A Switched-Capacitor Implementation of Short-Term Synaptic Dynamics

Marko Noack, Christian Mayr, Johannes Partzsch, Michael Schultz and René Schüffny Endowed Chair for Highly Parallel VLSI Systems and Neuromorphic Circuits Institute of Circuits and Systems, Technische Universität Dresden, Germany Email: marko.noack@tu-dresden.de

Abstract—In this paper we present a novel switched-capacitor implementation of short-term synaptic dynamics with simultaneous depression and facilitation. The developed circuit model is a modified version of a model of neurotransmitter release derived from biological measurements. Despite the simplicity of the circuit the rich dynamics of the original model can be delivered. By completely relying on SC techniques for all calculations, our circuit is significantly less sensitive to process variations and easier to calibrate than commonly employed subthreshold circuits. The circuit makes use of a technique for minimizing leakage effects allowing for real-time operation with time constants up to several seconds. Functionality and robustness of the circuit are verified by simulations and comparisons to the original model.

Index Terms—Switched-capacitor, short-term synaptic dynamics, depression, facilitation, silicon synapse.

I. INTRODUCTION

Biological synapses employ a range of short-time adaptation mechanisms in their stimulus transmission. This so-called short-term plasticity has been identified as a crucial constituent of neural information processing, allowing for temporal filtering [1] and pattern classification in attractor networks [2]. The quantal model of neurotransmitter release introduced in [3] is a well-established model of these synaptic dynamics that has been directly derived from biological evidence. Furthermore, it has been thoroughly analyzed in terms of information processing [4], [5].

Despite the functional importance of biologically realistic short-term dynamics, there are only few neuromorphic circuits implementing simultaneously acting short-term depressing and facilitating mechanisms as described in [3]. The circuits presented in [6] produce either depression or facilitation, while the more complex circuit presented in [7] is at least capable of switching between the two. A combination of facilitation and depression mechanisms is shown in [8], but this implementation is rather aimed at hardware efficiency than biological relevance. In [9], we prepared the model in [3] for circuit design, where the depression part is implemented with OTA-C circuits.

When it comes to neuromorphic implementations of neural dynamics, the main challenge is replication of large time constants in the order of milliseconds up to several seconds. Most circuits make use of transistors working in the subthreshold region [6], [8], exploiting the low drain currents in this regime to discharge capacitors over a long period of time. A serious disadvantage resulting from this approach is the sensitivity to threshold voltage mismatch, resulting in large variations on the fabricated chip [10]. As an alternative, the time base of the circuits may be accelerated for increased current amplitudes [7], but this prevents interfacing to real-time operating sensors and biological substrates.

In this paper, we aim at tackling both the modeling and circuit implementation problems introduced above. We take the full quantal model of [3] and adapt it for an optimized circuit implementation, deriving equations for mapping parameters to the adapted model. By this modification, we retain the full power of the original model while greatly reducing circuit effort. In contrast to the OTA-C based circuit presented in [9], we employ switched-capacitor (SC) circuit techniques for realizing large time constants, whose behaviour is mainly determined by capacitance ratios instead of absolute transistor parameters, which greatly decreases their mismatch sensitivity. This approach has already been successfully applied to neuromorphic neuron implementations [11], [12]. We carried out a sample design in 180 nm CMOS, relying on SC techniques for all calculations required by the model to maximally reduce reliance on analog performance. The main limit on the achievable time constants are the leakage currents through the employed switches, which can be minimized with simple circuit extensions to reach values in the order of seconds. With all these characteristics, our proposed circuit can be easily ported to sub-100 nm technologies, taking full advantage of their higher integration density, while still being easily controllable by digital logic.

In section II we present our modified model of shortterm synaptic dynamics, obviating the need for expensive multiplication circuits. The implemented design is presented in section III followed by simulation results in section IV.

