

Integrating Visual Perception with Decision Making in Neuromorphic Fault-tolerant Quadruplet-spike Learning Framework

Shuangming Yang, *Member, IEEE*, Haowen Wang, Yanwei Pang, *Senior Member, IEEE*, Yaochu Jin, *Fellow, IEEE*, Bernabe Linares-Barranco, *Fellow, IEEE*

Abstract—The brain possesses the remarkable ability to seamlessly integrate perception with decision-making within a dynamically changing environment in a fault-tolerant, end-to-end manner. This extraordinary capability offers a compelling solution for brain-inspired intelligence, replete with the advantages of end-to-end decision-making: robustness, high accuracy, real-time responsiveness, autonomous intelligence, and a high degree of biological plausibility. Neuromorphic computing stands as a promising avenue for brain-inspired intelligence through the harmonious co-design of algorithms and hardware, aimed at unlocking full potential. This paper introduces a comprehensive neuromorphic computing framework for end-to-end intelligence. It introduces the quadruplet spike-timing-dependent plasticity, which serves as a cornerstone for perceptual to decision-making tasks. A fault-tolerant neuromorphic routing strategy is presented to fortify the framework's robustness. Empirical results underscore its impressive attributes with high accuracy, robustness, fault tolerance, and minimal computational latency when orchestrating end-to-end decision-making alongside visual perception. This study marks a pioneering effort in unified, fault-tolerant neuromorphic framework engineered for brain-inspired end-to-end intelligent tasks, merging visual perception with adaptive decision-making. Such an endeavor is profoundly meaningful, as it propels the development of artificial general intelligence, holding vast implications for the field's advancement.

Index Terms—Neuromorphic computing, spiking neural network (SNN), fault-tolerant, STDP, decision making.

I. INTRODUCTION

The superior computational power of the biological neural system of the human brain inspires researchers to develop

intelligent computing algorithms and systems that can process information in a more accurate, reliable, and flexible manner [1], [2]. The brain has the capability of integrating visual perception with decision-making efficiently in a dynamically changing environment [3], [4]. An efficient realization of the end-to-end intelligent behaviors from visual input to action selection is meaningful for comprehending brain mechanisms and developing artificial intelligence [5], which is also essential and necessary for the development of high-performance brain-inspired intelligence.

The edge computing of intelligent agents has been successfully applied in various fields, such as autonomous driving, unmanned vehicles, and healthcare robots [36]. When executing real-world tasks, the complex environment requires the intelligent agent to robustly adapt to and be fault-tolerant with the complex environments with different kinds of disturbance, such as electronic interference and night darkness. Precise decision-makings should be completed when equipped with a low-cost device, and should be realized with quick response after accurate visual perception. Therefore, there exists an urgent requirement to develop a novel powerful approach that can adapt to complex environment with strong robustness, strong fault tolerance, high accuracy, high response speed, and lightweight efficient computation.

Neuromorphic computing has the potential to address the aforementioned challenges. It has been developing biologically inspired, event-driven, and time-based algorithms in an efficient manner [20], [21], [43]. A number of neuromorphic systems have been presented based on SNNs, including HiAER, SpiNNaker, Neurogrid, and TrueNorth [22]-[25]. Neuromorphic systems have been developed based on dedicated logic and memory to digitally simulate neurons and synapses with higher computational precision [26]-[28]. In terms of the fault-tolerant techniques of the neuromorphic systems, Fukushima et al. proposed a fault-tolerant routing algorithm in two-dimensional network-on-chip (NoC) topology without using virtual channels (VC), and a new routing algorithm for non-VC routers in NoC network is proposed [55]-[56]. Putra et al. presented a new architecture to mitigate the adverse effects of faults for on-chip and off-chip memory with flexibility and energy efficiency [57]. Ikechukwu et al. presented a neuromorphic system based on 3D-NoC fault tolerant architecture, taking the spiking neurons with STDP as the processing elements [52]-[54]. Baek et al. presented FlexLearn, an on-chip learning engine that offers flexibility and enables rapid and efficient brain simulations. It achieves its high flexibility by supporting a range of biologically plausible

This work was supported partly by the National Key Research and Development Program of China (Grant No. 2022ZD0160405), and partly by the National Natural Science Foundation of China (Grant No.62376185, and 62176179) (*Corresponding author: Shuangming Yang*)

Shuangming Yang and Yanwei Pang are with School of Electrical and Information Engineering, Tianjin University, Tianjin, 300072 China, and are also with Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China. Haowen Wang is with School of Electrical and Information Engineering, Tianjin University, Tianjin, 300072 China.

Yaochu Jin is with the Chair of Nature Inspired Computing and Engineering, Faculty of Technology, Bielefeld University, D-33619 Bielefeld, Germany. B.Linares-Barranco is with the Microelectronics Institute of Seville, Seville 41092, Spain.

sub-rules that can be combined to simulate diverse target learning rules. To develop FlexLearn, they begin by identifying 17 representative sub-rules that modify synaptic weights using different approaches. Subsequently, FlexLearn design and streamline specialized datapaths dedicated to these sub-rules, carefully considering their interdependencies to maximize parallelism [58]. Investigations into the combination of FlexLearn with the learning algorithms under consideration of visual perception and decision making could provide valuable insights into enhancing the flexibility and efficiency of on-chip learning in neuromorphic computing. These works significantly facilitate the development and applications of the spike-based computing and neuromorphic engineering.

Spiking neural network (SNN) is the key model of neuromorphic computing, which is inspired by spike-based processing mechanism in real biological neural systems [6]-[10], [45]. The neuromorphic computing approach can support the end-to-end intelligent performance by coupling perception, learning, and decision making in real time without specifying a large number of hyper-parameters. Its good openness enables it to expand more brain-inspired intelligent behaviors on scalable parallel computing architectures.

Algorithm implementations on neuromorphic chips, such as Intel's Loihi chips, have garnered significant attention in recent years [46]-[48]. These specialized hardware platforms aim to replicate the functionality and efficiency of the human brain, facilitating the acceleration of neuromorphic computation and the execution of neural network models in a highly parallel and energy-efficient manner. For instance, Viale et al. employ a SNN coupled with an event-based camera to address a critical challenge in autonomous driving: the classification of cars versus other objects on the Loihi chip [46]. Massa et al. extensively investigate the process of converting a deep neural network into an SNN, with a specific focus on the Intel Loihi platform [47]. Additionally, to leverage the advantages of event-based sensors, they develop a pre-processing technique that enables seamless integration of these sensors within the deep neural network framework. They evaluate their method using the DvsGesture dataset, showcasing its efficacy in incorporating event-based sensor data into the domain of deep neural networks. Angelidis et al. construct a central pattern generator model by leveraging naturally emerging dynamical systems that arise from recurrent neural populations within the neural engineering framework [48]. They evaluate their system on two neuromorphic platforms, namely SpiNNaker and Loihi, indicating that spiking algorithms of this kind hold great potential for harnessing the theoretical advantages offered by neuromorphic hardware, including enhanced energy efficiency and computational speed. Although not implemented on neuromorphic hardware, Eliasmith et al. present a large-scale SNN model capable of performing eight distinct perceptual, motor, and cognitive tasks, including visual perception and decision making, based on the neural engineering framework [49]. Deng has developed a comprehensive model description framework and processing architecture known as Tianjic, offering a unified approach encompassing the entire software-to-hardware stack [50]. This integration enables seamless coordination between different components,

exemplified by hybrid-paradigm examples, i.e., a multimodal unmanned bicycle with both the decision making and visual perception capabilities. These examples demonstrate the immense potential of the unified framework in achieving versatile and efficient solutions. Unlike the previous research, in this study, we provide insights into the effective integration of SNN models and fault-tolerant neuromorphic architectures, aiming to realize a fault-tolerant neuromorphic computing framework with an end-to-end spike-based perception-to-decision-making capabilities. Our research showcases the promising prospects of leveraging advancements in neuromorphic architecture to achieve superior and fault-tolerant performance when executing brain-inspired end-to-end intelligent tasks by integrating visual perception with adaptive decision making.

The contributions of this study include three aspects. First, a unified neuromorphic framework is presented to realize the end-to-end intelligence from visual perception with decision making. The use of spiking representation makes the presented framework more biologically plausible and computationally efficient. Second, a quadruplet spike-timing-dependent plasticity (STDP) is first presented as a selection reference for spike-based perception-to-decision-making task. Third, a novel neuromorphic fault-tolerant algorithm is presented, which is suitable for neuromorphic implementation and enhances the communication performance compared to other solutions. The layout of this paper is summarized as follows. The network model is presented in Section II, and Section III describes the fault-tolerant routing strategy in neuromorphic framework. Digital neuromorphic architecture is described in Section IV, and Section VI discusses the proposed study, and concludes this paper finally.

II. NETWORK MODEL

A. Scenario setup

Imagine there are two closed houses, whose doors are marked with sign "1" and "2" respectively. In house A and house B, there are two rooms with sign "3" and "4" marked on each door. In both rooms, there are two bottles marked "5" and "6" respectively. Only one bottle contains a reward, and the other is empty. A monkey does not know any information or knowledge of the marks at the initial state. It should go through an experience to classify the different marks on doors or bottles by a series of epochs. In addition to the accurate classification of each sign, it should use the signs marked on each door to learn and make the decision that which house, room and bottle it should get in and get after a number of trials. After a period of learning trials, it finally get the knowledge to comprehend each sign and the experience to make the optimal decision.

