ARTICLE Communicated by Christal Gordon

Synaptic Dynamics in Analog VLSI

Chiara Bartolozzi
chiara@ini.phys.ethz.ch
Giacomo Indiveri
giacomo@ini.phys.ethz.ch
Institute for Neuroinformatics, UNI-ETH Zürich, Zürich, Switzerland

Synapses are crucial elements for computation and information transfer in both real and artificial neural systems. Recent experimental findings and theoretical models of pulse-based neural networks suggest that synaptic dynamics can play a crucial role for learning neural codes and encoding spatiotemporal spike patterns. Within the context of hardware implementations of pulse-based neural networks, several analog VLSI circuits modeling synaptic functionality have been proposed. We present an overview of previously proposed circuits and describe a novel analog VLSI synaptic circuit suitable for integration in large VLSI spike-based neural systems. The circuit proposed is based on a computational model that fits the real postsynaptic currents with exponentials. We present experimental data showing how the circuit exhibits realistic dynamics and show how it can be connected to additional modules for implementing a wide range of synaptic properties.

1 Introduction

Synapses are highly specialized structures that, by means of complex chemical reactions, allow neurons to transmit signals to other neurons. When an action potential generated by a neuron reaches a presynaptic terminal, a cascade of events leads to the release of neurotransmitters that give rise to a flow of ionic currents into or out of the postsynaptic neuron's membrane. These excitatory or inhibitory postsynaptic currents (EPSC or IPSC, respectively) have temporal dynamics with a characteristic time course that can last up to several hundreds of milliseconds (Koch, 1999).

In computational models of neural systems, the temporal dynamics of synaptic currents have often been neglected. In models that represent the information with mean firing rates, synaptic transmission is typically modeled as an instantaneous multiplier operator (Hertz, Krogh, & Palmer, 1991). Similarly, in pulse-based neural models, where the precise timing of spikes and the dynamics of the neuron's transfer function play an important role, synaptic currents are often reduced to simple instantaneous charge impulses. Also in VLSI implementations of neural systems, silicon

synapses have often been reduced to simple multiplier circuits (Borgstrom, Ismail, & Bibyk, 1990; Satyanarayana, Tsividis, & Graf, 1992) or constant current sources activated only for the duration of the presynaptic input pulse (Mead, 1989; Fusi, Annunziato, Badoni, Salamon, & Amit, 2000; Chicca, Badoni, et al., 2003).

Within the context of pulse-based neural networks, modeling the detailed dynamics of postsynaptic currents can be a crucial step for learning neural codes and encoding spatiotemporal patterns of spikes. Leaky integrate-and-fire (I&F) neurons can distinguish between different temporal input spike patterns only if the synapses stimulated by the input spike patterns exhibit dynamics with time constants comparable to the time constant of the neuron's membrane potential (Gütig & Sompolinsky, 2006).

Modeling the temporal dynamics of each synapse in a network of I&F neurons can be onerous in terms of CPU usage for software simulations and in terms of silicon real estate for dedicated VLSI implementations. A compromise between highly detailed models of synaptic dynamics and no dynamics at all is to use computationally efficient models that account for the basic properties of synaptic transmission. A very efficient model that reproduces the macroscopic properties of synaptic transmission and accounts for the linear summation property of postsynaptic currents is the one based on pure exponentials proposed by Destexhe, Mainen, and Sejnowski (1998). Here we propose a novel VLSI synaptic circuit, the diff-pair integrator (DPI), that implements the model proposed in Destexhe et al. (1998) as a log-domain linear temporal filter and supports a wide range of synaptic properties, ranging from short-term depression to conductance-based EPSC generation.

The design of the DPI synapse is inspired by a series of similar circuits proposed in the literature that collectively share many of the advantages of our solution but individually lack one or more of the features of our design. In the next section, we present an overview of previously proposed synaptic circuits and describe the DPI synapse pointing out the advantages that the DPI offers over each of them. In section 3 we present experimental data from a VLSI chip showing the properties of the circuit in response to a single pulse and to sequences of spikes. In section 4 we show how the DPI is compatible with additional circuits used to implement various types of synaptic dynamics, and in section 5, we discuss possible uses of the DPI circuit in massively parallel networks of I&F neurons implemented on single or multichip neuromorphic systems.

2 Synaptic VLSI Circuits

Synaptic circuits translate presynaptic voltage pulses into postsynaptic currents injected in the membrane of their target neuron, with a gain typically referred to as the *synaptic weight*. The function of translating "fast" presynaptic pulses into long-lasting postsynaptic currents, with elaborate

temporal dynamics, can be efficiently mapped onto silicon using subthreshold (or weak-inversion) analog VLSI (aVLSI) circuits (Liu et al., 2002). In typical VLSI neural network architectures, the currents generated by multiple synapses are integrated by a single postsynaptic neuron circuit. The neuron circuit carries out a weighted sum of the input signals, produces postsynaptic potentials, and eventually generates output spikes that are typically transmitted to synaptic circuits in further processing stages. A common neuron model used in VLSI spike-based neural networks is the *point neuron*. With this model, the spatial position of the synaptic circuits connected to the neuron is not relevant, and the currents produced by the synapses are summed linearly into the single neuron's membrane capacitance node. Alternatively, synaptic circuits (including the one presented in this article) can be integrated in multicompartmental models of neurons, and the neuron's dendrite, comprising the spatial arrangement of VLSI synapses connected to the neuron, implements the spatial summation of synaptic currents (Northmore & Elias, 1998; Arthur & Boahen, 2004).

Regardless of the neuron model used, one of the main requirements for synaptic circuits in large VLSI neural networks is compactness: the less silicon area is used, the more synapses can be integrated on the chip. On the other hand, implementing synaptic integrator circuits with linear response properties and time constants of the order of tens of milliseconds can require substantial silicon area. Therefore, designing VLSI synaptic circuits that are compact and linear and model relevant functional properties of biological synapses is a challenging task still being actively pursued. Several subthreshold synaptic circuit designs have been proposed (Mead, 1989; Lazzaro, 1994; Boahen, 1998; Fusi et al., 2000; Chicca, Indiveri, & Douglas, 2003; Shi & Horiuchi, 2004a; Gordon, Farquhar, & Hasler, 2004; Hynna & Boahen, 2006) covering a range of trade-offs between functionality and complexity of temporal dynamics versus circuit and layout size. Some of the circuits proposed require floating-gate devices (Gordon et al., 2004) or restrict the signals used to a very limited dynamic range (Hynna & Boahen, 2006) to reproduce in great detail the physics of biological synaptic channels. Here we focus on the synaptic circuits that implement kinetic models of synaptic transmission functionally equivalent to the one implemented by the DPI, which can be directly integrated into large arrays of address-event-based neural networks (Lazzaro, 1994; Boahen, 1998).

2.1 Pulsed Current-Source Synapse. The pulsed current-source synapse, originally proposed by Mead (1989) in the late 1980s, was one of the first synaptic circuits implemented using transistors operated in the subthreshold domain. The circuit schematics are shown in Figure 1 (left); it consists of a voltage-controlled current source activated by an active-low input spike. In VLSI pulsed neural networks, input spikes are typically brief digital voltage pulses that last at most a few microseconds. The output of this circuit is a pulsed current I_{syn} that lasts as long as the input spike.

