



HAL
open science

Sub-0.3V CMOS neuromorphic technology and its potential application

Francois Danneville, Kevin Carpentier, Ilias Sourikopoulos, Michel Paindavoine, Christophe Loyez

► **To cite this version:**

Francois Danneville, Kevin Carpentier, Ilias Sourikopoulos, Michel Paindavoine, Christophe Loyez. Sub-0.3V CMOS neuromorphic technology and its potential application. 2021 International Conference on Content-Based Multimedia Indexing (CBMI 2021), Jun 2021, Lille, France. 10.1109/CBMI50038.2021.9461899 . hal-03325599

HAL Id: hal-03325599

<https://hal.science/hal-03325599v1>

Submitted on 10 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sub-0.3V CMOS neuromorphic technology and its potential application

François Danneville¹, Kevin Carpentier², Ilias Sourikopoulos², Michel Paindavoine³, Christophe Loyez¹

¹ USR CNRS 3380 – IRCICA and UMR CNRS 8520, IEMN, Lille, France

² SATT NORD, 25 Avenue Charles Saint-Venant, Lille, France

³ YUMAIN, 14H rue Pierre de Coubertin, 21000 Dijon, France

Abstract—The aim of this paper is to present a sub-0.3 V neuromorphic technology developed for spiking neural network design and its potential application. The main properties of the developed ultra low power (ULP) artificial neuron are first recalled. A description of ULP synapses follows that includes the plasticity scheme. The neuromorphic toolbox is then used to design a basic circuit allowing oriented edges classification. The circuit is made of 40 neurons and 108 plastic synapses, its consumed silicon core area is 0.025 mm² and the overall power consumption of 5 nW. Finally, the deployment of the technology within an industrial context to fabricate highly energy efficient Spike-based visual sensor is discussed.

Keywords—ultra low power; artificial neuron and synapses; spiking neural networks; oriented edges; Spike-based visual sensor

I. INTRODUCTION

For many applications, energy efficiency of information processing and communication becomes one of the main quality factors [1]. The current technology, based on binary coding, Von Neumann architecture and CMOS technology, has now reached its limits and the scaling down rules (R. Dennard, 1974) that guided the evolution of the micro- and nanoelectronics for almost 40 years cannot be anymore applied today. It becomes thus urgent to propose new paradigms of information processing capable of reducing the energy consumption in a drastic way while improving the performances as well. The connection of recent advances in neuroscience and nanodevice manufacturing suggests the possibility of designing and manufacturing radically new architectures with efficiency exceeding several orders of magnitude the current systems. This approach has extraordinary potential because all living beings are organized to minimize energy consumption and optimize performance. These bio-inspired or neuromorphic architectures incorporate the principles of operation currently known of the brain, that is large network of interconnected neurons and information coding by pulses. Artificial networks of silicon neurons and synapses are excellent candidates to build neuro- or bioinspired architectures compatible with current CMOS technology while bringing required efficiency and adaptability. They are also well suited for the implementation of synaptic mechanisms such as LTP (Long Term Potentiation) and LTD (Long Term Depression) as well as supervised/unsupervised learning algorithms

known from neuroscience as «Spike -Timing -Dependent Plasticity» (STDP).

In this context, we will briefly recall in this paper the properties of an ultra low power (ULP) artificial neuron (AN) [2] that we have developed, a preliminary step that was required for the development of our ULP neuromorphic technology. Then, the design of ULP plastic synapse [3] will be presented, including the plasticity scheme. In order to highlight the capability of this ULP neuromorphic technology, this “toolbox” (that is, constituted of artificial neuron and plastic/non-plastic synapse) is afterwards used to design a medium scaled spiking neural network (SNN), able to select edges orientations through a supervised learning. The last part describes the application of this technology, with special emphasis on spike-based smart visual sensor design.

II. ULP NEURON

The ULP AN that was developed is shown in Fig. 1. The circuit mimicks the equations of Morris-Lecar biological model [4], operates under a supply voltage VDD of 200-300 mV for which all MOS transistors behave in deep subthreshold operation, ensuring an ultra low standby power.

