Neural networks beyond the mean-field paradigm

Elena Agliari [a,b], Adriano Barra[b], Andrea Galluzzi [c], Francesco Guerra [b,d], Daniele Tantari [e],Flavia Tavani [f]

[a] Università Campus Biomedico di Roma
[b] Dipartimento di Fisica, Sapienza università di Roma
[c] Dipartimento di Matematica, Sapienza università di Roma
[d] Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Roma
[e] Dipartimento di Matematica, Scuola Normale Superiore di Pisa
[f] Dipartimento Scienze di Base e Applicate all'Ingegneria (SBAI), Sapienza università di Roma
e-mail: flavia.tavani@sbai.uniroma1.it

Keywords: (neural networks, serial processing parallel processing)

The seminal paper "Neural Networks and physical systems with emergent collective computational abilities" by John Hopfield (1982) and its statistical mechanical treatment by Amit, Gutfreund and Sompolinsky (1985) still play as "*harmonic oscillators"* in Artificial Intelligence: crucially, in their picture, "associative memory" emerges as a collective feature of neurons. It showed how, using the mean- field approximation, every neuron interacts with all the others in the networks in order to perform a ``serial processing", namely it is able to retrieve one pattern of information per time. The main limitation oft he mean-field approximation is that it does not take into account the mutual distance between the elements in the network, but it has been clarified that in real-world networks (for example the brain network) the elements interact each other as strongly as their distance is lower. This presentation aims to show how, overcoming the mean field approximation, introducing the hierarchical structure on the network, and showing how this permits to obtain a structure able to perfom both serial and parallel processing.

We start from the most famous hierarchical model, Dyson's Hierarchical Model (or DHM) [1], is built up recursively, starting from a couple of nodes, linked with the same strength, and then creating an identical copy of the existing dimer, and connecting them with weaker bonds. Iterating this procedure, one creates at every step an identical copy of the existing structure, and set up the new links. The resulting network has two features: the first one is that two nodes connected at the d-th step of the algorithm are considered at distance d (so one can define a matrix, whose elements are the distances between nodes), and the second one is that links that connect nodes at distance d have stronger weight than links connecting nodes at distance d=d+1, …, K, with K the total number of steps of the procedure (see Fig. 1). First, we approach these systems à la Mattis, by thinking at the Dyson model as a single-pattern hierarchical neural network. One step forward, we extend this scenario toward multiple stored patterns by implementing the Hebb prescription for learning within the couplings. This results in a Hopfield-like networks constrained on a hierarchical topology, for which, restricting to the low storage regime (where the number of patterns grows at most logarithmical with the amount of neurons) we give an explicit expression of its mean field bound and of the related improved bound. Our main finding is that embedding the Hebbian rule on a hierarchical topology allows the network to accomplish both serial and parallel processing (see Fig 2). By tuning the level of fast noise affecting it, or triggering the decay of the interactions with the distance among neurons, the system may switch from sequential retrieval to multitasking features and vice versa. However, as these multitasking capabilities are basically due to the vanishing "dialogue" between spins at long distance, such an effective penury of links strongly penalizes the network's capacity, which results bounded by the low storage [2], [3], [4].

