A Fully-integrated Gesture and Gait Processing SoC for Rehabilitation with ADC-less Mixed-signal Feature Extraction and Deep Neural Network for Classification and Online Training

Yijie Wei, Qiankai Cao, Jie Gu
Department of Electrical and Computer Engineering
Northwestern University,
Evanston, IL

Kofi Otseidu
Intel Corporation
Hudson, MA

Levi Hargrove
Shirley Ryan AbilityLab
Chicago, IL

***Abstract*— An ultra-low-power gesture and gait classification SoC is presented for rehabilitation application featuring (1) mixed-signal feature extraction and integrated low-noise amplifier eliminating expensive ADC and digital feature extraction, (2) an integrated distributed deep neural network (DNN) ASIC supporting a scalable multi-chip neural network for sensor fusion with distortion resiliency for low-cost front end modules, (3) on-chip learning of DNN engine allowing in-situ training of user specific operations. A 12-channel 65nm CMOS test chip was fabricated with 1µW power per channel, less than 3ms computation latency, on-chip training for user-specific DNN model and multi-chip networking capability.**

Keywords—edge processing; deep neural network; inter-chip communication; mixed-signal feature extraction; on-chip learning

I. Introduction

The rapid developments of artificial intelligence create significant demands on low power wearable electronics integrating a large number of heterogeneous sensor components and built-in machine learning capability for human activity assistance, e.g. rehabilitation of amputees. It is reported that over 156,000 patients in the U.S. suffer from the loss of lower or upper-limbs, which provides constant demand of low power devices with built-in artificial intellgiience for patient assistance [1]. To achieve higher efficiency and lower power consumption, the computing tasks are being pushed toward sensor nodes where high dimensional sensor data are directly processed at the sensor edge through on-chip classifiers eliminating the expensive off-chip data communication. For rehabilitation application, accurate human gesture and gait classification holds the most critical roles in detecting user's intent for the operation of prosthetic devices. To improve classification accuracy of user's intent, two areas of improvements are currently being pursued. Firstly, at analog front end, sensor fusion techniques with large numbers of heterogenous sensors such as EMG sensors, strain sensors, and accelerometers are being incorporated for gesture or gait classification. As shown in Fig. 1, 70 channels of heterogeneous sensors were used to obtain accurate classification of user's gesture [2]. Secondly, deep neural network (DNN) based advanced machine learning methods are being explored to achieve higher level of accuracy [3]. However, sensor fusion techniques and use of deep learning method pose significant challenges to an integrated system-on-chip (SoC): (1) Due to the large numbers of front-end channels to be supported, significant area and power costs are consumed for data conversion from analog-digital-converter (ADC) and digital feature extractions; (2) communication and computing bottleneck is easily formed when data from large number of channels are being processed by conventional centralized classifier ASIC; (3) As the DNN classifiers need to be adaptive to user's body characteristics and wearing configuration, e.g. location of sensors, online training is highly desirable but faces significant challenges due to limited on-chip memory and low precision in embedded ASIC chips. In addition, prosthetic device sets a stringent millisecond classification latency requirement, much faster than existing ECG, EEG based classification.

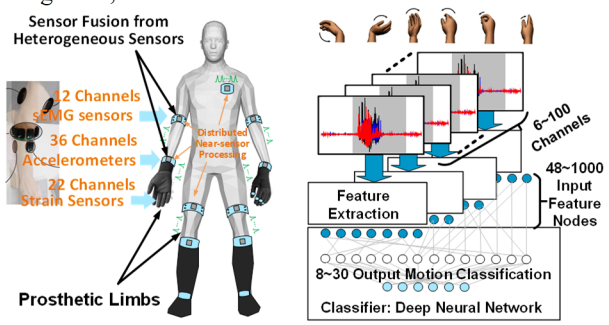

Fig. 1. Illustration of gesture and gait classification with sensor fusion using near-sensor edge device with distributed neural network classifier.

