ANN lecture 6

Neural computation and implementation

Anders Lansner CB(N)

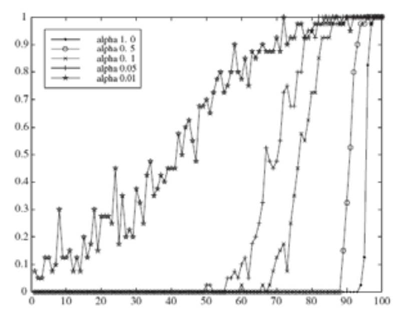
Synposis lecture 6

- Extensions of attractor networks
  - Sparse activity
  - Diluted W
  - Modularity
  - Forgetting - Leaky weights
  - Hidden layer
  - Adaptation
- Internal representation
  - Fuzzy coding
  - Invariant recognition
- Implementation of brain-like computing
  - incl NN hardware

Extension 4
Continuous learning, forgetting

- "leaky" weight updating $\frac{dw_{ij}}{dt} = \frac{\xi_i^{\mu}\xi_j^{\mu} - w_{ij}}{\tau_w}$
- Time constant for weights, τ_w
  - Avoids "catastrophic forgetting"
  - "Palimpsest memory"
  - Long-term >> short-term >> working memory
  - Memory consolidation
    - Prefrontal cortex ↔ Hippocampus ↔ Parietalcortex
    - Dream sleep …

Forgetful attractor memory



A good description of human memory?

List learning...

Standby!

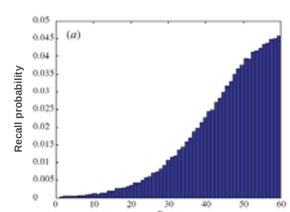
END of list

In the list?

boat
hitler
airplane
bus
rollerblades
tomato

Relevance modulation of memory

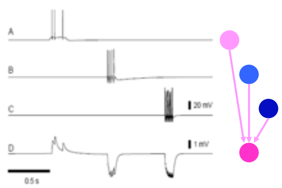
- Print-now for each pattern, τ_w
  - Memory modulation, emotion/relevance, dopamin D1



Good model of LTM, ITM, WM …?

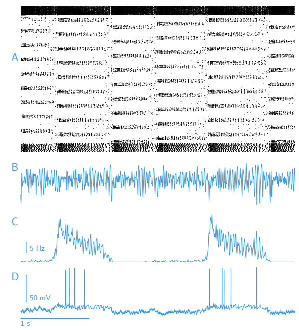
Extension 6: Adaptation

- Neurons – properties
  - Current – Frequency
  - Adaptation
  - Synaptic depression
- What happens in an attractor network?
  - (Hopfield network sticks in minima)
  - Sensitivity to new input
  - Quasistable attractors
  - Attractor wandering

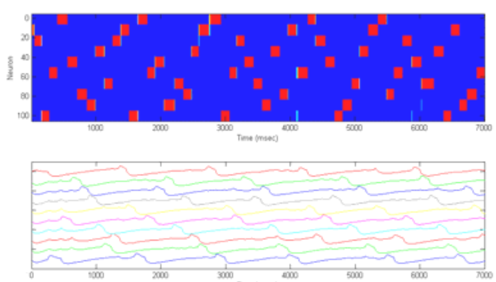


Extension 6: Adaptation

- Neurons – properties
  - Current – Frequency
  - Adaptation
  - Synaptic depression
- What happens in an attractor network?
  - (Hopfield network sticks in minima)
  - Sensitivity to new input
  - Quasistable attractors
  - Attractor wandering



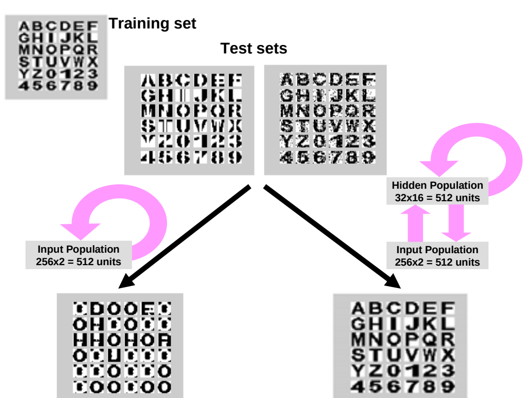
Attractor ANN with adaptation



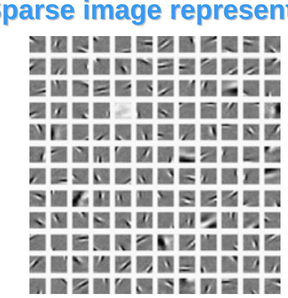
Good storage capacity etc, but …

- Remaining limitations?
- Missing functionality?

