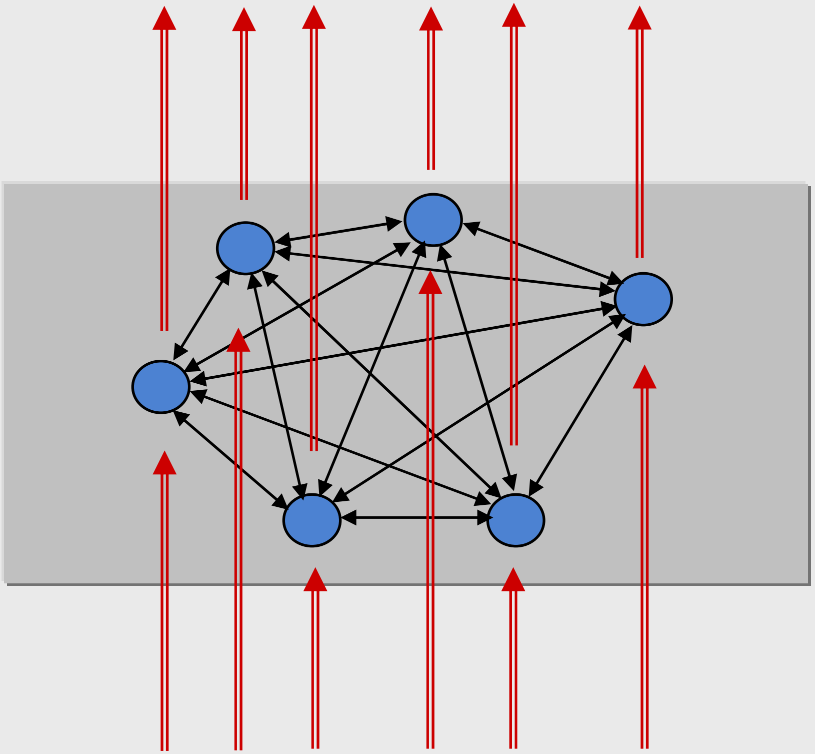
h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES

Computational Intelligence

Kapitel 3: Neuronale Netze

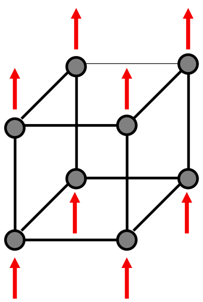
Teil C: Hopfield-Netze



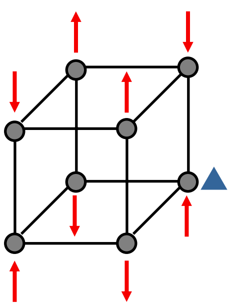
Dr. Norbert Waleschkowski

h_da Fachbereich Informatik
Sommersemester 2011
Master-Studiengang

The Spinglas Theory (1)



ferro magnetic domain



anti ferro magnetic domain

(Remark: In solid state theory solid bodies with an erratic arrangement of its atoms are called spin glasses.)

- The Hopfield network is motivated by phenomena from solid-state physics and statistical mechanics, resp. - the so-called spin glasses.
- The spin glass model describes the magnetic behaviour of solid bodies which is caused by the interaction of its atoms.
- The ferro magnetism – a crystal property – is based upon such interactions. The most salient feature of such ferro magnetic stuff is that even a very weak outer field force causes a very strong magnetisation.
- This impact is based upon the existence of magnetic and unmagnetic atoms. Each atom is associated with a so-called Ising-spin S_i, which indicates the magnetic orientation of the atom:

$$S_i = +1 \text{ or } S_i = -1$$

- There is a mutual interference between the atoms which leads to domains within the solid state bodies in which neighbouring spins are arranged in a parallel (ferro magnetic domain) or anti parallel (anti ferro magnetic domain) manner.
- Within spin glasses ferro magnetic and anti ferro magnetic domains are mixed up. If perturbations occur single spins will be adjusted to its correct orientation due to the forces of its neighbouring atoms.

The Spinglas Theory (2)

- Amount and direction of the forces between two atoms i and j depend in a complex manner on the configuration of all atoms, including the non-magnetic ones. These forces can be represented by coupling coefficients w_{ij}.

- For the coupling coefficients the following constraints apply:

  (i) $w_{ij} = w_{ji}$ for all i,j

  (Symmetry property; S_i has an impact to S_j and vice versa.)

  (ii) $w_{ii} = 0$ for all i

  (No direct self coupling; through the behaviour of the complete system there is indirect and collateral feedback, resp.)

- The state of a spin S_i is described by the equation

$$S_i(t+1) = sign\left(\sum_{j,j \neq i} w_{ij} S_j(t) - \Theta_i \right)$$

where Θ_i is a local field of i.

sign: $\Re \rightarrow \{-1,+1\}$ is the sign function.