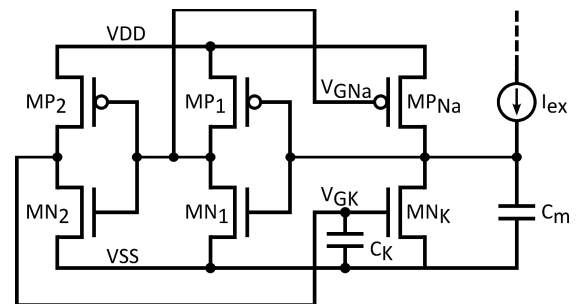


Fig. 1. ULP artificial neuron circuit

All details related to the AN behavior may be found in [2]; we will recall in the following the driving rules used to achieve an outstanding energy efficiency. Indeed, the energy consumed by the AN during one spike generation is related to the capacitance value times the square voltage variation across the capacitance (note that there are two capacitances for the AN, C_K and C_m , the latter being the membrane capacitance). As a result, in order to achieve a high energy efficiency (and ultra low power consumption), in addition to decrease the supply voltage VDD to 200-300 mV (which “de facto” limits

spike voltage swing), the capacitances value were scaled down. Following this guideline, the ULP AN performs well against state of the art [2], featuring an energy efficiency of few fJ/spike (two to three orders of magnitude lower than other CMOS silicon neurons), along a standby power of few tens pW and an area consumption of 35 μm^2 . Typical voltage waveforms are shown in Fig. 2 (for this plot, the voltage spike magnitude is 112 mV).

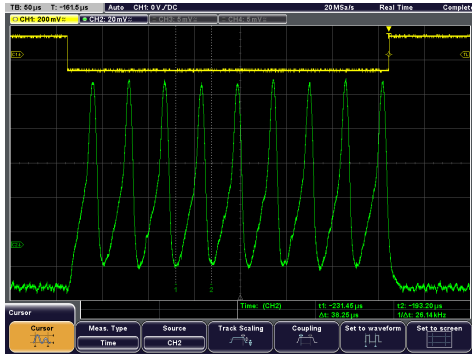


Fig. 2. Typical output waveform under constant current excitation. Excitation current = 150 pA, $V_{\text{peak-peak}} = 112$ mV, Spike frequency = 26 kHz.

III. ULP SYNAPSES

Having recalled some of the ULP artificial neuron features in Section II, this section describes the development of ULP synapses. Note the scheme used to realize these synapses are known for a long time [5], but special emphasis was made to decrease the power consumption of the excitatory synapse.

A. Basic synapses

Before presenting the basic synapses architecture, a first step has been to generate the appropriate control voltages, aimed to control synapses behavior. It was achieved by simply connecting two inverters cascaded in series, whose input is connected to the AN membrane voltage node. Such a circuit permits to generate steep voltage pulses $\overline{V_{out}}$ and V_{out} at the output of the inverters, respectively, as pictured in Fig. 3.

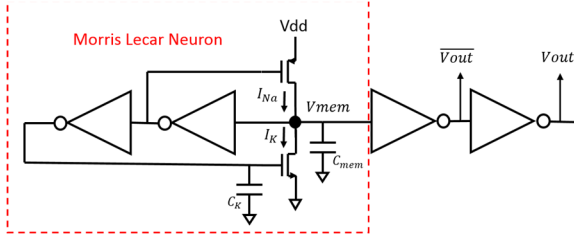


Fig. 3. Circuit to generate $\overline{V_{out}}$ and V_{out}

The basic synapses architectures are shown in Fig. 4.

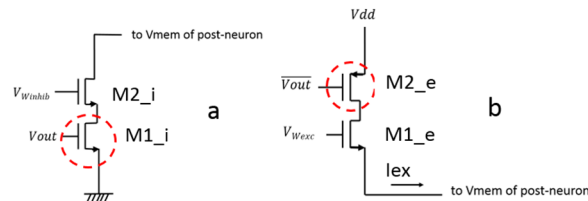
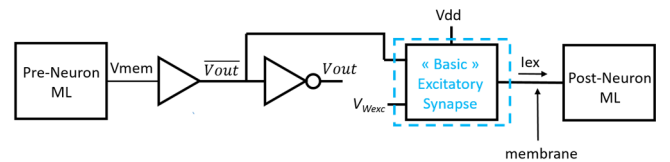


