Adaptive R-Peak Detection on Wearable ECG Sensors for High-Intensity Exercise

Elisabetta De Giovanni, Tomas Teijeiro, Grégoire P. Millet, and David Atienza, *Fellow, IEEE*

Abstract*—*Objective: **Continuous monitoring of biosignals via wearable sensors has quickly expanded in the medical and wellness fields. At rest, automatic detection of vital parameters is generally accurate. However, in conditions such as high-intensity exercise, sudden physiological changes occur to the signals, compromising the robustness of standard algorithms.** ***Methods:*** **Our method, called BayeSlope, is based on unsupervised learning, Bayesian filtering, and non-linear normalization to enhance and correctly detect the R peaks according to their expected positions in the ECG. Furthermore, as BayeSlope is computationally heavy and can drain the device battery quickly, we propose an online design that adapts its robustness to sudden physiological changes, and its complexity to the heterogeneous resources of modern embedded platforms. This method combines BayeSlope with a lightweight algorithm, executed in cores with different capabilities, to reduce the energy consumption while preserving the accuracy.** ***Results:*** **BayeSlope achieves an F1 score of 99.3% in experiments during intense cycling exercise with 20 subjects. Additionally, the online adaptive process achieves an F1 score of 99% across five different exercise intensities, with a total energy consumption of 1.55±0.54 mJ.** ***Conclusion:*** **We propose a highly accurate and robust method, and a complete energy-efficient implementation in a modern ultra-low-power embedded platform to improve R peak detection in challenging conditions, such as during high-intensity exercise.** ***Significance:*** **The experiments show that BayeSlope outperforms state-of-the-art QRS detectors up to 8.4% in F1 score, while our online adaptive method can reach energy savings up to 38.7% on modern heterogeneous wearable platforms.**

***Index Terms*—Adaptive R peak detection, algorithm robustness, biosignal processing, edge computing, energy-accuracy trade-off, heterogeneous nodes, intense exercise, machine learning, ultra-low power computing, wellness.**

Manuscript received 1 July 2022; accepted 6 September 2022. Date of publication 9 September 2022; date of current version 20 February 2023. This work was supported in part by the Swiss NSF MLEdge Project under Grant 200020_182009, in part by the MyPreHealth Project funded by Hasler Stiftung under Grant 16073, and in part by the H2020 DeepHealth Project under Grant GA 825111. The work of Tomas Teijeiro was supported by Maria Zambrano fellowship under Grant MAZAM21/29 from the University of Basque Country and the Spanish Ministry of Universities funded by the European Union-NextGenerationEU. *(Corresponding author: David Atienza.)*

Elisabetta De Giovanni is with the Embedded Systems Laboratory, EPFL, Switzerland.

Tomas Teijeiro is with the Department of Mathematics, University of the Basque Country (UPV/EHU), Spain.

Grégoire P. Millet is with the Institute of Sport Sciences, University of Lausanne, Switzerland.

David Atienza is with the Embedded Systems Laboratory, EPFL, Switzerland (e-mail: david.atienza@epfl.ch).

This article has supplementary downloadable material available at https://doi.org/10.1109/TBME.2022.3205304, provided by the authors.

Digital Object Identifier 10.1109/TBME.2022.3205304

I. Introduction

In RECENT years, increasing healthcare costs [1] and hospital overcrowding have pushed new technological advances to improve remote wellness monitoring, and enable early intervention and prevention [2]. In addition, population aging and the resulting higher incidence of noncommunicable diseases (NCDs) create the need for long-term health monitoring. For these reasons, there is an increasing demand for applications working on wearable platforms that continuously and remotely monitor biosignals, such as electrocardiogram (ECG) [3] or photoplethysmography (PPG), and extract relevant health parameters from them. Furthermore, daily physical activity is highly recommended [4] to prevent NCDs, and in particular high-intensity interval training (HIIT) are postulated as a good alternative to moderate intensity for health improvement [5].

As we illustrate in Section II, during intense physical exercise sudden physiological changes occur, such as short RR intervals, high breathing frequency and noise from respiratory sinus arrhythmia, or sympathetic activation, amongst others [6]. These changes induce artifacts or noise that are in general not properly detected by standard algorithms, leading to a need for using advanced computing techniques, such as machine learning and online adaptivity, to improve the robustness of the analysis. However, with the advancement of new complex algorithms comes the question of managing constrained resources in wearable sensor nodes (WSNs), and the consequent toll on energy consumption.

In fact, the implementation of complex biomedical applications in traditional WSNs can cause a significant draining of platform resources leading to frequent device charging [7]. Furthermore, various algorithmic optimizations implemented to lower the device energy consumption can lead to a decrease in the output accuracy of the algorithm [8]. With the advent of modern ultra-low power (ULP) platforms [9] and their capabilities, the trade-off between optimizing the device resources to lower the energy consumption and maintaining a highly accurate output has become more attainable [10], [11]. Nevertheless, in the context of complex biomedical applications for WSN-based wellness monitoring, the designer has to consider new challenges to achieve an optimal energy-accuracy trade-off. First,

in the acquired biosignals of various pathologies or physical conditions sudden events occur, which traditional algorithms can miss or misinterpret (e.g., atrial fibrillation (AF) or intense physical exercise) [6], [12]. For this reason, their robustness is compromised and severely affects the reliability of the wellness progress in the long term. Second, the static nature of traditional algorithms do not include the adaptive management of platform resources at run time according to the complexity of the application, which has recently become a need in WSNs design. New methods tackle self-aware applications at the algorithmic level applying a multi-layer classification or detection system with increasing complexity [13], [14]. Based on the confidence of the low complexity classifiers, the algorithm decides whether to execute a more complex layer and, therefore, consumes more energy. However, these methods are targeted to traditional homogeneous platforms, and some do not consider the error in the pathological events detection.

