ANNfk Lecture 3
Brains and ANNs

Anders Lansner, CB(N)

Goals of lectures 3 & 6

- Beyond ANN as a tool

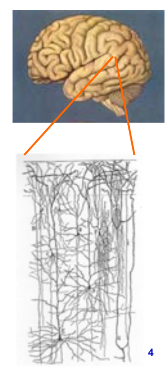
- Modern brain theory, relation to ANN
- Biorealistic model of neocortex
- Attractor networks, as CAM and for optimization
- Lessons from biology?
- Technological potential of brain-like computing
- NN hardware

ANN yesterday … today

- First ANN wave 1945 – 1969
- Second wave 1982 – 1995
- Today
  - ANN applications
  - AI-ANN research scattered
    - E.g. paradigm shift in Computer Vision
  - Computational neuroscience expanding
  - Neuroinformatics
    - Databases, Simulators, Brain imaging and Visualisering
    - ”Neuro-IT”
  - EU, NSF, DARPA, …, Japan
    - ”Reverse engineering the brain”
  - Startups …
- Third ANN wave?
  - Truly brain-like computing and architectures?
  - Artificial Nervous Systems, Artificial Brains

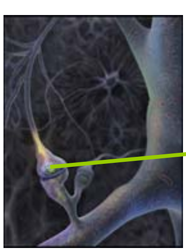
Our brain = ”alien technology”

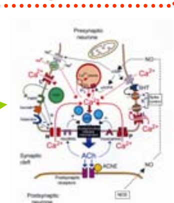
- Volume: 1.5 liters, 80% cortex
- Cerebral cortex
  - Dimension: 3 mm × 2 × 1200 cm^2
  - N:o neurons: 20 10^9
  - N:o synapses: 10^{14}, sparseness 10^{-6}
- Fiber length: 10 million km
  - >20 × distance Moon-Earth
  - 6000km/cm^2
  - 90 nm chip: 10 km/cm^2
- N:o messages/s: 10^{13}
- ”RAM”: >10^{15} bits
- Ops/s: >10^{19} Ops
- Power consumption: 30 W
  - Light bulb, PC processor



Brain networks – multi-scale view

The synapses == memory and compute nodes





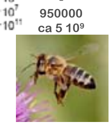
Graham Johnson Medical Media, Boulder, Colorado

Size of real neuronal networks

- In reality
  - Several types of different neurons
  - Huge numbers
  - Modular and layered

Table 1. The cortex data is summarized for a number of mammals [23]. All types of neurons are included in the counts.

	Human	Macaque	Cat	Rat	Mouse	Honey bee
Cortex Area (mm^2)	$2.4 \cdot 10^5$	$2.5 \cdot 10^4$	$8.3 \cdot 10^3$	$6.0 \cdot 10^2$	$2.5 \cdot 10^2$	-
Neurons	$2.0 \cdot 10^{10}$	$3.0 \cdot 10^9$	$6.0 \cdot 10^8$	$5.0 \cdot 10^7$	$2.0 \cdot 10^7$	950000
Synapses	$1.5 \cdot 10^{14}$	$2.2 \cdot 10^{13}$	$4.5 \cdot 10^{12}$	$4.0 \cdot 10^{11}$	$1.6 \cdot 10^{11}$	ca 5 10^9



- Quite similar over areas and species!
- Linear scaling also of synapses
- Very sparse connectivity

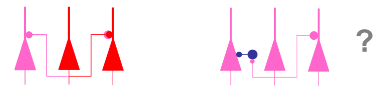
Biological nervous systems excel in …

- Real-time perception and control
- Sensor ”technology”
  - Sensor arrays, high sensitivity
- Holistic perception
  - Exploits context, multi-modality
  - Figure-background segmentation, information fusion
- Parallelism, distributed memory
  - Scalability (mouse to man)
  - N = 10^{11}/ 10^{15}
- Compactness
  - Low power dissipation (< 50W)
  - Slow, stochastic elements
  - Fault tolerance
- Can we understand HOW?

