Power-Efficient and Accurate Texture Sensing Using Spiking Readouts for High-Density e-Skins

Mark Daniel Alea[1], Ali Safa[1,2], Jonah Van Assche[1], Georges G.E. Gielen[1,2]
[1]*ESAT, KU Leuven,* [2]*imec, 3001, Leuven, Belgium*
{Mark.Alea,Ali.Safa,Jonah.Vanassche,Georges.Gielen}@kuleuven.be

***Abstract*—Fine-grain tactile sensing has recently gained much attention in robotics applications where the manipulation of potentially fragile objects must be provided. This has led to the emergence of electronic skin (e-skin) sensors, usually implemented with conventional frame-based acquisition chains. In addition, prosthetics applications require e-skins with human-level, sub-millimeter spatial resolution. This paper proposes to study two types of spike-based e-skin readout circuits, based on conventional and neuromorphic level crossing architectures. Compared to prior frame-based, coarse spatial resolution readout schemes, a sub-millimeter spiking e-skin scheme is modeled and compared to its frame-based counterpart in terms of power consumption and texture classification accuracy, using a Spiking Neural Network. Our analysis shows that the sparsity-inducing spike-based solutions achieve one order of magnitude lower power consumption while reaching a higher classification accuracy (87.92%) compared to the frame-based readout (74.58%).**

***Index Terms*—electronic skin, spiking readout, level crossing**

I. Introduction

In recent years, fine-grain tactile sensing through *electronic skins* (e-skins) has gained much attention in applications such as *object sensing* for humanoid robotics [1]. Indeed, contact parameters such as roughness, texture, weight, and slip are difficult to infer from mere visual data, demanding the integration of *high-resolution* e-skins on robotic arms (see Fig. 1 a). The spatial resolution of current e-skin sensors, however, still falls short of the ~ 0.4 mm resolution found in human skin [2]. Future prosthetic hand applications would likely require human-like sub-millimeter e-skin resolutions.

In the past years, research in e-skins has principally been devoted to the development of tactile sensors with readout electronics mainly implemented using off-the-shelf components having over-designed precision and speed [3]–[5]. This results in readout circuits suffering from large area and power consumption, which becomes an even greater issue when targeting sub-millimeter-resolution, high-density e-skin.

In order to cope with the fast sensor array scan rate (1-10 kHz) required to detect fine spatio-temporal stimuli [4], the skin sensors in each array are either polled sequentially (*i.e.*, one by one) [3], [5] or in a column-based fashion [4], [6]. As shown in Fig. 2 a) for *conventional* readouts, the analog sensor output is multiplexed onto a central ADC for conversion, producing synchronous frames of tactile information that

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 861166.

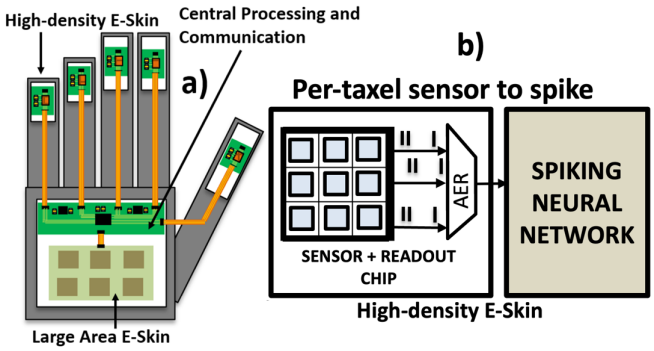

Fig. 1: *High-density electronic skin: a) Equipping a robotic hand with a combination of a large area e-skin with low spatial resolution e-skin on the fingertips allows classification of macro and micro tactile features. The chip-scale e-skin at the fingertips is connected to a central board with SNN hardware. b) The high-density e-skin converts sensor signals to spikes at taxel level.*

