



**The Abdus Salam
International Centre for Theoretical Physics**



1860-9

**Borsellino College 2007. Spike Trains to Actions: Brain Basis of
Behavior**

3 - 14 September 2007

Evolution of hippocampus and neocortex (Spatial memory codes in CA3)

Alessandro TREVES
Cognitive Neuroscience Sector, SISSA, Trieste, Italy

Trieste, Sept 3rd, 2007

THE CA3 NETWORK AS A MEMORY STORE FOR SPATIAL REPRESENTATIONS

Gergely Papp¹, Menno P Witter^{2,3} and Alessandro Treves^{1,2}

¹SISSA, Cognitive Neuroscience Sector, Trieste, Italy

²NTNU, Kavli Institute for System Neuroscience and Centre for the Biology of Memory,
Trondheim, Norway

³VUMC, Department of Anatomy & Neuroscience, Amsterdam, The Netherlands

Short title: Spatial memory codes in CA3

Address all correspondence to: Alessandro Treves, SISSA, via Beirut 2, I-34014 Trieste, Italy

tel. +39-040-3787623, fax. +39-040-3787615, email: ale@sissa.it

Total number of words in text: 9,237

Total number of figures: 5, in b & w

Acknowledgements: we are grateful to all colleagues at the Centre for the Biology of Memory and to Floris Wouterlood and Piet Hoogland at the VUmc, and to two reviewers for their helpful suggestions.

Abstract.

Comparative neuroanatomy suggests that the CA3 region of the mammalian hippocampus is directly homologous with the medio-dorsal pallium in birds and reptiles, with which it largely shares the basic organization of primitive cortex. Autoassociative memory models, which are generically applicable to cortical networks, then help assess how well CA3 may process information and what are the crucial hurdles it may face. The analysis of such models points at spatial memories as posing a special challenge, both in terms of the attractor dynamics they can induce and of how they may be established. Addressing such challenge may have favoured the evolution of elements of hippocampal organization observed only in mammals.

1. Introduction: a comparative perspective

The forebrain of vertebrates shows remarkable morphological variation and specialized adaptations, presumably related to differences in environment. Yet, in all species, part of the brain apparently is involved in the formation of map-like representations of the environment. The need to localize distributed resources such as food, shelter and so on is an essential component of the evolutionary success of freely moving species. It is quite likely that the formation of map-like representations strongly depend on the potential to generate relational representations of environmental features (of course, in order to navigate, animals do not simply rely on these relational maps between landmarks, but also make use of e.g. idiothetic cues).

In all species studied, including teleost fish, amphibians, reptiles and mammals, relational representations are mediated by structures in the brain that embryologically derive from the most medial part of the telencephalic anlage (Rodriguez et al., 2002). In mammalian species, this derivative is called the hippocampal formation and it is of interest that a correlation apparently exists between the volume of the hippocampal formation and successful spatial performance. It is therefore not surprising that the mammalian hippocampal formation is considered to represent one of the phylogenetically oldest cortical areas. Comparative neuroanatomical studies, in conjunction with embryological studies as well as data on gene expression patterns have indicated that the hippocampal formation in mammals most likely is homologous to parts of the medial cortex in reptiles (Ulinsky, 1990, Lopez-García and Martínez-Guijjaró, 1988, ten Donkelaar, 2000). The hippocampal formation originates from a dorsomedially positioned anlage in the developing brain which shows a gross overall similarity with the dorsomedial cortex in reptiles (Stephan, 1975). In birds, the cortex of the medial surface of the pallium merges with more ventrally located pallial structures such that the apparent typical layered structure of the cortex no longer is apparent.

The overall wiring of the mammalian hippocampal formation (Amaral and Witter, 1989) will not be reviewed here. Suffice it to note that, both with respect to its extrinsic as well as its intrinsic connections, it shows remarkably strong similarities even when such different species as the mouse and monkey are compared. This indicates that there may be some ecological benefit for this stabilized connectivity. In order to gain insight into the functional relevance of this particular organization of the hippocampal system, only a few approaches are at hand. Either we test the effects of loss of parts of the system, for example in lesioned animals (Kesner et al, 2002), or we adopt computational modeling approaches, in which, at the cost of many necessary simplifications, network organization can be varied at will. Both approaches have to be informed by an appreciation of the organization of homologous structures in non-mammalian species of vertebrates, which indicates which changes and refinements may have been driven by evolutionary pressures.

2. The medial pallium in reptiles and in birds

2.1. Structure and connectivity of the mediodorsal cortex of reptiles.

The cortex in reptiles is generally divided into mediodorsal cortex, dorsal and lateral cortex. The mediodorsal cortex is further subdivided into a small-celled part and a large-celled one, also referred to as medial and mediodorsal cortex (Fig. 1a,b). Like the cortex that makes up the mammalian hippocampus, within the reptilian cortex three different layers are distinguished, an outer molecular layer, the cell layer, and, bordering the underlying white matter, the polymorph layer (Fig. 1d). In general the small celled portion of the reptilian cortex is considered to be comparable to the dentate gyrus of the mammalian hippocampus. The large celled part is taken to be comparable to the CA-fields, but in reptiles no further subdivisions have been distinguished, for example between a CA3-like and a CA1-like region.

Neurons in the small celled portion are pyramidal or spherical, with an average diameter of about 6-18 μm . They are closely packed and extend dendrites into the molecular layer as well as into the polymorph layer. Similar to what is seen in the dentate gyrus in mammals, a majority of the dendrites do extend into the molecular layer and the first bifurcation is close to the soma. In contrast to the mammalian situation though, the cell types appear to be more variable, such that a total of 6 different cell types have been described within the cell layer, of which only some contribute to projections targeting adjacent portions of the cortex, mainly to the large celled part of the mediodorsal cortex (Wouterlood, 1981; Olucha et al., 1988). These projections are strongly Timm positive, strikingly similar both morphologically as well as in terms of overall distribution to the mossy fiber projection in mammals (Olucha et al., 1988; Lopes Garcia and Martinez-Guijaro 1988). Although not much is known about the different local circuitry neurons in the reptilian brain, neurons have been described in both the molecular layer and the deep polymorph layer.

The cell layer of the large celled portion comprises mainly one cell type, which shows a polygonal or pyramidal shape, with large apical dendrites extending into the molecular layer as well as basal dendrites extending into the polymorph layer and adjacent white matter. Short basal dendrites extend from the soma parallel to the cell layer. Additional neurons in the molecular and polymorph layer have been described as well (Wouterlood, 1981, ten Donkelaar, 2000). Note that in several mammalian species, the anterior (supracallosal) continuation of the hippocampus, indusium griseum and tenia tecta (considered the olfactory hippocampus) shows similarities to the lizard medial cortex, where the dentate and CA fields form a continuous sheet of cells with two morphologies, granule and pyramidal (Stephan, 1975; Wyss and Sripanidkulchai 1983; Shipley and Adamek, 1984; Kunzle 2004). Projections distribute widely in the cortex, including a return projection to the small celled part. Most notably, widespread dense intrinsic projections have been described distributing both to apical as well as basal dendrites, reminiscent of the associative connectivity in the mammalian CA3 (Hoogland and Vermeulen-Vanderzee 1993; Olucha et al., 1988).

2.2. Structure and connectivity of the dorsomedial telencephalon of birds.

In birds, the dorsomedial telencephalon is considered the homologue of the hippocampal region of mammals. However, a detailed comparison between subdivisions of the avian and mammalian hippocampus, and even between birds and reptiles, is still speculative. According to the detailed

descriptions of Ariens Kappers et al (1936), there is a layered cortical structure that might be considered the avian hippocampus and parahippocampal region. In the chicken, the dorsomedial cortex comprises a superficial plexiform layer, a granular layer and a periventricular or polymorph layer (Molla et al., 1986). Interestingly, according to the latter authors, the cells in the granular layer are pyramidal or bipyramidal cells, similar to what has been reported for the mediodorsal cortex in a number of reptilian species. On the basis of afferent and efferent connections, it can also be argued that the dorsomedial telencephalic area in birds is comparable to the hippocampal formation and parahippocampal region in mammals (see Dubbeldam 1998 for a review).

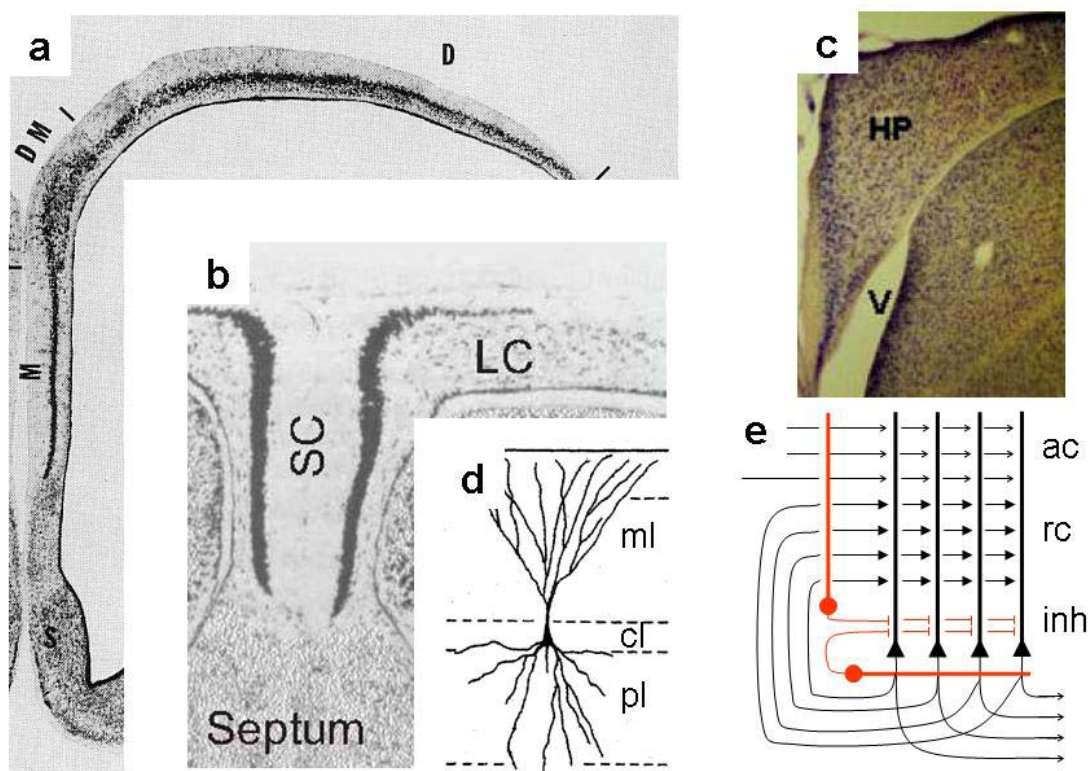


Figure 1. The reptilian and avian forebrain and the generic cortical architecture of autoassociative memory networks. Partial cross sections through cerebral emispheres of a crocodile (a), a lizard (b) and a zebra finch (c). Pyramidal cell positioned in phylogenetically ancient layered cortex (d) and autoacciative network based on plastic recurrent connections (e). Abbreviations (by panel): D, dorsal; DM, dorso-medial; M, medial; LC, large celled; SC, small celled; HP, hippocampus; V, ventricle; cl, cell layer; ml, molecular layer; pl, polymorph layer; aff, afferent connections; inh, (feedforward and recurrent) inhibition; rc, recurrent connections. (a), (b) and (d) redrawn from Gloor (1997); (c) from Sadananda (2004).

