

A 1.06 μ W Smart ECG Processor in 65 nm CMOS for Real-Time Biometric Authentication and Personal Cardiac Monitoring

Shihui Yin, *Student Member, IEEE*, Minkyu Kim, *Student Member, IEEE*,
Deepak Kadetotad, *Student Member, IEEE*, Yang Liu, *Member, IEEE*, Chisung Bae, *Member, IEEE*,
Sang Joon Kim, *Member, IEEE*, Yu Cao, *Fellow, IEEE*, and Jae-sun Seo, *Senior Member, IEEE*

Abstract—Many wearable devices employ sensors for physiological signals (e.g. electrocardiogram or ECG) to continuously monitor personal health (e.g. cardiac monitoring). Considering private medical data storage, secure access to such wearable devices becomes a crucial necessity. Exploiting the ECG sensors present on wearable devices, we investigate the possibility of using ECG as the individually unique source for device authentication. In particular, we propose to use ECG features towards both cardiac monitoring and neural-network-based biometric authentication. For such complex functionalities to be seamlessly integrated in wearable devices, an accurate algorithm must be implemented with ultra-low power and a small form factor.

In this paper, a smart ECG processor is presented for ECG-based authentication as well as cardiac monitoring. Data-driven Lasso regression and low-precision techniques are developed to compress neural networks for feature extraction by 24.4 $\times$. The 65nm testchip consumes 1.06 μ W at 0.55 V for real-time ECG authentication. For authentication, equal error rates of 1.70%/2.18%/2.48% (best/average/worst) are achieved on the in-house 645-subject database. For cardiac monitoring, 93.13% arrhythmia detection sensitivity and 89.78% specificity are achieved for 42 subjects in MIT-BIH arrhythmia database.

Index Terms—Biometric authentication, ECG, cardiac monitoring, arrhythmia detection, Lasso regression, weight compression, sparse neural network.

I. INTRODUCTION

WEARABLE devices are becoming ubiquitous in our daily lives, many of which featuring ability to sense and process our physiological signals including photoplethysmogram (PPG), electrocardiogram (ECG), bio-impedance, etc. Smart watches such as Apple Watch Series 3 [1] and Samsung Gear S3 [2] can measure our heart rate by analyzing PPG signals. However, to obtain further cardiac health information beyond heart rate, ECG signals can be analyzed. By inspecting the rhythm/shape of ECG beats, initial assessment of cardiovascular diseases can be performed. Due to privacy concerns, instead of processing ECG signals on cloud servers, performing ECG-based cardiac monitoring on local wearable devices has gained a lot of attention.

Several custom ECG processors [3]–[5] have demonstrated arrhythmia detection by continuously monitoring heart rate

variability or frequency spectrum of ECG signals. In addition, detection of abnormal ECG pulse shapes has been presented in [6]. In the commercial market, several wearable ECG recording devices are available for ambulatory monitoring such as Zio patch [7], KardiaBand [8], and Simband [9].

On the other hand, privacy concerns on personal health data necessitate enhanced security to access such wearable devices. Authentication based on biometrics such as fingerprint [10], iris [11], gesture [12], etc., is becoming increasingly popular for wearables. ECG signal has emerged as an attractive biometric modality, due to two key advantages compared to other existing methods: (1) ECG originates from the electrical activity of the heart, thus providing intrinsic liveness proof, and (2) ECG authentication is difficult to spoof, since the ECG signal cannot be easily replicated. In recent years, many prior works proposed ECG-based authentication methods using various machine learning algorithms [13]–[17].

A few ECG-based authentication works have been reported in hardware. The authors of [15] implemented an ECG authentication deep neural network on FPGA, consuming ~ 1 MB memory and 256 mW power. In [16], a cross-correlation based ECG authentication algorithm was implemented on the ARM microcontroller unit (MCU) in a wearable watch. However, both works evaluated authentication only on relatively small databases (90 subjects for [15] and 28 subjects for [16]). Furthermore, considering severe power constraints of wearable devices, compared to FPGAs/MCUs, energy-efficient ASIC processors are desired. Since both ECG authentication and cardiac monitoring share certain computation modules, it is natural to integrate both functionalities in a single chip.

In this paper, we present an ultra-low-power smart ECG processor that performs both ECG-based authentication and cardiac monitoring including arrhythmia detection and anomaly detection [18]. Fig. 1(a) shows the high-level illustration of the proposed ECG processor. The ECG processor was designed for watch-type wearable devices such as smart watches or wristbands. We evaluated the ECG-based authentication performance on three ECG databases: MIT-BIH NSRDB [19] (18 subjects) and ECG-ID [20] (90 subjects) and an in-house ECG database (645 subjects). The arrhythmia detection performance was evaluated on MIT-BIH arrhythmia database [21], targeting atrial premature contraction (APC) or premature ventricular contraction (PVC). Our main contributions include:

- 1) This paper presents, to the best of our knowledge, the

S. Yin, M. Kim, D. Kadetotad, Y. Cao and J. Seo are with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, USA (e-mail: jaesun.seo@asu.edu).

Y. Liu is with Samsung Research Center, Beijing, China. C. Bae and S. J. Kim are with Samsung Advanced Institute of Technology, Suwon, Korea.

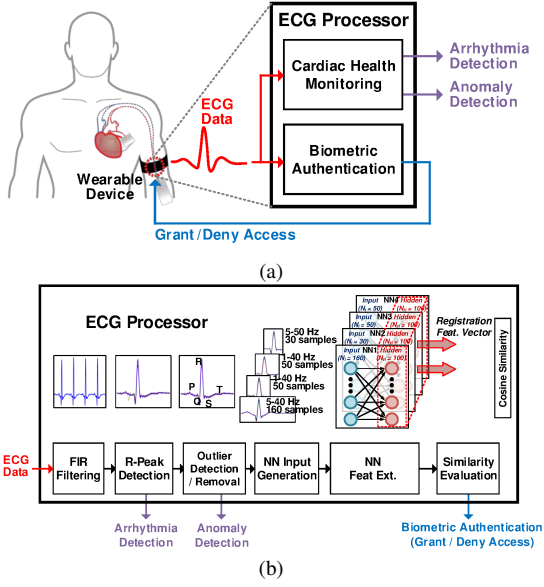


Fig. 1: (a) The proposed ECG processor integrates two functionalities: cardiac health monitoring and biometric authentication. (b) Computation flow of the ECG processor.

first ECG processor that integrates both ECG-based authentication and cardiac monitoring functionalities.

- 2) A novel neural network based ECG feature extraction method is implemented and optimized for on-chip weight memory storage via data-driven Lasso regression. Aided by this, the ECG processor only consumes 1.06 μW at 0.55 V for real-time ECG authentication.
- 3) ECG authentication accuracy is verified on a large 645-subject database considering temporal variability of ECG signals, and achieves a low equal error rate (EER) of $<2.5\%$, demonstrating its practical feasibility.
- 4) For cardiac monitoring, we achieve 93.13% arrhythmia detection sensitivity with a specificity of 89.78% for 42 subjects in the MIT-BIH arrhythmia database.

II. SMART ECG PROCESSOR OPERATIONS

The input to the ECG processor is single-lead raw digitized ECG signal sampled at 250 Hz with 13-bit precision, which goes through FIR filtering, R-peak detection, outlier detection/removal, normalization, neural network based feature extraction and cosine similarity evaluation. Fig. 1(b) illustrates the computation flow of these modules. The proposed ECG processor supports two ECG functionalities with shared hardware: (1) authentication and (2) cardiac health monitoring.

