A Gesture Classification SoC for Rehabilitation With ADC-Less Mixed-Signal Feature Extraction and Training Capable Neural Network Classifier

Yijie Wei, *Graduate Student Member, IEEE*, Qiankai Cao, *Graduate Student Member, IEEE*, Kofi Otseidu, *Member, IEEE*, Levi J. Hargrove, *Member, IEEE*, and Jie Gu, *Senior Member, IEEE*

***Abstract*— This article presents a fully integrated gesture and gait classification system-on-chip (SoC) for rehabilitation application. In order to reduce the power consumption and area cost on the analog front end, special analog-to-digital converter (ADC)-less mixed-signal feature extraction (MSFE) circuits were designed to directly generate eight commonly used time-domain features to eliminate the area cost of ADC. A fully connected neural network classifier was implemented supporting: 1) on-chip learning to deliver user-specific training for better classification accuracy; 2) dedicated neural network layer to support gait classification; and 3) multi-chip data communication, which transfers only low-dimensional features from the neural network to minimize the communication bottleneck in a sensor fusion environment. A 12-channel test chip was fabricated in a 65-nm low-power process to demonstrate the proposed techniques. The measurements show an average power of 1 μW per channel and a 3-ms computational latency as required by the stringent rehabilitation requirement. In addition, the MSFE circuits achieve 3× saving of area compared with the conventional approach, while the communication bandwidth was reduced by 100× due to the transferring of only low-dimensional feature data from the neural network among multiple chips.**

***Index Terms*— Biomedical devices, edge device, inter-chip communication, mixed-signal feature extraction (MSFE), neural network classifier, on-chip training.**

I. Introduction

THERE are approximately 2 million amputees living in USA, and the number grows by approximately 185 000 per year [1]. The prosthetic arms and legs embedded with real-time gesture classification capability provide a solution to bring amputees' life back to normal [2]. For realizing real-time gesture recognition as well as some common operations of health monitoring, multiple noninvasive sensors for biomedical signals, such as the electrocardiogram (ECG), surface electromyogram (sEMG), photoplethysmography (PPG), and bio-impedance (Bio-Z), have been utilized to infer users' activities or health information, such as limb movement, heart rate, respiration rate, gait, and mood [3]–[5]. Fig. 1(a) shows an example where an amputee was able to walk on a ramp by wearing an sEMG controlled motorized prosthetic leg [6]. Fig. 1(b) shows a gesture signal acquisition system with sensor fusion techniques containing 12 channels sEMG sensors on the upper arm, 36 channels of accelerometers on the forearm, and 22-channel strain sensor on the hand [7]. The use of sensor fusion techniques, i.e., heterogeneous sensors such as sEMG and accelerometers, brings enhanced classification results compared with homogenous sensors [8]. Note that typical sEMG has a signal level from 0 to 2 mV [9]–[11] while the accelerometer sensors can output a large signal swing of hundreds of millivolts to volts [12], leading to the requirement of programmable gains at the analog front end to accommodate different signal conditioning. Fig. 1(c) shows the signal propagation flow of a rehabilitation system based on a fully connected neural network classifier. Multiple analog channel signals pass through digital feature extraction circuits to generate the corresponding features that are further processed by a neural network classifier for final gesture labels. All computation process needs to finish within 5–15 ms to fulfill the stringent latency requirements of rehabilitation applications [13].

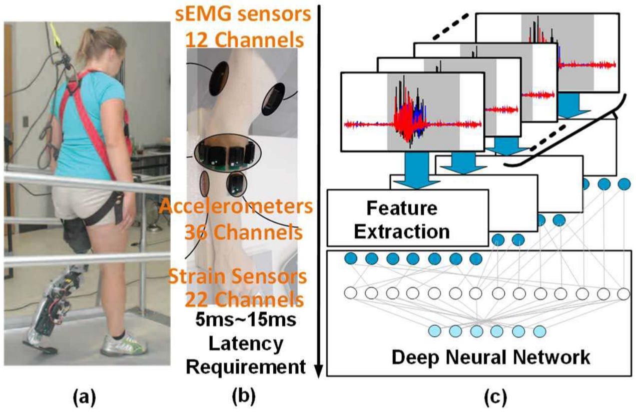


Fig. 1. Bio-signal-based gesture and gait classification system overview with (a) amputee with a motorized prosthetic leg [6]. (b) Use of multiple different types of sensors on the forearm and lower arm [7]. (c) Typical classification flow with the neural network as the classifier.

Manuscript received June 12, 2020; revised August 22, 2020 and October 23, 2020; accepted November 16, 2020. Date of publication December 18, 2020; date of current version February 24, 2021. This article was approved by Guest Editor Qun Jane Gu. This work was supported in part by the National Science Foundation under Grant CNS-1816870. *(Corresponding author: Yijie Wei.)*

Yijie Wei and Qiankai Cao are with the Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208-0001 USA (e-mail: yijiewei2019@u.northwestern.edu).

Kofi Otseidu is with the Intel Corporation, Hudson, MA 01749 USA.

Levi J. Hargrove is with the Shirley Ryan Ability Lab, Chicago, IL 60611 USA.

Jie Gu is with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208 USA.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSSC.2020.3041288.

Digital Object Identifier 10.1109/JSSC.2020.3041288

0018-9200 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Northwestern University. Downloaded on September 20,2021 at 19:30:45 UTC from IEEE Xplore. Restrictions apply.

In order to process the high volume of data transmitted from the multi-channel signals used for rehabilitation application, a powerful microprocessor is conventionally utilized, such as the OMAP microprocessor from Texas Instrument, Dallas, TX, USA [14]. Unfortunately, such a powerful microprocessor consumes hundreds of milliwatts of power, leading to a significant burden on the battery life. To extend the battery life and the capability of biomedical devices, a growing number of developments are utilizing ultra-low-power embedded microcontrollers to realize modern machine learning techniques. For instance, an emotion classifier based on linear discriminator (LDA) using the STM32 microcontroller was proposed as a low-cost solution with a power budget of 35 mW [15]. Specially designed application-specific integrated circuits (ASIC) with embedded machine learning support have also been developed. For example, a fully integrated system-on-chip (SoC) with embedded non-linear support vector machine (SVM) for seizure detection was demonstrated, achieving a 95% seizure accuracy rate with 1.83 μJ/Class [16]. A 16-channel fully integrated seizure detection and stimulation SoC using extracted time-domain features was developed with 16-channel intracranial electroencephalography (iEEG) analog front end and 8-bit successive approximation register (SAR) analog-to-digital converter (ADC) with a power consumption of 0.92 μW/Channel [17]. A seizure detection processor powered by gradient-boosted decision tree with 41.2 nJ/Class was implemented with 1 mm^2 in 65-nm technology [18]. An eight-channel closed-loop neural-prosthetic SoC with linear least-square (LLS) classifier and wireless power and data transmission capability was presented in [19]. A real-time EEG-based emotion recognition system with a multiphase convolutional neural network (CNN) on-chip binary classification processor was implemented in [20].

While significant efforts have been delivered to reduce the power and area of the SoC chip for biomedical applications, one of the design bottlenecks is the requirement of ADC. Especially, the area cost of ADC becomes high when a large number of input channels are to be supported as the case for sensor fusion applications [21]. Multiplexing ADC can mitigate the area cost but may face additional challenges, such as higher design complexity, increased ADC sampling rates, and issues of channel crosstalk [16]. In addition, the analog front-end circuits also suffer from limited dynamic range from the low-noise amplifier (LNA) and ADC, especially under the condition of stimulation. To extend the dynamic range, a voltage-controlled oscillator (VCO)-based front-end amplifier was previously proposed for neural recording, leading to a significant enhancement of the dynamic range under the influence of stimulation artifact leading to a high linear input range of ±50 mV [22]. However, in the above designs, features still have to be extracted by a separate ASIC module. In this work, we propose a further simplification of the architecture by combining ADC and feature extraction circuits using mixed-signal feature extraction (MSFE) circuits as will be described later.

In addition to the power and area issues mentioned earlier, for multi-channel sensor fusion around the human body, the data from the multiple sensing channels need to be transferred at the same time to the central processor or digital classifier for gesture classification. Such a configuration may create a communication bottleneck and routing congestion at the central processor site. In this work, we propose to apply the concept of near-sensor computing by embedding partial processing at distributed sensor nodes. Only low-dimensional data are to be transferred among the chips to significantly reduce the communication data traffic under the sensor fusion environment.