Two issues appear relevant here. First, the question arises whether or not in birds, hippocampal and parahippocampal subdivisions are apparent that show cytoarchitectonic and/or connectional similarities to those described in mammals; and second, it is of interest to assess whether the typical unidirectional connectivity seen in the mammalian hippocampal system is present in birds. Unfortunately, staining the avian brain for the distribution of Zinc with the Timm stain, which in

mammals and reptiles stains the projection from neurons in the granular cell layer to a group of large pyramidal cells, does not provide useful clues. In birds the Timm stain only results in rather diffuse and weak staining (Faber et al., 1989, Montagnese et al., 1993, 1996).

A case of mistaken identity?

Two recent studies addressed both issues by studying the connectional organization of the dorsomedial telencephalic domain of the pigeon in much detail (Kahn et al., 2003; Atoji and Wild 2004). Both studies describe, within this part of the avian brain, a V-shaped ventral medial region, subdivided into a lateral and medial blade of neurons which enclose an area called central or triangular region (cp. Fig.1c). This ventral region is bordered more dorsally by an area referred to as dorsomedial cortex, followed by a more dorsolaterally positioned dorsolateral cortex. Both follow the original descriptions of Karten and Hodos (1967), supplemented with neurochemical patterns as described by Erichsen et al (1991). On the basis of the connectional data, both papers describe a series of connections comparable to the mammalian trisynaptic circuit. Unfortunately, the results lead Kahn et al (2003) to suggest that the ventral domain is comparable to CA1, similar to suggestions based on anterograde tracing in the zebra finch (Székely and Krebs 1996), whereas Atoji and Wild (2004) suggest that this region actually represents the dentate gyrus. In terms of position and shape the suggestion that the ventral region actually is the “dentate” of the avian brain is the more apparent one (see also Atoji and Wild, 2006); however, no clear conclusions can be stated here at this point in time. Moreover, both in birds and in reptiles, the dorsomedial cortex issues projections to the lateral septum, similar to what is reported for CA3 and CA1 in mammals. Irrespective of whether the ventral or dorsomedial areas should be considered the homologue of the mammalian dentate gyrus, neither of the connectional studies indicate that in the avian brain unidirectional connections of one cortical area to another are present that form large, Zinc-positive terminals that may function as detonator synapses, as proposed for the synapses on the mossy fibers to CA3 cells in mammals (Andersen, 1962). In all reports, bidirectional connections are in fact predominant.

Physiology and plasticity of the avian hippocampal formation

In comparison to the vast amount of neurophysiological data in mammals, there are relatively few studies analyzing the activity of neurons in the avian hippocampal formation, and they point at broad similarities with the spatially selective activity of mammalian hippocampal neurons. These similarities are consistent with the intensely studied role of the avian hippocampus in episodic and spatial memory (Krebs et al., 1989; see the discussion by Clayton et al., 2003; Healy et al, 2007). As reviewed by Bingman and Sharp (2006), location specific cells, “arena-off” cells and path cells have been described in the pigeon hippocampus, and even some putative “grid-like” cells. Recordings have also revealed the presence of oscillatory activity, in birds, similar to the theta rhythm of rodents, though slower (Siegel et al., 2000). Different cell types, however, could not be reliably localized to distinct subdivisions within the hippocampus (see Hough and Bingman, 2004), or to a distinct layer within a subdivision, with only some distinction based on electrophysiological cell properties, perhaps corresponding to principal cells and interneurons (Siegel et al., 2002). Units in the ventral regions might tend to be similar in spike width and frequency to the pyramidal cells of CA regions in the rat, while units in the dorsocaudal region might be more similar, in terms of firing pattern, to the granule cells of the mammalian DG, perhaps in support of the conclusions of Kahn et al (2003). If anything, there is clearer evidence for lateralization, e.g. path cells are found on the left side, and location specific cells show more prominent selectivity on the left (Siegel et al., 2006).

Forms of synaptic plasticity potentially involved in learning and memory such as LTP and LTD have also been discovered in the avian brain (Scott and Bennett, 1993, Wieraszko and Ball, 1993, Margrie et al, 1998). There is also evidence for the formation of new synapses after training (Ünal et al., 2002), as well as considerable long-standing evidence for neurogenesis, which antedates parallel evidence more recently found in mammals. Nevertheless, in contrast to the extensive literature available for rodents, synaptic plasticity has not been described in subregional detail.

Homology and beyond

In conclusion, our current understanding does not allow us to draw firm conclusions concerning structures in the avian brain that would represent the mammalian dentate gyrus, despite the homology between the overall 'hippocampal region' in the two lineages (Colombo and Broadbent, 2000). Importantly, some aspects of the connectivity are quite different, leading one to doubt the relevance of forcing a correspondence at the subregional level, given hundreds of million years of separate evolution (Striedter, 2005). The salient exception is that, as for the reptilian one, the avian hippocampus appears to retain a connectional system comparable to the autoassociative system of the mammalian CA3 field (Fig. 1e). This suggests that a fundamental associative memory function of the primitive cortex (Braitenberg and Schüz, 1991) may have been largely preserved in all amniotes, based on its recurrent architecture and associative synaptic plasticity, though perhaps made more complex and powerful in different ways in reptilian, in avian and in mammalian derivatives. This simplifying perspective, that regards the CA3 field as part of the common amniotic heritage, and DG and CA1 as novel mammalian concoctions (although with similarities, particularly in the distribution of Zinc, with reptiles) informs the computational analyses described next.

3. The CA3 region as an autoassociative memory

David Marr (1971) brilliantly synthesized ideas about the role of the hippocampus in memory formation, which he may have subsumed, but indirectly, from neuropsychological studies, and took them as the starting point to understand the organization of hippocampal circuits. This ‘structure-from-function’ theoretical research program has been enormously influential, even though the details of his modeling approach are difficult to appraise. For example, Marr eloquently emphasized, in words, the “collateral effect” i.e. the potential role in pattern completion of recurrent connections, prominent among CA3 pyramidal cells (Amaral et al, 1990); but his own model was not really affected by the presence of such collaterals, as shown later by careful meta-analysis (Willshaw and Buckingham, 1990). Marr thought in terms of discrete memory states, and devoted an entire section of his paper to ‘capacity calculations’, which indicates that he realized how an estimate of the maximum number of activity patterns retrievable from a memory network could be a central contribution of mathematical models. The simple model he considered, with binary processing units and binary synaptic weights, is endowed with the capacity to retrieve up to p_c patterns, where $p_c (a_{CA})^2 \ll 1$, and a_{CA} is the activity level of CA pyramidal units, i.e. the fraction of units active in a pattern (which generalizes to the sparsity parameter for non binary patterns). Marr stated that with an activity level $a_{CA} \approx 0.001$, the hippocampus could store and retrieve of the order of $p_c \approx 100,000$ memories, which he reckoned was a reasonable number, at least for a temporary memory store. Those ‘doubly binary’ models cannot be easily related to real CA3 neuronal networks, however, and even if they were, the observed mean sparsity $a_{CA3} \approx 0.03-0.04$ (see e.g. Papp and Treves, 2007) would lead to a rather dismal capacity of about $p_c \approx 100$ patterns. Although the paper by Marr was nearly simultaneous with the discovery of place cells (O’Keefe and Dostrovsky, 1971) and with that of long term synaptic potentiation (Bliss and Lømo, 1973; cited as a note added in proof) for a long time Marr did not seem to inspire further theoretical analyses - with the exception of an interesting discussion of the collateral effect in a neural network model (Gardner-Medwin, 1976).

3.1. The Dentate Gyrus as a generator of CA3 activity

With their 1987 review, McNaughton and Morris re-examined and revived the Marr framework, discussing several ‘Hebb-Marr’ associative memory model architectures and whether they resembled hippocampal networks. The operation of such models of *attractor* networks (Amit, 1989) can be more readily analyzed if the memory patterns to be stored are assigned ‘by hand’, rather than self-organized under the influence of on-going inputs. One can imagine that a system of strong one-to-one connections from another area may effectively ‘transfer’ a pattern of activity from the other area, where it is determined by some unspecified process, to the associative memory network. McNaughton and Morris (1987) took the strong “detonator” synapses on the MF projections from DG to CA3 (Andersen, 1962) as an approximate implementation in the real brain of such one-to-one connections. The distributions of activity to be stored in memory would be effectively generated in DG, perhaps by expansion recoding (as hypothesized for granule cells in the cerebellum) and then simply transferred to CA3.

As clarified by Treves and Rolls (1992), for the dentate gyrus to ‘impose’ a novel pattern of activity onto CA3, it need not transfer its own, and one-to-one connections are not necessary: they would not be very explanatory either, in that they would simply anticipate the question of how to generate an appropriate pattern of activity back to DG, rather than resolving it. What matters, instead, is that the MF synapses be strong, sparse, and conveying sparse activity from DG. This is sufficient to

effectively select a limited ensemble of CA3 cells to represent a new memory, unrelated to ensembles which are co-activated in other, previously stored memories, and which would tend to be reinstated by the collateral effect. The competition between the MF projections, forcing a novel ensemble, and the recurrent connections, reinstating fragments of previously stored ones, is a quantitative balance, which has to be shifted towards the MF inputs only when the system is in 'storage mode'.

When retrieving a previously stored memory representation, the input to the recurrent network has to relay a cue, expressed as a pattern of activity on the afferent fibers correlated with the pattern at the time of storage of a particular memory. Simple algebra shows that such correlation tends to be washed away unless the afferents make many associatively modifiable synapses onto each receiving cells, synapses that have to be modified during storage (Treves and Rolls, 1992). The CA3 region would then work better as a recurrent associative memory, implementing pattern completion thanks to the collateral effect (Rolls, 1989), if it had two separate afferent systems: one particularly strong at the time of storage, relaying sparse activity, and possibly weak at retrieval – which can be identified with the mossy fibers – and one relayed by many synapses, highly plastic but weak during storage, and transmitting at retrieval – to be identified with the perforant path.

