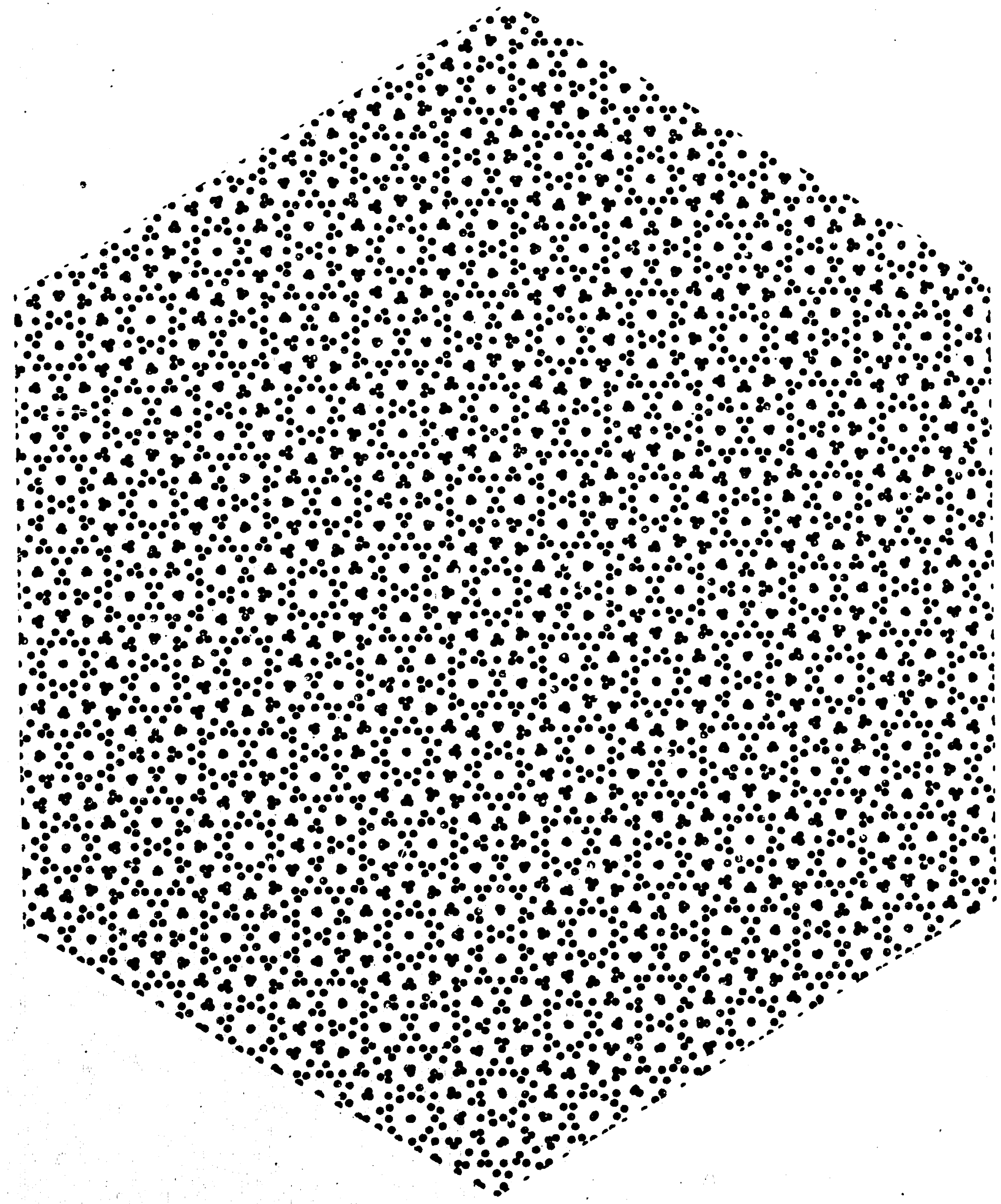
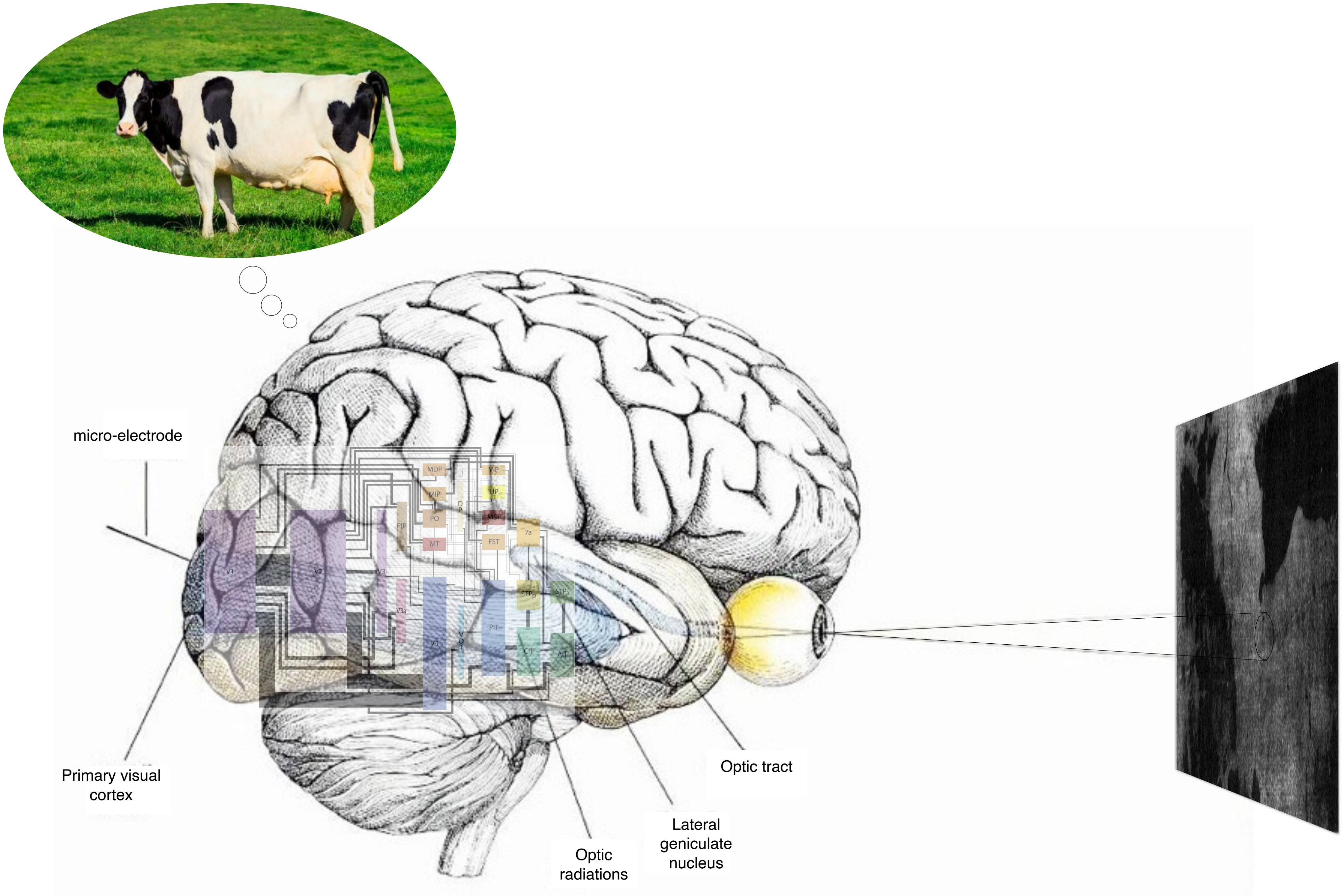
Attractor neural networks

