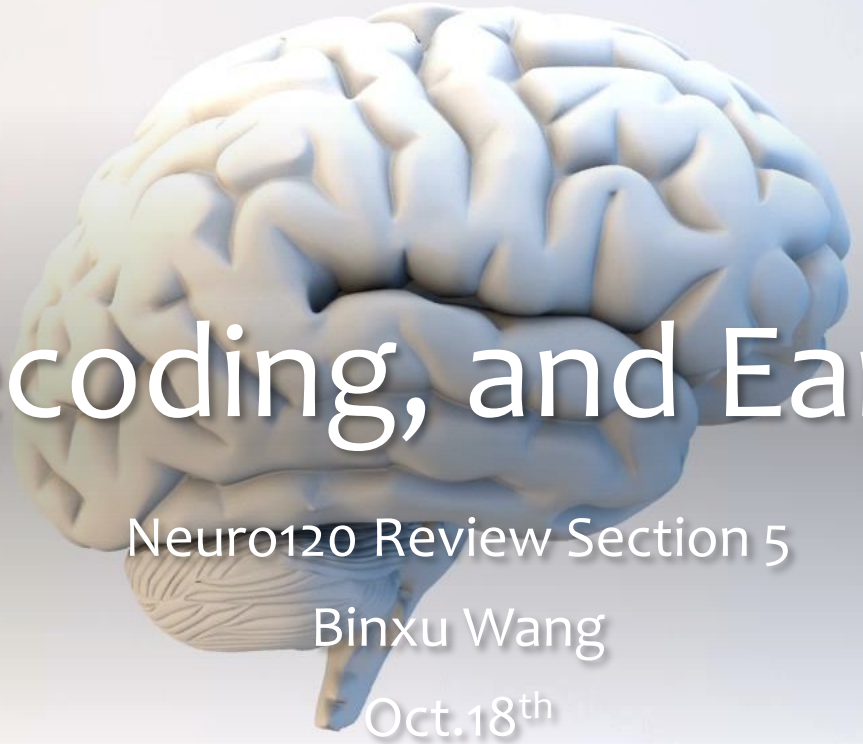


PCA, Decoding, and Early Vision

Neuro120 Review Section 5

Binxu Wang

Oct.18th

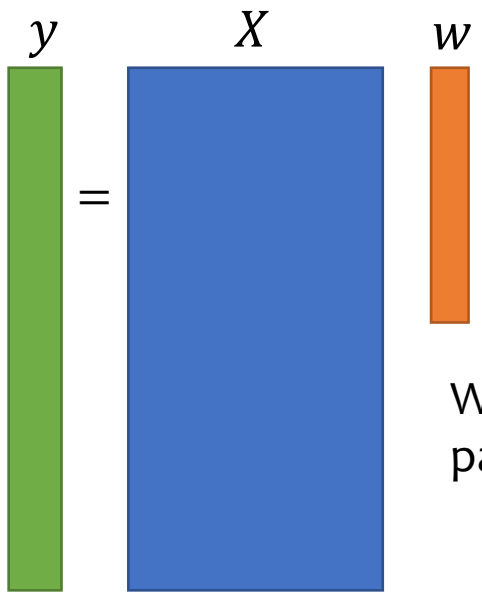
Outline

- Neural Decoding (P2)
  - Math: Linear regression
- Dimensionality reduction (P1)
  - Math: PCA
- Model of early vision (P3)
  - Math: DoG kernel, Convolution, Padding etc.

Linear Regression (P2)

- Our model

$$y = Xw + \epsilon$$

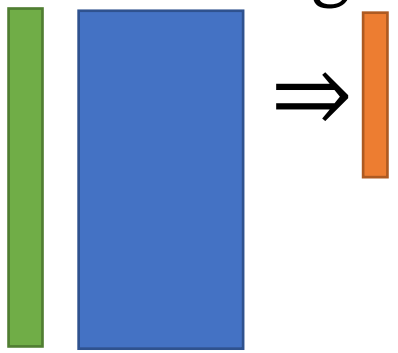


Weights / parameter

Response / Target

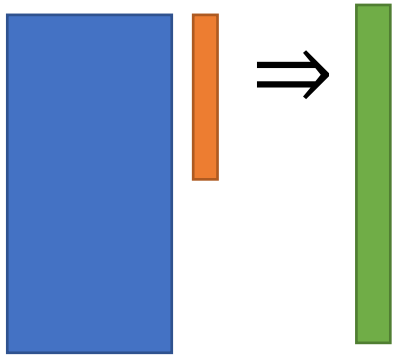
Predictor data / Regressor

- Given data and target => weight



Fitting, Learning, Training …

- Given data and weight => "predicted" response



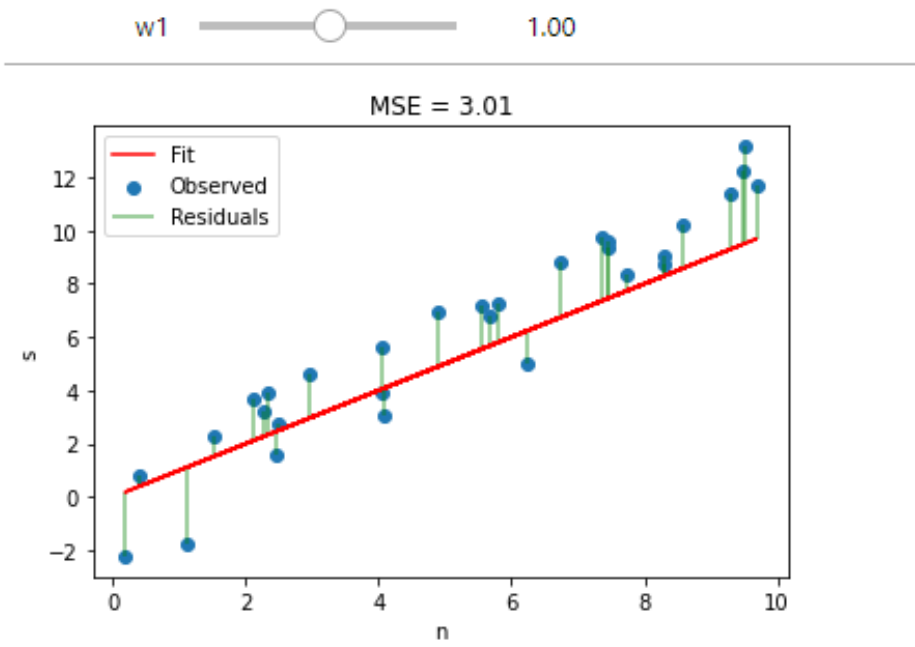
Prediction, Inference, Forward …

Linear Regression (P2)

- How do you get the w?
  - General method: set an error function, minimize prediction error.
    - Gradient descent=> Deep learning

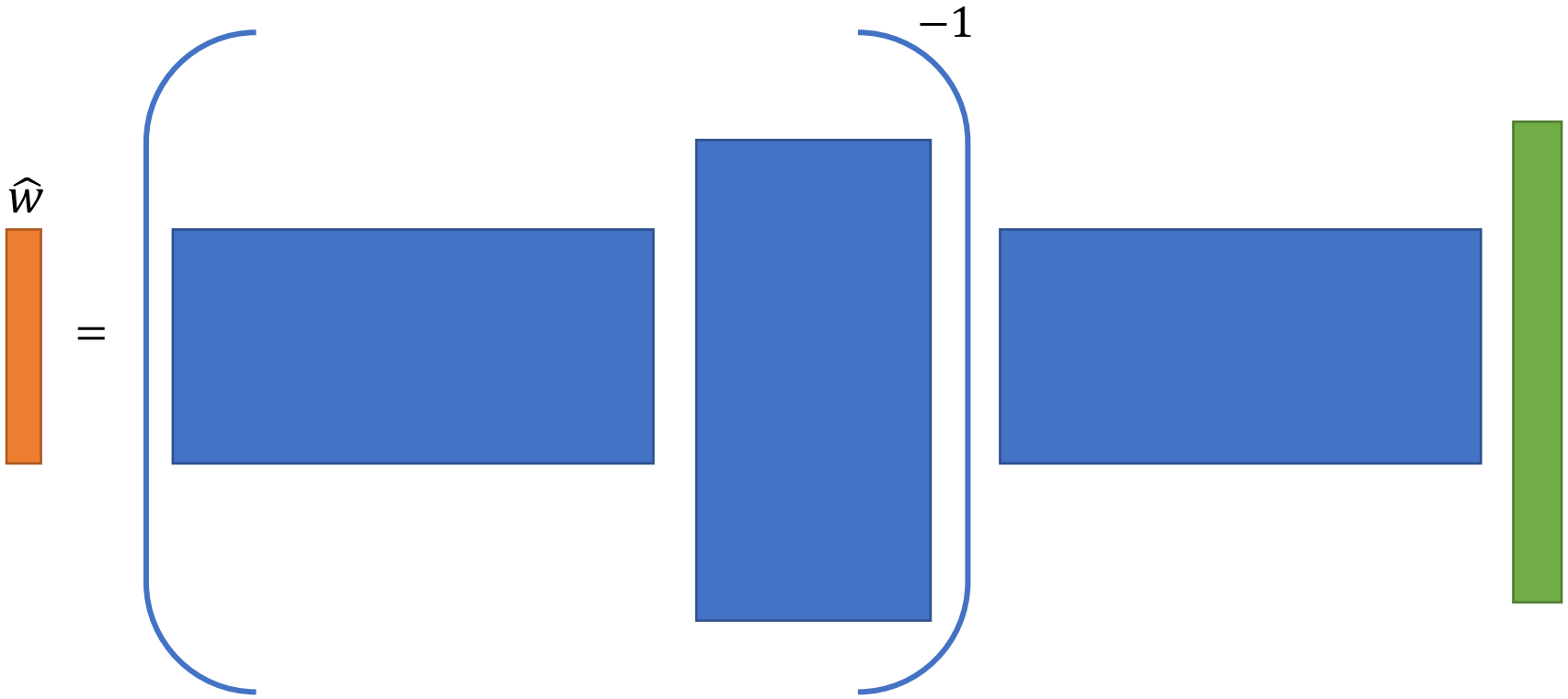
$$\arg\min_{w} \|Xw - y\|_2^2$$

  - Few cases: analytical solution!



Solution to Ordinary Least Square

$$\hat{w} = (X^T X)^{-1} X^T y$$



Quick way to "derive" the equation.

If $y = Xw$

Then,

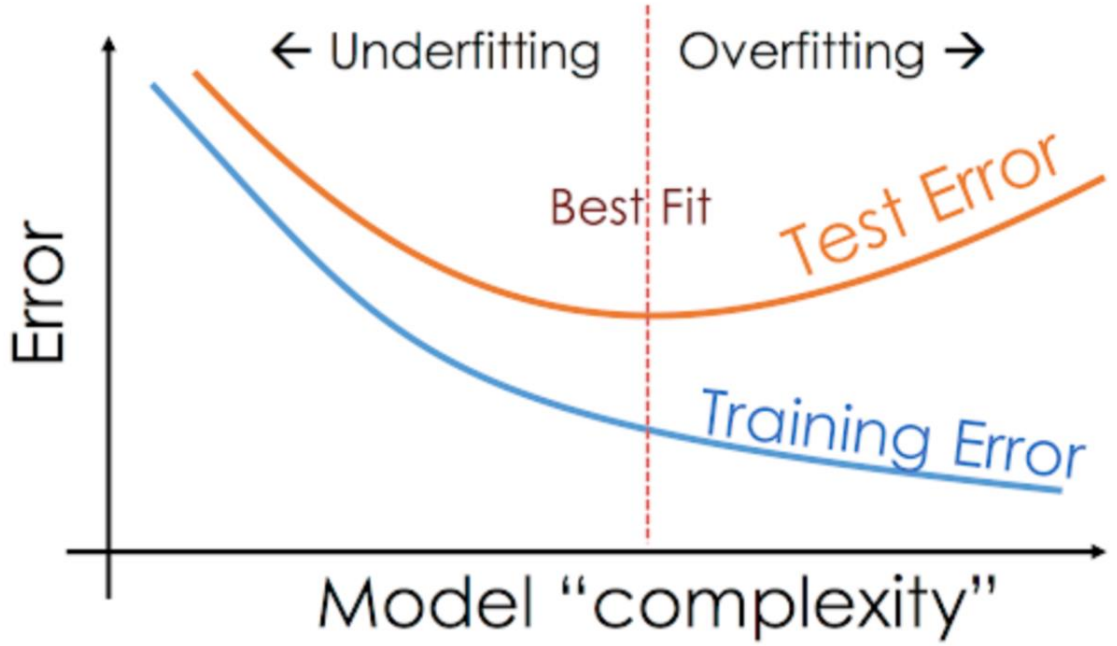
$$X^T y = X^T X w$$

$$w = (X^T X)^{-1} X^T y$$

$$\hat{w} = (X^T X)^{-1} X^T y$$

Importance of Train-Test Split

- Why do we separate the test set?
  - Test error do not always follow training error.
  - Key for model / parameter selection (not showed)
  - What about PCA ?