Synaptic plasticity and Hebbian synapses

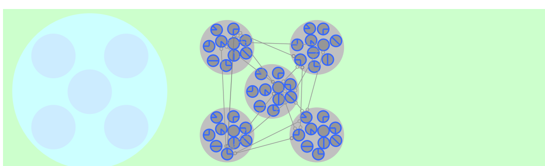
Hebb D O, 1949: The Organization of Behavior

- Experience dependent connectivity
- Co-activation of nerve cells ⇒ enhanced connection/synapse



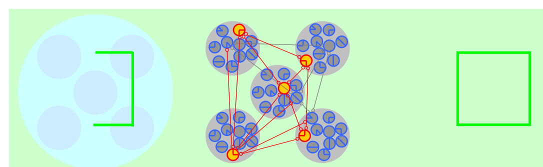
- Hebb’s hypothesis verified!
- Synaptic long-term potentiation (LTP), 1973, Hebbian, 1978
  - LTD, long-term depression
  - ”Spike-timing dependent plasticity” (STDP)!

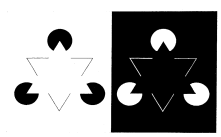
Hebbian cell assemblies – formation



Hebbian cell assemblies – dynamics

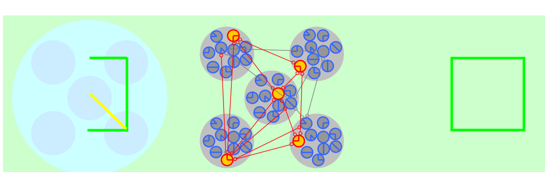
Perceptual/Pattern completion



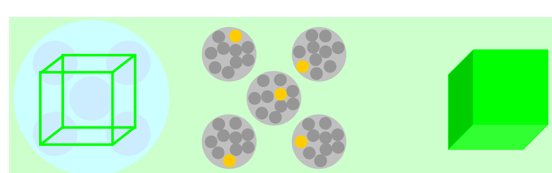


Hebbian cell assemblies – dynamics

Perceptual rivalry

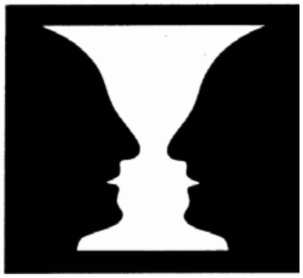


Hebbian cell assemblies – bistabilty



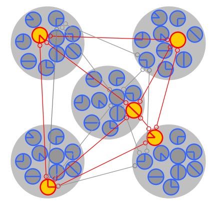
Necker cube

Hebbian cell assemblies – bistabilty



- Mechanism?
- Synaptic depression + neuronal adaptation?

Hebbian synapses and cell assemblies



- Cell assembly = mental object
- Gestalt perception
  - Figure-background segmentation
  - Perceptual completion
  - Perceptual rivalry
- After activity ≈ 100 ms – 1 s
- Association chains
  - + Central Executive → Thinking
- Suggests kind of ANN?

Mathematical instanciation of Hebb's theory
The Hopfield network

$$C_i \frac{du_i}{dt} = \sum_j W_{ij} v_j - \frac{u_i}{R_i} + i_i$$

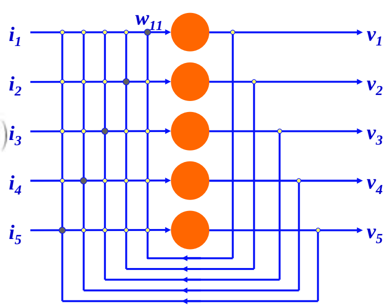
$$u_i = g_i^{-1}(v_i).$$

$$v_i = g_i(u_i) = \frac{1}{2}\left(1 + \tanh\left(\frac{u_i}{T}\right)\right)$$

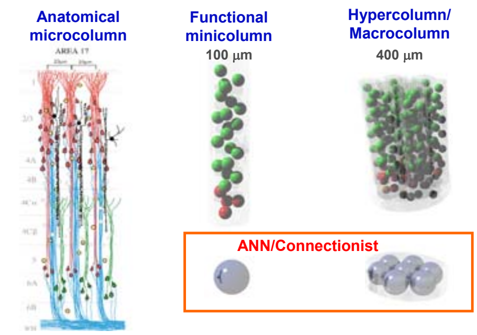
"Hebbsk inlärning"

$$w_{ij} = \frac{1}{N}\sum_{\mu=1}^{M} \xi_\mu \xi_\mu^T - \frac{M}{N} I$$

$$(w_{ii} = 0)$$



Neuron == ANN unit?



