

Mayr C, Noack M, Partzsch J, Schüffny R (2010) Replicating experimental spike and rate based neural learning in CMOS. In: IEEE International Symposium on Circuits and Systems ISCAS 2010, pp 105–108

Copyright 2010 IEEE. Reprinted, with permission from the above source

# Replicating Experimental Spike and Rate Based Neural Learning in CMOS

Christian Mayr, Marko Noack, Johannes Partzsch and René Schüffny  
 Institute of Circuits and Systems, University of Technology Dresden, Germany  
 Email: mayr, noack, partzsch, schueffn@iee.et.tu-dresden.de

**Abstract**—The computational function of neural networks is thought to depend primarily on the learning/plasticity function carried out at the synapse. Neuromorphic circuit realizations have taken this into account by implementing a variety of synaptical processing functions, with most recent synapse circuits replicating some form of Spike Time Dependent Plasticity (STDP). However, STDP is being challenged by older rate-dependent learning rules as well as by biological experiments exhibiting more complex timing rules (e.g. spike triplets) as well as simultaneous rate- and timing dependent plasticity. In this paper, we present a circuit realization of a plasticity rule based on the postsynaptic neuron potential as well as the transmission profile of the presynaptic spike [1]. To the best of our knowledge, this is the first circuit realization of synaptical behaviour which moves significantly beyond STDP, replicating the triplet experiments of Froemke and Dan [2], the combined timing and rate experiments of Sjoestrom et al. [3], as well as conventional BCM behaviour [4].

## I. INTRODUCTION

Based on experimental results, various mathematical expressions of the synaptical learning function have been proposed in literature, such as the Bienenstock-Cooper-Munroe (BCM) rule [4], STDP [5], or more involved spike timing [2] and membrane voltage [3] dependent rules. Several of those have been implemented in neuromorphic hardware to replicate their learning functions and/or computational properties [6], [7], [8], [9]. In this paper, the circuit implementation of a novel learning rule [1] based on local synaptic state variables in a UMC 130nm CMOS technology is presented. This rule can reproduce a variety of recent experimental results and is simple to implement, since it derives most of its dynamics from the neuron and the reconstruction of the incoming pulse, so the synapse complexity itself can be kept to a minimum. In Sec. II, the learning rule is introduced, with circuit realizations of the waveforms necessary for its implementation given in Sec. III-A. The synaptical circuit realizing the computation of the learning function is described in Sec. III-B. Sec. IV details results obtained when simulating the complete circuit realization using various major experimental protocols.

## II. THE BCM PLASTICITY RULE

Our circuit implementation is based on a BCM plasticity rule [1] that combines the well-established BCM formulation [4] with spike-based plasticity mechanisms. Thereby, presynaptic spikes are detected by a conductance variable  $g(t)$ , which is representative of the postsynaptic current (PSC) evoked by a presynaptic spike arriving at the synapse. Postsynaptic spikes are read out via the neuron's membrane potential

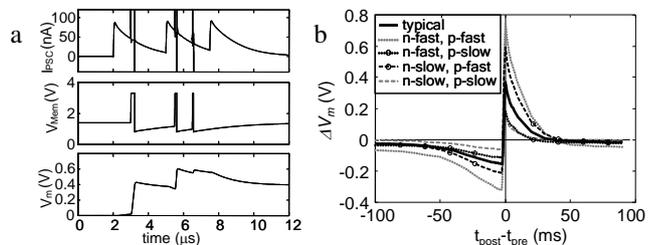


Fig. 1. (a) progress of the weight induced by a sample spike train (circuit time); (b) weight change produced by the circuit for spike pairings, simulated at typical and corner cases (biological time).

$u(t)$ . These variables are combined in a BCM-like fashion to determine the synaptic weight  $m(t)$  [1]:

$$\frac{dm(t)}{dt} = (u(t) - \Theta_u) \cdot g(t), \quad (1)$$

with a voltage threshold parameter  $\Theta_u$ . Normally, this threshold is equal to the membrane's resting potential, but it may be varied to change the ratio of weight increase (potentiation) to decrease (depression), thus corresponding to the frequency threshold in the original BCM rule.