A ‘puzzle picture’
Dallenbach (1951)

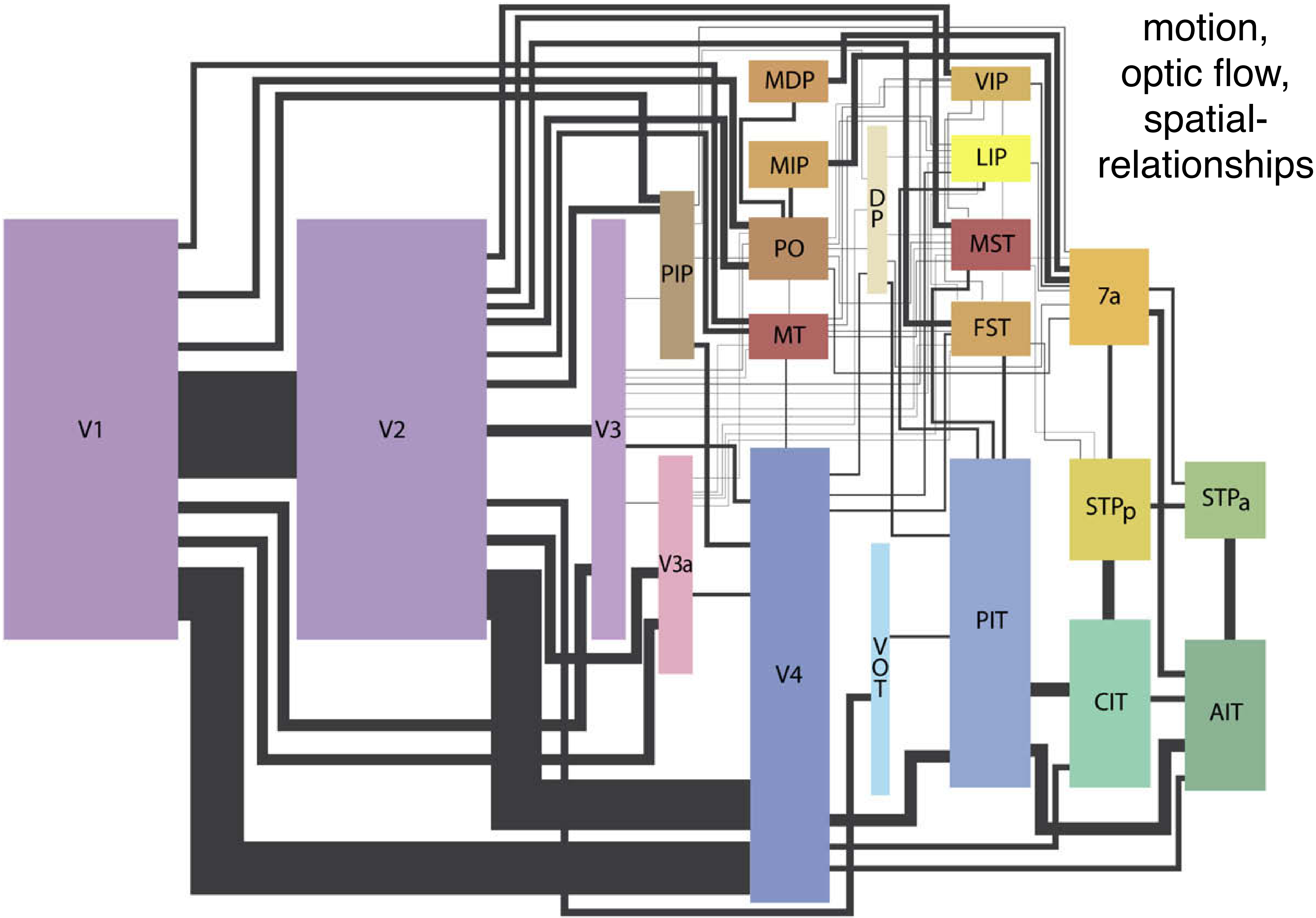


from Marroquin (1976)