Meta-note on devising analysis methods

LNP, Regression, PCA, deep learning …

1. Think about what you want, what's the desired properties?
   - Predict output y from input X

2. Formulate it in math, mostly a maximization or minimization problem
   - Minimize the prediction error
   - Here we choose the L2 error: $\hat{w} = \arg\min_{w} \|y - Xw\|_2^2$

3. Solve it with optimization or algebra
   - Ordinary least square solution: $\hat{w} = (X^T X)^{-1} X^T y$

4. What if 3 fails,
   - Call your mathematician & CS friends …
   - Go back to 2, reformulate the problem (you cannot want what you want… so want something easier to get)

Forms of Neural Code

- Rate vs Temporal Coding

- Neural Doctrine vs Population Doctrine
  - Functional unit

Mixed Selectivity and Interpretability

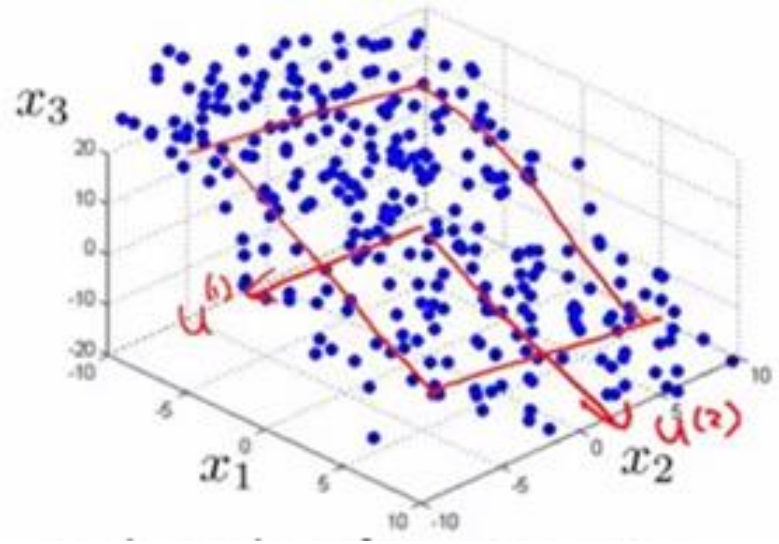
Why Dimension Reduction? (P1)

- Why: Find lower dimension *representation* that
  - Preserve information for analysis, visualization
  - Compress data
  - Assist downstream analysis tasks.
- It's possible because
  - Correlation

Principal Component Analysis (PCA)

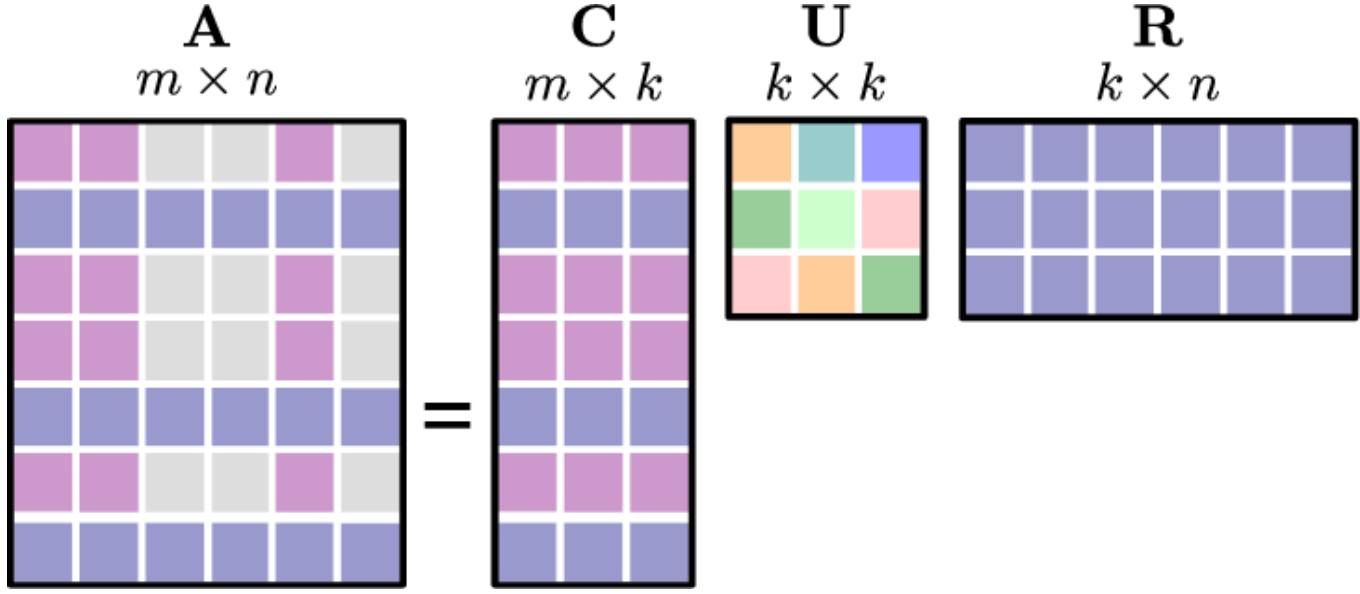
P1

- Multiple views of the same method
  - Optimal embedding of data points (Geometric)
  - Matrix decomposition / Low rank approximation / Data compression (Algebraic)



Advanced methods generalize from these two views

- Geometry: Manifold learning (LLE, Isomap, UMAP, tSNE)
- Algebra: Decomposition (NMF)

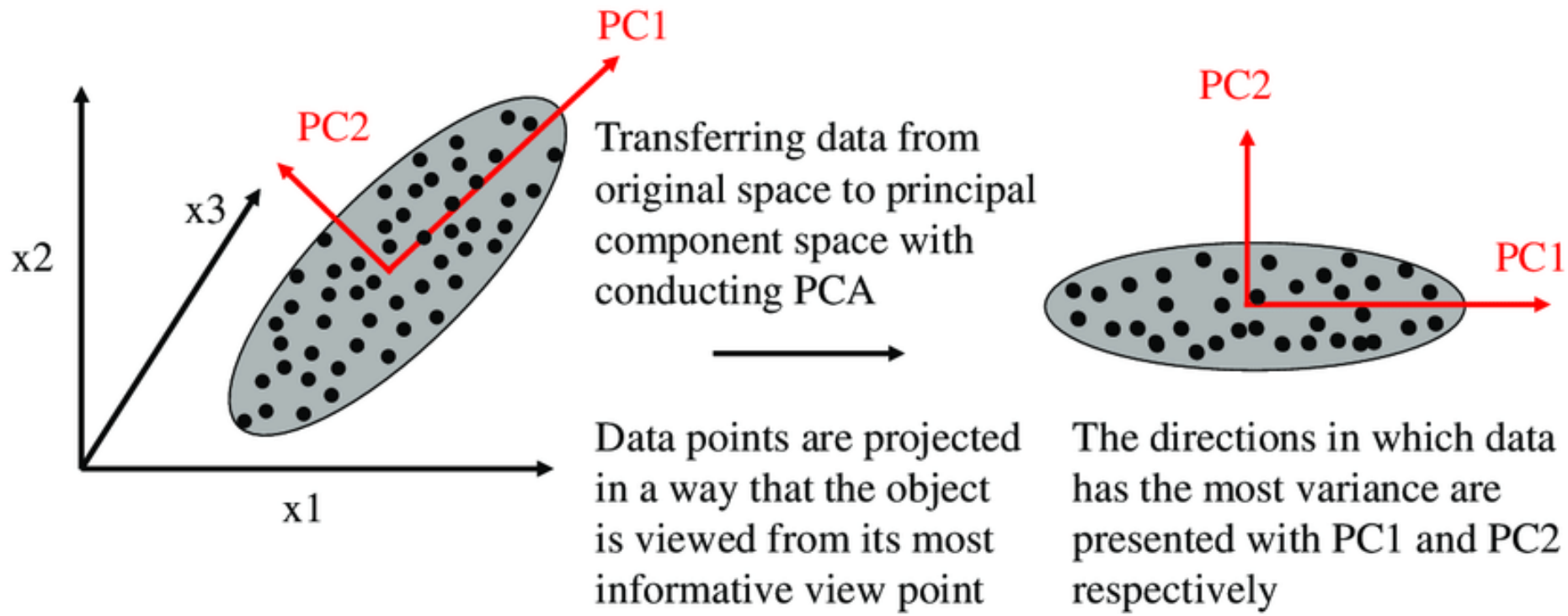


Geometric Embedding View

- Problem setup:
  - Find a sequence of directions v_i that preserve the most variance of the data after projection.

$$\arg\max_{v} Var(Xv_1)$$

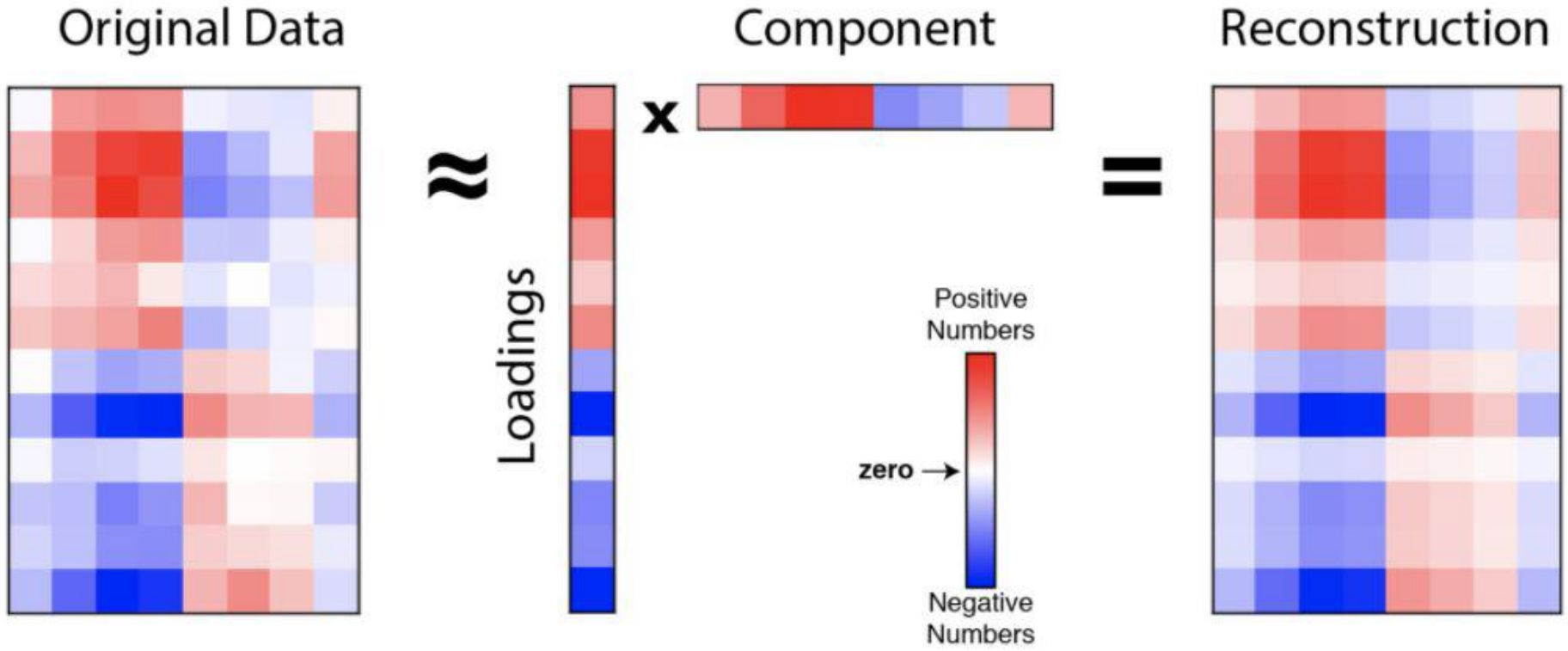
$$\arg\max_{v_2} Var(Xv_2)\text{ , given } v_2^T v_1 = 0$$