As can be seen from Eq. 1, the synaptic weight is always increased if presynaptic activity coincides with a membrane voltage above the threshold  $\Theta_u$ . Conversely, the weight is decreased if the membrane voltage is below  $\Theta_u$  at presynaptic activity (cf. Fig. 1a). In particular, a postsynaptic spike, due to its high positive amplitude, leads to a steep weight increase, whereas the reset of the membrane below resting potential after a postsynaptic spike (hyperpolarization) results in elongated weight decrease. However, these weight changes are subject to presynaptic activity, reflected in  $g(t)$ . In the model, this variable is set to a fixed value at each presynaptic spike and decays exponentially. Likewise, the membrane potential is set to a fixed value below rest after each postsynaptic spike and decays exponentially afterwards. The interaction of all these mechanisms is shown in Fig. 1a, extracted from our circuit implementation, with  $u(t)$  corresponding to  $V_{Mem}$ ,  $g(t)$  to  $I_{PSC}$ , and  $m(t)$  to  $V_m$ . Note that this implementation operates at an acceleration of  $10^4$  compared to biological time. However, for easier comparison with experimental results in the rest of this paper, we convert the time base used in our circuit simulations back to biological time (cf. Fig. 1b).

The above rule exhibits the typical STDP behaviour for spike pairings, as is analytically shown in [1]: If a presynaptic

spike occurs,  $g(t)$  is high, so that a postsynaptic spike occurring shortly afterwards will strongly increase the synaptic weight, before the membrane hyperpolarization leads to a slight weight decrease. If the order of the spikes is reversed, no potentiation occurs because  $g(t)$  is zero at the postsynaptic spike, so that the weight is only decreased. This leads to the temporal asymmetry seen in classical STDP experiments [5]. Fig. 1b shows the corresponding time window of our circuit. Note that the areas under the potentiation and depression part of the time window are roughly equal, which is a typical property of STDP [5]. For our rule, this results in equal area under the pulse and the hyperpolarization curve, meaning that the pulse amplitude needs to be much higher than that of the hyperpolarization, which is biologically realistic [1].

### III. CIRCUITS FOR THE BCM PLASTICITY RULE

The waveforms necessary for the plasticity rule of [1] are generated in the presynaptic as well as the postsynaptic neuron. This reduces the synapse itself to the simple multiplication of Eq. 1, which is very advantageous in VLSI implementations. Especially for conventional matrix implementations [9], [7], where the number of synapses scales quadratically with the number of neurons, transferring functionality from the synapse to the neuron and PSC reconstruction situated at the edge of the matrix [9] significantly reduces the overall complexity. Also, as long as the multiplication is carried out with sufficient matching between synapses, all other error sources (e.g. the STDP time windows) are contained in the waveform generation circuits. Since those scale linearly with the number of neurons, it is computationally feasible to adjust them on a much finer level than the synapses.

#### A. Neuron and PSC generation

The neuron circuit proposed for validating the learning rule from above is a leaky integrate-and-fire neuron, which is able to reproduce the waveform of  $V_{Mem}$  in Fig. 1a. The circuit diagram is shown in Fig. 2a. Features of this neuron are a tunable resting potential unequal to zero, adjustable membrane time constant, threshold voltage and duration of an action potential. An adaptation mechanism which modulates the pulse duration against the time between current and previous spike [1] is also implemented.

1) *operation*: All currents flowing into the neuron are integrated on capacitance  $C_{mem}$  and increase the neuron's membrane potential  $V_{mem}$ . When reaching a certain threshold  $V_{thr}$  an action potential is generated. For the duration of a spike, digital signals  $SPK$  and  $\overline{SPK}$  are set. After this the membrane potential is reset to the voltage  $V_{reset}$  below resting potential and runs through a hyperpolarization curve.

