

Scalable Digital Neuromorphic Architecture for Large-Scale Biophysically Meaningful Neural Network With Multi-Compartment Neurons

Shuangming Yang, *Student Member, IEEE*, Bin Deng^{ID}, *Member, IEEE*, Jiang Wang^{ID}, Huiyan Li, Meili Lu^{ID}, Yanqiu Che^{ID}, *Member, IEEE*, Xile Wei^{ID}, and Kenneth A. Loparo, *Life Fellow, IEEE*

Abstract—Multicompartment emulation is an essential step to enhance the biological realism of neuromorphic systems and to further understand the computational power of neurons. In this paper, we present a hardware efficient, scalable, and real-time computing strategy for the implementation of large-scale biophysically meaningful neural networks with one million multi-compartment neurons (CMNs). The hardware platform uses four Altera Stratix III field-programmable gate arrays, and both the cellular and the network levels are considered, which provides an efficient implementation of a large-scale spiking neural network with biophysically plausible dynamics. At the cellular level, a cost-efficient multi-CMN model is presented, which can reproduce the detailed neuronal dynamics with representative neuronal morphology. A set of efficient neuromorphic techniques for single-CMN implementation are presented with all the hardware cost of memory and multiplier resources removed and with hardware performance of computational speed enhanced by 56.59% in comparison with the classical digital implementation method. At the network level, a scalable network-on-chip (NoC) architecture is proposed with a novel routing algorithm to enhance the NoC performance including throughput and computational latency, leading to higher computational efficiency and capability in comparison with state-of-the-art projects. The experimental results demonstrate that the proposed work can provide an efficient model and architecture for large-scale

Manuscript received March 6, 2018; revised September 7, 2018 and December 24, 2018; accepted February 13, 2019. Date of publication March 18, 2019; date of current version January 3, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61471265, Grant 61501330, and Grant 61771330 and in part by the Tianjin Municipal Special Program of Talents Development for Excellent Youth Scholars under Grant TJTZJH-QNBJRC-2-21. (*Corresponding author: Xile Wei.*)

S. Yang, B. Deng, J. Wang, and X. Wei are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: yangshuangming@tju.edu.cn; dengbin@tju.edu.cn; jiangwang@tju.edu.cn; xilewei@tju.edu.cn).

H. Li is with the School of Automation and Electrical Engineering, Tianjin University of Technology and Education, Tianjin 300222, China (e-mail: lhy2740@126.com).

M. Lu is with the School of Information Technology Engineering, Tianjin University of Technology and Education, Tianjin 300222, China (e-mail: meililu@tute.edu.cn).

Y. Che is with the Department of Neurosurgery, Penn State College of Medicine, Hershey, PA 17033 USA (e-mail: yche@pennstatehealth.psu.edu).

K. A. Loparo is with the Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH 44106 USA (e-mail: kenneth.loparo@case.edu).

This article has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the author.

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNNLS.2019.2899936

biologically meaningful networks, while the hardware synthesis results demonstrate low area utilization and high computational speed that supports the scalability of the approach.

Index Terms—Compartmental neuron (CMN) model, field-programmable gate array (FPGA), network on chip (NoC), neuromorphic engineering, spiking neural network (SNN).

I. INTRODUCTION

DEEP understanding of the structural and dynamic complexity of the human brain depends on the development of large-scale, anatomically detailed models of the brain network, which can realistically reveal the mechanisms on how neuronal and synaptic processes interact to generate the collective behaviors of the brain [1]–[3]. A core problem in neuroscience is to improve the understanding of the dynamics and coding within neuronal morphology and their role in large neural networks, i.e., the mechanisms that neuronal morphological properties contribute to information coding and behavior [4]–[6]. The challenge is to simultaneously incorporate morphological details that are responsible for complex neural dynamics with the high-level scale of the network that maintains global activities of the brain in an efficient manner.

Neurons are spatially distributed and heterogeneous units, and the dynamics and computations of single neurons play vital roles within large neural networks [4]. Although the “integrate-and-fire (IF)” model can provide a simple mechanistic model of basic neural operations [7], this type of point neuron model is far from accurate in representing neural behaviors. Conductance-based point neuron models, such as the classical Hodgkin–Huxley model, also ignore the spatial and morphological structure of a spiking neuron [11]–[13]. Previous studies have shown that the geometrical structure of a neuron, especially neural dendritic computation, is significant for information processing [14]–[17]. The neural dendrites have been shown to be essential for computations at molecular, cellular, and even behavioral levels [20]–[24]. Thus, quantitative spiking neuron models have used dendritic computation to reproduce more biologically meaningful dynamics [8]–[10]. Although detailed compartmental models can reproduce the biological dynamics of single neurons quite accurately, their high dimensionality and intricate architecture rule out mathematical understanding of the emergent properties. Besides, they are computationally

expensive and, thus, are not well suited for large-scale implementation of spiking neural networks (SNNs) [4]. Reduced compartmental models with only a few dendritic compartments can overcome these challenges and are sufficient to investigate soma-to-dendritic interactions that are related to real spiking and bursting activities [18], [19]. However, in comparison with simple point neuron models, there are four major challenges in large-scale neuromorphic computations using compartmental neuron (CMN) models: how to more efficiently calculate the dendritic part with high nonlinearity; how to deal with the conflict between hardware resource limitation and the increasing complexity of CMN models; how to more efficiently reproduce high nonlinear ionic channel dynamics in the compartmental model; and how to implement large-scale biologically meaningful networks with more enhanced hardware performance.

The last decade has seen great progress in furthering our understanding of brain dynamics, including several studies on large-scale modeling and implementation of SNNs [3], [25], [26]. In general, the large-scale implementations are aimed at several aspects, including the formulation of hypotheses regarding the function of neural systems, validation of self-consistency in the description of phenomenon or function, neural computation instead of traditional computing structures, and biologically inspired engineering applications [27], [28], [30], [31]. The BrainScaleS, TrueNorth, and SpiNNaker projects used point neuron models that cannot reproduce the detailed dendritic dynamics of neurons. The Neurogrid project used CMN models for the reproduction of detailed biological behaviors [29], but an IF-type neuron model was used as the soma, which cannot regenerate the cellular biological characteristics across the membrane. Although the blue brain project used the CMN model for large-scale network simulation, it cannot simulate the neural network in real time with low computational efficiency due to the von Neumann calculation architecture [25]. Therefore, novel designs and techniques should be explored to solve the four previously mentioned challenges in neuromorphic computation and to achieve a balance between scalability and computational speed. Existing neuromorphic projects have not provided a solution to these challenges.

This paper focuses on the real-time scalable implementation of large-scale biologically meaningful neural networks using the compartmental model, considering neuronal morphological properties, with the high programming capability to provide greater model flexibility in comparison with analog designs of current advanced techniques. Both the detailed neuronal dynamics and the large-scale network properties are considered by using a set of digital implementation techniques. The remainder of this paper is organized as follows. Section II describes the general overview of the proposed system and the efficient CMN model. In Section III, the hardware design for the digital neuromorphic system is presented. The experimental results are presented in Section IV, which includes the implementation results of the digital neuromorphic system, validation of the model dynamics, and hardware performance analysis. Section V makes a comparison with the state-of-the-art techniques and explores the impact of neuronal

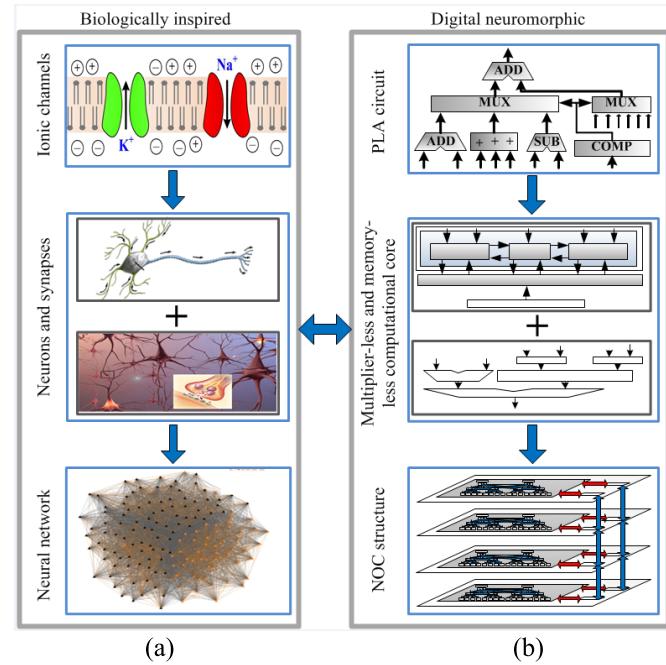


Fig. 1. General overview of the digital neuromorphic system. (a) Biologically inspired point of view. (b) Digital neuromorphic point of view.

morphology on the dynamics of neuromorphic SNNs. Finally, this paper is concluded with future works in Section VI.

II. GENERAL OVERVIEW OF THE PROPOSED DIGITAL NEUROMORPHIC SYSTEM

A. General Overview of the System

The proposed digital neuromorphic system is biologically meaningful, which uses a multicompartment conductance-based neuron model to reproduce biologically plausible dynamical characteristics. It is a high-performance simulator aiming at real-time simulation of large-scale biologically meaningful neural networks. Inspired by neuroscience, the key architecture abstraction is a multicore digital neuromorphic system that is efficient, scalable, and flexible. As shown in Fig. 1, the focus is on large-scale SNN with compartmental dynamics and ionic properties of a single cell included. At the ionic level, dynamics of ionic currents responsible for spike generation are considered. At the morphology level, a multi-CMN model that recalibrates dendritic channel densities to get optimal spike encoding of external stimulation is used. In addition, a key component is modeling the plasticity, which enables neurons to learn over time through changes in synaptic sensitivity and through generation of new synaptic connections.

From the computational perspective, the basic building block is a digital circuit for realizing the nonlinear functions of a multi-CMN model. The general method to implement the nonlinear functions in digital circuits is to use a lookup table (LUT). However, this requires significant hardware resources, for example, an LUT with a 10-bit address and 30-bit data requires up to 30 720 bits. Thus, the cost of LUTs in the implementation needs to be reduced. In addition,

TABLE I
EQUATIONS IN THE TWO-CMN MODEL

Function	Expressions	Function	Expressions
$I_{d\text{Leak}}$	$g_L \cdot (V_d - V_L)$	I_{Na}	$g_{Na} \cdot m_\infty(V_s) \cdot (V_s - V_{Na})$
I_{KDR}	$g_{KDR} \cdot w \cdot (V_s - V_K)$	I_{DS}^{in}	$g_c \cdot (V_d^{\text{in}} - V_s^{\text{in}})$
$I_{s\text{Leak}}$	$g_L \cdot (V_s - V_L)$	I_{Ca}	$g_{Ca} \cdot s^2 \cdot (V_d - V_{Ca})$
$m_\infty(V_s)$	$0.5 \cdot [1 + \tanh(\frac{V_s - \beta_m}{\gamma_m})]$	$w_\infty(V_s)$	$0.5 \cdot [1 + \tanh(\frac{V_s - \beta_w}{\gamma_w})]$
$\tau_w(V_s)$	$1 / \cosh(\frac{V_s - \beta_w}{2\gamma_w})$		

the constrained number of available fast multipliers on a field-programmable gate array (FPGA) device limits the design and realization of a complex neuron and large-scale SNN. Multipliers introduce delay and require power consumption and hardware cost. In this paper, multiplier-less and the memory-less techniques are proposed to implement the CMNs and synapses. The dedicated hardware architecture for the multi-CMN model is also presented. An improved hierarchical network-on-chip (NoC) architecture is designed for the scalable and efficient implementation and uses the address-event representation (AER) method for data communication with a novel routing algorithm. A multichip scheme is used to realize the 3-D NoC interconnection structure with the hierarchical NoC topology on each chip.

