BIDIRECTIONAL COMPLETION OF CELL ASSEMBLIES IN THE CORTEX

Friedrich T. Sommer, Thomas Wennekers and Günther Palm

Department of Neural Information Processing
University of Ulm, 89069 Ulm, Germany

INTRODUCTION

Reciprocal pathways are presumedly the dominant wiring organization for cortico-cortical long range projections[5]. This paper examines the hypothesis that synaptic modification and activation flow in a reciprocal cortico-cortical pathway correspond to learning and retrieval in a bidirectional associative memory (BAM): Unidirectional activation flow may provide the fast estimation of stored information, whereas bidirectional activation flow might establish an improved recall mode. The idea is tested in a network of binary neurons where pairs of sparse memory patterns have been stored in bidirectional synapses by fast Hebbian learning (Willshaw model). We assume that cortical long-range connections shall be efficiently used, i.e., in many different hetero-associative projections corresponding in technical terms to a high memory load. While the straight-forward BAM extension of the Willshaw model does not improve the performance at high memory load, a new bidirectional recall method (CB-retrieval) is proposed accessing patterns with highly improved fault tolerance and also allowing segmentation of ambiguous input. The improved performance is demonstrated in simulations. The consequences and predictions of such a cortico-cortical pathway model are discussed. A brief outline of the relations between a theory of modular BAM operation and common ideas about cell assemblies is given.

THE MODEL

Cortical circuitries with strong feedback are the most probable structures realizing Hopfield like associative memory modules[8]. Where anatomy ensures strong direct feedback, Hebbian synaptic learning[7], as verified experimentally[6], enhances the symmetry of the synaptic connections, since reciprocal synapses between a pair or a small group of simultaneously active neurons should be strengthened by roughly the same amount. Columnar networks of strongly interconnected pyramidal cells formed by local intra-cortical connections[3] have been ascribed to this kind of auto-associative memory function[1]. Recent systematic evaluations of anatomical tracer studies in animals shed light on the organization of cortico-cortical long-range connections: Strong reciprocal pathways between two cortical areas x and y – independent from their distance – have turned out to be a dominating connection scheme[5]. These pathways terminate in a patchy way and, strikingly, the patch size coincides with column size[12].

We model a reciprocal cortico-cortical pathway as a densely connected bidirectional hetero-associative memory (BAM), and do not consider explicitely the auto-associative intra-columnar connections (but see result section). Two pools of neurons x and y are fully bidirectionally connected. Synaptic modification is modeled by Hebbian learning of M pattern pairs: $\{(x^\nu, y^\nu) : x^\nu \in \{0,1\}^n, y^\nu \in \{0,1\}^m, \nu = 1, ..., M\}$. We use randomly generated memory patterns with constant numbers of '1'-entries, i.e., $|x^\nu| = \sum_i x_i^\nu = a, |y^\nu| = b \; \forall \; \nu$. Since in the cortex low activation is observed, the memory patterns are sparse, i.e., $a, b << n, m$. For this case the Willshaw model[16, 18] is efficient, where only minimum synaptic depth is required by binary clipping of the outer product sum:

$$C_{ij} = \sup_\nu x_i^\nu y_j^\nu. \tag{1}$$

During retrieval an associative memory maps an initial pattern $x(0)$ closest to a memory pattern x^ν onto a pattern close to y^ν. In the Willshaw model each neuron j forms the dendritic potential

$$d(0)_j := [Cx(0)]_j = \sum_i C_{ij} x(0)_i, \tag{2}$$

and determines its activity value by threshold comparison

$$y(1)_j = H(d(0)_j - \theta) \;\; \forall j, \tag{3}$$

with the global threshold value θ and $H(x)$ denoting the Heaviside function. In the following eqs. (2) and (3) are denoted as *simple retrieval*.

