INTERNATIONAL COMPUTER SCIENCE INSTITUTE

1947 Center St. • Suite 600 • Berkeley, California 94704-1198 • (510) 643-9153 • FAX (510) 643-7684

Rapid learning of binding-match and binding-error detector circuits via long-term potentiation

Lokendra Shastri*

TR-97-037

October 1997

Abstract

It is argued that the memorization of events and situations (episodic memory) requires the *rapid* formation of neural circuits responsive to binding errors and binding matches. While the formation of circuits responsive to binding matches can be modeled by associative learning mechanisms, the rapid formation of circuits responsive to binding errors is difficult to explain given their seemingly paradoxical behavior; such a circuit must be *formed* in response to the occurrence of a binding (i.e., a particular pattern in the input), but subsequent to its formation, it must not fire anymore in response to the occurrence of the very binding (i.e., pattern) that led to its formation. A plausible account of the formation of such circuits has not been offered. A computational model is described that demonstrates how a transient pattern of activity representing an event can lead to the rapid formation of circuits for detecting bindings and binding errors as a result of long-term potentiation within structures whose architecture and circuitry are similar to those of the hippocampal formation, a neural structure known to be critical to episodic memory. The model exhibits a high memory capacity and is robust against limited amounts of diffuse cell loss. The model also offers an alternate interpretation of the functional role of region CA3 in the formation of episodic memories, and predicts the nature of memory impairment that would result from damage to various regions of the hippocampal formation.

*This work was partially funded by ONR grant N00014-93-1-1149. Thanks to Jerry Feldman, Steve Greenberg, Robert Knight, Art Shimamura, and members of the NTL and SHRUTI projects at ICSI for discussions. *E-mail:* shastri@icsi.berkeley.edu

1 Introduction

We often remember our experiences in terms of events and situations that 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 has been variably referred to as *episodic* (Tulving, 1983), *declarative* (Squire, 1992; Cohen & Eichenbaum, 1993), *explicit* (Graf & Schacter, 1985), and *locale* memory (O'Keefe & Nadel, 1978) and it is well known that the hippocampal formation (HF) serves a critical role in its formation (Squire, 1992; Cohen & Eichenbaum 1993; Treves & Rolls, 1994). A remarkable feature of such episodic memory is that it is formed rapidly — typically as a result of a single occurrence.

An event or a situation can be viewed as a *relational instance* consisting of a collection of *bindings* between the *roles* of a generic relation and the *entities* that fill these roles in the given instance.[1] For example, the event "John gave Mary a book on Tuesday in the library"[2] may be expressed as an instance of the generic relation GIVE with the following *role*-`entity` bindings:

($\langle$*giver*=`John`$\rangle$,
$\langle$*recipient*=`Mary`$\rangle$,
$\langle$*give-object*=`a-Book`$\rangle$,
$\langle$*temporal-location*=`Tuesday`$\rangle$,
$\langle$*location*=`Library`$\rangle$)

As discussed below, the persistent encoding of an event or a situation, i.e., its *memory trace*, must satisfy several representational requirements. In particular, it must be capable of detecting matches as well as mismatches (errors) between bindings in the memorized event and those in a cue. Hence the memorization of an event or a situation must lead to rapid changes in neural circuits that enable the subsequent detection of bindings and binding errors by episodic memory.

While the formation of neural circuits responsive to binding matches can be readily modeled by associative learning mechanisms, the rapid formation of neural circuits responsive to binding errors is more difficult to explain in view of their seemingly paradoxical behavior; such a circuit must be *formed* in response to the occurrence of a binding, but subsequent to its formation, it must *not* respond anymore to the occurrence of the very binding that led to its formation.

This article describes a computational model that demonstrates how a transient pattern of activity representing an event or situation can lead to the rapid formation of circuits for detecting binding errors and binding matches as a result of long-term potentiation (LTP) (Bliss & Collingridge, 1993; Lynch & Ambros-Ingerson, 1993) and long-term depression (LTD) (Linden, 1994; Artola & Singer, 1993; Derrick & Martinez, 1996) within structures

[1] The conception of events and situation as collections of role-entity bindings is well founded. Extensive work in cognitive science and linguistics has demonstrated how declarative aspects of conceptual knowledge can be expressed in terms of suitable relational structures composed of role-entity bindings such as frames, schemas, and predicates (e.g., see Jackendoff, 1990; Pinker, 1989).

[2] This sentence is simply meant to be a *description* of an event. An agent may directly experience such an event (say via the visual modality) or be informed of its occurrence (e.g., via verbal or written use of language).

whose architecture and circuitry are similar to those of the HF. The proposed model exhibits a high capacity for memorizing bindings and is robust against diffuse cell loss.

The proposed model offers an alternative interpretation of the functional role of the dentate gyrus (DG) and hippocampal region CA3. While most models of the HF view CA3 as an associative or auto-associative memory (Marr, 1971; Treves & Rolls, 1994; Hasselmo, 1997; O'Reilly & McClelland, 1996), this work suggests that a key representational role of CA3 in humans might be the detection of binding errors. The model also makes several predictions about the nature of memory impairment that would result from focal damage to specific regions of the HF.

The structure for the formation of circuits for detecting bindings and binding errors described here has been embedded within SMRITI, a detailed model of episodic memory formation (Shastri, 1997a, 1997b). The complete model demonstrates how the HF might rapidly encode an event or a situation as a persistent structure capable of supporting rapid recognition and recall.[3]

The significance of circuits responsive to binding errors described below and the proposed manner in which such circuits can be formed rapidly extends beyond the encoding of episodic memories. A binding-error detector circuit can perform the generic function of *coincidence error* detection, since such a circuit is formed when two patterns A and B occur concurrently, and once formed, it fires whenever A occurs without being accompanied by B. The firing of such circuits can also signify a failure of expectation, and hence, such circuits can form the basis of a system for novelty detection, a function attributed to the HF (Knight, 1996).

1.1 Episodic memory traces must be responsive to binding errors as well as bindings matches

We construe and memorize our experiences as events and situations that describe who did what to whom where and when, or describe states of affairs wherein multiple entities occur in particular configurations. Such construals can be viewed as relational instances consisting of a collection of bindings between the roles of a generic relation and the entities that fill these roles in a given situation.[4]

The fact that an event or a situation is essentially a collection of role-entity bindings gives rise to a number of representational requirements. First, the memory trace of an event or situation (henceforth, an event) must encode, and subsequently respond to, specific role-entity bindings. A memory trace that only associates various entities appearing in an event but does not encode which entity fills which role will not function properly since it would be unable to distinguish between distinct events such as "John gave Mary a book" and "Mary gave John a book" if they involve the same set of entities. Hence the memory trace of an event must incorporate circuits for detecting role-entity bindings.

Second, the memory trace of an event must also recognize partial cues. For example, the encoding of the event "John gave Mary a book on Tuesday in the library" should

[3] Here recognition refers to the processing of *yes-no* questions such as "Did John give Mary a book?" and recall refers to the processing of *wh* questions such as "What did John give Mary?"

[4] Note that two levels of bindings are involved: (1) entities occurring in an event must be bound to the respective roles they fill in the event, and (2) all the role-entity bindings pertaining to an event must be grouped together (i.e., contextualized) and distinguished from role-entity bindings pertaining to other events.

respond positively to the cue: "Did John give someone a book?" and "Did someone give Mary something in the library?" Third, the encoding of an event should not match a cue that specifies an incompatible binding even if the cue contains several other bindings that match the memorized event. For example, the encoding of the event "John gave Mary a book on Tuesday in the library" should not match a cue such as "Did John give Susan a book on Tuesday in the library?" even though the latter has a high degree of match with the memorized event. Observe that the cue refers to the same generic relation as the memorized event and shares 4 out of 5 bindings with the latter. Yet we would expect our episodic memory system to be capable of treating these two patterns as distinct events. The integrity of episodic memory critically depends on its ability to support this strong form of *pattern separation*, whereby it can distinguish between events that differ in a single role-entity binding even if they share a number of other role-entity bindings. The issue of pattern separation has been addressed by a number of researchers (e.g., Marr, 1971; McNaughton & Morris 1987; O'Reilly & McClelland, 1996). These models, however, do not view patterns as collections of role-entity bindings and do not attempt to separate patterns based on a single binding mismatch.

The requirement that a memory trace respond to highly partial cues but at the same time reject a highly similar cue if it specifies an incompatible binding entails that the memory trace of a relational instance must be capable of detecting binding *errors* as well as binding matches. This requirement cannot be satisfied by an encoding that only detects binding matches since such an encoding would be incapable of distinguishing between an *unspecified* binding and an *incorrect* binding. Consider any encoding of $(R_1 : \langle r_1 = a\rangle, \langle r_2 = b\rangle)$ that can detect *only* binding matches. To such an encoding, the partial pattern $(R_1 : \langle r_1 = a\rangle)$ and the erroneous pattern $(R_1 : \langle r_1 = a\rangle, \langle r_2 = c\rangle)$ would appear identical since both contain one matching binding.[5]

Associative memories versus episodic memories

The ability to distinguish between similar instances on the basis of incompatible bindings should be contrasted with the "error correction" ability of associative memories (e.g., Willshaw, Buneman, & Longuet-Higgins, 1969; Hopfield, 1982; Hinton & Anderson, 1981) which enables them to match *similar*, though *distinct*, patterns. These two abilities are complementary; while both associative and episodic memories must respond to partial cues, the strict separation of distinct, though similar, patterns is appropriate for the episodic memory system and the matching of similar, though distinct, patterns (error correction) is suitable for similarity driven *taxon* memory systems (O'Keefe & Nadel, 1978).

Episodic memories in concert with other forms of memories give rise to flexible memory retrieval

Though episodic memory only encodes specific events and situations and enforces strict pattern separation it can support flexible memory retrieval in concert with other forms of memory while still maintaining strict pattern separation. For example, episodic memory

[5] A strictly conjunctive encoding can only match complete patterns, and hence, can only perform recognition. While an encoding that uses soft (or graded) conjunction can respond to partial cues, it cannot reject cues that contain incompatible bindings.

operating in tandem with a memory system for representing categories and their superordinate and subordinate relationships can match the cue "Did John buy a car?" to the memorized event "John bought a Rolls Royce" because the category representation system can map "a car" to "a Rolls Royce" (since a Rolls Royce is a type of car). The interaction between these two memory systems allows a flexible specification of a role-filler but retains the ability to reject any cue that specifies an incorrect binding. Furthermore, such a combined memory system can also memorize facts such as "Kangaroos live in Australia" in which roles are filled by generic categories rather than by individual entities.

1.2 Rapid learning of circuits responsive to binding errors is problematic

As discussed above, the proper functioning of episodic memory requires circuits responsive to binding errors as well as circuits responsive to binding matches. But while the rapid learning of circuits for detecting binding matches can be explained by associative and recruitment learning mechanisms (see below), the rapid learning of circuits responsive to binding errors is problematic given their paradoxical functional behavior and a plausible account of their formation has not been offered.

In order to see why the formation of a binding-error detector circuit is problematic, consider the memorization of an event wherein the role r is bound to the entity f. The memory trace formed in response to this event must include circuits for detecting an error in the binding $\langle r = f \rangle$. But this poses a problem. Such a binding-error detector circuit must be learned in response to the concurrent activation of r and f in the transient encoding of the observed event. But subsequent to its formation, such a circuit must *not* fire anymore in response to the concurrent activity of r and f — *the very activity that led to its formation* (since the occurrence of such an activity signifies a binding match and not a binding error).