- The behaviour of a spin glass over time axis is called dynamics of the spin glass.

The Spinglas Theory (3)

- Under the assumptions discussed above one can prove that there is a term E with

  $$E(\bar{S}) = -1/2 \sum_i \sum_j w_{ij} S_i S_j + \sum_i S_i \Theta_i$$, where $\bar{S}$ represents the spin glass.

- For the sake of simplicity we choose $\Theta_i = 0$ so that the following formula remains:

  $$(*) \qquad E(\bar{S}) = -1/2 \sum_i \sum_j w_{ij} S_i S_j$$

- For the term E the following applies : E cannot increase over time!

- <u>Proof</u>: Each change of state of a spin S_i may be expressed by

  $$\Delta S_i = 2\, sign\, (\sum_j w_{ij} S_j) \in \{-2,+2\}$$

  Because $w_{ij} = w_{ji}$ and inserting into (*)

  $$\Delta E = -\tfrac{1}{2} \Delta S_i \sum_j w_{ij} S_j$$

  $$= -sign(\sum_j w_{ij} S_j) \sum_j w_{ij} S_j$$

  $$= -\left| \sum_j w_{ij} S_j \right|$$

  ≤ 0 applies.

The Spinglas Theory (4)

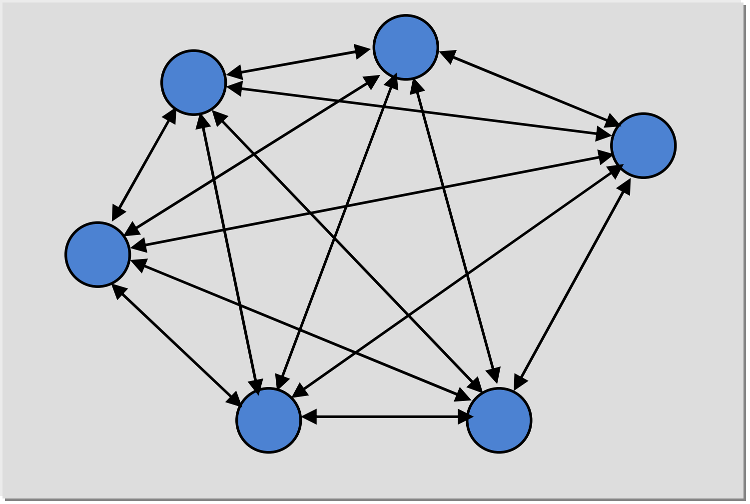
- E decreases as long as a fixed point, i.e. a stable state has been reached.
- E is called Energy resp. Hamilton function. E is a measure for the energy stored within the system. The system tries to minimize its energy.
- One can think of E as potential mountains over the state space of all possible binary vectors s. Each stable state of the system is associated with a local minimum of the function E.
- A local minimum is also called attractor state. At the beginning the system resides in an attractor valley and is attracted downwards into the bottom of the valley.

Hopfield, John: Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences, USA, Vol. 79, 1982, pp. 2554 – 2558

Hopfield, John: Neurons with graded response have collective compu-tational properties like those of two-state-neurons. Proceedings fo the National Academy of Sciences, USA, Vol. 81; 1984, pp. 3088 - 3092

The Discrete Hopfield Network (1)

- The American physicist John Hopfield proved in 1982, that a certain class of back coupled network has a strong analogy to spin glasses.
- A Hopfield network consists of n binary elements which are completely interconnected. The elements have only the values 0 and 1 as valid output values and a threshold value Θ_i . The states of the network are positioned at the vertices of the n-dimensional hyper cube $[0,1]^n$.



- A connection in a Hopfield network is always bi-directional. I.e. if N_i is connected to N_j then N_j is also connected to N_i.

The Discrete Hopfield Network (2)

- For the weights the following constraints apply

  $w_{ij} = w_{ji}$ (symmetry property)

  $w_{ii} = 0$ (no direct self back coupling).

- A Hopfield network with n neurons has $O(n^2)$ connections.

- The input vector is the total activity vector of the network at the beginning of the processing. Then the system is left to its own dynamics. The output is the total activity of the network at the end of the process if the network has reached some stable state.

- The state of a neuron is equivalent to the Ising-spin.

  $S_i = -1 \Leftrightarrow a_i = 0$

  $S_i = +1 \Leftrightarrow a_i = 1$

- The net input of a neuron N_i is $net_i = \sum_j w_{ij} o_j$

- For the whole networks applies: $a_i(t+1) = \begin{cases} 1 & \text{if } net_i(t) > \Theta_i \\ 0 & \text{if } net_i(t) < \Theta_i \\ a_i(t) & \text{if } net_i(t) = \Theta_i \end{cases}$

- A connection in a Hopfield network is always bi-directional.
  I.e. if N_i is connected to N_j then N_j is also connected to N_i.