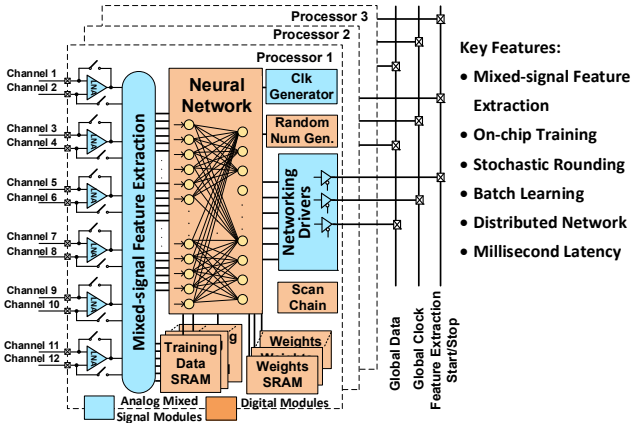

Fig. 2. Top-level chip architecture in this work including LNA, mixed-signal feature extraction, DNN engine and on-chip training circuitry.

To address these challenges, this work proposes a fully-integrated SoC with both analog front end and integrated deep neural network processor with on-chip training capability. The novelty of this work includes: (1) a mixed-signal feature extraction circuit leading to more than 9X reduction in area by elimination of both ADC and digital feature extraction circuits; (2) Empowered by the integrated DNN classifier, the distortion and gain loss from the low-cost low-power analog front end are tolerated from the proper training of the deep neural network classifier leading to significant area and power saving; (3) The design of DNN classifier allows a single large neural network to be split among various sensor nodes leading to 72X reduction of data traffic in sensor fusion environment; (4) On-chip 8-bit in-situ training is enabled with stochastic rounding technique allowing user specific, usage specific adjustment of the classifier operation.

II. Top-level Chip Architecture

Fig. 2 shows the overall implementation of the gesture and gait classification SoC which consists of 6-channels of differential low

978-1-7281-6031-3/20/$31.00 ©2020 IEEE

noise amplifier (LNA) or 12 single-ended LNA bypassed channels for sensor fusion, along with ADC-less mixed-signal feature extraction circuits, integrated deep neural network engine with chip-chip communication and online learning capability. The on-chip neural network contains four fully-connected layers with 12 input-layer neurons, 24 second-layer neurons, 18 output neurons for gesture classification and an additional layer of 18 neurons for gait classification. In addition, the neural network can be interconnected to scale up into a larger network through multiple chips. Low precision on-chip learning with batch processing engine using stochastic rounding, and on-chip training data SRAM, is implemented to enable in-situ training for specific users.

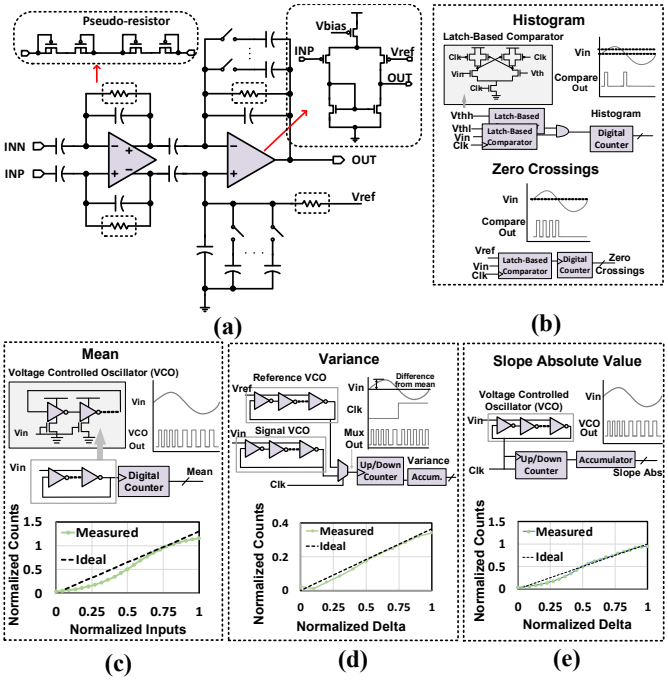

Fig. 3 (a) LNA design with differential-to-single convertion at the second stage amplifier; Mixed-signal feature extraction circuits for (b) histogram and zero crossing, (c) mean, (d) variance, (d) absolute slope sign change.