- Many and correlated patterns!
- Time – sequences?



Sparse image representation



"Over complete" = n:o basis functions > n:o pixels in a patch

"Sparse basis functions for the patches were extracted from training images by an iterative learning procedure. Most of the functions are oriented, localized and cover a limited range of spatial frequencies, like the receptive fields of neurons in the visual cortex."

Cortex Layer IV

Image courtesy of Bruno A. Olshausen.

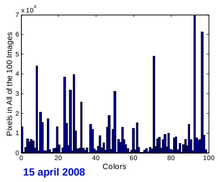
Example – image reconstruction

Images from Colombia University Image Library COIL-100.

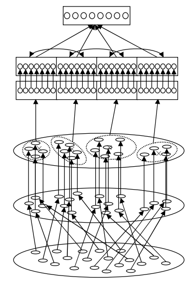
Resolution 128×128 pixel

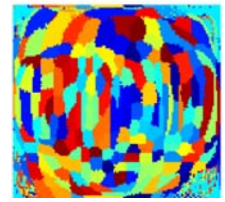
100 images stored in ANN with 16384 hypercolumns of size M = 100

Largest. N = 1.6 10^6, N_W = 2.0 10^{11}



Self-organization of image patches





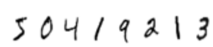
Extension 5

Attractor memory with hidden layer

- Without hidden layer
  - More patterns → unstable attractors
    - ”Catastrophic forgetting” if
      - Large number of random patterns, few correlated patterns
- With hidden layer
  - Self-organization – Competitive learning
    - ”Feature extraktion”
    - Unsupervised learning, can run continuously
      - Compensation for e.g. sensor drift
  - Pattern decorrelation
  - → storage capacity not limited by N_{in}
  - → improved classification rate (in FF configuration)
  - Previous capacity relations now holds for the hidden layer …

Example: Insect brain-inspired ANN

- Fast and robust learning by reinforcement signals in the insect brain
  - Ramón Huerta, Thomas Nowotny (2008), Centre for Computational Neuroscience an Robotics, Dept of Informatics, University of Sussex, Brighton, UK
- Basic setup
- Evaluated on MNIST dataset

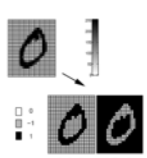


Insect brain-inspired ANN cont´d

+ Type 1 reinforcement
  - Connection update only on error
  - 75%
+ Type-II reinforcement
  - No strengthening if already correct answer
  - 80%
+ Regulation of KC activity level
  - [1% 10%]
  - 85%
+ On-Off cells
  - [2% 5%]
  - 91%
+ Type-III reinforcement
  - Decrease synapses firing wrong output
  - 97,5%

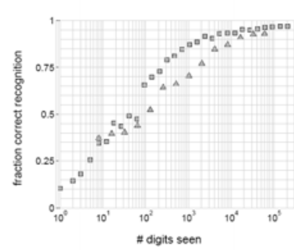
$$w_{ij}(n+1) = H(z_i, y_j, w_{ij}(n)),$$

$$H(z,y,w) = \begin{cases} w+1 & z=1,\ y=1,\ \xi < p_+, \\ [w-1]_+ & z=1,\ y=0,\ \xi < p_-, \\ w & \text{otherwise.} \end{cases}$$



Insect brain-inspired ANN cont´d

- Results for different variants
- Comparison to SVM
- A possible ANNfk project!