The Discrete Hopfield Network (3)

- The threshold Θ_i corresponds to a local field which tries to direct the spin into a certain direction. For the sake of simplicity we choose here $\Theta_i = 0$.
- The output function o is the identity: $o_i = id\ a_i = a_i$.
- In analogy to spin glasses there exists an energy- or a Hamilton function E, resp.:

$$E = -\frac{1}{2}\sum_i \sum_j w_{ij} o_i o_j$$

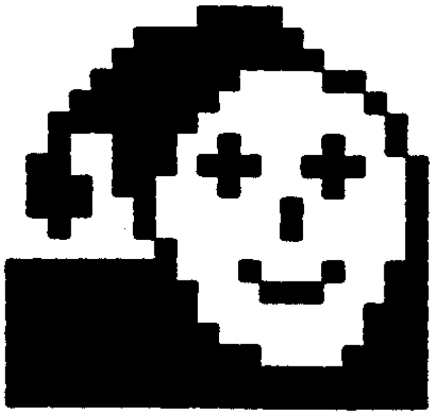
- A learned pattern should correspond to a local minimum of the Hamilton function. But how to determine the weights in order to find the appropriate minima of the Hamilton function? By choosing the weights properly one is able to form the shape of the potential mountains in the desired manner and to place the minima appropriately.
- If the weights have been chosen in such a way that the bottoms of the valley are associated to the input training patterns then the dynamics of the Hopfield network makes it happen that each input pattern is attracted into the bottom of the valley is has fallen into.
- If the input patterns are sufficiently uncorrelated (i.e. if the similarities between the input vectors are low) one can calculate the weights as follows:

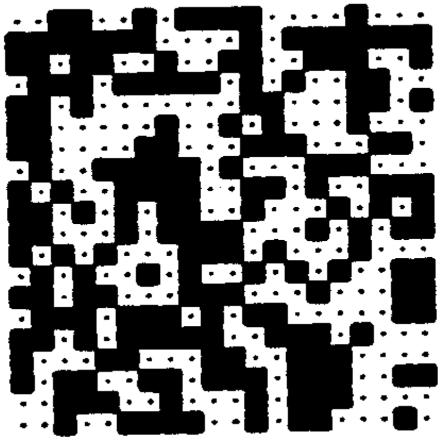
The Discrete Hopfield Network (4)

- Let $\mathbf{x}^l,\ l \in \{1,\ldots,m\}$ training patterns acc. to $\mathbf{x}^l = (x_1^l, x_2^l, x_3^l, \ldots, x_n^l)$.
- Calculate weights as follows:

$$w_{ij} = \frac{1}{n}\sum_{l=1}^{m} x_i^l x_j^l,\quad n = \text{number of neurons}$$

- After this “learning phase” the network can go to work.
- Example: A NN shall recognize and restore noisy and incomplete pictures. 20 pictures of the following kind are given.



