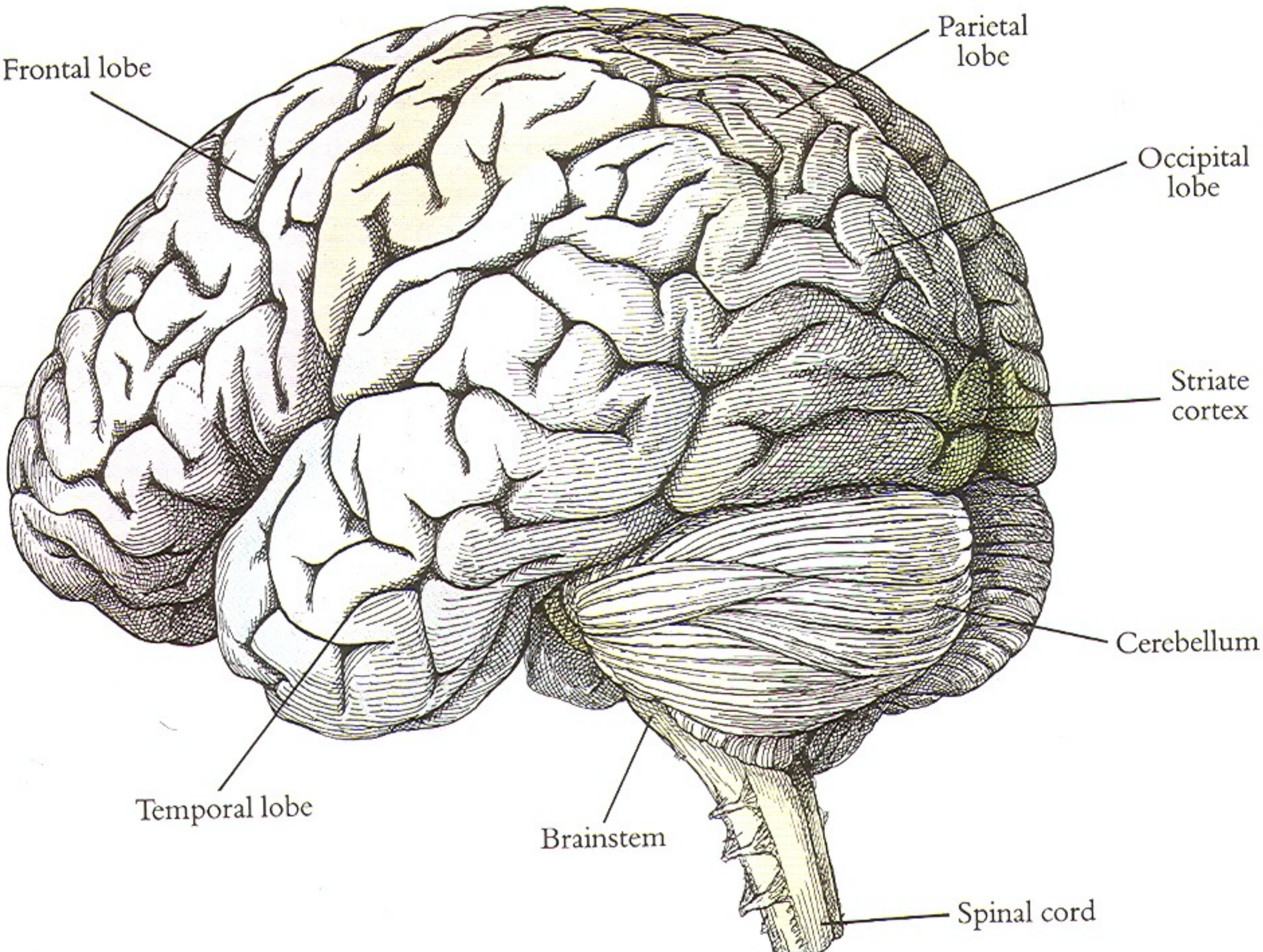
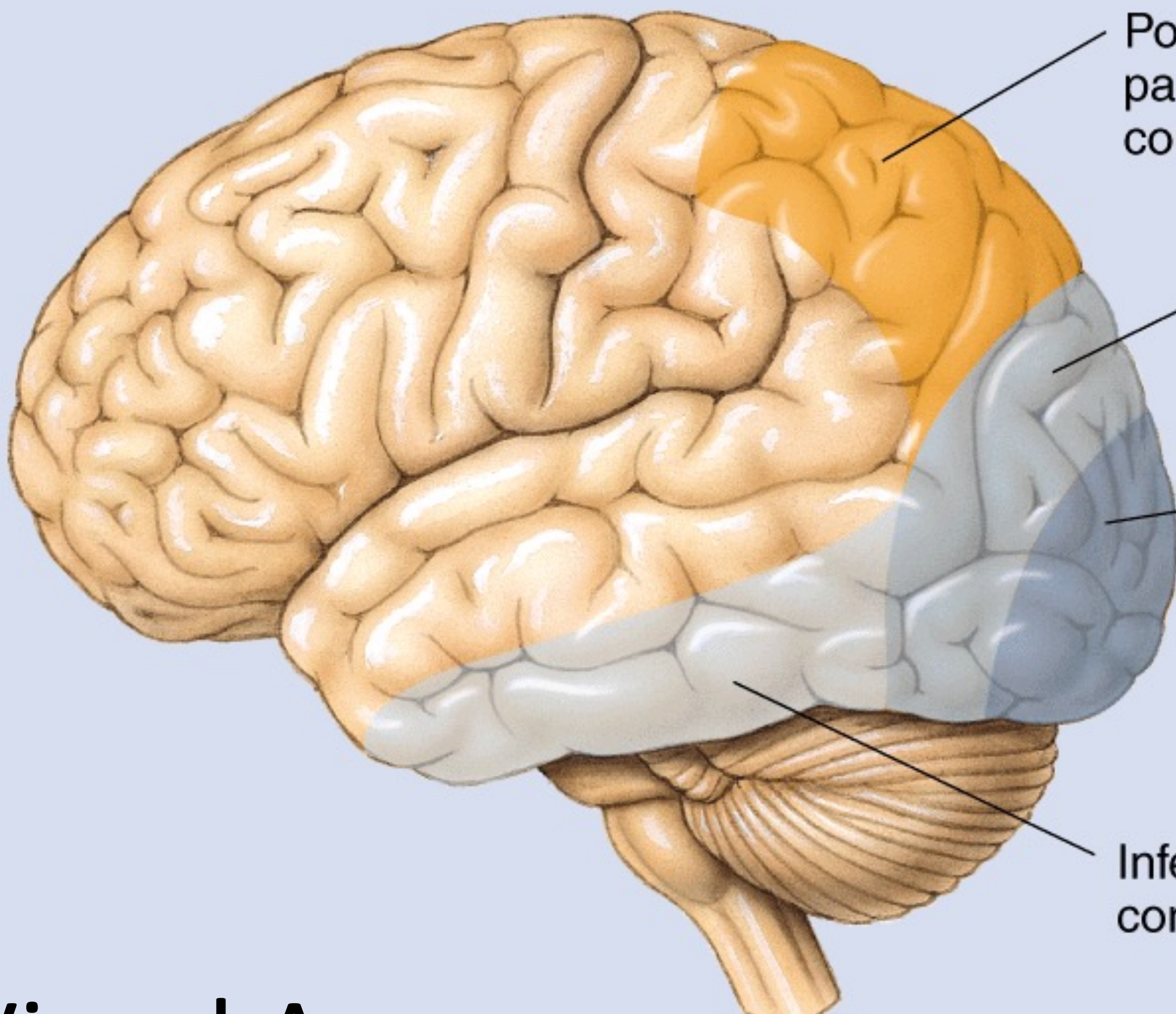
Human vision

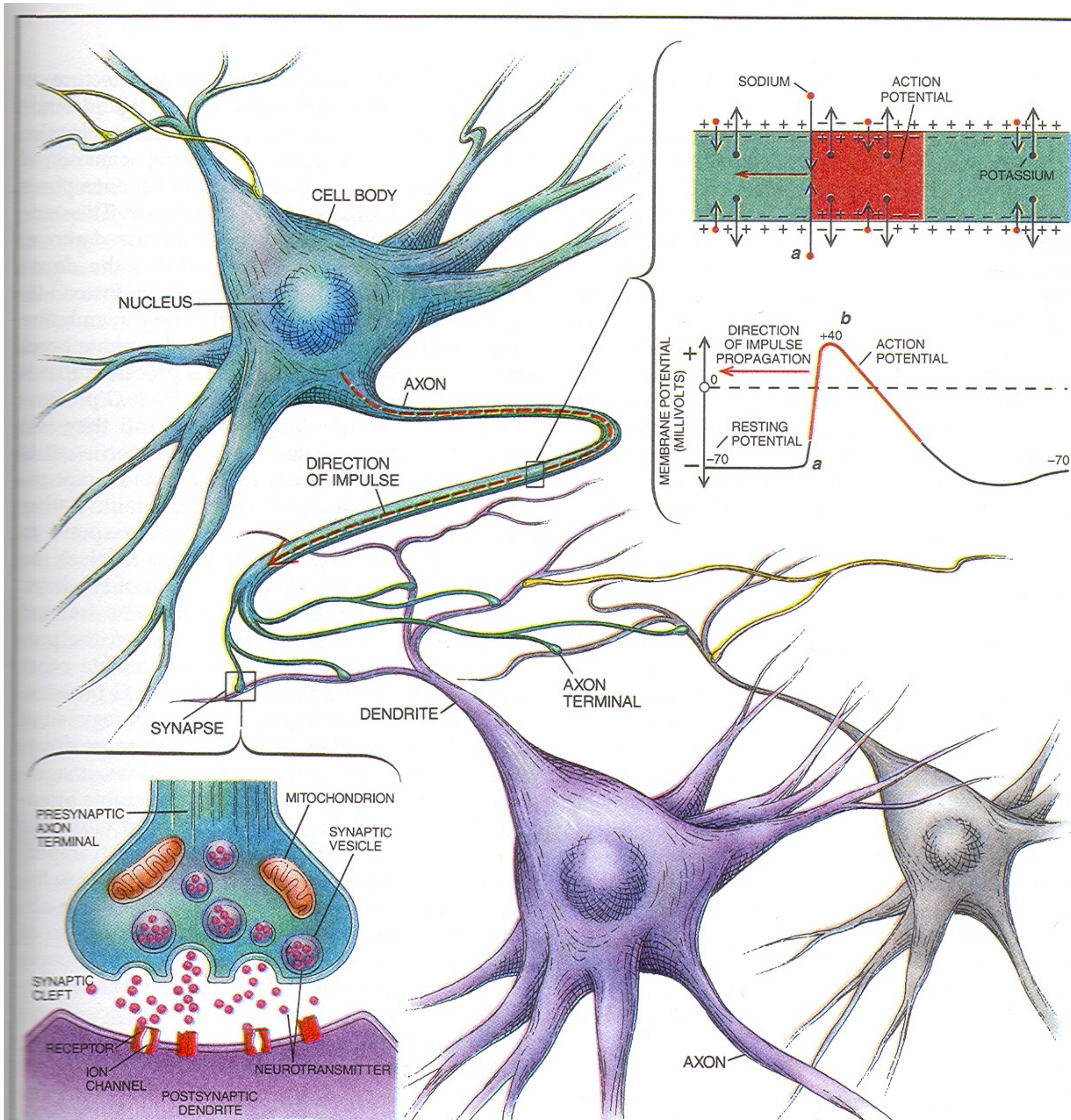
Jitendra Malik
U.C. Berkeley


Frontal lobe
Parietal lobe
Occipital lobe
Striate cortex
Cerebellum
Temporal lobe
Brainstem
Spinal cord