II. MODEL OF SHORT-TERM SYNAPTIC DYNAMICS

A. Original Quantal Model

The quantal model of neurotransmitter release introduced in [3] describes the amplitude of excitatory postsynaptic currents (PSC) as a function of the timing of presynaptic spikes and their history. Equations (1) – (3) show the iterative description of the model, where n is the number of the spike and Δt_n is

the time between *n*-th and (n + 1)-th spike.

$$u_{n+1} = u_n \cdot (1 - U) \cdot e^{-\frac{\Delta t_n}{\tau_{facil}}} + U \tag{1}$$

$$R_{n+1} = R_n \cdot (1 - u_n) \cdot e^{-\frac{\Delta t_n}{\tau_{rec}}} + 1 - e^{-\frac{\Delta t_n}{\tau_{rec}}}$$
(2)

$$PSC_n = A \cdot R_n \cdot u_n \tag{3}$$

Two mechanisms acting simultaneously are modulating the PSC amplitude. An amplifying mechanism called facilitation is modeled by variable u, which increases by a certain amount at every presynaptic spike and decays back with time constant τ_{facil} to its resting state U. At high spiking frequencies u_n saturates at 1. R, which describes a depression mechanism is decreased at presynaptic spikes by a certain amount, which is influenced by u. The depression recovers to its resting state 1 with time constant τ_{rec} . At high frequencies R_n saturates at 0. The amplitude of the n-th PSC is calculated by a multiplication of the two variables u_n and R_n and a scaling factor A, which represents the absolute synaptic efficacy.

B. Proposed Model

In order to develop a neuromorphic circuit which is capable of reproducing the quantal model [3] the approach presented in [9] could be taken. Therefore, exponentially decaying voltage curves have to be generated which are updated and triggered at the occurrence of presynaptic spikes. The major drawback of this approach concerning a switched-capacitor implementation is the need for a wide-range voltage multiplier for calculating the product of u_n and R_n . Existing multipliers are rather complex, very area consuming [13] or need large operational amplifiers driving resistive loads [14]. In contrast, our proposed model is capable of approximately reproducing the original quantal model without any multiplier circuit and with a minimum effort on analog circuitry in general.

The iterative description of the proposed model is shown in eqs. (4) - (6).

$$\tilde{u}_{n+1} = \tilde{u}_n \cdot (1 - \tilde{U}) \cdot e^{-\frac{\omega n}{\tilde{\tau}_{facil}}} + \tilde{U}$$
(4)

$$\tilde{R}_{n+1} = ((1-\alpha) \cdot \tilde{R}_n + \alpha \cdot \tilde{u}_n) \cdot e^{-\frac{\Delta t_n}{\tilde{\tau}_{rec}}}$$
(5)

$$P\tilde{S}C_n = \tilde{A} \cdot (\tilde{u}_n - \tilde{R}_n) \tag{6}$$

From Eq. (1) it is obvious that \tilde{u} is equivalent to u. Only a parameter mapping has to be performed from U to \tilde{U} . In contrast to the original model, where R_{n+1} depends on the product of u_n and R_n , \tilde{R}_{n+1} only depends on a weighted sum of \tilde{u} and \tilde{R} , which simplifies the circuit enormously. Factor α determines the strength of the depression. Here \tilde{R} is inverted to R meaning the 'recovered' state is at $\tilde{R} = 0$. This implies that large values of \tilde{u} lead to a stronger increase of \tilde{R} which has the same effect as in Eq. (2) where large values of u lead to a stronger decrease of R. Also it can be seen that \tilde{R} never exceeds \tilde{u} so that the difference $(\tilde{u} - \tilde{R})$ is always positive. The amplitude of the postsynaptic current $P\tilde{S}C$ is determined by this difference scaled by \tilde{A} . The time course of the PSC is modeled by an exponential decay with time constant $\tilde{\tau}_{PSC}$.



Fig. 1. Normalized steady-state amplitude of PSC at constant presynaptic pulse rates for the quantal model and the proposed model. The peak frequency is marked with Θ . Used parameters are U = 0.03, $\tau_{rec} = 130$ ms, $\tau_{facil} = 530$ ms, $\tilde{U} = 0.055 \alpha = 0.44$, $\tilde{\tau}_{rec} = 87$ ms, $\tilde{\tau}_{facil} = 864$ ms.