B. Multicompartment Conductance-Based Neuron Models

In the proposed neuromorphic system, we use a two-CMN model comprised of a somatic compartment and a dendrite compartment. Mathematically, the model consists of three variables as

$$\begin{cases} C_m \frac{dV_S}{dt} = -I_{s\text{Leak}} - I_{Na} - I_{KDR} + \frac{I_{DS}^{\text{in}}}{p} + I_{app} \\ C_m \frac{dV_d}{dt} = -I_{d\text{Leak}} - I_{Ca} - I_{syn} - \frac{I_{DS}^{\text{in}}}{1-p} + \frac{I_d}{1-p} \\ \frac{dw}{dt} = \phi \frac{w_\infty - w}{\tau_w(V_S)} \end{cases} \quad (1)$$

where V_S and V_d are the deviations of the somatic and dendritic membrane potentials and I_d is the stimulation currents applied on the dendrite divided by the total neuronal membrane area. The applied direct stimulation current to the soma is I_{app} . The current I_{DS}^{in} is the internal current between soma and dendrite. The parameter p is the proportion of neuronal area taken up by the soma. Other relationships are given in Table I. The parameter values based on Prescott *et al.* [32] are presented in Table S1 in the supplementary material.

C. Synapse Model With Spike-Timing-Dependent Plasticity Learning Rule

A series of empirical investigations has shown small-world topologies in many structural and functional networks of the brain [33], [34]. Thus, we introduce the small-world network topology to implement the large-scale biologically meaningful neural network. The synaptic current delivered from

neuron j to a synaptically coupled neuron i at time $t \geq t_j$ is governed by

$$I_{ij}^{\text{syn}} = \sum_{j=1(j \neq i)}^N g_{ij} C_{ij} s_j(t)(V_i - V_{\text{syn}}) \quad (2)$$

where C_{ij} is a connectivity matrix and g_{ij} represents the coupling strength. If neuron j couples to neuron i , then $C_{ij} = 1$ and $C_{ji} = 1$, otherwise $C_{ij} = C_{ji} = 0$ and $C_{ii} = 0$. The value of C_{ij} is determined by the network connection probability p_{con} based on the Newman–Watts procedure [52]. The synaptic current is based on the value of reversal potential V_{syn} of the synapse. The synaptic variables s_j are governed by

$$\dot{s}_j(t) = \alpha(V_j(t))(1 - s_j(t)) - \beta s_j(t) \quad (3)$$

$$\alpha(V_j(t)) = \alpha_0 / (1 + e^{-V_j(t)/V_{\text{thr}}}) \quad (4)$$

where the synaptic recovery function $\alpha(V_j)$ is described by the Heaviside function. The threshold V_{thr} is set to 0.05, and the parameter values α_0 and β are set to 2 and 1, respectively. The synaptic conductance g_{ij} from the j th neuron to the i th one can be changed using the doublet spike-timing-dependent plasticity (D-STDP) function $F(\Delta t)$, which is defined as

$$g_{ij} = g_{ij} + \Delta g_{ij} \quad (5)$$

$$\Delta g_{ij} = g_{ij} F(\Delta t) \quad (6)$$

$$F(\Delta t) = \begin{cases} A_+ \exp(-|\Delta t|/\tau_+) & \text{if } \Delta t > 0 \\ -A_- \exp(-|\Delta t|/\tau_-) & \text{if } \Delta t < 0 \\ 0 & \text{if } \Delta t = 0 \end{cases} \quad (7)$$

where t_i (or t_j) represents the spiking time of the i th (or j th neuron). The variable $\Delta t = t_i - t_j$ and adjusting the rates A_+ and A_- constrain the maximum value of synaptic modification. The values $\tau_+ = \tau_- = 20$ ms represent the temporal window over which strengthening and weakening phenomenon occurs.

Previous studies show that the D-STDP learning rule cannot reproduce the experimental outcomes considering higher order spike patterns and cannot reproduce the observed weight based on repetition frequency of spikes pairs [61], [62]. In order to make the implemented digital neuromorphic system more biologically meaningful, the triplet STDP (T-STDP) algorithm is considered in this paper, which is described as [54]

$$F(\Delta t) = \begin{cases} e^{(-\Delta t_1)/\tau_+} (A_2^+ + A_3^+ e^{(-\Delta t_2)/\tau_y}), & \text{if } t = t_{\text{post}} \\ -e^{(-\Delta t_1)/\tau_-} (A_2^- + A_3^- e^{(-\Delta t_2)/\tau_x}), & \text{if } t = t_{\text{pre}} \end{cases} \quad (8)$$

where A_2^+ and A_2^- represent the amplitude of the weight change when a pre–post or post–pre pair occurs, respectively. A_3^+ and A_3^- represent the amplitude of the triplet term for potentiation and depression, respectively. Time differences are described as

$$\begin{cases} \Delta t_1 = t_{\text{post}}(n) - t_{\text{pre}}(n) \\ \Delta t_2 = t_{\text{post}}(n) - t_{\text{post}}(n-1) \\ \Delta t_3 = t_{\text{pre}}(n) - t_{\text{pre}}(n-1) \end{cases} \quad (9)$$

where τ_+ , τ_- , τ_x , and τ_y are time constants for the spikes pairs. For hippocampal cortex data set, (9) can be simplified to

$$\Delta w(t) = \begin{cases} e^{\left(\frac{-\Delta t_1}{\tau_+}\right)} \left(A_2^+ + A_3^+ e^{\left(\frac{-\Delta t_2}{\tau_y}\right)} \right), & \text{if } t = t_{\text{post}} \\ e^{\left(\frac{-\Delta t_1}{\tau_-}\right)} \left(A_2^- \right), & \text{if } t = t_{\text{pre}} \end{cases} \quad (10)$$

which is called Hippocampal triplet spike-timing-dependent plasticity (HT-STDP) in this paper.

D. Modified CMN Neuron Model

The conventional hardware approach to deal with the problem of high nonlinearity in CMN models is to use the LUT technique [35], [36]. But the LUT-based design will require significant memory resources on chip. Thus, a novel method is needed to deal with the complex calculations required for CMN models as well as to reduce the cost for resources. A cost-efficient multi-CMN model is proposed, which is modified from the original CMN model as follows:

$$\begin{cases} f_{\text{Na}} = g_{\text{Na}} \cdot m_\infty(V_S) \cdot (V_S - V_{\text{Na}}) \\ f_{\text{winf}} = 0.5 \cdot [1 + \tanh((V_S - \beta_m)/\gamma_m)] \\ f_{\text{taow}} = \cosh((V_S - \beta_w)/(2\gamma_w)). \end{cases} \quad (11)$$

The modified CMN model is given by

$$\begin{cases} C \frac{dV_S}{dt} = -I_{\text{sLeak}} - f_{\text{Na}}(V_S) - I_{\text{KDR}} + \frac{I_{\text{DS}}^{\text{in}}}{p} + \frac{I_s}{p} + I_{\text{app}} \\ C_m \frac{dV_d}{dt} = -I_{\text{dLeak}} - I_{\text{Ca}} - I_{\text{syn}} - \frac{I_{\text{DS}}^{\text{in}}}{1-p} + \frac{I_d}{1-p} \\ \frac{dw}{dt} = \phi(f_{\text{winf}}(V_S) - w) * f_{\text{taow}}(V_S). \end{cases} \quad (12)$$

The nonlinear functions in the original model are replaced by piecewise linear functions with m segments, which can be expressed as

$$f_i(V_S) = \begin{cases} k_1 V_S + b_1 & \text{when } V_S < \frac{b_2 - b_1}{k_1 - k_2} \\ k_2 V_S + b_2 & \text{when } \frac{b_2 - b_1}{k_1 - k_2} \leq V_S < \frac{b_3 - b_2}{k_2 - k_3} \\ \dots \\ k_n V_S + b_n & \text{when } V_S \geq \frac{b_n - b_{n-1}}{k_{n-1} - k_n} \end{cases} \quad (13)$$

where $i \in \{\text{Na}, \text{taow}, \text{winf}\}$. The parameters k_n and b_n are the slope and intercept of the piecewise linear functions, respectively ($n = 1, 2, \dots, m$). The value of k_n is represented as a power of 2 so that it can be implemented using shift and addition/subtraction operations. The criterion to evaluate the approximating error of the modified functions is defined as

$$\text{CFRE} = \frac{1}{m} \sqrt{\sum_{i=1}^m \frac{(f_{\text{ori}}(i) - f_{\text{mod}}(i))^2}{f_{\text{ori}}(i)^2}} \quad (14)$$

where m is the total number of sampling points. An exhaustive search algorithm is used in the approximation of the modified functions to obtain the coefficients values. If the value of the cost function is large, a segment will be added to the piecewise linear function until the accuracy can be guaranteed.

TABLE II
ERROR EVALUATION OF THE MODIFIED FUNCTIONS

PLA functions	CF_{RE}	CF_{NMAE}	CF_{MAE}
f_{Na}	0.0720	0.0012	0.705
f_{winf}	0.0039	0.000149	0.4722
f_{taow}	0.0184	0.0183	0.0101

The detailed parameter values of the modified piecewise linear approximation (PLA) functions are given in Table S2 in the supplementary material. In order to further quantify the relative error, the cost functions CF_{NMAE} and CF_{MAE} are defined as

$$\begin{cases} CF_{NMAE} = \frac{\frac{1}{\lambda} \sum_{i=1}^{\lambda} |\text{ERR}_{\text{ABS}}|}{(f_{\max} - f_{\min})} \\ CF_{MAE} = \frac{1}{M} \sum_{i=1}^M |e_i|. \end{cases} \quad (15)$$

In the first equation, f_{\max} is the maximum value of the modified function and f_{\min} is the minimum value. The absolute error $|e_i| = |f_{\text{ori}} - f_{\text{in}}|$ denotes the error between the predicted and the values. The parameter $\lambda = 2000$ is the number of sampling points to get the value of CF_{NMAE} . The detailed error analysis results of the three modified functions are given in Table II.

III. HARDWARE DESIGN FOR THE DIGITAL NEUROMORPHIC SYSTEM

A. Digital Implementation of the CMN Neuron Model

The top-level design of the multicompartment neuron processor (MNP) is shown in Fig. 2(a). The time-multiplexed technique is used in the proposed MNP to scale up the neuron number of the neuromorphic system. The Euler method is used in consideration of both high precision cost and low resource cost. The digital structure of the CMN model is shown in Fig. 2(b), which has three pipelines for the three variables. The number of stages in “ V_S ” pipeline is represented by V_{S_stage} . The digital architectures of the “ V_S ,” “ V_d ,” and “ w ” pipelines are depicted in Fig. 2(c)–(e), respectively. Digital implementation of the T-STDP algorithm is shown in Fig. 2(f). The implementation of the HT-STDP algorithm is based on the T-STDP and simpler than it. Detailed descriptions of the MNP are presented in Figs. S1 and S2 in the supplementary material.