► **Visual Areas of the Human Cerebral Cortex**



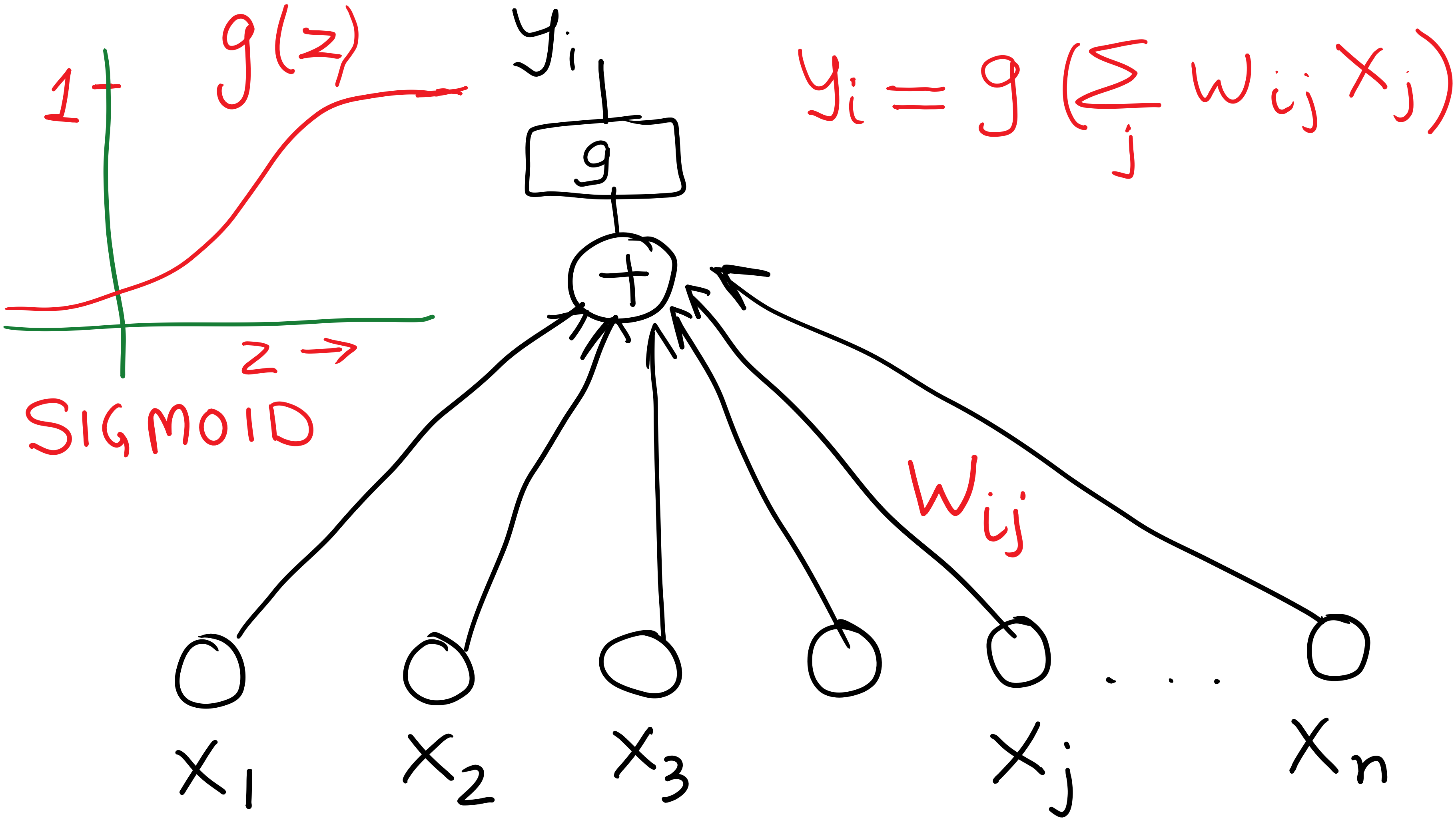
Visual Areas

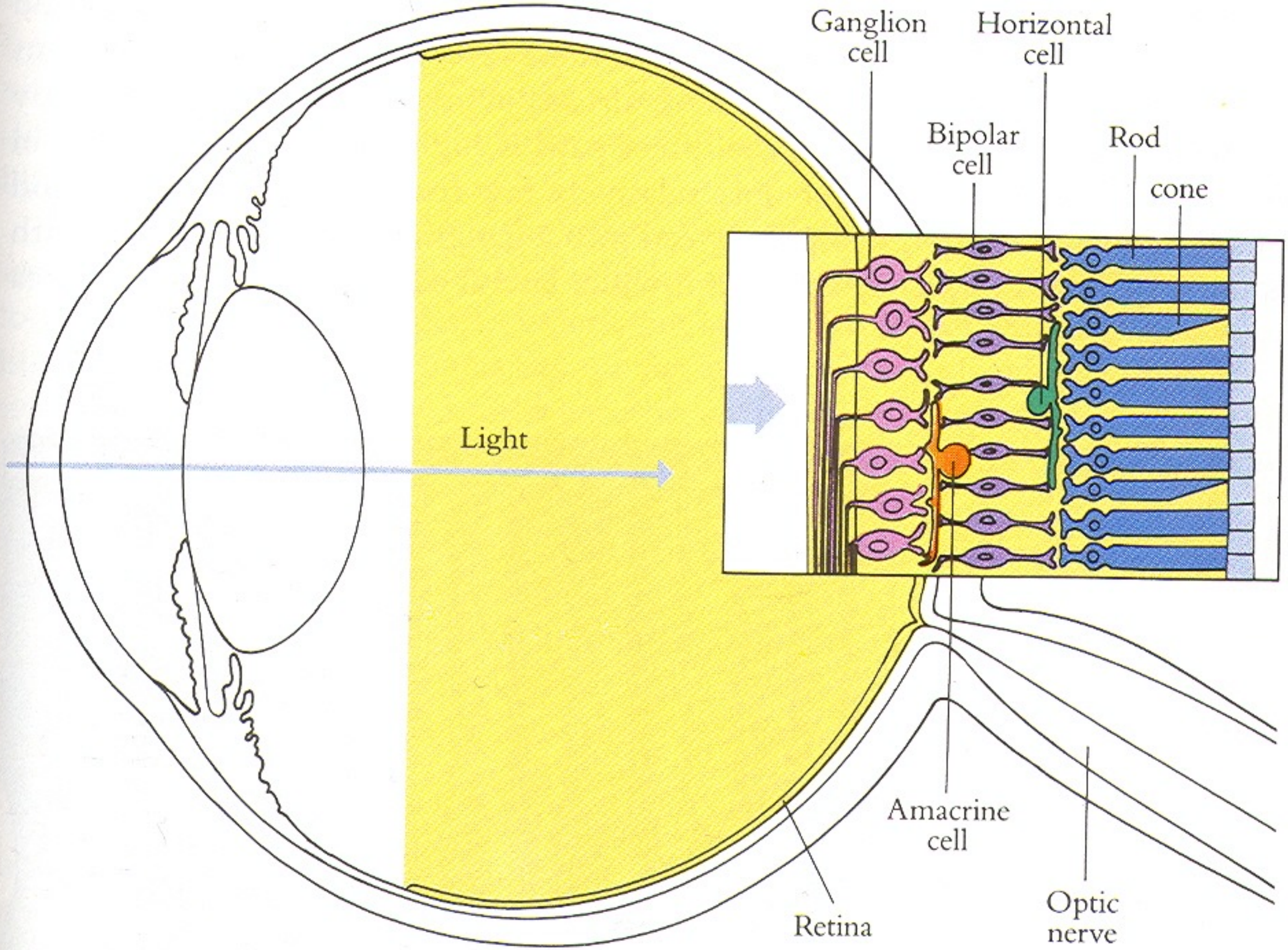


How Neurons Communicate

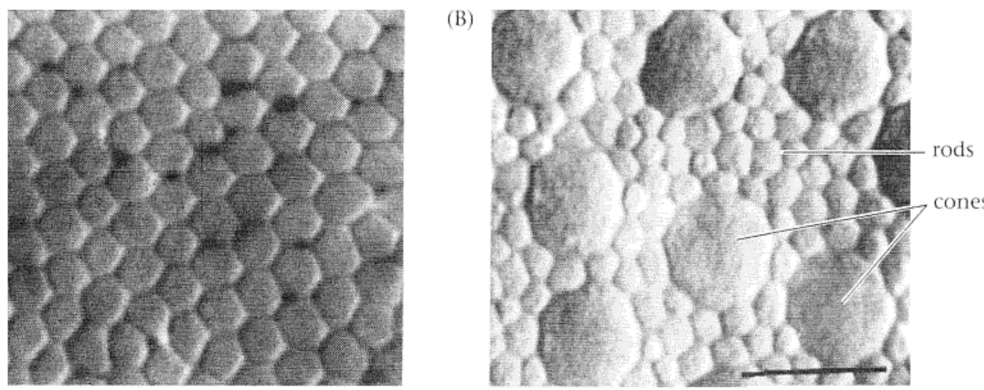
Mathematical Abstraction

$$y_i = g\left(\sum_j w_{ij} x_j\right)$$

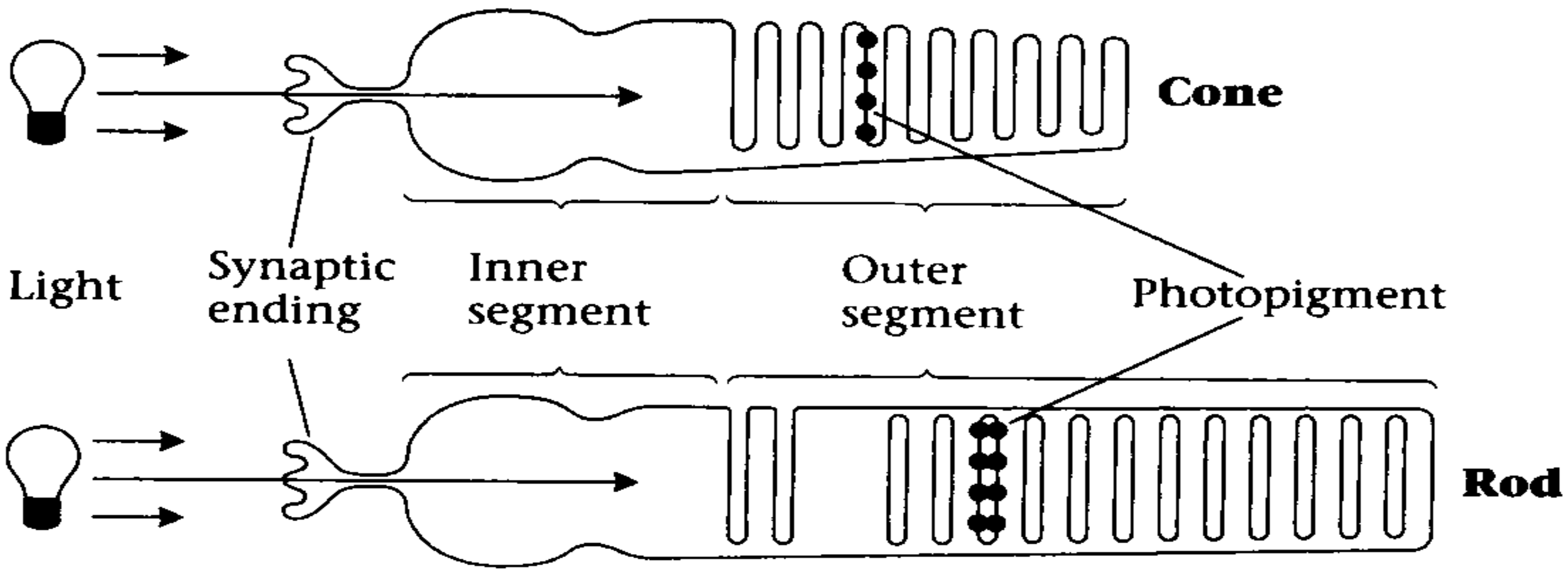



Ganglion cell
Horizontal cell
Bipolar cell
Rod
cone
Light
Amacrine cell
Retina
Optic nerve

