Deep Learning

Jonathon Hare

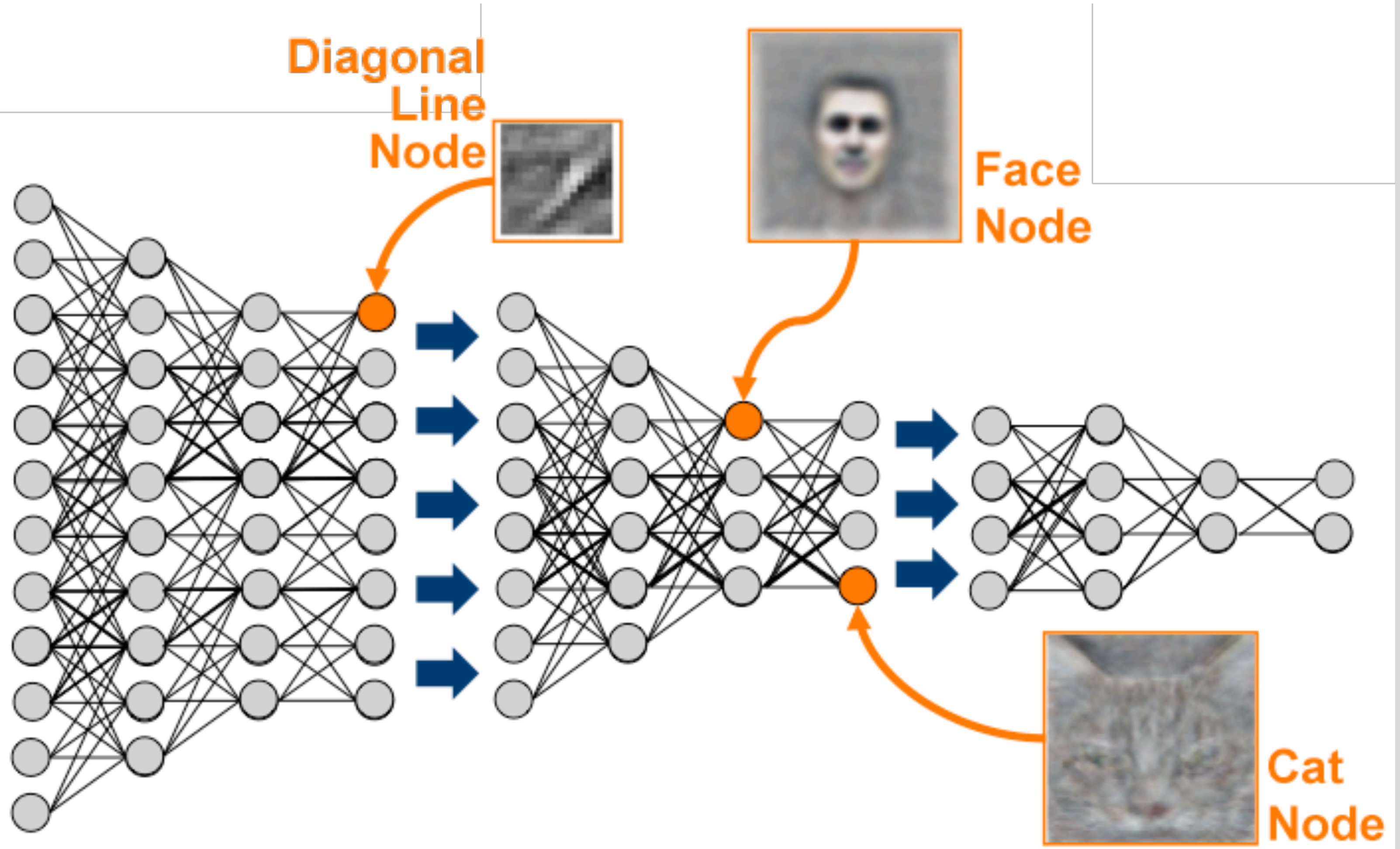
Contents

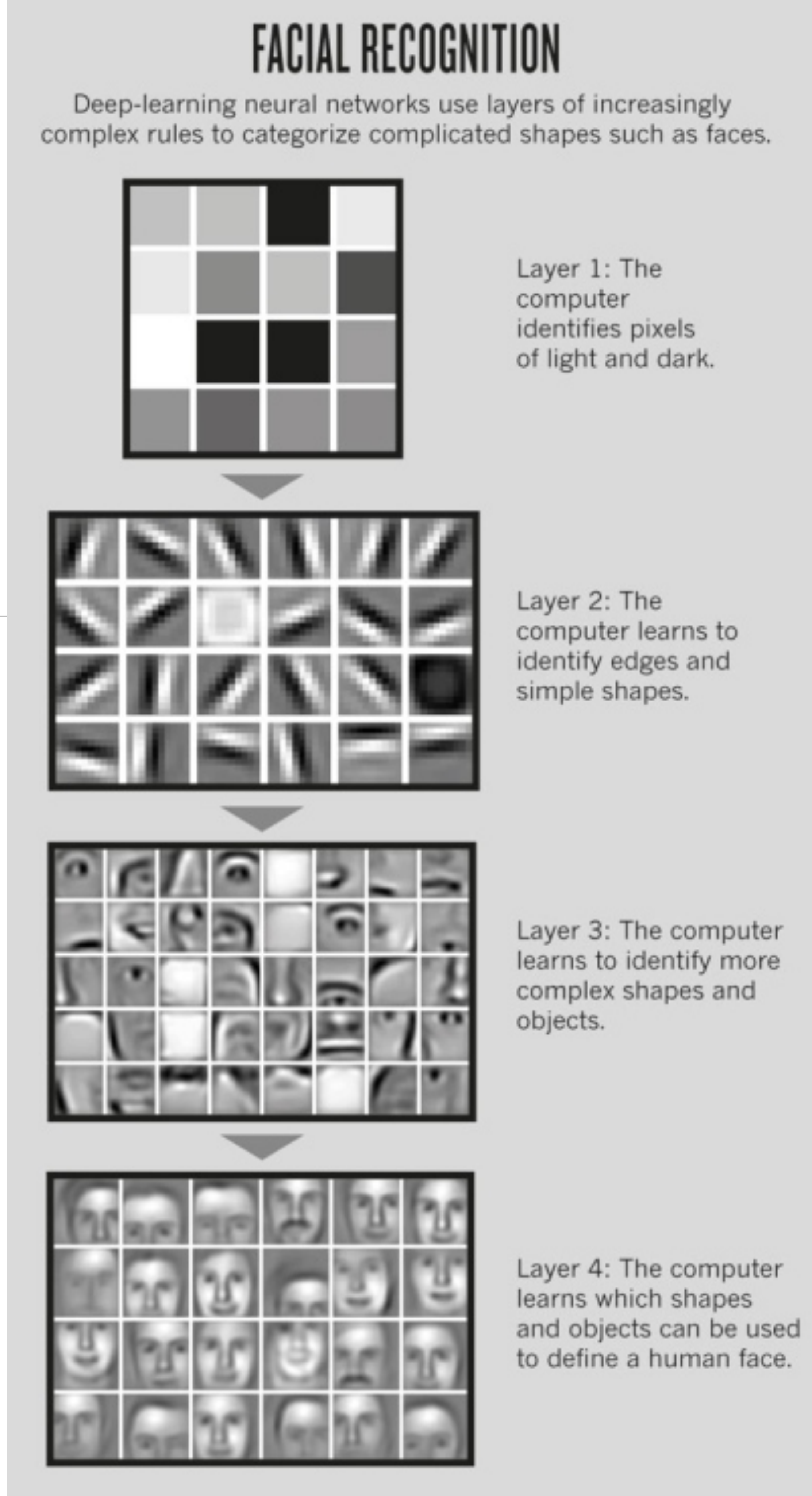
- What is “Deep Learning”
- Motivation and practical issues
- Techniques and nomenclature
  - CNNs, RNNs, Backpropagation, Loss functions
- Technologies

What is “Deep Learning”?

