A Spike-Based Neuromorphic Stereo Architecture for Active Vision

Nicoletta Risi, Alessandro Aimar, Elisa Donati, Sergio Solinas, Giacomo Indiveri
Institute of Neuroinformatics, University of Zurich and ETH Zurich

Abstract

The problem of finding stereo correspondences in binocular vision is solved effortlessly in nature and yet is still a critical bottleneck for artificial machine vision systems. As temporal information is a crucial feature in this process, the advent of event-based vision sensors and dedicated event-based processors promises to offer an effective approach to solve stereo-matching. Indeed, event-based neuromorphic hardware provides an optimal substrate for biologically-inspired, fast, asynchronous computation, that can make explicit use of precise temporal coincidences. Here we present an event-based stereo-vision system that fully leverages the advantages of brain-inspired neuromorphic computing hardware by interfacing event-based vision sensors to an event-based mixed-signal analog/digital neuromorphic processor. We describe the multi-chip sensory-processing setup developed and demonstrate a proof of concept implementation of cooperative stereo-matching that can be used to build brain-inspired active vision systems.

1 Introduction

Biological and artificial binocular visual systems rely on stereo-vision processes to merge the visual information streams. This implies solving the stereo-matching problem, i.e. finding correspondent points in two slightly shifted views (Cumming and Parker, 1997). Typical applications in robotics that can benefit from stereo vision include navigation in unknown environment, object manipulation and grasping. However, current machine-vision approaches still lag behind their biological counterpart mainly in terms of bandwidth and power consumption (Steffen et al., 2019; Tippetts et al., 2016). Classical methods are based on frame-based vision sensors. The main challenges of frame-based algorithms are spatial redundancy and temporal information loss due to the intrinsic nature of fixed-rate processing. This affects latency, throughput, and power consumption, making frame-based approaches difficult to integrate into mobile platforms. Biological systems, on the other hand, seem to efficiently solve the stereo-matching problem by using space-variant and asynchronous space-time sampling (Polimeni and SchwartzÝ, 2001). Space-variant resolution refers to a non-uniform distribution of retinal photoreceptors, with higher density in the center (fovea) and a decreasing density towards the periphery. Asynchronous instead refers to event-driven, self-timed sensing, and processing. Therefore a massively parallel, asynchronous, event-based chain, from sensing to processing, seems to be a promising solution for more robust and efficient architectures of stereo vision.
In this context, neuromorphic hardware has proven to be an effective substrate (Indiveri, Corradi, and Qiao, 2015). To date, the emerging field of event-based stereo-vision has shown successful approaches that interface Spiking Neural Networks (SNNs) with neuromorphic event-based sensors in order to build real-time event-based visual processing systems (Mahowald, 1994a; Osswald et al., 2017). Inspired by the retinal ganglion cells, the neuromorphic vision sensors broadcast information, independently for all the pixels, only in response to significant changes in illumination, which results in a low-power, low-latency, event-driven and sparse input stream (Posch, Matolin, and Wohlgenannt, 2010; Berner et al., 2013; Lichtsteiner, Posch, and Delbruck, 2008). Spiking neurons, in turn, can process signals using temporal information and therefore, can take full advantage of an event-based input stream to solve the stereo-matching problem.

However, although several biologically-inspired implementations of stereo vision (Osswald et al., 2017; Mahowald, 1994b; Kaiser et al., 2018; Dikov et al., 2017; Piatkowska, Belbachir, and Gelautz, 2013; Piatkowska, Kogler, et al., 2017) have extensively been explored,

Appearing in Proceedings of the Workshop on Robust Artificial Intelligence for Neurorobotics (RAI-NR) 2019, University of Edinburgh, Edinburgh, United Kingdom. Copyright 2019 by the authors.

