoping, University College London
Aug 21, 2014
At EU advanced course in computational neuroscience, FIAS, Frankfurt, Germany

The three-stage framework of vision (see Zhaoping 2014)



Visual inputs

Encoding

Selection

Decoding

for motor action or cognitive decisions

(e.g., by retinal neural activities)

(attentional selection, often by saccading to selected locations)

(e.g., face recognition)

High data rate
~25 frame/second
~25 Megabytes/second

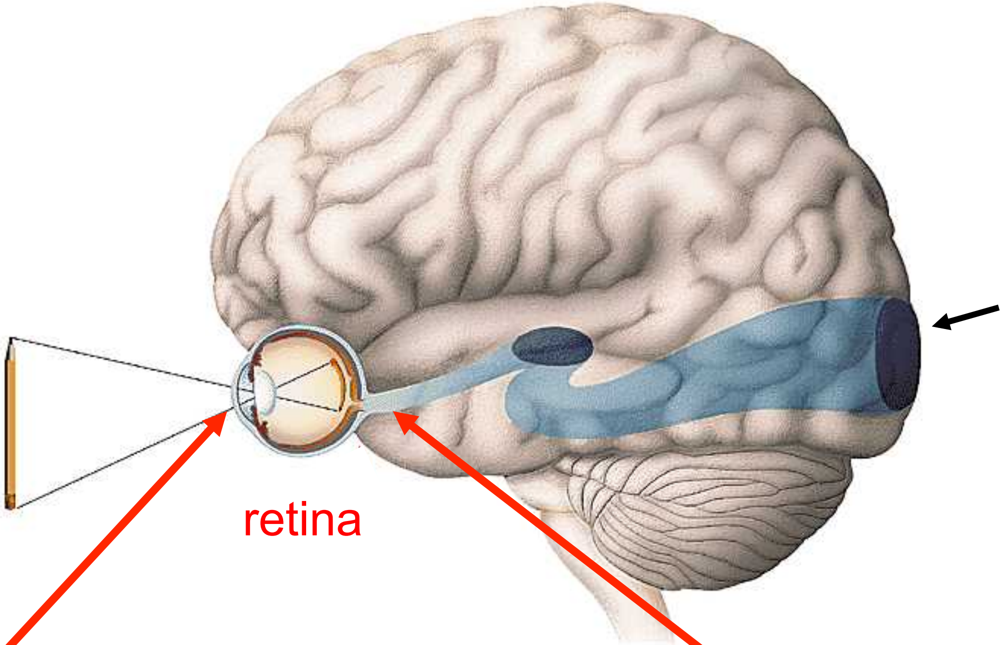
Attentional bottleneck
40 bits/second
(Sziklai 1956)

Represent visual input optimally, e.g., efficient encoding coding

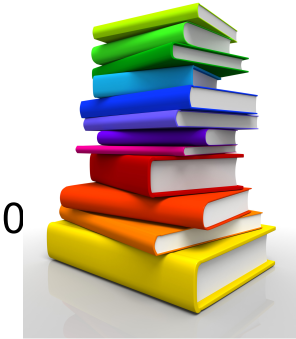
Select a small fraction of visual input for deeping processing.

Inferring visual scene properties.

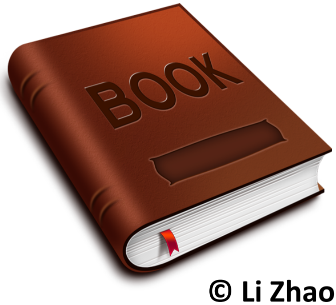
Information bottlenecks in the visual pathway:



10^9 bits/second (Kelly 1962)
~ 25 frames/second, 2000x2000 pixels, 1 byte/pixel



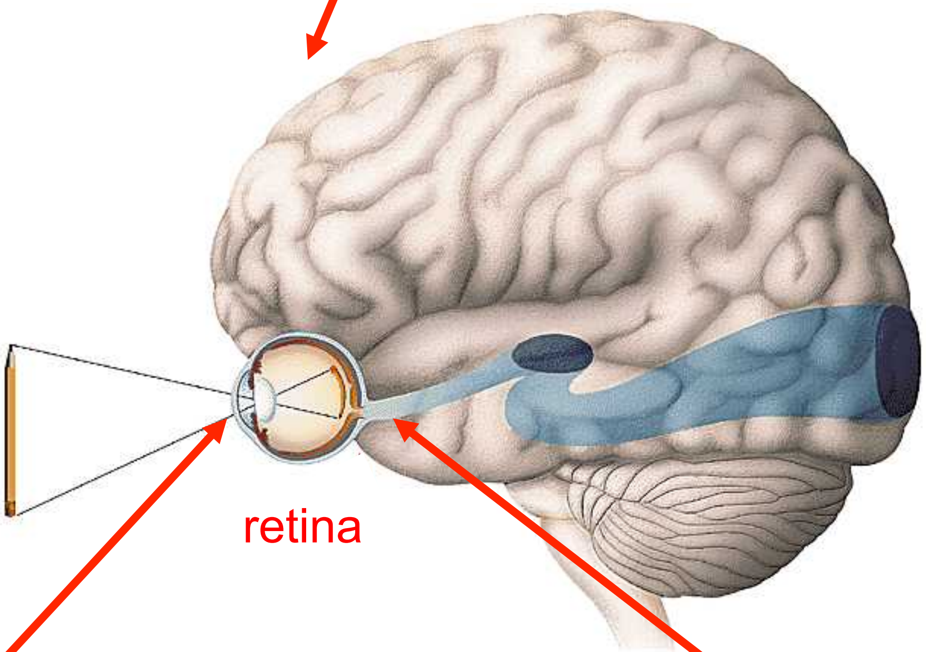
10^7 bits/second
~ 10^6 neurons ,
~10 spikes/neuron
~1 bit/spike



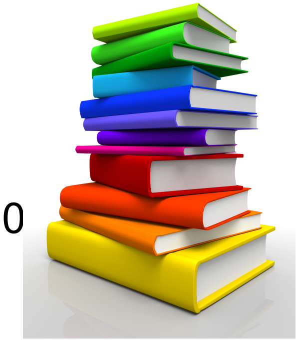
Information bottlenecks in the visual pathway:

Attentional bottleneck ~ 40 bits/second

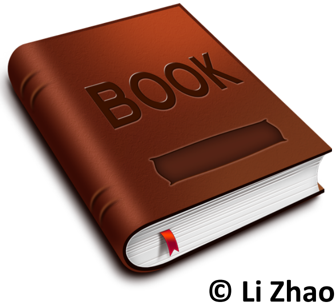
> “To be or not to be,
> This is the question ..”



10^9 bits/second (Kelly 1962)
~ 25 frames/second, 2000x2000 pixels, 1 byte/pixel

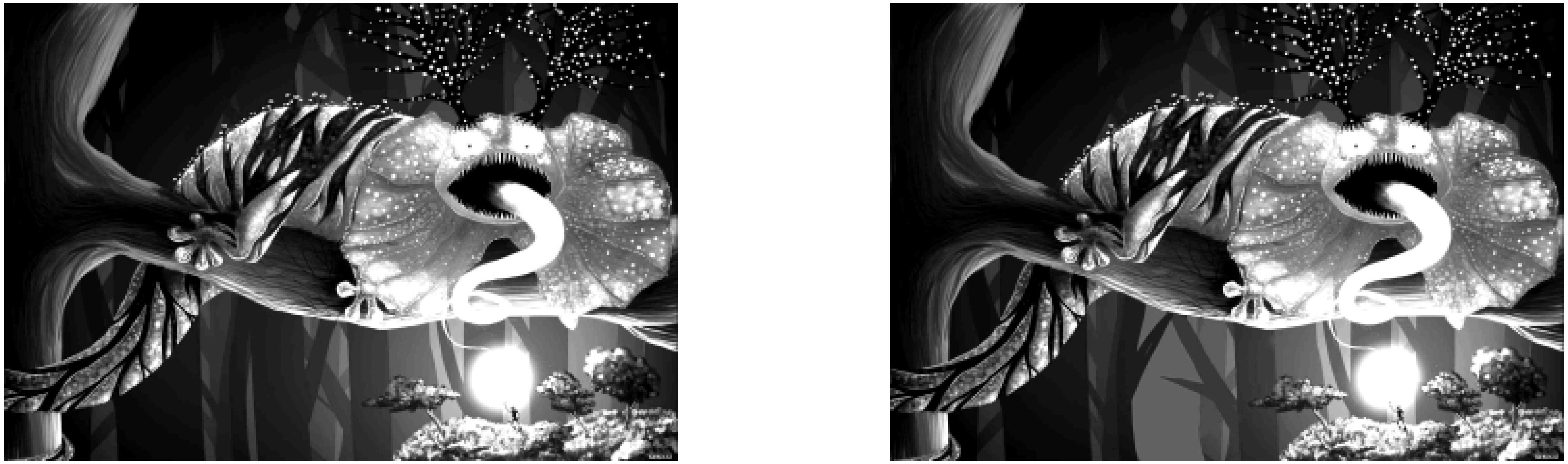


10^7 bits/second
~ 10^6 neurons ,
~10 spikes/neuron
~1 spike/second

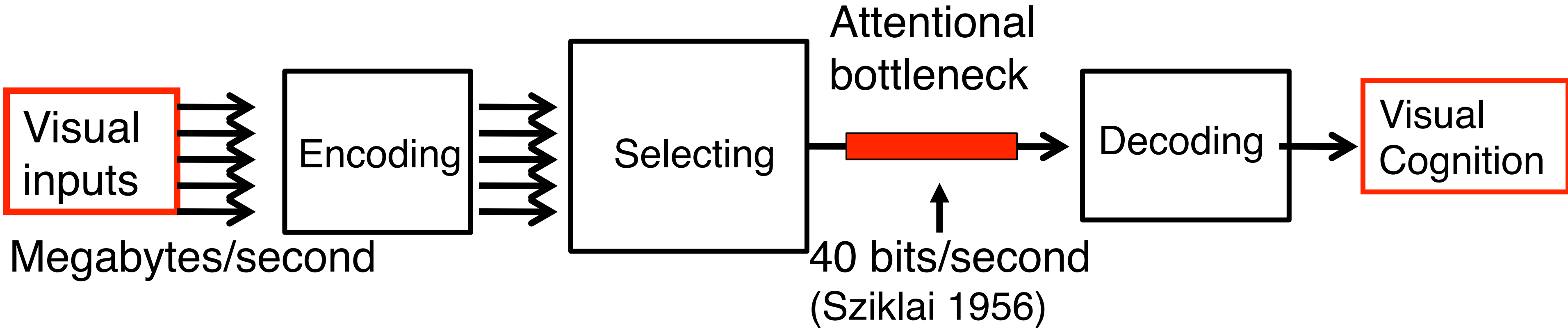


Demo of information deletion --- change blindness

Inattentional blindness — spotting the difference between the two images



We are blind to almost everything except the tiny bit that we pay attention to!



Visual inputs

Megabytes/second

Encoding

Selecting

Attentional bottleneck

40 bits/second (Sziklai 1956)

Decoding

Visual Cognition

Which brain areas for each of the stages?

Visual inputs

Encoding

Selection

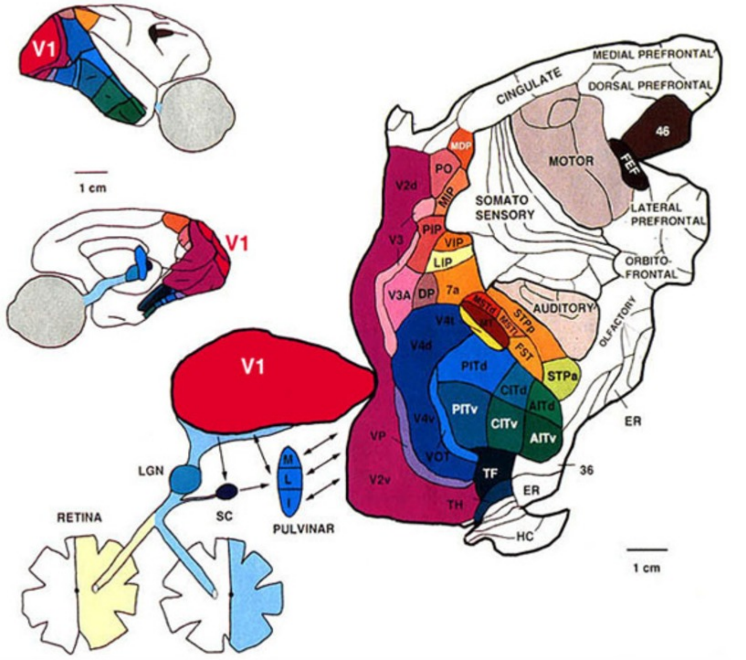
Decoding

for motor action or cognitive decisions

(e.g., by retinal neural activities)

(attentional selection, often by saccading to selected locations)

