INTUITIVE

INnovative Network for Training in ToUch InteracTIVE Interfaces

Grant agreement: #861166 H2020-MSCA-ITN-2019
Start date: 1 October 2019 End date: 30 September 2023

Deliverable reporting document

Deliverable no: D4.6		WP: 4
Deliverable title: Low-power sensory readout and electronic interface	Type: Report	Dissemination level: Public
Due Delivery date: 30 September 2021		Date delivered: 15 October 2021 Revised: 15 March 2022

Description:
Architectural study of time-based sensor readout in CMOS. Circuit design and layout of a low-power time-based sensor interface in CMOS.

INTUITIVE deliverable report

Low-power sensory readout and electronic interface

Mark Alea, Prof. Georges Gielen, KU Leuven

Architectural study of time-based sensor readout in CMOS.
Circuit design and layout of a low-power time-based sensor interface in CMOS.

Summary

The ESR's contribution to the INTUITIVE project is to carry out an architectural study of time-based sensor readout in CMOS for the INTUITIVE e-skin application, followed by integrated circuit design and layout of a low-power time-based sensor interface in CMOS. In the first reporting period, an extensive literature review of suitable sensor readout architectures for electronic skin has been conducted. Considering the dense sensor array envisioned for electronic skin, event-driven sensor conversion techniques with some form of sampling rate reduction have been investigated, as they would translate in a reduction of the power consumption of the overall system. Furthermore, a taxel-wise readout array eliminates the need for both fast polling and sampling rates on an otherwise multiplexed ADC. By applying event-driven neuromorphic conversion, sensor data are sampled, quantized and transmitted at a sub-Nyquist rate while preserving the temporal information content. It is therefore no surprise that biological skin, with millions of years of evolution and confined to spend the least amount of energy possible, in fact has come to implement a similar event-driven conversion of tactile input. By mimicking how biology converts and processes tactile information into a hardware-efficient and power-efficient implementation, the derived sensor readout is estimated to consume orders of magnitude lower power consumption than the current state of the art of electronic skin, while maintaining sufficient fidelity in its sensor value encoding. In the second reporting period, the full circuit-level design and layout of the event-driven taxel readout circuit has been completed, as will be reported below.

1. Existing electronic skin readout solutions

The research into electronic skins so far has focused mainly on developing the tactile sensors able to transduce normal force [1-2], shear [3-4], among many parameters. The development of tactile sensors is not trivial for many reasons. One reason is, in contrast to existing biological-inspired sensors such as artificial retinas [5] and cochleae [6], artificial skins need to interact with the environment it is sensing, thus making it prone to wear and tear. Secondly, current fabrication processes that allow sensors to have skin-like properties, such as force transduction, flexibility, and viscoelasticity, limit the spatial resolution achievable on the sensors. Therefore, the implementation of skin sensor readouts has mainly been relegated to using custom off-the-shelf centralized electronics with overdesigned precision and speed [1,7-8]. While this allows for quick measurements and characterizations of skin sensors in existing designs, the readout circuits suffer from large area and power consumption. Although this may be sufficient for existing skin sensor designs with <100 sensors and coarse spatial resolution, this becomes an issue for emerging large-area electronic skin with fine spatial resolution and 2-3 orders magnitude more sensors, as targeted in INTUITIVE.

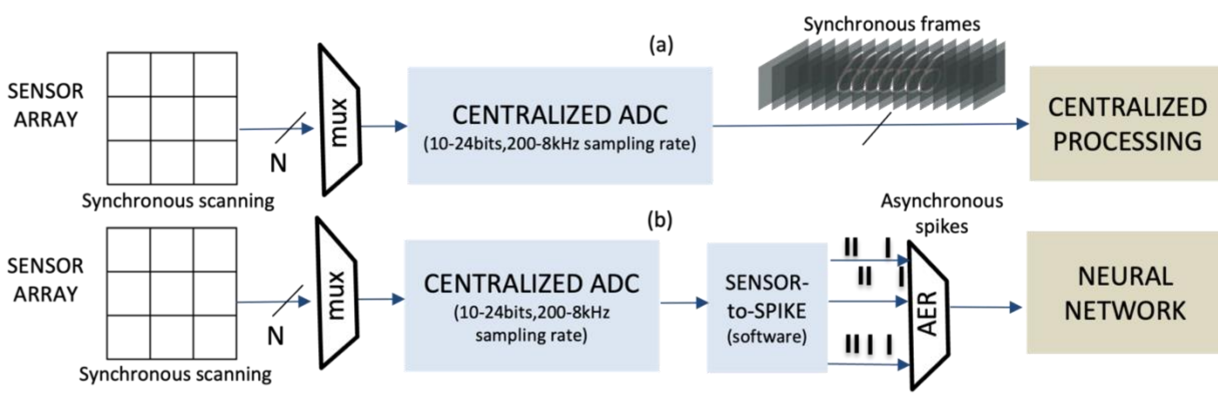


Fig.1. State-of-the-art electronic skin readout solutions can be classified in two major categories: (a) synchronous frame generation using a central ADC and further digital processing, and (b) asynchronous spike generation in software followed by a spiking neural network.