- Learning using Neural Networks
  - But…
    - Lots of layers
      - **(network is deeper through layers than a simple MLP)**
      - Focus on letting the network learn the features rather than manually engineering them
    - Or, network makes use of its past predictions
      - **(network is deep through time)**

Deep learning: learning layers of features

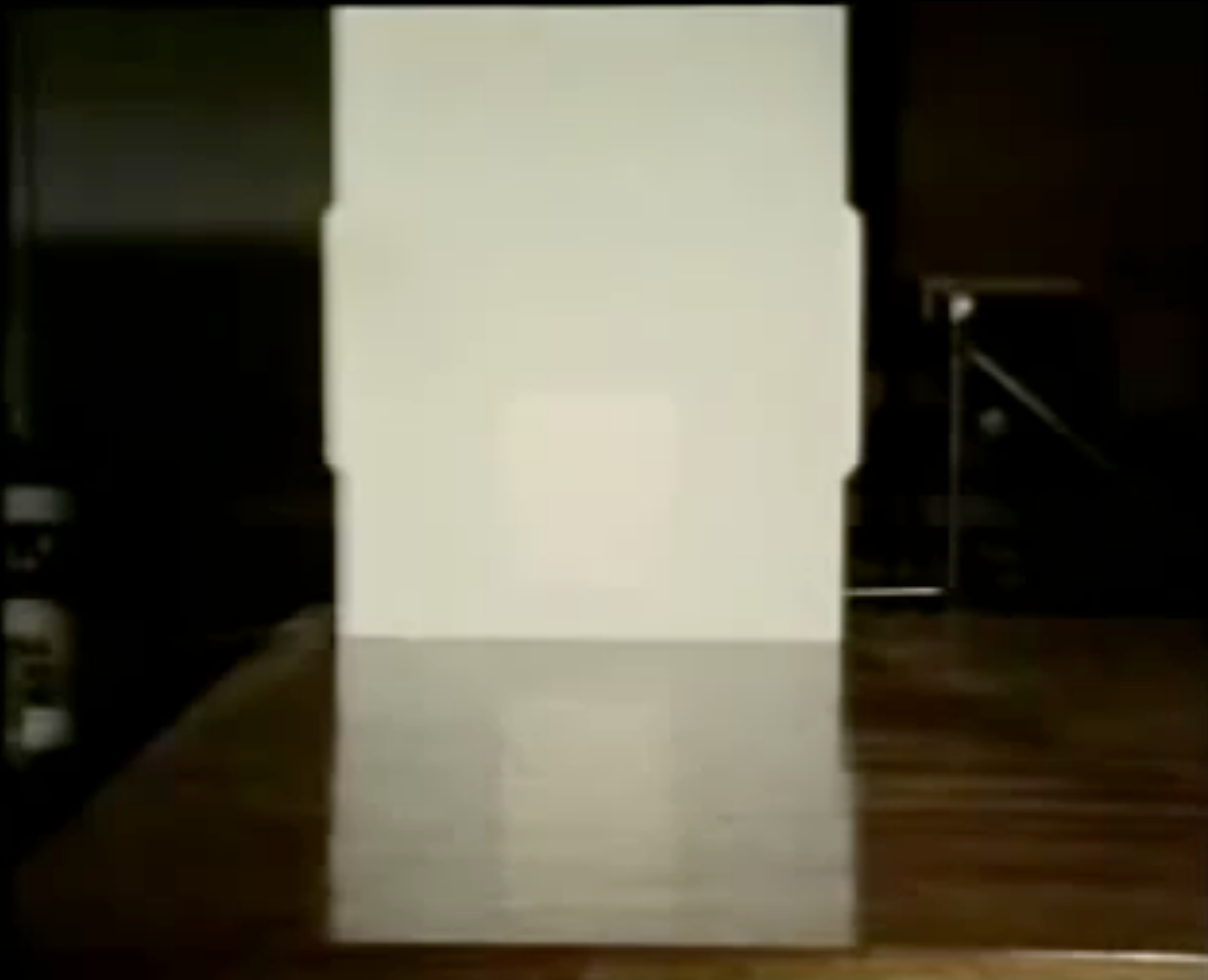




Motivation

Is vision innate or acquired?