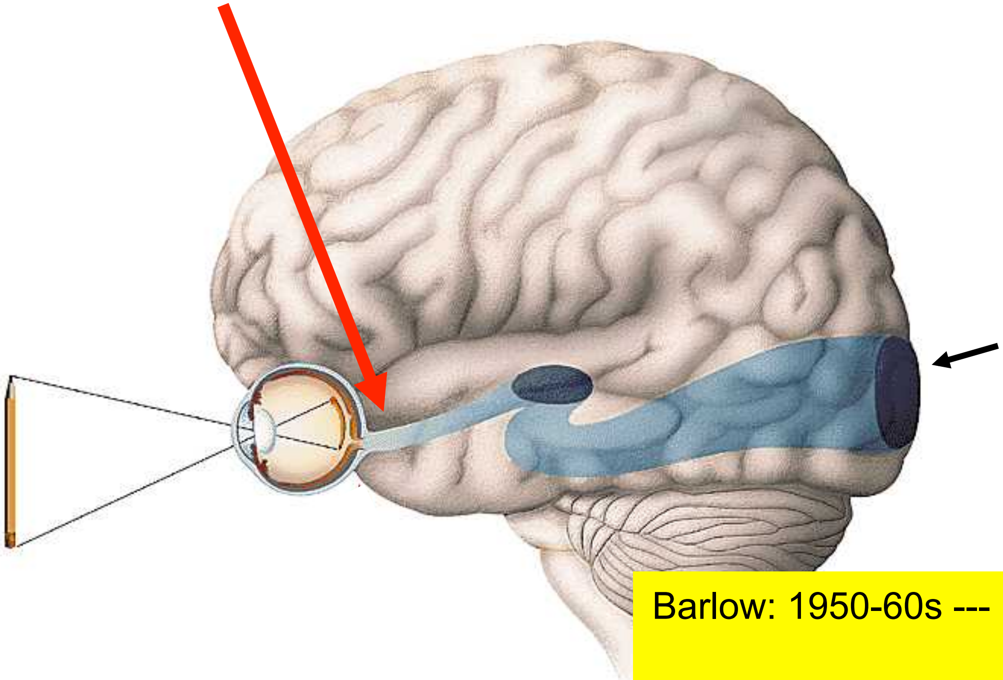
(e.g., face recognition)



Focus today: Efficient encoding in early vision

One hypothesis for early vision

Understanding early visual encoding by
data compression --- maximize information given limited channel capacity



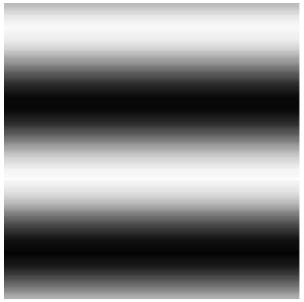
Barlow: 1950-60s --- redundancy reduction.

Laughlin, Linsker, Atick, Redlich, Li, van Hateran, etc. 1980-90s mathematical (information theory) formulation and derivation/prediction

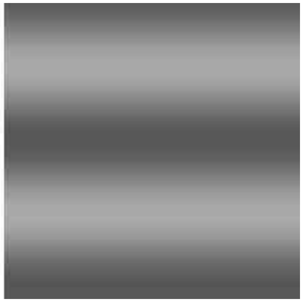
Bell & Sejnowski, Olshausen & Field etc, 1990s, computer simulations.
Also others, after 2000

Example in ocular coding

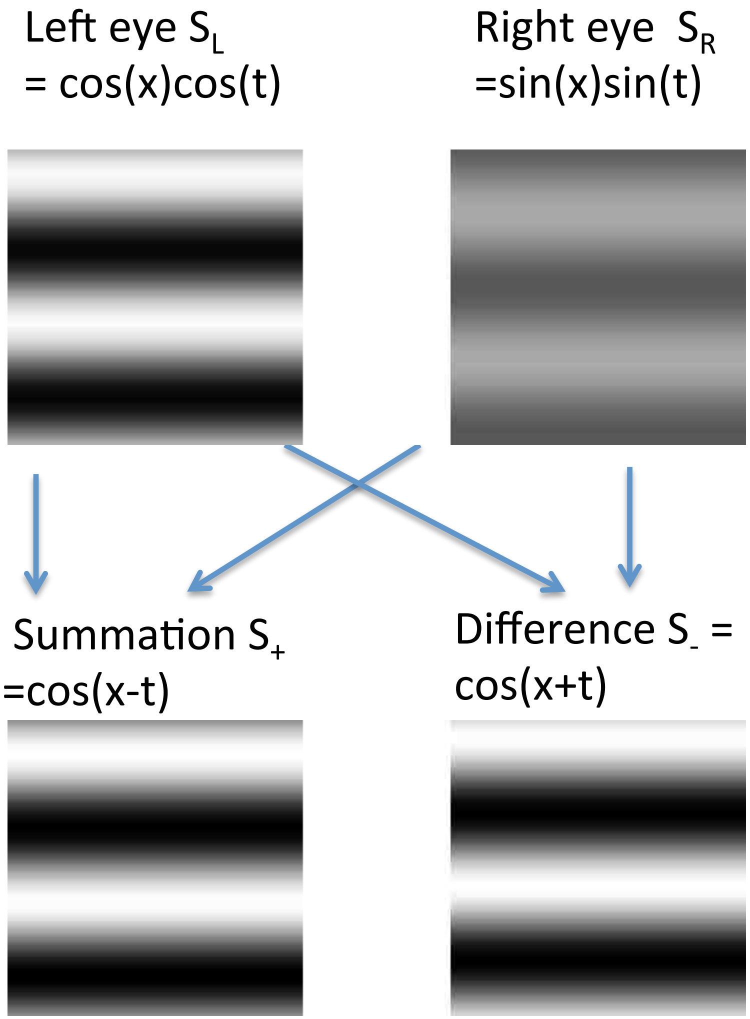
Left eye S_L
$= \cos(x)\cos(t)$



Right eye S_R
$=\sin(x)\sin(t)$



Example in ocular coding



Left eye S_L
$= \cos(x)\cos(t)$

Right eye S_R
$=\sin(x)\sin(t)$

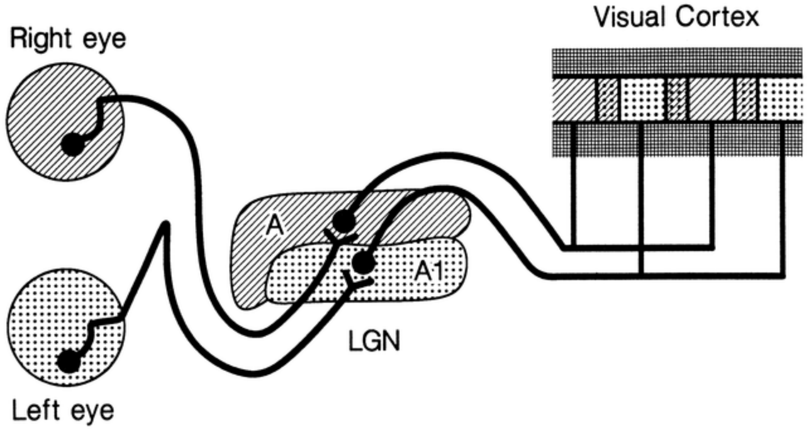
Summation S_+
$=\cos(x-t)$

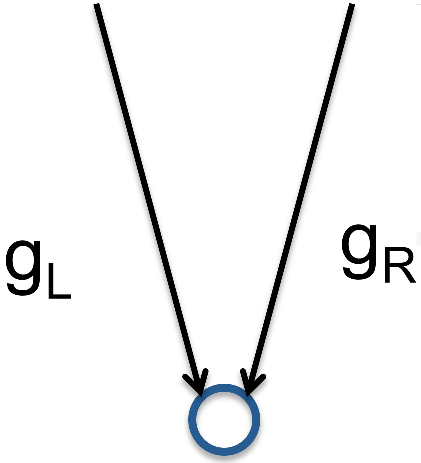
Difference $S_- =$
$\cos(x+t)$

Stereo input: S_L, S_R on the retina

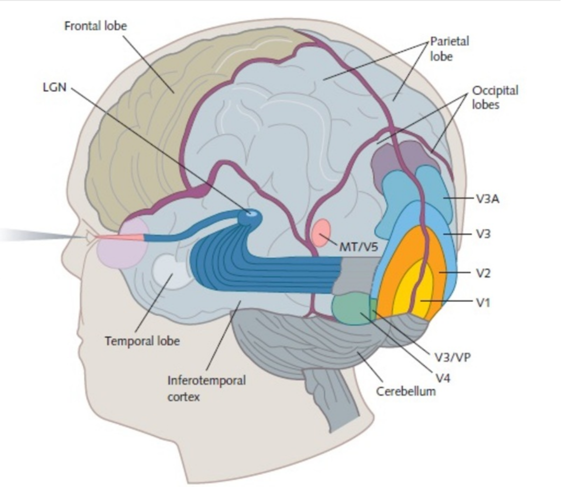
S^L

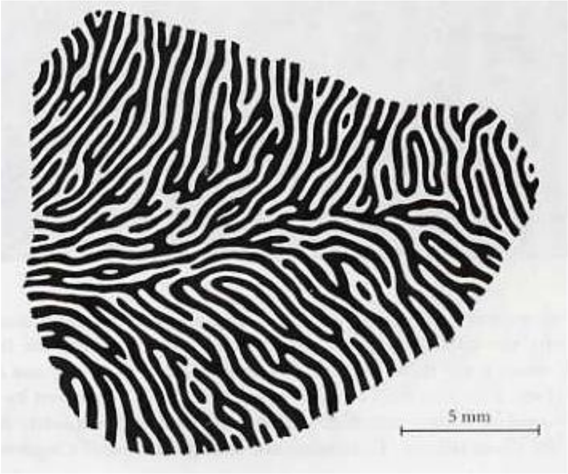
S^R





A neuron in V1





Ocular dominance columns

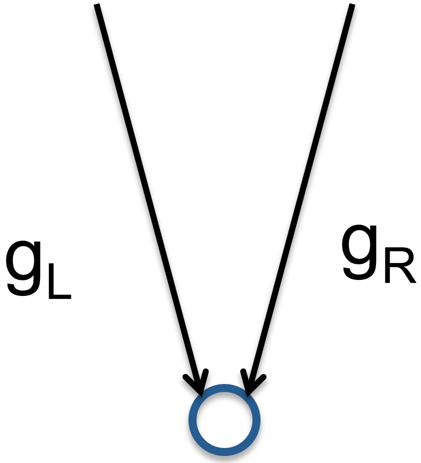
Encoded

Output $=g_L S_L + g_R S_R$

Stereo input: S_L, S_R on the retina

Formulation:
Input (S_L, S_R)

Neural encoded signals:
$g_L S_L + g_R S_R$



Question: Can we understand the encoding (g_L, g_R) from efficient coding point of view?

A neuron in V1

Encoded
Output $=g_L S_L + g_R S_R$

Stereo input: S_L, S_R on the retina

S^L

S^R

There is redundancy in the input from the two eyes, they are highly correlated with each other.

e.g., at one pixel or one Fourier amplitude of particular frequency

$$R^S \equiv \begin{pmatrix} \langle S_L^2 \rangle & \langle S_L S_R \rangle \\ \langle S_R S_L \rangle & \langle S_R^2 \rangle \end{pmatrix} = \langle S_L^2 \rangle \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$$

This correlated has been measured (Li and Atick 1994)

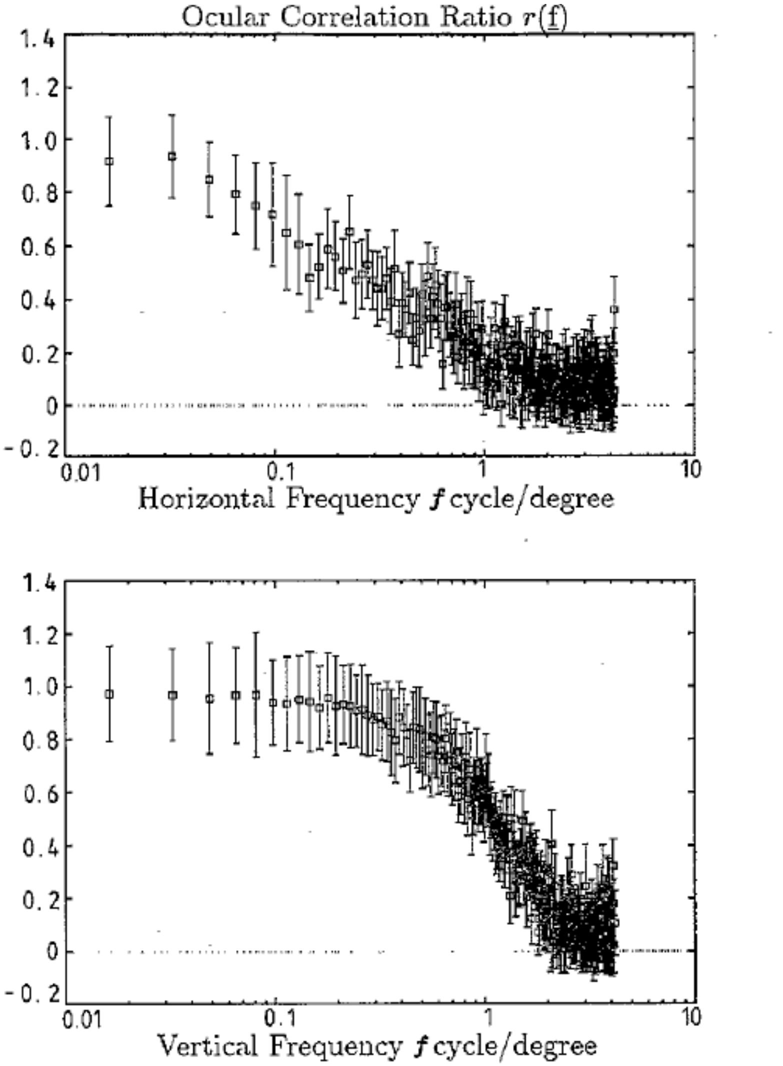


Figure 1. Measured $r(f)$ (data points with error bars) function from stereo images for slices in horizontal $f = (f, 0)$ and vertical $f = (0, f)$ frequencies

Stereo input: S_L, S_R on the retina

S^L

S^R

There is redundancy in the input from the two eyes, they are highly correlated with each other.

e.g., at one pixel or one Fourier amplitude of particular frequency

↓ 1 byte

↓ 1 byte

Because of redundancy, total information < 2 bytes.