To satisfy the fast sensor array scan rate (1-10 kHz) required to detect fine spatiotemporal tactile stimuli [7], skin sensors in an array are polled sequentially one by one [1,8], or column-based [7]. As shown in Fig. 1(a), the analog sensor output is then multiplexed into a central ADC for conversion, producing synchronous frames of tactile information that are periodically sent for processing.

The periodic high rate of sensing and transmission of tactile frames despite possibly sparse tactile stimuli is power-inefficient. Furthermore, increasing the sensor count for future large-scale electronic skins result in increasingly faster sensor polling rates to satisfy the scan rate required. Similar to the insights developed from early implementations of artificial retinas, event-based sensing of sparse visual input leads to power-efficient readout electronics that scale well with the rising sensor count. Having individual pixels respond independently to the visual stimuli, the effective sensing and data transmission rate scales well with the information rate of the visual stimuli.

The advent of using machine learning (ML) techniques in computation offers energy reduction over conventional digital processing [9]. Convolutional neural networks (CNN) deployed in event vision [10], through training, classify visual stimuli with good accuracy despite the sensors having large nonlinearity and mismatch. Spiking neural networks (SNN), while being a relatively newer development, are now slowly becoming at par with CNN classification performance [10]. SNNs are expected to be more power-efficient due to their sparse and asynchronous activity, scaling well with the sparsity of the input [11]. Similar to biological neural networks, SNNs use asynchronous spikes as inputs instead of synchronous frames. A sensor-to-spike encoder is therefore highly compatible as input to SNNs, while also being power-efficient in its sensing due to its event-driven nature.

Existing spike encoders (see Fig. 1(b)) are mostly implemented in software by post-processing uniform rate-quantized sensor values into spikes. Thus, real-time spike conversion of sensor values from a dense array today is hardware-inefficient -- relying on a single ADC with overdesigned precision, sequentially polling sensor values at a high rate. Therefore, while the resulting spike rate is low (10-100 Hz), the ADC sampling rate is still high (1-100 kHz), despite having sparse tactile stimuli as input. The sensor readout developed and proposed in the INTUITIVE project implements sensor-to-spike encoding directly and efficiently in hardware, eliminating the required software processing and the unnecessarily high input sampling rate.

2. Proposed event-driven e-skin readout architecture

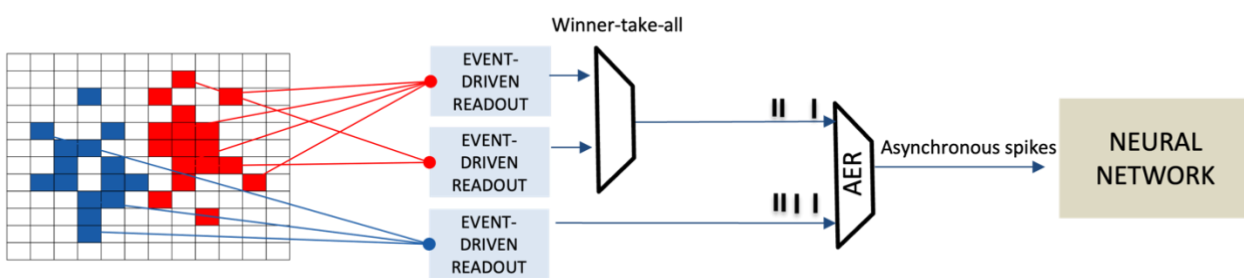


Fig.2. Proposed event-driven sensor readout architecture with spatial compression.

The proposed sensor readout (Fig. 2) mimics biology in both its sensor conversion and data processing. In mammalian skins, touch is communicated to the nervous system and eventually to the brain by way of asynchronous spikes – pulse trains with temporal patterns that code the input sensor signal. The brain then processes incoming spikes through a power-efficient, asynchronous, and densely connected spiking neural network (SNN). State-of-the-art SNNs implemented on chip (in other applications) have been demonstrated to perform computation at orders of magnitude lower power consumption [9] than conventional digital processing. Thus, in the proposed sensor readout, continuous-time tactile sensor values are encoded into asynchronous, discrete-time spikes that are then transmitted to a spiking neural network for tactile stimulus classification. Furthermore, taxel-based event-driven readout converts information only from sensors whose values have changed. This reduces the sensing and transmission rates for sparse stimuli, while maintaining a high sampling rate capability during fast-changing inputs.