Furthermore, for the classification of the physiological data, the system setting or model weights are typically calibrated or trained offline from desktop computers [23]. However, in the real use case, the signal level and signal features on each channel might vary due to motion artifacts and sweat conditions [24]. All these uncertainties affect the final classification accuracy. To mitigate such issues, an online training capability is highly desirable so that the device can update the weights for a particular user under particular signal conditions.

To address the challenges described earlier, this work proposes a fully integrated SoC integrating LNA, MSFE circuit, and a distributed neural network classifier with on-chip training capability. The contributions of this work are highlighted as follows.

1) This work proposes a novel ADC-less MSFE circuit that directly extracts eight time-domain features including mean, variance, slope absolute, zero-crossing, and histograms leading to a 3× area reduction compared with conventional ADC design [16], [22].
2) A multi-chip distributed neural network is proposed to achieve up to 100× communication data reduction in a three-chip usage case.
3) On-chip 8-bit training was enabled in our proposed neural network classifier by storing training data on the chip and performing randomized batch training with stochastic rounding. The user-specific training allows up to 13% improvement to the classification accuracy compared with a generically trained global user model.

This article is an extension based on the conference presentation in [25] and the earlier analysis focused presentation in [26], which does not have a fully integrated solution, complete neural network implementation with gait classification, and online learning capability.

The rest of this article is organized as follows. The LNA and the ADC-less MSFE circuits are introduced in Section II. Section III presents the analysis of the overall scheme of the proposed distributed neural network classifier. The SoC top-level architecture and implementation are shown in Section IV. Section V presents the measurement results and analysis of the proposed SoC. A comparison with related works is given in Section VI and followed by the conclusion in Section VII.

II. LNA and Mix Signal Feature Extraction Circuit

A. Low Noise Amplifier

To support direct sensing of biomedical signals such as sEMG signals from patients, an LNA was implemented in this work. Typical EMG signals have an amplitude in the

Authorized licensed use limited to: Northwestern University. Downloaded on September 20,2021 at 19:30:45 UTC from IEEE Xplore. Restrictions apply.

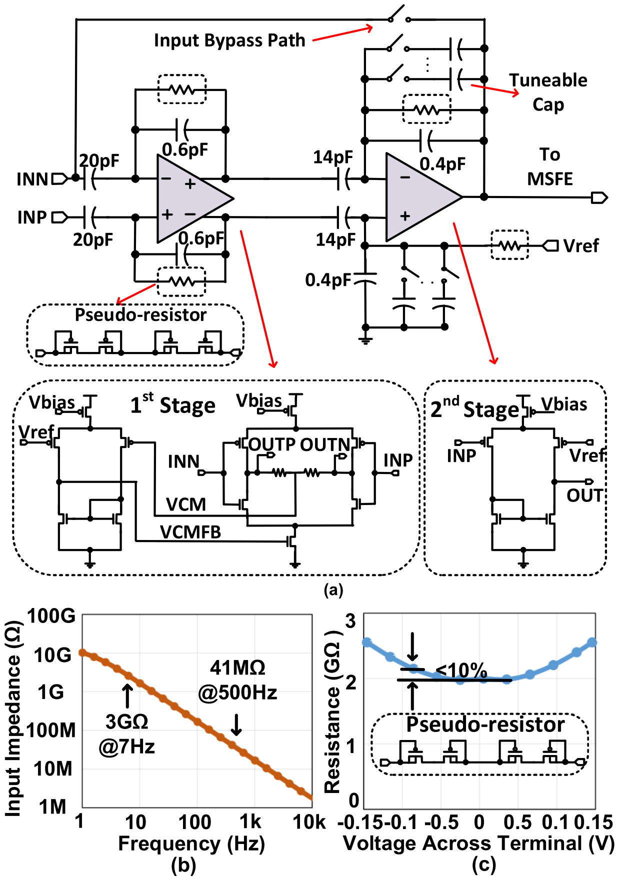


Fig. 2. (a) LNA design with common-mode feedback and differential-to-single conversion at the second stage amplifier. (b) Simulated input impedance of the LNA. (c) Resistance of pseudoresistors with different voltages applied across terminals.

range of 0–2 mV, and the frequency band is from 20 to 500 Hz, according to [9]. The output requirement of the LNA is defined by the requirement from the following MSFE circuits (described in Section II-B) with a full-scale range of around 200 mV and centered around 400 mV and requires a minimum 43-dB gain. When dry electrodes are used, large tissue-electrode impedance from 1 to 5 MΩ requires high input impedance [27] and a higher amplifier gain. Hence, the LNA was designed with capacitive coupling for high input impedance. Meanwhile, a bypass circuit was also added to support large external signals from non-biological signals, e.g., analog signals from an accelerometer, gyroscope, and so on, to enable the support of sensor fusion technology [21].

Fig. 2(a) shows the schematic of LNA that contains a fully differential LNA in series with a gain-programmable differential-input-single-ended output amplifier, which is slightly different from conventional fully differential design for driving ADC [28]. The use of a differential-input-single-ended output amplifier is driven by the need of the MSFE circuits. Metal–oxide–metal (MOM) capacitor and PMOS-based pseudoresistors are used for bandwidth control. Simulation in Fig. 2(b) shows the capacitive coupling provides high input impedance from 3 GΩ to 41 MΩ in 7–500 Hz to support the high impedance by dry electrodes, and the signal attenuation is lower than 1 dB at 500 Hz when the 5-MΩ electrode impedance was considered. The feedback capacitors with a conjunction of the PMOS pseudoresistors deliver the lower bound of the passband to around 7 Hz. Simulation results in Fig. 2(c) show that the resistance of four series-connected PMOS pseudoresistor only varies by 10% within the target working region from +100 to −100 mV across the two terminals. A common-mode feedback amplifier (CMFB) was implemented in the first stage of LNA and helped to achieve a common-mode rejection ratio (CMRR) of −90 dB at 60 Hz [29] with 25-dB gain. The tunable capacitor in the second-stage amplifier supports a 3-bit programmable gain step with a total gain of 32 dB, which is adequate to support the signal ranges from EMG and ECG. The input-referred noise is simulated at 9 μVrms, which is also sufficient for EMG and ECG classification applications. The dc level at the output stage is defined by an on-chip generated reference voltage to fulfill the input requirement of the MSFE circuit. The use of single-end output at the second stage leads to small degradation of CMRR of 6 dB. In addition, to support sensor fusion, large external signals can be directly connected to MSFE bypassing the LNA. The LNA is working under 1-V supply with the first and second stages consuming 240 and 80 nW, respectively. The area of LNA for each channel is 0.035 mm^2, which is mainly determined by the MOM capacitors and similar to related works [30].

B. ADC-Less MSFE Circuits

The conventional analog front end normally implements an ADC to convert the analog signal into the digital domain for further processing. However, the area and power cost of an ADC is significant. In this work, we proposed an MSFE circuit, which replaces both conventional ADC and digital feature extraction circuits.

Fig. 3(a) and (b) shows the proposed MSFE circuit diagram and operating waveforms of each type of circuits, respectively. Each analog channel extracts eight corresponding time-domain features: mean, variance, slope absolute value, zero crossing, and four histograms from four voltage levels [32]. All feature circuits only contain simple VCO, multiplexers (MUXs), comparators, and digital counters. Each feature counters and accumulators update and clear in every 100-ms period.

The "mean" feature circuits consist of a VCO and a counter that accumulate the pulses of the VCO to provide an average count in the 100-ms sampling windows. The "variance" feature fetches the mean feature VCO clock with another fixed-frequency reference VCO in conjunction with a bidirectional counter to accumulate the distance to the reference frequency over time. The "slope absolute" feature uses a bi-directional counter, which compares the difference in voltage at every 1-ms window. The result of the absolute value of this difference is accumulated over 100-ms windows. A clock gating circuit was added into VCOs and the MUXs to prevent glitches during input transition and counts update. 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 input voltage range

Authorized licensed use limited to: Northwestern University. Downloaded on September 20,2021 at 19:30:45 UTC from IEEE Xplore. Restrictions apply.