Hence, the raw input is highly inefficient in information representation.

This correlated has been measured (Li and Atick 1994)

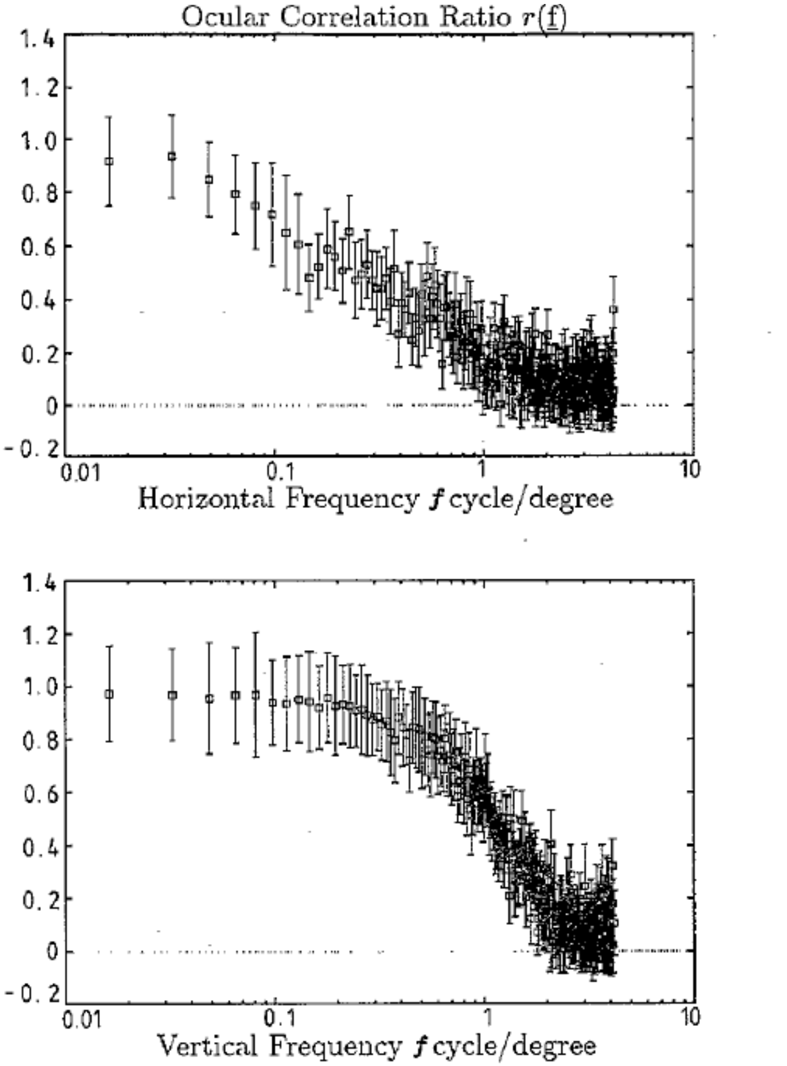


Figure 1. Measured $r(f)$ (data points with error bars) function from stereo images for slices in horizontal $f = (f, 0)$ and vertical $f = (0, f)$ frequencies

Stereo input: S_L, S_R on the retina

S^L

S^R

Coding transform K:

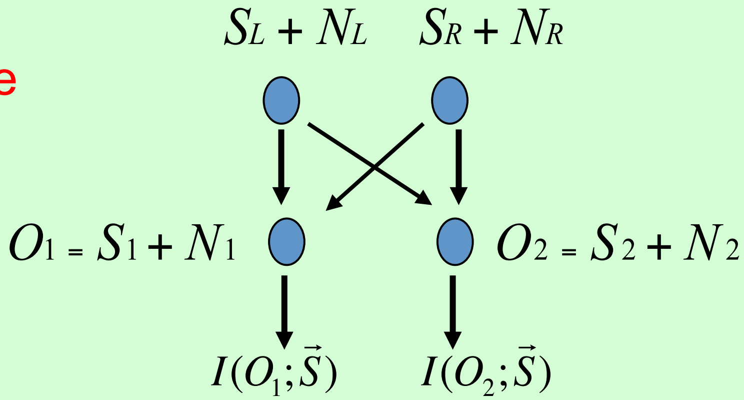
$$O_i = [\sum_j \mathsf{K}_{ij}(S_j + N_j)] + (N_o)_i$$

Information extracted by O about S:

$$I(\vec{O};\vec{S})$$

Neural cost: e.g. the firing rates in O, dynamic range of O, metabolic energy, etc.

One possible solution: encode to two new outputs: O_1 and O_2 which are not redundant.



$$I(O_1;\vec{S}) + I(O_2;\vec{S}) \geq I(\vec{O};\vec{S})$$

With no redundancy between O_1 and O_2

$$I(O_1;\vec{S}) + I(O_2;\vec{S}) = I(\vec{O};\vec{S})$$

Efficient coding:

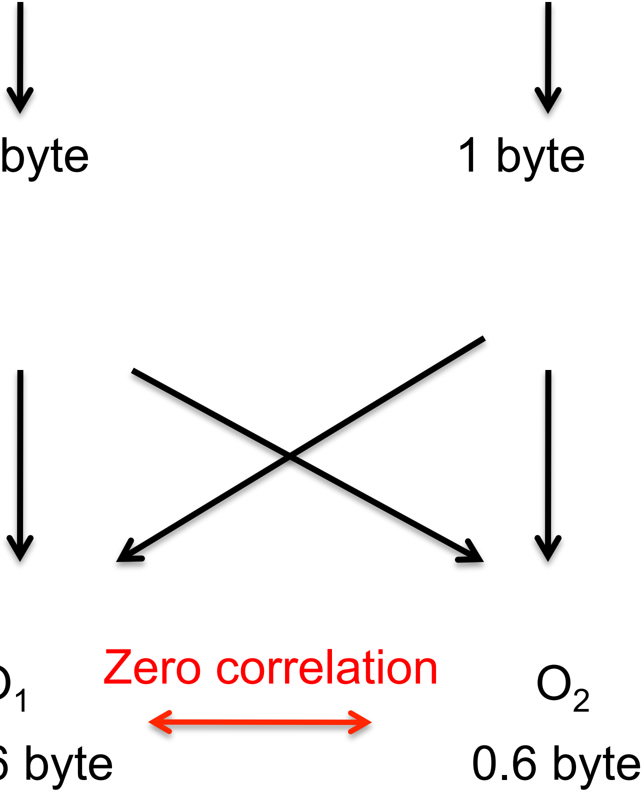
Find K to maximize $I(\vec{O};\vec{S})$ while minimize cost.

(I will show the gist of the solution by some hand waving, see the book for the more rigorous treatment).

Stereo input: S_L, S_R on the retina

S^L

S^R



1 byte

1 byte

O_1

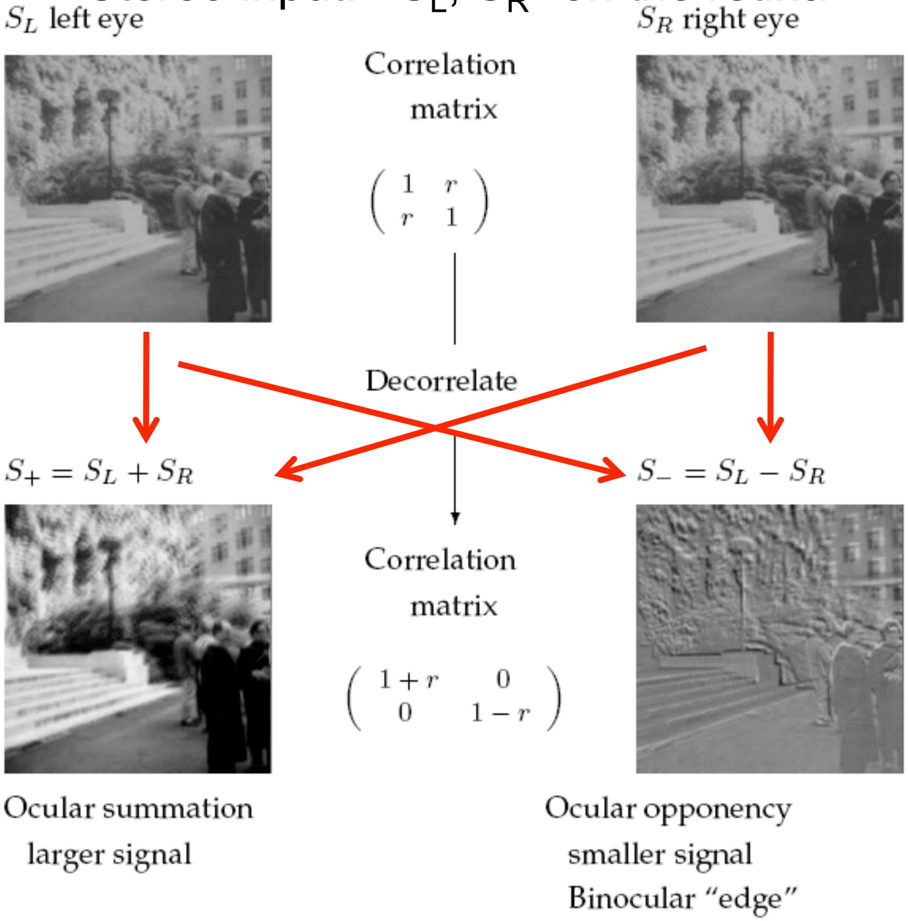
0.6 byte

Zero correlation

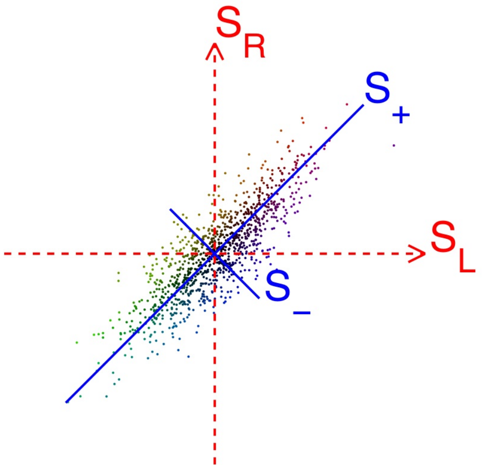
O_2

0.6 byte

Stereo input: S_L, S_R on the retina



Encoding is like a coordinate rotation



Note:

S_+ is binocular,
S_- is more monocular-like.

S_+ and S_- are eigenvectors or principal components of the correlation matrix R^S, with eigenvalues
$<S^2_\pm> = (1\pm r) <S_L{}^2>$

Consider redundancy and encoding of stereo signals

S^L

S^R

Redundancy is seen at correlation matrix (between two eyes)

$$R^S \equiv \begin{pmatrix} \langle S_L^2 \rangle & \langle S_L S_R \rangle \\ \langle S_R S_L \rangle & \langle S_R^2 \rangle \end{pmatrix} = \langle S_L^2 \rangle \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$$

0< r < 1.

When gaussian: $P(S_L, S_R) \sim \exp(-\sum_{ij} S_i S_j / (R^S)^{-1}_{ij} / 2)$

$$\longrightarrow \quad P(S_L, S_R) \neq P(S_L) P(S_R)$$

Factorial encoding: $S_+ \propto S_L + S_R,\qquad S_- \propto S_L - S_R$

$$\langle S_+ S_- \rangle = 0, \qquad P(S_+, S_-) = P(S_+) P(S_-)$$

Gain control of the de-correlated channels

S_L left eye

S_R right eye

Correlation matrix

$$\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$$

Decorrelate

Correlation matrix

$$\begin{pmatrix} 1+r & 0 \\ 0 & 1-r \end{pmatrix}$$

$S_+ = S_L + S_R$

Ocular summation
larger signal

$S_- = S_L - S_R$

Ocular opponency
smaller signal
Binocular "edge"

g_+S_+

Gain control

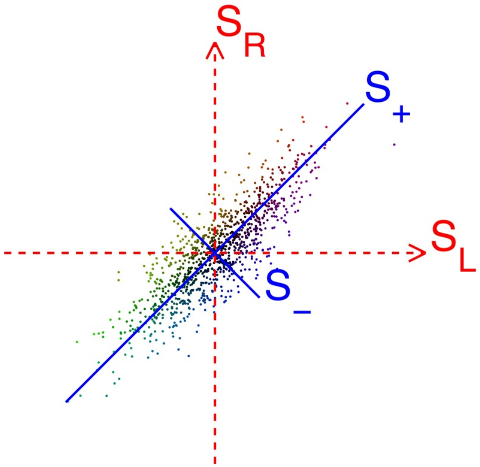
(e.g., whitening)

