

Fan-In Analysis of a Leaky Integrator Circuit Using Charge Transfer Synapses

Thomas Dowrick^a, Liam McDaid^b, Stephen Hall^c

^aDepartment of Medical Physics & Biomedical Engineering, University College London, previously at University of Liverpool

^bSchool of Computational Neuroscience & Intelligent Systems, Ulster University

^cDepartment of Electrical Engineering & Electronics, The University of Liverpool

Abstract— It is shown that a simple leaky integrator (LI) circuit operating in a dynamic mode can allow spatial and temporal summation of weighted synaptic outputs. The circuit incorporates a current mirror configuration to sum charge packets released from charge transfer synapses and an n-channel MOSFET, operating in subthreshold, serves to implement a leakage capability, which sets the decay time for the postsynaptic response. The focus of the paper is to develop an analytical model for fan-in and validate the model against simulation and experimental results obtained from a prototype chip fabricated in the AMS 0.35 μ m mixed signal CMOS technology. We show that the model predicts the theoretical limit on fan-in, relates the magnitude of the postsynaptic response to weighted synaptic inputs and captures the transient response of the LI when stimulated with spike inputs.

Index Terms—neuromorphic circuits, fan-in, spiking neural network, leaky integrator, charge transfer synapse, CMOS

1. Introduction

SPIKING NEURAL NETWORKS (SNN) implemented in hardware are an increasingly popular area, both in research and in commercial settings. Spiking neurons encode information in the timing of single spikes, and not just in their statistical firing rate [1]. Recent neuroscience research has shown that SNNs mimic neuron behaviour on a level more closely related to biology and so have the propensity for powerful computational ability compared to classic artificial neural networks.

Artificial spiking neural networks can be implemented with either software or hardware approaches. Several software simulators [2] have been developed to simulate SNNs and allow investigation of the role played by spike-timing in the field of computational neuroscience. However, software simulations with general-purpose platforms require high computational cost with no guarantee of real-time performance. Even the latest supercomputer to date has not shown capability for achieving real-time and detailed simulations for a large-scale SNN over multiple cortical areas. Several computational systems based on FPGAs, GPUs and ARM processor cores [3–5] have been developed for hardware accelerated simulation which could offer such capability at the expense of large silicon area and low energy efficiency.

Implementing SNN with dedicated hardware however, has a number of important advantages over software solutions. The major advantage is high speed computation with inherent high parallelism and distributed computing ability. There has been extensive activity in the development of hardware SNN including digital, analogue and hybrid implementations. Good reviews of the current progress of hardware SNN development can be found in [6–8]. Different neuron models have been used in current hardware SNN projects, varying from very detailed conductance-based models to simpler leaky integrate and fire versions. Conductance-based models emulate biophysical ion channels and hence are more faithful to biology. The integrate and fire models are less realistic but require fewer transistors. Their compact layouts and low energy consumption allow designers to balance the accuracy with a higher number of neurons in the network and hence are more scalable. Each of the different implementations offers some trade-off between scalability, latency and biological realism. Finding an appropriate balance between these three elements represents one of the key challenges of hardware SNNs.

This paper describes the dynamics of a leaky integrator (LI) circuit that aggregates the output of multiple charge transfer synapses (CTS). The LI is verified using simulation/experimental results and an analytical model is developed to relate the postsynaptic response to the fan-in, n . The remainder of this paper is organized as follows. Section II presents a comparison between digital and analogue hardware approaches to SNN implementation while section III presents a description of the operation of the CTS. Section IV details the fan-in model for two distinctly different operating conditions while section V presents both experimental and simulations results to support the model. Section VI presents a discussion of the work followed by a conclusion in section VII. Derivations of equations and details of processing can be found in the appendices.

Submitted for review 19/10/2017. Revised 24/05/2018 The authors acknowledge funding of the project by the Engineering and Physical Science Research Council (EPSRC), project EP/F05551X/1. T. Dowrick also thanks EPSRC for funding his PhD study.

T. Dowrick is with the Department of Medical Physics & Biomedical Engineering, University College London and was previously with The University of Liverpool (Corresponding author. e-mail: t.dowrick@ucl.ac.uk).

S. Hall is with the Department of Electrical Engineering & Electronics at The University of Liverpool.

L. McDaid is with the School of Computational Neuroscience & Intelligent Systems at Ulster University.

2. Background

Hardware SNNs can be classified as analogue, digital or mixed signal hybrids. In analogue SNNs, neural signals such as synaptic weights and membrane status are mathematically presented as a set of continuous values of voltage, current, or charge [9–13]. This differs from digital SNNs which uses discrete quantities to represent signals [5]. The most important benefit of using an analogue approach is that neural functions such as temporal/spatial summation and weighting can be performed efficiently in real time using much less power and area than equivalent digital adders and multipliers. However, due to process variations, device non-linearity and noise disturbance in analogue VLSI system, the computational functions in analogue SNN are not precise and uniform across chips. Therefore, on-chip learning schemes and fault-tolerant architecture are crucial so that noise and process variations do not affect the performance [14,15].

Weight storage represents another challenge in hardware SNN. The use of digital random-access memory (RAM) allows weight information to be retained statically and also updated dynamically. However, only quantized weights can be stored and the weight update is normally based on a synchronous clock signal which requires more power than analogue memory. Both pure digital SNNs, and mixed signal SNNs can employ digital RAM to store the weight information. The major drawback of this mixed signal approach is that it requires a digital-analogue converter (DAC) per synapse. This increases the layout area and power consumption and also introduces conversion error and noise. Analogue non-volatile memory using floating gate devices offers long-term weight storage and also a continuous weight voltage [16,17]. Unfortunately, weight update with floating gate is complex and is inherently slow. Moreover, it requires the use of high-voltage and an expensive manufacturing process. Another type of analogue memory is based on charging/discharging a capacitor which allows rapid weight changes but no long-term memory. In order to hold the charge in the weight capacitors for a long period of time, op-amps with negative feedback, or similar variations, need to be integrated with the weighted capacitors. This again increases the required layout area and power and severely limits scalability. While floating gate devices could be the ultimate solution for large scale SNN with on-chip learning, the disadvantages of floating gates means that digital RAMs with DAC, and charging/discharging capacitors are more widely used in practice.

