Knowledge Processing with Neural Networks

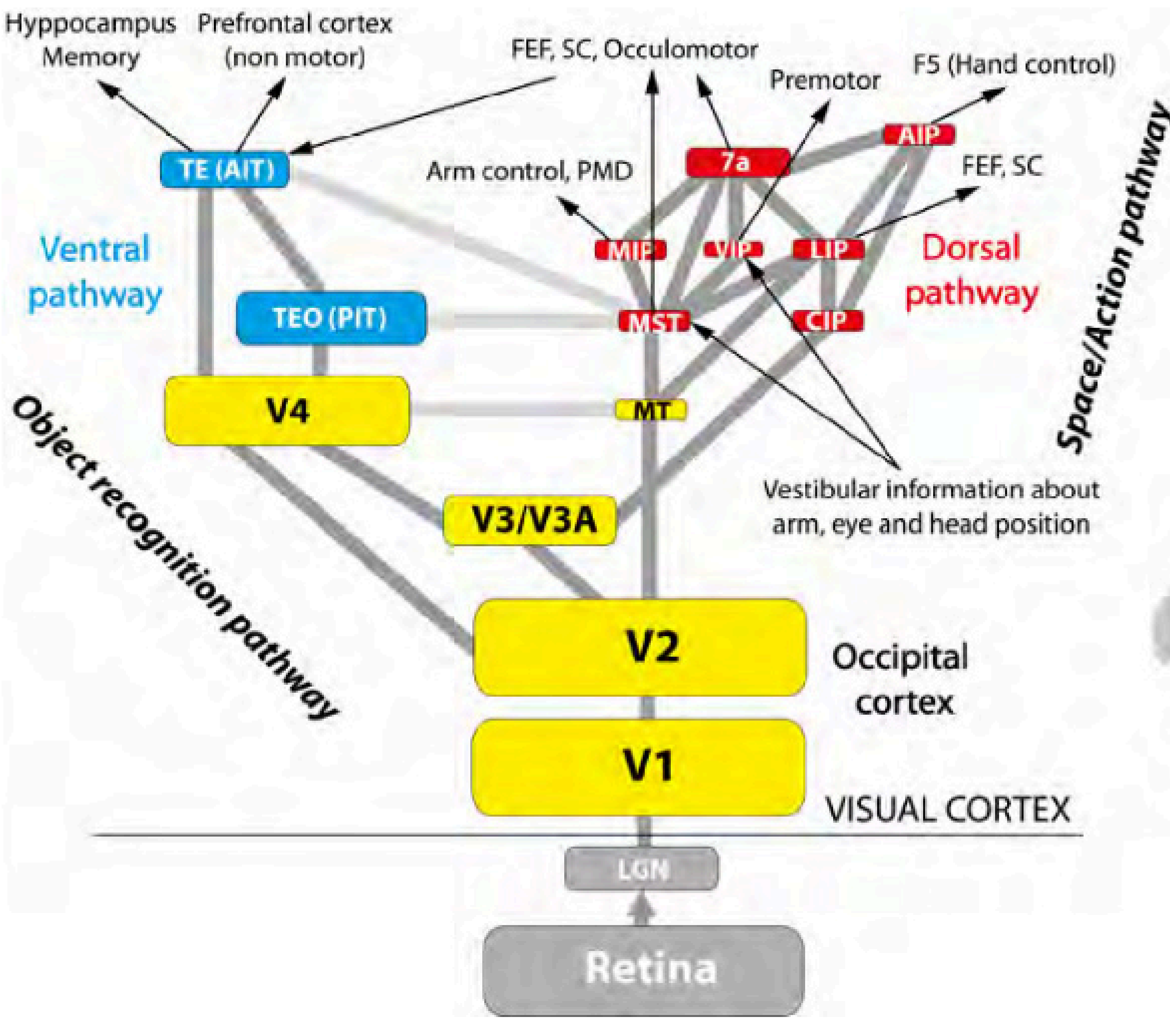
Lecture 9
Visual Frame of Reference Transformations

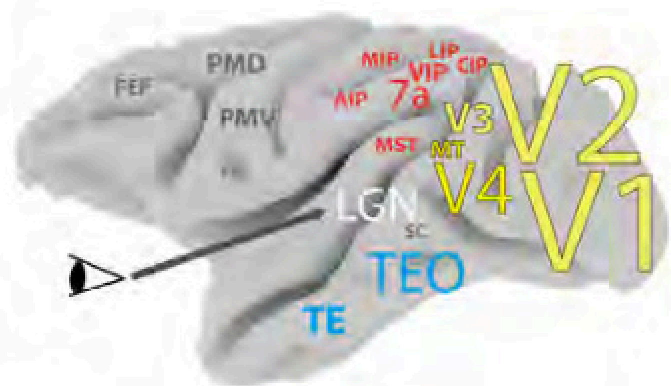
http://www.informatik.uni-hamburg.de/WTM/

Generative Models – Summary

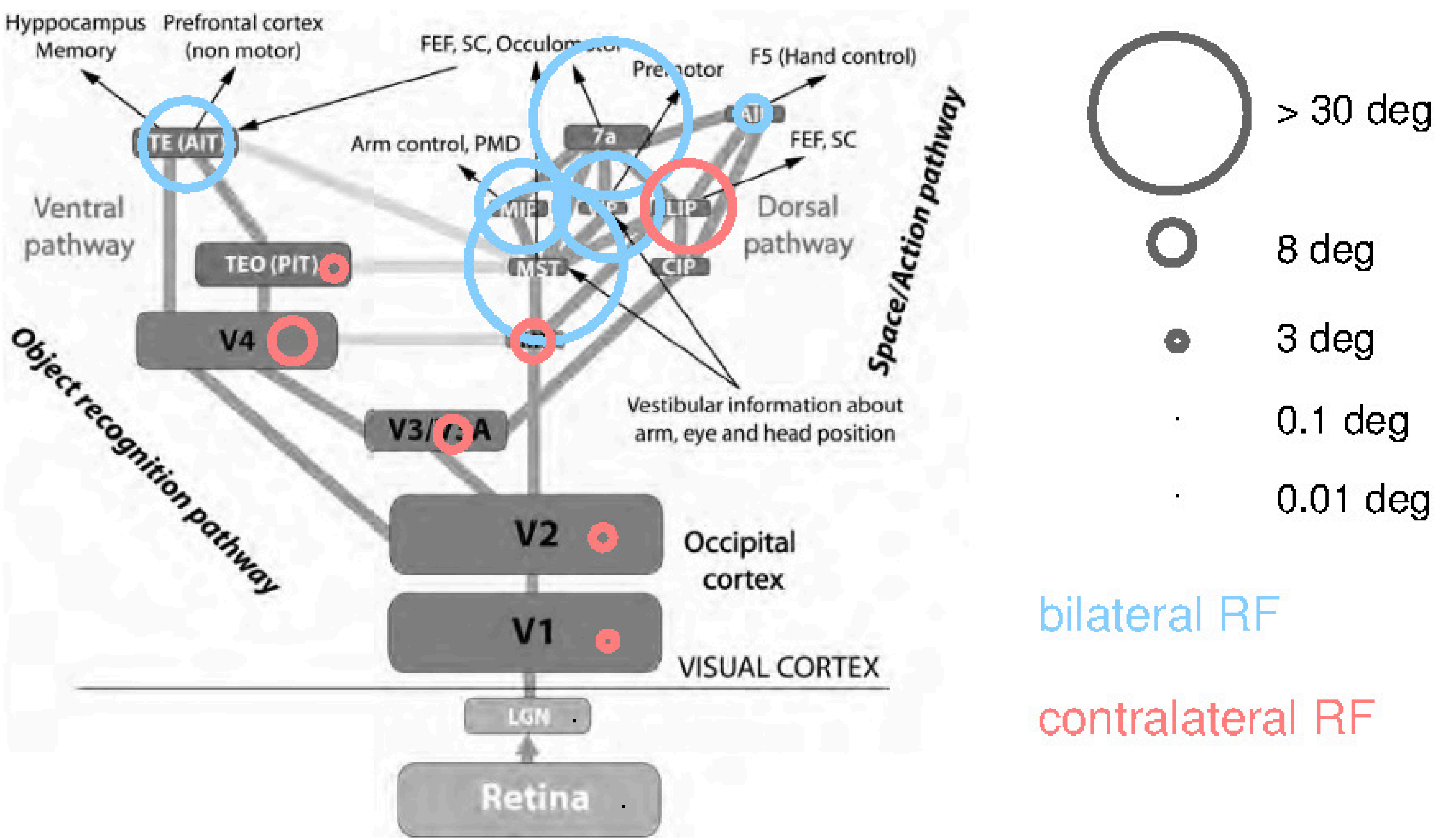
- hierarchical visual system with growing abstraction
- weight matrices transform the representations
- generative model for learning
- prior constraints on hidden encoding:
  - few hidden neurons – PCA
  - weight constraints
  - sparse hidden activations – ICA
  - non-negativity – NMF
  - denoising
  - winner code – vector quantization, K-means, Kohonen
- innate face preferences
- peripheral / foveal preferences
- slow / fast responses

