6

Interaction Between the Spatio-Temporal Learning Rule (Non Hebbian) and Hebbian in Single Cells: A Cellular Mechanism of Reinforcement Learning

Minoru Tsukada
Tamagawa University
Japan

1. Long term potentiation(LTP), depression(LTD) and Hebbian type learning rule

Hebb (1949) formulated the idea that modification is strengthened only if the pre- and post-synaptic elements are activated simultaneously (Fig.1). Experimentally, long term potentiation (LTP) and long term depression (LTD) are generally considered to be the cellular basis of learning and memory. Bliss & Lømo (1973) first found that high- frequency electrical stimulation ("tetanus," 100-500 Hz) effectively produced LTP in the hippocampal

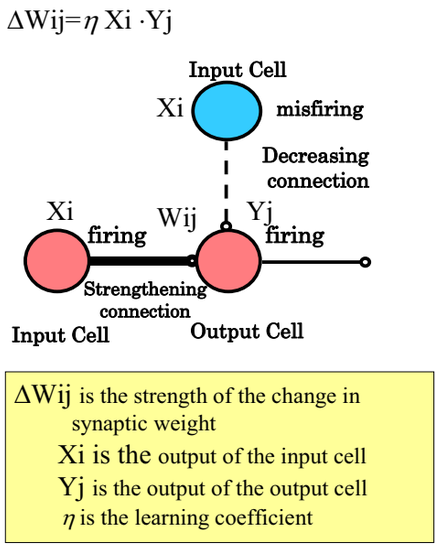

Fig. 1. The Hebbian Learning Rule

Source: Reinforcement Learning: Theory and Applications, Book edited by Cornelius Weber, Mark Elshaw and Norbert Michael Mayer
ISBN 978-3-902613-14-1, pp.424, January 2008, I-Tech Education and Publishing, Vienna, Austria

CA1 pyramidal cells. Recently, a series of experiments provided direct empirical evidence of Hebb's proposal (Markram et al., 1997; Magee & Johnston, 1997; Zhang et al., 1998; Feldman, 2000; Boettiger & Doupe, 2001; Sjostrom, 2001; Froemke & Dan, 2002). These reports indicated that synaptic modification can be induced by repetitive pairing of EPSP and back-propagating action potentials (BAPs). Pre-synaptic spiking within tens of milliseconds before postsynaptic spiking induced LTP whereas the reverse order resulted in LTD. This spike-timing-dependent LTP/LTD has been confirmed by using pyramidal cell pairs in hippocampal cultures, in which they found an asymmetric profile of LTP and LTD in relation to the relative timing between EPSPs and BAPs (Debanne & Thompson, 1998; Bi & Poo, 1998).

The influence of location dependency of synaptic modification along dendritic trees was examined in the CA1 area of rat hippocampal slices (Tsukada et al., 2005). A pair of electrical pulses was used to stimulate the Schaffer-commissural collaterals (SC) and stratum oriens (SO). Then we estimated the profile of LTP and LTD at a layer specific location from the proximal to distal region of the stratum radiatum.

Figure 2 shows the optical imaging results of LTP and LTD induced by a series of different spike timing (τ). The widest and strongest LTP was observed when simultaneous stimuli (τ= 0 ms) were applied. LTP decreased rapidly in space and time as the absolute value of relative timing increased to 15 ms on both sides. Accordingly, LTP was induced when back-propagating spikes (*Stim B*) were applied within a time window of 15 ms before and after the onset of *Stim A*, whereas LTD was induced on both sides at |τ| = 20 ms. Outside the 50 ms time window, synaptic modification disappeared. These instances of LTP and LTD showed a globally symmetric window of spike timing similar to a "Mexican hat function."