In the authentication mode, the ECG processor works in two stages: registration and identification. In the registration stage, 30 continuous ECG beats of the user are acquired, among which the outliers are detected and removed. The feature vectors (FVs) of all remaining valid ECG beats are extracted by neural networks (NNs), and the mean FV is registered. In the identification stage, the ECG processor acquires 4 valid (outlier-free) ECG beats, and the FVs of the 4 ECG beats are extracted using the same NNs. If the mean of the 4 newly extracted FVs is sufficiently similar to the registered mean

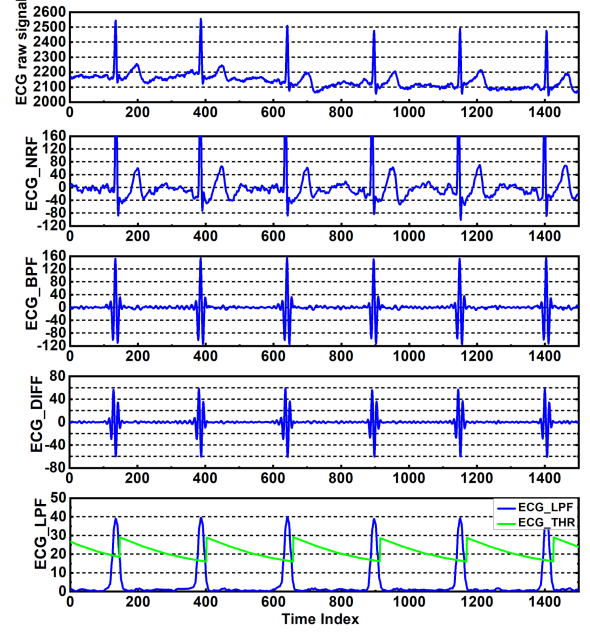


Fig. 2: Representative ECG waveforms. From top to bottom: ECG raw signal, 256-tap FIR NRF output, 42-tap FIR BPF output, differentiator output, and 11-tap FIR LPF output (green line: dynamic threshold).

FV, then the user is accepted; otherwise, the user is rejected. Acquiring 4 valid ECG beats for identification achieves a good balance between the identification time and authentication accuracy, where EER is 1.1% in software simulation using NNs with floating-point precision and without Lasso compression. If we use only 3 and 2 valid beats for authentication, we observed that the EER noticeably degrades from 1.1% to 2.0% and 3.9%, respectively.

In the cardiac monitoring mode, R-peak detection and outlier detection modules are used to detect arrhythmia (irregular ECG rhythm) and anomaly (abnormal ECG shape), respectively. Fig. 3 shows the overall architecture and operations, which are described further in the following subsections. We explored the design space with different choices of the FIR filtering frequency bands, data segmentation, network topology, activation function, similarity metric, etc., using a forward greedy and backward greedy algorithm [22]. In the forward greedy algorithm, we evaluated each candidate model (a neural network and its preprocessing) on the first half of ECG records of the in-house 645-subject ECG databases, and incrementally augment the pool of models with a model that achieves the highest accuracy when combined with other ones in the pool. In the backward greedy algorithm, we incrementally remove a model such that the remained models in the pool achieve the best accuracy. Four final models remained in the pool were selected as the final choice.

A. FIR filtering

A 256-tap FIR noise rejection filter (NRF) with cutoff frequency of 1-40 Hz is designed to reject both high frequency noise as well as DC wandering of the raw ECG signal. The NRF output ("ECG_NRF") is further filtered by a 40-tap FIR

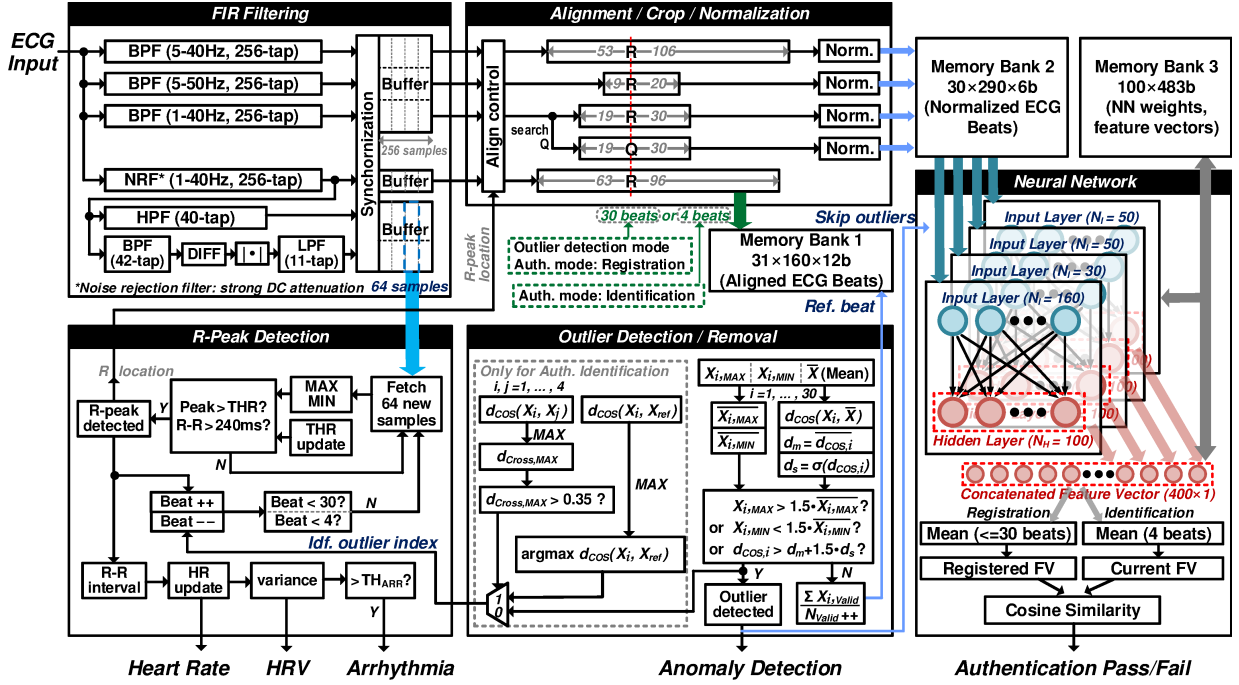


Fig. 3: The overall architecture, dataflow, and computations of the proposed ECG processor.

high pass filter (HPF), as well as a cascade of 42-tap FIR band pass filter (“ECG_BPF”), differentiator (“ECG_DIFF”) and a 11-tap FIR low pass filter (“ECG_LPF”). Simulated waveforms of these filter outputs are shown in Fig. 2. In parallel, we have three additional 256-tap FIR BPFs with cutoff frequencies of 5-40 Hz, 1-40 Hz, and 5-50 Hz. These filters extract ECG information in different frequency ranges, which will be later used for 4 separate NNs. Let us denote these three filters as BPF_5_40, BPF_1_40, and BPF_5_50, respectively.

B. R-peak detection

The outputs of LPF, HPF and four 256-tap FIR BPFs will be buffered in different but time-aligned consecutive 64-sample windows. The HPF output is used to accurately determine the maximum/minimum peaks in a 64-sample window. The LPF output is compared with a dynamic threshold (“ECG_THR”) to detect the R-peak of ECG beats within a window [23], as shown in Fig. 2 (bottom). When a valid R-peak is detected, we extract a 160-sample segment from the buffer for the outputs of NRF and BPF_5_40 aligned at R-peak, two 50-sample segments from the buffer for the output of BPF_1_40 (one aligned at R-peak, the other aligned at Q-point), and a 30-sample segment from the buffer for the output of BPF_5_50 aligned at R-peak. The aligned 30 or 4 ECG beats after NRF are stored in on-chip memory for outlier detection.