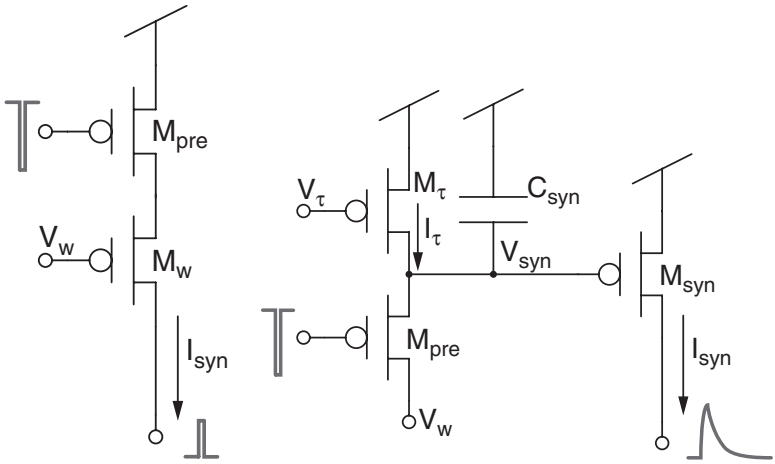

Figure 1: (Left) Pulsed current-source synaptic circuit. (Right) Reset-and-discharge synapse.

Assuming that the output p-FET M_w is saturated (i.e., that its V_{ds} is greater than $4U_T$), the current I_{syn} can be expressed as

$$I_{syn} = I_0 e^{-\frac{\kappa}{U_T}(V_w - V_{dd})}, \tag{2.1}$$

where V_{dd} is the power supply voltage, I_0 the leakage current, κ the subthreshold slope factor, and U_T the thermal voltage (Liu et al., 2002).

This circuit is extremely compact but does not integrate input spikes into continuous output currents. Whenever a presynaptic spike reaches M_{pre}, the postsynaptic membrane potential undergoes a step increase proportional to I_{syn}. As integration happens only at the level of the postsynaptic I&F neuron, input spike trains with the same mean rates but different spike timing distributions cannot be distinguished. However, given its simplicity and compactness, this circuit has been used in a wide variety of VLSI implementations of pulse-based neural networks that use mean firing rates as the neural code (Murray, 1998; Fusi et al., 2000; Chicca, Badoni, et al., 2003).

2.2 Reset-and-Discharge Synapse. In the early 1990s, Lazzaro (1994) proposed a synaptic circuit where the duration of the output EPSC $I_{syn}(t)$ could be extended with respect to the input voltage pulse by means of a tunable exponential decay (see also Shi & Horiuchi, 2004b, for a recent application example). This circuit, shown in Figure 1 (right), comprises three p-FET transistors and one capacitor; the p-FET M_{pre} is used as a digital switch that is turned on by the synapse's input spikes; the p-FET M_τ is

operated in subthreshold and is used as a constant current source to linearly charge the capacitor C_{syn}; the output p-FET M_{syn} is used to generate an EPSC that is exponentially dependent on the V_{syn} node (assuming subthreshold operation and saturation):

$$I_{syn}(t) = I_0 e^{-\frac{\kappa}{U_T}(V_{syn}(t)-V_{dd})}. \tag{2.2}$$

At the onset of each presynaptic pulse, the node V_{syn} is (re)set to the bias V_w. When the input pulse ends, the p-FET M_{pre} is switched off, and the node V_{syn} is linearly driven back to V_{dd}, at a rate set by I_τ / C_{syn}. For subthreshold values of $(V_{dd} - V_w)$, the EPSC generated by an input spike is therefore

$$I_{syn} = I_{w0} e^{-\frac{t}{\tau}}, \tag{2.3}$$

where $I_{w0} = I_0 e^{-\frac{\kappa}{U_T}(V_w - V_{dd})}$ and $\tau = \frac{\kappa I_\tau}{U_T C_{syn}}$.

In general, given a generic spike sequence on n spikes,

$$\rho(t) = \sum_i^n \delta(t - t_i), \tag{2.4}$$

the response of the reset-and-discharge synapse can be formally expressed as

$$I_{syn}(t) \;=\; I_{w0} e^{-\frac{t}{\tau}} \cdot \int_0^t \delta(\xi - t_n) e^{\frac{\xi}{\tau}} d\xi \;=\; I_{w0} e^{-\frac{(t-t_n)}{\tau}}. \tag{2.5}$$

Although this synaptic circuit produces an EPSC that lasts longer than the duration of its input pulses and decays exponentially with time, its response depends on only the last (nth) input spike. This nonlinear property of the circuit fails to reproduce the linear summation property of postsynaptic currents often desired in synaptic models and makes the theoretical analysis of networks of neurons interconnected with this synapse intractable.

2.3 Linear Charge-and-Discharge Synapse. In Figure 2 (left), we show a modification of the reset-and-discharge synapse that has often been used by the neuromorphic engineering community and was recently presented in Arthur and Boahen (2004). Here the presynaptic pulse, applied to the input n-FET M_{pre}, is active high. Assuming that all transistors are saturated and operate in subthreshold, the circuit behavior is the following. During an input pulse, the node $V_{syn}(t)$ decreases linearly, at a rate set by the net current $I_w - I_\tau$, and the synapse EPSC $I_{syn}(t)$ increases exponentially (charge phase). In between spikes, the $V_{syn}(t)$ node is recharged toward V_{dd} at a rate set by I_τ,

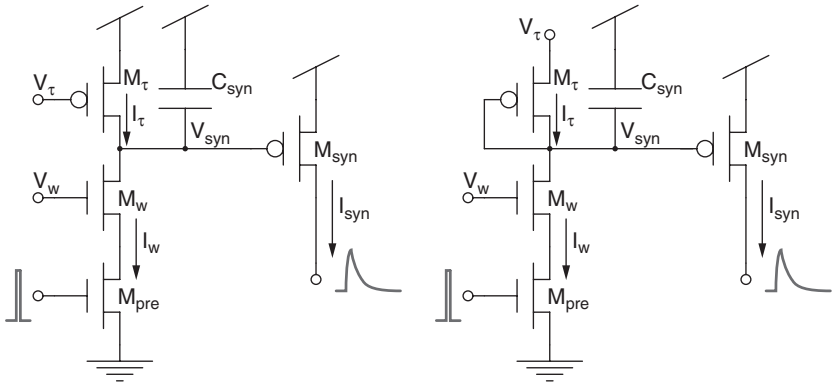

Figure 2: (Left) Linear charge-and-discharge synapse. (Right) Current mirror integrator synapse.

and $I_{syn}(t)$ decreases exponentially with time (discharge phase). The circuit equations that describe this behavior are

$$I_{syn}(t) = \begin{cases} I_{syn}^{-} e^{+\frac{(t-t_i^-)}{\tau_c}} & \text{(charge phase)} \\ I_{syn}^{+} e^{-\frac{(t-t_i^+)}{\tau_d}} & \text{(discharge phase),} \end{cases} \tag{2.6}$$

where t_i^- is the time at which the ith input spike arrives, t_i^+ the time at which it ends, I_{syn}^- the initial condition at t_i^-, I_{syn}^+ the initial condition at t_i^+, $\tau_c = \frac{U_T C_{syn}}{\kappa(I_w - I_\tau)}$ is the charge phase time constant, and $\tau_d = \frac{U_T C_{syn}}{\kappa I_\tau}$ the discharge phase time constant.

Assuming that each spike lasts a fixed brief period Δt, and considering two successive spikes arriving at times t_i^- and t_{i+1}^-, we can then write

$$I_{syn}(t_{i+1}^-) = I_{syn}(t_i^-) e^{\Delta t \left(\frac{1}{\tau_c} + \frac{1}{\tau_d}\right)} e^{-\frac{(t_{i+1}^- - t_i^-)}{\tau_d}}. \tag{2.7}$$

From this recursive equation, we derive the response of the linear charge-and-discharge synapse to a generic spike sequence $\rho(t)$ of n spikes

$$I_{syn}(t) = I_0 e^{n\Delta t \left(\frac{1}{\tau_c} + \frac{1}{\tau_d}\right)} e^{-\frac{t}{\tau_d}}, \tag{2.8}$$

assuming as the initial condition $V_{syn}(0) = V_{dd}$.