Achieving the high level of connectivity seen in biological systems is another challenge for the large scale neural networks. It is obvious that the digital approach has more potential as connections are more flexible and are less susceptible to noise. A commonly-used digital asynchronous communication protocol is the address-event representation (AER) system which allows massive connections between neurons; even across different chips [18]. In an AER system, every digital spike event is encoded with the identity/address of the sender neuron and transmitted over a common commutation bus. The address decoder selects the appropriate synapse to receive the spike. The disadvantage of the AER protocol is that the firing rate or communication speed is limited due to bus sharing and time multiplexing.

The mixed signal approach is the most common, where analogue circuitry is used to implement synaptic and neuronal dynamics, with inter-neuron communication handled by digital circuitry. The large number of synapses/neurons desired in neuromorphic systems means that the silicon area and energy required by the analogue components is typically far greater than the digital and that improvements to the efficiency of synapse/neuron implementations can yield significant savings.

3. Charge Transfer Synapse

The authors have previously reported an excitatory charge transfer synapse (CTS) capable of implementing synaptic depression and producing biological plausible post synaptic potentials (PSPs) [19]. Although the proposed LI circuit is compatible with other synaptic implementations, the CTS uses a single puff of “weighted” charge to map a spike input to a postsynaptic response, as opposed to other synaptic implementations which are circuit based and use either voltage or current [9–13]. For this reason the CTS is extremely compact, energy efficient and consequently scalable in readily available CMOS technology. The CTS (Figure 1) consists of three MOS capacitors (M1, M2 and M3) in series, a current mirror integrator (M4, M5) and a ‘leakage’ transistor operating in subthreshold mode (M6): M4, M5 and M6 therefore forms the proposed Leaky Integrator (LI) circuit. The synaptic weight is set by V_w and reflected in the magnitude of the charge package stored in the channel of M2. Once a presynaptic spike occurs at the V_{PRES} input the weighted charge packet, Q_w , is transferred to node $V_{IN}(t)$, by charge sharing between M2 and the capacitance $C_{IN}(n)$ where n is the number of synapses (fan-in) summed at the input to the LI circuit. The charge packet in the channel of M2 is subsequently replenished by M1, at a rate determined by V_p . M1 will usually be biased in sub-threshold and effectively controls the recovery of the weight charge in the channel of M2. The sudden drop in $V_{IN}(t)$, due to the spike at V_{PRES} , turns on M5 thus charging the output node $V_{PSP}(t)$. However, this voltage is simultaneously discharged by the leakage current flowing in M6, where V_{LEAK} controls the rate of leakage and allows for a tunable decay time. In biological terms $V_{PSP}(t)$ is the post-synaptic potential (PSP) of the neuron. A more comprehensive analysis of the CTS synapses can be found in [19]. In this paper we investigate temporal/spatial summation properties of the LI circuit. Additionally, a model for the fan-in, n , of the LI circuit is developed and compared to simulation and experimental results.

4. Fan-in Model

In this section, a model is presented which shows the dependency of $V_{PSP}(t)$ on the fan-in, n and the weight voltage, V_W . The model is validated by comparison with Cadence simulation and experimental results on fabricated circuits. Transistors M4 and M5 are assumed to be operating below threshold at all times and under the condition that the drain-substrate voltage, V_{DS} is greater than about three thermal volts (~ 75 mV) such that V_{DS} -dependence can be ignored.

Consider n CTS connected in parallel (Figure 1), where node $V_{IN}(t)$ is the summing node and is common to all n CTS. The output currents from all n CTS are summed at $V_{IN}(t)$ and the resulting current is mirrored in M5, charging the $V_{PSP}(t)$ node. Under quiescent conditions, M6 pulls down the V_{PSP} node to $\sim 0V$. However, with one or more synapses active the V_{PSP} node will be charged by M5. The node capacitance C_{PSP} is made up of the capacitance of the n^+ drain regions of M5 and M6, and parasitic capacitances associated with the layout. Any voltage dependencies of C_{PSP} are neglected for this analysis. An estimated value for C_{PSP} can be found analytically [19] and is approximately 3fF.

Prior to the application of the presynaptic pulse V_{PRES} , $V_{IN}(t)$ will be less than V_{DD} by a constant DC offset voltage, V_{OS} , due to the need for M4 to supply leakage current, I_L , to all n synapses. Thus:

$$V_{OS} = m_p V_t \ln \left(n \frac{I_L}{I_{Op}} \right) \quad (1)$$

where I_{Op} is the pre-exponential constant of M4, m_p is the subthreshold slope factor and V_t is the thermal voltage. Values for all parameters are listed in Appendix 2. When a presynaptic spike arrives at V_{PRES} , the charge stored in the channel of M2, due to the weight voltage V_W , is transferred to the $V_{IN}(t)$ node and reduces the voltage by an amount:

$$\Delta V_{IN}(t) = \frac{Q_W}{C_{IN}(n)} = \frac{C_{OX}(V_W - V_{Tn})}{C_{IN}(n)} \quad (2)$$

where $C_{IN}(n)$ is the associated diffusion capacitance given by [19]:

$$C_{IN}(n) = nC_{INeq} + 2C_{gs} + C_{int1} + C_{int2}(n-1) \quad (3)$$

C_{INeq} is the diffusion capacitance of the reverse biased n^+/p region at the drain of M3, $2C_{gs}$ is the gate capacitances of M4 and M5 and C_{int1} and C_{int2} are the parasitic capacitances associated with the metal interconnects. The total gate-source voltage of M5, labelled $V_{IN}'(0)$ ($= V_{DD} - V_{IN}(0)$), can be expressed as:

$$V_{IN}'(0) = V_{OS} + \Delta V_{IN}(0) \quad (4)$$

The voltage V_{OS} and hence $V_{IN}'(0)$ will increase with n . However, even for large values of n the maximum value of $V_{IN}'(0)$ remains well below the threshold voltage of M4/M5 ($V_{IN}'(0) = 0.6V$ for $n = 10000$, $V_W = 3V$), justifying the previous claim that they operate in subthreshold at all times. Following an input spike at V_{PRES} , the $V_{IN}(t)$ node will relax back to the quiescent value, V_{OS} , as it is charged by current through M4. The time dependence of $V_{IN}'(t)$ can be derived as follows:

$$C_{IN} \frac{dV_{IN}'(t)}{dt} = -I_{Op} \exp \left(\frac{V_{IN}'(t)}{m_p V_t} \right) \quad (5)$$

Note that the voltage across C_{IN} is $V_{IN}(t)$ so the variable in the LHS of (5) has been changed according to $dV_{IN}(t)/dt = -dV_{IN}'(t)/dt$. By integrating (5), rearranging for the time, t , and substituting appropriate values for V_{IN}' , an expression for the rise time of V_{PSP} , τ_R , is found:

$$\tau_R = \left[\frac{I_{Op}}{m_p V_t C_{IN}} \right]^{-1} \left\{ \frac{I_{On}}{I_{Op}} \exp \left(-\frac{V_{LEAK}}{m_n V_t} \right) - \exp \left(-\frac{V_{IN}'(0)}{m_p V_t} \right) \right\} \quad (6)$$

where I_O and m are the pre-exponential constant and slope factor of M4/M5/M6, V_t is the thermal voltage, C_{IN} is the capacitance at the $V_{IN}(t)$ node and $V_{IN}'(0)$ is the maximum gate source voltage of M4/M5, immediately after the CTS has fired. A full derivation is included in Appendix 1.

Now consider the time response of the output node voltage $V_{PSP}(t)$. The voltage $V_{PSP}(t)$ will rise rapidly as C_{PSP} is charged by the current delivered by M5 which is controlled by $V_{IN}'(t)$. In addition, M6, controlled by V_{LEAK} , acts to slowly discharge C_{PSP} . From Kirchoff's current law, C_{PSP} will charge according to:

$$C_{PSP} \frac{dV_{PSP}(t)}{dt} = I_{Op} \exp\left(\frac{V_{IN}(t)}{m_p V_t}\right) - I_{M6} \quad (7)$$

Integrating (7) and solving for $t = \tau_R$ gives the maximum value of V_{PSP} (full derivation in Appendix 1):

$$V_{PSPMAX} = m_p V_t \frac{C_{IN}}{C_{PSP}} \ln \left[1 + \frac{I_{Op}}{m_p V_t C_{IN}} \exp\left(\frac{V_{IN}(0)}{m_p V_t}\right) \tau_R \right] - \frac{I_{M6}}{C_{PSP}} \tau_R \quad (8)$$

Counter-intuitively, V_{PSPMAX} will increase with the fan-in, n , due to the influence of V_{OS} (1), which increases as more synapses are connected, causing the DC level at the $V_{IN}(t)$ node to decrease and so increasing the initial gate-source voltage of M4.

5. Experimental Setup

In order to confirm the correct operation of the CTS and to validate the developed model, a range of simulation and experimental measurements were made to measure the dependence of V_{PSPMAX} and τ_R on both the level of fan-in, n , and the CTS weight voltage, V_W . Several leaky integrator (LI) circuits (Figure 1) were fabricated in a 0.35um CMOS process from AMS, with $n = 1, 5, 9$ and 20. All transistors were sized with an aspect ratio (W/L) of 1. Additional test structures (MOS capacitors and MOSFETs with dimensions of $50\mu\text{m} \times 100\mu\text{m}$, $100\mu\text{m} \times 100\mu\text{m}$ and $200\mu\text{m} \times 100\mu\text{m}$) were included on the chip to allow for device characterization and extraction of key process parameters (oxide thickness, doping levels etc.).

A custom PCB was fabricated to house the packaged IC, with BNC connections for each input/control voltage. A 4-terminal DC voltage source was used to provide V_{DD} and the CTS control voltages (V_W , V_P , and V_{LEAK}); a signal generator provided pre-synaptic input pulses at V_{PRES} (3.3V amplitude, 1 ns rise time, 5 ns duration) and an Agilent oscilloscope (MSOX2024A) was used to capture output waveforms from the $V_{PSP}(t)$ node together with the exact timings of the presynaptic inputs at V_{PRES} . To avoid excessive capacitive loading of the V_{PSP} node during measurement, an on-chip source-follower voltage buffer was realised at the V_{PSP} node, from which output recordings were made. By adjusting the values of V_W , V_{LEAK} and the inter spike interval (ISI) between input spikes at the V_{PRES} node, PSPs with different shapes and characteristics were recorded at that node.