Mouse to Man?

Applications of attractor networks

- Mathematical cortex model
  - Associative memory model
  - Content Addressable Memory (CAM)
    - Storage capacity
    - Sparse activity
    - Diluted connectivity
  - Lots of modeling done
    - Connectionist → Biophysically detailed
- Combinatorial optimization
  - E.g. TSP

CAM – different (attractor) networks

- Hopfield
  - Potts network – modular, normalization
- Willshaw-Palm
  - Binary W
- …
- BCPNN
  - Bayesian probabilistic, modular, normalization
- How many memories can cortex store?
- How many does it need to store? Forgetting?

Performance measure – storage capacity

- Number of retreivable (random) patterns
- Bits/connection (free parameter)

- Noise analysis
- Information theoretical analysis
- ”Replica theory”

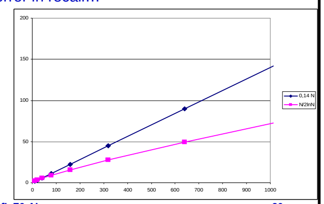
Hopfield networks – storage capacity

- Random patterns, 50% activity
- ”Crosstalk”, variance analysis

$M_c = 0,14N$ with error in recall…

$M_{\max} = \dfrac{N}{2\ln N}$ without error in recall…

- Much analysis exists



Hopfield networks – storage capacity

Information theoretic estimate

- N:o bits in W (symmetric) $bits_W = \dfrac{N(N-1)}{2}$
- N:o bits/retreived pattern $\binom{N}{m} = different\,(binary)\,patterns\ possible$

$$bits_P = bits/pattern = \log_2 \binom{N}{m}$$

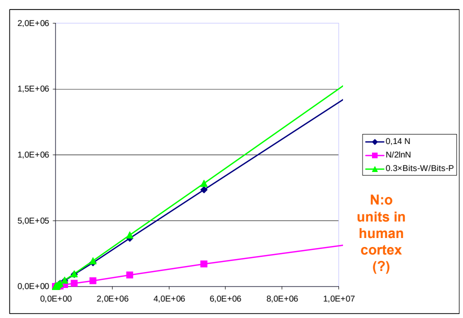
$$Z_{\max} = k\frac{bits_W}{bits_P}$$

- *k* = 0.2 – 0.3 (bits stored by each synapse)
- Reasonable for standard Hopfield
  - … and for attractor networks generally!

Efficient memory!

Hopfield networks – storage capacity

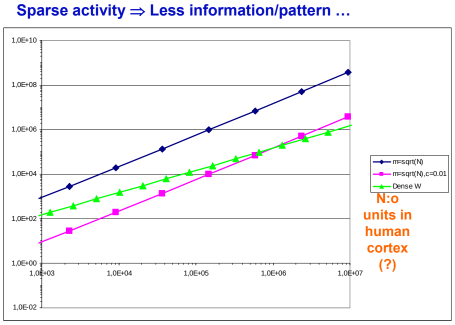
Information theoretic estimate



Hopfield networks – storage capacity

Sparse activity a = sqrt(N), 10^{-2} connectivity

Sparse activity ⇒ Less information/pattern …



Some further remarks

- Skewed random patterns
  - Common in real world data
    - E.g. varying feature detector selectivity
  - Willshaw-Palm does not work
  - Hopfield neither …
    - High utility units gets strongly connected
    - Convergens to a common pattern
  - BCPNN works well
    - High utility units gets weakly connected
  - …
- ”Dynamic sparsity”
  - Synaptic spine dynamics …
  - Improves bits/synapse capacity
- Cortical modularization still uncertain

Combinatorial optimization using attractor networks – useful at all??

