Exploring Information-Theoretic Criteria to Accelerate the Tuning of Neuromorphic Level-Crossing ADCs

Safa, Ali; Van Assche, Jonah; Frenkel, Charlotte; Bourdoux, Andre; Catthoor, Francky; Gielen, Georges

DOI
10.1145/3584954.3584994

Publication date
2023

Document Version
Final published version

Published in
Proceedings of the 2023 Annual Neuro-Inspired Computational Elements Conference, NICE 2023

Exploring Information-Theoretic Criteria to Accelerate the Tuning of Neuromorphic Level-Crossing ADCs

Ali Safa
ali.safa@imec.be
imec and ESAT KU Leuven
Leuven, Belgium

Jonah Van Assche
jonah.vanassche@esat.kuleuven.be
MICAS, KU Leuven
Leuven, Belgium

Charlotte Frenkel
c.frenkel@tudelft.nl
Delft University of Technology
Delft, The Netherlands

André Bourdoux
andre.bourdoux@imec.be
imec
Leuven, Belgium

Francky Catthoor
francky.catthoor@imec.be
imec and ESAT KU Leuven
Leuven, Belgium

Georges Gielen
georges.gielen@kuleuven.be
imec and ESAT KU Leuven
Leuven, Belgium

ABSTRACT

Level-crossing analog-to-digital converters (LC-ADCs) are neuromorphic, event-driven data converters that are gaining much attention for resource-constrained applications where intelligent sensing must be provided at the extreme edge, with tight energy and area budgets. LC-ADCs translate real-world analog signals (such as ECG, EEG, etc.) into sparse spiking signals, providing significant data bandwidth reduction and inducing savings of up to two orders of magnitude in area and energy consumption at the system level compared to the use of conventional ADCs. In addition, the spiking nature of LC-ADCs make their use a natural choice for ultra-low-power, event-driven spiking neural networks (SNNs). Still, the compressed nature of LC-ADC spiking signals can jeopardize the performance of downstream tasks such as signal classification accuracy, which is highly sensitive to the LC-ADC tuning parameters. In this paper, we explore the use of popular information criteria found in model selection theory for the tuning of the LC-ADC parameters. We experimentally demonstrate that information metrics such as the Bayesian, Akaike and corrected Akaike criteria can be used to tune the LC-ADC parameters in order to maximize downstream SNN classification accuracy. We conduct our experiments using both full-resolution weights and 4-bit quantized SNNs, on two different bio-signal classification tasks. We believe that our findings can accelerate the tuning of LC-ADC parameters without resorting to computationally-expensive grid searches that require many SNN training passes.

KEYWORDS

LC-ADC, Information criteria, event-based sampling, Spiking Neural Networks

NICE 2023, April 11–14, 2023, San Antonio, TX, USA

ACM ISBN 978-1-4503-9947-0/23/04…$15.00
https://doi.org/10.1145/3584954.3584994

ACM Reference Format:
Ali Safa, Jonah Van Assche, Charlotte Frenkel, André Bourdoux, Francky Catthoor, and Georges Gielen. 2023. Exploring Information-Theoretic Criteria to Accelerate the Tuning of Neuromorphic Level-Crossing ADCs. In *Neuro-Inspired Computational Elements Conference (NICE 2023), April 11–14, 2023, San Antonio, TX, USA*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3584954.3584994

1 INTRODUCTION

In recent years, the use of neuromorphic *level-crossing analog-to-digital converters* (LC-ADCs) is gaining an increasing attention for sensing applications in extreme-edge devices, where latency, area and energy consumption must be kept as low as possible [18, 23], while maintaining the target performance metric (e.g., classification accuracy). Indeed, LC-ADCs are event-driven converters that transform their analog input signals into resource-efficient spike trains, emitting spikes only when the change in the input signal exceeds a certain threshold [23]. Hence, the use of LC-ADCs can result in up to two orders of magnitude reductions of data bandwidth and system energy consumption, especially when input signals are burst-like or sparse in time [18, 22]. Since many biological signals, such as electrocardiogram (ECG) and electroencephalogram (EEG) data, exhibit burst-like sparsity properties, a lot of attention has been paid to the application of LC-ADCs for ultra-low-power biomedical monitoring applications [8].

In addition, the spike trains produced by LC-ADCs align with spiking neural networks (SNNs) [9, 18], which differ from conventional frame-based deep neural networks (DNNs) through their use of event-driven spiking neurons [9]. LC-ADCs are especially suited for SNNs since their *sparse* spike trains can induce a *sparse* SNN activity, whereby reducing the overall energy consumption.

Still, the compressed nature of LC-ADC spiking signals can significantly affect *task performance*. It is therefore paramount to wisely tune the LC-ADC parameters such as the spiking threshold *offset* and *decay* rate, when seeking a high SNN signal classification accuracy and efficiency at the system level.

Since performing a fine-grained grid search can be lengthy as it involves a large number of SNN training procedures, several studies have highlighted the use of reconstruction error metrics for *quickly* assessing the quality of the LC-ADC spike encoding given its tuning parameters [14, 16, 20, 25], without the need for SNN training and testing. But to the best of our knowledge, the

use of other information metrics taking into account the LC-ADC spiking density (or *complexity*) has not yet been reported in literature. Indeed, typical reconstruction error metrics [14, 16, 20, 25] do not penalize for the model complexity, even though a larger spike density may increase the chances of over-fitting and hence, jeopardize SNN accuracy. Therefore, our goal in this paper is to explore the use of popular information-theoretic criteria such as the Bayesian, Akaike and corrected Akaike criteria [1, 4, 5] with the aim of tuning the LC-ADC parameters in order to optimize downstream SNN classification accuracy.

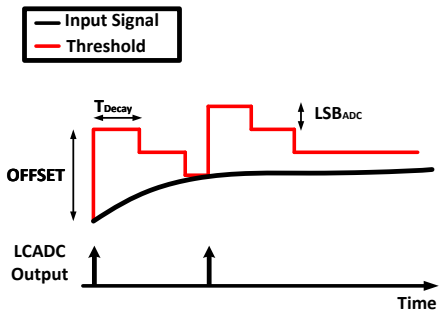

Figure 1: ***Working principle of the adaptive threshold LC-ADC.***

The contributions of this paper are the following:

(1) We provide an *information-theoretic* method for tuning the spiking parameters of neuromorphic LC-ADCs in order to maximise the accuracy of *spiking signal* classification tasks after the LC-ADC encoding.
(2) We experimentally demonstrate our method on two biosignal classification tasks: *electrocardiogram* (ECG) PQRST detection and *epilepsy* detection, using both Support Vector Machines (SVMs) and Spiking Neural Networks (SNNs) with both 4-bit weights (for hardware implementation) and floating point weights (on computer).
(3) We show that a *clear correlation* exists between tuning the LC-ADC to optimize the various information criteria and maximising the SVM and SNN accuracy.

