8th International IEEE EMBS Conference on Neural Engineering
Shanghai, China, May 25 - 28 , 2017

Neuromorphic Spike Data Classifier for Reconfigurable Brain-Machine Interface

Amir Zjajo, Sumeet Kumar, Rene van Leuken

Abstract—**In this paper, we propose a reconfigurable neural spike classifier based on neuromorphic event-based networks that can be directly interfaced to neural signal conditioning and quantization circuits. The classifier is set as a heterogeneity based, multi-layer computational network to offer wide flexibility in the implementation of plastic and metaplastic interactions, and to increase efficacy in neural signal processing. Built-in temporal control mechanisms allow the implementation of homeostatic regulation in the resulting network. The results obtained in a 90 nm CMOS technology show that an efficient neural spike data classification can be obtained with a low power (9.4 μW/core) and compact (0.54 mm^2 per core) structure.**

I. Introduction

Neural spike patterns are the fundamental means by which neurons process and transmit information in the nervous system. Consequently, extracting information from neural recordings is prerequisite for understanding the cortical structures, a better perception of stark brain disorders such as Alzheimer's and Parkinson's diseases, epilepsy and autism, or for reestablishment of sensory (e.g., hearing and vision) and motor (e.g., movement and speech) functions. However, very frequently a high-density microelectrode arrays in multichannel brain-machine interface (BMI) record the patterns from multiple surrounding neurons, e.g., due to the background activity of other neurons, slight perturbations in electrode position, external electrical or mechanical interference.

Subsequently, clustering spike-derived features is a challenging task due to the contaminating noise and superimposed potentials in the recorded patterns [1]; the degree of overlap between the annotated clusters increases as a function of the noise variance. In addition, classifier performance could deteriorate over time due to the changes in task conditions and the surrounding environment. Similarly, convenience and usability of BMI technologies are severely limited due to the required extensive individualized calibration. The neuromorphic event-based neuron network, however, due to ability to learn by example, parallelism of the operation, associative memory, multifactorial optimization, and extensibility, is inherent choice for compact and low power cognitive systems that learn and adapt to the changes in the statistics of the complex neural signals [2].

This research was supported in part by the European Union and the Dutch government, as part of the CATRENE program under Heterogeneous INCEPTION project.

A. Zjajo, S. Kumar and R. van Leuken are with Circuits and Systems Group, Delft University of Technology, Mekelweg 4, 2628 CD, Delft, The Netherlands (e-mail: amir.zjajo@ieee.org).

In this paper, we propose a 64-channel, cognitive, kernel-based neural spike classifier based on neuromorphic event-based networks. The power-efficient, reconfigurable, neural-inspired clustering in BMI is achieved with heterogeneity based multi-layer computation network configuration. Built-in temporal control mechanisms allow the implementation of homeostatic regulation in the resulting network. The results obtained in a 90 nm CMOS technology show that an efficient neural spike data classification can be obtained with a low power (less than 9.4 μW/core, corresponding to a 16.7 μW/mm^2 of power density), in the compact, a low resource usage (0.54 mm^2/core area) structure.

II. Neuromorphic Spike Data Classifier

A. Neural Spike Pattern Classification

Signal information in the brain is encoded by patterns of neural spike activity occurring over populations of neurons. The synapses, i.e. connection to the subsequent neurons, adjust their weight (synaptic conductance), and subsequently their task, in accordance with the neural spikes they receive. Let $K(t)$ be a fixed kernel that includes both pre- and postsynaptic factors. We consider $\Lambda^+=\{\lambda^+\}$ (target neural spike patterns) that have to be separated from $\Lambda^-=\{\lambda^-\}$ (background patterns). We define t_{ij} as the time at which spike arrive at the synapse, i.e. the delay of the j-th spike of the i-th neuron. At each discrete point in time $t_\eta=\Delta\eta$, we evaluate $f(t)=[f_1(t),f_2(t),...,f_N(t)]$, where $f_i(t)=\Sigma_j K(t-t_{ij})$. The objective of classification (Fig. 1) is to find a hyperplane $\mathcal{H}(w,b)$ specified by $w_1f_1+w_2f_2+\ldots+w_Nf_N-b=0$ that separates at least one point of each pattern from all the points $f(t_\eta)$ [3]-[4]