III. LNA AND MIXED-SIGNAL EXTRACTION CIRCUITS

A. *Front-end LNA and Feasture Extraction Circuit Design*

To support a wide range of front-end sensors, a gain-programmable two-stage LNA is implemented as shown in Fig. 3(a). The programmable feedback capacitors provide a programmable gain up to 57dB supporting a large variety of sensory signals such as EMG, strain sensors, inertia sensors, etc. PMOS pseudo resistors are used to provide high resistance to establish a bandwidth from 5Hz to 3kHz supporting EMG, ECG and other commonly used sensors. The LNA has an input impedance from 3GΩ to 27MΩ within the signal band supporting dry electrodes. Input referred noise of LNA is 9μVrms sufficiently low for EMG signals for gesture classification. Different from conventional support of differential signals for ADC, a differential-to-single end conversion was made at the second stage amplifier to satisfy the signal level required at the mixed-signal feature extraction circuits. The DC level of output signal is set from the feedback structure of the amplifier and adjustable using an on-chip generated reference voltage. The LNA can be by passed for non-EMG types of large signals generated from off-chip sensors.

Fig. 3(b) shows the implemented mixed-signal feature extraction circuits eliminating both conventional ADC and digital feature extraction circuits. Totally eight time-domain features: mean, variance, slope absolute value, zero crossings and four bins of histograms are extracted within 200ms overlapped windows. The feature extraction circuits use simple mixed-signal circuits consisting of only voltage-controlled oscillator (VCO), comparators and counters. The mean feature consists of a VCO and a counter which calculates the averages counts of the VCO within the 100ms windows. The variance feature uses a VCO with another reference VCO in conjunction with a bidirectional counter to accumulate the distance from the mean over time. The slope absolute value feature uses a bi-directional counter which compares the difference in voltage between two 1-millisecond windows. The results of the absolute value of this difference is accumulated over many windows. The histogram features contain four bins and use clocked comparators to count the number of times the voltage falls within a bin. The zero-crossing feature is similar to the histogram with bin threshold set to the reference voltage. The simple design of mixed-signal feature extraction circuits replaces the area costly ADC design showing more than 9X reduction of area as compared with prior work in [4, 5]. All VCO operates at subthreshold voltage from around 300mV to 500mV from LNA output. VCO runs at 20kHz to 100kHz offering high sampling rate for the incoming signals.

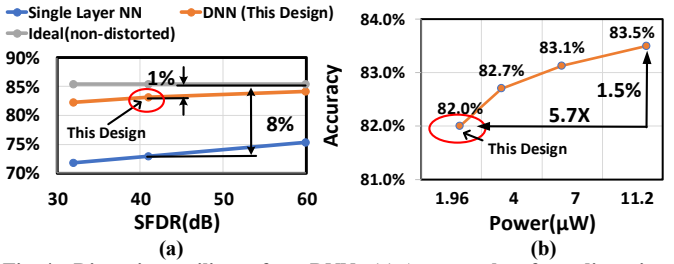

Fig. 4. Distortion resiliency from DNN. (a) Accuracy loss from distortion of VCO non-linearity; (b) Accuracy loss from LNA power reduction.

B. *Distortion Resiliency from Deep Neural Network*

Fig. 3 shows that the measured mean and variance features exhibiting strong distortion from the VCOs due to the subthreshold operation of VCO producing strong 2nd and 4th order distortion which leads to the collapse of feature spaces and degradation from linear classifiers. The distortion from VCO is studied by characterizing VCO as an ADC in term of spur-free dynamic range (SFDR) at near full-scale output range and evaluating the accuracy impact through classifiers. As shown in Fig. 4 (a), such a distortion leads to significant degradation, 8% from simple single layer neural network or linear SVM. However, proper training of the deep neural network using feature characteristics of the mixed-signal circuits can effectively mitigate the distortion impact through the nonlinear sigmoid operation of multi-layer neural network and hence recovers most of the accuracy loss from the mixed-signal distortion. As shown in Fig. 4(a), the VCO of feature extraction circuit in this work has a SFDR of 42 dB due to distortion. However, the neural network reduces the accuracy loss to only 1% as compared with 4% in single layer neuron network. As a result, significant relaxation in LNA and feature extraction circuit's performance, e.g. linearity, is resulted. Fig. 4(b) shows the accuracy impact versus total six channels of LNA power. With 5.7X power reduction of LNA, the accuracy is only dropped by 1.5% due to gain loss. In addition, compared with the digital feature extraction generated by ASIC flow, the mixed-signal feature extraction circuit in this work achieves 2X area saving.