B. Hierarchical network architecture

In this study, a hierarchical network architecture is presented as shown in Fig. 1, which includes a visual perception layer (VPL) and a decision-making layer (DML). The VPL component is composed of four layers, including the input layer, excitatory layer, inhibitory layer, and output layer. The input layer is an encoding layer, in which one neuron is for one image pixel. The input is encoded as a Poisson spike train, which is the

input to the excitatory layer based on the intensity of the input image pixels. The excitatory layer is connected to the input layer in a one-neuron-to-one-neuron manner. Inhibitory neurons connect back to the excitatory layer except for the one receiving excitatory connection. This winner-take-all mechanism realizes the lateral inhibition between the excitatory and inhibitory layers. The DML component contains three layers. The first layer contains six neurons, which represent different modality of information A, B, C, D, X, Y. The first layer is connected to eight excitatory neurons using all-to-all form in the second layer, and inhibitory connection in this layer inhibits each other in spite of themselves. The second layer is coupled with the third layer based on all-to-all connection manner. The third layer includes two neurons, which inhibit each other but also not inhibiting themselves. The two neurons in the DML layer represent two types of actions, such as turning left and turning right, exploitation and exploration, and also digging and moving respectively. In the presented model, the selection can only exist in two places and no intermediary places exist.

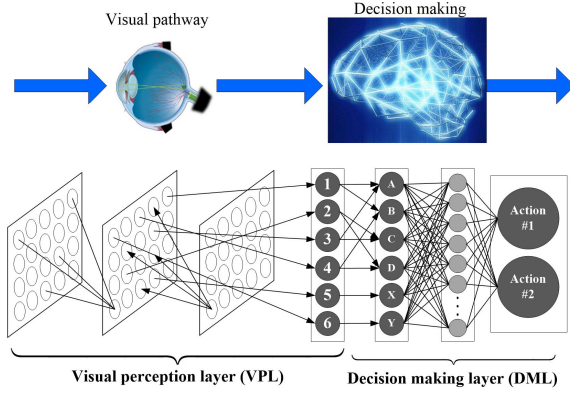


Fig. 1. Hierarchical network architecture with spiking neurons.

C. Neuron model in VPL

Various models are proposed to reproduce and mimic the information propagation and neural dynamics observed in electrophysiological experiments in the last few decades. Considering biological similarity, the leaky integrate-and-fire (LIF) model with the conductance of the synapse is used. The variable of membrane voltage V is described by

$$\tau \frac{dV}{dt} = (E_{rest} - V) + g_e(E_{exc} - V) + g_i(E_{inh} - V) \quad (1)$$

where E_{rest} represents the resting membrane potential. E_{exc} and E_{inh} represent the membrane potentials of excitatory and inhibitory synapses, and g_e and g_i represent the conductance of excitatory and inhibitory synapses. A time constant τ is used to simulate biological neurons, and this constant of the excitatory neuron is longer than inhibitory neuron. When V exceeds the threshold V_{th} , the neuron fires and V will be reset at V_{reset} . The information processing mechanism of a simple LIF neuron model is shown in Fig. 2, which shows the effects of synaptic weights and time on the membrane potentials V_1 , V_2 and the process of lateral reset connections and refractory period.

Synapse model is based on conductance changes, and their conductance increases when a presynaptic spike is input into

the synapse. In contrast, the excitatory and inhibitory conductance decays with time exponentially:

$$\begin{cases} \tau_{ge} \frac{dg_e}{dt} = -g_e \\ \tau_{gi} \frac{dg_i}{dt} = -g_i \end{cases} \quad (2)$$

where τ_{ge} and τ_{gi} represent the excitatory synaptic transmission time constant and the inhibitory synaptic conduction time constant, respectively. Homeostasis mechanism in LIF model and the neuron model in DML are described in detail in Supplementary material.

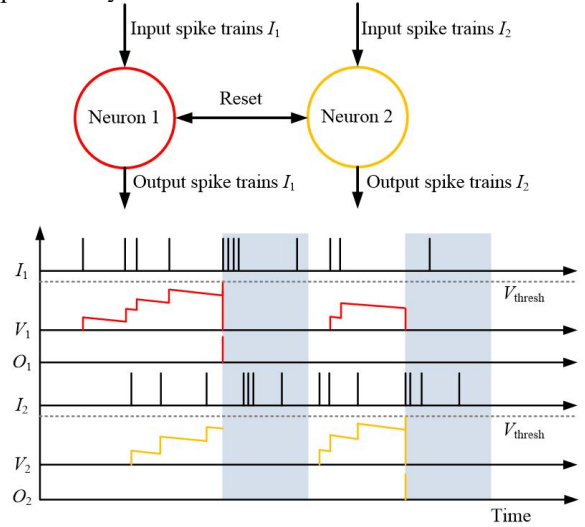


Fig. 2. Evolution process of a LIF neuron.

D. Q-STDP learning rule

In biological neural system, synapses are specialized component permitting the transfer of signals between neurons. Learning is realized by the synaptic weight updating using the neural activity. In D-STDP algorithm, potentiation exists if a postsynaptic spike occurs before a presynaptic spike. The temporal learning window governs the weight changes, in which the formulation of D-STDP can be described as follows

$$\Delta w = \begin{cases} \Delta w^+ = A^+ e^{\left(\frac{-\Delta t}{\tau^+}\right)}, & \text{if } \Delta t \geq 0 \\ \Delta w^- = A^- e^{\left(\frac{-\Delta t}{\tau^-}\right)}, & \text{if } \Delta t < 0 \end{cases} \quad (7)$$

where $\Delta t = t_{\text{post}} - t_{\text{pre}}$ denotes the timing distance between a post and a pre-synaptic spike. The constant τ^+ and τ^- denote the time constant of the learning window, and A^+ and A^- represent the maximal weight update for potentiation and depression. In addition, D-STDP including spike triplets induces triplet-based STDP (T-STDP). It is described as follows

$$\begin{cases} \Delta w^+ = e^{\left(\frac{-\Delta t_1}{\tau^+}\right)} \left(A_2^+ + A_3^+ e^{\left(\frac{-\Delta t_2}{\tau_y}\right)} \right), & \text{if } t = t_{\text{post}} \\ \Delta w^- = -e^{\left(\frac{-\Delta t_1}{\tau^-}\right)} \left(A_2^- + A_3^- e^{\left(\frac{-\Delta t_3}{\tau_x}\right)} \right), & \text{if } t = t_{\text{pre}} \end{cases} \quad (8)$$

where A_2^+ and A_2^- represent the weight updating amplitude with a pre-post or a post-pre pair respectively. Besides, A_3^+ and A_3^- represent triplet amplitude for potentiation and depression. However, the T-STDP rule contains exponential functions, and cannot be realized on digital neuromorphic system directly.

In order to consider the experimental outcomes based on quadruplets of spikes into account, a Q-STDP learning rule is proposed in this study. Previous studies have proposed Q-STDP algorithm with exponential functions, which cannot be implemented by logic elements on digital hardware directly. Therefore, this study presents the Q-STDP learning rule with differential equations without nonlinear functions. Six additional variables, r_1 , o_1 , r_2 , o_2 , r_3 , and o_3 , are used in the Q-STDP learning rule. They determine exponentially decaying potentiation and depression potentials, which are formulated as

$$\begin{cases} \frac{dr_1}{dt} = -\frac{r_1}{\tau_{x1}}, \text{ if } t = t_{pre}, r_1 = 1 \\ \frac{do_1}{dt} = -\frac{o_1}{\tau_{y1}}, \text{ if } t = t_{post}, o_1 = 1 \\ \frac{dr_2}{dt} = -\frac{r_2}{\tau_{x2}}, \text{ if } t = t_{pre}, r_2 = 1 \\ \frac{do_2}{dt} = -\frac{o_2}{\tau_{y2}}, \text{ if } t = t_{post}, o_2 = 1 \\ \frac{dr_3}{dt} = -\frac{r_3}{\tau_{x3}}, \text{ if } t = t_{pre}, r_3 = 1 \\ \frac{do_3}{dt} = -\frac{o_3}{\tau_{y3}}, \text{ if } t = t_{post}, o_3 = 1 \end{cases} \quad (9)$$

Weight change is defined as follows

$$\Delta w = \begin{cases} -o_1(t) [A_1^- + A_2^- r_2(t-\varepsilon) + A_3^- r_3(t-\varepsilon) r_2(t-\varepsilon)], \\ \text{ if } t = t_{pre} \\ +r_1(t) [A_1^+ + A_2^+ o_2(t-\varepsilon) + A_3^+ o_3(t-\varepsilon) o_2(t-\varepsilon)], \\ \text{ if } t = t_{post} \end{cases} \quad (10)$$

where parameters τ_{x1} , τ_{y1} , τ_{x2} , τ_{y2} , τ_{x3} and τ_{y3} are the time constant of the potentiation decay (r_1 , r_2 , r_3) or depression (o_1 , o_2 , o_3) potentials respectively. The parameters A_2^- , A_3^- , A_4^- , A_2^+ , A_3^+ , and A_4^+ are learning rates of the Q-STDP rule.

III. FAULT-TOLERANT SPIKE ROUTING

Previous research has focused on the fault-tolerant routing scheme of NoC topology. The routing system based on virtual channel divides a physical link into multiple virtual channels, but the control circuit is complex and the hardware power consumption is high. In addition, the routing system based on path location needs a large number of routing tables, resulting in high hardware costs. On the contrary, the bypass method divides the fault node into non overlapping areas, and uses the links around the normal node and the fault area to form a new path. When the packet reaches the fault area, it will bypass the new routing path according to certain rules, so as to avoid

unstoppable failures. Therefore, a spike fault-tolerant routing algorithm based on neuromorphic network is proposed.