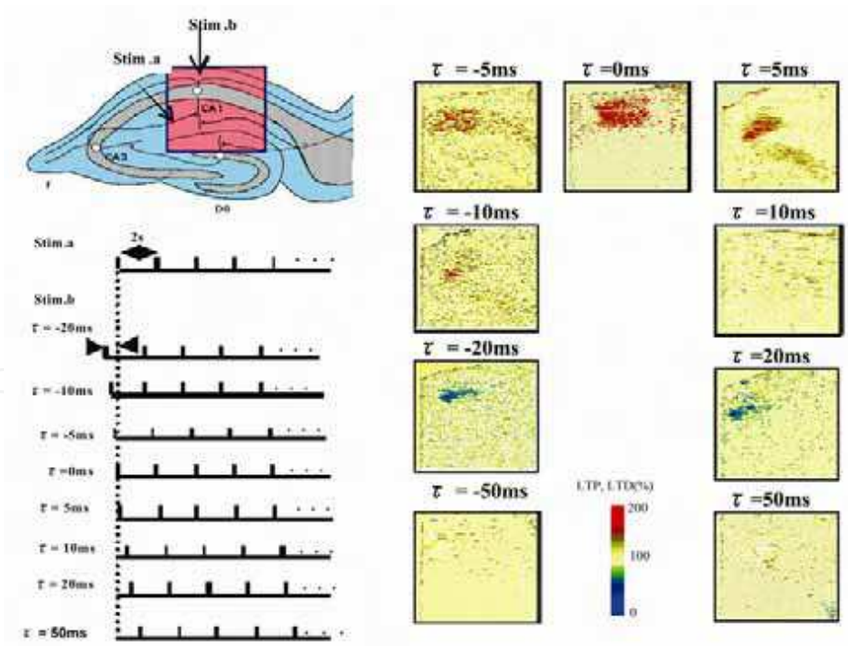

Fig. 2. Input-Output Timing Dependent LTP/LTD

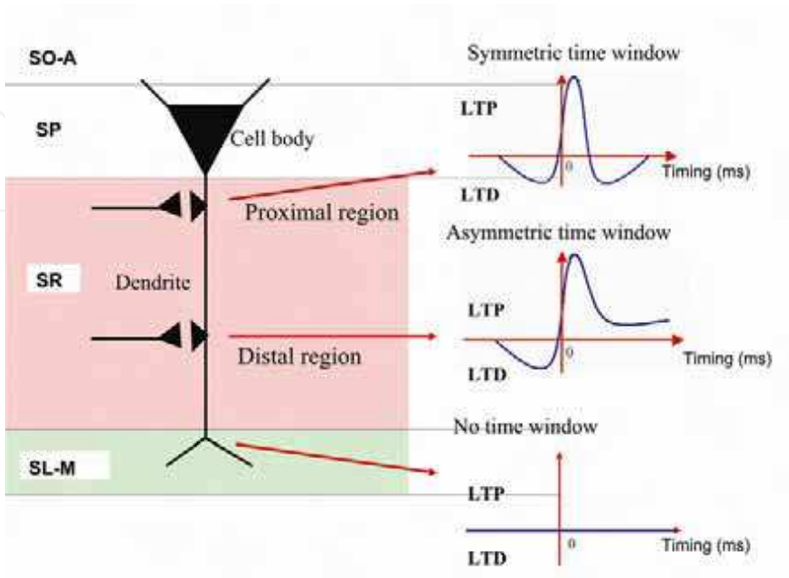

Fig. 3. Layer-specific profiles of LTP and LTD

We tested the location dependence of synaptic modification along dendritic trees. A symmetric window was obtained at the proximal region of the SR where GABAergic interneurons are projected, while an asymmetric window was obtained at the distal region of the SR where there is no projection of GABAergic interneurons (Fig 3).

The region-specific profiles of LTP/LTD depend on the network with or without the inhibitory projection on the layer of SR. Two factors of "after hyperpolarization" of spikes and "region-specific projection of inhibitory interneurons", which organize "lateral inhibition" for timing τ, underlie the "symmetric" profile for timing τ, while one factor of "after hyperpolarization" of spikes serves to organize the "asymmetric" profile. The "symmetric" profile, with a sharp window for τ, works as a coincidence detector between the input of CA3 Shaffer collaterals and the output of CA1 pyramidal cells. The time window corresponds to the time interval of a gamma cycle under the assumption that sequence information is processed in a time scheme of several gamma cycles (local) in a theta cycle (global) (Lisman, 1989; Aihara et al., 2000). On the other hand, the "asymmetric" profile, with a broad time window after τ= 0 ms, is able to integrate sequence information ("temporal summation") or to code phase information. This difference between the distal region and the proximal region of SR was seen in the results of temporal-pattern-dependent LTP using optical imaging of CA1 area in hippocampal slices (Aihara et al., 1997; Aihara et al., 2005). The sensitivity of LTP to the temporal pattern is even higher in the distal region than in the proximal region (Aihara et al., 2005). These results suggest an important function of memory processing depending on the synaptic localization on dendrites of CA1 pyramidal cells.

2. Spatio-temporal learning rule (non Hebbian)

The spatiotemporal learning rule (STLR), proposed as a non-Hebbian type by Tsukada et al. (1996) consisted of two defining factors: (a) cooperative plasticity (Input-Input timing coincidence) without a postsynaptic spike and (b) temporal summation (Fig 4).

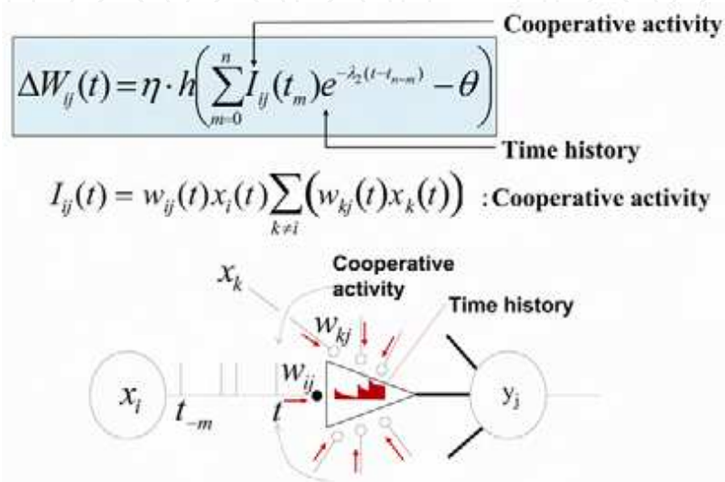

Fig. 4. Spatio-Temporal Learning Rule (STLR)

