End-to-End Optimization of High-Density e-Skin Design: From Spiking Taxel Readout to Texture Classification

Jiaqi Wang[1*], Mark Daniel Alea[1*], Jonah Van Assche[1], Georges G.E. Gielen[1,2]
ESAT-MICAS, KU Leuven, Kasteelpark Arenberg 10, 3001 Leuven (Heverlee), Belgium
email:jiaqi.wang@kuleuven.be
[1]*ESAT, KU Leuven,* [2]*imec, 3001, Leuven, Belgium*
{Jiaqi.Wang,Mark.Alea,Jonah.Vanassche,Georges.Gielen}@kuleuven.be
*These authors contributed equally

***Abstract*—Spiking readout architectures are a promising low-power solution for high-density e-skins. This paper proposes the end-to-end model-based optimization of a high-density neuromorphic e-skin solution, from the taxel readout to the texture classification. Architectural explorations include the spike coding and preprocessing, and the neural network used for classification. Simple rate coding preprocessing to spiking outputs from a modeled low-resolution on-chip spike encoder is demonstrated to achieve a comparable texture classification accuracy of 90 % at lower power consumption compared to the state of art. The modeling has also been extended from single-channel sensor recording to time-shifted multi-taxel readout. Applying this optimization to an actual tactile sensor array, the classification accuracy is boosted by 63 % for a low-cost FFNN using multi-taxel data. The proposed Spike-based SNR (SSNR) and Spike Time Error (STE) metrics for the taxel readout circuitry are shown to be good predictors of the accuracy.**

***Index Terms*—electronic skin, spiking readout, neuromorphic, end-to-end optimization, spike encoding metrics**

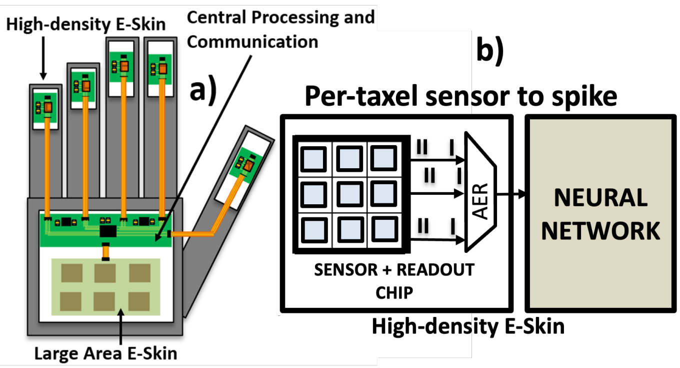


Fig. 1: *High-density electronic skin: a) Equipping a robotic hand with a combination of a large-area e-skin with high-spatial resolution e-skin on the fingertips allows classification of macro and micro tactile features. The chip-scale e-skin at the fingertips is connected to a central board with Neural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

I. Introduction

To enable next-generation robotic and prosthetic applications, artificial hands require fine-grained sensing modalities, such that they can sense and analyze their environment in a human-like fashion. As contact parameters during object manipulation are hard to infer from mere visual data, high-resolution tactile sensing is required [1]. Distributed tactile sensors fulfill the same function as the skin does for humans and therefore is also referred to as *electronic skin* (e-skin), as shown in Fig. 1. Such e-skin consists of an array of sensors that measure parameters such as the normal/shear force, humidity, and temperature. Using data collected from these sensors, a processing unit must classify the objects with which the robotic hand is interacting. In this paper, the focus is on normal force sensing for texture classification of objects through capacitive sensors – chosen because of their simple structures and readout electronics, and their ability to detect static force [2].