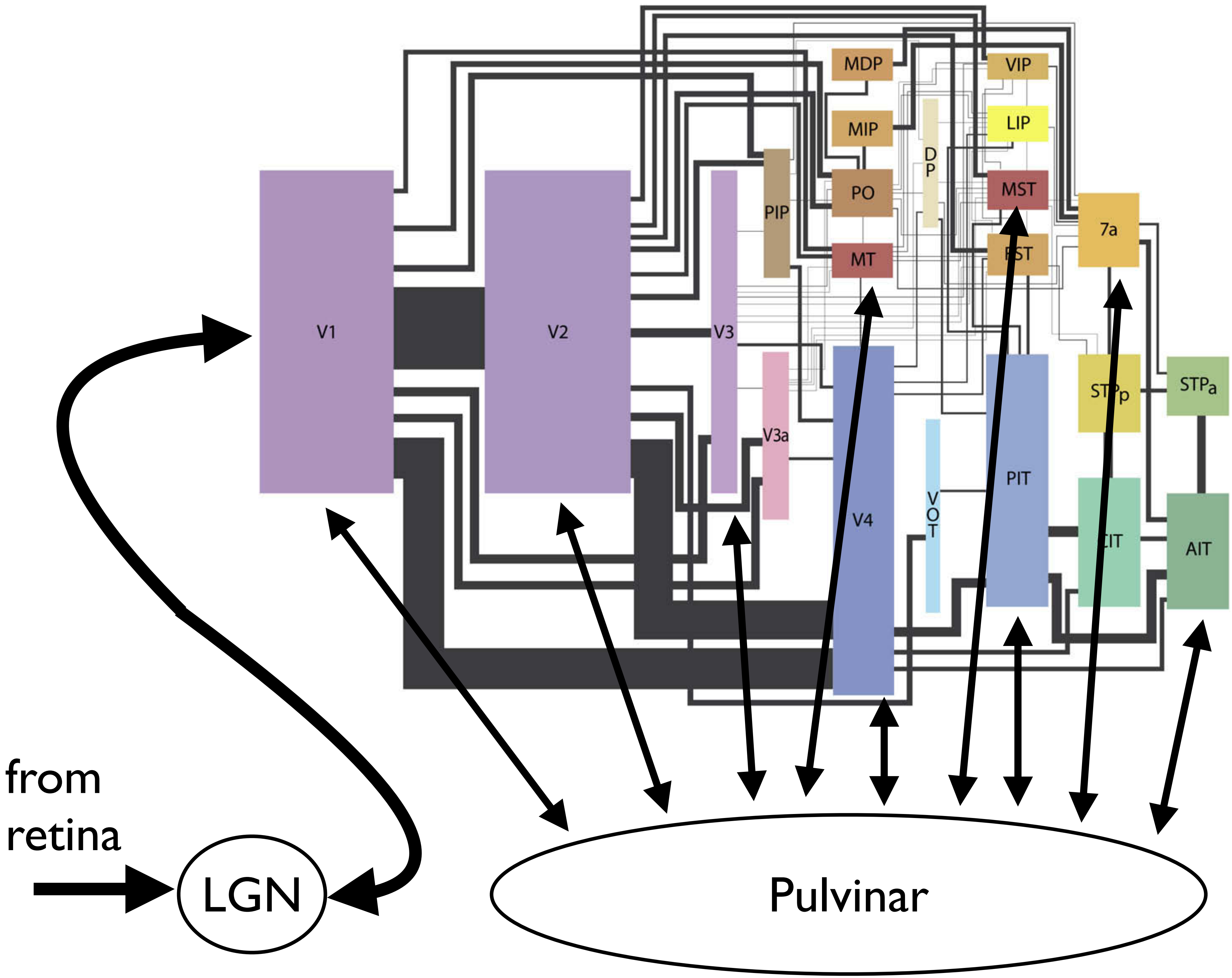
Andy Goldsworthy artwork


micro-electrode
Primary visual cortex
Optic radiations
Lateral geniculate nucleus
Optic tract

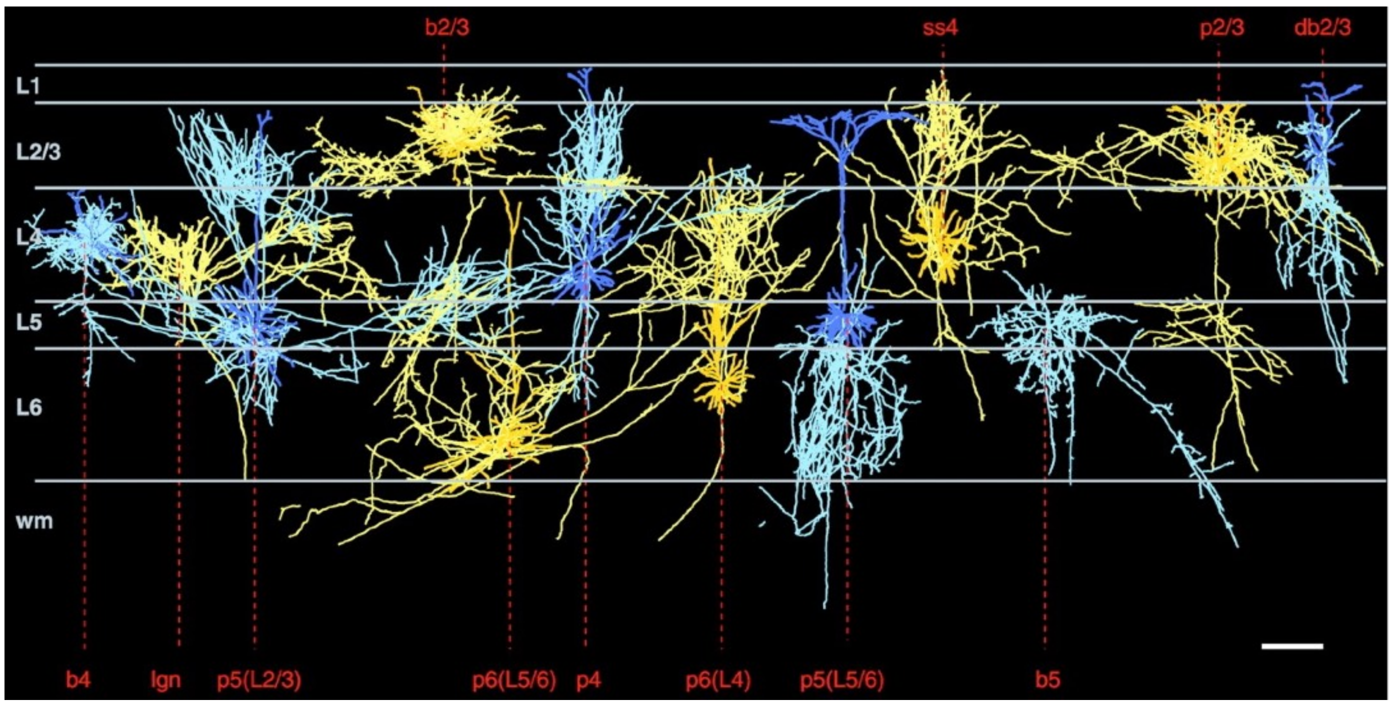
Wallisch & Movshon (2008)