The photoreceptor mosaic: rods and cones are the eye’s pixels

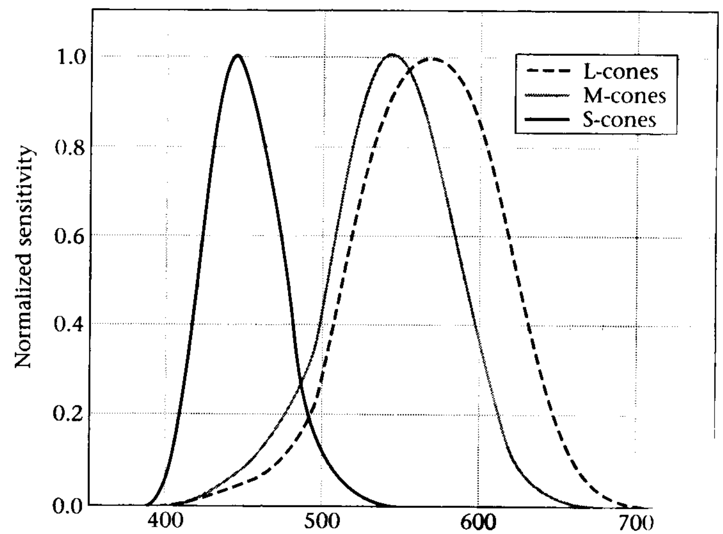


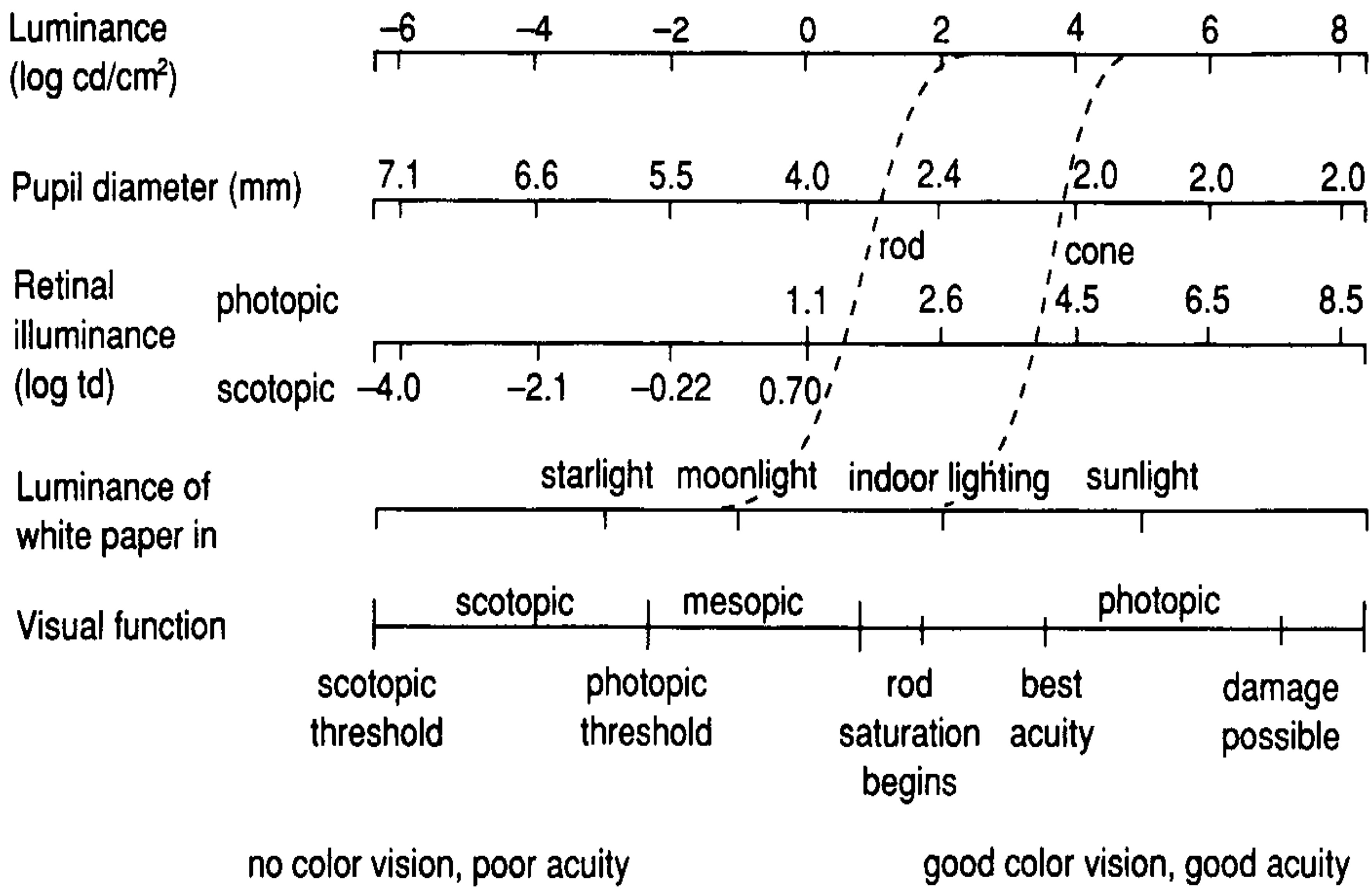
Cones and Rods

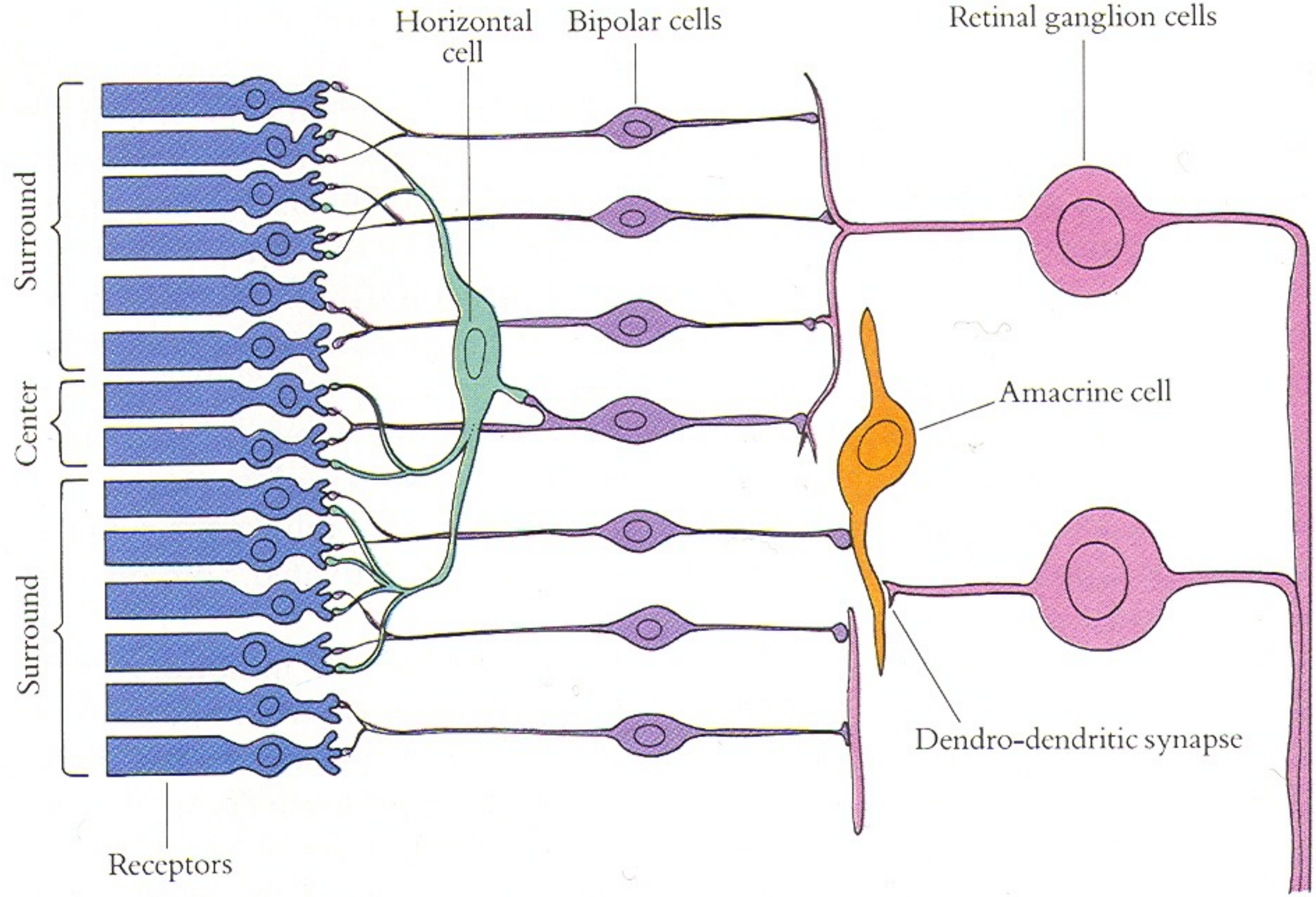


After dark adaptation, a single rod can respond to a single photon

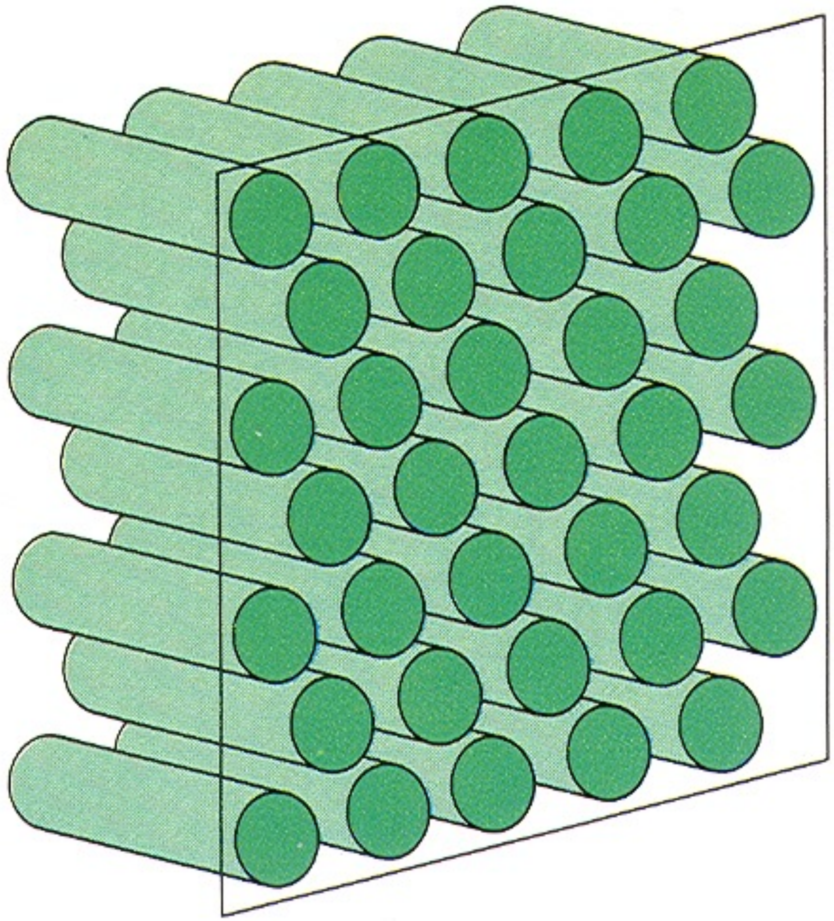
The three cone types have different spectral sensitivity functions



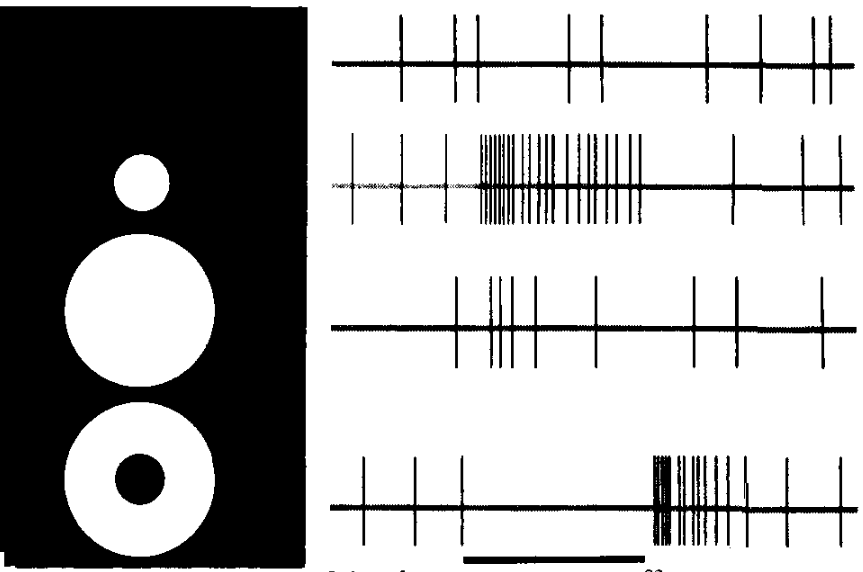

Luminance (log cd/cm²)
–6 –4 –2 0 2 4 6 8
Pupil diameter (mm)
7.1 6.6 5.5 4.0 2.4 2.0 2.0 2.0
rod
cone
Retinal illuminance (log td)
photopic
1.1 2.6 4.5 6.5 8.5
scotopic
–4.0 –2.1 –0.22 0.70
Luminance of white paper in
starlight moonlight indoor lighting sunlight
Visual function
scotopic mesopic photopic
scotopic threshold
photopic threshold
rod saturation begins
best acuity
damage possible
no color vision, poor acuity
good color vision, good acuity

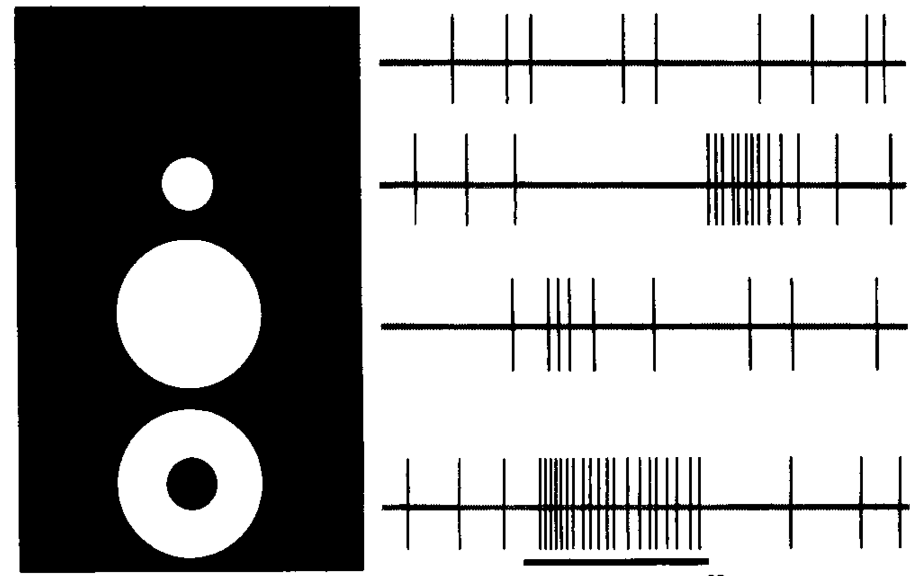

Horizontal cell
Bipolar cells
Retinal ganglion cells
Surround
Center
Surround
Amacrine cell
Dendro-dendritic synapse
Receptors

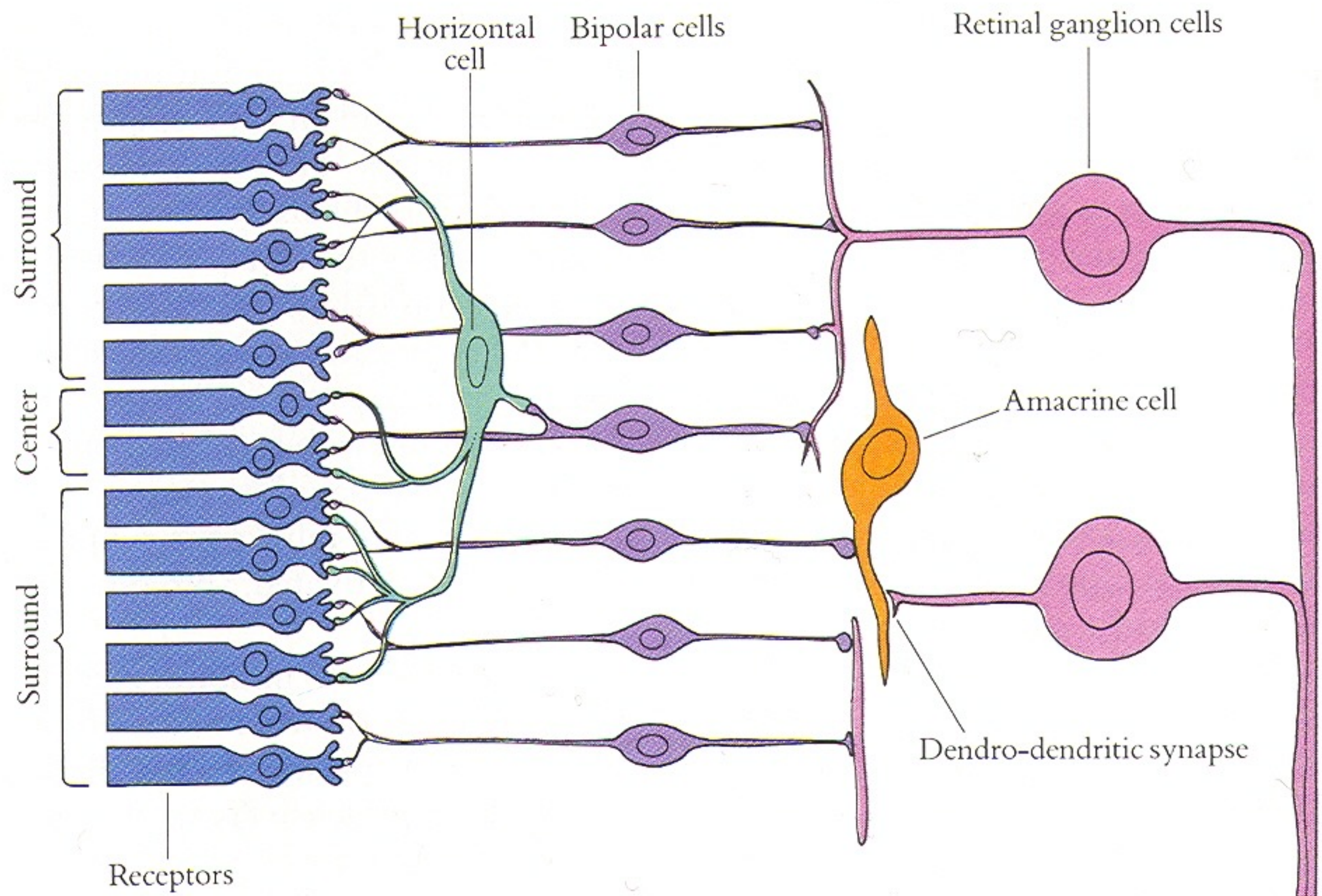
Stage 1
(rods and cones)



ON and OFF cells in retinal ganglia






Horizontal cell
Bipolar cells
Retinal ganglion cells
Surround
Center
Surround
Amacrine cell
Dendro-dendritic synapse
Receptors

Receptive Fields

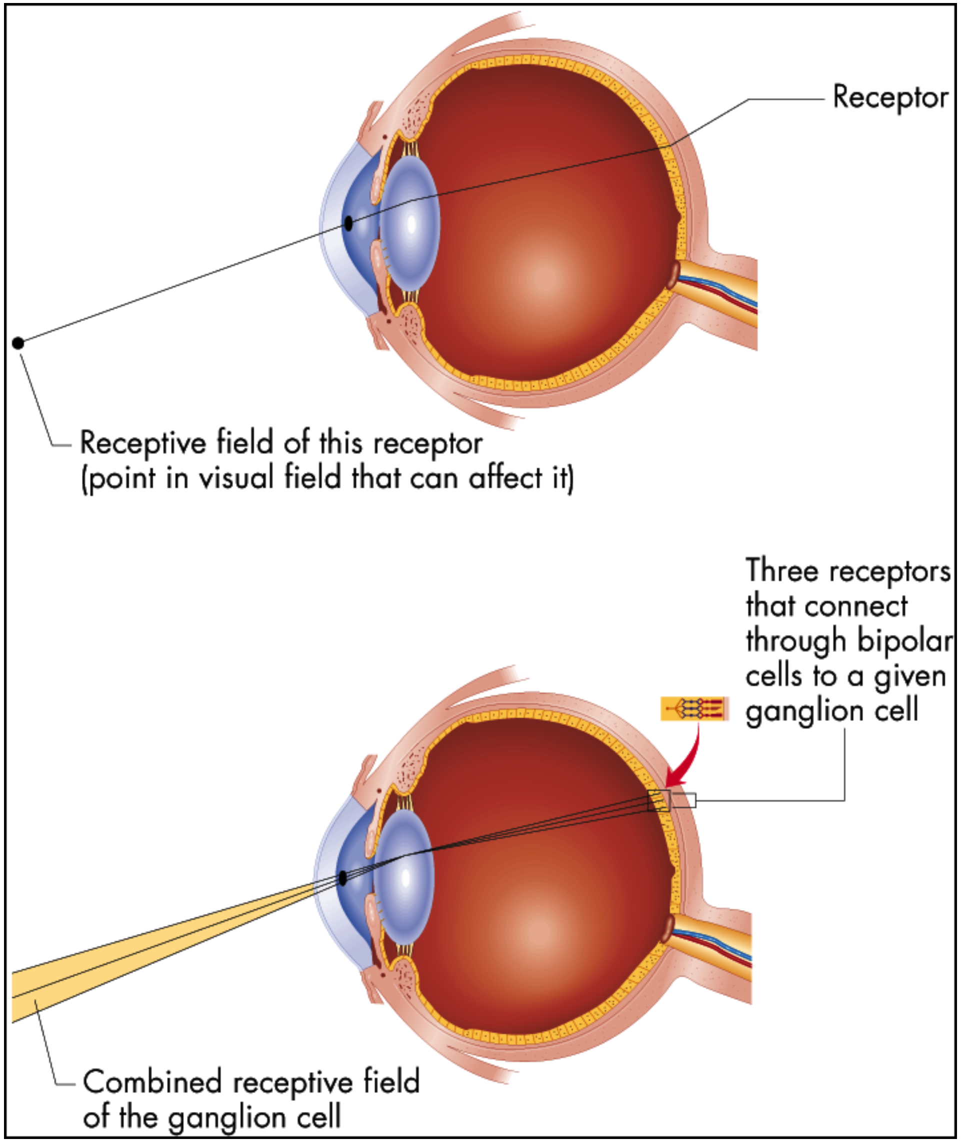
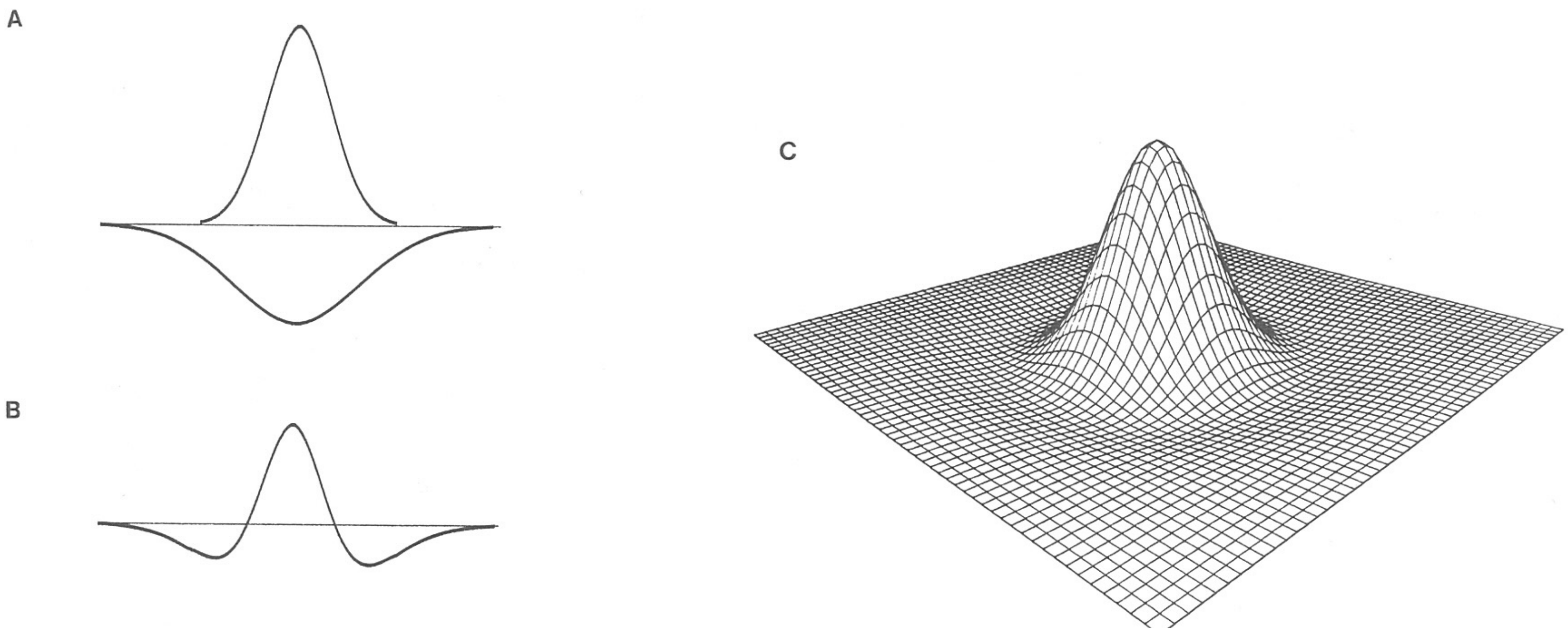


Figure 6.16 Receptive fields
The receptive field of a receptor is simply the area of the visual field from which light strikes that receptor. For any other cell in the visual system, the receptive field is determined by which receptors connect to the cell in question.