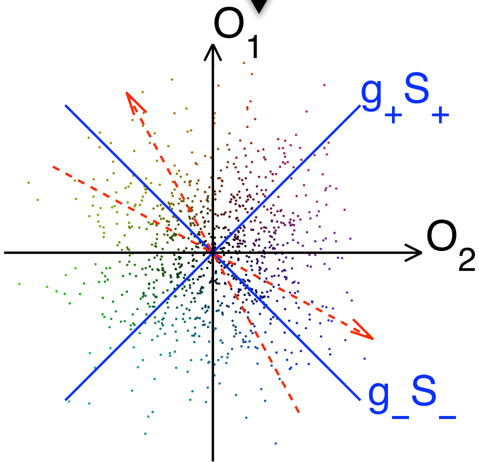
$$g_\pm \propto \left\langle S_\pm^2 \right\rangle^{-1/2}$$

g_-S_-

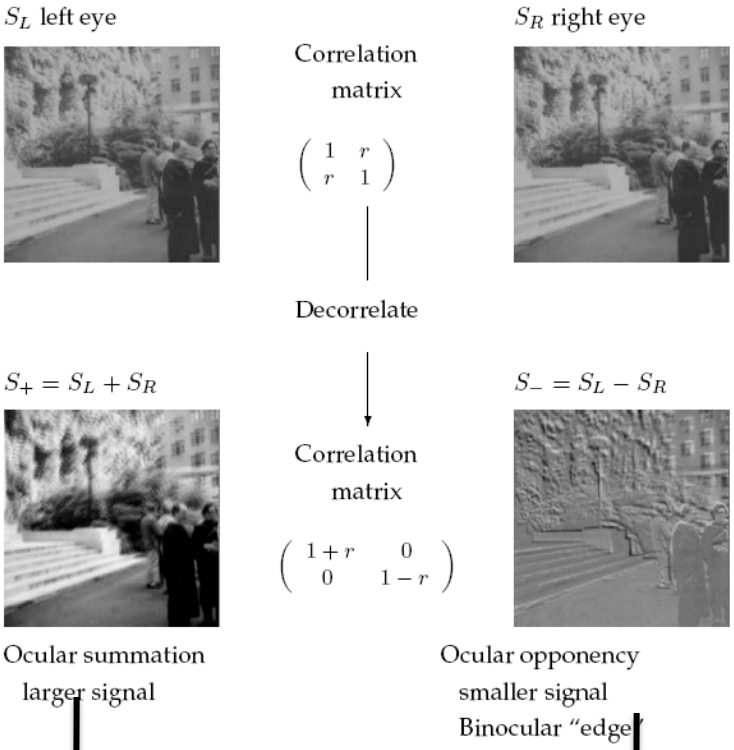
$g_+ < g_-$, since S_+ channel has larger signals

Encoding is like a coordinate rotation

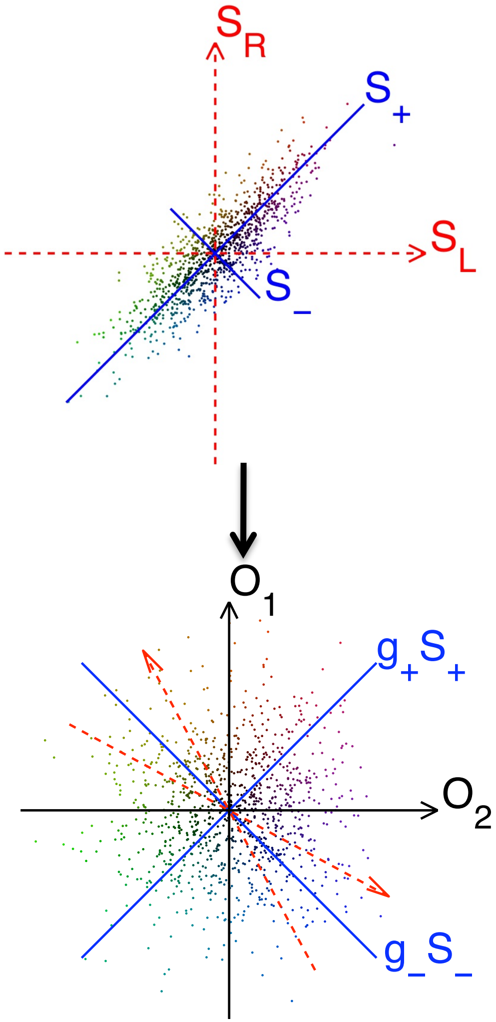




Then the whole efficient stereo coding:



Encoding is like a coordinate rotation



g_+S_+

Gain control (e.g., whitening)

$$g_\pm \propto \left\langle S_\pm^2 \right\rangle^{-1/2}$$

g_-S_-

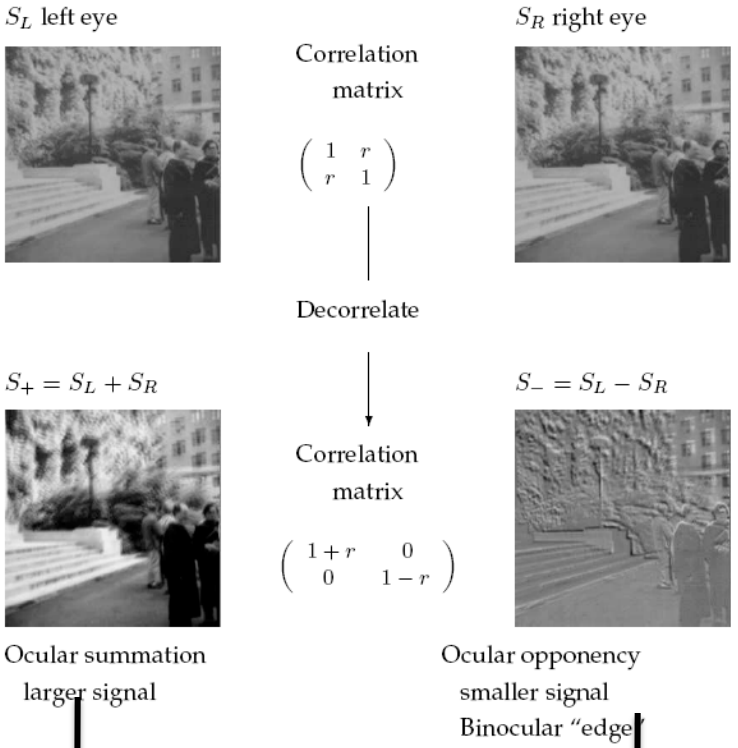
Multiplex

$$O_1 = g_+S_+ - g_-S_-$$
$$= (g_+ - g_-)S_L + (g_+ + g_-)S_R$$

$$O_2 = g_+S_+ + g_-S_-$$
$$= (g_+ + g_-)S_L + (g_+ - g_-)S_R$$

Cortical neurons

Then the whole efficient stereo coding:



g_+S_+

Gain control (e.g., whitening)

$$g_\pm \propto \left\langle S_\pm^2 \right\rangle^{-1/2}$$

g_-S_-

Multiplex

$$O_1 = g_+S_+ - g_-S_-$$
$$= (g_+ - g_-)S_L + (g_+ + g_-)S_R$$

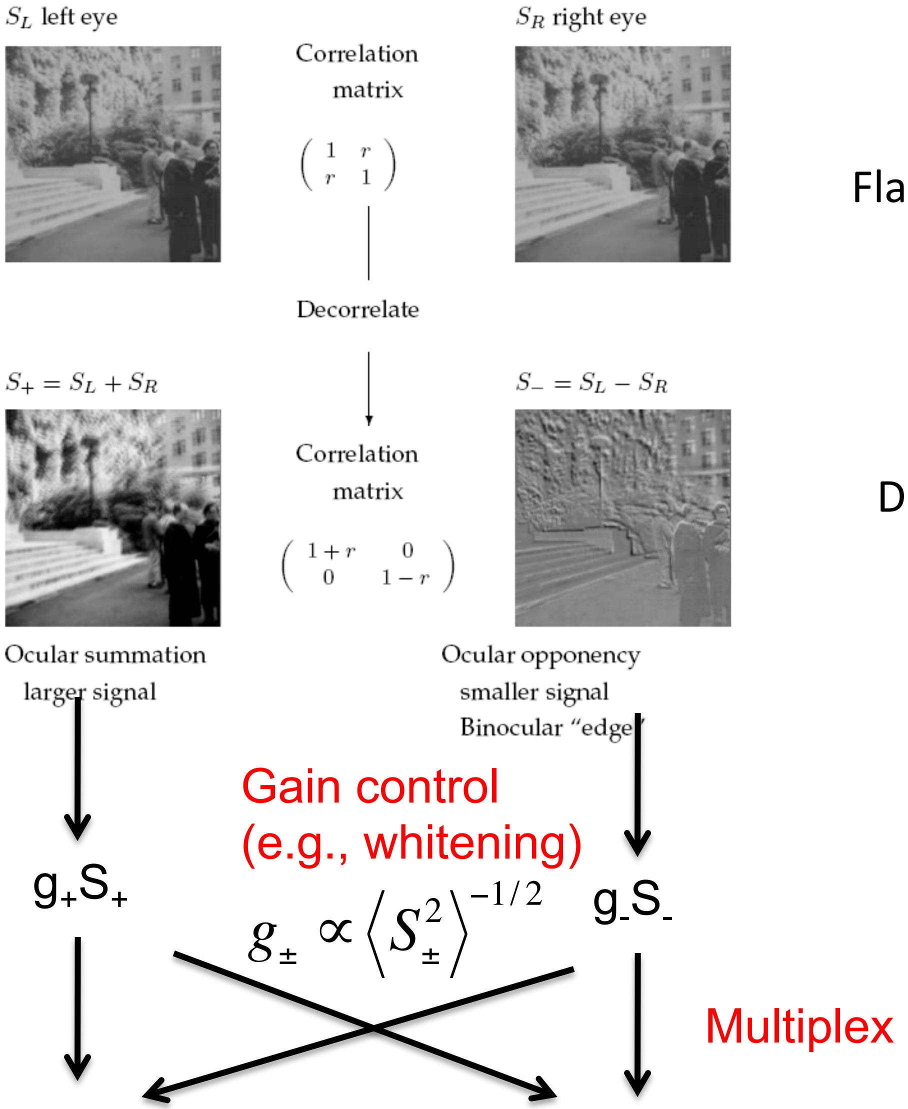
$$O_2 = g_+S_+ + g_-S_-$$
$$= (g_+ + g_-)S_L + (g_+ - g_-)S_R$$

Cortical neurons

Let us probe these two underlying channels

$$O = g_L S_L + g_R S_R$$
$$\rightarrow g_+ = (g_L + g_R)/2$$
$$g_- = (g_L - g_R)/2$$

Then the whole efficient stereo coding:



Gain control
(e.g., whitening)

g_+S_+

g_-S_-

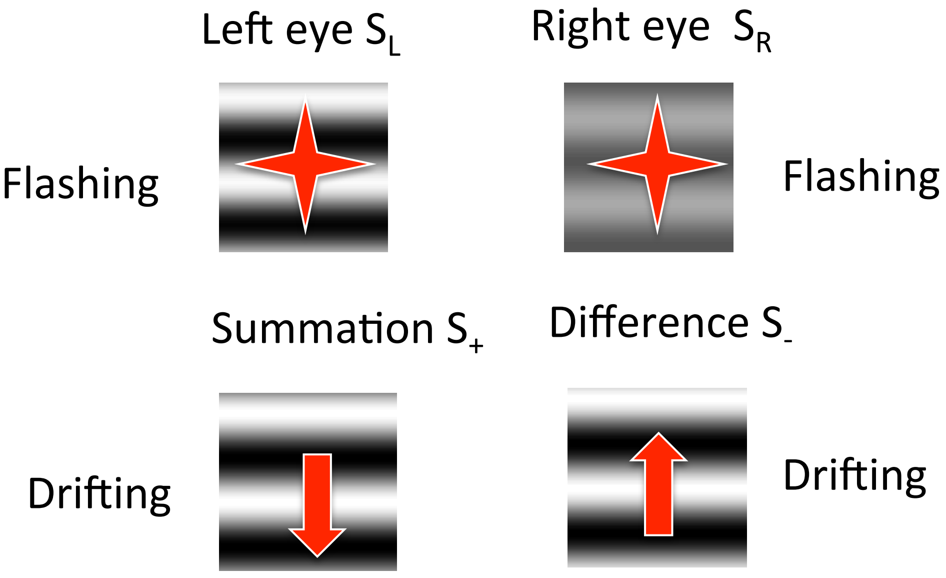
$$g_\pm \propto \left\langle S_\pm^2 \right\rangle^{-1/2}$$

Multiplex

$$O_1 = g_+S_+ - g_-S_-$$
$$= (g_+ - g_-)S_L + (g_+ + g_-)S_R$$

$$O_2 = g_+S_+ + g_-S_-$$
$$= (g_+ + g_-)S_L + (g_+ - g_-)S_R$$

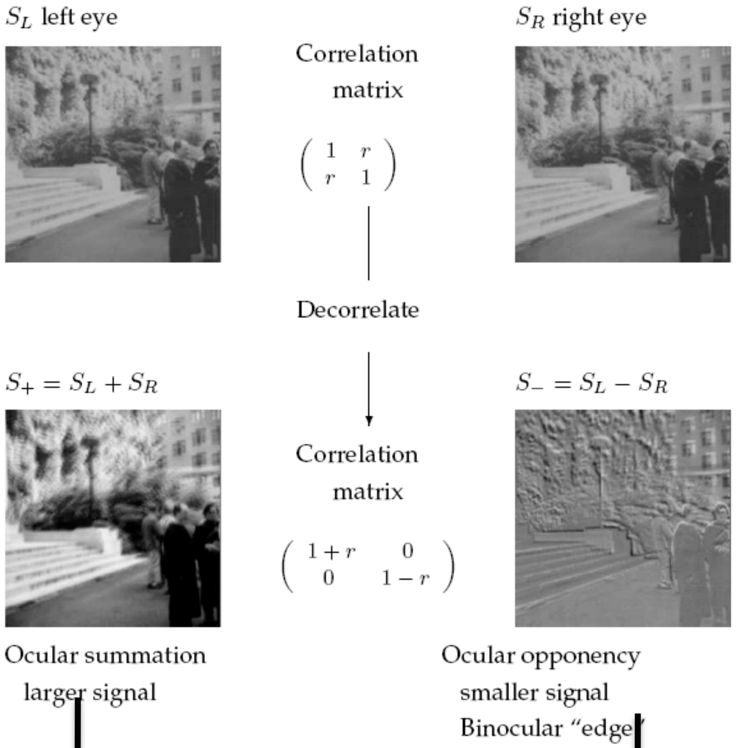
Cortical neurons



The chance of seeing this drift should increase/decrease with

g_+/g_-

Prediction: adaptation of stereo coding to different input statistics.