In the proposed model, a binding-error detector circuit for the binding $\langle r = f \rangle$ is required to fire whenever r fires without the coincident firing of f. The rationale for this behavior is as follows: The firing of r signifies that the role r is "active" and has been bound to some entity. But the absence of coincident firing of f signifies that the entity f is not bound to r. Hence the firing of r without the coincident firing of f means that r has been bound to some entity other than f. This constitutes a binding error with respect to r and should lead to the firing of the binding-error detector circuit for the binding $\langle r = f \rangle$. The lack of firing of r, on the other hand, simply means that no binding has been specified for r. Consequently, the firing of f without the firing of r does not constitute a binding error with respect to r and should not lead to the firing of the circuit.

1.3 Computational models of rapid learning in random nets

A number of computational models of memory formation in the HF have been proposed dating back to Marr (1971) (the special issue of *Hippocampus*, **6** (6), 1996, and references therein, provide pointers to a number of prominent models). These models have greatly contributed to our understanding of potential hippocampal function and lead to a number of insights about associative memories, but they have not addressed some of the representational problems discussed in Section 1.1. These models typically view a memorized item as a "pattern" or a feature vector and not as a collection of role-entity bindings, and they do

not make a sharp distinction between associative and episodic memory functions outlined above.

The rapid formation of chunking and binding "nodes" has been modeled within a computational framework by several researchers (e.g., Feldman, 1982; Shastri, 1988; Diederich, 1989; Valiant, 1994; Chover, 1996), but an explanation of how circuits for detecting binding errors may be formed rapidly has not been proposed. Feldman (1982) showed that a "conjunctive" concept consisting of several constituents distributed over a randomly connected neural net may be learned by binding these constituents together via a chunking node that happens to receive links from all the constituents (see Wickelgren, 1979). Feldman referred to this form of learning within random networks as "recruitment learning" and showed that suitably connected chunking nodes may be found with high probability by making appropriate assumptions about network connectivity. Shastri (1988) extended the notion of recruitment learning to structured concepts. He treated a concept as a collection of attribute-value bindings (akin to roles-filler bindings) and suggested a two stage memorization process. In the first stage, "binder" nodes are recruited for each attribute-value binding of a concept. In the second stage, these binder nodes are chunked together by the recruitment of a concept node. Diederich (1989) showed how this form of structured recruitment learning can be used to learn new concepts expressed as modifications of existing concepts. Valiant (1994) described several algorithms for recruitment learning of conjunctive concepts as well as concepts that are collections of role-entity bindings and presented a quantitative analysis of these algorithms using plausible assumptions about connectivity in the neocortex. Chover (1996) described a system for associating stimulus and response pairs using a correlation layer that receives links from, and sends links back to, stimulus and response cells.

All of the above models, however, only deal with "positive knowledge", that is, the structures recruited by these models incorporate nodes that can actively signal a binding match, but they do not include nodes or circuits that can *actively signal* a binding mismatch. Consequently, these structures cannot support the strong form of pattern separation required for the proper functioning of episodic memory whereby the memory trace of a memorized event does not respond to any cue that specifies an erroneous role-entity binding, even if the cue shares a large number of other role-entity bindings with the memorized instance.[6]

1.4 Outline of the paper

Section 2 describes the computational abstraction of cells and synapses used in the proposed model. Section 3 describes the overall architecture of the composite structure that supports the formation of binding detectors and binding-error detectors, and Section 4 specifies how role-entity bindings associated with an event are expressed as a transient pattern of rhythmic activity. The internal architecture of region BIND and BED and the rapid

[6] A number of researchers have proposed formal as well as computational models for representing bindings using various types of convolution, tensor product, matrix multiplication, and XOR operations (see Plate, 1994 for an extensive review). These models also fall within the general category of associative memory models and incorporate both error correction and pattern separation properties within the same system. Consequently, they do not exhibit the strong pattern separation property discussed above. These models also differ from the proposed model in that they are specified at a relatively abstract computational level and their mapping to neurally plausible circuits and structures is not specified.

formation of binding detectors and binding-error detectors in these region in response to such transient rhythmic activity is described in Sections 5 and 6, respectively. Section 7 identifies the correspondence between the proposed model and the architecture and circuitry of the hippocampal formation, and Section 8 situates the proposed model in the context of SMRITI, a model of episodic memory formation in the hippocampal system. A quantitative analysis of the model's memory capacity, and its robustness against diffuse cell loss and cross-talk are presented in Section 9. Section 10 lists several predictions of the model and Section 11 makes some general observations.

2 LTP, LTD and the emergence of functionally selective cells

LTP and LTD refer to long-term activity dependent changes in synaptic strength and are believed to underlie memory formation (Bliss & Collingridge, 1993; Lynch & Ambros, 1993; Linden, 1994; Artola & Singer, 1993).

LTP typically results from the pairing of presynaptic activity with postsynaptic depolarization. Its induction involves the unusual receptor NMDA which is activated by the neurotransmitter glutamate, but only if the postsynaptic membrane is already depolarized. Once the NMDA receptor is activated, calcium ions flood into the postsynaptic cell and lead to a complex series of biochemical changes that result in the induction of LTP.[7] The two conditions required for the activation of the NMDA receptor can be brought about by a brief but high frequency burst of activity, or by multiple converging inputs arriving at a cell in close temporal proximity. The long-term increase in the efficacy of a synapse resulting from brief but high-frequency activity at the synapse is referred to as *homosynaptic* LTP. The long-term increase in efficacy of synapses resulting from convergent activity arriving at multiple synapses sharing the same postsynaptic cell is called *associative* LTP.

In addition to potentiation, synapses can also undergo long-term depression (LTD). A synapse receiving no presynaptic activity can undergo heterosynaptic LTD if other synapses of the same postsynaptic cell receive high frequency presynaptic activity. 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. Finally, prolonged low frequency stimulation of a synapse can lead to its homosynaptic LTD.

The proposed model for the formation of binding and binding-error detector circuits uses idealized forms of associative LTP and heterosynaptic LTD. These idealizations are described below. The use of associative LTD resulting from out-of-phase activity (Stanton & Sejnowski, 1989) was investigated in one version of the model. Although this form of LTD is not essential for the formation of binding and binding-error detectors, it does reduce the probability of the formation of *ambiguous* detectors (these are discussed below). The model described below does not deal with reversal of LTP and LTD and assumes that only naive synapses, that is, synapses that are neither potentiated nor depressed, can undergo LTP and LTD. The reversal of LTP is important for modeling forgetting and is considered in the detailed episodic memory model described in Shastri (1997b).

[7] The LTP of synapses formed by mossy fiber on CA3 pyramidal cells has a different biological basis than the NMDA based LTP mentioned above (Zalutsky & Nicoll, 1990).

2.1 Computational modeling of LTP and LTD

Computational abstraction of cell behavior

A cell is modeled as a highly idealized integrate-and-fire neuron. The goal has been to keep the model deliberately simple and discrete in order to facilitate its analysis and simulation, while still capturing some essential temporal aspects critical for modeling LTP, LTD, and the expression of dynamic bindings via temporal synchrony.

The spatio-temporal integration of activity arriving at the synapses of a (postsynaptic) cell is modeled as follows. Each synapse, s_i, of the cell has a weight, $w_i(t)$ (see below). The postsynaptic potential, $psp_i(t)$, resulting from the arrival of presynaptic activity,[8] $a_i(t)$, at s_i, at time t, is given by $a_i(t) * w_i(t)$.

Each psp_i is assumed to have a life-time of ω. In other words, the influence of activity $a_i(t)$ arriving at time t_0 persists from time t_0 to $t_0 + \omega$. Here ω may be viewed as the window of integration, or the window of synchrony which measures the maximum amount by which two incident activities may lead/lag and still be treated as being synchronous by the postsynaptic cell. That is:

$$psp_i(t) = \sum_{(1 \leq \tau \leq \omega)} a_i(t-\tau) w_i(t-\tau)$$

The combined effect of presynaptic activity on the cell is modeled as the potential, $pot(t)$, which equals the sum of individual $psp_i(t)$'s. That is:

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

where i ranges over all synapses of the cell. We will use "potential" and "weighted sum of activity" interchangeably.

Each cell has a *firing threshold*, θ_f. A cell fires at time t, if $pot(t) \geq \theta_f$, and produces an output which arrives at synapses downstream from the cell at time $t+d$ (d is the propagation delay and is usually assumed to be 1). The activity arriving at time $t + d$ at a synapse of a postsynaptic cell is the same as the output of the presynaptic cell at time t.

Synapses

Synapses formed by the same projection are assumed to have similar attributes (see below). This is modeled by associating a distinct synaptic *type* with all the synapses formed by a given projection.

A synapse can be in one of three *states*, namely, *naive*, *potentiated*, or *depressed*. The weights of all synapses of a particular type, and in a given state, are assumed to lie within a restricted band. For a given synaptic type, the weight bands associated with the three states are assumed to be disjoint. The weight bands associated with a synaptic state may, however, vary from one synaptic type to another. For example, the weights of naive synapses formed on principal cells in BED by afferents from BIND may have a higher range of values than the weights of naive synapses formed by afferents from ROLE.

[8] Such "activity" may correspond to a single action potential or a brief high-frequency burst of spikes.

Computational modeling of LTP

The occurrence of LTP is governed by the following parameters: the *potentiation threshold* θ_p, the *weight increment* Δw_{ltp}, the *repetition factor* κ, and the *maximum inter-activity interval* τ_{iai}.

Consider a set of neighboring[9] synapses $s_1, \ldots, s_n$ sharing the same postsynaptic cell. Convergent presynaptic activity at $s_1, \ldots, s_n$ can lead to associative LTP of naive s_i's and increase their weights by Δw_{ltp} if the following conditions hold:

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

   The above entails that the presynaptic activity arriving at $s_1, \ldots, s_n$ is synchronous, that is, the maximum lead/lag in incident arriving at any pair of synapses is no more than ω.

2. Such synchronous presynaptic activity recurs (repeats) at least κ times.

3. The interval between two *successive* arrivals of presynaptic activity at a synapse during the above repetition is at most τ_{iai} time units.

Computational modeling of LTD

Heterosynaptic LTD is modeled in an analogous manner using five parameters. These are: the potentiation threshold θ_p, the weight decrement Δw_{ltd}, the repetition factor κ, the maximum inter-activity interval τ_{iai}, and the *ubiquity of LTD* ζ. 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 Δw_{ltd}. As in the case of LTP, θ_p dictates the minimum weighted sum of synchronous activity that neighboring synapses of the postsynaptic cell must receive, and κ specifies the number of times such presynaptic activity must recur in order to induce heterosynaptic LTD of naive inactive synapses. Also as before, τ_{iai} specifies the maximum permissible gap between the successive arrival of presynaptic activity. The parameter ζ specifies the fraction of inactive naive synapse that undergo LTD when the above conditions are met. Thus ζ provides a simple computational mechanism for controlling the ubiquity 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.

The proposed model can be realized with different values of ζ in the interval $[0, 1]$. Section 9 presents quantitative results that point out the differences in the model's behavior stemming from different choices of ζ. A variant of the proposed model uses a synaptic modification rule similar in spirit to the BCM rule (Bienenstock, Cooper, & Munro, 1982). In this variant of the model, the potentiation threshold of a postsynaptic cell is variable and increases upon the potentiation of its synapses. Hence, the potentiation of a certain number of synapses can raise the potentiation threshold of the postsynaptic cell to a sufficiently high level and thereafter, make the potentiation of its remaining naive synapses very

[9] In the model described here, all synapses sharing the same postsynaptic cell are assumed to be neighboring synapses. In subsequent work we plan to use the notion of neighborhood in a non-trivial manner.

unlikely. In computational terms, assuming a steep increase in the potentiation threshold of a postsynaptic cell after the potentiation of a small number of its synapses is analogous to assuming $\zeta \approx 1$ (i.e., widespread heterosynaptic LTD), whereby the LTP of a small number of synapses can lead to the LTD of all remaining naive synapses.