only a few solutions fully exploit the advantages of parallel computation, with an end-to-end neuromorphic architecture that can replace traditional Von Neumann architectures. In (Osswald et al., 2017), a proof of concept of the Marr and Poggio cooperative network (Marr and Poggio, 1976; Marr and Poggio, 1977; Marr and Poggio, 1979) is implemented on a reconfigurable online learning spiking neuromorphic processor (ROLLS, (Qiao et al., 2015)); Dikov et al., (2017) propose an alternative implementation that reduces false positive stereo-matches on a scalable neuromorphic computing platform, SpiNNaker (Furber et al., 2014); more recently, Andreopoulos et al., (2018) introduced a neuromorphic architecture of 3D vision using multiple TrueNorth processors (Sawada et al., 2016).

Following up on the work from (Osswald et al., 2017), in this paper we introduce a proof of concept for a fully asynchronous, brain-inspired implementation of the cooperative model for stereo-vision. To this end, we replaced the ROLLS chip used in (Osswald et al., 2017) with a multicore mixed-signal Very Large Scale Integration (VLSI) architecture, and we built an event-based interface that supports inputs from a pair of neuromorphic sensors. Compared to the previous work, the current implementation features a more robust, biologically-inspired coincidence detection mechanism implemented directly on chip and allows to scale up the size of the network. Here we describe the interface implemented on FPGA, and we show a proof of concept of the cooperative network on the scalable neuromorphic processor.

2 Methods

The stereo-vision architecture introduced here combines two event-based sensors, the Dynamic and Active Pixel Vision Sensor (DAVIS) (Berner et al., 2013), and a VLSI analog/digital Dynamic Neuromorphic Asynchronous Processor (DYNAP) (Moradi et al., 2018). As a prototype, we designed the interface between sensing and processing on a dedicated Field Programmable Gate Array (FPGA) device (Xilinx Kintex-7 FPGA on the OpalKelly XEM7360).

2.1 Event-based Sensing

As opposed to classical frame-based cameras, event-based sensor encodes information with lower latency and redundancy. Inspired by the biological photoreceptors, the neuromorphic pixels operate independently and send out asynchronous events in response to significant changes in illumination using an event-based data protocol Address Event Representation (AER) (Deiss, Douglas, and Whatley, 1998). Overall, this results in a fast acquisition with low latency and high temporal resolution (approximately 10 μs). Compared to the original DVS (Lichtsteiner, Posch, and Delbruck, 2008), the DAVIS sensor features a higher spatial resolution (240×180) and adds an APS readout.
In the proposed architecture, the two DAVIS sensors are mounted on a stereo-setup (see Fig. 1) and separated by a baseline distance of about 6 cm, which is similar to the pupillary distance of humans ($\approx$ 65 mm). Events are sent separately from both retinas to an FPGA using the AER protocol.

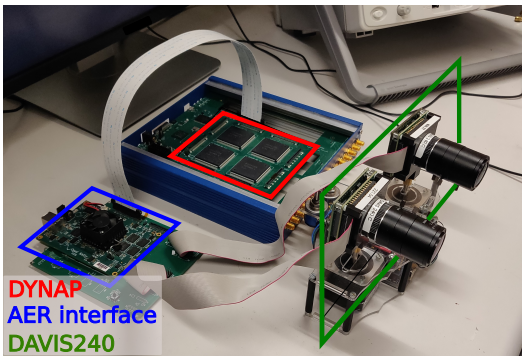

Fig. 1: The neuromorphic stereo-vision setup

2.2 Sensing-Processing FPGA interface

Fig. 2 shows the main modules of the sensing-processing interface. The communication to/from the FPGA is based on a 4-phase handshake protocol, handled by the Handshake Receiver (HSR). The Metastability Synchronizer (MSC) prevents metastability issues using a chain of two Flip-Flops on the input signals. For the correct functioning of the network, a pre-processing element (PEL) reduces the input resolution to a 32×32 array to redirect the AER events to the destination chip on the neuromorphic processor. Specifically, inspired by the biological space-variant and asynchronous space-time sampling, two output modalities are supported: *spatial downsampling*, which allows for coarse spatial resolution but with an extended field of view, and *fovea selection*, which forwards only the central 32×32 array of pixels of the neuromorphic sensors. The pre-processed events are thus forwarded to a small FIFO with 8 entries, in charge of absorbing the pipeline stall due to the successive muxing stage. The DAVIS Input Selector (DIS) module muxes the data using a round-robin scheme and forwards them to the Handshake Sender (HSS), which handles the output handshake with the neuromorphic processor.