The argument requires DG activity to be sparse or moderately sparse (values of $a_{DG} = 0.1$ or 0.02 produce nearly identical results in the mathematical model) but it does not constrain further the form of granule cell activity, e.g. in its spatial correlates. It does predict, however, that since the useful role of the 'duplicated' DG input is only in establishing new CA3 representations, lesioning DG or blocking MF transmission should have no effect on memory retrieval. The prediction is so far consistent with behavioral results obtained in two independent experimental approaches (Lassalle et al, 2000; Lee and Kesner, 2004).

3.2. The function of neuromodulation

A selective modulation of the activity (and plasticity) of specific synaptic systems may be effected by acetylcholine (ACh), exploiting the orderly arrangement of pyramidal cell dendrites in the cortex, which allows for differential action on the synapses distributed in distinct layers (Hasselmo et al., 1994). Acetylcholine is one of several very ancient neuromodulating systems, well conserved across vertebrates, and it may have operated in this way already in the early reptilian cortex, throughout its subdivisions. Hasselmo has emphasized this likely role of ACh in memory, with a combination of slice work and neural network modeling (Hasselmo et al., 1995, 1996). This work has been focused on the hippocampus - originally, the medial wall - and on piriform cortex - originally, the lateral wall. The proposed mechanism, however, has no reason to be circumscribed to these regions, and it could well operate across cortical systems involved in memory storage. A drawback of relying on ACh modulation alone is that it requires an active process that distinguishes storage from retrieval periods, and regulates ACh-release accordingly. Combining ACh modulation with MF orthogonalization may allow the hippocampus to operate efficiently also during behaviors, such as exploration, when storage and retrieval are perhaps largely admixed. In addition, the more complex circuitry allows for further neuromodulatory control, e.g. through dopamine release (Kobayashi and Suzuki, 2006). In refining the architecture of the hippocampus, therefore, it could be that mammals have devised a more efficient memory process, which augments rather than replacing the earlier one based on neuromodulation.

3.3. Quantitative analyses of storage capacity

Such a hypothesis leads to predict quantitative rather than qualitative effects for most manipulations, and hence it has to be articulated in terms of network models that allow quantitative measures. Unfortunately the rudimentary binary-synapse model considered by Marr (1971) does not allow a meaningful correspondence with experimental measures, as indicated by the many predictions at the end of his paper, which have remained untested. The associative memory model introduced by John Hopfield (1982), on the other hand, proved suitable for a theoretical breakthrough when it was analyzed by Amit, Gutfreund and Sompolinsky (1987). They were able to calculate its storage capacity by applying mathematical techniques from statistical physics, which could later be applied to biologically more plausible versions of the model. The generic result is that a network with C recurrent connections per unit can retrieve, i.e. complete, up to

$$p_c \approx 0.2-0.3 C/[a \ln (1/a)], \quad (1)$$

patterns of activity, where a is their sparsity value (in the sparse, $a \ll 1$ regime; see Treves and Rolls, 1991). Each such representation can contain at least

$$i \approx N a \ln (1/a) \text{ bits} \quad (2)$$

of new information, where N is the number of units. This leads to a maximum total amount of information, cumulating the contribution of all memory patterns, which is proportional to the number of synapses NC , and which does not depend much on sparsity

$$I_{\max} \approx 0.2-0.3 N \cdot C \text{ bits.} \quad (3)$$

Efficient use of the retrieval capacity of the CA3 recurrent network requires that its pyramidal units encode as much new information in a pattern of activity as the amount, estimated in Eq.(2), which can be later retrieved; just as efficient use of cash lending machines requires one to deposit in the corresponding account the wherewithal, that one wants later to withdraw. The challenge for afferent inputs is then to prevail during storage over the recurrent connections, which do not impart new contents to a pattern of activity to be stored, but to let them carry out pattern completion at retrieval. Analytical estimates, derived for a simple model with discrete attractor states (Treves and Rolls, 1992), suggest that this challenge can be met by afferent inputs with the characteristics and strength of the mossy fibers (Mori et al, 2007), but not by those, conveyed by the perforant path to CA3, relayed by synapses which are presumably similar to recurrent synapses, but fewer in number. To make full contact with recordings of CA3 activity, however, the argument has to be generalized and applied to models in which CA3 units encode spatial representations, not just discrete attractor states.

4. CA3 as a memory store for spatial ‘charts’

Initially, the quantitative neural network analyses cited above had been formulated in terms of fully connected recurrent architectures and discrete memory states, conceived - in the limit of no fluctuations - as points in a multidimensional space, in which each component corresponds to the firing rate or in general to the activity of one unit (Hopfield, 1982). Thus, the salient spatial character of hippocampal memory correlates was provisionally neglected, to take advantage of the formal models based on discrete attractor states. The very same autoassociator architecture, however, may subserve both the storage of discrete memories as point-like attractor states, or of more complex memories, including continuous attractors, when network dynamics converges to fixed points which are not isolated, but continuously arranged on one or more low-dimensional manifolds embedded in the high-dimensional activity space. Simple examples of continuous attractors are present in models of orientation selectivity by horizontal interaction in visual cortex (Sompolinsky and Shapley, 1997) or of the head direction system (Skaggs et al., 1995). These models do not store information in long term memory, and their fixed points comprise a single (in these particular cases, 1D) manifold. Samsonovich and McNaughton’s multiple chart model (1997) demonstrated instead, in the context of a model for path integration, how one could conceive of fixed points organized in multiple 2D continuous manifolds, each of which maps the animal’s position in a distinct environment.

4.1. Context discrimination and exact localization from CA3 activity

The model conceptually distinguishes two types of information that can both be extracted from CA3 activity: context discrimination (“which chart does the current activity pattern belong to?”) and exact localization within a context (“where in the chart?”). The distinction may be fuzzier for an animal, who navigates among a succession of local contexts without clear boundaries, but it is very helpful at a theoretical level. For both types of information, part may be present, at any given time, in the inputs to CA3, and part may be retrieved there through pattern completion. Path integration, and the predictive coding of future locations, to the extent that they are implemented within CA3, can be conceptualized as continuously sliding activity within the manifold spanned by a single chart; whereas the retrieval of the current context from partial cues corresponds to chart selection, i.e. to pattern completion in a slightly more general sense, in which the attractor is a full chart (Fig.2A).

Within a chart, population activity on the surface of this attractor can therefore be conceived in a form of a “bump”, and it is then the position of the bump that codes for the position of the animal (Fig.2B). On the surface of strictly continuous attractors there is no resistance to drift, hence population activity is only marginally stable: a bump smoothly follows any variation of the input signal.

Individual cells may not be responsive anywhere on a chart, or they may have the characteristic place fields. Sharp (1991) demonstrated, using a simple neural network model with only a generic resemblance to hippocampal networks, how place cell-like responses could arise from the association of different local views, and how the direction-dependent place fields typical of 1D environments may become direction-insensitive in 2D through such associational mechanism. Place fields *per se* arise in networks of different architecture and plasticity, which are made to process spatial information (Burgess and O’Keefe, 1996; Treves et al., 1992; Zipser, 1985), so their mere appearance does not usefully constrain hypothetical hippocampal operations. Their relative

quality, however, can give important insight on these operations, particularly as concerns exact localization within a context. One can assess the quality of place fields directly, e.g. visually from firing rate maps but, for a quantitative appraisal of how encoding quality translates into network function, it is convenient to assess it in terms of localization measures.

As shown by Wilson and McNaughton (1993), one can easily decode, with reasonable accuracy, the position of the rat in a recording box, from a sufficient number of simultaneously recorded units. One constructs firing rate vectors at any time t , as vectors $\{r_i(t)\}$ where each component r_i is the firing rate of one unit; then one may simply average all vectors expressed when the rat is in position (x,y) , to extract a ‘template’ or mean vector $\{r_i(x,y)\}$; and finally by finding the best match for a vector at time t among all templates, one can assign a decoded position to time t . More sophisticated procedures yield somewhat better performance. The very same decoding procedure can be applied to data obtained by simulating simple network models. Moreover, if a rat, real or simulated, has explored several environments, one may measure both the accuracy of decoding its position in each environment, and of decoding which environment the rat is in, both with the same procedure and using multiple sets of templates, one per environment.

4.2. Fractured attractors tend to collapse

The mathematical analysis of attractor states applies, strictly speaking, only to networks with infinitely many cells, whose place fields are arranged with infinite regularity to tile each chart. In this limit case there is no resistance to drift on the surface of a continuous attractor. This also implies that such attractors are susceptible to noise, which would push a bump away from its coding location. In a real system, however, besides noise there are also several sources of *disorder*: even when there is only one map in memory, disorder arises from the finite size of the system, from the partial connectivity, from non-exact assignment of connection weights, etc. It was pointed out by Tsodyks and Sejnowski (1995) that the continuity of a 1D ‘ring’ attractor would be broken already by the moderate irregularity associated with assigning units to code random positions along the ring, instead of equally spaced positions. In statistical physics terms, the free energy is not quite constant along the ring: it has valleys at slightly lower levels where the units happen to be more concentrated; so that the network relaxes its activity in one of the discrete states at the bottom of the valleys, rather than being indifferently movable anywhere on the ring. To stabilize network activity at arbitrary positions requires, for example, supra-linear activation functions (Stringer et al, 2002a) of the type that might be implemented by NMDA-receptor currents (Lisman et al., 1998). The same phenomenon manifests itself in a different fashion if one attempts to create a continuous 1D attractor not on a ring, which has no boundary, but on a linear segment, which has two extreme points. If equally spaced patterns are stored through a standard ‘Hebbian’ learning rule, they need to be encoded with different weights, heavier at the extremes, otherwise the extremes collapse onto the median point (Blumenfeld et al., 2006).

With 2D would-be continuous attractors, the relevant case for the chart model, the effects of inhomogeneity in the distribution of place field centers and of boundary conditions are much stronger than in 1D. Even with networks of considerable size, to establish reasonably smooth 2D attractors, such that activity can settle at almost arbitrary positions, one needs special learning rules (Kali & Dayan, 2000) or external factors such as position-dependent threshold (Stringer et al., 2002b) or gain modulation (Roudi & Treves, 2006). As a result, a bump, without further sensory input, would slowly drift away from its coding location and stabilize in a local or in the global minimum of the free-energy surface (Hamaguchi et al., 2006). This drift may, however, be much

slower than pattern completion. That is, efficient coding of a position is possible on the timescale between the attraction to the surface and before the bump drifts away by a reasonable amount (Fig.2A). Interestingly, since the bump is sustained by the attraction to the surface, the bump does not disintegrate. When more than one chart is stored, disorder is increased. Attractor ‘collapse’ is then accelerated slightly at each extra chart added. Close to storage capacity, suddenly all charts become unstable, that is, bumps drift away very fast, pattern completion may occur to ‘wrong’ charts and, finally, no convergence is observed at all. Interestingly, such malfunctions may occur independently for different locations in an environment, thus part of a chart may be recalled perfectly, while another part may be lost. These effects may be very difficult to assess in the real hippocampus, which normally operates under the influence of afferent inputs. Unless one designs ad hoc experiments to let place cell activity sustain itself (Jarosiewicz and Skaggs, 2004), it is much easier to understand the issue by studying retrieval dynamics through network simulations.