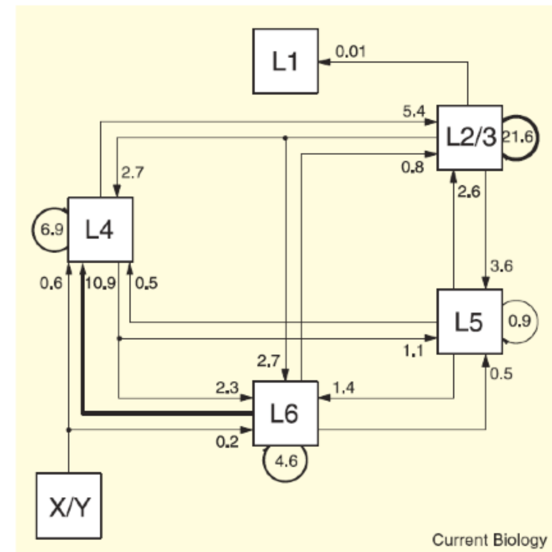
lines, edges surface boundaries ? objects faces


MDP
VIP
MIP
LIP
DP
PO
MST
PIP
MT
FST
7a
V1
V2
V3
STPp
STPa
V3a
V4
VOT
PIT
CIT
AIT
from
retina
LGN
Pulvinar

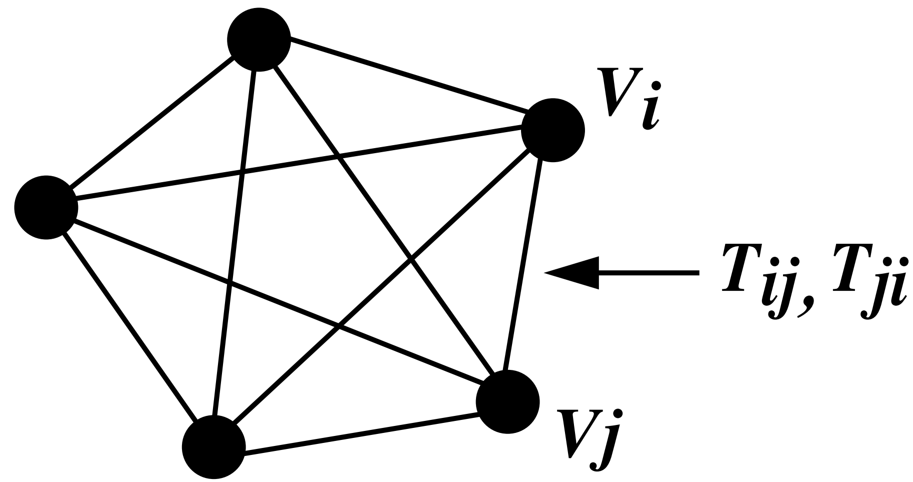
Recurrent neuronal circuits in the neocortex
(Douglas & Martin 2007)



(Binzegger, Douglas & Martin, 2004)



(Douglas and Martin, 2007)



Dynamics:

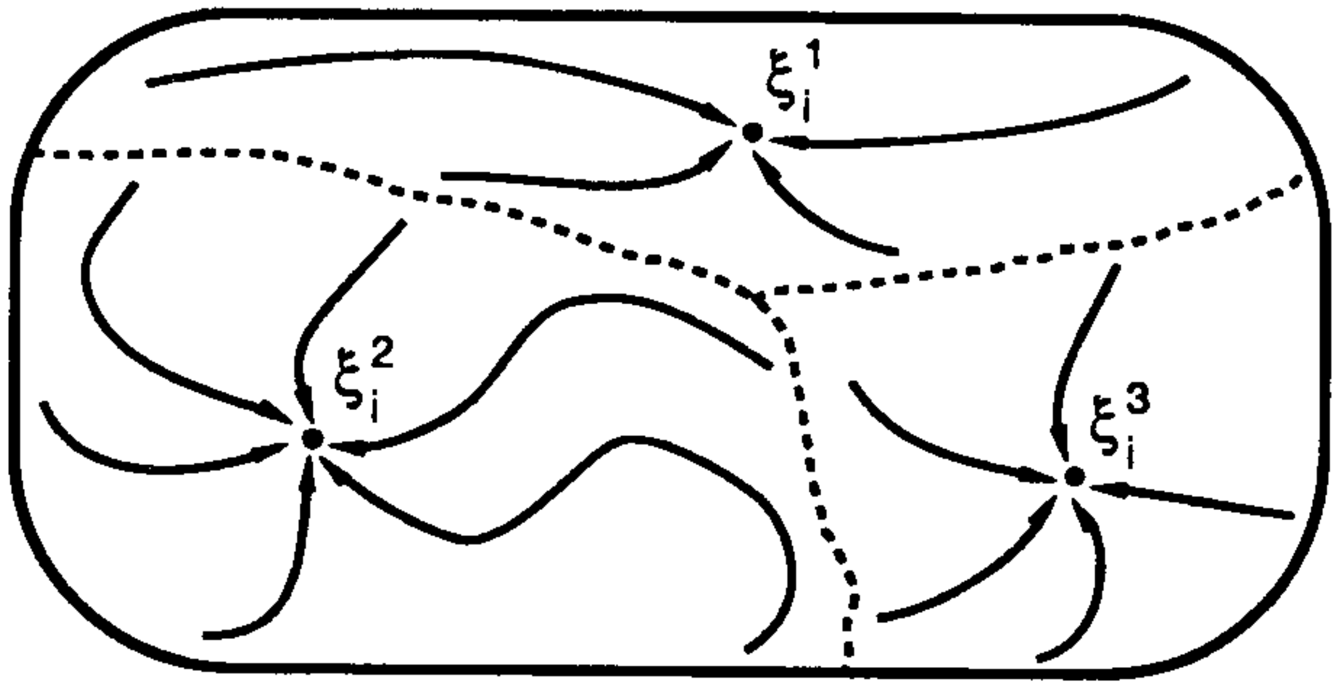
$$U_i = \sum_j T_{ij}\, V_j$$

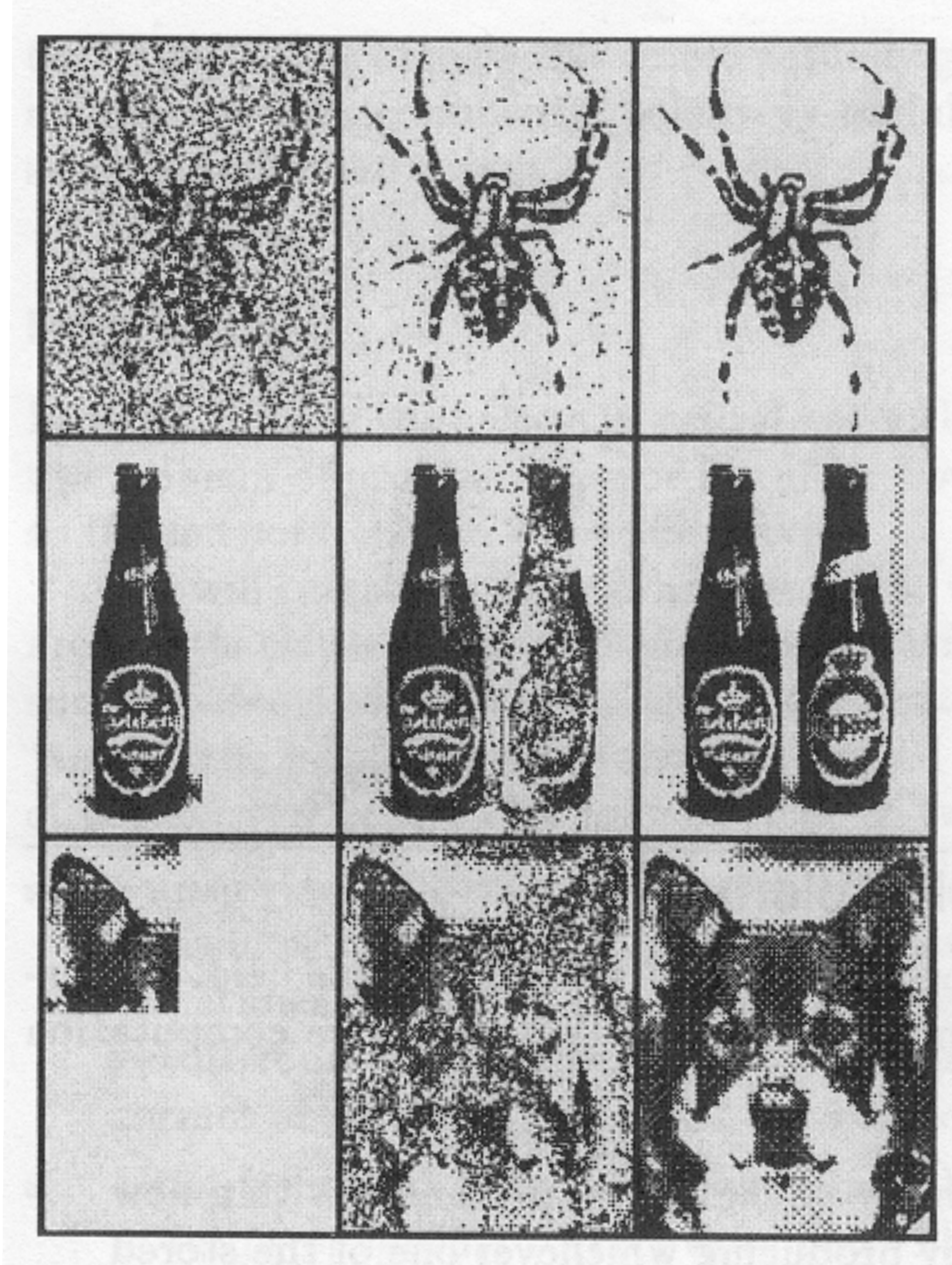
$$V_i = \mathrm{sign}(U_i)$$

Energy function:

$$E = -\frac{1}{2} \sum_{i,j \neq i} T_{ij}\, V_i\, V_j$$

Basins of attraction





input ⟶ recall

Outer-product (Hebb) rule

$$
\begin{aligned}
T_{ij} &= \sum_{\alpha} P_i^{(\alpha)} \, P_j^{(\alpha)} \\
&= P_i^{(1)} \, P_j^{(1)} + P_i^{(2)} \, P_j^{(2)} + P_i^{(3)} \, P_j^{(3)} + \ldots
\end{aligned}
$$

or $$\mathbf{T} = \mathbf{P}^{(1)} \, {\mathbf{P}^{(1)}}^T + \mathbf{P}^{(2)} \, {\mathbf{P}^{(2)}}^T + \mathbf{P}^{(3)} \, {\mathbf{P}^{(3)}}^T + \ldots$$

Thus

$$
\begin{aligned}
\mathbf{U} &\cong (\mathbf{P}^{(1)} \, {\mathbf{P}^{(1)}}^T + \mathbf{P}^{(2)} \, {\mathbf{P}^{(2)}}^T + \mathbf{P}^{(3)} \, {\mathbf{P}^{(3)}}^T + \ldots) \, \mathbf{V} \\
&= \mathbf{P}^{(1)} \, (\mathbf{P}^{(1)} \cdot \mathbf{V}) + \mathbf{P}^{(2)} \, (\mathbf{P}^{(2)} \cdot \mathbf{V}) + \mathbf{P}^{(3)} \, (\mathbf{P}^{(3)} \cdot \mathbf{V}) + \ldots
\end{aligned}
$$