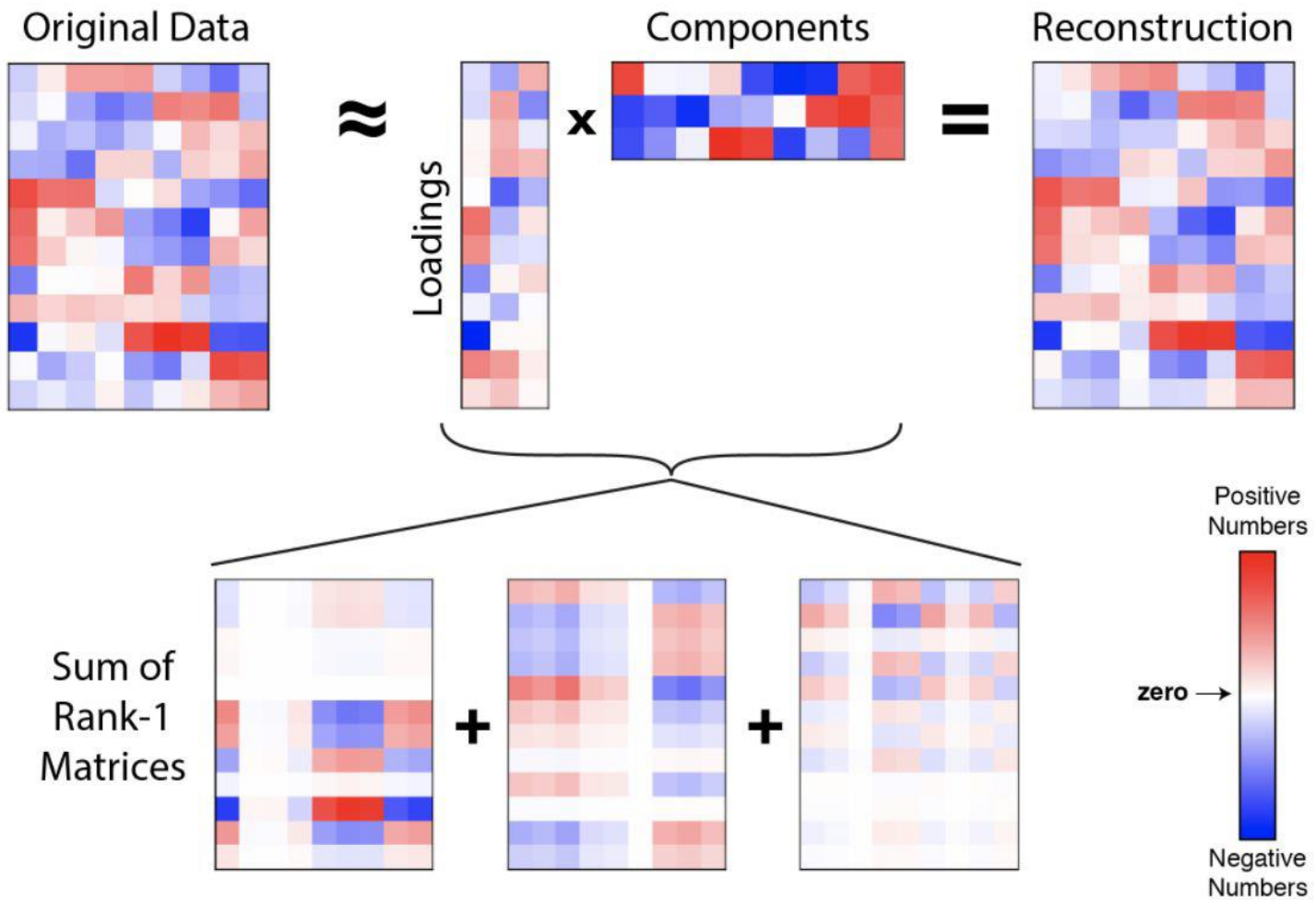
Matrix Decomposition / Approximation / Compression View

- Problem setup:
  - Find basis that make the reconstruction on this basis closest to original data.

$$\arg\min_{K,V} \|X - KV^T\|_F^2 \ , V^T V = I$$



Matrix Decomposition / Approximation



Rank of reconstructed data = number of components

Solution to PCA

- Eigen decomposition of $X^T X$
  - $X^T X = V \Lambda V^T$, V is orthogonal basis, Λ is diagonal.

- Singular value decomposition of X
  - $X = U \Sigma V^T$, U, V are orthogonal basis, Σ is diagonal.

- Results
  - The eigenvectors / singular vectors $V = [v_1, v_2, v_3 \ldots]$ provide the 1st to last principal components.

  - **Eckart–Young–Mirsky theorem**: Best rank r approximation of a matrix comes from the top r singular vectors.

$$X \approx \sum_i \sigma_i u_i v_i^T$$

Meta-note on devising analysis methods

Regression, PCA, LNP, deep learning …

1. Think about what you want, what's the desired properties?
   - Want a "good" low-dim basis / subspace

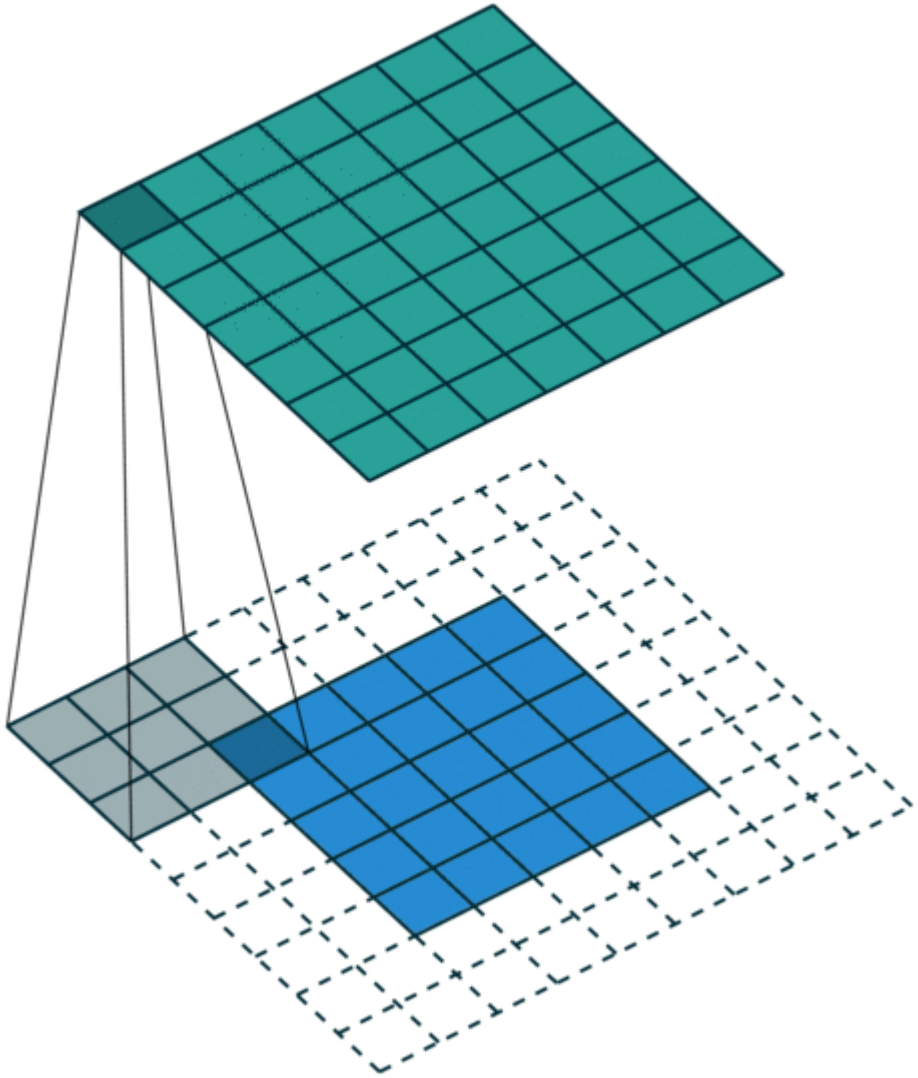
2. Formulate it in math, mostly a maximization or minimization problem
   - Minimize residue, maximize variance: $\arg\max_{v} Var(Xv)$,
     - $\arg\min_{K,V} \|X - KV^T\|_F^2$

3. Solve it with optimization or algebra
   - Solution to PCA: SVD of $X \Rightarrow$ basis V

Connection between Linear Regression and PCA?

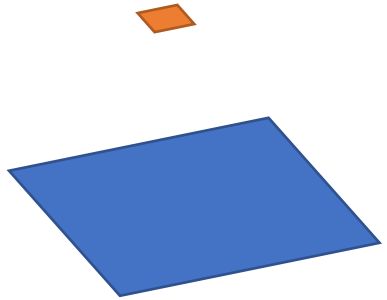
Convolution (P3 b-e)

- Which form of multiplication it is, between kernel and patch?
  - Dot (vector dot). Or element wise product and sum
- How to deal with borders?
  - Padding with “pseudo values”.
  - (“valid” parameter in `convolve2d`)

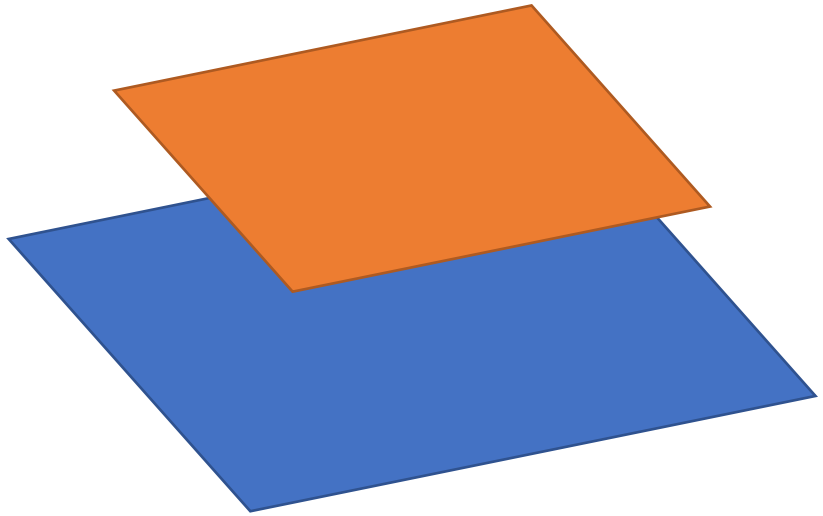


Convolution vs Linear Nonlinear Model
(P3 c vs e)

- Linear Nonlinear Model:
  - Image Patch → Scalar output
  - One neuron linearly combine / weight the patch in the RF



- Convolution:
  - Image Patch → “feature map” / image output
  - Population of neurons with same property processing / responding to the whole image



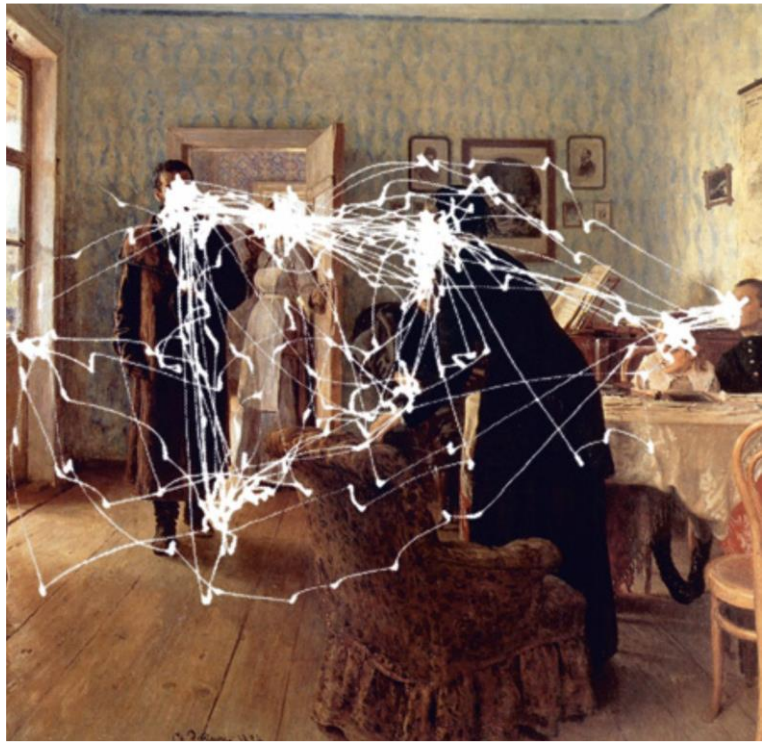
Vision

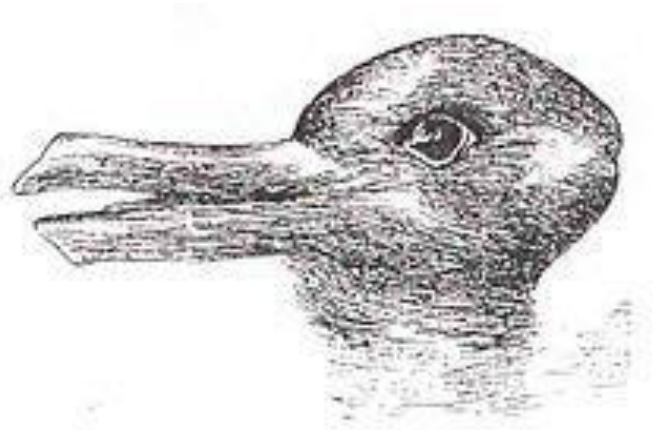
Vision is important

- It affords so many functions (visual question answering):
  - Object recognition
  - Face identification
  - Physical inference
  - Social inference
  - Reconstruct events from static images