This paper is organized as follows. An overview of the LC-ADC model used throughout our work is provided in Section 2. The various *information criteria* investigated in this work are covered in Section 3. Experimental results are shown in Section 4. A discussion on the obtained results is given in Section 5. Finally, conclusions are provided in Section 6.

2 LEVEL-CROSSING ADC MODEL

In any sensor application, an interface that converts analog signals to a digital representation is required. For neuromorphic sensing systems, this conversion can be done by means of signal-dedicated hardware such as dynamic vision sensors (DVS) and spiking cochlea [12, 28]. However, these sensor types are signal specific, and can only be used for image and audio signals. A more generic type of neuromorphic sensing interface is the level-crossing analog to digital converter (LC-ADC), also called analog-to-spike converter [9, 23, 26]. Unlike conventional sensor systems, where the ADC samples the sensor signal at a fixed (Nyquist) rate that is determined by the maximum sensor signal bandwidth, LC-ADCs sample the signal only whenever a threshold value is exceeded [9, 23, 26]. This makes the output of the circuit *event-driven*, with the circuit dynamic power consumption and data rate scaling according to the input signal activity. For low-duty-cycle signals such as ECG or EEG, this gives up to two orders or magnitude energy saving at the ADC and system levels [22]. The ADC output of a LC-ADC is represented in a spiking format with both positive and negative polarities [9], making it an ideal candidate for interfacing to SNNs.

The resolution of the LC-ADC has an effect on the total data rate and hence, on the power consumption of the neuromorphic system [9, 22]. To lower the data rate, while simultaneously having a high ADC resolution, adaptive-threshold LC-ADCs have been proposed [2, 26]. Adaptive-threshold LC-ADCs dynamically change the threshold over time (e.g., by a step function [26] or an exponential decay [2]). Fig. 1 shows the working principle of such adaptive threshold LC-ADC, using a decaying step function.

Fig. 2 shows a high-level behavioral model for the adaptive-threshold LC-ADC used in this work, implementing a decaying step function [26]. The LC-ADC architecture of Fig. 2 is an enhanced version of our previously presented LC-ADC converter [23], and also includes the adaptive threshold mechanism now. Our model is implemented in MATLAB, and can be used to convert any data set of signals into a spiking dataset that contains the same artefacts as an LC-ADC implemented in hardware (without the need of generating the data from an actual LC-ADC chip).

In the model, the main circuit *non-idealities* are included such as thermal noise from the sample stage and the comparator ($V_{n,kTC}$ and $V_{n,comp}$), DAC mismatch ($V_{e,DAC}$) and the offset of the comparator ($V_{os,comp}$). In Fig. 1, the behavior of the model (for the *positive* polarity threshold) is depicted. After a sample is generated by the ADC, the threshold is reset to the maximum ***offset*** value (this is a *programmable* value, for example 5 LSB). After a certain *programmable* ***decay*** time T_{Decay} have passed, the offset is lowered with 1 LSB. If the signal crosses the threshold value, a sample is again generated and the offset is reset again. If the signal does not cross the threshold value, the ADC decreases the threshold till the offset value is equal to 1 LSB and remains equal to this value until a new threshold crossing is registered. The ADC compares the input signal in 4 phases (see $\Phi_{1,2,3,4}$ in Fig. 2) to two thresholds (a positive and negative threshold), and depending on which threshold is crossed, the LC-ADC generates a spike with either positive or negative polarity.

While the compressed nature of the LC-ADC spiking output helps to reduce the data rate (lowering area and power consumption at the system level), it has been observed that the accuracy of downstream classifiers such as SNNs is highly sensitive to the choice of the LC-ADC parameters [2]. In Section 3, we propose to study the use of popular information criteria for the tuning of the two main parameters of the adaptive-threshold LC-ADC of Fig. 2: the programmable *offset* and the *decay* values (both taking positive integer values between 1 and 10 LSBs).

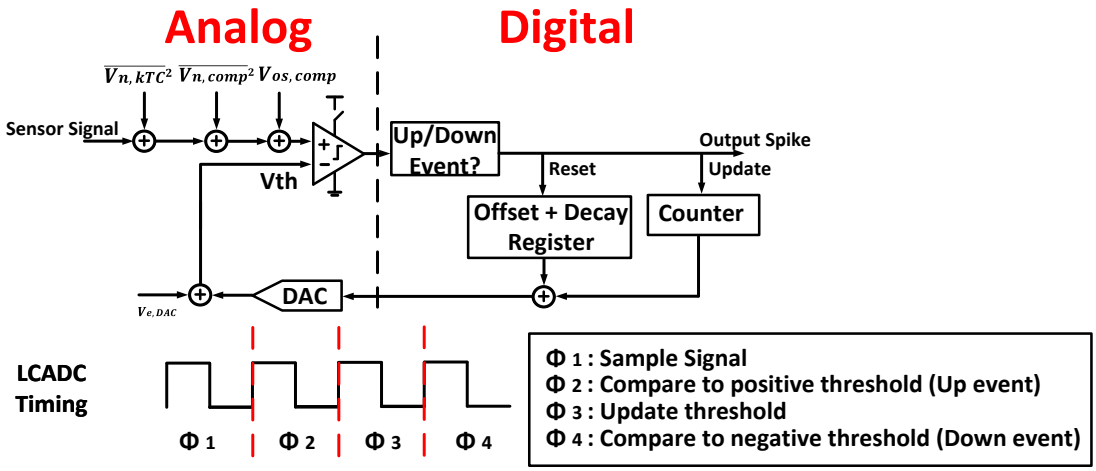

Figure 2: ***High-level behavioral model of the adaptive-threshold LC-ADC used in this work. The LC-ADC timing is also shown. The LC-ADC working phases are noted as*** $\Phi_{1,2,3,4}$.

3 TUNING VIA INFORMATION-THEORETIC CRITERIA

This Section introduces the various information criteria used during our experiments and their application for LC-ADC tuning.