A. Fault tolerance for neuromorphic systems

The reliability of neuromorphic systems is a crucial concern for ensuring the robustness of hardware implementations of SNNs. Consequently, it becomes imperative to explore stable fault-tolerant technologies during the design process of neuromorphic systems. The spiking sequences within the network are transmitted as packets through the links present in the NoC system. Throughout the neuromorphic hardware implementation process, various errors may arise, which can be classified into three main types: soft faults, hard faults, and transition defects. Soft faults primarily pertain to transition delays, such as instances where the resistance value increases, causing the membrane potential of the post-neuron to fail in generating timely spiking outputs. On the other hand, hard faults denote transition failures, such as when an interruption in a link disrupts the transmission of spiking packets. In the case of faults affecting the connection link between two interconnected neurons, the postsynaptic neuron either enters a silent state or a near-silent state. The silent state manifests as a resting potential, wherein no output is generated, while the near-silent state signifies low firing rates. Hard faults impede the transmission of information, resulting in reduced firing activity that adversely affects the performance of SNN models. Therefore, we propose a novel fault-tolerant routing algorithm specifically designed to address hard faults encountered during the neuromorphic hardware implementation.

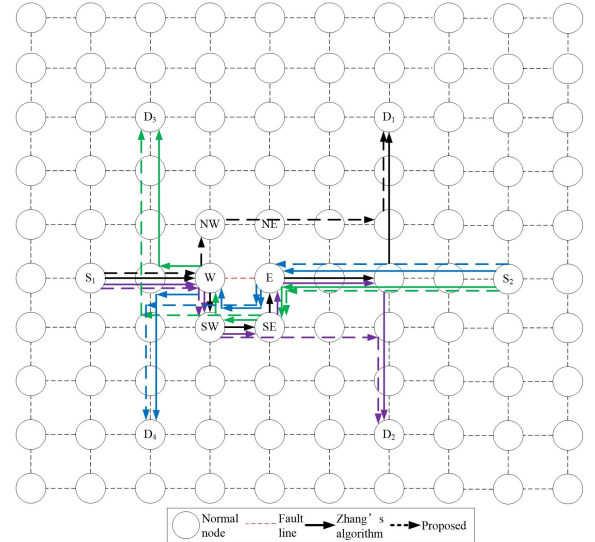


Fig. 3. The LFNT routing scheme along X direction.

B. Link-fault neuromorphic tolerant (LFNT) algorithm for spike routing

Fault link in the case of neuromorphic information routing along X direction is divided into two types. The first case is that the fault link is horizontal, where its direct adjacent nodes are W and E nodes. The second case is S and D nodes are in the east and west of the fault link, where $x_D \leq x_W < x_E \leq x_S$ or $x_S \leq x_W < x_E \leq x_D$. Besides, the S node is on the line of the fault link.

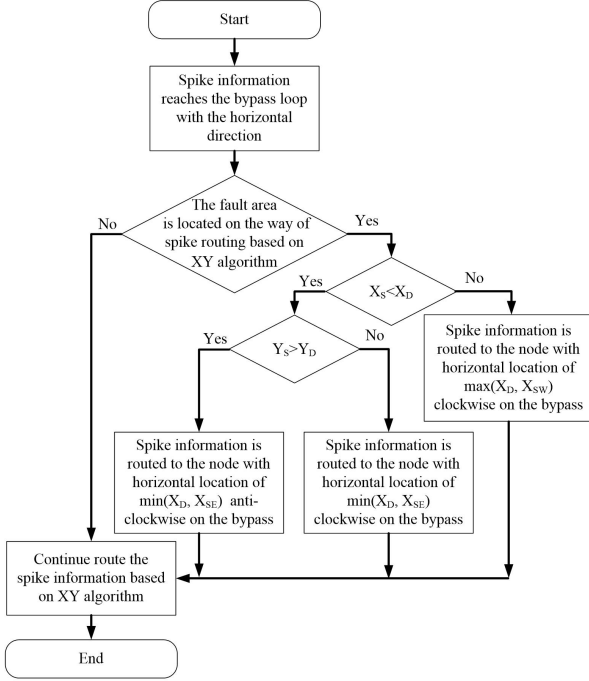


Fig. 4. Optimized bypass routing algorithm for spike information along X direction.

When the information packet meets the link fault along X direction, the spiking information carried on packet should be routed along the bypass loop for half a cycle. Building upon this strategy, it is important to note that the information packet must make two additional turns along the X direction. While this approach may enhance the bypass distance and impose an increased load on the left half, it also introduces a prolonged recovery time. Nonetheless, it is crucial to avoid excessively long recovery periods as they can result in extended latency, adversely affecting the timely generation of spiking activity in postsynaptic neurons. Given that temporal information plays a critical role in learning and information processing of SNNs that rely on learning based on STDP rules and computing based on spike events, prolonged recovery time significantly undermines the performance of perception and decision making tasks. After reaching the line of the source node, the data will be routed to the destination node D based on XY routing algorithm. The cases are demonstrated as the routing pathways from S_1 to D_1/D_2 and from S_2 to D_5 in Fig. 3. In these cases, for reducing the routing distance of the bypass loop and the routing distance from source to destination nodes, the presented bypass strategy optimizes the bypass strategy along the X direction. When information packet reaches bypass loop, the location relationship between current node C and the destination node D is judged. When $x_S \leq x_W \leq x_E \leq x_D$, the information packet is routed clockwise along the bypass loop in the case of $y_S < y_D$, and routed based on XY algorithm after reaching NE node. The information packet is routed anticlockwise if $y_S \geq y_D$, and the information packet is routed based on XY algorithm after the arrival of SE node. When $x_S \geq x_W > x_E \geq x_D$, the information packets are routed clockwise, and based on XY algorithm after reaching SW node. It shows that the presented bypass strategy can reduce one hop of the routing distance on the bypass loop,

and the routing distance from source to destination nodes can be reduced by 2 hops except the cases of $x_S \leq x_W \leq x_E \leq x_D$ and $y_S < y_D$. Reducing the number of hops means saving time of fault recovery, which is critical for the training performance of SNN based on STDP rules. The optimization induced by LFNT algorithm can not only reduce the data load on the bypass loop, but also maintain the routing distance from source to destination nodes. The detailed LFNT algorithm along X direction is described in Fig. 4. The symbols of X_S , X_D , X_{SW} , and X_{SE} represent the horizontal ordinates of the nodes S, D, SW, and SE respectively. In addition, the symbols of Y_S and Y_D represent the vertical ordinates of the nodes S and D. As a result, the preservation of essential spike timing information on the neuromorphic architecture is effectively ensured, enabling the implemented SNN to learn based on STDP rules and process spike-based information with minimal errors.

Another case is that there exists link fault along Y direction when the S and D nodes are located with the relative positions in the following two cases. The first case is that the fault link is longitudinal, where the direct adjacent nodes are S and N nodes. The second case is the S and D nodes are on the south and north sides of fault link, where $y_D \leq y_{South} < y_N \leq y_S$ or $x_S \leq x_W < x_E \leq x_D$, and D node is in the column of fault link with $x_D = x_N = x_S$. LFNT algorithm for spike routing along Y direction, algorithm procedure and deadlock-free proof are described in Supplementary material.

In light of the aforementioned fault-tolerant spike routing algorithm for neuromorphic architecture, it offers an effective optimization scheme for selecting bypass loops, thereby minimizing the recovery time when neuromorphic fault occurs. It is worth noting that an excessively prolonged error recovery time can result in extended latency periods, thereby impeding the timely generation of spiking activity in postsynaptic neurons. Given that temporal information plays a crucial role in SNNs that rely on STDP rules for learning and spikes processing, an extended recovery time significantly undermines perception and decision-making performance. To address these concerns, the LFNT routing algorithm proposed for spike routing effectively reduces fault recovery time, consequently minimizing communication delays induced by architecture fault. By ensuring fault-tolerant performance without compromising routing information, the proposed algorithm contributes to the overall robustness and reliability of neuromorphic systems.

IV. DIGITAL NEUROMORPHIC ARCHITECTURE

A. Digital neuromorphic architecture of the proposed spiking network

In order to efficiently realize the system on BiCoSS system, a time-multiplexing method is used to leverage the high clock frequency of the digital neuromorphic circuit. The time-multiplexing technique is limited by the available on-chip memory resource. Off-chip memory can solve this problem, but the bandwidth constraint will limit its realization and it requires more complex architecture for a dedicated memory controller. Fig. 5 shows the digital neuromorphic architecture of the proposed system, including an input layer with 36 Poisson

neurons, a time-multiplexed system containing an excitatory, an inhibitory, and an output layer. Multipliers are replaced by multiplexers so that significant hardware resources on FPGA can be saved. Each input pixel is responsible to control a 2-input multiplexer. An input buffer is used to store the input digits transiently, and the updating weight buffer outputs the weight values into the multiplexers, and then the computed values are summed up using a parallel adder. The time-multiplexed system is shown in Fig. 5(c). There are four computing components in this system, containing an excitatory physical neuron, an inhibitory physical neuron, an excitatory weight buffer and an inhibitory weight buffer. A global counter is used to control the four modules simultaneously. The excitatory and inhibitory weight buffers store the values of the synaptic weight strength for time-multiplexing method. By using this technique, we can realize a larger neural network using limited memory resource.