Emergence of functionally selective cells and circuits

In the proposed model, a cell receives a number of afferents, and hence, can potentially participate in a number of functional circuits. If however, the weights of selected synapses of the cell increase via LTP and the weights of other synapses decrease via LTD, the cell can become highly selective and participate in only a small number of functional circuits. Thus LTP and LTD provide a promising neural mechanism for the formation of functional structures within quasi-random networks.

3 A structure for the formation of circuits responsive to bindings and binding errors

The composite structure that supports the rapid formation of circuits responsive to binding matches and binding errors consists of four regions: ROLE, ENTITY, BIND and BED (see Figure 1(a)). Region ENTITY projects to region BIND, region ROLE projects to both regions BIND and BED, and region BIND projects to region BED. All these projections are diffuse and dense with the exception of the projection from BIND to BED which is diffuse but sparse. Each role and entity is encoded by a small ensemble of cells in the ROLE and ENTITY regions, respectively. Cells within an ensemble are dispersed within a region and each cell belongs to at most one ensemble. A schematic of the projections from ROLE and ENTITY to BIND is shown in Figure 1(b)).

There is a direct correspondence between the model structure described above and the HF. Thus ROLE and ENTITY regions correspond to subregions of the entorhinal cortex (EC), the BIND region corresponds to the dentate gyrus (DG), and the BED region to field CA3 of the hippocampus. The projections from high-level cortical areas to ROLE and ENTITY correspond to the well known cortical projections to EC (Van Hoesen, 1982; Insausti, Amaral & Cowan, 1987; Suzuki & Amaral, 1994). The dense and diffuse projections from ROLE and ENTITY to BIND and BED correspond to the dense and diffuse projections along the perforant path from EC to DG and CA3, respectively (Amaral, Ishizuka & Claiborne, 1990; Amaral & Witter, 1995). Similarly, the sparse but diffuse projection from BIND to BED corresponds to the sparse mossy fiber projection from DG to CA3.

4 The transient representation of role-entity bindings

The model assumes that an experience construed as a relational instance is expressed as a transient pattern of rhythmic activity over distributed high-level cortical circuits (HLCCs). These HLCCs project to cells in ENTITY and ROLE regions and, in turn, induce transient patterns of rhythmic activity within these regions. Figure 2 is an idealized depiction of the transient activity induced in ENTITY and ROLE regions by HLCCs to convey the relational

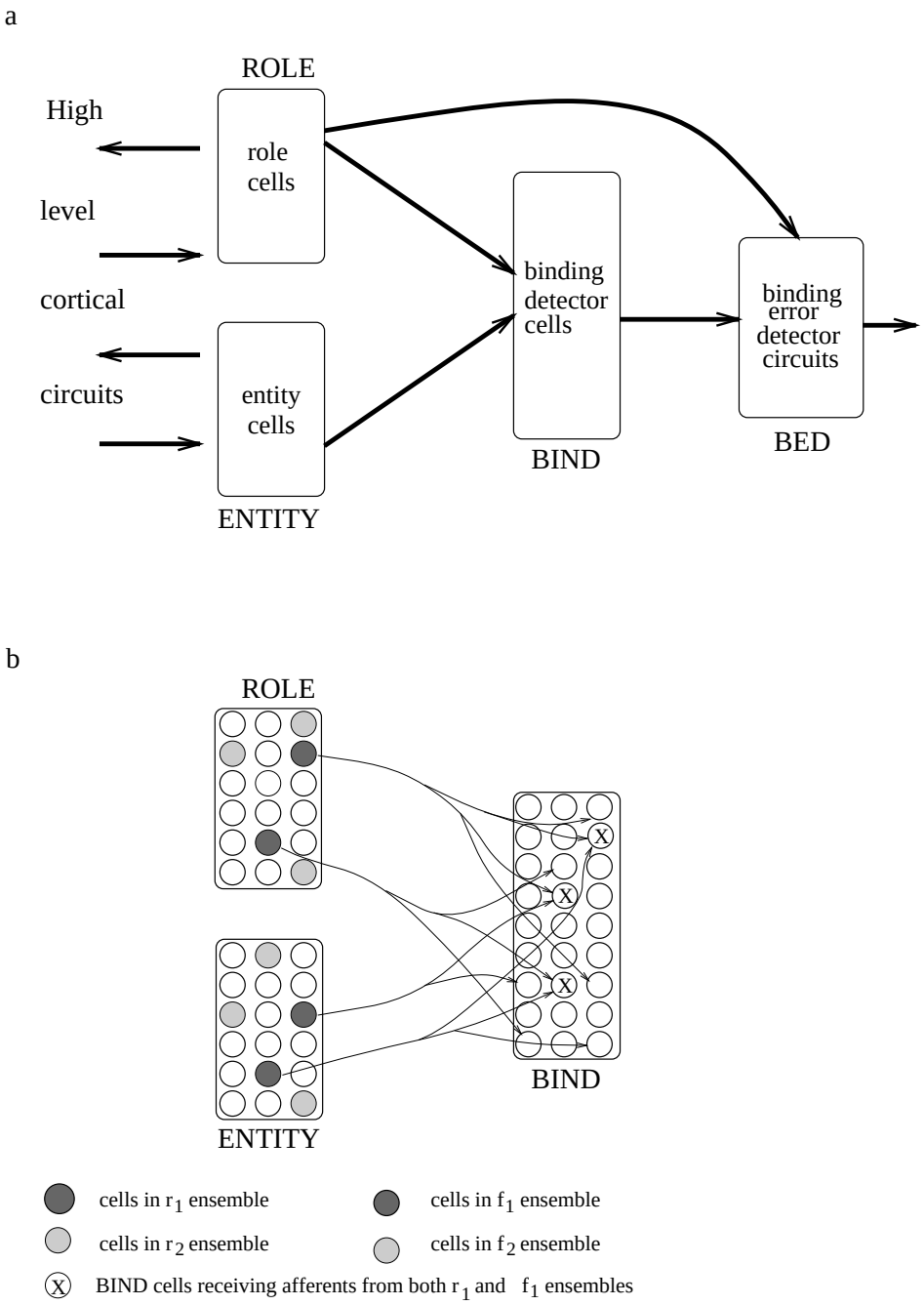

Figure 1: (a) The composite structure for the formation of circuits for detecting bindings and binding errors. Arrows indicate projections. All projections are diffuse and dense with the exception of the projection from BIND to BED which is diffuse but sparse. Each role and entity is encoded by a small ensemble of cells in the ROLE and ENTITY regions, respectively. These ensembles are dispersed and non-overlapping. Binding detector cells and binding-error detector circuits are formed in regions BIND and BED, respectively. (b) This schematic depicts the cell 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 depicted.

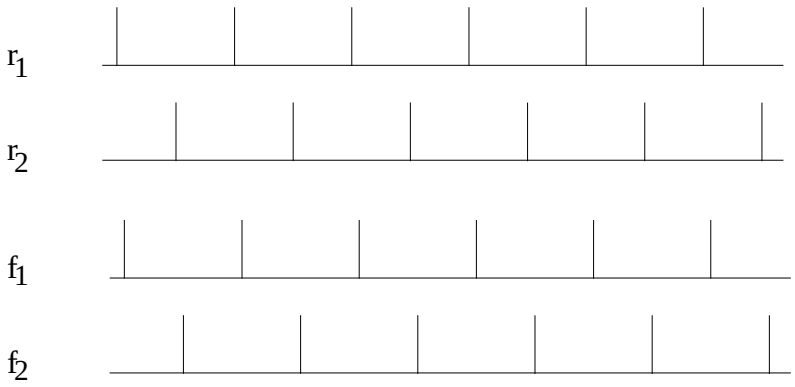

Figure 2: The transient encoding of a relational instance RI given by: $(\langle r_1 = f_1 \rangle, \langle r_2 = f_2 \rangle)$. Here r_1 and r_2 are roles, and f_1 and f_2 are entities bound to r_1 and r_2, respectively. Each spike in the illustration signifies the synchronous firing of a cell ensemble. Cells in the r_1 and f_1 ensembles fire in synchrony and so do cells in the r_2 and f_2 ensembles.

instance RI: $(\langle r_1 = f_1 \rangle, \langle r_2 = f_2 \rangle)$. In the above, 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 quasi-synchronous firing of a cell ensemble. 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. Thus a role-entity binding is expressed by the synchronous firing of the cell ensembles associated with the bound role and entity (von der Malsburg, 1986; Ajjanagadde & Shastri, 1991; Singer & Gray, 1995). In general, the transient encoding of a relational instance with n distinct entities participating as role-fillers involves n interleaved quasi-periodic activities. Such a spatio-temporal encoding enables multiple role-entity bindings to be expressed and propagated concurrently without cross-talk (Shastri & Ajjanagadde, 1993).[10] The proposed model explains how such a transient encoding of a relational instance may be transformed rapidly into persistent circuits for detecting bindings and binding errors via LTP and LTD.

5 Internal structure of BIND and the formation of cells for detecting binding matches

BIND contains two kinds of cells: principal cells and Type-1 inhibitory interneurons. Each principal cell receives afferents from a number of cells in ROLE and ENTITY regions and makes

[10] As discussed in Shastri & Ajjanagadde, (1993), $n \leq \pi/\omega$, where $1/\pi$ is the lowest frequency at which synchronous propagation of activation can be sustained and ω is the window of synchrony, that is, the maximum amount by which two incident activities may lead/lag and still be treated as synchronous by the postsynaptic cell. It was speculated that the transient expression of role-entity bindings involves γ band activity, and hence, π varies roughly between 15—35 milliseconds with smaller values of π being associated with smaller values of π (cf. Lisman & Idiart, 1995).

The presentation of an event to the HF involves a repetitious presentation of the bindings pertaining to the event and such repeated blocks may correspond to θ band activity. Thus the transient expression of an event may involve both γ and θ band activity; the former for expressing bindings and the latter for grouping together the bindings pertaining to an event.

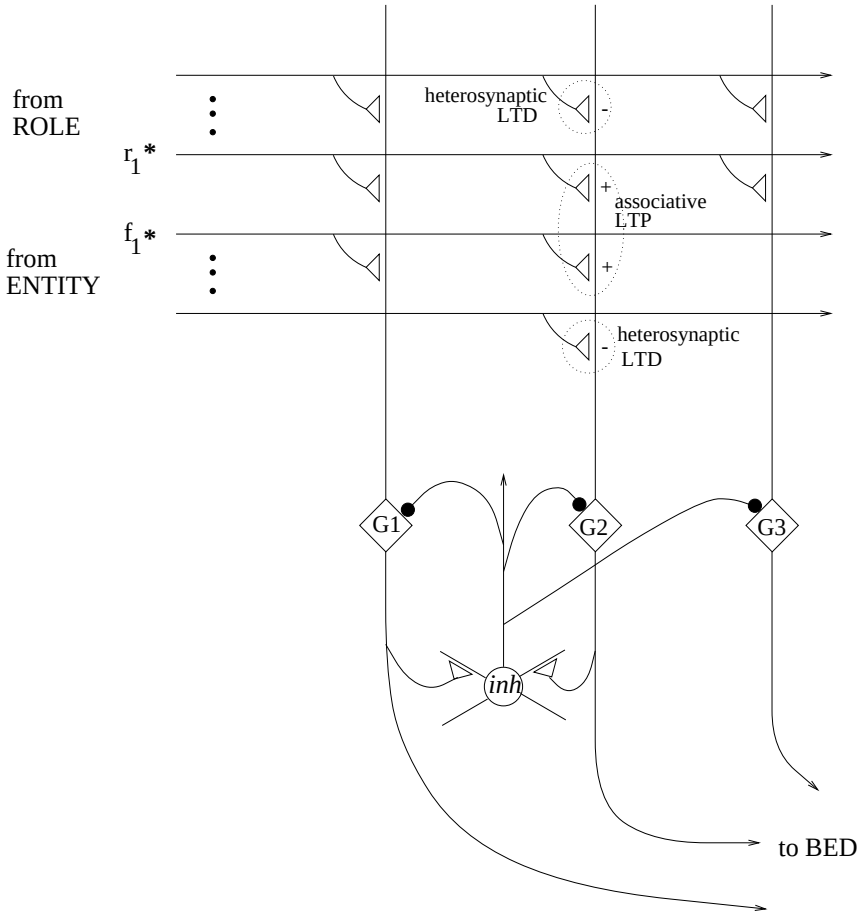

Figure 3: Internal structure of BIND. The region consists of principal cells and inhibitory interneurons (Type-1) and receives dense and diffuse afferents from ROLE and ENTITY regions. Principal cells and Type-1 interneurons form feedback and feedforward inhibitory circuits that limit the number of cells whose synapses undergo LTP. A principal cell receiving synchronous activity along afferents from role and entity cells can be recruited to be a binding detector for the role-entity pair. In the illustration, G1—G3 are principal cells and inh is a Type-1 interneuron. Afferents 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 $\langle r_1 = f_1 \rangle$. In general several cells are recruited as binding detector cells for each role-entity binding. This illustration assumes that the weighted sum of activity arriving along afferents from r_1 and f_1 is at least θ_p.

synaptic contacts on a number of interneurons. The interneurons in turn make contacts on a number of principal cells, thereby forming feedback and feedforward inhibitory circuits within BIND (see Figure 3).