3.1 Bayesian and Akaike Information Criteria

When modelling systems and signals, different models with different complexity (e.g., number of parameters of a regression model, the sparsity of regressed output,...) can be selected [4]. Usually, the higher the complexity of the model, the lower the modelling error for the signals used to fit the model parameters. On the other hand, a higher model complexity generalizes poorly on new signals (i.e., the well-known over-fitting problem) [1]. Information-theoretic criteria such as the Bayesian Information Criteria (BIC) [4] and the Akaike Information Criteria (AIC) [1] are popular techniques that have been used to select the best model candidates based on the *trade-off* between modelling error and model complexity.

Assuming independent and identically-distributed (i.i.d) Gaussian noise, the BIC is given by [4]:

$$\mathrm{BIC} = N_s \log\left(\frac{\sum_{i=1}^{N_s}(\tilde{s}_i - s_i)^2}{N_s}\right) + \kappa \log(N_s) \tag{1}$$

where $s_i, i = 1, ..., N_s$ is the *true* discrete-time signal being encoded, $\tilde{s}_i$ is the *reconstructed* signal from the encoder output, κ is the model complexity (i.e., the number of *non-zero* model parameters) and $\sum_{i=1}^{N_s}(\tilde{s}_i - s_i)^2$ represents the model reconstruction error. Similarly, the AIC is given by [1]:

$$\mathrm{AIC} = N_s \log\left(\frac{\sum_{i=1}^{N_s}(\tilde{s}_i - s_i)^2}{N_s}\right) + 2\kappa \tag{2}$$

Intuitively, the model that gives the ***lowest*** BIC or AIC is the one that provides the best trade-off between modelling error ($\sum_{i=1}^{N_s}(\tilde{s}_i - s_i)^2$ in Eq. 1, 2) and over-fitting (through the κ-dependent complexity term in Eq. 1, 2). Therefore, the BIC and AIC can be used to tune the parameters of a given model by selecting the set of parameters that minimizes the BIC or AIC. Even though playing a similar role, BIC and AIC have been derived based on different prior assumptions [1, 4]. It is therefore customary to investigate both of their effect in literature [4].

In addition, later studies found that the AIC can be biased when N_s is small (i.e., the minimum AIC can suffer from an offset with regard to the model complexity). The *corrected* AIC (noted $\mathrm{AIC_c}$) has then been proposed [5]:

$$\mathrm{AIC_c} = N_s \log\left(\frac{\sum_{i=1}^{N_s}(\tilde{s}_i - s_i)^2}{N_s}\right) + 2\kappa + \frac{2\kappa^2 + 2\kappa}{N_s - \kappa - 1} \tag{3}$$

where the last term is a correction term reducing the AIC bias.

3.2 LC-ADC decoder and complexity measure

Let us now investigate how the BIC, AIC and $\mathrm{AIC_c}$ can be used to automatically tune the *decay* and *offset* parameters of our adaptive-threshold LC-ADC (see Section 2) given arbitrary sensory data as input.

3.2.1 Decoder. In order to measure the BIC, AIC and $\mathrm{AIC_c}$ of our LC-ADC given its *decay* and *offset* parameters, we must find a way to reconstruct *a posteriori* the input signal from the event-driven spiking output of the LC-ADC. We therefore look at the LC-ADC as an *encoder* that converts the input vector $s_i, i = 1, ..., N_s$ into a vector of spikes $e_i, i = 1, ..., N_s$ (where i represents the discrete time index). Hence, we must propose a *decoder* that converts the LC-ADC output $e_i, i = 1, ..., N_s$ into the reconstructed input signal $\tilde{s}_i, i = 1, ..., N_s$.

We build our decoder as follows. First, we apply a low-pass filter to $e_i, i = 1, ..., N_s$ in order to reject the spurious frequencies associated with the discontinuity of the spiking signal. Doing so, we retrieve an intermediate signal r_i:

$$r_{i+1} = \beta r_i + (1 - \beta)e_i, \forall i = 1, ..., N_s \tag{4}$$

where $r_0 = 0$, $\beta = \exp(-\omega_c T_{sim})$ sets the cutoff frequency ω_c of the low-pass filter, and T_{sim} is the simulation period. The cutoff frequency ω_c is signal-dependent and must be tuned in order to cover the bandwidth of the original input signal s_i fed to the LC-ADC.

Then, we must adjust the *amplitude* and *offset* of the intermediate signal r_i in order to compensate for the attenuation induced by the low-pass filtering steps. The amplitude gain a and offset b can be found via the *least-squares* method:

$$a, b = \arg\min_{a,b} \sum_{i=1}^{N_s} (ar_i + b - s_i)^2 \tag{5}$$

By setting the derivative of $\sum_{i=1}^{N_s}(ar_i + b - s_i)^2$ with regard to a and b to zero, we find:

$$a = \frac{\sum_{i=1}^{N_s}(r_i - \mu_r)(s_i - \mu_s)}{\sum_{i=1}^{N_s}(r_i - \mu_r)^2}, \quad b = \mu_s - a\mu_r \tag{6}$$

where $\mu_{r,s}$ respectively denote the mean value of r_i and s_i. Finally, the reconstructed signal $\tilde{s}_i$ is found as:

$$\tilde{s}_i = ar_i + b \quad \forall i = 1, ..., N_s \tag{7}$$

Fig. 3 and 4 show our reconstruction process using as respective input s_i an ECG signal from the MIT-BIH arrhythmia dataset [13] and an electroencephalogram (EEG) signal from the epilepsy seizure detection dataset proposed in [3]. We respectively use a cut-off frequency of $\omega_c = 2\pi \times 2.5$ rad/s and $\omega_c = 2\pi \times 1.25$ rad/s for the ECG and EEG signals (hand-tuned heuristically based on the obtained reconstructions $\tilde{s}_i$).