Fig. 4. Basic synapses

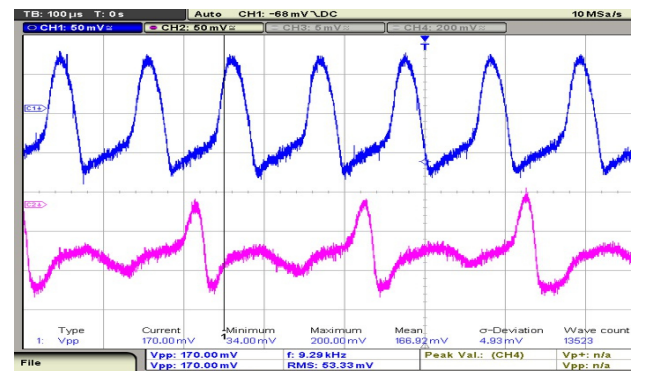
The inhibitory synapse corresponds to Fig. 4a while the excitatory one to Fig. 4b. Looking to the inhibitory synapse, the n type transistor $M1_i$ has its source connected to the ground. The “pulse-shaped” voltage V_{out} of the “pre-neuron”, that reflects the membrane voltage, is connected to the gate of $M1_i$ (which behaves as a “switch”). When the “pre-neuron” is in its resting state, $V_{out} = 0V$ and $M1_i$ is in its “OFF” state; thus the inhibitory synapse is “transparent” (no influence on the “post-neuron”). When the “pre-neuron” generates a spike, V_{out} rises to V_{dd} during the duration of the spike, and $M1_i$ switches in the “ON” state. As a result, an inhibitory current is flowing out the post-neuron down to the ground. This current contributes to an “hyperpolarization” of the post-neuron, its membrane voltage decreasing, $M2_i$ is connected in series with $M1_i$, and the voltage V_{Winhib} reflects the “weight” of the inhibitory synapse.

If we consider now the excitatory synapse, $M1_e$ behaves as well as a switch while $M2_e$ reflects the “weight” of the synapse. The source of $M2_e$ being connected to V_{dd} , when a spike is generated by the pre-neuron, $\overline{V_{out}} = 0V$, and an excitatory current is flowing down towards the membrane of the post-neuron which contributes to a depolarization of the post-neuron (that is, the membrane potential is rising). The voltage V_{Wex} reflects the “weight” of the excitatory synapse.

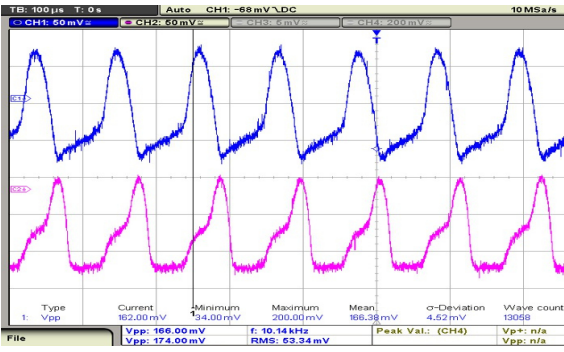
In order to verify the correct behavior of an excitatory synapse, the circuit shown in Fig. 5a was designed and experimentally investigated. It consists of an association of pre-neuron to a post-neuron with an excitatory synapse in between, as the one described in Fig. 4b. For this circuit, in Fig. 5a, the supply voltage node was connected to V_{dd} . As shown in Fig. 5b, when the synaptic weight is too weak, the post-neuron is firing after the occurrence of two pre-neuron spikes, because V_{Wex} was too weak. When V_{Wex} is reaching a sufficient level (Fig. 5c), a post-neuron spike is generated for each pre-neuron spike occurrence.



a



b



C

Fig. 5. Association « pre-neuron / excitatory synapse / post-neuron »

The spikes type in Fig. 5c correspond to “tonic spiking”, but biology features a variety of spike waveforms. For instance, in order to emulate a “bursting mode” with the synapses prior described, a basic neural circuit is to associate using a close loop a “fast” excitatory neuron connected to a “slow” inhibitory neuron; it turns out that when the “fast” excitatory neuron is excited by a current step, the “bursting” mode can be emulated [6].