Vision is complex

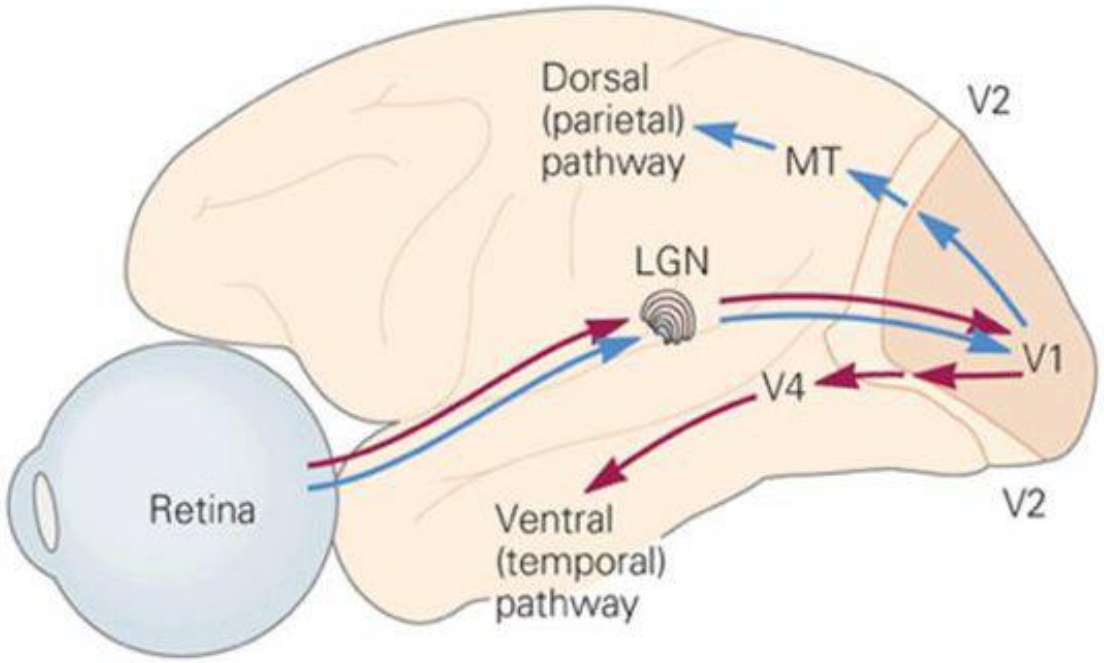
- Active sampling aspect of vision
- Influence from internal models / prior / belief / cognitive state.
  - More than mapping stimuli → response.
- It's an "Illusion".