The potentiation threshold, θ_p, of principal cells is sufficiently high, and hence, LTP of a synapse occurs only if multiple synapses of the postsynaptic cell receive coincident presynaptic activity. Moreover, the firing threshold, θ_f, of principal cells is such that a cell does not fire unless it receives impulses at multiple potentiated synapses (a possible set of values for θ_p, θ_f, and synaptic weights of naive and potentiated synapses is given in Section 9).

The transient encoding of the relational instance RI shown in Figure 2 leads to the following events in BIND (refer to Figure 3). 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 afferents from *both* r_1 and f_1 cells.[11] At the same time, some of the inactive naive synapses of these principal cells 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 detector cells for the binding $\langle r_1 = f_1 \rangle$ and we will refer to such cells as $\langle r_1 = f_1 \rangle$ cells.

Since θ_f of principal cells is such that a cell does not fire unless it receives impulses at multiple potentiated synapses, impulses arriving along naive synapses formed by afferents from ROLE and ENTITY do not lead to the firing of a $\langle r_1 = f_1 \rangle$ cell. Similarly, isolated activity arriving at a potentiated synapse from a ROLE or an ENTITY cell does not lead to the firing of a $\langle r_1 = f_1 \rangle$ cell. However, the coincident arrival of impulses at potentiated synapses from role and entity cells satisfies θ_f and causes such a cell to fire in close temporal proximity of presynaptic activity. Thus a $\langle r_1 = f_1 \rangle$ cell fires when r_1 and f_1 cells fire in synchrony and behaves as a binding detector cell for the role-entity binding $\langle r_1 = f_1 \rangle$.

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

Deviations from the norm

The process by which binding detector cells are formed is susceptible to four potential problems. First, in order to form binding detector cells $\langle r_i = f_j \rangle$, there should exist cells that receive afferents from both r_i and f_j cells. Given the quasi-random nature of connectivity this cannot be *guaranteed*, and hence, it is possible that the process may fail to commit any binding detector cells in BIND for a given binding.

Second, there may exist a cell that receives sufficient activity ($\geq \theta_p$) along afferents from r_i cells alone, but which does not receive any afferents from f_j cells. Since there is sufficient synaptic activity, the synapses formed on this cell by the afferents from r_i cells may undergo LTP. If this happens, this cell may behave as a *spurious* binding detector cell

[11] As discussed below, LTP may also occur at synapses of a postsynaptic cell that receives afferents only from r_1 cells or only from f_1 cells. However, the probability of this happening is sufficiently small because LTP requires coincident activity at several synapses of a postsynaptic cell and the probability that *all* of these synapses receive afferents only from r_1 cells, or only from f_1 cells, is very small (see Section 9).

and fire in response to the firing of r_i cells alone, even if there is no coincident activity of f_j cells. Similarly, a cell receiving sufficient activity along afferents from f_j cells and none from any of the r_i cells may undergo LTP and behave as a spurious binding detector cell.

Third, an *ambiguous* binding detector may be formed during the memorization of a relational instance if the same cell receives afferents from role and entity cells corresponding to more than one role-entity pairs bound in the relational instance being memorized.

Fourth, the same cell may get recruited as a binding detector cell for multiple bindings. Under circumstances mentioned below, this may *compromise* the binding-detector function of the cell and cause it to produce false-positive responses. Consider a principal cell that has been recruited as a binding detector cell for the bindings $\langle r_i = f_k \rangle$ and $\langle r_i = f_l \rangle$ which involve the same role r_i but distinct entities f_k and f_l. This cell will behave as the detector for the binding $\langle r_i = (f_j \vee f_k) \rangle$. In other words, it will fire in response to the binding $\langle r_i = f_k \rangle$ as well as the binding $\langle r_i = f_l \rangle$. Such promiscuous firing will, in turn, cause binding-detector circuits downstream in BED to malfunction since they will receive the same binding-match signal in two different situations. The functioning of a binding detector cell $\langle r_i = f_k \rangle$ can also be compromised if it gets recruited as a binding detector for $\langle r_j = f_l \rangle$, where $r_i \neq r_j$, provided r_i and r_j are roles of the *same* generic relation.[12]

As discussed in Section 9, the probability of occurrence of the above problems is extremely small if we make plausible assumptions about the nature of connectivity between regions ROLE, ENTITY and BIND, the internal architecture of these regions, and the values of θ_p and synaptic weights of naive and potentiated synapses. For example, consider the problem of finding suitably connected cells to serve as binding detectors for a given role-entity binding. As discussed in Section 9, the probability of *not* finding a suitable cell is vanishingly small if one assumes plausible values for the size of projective fields (PFs) of role and entity cells.[13] This low failure probability however, has a potential drawback because PF sizes that practically guarantee the existence of suitably connected cells also entail that the expected number of such cells will be very large. This is undesirable since it could lead to too many cells being recruited for each role-entity binding. While the recruitment of several cells to perform a given function has obvious advantages related to redundancy and robustness, the recruitment of too many cells is undesirable since it increases the probability that a given cell will become a binding detector cell for multiple bindings. As discussed above, such sharing of cells can increase the probability that a cell will malfunction. This problem of over recruitment however, can be partially alleviated by inhibitory feedback and feedforward local circuits formed by principal cells and inhibitory interneurons because these circuits act as soft-winner-take-all networks (soft-WTA) and only allow synapses of a limited number of cells — from among the large pool of candidate cells that receive afferents from appropriate role and entity cells — to undergo LTP (cf. Marr, 1971; Feldman, 1982; McNaughton & Morris, 1987).

[12]The episodic memory system is expected to receive a query or a cue pertaining to a single generic relation at a time. Consequently, potential cross-talk between binding detectors pertaining to roles of two different generic relations is harmless.

[13]The projective field of a cell c refers to the collection of target cells that receive afferents from c.

6 The internal structure of BED and the formation of circuits responsive to binding errors

BED contains principal cells and two types of inhibitory interneurons (Type-1 and Type-2). The principal cells and interneurons form two types of local circuits. The first of these involve Type-1 interneurons and perform the same function as that performed by local inhibitory circuits in BIND; they limit the number of principal cells whose synapses undergo LTP. The second type of circuits involve Type-2 interneurons and lead to the formation of binding-error detector circuits.

Each principal cell receives afferents from a number of cells in ROLE and BIND and sends collaterals to neighboring Type-2 interneurons. Type-2 interneurons in turn make contacts on neighboring principal cells. If a principal cell receives an inhibitory contact from a Type-2 interneuron, then the likelihood that the principal cell also sends a collateral back to the same interneuron is high. Consequently, there exist a large number of feedback circuits consisting of a principal cell and a Type-2 interneuron. One such feedback circuit consisting of principal cell P and Type-2 interneuron *int* is depicted in Figure 4(a). As a matter of convention, we will refer to *int* as a *satellite* of P. Typically, each principal cell will have several satellites and and each Type-2 interneuron will be a satellite of numerous principal cells.

The projection from BIND to BED is such that given a principal cell P and one of its satellite *int*, if P receives an afferent from a cell b in BIND, then it is likely that *int* also receives an afferent from b. This sort of connectivity could arise naturally if axonal fibers make *en passant* connections with both principal and Type-2 cells within a neighborhood.

The potentiation threshold, θ_p, of principal cells is such that LTP of synapses occurs only if multiple synapses of a postsynaptic cell receive coincident presynaptic activity. Similarly, θ_p of a Type-2 interneuron is such that LTP of its synapses formed by afferents from BIND cells occurs only if the interneuron also receives coincident activity at several synapses. The synapses formed on Type-2 interneurons by collaterals from principal cells are not required to undergo LTP.

The firing thresholds (θ_f's) of cells in BED are such that a principal cell does not fire unless it receives impulses at potentiated synapses from ROLE and/or BIND cells, and a Type-2 interneuron does not fire unless it receives impulses at potentiated synapses from BIND cells. Moreover, inhibitory input from a Type-2 interneuron is sufficient to block a principal cell from firing — irrespective of any excitatory inputs received by the latter. A set of possible values for weights and thresholds are given in Section 9.

Figure 4(a) shows a principal cell P receiving afferents from r_1 cells in ROLE, and $\langle r_1 = f_1 \rangle$ cells in BIND. Given the dynamic encoding of RI (Figure 2), these afferents convey synchronous activity and this leads to the associative LTP of P's synapses receiving afferents from $\langle r_1 = f_1 \rangle$ and r_1 cells.[14] At the same time, some of the inactive naive synapses of P receiving afferents from BIND cells and ROLE cells undergo heterosynaptic LTD. After the potentiation of its synapses, P fires upon receiving activation from r_1 cells and/or $\langle r_1 = f_1 \rangle$ cells (Figure 4(b)).

[14] As discussed below, LTP may also occur at synapses of a postsynaptic cell that receives afferents only from r_1 cells or only from $\langle r_1 = f_1 \rangle$ cells. However, the probability of this happening is very small (see Section 9).