Size of Visual Areas

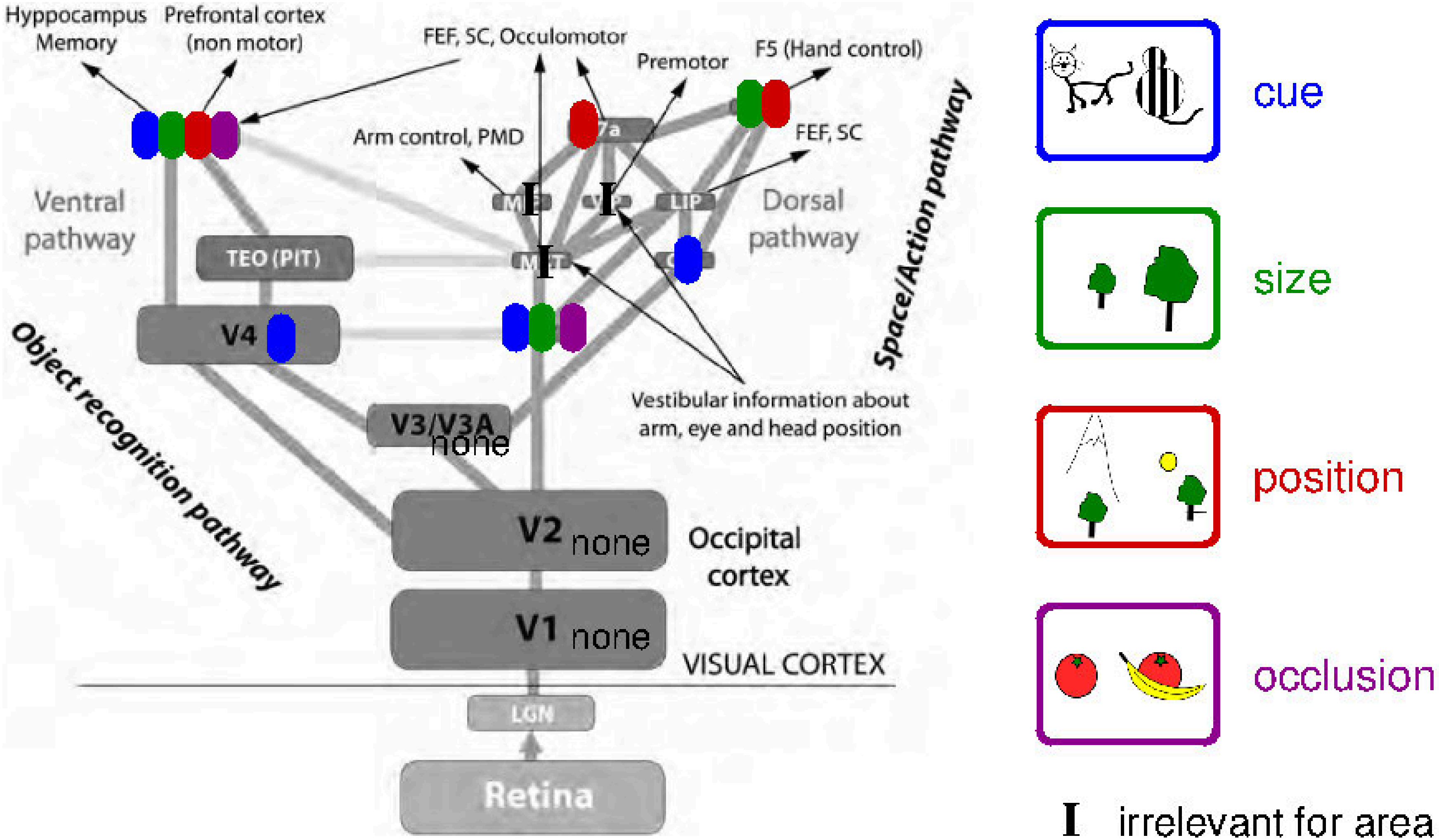




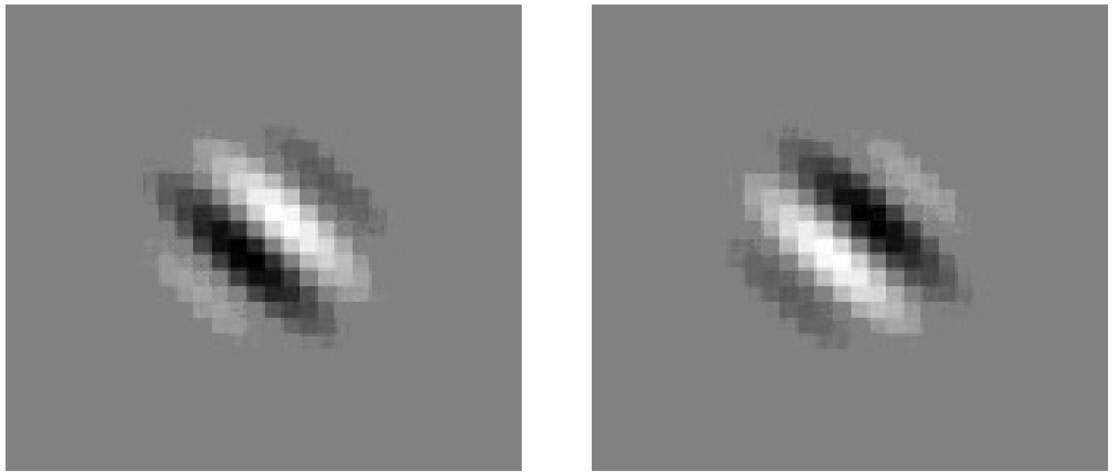
Receptive Field Sizes at 5° eccentricity



Invariances in Visual Areas

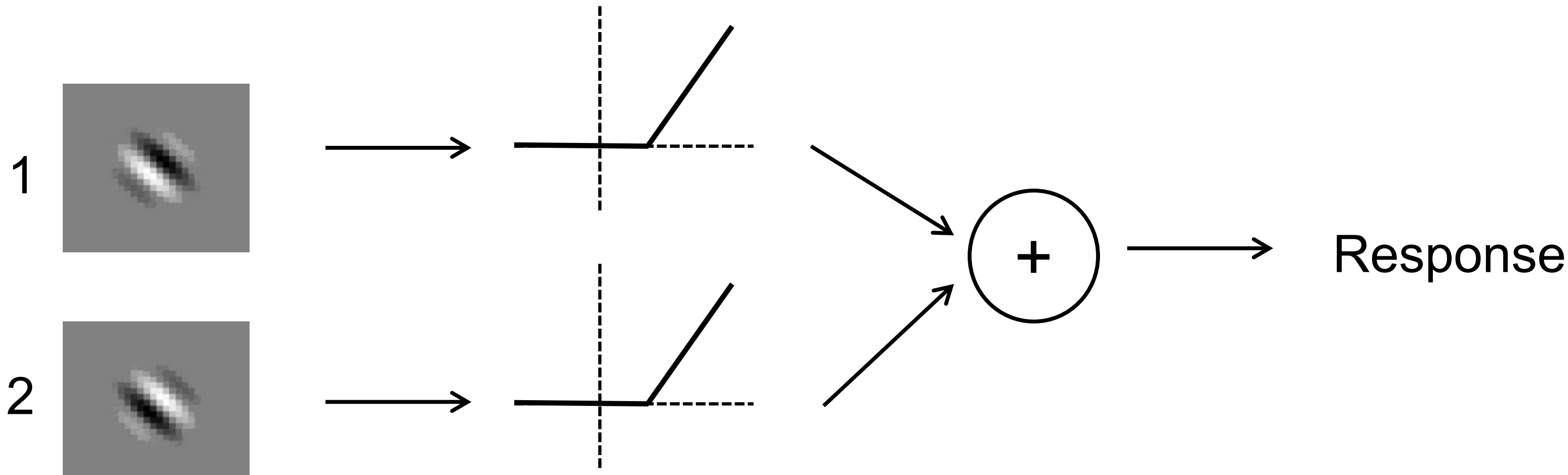


Invariances in V1



- Complex Cells in V1 are orientation selective (like Simple Cells), but fire independent of spatial phase (e.g. if a black/white edge turns to white/black)
- Cannot be explained by linear superposition of receptive fields, or of linear responses (both would result to zero)

Non-Generative Models: Invariances

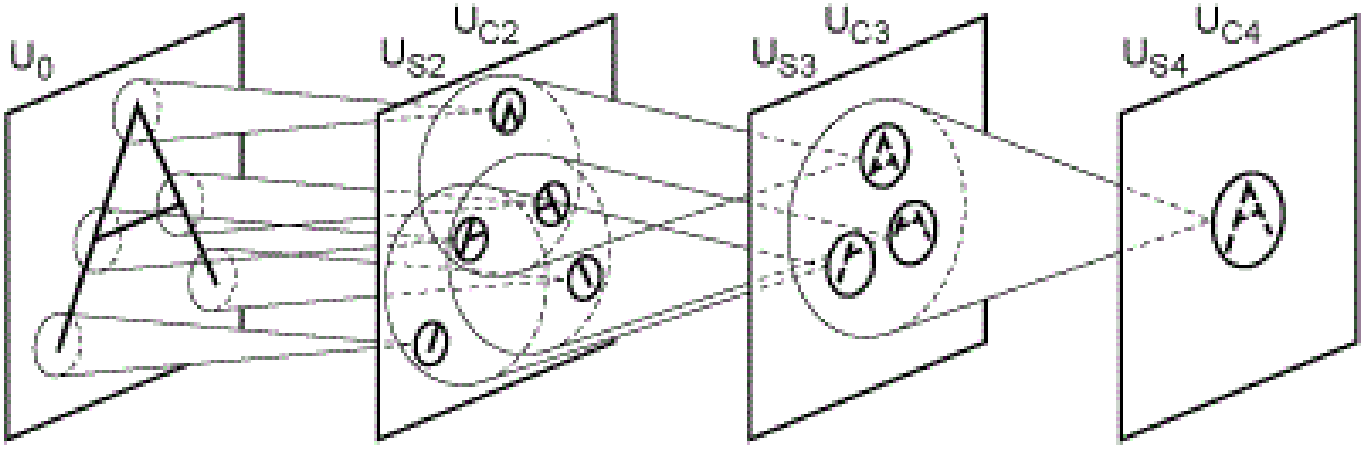


- Non-linear simple cell:
- if simple cell 1 fires, simple cell 2 will be OFF

  ← construct a 2nd-layer complex cell that fires in both cases: 2nd-layer neuron sums output of simple cells 1 & 2 (HMAX operator may be useful)

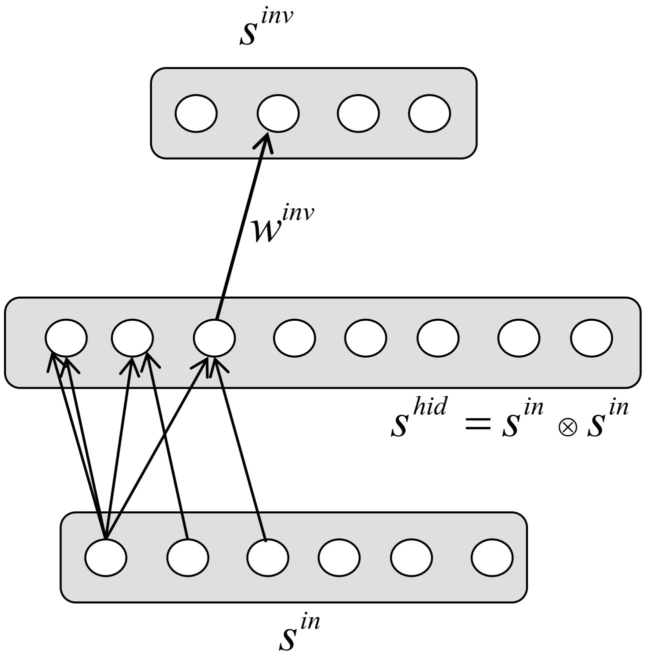
  → 2nd-layer neuron is orientation-selective, but shift-invariant