A POSFET-based piezoelectric force sensor array is chosen as taxel sensor because of its sub-millimeter spatial resolution [1] and its compatibility as an integration step with standard CMOS wafers. This production can be done at wafer level by the group of Prof. Leandro Lorenzelli at the Fondazione Bruno Kessler (FBK) in the INTUITIVE project. Current sensor fabrication techniques, bound by the distance between two neighboring piezo sensors, sets the minimum possible taxel area to 200 μm x 200 μm, as shown in Fig. 3. Assuming a 2 mm x 2 mm sensing area, this translates to 100 sensors in total. While the sensor density is only half that of human fingertips and 10 times that of palms, the density would still be the highest reported in literature.

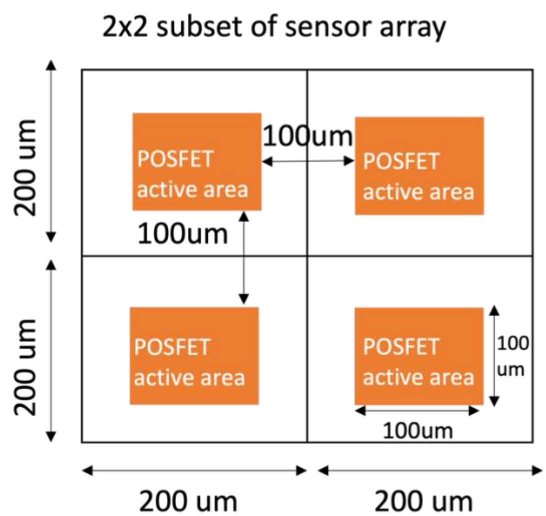


Fig.3. A 2x2 subset of the sensor array highlighting the taxel dimensions and spacing.

The individual readout channel must have a very low area and needs to be smaller than the taxel. However, since the readout output is processed through an SNN and is not reconstructed post spike conversion for conventional digital processing, the required sensor encoding precision is reduced. This allows for using simpler sensor readout circuits with reduced area and power consumption.

By mimicking biological skins, the proposed sensor readout is capable of both temporal and spatial compression. Temporal compression provides sub-Nyquist sampling and reduced data transmission rates due to its event-driven nature. Spike encoders implemented fully in hardware

are inherently event driven. At the same time, spatial compression is achieved by mimicking the skin's complex receptive fields (CRFs). Random spatial sampling and compression of multiple sensor inputs to a single spike train result in a reduced overall data transmission rate and increased robustness to wear and tear. This will now be explained in more detail.

2.1. Temporal compression

Spike encoders such as in [5] offer temporal compression of sensor information, where - instead of in Nyquist-rate conversion - the sensor value is only converted during level crossings (i.e. information rate conversion).

Typical hardware-implemented spike encoders such as the one used in artificial retina [5] generate either a bipolar spike (ON or OFF) if the change in the sensor value exceeds a certain threshold. Other spike encoders use a similar approach but differ on how the threshold is generated, either tracked through a moving baseline following the previous level crossing [12-13] or a baseline generated from average values of past inputs with a certain moving time window [12].

In [14], a unipolar spike encoder based on the Simplified Response Model (SRM) [15] of a neuron is conceptualized. In the SRM interpretation of neuronal encoding, the neuron is modelled as a series of linear filters with feedback. As shown in Fig. 4, this resembles a level-crossing data converter with the threshold derived from the low-pass-filtered output spikes. Instead of a constant threshold in-between level crossing, the threshold is an exponentially decaying value mimicking the neuron's refractory response and spike-induced membrane voltage change [16]. As far as we know, this proposed spike encoder has not yet been implemented and validated experimentally on-chip. It will therefore be designed and implemented as part of the INTUITIVE project activities.