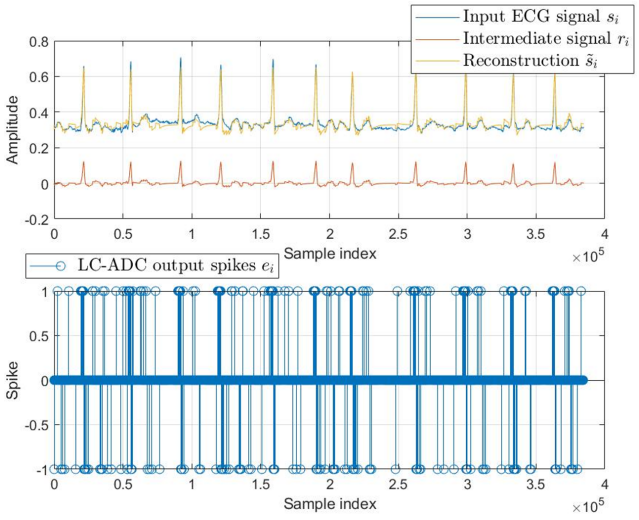

Figure 3: *Input ECG signal reconstruction. The input signal s_i is fed to the LC-ADC which outputs the spiking signal e_i. The decoding process of Section 3.2.1 reconstructs the input by i) low-pass filtering to get the intermediate signal r_i via (4) and ii) gain compensation to obtain the reconstruction $\tilde{s}_i$ via (7).*

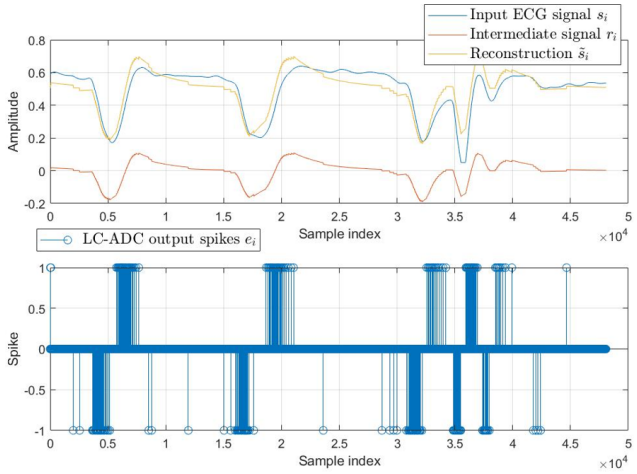

Figure 4: *Input EEG signal reconstruction. The reconstruction steps are similar to Fig. 3.*

3.2.2 Complexity measure. Since we consider the architecture of the LC-ADC fixed, the complexity of the encoding-decoding system corresponds to the number of variable non-zero parameters in the decoding process (4)-(7). The gain compensation scheme (7) being always fixed (i.e., always has two non-zero parameters a and b), we turn our attention to the low-pass filtering step (4).

Under the light of (4), it is clear that the only variable source of non-zero parameters corresponds to the number of non-zero values in $e_i, i = 1, ..., N_s$ (i.e., the number of LC-ADC spikes). Therefore, we compute κ in BIC (1), AIC (2) and AIC_c (3) as the number of non-zero values in $e_i, i = 1, ..., N_s$.

We now have all the tools necessary to evaluate the various information criteria by *i)* computing the model reconstruction error $\sum_{i=1}^{N_s}(\tilde{s}_i - s_i)^2$ in (1)-(3) via the reconstructed signal (7) and *ii)* computing the complexity term κ as the number of spikes in the LC-ADC output e_i.

We can then perform these steps for each choice of *decay* and *offset* parameter and report the BIC, AIC and AIC_c in function of LC-ADC *decay* and *offset* (see Fig. 5)

4 EXPERIMENTAL RESULTS

4.1 Dataset preparation

Since biological signal processing has been proposed as a well-suited task for LC-ADCs due to their intrinsic sparsity in time [9, 18, 22], we consider two different bio-signal classification datasets in this work: *1)* 5-class ECG labelling of P, Q, R, S and T-type peaks using the MIT-BIH arrhythmia dataset [13] and *2)* 2-class epilepsy seizure detection using the dataset proposed in [3].

4.1.1 ECG PQRST labelling. The MIT-BIH arrhythmia dataset [13] features a collection of long ECG recordings from different patients (each recording lasts around 30 minutes). We choose recording *101* as our training sequence and recording *201* as the independent test sequence acquired on a different patient. This balanced train-test partitioning makes the learning problem challenging since the training set is *not* significantly larger than the test set (as it is often

the case), while keeping the computational time of the baseline grid searches with quantization-aware SNN training to an acceptable duration.

We encode the 30-min train/test ECG sequences using our MATLAB LC-ADC model, for various (*offset*, *decay*) settings in a 10×10 grid (i.e., both *offset* and *decay* $\in \{1, ..., 10\}$ LSB with increment step 1), and with simulation time-step $T_{sim} = 2.08 \times 10^{-5}$ s. We use the PQRST labelling MATLAB toolbox recently proposed in [19] to automatically label the ECG sequences in order to obtain the location of the P, Q, R, S and T-peaks. Doing so, we obtain 100 train and test spiking sequences, for each combination (*offset*, *decay*), together with their labels (i.e., the time step at which a P, Q, R, S or T-peak was found).

Then, we partition both the train and test sequences into chunks of spiking data by making each chunk correspond to the time window spanned right before each PQRST label. For example, if a P-type label is found at time step i_p, the complete ECG spike train $e_i, \forall i = 1, ..., N_s$ is cut into the chunk $e_j, \forall j = i_p - L, ..., i_p$, corresponding to the spiking activity that happened *right before* the time step when the P-type peak was found. The time length L is set to 80 steps, as a good balance between performance and computation time.

Doing so, we obtain a 5-class dataset composed of 7350 training patches and 6484 test patches of LC-ADC-encoded ECG data, for each choice of (*offset*, *decay*) parameters.

4.1.2 Epilepsy detection. The epilepsy detection dataset in [3] features 11500 recordings of EEG data chunks over 1.025 s. Each chunk is already labelled as either corresponding to an epileptic seizure or either to an non-epileptic category. We employ the same LC-ADC encoding setup as in the ECG case of Section 4.1.1 and obtain an encoded *spiking* dataset for each choice of (*offset*, *decay*), following the same 10×10 grid (both *offset* and *decay* $\in \{1, ..., 10\}$ LSB with increment step 1). Finally, we randomly split the dataset into a training and a test set using a 70%-30% split.

Next, we will use the ECG and EEG datasets to explore the applicability of the information criteria given in Section 3 for the identification of the LC-ADC parameters maximizing classification accuracy.

4.2 Linear separability assessment

Before exploring the correlation between the information criteria and SNN accuracy, we first verify if there is a correlation between the information criteria and the *linear separability* of the spiking data, in order to prepare the ground for further comparison between the linear and non-linear (SNN) case in Section 5. Linear separability denotes *how well* the dataset, encoded via LC-ADC, can be partitioned using *linear hyper-planes* and therefore, indicates the quality of the LC-ADC encoding [24].