In auto-associative memories usually iterative retrieval is employed[8]. In parallel update schemes simple retrieval is iterated until a stable pattern is reached which is the retrieval result. Iterative retrieval improves the Willshaw model, if combined with an activity dependent threshold strategy or with a Boolean ANDing ($\wedge$) in each component between the old and the new input pattern applied after the second iteration step[13], i.e., $r \geq 1$:

$$x(r+1)_j = x(r)_j \wedge H(d(r)_j - \Theta) \; \forall j. \tag{4}$$

Biologically, Boolean ANDing could be realized by after-depolarization effects, where the cell response is fascilitated by its recent previous activation *(cell fascilitation)*[4, 9, 2].

Crosswise Bidirectional Retrieval

Iterative retrieval can be understood as a kind of high precision recall mode[1] which will now be extended to the hetero-associative case. At higher memory load the input error tolerance of simple retrieval is small, see Fig. 3. For bidirectional iteration of the retrieval as in the Kosko BAM[10] the modification (4) can be first applied after the third step which turns out to be too late to achieve significant retrieval error reduction[14]. Since in the hetero-associative matrix row and column sums are different, retrieval methods combining both in each step might be superior to a sequence of simple retrieval steps (2): In a decision whether a y-neuron is active in a particular memory pattern, we include lateral constraints between different y-units – also stored in the matrix – by the use of second order quantities, denoted as *conditioned links*, see Fig. 1.

A retrieval process based on this idea should return a pattern where '1'-components are connected by the highest conditioned links. We approximate such a process by using

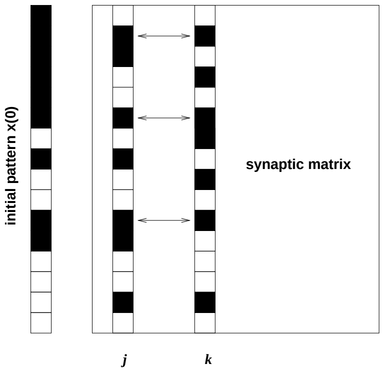

Figure 1. Conditioned link: Black squares correspond to '1'-entries in the pattern and columns. The link between y-units j and k conditioned by the pattern $x(0)$ is defined as overlap between matrix columns restricted on the rows $\{i : x(0)_i = 1\}$:

$$w_{jk}^{x(0)} := \sum_i C_{ij} C_{ik} x(0)_i$$

Provided $x(0)$ is a fair estimation of a learning pattern x^ν then a high conditioned link indicates high probability that $y_j^\nu = y_k^\nu = 1$. In the example $w_{jk}^{x(0)} = 3$.

conditioned links as auto-associative weights for the y-units. The resulting dendritic potentials can be transformed by $\sum_k w_{jk}^{x(r)} y(r\text{–}1) = \sum_{i \in x(r)} C_{ij} [C^T y(r\text{–}1)]_i$ (and similiarly in the x-layer). Obviously, they can be computed by a forth and back propagation through the ordinary synaptic matrix. The use of conditioned links in updates of both layers yields a new retrieval scheme. We call it *crosswise bidirectional (CB) retrieval* since matrix rows and columns are evaluated simultaneously:

$$y(r+1)_j = H\Big(\sum_{i \in x(r)} C_{ij} [C^T y(r-1)]_i - \Theta \Big) \tag{5}$$

$$x(r+1)_i = H\Big(\sum_{j \in y(r)} C_{ij} [C x(r-1)]_j - \Theta' \Big) \tag{6}$$

For $r = 0$ pattern $y(r\text{–}1)$ has to be replaced by the simple retrieval result $H([Cx(0)] - \theta)$, for $r > 2$ Boolean ANDing with results from timestep $r - 1$ can be applied.