To mimic the functionality of human skins, the e-skin requires the monitoring and transmitting of information from thousands of distributed sensing elements. For conventional readouts (shown in Fig. 2(a)), the analog sensor output is multiplexed onto a central ADC for conversion, producing synchronous frames of tactile information that are periodically sent for processing [3]. This high-rate, periodic sensing and transmission of the tactile frames is, however, power-inefficient [4], even though tactile stimuli may be sparse in time. In state-of-the-art e-skins [5], [6], the communication bandwidth is reduced by encoding the sensor data into sparse "events" – referred to as *spikes* – which are then transmitted asynchronously. Such event-based encoding of sensor data, however, is typically performed at the software level. Therefore, while an event-based approach is used on the data transmission, the sensor data are still converted in a frame-based approach, as shown in Fig. 2(b). State-of-the-art e-skin solutions often use off-the-shelf components with overdesigned specifications, which often results in a large system power consumption [5], [7] and a low spatial resolution [8], which prohibits fine-grained tactile sensing of objects. An interesting approach to tactile sensing is to use neuromorphic sensors with sensor-to-spike encoding performed at hardware level [9], as shown in Fig. 2(c). This neuromorphic system offers a lot of potential for high-density e-skin applications with high spatial resolution and low system power consumption.

978-3-9819263-7-8/DATE23/© 2023 EDAA

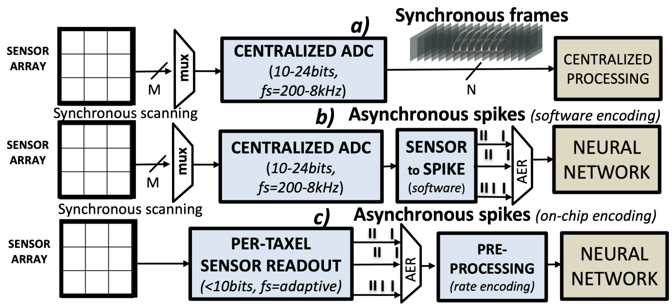
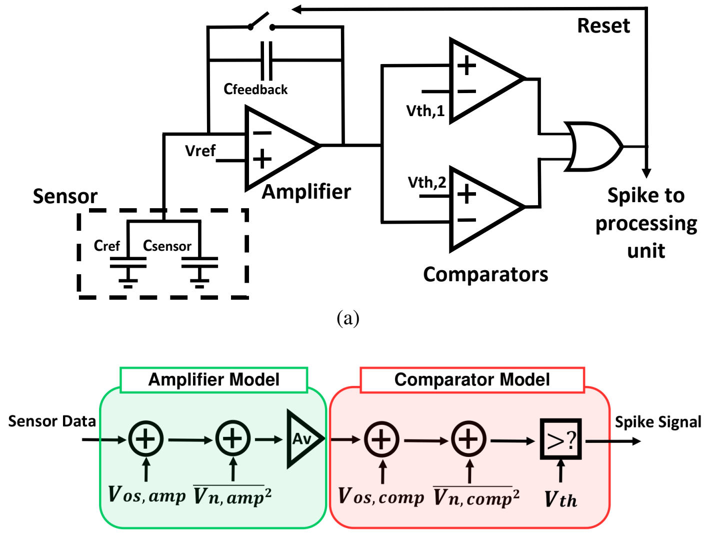


Fig. 2: *State-of-the-art electronic skin readouts can be classified in three major categories: a) synchronous frame generation using a central ADC and further digital processing; b) asynchronous spike generation in software followed by a spiking neural network; c) asynchronous spike generation performed on-chip at per-taxel level. The spike count is accumulated to form the rate code and be compatible with the frame-based input of the neural network*



(b)

Fig. 3: *a) Proposed capacitive tactile sensor with spiking readout circuit, and b) high-level model for system simulations.*

This paper describes the end-to-end optimization of such a neuromorphic e-skin design, from the taxel readout to the texture classification. First, in Section II, a model for the spiking neuromorphic taxel readout is derived. This model includes the effect of hardware non-idealities and the encoding method of the spikes. Then, two metrics for the accuracy of the taxel readout are proposed: the Spike-based SNR (SSNR) and the Spike Timing Error (STE). In Section III, the modeling of a taxel array is discussed, as well as the preprocessing method used to feed the spikes to the processing algorithms, *i.e.* both a convolutional neural network (CNN) and a feedforward neural network (FFNN). Section IV then discusses the results of the e-skin parameter optimization and compares its performance to prior work on texture recognition. Section V concludes the paper.