Fukushima's Neocognitron



alternating layers of feature-detecting S-cells (simple-like) and C-cells (complex-like) over the hierarchy of the visual system

Slow Feature Analysis



Hidden layer pre-determined

$$s^{hid} = ((s_1^{in})^2, s_1^{in} s_2^{in}, s_1^{in} s_3^{in}, \ldots, (s_n^{in})^2)$$

Linear "invariant" output

$$s_j^{inv} = \sum_k w_{jk}^{inv} s_k^{hid}$$

Output shall change slowly

$$\left\langle \frac{\partial s_j^{inv}}{\partial t} \right\rangle_t \overset{!}{=} \min$$

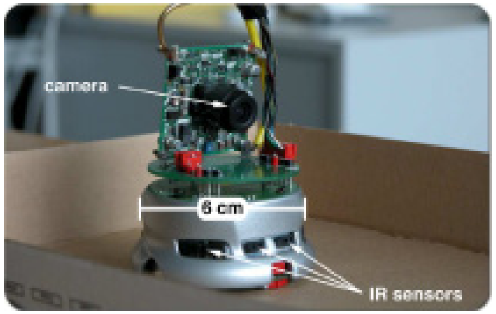
Constraints: $\langle s_j^{inv} \rangle_t = 0$ (zero mean)

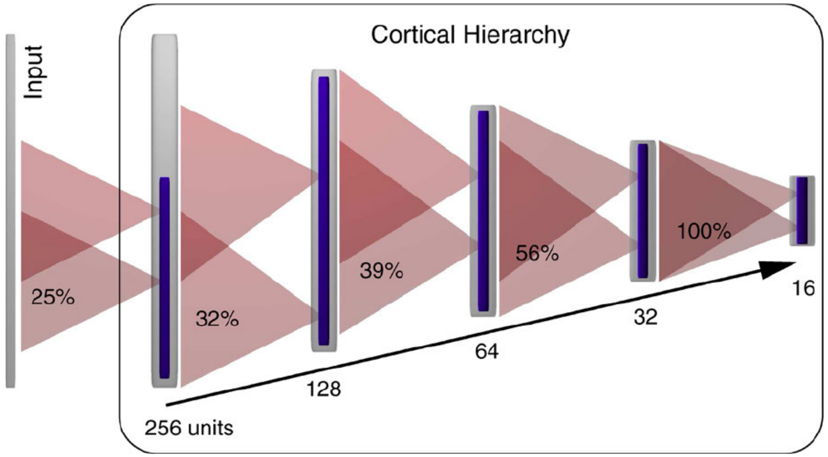
$\langle (s_j^{inv})^2 \rangle_t = 1$ (unit variance)

$\forall i < j : \langle s_i^{inv} s_j^{inv} \rangle_t = 0$ (decorrelation and order)

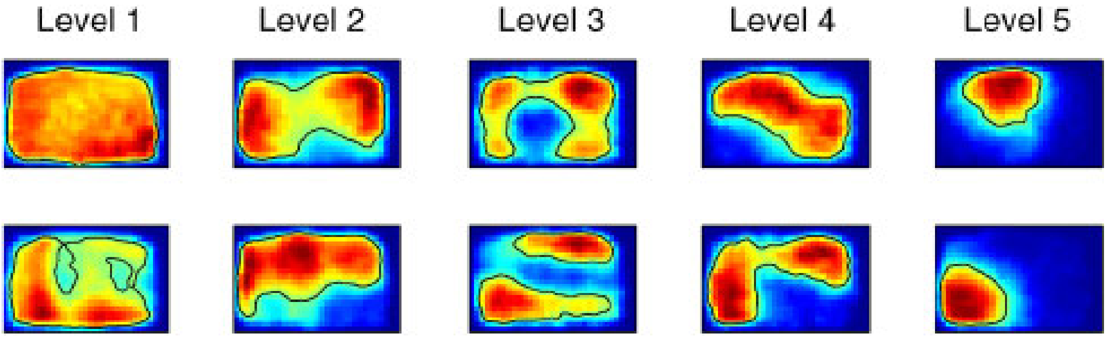
→ Obtain w^{inv} using linear algebra

Place Cells





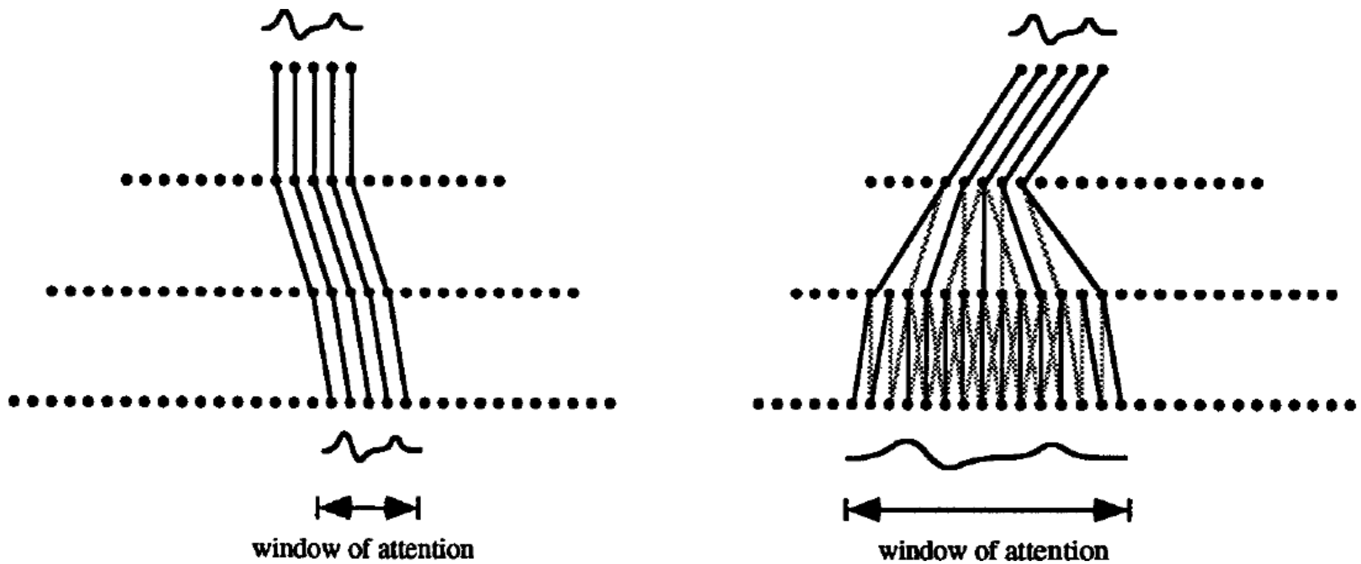
Robot turns a lot ⇒ direction-independent place fields



Invariant Representations – Summary

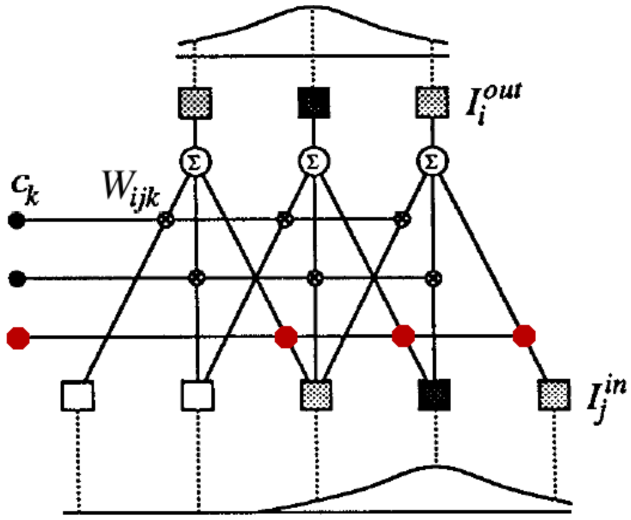
- learning invariances by temporal averaging ("trace rule", "slowness")
- spatial transformations: shift, scale, rotation, shape
- other transformations: light, shading, color
- information loss – no strict generative model
- need to code features and their transformations
  - "what & where" models (for the shift-transformation)
  - 3-way model needed:
    - Input
    - what's there
    - where is it / where do I want to look

Dynamic Routing



- rich physical connectivity must exist
- sparse “logical” connectivity dynamically activated
- partial routing between layers – full routing over entire hierarchy

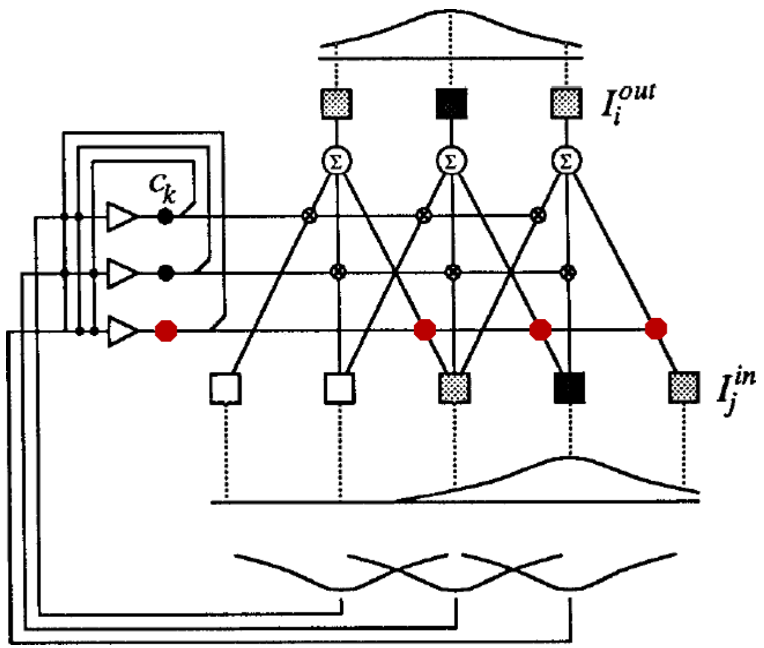
Dynamic Routing



$$I_i^{out} = \sum_k \sum_j W_{ijk} c_k I_j^{in}$$

- Higher-order connections W
- Only one c_k should be active

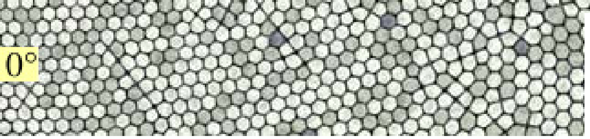
Dynamic Routing



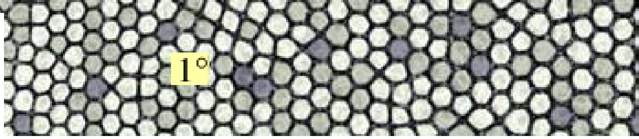
$$I_i^{out} = \sum_k \sum_j W_{ijk} c_k I_j^{in}$$