The synapses that have been described lead to a post-synaptic potential whose duration is equal to those of the pre-neuron spike. Nevertheless, in biology, the duration of post-synaptic potential may last up to 10 times those of a spike. Moreover, the excitatory synapse presents in standby state an excitatory leakage current which is not negligible.

B. Ultra low power synapses

In order to implement excitatory synapses both featuring a more realistic dynamic and an ULP standby power, the architecture presented in Fig. 6 was developed.

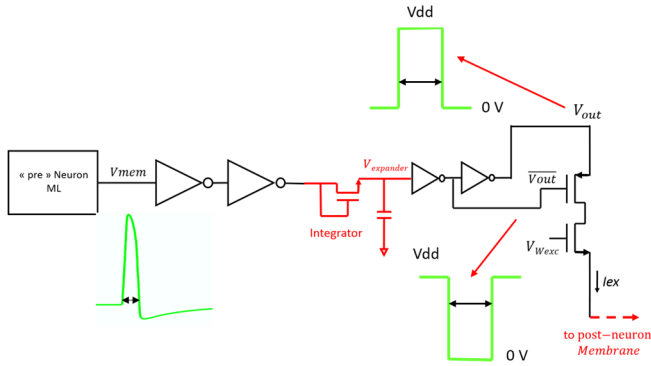


Fig. 6. Ultra low power synapses.

First, the role of the RC integrator (red in Fig. 6) is to extend the duration of V_{out} (or $\overline{V_{out}}$); as such, it can be viewed as an “expander” circuit. As illustrated, this expander is made of an active load (that is, a load made with a transistor) connected itself to a capacitor. Second, in order to decrease the power consumption of the synapse when operating in its “standby” mode (that is, when a pre-neuron does not generate a spike), the p-type MOSFET has its source connected to V_{out} instead to be connected to V_{dd} . This p-type transistor behaves also as a switch, as for a “regular” excitatory synapse such the

one pictured in Fig. 4b, but this time with (its) input gate-to-source voltage controlled by V_{out} and $\overline{V_{out}}$. The synapse behaves as follows: (i) when the “pre” neuron does not generate spikes, $V_{out} = 0V$ while $\overline{V_{out}} = V_{dd}$; it turns out that the p-type MOSFET is strongly pinched-off ($V_{sg} < 0V$), which leads to an ULP “standby” power consumption as low as the femtoWatt (fW). It is to be stressed that source and drain for the n-type MOSFET are reversed and that the current I_{ex} (Fig. 6) is negative, which means that I_{ex} flows down the ground. The synapse rather acts as an “inhibitory” synapse, preventing the post-neuron to generate any spike, (ii) when the “pre” neuron generates spikes, $V_{out} = V_{dd}$ while $\overline{V_{out}} = 0V$, the p-type MOSFET is conductive, and the synapse acts as a regular excitatory synapse.

In order to verify the correct behavior of this architecture, the temporal summation of a spikes train (made of four spikes) generated by the pre-neuron was carried out in simulation, to observe the time-varying post-neuron membrane voltage (PNMV). For this purpose, the circuit shown in Fig. 6 was used, a ML post-neuron being connected to the output synapse node (where I_{ex} is indicated Fig. 6).

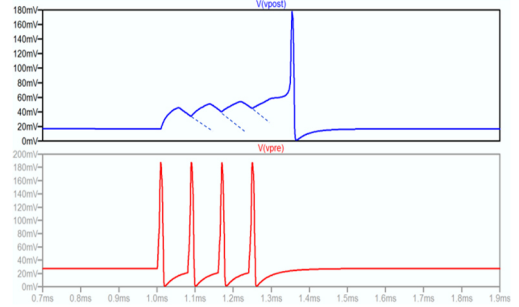


Fig. 7. Simulated membrane voltage for pre-neuron (bottom) and post-neuron (top) to verify temporal summation.

The temporal summation is clearly observed when carefully looking at PNMV: after the first pre-neuron spike, a post excitatory synaptic potential is observed. The latter has no sufficient time to completely relax before the second pre-neuron spike occurs, and the post neuron membrane voltage kept on increasing. After the fourth spike, the post-neuron membrane voltage is reaching its firing threshold and a post-neuron spike is generated.

