

Robust Learning and Recognition of Visual Patterns in Neuromorphic Electronic Agents

39

Abstract—Mixed-signal analog/digital neuromorphic circuits are characterized by ultra-low power consumption, real-time processing abilities, and low-latency response times. This makes them promising for robotic applications that require fast and power efficient computing. However, the unavoidable variance inherently existing in the analog circuits makes it challenging to develop neural processing architectures able to perform complex computations robustly. In this paper, we present a spiking neural network architecture with spike-based learning that enables robust learning and recognition of visual patterns in noisy silicon neural substrate and noisy environments. The architecture is used to perform pattern recognition and inference after a training phase with computers and neuromorphic hardware in the loop. We validate the proposed system in a closed-loop hardware setup composed of neuromorphic vision sensors and processors, and we present experimental results that quantify its real-time and robust perception and action behavior.

Index Terms—Neuromorphic computing, noisy spiking neural networks, robust object recognition, unsupervised learning

I. INTRODUCTION

Neuromorphic engineering is concerned with emulating the dynamics of biological neurons and synapses in silicon [1], as well as the organizing and computing principles of real neural processing systems. A main goal of these studies is to take advantage of the unique features of the brain, such as low-power consumption, massive parallelism, and low-latency processing, in order to perform efficient cognitive computations. In contrast to the classical von Neumann architecture of state-of-the-art digital computers, in neuromorphic hardware, memory and processing are co-localized in the synapses and neurons present in such devices. Previously developed neuromorphic computing hardware devices are using either asynchronous mixed-signal analog/digital [2]–[4] or purely digital circuits. Asynchronous neuromorphic systems process data and transmit signals only if and when events (spikes) are produced. Unlike purely digital approaches, the mixed-signal analog/digital approach has recently produced promising technologies for implementing computing architectures based on silicon neurons and synapses which exhibit dynamics that are similar to their biological counterparts [5].

Similar to the widely observed variance across biological neural networks [6], variability exists in all analog spiking neurons due to the unavoidable circuit noise and manufacturing mismatch, which has an especially strong effect on neurons' behavior when the circuits are working in the sub-threshold domain [7], [8]. This makes the implementation of desired behaviors and computations very challenging [9]. However, it has been observed in the brain, that despite the noisy neural substrate and environment, the behavior of biological spiking

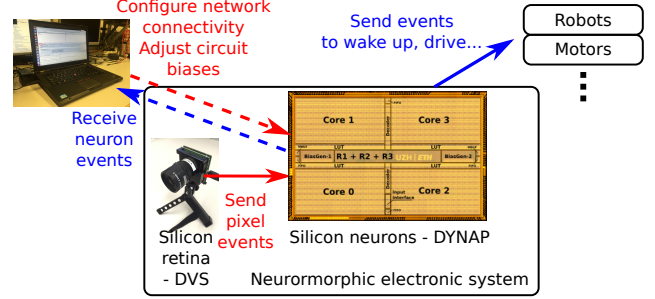


Fig. 1. The neuromorphic electronic system consists of DVS sensors and DYNAP chips. A computer is used to configure the sensors and chips, monitor their activities, or perform the learning algorithm. Once a network is set-up or the training is finished, the computer can be disconnected. The perception and processing on neuromorphic sensors and processors are ultra-low-power.

neural networks remains robust and reliable. It raises the question: can a proper network connectivity enable mixed-signal spiking neural systems to perform computations with high robustness despite their noise and mismatch?

To address this question we present a spiking neural architecture that comprises three biologically plausible mechanism geared toward variability reduction and robust computation: a dis-inhibition mechanism to reduce the effect of noise and enable robust feature detection; an up- and down-scaling mechanism of connections that leads to the spatial invariance during recognition; and a group of Neural State Machine (NSM) structures [10], [11] to perform unsupervised learning. In addition, we present an on-line spike-based learning rule for both excitatory and inhibitory plastic synapses that enables the network to memorize trained visual patterns and perform visual pattern recognition. Along with the innate ultra-low-power and even-driven computing paradigm of the mixed-signal analog/digital devices, the high robustness achieved on such hardware makes it possible to build an always-on and massively distributed system. Such a system reports an event or takes an action only if and when a visual stimulus is recognized as a pre-trained pattern; otherwise the network keeps running with an ultra-low power consumption. In Section II, we describe the architecture of the mixed-signal neuromorphic hardware. In Section III, we present the network architecture and the learning rule. In Section IV, we validate the network in a real-time real-world tasks.