The EPSC dynamics depend on the total number of spikes n received at time t and on the circuit's time constants τ_c and τ_d. If we denote the input

spike train frequency at time t with $f = (n/t)$, we can express equation 2.8 as

$$I_{syn}(t) = I_0 e^{-\frac{\tau_c - f\Delta t(\tau_c + \tau_d)}{\tau_c \tau_d} t}. \tag{2.9}$$

The major drawback of this circuit, aside from its not being a linear integrator, is that if the argument of the exponential in equation 2.9 is positive (i.e., if $f > \frac{1}{\Delta t}\frac{I_\tau}{I_w}$), the output current increases exponentially with time, and the circuit's response saturates: $V_{syn}(t)$ decreases all the way to Gnd, and $I_{syn}(t)$ increases to its maximum value. This can be a problem because in these conditions, the circuit's steady-state response does not encode the input frequency.

2.4 Current-Mirror-Integrator Synapse. In his doctoral dissertation, Boahen (1997) proposed a synaptic circuit that differs from the linear charge-and-discharge one by a single node connection (see Figure 2) but that has a dramatically different behavior. The two transistors $M_\tau - M_{syn}$ of Figure 2 (right) implement a p-type current mirror, and together with the capacitor C_{syn}, they form a current mirror integrator (CMI). The CMI synapse implements a nonlinear pulse integrator circuit that produces a mean output current I_{syn} that increases with input firing rates and has a saturating nonlinearity with maximum amplitude that depends on the circuit's synaptic weight bias V_w and on its time constant bias V_τ.[1]

The CMI response properties have been derived analytically in Hynna and Boahen (2001) for steady-state conditions. An explicit solution of the CMI response to a generic spike train, which does not require the steady-state assumption, was also derived in Chicca (2006). According to the analysis presented in Chicca, the CMI response to a spike arriving at t_i^- and ending at t_i^+ is

$$I_{syn}(t) = \begin{cases} \dfrac{\alpha I_w}{1 + \left(\frac{\alpha I_w}{I_{syn}^-} - 1\right) e^{-\frac{(t - t_i^-)}{\tau_c}}} & \text{(charge phase)} \\ \dfrac{I_w}{\frac{I_w}{I_{syn}^+} + \frac{(t - t_i^+)}{\tau_d}} & \text{(discharge phase),} \end{cases} \tag{2.10}$$

where t_i^-, t_i^+, I_{syn}^-, and I_{syn}^+ are the same as defined in equation 2.6, $\alpha = e^{\frac{(V_\tau - V_{dd})}{U_T}}$, $\tau_c = \frac{C_{syn} U_T}{\kappa I_w}$, and $\tau_d = \alpha \tau_c$.

[1] The CMI does not implement a linear integrator filter; therefore the term *time constant* is improperly used. We use it in this context to denote a parameter that controls the temporal extension of the CMI's impulse response.

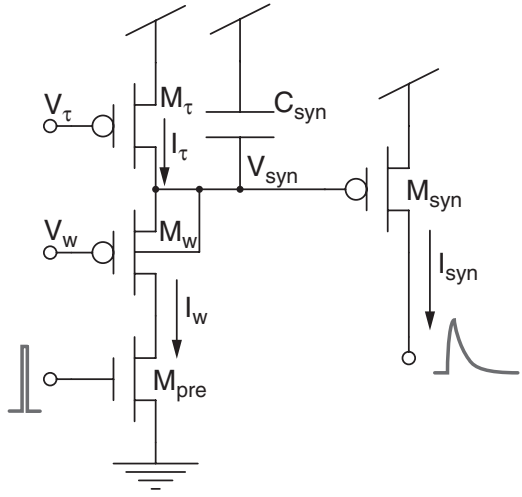

Figure 3: Log-domain integrator synapse.

During the charge phase, the EPSC increases over time as a sigmoidal function, while during the discharge phase, it decreases with a $1/t$ profile. The discharge of the EPSC is therefore extremely fast compared to the typical exponential decay profiles of other synaptic circuits. The parameter α (set by the V_τ bias voltage) can be used to slow the EPSC response profile. However, this parameter affects both the length of the EPSC discharge profile and the maximum amplitude of the EPSC charge phase: longer response times (larger values of τ_d) produce higher EPSC values.

Despite these problems and although the CMI cannot be used to linearly sum postsynaptic currents, this circuit was very popular and has been extensively used by the neuromorphic engineering community (Boahen, 1998; Horiuchi & Hynna 2001; Indiveri, 2000; Liu et al., 2001).

2.5 Log-Domain Integrator Synapse. More recently Merolla and Boahen (2004) proposed another variant of the linear charge-and-discharge synapse that implements a true linear integrator circuit. This circuit (shown in Figure 3) exploits the logarithmic relationship between subthreshold MOSFET gate-to-source voltages and their channel currents and is therefore called a *log-domain* filter. The output current I_{syn} of this circuit has the same exponential dependence on its gate voltage V_{syn} as all other synapses presented (see equation 2.2). Therefore, we can express its derivative with respect to time as

$$\frac{d}{dt} I_{syn} = -I_{syn} \frac{\kappa}{U_T} \frac{d}{dt} V_{syn}. \tag{2.11}$$

During an input spike (charge phase), the dynamics of the V_{syn} are governed by the equation: $C_{syn}\frac{d}{dt}V_{syn} = -(I_w - I_\tau)$. Combining this first-order differential equation with equation 2.11, we obtain

$$\tau \frac{d}{dt} I_{syn} + I_{syn} = I_{syn} \frac{I_w}{I_\tau}, \tag{2.12}$$

where $\tau = \frac{C_{syn} U_T}{\kappa I_\tau}$. The beauty of this circuit lies in the fact that the term I_w is inversely proportional to I_{syn} itself:

$$I_w = I_0 e^{-\frac{\kappa(V_w - V_{syn})}{U_T}} = I_0 e^{-\frac{\kappa(V_w - V_{dd})}{U_T}} e^{\frac{\kappa(V_{syn} - V_{dd})}{U_T}} = I_{w0} \frac{I_0}{I_{syn}}, \tag{2.13}$$

where I_0 is the leakage current and I_{w0} is the current flowing through M_w in the initial condition, when $V_{syn} = V_{dd}$. When this expression of I_w is substituted in equation 2.12, the right term of the differential equation loses the I_{syn} dependence and becomes the constant factor $\frac{I_0 I_{w0}}{I_\tau}$.

Therefore, the log-domain integrator transfer function takes the form of a canonical first-order low-pass filter equation, and its response to a spike arriving at t_i^- and ending at t_i^+ is

$$I_{syn}(t) = \begin{cases} \dfrac{I_0 I_{w0}}{I_\tau}\left(1 - e^{-\frac{(t - t_i^-)}{\tau}}\right) + I_{syn}^- e^{-\frac{(t - t_i^-)}{\tau}} & \text{(charge phase)} \\ I_{syn}^+ e^{-\frac{(t - t_i^+)}{\tau}} & \text{(discharge phase).} \end{cases} \tag{2.14}$$

This is the only synaptic circuit of the ones described up to now that has linear filtering properties. The same silicon synapse can be shared to sum the contributions of spikes potentially arriving from different sources in a linear way. This could save significant amounts of silicon real estate in neural architectures where the synapses do not implement learning or local adaptation mechanisms and could therefore solve many of the problems that have hindered the development of large-scale VLSI multineuron chips up to now. However, this particular circuit has two drawbacks. One problem is that the VLSI layout of the schematic shown in Figure 3 requires more area than the layout of other synaptic circuits, because the M_w p-FET has to live in an "isolated well" structure (Liu et al., 2002). The second, and more serious, problem is that the spike lengths used in pulse-based neural network systems, which typically last less than a few microseconds, are too short to inject enough charge in the membrane capacitor of the postsynaptic neuron to see any effect. The maximum amount of charge possible is $\Delta Q = \frac{I_0 I_{w0}}{I_\tau}\Delta t$, and I_{w0} cannot be increased beyond subthreshold current limits (of the order of nano-amperes); otherwise, the log domain properties of the filter break down (note that also I_τ is fixed, to set the desired time constant