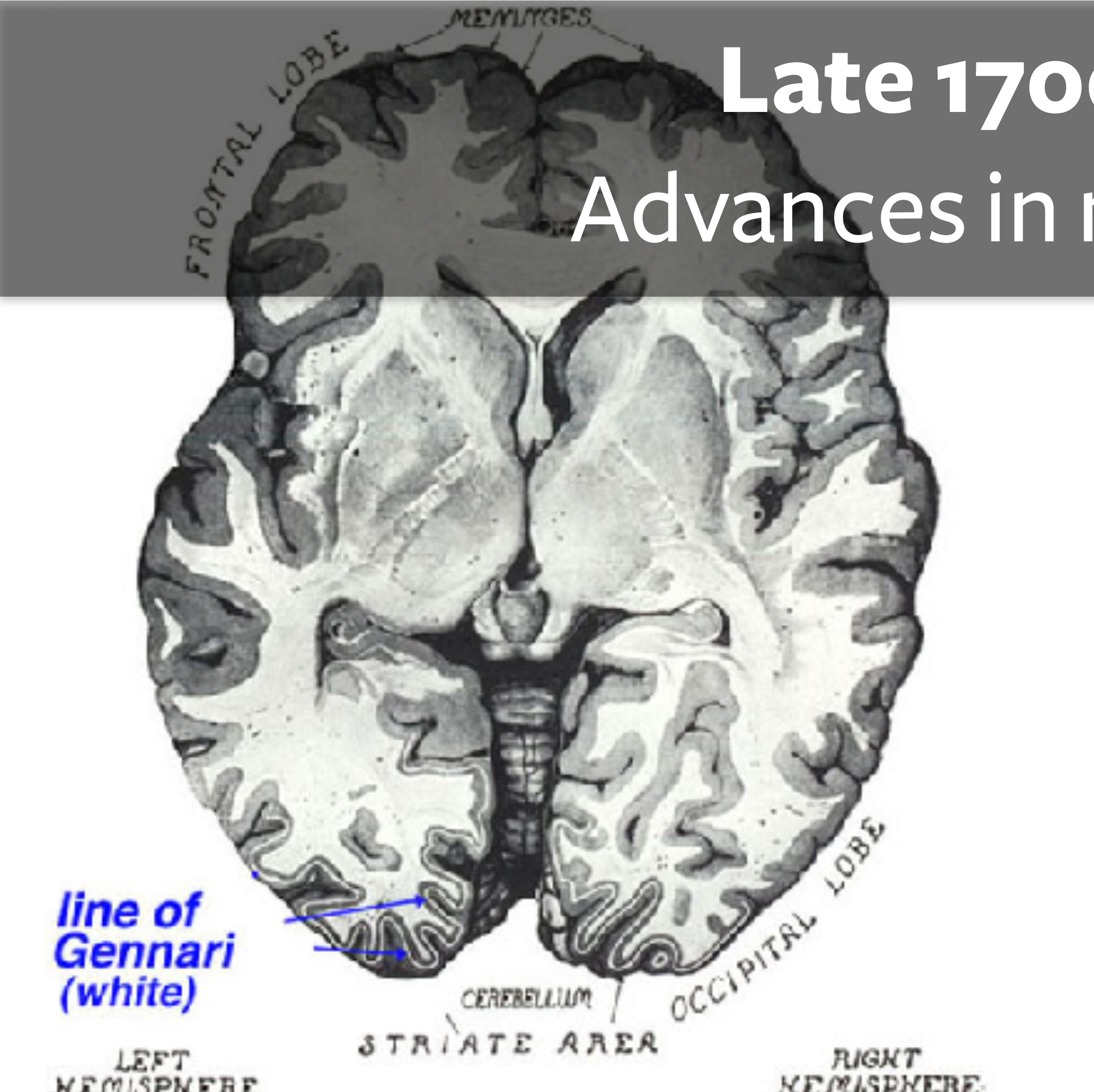
Colin Blakemore

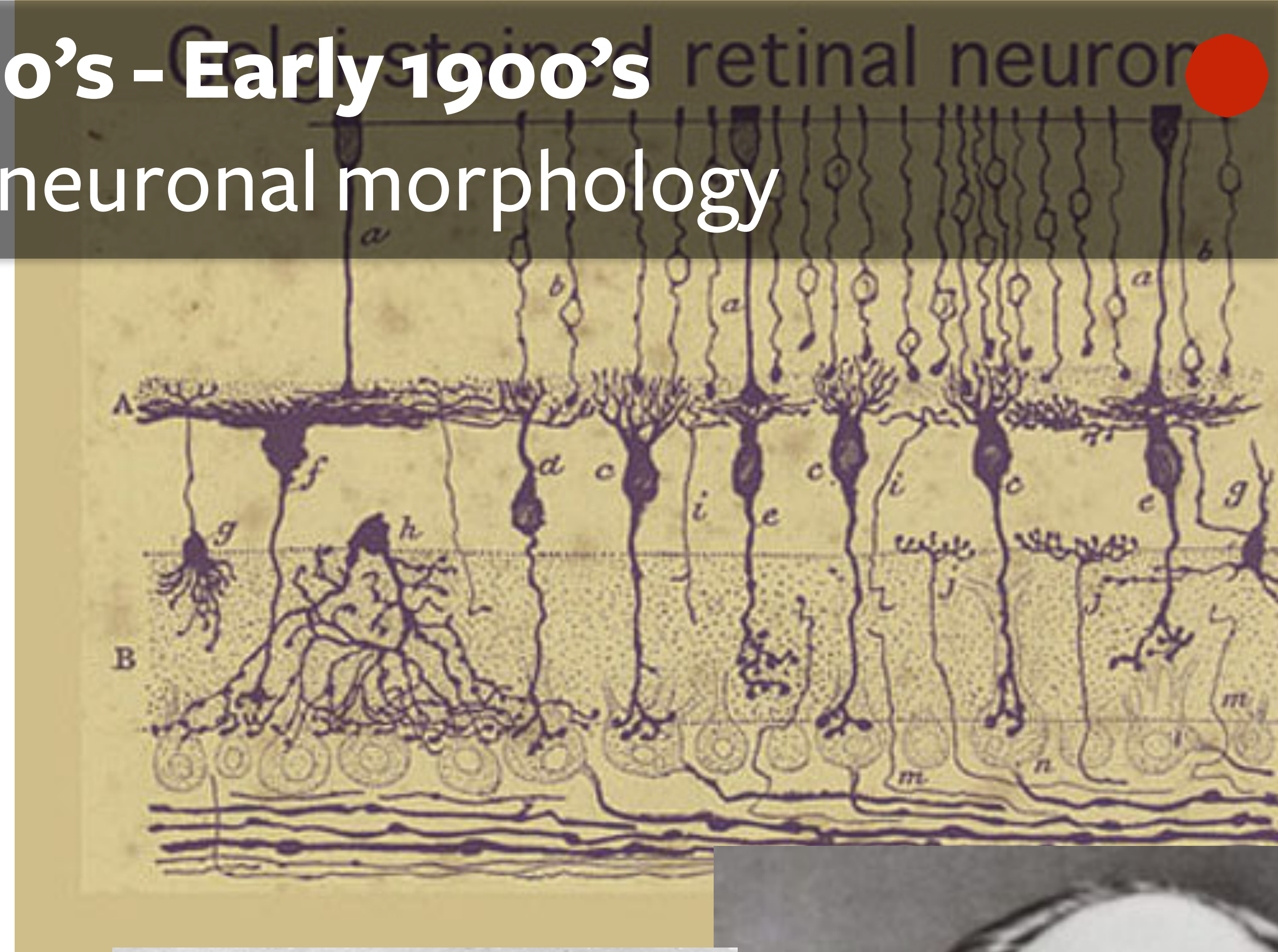


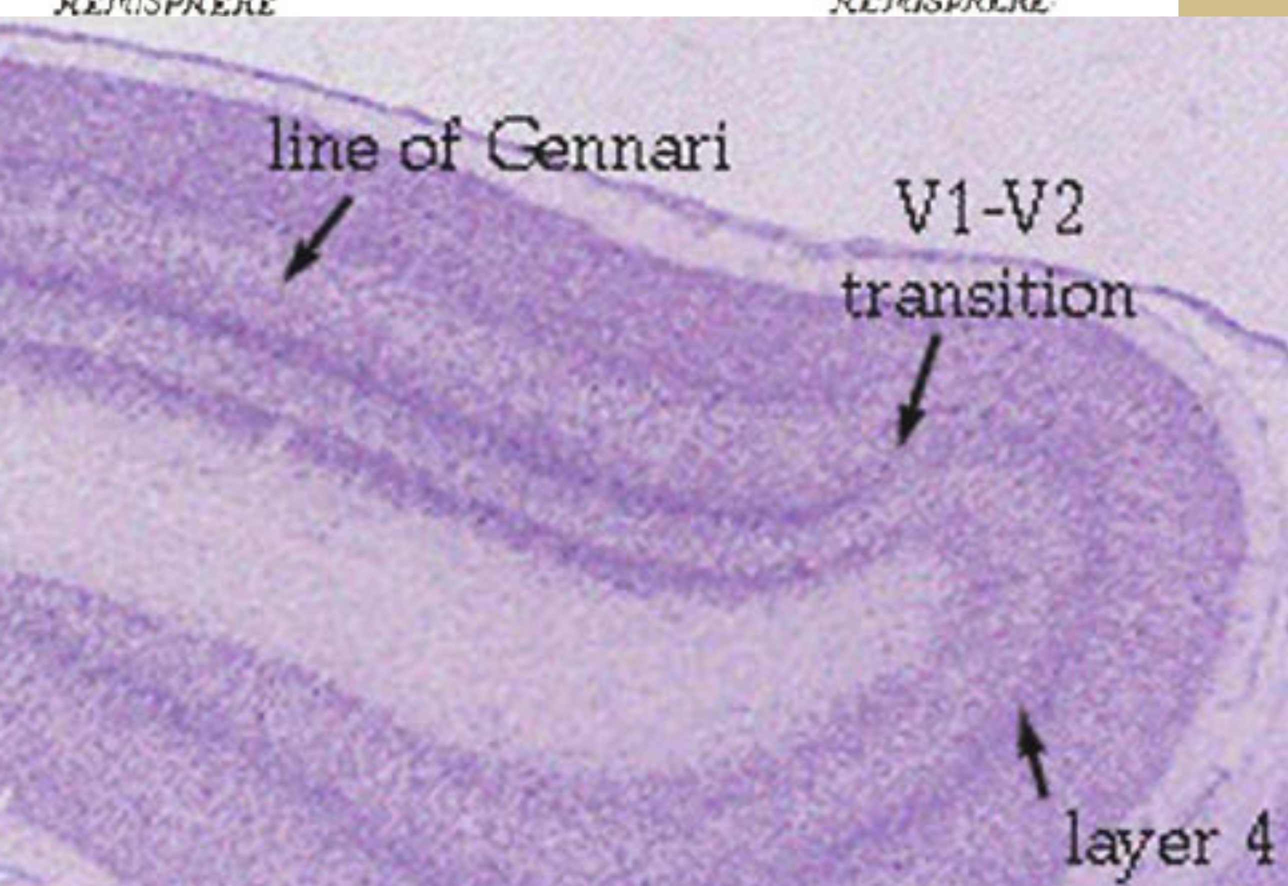
Historical Context

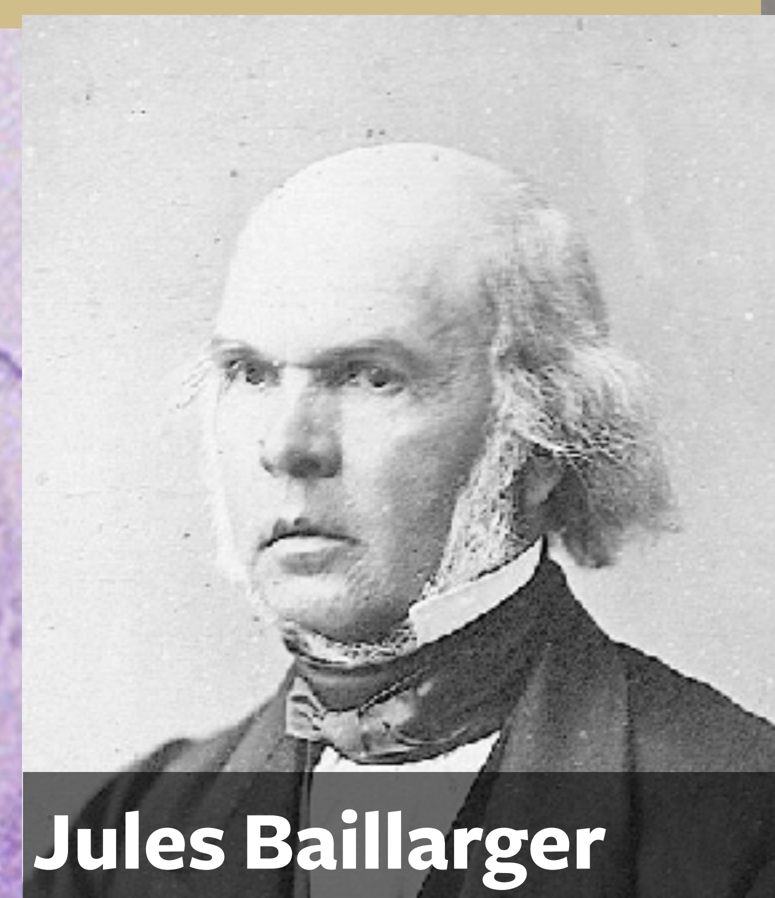
Late 1700's - Early 1900's

Advances in neuronal morphology