To do so, we employ a *linear* SVM classifier [15] and train it on the spiking LC-ADC output, for each choice of *decay* and *offset*. We perform our experiments on both the *ECG* and the *EEG* datasets. Note that the goal here is not to attain maximal classification accuracy but rather, to transparently check the linear separability of the LC-ADC output (vs. *non-linearity* of SNNs which attains higher accuracy but jeopardizes the *transparent* assessment of linear separability).

Since the SVM we use is not a recurrent model, we feed at once the chunks of spiking ECG and EEG data obtained in Section 4.1 as input vectors to our SVM. We train and test the SVM for each choice of LC-ADC *decay* and *offset* and report the obtained test accuracies as heat-maps in Figs. 5 and 6.

Fig. 5 and 6 clearly shows that a correlation exists between the locations of maximal linear SVM accuracy (in red) and the locations of minimum BIC, AIC and AIC_c in dark blue (with Pearson correlations for all criteria above 0.98 in the ECG case and above 0.32 in the EEG case). The maximal linear SVM accuracy is 60.8% in the ECG case, and 64.9% in the EEG case.

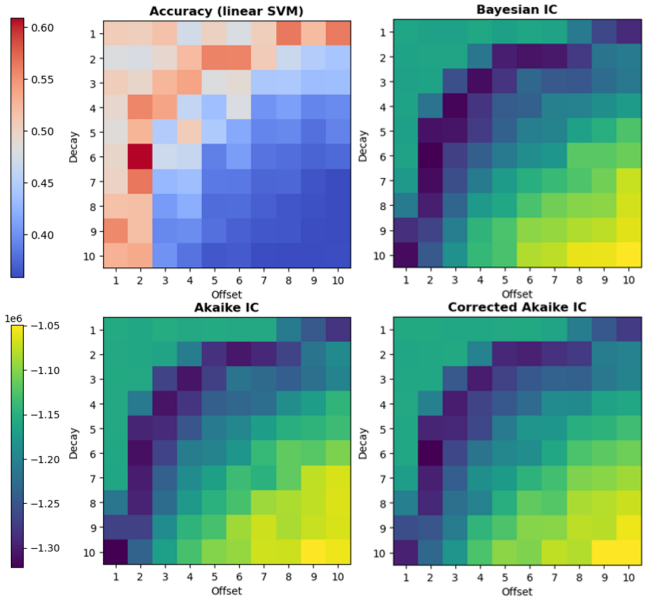

Figure 5: *ECG dataset. Correlation between Linear SVM Accuracy, BIC, AIC and* AIC_c*, in function of LC-ADC decay and offset. A clear correlation exists between the coordinates of maximal accuracy and the coordinates of minimal BIC, AIC and* AIC_c*. The maximum accuracy is 60.8 %.*

4.3 SNN classification assessment

Section 4.2 showed the correlation between the information criteria and the linear separability of the data. Here, we explore whether such correlation still holds when using an SNN as downstream classifier. We verify this under two cases: *1)* when using quantized weights (for SNN implementation in low power neuromorphic chips) and *2)* when using standard floating point weights.

As reference neuromorphic architecture design, we consider the constraints found in the ODIN chip [6] where synapses are quantized to 4-bit precision. In addition, the maximum number of neurons supported by ODIN is 256, which sets the total number of neurons in our proposed SNN design (see Fig. 7). In addition, we exclusively use *Integrate and Fire* (IF) neurons instead of *Leaky* IF to further reduce hardware overheads.

4.3.1 Using 4-bit weights. We train the SNN architecture of Fig. 7 via back-propagation through time (BPTT) [27] using the SLAYER

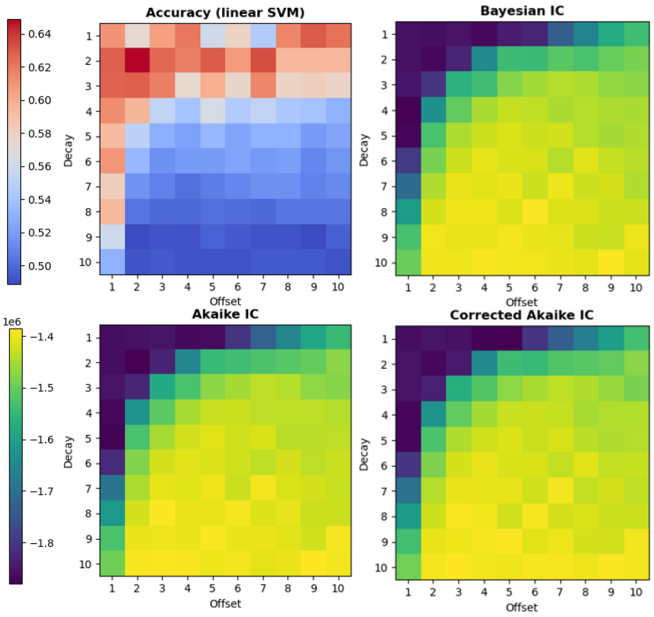

Figure 6: ***EEG dataset. As in the ECG case (Fig. 5), a correlation can be clearly remarked between the information criteria and the SVM accuracy. The maximum accuracy is 64.9 %.***

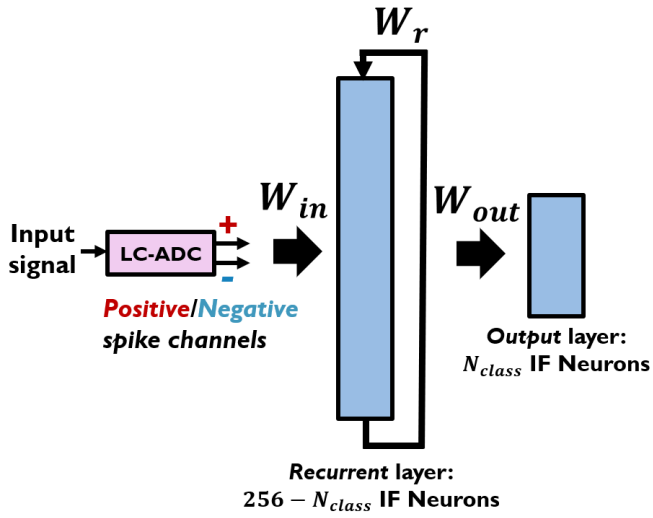

Figure 7: ***Recurrent SNN architecture. The input signal (ECG, EEG) is encoded by the LC-ADC to spike trains of positive and negative polarity. Then, these 2-dimensional spike train are fed to a recurrent layer of IF neurons via the fully-connected weight matrix W_{in}. Recurrence is provided through the weight matrix W_r. Finally, the output of the recurrent layer is fed to an output layer composed of N_{class} IF neurons via the weight matrix W_{out}, where N_{class} is the number of classes in the dataset, also setting the number of neurons in the recurrent layer. Doing so, all 256 neurons in the ODIN chip are used.***

surrogate gradient technique [21] on the 5-class ECG dataset (see Section 4.1) and we report in Fig. 8 the *test accuracy* obtained in function of the LC-ADC parameters *offset* and *decay*. We use the Adam optimizer [11] with decay parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$ and learning rate $\eta = 3 \times 10^{-4}$ for a total of 100 training epochs. Following prior SNN training studies such as [17], we initialize all weights using the *Uniform Xavier* initialization method [7].