15

INFORMS Journal on Computing
Vol. 11, No. 1, Winter 1999

0899-1499/99/1101-0015 $05.00
© 1999 INFORMS

Neural Networks for Combinatorial Optimization: A Review of More Than a Decade of Research

KATE A. SMITH / *School of Business Systems, Monash University, Clayton, Victoria, 3168, Australia*
Email: ksmith@bs.monash.edu.au

- Compares Hopfield networks (and SOM) to ”meta-heuristic” methods
- Hopfield and Tank, TSP, 1985
  - Energy function
  - … equivalent to the function to be minimized
  - ”Constraints” included as penalty terms
  - Legal & good solutions desirable

Combinatorial optimization using attractor networks - critisism

- Wilson and Pawley, TSP, 1988
  - Could not reproduce Hopfield & Tank 1985
  - Only fraction of tours valid, not very short
  - ”our simulations indicate that Hopfield and Tank were very fortunate in the limited number of TSP simulations they attemtped”
- ➔ Method in doubt! ➔ heuristic methods

Combinatorial optimization using attractor networks – improvements

- Reformulation of energy function ➔ More legal solutions
  - ”Valid subspace approach”
- More optimal solutions
  - Stochastic ”annealing”
    - E.g. Bolzmann-Cauchy machine
  - Units with chaotic dynamics
- Dedicated hardware, FPGA
- ”The initial problems that have plagued the reputation of the Hopfield network have now been solved.”

Combinatorial optimization using attractor networks – applications

- Comparison to heuristic methods often lacking …
- TSP
- Assignment problems
  - N tasks to N persons
    - Job-shop scheduling
    - Other man-power scheduling
  - Quadratic
    - VLSI design, channel assignment in telecom
- ”Constraint satisfaction”
  - N-Queen
- Scheduling
  - E.g airline crews
- Stock cutting, Partitioning, Bin packing
- Integer programming

Recent TSP evaluation

2006 International Joint Conference on Neural Networks
Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada
July 16-21, 2006

A Columnar Competitive Model with Simulated Annealing for Solving Combinatorial Optimization Problems

Eu Jin Teoh, *Student Member, IEEE*, Huajin Tang, *Member, IEEE*, and Kay Chen Tan, *Member, IEEE*

- WTA modules (columns)
  - Ensuring exactly one visit/step
  - Some invalid states disabled (may have 2 ones on row)

$$E(\mathbf{v}) = \frac{K}{2}\sum_x\sum_i(v_{x,i}\sum_{j\neq i}v_{x,j}) + \frac{1}{2}\sum_x\sum_{y\neq x}\sum_i d_{xy}v_{x,i}(v_{y,i+1}+v_{y,i-1}), \quad (6)$$

$$W_{xi,yj} = -\{K\delta_{xy}(1-\delta_{ij}) + d_{xy}(\delta_{i,j+1}+\delta_{i,j-1})\} \quad (7)$$

Recent TSP evaluation, cont’d

- $K>2d_{max} - d_{min}$
  - Converges
  - Always valid tour
- $K<2d_{min} - d_{max}$
  - Never valid tour
- Solution quality improves with <u>decreasing</u> K …

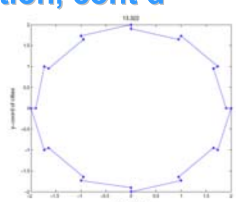


TABLE I
THE PERFORMANCE OF ORIGINAL HOPFIELD MODEL FOR THE 24-CITY EXAMPLE

C	Valid	Invalid	Good	Minimum length	Average length
10	302	198	111	35.2769	50.9725
1	285	215	108	26.2441	37.3273
0.1	366	134	168	25.2030	36.3042
0.01	360	140	159	27.5906	36.9118
0.001	372	128	153	21.6254	32.5432