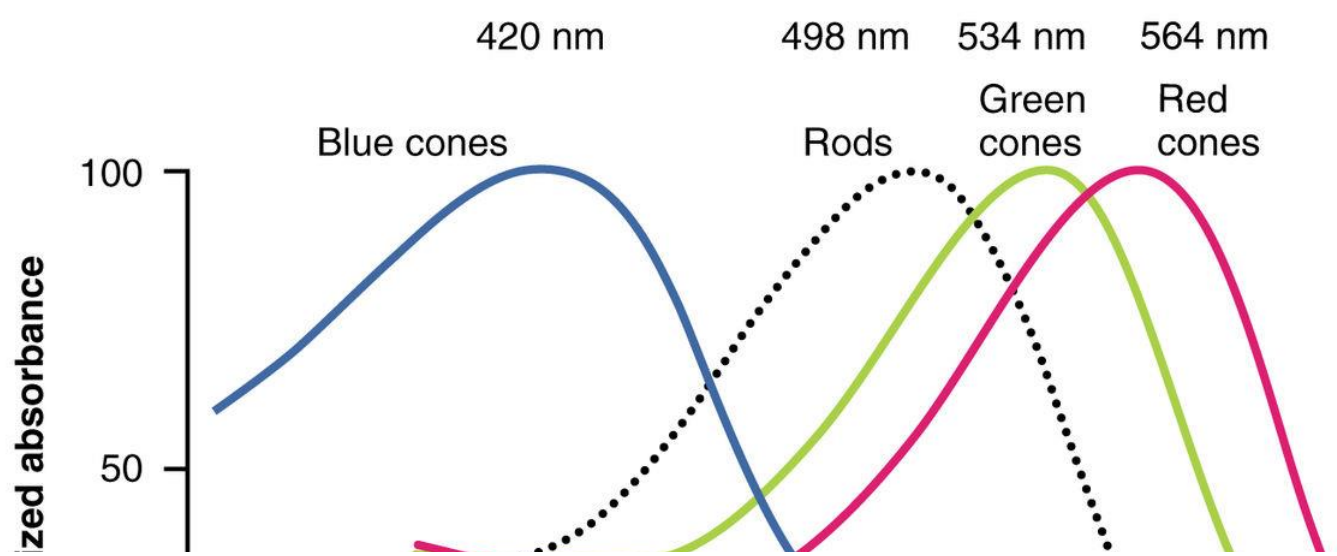
Early Vision

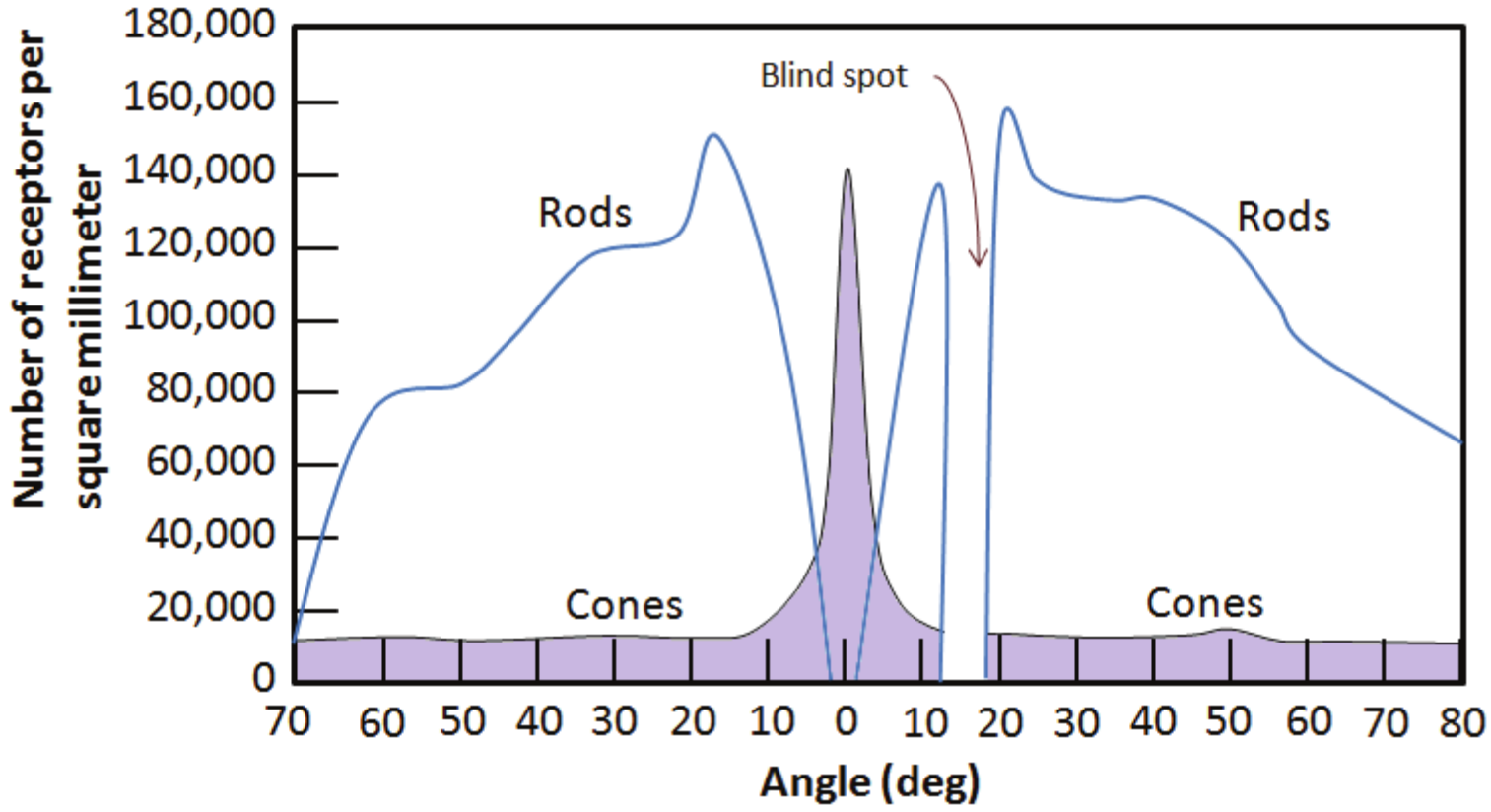
- Retina -> LGN -> V1



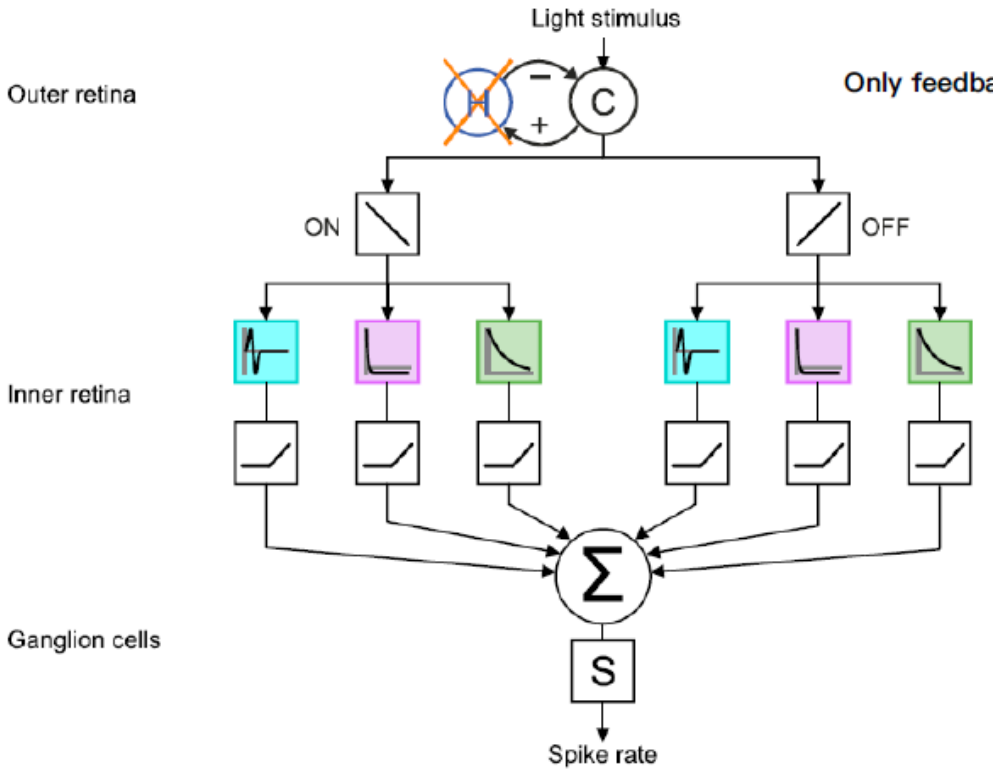
Photoreceptors

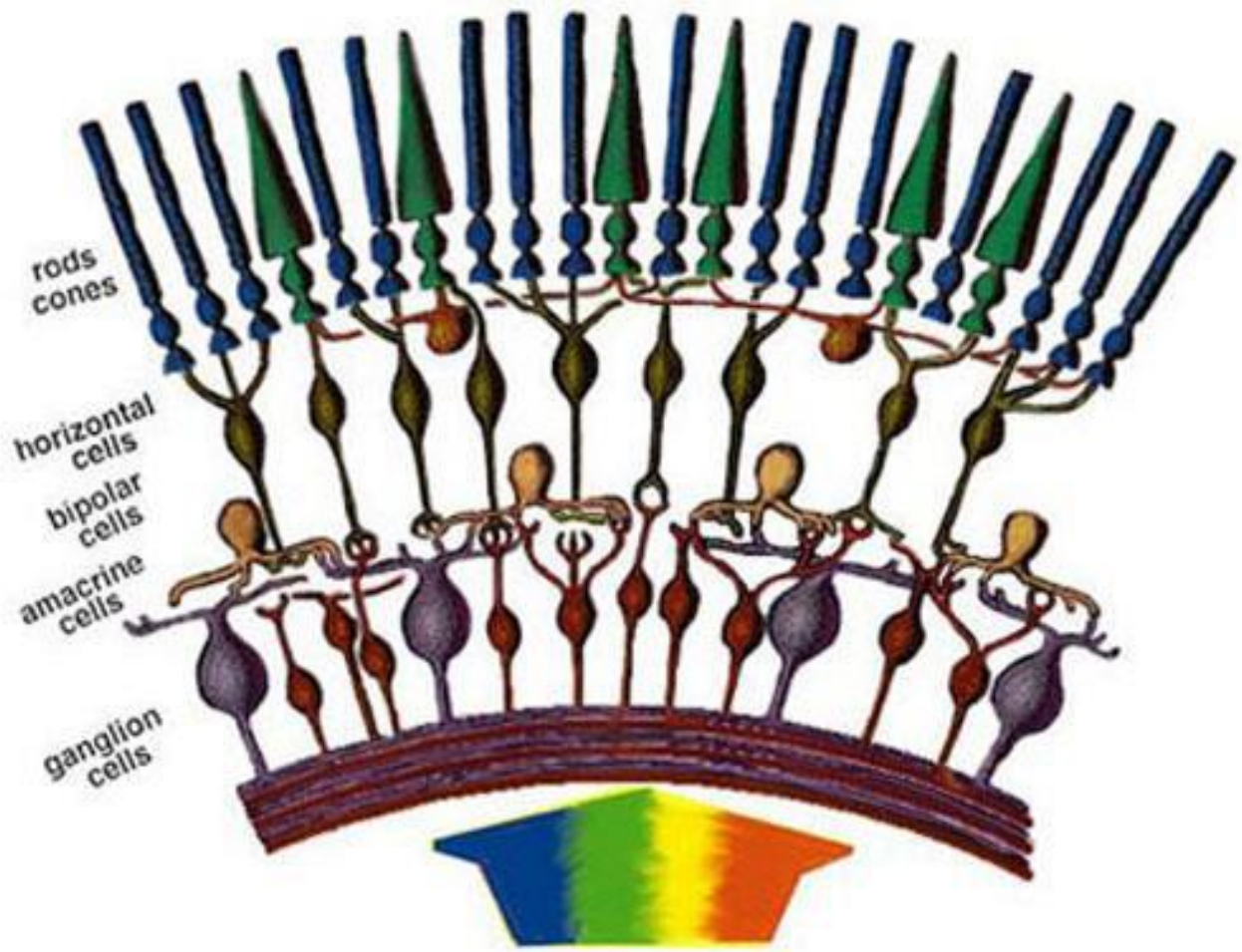
- Rods and Cones
  - Cones: S, M, L
  - Rods:
- Spatial distribution ~ eccentricity





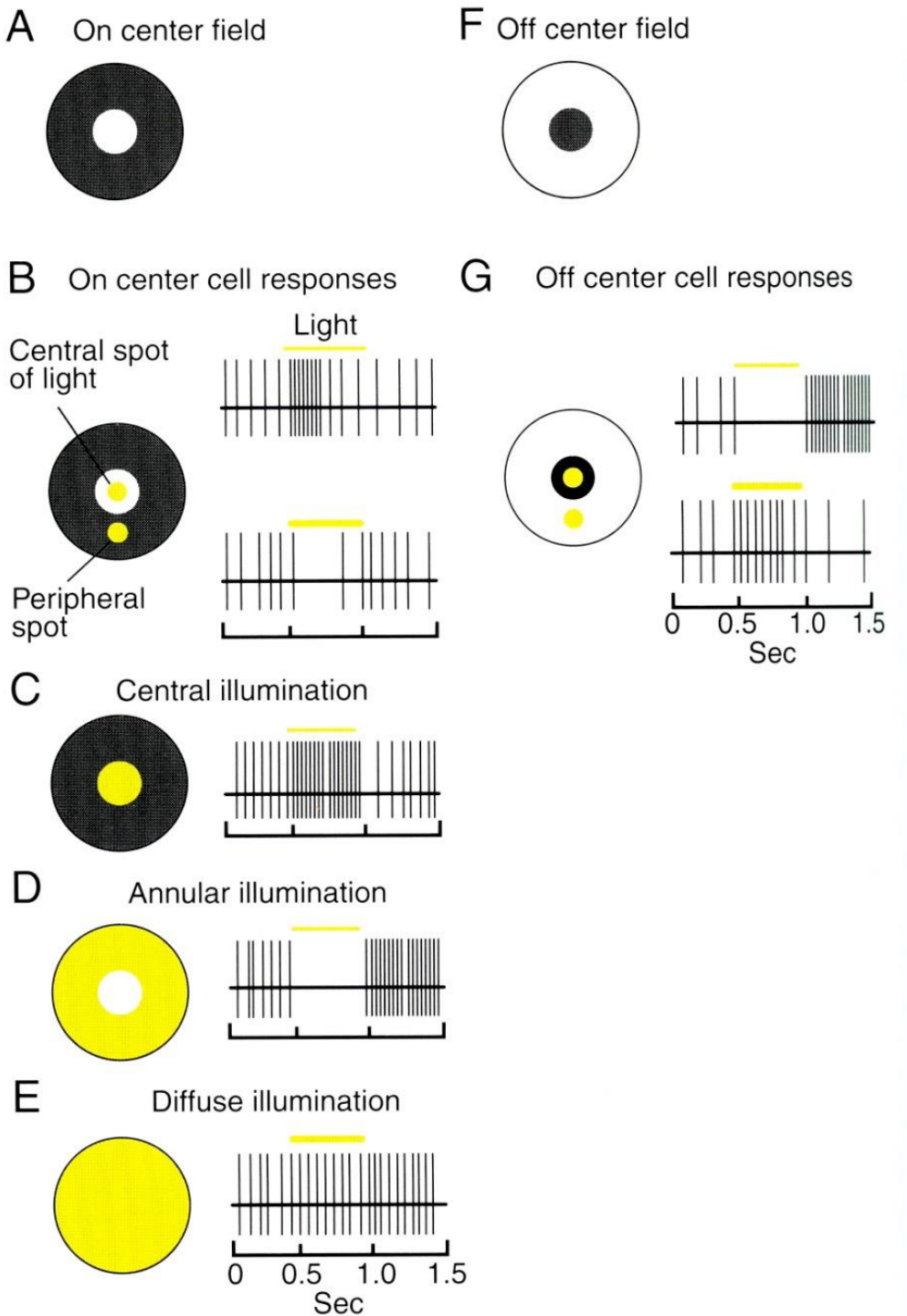
Some complicated retinal circuitry…





Response Properties of RGCs

- Spatial filtering: Center surround RF
- Temporal filtering: ON or OFF



Major Cell Types of RGC

Parvo-Cells

- *(aka Midget RGC)*
- More numerous (70%)
- Smaller receptive fields
- Higher spatial resolution
- Static/slow change, Low temporal frequency
- Low contrast sensitivity
- Chromatic information

P (ON, OFF)
1-2 pairs at each sampling locus

Magno-Cells

- *(aka Parasol RGC)*
- Less numerous (10%)
- Larger receptive fields (3-fold)
- Lower spatial resolution
- Rapid change, High temporal frequency
- High contrast sensitivity
- Achromatic

M (ON, OFF)
1-2 pairs at each sampling locus

K-Cells

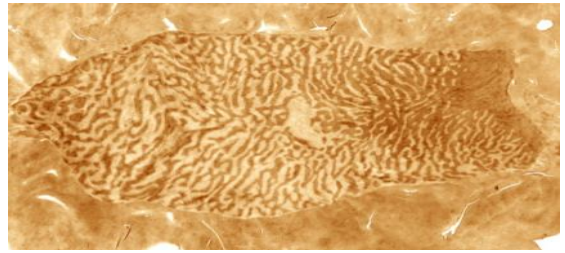
- (8%)
- Bistratified
- Spatial-, temp properties overlap with M, P

Quote from Marge Livingstone:
"Your color vision system is slow from the start!"
"Your color vision is of lower spatial resolution."
"Motion system does not know color."
https://youtu.be/oKCtrVtz_zA

From lecture slides of Andreas Burkhalter

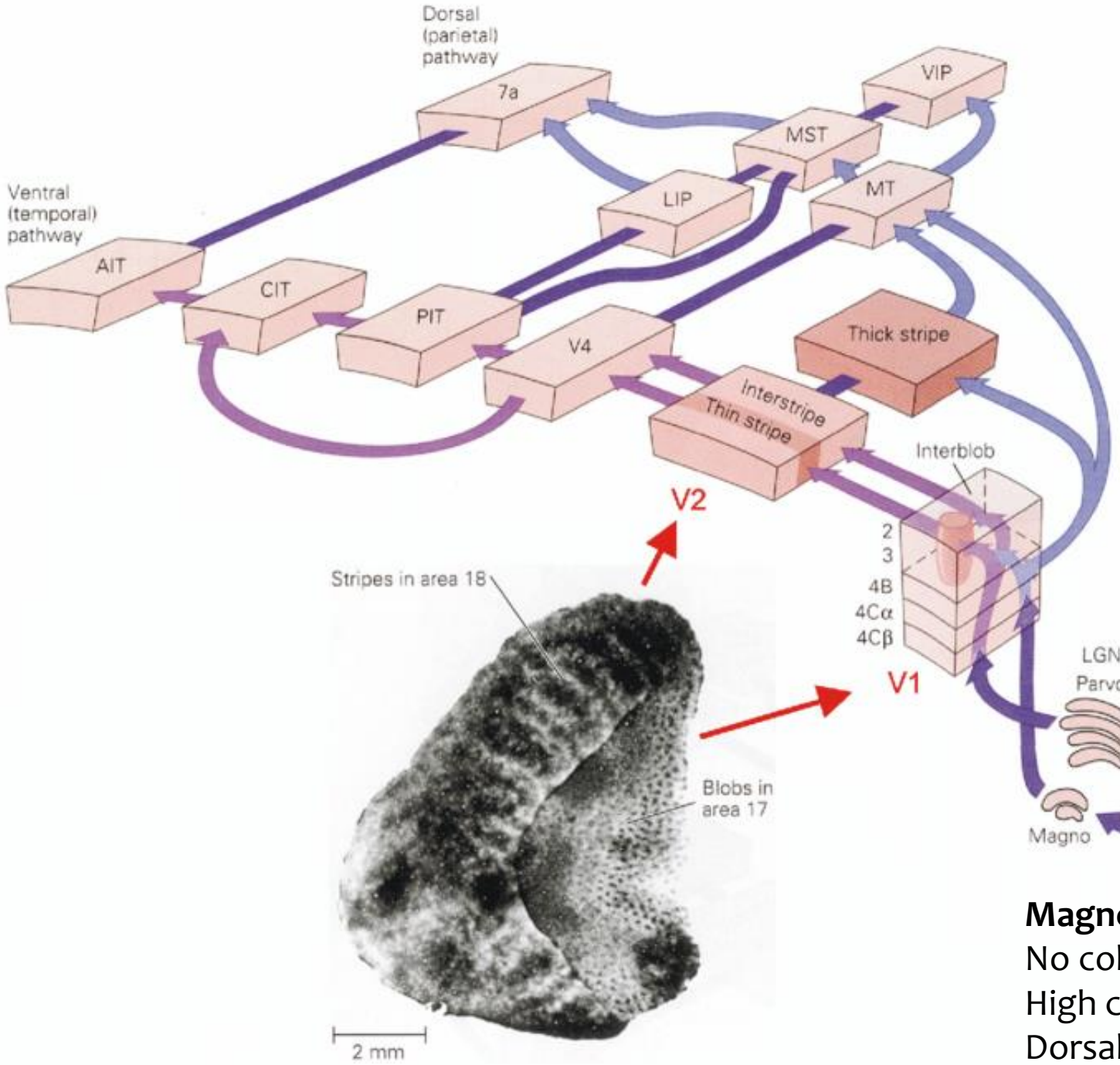
Eye-specific layers terminate in ocular dominance columns (bands)

P,M channel are fed forward in parallel.