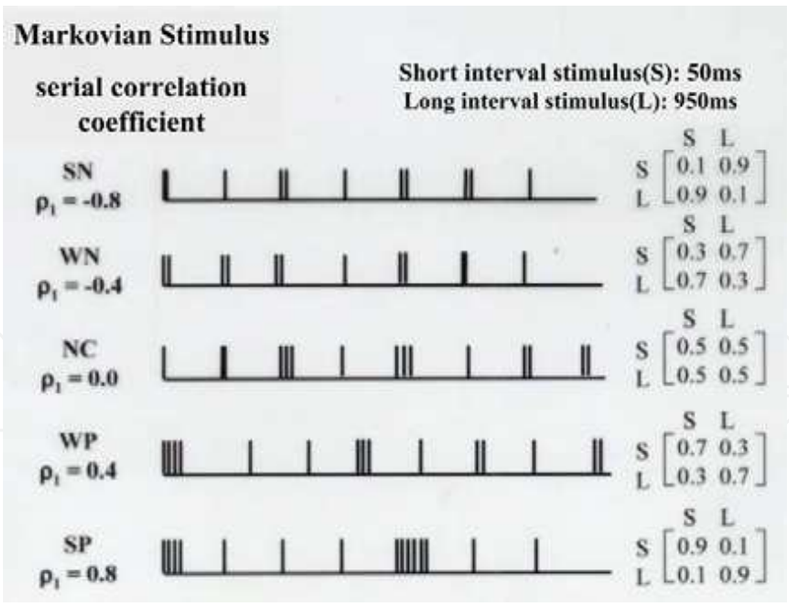

Fig. 5a. Temporal Pattern Stimuli

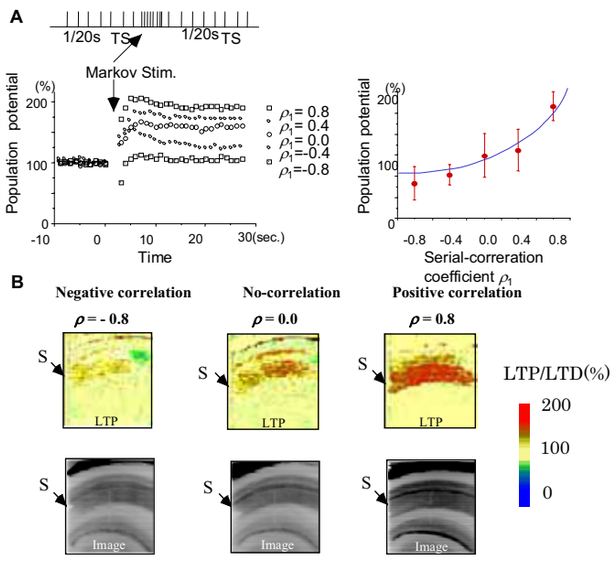

Fig. 5b. Temporal Sequence Pattern Dependent LTP/LTD - Effects of Markov Chain Stimulus –

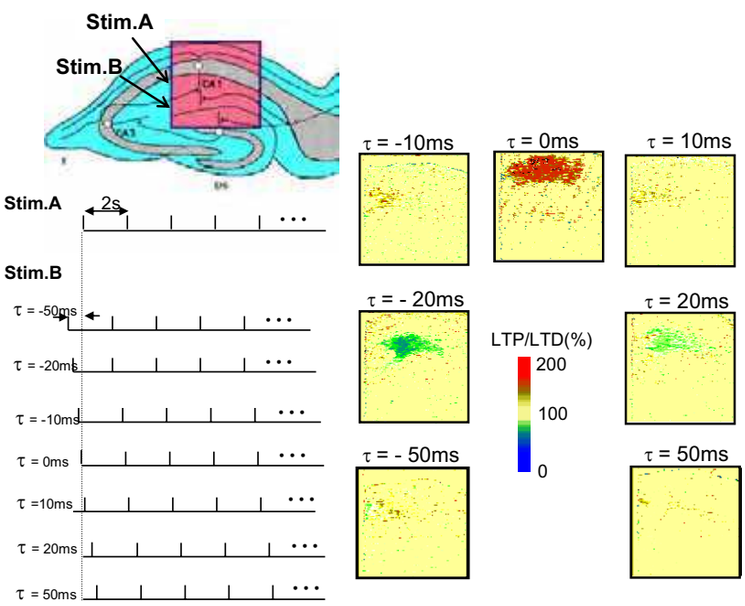

Fig. 6. Input-Input Timing-Dependent LTP/LTD

The neurophysiological evidence of "temporal summation" was obtained by applying temporal stimuli (Markov chain stimuli) to Schaffer collaterals of CA3 (Tsukada et al., 1994; 1996; Aihara et al., 2000) (Fig 5a,b). The cooperative plasticity (Fig 6) was measured by using two stimulus electrodes to stimulate the Schaffer-commissural collaterals (SC). First, electrode A was stimulated at 2 s intervals, but this did not produce any change in the synapse. Electrode B was then stimulated at a range of -50ms to 50ms with respect to electrode A. When the stimuli from both electrodes were simultaneous (relative timing τ=0), an extremely large plasticity appeared, but when it was shifted by 10 ms, there was almost no activity, and if it shifted another 10 ms, LTD appeared instead of LTP. When the relative timing was shifted 50 ms, it returned to normal. These data show that a time window exists in response to the relative timing τ. That is, the existence of a Mexican hat-shaped time window at the range of τ=±50ms. The coincidence of spike timing of Schaffer-collateral-paired stimulation of CA3 played a crucial role in inducing associative LTP (Tsukada et al., 2007). However it remains to be clarified whether the associative LTP is independent of back-propagated action potentials (BAPs) or not.