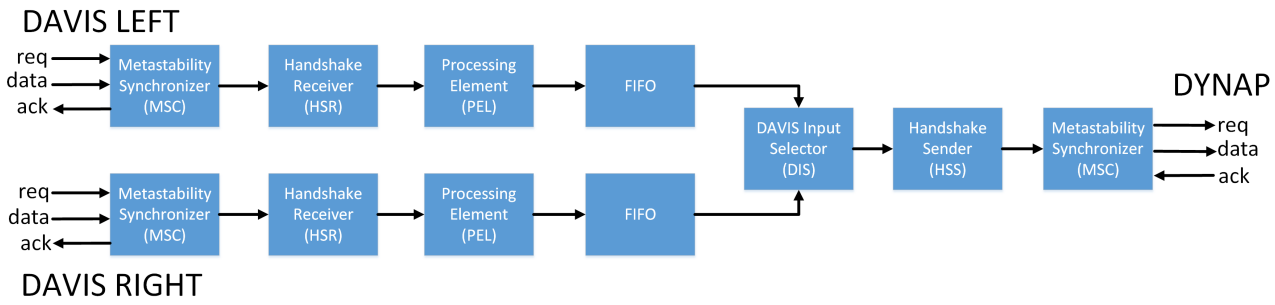

Fig. 2: Overview of the implemented FPGA architecture

2.3 Event-based Processing

As the natural interface of event-based sensing is event-based processing, here we implemented the cooperative stereo network using a multi-core asynchronous mixed-signal neuromorphic processor (DYNAP), fabricated using standard 0.18 μm1P6M CMOS technology (Moradi et al., 2018). Each core comprises 256 adaptive exponential integrate-and-fire (AEI&F) silicon neurons that emulate the biophysics of their biological counterpart, and four different dedicated analog circuits that mimic fast and slow excitatory/inhibitory synapse types (Chicca et al., 2014). Each neuron has a Content Addressable Memory (CAM) block, containing 64 programmable entries allowing to customize the on-chip connectivity. A fully asynchronous inter-core and inter-chip routing architecture allows flexible connectivity with microsecond precision under heavy systems loads. Digital peripheral asynchronous input/output logic circuits are used to receive and transmit spikes via an AER communication protocol, analogous to the one used for the event-based input stream. As a result, the proposed implementation leads to a prototype for a fully asynchronous pipeline of event-based stereo-vision.

2.4 The Spiking Neural Network Model

The SNN implemented on the DYNAP follows the structure presented in (Osswald et al., 2017). It consists of three neuronal populations: the retina, the coincidence detectors, and the disparity detectors. Pairs of epipolar retina cells, here implemented with left and right DAVIS, project excitatory connections to the coincidence detector layer which encodes temporal coincidences among pairs of interocular events. However, coincidence neurons are also sensitive to false matches. For instance, two different stimuli moving simultaneously but at different depths would erroneously be perceived as a true target, thus leading to the encoding of wrong disparities. Therefore, temporal information is crucial but clearly not sufficient to correctly solve the stereo-matching problem. This ambiguity is reduced in the disparity layer by means of recurrent inhibition, and excitatory and inhibitory projections from the coincidence layer, which results in a Winner-Takes-All (WTA) mechanism that implements the matching constraints of cooperative algorithms (Mahowald, 1994b; Marr and Poggio, 1976). Each coincidence and disparity neuron is assigned a triplet of coordinates, a horziontal and vertical cyclopean position (x, y) and a disparity value (d), which determines the neuron representation of a location in the 3D space. Therefore, disparity spikes provide evidence for a potential target in the correspondent 3D position.

