

available at [www.sciencedirect.com](http://www.sciencedirect.com)[www.elsevier.com/locate/brainres](http://www.elsevier.com/locate/brainres)


---



---

**BRAIN  
RESEARCH**


---



---



---

**Research Report**
**Lateral thinking, from the Hopfield model to cortical dynamics**
**Athena Akrami, Eleonora Russo, Alessandro Treves\***

 SISSA, Cognitive Neuroscience sector, via Bonomea 265, 34136 Trieste, Italy
 

---

**ARTICLE INFO**
**Article history:**

Accepted 13 July 2011

Available online 23 July 2011

**Keywords:**

Neural computation

Associative memory

Cortical dynamics

Modular network

**ABSTRACT**

Self-organizing attractor networks may comprise the building blocks for cortical dynamics, providing the basic operations of categorization, including analog-to-digital conversion, association and auto-association, which are then expressed as components of distinct cognitive functions depending on the contents of the neural codes in each region. To assess the viability of this scenario, we first review how a local cortical patch may be modeled as an attractor network, in which memory representations are not artificially stored as prescribed binary patterns of activity as in the Hopfield model, but self-organize as continuously graded patterns induced by afferent input. Recordings in macaques indicate that such cortical attractor networks may express retrieval dynamics over cognitively plausible rapid time scales, shorter than those dominated by neuronal fatigue. A cortical network encompassing many local attractor networks, and incorporating a realistic description of adaptation dynamics, may be captured by a Potts model. This network model has the capacity to engage long-range associations into sustained iterative attractor dynamics at a cortical scale, in what may be regarded as a mathematical model of spontaneous lateral thought.

*This article is part of a Special Issue entitled: Neural Coding.*

© 2011 Elsevier B.V. All rights reserved.

---

**1. Introduction: a universal cortical transaction?**

Information-processing models of cognitive functions, a most productive approach developed over the last few decades, have usually described those functions in terms of sequences of specialized routines, conceptually akin to components of a complex computer code. For example, the computation of the trajectory to reach a particular goal in space may be described as entailing the transformation of spatial information arriving through the senses from sensor-based to allocentric coordinates, then the construction or extraction from a memory store of the relevant map of the environment, then the geometric calculation of the available paths connecting the current position and the goal, and of their properties, such as

time needed, energy expenditure, chances of failure (Hikosaka et al., 1999; Kawato, 1999; Tanji, 2001; Wolpert, 1997). Reading written text, instead, may be described as entailing the extraction of line and corner elements, the recognition of the abstract invariants characterizing each letter, the composition of individual letters with error correction to form meaningful words, a further error correction stage that takes into account neighboring words, and a cascade of higher-level lexical and semantic processes (Plaut, 1999). Associative processes have often been seen as alternative side paths to the orderly usage of such specialized routines, “lateral thinking” that may occasionally provide a shortcut to, and more frequently derail, the successful execution of a task.

The ready availability of functional imaging techniques has encouraged the further elaboration of such information

---

\* Corresponding author. Fax: +39 040 3787528.

E-mail address: [ale@sissa.it](mailto:ale@sissa.it) (A. Treves).URL: <http://www.sissa.it/~ale/> (A. Treves).

processing models, promising to assign distinct cortical areas as the theatres for the operation of several of the routines. Yet, activation patterns observed with fMRI cannot resolve single neuron activity, which would be necessary in order to test information processing models at the algorithmic level. Evidence from neurophysiological recordings in brain slices, in rats and in macaques, and sparsely in human patients, on the other hand, have essentially provided no evidence for any other neuronal operation taking place in the cortex, other than associative processes: associative synaptic plasticity and associative retrieval. The hypothesis has to be entertained, therefore, that the cerebral cortex may contribute nothing but associative network processes, although they may be “dressed up” in different guises depending on the connectivity of each cortical area and on the codes it expresses. For such a hypothesis to be subject to validation or falsification, however, the notion of associative processes has to be made precise, for example in terms of a mathematically defined network model.