- Higher-order connections W
- Only one c_k should be active
- c_k inhibit each other

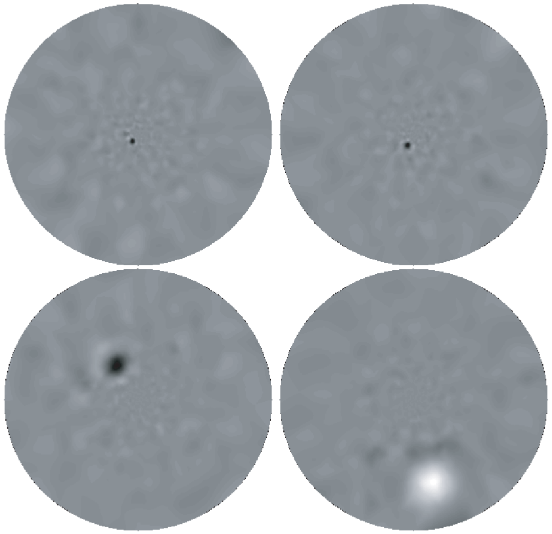
- c_k may be activated by I^{in} (bottom-up attention)
- c_k may be activated voluntarily (top-down attention)

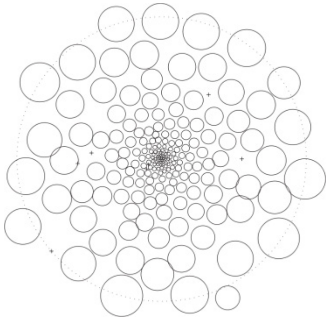


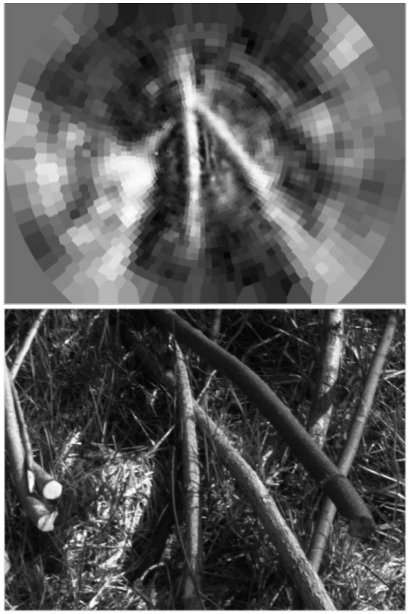
Foveated Vision



- retinal receptive field sizes increase with eccentricity





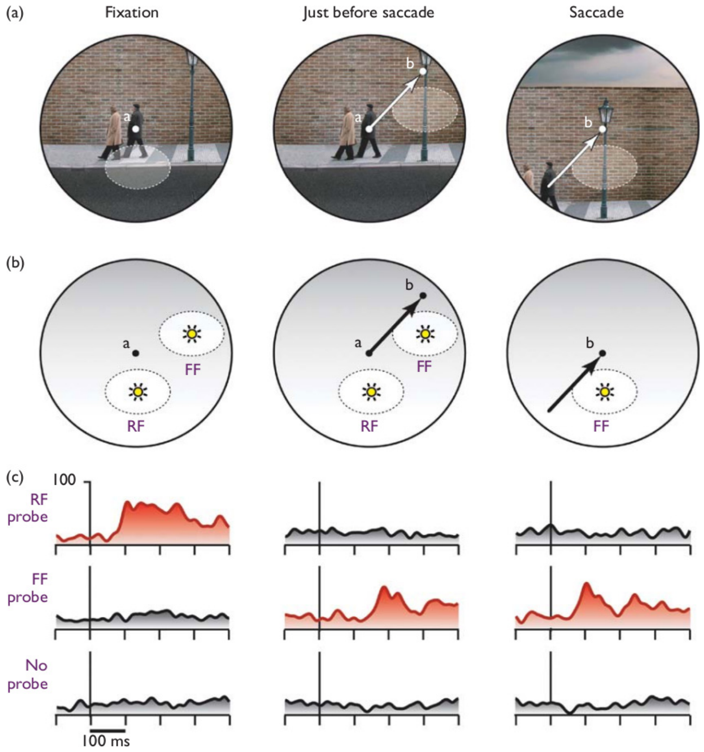


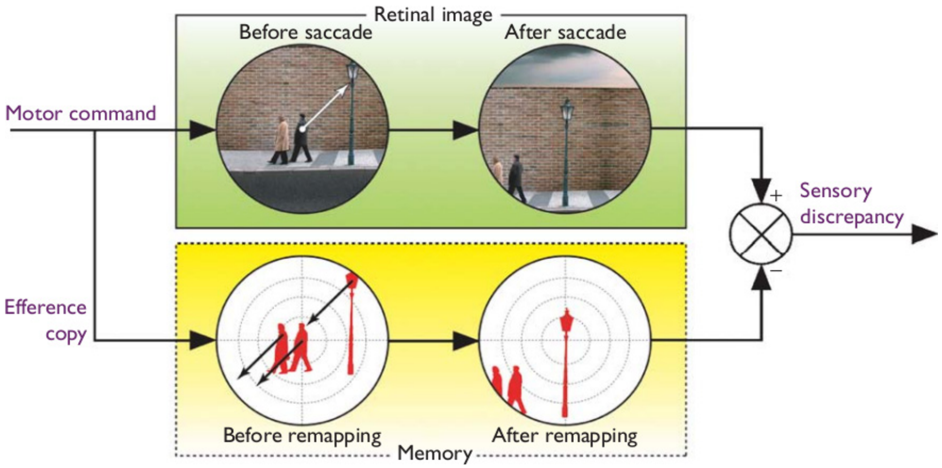
→ little information in the periphery to route (covert attention)

→ we do saccades to retrieve visual detail (overt attention)

Predictive Remapping

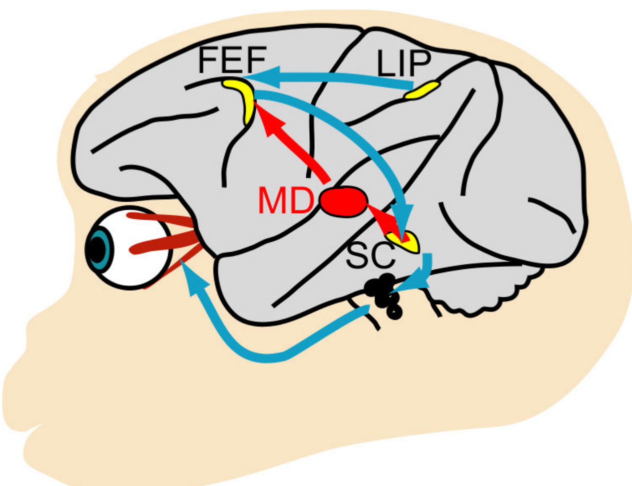
- just before a saccade, a neuron in FEF (frontal eye field) will shift its receptive field to its predicted position after the saccade