II. METHODS

The setup used to train and configure the neuromorphic electronic system proposed is illustrated in Fig. 1. It is

composed of a neuromorphic vision sensor - the silicon retina chip [12] embedded in the Dynamic Vision Sensor (DVS) (DAVIS240C) and a group of neuromorphic processors - 4 Dynamic Neuromorphic Asynchronous Processor (DYNAP) chips [3]. The neuromorphic chip on both of them consumes mW ultra-low power consumption. The DVS emulates the dynamics of biological retina cells in silicon using mixed-signal analog/digital technologies. There are 240×180 pixels integrated on each chip. Each of them independently detects the illumination intensity change in a small area of the visual scene. It captures fast moving objects from the environment in a wide range of lighting conditions. The DYNAP integrates 1024 silicon neurons on each chip. Every neuron features 64 programmable incoming and 4096 programmable outgoing synapses. The neurons realize adaptive exponential integrate-and-fire dynamics with biologically realistic time constants [13]. The synapses are non-plastic but can be trained with a computer in the loop. The parameters of the neurons and their connectivity are configurable via on-chip analog bias generators and digital latches. The events generated by the DVS silicon retina chip are sent to the silicon neurons on-chip using the Address-Event Representation (AER) protocol. Although in this prototype setup the events are relayed by two Field Programmable Gate Array (FPGA) chips, for rapid prototyping and convenience, the connection between the chips can be made directly with parallel cables using the AER protocol, and therefore removing the need of power-hungry glue-logic.

In this setup we use a computer is used to configure the network, implement the on-line learning algorithm, and monitor the neural activity. Whenever a neuron sends a spike to the computer, the address of the neuron that emits the spike and the timing of when the spike is emitted is written into a ring buffer, for example, at position S . After that, the learning algorithm will read the ring buffer from the position $S - 1$ to $S - N$ to go through the history of received spikes. The parameter N determines the maximum average timing difference (i.e., the time window) between the events stored at $S - N$ and S (50ms in this implementation). For each pair of the newly arrived spikes and the stored spike, the algorithm checks a look-up table which stores the information of whether two neurons are connected. If there is a plastic synapse between this neuron pair, a synaptic weight change will be calculated according to the current weight and the timing difference between the emitted two spikes.

As for biological retinas, illumination intensity change is critical for the silicon retina to perceive the environment. To ensure that there is retinal activity also for static scenes biological vision systems resort to two types of eye movements: saccades [14] and microsaccades [15]. In our setup we simulate these eye movements by fixing the DVS, and moving the objects in front of it. The on-line learning algorithm monitors the generated events from DVS and simulates the motor signal that controls the saccade. The mean rate of the generated events represent the speed of the relative movement between the silicon retina and visual patterns. We use this

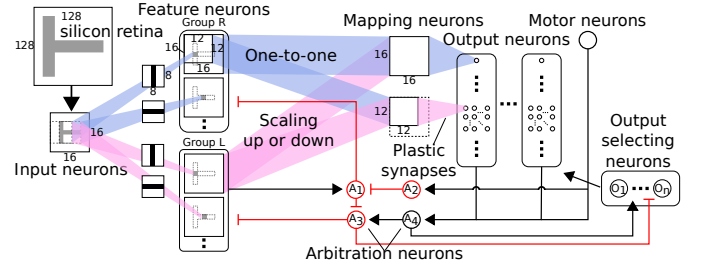


Fig. 2. The network architecture. The network is composed of an input layer, a feature layer, and an output layer. The neurons of these layers form a learning pathway and a recognition pathway. At the bottom of the figure, there is an arbitration mechanism implemented by neuron populations A_1 , A_2 , A_3 , and A_4 that controls which pathway will become dominant. In this paper, each of the populations A_1 , A_2 , A_3 , and A_4 is implemented by 4 silicon neurons to make their behavior robust.