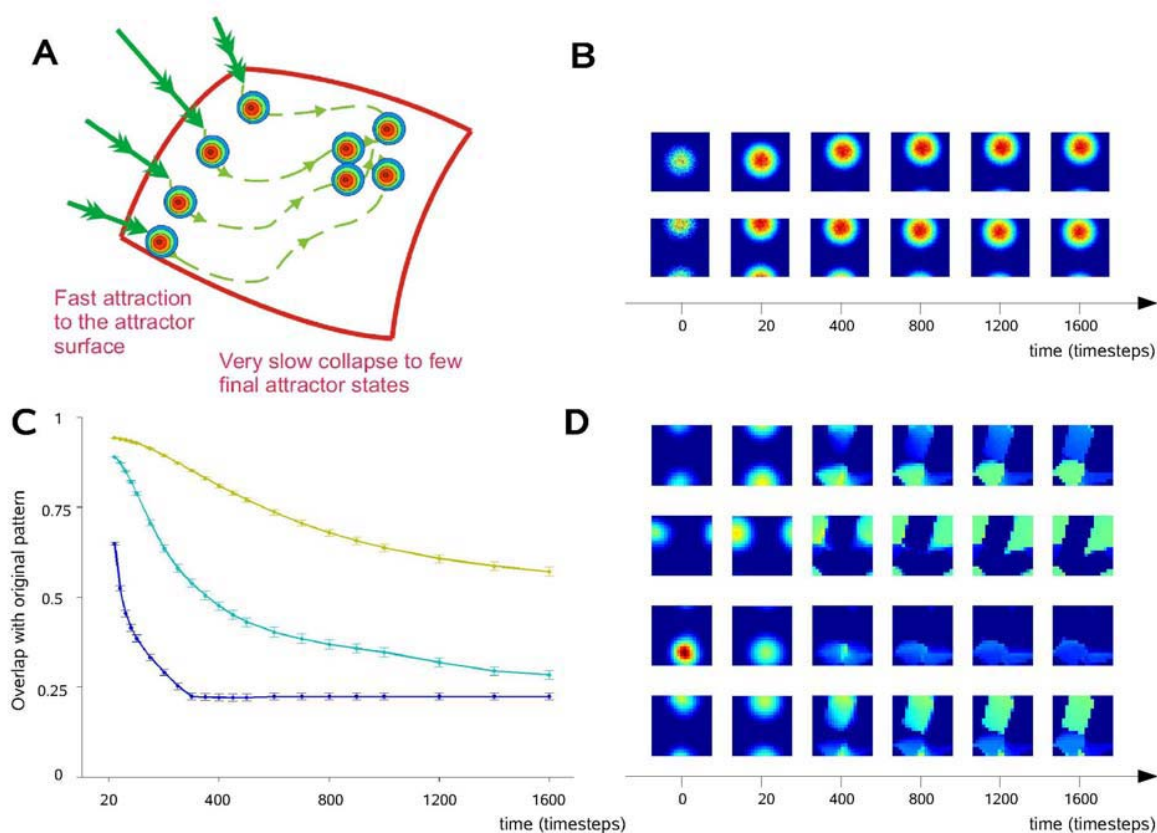


Figure 2. Continuous attractors in theory and in practice. **A)** Continuous attractors allow for rapid pattern completion, similar to that of discrete ones. As a result, a bump of activity forms on an attractor surface, to codes for position in the corresponding environment. As shown by the green arrows, the metrics of real space is conserved during pattern completion: similar inputs cause bump formation on nearby locations on the attractor. Theoretically continuous attractors in practice contain privileged spots, which in the absence of ongoing input act as discrete attractors and ‘attract’ bumps from their vicinity. In a small system, this effect can take the form of a complete attractor collapse. The timescale of pattern completion and that of drift or collapse can however be separated. **B)** Sample network population responses, coding for two different locations on the surface of a torus (periodic boundary conditions). At time step 0 we see the summation of afferent inputs only. Pattern completion is effectively accomplished between 0 and 20 timesteps (roughly

250 msec of real time), leading to bumps at the exact location of the input. Several hundred timesteps later, the two example bumps, coding initially for two locations far apart, drift into the same position. **C)** Drift speed and distance in three networks differing only in size (blue: $N=400$ units, cyan: $N=1600$, green: $N=4900$ units; feed-forward connections per unit: $C_{\text{aff}}=0.08N$; recurrent connections per unit: $C_{\text{rc}}=0.24N$). The graph shows the average overlap between ongoing population activity and the original activation caused by the sensory input. In larger networks, the drift is slower and covers less distance, as shown by the large final overlap value, indicating more privileged spots than in smaller networks, i.e. a better approximation of the continuous attractor limit. **D)** Place-fields of 4 randomly chosen CA3 units. At timestep 0 (sensory input alone) and 20 (after pattern completion, but before attractor collapse) place-fields are generally round in shape, and present a smooth firing rate map. As population bumps reach their final state on the attractor surface (panel **B**), place-fields lose their smooth shape and smoothness, and tend to fire at the same rate in large portions of the environment, with sharp boundaries, which are common to many other units. The borders of such regions define the basin of attraction of a final (collapsed) state.

Simulations of a medium-size self-organizing hippocampal model demonstrate that, if ad hoc mechanisms are not introduced, rather than 2D charts one obtains ‘wrinkled’ attractors, where activity can settle in only one or a few positions (Papp and Treves, 2006). In these simulations, a virtual rat explores one or several virtual environment, realized as square boxes with toroidal boundary conditions (the 2D equivalent of a ring), while a population of CA3 units develops representations of each environment (Treves, 2004). If the unsupervised self-organization of a chart is produced solely by a ‘Hebbian’ learning rule, place-cell-like responses emerge spontaneously in CA3, but many units end up having one of a few available fields, or no field: the corresponding attractor has collapsed. These are finite size effects, and they can be alleviated by scaling up the number of units and the connectivity in the simulated network (Fig.2C); by constraining activity to be less sparse than experimentally observed, e.g. $a_{\text{CA3}} \approx 0.2$ rather than $a_{\text{CA3}} \approx 0.03$; and by keeping inhomogeneities to a minimum; but even with thousands of CA3 units the resulting attractors remain far from the ideal charts realized in mathematical models (see also Fig.3 below).

This ‘attractor collapse’ scenario is particularly relevant for threshold-linear units with no saturation, which may model neuronal activity far from saturation. Simulations with effectively binary units, instead, with activity determined by a sigmoidal activation between 0 and 1, yield bumps of activity which are much more stable, and show virtually no drift in the absence of inputs (Papp and Treves, unpublished observations). In the sigmoidal model, however, most units in a stable state have activity close to 1 or 0 (that is, maximum or zero activity), and such an activity profile is incapable to describe real place-fields, where activity changes gradually over space (see also Roudi and Treves, 2006).

4.3. How to iron out two-dimensional charts?

In the multiple chart model, exploration of a new environment must lead to the formation of a new chart. A number of questions then arise: how is the new chart laid out? *ab initio* or using some pre-wired connectivity? are inputs from a population of grid cells useful? can the process be aided by the dentate, as for discrete attractors? is the final chart as smooth as the ideal concept of a continuous attractor implies? how many charts can possibly accumulate in a single recurrent network? These are very much issues of current research, and in the following we only discuss some of the results already obtained within the modeling approach.

In relation to the last question, discrete memory states models point at *sparsity* as the crucial representational parameter that influences memory capacity (besides the anatomical one of the connectivity per unit), as illustrated by Eq.(1) above. Hippocampal space-related activity lends itself to the measurement of sparsity values, e.g. defined as the square of the mean firing rate across a population of units at a given instant, divided by the mean of the square firing rate (Treves and Rolls, 1991). Applying such measure to individual cells recorded in a freely foraging task (courtesy of Jill and Stefan Leutgeb) one obtains values for CA3 in the sparse range, $a_{CA3} \approx 0.02-0.06$ (Papp and Treves, 2007). The storage capacity of a multi-chart recurrent autoassociator was analyzed by Battaglia and Treves (1998), who extracted a simple ‘rule-of-the-thumb’ for assessing the memory load of a chart. A chart which maps a finite environment onto the activity of place-cell-like units is equivalent, capacity-wise, to as many discrete attractor states as there are locations, in the environment, for which the activity vectors are pair-wise decorrelated. If the 2D environment is represented by place-cell-like units, which are quiescent outside their place field, the decorrelation radius is roughly the radius of the typical place field, which is itself proportional to the linear size of the environment times the square root of the sparsity of the neural representation. Thus, if some dozen typical CA3 fields, say, ‘fit’, once properly juxtaposed, in a typical rat recording box, the memory load of the chart corresponding to that box is roughly equivalent to a dozen discrete memories of equal sparsity. The number of such charts, or distinct environments of the size of a typical recording box, that can be held simultaneously in the network, is theoretically limited by the critical value

$$p_{charts} \approx 0.1 C / \ln(1/a) \quad (4)$$

(see Figs.1, 2 in Battaglia and Treves, 1998). This is still a huge number, of the order of several hundreds, given the recurrent connectivity of the CA3 network in rodents (Amaral et al, 1990). The apparent paradox that fewer charts can be stored if they are sparser (a lower a parameter makes the denominator larger) can be understood by considering that sparser activity, in a large net, leads to better spatial resolution, and hence requires more discrete fixed points attractors to cover, as effectively smaller tiles, the whole environment. This chart capacity again respects the associative memory theoretical upper bound alluded to above, in that the maximum amount of information that can be retrieved per synapse is about 0.15 bits, as shown in Fig.5 of Battaglia and Treves (1998). It remains to be seen, however, whether in practice the upper bound can be approached.

4.4. The challenge of confronting grid cells

In the simulations of Fig.2, the inputs to CA3 units were from other units that themselves had place-field-like properties. One may expect that introducing a model of mEC grid cells as an input station to CA3 would help produce better continuous attractors, given the regularity of experimentally observed grid fields, and the accurate localization they allow (Fyhn et al., 2004; Hafting et al., 2005). Taken as an isolated mEC-CA3 feedforward network, the transformation from grid cells to place cells is similar to a 2D Fourier transform, requiring only a few grid cells of different spatial frequencies but otherwise straightforward – if the grid fields are precisely aligned and the feed-forward connection weights are assigned pre-calculated values (Solstad et al, 2006). In fact, retrieval dynamics proceed smoothly from mEC-like grid inputs, if the CA3 representation has earlier been established with the aid of strong DG inputs as in Fig.3.