$$\begin{aligned}&\forall\lambda\in\lambda^+,\exists t_\eta:\sum_i w_i\sum_j K(t-t_{ij})-b\geq 0\\&\forall\lambda\in\lambda^-,\exists t_\eta:\sum_i w_i\sum_j K(t-t_{ij})-b<0\end{aligned}\tag{1}$$

where coefficients of the corresponding hyperplane are designated as synaptic weights on the dendrite of an integrate-and-fire neuron. For rate and synchrony encoded input spike patterns, the average synaptic activation $1/T\int^T\sum_{tij}K(t-t_{ij})\,dt$ is directly proportional to the input arriving at that synapse, x_{ij}, where T is the pattern duration [4]. The kernel functions $K(t-t_{ij})$ (double exponential kernel) is defined as

$$K(t-t_{ij})=\kappa_0[\exp(-(t-t_{ij})/\tau_f)-\exp(-(t-t_{ij})/\tau_r)]\tag{2}$$

where κ_0 is the normalization constant, and τ_f and τ_r are the fall and rise time constants, respectively.

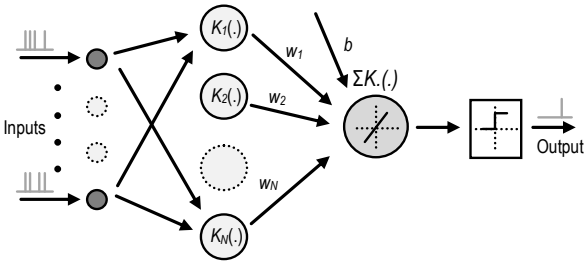

Figure 1: Time-encoded, event-based spike processing network structure. Inputs are connected to a postsynaptic neuron through weighted synapses. The kernels acts as a hidden layer element, and integrate the synaptic signals. Dendritic signals are summed at the soma; if the signals exceed a threshold, the axon releases a spike.

Classification of λ patterns for one-versus-all case is completed with a winner-takes-all scheme (in combination with spike timing dependent plasticity learning), which indicates that the neurons inhibit each other having only one winner, i.e. a group of recurrent neurons cooperate and compete with each other, and the classifier with the highest output function allocates the class. The neuron responds to a λ^+ pattern by firing at least one action potential, and remains inactive when driven by a λ^- pattern. The inputs to the neurons for each class are received from a pair of excitatory and inhibitory dendrites.

B. Neuromorphic Core

Reconfigurable neuromorphic networks, typically, consist of the circuits that only partially include dendritic, and subsequently, synaptic properties. However, increased experimental evidence indicates existence of a large variety of dendritic channels [5], which alter synaptic response by amplification, regulation, the dendritic structure scaling, etc. Accordingly, a synaptic circuit can be computationally much more powerful [6] than just as a simple point processing unit. A typical synaptic learning-array [7] is illustrated in Fig. 2a). The changes in synaptic morphology affect synaptic dynamics and efficacy, and consequently, lead to alterations in network signal processing capabilities. Each synapse multiplies its column input with its weight, and outputs a current. The synapse output currents are summed along the row. The error signal compels the time-averaged sum of the row-synapse weights to be a constant, and forms a bound on the row weights by forcing the synapses to contend for weight value.

In contrast, implemented synapse circuit [8] (Fig. 2b) includes, additionally, multi-compartment dendrites [6], and two postsynaptic back propagating signals [8]-[9] to model local and global (lateral) postsynaptic influences, respectively. This increase in dimensionality allows more states and transitions, offering more flexibility in the implementation of plastic and metaplastic interactions. The model allows assignment of various neuronal characteristics, e.g. axonal and dendritic delays, synaptic transfer functions, offers computation of the optimal input-output transfer function by means of calculated dendritic weights, allows spike train kernel convolution in the time domain. Consequently, by using structural plasticity we reduce the memory requirement to read, to store and access connection information, respectively.

At the circuit input, two receptors are available: NMDA receptor offers activity-dependent modifications of synaptic weight w, denoted as conductance g_m, while AMPA receptor mediates a fast (glutamatergic) synaptic current to drive the soma. The NMDA conductance is a filtered version of the AMPA conductance with a slower rise and a longer tail. In the post-synaptic part, the temporal summation of a back propagating spikes, i.e. dendrite and soma spikes [10], respectively, is completed. If groups (bursts) of dendritic spikes are sufficiently strong to drive the soma, the neuron will generate action potentials; resulting spike is back propagated into the dendrite [9]. We model this global (lateral) inhibitory back propagated current signal as $K_{bp,inh}(t\text{-}t^n_{last})$, where the last postsynaptic spike is generated by the n-th synapse at t^n_{last}. The back propagated dendrite and soma signals are multiplied and added to NMDA receptor signals to form the weight control signal.