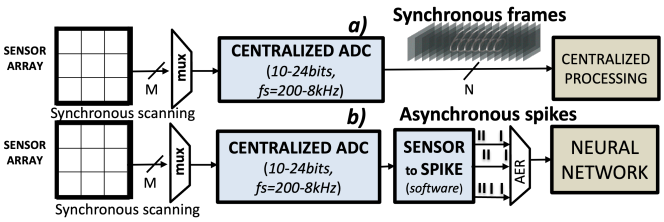

Fig. 2: *State-of-the-art electronic skin readouts can be classified in two major categories: a) synchronous frame generation using a central ADC and further digital processing; b) asynchronous spike generation in software followed by a spiking neural network.*

are periodically sent for processing. This high-rate, periodic sensing and transmission of the tactile frames is, however, power-inefficient, despite the fact that tactile stimuli may be sparse in time (this *sparsity* typically being ignored in the conventional frame-based readout of Fig. 2 a).

Therefore, in contrast to previously-presented *frame-based* e-skin readouts, the use of neuromorphic, *spike-based* e-skin readout circuits is investigated in this paper. As opposed to previously studied spiking e-skins, where the sensor to spike conversion is performed in software after ADC acquisition [7]–[9] (see Fig. 2 b), a high-density e-skin readout with taxel-wise sensor to spike conversion implemented directly in hardware (*i.e.*, on-chip) is modeled, resulting in a more power-

efficient design.

We also study the impact of the sensor to spike conversion data rate on the *texture classification accuracy*, by feeding texture signals from the popular *Kylberg* texture dataset [10] into our neuromorphic e-skin readout. In order to perform spike-based classification, we use a recurrent *Spiking Neural Network* (SNN) trained using the *Surrogate Gradient* method [11] and achieving a top accuracy of 87.92%.

The paper is organized as follows. The *spiking* taxel circuits studied in this paper are presented in Section II. A comparison between frame-based and spiking readouts in terms of power consumption and texture classification accuracy is modeled in Section III. Conclusions are drawn in Section IV.

II. Taxel Circuit Design

A. Conventional Level-Crossing Sampling

Prior work has shown that event-based sensing of time-sparse input signals greatly helps making the readout electronics more power-efficient, which scales well with the ever increasing sensor counts [12]. In addition, it has been shown that deploying individual taxels that respond independently to the stimuli results in an efficient scaling between the actual information carried by the stimuli and the data transmission rate at the output of the sensor [13].

Therefore, *Level-Crossing Sampling* (LCS) converters [14] have been used as a straightforward way for converting the input signals into event-based spike trains. As shown in Fig. 3, the LCS converter generates a spike sample every time the input exceeds a certain threshold. Then, the sequence of level crossings across multiple taxels is encoded as a sequence of the taxel addresses using the popular *Address Event Representation* (AER) interface, as shown in Fig. 1 b) [13].

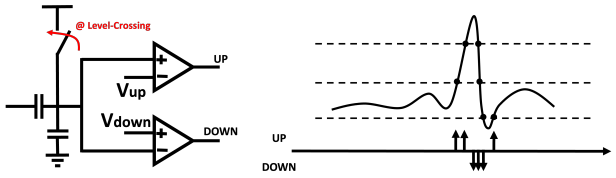

Fig. 3: *Analog to Spike Conversion using a non-uniform sampling/level-crossing converter.*

B. Neuromorphic Level-Crossing Sampling

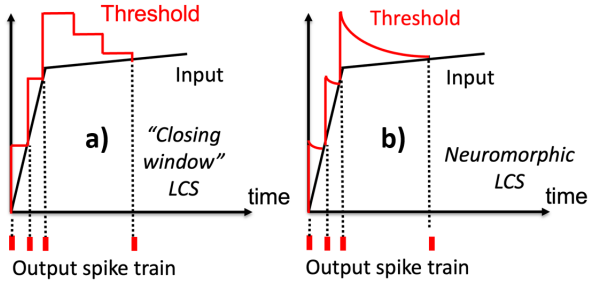

Fig. 4: *The Closing Window Level-Crossing Sampling in a) approximates how a leaky neuron in b) generates spikes.*

A common limitation with previously proposed LCS designs is the lack of level crossings during a slow-moving input. Such behavior is expected as LCS converters are designed to detect the signal change and not the DC component. For signals that are sparse in time, the lack of level crossings results in a reduced signal conversion accuracy. Increasing the ADC resolution N to also detect the small signal changes requires a $\sim 2^{2N}$ increase in power consumption [15]. Therefore, solutions such as adding a triangular dither [16] or an adaptive threshold [17] have been proposed.