Crucially, we perform *quantization-aware training* [10] where the SNN weights are quantized to 4-bit precision, while clipping the maximum *absolute* weight values to 0.05 (adjusted empirically). Quantization-aware training uses the *4-bit weights* for the inference pass and fine-tunes the *full-precision weights* using the back-propagated errors obtained during the 4-bit weight inference pass.

As loss function, we opt for the standard *cross-entropy* as widely used in most SNN classification problems [17]. We perform all our experiments using *pytorch* with an NVIDIA V100 GPU using a batch size of 128. Fig. 8 compares the accuracy on the *test set* with the AIC_c already shown in Fig. 5

We observe in Fig. 8 that a slight *bias* exists between the region of minimum BIC, AIC, AIC_c and the region of maximal SNN accuracy in Fig. 8. Indeed, the region of maximal SNN accuracy is found for slightly larger values of (*offset*, *decay*) compared to the region of minimum BIC, AIC (in Fig. 5) and AIC_c. Still, the use of information criteria already provides a useful estimation of the region of (*decay*, *offset*) pairs that induce maximal SNN accuracy. The existence of this bias will be further discussed in Section 5.

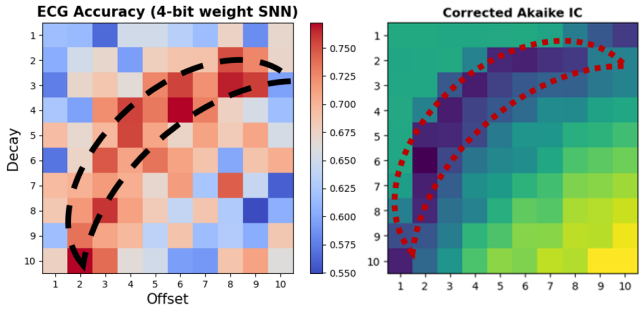

Figure 8: ***ECG test accuracy using our SNN with 4-bit weights, in function of LC-ADC decay and offset. Similar to the linear SVM case (Section 4.2), the coordinates of maximum accuracy correlate well with the coordinates of minimum BIC, AIC and AIC_c in Fig. 5. The top accuracy is 77.3%. The banana-shaped delimitations were manually drawn for indication purposes.***

4.3.2 Using floating point weights. Finally, we close our experimental investigations by exploring whether the correlation between the information criteria and SNN accuracy still holds when using regular floating point weights. We train the SNN of Fig. 7 in a similar setting as in Section 4.3.1 on the 2-class Epilepsy detection dataset (see Section 4.1) and we report in Fig. 9 the *test accuracy* obtained in function of the LC-ADC parameters *offset* and *decay*. All training parameters are the same as in Section 4.3.1 with the difference that we use at most 10 training epochs with no weight quantization step. Fig. 9 compares the accuracy on the *test set* with the AIC_c already shown in Fig. 6

As in the case of Fig. 8, Fig. 9 shows that a *bias* still remains between the region of minimum BIC, AIC (in Fig. 6), AIC_c and the region of maximal accuracy. Still, using the information criteria enables a first crude identification of the possible *offset* and *decay* values expected to induce a higher SNN classification accuracy.

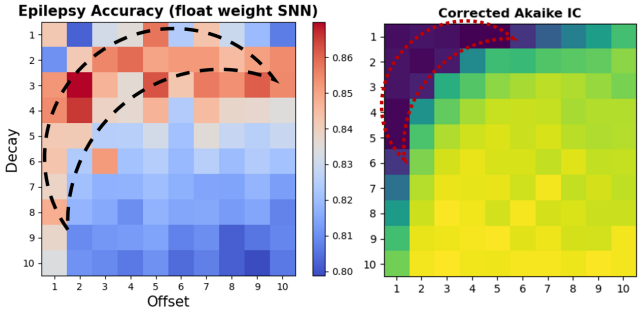

Figure 9: ***EEG test accuracy (using 32-bit float weights). Similar to the linear SVM case (Section 4.2), the coordinates of maximum accuracy correlate well with the coordinates of minimum BIC, AIC and*** $\mathbf{AIC_c}$ ***in Fig. 6. Top accuracy is 86.9%.***

5 DISCUSSION

In this paper, we have studied how popular information criteria used in statistical model selection theory can be applied for tuning LC-ADC parameters, in order to optimize the accuracy of downstream SNN classifiers. This study has been conducted with the aim of finding a method that would enable neuromorphic system designers to efficiently tune the spiking parameters of LC-ADCs given a specific class of signals to process (e.g., ECG, EEG and so on), *without* the need for explicit SNN training in lengthy grid searches. In Section 4, we have shown that a clear correlation exists between the (*decay*, *offset*) coordinates of maximal SNN accuracy and the different information criteria assessed in this paper, *accelerating* the tuning of LC-ADC parameters towards their near-optimal values.

Still, we have remarked in Section 4.3 the existence of a small bias between the regions of maximal SNN accuracy and the regions of minimal BIC, AIC, AIC_c (see Figs. 5, 8 and 6, 9). On the other hand, it can be seen in Figs. 5 and 6 that this bias is *significantly less important* in the *linear* classification case (using a linear SVM, see Section 4.2). This difference between the linear and the SNN case might be caused by the fact that the SNN classifier is a *non-linear* model. Indeed, due to its non-linearity, it will be easier for the SNN to *over-fit* compared to the case of the linear SVM. Therefore, in order to avoid over-fitting, the SNN will require a *higher* LC-ADC spike train *sparsity*, leading to larger values of BIC, AIC and AIC_c (since a higher spiking sparsity will make the reconstruction error $\sum_{i=1}^{N_s}(\tilde{s}_i - s_i)^2$ grow in Eqs. 1-3). This explains why the SNN requires slightly *larger* values of (*offset*, *decay*) compared to the linear SVM, as both larger *offsets* and larger *decays* lead to higher LC-ADC sparsity, reducing the SNN over-fitting.