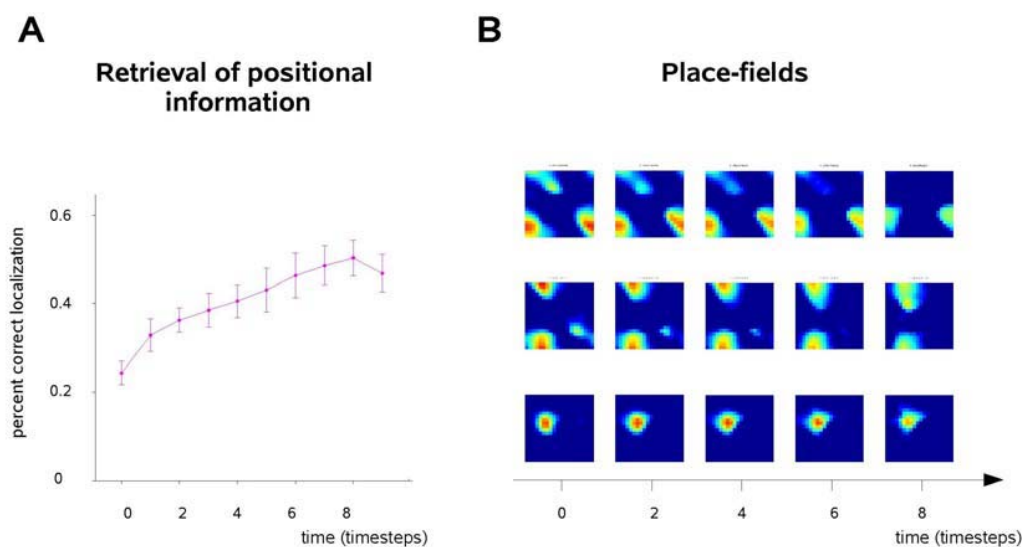


Figure 3. Characterization of the dynamics of recall. **A)** Percent correct localization increases as the information provided by afferent inputs is reverberated in the recurrent network for several iterations (roughly corresponding to 100 msec). Note that the afferent cue is partial, comprising only 20% of the activity pattern in the model input layer. The localization of the virtual rat, afforded by population activity at timestep 0 (summation of afferent inputs only) is already better than before learning (cp. Fig 4A), yet it is further improved as activity is iterated in the recurrent network (timesteps 1-9) and thus memory is recalled from the model CA3 network. **B)** Development of place-fields for three representative CA3 units. Iteration through the recurrent collaterals also decreases the number of resulting place fields, as small peaks gradually vanish. .

If the weights have to emerge from a self-organizing process, even the most effective algorithm, a variant of Independent Component Analysis subject to a sparsity constraint, was found to produce rather implausible place fields even when summing from 100 grid units (Franzius et al., 2007). The main difficulty, for the algorithm which has to structure the weights appropriately, is how to suppress the periodicity inherent in the grid fields, to lead to a single-peaked place field: a competitive learning algorithm may only reduce the mean number of peaks of the output units (Rolls et al., 2006). The difficulty is alleviated when the algorithm has available the input of an increasing number of grid units, of the order of thousands. Even in this situation, however, the self-organization of the feedforward weights is a delicate process, easily disrupted by the concurrent effect of any recurrent weights. In the presence of massively recurrent connectivity, as in CA3, no one has been able to demonstrate, so far, how direct grid cell inputs could be the driving force for the establishment of a new chart – all the more unlikely, if the recurrent connections are structured by the earlier storage of other charts, which interfere with the new one. It is reasonable to expect

that combining grid cell inputs with non periodic spatial inputs, representing e.g. the activity of cells in lateral entorhinal cortex (IEC) should help with the formation of place fields; but evidence about the spatial correlates of IEC activity is too fragmentary yet to effectively constrain computational models.

The dentate gyrus as a chart preprocessor

The recent characterization of multi-peaked place fields in the dentate gyrus (Leutgeb et al., 2007) provides a break-through for the development of models of chart formation in CA3. The observed fields resemble those produced by self-organization of feedforward inputs from grid-like-units (Franzius et al., 2007; Rolls et al., 2006), redefining the feedforward models as relevant for studying granule cell activity and its changes after different manipulations. At the same time, the question arises as to whether the observed fields in the dentate can serve as effective *inputs* to drive the establishment of new spatial representations in CA3. What computational models can attempt to resolve, is whether such inputs can, given an appropriate synaptic structure with MF characteristics, overcome the unwelcome effects of the prevailing recurrent CA3 circuitry, including the ‘wrinkling’ and eventual collapse of individual charts and the coalescence of new charts with previous ones.

With simulations, one can investigate the emergence of new charts in models with contrasting architecture, e.g. with and without a layer modeling the dentate gyrus. Fig.4 shows the results of simplified simulations, in which the same network was trained as a virtual rat explored a new environment. The DG layer, if present, is modeled here with single-field granule units and one-to-one detonator synapses to CA3 units, extrapolating somewhat current experimental evidence on the strength and indirect suppressive effect of MF activity (Mori et al, 2007), and is active only during training. At testing, DG is turned off, and CA3 cells, activated by the model perforant path and under the influence of recurrent collaterals, both modified during training, show scattered spatial responses before training, which self-organize into smoother fields after training (Fig.4).

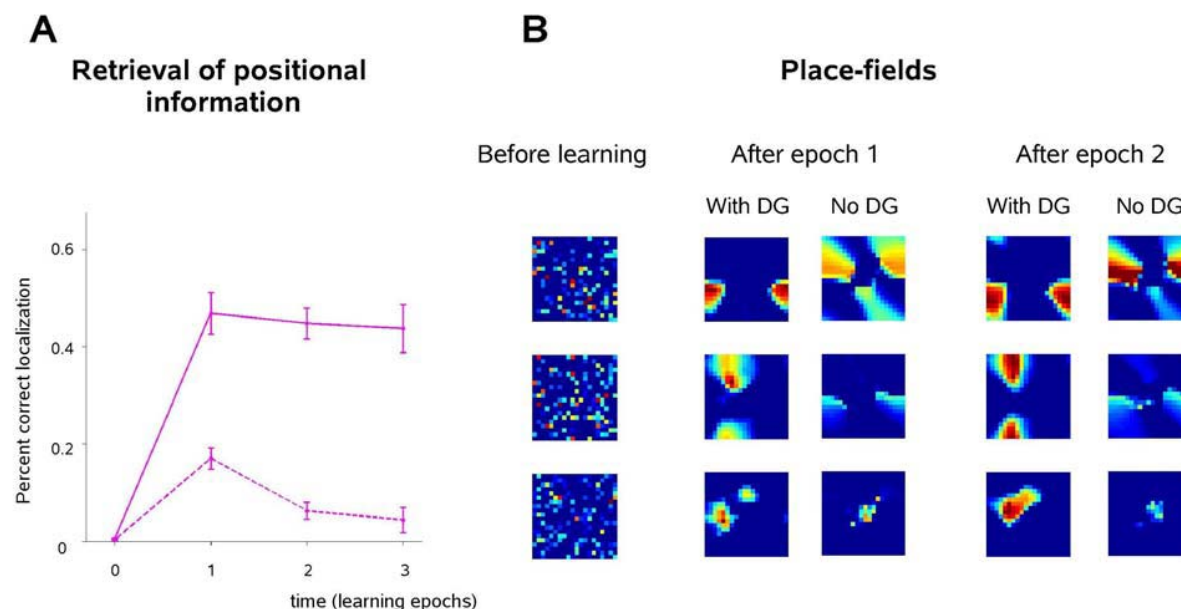


Figure 4. Development of CA3 place fields with and without a model DG array. **A)** Training the model in new environment with DG inputs (solid line) increases percent correct localization relative to training without DG (dashed line). Learning for one epoch (corresponding to ~10 minutes for a real rat) increases in both cases the accuracy afforded by decoding the entire population of 4900 CA3 units; further training without DG, however, is detrimental. **B)** Three examples of firing rate maps in the model CA3 population are shown before and after training. Without DG, CA3 fields do not develop single peaks (note that single peaks occurring across the toroidal boundaries appear as double when the torus is displayed as a square box). Note also the refinement in the fields with the second epoch of training.

The DG teaching input modifies, in this simulation, both the perforant path and the recurrent collateral weights, and both contribute to the response properties of CA3 units. As the perforant path, which at testing relays only a partial cue, is made to gradually fade over 10 x 12.5msec iterations, towards the end of the iteration cycle the cue is largely completed by the collateral effect (Fig.3).

Note that in these preliminary simulations, MF synapses are one-to-one, and DG units have single-peaked fields, unlike those observed by Leutgeb et al. (2007). Still, DG units have spatially continuous firing rate maps, which is the crucial element to generate continuous representations of space, and which is beyond the scope of the original Treves and Rolls (1992) argument. We suggest that a thorough quantitative analysis of information storage in a model CA3 network, operating

with and without dentate gyrus, is needed, to assess again any information theoretic advantage in forming new representations, this time in the form of charts of place-cell-like units. Since, unfortunately, the very 2D nature of charts makes a simplified mathematical analysis of information storage like the one in Treves and Rolls (1992) difficult to work out, computer simulations offer at present a practical approach to this issue.

In our simulations, considering one time step to be 12.5 msec, attractors tend to collapse over several seconds (Fig.2). Some deterioration in the accuracy of localization may appear already within the equivalent of 100-150 msec (Fig.3). In rodents, theta-oscillations pace activity over a similar time scale, and it might just be that the collapse phenomenon, because of the typically sustained afferent inputs (e.g. from mEC units) and because of the limited time for convergence, is not so relevant to the real hippocampus. Theta oscillations are however characteristic of exploratory behaviour, usually associated with memory encoding in the CA3 network, not retrieval; and the correspondence between real hippocampal dynamics in different behavioural conditions and the simplified dynamics of simulated models must be elaborated much better, probably with more complicated models, before dismissing the implications of attractor collapse.

At any rate, the collapse of attractors is intimately related to the recurrent nature of the CA3 network. If one makes partial 'lesions' to the recurrent connections, attractors become smoother. At the same time, however, one loses the benefits of the collateral effect (Fig.3). A possible solution to this problem, seemingly 'invented' by mammals as mentioned above, might be the insertion of a population of units right after CA3, as a post-processor, i.e. the CA1 feedforward network. In simulations to be reported elsewhere (Papp and Treves, 2007) we observe that CA1 processing may smooth the representation retrieved from CA3, thereby increasing its spatial information content (Treves, 2004).

5. Morphing memories and morphing environments

Do discrete boundaries between attractor states survive manipulations in which external correlates are intentionally interpolated, i.e. *morphed*, between pre-established representations? In simulations of the morphing experiments of Wills et al (2005) and Leutgeb et al (2005b), we aimed first to test the idea that correlations between encoded patterns may lead to a linearization of responses along the morphing sequence, as observed by Leutgeb et al in contrast to Wills et al (Papp and Treves, 2005). Second, we aimed to assess the relevance, in the model, of the ratio of recurrent to afferent connections in driving recall dynamics. The first issue is relevant, as a crucial difference between the two experiments involved the degree of orthogonalization of the spatial representations originally established (whether they led to complete remapping or not, see Leutgeb et al, 2005a). The second issue is relevant, in relation to the hypothesized effect of ACh, even if this is not explicitly modeled in the simulations. We proceed by first using simple, discrete patterns and then 2D charts.