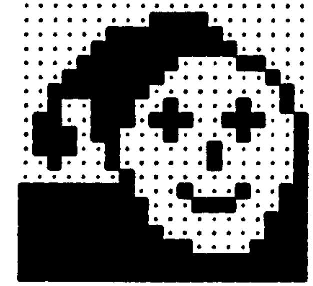


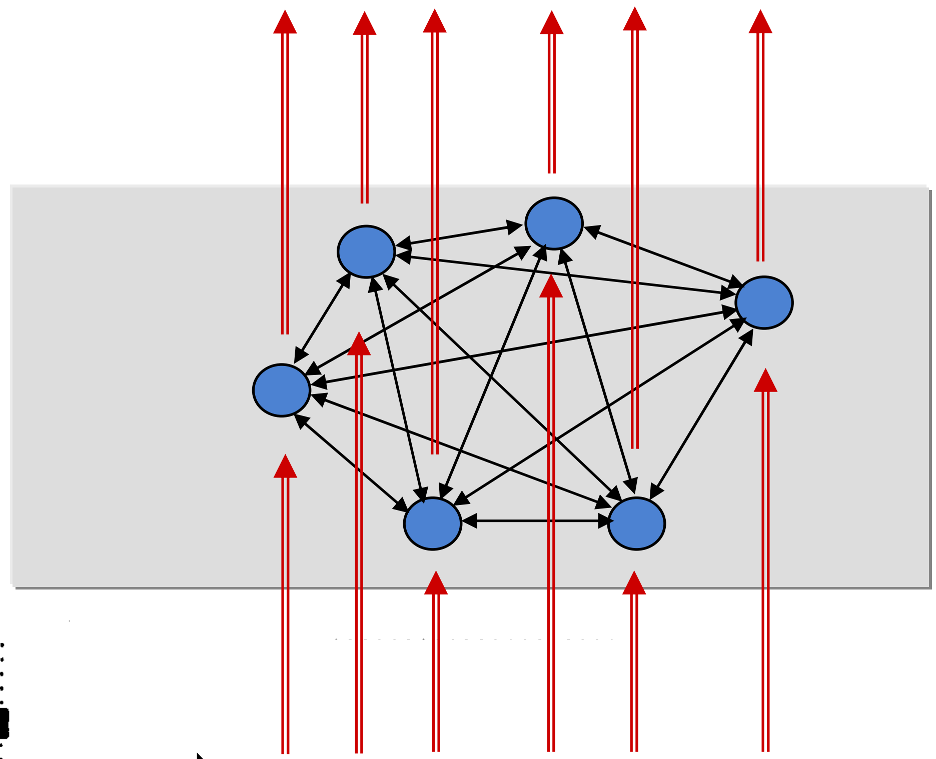
- These pictures consist of 20 x 20 = 400 black and white pixels. Each picture has 200 black and 200 white pixels. Picture 1 shows “a little man with a cap”, picture 2 is one of the 19 other random patterns.

The Discrete Hopfield Network (5)

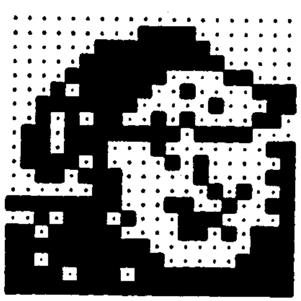
The network should recognize noisy patterns and identify the most similar one. Therefore a picture is transformed into a string of „0“ and „1“ with length 400. The string then is presen-ted to the network.

(0000....00001......00001.....11) **Dekodierung**





The network should recognize and complete a noisy pattern like the one in the lower left corner.



Kodierung (0000....10111......10011.....11)

The Discrete Hopfield Network (6)

- To solve this problem we choose a network of 400 neurons. The valid output values are 0 and 1 and represent a white or a black pixel. A picture is represented as a vector of length 400. (Remind that by such a transformation the neighbourhood information per pixel has been lost.)

- The 20 patterns have been chosen randomly. Hence, one can assume that they are sufficiently uncorrelated. Therefore the weights are calculated according to

$$w_{ij} = \frac{1}{400} \sum_{l=1}^{20} x_i^l x_j^l$$

- This leads to the following energy function, where = 0 for all $i \in \{1,\ldots,400\}$, because black and white pixels are distributed equally.

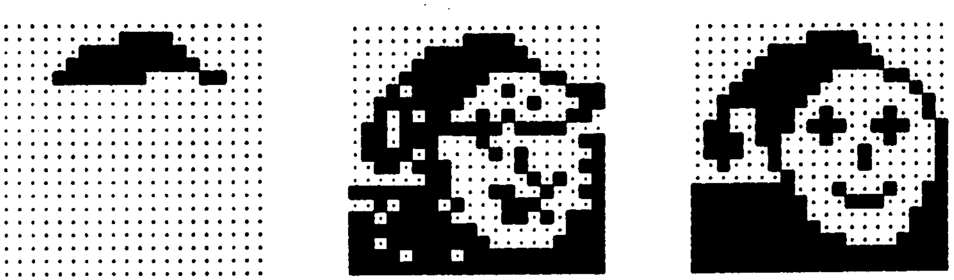
$$E = -\frac{1}{2} \sum_i \sum_j w_{ij} o_i o_j$$

- The next figures illustrate the reaction of the network for different input patterns.

The Discrete Hopfield Network (7)

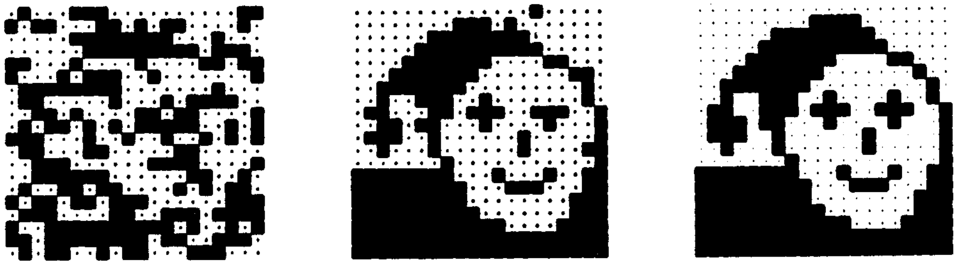
- Recognition process for input patterns that have been noised increasingly.