For the reasons above, in this work, we propose an online adaptive design of a new ECG R peak detection algorithm for wearable systems based on unsupervised machine learning, which exploits the capabilities and heterogeneity of modern ULP platforms. In the proposed design, we introduce for the first time BayeSlope, a slope-based R peak detector that applies a Bayesian filter, non-linear normalization, and a clustering technique to an ECG segment. In the literature, the use of slope-based QRS detectors has been extensive [15], [16]. There are examples of the use of the Kalman filter for smoothed estimation of the heart rate (HR), different than R peak detection, and using multiple signals [17]. However, many of these works target ambulatory monitoring. Hence, to the best of our knowledge, this is the first time that an R peak detection like BayeSlope is applied in the context of intense physical exercise. In fact, we test the proposed method with a dataset collected in collaboration with the Institut des sciences du sport de l'Université de Lausanne (ISSUL), where the subjects performed a maximal exercise test on a cycle ergometer till exhaustion. Our main contributions are:

- We propose a new highly accurate slope-based R peak detection method that is specifically aimed at high intensity exercise application scenarios. Our new R peak detection method, called BayeSlope, applies a Bayesian filter and a non-linear normalization to the input ECG signal. The combination of these signal processing techniques enhances and correctly detects the next R peak in the expected position on a peak-to-peak resolution. During high intensity exercise, BayeSlope outperforms the most popular state-of-the-art QRS detectors up to 8.4% in F_1 score, while being comparable during low intensity exercise.
- We pair the newly proposed algorithm with the REWARD algorithm, presented in [18], which is less complex though more prone to error if sudden events occur. To ensure the adaptive nature of the design, we propose an error detection routine applied to REWARD that triggers BayeSlope if REWARD fails.
- To apply adaptive management of resources at run time according to the algorithm's complexity, we implement the proposed method on an heterogeneous platform, that allows to run BayeSlope on a more capable core than the one where REWARD runs, which is simpler. In fact, the R peak detection step of REWARD is approximately $104\times$ less complex than BayeSlope when running on the same core. Hence, a simpler processor can handle it better, while a more powerful core handles better the more complex BayeSlope.
- The fully adaptive process has an F_1 score of up to 99.0%, comparable to always running BayeSlope, which achieves an F_1 score up to 99.3%, across five different exercise intensities. Moreover, our proposed adaptive process is up to 17.5% more accurate compared to running only REWARD, across the five exercise intensities. Finally, the adaptive method tailored for modern heterogeneous platforms can reach energy savings up to 38.7% compared to continuously executing BayeSlope. Therefore, the newly proposed adaptive design is the best solution for an optimal energy-accuracy trade-off for long-term wellness monitoring with latest wearable systems.

In Section II, we describe what occurs during intense physical exercise and the relevance of a highly accurate R peak detection in such conditions. In Section III, we present the new R peak detection algorithm and its adaptive design. In Section IV, we describe the protocol of the experiments and the platform used. Finally, in Sections V and VI, we present, respectively, the results and the conclusion of our analysis.

II. Background

To motivate the newly proposed method for autonomous wellness monitoring, let us consider the sudden changes occurring in the ECG during intense physical exercise. We will focus specifically on the R peak, as it is the basis for ECG analysis. Fig. 1 shows two segments of ECG acquired from a subject performing a maximal exercise test (c.f. Section IV). The segments were extracted from the initial rest condition (Fig. 1a) and a window of intense physical exercise, close to exhaustion (Fig. 1b). As shown in Fig. 1b, the peak-to-peak (RR) interval variability is significantly low compared to a rest condition. Moreover, the amplitude of the R peaks is highly variable. Therefore, when standard algorithms are used to detect the peaks in these conditions, their robustness is compromised. In this work, we consider one R peak detection algorithm that was presented in [18], called REWARD, as the standard base algorithm to build on and motivate our proposed online adaptive method. REWARD can detect peaks within a window of $1.75\,\mathrm{s}$ by adapting hysteresis thresholds based on the average and maximum (or minimum if R is negative) amplitude of the peaks within a window. This method works well when the amplitude variability is limited, as shown in Section V-A. However, during intense physical exercise, the RR interval decreases significantly and the amplitude highly varies between one peak and the other—within $1.75\,\mathrm{s}$, there are many peaks significantly different in amplitude—with the result that REWARD fails to detect smaller peaks. In Fig. 3, we show an example of an ECG segment extracted from the analyzed dataset and, specifically, a window where REWARD fails.

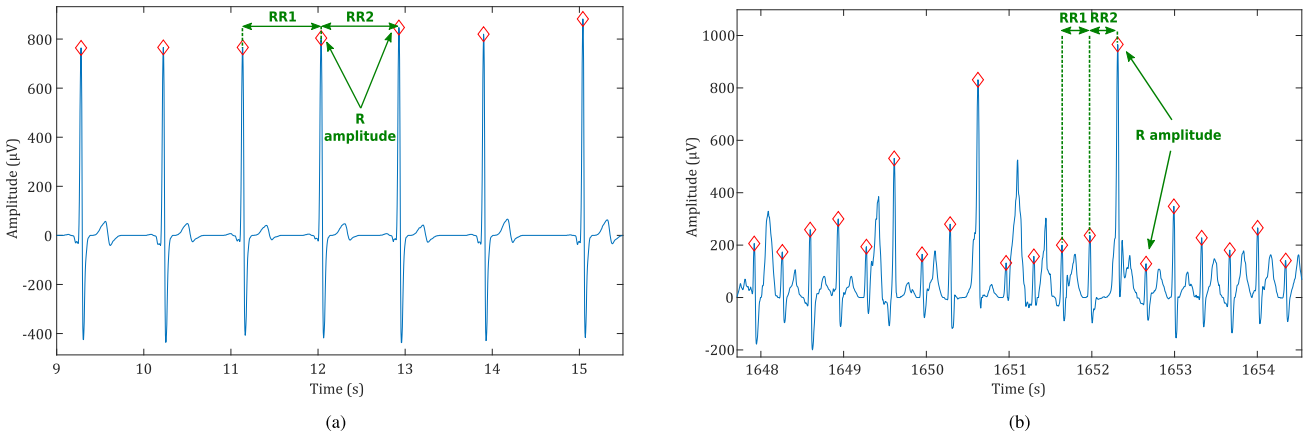

Fig. 1. Effects of intense physical exercise on ECG, and, specifically, the R peak amplitudes and RR interval variability, compared to rest. The ECG segments are extracted from the database presented in Section IV. Specifically, the segment in Fig. 1a is extracted from the first 3 min of rest of the maximal exercise test of Subject 3, starting at second nine. The segment in Fig. 1b is extracted close to exhaustion of Subject 3 during the maximal exercise test, starting at approximately 27 min. (a) Rest. (b) Intense physical exercise.

To capture the changes occurring during various intensities of physical exercise in the wellness context, there exists a gold standard protocol where subjects perform a maximal exercise test on a cycle ergometer or a treadmill till exhaustion. The subjects wear a gas mask that measures the volume of O_2 and CO_2 (VO_2, VCO_2) inhaled and exhaled [19]. Additionally, the protocol includes the acquisition and analysis of a single-lead ECG, from which specific heart rate variability (HRV) parameters can be extracted to help in the estimation of the so-called ventilatory thresholds (VT1, VT2) [20], and VO_2 max [21]. These three variables describe the cardiovascular and respiratory state during intense physical exercise. VT1 measures the hyperpnea (i.e., faster breathing) caused by the increased production of CO_2 for exercise intensities above the anaerobic threshold resulting in a non-linear increase in the ratio between ventilatory flow (VE) and VO_2. VT2 represents a phase where the hyperpnea is not enough to eliminate the CO_2, which remains constant, leading to a sharp increase of VE/VCO_2. Finally, VO_2 max is the final stage where exhaustion is reached and, consequently, a maximum oxygen uptake and HR. The determination of the ventilatory thresholds usually relies on an agreement between medical experts who evaluate the gas analysis and the HRV parameters to find the position of the thresholds [22].