C. Parameter Mapping

A characteristic property, which allows for comparing both models is the steady-state behavior at constant presynaptic firing rates. If both mechanisms, facilitation and depression, are acting simultaneously the steady-state amplitude PSC_{st} of the PSC shows a bell-shaped curve (see Fig. 1). In order to fit our model against the quantal model, parameters can be adjusted, so that the peaks of the curves are matching. The peak frequency Θ of the quantal model is determined by

$$\Theta \approx \frac{1}{\sqrt{U \cdot \tau_{facil} \cdot \tau_{rec}}} \,. \tag{7}$$

In the proposed model the peak frequency $\hat{\Theta}$ is

$$\tilde{\Theta} \approx \frac{1}{\sqrt{\tilde{U} \cdot \alpha \cdot \tilde{\tau}_{facil} \cdot \tilde{\tau}_{rec}}} \,. \tag{8}$$

The peak amplitude, normalized to the first PSC amplitude is

$$\frac{PSC_{st}(\Theta)}{PSC_1} \approx \sqrt{\frac{1}{4U} \cdot \frac{\tau_{facil}}{\tau_{rec}}} \tag{9}$$

for the quantal model and

$$\frac{P\hat{S}C_{st}(\hat{\Theta})}{P\tilde{S}C_{1}} \approx \frac{1}{\sqrt{4 \cdot \tilde{U} \cdot \alpha \cdot \frac{\bar{\tau}_{rec}}{\bar{\tau}_{facil}}} + \alpha \frac{\bar{\tau}_{rec}}{\bar{\tau}_{facil}}}$$
(10)

for the proposed one. With equations (7) - (10) the time constants of the proposed model can be derived to

$$\tilde{\tau}_{facil} = \tau_{facil} \cdot \left(1 + \frac{\alpha}{2} \sqrt{\frac{\tau_{rec}}{U \cdot \tau_{facil}}} \right) \tag{11}$$

$$\tilde{\tau}_{rec} = \tau_{rec} \cdot \frac{1}{\frac{\alpha^2 \cdot \bar{U}}{2U} \cdot \sqrt{\frac{\tau_{rec}}{U \cdot \tau_{facil}} + \alpha \cdot \frac{\bar{U}}{U}}}$$
(12)

The remaining parameters \tilde{U} and α , whose ranges are between 0 and 1, are used for better curve fitting. The absolute PSC amplitude has to be scaled with $\tilde{A} = A \cdot \frac{U}{U}$.

III. SWITCHED-CAPACITOR IMPLEMENTATION

The switched-capacitor circuit implementing the proposed model is depicted in Fig. 2(a). It consists of three analog blocks for facilitation, depression and PSC generation and a state machine (FSM) for controlling the switches. The values \tilde{u} and \tilde{R} are stored on capacitors C_u and C_R respectively. C_{PSC} stores the trace of the PSC.

A. Circuit Operation

As intended by equations (4) and (5) an exponential decay of \tilde{u} and \tilde{R} as well as of the PSC curve has to be performed. This is implemented by generic switched capacitor resistor emulations with switches S_{u1} and S_{u2} and capacitor C_{Ru} for the facilitation block, S_{R1} , S_{R2} and C_{RR} for depression and S_1, S_2 and C_{RPSC} for the PSC generation circuit, which discharge C_u , C_R and C_{PSC} towards ground. The corresponding time constants of decay $\tilde{\tau}_{facil}$, $\tilde{\tau}_{rec}$ and $\tilde{\tau}_{PSC}$ are determined by the switching frequency and can be adjusted externally. Therefore signals DECAY_u, DECAY_R and DECAY_PSC are provided, which send pulses to the state machine in certain intervals (see Fig. 2(b)). If the FSM receives one of these signals two non-overlapping switch phases are triggered. In the case of DECAY_u, first S_{u1} is switched on, which completely discharges C_{Ru} . After S_{u1} is switched off, S_{u2} is switched on leading to a charge equalization on C_{Ru} and C_u . The voltage V_u over C_u drops from an initial voltage V_{u0} to $V_{u0} \cdot \frac{C_u}{C_u + C_{Ru}}$. After *n* switching events a value of $V_{u0} \cdot \left(\frac{C_u}{C_u + C_{R_u}}\right)^n$ is reached. With $\left(\frac{C_u}{C_u + C_{R_u}}\right)^n = e^{-1}$ the number of switching events needed for one $\tilde{\tau}_{facil}$ period can be derived, so the pulse frequency of DECAY_u is determined by $f_{DECAY} = -(\tilde{\tau}_{facil} \cdot \ln \frac{C_u}{C_u + C_{Ru}})^{-1}$. Same applies for depression with and PSC generation circuits. In our design we chose $C_u/C_{Ru} = 1.12 \text{ pF}/32 \text{ fF} = 35/1$ and $C_R/C_{RR} = C_{PSC}/C_{RPSC} = 480 \text{ fF}/32 \text{ fF} = 15/1$. Thus, for $\tilde{\tau}_{facil} = 500 \text{ ms}$ signal DECAY_u is sent with a frequency of 71 Hz.