Case 1:
Only 25% of picture 1 are presented.



Completion of a fragmentic input pattern. Only 25% of the input pattern are presented (leftmost picture). Rem.: (Actually, appr. 62% are presented, because 50% of the remaining 75% are truely white.) After one cycle the original pattern is recognizable, one cycle later is has been completed correctly

Case 2:
Each pixel has been noised with a probability of 0,3.

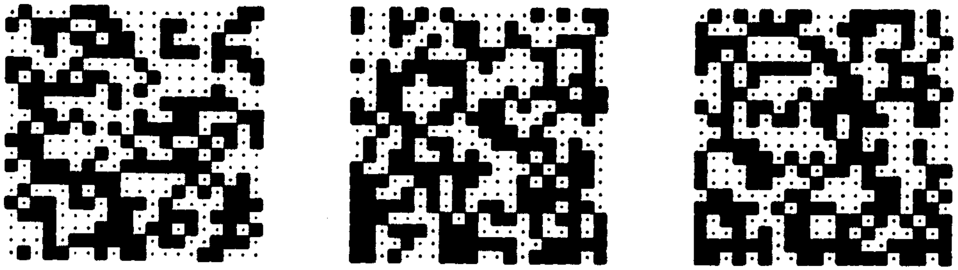


Restauration of a noised pattern. Input is again „the little man“. Each pixel was noised with a probability of 0.3. After a few cycles the original pattern has been completed correctly.

The Discrete Hopfield Network (8)

- Recognition process for input patterns that have been noised increasingly.

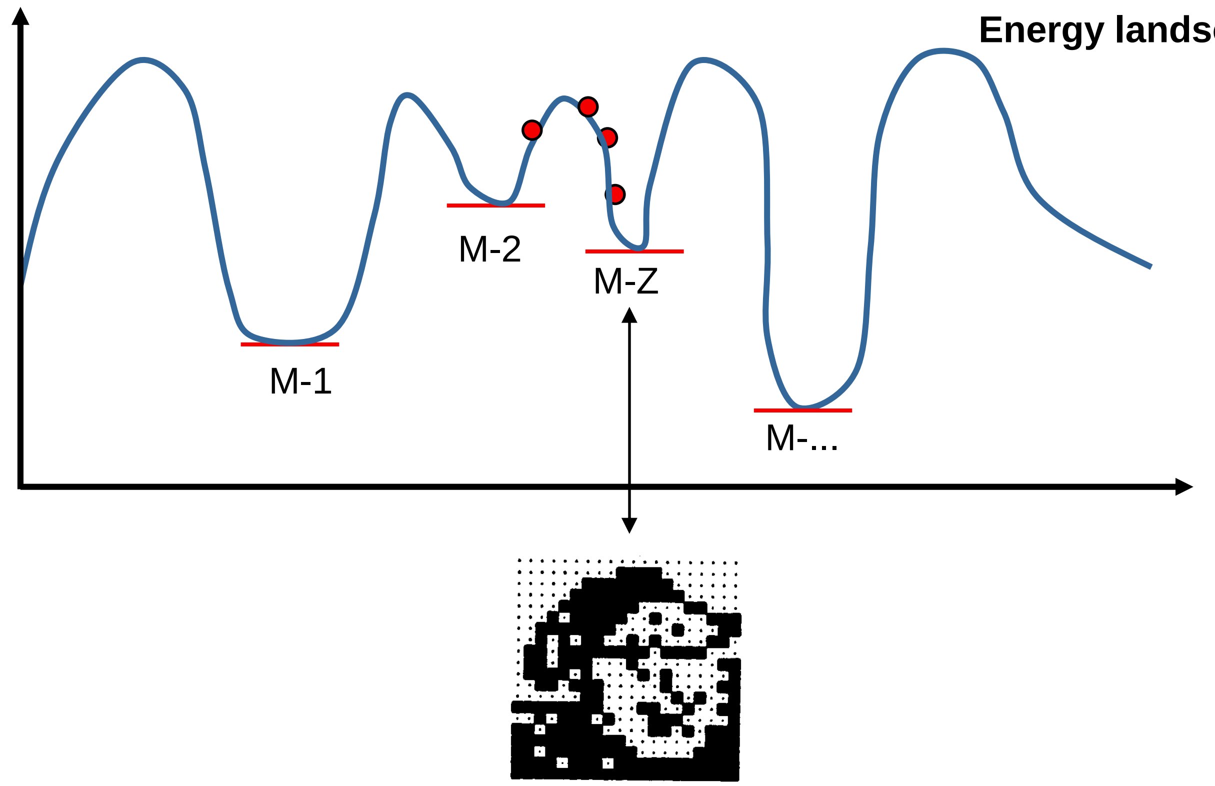
<u>Case 3:</u>
Each pixel was noised with a probability of 0.4.