6. Results

To confirm that the circuit produces sufficiently realistic PSPs, the voltage $V_{PSP}(t)$ was recorded in response to a series of spike trains with different inter spike intervals, (Figure 2) where each train consist of a five spike burst with ISI settings of (a) 1.5ms, (b) 1ms, (c) 500us and (d) 100us. The measured waveforms show a LI profile which is characteristic of biological synapses where the magnitude of $V_{PSP}(t)$ increases with decreasing ISIs and subsequently decays at a significantly slower rate dictated by V_{LEAK} . It should also be noted from Figure 2 that the rise time of a postsynaptic response and its amplitude is determined by the level of activity of the presynaptic neurons, which is again observed in real neurons [20]. The effect on the PSP of varying V_W and V_{LEAK} are illustrated in Figure 3, where additional experimental results, alongside simulated values (Cadence, spectreS) are presented. In (a), three pre- synaptic inputs were applied at $75\mu\text{s}$ intervals, with $\Delta V_{PSP}(t)$ equal to several hundred mVs and a fall time, set by V_{LEAK} , of $140\mu\text{s}$. In (b), 20 input spikes were applied at $10\mu\text{s}$ intervals, with lower values of V_W and V_{LEAK} used. The value of $\Delta V_{PSP}(t)$ is reduced to tens of mV and the fall time increases to 0.5ms.

In order to validate the predictions made for the values of V_{PSPMAX} and τ_R , further simulations of the CTS were carried out, looking specifically at the V_{IN} and V_{PSP} nodes. Figure 4 shows simulated V_{IN} and V_{PSP} waveforms as with n as a parameter, in response to a single synapse firing. As predicted by (1), the resting potential of V_{IN} increases with n . This is also reflected in an increased resting potential at V_{PSP} . This is not detrimental to the operation of the LI circuit as this offset, and indeed offsets due to process variations across a wafer, can be absorbed into the post-trained weight values. These results also justify the assumption that M4/M5 remain in subthreshold at all times. With $V_{DD} = 3.3\text{V}$, the maximum V_{GS} is 0.45V which is well below the 0.7V threshold. Figure 5 shows V_{PSPMAX} and τ_R extracted from these simulated results and compared to the predictions made by the theoretical model for V_{PSPMAX} (4) and τ_R (1): process parameters used in the model are given in Appendix 2. The values of V_{LEAK} and V_W were adjusted to give the best fit to the simulated data for both V_{PSPMAX} and τ_R . The minimum/maximum error between modelled and simulated values of τ_R and V_{PSPMAX} are 80ns/34 μs and 21mV/23mV respectively. Finally, modelled values for V_{PSPMAX} were compared to those obtained from measurement on the fabricated IC. Figure 7 shows experimental values for V_{PSPMAX} against V_W , for increasing values of n , alongside values predicted by (4). The experimental results confirm the prediction that V_{PSPMAX} will increase as n becomes larger and demonstrates a linear relationship with V_W .

7. Discussion

A theoretical analysis and experimental results, showing the effects of increasing the fan-in, n , have been presented for the proposed LI circuit. If the assumption that M4/M5 will operate in subthreshold at all times is to hold, then by substituting (1) into (4) and solving for $V_{IN}' = V_T$, a theoretical maximum value for n can be found. This analysis yields a value of n greater than 10^5 . A more practical approach is to consider the design trade-off between the total number of synapses and the amount of presynaptic activity required to trigger a postsynaptic spike. As n is increased, the contribution of a single spike to the value of V_{PSP} also increases. As an example, a block with 5 synapses and $V_w = 1V$ would require ~ 60 spikes in quick succession to reach a 1.5V postsynaptic spike threshold voltage, assuming a standard CMOS inverter output stage, while a block with 20 synapses would require half the amount (30), for the same effect. To maximize the headroom ($V_{OS} \rightarrow V_T$), a limit needs to be imposed on n to maintain V_{OS} small and therefore we envisage an architecture where in each layer the number of inputs would be shared equally across several LI circuit blocks each with n inputs. The output current from each LI block would then be summed into one leak transistor to set the PSP decay duration. It should be noted, that there is a practical limitation to the maximum fan-in, set by the density of interconnect which needs to be considered at layout. While the analysis presented here is based on the CTS [19], the fan-in model itself and the conclusions drawn from it, are applicable to other synapse implementations that have been reported in the literature. The ability to have a programmable postsynaptic response, using V_{Leak} , gives the LI circuit the capability of capturing the different receptor types which influences the duration of the postsynaptic responses. For example, with ionotropic receptors, which are ion channels with a binding site for the respective transmitter (ligand gated), the postsynaptic response lasts only a few milliseconds while with metabotropic receptors, which involves secondary messengers within the postsynaptic cell, the response time can be in the hundreds of milliseconds [20]. Furthermore, we have demonstrated above an excitatory postsynaptic responses (EPSPs) where the change in the membrane voltage is de-polarising following the influx of positively charged ions, typically sodium, into a neuron cell. Conversely, hyperpolarization of the cell membrane also occurs due to inhibitory synapses resulting in an inhibitory postsynaptic potentials (IPSPs) which result from the influx of negative ions, typically GABA, or Cl⁻. With the addition of another current mirror circuit to sink current, and a negative supply rail, the proposed LI block can effectively realise an IPSP response and hence since both types of postsynaptic potential responses are graded, the LI could therefore sum both EPSP and IPSP together to have a cumulative excitatory or inhibitory effect.

8. Conclusions

A theoretical model has been developed to explore the effects of increasing fan-in on the operation of a proposed LI circuit, based around a charge transfer synapse. The model relates both the magnitude of the postsynaptic response and its transient behaviour to fan-in. Results from test circuits fabricated in a 0.35 μ m CMOS process and simulations demonstrate the operation of the circuit and validate the model against accepted biophysical behavior. An upper limit can be predicted for the maximum fan-in level for which the circuit will still function as intended, $n = 10^5$. Given the magnitude of this number, the real limitation on the fan-in will be dictated by practical considerations such as interconnect density and silicon area. Based on this, the authors estimate that a realistic limit is of the order of tens of synapse per neuron cell.