If a presynaptic spike occurs (signal PRESPK is sent to the FSM), the values \tilde{u} , \tilde{R} and $P\tilde{S}C$ have to be updated. By definition the current value of \tilde{u} is represented by V_u after the update and the current value of \tilde{R} is represented by V_R before the update. This ensures that \tilde{u} decays towards \tilde{U} and \tilde{R} towards 0. Thus, V_u is updated first, then the PSC amplitude is calculated and finally V_R is updated.

In order to calculate Eq. (4) the FSM alternately switches S_{u3} and S_{u2} (see Fig. 2(b)), which increases V_u exponentially towards V_A and thereby implements a constant update factor $(1 - \tilde{U})$. The number of switching events is provided by the 6 bit wide signal UTIL. The relationship between \tilde{U} and UTIL depends on the capacitance ratio and can be expressed as

$$\text{UTIL} = \left\lfloor \frac{\log(1 - \tilde{U})}{\log(\frac{C_u}{C_u + C_{R_u}})} \right\rfloor . \tag{13}$$

After charging of C_u , switches S_4 and S_5 are closed and the voltage difference $(V_u - V_R)$ is stored on C_{PSC} , which corresponds to Eq. (6). Finally V_R is charged up towards V_u . As can be seen in Eq. (5) the strength of the charging is determined by α which is performed by alternately switching S_{R2} and S_{R3} similar to the facilitation mechanism. The



Fig. 2. (a) SC implementation of the proposed model of synaptic dynamics. Capacitors C_u and C_R store the values \tilde{u} and \tilde{R} respectively. C_{PSC} stores the trace of the PSC. (b) Diagram of switch signals for decay of V_u and an update at a presynaptic spike. Decay of V_R and V_{PSC} follow the same principle as for V_u .

number of switching events can be calculated by

$$ALPHA = \left\lfloor \frac{\log(1-\alpha)}{\log(\frac{C_R}{C_R+C_{RR}})} \right\rfloor .$$
(14)

Except for the state where S_4 and S_5 are closed, S_3 is always closed to get a defined voltage V_{PSC} by grounding capacitor C_{PSC} .

The voltage V_A controls the absolute synaptic efficacy Aand represents the upper limit for V_u and V_R . If V_A stays one threshold voltage below supply voltage, C_u , C_{Ru} , C_R and C_{RR} can be implemented as PMOS capacitors which then always operate in the strong inversion region. Also all switches can be implemented with single NMOS transistors and the buffer amplifiers do not need to provide rail-to-rail operation. As can be seen in Fig. 3 the amplifiers are implemented in a single ended single stage folded cascode topology to allow low supply voltage and a sufficiently high gain to reduce the output voltage offset. Since the state machine operates at a low clock frequency of about 160 kHz slew rate is not a major issue. This allows a compact design with a low power consumption.

B. Leakage Optimization

Due to the large time constants in our implementation, leakage effects at high impedance nodes have to be considered. The main source for leakage effects at C_u and C_R are the



PRESPK<0> PRESPK<1> PRESPK<N> ¥ ¥ ¥ DECAY GENERATOR SC CIRCUI PSC CIRCUI SC CIRCUI DECAY FSM FSM FSM DECAY . . . UTIL SC SC SC CIRCUIT CIRCUIT CIRCUIT AI PHA M NΝ NΝ PSC PSC PSC

Fig. 3. Amplifier in single ended single stage folded cascode topology used as buffer in Fig. 2.

switching transistors S_{R2} and S_{u2} in off-state. As shown in [15] the main leakage effects that appear at a turned off transistor are junction leakage, channel leakage and subthreshold leakage. They scale with the potentials between the transistor's terminals. Therefore, in the interval when neither a decay nor an update is performed, the voltages across C_u and C_R are fed back via switches S_{u4} and S_{R4} to C_{Ru} and C_{RR} respectively. So V_{DS} of the transistors implementing S_{u2} and S_{R2} is defined by gain and offset voltage of the buffers, which is close to 0. Consequently, subthreshold leakage of S_{u2} and S_{R2} is strongly reduced. As bulk and gate of the transistors are at 0 V while they are turned off, the leakage currents caused by reverse biased PN junctions at times of high V_{GB} and V_{DB} cannot be influenced. However, these are much smaller than the subthreshold leakage in the employed technology [15].

