# Attractor dynamics in an electronic neural network

Paolo Del Giudice[1] and Stefano Fusi[2]

[1] Physics Laboratory, Ist. Superiore di Sanità and INFN, Sez. Sanità, Rome Italy
[2] INFN, Sez. Roma1, Dip. Fisica, Universitá "La Sapienza", Rome Italy

**Abstract.** LANN27 is an electronic device implementing in discrete electronics a 27 neurons, fully connected attractor neural network with stochastic learning. We summarize in this paper some key features emerged by extensive tests performed to elucidate the neuronal collective dynamics, the learning dynamics and the memory capacity of the LANN27 device.

## Introduction

This paper describes the process of attractors formation in an electronic recurrent neural network which implements neural dynamics against a background of plastic synapses, whose dynamics is driven in turn by the neural activities. The synaptic dynamics (learning) is stochastic and is studied under various "environmental" conditions.

Stochastic learning was originally proposed [3] as an effective strategy to overcome the limitations of the memory capacity attainable in networks with a finite set of stable synaptic efficacies. The latter constraint appears to be a very general and natural one either in material implementations of neural models or in plausible biological modelling. In fact, stochastic learning is a natural way to affect learning in recurrent networks of spiking neurons.

LANN27 is an electronic implementation of a fully connected attractor neural network of 27 neurons and 351 synapses with stochastic learning. The methodological approach is explained and motivated in [2] (where a smaller pilot device is described), together with a detailed description of the single elements (neurons and synapses) of the network.

Here we give a brief account of the analysis of the system's dynamics in terms of the evolution of its space of states as a result of the learning process.

The extensive tests performed to this end provide both insight into the implications of a stochastic learning scheme for real, finite size, noisy networks and methodological suggestions for the next generation networks of spiking neurons implemented in VLSI chips, which are now under development.

## The LANN27 network

Relevant features of LANN27 are:

*1.* neurons are implemented with high gain amplifiers

*2.* synapses have 3 stable states, preserved indefinitely by a stochastic "refresh" mechanism. On short time scales upon the arrival of a stimulus, the synapse acts as a capacitor, integrating a Hebbian-like source term induced by the stimulus. This analog charging of the synapse can in turn result in synaptic transitions to one of the stable states, with probabilities that depend on a noise source associated with each synapse, and on the strength and duration of the stimulus (see Fig. 1).

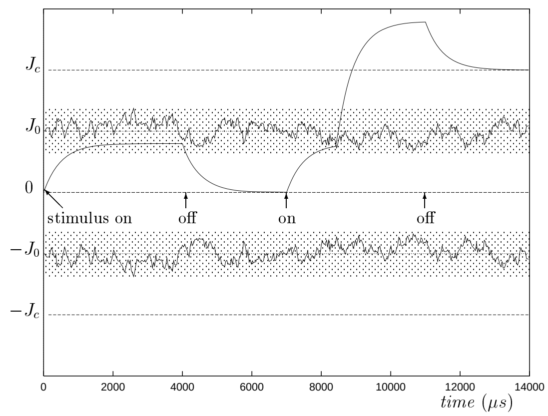


**Fig. 1.**
Example illustrating the stochastic learning mechanism. $(-J_c, 0, J_c)$ are the synaptic stable states; $(-J_0, J_0)$ are noisy thresholds fluctuating in the shaded strip. If the analog value of the synapse (solid plot), driven by the stimulus, does not reach the threshold, it decays to the previous stable state (0) upon removing the stimulus, while it is clipped to $J_c$ if the random threshold happens to go below it. An analogous mechanism governs the other transitions.

The transition probabilities associated with this stochastic discretization mechanism are made small enough to allow the network to maintain a significant memory span along the temporal flux of stimuli[4]. Indeed, in a network with a finite set of stable synaptic values, the probability that each synapse has to make a transition to one of these states, under the influence of an incoming stimulus, sets the average rate at which the information about past stimuli fades away as an effect of the new ones. Consequently the network exhibits a *palimpsest* property, and never undergoes the blackout catastrophe that occurs, for example, in the Hopfield or Willshaw models.

*3.* the network undergoes a double, unsupervised dynamics for neurons and synapses. Learning and retrieval are not logically separated; they are distinguished only by the time the stimulus persists and, consequently, by the chance it has to induce synaptic modifications. The neurons are the 'fast' dynamical variables, while the typical time scales for synaptic changes are much longer.