Acknowledgements

The authors would like to acknowledge Dr Shou Huang for obtaining some of the experimental results used in this work.

Appendix I – Derivation of Model Equations

AI-I derivation of rise time

Following the input spike, the V_{IN} node will relax back to the quiescent value, V_{OS} , as it is charged by current through M4. The time dependence of V_{IN}' can be derived:

$$C_{IN} \frac{dV_{IN}'(t)}{dt} = -I_{Op} \exp\left(\frac{V_{IN}'(t)}{m_p V_t}\right) \quad (A-1)$$

The V_{DS} dependence of I_{M4} can be neglected, as $V_{DS} > 50mV$ at all times. Integrating with respect to t :

$$\int_{V_{IN}'(0)}^{V_{IN}'(t)} \exp\left(-\frac{V_{IN}'(t)}{m_p V_t}\right) dV_{IN}'(t) = -\frac{I_{Op}}{C_{IN}} \int_0^t dt \quad (A-2)$$

$$-m_p V_t \exp\left(-\frac{V_{IN}'(t)}{m_p V_t}\right) \bigg|_{V_{IN}'(0)}^{V_{IN}'(t)} = -\frac{I_{Op}}{C_{IN}} t \quad (\text{A-3})$$

Arranging for $V_{IN}'(t)$:

$$\left[\exp\left(-\frac{V_{IN}'(t)}{m_p V_t}\right) - \exp\left(-\frac{V_{IN}'(0)}{m_p V_t}\right) \right] = \frac{I_{Op}}{m_p V_t C_{IN}} t \quad (\text{A-4})$$

$$\exp\left(\frac{V_{IN}'(t)}{m_p V_t}\right) = \left[\exp\left(-\frac{V_{IN}'(0)}{m_p V_t}\right) + \frac{I_{Op}}{m_p V_t C_{IN}} t \right]^{-1} \quad (\text{A-5})$$

rearranging (A-5) for time:

$$t = \left[\frac{I_{Op}}{m_p V_t C_{IN}} \right]^{-1} \left\{ \exp\left(-\frac{V_{IN}'(t)}{m_p V_t}\right) - \exp\left(-\frac{V_{IN}'(0)}{m_p V_t}\right) \right\} \quad (\text{A-6})$$

Following an input spike, the voltage at the output node, $V_{PSP}(t)$, will increase so long as $I_{M5} > I_{M6}$, after which $V_{PSP}(t)$ will be discharged through I_{M6} . By equating the currents at the equilibrium condition ($I_{M5} = I_{M6}$) and calculating the corresponding value of $V_{IN}'(t)$, it is possible to estimate the value of the rise time:

$$I_{Op} \exp\left(\frac{V_{IN}'(t)}{m_p V_t}\right) = I_{On} \exp\left(\frac{V_{LEAK}}{m_n V_t}\right) \quad (\text{A-7})$$

$$V_{IN}' = m_p \left[\frac{V_{LEAK}}{m_n} - V_t \ln\left(\frac{I_{On}}{I_{Op}}\right) \right] \quad (\text{A-8})$$

Substituting (A-8) into (A-6) and simplifying gives a value for the rise time, τ_R :

$$\tau_R = \left[\frac{I_{Op}}{m_p V_t C_{IN}} \right]^{-1} \left\{ \frac{I_{On}}{I_{Op}} \exp\left(-\frac{V_{LEAK}}{m_n V_t}\right) - \exp\left(-\frac{V_{IN}'(0)}{m_p V_t}\right) \right\} \quad (\text{A-9})$$

AI-II derivation of V_{PSPMAX}

$$C_{PSP} \frac{dV_{PSP}(t)}{dt} = I_{Op} \exp\left(\frac{V_{IN}'(t)}{m_p V_t}\right) - I_{M6} \quad (\text{A-10})$$

where $V_{PSP}(t)$ is considered to be > 50 mV, allowing the subthreshold current in M5 to be written without consideration of the V_{DS} dependence. Substituting (A-5) into (A-10) and separating variables:

$$\frac{dV_{PSP}(t)}{dt} = \frac{I_{Op}}{C_{PSP}} \left[\exp\left(-\frac{V_{IN}'(0)}{m_p V_t}\right) + \frac{I_{Op}}{m_p V_t C_{IN}} t \right]^{-1} \quad (\text{A-10})$$

$$\int_0^{V_{PSP}(t)} dV_{PSP}(t) = \frac{I_{Op}}{C_{PSP}} \int_0^t \left[\exp\left(-\frac{V_{IN}'(0)}{m_p V_t}\right) + \frac{I_{Op}}{m_p V_t C_{IN}} t \right]^{-1} - \frac{I_{M6}}{I_{Op}} dt \quad (\text{A-11})$$

Integrating:

$$V_{PSP}(t) = \frac{I_{Op}}{C_{PSP}} \left\{ m_p V_t \frac{C_{IN}}{I_{Op}} \ln \left(\exp \left(-\frac{V_{IN}'(0)}{m_p V_t} \right) + \frac{I_{Op}}{m_p V_t C_{IN}} t \right) - \frac{I_{M6}}{I_{Op}} t \right\} \quad (A-12)$$

$$= \frac{I_{Op}}{C_{PSP}} \left\{ m_p V_t \frac{C_{IN}}{I_{Op}} \ln \left(\exp \left(-\frac{V_{IN}'(0)}{m_p V_t} \right) + \frac{I_{Op}}{m_p V_t C_{IN}} t \right) - \frac{I_{M6}}{I_{Op}} t \right\} \quad (A-13)$$

$$- \frac{I_{Op}}{C_{PSP}} \left\{ m_p V_t \frac{C_{IN}}{I_{Op}} \ln \left(\exp \left(-\frac{V_{IN}'(0)}{m_p V_t} \right) \right) \right\}$$