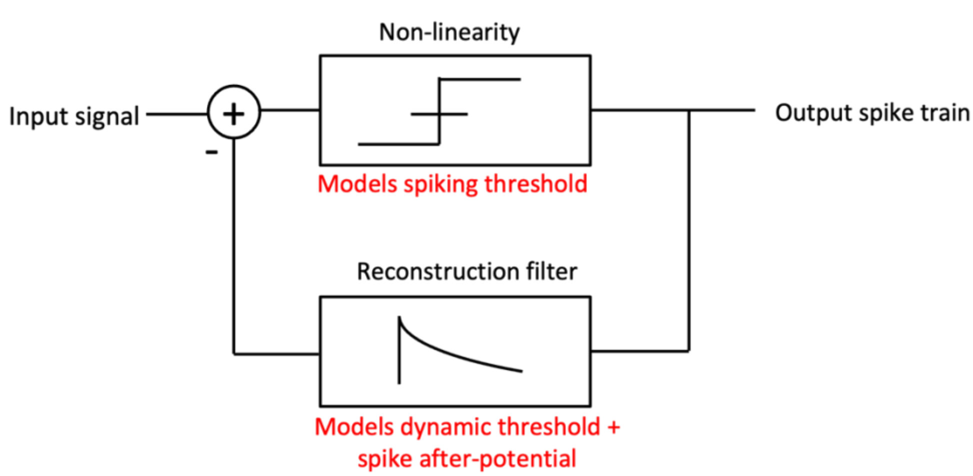
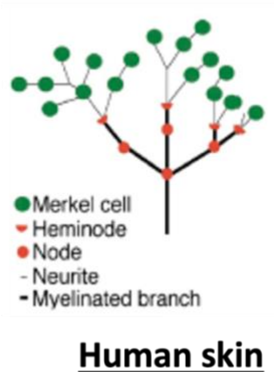


Fig.4. Simplified Response Model (SRM) interpretation of the neuron firing for e-skin implementation.

2.2. Spatial compression

(a)



(b)

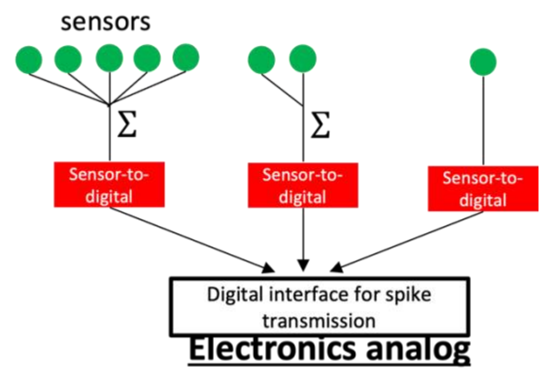


Fig.5. Mimicking the complex receptive fields (CRFs) (a) in the sensor readout using heminodes (b).

The biological skin offers a form of spatial compression of sensor information through complex receptive fields (CRFs), as shown in Fig. 5(a). CRFs are formed by random innervation of multiple mechanoreceptors by a single afferent [17-18], therefore multiple sensor information is compressed into a single spike train. It has been theorized that the CRFs allow for first order encoding of tactile stimulus [17], that is, spike trains generated by first-order neurons already reflect the tactile stimulus information such that the spike trains allow sufficient discriminability across different stimuli.

In this work, CRFs are mimicked by implementing heminodes, as shown in Fig 5(b). In biological skins, heminodes are areas of sensor value aggregation and spike generation [19]. Since the POSFET output value is a current, addition of multiple sensor values is done by routing individual POSFET currents to the heminode input current mirror by way of configurable switches.

3. Circuit implementation of the event-driven taxel readout

3.1. Taxel-to-spike conversion circuit

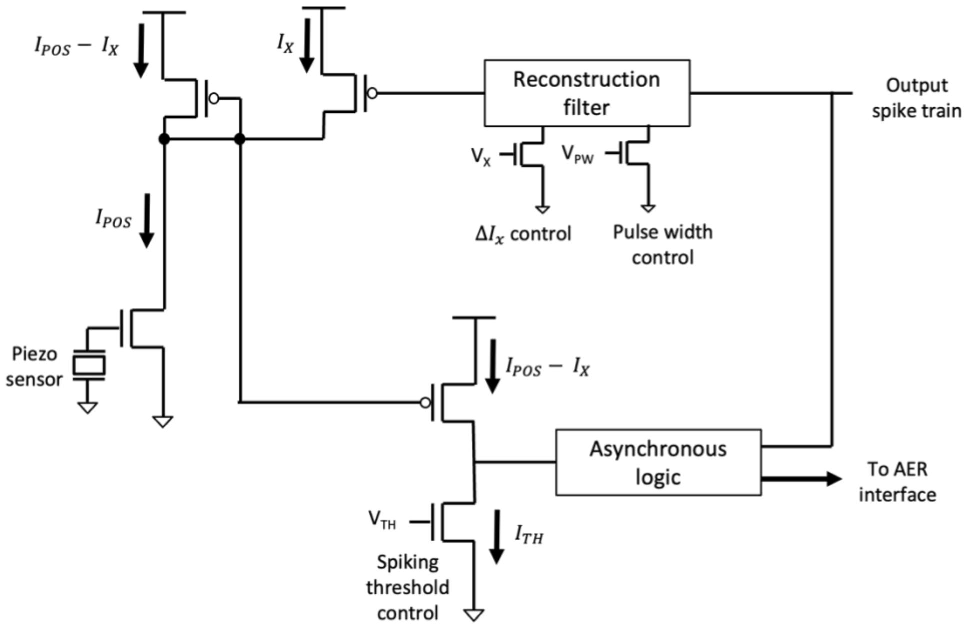


Fig.6. Taxel to spike readout circuit.