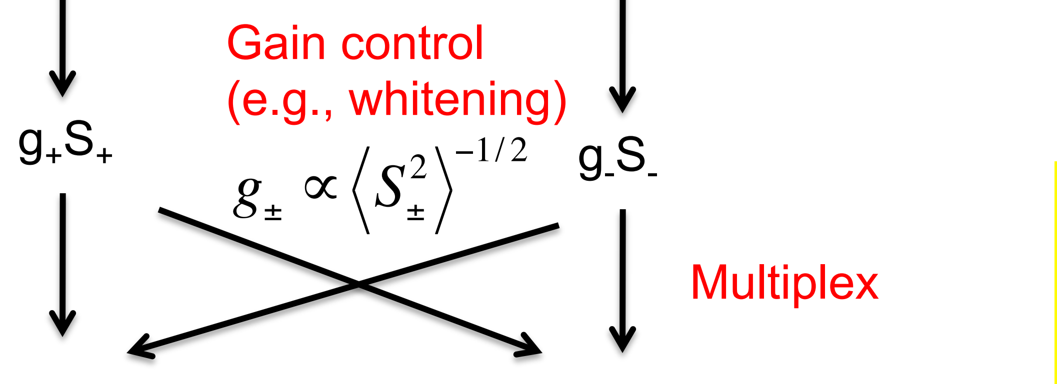
$$g_\pm \propto \left\langle S_\pm^2 \right\rangle^{-1/2}$$

Sensitivity to the channel signal

Signal power in the channel

$$\left\langle S_\pm^2 \right\rangle = (1 \pm r)\left\langle S_L^2 \right\rangle$$

Ocular correlation



g_+S_+

Gain control (e.g., whitening)

$g_\pm \propto \left\langle S_\pm^2 \right\rangle^{-1/2}$

g_-S_-

Multiplex

$$O_1 = g_+S_+ - g_-S_- = (g_+ - g_-)S_L + (g_+ + g_-)S_R$$

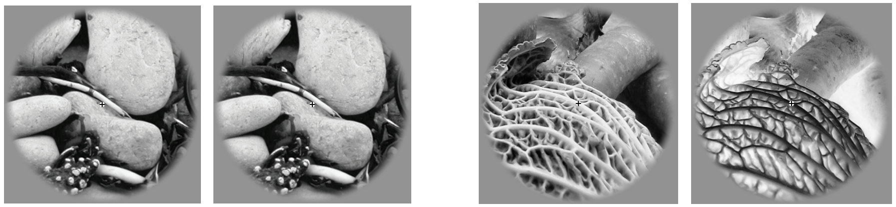
$$O_2 = g_+S_+ + g_-S_- = (g_+ + g_-)S_L + (g_+ - g_-)S_R$$

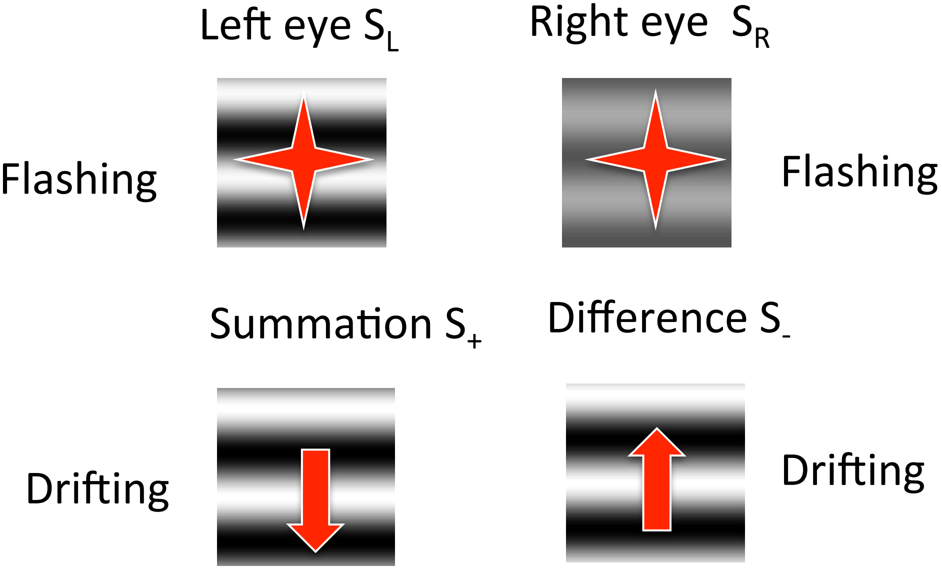
Cortical neurons

Adapting to inputs of high or low r should decrease or increase

g_+/g_-

Prediction: adaptation of stereo coding to different input statistics.





$$g_\pm \propto \left\langle S_\pm^2 \right\rangle^{-1/2}$$

Sensitivity to the channel signal

Signal power in the channel

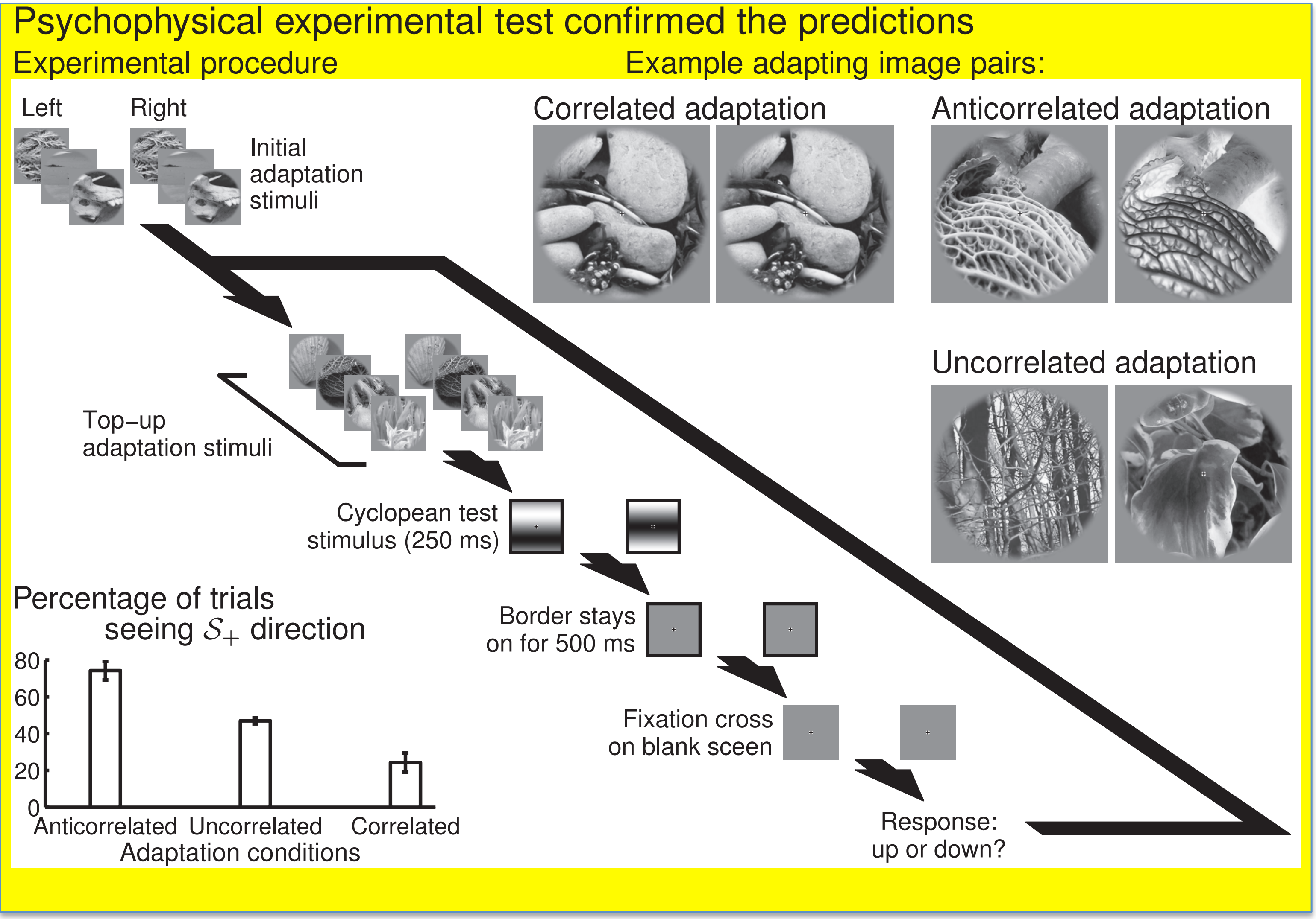
$$\left\langle S_\pm^2 \right\rangle = (1 \pm r)\left\langle S_L^2 \right\rangle$$

Ocular correlation

Prediction (May et al 2012): one should see less/more of the S_+ drift by adapting to stronger S_+/S_- input, i.e., ocularly correlated/anti-correlated input

Adapting to inputs of high or low r should decrease or increase g_+/g_-

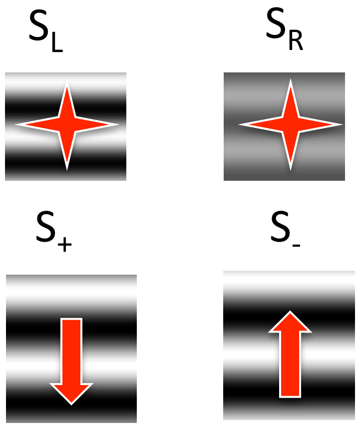
Confirming the prediction: (May, Zhaoping, Hibbard 2012)



Psychophysical experimental test confirmed the predictions

Experimental procedure

Example adapting image pairs:



Expositions on efficient coding:

Recipe for efficient coding

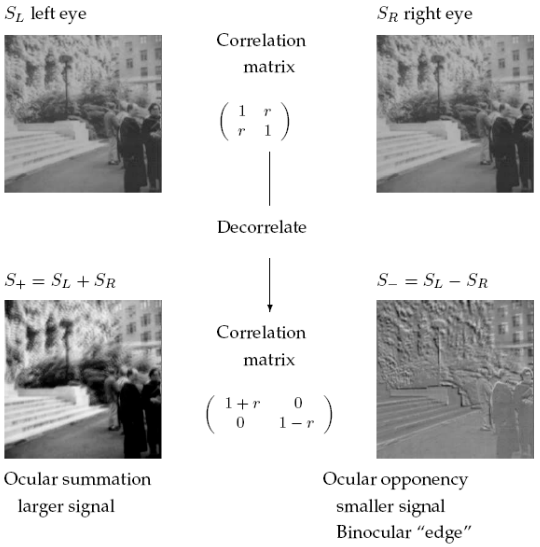
Efficient color coding

Efficient spatial coding

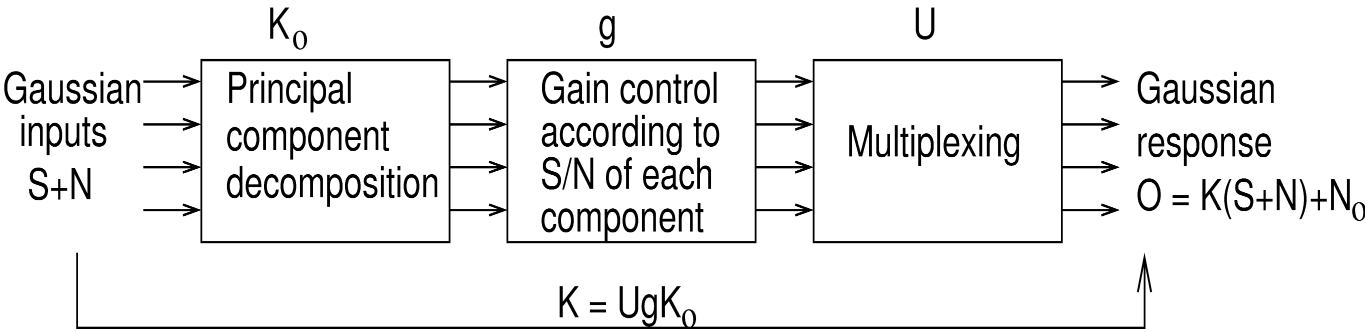
Efficient temporal coding, etc

How visual coding adapts to signal-to-noise of inputs

Efficient coding recipe:



Efficient coding K of Gaussian signals decomposed into three components: K_0, g, U.