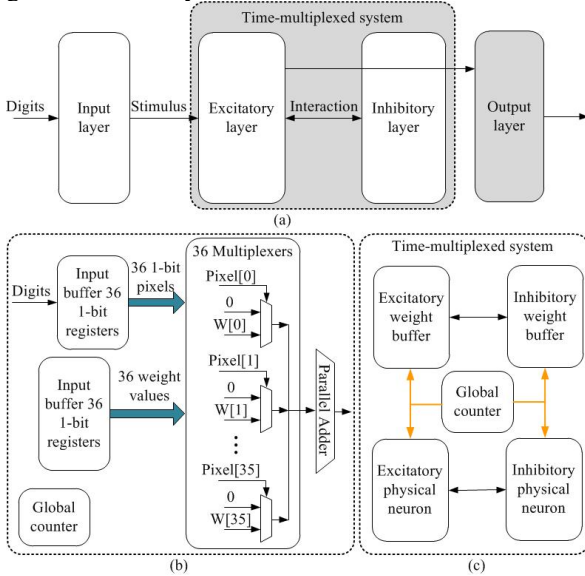


Fig. 5. Digital neuromorphic architecture of the proposed system. (a) Top-level entity of digital neuromorphic system. (b) The detailed architecture of the input layer. (c) Internal architecture of the time-multiplexed system.

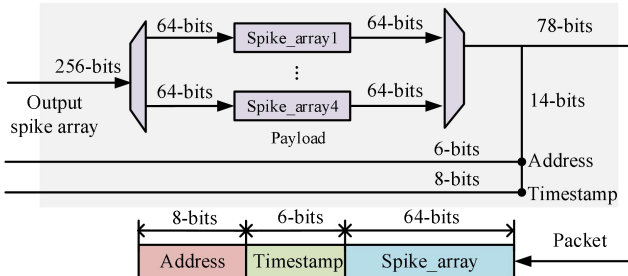


Fig. 6. The interface encoder for LFNT-based spike routing strategy.

B. Network Interface for LFNT routing

The network interface for LFNT routing is the communication interface link between LFNT and spiking nodes. It consists of two modules: the encoder and decoder. As shown in Fig. 6, the encoder receives spiking output vector and then encodes it into packet before sending to the postsynaptic spiking neuron. The following six bits indicates the address of

source node and the x and y direction each containing four bits. The next six bits is the time information about the spiking output of source neuron. The last 64 bits is used to represent the spike vector. Corresponding to the encoder, Fig. 7 describes the decoder of network interface between LFNT and spiking nodes, which receiving the packets, and then decoding it into spiking sequence before sending to the next node.

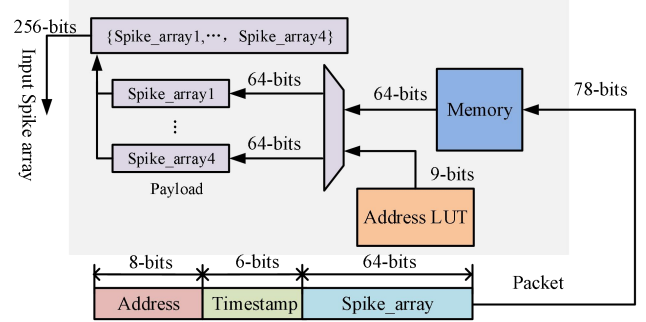


Fig. 7. The interface decoder for LFNT-based spike routing strategy.

V. EXPERIMENTAL RESULTS

Experimental results including learning capabilities on different classification dataset, communication efficiency and fault-tolerant capability are presented in this section. The neuromorphic network is implemented by digital neuromorphic FPGA chip, which uses Intel Stratix III EP3SL340 FPGA device. We first explore the learning performance of the VPL region in the presented neuromorphic network. To compare the performance of Q-STDP, T-STDP, and D-STDP, we conducted on-chip learning for all the learning algorithms on BiCoSS using an EP4CE115 FPGA. Due to the inability of digital circuits using logic and arithmetic computational elements to directly implement the nonlinear functions, we employed a piecewise linear approximation method to implement the exponential parts in D-STDP and T-STDP. This method is widely used to address such challenges. The implementation of the exponential parts in D-STDP and T-STDP using this approximation technique revealed that these algorithms are not hardware-friendly and may lead to reduced accuracy. An alternative approach involves using look-up tables for these exponential parts. However, both the time-division multiplexed technique for SNN implementation and the storage of weights and variables incur significant hardware memory costs. Consequently, implementing an end-to-end SNN model becomes resource-intensive for memory cost. In our study, we avoided using hardware memory resources by employing look-up tables to implement end-to-end SNNs with visual perception and decision-making functions. Instead, we utilized logic and arithmetic computational resources for the approximated computation of the exponential parts in the learning algorithms. Regarding the experimental results showing the superiority of Q-STDP over T-STDP, two reasons can be identified. Firstly, T-STDP incorporates exponential terms. To allocate memory resources for storing SNN network variables and weight parameters, as well as for time-division multiplexing of spiking neurons, we implemented T-STDP on neuromorphic hardware using logic and arithmetic computational resources. As a result, we approximated the

exponential terms using a piecewise linearization approach. However, this hardware-oriented approximation led to a decrease in the accuracy of T-STDP. Secondly, Q-STDP encompasses a greater number of learnable variables, enabling a potentially more flexible learning process for network weights. This may enhance flexibility facilitates weight updates during the plasticity computation process.

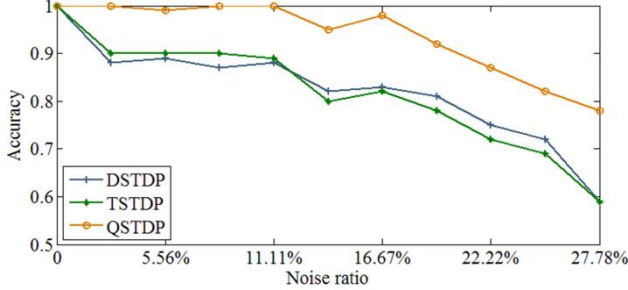


Fig. 8. Performance comparison of noise ratio on recognition performance with D-STDP, T-STDP and Q-STDP respectively.

As shown in Fig. 8, along with the increase of the noise ratio, the learning accuracy decreases. However, the neuromorphic network with Q-STDP obtains a higher learning accuracy in comparison with D-STDP and T-STDP learning rules. The presented Q-STDP rule is implemented on neuromorphic hardware, and on-chip learning is first realized during the training process. Then we fixed the synaptic weight to execute the end-to-end intelligence task to evaluate its performance in the testing phase. The intelligent agent is first trained to recognize the digits 1-6 with 6×6 pixels in each image. The camera input is preprocessed by a conventional convolution neural network, which simplify the image to 6×6 pixels. Then the decision-making process is trained and tested. The intelligent agent is trained to make the optimal decision. The training process of the proposed neuromorphic system can be considered as a two-phase training task. In the first phase, the proposed system is trained by the visual images until the VPL part of the proposed network can accurately recognize digits 1-6 with 6×6 pixels in each image. Only VPL module is employed, and DML module does not take part in this phase of training. In the second phase, the intelligent agent sequentially makes decision. If the intelligent agent reaches the reward, a positive feedback is obtained in the second phase. Otherwise, a negative feedback is presented. The second training phase continues until the decision making accuracy reaches the highest value. The presented two-phase training process is like the brain mechanism. Firstly, the human brain learns using STDP-based learning rule, and no evidence has proven the exist of the back propagation mechanism. Secondly, the brain will divide the image recognition and navigation tasks and learn each task respectively without learning at the same time. For example, we learn the digit in class and apply the knowledge in the daily navigation activities.

We first investigate the effects of the network scale on the recognition accuracy in the proposed neuromorphic network with T-STDP and Q-STDP learning rules, which are both trained using 5000 epochs. As shown in Fig. 9(a), with the increment of the network scale, the recognition accuracy is

improved until the network scale is larger than 100. Since the basic pixel number of image is 36, different colors of lines represent different levels of noise with 0%, 5.56% (1/36), 11.11%, 16.67%, 22.22%, and 27.78% ratios. The under-fitting means that more training epochs are required for the larger network. The learning capability of Q-STDP is shown in Fig. 9(b). The experimental result shows that the learning accuracy can be enhanced by using Q-STDP with different levels of noise. The learning capability of Q-STDP is superior to that of T-STDP along with the increment of network scale, and Q-STDP can help the network to avoid the over-fitting problem.

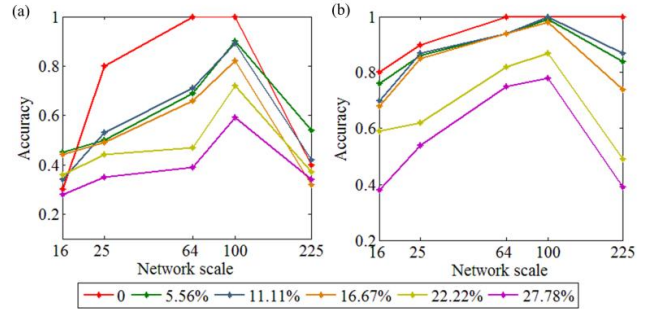


Fig. 9. Effects of the network scale on the learning performance. (a) Network with T-STDP learning rule. (b) Network with Q-STDP learning rule.

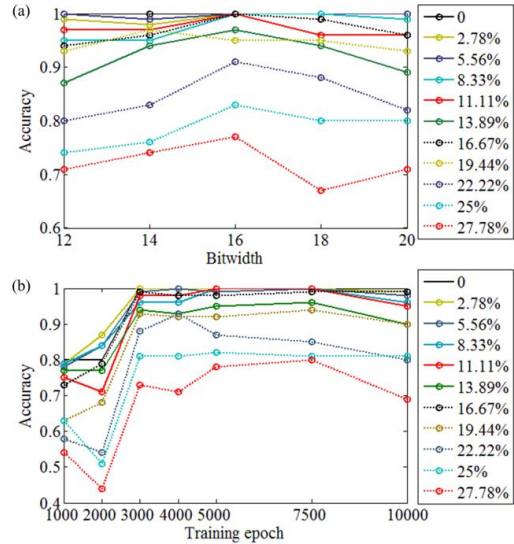


Fig. 10. Performance analysis of the proposed neuromorphic network with the Q-STDP learning rule.