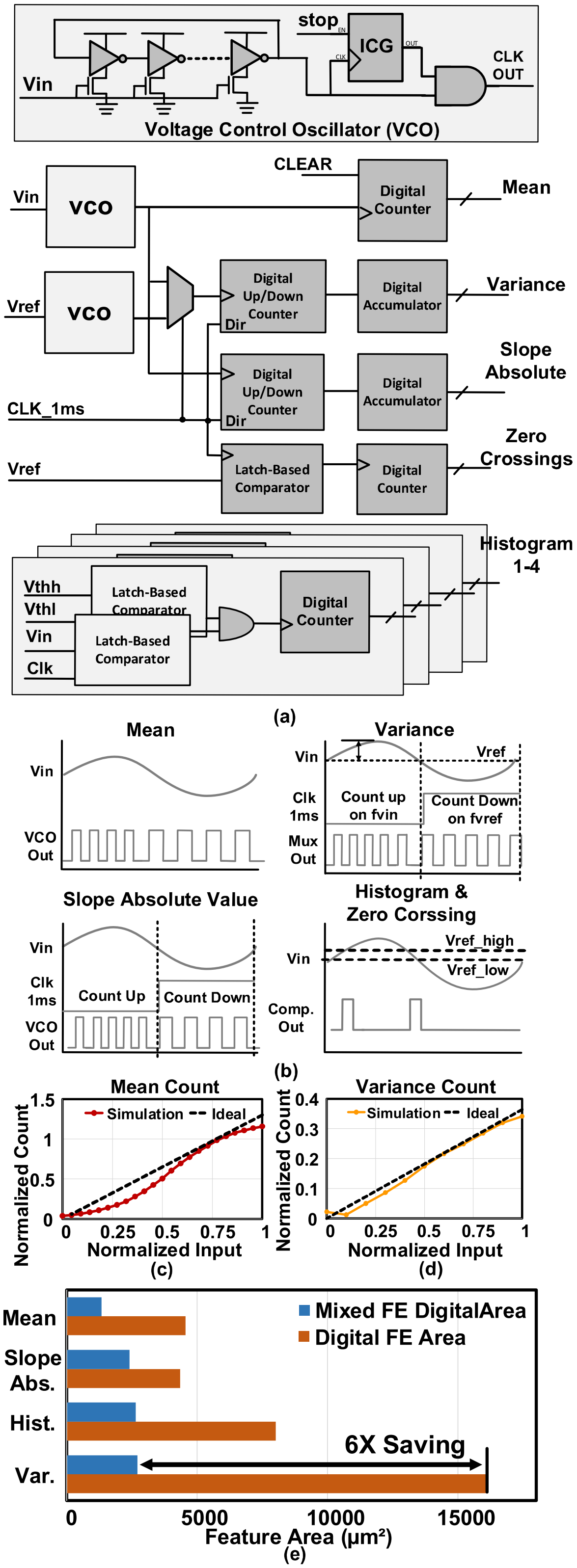


Fig. 3. MSFE circuits in this work. (a) Schematic of mean, variance, slop absolute, and histogram circuits. (b) Operation waveforms of mean, variance, slope absolute, histogram, and zero-crossing features. (c) Normalized mean feature counts with ideal counts. (d) Normalized variance feature counts with ideal counts. (e) Feature extraction area cost comparison between the conventional digital feature extraction circuit and the proposed MSFE circuit.

of all VCOs was set to around 300–500 mV by adjusting the gain and output reference voltage level on the LNA. The output clock frequency is around 20–100 kHz to provide a high sampling rate for signals. Note that the output voltage level and range of LNA was determined by the requirements of the feature extraction circuits. The VCO-based feature extraction circuits exhibit a tradeoff between power and the total number of bits of the extracted features. A higher voltage level will cause the VCO to run at a higher speed with more power consumption. A lower voltage will cause a too small number of counts, hence losing precision in feature extraction. The simulation shows a voltage range between 300 and 500 mV provides an adequate tradeoff between power and precision.

Fig. 3(c) and (d) shows the normalized simulation results of the VCO-based mean and variance feature counts versus the ideal counts, respectively. The simulation curves on both features show distortion, which comes from VCOs working in the subthreshold region. However, the neural network classifier can reduce most of the impact from the VCO distortion by applying distorted feature characteristics during the training process. Our analysis shows less than 1% accuracy impact from the feature distortion compared with the conventional ADC-based solution. The comparison is made with a reference setup using an 8-bit ADC sampling at 2 kHz with ideal feature extraction results feeding into a neural network classifier under ten different users from the Ninapro database.

Due to the asynchronous nature of MSFE circuit, a handshaking scheme was implemented to avoid capturing an erroneous transitional value by the digital back-end classifier. The neural network classifier first sends out a request signal to the MSFE circuits to stop its internal counters at the end of the sampling window. After all the counters are properly latched internally, the MSFE circuits send back a ready signal to notify the neural network to capture the new feature data and release the request signal for the next sampling window.

Fig. 3(e) shows the area cost comparison of the conventional digital feature extraction circuit with the proposed mixed feature extraction circuit in each feature in a 65-nm technology. The proposed feature extraction circuits lead to a 2× area saving on slope absolute feature and 6× area saving on variance feature.

Overall, the proposed 12-channel MSFE circuits occupy 0.01 mm^2/channel and 46 nW/channel. It leads to more than 3× area saving and power saving compared to prior work with conventional ADC circuits [16]. Note that the MSFE circuits serve the purpose of conventional ADC and the digital feature extraction circuits. Hence, a comparison with ADC in prior work is not an exact apple-to-apple comparison.

III. Neural Network Classifier

To classify the user's gesture based on physiological signals measured from the analog front end, we implemented a neural network classifier following the results from MSFE circuits. Different from prior work with inference only operation [16], [23], [31], the implemented classifier contains special features including: 1) online training that is powered

Authorized licensed use limited to: Northwestern University. Downloaded on September 20,2021 at 19:30:45 UTC from IEEE Xplore. Restrictions apply.

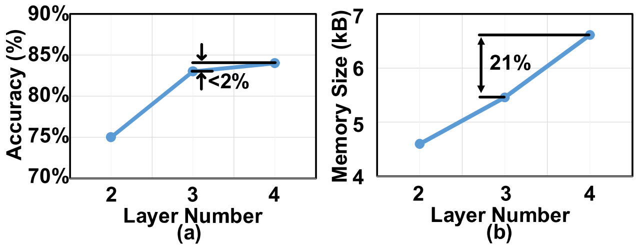


Fig. 4. Impact of layer numbers of the fully connected neural network on (a) classification accuracy and (b) corresponding weight memory size.

by stochastic rounding, random batch processing, and on-chip feature SRAM for training data storage; 2) distributed neural network computation supporting low dimensional data transfer for the chip to chip communication; and 3) an additional layer to support gait classification applications.

A. Analysis of Neural Network Classifier

Since the neural network classifier and associate SRAM will occupy large area to support neural network operation and weight storage, it is crucial to optimize the classifier's structure to save considerable silicon cost. The main optimization is based on the tradeoff between area, computation latency, classification accuracy, and power. In this design, an 8-bit precision was used, which shows only a small accuracy loss compared to the floating points model in the inference task.

Fig. 4 shows the quantized 8-bit classification accuracy results and the corresponding weight memory requirement with a different number of layers neural networks applied to the Ninapro database [7]. As shown in Fig. 4(b), the three-layer setting shows the optimum point setting with no significant accuracy improvements by adding additional layers, which will require an additional 21% weights memory overhead. Based on this analysis, a neural network with three layers is implemented.

B. Distributed Neural Network Architecture

As described in Section I, in the sensor fusion environment, many heterogeneous sensors and analog channels might work collaboratively to feed data into a classifier for a better classification result. These sensors and electrodes might be placed around the human body at various locations and sent back into a central classifier, leading to a communication bottleneck from a large number of channels around the classifier. A solution to reduce the communication data traffic is to reduce the data size by preprocessing all the incoming signals locally. It is possible to divide a fully connected neural network and place all the SoCs near the sensors to process signals locally and then transfer the internal calculation results to the main SoC to finish final processing works.

Fig. 5(a) shows the proposed distributed neural network architecture in a three SoC working scenario. Each individual neural processor is identical and supports 12 input channels and has four fully connected layers with the first three layers for gesture classification and the last layer for gait classification. Each hidden layer has 24 neurons, and the output and gait layers have 18 neurons each. When working in single-processor mode, all the incoming signals are converted into features and passed through all layers within the single chip to generate a classification label. In the multi-chip mode, all incoming features in all processors pass through their local hidden layer first. Then, the lower dimensional data are sent to the global output layer through a communication channel to generate the final collaborative classification result.