5.1. Morphing discrete patterns

The model simulated in this and the following section includes 2 layers: ‘entorhinal cortex’, serving as the input, and CA3, the actual attractor network. Each of them is comprised of 3600 threshold-linear units (Treves, 1991), arranged on a 60 by 60 grid with periodic boundary conditions. Connectivity is partial: 240 units randomly chosen in the EC array connect to any CA3 unit, which also receives from a random assortment of 400 fellow CA3 units. Discrete patterns or spatial charts are not the product of extensive training, but rather they are ‘assigned’ to the CA3 network, in a crude model of one-shot associative learning, as in the Hopfield (1982) model. A simple Hebbian learning rule is used, as in Treves (2004). Patterns are set also in the EC layer, and associated with those in CA3 with the same one-shot learning rule. During testing, a full input pattern is presented in EC, and the network is updated for several hundred timesteps.

Different input patterns in EC, e.g. A and B (representing for example the circle and square box used in the experiments), were created by scrambling EC units among themselves. Morphs were then set up by setting the activity of a fraction of EC units to their activity in one pattern and the remaining cells to their activity in another pattern. Changes from one full pattern to the first morph, between the morphs, and between the last morph and the second full pattern was set to be equispaced. Correlated initial patterns were created with an incomplete scrambling, thus a number of EC units had in this case the same representation for A and B, whereas the other were allowed to have different representation.

The storage capacity of the network was first established and was found to be around 60 independent patterns. To stay well below storage capacity, only 4 discrete patterns were used in the morphing simulation. Results were mapped as a function of two parameters, the recurrent-to-afferent strength and the degree of correlation between A and B. In Fig.5A four representative examples are shown for these two parameters. As in Leutgeb et al, 5 intermediate morphs were used, denoted as AAAB, AAB, AB, ABB and ABBB respectively. The reference case, with ‘normal’ recurrent strength and no correlation between A and B, is shown in panel 3, and corresponds to a sudden transition between the two discrete attractors, for A and for B. Decreasing the recurrent strength by a factor of 10 (Fig 5A, panel 1), the transition becomes linear. A very slight trace of pattern completion can be still detected (as overlap scores for AAAB and AAB

above the line connecting those for A and B), showing that although the strength of recurrent weights is decreased they are still not completely ineffective. Introducing a correlation of 0.3 (that is, 30% of EC units are kept to have the same representation in A and B) in the case with strong recurrent weights (panel 4) strongly increases the overlap between the representation of A and B, and the sudden transition is lost, replaced by an accelerating decrease in overlap scores over the morphing sequence. Thus introducing correlations between attractors has a similar effect to reduced recurrent ratios. Finally, again decreasing the effect of recurrent connections while also keeping the elevated correlation (Fig 5A, panel 2), a smooth linear transformation appears between similar attractor states, with overlap scores that remain high even between A and B.

Further simulations show that an even higher degree of correlation leads to a complete collapse of the memory states for A and B. The collapse, much like the collapse within a chart, described in the previous section, results from the action of recurrent collateral connections, as shown by the finding that removing such connections altogether the representations of A and B remain different, and the transition over the morphing sequence linear. Like the collapse within a chart, it will not manifest itself when the network is sufficiently driven, that is steered, by afferent inputs. Note that recurrent weights in these simulations reflect the storage of only the extreme patterns, unlike the model analysed by Blumenfeldt et al (2006).

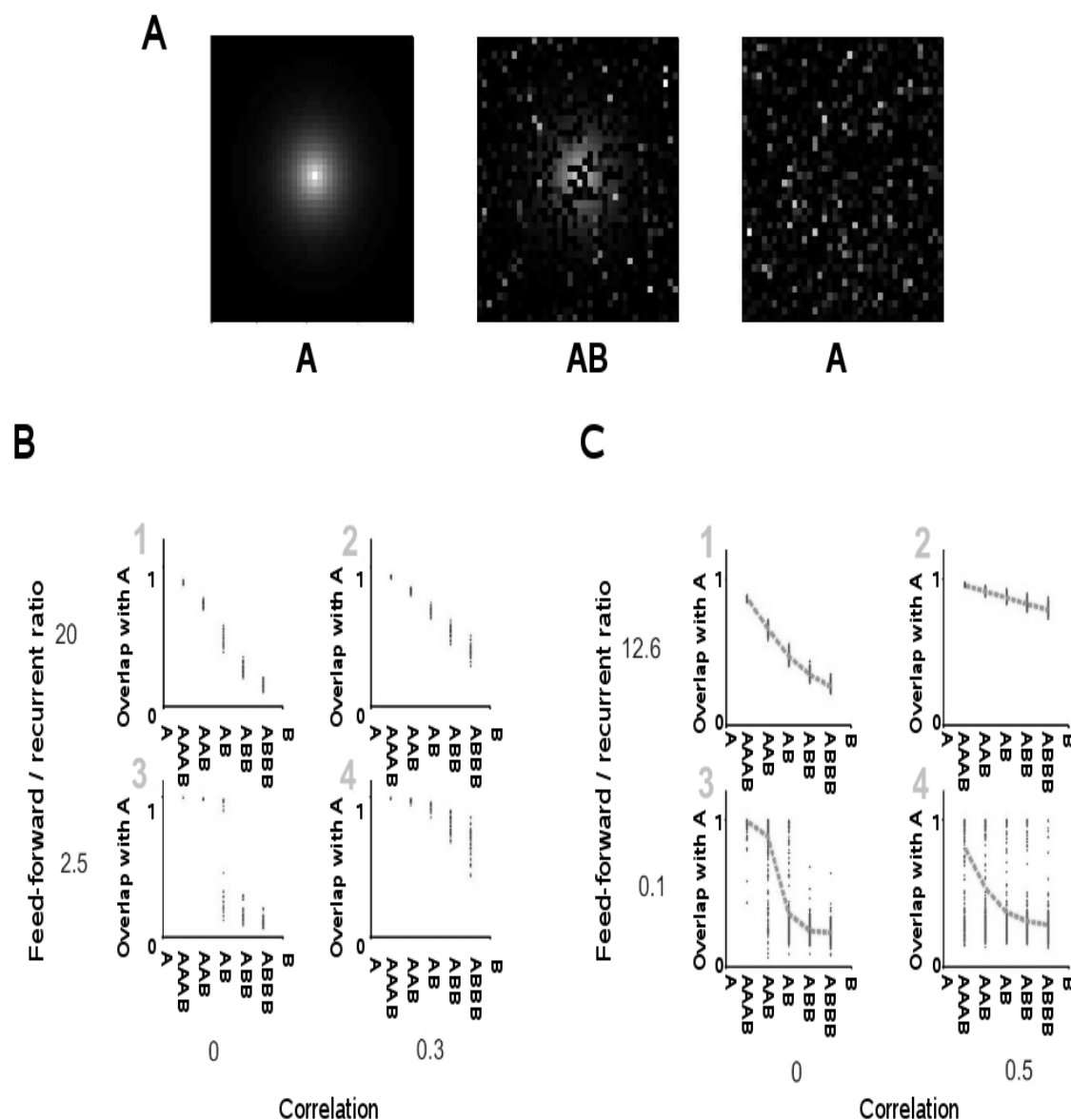


Figure 5. Simulation of the morphing experiments. **A)** Morphing environments. In our simulations, the input array is arranged on the nodes of a (virtual) square grid, and it encodes a representation of space in terms of self-similar single-peak place fields – so the center of environment A is represented by a bump of activity in the center the grid. The position of the input units on the grid is reshuffled in environment B while, to model gradual changes from one environment to the other, a partial reshuffle of half the units encodes environment AB. The other 4 morphed environments were obtained by shuffling 1/6, 2/6, 4/6 and 5/6 of the units. **B)** Morphing discrete input patterns with different feed-forward / recurrent ratio and different levels of correlation between patterns A and B. Relatively suppressed feed-forward weights and no correlation (panel B.3) lead to a sudden transition during morphing. Stronger sensory input linearize the transition (B.1). Correlations, along with suppressed feed-forward weights both linearize the transition and decrease the distance between the representation of environments A and B (B.4). The presence of strong inputs and of correlations between A and B results in a linear transition with morphing, resembling the results of

Leutgeb et al 2005b. Dots: overlap between population response to a morphed pattern, and to pattern A, across repeated simulations. C) Similar results were obtained with morphing charts instead of discrete patterns. Weak sensory inputs and the absence of correlation between chart A and B lead to a sudden transition between A and B (C.3). Correlations are found to effectively linearize the transition during morphing (C.4), as well as an increased input strength (C.1). Strong inputs together with correlations lead to linearization and to less distance between the representation of A and B. Dots: overlap of population responses at corresponding locations in environment A and in a morphed environment. Dashed line: the mean overlap between environment A and the morphs.

5.2. Morphing charts

Simulations with 2D representations were conducted with the same procedure as for the discrete attractor case, however in a more complex situation, where spatial charts are stored in the network instead of discrete patterns. To attain the larger capacity needed to store charts, a network comprised of 4900 CA3 units arranged on a 70 by 70 grid was used, with 400 afferent inputs and 1200 recurrent connections each. The network was found to store up to 8 different charts before a clear capacity breakdown, and 4 charts were used in the morphing simulations, to stay below storage capacity. Using periodic boundary conditions, each chart was arranged on a torus, and a scrambling procedure was used on the torus to obtain morphed environments. Connection weights were precalculated as a sum of contributions from the different charts stored, in each decreasing as a function of the distance between the position of the pre- and post-synaptic unit in the chart. Parameters were chosen through an extensive search, to find conditions that would allow the slowest drift of population activity, i.e. the best stability in the chart.

In order to set up the chart, the centers of place fields in EC and in CA3 were assigned prior to setting up the weights. Each unit, both in EC and in CA3, was simply set to have a field in all of the environments (decreasing the number of units coding for an environment decreases the stability of the resulting memory, as measured by percent correct localization). Note therefore that, in these simulations, for simplicity each EC unit was set to have one peak, unlike the simulations with grid units described earlier. The morphing effect discussed here is an effect endogenous to the recurrent CA3 network, and thus likely to emerge irrespective of the input pattern in EC.