II. Spiking Readout Model and Metrics

A. Continuous-time Capacitive Readout Model

The capacitive readout model simulates the spike train generation due to the changes in the normal force transduced by the taxel. The continuous-time (CT) readout circuit is composed of a front-end circuit and comparators, as shown in Fig. 3(a). The choice of this architecture is because of its simple architecture and its CT nature, allowing for asynchronous spike generation. A change in the capacitive taxel output corresponds to a change in the output voltage of the frontend circuit. Such change results in spikes when the corresponding change exceeds a pre-defined threshold value. The signs of the spikes are decided by the direction of the output voltage change.

The readout circuit and its spike generation have been modeled in MATLAB. The model accepts an array of discrete samples as input, representing the transient behavior of the C_{sensor} output. Readout parameters, such as the frontend circuit gain and the threshold, can be configured. The spike train output is stored as an array of time stamps representing the spike times. Readout circuit non-idealities, such as the amplifier offset, the amplifier noise, the comparator offset, and the comparator noise can be added to the model, as shown in Fig. 3(b).

B. Generating Multi-Taxel Signals from Single-Channel Recordings

A key limitation with publicly-available tactile datasets of texture recording is that they mostly contain sensor signals recorded from a single channel. Therefore, this limits classification algorithms that use recordings from multiple sensors in an array. Previous modeling of human fingertips [10], [11] provides evidence that a population of neurons cooperates to convey tactile sensory information during texture scanning. Therefore, in this paper, a methodology to generate multi-taxel signals from a single channel recording during texture scanning is presented (Fig. 4). To simulate a texture being scanned across the sensors in an array, time delays are introduced to the simulated sensor outputs across the y-axis (see Fig. 4). The unit time delay t_{delay} is proportional to the ratio of the spatial resolution and the texture scanning speed. To generate simulated sensor output across the x-axis (see Fig. 4), the texture profile across the x-axis is assumed to have the same profile across the y-axis. The multi-taxel signal outputs are then individually used as input to the taxels.

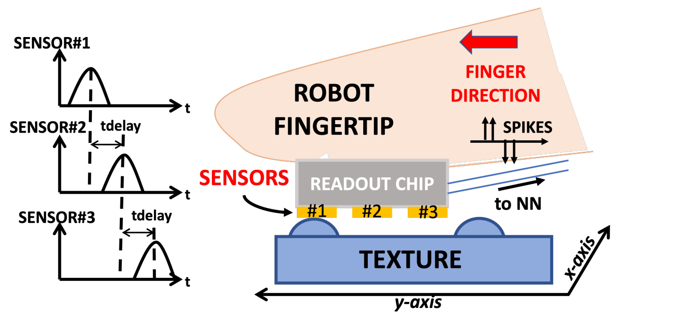


Fig. 4: *To simulate multi-taxel signals from single-channel recordings, time delays are introduced. The time delays model the delay in which consecutive sensors see the same texture profile during finger scanning.*

C. Error Metrics

While for a conventional sensor readout system, well-defined accuracy metrics exist for hardware designers, such as the Signal to Noise Ratio (SNR) and Signal to Error Ratio (SER), there exist no widely-accepted metrics for readouts with spike-based output [12]. Moreover, since the goal of such systems typically is to be directly processed by classification algorithms, the question is which accuracy metrics at the sensor level can predict well the classification performance of the processing algorithm.

The proposed spike encoding metrics are based on the difference between the REAL spike train generated from the spike encoder model with non-idealities and the TRUE spike train generated by the ideal spike encoder. Therefore, the two metrics quantify the random and systematic variations on the generated spike trains due to the readout non-idealities. Such spike train variations typically degrade the classification accuracy due to overfitting. An example of the generation of false spikes by a *noisy* spike encoder model can be seen in Fig. 5.

The following two error metrics are proposed:

1) Spike-based SNR: The SSNR metric is a rate coding-based metric since it only deals with the deviation in the spike rate due to the non-idealities. As shown in equation (2), the SSNR is defined as the ratio of the number of TRUE spikes to the number of FALSE spikes, which equals the difference between REAL spikes and TRUE spikes:

$$SSNR = 20 * log\frac{N_{true\ spike}}{abs(N_{real\ spike} - N_{true\ spike})}[dB] \quad (2)$$