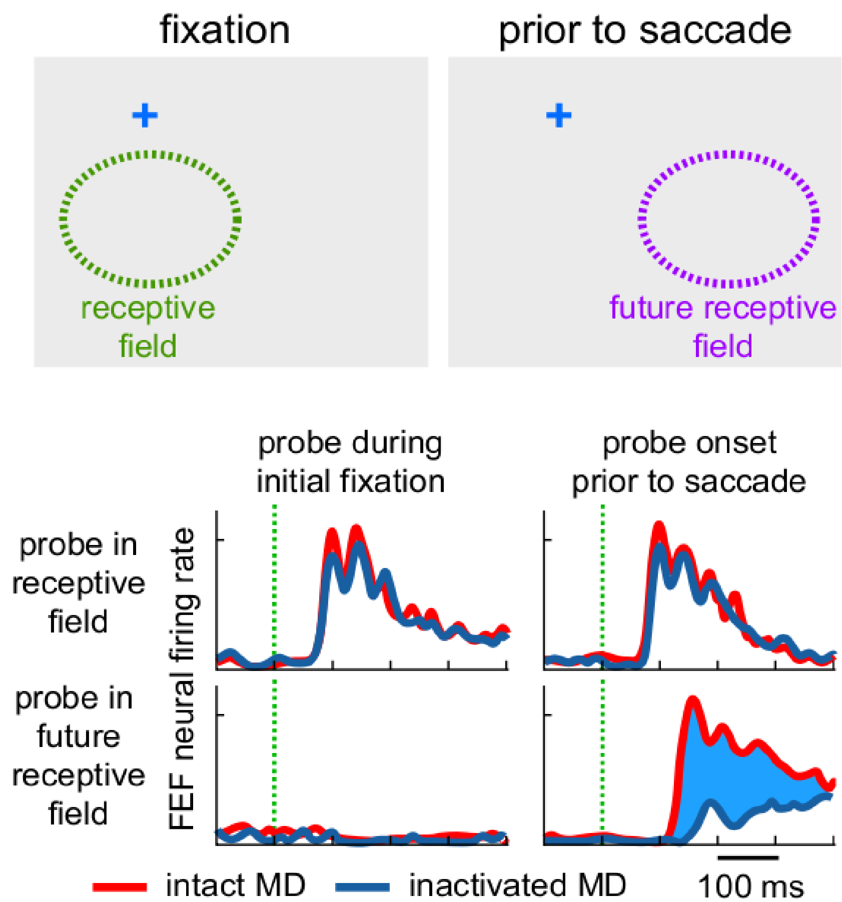




Predictive Remapping

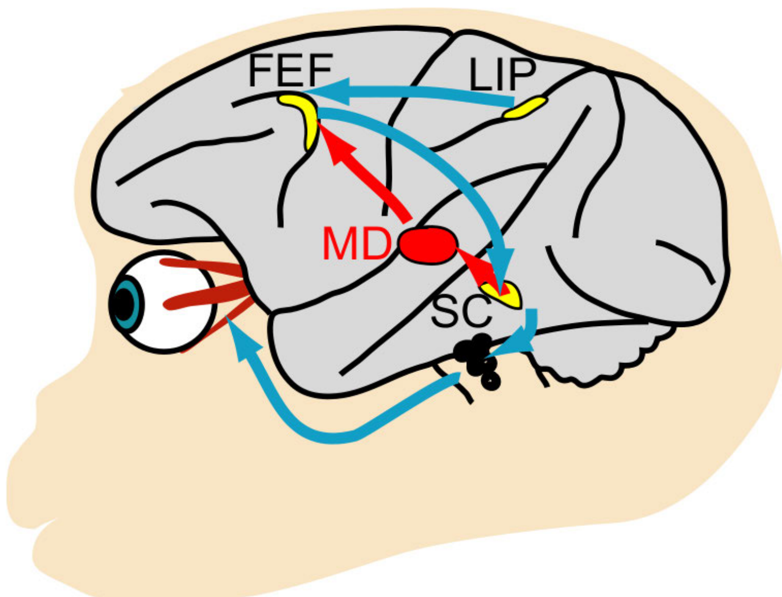
- activation of MD neurons (medial dorsal nucleus of the thalamus) is required to remap the FEF neuron's receptive field

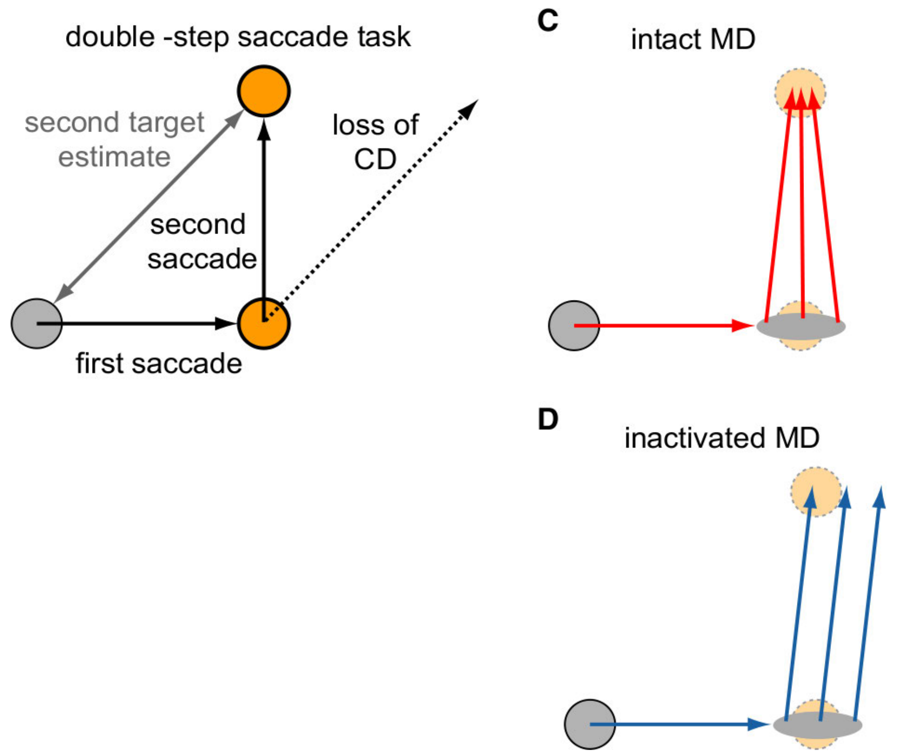




Predictive Remapping

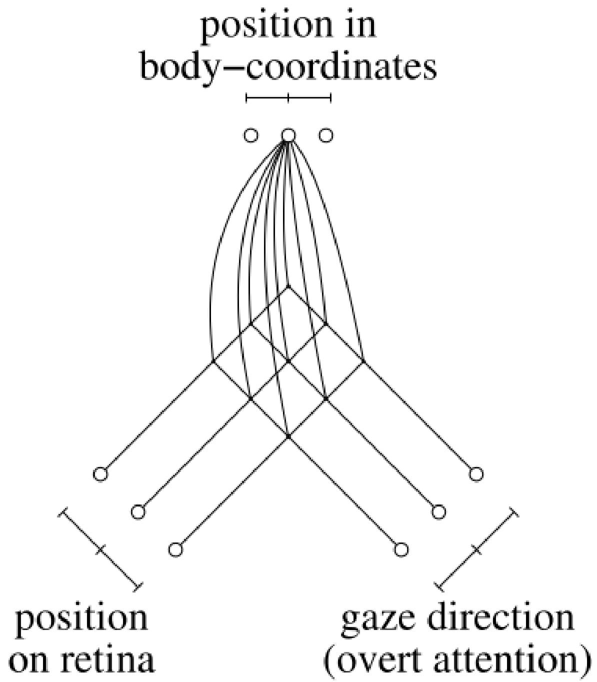
- double saccade task: perform memory-guided saccades to two briefly flashed targets

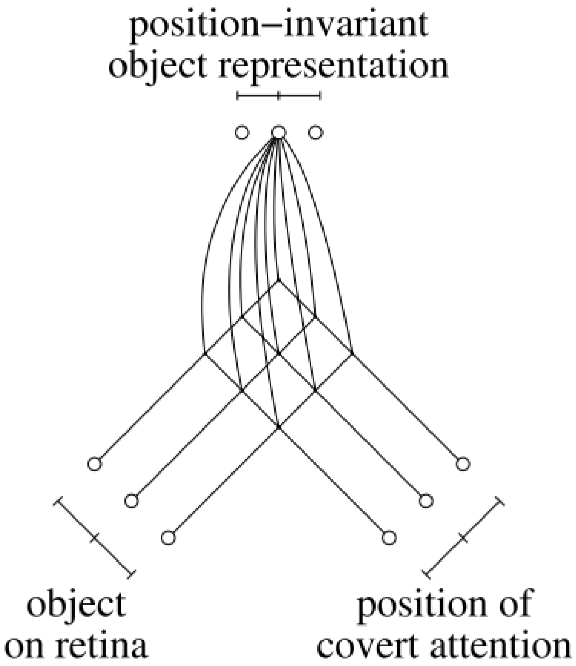




- MD transmits motor signals from the first saccade to update the sensory representation to guide the second saccade

Remapping & Transformations

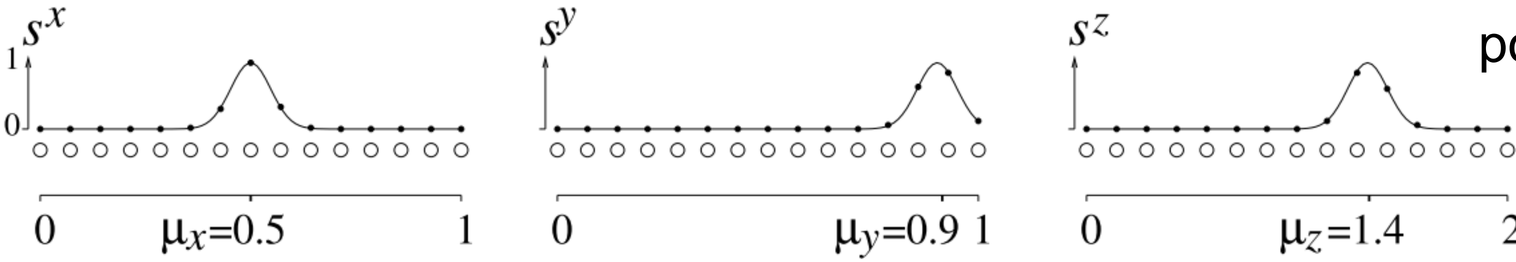




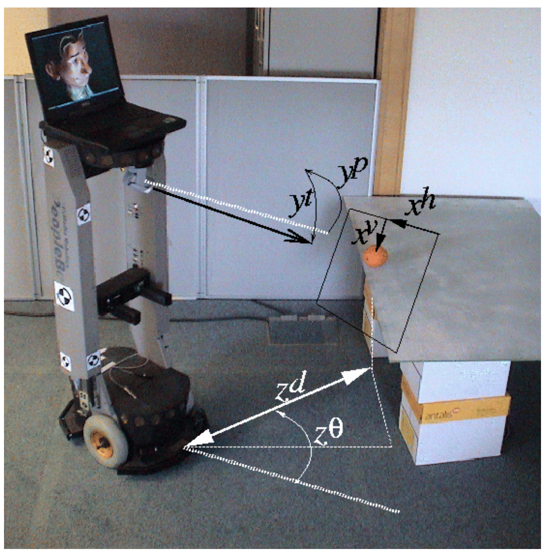
shift all object features to a defined frame for recognition (i.e. “subtract” eccentricity)

assess an object’s position in a motor-relevant coordinate frame (i.e. add retinal+gaze position)