B. Digital Architecture of the Large-Scale Neural Network Based on Improved Butterfly Fat Tree Topology

As shown in Fig. 3(a), a 3-D NoC architecture was proposed for the scalable implementation of large-scale SNNs with CMN models, which uses the improved butterfly fat tree (IBFT) NoC topology. The layered approach divides 3-D NoC into two parts: the horizontal IBFT layers and the vertical connectors. Each horizontal NoC layer is realized based on an FPGA, and the high speed terasic connectors connector on the DE3 development board is used as the vertical connector. The number of horizontal NoC layers is not restricted. First in first out (FIFOs) are used to buffer

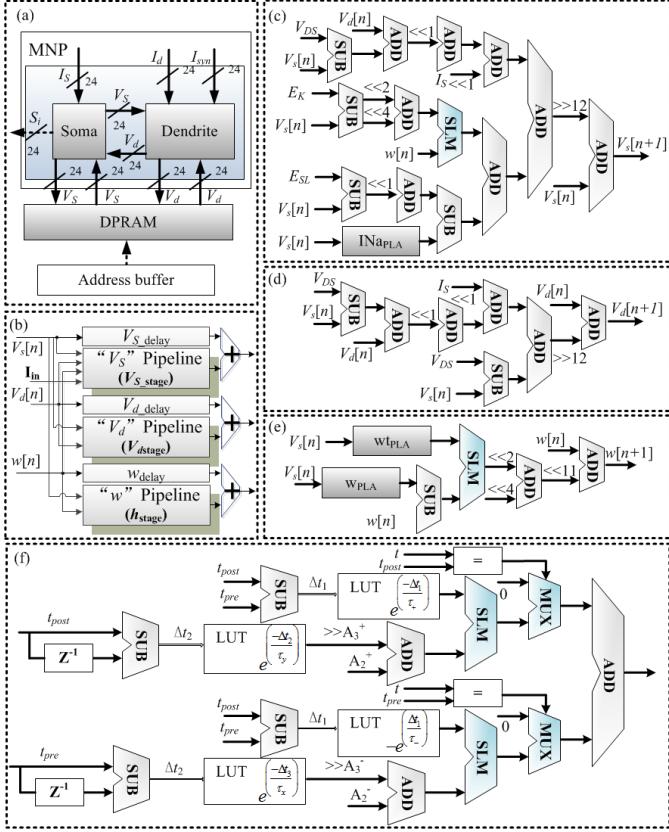


Fig. 2. Digital implementation of the proposed MNP. (a) Top-level design of the MNP. (b) Pipeline architecture of the MNP. (c) V_s pipeline in the MNP. (d) V_d pipeline of the MNP. (e) w pipeline of the MNP. (f) Digital implementation of the T-STDP algorithm.

interboard data. The transmitted data flow uses 26-bit packet, which contains 8-bit AER data, 3-bit chip address, 3-bit layer address, 6-bit node address, and 6-bit timestamp. The interboard and intraboard data communication use the same data packet. The AER communication protocol is used to represent synaptic connectivity in a dynamically reconfigurable manner by routing address events through synaptic routing tables in off-chip memory, which maps presynaptic source addresses to postsynaptic destination addresses with synaptic parameters. The synaptic routing tables are implemented using a 1-Gb DDR2 synchronous dynamic random-access memory (SDRAM) for each FPGA, with each SDRAM interfacing through a dedicated bus and independent memory controller. Fig. 3(b) depicts the floor planning of an IBFT topology in 3-D NoC structure. The available hardware determines the number of MNP to build. For the case with 64 MNPs, it requires three levels of routers.

C. Router Architecture and Algorithm for 3-D IBFT NoC Topology

The proposed router is composed of three modules, which are the buffer module, the arbiter module, and the crossbar module between the I/O ports. The buffer module stores the AER data, and the arbiter module determines the output directions and order of each data packet. Conventional router structure in butterfly fat tree (BFT) topology contains six I/O ports, and I/O ports are connected by the crossbar module

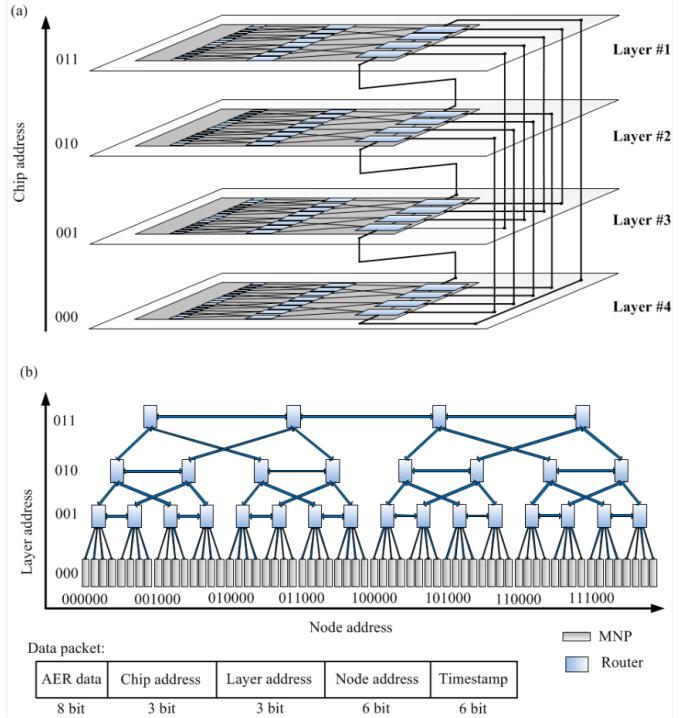


Fig. 3. Topology of the presented 3-D IBFT for large-scale neuromorphic realization. (a) Scalable connection structure for the proposed 3-D NoC system. (b) Digital architecture on each chip.

and the multiplexer. In the IBFT topology, the number of I/O ports in the top-layer routers is reduced, which can save the hardware resource.

The architecture of the second-layer routers is shown in Fig. 4(a), which contains six input ports and six output ports. Each port has six virtual channels, and each channel is used as a FIFO buffer. The selection of a virtual channel is determined by the arbiter module. The router algorithm of the IBFT architecture is a deterministic nonshortest path router algorithm. The routing path obtained from this algorithm is not guaranteed to be the shortest path. The node code of the router will be compared to the node code of the destination node. The pseudocode of the router for the IBFT topology is shown in Fig. 4(b), where NAD represents the node address of the destination, and the node code of the current router is encoded as $C(i, j)$. The address of the child router, neighborhood router, and parent router of the current router can be searched in the prestored LUT, which is presented in Table S3 in the supplementary material. In order to reduce the routing latency of the proposed system, the local load mode is preferred in the design, which means the probability of the information routing between the source and destination nodes will decrease when the distance of the two nodes is increasing.

IV. EXPERIMENTAL RESULTS

As mentioned above, the assessment of the performance of the proposed digital neuromorphic system has to consider three major factors: emulation performance, dynamical accuracy of the modified CMN model, and the hardware performance of the large-scale biologically meaningful neural network. In the following, both the timing and utilization analysis will

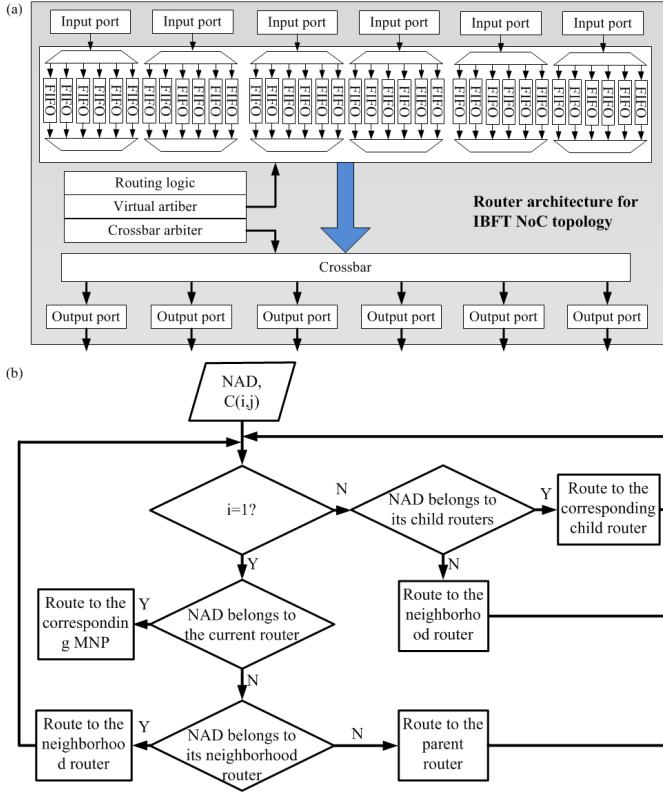


Fig. 4. Routing in the proposed IBFT architecture. (a) Router architecture for the IBFT NoC topology. (b) Pseudocode of the routing algorithm in the IBFT NoC network.

be considered. The dynamical characteristics of the proposed CMN model will also be investigated and compared to the original model. All the proposed results were verified after synthesis and implementation.

A. Hardware Experimental Results for Large-Scale Biophysically Meaningful Neural Network

The proposed real-time system for large-scale biologically meaningful neural network uses four Altera Stratix III EP3SL340 FPGA devices. The experimental results of the outputs of the proposed system are shown in Fig. 5. The results are the firing activities of the soma, dendrite, and evolved activities of the slow variable, which are randomly chosen in the large-scale biologically meaningful network. The experimental result shows that biological behaviors can be reproduced accurately in real time.

In order to conduct the validation of the proposed implementation, the bit-level fixed-point simulation was presented and compared with the software-based simulation results using MATLAB. For the bit-level simulation, all hardware components were designed and realized based on the very-high-speed integrated circuit hardware description language. The comparison results in Fig. 6 show that the proposed implementation obtains high computational precision. The main cause of the deviation is the fixed-point calculation of the FPGA implementation. However, by increasing the number of bits in the proposed system or adjusting the parameter values of the large-scale biologically meaningful network, the small difference can be further reduced.

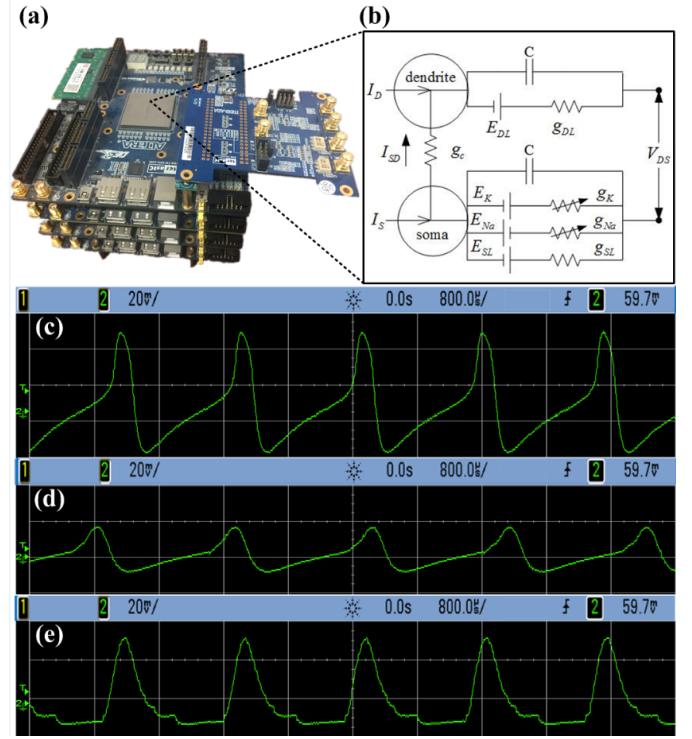


Fig. 5. Experimental results on oscilloscope device. The time division is 2 ms, and the amplitude division is 100 mV. (a) Proposed digital neuromorphic system implemented by four DE3 340 development boards. (b) Equivalent circuit of the multi-CMN model. (c) Firing evolution of V_S . (d) Firing evolution of V_d . (e) Firing evolution of w .