2) Spike Timing Error: The STE metric is more related to the spike temporal code. It measures differences in the spike rates but also quantifies the deviation in the spike timing due to the non-idealities. The closest FALSE spike to each TRUE spike is found. As shown in equation (3), the standard deviation of the time difference Δt depicts how dispersed Δt is. To eliminate the effect of the number of spikes on the time difference, the standard deviation of the spike times is normalized by multiplication with the number of spikes. The STE is defined as follows:

$$STE = \frac{\sum(\Delta t - \overline{\Delta t})^2}{N_{real\ spike}} * (N_{real\ spike} - N_{true\ spike})[s^2] \quad (3)$$

III. Texture Classification using Neural Networks

A. Texture Dataset

A 12-class publicly-available tactile dataset of texture recordings, VIBTAC12 [13], is used as sensor input. Each texture class contains 10 unique accelerometer recordings. Due to the lack of available texture datasets recorded with a tactile capacitive sensor, the VIBTAC12 dataset is used instead. To model the normal force captured by a capacitive sensor, the z-axis output of the three-dimensional accelerometer data in the VIBTAC12 dataset is selected.

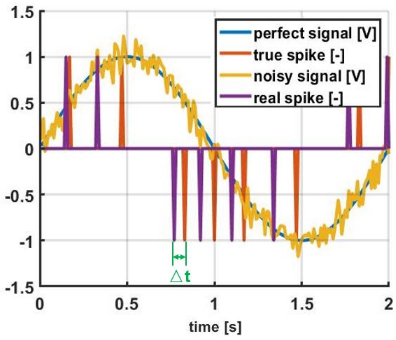


Fig. 5: *The TRUE spike train is generated with the ideal readout circuit model while the REAL spike train is generated while considering readout non-idealities*

B. Spike Pre-processing

Feedforward neural networks (FFNN) and convolutional neural networks (CNN) are applied to perform texture classification tasks. However, since both FFNN and CNN accept frames as input instead of spikes, the spatio-temporal spike trains to be generated are pre-processed to generate a frame, as shown in Fig. 6. For every sensor, the spike train is divided into time windows. The number of spikes during each time window is then accumulated, forming a matrix of the instantaneous rate codes across time windows. The rate code matrices of individual taxels are then concatenated in a manner that preserves the spatial location of individual taxels. This is done by concatenating the rate code matrices of neighboring taxels.

C. Neural Network Implementation

The size at the input layer depends on the number of taxels $\times$ the number of time windows. In our experiments, we use 9 taxels and 40 time windows. Meanwhile, the number of output neurons depends on the number of texture classes (12 in our example). Both networks consist of *ReLU* neurons at the hidden layer, followed by spike accumulation and a *Softmax* output layer [9]. The CNN has two convolution layers and two *2x2* pooling layers. The first convolution layer has 16 output channels, while the second layer has 32. Both convolution layers use *3x3* kernels and padding=1.

Hyperparameter tuning yields optimum values of the training epochs (100), the learning rate ($1e^{-3}$), and the number of neurons in the hidden layer (100). A fair comparison is guaranteed by application of the same hyperparameters to all the FFNN and CNN classifications performed in this paper . The mean classification accuracy and the standard deviation are obtained using a 10-fold cross-validation [14]. The *Adam* optimizer [15] is used to train the network.

IV. Optimization Results

A. Optimizing e-Skin Parameters

Enabled by the end-to-end e-skin model, it is possible to tune design parameters related to both the spiking readout and the rate coding preprocessing and to predict its immediate effect on the texture classification performance. In contrast to current

