

In *Emergent neural computational architectures based on neuroscience*, J. Austin, S. Wermter, and D. Wilshaw (Eds.), Lecture Notes in Computer Science, Lecture Notes in Artificial Intelligence Springer-Verlag, Berlin.

## Biological Grounding of Recruitment Learning and Vicinal Algorithms in Long-term Potentiation

**Lokendra Shastri**

International Computer Science Institute

1947 Center Street, Suite 600

Berkeley, CA 94704, USA

TEL: (510) 642-4274; FAX: (510) 643-7684

shastri@icsi.berkeley.edu

### Abstract

Biological networks are capable of gradual learning based on observing a large number of exemplars over time as well as of rapidly memorizing specific events as a result of a single exposure. The focus of research in neural networks has been on gradual learning, and the modeling of one-shot memorization has received relatively little attention. Nevertheless, the development of biologically plausible computational models of rapid memorization is of considerable value, since such models would enhance our understanding of the neural processes underlying episodic memory formation. A few researchers have attempted the computational modeling of rapid (one-shot) learning within a framework described variably as *recruitment learning* and *vicinal algorithms*. Here it is shown that recruitment learning and vicinal algorithms can be grounded in the biological phenomena of long-term potentiation and long-term depression. Toward this end, a computational abstraction of LTP and LTD is presented, and an “algorithm” for the recruitment of *binding-detector* (or *coincidence-detector*) cells is described and evaluated using biologically realistic data.

## 1 Introduction

Biological neural networks are capable of slow gradual learning as well as rapid one-shot memorization. The former involves an exposure to a large number of exemplars and leads to the acquisition of perceptual-motor skills, category formation, language skills, and certain types of semantic knowledge. In contrast, one-shot memorization can result from a single exposure to an example, and underlies, among other things, the acquisition of “episodic memories” of everyday events, and memories of faces.

The primary focus of research in neural network models has been on slow gradual learning, and the modeling of one-shot memorization has received relatively little attention. Nevertheless, the development of biologically plausible computational models of rapid memorization is of considerable value, since such models would enhance our understanding of the neural processes underlying memory formation and retrieval, and could lead to the design of robust episodic memory modules for autonomous agents, and perhaps, to the development of memory prosthesis for brain injured humans.

A few researchers have attempted the computational modeling of rapid one-shot learning within a framework described variably as *recruitment learning* [10, 22, 8, 24, 11, 20] and *vicinal algorithms* [37]. In simple terms, recruitment learning can be described as follows: Learning occurs within a network of randomly connected nodes. Recruited nodes are those nodes in the network that have acquired a distinct “meaning” (or functionality) by virtue of their *strong*

interconnections to other recruited nodes and/or other sensorimotor (i.e., input/output) nodes. Nodes that are not yet recruited can be viewed as “free” nodes. Such nodes are connected via weak links to a large number of free, recruited, and/or sensorimotor nodes. These free nodes form a primordial network from which suitably connected nodes may be recruited for representing new items. For example, a novel concept  $y$  which can be expressed as a conjunct of existing concepts  $x_1$  and  $x_2$  can be memorized by (i) identifying free nodes that receive links from nodes representing  $x_1$  as well as nodes representing  $x_2$  and (ii) “recruiting” one or more such free nodes by strengthening the weights of links incident on such nodes from  $x_1$  and  $x_2$  nodes.

Feldman [10] showed that conjunctive concepts can be recruited with a high probability if one makes suitable assumptions about network connectivity. He presented a probabilistic analysis of recruitment learning based on the degree of connectivity and the number of intermediate layers in random interconnection networks. Shastri [22] extended the notion of recruitment learning to relational concepts. He treated a concept as a collection of attribute-value bindings and suggested a two-stage memorization process. In the first stage, *binder* nodes are recruited for each attribute-value binding in a concept. In the second stage, these *binder* nodes are joined together by the recruitment of another conjunctive node. Diederich [8] showed how this form of structured recruitment learning can be used to learn new concepts expressed as modifications of existing concepts. Valiant [37] proposed a formal “neuroidal model” and described several algorithms for the recruitment learning of conjunctive and relational concepts. He also presented a quantitative analysis of these algorithms using plausible assumptions about connectivity in the neocortex. Valiant referred to these algorithms as “vicinal algorithms.”

While general arguments in support of the neural plausibility of recruitment learning and vicinal algorithms have been presented in the past (see [10, 37]), a specific neural correlate of such learning has not been proposed. In this paper it is shown that recruitment learning can be firmly grounded in the biological phenomena of *long-term potentiation* (LTP) and *long-term depression* (LTD) that involve rapid, long-lasting, and highly specific changes in synaptic strength. Toward this end, a computational abstraction of LTP and LTD is proposed, and an “algorithm” for the recruitment of binding-detector (or *coincidence detector*) cells is described and evaluated using biologically realistic data about region sizes and cell connectivity. In the proposed grounding, the specification of a vicinal algorithm amounts to choosing a suitable network architecture and a set of appropriate parameter values for the induction of LTP and LTD.

The rest of the paper is organized as follows: Section 2 briefly reviews the phenomena of LTP and LTD. Section 3 describes a computational abstraction of cells, synapses, LTP, and LTD. Section 4 describes how a transient pattern of activity can lead to the recruitment of binding-detector cells as a result of LTP (and optionally, LTD) within quasi-random network structures. Finally, section 6 presents some concluding remarks.

## 2 Long-term potentiation and depression

Long-term potentiation (LTP) refers to a long-term increase in synaptic strength<sup>1</sup> resulting from the pairing of presynaptic activity with postsynaptic depolarization [5, 14] LTP was first observed in the rabbit hippocampal formation, and has since been observed in synapses along many excitatory pathways in the mammalian brain. Recent evidence strongly suggests that LTP plays a direct causal role in learning and memory formation (e.g., [33, 21]).

The most extensively studied form of LTP involves the unusual receptor NMDA<sup>2</sup> (*N*-methyl-D-aspartate) which is

---

<sup>1</sup>A synapse is the site of communication between two cells. Typically, a synapse is formed when an axonal (output) fiber emanating from a “presynaptic” cell makes contact with the dendrites (input structure) of a “postsynaptic” cell.

A synapse can be excitatory or inhibitory. The arrival of activity at an excitatory synapse from its presynaptic cell leads to a depolarization of the local membrane potential of its postsynaptic cell and makes the postsynaptic cell more prone to firing. In contrast, the arrival of activity at an inhibitory synapse leads to a hyperpolarization of the local membrane potential of the postsynaptic cell and makes the postsynaptic cell less prone to firing. The strength of an excitatory (or inhibitory) synapse determines the degree of depolarization (or hyperpolarization) that will result from a given presynaptic activity. The greater the synaptic strength, the greater the depolarization (hyperpolarization).