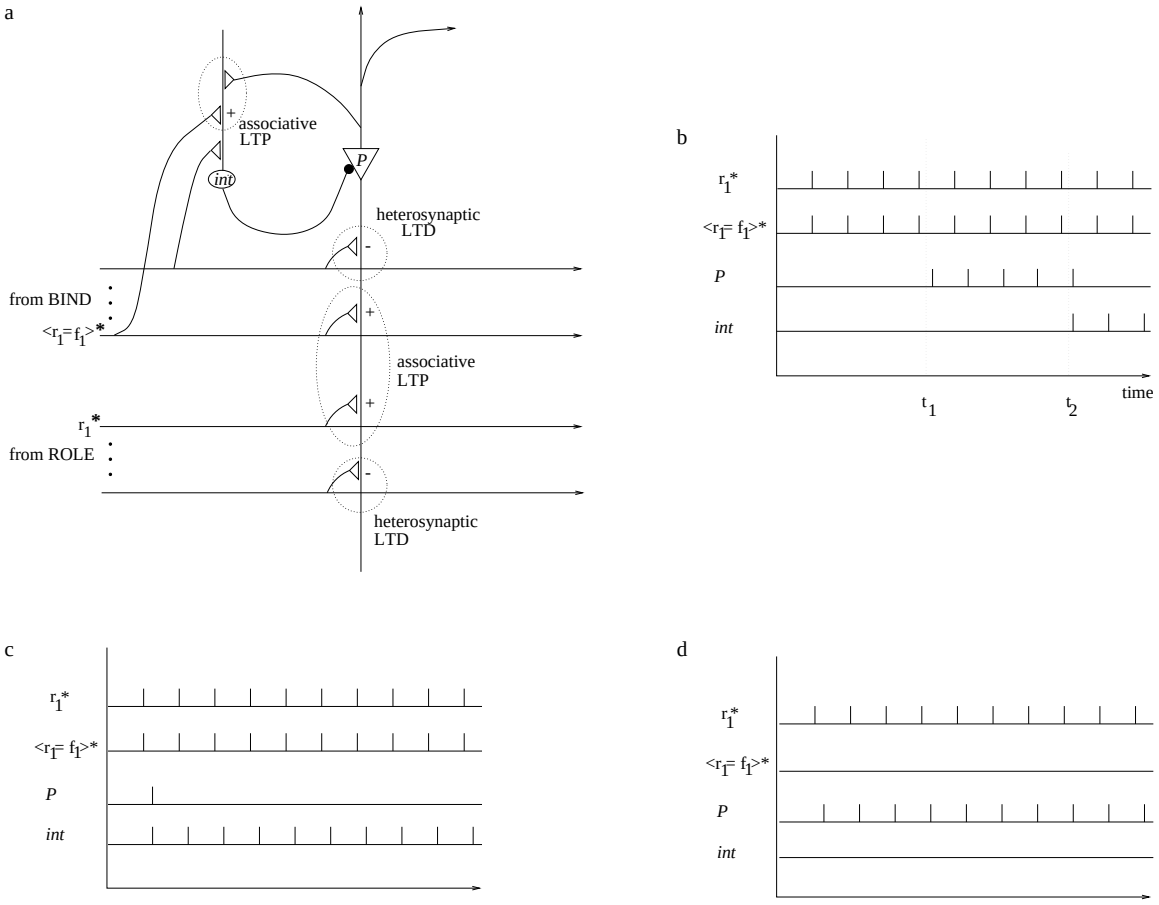

Figure 4: (a) A local inhibitory feedback circuit consisting of a principal cell P and a Type-2 interneuron *int* in BED. Type-1 inhibitory interneurons are not shown. Typically, each principal cell and inhibitory interneuron participates in several such feedback circuits. The input labeled r_1* refers to afferents from r_1 cells in ROLE and that labeled $\langle r_1 = f_1 \rangle$* refers to afferents from $\langle r_1 = f_1 \rangle$ cells in BIND. The arrival of coincident activity along r_1* and $\langle r_1 = f_1 \rangle$* causes LTP and LTD of synapses as shown, and thereby, results in the formation of a binding-error detector circuit consisting of P and *int*, for the binding $\langle r_1 = f_1 \rangle$. Such a circuit is referred to as $bed(\langle r_1 = f_1 \rangle)$. Typically, several such circuits are formed for each binding. (b) A schematic representation of the activity of P and *int* during the above process. The LTP and LTD of P's synapses has occurred by time t_1, and the LTP of *int*'s synapse has occurred by time t_2. It is assumed that the weighted sum of activity along r_1* and f_1* exceeds θ_p and that κ is 4. (c,d) A schematic representation of the response of P and *int* subsequent to their forming a binding-error detector circuit. The sustained firing of P signals a binding error. (c) The response to a cue that specifies the binding $\langle r_1 = f_1 \rangle$ and (d) the response to a cue that binds r_1 to an entity other than f_1.

Subsequent volleys of inputs from r_1 cells and/or $\langle r_1 = f_1 \rangle$ cells cause P to fire and result in impulses arriving at the synapse between P and *int*. Now *int* is already receiving activation from $\langle r_1 = f_1 \rangle$ cells and the arrival of concurrent activity from P leads to the associative LTP of the synapse at which *int* receives activation from $\langle r_1 = f_1 \rangle$ cells.[15] After the potentiation of this synapse, activation arriving from the $\langle r_1 = f_1 \rangle$ cell is sufficient to fire *int* and cause the inhibition of P.

At the end of the above sequence of events, the circuit consisting of P and *int* becomes a binding-error detector circuit for the binding $\langle r_1 = f_1 \rangle$ and will be referred to as a $bed(\langle r_1 = f_1 \rangle)$ circuit. In future, P does not fire whenever role r_1 is bound to entity f_1 in the relational instance currently active in the ROLE and ENTITY regions because the synchronous firing of r_1 and f_1 cells activates $\langle r_1 = f_1 \rangle$ cells in BIND, which activate *int*, which in turn inhibits P (Figure 4(c)). However, P fires whenever the firing of r_1 cells is *not* accompanied by the synchronous firing of $\langle r_1 = f_1 \rangle$ cells. In other words, P fires whenever the relational instance currently active in the ROLE and ENTITY regions binds r_1 to any entity other than f_1 (Figure 4(d)). A similar process leads to the formation of $bed(\langle r_2 = f_2 \rangle)$ circuits that act as binding-error detectors for the binding $\langle r_2 = f_2 \rangle$. In general, multiple binding-error detector circuits are formed for each binding.

The process of binding-error detector circuit formation described above can also suffer from the sort of problems discussed in the context of the formation of binding detector cells in BIND. These include the possibility that for a given binding, there may not exist any suitably connected cells to form a binding-error detector circuit, spurious and ambiguous binding-error detector circuits may get formed during the memorization process, and the sharing of cells among several binding-detector circuits can compromise the functioning of such circuits.

The formation of a spurious binding-error detector circuit may occur if a principal cell P does not receive afferents from any $\langle r_1 = f_1 \rangle$ cells but receives sufficient activity ($\geq \theta_p$) along afferents from r_1 cells. Such a situation would typically lead to a spurious binding-error detector circuit.[16] Upon commitment, such a cell would fire and signal a binding-error whenever the cue specifies *any* binding involving the role r_1, irrespective of the specified role-filler. A principal cell may also form a spurious binding-error detector circuit, in spite of receiving afferents from both r_1 and $\langle r_1 = f_1 \rangle$ cells, if it does not have any satellites *or* if afferents from $\langle r_1 = f_1 \rangle$ cells do not make contact with *any* of its satellites. Such a cell would also fire and signal a binding error whenever the cue specifies *any* binding involving the role r_1, irrespective of the specified role-filler. Finally, a spurious binding-error detector circuit may get formed if a cell does not receive any afferents from r_1 cells, but receives sufficient activity along afferents from $\langle r_1 = f_1 \rangle$ cells. In this case, the cell would not behave as the principal cell of a $bed(\langle r_1 = f_1 \rangle)$ circuit since its firing does not depend on r_1. As discussed in Section 9, however, the ratio of the expected number of proper binding-error

[15] θ_p of a Type-2 interneuron is such that LTP of its synapses formed by afferents from BIND cells occurs only if the interneuron receives coincident activity at several synapses. Furthermore, the weights of naive synapses formed on Type-2 interneurons by BIND cells are low relative to the weights of synapses formed on these interneurons by principal cells in BED. Hence, it is highly likely that LTP of a synapse formed by a $\langle r_1 = f_1 \rangle$ cell on *int* will occur only if *int* also receives coincident activity from P (or some other principal cell).

[16] Such a cell may form a proper binding-error detector circuit, in spite of not receiving afferents from any $\langle r_1 = f_1 \rangle$ cells, if an afferent from one of the $\langle r_1 = f_1 \rangle$ cells makes contact with one of the satellites of P.

detector circuits formed to the expected number of spurious binding-error detector formed is extremely high, and hence, the episodic memory can function properly with a very high probability.

As in the case of BIND, an *ambiguous* binding-error detector may be formed during the memorization of a relational instance if the same cell receives afferents from role cells and binding detector cells corresponding to more than one role-entity pairs bound in the relational instance being memorized. Again, as discussed in Section 9, the probability of occurrence of such ambiguous detectors is extremely small.

The functioning of a binding-error detector circuit may be compromised if the principal cell associated with this circuit also undergoes LTP in response to additional bindings. Consider a binding-error detector circuit for the binding $\langle r_i = f_k \rangle$. Let its principal cell be *princ*. The functioning of this circuit may be disrupted if *princ* undergoes LTP in response to the binding $\langle r_j = f_l \rangle$ and one of the following two conditions hold: (1) $r_i = r_j$ and $f_k \neq f_l$, OR (2) $r_i \neq r_j$, but r_i and r_j are roles of the *same* generic relation.

Since there could be many more primary cells than Type-2 inhibitory interneurons, each Type-2 interneuron would typically participate in multiple binding-error detector circuits. The sharing of interneurons among multiple binding-error detector circuits, however, can cause such circuits to malfunction. Consider, the situation where the same Type-2 interneuron is shared among two binding-error detector circuits $bed(\langle r_i = f_k \rangle)$ and $bed(\langle r_i = f_l \rangle)$, which involve the same role but distinct entities as role-fillers. Because of the shared interneuron, both these circuits will behave as binding-error detectors for the binding $\langle r_i = (f_j \vee f_k) \rangle$. In other words, $bed(\langle r_i = f_k \rangle)$ will be satisfied by the binding $\langle r_i = f_k \rangle$ as well as by the binding $\langle r_i = f_l \rangle$. Similarly, $bed(\langle r_i = f_l \rangle)$ will be satisfied by the binding $\langle r_i = f_l \rangle$ as well as by the binding $\langle r_i = f_k \rangle$. The functioning of a binding-error detector circuit $bed(\langle r_i = f_k \rangle)$ can also be disrupted by a binding-error detector circuit $bed(\langle r_j = f_l \rangle)$, where $r_i \neq r_j$, if the two circuits share a Type-2 interneuron. However, the disruption in the functioning of $bed(\langle r_i = f_k \rangle)$ caused by $bed(\langle r_j = f_l \rangle)$ can lead to an erroneous response to recognition or recall queries *only* under the following circumstances: (i) r_i and r_j are roles of the *same* generic relation, and (ii) the relational instance in which the binding $\langle r_i = f_k \rangle$ occurred also included the binding $\langle r_j = f_l \rangle$. [17]

As reported in Section 9, the probability that the detection of binding errors will be disrupted as a result of shared interneurons is extremely small. This low probability results, in part from the fact that shared interneurons lead to problems only in the specific situations mentioned above and several binding-error detection circuits are formed for each binding.

[17] A brief justification of conditions (i) and (ii) is as follows. The episodic memory system is expected to receive a query or a cue pertaining a single generic relation at a time. Consequently, potential cross-talk between binding-error detectors pertaining to roles of two different generic relations is harmless. Hence, condition (i).

Assume that *RI* included the binding $\langle r_j = f_m \rangle$, where $f_m \neq f_l$, and hence, condition (ii) is not met. Now consider any cue that specifies the binding $bed(\langle r_j = f_l \rangle)$. Such a cue would disrupt the circuit $bed(\langle r_i = f_k \rangle)$ and prevent it from producing an error signal even if the cue specifies the wrong binding for the role r_i. However, such a cue would cause the binding-error detector circuits for the binding $bed(\langle r_j = f_m \rangle)$ to detect an error and fire since the cue specifies the binding $bed(\langle r_j = f_l \rangle)$ and $f_m \neq f_l$. As explained in (Shastri, 1997b) the firing of these circuits would generate a binding-error signal that would block a yes response downstream in area CA1, and thereby render the disruption of the circuit $bed(\langle r_i = f_k \rangle)$ inconsequential.

7 Relationship between the hippocampal formation and the model structure

As pointed out in Section 3 , there is a correspondence between the gross architecture of the structure described above and the hippocampal formation (HF) with ROLE and ENTITY regions corresponding to subregions of EC, and BIND and BED regions corresponding to DG and CA3, respectively.