Since the proposed neuromorphic network uses the fixed-point representation, the effects of the bit width for the fractional part of the trained synaptic weight on the final recognition accuracy are explored in Fig. 10(a). Different colors of lines represent different levels of noise with 0%, 5.56%, 11.11%, 16.67%, 22.22%, and 27.78% ratios. Interestingly, the recognition accuracy increases along with the bitwidth increasing, and then decreases when the bitwidth is more than 16 bit. It shows that a computation precision with 16 bit width can reach the maximum classification accuracy. Fig. 10(b) shows the convergence learning rate based on Q-STDP rule. It shows that training within 3000 epochs can fast reach a

high-level of classification accuracy for all the conditions of noise rates.

Comparisons with previous studies can help to analyze the relationship between the fault-tolerant feature of the NoC and the overall performance of the SNN in more depth. The fault-tolerant mechanism implemented in the NoC plays a crucial role in ensuring reliable communication and information processing within the SNN, ultimately impacting its end-to-end performance. By examining the results presented in Fig. 11(a) and (b), we can observe the effects of different fault rates on the average latency and throughput, comparing them with the results of previous studies [56] and [57]. The experimental setup involves varying levels of fault rates, and the performance metrics are evaluated in terms of average latency and throughput. It is evident that as the fault rate increases, the throughput gradually decreases for all algorithms, as shown in Fig. 11(a). This decrease in throughput is primarily attributed to the increased number of spikes being processed to accomplish the end-to-end visual perception and decision making task in the SNN. Furthermore, Fig. 11(b) demonstrates that the average latency experiences a noticeable increase of 22% when the fault rate reaches 30%. This rise in average latency is a direct result of the increased spike count propagated through the neuromorphic routing caused by fault-induced delays. The proposed fault-tolerant scheme proves to outperform the previous works [56] and [57] even under high fault rates of up to 30%. These findings provide valuable insights into how the fault-tolerant mechanism of the NoC significantly impacts the end-to-end performance of the SNN. To delve deeper into the impact of the fault-tolerant NoC on the SNN's performance, Fig. 11(c) analyzes the decision-making accuracy across various spike arrival windows, including 0.1 μ s, 0.12 μ s, 0.14 μ s, and so on. The proposed method consistently achieves higher decision-making accuracy compared to the referenced studies [55] and [57]. The improved accuracy can be attributed to the increased number of spikes processed within the spike arrival window, leading to better decision-making capabilities. It is noteworthy that at a spike arrival window of 0.16 μ s, all methods reach saturation, indicating that no further improvement in decision-making accuracy can be achieved beyond this point. Additionally, Fig. 11(d) provides insights into the average computing time of the proposed method compared to the comparative studies. Notably, the proposed method consistently exhibits lower average computing time across different spike arrival windows, such as 0.10 μ s, 0.12 μ s, and 0.14 μ s, when compared to study [55]. When compared to study [57], the average computing time of the proposed method varies, being slightly larger in certain cases but significantly lower in others. These findings highlight the influence of the fault-tolerant NOC mechanism on the overall computational efficiency of the SNN, thereby impacting its end-to-end performance. Furthermore, Fig. 11(e) demonstrates the decision-making accuracy for different fault rates. The proposed method and study [55] maintain saturation accuracy from 0% to 30% fault rate, whereas the decision-making accuracy of study [56] shows a significant decrease of 55.10%. However, at a 40% fault rate, the proposed method achieves substantially higher accuracy compared to both study [55] and

study [56]. This analysis further emphasizes the importance of the fault-tolerant feature in maintaining the SNN's performance even under challenging conditions. Lastly, Fig. 11(f) provides insights into the average computing time of the proposed method compared to the comparative studies. The proposed method consistently showcases lower average computing time, even at a 40% fault rate, outperforming study [56] and study [55]. These results highlight the effectiveness of the fault-tolerant NoC mechanism in achieving efficient information processing within the SNN. In conclusion, the thorough analysis of the proposed system highlights the significant impact of the fault-tolerant NoC feature on the end-to-end performance of the SNN. The comparisons with previous studies reveal the superiority of the proposed method, demonstrating its ability to deliver high accuracy while operating within smaller spike arrival windows and under faulty conditions. The fault-tolerant NoC mechanism plays a critical role in ensuring reliable communication, minimizing latency, and maximizing throughput, thereby enhancing the overall performance of the SNN.

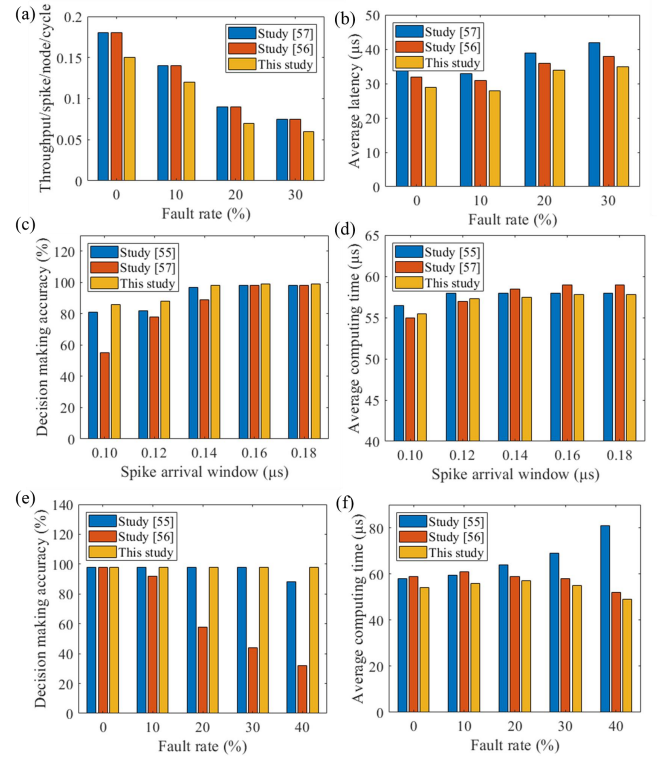


Fig. 11. Comparison with other work over various fault rate and spike arrival window. (a) Throughput. (b) Average latency. (c) Decision making accuracy over various spike arrival window. (d) Average computing time over various spike arrival window. (e) Decision making accuracy over various fault rate. (f) Average computing time over various fault rate.

We further conducted ablation experiments to investigate the effects of the proposed LFNT strategy on the improvement of spiking rate loss and accuracy loss in STDP-based learning under different fault rates. The objective was to reveal the relationship between the fault-tolerant feature of the LFNT strategy and the intrinsic mechanisms of SNNs, as illustrated in Fig. 12(a) and Fig. 12(b). From Fig. 12(a), it is observed that faulty link transmissions result in a decrease in neuron spiking

rate, which can impact the performance of the SNN model. From a robustness perspective, neurons with lower spiking rates are more susceptible to noise and firing delays, leading to increased spiking rate variance and affecting the robustness of the SNN model. Faulty links cause the loss of some information carried by routing packets, and the changes in spiking rate can affect visual perception performance and result in erroneous decisions. Fig. 12(b) presents the accuracy loss in STDP-based learning under different fault rates, measured by visual perception accuracy. It can be observed that time information is crucial for SNNs learning based on the STDP rule. The loss of partial time information (without the LFNT strategy) significantly reduces the learning performance of STDP. These findings highlight the importance of the LFNT strategy's fault-tolerant feature in relation to the spiking rate and learning accuracy loss in SNNs. The relationship between the LFNT strategy and the intrinsic mechanisms of SNNs provides valuable insights into the performance improvements of SNNs under fault conditions.

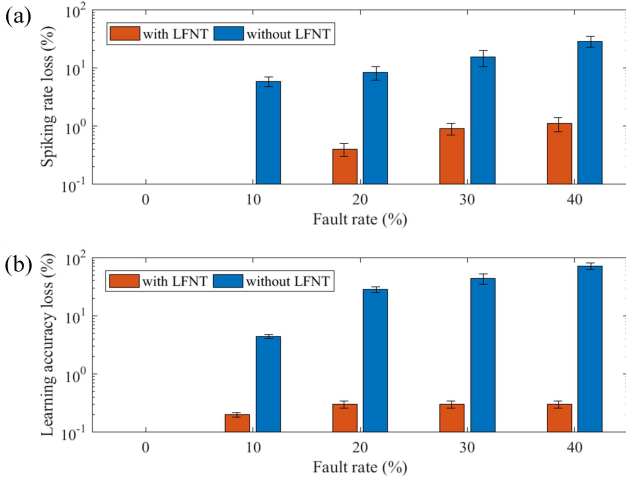


Fig. 12. Ablation study to evaluate the improvement of the proposed LFNT strategy on intrinsic performance of implemented SNNs. (a) Effects of LFNT on spiking rate loss. (b) Effects of LFNT on STDP-based learning accuracy loss.

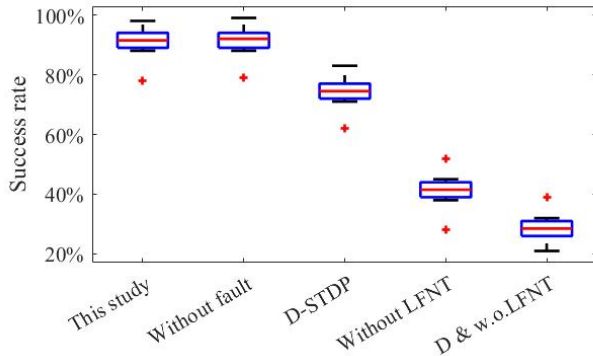


Fig. 13. Experimental results of the application of the proposed neuromorphic network.