$$\mathbf{V} = \mathrm{sgn}(\mathbf{U})$$

Capacity vs. error rate

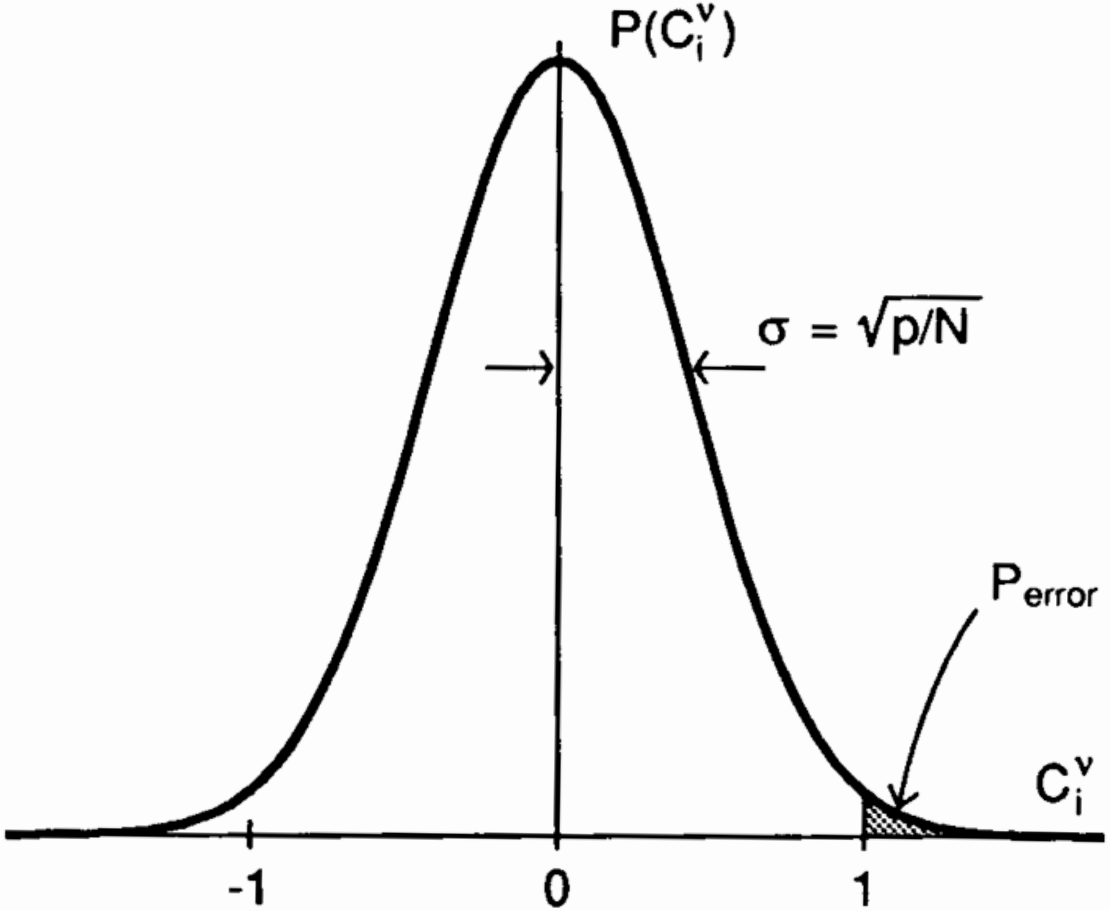
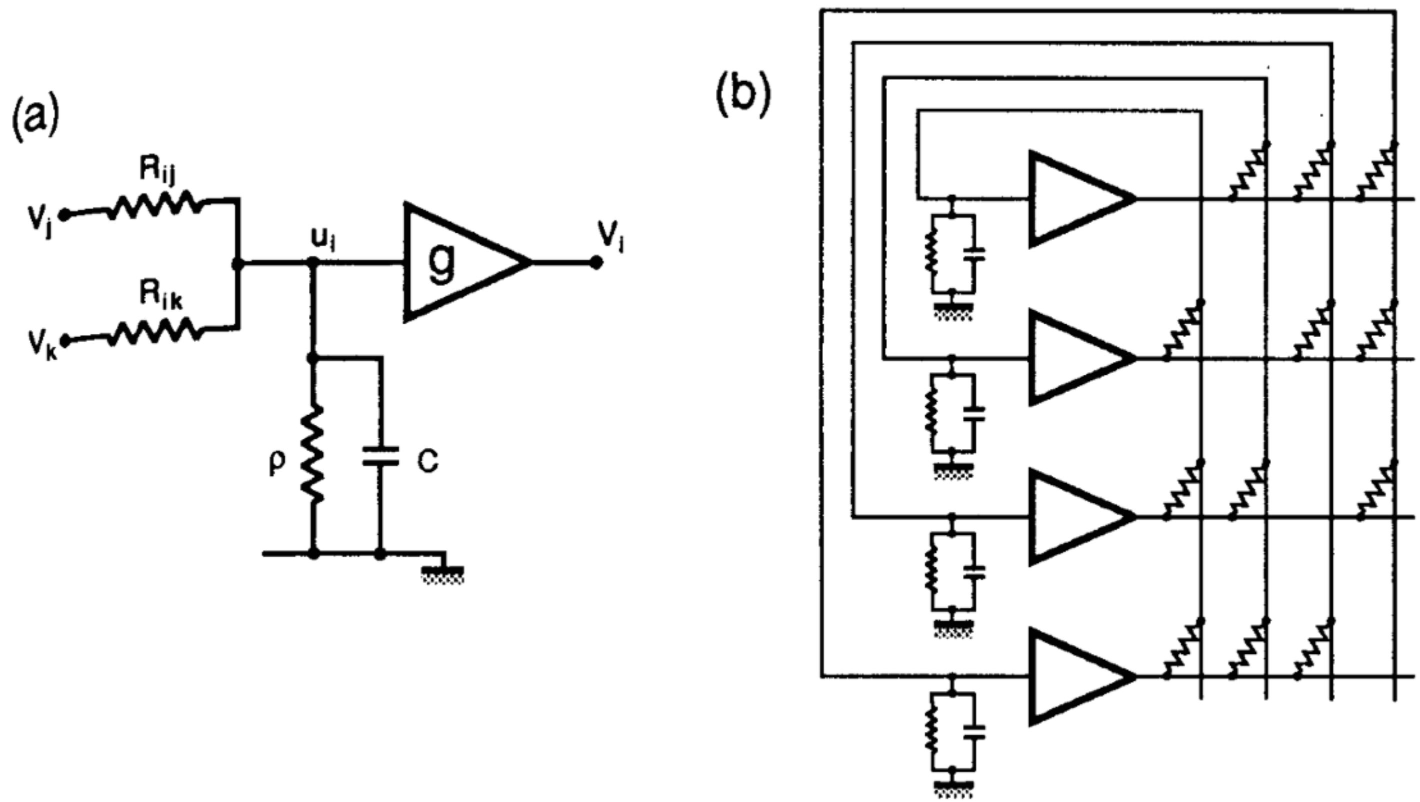