C. Arrhythmia detection

Among the various types of arrhythmia that concern cardiac health, we focus on the detection of sudden change in the R-R interval. This can be caused by premature contractions such as atrial premature contraction (APC) and premature ventricular contraction (PVC), for which example ECG waveforms from

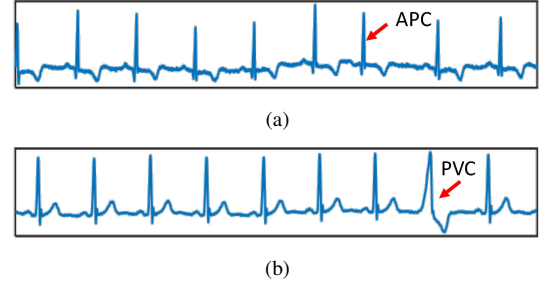


Fig. 4: ECG waveforms of (a) atrial premature contraction (APC) and (b) premature ventricular contraction (PVC) are shown, from record 100 and 116 of [21], respectively.

the MIT-BIH arrhythmia database [21] are shown in Fig. 4. We first estimate the instantaneous heart rate as the inverse of R-R interval. Then, the instantaneous heart rate variability (HRV) is estimated by computing the standard deviation of the past three consecutive heart rates. If the instantaneous HRV is above a threshold, we define that an arrhythmia is detected. If needed, the stored ECG beats can be exported for further diagnosis by cardiologists. Our arrhythmia detection algorithm is similar to that in [5], where HRV in a non-overlapping 10-second time window is estimated. In our ECG processor, HRV in a overlapping window of four ECG beats is estimated, which can more precisely locate the ECG beat when arrhythmia occurs.

D. Outlier detection/removal

Following the R-peak and arrhythmia detection module, the outlier detection module is present. It is notable that we employ the same outlier detection hardware for both authentication and cardiac monitoring modes, but in an opposite manner. For authentication, the objective is to detect the outliers and remove them, to form a representative (outlier-free) set of ECG beats for personal ECG feature extraction; for cardiac

monitoring, we actually detect the outliers and save them, since they are considered as potential anomalies for the individual's cardiac health. Further details of the outlier detection/removal module for both operation modes are described below.

Authentication mode: In the authentication mode, after a certain number of ECG beats are obtained, we find outliers among the collected ECG beats and discard them, in order to acquire an outlier-free, representative ECG beat set for each individual. The outlier removal algorithm is adopted from the techniques proposed in [24]. An ECG beat is defined as an outlier when any of the following three conditions is satisfied:

- 1) $X_{max} > 1.5\tilde{X}_{max}$,
- 2) $X_{min} < 1.5\tilde{X}_{min}$,
- 3) $d_{cos} > \bar{d}_{cos} + \gamma \cdot \sigma$,

where X_{max}/X_{min} are the maximum/minimum value of the ECG beat, respectively, and $\tilde{X}_{max}/\tilde{X}_{min}$ are the median of maximum/minimum values of all acquired ECG beats. d_{cos} is the cosine distance of the ECG beat to the mean ECG beat, $\bar{d}_{cos}$ is the mean of the cosine distances of all ECG beats to the mean ECG beat, and σ is the standard deviation of the cosine distances. γ is a configurable parameter in our hardware design, whose value can be 0.25, 0.5, 1 or 2. Smaller γ represents more stringent outlier removal. Depending on the scale and noise level of the target ECG data, different γ values might be desired. In our experiments, 0.5 is found to be the optimal γ value for all three ECG databases we investigated. The cosine distance between two ECG beats is defined as:

$$d_{cos}(\mathbf{X}_1, \mathbf{X}_2) = 1 - \frac{\mathbf{X}_1^T \mathbf{X}_2}{\|\mathbf{X}_1\|_2 \|\mathbf{X}_2\|_2}, \quad (1)$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are two 160-sample ECG beats.

To ensure that sufficient ECG beats are used in the identification stage, in case an outlier is detected and discarded, new ECG beats are continuously read in until four ECG beats are collected. With outlier removal, EER improves from 2.60% to 1.70% for the in-house 645-subject ECG database.

Cardiac monitoring mode: In the cardiac monitoring mode, in addition to detecting abnormal heart rate, we can also detect abnormal ECG pulse shape. As aforementioned, outliers among acquired ECG beats will be detected and removed in the authentication mode for better feature extraction quality. By reusing this identical outlier detection module, outliers in every 30 ECG beats will be continuously detected and reported in the cardiac monitoring mode. The ECG beats detected with abnormal ECG pulse shapes might be classified as an anomaly for cardiac health, as shown in Fig. 4(b). As needed, all 30 ECG beats can be exported for further analysis by cardiologists. Similar to the authentication mode, cosine distance threshold parameter γ can be varied to adjust the criteria for anomaly detection in the cardiac monitoring mode.

E. Normalization

Before the four ECG streams are sent to four corresponding NNs, we normalize the data such that NN input values are bounded within a certain range. Normalization is performed in two steps. First, the aligned ECG beats are normalized to zero-mean and unit-variance across different beats. Second, we

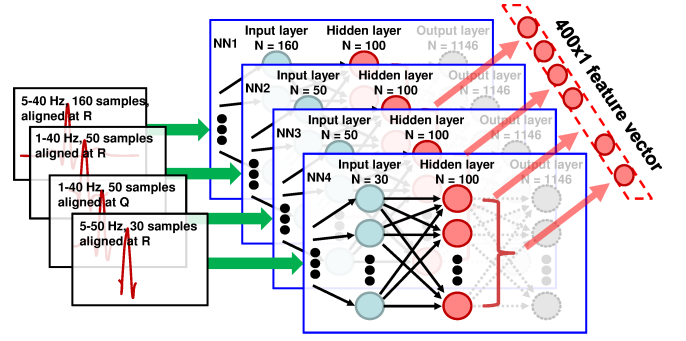


Fig. 5: Four parallel neural networks for ECG feature extraction. Four 100×1 FVs are concatenated to form a 400×1 FV.

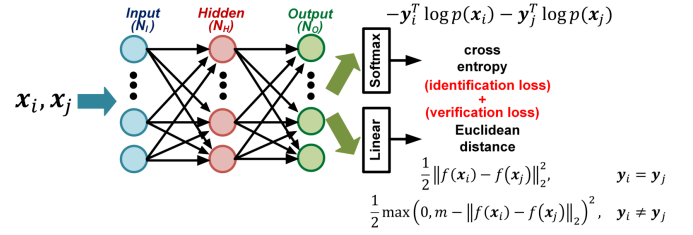


Fig. 6: Identification and verification loss function used for training the neural networks.

normalize the ECG beats by the mean and standard deviation of ECG beats used for NN training. Then, ECG beats become normalized in both inter-beat and intra-beat dimensions.

F. Neural network based feature extraction

As shown in Fig. 5, four parallel NNs with input layer, one hidden layer, and output layer are designed to extract features from different frequencies and alignment (e.g. aligned at Q versus R). For each NN, there are 100 hidden layer neurons and 1,146 output layer neurons. The number of input neurons varies from 30 to 160, depending on the number of samples. The activation function of the hidden layer is $\tanh(x)$. The NNs were trained on the first half of ECG records of the in-house 645-subject ECG database with 10-fold cross-validation. Each subject has a different number of ECG records acquired over different time, and the total number of ECG records is 1,146. We could map different ECG records from the same subject to the same output neuron or to different output neurons. We tried both mapping schemes and found that the latter one leads to better EER, as it could extract ECG features in the hidden layer more effectively when we spread out each ECG record's data to different neurons. Therefore, we employed 1,146 output neurons in the NNs.

Training: We first pre-train a two-layer deep belief network as the initial weights values of NN [25]. Then, we use the identity labels of samples as the supervision information for fine tuning. After training is done, the intention is to use the hidden layer output as the feature descriptor.

As illustrated in Fig. 6, two loss functions are collectively employed to improve the accuracy: identification loss function and verification loss function [22]. The identification loss function maximizes the difference of ECG features from different users. The cross entropy value after a softmax layer

is evaluated for the identification loss. On the other hand, the verification loss function minimizes the ECG feature variation from the same user, considering the temporal variation in ECG signals from the same user. Pairs of ECG data x_i, x_j are fed to the NNs. If they are from the same subject, the Euclidean distance between their network outputs is minimized; otherwise, the Euclidean distance is maximized to be greater than a specified threshold m .