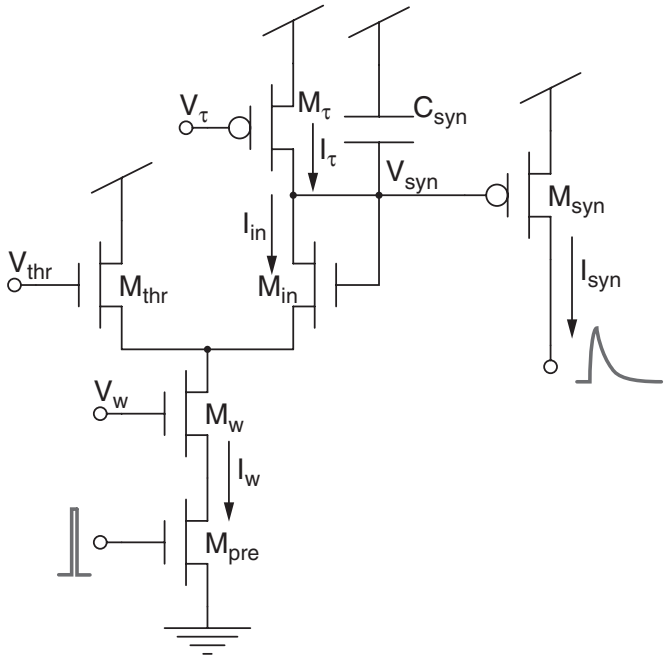

Figure 4: Diff-pair integrator synapse.

τ). A possible solution is to increase the fast (off-chip) input pulse lengths with on-chip pulse extenders (e.g., with CMI circuits). But this solution requires additional circuitry at each input synapse and makes the layout of the overall circuit even larger (Merolla & Boahen 2004).

2.6 Diff-Pair Integrator Synapse. The DPI circuit that we designed solves the problems of the log domain integrator synapse while maintaining its linear filtering properties, thus preserving the possibility of multiplexing in time spikes arriving from different sources. The schematic diagram of the DPI synapse is shown in Figure 4. This circuit comprises four n-FETs, two p-FETs, and a capacitor. The n-FETs form a differential pair whose branch current I_{in} represents the input to the synapse during the charge phase. Assuming subthreshold operation and saturation regime, the diff-pair branch current I_{in} can be expressed as

$$I_{in} = I_w \frac{e^{\frac{\kappa V_{syn}}{U_T}}}{e^{\frac{\kappa V_{syn}}{U_T}} + e^{\frac{\kappa V_{thr}}{U_T}}}, \tag{2.15}$$

and multiplying the numerator and denominator of equation 2.15 by $e^{-\frac{\kappa V_{dd}}{U_T}}$, we can express I_{in} as

$$I_{in} = \frac{I_w}{1 + \left(\frac{I_{syn}}{I_{gain}}\right)}, \tag{2.16}$$

where the term $I_{gain} = I_0 e^{-\frac{\kappa(V_{thr} - V_{dd})}{U_T}}$ represents a virtual p-type subthreshold current that is not tied to any p-FET in the circuit.

As for the log-domain integrator, we can combine the C_{syn} capacitor equation $C_{syn}\frac{d}{dt}V_{syn} = -(I_{in} - I_\tau)$ with equation 2.11 and write

$$\tau \frac{d}{dt} I_{syn} = -I_{syn}\left(1 - \frac{I_{in}}{I_\tau}\right), \tag{2.17}$$

where (as usual) $\tau = \frac{C U_T}{\kappa I_\tau}$. Replacing I_{in} from equation 2.16 into equation 2.17, we obtain

$$\tau \frac{d}{dt} I_{syn} + I_{syn} = \frac{I_w}{I_\tau} \frac{I_{syn}}{1 + \left(\frac{I_{syn}}{I_{gain}}\right)}. \tag{2.18}$$

This is a first-order nonlinear differential equation; however, the steady-state condition can be solved in closed form, and its solution is

$$I_{syn} = \frac{I_{gain}}{I_\tau}(I_w - I_\tau). \tag{2.19}$$

If $I_w \gg I_\tau$, the output current I_{syn} will eventually rise to values such that $I_{syn} \gg I_{gain}$, when the circuit is stimulated with an input step signal. If $\frac{I_{syn}}{I_{gain}} \gg 1$ the I_{syn} dependence in the second term of equation 2.18 cancels out, and the nonlinear differential equation simplifies to the canonical first-order low-pass filter equation:

$$\tau \frac{d}{dt} I_{syn} + I_{syn} = \frac{I_w I_{gain}}{I_\tau}. \tag{2.20}$$

In this case, the response of the DPI synapse to a spike arriving at t_i^- and ending at t_i^+ is

$$I_{syn}(t) = \begin{cases} \frac{I_{gain} I_w}{I_\tau}\left(1 - e^{-\frac{(t - t_i^-)}{\tau}}\right) + I_{syn}^- e^{-\frac{(t - t_i^-)}{\tau}} & \text{(charge phase)} \\ I_{syn}^+ e^{-\frac{(t - t_i^+)}{\tau}} & \text{(discharge phase)} \end{cases}. \tag{2.21}$$

The solution of the DPI synapse is almost identical to the one of the log-domain integrator synapse, described in equation 2.14. The only difference is that the term I_0 of equation 2.14 is replaced by I_{gain}. This scaling factor can be used to amplify the charge phase response amplitude, therefore solving the problem of generating sufficiently large charge packets sourced into the neuron's integrating capacitor for input spikes of very brief duration, while keeping all currents in the subthreshold regime and without requiring additional pulse-extender circuits. In addition, the layout of DPI does not require isolated well structures and can be implemented in a very compact way.

As for the log-domain integrator synapse described in section 2.5, the DPI synapse implements a low-pass filter with linear transfer function (under the realistic assumption that $I_w \gg I_\tau$). Although it is less compact than the synaptic circuits described in sections 2.1, 2.2, 2.3, and 2.4, it is the only one that can reproduce the exponential dynamics observed in excitatory and inhibitory postsynaptic currents of biological synapses (Destexhe et al., 1998), without requiring additional input pulse-extender circuits. Moreover, the DPI synapse we propose has independent control of time constant, synaptic weight, and synaptic scaling parameters. The extra degree of freedom obtained with the V_{thr} parameter can be used to globally scale the efficacies of the DPI circuits that share the same V_{thr} bias. This feature could in turn be employed to implement global homeostatic plasticity mechanisms complementary to local spike-based plasticity ones acting on the synaptic weight V_w node (see also section 4). In the next section, we present experimental results from a VLSI chip comprising an array of DPI synapses connected to low-power leaky I&F neurons (Indiveri, Chicca, & Douglas, 2006) that validate the analytical derivations presented here.