IV. DEEP NEURAL NETWORK DESIGN AND INTER-CHIP COMMUICATION FOR A SCALABLE OPERATION

Channels from sensors in close proximity contain a locality of information as opposed to sensors at a distance. To take advantage of this locality, a distributed neural network is specially developed. This architecture allows the neural network to be divided among

several neural network processors reducing routing and data communication congestion at a single central classifier.

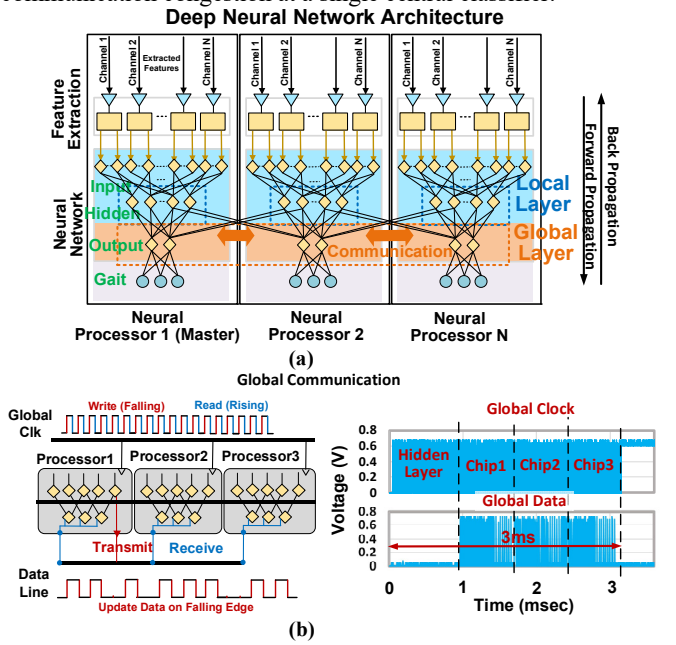

Fig. 5. (a) Deep neural network topology used in this work; (b) Inter-chip communication and measured waveform for three chip operation.

Fig. 5(a) shows the neural network architecture with local and global neural network layers. Each distributed neural processor contains its own private hidden layer and communicates with adjacent sensor and processing nodes only at a global output layer reducing the connections among local neurons. Both local and global layers are fully connected. For a sensor node number of N, a reduction of N times of weights is achieved through the proposed distributed architecture. For bit width B, the communication traffic reduction through the proposed architecture is on the order of 3×N×B. For instance, the separation of local and global layer for 3 chips leads to a 66% memory reduction and 72X saving of communication latency with a small accuracy loss of about 2%. As a result, sensor fusion signals do not need to be routed to a centralized location and only low dimensional feature data processed after local neural networks are transmitted across chips.

Fig. 5(b) shows the networking scheme of the distributed neural network. Global wires, including a start/stop signal, a single clock and a single data line are routed for communication. A master chip is programmed to provide a global clock and each chip is programmed to sequentially send its hidden layer neuron output through the global single-bit data line. While one chip is transmitting, all chips are receiving the data, and rising and falling edges of global clocks are alternatively used for transmitting and receiving to tolerate clock frequency mismatch of individual chips. A networking operation of 36 channels can be successfully processed by three chips within 200us~3ms at a supply from 0.9V to 0.6V. After communication is complete, the master chip will complete the process to produce the final classification result.