C. Scalability

The circuit is designed for a special architecture of neuromorphic synapse arrays presented in [16], where the circuits generating the PSCs are separated from the individual synapses. Thus, several synapses receive the same presynaptic input, which is usually the case in neural networks. This leads to a smaller number of PSC circuits and smaller synapses, significantly improving the scalability of the overall system. As depicted in Fig. 4 the PSC circuits with the same parameterization can share one decay generation circuit to keep them as small as possible.

IV. RESULTS

Figure 5 shows the simulation results of PSC amplitudes of the original quantal model [3], the proposed model and a circuit simulation for a sequence of constant pulse rates of 15 Hz, 30 Hz, 80 Hz and 15 Hz with instantaneous transitions. All curves are normalized to the first PSC amplitude. As can be seen the proposed model well approximates the quantal model in terms of steady state amplitudes and the behavior at spike frequency transitions. At frequencies up to 30 Hz the circuit simulation results match the model results. At higher frequencies the PSC amplitudes underlie deterministic

Fig. 4. System of several PSC circuits receiving pulse signals from the decay generation circuit. The PSC circuits consist of the SC circuit shown in Fig. 2(a) and a finite state machine for controlling the switches. PSCs are triggered by the input signal PRESPK.

jumps which are caused by time discretization of the decay mechanism of V_u and V_R , since at these presynaptic spike frequencies the pulse frequency of the signal DECAY_u is in the same order of magnitude. In order to suppress the jumps and to get a finer discretization the ratio of capacitances C_{Ru}/C_u can be made smaller which would raise two issues: first, the counter in the FSM, which controls the update of V_u has to be larger since more switching phases are needed to charge the capacitor C_u and second, the matching between these capacitors gets worse. But since future technologies tend to smaller transistor sizes, the digital building blocks carry less weight and could be more complex.

An example spike train, which shows both facilitation and depression, is depicted in Fig. 6. The voltage V_{PSC} has been sampled on valid data only, meaning the state where S_4 and S_5 are closed is excluded because C_{PSC} is not grounded in this moment. In order to verify the robustness of the circuit concerning device mismatch and process variations a Monte Carlo simulation with 100 runs has been performed. The amplitude deviations, which are similar to those in subthreshold circuits [17], are mainly caused by the amplifiers whereas the time constants can be robustly reproduced due to the well matched capacitors and the externally provided switching frequency. Note that due to the simplification of the model by replacing the multiplication in Eq. 2 with the difference in Eq. 5 some cases can be covered less accurately. Thus, the response to a pulse train when $\tau_{facil} = \tau_{rec}$ is slightly smoother than shown in [18]. However, most cases can be replicated well.

For the design in a 180 nm CMOS technology the required chip area is approximately 2400 μ m² including the FSM, which consumes about 75 % of the entire area. When moving towards smaller technologies shrinking of analog components like capacitors and switches is limited, whereas the digital parts scale fully. Thus, in a 65 nm technology, the area could be shrunk to 850 μ m², which is slightly smaller than the depressing synapse presented in [17].



Fig. 5. Simulation results of the quantal model [3], the proposed model and the switched-capacitor circuit with a sequence of instantaneous spike frequency transitions. Parameters are the same as in Fig. 1.



Fig. 6. Simulation results of V_{PSC} for an example spike train including results of a Monte Carlo simulation with 100 runs. The grey lines represent the largest deviations from the nominal simulation. Parameters are U = 0.081 $\alpha = 0.36, \, \tilde{\tau}_{rec} = 500 \, \mathrm{ms}, \, \tilde{\tau}_{facil} = 100 \, \mathrm{ms}.$

V. CONCLUSION

In this paper we presented an SC implementation of synaptic dynamics derived from a biologically realistic model. Our model adaptation reduces the implementation complexity to simple subtractions, while keeping the computational richness of the original quantal model. The circuit allows for real-time operation and is robust in terms of process variation on the chip. Especially the time constants, which govern dynamic behavior of a neural network, exhibit very little spread. Since digital building blocks dominate the overall circuit area, our design will scale well with advanced CMOS technologies.