The internal circuitry of BIND and BED regions assumed in the model also matches the local circuitry of DG and CA3. The principal cells in BIND and BED correspond to DG granule cells and CA3 pyramidal cells, respectively, and the interneurons in these regions correspond to inhibitory interneurons in DG and CA3, respectively (Kawagachi & Hama, 1987; Braak, 1974; Anderson, Gross, Lomo & Sveen, 1969; Schwartzkroin, Scharfman & Slovitor, 1990). DG granule cells and inhibitory interneurons in DG are known to form feedback and feedforward circuits of the sort formed by principal cells and Type-1 interneurons in BIND (Schwartzkroin, Scharfman & Slovitor, 1990; Amaral & Witter, 1995). Similarly, CA3 pyramids and inhibitory interneurons are known to participate in local circuits of the type assumed in BED. This includes the generic inhibitory feedback and feedforward circuits involving principal cells and Type-1 interneurons as well as the more specific disynaptic feedback inhibitory circuits involving primary cells and Type-2 interneurons (Miles & Wong, 1984; Anderson, Gross, Lomo & Sveen, 1969).

The requisite LTP and LTD at synapses in BED formed by afferents from BIND is analogous to the occurrence of associative LTP and heterosynaptic LTD of mossy fiber — CA3 synapses (Derrick & Martinez, 1996) and the LTP of perforant path — CA3 synapses (Zalutsky & Nicoll, 1990). The model posits that LTP occurs at a synapse formed by mossy fibers on a CA3 interneuron if the latter receives coincident activation from a CA3 pyramidal cell.

8 Relation of present work to a model of hippocampal memory formation

The structure for the formation of binding and binding-error detector circuits described above has been embedded within a detailed model of episodic memory formation (Shastri, 1997a, 1997b). The proposed model, SMRITI, demonstrates how the HF might rapidly encode an event or a situation as a persistent structure capable of supporting rapid recognition and recall. Figure 5 describes the functional architecture of the model and its mapping to the HF. The construal of an event or a situation is expressed as a transient pattern of activity over HLCCs. Projections from HLCCs to EC (Van Hoesen, 1982; Insausti, Amaral & Cowan, 1987; Suzuki & Amaral, 1994) convey this activity to EC from where it propagates around the complex loop consisting of EC, DG, CA3, CA2, CA1, SC, and EC, and triggers a sequence of synaptic changes involving LTP and LTD. As a result of these changes, the relational instance is transformed from a transient pattern of activity into a persistent structural encoding consisting of the functional units enumerated in Figure 5. The activity arriving at EC from CA1 and SC, and thereafter propagating back to the HLCCs, constitutes the response of the HF.

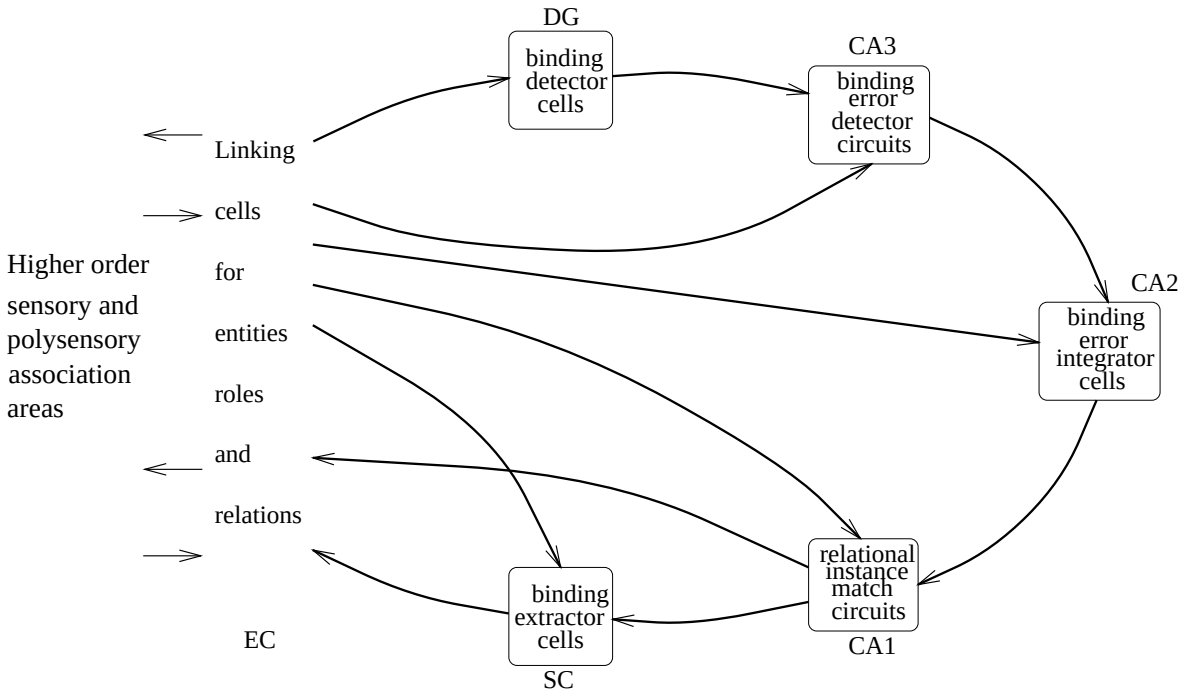

Figure 5: The functional architecture of a model of hippocampus-based episodic memory system. EC provides linking cells for connecting the cortical representations of generic relations, roles, and entities with the hippocampal system, cells in DG act as detectors for role-entity bindings, circuits in CA3 serve as binding-error detectors, and CA2 cells integrate outputs of binding-error detectors pertaining to a single memorized relational instance and convey it to CA1. CA1 provides the locus for matching memorized relational instances with relational instances presented to EC by some high-level cortical circuit, and cells in the subicular complex act as binding extractor cells for the selective activation in EC of linking cells of entities that fill specific roles in a given relational instance.

9 A Quantitative analysis of capacity and robustness

This section presents results of a quantitative analysis of the model using gross, but plausible, estimates of the size of various regions and projective fields (PFs). Specifically, the following probabilities and expected values have been approximated for regions BIND (DG) and BED (CA3).

1. The probabilities that *no* cell or circuit with suitable connections will be found for recruitment as binding and binding-error detectors, respectively, during the memorization of a relational instance (P_{fail}).

2. The expected number of cells and circuits that will receive appropriate connections, and hence, will be *good* candidates for recruitment during the memorization of a relational instance ($\mathrm{E}\langle good.cand \rangle$).

3. The expected number of cells and circuits that will receive sufficient but inappropriate inputs, and hence, will be *spurious* candidates for recruitment during the memorization of a relational instance ($\mathrm{E}\langle spurious.cand \rangle$).

4. The expected number of cells and circuits that will receive inputs pertaining to more than one role-entity pair bound in the relational instance being memorized, and hence, will be *ambiguous* candidates for recruitment during the memorization process ($\mathrm{E}\langle ambiguous.cand \rangle$).

5. Assuming that only a certain number (Est_{rec}) of candidate cells get recruited, the expected number of "good" cells and circuits recruited (in contrast to the number of spurious and ambiguous cells and circuits recruited) during the memorization of a relational instance ($\mathrm{E}\langle good.recruit \rangle$).

6. The expected number of binding detector cells whose functioning will be compromised as a result of the same principal cell in BIND undergoing LTP in response to multiple bindings ($\mathrm{E}\langle compromised \rangle$ for BIND).

7. The expected number of binding-error detector circuits whose functioning will be compromised as a result of (i) the sharing of Type-2 interneurons among several binding-error detector circuits ($\mathrm{E}\langle compromised.T2 \rangle$), (ii) the same principal cells in BED undergoing LTP in response to multiple bindings ($\mathrm{E}\langle compromised.princ \rangle$), and (iii) the disruption of binding detector cells in BIND ($\mathrm{E}\langle compromised.bind \rangle$). The total number of binding-error detector circuits expected to be disrupted by the cumulative effect of the above ($\mathrm{E}\langle compromised \rangle$).

Table 1 displays the above quantities calculated with $\zeta = 0$. This means that *none* of the synapses undergo heterosynaptic LTD as a result of the LTP its synapses. This condition allows a maximum sharing of cells among various binding detectors and binding-error detectors, and hence, leads to a maximal disruption in the functioning of these cells and circuits. Consequently, these results provide a measure of the system's performance under the most adverse set of conditions vis-a-vis cross-talk resulting from the sharing of cells.

Region	Statistic	Tabula rasa 0 bindings	100,000 bindings	150,000 bindings	200,000 bindings
BIND (DG)	P_{fail}	$< 10^{-18}$	$< 10^{-18}$	$< 10^{-18}$	$< 10^{-18}$
	$\mathrm{E}\langle good.cand \rangle$	120.9	120.9	120.9	120.9
	$\mathrm{E}\langle spurious.cand \rangle$	3.2	3.2	3.2	3.2
	$\mathrm{E}\langle ambiguous.cand \rangle$	$9.8 * 10^{-4}$	$9.8 * 10^{-4}$	$9.8 * 10^{-4}$	$9.8 * 10^{-4}$
	Est_{rec}	40	40	40	40
	$\mathrm{E}\langle good.recruit \rangle$	39.0	39.0	39.0	39.0
	$\mathrm{E}\langle compromised \rangle$	0.0	5.3	7.7	10.0
BED (CA3)	P_{fail}	$< 10^{-18}$	10^{-18}	10^{-18}	10^{-18}
	$\mathrm{E}\langle good.cand \rangle$	51.7	51.7	51.7	51.7
	$\mathrm{E}\langle spurious.cand \rangle$	2.7	2.7	2.7	2.7
	$\mathrm{E}\langle ambiguous.cand \rangle$	$9.9 * 10^{-4}$	$9.9 * 10^{-4}$	$9.9 * 10^{-4}$	$9.9 * 10^{-4}$
	Est_{rec}	10	10	10	10
	$\mathrm{E}\langle good.recruit \rangle$	9.5	9.5	9.5	9.5
	$\mathrm{E}\langle compromised.T2 \rangle$	0	0.43	0.63	0.83
	$\mathrm{E}\langle compromised.princ \rangle$	0	0.04	0.05	0.07
	$\mathrm{E}\langle compromised.bind \rangle$	0	1.34	1.93	2.49
	$\mathrm{E}\langle compromised \rangle$	0	1.75	2.48	3.16

Table 1: Quantitative evaluation of binding detector cell formation in BIND (DG) and binding-error detector circuit formation in BED (CA3). Here ζ is assumed to be 0, and hence, principal cells and Type-2 interneurons can participate in the encoding of a potentially large number of bindings. The interpretation of P_{fail}, $\mathrm{E}\langle good.cand \rangle$, $\mathrm{E}\langle spurious.cand \rangle$, $\mathrm{E}\langle ambiguous.cand \rangle$, Est_{rec}, $\mathrm{E}\langle good.recruit \rangle$, $\mathrm{E}\langle compromised.T2 \rangle$, $\mathrm{E}\langle compromised.princ \rangle$, $\mathrm{E}\langle compromised.bind \rangle$, and $\mathrm{E}\langle compromised \rangle$ is given in the text. These values are computed for four system states — after the system has encoded 0 (tabula rasa), 100,000, 150,000 and 200,000 distinct bindings, respectively.