The receptive field of a retinal ganglion cell can be modeled as a "Difference of Gaussians"



$$G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}}$$

Convolving an image with a filter

10	20	20	20
10	20	20	20
10	20	20	20
10	20	20	20

$$* \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}$$

Result is a new array

Convolution is implemented by "flip and drag". Here let us flip

$$\begin{bmatrix} -1 & 0 & 1 \end{bmatrix}$$

Each output unit gets the weighted sum of image pixels

$\overset{1}{10}$	$\overset{0}{20}$	$\overset{-1}{20}$	20
10	20	20	20
10	20	20	20
10	20	20	20

multiply pointwise and add

$$1 \times 10 + 0 \times 20 - 1 \times 20$$

$$= -10$$

	−10		

Each output unit gets the weighted sum of input units

	1	0	-1
10	20	20	20
10	20	20	20
10	20	20	20
10	20	20	20

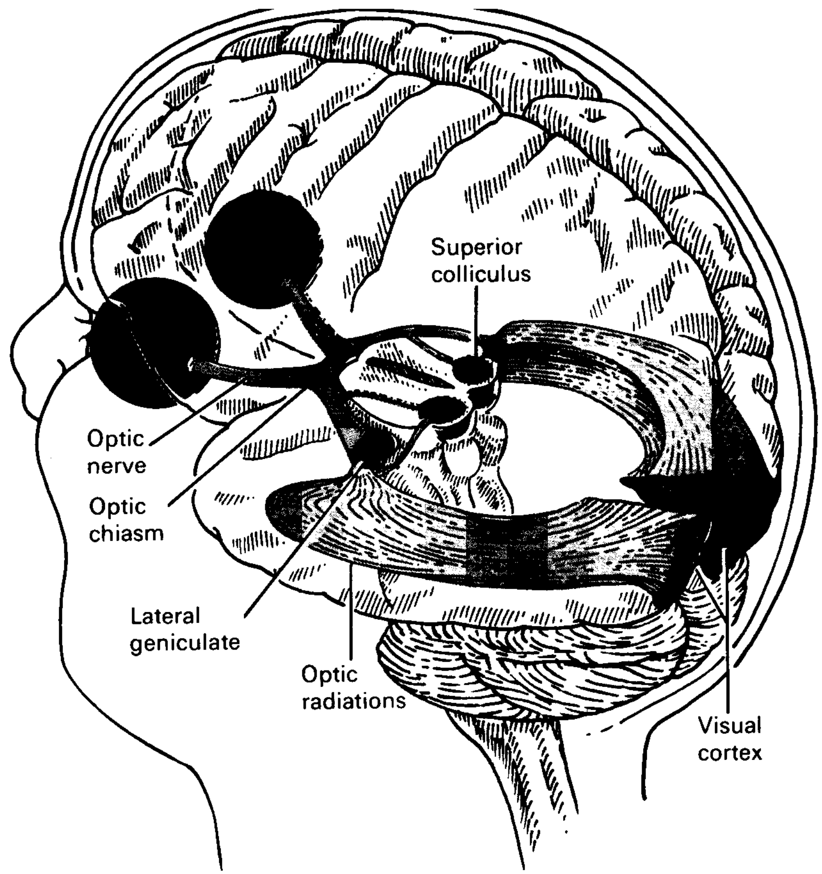
Slide mask & repeat

$$1 \times 20 + 0 \times 20 - 1 \times 20 = 0$$

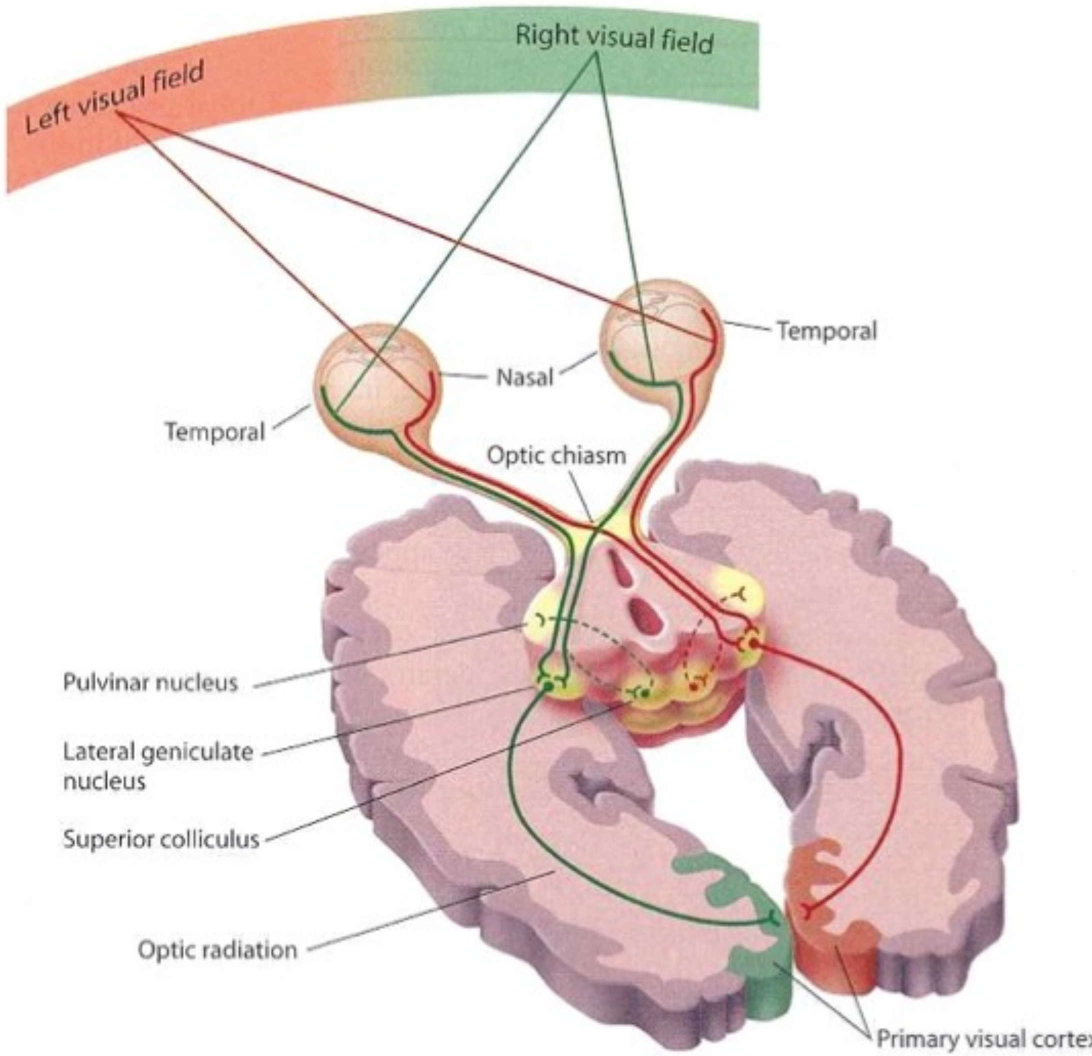
	-10	0	

and so on..

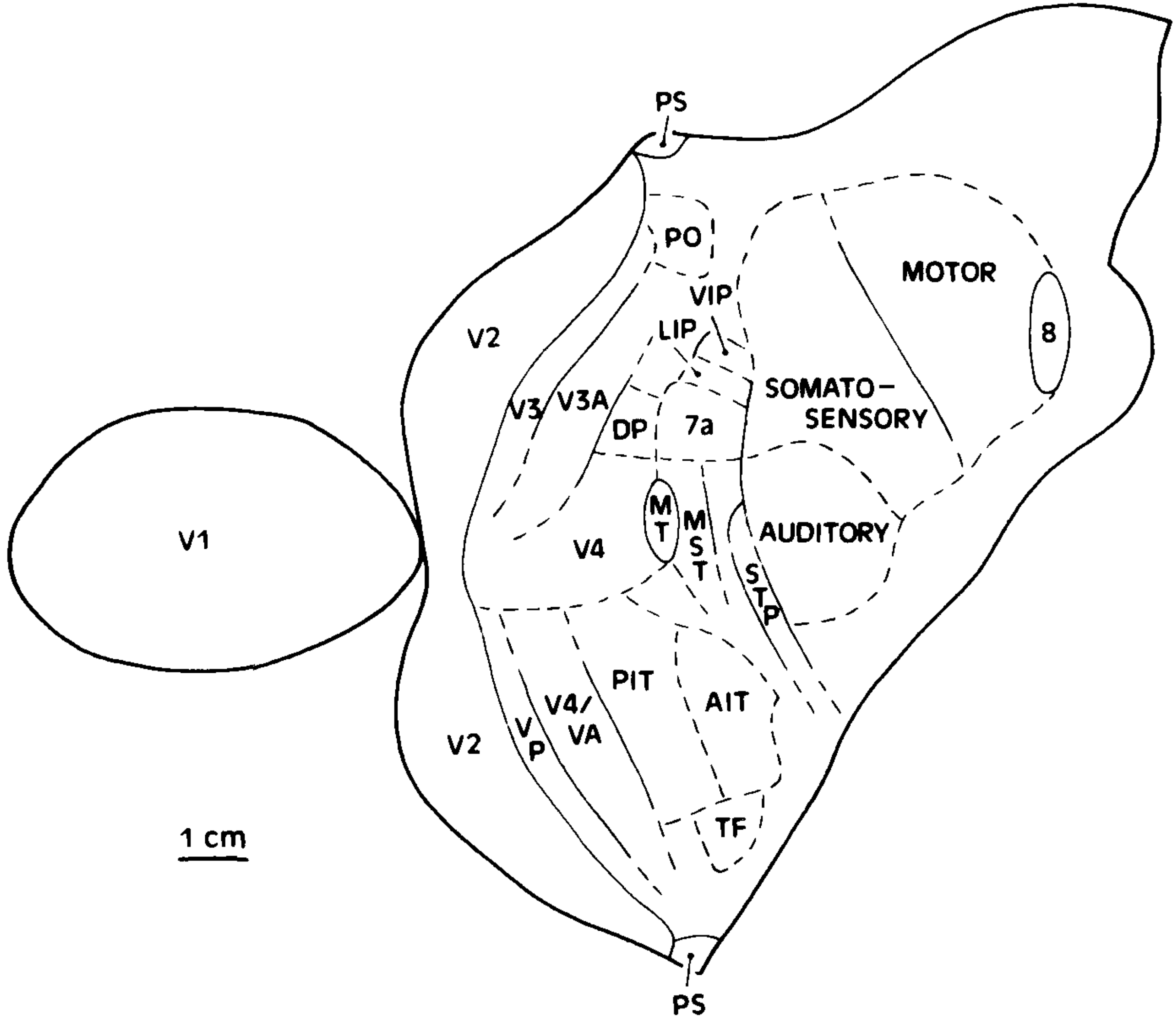
We can think of this weighting function as the receptive field of the output unit

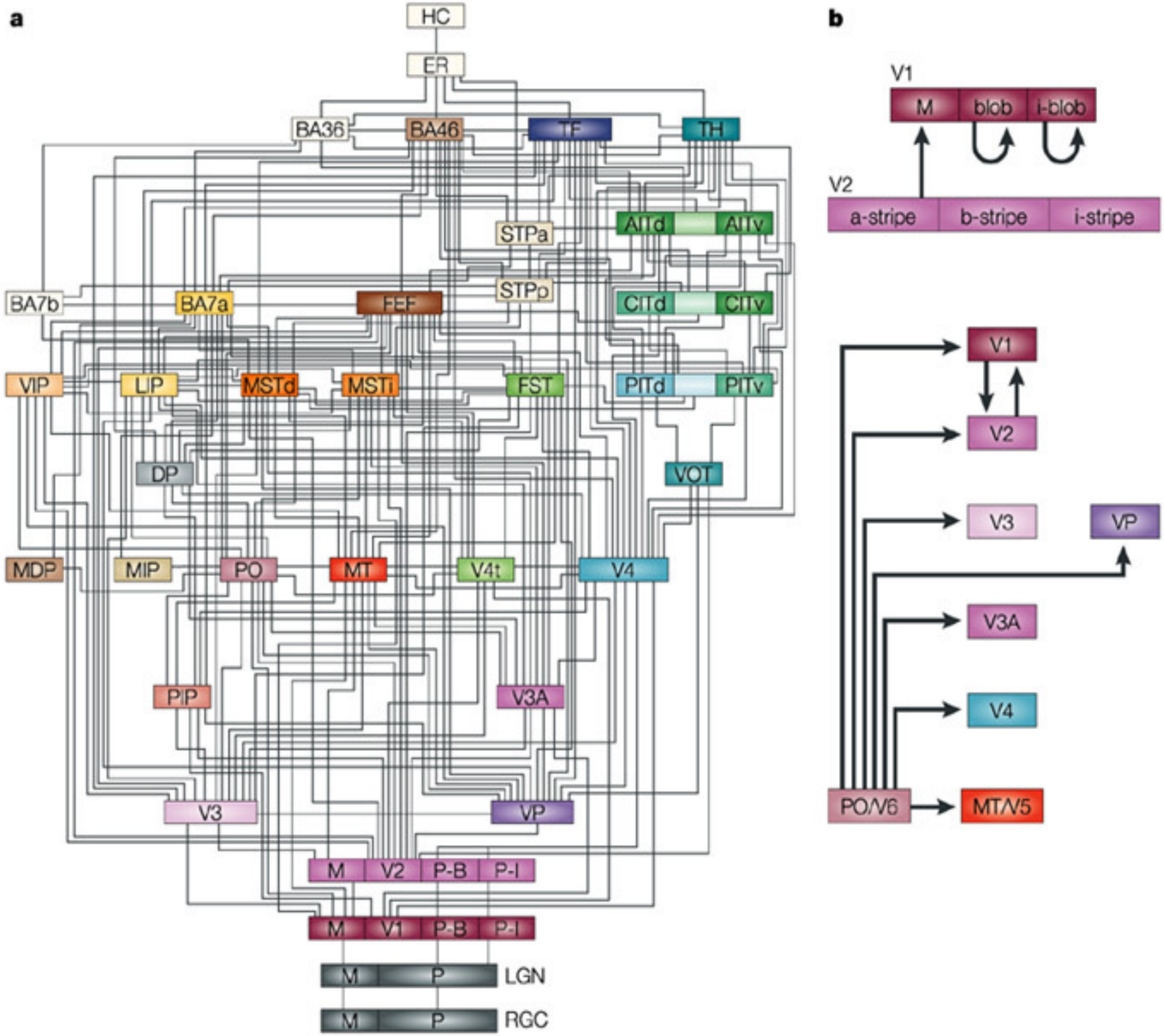

Superior colliculus
Optic nerve
Optic chiasm
Lateral geniculate
Optic radiations
Visual cortex