C. Synapse plasticity

Synapse plasticity is a key point, required to make a SNN learning through a supervised or unsupervised scheme. For this purpose, the plasticity scheme implemented in [7] was used. It is based on the so-called spike time dependent plasticity (STDP), encountered in biology, and widely used in SNN. STDP is a process for which the connection strength between a pre-neuron and post-neuron is adjusted through the relative timing between pre-neuron and post-neuron spikes. The STDP scheme that was implemented has the following features:

- First, a voltage “ $V_{control}$ ”, issued from a circuit comprising a spike expander and an integrator, will increase (starting from 0 V), whenever a post-neuron is firing after a pre-neuron (providing the delay

between spikes falls within a defined time-window); this voltage is applied to a SRAM

- After a sufficient number of pre-post neurons pairing, $V_{control}$ is reaching a threshold value for which the SRAM state changes. The SRAM output which switches from 0 V to V_{dd} corresponds to a binary synaptic weight and is connected to V_{wex} node voltage of the excitatory synapse (Fig. 6)
- If V_{wex} is equal to 0 V, the excitatory synapse is “non connected” while if it is equal to V_{dd} , long term potentiation has been achieved

IV. MEDIUM SCALED SNN: ORIENTED EDGES RECOGNITION

In order to proceed to a first circuit design of the ULP neuromorphic technology presented in sections II and III, that gathers neurons and synapses, a “medium” scaled SNN intended to recognize oriented edges was designed. The aim was to fabricate a circuit which integrates different building blocks, using both non-plastic (fixed weight) and plastic (weight obtained through STDP). The learning scheme implemented was a supervised one.

A. SNN architecture

The SNN architecture is shown in Fig. 9. The circuit is made of 9 “emulated” input pixels: these input pixels simply correspond to neurons that are excited through external excitatory current. The SNN itself corresponds to a reservoir architecture which gathers 27 neurons. Each input pixel is randomly connected to 5 neurons of the reservoir through fixed weight (equal to V_{dd}) excitatory synapses. The 27 neurons constituting the reservoir are interconnected in between them, as it is represented in Fig. 9, through synapses having fixed weight as well; 80% of the synapses are excitatory while the other 20% are inhibitory, randomly distributed within the reservoir.

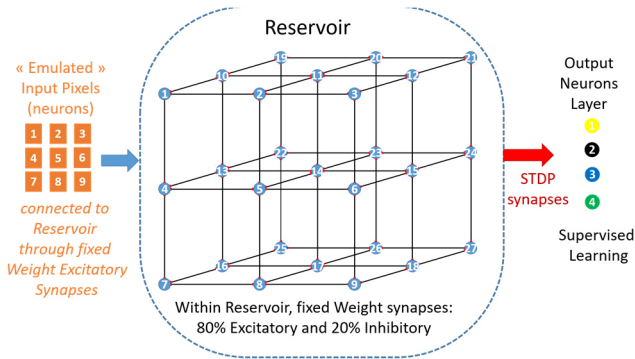


Fig. 9. Neural circuit realized to verify temporal summation.

At the output side, each neuron of the reservoir is connected to the four output neurons through plastic excitatory synapses, using the plasticity scheme explained in section III.C. Fig. 10 shows the SNN layout. The core area

surrounded in dashed red line is $250 \times 100 \mu\text{m}^2$.

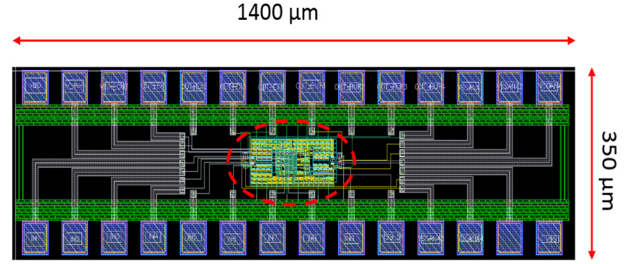


Fig. 10. Layout for the SNN designed for edges recognition.