Once the readout architecture defined, as described above, the circuit implementation has been started as part of the INTUITIVE project. This work has resulted in a prototype chip implementation for feasibility demonstration. In our work, a neuron-like spike encoder with exponentially decaying threshold is being explored. The spike encoder is based on the SRM-based model of neuronal encoding. As shown in Fig. 6, through negative feedback, the difference between the input and the effective threshold is quantized by a comparator, producing an impulse output that mimics spikes. Since an output spike is only generated when the input and the effective threshold is within a certain fixed value, the spike encoder is event-driven, and level-crossing based [20]. The effective threshold can be seen as a low-pass-filtered version of the output spike train.

As mentioned, nature's way of encoding environmental stimuli has been refined and optimized through evolution to maintain sufficient encoding fidelity while consuming the lowest energy possible. As such, we see that analogy-to-digital conversion techniques developed without biomimicry in mind are, in fact, already prevalent in nature. For instance, the exponentially decaying reconstruction filter output can be thought of as a dither-adding mechanism similar to [21], increasing the sampling rate for slowly varying signals. In addition, the decaying filter can be thought of as a neuron-like equivalent of adaptive level crossing [22] where the threshold starts out large and then progressively decreases.

3.1.1. Current-based operation

The sensor readout uses current-based operation to do the sensor transduction, the current subtraction in the negative feedback, and consequently, the level-crossing detection. One reason is because the POSFET sensor used in the taxels translates force transduction into a voltage output across the piezoelectric layer [1]. This piezoelectric layer usually requires a high input impedance interface which is typically implemented using a single MOS transistor. Therefore, assuming a small voltage swing at the gate input of the MOS transistor, an equivalent output current swing is then provided. Current subtraction to implement the negative feedback is then easily done through current mirrors. Another reason is because the reconstruction filter, when implemented with a current output, requires simple circuits with low bias currents. Furthermore, the addition of multiple sensor outputs in implementing the CRFs is easily done by way of current mirrors. The routing of currents on-chip is less sensitive to the parasitic resistance in the metal wirings, allowing for long routing traces as required for flexibility in the CRF configurations.

3.1.2. Reconstruction filter

In actual biological neurons, the reconstruction waveform or the effective threshold for spiking is often modelled as the sum of both the dynamic threshold and the spike after-potential [Gerstner]. Both the dynamic threshold and the spike after-potentials are modelled with multiple exponentially decaying kernels with different time constants, ranging from tens to hundreds of milliseconds [16]. For small area, the reconstruction filter in the proposed readout has only one exponentially decaying time constant.

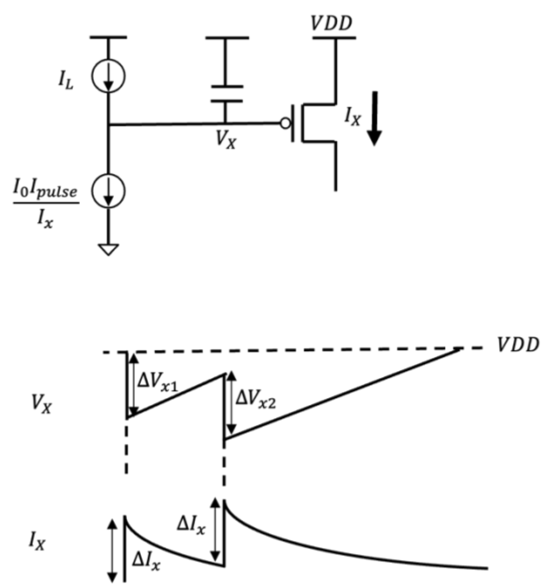


Fig.7. Generation of the effective threshold current through a subthreshold PMOS transistor.

Similar to silicon neurons [23], the exponentially decaying kernel seen in membrane filters is implemented with log-domain subthreshold circuits. One reason is that the small subthreshold currents yield biologically plausible time constants without using very large capacitors. Secondly, the exponential dependence of the transistor current on the gate-source voltage allows easy multiplication and division of currents, which is important to establish an additive increase on the reconstruction filter output during output spikes.