TABLE II
THE PERFORMANCE OF CCM FOR THE 24-CITY EXAMPLE

K	Valid	Invalid	Good	Minimum length	Average length
d_{max}	438	62	248	13.3220	18.1172
$d_{max} + d_{min}$	453	47	256	13.3220	17.9778
$2d_{max} + d_{min}$	500	0	18	15.1856	24.1177

Recent TSP evaluation, cont'd

- Order of 10 iterations
- + Simulated Annealing

TABLE III
THE PERFORMANCE OF CCM WITH SA FOR THE 24-CITY EXAMPLE

ϵ	Valid	Invalid	Good	Minimum length	Average length
0	500	0	18	15.1856	24.1177
1	500	0	203	13.3404	22.2688
2	500	0	218	13.3220	19.9771
3	500	0	259	13.3220	18.7153
5	500	0	256	13.3220	18.6785
10	500	0	301	13.3220	18.4080
15	500	0	307	13.3220	18.3352
20	500	0	289	13.3220	18.4643

TABLE IV
THE PERFORMANCE OF CCM FOR THE 48-CITY EXAMPLE

K	Valid	Invalid	Good	Minimum length	Average length
d_{max}	90	10	0	24.6860	32.7950
$d_{max} + d_{min}$	92	8	0	24.9202	33.8892
$2d_{max} - d_{min}$	100	0	0	35.9438	42.4797

Recent TSP evaluation, cont'd

- Larger problems …
- Perf. ratio = minimum/average

TABLE V
THE PERFORMANCE OF CCM WITH SA FOR VARIOUS CITY SIZES

City size	Simulations	Valid	Invalid
30	250	250	0
100	250	250	0
500	100	100	0
1000	50	50	0
3000	5	5	0

- Other tours
- Better SA
- Map coloring problem

Possible ANNfk project

TABLE VI
THE PERFORMANCE OF CCM WITH SA FOR VARIOUS CITY SIZES
(CONTINUATION FROM TABLE ABOVE)

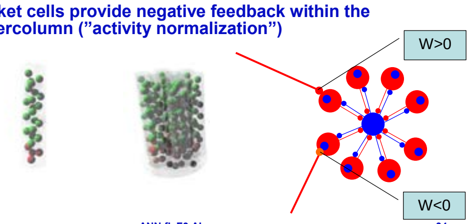
City size	Minimum length	Average length	Perf. ratio
30	13.6922	22.3159	0.6136
100	28.2552	35.6920	0.7916
500	257.1380	283.2810	0.9060
1000	454.3827	490.6963	0.9260
3000	1395.5	1454.2	0.9455

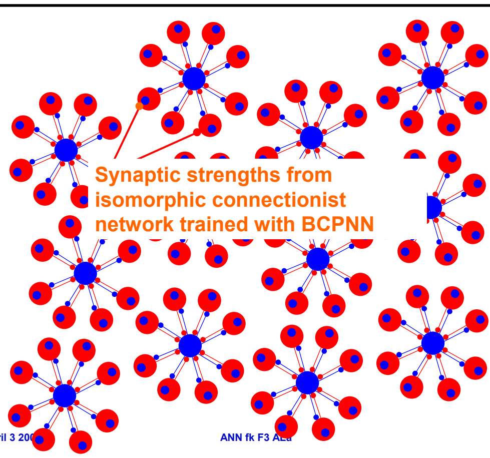
Back to biology: Attractor memory and cortex

- Biological attractor memories?
  - Olfactory cortex, Hippocampus
    - Recurrent connectivity
    - Inspiration for CAM models
  - Entire neocortex: Multiple connected attractor networks?
  - …
- →Biophysically detailed model of neocortex
  - Structure as "modular Hopfield network"