2) *leakage*: The leakage module in Fig. 2b basically consists of the OTA structure presented in [10]. A negative feedback lets the circuit act like a resistor with resistance  $R_{leak}$ , which can be adjusted by changing the bias current  $I_{leak}$ . The resting potential is controlled by  $V_{rest}$  at the noninverting input. If no external currents are flowing, the

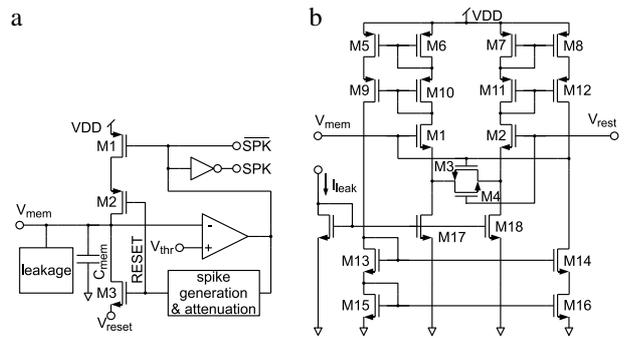


Fig. 2. (a) top level of neuron (b) leakage circuit

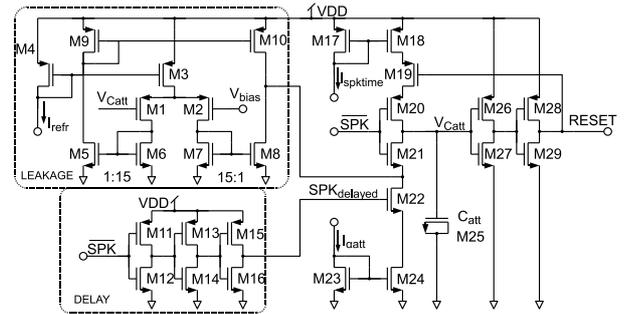


Fig. 3. Spike generation and attenuation circuit: The LEAKAGE circuit generates an exponential decay of  $V_{Catt}$  towards  $V_{bias}$ . The DELAY circuit delays and inverts signal  $\overline{SPK}$  to  $SPK_{delayed}$ .

membrane potential discharges exponentially with time constant  $\tau_{mem} = R_{leak} \cdot C_{mem}$  towards  $V_{rest}$ . By using a balanced OTA, the symmetry is improved and the offset is reduced [11]. This design is successfully applied in [6]. Cascode current mirrors (M5 - M16) fully reduce the offset. To allow larger time constants the current mirrors divide the output current by 15. Dynamic range and linearity are improved by using an active source degeneration topology (M3/4) [11].

3) *spike generation and attenuation*: An attenuation of postsynaptic spikes was introduced in [1], which reduces the pulse area via an exponential dependence on the interval between the last and the current spike [2]. If an action potential is represented by a rectangle, the pulse area can be varied by changing height or width. However, to exploit a maximal range for sub-threshold activity,  $V_{thr}$  is very high. Since the plasticity rule is based on a single state variable representing both the neurons' subthreshold behaviour as well as the output pulse, the height of a pulse would have to be in the limited range between  $V_{thr}$  and  $VDD$ . Thus, it is more reasonable to change the pulse area by adjusting the duration of the pulse. Core of the spike generation and adaptation circuitry is the capacitance  $C_{att}$ , which defines via an exponential decay of  $V_{Catt}$  the pulse duration adaptation. In steady state  $\overline{SPK}$  is high,  $SPK_{delayed}$  is low and  $C_{att}$  is charged up to  $V_{bias}$ . In this case  $RESET$  is low. If the membrane potential increases and reaches the threshold voltage  $V_{thr}$ , signal  $\overline{SPK}$  becomes low and stops all currents flowing out of  $C_{att}$ . Now the voltage across  $C_{att}$  increases linearly by charging with a constant

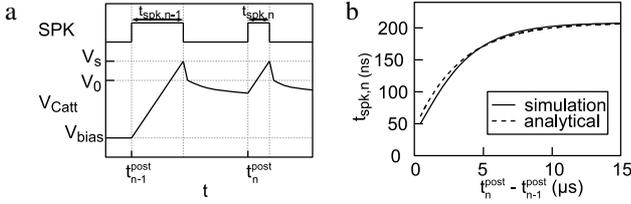


Fig. 4. (a) schematic time response diagram of  $V_{C_{att}}$  and  $SPK$  (b) pulse duration of the current spike as function of the interspike interval

current  $I_{spktime}$ . With reaching the threshold voltage  $V_s$  of the following inverter (M26, M27), signal  $RESET$  becomes high, stops all currents flowing into  $C_{att}$  and resets the membrane. The time to charge  $C_{att}$  determines the spike duration  $t_{spk,n}$ :

$$t_{spk,n} = C_{att} \frac{V_s - V_{C_{att}}(t_n^{post})}{I_{spktime}} \quad (2)$$