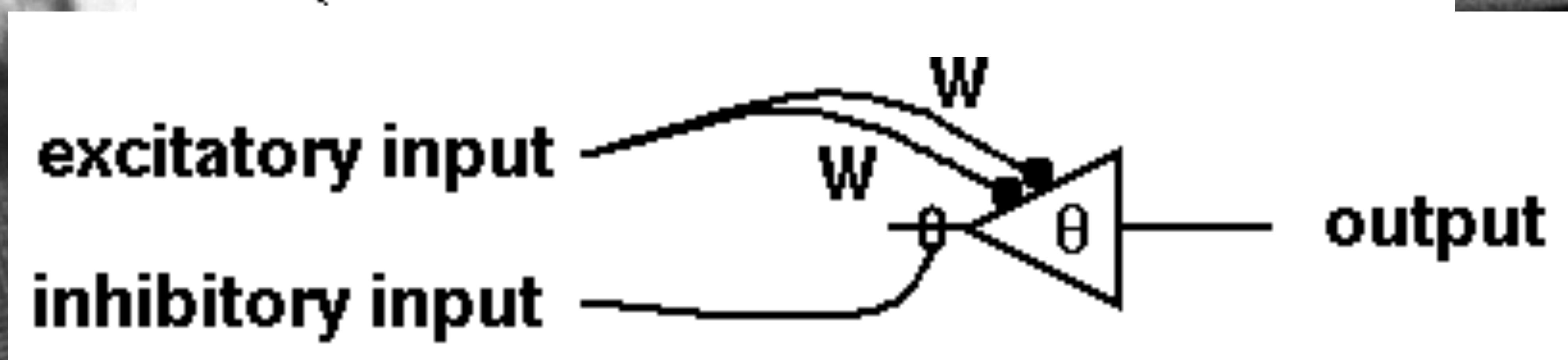


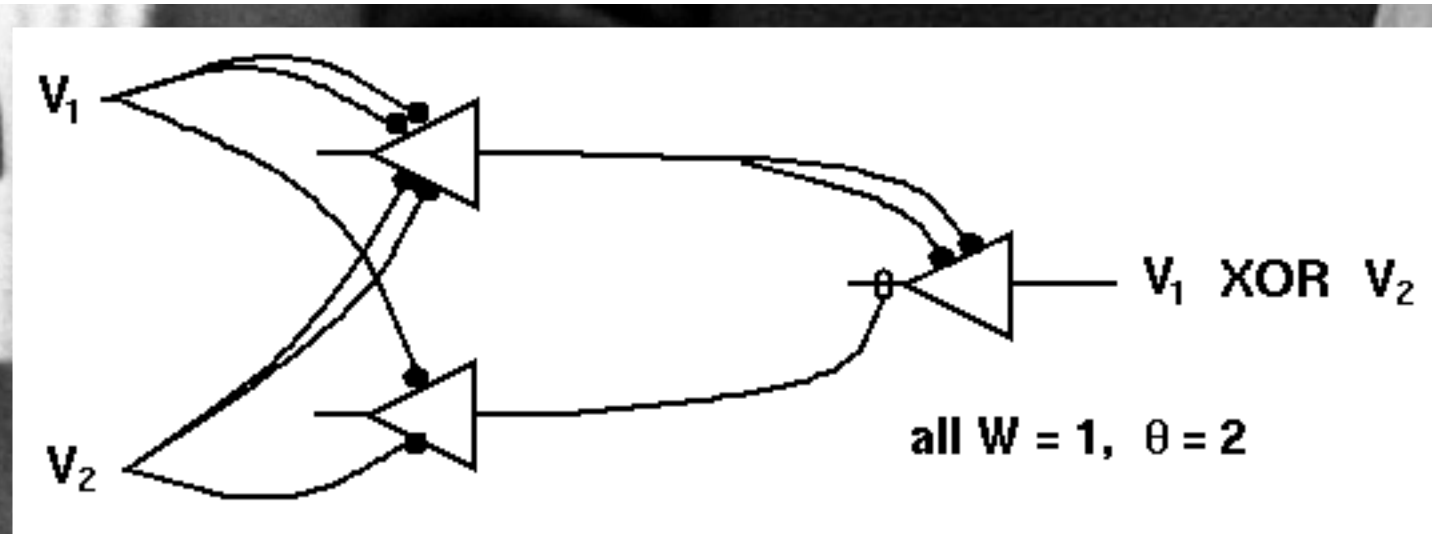

Jules Baillarger


Santiago Ramón y Cajal

1943
McCulloch-Pitts Artificial Neuron

$$V_i = \begin{cases} 1 & : \sum_j W V_j \geq \theta \text{ AND no inhibition} \\ 0 & : \text{otherwise} \end{cases}$$