<sup>2</sup>Not all forms of LTP are NMDA receptor-dependent. The LTP of synapses formed by mossy-fibers on CA3 pyramidal cells is a case in point

activated by the excitatory neurotransmitter glutamate, but only if the postsynaptic membrane is sufficiently depolarized. In the absence of adequate depolarization, NMDA receptor-gated channels remain blocked by magnesium ions in spite of glutamate being bound to the receptor. Adequate depolarization of the postsynaptic membrane, however, expels the magnesium ions and unblocks the channels. Once the channels are unblocked, calcium ions flood into the dendritic spine of the postsynaptic cell and trigger a complex series of biochemical changes that result in the induction of LTP.

The two conditions required for the activation of NMDA receptor, namely, presynaptic activity and strong postsynaptic depolarization, together entail that the LTP requires the concurrent arrival of activity at several synapses of the postsynaptic cell. This is referred to as the *cooperativity* property of LTP.

Several properties of LTP make it suitable for serving as the basis for one-shot recruitment learning. First, it is induced rapidly — within a few seconds, and is fully present within 20-30 seconds. Second it is long lasting. Third, the cooperativity property of LTP makes it an ideal mechanism for transforming a *transient* expression of a relationship between two items or more (encoded as the coherent activity of the ensembles representing these items) into a *persistent* expression of this relationship (encoded via long-term changes in the efficacy of synapses linking the ensembles representing these items). Finally, LTP is synapse specific, and hence, it can express highly specific bindings and correlations.

LTP resulting from the arrival of coincident activity along afferent fibers belonging to a single pathway is referred to as *homosynaptic* LTP. If the arrival of coincident activity along two independent pathways, *A* and *B*, leads to the LTP of synapses formed by fibers of *A*, but the arrival of activity along fibers of *A* alone does not, then the LTP of synapses formed by fibers of *A* is referred to as *associative* LTP [12, 6]. We will, however, distinguish between associative and homosynaptic LTP based on the *representational* distinctiveness of the afferent sources whose cooperative activity leads to LTP. Thus we will use the qualifier “homosynaptic” to refer to LTP resulting from the arrival of coincident activity along afferents fibers emanating from cells representing the *same* item, and we will use the qualifier “associative” to refer to LTP resulting from the arrival of coincident activity along two sets of afferent fibers, with each set emanating from cells representing a distinct item.

In addition to LTP, synapses along key excitatory pathways in the mammalian hippocampal formation have been shown to undergo long-term depression (LTD) [3, 13]. A synapse receiving no presynaptic activity can undergo *heterosynaptic* LTD if other synapses of the same postsynaptic cell receive strong presynaptic activity. In other words, the absence of presynaptic activity in the presence of strong postsynaptic activity can lead to heterosynaptic LTD of a synapse. A synapse may undergo *associative* LTD upon receiving presynaptic activity that is out of phase with strong rhythmic activity converging on other synapses of the postsynaptic cell [34]. Finally, prolonged low frequency stimulation of a synapse can lead to its homosynaptic LTD [9].

### 3 A Computational Abstraction of LTP and LTD

The computational abstraction of LTP and LTD proposed here is an highly simplified idealization of the complex biophysical processes underlying the induction and expression of LTP and LTD. This abstraction is guided by two considerations. First, the abstraction should be rich enough to capture temporal aspects critical for modeling LTP and LTD. Second, the abstraction should be *discrete* and minimal so as to facilitate quantitative analyses and efficient computer simulations of large-scale neuronal networks.

---

[18].

### 3.1 Cells

A cell is modeled as an idealized integrate-and-fire neuron (e.g., see [16]), and the spatio-temporal integration of activity arriving at a cell is modeled as follows:

Let  $a_i(t)$  be a measure of presynaptic activity occurring at synapse  $s_i$  of the cell at time  $t$ . In biophysical terms,  $a_i(t)$  may correspond to the number of spikes arriving at  $s_i$  within a unit time interval anchored at  $t$ . Thus the arrival of a high-frequency spike-burst at  $s_i$  would correspond to a high value of  $a_i(t)$ . Let  $w_i(t)$  refer to the weight of synapse  $s_i$  at time  $t$ .

The postsynaptic potential,  $psp_i(t | a_i(t_0))$ , resulting from the presynaptic activity at  $s_i$  at time  $t_0$  is modeled as a piecewise linear function consisting of a rising (ramp-up) segment, a flat (plateau) segment, and a falling (decay) segment. That is:

$$psp_i(t | a_i(t_0)) = \begin{cases} m_r * (t - t_0) & t_0 \leq t < (t_0 + \Delta T_r) \\ m_r * \Delta T_r & (t_0 + \Delta T_r) \leq t < (t_0 + \Delta T_{rs}) \\ m_r * \Delta T_r + (t - \Delta T_{rs}) * m_f & (t_0 + \Delta T_{rs}) \leq t < (t_0 + \omega_{int}) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where  $m_r$  is the slope of the rising segment and is given by  $(a_i(t_0) * w_i(t_0)) / \Delta T_r$ ,  $\Delta T_r$  is the duration of the rising segment,  $\Delta T_s$  is the duration of the flat segment,  $\Delta T_{rs}$  equals  $(\Delta T_r + \Delta T_s)$ ,  $m_f$  is the slope of the falling segment, and  $\omega_{int}$  is the window of temporal integration denoting the maximum amount by which two incident activities may lead/lag and still be summated by the postsynaptic cell. Note that

$$\omega_{int} = \Delta T_{rs} + (m_r * \Delta T_r) / m_f \quad (2)$$

The postsynaptic potential at time  $t$  attributable to  $s_i$ ,  $psp_i(t)$ , can be obtained by summing the effect of all the activity arriving at  $s_i$  during the past  $\omega_{int}$  time units. Thus

$$psp_i(t) = \sum_{(0 \leq \tau < \omega_{int})} psp_i(t | a_i(t - \tau)) \quad (3)$$

and  $pot(t)$ , the cell's potential at time  $t$  resulting from the combined effect of presynaptic activity at all its synapses equals:

$$pot(t) = \sum_i psp_i(t) \quad (4)$$

where  $i$  ranges over all synapses of the cell.

A cell has a firing threshold,  $thresh_f(t)$ , with a resting value of  $\theta_f$ . A cell fires at time  $t$  if  $pot(t) \geq thresh_f(t)$ , and produces an action potential (spike). This spike arrives at synapses downstream from the cell at time  $t + d$ , where  $d$  is the propagation delay.

After a cell fires, it enters a refractory state for a duration  $\omega_{ref}$ . During this interval, the cell does not fire irrespective of its inputs. That is,

$$thresh_f(t) = \begin{cases} +\infty & \text{if cell has fired during the interval } [t - \omega_{ref}, t - 1] \\ \theta_f & \text{otherwise} \end{cases} \quad (5)$$

Some cell-types can have two firing modes: *supra-active* and *normal*. These modes are associated with firing thresholds  $\theta_{sf}$  and  $\theta_f$ , respectively, ( $\theta_{sf} > \theta_f$ ), and output levels  $O_2$  and  $O_1$ , respectively, ( $O_2 > O_1$ ). Neurally, the

*supra-active* mode corresponds to a high-frequency burst response such as the complex spike burst response generated by hippocampal pyramidal cells, and the *normal* mode corresponds to a simple spike response consisting of isolated spikes. The proposed abstraction of the distinction between a complex spike burst response and a simple spike response based on firing thresholds and output levels is a gross simplification. But for suitable choices of parameter values, this simple abstraction offers a computationally inexpensive, yet functionally adequate, means of modeling the two distinct response modes of certain cells.