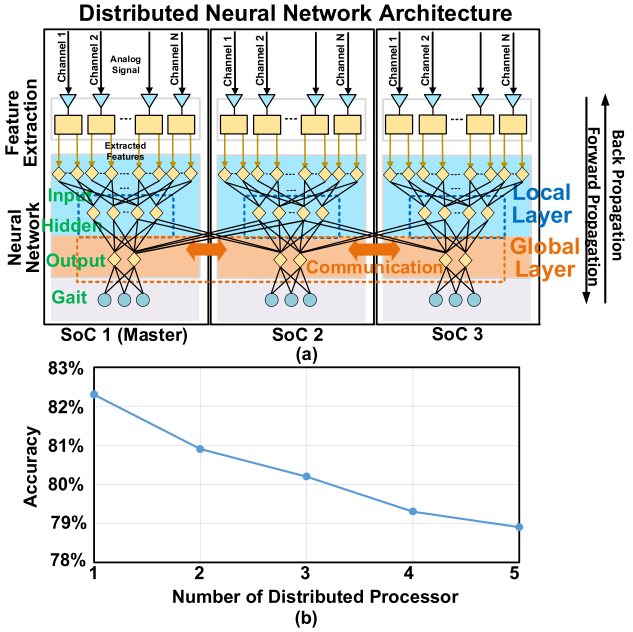


Fig. 5. (a) Neural network architecture used in this work in a three SoCs working scenario. (b) Single-user inference accuracy with different numbers of distributed processors.

As shown in Fig. 5(a), in our neural network, the input and hidden layers are only connected locally instead of fully connected across chips. This connection reduction reduces the data transferring among the chips. Fig. 5(b) shows the classification accuracy with the number of distributed chips changed from single chip to five chips. Due to the reduction of some neural network connections, a small loss of accuracy in the neural network was observed compared to the fully connected neural network. An accuracy loss of about 2% is observed on a three-chip distributed scheme. A five-chip distributed network can be supported in this design due to the accuracy drop and the limited total SRAM space for the output layer's weight. This structure allows decentralized routing at classifiers avoiding signal congestion at the digital back end and reducing the communication traffic significantly. In an example of a 36 analog input channels system, the raw sensor data from an 8-bit, 2-kHz sampling rate ADC to the centralized neural network in every second can be calculated as 36 Ch × 2 kHz × 8 bit = 576 kbit. By applying the proposed near sensor SoC, all input signals are turned into low-dimensional data by MSFE circuits and distributed neural networks. The total data traffic per second is 1 s/100 ms × 24 neurons × 3 SoCs × 8 bit = 5.76 kbit, which is reduced by 100× compared to the conventional scheme, which transmits the raw data.

Fig. 6 shows the communication protocol of the proposed distributed deep neural network. There are three shared wires include a global clock, a data signal, and a sampling window

Authorized licensed use limited to: Northwestern University. Downloaded on September 20,2021 at 19:30:45 UTC from IEEE Xplore. Restrictions apply.

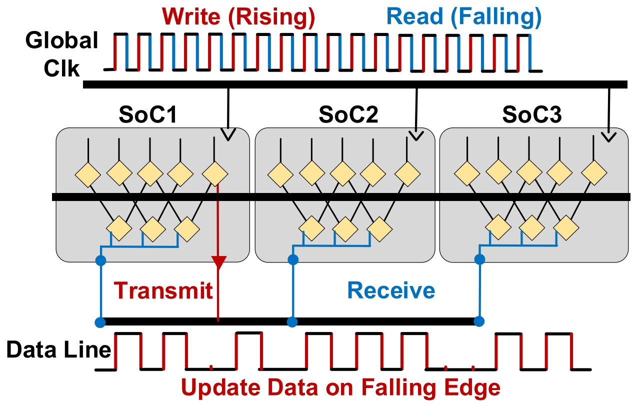


Fig. 6. Communication protocol for the distributed neural network.

signal across all neural processors. Each processor is assigned a unique chip ID, which also defines the master and slave relationship. The master chip is responsible for providing a sampling window signal and data line signal to synchronize all other processors. The classification sequence starts at the rising or falling edge of the sampling windows signal. All processors start to process the received features in the last sampling window to their local layer and sequentially transmit their local hidden layer neuron output by following the order of chip ID. The sender processor transmits data signals at the rising edge of the global clock, whereas all receivers latch data signal at the falling edge. In the end, the master processor will proceed with all the received data to the output neural layer to finish the final classification. For a three-chip network that supports 36 input channels, after each sampling window, the total communication and computation time can be finished within 3 ms under a 0.6-V supply, meeting the stringent latency requirement from rehabilitation.

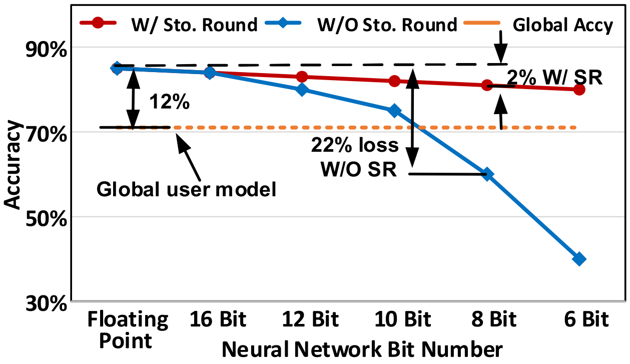


Fig. 7. Comparison of classification accuracy with and without stochastic rounding with different precisions of the neural work. Accuracy using global users' generic weights from floating-point training is also shown in comparison with user-specific training.

Algorithm 1 Stochastic Rounding

```
Procedure Stochastic_Rounding(delta_input_list, random_number, random_threshold, delta_threshold, delta_output_list)
foreach delta_input ∈ delta_input_list do
  if abs(delta_input) < delta_threshold then
    if random_number < rand_threshold then
      Temp_value ←round (delta_input)
      delta_output_list ←FLIP_LSB(Temp_value)
    else
      delta_output_list ←round(delta_input)
    end if
  else
    delta_output_list ←round(delta_input)
  end if
end for
return delta_output_list //return new output delta
```

C. On-Chip Learning by Stochastic Rounding and Randomized Batch Processing

In most existing systems, the neural network training process was performed on an external computer with floating point. However, the classifier is sensitive to the users' signal characteristics and the location of the sensors. A capability of on-chip learning allows user-specific training and adaptation to the change of sensor placement on the body. This work proposed a low-precision on-chip learning scheme by applying stochastic rounding and randomized batching training to the backpropagation training process. Algorithm 1 describes the stochastic rounding operation implemented in this work. When the higher precision weight difference during backpropagation is lower than the preset threshold, the least significant bit (LSB) of the new weight will be randomly flipped. The random numbers are provided by a 32-bit linear feedback shift register (LFSR) as a pseudorandom number generator on the chip. Our analysis of neural network training shows adequate randomness from the simple LFSR circuits. Fig. 7 shows the gesture classification accuracy comparison in the Ninapro database. The accuracy loss of 8-bit training was reduced from 22% without stochastic rounding to only 2% with stochastic rounding rendering a significant saving of the chip area due to the 8-bit implementation.

Fig. 7 also shows the benefit of user-specific training, which can lead to 12% accuracy improvements compared to using a generic weight for the classification from ten users data set from the Ninapro database (referred to as "global user model"). To support the user-specific on-chip learning, a large amount of data points is required to be saved on chip and then randomly processed during training. However, it is very expensive to have all the training data stored in on-chip SRAM. Fig. 8(a) shows the memory space occupation by the number of samples with the corresponding classification accuracy. In this design, an on-chip training SRAM was used to store extracted feature values from up to 256 input samples at one time, which reduces the memory size by 128× by trading off 3% classification accuracy. During the training process, the neural network randomly selects input samples using the on-chip LFSR-based random number generator. For each batch of 256 samples, the chip runs for six epochs. Each epoch contains randomly fetched samples for 256 times. The overall accuracy after four batches of training achieves around 82%, which is close to the target training accuracy. Compared

Authorized licensed use limited to: Northwestern University. Downloaded on September 20,2021 at 19:30:45 UTC from IEEE Xplore. Restrictions apply.