After  $\overline{SPK}$  changes back to high, another leakage circuit similar to Fig. 2b lets  $V_{C_{att}}$  decay back exponentially towards  $V_{bias}$  with a controllable time constant  $\tau_{refr}$ . If a second spike occurs shortly after the first one, the pulse duration is shortened since the time to charge  $C_{att}$  is shorter (see Fig. 4a). This adaptation mechanism is adjustable via the current  $I_{\alpha_{att}}$  which discharges  $C_{att}$  to  $V_0$  after a pulse for a fixed time period determined by the time difference between  $\overline{SPK}$  and  $SPK_{delayed}$ . Since this time is short compared to  $\tau_{refr}$ ,  $V_0$  can be used as new starting value for the exponential decay:

$$V_{C_{att}}(t) = V_{bias} + (V_0 - V_{bias})e^{-\frac{t - t_{n-1}^{post}}{\tau_{refr}}} \quad (3)$$

Inserting (3) in (2) results in an attenuation of  $t_{spk}$  similar to [1] (see Fig. 4b). To keep the attenuation from producing an insufficient pulse width for subsequent stages, the RESET signal is additionally gated by the state of the digital pulse registration following the neuron.

4) *PSC generation*: The exponentially decaying PSC current required by the learning rule is realized by employing another instance of the leakage circuit of Fig. 2b in a pulsed configuration. With each incoming (presynaptic) spike, its capacitance is set to a defined value below its resting potential, and the exponentially decreasing current generated by the leakage circuit charges the capacitance back to its resting potential. This is mirrored to the synapse circuit as PSC current. The basic PSC waveform generated by this circuit (Fig. 1a, upper curve) is quite similar to other PSC time course reconstructions in literature [9]. PSC and neuron will each occupy about  $1400\mu m^2$  in the chosen UMC 130nm process.

## B. Synapse

The circuit shown in Fig. 5 approximates the multiplication of the difference between membrane potential and voltage threshold  $\Theta_u$  with the PSC current as shown in Eq. 1. It is similar in realization to the learning rule of [12], with a current difference computation carried out in the differential pair MN5/6, while the multiplication is based on the proportionality of this current difference to the tail current. The

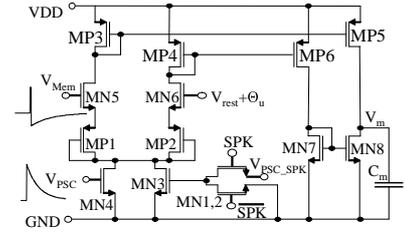


Fig. 5. Circuit carrying out the difference computation and multiplication of Eq. 1 based on the waveforms of Fig. 1a. Waveforms are generated by the circuits described in Sec. III-A

tail current of the differential pair is accordingly provided as the exponentially decaying PSC current via MN4. As in the neuron, source degeneration [11] is used in the form of MP1/2 to extend the linear range of the differential pair and thus the quality of the  $V_{Mem} - \Theta_u$  computation. Since the neuron circuit has a membrane resting potential at approximately half the supply voltage, the term  $\Theta_u$  of Eq. 1 has to be offset by this resting potential, arriving at  $(V_{rest} + \Theta_u)$  for the gate voltage of MN6. The current in both differential paths is mirrored out via MP3/4, while MN7/8 and MP5/6 compute the current difference which is used to charge the weight capacitance  $C_m$  to arrive at the weight-proportional voltage  $V_m$ . Since this constitutes a volatile weight storage, it is planned to extend this linearly by a secondary digital weight storage similar to [9], [7], which is incremented/decremented when the weight value at the capacitance is driven to one of its extremes.

