

PreSyNC: Hardware realization of the Presynaptic Region of a Biologically Extensive Neuronal Circuitry

Rounak Chatterjee*, Souradeep Chowdhury*, Soham Mondal†, Arnab Raha‡, Janet Paluh§, Amitava Mukherjee¶

*Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata, India.

†Memory Solutions, Samsung Electronics Co., Ltd., Bangalore, India.

‡Advanced Architecture Research, Intel Edge.AI, Intel Corporation, Santa Clara, USA.

§Colleges of Nanoscale Science and Engineering, Nanobioscience, SUNY Polytechnic Institute, Albany, New York, USA.

¶Department of Computer Science and Engineering, Adamas University, Kolkata, India.

Corresponding author email: rounakchatterjee007@gmail.com

Abstract—Spiking Neural Networks (SNN) have gained relevance in recent times, due to their ability to mimic the biological nature to communicate and process sparse asynchronous binary signals in a massively parallel fashion. SNN based neuromorphic hardware exhibits highly desired favourable properties such as low power consumption, fast inference, and event-driven information processing. A recognized challenge of standard SNN neuron models is their limited capabilities in biological applications, such as applying neural networks to study network responses arising from variations pertaining to damage, external influence or disruptions in channel transfer dynamics.

This paper presents Pre-Synaptic Neuronal Circuit (PreSyNC), a high performance hardware realization of an input-specific presynaptic region of a generic neuron, without abstraction of primary intra-neuronal parameters. PreSyNC is configured to operate on three precision modes: IEEE 754 single precision, double precision and the recently developed universal number posit number system. The developed hardware design is compared to current standards of SNN neuron models as well as biological models in terms of flexibility, resource efficiency and damage modelling capability. Error margins as low as 0.9% were obtained and suggest the capability of our hardware in handling applications involving large scale neuron networks. These architectures are synthesized on 45 nm process technology where they all operate at a minimum frequency of approximately 1GHz. The three precision modes are compared based on power, accuracy, and sensitivity handling and are expected to benefit implantation oriented applications such as neural prosthesis and Human-Computer Interaction (HCI). The posit-based implementation outperforms the rest of the operating modes in terms of RMS error, while having 26.3% less area and 25.2% less power consumption compared to double precision implementation. These new architectures can be expanded in the future with various post-synaptic inputs to open up a broader understanding of biological systems and other applications.

Index Terms—Neuromorphic computation, presynaptic region, neural networks, posit arithmetic

I. INTRODUCTION

Neurons act as computational engines of the brain, which they accomplish by exchanging information with other neurons via synapses and experience-dependent plasticity. They are dissimilar from most other cells in the body in that they have their distinct morphological axonal and dendritic regions provide each with specific polarization-incorporating functions that contribute to directional cell communication. Despite the wide variation in neuron type and morphology, general fundamental structural and functional features of the neuron can be exploited to design biologically realistic mathematical models of a neuron which closely resemble biologically obtained data [1]–[3]. A long-term change in the synaptic weight associated with a neuron provides a physiological substrate for learning and memory, whereas short-term changes support synaptic computations [4]. The effect of an action potential transmitted from one neuron to another depends on the history of neural activity at either or both sides of the synapses such that their effect can last from milliseconds to months [5]. Synapses in these processes act as filters gateways that selectively and unreliable filter the flow of information between pre and postsynaptic region of a neuron.

However, since experience-dependent plasticity data in biological systems obtained as such is reflective of trends specific to regions and neural systems, there is an overall lack of a general widely applicable neuron model that can be used to compare systems. Moreover, a neuron on its own cannot perform any significant function. Modelling is critical since most of the important functions of a neuron involved with communication and coordination require a large number of neurons and synapses. Due to these reasons, any form of analysis involving biologically accurate neural networks must be carried out with the aid of simulation.

Many software-based biologically accurate SNN models have been developed such as NEURON, NEST, Brian, NEMO, PYNN and HRLSIM. While providing accurate results, these software-based simulations generally encounter extraordinarily high computational costs while performing complex numerical simulations. Hence modern

computers still fail to obtain real-time performance when they come to simulate large-scale neural networks. For instance, a one-second simulation of a neuronal network composed of 8 million integrate and fire neurons and 4 billion synapses processed on the Gene Rack, a supercomputer with 2048 processors, requires 80 minutes to compute [6], versus the nearly instantaneous response capability of a biological neural network.

Due to essential capabilities to obtain the balance of resource management and real-time speed required by the user, hardware platforms for large-scale simulation of neural networks must be developed. Recent models suggested for successful implementation of neuron models include the Time Machine approach [7], SpinNaker [8], Neurogrid [9], BrainScales [10], etc. . BrainScales is a wafer-scale neuromorphic system, in which each wafer contains 48 reticles with eight High-Count Analog Neural Network (HiCANN) dice. Each such HiCANN die has the capability to emulate 512 adaptive exponential neuron models. SpiNNaker is a one-million-core supercomputer, developed exclusively for massively-parallel real-time simulation of large-scale neural networks, making it one of the largest digital neuromorphic platforms to date. Neurogrid is a real-time system consisting of over one million quadratic integrate-and-fire neurons, in which neuron and synapse dynamics are emulated using analog circuits and communications are performed by digital means, for simulation of over a billion synapses. However, since these models have been developed with the intention of utilization in Spiking Neural Network (SNN) architectures, they only approximate the effect of the numerous parameters in a neuron. Such exclusive abstraction of intracellular dynamics of a neuron that ignores other biologically modulated parameters is restricted in applications and cannot be used to simulate the effect of variations of these individual parameters in the neuron and its networks.

In PreSyNC, we focus on the presynaptic region of the intra-neuron communication network and use a retrograde messenger mediated plasticity (RMMP) model as opposed to glial cell-mediated model. In the RMMP bipartite neuron system there are no dynamics considered in the synapse and hence the synaptic current from the presynaptic region is the same as the synaptic current entering the postsynaptic region. We consider the dynamics of retrograde messengers that travel across the synapse via diffusion to be an input parameter in our model. We consider three different hardware precision modes integrated in our core for accuracy and handling, investigating three standard number systems widely accepted in computer architecture: 32 bit posit with 4 bit exponent size, IEEE 754 single precision floating point and IEEE 754 double precision floating point. These models are compared on the basis of stability as well as spiking values of the individual variables and the deviation from that of software results, to find the most efficient and economical system. In a biological system, all the parameter equations have very high sensitivity. To our knowledge this is the first work to implement accurately the presynaptic region of a neuron in detail in hardware.

We mapped and synthesized the PreSyNC hardware designs on 45 nm process. Simulation results show that differential forms of these equations reproduce the expected characteristics of the presynaptic region of a neuron, while conveniently transformed into discretized form. Synthesis results show that using posit operation core gives us performance slightly better than IEEE 754 double precision while having more accuracy than IEEE single precision operation core which affirms the claim of posit being a possible replacement of floats. Also this proposal remains the only available work in terms of real time application of posits in neural hardware design.

II. MODELLING

A. Presynaptic RMMP neuron model

We consider the neuron model proposed by Faghini-Moustafa [11] that describes the processes involved in synaptic plasticity in a synapse mediated by retrograde messengers (RMs). The model, described in Fig.1, demonstrates the synaptic current release from the presynaptic region when activity spike and retrograde messenger are provided to it. The amount of RM that diffuses into the presynaptic neuron consists of RMtrace. Activity trace of input spike considers the effect of plasticity and assigns the effect of latency to the input impulse train. RMtrace along with the activity trace of input spike decide the amount of inhibitory complex released. This inhibitory complex decides the effective release of synaptic vesicles from the presynaptic region. Thus, inhibitory complex concentration along with The concentration of neurotransmitter release and the activity trace of input spike decide the magnitude associated with the released synaptic current. This value is the synaptic weight and its rate of change gives the synaptic efficacy, a parameter that describes the strength of plasticity.

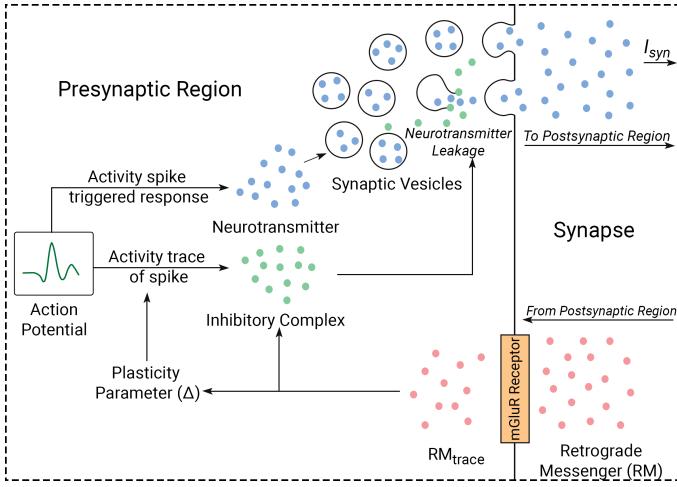


Fig. 1. Pathways for Synaptic Transmission from presynaptic region in a retrograde messenger mediated plasticity (RMMP) neuron model

Table I specifies the values of constants for the system of equations considered by us in the PreSyNC model, which are taken directly from [11]. However, the equilibrium point of inhibitory complex is adjusted because the value considered in [11] has a very high threshold and is impractical, to a value where the effects of the impulses can be observed distinctly. The state variables of the presynaptic region: effective strength of feedback of Retrograde Messenger (RMtrace), Inhibitory complex concentration (Inh) and activity trace of spike (C) and the concentration of neurotransmitter (D) are defined using the Tsodyks Markram Model as:

$$\frac{d}{dt} RM_{trace} = -\frac{RM_{trace}}{\tau_r} + RM \quad (1)$$

$$\frac{d}{dt} Inh = -\frac{Inh}{\tau_{inh}} + RM_{trace}.C \quad (2)$$

$$\frac{d}{dt} C = -\frac{1}{\tau_c} [C + \Delta \delta(t - t_p)] \quad (3)$$

$$\frac{d}{dt} D = -\frac{1}{\tau_d} [D + \sum \delta(t - t_d)] \quad (4)$$

where, τ_r is the time constant associated with the influx of RMs, τ_{inh} is the time constant associated with inhibition of release of neurotransmitters, t_p is the time when an impulse is received at the presynaptic neuron, RM is the concentration of retrograde messenger present in the postsynaptic region and τ_c and τ_d are the time constants associated with the biological latency in activity spike and neurotransmitter release respectively. Equations 1-3 give the resulting effect of the action potential on the concentration of the inhibitory complex. Equation 4 gives the concentration of neurotransmitter released from the presynaptic neuron which depends on the rate at which impulses enter. Δ is the parameter for hebbian plasticity and is calculated according to Table II.

TABLE I
SPECIFIC VALUES OF TIMESCALES, THRESHOLD VALUES AND CONTROL PARAMETERS IN NEURON MODEL

SI	Attributes	Values	Units
1.	Time constant for Activity Spike ^a	0.1	seconds
2.	Time constant for RMtrace ^a	0.4	seconds
3.	Time constant for Inhibitory Complex conc. ^a	0.1	seconds
4.	Time constant for Neurotransmitter conc ^a	0.02	seconds
5.	Time constant for Synaptic Current ^a	0.1	seconds
6.	Initial Probability (P_{init})	0.25	none
7.	Retrograde Messenger Influx (RM)	0.691	M
8.	Retrograde Messenger threshold (RMrest)	0.691	M

^aValues vary from cell to cell [11]

TABLE II
SPECIFIC VALUES OF TIMESCALES, THRESHOLD VALUES AND CONTROL PARAMETERS IN NEURON MODEL

u[n]	$sgn(RM_{trace} - RM_{rest})$	Δ
1	1	1
1	0	-1
0	1	-1
0	0	0

The probability of inhibition and release of neurotransmitter, concentration of neurotransmitter and synaptic current are written as:

$$P_{inh} = e^{-\frac{0.0000001}{Inh} - 1} \quad (5)$$

$$P_{rel} = P_{init}(1 - P_{inh}) \quad (6)$$

Here, the probability of inhibition is a function dependent only on the concentration of the inhibitory complex. The probabilities considered here decide the synaptic weight of a neuron i.e., the amount by which the transmission of an impulse through the synapse is magnified. The rate of change of synaptic weight is called synaptic efficacy and is calculated as:

$$\frac{d}{dt} \omega = P_{rel}.C.D \quad (7)$$

Hence, the resulting synaptic current transmitted can be written as:

$$I_{syn}(t) = \omega \sum \frac{t - t_p}{\tau} e^{\frac{t - t_p}{\tau}} \delta(t - t_p) \quad (8)$$

B. Modified PreSyNC model

The equations, defined in the above sub-section, are optimized to improve the computational efficiency of the model and reduce its implementation cost by adopting polynomial expansion of functions to simplify complex and lengthy functions. In the 1st order Tsodyks Markram differential equations (1)-(4), since all the equations are linear, there is no requirement to make any adjustments to them.

For determining the probability of inhibition, we encounter the implementation of exponential function. On using CORDIC algorithm to implement the same, the area of the hardware is increased by a large amount. Also since the CORDIC algorithm operates on convergence, this would lead to an increase in time taken per process for computation. Hence, we solve by using an rectangular hyperbolic function that closely resembles the curve in the region of operation.

$$P_{inh} = \frac{-0.00007}{Inh - 0.00001} + 1.1 \quad (9)$$

For synaptic current, we consider the summation segment as a reset enabled function. Thus, we obtain a pair of equations Y and Z, which generate the form of unweighted synaptic current.

$$Y = \frac{t - t_p}{\tau} e^{\frac{t - t_p}{\tau}} \quad (10)$$

$$Z = \tau e^{\frac{t - t_p}{\tau}}$$

For a pre-fixed frequency of t_{p0} interval, the summation symbol in (8) is removed without any loss of accuracy by considering the summation of each power for a finite number of terms. However, for PreSyNC model to function for any form of t_p provided, we needed to make some approximations. Adoption of a piece-wise approach leads to a trade-off between area efficiency and error efficiency. Also the properties of the curve are lost at every new impulse completely, which leads to significant error in output, unless more functions are added to compensate the loss. But this almost doubles the area. Hence,

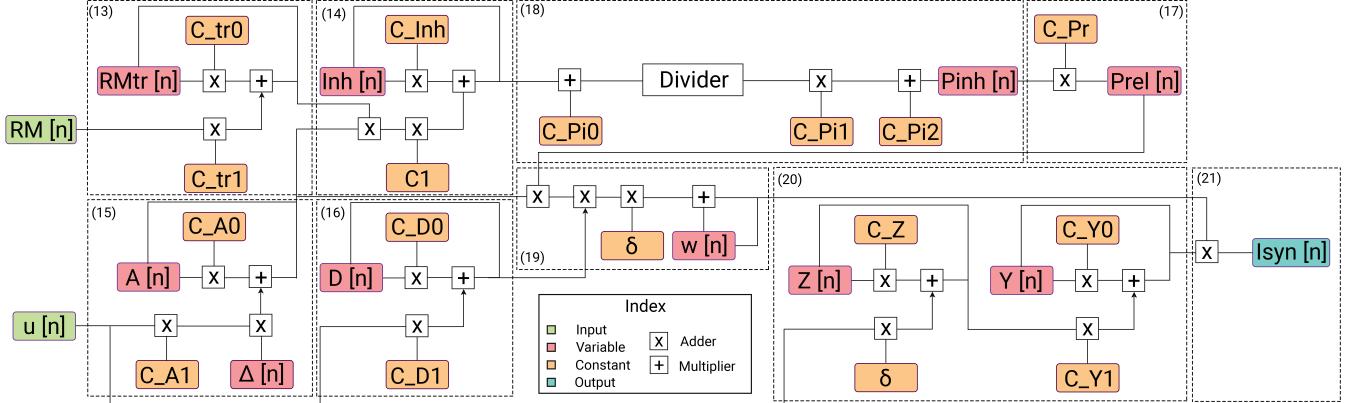


Fig. 2. Schematic diagram for hardware realisation of discretized PreSync neuron model. The different blocks in the diagram are numbered in accordance with the equation they are representing. The adders and multipliers used are pipelined for the different precision modes to different stages

the alternative is to go for differential form of these equations. As such the impulse driven activation can be enabled, without removing the value of Y completely before the activation. For this particular set of equations, activation conditions can be met simply by adding 1 to Z , when an impulse arrives in each cycle. Such a form is more robust in error handling for discretized calculations. Based on above technique, $Isyn$ is calculated as follows:

$$\begin{aligned} \frac{dY}{dt} &= \frac{Z - Y}{\tau} \\ \frac{dZ}{dt} &= \frac{-Z}{\tau} + u[n] \end{aligned} \quad (11)$$

$$Isyn = \omega \cdot Y \quad (12)$$

III. SYSTEM ARCHITECTURE

A. Discretization of modified PreSync model

The majority of equations in modified PreSync model are differential or linear in nature, which greatly simplifies the hardware realisation of the model. However, we require one division module in our design. Divider modules rarely reach timing and power requirements and hence define the upper bounds in terms of operation for the hardware realisation. In the Tsodyks Markram model functions (1)-(4), the inverse of the time constant is used directly.

$$\begin{aligned} RMtrace[n+1] &= RMtrace[n] \\ &+ \delta \left(-RMtrace[n] \cdot \frac{1}{\tau_r} + RM[n] \right) \end{aligned} \quad (13)$$

$$RMtrace[n+1] = C_tr0.RMtrace[n] + C_tr1.RM[n]$$

$$\begin{aligned} Inh[n+1] &= Inh[n] \\ &+ \delta \left(-Inh[n] \cdot \frac{1}{\tau_{inh}} + RMtrace[n] \cdot C[n] \right) \end{aligned} \quad (14)$$

$$Inh[n+1] = C_Inh0.Inh[n] + C_Inh1.RMtr[n].C[n]$$

$$A[n+1] = A[n] - \delta \frac{1}{\tau_c} (A[n] + \Delta[n].u[n]) \quad (15)$$

$$A[n+1] = C_A0.A[n] + C_A1.\Delta[n].u[n]$$

$$\begin{aligned} D[n+1] &= D[n] - \delta \frac{1}{\tau_d} (D[n] + \sum u[n]) \\ D[n+1] &= C_D0.D[n] + C_D1.u[n] \end{aligned} \quad (16)$$

The discretized equations of synaptic current, synaptic weight and probability of neurotransmitter release can be obtained with minimal changes as:

$$P_{inh}[n] = -0.0007/(Inh[n] - 0.00001) + 1.1 \quad (17)$$

$$P_{rel}[n] = 0.25 (1 - P_{inh}[n]) \quad (18)$$

$$\omega[n+1] = \omega[n] + \delta (P_{rel}.A[n].D[n]) \quad (19)$$

$$Y[n] = C_Y0.Y[n] + C_Y1.Z[n] \quad (20)$$

$$Z[n] = C_Z0.Z[n] + C_Z1.\delta.u[n]$$

$$Isyn[n] = \omega[n].Y[n] \quad (21)$$

Figure 2 demonstrates the correspondence of the system of equations (13)-(21) to the different modules in the hardware realisation of the discretised PreSync model. In these equations, the value of δ is calculated as the number of intermediate points at which hardware design processes result in a duration of one second. We set this value to be 1000 for proposed design and compare results against a software simulation developed for PreSync RMMP model, in which δ is 10 times that of the corresponding hardware realisation.

B. Error Metrics

For the PreSync hardware realisation under consideration, all the concentration changes are highly sensitive processes. The primary requirement for a replication of the intra-neuron processes is that the net concentration change per impulse for all variables is the same. This is in keeping with the fact that by the law of conservation of mass, the net molecular release for all concentration variables used in this model, every time there is a triggered release, must be the same for hardware design as it is for the software simulations. Hence, we consider the cumulative relative difference in the hardware realization and software simulation to be the metric for biological accuracy. The area average error (AAE) hence calculated is observed for a single impulse or an impulse train.

$$AAE = \frac{\int_0^{t_p} f(t)_{Hardware} dt - \int_0^{t_p} f(t)_{simulation} dt}{\int_0^{t_p} f(t)_{simulation} dt} \quad (22)$$

To calculate the error in translation of the equations in context of hardware design, we consider root mean square error (RMSE) between the hardware generated values of the variables and their counterpart software generated original values.

$$RMSE = \sqrt{\frac{\int f(t)_{Hardware} - f(t)_{simulation}}{n}} \quad (23)$$

IV. SYSTEM DESIGN

A. Flexibility of PreSync Design

In Fig.3, the proposed PreSync model is compared with standard phenomenological neuron models used in SNN architectures as well as biological models that are widely accepted. We assume that no additional circuitry is added for modelling special cases in our comparison since it is not being reproducible on a large scale. Hence, we compare the effectiveness of the proposed PreSync hardware design with its counterparts on their variation modelling adaptability. The model that requires the least amount of pre-processing of data before being fed into a system is considered the most efficient.

We consider three primary types of neuron models. Phenomenological neuron models, that are usually involved in designing SNN architectures, use simplified forms of these equations with a goal for obtaining the output response and synaptic weights for a neuron. However, such an abstract neuron model can not be used for studying the channel or molecular dynamics of a neuron. For the said comparison, we consider the flow of information in Clopath model [12], frequently used in NEST simulations. Postnov neuron

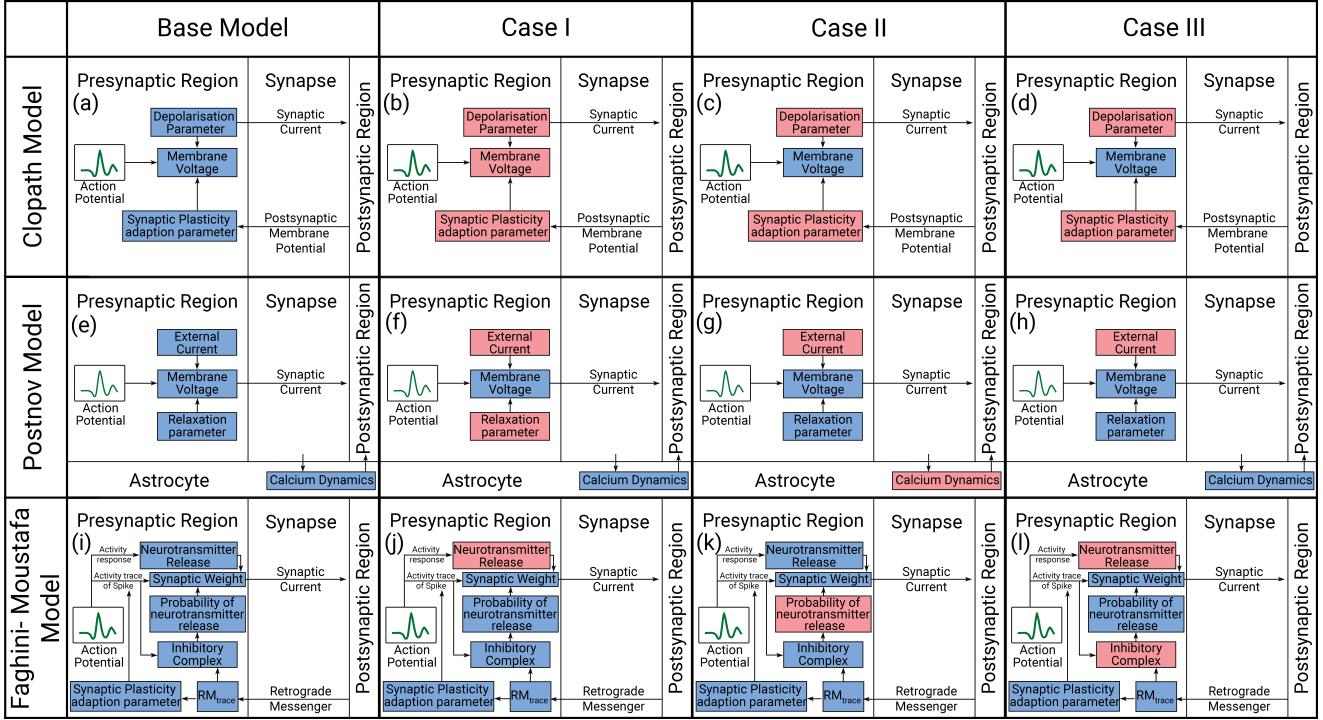


Fig. 3. Canonical flow of information in (a)Clopath Model, (e)Postnov Model and (i)Faghini-Moustafa Model. The figure exhibit the variables for which information is being fed to the model (represented in red boxes) for 3 separate cases. Case I (b)(f)(j) and III (c)(g)(k) considers variations due to neuronal damage or external influence, whereas case II (d)(h)(l) considers the variations in synaptic transfer dynamics for the three models respectively.

model [13], one of the widely accepted tripartite neuron models, provides a more elaborate analytical description of pathways controlling synaptic weight dynamics. This is achieved by defining molecular dynamics in the synapse and astrocyte, while using FitzHugh Nagumo phenomenological neuron model for describing neural dynamics in the postsynaptic and presynaptic neuron. Thus the Postnov model's response, while can be achieved for different variations in molecular dynamics of the presynaptic region, cannot be achieved by direct input of data. The proposed PreSyNC model is an analytical bipartite neuron model, which describes the molecular dynamics of intra-neuron parameters as well as channel dynamics of the neuron. These three models are compared amongst each other in terms of their adaptability for modelling various cases, as shown in Fig.3.

Case I : It is considered that the neurotransmitter dynamics are being affected by other molecular processes in the presynaptic neuron and the resulting concentration of neurotransmitter release is available. For the Clopath neuron model, the model must be simulated and the model gain parameters for Long term Depression (LTD) and Long term Potentiation (LTP) must be changed to appropriately reproduce the effects. For the Postnov neuron model, the external current can be described as a function of neurotransmitter release and reproduce the effects of the neurotransmitter dynamics. However, for the PreSyNC model, the information on concentration of neurotransmitter release can be directly be fed to the model.

Case II: It is considered that the synaptic channel transfer dynamics are being affected by other molecular processes restricting the outflow of neurotransmitter into the synapse and the resulting reduction in synaptic current is available. For Clopath neuron model, the model must be simulated and the model gain for synaptic weight parameters must be changed to appropriately reproduce the effects. For Postnov neuron model, the equations in the synapse and the astrocyte require to be adjusted accordingly. However, for PreSyNC model, we can alter the probability of inhibition function to include the effect of synaptic channel dynamics. Here, even though we are not directly feeding the data, it is still the model that requires minimal changes. For the above two cases, we observe that there are no additional dependencies in the interactions amongst the variables.

Case III: It is considered that the inhibition complex dynamics is a function of neurotransmitter release. For Clopath neuron model, the model must be simulated and the model gain for synaptic weight parameters must be changed to appropriately reproduce the effects. For Postnov neuron model, the external current as well as relaxation

parameter constants must be fed to the system. Since the variables whose dynamics is altered are not directly related by default, for the PreSyNC model, if we have the data regarding neurotransmitter release dynamics only, this case cannot be simulated without additional circuitry. However, if the inhibitor dynamics data are generated independently and fed to the model, we can simulate the remaining variables.

B. Hardware Realization of PreSyNC

This section provides a detailed description of the hardware realization of the proposed discretized presynaptic neuron model. The computation core has been developed to work with three different number systems namely IEEE-754 single precision format, IEEE-754 double precision format and recently developed universal number system posit number system. IEEE-754 format is the widely accepted format for floating-point computation whereas posit is a recent development in numerical computation [14] [15]. Here, 32-bit posit numeral with exponent size 4 is selected, as this particular type-III unum representation requires a large mantissa width of 25-bit (N-ES-3), that is comparable to that of single precision floating point format (24-bits). N denotes word size and ES denotes exponent size [16]. The 32-bit word size is chosen for the posit representation in order to contrast its output with 64-bit IEEE-754 float representation to find if comparable accuracy can be obtained with fewer bits [16].

Fig. 4 describes the different stages of pipelining in terms of registers or flip-flops in the arithmetic units for the three precision modes considered. The designs have been pipelined extensively and even arithmetic operator pipelining has been done for mantissa operations for all three precision modes by usage of custom-made, non-compressed, tree structured, pipelined, moderately large sized and power consuming high speed, Wallace tree multiplier. The largest combinational unit in the posit and floating point multiplier architecture is 1 mantissa multiplication of respective mantissa widths, whereas the largest one in division is the iterative Newton-Raphson division method which consists of 3 mantissa width multiplications for both posit and float. All of these multiplications are done using custom implemented Wallace tree unsigned multiplier for gaining greater speed and to meet acceptable operating frequency. The highly configurable posit hardware exhibit a lot higher complexity which has been designed to handle all combinations of dynamic posit word size with variable exponent sizes. The posit number system is composed of a run-time varying exponent component, which is defined by a composition of varying length "regime-bit" and "exponent-bit" (with

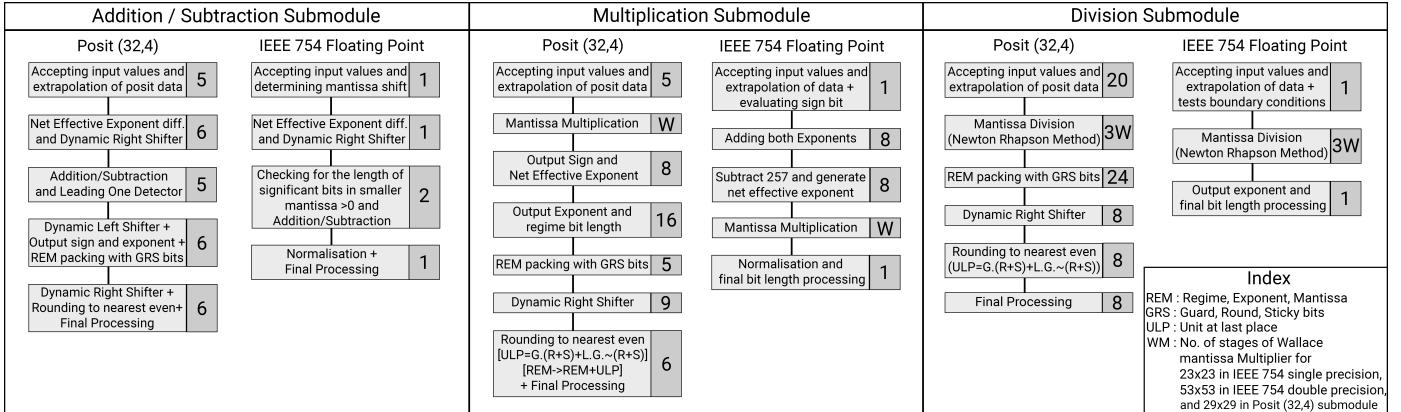


Fig. 4. Number and location of pipelining stages for different precision modes in hardware realization of PreSyNC model. The relationship between the number of stages in wallace tree mantissa multiplier for IEEE 754 single precision floating point, IEEE 754 double precision floating point and posit (32,4) precision mode can be expressed as $x:2x:x+5$.

TABLE III
PERFORMANCE PARAMETERS OF THE DIFFERENT CORES FOR PROPOSED
HARDWARE REALISATION

Sl. No.	Operation Core for Hardware Realisation	No. of Gates	Latency (ns)	Operating frequency
1.	IEEE 754 Single Precision	949K	1.03	970MHz
2.	IEEE 754 Double Precision	2737K	1.03	970MHz
3.	(32,4) posit	2163K	1.11	900MHz

TABLE IV

OPERATION PARAMETERS OF THE DIFFERENT CORES FOR PROPOSED
HARDWARE REALISATION

Sl. No.	Precision mode for Hardware Realisation	Area (mm^2)	Power (W)	Speed-up by pipelining
1.	IEEE 754 Single Precision	0.757	0.92	3.88
2.	IEEE 754 Double Precision	2.183	2.68	3.88
3.	(32,4) posit	1.725	2.14	3.60

a maximum size of ES bits, the exponent size). This in effect also makes the fraction part to vary at run-time in size and position. These run-time variations make posit arithmetic units very complex. Large number of pipelined stages/registers needed to be introduced to posit arithmetic units in order to break combinational loop and to incrementally meet timing goal. Table III and IV show the comparison of performance and operation parameters across all these modes.

The three different hardware designs of PreSyNC model, with exact same input configurations, are simulated in Questasim 10.0b Simulator by simulation scripts and output results are functionally verified by self developed utility scripts in C++ v.11 and Python 3.8. For simulation as well as the hardware design for the given system of equations, we consider a constant input of Retrograde messenger from the presynaptic region such that it is always just above the threshold for activation of synaptic plasticity parameter. We implemented PreSyNC in RTL which are then mapped and synthesized using Synopsys Design Compiler on 45 nm OpenNangate technology. The power consumption is derived using Synopsys Power Compiler. Standard operating frequencies of almost 1 GHz is met for all the designs in 45nm ASIC.

For low power, economical and optimized implementation of most widely used single precision float, we have made use of reduced complexity hardware of base arithmetic units. Complexity of other numerous sub-units are also significantly reduced. This brought down the power and area footprints by orders of magnitude than the base version to 13.75mW and $0.292mm^2$ respectively, by reduction of the number of intermediate registers and less frequent switching of intermediate variables. Through extensive pipelining, single precision floating point optimized version is able to meet acceptable operating frequency of 1030 MHz.

V. RESULTS FROM HARDWARE REALIZATION OF THE PRESYNAPTIC MODEL

A. Error Handling

In the outputs obtained from the hardware realisation of PreSyNC neuron model, a latency in the results is observed. The average latency is in the range of 0.015–0.017s, which is an effect of pipelining. For this very reason, the first values obtained within this period must be

TABLE V
SPECIFIC VALUES OF TIMESCALES, THRESHOLD VALUES AND CONTROL
PARAMETERS IN NEURON MODEL FOR MINIMUM MAE

Sl	Attributes	Values	Units
1.	Time constant for Activity Spike	0.1052	seconds
2.	Time constant for RMtrace	0.42	seconds
3.	Time constant for Inhibitory Complex conc.	0.105	seconds
4.	Time constant for Neurotransmitter conc	0.0253	seconds
5.	Time constant for Synaptic Current	0.94	seconds
6.	Initial Probability (P_{init})	0.25	none

rejected and the value of ω is fixed at 0, to remove the effect of default values of ω from influencing the nature of the curve. The error is less than 1% for architectures designed with any of the three operation cores, which are sufficiently low and the output is barely affected and hence acceptable. On comparing the RMSE and AAE of the different hardware precision modes for Isyn, it is observed that the error is the least for posit representation.

B. Sensitivity

For our hardware design with precision mode IEEE-754 Single Precision floating point, we observe that the amplitudes of the variables are higher in the hardware realisation of our models than that of software results. This is a result of using differential forms of the equations to develop the PreSyNC hardware realisation. Hence, the error response of the model is directly dependent upon the value of δ utilised to develop the operational constants of the hardware design. Table V references the changes to the operation constants that would minimise the RMS error of the models. However, it is observed that making this change would result in higher AAE. Since for the PreSyNC neuron model, the conservation of molecules involved is of higher importance from a biological perspective, we disregard the amplitude error in favour of lower AAE. However, for a system requiring precision in translated equations, it can be achieved by using the constant values in Table V.

We observe that the sensitivities of the hardware system are affected noticeably for the different precision modes. In Fig. 5, the variations across the three precision cores for identical set of equations can be observed. It can be seen that from the comparison of output synaptic current (Isyn) in these models, the posit (32,4) precision mode functions with higher accuracy.

C. Efficiency

From Fig. 6, it is observed that while posit (32,4) architecture have accuracy comparable to as well as lesser area IEEE 754 Double Precision floating point architecture than IEEE Single Precision floating point architectures. However, its power consumption is almost triple than that of IEEE Single Precision floating point architecture. Posit (32,4) achieves better accuracy than Floating point 64 bit with lower power and area overhead. Posit(32,4) also meets almost same timing and has lower $Area * Power * Delay$ product. The optimized low power, economical implementation of Single Precision IEEE-754 has far greater performance than all of them with accuracy errors within tolerable range.

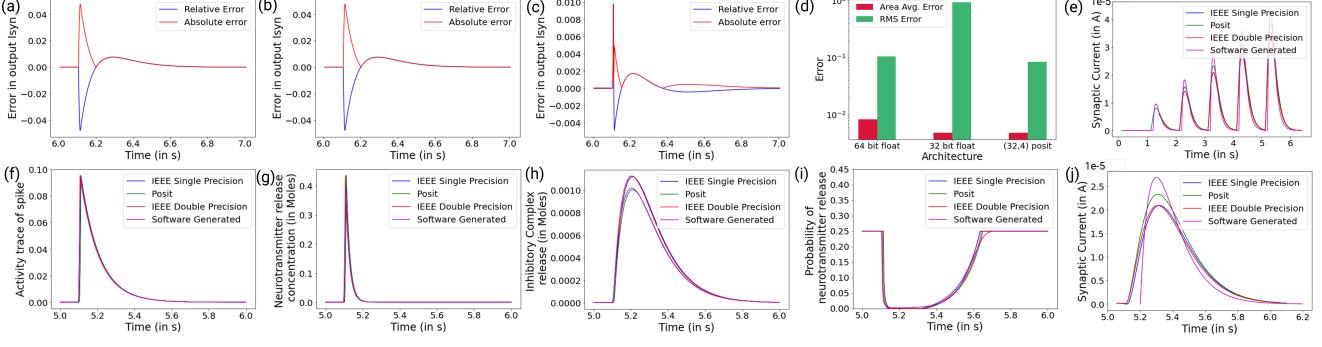


Fig. 5. Absolute and Relative Error obtained for (a) IEEE Single Precision, (b) IEEE Double Precision and (c) the universal posit (32,4) execution cores in the hardware architecture is shown in the figure. Their RMS Error and Area Avg. Error are compared in (d). The results obtained for these separate operation cores are compared with simulation results of RMMP model in (e) for the output synaptic current for an impulse train and (f)-(j) for the response of primary concentration variables involved for one impulse.

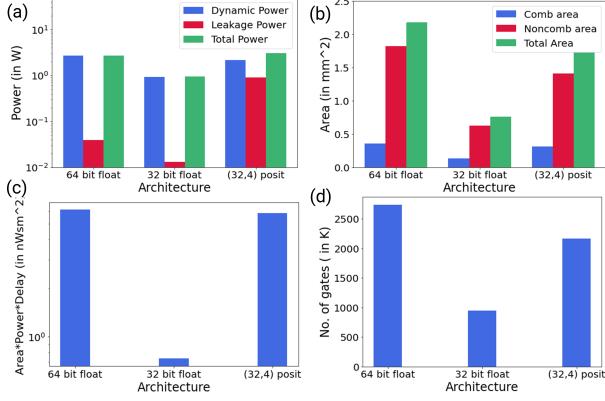


Fig. 6. Comparison of operation and performance parameters across the three operation cores of PreSync hardware realisation

VI. NETWORK MODEL

A. Long Term Potentiation

The long-term potentiation (LTP) indicates a continuous intensification of signal transmission between presynaptic and postsynaptic neurons based on recent patterns of synaptic activity. LTP refers to a permanent increase in synaptic strength and is known as a molecular mechanism of learning and memory in our Central Nervous System [4], [5]. For the PreSync neuron model, the influx of retrograde messenger into the presynaptic region is considered to be a constant value [Table I] such that it is just above the limiting threshold. If this value is increased, then the gain in synaptic weight corresponding to an input impulse is also larger.

B. Extension to PreSync Model

This PreSync model does not involve any dynamic processes in the synapse. Hence, a bipartite system that defines the postsynaptic region of a neuron is exhaustive and sufficient to create a functional neuron system. A network of such a neuron system can be used for information processing via SSN processing methods.

Fig. 7 describes a neuron network constructed using this an extenuron and damage parameters are being fed to it via an external output. The results are obtained across the neurons at the extremities of the neuron network to find the effective variation to their output in response to the input to Neuron 2.

VII. CONCLUSION

There is an increasing need for high adaptability for modelling various scenarios that we meet here by an elaborate description of presynaptic neural dynamics pathways in the RMMP model. Our proposed PreSync model has lower hardware costs and low error margins compared to the RMMP neuron model that makes it optimal for large scale digital implementation in neuronal network architectures. Hardware results show that posit (32,4) outperforms even IEEE Single Precision mode in terms of minimum AAE. The optimized low power, economical implementation of Single Precision IEEE-754 has the greatest performance with accuracy errors within tolerable range.

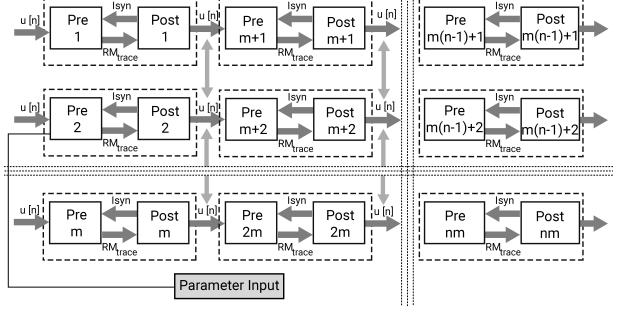


Fig. 7. Extended Network Model developed along with the postsynaptic region of a neuron. Parameter input refers to the parameter whose variation effects in a network are being studied.

Hence, posit version has room for extensive scaling for high precision and error-intolerant applications while the optimized Single Precision Float version is most suitable for high performance applications.

REFERENCES

- Citri, A. and Malenka, R. "Synaptic Plasticity: Multiple Forms, Functions, and Mechanisms." *Neuropsychopharmacology* vol. 33, pp. 18–41, 2008.
- Shunsuke S. and Shinji D. "Response characteristics of the BVP neuron model to periodic pulse inputs", *Mathematical Sciences*, vol. 112, issue 2, pp. 243–259, 1992.
- Nagumo J., Sato S. "On a response characteristic of a mathematical neuron model." *Kybernetik*, vol. 10.3, pp. 155–164, 1972.
- Gaël D. and Dominique D. "Long-Term Plasticity of Intrinsic Excitability: Learning Rules and Mechanisms" *Cold Spring Harbor Lab.: Learn. Mem.*, vol. 10, pp. 456–465, 2003.
- Tetzlaff C., Kolodziejczyk C., Markelic I., and Wörgötter F. "Time scales of memory, learning, and plasticity." *Biol. Cybern.*, vol. 106, pp. 715–726, 2012.
- L. Wittie, H. Memelli, and R. Welsch, "Billion neuron memory models in slender blue genes," in Program 208.30/MMM21, 2010 Neuroscience Meeting Planner. Society for Neuroscience San Diego, CA, 2010, p. 1.
- V. Garg, R. Shekhar and J. G. Harris, "Spiking Neuron Computation With the Time Machine," in *IEEE Transactions on Biomedical Circuits and Systems*, vol. 6, no. 2, pp. 142–155, April 2012.
- S. B. Furber, F. Galluppi, S. Temple, and L. A. Plana, "The spinnaker project," *Proceedings of the IEEE*, vol. 102, no. 5, pp. 652–665, 2014.
- B. V. Benjamin, et al., "Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations," *Proceedings of the IEEE*, vol. 102, no. 5, pp. 699–716, 2014.
- T. Pfeil, et al., "Six networks on a universal neuromorphic computing substrate," *Frontiers in neuroscience*, vol. 7, p. 11, 2013.
- Faghhihi, F., and Moustafa, A. A. "The dependence of neuronal encoding efficiency on Hebbian plasticity and homeostatic regulation of neurotransmitter release". *Front. Cell. Neurosci.* 9:164, 2015.
- Clopath, C., et al. "Connectivity reflects coding: a model of voltage-based STDP with homeostasis." *Nat. Neurosci.* vol 13, 344–352, 2010
- Postnov, D. E., et al., "Dynamical patterns of calcium signaling in a functional model of neuron-astrocyte networks," *J. Biol. Phys.*, vol. 35.4, pp. 425–445, 2009
- M. K. Jaiswal and H. K. - So, "PACoGen: A Hardware Posit Arithmetic Core Generator" in *IEEE Access*, vol. 7, pp. 74586–74601, 2019
- J. L. Gustafson, I. T. Yonemoto, "Beating Floating Point at its own game: Posit Arithmetic" *Proceedings of the IEEE*, vol. 102, no. 5, pp. 699–716, 2014
- John L Gustafson. *The End of Error: Unum Computing*, vol. 24. CRC Press, 2016.