Physics of Neural Networks

W. Kinzel

Institute for Theoretical Physics
Justus Liebig University, Giessen

A new field of research — the physics of neural networks — has grown up within theoretical physics over the last few years. Although still in its infancy it has produced some impressive results, especially a basic understanding of the cooperative behaviour of neural networks which is useful in neurobiology as well as in computer science.

The material basis of our thinking, intelligence and creativity are 10^{13}-10^{14} nerve cells *(neurons)* which in our brain are densely packed into a grey substance weighing about 1.5 kg. Each neuron has — like the root of a tree — highly branched dendrites that collect information from about 10000 other nerve cells. This huge network, which on a microscopic scale looks rather homogeneous and disordered, is able to process information. The properties of the brain obviously arise from the cooperation of a very large number of neurons. But is it possible to understand complex processing of information on the basis of nerve cells and their interactions?

Physicists are, of course, not the first and only people to think about neural networks. Much is known about single cells and their interactions. A neuron emits electrical pulses, and the mechanism by which the pulses move along nerve fibres *(axons)* has been explained. Complex biochemical processes in the contact regions *(synapses)* between two neurons transmit the pulses to the connected neurons. Each neuron integrates all of the transmitted pulses to give a local electrical potential. The neuron fires several electrical pulses on its own if this potential exceeds some threshold value; otherwise its firing rate is low.

Hence, many details of the **single** element and its interactions are known. But the cooperation of many neurons is far from understood. Is there a single cell which only fires if you see your grandmother? How is the memory of your grandmother stored in your neural network? Is it in a single synapse or distributed throughout all parts of your brain? What happens when you meet a person you know, when, at the same time, all associations with this person seem to be ready for recall? How does a neural network recognize the face of a known person: how is the necessary information stored and learnt? All of these questions cannot yet be answered in terms of the cooperative behaviour of many neurons.

Brains and Computers

We know of course how to build machines from electrical switches to process information, namely computers. But even a modern computer is very different from a neural network in its architecture as well as in its functionning. A computer has a central processing unit which follows, step by step, program commands written by a human being beforehand. The output is stored in what can be considered as numbered boxes.

The brain, on the other hand, is a network. All of its units are simultaneously sending and reacting to signals. Data are distributed throughout the network. So whereas even a child recognizes a face in a fraction of a second without any effort, a modern supercomputer in spite of being able to compare, add or sort a huge amount of data extremely rapidly, cannot perform the same task with hours of computation time.

Neurocomputers

Although the higher functions of the brain are not yet understood on the basis of the dynamics of a neural network, one may examine information processing in networks of simple switching elements. In fact, several such networks had already been analysed and applied to simple tasks in the 1960's [1]. These networks contain sets of input and output units, and the network can adapt to input/output examples which are presented. This adaption occurs by synaptic plasticity, *i.e.* the synaptic weights between input and output ''neurons'' adjust to a presented example *via* simple mechanisms. A neurocomputer therefore ''learns'' an input/ouput map from examples. Two learning rules are well known: Rosenblatt's ''Perceptron'' and Widrow's linear rule ''Adaline'' [1].

In 1969, an influential book by Minsky and Papert cooled down the excitement surrounding the application of artificial neural networks to complex computational tasks [2]. A mathematical analysis showed that a **single** layer of synapses between input and output, without any additional (hidden) neurons, could only perform linear separable functions — a very limited class of mappings.

However, the initial excitement has been revived these last few years due to the fact that multilayer networks can be studied nowadays by simulating them on supercomputers; and that it should be possible to build networks on modern electronic hardware. **Multilayer** networks are the key because they can realize any input/output map.

Associative Memory

J.J. Hopfield pointed out that this type of a simple network of formal neurons can act as an associative memory capable of storing many patterns [3]. A pattern is a set of bits ξ_i^ν coded as $\xi_i^\nu = +1$ or $\xi_i^\nu = -1$ where ν labels one of p stored patterns and i is the index of the neuron. Using linear algebra, the synaptic weights are given by [1]:

$$J_{ij} = (1/N) \sum_\nu \xi_i^\nu \xi_j^\nu$$

For a set of p random patterns $\{\xi_i^\nu\}$, one may study the dynamics of N neurons

---

Professor Wolfgang Kinzel has been four years with the Institute for Theoretical Physics, Justus Liebig University, D-6300 Giessen. A theoretical solid state physicist, he graduated with a Ph.D. from the University of Köln in 1978. He then worked as a postdoctoral fellow in the Physics Department, University of Washington, Seattle, USA before spending nine years in the Solid State Institute at KFA Jülich. Interested in the statistical mechanics of disordered systems, he heads a research group supported by both government and industry studying neural network theory.