Classification: Classification is performed with feedforward propagation using the trained NNs. Four 100×1 FVs are extracted from the hidden layer of four NNs and concatenated to form a 400×1 FV. The average 400×1 FV over all valid beats is considered as the final FV. Since the hidden layer is directly used for FVs, weights between hidden and output layers are not required for classification (only used in training).

G. Cosine similarity evaluation

In identification stage, cosine similarity is computed between the current (identification) FV and the registered FV:

$$\cos_{sim} = \frac{FV_{new}^T FV_{reg}}{\|FV_{new}\|_2 \|FV_{reg}\|_2}. \quad (2)$$

When $\cos_{sim}$ is above the pre-defined threshold, the user will be accepted; otherwise, the user will be rejected.

III. LOW-POWER DESIGN OPTIMIZATION

Power consumption is critical for our ECG processor due to the limited power budget of wearable devices. The ECG processor employs extensive clock gating based on module-level activity, and is optimized from software and hardware perspectives to substantially reduce the power consumption.

A. Selective precision in hardware design

To reduce the power and area of the ECG processor, we adopted fixed-point precision representation. The input raw digitized ECG exhibits 13-bit precision. However, there is no need in keeping this precision throughout the entire signal processing flow. To optimally reduce the power without hampering accuracy, we selectively reduced the precision for different modules to the lowest before accuracy degradation occurs. The final precision values optimized for FIR filter coefficients, FIR filter signals, R-peak detection, normalization, NN based feature extraction, and similarity evaluation modules are 8-bit, 13-bit, 13-bit, 11-bit, 12-bit, 6-bit and 9-bit, respectively. Note that the 32-bit floating-point NN weights are quantized to only 6-bit, reducing the weight storage by $5.3 \times$.

B. FIR filter design

We adopted the systolic architecture [26] for our direct form FIR filters design to improve the throughput. Since our FIR filters are linear-phase, the coefficients are always symmetrical around the center coefficient. Exploiting this, we used pre-adder to sum the samples associated with two symmetric coefficients, halving the number of necessary multipliers. In addition, we reduced the precision of the FIR coefficients to 8-bit. Consequently, a portion of the heading and ending

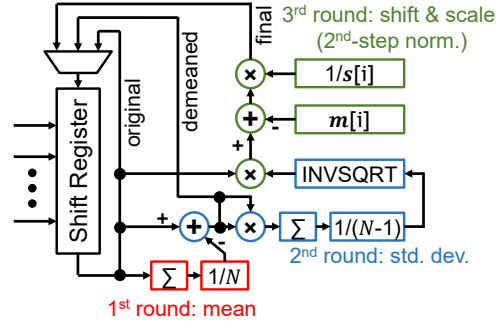


Fig. 7: Normalization module based on three rotation rounds.

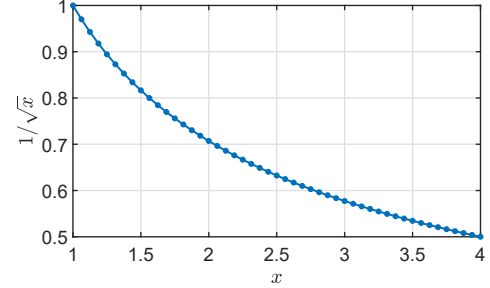


Fig. 8: Piecewise linear approximation of $1/\sqrt{x}$ with 48 uniformly spaced segments on $[1, 4]$.

coefficients were quantized to zeros, which effectively reduced the orders of the FIR filters. Nonetheless, no significant impact was found on the filtering effect by discarding these coefficients. Therefore, we implemented reduced-order FIR filters to reduce area and power. For example, the 256-tap NRF, BPF_5_40, BPF_1_40 and BPF_5_50 were simplified to 148-tap, 178-tap, 178-tap and 178-tap, respectively. Delay lines were added to synchronize signals across different channels.

C. Normalization module design

As mentioned in Section II.E, the normalization of ECG beats consists of two steps. First, we compute the mean and standard deviation of an ECG beat. Then, we subtract the ECG beat by the mean and divide by the standard deviation. After this first step normalization, each ECG beat becomes zero-mean and unit-variance. Finally, we subtract the normalized ECG beat by a global mean vector and divide by a global standard deviation vector. We could have processed all samples in an ECG beat in parallel to reduce the normalization latency to a few clock cycles. However, since normalization is performed while reading in new ECG beats and the ECG beat rate is relatively low (60-100 bpm), the normalization module can tolerate higher latency (e.g. a few hundred clock cycles).

We designed a rotation-based normalization module (Fig. 7), which is time-multiplexed for three rounds of computation. The ECG beat is fed into a shift register from QRS alignment module. In the first round of rotation, the mean of the ECG beat is computed, and in the second round, the standard deviation of the ECG beat is computed. In the third round of rotation, the second-step normalization is completed. By reusing the data path for all samples in an ECG beat, only three multipliers are required for the normalization of one channel.

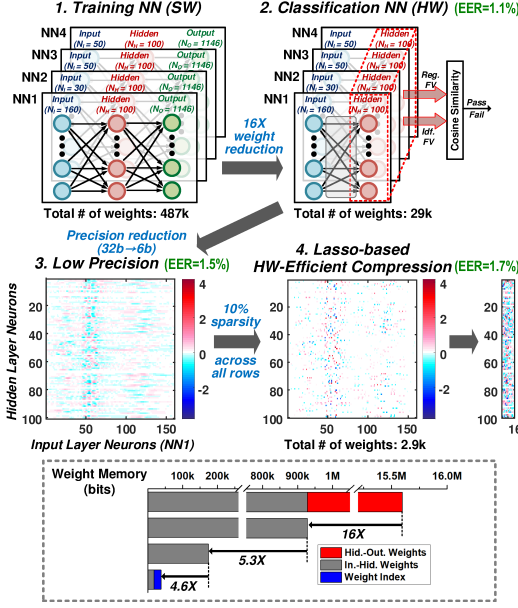


Fig. 9: Summary of NN compression. 1: software model; 2: output layer removal; 3: quantization; 4: Lasso regression.

D. Inverse of square root module design

When we compute the cosine distance in outlier detection module, the standard deviations in normalization and arrhythmia detection module or the cosine similarity, we need to evaluate the inverse of square root (INVSQRT) function ($1/\sqrt{x}$), which is a nonlinear function. We approximated this nonlinear function with a piecewise linear function. In particular, we pre-computed the values (offset and slope) for 48 uniform-length segments between $[1, 4)$ as shown in Fig. 8. The input of the INVSQRT function has 14-bit precision. The six most significant bits (MSBs) of the input are used for segment index. The INVSQRT value for the input range beyond $[1, 4)$ is obtained by bit-shifting. For example, if the input is 0.2, we first left-shift the input by 4 bits (i.e. $0.2 \times 2^4 = 3.2$), to make the input fall within the valid range of $[1, 4)$, then we need to left-shift the INVSQRT output by 2 ($= 4/2$) bits. In general, the amount of shifting at the output is half of the amount of shifting at the input, due to the fact that:

$$1/\sqrt{x} = 1/\sqrt{x \times 2^{2k}} \times 2^k. \quad (3)$$

E. Lasso regression based sparsification of neural networks

As mentioned in Section II.F, the output layers of four parallel NNs are discarded as we extract the ECG features from the hidden layers, which reduces the number of weights by $16\times$. The input dimension of each NN is 160, 50, 50 and 30, respectively. The hidden layer dimension is 100 for all four NNs. The remaining fully-connected weight matrices have $(160+50+50+30) \times 100 = 29,000$ weights, which could still exceed on-chip storage capacity of wearable devices.