Anatomy of Pathway to Visual Cortex



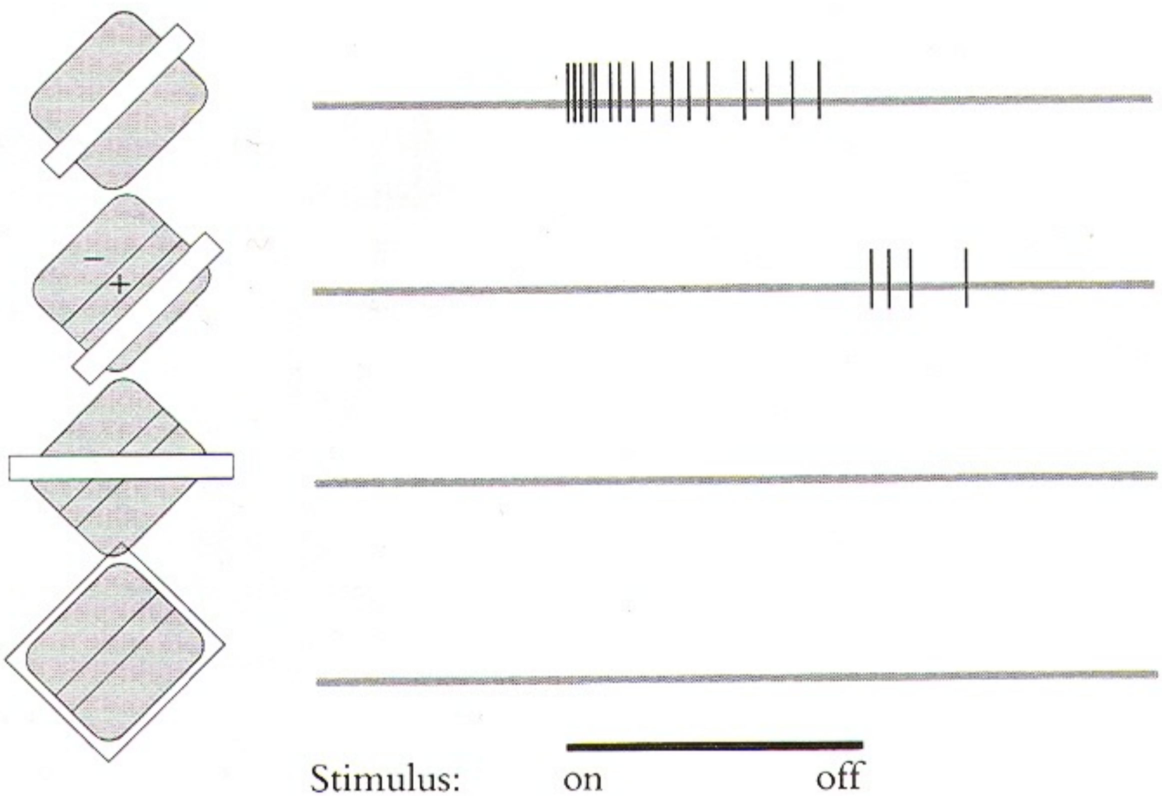
Visual Processing Areas



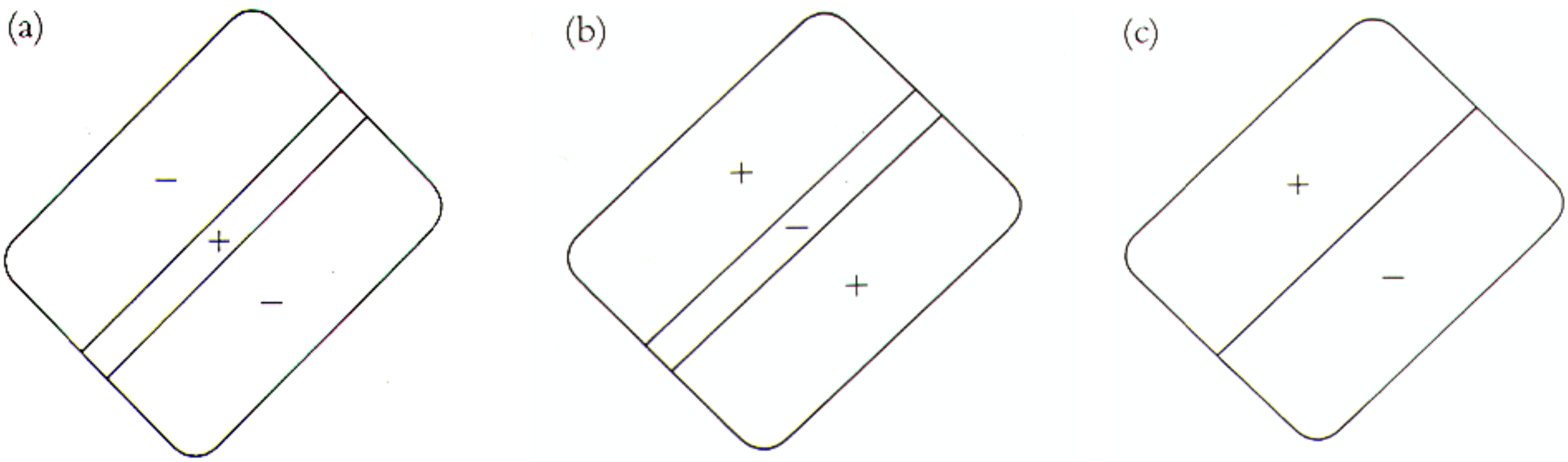


Nature Reviews | Neuroscience

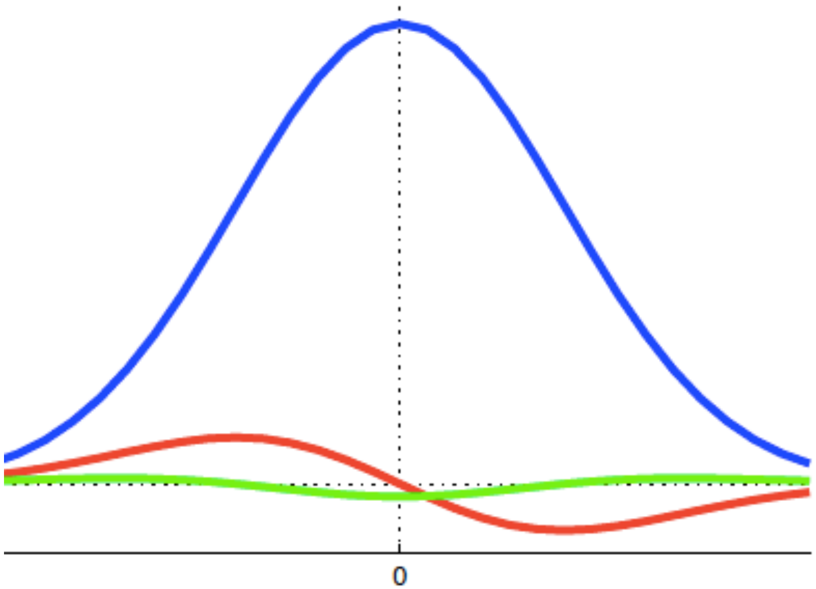
Orientation Selectivity in V1



Receptive fields of simple cells (discovered by Hubel & Wiesel)



The 1D Gaussian and its derivatives



$$G_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

$$G'_\sigma(x) = \frac{d}{dx} G_\sigma(x) = -\frac{1}{\sigma}\left(\frac{x}{\sigma}\right) G_\sigma(x)$$

$$G''_\sigma(x) = \frac{d^2}{dx^2} G_\sigma(x) = \frac{1}{\sigma^2}\left(\frac{x^2}{\sigma^2} - 1\right) G_\sigma(x)$$

$G'_\sigma(x)$'s maxima/minima occur at $G''_\sigma(x)$'s zeros. And, we can see that $G'_\sigma(x)$ is an odd symmetric function and $G''_\sigma(x)$ is an even symmetric function.

Oriented Gaussian Derivatives in 2D

$$f_1(x,y) = G'_{\sigma_1}(x)G_{\sigma_2}(y) \tag{10.4}$$
$$f_2(x,y) = G''_{\sigma_1}(x)G_{\sigma_2}(y) \tag{10.5}$$

We also consider rotated versions of these Gaussian derivative functions.

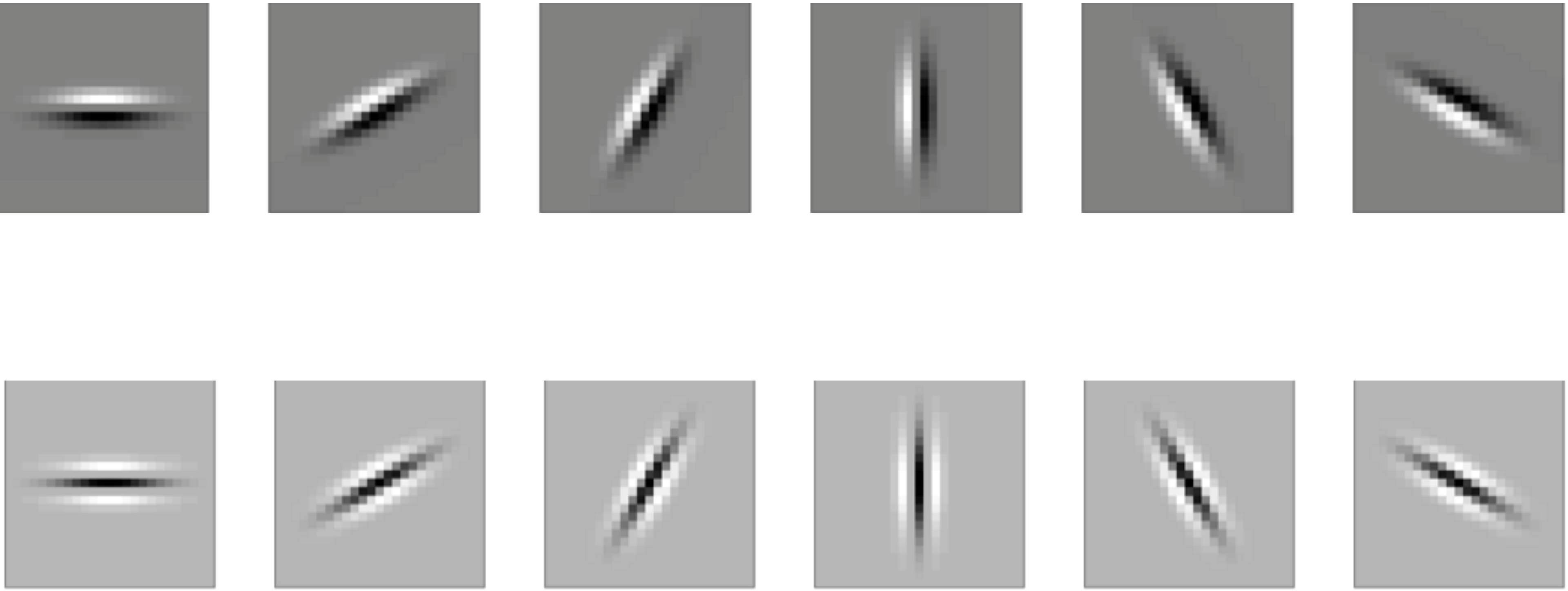
$$Rot_\theta f_1 = G'_{\sigma_1}(u)G_{\sigma_2}(v) \tag{10.6}$$
$$Rot_\theta f_2 = G''_{\sigma_1}(u)G_{\sigma_2}(v) \tag{10.7}$$

where we set

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

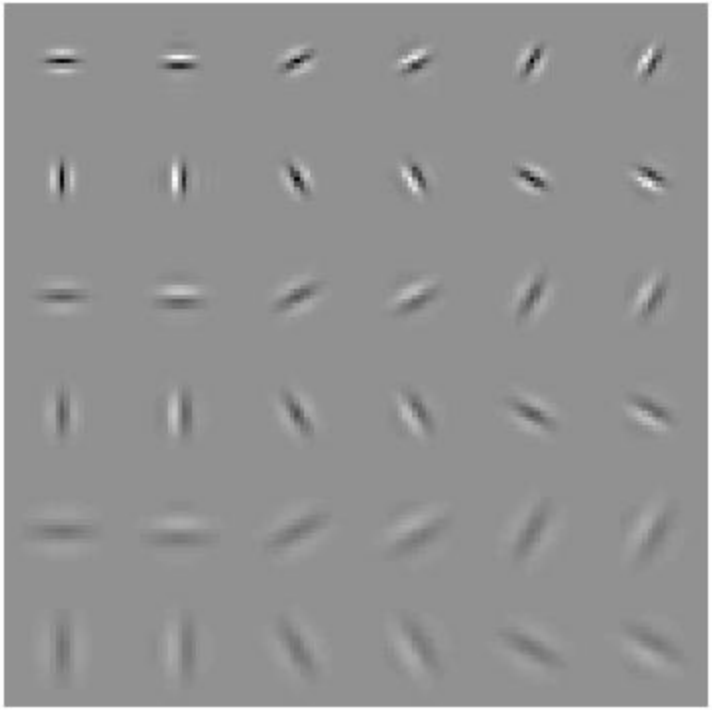
These are useful when we convolve with 2D images, e.g. to detect edges at different orientations.