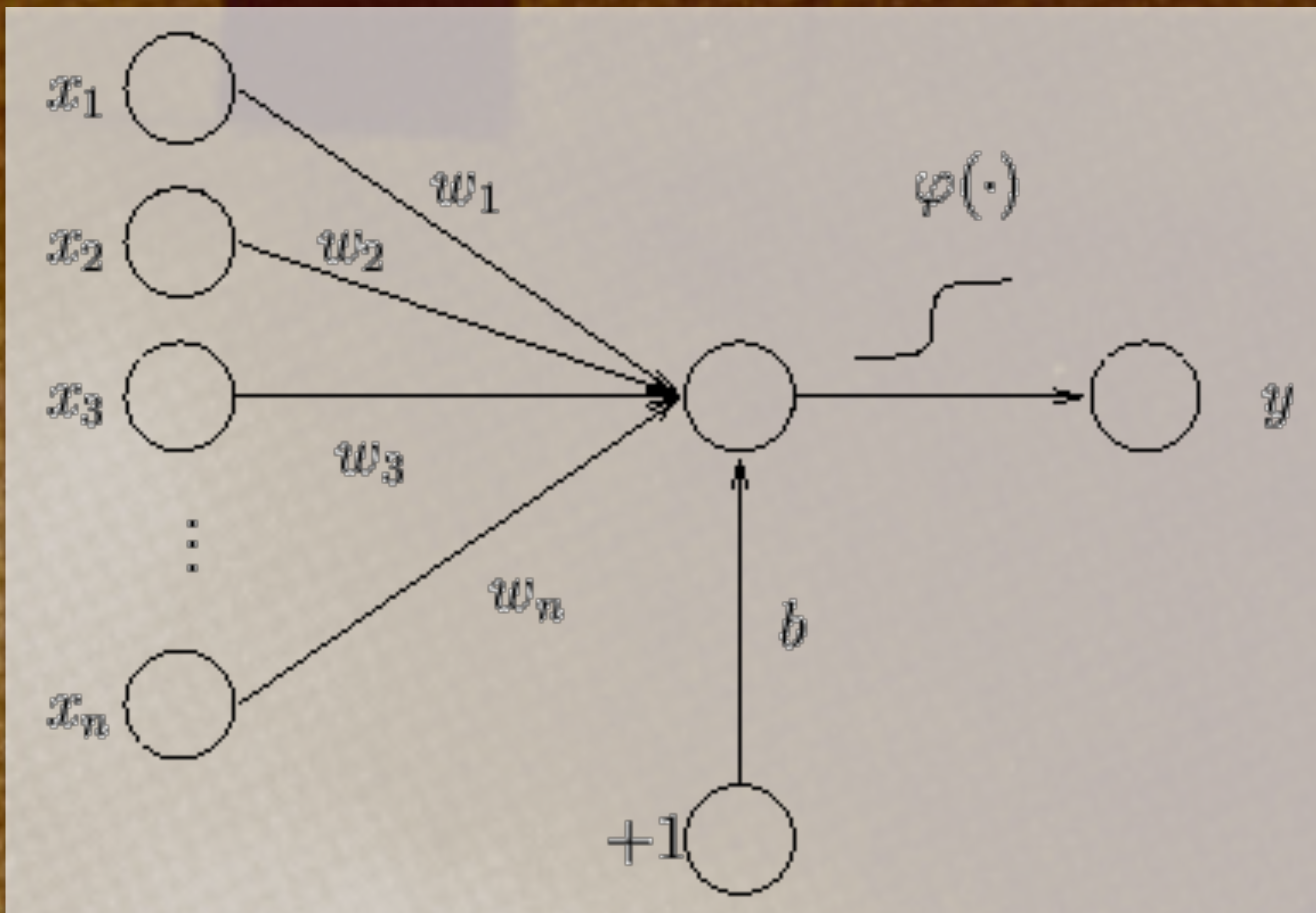


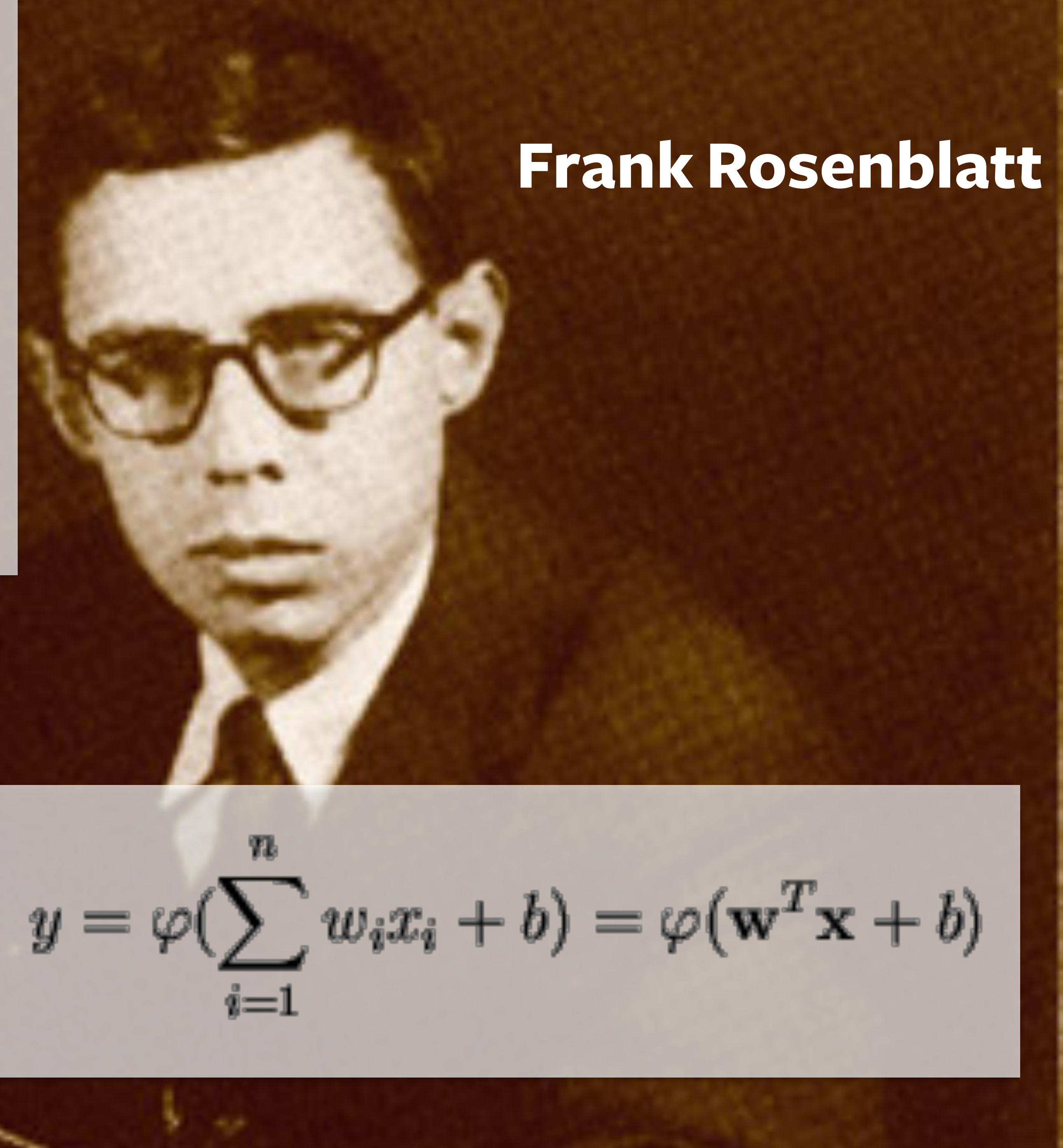


Warren McCulloch

Walter Pitts

1958
Rosenblatt's Perceptron

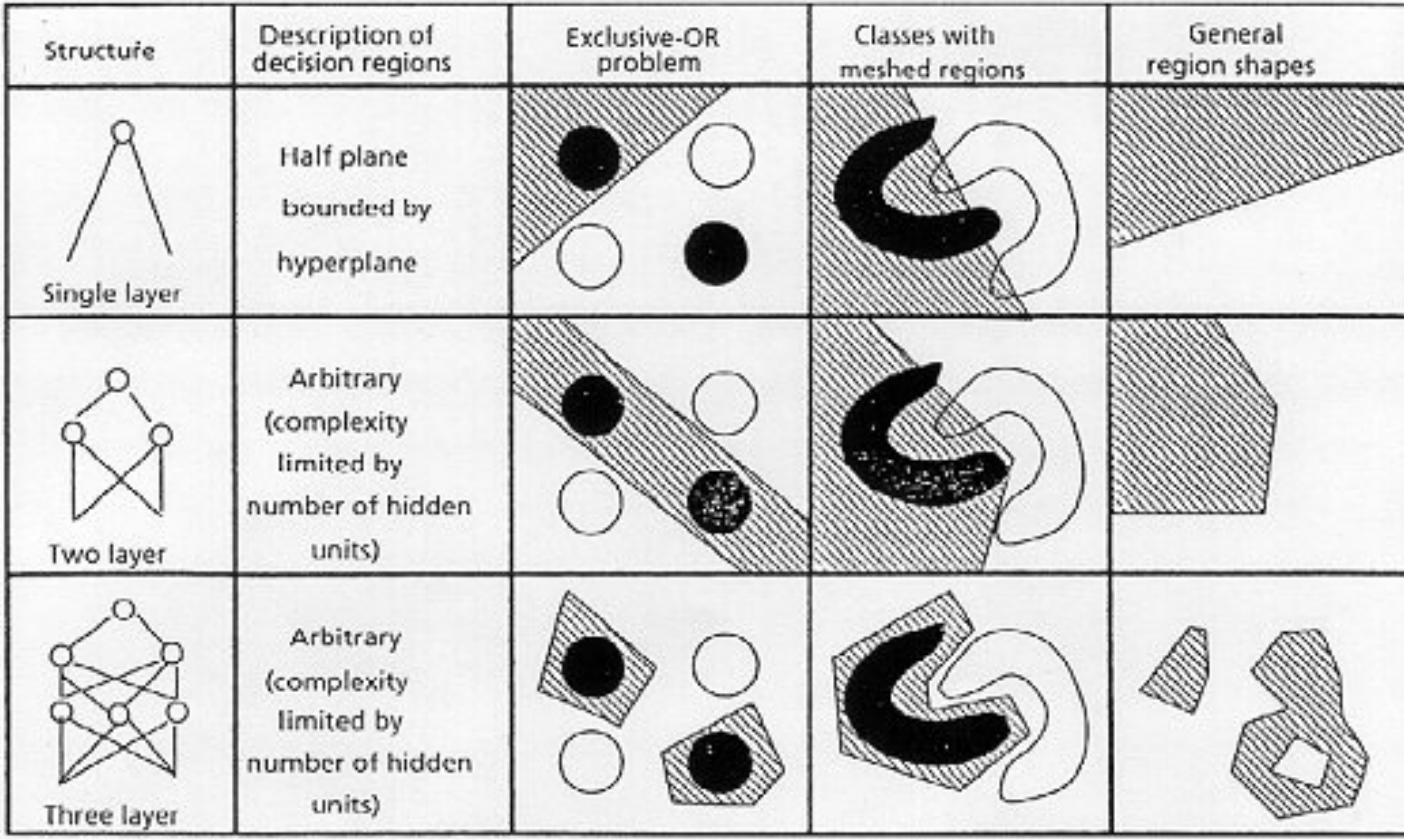




Frank Rosenblatt

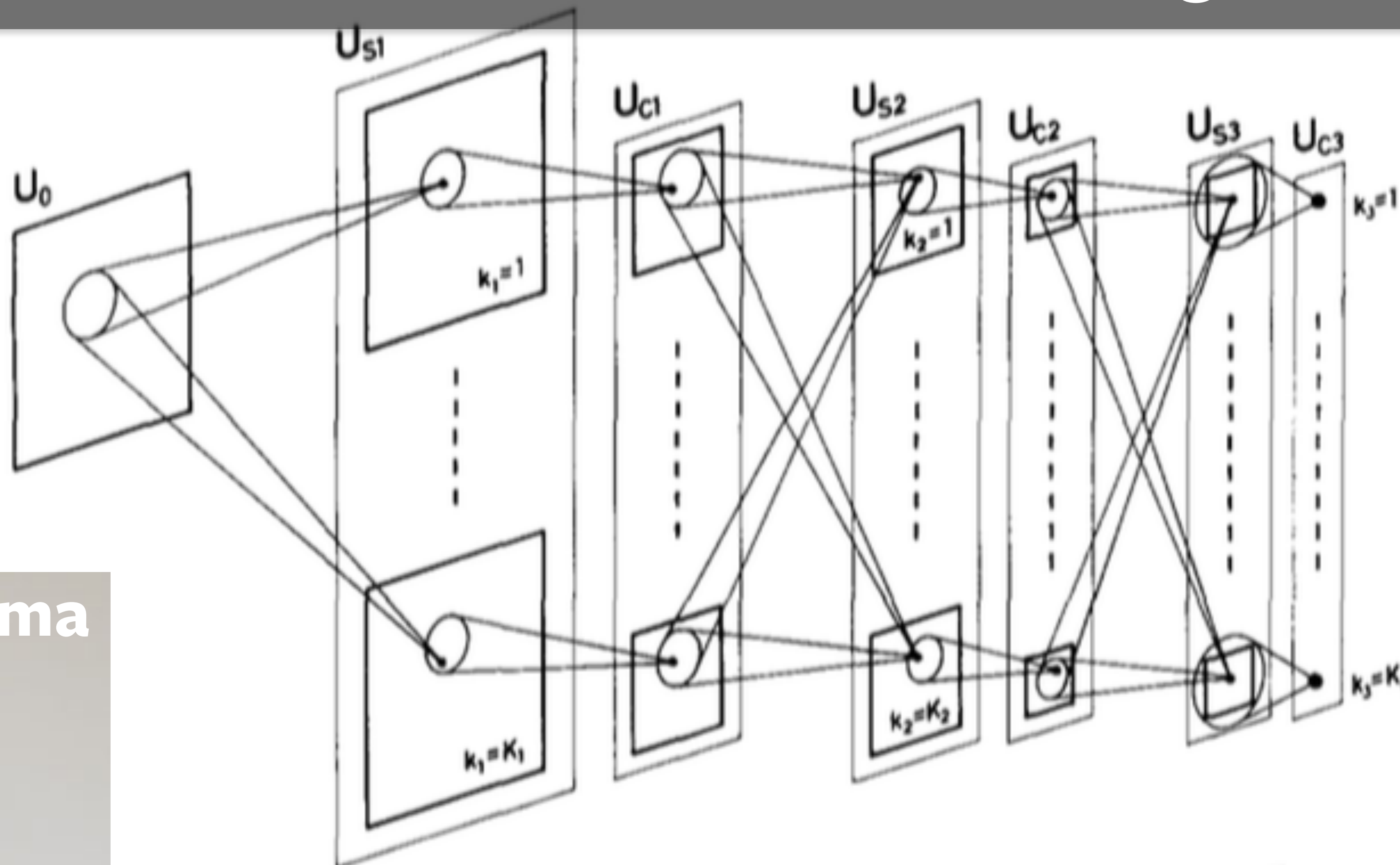