The HRV analysis uses the RR time series of an ECG signal to extract time and frequency domain features, which can be used for a direct estimation of VT1, VT2 and VO_2 max [20]. The current methods for this estimation are performed in post-processing with the help of medical experts and, usually, require interpolation and correction of the RR time series. The R peak detection needs to be accurate, robust and adapt at run time to sudden changes to ensure the correct comparison between ventilatory measurements and the RR time series. Therefore, we propose BayeSlope, a new highly accurate and robust R peak detection algorithm for wearable sensors, which is paired to REWARD. BayeSlope is much more complex than REWARD, and, if run continuously, it can drain the device battery. For this reason, we additionally propose a real-time strategy that automatically adapts the algorithm's complexity and the resources assigned based on the robustness of REWARD, for an enhanced energy-accuracy trade-off in latest ultra-low power autonomous wearable systems.

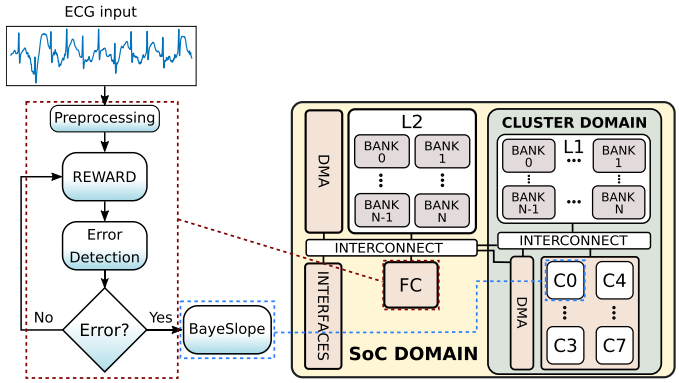

Fig. 2. On the left, data-flow diagram of the adaptive R peak detection algorithm with a raw ECG input. REWARD refers to the low complexity R peak detection presented in [18]. BayeSlope is a new slope-based R peak detection algorithm presented in this work. On the right, the PULP-based [9] architecture used for the analysis. Preprocessing, REWARD, and error detection run on the SoC domain in the fabric controller (FC), while BayeSlope runs on the cluster domain, in one core of the cluster (CL) of eight cores.

III. Adaptive R Peak Detection in Modern Wearable Sensors

One of the main problems in the context of edge computing in WSNs is minimizing energy consumption while maximizing output accuracy. In this section, we describe our proposed method to detect R peaks from a single-lead ECG that optimizes the energy-accuracy trade-off with a two-level adaptive method. Fig. 2 shows the data-flow diagram of the full process and the architecture where the algorithm is implemented [9]. The two-level adaptivity consists in the robustness and complexity of the two different R peak detection algorithms, namely REWARD [18] and BayeSlope. As described in Section II, REWARD uses hysteresis thresholds that are adapted to each

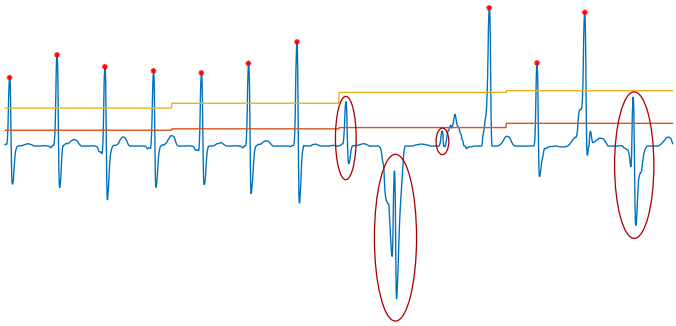

Fig. 3. Missed peaks (in ellipses) in REWARD R peak detection using hysteresis thresholds (in orange and yellow) based on the ECG window morphology. The segment was extracted from Subject 3 of the used dataset (c.f., Section IV-A).

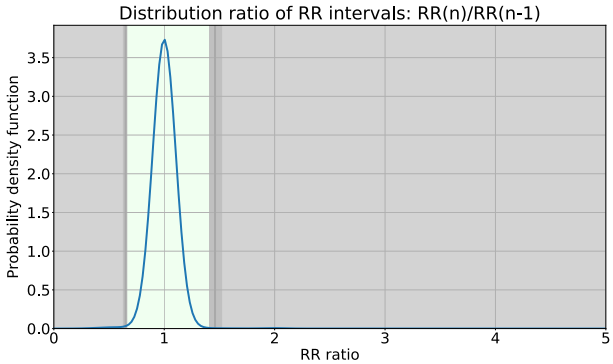

Fig. 4. RR ratio distribution for the full dataset acquired for this analysis. The shadowed grey areas represent the range of RR ratio for which an error is detected, while the light green area represents the range of RR ratio for which there is no error in the R peaks. Moreover, the darker grey areas and the vertical lines represent the full range of percentile thresholds for all the subjects in the used dataset (c.f., Section IV-A).

window of 1.75 s. However, this window resolution is too small to capture peak-to-peak sudden changes. For this reason, we introduce BayeSlope that implements peak normalization through a generalized logistic function and a Bayesian filter to enhance and compute the expected position of the next R peak. This expectation is the basic guide for BayeSlope, and is based on the assumption of local stability in the RR interval. Therefore, if there are relevant arrhythmias in the ECG beyond occassional ectopic beats, BayeSlope will offer no advantage over classical slope-based methods. Once the peaks are normalized and enhanced, the method applies a k-means clustering to divide the ECG samples in two clusters represented by two centroids, one corresponding to the baseline and one to the R peaks. The first level of adaptivity consists in adapting the robustness of the algorithm, as shown in the data-flow of Fig. 2. The output of REWARD is fed to an error detection method that checks when REWARD is probably failing to detect R peaks and, in this case, triggers the more accurate and robust BayeSlope. The second level of adaptivity consists in adapting the resources to the complexity of the method. In fact, BayeSlope is a more complex algorithm, hence, it benefits from execution on the cluster of cores in the platform, which includes eight cores with higher Instructions-Per-Cycle (IPC) and floating point units. Secondly, the main core, which is a simpler core, can handle the less complex REWARD in a more energy-efficient way. In the next sections, we describe first the different blocks shown in Fig. 2. Then, the higher-level design within the heterogeneous platform is presented.