3 Experimental Results

We fabricated a prototype chip in standard AMS 0.35 μm CMOS technology comprising the DPI circuit and additional test structures to augment the synapse's functionality. Here we present experimental results measured from the basic DPI circuit of Figure 4, while the characteristics and measurements from the additional test circuits are described in section 4. In Figure 5, we show a picture of the synaptic circuit layout. The full layout occupies an area of 1360 μm^2. These types of synaptic circuits can therefore be used to implement networks of spiking neurons with a very large number of synapses on a small chip area. For example, in a recent chip, we implemented a network comprising 8192 synapses and 32 neurons (256 synapses per neuron) using an area of only 12 mm^2 (Mitra, Fusi, & Indiveri, 2006). The silicon area occupied by the synaptic circuit can vary significantly, as it depends on the choice of layout design solutions. More conservative solutions use large transistors, have lower mismatch, and require more area. More aggressive solutions require less area, but multiple instances of the

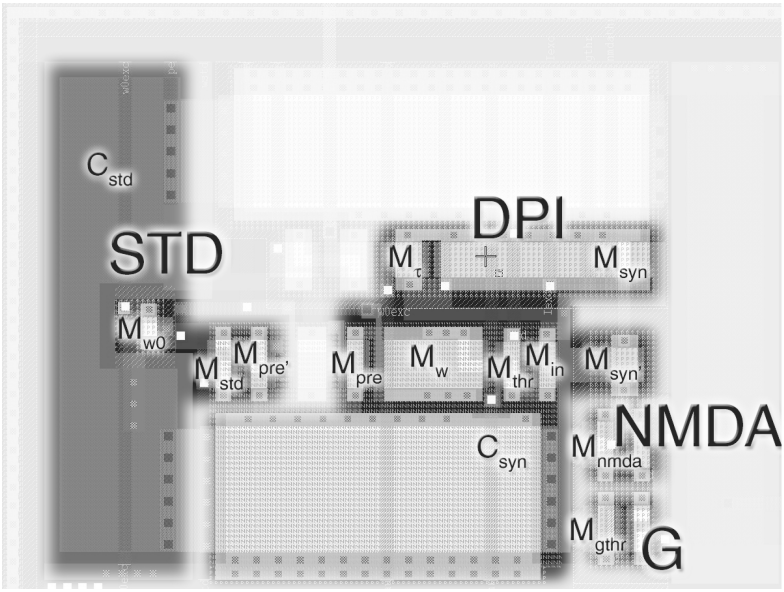

Figure 5: Layout of the fabricated DPI synapse and additional circuits that augment the synapse's functionality. The schematic diagram and properties of the STD, NMDA, and G blocks are described in section 4.

same layout cell produce currents with larger deviations. The layout of Figure 5 implements a very conservative solution.

To validate the theoretical analysis of section 2.6, we measured the DPI step response and fitted the experimental data with equation 2.21. In Figure 6 (left), we plot the circuit's step response for different synaptic weight V_w bias values. The rise and decay parts of the data were fitted with the charge phase and discharge phase parts of equation 2.21 using slightly different parameters for the estimated time constant. The small differences in the time constants are most likely due to leakage currents and parasitic capacitance effects, not considered in the analytical derivations. These results, however, show that the DPI time constant does not depend on V_w and can be independently tuned with V_τ.

Silicon synapses are typically stimulated with trains of pulses (spikes) of very brief duration, separated by longer interspike intervals (ISIs). It can be easily shown from equation 2.20 that when the DPI is stimulated with a spike train of average frequency f_{in} and pulse duration Δt, its steady-state response is

$$< I_{syn} >= \left(\frac{I_{gain} I_w}{I_\tau} \right) \Delta t f_{in}. \tag{3.1}$$

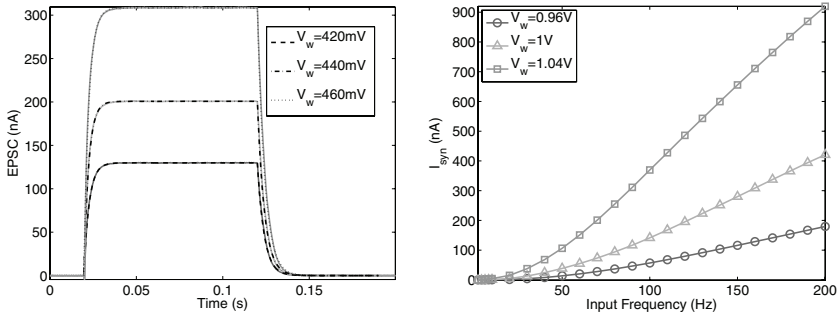

Figure 6: DPI circuit response properties. (Left) Step response for three different values V_w. The response is fitted with equation 2.21, and the fitting functions (dotted, and dashed lines) are superimposed to the measured data. The time constants estimated by the fit are $\tau = 3$ ms for the charge phase and $\tau = 4$ ms for the discharge phase. (Right) Response to spike trains of increasing frequencies. The output mean current is linear with the synaptic input frequency, and its gain can be changed with the synaptic weight bias V_w.

We also verified this derivation by measuring the mean EPSC of the circuit in response to spike trains of increasing frequencies. In Figure 6 (right), we show the $i - f$ curve for typical biological spiking frequencies, ranging from 10 to 200 Hz. The mean output current is linear over a wide range of input frequencies (extending well beyond the ones shown in the plot).

4 Synaptic Dynamics

The results of the previous sections showed how the DPI response models the EPSC generated by biological excitatory synapses of AMPA type (Destexhe et al., 1998). Inhibitory ($GABA_a$) type synapses can be easily emulated by using the complementary version of the DPI circuit of Figure 4 (with a p-type diff-pair, and n-type output transistor). Additional circuits can be attached to the DPI synapse to extend the model with additional features typical of biological synapses and implement various types of plasticity. For example, by adding two extra transistors, we can implement voltage-gated channels that model NMDA synapse behavior. Similarly, by using two more transistors, we can extend the synaptic model to be conductance based (Kandel, Schwartz, & Jessell, 2000). Furthermore, the DPI circuit is compatible with previously proposed circuits for implementing synaptic plasticity, on both short timescales with models of short-term depression (STD) (Rasche & Hahnloser, 2001; Boegerhausen, Suter, & Liu, 2003) and on longer timescales with spike-based learning mechanisms, such as spike-timing-dependent plasticity (STDP) (Indiveri et al., 2006). Finally the DPI's extra degree of freedom for modifying the overall gain of the synapse

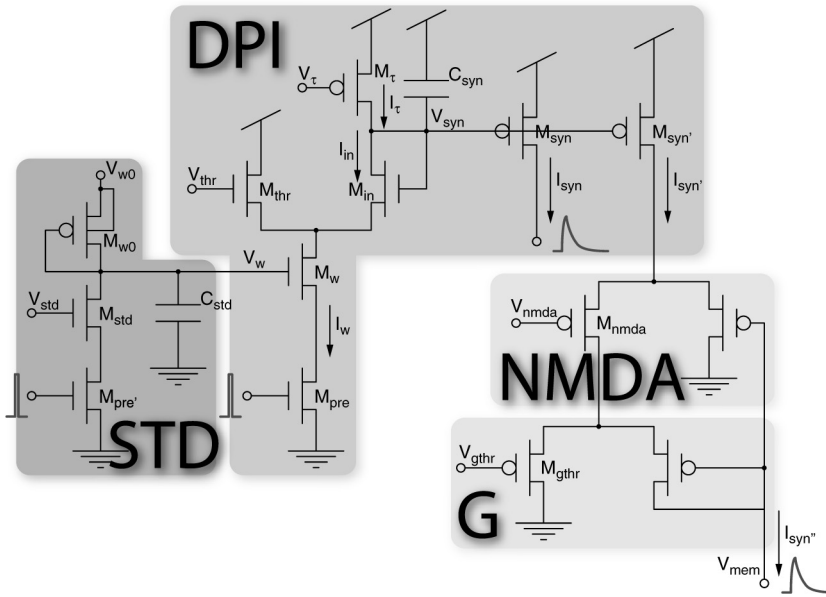

Figure 7: Schematic diagram of the DPI connected to additional test circuits that augment the synapse's functionality. The names of the functional blocks correspond to the ones used in the layout of Figure 5: The STD block comprises the circuit modeling short-term depression of the synaptic weight, the NMDA block comprises the transistors modeling NMDA voltage-gated channels, and the G block includes transistors that render the synapse conductance based.

either with V_{thr} or with V_w allows the implementation of synaptic homeostatic mechanisms (Bartolozzi & Indiveri, 2006), such as global activity dependent synaptic scaling (Turrigiano, Leslie, Desai, Rutherford, & Nelson, 1998).