Synaptic dynamics is reproduced in real-time using arrays of pulse (spike) integrators. The log domain integrator [11] circuit models slow NMDA receptor-mediated currents, while the differential pair integrator circuit [12] models fast AMPA mediated current. This sub-circuit is functionally equivalent to the one described in theoretical models [5], and often used in computational neuroscience. The transconductance amplifier in the synapse is a typical differential pair amplifier with active loading, and enable/disable capability for power-efficient operation. Employed hysteretic differentiator circuit with an exponential resistive element offers large range of the time-constants of the feedback loop (over several orders of magnitude). The soma and axon hillock are implemented as the adaptive conductance-based integrate-and-fire neuron circuit [12] for compact description of neuronal firing dynamics. Repeating synapse and neuron units form a neuromorphic core (Fig. 2c) for a learning network. The array structure includes electronic synapses at the junctions, and the row of the circuits at the periphery of the array, which mimic the action of the soma and axon hillock of biological neurons.

III. Experimental Results

Design simulations on the transistor level were performed at body temperature (37 ºC) on Cadence Virtuoso using industrial hardware-calibrated TSMC 90nm CMOS technology. The synapse circuit consumes 53 pW at 300 Hz (e.g. in the range of local-field potentials), leading to an energy of 180 fJ per synaptic event. Higher speed version at 20 kHz (e.g. action potentials indicating single-cell activity) consumes 3.2 nW, which corresponds to 160 fJ per synaptic event. The active area of the synapse circuit is 15 μm×9 μm. The area of the synapse including capacitors and dendrites is 15 μm×14 μm. The soma necessitates significant current to create adequate positive feedback; the level is defined by the maximum synaptic current multiplied with the square root of the number of inputs. The digital spikes generated by the conductance-based neuron circuit are very narrow (~250 ns). The circuit average power consumption during this period is 2.1 pJ/spike. The average power dissipation measured throughout the whole current integration and action-potential generation phase is 147 pJ over 100 ms (for an average firing rate of 10 Hz). The neuron circuit has an area of 54 μm^2.

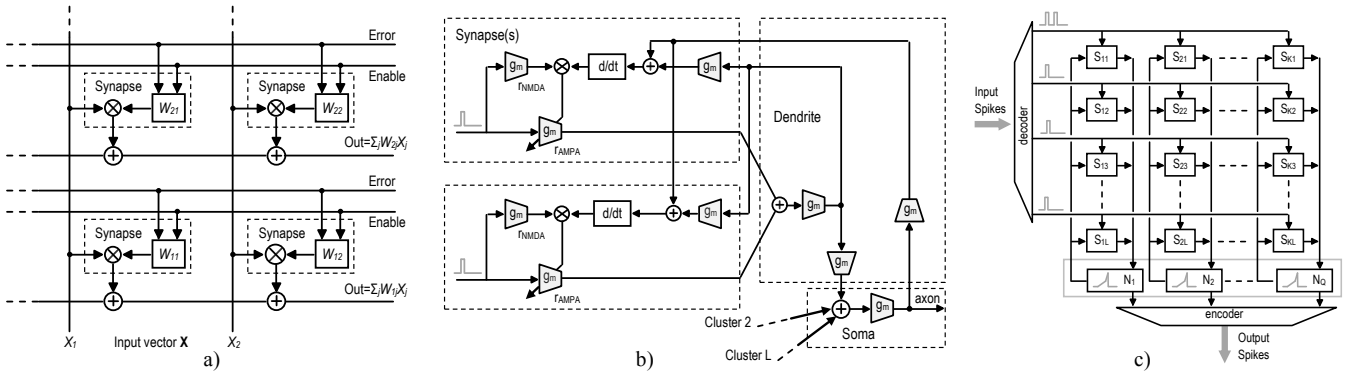

Figure 2: a) Typical learning-array block diagram [7], b) conceptual diagram of the implemented neuromorphic synapse, and a dendrite (input) and an axon soma (output) of a biological neuron, c) internal blocks of the neuromorphic core with dedicated synapse.