*4.* the Hebbian self-organization of the (symmetrical) synapses results in an attractor structure of the network's state space. A learnt stimulus (attractor) acts as the *prototype* of a class, the latter being the set of all input stimuli that lead the network to that attractor under the neural dynamics. All the stimuli belonging to the same class define the *basin of attraction* of their prototype. Learning is the process of slow synaptic modification that translates the statistics of the stream of input stimuli into the attractor structure.

*5.* the temporal and spatial statistics of the flux of stimuli is in principle unconstrained.

*6.* the electronic implementation is completely analog and asyncronous.

**Monitoring the state space**

In the tests we performed, the gain of the amplifiers implementing the neurons was chosen high enough to let us consider them effectively as binary $\pm 1$ Hopfield-like neurons, which allows us to define the state space as the 27-dimensional hypercube. The high dimensionality of the space of states makes it difficult to monitor the learning process sculpting the attractors and the basins around them.

We propose a graphical approach to capture and describe the development of attractors during the learning process. This description allows to project the 27-dimensional space of network state space onto a set of planes, in such a way that euclidean distances in the plane between each of all possible configurations, and those corresponding to class prototypes, be equal to the corresponding Hamming distances. The logical scheme of this graphical representation is given in Fig.2.

As we mentioned above, learning and retrieval regimes are only distinguished by the time the stimuli persist; in testing the behaviour of LANN27, we characterize the input streams of stimuli as controlled sequences of "learning" and/or "retrieval" presentations, meaning stimuli persisting for a long or short time, respectively.

During a learning session we perform at fixed intervals a retrieval phase: for each prototype, many patterns at increasing distance from it are presented to the network for a time short enough not to induce synaptic transitions. We sample this space by assigning a number of patterns, to be generated at random and presented for retrieval to the network, for each of the possible distances from 0 to 27. Following the removal of each pattern, the neural configuration to which the network relaxes is recorded. Next, for each pair of prototypes, we project the state space onto a plane, to expose the structure of the basins of attraction (see Fig.2). Two points are set on the plane, representing two of the prototypes (their euclidean distance on the plane being proportional to their Hamming distance). Then two sequences of 28 circles each are drawn, centered around each prototype. All the patterns whose representative point lie on a circle are at fixed Hamming distance from the prototype represented by the center of the circle. Each of the possible intersections of the circles centered around the two prototypes represents the set of all patterns at a fixed Hamming distance from the two prototypes. Then we make the graphical construction shown in Fig.2, in which each diamond shaped region corresponds to one of the intersections.

We associate each attractor with one of the fundamental colour components (red-R, green-G, blue-B). For each diamond (for each pair of distances from the two prototypes) we measure the number $n_1$ of times in which the input patterns with those particular distances from the prototypes make the network relax to the first prototype, and the numbers $n_2$ and $n_3$ for the other prototypes; Z is the number of stimuli for which the network did not converge to any attractor within a fixed tolerance. The color of the corresponding box with given distances from the two prototypes is calculated, in RGB components, as: $R = n_1/D$, $G = n_2/D$, $B = n_3/D$, where $D = n_1 + n_2 + n_3 + Z$. So, for example, the presence of a bright red diamond implies that the vast majority of the corresponding input patterns led the network to the 'red attractor', while a dim green diamond indicates that many times the network relaxed far from all the three attractors, into the green one in the other cases.

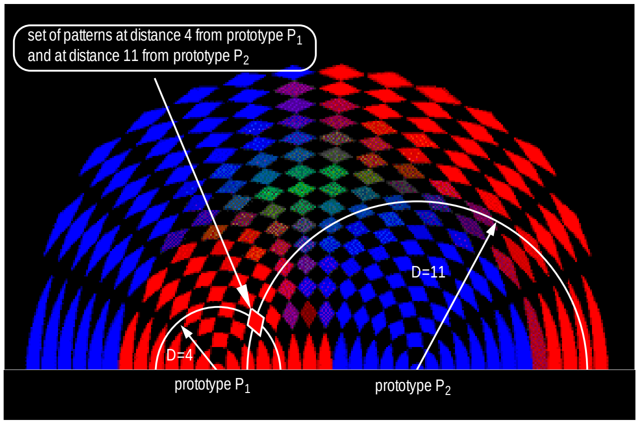


**Fig. 2.**
Scheme of the graphical representation of the attractors' basins of attraction. The drawing in white is superposed on an actual sampling of a network that has learnt three prototypes, to highlight the logical scheme. See text for explanation.