To reduce the power/area of NNs, we propose to sparsify weight matrices by Lasso regression. Denote the original trained dense NN weight matrix between the input layer (m neurons) and the hidden layer (n neurons) as $\mathbf{W}_{\text{ori}}$ ($m \times n$).

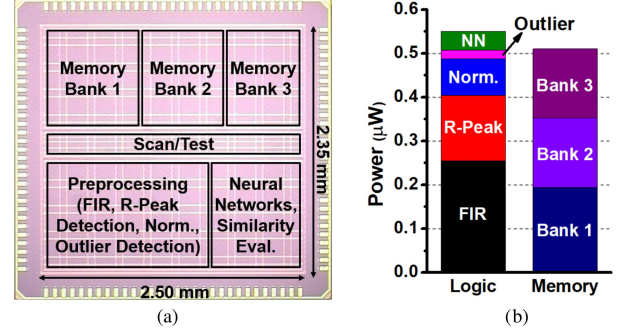


Fig. 10: (a) Chip micrograph and (b) power breakdown.

Given a sufficient number (p) of representative input samples $\mathbf{X}$ ($p \times m$) to the NNs, the weighted sum for the hidden layer will be:

$$\mathbf{Y} = \mathbf{X} \times \mathbf{W}_{\text{ori}}. \quad (4)$$

We seek a sparse weight matrix $\mathbf{W}^*$ such that

$$\mathbf{X} \times \mathbf{W}^* \approx \mathbf{Y}. \quad (5)$$

Lasso regression is a well-known algorithm [27] to find a sparse solution to (5). To sparsify the NN weights, we formulate n Lasso regression problems as follows:

$$\min_{\mathbf{w}_i^*} \|\mathbf{X} \times \mathbf{w}_i^* - \mathbf{X} \times \mathbf{w}_{\text{ori},i}\|_2 + \lambda_i \|\mathbf{w}_i^*\|_1, i = 1, \dots, n, \quad (6)$$

where $\mathbf{w}_i^*$ and $\mathbf{w}_{\text{ori},i}$ are the i -th column of $\mathbf{W}^*$ and $\mathbf{W}_{\text{ori}}$, respectively, and λ_i is the regularization parameter in each Lasso regression problem. For the convenience of hardware implementation and efficient storage of sparse weight matrices, each $\mathbf{w}_i^*$ is preferred to have the same number of non-zero weights. Typically, the larger λ_i is, the sparser $\mathbf{w}_i^*$ will be.

For each Lasso regression problem, we conduct a binary search to find the smallest λ_i for each $\mathbf{w}_i^*$ column, so that the same target sparsity is achieved for all columns. For $10\times$ weight compression, Lasso regression results in better EER of 1.70%, compared to the EER of 3.31% that is obtained when magnitude-based pruning is performed to reach the same sparsity. With $10\times$ weight reduction, we achieved $4.6\times$ memory compression (including overhead for weight indices) with only 0.2% EER degradation. Starting from the software NN model, Fig. 9 summarizes the overall NN weight reduction of $390\times$, collectively achieved by output layer removal, low-precision quantization, and Lasso-based compression.

The extracted features from four NNs are evaluated in parallel. The weighted sum of one hidden neuron is computed in one cycle. 160 -to- 16 , 50 -to- 5 , 50 -to- 5 and 30 -to- 3 multiplexers are used in the four NNs, to select corresponding partial inputs according to the sparse weight indices for each hidden neuron. In total, only 29 multipliers are required for the four NNs. After the bias is added, the weighted sum goes through a 128-entry look-up table based $\tanh(x)$ module that only stores the positive parts, exploiting the rotational symmetry of $\tanh(x)$.

IV. MEASUREMENT RESULTS

The prototype chip is implemented in 65 nm LP CMOS (Fig. 10(a)). Total on-chip memory is 19.5 kB, out of which 4.6

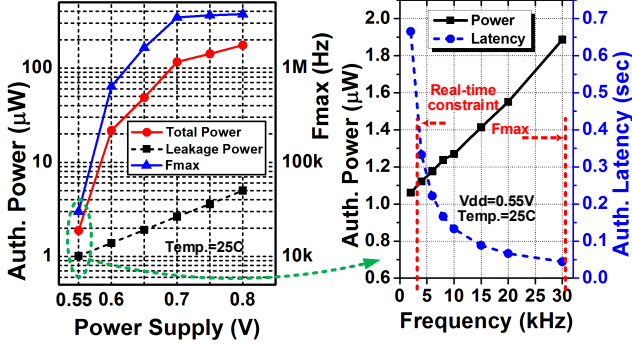


Fig. 11: Measurement results of power, frequency and latency for authentication.

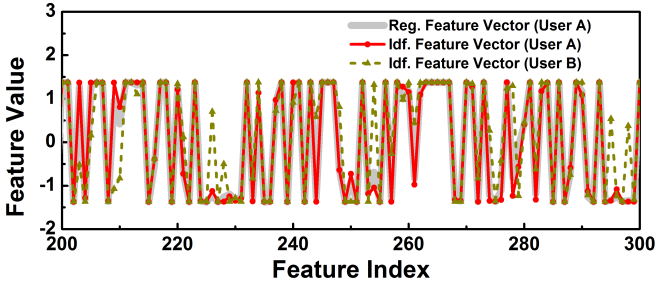


Fig. 12: Measurement results of registration/identification FVs (features 201-300 are shown) from same and different users.

kB is used for the NN weights. Measured power consumption of the ECG processor is $1.06 \mu\text{W}$ at 0.55 V supply and 2 kHz clock frequency, when it performs continuous real-time authentication. For cardiac monitoring, the ECG processor consumes $0.83 \mu\text{W}$ (at 1 kHz) for arrhythmia detection and $0.88 \mu\text{W}$ (at 3 kHz) for anomaly detection at 0.51 V supply. All measurements are performed at room temperature. Fig. 10(b) shows the power breakdown, based on module-level power percentages obtained from post-layout simulation.

A. Biometric authentication

The maximum clock frequency scales down with lower supply voltages as shown in Fig. 11 (left). Fig. 11 (right) shows the tradeoff between authentication latency and power consumption at 0.55 V as we sweep the clock frequency. The minimum clock frequency for real-time authentication is 2 kHz . The combined latency of outlier detection, NN, and similarity evaluation modules dictates that the clock frequency should not be too low; otherwise, user will need to wait for more than one second from the moment when the last ECG beat is acquired till the final authentication result comes out.

We tested the ECG processor with real-time ECG signals from a few volunteers. Fig. 12 shows the registration and identification FVs for one volunteer (user A), and the identification FV for another volunteer (user B). It can be seen that there is a very small difference between the registration and identification FVs of user A, while there is sufficient difference between the registration FV of user A and identification FV of user B. As a result, only user A is authenticated during identification when user A's ECG data is registered.

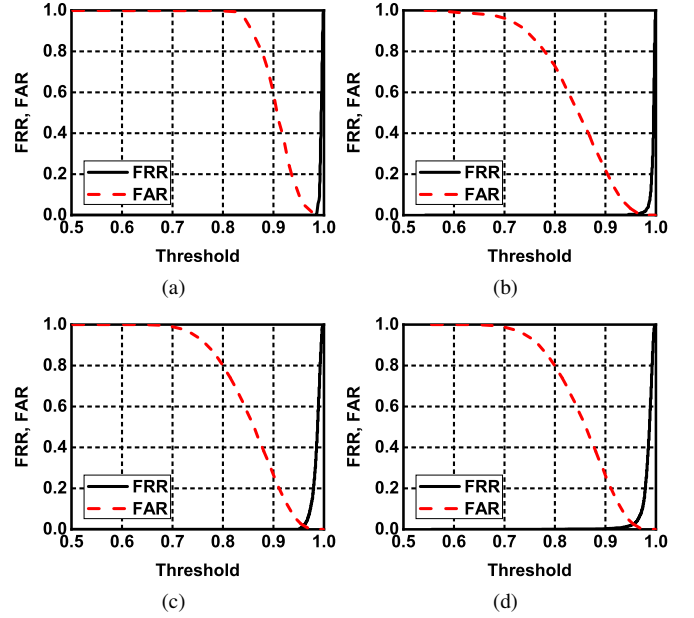


Fig. 13: Measurement results of FAR and FRR of ECG authentication for three databases and different starting points are shown: (a) MIT-BIH NSRDB with 18 subjects (EER=0.10%), (b) ECG-ID database with 90 subjects (EER=0.74%), (c) in-house ECG database with 645 subjects with authentication starting at 4 seconds (results in best EER of 1.70%), and (d) in-house ECG database with 645 subjects with authentication starting at 1 second (results in worst EER of 2.48%).

