Neuromorphic Context-dependent Learning Framework with Fault-tolerant Spike Routing

Shuangming Yang, *Member, IEEE*, Jiang Wang, Bin Deng, Mostafa Rahimi Azghadi, *Senior Member, IEEE*, Bernabe Linares-Barranco, *Fellow, IEEE*

***Abstract*—Neuromorphic computing is a promising technology that realizes computation based on event-based spiking neural networks (SNNs). However, fault-tolerant on-chip learning remains a challenge in neuromorphic systems. This study presents a scalable neuromorphic fault-tolerant context-dependent learning (FCL) framework with reinforcement learning mechanism, which integrates both learning and fault-tolerant capabilities in a unified system. We show how this system can learn associations between stimulation and response in context-dependent tasks, which are inspired by the biological hippocampus-mPFC network. Furthermore, we demonstrate how our novel fault-tolerant neuromorphic spike routing scheme can avoid multiple fault nodes successfully, and can enhance the maximum throughput of the neuromorphic network by 0.9% to 16.1% in comparison with previous studies. By utilizing the real-time computational capabilities and multiple-fault tolerant property of the proposed system, the neuronal mechanisms underlying the spiking activities of neuromorphic networks can be readily explored. In addition, the proposed system can be applied in real-time learning and decision-making applications, brain machine integration, as well as the investigation of the brain cognition during learning.**

***Index Terms*—context-dependent learning, neuromorphic computing, spiking neural network (SNN), brain inspired, fault tolerant**

I. Introduction

The encoding and remembering of an event context relies on the episodic memory of the brain when observing an object or item [1]. The context can be an absolute time, relative time based on other events that happened before or after, or a specific place [2-3]. Previous studies have shown that the interactions of neocortical and hippocampal circuits can enable contextual learning during an item-reward association task [4-5].

The hippocampal firing activities are affected by the context at which a current task is performed [6]. In a sampling or an encoding phase of the task, the selectivity appears among the hippocampal CA1 neurons during a discrete delayed nonmatch-to-place task [7]. In addition, neural firing activities depend on the start or anticipated end location of a trajectory [8]. A previous experiment has revealed that neurons in hippocampus develop selectivity towards specific items in an abstraction of spatial context [4]. The generation of behaviors based on the contextual representations depends on both the hippocampus and medial prefrontal cortex (mPFC) [9]. These studies reveal that context encoding may involve the interaction between hippocampus and mPFC.

The experimental setup and preparation for the above-mentioned studies are usually time-prohibitive and involve interacting with live subjects. One approach to facilitate these experiments and improve our understanding of the context-dependent learning is to build a computational model utilizing brain-inspired SNN models with neural spike representation. In this study, the concept of context-dependent learning refers to the field of neuroscience, and concept means the environmental sensory cues that are processed and learned in the hippocampus [53]-[58]. Gulli *et al.* used monkeys to complete an associative memory task in the virtual environment for the investigation of the context-dependent representation of objects and space in hippocampus [53]. The context was defined by a texture applied to the maze walls. Zhao *et al.* suggested that neurons in the hippocampus undergo context-dependent learning because they inherit different input patterns from pre-synaptic areas in different contexts, which occurs during decision making and navigation tasks [54]. Lee *et al.* pointed out that navigation, context-dependent learning and episodic memory are produced in a recurrent collateral circuitry in the hippocampus [55]. In addition, the mechanisms context-dependent learning are explored in a series of neuroscience studies, using both behavioral paradigm and physiological observation [56]-[58]. The proposed model with FCL framework can then be implemented on a neuromorphic architecture, not only to better understand the brain, but also to use it in various categories of applications such as robotics and brain computer interfaces.

This work was supported partly by the National Natural Science Foundation of China (Grant Nos. 62071324, 62006170), and in part by China Postdoctoral Science Foundation (Grant No. 2020M680885).

Shuangming Yang, Jiang Wang, and Bin Deng are with School of Electrical and Information Engineering, Tianjin University, Tianjin, 300072 China.(corresponding email: yangshuangming@tju.edu.cn).

Mostafa Rahimi Azghadi is with the College of Science and Engineering, James Cook University, Townsville, QLD 4814, Australia (e-mail: mostafa.rahimiazghadi@jcu.edu.au).

B. Linares-Barranco is with the Microelectronics Institute of Seville, Seville 41092, Spain (e-mail: bernabe@imse-cnm.csic.es).

Neuromorphic computing is a promising approach towards non-von Neumann systems for neuroscience and artificial intelligence applications including the formulation of hypotheses regarding the function of neural systems, validation of self-consistency in the description of neural phenomena or function, neural computation instead of traditional computing structures, and biologically inspired engineering applications [10, 12, 13, 46]. Some of the large-scale neuromorphic systems used in these applications include BrainScaleS, TrueNorth, and SpiNNaker [10-13]. In addition to these large-scale systems, several other neuromorphic systems have been developed in the literature [14, 44, 48]. We have previously developed large-scale conductance-based spiking neural networks (LaCSNN), which is a digital neuromorphic system designed for simulating SNNs using multicasting address event representation (AER) with 3D network-on-chip (NoC) architecture [14]. Compared to state-of-the-art neuromorphic designs with similar capabilities, LaCSNN provides significant benefits in both biological accuracy and reconfigurability [14]. It is able to realize a large-scale SNN with one million biologically plausible neurons in real time.

When developing any large-scale neuromorphic system, such as those mentioned above, two main capabilities are required. These include online learning capability and fault-tolerant operation capability. It is also important to implement a system that integrates these two capabilities. This study focuses on implementing a neuromorphic system named FCL, for modeling large-scale SNNs with online fault-tolerant context-dependent learning. It abstracts the mechanisms from both hippocampus and mPFC, and realizes the learning capability in a context-dependent task responding to item reward. To the best of our knowledge, this paper presents the first scalable fault-tolerant context-dependent learning framework.

The remainder of the paper is organized as follows. Section II introduces the fault-tolerance considerations in neuromorphic systems. The implemented network model for context-dependent learning is presented in Section III, while Section IV describes our digital neuromorphic architecture in detail. Section V proposes the fault-tolerant algorithm and methodology for the presented neuromorphic system. Experimental results of the digital neuromorphic system are presented in Section VI. Section VII discusses the advantages of the proposed neuromorphic model compared to state-of-the-art. The paper is concluded with discussion on future works, in Section VIII.

II. Fault-tolerance Considerations in Neuromorphic Systems

State-of-the-art neuromorphic systems have used different architectures for the realization of SNNs, which are shown in Fig. 1. For the non-fault-tolerant neuromorphic systems, there are three conventional architectures, including shared bus, 2D NoC, and 3D NoC. The shared bus architecture can support both multicast and broadcast routing with low-cost SNN models, but is constrained by its limited scalability [15], [16]. A number of studies have focused on the 2D NoC architecture for neuromorphic systems, including H-NoC [17], Neurogrid [11], HiAER [13], and Truenorth [18]. The H-NoC architecture is based on the EMBRACE system using the leaky integrate-and-fire neuron model [17]. Analog implementation is used to realize the calculation of the ionic dynamics in Neurogrid and HiAER systems [11], [13], while fully digital method is used in SpiNNaker, Truenorth and Tianjic [10], [18], [19]. To enable the implementation of larger scale SNNs, 3D NoC architecture is used in several works. In the first work, a multicast AER architecture is developed with biologically plausible conductance-based neuron models on LaCSNN system, which has the intrinsic mechanisms underlying the neuronal spiking activities within large-scale multi-nucleus networks [14]. In other works, a multicast routing scheme is used in 3D mesh NoC architecture in KMCR [20], and multi-compartment conductance-based neuron models are used in the IBFT-based CMN system, which can further enhance the system scalability in comparison with the previous works [21].

As more neuromorphic designs are developed, fault-tolerance becomes essential and critical for reliable neuromorphic computing. Recently, a number of studies have focused on fault-tolerant neuromorphic designs. Notably, SpiNNaker system presents a novel routing strategy to deal with the problems of congested or broken links in its digital neuromorphic architecture [22]. FTSP-KMCR proposes a multicast fault-tolerant architecture to implement a neural engineering framework [23]. However, there is a lack of fault-tolerant learning methodology for neuromorphic computing, especially based on brain-inspired learning mechanisms. To that end, this study proposes a fault-tolerant context-dependent learning (FCL) model as well as an AER multicast routing strategy on the IBFT architecture, which enables fault-tolerant context-dependent neuromorphic learning.

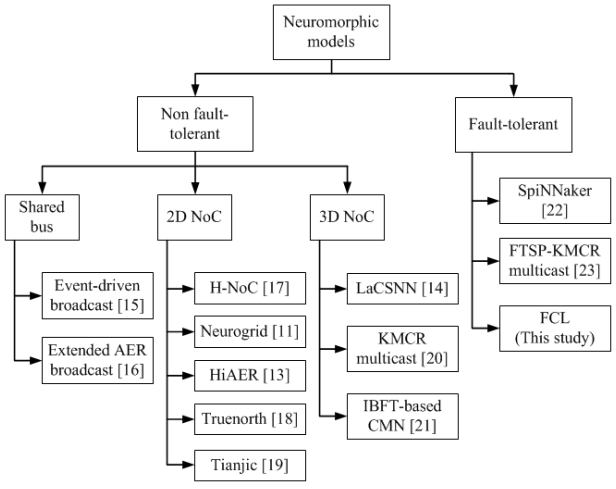


Fig. 1. Overview of the current neuromorphic models considering fault-tolerant properties.

III. Network Model for Context-dependent Learning

The network model for the implemented context-dependent learning mechanism is shown in Fig. 2 (a). This network is composed of three layers including sensory, hippocampal and motor layer. The neurons and learning synapses in these layers are explained in the following subsections.

A. Neuron Model

The neuron model used in our implementation is based on the leaky integrate-and-fire (LIF) model. In this model, the membrane potential V_i is governed by capacitance C and driven by the input current I_j, and leaking current is affected by a leaky channel of conductance G_l. The membrane has the resting potential V_{rest} and is influenced by small fluctuations of a noise term η. The noise term here denotes a random variable $\eta \in N(\mu, \sigma)$ based on a Gaussian distribution with mean value μ=0 and standard deviation σ. The dynamic equation of the membrane voltage can be expressed as

$$C\frac{dV_i}{dt} = G_l\left(V_i - V_{rest}\right) + I_k + \eta\,, \tag{1}$$

where i=1…n_k and $k \in$ {*sensory*, *hippo*, *motor*} is one of the network layers. The input current for the sensory layer is $I_{sensory}$=1.00 nA, while it is I_{hippo}=0.98 nA for the hippocampal layer, and I_{motor}=0.96 nA for the motor layer. The model parameter values are listed in Table I, accordingly. The model parameter values fit the empirical behavioral data of the context-dependent task (Komorowski et al., 2009).