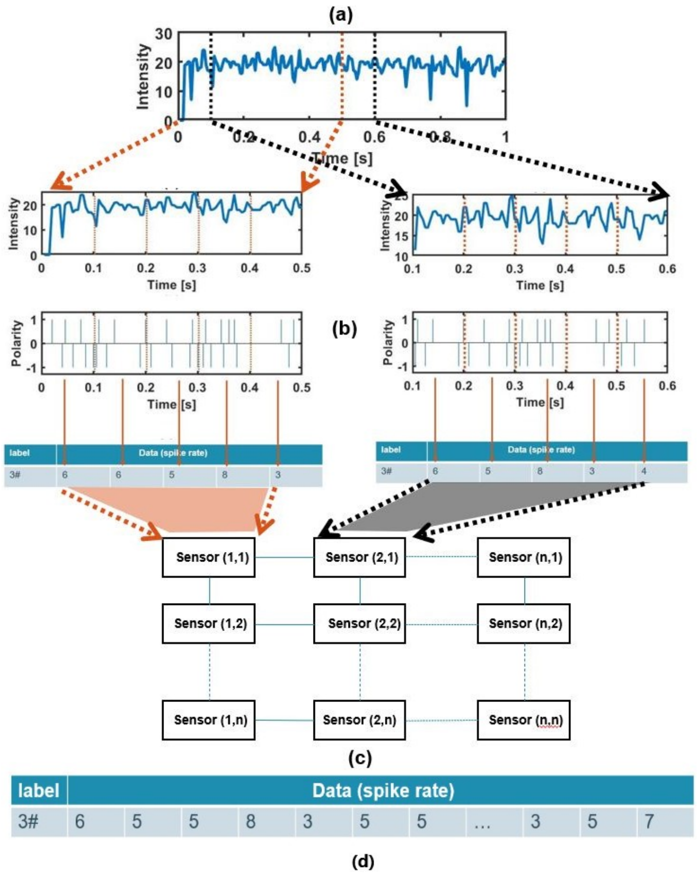


Fig. 6: *Rate encoding preprocessing of spikes: a) single-channel recording is used to simulate multi-taxel signals through time delays; b) spike train output generated by the spike encoder model is divided into time windows; c) spikes are accumulated per window while discarding the polarity, forming an instantaneous rate code across several windows; (d) multi-taxel frame information is assembled while preserving the spatial location of each sensor.*

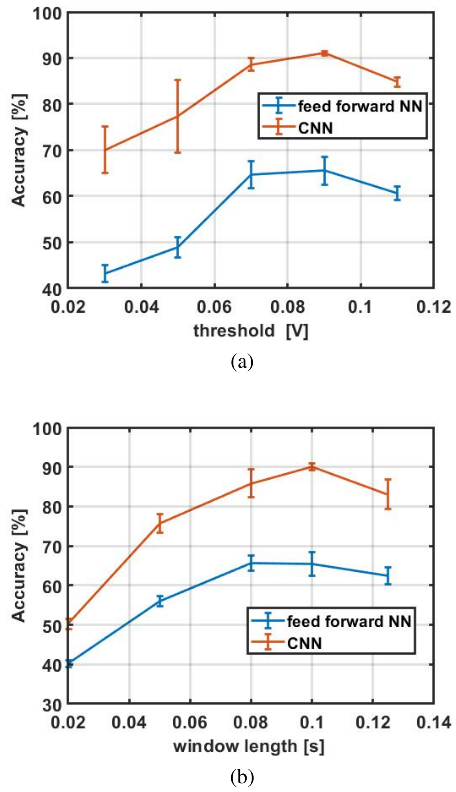


Fig. 7: *Classification accuracy versus (a) the spiking readout threshold and (b) the preprocessing time window length.*

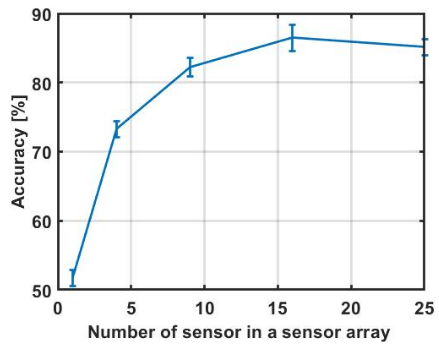


Fig. 8: *Accuracy vs number of sensors in a sensor array for feed forward neural network. Spike train is generated by continuous-time capacitive readout model.*

e-skin designs where off-the-shelf components are used, an e-skin pipeline designed on a custom ASIC offers significantly more design freedom and consumes less power. Guided by the end-to-end model optimization, this freedom allows the design of high-density e-skins with relaxed readout requirements and simpler, hardware-friendly spike preprocessing, as will be discussed below.