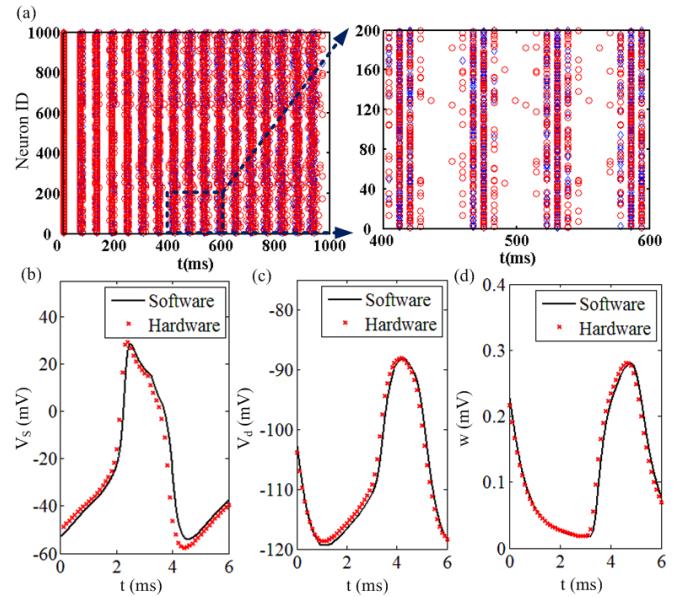


Fig. 6. Comparison between *in silico* and hardware simulations. (a) Network dynamics characterization: comparison between *in silico* and hardware simulations. (b) Comparison of variable V_S between the proposed architecture and MATLAB. (c) Comparison of variable V_d . (d) Comparison of variable w .

The mean firing rate (MFR) is investigated using the software and hardware models as plotted in Fig. 7. Two neuronal morphology parameter values are selected, which are $p = 0.45$ and $p = 0.8$, respectively. Fig. 7(a1) and (b1) shows that the MFR will increase as the morphology parameter increases. In the case where $p = 0.45$, MFRs

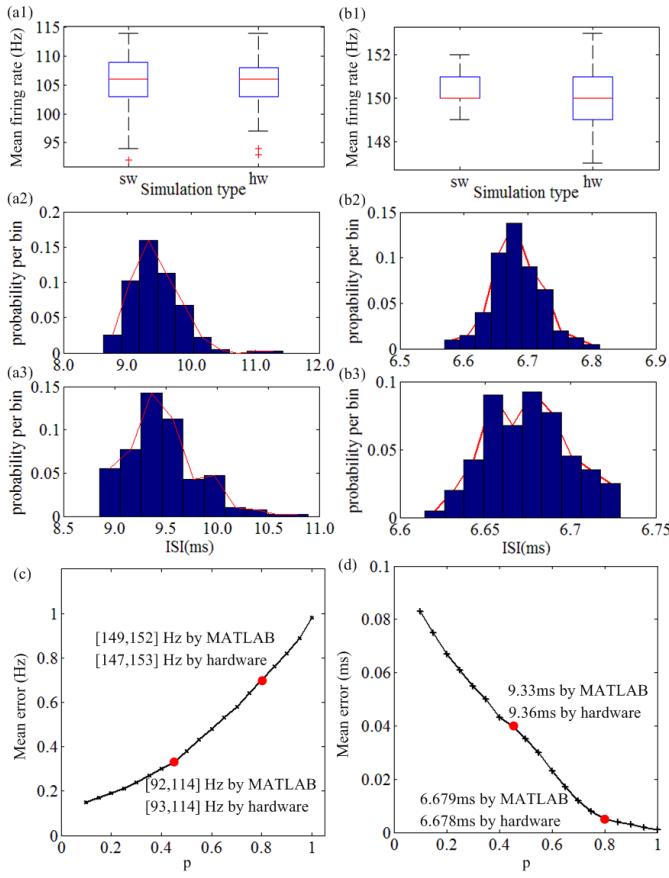


Fig. 7. Network dynamics characterization: comparison between software and hardware simulations. (a1) Box plot of MFR that is evaluated over the whole neurons of the network with the morphology parameter $p = 0.45$. (a2) ISI distributions of the software-based simulation results with the morphology parameter $p = 0.45$. (a3) ISI distributions of the hardware-based simulation results with the morphology parameter $p = 0.45$. (b1) Box plot of MFR that is evaluated over the whole neurons of the network with the morphology parameter $p = 0.8$. (b2) ISI distributions of the software-based simulation results with the morphology parameter $p = 0.8$. (b3) ISI distributions of the hardware-based simulation results with the morphology parameter $p = 0.8$. (c) Mean error of the network MFRs between software and hardware. (d) Mean error of the ISI between software and hardware.

are [92, 114] Hz in the MATLAB simulation of the model and [93, 114] Hz in the hardware implementation. For $p = 0.8$, MFRs are [149, 152] Hz in the software and [147, 153] Hz in the hardware, respectively. The interspike interval (ISI) distributions were also evaluated. Fig. 7(a2) and (b2) are relative to the software model with $p = 0.45$, and Fig. 7(a3) and (b3) are relative to the hardware model with $p = 0.8$. Good consistency is obtained between the two implementations. When $p = 0.45$, the temporal positions of the peaks of the curves are 9.33 ms for software and 9.36 ms for hardware, respectively. When $p = 0.8$, the software and hardware simulations have the temporal position of the peaks of the curves at 6.679 and 6.678 ms, respectively.

The computational error between the MATLAB and the proposed hardware simulations is proposed in Fig. 7(c) and (d). Fig. 7(c) shows the error of the MFR with the morphological parameter increasing across its range of values. In the case when $p = 0.45$, MFRs are [92, 114] Hz by MATLAB and [93, 114] Hz by hardware. For $p = 0.8$, MFRs are

[149, 152] Hz by software and [147, 153] Hz by hardware, respectively. It shows that the mean error is augmented as the morphological parameter increases. This is because the firing rate is affected by the parameter. Fig. 7(d) shows the mean error of the ISI between software and hardware. When $p = 0.45$, the peak values of ISI are 9.33 ms for software and 9.36 ms for hardware, respectively. When $p = 0.8$, the software and hardware simulations have the peak values of ISI at 6.679 and 6.678 ms, respectively. As the morphological parameter increases, the mean error of ISI will be decreased.

B. Dynamical Behaviors of the CMN Model

In order to investigate the dynamical characteristics of the CMN model, a bifurcation analysis is presented. It reveals that there exists a saddle-node on invariant circle (SNIC) bifurcation, which is composed of two trajectories called heteroclinic trajectories connecting the saddle and the node. There is also a supercritical Andronov–Hopf bifurcation, which shows a loss of stability. As shown in Fig. 8(a1) and (a2), the bifurcation diagram of the CMN neuron model is consistent with the original model. In addition, we compare the firing rates between simulations of the original and CMN models with an externally applied current. As shown in Fig. 8(b1) and (b2), the original neuron model will change from quiescent state to the repetitive spiking state with the current $I_{app} = 67.0$ pA, and the CMN neuron model will change its state with $I_{app} = 68.0$ pA. It shows that the supercritical Andronov–Hopf bifurcation point of the modified model shifts right compared to the original model. As the applied current increases, the biological activities of the CMN model turn into a resting state slower than the original model, with a longer period of time for subthreshold membrane potential oscillation. It means a period of resting state is converted to subthreshold membrane potential oscillation right after the supercritical Andronov–Hopf bifurcation point by using the proposed CMN model. However, it will not influence the firing state of a single neuron qualitatively and the biological dynamics in neural networks. Besides, the value of the applied current in this paper will not be close to the supercritical Andronov–Hopf bifurcation point. The value range of the applied current is around and slightly larger than the SNIC bifurcation point, which is consistent with the biological activities in the human brain. Thus, the deviation of bifurcation points can be accepted. The deviation of the bifurcation point between two models is caused by the PLA functions used in the CMN model, especially the approximation in the computation of sodium ionic current I_{Na} . The increment of the number of segments of the PLA function f_{Na} makes the modified function more close to the original one, which could induce more closed bifurcation points in comparison with the original model. However, it will cost more hardware resource with the precision of PLA functions increasing. It can be found that under the external applied current stimulus, the CMN can exhibit either quiescent state or repetitive spiking state with a variation of the stimulation current. The firing rates of these two neuron models under different simulations are also consistent with each other. Fig. 8(c1), (d1), and (e1) shows the firing patterns under resting state, regular spiking state, and fast spiking state,

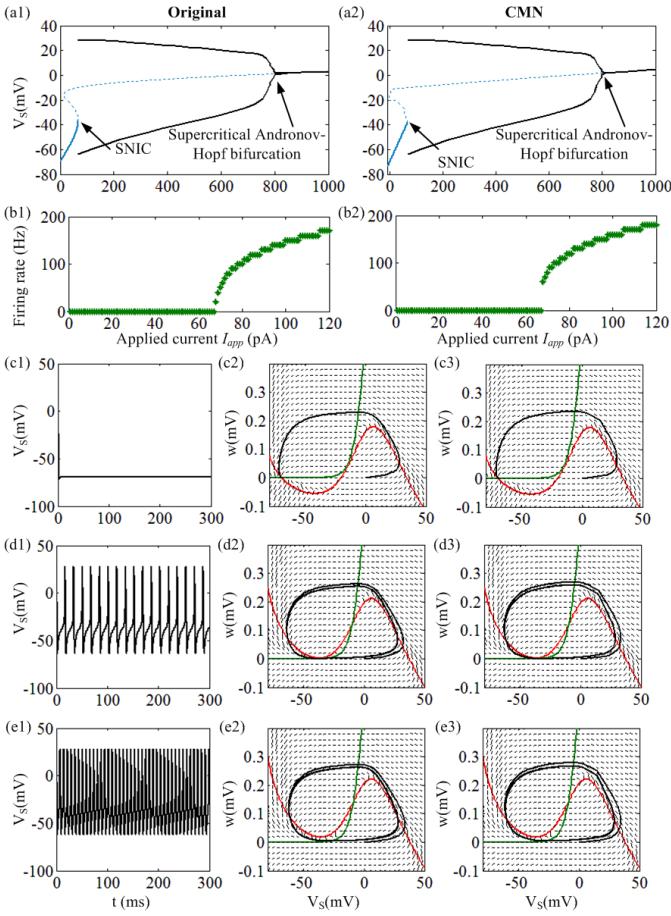


Fig. 8. Dynamical investigation of the CMN model. (a1) Bifurcation diagram of the original model. (a2) Bifurcation diagram of the CMN model. (b1) Average spiking frequency and applied current of the original model. (b2) Average spiking frequency and applied current of the CMN model. (c1) Neural firing under the resting state. (c2) Phase portraits of the original model under the resting state. (c3) Phase portraits of the CMN model under the resting state. (d1) Neural firing under the regular spiking state. (d2) Phase portraits of the original model under the regular spiking state. (d3) Phase portraits of the CMN model under the regular spiking state. (e1) Neural firing under the fast spiking state. (e2) Phase portraits of the original model under the fast spiking state. (e3) Phase portraits of the CMN model under the fast spiking state.

respectively. Fig. 8(c2) and (c3) shows the phase portraits of the original and the CMN models, revealing that no cross point in the phase plane will lead to a quiescent state. Fig. 8(d2) and (d3) shows that a stable limit cycle occurs with the state of tonic spiking. As shown in Fig. 8(e2) and (e3), the firing rate of the neuron model will be increased with the enhancement of the applied stimulation currents. Furthermore, the CMN model and the original model maintain a high consistency of their phase portraits with different levels of applied currents.