TABLE I

Parameter values of neuron model.

Parameter	Description	Value
C	Membrane capacitance	5.5×10^{-9} F
G_l	Leaky membrane conductance	10×10^{-9} S
V_{peak}	Peak membrane potential	0 mV
V_{th}	Threshold membrane potential	-50 mV
V_{reset}	Reset membrane potential	-70 mV
σ	Standard deviation of Gaussian noise	1 μV per step

B. Synaptic connections and learning algorithm

In order to realize weight adaptation for learning, spike-timing dependent plasticity (STDP) rule for synaptic modification is used [24]. Synaptic weights between the sensory layer, the hippocampal layer, and the motor layer are modified. The rule employs the time difference Δ between the pre-synaptic and post-synaptic spikes. If the pre-synaptic spike occurs before the post-synaptic spike, it induces a positive time difference $\Delta > 0$, resulting in a synaptic long term potentiation (LTP). When the pre-synaptic spike happens after the post-synaptic spike, this results in a negative time difference $\Delta < 0$, inducing a synaptic long term depression (LTD). This effect occurs within a small time window of ≈20 ms, with the weight dynamic range between $w_{\min}$=0 and $w_{\max}$=1. The STDP learning rule can be implemented as a differential equation as follows

$$\tau_w \frac{dW_{ij}^{exc}}{dt} = \left(w_{\max} - W_{ij}^{exc}\right)\cdot A_+ \exp\left(-\Delta/\tau_+\right) - \left(w_{\min} - W_{ij}^{exc}\right)\cdot A_- \exp\left(+\Delta/\tau_-\right)\,, \tag{2}$$

where i=1…n_k, j=1…n_l, $k \in$ {*sensory*, *hippo*, *motor*}, $l \in$ {*sensory*, *hippo*, *motor*} and $k \neq l$. Indices "i" and "j" represent neurons from two connected layers, for instance, the hippocampal layer with motor layer. The weight alterations, which may result in LTP and LTD are controlled by the time constants τ_w, τ_+, and τ_- respectively. The parameter values of the implemented STDP learning rule are listed in Table II.

TABLE II

Parameter values of spike-timing dependent plasticity (STDP).

Parameter	Description	Value
τ_+	Pre- before post-synaptic spike time constant	10 ms
τ_-	Pre- after post-synaptic spike time constant	10 ms
A_+	Pre- before post-synaptic spike amplitude	+1.2
A_-	Pre- after post-synaptic spike amplitude	-0.4
$w_{\min}$	Minimum activation for synaptic weight	0.0
$w_{\max}$	Maximum activation for synaptic weight	1.0
τ_w	Learning rate for weight adaptation	10 ms

Neural spikes are transmitted between layers via the excitatory weights W^{exc} and inhibitory weights W^{inh} as shown in Fig. 2(a). Winner take all (WTA) rule is used at the receiving terminal to generate a current of I_{hippo}=0.98 nA for neurons in the hippocampal layer or I_{motor}=0.96 nA for neurons in the motor layer. The WTA rule is defined as follows

$$I_{j^*} = I_k \;\text{ if } j^* = \arg\max_j \left\{ \begin{array}{l} \sum_{i=1}^{n_k}\left(V_i - V_{reset}\right)W_{ij}^{exc} \\ -\sum_{\substack{i=1 \\ i\neq j}}^{n_k}\left(V_i - V_{reset}\right)W_{ij}^{inh} \end{array} \right\}, \quad \text{and } j = 1...n_l \tag{3}$$

where the input current $I_j = 0$. Similar to Eq. 2, here $k \in$ {*sensory*, *hippo*, *motor*}, and $l \in$ {*sensory*, *hippo*, *motor*}.

C. Network architecture

The presented SNN model is inspired by visual, odors, and tactile sensory inputs in the form of binary vectors [26]. The

sensory signals are delivered through six input neurons, i.e. $n_{sensory}$=6. Four of them provide context-place information, and the other two provide item information. As shown in Fig. 2(a), the first input neuron is activated with context-place A1 and the second input neuron is activated with context-place A2, while the third and fourth neurons are activated with context-place B1 and B2, respectively. The fifth and sixth neurons are activated with item information *X* and *Y,* respectively. The input neurons are connected to hippocampal neurons using adaptive weights W^{exc} to represent excitatory connections in an all-to-all setting. The hippocampal neurons have inhibitory connections W^{inh} among them, but not inhibiting themselves. Fig. 2(a) only draws the connections of 1st hippocampal cell and the 1st output cell. The hippocampal cells are connected to two motor output neurons with adaptive weights using all-to-all connectivity. All the weights are uniformly randomly initialized with values in the range 0 to 1. The two cells in the motor layer represent the action "digging" and "moving", respectively.

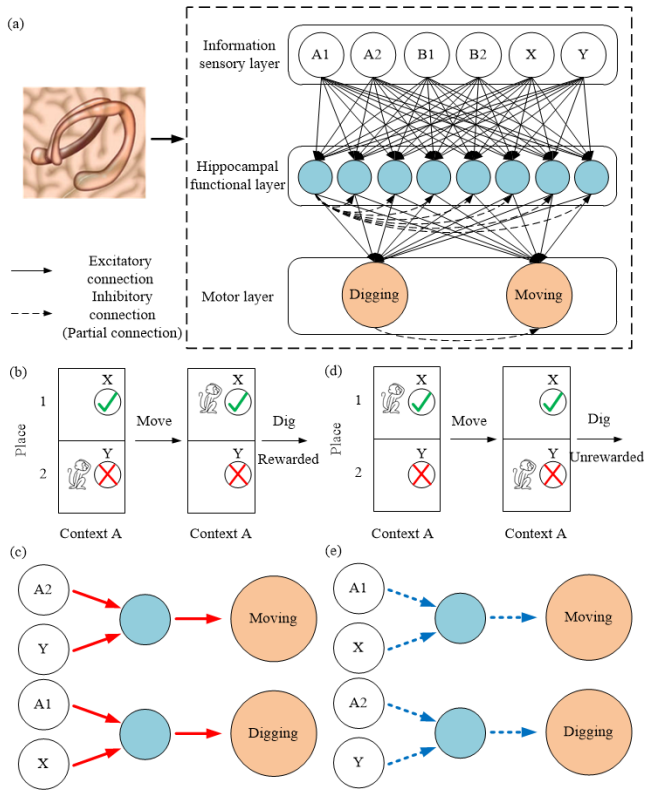


Fig. 2. The proposed learning scheme for context-dependent task. (a) The schematic structure of the implemented SNN model for the context-dependent task. (b) The rewarded action sequence. (c) Rewarded action sequence leads to synaptic enhancement denoted by red solid arrow. (d) The non-rewarded action sequence. (e) Non-rewarded action sequence leads to synaptic depression denoted by blue dotted arrow.

Fig. 2(b) shows a cartoon of the context-dependent learning at hand. Here, a model monkey can only move between place 1 and 2 in either context A or B, without intermediary places. It can perform no action to change the context, but the context can change randomly between trials. Some trials can start the model in context A and others in context B. Table III lists the model parameters used in our experimentation.

TABLE III

PARAMETER VALUES OF THE SNN MODEL USED IN THE TARGETED CONTEXT-DEPENDENT LEARNING TASK.

Parameter	Description	Value
T_{trial}	Maximum time interval for a trial	10 ms
T_{replay}	Maximum time interval for replay	400 ms
Δt	Time increment per simulation step	+1.2
$I_{sensory}$	Input current for sensory neuron	1.0 nA
I_{hippo}	Input current for hippocampus neuron	0.98 nA
I_{motor}	Input current for motor neuron	0.96 nA

D. Learning of the context-dependent task

The sequence of monkey actions can be represented using Fig. 2(b) and Fig. 2(d). The two square areas in the figure represent place 1 and place 2 of context A. In each square area, there is a ball representing item, item X and item Y respectively. If there is a green check in the ball, it means that the reward is hidden under this item. If there is a red cross in the ball, it means that there is no hidden reward. As shown in Fig. 2(b), a case of a rewarded action sequence is A2Y, move, A1X, dig, and receive a reward. In the examples in Fig. 2(b) and Fig. 2(c), the learning in the first stage is assumed to occur to establish the correct connection to activate the first hippocampal neuron, inducing the synaptic activation of the neurons encoding the action "move". This neuron spikes several times, inducing the action "move" executed by the monkey. After moving, the monkey can sense the place A1 and item X, which activates a hippocampal neuron because of the established connection, resulting in the activation of the neuron coding the action "dig", and the monkey obtain the reward. Since in the first stage, the monkey has dug a reward, this action can be enhanced in the next stage, which is represented by the red solid lines as shown in Fig. 2(c). For the procedure without reward, the monkey is first located at A1X as shown in Fig. 2(d), and then moved to the location of A2Y and dug. Since the reward is at A1X, the monkey cannot obtain the reward. Thus, this action will be depressed in the next stage, which is represented by the blue dashed lines in Fig. 2(e).

IV. Digital Neuromorphic Architecture

A. Network-on-chip (NoC) architecture

The proposed FCL model is realized using FPGA and evaluated based on its average spike latency and throughput. Here, the FCL model for the targeted context-dependent learning task (shown in Fig. 2) is mapped onto the LaCSNN digital neuromorphic system in a layer-to-layer fashion, where our proposed routing algorithm is combined with the 3D mesh NoC topology. The mapping strategy of SNNs onto NoC-based

neuromorphic systems is critical for neuromorphic applications, because it significantly influences both the overall performance and the power consumption. Fig. 3 shows the mapping of the context-dependent task to the 3D NoC architecture, in which neurons in the same network layer are mapped onto the same architecture layer. In a previous study with fault-tolerant spike routing, neurons can only send spikes to the next layer [25]. In this study, this limitation is eliminated by using a novel mapping and NoC architecture. In the first layer, six neuron units representing the six neurons in the information sensory layer are distributed in parallel, and are fully connected with the neuron units in the second layer. In the second layer, the neuron units are implemented on an 8×8 NoC architecture, which uses mesh-based multicasting AER strategy. The moving and digging neurons are implemented digitally in the third layer. All the three network layers are fully connected. The spike event packet transferred in this network contains 22 bits, which include 3-bit layer ID, 1-bit AER data, 3-bit Y_dest address, 3-bit X_dest address, and 12-bit Timestamp.