Input is again „the little man“. Now each pixel was noised with a probability of 0.4. Now the network can‘t complete the original pattern and converges against one of the other 19 random patterns.

The Energy of a Hopfield Network

- As physical systems do Hopfield networks have an „inner“ energy.
- The energy landscape is formed by the weights.
- The dynamics of a system leads to a minimization of this energy until a local energy minimum (attractor valley) has been reached.



The Discrete Hopfield Network (9)

- Processing:
- The processing of a NN can be carried out following different criteria:
- Presentation of a pattern , then
  - (a) 1. Select a neuron randomly
    2. Determine state of the neuron
    3. Proceed with 1.
  - (b) Process per cycle all neurons in a sequence
  - (c) Parallel processing of all neurons (can also be done in a sequence)
- The result is available
  - after cancelling the processing of a certain number of steps
  - after a stable state has been reached
- The result is the whole activity pattern of the network after the processing has been finished.

The Discrete Hopfield Network (10)

- Spurious minima:
- Very likely it can happen that the energy function has minima which do not correspond to a learned pattern. Such minima are called spurious minima. If a very noisy pattern is presented then the network can reach such a state.
- The more similar input patterns are, i.e. the more correlated these patterns are, the more spurious minima of the energy function are generated by the calculation of the weights.
- A measure of the correlation of the patterns is the Hamming distance. Consider two binary vectors $\boldsymbol{x}_l$ and $\boldsymbol{x}_k$. The Hamming distance then is defined according to

$$h = \sum_{i=1}^{n} \left| x_i^l - x_i^k \right|$$

- i.e. in how many bits are the vectors different?
  - *(1)* $h = 0$ → vectors are identical
  - *(2)* $h = n$ → vectors are totally different.
- Properties of the Hamming distance:
  - *(1)* $h \leq n/5$ → minima come nearer to each other; instability increases
  - *(2)* $h \leq n/10$ → minima are coalesced.

The Discrete Hopfield Network (11)

- How strong can be the degree of noise of a pattern to be still recognized correctly?
- $s = h/n$ is called signal-noise-proportion. If s decreases then the probability for the correct recognition of the input pattern increases and vice versa.
- Spurious minima may become “unlearned” to the attractor valley may become smaller
  - (1) by presenting “white noise”
  - (2) by correction of all weights by $\Delta w_{ij} = -k(x_i^w x_j^w)$ $(0<k<1)$ where $\mathbf{x}_w$ is the spurious minimum.
- <u>Capacity of Hopfield networks:</u>

  The number of training patterns should not be too large, otherwise the energy landscape may become too complex and may have got too many minima. If cap = m/n denotes the capacity, where m is the number of learned patters and n is the number of neurons then the critical transition is appr. $Cap \sim 0{,}15$.

Hopfield, John; Tank, D.W.: „Neural“ computation of decisions in optimization theory. Biological Cybernetics 52, 1985, pp. 141 – 152

Hopfield Networks and the Travelling Salesman Problem (1) (The Algorithm of Hopfield and Tank)

- TSP (Travelling Salesman Problem): A salesman shall visit n cities using the shortest tour. He is not allowed to visit one city twice or more. It is not allowed to pass a city. How to find the shortest tour?
- The TSP shall be solved by using a Hopfield network.
- While the energy E is minimized by the system dynamics, the systems passes a route within the state space.
- By correlating the cost function of the optimization problem with the energy function, minimizing the energy is equivalent to minimizing the costs.
- This was the underlying idea of Hopfield and Tank (1985) in order to attack the TSP using the Hopfield model. Assumption is a symmetric cost matrix to apply the algorithm to the TSP.
-

	A	B	C	D	E
A	0	d_{AB}	d_{AC}	d_{AD}	d_{AE}
B	d_{BA}	0	d_{BC}	d_{BD}	d_{BE}
C	d_{CA}	d_{CB}	0	d_{CD}	d_{CE}
D	d_{DA}	d_{DB}	d_{DC}	0	d_{DE}
E	d_{EA}	d_{EB}	d_{EC}	d_{ED}	0

The distance matrix provides the distance between any pair of 2 cities X and Y. The distance matrix is symmetric, i.e. $d_{XY} = d_{YX}$.

Hopfield Networks and the Travelling Salesman Problem (2)

- The assignment matrix below provides information about the position of a city within a tour.