Oriented Gaussian First and Second Derivatives

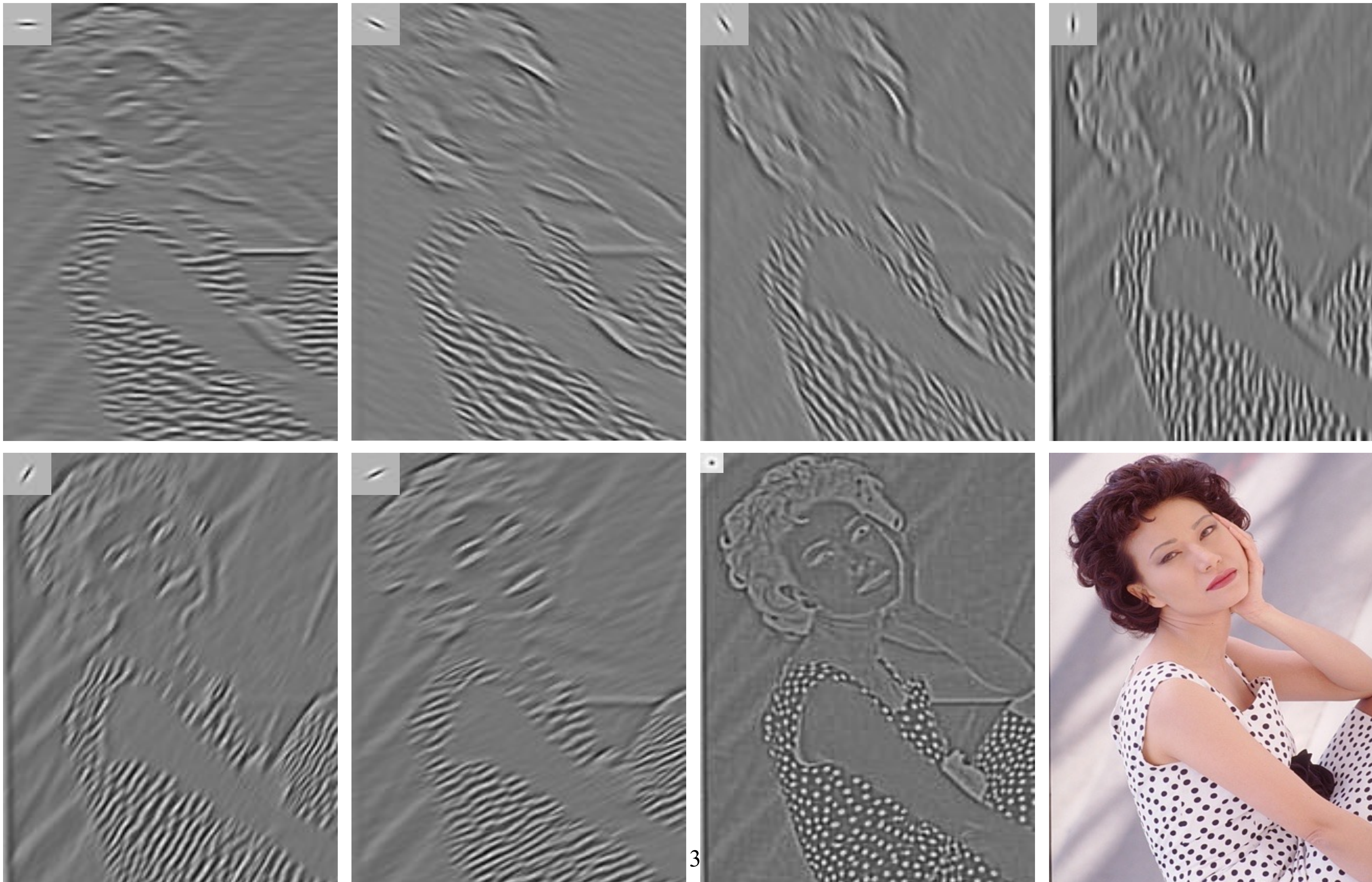


Modeling simple cells

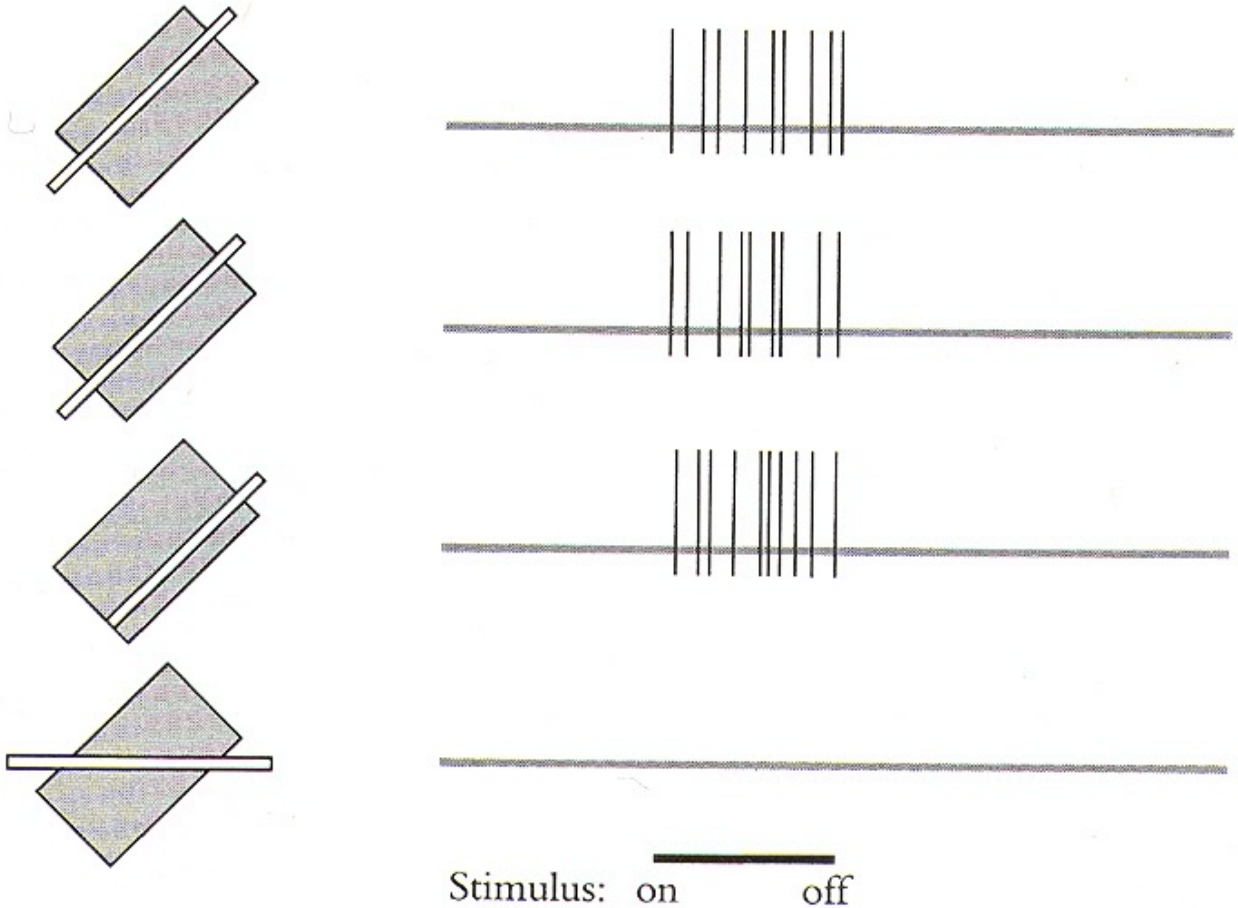
- Elongated directional Gaussian derivatives
- Gabor filters could be used instead
- Multiple orientations, scales



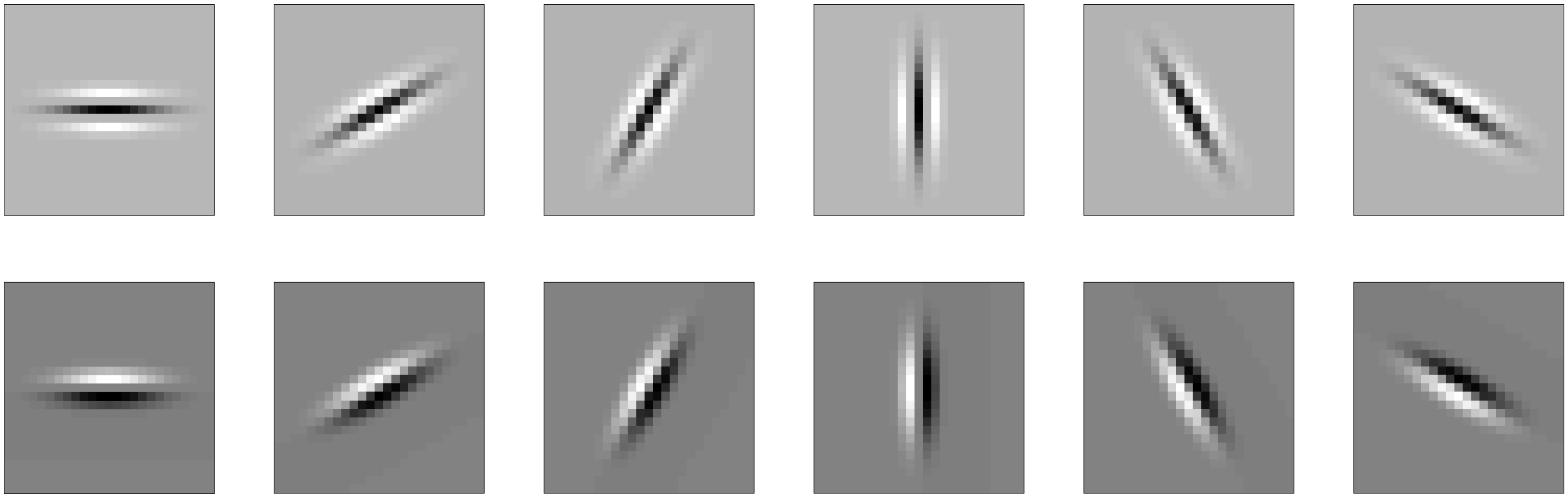
Filter Outputs



Receptive fields of complex cells

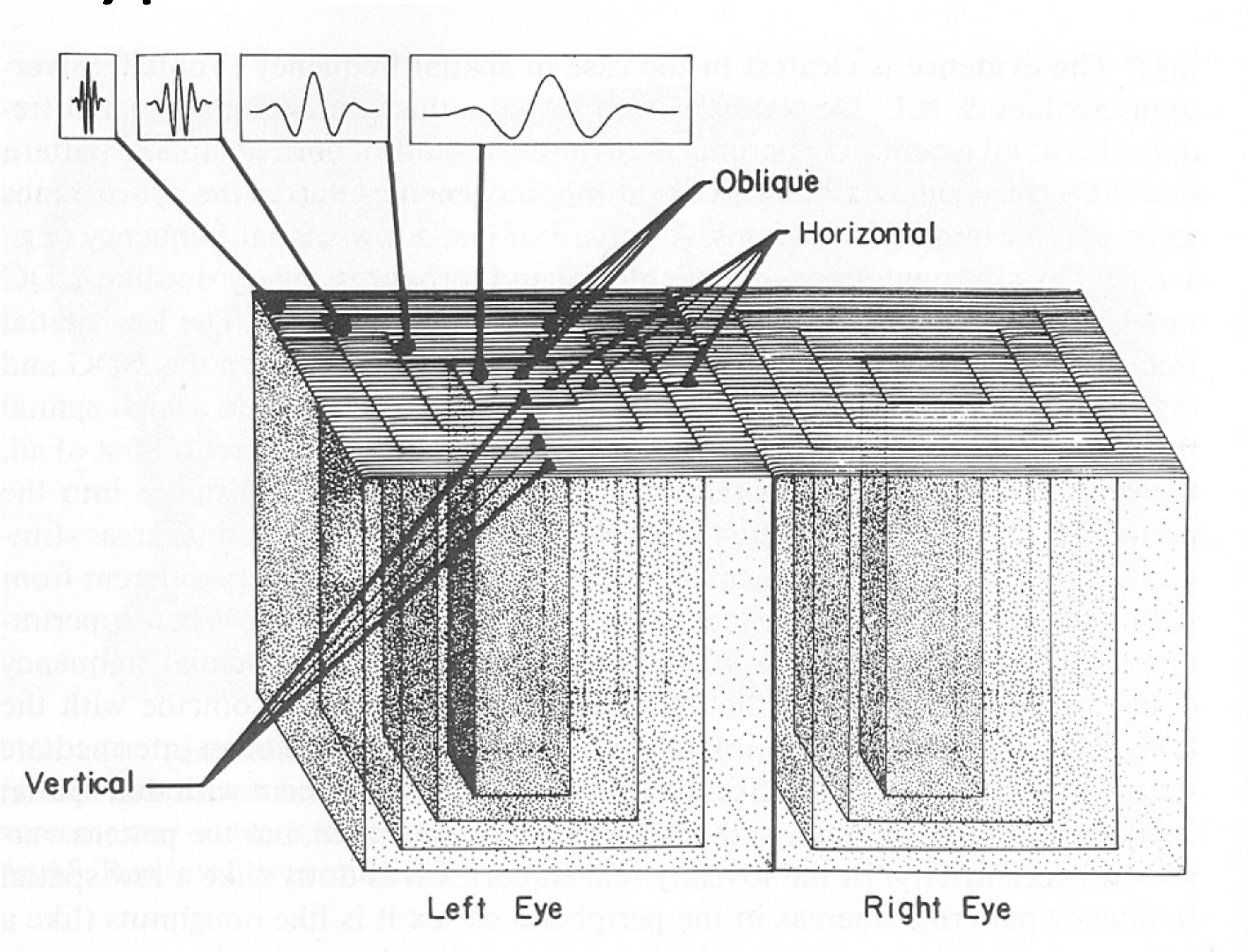


Orientation Energy

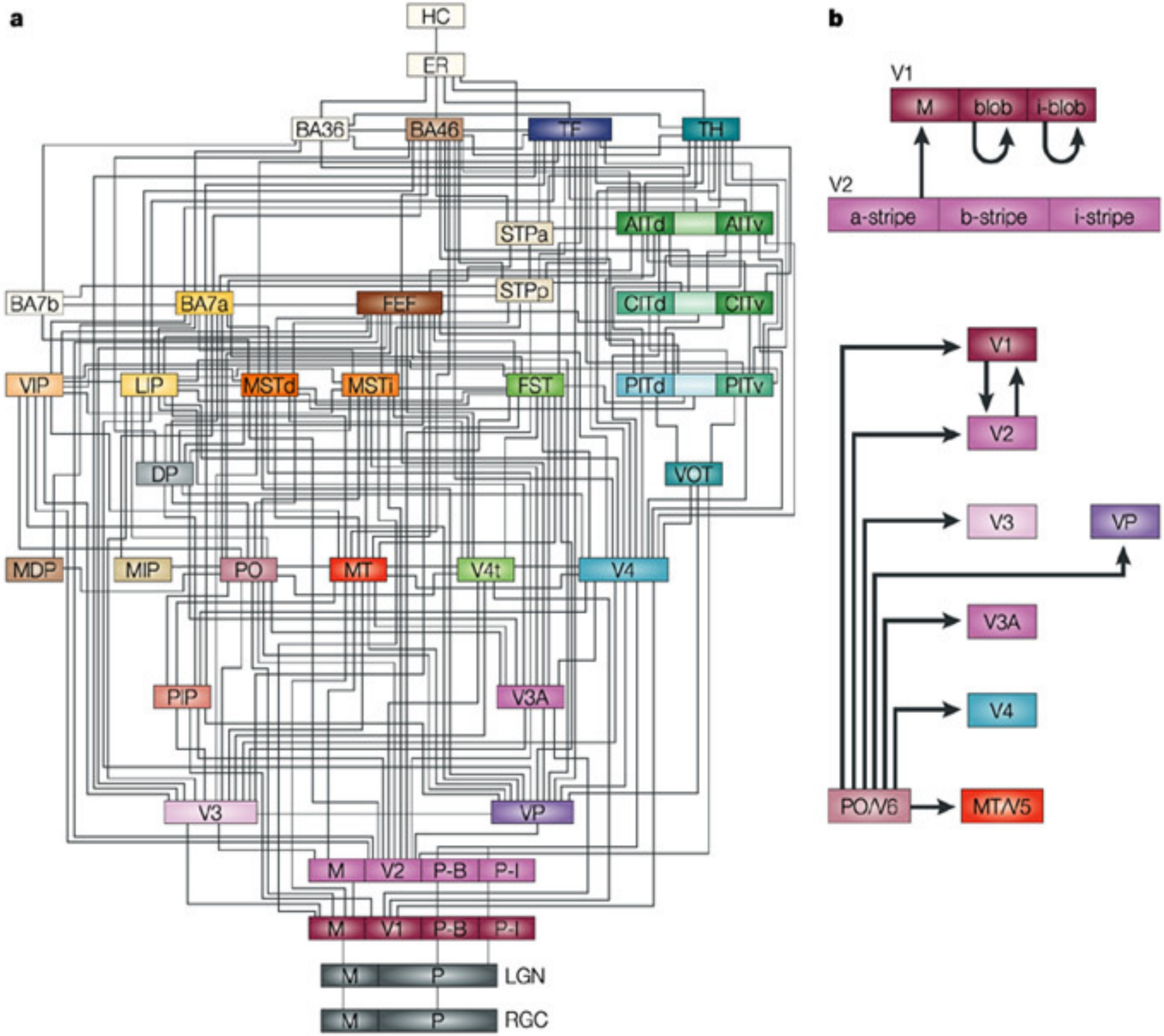


- $OE = (I * f_{odd})^2 + (I * f_{even})^2$
- Can be used to model complex cells, as this is insensitive to phase
- Multiple scales