In a case of recall with strong recurrences and no correlation (Fig.5B, panel 3), the mean overlap for morphs AAAB and AAB shows pattern completion towards morph A, whereas for ABB and ABBB the CA3 units tend to converge towards the representation of B. In middle morph, some locations are coded as more similar to those in A, while others to those in B. In morphs AAB and ABB, in fact, there are already some locations acquiring the representation of the extreme shape that is more distant in terms of the morphing sequence. Still, the representation in the majority of spatial locations switches abruptly between the morphs AAB and AB. Individual place fields mostly follow the same dynamics, with most often global remapping between morphs AAB and AB. Decreasing the strength of recurrent connections leads again to a linearization of the transition (panel 1). There is now little sign of pattern completion, however, unlike that seen with discrete attractors. Place fields change smoothly over the whole morphing sequence (data not shown). Again, that implies that a network predominantly driven by external inputs, e.g. from mEC, is unlikely to show the features characterizing attractor dynamics, be they advantageous, like pattern completion, or disadvantageous, like attractor collapse.

Introducing correlation (panel 4) between the neural representations of A and B formed by CA3 also abolishes the sudden transition. As shown by the scatter of individual dots in panel 4, individual locations change their population code quasi independently over the morphing sequence, similar to experimental results by Leutgeb et al (2005b). In the model, however, no unit acquired a place field only in some intermediate morphs, because in the model only the extremes A and B were used in setting connection weights, and in A and B each unit had a place field. Decreasing the relative strength of the recurrent connections further linearizes the morphing curve (panel 2), with the overlap between A and B remaining quite elevated, again consistent with the findings of Leutgeb et al. Place-field again changed smoothly.

A conclusion from these morphing simulations is that correlations in the environment, as well as between discrete stimuli represented in a strongly recurrent network, lead to attractor collapse, a different manifestation of the same phenomenon apparent with the storage of continuous attractors, discussed above. The drift to discrete attractor locations within a would-be continuous attracting manifold is augmented, when multiple manifolds are stored, by their collapsing onto each other. Both phenomena remain latent, when the network is, broadly speaking, steered by afferent inputs: a weakly recurrent network is less subject to attractor collapse, but then population activity tends to track ongoing inputs anyway, resulting in essentially linear transitions along a morphing sequence. The self-organized formation of a single continuous attractor and the retrieval of pre-established representations from morphed inputs therefore both face the same challenge, in the recurrent CA3 network, of separating patterns that are separated in the inputs. When multiple spatial charts are to be self-organized from scratch, the difficulty for the CA3 network is redoubled, and further computational work may help assess to what extent the Dentate Gyrus may help meet the challenge.

6. Conclusions

As originally proposed by David Marr (1971), and then by McNaughton and Morris (1987) and by Rolls (1989), the general theory of autoassociative memory networks provides a most useful model with which to gain insight into the information processing operations performed by the hippocampal CA3 network. A comparison with the organization of the medial telencephalon in reptiles and birds in fact indicates in CA3, in contrast to other hippocampal subfields in mammals, the most preserved traits of the archetypical organization of primitive cortex. This suggests that in functional terms, too, the operation of CA3 may be the straightforward evolutionary derivative of an archetypical storage site for complex memories, involving relations between disparate elements, and capable of autoassociative retrieval from arbitrary partial cues. The mathematical analysis introduced for the Hopfield (1982) model by Amit, Gutfreund and Sompolinsky (1987), as well as computer simulations on more realistic versions of the model, allow for a quantitative appraisal of how efficiently CA3 may function as an autoassociative memory network, in particular in terms of the crucial constraint of storage capacity. These approaches reveal that acquiring new memory representations in general, and acquiring spatial representations in particular, even more when derived from grid units, pose hard challenges to CA3. Meeting these challenges may have required, in mammals, the evolution of additional specialized circuitry, including at least the dentate gyrus as a preprocessing stage to CA3. It remains to be explored whether other, distinct evolutionary adaptations may have met similar challenges in the avian brain, given the intriguing suggestion that birds, unlike reptiles, appear to have evolved similar slow-wave sleep dynamics as mammals (Rattenborg, 2007).

References

- Amaral, D.G. and Witter, M.P..1989 The three-dimensional organization of the hippocampal formation: a review of anatomical data. *Neuroscience*. **31**:571-91.
- Amaral DG, Ishizuka N. and Claiborne B. 1990. Neurons, numbers and the hippocampal network. *Prog Brain Res* **83**:1-11.
- Amit, D.J. 1989. *Modeling Brain Function*. Cambridge, UK: CambridgeUP.
- Amit, D.J., Gutfreund, H., Sompolinsky, H. 1987. Statistical mechanics of neural networks near saturation. *Ann Phys (N.Y.)* **173**:30-67
- Andersen, P. and Loyning, Y. 1962. Interaction of various afferents on CA1 neurons and dentate granule cells. *Colloq Int CNRS* **107**:23– 45.
- Ariens-Kapper C.U, Huber G C and Crosby E C 1936 *The comparative anatomy of the nervous system of vertebrates, including man*. New York: Hafner Publishing
- Atoji, Y. and Wild, J.M. 2004. Fiber connections of the hippocampal formation and septum and subdivisions of the hippocampal formation in the pigeon as revealed by tract tracing and kainic acid lesions. *J. Comp. Neurol.* **475**:426-61.
- Atoji, Y. and Wild, J.M. 2006. Anatomy of the avian hippocampal formation. *Rev. Neurosci.* **17**:3-15.
- Bingman, V.P. and Sharp, P.E. 2006. Neuronal implementation of hippocampal-mediated spatial behaviour: a comparative evolutionary perspective. *Behav. Cogn. Neurosci. Rev.* **5**:80-91.
- Blumenfeld, B., Preminger, S., Sagi, D. and Tsodyks, M. 2006. Dynamics of memory representations in networks with novelty-facilitated synaptic plasticity. *Neuron*, **52**, 383-394 (2006).
- Battaglia, F.P. and Treves, A. 1998. Attractor neural networks storing multiple space representations: a model for hippocampal place fields. *Phys Rev E* **58**:7738-7753.
- Bliss, T.V. and Lomo, T. 1973. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J Physiol (Lond)* **232**:331-56.
- Braitenberg V and Schüz A 1991. *Anatomy of the Cortex*. Berlin, Germany: Springer Verlag.
- Burgess, N. and O'Keefe, J. 1996. Neuronal computations underlying the firing of place cells and their role in navigation. *Hippocampus* **6**:749-62.
- Clayton, N.S., Bussey, T.J. and Dickinson, A. 2003. Can animals recall the past and plan for the future? *Nat. Rev. Neurosci.* **4**:685-91.

Colombo, M and Broadbent, N. 2000. Is the avian hippocampus a functional homologue of the mammalian hippocampus? *Neurosci. and biobehav. reviews* **24**:465-484.

Dubbeldam, J.L. 1998. The neural substrate for 'learned' and 'nonlearned' activities in birds: a discussion of the organization of bulbar reticular premotor systems with side-lights on the mammalian situation. *Acta Anat (Basel)*. **163**:157-72.

Erichsen, J.T. and Bingman, V.P., Krebs, J.R. 1991. The distribution of neuropeptides in the dorsomedial telencephalon of the pigeon (*Columba livia*): a basis for regional subdivisions. *J Comp Neurol* **314**: 478-492

Faber, H., Braun, K., Zuschratter, W. and Scheich, H. 1989. System-specific distribution of zinc in the chick brain. A light- and electron-microscopic study using the Timm method. *Cell Tissue Res.* **258**:247-57.

Franzius, M., Vollgraf, R. and Wiskott, L. 2007. From grids to places. *J. Comput. Neurosci.* **22**:297-9

Fyhn, M., Molden, S., Witter, M.P., Moser, E.I., Moser, M-B. 2004 Spatial representation in the entorhinal cortex. *Science* **305**:1258-64.

Gardner-Medwin, A.R. 1976. The recall of events through the learning of associations between their parts. *Proc R Soc London B* **194**:375-402.

Gloor, P. 1997 *The Temporal Lobe and Limbic System* (Oxford University Press, New York).

Hafting, T., Fyhn, M., Molden, S., Moser, M-B. and Moser, E.I. (2005) Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**:801-806.

Hamaguchi, K. and Hatchett, J.P.L. 2006. Analytic solution of neural network with disordered lateral inhibition. *Phys. Rev. E* **73**: 051104

Hasselmo, M.E. and Schnell, E. 1994. Laminar selectivity of the cholinergic suppression of synaptic transmission in rat hippocampal region CA1: Computational modeling and brain slice physiology. *J Neurosci.* **14**:3898-3914.

Hasselmo, M.E., Schnell, E. and Barkai, E. 1995. Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *J. Neurosci.* **15**:5249-5262.

Hasselmo, M.E., Wyble, B. and Wallenstein, G. 1996. Encoding and retrieval of episodic memories: role of cholinergic and GABAergic modulation in the hippocampus. *Hippocampus* **6**:693-708.

Healy, S.D., de Kort, S.R. and Clayton, N.S. 2003. The hippocampus, spatial memory and food hoarding: a puzzle revisited. *Trends Ecol. Evol.* **20**:17-22.

- Hoogland, P.V. and Vermeulen-Vanderzee, E. 1993. Medial cortex of the lizard *Gekko gecko*: a hodological study with emphasis on regional specialization. *J Comp Neurol* **331**: 326-338.
- Hopfield, J.J. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci USA* **79**:2554-2558.
- Hough, G.E. and Bingman, V.P. 2004. Spatial response properties of homing pigeon hippocampal neurons: correlations with goal locations, movement between goals, and environmental context in a radial-arm arena *J Comp Physiol A* **190**:1047-62
- Jarosiewicz, B. and Skaggs, W.E. 2004. Hippocampal place cells are not controlled by visual input during the small irregular activity state in the rat. *J. Neurosci.* **24**: 5070-7.
- Kahn, M.C., Hough, G.E., ten Eyck, G.R. and Bingman, V.P.. 2003. Internal connectivity of the homing pigeon (*Columba livia*) hippocampal formation: an anterograde and retrograde tracer study. *J Comp Neurol* **459**: 127-141.
- Káli, Sz. and Dayan, P. 2000. The involvement of recurrent connections in area CA3 in establishing the properties of place fields: A model. *J. Neurosci.*, **20**:7463-7477.
- Karten, H.J. and Hodos W. 1967. *A stereotaxic atlas of the brain of the pigeon (Columba livia)*. Baltimore: John Hopkins University Press.
- Kesner RP, Gilbert PE and Lee I (2002) Subregional analysis of hippocampal function in the rat. In: *Neuropsychology of Memory* (Squire LR, Schacter DL eds), 3rd Ed., pp395-411. New York, USA: Guilford Press.
- Kobayashi, K. and Suzuki, H. 2006. Dopamine selectively potentiates hippocampal mossy fiber to CA3 synaptic transmission. *Neuropharmacology*, **52**: 552-61.
- Krebs, JR, Sherry, DF, Healy, SD, Perry, H and Vaccarino, AL. 1989. Hippocampal specialization of food-storing birds. *Proc. Natl. Acad. Sci.U.S.A.* **86**:1388-1392.
- Künzle, H. 2004. The hippocampal continuation (indusium griseum): its connectivity in the hedgehog tenrec and its status within the hippocampal formation of higher vertebrates. *Anat Embryol* **208**: 183-213.
- Lassalle, J-M., Bataille, T. and Halley, H. 2000. Reversible inactivation of the hippocampal mossy fiber synapses in mice impairs spatial learning, but neither consolidation nor memory retrieval, in the Morris navigation task. *Neurobiol Learn Mem* **73**:243-257.
- Lee, I. and Kesner, R.P. 2004. Encoding versus retrieval of spatial memory: double dissociation between the dentate gyrus and the perforant path inputs into CA3 in the dorsal hippocampus. *Hippocampus*. **14**:66-76.