As shown in Fig. 7, the effective threshold current is generated by a reconstruction filter through a subthreshold PMOS transistor. Due to the exponential dependence of the PMOS output current on the gate overdrive, a simple linearly varying gate voltage would provide an exponentially decaying output current. This is implemented by injecting a DC current into a capacitor (pushing the gate voltage closer to the supply). The additive increase on the threshold during an output spike is implemented by discharging the capacitor (pushing the gate voltage closer to GND) with an impulse current, or practically, a pulse current with a very small pulse width. The consequent decrease on the capacitor voltage is proportional to both the current magnitude and the pulse width. To make sure that the consequent jumps on the output current are additive in nature, the pulse current I_{pulse} is scaled by the output current I_X.

The POSFET output current minus the reconstruction filter output is then compared with a fixed threshold using a simple 2-transistor current comparator (see Fig. 6) like the pixel circuits in event cameras [5].

3.2. Heminode circuit and routing

As shown in Fig. 8, CRFs are implemented on-chip by way of heminodes. The combination of multiple sensor values is possible by routing the currents to either a heminode or a local taxel circuit. This is done by way of digitally configurable analog multiplexers. Heminodes consist of the same circuits as seen in a taxel circuit (Fig. 6) but they experience a larger signal swing.

Flexibility in the heminode configuration will be implemented in the prototype chip implementation using a digitally controlled switching network (see Fig. 8).

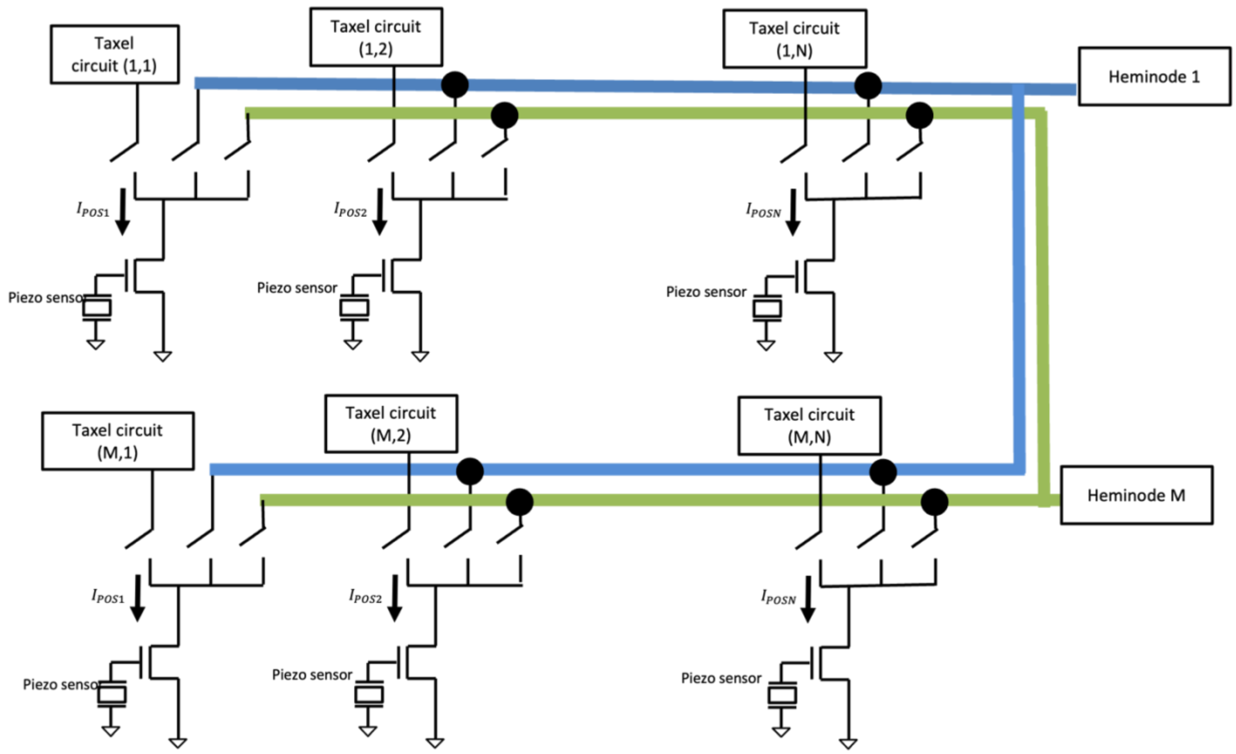


Fig. 8. Heminode configuration and routing.