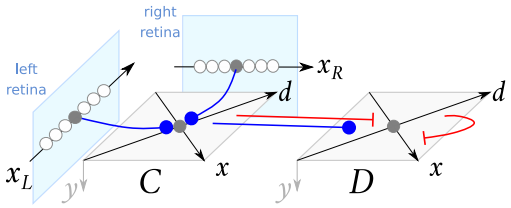

Fig. 3: One layer of the Spiking Neural Network: one epipolar line for each retina projects to the corresponding layers of coindcidence (C) and disparity detectors (D). Excitatory (blue) and inhibitory (red) connections are shown. Adapted from (Osswald et al., 2017).

2.5 Neuromorphic Hardware Implementation

The entire pipeline of visual information processing was designed to be a scalable neuromorphic architecture. Here we built a proof of concept where the field of view is restricted to one epipolar line of 32 cells for each retina, which narrows the coincidence and disparity layers down to a two-dimensional 32×32 population (see simplified scheme in Fig. 3). Since temporal coincidence detection is a key component of our model, we carefully emulated and further optimized the low power method exploited by biological brains. We detect precise coincidence of spike events by combining the mechanism of supra-linear, dendritic summation of synaptic events with slow and

fast synaptic time constants. As in biological brains, AMPA synaptic currents can boost the effect of slow NMDA synapses when both synaptic inputs are close in time (González, 2011). Coincidence detectors are emulated on chip exploiting the nonlinear properties of the dedicated analog synapse circuit block which mimics the biological NMDA voltage-gating dynamics. Specifically, each coincidence detector is connected to one of the corresponding input retina cells via the slow (NMDA-like) synapse and to the other one via the fast (AMPA-like) synapse circuit block. An example of coincidence detection emulated on chip is shown in Fig. 4 This results in a more robust coincidence detection mechanism, less sensitive to device mismatch, which is a crucial feature of subthreshold mixed-signal neuromorphic processors.

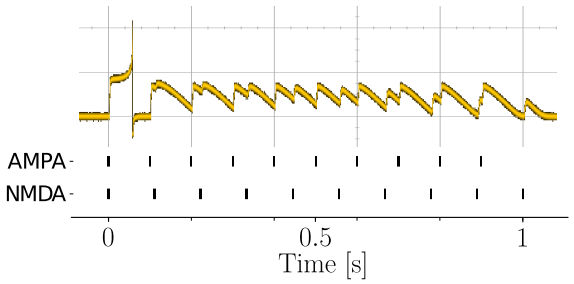

Fig. 4: Emulation of coincidence detection: recorded membrane potential of a silicon neuron that receives two input spike trains through AMPA-like and NMDA-like synaptic circuits. The silicon neuron responds only to temporal coincidences within a time window of 10 ms (first pair of input spikes).

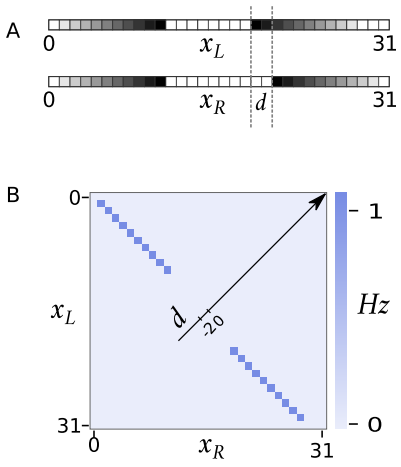

Fig. 5: Simulated stimuli: temporal image representation, with black pixels representing more recent events (A); theoretically predicted output (ground-truth) of the disparity layer given the input in A (B).