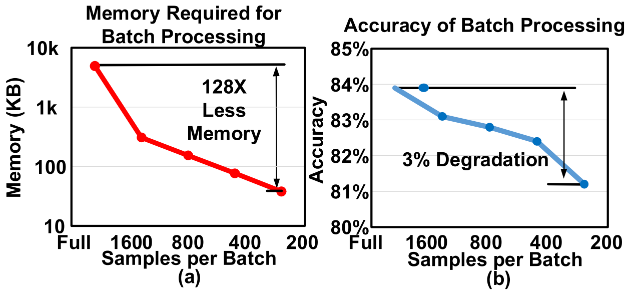


Fig. 8. Tradeoff of user-specific on-chip training. (a) Memory cost versus batch size. (b) Accuracy versus batch size.

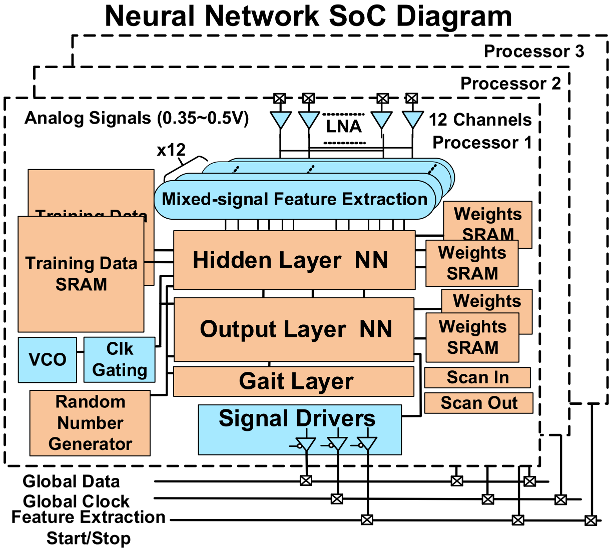


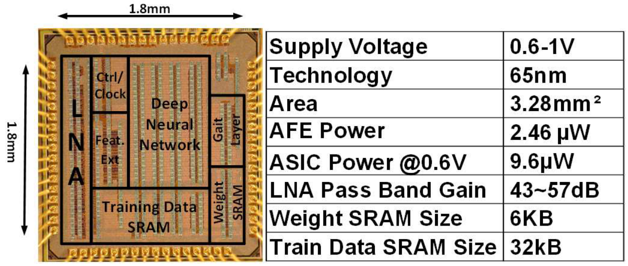
Fig. 9. Top-level chip architecture of the gesture classification SoC.

with offline user-specific training in floating point, due to the limited data points in each running batch, lower training precision, and limited running epochs, 1%–3% accuracy drop was observed on user-specific online training compared with ideal floating-point offline training. However, online training helps mitigate artifacts, such as electrode displacement or sweat conditions.

IV. SoC Architecture

Fig. 9 shows the overall top-level architecture of the implemented gesture and gait classification SoC. In each chip, a total of six differential-input single-end output LNAs are integrated. The external analog sensor signals can also be directly sent into mixed feature extraction circuits. The MSFE circuits generate eight time-domain features per analog channel with a total of 96 features. In the bypass mode, a total of 12 analog single-ended channels are supported. The 8-bit neural network classifier contains 4 fully connected layers with 12 neurons in the input layer, 24 neurons in the hidden layer and 18 neurons in the output layer. In addition, another 18 neurons are added at the last layer for gait classification.

The weight SRAM stores the weights for all neurons and up to 256 feature data can be stored for on-chip training. Global data, global clock, and sampling window data line are used for chip-wise communication up. The LFSR-based pseudorandom number generator is used to support on-chip stochastic rounding and randomized batch training.



Supply Voltage	0.6-1V
Technology	65nm
Area	3.28mm^2
AFE Power	2.46 µW
ASIC Power @0.6V	9.6µW
LNA Pass Band Gain	43~57dB
Weight SRAM Size	6KB
Train Data SRAM Size	32kB

Fig. 10. Die photograph and specification.

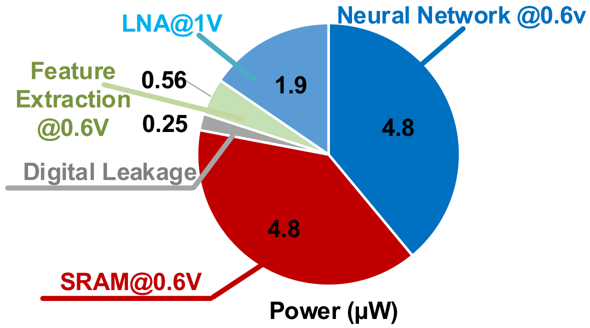


Fig. 11. Measured SoC power breakdown in inference mode.

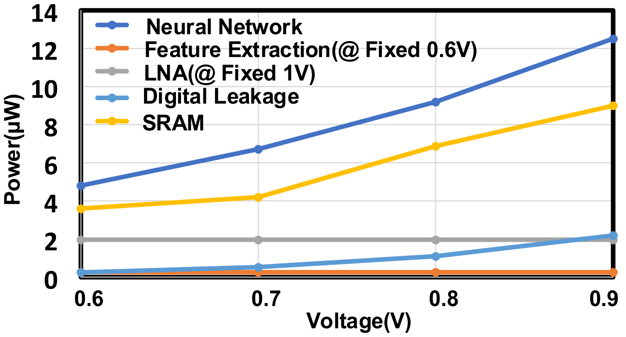


Fig. 12. Power breakdown at different digital supply voltages in inference mode. Power is reported as duty-cycled average power.

V. Measurement Result and Analysis

The proposed gesture classification SoC was fabricated in 65-nm low-power CMOS technology. Fig. 10 shows the die photograph and chip specifications. The total area is 3.24 mm^2. The analog module, including LNA and feature extraction circuits, works at a supply voltage of 1 V, while all other digital modules, including neural network and SRAMs, work down to 0.6 V. The power breakdown in this setting was shown in Fig. 11 for a single-processor mode on the inference task. The total power was measured to be 12.31 μW where neural network and SRAM dominate 80% of total power consumption, whereas the analog front-end and feature extraction circuits consume the rest 20% of total power. The average power per channel is 1.1 μW/channel in total.

Fig. 12 shows the power scaling with digital supply voltages. The digital power consumption scales with the supply voltage with a tradeoff on the computation latency. The power

Authorized licensed use limited to: Northwestern University. Downloaded on September 20,2021 at 19:30:45 UTC from IEEE Xplore. Restrictions apply.

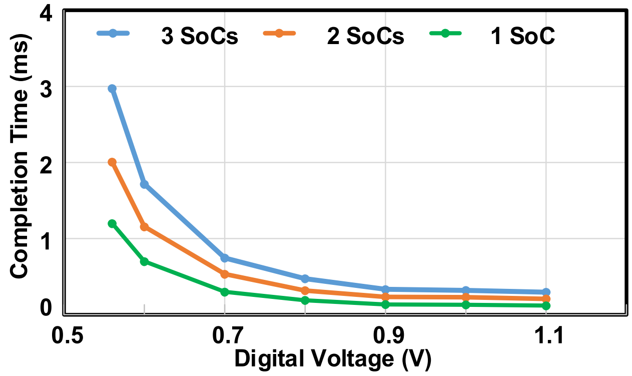


Fig. 13. Inference completion time in three distributed working schemes with various digital supply voltages.

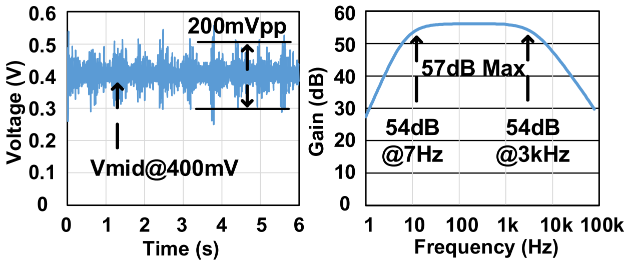


Fig. 14. (a) Measured lower limb EMG signals after LNA. (b) Measured LNA ac gain.

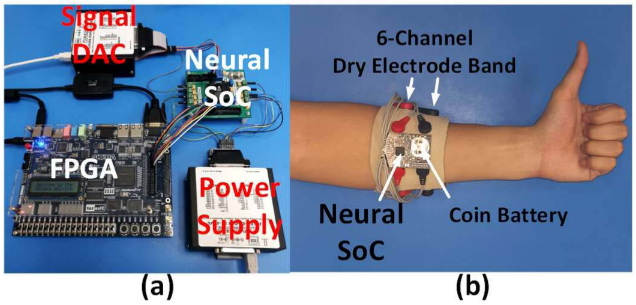


Fig. 15. Testing setup. (a) Setup for replaying recorded database. (b) Real demonstration on human arm with dry electrode arm band.

of the neural network and SRAM was reduced by 2× from 0.9 to 0.6 V. Fig. 13 shows the inference completion time versus the digital voltage applied. Three processors' operation can be finished within 3 ms when all processors are working at 0.6 V, meeting the latency requirement for rehabilitation.