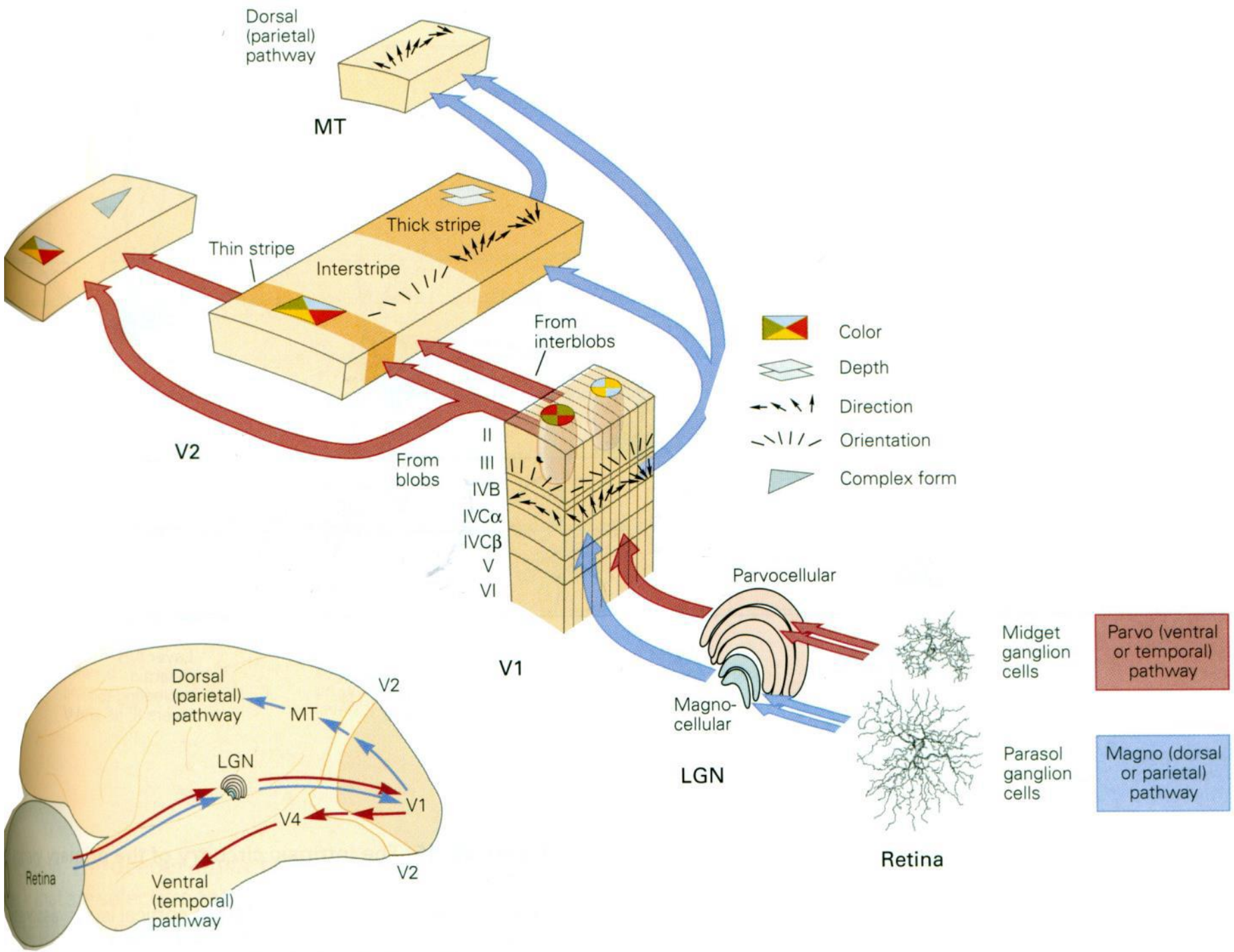
From lecture slides of Andreas Burkhalter

P and M pathway in the whole visual system

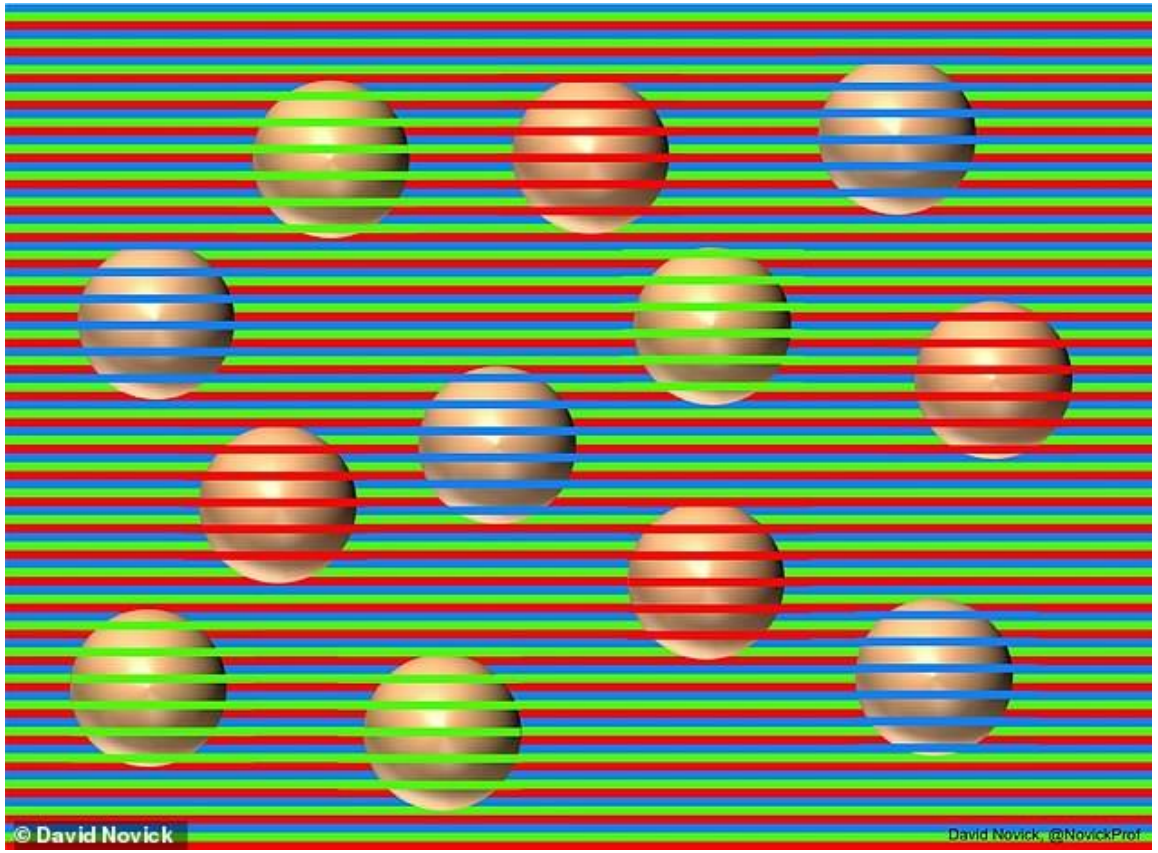


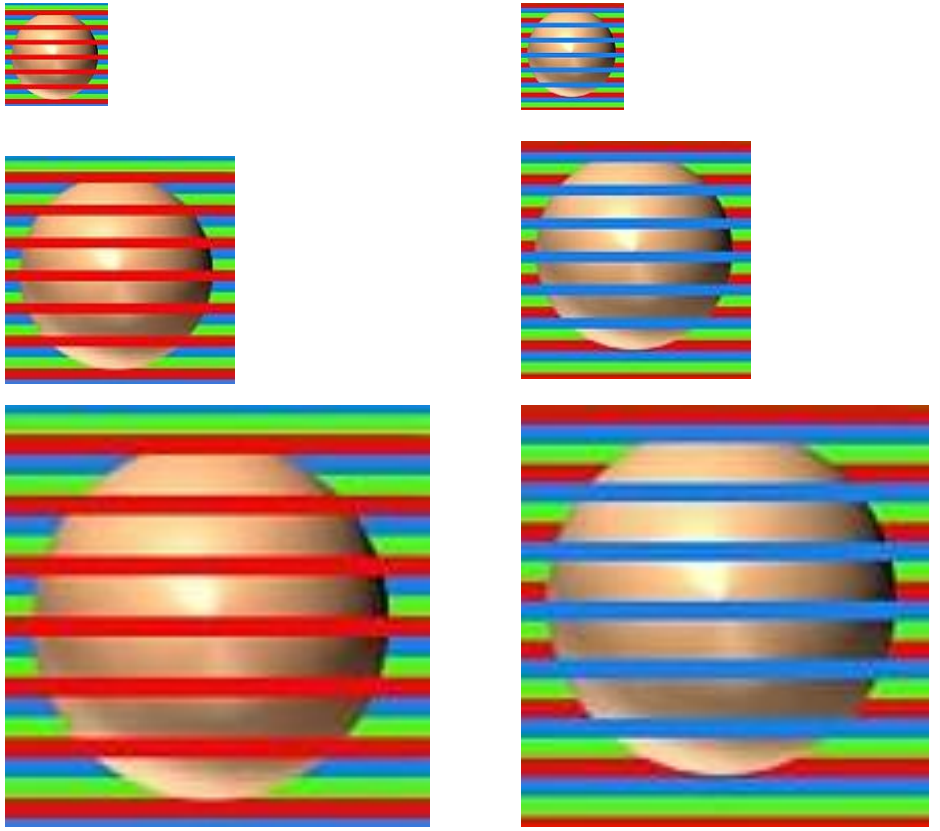
Parvo
Color; High acuity; Slow
Low contrast sensitivity
Ventral

Magno
No color; Low acuity; Fast
High contrast sensitivity
Dorsal

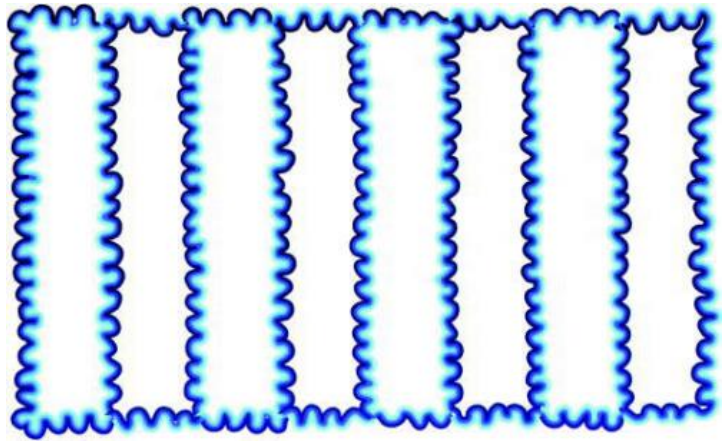

Dorsal (parietal) pathway
MT
Thick stripe
Thin stripe
Interstripe
From interblobs
From blobs
V2
II
III
IVB
IVCα
IVCβ
V
VI
V1
Color
Depth
Direction
Orientation
Complex form
Parvocellular
Magno-cellular
LGN
Midget ganglion cells
Parasol ganglion cells
Parvo (ventral or temporal) pathway
Magno (dorsal or parietal) pathway
Retina
Dorsal (parietal) pathway
MT
V2
LGN
V1
V4
Retina
Ventral (temporal) pathway
V2