3 Results

3.1 Control Stimulus

Prior to validation with a real-time scenario, here we assess the performance of the event-based neuromorphic implementation of stereo-matching with a synthetic ground-truth. Specifically, an input spike train generated on FPGA is used to simulate two input event streams, which mimic two stimuli moving in opposite directions at two different but constant disparities. The correspondent temporal image representation is shown in Fig. 5A. The correspondent ground truth of stereo-matches is shown in Fig. 5B.

3.2 Stereo Matching

To test the network performance on the hardware setup, we monitored the activity of the network across 100 presentations of the input stimuli (see Fig. 6). The coincidence detectors successfully detect the temporal matches, i.e. an action potential arises only when the input events from the retina cells are coincident in time (compare inserts A and B). However, coincidence detectors still respond to false targets which would correspond to stimuli moving at constant cyclopean positions but across different depths. This ambiguity is instead solved in the disparity layer.

In order to quantify the stereo-matching performance, we computed the PCM (percentage of correct matches) (Osswald et al., 2017) over windows of 100 ms, averaged across trials. As shown in Fig. 7, the PCM of disparity detectors (D) is larger than the one of coincidence detectors (C), suggesting that the WTA is effectively reducing the number of false matches from the coincidence to the disparity layer.

4 Discussion

Here we describe a proof of concept in hardware of the cooperative network presented in (Osswald et al., 2017). Thanks to the balance of feedforward excitation/inhibition and recurrent inhibition, the disparity detectors exhibit an average PCM that stays close to 1 and always above the PCM of coincidence detectors, thus successfully reducing the number of false matches. Having tested the network dynamics with a synthetic ground-truth, the next step for validating the stereo-network architecture will be to stimulate the silicon neurons with a real-time input stream from the neuromorphic sensors. Moreover, although not yet exploited in this proof-of-concept implementation, the dual output modality supported by the pre-processing element (PEL) of the AER interface would allow optimizing computational resources on the neuromorphic proces-

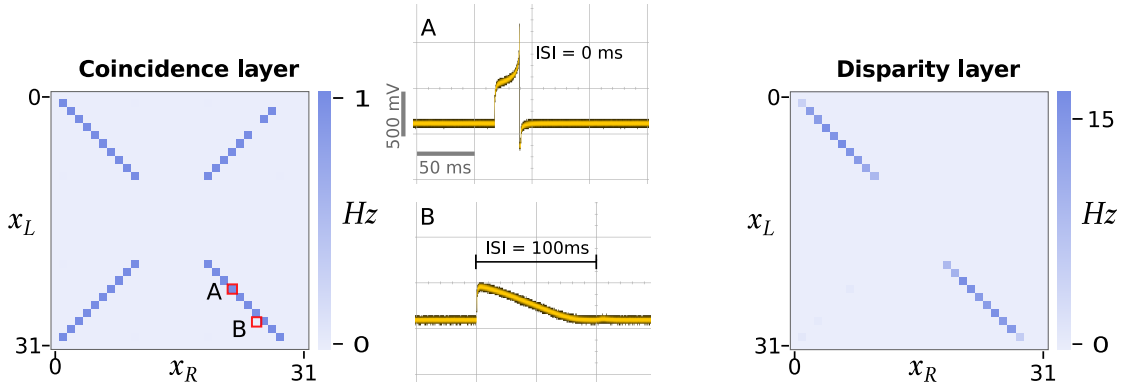

Fig. 6: Emulation of one layer of the stereo network on DYNAP: mean firing rate across 100 trials of both coincidence (left panel) and disparity (right panel) layer. Only input pair of events with ISI=0 (Interstimulus Spike Interval) give rise to an action potential in the coincidence layer (compare the recorded membrane potential of two silicon neurons in inserts A and B).