K_0

g

U

Gaussian inputs S+N

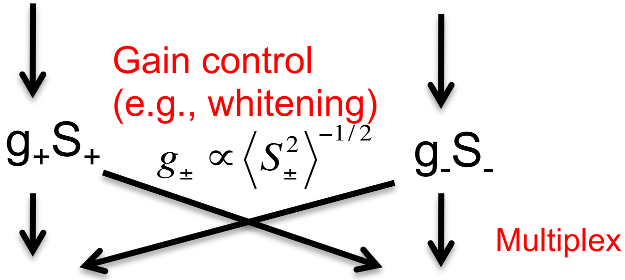
Principal component decomposition

Gain control according to S/N of each component

Multiplexing

Gaussian response $O = K(S+N)+N_0$

$K = UgK_0$



g_+S_+

Gain control (e.g., whitening)

$$g_\pm \propto \left\langle S_\pm^2 \right\rangle^{-1/2}$$

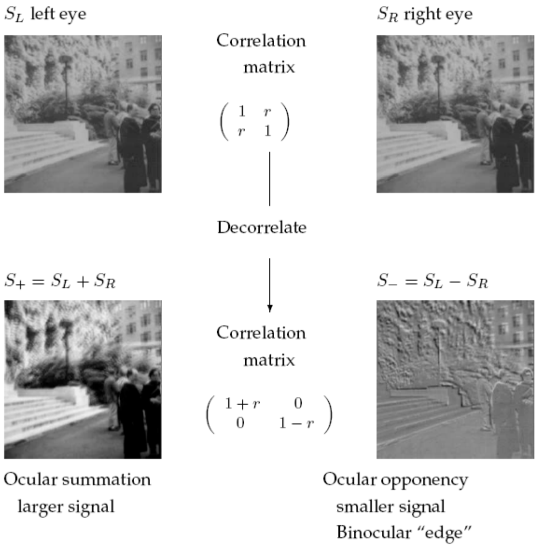
g_-S_-

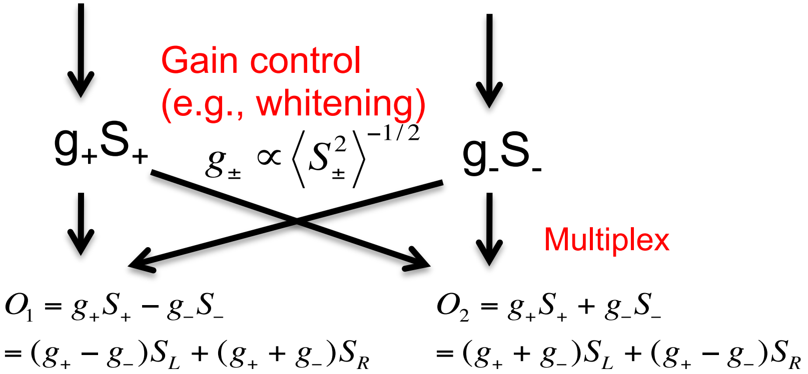
Multiplex

$$O_1 = g_+S_+ - g_-S_-$$
$$= (g_+ - g_-)S_L + (g_+ + g_-)S_R$$

$$O_2 = g_+S_+ + g_-S_-$$
$$= (g_+ + g_-)S_L + (g_+ - g_-)S_R$$

Applying to efficient spatial coding by analogy:





Gain control
(e.g., whitening)

g_+S_+

$g_{\pm} \propto \left\langle S_{\pm}^2 \right\rangle^{-1/2}$

g_-S_-

Multiplex

$$O_1 = g_+S_+ - g_-S_- = (g_+ - g_-)S_L + (g_+ + g_-)S_R$$

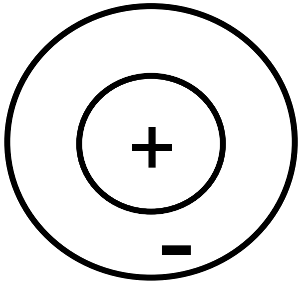
$$O_2 = g_+S_+ + g_-S_- = (g_+ + g_-)S_L + (g_+ - g_-)S_R$$

Left eye, Right eye
→ One pixel's input, another pixel's input

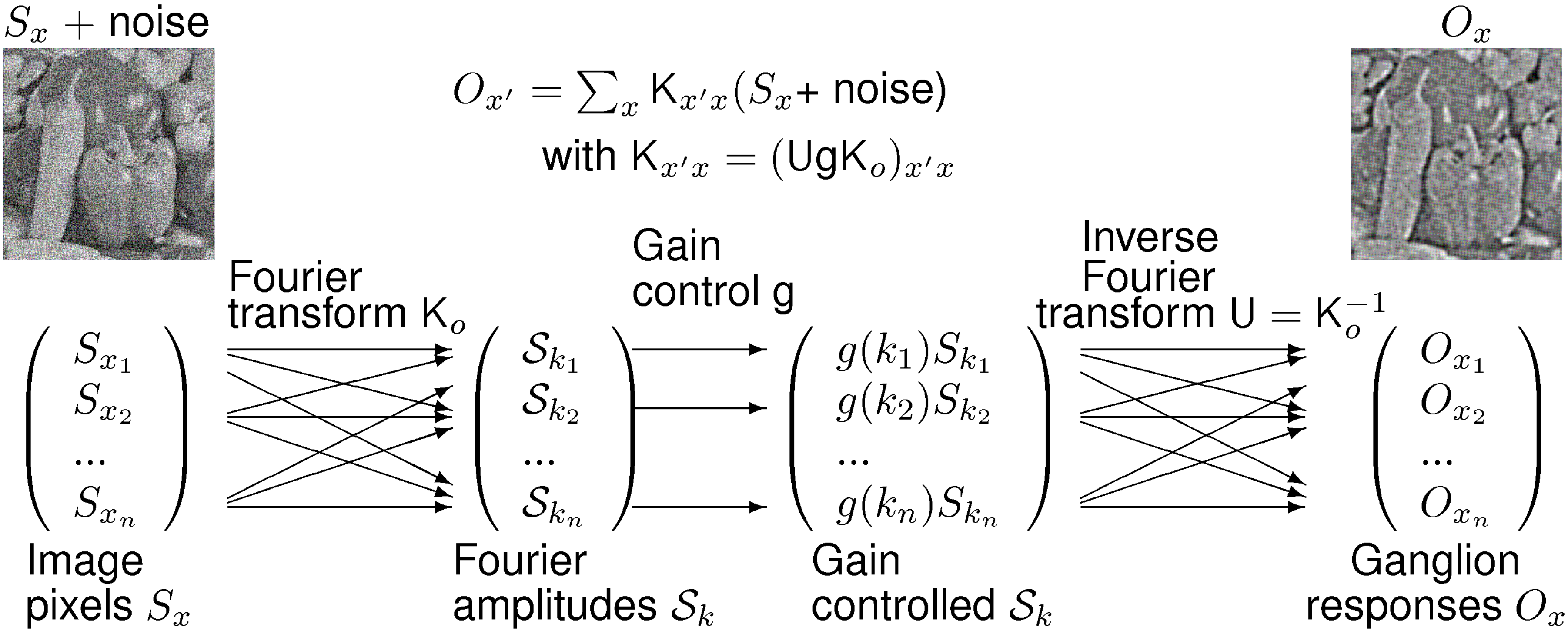
Decorrelation:
→ (1) Spatial summation/average
→ (2) Spatial contrast

Gain control:
→ Higher sensitivity to spatial contrast (when signal to noise is high)

Multiplex:
→Center surround receptive-field



More mathematically:



S_x + noise

O_x

$$O_{x'} = \sum_x \mathrm{K}_{x'x}(S_x + \text{noise})$$

with $\mathrm{K}_{x'x} = (\mathrm{UgK}_o)_{x'x}$

Fourier transform K_o

Gain control g

Inverse Fourier transform $\mathrm{U} = \mathrm{K}_o^{-1}$

$$\begin{pmatrix} S_{x_1} \\ S_{x_2} \\ \dots \\ S_{x_n} \end{pmatrix} \quad \begin{pmatrix} \mathcal{S}_{k_1} \\ \mathcal{S}_{k_2} \\ \dots \\ \mathcal{S}_{k_n} \end{pmatrix} \quad \begin{pmatrix} g(k_1)S_{k_1} \\ g(k_2)S_{k_2} \\ \dots \\ g(k_n)S_{k_n} \end{pmatrix} \quad \begin{pmatrix} O_{x_1} \\ O_{x_2} \\ \dots \\ O_{x_n} \end{pmatrix}$$

Image pixels S_x

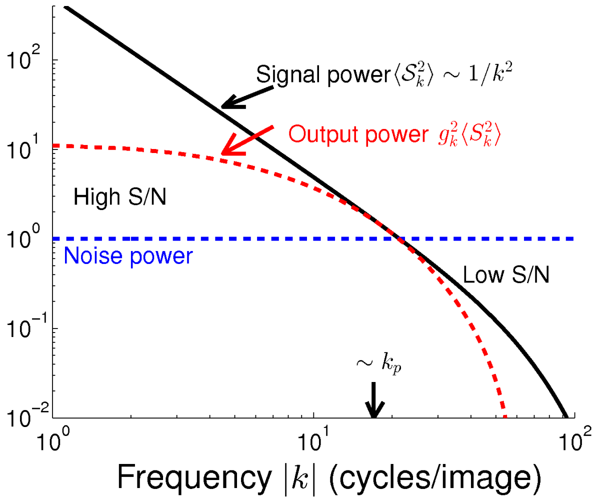
Fourier amplitudes $\mathcal{S}_k$

Gain controlled $\mathcal{S}_k$

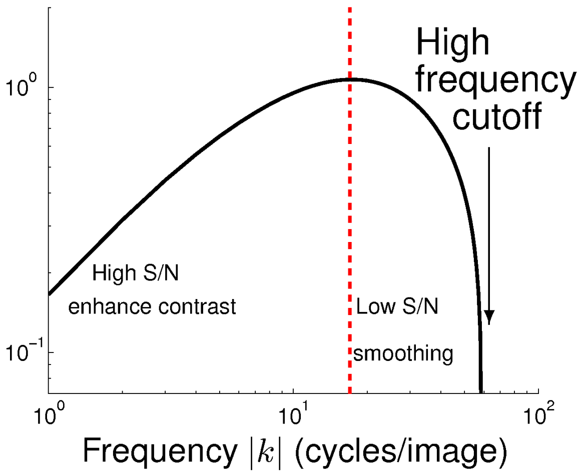
Ganglion responses O_x

Figure from "Understanding vision: theory, models, and data", by Li Zhaoping, Oxford University Press, 2014

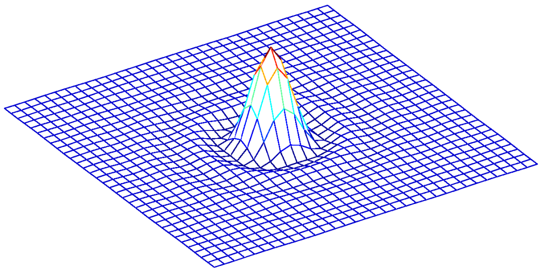
A: Power of input, $\langle|\mathcal{S}_k|^2\rangle$, output $g_k^2\langle\mathcal{S}_k^2\rangle$, and noise $\langle\mathcal{N}_k^2\rangle$ versus k.



B: The optimal gain g_k (or $g(k)$) versus k



C: The spatial receptive field



D: Visualizing your own $g(k)$

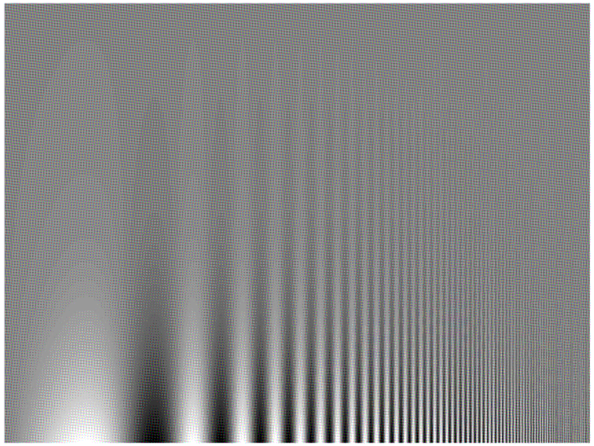
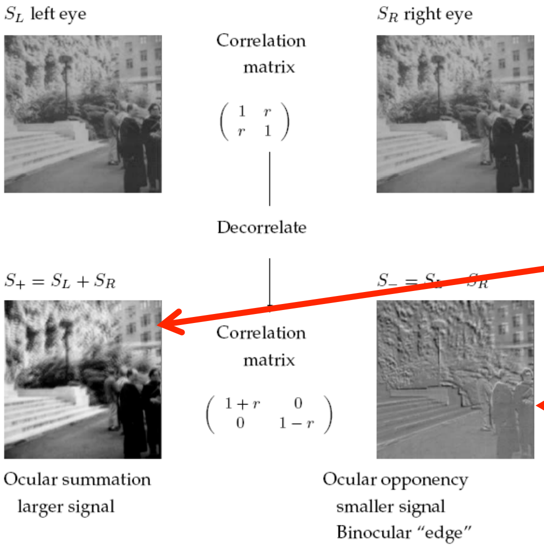


Figure from "Understanding vision: theory, models, and data", by Li Zhaoping, Oxford University Press, 2014