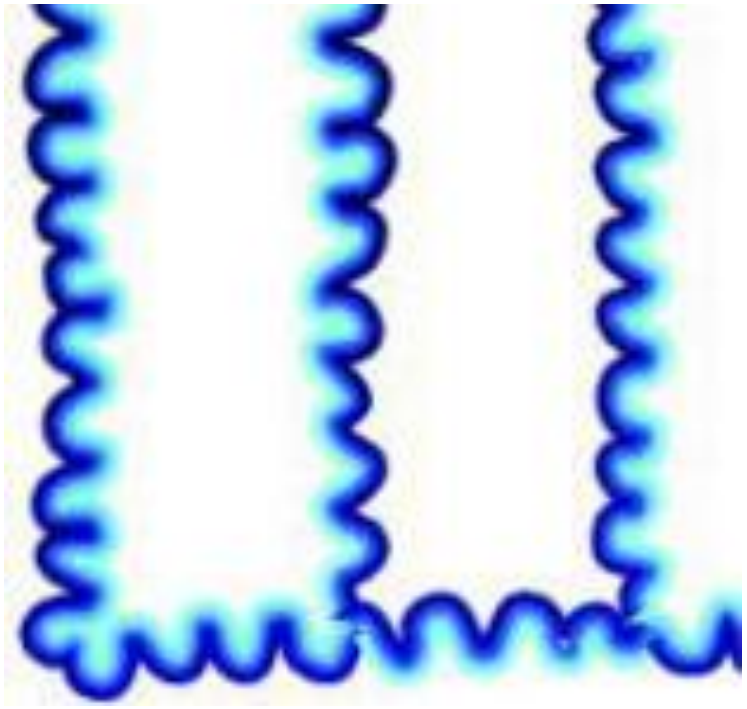
Color vision with lower resolution…



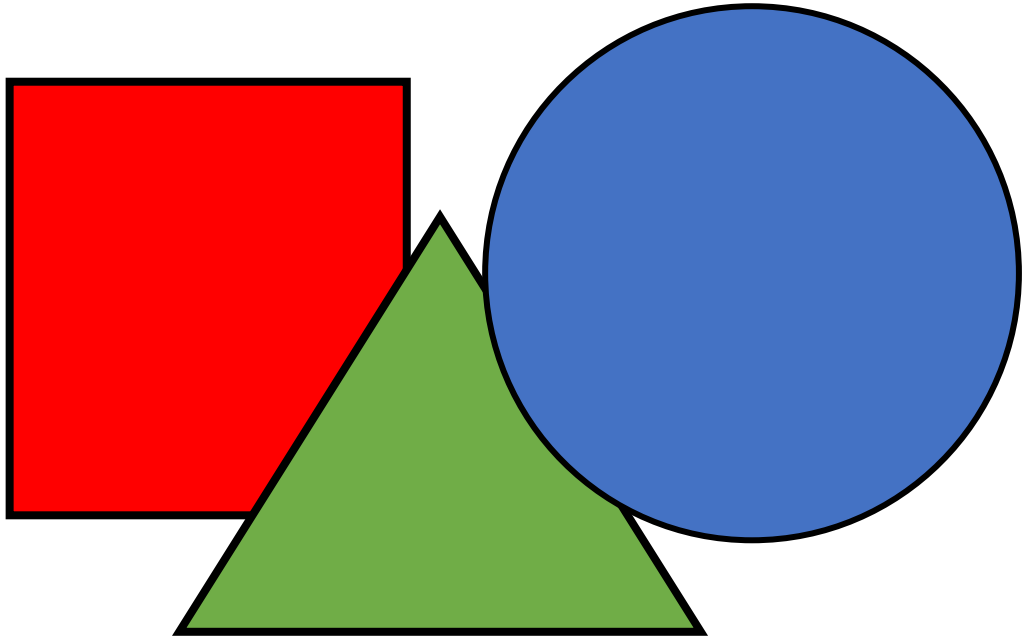


Color vision with lower resolution…

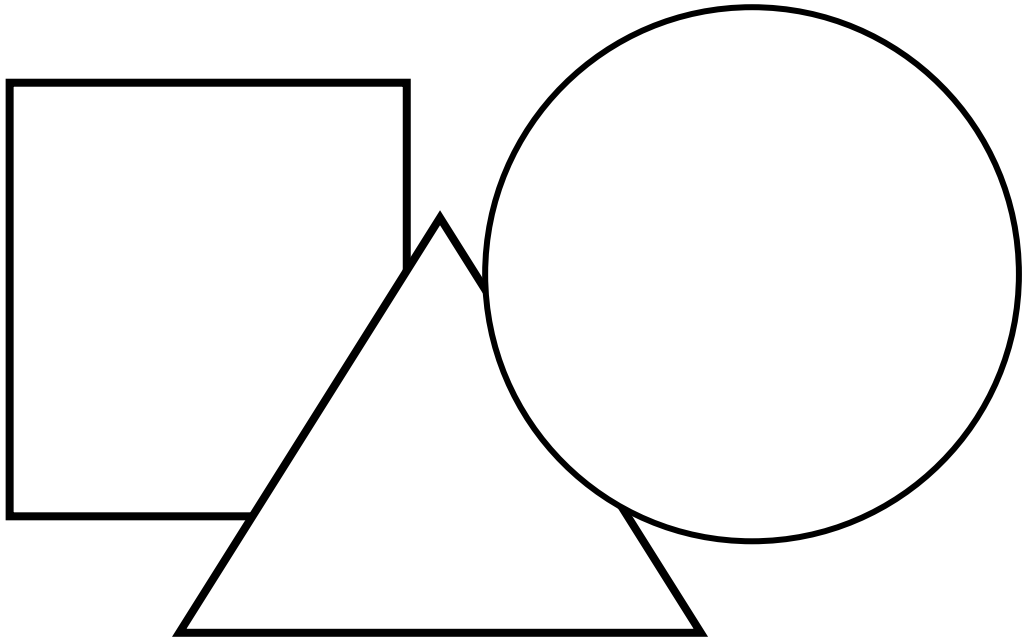


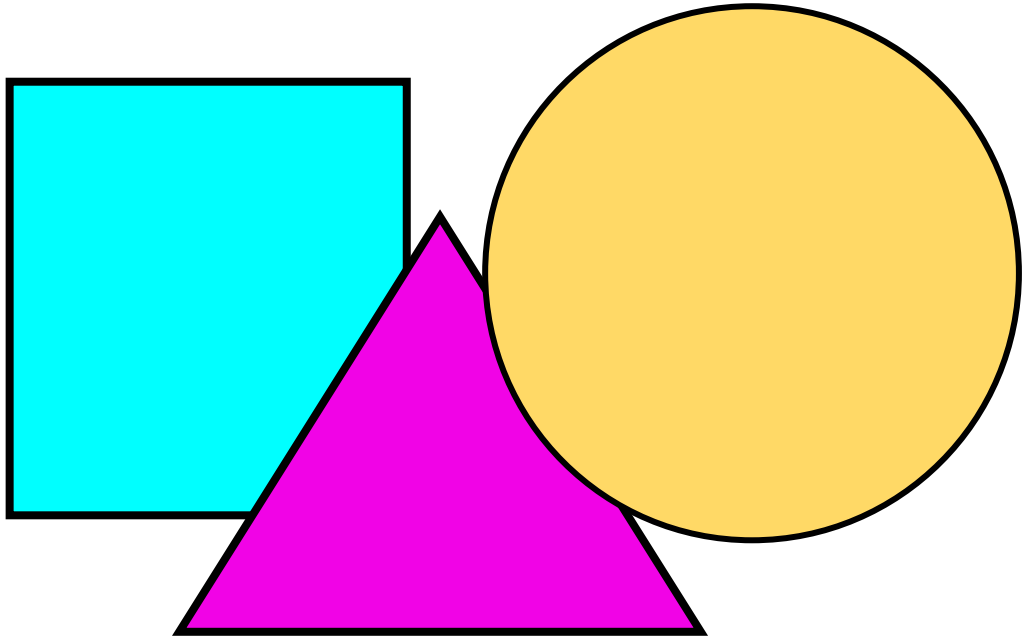


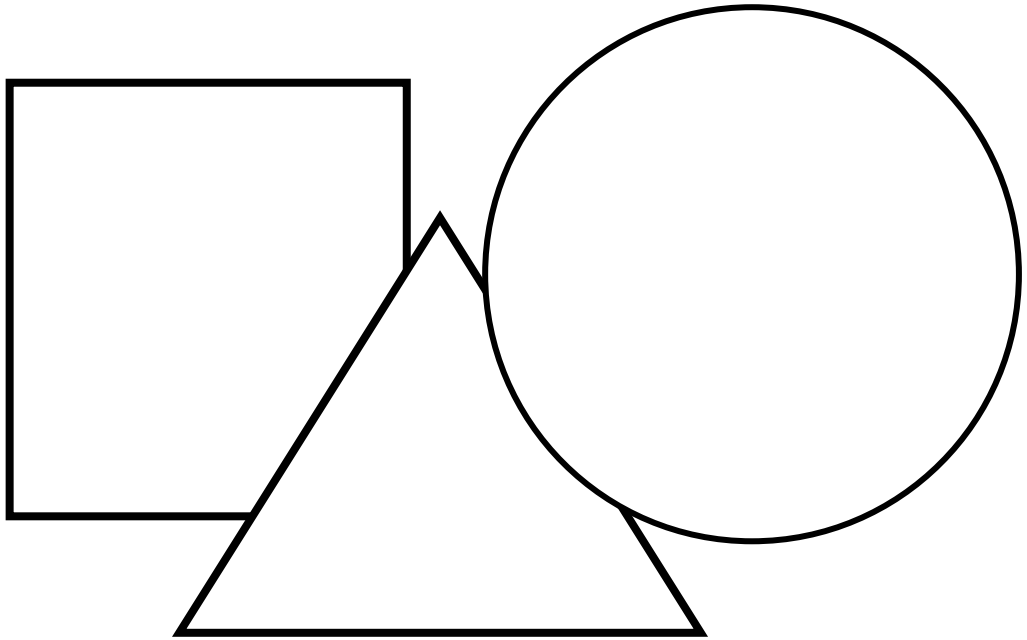
https://www.osapublishing.org/josaa/fulltext.cfm?uri=josaa-22-10-2207&id=85412