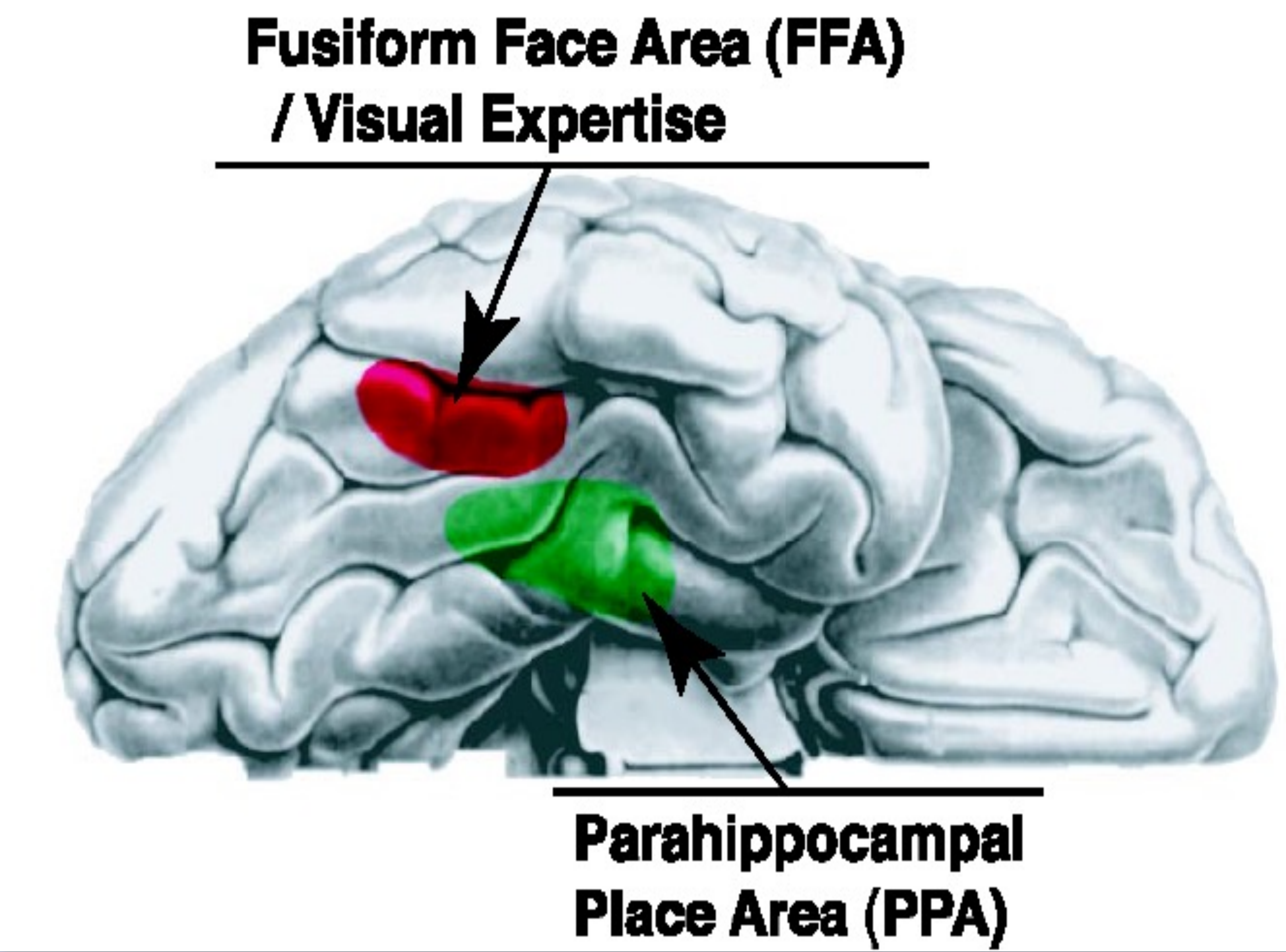
Hypercolumns in visual cortex



Model of Striate Module in Monkeys

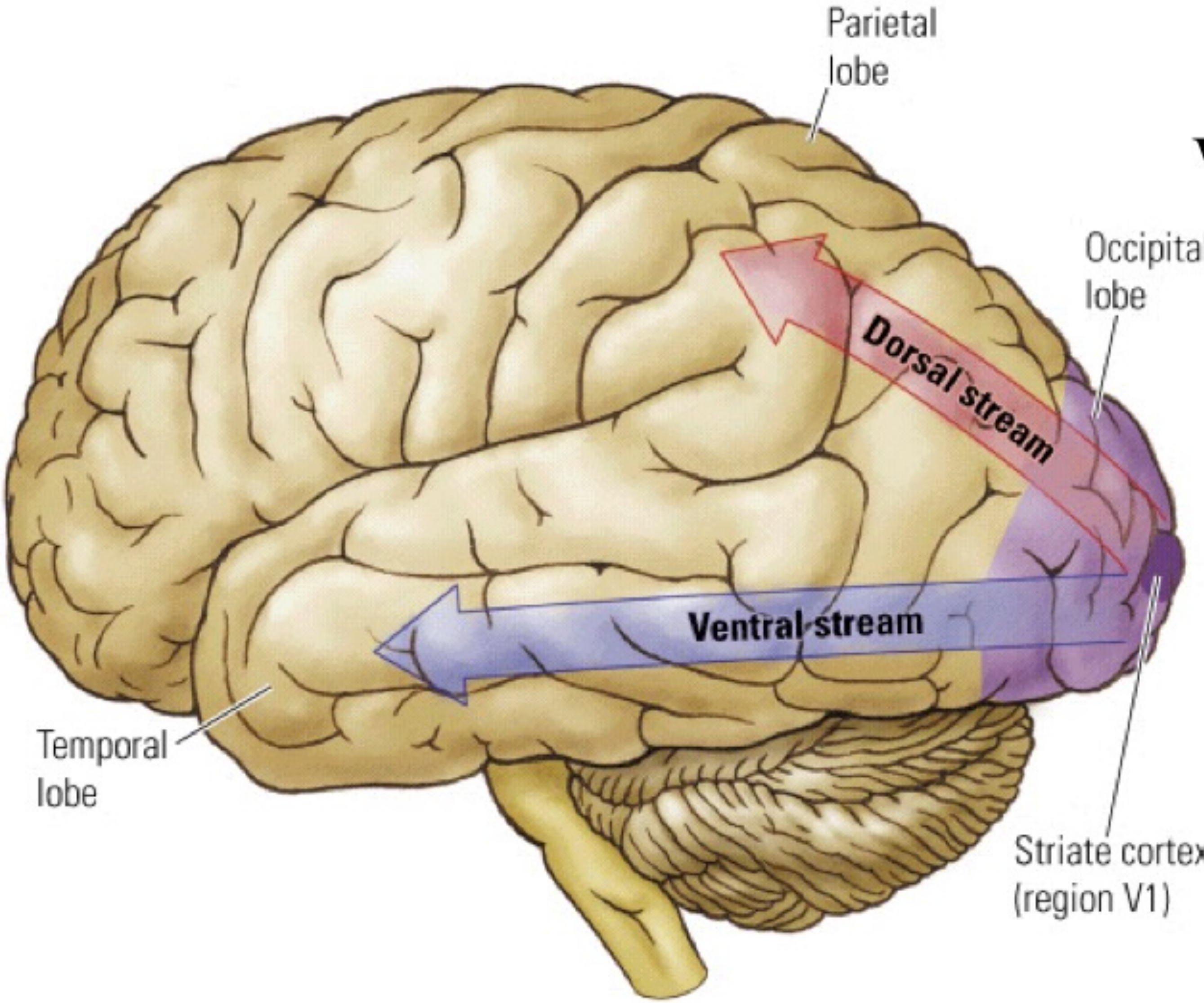


Nature Reviews | Neuroscience



Fusiform Face Area (FFA)
/ Visual Expertise
Parahippocampal
Place Area (PPA)

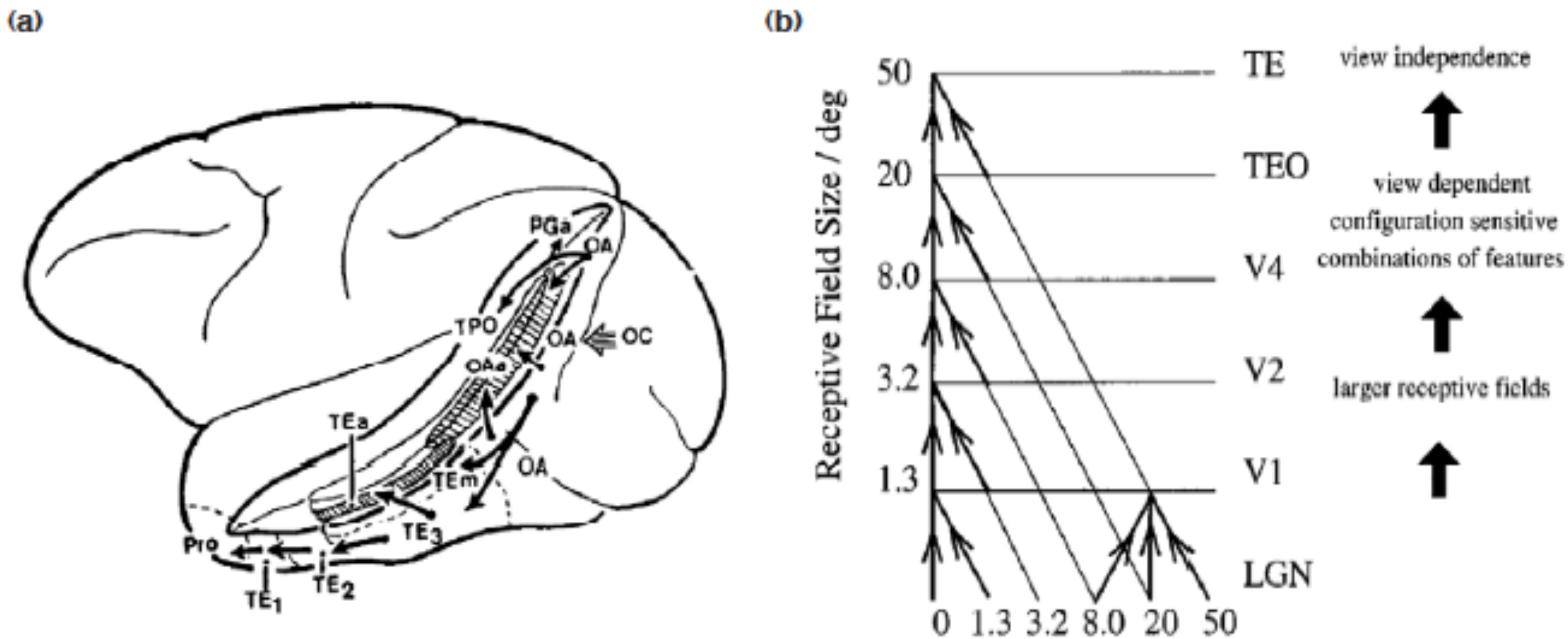
Dorsal and Ventral Streams



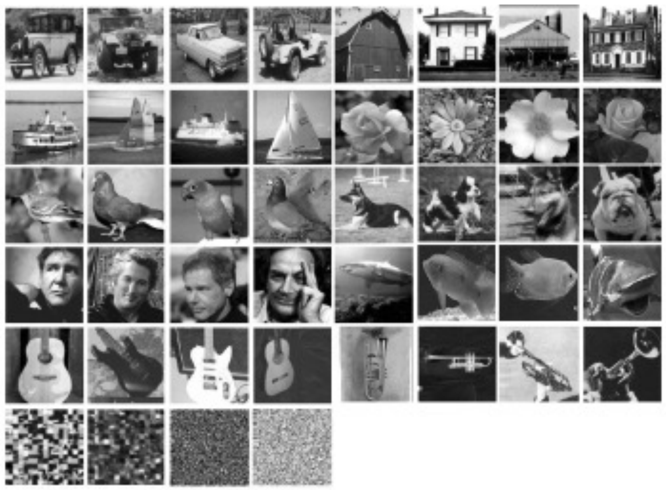
Where

What

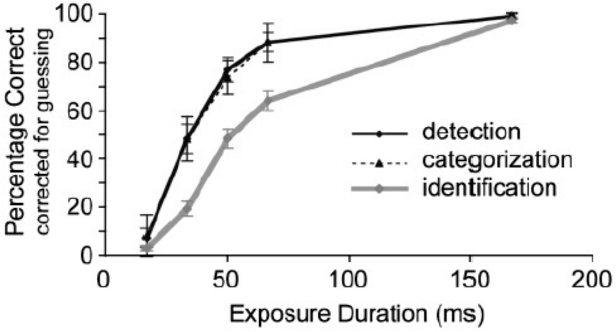
Rolls et al (2000) model of ventral stream



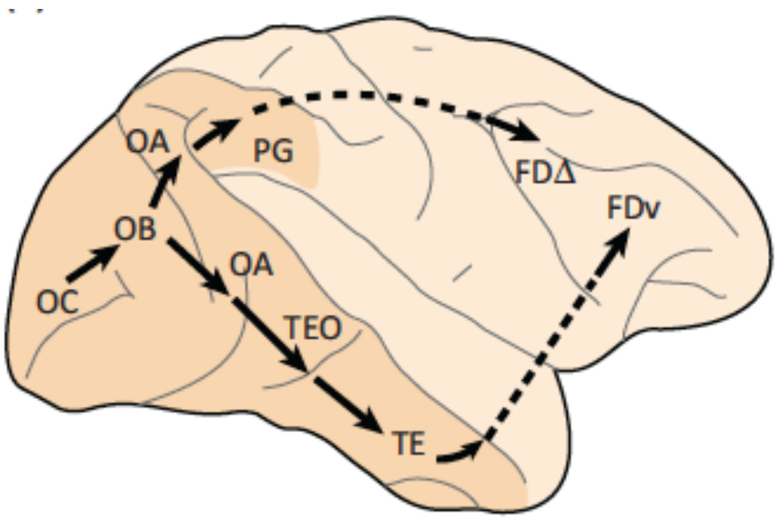
Object Detection can be very fast

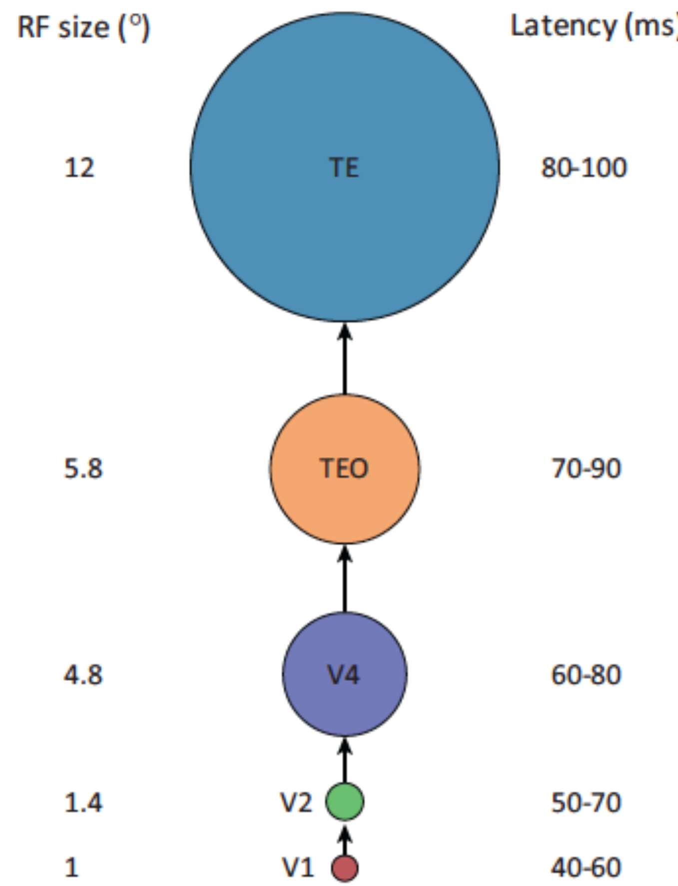


- On a task of judging animal vs no animal, humans can make mostly correct saccades in 150 ms (Kirchner & Thorpe, 2006)
  - Comparable to synaptic delay in the retina, LGN, V1, V2, V4, IT pathway.
  - Doesn't rule out feed back but shows feed forward only is very powerful
- Detection and categorization are practically simultaneous (Grill-Spector & Kanwisher, 2005)



Feed-forward model of the ventral stream





Intrinsic & Extrinsic Connectivity of the Ventral Stream
(Kravitz, Saleem, Baker, Ungerleinder, Mishkin, TICS, 2013)