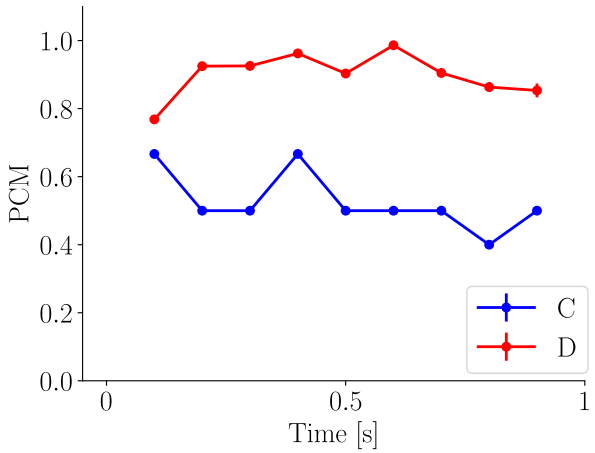

Fig. 7: Percentage of Correct Matches (PCM) for the coincidence layer (C) and the disparity layer (D): average measure across 100 trials.

sor via multiplexing. Indeed, the parallel AER connection between the DYNAP and the FPGA supports both input/output stream. Thus, by rerouting spikes from the chip to the AER interface, it would be possible to select the processing modality of the PEL, able to drive an active sensory-processing system. Inspired by the biological space-variant and asynchronous space-time sampling, this approach could lead to more effective behaviors of stereo vision in robotics, e.g. when having to cope with constraint resources, such as computational cost and power consumption.

5 Conclusion

We presented a mixed-signal neuromorphic architecture for event-based stereo-vision. The proposed implementation aims at fully leveraging the advantages of brain-inspired parallel computation by interfacing neuromorphic sensing and processing together. An event-based digital interface was built to handle the AER handshake between the silicon retina and the DYNAP chip while preserving the temporal information of the input events streams. The long-term goal is to include the finalized architecture as a brain-inspired sensory building block of active stereo-vision system.

Acknowledgements

The authors would like to thank Marc Osswald and Dmitrii Zendrikov for the fruitful discussions, and Chenxi Wu for contributing to the AER interface design. This work is supported by the ERC Grant "NeuroAgents" (724295).

References

Cumming, B. G. and A. J. Parker (1997). "Responses of primary visual cortical neurons to binocular disparity without depth perception". In: *Nature*. ISSN: 00280836.

Steffen, L. et al. (2019). "Neuromorphic Stereo Vision: a Survey of Bio-inspired Sensors and Algorithms". In: *Frontiers in Neurorobotics* 13, p. 28.

Tippetts, B., D. J. Lee, K. Lillywhite, and J. Archibald (2016). "Review of stereo vision algorithms and their suitability for resource-limited systems". In: *Journal of Real-Time Image Processing* 11.1, pp. 5–25.

Polimeni, J. and E. L. SchwartzÝ (2001). "Space-time adaptive image representations: Data Structures, Hardware and Algorithms". In: *Defining a Motion Imagery Research and Development Program workshop, Virginia's Center for Innovative Technology: published Nov*. Vol. 20. Citeseer.

Indiveri, G., F. Corradi, and N. Qiao (2015). "Neuromorphic Architectures for Spiking Deep Neural Networks". In: *Electron Devices Meeting (IEDM), 2015 IEEE International*. IEEE, pp. 4.2.1–4.2.14.

Mahowald, M. (1994a). "Analog VLSI chip for stereocorrespondence". In: *International Symposium on Circuits and Systems, (ISCAS), 1994*. Vol. 6, pp. 347–350.

Osswald, M., S.-H. Ieng, R. Benosman, and G. Indiveri (2017). "A spiking neural network model of 3Dperception for event-based neuromorphic stereo vision systems". In: *Scientific reports* 7.40703, pp. 1–11.

Posch, C., D. Matolin, and R. Wohlgenannt (2010). "A QVGA 143 dB dynamic range asynchronous address-event PWM dynamic image sensor with lossless pixel-level video compression". In: *International Solid-State Circuits Conference Digest of Technical Papers, ISSCC 2010*. IEEE, pp. 400–401.