### 3.2 Projection

A *projection* refers to the set of links emanating from cells in a source region and impinging on cells in a target region. It is assumed that all the synapses formed by a projection are of the same *type* and have similar attributes.

### 3.3 Synapses

A synapse can be in any one of following three states: *naive*, *potentiated*, or *depressed*. The state of a synapse signifies its strength (weight). For a given synaptic type, the weights of all synapses in a given state lie within a restricted band. The weight bands associated with different states are disjoint. The weight bands associated with a synaptic state may differ from one synaptic type to another.

### 3.4 Computational modeling of LTP

The induction of LTP is governed by the following parameters: the *potentiation threshold*  $\theta_p$ , the *weight increment*  $\Delta w_{ltp}$ , the *repetition factor*  $\kappa$ , and the *maximum inter-activity interval*  $\tau_{iai}$ .

Consider a set of neighboring synapses  $s_1, \dots, s_n$  sharing the same postsynaptic cell. Convergent presynaptic activity at  $s_1, \dots, s_n$  can lead to LTP of naive  $s_i$ 's and increase their weights by  $\Delta w_{ltp}$  if the following conditions hold:

1.  $\sum_{1 \leq i \leq n} psp_i(t) \geq \theta_p$

Note that in order to summate, the presynaptic activity arriving at  $s_1, \dots, s_n$  must be “synchronous”, that is, the maximum lead/lag in incident activity at any pair of synapses should be no more than  $\omega_{int}$ .

2. Such synchronous presynaptic activity recurs (repeats) at least  $\kappa$  times.
3. The interval between two *successive* arrivals of presynaptic activity at a synapse during the above repetition is at most  $\tau_{iai}$  time units. In other words, successive volleys of synchronous activity should not be more than  $\tau_{iai}$  apart.

Note that associative and homosynaptic LTP are modeled in an analogous manner. The difference between homosynaptic and associative LTP is simply this: In the case of homosynaptic LTP, the activity leading to LTP emanates from a cell ensemble representing a single item. In the case of associative LTP such activity emanates from multiple cell ensembles representing more than one item.

### 3.5 Computational modeling of LTD

Heterosynaptic LTD is also modeled similarly using five parameters. These are: the potentiation threshold  $\theta_p$ , the weight decrement  $\Delta w_{ltd}$ , the repetition factor  $\kappa$ , the maximum inter-activity interval  $\tau_{iai}$ , and the *propensity of LTD*  $\zeta$   $0 \leq \zeta \leq 0$ . When naive or potentiated synapses of a postsynaptic cell receive convergent presynaptic activity, neighboring inactive naive synapses of the postsynaptic cell undergo heterosynaptic LTD and their weights decrease

by  $\Delta w_{ltd}$ . As in the case of LTP,  $\theta_p$  dictates the minimum weighted sum of synchronous activity that neighboring synapses of the postsynaptic cell must receive, and  $\kappa$  specifies the number of times such presynaptic activity must recur in order to induce heterosynaptic LTD of naive inactive synapses. Also as before,  $\tau_{iai}$  specifies the maximum permissible gap between the successive arrival of presynaptic activity. The parameter  $\zeta$  specifies the fraction of inactive naive synapse that undergo LTD when the above conditions are met. Thus  $\zeta$  provides a simple computational mechanism for controlling the prevalence of heterosynaptic LTD. A value of  $\zeta = 0$  means that there is no heterosynaptic LTD and a value of  $\zeta = 1$  means that a single occurrence of LTP can lead to the heterosynaptic LTD of all inactive naive synapses of the postsynaptic cell.

### 3.6 Modeling neuromodulation

The effect of neuromodulators on the response of a cell and on the induction of LTP at a synapse is a complex phenomena. In the present proposal, these effects are modeled by positing an additional input (or bias) that modifies the firing thresholds ( $\theta_f$  and  $\theta_{sf}$ ) of a cell and the potentiation threshold ( $\theta_p$ ) of a synapse.

### 3.7 Emergence of cells and circuits responsive to specific functionalities

LTP and LTD can transform random networks into structures consisting of cells tuned to specific functionalities. Typically, a cell receives a large number of inputs (afferents), and hence, can potentially participate in a large number of functional circuits. If, however, the weights of selected synapses on the cell increase via LTP (and, optionally, the weights of other synapses decrease via LTD) the cell can become more selective and participate in a limited number of functional circuits. Thus LTP and LTD provide a promising neural mechanism for the recruitment of structures with specific functionalities within quasi-random networks.

Let us consider the simplest of all cases where a cell  $B$  (see Fig. 1) becomes *linked* to a cell ensemble  $A$  as a result of LTP (cell  $B$  is linked to a cell ensemble  $A$  if the firing of a significant number of cells in  $A$  leads to the firing of  $B$ ). Under appropriate conditions, the firing of cells in  $A$  would lead to the homosynaptic LTP of synapses formed by afferents from  $A$  impinging on  $B$  and (optionally) to the heterosynaptic LTD of some of the inactive synapses formed by the afferents from other cells impinging on  $B$ . The strengthening of synapses from  $A$  to  $B$  would increase the likelihood that  $B$  fires whenever cells in  $A$  fire. This would result in cell  $B$  becoming linked to cell ensemble  $A$ . Additionally, the weakening of  $B$ 's synapses formed by cells not in  $A$  would lower the likelihood that  $B$  fires in circumstances when cells in  $A$  do not fire.

In the following section we illustrate how transient activity propagating through neural circuits can automatically lead to the recruitment of binding-detector (coincidence detector) cells.

## 4 Recruitment of binding-detector cells

Our ability to remember events in our daily life demonstrates our capacity to rapidly acquire new memories. Typically, such memories record who did what to whom where and when, or describe states of affairs wherein multiple entities occur in particular configurations. This form of memory is often referred to as episodic memory [36], and there is a broad consensus that the hippocampal formation and neighboring areas in the medial temporal lobes serve a critical role in its formation [19, 32, 7, 35].