S_i in a network with synapses J_{ij}. Fig. 1 shows what happens. If the initial state $S_i(t=0)$ has some overlap with one of the patterns ν, say with the pattern "A", then the neurons relax to the complete state "A". Initially incomplete information is restored completely. This is a cooperative effect: the system of N neurons moves downhill in the energy landscape given by the couplings J_{ij}. The synaptic weights that were selected obviously created a landscape with a broad valley for each of the p patterns.

The properties of the stationary states for the Hopfield model have been calculated by Amit, Gutfreund and Sompolinsky using spin glass theory [5]. The system is described by a free energy which is averaged over the set of random patterns. By employing a mathematical trick (the replica method [3, 4]), the infinitely large system ($N \to \infty$) is described in terms of order parameters that are found using implicit equations (saddle points).

The final solution shows that the system can store infinitely many patterns with the same synapses. For

$$P = \alpha N$$

one obtains a coefficient of maximum storage capacity α (for $N \to \infty$) of $\alpha = \alpha_c = 0.14$. For higher values of α, the associative memory suddenly disappears completely as with a first order phase transition in a magnetic model.

Many details and variants of the Hopfield model have been studied using exact solutions (different choices of synapses; correlated patterns; the effects of external fields, thermal noise, and static synaptic noise; dilution of synapses; multistate neurons; forgetting, *etc.*). In some special cases, *e.g.* for the extreme dilution model for layered feedforward networks [6], the dynamics could be solved exactly.

Learning

An important property of a neural network is its ability to learn specific tasks. The psychologist D.O. Hebb suggested in 1949 that synaptic plasticity was the essential learning mechanism: each synaptic weight adjusts itself according to the activities of the two adjacent neurons which it connects.

Neurocomputers mainly use a simple algorithm called "back propagation" to adjust their synapses to presented examples: the least-squares deviation of the desired output from the actual output is minimized by exploiting a gradient descent.

Physicists have recently described some properties of learning algorithms using the exact analytical methods of statistical mechanics [7]. Although up to now only a single layer of synapses can be handled, much insight into the dynamics of synapses has been obtained.

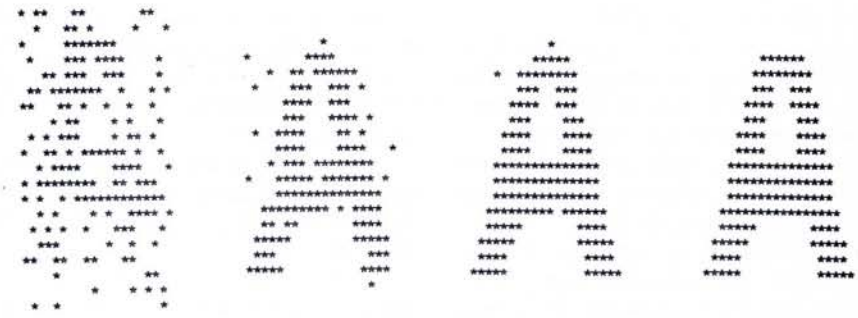

Fig. 1 — A network of 20 × 20 neurons which has learnt 30 patterns. One of them is the pattern "A" and the remaining 29 are random. The dots represent active neurons (passive ones and the synapses are not shown). After four only adaptions per neuron, a noisy initial state on the left relaxes to the corresponding learnt activation pattern on the right.

It was in 1987 that the late E. Gardner first showed how to apply statistical mechanics to synaptic plasticity [8]. She succeeded in calculating the volume in n-dimensional synaptic space of all possible synaptic weights $\{J_{ij}\}$ which stores a set of $p = \alpha N$ patterns $\{\xi_i^\nu\}$. As one increases the number of stored patterns, this volume shrinks to zero and the corresponding associative memory reaches its maximum storage capacity α_c. Thus, for a larger number of stored patterns the network does not function any longer.

For random patterns, Gardner found $\alpha_c = 2$, *i.e.* a single network of N neurons can store $p = 2N$ patterns (compared to $\alpha_c = 0.14$ for the Hopfield model). The maximum capacity α_c increases further for correlated patterns.