1) Spiking Readout: The parameters related to the spike encoder model, such as the threshold voltage and the number of sensors, are tuned. As shown in the texture classification accuracy graph of Fig. 7(a), a sweep of the threshold voltage shows an optimum threshold value of around 0.09V for both the FFNN and the CNN. This translates into an equivalent bit resolution of around 3.5 bits, which is significantly more relaxed than the commercial ADCs used in state-of-the-art e-skin (10-24 bits) [5], [11]. Interestingly, a further threshold/resolution decrease translates again into diminishing accuracy. A possible explanation is that having too many spikes can jeopardize the back-propagation, as seen in [9].

The effect of the number of sensors on the classification accuracy is also modeled, as shown in Fig. 8. An increasing number of correlated taxels (time-shifted only) leads to more input data for the neural networks, thus to better training results and higher classification accuracy. By increasing the number of taxels from one to 16, the accuracy increased by 63%. Interestingly, the classification accuracy saturates at >16 taxels to about 90%. Possibly, the increase in the input layer size due to the increased number of taxels results in a more complex network model, making it less robust to generalization and more prone to overfitting.

2) Pre-Processing: An important pre-processing parameter – the time window length – is also tuned, as shown in Fig. 7(b). A shorter time window results in more time windows, therefore, increasing the input layer size of the NN. Meanwhile, a longer time window translates into a reduction in the temporal information carried by the rate code matrix. This is because the accumulation of spikes per time window inherently results in the loss of fine temporal information carried by the spike train, as the individual spike times of accumulated spikes inside a window are discarded. For both the FFNN and the CNN, a time window length of 0.1 seconds results in the maximum classification accuracy. Interestingly, reducing the size of the window length results in a lower accuracy. Similar to what is seen in Section IV-A1, the increase in the input layer size due to a shorter time window length results in a more complex network model that is more prone to overfitting.

B. Correlation between Metrics

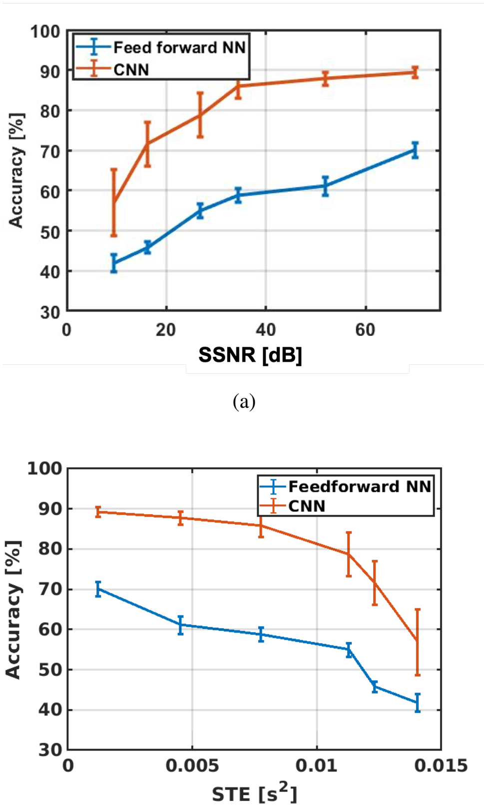
As discussed in Section II-C, an important contribution of this paper is to derive signal encoding metrics that predict well the classification performance. A single taxel with a constant threshold is tested to investigate the correlation between the proposed spike-based error metrics and the traditional texture classification accuracy. As shown in Fig. 9 (a) and (b), there is a strong positive correlation between the classification accuracy and the SSNR, while the accuracy and STE have a strong negative correlation. Such correlation is expected as the SSNR measures the robustness of the rate code to the noise, while the STE quantifies the amount of noise on the spike train through variations in the spike timing. Through the proposed metrics, an alternative way of validating the signal encoding fidelity of e-skin designs is developed, independent from the neural network hyperparameters and its complex tuning.

C. Multi-Taxel Readout Consideration

As seen in Section IV-A1, the number of taxels in the array helps to increase the classification accuracy. However, this characterizations assumes an ideal, uniform taxel response. As seen in Fig. 10, introducing mismatch on the taxels results in a reduction in the classification performance. The taxel non-uniformity is modeled as mismatch on the thresholds due to, for instance, random intra-die variations, resulting in some sort of fixed pattern noise (FPN). To reduce the mismatch, the taxel sizing will have to be increased, possibly limiting both the minimum spatial resolution and the number of taxels achievable while maintaining a robust classification performance.