The persistent encoding of an event must be capable of encoding role-entity *bindings*. Consider the event *described* by “John gave Mary a book in the library on Tuesday”. This event cannot be encoded by simply forming a conjunctive association between “John”, “Mary”, “a book”, “Library”, “Tuesday” and “give” since such an encoding would be indistinguishable from that of the event described by “Mary gave John a book in the library on Tuesday”. In order

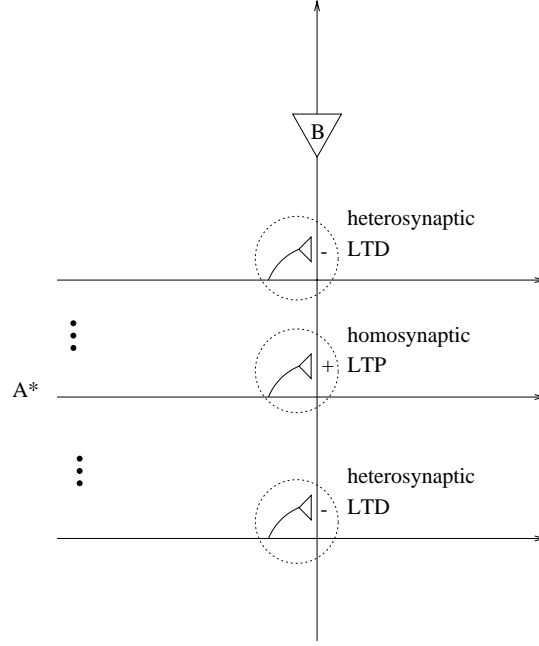


Figure 1: Cell  $B$  becomes linked to ensemble  $A$ . The label  $A^*$  attached to one of the afferents of  $B$  indicates that (i) the source of this afferent is a cell in ensemble  $A$  and (ii) this source cell is firing. Only a single afferent from ensemble  $A$  to  $B$  is shown. In general,  $B$  may have to receive several afferents from cells in  $A$  in order to become linked to  $A$ .

to make the necessary distinctions, the encoding of an event should specify the *bindings* between the *entities* participating in the event and the *roles* they play in the event. For example, the encoding of the event in question should specify the following *role-entity* bindings: ( $\langle \text{giver}=\text{John} \rangle$ ,  $\langle \text{recipient}=\text{Mary} \rangle$ ,  $\langle \text{give-object}=\text{a-Book} \rangle$ ,  $\langle \text{temporal-location}=\text{Tuesday} \rangle$ ,  $\langle \text{location}=\text{Library} \rangle$ ).

As explained in [27], it is possible to evoke a fleshed out representation of an event by “retrieving” the bindings pertaining to the event and activating the web of semantic and procedural knowledge with these bindings. Thus cortical circuits encoding generic “knowledge” about actions such as *give* and entities such as *persons*, *books*, *libraries*, and *Tuesday* can recreate the necessary gestalt and details about the event “John gave Mary a book on Tuesday in the library” upon being activated with the above bindings. This view is supported by work on “reflexive reasoning” [28, 25, 30] and “executing schemas” [4, 29].

In view of the above, the recruitment of binding-detectors is expected to be a critical step in the memorization of episodic memory. The following describes how such binding-detector cells can arise spontaneously and rapidly within a biologically motivated network structure as a result of LTP (and optionally, LTD).

#### 4.1 A structure for the encoding of binding-detector cells

A structure for the rapid formation of cells responsive to binding matches consists of three regions: *ROLE*, *ENTITY*, and *BIND* (see Fig. 2(a)). Regions *ROLE* and *ENTITY* are assumed to have 750,000 primary (excitatory) cells each, while region *BIND* is assumed to have 15 million cells. Regions *ROLE* and *ENTITY* have dense projections to region *BIND*, with each cell in *ROLE* and *ENTITY* regions making 17,000 connections with cells in region *BIND*. In other

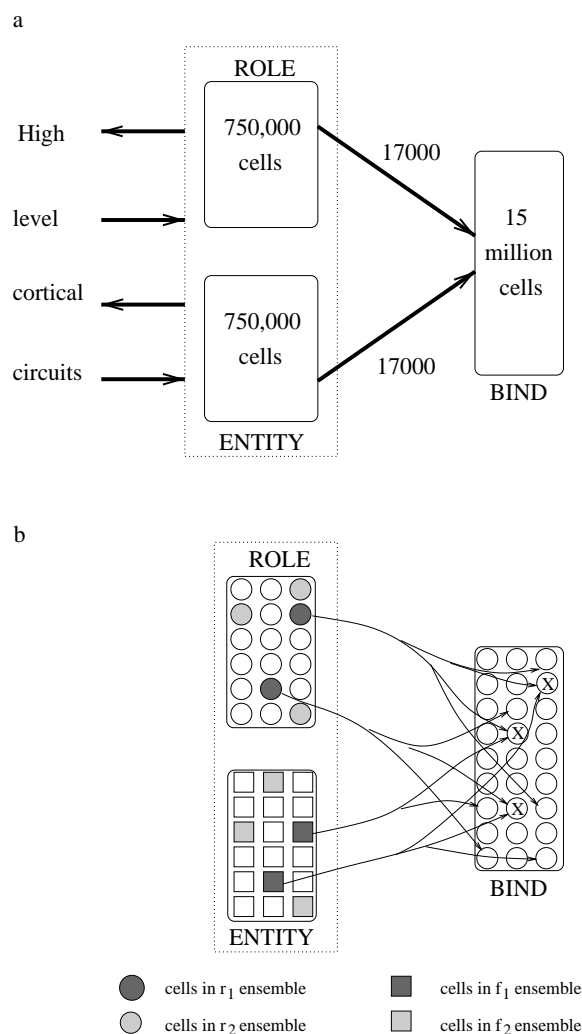


Figure 2: (a) A structure for the formation of binding-detector cells. Arcs indicate projections and the number on an arc indicates the projective field size. These projections are assumed to be uniformly distributed over the BIND region. Each role and entity is encoded by a small ensemble of cells in the ROLE and ENTITY regions, respectively. Cells in role and entity ensembles are also assumed to be distributed uniformly within regions ROLE and ENTITY, respectively. Binding-detector cells are recruited in region BIND. It is assumed that the ROLE and ENTITY regions lie in the entorhinal cortex and the region BIND corresponds to the dentate gyrus (a part of the hippocampus). The projective field and region sizes are based on [2, 40]. (b) A schematic depiction of the ensembles of roles  $r_1$  and  $r_2$  and entities  $f_1$  and  $f_2$ . Only links from cells in  $r_1$  and  $f_1$  ensembles to cells in BIND are shown. Cells marked with an “X” are candidates for recruitment as *binder* cells for the binding  $\langle r_1 = f_1 \rangle$ .

words, the projective field (PF) size<sup>3</sup> of the ROLE to BIND projection as well as the ENTITY to BIND projection is 17,000. It is assumed that these projections are uniformly distributed over BIND.

Each role and entity is encoded by a small ensemble of cells in the ROLE and ENTITY regions, respectively. Cells in role and entity ensembles are also assumed to be distributed uniformly within regions ROLE and ENTITY, respectively. Note that a cell in ROLE may belong to multiple role ensembles, and a cell in ENTITY may belong to multiple entity ensembles.

## 4.2 A neural correlate of the structure for encoding binding-detectors

There is a direct correspondence between the model structure described above and the interaction between the entorhinal cortex (EC) which is a region in the medial temporal lobe, and the dentate gyrus (DG) which is a component of the hippocampal formation. The ROLE and ENTITY regions correspond to subregions of the EC and the BIND region corresponds to the DG. The projections from high-level cortical areas to ROLE and ENTITY regions correspond to the well known cortical projections to EC [38]. The dense projections from ROLE and ENTITY to BIND correspond to the dense projections from EC to DG [2]. Moreover, the projective field and region sizes shown in Fig. 2(a) are based on anatomical findings presented in [2, 40] (see [27] for a detailed discussion).