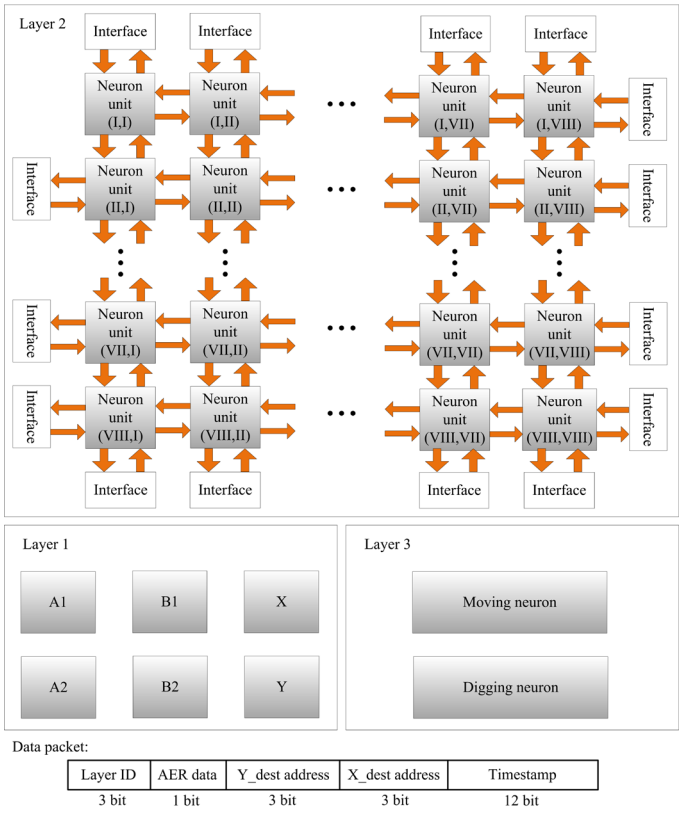


Fig. 3. 3D NoC architecture of the proposed neuromorphic network.

The detailed digital architecture of the neuron unit is shown in Fig. 4(a). It contains a neuron processor, a fault-tolerant router, a synapse unit, and a configuration unit. Each router has six ports including up, down, north, west, east and south to route the AER packets to another neuron unit. Compared to a previous study [27], the proposed architecture has three features including 1) computation using events with synaptic weighting; 2) implementation of physical synapses; and 3) fault-tolerant neuromorphic routing capability. Therefore, the proposed architecture is more suitable for the neuromorphic SNN computation aiming at complicated cognitive behaviors, such as context-dependent learning.

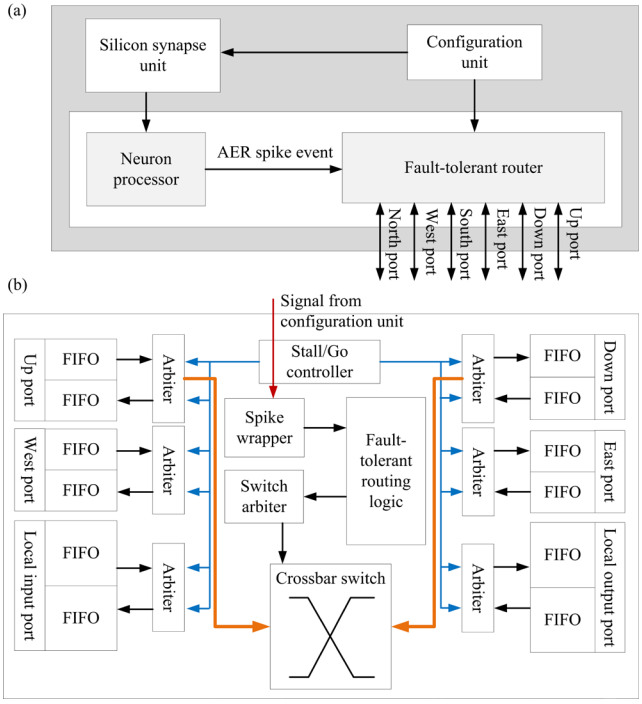


Fig. 4. The detailed digital neuromorphic architecture of the neuron unit and routers in the proposed decision-making spiking network. (a) Digital neuromorphic architecture of the neuron unit. (b) Digital neuromorphic architecture of the fault-tolerant router.

B. Fault-tolerant router architecture of the proposed neuromorphic network

Router is critical in the proposed neuromorphic architecture, because it plays vital roles in achieving the fault-tolerance targeted. The fault-tolerant multicast 3D router architecture implemented in the neuron units of our system is shown in Fig. 4(b). At the first stage, the spike events are received from the four neighboring nucleus processors and their packets are stored in the input buffer before being processed. The spike wrapper unit is used to convert a single spike event into a valid AER spike packet using the information received from the configuration processor. This processor can be started at any time based on the neuronal connectivity. The configuration processor contains four types of registers, which are chip address register, layer address register, node address register and timestamp register. Incoming spike events and the corresponding deliver-at time are stored in the on-chip memory after the deliver-at time stamps are reached. Then the source address of the packet is extracted and calculated to determine the output port. The fault-tolerant routing calculation, which will be introduced in Section V, is then used to route the packet. The switch arbiter is used using least recently served priority to provide fast computation, inexpensive implementation and strong fairness as presented in previous studies [28]. Finally, the packet is sent to the desired output port through the crossbar.

The crossbar switch is controlled by the switch arbiter and implemented by multiplexers.

C. Digital architecture of the conductance-based LIF neuron model

In this study Euler method is used for the discretization of the neuron model since it can save hardware resources compared to Runge-Kutta method. Based on Euler method, the original equations can be transformed into the following equations:

$$\begin{cases} V(n+1) = \Delta t \cdot \left(\frac{1}{C} \left(G_l \left(V_i - V_{rest} \right) + I_k + \eta \right) \right) + V(n) \\ W_{ij}^{exc}(n+1) = \Delta t \cdot \begin{pmatrix} \left(w_{\max} - W_{ij}^{exc} \right) \cdot A_+ \exp\left(-\Delta / \tau_+ \right) \\ -\left(w_{\min} - W_{ij}^{exc} \right) \cdot A_- \exp\left(+\Delta / \tau_- \right) \end{pmatrix} + W_{ij}^{exc}(n) \end{cases} \quad (4)$$

The digital architecture of the neuron model in the information sensory, hippocampal and motor layers is shown in Fig. 5(a). Multipliers are extravagant hardware resources in digital design, and are usually avoided as much as possible to gain energy and hardware cost benefits. Thus, in the proposed digital architecture, shift logic multipliers (SLM) are used to replace multipliers to realize multiplication operations. Fig. 5(b) shows the detailed digital implementation of the SLM block, which is used to replace the multipliers in this study.

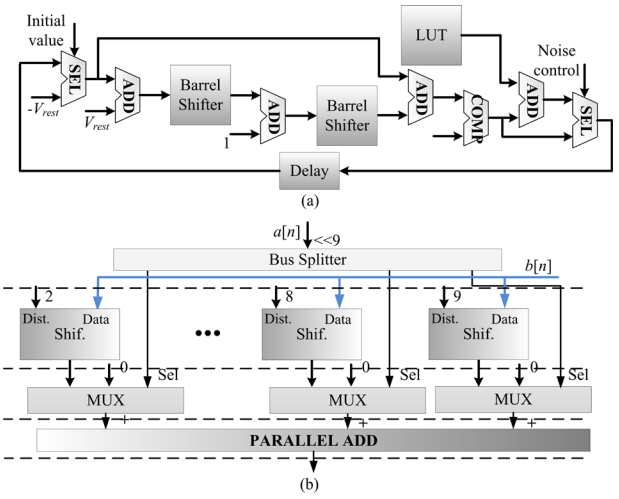


Fig. 5. Detailed digital architecture of the neuron processor and the SLM block. (a) Detailed digital implementation of the neuron unit. (b) Detailed digital implementation of the SLM block.

D. Digital architecture of the synapse module implementing STDP

The learning capability of the proposed FCL model is based on STDP learning algorithm. The detailed digital implementation of this learning algorithm is shown in Fig. 6. As demonstrated in Fig. 6(a), the pre-synaptic and post-synaptic spike timings "Timepre" and "Timepost" are first calculated, which are then used in the digital implementation of the STDP learning rule shown in Fig. 6(b). Here, LUTs are used to calculate the exponential part in the STDP algorithm, and barrel shifters are used to replace the required multipliers. These components help to significantly cut down the hardware resource cost and power consumption.

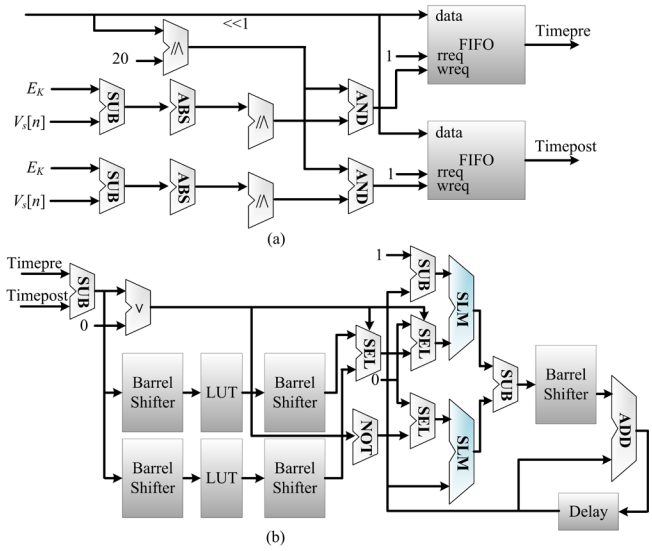


Fig. 6. Digital architecture of the synapse. (a) Detailed digital implementation for the computation of "Timepre" and "Timepost". (b) Digital implementation of the STDP weight updating module.

V. Fault-tolerant Spike Routing

Several previous studies have investigated fault-tolerant routing schemes for NoC topologies [22, 23]. These schemes are based on different approaches including virtual channels, path-finding, and bypass methods to perform efficient routing in the presence of faulty nodes. The virtual channels based, path-finding based and bypass-based routing schemes are different in terms of the hardware resource cost and circuit complexity. The routing schemes based on virtual channels divide a single physical link into several virtual channels, but they require complex control circuitry and cost large hardware resource and power consumption. The routing schemes based on path-finding methods need a large number of routing tables, which induce large hardware resource cost. Contrarily, bypass methods divide the fault nodes into non-overlapping fault areas, and make use of the normal nodes and links around the fault areas to form a new routing path. When the data packet reaches a fault area, it will be bypassed along a new routing path based on certain rules, thus avoiding the fault without deadlock. Here, we utilize the bypass method without virtual channels to perform fault-tolerant routing in the proposed neuromorphic architecture for the targeted context-dependent learning task. Built-in self-test (BIST) technique is used to get the location information of the fault node. Furthermore, the realization of the load balancing in the bypass loop and the reduction of the communication latency can be performed by optimizing the routing distance of the spike events.