As explained in Sec. II, the spike amplitude needs to be much higher than the hyperpolarization amplitude for a realistic STDP curve. To achieve equal areas, the relation between these amplitudes must be the inverse of the relation between spike duration (approx. 2ms biol.) and the membrane time constant of the refractoriness period (approx. 34ms biol.). Thus, the spike amplitude has to be 17 times higher than the hyperpolarization amplitude. Since the supply voltage range is used almost entirely for implementing the subthreshold neuron dynamics (including hyperpolarization), a simple scaling of the spike is not possible. Instead, the PSC current provided by the exponential PSC reconstruction is scaled by this amount and added to the ordinary PSC current  $I_{PSC}$  via the  $V_{PSC\_SPK}$  signal at MN3. A digital pulse from the neuron signals an action potential to the synapse at port SPK (see Fig. 2a), activating the increased PSC current only during action potentials. A second inverted signal  $\overline{SPK}$  pulls the gate of MN3 to ground after the action potential to reduce charge storage effects. This switching can be seen in the transients of  $I_{PSC}$  in Fig. 1a. The synapse has been realized with an area of approx.  $350\mu m^2$  in the chosen UMC 130nm process.

## IV. RESULTS

Basic STDP behaviour for the overall circuit (i.e. synapse, neuron, PSC reconstruction) is shown in Fig. 1b, also included is a sweep of the process corners for the transistors. As can be seen, the STDP curve [5] is robustly replicated, with especially the time constants almost invariant to process mismatch, while

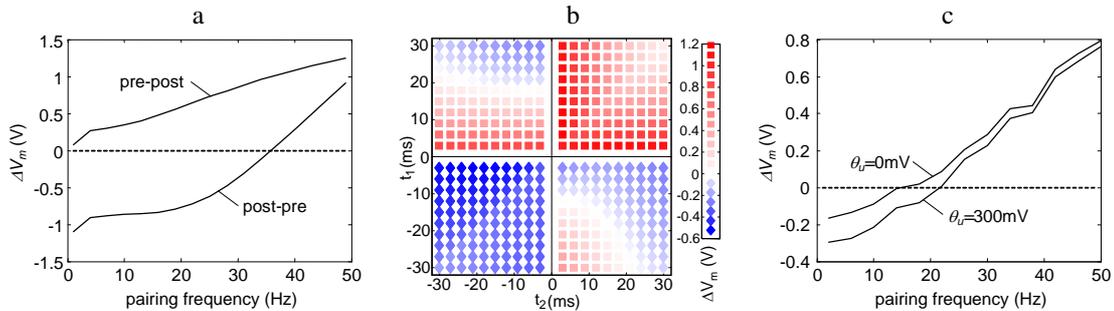


Fig. 6. Results for replicating various experimental paradigms using the circuit realization of our learning rule, with results plotted as the voltage difference at the weight capacitance before and after the experiment: (a) pairs of pre- and postsynaptic spikes at fixed time differences of 10ms in either pre-post or post-pre order, subject to a sweep of the pairing repetition frequency [3] (b) Triplets of spikes consisting of two pre- and one postsynaptic spike or the opposite, results plotted in similar conventions as in [2], weight increase denoted in red squares, decrease in blue diamonds (c) pairs of pre- and postsynaptic spikes with a timing difference drawn from a Gaussian distribution with  $\mu = 0ms$  and  $\sigma = 7ms$  [3]. To remove the resultant stochastic variability in the experimental results, the mean of 20 repetitions of this experiment was plotted. In order for the remaining variability of the curve not to mask the effect of different  $\Theta_U$  settings, an identical instance of the stochastic protocol was employed for both  $\Theta_U$ 's, leading to similar curve shapes.

the actual weight increase/decrease varies by about a factor of 4. However, this could easily be adjusted via parameter settings of the respective neuron and/or PSC generation. Also, since these curves represent the  $6\sigma$  ranges of the process, actual deviations across a single neuromorphic ASIC are expected to be significantly less.

As shown in Fig. 6a, the combination of spike pairings at varying time differences with a sweep of the repetition frequency of those pairings [3] is also replicated. The extension of spike pairings to the well-known triplet paradigm [2] can also be reproduced by the learning rule (Fig. 6b). A sample BCM-type protocol [3] is shown in Fig. 6c. Note especially the sliding of the frequency threshold (i.e. the point where the weight crosses from decrease to increase with spike frequency) when the voltage threshold is varied. Thus, this feature of the BCM theory [4] is also replicated using our circuit.