A. Preprocessing, REWARD and Error Detection

A standard R peak detection algorithm requires several steps of preprocessing of the ECG input signal. In this case, the input is a single-lead ECG where a morphological filtering (MF) is applied to remove the baseline and high frequency noise [23]. Then, the signal is enhanced by applying the Relative-Energy (Rel-En) method, which amplifies the most dominant peaks [18]. This preprocessing method is part of the REWARD algorithm presented in [18]. The second part of the algorithm searches for the R peak in a window of 1.75 s using hysteresis thresholds based on the ECG morphology within the window. However, during intense physical exercise, the interval between two R peaks (i.e., RR interval) decreases significantly and sudden changes in amplitude occur. Therefore, within a window of analysis, many peaks can be missed, as shown in Fig. 3. Moreover, right after exhaustion during a maximal exercise test, there can be an increase of the T wave amplitude—the wave after the QRS complex that represents the repolarization of the heart ventricles—often significantly more dominant than the R peak itself, and a decrease of the RT interval. In these conditions, REWARD fails in detecting very small peaks as the hysteresis thresholds are skewed by the higher amplitude variability of the peaks within the window. However, it performs extremely well if these events do not occur as demonstrated in [18].

For this reason, we propose a statistical method to identify potential errors in the R peak detection within a window of 1.75 s by analyzing the distribution of the ratio $\frac{RR(n)}{RR(n-1)}$, where $n = 0, 1, 2\ldots$, of all the data acquired. This distribution has been computed offline using the results of BayeSlope, since it is the most accurate (c.f., Section V). However, to avoid data snooping, for each subject, the RR ratio distribution is computed with a leave-one-out (LOO) strategy, in which the analyzed subject is not included in the distribution. The RR ratio can capture sudden changes with a three-peak resolution, such as missing peaks, additional wrong peaks (e.g., T wave), and highly noisy signal segments. We want to underline that the main objective of the error detection mechanism is to balance the execution of REWARD and BayeSlope to optimize the energy consumption. Thus, this mechanism has a marginal influence on the R peak detection accuracy. There is only one situation, thoroughly explained in Section V, in which a mistake in the error detection can lead to misdetected peaks.

First, the method computes offline the RR intervals and the corresponding RR ratio sequence used for the distribution from all the subjects, except the one that is being analyzed. Then, for each subject, if at least one value of RR ratio computed within each window falls in the tails of the distribution (below the 0.5 or above the 99.5 percentiles of the RR ratio distribution, respectively), the algorithm detects an error. This is performed in the online phase of the error detection applied to the output

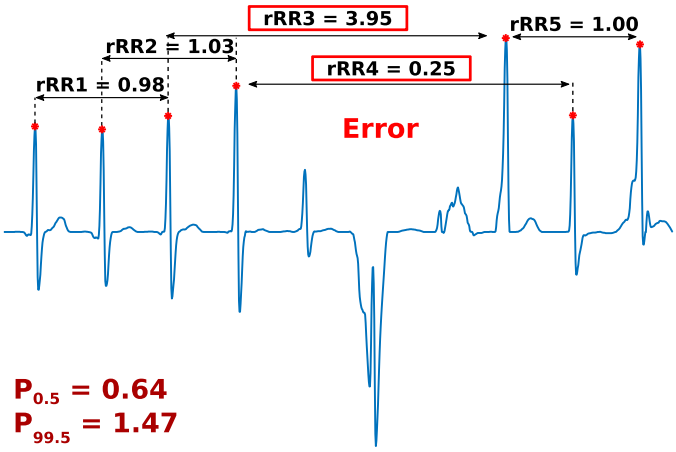

Fig. 5. Result of error detection on example ECG extracted from Subject 3 of the dataset used (c.f., Section IV-A). The values of the RR ratio are computed on a resolution of three R peaks. Considering the percentile thresholds for the analyzed subject (bottom left), the method can detect an error where REWARD fails (in the red boxes).

R peaks of REWARD. Fig. 4 shows the distribution considering all the subjects analyzed in this work. We report the overall distribution for convenience, but note that it is not the one used in our proposed online error detection. Moreover, the right tail is longer than reported on the figure, as it is redundant. In fact, the percentile thresholds with the LOO strategy are:

$$P_{0.5} = 0.65 \pm 0.02; P_{99.5} = 1.46 \pm 0.06; \quad (1)$$

Even though the distributions are different, the standard deviation among the subjects is quite small, suggesting that the RR ratio has a low inter-patient variability. On the other hand, the range of the distribution suggests a high intra-patient variability. Considering these values of thresholds, Fig. 4 shows shadowed areas in grey, which represent the range of RR ratio for which an error is detected. Moreover, this figure shows a light green area that represents the range of RR ratio for which no error in the R peaks exists. Finally, the darker grey areas and the vertical lines represent the full range of percentile thresholds reported in (1). Therefore, if we consider the ECG example in Fig. 3, the error detection results are shown in Fig. 5, where the values of the RR ratio over the segment are reported. Considering the percentile thresholds $P_{0.5} = 0.64$ and $P_{99.5} = 1.47$ for the analyzed subject, the method can detect an error where REWARD fails. The last peak in Fig. 3 where there is an error would be detected in the next window. It is worth mentioning that there are genuine physiological events, such as ectopic heartbeats, that can cause a variation in the RR ratio in the tails of the distribution, and therefore they will be detected as "errors". These events will lead to trigger the BayeSlope algorithm, but it does not mean that they will be ignored or undetected, as long as they fit the detection conditions of BayeSlope.

B. BayeSlope: Adaptive Slope-Based R Peak Detection

Once an error is detected, a more accurate adaptive R peak detection, BayeSlope, is triggered. This newly proposed method

Algorithm 1: BayeSlope R Peak Detection.

1: **Input**: windows of RelEn signal, s (μV)
2: **Output**: R peaks, r
3: Initialize centroids: $hcentr$ = percentile(diff(s), 99) and $lcentr = 1$
4: $min\_rr\_dist = 240$ ms; $max\_qrs\_dur = 140$ ms; ▷ Constant parameters
5: Initialize: $mu = 75$ bpm; $sd = 100$ ms; $zeroctr = 0$; $qrs\_init = 0$; $label = 0$; in_qrs = false;
6: **for** $i = 2, \ldots$.length(s)**do**
7: $s2[i] = s[i] - s[i-1]$; ▷ Derivative approximation
8: x = abs($s2[i]$);
9: $bf[i]$ = gaussian($i - last\_peak, mu, sd$); ▷ Bayesian filter
10: $bt[i]$ = genlogfun($x, param\_logfun$); ▷ Sigmoid normalization
11: $st[i] = \max(x, bt[i] * bf[i])$; ▷ Normalize signal
12: Update $hcentr$ and $lcentr$ each as the new mean of their cluster
13: **if** in_qrs **then** ▷ Peak search
14: **if** $label = 0$ **then**
15: $zeroctr+ = 1$;
16: **else**
17: $zeroctr = 0$;
18: **end if**
19: **if** $zeroctr = 30$ OR $i - qrs\_init > max\_qrs\_dur$**then**
20: max_min_slope = argmaxmin($st * sign(s2)$);
21: Search for new_peak within max_min_slope
22: $r[i] = new\_peak$;
23: **end if**
24: **else**
25: **if** $label = 1$ AND $i > last\_peak + min\_rr\_dist$**then**
26: in_qrs = true;
27: $qrs\_init = i$;
28: **end if**
29: **end if**
30: **end for**

applies a non-linear normalization of the signal and a Bayesian filter to enhance high slope areas which are near the expected position of the next peak according to the current HR. These high slope areas are assumed to belong to the QRS complex, and to distinguish them from low slope areas, the approach relies on a clustering method based on K-means.