Simplifying:

$$V_{PSP}(t) = \frac{I_{Op}}{C_{PSP}} \left\{ m_p V_t \frac{C_{IN}}{I_{Op}} \ln \left(\frac{\exp \left(-\frac{V_{IN}'(0)}{m_p V_t} \right) + \frac{I_{Op}}{m_p V_t C_{IN}} t}{\exp \left(-\frac{V_{IN}'(0)}{m_p V_t} \right)} \right) - \frac{I_{M6}}{I_{Op}} t \right\} \quad (A-14)$$

$$= \frac{I_{Op}}{C_{PSP}} \left\{ m_p V_t \frac{C_{IN}}{I_{Op}} \ln \left(1 + \frac{I_{Op}}{m_p V_t C_{IN}} \exp \left(\frac{V_{IN}'(0)}{m_p V_t} \right) t \right) - \frac{I_{M6}}{I_{Op}} t \right\} \quad (A-15)$$

$$= m_p V_t \frac{C_{IN}}{C_{PSP}} \ln \left[1 + \frac{I_{Op}}{m_p V_t C_{IN}} \exp \left(\frac{V_{IN}'(0)}{m_p V_t} \right) t \right] - \frac{I_{M6}}{C_{PSP}} t \quad (A-16)$$

To find the maximum value, V_{PSPMAX} , solve (A-16) for $t = \tau_R$ (A-9).

Appendix II – Model Parameters

Values for the model parameters were extracted from measurements made on large area ($\sim 100 \mu\text{m} \times 100 \mu\text{m}$) test structures fabricated on the same IC as the CTS and neuron circuits. All values fell within the expected range of values for the AMS 0.35 μm fabrication process used. A number of capacitance parameters were estimated empirically (for experimental results) or by simulation, these are indicated by *.

	Experimental	Model
m_p	1.71	1.71
m_n	1.73	1.73
t_{ox}	15 nm	15 nm
V_{Tn}	0.7 V	0.7 V
I_{Op}	9 fA	9 fA
I_{On}	11 fA	11 fA
I_L	10 pA	10 pA
C_{PSP}^*	3 fF	3 fF
C_{int1}^*	2.77 fF	0.1 fF
C_{int2}^*	0.2 fF	0.1 fF

Table A-1. Model parameters used. For capacitances, different values were used when comparing the model with either simulated or experimental results. All other model parameters remained the same.

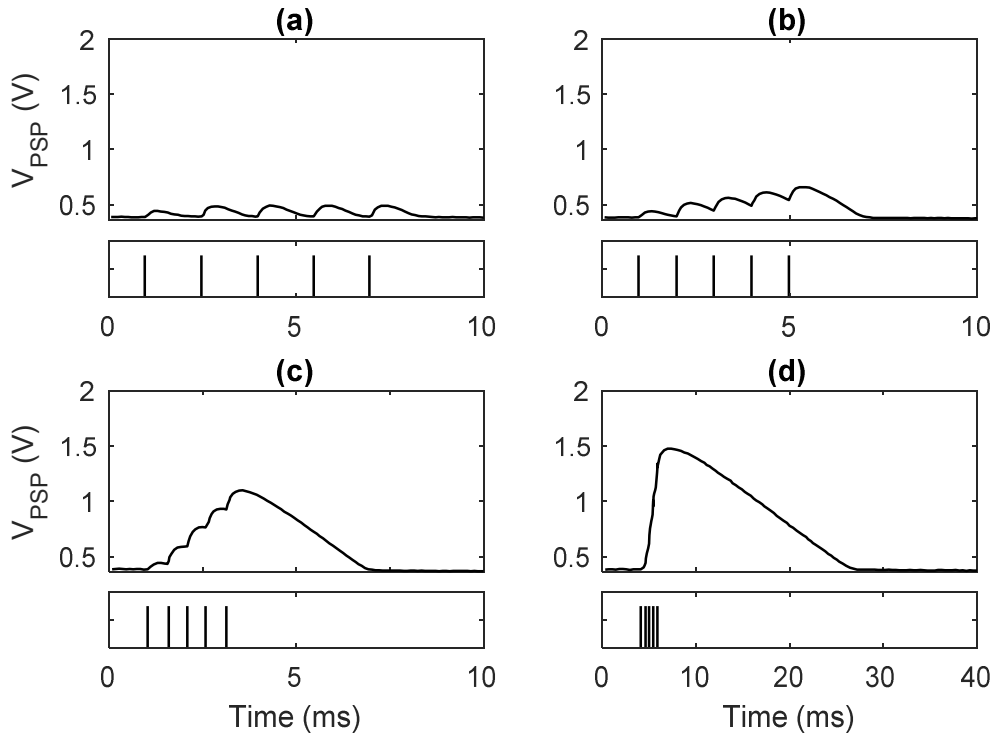


Figure 2. Circuit measurement results of typical synapse response to sequences of spikes with different ISI settings, (a) $ISI=1.5$ ms, (b) $ISI=1.0$ ms, (c) $ISI=500$ μ s and (d) $ISI=100$ μ s. Lower plots show the spike firing time. On the arrival of each input spike, a discrete amount of charged is transferred to the V_{PSP} node, raising the voltage. Over time, the charge at V_{PSP} decays through M6. As the ISI is decreased, additional synaptic charge is delivered before the previous charge packet has decayed, increasing the maximum V_{PSP} voltage.