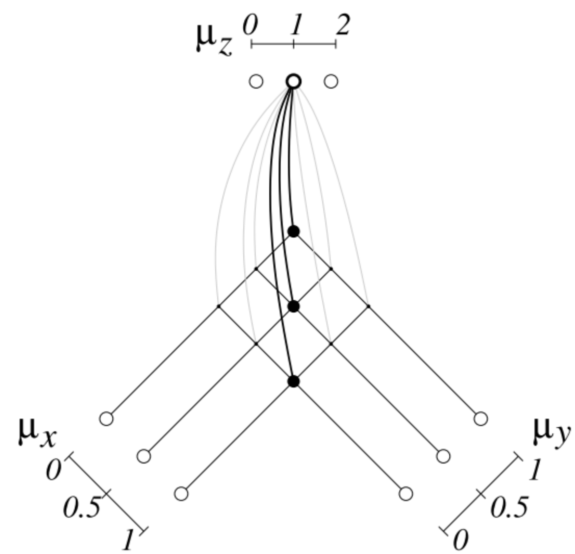
Frame of Reference Transformations



population code



$$s_i^z = \sum_{jk} w_{ijk} s_j^x s_k^y$$



training with slowness: similar s^z response for consecutive views

Frame of Reference Transformations

$$E = \frac{1}{2}\sum_{\mu}^{data}\sum_{j}^{hid} s_j^{hid}(x,y)\sum_{k}^{hid} h(|\,j-k\,|)(\bar{x}_i\bar{y}_l - w_{kil})^2$$

on-line algorithm (Kohonen-like)

initialize

w_{jil}= small random variables

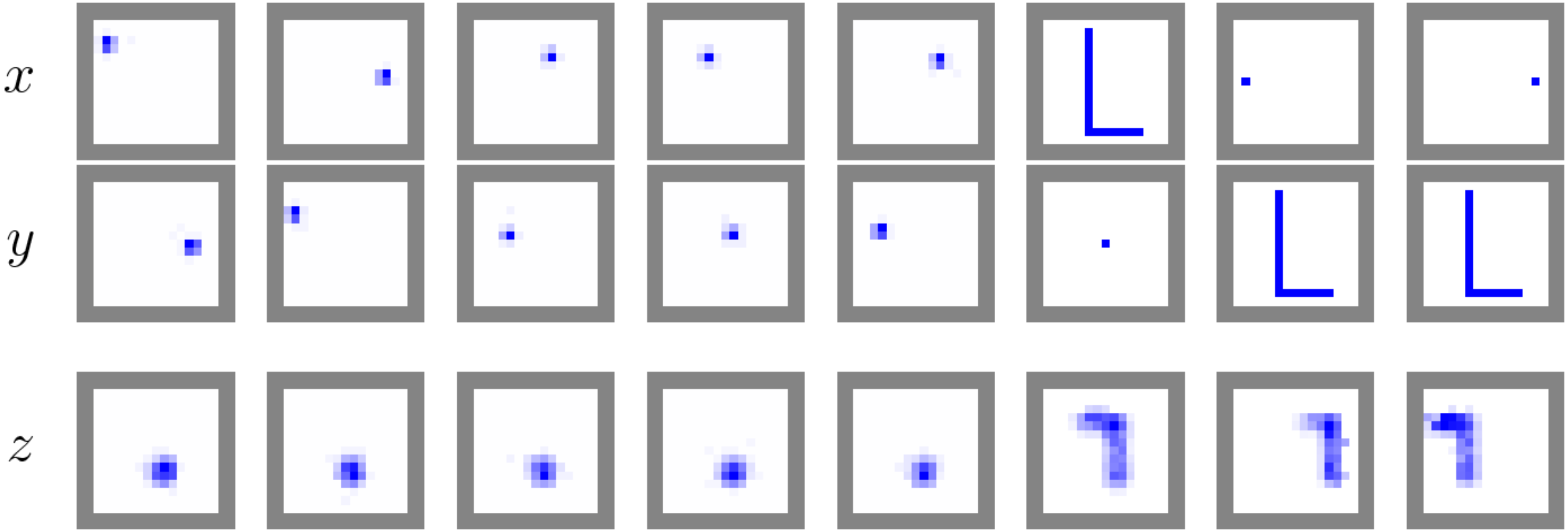
repeat

- choose random sum location z, random matching input x,y
- assign winning unit $j^* = \text{argmax}_k\, w_{kil}\, x_i\, y_l$
- give another random input pair $\bar{x}, \bar{y}$ that matches z
- learn winner and surround

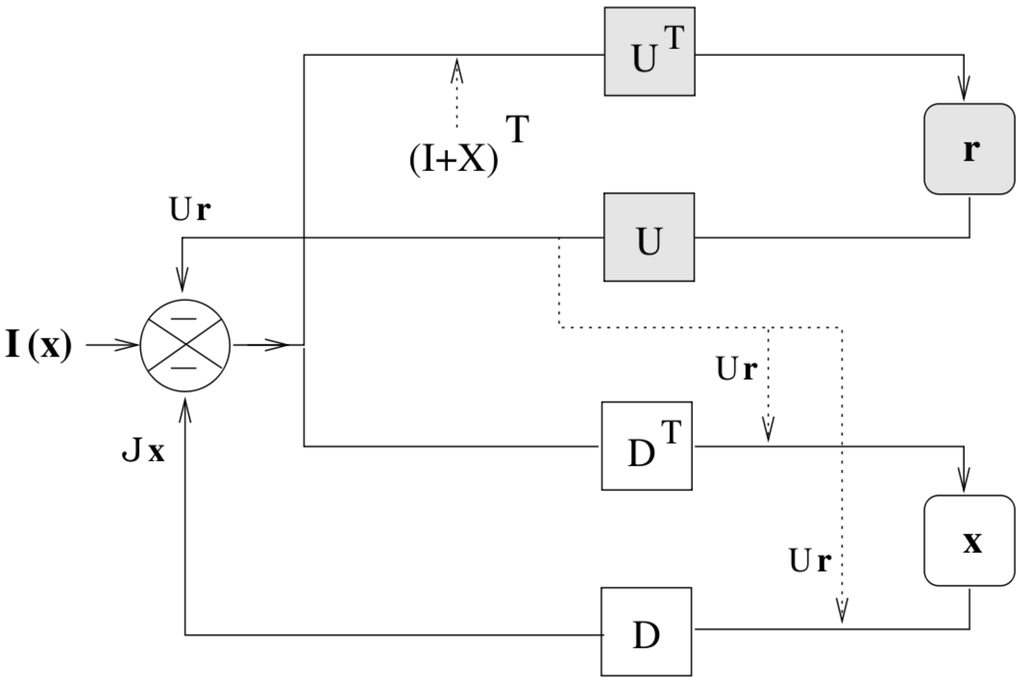
$$\Delta w_{kil} = \varepsilon h(|\,j^* - k\,|)(\bar{x}_i\bar{y}_l - w_{kil})$$

- Gaussian neighbour function h = h(σ) on map layer
- Interaction with σ is reduced during learning

Frame of Reference Transformations



What-Where Estimation (Bi-Linear)



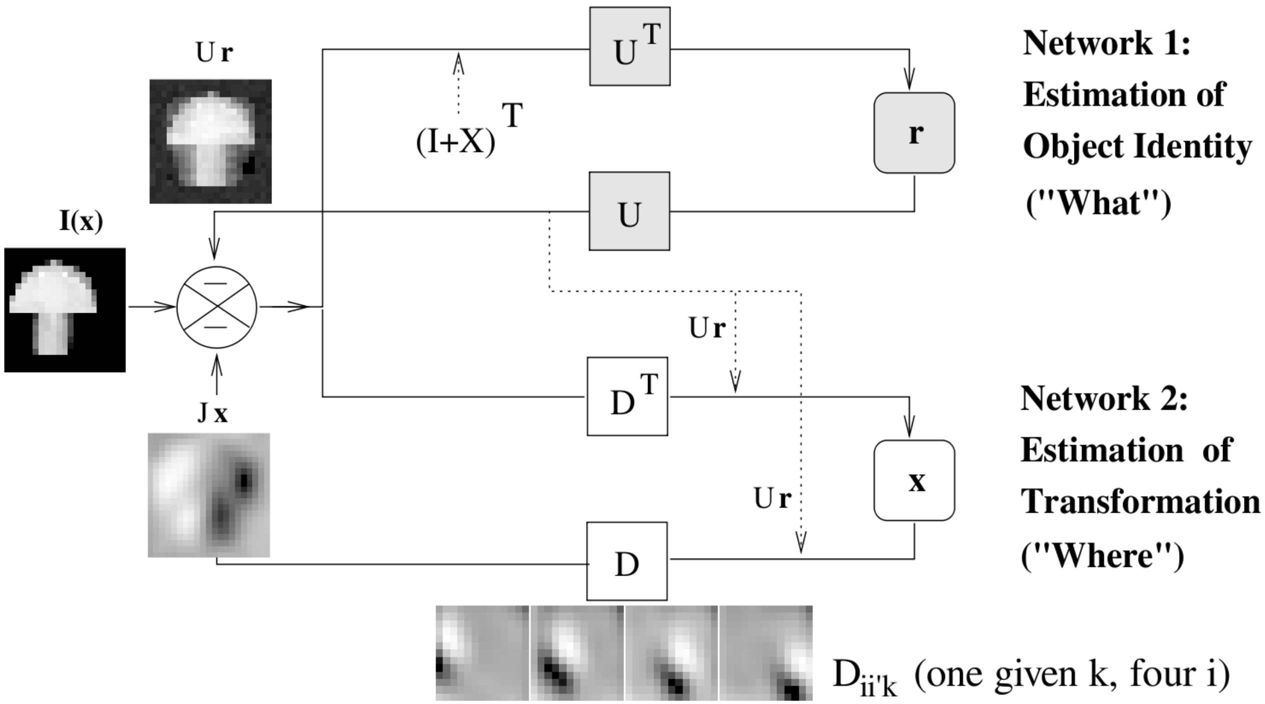
Network 1:
Estimation of
Object Identity
("What")

Network 2:
Estimation of
Transformation
("Where")

$$I^{out} = Ur + \underbrace{D\mathrm{I}}_{J=\frac{\partial I}{\partial x}} x$$

$$I_i^{out} = \sum_j U_{ij} r_j + \sum_k \sum_{i'} D_{ii'k} \left(\sum_j U_{i'j} r_j\right) x_k$$

What-Where Estimation (Bi-Linear)