Algorithm 1 describes the main steps of BayeSlope. The method takes as input the Rel-En signal window, s, and it outputs the vector of R peaks detected. The algorithm is derivative-based, and considers two clusters that represent the high and low slope areas of the signal. Then, the two centroids are initialized beforehand, as shown in Line 3, where $hcentr$ is the 99^{th} percentile of the derivative of s and $lcentr = 1$. When a new sample is assigned to a cluster it is labeled as 1 if it is colser to $hcentr$, or 0 if it is closer to $lcentr$. Two windows of 1.75 s

are used for the $hcentr$ initialization to account for enough peaks even at rest and avoid errors due to signal noise. Then, the algorithm initializes all the other parameters needed, constant and varying, in Lines 4–5. The values for these parameters were chosen based on common physiological constraints [24], and not on any values observed in the data. Thus, we can be confident that the values are general enough for any ECG obtained from an adult human.

The main process starts by considering the derivative of s and computing its absolute value, x, in Lines 7–8. We apply this initial transformation to enhance the maximum and minimum slopes of the original signal s. This choice follows the assumption that the R peak is located in general within the maximum upward and downward deflections within an ECG signal. Next, the method computes the Bayesian filter (Line 9), which is a Gaussian centered on the expected peak, mu, with standard deviation sd, two parameters computed based on the last five peaks. The selection of such a small number of peaks is to make the Bayesian filter responsive to the potential fast variations of the RR interval near high intensity limits [25], which are precisely the regions of greatest interest to our algorithm. Then, in Line 10, the algorithm computes the generalized logistic function [26] with input x and its parameters computed based on the last $hcentr$ and $lcentr$. The sigmoid varies between 0 and the value of the higher k-means centroid, $hcentr$. The sigmoid and the Bayesian filter are used to normalize the peak or, specifically, to increase the amplitude of expected small peaks, as shown in Line 11 and Fig. 6. If the analyzed sample does not reach the computed threshold, the function does not increase its value. When the input is approximately double the value of the lowest centroid, st (in Algorithm 1) reaches the threshold between $lcentr$ and $hcentr$, which determines the value of $label$. In Fig. 6, the expected location of the peak (i.e., the prior expectation) is depicted with the Gaussian centered on it. In this case, the original peak (i.e., observed value) in x is small, and it is enhanced by the Gaussian multiplied by the sigmoid function, leading to the posterior estimation st. This situation is shown in the posterior estimation rectangle.

Once the signal is normalized, the algorithm starts a peak search within a QRS complex—the ECG main wave—in Lines 13–29. This procedure is also illustrated in Fig. 6. The distance between QRS complexes must be more than min_rr_dist, according to standard physiological characteristics and the sample that starts the QRS complex (qrs_init in Line 25) must belong to the cluster represented by $hcentr$ (i.e., $label = 1$). Within the QRS complex, the algorithm waits till it reaches its maximum duration according to physiology (max_qrs_dur) or for enough samples ($zeroctr = 30$) labeled 0 that represent the end of the QRS complex (Lines 14–19). Once within this interval (Lines 20–22), the algorithm computes the maximum and minimum of the function $st * sign(s2)$ representing the maximum upslope and downslope of the original signal. The sign function is used in case these values fall in the Q, S or T wave, which are not distinguished if only st is used, as it is positive by definition. Finally, the new_peak is found and stored in the vector r.

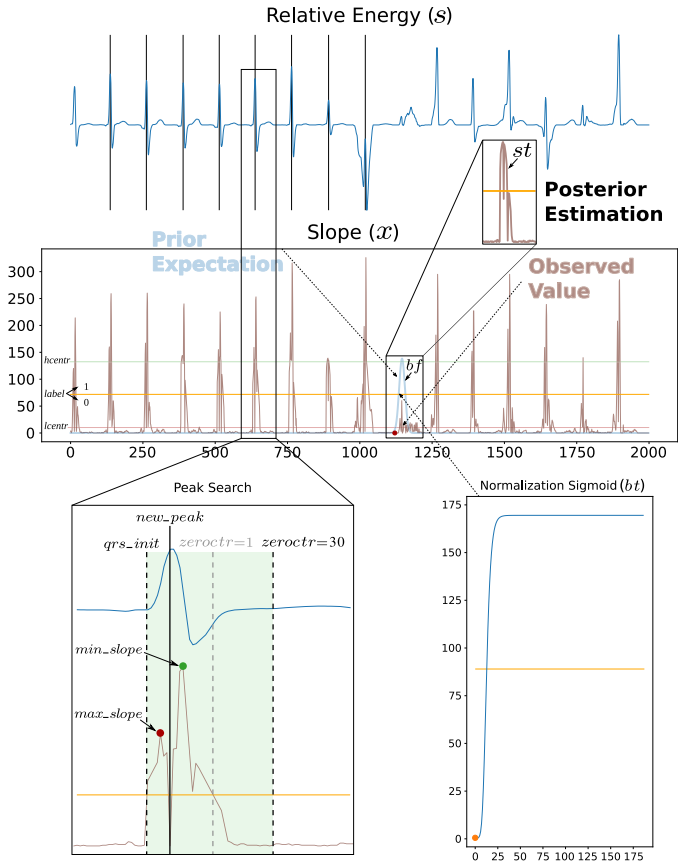

Fig. 6. Illustration of the main features of the BayeSlope algorithm. The bayesian filter (i.e., prior estimation), together with the generalized logistic function (normalization sigmoid) allows to locally enhance the observed slope value (x), leading to a posterior estimation that goes above the detection threshold. The bottom-left part of the figure illustrates the peak search procedure that is triggered once the detection threshold is surpassed. All symbol names correspond to those introduced in Algorithm 1.