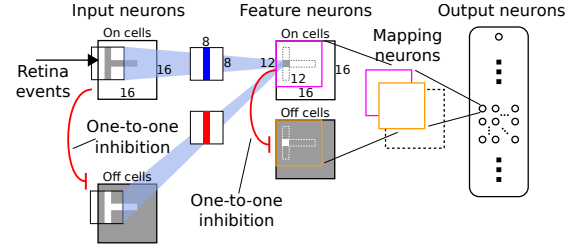


Fig. 3. Detailed structure of the feed-forward pathway from input neurons to output neurons. To simplify the illustration, here we only show the neurons of the recognition pathway while the learning pathway has the same on and off cells and the same connectivity principle. The on cells receive events from the previous layer, whereas the off cells tend to fire spontaneously due to a constantly supplied stimuli. However, they receive a strong inhibition from the on cells in a one-to-one manner.

average firing rate to generate events that represent the motor signal for the neuromorphic agent.

III. SPIKING NEURAL NETWORK ARCHITECTURE

The architecture we propose is illustrated in Fig. 2. It has a feed-forward structure consisting of three layers of neurons. It starts from 16×16 input layer neurons. We select the center 128×128 pixels on the silicon retina and down-sample them to match the 16×16 input neurons. A feature layer succeeds the input layer. Due to the limited number of neurons on the prototype chip, here we choose to configure the feature neurons to detect two features: horizontal bars and vertical bars. Neurons in the feature layers receive events from the input layer in a convolutional manner with kernel size 8×8 and stride 1. The output layer consists of multiple groups of neurons, each of which learns to recognize a different visual pattern. Which group will learn a new pattern is controlled by a group of output selecting neurons.

Neurons in the network form two pathways. One pathway is for learning and the other one is for inference, denoted as the blue and pink colors in Fig. 2 respectively. Each pathway has its own group of feature neurons: neurons L and R for the learning and recognition pathways respectively. For recognition, the network is connected in a convolutional manner with different kernel sizes. For learning on the contrary, there

is an up- and down-scaling mechanism through which all the feature neurons are connected to each output neuron and the spatial information is maintained. However, the connections are not direct. There is a group of mapping neurons that relay the spikes from the feature neurons to the output neurons. The learning takes place at the plastic synapses between the mapping neurons and the output neurons. The detailed structure and the learning rule are presented as following:

a) *Dis-inhibition mechanism*: The neurons that are connected to the next layer, namely the input neurons and feature neurons, consist of on and off cells. The on and off cells are excitatory and inhibitory neurons respectively. They cooperatively perform a dis-inhibition mechanism that makes the recognition robust to the noise from inputs and the intrinsic mismatch of analog circuits. The connectivity of the on and off cells is illustrated in Fig. 3. The on and off cells have the same connectivity manner to the next layer as what is shown in Fig. 2. However, only the on cells receive event from the previous layer whereas the off cells are not connected to the previous layer but receive a one-to-one inhibition from the on cells. In addition to the inhibition, the off cells receives excitation from a constantly supplied stimuli. In this way, the on and off cells always exhibit opposite activities.

b) *Arbitration mechanism*: Controlled by the arbitration neurons A_1, A_2, A_3 , and A_4 at the bottom of Fig. 2, the two pathways take turns to perform the learning and recognition of visual patterns. The learning pathway gains and maintains the opportunity to start and finish the learning as follows. Within the learning pathway, given a visual pattern, the feature neurons L will start to fire only if the neurons A_3 and A_4 are silent. This happens when the following condition is satisfied: the output layer does not recognize the visual pattern, and the simulated eye-movement motor neuron stops to fire. This implies that the visual pattern becomes stable at the input layer. If the neurons A_3 and A_4 are firing, they will also reset the output selecting neurons, that send spikes to the output neurons in order to start the learning. Thus, it ensures that if the visual pattern is already learned, the learning phase will not start again. Once the neuron group L start to fire, they will excite the neurons A_1 to ensure that the neuron group R is silent during the learning process. Because A_1 inhibits the neurons A_3 , no more inhibition can be given to the neuron group L . The arbitration neurons A_2 inhibit A_1 when the motor neurons fire. It ensures that after learning, when the silicon retina or the visual pattern starts to move, L will be inhibited by A_4 .