A measure of the size of the basin of attraction of a stored pattern $\{\xi_i^\nu\}$ is its stability

$$\Delta_{i\nu} = \xi_i^\nu \sum_j J_{ij} \xi_j^\nu \,/\, (\sum_j J_{ij}^2)^{1/2}$$

Gardner has calculated the maximum possible stability Δ as a function of the storage capacity α. As α approaches α_c then Δ decreases to zero.

The phase space analysis shows that synaptic matrices exist for $\alpha < \alpha_c$. But there remains the problem of how to construct algorithms with which the network can automatically find these synaptic weights. As already mentioned, there exist two algorithms which can be applied to the attractor networks studied by physicists: the Perceptron and the Adaline. For the former mechanism there even exists a convergence theorem: if a matrix J_{ij} exists which stores a set of p patterns then the algorithm is guaranteed to find one of these matrices. Krauth and Mezard constructed a version of the Perceptron which is even able to find the matrix of optimal stability [9].

Simulations of networks have in fact recently demonstrated several interesting applications. So-called "neurocomputers" have learnt to read written text aloud, to make a medical diagnosis, to play backgammon, to recognize symmetries in patterns, to balance a stick, to detect defects from listening to the noise of motors, to move a truck and trailer backwards, *etc.* The remarkable feature in each case is that a network does not need a program but learns a task by itself using examples.

Statistical Mechanics

Physicists' interest in neural networks stems largely from the analogies between such networks and models used in statistical mechanics. These analogies were first pointed out by Little in 1974, and again in 1982 when J.J. Hopfield explained the properties of a network which acts as an associative memory in terms of its similarity to relaxation processes in spin glasses [3].

Spin glasses are disordered magnetic materials such as certain gold-iron alloys [4]. At low temperatures, the magnetic moments freeze into a complex magnetic structure and, owing to competing interactions, the system adopts one of a large number of possible disordered structures. They have lately been considered as paradigms for many different types of systems which have a huge number of possible stable states.

Models for spin glasses have been studied intensively since 1975 when S.F. Edwards and P.W. Anderson introduced and solved a simple Ising model with random, competing interactions. The first solution invoked a mean-field approximation, but shortly afterwards D. Sherrington and S. Kirkpatrick (SK) presented a solution for the corresponding model for couplings over an infinite range. It took a further five years before

G. Parisi found a solution which is believed to be exact. The SK model exhibits a phase transition to a disordered, frozen structure at low temperatures. Many different low temperature phases are possible, and there exist complex, finely detailed structures for the various phases [3, 4].

The properties at equilibrium of the SK model are therefore well understood. However, dynamic behaviour, particularly relaxation far from equilibrium, still cannot be described analytically and is known only from computer simulations.

The theory of spin glasses has been applied to models of neural networks: the neurons are the spin variables, and the excitatory and inhibitory synapses are the ferromagnetic and antiferromagnetic interactions [3]. If one considers an associative memory as a memory which stores random patterns, then patterns are equivalent to disorder in the spin glass couplings.

Theoretical physics is contributing greatly to a general understanding of the cooperative behaviour of neural networks: using a limited number of essential mechanisms in simple models which can be solved exactly, infinitely large systems can be treated and the properties of typical systems calculated (*i.e.* an average over all possible patterns stored in an associative memory). Physicists have been able as a result to formulate generalised quantitative laws for networks.

The Mathematical Model

As long ago as 1943, McCullogh and Pitts reduced the complex electrical and biochemical mechanisms of a single neuron to a very simple mathematical form. A neuron can take two states described by $S_i = +1$ and $S_i = -1$:

$S_i = +1$, neuron i fires

$S_i = -1$, neuron i is quiet.

The synaptic contact from neuron j to neuron i is described by a real number J_{ij} which is positive for an excitatory and negative for an inhibitory synapse. Each neuron collects signals from many others, thus generating an electrical potential,

$$h_i = \sum_j J_{ij} S_j$$

If h_i is larger than some threshold $\theta = 0$, $S_i(t+1) = \text{sign}[h_i(t)]$ where t is the time.

For symmetric synapses, the $J_{ij} = J_{ji}$ dynamic relationship minimizes the function

$$H = -\sum_i S_i h_i = -\sum_{ij} J_{ij} S_i S_j$$

H therefore plays the role of an energy in the corresponding model for a magnetic system. If the synaptic weights J_{ij} are random then the model is exactly the SK model for a spin glass and the neural activity, $\{S_i\}$, relaxes into one of the many valleys of the complex energy landscape given by H.