We evaluated the authentication performance on three ECG databases. Note that clinical studies typically use 12-lead ECG signals to obtain spatial information of the heart's electrical activity. However, since we focus on wearable applications, which can practically integrate only single-lead ECG sensors, only single-lead or single-channel ECG signals are used for all three databases below. "Lead I" represents one of the 12-lead ECG orientations, namely right arm (-) to left arm (+).

- MIT-BIH normal sinus rhythm database (MIT-BIH NSRDB) [19], [28] includes 18 subjects (5 men, age 26-45, and 13 women, age 20-50) and ECG data of two channels. We used the first channel in our testing.
- ECG-ID database [20] includes 90 subjects (44 men and 46 women, age 13-75) and data of one channel (lead I).
- In-house ECG database includes 645 subjects half men and half women. The ECG data has only one channel (lead I, electrodes on the wristband sense ECG signals from right index finger (-) to left wrist (+)), digitized by analog front end (AFE) chip ADS1292R [29] on a custom wristband. The subjects wearing the custom wristband sat calmly during ECG acquisition, and each ECG record is 2-minute long. The first half of the ECG records were used for NN training, and the second half of the ECG records were used for ECG authentication evaluation.

To evaluate the authentication accuracy, false acceptance rate (FAR), false rejection rate (FRR) and equal error rate (EER) are measured. FAR is the rate at which a wrong user is accepted; FRR is the rate at which a correct user is rejected; EER is the error rate when FAR and FRR are identical.

We perform registration once and identification three times for every subject. If a database has n subjects, we perform ECG authentication experiments for $n \times 3n$ times, where there are $3n$ same-user identification attempts and $3n \times (n - 1)$ different-user identification attempts. By comparing $n \times 3n$ cosine similarity values with the threshold, we obtain FRR as the ratio of the number of the same-user cosine similarities that is less than the threshold, to the total number of the same-user cosine similarities ($3n$). FAR is obtained as the ratio of the number of the different-user cosine similarities that is greater than or equal to the threshold, to the total number of the different-user cosine similarities ($3n \times (n - 1)$). By increasing the threshold, FAR decreases while FRR increases. Fig. 13 shows the FAR/FRR for a range of threshold values for MIT-BIH NSRDB, ECG-ID, and 645-subject in-house databases. The EER values are: 0.10% for MIT-BIH NSRDB, 0.74% for ECG-ID, and 1.70%/2.18%/2.48% (best/average/worst) for in-house ECG database. These measured EERs are comparable to EERs of recent fingerprint (0.8% [30]) and iris (0.82% [31]) based authentication algorithms.

Temporal variability exists in ECG signals for any individual, which poses a critical challenge for ECG authentication accuracy. To evaluate this, for the in-house ECG database, we varied the starting point of registration and identification among ~ 8 seconds in the ECG records for 645 individuals, and measured EERs for these experiments. As shown in Fig. 14, the best EER is 1.70%, the worst EER is 2.48%, and the average EER for all 9 temporal experiments is 2.18%.

B. Arrhythmia detection

We tested arrhythmia detection with ECG recordings from MIT-BIH arrhythmia database [21], which includes 48 half-hour excerpts of two-channel ambulatory ECG recordings from 25 men (age 32-89) and 22 women (age 23-89). We used the ECG data from channel 1. Fig. 15(a) shows successful arrhythmia detection for record 100 from [21]. When the PVC beat occurs, the HRV rises above the threshold (7.5 bpm) for a short period, which promptly turns on the detection signal.

In arrhythmia detection mode, the detection results are reported every heart beat based on the R-R intervals for the past 4 heart beats. Therefore, starting from the 4th ECG beat, we can compare the arrhythmia detection results with annotations made by experts in the database to evaluate our arrhythmia detection accuracy. As mentioned in Section II.C, we focus on detection of APC and PVC, which are the two

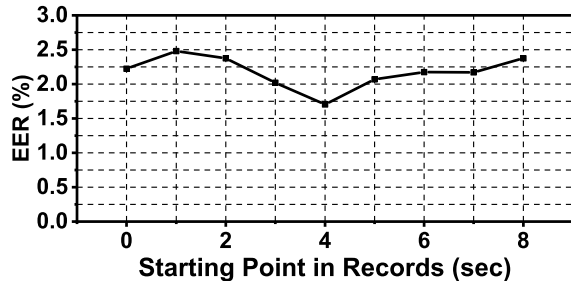


Fig. 14: EER results on 645-subject in-house ECG database with varying starting points for registration and identification.

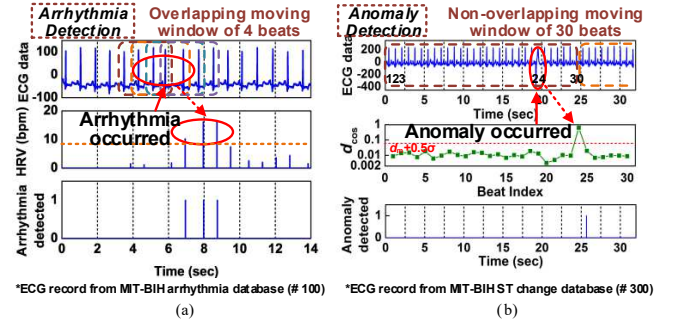


Fig. 15: Measurement results of (a) arrhythmia detection and (b) anomaly detection.

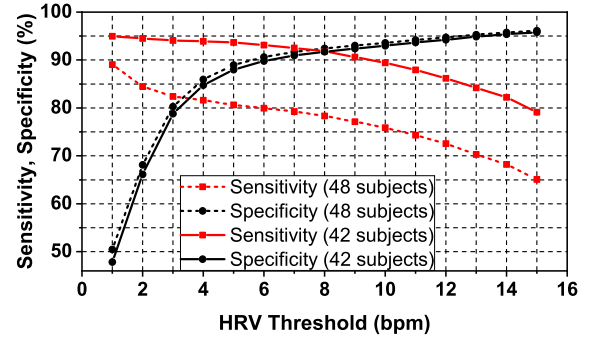


Fig. 16: Measurement results of arrhythmia detection (APC or PVC) for subjects in MIT-BIH arrhythmia database [21].

most common premature contractions [32], [33], and can be occasionally caused by heart diseases. The configurable HRV threshold can be a value between 0 and 15.9375 bpm.

Fig. 16 shows arrhythmia detection sensitivity (true positive rate) and specificity (true negative rate) measurement results for 48/42 subjects in the MIT-BIH arrhythmia database [21]. For all 48 subjects in [21], when the HRV threshold is 5 bpm, the sensitivity of APC or PVC detection is 80.65%, while the specificity is 88.98%. We noticed that 6 subjects (record 207, 208, 209, 213, 223, 232) contain hundreds of consecutive premature contraction beats that cannot be detected by our ECG processor, since HRV would be small in those cases. When these 6 subjects are excluded, the sensitivity for 42 subjects improve considerably, which shows that our ECG processor can well detect sudden HRV changes. At the HRV threshold of 6 bpm, the sensitivity is 93.13% and the specificity is 89.78%.