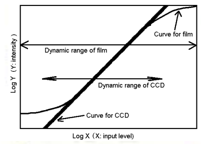


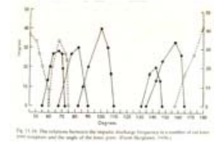
Extension 7
Time sequences

- Research remains …
  - Little known from biology
- Hidden units detecting change
- Slow, asymmetric Hebbian synapses

Value representation in CNS

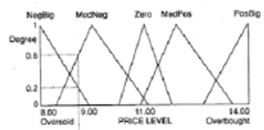
- Engineering: Linear sensors with high sensitivity&selectivity, broad dynamic range, …
- Fuzzy coding, population codes and interval sensors
  - Gaussian Mixture Models (GMM)
- Biological sensor arrays and motor units
  - Orientation coding in vision, Frequency coding in audition
  - Joint angle sensors
  - Motor unit pool and reqruitment, Populations codes in M1
  - …
- Fits into computational models of cortical processing
  - Hyper/Macrocolumns in cortex





”Fuzzy representation” – neurofuzzy

- Fuzzy variables
  - Linguistic, nominal
  - Interval coding
- Manual subdivision
- Graphic mixture model
  - Automatic, EM
  - Competitive learning



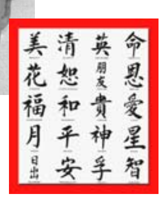
Hypercolumn with ? units

- Early popular in Japan
  - In consumer products
    - Refrigerators
    - Egg boilers
    - …
  - Soft train braking (Shinkhansen)
- Structure – property relations of chemical substances
  - Astra-Zeneca

Invariant recognition

- Object transformations
- Images
  - Translation, size, rotation 2D,3D
  - Light conditions, e.g. colour of ambient light
- Sound
  - Female/Male voices, e.g. Soprano/Baritone
    - Tempo
- NOT built-in for ANN
  - ”iconic” representation ➔ lacks invariance…
- What about our own brain?

Invariant perception?



Invariance?

MYTOMAN

FANTAST

Invariance?

MYTOMAN

FANTAST

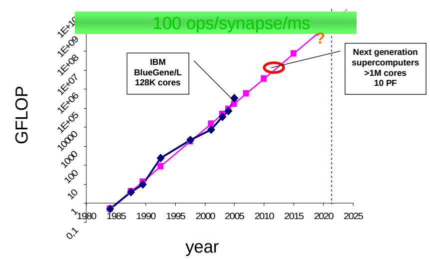
Two ways to achieve invariant recognition

- Haykin 29ff
- Built in, needs domain knowledge/evolution
  - Invariant feature detectors
  - For instance, 30º 60º 90º corners independent of position, orientation, size,
  - + feature histogram
  - Time warping of speech signals
- Through training
  - Learn characteristic ”views”
  - Interpolation between these
  - E.g. fighter aircrafts from different angles
  - ”Slow feature analysis”, SFA
    - Wiscott L. and Sejnowski, T. 2002
      - Slow Feature Analysis: Unsupervised Learning of Invariances, Neural Computation 14, 715 - 770
  - Sequence learning (use correlactions over time) during continuous observation

Brain-like technology - prerequisites

- I: Understanding of the brain’s operating principles
- II: Efficient implementation
  - Real time
  - Compact
  - Low power
  - Cheap

Computing Power
Moore’s law …



How to build an Artificial Brain?

- CCortex, Numenta, IBM are trying …

- Cluster computers
- Accelerator cards – graphics/Cell processor
- FPGAs
- ANN hardware - dedicerat kisel
  - Digital/Analog

- Biochip
- Molecular scale computing
- …

Scalable cluster implementation

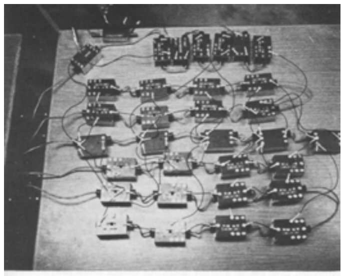
- Demands on
  - Memory
  - Memory bandwidth
  - Computation
  - Communication
    - MPI
    - Spiking 0/1 unit output
      - Address Event Representation (AER)
    - NOT limiting
- $N = 1.6\ 10^6$, $|W| = 2.0\ 10^{11}$
  - 256 nodes on PDC/Lenngren
  - Real time! learning/recall ca 0.5 sec/pattern
  - Highly efficient code

ANN hardware

- **Why?**
  - Real time operation, compactness, low power
- **Preconditions**
  - Inherent massive parallelism
  - Locality in computation
  - Tolerance to hardware faults
  - Tolerance to noise in data and components
- **How?**

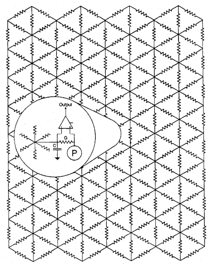
Early ANN hardware

- 1977...
  - 12 neurons, 36 synapses



Analog VLSI?