Only local dendritic depolarization at synaptic sites, such as theta-burst stimulation, can induce homosynaptic LTP evoked in the conditioning pathway by application of the associative pairing protocols to Schaffer collaterals even in the absence of BAP (in the presence of low TTX) (Golding et al., 2002). Robust homosynaptically induced LTP is observed in both the absence and presence of low TTX in the conditioning pathway (Fig.7a ; Tsukada et al., 2007). These results suggest that homosynaptic LTP by the present pairing protocol is induced under the condition of inhibiting activation of dendritic Na^+ channels.

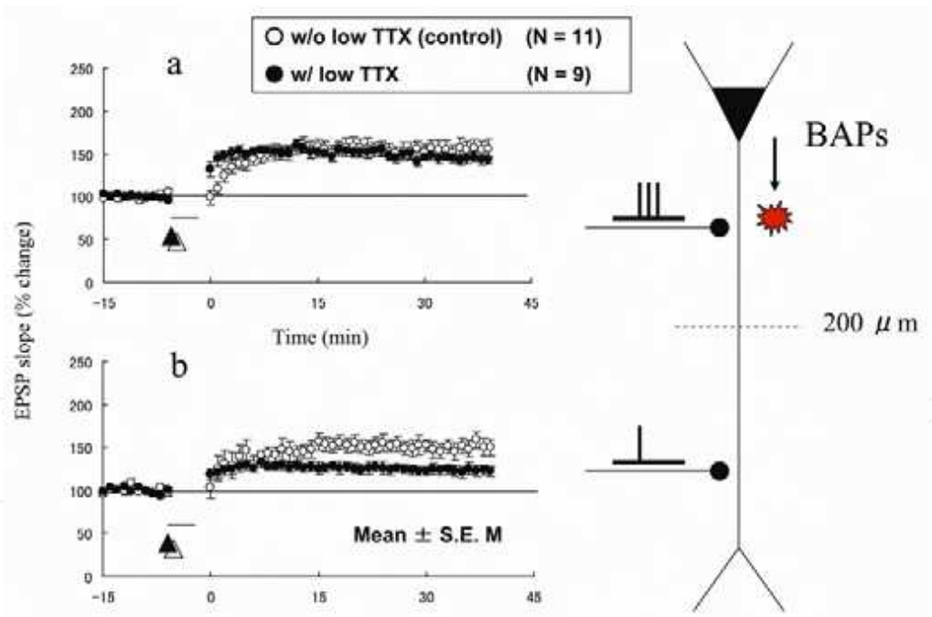

Fig. 7a/b. Input-input timing-dependent LTP can be induced independent of backpropagating action potentials (BAPs)

However, in the same preparation, the magnitude of the heterosynaptically induced LTP in association with conditioning bursts is reduced, while a considerable amount of the LTP was preserved in the presence of low TTX (Fig.7b). Homo-synaptic and hetero-synaptic associative LTP can be induced under conditions of inhibited BAPs, even in the absence of a cell spike. If the two inputs synchronize at the dendritic synapse of CA1 pyramidal cells, then the synapse is strengthened, and the functional connection is organized on the dendrite. If the two inputs are asynchronous then the connection is weakened. A schematic representation was drawn in Fig.8. The functional connection/disconnection depends on the input-input timing dependent LTP (cooperative plasticity). This is different from the Hebbian learning rule, which requires coactivity of pre- and post-cell. However, the magnitude LTP is also influenced by BAPs. From these experimental results, it can be concluded that the two learning rules, STLR and HEBB, coexist in single pyramidal neurons of the hippocampal CA1 area.

STLR (non-Hebbian) incorporates two dynamic processes: fast (10 to 30 ms) and slow (150 to 250 ms). The fast process works as a time window to detect spatial coincidence among various inputs projected to a weight space of the hippocampal CA1 dendrites, while the slow process works as a temporal integrator of a sequence of events. In a previous paper in which parameters were fitted to the physiological data of LTP's time scale (Aihara et al., 2000), the decay constant of fast dynamics was identified as 17 ms, which matches with the period of hippocampal gamma oscillation. The decay constant of the slow process is 169 ms, which corresponds to a theta rhythm. This suggests that cell assemblies are synchronized at two time scales in the hippocampal- cortical memory system, and this is closely related to the memory formation of spatio-temporal context.