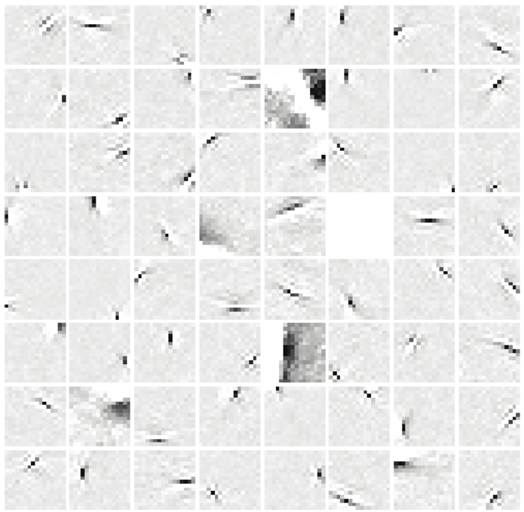


$$I^{out} = Ur + \underbrace{DI}_{J=\frac{\partial I}{\partial x}} x$$

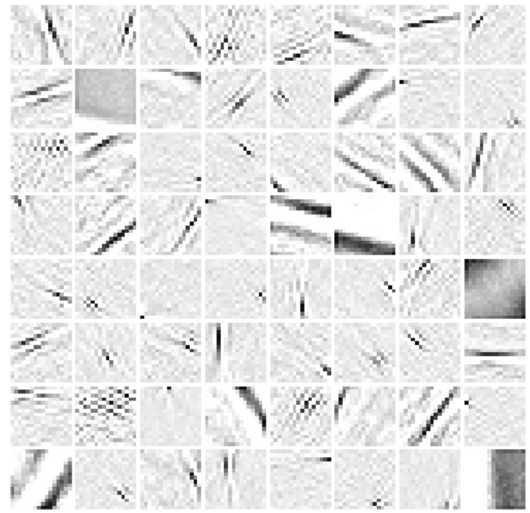
$$I_i^{out} = \sum_j U_{ij} r_j + \sum_k \sum_{i'} D_{ii'k} (\sum_j U_{i'j} r_j) x_k$$

What-Where Estimation (Horizontal Product)

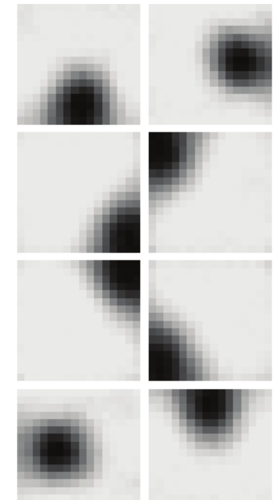
U from pure ICA



U



W

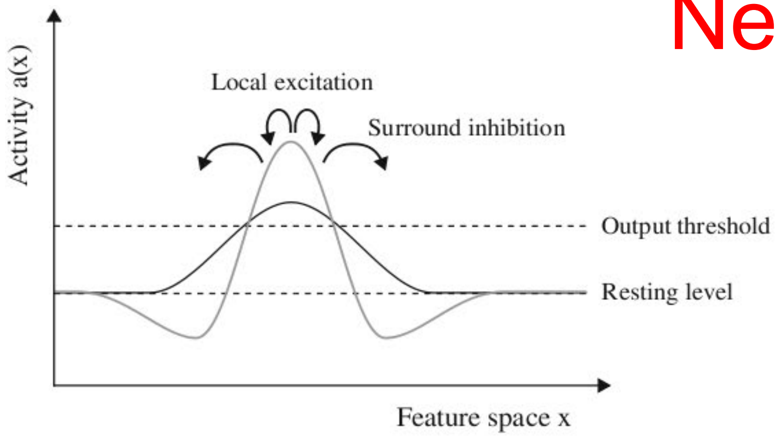


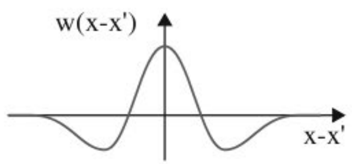
$$I^{out} = Ur \otimes Wx$$

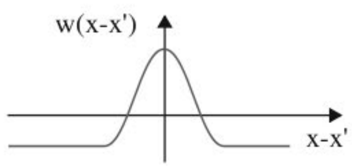
$$I_i^{out} = \left(\sum_j U_{ij} r_j\right)\left(\sum_k W_{ik} x_k\right)$$

- W is undercomplete
- W and x have only positive entries

Neural Fields



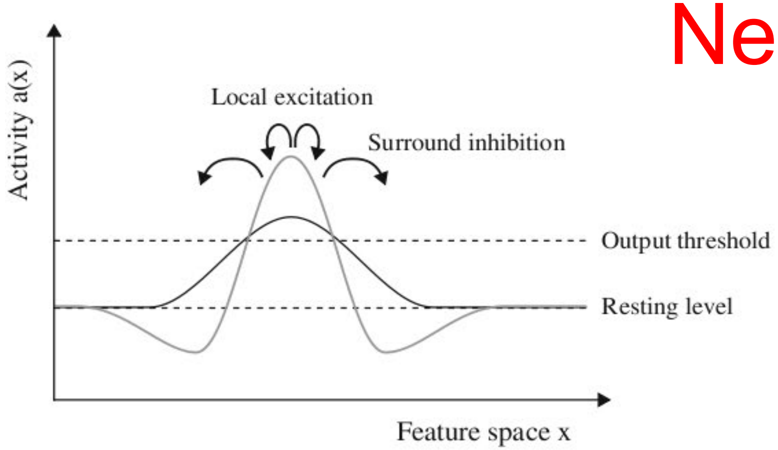


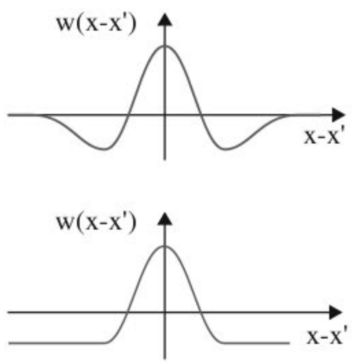


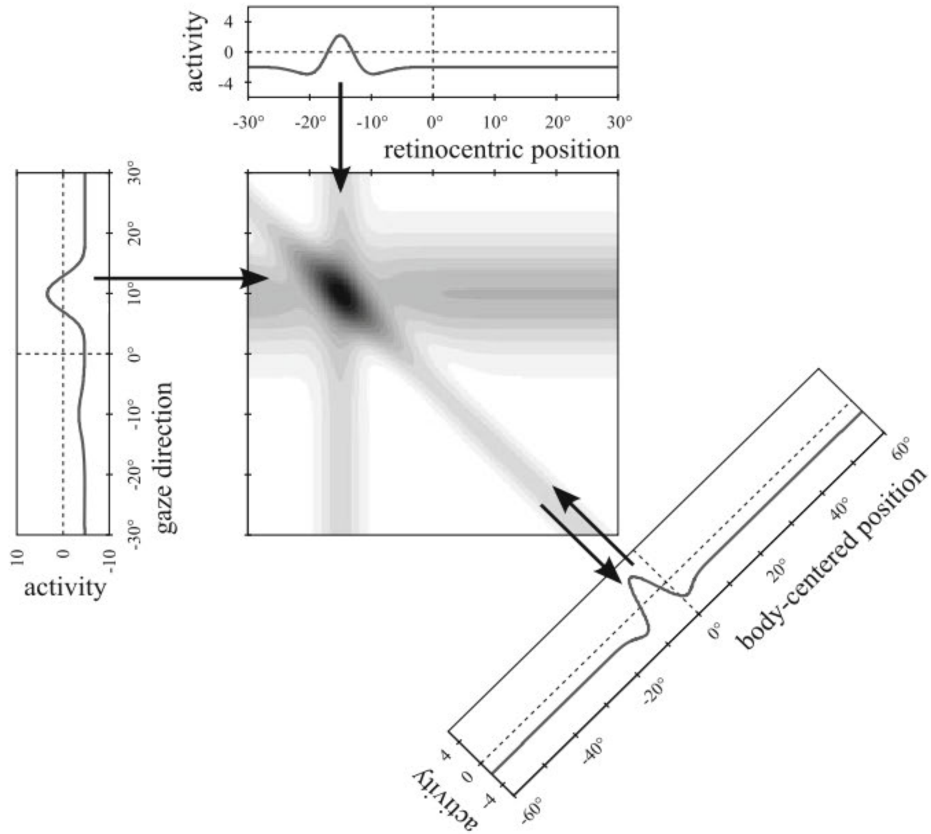
- update of a neural field (time constant τ, additional input I):

$$\tau \frac{\partial s_i(t)}{\partial t} = -s_i(t) + \sum_j w_{ij} f(s_j(t)) + I_i(t)$$