The reconfigurable network with a neuromorphic core, which includes 64 integrate-and-fire neurons and 2048 synapses (32 pre neuron), occupies an area of 0.56 mm^2. With all neurons active (i.e. worst case scenario) at 20 kHz firing rate, with each event broadcast within one neuromorphic core, the classifier (including communication) consumes 9.4 μW of power from an 1 V supply voltage. In Table I, we compare the state of the spike sorting systems to this work.

The classifier training is performed by offering multiple trials, which comprise signals attained from grouped neural recordings, conjointly with teacher (Poisson spike trains) signals generated externally. The test dataset is based on BMI recordings from the human neocortex and basal ganglia (Fig. 3a). We designed a model-based analysis to calculate the rate of patterns in spiking activity. Expending a maximum entropy principle [13] with a Markovian assumption, we obtain a model that accounts for temporal pairwise correlations among neurons (Fig. 3b and Fig. 3c). In network simulations, the synaptic weight matrices are initialized to the values derived from the state of the learning synapses at the end of training procedure. External input is then applied to the network with learning rule mediating synaptic plasticity; synapses contain active channels, which respond to the local conditions offering the apparatus to perform non-linear local computation. The spiking activity of the network consists of a superposition of all temporal spike sequences, which it is concurrently recalling. After encoding (through spike timing dependent plasticity in synapses), a pattern of active neurons is formed in a winner-take-all arrangement (due to the strong recurrent inhibition). Synchrony of the network is illustrated in Fig. 4a) and Fig. 4b). When synchronized at 300 Hz, the network receive approximately 600k events per second, and transmit 18.75k events/s.

	[14]	*[15]*	*[16]*	*[17]*	***[this work]***
Technology [nm]	*65*	*90*	*65*	*65*	*90*
Programmability	*no*	*yes*	*no*	*yes*	*yes*
V_{DD} [V]	*0.27*	*1*	*0.3*	*0.4*	*1*
# channels	*16*	*128*	*1*	*128*	*64*
P. Dens. [μW/mm^2]	*60.9*	*9.8*	*43.4*	*15.5*	*16.7/core*
Power [μW]	*75*	*87*	*2.17*	*41*	*9.4/core*
Area [mm^2]	*1.23*	*8.9*	*0.05*	*2.64*	*0.56/core*

TABLE I- COMPARISON WITH PRIOR ART.

The information encoded in the spike trains is classified with a reconfigurable learning network as illustrated in Fig. 4c), where the bold line represent decision boundary. Positive and negative classification outputs in (1) are separated based on the discrimination threshold determined by maximizing classifiers performance. The classifier size, i.e. required number of the synapses and the neurons, is (partly) governed by the complexity and adaptability of the classification task. Additionally, the kernel dynamics, i.e. the response of the membrane potential to spike arrival at an excitatory or inhibitory synapse, influence spike pattern classification. We compared the classification accuracy, and the energy consumption of the neuromorphic classifier (Fig. 5a and Fig. 5b) with the general (primal) support vector machines (SVM), and with the optimized multiclass SVM [17]. The performance is quantified using the effective accuracy, i.e., total spikes classified versus spikes correctly classified (excluding spike detection). The SVM spike sorting performance has been summarized and benchmarked (Fig. 5c) versus four different, relatively computationally-efficient methods for spike sorting. The number of support vectors (typically) scales linearly with the size of the training set, i.e. the number of operations to perform, and, accordingly, the energy per classification scales with the size of the training set. Subsequently, for large-scale tasks such as multi-channel EEG with preserved spatiotemporal information, SVMs can be computationally (and qua energy consumed) demanding. In contrast, neuromorphic network with v neurons and ι inputs evaluates a test sample with the weighted (synaptic) coding requiring only $O(v\times\iota)$ synaptic events, i.e. the number is effectively independent of the size of training set.

IV. CONCLUSION

In this paper, we propose a programmable neural spike classifier based on neuromorphic event-based networks for 64-channel spike sorting system that tracks the evolution of clusters in real-time, offers high re-configurability, and can be directly interfaced locally to neural signal conditioning and quantization circuits. The results obtained in a 90 nm CMOS technology show that an efficient neural spike data classification can be obtained with a low power (less than 9.4 μW/core, corresponding to a 16.7 μW/mm^2 of power density), in the compact, a low resource usage (0.54 mm^2/core area) structure.