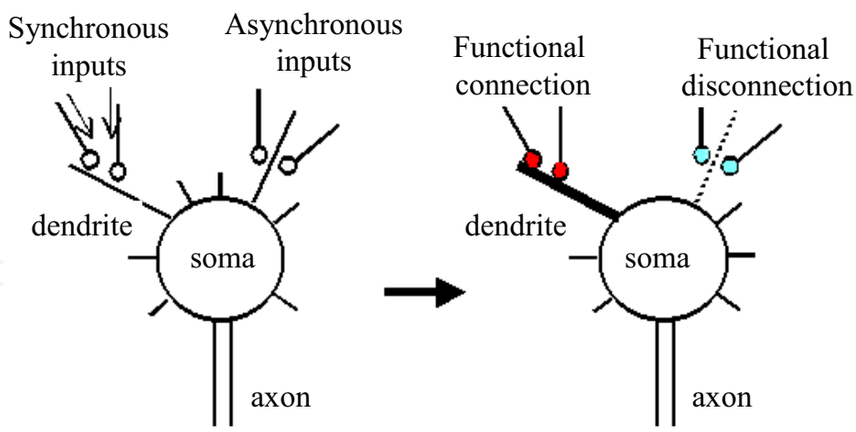

Fig. 8. A schematic presentation of synaptic potentiation or depression by the synchronous or asynchronous inputs

3. The functional differences between STLR and HEBB

Two rules are applied to a single-layered feed-forward network with random connections (Fig 9a) and their abilities to separate spatiotemporal patterns are compared with those of other rules, including the Hebbian learning rule and its extended rules (Tsukada & Pan, 2005). The elements of input patterns are connected to each neuron through a separate weight w_{ij} (i =1,2,...,N, j =1,2, ...,N). The potential of each neuron depends both on a weighted sum of the simultaneously provided inputs (spatial summation) and inputs arrived in the near past (temporal summation).
The above mentioned functions are expressed in the following equations.

Spatial summation:

$$p_i(t_n) = \sum_{j=1}^{N} w_{ij}(t_n) x_j(t_n) \tag{1}$$

Temporal summation:

$$y_i(t_n) = \sum_{m=0}^{n} p_i(t_m) e^{-\lambda_1 (t_n - t_m)} \tag{2}$$

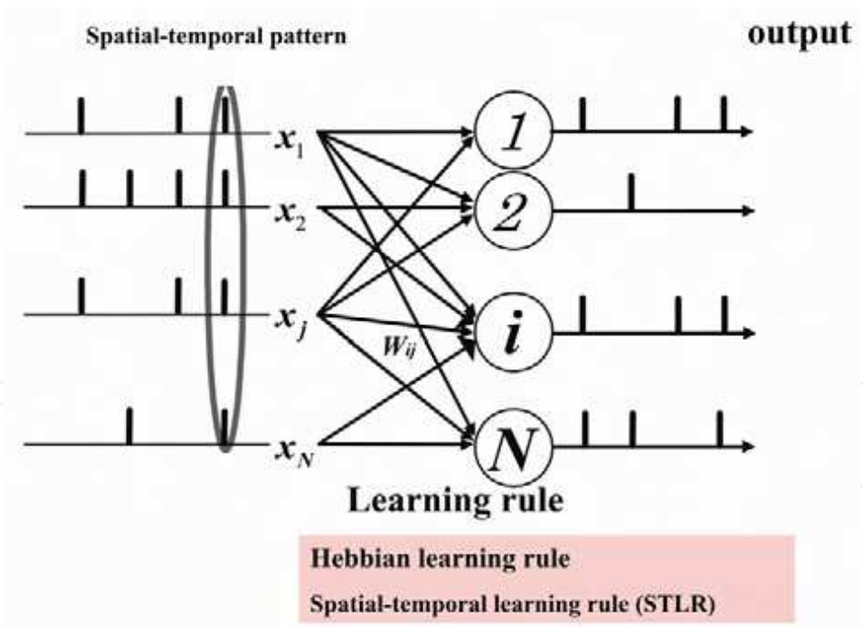

Fig. 9a. All Connected 1 Layer Neural Network

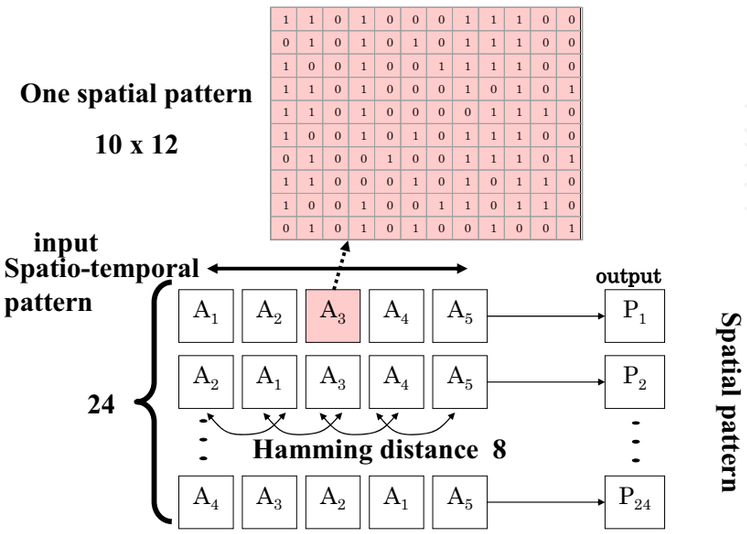

Fig. 9b. The Spatio-Temporal Pattern of Inputs

And the output of the neuron:

$$r_i(t_n) = f(y_i(t_i) - \theta_1) \tag{3}$$

where a set of label x_1, x_2, ..., x_N are inputs to neurons, $x_i(t_n)$ is an input to i neuron at time t_n (n=1,2, ..., n), $w_{ij}(t_n)$ is a synaptic weight from neuron j to neuron i at time t_n, $y_i(t_n)$ is the potential of neuron i at time t_n. $r_i(t_n)$ is its output, λ_1 is time decay constant of temporal summation, θ_1 is threshold. The output function of neurons is defined as:

$$f(u) = \begin{cases} 1 & u > 0 \\ 0 & u \le 0 \end{cases} \tag{4}$$

The spatiotemporal pattern used in this simulation consists of 5 frames of spatial patterns (Fig. 9b), i.e., A_1, A_2, A_3, A_4, A_5 (A_i is a spatial frame).
Every frame consists of N elements (N=120) and each element is chosen as "1" or "0" randomly, but the total number of "1"s is maintained throughout the various spatial patterns (in this simulation, half of the elements in one spatial frame are “1”, and the other half are “0”). The Hamming distance (HD) between every two spatial patterns is 8 bits (if not specified in the simulation). In some cases it is 2 or 24 bits (mentioned). Calculating all of the permutations of 4 spatial patterns, 24 spatiotemporal patterns were grouped as a training set. The last frame of each spatiotemporal pattern is the same (A_5). During the learning process, the 24 spatiotemporal patterns in the training set were learned by each neural network under the same initial conditions. After finishing the learning course, a test pattern (same as the learned pattern) was applied to the networks to attain an output-pattern (for

each learning rule, the threshold of neurons θ_1 is set so that about half of the elements in the output-pattern are "1"). We compared HDs between output-patterns for each learning rule. The averaged HD is often adopted to compare the ability of discriminating spatiotemporal patterns, which is defined as:

$$\text{averaged HD} = \frac{\sum(\text{number of pairs} * \text{HD of this pair})}{\sum \text{number of pairs}}$$

Three learning algorithms were used to train each of 24 spatiotemporal input patterns in single-layer neural network models. Each of the neural networks had the same initial condition. The differentiation of output-patterns represented in learned networks was analyzed by their Hamming distances (Fig.10a). HEBB produced the same output pattern, with a Hamming distance of zero, for all of the different spatiotemporal input patterns (Fig. 10a). This proves that the Hebbian learning rule cannot discriminate different spatiotemporal input patterns. Covariant Hebbian rule showed a slight improvement in their pattern separation ability (Fig.10a). The spatiotemporal learning rule had the highest efficiency in discriminating spatiotemporal pattern sequences (Fig.10a). The novel features of this learning rule were induction of cooperative plasticity without a postsynaptic spike and the time history of its input sequences. According to the Hebbian rule, connections strengthen only if the pre- and post-synaptic elements are activated simultaneously, and thus, the Hebbian rule tends to map all of the spatio-temporal input patterns with identical firing rates into one output pattern. HEBB has a natural tendency to attract analogous firing patterns into a representative one, in the simple word "pattern completion". In contrast, the spatio-temporal rule produces different output patterns depending on each individual input pattern. From this, the spatiotemporal learning rule has a high ability in pattern separation, while the Hebbian rule has a high ability in pattern completion. Finally, the network trained by the spatiotemporal learning rule produced the widest bimodal-distribution of Hamming distance (Fig10b), which shows that it has the highest efficiency in pattern separation.