C. Adaptive Design in Modern Heterogeneous Platforms

As shown in Fig. 2, the modules of the algorithm run in different cores of the wearable computing architecture according to the complexity of the corresponding module. The wearable architecture used in this work is based on one of the evolutions of the open-source PULP platform [27], called Mr.Wolf [9]. The PULP structure consists of a main streamlined processor, the fabric controller (FC), and an 8-core parallel compute cluster (CL). Moreover, PULP includes a direct memory access (DMA) that can transfer data to a multi-banked 512 KiB L2 memory during acquisition time or from L2 to a shared multi-banked 64 KiB L1 memory, which has a single-cycle latency in the cluster side. Both FC and CL are power-gated while the DMA fills the required L2 memory bank during sample acquisition. The FC is clock-gated when the CL is active, and each of the cores in the CL can be independently clock-gated to reduce dynamic power. Mr.Wolf includes a core for the FC (Zero-riscy) that is simpler than the RI5CY cores of the CL, but it has a lower IPC. On the other hand, the cores of the CL have more

capabilities [28]. Therefore, this work considers the Mr.Wolf architecture by selectively using the FC and one core of the CL.

Considering this design, the modules of preprocessing (MF), REWARD (which includes Rel-En and R peak detection via hysteresis thresholds), and error detection run in the FC. REWARD is a very lightweight integer-based algorithm, as demonstrated in [18]. In a preliminary analysis, considering the dataset (c.f., Section IV-A), we performed a test executing the R peak detection step of REWARD on the FC and on one core of CL. The algorithm executed on CL is $1.23\times$ slower (in terms of execution time) and consumes $1.35\times$ more energy. Therefore, REWARD benefits from running on the FC, which is a simpler core, clocked at a higher frequency (170 MHz vs. 110 MHz of the CL). On the contrary, BayeSlope is about $100\times$ more complex than REWARD, hence, it benefits from running on a more advanced core with higher IPC. This helps to meet real-time constraints and limits the amount of time the system is active. Additionally, the Gaussian and the generalized logistic function of BayeSlope are implemented in floating-point. Since the FC does not have a floating-point unit, BayeSlope should be converted to fixed-point representation. Therefore, we adapted these functions of BayeSlope to employ fixed-point arithmetic, using a 32-bit representation with 1 sign bit, 15 integer bits, and 16 decimal bits. The results reveal that during the clustering step the algorithm quickly reaches the maximum range representable (i.e., approximately within 15 s of signal processing), with a consequent drop in accuracy. In contrast, this does not occur in the 32-bit floating-point representation as the maximum range is reached after approximately 27 h of signal processing. Therefore, we decided to implement BayeSlope on one core of the CL (RI5CY), which has a floating point unit and higher IPC.

After the signal filtering and REWARD running on the FC, the error detection (also running on the FC) checks the accuracy of the R peaks output. If an error is detected, the DMA transfers the necessary buffer of data from L2 to L1 ready for the core in the CL. Conversely, the FC is clock-gated. Since BayeSlope needs an initialization of the R peaks of two windows of 1.75 s, the previous window error needs to be checked. If the error in the previous window is 0, then the DMA transfers two windows, otherwise it transfers only one. This is an optimization applied in case REWARD fails more frequently and to avoid recomputing the same window. BayeSlope runs on the CL while the FC is clock-gated. The final output is the combination of correct R peaks from REWARD and BayeSlope. The full code for the adaptive R peak detection has been published as open-source software[1].

IV. Experimental Setup

A. Database Acquisition Protocol

The database was acquired considering 22 subjects performing an incremental test to exhaustion on a cycle ergometer for an average of 30 minutes each until VO_2 max was reached, plus at least 1 minute post-exercise. The power of the cycle ergometer was increased every 3 min by 30 W, after initial 3 min of rest.

[1]https://c4science.ch/source/adaptive_rpeak_det_public

Fig. 7. Sketch of the BIOPAC [29] sensors positioning (left) during the experiment and the sensors themselves (right).

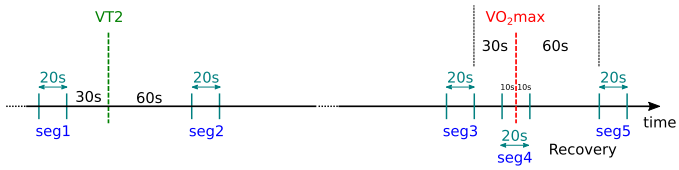

Fig. 8. Position in time through the maximal exercise test of the five 20-second segments extracted from the full ECG of each subject. The numbers of the segments are sorted in time in ascending order. The first segment was extracted 30 s before VT2 and corresponds to heavy intensity; the second 60 s after VT2 (severe intensity); the third 30 s before VO_2 max (highly severe intensity up to exhaustion); the fourth at the moment of exhaustion (centered in VO_2 max); the fifth 60 s post-exercise, i.e. during the recovery after exhaustion.

Moreover, a three-minute recovery period was recorded right after exhaustion. A single-lead ECG sampled at 500 Hz was acquired using the BIOPAC system [29], together with other biosignals and oxygen uptake measurements that were not used for this work. Fig. 7 shows a sketch of the positioning of biosignals sensors and the equipment used. The protocol was ethically approved by the Commission Cantonale (VD) d'Ethique de la Recherche sur l'Etre Humain (CER-VD), with reference 2016-00308, on 01/03/2018. For the experiments, the ECG was downsampled to 250 Hz since REWARD was validated only for this frequency in [18]. Two of the 22 subjects were discarded because one did not complete the protocol and for the second one the majority of the recording was corrupted. Therefore, the statistics and analysis were performed on 20 subjects. Next, five 20-second segments were extracted from the full ECG of each subject to be manually annotated by experts. The 20-second duration has been selected as a minimum to enable posterior HRV analysis [20]. The annotations were initially made by an engineer with background on ECG analysis, and then individually assessed by a cardiologist. A consensus was achieved among both annotators after just one iteration. The annotation process was done on PDF files with the standard ECG grid of $0.2s \times 0.5\,mV$, as shown in Fig. 10. The segments were chosen based on the different phases of the maximal exercise test, namely, considering higher intensities of exercises where it is more likely that sudden changes occur. Then, these segments were extracted and reported in the order shown in Fig. 8: the first, 30 s before VT2; the second, 60 s after VT2; the third, 30 s before VO_2 max (exhaustion); the fourth, at the moment of exhaustion (centered in VO_2 max); finally, the fifth, 60 s after VO_2 max, i.e. during the recovery after exhaustion. The segments at rest were ignored since REWARD performs very well in this condition, and there is no need to

run BayeSlope. Moreover, also the segments near VT1 are not considered as they represent lower intensities of exercise for which the performance of REWARD is satisfactory. Only one out of 100 segments was not extracted and annotated (subject 9, segment during recovery). In fact, the recording for this subject was stopped right after exhaustion (instead of after three minutes of recovery expected by the protocol) and it was not possible to have a 20-second segment $60\,\mathrm{s}$ after VO_2 max. The input segments to the peak detection were extracted considering the $20\,\mathrm{s}$ given to the experts and going backward of $0.6\,\mathrm{s} + 0.95\,\mathrm{s} + 1.75\,\mathrm{s}$, which represents, respectively, the initial delay of the MF, the initial delay of Rel-En, and one additional window of analysis for BayeSlope initialization; and going forward $1.75\,\mathrm{s}$ to avoid missing the last peaks. Thus, each segment is approximately $25\,\mathrm{s}$ long. The accuracy of R peak detection is measured according to the standard tolerance of $150\,\mathrm{ms}$ between the detected peaks and the manual annotations [30]. We also report for each subject the mean and standard deviation of the time difference between the two. We first compare the accuracy of BayeSlope against a broad spectrum of state-of-the-art methods, including:

1) The Pan-Tompkins algorithm [31], which is the most widely used QRS detector in the literature. It is based on a combination of bandpass filtering and differentiation.
2) A variation of the Engelse-Zeelenberg (EngZee) method, which is based on a heavy difference filter and which is considered particularly robust to noise and artifacts [32].
3) The GQRS detector [33], which belongs to the matched-filter family of methods.
4) A more modern method based on adaptive thresholding on the Stationary Wavelet Transform (SWT) of the ECG [34].

We believe these four methods give a fair and complete overview of the different families of approaches that are typically used in embedded implementations. All the experiments were done with open-source implementations of the algorithms (specifically, [33] for GQRS and [35] for the other methods). Then, we also compare the accuracy of our algorithms in the following three designs:

a) Preprocessing (MF) and always running REWARD (Rel-En + peak detection);
b) Preprocessing (MF + Rel-En) and always running the newly proposed BayeSlope;
c) Our proposed adaptive design including preprocessing (MF), REWARD (Rel-En + peak detection), error detection and running BayeSlope only when REWARD fails.

All the segments used in the experiments, as well as the manual annotations, have been published as an open dataset [36].

B. Test Benches on the Heterogeneous Platform

The three designs are mapped on the Mr.Wolf platform to estimate their overall energy consumption and perform the energy-accuracy analysis. In all test benches, the preprocessing always runs on the FC. The first two test benches consist of 1) REWARD running on the FC with the CL power-gated, and 2) BayeSlope always running on the CL. The third test bench consists of the fully adaptive process, including the error detection, with REWARD running on the FC and BayeSlope running on CL when REWARD fails. Each of the test benches is applied to the 99 segments described in Section IV-A.

To measure the execution time of the three configurations, we used the open PULP platform [27]. PULP provides an SDK to run RTL simulations, using Modelsim, in order to obtain a cycle-accurate profiling. To estimate the energy consumption of our proposed system, we use the power numbers reported for a chip based on the PULP architecture implemented in TSMC $40\,\mathrm{nm}$ LP CMOS technology, namely, Mr.Wolf [9]. We consider the lowest energy point of the platform, at $0.8\,\mathrm{V}$. The platform requires $3.6\,\mu\mathrm{W}$ [37] when power-gated[2] and $12.6\,\mu\mathrm{W}$ with full L2 retention. To implement better memory management of the activated banks as done in [38], we reduce the L2 to 128 KiB, with a resolution of 16 KiB per memory bank, since the application does not need more memory. When the System-on-Chip (SoC) architecture of PULP is active, it consumes $0.98\,\mathrm{mW}$ with its main processor clock-gated, and $6.66\,\mathrm{mW}$ while operating at $170\,\mathrm{MHz}$. Once the CL is activated, it consumes $0.61\,\mathrm{mW}$ with all the cores clock-gated and $18.87\,\mathrm{mW}$ with the eight cores running at $110\,\mathrm{MHz}$.

The three designs are compared first in terms of accuracy, then energy consumption of their mapping on Mr.Wolf, and then in their energy-accuracy trade-off for all the subjects and as a summary for worst, average and best cases.

V. Experimental Results

A. Accuracy Analysis of the Test Benches

In Fig. 9, we report the percent of the error rate (ErrRate%) in the peak detection of the three designs, described in Section IV-A, and its evolution through the type of segments for three example subjects. These examples illustrate three cases within the worst, best, and average groups in terms of accuracy of the new algorithm, BayeSlope, and the fully adaptive design (REWARD+Error detection (ErrDet)+BayeSlope) compared to REWARD. ErrRate% is computed as (1-F_1)*100, where F_1 score measures the peak detection performance as:

$$F_1 = \frac{TP}{TP + \frac{1}{2} \cdot (FP + FN)} \tag{2}$$

where TP is the set of the correctly detected peaks that match the manual annotations. FN represents all the misdetected peaks by the algorithm. FP is the set of all the peaks in the algorithm that do not match any manual annotation. The different segments shown in Fig. 9 represent increasing exercise intensities till the recovery after exhaustion (segment 5), as described in Section IV. In Fig. 9a, Subject 7 has one of the worst error rates for the new algorithm, and the reason is that segment 3 is quite noisy. The quality of the segment is shown in Fig. 10, where the amplitude of the ECG has a high variability due to changes caused by the exercise intensities near exhaustion (segment 3 is before VO_2 max). However, BayeSlope and its adaptive design, with an F_1

[2]As reported for GAP-8 [37], which is an industrial version of PULP with state-of-the-art deep sleep optimizations not yet included in Mr.Wolf, its academic counterpart.

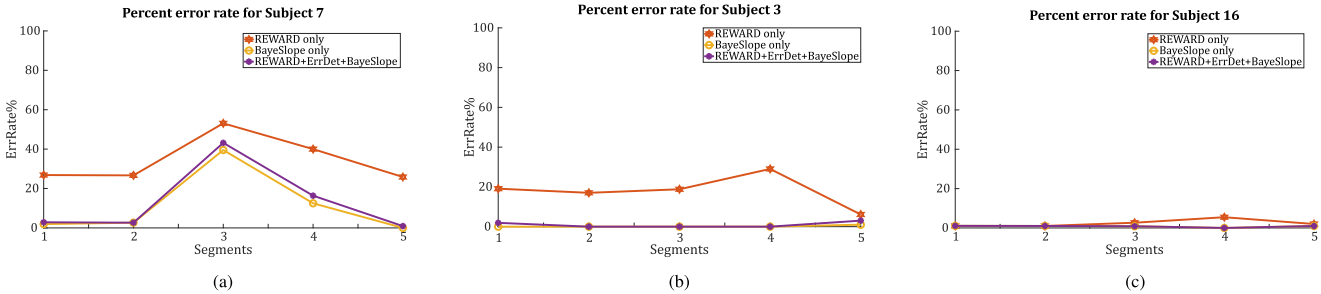

Fig. 9. Percent error rate with respect to the manual annotations of the three designs described in Section IV-A, for the worst, average, and best case subjects along five segments of increasing exercise intensities. (a) Worst case. (b) Average case. (c) Best case.