V. On-Chip In-situ Training of Neural Network

On-chip learning remains very challenging due to the limitation of bit precision and memory space available on the chip. As a result, there is a lack of prior demonstration of on-chip learning for embedded SoC. This work mitigates the precision and memory limitation using stochastic rounding and batching processing.

A. Low Precision with Stochastic Rounding

Neural network training was conventionally performed with floating point operation offline using a PC. Fig. 6(a) shows the precision loss for gesture classification in Ninapro database as number of bits drop [2]. A 22% loss of classification accuracy is observed with 8-bit precision using backpropagation due to diminishing gradient. The use of stochastic rounding reduces the accuracy loss to only 2% enabling 8-bit on-chip learning. The stochastic rounding was realized by randomly flipping LSB of a weight bit using an on-chip random number generators from linear feedback shift register (LFSR).

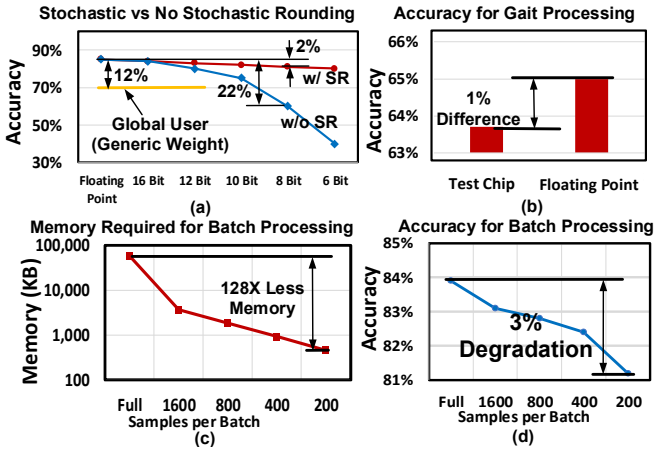

Fig. 6. Accuracy of neural processor with (a) stochatic rounding, (b) gait processing, (c) memory usage vs. batch size, (d) accuracy vs batch size.

B. On-chip Stochastic Batch Processing

On-chip training is highly desirable because each user has his own charateristics of physiological signals and the sensor attachement varies case by case. The use of user-specific training can improve the accuracy from globally trained generic weight by 12% as shown in Fig. 6(a). To enable on-chip learning, each chip is preloaded with a globally trained generic weights at lower accuracy. Users are instructed to perform motions with designated labels. However, a single run of labeled data produces very low accuracy. Multiple batch processing with randomized data sequences allows data to be repeatedly used for training A random number generator based on LFSR is used to randomize the training sequence for each batch during learning. In this design, each batch is run for 6 epoches, to avoid overfitting. However, the batch processing requires all data to be saved on-chip leading to significant memory area overhead. As a result, a tradeoff between on-chip training accuracy and memory space is observed as shown in Fig. 6(c) and (d). As a gesture classification of above 70% is generally satisfactory for the target application, this work trades off memory space with accuracy loss. An on-chip learning memory of 256 examples for each batch run was used with 128X reduction of memory space compared with full dataset with an accuracy loss up to 3%. After a batch of 256 examples are stored onto the chip, the chip runs a batch training with 6 epochs of 256 samples. Four batches of learning are run to reach a final accuracy of about 82%. An additional layer is added on top of neural network for gait classification. By classification of a temporal sequence of gesture operation up to 24 motions, a gait of 5 seconds of continuous motion can be further classified with an accuracy loss of within 1% from off-line PC.

VI. Measurement Results

The test chip was fabricated in a 65 nm low power process. Fig.7(a) and (b) show the measured inference waveforms and computing latency from both feature extraction and neural network operations. The computation including communication is completed within 0.3~3ms scalable with supply voltages meeting the required latency for prothestic limbs of 10~15ms [1]. Fig. 7(c) and (d) show inference power breakdown versus supply voltages. Digital power from neural network and the SRAM dominates the inference operation due to the stringent latency requirement. The