D. Comparison with State-of-the-art e-Skins

The resulting design is compared with state of art e-skin designs for texture classification in Table I. A key insight is that, through an end-to-end optimization coupled with the design freedom of an ASIC design, the e-skin achieves a comparable texture classification accuracy of 90% despite a low precision spike encoder of 3.5 bits. Furthermore, the data rate required to transmit information is reduced by up to three orders of magnitude (compared to [18]) due to the event-based spike encoding. Lastly, a hardware-friendly rate code preprocessing through spike accumulation is performed, in contrast to complex sofware-based feature extractions [18], [19]. While a higher classification accuracy is seen in cited work, it is to be noted that previous work uses less texture classes for classification.



(a)

(b)

Fig. 9: *Classification accuracy versus (a) Spike-based SNR (SSNR) and (b) Spike Timing Error (STE)*

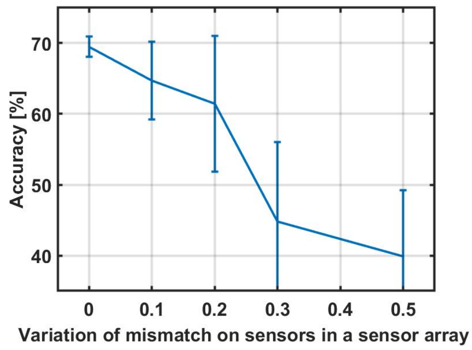


Fig. 10: *Classification accuracy versus the standard deviation of the mismatch between taxels.*

V. Conclusions

This paper has presented the optimization of a high-density e-skin design, from the spiking taxel readout to the spike preprocessing and the texture classification hardware (FFNN and CNN). The model simulates and generates multi-taxel data from a single channel sensor recording, increasing the classification accuracy by 63%. Two proposed spike-based signal encoding metrics have been demonstrated to predict well the classification accuracy. The model shows that a relaxed spiking

TABLE I: Comparison of state-of-the-art texture classification systems for robotic applications.

Reference	Sensor Readout (ADC Resolution, Sampling Rate)	Hardware Setup	No. of Sensors	Estimated Data Rate	Preprocessing	Classifier	No. of Classes	Classification Accuracy (Chance)
This work	On-chip Spikes (3.5b,variable)	End-to-End Model (Hardware Spike Encoding)	9	51.32 Hz	Rate code	CNN	12	90% (8.33%)
Ref [16]	Frames (10b,9.6 kHz)	Off-the-shelf Sensors + ADC, Classification in software	4 (Multi-modal)	96 kHz	Cumulative Multi-Bandpower (in software)	Multi-layer Perceptron (MLP)	12	79% (8.33%)
Ref [17]	Software Spikes (unreported,300 Hz)	In-house MEMS sensors + Commercial ADC, Classification in software	6	100 Hz	None	Extreme Learning Machine (ELM) chip	10	92% (10%)
Ref [11]	Software Spikes (24b,380 Hz)	In-house piezoresistive sensors + Commercial ADC, SNN Classification in software	4	9.12kHz (Estimated)	None	SNN (w/ Calcium plasticity) + KNN	10	94.2% (10%)
Ref [18]	Software Spikes (10b,200 Hz)	In-house sensors + Commercial ADC, SNN Classification in software	2 (Multi-modal)	115.2 kHz (Bluetooth baudrate)	Wavelet Transform	SNN (Tempotron)	2	99.45% (50%)
Ref [19]	Software Spikes (24b,300 Hz)	In-house piezoresistive sensors + Commercial ADC, Classification in software	4	28.8 kHz (Estimated)	Wavelet Transform + Cross-Wavelet Transform	Transform + KNN	3	97.6% (33.33%)

readout resolution (*3.5 bits*), a small network (100 neurons in the hidden layer) with a simple rate coding pre-processing of spikes can achieve a texture classification accuracy of 90 % using a CNN, which is comparable to the state of the art, but at a lower power consumption.

ACKNOWLEDGMENT

The authors would like to thank Ali Safa of *imec* and *KU Leuven* for sharing his insights. This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 861166.