In Figure 7, we show the schematics of the extension circuits mentioned above implemented on the test chip (with the exception of the STDP and homeostatic circuits). In the next paragraphs, we describe the behavior of these additional circuits, characterized by measuring the membrane potential V_{mem} of a low power leaky I&F neuron (Indiveri et al., 2006) that receives in input the synaptic EPSC.

4.1 NMDA Synapse. With the DPI we reproduce phenomenologically the current flow through ionic ligand-gated membrane channels that open and let the ions flow across the postsynaptic membrane as soon as they sense the neurotransmitters released by the presynaptic boutons (e.g., AMPA channels). Another important class of ligand-gated synaptic

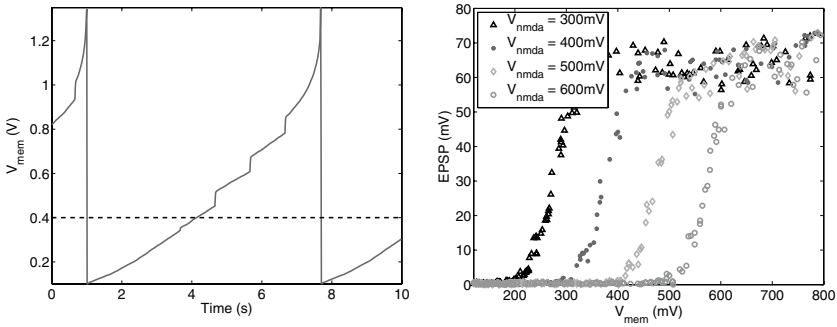

Figure 8: NMDA-type synapse response properties. (Left) Membrane potential of an I&F neuron connected to the synapse and stimulated by a constant injection current. The NMDA threshold voltage is set to $V_{nmda} = 400$ mV. The small bumps in V_{mem} represent the excitatory postsynaptic potentials (EPSPs) produced by the synapse, when $V_{mem} > V_{nmda}$, in response to the presynaptic input spikes. (Right) EPSP amplitude versus the membrane potential, for increasing values of the NMDA threshold V_{nmda} and for a fixed value of V_w.

channels, the NMDA receptors, is, in addition, voltage gated; these channels open to let the ions flow only if the membrane voltage is depolarized above a given threshold while in the presence of its neurotransmitter (glutamate). We can implement this behavior by exploiting the thresholding property of the differential pair circuit, as shown in Figure 7; if the node V_{mem} is lower than the externally set bias V_{nmda}, the output current I_{syn} flows through the transistor M_{nmda} in the left branch of the diff-pair and has no effect on the postsynaptic depolarization. On the other hand, if V_{mem} is higher than V_{nmda}, the current flows also into the membrane potential node, depolarizing the I&F neuron, and thus implementing the voltage-gating typical of NMDA synapses.

In Figure 8, we show the results measured from the test circuit on the prototype chip: we stimulate the synapse with presynaptic spikes, while also injecting constant current into the neuron's membrane. The synapse's EPSC amplitude depends on the difference between the membrane potential and the NMDA threshold V_{nmda}. As expected, when V_{mem} is smaller than V_{nmda}, the synaptic current is null, and the membrane potential increases solely due to the constant injection current. As V_{mem} increases above V_{nmda}, the contribution of the synaptic current injected with each presynaptic spike becomes visible. The time constant of the DPI circuit used in this way can be easily extended to hundreds of milliseconds (values typical of NMDA-type synaptic dynamics) by increasing the V_τ bias voltage of Figure 7. This allows us to faithfully reproduce both the voltage-gated and temporal dynamic properties of real NMDA synapses. It is important to be able to implement

these properties in our VLSI devices because there is evidence that they play an important role in detecting coincidence between the presynaptic activity and postsynaptic depolarization for inducing long-term-potentiation (LTP) (Morris, Davis, & Butcher, 1990). Furthermore, the NMDA's synapse stabilizing role, hypothesized by computational studies within the context of working memory (Wang, 1999), could be useful for stabilizing persistent activity of recurrent VLSI networks of spiking neurons.

4.2 Conductance-Based Synapse. So far we have reproduced the total current flowing through the synaptic channels independent of the postsynaptic membrane potential. However, in real synapses, the current is proportional to the difference between the postsynaptic membrane voltage and the synaptic ion reversal potential E_{ion}:

$$I_{syn} = g_{syn}(V_{mem} - E_{ion}). \tag{4.1}$$

Exploiting once more the properties of the differential pair circuit, we can model this dependence with just two more transistors (see G block of Figure 7), and obtain a behavior that, to first-order approximation, is equivalent to that described by equation 4.1. Formally, the conductance-based synapse output is:

$$I_{syn''} = I_{syn'} \frac{1}{1 + e^{\frac{\kappa}{U_T}(V_{mem} - V_{gthr})}}, \tag{4.2}$$

so if we consider the first-order term of the Taylor expansion, when $V_{mem} \cong V_{gthr}$, we obtain

$$I_{syn''} = \frac{I_{syn'}}{2} + g_{syn}(V_{mem} - V_{gthr}), \tag{4.3}$$

where the conductance term $g_{syn} = I_{syn'} \frac{\kappa}{4U_T}$.

In Figure 9 we plot the EPSPs measured from the I&F neuron connected to the conductance-based synapse for different values of V_{gthr}. These experimental results show that our synapse can reproduce the behavior of conductance-based synapses. This behavior is especially relevant in inhibitory synapses, where the dependence expressed in equation 4.1 results in shunting inhibition. Computational and biological studies have attributed different roles to shunting inhibition, such as logical AND-NOT) Koch, Poggio, & Torre, 1983), and normalization (Carandini, Heeger, & Movshon, 1997) functions. Evidence for these and other hypotheses continues to be the subject of further investigation (Anderson, Carandini, & Ferster, 2000; Chance, Abbott, & Reyes, 2002). The implementation of shunting inhibition in large arrays of VLSI synapses and spiking neurons provides

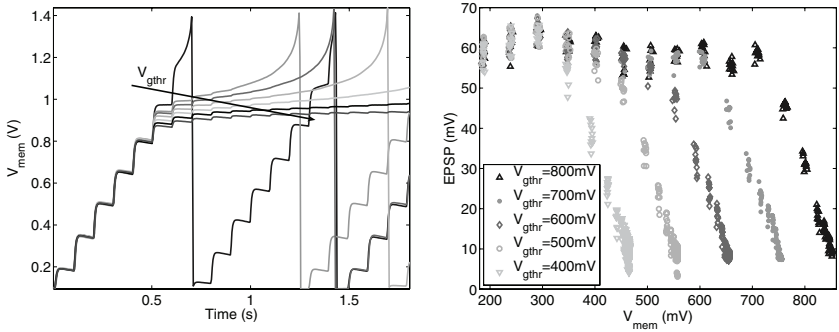

Figure 9: Conductance-based synapse. (Left) Membrane potential of the I&F neuron stimulated by the synapse for different values of the synaptic reversal potential V_{gthr}. (Right) EPSP amplitude as a function of V_{mem} for different values of V_{gthr}.

an additional means for exploring the computational role of this computational primitive.