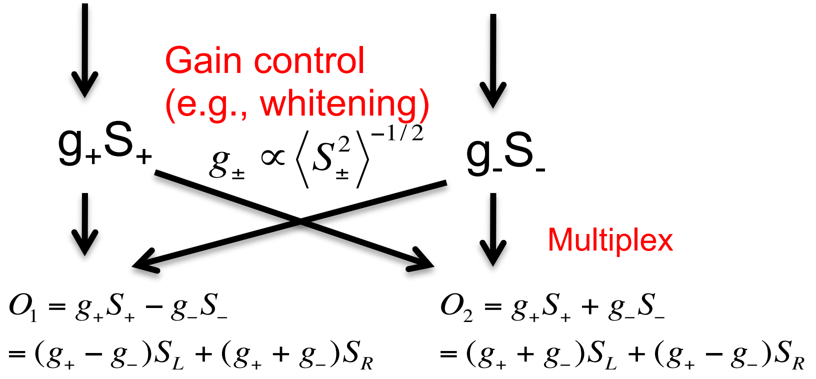
Efficient color coding as an analog to efficient stereo coding:



Analogy in color vision:

Luminance channel

Chromatic channel



$$O_1 = g_+S_+ - g_-S_- \\ = (g_+ - g_-)S_L + (g_+ + g_-)S_R$$

$$O_2 = g_+S_+ + g_-S_- \\ = (g_+ + g_-)S_L + (g_+ - g_-)S_R$$

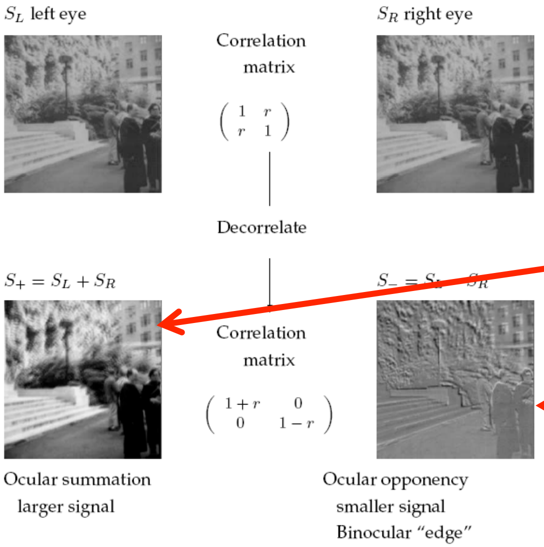
Left eye, Right eye
→ Red cone, Green Cone

Decorrelation:
→ (1) Luminance channel
→ (2) Chromatic channel

Gain control:
→ Higher sensitivity to chromatic channel (when signal to noise is high)

Multiplex:
→?

Efficient color coding as an analog to efficient stereo coding:



Analogy in color vision:

Luminance channel

Chromatic channel

Left eye, Right eye
→ Red, Green

or
→ Yellow, Blue

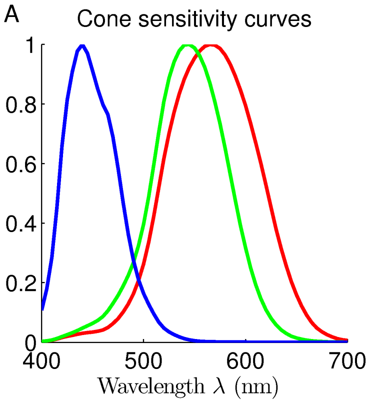
Gain control (e.g., whitening)

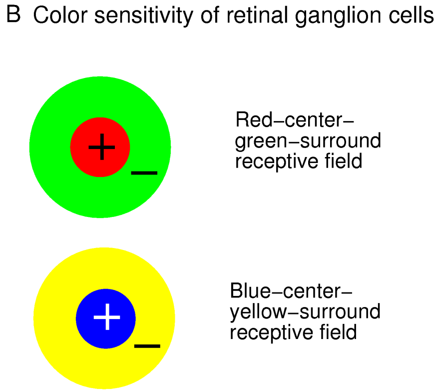
g_+S_+ $\quad g_\pm \propto \left\langle S_\pm^2 \right\rangle^{-1/2}$ $\quad g_-S_-$

Multiplex

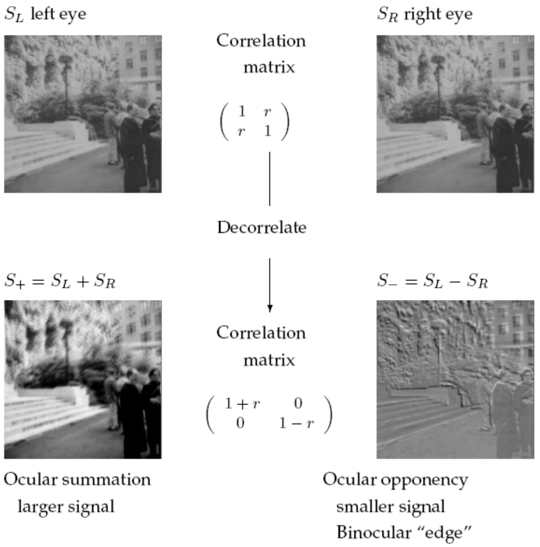
$$O_1 = g_+S_+ - g_-S_-$$
$$= (g_+ - g_-)S_L + (g_+ + g_-)S_R$$

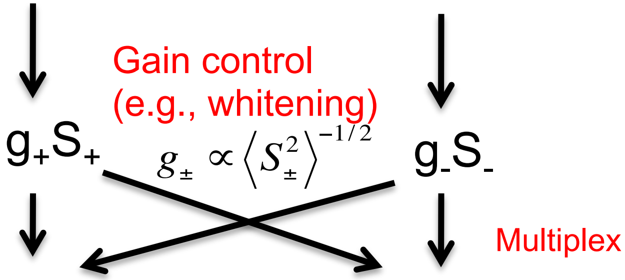
$$O_2 = g_+S_+ + g_-S_-$$
$$= (g_+ + g_-)S_L + (g_+ - g_-)S_R$$





Applying to efficient temporal coding by analogy:





$$O_1 = g_+S_+ - g_-S_- = (g_+ - g_-)S_L + (g_+ + g_-)S_R$$

$$O_2 = g_+S_+ + g_-S_- = (g_+ + g_-)S_L + (g_+ - g_-)S_R$$

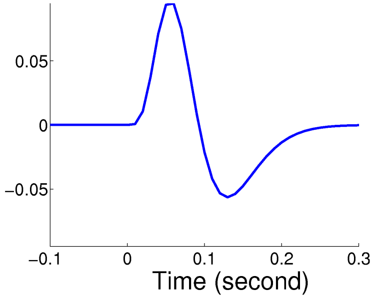
Left eye, Right eye
→ One time's input, next time's input

Decorrelation:
→ (1) Temporal summation/average
→ (2) temporal contrast

Gain control:
→ Higher sensitivity to transient than to sustained input (when signal to noise is high)

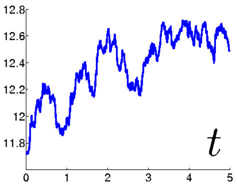
Multiplex: temporal filter

Impulse response function of a model neuron



More mathematically:

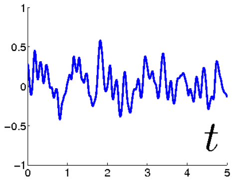
S_t + noise

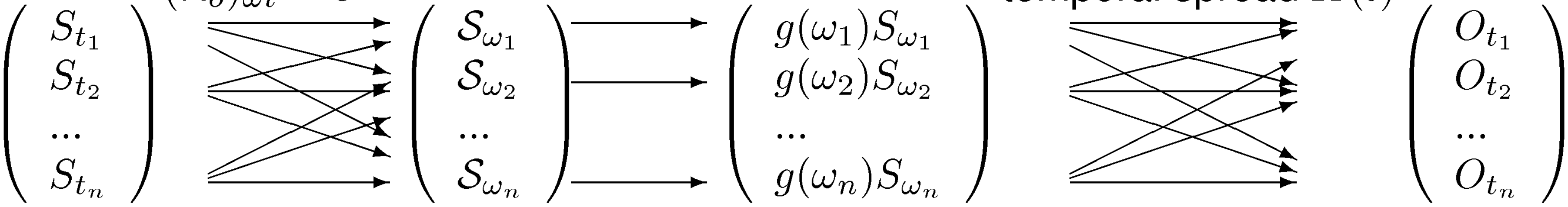


$$O_{t'} = \sum_t \mathrm{K}_{t't}(S_t + \text{noise}) \quad , \quad \text{with } \mathrm{K}_{t't} = (\mathrm{UgK}_o)_{t't}$$

$$= \sum_t K(t'-t)(S_t + \text{noise})$$

O_t





Fourier transform
$(\mathrm{K}_o)_{\omega t} \sim e^{-i\omega t}$

Gain
control g

$\mathrm{U}_{t\omega} \sim e^{i\omega t + i\phi(\omega)}$
for a causal, minimal
temporal spread $K(t)$

$$\begin{pmatrix} S_{t_1} \\ S_{t_2} \\ \ldots \\ S_{t_n} \end{pmatrix} \qquad \begin{pmatrix} \mathcal{S}_{\omega_1} \\ \mathcal{S}_{\omega_2} \\ \ldots \\ \mathcal{S}_{\omega_n} \end{pmatrix} \qquad \begin{pmatrix} g(\omega_1)\mathcal{S}_{\omega_1} \\ g(\omega_2)\mathcal{S}_{\omega_2} \\ \ldots \\ g(\omega_n)\mathcal{S}_{\omega_n} \end{pmatrix} \qquad \begin{pmatrix} O_{t_1} \\ O_{t_2} \\ \ldots \\ O_{t_n} \end{pmatrix}$$

Temporal
inputs S_t

Temporal Fourier
amplitudes $\mathcal{S}_\omega$

Gain
controlled $\mathcal{S}_\omega$

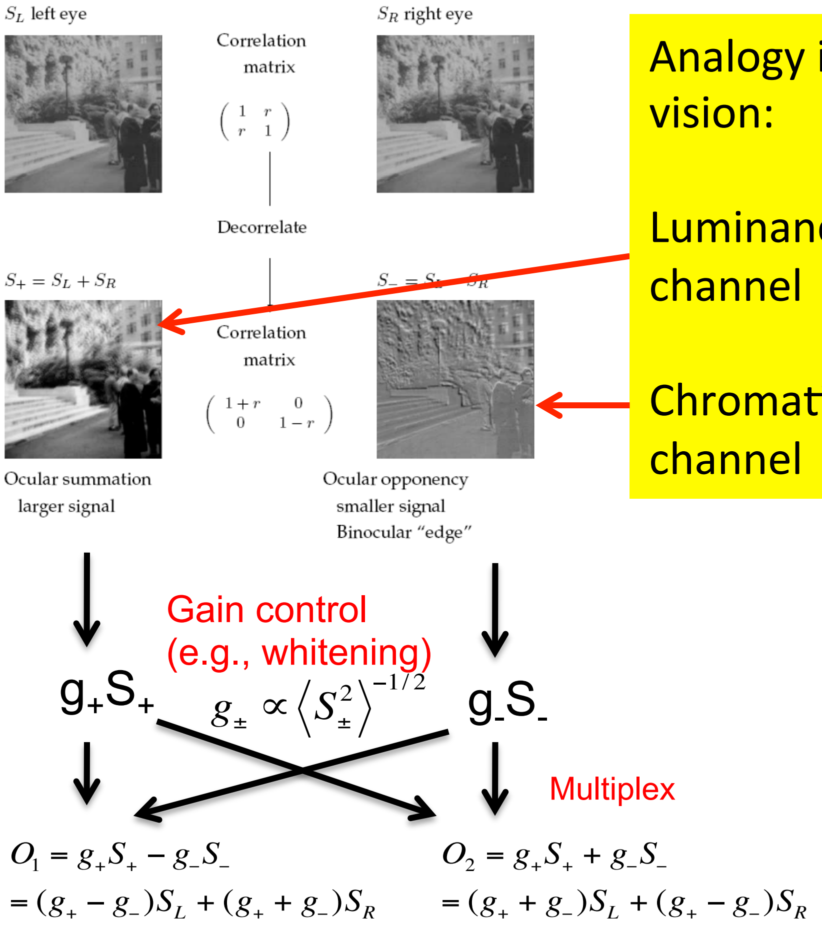
Temporal
output O_t

Figure from "Understanding vision: theory, models, and data", by Li Zhaoping, Oxford University Press, 2014

Interactions between visual coding in different sensory dimensions: space, time, stereo, color

See the book for details.