RESULTS

Memory Performance

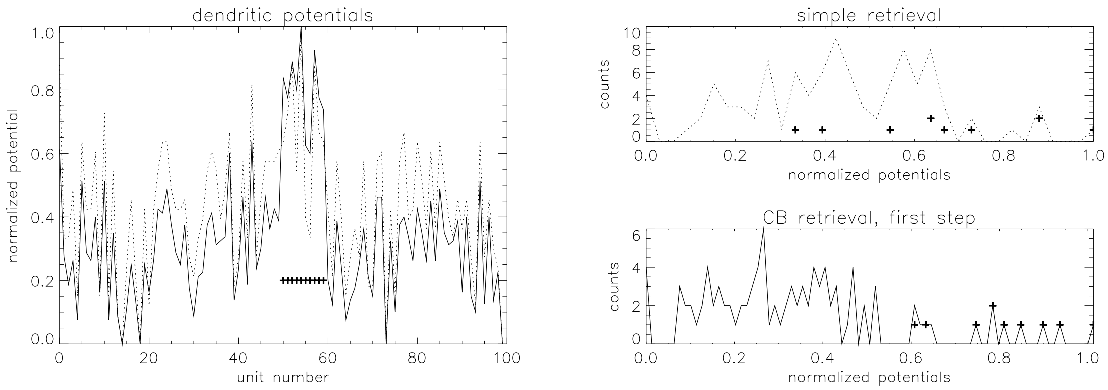

Figure 2. 36 stored pattern pairs with $n = m = 100$ and activities $a = b = 10$ resulting in a memory load of $p(C_{ij} = 1) = 0.3$: Dendritic potentials in the y units and their histograms obtained with an initial pattern corrupted by 35 wrong one-components. Crosses denote one-components in the addressed learning pattern.

Fig. 2 demonstrates in a small network with 100 neurons in each layer that already the first CB retrieval step produces highest dendritic potentials at sites of '1'-components in the addressed learning pattern whereas for simple retrieval a threshold detection of the '1'-components is impossible. Fig. 3 shows at a memory load

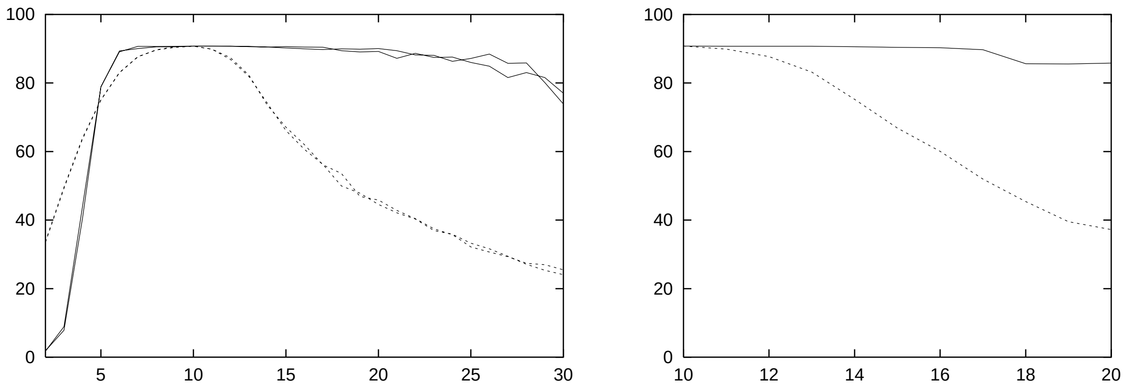

Figure 3. Mean output transinformation about a memory pattern for $n = m = 2000$, $M = 15000$, and again $a = b = 10$. Simple retrieval: dotted, CB-retrieval: solid. The x-axes display the address activity: $|x(0)| = 10$ corresponds to an errorfree memory pattern, lower activities are due to "miss" errors, higher activities are due to "add" errors. Left: Retrieval from addresses with random noise, two trials using different sets of randomly generated memory patterns. The output activities are adjusted near $|y| = b$ by threshold setting. Right: Retrieval from addresses where add errors are '1'-components of a second learning pattern.

$p(C_{ij} = 1) = 0.3$ retrieval results for addresses distorted either by random noise (left), or by a second learning pattern (right) where the address contains one learning pattern and '1'-components of a second learning pattern are successively added with increasing abscissa. The information in a memory pattern with $m = 2000$ and $b = 10$ is 90.8 bits. Lower transinformation values indicate output patterns disrupted by a nonvanishing retrieval error rate. For errorfree addresses, i.e., $|x(0)| = 10$, both retrieval methods work errorfree. However, with increasing input noise the transinformation in the simple retrieval results drops much faster than in the CB retrieval outputs*. With superimposed inputs simple retrieval behaves similiarly poor as with random noise. CB-retrieval does quite well, even if a complete superposition of two learning patterns is presented and yields almost perfectly one of the learning pattern pairs. Subsequently, the other pattern can be retrieved by the original address where the '1'-components from the first pattern pair have been deleted.