c) *Spike-based learning rule*: The learning happens on the plastic synapses between the mapping neurons and the output neurons. It is performed on the computer in real time with the neuromorphic chips in the loop. The learning of a plastic synapse is triggered by every spike of the post-synaptic neuron. The learning rule for excitatory synapses is Hebbian-like. It can be described as:

$$w_{ij}(t) = \max((w_{ij}(t) + \alpha \Delta w_{ij}(t) \cdot S_j(t), 0) \quad (1)$$

$$\Delta w_{ij}(t) = \beta \sum_{t-\Delta t < t' < t} S_i(t') - 1 \quad (2)$$

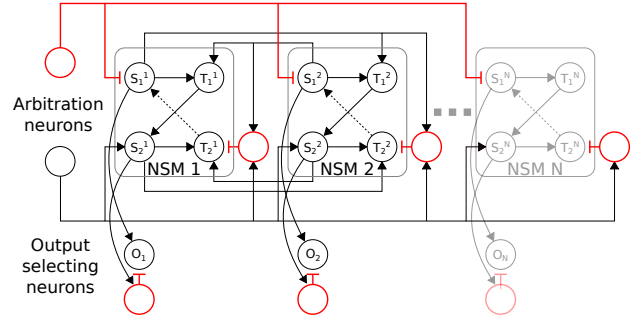


Fig. 4. Network structure that implements the automatic selection of output neurons during training. The NSM structure is modified from [10]. Here within each NSM, the connections between S_1 and T_1 , and S_2 and T_2 neurons, as well as the winner-take-all structure built upon S_1 and S_2 neurons are simplified for visualization. It only shows the connectivity between NSM 1 and NSM 2 as an example. Every pair of NSMs has the same connectivity. The dotted line represents plastic synapses. Each circle represents a neuron population of 4 neurons that has the same connectivity.

where i and j denote the indexes of the pre-synaptic and post-synaptic neurons respectively. $\alpha > 0$ if the pre-synaptic neuron i is an excitatory neuron, and $\alpha < 0$ if neuron i is an inhibitory neuron. β denotes a learning rate. $w_{ij}(t_0) = 0$, where t_0 is when the learning starts. $S_i(t) = \sum_k \delta(t - t_k^i)$ denotes whether there is a spike generated by neuron i at time t , where t_k^i represents the timing of spikes and $\delta(x) = 1$ when $x = 0$ otherwise $\delta(x) = 0$. Once the weight change of a synapse accumulated to be larger than a predefined threshold, the new synaptic weight is updated on the chips. Otherwise, it will stay as the previous state. Thus, the learned weights on chip can be considered as a binarized weight matrix. When a neuron is activated longer than a threshold period (e.g., > 80 ms), the learning of its incoming synapses will stop and learning is finished. After learning, the synapses are able to memorize the pre-synaptic activity that was shown to them. This learning rule ensures that during learning, only the firing rate but not the firing-or-not activity of the post-synaptic neurons will be changed according to the update of synapse weights. This makes the learning more robust and predictable.

d) *Unsupervised learning*: The output selecting neurons are controlled by a plastic winner-take-all mechanism as illustrated in Fig. 4. It is implemented with multiple NSM structures described in [10]. The network connectivity ensures that anytime only one NSM can stay at state S_1 and all the others are at state S_2 . Once a NSM wins the competition, it will self-maintain its activity and the synapses that are coming into its S_1 neurons will decrease their weights. Therefore, next time, this state will not be selected as the winner again. In detail, each NSM could stay at either state S_1 or S_2 , meaning that the neurons denoted as S_1 or S_2 are firing respectively. The neurons T_1 and T_2 will fire only if the two groups of in-coming synapses both receive spikes. The interconnected NSMs carry out a winner-take-all behavior, because they always compete with each other. Anytime only one NSM can stay at state S_1 , while the others are at state S_2 . After the competition, S_1 neurons of the winner NSM will decrease their

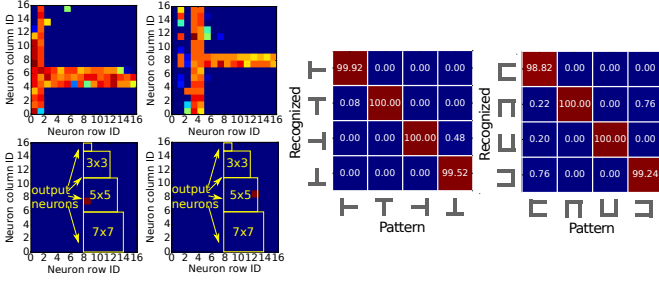


Fig. 5. A 2-dimensional histogram of the neural firing activity of the network in a duration of 20 ms. Left: recorded neural activity of the input and output layer. Right: performance of the network in two simple tasks. Each element in the confusion matrix represents the percentage of spikes that belong to each of the 4 groups of output neurons compared to the total number of spikes when a pattern is shown to the neuromorphic system.