C. Analysis of the Hardware Performance

In order to evaluate the hardware performance of the CMN model, a set of evaluation criteria are presented, which considers both the resource cost and the operating frequency. The delay CF_D is described as

$$CF_D = 1/f = f_{\text{Available}}/f_{\text{Used}} \quad (16)$$

TABLE III
HARDWARE RESOURCE COST OF THE PROPOSED MODELS

Resource	Available	Original	CMN model
Combinational ALUTs	270400	7764 (3%)	2970 (1%)
Memory ALUTs	135200	1888 (1%)	0 (0%)
Dedicated logic registers	270400	471 (<1%)	94 (<1%)
Block memory bits	16662528	0 (0%)	0 (0%)
DSP block 18-bit elements	576	0 (0%)	0 (0%)
CF_D	--	1.8406	1.3620
Resource	D-STDP	T-STDP	HT-STDP
Combinational ALUTs	54	360	241
Dedicated logic registers	0	88	69

where f is the working frequency of the neuron model. It should also be noted that the number of neurons is limited to the available number of embedded multipliers for an FPGA device, which means the network scale of the large-scale biologically meaningful network will be limited by the EM9Es. Thus, the digital multiplier-less implementation of the CMN model makes significant contributions to the implementation of the large-scale biologically meaningful neural network. The limitation of the proposed method lies in the computing deviation of the neuronal biological activities due to the PLA, which requires an extra investigation of the neural dynamics to guarantee the accurate reproduction of the neural biological behaviors. In order to make a fairer comparison between the original and CMN models, the DSP blocks are replaced by combinational ALUTs and block memories are replaced by memory ALUTs when implementing the original model. As shown in Table III, the proposed CMN model can reduce the hardware resource cost of combinational ALUTs, memory ALUTs, and dedicated logic registers, and increase the working speed. There is a 35.14% speedup improvement compared to the LUT-based implementation of the original model, and this is essential because increasing the working frequency means that more virtual neurons can be implemented on a fixed number of physical neurons with the same sampling time. In addition, the hardware resource costs of D-STDP, T-STDP, and Q-STDP are shown in Table III.

In order to validate enhancement owing to the proposed NoC method, we conducted experiments with a straightforward method based on flat hierarchy presented by Vogelstein *et al.* [56] and Serrano-Gotarredona *et al.* [57]. A flat system requires a unique global address for each source node, which will severely reduce the total number of neurons the system can realize. In the hierarchical IBFT-based system, only the next destination for the source node is needed. For the experiments of measured latency, the Poisson spike generator is used to route 40 000 neural events across the proposed architecture at 40M synaptic outputs per second (SynOPS). As shown in Fig. 9(a), for the flat topology, latency is measured from data on each of the 64 MNPs with local 2000 fan-outs. The same network is partitioned into eight parts based on the four-level IBFT topology. A shorter tail and narrower distribution is observed in the case of IBFT topology. For the IBFT hierarchy as shown in Fig. 9(b), network latency was measured from data across eight groups with 250 fan-outs. The IBFT topology obtains a significant reduction in maximum latency. Measured event latency as a function of

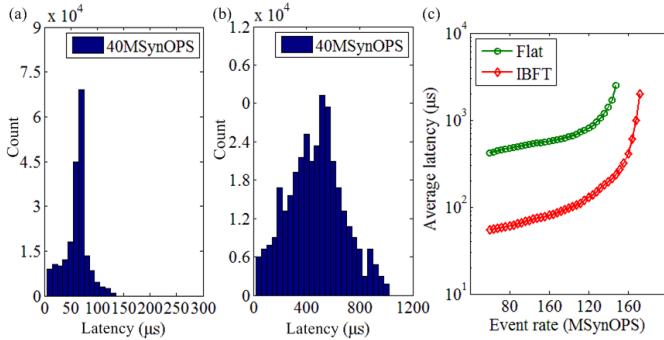


Fig. 9. Comparison between the IBFT and flat structures. (a) Measured latency of IBFT topology. (b) Measured latency of flat topology. (c) Average event latency measured as a function of synaptic event rate for flat and IBFT of the proposed SNN.

synaptic event rate for both flat and IBFT hierarchies is shown in Fig. 9(c). It shows that the average latency of the flat topology is eight times larger than IBFT topology, and the proposed topology can support an eightfold greater overall event throughput across the network. In addition, the extension from flat topology to IBFT topology is critical in scaling up neuromorphic systems toward levels of synaptic connectivity approaching that of the human central nervous system.

From the system point of view, the NoC topology determines a few essential points including the coding method of the router and MNP and the router and mapping algorithms, which further influences the latency, resource cost, and power consumption of the chip. In comparison with computer networks, NoC has a higher requirement for the hardware resource cost because it influences area and power consumption directly. Thus, the increment of resource utilization is required in NoC design. A cost function of the buffer utilization rate is defined as follows:

$$CF_{BUR} = \frac{\sum_{i=1}^{tot} [\sum_{j=1}^{N_R} N_{BSS}^{i,j} + \sum_{j=1}^{N_{IP}} N_{BSN}^{i,j}]}{NVC \cdot NPPB \cdot (NPL + N_{IP}) \cdot t_{tot}} \quad (17)$$

where NVC and $NPPB$ describe the number of virtual channels and the number of packets per buffer, respectively, and NPL represents the total number of physical links. N_{IP} is the total number of IP in the NoC structure. t_{tot} describes the total time from the first packet to the last, N_R represents the total number of routers in the NoC architecture, and $N_{BSS}^{i,j}$ and $N_{BSN}^{i,j}$ represent the buffer slots used in the switch and buffer slots used in a node, respectively. Fig. 10(a) shows the comparison of the buffer utilization between the BFT and the proposed IBFT architectures. When the load rate is low, the buffer utilization of these two architectures is the same. With the increase in the load rate, the buffer utilization of the IBFT architecture becomes higher than the BFT architecture. It demonstrates that the IBFT structure can obtain higher buffer utilization than the conventional BFT one.

In order to illustrate the computational efficiency and scalability of the proposed system, we compare the proposed work with the CPU-based version and our previous study using the time-multiplexed technique [37]. A cost function for the computational efficiency can be described as

$$Q = t_{bio}/t_{exp} \quad (18)$$

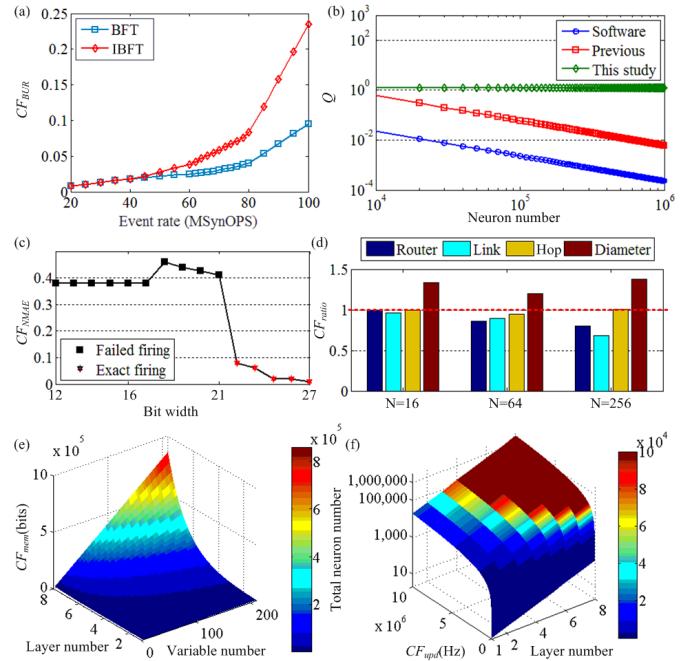


Fig. 10. Evaluation of the hardware performance. (a) Comparison of the buffer utilization between BFT and the proposed IBFT. (b) Comparison of the computational efficiency with other alternative approaches. (c) Precision analysis of the proposed system. (d) Evaluation of the NoC performance of the proposed architecture. (e) Analysis of the memory cost on the proposed system. (f) Analysis of the total implemented neuron number on the proposed system.

where t_{exp} and t_{bio} are the experimental computational time on the simulation system and the biological activity time, respectively. The CPU-based version ran on an Intel Core2 2.4-GHz CPU. As shown in Fig. 10(b), the proposed work achieves higher computational speed and higher scalability compared to the other hardware versions. When the neuron number is increased to 10^6 , the computational efficiency Q is 204.35 times larger than that in the previous study [37], and 5351 times larger than the software-based simulation. Although the network scale increases significantly, the computational efficiency of the proposed digital neuromorphic system will not be influenced and real-time simulation at a high level is maintained.

A fixed-point representation consists of a number of integers and fractional bits. In order to present the precision analysis, the integer part of the data representation is fixed according to the data value, and the fractional part is changed to adapt itself to the dynamical behaviors. For a real value x , the bit number of the integer part a is defined as

$$a = \log_2(\text{ceil}|x| + 1) \quad (19)$$

where $\text{ceil}(x)$ is the roundup value of x . We added an extra bit (+1) for signed operations to avoid overflow for all of the variables and parameters in the model.

The fractional part of a fixed-point representation is regarded as the precision of the system. In Fig. 10(c), we show a precision analysis between the results obtained in software and hardware. The multiprecision analysis guarantees the exact reproduction of the neural dynamics. The multiprecision

analysis is obtained from the simulation at different fixed-point representations, and the reproduced membrane potentials are compared in software and hardware by using the criterion CF_{NMAE} in (15). When the word length is more than 21 bits, exact firing dynamics can be obtained. In this paper, in order to obtain accurate biological behaviors, the bit numbers of the integer part for variables and parameters are determined based on (19), and the bit width of the fractional part is 24 bits according to the precision analysis shown in Fig. 10(c). Although higher bit width can enable more preferable computational precision, the further increment of the bit width will induce the higher hardware resource cost, which will limit the network scale and increase the power consumption on a single chip. We are supposed to find the minimum bit width with high computational precision. Thus, there is a tradeoff between the computational precision and hardware resource cost.

As shown in Fig. 10(d), in order to further evaluate and compare the NoC performance of the presented IBFT architecture with the original BFT, we use CF_{ratio} , which is the performance of IBFT-to-BFT ratio. The numbers of routers, links, average hop, and diameter are considered. The reduction of these factors means a decrease in the complexity of the NoC architecture. In comparison with the original BFT architecture, the performance of IBFT is enhanced considering the router, link, and average hop, without reduction of the NoC diameter.

On-chip memory is a critical factor for the large-scale implementation of SNNs. In order to evaluate the network scale on chip, we further investigate the memory cost CF_{mem} in the IBFT architecture. The relationship of CF_{mem} , the layer number of IBFT, and a variable number of the CMN model is explored as shown in Fig. 10(e). The upper bound of the variable number is set to 209, which is based on a detailed 19-compartment cable model proposed by Traub *et al.* [10]. The maximum layer number is considered to be 8 because a large layer number is unavailable for the implementation of the large-scale biologically meaningful network. In fact, there is a conflict between the memory resource cost, variable number in the neuron model, and the layer number of NoC.

The relationship between total available neuron number, network updating speed CF_{upd} , and layer number is investigated in Fig. 10(f), which shows the possibility to realize a large-scale biologically meaningful network with the proper digital neuromorphic design. The layer number determines the parallelism of the neuromorphic system. The updating speed CF_{upd} is based on the multiplexed time of the MNPs and the operating frequency of the digital network. The total neuron number computed on the digital system increases with the increment of both layer number and updating speed.

When the local communication burden becomes higher, communication bottleneck would occur. Thus, it is quite necessary to deal with the bottleneck problems caused by high local communication burden effectively. The comparison of the network throughput between original BFT and presented IBFT topologies under the local burden condition is shown in Fig. 11(a), which reveals the system performance of bandwidth. When the event rate of the network is less than 100MSynOPS, the throughputs of both architectures are consistent with each other. When the event rate is larger than

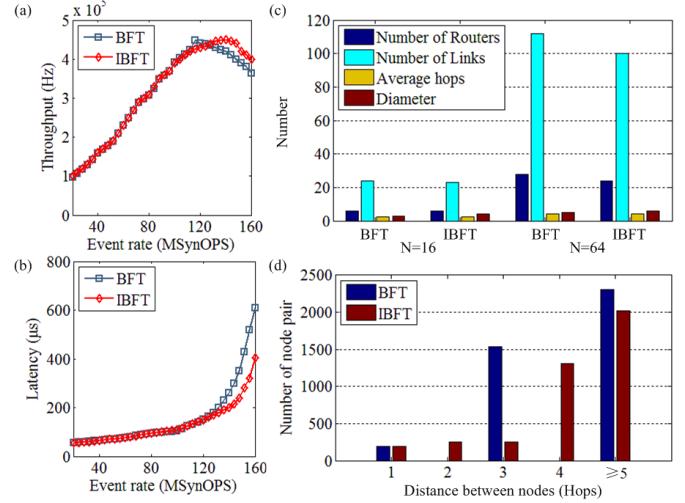


Fig. 11. Comparison of the NoC performance between BFT and IBFT architectures. (a) Throughput. (b) Latency. (c) Comparison of the resource cost. (d) Comparison of the distance between nodes.