Thus, CB-retrieval exhibits much more fault tolerance than simple retrieval and can also be used for a segmentation of ambiguous inputs. The achieved experimental information capacity in the parameter range of Fig. 3 is strikingly high, reaching the order of the theoretical expectations of 0.52 bit/synapse. The higher performance is due to better exploitation of information stored in the Hebbian learning matrix and does not require more complex learning procedures[11] or dummy augmentation in the pattern coding[17] as has been proposed earlier to improve BAM. For a theoretical treatment and more detailed simulation results, see[15].

Implications and questions raised by the model

We have shown that information stored by a rough (binary clipped) hetero-associative Hebbian learning process can be recalled with significantly increased precision by bidirectional retrieval. It provides the fault tolerance and the ability of segmentation one would expect from associative memory realized in a biological system.

Our model of a reciprocal pathway assumes the existence of short closed loops where symmetric weights can be formed by Hebbian learning. This would be provided by a patchwise correspondence of both directions of a pathway, which has not

*The range of very small input activities is irrelevant since there retrieval errors are already high with both methods.

yet been examined experimentally. The time-delays in long reciprocal pathways can be significantly higher than in local connections. Therefore, the onset of the predicted hetero-associative high precision recall should be accompanied by a considerable slow-down in performance, although only a few iterations are required in the model. It may be related to cognitive or perceptual tasks, such as those where γ-oscillations where observed in behaving monkeys.

High precision recall can be realized by combinations of cell fascilitation and the use of quantities like conditioned links. How conditioned links can computed in detail by biological mechanisms in cortical pyramidal cells requires more detailed model assumptions. The present model, however, provides some hints which of its parts have to be refined: Eq. (5), for instance, prescribes the population of active neurons at one side of the pathway $x(t)$ for a short time to be fixed and to echo the signals from the other side (with linear dependence on the input). This could be realized by the interplay between fast and slow dendritic-neuronal processes: Increased slow components in the dendritic potentials would determine the population $x(t)$. Patterns of coinciding single spikes $y(t-1)$ affect the spike-rates in this cell population. The spike-rates influence the slow components in the dendritic potential and by this the new population of firing cells in $y(t+1)$, see Fig. 4. Another possible realization of eq. (5) is that the columnar auto-associative short-range connections stabilize the population $x(t)$.

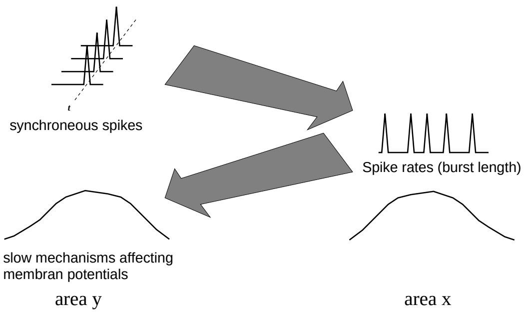

Figure 4. Formation of conditioned links in a cortico-cortical pathway: Increased slow components in the dendritic potentials of neurons in x determine the population of firing neurons. Coinciding activities on a fast timescale in area y affect the spikerates (with linear dependence on the input) in the x-cell population. The spikerates influence the slow components in the dendritic potential and by this the new population of firing cells in area y.

CONCLUSIONS

Three types of associative memory operations might play an important role in the cortex: auto-associative memory (AAM) using intra-cortical short range connections, one-way hetero-associative memory (HAM) on unidirectional cortico-cortical pathways, and bidirectional associative memories (BAM) on reciprocal cortico-cortical pathways.