In this paper, we explore a novel approach to the problem of sampling slow tactile inputs by borrowing insights from neural encoding. Instead of a *step-wise* decreasing threshold used in common LCS circuits [17], we propose to explore the use of an *exponentially-decaying* threshold instead, as a more faithful model for biological neural encoding [18], [19]. Fig. 4 compares our proposed exponentially-decaying threshold to the common *step-wise* dynamics.

III. Results

A. Power Consumption Analysis of Spikes vs. Frames

In this Section, the power consumption analysis for spike-based readout circuits is provided when multiple input channels must be sensed and converted. Prior analysis has shown the increase in power efficiency obtained when spike-based readouts are used [15], but these previous results assumed a single input channel with no *analog front end* (AFE) stage. For small sensor signals with high output impedance, such as the piezoelectric sensor used in this application, an AFE with gain and signal conditioning is needed.

In order to compare our event-based readout techniques against the frame-based counterpart in a multi-channel setting, detailed power consumption models of the following readout circuits are implemented: *i)* the spiking readout with regular LCS taxels and without a closing window; *ii)* the spiking readout with the proposed *neuromorphic* LCS taxels (see Section II-B) and *iii)* a *conventional* frame-based readout with sensors multiplexed into a single ADC (with a typical figure of merit of around 10 fJ per conversion step). Fig. 5 illustrates these models, assuming a 64-taxel array, which is comparable to state of the art large area e-skins, at the input of each model.

Fig. 6 shows the result of this power consumption analysis. At low signal-to-noise ratio (SNR) (right bars in Fig. 6), we observe that the *multi-channel* spike-based readout with *AFE modelling* is more power-efficient than the conventional frame-based counterpart. This result is in line with prior *single-channel* analysis, where AFE power consumption was omitted [15].

On the other hand, at *high* SNR (left bars in Fig. 6), the power consumption of the *per-taxel* AFE in *spike-based readouts* dominates the system power consumption, making the spike-based solution less power-efficient compared to the frame-based counterpart, despite the savings in transmission power due to the *compressing* transfer of spiking data. In addition, Fig. 7 shows the allowed RMS value of the AFE output noise (cross-over point) as a function of the equivalent

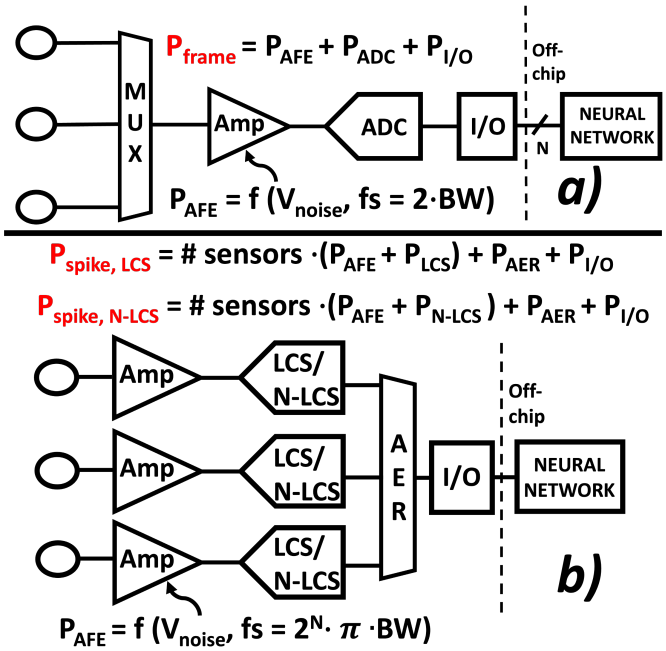

Fig. 5: *Blocks considered for power consumption modelling: In contrast to a frame-based readout in a) where the analog front-end (AFE) circuit is shared among all sensors, the spike-based readout in b) requires one AFE per taxel.*