The circuit was designed using TSMC CMOS 65 nm technology, and experimentally investigated. 28 pads are used to both set the supply voltages and command the SNN:

- 9 pads to control the 9 input neurons representing the pattern,
- 4 pads to control the 4 output neurons, to achieve the supervised learning,
- 4 pads used to readout the 4 output neurons,
- 2 pads (V_{DD} and V_{SS}) used to supply the network,
- 3 pads used to control the weight of synapses,
- 1 pad to initialize plastic synapses,
- 5 pads to bias the output buffers

B. Learning procedure

The learning procedure is detailed in Fig. 11, for the left to right diagonal edge learning. The resting state of the circuit corresponds to Fig. 11a for which only the central pixel is excited; in such a situation, the membrane voltages corresponding to output neurons labelled #1, #3 and #4 are 0 V (only 3 oriented edges were learned). Then, as pictured in Fig. 11b, the neurons mimicking “pixels” at SNN input, corresponding to a left to right diagonal edge, are simultaneously excited at the mean time of the output neuron labelled #1 at the output (supervised learning); output neurons labelled #3 and #4 are kept silent. As shown in Fig. 11c, output neuron #1 promptly learns, generating spikes. If now the supervision is disrupted (Fig. 11d), output neuron #1 is still spiking, indicating that the left to right diagonal edge pattern has been learned.

The above procedure was repeated to learn two other patterns and the results are shown in Fig. 12.

Starting from the resting state (Fig. 12a), the three learned patterns are successively presented at the input of the SNN circuit. Fig. 12b-d clearly indicate that output neurons labelled #1, #3 and #4 specialized themselves for these different oriented edges.