The application of the proposed neuromorphic network in robot vehicle is designed as follows. The vehicle is first placed in the straight extension line of the digits while facing the midpoint of each pair of digits. Before the vehicle receives the digits images, it does not any decision, and just goes along the

straight extension line towards the digits. Before the computation of the VPL module, the vehicle has an additional judgement module to judge whether have seen the digits. After the digits have been seen by the vehicle, it triggers the DML module to make an action selection and finally controls the vehicle to turn either left or right. The detailed experimental design is presented in Supplementary materials. Then the decision making process is trained and tested. The experimental results are shown in Fig. 13, where five situations are considered. The first one is the presented neuromorphic network model with LFNT strategy, and the second one is the neuromorphic network without any fault. The third situation is to replace the proposed Q-STDP with conventional D-STDP. The fourth one is blocked by two links but without any fault-tolerant strategy. The fifth condition is the D-STDP situation without LFNT strategy (D & w.o.LFNT). It shows that the presented fault-tolerant neuromorphic framework can enhance the system performance significantly, which enables the controlled agent more robust and stable. Besides, in comparison with the network without fault, the proposed framework also has nearly consistent success rate. It reveals that the proposed method can significantly inhibit the negative impact on the success rate due to the neuromorphic fault.

VI. DISCUSSIONS AND CONCLUSIONS

This study presents a unified neuromorphic framework to integrate visual perception with decision making, which is meaningful for both neuroscience and artificial intelligence. As shown in Fig. 1, a hierarchical network architecture is presented, which contains VPL and DML for functioning the unified framework. The presented framework is inspired by neural system and implemented by SNN representation. A scenario setup is designed to test the functional effectiveness of the presented network. In the VPL region, a LIF dynamics with homoeostasis mechanism is presented as shown in Fig. 2. A novel Q-STDP learning rule is presented and applied in the presented network. It uses the precise timing of quadruplet spikes, which is a biologically plausible information coding scheme. The neural learning parameters in the hidden layer are trained online using Q-STDP mechanism. Since neuromorphic computing suffers from the inevitable faults, which hinders the spiking network stable and accuracy. As shown in Fig. 3, a novel LFNT algorithm is described, with their pseudo codes shown in Fig. 4. The top-level design of the presented neuromorphic system is shown in Fig. 5. The proposed LFNT algorithm takes into account the end-to-end information processing paradigm of SNN. It aims to minimize bypass routing time and facilitate spike routing recovery while preserving crucial time-sensitive information that significantly impacts SNN model performance. This fault-tolerant strategy effectively prevents information loss and spike timing errors caused by routing failures within SNN networks during learning and information processing, thereby mitigating perception and decision-making errors resulting from routing failures. The network interface's encoder and decoder are illustrated in Fig. 6 and Fig. 7, respectively. Fig. 8, Fig. 9 and Fig. 10 demonstrate that the Q-STDP learning rule outperforms the alternative D-STDP and T-STDP principles. As

demonstrated in Fig. 11, the average latency experiences a noticeable increase when the fault rate reaches 30%, primarily caused by fault-induced delays in the neuromorphic routing. The proposed LFNT scheme outperforms previous works even under high fault rates of up to 30%, highlighting its effectiveness. The proposed LFNT method consistently achieves higher decision-making accuracy compared to referenced studies across various spike arrival windows, attributed to the increased number of spikes processed within the window. The average computing time of the proposed method is consistently lower compared to comparative studies, emphasizing the impact of the LFNT strategy on the computational efficiency of the SNN. The proposed LFNT method maintains high decision-making accuracy from 0% to 30% fault rate, while other studies show significant decreases. At a 40% fault rate, the proposed LFNT method achieves substantially higher accuracy. The LFNT contributes to maintaining the SNN's performance even under challenging conditions. The proposed LFNT consistently showcases lower average computing time, even at a 40% fault rate, outperforming comparative studies. Therefore, the proposed LFNT strategy significantly impacts the end-to-end performance of the SNN, improving reliability, reducing latency, maximizing throughput, and achieving efficient information processing. LFNT demonstrates superiority in terms of accuracy, spike arrival windows, and fault tolerance compared to previous studies. Fig. 12 indicates that the LFNT strategy has a significant impact on reducing spiking rate loss and accuracy loss in STDP-based learning under different fault rates in SNNs. Faulty link transmissions were observed to decrease the neuron spiking rate, which can negatively affect the overall performance of the SNN model. Neurons with lower spiking rates were found to be more vulnerable to noise and firing delays, leading to increased spiking rate variance and impacting the robustness of the SNN model. The loss of information carried by routing packets due to faulty links and the resulting changes in spiking rate can affect visual perception performance and lead to erroneous decisions. The accuracy loss in STDP-based learning, as measured by visual perception accuracy, was influenced by the availability of time information. The absence of partial time information, without the LFNT strategy, significantly reduced the learning performance of STDP. These findings underscore the importance of the LFNT strategy's fault-tolerant feature in mitigating spiking rate and learning accuracy losses in SNNs. Understanding the relationship between the LFNT strategy and the intrinsic mechanisms of SNNs provides valuable insights into enhancing the performance of SNNs in fault conditions. The application in robot vehicle is demonstrated in Fig. 13, which shows a good performance by using the presented neuromorphic framework. Overall, the major contributions of this study include four aspects. Firstly, a novel unified framework with hierarchical SNN architecture is presented, which successfully integrates the visual perception and decision making. Secondly, a novel learning algorithm with quadruplet spiking mechanism is presented, which shows high learning capability compared to the D-STDP and T-STDP learning rules. Thirdly, a novel LFNT algorithm is presented to

enable the proposed neuromorphic network fault tolerant. Fourthly, a real-time application is realized on the robot vehicle, which demonstrates the computational efficiency and real-time capability of the proposed neuromorphic network.

The presented SNN algorithm, which integrates visual perception with decision making, exhibits three aspects that reflect non von Neumann architecture. Firstly, it harnesses inherent parallelism with multiple computational cores (spiking neurons), aligning with the distributed and massively parallel nature of neuromorphic systems based on non von Neumann architecture. SNNs, being a key component of non von Neumann architectures, possess massive parallelism, allowing for concurrent processing of information, similar to the distributed nature of non von Neumann systems. By utilizing the spiking neurons and synaptic connections, our algorithm exploits this parallelism to process visual input and make decisions concurrently, thereby exhibiting the characteristics of non von Neumann architectures. This enables efficient, real-time, and scalable implementation of large-scale neuromorphic networks, particularly advantageous for time-sensitive applications like sensory processing, robotics, and real-time decision making. Secondly, the event-driven processing nature in the presented SNN algorithm enables low-power operation on non von Neumann architecture through asynchronous spike-based computation and communication. This differs from traditional von Neumann architectures that rely on discrete-time processing with explicit synchronization and continuous processing. By incorporating event-driven processing, our algorithm aligns with the principles of non von Neumann architectures that aim to reduce power consumption and efficiently process information based on event-driven triggers. The presented work significantly reduces energy consumption, a crucial factor in edge computing and mobile applications where power efficiency is paramount. By tailoring algorithms to the underlying hardware, the full potential of neuromorphic systems can be unlocked, achieving superior performance in terms of speed, energy efficiency, fault-tolerance, and scalability. Thirdly, neuromorphic NoC technique is employed without the need for transferring data between separate memory and processing units by a bus on conventional Von Neumann architecture. By integrating computation and memory within the NoC architecture, our work showcases the efficiency and data locality benefits, akin to the principles of non von Neumann architectures.

Neuromorphic computing represents a promising approach to achieving brain-inspired intelligence. The field of neuromorphic architecture and hardware implementation has attracted increasing attention in the context of real-time artificial intelligence tasks, primarily due to the low energy consumption and silicon area costs associated with these studies [22]-[33]. Notable examples of digital neuromorphic systems include TrueNorth, SpiNNaker, Loihi, and Tianjic, all of which enable the realization of reconfigurable SNNs on digital hardware. However, it is important to note that TrueNorth lacks online learning capabilities [23]-[24]. Examining Table I, we observe that although there are several neuromorphic computing studies employing various hardware

architectures, with the exception of SpiNNaker, none of them specifically consider fault-tolerant capabilities [22], [25], [29]-[33]. In this study, a novel neuromorphic architecture has been proposed, integrating fault-tolerant capabilities and online learning ability, thereby enabling more intricate and intelligent cognition. This advancement addresses the current limitations in existing neuromorphic systems.

In addition to the comparison of neuromorphic hardware, it is essential to acknowledge the existing neuromorphic algorithms developed for visual perception and decision-making tasks. It should be emphasized that neuromorphic computation is not restricted to computation solely on neuromorphic chips. An example is Spaun, which, although not implemented on neuromorphic hardware, has successfully executed the functions required for cognitive tasks. Spaun flexibly coordinates neural groups based on cognitive tasks, including visual perception and decision making [49]. Croon et al. introduced SNN solutions to insect-inspired artificial intelligence for autonomous robots, which can fully implement visual perception and decision making functions on a neuromorphic processor [51]. Other notable achievements include Truenorth [23], DYNAP [29], and Loihi [30], which efficiently implement visual perception functions on digital neuromorphic architectures, and LaCSNN [34], which has demonstrated its decision-making capabilities. Researchers have also conducted real-time visual perception and decision-making models based on SNNs using SpiNNaker [24]. Additionally, BiCoSS has implemented a multi-granular hierarchical heterogeneous neuromorphic architecture that unifies both visual perception and decision-making tasks within a single system [35]. Furthermore, Tianjic has realized a cross-paradigm neuromorphic architecture, showcasing the application of a hybrid paradigm in end-to-end visual perception and decision-making tasks [50]. To achieve stable and effective implementation of SNN algorithms for end-to-end cognition tasks on digital neuromorphic systems, it is imperative to develop fault-tolerant strategies for neuromorphic architectures. These strategies would further enhance the performance of these representative research endeavors.