If a neuron S_i aligns to its local potential h_i with a probability P (owing to some high frequency noise), and if P is given by

$$P = 1/[1+\exp(-h_i S_i/T)]$$

then the corresponding spin model relaxes to thermal equilibrium at some temperature T.

Learning Times

A problem of practical importance is the speed of learning algorithms. In many applications of neural networks the set of examples has to be presented in many thousands of iterations before the synaptic matrix has converged to one of the desired solutions. Again the methods of statistical mechanics gave exact solutions to this problem for the special case of a single layer. For both of the Adaline and the Perceptron of optimal stability, Opper has calculated the distribution of learning times exactly [10]. He finds a slowing down of the learning speed if the associative memory is loaded close to its maximal storage capacity α_c: the corresponding average learning time diverges as $(\alpha_c - \alpha)^{-2}$ if α approaches α_c from below.

Based on this mathematical insight into the structure and the dynamics of the Adaline and Perceptron rules, Anlauf and Biehl recently suggested, and investigated, a combination of these two rules, called the "Adatron" which converges very rapidly to the synaptic matrix of optimal stability [11]. The author hopes to apply this approach to multilayer networks.

Summary

While information processing in our brain still remains a mystery, the investigation of simple mathematical models, incorporating only a few essential features of a real network, shows that information processing can emerge as a cooperative effect owing to the interaction of many basic elements.

Analyses of neural networks have yielded quantitative results describing several striking properties:

1. A network comprising two state elements (neurons) which are totally connected by synapses *operates* quite differently from a modern computer. Needing neither a central processing unit nor a program, it operates automatically *via* the parallel, mutual interaction of all its elements. The restoration of a noisy pattern results from cooperation between many neurons.
2. The same holds for *learning*: using a simple mechanism, the synaptic weights slowly adjust to presented examples. But the network does not only learn the examples — it can also generalize to some extent.
3. Patterns are not *stored* in numbered locations as in a computer, but distributed over the synapses. Each synapse contains some information about all of the stored information and needs the interaction of a large number of competing couplings to store many patterns in a single synaptic matrix.
4. To *retrieve* stored information a network requires a partial, incomplete pattern as input. Its memory is therefore content addressable and associative, in contrast to a computer which needs the number of the corresponding location to retrieve data.
5. *Access* to a stored pattern is extremely rapid (as demonstrated by restoration of the noisy "A" pattern in four steps per neuron). Hardware realizations of networks would perform tasks of this type in microseconds or less.
6. A network is extremely *fault tolerant*. Even after the destruction of large fractions of neurons and synapses the system is still working, albeit with a larger error and less storage capacity.
7. Neurons work in *parallel*, but they do not have to switch synchronously. If they work with some defined probability (corresponding to thermal noise) the network's performance improves.
8. The predicted properties of the network are very *insensitive* to details of the model. For example, even if the synaptic weights are bounded or restricted to binary values, an associative memory changes properties gradually.

REFERENCES

[1] Kohonen T., *Selforganization and Associative Memory* (Springer Verlag, Berlin) 1988.

[2] Minsky M. and Papert S., *Perceptrons* (MIT Press, Cambridge, USA) 1988.

[3] Mezard M., Parisi G. and Virasoro M.A., *Spin Glass Theory and Beyond* (World Scientific, Singapore) 1987.

[4] Binder K. and Young A.P., *Rev. Mod. Phys.* **58** (1986) 801.

[5] Amit D.J., Gutfreund H. and Sompolinsky H., *Ann. Phys.* **173** (1987) 30.

[6] Domany E., Kinzel W. and Meir R., *J. Phys.* **A22** (1989) 2081.

[7] Kinzel W. and Opper M. in: *Physics of Neural Networks*, Eds L.V. Hemmen, E. Domany and K. Schulten (Springer Verlag, Berlin) 1989.

[8] Gardner E., *J. Phys.* **A21** (1988) 257.

[9] Krauth W. and Mezard M., *J. Phys.* **A20** (1987) L745.

[10] Opper M., *Phys. Rev.* **A30** (1988) 3824; *Europhys. Lett.* **8** (1989) 389.

[11] Anlauf J. and Biehl M., *Europhys. Lett.* **10** (1989) 687.