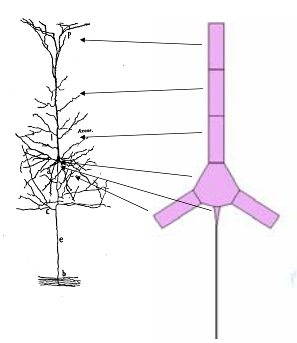
Units/Minicolumns, Hypercolumns

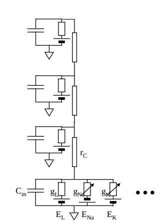
- Unit => cortical minicolumn
  - 30 pyramidal cells
  - 2 "regular spiking non-pyramidal cells" (RSNP)
    - "double bouquet", …
  - "Functional minicolumn"
- Hypercolumns with ca 100 minicolumns = 3000 cells
- Basket cells provide negative feedback within the hypercolumn ("activity normalization")





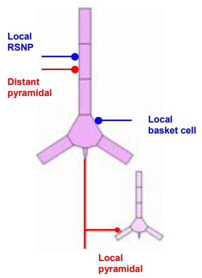
Biophysically detailed cell modell





Hodgkin-Huxley formalism
Na, K, K_{Ca}, Ca-channels
Ca_{AP} and Ca_{NMDA} pools
Fast synaptic transmission:
Glutamatergic (50/50 AMPA-NMDA)
GABA-ergic
Mostly data from McCormick et al.
SPLIT simulator

Synapse properties

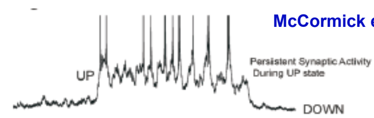


Tsodyks, Uziel, Markram 2000

- Realistic amplitude of PSP:s in largest network model
- Asymmetric cell-cell connectivity
- 2D geometry ⇒ delays, 0.1 - 1m/s

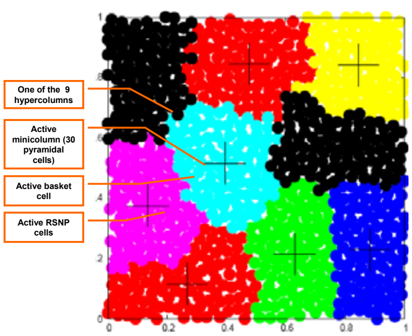
Some questions we had

- Is cortical recurrent connectivity strong enough?
  - … to support pattern completion and rivalry
- How much of spike synchronicity, gamma rhythm, theta rhythm, … etc
- Can the model reproduce ”UP/DOWN-states” observed *in vitro* in cortical slices?
- Model of cortex layer 2/3, horizontal connectivity



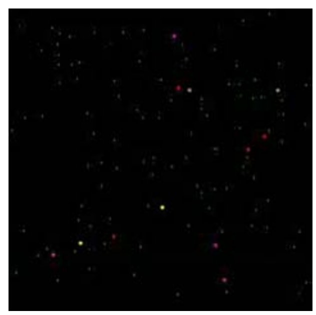
McCormick et al, 2003

Network layout



- 1x1 mm patch
- 9 hypercolumns
- Each hypercolumn
  - 100 minicolumns
  - 100 basket cells
- 100 patterns stored
- 29700 neurons
- 15 million synapses

9 hypercolumns



- 1x1 mm patch
- 9 hypercolumns
- Each hypercolumn
  - 100 minicolumns
  - 100 basket cells
- 100 patterns stored
- 29700 neurons
- 15 million synapses

100 hypercolumns

- 330000 neurons
- 161 million synapses

≈ 4x4 mm

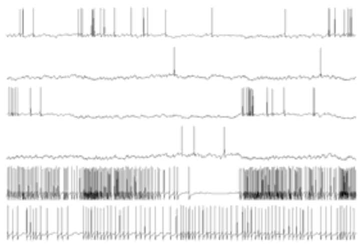
8 rack BG/L simulation

- 22x22 mm cortical patch
  - 22 million cells, 11 billion synapses
- 8K nodes, co-processor mode
  - used 360 MB memory/node
- Setup time = 6927 s
- Simulation time = 1 s in 5942 s
- >29000 cpu hours
- Massive amounts of output data
- 77 % of linear speedup
  - Point-point communication slows (?)
- Currently (inofficial) world record!