in-coming synaptic weights according to the same learning rule as discussed, except that initially $w_{ij}(t_0) > 0$. Next time this NSM will not become the winner again. After learning a visual pattern, the arbitration neurons A_3 and A_4 send spikes to reset all the NSMs to state S_2 . This prepares the learning of a new pattern. Because each group of S_1 neurons excites a group of output neurons, different groups of the output neurons will take turns to be activated for different visual patterns. This enables the learning to be performed in an unsupervised manner.

IV. EXPERIMENTAL RESULTS

In the following experiments, objects either move with a speed in the range of [60, 1.2k]pps (pixels per second), or constantly shake in front of the silicon retina. Here the pps represents how fast the perceived visual pattern moves on the silicon retina. The value 60pps is the measured minimum speed, below which the generated events are not strong enough for the network to recognize. The value 1.2kpps is the measured maximum speed, above which the time constants of the neurons and synapses are not fast enough to distinguish the visual patterns.

a) Performance and robustness: Because of the limited number of neurons on the prototype chips, every time we train the network with 4 different visual patterns composed of horizontal and vertical bars, namely a 'T' shape or a 'U' shape symbol with different directions. Due to the up- and down-scaling mechanism within the connections, the position and distance information of the input pattern can be reserved during recognition and is represented by the output neurons. To test the recognition performance of the network after learning, each pattern is shown to the silicon retina in a duration of 1000s. Fig. 5 shows that the network is robust to the noise and mismatch of neuromorphic devices and the environment. We choose to use these simple patterns due to the limited number of silicon neurons on the prototype chip. The same principle can be scaled up for complex patterns and features.

b) Power Consumption: Without incoming events, the spiking neural network architecture implemented on neuromorphic devices only consumes static power. The static power dissipation is 945 μ W for the DYNAP chip [16]. The main

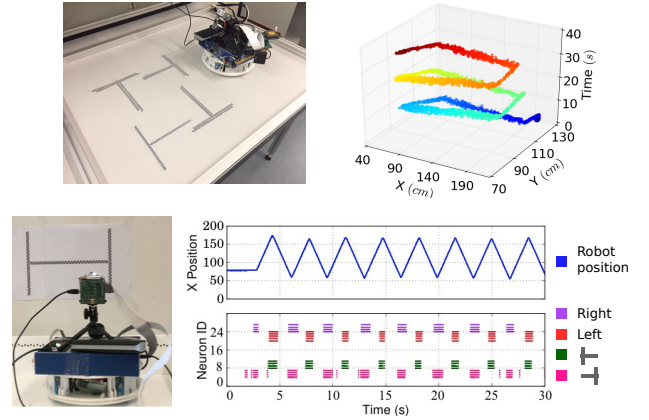


Fig. 6. Real-time real-world experimental results on an omni-directional robot. The silicon retina and neuromorphic processors are mounted on the robot. The motor neurons on chip send out events to drive the robot's motors to go either forward, backward, left or right. Up: Robot navigates in the arena by recognizing the surface markings on the ground. Down: The robot keeps changing between left and right movements.

source of the dynamic power consumption is due to neurons firing and spikes generation. For generating every spike, the neurons use 2.8pJ [16]. The average mean firing rate of the neurons is 41.76Hz when no visual pattern is given, and 55.73Hz when visual patterns are given to the input layer. According to [12], the silicon retina chip consumes less than 5mW when the input activity is not very intensive. Thus the total power consumption of this system is estimated to be less than a few mW.

c) Real-time real-world tasks: The time cost for recognizing each pattern depends on the properties of neurons and synapses (i.e., time constant, firing threshold, and the length of Post-Synaptic Potentials (PSPs), etc.) on the feed-forward pathway. Each neuron takes time to integrate evidence from the asynchronous input. This delayed time, however, enables a temporal invariance in the range from tens of to several ms. It is essential for real-world tasks where the robot or object is not moving smoothly. We tested two scenarios in which the robot moved in an arena, continuously detecting and recognizing patterns that are associated with actions accordingly. Fig. 6 illustrates the neural activity and the robot trajectory. It shows that our spiking neural network architecture performs well in the real-world noisy environment.