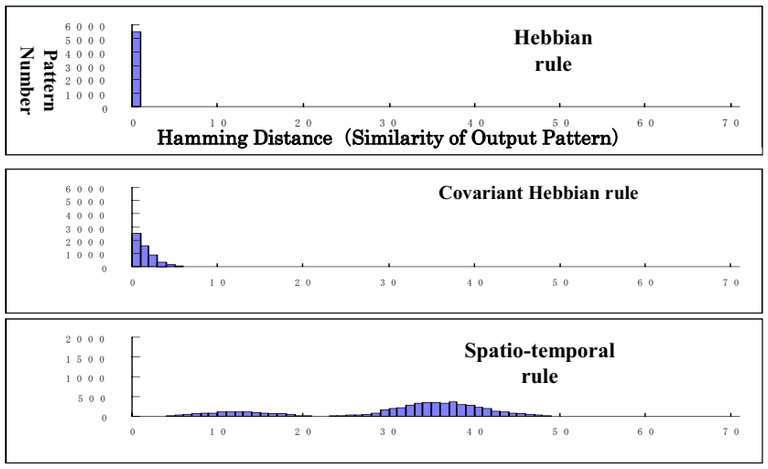

Fig. 10a. Output Pattern Distribution

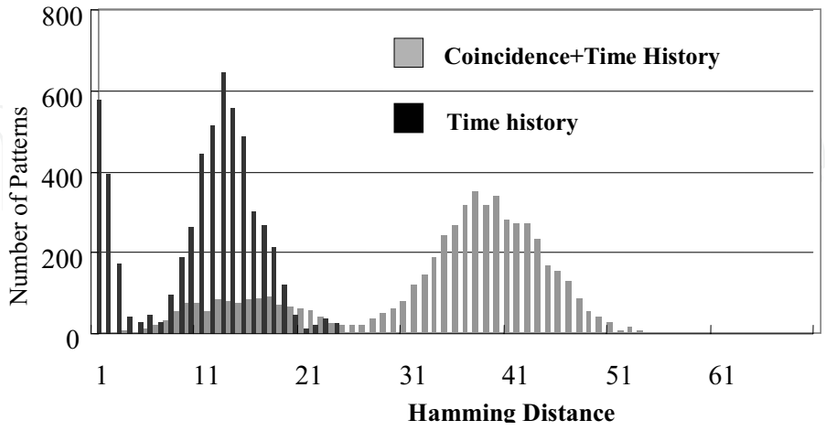

Fig. 10b. The Effect of Input Timing-Coincidence on Output Distribution

The two factors responsible for the high efficiency in pattern separation are spatial coincidence and temporal summation. The network trained by the learning rule without spatial coincidence produced the one-modal distribution. From this fact, we can conclude that the distribution in the longer range of the bimodal distribution (Fig.10b) in the histogram is generated by the spatial coincidence factor while the distribution in the short range is generated by the spatiotemporal summation. Thus, the ability of separating patterns in the network can be improved by introducing two factors: spatiotemporal summation and spatial coincidence, but the latter is more important.

4. Interaction of both rules in a dendrites-soma system

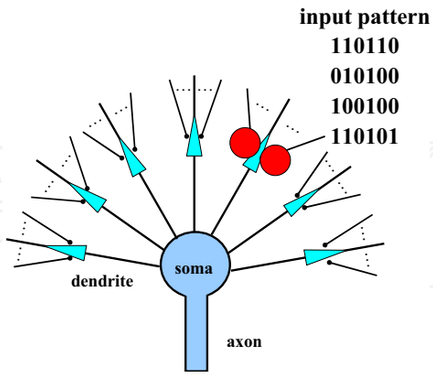

Fig. 11a. The Change in Synaptic weight according to Hebbian Learning Rule

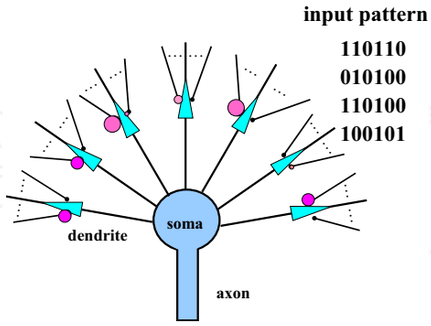

Fig. 11b. The Change in Synaptic Weight according to the Spatio-temporal Learning Rule

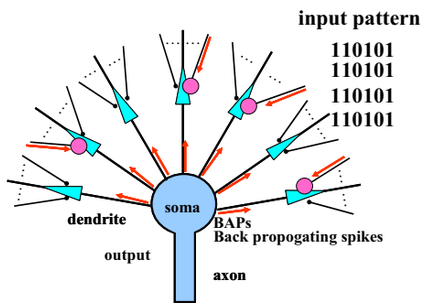

Fig. 11c. The Function of Local (dendrite) –Global (soma) Interaction and the Role of Back Propagating Spikes (BAPs)

The extension of the theoretical simulation results imply that this phenomenon occurs in a dendrites-soma system in a single pyramidal cell with many independent local dendrites in the CA1 area of the hippocampus. This system includes a spine structure, NMDA receptors, and sodium and calcium channels. The pyramidal cell integrates all of these local dendrite functions. The spatiotemporal learning rule and the Hebbian rule coexist in single pyramidal neurons of the hippocampal CA1 area (Tsukada et al., 2007). The Hebbian rule leads to the pattern completion and the spatiotemporal learning rule leads to the pattern separation.

Schematic illustrations were drawn in Figure-11abc. HEBB leads to pattern completion (Fig.11a). In contrast, STLR leads to pattern separation (Fig.11b). In the spatiotemporal learning rule, synaptic weight changes are determined by the "synchrony" level of input neurons and its temporal summation (bottom-up) whereas in the Hebbian rule, the soma fires by integrating dendritic local potentials or by top-down information such as environmental sensitivity, awareness, and consciousness. The coexistence of the spatiotemporal learning rule (local information) and the Hebbian rule (global information) on the neuronal level may support this dynamic process that repeats itself until the internal model fits the external environment (Fig 11c). The dendrite-soma interaction (Fig 11c) in pyramidal neurons of the hippocampal CA1 area can play an important role in the context formation of policy, reward, and value in reinforcement learning.

5. Mechanisms of reinforcement learning in single cells

The role of soma spiking in relation to top-down information raises a number of interesting computational predictions. Hippocampal theta is one of the candidates of top-down information which is driven by the medial septum (Buzsaki et al.,1983). The theta stimulation of adult rat hippocampal synapses can induce LTP (Thomas et al.,1998). Another candidate is extrinsic modulation by acetylcholine, serotonin, norepinephrine and dopamine. They can alter neuronal throughput and BAPs (so-called "meta-plasticity") in such a way that these transmitters diffuse broadly.

6. References

Aihara, T.; Tsukada, M.; Crair, MC. & Sinomoto, S. (1997). Stimulus-Dependent Induction of Long-Term Potentiation in CA1 Area of the Hippocampus: Experiment and Model. *Hippocampus* 7, p.416-426.