chip operates down to 0.6V consuming about 1.1µW/channel or total 12.31 µW. Fig. 7(e) and (f) show on-chip training waveform, training time and power consumption. Each batch with 6 epochs takes about 36 second to run and consumes 600uW peak power at 0.9V and scales with digital supply voltages. To be able to evaluate the chip performance with more complete sensor fusion environment in comparison with existing publication, multiple DAC boards were used to replay the digitally recorded sensor waveforms from existing database. Various large signals were sent directly into the mixed-signal feature extraction circuits bypassing LNA due to the large signal nature of the sensor. Fig. 7(e) shows the inference results across three databases with the "Rehab" database obtained from 20 real-life amputee patients from our collaborator hospital. Classification accuracy loss remains within 2% for all use cases with 8-bit precision. Fig. 8 shows die photo and a simple demonstration setup with EMG channels. The gestures can be accurately classified through 6 EMG dry-electrode channels.

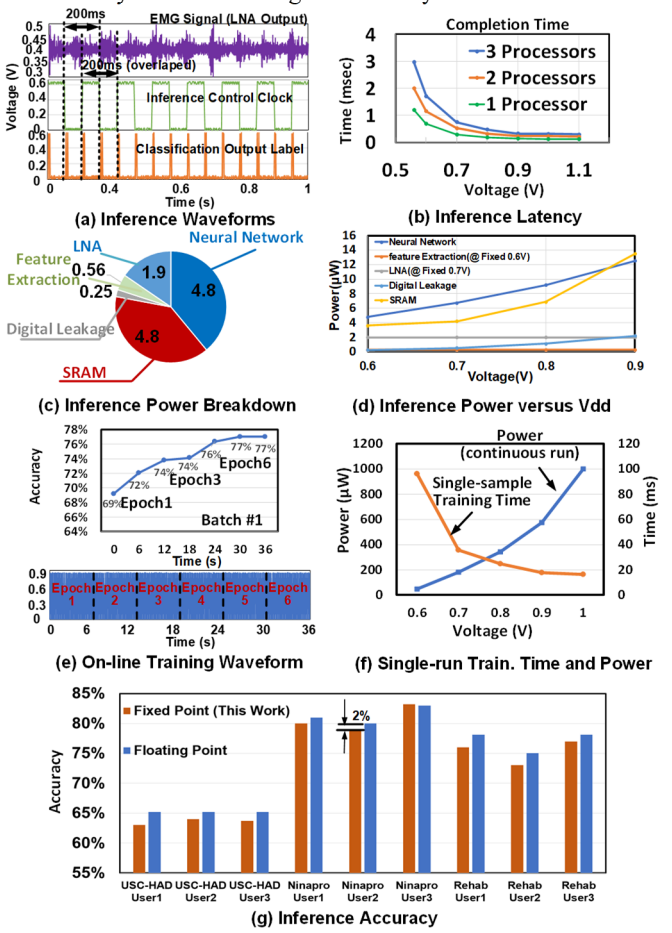

Fig. 7. Measurement results. (a) Inference waveforms; (b) Inference latency; (c) Inference power breakdown; (d) Inference power versus vdd; (e) On-line training waveform and accuracy; (f) On-line training time (single sample) versus power; (g) Accuracy across various data set. Ninapro [2] and Rehab are for gesture. USC-HAD is for gait [6].

Comparison in Table 1 is made on both digital and analog front-end design with recent physiological signal processors. Most of prior work only contains either analog front-end [5, 8] or digital backend [7]. A fully-integrated SVM based SoC for seizure detection was reported in [4], but does not support on-chip batch training from integrated deep neural network as in this work. Compared with [5, 8], our analog front-end design takes less area, runs at a faster speed and consumes the similar or smaller power (Note feature extraction power is also included in our power number while not in previous work). Compared with VCO based front-end design [5], besides area saving, the calibration for distortion has been eliminated due to the use of neural network resulting power/area saving. Compared to neural network architectures in [7], this work consumes less or similar power at digital back-end while operating at significantly faster classification speed with millisecond total computing latency versus seconds' latency in [7] and [4]. Furthermore, for the first time, on-chip deep neural network training through back-propagation and multi-chip networking capability were enabled for physiological signal processing.