V. CONCLUSION

We presented a spiking neural network architecture and an on-line off-chip training method that enable robust learning and recognition of visual patterns in noisy spiking neural networks and noisy environment. We demonstrated a pattern recognition tasks in a closed-loop system composed of asynchronous neuromorphic sensors, processors, and robotic agents. In addition to solving practical engineering problems, the proposed network architecture might shed light on how the neurons are organized to carry out robust computation in biological networks.

REFERENCES

- [1] C. Mead, "Neuromorphic electronic systems," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1629–36, 1990.
- [2] N. Qiao, H. Mostafa, F. Corradi, M. Osswald, F. Stefanini, D. Sumislawska, and G. Indiveri, "A re-configurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128k synapses," *Frontiers in Neuroscience*, vol. 9, no. 141, pp. 1–17, 2015.
- [3] S. Moradi, N. Qiao, F. Stefanini, and G. Indiveri, "A scalable multicore architecture with heterogeneous memory structures for dynamic neuromorphic asynchronous processors (DYNAPs)," *Biomedical Circuits and Systems, IEEE Transactions on*, pp. 1–17, 2017.
- [4] "Brain-inspired multiscale computation in neuromorphic hybrid systems (BrainScaleS)," FP7 269921 EU Grant, 2011–2015.
- [5] E. Chicca, F. Stefanini, C. Bartolozzi, and G. Indiveri, "Neuromorphic electronic circuits for building autonomous cognitive systems," *Proceedings of the IEEE*, vol. 102, no. 9, pp. 1367–1388, 9 2014.
- [6] A. A. Faisal, L. P. Selen, and D. M. Wolpert, "Noise in the nervous system," *Nature reviews neuroscience*, vol. 9, no. 4, pp. 292–303, 2008.
- [7] S.-C. Liu, J. Kramer, G. Indiveri, T. Delbruck, and R. Douglas, *Analog VLSI: Circuits and Principles*. MIT Press, 2002.
- [8] J. Binas, G. Indiveri, and M. Pfeiffer, "Spiking analog VLSI neuron assemblies as constraint satisfaction problem solvers," in *International Symposium on Circuits and Systems, (ISCAS), 2016*. IEEE, 2016, pp. 2094–2097.
- [9] T. Pfeil, A. Grübl, S. Jeltsch, E. Müller, P. Müller, M. Petrovici, M. Schmuker, D. Brüderle, J. Schemmel, and K. Meier, "Six networks on a universal neuromorphic computing substrate," *Frontiers in Neuroscience*, vol. 7, 2013.
- [10] D. Liang and G. Indiveri, "Robust state-dependent computation in neuromorphic electronic systems," in *Biomedical Circuits and Systems Conference, (BioCAS), 2017*. IEEE, Oct. 2017, pp. 108–111.
- [11] E. Neftci, J. Binas, U. Rutishauser, E. Chicca, G. Indiveri, and R. Douglas, "Synthesizing cognition in neuromorphic electronic systems," *Proceedings of the National Academy of Sciences*, vol. 110, no. 37, pp. E3468–E3476, 2013.
- [12] C. Brandli, R. Berner, M. Yang, S.-C. Liu, and T. Delbruck, "A 240×180 130 dB $3 \mu\text{s}$ latency global shutter spatiotemporal vision sensor," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 10, pp. 2333–2341, 2014.
- [13] G. Indiveri, E. Chicca, and R. Douglas, "A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity," *IEEE Transactions on Neural Networks*, vol. 17, no. 1, pp. 211–221, Jan 2006.
- [14] H. Deubel and W. X. Schneider, "Saccade target selection and object recognition: Evidence for a common attentional mechanism," *Vision research*, vol. 36, no. 12, pp. 1827–1837, 1996.
- [15] S. Martinez-Conde, S. L. Macknik, and D. H. Hubel, "The role of fixational eye movements in visual perception," *Nature Reviews Neuroscience*, vol. 5, no. 3, p. 229, 2004.
- [16] G. Indiveri, F. Corradi, and N. Qiao, "Neuromorphic architectures for spiking deep neural networks," in *Electron Devices Meeting (IEDM), 2015 IEEE International*. IEEE, Dec. 2015, pp. 4.2.1–4.2.14.