## 4.3 The transient representation of role-entity bindings

It is assumed that the bindings constituting an event are expressed as a transient pattern of rhythmic activity over distributed high-level cortical circuits (HLCCs) [39, 1, 28, 31]. These HLCCs project to cells in ENTITY and ROLE regions and, in turn, induce transient patterns of rhythmic activity within these regions. Fig. 3 is an idealized depiction of the transient activity induced in ENTITY and ROLE regions by HLCCs to convey the relational instance  $RI: (\langle r_1 = f_1 \rangle, \langle r_2 = f_2 \rangle)$ . Here  $r_1$  and  $r_2$  are roles, and  $f_1$  and  $f_2$  are entities bound to  $r_1$  and  $r_2$ , respectively. Each spike in the illustration signifies the synchronous firing of a cell ensemble. It is shown that cells in the  $r_1$  and  $f_1$  ensembles are firing in synchrony, and so are cells in the  $r_2$  and  $f_2$  ensembles. The firing of cells in the  $r_1$  and  $f_1$  ensembles, however, is desynchronized with the firing of cells in the  $r_2$  and  $f_2$  ensembles. This desynchronization is assumed to be  $\geq \omega_{int}$  time units. Note that the dynamic encoding of  $RI$  can be viewed as a periodic pattern consisting of two phases:  $\rho_1$  and  $\rho_2$ . Here  $\rho_1$  and  $\rho_2$  are mere labels and the ordering of phases has no significance.

In effect, a role-entity binding is expressed by the synchronous firing of the cell ensembles associated with the bound role and entity [1, 28]. In general, the transient encoding of a relational instance with  $n$  distinct entities participating as role-fillers involves  $n$  interleaved quasi-periodic activities having a period  $\pi$ . It is assumed that  $\pi \leq \tau_{iai}$  time units. Such a spatio-temporal encoding enables multiple role-entity bindings to be expressed and propagated concurrently without cross-talk [28].

The following section explains how such a transient encoding of a relational instance may be transformed rapidly into persistent circuits for detecting bindings.

## 4.4 Recruitment of binder cells for memorizing role-entity bindings

BIND contains two kinds of cells: principal cells (these correspond to granule cells in the dentate gyrus) and Type-1 inhibitory interneurons.<sup>4</sup> Each principal cell receives afferents from a number of cells in ROLE and ENTITY regions and makes synaptic contacts on a number of interneurons. The interneurons in turn make contacts on a number of principal

<sup>3</sup>The set of cells in the target region that receive links from a cell  $c$  in the source region is referred to as the projective field (PF) of  $c$ . The PF size of  $c$  refers to the number of synapses formed by  $c$  with cells in the target region.

<sup>4</sup>The synapses formed by inhibitory interneurons on other cells have negative weights.

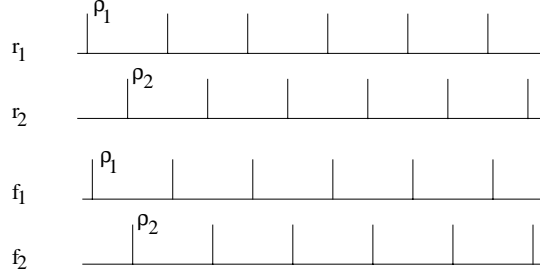


Figure 3: The transient encoding of a relational instance  $RI$  given by the bindings:  $(\langle r_1 = f_1 \rangle, \langle r_2 = f_2 \rangle)$ . Here  $r_1$  and  $r_2$  are roles, and  $f_1$  and  $f_2$  are entities bound to  $r_1$  and  $r_2$ , respectively, in  $RI$ . Each spike in the illustration signifies the synchronous firing of a cell ensemble. Cells in the  $r_1$  and  $f_1$  ensembles fire in synchrony and so do cells in the  $r_2$  and  $f_2$  ensembles. The firing of cells in the  $r_1$  and  $f_1$  ensembles, however, is desynchronized with the firing of cells in the  $r_2$  and  $f_2$  ensembles. This desynchronization is assumed to be  $\geq \omega_{int}$  time units. Moreover, the period of firing,  $\pi$ , is assumed to be  $\leq \tau_{iai}$  time units. The dynamic encoding of  $RI$  can be viewed as a periodic pattern consisting of two *phases*:  $\rho_1$  and  $\rho_2$  (the *order* in which these phases appear has no significance).

cells, thereby forming inhibitory circuits within BIND (see Fig. 4). The significance of inhibitory interneurons will be explained later.

The potentiation threshold,  $\theta_p$ , of principal cells is sufficiently high, and hence, LTP of a synapse occurs only if multiple synapses of a postsynaptic cell receive coincident presynaptic activity.<sup>5</sup> Moreover, the response threshold,  $\theta_f$ , of principal cells is such that a cell does not fire unless it receives impulses at multiple potentiated synapses. A set of values for  $\theta_p$ ,  $\theta_f$ , synaptic weights, and other parameters of LTP and LTD are given below.<sup>6</sup>:

$$\theta_f = 1700; \quad \theta_p = 890;$$

$$\text{naive weight band} = 100-110;$$

$$\Delta_{ltp} = 100; \quad \Delta_{ltd} = 50;$$

$$\zeta_{ltp} = 1; \quad \zeta_{ltd} = 0;$$

$$\kappa = 4; \quad \omega_{int} = 5; \quad \omega_{ref} = 2.$$

The choice of  $\tau_{iai}$  is governed by  $\omega_{int}$  and the number of role-entity bindings in an event. Thus any  $\tau_{iai} \geq \omega_{int} * n$ , where  $n$  is the number of bindings in the event, is appropriate.

The transient encoding of the relational instance  $RI$  shown in Fig. 3 leads to the following events in BIND (refer to Fig. 4). The synchronous firing of cells in the  $r_1$  and  $f_1$  ensembles (henceforth,  $r_1$  and  $f_1$  cells) leads to the associative LTP of active synapses of principal cells receiving sufficient afferents from  $r_1$  and  $f_1$  cells. At the same time, depending on the value of  $\zeta_{ltd}$ , some of the inactive naive synapses of these principal cells may undergo heterogeneous LTD. The LTP of synapses formed by afferents arriving from  $r_1$  and  $f_1$  cells makes these principal cells behave as binding

<sup>5</sup>Here and elsewhere in the paper, “coincidence” is defined with reference to  $\omega_{int}$ , the window of temporal integration. See Section 3.

<sup>6</sup>LTD does not play a critical role in the recruitment of binding-detector cells described here. It might, however, play an important role in the recruitment of other functional circuits.