Djurfeldt, M., M. Lundqvist, C. Johansson, M. Rehn, Ö. Ekeberg and A. Lansner (2008). "Brain-scale simulation of the neocortex on the IBM Blue Gene/L supercomputer." *IBM J Res Dev.* 52: 31-41.

The three different cell types

3 sec simulation

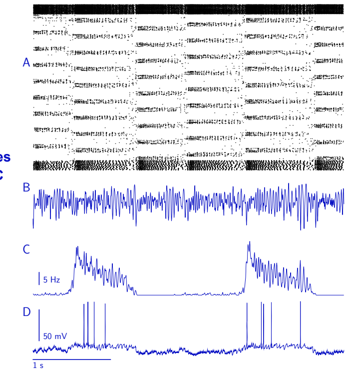


Pyramidal

RSNP

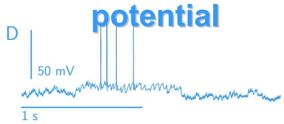
Basket

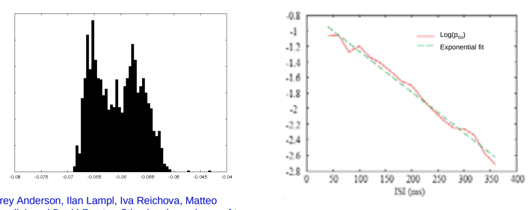
- 2000+ neurons
- 250000+ synapses
- 5 s = 600 s on PC



Lundqvist M, Rehn M, Djurfeldt M and Lansner A (2006). Attractor dynamics in a modular network model of the neocortex. *Network: Computation in Neural Systems*: 17, 253-276

Bimodal membrane potential



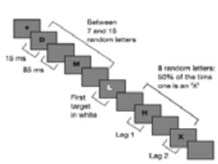


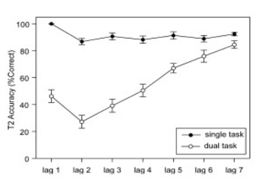
Jeffrey Anderson, Ilan Lampl, Iva Reichova, Matteo Carandini, and David Ferster. Stimulus dependence of two-state fluctuations of membrane potential in cat visual cortex. *Nat. Neurosci.*, 3(6):617–621, 2000.

Perception and associative memory performance

- Pattern reconstruction
  - Figure-background
  - Pattern completion and rivalry
  - 50 – 100 ms
- Sustained after-activity
  - 150 ms – 2 sec
  - $NMDA_{Ca}$, K_{Ca} modulation
- Robust to parameter changes and scaling
- Cortical long-range recurrent excitation strong enough to support attractor dynamics
- Spiking statistics as *in vivo*
- Can even reproduce cognitive phenomena…
  - … like ”Attentional Blink”

What is Attentional Blink?

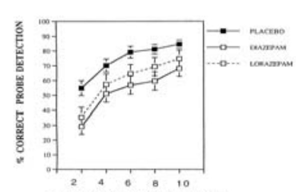


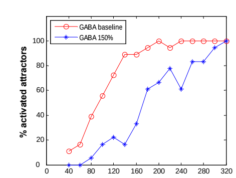


Adapted from Kranczioch, Debener, Engel 2003

- Occurs when two expected stimuli are presented < 500 ms apart
- Experimentally, stimuli are changed every 100ms; exposure 85-100ms
- Can occur in vision and audition, likely any modality
- A cognitive component of general reaction time
- Effect is diminished when interval is <= 100ms (lag-1 sparing)

Attentional blink – effect of GABA↑





- Attractor activation correlates with percentage of correct probe detections
- Time scales different but qualitatively similar results

Ongoing and planned work

- Add a layer 4
  - Selective feature detectors
  - V1 model
- Add a Layer 5
  - Martinotti cells, delayed inhibition to superficial layers
  - Pyramidals, cortico-cortical connections
- Overlapping stored memories
- Analysing dynamics, spiking statistics, conductances, intracellular potentials
- Better synthetic VSD, BOLD signals