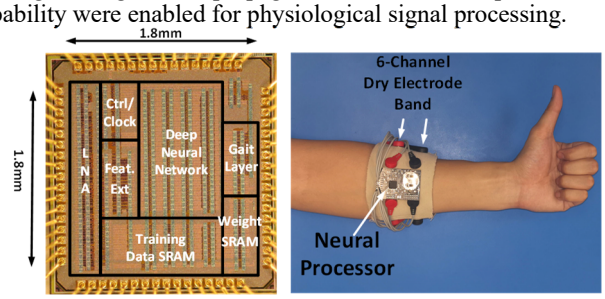

Fig. 8. Die photo and a demonstration setup with EMG signals.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under grant number CNS-1816870.

TABLE I. COMPARISON WITH PRIOR WORKS

		[7] VLSI '17	[4] ISSCC '13	[5] ISSCC'16	[8] VLSI'16	This Work
	Technology	65 nm	180 nm	40 nm	180 nm	65 nm
	Total Area	5.87 mm^2	25 mm^2	0.14 mm^2	1.1 mm^2	3.28 mm^2
	Supply Voltage	0.55V	1.0-1.8V	0.45V-1.2V	0.8V	0.55V-1.0V
	Application	ECG Biometric Authentication	EEG Seizure Detection	LFP Signal Disorder Detection	EEG Feature Extraction	EMG, etc. Motion Recognition
Processor	Memory Size	19.5 kB	96kB	-	-	39 kB
	Clock Freq	2 kHz	1kHz	-	-	100kHz-3MHz
	Latency	1 s	2 s			< 5 ms
	Total Power	1.06 µW	156 µW	-	-	9.6 µW
	Power/Channel	1.06 µW	19 µW	-	-	800 nW
	On-Chip Classifier	DNN	SVM	-	-	DNN
	On-Chip Learning	No	No	-	-	Yes
Analog Front End	Topology	-	LNA+ADC	VCO	LNA+ADC	LNA+VCO
	# of Channels		8	4	16	6/12
	LNA Area/Channel		0.63 mm^2	N/A	0.04 mm^2	0.035mm^2
	ADC/FE Area/Channel	-	0.48 mm^2	0.1 mm^2	N/A	0.011mm^2
	Sampling Rate	-	1kHz	1 kHz	4 kHz	20-100kHz
	Power/Channel	-	740 nW	7 µW	240 nW	326nW

REFERENCES

[1] A J Young, et. al, "Analysis of using EMG and mechanical sensors to enhance intent recognition in powered lower limb prostheses," in *Journal of Neuural Engineering*, 2014.

[2] Manfredo Atzori, et al., "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," in *Scientific Data*, 1:140053 doi: 10.1038/sdata.2014.53, Dec. 2014.

[3] Richard B. Woodward, John A. Spanias, Levi J. Hargrove, "User Intent Prediction with a Scaled Conjugate Gradient Trained Artificial Neural Network for Lower Limb Amputees Using a Powered Prosthesis", *IEEE International Conf. in Medicine and Biology Society (EMBC)*, 2016.

[4] Jerald Yoo, et al., "A 1.83µJ/Classification Nonlinear Support-Vector-Machine-Based Patient-Specific Seizure Classification SoC," in *ISSCC*, 2013.

[5] Wenlong Jiang, et al, "A +-50mV Linear-Input-Range VCO-Based Neural-Recording Front-End with Digital Nonlinearity Correction," in *ISSCC*, 2016.

[6] Mi Zhang and Alexander A. Sawchuk, "USC-HAD: A Daily Activity Dataset for Ubiquitous Activity Recognition Using Wearable Sensors" in *UbiComp*, Sept. 2012, pp 1036-104.

[7] Shihui Yin, et al., "A 1.06 µW Smart ECG Processor in 65 nm CMOS for Real-Time Biometric Authentication and Personal Cardiac Monitoring," in Symposium On *VLSI Circuits*, 2017.

[8] Mahsa Shoaran, et. al, "A 16-Channel 1.1mm^2 Implantable Seizure Control SoC with Sub-uW/Channel Consumption and Closed-Loop Stimulation in 0.18um CMOS", Symposium on VLSI Circuits, 2016