120MSynOPS, the buffer of BFT is limited, which induces the packet loss and constrains the communication throughput. However, the distance between nodes in IBFT architecture is shorter than BFT, which cuts down the routing time of data flow. The packet can be transmitted at a larger speed and packet loss could be avoided to a certain extent. Thus, the throughput of the IBFT is larger than the conventional BFT under large event rate. Fig. 11(b) shows the comparison of the communication latency. The latencies of the BFT and IBFT architectures are consistent with each other. As the event rate becoming larger and exceeds 120MSynOPS, the data packet waiting in buffer occurs in BFT and increases the communication latency, which induces a larger latency in BFT. In comparison with BFT, communication nodes in IBFT are more close to each other, which means fewer routers will be used in the data transmission. Thus, the waiting time of packets could be reduced and the communication latency can be cut down effectively.

As shown in Fig. 11(c), for an NoC network with node number $N = 16$, the numbers of routers and links in the IBFT architecture are reduced by 14.3% and 10.7%, respectively. For the condition $N = 64$, the numbers could be cut down by 20% and 32.1%. Thus, the NoC complexity and the hardware resource cost can be cut down effectively. The average hops of both NoC architectures are similar to each other. The network diameter of IBFT is larger than BFT, which means a larger latency would occur under the global communication condition. As shown in Fig. 11(d), the communication distance of BFT can be divided into three categories, which are 1 hop, 3 hops, and more than 5 hops. The corresponding numbers of node pairs are 192, 1536, and 2304, respectively. IBFT architecture has more detailed distance categories: 1 hop with 192 node pairs, 2 hops with 256 node pairs, 3 hops with 256 node pairs, 4 hops with 1312 node pairs, and more than 5 hops with 2016 node pairs. More detailed distance categories are helpful for the NoC design when dealing with high local

burden. It means that IBFT can enhance the NoC connectivity effectively.

V. DISCUSSION

Several research groups have simulated sizeable SNNs, with the ultimate aim toward the emulation of the entire human brain. One of the motivations for large-scale SNN simulation is to gain a better understanding of how the brain works. This kind of study requires a more detailed and computationally efficient model of neuron behavior. A critical approach toward the comprehension of brain dynamics is to build a large-scale biologically plausible network with detailed neural activities. The required model should provide insight from biochemistry and neurochemical behavior of individual cells to the behavior of networks of neurons in the cortex and other parts of the brain.

Many projects around the world have successfully simulated large-scale SNNs, which includes SpiNNaker, Blue Brain, Cognitive Computing via Synaptronics and Supercomputing (C2S2) SyNAPSE, BrainScaleS, Neurogrid, IF array transceivers (IFAT), and Truenorth [2], [25], [28]–[31], [38]. The SpiNNaker project uses low-power ARM CPUs to simulate the large-scale SNN, which can realize arbitrary kind of neuron models due to its programmability [28]. It provides significant contributions and powerful platform for neural computing and is more preferable to support point neuron models rather than complex biologically meaningful neuron models due to its hardware resource constraint on a single chip. The Blue Brain project constructed a 10 000 neuron model of a neocortical column, whose simulation time is ten times slower than biological neurons [25]. This model attempts to account for the 3-D morphology of the neurons and cortical column, using about one billion triangular compartments. The “C2S2” project uses a single-compartment spiking Izhikevich neuron model, running on IBM’s Dawn Blue Gene/P supercomputer [38]. BrainscaleS focuses on perceptual systems, which uses AdEXP neuron model for the real-time large-scale brain network simulations [30]. In contrast to the point neuron model used in C2S2 and the thousands of compartments used in Blue Brain, Neurogrid uses an approach as a compromise to provide reasonable accuracy without excessive complexity. A quadratic IF model is used for the somatic compartment and four Hodgkin–Huxley channels are used to model the dendritic compartments. Izhikevich *et al.* [3] used 22 different types of multi-CMNs in the cortex and thalamus and exhibited phenomena that are known to exist in the human brain, such as spontaneous activity and rhythms of spiking activity. The programmability is one of the most critical factors considered in the design of hardware systems, as it provides the capability to change topology, parameters, synaptic model, neuron model, and other vital components of SNNs. Another issue of large-scale simulations is the computational speed. Because biological intelligence is assumed to emerge through real-time interactions between an organism and a real environment [10], real-time computational power is essential for the application of a neuromorphic system. Compared to current major projects and considering biological plausibility, programmability, and real-time computational speed, this paper provides biological

TABLE IV
COMPARISON WITH THE PREVIOUS DIGITAL NEUROMORPHIC STUDIES

Study	Hardware	Model	#Neurons
Glackin <i>et al.</i> , 2005	FPGA	IF	4000
Pearson <i>et al.</i> , 2007	FPGA	LIF	1120
Cassidy <i>et al.</i> , 2007	FPGA	LIF	51
Jin <i>et al.</i> , 2008	ARM	Izhikevich	1000
Thomas and Luk, 2009	FPGA	Izhikevich	1024
Fidjeland <i>et al.</i> , 2010	GPU	Izhikevich	55000
Igarashi <i>et al.</i> , 2011	GPU	Hodgkin–Huxley type	1100
Ambroise <i>et al.</i> , 2013	FPGA	Izhikevich	167
Yamazaki <i>et al.</i> , 2013	GPU	LIF	100,000
Shimada <i>et al.</i> , 2015	FPGA	Compartmental cellular automaton	Single
Cheung <i>et al.</i> , 2016	FPGA	Izhikevich	100,000
Luo <i>et al.</i> , 2016	FPGA	LIF	100,000
Pani <i>et al.</i> , 2017	FPGA	Izhikevich	1440
This study	FPGA	Compartmental Hodgkin–Huxley type	1000,000

meaningful simulation, system flexibility, and the capability of applications in the fields of biomedical and other branches of engineering.

The proposed compartmental Hodgkin–Huxley-type neuron model used in this paper is capable of investigating the mechanisms of soma-to-dendritic interactions that are essential in neural systems [18], [19]. It is biologically meaningful, which plays a pivotal role in the synchronization of neuronal ensembles and also makes them likely agonists of pathological brain activity. Previous studies have proposed various more complicated high-order synaptic learning rules such as triplet and quadruplet models [54], [55]. These models would incur strong penalties in area efficiency and limit the network scale of the neuromorphic system. In contrast, the STDP learning rule used in the proposed neuromorphic system is one of the most recognized learning mechanisms, which has been implemented and utilized in various theoretical models and neuromorphic systems.

A. Comparison With the State-of-the-Art Digital Neuromorphic Studies

Although there are different approaches in neuromorphic engineering and large-scale brain simulation projects, this paper focuses on the digital neuromorphic computing method. Analog-based implementations are constrained by higher cost, lower precision, inflexibility, and significantly longer development cycles. In comparison with analog circuits, digital implementations are more flexible, accurate, and generally require less development time. Previously several digital neuromorphic works have focused on the high-performance implementation of SNNs, including methods based on DSP, ARM, GPU, and FPGA [39]–[41]. Table IV summarizes the state-of-the-art related to digital neuromorphic studies. Although not exhaustive, it presents a good overview of current work in the area. Compared to the other two alternatives, the FPGA provides researchers with the possibility to reconfigure the device by full reload of the configuration bitstream [42], [43]. The presence of IP cores also enables the creation of multiprocessor systems on chip on the FPGA device [44]. Some of the previous digital neuromorphic works used simplified custom neuron models

such as IF or leaky integrate and fire (LIF) [45], [46]. Fidjeland and Shanahan [47] presented the GPU-based SNN with the Izhikevich model. Ambroise *et al.* [48] presented an approach to implementing 167 neurons with a smaller FPGA device. Pani *et al.* [49] successfully implemented the fully connected SNN with 1440 Izhikevich neurons, which is meaningful for digital neuromorphic engineering. Shimada and Torikai [53] presented a multi-CMN model based on cellular automaton algorithm. However, it loses some vital neural dynamics including ionic dynamics and bifurcation dynamics. Increasing numbers of digital neuromorphic studies have focused on larger-scale SNNs [40], [42], [50]. Luo *et al.* [42] used the LIF model to implement the cerebellum neural network with 100 000 neurons, which reproduce the passage-of-time dynamics in real time. Cheung *et al.* [50] presented the NeuroFlow architecture, which can represent the state-of-the-art in the field of digital neuromorphic implementation of SNNs. However, they only focused on the point Izhikevich neuron model, which ignores the morphology of the neuron. In the proposed work, both the biological plausibility and the network scale are considered. The compartmental Hodgkin–Huxley-type neuron model is used, which can reproduce the essential dynamics of the dendrites. Higher biological plausibility and real-time computational speed are obtained in the proposed study. The proposed study is also programmable, which enables the reconfiguration of the network model and enhances the system flexibility to be used in various applications.

There have been previous studies that investigated the multiplier-less realization of the neuron model based on PLA methods [37], [59]–[61]. However, few of these works have presented a method for CMN models with biologically meaningful dynamics. Most of them focused on the PLA design of spike-based models such as LIF and Izhikevich models, which cannot describe the biological behaviors of neurons in the central nervous system [37], [58], [59]. From the computational point of view, there exists a significant difference from the Hodgkin–Huxley-type models because the nonlinear part required to be approximated in these models is the multiplication between the same variables. However, the nonlinear parts in Hodgkin–Huxley-type models are included in multiplication between the same and different variables and hyperbolic functions. Thus, only using the methods proposed in these studies cannot deal with the challenges. Another study has presented two PLA methods for the Morris–Lecar model that is conductance-based with ionic dynamics [60]. Since only the comparison of PLA methods is meaningless, a comparison of the multiplier-less implementation results is presented in Fig. 12(a) and (b), in which the same segments are guaranteed using these methods. As shown in Fig. 12(a), method 1 in [60] costs the most DSP elements, and the most memory bits are cost using method 2. The higher resource cost will severely limit the neuron number in a large-scale network and increase the power consumption of the digital system. As shown in Fig. 12(b), the higher computational error is obtained using method 2 in [60]. Thus, the proposed method is more suitable for the digital realization of the large-scale biologically meaningful neural network.

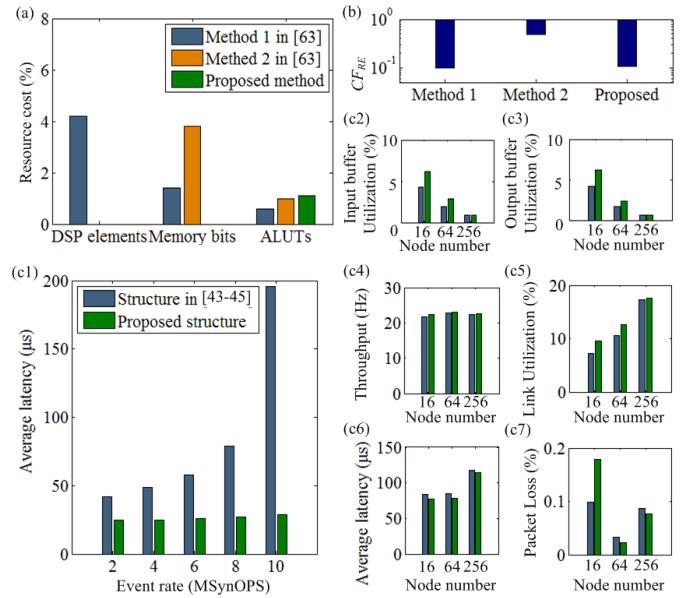


Fig. 12. Performance evaluation of the proposed methods in this paper. (a) Comparison of the resource cost. (b) Comparison of the computational error. (c1) Comparison of the average latency. (c2) Comparison of the input buffer utilization. (c3) Comparison of the output buffer utilization. (c4) Comparison of the throughput. (c5) Comparison of the link utilization. (c6) Comparison of the average latency. (c7) Comparison of the packet loss.