REFERENCES

[1] Bartolozzi, C., Natale, L., Nori, F. et al. Robots with a sense of touch. Nature Mater 15, 921–925 (2016). https://doi.org/10.1038/nmat4731

[2] X. Karagiorgis, M. Ntagios, P. Skabara and R. Dahiya, "3D-printed elastomer foam-based soft capacitive pressure sensors," 2022 IEEE International Conference on Flexible and Printable Sensors and Systems (FLEPS), 2022, pp. 1-4, doi: 10.1109/FLEPS53764.2022.9781553.

[3] L. Pinna, A. Ibrahim and M. Valle, "Interface Electronics for Tactile Sensors Based on Piezoelectric Polymers," in IEEE Sensors Journal, vol. 17, no. 18, pp. 5937-5947, 15 Sept.15, 2017, doi: 10.1109/JSEN.2017.2730840.

[4] 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.

[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] Tee, Benjamin Chortos, Alex Berndt, André Nguyen, Amanda Tom, Ariane McGuire, Allister Lin, Ziliang Tien, Kevin Bae, Won-Gyu Wang, Huiliang (Evan) Mei, Phil Chou, Ho-Hsiu Cui, Bianxiao Deisseroth, Karl Ng, Tse-Nga Bao, Zhenan. (2015). A skin-inspired organic digital mechanoreceptor. Science. 350. 313-316. 10.1126/science.aaa9306.

[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] K. E. Friedl, A. R. Voelker, A. Peer and C. Eliasmith, "Human-Inspired Neurorobotic System for Classifying Surface Textures by Touch," in IEEE Robotics and Automation Letters, vol. 1, no. 1, pp. 516-523, Jan. 2016, doi: 10.1109/LRA.2016.2517213.

[9] M.D. Alea, A. Safa, J. Van Assche, and G. Gielen, "Power-Efficient and Accurate Texture Sensing Using Spiking Readouts for High-Density e-Skins," in IEEE Biomedical Circuits and Systems (BioCAS), Oct. 2022, in press.

[10] Saal, Hannes Delhaye, Benoit Rayhaun, B.C. Bensmaia, Sliman. (2017). Simulating tactile signals from the whole hand with millisecond precision. Proceedings of the National Academy of Sciences of the United States of America. 114. E5693-E5702. 10.1073/pnas.1704856114/-/DCSupplemental.

[11] 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.

[12] M. Yang, S. -C. Liu and T. Delbruck, "Comparison of spike encoding schemes in asynchronous vision sensors: Modeling and design," 2014 IEEE International Symposium on Circuits and Systems (ISCAS), 2014, pp. 2632-2635, doi: 10.1109/ISCAS.2014.6865713.

[13] Olcay Kursun, Ahmad Patooghy, March 26, 2020, "VibTac-12: Texture Dataset Collected by Tactile Sensors", IEEE Dataport, doi: https://dx.doi.org/10.21227/kwsy-x398.

[14] 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.

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

[16] O. Kursun and A. Patooghy, "An Embedded System for Collection and Real-Time Classification of a Tactile Dataset," in IEEE Access, vol. 8, pp. 97462-97473, 2020, doi: 10.1109/ACCESS.2020.2996576.

[17] M. Rasouli, Y. Chen, A. Basu, S. L. Kukreja and N. V. Thakor, "An Extreme Learning Machine-Based Neuromorphic Tactile Sensing System for Texture Recognition," in IEEE Transactions on Biomedical Circuits and Systems, vol. 12, no. 2, pp. 313-325, April 2018, doi: 10.1109/TBCAS.2018.2805721.

[18] Navaraj, William Dahiya, Ravinder. (2019). Fingerprint-Enhanced Capacitive-Piezoelectric Flexible Sensing Skin to Discriminate Static and Dynamic Tactile Stimuli. Advanced Intelligent Systems. 1. 10.1002/aisy.201900051.

[19] C. M. Oddo, M. Controzzi, L. Beccai, C. Cipriani and M. C. Carrozza, "Roughness Encoding for Discrimination of Surfaces in Artificial Active-Touch," in IEEE Transactions on Robotics, vol. 27, no. 3, pp. 522-533, June 2011, doi: 10.1109/TRO.2011.2116930.