Adaptation of efficient coding to input signal-to-noise



Gain control (e.g., whitening)

g_+S_+

$g_\pm \propto \left\langle S_\pm^2 \right\rangle^{-1/2}$

g_-S_-

Multiplex

$$O_1 = g_+S_+ - g_-S_- \\ = (g_+ - g_-)S_L + (g_+ + g_-)S_R$$

$$O_2 = g_+S_+ + g_-S_- \\ = (g_+ + g_-)S_L + (g_+ - g_-)S_R$$

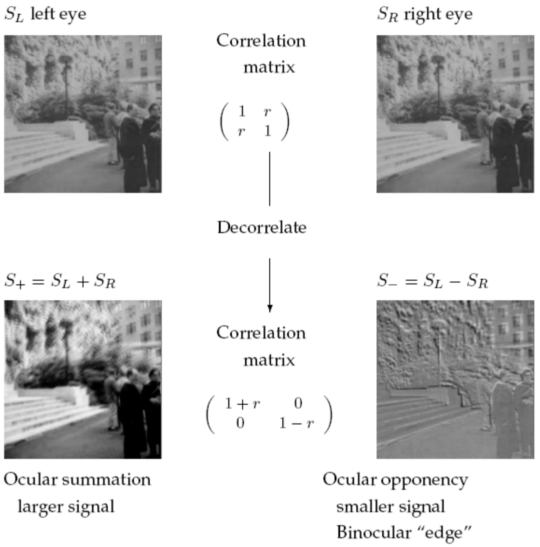
Analogy in color vision:

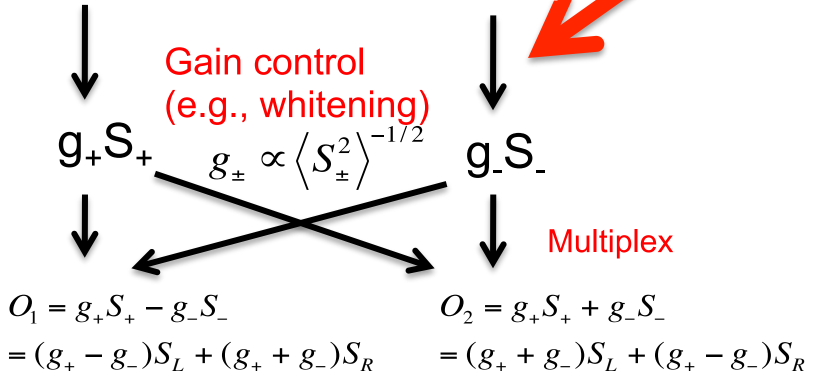
Luminance channel

Chromatic channel

In dim light, sensitivity to color drops, i.e., g_- drops

Adaptation of efficient coding to input signal-to-noise





Gain control
(e.g., whitening)

g_+S_+

$g_\pm \propto \left\langle S_\pm^2 \right\rangle^{-1/2}$

g_-S_-

Multiplex

$$O_1 = g_+S_+ - g_-S_-$$
$$= (g_+ - g_-)S_L + (g_+ + g_-)S_R$$

$$O_2 = g_+S_+ + g_-S_-$$
$$= (g_+ + g_-)S_L + (g_+ - g_-)S_R$$

Why?

Gain control:

(1) Whitening is only for zero noise input

(2) When input noise is high, give lower gain

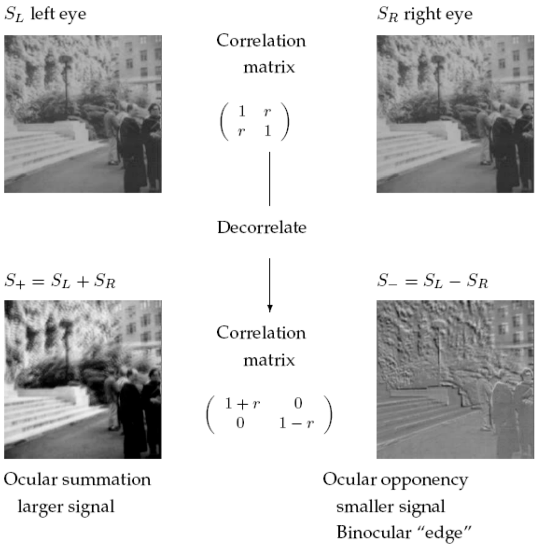
General solution

$$g_k^2 \propto \text{Max}\left\{[\frac{1}{2}\frac{\langle S_k^2\rangle}{\langle S_k^2\rangle+\langle N^2\rangle}(1+\sqrt{1+\frac{4\lambda}{(\ln 2)\langle N_o^2\rangle}\frac{\langle N^2\rangle}{\langle S_k^2\rangle}})-1],0\right\}$$

In dim light, sensitivity to color drops i.e., g_- drops

sensitivity to stereo depth also drops

Adaptation of efficient coding to input signal-to-noise



S_L left eye

S_R right eye

Correlation matrix

$$\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$$

Decorrelate

$S_+ = S_L + S_R$

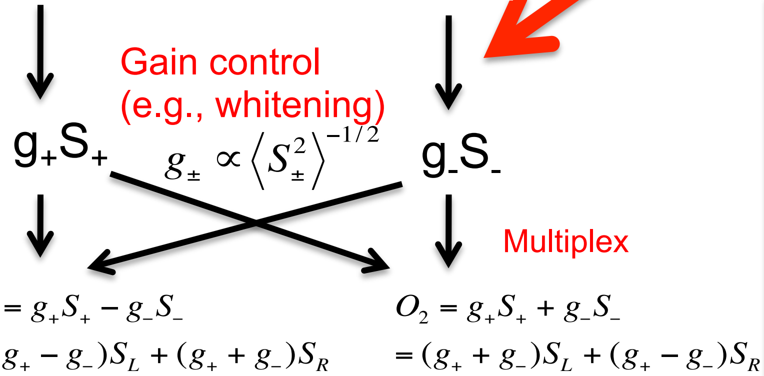
$S_- = S_L - S_R$

Correlation matrix

$$\begin{pmatrix} 1+r & 0 \\ 0 & 1-r \end{pmatrix}$$

Ocular summation
larger signal

Ocular opponency
smaller signal
Binocular "edge"



g_+S_+

Gain control
(e.g., whitening)

$g_\pm \propto \left\langle S_\pm^2 \right\rangle^{-1/2}$

g_-S_-

Multiplex

$O_1 = g_+S_+ - g_-S_-$
$= (g_+ - g_-)S_L + (g_+ + g_-)S_R$

$O_2 = g_+S_+ + g_-S_-$
$= (g_+ + g_-)S_L + (g_+ - g_-)S_R$

Why?

Gain control:

(1) Whitening is only for zero noise input

(2) When input noise is high, give lower gain

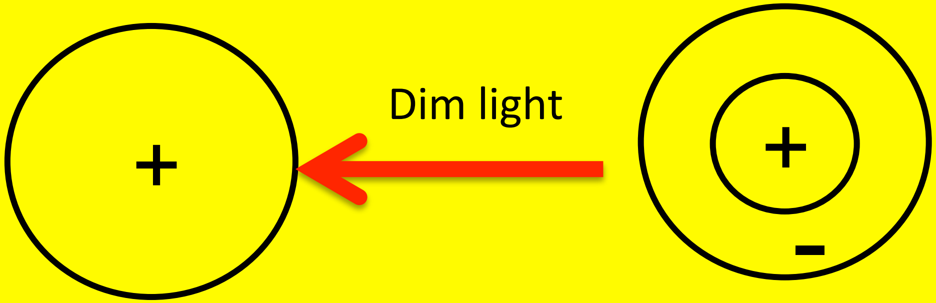
General solution

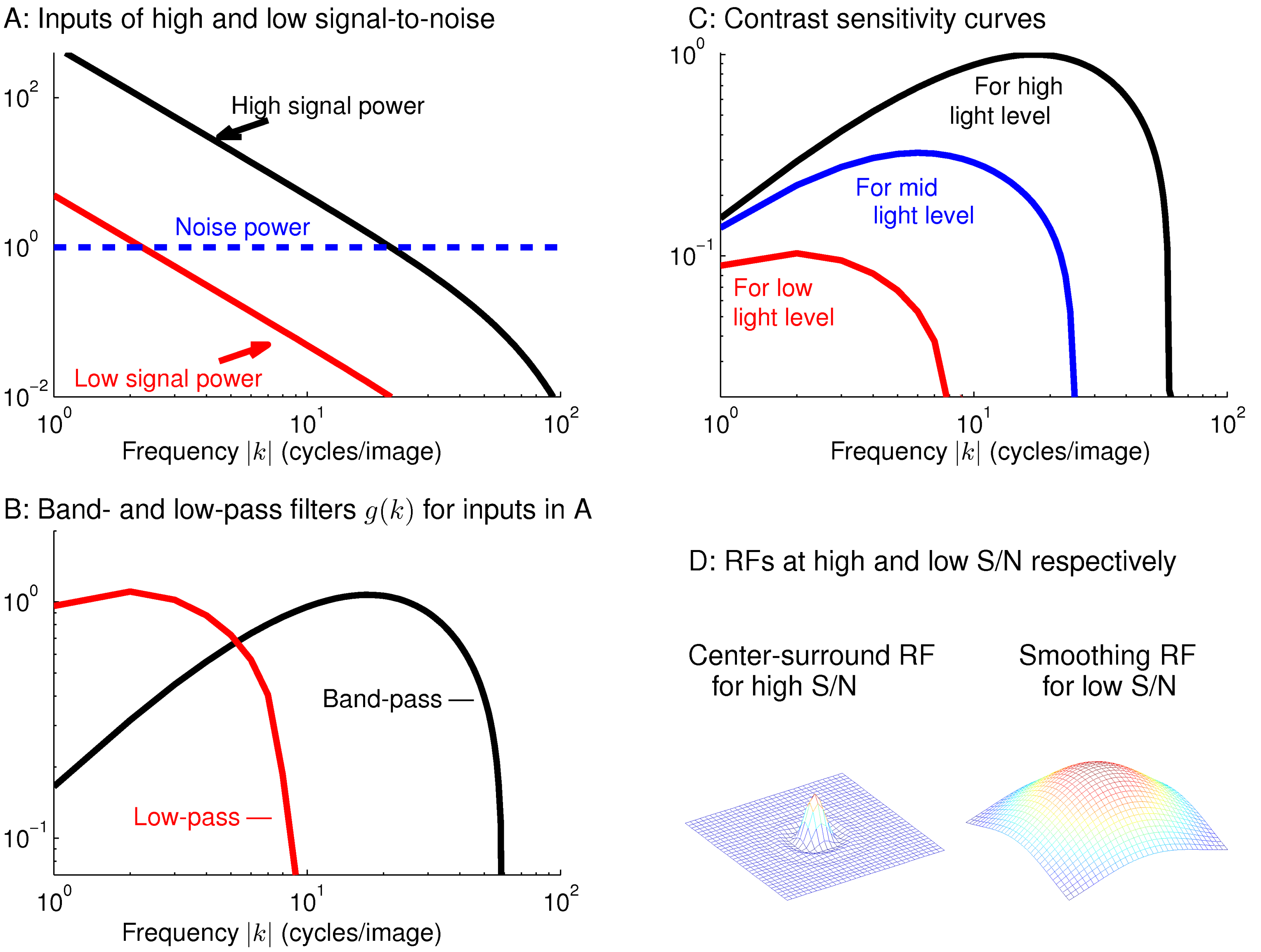
$$g_k^2 \propto \text{Max}\left\{[\frac{1}{2}\frac{\langle S_k^2\rangle}{\langle S_k^2\rangle + \langle N^2\rangle}(1+\sqrt{1+\frac{4\lambda}{(\ln 2)\langle N_o^2\rangle}\frac{\langle N^2\rangle}{\langle S_k^2\rangle}})-1], 0\right\}$$

Apply to spatial coding

Gain control:
→ Higher sensitivity to spatial contrast
(when signal to noise is high)

Multiplex:
→Center surround receptive-field





A: Inputs of high and low signal-to-noise

B: Band- and low-pass filters $g(k)$ for inputs in A

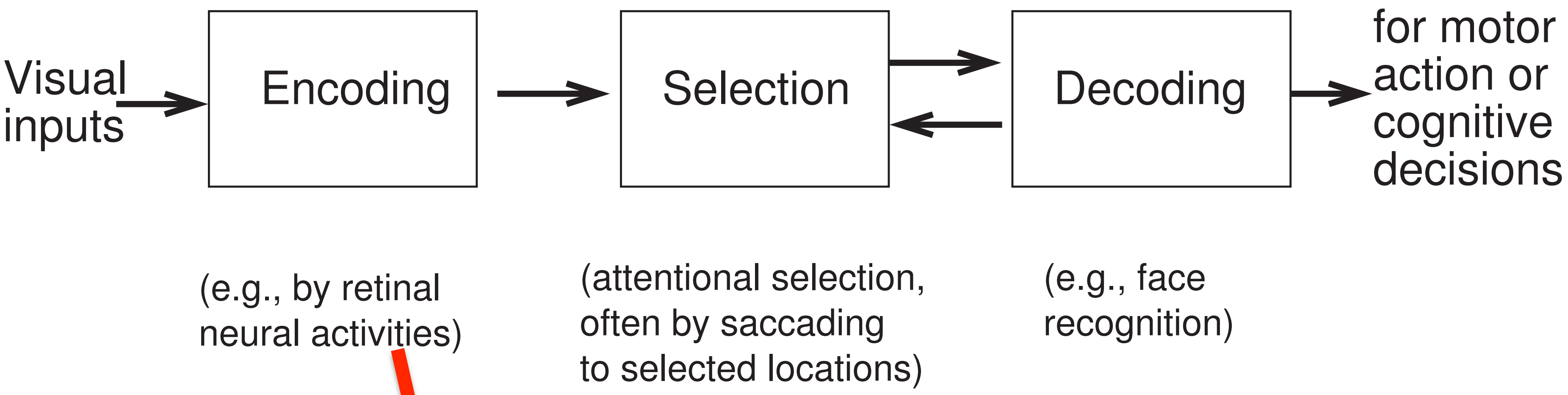
C: Contrast sensitivity curves

D: RFs at high and low S/N respectively

Figure from "Understanding vision: theory, models, and data", by Li Zhaoping, Oxford University Press, 2014

Summary:

(see Zhaoping 2014)



Visual inputs → Encoding → Selection ⇄ Decoding → for motor action or cognitive decisions

(e.g., by retinal neural activities)

(attentional selection, often by saccading to selected locations)

(e.g., face recognition)

Efficient coding for early vision:

Encoding depends on input statistics

Use it to understand (design) receptive fields optimally

Understand and predict experimental data.