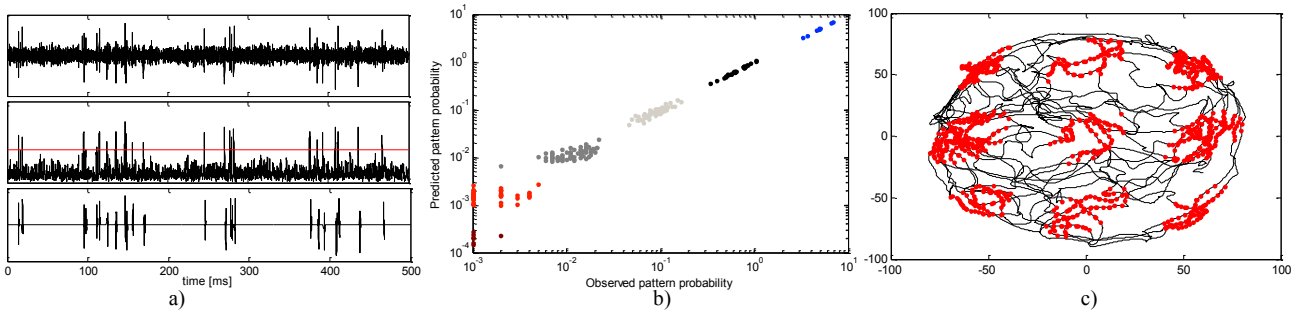

Figure 3: a) Spike detection from continuously acquired data, the *y* axis is arbitrary; top: raw signal, middle: threshold (line) crossings of a local energy measurement with a running window of 1ms, and bottom: detected spikes, b) population patterns in distributed pattern activity of the acquired data, c) trajectory (black) with spike locations superimposed (red).

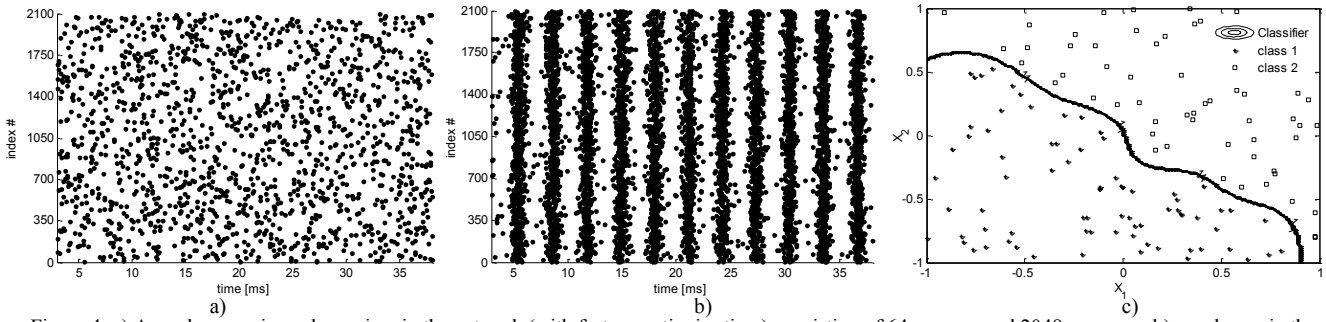

Figure 4: a) Asynchronous irregular regime in the network (with fast synaptic rise time) consisting of 64 neurons and 2048 synapses, b) synchrony in the network (with slow synaptic rise time), c) the separation hypersurface - two classes graphical representation example; bold line represent decision boundaries.