	1.	2.	3.	4.	5.
A	0	0	1	0	0
B	1	0	0	0	0
C	0	0	0	0	1
D	0	1	0	0	0
E	0	0	0	1	0

- The first row vector means that city A is visited as third city within the tour. The complete tour then is *B-D-A-E-C*.
- An NN then needs for a n-city-TSP n^2 neurons.
- Before we determine the weight matrix, we will consider the energy function and try to pose appropriate demands. Later on we will derive the weights.
- There is a need for an energy function that meets the following demands:

Hopfield Networks and the Travelling Salesman Problem (3)

- Energy minima must prefer states that meet the following criteria:
  (1) A city occurs only once within a tour, i.e. there is only one active neuron in a row.
  (2) At a time only one city may be visited, i.e. there is only one active neuron in a column.
  (3) Within a regular tour all cities must be visited, i.e. there must be (at least) n active neurons.
  (4) The shorter the length of a tour the better.
- Furthermore, we will declare the following convention:
- The output value of a city is denoted by V_{Xi} where X denotes a city and i represents the position within a tour:

$V_{Xi} = 1$ if city X was visited as i-th city within the tour

otherwise $V_{Xi} = 0$ (if X is not visited as i-th city)

Hopfield Networks and the Travelling Salesman Problem (4)

- As a result we get the following energy function, where A, B, C and D > 0.

$$
\begin{aligned}
E = {} & \tfrac{A}{2} \sum_{X} \sum_{i} \sum_{\substack{j \\ j \neq i}} V_{Xi} V_{Xj} \\
& + \tfrac{B}{2} \sum_{i} \sum_{X} \sum_{\substack{Y \\ Y \neq X}} V_{Xi} V_{Yi} \\
& + \tfrac{C}{2} \left(\sum_{X} \sum_{i} V_{Xi} - n \right)^2 \\
& + \tfrac{D}{2} \sum_{X} \sum_{\substack{Y \\ X \neq Y}} \sum_{i} d_{XY} V_{Xi} \left(V_{Y,i+1} + V_{Y,i-1} \right)
\end{aligned}
$$

- Let's consider the 4th term. If $i = 1$ then $V_{Y,i-1} = V_{y0}$ and if $i = n$ then $V_{Y,i+1} = V_{Y,n+1}$. Therefore we define $V_{y0} = V_{yn}$ and $V_{Y,n+1} = V_{Y1}$.
- The i-th term of the energy function corresponds with criteria (i).

Hopfield Networks and the Travelling Salesman Problem (5)

- Analysis of the term $d_{XY}V_{Xi}(V_{Y,i+1}+V_{Y,i-1})$
- There is exactly one “i” with $d_{XY}V_{Xi}\neq 0$, where X and Y are fixed. This “i” is denoted by iX. Then the following applies

$$\sum_i d_{XY}V_{Xi}=d_{XY}V_{X,ix}=d_{XY}$$

- The distance d_{XY} should only be added if the tour goes from X to Y or vice versa. I.e., either $V_{Y,ix-1}\neq 0$ or $V_{Y,ix+1}\neq 0$ applies. Thus $(V_{Y,ix-1}+V_{Y,ix+1})\neq 0$
- Then $d_{XY}V_{Xi}(V_{Y,ix-1}+V_{Y,ix+1})\neq 0$
- As an overall result

$$d_{XY}V_{Xix}\left(V_{Y,ix+1}+V_{Y,ix-1}\right)$$
$$=\sum_i d_{XY}V_{Xi}\left(V_{Y,ix+1}+V_{Y,ix-1}\right)$$ applies.

- Summation for a fixed X over $Y,\ Y\neq X$, followed by summation over X leads to the 4th term.
- The coefficients A,B, C and D have to be chosen properly. If, for instance, one of these coefficients is significantly greater than the others, then there is a tendency of the network to prefer minimizing the corresponding term.
- Let’s consider the fourth term. If i = 1 then $V_{Y,i-1} = V_{y0}$ and if i = n then $V_{Y,i+1} = V_{Y,n+1}$. Therefore we define $V_{y0} = V_{yn}$ and $V_{Y,n+1} = V_{Y1}$.
- The i-th term of the energy function corresponds with criteria (i).