bit resolution. Increasing the bit resolution in both LCS and N-LCS limits the minimum AFE output noise allowed to maintain a lower system power consumption than a frame-based readout.

Therefore, the analysis in Fig. 6 shows that the use of spike-based readout is mainly suitable in *challenging* signal acquisition scenarios where the input SNR is low.

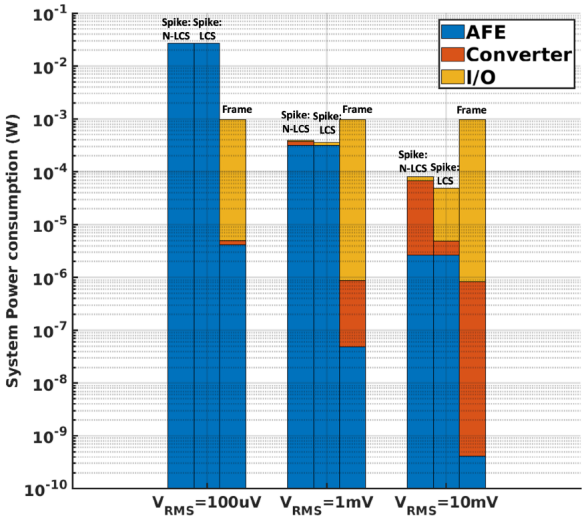

Fig. 6: *System power consumption breakdown for spike- and frame-based readouts, at three different AFE noise RMS target.*

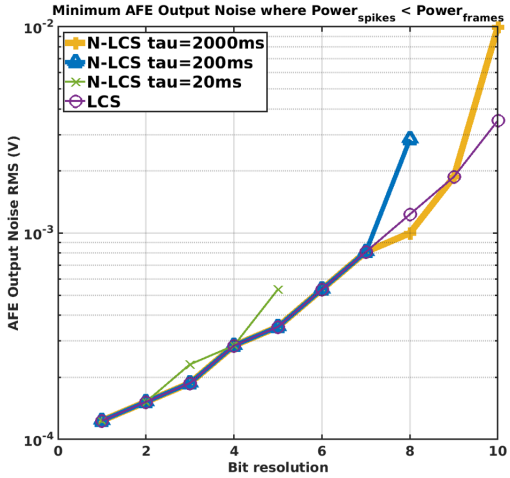

Fig. 7: *Minimum AFE output noise above which the spiking readout is more power efficient than the frame-based one as a function of the resolution. The lines halting below 10 bits mean that the frame-based power efficiency is better from that bit resolution onward. Plots for different threshold decay constants tau (Fig. 4) are shown.*

B. Texture Recognition Experiments

Next, the impact of the readout spiking data rate on the texture recognition accuracy is investigated, by connecting the spiking output of our tactile skin readout to a recurrent SNN. We use the popular 6-class *Kylberg* texture dataset [10] as input to a 64-taxel readout with 200 μm spatial resolution. Texture scanning is simulated by assembling the pixel values from left to right into a continuous signal (see Fig. 8 a).

As SNN architecture, we consider a 1-hidden-layer recurrent SNN with 256 *Integrate and Fire* (IF) neurons [11], followed by spike accumulation and a *SoftMax* output layer (see Fig. 8 c). This choice is motivated by the fact that *i)* using a low-complexity 1-hidden-layer network enables a fair analysis of the front-end hardware parameter effect, and *ii)* such resource-constrained network can be directly deployed on existing ultra-low-power SNN accelerators [14], [20].