C. Anomaly detection

Anomaly detection is aimed at detecting abnormal ECG pulse shapes among normal ones, even if the heart rate is normal. We tested the ECG processor with ECG recordings with such abnormal pulse shapes. Fig. 15(b) shows the measurement results for an ECG recording (record 300) from MIT-BIH ST change database [34]. It can be seen that, although the heart rate is regular, the 24th ECG beat appears to be very different in the ECG pulse shape, and its cosine distance (from simulation) to the mean ECG beat among the 30 beats is much higher than others and is higher than the threshold. Also, the minimum value of this 24th is much lower than the 1.5 times the median minimum value. Therefore, this beat is detected as

TABLE I: Comparison with Prior ECG Processor Works

	[4]	[5]	[6]	This Work	[15]	[16]
Technology	180nm	65nm	90nm	65nm	Artix-7 FPGA	Cortex-M MCU
Supply Voltage	0.7 V	0.4 V	0.5-1.0 V	0.55 V	1.0 V	Unknown
Include AFE	No	Yes	Yes	No	No	No
Power (Digital)	1.26 μ W (Arr.)	45 nW (Arr.)	7-32.8 μ W (M.I.)	1.06 μ W (Auth.)	256 mW (Auth.)	Unknown
Clock Frequency	0.25-1 kHz	10 kHz	8-32 kHz	2 kHz	50 MHz	168 MHz
Memory Size	10.5 kB	3.7 kB	20.0 kB	19.5 kB	979.1 kB	Unknown
Arrhythmia Detection	Yes	Yes	Yes	Yes	No	No
Anomaly Detection	No	No	Yes	Yes	No	No
Authentication	No	No	No	Yes	Yes	Yes
Auth. EER/FAR/FRR	N/A	N/A	N/A	0.74% (ECG-ID) 1.70%-2.48% (645 subjects)	0.06% (ECG-ID)	FAR = 5.2% FRR = 1.9% (28 subjects)

Arr.: Arrhythmia, Auth.: Authentication, M.I.: myocardial infarction

an outlier and reported by the ECG processor. The anomaly detection result will be updated every 30 ECG beats.

D. Comparison with related works

Table I provides the comparison of our proposed ECG processor against other related works. AFE is not included in our processor, so only the digital power of prior works are reported in Table I. While the cardiac monitoring part compares similarly to prior works, note that anomaly detection is much more complex than arrhythmia detection.

On the authentication side, our work is the first ASIC that implements ECG-based biometric authentication. The authors of [15] achieved a very low EER on the ECG-ID database. However, they trained different NNs for different users, while we used a fixed common NN for all users across different ECG databases. In addition, our ECG processor is tested on a much larger in-house database of 645 subjects, together with temporal variability in the ECG records. Only our ECG processor supports both cardiac monitoring and authentication in a single chip with $\sim 1 \mu$ W power for real-time operation.

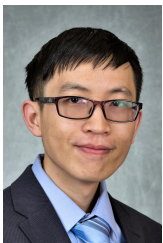
V. CONCLUSION

We presented an ultra-low-power ECG processor that performs both biometric authentication and personal cardiac monitoring. Aided by output layer removal, low precision, and Lasso-based compression, a total of $390 \times$ NN memory is reduced compared to software. We achieved $<2.5\%$ EER for a 645-subject in-house ECG database, consuming 1.06 μ W power for real-time ECG authentication. For arrhythmia detection, we achieved 93.13%/89.78% sensitivity/specificity for 42 subjects in MIT-BIH arrhythmia database. The proposed ECG processor enables secure access and cardiac monitoring in wearable devices with stringent power/area constraints.

REFERENCES

- [1] Apple, "Watch Series 3," <https://www.apple.com/apple-watch-series-3/>.
- [2] Samsung, "Gear S3," <https://www.samsung.com/global/galaxy/gear-s3/>.
- [3] K. H. Lee and N. Verma, "A low-power processor with configurable embedded machine-learning accelerators for high-order and adaptive analysis of medical-sensor signals," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 7, pp. 1625–1637, July 2013.
- [4] H. Kim, R. F. Yazicioglu, T. Torfs, P. Merken, H. J. Yoo, and C. V. Hoof, "A low power ECG signal processor for ambulatory arrhythmia monitoring system," in *IEEE Symposium on VLSI Circuits*, June 2010, pp. 19–20.
- [5] Y. P. Chen, D. Jeon, Y. Lee, Y. Kim, Z. Foo, I. Lee, N. B. Langhals, G. Kruger, H. Oral, O. Berenfeld, Z. Zhang, D. Blaauw, and D. Sylvester, "An Injectable 64 nW ECG Mixed-Signal SoC in 65 nm for Arrhythmia Monitoring," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 1, pp. 375–390, Jan 2015.
- [6] S. Y. Hsu, Y. Ho, P. Y. Chang, C. Su, and C. Y. Lee, "A 48.6-to-105.2 μ W machine learning assisted cardiac sensor SoC for mobile healthcare applications," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 4, pp. 801–811, April 2014.
- [7] iRhythm, "Zio XT Patch," <http://irhythmtech.com/zio-services.php>.
- [8] AliveCor, "KardiaBand," <https://store.alivecor.com>.
- [9] Samsung, "Simband," <https://www.simband.io>.
- [10] Token, "The Ring," <https://tokenring.com/the-ring>.
- [11] J. Lee, S. Noh, K. R. Park, and J. Kim, "Iris recognition in wearable computer," *Biometric Authentication*, vol. 3072, pp. 475–483, 2004.
- [12] J. Chauhan, H. J. Asghar, A. Mahanti, and M. A. Kaafar, "Gesture-based continuous authentication for wearable devices: The smart glasses use case," in *Applied Cryptography and Network Security*, 2016, pp. 648–665.
- [13] A. D. C. Chan, M. M. Hamdy, A. Badre, and V. Badee, "Wavelet distance measure for person identification using electrocardiograms," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 2, pp. 248–253, February 2008.
- [14] I. Odinaka, P. H. Lai, A. D. Kaplan, J. A. O'Sullivan, E. J. Sirevaag, S. D. Kristjansson, A. K. Sheffield, and J. W. Rohrbach, "ECG biometrics: A robust short-time frequency analysis," in *IEEE International Workshop on Information Forensics and Security*, December 2010, pp. 1–6.
- [15] A. Page, A. Kulkarni, and T. Mohsenin, "Utilizing deep neural nets for an embedded ECG-based biometric authentication system," in *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, October 2015, pp. 1–4.
- [16] S. J. Kang, S. Y. Lee, H. I. Cho, and H. Park, "ECG Authentication System Design Based on Signal Analysis in Mobile and Wearable Devices," *IEEE Signal Processing Letters*, vol. 23, no. 6, pp. 805–808, June 2016.
- [17] R. D. Labati, E. Munoz, V. Piuri, R. Sassi, and F. Scotti, "Deep-ECG: Convolutional Neural Networks for ECG Biometric Recognition," *Pattern Recognition Letters*, March 2018.
- [18] S. Yin, M. Kim, D. Kadetotad, Y. Liu, C. Bae, S. J. Kim, Y. Cao, and J. sun Seo, "A 1.06 μ W Smart ECG Processor in 65nm CMOS for Real-Time Biometric Authentication and Personal Cardiac Monitoring," in *IEEE Symposium on VLSI Circuits*, 2017, pp. C102–C103.
- [19] The MIT-BIH Normal Sinus Rhythm Database, doi: 10.13026/C2NK5R.
- [20] T. S. Lugovaya, "Biometric human identification based on electrocardiogram," Master's thesis, Faculty of Computing Technologies and Informatics, Electrotechnical University "LETI", Saint-Petersburg, Russian Federation, 2005.
- [21] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