Fig. 14(a) shows the measured lower limb EMG signal when dry electrodes were applied at the LNA gain setting of 46 dB. Fig. 14(b) shows the measured gain plot of the LNA, which shows the lower −3-dB bandwidth at around 7 Hz and 3 kHz, which are sufficient to cover most frequency components in EMG and ECG signals according to [9]–[11].

Fig. 15(a) shows the measurement setup when a digitally recorded database is used. All recorded gesture signals from multiple databases were reproduced by a USB-DA12-8A digital-to-analog converter (DAC) and sent into the chip. Fig. 15(b) shows the coin battery-powered demonstration setup with six-channel dry electrode attached to SoC LNA input. User-specific gestures have been correctly classified from this demonstration setup.

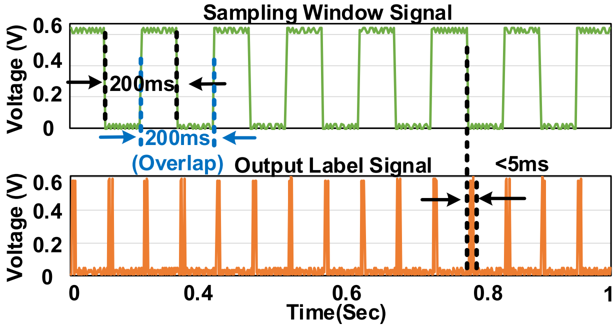


Fig. 16. Measured inference waveforms showing sampling window signal (top) with classification output label signal (bottom) during single SoC inference mode.

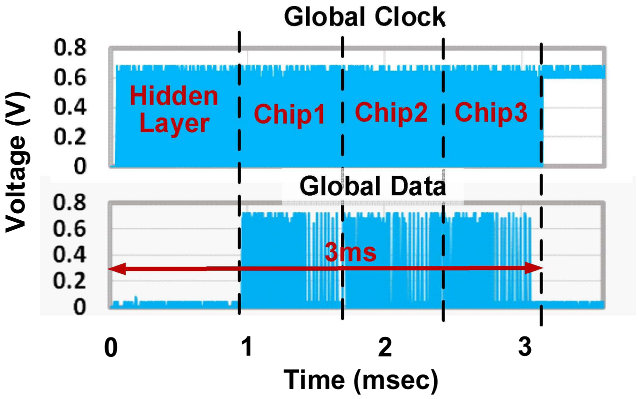


Fig. 17. Measured data signal and global clock signal for three-chip operations.

Fig. 16 shows the measured waveforms of the sampling window signal and classification output signals on a single processor. Overlapped 200-ms windows were used for each inference task. More specifically, in every 100 ms, the feature extraction results were sampled using rising or falling edge of sampling window signal to collect partial feature values that were combined with feature values from the previous 100 ms to form the new feature data for classification. The output label signals from the neural network computation were generated with less than 5-ms latency after the toggling of the sampling window signals at every 100 ms. Since the feature would only be available after each sampling window, the system computation latency was measured after each sampling window is finished.

Fig. 17 shows the measured global clock signal and the global data signal shared by three distributed processors during the inference process. In the beginning, all processors processed the extracted features locally in the local hidden layer. The processors then sequentially sent out the internal results to the master processor to finish the inference process. As shown in Fig. 17, in this measurement, the inference task from three processors completed within 3 ms.

The on-chip training process starts with initial weights trained using a global user database, which can achieve only around 69% accuracy. The 256 user-specific data samples

Authorized licensed use limited to: Northwestern University. Downloaded on September 20,2021 at 19:30:45 UTC from IEEE Xplore. Restrictions apply.

TABLE I
COMPARISON TABLE WITH RELATED WORKS

		[23] 17'VLSI	[16] 17' TBioCAS	[22] 17'JSSC	[17] 16'VLSI	[31] 20'CICC	[19] 14'JSSC	This work
Overview	**Process**	**65 nm**	**180 nm**	**40 nm**	**180 nm**	**180 nm**	**180 nm**	**65 nm**
	Area	**5.87 mm²**	**25 mm²**	**0.14 mm²**	**1.1 mm²**	**16 mm²**	**13.47 mm²**	**3.28 mm²**
	Supply Voltage	**0.55 V**	**1.0-1.8 V**	**0.45-1.2 V**	**0.8 V**	**1 V**	**1.8 V**	**0.55-1 V**
	Application	**ECG Biometric Authenticat ion**	**EEG Seizure Detection**	**LFP Signal recording**	**iEEG Seizure Stimulatio n**	**Ear EEG Emotion Detection**	**EEG Seizure Detection Stimulation**	**EMG Motion Recognition**
Processor	**Memory Size**	**19.5 kB**	**96 kB**	**-**	**-**	**16 kB**	**N/A**	**39 kB**
	Clock Freq	**2 kHz**	**1 kHz**	**-**	**-**	**N/A**	**3.125 MHz- 31.25 MHz**	**100 K- 3 MHz**
	Latency	**1 s***	**2 s***	**-**	**-**	**0.05 s**	**1 s***	**<5 ms**
	Total Power	**1.06 μW**	**156 μW**	**-**	**-**	**10.13 μJ/Class**	**77.91 μJ/Class**	**9.6 μW**
	Power / Ch	**1.06 μW**	**19 μW**	**-**	**-**	**N/A**	**N/A**	**800nW**
	Classifier	**DNN**	**SVM**	**-**	**-**	**DNN**	**LLS**	**DNN**
	On-chip Learning	**No**	**No**	**-**	**-**	**No**	**No**	**Yes**
ADC	**Topology**	**-**	**SAR ADC**	**VCO**	**SAR ADC**	**SAR ADC**	**DMSAR ADC**	**MSFE**
	Area/Ch	**-**	**0.035 mm²**	**0.135 mm²**	**N/A**	**0.15 mm²**	**0.0625 mm²**	**0.011 mm² (MSFE)**
	Sampling Rate	**-**	**1 kHz**	**1 kHz**	**4 kHz**	**500 Hz**	**62.5 kHz**	**20 k- 100 kHz**
Analog Frontend	**# of Ch**	**-**	**8**	**4**	**16**	**2**	**8**	**6/12**
	LNA Area/Ch	**-**	**0.63 mm²**	**N/A**	**0.04 mm²**	**N/A**	**0.25 mm²**	**0.035 mm²**
	Power/Ch	**-**	**7.76 nW**	**7 μW (total)**	**<1 μW (total)**	**1.63 μW**	**6.71 μW**	**326 nW**

***Latency number in [16, 19, 23] includes the sampling time.**

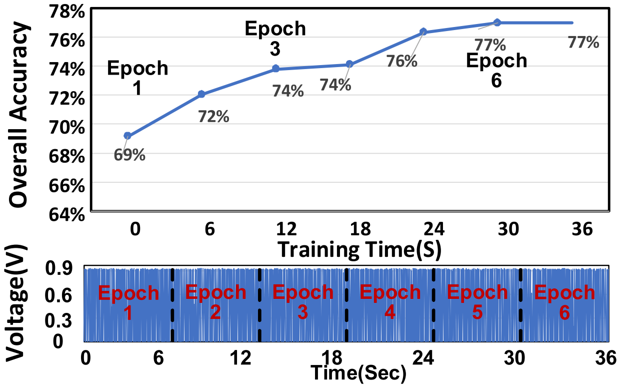


Fig. 18. User-specific on-chip training accuracy with single batch running for six epochs (top) and measured training iteration clock signal (bottom).

of extracted features were loaded into the on-chip training memory for each batch of training data. Fig. 18 shows the accuracy rate and measured training iteration clock signal. In each training epoch, 256 iterations of forward and backward propagation were proceeded, which takes about 6 s. After each training epoch until the sixth epoch, the classification accuracy can increase to 77%. By apply another 3 batches with a total of 24 epochs, the single-user classification accuracy can achieve 82%, which is 13% higher than the accuracy from the global user weights. The training power consumption versus supply voltage with a tradeoff of training time in every training iteration is shown in Fig. 19, and the training time of every iteration takes near 0.1 s when the neural network is working at 0.6 V with a power consumption of 46 μW. To reduce the total training time, a faster training speed can be achieved to 18 ms per iteration by consuming 600-μW power at a digital voltage setting at 0.9 V. Note that power consumption is based on a continuous operation without duty cycling.

