Microelectronics Journal 68 (2017) 23–31

Contents lists available at ScienceDirect

Microelectronics Journal

journal homepage: www.elsevier.com/locate/mejo

A dummy cell added neural network using in pattern recognition for prevention of failed events

Cheng Li, Yun-Heub Song*

Department of Electronics and Computer Engineering, Hanyang University, Seoul 04763, South Korea

ARTICLE INFO

Keywords:
Neural network
Integrate-and-fire
Pattern recognition
Neuromorphic

ABSTRACT

Brain-inspired neuromorphic computing systems are receiving significant attention. A typical neuromorphic computing system is the neuron network, whose basic performance is the integrate-and-fire operation. However, latency issues can occur if the integrated signal is not sufficient during the integration process, the integration time is too long, or no firing occurs. In this paper, we propose a dummy cell added neural network to ensure complete I & F operation. The dummy cell compensates the weak signals to ensure a complete I & F operation and to modulate the integration time; but makes negligible influence on the strong signals. The firing rate of a weak signal increases from 80% to 100%. Finally, we analyzed the external area consumption of dummy cells, it can be reduced as small as a few thousandths with large number of input neurons. This proposed scheme can be used in pattern recognition to increase reliability and modulate the integration time.

1. Introduction

Conventional computing systems are limited by the von Neumann bottleneck, where the central processing unit (CPU) is physically separated from the memory. Recently, researchers have f\ocused on the brain-inspired neuromorphic computing system, which co-localizes memory with computation [1]. This is typically analyzed as a network of neurons and performs read/write operations in parallel instead of the conventional one-by-one access mechanism of the von Neumann system. The neural networks were initially realized via software; however, the power consumption is excessive. To reduce the power consumption and retain the brain-like parallel computation, a type of specialized hardware neural network (HNN) is implemented with complementary metal-oxide-semiconductor (CMOS) technology. The basic elements of a neural network are neurons and synapses. The primary limitation of mass integration in the CMOS-based HNN is the area of the synapse (the human brain is believed to have $\sim 10^{10}$ neurons and $\sim 10^{14}$ synapses). The emergence of memristive synapse has promoted the development of HNN for the excellent scalability and synapse-like characteristics of memristors. Several types of memristive synapses have been implemented, including resistive random access memory (ReRAM) [2–10], spin-transfer torque magnetic random-access memory (STT-MRAM) [11–13], and phase-change material (PCM) devices [14–22]. We prefer to use PCM as the synapse for its excellent properties such as multi-level resistance values, strong data retention, high endurance, promising reliability, CMOS compatibility, and technological maturity.

The conventional neuron circuit in full CMOS HNN is designed to faithfully imitate the biological neuron, which shows ion movement and provides a specific spike shape [23–26]. With the emergence of the memristive synapse, the neuron circuit is required to be designed to fit the synaptic behaviors of memristors [5,8–10]. The basic operation of a neuron is integrate-and-fire (I & F), a model of which is shown in Fig. 1. The input signals (from $\times_1$ to x_n) generated from pre-synaptic neurons (PREs) are weighted using correlated synaptic weights (from w_{1j} to w_{nj}). The weighted signals are then integrated in a post-synaptic neuron (POST) based on the transfer function (Eq. (1)) to generate an integrated signal (ζ_j), which is then compared with a threshold value (??). Once the integrated signal exceeds the threshold, the POST then fires based on the activation function (Eq. (2)).

$$\zeta_j = \sum_{i=1}^{n} (x_i \bullet w_{ij}) \tag{1}$$

$$O_j = \varphi(\zeta_j, \theta) \tag{2}$$

There is a latent issue during the integration process: if the integrated input signal is insufficient, the neural network takes too long to integrate before firing, or does not fire at all. In this paper, we propose a neural network with a dummy input neuron that works together with other input neurons to complement the insufficient integrated signal. Complete I & F operation is guaranteed with the dummy input neuron, and the integration time is limited to a desired range.

* Corresponding author.
E-mail address: yhsong2008@hanyang.ac.kr (Y.-H. Song).

http://dx.doi.org/10.1016/j.mejo.2017.08.013
Received 11 April 2017; Received in revised form 11 August 2017; Accepted 26 August 2017

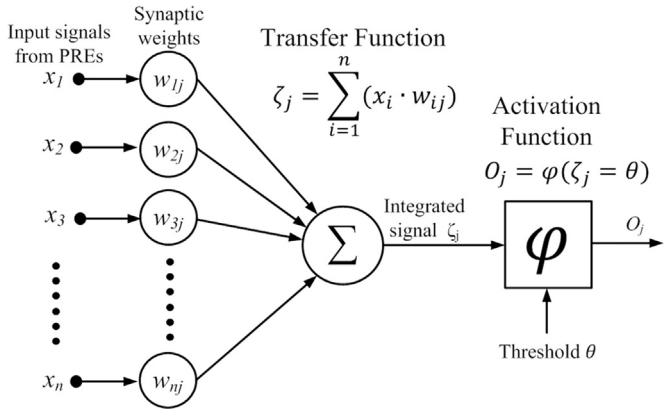

Fig. 1. Model of the integrate and fire (I & F) operation.

2. Neural network used for pattern recognition

2.1. Classical neural network

Pattern recognition is a typical application of neural networks. A 2-layer classical neural network (CNN) with n PREs and m POSTs is actualized by connecting each PRE with every POST via a synapse, as shown in Fig. 2(a). Each PRE, or input neuron (from IN_1 to IN_n), correlates to one pixel of the input pattern and provides specified signals based on whether the pixel is on or off. The POSTs, or output neurons (from ON_1 to ON_m), are implemented via leaky integrate-and-fire (LIF) neurons (Fig. 3). A crossbar of the memristive device acts as the synapse (called 1R-synapse), in which each synapse is realized by one memristor. A 2R-structure synapse was recently proposed to overcome the asymmetry between long-term-potentiation (LTP) and long-term-depression (LTD) [7,14,15]. This type of synapse consists of two memristive devices, one for LTP, which makes a positive distribution on the integration process, and the other is for LTD, which makes a negative distribution on the integration process. The network with 2R-structure synapses is shown in Fig. 2(b). An external control circuit (ECC) is necessary to perfect the function of the neural network, and includes a winner-takes-all (WTA) mechanism and synchronization of the pre- and post-synaptic spikes in spike-timing-dependent-plasticity (STDP) learning.

2.2. Latency issues

In practical pattern recognition applications, the HNN is expected to be capable of processing massive complex patterns, among which the amount of on-pixels varies significantly. In some cases, patterns with a fewer number of on-pixels generate a small integrated input signal that is insufficient and incapable of driving the I & F neuron into the firing mode. As an example, three patterns with 30 pixels are shown in Fig. 4(a). The amount of on-pixels are 8, 12, and 14 for the patterns representing digits “1”, “9”, and “8”, respectively. We research the strength of the integrated signals through a statistical approach. In this work, some resistive devices act as the synapses, 30 PREs provide the required voltage, and one POST receives the integrated signal. Initially, both sides of the synapses remain at the rest voltage V_{rest}. As soon as the integration process begins, the PREs correlating to on-pixels provide a pulse with a voltage that is 30 mV higher than V_{rest} as the excitatory signal, and the PREs correlating to off-pixels retain V_{rest} as the inhibitory signal. A current is then generated through the synapses correlating to the on-pixels and the summation of the currents is the integrated signal, I_{in}, which can be rewritten as Eq. (3). Here, V_i and G_i are the voltage difference and the conductance of i^{th} synapse, respectively. The resistance of resistive synapses are distributed uniformly

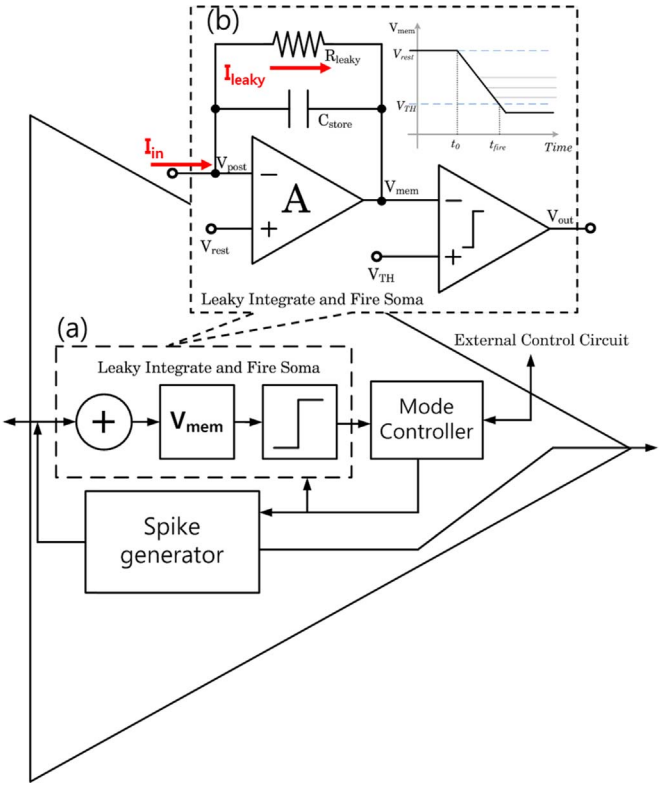

Fig. 3. (a) Basic blocks of an I & F neuron and (b) schematics of a simplified LIF soma.

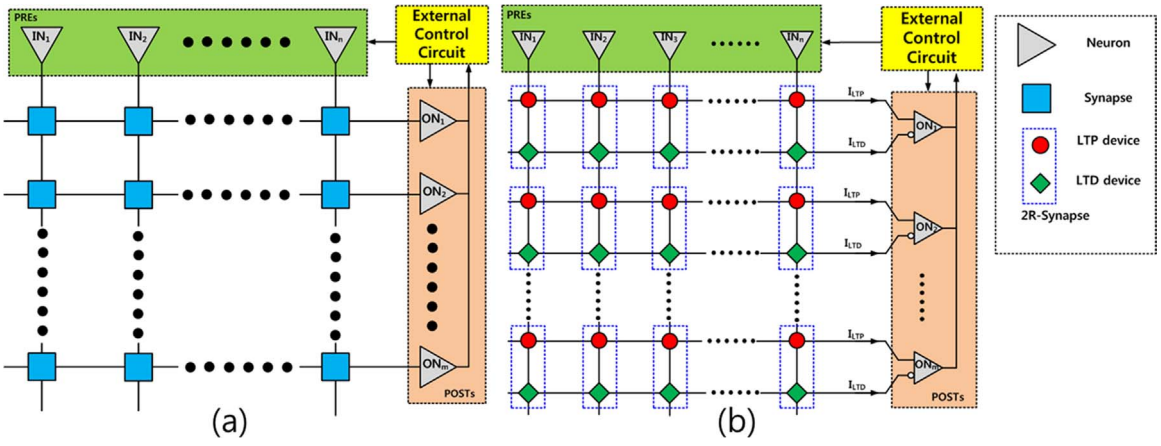

Fig. 2. A 2-layer classical neural network with (a) 1R-synapses and (b) with 2R-synapses.