- [22] Y. Liu, X. Feng, C. Zhang, C. Bae, and S.-J. Kim, "Electrocardiogram (ECG) authentication method and apparatus," US Patent Application 20170188971A1. [Online]. Available: <https://patents.google.com/patent/US20170188971A1/en>
- [23] C. Choi, Y. Kim, and K. Shin, "A PD control-based QRS detection algorithm for wearable ECG applications," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, August 2012, pp. 5638–5641.
- [24] L. Wiclaw, Y. Khoma, P. Faat, D. Sabodashko, and V. Herasymenko, "Biometric identification from raw ECG signal using deep learning techniques," in *IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, vol. 1, September 2017, pp. 129–133.
- [25] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [26] J. Bombardieri, "Systolic pipeline architectures for symmetric convolutions," *IEEE Transactions on Signal Processing*, vol. 40, no. 5, pp. 1253–1258, May 1992.
- [27] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society (Series B)*, vol. 58, pp. 267–288, 1996.
- [28] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. C. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng, and H. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, June 2000.
- [29] Texas Instruments, "ADS1292R," <http://www.ti.com/product/ADS1292R>.
- [30] K. K. M. Shreyas, S. Rajeev, K. Panetta, and S. S. Agaian, "Fingerprint authentication using geometric features," in *IEEE International Symposium on Technologies for Homeland Security (HST)*, April 2017, pp. 1–7.
- [31] A. Uka, A. Roçi, and O. Koç, "Improved segmentation algorithm and further optimization for iris recognition," in *IEEE EUROCON - 17th International Conference on Smart Technologies*, July 2017, pp. 85–88.
- [32] A. Iwasa, M. Hwa, A. Hassankhani, T. Liu, , and S. M. Narayan, "Abnormal heart rate turbulence predicts the initiation of ventricular arrhythmias," *Pacing and Clinical Electrophysiology*, vol. 28, no. 11, pp. 1189–1197, 2005.
- [33] U. Ofoma, F. He, M. L. Shaffer, G. V. Naccarelli, and D. Liao, "Premature cardiac contractions and risk of incident ischemic stroke," *Journal of the American Heart Association*, vol. 1, no. 5, October 2012.
- [34] P. Albrecht, "ST segment characterization for long term automated ecg analysis," Ph.D. dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 1983.



Shihui Yin Shihui Yin (S'15) received the B.S. degree in microelectronics from Peking University, Beijing, China, in 2013, and the M.S. degree in electrical engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 2015. Currently, he is currently working towards the Ph.D. degree in electrical engineering at Arizona State University, Tempe, AZ, USA. His research interest includes low power biomedical circuit and system design, and energy-efficient hardware design for machine learning and neuromorphic computing.

Mr. Yin was a recipient of University Graduate Fellowship from Arizona State University in 2015 and IEEE Phoenix Section Student Scholarship for the year 2016.



Minkyu Kim Minkyu Kim (S'15) received the B.E. degree in electronics and electrical engineering from Pusan National University in 2005, and the M.S. degree in electrical engineering from Pohang University of Science and Technology in 2007, all Republic of Korea. During 2007 and 2013, he was with LG Display Co. Ltd, Republic of Korea, where he worked on development of timing controller chip and image processing for high-quality display. He is currently working towards the Ph.D. degree in electrical engineering from Arizona State University.

His research interests include efficient hardware design and application of machine learning and neuromorphic algorithms.



Deepak Kadetotad Deepak Kadetotad (S'15) received the B.E. degree in electronics and communication engineering from M.S. Ramaiah Institute of Technology, Karnataka, India, in 2013. He is currently working towards the Ph.D. degree in electrical engineering from Arizona State University, Tempe, AZ, USA. His research interests include hardware design and application of machine learning and neuromorphic algorithms.

Mr. Kadetotad was the recipient of the LSI Chairman's International scholarship from 2009 to 2013.



Yang Liu Yang Liu received the B.S. degree from University of Science and Technology of China (USTC), in 2009, and the Ph.D. degree from Institute of Automation, Chinese Academy of Sciences, in 2014, in pattern recognition and intelligent systems. Since 2014 she has been a researcher with the Samsung Advanced Institute of Technology, China. Her research interests include computer vision, deep learning, and intelligent systems.



Chisung Bae Chisung Bae received the B.S. degree from Pohang University of Science and Technology (POSTECH), South Korea, in 2004, and the M.S. and Ph.D. degrees from Korea Advanced Institute of Science and Technology (KAIST), South Korea, in 2006 and 2010, respectively, all in electrical engineering. Since 2010 he has been a principal researcher with the Samsung Advanced Institute of Technology, South Korea. His research interests include ultra-low-power SoC design for emerging embedded systems, bio-medical signal processing

systems, and wireless communication systems.



Sang Joon Kim Sang Joon Kim received the Bachelors degree in electrical engineering from Seoul National University, Seoul, Korea (ROK), in 2001, and the Masters degree in Applied Mathematics and the Ph.D. in Engineering Science from School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, in 2007 and 2009, respectively.

Since 2009, he has been with Samsung Advanced Institute of Technologies as a principle researcher. His current research interests include bio signal processing and untral low power system architecture from information theory perspective.



Yu Cao (S'99-M'02-SM'09-F'17) received the B.S. degree in physics from Peking University in 1996. He received the M.A. degree in biophysics and the Ph.D. degree in electrical engineering from University of California, Berkeley, in 1999 and 2002, respectively.

He worked as a summer intern at Hewlett-Packard Labs, Palo Alto, CA in 2000, and at IBM Microelectronics Division, East Fishkill, NY, in 2001. After working as a post-doctoral researcher at the Berkeley Wireless Research Center (BWRC), he

is now a Professor of Electrical Engineering at Arizona State University, Tempe, Arizona. He has published numerous articles and two books on nano-CMOS modeling and physical design. His research interests include physical modeling of nanoscale technologies, design solutions for variability and reliability, reliable integration of post-silicon technologies, and hardware design for on-chip learning.

Dr. Cao was a recipient of the 2012 Best Paper Award at IEEE Computer Society Annual Symposium on VLSI, the 2010, 2012, 2013, 2015 and 2016 Top 5% Teaching Award, Schools of Engineering, Arizona State University, 2009 ACM SIGDA Outstanding New Faculty Award, 2009 Promotion and Tenure Faculty Exemplar, Arizona State University, 2009 Distinguished Lecturer of IEEE Circuits and Systems Society, 2008 Chunhui Award for outstanding oversea Chinese scholars, the 2007 Best Paper Award at International Symposium on Low Power Electronics and Design, the 2006 NSF CAREER Award, the 2006 and 2007 IBM Faculty Award, the 2004 Best Paper Award at International Symposium on Quality Electronic Design, and the 2000 Beatrice Winner Award at International Solid-State Circuits Conference. He has served as Associate Editor of the IEEE Transactions on CAD, and on the technical program committee of many conferences.



Jae-sun Seo (S'04-M'10-SM'17) received the B.S. degree in electrical engineering from Seoul National University, Seoul, South Korea, in 2001, and the M.S. and Ph.D. degrees in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 2006 and 2010, respectively.

From 2010 to 2013, he was with IBM T. J. Watson Research Center, Yorktown Heights, NY, USA, where he worked on cognitive computing chips under the DARPA SyNAPSE Project and energy-efficient integrated circuits for high-performance

processors. In 2014, he joined the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, USA, as an Assistant Professor. In 2015, he was with the Intel Circuits Research Lab as a Visiting Faculty. His current research interests include efficient hardware design of machine learning and neuromorphic algorithms and integrated power management.

Dr. Seo was a recipient of the Samsung Scholarship during 2004–2009, the IBM Outstanding Technical Achievement Award in 2012, and the NSF CAREER Award in 2017.