4.3 Synaptic Plasticity. In the previous sections, we showed that our circuit can model biologically realistic synaptic current dynamics. The synaptic main feature exploited in neural networks, though, is plasticity: the ability of changing the synaptic efficacy to learn and adapt to the environment. In neural networks with large arrays of synapses and neurons, usually (Indiveri et al., 2006; Mitra et al. 2006; Arthur & Boahen, 2004; Shi & Horiuchi, 2004b) all the synapses belonging to one population share the same bias that sets their initial weight.[2] In addition each synapse can be connected to a local circuit for the short- and/or long-term modification of its weight. Our silicon synapse supports all of the short-term and long-term plasticity mechanisms for inducing long-term potentiation (LTP) and long-term depression (LTD) in the synaptic weight that have been proposed in the literature. Specifically, the possibility of biasing M_w with subthreshold voltages on the order of hundreds of mV makes the DPI compatible with many of the spike-timing-dependent plasticity circuits previously proposed (Indiveri et al., 2006; Mitra et al., 2006; Arthur & Boahen, 2006; Bofill, Murray, & Thompson, 2002).

Similarly the DPI synapse is naturally extended with the short-term depression circuit proposed by Rasche and Hahnloser (2001), where the synaptic weight decreases with increasing number of input spikes and recovers during periods of presynaptic inactivity. From the computational point of

[2]The initial weight V_w can be set by an external voltage reference or by on-chip bias generators.

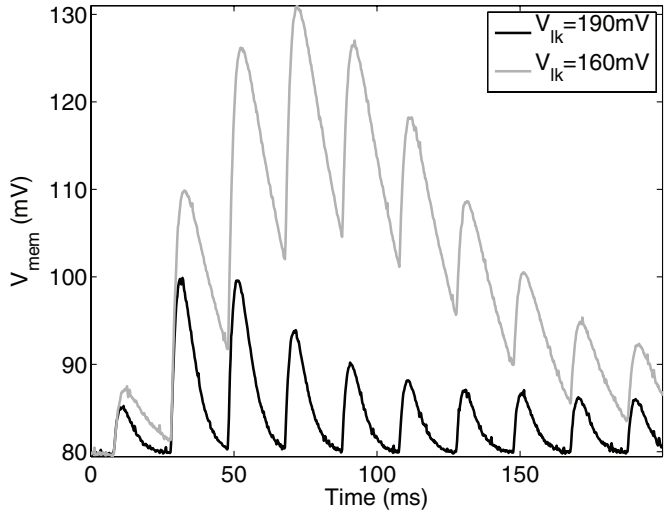

Figure 10: Short-term depression: Membrane potential of the leaky I&F neuron, when the short-term depressing synapse is stimulated with a regular spike train at 50 Hz. The different traces of the membrane potential correspond to different values of the leakage current of the neuron. Note how (from the second spike on) the EPSP amplitude decreases with each input spike.

view, STD is a nonlinear mechanism that plays an important role for implementing selectivity to transient stimuli and contrast adaptation (Chance, Nelson, & Abbott, 1998). In Figure 10, we show the EPSPs of the I&F neuron connected to the synapse, having activated the STD block of Figure 7. These results confirm the compatibility between the DPI and the STD circuits and show qualitatively the effect of short-term depression. Quantitative considerations and comparisons to short-term depression computational models have already been presented elsewhere (Rasche & Hahnloser, 2001; Boegerhausen et al. 2003).

Another valuable property of biological synapses is the homeostatic mechanism known as activity-dependent synaptic scaling (Turrigiano et al., 1998). It acts by scaling the synaptic weights in order to keep the neurons, firing rate within a functional range in the face of chronic changes of their activity level while preserving the relative differences between individual synapses. As demonstrated in section 2.6 and Figure 11, we can scale the total synaptic efficacy of the DPI by independently varying either I_w or I_{gain} (see also Figure 6 (left)). We can exploit these two independent degrees of freedom for learning the synaptic weight V_w with "fast" spike-based learning rules, while adapting the bias V_{thr} to implement homeostatic synaptic scaling, on much slower timescales. A control algorithm that exploits the

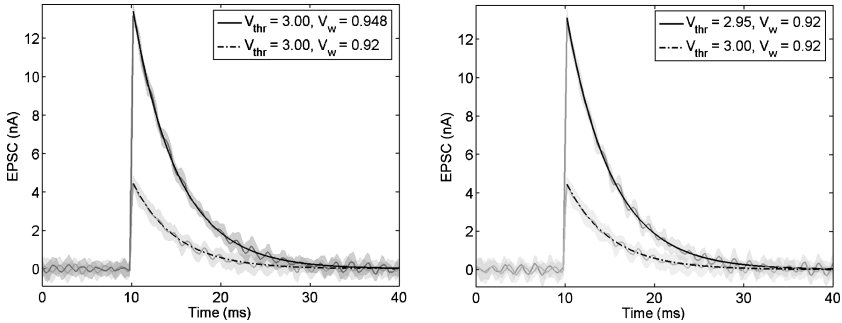

Figure 11: Independent scaling of EPSC amplitude by adjusting either V_{thr} or V_w. The plots show the time course of mean and standard deviation (over 10 repetitions of the same experiment) of the current I_{syn}, in response to a single-input voltage pulse. In both plots, the lower EPSC traces share the same set of V_{thr} and V_w, in (Left) The higher EPSC is obtained by increasing V_w and (right) by decreasing V_{thr}, with respect to the initial bias set. Superimposed to the experimental data, we plot theoretical fits of the decay from equation 2.21. The time constant of all plots is the same and equal to 5 ms.

properties of the DPI to implement the activity-dependent synaptic scaling homeostatic mechanism has been recently proposed by Bartolozzi and Indiveri (2006).

5 Conclusion

We have proposed a new analog VLSI synapse circuit (the DPI of section 2.6) useful for implementing postsynaptic currents in neuromorphic VLSI networks of spiking neurons with biologically realistic temporal dynamics. We showed in analytical derivations and experimental data that the circuit proposed matches detailed computational models of synapses. We compared our VLSI synapse to previously proposed circuits that implement an equivalent functionality and derived analytically their transfer functions. Our analysis showed that the DPI circuit incorporates most of the strengths of previously proposed circuits, while providing additional favorable properties. Specifically, the DPI implements a linear integrator circuit with two independent tunable gain parameters and one independently tunable timeconstant parameter. The circuit's mean output current encodes linearly the input frequency of the presynaptic spike train. As the DPI performs linear temporal summation of its input spikes, it can be used for processing multiple spike trains generated by different sources multiplexed together, modeling the contribution of many different synapses that share the same weight.

Next to being linear and compact, this circuit is compatible with existing implementation of both short-term and long-term plasticity. The favorable features of linearity, compactness, and compatibility with existing synaptic circuit elements make it an ideal building block for constructing adaptive dynamic synapses and implementing dense and massively parallel networks of spiking neurons capable of processing spatiotemporal signals in real time. The VLSI implementation of such networks constitutes a powerful tool for exploring the computational role of each element described in this work, from the voltage-gated NMDA channels, to shunting inhibition, and homeostasis, using real-world stimuli while observing the network's behavior in real time.

Acknowledgments

This work was supported in part by the EU Grants ALAVLSI (IST-2001-38099) and DAISY (FP6-2005-015803) and in part by the ETH TH under Project 0-20174-04. The chip was fabricated via the EUROPRACTICE service. We thank Pratap Kumar for fruitful discussions about biological synapses.

References

Anderson, J., Carandini, M., & Ferster, D. (2000). Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. *Journal of Physiology, 84*, 909–926.

Arthur, J., & Boahen, K. (2004, July). Recurrently connected silicon neurons with active dendrites for one-shot learning. In *IEEE International Joint Conference on Neural Networks* (Vol. 3, pp. 1699–1704). Piscataway, NJ: IEEE.

Arthur, J., & Boahen, K. (2006). Learning in silicon: Timing is everything. In Y. Weiss, B. Schölkopf, & J. Platt (Eds.), *Advances in neural information processing systems, 18*. Cambridge, MA: MIT Press.