In addition, the NoC-based method for SNNs has been investigated by Luo *et al.* [42] and Carrillo *et al.* [44]. Most of the previous works use the mesh-based topology. A detailed comparison between the structure in [42]–[44] and the presented structure at 40MSynOPS is shown in Fig. 12(c1)–(c7), considering average latency, input and output buffer utilization, the throughput of data packets, link utilization, and packet loss. It demonstrates that the performance of the proposed IBFT structure is superior to the structure in previous works.

B. Impact of the Neuronal Morphology on the Dynamics of Neuromorphic SNNs

In order to clarify the roles of dendritic computation on the neural network dynamics, we further investigate the effects of morphological parameters on network dynamics. As shown in Fig. 13(a), the firing rate f is influenced by both the cellular geometric parameter p and the dendritic stimulation current I_d . Under the stimulation of external currents I_d , the neuron can be in either a quiescent state or a repetitive spiking state as p changes from 0.01 to 0.99. For $p < 0.06$ and $I_d < 40$, there is no repetitive spike generation. For a higher I_d , the firing rate f will be larger. Fig. 13(b) shows the relationship between firing rate f and geometric parameter p with different values of I_d . It can be seen that the firing rate increases with the increase in the geometric parameter with fixed I_d and also increases with the increment of the stimulation current I_d . We also quantify phase-zero synchronization of a network by calculating the bursting measure B , which is defined as

$$B = \left(\frac{\sqrt{\langle \tau^2 \rangle - \langle \tau \rangle^2}}{\langle \tau \rangle} - 1 \right) \frac{1}{\sqrt{N}} \quad (20)$$

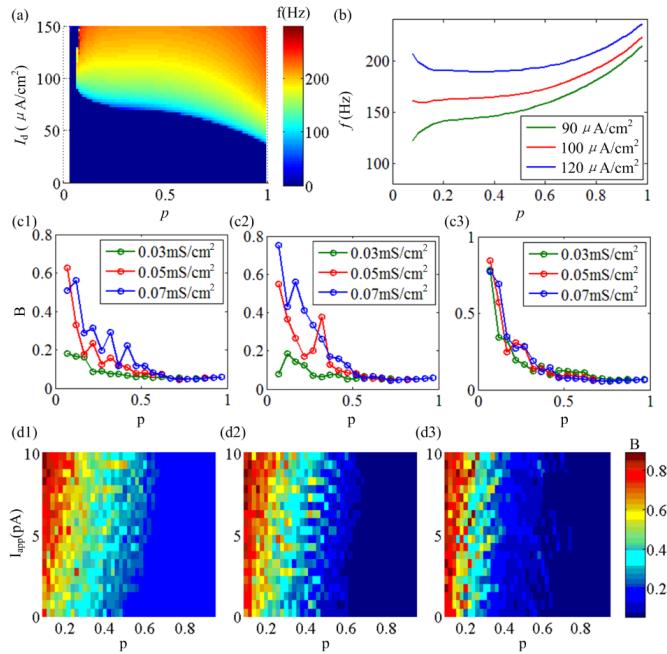


Fig. 13. Dynamical analysis of the neuromorphic SNN. (a) Influence of the dendritic stimulation current and the geometric parameter on the firing rate of the multi-CMN. (b) Influence of the geometric parameter on the neural firing rate with certain dendritic stimulation current. (c1) Relationship between geometric parameter and network synchronization with different coupling strengths when the network connection probability $p_{\text{con}} = 0.3$. (c2) Relationship between geometric parameter and network synchronization with different coupling strengths when the network connection probability $p_{\text{con}} = 0.5$. (c3) Relationship between the geometric parameter and network synchronization with different coupling strengths when the network connection probability $p_{\text{con}} = 0.7$. (d1) Influence of the applied current and the geometric parameter on the network synchronization when $I_d = 90$. (d2) Influence of the applied current and the geometric parameter on the network synchronization when $I_d = 100$. (d3) Influence of the applied current and the geometric parameter on the network synchronization when $I_d = 120$.

where τ_i represents the time difference between spikes i and $i + 1$. The symbol $\langle \cdot \rangle_t$ denotes an average over time.

In Fig. 13(c1)–(c3), the stimulation current $I_d = 90, 100, 120 \mu\text{A}/\text{cm}^2$, and the network connection probability p_{con} is 0.3, 0.5, and 0.7, respectively. For this setup, $B = 0$ represents random spiking and $B = 1$ represents perfect locking at zero phases between all neurons for a large number of total spikes and neurons. The coupling strength $g_{ij} = 0.03, 0.05$, and $0.07 \text{ mS}/\text{cm}^2$, respectively. Results reveal that the increment of the coupling strength will lead to network synchronization. A larger connection probability and stimulation current will reduce the influence of the coupling strength on network synchronization. In Fig. 13(d1)–(d3), the dendritic stimulation current $I_d = 90, 100$, and $120 \mu\text{A}/\text{cm}^2$, respectively. Results show that the increment of dendritic stimulation will induce a reduction of network synchronization. Network synchronization will be reduced as the geometric parameter increases. Larger applied currents on the soma will lead to slightly enhanced network synchronization.

VI. CONCLUSION

This paper presented a digital neuromorphic architecture for the large-scale biologically meaningful neural network with

biophysically plausible dynamics, and scalable efficient implementation on FPGAs, with both cellular and network levels included in the model. As a proof-of-concept, a randomly connected SNN with one million compartmental conductance-based neurons was implemented in an IBFT architecture on high-end Altera Stratix III FPGAs. From the cellular point of view, both the neuronal compartment and the synapse are individually programmable. A modified CMN neuron model is presented using PLA methods, which can generate accurate dynamical characteristics of the CMN model. Using a set of criteria, we demonstrated computational speed can be increased by 56.59% from the original model to the proposed CMN model resulting from the corresponding reduction of computation complexity using the PLA method. By combining the IBFT topology with a novel routing algorithm, the large-scale biologically meaningful neural network can be realized with both the throughput and computational latency enhanced. Considering biological plausibility, programmability, computational efficiency, and network scale, this paper is superior to the state-of-the-art projects and digital neuromorphic studies for large-scale SNNs. Although not pursued here, future works include efficient routing for interboard communication, efficient learning for digital neuromorphic SNN, and efficient logic partitioning.

ACKNOWLEDGMENT

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

REFERENCES

- [1] C. Eliasmith *et al.*, “A large-scale model of the functioning brain,” *Science*, vol. 338, no. 6111, pp. 1201–1205, Nov. 2012.
- [2] P. A. Merolla *et al.*, “Amillion spiking-neuron integrated circuit with a scalable communication network and interface,” *Science*, vol. 345, no. 6197, pp. 668–673, Aug. 2014.
- [3] E. M. Izhikevich and G. M. Edelman, “Large-scale model of mammalian thalamocortical systems,” *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 9, pp. 3593–3598, Mar. 2008.
- [4] A. V. Herz, T. Gollisch, C. K. Machens, and D. Jaeger, “Modeling Single-neuron dynamics and computations: A balance of detail and abstraction,” *Science*, vol. 314, no. 5796, pp. 80–85, Oct. 2006.
- [5] O. Sporns, G. Tononi, and R. Kötter, “The human connectome: A structural description of the human brain,” *PLoS Comput. Biol.*, vol. 1, no. 4, p. e42, Sep. 2005.
- [6] E. Marder and A. L. Taylor, “Multiple models to capture the variability in biological neurons and networks,” *Nature Neurosci.*, vol. 14, no. 2, pp. 133–138, Jan. 2011.
- [7] B. W. Knight, “Dynamics of encoding in a population of neurons,” *J. Gen. Physiol.*, vol. 59, no. 6, pp. 734–766, Jun. 1972.
- [8] G. Halnes, S. Augustininaite, P. Heggelund, G. T. Einevoll, and M. Migliore, “A multi-compartment model for interneurons in the dorsal lateral geniculate nucleus,” *PLoS Comput. Biol.*, vol. 7, no. 9, Sep. 2011, Art. no. e1002160.
- [9] P. F. Pinsky and J. Rinzel, “Intrinsic and network rhythmogenesis in a reduced traub model for CA3 neurons,” *J. Comput. Neurosci.*, vol. 1, nos. 1–2, pp. 39–60, Jun. 1994.
- [10] R. D. Traub, R. K. Wong, R. Miles, and H. Michelson, “A model of a CA3 hippocampal pyramidal neuron incorporating voltage-clamp data on intrinsic conductances,” *J. Neurophysiol.*, vol. 66, no. 2, pp. 635–650, Aug. 1991.
- [11] T. B. Kepler, L. F. Abbott, and E. Marder, “Reduction of conductance-based neuron models,” *Biol. Cybern.*, vol. 66, no. 5, pp. 381–387, Mar. 1992.
- [12] C. Morris and H. Lecar, “Voltage oscillations in the barnacle giant muscle fiber,” *Biophys. J.*, vol. 35, no. 1, pp. 193–213, Jul. 1981.