4. Chip prototype

The design and layout of the prototype chip implementation has been completed for tapeout in March 2022. The chip is set to be fabricated in a 0.18-µm TSMC CMOS process. Since wafers are required for sensor deposition, wafers rather than individual dies had to be ordered. Delivery of the wafers is expected sometime between June to August 2022. The wafers will then be sent for sensor deposition at Fondazione Bruno Kessler (FBK) and chip thinning at the University of Glasgow (UoG). Electrical testing and characterization of the chip is then planned afterwards in the IC measurement facilities at KU Leuven. Mechanical e-skin testing is planned to be conducted at different sites, including KU Leuven, FBK, and UoG.

As planned, the readout is a taxel-wise, event-based sensing system with spikes as outputs. The spikes are to be read and processed by a spiking neural network. The readout achieves a state-of-the-art density of 200-µm sensing spatial resolution enabled by a) dense piezoelectric sensor array deposition compatible with CMOS technology, and b) area-efficient taxel circuits.

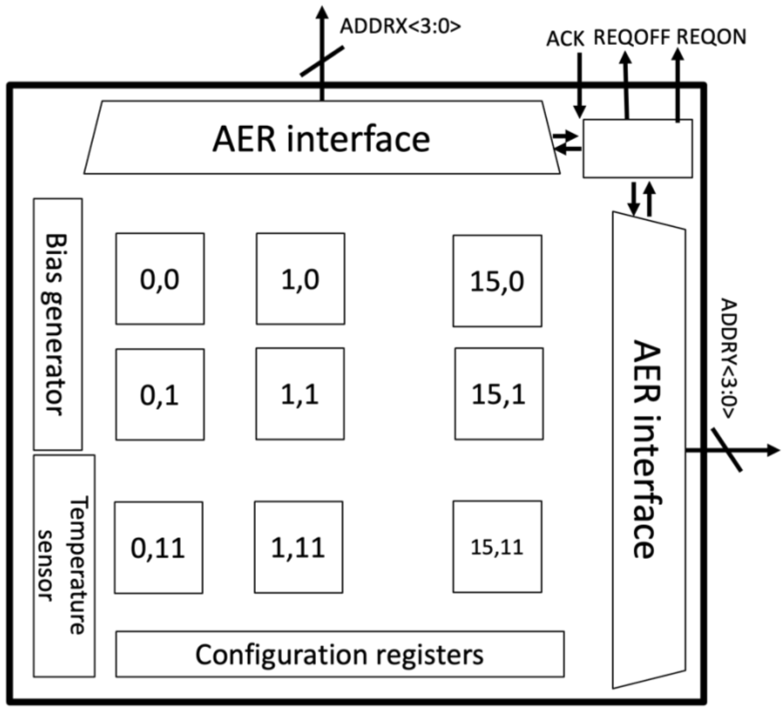


Fig. 9. Chip prototype

The chip prototype, shown in Fig. 9, contains a taxel-wise readout of 12x16 piezoelectric sensors. The output spikes of the sensors are read asynchronously by a digital interface, called Address Event Representation (AER). The spikes, encoded into an equivalent 8-bit address (4-bit row and 4-bit column addresses) are then transmitted off-chip. Since the interface from the sensor to the receiver is asynchronous, request (REQ) and acknowledge (ACK) handshaking is used. By using an AER interface, the chip is compatible with existing spiking neural net (SNN) hardware for classification of the taxel input.

As shown in Fig. 8 above, spatial compression similar to the heminode function is possible by allowing programmability of the sensor to neuron connections. Such connection can be programmed through the digitally configurable registers indicated in Fig. 9. The registers are then used to control the analog multiplexers that connect the sensor to the neurons.

A PTAT temperature sensor is also included on-chip to add extra information about the tactile stimulus to be classified. Programmable bias generators are also included on-chip, allowing for a fine control of the spike encoding parameters on the prototype.

In the next reporting period, the measurement setup will be prepared, and the fabricated and postprocessed chip samples will be measured, both electrically and in terms of e-skin performance.

References

[1] R. S. Dahiya, A. Adami, L. Pinna, C. Collini, M. Valle and L. Lorenzelli, "Tactile Sensing Chips With POSFET Array and Integrated Interface Electronics," in IEEE Sensors Journal, vol. 14, no. 10, pp. 3448-3457, Oct. 2014.

[2] W. Navaraj, and R. Dahiya, "Fingerprint-Enhanced Capacitive-Piezoelectric Flexible Sensing Skin to Discriminate Static and Dynamic Tactile Stimuli", in Adv. Intell. Syst., 1: 1900051, 2019.

[3] J. Park, Y. Lee, J. Hong, M. Ha, Y.D. Jung, H. Lim, S.Y. Kim, and H. Ko, "Giant tunneling piezoresistance of composite elastomers with interlocked microdome arrays for ultrasensitive and multimodal electronic skins", in ACS Nano. 8(5):4689-97, 2014 May 27.