### Attractors structure

In order to understand the behaviour of `LANN27` when exposed to various streams of stimuli, we adopted two 'protocols' of stimuli presentation, to shed light on two aspects of the learning process:

*incremental protocol*: 3 random prototypes (27-bit words) are generated subject to the constraint that each pair should differ by not less than 12 neuron states. In a first stage only the first prototype is repeatedly presented for learning, i.e. for a time long enough to induce significant changes in the synaptic configuration (the duration of the stimulus is such that the transition probability is $\simeq 5\%$). Then, in the second phase the first two prototypes are presented in random permutations, and in the third phase all three prototypes. With this protocol the gradual structuring of the attractors as the input from the environment gets more and more complex is made particularly clear (see Fig.3, left).

*palimpsest protocol*: Three prototypes are generated as in the previous case. Two prototypes are presented in random permutations for learning; after that the second and the third prototypes are presented for learning (the first prototype is no longer presented). It is clear from Fig.3 (right) that the network adapts to the temporal flux of stimuli, forgetting those which are not seen for a long time. The basins of attractions of the prototypes no longer seen are gradually shrunk and finally disappear. This 'palimpsest' property allows the network to make room for new information, discarding what is no longer present in the environment.

The small size of the network implies a low value for its "memory capacity" (which also limits the variety of possible tests). For the Hopfield model, in the limit of an infinite network, crossing the limit of capacity results in an abrupt change in the dynamics of the system: above it none of the memories embedded in the synaptic matrix is a stable state of the dynamics, and the network undergoes a transition to a complete blackout state. When the number of neurons is finite, the situation is more complex and becomes subtle. The criterion adopted to define the limit of capacity of our network is the following: given a learnt pattern $\xi$, and the attractor $s^{\xi}$ that the network develops for it, if all the patterns

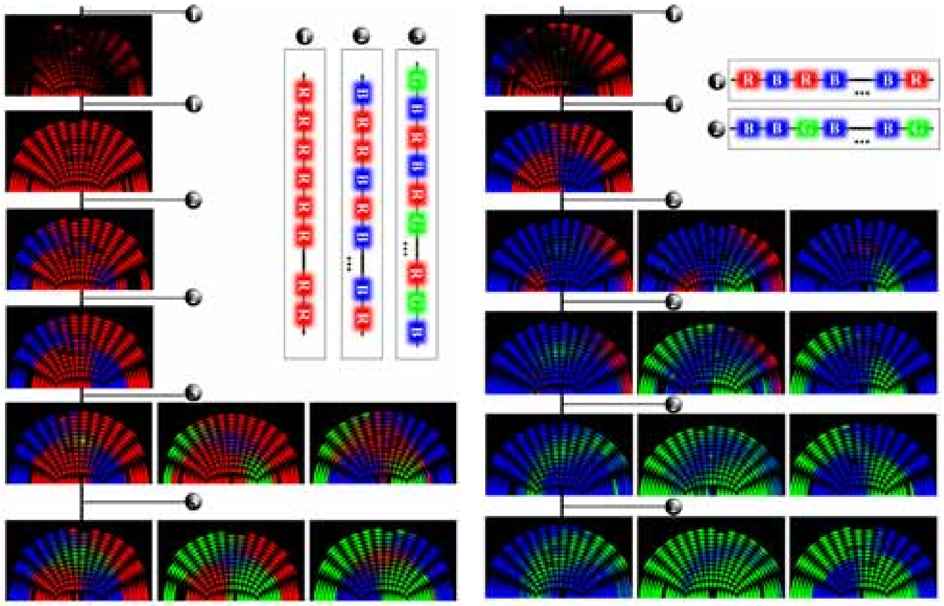

**Fig. 3.** Learning behaviour of `LANN27` for the incremental protocol (left) and for the palimpsest protocol (right). The state space is monitored by performing a retrieval cycle every 5 learning steps (small spheres in the figure represent the sequence of learning presentations illustrated in the corresponding rectangle). *Incremental protocol*: following the presentations of the 'red' prototype, the corresponding attractor grows larger and larger, eventually covering most of the state space with its basin of attraction. Then 'red' and 'blue' prototypes first, and finally the three prototypes, are presented in random permutations. The two last rows contain the three projections of the state space corresponding to the three different pairs of prototypes. *Palimpsest protocol:* in a first stage (for the first two retrieval cycles) only 'red' and 'blue' prototypes are presented for learning. Then the 'red' prototype is not presented any more, and one more stimulus (the 'green' prototype) enters the environment. The 'red' attractor gradually fades away while the novel prototype extends its basin of attraction.