TABLE I
COMPARISON WITH THE STATE-OF-THE-ART WORKS FOR NEUROMORPHIC
HARDWARE ARCHITECTURES

Studies	Fault-tolerant	Learning
Truenorth [23]	No	No
SpiNNaker [24]	Yes	Yes
Neurogrid [22]	No	No
HiAER [25]	No	No
DYNAP [29]	No	Yes
Loihi [30]	No	Yes
DeepSouth [31]	No	No
Study [32]	No	No
Study [33]	No	Yes
LaCSNN [34]	No	Yes
BiCoSS [35]	No	Yes
Tianjic [50]	No	Yes
Study [44]	No	Yes
This study	Yes	Yes

As to future works, the neural coding is based on not only synaptic inputs, but also on the neural dynamical characteristics [37]. Thus, to consider the detailed neural dynamics may induce a higher performance. For example, Zhang et al. presented a novel algorithm with dynamic threshold to be robust to noise [38]. In addition, another further application of the proposed study is the brain-machine integration [39]-[41]. Neuromorphic systems with online learning and cognitive functions can enable the self-adaption and control of the intelligent machines that are interfaced with biological brain. Previous study has revealed that on-chip online learning makes the machine to decode signals from mammalian brain to control signals of an external device [39]. In conclusion, this study presents a unified spike-based framework for the integration of visual perception with decision making. Besides, a novel Q-STDP learning rule with LFNT fault-tolerant algorithm is presented, which enables the presented model online learning and fault tolerant. Compared to other studies, this study is superior considering different aspects of the computational capability. Therefore, this work can be applied in various kinds of fields, including autonomous robotic control, brain-machine integration, and mechanisms comprehension for neuroscience community.

ACKNOWLEDGMENT

The authors would like to thank the editor and the reviewers for their critical and constructive comments and suggestions.

REFERENCES

- [1] T. P. Lillicrap, A. Santoro, L. Marris *et al*, "Backpropagation and the brain," *Nat. Rev. Neurosci.*, pp. 1-12, 2020.
- [2] J. C. R. Whittington and R. Bogacz, "Theories of error back-propagation in the brain," *Trends Cognit. Sci.*, 2019.
- [3] J. P. Gallivan, C. S. Chapman, D. M. Wolpert *et al*, "Decision-making in sensorimotor control," *Nat. Rev. Neurosci.*, vol. 19, no. 9, pp. 519-534, 2018.
- [4] R. G. O'Connell, M. N. Shadlen, K. F. Wong-Lin *et al*, "Bridging neural and computational viewpoints on perceptual decision-making," *Trends Neurosci.*, vol. 41, no. 11, pp. 838-852, 2018.
- [5] A. P. Johnson, J. Liu, A. G. Millard *et al*, "Time-multiplexed system-on-chip using fault-tolerant astrocyte-neuron networks," *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1076-1083, 2018.
- [6] P. J. Vance, G. P. Das, D. Kerr, *et al*, "Bioinspired approach to modeling retinal ganglion cells using system identification techniques," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1796-1808, 2017.
- [7] S. Yang, B. Chen. "SNIB: Improving Spike-Based Machine Learning Using Nonlinear Information Bottleneck," *IEEE Trans. Syst. Man, Cybern. Syst.*, 2023.
- [8] S. Yang, Y. Pang, H. Wang, *et al*. "Spike-driven multi-scale learning with hybrid mechanisms of spiking dendrites," *Neurocomputing*, 542: 126240, 2023.
- [9] S. Yang, J. Tan, B. Chen. "Robust spike-based continual meta-learning improved by restricted minimum error entropy criterion," *Entropy*, 2022, 24(4): 455.
- [10] S. Yang, B. Linares-Barranco, B. Chen. "Heterogeneous ensemble-based spike-driven few-shot online learning," *Front. Neurosci.*, 2022, 16: 850932.
- [11] M. Mozafari, S. R. Kheradpisheh, T. Masquelier, *et al*. "First-spike-based visual categorization using reward-modulated STDP," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 12, pp. 6178-6190, 2018.
- [12] L. Zhang, H. Tao, C. Holt *et al*, "A critical window for cooperation and competition among developing retinotectal synapses," *Nature*, vol. 395, no. 6697, pp. 37-44, 1998.
- [13] A. D. Rast, S. V. Adams, S. Davidson, *et al*. "Behavioral learning in a cognitive neuromorphic robot: an integrative approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 12, pp. 6132-6144, 2018.