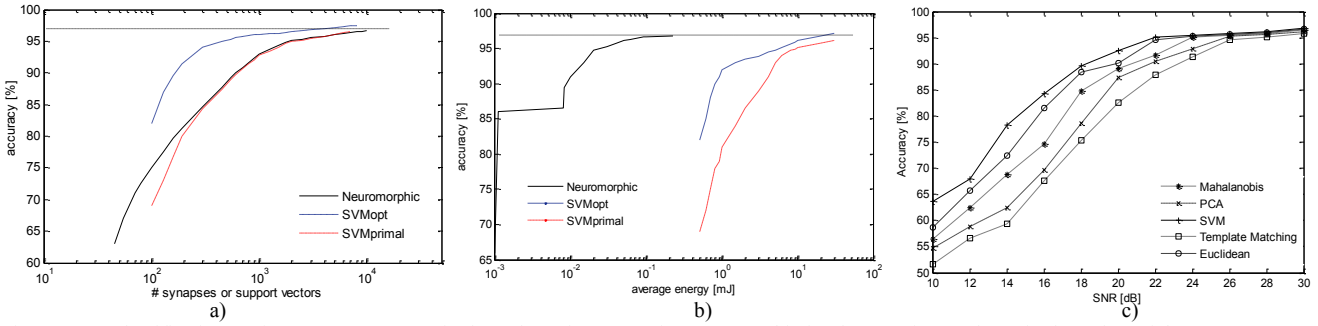

Figure 5: a) Classification performance versus required number of synapses in neuromorphic implementation, and required number of the support vectors in the general primal SVM and the optimized multiclass SVM, b) classification performance versus required average energy per pattern, c) effect of SNR on spike sorting accuracy of the BMI system.

REFERENCES

[1] K.H. Lee, N. Verma, "A low-power processor with configurable embedded machine-learning accelerators for high-order and adaptive analysis of medical-sensor signals", *IEEE J. Solid-State Circ.*, vol. 48, no. 7, pp 1625-1637, 2013.

[2] F. Corradi, G. Indiveri, "A neuromorphic event-based neural recording system for smart brain-machine-interfaces", *IEEE Trans. Biomed. Circ. Syst.*, in press, DOI: TBCAS.2015.2479256.

[3] M. Ambard, S. Rotter, "Support vector machines for spike pattern classification with a leaky integrate-and-fire neuron", *Frontiers Comp. Neursc.*, vol. 6, pp. 1-13, 2012.

[4] S. Hussain, A. Basu, "Morphological learning in multicompartment neuron model with binary synapses", *Proc. IEEE Int. Symp. Circ. Syst.*, pp. 2527-2530, 2016.

[5] I. Segev, M. London, "Untangling dendrites with quantitative models", *Science*, vol. 290, pp. 744-750, 2000.

[6] Y. Wang, S.-C. Liu, "A two-dimensional configurable active silicon dendritic neuron array", *IEEE Trans. Circ. Syst.-I: Reg. Papers*, vol. 58, no. 9, pp. 2159-2171, 2011.

[7] C. Diorio, "Neurally inspired silicon learning: from synapse transistors to learning arrays", *California Institute of Technology,* USA, PhD Thesis, 1997.

[8] M. Tamosiunaite, B. Porr, F. Wörgötter, "Self-influencing synaptic plasticity: recurrent changes of synaptic weights can lead to specific functional properties", *J. Comput. Neurosci.*, vol. 23, pp. 113-127, 2007.

[9] R.C. Froemke, Y. Dan, "Spike-timing-dependent synaptic modification induced by natural spike trains", *Nature*, vol. 416, pp. 434-438, 2002.

[10] S.R. Williams, G.J. Stuart, "Role of dendritic synapse location in the control of action potential output", *Trends Neurosci.*, vol. 26, no. 3, pp. 147-154, 2003.

[11] P. Merolla, K. Boahen, "A recurrent model of orientation maps with simple and complex cells", S. Thrün, et al. (eds.), *Adv. Neur. Infor. Proc. Syst.*, vol. 16, pp. 995-1002, 2004.

[12] C. Bartolozzi, G. Indiveri, "Synaptic dynamic in analog VLSI", *Neur. Comp.*, vol. 19, no. 10, pp. 2581-2603, 2007.

[13] J.N. Kapur, *Maximum entropy models in science and engineering*, Wiley, New York, 1990.

[14] V. Karkare, S. Gibson, D. Marković, "A 75-µW, 16-channel neural spike-sorting processor with unsupervised clustering", *IEEE J. Solid-State Circ.*, vol. 48, no. 9, pp. 2230-2238, 2013.

[15] T.-C. Ma, T.-C. Chen, L.-G. Chen, "Design and implementation of a low power spike detection processor for 128-channel spike sorting microsystem", *Proc. IEEE Int. Conf. Acous., Speech Sig. Process.,* pp. 3889-3892, 2014.

[16] Z. Jiang, Q. Wang, M. Seok, "A low power unsupervised spike sorting accelerator insensitive to clustering initialization in sub-optimal feature space", *Proc. IEEE Des. Autom. Conf.*, pp. 1-6, 2015.

[17] A. Zjajo, R. van Leuken, "A 41 µW real-time adaptive neural spike classifier", *Proc. IEEE Biomed. Health Inform.*, pp. 489-492, 2016.