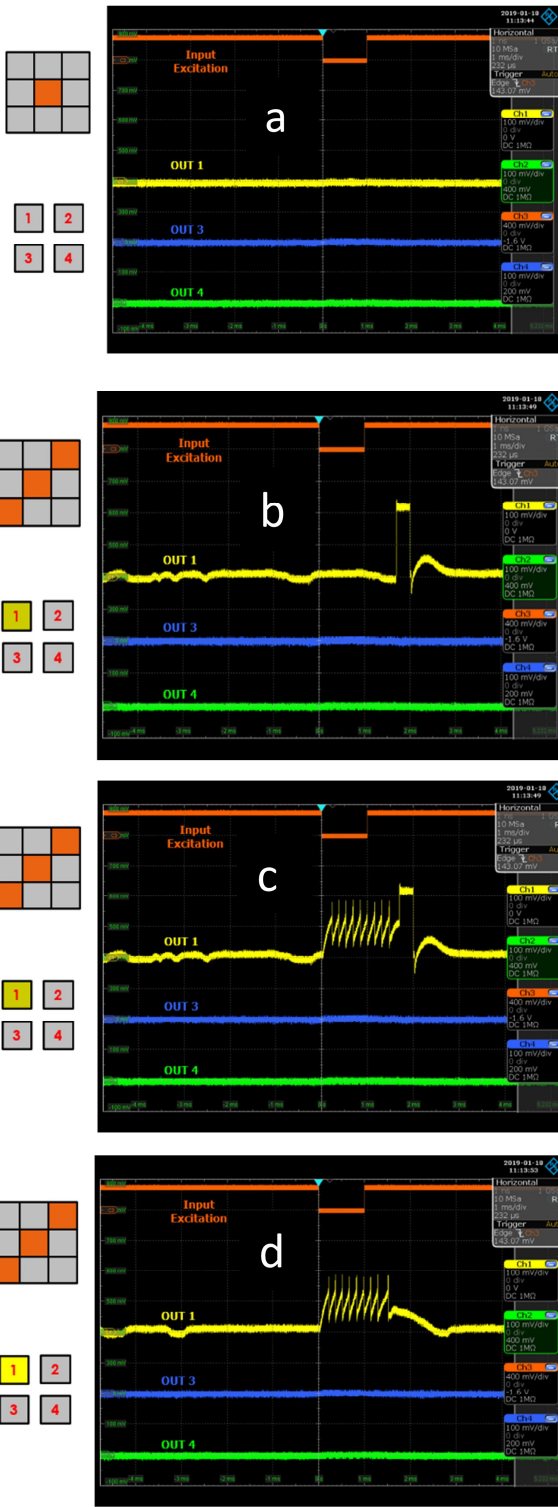


Fig. 11. "Left to right diagonal edge" learning

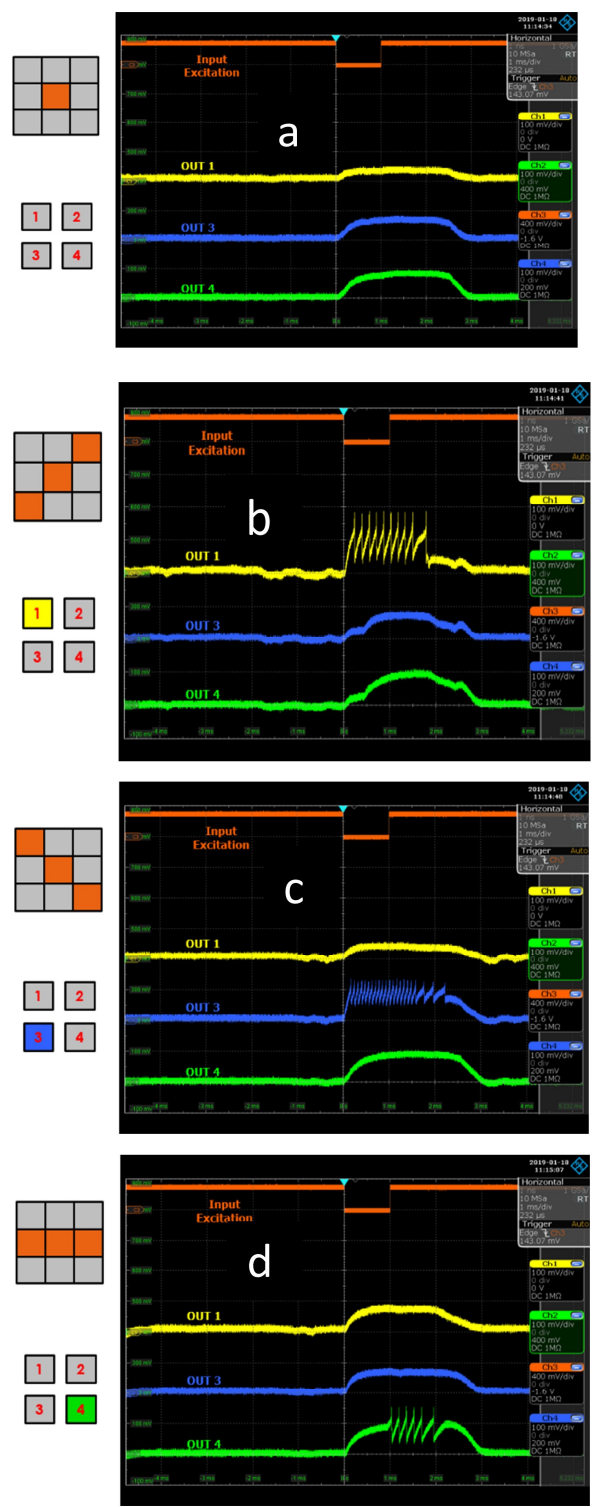


Fig. 12. Illustration of SNN edges pattern recognition.

C. Power Consumption: discussion

Power consumption was the main goal to investigate the ULP neuromorphic technology. After learning, the different experimental DC currents and overall power consumption were carefully measured for the SNN. Results are provided in Table I, both in standby condition and when a pattern is presented, for $V_{dd} = 300$ mV. From Table I, it is obvious that the main power consumption comes from neurons and SRAM (used to memorize the weights after learning). The increase of power consumption when a pattern is presented is due to spikes generated within the SNN. The SNN gathers 40 neurons and 108 SRAM.

TABLE I.

	Total DC current for all neurons and all SRAM	Total current for all synapses	SNN Power Consumption
After learning, Standby Power	9.9 nA	5 nA	4.48 nW
After learning, Pattern presented	11.8 nA	5 nA	5.07 nW

From this table, assuming that the SRAM power consumption is half those of a neuron, the stand alone neuron power consumption is around 40 pW. This number is in line with the power consumption reported in [2]. From these numbers, it can be concluded that the ULP neuromorphic technology correctly scales regarding the power consumption.