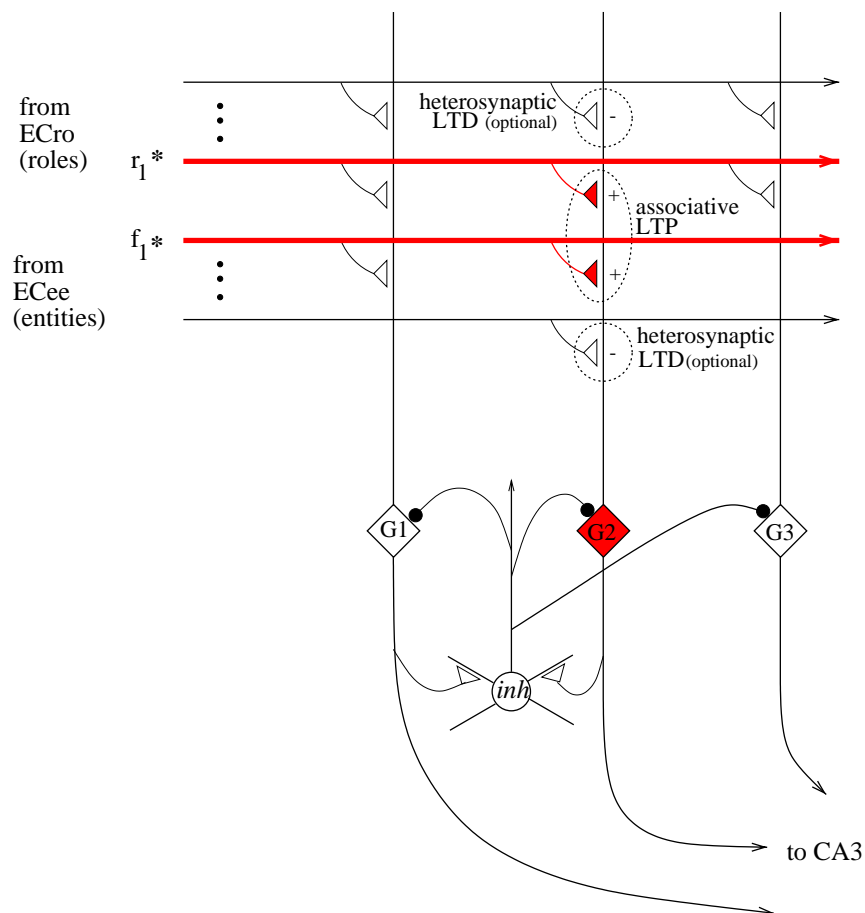


Figure 4: The BIND region consists of principal cells and inhibitory interneurons (Type-1). In the illustration, G1—G3 are principal cells and *inh* is a Type-1 interneuron. Afferents (incoming links) labeled  $r_1^*$  and  $f_1^*$  are from cells in the ensembles for role  $r_1$  and entity  $f_1$ , respectively. Since G1 and G2 receive synchronous activity along afferents from  $r_1$  and  $f_1$  cells, they are *candidates* for becoming binding-detector cells for the binding  $\langle r_1 = f_1 \rangle$ . It is assumed that the inhibition from *inh* prevents the LTP of G1's synapses, and only G2 becomes a binding-detector cell for  $\langle r_1 = f_1 \rangle$ . Filled blobs denote inhibitory synapses, and the size of a filled blob is meant to convey the strength of the (inhibitory) synapse. See text for additional details.

detector cells for the binding  $\langle r_1 = f_1 \rangle$  and we will refer to such cells as  $binder(\langle r_1 = f_1 \rangle)$  cells.<sup>7</sup>

The claim that  $binder(\langle r_1 = f_1 \rangle)$  cells behave as binding-detector cells for  $\langle r_1 = f_1 \rangle$  is substantiated quantitatively in Section 5, but let us examine why these cells behave in the desired manner. Note that a  $binder(\langle r_1 = f_1 \rangle)$  cell will fire in response to the synchronous firing of  $r_1$  and  $f_1$  cells since the connectivity between  $r_1$  and  $f_1$  cells and a principal cell required for the latter's recruitment as a  $binder$  cell during the memorization of  $\langle r_1 = f_1 \rangle$  also suffices for the latter's firing during the retrieval of  $\langle r_1 = f_1 \rangle$ . At the same time, since  $\theta_f$  is quite high (1700), a  $binder(\langle r_1 = f_1 \rangle)$  cell is unlikely to fire as a result of stray impulses arriving at its synapses.

Similar LTP and LTD events occur at the synapses of principal cells that receive coincident activity along afferents from  $r_2$  cells in ROLE and  $f_2$  cells in ENTITY, and lead to their recruitment as  $binder(\langle r_2 = f_2 \rangle)$  cells. A  $binder(\langle r_2 = f_2 \rangle)$  cell fires whenever  $r_2$  cells in ROLE and  $f_2$  cells in ENTITY fire in synchrony and behaves as a binding detector cell for the role-entity binding  $\langle r_2 = f_2 \rangle$ .

## 4.5 Encoding and Response Times

The time required for the recruitment of  $binder$  cells is given by  $\kappa * \tau_{iai}$ . If we assume that the rhythmic activity encoding dynamic bindings corresponds to  $\gamma$  band activity (ca. 40 Hz) we get  $\tau_{iai} \simeq 25$  msec. Assuming a plausible value of  $\kappa$  to be 4 suggests that  $binder$  cells can be recruited in about 100 msec. The time required for  $binder$  cells to respond to a retrieval cue is at most  $\tau_{iai}$ . Thus both the recruitment and response times of the proposed model are consistent with the requirements of rapid (one-shot) memorization and recognition.

## 4.6 Potential problems in the formation of binder cells

The process by which  $binder$  cells are formed is susceptible to several problems. First, in order to form  $binder(\langle r_i = f_j \rangle)$  cells there should exist cells that receive afferents from both  $r_i$  and  $f_j$  cells. Given the random connectivity between the regions ROLE and ENTITY and BIND this cannot be *guaranteed*. Second, there may exist cells that receive sufficient activity ( $\geq \theta_p$ ) along afferents from  $r_i$  cells alone. Upon recruitment, such an *ill-formed* cell would produce false-positive responses since it will fire in response to the firing of  $r_i$  cells alone, even if there is no coincident activity of  $f_j$  cells. Similarly, cells receiving sufficient links from  $f_j$  cells alone could also be recruited as ill-formed  $binder$  cells. Third, the same cell may get recruited as a  $binder$  cell for multiple bindings. Consider a cell that gets recruited as a  $binder$  cell for two bindings  $\langle r_i = f_k \rangle$  and  $\langle r_j = f_l \rangle$ . This cell will fire in response to subsequent inputs containing either of these two bindings as well as the bindings:  $\langle r_i = f_l \rangle$ , and  $\langle r_j = f_k \rangle$ . Consequently, other cells connected downstream to this cell could receive false-positive binding-match signals in certain circumstances.