Bartolozzi, C., & Indiveri, G. (2006). Silicon synaptic homeostasis. In *Brain Inspired Cognitive Systems 2006*. [CD]

Boahen, K. A. (1997). *Retinomorphic vision systems: Reverse engineering the vertebrate retina*. Unpublished doctoral dissertation, California Institute of Technology.

Boahen, K. (1998). Communicating neuronal ensembles between neuromorphic chips. In T. S. Lande (Ed.), *Neuromorphic systems engineering* (pp. 229–259). Norwell, MA: Kluwer Academic.

Boegerhausen, M., Suter, P., & Liu, S.-C. (2003). Modeling short-term synaptic depression in silicon. *Neural Computation, 15*(2), 331–348.

Bofill, A., Murray, A., & Thompson, D. (2002). Circuits for VLSI implementation of temporally asymmetric Hebbian learning. In T. G. Dietlerich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems, 14*. Cambridge, MA: MIT Press.

Borgstrom, T., Ismail, M., & Bibyk, S. (1990). Programmable current-mode neural network for implementation in analogue MOS VLSI. *IEE Proceedings G, 137(2)*, 175–184.

Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience, 17*(21), 8621–8644.

Chance, F., Abbott, L., & Reyes, A. (2002). Gain modulation from background synaptic input. *Neuron, 35*, 773–782.

Chance, F. S., Nelson, S. B., & Abbott, L. F. (1998). Synaptic depression and the temporal response characteristics of V1 cells. *Journal of Neuroscience, 18*(12), 4785–99.

Chicca, E. (2006). *A neuromorphic VLSI system for modeling spike-based cooperative competitive neural networks*. Unpublished doctoral dissertation, ETH Zurich, Zurich, Switzerland.

Chicca, E., Badoni, D., Dante, V., D'Andreagiovanni, M., Salina, G., Fusi, S., et al. (2003). A VLSI recurrent network of integrate-and-fire neurons connected by plastic synapses with long term memory. *IEEE Transactions on Neural Networks, 14*(5), 1297–1307.

Chicca, E., Indiveri, G., & Douglas, R. (2003). An adaptive silicon synapse. In *Proc. IEEE International Symposium on Circuits and Systems* (pp. I-81–I-84). Piscataway, NJ: IEEE.

Destexhe, A., Mainen, Z., & Sejnowski, T. (1998). Kinetic models of synaptic transmission. In C. Koch & I. Segev (Eds.), *Methods in neuronal modelling, from ions to networks* (pp. 1–25). Cambridge, MA: MIT Press.

Fusi, S., Annunziato, M., Badoni, D., Salamon, A., & Amit, D. J. (2000). Spike-driven synaptic plasticity: Theory, simulation, VLSI implementation. *Neural Computation, 12*, 2227–2258.

Gordon, C., Farquhar, E., & Hasler, P. (2004, May). A family of floating-gate adapting synapses based upon transistor channel models. In *2004 IEEE International Symposium on Circuits and Systems* (Vol. 1, pp. 317–1320). Piscataway, NJ: IEEE.

Gütig, R., & Sompolinsky, H. (2006). The tempotron: A neuron that learns spike timing–based decisions. *Nature Neuroscience, 9*, 420–428.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computation*. Reading, MA: Addison-Wesley.

Horiuchi, T., & Hynna, K. (2001). A VLSI-based model of azimuthal echolocation in the big brown bat. *Autonomous Robots, 11*(3), 241–247.

Hynna, K., & Boahen, K. (2001). Space-rate coding in an adaptive silicon neuron. *Neural Networks, 14*, 645–656.

Hynna, K. M., & Boahen, K. (2006, May). Neuronal ion-channel dynamics in silicon. In *2006 IEEE International Symposium on Circuits and Systems* (pp. 3614–3617). Piscataway, NJ: IEEE.

Indiveri, G. (2000). Modeling selective attention using a neuromorphic analog VLSI device. *Neural Computation, 12*(12), 2857–2880.

Indiveri, G., Chicca, E., & Douglas, R. (2006). A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity. *IEEE Transactions on Neural Networks, 17*(1), 211–221.

Kandel, E. R., Schwartz, J., & Jessell, T. M. (2000). *Principles of neural science*. New York: McGraw-Hill.

Koch, C. (1999). Synaptic input. In M. Stryker (Ed.), *Biophysics of computation: Information processing in single neurons* (pp. 85–116). New York: Oxford University Press.

Koch, C., Poggio, T., & Torre, V. (1983). Nonlinear interactions in a dendritic tree: Localization, timing, and role in information processing. *PNAS, 80*, 2799–2802.

Lazzaro, J. P. (1994). Low-power silicon axons, neurons, and synapses. In M. E. Zaghloul, J. L. Meador, & R. W. Newcomb (Eds.), *Silicon implementation of pulse coded neural networks* (pp. 153–164). Norwell, MA: Kluwer.

Liu, S.-C., Kramer, J., Indiveri, G., Delbruck, T., Burg, T., & Douglas, R. (2001). Orientation-selective aVLSI spiking neurons. *Neural Networks, 14*(6/7), 629–643.

Liu, S.-C., Kramer, J., Indiveri, G., Delbrück, T., & Douglas, R. (2002). *Analog VLSI: Circuits and principles*. Cambridge, MA: MIT Press.

Mead, C. (1989). *Analog VLSI and neural systems*. Reading, MA: Addison-Wesley.

Merolla, P., & Boahen, K. (2004). A recurrent model of orientation maps with simple and complex cells. In S. Thrün, L. K. Saul, & B. Schölkopf (Eds.), *Advances in neural information processing systems, 16* (pp. 995–1002). MIT Press. Cambridge, MA:

Mitra, S., Fusi, S., & Indiveri, G. (2006, May). A VLSI spike-driven dynamic synapse which learns. In *Proceedings of IEEE International Symposium on Circuits and Systems* (pp. 2777–2780). Piscataway, NJ: IEEE.

Morris, R., Davis, S., & Butcher, S. (1990). Hippocampal synaptic plasticity and NMDA receptors: A role in information storage? *Philosophical Transactions: Biological Sciences, 329*(1253), 187–204.

Murray, A. F. (1998). Pulse-based computation in VLSI neural networks. In W. Maass & C. M. Bishop (Eds.), *Pulsed neural networks* (pp. 87–109). Cambridge, MA: MIT Press.

Northmore, D. P. M., & Elias, J. G. (1998). Building silicon nervous systems with dendritic tree neuromorphs. In W. Maass & C. M. Bishop (Eds.), *Pulsed neural networks* (pp. 135–156). Cambridge, MA: MIT Press.

Rasche, C., & Hahnloser, R. (2001). Silicon synaptic depression. *Biological Cybernetics, 84*(1), 57–62.

Satyanarayana, S., Tsividis, Y., & Graf, H. (1992). A reconfigurable VLSI neural network. *IEEE J. Solid-State Circuits, 27*(1), 67–81.

Shi, R., & Horiuchi, T. (2004a). A summating, exponentially-decaying CMOS synapse for spiking neural systems. In S. Thrün, L. Saul, & B. Schölkopf (Eds.), *Advances in neural information processing systems, 16*. Cambridge, MA: MIT Press.

Shi, R., & Horiuchi, T. (2004b). A VLSI model of the bat lateral superior olive for azimuthal echolocation. In *Proceedings of the 2004 International Symposium on Circuits and Systems (ISCAS04)* (Vol. 4, pp. 900–903). Piscataway, NJ: IEEE.

Turrigiano, G., Leslie, K., Desai, N., Rutherford, L., & Nelson, S. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, 391, 892–896.

Wang, X. (1999). Synaptic basis of cortical persistent activity: The Importance of NMDA receptors to working memory. *Journal of Neuroscience, 19*, 9587–9603.

Received May 11, 2006; accepted September 27, 2006.