Aihara, T.; Tsukada, M. & Matsuda, H. (2000). Two dynamic processes for the induction of long-term in hippocampal CA1 neurons. *Biol. Cybern.* 82: 189-195.

Aihara, T.; Kobayashi, Y. & Tsukada, M. (2005). Spationtemporal visualization of long-term potentiation and depression in the hippocampal CA1 area. *Hippocampus* 15: 68-78.

Bi, G. & Poo, M. (1998). Synaptic modifications in cultured hippocampal neurons; dependence on spike timing, synaptic strength, and postsynaptic type. *J. Neurosci* 18:10464-10472.

Bliss, TP. & Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anesthetized rabbit following stimulation of perforant path, *The Journal of Physiology,* 232: 331-356.

Boettiger, CA. & Doupe, AJ. (2001). Developmentally restricted synaptic plasticity in a songbird nucleus required for song learning. *Neuron* 31: 809-818.

Buzsaki, G.; Leung, L. & Vanderwolf, CH. (1983). Celluler bases of hippocampal EEG in the behaving rat. *Brain Res. Rev.* 6: 169-171.

Debanne, D. & Thompson, SM. (1998). Associative long-term depression in the hippocampus in vitro. *Hippocampus* 6:9-16.

Feldman, DE. (2000). Timing based LTP and LTD at vertical inputs to layer II/III pyramidal cells in rat barrel cortex. *Neuron* 27:45-56.

Froemke, RC. & Dan, Y. (2002). Spike-timing-dependent synaptic modification induced by natural spike trains. *Nature* 416:433-438.

Golding, NL.; Staff, NP. & Spruston, N. (2002). Dendritic spikes as a mechanism for cooperative long-term potentiation. *Nature* 418(6895):326-31.

Hebb, DO. (1949). *The organization of behavior.* John Wiley, New York

Lisman, J. (1989). A mechanism for Hebb and the anti-Hebb processes underlying learning and memory. *Proc Natl Acad Sci USA* 86: 9574-9578.

Magee, JC. & Johnston, D. (1997). A synaptically controlled, associative signal for Hebbian plasticity in hippocampal neurons. *Science* 275(5297):209-13.

Markram, H.; Lubke, J.; Frotscher, M. & Sakmann, B.(1997). Reguration of synaptic efficacy by coincidence of postsynaptic Aps and EPSPs. *Science* 275:213-215.

Sjostrome, PJ. (2001). Rate timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron* 32:1149-1164.

Thomas, MJ.; Watabe, AM.; Moody, TD.; Makhinson, M. & O'Dell, TJ. (1998). Postsynaptic complex spike bursting enables the induction of LTP by theta frequency synaptic stimulation. *J Neurosci* 18(18):7118-26.

Tsukada, M.; Aihara, T.; Mizuro, M.; Kato, H. & Ito, K. (1994). Temporal pattern sensitivity of long-term potentiation in hippocampal CA1 neurons. *Biol. Cybern.* 70: 495-503.

Tsukada, M.; Aihara, T.; Saito, H. & Kato, H. (1996). Hippocampal LTP depends on spatial and temporal correlation of inputs. *Neural Networks* 9: 1357-1365.

Tsukada, M. & Pan, X. (2005). The spatiotemporal learning rule and its efficiency in separating spatiotemporal patterns. *Biol. Cybern.* 92: 139-146.

Tsukada, M.; Aihara, T.; Kobayashi, Y. & Shimazaki H. (2005). Spatial analisis of spike-timing-dependant LTP and LTD in the CA1 area of hippocampal slices using optical imaging. *Hippocampus* 15: 104-109.

Tsukada, M.; Yamazaki, Y. & Kojima, H. (2007). Interaction between the Spatio-Temporal Learning Rule (STLR) and Hebb type (HEBB) in single pyramidal cells in the hippocampal CA1 Area. *Cognitive Neurodynamics* 1: 157-167.

Zhang, LI.; Tao, HW.; Holt, CE.; Harris, WA. & Poo, M. (1998). A critical window for cooperation and competition among developing retino-tectal synapses. *Nature* 395: 37-44.

Reinforcement Learning

Edited by Cornelius Weber, Mark Elshaw and Norbert Michael Mayer

ISBN 978-3-902613-14-1
Hard cover, 424 pages
Publisher I-Tech Education and Publishing
Published online 01, January, 2008
Published in print edition January, 2008

Brains rule the world, and brain-like computation is increasingly used in computers and electronic devices. Brain-like computation is about processing and interpreting data or directly putting forward and performing actions. Learning is a very important aspect. This book is on reinforcement learning which involves performing actions to achieve a goal. The first 11 chapters of this book describe and extend the scope of reinforcement learning. The remaining 11 chapters show that there is already wide usage in numerous fields. Reinforcement learning can tackle control tasks that are too complex for traditional, hand-designed, non-learning controllers. As learning computers can deal with technical complexities, the tasks of human operators remain to specify goals on increasingly higher levels. This book shows that reinforcement learning is a very dynamic area in terms of theory and applications and it shall stimulate and encourage new research in this field.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen