2D-Motion Detection using SNNs with Graphene-Insulator-Graphene Memristive Synapses

Shubham Pande, Karthi Srinivasan, Suresh Balanethiram, Bhaswar Chakrabarti, Anjan Chakravorty

2022 29th IEEE International Conference on Electronics, Circuits and Systems (ICECS) | 978-1-6654-8823-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICECS202256217.2022.9970889

***Abstract*—The event-driven nature of spiking neural networks (SNNs) makes them biologically plausible and more energy-efficient than artificial neural networks. In this work, we demonstrate SNN-based motion detection of an object in a two-dimensional visual field. The network architecture presented here is biologically plausible and uses CMOS-based analog leaky integrate-and-fire neurons and ultra-low power RRAM synapses. We have investigated the role of visual field optimization on network performance in terms of accuracy and power cost. Also, noise immunity of the proposed architecture is studied by injecting random fluctuations across the spatial locations in the visual field. Detailed transistor-level SPICE simulations show that the proposed structure can accurately and reliably detect complex motions of an object in a two-dimensional visual field.**

***Index Terms*—RRAM, Synapse, LIF Neuron, Compact model, SNN, Motion Detection**

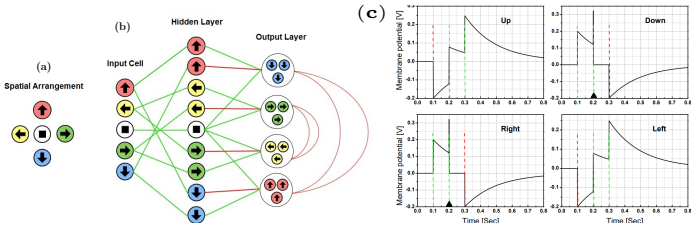

Fig. 1: Unit cell architectures: (a) spatial arrangement of the input layer neuron (b) modified proposed network showing the neurons in the input, hidden, and output layers; LIF neurons represented by circles are interconnected by RRAM-based synapses shown by lines. (c) The membrane potential evolution of the output neurons when the stimulus is moving diagonally from Left-Up to Right-Down. Vertical dotted lines indicate input spikes (green for excitatory and red for inhibitory). Triangle markers indicate output spikes.

I. Introduction

In the field of computer vision, motion detection has been traditionally performed using machine learning algorithms on hardware that are based on the von Neumann architecture. As modern-day microprocessors face the infamous memory bottleneck issue, these approaches often turn out to be power-hungry [1]. On the other hand, spiking neural networks (SNNs) emulate the dynamics of biological neural networks in electronic circuits and deliver low-power, inherently parallel computation. Event-driven computation thus becomes an alternative in realizing energy-efficient hardware for numerous cognitive tasks [2]. The emergence of novel non-volatile memory technologies such as resistive random-access memories (RRAM), phase-change memories, and spin-transfer torque RAMs has further accelerated the development of SNNs [3]. Several techniques have been proposed in the past for motion detection using SNNs [4–8]. In [4], a novel temporal coding scheme has been used to encode the real-time motion into a series of spikes, and a gradient-descent learning algorithm is utilized for training the two-layered SNN. Various studies, such as probabilistic feed-forward synapses [6], rate-based Hebbian learning [7], and spike-timing-dependent plasticity, [8] have been proposed to realize directional selectivity and solve the motion detection problem. These approaches have shown promising results. However, they often come up with hardware implementation challenges. On the other hand, considering biological neural systems' high efficacy and accuracy, authors in [9] have shown insect-inspired elementary motion detection. Onboard computational bandwidth is limited in applications such as robotics and compact visual surveillance. Therefore, it is envisaged that the ability to mimic an insect's navigation system would significantly benefit these applications. In this work, we modify the network presented in [9] and extend its capability to detect complex motions. Furthermore, we investigate the role of visual field optimization on network performance in terms of accuracy and power cost. Finally, we evaluate the noise immunity of the proposed network by injecting noise into the visual field at a random spatial location. Our proposed SNN uses CMOS-based analog leaky integrate-and-fire (LIF) neurons and ultra-low-power, forming-free RRAMs as synapses. The SPICE implementation of our network uses an experimentally validated physics-based Verilog-A model for the synapses and transistor-level implementations for the neurons.

II. Network Architecture and Its Operation

A delay and correlate network was presented in [9]. The basic idea is to introduce a temporal delay between two spatially adjacent inputs followed by a downstream mechanism for detecting the spatio-temporal correlations. The spatial arrangement of the input layer of a unit cell is shown in Fig. 1(a). The neuron at the center (white circle) is denoted as a neutral neuron as it is not pointing to any of the cardinal directions. Neurons in the input layer are connected to the output layer neurons through a hidden layer, as shown in Fig. 1(b). We introduce lateral inhibition (indicated by the red lines) between Up-Down and Left-Right output neurons to emphasize that an object cannot simultaneously move in the Up-Down or Left-Right direction. This modification reduces the spiking activity for the output neurons that are not meant to fire for the given excitation, thus improving the network accuracy and energy cost. Additionally, we use N number of (instead of a single) output neurons per direction to make the network capable of detecting objects moving with a wide range

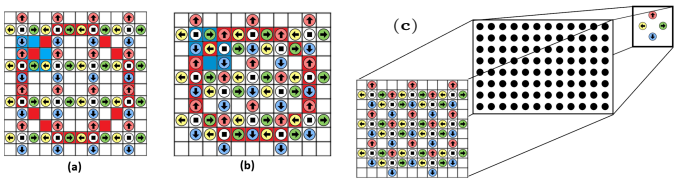

Fig. 2: The tessellation pattern consists of: (a) 12×12 visual field with 16 unit cells (VF_1); (b) 15 unit cells, arranged in a staggered manner for a 10×11 visual field (VF_2). Red boxes denote the locus of the center of the 3×3 square stimulus, while the stimulus itself is shown in blue. (c) Structure of the implemented SNN. Here neurons (shown by circles) in one layer are connected to the neurons in other layers through synapses (not shown to avoid cluttering).

of frequencies. The value of N and the time constants of each output neuron can be adjusted to meet the application-specific requirements.

Input layer neurons in Fig. 1(b) denote the spatial activity of the object in the visual field and encode it by introducing the temporal delay between the activities in the spatially adjacent region. This information from the input layer is passed on to the hidden layer through synapses. Appropriately adjusted time constants of the hidden layer neurons coupled with the information received from the input layer generate a unique spatio-temporal pattern encoding the spatial activity at the input layer. In this case, the neutral neuron's time constant is shorter than the other four neurons. Finally, the spatio-temporal pattern generated at the hidden layer is transmitted to the output layer. The incoming spikes to the output layer may have an excitatory (green lines) or inhibitory (red lines) effect on the output layer. The output layer neurons are designed with threshold voltages and time constants such that they will fire if they receive two consecutive excitatory spikes within a short time interval. However, if an inhibitory spike is received shortly before the first spike, the membrane voltage is reduced sufficiently, and the excitatory pair of spikes do not cause an output spike. Fig. 1(c) illustrates the working of the unit cell when the object moves diagonally from Left-Up to Right-Down direction across an input cell. The input neurons get excited in the order: Left and Up → Center → Right and Down. The temporal delays in the activities at the spatially adjacent location are encoded at the input layer and are coupled with the appropriately adjusted time constants of the hidden layer neurons. This results in a spike pattern at the output neurons as follows: Down and Right receive two excitatory spikes followed by one inhibitory spike; Up and Left receive one inhibitory spike followed by two excitatory spikes. Hence, the Right and Down neurons spike for this input sequence while the other two do not.

The input cell shown in Fig. 1(a) is used to tessellate both the visual fields, as shown in Fig. 2(a)&(b). In order to investigate the impact of tessellation pattern across the visual field, we consider two visual fields, VF_1 and VF_2. Each input cell has its own patch of hidden layer neurons connected to a common output layer, as shown in Fig. 2(c). The network utilizing VF_1 (VF_2) has a total of 80 (75) input neurons, 144 (135) hidden neurons, and 4N output neurons, where N varies from 1 to 5 for performance comparison. A total of 336 (315) synapses are required to realize the connections between the neurons.

III. SPICE Implementation

A. Synapse Model

We demonstrated "forming-free" operation, with the ultra-low operating current, in graphene-insulator-graphene (GIG) RRAM devices [10]. The forming step, and high operating currents often demand bulky and energy-hungry peripheral circuits, therefore, our GIG RRAM devices are promising for low power application. The electrical equivalent schematic of the RRAM used for the simulations is shown in Fig. 3(a)(top). The model framework assumes the existence of a pre-formed filament in the top TiO_2 and the bottom TiO_x layers because of the forming-free nature of the device. The total series resistance offered by the graphene electrode and TiO_2 layer is modeled as a lumped resistor, R_{gr} + R_{TiO_2}. The resistive switching behavior in this device is attributed to the formation and rupture of a conductive filament in the Al_2O_3 layer (as shown in Fig. 3(b)), and it is modeled as a series connection of variable resistor ($R_{Al_2O_3}$) and controlled current source (I_{gap}). Fig. 3(c) shows the simulated current-voltage characteristics of GIG devices. One can look into our earlier work for a detailed discussion on model development and related parameters [11].

B. Analog LIF neuron

The schematic of the LIF neuron used in this work is shown in Fig. 3(a)(bottom). The currents I_{in} coming from the synapses (connected at the input terminal of the LIF neuron) charge up the capacitor. An internal state variable, the membrane potential $V_m(t)$, characterizes each spiking neuron. When this membrane potential exceeds a fixed threshold (V_{th}), a spike is sent out, and the capacitor is discharged to a reset potential (V_{reset}) using the voltage-controlled NMOS

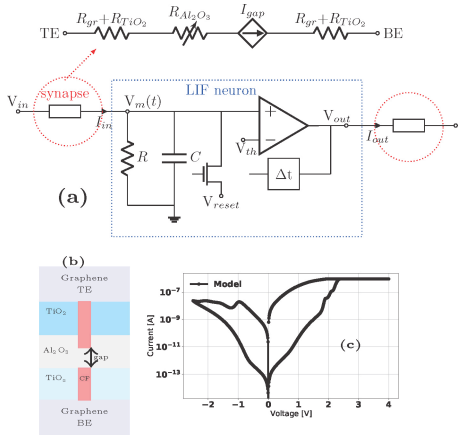

Fig. 3: (a) The electrical equivalent of the RRAM used in this study [top]. Schematic of a Leaky Integrate-and-Fire (LIF) neuron. (b) 2-D schematic of the GIG devices (c) Current-voltage characteristics obtained using our model

switch. A comparator and a delay block are used to generate the constant-width output spikes. The neuron block is implemented using TSMC's 180 nm CMOS standard cell library.

C. Stimulus Generation

The input spikes denoting the object's motion are generated using a peripheral circuit that mimics a dynamic vision sensor (DVS). Each pixel in a DVS is sensitive to relative temporal contrast in the luminous intensity, and upon detection of a sufficiently significant change, it sends out a spike [12]. To emulate this behavior, we treat each event as a vector $e = (x, y, t)$, where x and y define the pixel location in the visual field, and t denotes the time of the event. The moving object's trajectory is modeled using the appropriate mathematical function, and it is used to excite the input layer.

IV. Results

A SPICE simulation of the network shown in Fig. 2(c) is done using Cadence Virtuoso. The synaptic conductance values, LIF neuron's time constants, and threshold voltages are analytically obtained. However, for large networks, one could use bio-inspired algorithms such as genetic algorithms for obtaining relevant network parameters. The object is assumed to be a square block of 3×3 pixels shown in blue in Fig. 2(a), (b). The red color denotes the circular locus of the object's center. Whenever the object is at a specific location, all the neurons that the object covers produce a spike that propagates through the network's hidden layers, as shown in Fig. 2(c). We expect a significant increase in the firing activity of the output neuron when the object moves across the visual field in its corresponding direction. The output firing rates for circular and eight-shaped trajectories are shown in Fig. 4(a), (b). For a circular input path, the four output neurons produce sinusoidal firing with a rate having relative phase lags of 90° between them, as shown in Fig. 4(a). This can be understood by considering the UP-DOWN and LEFT-RIGHT neuron pairs as two orthogonal basis elements encoding the direction of motion. The stimulus travels rightwards from 0 to T/2, leftwards from T/2 to T, upwards from 3T/4 to T/4, and downwards from T/4 to 3T/4, where T is the time period. The firing rates of the corresponding output neuron reach their maxima at the midpoint of the time intervals mentioned above. Similarly, for an eight-shaped input path, the firing rate of the UP-DOWN neurons is half of that of the LEFT-RIGHT neurons. Here, the stimulus makes two oscillations along the horizontal direction for every oscillation along the vertical direction, as shown in Fig. 4(b). The instantaneous firing rate (f^o) is calculated by filtering the spike trains using a low-pass filter as $f^o(t) = h(t) * S(t)$, where * denotes the convolution operation. Here $S(t)$ refers to the spike train (a sequence of impulses at the locations of the spikes) and $h(t) = \lambda(e^{-t/\tau_1} - e^{-t/\tau_2})u(t)$ denotes the impulse response of the aforementioned filter with τ_1 and τ_2 as constants that are set to the mean and twice the mean of the output neuron's time constants. λ is a normalization constant and $u(t)$ is the unit step function. The ideal spiking rates (f^o_{ideal}) corresponding to the four output neurons are obtained from the parameterization $(x(t), y(t))$ of the path for the considered trajectories (circular and eight shaped) as

$$f^o_{i,ideal}(t) = \frac{f_{max}}{2}\left|\frac{\dot{p}(t)}{\dot{p}_{max}} + 1\right| \tag{1}$$

where i takes on four values depending on the output channel (UP, DOWN, LEFT, RIGHT). For UP and DOWN (LEFT and RIGHT), the directional velocity $\dot{p}(t)$ and its maximum value $\dot{p}_{max}$ are replaced by $\dot{y}(t)$ and $\dot{y}_{max}$ ($\dot{x}(t)$ and $\dot{x}_{max}$), respectively. Here f_{max} is the maximum firing rate of the output neurons. The ideal firing rates reach their maximum when the object moves at the maximum velocity along the preferred direction and go to zero when the object moves against the preferred direction. In case of no movement either along or against, they settle to half of the maximum rate. The firing rate obtained from the SPICE simulations (solid lines) is in good agreement with the ideal firing rate (dotted lines), as observed in Fig. 4(a) & (b).

Next, we use N (instead of one) output neuron(s) per direction to make the network sensitive to a wide range of moving object frequencies. The network with five output neurons per direction, with logarithmically distributed time constants ranging from 5 ms to 500 ms, is found to perform better, as shown in Fig. 4(c). The accuracy score of the network is

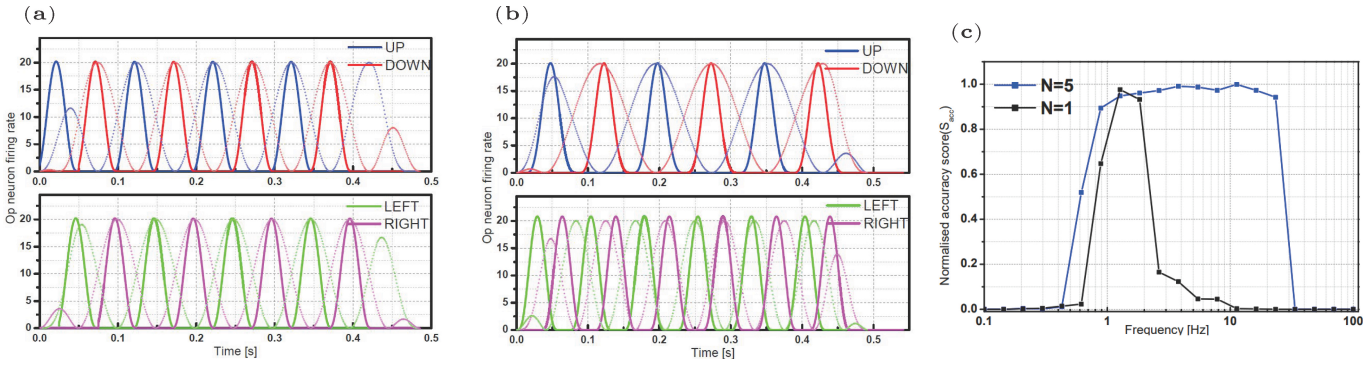

Fig. 4: Ideal (dotted lines) and SPICE simulated (solid lines) firing rates of the output neuron for an object moving in (a) a circular path and (b) an eight-shaped path. These results correspond to VF_1. (c) Normalized accuracy score versus input motion frequency: performance comparison of single output neuron with a single time constant (500 ms) and a larger number of output neurons with a range of time constants (5 ms-500 ms).

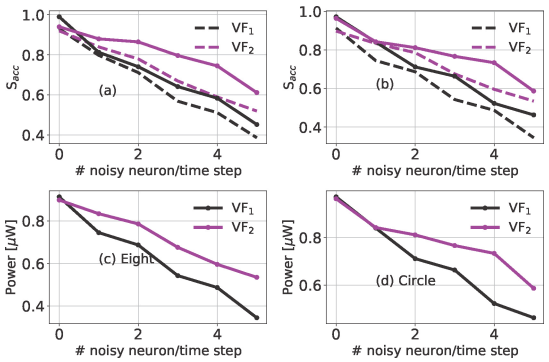

Fig. 5: Normalized accuracy score versus the number of noisy neurons for (a) a circular path (b) an eight-shaped path. The accuracy score is shown for both VF_1 and VF_2 with (solid line) and without (dashed line) lateral inhibition. Power consumption comparison for (c) eight and (d) circular shaped path for both VF_1 and VF_2 considering lateral inhibition.

defined as

$$S_{acc} = \frac{1}{4}\sum_{i}\left(1 - \frac{|f^{o}_{i,ideal}(t) - f^{o}_{i,measured}(t)|^2}{|f^{o}_{i,ideal}(t)|^2}\right). \quad (2)$$

The design of the visual field plays a pivotal role in the performance of the network. Therefore, we consider two visual fields for the analysis: VF_1 and VF_2. To evaluate the noise immunity of the network, we introduce noisy neurons (randomly located) per time-step across the visual field during the simulation. Fig. 5(a), (b) show the accuracy score comparison against the number of noisy neurons for both the visual fields. As the number of noisy neurons per time-step increases, the accuracy score is expected to decrease; however, this reduction is less in the case of VF_2 than VF_1. VF_2 shows better noise immunity for both circular and eight-shaped paths than VF_1. We attribute this improvement in noise immunity of VF_2 to the reduced sparsity in the visual field and better SNR in comparison with VF_1. Also, it can be noted that the presence of lateral inhibition has improved the accuracy scores significantly in both the visual fields. Finally, we compare the power consumption associated with both the VF_1 and VF_2, shown in Fig. 5(c), (d). Reduced sparsity in VF_2 has led to improvement in SNR and consequently in the accuracy score; however, this improvement is accompanied by relatively higher power consumption in VF_2 than VF_1.

V. Conclusion

We have presented a SPICE simulation framework for SNN-based motion detection. We design our network using an experimentally validated physics-based synapse model and transistor-level implementation of LIF neurons. We demonstrate improved network performance by including lateral inhibition and a pool of N output neurons per direction. The proposed network can accurately encode the object's direction of movement over a significant frequency range of the moving objects. Also, we discuss the trade-off involved in the design of visual fields. A sparser visual field is found to be advantageous from a power consumption viewpoint; however, they might underperform in terms of accuracy and noise immunity.

VI. Acknowledgment

This work is partially funded by Science and Engineering Research Board (SERB), a statutory body of the Department of Science and Technology, Government of India, through the project SRG/2020/001126, and Prime Minister Research Fellowship.

References

[1] S. A. McKee. Reflections on the memory wall. Association for Computing Machinery, 2004. doi: 10.1145/977091.977115.

[2] Michael Pfeiffer and Thomas Pfeil. Deep learning with spiking neurons: Opportunities and challenges. *Frontiers in Neuroscience*, 12, 2018.

[3] S. Yu. Neuro-inspired computing with emerging non-volatile memorys. *Proceedings of the IEEE*, 106(2):260–285, 2018. doi: 10.1109/JPROC.2018.2790840.

[4] J. Yang, Q. Wu, M. Huang, and T. Luo. Real time human motion recognition via spiking neural network. 366:012042, jun 2018. doi: 10.1088/1757-899x/366/1/012042.

[5] G. Orchard, R. Benosman, R. Etienne-Cummings, and N. V. Thakor. A spiking neural network architecture for visual motion estimation. In *2013 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 298–301, 2013. doi: 10.1109/BioCAS.2013.6679698.

[6] N. J. Buchs and W. Senn. Learning direction selectivity through spike-timing dependent modification of neurotransmitter release probability. *Neurocomputing*, 38:121–127, 2001.

[7] J. C. Feidler, A. B. Saul, A. Murthy, and A. L. Humphrey. Hebbian learning and the development of direction selectivity: the role of geniculate response timings. *Network: Computation in Neural Systems*, 8(2):195–214, jan 1997. doi: 10.1088/0954-898x˙8˙2˙006.

[8] N. J. Buchs and W. Senn. Spike-based synaptic plasticity and the emergence of direction selective simple cells: simulation results. *Journal of computational neuroscience*, 13(3):167–186, 2002.

[9] T. Dalgaty, E. Vianello, D. Ly, G. Indiveri, B. Salvo, E. Nowak, and J. Casas. *Insect-Inspired Elementary Motion Detection Embracing Resistive Memory and Spiking Neural Networks*, pages 115–128. 07 2018. ISBN 978-3-319-95971-9. doi: 10.1007/978-3-319-95972-6˙13.

[10] B. Chakrabarti, T. Roy, and E. M. Vogel. Nonlinear switching with ultralow reset power in graphene-insulator-graphene forming-free resistive memories. *IEEE Electron Device Letters*, 35(7):750–752, 2014. doi: 10.1109/LED.2014.2321328.

[11] L. H. Reddy, S. Pande, T. Roy, E. M. Vogel, A. Chakravorty, and B. Chakrabarti. A spice compact model for forming-free, low-power graphene-insulator-graphene reram technology. *Emergent Materials*, 4:1055 – 1065, 2021.

[12] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. A 128x128 120 db 15μs latency asynchronous temporal contrast vision sensor. 2006.