V. POTENTIAL APPLICATION: ULP SPIKE-BASED VISUAL SENSOR

Regarding the performance of our ULP neuromorphic technology, one can expect to develop highly energy efficient spike-based visual sensor. Such a development is on-going, through a collaboration with an industrial partner. Currently, the more powerful technology is the so-called Address Event Representation (AER) visual sensor technology [8]. Though this AER based technology performs very well, the power consumption per pixel (PCPP) is around 100 nW, which is more than three times higher than the PCPP featured by a human being retina (which features a PCPP of 30 pW, see Table 3.1 in [9]). The (main) reason lies in the supply voltage used to develop AER visual sensor, which stands higher than 1V for this technology.

In this context, our neuromorphic technology is used to develop Spike-based visual ULP sensor. Indeed, in addition to neurons and plastic/non-plastic synapses, an ULP optical sensor was developed [10] that operates luminance transduction into spikes generation. Work is in progress to fabricate a $64 * 64$ pixels demonstrator, able to extract luminance (contrast) or edges (using ON cells) from a fixed or video visual scene.

Some challenges obviously remain, to cite a few: (i) contain the technology variability, (ii) design the interface with the digital world. When fixed, the time varying spikes generated by the visual sensor will be processed by low power FPGAs/Processors to perform various supervised / non-supervised learning. Full analog implementation will also be considered in the framework of dedicated applications.

VI. CONCLUSION

In this paper, the scaling progress of an ULP neuromorphic technology recently developed has been proposed. The main features of the ULP artificial neuron was first recalled, followed by those of ULP artificial synapses with special emphasis on their dynamic and plasticity. Hence, the ULP neuromorphic toolbox was used to design a medium scaled SNN, aimed to oriented edges classification, the power consumption of # 5nW, in line with the stand alone power consumption reported for a stand alone artificial neuron [2]. Finally, a discussion on the opportunity to use this ULP neuromorphic technology to develop highly energy efficient spike-based visual sensor was made.

ACKNOWLEDGMENTS

The authors acknowledge IRCICA, SATT NORD and YUMAIN company for the financial support. The personnel of the “Centrale de Caractérisation” of IEMN for its contribution to the electrical measurements of the chips. The authors wish also to acknowledge A. Cappy with whom they had fruitful discussions.

REFERENCES

- [1] Greenpeace : « How dirty is your data – A look at your energy choices that power cloud computing » 2011 <https://storage.googleapis.com/planet4-international-stateless/2011/04/4cceba18-dirty-data-report-greenpeace.pdf>
- [2] I. Sourikopoulos, S. Hedayat, C. Loyez, F. Danneville, V. Hoel, E. Mercier and A. Cappy A., “A 4- μ m/spike artificial neuron in 65 nm CMOS technology”, *Frontiers in Neuroscience*, Volume 11, Article 123, March 2017
- [3] F. Danneville, A. Cappy, I. Sourikopoulos, C. Loyez, “Switched artificial synapses”, Patent WO2019219618 (A1) – 2019-11-21
- [4] C. Morris and H. Lecar, “Voltage oscillations in the barnacle giant muscle fiber”, *Biophys. J.* 35, 193–213, 1981.
- [5] C. Bartolozzi, G. Indiveri, “Synaptic Dynamics in Analog VLSI », *Neural Computation* 19, 2581–2603 (2007)
- [6] F. Danneville, I. Sourikopoulos, S. Hedayat, C. Loyez, V. Hoël, and A. Cappy “Ultra low power analog design and technology for artificial neurons”, Proceedings of 31st Annual IEEE Bipolar/BiCMOS Circuits and Technology Meeting, BCTM 2017, Miami, FL, USA, october 19-21, 2017
- [7] J. B. Arthur, and K. Boahen, “Learning is silicon: timing is everything”, in *Advances in neural information processing systems*, 18 January 2005
- [8] Guillermo Gallego et al, « Event-based Vision: A Survey », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, early article 2020
- [9] S-C Liu, T. Delbruck, G. Indiveri, A. Whatley, R. Douglas, “Event-based Neuromorphic Systems”, Book edited by WILEY, ISBN: 9780470018491, 2015
- [10] A. Cappy, F. Danneville, Hoël V., Loyez C., Sourikopoulos, “Optical sensor”, Patent WO2019081562, 02-05-2019