$$y = \varphi\left(\sum_{i=1}^{n} w_i x_i + b\right) = \varphi(\mathbf{w}^T \mathbf{x} + b)$$

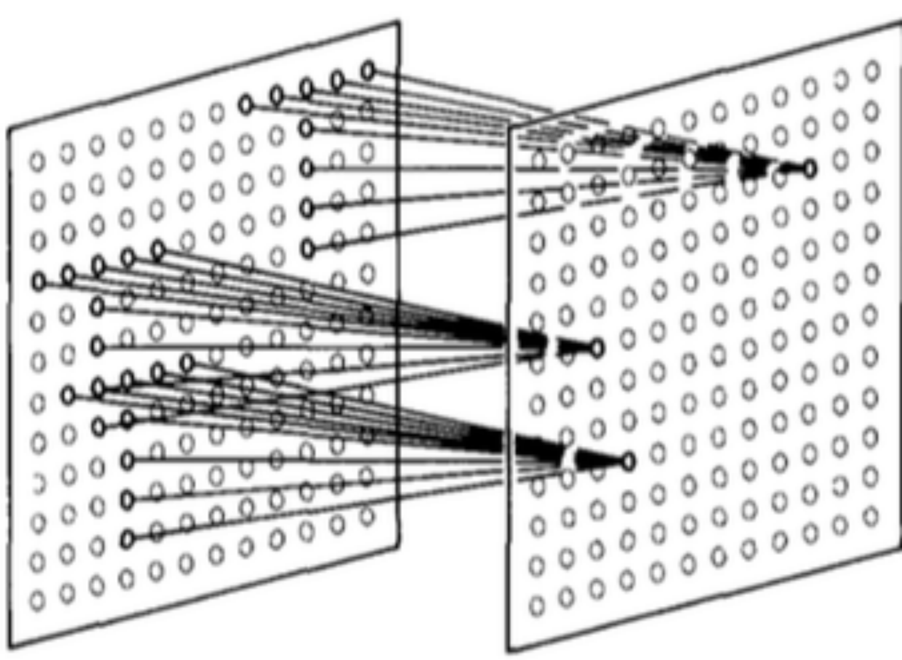

MARK I PERCEPTRON
CORNELL AERONAUTICAL LABORATORY, Inc.



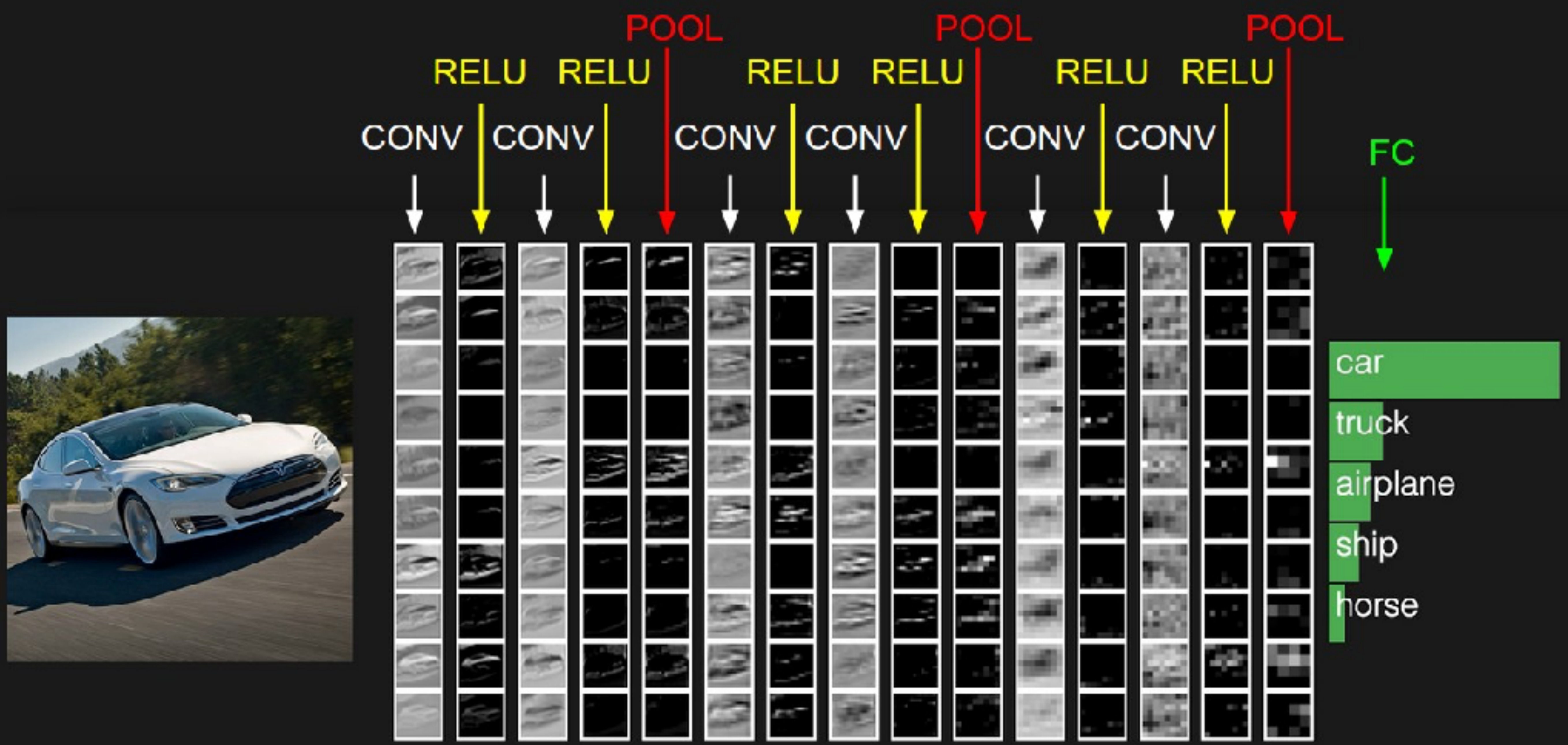
Structure
Description of decision regions
Exclusive-OR problem
Classes with meshed regions
General region shapes
Single layer
Half plane bounded by hyperplane
Two layer
Arbitrary (complexity limited by number of hidden units)
Three layer
Arbitrary (complexity limited by number of hidden units)