Neural Fields



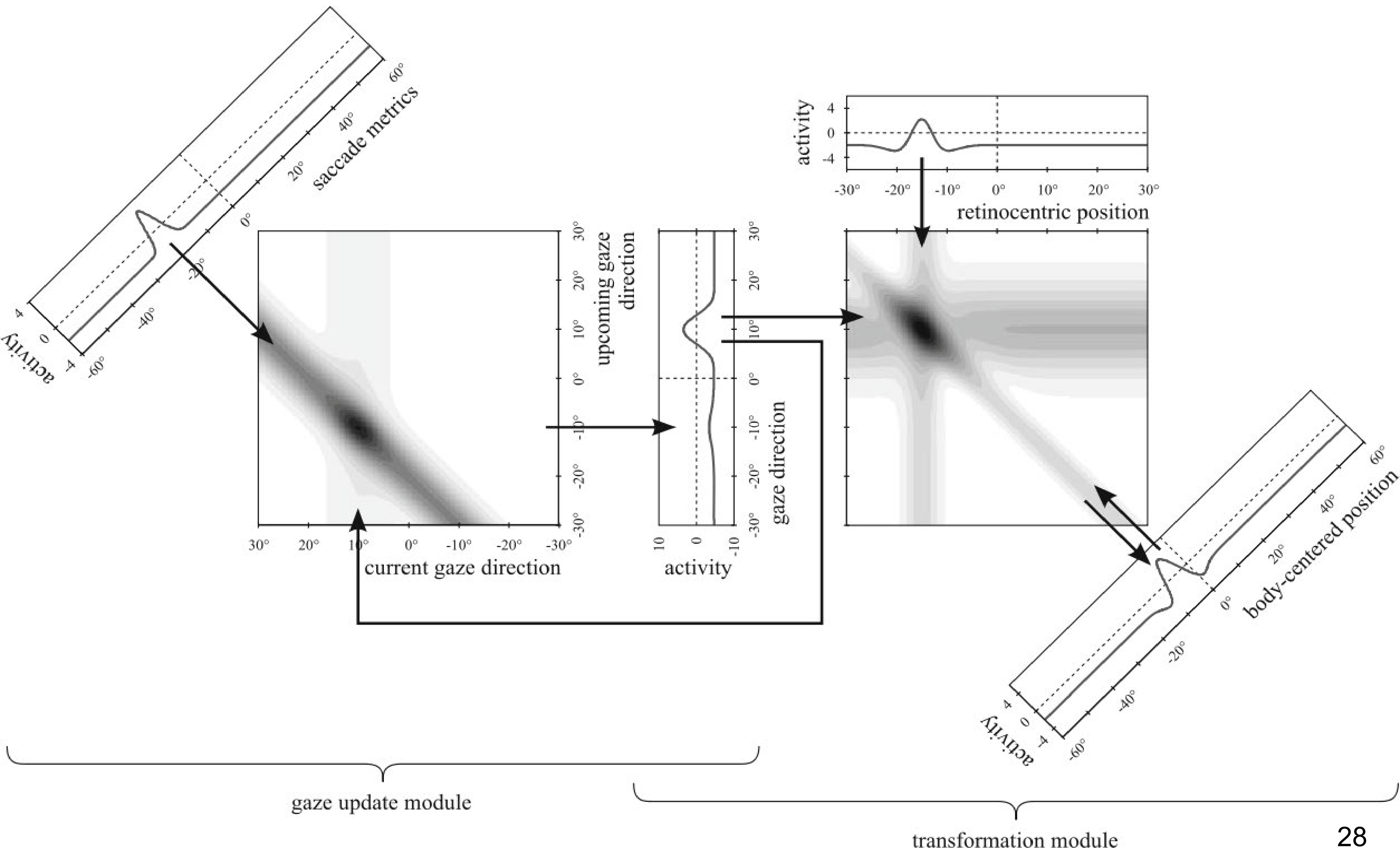


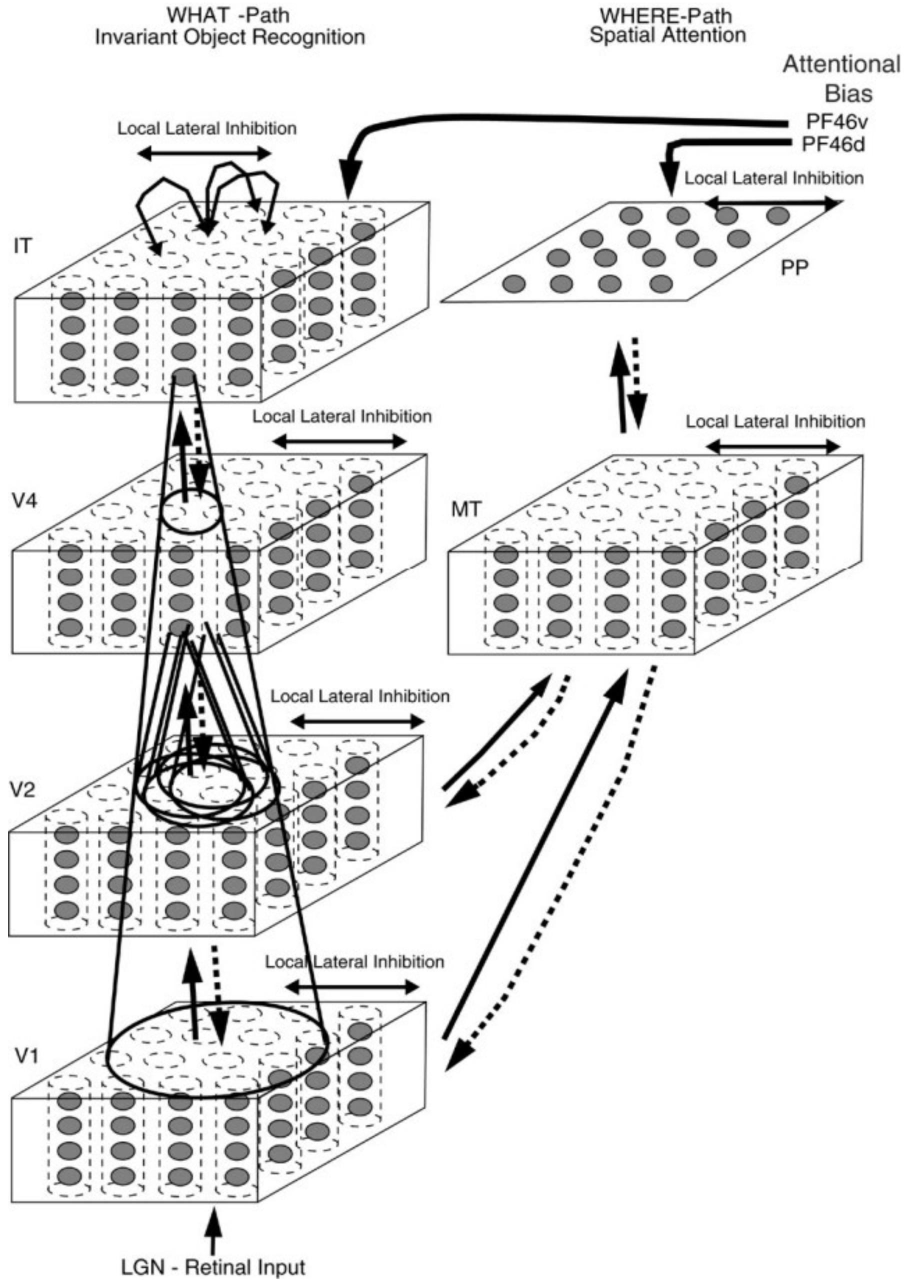


- update of a neural field (time constant τ, additional input I):

$$\tau \frac{\partial s_i(t)}{\partial t} = -s_i(t) + \sum_j w_{ij} f(s_j(t)) + I_i(t)$$

Neural Fields

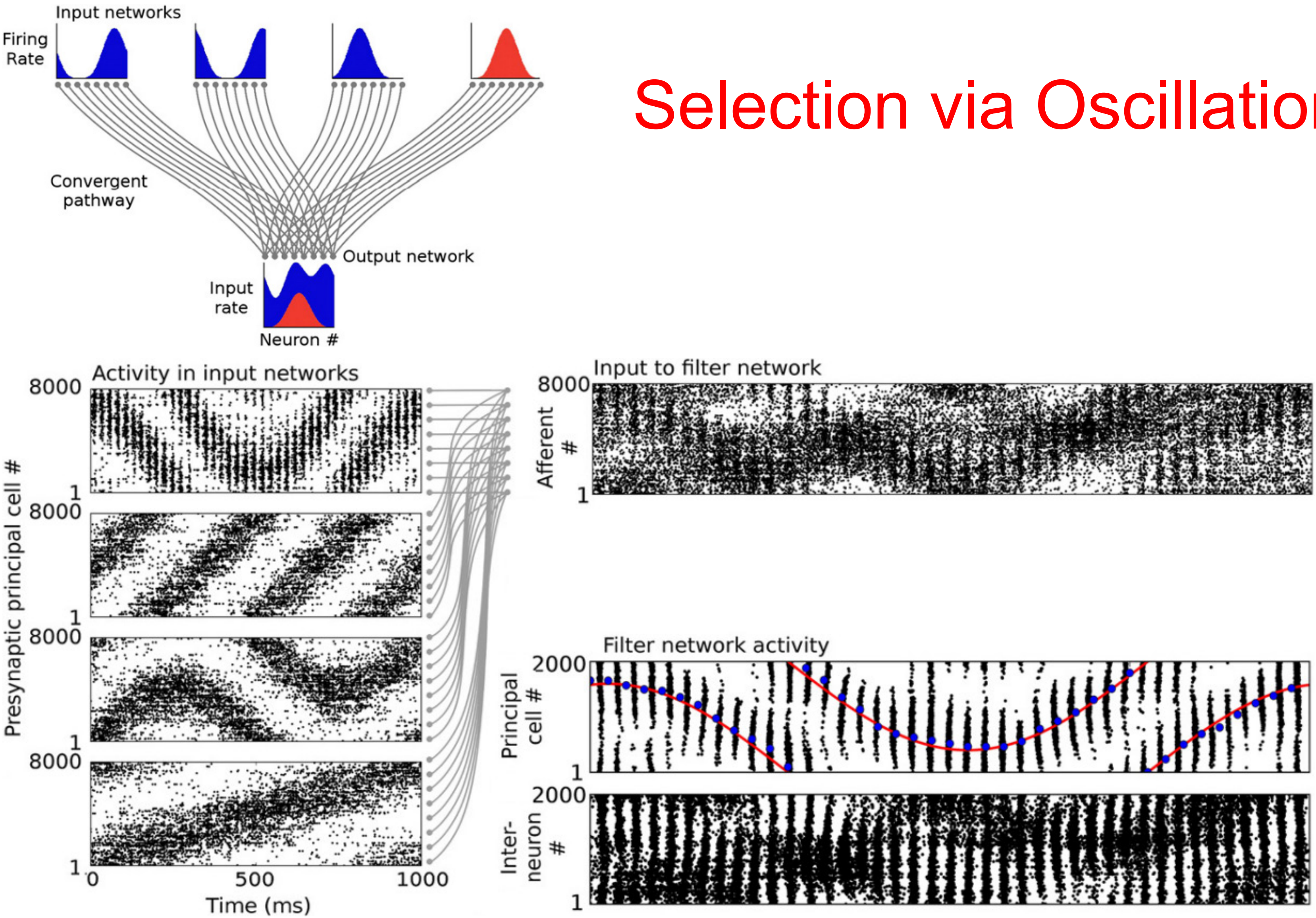




Attentional Selection

- hierarchical model with "gradually" increasing invariance
- competition via neural field within any layer
- topographic lateral connections between "what" and where" stream
- attentional bias to "what" or to "where" possible

Selection via Oscillations



- oscillating neurons select inputs that have matching frequency

“What” & “Where” Models – Summary

- non-linearities required for modulation/control
  - e.g. higher-order neurons, recurrent competition, oscilations
- computationally demanding
- reduced information: non-routed information is discarded
  - e.g. limit field of view ⇔ spatial attention
- some models learn with the generative model paradigm
- simultaneous feature extraction and control is problematic
  - control- and invariance-mechanisms may overlap
- models should better exploit foveation & active vision
- biological models of routing not used in practice (rather, AI-like sliding search window for object recognition)

Note:

- Oral Exams:

  Thu 2nd Oct 2014; 10-13 & 14-16; F-230

- Registration:

  Mon 23rd Jun – Thu 3rd Jul 2014; 9-15; Studienbüro;

  (exception: Studienbüro closed on Friday 27th Jun)