Region	Statistic	Tabula rasa 0 bindings	100,000 bindings	150,000 bindings	200,000 bindings
BIND (DG)	P_{fail}	$< 10^{-18}$	$< 10^{-18}$	$< 10^{-18}$	$< 10^{-18}$
	$\mathrm{E}\langle good.cand \rangle$	120.9	88.6	72.5	56.4
	$\mathrm{E}\langle spurious.cand \rangle$	3.2	2.3	1.9	1.5
	$\mathrm{E}\langle ambiguous.cand \rangle$	$9.8 * 10^{-4}$	$7.1 * 10^{-4}$	$5.8 * 10^{-4}$	$4.5 * 10^{-4}$
	Est_{rec}	40	40	40	40
	$\mathrm{E}\langle good.recruit \rangle$	39.0	39.0	39.0	39.0
	$\mathrm{E}\langle compromised \rangle$	0	0	0	0
BED (CA3)	P_{fail}	$< 10^{-18}$	$7.3 * 10^{-15}$	$1.1 * 10^{-10}$	$1.5 * 10^{-6}$
	$\mathrm{E}\langle good.cand \rangle$	51.7	32.6	23.0	13.4
	$\mathrm{E}\langle spurious.cand \rangle$	2.7	1.7	1.2	0.7
	$\mathrm{E}\langle ambiguous.cand \rangle$	$9.9 * 10^{-4}$	$6.3 * 10^{-4}$	$4.4 * 10^{-4}$	$2.6 * 10^{-4}$
	Est_{rec}	10	10	10	10
	$\mathrm{E}\langle good.recruit \rangle$	9.5	9.5	9.5	9.5
	$\mathrm{E}\langle compromised.T2 \rangle$	0	0.43	0.63	0.83
	$\mathrm{E}\langle compromised.princ \rangle$	0	0	0	0
	$\mathrm{E}\langle compromised.bind \rangle$	0	0	0	0
	$\mathrm{E}\langle compromise \rangle$	0	0.43	0.63	0.83

Table 2: Quantitative evaluation of binding detector cell formation in BIND (DG) and binding-error circuit formation in BED (CA3). Here ζ has been assumed to be 1, and hence, each principal cell participates in the encoding of at most one binding (Type-2 interneurons are, however, shared as before). The interpretation of P_{fail}, $\mathrm{E}\langle good.cand \rangle$, $\mathrm{E}\langle spurious.cand \rangle$, $\mathrm{E}\langle ambiguous.cand \rangle$, Est_{rec}, $\mathrm{E}\langle good.recruit \rangle$, $\mathrm{E}\langle compromised.T2 \rangle$, $\mathrm{E}\langle compromised.princ \rangle$, $\mathrm{E}\langle compromised.bind \rangle$, and $\mathrm{E}\langle compromised \rangle$ is given in the text. These values are computed for four system states — after the system has encoded 0 (tabula rasa), 100,000, 150,000 and 200,000 distinct bindings, respectively.

Table 2 displays the above quantities calculated with $\zeta = 1$. This means that *all* of the inactive synapses of a postsynaptic cell undergo heterosynaptic LTD upon the LTP of one of its synapses. This condition minimizes the sharing of principal cells among multiple binding detectors and binding-error detectors (Type-2 interneurons are shared, however, as before), and hence, it minimizes the disruption in the functioning of these cells and circuits. Consequently, these results provide a measure of the system's performance under a highly favorable set of conditions vis-a-vis cross-talk resulting from the sharing of cells. Observe that while the $\zeta = 1$ condition minimizes the sharing of principal cells and leads to a minimal disruption of cell function, it also reduces the capacity of the memory system since it allows any given cell or circuit to be associated with at most one binding.

For intermediate values of ζ, the performance of the system will lie in between the above two levels shown in Tables 1 and 1.

It is assumed that events and situations are expressed in terms of 500 generic relations involving a total of 2000 roles, and 50,000 entities can serve as role-fillers. The events and the associated bindings being memorized are assumed to be uniformly distributed over the space of possible events and bindings.

Four sets of estimates for four states of episodic memory utilization are displayed in Tables 1 and 2. These states correspond to 0, 100,000, 150,000, and 200,000 bindings, respectively, being *already encoded in memory.* Assuming that, on an average, each event consists of four bindings, these states correspond to 0, 25,000, 37,500, and 50,000 events, respectively, being already encoded in memory.[18]

The above calculations assume that DG and CA3 contain 15 million and 2.7 million cells, respectively, and that the PF sizes for the projections from DG to CA3, EC to DG, and EC to CA3 are 21, 17,000, and 6000, respectively. These choices are based in part on data provided in (Amaral, Ishizuka & Claiborne, 1990; Amaral & Witter 1995; West, 1990).[19] For ease of analysis, it is assumed that these PFs are uniformly distributed over their respective target regions.

A role and entity ensemble in the ROLE and ENTITY subregions of EC is assumed to contain 30 and 10 cells, respectively. It is also assumed that on an average (mean), the number of candidate cells in BIND and the number of candidate circuits in BED, actually recruited (Est_{rec}) during the memorization of a relational instance is 40 and 10, respectively.[20] Recall that soft-WTA networks formed by Type-1 inhibitory interneurons are expected to limit the number of cells actually recruited from the pool of candidate cells. The soft-WTA effect of Type-1 interneurons has not been modeled in detail in this work, however, functionally similar networks has been studied extensively by a number of researchers since Marr (1971).

Each principal cell in BED sends collaterals to several Type-2 inhibitory interneurons. It is assumed that on an average 6 of these Type-2 interneurons also send inhibitory links back to the principal cell. Thus each principal cell has about 6 satellites (i.e., it participates in

[18] Four bindings in an event would allow the specification of who (did what) to whom, where, and when. Note that "what" corresponds to the generic relation being instantiated and does not count as a binding.

[19] The size of the DG to CA3 PF is the mean of the 14–28 range specified in (Amaral & Witter, 1995).

[20] If $\zeta = 0$, the expected number of candidate cells and circuits does not change with memory load, and hence, the values of Est_{rec} are expected to remain constant. If $\zeta = 1$, the expected number of candidate cells and circuits decreases as memory load increases. Consequently, the values of Est_{rec} would be expected to decrease. However, these values have been treated as constant in the $\zeta = 1$ condition also to simplify calculations.

about 6 local feedback circuits of the sort described in Figure 4). Furthermore, each Type-2 interneuron is assumed to receive inputs from about 100 principal cells. The probability that a cell in BIND that makes contact with a principal cell P, also makes contact with a given satellites of P, is assumed to be 0.4, and the probability that a Type-2 interneuron receives a collateral from a principal cell, given that it makes an inhibitory contact on the principal cell, is assumed be close to 1.0.

The potentiation thresholds, θ_p's, for principal cells in BIND and BED are set to 760 and 500, respectively, and the firing thresholds, θ_f's, for these cells are set to 1900 and 800, respectively. The values of θ_p and θ_f for Type-2 interneurons in BED are 1400 and 1800, respectively.

The naive weight bands for projections from ROLE to BIND, ENTITY to BIND, ROLE to BED, and BIND to BED are set to [200,220], [360,380], [100,120], and [400,420], respectively, the potentiated weight bands for these projections are set to [500,520], [900,920], [800,820], and [800,820], respectively, and the depressed weight bands are set to [40,45], [70,80], [20,35], and [80,90], respectively. The weight of a link from a principal cell to a Type-2 interneuron and the weight of a link from a Type-2 interneuron to a principal cell in BED is assumed to be 1000 and -5,000 respectively. The weight bands of naive, potentiated, and depressed synapses from BIND to Type-2 interneurons are also assumed to be [400,420], [800,820], and [80,90], respectively.

Capacity

As indicated by the results shown in Tables 1 and 2, the failure probabilities are extremely low. Even when $\zeta = 1$, which corresponds to a worse case scenario for capacity, P_{fail} for BIND and BED is less than 10^{-18} and $2 * 10^{-6}$, respectively. Moreover, even after a large number of bindings have been encoded, the expected number of good candidate cells (or circuits) available for encoding a given binding remains sufficiently high, and so does the ratio of good candidate cells (or circuits) to spurious candidate cells (or circuits). Thus even after 200,000 bindings have been memorized, the expected number of cells suitable for recruitment as binding detector cells for a given binding is 56.4, and the expected number of circuits suitable for recruitment as binding-error detector circuits is 13.4. The expected number of spurious cells in these cases is 1.5 and 0.7, respectively. The number of ambiguous cells is also extremely small. Consequently, a majority of binding detector cells (30 out of 40) and binding-error detector circuits (9.5 out of 10) formed in response to a binding are expected to function correctly.

Robustness against cross-talk resulting from sharing of cells and the formation of spurious cells

Tables 1 and 2 also indicate that the expected number of copies of a given binding detector or binding-error detector disrupted due to the sharing of principal cells in BIND and principal cells and Type-2 interneurons in BED remains low. Even in the case where $\zeta = 0$ which corresponds to a worse case situation from the point of view of cross-talk and interference among shared cells, only 10 out of 40 binding detector cells associated with a given binding are expected to be compromised as a result of memorizing 200,000 other bindings. Similarly, only 3.16 out of 10 binding-error detector circuits associated with a given binding

Region	X_{min}	$P_{fail}(X_{min})$	E$\langle loss, X_{min}\rangle$ out of 200,000
BIND (DG)	39	$6.1 * 10^{-2}$	12,147
($Est_{rec} = 40$)	37	$6.9 * 10^{-4}$	137
	35	$2.9 * 10^{-6}$	0.6
BED (CA3)	9	$4.3 * 10^{-3}$	853
($Est_{rec} = 10$)	7	$2.0 * 10^{-6}$	0.4
	5	$2.0 * 10^{-10}$	0.00004

Table 3: Effect of a 1%cell loss on the encoding of binding and binding error detectors. Est_{rec} indicates that the number of copies that are estimated to be recruited for each functional unit. X_{min} indicates different assumptions about the *minimum* number of copies of a functional unit that must survive for the functional unit to function properly. $P_{fail}(X_{min})$ is the probability that less than X_{min} copies will survive as a result of a 1% diffuse cell loss. E$\langle loss, X_{min}\rangle$ specifies the expected number of functional units that will be lost out of a total of 200,000 as a result of a 1% cell loss assuming that at least X_{min} copies of each functional unit must survive the loss.

are expected to be compromised as a result of memorizing 200,000 other bindings. Here 3.16 takes into account not only the disruptive effect of shared principal cells and Type-2 interneurons in BED, but also the disruptive effect of compromised binding detector cells in BIND.

Robustness against diffuse cell loss

Each binding and binding-error detector — i.e., each functional unit — is redundantly encoded by multiple cells and circuits. Moreover, these cells and circuits are physically dispersed over a region. Thus the memorization process leads to the formation of multiple physically dispersed copies of each functional unit. Consequently, the probability that limited amounts of diffuse cell loss will destroy a functional unit is extremely small. Table 3 summarizes the effect that a 1% diffuse cell loss will have on binding and binding-error detectors.[21] The odds that more that 5 out of 40 binding detector cells associated with a given binding will be lost due to a diffuse 1% cell loss are less than 3 in a million. Similarly, the odds that more than 3 out of 10 binding-error detector circuits associated with a given binding will be lost due to a similar loss are less that 2 in a million. Thus the encoding of binding and binding-error detectors is quite robust against diffuse cell loss.

10 Predictions

The model predicts that damage to EC will lead to loss of cells linking the HF to HLCCs, and hence, cues involving entities and roles whose linking cells are destroyed will not be processed correctly subsequent to the damage. If the linking cells of any of the entities specified in a cue are damaged as a result of damage to region ENTITY of EC, the system

[21] A 1% cell loss is quite significant and is analogous to a loss of a billion cortical cells.

will produce a "don't know" response. Behaviorally, this corresponds to forgetting. If on the other hand, the linking cells of entities mentioned in a cue are (partially) intact, but the linking cells of the roles mentioned in the cue are damaged due to damage to region ROLE of EC, the system will produce a "false positive" response. Behaviorally, this is analogous to the existence of spurious memories. Queries involving entities and roles whose linking cells are intact, will continue to be processed normally. Massive damage to EC however, will result in a massive loss of linking cells and lead to a catastrophic failure in the recognition and recall of existing memories. EC is one of the first areas to be affected by Alzheimer's disease and the predicted behavior is consistent with the nature of memory deficits observed during the progression of the disease.