- Leutgeb S, Leutgeb JK, Barnes CA, Moser EI, McNaughton BL and Moser M-B (2005a) Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science* 309:619-23.
- Leutgeb, J.K., Leutgeb, S., Treves, A., Meyer, R., Barnes, C.A., McNaughton, B.L., Moser, M-B., and Moser, E.I. 2005b. Progressive transformation of hippocampal neuronal representations in “morphed” environments. *Neuron* **48**:345-58
- Leutgeb, J.K., Leutgeb, S., Moser, M-B. and Moser, E.I. 2007. Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science* **315**:961-6.
- Lisman, J. E., Fellous, J.M. and Wang, X.J. 1998. A role for NMDA-receptor channels in working memory *Nature Neurosci.* **1** 273–5.
- Lopez-Garcia, C. and Martinez-Guijjarro, FJ. 1988. Neurons in the medial cortex give rise to TIMM-positive boutons in the cerebral-cortex of lizards. *Brain Res* **463**:205-217.
- Margrie, TW, Rostas, JA, Sah, P. 1998. Long-term potentiation of synaptic transmission in the avian hippocampus. *J. Neurosci.* **18**:1207-1216.
- Marr, D. 1971. Simple memory: a theory for archicortex. *Philos Trans R Soc Lond B Biol Sci* 262:23-81.
- McNaughton, B.L. and Morris, R.G.M. 1987. Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends Neurosci* **10**:408-415.
- Molla, R., Rodriguez, J., Calvet, S., et al. 1986. Neuronal types of the cerebral-cortex of the adult chicken (*Gallus gallus*) – A golgi study. *Journal fur Hirnforschung* **27**:381-390
- Montagnese, C.M., Geneser, F.A. and Krebs, J.R.. 1993. Histochemical distribution of zinc in the brain of the zebra finch (*Taenopygia guttata*). *Anat Embryol (Berl)*. **188**:173-87.
- Montagnese, C.M., Krebs, J.R. and Meyer, G. 1996. The dorsomedial and dorsolateral forebrain of the zebra finch, *Taenopygia guttata*: a Golgi study. *Cell Tissue Res.* **283**:263-82.
- Mori, M., Gähwiler, B.H. and Gerber, U. 2007. Recruitment of an inhibitory hippocampal network after bursting in a single granule cell. *Proc. Natl. Aca. Sci. USA* **104**: 7640-7645 .
- O'Keefe, J. and Dostrovsky, J. 1971. The hippocampus as a spatial map: preliminary evidence from unit activity in the freely moving rat. *Brain Res* **34**:171-5.
- Olucha, F., Martinez-Garcia, F., Poch, L., Schwerdtfeger, W.K. and Lopez-Garcia, C. 1988. Projections from the medial cortex in the brain of lizards: correlation of anterograde and retrograde transport of horseradish peroxidase with Timm staining. *J Comp Neurol* **276**: 469-480.
- Papp, G. and Treves, A. 2005. What produces attractor effects in the rat hippocampus? *Soc. Neurosci. Abs.* **198**:12

Papp, G. and Treves, A. 2006. Continuous attractors come fragmented. *Soc Neurosci. Abs.* **68**:13.

Papp, G. and Treves, A. 2007. Network analysis of the significance of hippocampal subfields. In: *Hippocampal place-fields: Relevance to Learning and Memory*. (ed Sh. Mizumori), Oxford University Press.

Rattenborg N.C. 2007. Response to commentary on evolution of slow-wave sleep and palliopallial connectivity in mammals and birds: A hypothesis. *Brain Research Bulletin* **72**: 187–93

Rodriguez, F., Lopez, J.C., Vargas, J.P., Gomez, Y., Broglio, C. and Salas, C. 2002. Conservation of spatial memory function in the pallial forebrain of reptiles and ray-finned fishes. *J Neurosci.* **22**:2894-903.

Rolls, E.T. 1989. Functions of neuronal networks in the hippocampus and cerebral cortex in memory. In: *Models of Brain Function* (ed Cotterill R), pp. 15-33. Cambridge, UK: Cambridge UP.

Rolls, E.T., Stringer, S.M. and Elliot, T. 2006. Entorhinal cortex grid cells can map to hippocampal place cells by competitive learning. *Network* **17**:447-65.

Roudi, Y. and Treves, A. 2006 Localized activity profiles and storage capacity of rate-based autoassociative networks. *Phys Rev E* **73**: 061904

Samsonovich, A. and McNaughton, B.L. 1997. Path integration and cognitive mapping in a continuous attractor neural network model. *J Neurosci* **17**:5900.

Sadananda M (2004) Adult neurogenesis in the brain of higher vertebrates: implications of the paradigm shift. *Current Science* **87**:297-307.

Scott, TR and Bennett, MR. 1993. The effect of ions and second messengers on long-term potentiation of chemical transmission in avian ciliary ganglia. *Brit. J. Pharmacol.* **110**:461–69.

Sharp, P.E. 1991. Computer simulation of hippocampal place cells. *Psychobiology* **19**:103-115.

Shiple, M.T. and Adamek, G.D. 1984 The connections of the mouse olfactory bulb: a study using orthograde and retrograde transport of wheat germ agglutinin conjugated to horseradish peroxidase. *Brain Res Bull.* **12**:669-88.

Siegel, JJ, Nitz, D and Bingman VP. 2000. Hippocampal theta rhythm in awake, freely moving homing pigeons. *Hippocampus* **10**:627-631

Siegel, JJ, Nitz, D and Bingman VP. 2002. Electrophysiological profile of avian hippocampal unit activity: a basis for regional subdivisions. *J. Comp. Neurol.* **445**:256-268

Siegel, JJ, Nitz, D and Bingman VP. 2006. Lateralized functional components of spatial cognition in the avian hippocampal formation: evidence from single-unit recordings in freely moving homing pigeons. *Hippocampus.* **16**:125-40.

Skaggs, W.E., Knierim, J.J., Kudrimoti, H.S. and McNaughton, B.L. 1995. A model of the neural basis of the rat's sense of direction. *Adv Neural Inf Process Syst* **7**:173-80.

Solstad, T., Moser, E. I and Einevoll, G.T. 2006. From grid cells to place cells: a mathematical model. *Hippocampus* **16**:1026-31.

Sompolinsky, H. and Shapley, R. 1997. New perspectives on the mechanisms for orientation selectivity. *Curr Opin Neurobiol* **7**:514-22.

Stephan, H. 1975. Allocortex. In: *Handbuch der mikr. Anatomie des Menschen*, (Bargmann, W. ed.) vol 4, Teil 9. p998. Springer, Berlin, Heidelberg, New York

Striedter, G.F. *Principles of Brain Evolution*. 2005. Sinauer Associates, U.S.A.

Stringer, S.M., Trappenberg, T.P., Rolls, E.T. and de Araujo, I.E.T. 2002. Self-organizing continuous attractor networks and path integration: one-dimensional models of head direction cells, *Network: Computation in Neural Systems*, **13**: 217-242.

Stringer, S.M., Rolls, E.T., Trappenberg, T.P. and de Araujo, I.E.T. 2002. Self-organizing continuous attractor networks and path integration: two-dimensional models of place cells, *Network: Computation in Neural Systems*, **13**: 429-446.

Székely, A.D. and Krebs, J.R. 1996. Efferent connectivity of the hippocampal formation of the zebra finch (*Taeniopygia guttata*): an anterograde pathway tracing study using *Phaseolus vulgaris* leucoagglutinin. *J. Comp. Neurol.* **368**: 198-214.

ten Donkelaar, H.J. 2000. Major events in the development of the forebrain. *Eur. J. Morphol.* **38**:301-8.

Treves A, Miglino O and Parisi D (1992) Rats, nets, maps and the emergence of place cells. *Psychobiology* **20**:1-8.

Treves, A. and Rolls, E.T. 1991. What determines the capacity of autoassociative memories in the brain? *Network* **2**:371-97.

Treves, A. and Rolls, E.T. 1992. Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus* **2**:189-99.

Treves, A. 2004. Computational constraints between retrieving the past and predicting the future, and the CA3-CA1 differentiation. *Hippocampus* **14**:539-556.

Treves A (2005) Frontal latching networks: a possible neural basis for infinite recursion. *Cogn Neuropsychol* **22**:276-91

Tsodyks, M. and Sejnowski, T. 1995. Associative memory and hippocampal place cells *Int J Neural Systems* **6**:81-86.

- Ulinsky PS (1990) The cerebral cortex of reptiles. In: *Cerebral Cortex*, vol. 8A: Comparative Structure and Evolution of Cerebral Cortex (Jones EG, Peters A, eds), pp139-215. New York, USA: Plenum Press.
- Ulinsky, P.S.1990. Nodal events in the forebrain evolution. *Neth. J. Zool.* **40**:215-240.
- Ünal, B., Bradley, PM, Sahin, B., Canan, S., Aslan, H., Kaplan, S. 2002. Estimation of numerical density and mean synaptic height in chick hippocampus 24 and 48 hours after passive avoidance training. *Dev. Brain Res.* **136**:135-144.
- Wierazko, A and Ball, GF. 1993. Long-term potentiation in the avian hippocampus does not require activation of the N-methyl-D-aspartate (NMDA) receptor. *Synapse* **13**:173–78.
- Wills, T.J., Lever, C., Cacucci, F., Burgess, N. and O’Keefe, J. 2005. Attractor dynamics in the hippocampal representation of the local environment. *Science* **308**:873-6.
- Willshaw D and Buckingham J. 1990. An assessment of Marr's theory of the hippocampus as a temporary memory store. *Philos Trans R Soc Lond B Biol Sci* **329**:205-15.
- Wilson, M.A. and McNaughton, B.L. 1993. Dynamics of the hippocampal ensemble code for space. *Science* **261**:1055-8.
- Wouterlood, F.G. 1981. The structure of the mediodorsal cerebral cortex in the lizard *Agama agama*: a Golgi study. *J. Comp. Neurol.* **196**:443-58.
- Wyss,J.M. and Sripanidkulchai, K. 1983. The indusium griseum and anterior hippocampal continuation in the rat. *J Comp Neurol* **219**: 251-271.
- Zipser, D. 1985. A computational model of hippocampal place fields. *Behav.l Neurosci.* **99**:1006-18.