Using biologically motivated values of various system parameters it is shown in Section 5 that the probability of not finding cells for recruitment as  $binder$  cells is vanishingly small. The problem of too many cells becoming recruited for a binding turns out not to be very serious in the case under consideration, and hence, is not discussed here. In general, however, this problem can be alleviated by inhibitory feedback and feedforward local circuits formed by principal cells and Type-1 inhibitory interneurons. These inhibitory circuits act as soft-WTA and only allow synapses of a limited number of cells to undergo LTP (cf. [15, 17]). Furthermore, it is shown that the ensemble response of binder cells is highly robust, in spite of the possibility that ill-formed and overlapping  $binder$  cells can be recruited.

---

<sup>7</sup>To be precise, a cell is deemed to be recruited as a  $binder(\langle r_1 = f_1 \rangle)$  cell if during the memorization of  $\langle r_1 = f_1 \rangle$ , the cell's synapses undergo LTP and the cell fires. The firing of the cell at the time of its recruitment is crucial if the cell is to become part of functional circuits lying downstream from the cell.

## 5 Quantitative Results

The following quantities have been calculated analytically<sup>8</sup> using the region and projective field sizes described in Section 4.1, the cell, synapse, and LTP parameters described in Section 4.4, and by assuming that each role and entity ensemble contains 600 cells.

1.  $P_{fail}$ , the probability that for a given binding *no* cells will be found in BIND (DG) for recruitment as binding detector cells is less than  $< 10^{-18}$ . The probability that the system would be unable to encode a binding is essentially zero.
2. The expected number of cells in BIND (DG) that will receive appropriate connections and will be candidates for recruitment for a binding is 195. Thus a fairly large number of cells are recruited as *binder* cells for each binding.
3. The expected number of *binder* ( $\langle r1 = f1 \rangle$ ) cells that will fire in response to various retrieval *cues* are shown in Fig. 5. Note that *binder* ( $\langle r1 = f1 \rangle$ ) cells
  - respond robustly to the matching binding  $\langle r1 = f1 \rangle$
  - produce an extremely weak response to erroneous, but partially related bindings of the form  $\langle r1 = fx \rangle$  and  $\langle rx = f1 \rangle$  (where  $fx$  is any entity other than  $f1$ , and  $rx$  is any role other than  $r1$ )
  - produce essentially no response to unrelated bindings.

Note that this performance holds irrespective of the number of bindings memorized in BIND.

Since each binding is redundantly encoded by multiple cells, and since these cells are physically dispersed in the BIND region, the probability that a cell loss will destroy many *binder* cells for any given binding remains extremely small. In particular, a loss of  $x\%$  of the cells in region BIND will lead to an expected loss of only  $x\%$  of the 195 *binder* cells for a given binding. Thus the memorization of binding-detectors is robust with respect to cell loss (e.g., a cell loss of 10% will leave about 175 *binder* cells of any given binding intact).

The results in Fig. 5 are based on  $\zeta_{ltd} = 0$  (i.e., no LTD). This condition results in a maximal sharing of *binder* cells among different bindings, and hence, these results provide a measure of the system's performance under conditions of maximal cross-talk. A non-zero value of  $\zeta_{ltd}$  would reduce cross-talk, but it would also lead to a gradual reduction in the number of cells available for recruitment as more and more bindings are memorized.

## 6 Conclusion

A grounding of recruitment learning and vicinal algorithms in the biological phenomena of LTP and LTD has been described. A realization and specification of a vicinal algorithm using LTP has been illustrated by showing how *binder* cells responsive to specific role-entity bindings can be memorized rapidly in response to a transient pattern of activity encoding the bindings. Using biologically plausible values for the number of cells in the ROLE, ENTITY and BIND regions, and the density of projections from the ROLE and ENTITY regions to BIND, it has been shown that the existence of suitable *binder* cells for encoding arbitrary role-entity bindings is practically certain. It has also been shown that the interference between *binder* cells for different bindings remains extremely low, and that the encoding of *binder* cells is robust with respect to cell loss.

The encoding of binding detectors is just one step in the acquisition of episodic memory. As argued in [23, 27] a proper encoding of episodic memory also requires the recruitment of *binding-error detector* circuits, *binding-error*

---

<sup>8</sup>The bases of these calculation are discussed in [26].

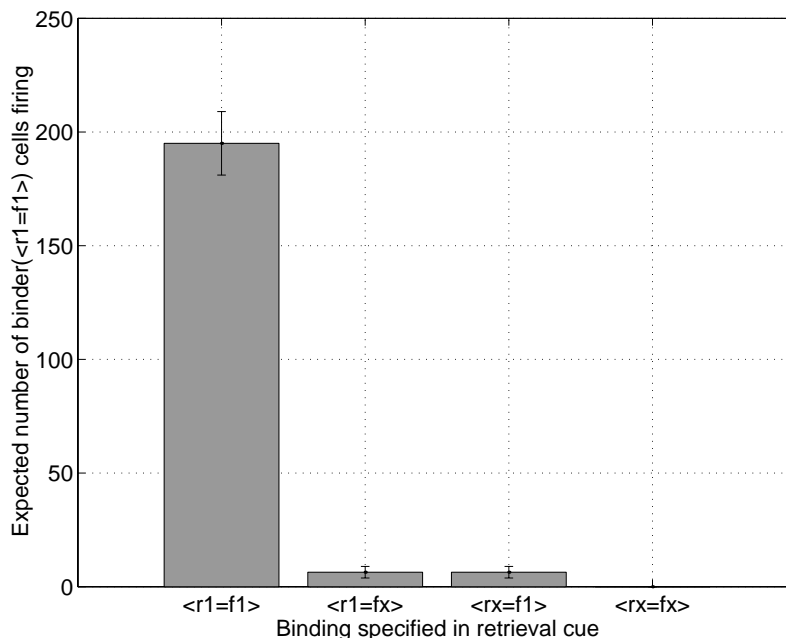


Figure 5: Response of  $binder(\langle r1 = f1 \rangle)$  cells to bindings in a retrieval cue. Here  $fx$  refers to an entity other than  $f1$  and  $rx$  refers to a role other than  $r1$ .

*integrator* cells, *relational instance match indicator* circuits, and *binding-reinstator* cells. Shastri [24, 27] has proposed a model of episodic memory formation that suggests how cells and circuits realizing these functional units can be recruited rapidly via LTP and LTD within quasi-random networks whose architecture and circuitry resembles that of the hippocampal formation. The resulting memory trace can differentiate between highly similar events and respond to partial cues.

Several of the functional units required for encoding an episodic memory trace are more complex than the *binder* cells discussed in this article [24]. The recruitment of these units requires the full-range of features included in the abstraction of LTP and LTD (e.g., spike versus burst firing modes) and further illustrates how interesting vicinal algorithms can arise from a suitable choice of network architecture and parameters for the induction of LTP and LTD.

The investigation of biologically grounded recruitment learning algorithms can increase our understanding of the neural processes underlying memory formation and retrieval. In the long-term, this research may also lead to the design of robust memory modules for autonomous agents, and perhaps, eventually, to the development of memory prosthesis for brain injured humans.

## Acknowledgment

This work was partially funded by NSF grants 9720398 and 9970890.

## References

- [1] Ajjanagadde, V., Shastri, L.: Rules and Variables in Neural Nets. *Neural Computation*, 3 (1991) 121–134.