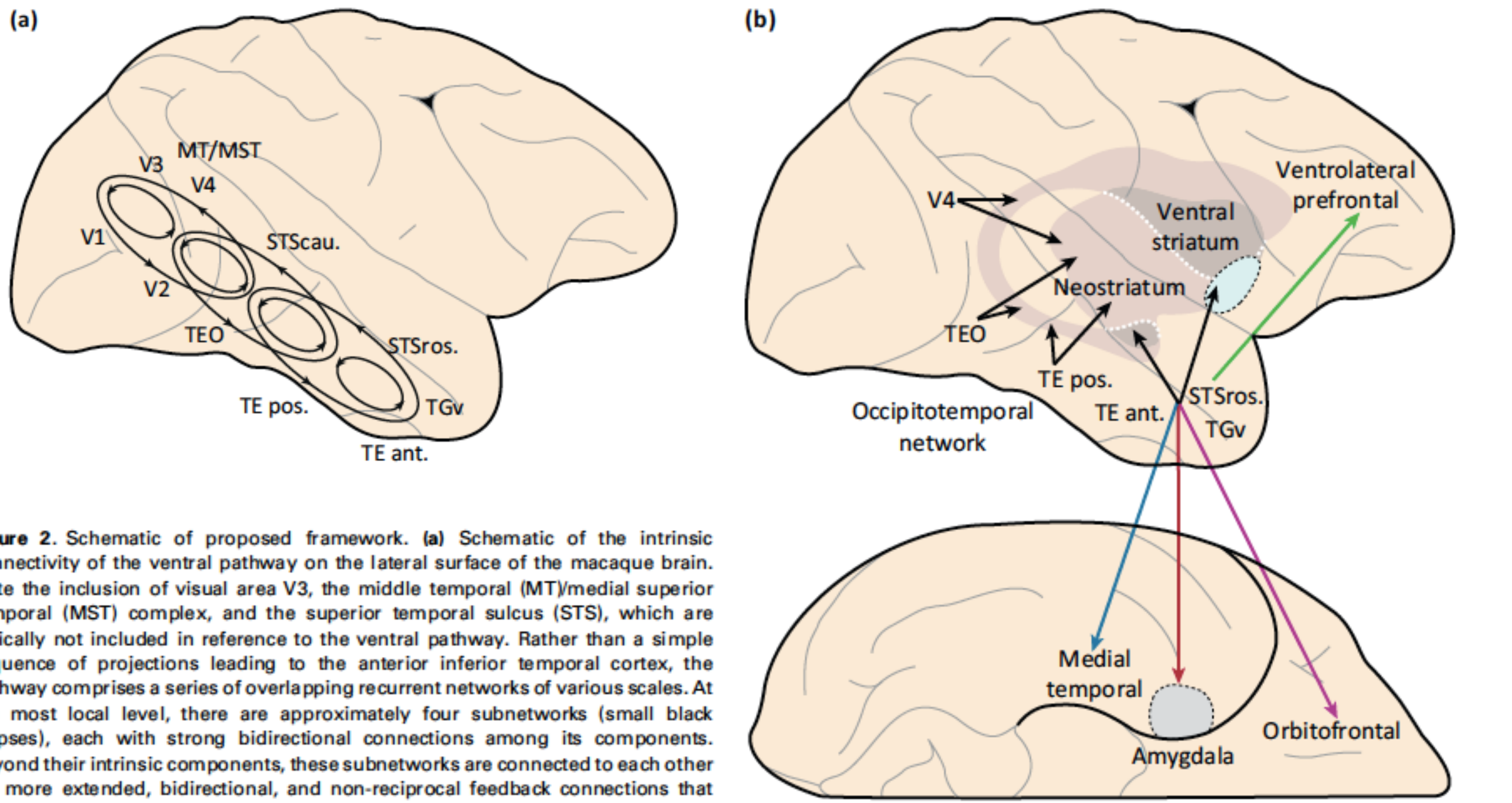


Figure 2. Schematic of proposed framework. **(a)** Schematic of the intrinsic connectivity of the ventral pathway on the lateral surface of the macaque brain. Note the inclusion of visual area V3, the middle temporal (MT)/medial superior temporal (MST) complex, and the superior temporal sulcus (STS), which are typically not included in reference to the ventral pathway. Rather than a simple sequence of projections leading to the anterior inferior temporal cortex, the pathway comprises a series of overlapping recurrent networks of various scales. At the most local level, there are approximately four subnetworks (small black ellipses), each with strong bidirectional connections among its components. Beyond their intrinsic components, these subnetworks are connected to each other via more extended, bidirectional, and non-reciprocal feedback connections that bypass intermediate regions (large black ellipses). **(b)** A summary of the extrinsic connectivity of the ventral pathway. At least six distinct pathways emanate from the occipitotemporal network. The occipitotemporo-neostriatal pathway (black

What can we learn?

- Neurons show increasing specificity higher in the visual pathway
- V1 simple and complex cells are orientation-tuned
- Convolution with a linear kernel followed by simple non-linearities is a good model for computation in retina, LGN and V1, but beyond that we do not have satisfactory computational models
- Good designs of visual systems are likely to be hierarchical and "mostly" feedforward

Neuroscience & Computer Vision Features

- Hubel & Wiesel’s finding of orientation selective simple and complex cells in V1 inspired features such as SIFT and HOG.
- A feed-forward view of processing in the ventral stream with layers of simple and complex cells led to the neocognitron and subsequently convolutional networks.
- We now know that the ventral stream is much more complicated with bidirectional as well as feedback connections. So far this has not been exploited much in computer vision

Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position

Kunihiko Fukushima

NHK Broadcasting Science Research Laboratories, Kinuta, Setagaya, Tokyo, Japan

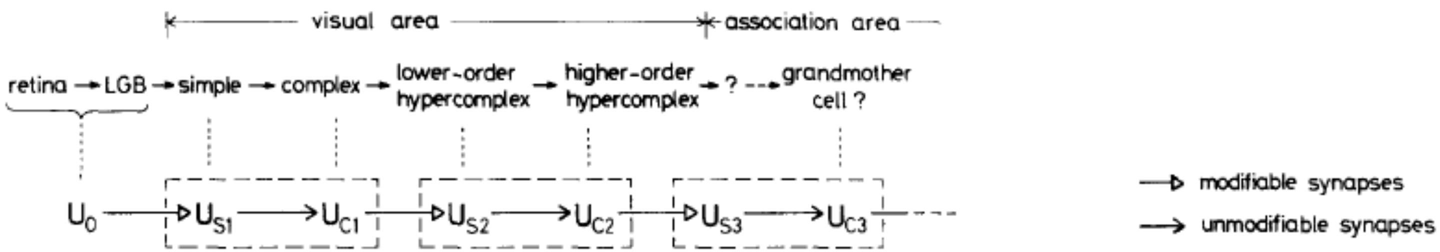


Fig. 1. Correspondence between the hierarchy model by Hubel and Wiesel, and the neural network of the neocognitron

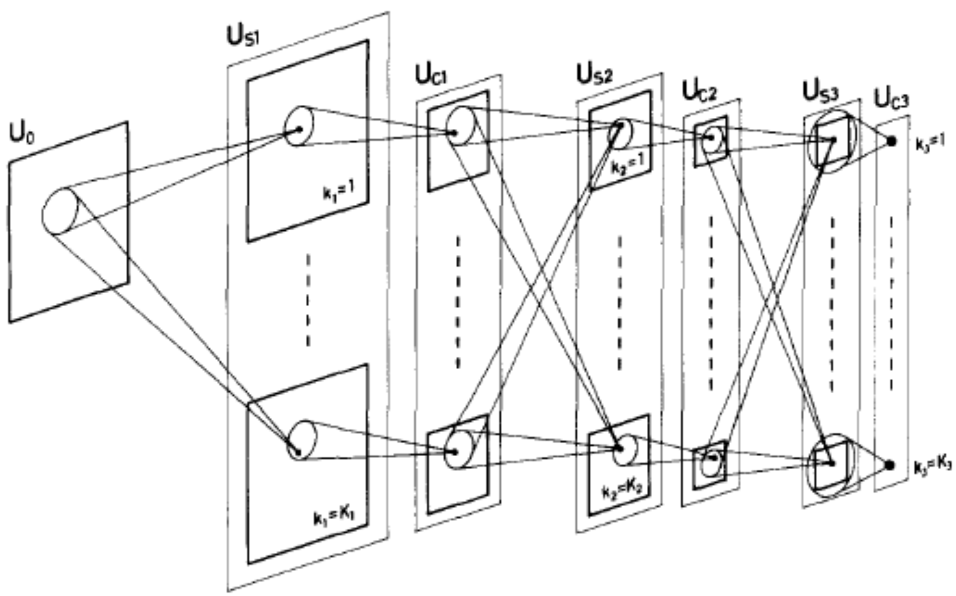
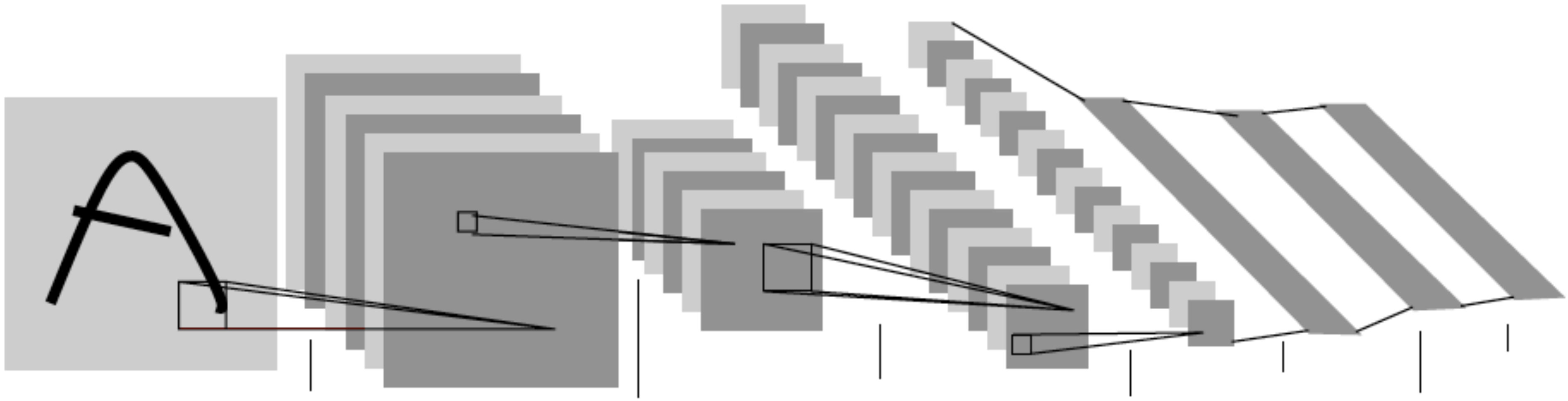


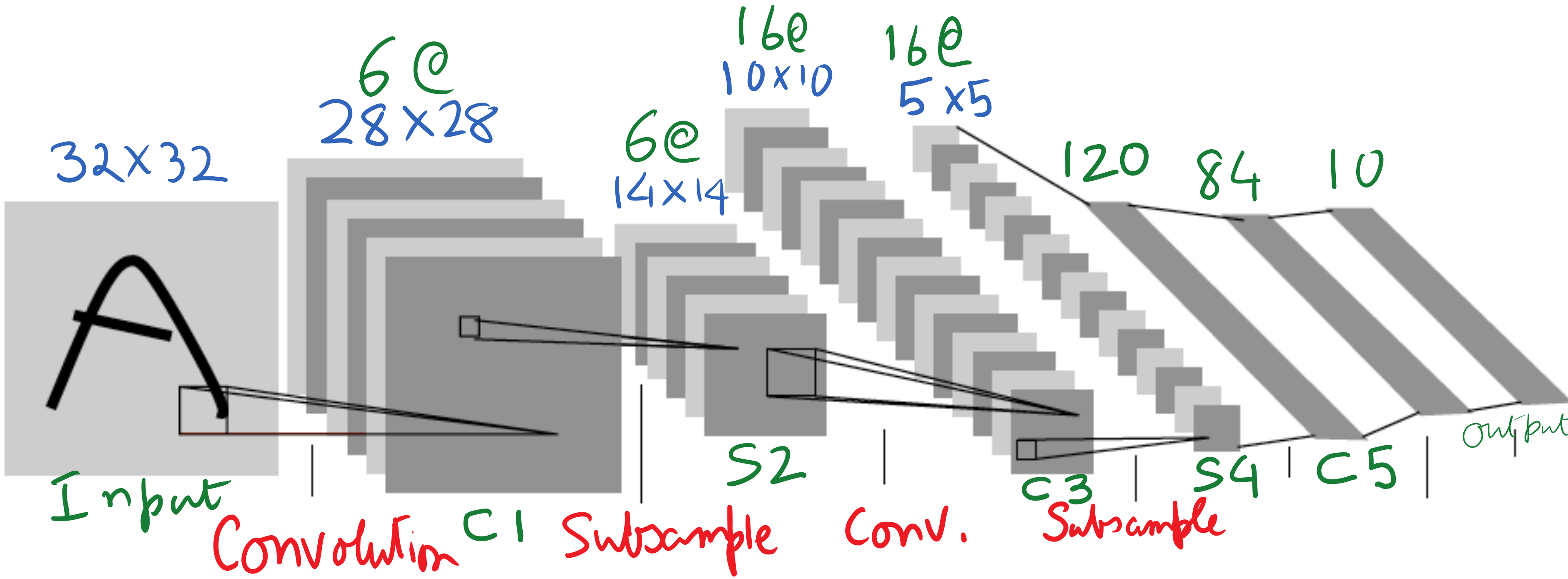
Fig. 2. Schematic diagram illustrating the interconnections between layers in the neocognitron

Biol. Cybernetics 36, 193–202 (1980)

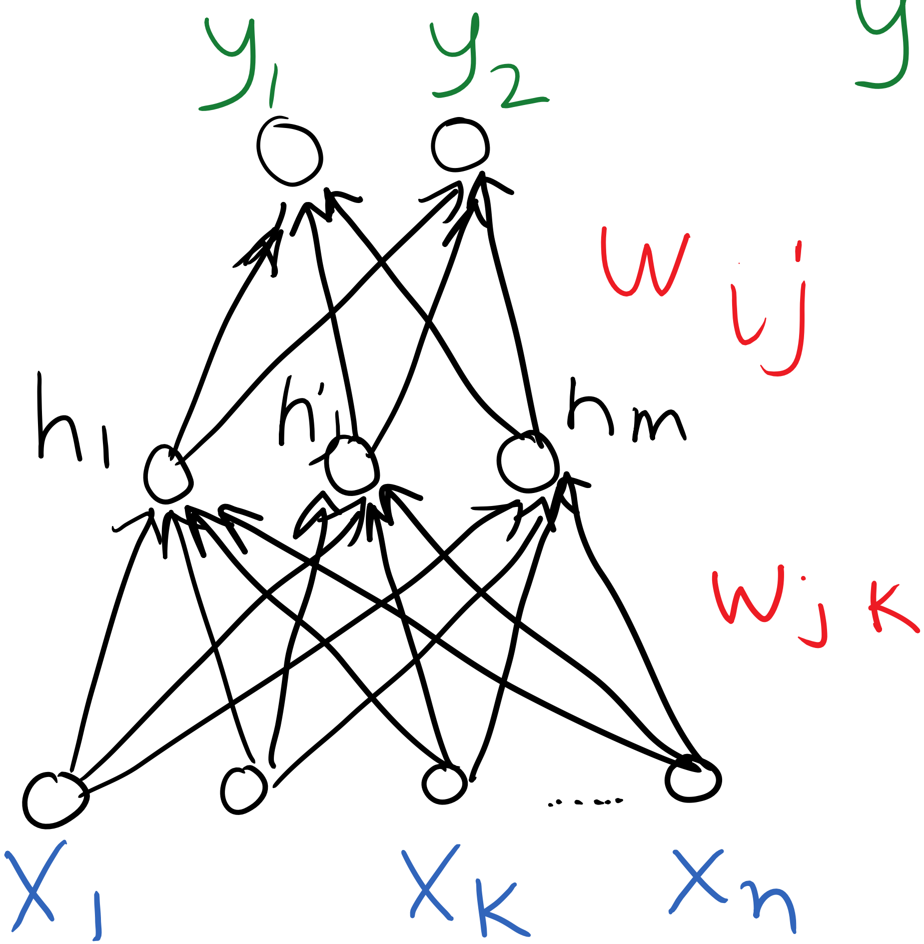
Convolutional Neural Networks (LeCun et al)



Convolutional Neural Networks (LeCun et al)



Training multi-layer networks



$$y_i = g\left(\sum_j w_{ij} h_j\right)$$

$$h_j = g\left(\sum_k w_{jk} x_k\right)$$

Minimize $\left(y_i - y_i^{desired}\right)^2$ by suitable choice of w's