A. Multiple-fault tolerant neuromorphic (MFTN) algorithm

In the proposed algorithm, XY routing strategy is used, which means the neural information is routed first along the X direction, then along the Y direction. So the proposed algorithm is divided into two parts, which are along X and Y directions respectively. In a previous study, Chen and Chiu have presented an essential fault-tolerant solution for NoC design [49]. However, Chen's algorithm requires significant computations and a large area, with a low node utilization. In addition, its communication load is heavy and its latency is high. On the contrary, our presented MFTN algorithm modifies and expands the single-fault bypass method in Zhang's algorithm [47] to improve the multiple-fault situation. The MFTN algorithm along both X and Y directions will be introduced and compared to Zhang's algorithm to further illustrate its improvement based on Zhang's algorithm in this section. The experimental results will be shown in Section VI and compared to Chen's algorithm to demonstrate its performance improvement.

In the case of multiple-fault tolerant schemes, there are two scenarios. In the first scenario, the source and destination nodes are located on different sides of the fault region and the source node is located in one of the rows of the fault region. For examples, see source 1 (S1) and Destination 3 (D3), or S2 and D5 in Fig. 7. In the second case, the source node is located within the rows of the fault region and the destination node is located in the columns of the fault region. For example of this case, see the positions of S1 and D2 in Fig. 7. As shown in this figure, we first delineate the fault regions that need to be bypassed, which include the fault nodes and the unsafe nodes surrounded by the fault nodes. The coordinate information of the four SW, NW, SE and NE nodes is transmitted to all the normal nodes in the corresponding column by all the nodes in the bypass loop, and stored in the on-chip memory to determine whether the routing process passes through the fault region. In Fig. 7, the solid arrow is based on Zhang's algorithm [47], and the dotted line arrow is based on the proposed MFTN routing algorithm. In comparison with Zhang's algorithm, the proposed algorithm optimizes the bypass strategy, which decreases the routing distance accordingly. For example, in the case from S1 to D4, Zhang's algorithm will turn up at SE, which induces longer distance. In contrast, the proposed MFTN algorithm will provide a direct route to D4 node. The pseudo code of the proposed MFTN algorithm is shown in Fig. 8. In the proposed pseudo code, C and D represent the current and destination nodes respectively. In the XY algorithm, the neural information will be routed along the X direction at first, and then routed along the Y direction. It is majorly for the case where the source and destination nodes are located on east and west sides of the fault region, separately. The bypass loop is separated from the original pathway when certain conditions are satisfied, therefore the load of the bypass loop and the routing distance are decreased.

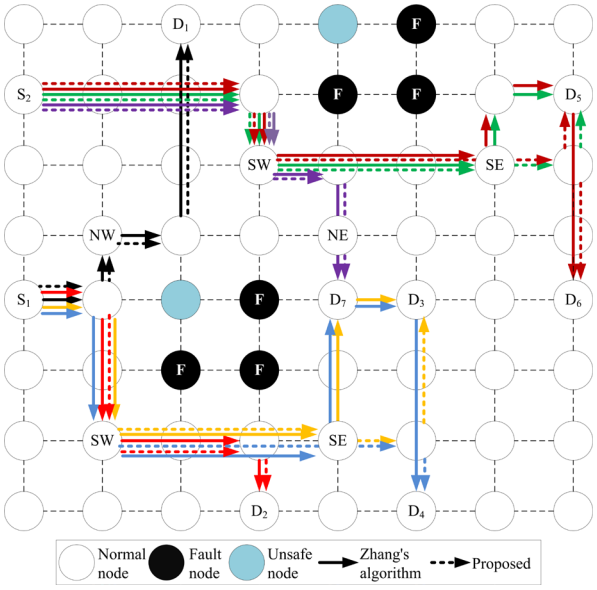


Fig. 7. The bypass route of the spike event along X direction.

Pseudo code for single-fault neuromorphic tolerant algorithm along X direction

```
if (C is destination D)
      {Processing the spike event ; }// Reaching the destination node
else if (fault information memory of C is empty) {
            Continue route according to XY algorithm;
            }// If the fault information is NULL, routing based on XY algorithm
else if (fault information memory of C will not affect routing from S to D ) {
            Continue route according to XY algorithm;
            }
          // If the fault region is not located on the pathway of XY routing,
continue routing based on XY algorithm
else if (X_SW<=X_D<=X_SE && Y_SW<=Y_S<=Y_NW){
            Route according to Zhang algorithm;
            }
else if (Y_SW<=Y_S<=Y_NW && (X_D<X_SW<X_SE<=X_S ||
X_S<=X_SW<X_SE<X_D) ){
                  // When satisfying the first case, begin optimization mode
            if (!(X_D<X_SW<X_SE<=X_C || X_C<=X_SW<X_SE<X_D)){
                  Route to X_D according to Zhang algorithm;
                        // Routing to the column of D based on Zhang's
algorithm
                  Continue route according to XY algorithm;// Routing spike
event based on XY algorithm
                  }// C and D do not locate on the both sides of the fault region
            else {
                  do {
                        Route to next node according to Zhang algorithm;
                              // Routing to the next node based on Zhang's
algorithm
                  }while (!(X_C==X_SW<X_S || X_C==X_SE>X_S));
                  Continue route according to XY algorithm;// Continue routing
based on XY algorithm
                  }// C and D locate on the both sides of the fault region
            }
```

Fig. 8. Pseudo code for MFNT algorithm along X direction. C and D represent the current and destination nodes respectively. In the XY algorithm, the neural information will be routed along the X direction at first, and then routed along the Y direction.

The situation for the bypass along Y direction in the case of multiple fault nodes is that the destination and source nodes are located on south and north sides of the fault region respectively

and the destination node is located within the columns of the fault region. According to Zhang's algorithm, more change of routing direction is required along X direction, which induces the increment of the bypass distance and the load is majorly on the left side of the bypass loop. When current node (C), is on the north side of the bypass, the spike event is first routed to the nearest corner based on XY algorithm and then bypassed based on Zhang's algorithm. As shown in Fig. 9, if the current node is on the south side of the bypass loop, the spike event is first routed to the SW node due to the prohibition of the turning on the NE corner, and then bypassed based on Zhang's algorithm. The pseudo code for the proposed MFNT algorithm along Y direction is shown in Fig. 10.

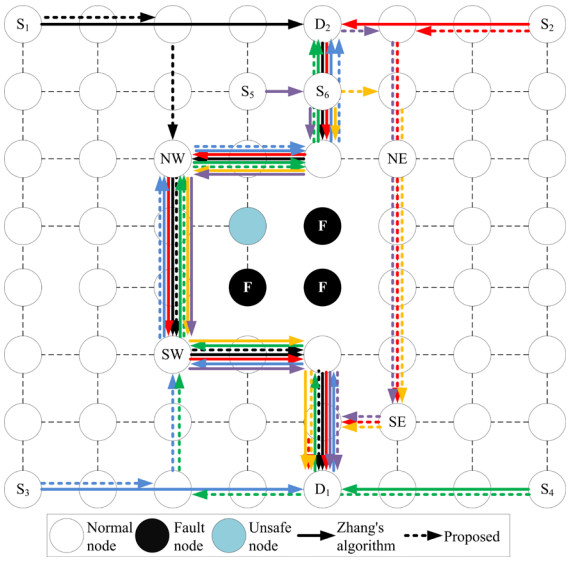


Fig. 9. The bypass route of the spike event along Y direction.

Pseudo code for multiple-fault neuromorphic tolerant algorithm along Y direction

```
if (C is destination D)
      {Processing the spike event ; }// Reaching the destination node
else if (fault information memory of C is empty) {
                Continue route according to XY algorithm;
          }// If the fault information is NULL, routing based on XY
algorithm
else if (fault information memory of C will not affect routing from S to D ) {
                Continue route according to XY algorithm;
          }
       // If the fault region is not located on the pathway of XY routing,
routing spike event based on XY algorithm
else if (X_SW<=X_D<=X_SE && ((Y_S>Y_NW &&
Y_D<=Y_SW)||(Y_S<Y_SW && Y_D>=Y_NW))){          if
(Y_C>Y_D){
                if (abs(X_C-X_NW)>abs(X_C-X_NE)){
                      Route to NE according to XY algorithm;          }
                else {
                      Route to NW according to XY algorithm;
    }
          }
          else {
                Route to SW according to XY algorithm;      }
          Route to X_D according to Zhang algorithm;
          Continue route according to XY algorithm;
          }
```

Fig. 10. Pseudo code for MFNT algorithm along Y direction.

B. Deadlock-free fault-tolerant routing

According to a previous study [45], the necessary and sufficient condition for any routing algorithms to be deadlock-free is that there is no loop in its corresponding Component dependency graph (CDG).

In the case of networks without any fault nodes, since the turn from Y to X direction is prohibited, there is no cycle in the CDG of the mesh-based network. When the fault region is located inside the network, the proposed MFTN algorithm adds turns from Y to X direction on the northwest, southwest and southeast corners of the bypass loop, and removes the turn from X to Y direction on the northeast corner. Since this turn is removed, no cycle will exist and the proposed algorithm is deadlock-free.

When the fault region is located on the edge the mesh network, the proposed algorithm adds the turns from Y to X direction on the vertex of the bypass loop in order to make spike events avoid the fault region. Since the bypass loop is not a cyclic link, it will not result in a deadlock. Therefore, the proposed MFTN algorithm is completely deadlock-free. It is also independent of the area and location of the fault region, and the NoC scale.

VI. Experimental Results

In this section, we present experimental results implemented on the digital neuromorphic system LaCSNN [14], which is modified to include the proposed MFTN algorithm and implement the required context-dependent learning task. First, the experimental oscilloscope outputs are displayed in Fig. 11. This figure shows the firing activities of the neurons in the three different layers of the implemented neuromorphic network realizing the targeted context-dependent learning task. Fig. 11(a) shows the spiking activities of all the six neurons in the information sensory layer. The first four neurons spike alternatively due to the specific input combination, and the last two neurons fire alternatively due to the WTA mechanism. Fig. 11(b) shows the spiking activities of eight neurons randomly chosen from the hippocampal functional layer. This figure shows that only one neuron spikes at any given time due to the WTA mechanism. Fig. 11(c) shows the firing activities of each neuron in the motor layer, in which the two neurons spike alternatively. These real-time millisecond-scale spiking activities reveal that biological behaviors can be reproduced accurately in real time.

In addition to the behavioral analysis of the neurons firing patterns, their selectivity is also evaluated rigorously in 200 runs, each with different weight initialization and random noise. In order to characterize the selectivity of the neurons in the proposed neuromorphic network, a criteria of selectivity index (SI) is defined as follows:

$$SI = \left(n - \sum_{i=1}^{n} \lambda_i / \lambda_{pref} \right) \Big/ (n-1) \tag{5}$$

where n represents the stimulus events, λ_i represents the firing

rate in response to the ith stimulus event for a single neuron, and λ_{pref} represents the preferred stimulus event for the same neuron. The variable λ_{pref} is calculated based on the maximum firing rates of all the experienced stimulus events for each neuron. For the case of place selectivity, n=4 because the context-dependent task contains four physically different places, including A1, A2, B1, and B2. In order to implement the place selectivity, the mean firing rate for when the toy monkey encounters item X and Y is combined when the toy monkey is in each of these four places. In the case of item selectivity, n=2 because two different items exist in the proposed context-dependent learning task.

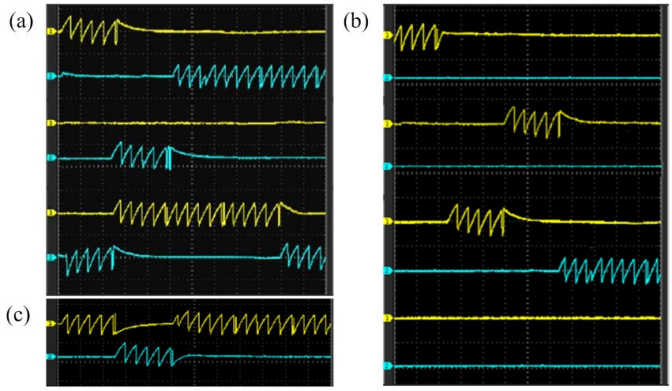


Fig. 11. Experimental oscilloscope output results. The time division is set to ms, while the amplitude division is mV. (a) Firing activities in the information sensory layer. (b) Firing activities of eight randomly chosen neurons in the hippocampal functional layer. (c) Firing patterns of the two neurons in the motor layer.

For the SI calculations, only neurons in the hippocampal functional layer are explored because this layer determines the selectivity capability. For all values of SI, the mean and standard deviation are calculated over four 50 trials as shown in Fig. 12. In this figure, four cases are investigated including (i) a neuromorphic network implementing the network shown in Fig. 2(a) but without the proposed fault-tolerant routing mechanisms and without faulty nodes (the first row of Fig. 12); (ii) the neuromorphic network without the proposed fault-tolerant routing mechanisms but with faulty nodes (the second row); (iii) the proposed FCL framework without faulty nodes (the third row); and (iv) the FCL framework with faulty nodes (the fourth row). As shown in Fig. 12, the mean value of SI for place selectivity is around 0.8 and remains constant in the four different cases. The mean value of SI for item selectivity begins from around 0.8 and increases to around 1.0, while the mean value of context selectivity begins around 0.7 increasing to around 1.0. These values suggest that the place selectivity is at a continuous rate, but item and context selectivity improve during the context-dependent learning task. Overall, Fig. 12 shows that, compared to a network without fault-tolerant routing, the proposed FCL network can solve the fault problems, successfully. Below, we present more experimental results, in all of which, the FCL framework with faulty nodes is used.

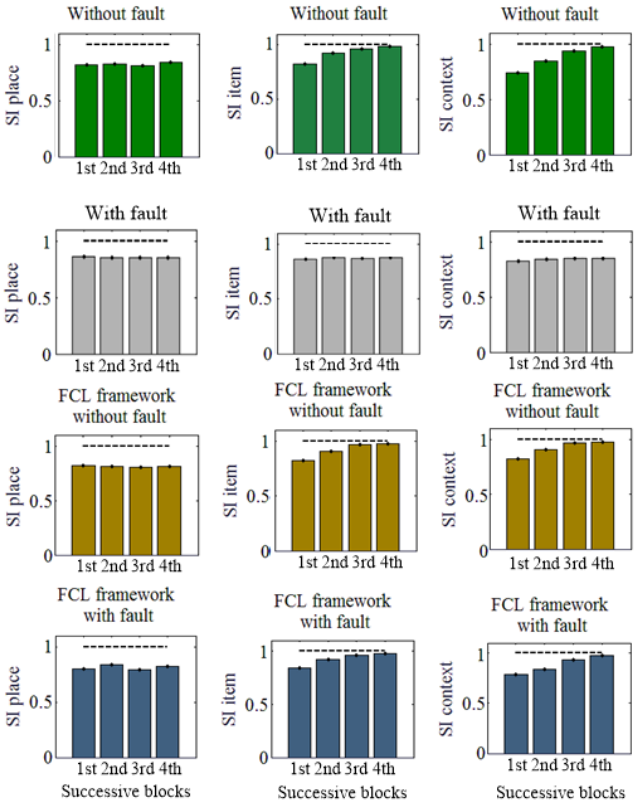


Fig. 12. Selectivity of the neuromorphic network shown in Fig. 2(a) without the proposed MFTN algorithm in two cases of without and with faults (first two rows) and the proposed FCL framework with and without faults for different parts of the context-dependent learning task, i.e. place, item, and context selection. Here, the x axis shows four successive blocks each with 50 trials.

A. Neuromorphic context-dependent learning capability

In order to evaluate the learning performance of the proposed network, a criteria representing the weight change significance is defined as follows:

$$B_{ij} = 4\left(W_{ij} - 0.5\right)^2 \tag{6}$$

where $i \in 1, 2, \ldots, n_{sensory}$, and $j \in 1, 2, \ldots, n_{hippo}$. The equation is defined so that the resulting value of B_{ij} is large when the weight change is large, and it is small when the weight change has been small. Note that, the weight variation is larger when higher spiking activities occur within the network.

In the performed experiments, the proposed neuromorphic network is trained in 100 epochs, each containing 130 trials. There are two phases in a trial. During the first phase, the model monkey explores the environment. After the first phase of a trial, the second phase replays the action sequence that is generated in the first phase. As shown in Fig. 13(a), the accuracy of the network is improved with increase in training epochs. In order to further enhance the learning capability, three improvement schemes can be investigated in the base network. The first scheme is to increase the number of layers in the proposed network. Full connections with excitatory synapses are then used between these layers, and lateral connections with

inhibitory synapses are used on each layer. By doing this, the deep learning capability of the proposed network can be investigated. The results of this change shown in Fig. 13(b), demonstrate that the learning accuracy drops and stays around 60%. This means that the learning capability of the proposed neuromorphic network cannot be enhanced by increasing the number of layers of the hippocampal functional network. The second scheme is to increase the number of neurons within the hippocampal functional layer to explore whether the increasing neuron number will improve the learning capability.

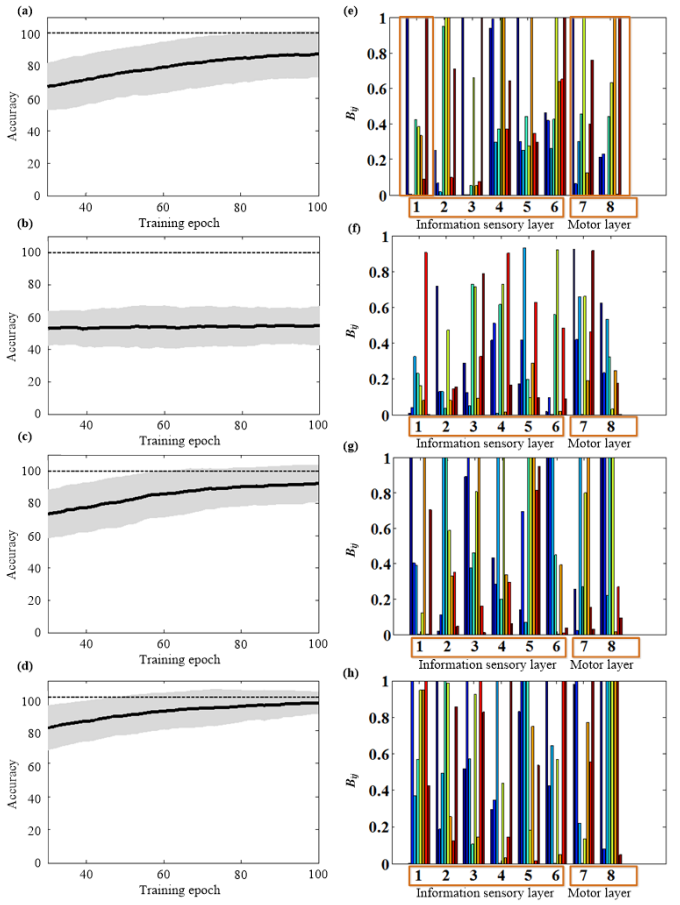


Fig. 13. Learning performance of the FCL framework with different network structures. (a) Learning accuracy of the original FCL framework trained in 100 epochs. (b) Learning accuracy by adding another layer of spiking neurons. (c) Learning accuracy with 16 neurons in hippocampal functional layer. (d) Learning accuracy with 64 neurons in hippocampal functional layer. (e) Corresponding weight changes with the original FCL framework. (f) Corresponding weight changes by adding another layer of spiking neurons. (g) Corresponding weight changes with 16 neurons in hippocampal functional layer. (h) Corresponding weight changes with 64 neurons in hippocampal functional layer.

Fig. 13(c) and Fig. 13(d) show the learning accuracy of the networks whose hippocampal functional layers contain 16 and 64 neurons, respectively. As shown in Fig. 13(d), the network with the hippocampal functional layer containing 64 neurons has the highest learning capability, which induces higher learning accuracy and higher learning speed compared to the other three schemes. As shown in Fig. 13(d) and Fig. 13(h), when the learning capability of the network is higher, the value of B_{ij} is larger. In these figures, the first six neurons are in the information sensory layer, and the last two are in the motor layer. We randomly select eight synapses connected to the chosen neurons, and evaluate the value of B_{ij} on those synapses. As shown in Fig. 13(h), when training does not actively happen, the proposed network does not experience significant weight changes. Thus the learning capability is not available in this network architecture.

B. Context-dependent learning analysis

In the implemented context-dependent learning task, there are eight input combinations according to the place, item, and context, which require eight hippocampal neurons to process the information. These eight combinations are shown on the x axis of Fig. 14(a). In this figure, each of the eight neurons (6 in the input sensory layer and two in the output motor layer) shown on the y axis, can only account for one input combination. The problem with this naive connectivity setting is that two or more neurons may learn the same combination, while others do not learn any combination. This results in the loss of useful information processed by the proposed network. In addition, with the combination number increasing, the network will be enlarged inducing larger hardware resource cost. Therefore, a different scheme with STDP learning rule is used. As shown in Fig. 14(b), each neuron can recognize two situations, and all the information can be learned. The neuron number of the hippocampal functional layer is increased to 16 and 40 respectively as shown in Fig. 14(c) and Fig. 14(d). The network can recognize more combinations of situations, thus more information combination can be learned. In addition, as shown in Fig. 14(b-d), the activated neuron numbers are increased along with the increasing neuron number in the hippocampal layer, which results in the increasing learning accuracy as shown in Fig. 13(c) and Fig. 13(d).

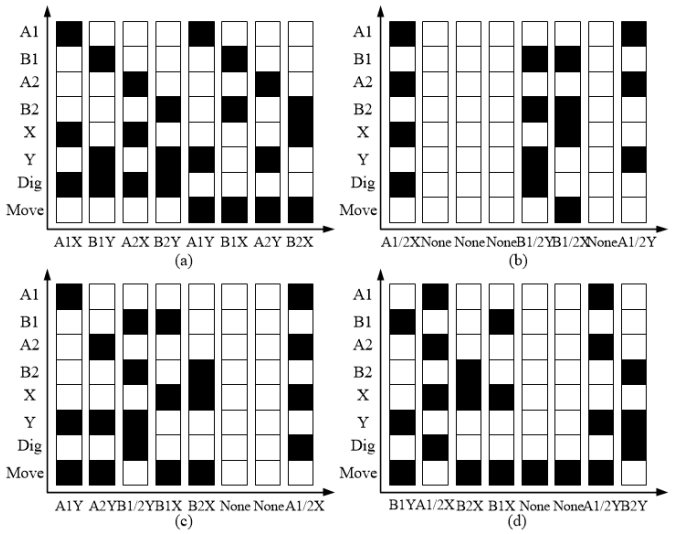


Fig. 14. Spiking situations with different inputs in the context-dependent learning task.

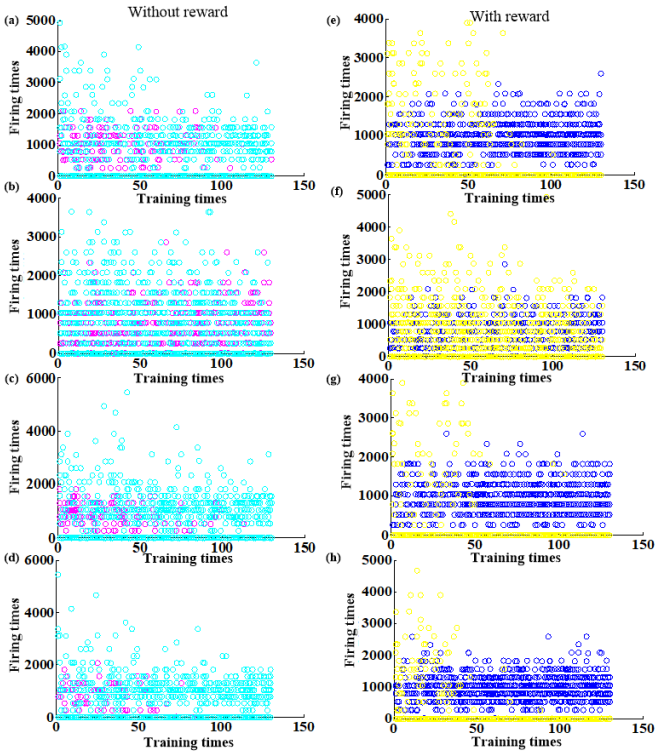


Fig. 15. Spiking activities of the proposed neuromorphic network with different layer number and neuron numbers in the hippocampal functional layer. (a) Raster plot with the original FCL framework without reward. (b) Raster plot with 2 layers in the hippocampal functional layer without reward. (c) Raster plot with 16 neurons in the hippocampal functional layer without reward. (d) Raster plot with 64 neurons in the hippocampal functional layer without reward. (e) Raster plot with the original FCL framework with reward. (f) Raster plot with 2 layers in the hippocampal functional layer with reward. (g) Raster plot with 16 neurons in the hippocampal functional layer with reward. (h) Raster plot with 64 neurons in the hippocampal functional layer with reward.

In order to analyze the difference in the learning accuracy of the various networks show in Fig. 14, the firing activities of these networks are assessed. In Fig. 15(a-d), the purple circles represent the number of spikes the digging neuron fires during training time without reward, and the light blue circles indicate the number of spiking activities with “moving” action without reward. The spike numbers of the “moving” and “digging” actions with reward are represented by yellow and dark light circles respectively in Fig. 15(e-h). As shown in Fig. 15(a), the number of false (unrewarded) “digging” actions are considerably large in the beginning of the learning (more purple circles). As the training continues, the unrewarded (false) actions begin to decrease. The same tendency occurs in Fig. 15(e), in which the number of the “moving” actions (yellow circles) decreases with the training time increasing. Fig. 15(b) and 15(f) show the same results but for a larger network with 2 layers in the hippocampal functional layer, respectively. These figures show that a larger network cannot improve the learning capability. The right and wrong actions occur alternatively without regularity in this case. Fig. 15(c-d) and Fig. 15(g-h) show the network with X neurons in its hippocampal functional layer, where X=16 for Fig. 15(c) and Fig. 15(g) and X=64 for Fig. 15(d) and Fig. 15(h). These figures suggest that the false actions can be reduced with the neuron number increasing. This enhances the context-dependent learning capability, which is consistent with the experimental results shown in Fig. 13. In addition, by comparing Fig. 15(a) and Fig. 15(b), it shows that the tendency towards firing of certain kinds of neurons in the network with reward will be stronger compared to the network without reward.

C. Performance analysis of the proposed fault-tolerant algorithm

The performance analysis of our proposed fault-tolerant algorithm for multiple fault nodes is presented in Fig. 16. Here, the northeast node on the bypass loop is defined as the reference node. As shown in Fig. 16(a), the performance of the proposed MFTN algorithm is better than Chen's algorithm [49] with different locations of faulty regions in the (8, 8) mesh network of neurons. As shown in Fig. 16, these algorithms follow a similar trend. The latency is the highest when the faulty region is in the network center, but it is the lowest when the fault region is located in the network vertex. This is because more nodes are affected when the fault region is in the network center, inducing the most influence on the data transmission latency. The figure shows that, no matter where the fault region is located, the latency performance with the proposed MFTN algorithm is better than Chen’s. When the communication latency is 80 μs, the improvements of the maximum event rate are 8.3%, 1.9% and 0.9% respectively when the reference node is located in network center (4, 4), network edge (5, 7) and network vertex (7, 2), respectively. Therefore, the best performance achieved using the proposed MFTN algorithm is when the fault region is located in the network center. This is because when the fault area is located in the network center, the number of source and destination nodes satisfying the optimization condition is the largest, therefore the improvement performance is the most obvious. When the fault nodes are located at the network vertex, the number of source and destination nodes satisfying the optimization condition is the lowest.

With the enlargement of fault area, compared with Chen's algorithm, the proposed algorithm provides more significant latency advantage. In Fig. 16(b), the latency of the proposed algorithm is compared with Chen’s algorithm when the fault area is 2×2 (the location of the reference node is (4,4)), 2×3 (the location of the reference node is (4,4)), and 2×4 (the location of the reference node is (6,5)). Fig. 16(c) shows the same fault area size and reference node location as in part (b), but with fault regions changing along vertical direction. The results show that no matter the fault area enlarges along the horizontal or vertical directions, the network latency of the proposed algorithm is better than Chen's algorithm.

When the network delay is 100 μs, compared with Chen's algorithm, the saturation injection rate increases by 5.2%, 11.7% and 16.1% respectively when the fault area enlarges horizontally (Fig. 16(b)); however, when the fault area enlarges vertically, the saturation injection rate increases by 5.2%, 5.5% and 6.1%, respectively (see Fig. 16(c)). This is because when the data encounters the fault region along the Y direction, the

proposed algorithm cannot reduce the distance from the source node to the destination node and the transmission length on the bypass. It also allocates part of the original data that needs to bypass along the left half of the fault region to the right half, to improve the load balance on the bypass loop. Furthermore, when the fault area enlarges along the horizontal direction, the number of nodes needed to bypass the fault area along the X direction decreases, and the number of nodes needed to bypass the fault area along the Y direction increases. Therefore, the network performance of the proposed algorithm is more prominent when the fault area increases along the horizontal direction.

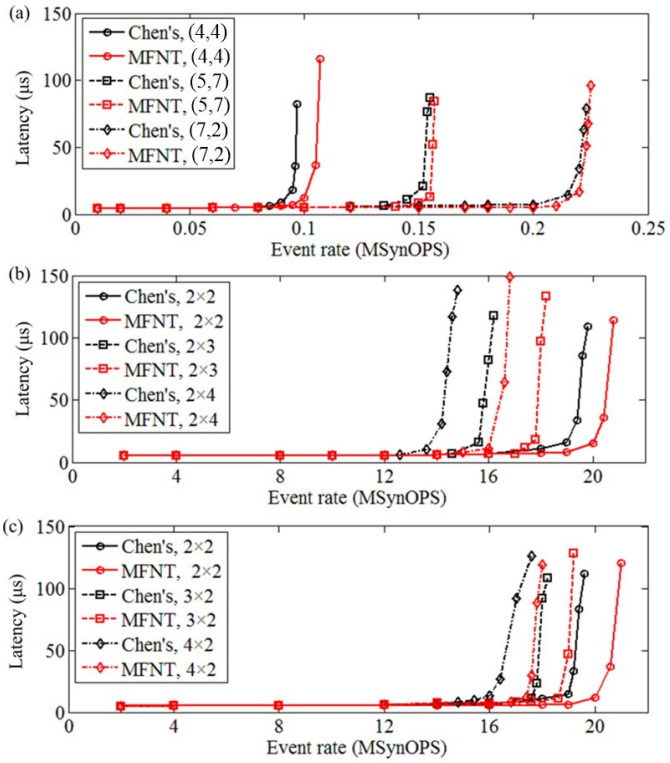


Fig. 16. Performance analysis of the fault tolerant algorithms for multiple fault nodes. (a) Communication latency with different locations of faulty regions in the (8, 8) Mesh network of neurons. (b) Communication latency with different areas of fault regions changing along horizontal direction. (c) Communication latency with different areas of fault regions changing along vertical direction.

VII. Discussions

In this study, a neuromorphic context-dependent learning framework is proposed with a novel multiple-fault-tolerant spike routing scheme. In order to implement the targeted learning framework in hardware, the digital neuromorphic system LaCSNN is used, which uses FPGAs to implement neuromorphic models. With the advantages of low energy consumption, high reconfigurability, parallel processing capability, and fast time to market [29]-[36], field programmable gate array (FPGA) implementations show promising potential for high performance neuromorphic systems [50], [51]. In summary, there are three critical points to be discussed to highlight the contributions of the proposed framework, and to present potential directions for future studies.

Firstly, a fault-tolerant neuromorphic architecture with a novel multicast routing scheme is presented, which is scalable and applicable to a variety of neuromorphic applications. The bypass method without virtual channels is used in the proposed fault-tolerant routing scheme implemented in the neuromorphic network in this study. In order to comprehensively compare our work with state-of-the-art, a multi-fault routing scenario is considered. As shown in Fig. 16, in comparison with a previous fault-tolerant routing scheme using bypass method [49], the improvements achieved in the event rate, when communication latency is 80μs, are 5.3%, 1.9% and 0.9% when the fault region is located in the network center, network edge, and network vertex, respectively. Furthermore, the event rate of the proposed fault-tolerant algorithm can be improved by 5.2%, 11.7% and 16.1% when the fault area is enlarged along horizontal direction, and improved by 5.2%, 5.5% and 6.1% with the fault area increasing in size along vertical direction.

Secondly, a digital neuromorphic context-dependent learning model inspired by hippocampus-mPFC pathway is proposed and implemented in this study. The neural mechanisms underlying the spiking activities of the hippocampus-mPFC network and context-dependent learning are fully investigated using the proposed neuromorphic FPGA framework. Fig. 12 shows the neuronal selectivity of the network, which reveals that place selectivity rate remains constant during context-dependent learning tasks, while item and context selectivity improve with learning. Fig. 13 shows that the learning performance can be enhanced by increasing the number of neurons in the hippocampal functional layer, while Fig. 15 displays the spiking activities of the neurons underlying the context-dependent learning and confirms that more neurons in the hippocampal layer can lead to better context-dependent learning.

Thirdly, our designed neuromorphic system integrates the proposed fault-tolerant routing capability with the hippocampus-mPFC pathway inspired context-dependent learning, in a unified framework. As shown in Table IV, although there are several digital neuromorphic systems with different types of neuron and synapse models and learning algorithms, none of them except Spinnaker [10], considers the important issue of faults and designing a fault-tolerant system [10], [14], [18], [21], [29]-[30], [36]-[44], [52]. Besides, to the best of our knowledge, there has been no previous implementation of a fault-tolerant framework for brain-inspired context-dependent learning. As shown in Fig. 12, Fig. 13 and Fig. 15, based on the proposed fault-tolerant framework, the context-dependent learning can perform flawlessly, while being affected by faulty nodes and regions of different sizes and locations.

TABLE IV

COMPARISON OF THIS STUDY WITH STATE-OF-THE-ART DIGITAL NEUROMORPHIC SYSTEMS.

Studies	Contribution	Learning	Fault-tolerant
Kim et al., 2012 [43]	Memristor synapse	Yes	No
Ambroise et al., 2013 [37]	Izhikevich network	No	No
Moore et al., 2012 [39]	Bluehive	Yes	No
Furber et al., 2014 [10]	SpiNNaker	Yes	Yes
Merolla et al., 2014 [18]	Truenorth	Yes	No
Yang et al., 2015 [29]	Basal ganglia	No	No
Qiao et al., 2015 [44]	ROLLS	Yes	No
Cheung et al., 2016 [40]	NeuroFlow	Yes	No
Luo et al., 2015 [30]	Cerebellar network	Yes	No
Kim et al., 2016 [38]	Neurocube	Yes	No
Pani et al., 2017 [41]	Izhikevich network	No	No
Yang et al., 2018 [14]	LaCSNN	Yes	No
Yang et al., 2018 [36]	Dopamine network	No	No
Wang et al., 2018 [42]	Cortex simulator	No	No
Yang et al., 2019 [21]	IBFT-based CMN	Yes	No
This study	FCL framework	Yes	Yes

The proposed high-performance fault-tolerant digital neuromorphic system helps conveniently prove any neuromorphic study concept. However, other implementation technologies such as analog neuromorphic chips and memristive designs can be also investigated to implement the proposed FCL framework. Besides, due to its fault-tolerant capability, the proposed neuromorphic framework presents a versatile platform for the study of the neuronal mechanisms of many biologically inspired spiking neural networks for cognitive behaviors such as motor learning and visual recognition. It can also be explored to be applied in robotic decision-making tasks, and other applications including unmanned aerial vehicles, and brain-machine interfaces.

VIII. CONCLUSIONS

This study presented a brain-inspired framework for context-dependent learning tasks implemented on the digital neuromorphic system LaCSNN. The proposed framework uses SNN models for information processing and STDP rule for learning. These are directly inspired by the mechanisms of biological hippocampus-mPFC networks. Experimental results show that the context-dependent learning can be conducted in real time. In addition, a fault-tolerant spike routing algorithm was proposed to make the proposed neuromorphic system for context-dependent learning prone to faulty hardware. We demonstrated that, various fault scenarios cannot impact the learning capability of the proposed system and the targeted context-dependent learning can be performed flawlessly. In addition, using the proposed novel routing method, the average latency and the maximum throughput of the system was shown to be significantly improved, compared to previous routing strategies. Furthermore, unlike many previous digital neuromorphic systems shown in Table IV, the proposed system in this work is one of the few neuromorphic systems that have fault-tolerant capabilities. The proposed fault-tolerant spike routing scheme will be applied to other brain-inspired computing tasks in neuromorphic systems.

NOMENCLATURE

Abbreviation	Meaning	Abbreviation	Meaning
SNN	Spiking neural network	FCL	Fault-tolerant context-dependent learning
FPGA	Field programmable gate array	mPFC	Medial prefrontal cortex
LaCSNN	Large-scale conductance-based spiking neural network	AER	Address event representation
NoC	Network on chip	LIF	Leaky integrate-and-fire
STDP	Spike-timing dependent plasticity	LTP	Long term potentiation
LTD	Long term depression	WTA	Winner take all
SLM	Shift logic multipliers	BIST	Built-in self-test
MFTN	Multiple-fault tolerant neuromorphic	CDG	Component dependency graph

REFERENCES

[1] D. Shohamy, A. D. Wagner, "Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events," *Neuron*, vol. 60, no. 2, pp. 378-389, Oct. 2008.

[2] C. J. MacDonald, K. Q. Lepage, U. T. Eden, *et al*., "Hippocampal "time cells" bridge the gap in memory for discontiguous events," *Neuron*, vol. 71, no. 4, pp. 737-749, Aug. 2011.

[3] B. E. Pfeiffer, D. J. Foster, "Hippocampal place-cell sequences depict future paths to remembered goals," *Nature*, vol. 497, no. 7447, pp. 74, May. 2013.

[4] R. W. Komorowski, J. R. Manns, H. Eichenbaum, "Robust conjunctive item–place coding by hippocampal neurons parallels learning what happens where, " *J. Neurosci.*, vol. 29, no. 31, pp. 9918-9929, 2009.

[5] O. Bichler, W. Zhao, F. Alibart, et al., "Pavlov's dog associative learning demonstrated on synaptic-like organic transistors," *Neural computation*, vol. 25, no. 2, pp. 549-566, 2013.

[6] E. R. Wood, P. A. Dudchenko, R. J. Robitsek, *et al*., "Hippocampal neurons encode information about different types of memory episodes occurring in the same location," *Neuron*, vol. 27, no. 3, pp. 623-633, 2000.

[7] J. C. Stanley, R. L. Elsom, P. C. Calder, *et al*., "UK Food Standards Agency Workshop Report: the effects of the dietary n-6: n-3 fatty acid ratio on cardiovascular health," *Br. J. Nutr.*, vol. 98, no. 6, pp. 1305-1310, Dec. 2007.

[8] J. Ferbinteanu, and M. L. Shapiro, "Prospective and retrospective memory coding in the hippocampus," *Neuron*, vol. 40, no. 6, pp. 1227-1239, Dec. 2003.

[9] I. Lee, and F. Solivan, "The roles of the medial prefrontal cortex and hippocampus in a spatial paired-association task," *Learn. Mem*., vol. 15, no. 5, pp. 357-367, Jan. 2008.

[10] S. B. Furber, F. Galluppi, S. Temple, *et al.*, "The spinnaker project," *Proc. IEEE*, vol. 102, no. 5, pp.652 -665, May. 2014.

[11] B. V. Benjamin, P. Gao, E. McQuinn, *et al*., "Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations." *Proc. IEEE*, vol. 102, no. 5, pp. 699-716, May. 2014.

[12] M. A. Petrovici, B. Vogginger, P. Müller, *et al*., "Characterization and compensation of network-level anomalies in mixed-signal neuromorphic modeling platforms." *PLoS One*, vol. 9, no. 10, pp. E108590, Oct. 2014.

[13] J. Park, T. Yu, S. Joshi, *et al*., "Hierarchical address event routing for reconfigurable large-scale neuromorphic systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2408-2422, Jul. 2016.

[14] S. Yang, B. Deng, J. Wang, *et al*., "Real-time neuromorphic system for large-scale conductance-based spiking neural networks," *IEEE Trans. Cybern*., vol. 49, no. 7, pp. 2490-2503, Apr. 2018.

[15] A. Mortara, E. A. Vittoz, P. A. Venier, "A communication scheme for analog VLSI perceptive systems," *IEEE J. Solid-State Circuits*, vol. 30, no. 6, pp. 660-669, Jul. 1995.

[16] J. Lazzaro, J. Wawrzynek, "A multi-sender asynchronous extension to the AER protocol," *Proc. 16th Conf. Adv. Res. VLSI. IEEE*, Apr. 1995.

[17] J. Lazzaro, J. Wawrzynek, "Scalable hierarchical network-on-chip architecture for spiking neural network hardware implementations," *IEEE Trans. Parallel Distrib. Syst*., vol. 24, no. 12, pp. 2451-2461, Dec. 2013.

[18] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science*, vol. 345, no. 6197, pp. 668-673, Aug. 2014.

[19] J. Pei, L. Deng, S. Song, *et al*., "Towards artificial general intelligence with hybrid Tianjic chip architecture," *Nature*, vol. 572, no. 7767, pp. 106-111, Jul. 2019.

[20] T. H. Vu, A. B. Abdallah, "Low-Latency K-Means Based Multicast Routing Algorithm and Architecture for Three Dimensional Spiking Neuromorphic Chips,"*2019 IEEE Int. Conf. BigComp IEEE*, Apr. 2019.

[21] S. Yang, B. Deng, J. Wang *et al*., "Scalable Digital Neuromorphic Architecture for Large-Scale Biophysically Meaningful Neural Network With Multi-Compartment Neurons," *IEEE Trans. Neural Netw. Learn Syst*., vol. 31, no. 1, pp. 148-162, Mar. 2019.

[22] J. Wu, and F. Steve, "A multicast routing scheme for a universal spiking neural network architecture," *Comput. J.*, vol. 53, no. 3, pp. 280-288, Apr. 2009.

[23] T. H. Vu, O. M. Ikechukwu, A. B. Abdallah, "Fault-tolerant spike routing algorithm and architecture for three dimensional noc-based neuromorphic systems," *IEEE Access*, vol. 7, pp. 90436-90452, Jun. 2019.

[24] G. Bi, M. Poo, "Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type," *J. Neurosci*., vol. 18, no. 24, pp. 10464-10472, Dec. 1998.

[25] T. H. Vu, O. M. Ikechukwu, A. B. Abdallah, "Fault-tolerant spike routing algorithm and architecture for three dimensional noc-based neuromorphic systems," *IEEE Access*, vol. 7, pp. 90436-90452, Jun. 2019.

[26] M. C. Hagen, O. Franzén, F. McGlone, et al., "Tactile motion activates the human middle temporal/V5 (MT/V5) complex," *European J. Neurosci.*, vol. 16, no. 5, pp. 957-964, 2002.

[27] C. Zamarreño-Ramos, A. Linares-Barranco, T. Serrano-Gotarredona, *et al*, "Multicasting mesh AER: a scalable assembly approach for reconfigurable neuromorphic structured AER systems. Application to ConvNets," *IEEE Trans. Biomed. Circuits Syst*., vol. 7, no. 1, pp. 82-102, Jun. 2012.

[28] A. B. Abdallah, "Advanced multicore systems-on-chip: architecture, on-chip network, design," *Springer*, 2018.

[29] S. Yang, J. Wang, S. Li, *et al*., "Cost-efficient FPGA implementation of basal ganglia and their Parkinsonian analysis," *Neural Netw.*, vol. 71, pp. 62-75, Nov. 2015.

[30] J. Luo, G. Coapes, T. Mak, *et al*., "Real-time simulation of passage-of-time encoding in cerebellum using a scalable FPGA-based system," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 3, pp. 742-753, Oct. 2015.

[31] S. Yang, X. Wei, B. Deng, *et al*., "Efficient digital implementation of a conductance-based globus pallidus neuron and the dynamics analysis," *Physica A*, vol. 494, no. 484-502, Mar. 2018.

[32] K. Akbarzadeh-Sherbaf, B. Abdoli, S. Safari, *et al*., "A scalable FPGA architecture for randomly connected networks of Hodgkin-Huxley neurons," *Front. Neurosci.*, vol. 12, pp. 698, Oct. 2018.

[33] S. Yang, J. Wang, S. Li, *et al*., "Digital implementations of thalamocortical neuron models and its application in thalamocortical control using FPGA for Parkinson′ s disease, " *Neurocomputing*, vol. 177, pp. 274-289, Feb. 2016.

[34] B. Sen-Bhattacharya, T. Serrano-Gotarredona, L. Balassa, *et al*., "A spiking neural network model of the lateral geniculate nucleus on the SpiNNaker machine," *Front. Neurosci*., vol. 11, pp. 454, Aug. 2017.

[35] S. Yang, X. Hao, B. Deng, *et al*., "A survey of brain-inspired artificial intelligence and its engineering," *Life Res*., vol. 1, no. 1, pp. 23-29, Jul. 2018.

[36] S. Yang, J. Wang, Q. Lin *et al*., "Cost-efficient FPGA implementation of a biologically plausible dopamine neural network and its application," *Neurocomputing*, vol. 314, pp. 394-408, Jul. 2018.

[37] M. Ambroise, T. Levi, Levi, Y. Bornat and S. Saighi, "Biorealistic spiking neural network on FPGA," *in Proc. 2013 47th Annu. Conf. Inf. Sci. Syst*., pp. 1-6. Oct. 2013.

[38] D. Kim, J. Kung, S. Chai, *et al*., "Neurocube: A programmable digital neuromorphic architecture with high-density 3D memory," *2016 ACM. IEEE 43rd Annu. Int. Sym. Comput. Archit.*, pp. 380-392, 2016.

[39] S. W. Moore, P. J. Fox, S. J. Marsh, *et al*., "Bluehive-a field-programable custom computing machine for extreme-scale real-time neural network simulation," *2012 IEEE 20th Int. Symp. Field-Program. Cust. Comp. Mach.*, May. 2012.

[40] K. Cheung, S. R. Schultz, W. Luk, "NeuroFlow: a general purpose spiking neural network simulation platform using customizable processors," *Front. Neurosci*., vol. 9, pp. 516, Jau. 2016.

[41] D. Pani, P. Meloni, G. Tuveri, F. Palumbo, P. Massobrio and L. Raffo, "An FPGA platform for real-time simulation of spiking neuronal networks," *Front. Neurosci*., vol. 11, no. 90, Feb. 2017.

[42] R. M. Wang, C. S. Thakur, A. Schaik, "An FPGA-based massively parallel neuromorphic cortex simulator," *Front. Neurosci*., vol. 12, pp. 213, Apr. 2018.

[43] Y. Kim, Y. Zhang, P. Li, "A digital neuromorphic VLSI architecture with memristor crossbar synaptic array for machine learning," *2012 IEEE Int. SOC Conf.*, pp 328-333, 2012.

[44] N. Qiao, H. Mostafa, F. Corradi, *et al*., "A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128K synapses," *Front. Neurosci.*, vol. 9, pp. 141, Apr. 2015.

[45] W. J. Dally, C. L. Seitz, "Deadlock-free message routing in multiprocessor interconnection networks," *IEEE Trans. Comput.*, vol. 36, no. 5, pp. 547-553, 1988.

[46] D. Mikael, E. Örjan, and L. Anders, "Large-scale modeling—A tool for conquering the complexity of the brain," *Frontiers Neuroinform.*, vol. 2, p. 1, Apr. 2008.

[47] Z. Zhang, A Greiner and S Taktak, "A Reconfigurable Routing Algorithm for a Fault-Tolerant 2D-Mesh Network-on-chip," *Design Automation Conference*, 2008: 441-446

[48] A. Neckar, S. Fok, B. V. Benjamin, *et al.*, "Braindrop: A mixed-signal neuromorphic architecture with a dynamical systems-based programming model," *Proceedings of the IEEE*, vol. 107, no. 1, pp. 144-164, 2018.

[49] K. Chen and G. Chiu, "Fault-tolerant routing algorithm for meshes without using virtual channels," *J. Inform. Sci. Eng.*, vol. 14, no. 4, pp. 765-783, 1998.

[50] M. Heidarpur, A. Ahmadi, M. Ahmadi, *et al.*, "CORDIC-SNN: On-FPGA STDP Learning With Izhikevich Neurons," *IEEE Trans Circuit Syst I: Regular Papers*, vol. 66, no. 7, pp. 2651-2661, 2019.

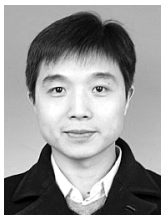
[51] C. Lammie, T. J. Hamilton, A. van Schaik, M. R. Azghadi, "Efficient FPGA implementations of pair and triplet-based STDP for neuromorphic architectures," *IEEE Trans Circuit Syst I: Regular Papers*, vol. 66, no. 4, pp. 1558-1570, 2018.

[52] H. Asgari, B. M. N. Maybodi, M. Payvand, *et al.*, "Low-Energy and Fast Spiking Neural Network For Context-Dependent Learning on FPGA," *IEEE Trans on Circuits Syst II: Express Briefs*, 2020.

[53] R. A. Gulli, L. R. Duong, B. W. Corrigan, *et al.*, "Context-dependent representations of objects and space in the primate hippocampus during virtual navigation," *Nat. Neurosci.*, vol. 23, no. 1, pp. 103-112, 2020.
[54] X. Zhao, Y. Wang, N. Spruston, *et al.*, "Membrane potential dynamics underlying context-dependent sensory responses in the hippocampus," *Nat. Neurosci.*, vol. 23, pp. 881-891, 2020.
[55] I. Lee I, D. Yoganarasimha, G. Rao, *et al.*, "Comparison of population coherence of place cells in hippocampal subfields CA1 and CA3," *Nature*, vol. 430, no. 6998, pp. 456-459, 2004.
[56] J. Waider, S. Popp, B. Mlinar, *et al.*, "Serotonin deficiency increases context-dependent fear learning through modulation of hippocampal activity," *Front. Neurosci.*, vol. 13, no. 245, 2019.
[57] J. F. Guzowski, J. J. Knierim, E. I. Moser, "Ensemble dynamics of hippocampal regions CA3 and CA1," *Neuron*, vol. 44, no. 4, pp. 581-584, 2004.
[58] R. S. Ross, K. R. Sherrill, C. E. Stern, "The hippocampus is functionally connected to the striatum and orbitofrontal cortex during context dependent decision making," *Brain Res.*, vol. 1423, pp. 53-66, 2011.

Shuangming Yang received his M.S. degree and Ph.D. degree from Tianjin University, Tianjin, China in 2016 and 2020 respectively. He is currently an assistant professor in the School of Electrical and Information Engineering, Tianjin University. His research interests include neuromorphic engineering, neural system modeling, neural engineering, brain-inspired computing, and machine learning. He is currently a Review Editor for *Frontiers in Neuroscience*.

Jiang Wang was born in China, 1964. He received the Master degree in Power and automation engineering from University of Tianjin, Tianjin, China in 1989. He received the PhD degree in University of Tianjin in1996. He is a professor in School of Electrical and Information Engineering, Tianjin University. His research interests are nonlinear dynamical systems, neuroscience, and information processing and detecting.



Bin Deng received the Ph. D degree in electrical engineering from Tianjin University, China, in 2007. He is a Professor in the School of Electrical and Information Engineering, Tianjin University. His research interests include dynamic analysis of neuron model, nonlinear analysis of neuron electrical information.

Mostafa Rahimi Azghadi (S'07–M'14, SM'19) received the Ph.D. degree in electrical and electronic engineering from The University of Adelaide, Australia. From 2012 to 2014, he was a Visiting Ph.D. Student with the Neuromorphic Cognitive System Group, Institute of Neuroinformatics, University and Swiss Federal Institute of Technology (ETH), Zurich, Switzerland. He was a recipient of the Doctoral Research Medal, and the Adelaide University Alumni Medal in 2014. He is currently a Senior Lecturer with the College of Science and Engineering, James Cook University, Townsville, Australia, where he researches neuromorphic engineering and brain-inspired architectures and develops custom hardware and software solutions for a variety of engineering applications ranging from medical imaging to precision agriculture. He was a recipient of several national and international awards and scholarships, such as the 2020 JCU Award for Excellence in Innovation and Change, a Queensland Young Tall Poppy Science Award in 2017, and the South Australia Science Excellence Awards in 2015.

Bernabe Linares-Barranco (M'90, S'06, F'10) received the B.S. degree in electronic physics, the M.S. degree in microelectronics, and a first Ph.D. degree in 1990 from the University of Sevilla, Sevilla, Spain, in 1986, 1987, and 1990, respectively, and a second Ph.D. degree from Texas A\&M University, College Station, TX, USA, in 1991. Since 1991, he has been with the Sevilla Microelectronics Institute (IMSE-CNM), from the Spanish Research Council (CSIC) of Spain, where he currently is Full Professor of Research and serves as Director of the Institute since February 2018. He has been a Visiting Professor/Fellow at The Johns Hopkins University, Baltimore, MD (USA), Texas A&M University, College-Station (USA), The University of Manchester (UK), and the University of Lincoln (UK). His recent interests are in Address-Event-Representation VLSI, real-time AER vision sensing and processing chips, memristor circuits, and extending AER to the nanoscale. He has received two IEEE Transactions Best Paper Awards, and has been an Associate Editor of the IEEE transactions on circuits and systems-II, the IEEE transactions on neural networks, and Frontiers in neuromorphic engineering. From 2011 to 2013, he was the Chair of the IEEE Circuits and Systems Society Spain Chapter, and became an IEEE Fellow in 2010.