- [2] Amaral, D.G., Ishizuka, N., Claiborne, B.: Neurons, numbers and the hippocampal network. In *Progress in Brain Research: Understanding the brain through the hippocampus*, J. Storm-Mathisen, J. Zimmer, & O.P. Ottersen (Eds.), 1–11, Elsevier Science, Amsterdam. (1990).
- [3] Artola, A., Singer, W.: Long-term depression of excitatory synaptic transmission and its relationship to long-term potentiation. *Trends in Neuroscience*, 16 (1993) 480–487.
- [4] Bailey, D., Chang, N., Feldman, J., Narayanan, S.: Extending Embodied Lexical Development. In the *Proceedings of the 20th Conference of the Cognitive Science Society*, Madison, WI. (1998) 84–89.
- [5] Bliss, T.V.P., Lomo, T.: Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology*, 232 (1973) 331–356.
- [6] Brown, T.H, Kairiss, E.W., & Keenan, C.L.: Hebbian Synapses: Biophysical Mechanisms and Algorithms. *Annual Review of Neuroscience*, 13 (1990) 475–511.
- [7] Cohen, N.J., Eichenbaum, H.: *Memory, Amnesia, and the Hippocampal System*. MIT Press, Cambridge, Massachusetts (1993).
- [8] Diederich, J.: Instruction and high-level learning in connectionist networks. *Connection Science*, 1 (1989) 161–180.
- [9] Dudek, S.M., Bear, M.F.: Bidirectional Long-Term Modification of Synaptic Effectiveness in the Adult and Immature Hippocampus. *Journal of Neuroscience*, 13 (1992) 2910–2918.
- [10] Feldman, J.A.: Dynamic connections in neural networks. *Bio-Cybernetics*, 46 (1982) 27–39.
- [11] Feldman, J.A., Bailley, D.: Layered hybrid connectionist models for cognitive science. In *Hybrid Neural System*, S. Wermter & R. Sun (Eds.), 14–27, Springer, Heidelberg (2000).
- [12] Levy, W.B. & Steward, O.: Synapses as associative memory elements in the hippocampal formation. *Brain Research*, 175 (1979) 233–245.
- [13] Linden, D.J.: Long-term synaptic depression in the mammalian brain. *Neuron*, 12 (1994) 457–472.
- [14] Malenka, R.C., Nicoll, R.A.: Long-term Potentiation - A Decade of Progress? *Nature*, 285 (1999) 1870–1874.
- [15] Marr, D.: Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society*, B 262 (1971) 23–81.
- [16] Maass, W., Ruf, B.: On Computation with Pulses. *Information and Computation*, 148 (1999) 202–218.
- [17] McNaughton, B.L., Morris, R.G.M.: Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends in Neuroscience*, 10 (1987) 408–415.
- [18] Nicoll, R.A., Malenka, R.C.: Contrasting properties of two forms of long-term potentiation in the hippocampus. *Nature*, 377 (1995) 115–118.
- [19] O’Keefe, J., Nadel, L.: *The hippocampus as a cognitive map*, Clarendon Press, Oxford (1978).
- [20] Page, M.: Connectionist Modeling in Psychology: A Localist Manifesto *Behavioral and Brain Sciences*, 23 (2000), In press.

- [21] Rioult-Pedotti, M.-S., Friedman, D., Donoghue, J.P.: Learning-induced LTP in Neocortex. *Science*, 290 (2000) 533-536.
- [22] Shastri, L.: *Semantic Networks: An evidential formalization and its connectionist realization*. (see p. 181–191). Morgan Kaufmann, Los Altos/Pitman Publishing Company, London (1988).
- [23] Shastri, L.: A Model of Rapid Memory Formation in the Hippocampal System, In the *Proceedings of the 19th Annual Conference of the Cognitive Science Society*, Stanford University, CA, (1997) 680–685.
- [24] Shastri, L.: Recruitment of binding and binding-error detector circuits via long-term potentiation. *Neurocomputing*, 26-27 (1999) 865–874.
- [25] Shastri, L.: Advances in SHRUTI — A neurally motivated model of relational knowledge representation and rapid inference using temporal synchrony. *Applied Intelligence*, 11 (1999) 79–108.
- [26] Shastri, L.: A biological grounding of recruitment learning and vicinal algorithms. Technical Report TR-99-009, International Computer Science Institute, Berkeley, CA. March, 1999.
- [27] Shastri, L.: From transient patterns to persistent structures: a computational model of rapid memory formation in the hippocampal system. In preparation.
- [28] Shastri, L., Ajjanagadde V.: From simple associations to systematic reasoning: A connectionist encoding of rules, variables and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences*, 16 (1993) 417–494.
- [29] Shastri, L., Grannes, D.J., Narayanan, S., Feldman, J.A.: A Connectionist Encoding of Schemas and Reactive Plans. In *Hybrid Information Processing in Adaptive Autonomous vehicles*, G.K. Kraetzschmar and G. Palm (Eds.), Lecture Notes in Computer Science, Springer-Verlag, Berlin (To appear).
- [30] Shastri, L., Wendelken, C.: Seeking coherent explanations – a fusion of structured connectionism, temporal synchrony, and evidential reasoning. In the *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society*, Philadelphia, PA. (2000) 453-458.
- [31] Singer, W.: Synchronization of cortical activity and its putative role in information processing and learning. *Annual Review of Physiology* 55 (1993) 349–74.
- [32] Squire, L.R.: Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review* 99 (1992) 195–231.
- [33] Tang, Y., Shimizu, E., Dube, G.R., Rampon, C., Kerchner, G.A., Zhuo, M., Liu, G., Tsien, J.Z.: Genetic enhancement of learning and memory in mice *Nature*, 401 (1999) 63 - 69.
- [34] Stanton, P.K. & Sejnowski, T.J.: Associative long-term depression in the hippocampus induced by hebbian covariance. *Nature*, 339 (1989) 215-218.
- [35] Treves, A & Rolls, E.T.: Computational analysis of the role of the hippocampus in memory. *Hippocampus*, 4 (1994) 374–391.
- [36] Tulving, E.: *Elements of Episodic Memory*. Clarendon Press, Oxford (1983).
- [37] Valiant, L.: *Circuits of the mind*, Oxford University Press, New York (1994).
- [38] Van Hoesen, G.W.: The primate hippocampus gyrus: New insights regarding its cortical connections. *Trends in Neuroscience*, 5 (1982) 345–350.

- [39] von der Malsburg, C.: Am I thinking assemblies? In *Brain Theory*, ed. G. Palm & A. Aertsen. Springer-Verlag, Berlin (1986).
- [40] West, M.J.: Stereological studies of the hippocampus: a comparison of the hippocampal subdivisions of diverse species including hedgehogs, laboratory rodents, wild mice and men. In *Progress in Brain Research: Understanding the brain through the hippocampus* J. Storm-Mathisen, J. Zimmer, & O.P. Ottersen (Eds.) 13–36, Elsevier Science, Amsterdam (1990).