The SNN training has been conducted during 500 epochs using the *Adam* optimizer [21] with learning rate 10^{-3} and batch size 32. The classification accuracy and standard deviation is assessed using 6-fold cross-validation [11]. Fig. 9 shows the texture classification accuracy and standard deviation as a function of the spiking data rate of our e-skin readout. Similarly, we also train a standard RNN with *ReLU* neurons on the non-spiking e-skin scanning data (see Fig. 8 a), using the same architecture as the SNN.

Three interesting effects can be observed in Fig. 9. First, we see that the system accuracy is highly dependent on the spiking data rate, set by adjusting the bit resolution of the spike encoder, as shown in Fig. 8 b). As expected, the accuracy increases when the data rate increases since less signal information is lost. On the other hand, we observe a decrease in accuracy when the data rate keeps growing. This is due to the fact that having too many spikes can jeopardize the SNN training during *Surrogate Gradient* back-propagation,

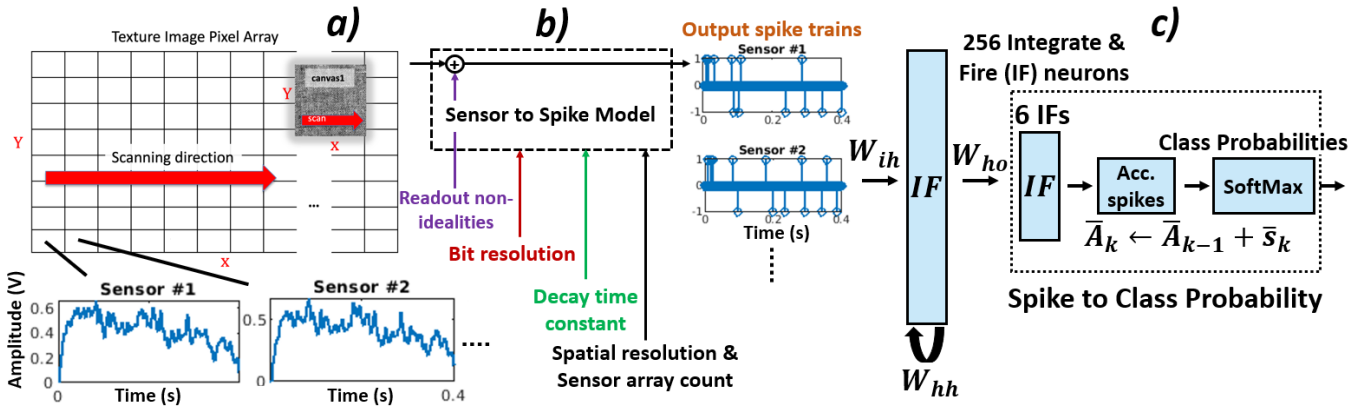

Fig. 8: *Spiking e-skin texture classification pipeline a) Time-varying sensor signals are obtained from the texture image dataset by simulating a texture scan. b) the spiking taxel conversion circuit model allows the sweeping of conversion parameters such as the bit resolution, c) The spike trains are fed to a recurrent SNN through the fully-connected weight matrix W_{ih}. Recurrence is introduced by the weights W_{hh}. The output layer consists of 6 IF neurons connected through W_{ho}, the spiking output $\bar{s}_k$ of which at time step k is accumulated into $\bar{A}$. $\bar{A}$ is transformed to class probabilities using the SoftMax operator.*