TABLE 2.1 Capacities

P_{error}	$p_{\max}/N$
0.001	0.105
0.0036	0.138
0.01	0.185
0.05	0.37
0.1	0.61

Hopfield network with analog units

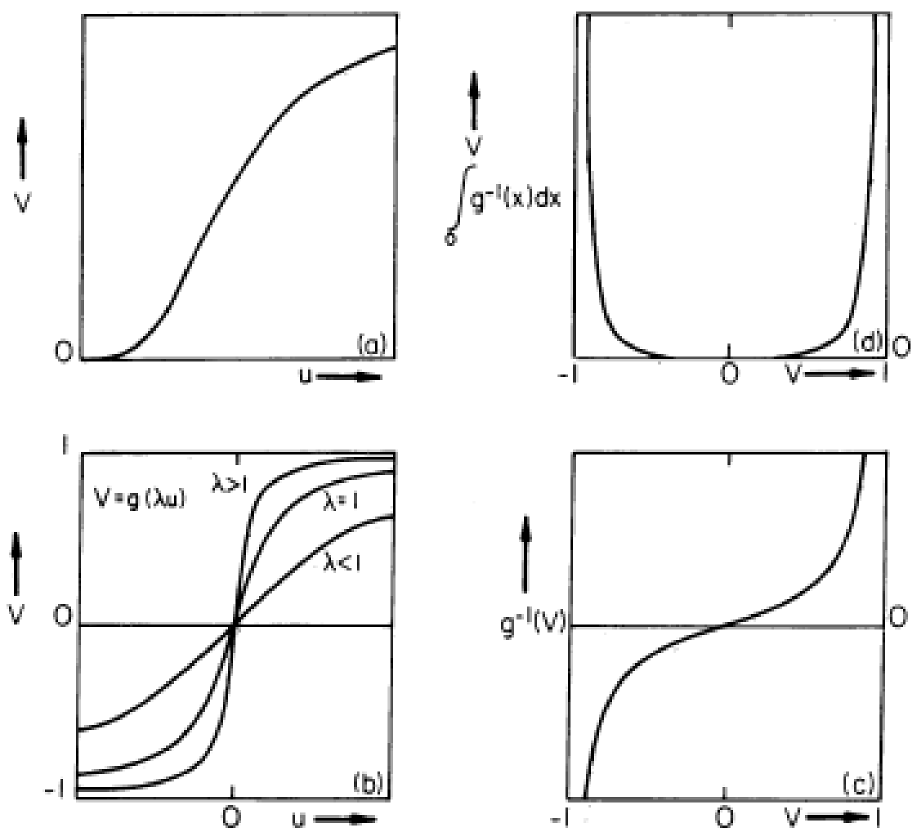


$$\tau\,\dot{u}_i + u_i = \sum_{j\neq i} T_{ij}\,V_j + I_i$$

$$V_i = g(u_i)$$

Lyapunov function

$$E = -\frac{1}{2}\sum_i \sum_{j\neq i} T_{ij}\, V_i\, V_j + \sum_i \int_0^{V_i} g^{-1}(V)dV - \sum_i V_i\, I_i$$



From Lyapunov function to dynamics

$$E = -\frac{1}{2}\sum_i \sum_{j\neq i} T_{ij}\, V_i\, V_j + \sum_i \int_0^{V_i} g^{-1}(V)dV - \sum_i V_i\, I_i$$

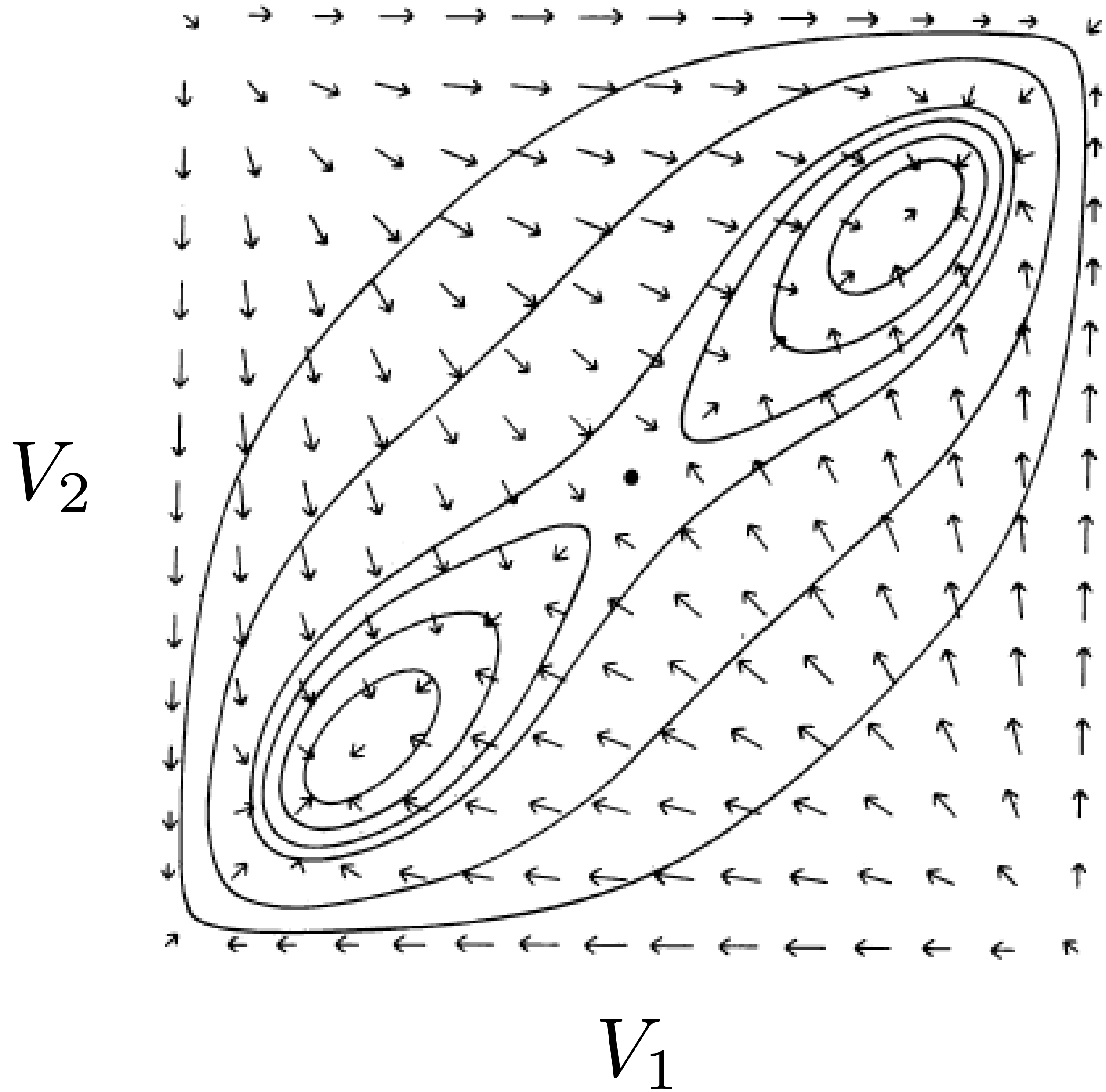
$$\frac{\partial E}{\partial V_k} = -\sum_{j\neq k} T_{kj}\, V_j + g^{-1}(V_k) - I_k$$