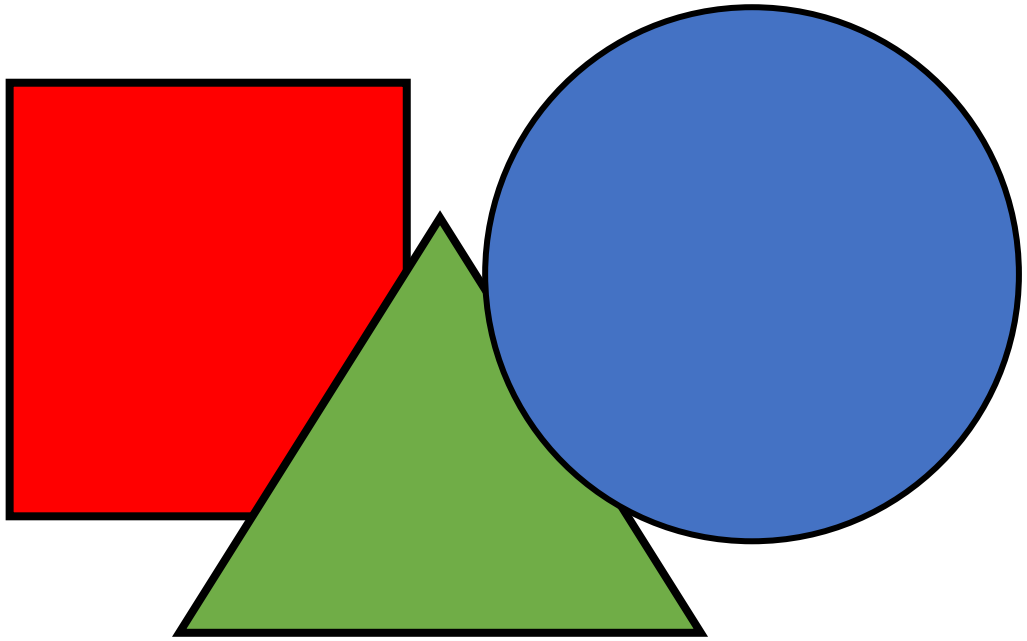


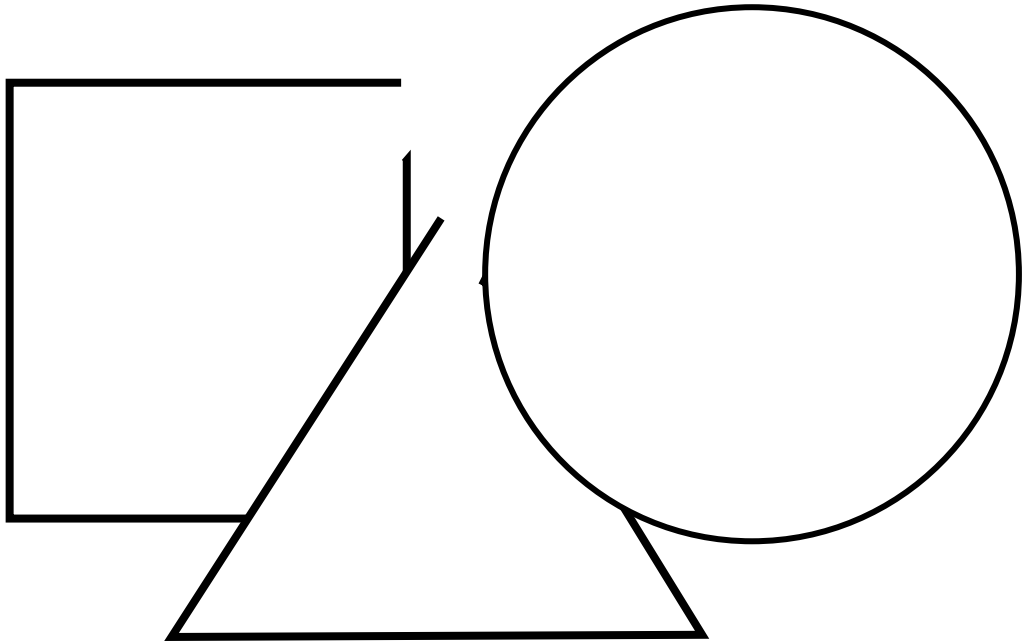
Recreated from Livingstone's





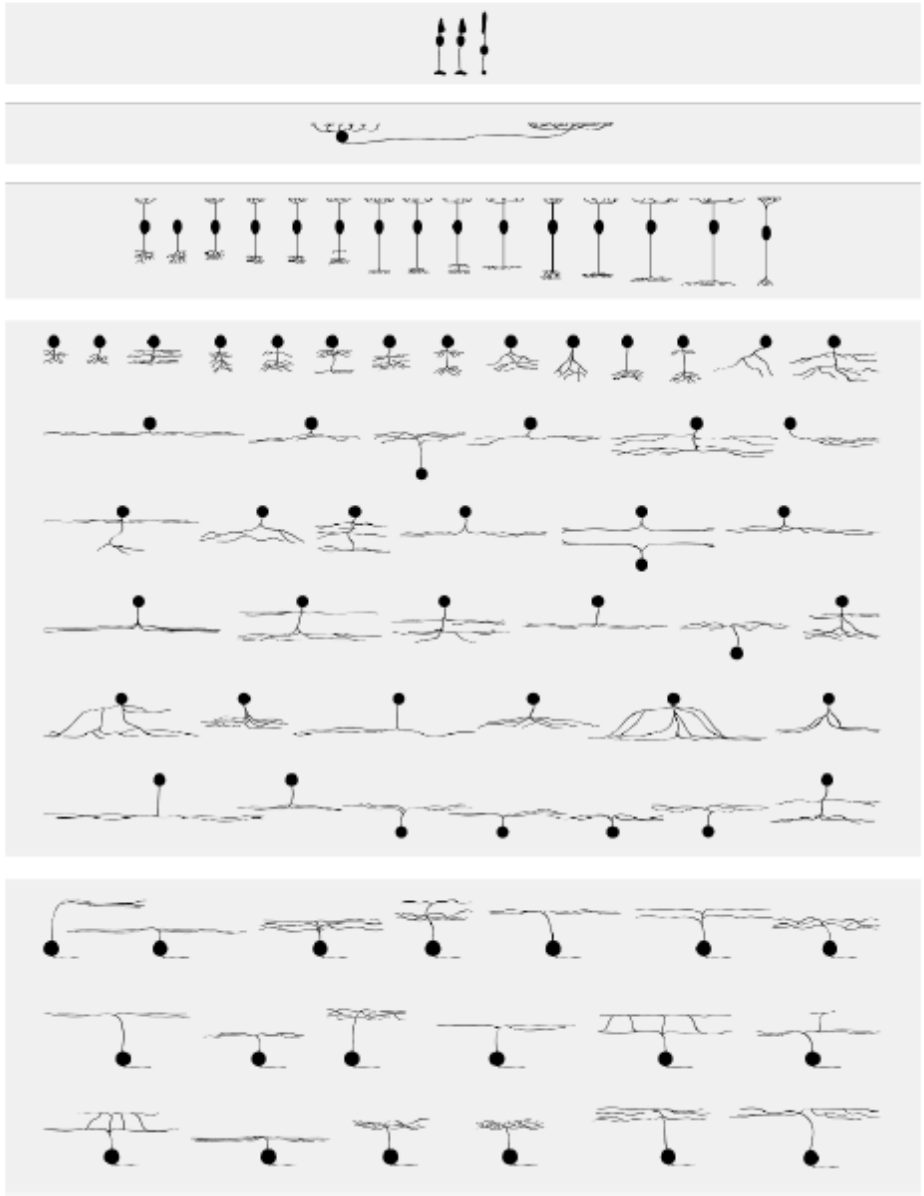






Neuronal diversity

	Cell types
Photoreceptors	3
Horizontal cells	1
Bipolar cells	15
Amacrine cells	>50
Ganglion cells	>40



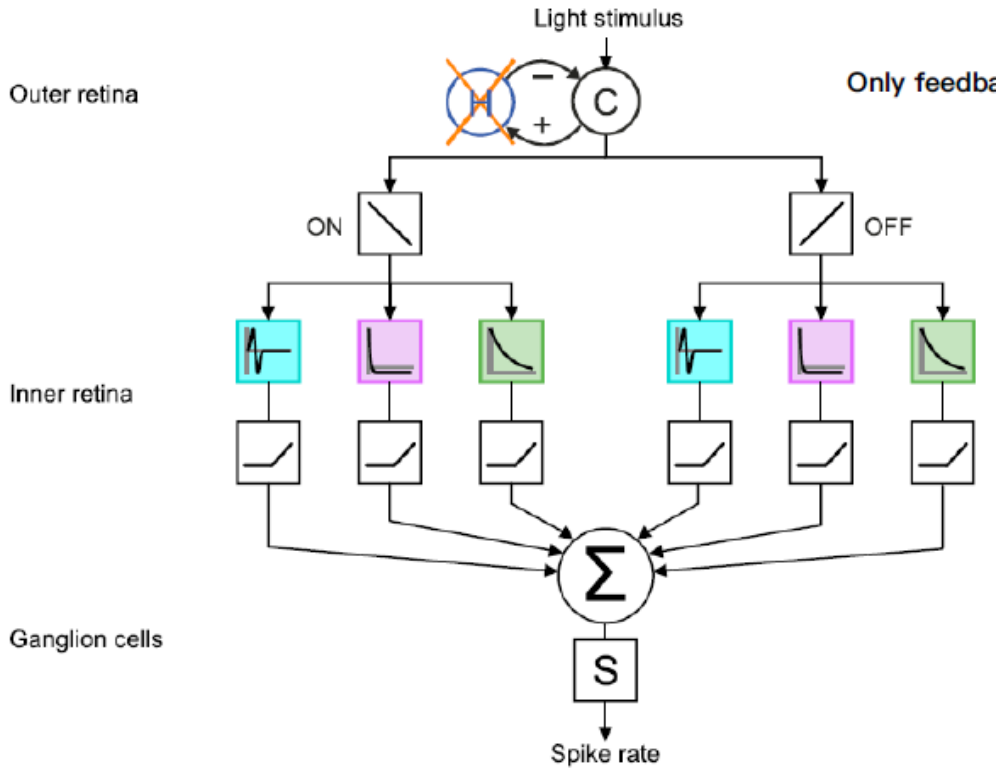
Retina 的celltype 理解的磁场清楚！比其他地方的circuit理解的好！

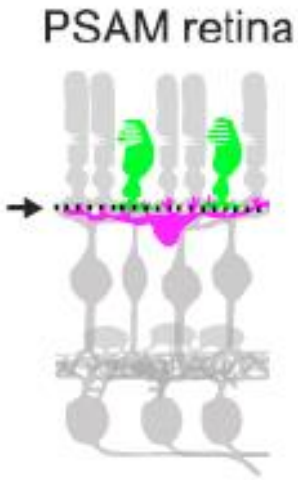
RGCs have a great diversity…

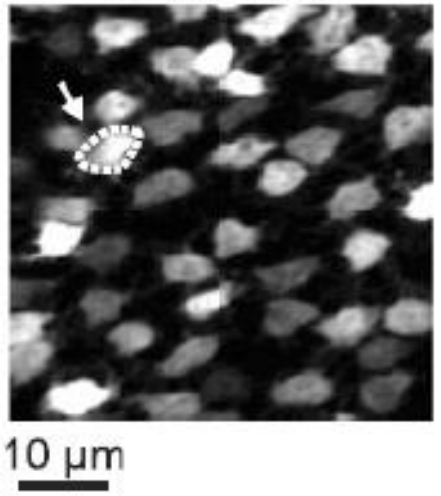
Diversity identified by

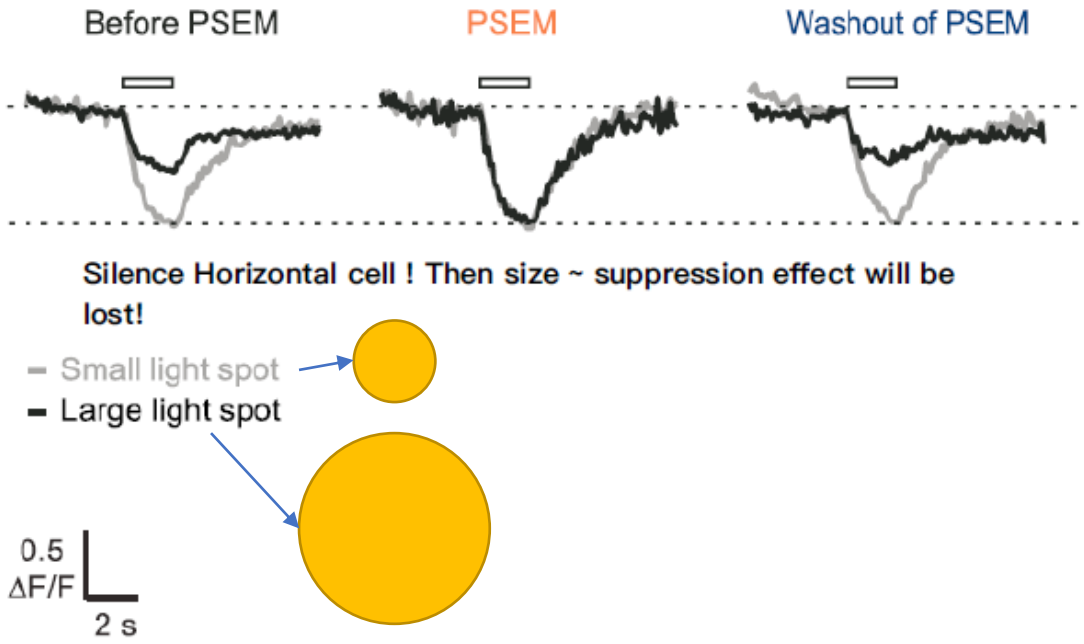
- Gene expression
- Morphology
- Response property
- Function (computation)

Surround Suppression Implemented by Horizontal Cells

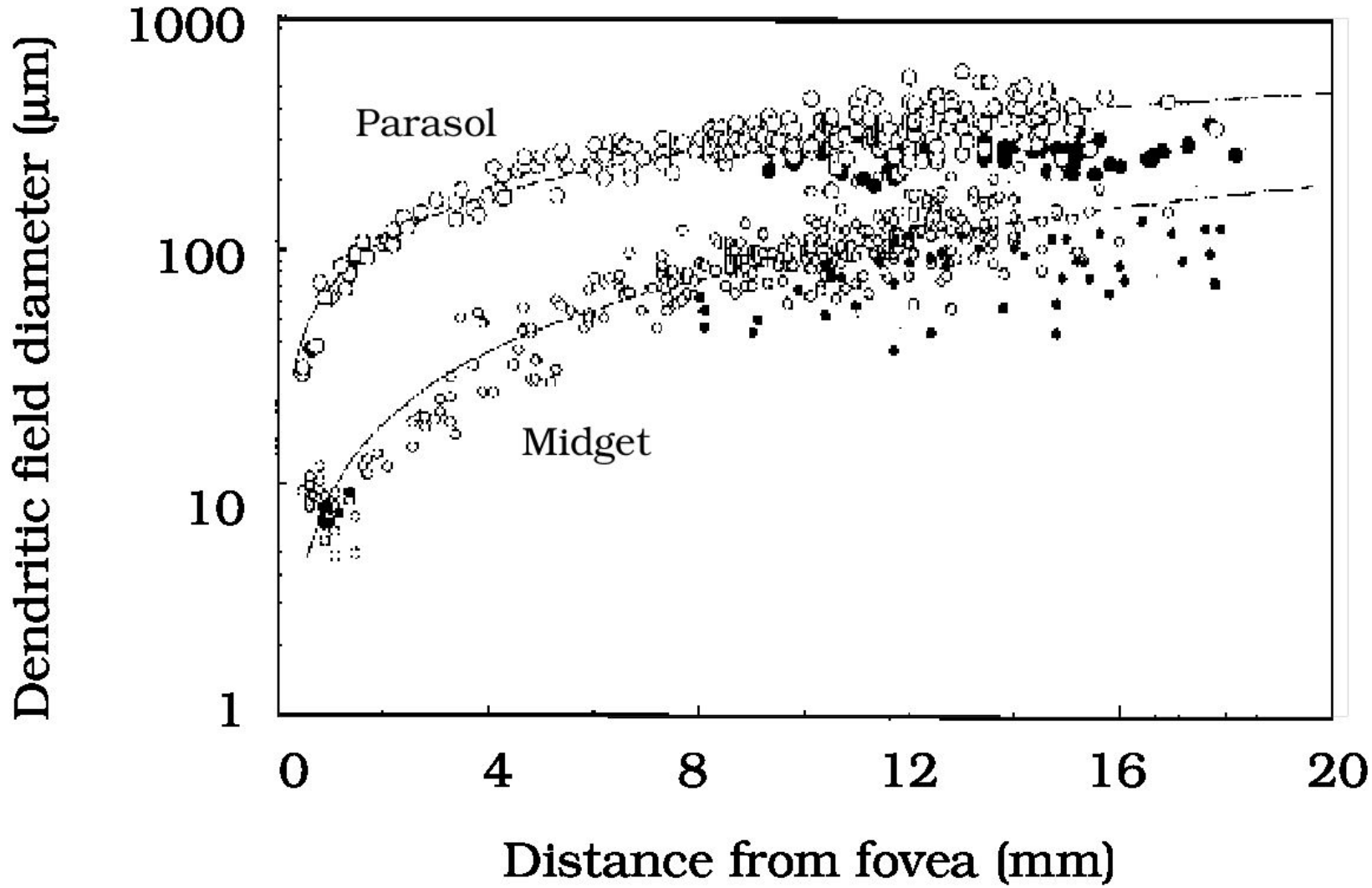








Drinnenberg et al., 2018


Dendritic field diameter (μm)
1000
100
10
1
Parasol
Midget
0
4
8
12
16
20
Distance from fovea (mm)