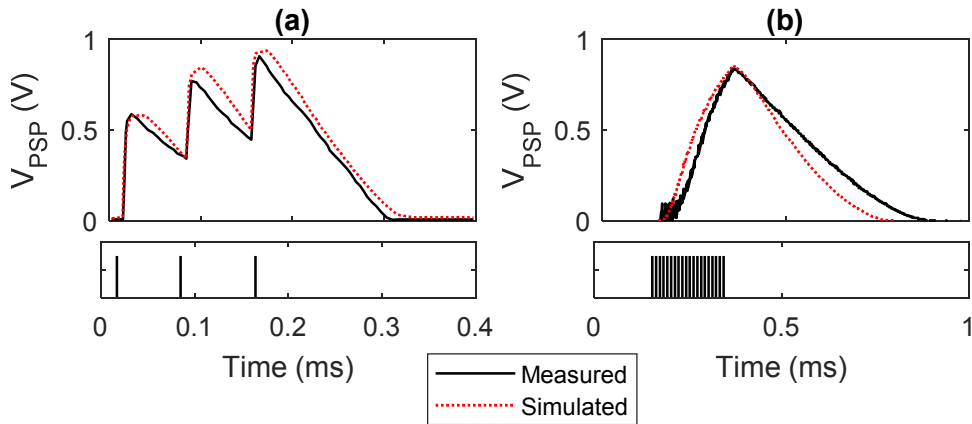


Figure 3. Comparison of measured and simulated PSPs. (a) PSP in response to 3 inputs, $ISI = 75$ μ s. Experimental values: $V_W = 0.95$ V, $V_{LEAK} = 0.36$ V. Simulation values: $V_W = 0.89$ V, $V_{LEAK} = 0.3$ V. V_{PSP} Fall time = 140 μ s. (b) PSP in response to 20 inputs, $ISI = 10$ μ s. Experimental values: $V_W = 0.9$ V, $V_{LEAK} = 0.3$ V. Simulation values: $V_W = 0.84$ V, $V_{LEAK} = 0.26$ V. V_{PSP} Fall time = 500 μ s.

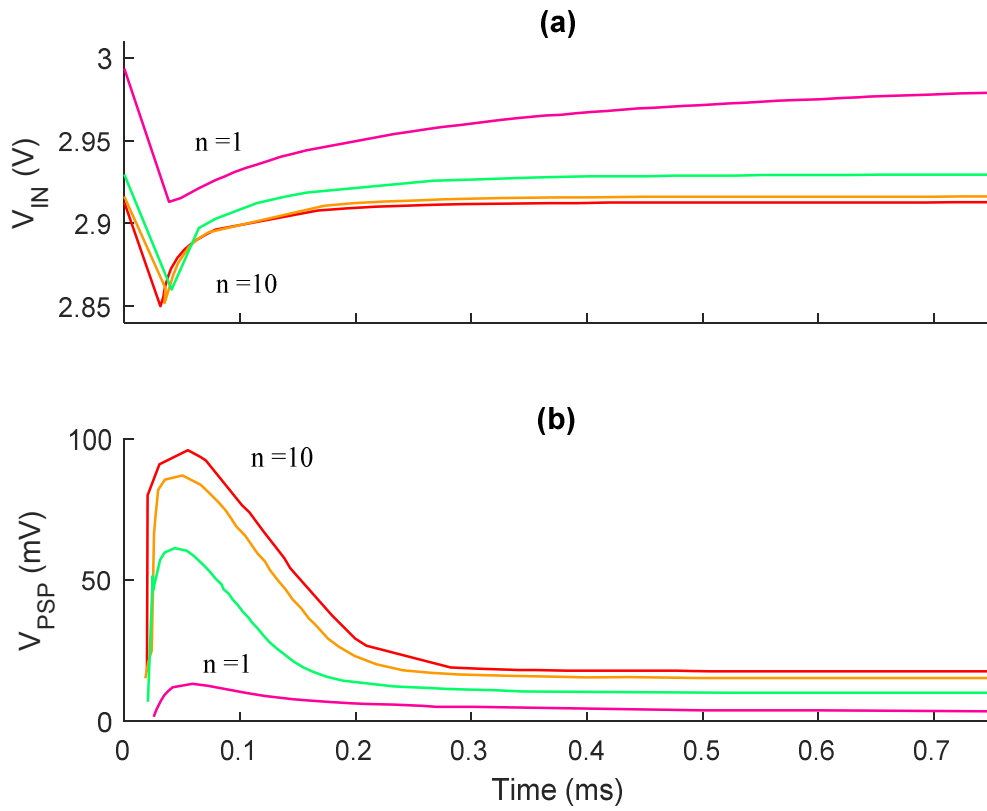


Figure 4. Simulation results showing (a) $V_{IN}(t)$ and (b) $V_{PSP}(t)$ in response to a single synaptic input arriving. Results shown for $n = 1, 2, 5, 10$. As n is increased, the resting potential at V_{IN} decreases, as predicted by (1), with a subsequent increase in the value of V_{PSPMAX} .

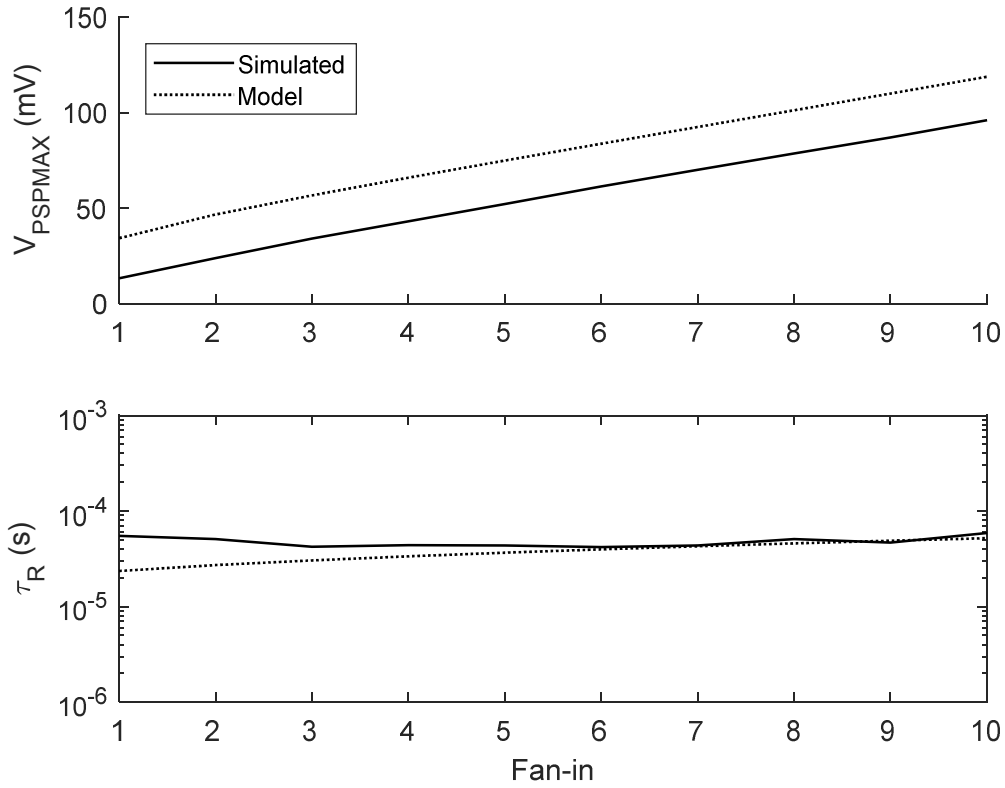


Figure 5. Simulated and modelled values for V_{PSPMAX} (4) and τ_R (1). Simulation parameters ($V_{LEAK} = 0.1$ V, $V_W = 1$ V). Model parameters ($V_{LEAK} = 0.22$ V,

$V_W = 0.8$ V). Process parameters given in Appendix II were used for the modelling equations. The minimum/maximum error between modelled and simulated values of τ_R and V_{PSPMAX} are 80 ns/34 μ s and 21 mV/23 mV respectively.

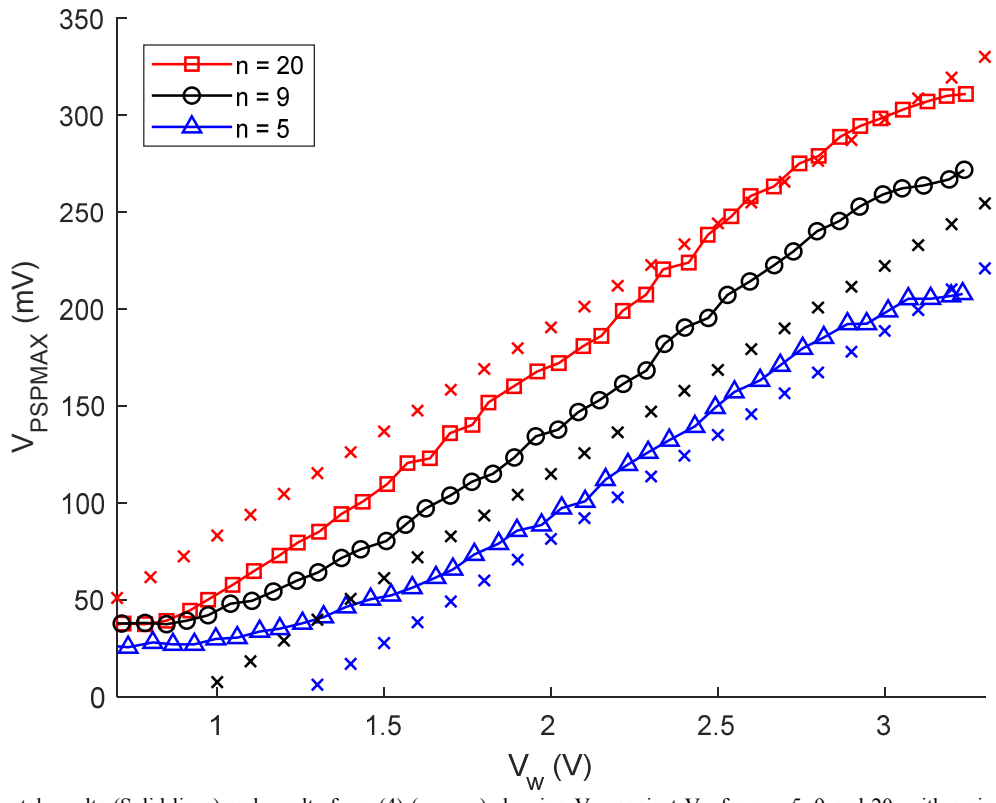


Figure 6. Experimental results (Solid lines) and results from (4) (crosses) showing V_{psp} against V_w for $n = 5, 9$ and 20 , with a single active synapse. Model parameters ($V_{LEAK} = 0.42$ V). Process parameters given in Appendix II were used for the modelling equations. The experimental results confirm the prediction that V_{PSPMAX} will increase as n becomes larger and demonstrates a linear relationship with V_w .