A number of prior work such as [14, 16, 20] propose to solely use the LC-ADC *mean square error* (MSE) in order to assess the quality of the spike trains produced by their signal-to-spike encoders (the smaller, the better). In contrast, this paper proposed the use of information criteria where *complexity* (i.e., density of spike trains κ in Eq. 1-3) also impacts the information metric. Indeed, we argue that using the MSE can be sub-optimal when the goal is to tune the spiking parameters for attaining maximal classification accuracy. To illustrate this, Fig. 10 compares the LC-ADC MSE on the ECG dataset against the use of AIC_c.

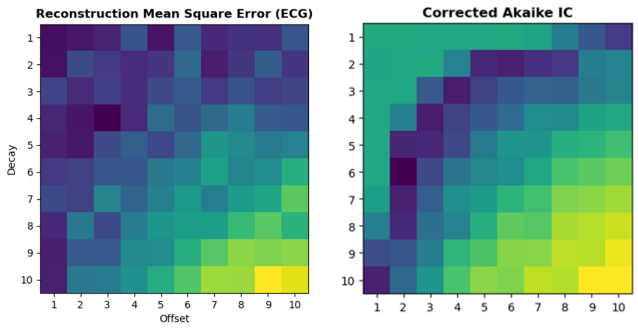

Figure 10: ***LC-ADC reconstruction MSE (left) compared to*** $\mathbf{AIC_c}$ ***(right). In contrast to the use of information criteria such as BIC, AIC and*** $\mathbf{AIC_c}$***, the MSE fails to indicate the banana-shaped region of maximal SVM and SNN accuracy on the ECG dataset (see Fig. 5 and 8).***

As expected, the smaller the *offset* and *decay* parameters, the lower the reconstruction MSE (since the LC-ADC becomes more sensible to instantaneous changes in the input signal). Crucially, the use of the LC-ADC MSE does *not* correlate with the regions of maximal SVM and SNN accuracy in Fig. 5 and 8. This is in contrast to the use of BIC, AIC and AIC_c where the banana-shaped region of maximal accuracy can be clearly identified (see Fig. 8).

Therefore, this observation indicates that the use of information-theoretic criteria covered in Section 3 might constitute *a better approach compared to MSE*, for tuning the LC-ADC parameters towards their near-optimal value without performing an explicit SNN training grid search. Still, in the case of the EEG dataset (see Fig. 6 and 9), we see that the various information criteria are monotonically decreasing when the *offset* and *decay* are adjusted towards smaller values, similar to how the MSE would behave. Therefore, this might also indicate that the superiority of information criteria such as BIC, AIC and AIC_c over the MSE can be *task-dependent*.

In summary, we strongly believe that the use of information criteria such as BIC, AIC and AIC_c can help accelerating the initial tuning of LC-ADC parameters, by quickly identifying the region of (*offset*, *decay*) values where accuracy is expected to be maximal. After identifying this initial region, a faster, *small-scale* search could then be performed to further fine-tune the LC-ADC parameters. In contrast to large-scale computationally-expensive fine-grained grid searches involving many SNN training passes, our methodology will greatly help accelerating design time.

6 CONCLUSION

This paper provides what is, to the best of our knowledge, a first link between the quality of the spiking signals produced by a neuromorphic LC-ADC and various information criteria used in statistical

model selection theory. After introducing the LC-ADC circuit model used in this work, we have provided a novel method for tuning the spiking parameters of neuromorphic components such as LC-ADCs using the Bayesian, Akaike and corrected Akaike criteria. Finally, we have illustrated our findings on two bio-signal classification tasks: ECG PQRST detection and epilepsy seizure detection, using both SVMs and SNNs as downstream classifiers. We strongly believe that the method provided in this paper will accelerate design time by enabling neuromorphic system designers to efficiently tune LC-ADC parameters near their optimal values for a given task, without the burden of lengthy grid-searches involving many SNN training phases.

ACKNOWLEDGMENTS

The authors thank Mark D. Alea, Lars Keuninckx, Yumming He and Ilja Ocket for useful discussions. This research received funding from the Flemish Government under the “Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen” programme, the European Union’s ECSEL Joint Undertaking under grant agreement n° 826655 - project TEMPO and the KUL internal grant "Time is money".

REFERENCES

[1] H. Akaike. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Control* 19, 6 (1974), 716–723. https://doi.org/10.1109/TAC.1974.1100705

[2] Mark Daniel Alea, Ali Safa, Jonah Van Assche, and Georges G. E. Gielen. 2022. Power-Efficient and Accurate Texture Sensing Using Spiking Readouts for High-Density e-Skins. In *2022 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. 359–363. https://doi.org/10.1109/BioCAS54905.2022.9948546

[3] Ralph Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian Elger. 2002. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical review. E, Statistical, nonlinear, and soft matter physics* 64 (01 2002), 061907. https://doi.org/10.1103/PhysRevE.64.061907

[4] K.P. Burnham and D.R. Anderson. 2002. *Model selection and multimodel inference: a practical information-theoretic approach.* Springer Verlag.

[5] Joseph E. Cavanaugh. 1997. Unifying the derivations for the Akaike and corrected Akaike information criteria. *Statistics & Probability Letters* 33, 2 (30 April 1997), 201–208. http://www.sciencedirect.com/science/article/B6V1D-3WNMX5B-X/1/fcad33e5e0d138a917cb28ab3e5b2d43

[6] Charlotte Frenkel, Martin Lefebvre, Jean-Didier Legat, and David Bol. 2019. A 0.086-mm^2 12.7-pJ/SOP 64k-Synapse 256-Neuron Online-Learning Digital Spiking Neuromorphic Processor in 28-nm CMOS. *IEEE Transactions on Biomedical Circuits and Systems* 13, 1 (2019), 145–158. https://doi.org/10.1109/TBCAS.2018.2880425

[7] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research, Vol. 9)*, Yee Whye Teh and Mike Titterington (Eds.). PMLR, Chia Laguna Resort, Sardinia, Italy, 249–256. https://proceedings.mlr.press/v9/glorot10a.html