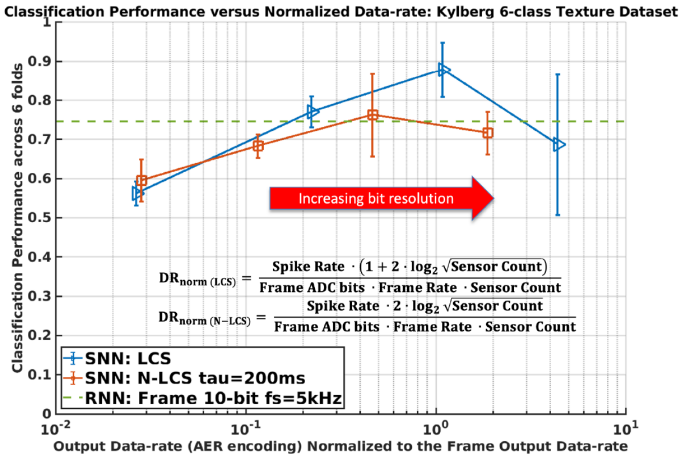

Fig. 9: *Texture classification accuracy in function of skin readout data rate. The data-rate for the spiking readout is given by the total spike rate multiplied by the row and column address bits. The LCS uses an extra bit for the spike polarity.*

leading to a decrease of model accuracy [11].

Secondly, it is interesting to note that the spike encoding and SNN pipelines (blue and orange curves in Fig. 9) lead to a higher accuracy compared to the baseline RNN for normalized data rates $> 10^{-1}$ in the *sub-Nyquist* region. This can be due to the sparsity in the spiking signals, which helps increasing the linear separability between data samples by providing an early compression to the sensory signals [11].

Lastly, it is shown that the N-LCS has a slightly lower accuracy compared to the LCS. However, it is important to note that the texture classification performed here uses fast sensor signals as input whereas, the N-LCS' inherent advantage is with slow input, as discussed in Section II-B. Future work on the comparison of the classification accuracies between the LCS and N-LCS using datasets with low-bandwidth sensor signals shall be performed.

Since the system power consumption scales with the spiking data rate, we see that the *spiking* e-skin readouts and SNN classifier can lead to an overall lower power consumption compared to the conventional RNN-based setup, while achieving a higher texture classification accuracy for normalized data rates between 10^{-1} and 1. A top accuracy of 87.92% is observed using the *LCS* signal-to-spike converter, compared to 74.58% for the frame-based readout.

IV. Conclusion

This paper has presented the modelling of a novel chip-scale, high-density e-skin solution targeting sub-millimeter spatial resolution. Compared to conventional frame-based systems, spike conversion is performed at the taxel level for reduced power consumption and direct compatibility with SNN processors. It has been shown that due to the per-taxel nature of the analog front end blocks, their consumption becomes dominant at higher SNR targets, meaning that the spike-based readout has higher power efficiency than the frame-based solution for lower SNR values. Finally, the impact of the spiking readout on texture recognition accuracy has been studied, achieving a superior accuracy of 87.92% over the frame-based counterpart. These investigations clearly indicate the advantages of using spiking readouts for high-density e-skin systems. While this work presents only the results of modeling the proposed e-skin spiking readout, the high-density spiking readout, with LCS and N-LCS taxels, has been designed and implemented on a standard sub-micron CMOS process.

Acknowledgment

The authors would like to thank L. Lorenzelli, A. Adami, F. Giacomozzi and I. Temel of *Fondazione Bruno Kessler*, Henrik Jörntell of *Lund University*, and C. Frenkel of *TU Delft* for sharing their insights.

REFERENCES

[1] Bhirangi, Raunaq Hellebrekers, Tess Majidi, Carmel Gupta, Abhinav. (2021). ReSkin: versatile, replaceable, lasting tactile skins.

[2] Jarocka, E., Pruszynski, A., Johansson, Roland. (2021). Human Touch Receptors Are Sensitive to Spatial Details on the Scale of Single Fingerprint Ridges. The Journal of Neuroscience. 41. JN-RM. 10.1523/JNEUROSCI.1716-20.2021.

[3] 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.

[4] 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.

[5] 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.

[6] Rasouli, M. Chen, Y. Basu, A. Kukreja, S., Thakor, N.v. (2018). An Extreme Learning Machine-Based Neuromorphic Tactile Sensing System for Texture Recognition. IEEE Transactions on Biomedical Circuits and Systems. PP. 1-13. 10.1109/TBCAS.2018.2805721.