Fig. 20 shows the accuracy comparison between the floating-point neural network and our 8-bit neural network quantized from user-specific offline-trained floating-point weights across different databases in various applications. The "USC-HAD" database [30] is for gait processing, whereas the "Ninapro" database [7] is for gesture classification. The "Rehab" database, which includes 20 amputee patients' data, is from our collaborator hospital, Shirley Ryan Ability Lab,

Authorized licensed use limited to: Northwestern University. Downloaded on September 20,2021 at 19:30:45 UTC from IEEE Xplore. Restrictions apply.

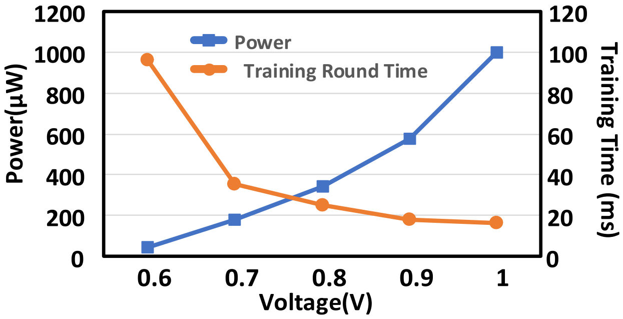


Fig. 19. On-chip training runtime per iteration and power consumption with various digital supply voltages.

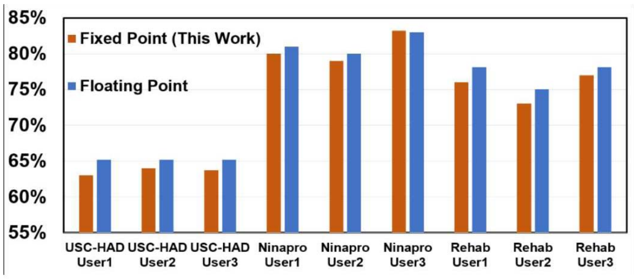


Fig. 20. Accuracy across various data set. Ninapro [7] and Rehab are for gesture classification. USC-HAD is for gait classification [30].

Chicago, IL, USA. The results show that the quantized user-specific classification only has around 2% accuracy loss compared to the floating-point trained neural network.

VI. Comparison and Discussion

Several related works for biomedical signal classification applications are compared in Table I with regard to analog front end and digital back end. The LNA and MSFE circuits in this work show significant benefits of smaller area and power per channel compared with conventional analog-to-digital conversion topology, such as SAR ADC and VCO-based ADC. Comparing work in [16] for EEG seizure detection, since the signal amplification requirement such as bandwidth, noise figure is less stringent for EMG signal, and an 18× area saving was achieved on LNA. More than 3× area saving and power saving was achieved on ADC with the MSFE circuits in this work. For the digital back end, deep neural network classifiers for ECG authentication and mood detection applications were proposed previously [23], [31]. An SVM-based seizure detection classifier was designed in [16], and an LLS-based seizure detection classifier was built in [19]. The computation latency in those prior works [16], [19], [23] was in the order of seconds, which is sufficient for seizure detection applications but is not sufficient to support the millisecond fast response requirement from rehabilitation applications in this work. In addition, only this work supports on-chip training and multi-chip networking scheme among these works.

VII. Conclusion

In this work, a gait and gesture classification SoC for rehabilitation applications is presented. Low-power area-efficient LNA was implemented to support sEMG with dry electrodes as well as different sensing devices, such as accelerometers. An ADC-less MSFE circuit was implemented to directly extract eight time-domain features, which lead to about 3× area saving compared with conventional schemes. A distributed neural network architecture with gait classification was proposed to help reduce 100× communication traffic when multiple SoC chips are used for classification. User-specific online training was also enabled by integrating on-chip SRAM, stochastic rounding, and randomized batch learning. A test chip was fabricated in 65-nm low-power CMOS technology and achieved about 1 μW/channel with a 3-ms computational latency after signal sampling, which meets the stringent requirement of rehabilitation applications.

References

[1] L. Kozak Jean and M. Owings, *National Center for Health Statistics (U.S.). (1998). Ambulatory and inpatient procedures in the United States*, Hyattsville, U.S. Dept. of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, Washington, DC, USA, 1995.

[2] B. H. Hu, N. E. Krausz, and L. J. Hargrove, "A novel method for bilateral gait segmentation using a single thigh-mounted depth sensor and IMU," in *Proc. 7th IEEE Int. Conf. Biomed. Robot. Biomechatronics (Biorob)*, Aug. 2018, pp. 807–812, doi: 10.1109/BIOROB.2018.8487806.

[3] J. A. Spanias, A. M. Simon, K. A. Ingraham, and L. J. Hargrove, "Effect of additional mechanical sensor data on an EMG-based pattern recognition system for a powered leg prosthesis," in *Proc. 7th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, Apr. 2015, pp. 639–642, doi: 10.1109/NER.2015.7146704.

[4] Y.-S. Shu *et al.*, "26.1 a 4.5 mm2 multimodal biosensing SoC for PPG, ECG, BIOZ and GSR acquisition in consumer wearable devices," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2020, pp. 400–402, doi: 10.1109/ISSCC19947.2020.9063112.

[5] E. J. Earley, A. A. Adewuyi, and L. J. Hargrove, "Optimizing pattern recognition-based control for partial-hand prosthesis application," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Chicago, IL, USA, Aug. 2014, pp. 3574–3577, doi: 10.1109/EMBC.2014.6944395.

[6] L. J. Hargrove *et al.*, "Intuitive control of a powered prosthetic leg during ambulation: A randomized clinical trial," *JAMA*, vol. 313, no. 22, p. 2244, Jun. 2015, doi: 10.1001/jama.2015.4527.

[7] M. Atzori *et al.*, "Building the ninapro database: A resource for the biorobotics community," in *Proc. 4th IEEE RAS EMBS Int. Conf. Biomed. Robot. Biomechatronics (BioRob)*, Rome, Italy, Jun. 2012, pp. 1258–1265, doi: 10.1109/BioRob.2012.6290287.

[8] B. Hu, E. Rouse, and L. Hargrove, "Fusion of bilateral lower-limb neuromechanical signals improves prediction of locomotor activities," *Frontiers Robot. AI*, vol. 5, p. 78, Jun. 2018, doi: 10.3389/frobt.2018.00078.

[9] H. Ghapanchizadeh, S. Ahmad, A. J. Ishak, and M. S. Al-Quraishi, "Review of surface electrode placement for recording electromyography signals," Biomed. Res.-Tokyo, Tokyo, Japan, Tech. Rep., 2017, pp. 1–7.

[10] G. L. Soderberg and T. M. Cook, "Electromyography in biomechanics," *Phys. Therapy*, vol. 64, no. 12, pp. 1813–1820, Dec. 1984, doi: 10.1093/ptj/64.12.1813.

[11] J. H. T. Viitasalo and P. V. Komi, "Signal characteristics of EMG during fatigue," *Eur. J. Appl. Physiol. Occupational Physiol.*, vol. 37, no. 2, pp. 111–121, 1977.

[12] ADXL354. *ADXL354 Datasheet and Product Info | Analog Devices*. Analog Devices, Norwood, MA, USA. Accessed: Dec. 1, 2020. [Online]. Available: https://www.analog.com/en/products/adxl354.html?doc=ADXL354_355.pdf#

[13] R. B. Woodward, J. A. Spanias, and L. J. Hargrove, "User intent prediction with a scaled conjugate gradient trained artificial neural network for lower limb amputees using a powered prosthesis," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Orlando, FL, USA, Aug. 2016, pp. 6405–6408, doi: 10.1109/EMBC.2016.7592194.

[14] OMAP3530, *OMAP3530 Data Sheet, Product Information and Support | TI.com*. Accessed: Dec. 1, 2020. [Online]. Available: https://www.ti.com/product/OMAP3530

[15] M. Padmanabhan, S. Murali, F. Rincon, and D. Atienza, "Energy-aware embedded classifier design for real-time emotion analysis," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 2275–2278, doi: 10.1109/EMBC.2015.7318846.

Authorized licensed use limited to: Northwestern University. Downloaded on September 20,2021 at 19:30:45 UTC from IEEE Xplore. Restrictions apply.