Hopfield Networks and the Travelling Salesman Problem (6)

Ullman,S.: Relaxation and Constrained Optimization by Local Processes, in: Computer Graphics and Image Processing 10, 1979, pp. 115-125

- As a consequence the network might find “solutions” where one or more criteria may be hurt, i.e. these solutions are irregular tours. That’s why the coefficients should be chosen properly in order to minimize each term, thus avoiding irregular tours.
- To determine the weights, E has to be transformed into a normalized form. The terms are transformed in such a way to combine quadratic and linear expressions. This can be achieved by means of the Lagrangian coefficients.
- Setting this new expression equal to the normalized form

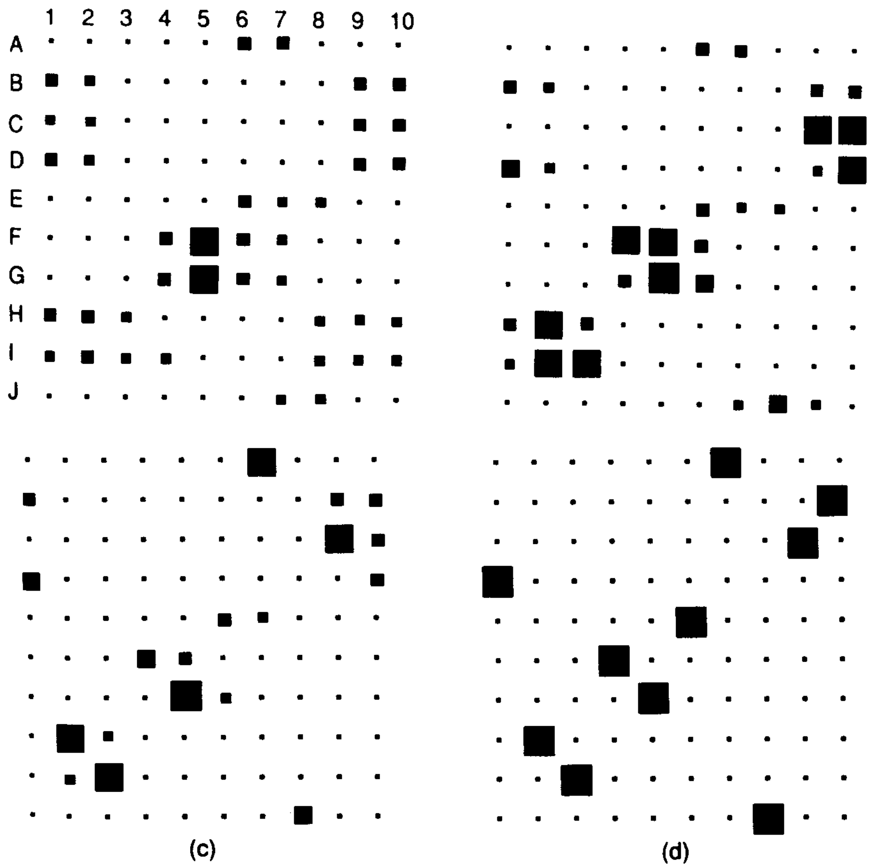
$$E = \frac{1}{2}\sum_X \sum_Y \sum_i \sum_j w_{Xi,Yj} o_i o_j - \sum_X \sum_i \Theta_i o_i$$

- leads to the following formula for determining the weights

$$\begin{aligned} w_{Xi,Yj} &= -A\delta_{XY}(1-\delta_{ij}) \\ &\quad -B\delta_{ij}(1-\delta_{XY}) \\ &\quad -C \\ &\quad -Dd_{XY}(\delta_{j,i+1}+\delta_{j,i-1}) \end{aligned}$$

Hopfield Networks and the Travelling Salesman Problem (7)

- Hopfield and Tank carried out tests with 10 and 30 cities. For 10 cities there exist 181.440 regular tours.
- Result of the tests: 16 of 20 results have been valid tours. 8 of these 16 solutions belong to the shortest tours near to the global optimum.
- Example for a 10-city tour:



Hopfield Networks and the Travelling Salesman Problem (8)

- When Hopfield and Tank attacked TSPs with 30 cities or more many problems occurred. The quality of the solutions could not compete with results achieved by other approaches like e.g. the Lin-Kernighan-algorithm. Hopfield and Tank argued that an insufficient determination of starting parameters and starting states of the network were responsible for the insufficient quality.
- One reason for the insufficient results is that Hopfield networks always converge against the next local minimum. They are unable to escape from an attractor valley. For energy functions with a great many of local minima, i.e., for bigger values of n, this normally leads to problems.
- Kohonen argues that Hopfield networks are being far apart from natural systems. They look engineered to a high degree. Many important para-meters have to be chosen by the developer and not by the system itself. Furthermore, the Hopfield approach is not robust in the sense, that for each new problem one has to find a new specific energy function. And the neurons don't „look" biological (e.g. symmetry of the weights).
- Another disadvantage of Hopfield networks is their low capacity. For n cities n^2 neurons and $O(n^4)$ connections are needed. These properties cannot be improved.
- Overall improvements can only be expected if there is a chance to escape from a local minimum.