Modeling Memory Consolidation (or Learning) in Long-Term Memory and Memory Recapitulation in Working Memory Through Spike-Time-Dependent-Plasticity-Enabled Spiking Neural Networks

Thesis submitted by

Amod Holla
2020EE10639

under the guidance of

Prof. Sumeet Agarwal, Indian Institute of Technology, Delhi
Prof. Debanjan Bhowmik, Indian Institute of Technology, Bombay

in partial fulfilment of the requirements
for the award of the degree of

Bachelor of Technology

Department Of Electrical Engineering
INDIAN INSTITUTE OF TECHNOLOGY DELHI

May 2024

THESIS CERTIFICATE

This is to certify that the project report titled, "Modeling Memory Consolidation (or Learning) in Long-Term Memory and Memory Recapitulation in Working Memory Through Spike-Time-Dependent-Plasticity-Enabled Spiking Neural Networks", submitted by Amod Holla (2020EE10639), to the Indian Institute of Technology, Delhi, for the partial fulfillment of the requirement of the Bachelor of Technology (B.Tech) degree, is a bona fide record of the work done by him under our supervision. We are fully aware of the contents of this report, its technical details and ethical integrity, and the report has not been submitted to any other Institute or University for the award of any degree or diploma.

Prof. Sumeet Agarwal
Professor
Dept. of Electrical Engineering
IIT-Delhi, 110016

Place: New Delhi

.....
Date: 5th May 2024

Prof. Debanjan Bhowmik
Professor
Dept. of Electrical Engineering
IIT-Bombay, 400076

Place: Mumbai

.....
Date: 5th May 2024

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisors Prof. Sumeet Agarwal and Prof. Debanjan Bhowmik for the opportunity to work on this project and also for the constant guidance and support, which enabled me to enjoy my work while learning a lot as well. In addition, I also thank the IIT Delhi HPC facility for computational resources.

ABSTRACT

KEYWORDS: Neuromorphic Computing; Crossbar Arrays; Long Term Memory; Working Memory; Memory Consolidation; Spiking Neural Network; Spike Time Dependent Plasticity

Memory consolidation in long-term memory (LTM) and recapitulation of the same memories in working memory (WM) (along with interference between those recapitulated memories and simultaneously occurring inputs from sensory organs) are important concepts in cognitive psychology and cognitive/ behavioural neuroscience. In this paper, we propose a model for these higher order mental functions using a network of neurons and synapses (biologically plausible spiking neural network (SNN)), where the models of neurons and synapse are at a single-nerve-cell level: leaky integrate fire (LIF) neuron, tempotron model of signal propagation in a synapse, and spike time dependent plasticity (STDP) of a synapse. We restrict ourselves to visual information (images of hand-written digits) for this purpose in the paper and hence model, using SNN, how the shapes of the digits are retained in the LTM (memory consolidation) while the human learns to identify/ classify the digit. Then we model how these digit impressions are recapitulated in the WM, which is connected to the LTM through synapses following the tempotron model. We model this both in the absence and presence of interfering visual inputs corresponding to other digits, incoming from sensory memory. We show that in the presence of interfering inputs, the impressions of recapitulated digit and that of the digit seen at the sensory organ interfere in the WM, as hypothesized in some seminal works on WM before. In addition, we perform a circuit-level design of the entire system and verify by simulation the functioning of the LIF neuron circuit, the tempotron synapse circuit, and crossbar arrays that represent the fully-connected and one-to-one connections that are present in our design, thus opening up the possibility of large scale SNNs being implemented efficiently on neuromorphic hardware.

The formatting is as (as far as the author is aware) per the current institute guidelines.

Contents

List of Figures

Chapter 1

INTRODUCTION

1.1 Long Term Memory, Working Memory, Memory Consolidation

Different aspects of human memory form an important subject of study in cognitive psychology and in cognitive and behavioural neuroscience. According to the modal model of memory, human memory can be partitioned into three parts: sensory memory, short-term memory, and long-term memory (LTM) [1]. Over years of study, the concept of short-term memory has evolved into that of working memory (WM), with WM further consisting of multiple parts: visuo-spatial sketchpad, phonological loop, and central executive [2].

When information is rehearsed again and again in short-term memory, then it gets stored over a long term or even permanently (until death) in LTM, and this can be considered memory consolidation or learning. Stored information from the LTM can be retrieved into short-term/ WM, which can be called memory recapitulation. But given that WM has limited capacity, what's recapitulated and retrieved temporarily into WM can interfere with what's incoming from the senses and passing through sensory memory [2, 3, 4, 5]. Typically, these concepts are explained in cognitive and behavioural neuroscience through a top-down approach involving monitoring and explanation of brain activities in large areas of the brain containing a huge number of neurons. For example, memory consolidation is explained through an interaction between the hippocampus and the neo-cortex of the brain [6].

1.2 Leaky Integrate Fire (LIF) Neuron, Tempotron Model of Synapse, Spike Time Dependent Plasticity (STDP) of Synapse

In the bottom-up view of neuroscience, models are instead formulated of isolated nerve cells/ neurons and networks of few neurons interconnected through synapses [7]. Leaky integrate fire (LIF) model of neuron, for example, is a simplistic model that determines the time of neuron spike under the influence of input current without going into the mechanism of the spike (more complicated models of the neuron like Hodgkin Huxley model explain that). Tempotron model of the synapse takes care of how signal propagates from the pre-neuron to the post-neuron through the synapse [8], while spike time dependent plasticity (STDP)

model of the synapse represents how the synaptic strength changes as a function of time difference between spikes occuring at the pre-neuron and the post-neuron [9].

In a seminal paper, Diehl and Cook connected STDP-enabled SNN with a higher level mental function: long-term learning of classification of handwritten digits [10]. They also showed that in this scheme, because of STDP property, the weights of the synapses connecting a given output neuron with the input neurons carry the impression of the shape of the digit which that particular neuron has been trained to identify.

1.3 Our Contributions

Most of the SNN-related works that have been reported following the seminal paper by Diehl and Cook [10] have focused on improving the SNN training algorithm to obtain high classification accuracy on more complicated data sets [12, 13] or on implementation of the proposed SNN on novel hardware using emerging non-volatile devices, also known as neuromorphic hardware [14, 15, 16]. But to the best of our knowledge, relatively much less work has been done on modeling other aspects of human memory (like the interactions between sensory memory and WM, or WM and LTM) using such SNNs, with the exception of some papers on SNN-based modeling of formation of polychronous patterns in working memory [17] and some on SNN-based modeling of formation of associations between different patterns fed to the brain [7, 18].

In this paper, we make the following contributions in that context (section-by-section layout of the paper is also presented in the process). In Chapter 2, we discuss our biologically plausible SNN design: we essentially model a combination of sensory memory, WM, and LTM units through multiple spiking neuron layers connected through synapses that follow the tempotron model [8]. Additionally, synapses in the LTM unit have STDP property [9].

In Chapter 3, we show how incoming images of digits through the sensory neurons/ memory eventually lead to learning of classification task of the digits at the LTM: memory consolidation. Synaptic weights of the SNN thereby form the impressions of the digits, much like in the work by Diehl and Cook [10].

In Chapter 4, we first show how long the signals are retained in the WM when the inputs to the sensory memory are turned off. The time dynamics in the tempotron model of the synapse is used for this purpose [8]. Next, we show how in the absence of incoming inputs from the sensory neurons/ memory, digit impressions already stored in the LTM are recapitulated in the WM. We also show that in the presence of interfering inputs, the impressions of recapitulated digit and that of the digit seen at the sensory organ interfere in the WM. This is the main novelty of our work here.

One of the peculiar aspects of recapitulation of images, text, etc. from LTM in the case of humans is that inputs from sensory organs are known to interfere with what are

recapitulated from LTM. Something similar is reported in the seminal work by Brooks [3], which Badeley also comments on in his book on WM [2]. According to Brooks and Badeley, visual information presented in real time through the senses of the subject (of the test) interfere with the spatial information visualized in the WM based on auditory or textual information (shapes and forms imagines based on speech or text) simultaneously presented to the subject [3, 2]. Baddeley proposes that this happens due to the limited storage capacity of the WM [2]. Here, we have taken inspiration from this finding but have modeled a simpler aspect of the brain's behaviour. Instead of studying the interference between spatial information presented through the senses to the brain and spatial information visualized/ reconstructed through text or speech presented to the brain, we have quite simply studied the interference between images of handwritten digits presented to the brain through the senses and images of some other handwritten digits recapitulated from the brain. We make this choice because it makes the design of the SNN to model the memory aspect simple; a SNN that converts textual information to visual/ spatial information (like imagining shapes from text) is far more complicated to design.

In addition to computational modeling of our SNN-WM network, there is also a need to design specialised neuromorphic hardware to make the SNN-WM network time and energy efficient. There is also the possibility of integrating our hardware-based SNN with new sensory neuromorphic modules, which makes our network multi-modal and biologically plausible. To this end, in Chapter 5, we design and simulate analog circuit designs for the LIF neurons and the tempotron circuits. Crossbar arrays of non-volatile memory (NVM) devices that contain the synaptic weights are designed for the two types of connections in our network - namely a fully-connected type (between the first and second layer of the long term memory), and the one-to-one type (sensory memory - working memory, working memory - long term memory). In chapter 6, we conclude the report.

Chapter 2

Working Memory Model based on Spiking Neural Networks

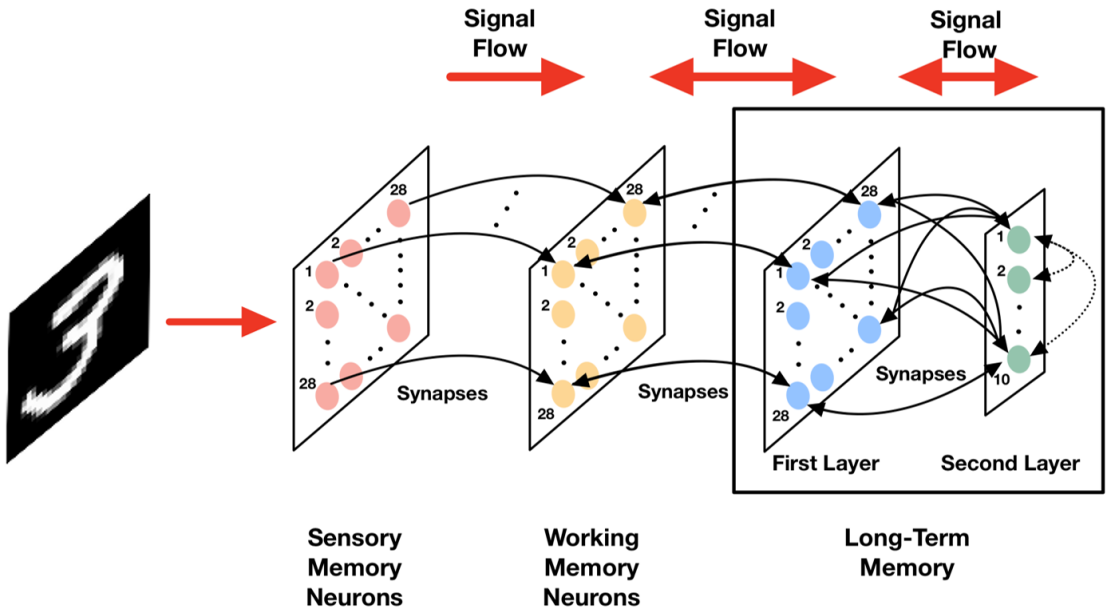

Figure 2.1: Schematic of the proposed SNN-based model of the memory system: 784 LIF input neurons act as the sensory memory, 784 LIF neurons serve as the WM, and a 2-layer STDP-enabled SNN (with 784 neurons in the first layer and 10 neurons in the second layer) serves as the LTM of the system.

Fig. 2.1 shows our proposed SNN-based model of the memory system. It consists of three functional parts: 784 input neurons that act as the sensory memory, 784 neurons that work as the WM, and a 2-layer SNN, with 784 pre-neurons and 10 post-neurons, which serves as the LTM of the system. The reason to choose 784 neurons is that we are working with the MNIST data set of handwritten digit images here, where each image is of the size 28×28 pixels [10].

All the neurons in the memory system follow the biologically motivated Leaky Integrate and Fire (LIF) model, where the membrane potential $v(t)$ of each neuron obeys the following equation:

$$C\frac{\partial v(t)}{\partial t} = -G_L(v(t) - E_L) + I(t) \tag{2.1}$$

Here, $I(t)$: input current to the neuron, G_L: membrane conductance, and E_L: resting membrane potential of a neuron. Once $v(t)$ reaches a chosen threshold potential (V_{th}), the neuron generates a spike, and $v(t)$ drops to E_L. We choose G_L = 30 nS, E_L = -70 mV, and

C = 500 fF for all the neurons in the design [14]. While V_{th} is fixed at $20mV$ for all the sensory neurons, WM neurons, and neurons of the first layer of LTM, V_{th} is a function of time for the neurons of the second layer of LTM. This is because we have considered that they exhibit the biologically plausible property of homeostasis [10]: every time a neuron spikes (say at time t_{spike}), V_{th} for that neuron increases by an amount $\Delta V_{th} = 7mV$ and then exponentially decays over time t, with a homeostasis time constant $\tau_{homeo} = 15\mu s$ as follows, until the neuron spikes again:

$$V_{th}(t) = V_{th}(t_{spike}) + \Delta V_{th} e^{-(\frac{t-t_{spike}}{\tau_{homeo}})}. \tag{2.2}$$

In our proposed model, sensory neurons are connected to the working-memory neurons through synapses such that the signal flows only one way, from sensory memory to WM. However, synapses that connect WM neurons with the first layer of LTM neurons and that connecting the first-layer of LTM neurons with the second layer of LTM neurons allow bidirectional signal flow (Fig. 2.1).

The biologically motivated tempotron model is used to capture this signal flow. If the neuron from which signal flows (numbered i) fires at a time $t_{spike,i}$, the receiving neuron, numbered j, receives a time-varying current given by the following equation:

$$I_{ji}(t) = I_0 w_{ji} (e^{(-\frac{t-t_{spike,i}}{\tau_M})} - e^{(-\frac{t-t_{spike,i}}{\tau_S}))} \tag{2.3}$$

, where $\tau_M = 10\mu s$ and $\tau_S = 2.5\mu s$ are the 2 time constants and w_{ji} is the weight stored in the synapse [8, 11, 14].

For the synapses connecting sensory neurons with WM neurons or that connecting WM neurons with LTM neurons, the weight values are fixed. However, the weight values connecting the first layer neurons (labelled i) in the LTM SNN with the second layer neurons (labelled j) are adjustable and they are updated using the previously mentioned STDP rule:

$$\Delta w_{ji} = \begin{cases} \Gamma_1 \left(1 - \frac{w}{w_{max}}\right)^{\mu} \exp\left(-\left(\frac{t_{spike,j} - t_{spike,i}}{\tau_1}\right)\right) \\ \qquad\qquad \text{if } t_{spike,j} \geq t_{spike,i} \\ -\Gamma_2 \left(\frac{w}{w_{max}}\right)^{\mu} \exp\left(-\left(\frac{t_{spike,i} - t_{spike,j}}{\tau_2}\right)\right) \\ \qquad\qquad \text{if } t_{spike,j} < t_{spike,i} \end{cases} \tag{2.4}$$

, where $\Gamma_1 = 9$, $\Gamma_2 = 15$, $\tau_1 = 10\mu s$, $\tau_2 = 20\mu s$, and $\mu = 1.7$ are constants associated with the above STDP equation [9, 10, 11, 14]. This learning rule depends only on the spike-times of the neuron of the first layer of LTM and of the neuron of the second layer of LTM, and thus it is very local in nature. The weight increases if the neuron of the first layer spikes before the neuron of the second layer ($t_{spike,j} \geq t_{spike,i}$), and is called synaptic potentiation. When the reverse occurs, the weight decreases and this is called synaptic depression.

The entire system in Fig. 2.1, including the neurons, tempotron synapses, the homeostasis property, and the STDP learning rule, have been modeled in our work using the high-level programming language Python, with the help of the Brian2 library [19].

Chapter 3

Training of Long Term Memory to Identify Handwritten Digits

The STDP-enabled SNN-based LTM has been trained here using different subsets of varying sizes from the MNIST main data set. The size of the data set varies from 1 image per class to 500 images per class. But for all the cases, testing/ inference has been carried out on the same randomly chosen set of 10 images per class. In order to prevent multiple neurons of the second layer of the LTM from spiking at the same time for a given image, they are connected to each other through synapses with constant negative weights. Thus is know as the winner take All (WTA) mechanism. Training is done in the partially supervised mode, as described in the report by Biswas *et al* [11] and Sahu *et al* [14]. In this mode, external bias currents are applied to the neurons of the second layer of LTM in order to allow their selective firing. When the input belongs to a particular digit/class, no inhibitory current is applied to the second layer neuron allotted for that digit (so that it fires), but an inhibitory current is applied to the other neurons of that layer (so that they don't fire).

For each training subset from the MNIST data set (with number of images per class varying from 1 to 500 as mentioned before), only one epoch of training is carried out, with all training samples shown to the network in that epoch. After that, for each training case, the neuron of second layer of LTM, corresponding to the digit '0', is fired, and the spike counts for each neuron of the first layer of the LTM are recorded. Then the number of spikes for each such neuron across the entire time period of this excitation (time $t = 0$ μs to $t = 110$ μs) is plotted in a 28×28 grid format in Fig. 3.1. The quality of the digit shape learning, as training set size is varied, can thus be studied.

For each training case in Fig. 3.1 (number of samples in the training sub-set is different for each case), train and test accuracy numbers are also recorded and are mentioned in Fig. 3.1. When the particular neuron of the second layer of the LTM that has been assigned to a particular digit fires for an input image of the same digit, then it is considered a success. If some other neuron fires instead, it is considered a failure. Thus, accuracy is determined.

From Fig. 3.1, we observe that as the network is exposed to more training samples for a given digit ('0' for this figure), the SNN learns more variations of the same digit. As a result, accuracy on new/ earlier unseen images of the digits (test accuracy) improves up to a certain point. But beyond a point, it goes down because the different variations of the same digit start superimposing on each other leading to a smearing effect. In our case here, relatively high train accuracy of 81.4% and peak test accuracy of 75% are obtained after

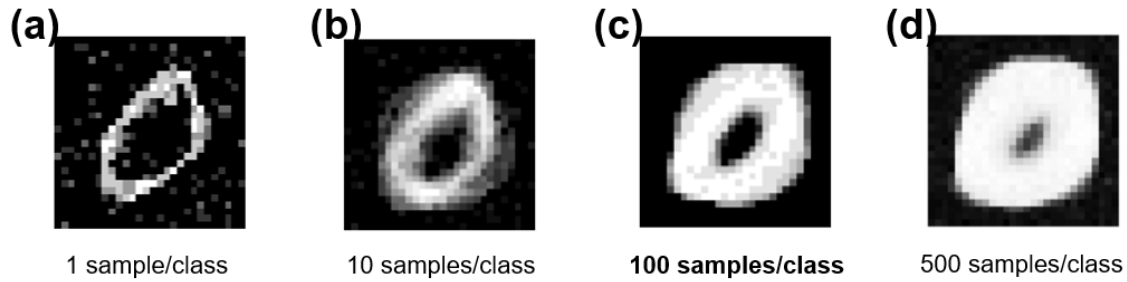

Figure 3.1: Time-averaged spike counts of each neuron of the first layer of the LTM SNN, over the entire time interval for which the neuron of the second layer of LTM SNN that corresponds to digit '0' is excited, are shown as a 28×28 grid. The SNN is trained with the following number of training samples per class (a) 1 (Train Accuracy: 90%, Test Accuracy 25%), (b) 10 (Train Accuracy: 61%, Test Accuracy 38%), (c) 100 (Train Accuracy: 81.4%, Test Accuracy 75%), and (d) 500 (Train Accuracy: 71.14%, Test Accuracy 69%).

training the network on 100 images per class/ digit. So, for all subsequent work in this paper, 100 images per class/ digit have been chosen for training the LTM network in this paper.

The smearing effect that's observed in the time-averaged spike counts of Fig. 3.1, happening due to presence of different variations of the same digit, can be further avoided by training the SNN on more closer variations of the same digit. The rationale for this design choice is the following: we are trying to model the LTM of humans here as they learn to identify digits. While humans do have the ability to identify the digit from different hand-written versions of the digit coming from different people (just like the MNIST database), they are always biased towards their own handwriting. When asked to write digits, every person writes them in a certain way, and thus they are exposed to more images of the digit with that variation compared to other variations that originate from other people's handwriting. We model this bias in the feedback next by training the same SNN of LTM on a different training sub-set from the MNIST data set.

In this sub-set, one representative image is chosen for each class/ digit from the original MNIST data set, and then 99 images closest to that image in terms of Euclidean distance are selected. Thus, a biased data set with 100 images per digit/ class is created for the purpose. Fig. 3.2 shows the spike counts of the neurons of the first layer of the LTM SNN when all ten neurons of the second layer corresponding to all ten digits are excited, for two cases: the SNN trained on an unbiased dataset as in Fig. 3.1, and and the SNN trained on the biased dataset constructed above. The smearing effect is much less in the case of the biased data set, and the digits are much more distinctly visible from the time-averaged spike counts in the case of the biased data set compared to the case of the unbiased data set.

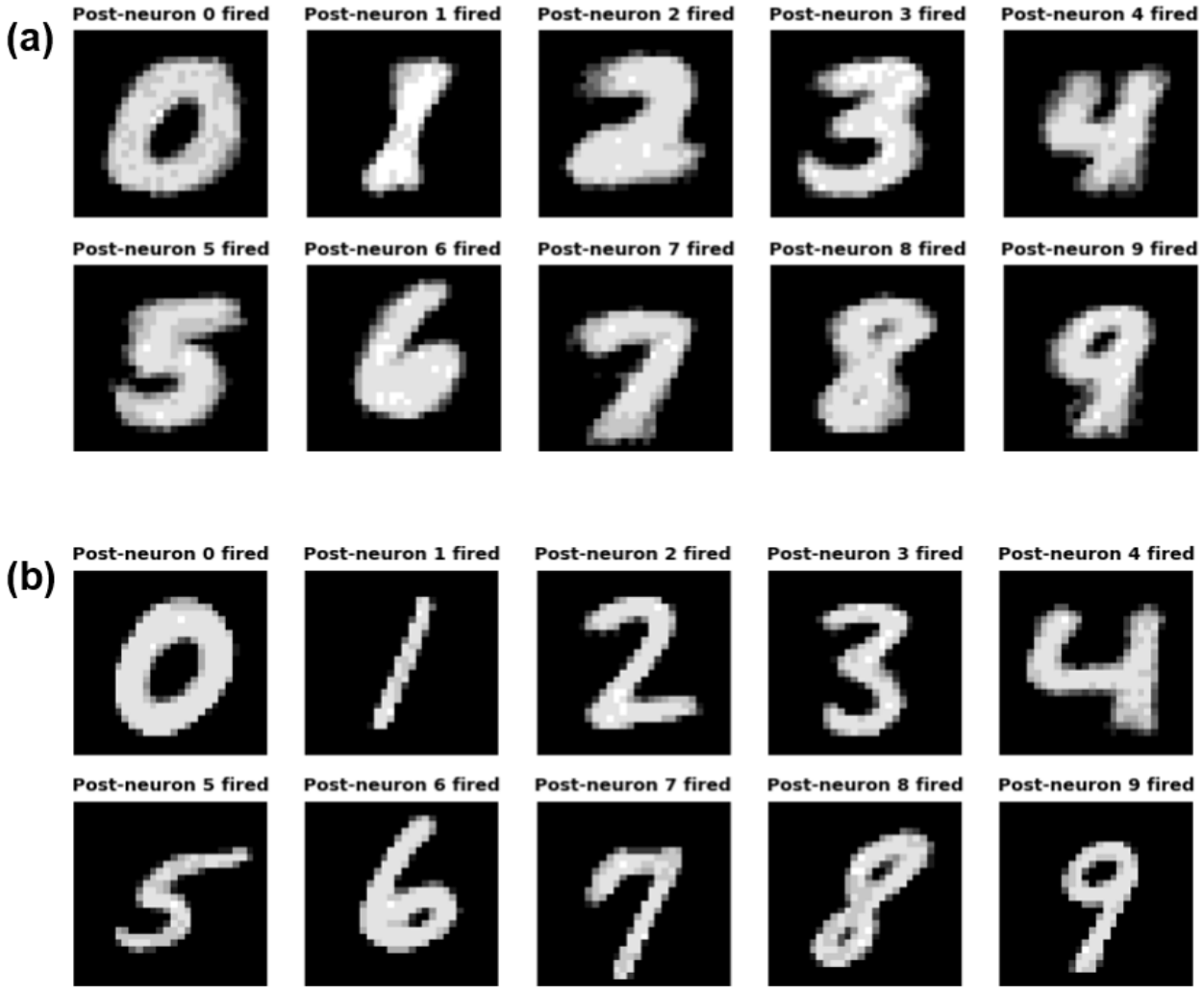

Figure 3.2: Time-averaged spike counts of each neuron of the first layer of the SNN-based LTM as a 28×28 grid when the neuron of the second layer, corresponding to each digit, is excited. The SNN has been trained on an unbiased data set in (a) and a biased data set in (b). Bright white: 10 spikes in the time interval, black: 0 spikes.

Chapter 4

Retention, Recapitulation and Interference in Working Memory

Having shown how the SNN representing the LTM can be trained to classify digits and how the neurons of that SNN spike to show the shapes of the digits, we now show some interesting properties related to the interactions between sensory memory, WM, and LTM. Similar properties are experienced by humans in their day-to-day lives, and we show here that our SNN model proposed here can indeed explain these experiences.

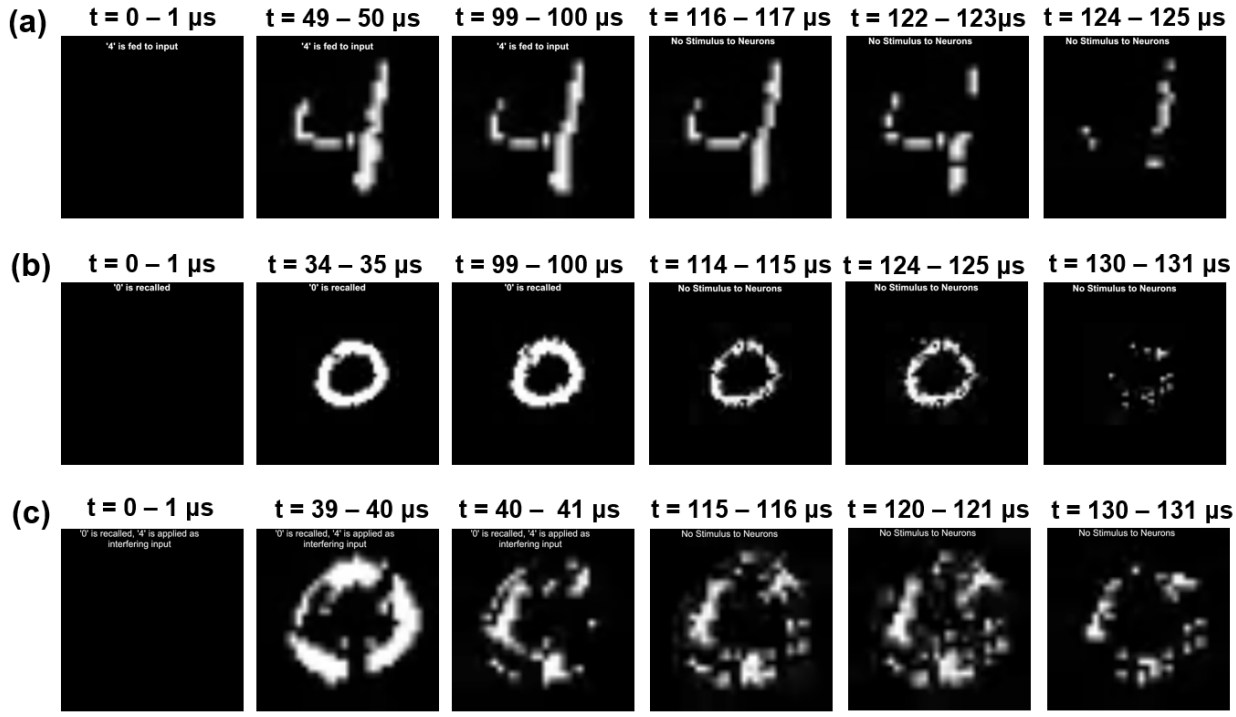

Figure 4.1: Spikes (over mentioned time intervals) of each visuo-spatial sketchpad neuron of the WM as a 28x28 grid when (a) an image of the digit '4' is applied to the sensory memory to test WM retention, (b) The digit '0' is recalled from the LTM to test the recapitulation abilities of the system, and (c) An image of the digit '4' is applied to the sensory memory and the digit '0' is recalled at the same time to demonstrate the interference phenomenon. In all three cases, the respective stimulus is applied to the neurons for a period of $110\mu s$, followed by a period of no stimulus for $110\mu s$. White pixel means at least one spike has been generated by the neuron in the given time interval; black means no spike has been generated during that time.

4.0.1 Retention

WM, earlier known as short-term memory, holds information for a short term only unlike LTM, as the name 'short-term' suggests. The retention capabilities of the WM neurons in our proposed SNN-based model have been tested by exposing the sensory memory neurons to a particular sample/ image from the MNIST dataset for a time period of time $t = 0$ μs to $t = 110$ μs. The input is subsequently switched off, leaving the neurons without any external stimuli for a time period of $t = 110$ μs to $t = 220$ μs. Throughout this whole period, at each time instant, which WM neurons spike are recorded. For each 1 μs time interval in this period, a 28×28 time-averaged spiking pattern plot is constructed, where each pixel takes a value of 255 if the corresponding WM neuron has fired at least once in that 1 μs time interval, and 0 otherwise. Fig. 4.1(a) shows such spiking patterns for 1 μ s intervals starting from $t = 0$ μs, $t = 49$ μs, and $t = 99$ μs (time instants when the input stimulus is present) and then for 1 μ s intervals starting from $t = 116$ μs, $t = 122$ μs, and $t = 124$ μs

From Fig. 4.1(a), it is observed that over the duration of the input stimulus (an image of the digit '4' for a time period of time $t = 0$ μs to $t = 110$ μs), the corresponding neurons in the WM fire, thus forming an image of the same digit every 5 μs. After the input has been switched off, the WM neurons continue to fire for a period of 14 μs, before the the spike pattern mimicking the shape of the digit vanishes.

This behavior can be attributed to the fact that the synapses between sensory memory and WM follow the tempotron model (equation 5.3) [8], which leads to a non-zero current getting injected into the WM neurons even after the sensory memory neurons stop firing. This means that after the input stimulus is removed at $t = 110\mu s$, the tempotron current to the WM neurons need not be 0 for $t > 110\mu s$. This is because of the remanent current injections induced by the sensory memory neurons that fired at $t < 110\mu s$, as per equation 5.3. Hence, it is also possible for the WM neurons to fire at $t > 110\mu s$. Thus, the WM neurons exhibit the property of retention, with the retention time depending upon the time constants τ_M and τ_S in equation 5.3.

4.0.2 Recapitulation

The recapitulation property of images in WM, or the recall of visual information from the LTM to the WM for immediate use, has been shown next. In the LTM, the neuron of the second layer corresponding to the digit '0' (as per the allotment during the training phase of the SNN for LTM as discussed before) is excited by applying an external bias current to that neuron. This current is applied for a time period of $t = 0$ μs to $t = 110$ μs. This current is subsequently switched off, leaving the neurons without any external stimuli for a time period of $t = 110$ μs to $t = 220$ μs. Throughout this period, the spikes at the neurons of the WM are recorded over 1 μs time intervals, as in the previous retention experiment.

Fig. 4.1(b) shows such spiking patterns for 1 μs intervals starting from $t = 0$ μs, $t =$

34 μs, and $t = 99$ μs (time instants when the input stimulus is present) and then for 1 μ s intervals starting from $t = 114$ μs, $t = 124$ μs, and $t = 130$ μs. The WM neurons clearly take the specific shape of the digit '0' as per what the LTM has learned based on the data set fed to it. The retention property of the WM is observed once again, with the digit '0' flashing even after the external input has been switched off.

4.0.3 Interference

The interference phenomenon in WM we study in this work is the interference between shapes of digits recalled from LTM and shapes of digits simultaneously fed through the sensory memory, as discussed in Section 1. This interference phenomenon has been modelled here by firing the second layer neuron (in the LTM) corresponding to the digit '0' through an external bias current. Simultaneously, the sensory memory neurons are exposed to an image of the digit '4'. These two phenomena happen together for a time period of $t = 0$ μs to $t = 110$ μs. Both of the external inputs are subsequently switched off, leaving the neurons without any external stimuli for a time period of $t = 110$ μs to $t = 220$ μs. As in the previous two experiments, the WM neuron spikes are recorded over 1 μs intervals. Fig. 4.1(c) shows the spiking patterns at $t = 0$ μs, $t = 39$ μs, and $t = 40$ μs (time instants when the input stimulus is present) and then for time 1 μs intervals starting at $t = 115$ μs, $t = 120$ μs, and $t = 130$ μs. The spike patterns of the WM neurons clearly show that both the digits are indecipherable now as opposed to the previous cases: Fig. 4.1(a),(b). This demonstrates the principle of interference in our SNN model of visual memory.

Chapter 5

Circuit Design of our Spiking Neural Network Based Working Memory

5.1 LIF Neuron

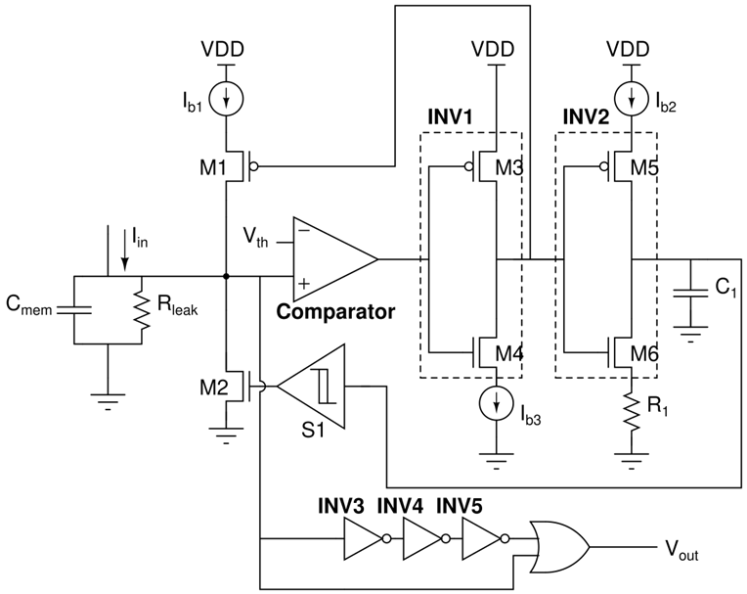

Figure 5.1: Design of LIF neuron circuit for our neuromorphic hardware.

Fig. 5.1 shows a CMOS-based LIF neuron circuit which we use in our design [20]. An input current I_{in} charges the capacitor C_{mem} along with the leakage resistor R_{leak}. V_mem is the membrane potential of the neuron, which is compared to a threshold potential V_{th} using a transconductor-amplifier-based comparator. When V_{mem} crosses the threshold potential, PMOS M1 turns on, and the current I_{b1} ($>> I_{in}$) rapidly increases the potential of the node V_{mem}. At the same time, capacitor C_1 is charged at a rate controlled by I_{b2}, and when the voltage across the capacitor exceeds the upper threshold of the schmitt trigger S1, NMOS M2 turns on, and the voltage V_{mem} is instantly brought down to ground. The capacitor discharges through the NMOS M6 and resistor R1, and when the voltage across C_1 falls below the lower threshold of the schmitt trigger, the voltage V_{mem} can increase through I_{in} again. This mechanism therefore results in the implementation of the refractory period of the neuron and is completely tunable. V_{mem} is then passed through a pulse-detector circuit consisting of $INV3$, $INV4$, $INV5$ and the OR-gate. V_{out} is therefore an active-low pulse between $VDD(= 1.8V)$ and GND.

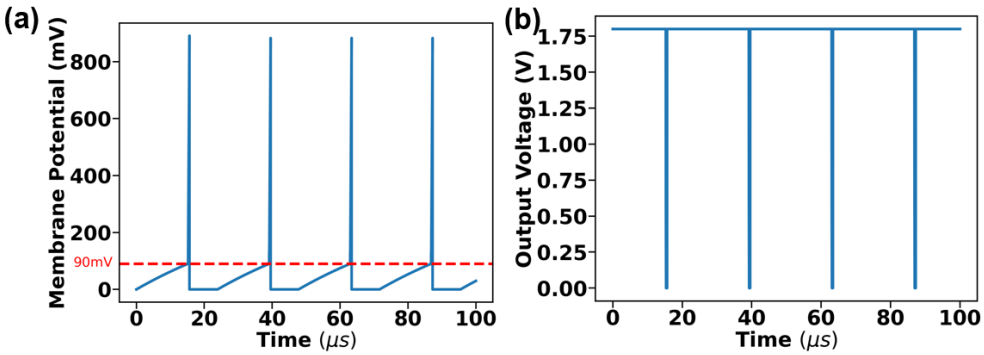

Figure 5.2: (a) Simulated V_{mem} and (b) Output voltage V_{out} for the LIF neuron circuit at an input current of $10nA$ and a threshold voltage of $90mV$. $R_{leak} = 33.33M\Omega$ and $C_{mem} = 5pF$.

SPICE simulations of the analog neuron circuit (Fig. 5.2) confirm its proper functioning. In line with our computational modeling, we choose $R_{leak} = 33.33M\Omega$, $C_{mem} = 5pF$ and $V_{th} = 90mV$.

5.2 Tempotron Circuit

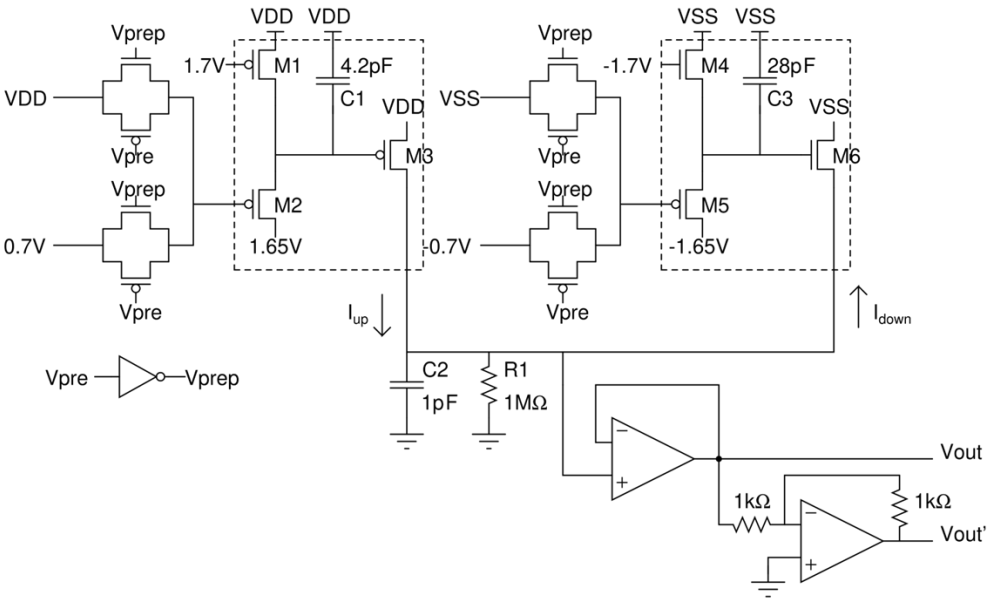

Figure 5.3: Design of the Tempotron circuit for our neuromorphic hardware

The circuit in Fig. 5.3 is a modified version of the Reset-and-Discharge Synapse [20]. NMOS M2 acts as a switch and is controlled by the spikes that are input to it. Initially, the voltage across the capacitor C1 is 0 and the gate voltage of M3 is at VDD. When a spike occurs (V_{pre} falls to 0V), M2 turns on, and the gate voltage of M3 is instantly set to 1.65V. After V_{pre} goes back to $1.8V$, M2 turns off and M1, which is operated in the subthreshold region as a constant current source, linearly charges capacitor $C1$, or equivalently, the gate

voltage of M3 linearly increases from $1.65V$ to $1.8V$. Since the MOSFET M3 operates in the subthreshold region, an exponential drain current is generated (I_{up}). The inverse process occurs in the subcircuit containing MOSFETs M4, M5, M6, and capacitor C_3, leading to a negative exponential current I_{down}. The net current flowing through resistor $R1$ is $I_{up} - I_{down}$, and the voltage across it is the biexponential waveform we seek. This voltage is passed through a voltage buffer and an inverting amplifier to generate the complementary biexponential outputs V_{out} and V'_{out}.

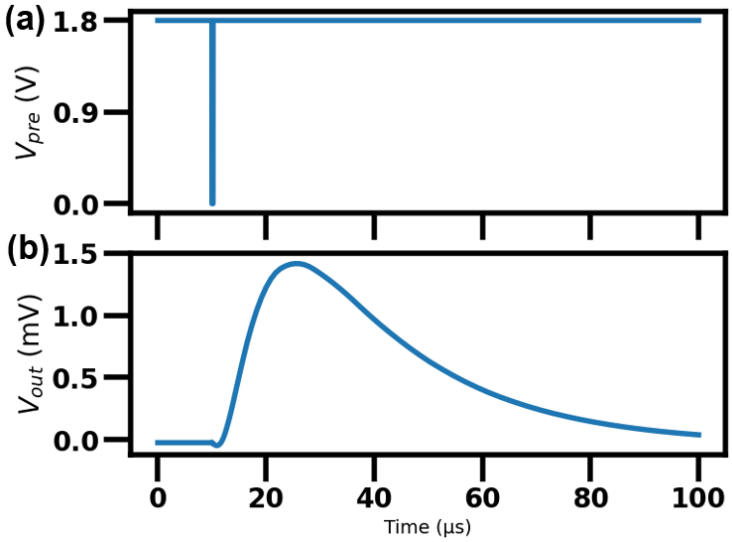

Figure 5.4: (a) An example input spike V_{pre} and (b) Output voltage V_{out} for the tempotron circuit.

SPICE simulations of the analog tempotron circuit (Fig. 5.4) confirm a bi-exponential output voltage in line with equation 2.3. Following equation 2.3, the experimentally obtained time constants τ_M and τ_S are $7\mu s$ and $2.5\mu s$ respectively.

5.3 Synaptic Connections between Neuron Groups

We now propose circuits to implement the weighted connections between neuron groups in our WM-SNN. There are namely 2 types of connections - (a) fully connected and (b) one-to-one.

First, there needs to be a mapping between the weight values (ranging from 0-900 in our computational modeling) and the actual synaptic conductance values ($166S - 332S$) [14]. The linear mapping between these is given by:

$$w = 900(332G - 1) \tag{5.1}$$

From equation 2.3, when a pre-neuron fires, the current injected into the post-neuron is

given by:

$$I = I_0 \cdot w \cdot exp(\cdots) \tag{5.2}$$

Replacing the weight with actual conductance values:

$$I = 900 \cdot 332 \cdot I_0 \left(G - \frac{1}{332} \right) \cdot exp(\cdots) \tag{5.3}$$

Equation 5.3 consists of 2 current terms, the first term can be interpreted as the current flowing across a synaptic conductance with a bi-exponential voltage across it, and the second term is the current flowing across an offset conductance with the complementary bi-exponential voltage across it. Currents through both these conductances add up, leading to the same biexponential current term as equation 5.2 flowing into the post-neuron.

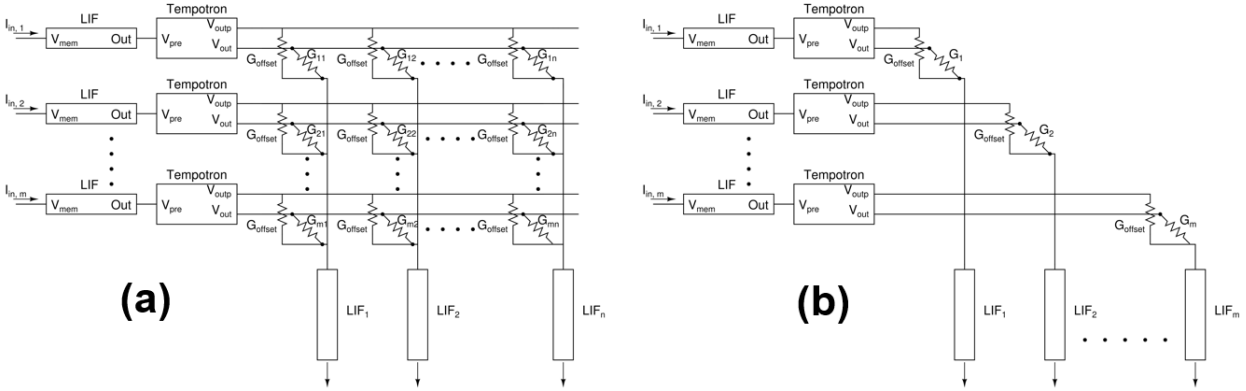

Figure 5.5: (a) A crossbar array implementing all-to-all connections between m pre-neurons and n post-neurons (b) A simplified pseudo-crossbar implementing one-to-one connections between m pre-neurons and m post-neurons

Fig. 5.5 shows the circuit implementations of the 2 types of connections that are part of our WM-SNN design. Here, the actual synapses are represented as resistors with constant conductance but can be easily replaced with emerging NVM devices with tunable conductance. The pre-neurons are placed in rows and the post-neurons are placed in columns in both designs. In Fig. 5.5(a), for a particular post neuron, the bi-exponential current terms due to all pre-neurons add up through the respective synaptic and offset resistors, thus implementing the all-to-all type of connection. In Fig 5.5(b), only one synapse cell exists between pre-neuron i and post-neuron i, thus implementing the one-to-one connection.

Chapter 6

Conclusion and Future Work

Thus, in this work, we have used a biologically plausible SNN design to model memory consolidation in LTM (learning of digits) and recapitulation and interference of digit shapes in WM. It is to be noted that while we restrict ourselves to the visual aspect of WM (visuo-spatial sketchpad) and learning of visual data in LTM, our SNN design can in principle be extended to the interaction between the phonological loop of WM and learning of speech and language in LTM [1, 2] with the help of more complicated SNN designs involving recurrent connections [21]. We have also proposed and simulated an analog circuit design for the entire system, with possibility of integration with NVM synaptic devices.

Our proposed SNN-based model here can be useful not only for developing better understanding of cognitive mechanism in the brain but also to develop novel neuromorphic hardware. Most neuromorphic hardware developed thus far, though being brain inspired in their design, are much like conventional artificial intelligence (AI) systems and only mimic humans in terms of the intelligence aspect: learning to classify images, recognize speech, complete sentences, etc. [15, 14, 16]. Despite the brain inspiration, most currently developed neuromorphic hardware cannot replicate human memory behaviour, as encapsulated by the modal model of memory described above, with the exception of a few recent works in that direction [22, 23]. If the SNN model proposed here is implemented on an emerging hardware platform consisting of some non-volatile memory devices (LTM) and short-term memory devices (short-term memory), then it can open a new direction in neuromorphic hardware research, ultimately leading to AI systems and robots that have very close-to-human behaviour.

Bibliography

[1] Goldstein, E.B.: Cognitive Psychology: Connecting Mind, Research, and Everyday Experience. 5th edn. Cengage Learning (2005).

[2] Baddeley, A.: Working Memory, Thought, and Action. 1st edn. OUP Oxford, Oxford (2007).

[3] Brooks L. R.: The suppression of visualisation by reading. Quarterly Journal of Experimental Psychology **19**, pp. 289–299 (1967). 10.1080/14640746708400105

[4] Buschman, T. J.: Balancing Flexibility and Interference in Working Memory. Annual Review of Vision Science **7**, pp. 367–388 (2021). 10.1146/annurev-vision-100419-104831

[5] McConnell, J., Quinn, J. G.: Interference in Visual Working Memory. The Quarterly Journal of Experimental Psychology Section A, **53**(1), pp. 53–67 (2000). 10.1080/713755873

[6] Squire, L. R., Genzel, L., Wixted, J. T., Morris, R. G.: Memory consolidation. Cold Spring Harb Perspect Biol. Aug 3; 7(8):a021766 (2015). 10.1101/cshperspect.a021766. PMID: 26238360; PMCID: PMC4526749

[7] Gerstner, W., Kistler, W.M., Naud, R., Paninski, L.: Neuronal Dynamics. 1st edn. Cambridge University Press, Cambridge (2014).

[8] Gütig, R., Sompolinsky, H.: The tempotron: a neuron that learns spike timing–based decisions. Nature Neuroscience 9, 420–428 (2006). https://doi.org/10.1038/nn1643

[9] Feldman, D. E.: The spike-timing dependence of plasticity. Neuron. 23;75(4):556-71 (2012). 10.1016/j.neuron.2012.08.001. PMID: 22920249; PMCID: PMC3431193

[10] Diehl, P., Cook, M.: Unsupervised Learning of Digit Recognition Using Spike-Timing-Dependent Plasticity. Frontiers in Computational Neuroscience. **9** (2015). 10.3389/fncom.2015.00099

[11] Biswas, A., Prasad, S., Lashkare, S., Ganguly, U.: A simple and efficient SNN and its performance and robustness evaluation method to enable hardware implementation. ArXiv:1612.02233 (2016). 10.48550/arXiv.1612.02233

[12] Sengupta, A., Ye, Y., Wang, R., Liu, C. , Roy, K.: Going Deeper in Spiking Neural Networks: VGG and Residual Architectures. Frontiers in Neuroscience. **13** (2019). 10.3389/fnins.2019.00095

[13] Lee, C., Sarwar, S. S., Panda, P., Srinivasan, G., Roy, K.: Enabling Spike-Based Backpropagation for Training Deep Neural Network Architectures. Frontiers in Neuroscience. **14** (2020). 10.3389/fnins.2020.00119

[14] Sahu, U., Goyal, K., Pandey, A., Bhowmik, D.: Spike Time Dependent Plasticity (STDP) Enabled Learning in Spiking Neural Networks Using Domain Wall Based Synapses and Neurons. AIP Advances **9**(12): 125339 (2019). 10.1063/1.5129729

[15] Sengupta, A., Banerjee, A., Roy, K.: Hybrid Spintronic-CMOS Spiking Neural Network with On-Chip Learning: Devices, Circuits, and Systems. Physics Review Applied **6** (6) (2016). 10.1103/PhysRevApplied.6.064003

[16] Guo, Y., Wu, H., Gao, B., Qian, H.: Unsupervised Learning on Resistive Memory Array Based Spiking Neural Networks. Frontiers in Neuroscience **13** (2019). 10.3389/fnins.2019.00812

[17] Szatmáry, B., Izhikevich, E. M.: Spike-Timing Theory of Working Memory. PLOS Computational Biology **6**(8):e1000879 (2010) 10.1371/journal.pcbi.1000879

[18] Hu, J., Tang, H., Tan, K. C., Gee, S. B.: A Spiking Neural Network Model for Associative Memory Using Temporal Codes. In: Handa, H., Ishibuchi, H., Ong, YS., Tan, K. (eds) Proceedings of the 18th Asia Pacific Symposium on Intelligent and Evolutionary Systems **1**. Proceedings in Adaptation, Learning and Optimization **1**. Springer, Cham (2015). 10.1007/978-3-319-13359-1_43

[19] Stimberg, M., Brette, R., Goodman, D. F.M.: Brian 2, an intuitive and efficient neural simulator. eLife **8**:e47314 (2019). 10.7554/eLife.47314

[20] Bartolozzi C, Indiveri G.: Synaptic dynamics in analog VLSI. Neural Comput. 2007 Oct;19(10):2581-603. 10.1162/neco.2007.19.10.2581

[21] Bellec, G., Salaj. D., Subramoney, A., Legenstein, R. and Maass, W.: Long short-term memory and learning-to-learn in networks of spiking neurons. Proceedings of 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montreal, Canada (2018). https://dl.acm.org/doi/10.5555/3326943.3327017

[22] Ji, X., Hao, S., Lim, K.G., Zhong, S. and Zhao, R.: Artificial Working Memory Constructed by Planar 2D Channel Memristors Enabling Brain-Inspired Hierarchical Memory Systems. Advanced Intelligent Systems **4**:2100119 (2022). 10.1002/aisy.202100119

[23] Ricci, S., Kappel, D., Tetzlaff, C., Ielmini, D., Covi, E.: Tunable synaptic working memory with volatile memristive devices. Neuromorphic Computing and Engineering (2023). **3**:044004. 10.1088/2634-4386/ad01d6