## V. CONCLUSION

Although no silicon implementation of the rule has yet been realized, the robustness of the general learning function can be seen from the corner simulation carried out for STDP learning in Fig. 1b. Since the learning rule is based on reusing waveform functionality which is already included in most neuromorphic systems [9], [6], circuit complexity is significantly reduced. Compared to [9], [8], [6], [13], no time constants (i.e. capacitances) are required for the STDP time windows, and the transistor count is reduced by a factor of 2-3. Our synapse is on par with the similar complexity-reduced approach of [7], while outperforming most other recent examples in terms of transistor and capacitor count. More importantly, even at this reduced complexity, our circuit outperforms all current synaptic learning implementations with respect to experiment reproduction, since these are only built to replicate STDP, which is incompatible with spike triplet results [2] and learning evoked by combinations of timing and rate [3].

## ACKNOWLEDGMENT

We thank W. Gerstner for fruitful discussions on plasticity. This work is supported by grant no. 15879 (FACETS), 6th

Framework Programme of the European Union. J. Partzsch has a graduate scholarship of the Konrad-Adenauer-Stiftung.

## REFERENCES

- [1] J. Partzsch, C. Mayr, and R. Schüffny, "BCM and membrane potential: Alternative ways to timing dependent plasticity," in *Proceedings of 15th International Conference on Neural Information Processing ICONIP08*, ser. LNCS, vol. 5506, 2008, pp. 137–144.
- [2] R. Froemke and Y. Dan, "Spike-timing-dependent synaptic modification induced by natural spike trains," *Nature*, vol. 416, pp. 433–438, 2002.
- [3] P. Sjöström, G. Turrigiano, and S. Nelson, "Rate, timing, and cooperativity jointly determine cortical synaptic plasticity," *Neuron*, vol. 32, pp. 1149–1164, 2001.
- [4] E. Bienenstock, L. Cooper, and P. Munro, "Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex," *Journal of Neuroscience*, vol. 2, no. 1, pp. 32–48, 1982.
- [5] G.-Q. Bi and M.-M. Poo, "Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type," *Journal of Neuroscience*, vol. 18, no. 24, pp. 10464–10472, 1998.
- [6] T. Koickal, A. Hamilton, S. Tan, J. Covington, J. Gardner, and T. Pearce, "Analog VLSI circuit implementation of an adaptive neuromorphic olfaction chip," *IEEE Transactions on Circuits and Systems-I: Regular Papers*, vol. 54, no. 1, pp. 60–73, 2007.
- [7] S. Mitra, S. Fusi, and G. Indiveri, "Real-time classification of complex patterns using spike-based learning in neuromorphic VLSI," *IEEE Trans. on Biomedical Circuits and Systems*, vol. 32-42, no. 1, 2009.
- [8] K. Saeki, R. Shimizu, and Y. Sekine, "Pulse-type hardware neural network with two time windows in STDP," in *Proceedings of 15th International Conference on Neural Information Processing ICONIP08*, ser. LNCS, vol. 5507, 2008, pp. 877–884.
- [9] J. Schemmel, A. Grübl, K. Meier, and E. Mueller, "Implementing synaptic plasticity in a VLSI spiking neural network model," in *International Joint Conference on Neural Networks IJCNN '06.*, 2006, pp. 1–6.
- [10] P. Kinget, M. Steyaert, and J. Spiegel, "Full analog CMOS integration of very large time constants for synaptic transfer in neural networks," *Analog Int. Circuits and Signal Processing*, vol. 2, pp. 281–295, 1992.
- [11] E. Sanchez-Sinencio and J. Silva-Martinez, "CMOS transconductance amplifiers, architectures and active filters: a tutorial," *IEEE Proceedings on Circuits, Devices and Systems*, vol. 147, no. 1, pp. 3–12, Feb 2000.
- [12] J. Schreiter, U. Ramacher, A. Heitmann, D. Matolin, and R. Schüffny, "Analog implementation for networks of integrate-and-fire neurons with adaptive local connectivity," in *Proc. of the 2002 12th IEEE Workshop on Neural Networks for Signal Processing*, 2002, pp. 657–666.
- [13] A. Bofill-i Petit and A. Murray, "Synchrony detection and amplification by silicon neurons with STDP synapses," *IEEE Transactions on Neural Networks*, vol. 15, no. 5, pp. 1296–1304, 2004.