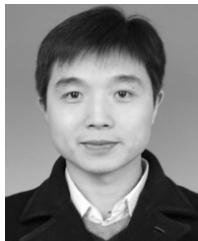
- [13] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *J. Physiol.*, vol. 117, no. 4, pp. 500–544, Aug. 1954.
- [14] C. Koch and I. Segev, "The role of single neurons in information processing," *Nature Neurosci.*, vol. 3, pp. 1171–1177, Nov. 2000.
- [15] E. M. Izhikevich, "Which model to use for cortical spiking neurons?" *IEEE Trans. Neural Netw.*, vol. 15, no. 5, pp. 1063–1070, Sep. 2004.
- [16] C. Koch, "Computation and the single neuron," *Nature*, vol. 385, no. 6613, pp. 1063–1070, Jan. 1997.
- [17] S. B. Laughlin and T. J. Sejnowski, "Communication in neuronal networks," *Science*, vol. 301, no. 5641, pp. 1870–1874, Sep. 2003.
- [18] A. Polksy, B. W. Mel, and J. Schiller, "Computational subunits in thin dendrites of pyramidal cells," *Nature Neurosci.*, vol. 7, no. 6, pp. 621–627, Jun. 2004.
- [19] N. Spruston, "Pyramidal neurons: Dendritic structure and synaptic integration," *Nature Rev. Neurosci.*, vol. 9, no. 3, pp. 206–221, Mar. 2008.
- [20] J. C. Magee, "Dendritic integration of excitatory synaptic input," *Nature Rev. Neurosci.*, vol. 1, no. 3, pp. 181–190, Dec. 2000.
- [21] T. Branco, B. A. Clark, and M. Häusser, "Dendritic discrimination of temporal input sequences in cortical neurons," *Science*, vol. 329, no. 5999, pp. 1671–1675, Sep. 2000.
- [22] H. Markram, "The blue brain project," *Nature Rev. Neurosci.*, vol. 7, no. 2, pp. 153–160, Feb. 2006.
- [23] A. S. Kuznetsov, N. J. Kopell, and C. J. Wilson, "Transient high-frequency firing in a coupled-oscillator model of the mesencephalic dopaminergic neuron," *J. Neurophysiol.*, vol. 95, no. 2, pp. 932–947, Feb. 2006.
- [24] A. Kepecs, X. J. Wang, and J. Lisman, "Bursting neurons signal input slope," *J. Neurosci.*, vol. 22, no. 20, pp. 9053–9062, Oct. 2002.
- [25] M. London and M. Häusser, "Dendritic computation," *Annu. Rev. Neurosci.*, vol. 28, pp. 503–532, Jul. 2005.
- [26] H. de Garis, C. Shuo, B. Goertzel, and L. Ruiting, "A world survey of artificial brain projects. part I: Large-scale brain simulations," *Neurocomputing*, vol. 74, nos. 1–3, pp. 3–29, Dec. 2010.
- [27] 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.
- [28] S. B. Furber, F. Galluppi, S. Temple, and L. A. Plana, "The SpiNNaker project," *Proc. IEEE*, vol. 102, no. 5, pp. 652–665, May 2014.
- [29] B. V. Benjamin *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.
- [30] M. A. Petrovici *et al.*, "Characterization and compensation of network-level anomalies in mixed-signal neuromorphic modeling platforms," *PLoS ONE*, vol. 9, no. 10, Oct. 2014, Art. no. e108590.
- [31] J. Park, T. Yu, S. Joshi, C. Maier, and G. Cauwenberghs, "Hierarchical address event routing for reconfigurable large-scale neuromorphic systems," *IEEE Trans. Neural. Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2408–2422, Oct. 2017.
- [32] S. A. Prescott, Y. De Koninck, and T. J. Sejnowski, "Biophysical basis for three distinct dynamical mechanisms of action potential initiation," *PLoS Comput. Biol.*, vol. 4, no. 10, Oct. 2008, Art. no. e1000198.
- [33] E. Bullmore and O. Sporns, "Complex brain networks: Graph theoretical analysis of structural and functional systems," *Nature Rev. Neurosci.*, vol. 10, no. 3, pp. 186–198, Mar. 2009.
- [34] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore, "A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs," *J. Neurosci.*, vol. 26, no. 1, pp. 63–72, Jan. 2006.
- [35] S. Sahin, Y. Bicerikli, and S. Yazici, "Neural network implementation in hardware using FPGAs," in *Proc. Int. Conf. Neural Inf. Process.*, vol. 4234, Oct. 2006, pp. 1105–1112.
- [36] S. Bellis *et al.*, "FPGA implementation of spiking neural networks—An initial step towards building tangible collaborative autonomous agents," in *Proc. IEEE Int. Conf. Field-Program. Technol.*, Dec. 2004, pp. 449–452.
- [37] S. Yang *et al.*, "Cost-efficient FPGA implementation of basal ganglia and their parkinsonian analysis," *Neural Netw.*, vol. 71, pp. 62–75, Nov. 2015.
- [38] J. Schemmel, D. Briiderle, A. Griegl, M. Hock, K. Meier, and S. Millner, "A wafer-scale neuromorphic hardware system for large-scale neural modeling," in *Proc. IEEE Int. Symp. Circuits Syst.*, May/Jun. 2010, pp. 1947–1950.
- [39] S. Yang *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.
- [40] T. Yamazaki and J. Igarashi, "Realtime cerebellum: A large-scale spiking network model of the cerebellum that runs in realtime using a graphics processing unit," *Neural Netw.*, vol. 47, pp. 103–111, Nov. 2013.
- [41] S. Yang *et al.*, "Efficient hardware implementation of the subthalamic nucleus–external globus pallidus oscillation system and its dynamics investigation," *Neural Netw.*, vol. 94, pp. 220–238, Oct. 2017.
- [42] J. Luo, G. Coapes, T. Mak, T. Yamazaki, C. Tin, and P. Degenaar, "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, Jun. 2016.
- [43] J. Liu, J. Harkin, L. P. Maguire, L. J. McDaid, J. J. Wade, and G. Martin, "Scalable networks-on-chip interconnected architecture for astrocyte-neuron networks," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 63, no. 12, pp. 2290–2303, Dec. 2016.
- [44] S. Carrillo *et al.*, "Advancing interconnect density for spiking neural network hardware implementations using traffic-aware adaptive network-on-chip routers," *Neural Netw.*, vol. 33, pp. 42–57, Sep. 2012.
- [45] J. F. Cassidy *et al.*, "The 2007 Nazko, British Columbia, earthquake sequence: Injection of magma deep in the crust beneath the anahim volcanic belt," *Bull. Seismological Soc. Amer.*, vol. 101, no. 4, pp. 1732–1741, Aug. 2011.
- [46] T. C. Pearson, A. E. Cetin, A. H. Tewfik, and R. P. Haff, "Feasibility of impact-acoustic emissions for detection of damaged wheat kernels," *Digit. Signal Process.*, vol. 17, no. 3, pp. 617–633, May 2007.
- [47] A. K. Fidjeland and M. P. Shanahan, "Accelerated simulation of spiking neural networks using GPUs," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2010, vol. 41, no. 3, pp. 1–8.
- [48] M. Ambroise, T. Levi, Y. Bornat, and S. Saighi, "Biorealistic spiking neural network on FPGA," in *Proc. 47th Annu. Conf. Inf. Sci. Syst. (CISS)*, Oct. 2013, pp. 1–6.
- [49] D. Pani, P. Meloni, G. Tuveri, F. Palumbo, P. Massobrio, and L. Raffo, "An FPGA platform for real-time simulation of spiking neuronal networks," *Frontiers Neurosci.*, vol. 11, p. 90, Feb. 2017.
- [50] K. Cheung, S. R. Schultz, and W. Luk, "NeuroFlow: A general purpose spiking neural network simulation platform using customizable processors," *Frontiers Neurosci.*, vol. 9, p. 516, Jan. 2016.
- [51] J. Lu, J. Yang, Y.-B. Kim, and J. Ayers, "Low power, high PVT variation tolerant central pattern generator design for a bio-hybrid micro robot," in *Proc. IEEE 55th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Aug. 2012, vol. 59, no. 5, pp. 782–785.
- [52] M. E. J. Newman and D. J. Watts, "Renormalization group analysis of the small-world network model," *Phys. Lett. A*, vol. 263, nos. 4–6, pp. 341–346, Dec. 1999.
- [53] N. Shimada and H. Torikai, "A novel asynchronous cellular automaton multicompartment neuron model," *IEEE Trans. Circuits Syst., II, Exp. Briefs*, vol. 62, no. 8, pp. 776–780, Aug. 2015.
- [54] R. Gopalakrishnan and A. Basu, "Triplet spike time-dependent plasticity in a floating-gate synapse," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 4, pp. 778–790, Apr. 2017.
- [55] M. R. Azghadi *et al.*, "A hybrid CMOS-memristor neuromorphic synapse," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 2, pp. 434–445, Apr. 2017.
- [56] R. J. Vogelstein, U. Mallik, J. T. Vogelstein, and G. Cauwenberghs, "Dynamically reconfigurable silicon array of spiking neurons with conductance-based synapses," *IEEE Trans. Neural Netw.*, vol. 18, no. 1, pp. 253–265, Jan. 2007.
- [57] R. Serrano-Gotarredona *et al.*, "CAVIAR: A 45k neuron, 5M synapse, 12G connects/s AER hardware sensory-processing–learning–actuating system for high-speed visual object recognition and tracking," *IEEE Trans. Neural Netw.*, vol. 20, no. 9, pp. 1417–1438, Sep. 2009.
- [58] M. Nouri, G. R. Karimi, A. Ahmadi, and D. Abbott, "Digital multiplierless implementation of the biological FitzHugh–Nagumo model," *Neurocomputing*, vol. 165, pp. 468–476, Oct. 2015.
- [59] S. Gomar and A. Ahmadi, "Digital multiplierless implementation of biological adaptive-exponential neuron model," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 4, pp. 1206–1219, Apr. 2014.
- [60] M. Hayati, M. Nouri, S. Haghiri, and D. Abbott, "Digital multiplierless realization of two coupled biological Morris–Lecar neuron model," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 62, no. 7, pp. 1805–1814, Jul. 2015.

- [61] H.-X. Wang, R. C. Gerkin, D. W. Nauen, and G.-Q. Bi, "Coactivation and timing-dependent integration of synaptic potentiation and depression," *Nature Neurosci.*, vol. 8, pp. 187–193, Jan. 2005.
 [62] H. Z. Shouval, S. Wang, and G. M. Wittenberg, "Spike timing dependent plasticity: A consequence of more fundamental learning rules," *Frontiers Comput. Neurosci.*, vol. 4, p. 19, Jul. 2010.



Shuangming Yang (S'18) received the B.S. degree from the Hebei University of Technology, Tianjin, China, in 2013, and the M.S. degree from Tianjin University, Tianjin, in 2016, where he is currently pursuing the Ph.D. degree with the School of Electrical and Information Engineering.

His current research interests include neuromorphic engineering, neural system modeling, neural control engineering, robotic control, and machine learning.



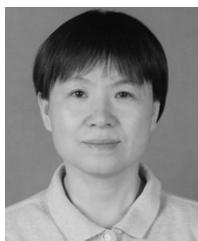
Bin Deng (M'18) received the Ph.D. degree in electrical engineering from Tianjin University, Tianjin, China, in 2007.

He is currently a Professor with the School of Electrical and Information Engineering, Tianjin University. His current research interests include dynamic analysis of neuron model and nonlinear analysis of neuron electrical information.



Jiang Wang was born in Hebei, China, in 1964. He received the master's degree in power and automation engineering and the Ph.D. degree from Tianjin University, Tianjin, China, in 1989, and 1996, respectively.

He is currently a Professor with the School of Electrical and Information Engineering, Tianjin University. His current research interests include nonlinear dynamical systems, neuroscience, and information processing and detecting.



Huiyan Li received the Ph.D. degree from Tianjin University, Tianjin, China, in 2007.

She is currently a Co-Professor with the School of Automation and Electrical Engineering, Tianjin University of Technology and Education, Tianjin. Her current research interests include nonlinear systems and neural networks.



Meili Lu was born in Hubei, China, in 1981. She received the B.S., M.S., and Ph.D. degrees from Tianjin University, Tianjin, China, in 2004, 2007, and 2013, respectively.

She is currently a Lecturer with the School of Information Technology Engineering, Tianjin University of Technology and Education, Tianjin. Her current research interests include analysis of neural modeling, neural control engineering, and analysis of neural dynamics.



Yanqiu Che (M'18) received the B.S., M.S., and Ph.D. degrees in control theory and engineering from Tianjin University, Tianjin, China, in 2003, 2005 and 2008, respectively.

From 2012 to 2015, he was a Visiting Research Associate with Penn State University, University Park, PA, USA. Since 2010, he has been an Associate Professor with the School of Automation and Electrical Engineering, Tianjin University of Technology and Education, Tianjin. He is currently with the Department of Neurosurgery, Penn State College of Medicine, Hershey, PA, USA. His current research interests include neurodynamics, neural control engineering, and brain-computer interface.



Xile Wei was born in Shaanxi, China, in 1975. He received the B.S., M.S., and Ph.D. degrees from Tianjin University, Tianjin, China, in 1997, 2004, and 2007, respectively.

He is currently an Associate Professor with the School of Electrical Engineering and Automation, Tianjin University, Tianjin. His current research interests include analysis of bioelectromagnetics effects, neuromorphic modeling, neural control engineering, analysis of neural dynamics, and nonlinear control theory.



Kenneth A. Loparo (LF'99) is currently a Nord Professor of engineering and the Chair of the Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH, USA.

His current research interests include stability and control of nonlinear systems with applications to modeling, simulation, and variability analysis of biological systems.

Dr. Loparo is a fellow of the American Institute for Medical and Biological Engineering.