Damage to DG granule cells will destroy binding detector cells. Consequently, binding error detector circuits in CA3 will not be inhibited, and will therefore, generate error signals even when a cue specifies correct bindings. This will lead to erroneous "don't know" responses. Behaviorally, this will correspond to forgetting.

The structure of binding-error detector circuits in CA3 entails that large scale loss of CA3 pyramids or loss of perforant path inputs to CA3 will prevent the generation of binding-error signals, and hence, lead to false positive responses during recognition and recall. Behaviorally, this will correspond to the existence of spurious memories. Large scale loss of mossy fiber inputs or loss of CA3 interneurons will prevent the blocking of binding-error signals, and hence, will lead to erroneous "don't know" responses. Behaviorally, this will correspond to forgetting. Finally, subjects with damage to CA1 and CA2, but with intact EC, DG and CA3 regions will continue to generate binding-error signals, and hence, continue to detect novelty even though they may be amnesic.

The detailed model SMRITI also predicts that damage to CA2 would lead to false positive responses, injury to CA1 would lead to catastrophic memory failure, and damage to SC would disrupt the recall of existing memories, while leaving the recognition of such memories intact.

11 Discussion

This work argues that the memorization of events and situations requires the rapid formation of circuits for detecting bindings and binding errors. It then demonstrates how a transient pattern of activity representing a relational instance can be transformed rapidly into persistent circuits for detecting bindings and binding errors within a structure whose architecture and circuitry matches that of the HF.

The resulting encoding is sparse and representationally compact, yet robust against cell loss due to its redundant and physically dispersed nature. Thus the coding offers the usual advantages of the so called "distributed representations" and at the same time possesses the representational compactness of the so called "localist representations". The model also suggests that activity arriving along a relatively small number of afferents may be sufficient for inducing LTP in a fully functioning HF.

This work offers an alternative interpretation of the functional role played by CA3 and its local inhibitory circuits. While many models of the HF view CA3 as an associative or auto-associative memory (Marr, 1971; Treves & Rolls, 1994; Hasselmo, 1997; O'Reilly & McClelland, 1996), this work suggests that a key representational role of CA3 in humans

may be the detection of binding errors. As summarized in Figure 5, these binding-error detector circuits acting in concert with other functional units in various regions of the HF enable it to perform memorization, recognition, and recall.

It has been suggested that inhibitory interneurons stabilize pyramidal cell activity (Marr, 1971; Buckmaster & Schwartzkroin, 1995; Schwartzkroin & Prince, 1980) and perhaps, subserve long-distance synchronization of hippocampal activity (Traub, Whittington, Stanford & Jeffreys, 1996; Sik, Ylinen, Penttonen & Buzsaki, 1994). This work suggests that additionally, CA3 interneurons may play a critical role in the formation of binding-error detector circuits critical to the proper functioning of episodic memory and novelty detection.

References

Ajjanagadde, V. & Shastri, L. (1991). Rules and Variables in Neural Nets. *Neural Computation* **3**, 121—134.

Amaral, D.G., Ishizuka, N. & Claiborne, B. (1990) Neurons, numbers and the hippocampal network. In *Progress in Brain Research: Understanding the brain through the hippocampus* (eds Storm-Mathisen, J., Zimmer, J. & Ottersen, O.P.) 1–11, Elsevier Science, Amsterdam.

Amaral, D.G. & Witter, M.P. (1995) Hippocampal Formation. In *The Rat Nervous System*, 2nd edn (ed Paxinos, G.) 443–493, Academic Press, London.

Anderson, P., Gross, G.N., Lomo, T., & Sveen, O. (1969) Participation of inhibitory and excitatory interneurons in the control of hippocampal cortical output. In *The Interneuron* (ed Brazier, M.A.B.) 415–465, University of California Press, Berkeley.

Artola, A. & Singer, W. (1993) Long-term depression of excitatory synaptic transmission and its relationship to long-term potentiation. *Trends Neurosci.* **16:** 480–487.

Bienenstock, E.L., Cooper, L.N., & Munro, P.W. (1982). Theory for the development of neural selectivity, orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience, 2*: 32—48.

Bliss, T.V.P. & Collingridge, G.L. (1993) A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361,** 31–39.

Braak, H. (1974) On the Structure of the Human Archicortex. I. The Cornu ammonis. A Golgi and Pigmentarchitectonic Study. *Cell and Tissue Res.* **152,** 349–383.

Buckmaster, P.S. & Schwartzkroin, P.A. (1995) Interneurons and inhibition in the Dentate Gyrus of the Rat *in vivo*. *Journal of Neuroscience, 15*(1): 774—789.

Chover, J. (1996) Neural Correlation via Random Connections. *Neural Computation* **8,** 1711–1729.

Cohen, N.J. & Eichenbaum, H. (1993) *Memory, Amnesia, and the Hippocampal System.* M.I.T. Press, Cambridge, Massachusetts.

Derrick, B.E. & Martinez Jr., J.L. (1996) Associative, bidirectional modifications at the mossy-fibre-CA3 synapse. *Nature* **381,** 429-434.

Diederich, J. (1989) Instruction and high-level learning in connectionist networks. *Connection Science* **1,** 161–180.

Diederich, J. & Hogan J.M. (1997) Constructive learning in connectionist semantic networks. In *Neural Networks and a new Artificial Intelligence,* (ed.) G. Dorffner.

Feldman, J. A. (1982) Dynamic connections in neural networks. *Bio-Cybernetics* **46,** 27–39.

Graf, P. & Schacter, D.L. (1985) Implicit and explicit memory for new associations in normals and amnesic subjects. *J. Exp. Psychol: LMC, 11* 501—518.

Hasselmo, M.E. (1997) Linking LTP to network function: A simulation of episodic memory in the hippocampal formation. In *Long-Term Potentiation*, vol 3. (eds Baudry, M. & Davis, J.) (MIT Press, Cambridge, Massachusetts) (in the press).

Hinton, G.E. & Anderson, J.A. (1981) (eds) *Parallel Models of Associative Memories* Erlbaum, Hillsdale, New Jersey.

Hippocampus **6,** 6 1996. Special Issue on Computational Models of Hippocampal Function in Memory.

Hopfield, J.J. (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* **79,** 2554–2558.

Insausti, R., Amaral, D.G. & Cowan, W.M. (1987) The entorhinal cortex of the monkey: II Cortical afferents. *J. Comp. Neurol.* **264,** 356–395.

Jackendoff, R. (1990) *Semantic Structures*, MIT Press, Cambridge, Massachusetts.

Kawagachi, Y. & Hama, K. (1987) Fast-spiking non-pyramidal cells in the hippocampal CA3 region, dentate gyrus and subiculum of rats. *Brain Res.* **425,** 351–355.

Knight, R.T. (1996) Contribution of human hippocampal region to novelty detection. *Nature* **382,** 256–259.

Linden, D.J. (1994) Long-term synaptic depression in the mamallian brain. *Neuron* **12,** 457–472.

Lisman, J. & Idiart, M.A.P. (1995) Storage of 7 ± 2 short-term memories in oscillatory subcycles. *Science 267,* 1512–1515.

Lynch, G. & Ambros-Ingerson, J. (1993) *Memory's Highway: Brain rules for encoding and organizing information.* Center for the neurobiology of learning and memory, University of California, Irvine, CA.

Marr, D. (1971) Simple memory: a theory for archicortex. *Phil. Trans. R. Soc.* B **262,** 23–81.

McNaughton, B.L. & Morris, R.G.M. (1987) Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends Neurosci.* **10,** 408–415.

Miles, R. & Wong, R.K.S. (1984) Unitary inhibitory synaptic potentials in the guinea-pig hippocampus *in vitro. J. Physiol.* **356,** 97–113.

O'Keefe, J. & Nadel, L. (1978) *The hippocampus as a cognitive map.* Clarendon Press, Oxford.

O'Reilly, R.C. & McClelland, J.L. (1994) *Hippocampal Conjunctive Encoding Storage, and Recall: Avoiding a Tradeoff.* Technical Report PDP.CNS.94.4, Carnegie Mellon University, Pittsburgh, PA.

Pinker S. (1989) *Learnability and Cognition: The Acquisition of Argument Structure* MIT Press, Cambridge, Massachusetts.

Plate, T. A. (1994) *Distributed Representations and Nested Compositional Structure* Ph.D. Thesis, Department of Computer Science, University of Toronto.

Schwartzkroin, P.A., Scharfman, H.E. & Sloviter, R.S. (1990) Similarities in circuitry between Ammon's horn and dentate gyrus: local interactions and parallel processing. In *Progress in Brain Research: Understanding the brain through the hippocampus* (eds Storm-Mathisen, J., Zimmer, J. & Ottersen, O.P.) 269–286, Elsevier Science, Amsterdam.

Schwartzkroin, P.A., & Prince, D.A. (1980) Changes in excitatory and inhibitory synaptic potentials leading to epileptogenic activity. *Brain Research,* 183 61—76.

Shastri, L. (1988) *Semantic Networks: An evidential formalization and its connectionist realization.* (see p. 181–191). Morgan Kaufmann, Los Altos/Pitman Publishing Company, London, 1988.

Shastri, L. (1997a) A Model of Rapid Memory Formation in the Hippocampal System, In the *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society,* 680–685, Stanford University, August 1997.

Shastri, L. (1997b) From transient patterns to persistent structures: a computational model of rapid memory formation in the hippocampal system. In preparation.

Shastri, L. & Ajjanagadde V. (1993) From simple associations to systematic reasoning: A connectionist encoding of rules, variables and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences* **16,** 417–494.

Sik, A., Ylinen, A., Penttonen, M., & Buzsaki, G. (1994) Inhibitory CA1-CA3-Hilar Region Feedback in the Hippocampus. *Science 265,* 1722–1724.

Singer, W. & Gray, C.M. (1995) Visual feature integration and the temporal correlation hypothesis. *Ann. Rev. Neurosci.* **18,** 555–586.

Squire, L.R. (1992) Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psych. Rev.* **99,** 195–231.

Stanton, P.K. & Sejnowski, T.J.(1989) Associative long-term depression in the hippocampus induced by hebbian covariance. *Nature* **339,** 215–218.

Suzuki, W.A. & Amaral, D.G. (1994) Perirhinal and parahippocampal cortices of the macaque monkey: Cortical Afferents. *J. Comp. Neurol.* **350,** 497–533.

Traub, R.D., Whittington, M.A., Stanford, I.M., & Jefferys, J.G.R. (1996) A mechanism for generation of long-range synchronous fast oscillations in the cortex. *Nature* **382,** 621–624.

Treves, A & Rolls, E.T. (1994) Computational analysis of the role of the hippocampus in memory. *Hippocampus* **4,** 374–391.

Tulving, E. (1983) *Elements of Episodic Memory.* Clarendon Press, Oxford.

Valiant, L. (1994). *Circuits of the mind.* New York : Oxford University Press.

Van Hoesen, G.W. (1982) The primate hippocampus gyrus: New insights regarding its cortical connections. *Trends Neurosci.* **5,** 345–350.

von der Malsburg, C. (1986) Am I thinking assemblies? In *Brain Theory* (eds Palm, G. & Aertsen, A.) 161–176. Springer-Verlag, Berlin.

West, M.J. (1990) 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* (eds Storm-Mathisen, J., Zimmer, J. & Ottersen, O.P.) 13–36, Elsevier Science, Amsterdam, 1990.

Wickelgren, W.A. (1979) *Learning and Memory* Prentice-Hall, Englewood Cliffs, New Jersey, 1979.

Willshaw, D.J., Buneman, O.P. & Longuet-Higgins, H.C. (1969) Non-holographic Associative Memory. *Nature* **222,** 960–962.

Zalutsky, R.A. & Nicoll, R.A. (1990) Comparison of two forms of long-term potentiation in single hippocampal neurons. *Science* **248,** 1619–1624.