- [14] J-P. Pfister and W. Gerstner, "Triplets of Spikes in a Model of Spike Timing-Dependent Plasticity," *J. Neurosci.*, vol. 26, no. 38, pp. 9673-9682, 2006.
- [15] J. Gjorgjieva, C. Clopath, J. Audett *et al.*, "From the Cover: A triplet spike-timing-dependent plasticity model generalizes the Bienenstock-Cooper-Munro rule to higher-order spatiotemporal correlations," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 48, pp. 19383, 2011.
- [16] E. Jokar, H. Soleimani, "Digital multiplierless realization of a calcium-based plasticity model," *IEEE Trans. Circuits Syst. II: Express Briefs*, vol. 64, no. 7, pp. 832-836, 2016.
- [17] H. X. Wang, R. C. Gerkin, D. W. Nauen *et al.*, "Coactivation and timing-dependent integration of synaptic potentiation and depression," *Nat. Neurosci.*, vol. 8, no. 2, pp. 187-193, 2005.
- [18] P. J. Sjöström, E. A. Rancz, A. Roth *et al.*, "Dendritic Excitability and Synaptic Plasticity," *Physiol. Rev.*, vol. 88, no. 2, pp. 769-840, 2008.
- [19] J. Gjorgjieva, C. Clopath, J. Audet *et al.*, "From the Cover: A triplet spike-timing-dependent plasticity model generalizes the Bienenstock-Cooper-Munro rule to higher-order spatiotemporal correlations," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 48, pp. 19383, 2011.
- [20] A. Taherkhani, A. Belatreche, Y. Li *et al.*, "A review of learning in biologically plausible spiking neural networks," *Neural Netw.*, vol. 122, pp. 253-272, 2020.
- [21] H. Shen, T. Wang, J. Cao, *et al.*, "Nonfragile dissipative synchronization for Markovian memristive neural networks: a gain-scheduled control scheme," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 30, no. 6, pp. 1841-1853, 2018.
- [22] B. V. Benjamin, P. Gao, E. McQuinn *et al.*, "Neurogrid: A Mixed-Analog-Digital Multipip System for Large-Scale Neural Simulations," *Proc. IEEE*, vol. 102, no. 5, pp. 699-716, 2014.
- [23] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza *et al.*, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science*, vol. 345, no. 6197, pp. 668-673, 2014.
- [24] E. Painkras, L. A. Plana, J. Garside *et al.*, "SpiNNaker: A 1-W 18-core system-on-chip for massively-parallel neural network simulation," *IEEE J. Solid-State Circuits*, vol. 48, no. 8, pp. 1943-1953, 2013.
- [25] J. Park, T. Yu, C. Maier *et al.*, "Hierarchical Address-Event Routing architecture for reconfigurable large scale neuromorphic systems," *International Symposium on Circuits & Systems*, 2012.
- [26] S. Yang, H. Wang, Y. Pang, *et al.*, "NADOL: Neuromorphic Architecture for Spike-driven Online Learning By Dendrites," *IEEE Trans. Biomed. Circ. Syst.*, 2023.
- [27] S. Yang, J. Tan, T. Lei, *et al.*, "Smart traffic navigation system for fault-tolerant edge computing of internet of vehicle in intelligent transportation gateway," *IEEE Trans. Intell. Transp. Syst.*, 2023.
- [28] B. Deng, Y. Fan, J. Wang, *et al.*, "Auditory perception architecture with spiking neural network and implementation on FPGA," *Neural Netw.*, 2023, 165: 31-42.
- [29] S. Moradi, N. Qiao, F. Stefanini *et al.*, "A Scalable Multicore Architecture With Heterogeneous Memory Structures for Dynamic Neuromorphic Asynchronous Processors (DYNAPs)," *IEEE Trans. Biomed. Circ. and Syst.*, vol. PP, no. 99, pp. 106, 2017.
- [30] M. Davies, N. Srinivasa, T. H. Lin *et al.*, "Loihi: A Neuromorphic Manycore Processor with On-Chip Learning," *IEEE Micro*, pp. 1-1, 2018.
- [31] R. M. Wang, C. S. Thakur and S. A. Van, "An FPGA-Based Massively Parallel Neuromorphic Cortex Simulator," *Front. Neurosci.*, vol. 12, pp. 213-231, 2018.
- [32] C. Kit, S. R. Schultz and L. Wayne, "NeuroFlow: A General Purpose Spiking Neural Network Simulation Platform using Customizable Processors," *Front. Neurosci.*, vol. 9, no. 19, pp. 516, 2016.
- [33] P. Danilo, M. Paolo, T. Giuseppe *et al.*, "An FPGA Platform for Real-Time Simulation of Spiking Neuronal Networks," *Front. Neurosci.*, vol. 11, pp. 90, 2017.
- [34] S. Yang, H. Wang, M. R. Azghadi, "Evolutionary Optimization of Neuromorphic Architecture for Low-power Cerebellum Prosthetic Instrumentation and Device in Biomedical Systems," *2023 IEEE International Instrumentation and Measurement Technology Conference*, 2023: 1-6.
- [35] S. Yang, B. Deng, J. Wang *et al.*, "BiCoSS: Toward large-scale cognition brain with multigranular neuromorphic architecture" *IEEE Trans. Neural Netw. Learn. Syst.*, 2021.
- [36] K. H. Chen and G. M. Chiu, "Fault-tolerant routing algorithm for meshes without using virtual channels," *J. Inform. Sci. Eng.*, vol. 14, no. 4, pp. 765-783, 1997.
- [37] M. Eugene and Izhikevich, "Simple model of spiking neurons," *IEEE Trans. Neural Netw.*, vol. 14, pp. 1569-1572, 2003.
- [38] M. Zhang, H. Qu, X. Xie *et al.*, "Supervised learning in spiking neural networks with noise-threshold," *Neurocomputing*, vol. 219, no. JAN.5, pp. 333-349, 2017.
- [39] F. Boi, T. Moraitis, V. DeFeo *et al.*, "A bidirectional brain-machine interface featuring a neuromorphic hardware decoder," *Front. Neurosci.*, vol. 10, pp. 563, 2016.
- [40] A. Basu, J. Acharya, T. Karnik *et al.*, "Low-power, adaptive neuromorphic systems: Recent progress and future directions," *IEEE J. Emerg. Sel. Top. Cir. Syst.*, vol. 8, no. 1, pp. 6-27, 2018.
- [41] Y. Wen, M. Li, J. Si, *et al.*, "Wearer-prosthesis interaction for symmetrical gait: a study enabled by reinforcement learning prosthesis control," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 28, no. 4, pp. 904-913, 2020.
- [42] J. Duato, "A new theory of deadlock-free adaptive multicast routing in wormhole networks," *IEEE Symposium on Parallel & Distributed Processing. IEEE*, 1994.
- [43] S. Luo, F. L. Lewis, Y. Song, *et al.*, "Accelerated adaptive fuzzy optimal control of three coupled fractional-order chaotic electromechanical transducers," *IEEE Trans. Fuzzy Syst.*, 2020.
- [44] S. Yang, J. Wang, N. Zhang, *et al.*, "CerebelluMorphic: Large-Scale Neuromorphic Model and Architecture for Supervised Motor Learning," *IEEE Trans. Neural Netw. Learn. Syst.*, 2021.
- [45] F. Paredes-Vallés, K. Y. W. Scheper, G. C. H. E. de Croon, "Unsupervised learning of a hierarchical spiking neural network for optical flow estimation: From events to global motion perception," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2051-2064, 2019.
- [46] A. Viale, A. Marchisio, M. Martina, G. Masera and M. Shafique, "CarSNN: An efficient spiking neural network for event-based autonomous cars on the Loihi neuromorphic research processor," *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, pp. 1-10, Jul. 2021.
- [47] R. Massa, A. Marchisio, M. Martina, *et al.*, "An efficient spiking neural network for recognizing gestures with a dvs camera on the loihi neuromorphic processor", *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, pp. 1-9, Jul. 2020.
- [48] E. Angelidis, E. Buchholz, J. Arreguit, *et al.*, "A spiking central pattern generator for the control of a simulated lamprey robot running on SpiNNaker and Loihi neuromorphic boards," *Neuromorphic Comput. Eng.*, vol. 1, no. 1, pp. 014005, 2021.
- [49] C. Eliasmith, T. C. Stewart, X. Choo, *et al.*, "A large-scale model of the functioning brain," *Science*, vol. 338, no. 6111, pp. 1202-1205, 2012.
- [50] L. Deng, G. Wang, G. Li, *et al.*, "Tianjic: A unified and scalable chip bridging spike-based and continuous neural computation," *IEEE J. Solid-State Circuits*, vol. 55, no. 8, pp. 2228-2246, 2020.
- [51] G. C. H. E. De Croon, J. J. G. Dupeyroux, S. B. Fuller and J. A. R. Marshall, "Insect-inspired AI for autonomous robots", *Science Robotics*, vol. 7, no. 67, pp. eab16334, 2022.
- [52] O. M. Ikechukwu, K. N. Dang, A. B. Abdallah, "On the Design of a Fault-Tolerant Scalable Three Dimensional NoC-Based Digital Neuromorphic System With On-Chip Learning," *IEEE Access*, vol. 9, pp. 64331-64345, 2021.
- [53] 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, 2019.
- [54] T. H. Vu, Y. Okuyama, and A. B. Abdallah, "Comprehensive analytic performance assessment and K-means based multicast routing algorithm and architecture for 3D-NoC of spiking neurons," *ACM J. Emerg. Technol. Comput. Syst.*, vol. 15, no. 4, pp. 1-28, Dec. 2019.
- [55] Y. Fukushima, M. Fukushi, S. Horiguchi, "Fault-tolerant routing algorithm for network on chip without virtual channels," *2009 24th IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems. IEEE*, pp. 313-321, 2009.
- [56] Y. Fukushima, M. Fukushi, I. E. Yairi, *et al.*, "A hardware-oriented fault-tolerant routing algorithm for irregular 2D-mesh network-on-chip without virtual channels," *2010 IEEE 25th International Symposium on Defect and Fault Tolerance in VLSI Systems. IEEE*, pp. 52-59, 2010.
- [57] R. V. W. Putra, M. A. Hanif, M. Shafique, "Respawn: Energy-efficient fault-tolerance for spiking neural networks considering unreliable memories," *2021 IEEE/ACM International Conference On Computer Aided Design (ICCAD). IEEE*, pp. 1-9, 2021.
- [58] E. Baek, H. Lee, Y. Kim and J. Kim, "Flexlearn: fast and highly efficient brain simulations using flexible on-chip learning", *Proc. 52nd Annu. IEEE/ACM Int. Symp. Microarchitect.*, pp. 304-318, 2019.



Shuangming Yang (Member, IEEE) received his Ph.D. degree from Tianjin University, Tianjin, China in 2020 and was a lecturer from 2020 to 2021. He is currently an associate professor in the School of Electrical and Information Engineering, Tianjin University. His research interests include artificial general intelligence, neuromorphic computing, brain-inspired computing, and computer vision.



Haowen Wang received the bachelor degree in control science and engineering from Xi'an university of technology, Xi'an, China, in 2022. She is currently pursuing for a master's degree at Tianjin University, Tianjin, China. Her current research interest is brain-inspired computing.



Yanwei Pang (Senior Member, IEEE) received the Ph.D. degree in electronic engineering from the University of Science and Technology of China, Hefei, China, in 2004. He is currently a Professor with Tianjin University, China, and also the Founding Director of the Tianjin Key Laboratory of Brain Inspired Intelligence Technology (BIIT), China. His research interests include object detection and image recognition, in which he has published 150 scientific articles, including 40 IEEE

TRANSACTIONS articles and 30 top conferences (e.g., CVPR, ICCV, and ECCV) papers. He is an Associate Editor of both IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS (T-NNLS) and Neural Networks (Elsevier) and a Guest Editor of Pattern Recognition Letters.



Yaochu Jin (Fellow, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in automatic control from Zhejiang University, Hangzhou, China, in 1988, 1991, and 1996, respectively, and the Dr.-Ing. degree in neuroinformatics from Ruhr-University Bochum, Bochum, Germany, in 2001.

He is currently an Alexander von Humboldt Professor of Artificial Intelligence with the Chair of Nature-Inspired Computing and Engineering, Faculty of Technology, Bielefeld University, Bielefeld, Germany, and a Distinguished Chair of Computational Intelligence with the Department of Computer Science, University of Surrey, Guildford, U.K. He was a Finland Distinguished Professor in Finland and a Changjiang Distinguished Visiting Professor in China. His main research interests include data-driven evolutionary optimization, trustworthy machine learning, multiobjective evolutionary learning, and evolutionary developmental systems.

Dr. Jin is the recipient of the 2018 and 2021 IEEE Transactions on Evolutionary Computation Outstanding Paper Award, and the 2015, 2017, and 2020 IEEE Computational Intelligence Magazine Outstanding Paper Award. He was named "Highly Cited Researcher" for 2019-2021 by Clarivate Analytics. He is currently the Editor-in-Chief of Complex & Intelligent Systems. He is a member of Academia Europaea.



Bernabé Linares-Barranco (Fellow, IEEE) received the B.S. degree in electronic physics, the M.S. degree in microelectronics, and the first Ph.D. degree in high-frequency OTA-C oscillator design from the University of Seville, Seville, Spain, in June 1986, September 1987, and June 1990, respectively, and the second Ph.D. degree in analog neural network design from Texas A&M University, College Station, TX, USA, in

December 1991.

From September 1988 to August 1991, he was a Graduate Student at the Department of Electrical Engineering, Texas A&M University. Since June 1991, he has been a Tenured Scientist at the "Instituto de Microelectrónica de Sevilla," IMSE-CNM-CSIC, Seville. In January 2003, he was promoted to Tenured Researcher, and in January 2004 to Full Professor. Since February 2018, he has been the Director with the "Instituto de Microelectrónica de

Seville." He has been involved in circuit design for telecommunication circuits, VLSI emulators of biological neurons, VLSI neural-based pattern recognition systems, hearing aids, precision circuit design for instrumentation equipment, and VLSI transistor mismatch parameters characterization and, over the past 25 years, has been deeply involved in neuromorphic spiking circuits and systems, with strong emphasis on vision and exploiting nanoscale memristive devices for learning. He is the Co-Founder of two start-ups, Prophesee SA, Paris, France (www.prophesee.ai) and GrAI-Matter-Labs SAS, San Jose, CA, USA (www.graimatterlabs.ai), both on neuromorphic hardware.

Dr. Linares-Barranco is a Co-Editor-in-Chief of the *Frontiers in Neuromorphic Engineering* since 2021.