1979
Convolutional Neural Networks & Neocognitron



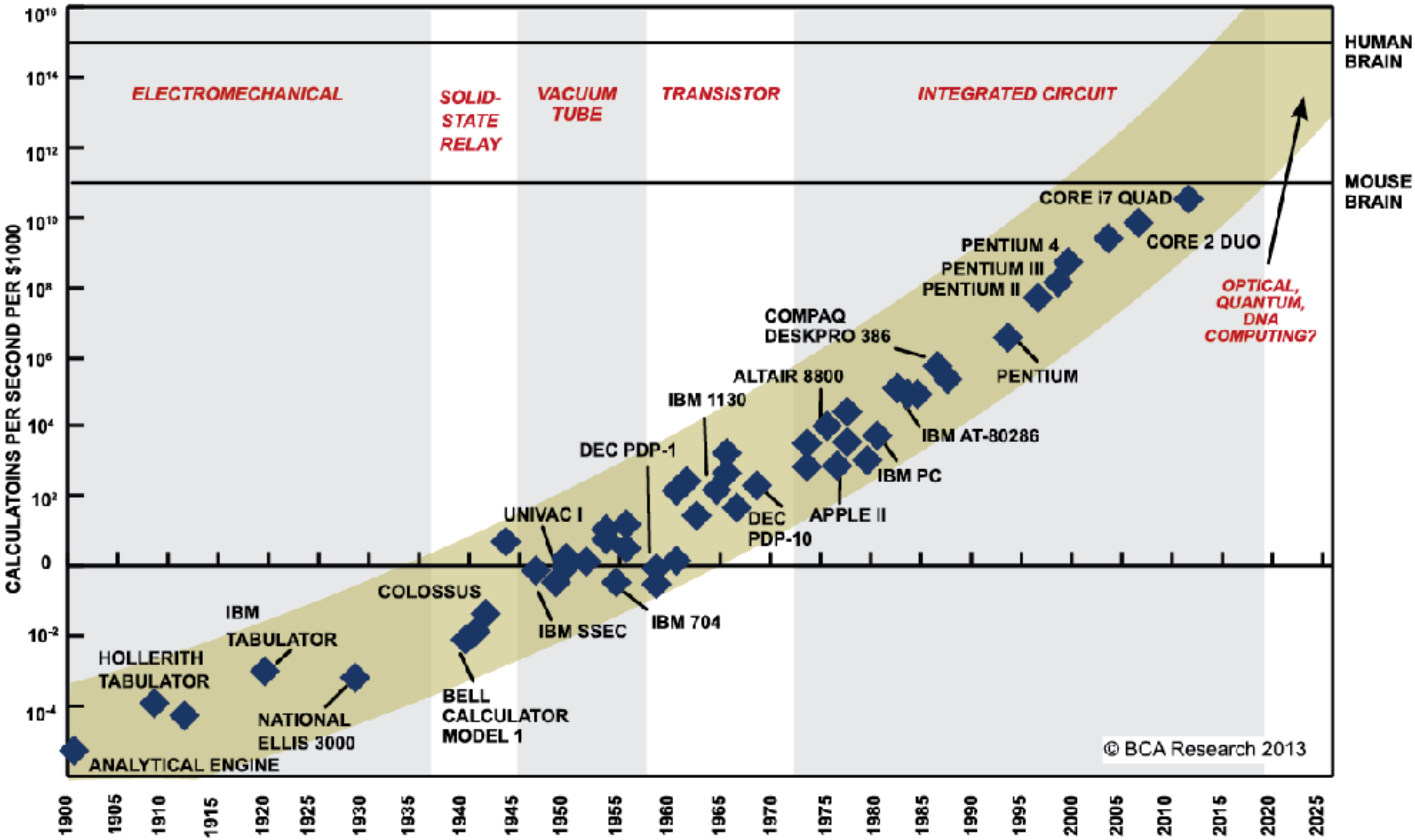


Kunihiko Fukushima



POOL
POOL
POOL
RELU
RELU
RELU
RELU
RELU
RELU
CONV
CONV
CONV
CONV
CONV
CONV
FC
car
truck
airplane
ship
horse

Moore’s Law



SOURCE: RAY KURZWEIL, "THE SINGULARITY IS NEAR: WHEN HUMANS TRANSCEND BIOLOGY", P.67, *THE VIKING PRESS*, 2006. DATAPOINTS BETWEEN 2000 AND 2012 REPRESENT BCA ESTIMATES.

The new Moore's Law: Computer's no longer get faster, just wider

Okay, but why do we care?

Overview of ideas and nomenclature

- Construction
- Learning - “Backpropagation”
- Common layers
- Typical Composition
- Training:
    - Optimisers
    - Avoiding overfitting: dropout and Batch Normalisation
    - Initialisation of weights

Dense Layers

Convolutional Layers

Optimisers

Dropout

Batch Normalisation

Initialisation

Technologies

Keras

PYTORCH

theano

Microsoft CNTK

TensorFlow

torch

neon

Caffe

mxnet

Lab 1: Learning to Deep Learn using Python, Keras, TensorFlow and a GPU

Lets get hands-on

- We'll learn:
  - How to use the MNIST dataset in Keras.
  - How to develop and evaluate a baseline neural network model for the MNIST problem.
  - How to switch the backends used by Keras and run your code on the GPU.
  - How to implement and evaluate a simple Convolutional Neural Network for MNIST.
  - How to implement a close to state-of-the-art deep learning model for MNIST.
  - How to serialise and deserialise trained models.
  - How to load your own image created outside of the MNIST dataset, and pass it through the network.
  - How to visualise the filters learned by the network.
  - How to implement networks with branching and merging.

Transfer Learning

Lab 2: Transfer Learning using Keras

Lets get hands-on

- We’ll learn:
    - How to use a pre-trained network as a deep feature extractor
    - How to train classifiers for more complex visual recognition tasks by transfer learning
    - How to perform “fine tuning” for domain adaption

Recurrent Nets

Lab 3: Recurrent Nets

Lets get hands-on

- We’ll learn:
  - How to train a generative language model
  - How to classify sequences