The Hopfield (1982) model meets the requirements for a mathematically well-defined model of associative memory retrieval, as it could be implemented in a local cortical network. Its cortical plausibility has been questioned, however, because of several dramatic simplifying assumptions it relied on, at least in the original version, as analyzed mathematically by Amit, Gutfreund and Sompolinsky (1985, 1987). Moreover, it is a simplified model of memory retrieval based on an even cruder model of associative memory storage. Over the nearly 3 decades since it was put forward, the effect of many of those simplifications has been analyzed, mathematically and with computer simulations, and overall it has been found not to alter the qualitative import of the model (see Rolls and Treves, 1998). In this contribution, based in part on a PhD Thesis (Akrami, unpublished) and including some original results, we discuss quantitatively, with computer simulations and with reference to recordings in monkeys, some of the crucial conceptual steps that bridge the gap between the Hopfield model and local cortical circuits, particularly with regard to how memory representations may be stored, and to the time scale for retrieval dynamics. The aim is to assess the validity of a yet more abstract model of a local cortical network—a single Potts unit—as a building block of models of extended cortical networks, which operate exclusively through associative processes.

## 2. Results

In the original Hopfield model, a memory item is retrieved from the network when neural activity, stimulated by a partial cue (usually given as a starting condition for the network), evolves into a pattern strongly correlated with one of the  $p$  representations which have been stored on synaptic weights. Daniel Amit (1995) and others have pointed at such “attractor dynamics” as a robust universal mechanism for memory retrieval in the cerebral cortex, and in the hippocampus. How smoothly can this retrieval operation proceed, and how wide the “basins of attraction” are of the  $p$  memory states, should depend on how memory representations are established during the storage phase, which determines whether other

attractors may hinder or obstruct retrieval. In the hippocampus, new memory representations are believed to be established under the dominant and decorrelating influence of the specialized dentate gyrus preprocessor, with its strong, sparse connectivity to CA3, so an *ad hoc* analysis is required (Cerasti and Treves, 2010). To assess, instead, how the storage process affects retrieval capacity in the cortex, where no dentate inputs are available, it is necessary to consider first the main factors that determine the effectiveness of attractor dynamics: connectivity, representational sparseness, the presence of noise.

### 2.1. Effective retrieval capacity with cortically realistic storage processes

A most important factor that determines retrieval is the degree of connectivity in the network. In the original Hopfield model the connectivity is complete, i.e. each of the  $N$  units in the network receives input from all other  $N-1$  units (Hopfield 1982). This simplifying assumption was linked to imposing symmetry on the coupling constants, that is, the synaptic weights, which in turn led to a great clarification of the properties of auto-associative neural networks (Hopfield 1982; Amit et al., 1985, 1987). The analysis of network performance derived from statistical physics and applicable in the “thermodynamic” limit  $N \rightarrow \infty$  can however be extended to the case where the number  $C$  of inputs per unit is smaller than  $N$ , but still it is regarded as very large,  $C \rightarrow \infty$  (Sompolinsky, 1986); and even in the so called “highly diluted” limit ( $C \rightarrow \infty$  but  $C/N \rightarrow 0$ ) considered by Derrida et al. (1987). The symmetry of the weights can likewise be discarded, leading to characterize some interesting dynamical properties of asymmetric networks (Derrida and Pomeau, 1986; Sompolinsky and Kanter, 1986; Derrida et al., 1987; Kree and Zippelius, 1987; Crisanti and Sompolinsky, 1987; Gutfreund et al., 1988). Overall the main insight gained by introducing incomplete or sparse connectivity,  $C < N-1$ , is rather simple: the capacity of networks with symmetric or asymmetric weights is primarily determined by  $C$ , and to a lesser extent by  $N$ . This result has been derived many years ago, for example from a signal-to-noise analysis, applicable when the synaptic weights encode  $p$  uncorrelated memory patterns represented as sparse activity distributions of sparsity  $a$  (Treves and Rolls, 1991). “Noise” here denotes the interference due to other memory patterns, what in the physics jargon is dubbed *quenched* noise. The analysis shows that the signal scales as  $C$ , and the noise as  $\sqrt{pC}$ . The relation between the maximum number  $p_c$  of patterns that can be turned into dynamical attractors, i.e. that can be associatively retrieved, and the number  $C$  of connections per receiving unit takes the form, for sparsely coded patterns (in the limit  $a \rightarrow 0$ ; Treves and Rolls, 1991)

$$p_c \sim k \frac{C}{a \ln(\frac{1}{a})}$$

where  $k$  is a numerical factor of order 0.1–0.2. In this limit  $p_c$  is independent of  $N$ .

The essential advantage introduced by the sparse connectivity, if randomly diluted, is that quenched noise has less of an opportunity to reverberate coherently. The signal from the

Download English Version:

<https://daneshyari.com/en/article/6264489>

Download Persian Version:

<https://daneshyari.com/article/6264489>

[Daneshyari.com](https://daneshyari.com)