ACKNOWLEDGMENT

This research has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 269459 (CORONET)

REFERENCES

- [1] L. Grande and W. Spain, "Synaptic depression as a timing device," Physiol, vol. 20, pp. 201-210, 2005.
- [2] J. Mejias and J. Torres, "Maximum memory capacity on neural networks with short-term synaptic depression and facilitation," Neur Comput, vol. 21, no. 3, pp. 851-871, 2009.
- H. Markram, Y. Wang, and M. Tsodyks, "Differential signaling via the [3] same axon of neocortical pyramidal neurons," PNAS, vol. 95, no. 9, pp. 5323-5328, 1998
- L. Abbott, G. Wade, and W. Regehr, "Synaptic computation," Nature, [4] vol. 431, no. 14, pp. 796-803, 2004.
- V. Matveev and X.-J. Wang, "Differential short-term synaptic plasticity and transmission of complex spike trains: to depress or to facilitate?' Cereb Cortex, vol. 10, pp. 1143–1153, 2000. [6] S.-C. Liu, "Analog VLSI circuits for short-term dynamic synapses,"
- EURASIP J. Appl. Signal Process., vol. 2003, pp. 620-628, 2003.
- J. Schemmel, D. Brüderle, K. Meier, and B. Ostendorf, "Modeling synaptic plasticity within networks of highly accelerated IF neurons, in ISCAS, may 2007, pp. 3367 -3370.
- [8] E. Chicca, G. Indiveri, and R. Douglas, "An adaptive silicon synapse," in Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on, vol. 1, may 2003, pp. I-81 - I-84 vol.1.
- M. Noack, C. Mayr, J. Partzsch, and R. Schüffny, "Synapse dynamics in CMOS derived from a model of neurotransmitter release," in 20th ECCTD 2011, 2011, pp. 197-200.
- [10] J. Pineda de Gyvez and H. Tuinhout, "Threshold voltage mismatch and intra-die leakage current in digital CMOS circuits," Solid-State Circuits, IEEE Journal of, vol. 39, no. 1, pp. 157 - 168, jan. 2004.
- [11] J. Elias and D. Northmore, "Switched-capacitor neuromorphs with wide-range variable dynamics," *Neural Networks, IEEE Transactions on*, vol. 6, no. 6, pp. 1542 -1548, nov 1995.
- [12] F. Folowosele, R. Etienne-Cummings, and T. Hamilton, "A CMOS switched capacitor implementation of the mihalas-niebur neuron," in BioCAS 2009, nov. 2009, pp. 105-108.
- [13] Z. Hong and H. Melchior, "Four-quadrant CMOS analogue multiplier," Electronics Letters, vol. 20, no. 24, pp. 1015 -1016, 22 1984.
- [14] N. Khachab and M. Ismail, "A nonlinear CMOS analog cell for VLSI signal and information processing," Solid-State Circuits, IEEE Journal of, vol. 26, no. 11, pp. 1689 -1699, nov 1991.
- [15] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, "Leakage current mechanisms and leakage reduction techniques in deepsubmicrometer CMOS circuits," IEEE Transactions on Electron Devices, pp. 1393-1400, 2000.
- [16] M. Noack, J. Partzsch, C. Mayr, S. Henker, and R. Schuffny, "Biologyderived synaptic dynamics and optimized system architecture for neuromorphic hardware," in MIXDES, 2010, pp. 219-224.
- [17] C. Bartolozzi and G. Indiveri, "Synaptic dynamics in analog VLSI," Neural Computation, vol. 19, no. 10, pp. 2581-2603, 2007.
- [18] Y. Wang, H. Markram, P. H. Goodman, T. K. Berger, J. Ma, and P. S. Goldman-Rakic, "Heterogeneity in the pyramidal network of the medial prefrontal cortex," Nature Neuroscience, vol. 9, no. 4, pp. 534-542, Mar. 2006