[8] Yuming He, Federico Corradi, Chengyao Shi, Ming Ding, Martijn Timmermans, Jan Stuijt, Pieter Harpe, Ilja Ocket, and Yao-Hong Liu. 2021. A 28.2 μC Neuromorphic Sensing System Featuring SNN-based Near-sensor Computation and Event-Driven Body-Channel Communication for Insertable Cardiac Monitoring. In *2021 IEEE Asian Solid-State Circuits Conference (A-SSCC)*. 1–3. https://doi.org/10.1109/A-SSCC53895.2021.9634787

[9] Yuming He, Federico Corradi, Chengyao Shi, Stan van der Ven, Martijn Timmermans, Jan Stuijt, Paul Detterer, Pieter Harpe, Lucas Lindeboom, Evelien Hermeling, Geert Langereis, Elisabetta Chicca, and Yao-Hong Liu. 2022. An Implantable Neuromorphic Sensing System Featuring Near-Sensor Computation and Send-on-Delta Transmission for Wireless Neural Sensing of Peripheral Nerves. *IEEE Journal of Solid-State Circuits* 57, 10 (2022), 3058–3070. https://doi.org/10.1109/JSSC.2022.3193846

[10] Benoit Jacob, Skirmantas Kligys, Bo Chen, Menglong Zhu, Matthew Tang, Andrew Howard, Hartwig Adam, and Dmitry Kalenichenko. 2018. Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[11] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. https://doi.org/10.48550/ARXIV.1412.6980

[12] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. 2008. A 128× 128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor. *IEEE Journal of Solid-State Circuits* 43, 2 (2008), 566–576. https://doi.org/10.1109/JSSC.2007.914337

[13] G.B. Moody and R.G. Mark. 1990. The MIT-BIH Arrhythmia Database on CD-ROM and software for use with it. In *[1990] Proceedings Computers in Cardiology*. 185–188. https://doi.org/10.1109/CIC.1990.144205

[14] Simon F. Müller-Cleve, Vittorio Fra, Lyes Khacef, Alejandro Pequeño-Zurro, Daniel Klepatsch, Evelina Forno, Diego G. Ivanovich, Shavika Rastogi, Gianvito Urgese, Friedemann Zenke, and Chiara Bartolozzi. 2022. Braille letter reading: A benchmark for spatio-temporal pattern recognition on neuromorphic hardware. *Frontiers in Neuroscience* 16 (2022). https://doi.org/10.3389/fnins.2022.951164

[15] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, Oct (2011), 2825–2830.

[16] Balint Petro, Nikola Kasabov, and Rita M. Kiss. 2020. Selection and Optimization of Temporal Spike Encoding Methods for Spiking Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems* 31, 2 (2020), 358–370. https://doi.org/10.1109/TNNLS.2019.2906158

[17] Ali Safa, Federico Corradi, Lars Keuninckx, Ilja Ocket, André Bourdoux, Francky Catthoor, and Georges G. E. Gielen. 2021. Improving the Accuracy of Spiking Neural Networks for Radar Gesture Recognition Through Preprocessing. *IEEE Transactions on Neural Networks and Learning Systems* (2021), 1–13. https://doi.org/10.1109/TNNLS.2021.3109958

[18] Ali Safa, Jonah Van Assche, Mark Daniel Alea, Francky Catthoor, and Georges G.E. Gielen. 2022. Neuromorphic Near-Sensor Computing: From Event-Based Sensing to Edge Learning. *IEEE Micro* 42, 6 (2022), 88–95. https://doi.org/10.1109/MM.2022.3195634

[19] Rohan Sanghavi, Fenil Chheda, Sachin Kanchan, and Sushma Kadge. 2021. Detection Of Atrial Fibrillation in Electrocardiogram Signals using Machine Learning. In *2021 2nd Global Conference for Advancement in Technology (GCAT)*. 1–6. https://doi.org/10.1109/GCAT52182.2021.9587664

[20] B. Schrauwen and J. Van Campenhout. 2003. BSA, a fast and accurate spike train encoding scheme. In *Proceedings of the International Joint Conference on Neural Networks, 2003.*, Vol. 4. 2825–2830 vol.4. https://doi.org/10.1109/IJCNN.2003.1224019

[21] Sumit Bam Shrestha and Garrick Orchard. 2018. SLAYER: Spike Layer Error Reassignment in Time. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (Montréal, Canada) *(NIPS'18)*. Curran Associates Inc., Red Hook, NY, USA, 1419–1428.

[22] Jonah Van Assche and Georges Gielen. 2020. Power Efficiency Comparison of Event-Driven and Fixed-Rate Signal Conversion and Compression for Biomedical Applications. *IEEE Transactions on Biomedical Circuits and Systems* 14, 4 (2020), 746–756. https://doi.org/10.1109/TBCAS.2020.3009027

[23] Jonah Van Assche and Georges Gielen. 2022. A 10.4-ENOB 0.92-5.38 μW Event-Driven Level-Crossing ADC with Adaptive Clocking for Time-Sparse Edge Applications. In *ESSCIRC 2022- IEEE 48th European Solid State Circuits Conference (ESSCIRC)*. 261–264. https://doi.org/10.1109/ESSCIRC55480.2022.9911506

[24] Vladimir N. Vapnik. 1995. *The nature of statistical learning theory*. Springer-Verlag New York, Inc.

[25] J. Wang, M Alea, J van Assche, and G Gielen. 2023. End-to-End Optimization of High-Density e-Skin Design: From Spiking Taxel Readout to Texture Classification. In *2023 Design, Automation & Test in Europe Conference & Exhibition (DATE), in press.*

[26] Colin Weltin-Wu and Yannis Tsividis. 2013. An Event-driven Clockless Level-Crossing ADC With Signal-Dependent Adaptive Resolution. *IEEE Journal of Solid-State Circuits* 48, 9 (2013), 2180–2190. https://doi.org/10.1109/JSSC.2013.2262738

[27] P.J. Werbos. 1990. Backpropagation through time: what it does and how to do it. *Proc. IEEE* 78, 10 (1990), 1550–1560. https://doi.org/10.1109/5.58337

[28] Minhao Yang, Chen-Han Chien, Tobias Delbruck, and Shih-Chii Liu. 2016. A 0.5V 55μW 64×2-channel binaural silicon cochlea for event-driven stereo-audio sensing. In *2016 IEEE International Solid-State Circuits Conference (ISSCC)*. 388–389. https://doi.org/10.1109/ISSCC.2016.7418070