[16] M. A. B. Altaf and J. Yoo, "A 1.83 μ J/classification, 8-channel, patient-specific epileptic seizure classification SoC using a non-linear support vector machine," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 1, pp. 49–60, Feb. 2016, doi: 10.1109/TBCAS.2014.2386891.

[17] M. Shoaran *et al.*, "A 16-channel 1.1 mm^2 implantable seizure control SoC with sub-μW/channel consumption and closed-loop stimulation in 0.18μm CMOS," in *Proc. IEEE Symp. VLSI Circuits (VLSI-Circuits)*, Honolulu, HI, USA, Jun. 2016, pp. 1–2, doi: 10.1109/VLSIC.2016.7573557.

[18] M. Shoaran, B. A. Haghi, M. Taghavi, M. Farivar, and A. Emami-Neyestanak, "Energy-efficient classification for resource-constrained biomedical applications," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 8, no. 4, pp. 693–707, Dec. 2018, doi: 10.1109/JETCAS.2018.2844733.

[19] W.-M. Chen *et al.*, "A fully integrated 8-Channel closed-loop neural-prosthetic CMOS SoC for real-time epileptic seizure control," *IEEE J. Solid-State Circuits*, vol. 49, no. 1, pp. 232–247, Jan. 2014, doi: 10.1109/JSSC.2013.2284346.

[20] W.-C. Fang, K.-Y. Wang, N. Fahier, Y.-L. Ho, and Y.-D. Huang, "Development and validation of an EEG-based real-time emotion recognition system using edge AI computing platform with convolutional neural network System-on-Chip design," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 9, no. 4, pp. 645–657, Dec. 2019, doi: 10.1109/JETCAS.2019.2951232.

[21] M. R. Siddiquee *et al.*, "Sensor fusion in human cyber sensor system for motion artifact removal from NIRS signal," in *Proc. 12th Int. Conf. Human Syst. Interact. (HSI)*, Richmond, VA, USA, Jun. 2019, pp. 192–196, doi: 10.1109/HSI47298.2019.8942617.

[22] W. Jiang, V. Hokhikyan, H. Chandrakumar, V. Karkare, and D. Marković, "A±50-mV linear-input-range VCO-based neural-recording front-end with digital nonlinearity correction," *IEEE J. Solid-State Circuits*, vol. 52, no. 1, pp. 173–184, Jan. 2017, doi: 10.1109/JSSC.2016.2624989.

[23] S. Yin *et al.*, "A 1.06-μ W smart ECG processor in 65-nm CMOS for real-time biometric authentication and personal cardiac monitoring," in *Proc. Symp. VLSI Circuits*, Kyoto, Japan, Jun. 2017, pp. C102–C103, doi: 10.23919/VLSIC.2017.8008563.

[24] A. N. Norali, M. H. M. Som, and J. Kangar-Arau, "Surface electromyography signal processing and application: A review," in *Proc. Int. Conf. Man-Mach. Syst. (ICoMMS)*, 2009, p. 1A4-1.

[25] Y. Wei, Q. Cao, J. Gu, K. Otseidu, and L. Hargrove, "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," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Boston, MA, USA, Mar. 2020, pp. 1–4, doi: 10.1109/CICC48029.2020.9075910.

[26] K. Otseidu, T. Jia, J. Bryne, L. Hargrove, and J. Gu, "Design and optimization of edge computing distributed neural processor for biomedical rehabilitation with sensor fusion," in *Proc. Int. Conf. Computer-Aided Design*, San Diego, CA, USA, Nov. 2018, pp. 1–8, doi: 10.1145/3240765.3240794.

[27] X. Zhang, Z. Zhang, Y. Li, C. Liu, Y. X. Guo, and Y. Lian, "A 2.89 μW dry-electrode enabled clockless wireless ECG SoC for wearable applications," *IEEE J. Solid-State Circuits*, vol. 51, no. 10, pp. 2287–2298, Oct. 2016, doi: 10.1109/JSSC.2016.2582863.

[28] C. J. Deepu, X. Zhang, W.-S. Liew, D. L. T. Wong, and Y. Lian, "An ECG-on-Chip with 535 nW/Channel integrated lossless data compressor for wireless sensors," *IEEE J. Solid-State Circuits*, vol. 49, no. 11, pp. 2435–2448, Nov. 2014, doi: 10.1109/JSSC.2014.2349994.

[29] S. Orguc, H. S. Khurana, H.-S. Lee, and A. P. Chandrakasan, "0.3 v ultra-low power sensor interface for EMG," in *Proc. 43rd IEEE Eur. Solid State Circuits Conf. (ESSCIRC)*, Sep. 2017, pp. 219–222, doi: 10.1109/ESSCIRC.2017.8094565.

[30] M. Zhang and A. A. Sawchuk, "USC-HAD: A daily activity dataset for ubiquitous activity recognition using wearable sensors," in *Proc. ACM Conf. Ubiquitous Comput.*, 2012, pp. 1036–1043.

[31] A. R. Aslam, T. Iqbal, M. Aftab, W. Saadeh, and M. A. Bin Altaf, "A10.13uJ/classification 2-channel deep neural network-based SoC for emotion detection of autistic children," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Boston, MA, USA, Mar. 2020, pp. 1–4, doi: 10.1109/CICC48029.2020.9075952.

[32] D. Tkach, H. Huang, and T. A. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition," *J. Neuroeng. Rehabil.*, vol. 7, no. 1, p. 21, 2010, doi: 10.1186/1743-0003-7-21.

Yijie Wei (Graduate Student Member, IEEE) received the B.E. degree in electrical engineering from the University of Mississippi, Oxford, MS, USA, and the North China University of Technology, Beijing, China, in 2017, and the M.S. degree in computer engineering from Northwestern University, Evanston, IL, USA, in 2019, where he is currently pursuing the Ph.D. degree.

His main research interests are in low-power circuit design, analog amplifiers, and mixed-signal integrated circuits.

Qiankai Cao (Graduate Student Member, IEEE) received the B.S. degree in measurement and control technology and instrumentation from the University of Electronic Science and Technology of China, Chengdu, China, in 2018. He is currently pursuing the Ph.D. degree in computer engineering with Northwestern University, Evanston, IL, USA.

He is currently a Researcher with Northwestern University, where he is working in the area of ultra-low-power design/algorithm for very large integrated, mixed-signal ICs. He is focusing on near sensor computing with a low-power algorithm, such as time-domain signal processing.

Kofi Otseidu (Member, IEEE) received the bachelor's degree in both computer science and electrical computer engineering from Cornell University, Ithaca, NY, USA, in 2016 and the master's degree in computer engineering from Northwestern University, Evanston, IL, USA, in 2019.

He is currently an Engineer with Intel Corporation, Hudson, MA, USA.

Levi J. Hargrove (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the University of New Brunswick (UNB), Fredericton, NB, Canada, in 2003, 2005, and 2007, respectively.

He is the Director and Scientific Chair of the Regenstein Center for Bionic Medicine at the Shirley Ryan Ability Lab, Chicago, IL, USA, and an Associate Professor with the Departments of Physical Medicine & Rehabilitation and the McCormick School of Engineering, Northwestern University, Evanston, IL. The primary goal of his research is to develop intuitively controlled bionic limbs for all levels of amputation.

Jie Gu (Senior Member, IEEE) received the B.S. degree from Tsinghua University, Beijing, China, in 2001 the M.S. degree from Texas A&M University, College Station, TX, USA, in 2003 and the Ph.D. degree from the University of Minnesota, Minneapolis, MN, USA, in 2008.

He worked as an I.C. Design Engineer with Texas Instruments, Austin, TX, from 2008 to 2010, focusing on ultra-low-voltage mobile processor design and integrated power management techniques. He was a Senior Staff Engineer with Maxlinear, Inc., Carlsbad, CA, USA, from 2011 to 2014, focusing on low-power mixed-signal broadband system-on-chip (SoC) design. He is currently an Assistant Professor with Northwestern University, Evanston, IL, USA. His research interests include low-power VLSI design and mixed-signal computing, cross-layer integrated power and clock management, and design of machine learning capable edge processing devices.

Dr. Gu was a recipient of the NSF CAREER Award. He has served as a Co-Chair of program committees and conference for numerous low-power design conferences and journals, such as ISPLED, DAC, ICCAD, and ICCD.

Authorized licensed use limited to: Northwestern University. Downloaded on September 20,2021 at 19:30:45 UTC from IEEE Xplore. Restrictions apply.