These networks display a richer phase diagram than their classical counterparts. In particular, these networks are able to perform serial processing (i.e. retrieve one pattern at a time through a complete rearrangement of the whole ensemble of neurons) as well as parallel processing (i.e. retrieve several patterns simultaneously, delegating the management of different patterns to diverse communities that build network, see Fig 1). The tune between the two regimes is given by the rate of the coupling decay and by the level of noise affecting the system. The price to pay for those remarkable capabilities lies in a network's capacity smaller than the mean field counterpart, thus yielding a new budget principle: the wider the multitasking capabilities, the lower the network load and vice versa. This may have important implications in our understanding of biological complexity.
One step forward, we analyzed also how many patterns the DHM arrives to retrieve. In fact, applying the same arguments used to show its capability to perform parallel processing, we showed that these networks have key motifs: in particular, we consider the dimer, i.e., the prototype of a loop-less reticular animal, and the square, i.e., the prototype of a loopy reticular animal, and we check whether magnetic configurations where spins associated to these motifs are misaligned with respect to the bulk are stable [5]. Not surprisingly, while the formers is found to be always unstable (i.e. there is no value of the tunable parameters defining the model that allows its stability), the latter has a range of stability. It is worth noting, however, that -as these motifs are by definition not-extensive (i.e. their sizes do not scale with the system size) - nor do they contribute to the model free energy in the thermodynamic limit, neither are they expected to be stable whenever a finite-amount fast noise is applied on the system. As a last remark we note that the Dyson model implies a modular architecture of the embedding structure and the reason for the stability of its loopy motifs lies exactly in the intrinsic modularity of the system: remarkably, modularity plays a major role even in real biological networks exactly those where the presence of motifs is expected.

[1]. F. Dyson, Existence of a phase transition in a one dimensional Ising ferromagnet, Comm. Math. Phys. 12, 91-107 (1969).

[2]. Metastable states in the hierarchical Dyson model drive parallel processing in the hierarchical Hopfield network, *Agliari E., Barra A., Galluzzi A., Guerra F., Tantari D., Tavani F.,* Journal of Physics A: Mathematical and Theoretical 48 (1), 015001 (2015).

[3]. Retrieval capabilities of hierarchical networks: From Dyson to Hopfield, *Agliari E., Barra A., Galluzzi A., Guerra F., Tantari D., Tavani F.,* Physical Review Letters 114, 028103 (2015).

[4]. Hierarchical neural networks perform both serial and parallel processing, *Agliari E., Barra A., Galluzzi A., Guerra F., Tantari D., Tavani F.,* Neural Networks 66, 22–35 (2015).

[5]. Multitasking cliques in hierarchical networks, *Agliari E., Barra A., Tavani F., Tirozzi B.,* Theory and Applications in Mathematical Physics [B. Tirozzi, G.Montani, E. Agliari, A. Barra, N. Carlevaro Eds., World Scientific Publishing (2015)].

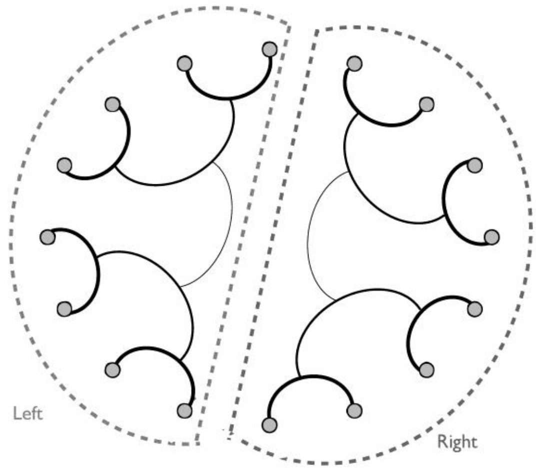

Figure 1. Schematic representation of the hierarchical topology where the associative network insists. Circles represent Ising neurons (N = 16 in this shapshot) while links are drawn with different thickness mimicking various interaction strengths: The thicker the line, the stronger the link.

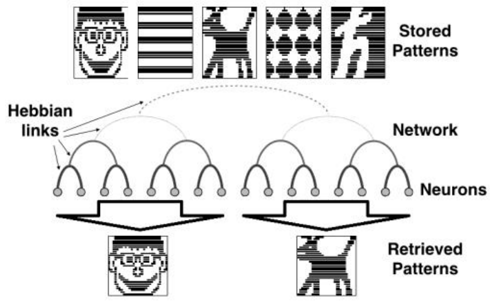

Figure 2. Example of parallel retrieval of patterns in a Hierarchical Hopfield Network with K=4 and N=16 total number of neurons.