Like in AAMs the feedback in a reciprocal pathway may allow recall procedures involving iterative retrieval providing higher precision at reduced speed. The idea is tested in a BAM model using the synaptic modification of the Willshaw model to store sparse binary random patterns. Simple bidirectional extension of the Willshaw model does not improve the performance. Indroducing the concept of conditioned links in BAM a new and very efficient BAM model for sparse memory patterns can be proposed. The high performance of the proposed CB-memory model and its segmantation ability for superimposed memory patterns is shown experimentally. We argue how the basic assumptions of our BAM model, namely symmetry of connections and high connectivity may be realized in a cortical pathway. Possible candidates for biological mechanisms corresponding to the computation of conditioned links might be the interplay of fast and slow dendritic processes and/or the influence of cortical short-range connections.

The proposed role of reciprocal cortico-cortical pathways as effective BAM devices supports Hebb's theory of concept representation by persistent distributed acticity patterns (cell assemblies): Fault tolerance and the capability to separate superpositions of several memory patterns – as provided by the CB-model – is crucial in a cell assembly theory[14]. Moreover, the proposed bidirectional retrieval in cortical pathways breaks up the monolithic cell assemblies of previous theories. Cell assemblies including many cortical areas can naturally emerge, if the areas are connected by reciprocal pathways. However, the cell assembly can be used in parts by restricting the recall on one or a few pathways involved (representing different facettes of a learned concept).

Acknowledgement: Two of the authors where supported by grants of the Deutsche Forschungsgemeinschaft: (F.T.S.) SO352/3-1 and (T.W.) PA268/8-1.

REFERENCES

1. D. A. Amit. The Hebbian paradigm reintegrated: Local reverberations as internal representations. *Behavioural and Brain Sciences*, 18:617 – 657, 1995.
2. A. Bibbig and T. Wennekers. Hippocamal two-stage learning and memory consolidation. In *Proceedings of the 13th European Meeting on Cybernetics and Systems Research*, 1996.
3. V. Braitenberg and A. Schüz. *The anatomy of the cortex. Statistics and Geometry.* Springer, Berlin Heidelberg New York, 1991.
4. M. Caesar, D. A. Brown, B. H. Gahwiler, and T. Knopfel. Characterization of a calcium-dependent current generating a slow afterdepolarization of CA3 pyramidal cells in rat hippocampal slice cultures. *European Journal of Neuroscience*, 5:560–569, 1993.
5. D. J. Felleman and D. C. Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1:1 – 47, 1991.
6. B. Gustafsson and H. Wigström. Physiological mechanisms underlying long-term potentiation. *Trends Neurosci.*, 11:156–162, 1988.
7. D. O. Hebb. *The Organization of Behaviour.* Wiley, New York, 1949.
8. J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, USA*, 79, 1982.
9. O. Jensen and J. E. Lisman. Novel lists 7 ± 2 known items can be reliable stored in an oscillatory short-term memory network: Interaction with long-term memory. *Learning & Memory*, 3:257–263, 1996.
10. B. Kosko. Adaptive bidirectional associative memories. *Applied Optics*, 26(23):4947–4971, 1987.
11. C.-S. Leung, L.-W. Chan, and E. Lai. Stability, capacity and statistical dynamics of second-order bidirectional associative memory. *IEEE Trans. Syst, Man Cybern.*, 25(10):1414–1424, 1995.
12. A. Schüz. Patchiness as a means to get a message across. *Trends in Neuroscience*, 17(9):365, 1994.
13. F. Schwenker, F. T. Sommer, and G. Palm. Iterative retrieval of sparsely coded associative memory patterns. *Neural Networks*, 9(3):445 – 455, 1996.
14. F. T. Sommer and G. Palm. Improved bidirectional retrieval of sparse patterns stored by Hebbian learning. *Submitted to Neural Networks*, 1997.
15. F. T. Sommer and G. Palm. Bidirectional retrieval from associative memory. In *Neural Information Processing Systems 10.* Lawrence Erlbaum Ass. Inc., 1998.
16. K. Steinbuch. Die Lernmatrix. *Kybernetik*, 1:36–45, 1961.
17. Y. F. Wang, J. B. Cruz, and J. H. Mulligan. Two coding stragegies for bidirectional associative memory. *IEEE Trans. Neural Networks*, 1(1):81–92, 1990.
18. D. J. Willshaw, O. P. Buneman, and H. C. Longuet-Higgins. Nonholographic associative memory. *Nature*, 222:960–962, 1969.