[7] Taunyazov, T., Sng, W., See, H., Lim, B., Kuan, J., Ansari, A., Tee, B., Soh, H. (2020). "Event-Driven Visual-Tactile Sensing and Learning for Robots," in Proceedings of Robotics: Science and Systems, Jul. 2020.

[8] D. A., M. M., N. A., C. E., V. M. and B. C., "Artificial Bio-Inspired Tactile Receptive Fields for Edge Orientation Classification," 2021 IEEE International Symposium on Circuits and Systems (ISCAS), 2021, pp. 1-5, doi: 10.1109/ISCAS51556.2021.9401749.

[9] Rongala, U. Mazzoni, A. Spanne, A., Jörntell, H., Oddo, C. (2019). Cuneate spiking neural network learning to classify naturalistic texture stimuli under varying sensing conditions. Neural Networks. 123. 10.1016/j.neunet.2019.11.020.

[10] Kylberg, G.. "The Kylberg Texture Dataset v. 1.0," Centre for Image Analysis, Swedish University of Agricultural Sciences and Uppsala University, External report (Blue series) No. 35. Available online at: http://www.cb.uu.se/ gustaf/texture/ and https://kylberg.org/datasets/

[11] A. Safa et al., "Improving the Accuracy of Spiking Neural Networks for Radar Gesture Recognition Through Preprocessing," in IEEE Transactions on Neural Networks and Learning Systems, doi: 10.1109/TNNLS.2021.3109958.

[12] T. Finateu et al., "5.10 A 1280×720 Back-Illuminated Stacked Temporal Contrast Event-Based Vision Sensor with 4.86μm Pixels, 1.066GEPS Readout, Programmable Event-Rate Controller and Compressive Data-Formatting Pipeline," 2020 IEEE International Solid-State Circuits Conference - (ISSCC), 2020, pp. 112-114, doi: 10.1109/ISSCC19947.2020.9063149.

[13] K. A. Boahen, "Point-to-point connectivity between neuromorphic chips using address events," in IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, vol. 47, no. 5, pp. 416-434, May 2000, doi: 10.1109/82.842110.

[14] Y. He et al., "A 28.2 μC Neuromorphic Sensing System Featuring SNN-based Near-sensor Computation and Event-Driven Body-Channel Communication for Insertable Cardiac Monitoring," 2021 IEEE Asian Solid-State Circuits Conference (A-SSCC), 2021, pp. 1-3, doi: 10.1109/A-SSCC53895.2021.9634787.

[15] J. Van Assche and G. Gielen, "Power Efficiency Comparison of Event-Driven and Fixed-Rate Signal Conversion and Compression for Biomedical Applications," in IEEE Transactions on Biomedical Circuits and Systems, vol. 14, no. 4, pp. 746-756, Aug. 2020, doi: 10.1109/TBCAS.2020.3009027.

[16] Wang, T., Wang, D., Hurst, P., Levy, B., Lewis, S. (2009). "A Level-Crossing Analog-to-Digital Converter With Triangular Dither." Circuits and Systems I: Regular Papers, IEEE Transactions on. 56. 2089 - 2099. 10.1109/TCSI.2008.2011586.

[17] 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, doi: 10.1109/JSSC.2013.2262738.

[18] Gerstner, Wulfram. (2008). Spike-response model. Scholarpedia. 3. 1343. 10.4249/scholarpedia.1343.

[19] 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.

[20] C. Frenkel, M. Lefebvre, J. -D. Legat and D. Bol, "A 0.086-mm^2 12.7-pJ/SOP 64k-Synapse 256-Neuron Online-Learning Digital Spiking Neuromorphic Processor in 28-nm CMOS," in IEEE Transactions on Biomedical Circuits and Systems, vol. 13, no. 1, pp. 145-158, Feb. 2019, doi: 10.1109/TBCAS.2018.2880425.

[21] Kingma, D., Ba, J.. (2014). "Adam: A Method for Stochastic Optimization."