- High speed, compact, power efficient
- Embedded – directly connceted to sensor/actuator
  - No A/D or D/A
- Efficient implementation of specific biological structures
  - Carver Mead, Caltech 1980's
  - Cochlea
  - Retina
    - Too fast ...!
  - → Synaptics Touchpad
  - → Optical mouse



- Noise tolerance of ANN crucial!

Analog VLSI - properties

- **Different categories**
  - No learning (e.g. Si retina)
  - Off-chip learning, downloading of W
  - On-chip learning, fast, continuous learning
    - Standard BP not suitable (>16 bits, FPU)
    - Stochastic gradient-decent have been used
- **Problems**
  - Temperature variations influences computation
  - ”Zero offset”, may not exceed 10% of signal (requires careful design and fabrication)
  - High cost of design and test
  - Memory hierarchy, off-chip memory too slow
    - 10^5 → 10^4
  - Slow off-chip communication

EU IP FACETS

From cortex physiology to analog VLSI 2005-2009

- ”... reverse engineering of the brain ...”
- 14 research groups, 100 MSEK/4 år
- Cortex physiology
- Small- and large-scale simulations
- Digital&Analog (”mixed signal”) VLSI
  - Wafer scale integration
- FACETS II now planned

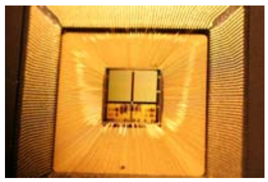
Process Technology	UMC 0.18 µm CMOS
Wafer Size	8 inch
Synapse Size	10 µm²
Synapses per Wafer	$1.5\ 10^7$
Synapse-to-Neuron Ratio	1000
Neurons per Wafer	$1.5\ 10^5$
Power of single neuron/synapse system	1 – 50 µW
Power of single NPU incl. digital overhead	about 100 W

Table 2 : Parameter list of the proposed FACETS facility

10^4 faster than real time ...

Simulator – VLSI – Cortex

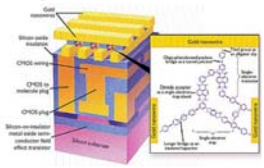
FACETS Stage I chip



	FACETS wafer (analog VLSI)	Human cortex
|N|	10^4	10^{10}
|W|	10^7	10^{14}
Synapse update time (sec)	10^{-7}	10^{-3}
Joule/synapse/update	10^{-15}	10^{-16}
Area/synapse (m^2)	10^{-10}	10^{-15}

Alternative HW implementations

- CMOL
  - Hybrid CMOS and molekylar circuits
  - ”nanoprinting”
  - stochastic computing
- Cell culture on chip
  - Implants!
- Quantum computer
- ...



Neuron culture on a chip

Future perspective

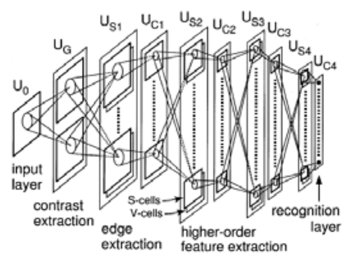
- Supercomputing power sufficient for real-time human brain simulation in 15-20 yrs
- Progress in brain science will stay very rapid
  - Brain mysteries likely to be largely uncovered in this period
  - + Quantitative computational models will be enabling
- Brain-inspired hardware development will parallel
  - Custom designs will be needed
  - A complementary breed of computers may emerge
- Potential for advanced and reliable computing using stochastic, imperfect HW
  - Molecular scale computing substrates?

Applications of Artificial brains?

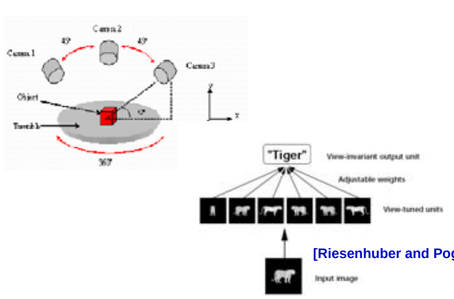
What about the singularity?

Neocognitronen

K Fukushima



Multipel-view based recognition



[Riesenhuber and Poggio, 2000]