which are nearest neighbours to $\xi$, when used as stimuli, make the network relax to the same $s^\xi$, we say that the network is below the limit of capacity. Notice that the attractor $s^\xi$ can in principle (and in fact sometimes does) differ from $\xi$; the only requirement is that the attractor coincides for each memory and all its nearest neighbours. The analysis, illustrated in Fig.4, shows that we can take 3 as the limit of capacity of our network. The histogram on the left in the figure is constructed as follows: for a fixed number $p$ of prototypes ($p = 1 \dots 4$) 200 sets of $p$ prototypes were generated and a learning session of 1500 presentations was performed for each of the sets. After each learning session, the network was presented briefly with each prototype and the whole set of 27 nearest neighbours patterns (one inverted neuron). The pattern retrieved by the network for each presentation is recorded, and the maximum Hamming distance is measured between the attractor retrieved for the prototype's nearest neighbours and the attractor retrieved when the prototype itself is presented. This is repeated for the $p$ prototypes of each set, and the greatest among the above $p$ distances is calculated, and a histogram of these maximum distances is constructed for each $p$. The histogram on the right is constructed in the same way, for the stable network configurations attracting the 351 patterns at Hamming distance 2 from the prototypes. It is apparent from the histograms that $p = 3$ can be considered as the value of the limit of capacity as defined in the text.

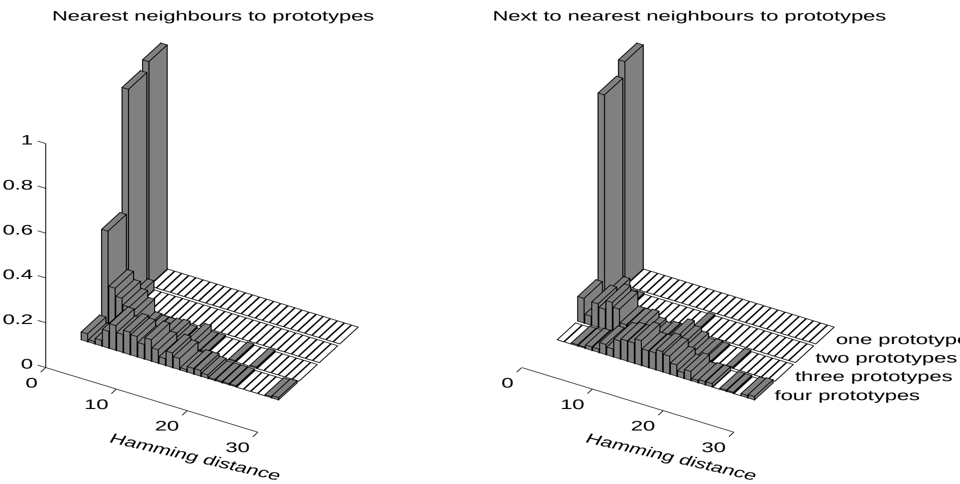


**Fig. 4.** The distribution of attractors for the patterns sourrounding the prototypes for different numbers of learnt prototypes.

*A version of the paper with colour figures can be found on the WEB at:*
`http://chimera.roma1.infn.it/FUSI/index_my_papers.html`
*At the same address a preprint is available with an extended report of this work.*

**Acknowledgements** We thank D. Amit, D. Badoni and M. Mattia, with whom part of this work was carried out, and D. Amit for many comments and suggestions that considerably improved the paper.

## References

1. D.J. Amit, "Modeling brain function", Cambridge University Press, 1989
2. D. Badoni, S. Bertazzoni, S. Buglioni, G. Salina, D.J. Amit and S. Fusi "Electronic implementation of an analogue neural network with stochastic transitions", *Network*, **6** 125 (1995)
3. D.J. Amit and S. Fusi "Learning in neural networks with material synapses", *Neural Computation* **6** 957 (1994)
4. N Brunel, F Carusi and S Fusi "Slow stochastic Hebbian learning of classes of stimuli", *submitted to Network*