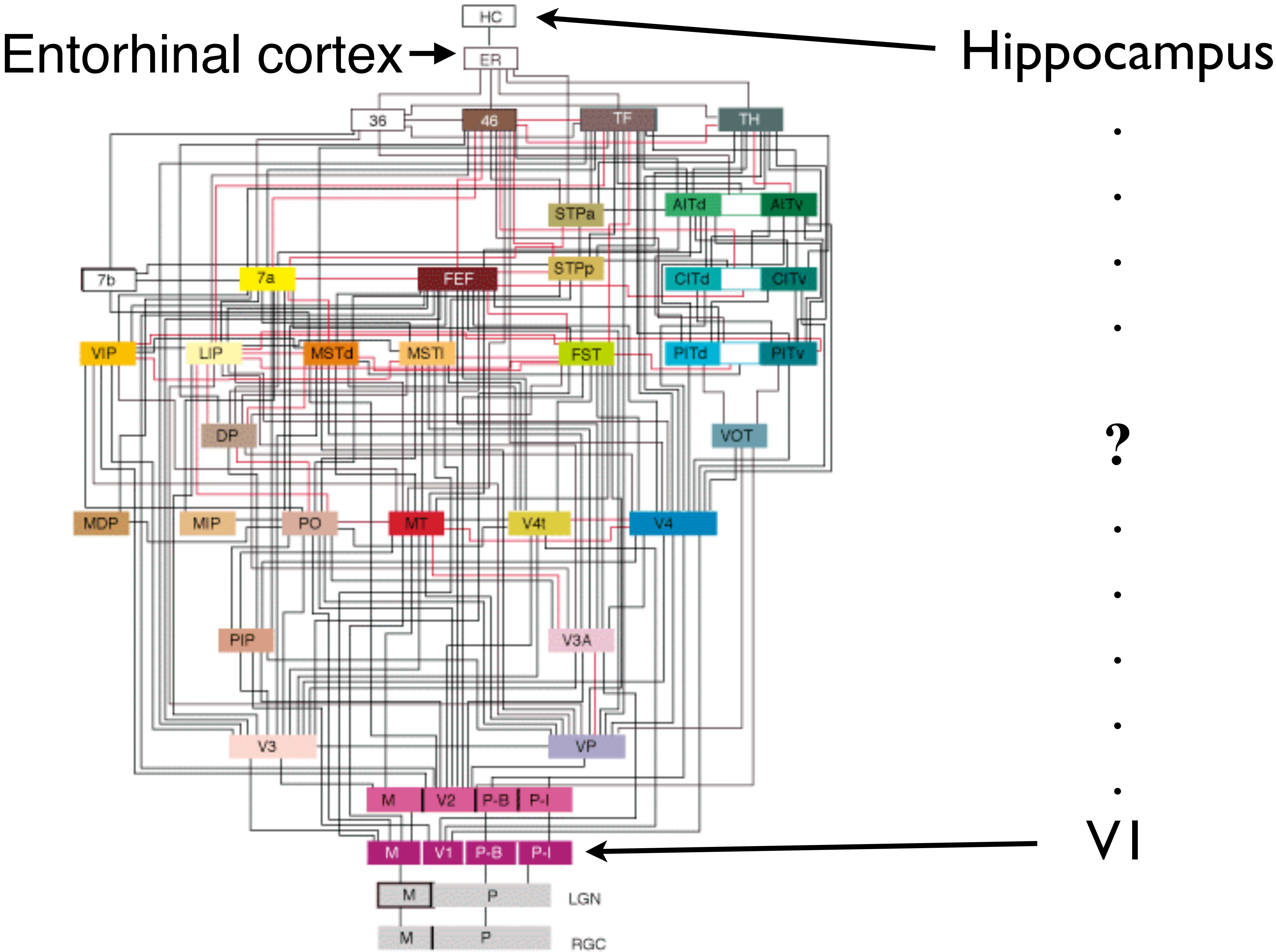
Let $\quad u_i = g^{-1}(V_i) \;\; \Rightarrow \;\; V_i = g(u_i)$

$$\dot{u}_i \propto -\frac{\partial E}{\partial V_i} = \sum_{j\neq i} T_{ij}\, V_j + I_i - u_i$$

Thus $\quad \dot{E} = \frac{\partial E}{\partial V}\,\frac{\partial V}{\partial u}\,\dot{u} < 0$

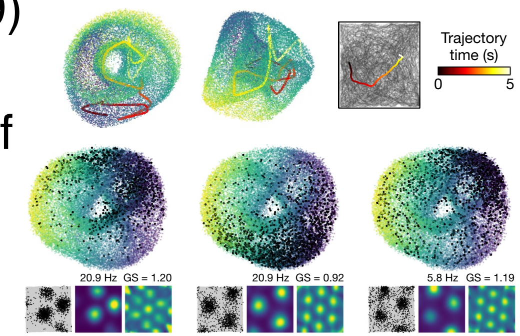
State space




Entorhinal cortex
Hippocampus
?
V1
HC
ER
36
46
TF
TH
STPa
AITd
AITv
7b
7a
FEF
STPp
CITd
CITv
VIP
LIP
MSTd
MSTl
FST
PITd
PITv
DP
VOT
MDP
MIP
PO
MT
V4t
V4
PIP
V3A
V3
VP
M
V2
P-B
P-I
M
V1
P-B
P-I
M
P
LGN
M
P
RGC

Place cells, grid cells, head-direction cells and continuous attractor neural networks

Gardner et al. (2022)
Toroidal topology



O'Keefe & Nadal (1978)
"The Hippocampus as a Cognitive Map"

Zhang (1996)
Continuous attractor model of head-direction cells

Burak & Fiete (2009)
Continuous attractor model of grid cells

1970 1980 1990 2000 2010 2020

O'Keefe & Dostrovsky (1971)

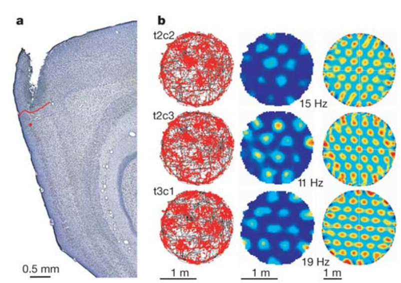
O'Keefe (1976)

"Place cells"

Ranck (1985);
Taube et al. (1990)

"Head-direction cells"

Hafting, Fyhn, Molden, Moser & Moser (2005)



"Grid cells"

Chaudhuri et al. (2019)
Ring topology

Seelig & Jayaraman (2015) - fly head-direction cells