[4] S. Sundaram, P. Kellnhofer, Y. Li, *et al,* "Learning the signatures of the human grasp using a scalable tactile glove," in Nature 569, 698–702, 2019.

[5] P. Lichtsteiner, C. Posch and T. Delbruck, "A 128$\times$ 128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor," in IEEE Journal of Solid-State Circuits, vol. 43, no. 2, pp. 566-576, Feb. 2008.

[6] M. Yang, C. Chien, T. Delbruck and S. Liu, "A 0.5V 55µW 64×2-channel binaural silicon cochlea for event-driven stereo-audio sensing," 2016 IEEE International Solid-State Circuits Conference (ISSCC), 2016, pp. 388-389.

[7] W.W. Lee, S.L. Kukreja, and N.V. Thakor, "Discrimination of Dynamic Tactile Contact by Temporally Precise Event Sensing in Spiking Neuromorphic Networks," in Frontiers in Neuroscience, vol. 11, 2017.

[8] W. W. Lee *et al.*, "A neuro-inspired artificial peripheral nervous system for scalable electronic skins," in Sci. Robot., vol. 4, no. 32, Jul. 2019, Art. no. eaax2198.

[9] K. Boahen, "A neuromorph's prospectus," in Computing in Science & Engineering, vol. 19, no. 2, pp. 14-28, Mar.-Apr. 2017.

[10] D. Zambrano, and S. M. Bohte, "Fast and Efficient Asynchronous Neural Computation with Adapting Spiking Neural Networks", in arXiv:1609.02053, 2016.

[11] G. Indiveri and Y. Sandamirskaya, "The Importance of Space and Time for Signal Processing in Neuromorphic Agents: The Challenge of Developing Low-Power, Autonomous Agents That Interact With the Environment," in IEEE Signal Processing Magazine, vol. 36, no. 6, pp. 16-28, Nov. 2019.

[12] N.Kasabov and E. Capecci, "Spiking neural network methodology for modelling, classification and understanding of EEG spatio-temporal data measuring cognitive processes," in Information Sciences, Volume 294, 2015, Pages 565-575, ISSN 0020-0255

[13] M. Zamani, M. Ronchini, H. A. Huynh, H. Farkhani and F. Moradi, "Flexible Energy-Efficient Implementation of Adaptive Spiking Encoder for Neuromorphic Processors," 2021 IEEE International Symposium on Circuits and Systems (ISCAS), 2021, pp. 1-5.

[14] Y. C. Yoon, "LIF and Simplified SRM Neurons Encode Signals Into Spikes via a Form of Asynchronous Pulse Sigma–Delta Modulation," in IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 5, pp. 1192-1205, May 2017, DOI: 10.1109/TNNLS.2016.2526029.

[15] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski, "Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition". Cambridge University Press, USA, 2014.

[16] S. Mensi, R. Naud, C. Pozzorini, M. Avermann M, C. Petersen, W. Gerstner, "Parameter extraction and classification of three cortical neuron types reveals two distinct adaptation mechanisms," in J. Neurophysiol. 107(6):1756-75, March 2012.

[17] J. Pruszynski and R. Johansson, "Edge-orientation processing in first-order tactile neurons," in Nat Neurosci 17**,** 1404–1409, 2014.

[18] E. Jarocka, J. Pruszynski, and R. Johansson, "Human touch receptors are sensitive to spatial details on the scale of single finger- print ridges," in J. Neurosci 41: 3622–3634, 2021.

[19] D.R. Lesniak, K.L. Marshall, S.A. Wellnitz, et al., "Computation identifies structural features that govern neuronal firing properties in slowly adapting touch receptors," in Elife;3:e01488, 2014

[20] Y. Tsividis, "Event-driven data acquisition and continuous-time digital signal processing," IEEE Custom Integrated Circuits Conference 2010, pp. 1-8, 2010.

[21] T. Wang, D. Wang, P. J. Hurst, B. C. Levy and S. H. Lewis, "A Level-Crossing Analog-to-Digital Converter With Triangular Dither," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 56, no. 9, pp. 2089-2099, Sept. 2009.

[22] C. Weltin-Wu and Y. Tsividis, "An Event-driven Clockless Level-Crossing ADC With Signal-Dependent Adaptive Resolution," in IEEE Journal of Solid-State Circuits, vol. 48, no. 9, pp. 2180-2190, Sept. 2013.

[23] E. Chicca, F. Stefanini, C. Bartolozzi and G. Indiveri, "Neuromorphic Electronic Circuits for Building Autonomous Cognitive Systems," in Proceedings of the IEEE, vol. 102, no. 9, pp. 1367-1388, Sept. 2014.