Berner, R., C. Brandli, M. Yang, S.-C. Liu, and T. Delbruck (2013). "A 240 × 180 10mW 12μs latency sparse-output vision sensor for mobile applications". In: *2013 Symposium on VLSI Circuits*. IEEE, pp. C186–C187.

Lichtsteiner, P., C. Posch, and T. Delbruck (2008). "A 128x128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor". In: *IEEE Journal of Solid-State Circuits* 43.2, pp. 566–576. ISSN: 0018-9200.

Mahowald, M. (1994b). *An Analog VLSI System for Stereoscopic Vision*. Boston, MA: Kluwer.

Kaiser, J. et al. (2018). "Microsaccades for neuromorphic stereo vision". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. ISBN: 9783030014179.

Dikov, G., M. Firouzi, F. Röhrbein, J. Conradt, and C. Richter (2017). "Spiking Cooperative Stereo-Matching at 2 ms Latency with Neuromorphic Hardware". In: ISBN: 978-3-319-63537-8.

Piatkowska, E., A. Belbachir, and M. Gelautz (2013). "Asynchronous Stereo Vision for Event-Driven Dynamic Stereo Sensor Using an Adaptive Cooperative Approach". In: *2013 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 45–50.

Piatkowska, E., J. Kogler, N. Belbachir, and M. Gelautz (2017). "Improved cooperative stereo matching for dynamic vision sensors with ground truth evaluation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 53–60.

Marr, D. and T. Poggio (1976). "Cooperative computation of stereo disparity". In: *Science* 194.4262, pp. 283–287. ISSN: 0036-8075, 1095-9203.

Marr, D. and T. Poggio (1977). *A Theory of Human Stereo Vision*. Tech. rep. MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTELLIGENCE LAB.

Marr, D. and T. Poggio (1979). "A Computational Theory of Human Stereo Vision". In: *Proceedings of the Royal Society of London. Series B, Biological Sciences* 204.1156, pp. 301–328. ISSN: 0080-4649.

Qiao, N. et al. (2015). "A Reconfigurable On-line Learning Spiking Neuromorphic Processor comprising 256 neurons and 128K synapses". In: *Frontiers in Neuroscience* 9.141, pp. 1–17.

Furber, S., F. Galluppi, S. Temple, and L. Plana (2014). "The SpiNNaker Project". In: *Proceedings of the IEEE* 102.5, pp. 652–665.

Andreopoulos, A., H. J. Kashyap, T. K. Nayak, A. Amir, and M. D. Flickner (2018). *A Low Power, High Throughput, Fully Event-Based Stereo System*. Tech. rep.

Sawada, J. et al. (2016). "Truenorth ecosystem for brain-inspired computing: scalable systems, software, and applications". In: *SC'16: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. IEEE, pp. 130–141.

Moradi, S., N. Qiao, F. Stefanini, and G. Indiveri (2018). "A Scalable Multicore Architecture With Heterogeneous Memory Structures for Dynamic Neuromorphic Asynchronous Processors (DYNAPs)". In: *Biomedical Circuits and Systems, IEEE Transactions on* 12.1, pp. 106–122.

Deiss, S., R. Douglas, and A. Whatley (1998). "A Pulse-Coded Communications Infrastructure for Neuromorphic Systems". In: *Pulsed Neural Networks*. Ed. by W. Maass and C. Bishop. MIT Press. Chap. 6, pp. 157–78.

Chicca, E., F. Stefanini, C. Bartolozzi, and G. Indiveri (2014). "Neuromorphic electronic circuits for building autonomous cognitive systems". In: *Proceedings of the IEEE* 102.9, pp. 1367–1388. ISSN: 0018-9219.

González, J. (2011). "Distinguishing linear vs. nonlinear integration in CA1 radial oblique dendrites: it's about time". In: *Frontiers in Computational Neuroscience*. ISSN: 16625188.