TABLE I

F_1 SCORE, PPV, SENSITIVITY (%) FOR THE THREE TEST BENCHES AND THE FIVE EXERCISE INTENSITIES COMPUTED ACROSS THE SUBJECTS

		Before VT2	After VT2	Before VO_2max	VO_2max	Recovery	**Total**
F_1 (%)	Pan-Tompkins	96.4	95.8	86.9	85.4	97.0	92.2
	EngZee	92.9	94.5	84.0	82.9	92.8	89.3
	GQRS	99.3	99.1	89.5	91.3	98.6	95.4
	SWT	96.8	97.3	89.2	88.5	97.2	93.7
	REWARD (RW)	92.1	90.9	78.7	80.2	92.5	86.7
	BayeSlope (BS)	**99.0**	**99.1**	**97.9**	**98.8**	**99.3**	**98.8**
	RW + ErrDet + BS	98.9	99.0	96.2	97.1	98.5	97.9
PPV (%)	Pan-Tompkins	99.7	99.8	98.3	98.1	99.4	99.1
	EngZee	100.0	99.6	99.4	99.5	100.0	99.7
	GQRS	99.6	99.5	99.4	99.7	100.0	99.6
	SWT	99.8	100.0	99.8	100.0	99.8	99.9
	REWARD (RW)	98.2	98.2	97.1	96.3	98.1	97.6
	BayeSlope (BS)	**98.6**	**98.6**	**98.9**	**98.6**	**98.6**	**98.7**
	RW + ErrDet + BS	98.3	98.4	97.3	96.2	97.5	97.5
Sensitivity (%)	Pan-Tompkins	93.3	92.2	77.9	75.6	94.7	86.1
	EngZee	86.8	89.9	72.8	71.0	86.6	80.9
	GQRS	99.0	98.7	81.4	84.2	97.2	91.6
	SWT	94.0	94.8	80.6	79.4	94.8	88.2
	REWARD (RW)	86.8	84.5	66.1	68.7	87.4	78.0
	BayeSlope (BS)	99.3	99.5	**96.9**	**98.9**	**100.0**	**98.9**
	RW + ErrDet + BS	**99.4**	**99.6**	95.2	98.0	99.6	98.3
Time (ms) from manual annotation	Pan-Tompkins	91.9±26.8	91.6±28.3	92.0±31.4	91.7±30.0	87.7±30.3	91.0±29.4
	EngZee	0.1±2.9	0.9±9.9	3.9±21.4	8.8±33.2	2.0±16.5	3.0±19.6
	GQRS	32.5±10.7	30.9±11.3	25.4±21.8	27.1±22.5	31.1±16.3	29.4±17.4
	SWT	60.6±18.6	61.4±18.8	63.4±24.2	65.1±23.0	59.3±18.5	61.9±20.8
	REWARD (RW)	0.6±8.4	1.1±11.2	10.4±35.7	20.6±48.6	9.2±34.8	8.1±32.0
	BayeSlope (BS)	**0.5±6.5**	**0.3±4.6**	4.9±24.0	**6.8±29.1**	**1.0±10.3**	2.9±18.6
	RW + ErrDet + BS	**0.5±6.5**	**0.3±4.6**	**4.3±22.3**	7.3±30.1	1.1±11.3	**2.8±18.5**

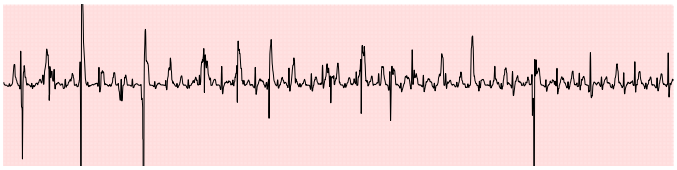

Fig. 10. ECG segment 3 (i.e., before VO $_2$ max) for Subject 7. The amplitude of the peaks is highly variable due to the changes in the exercise intensity. The signal is shown on a standard ECG sheet with a grid of $0.2\,\mathrm{s} \cdot 0.5\,\mathrm{mV}$.

score at approximately 60.5% and 56.8%, respectively, gains within 13.5% and 9.9% in performance, compared to REWARD. In Fig. 9b, Subject 3 represents an average case where REWARD has a lower error rate compared to the worst case (Subject 7), though still significant. In fact, the adaptive design performs significantly better, with an error rate up to 3%, slightly worse than BayeSlope. In Fig. 9c, Subject 16 is one of the best cases where REWARD fails only during more intense exercise (at exhaustion), with an error rate up to 5.5%, while BayeSlope has an error rate of only 1%.

Considering the five exercise intensities, a relevant summary of the algorithms' performance is depicted in Table I. Here, we report the F_1 score, sensitivity, and positive predictive value (PPV) of the four reference algorithms and the three test benches for each of the five types of segment computed across the subjects, as well as the mean and standard deviation of the time difference between each test bench output and the manual annotations. Our results show that BayeSlope is the most accurate of the three designs over all the performance parameters. In particular, in comparison with state-of-the-art algorithms, only GQRS achieves a comparable F_1 score during lower intensity exercises, while BayeSlope is superior to all the compared methods in the rest of the situations. We can see that the main benefit of our proposal comes from an increased sensitivity during intense physical exercise (before and after VO $_2$ max), which is precisely

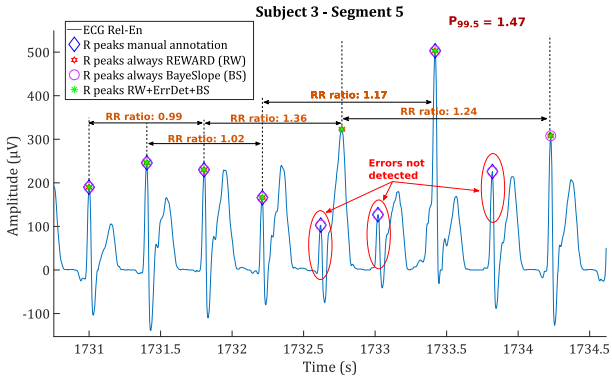

Fig. 11. ECG segment 5 (i.e., recovery after exhaustion) for Subject 3 with the R peaks from the three designs. For the misdetected small R peaks, the RR ratio is close to $P_{99.5}$ but not enough to trigger an error.

the primary flaw observed in the other methods. This supports our starting hypothesis, according to which general algorithms are not suited to handle the artifacts and sudden changes in the ECG arising from intense exercise. Indeed, the F_1 score and the sensitivity of REWARD during intense exercise are below acceptable medical standards, compared to less intense exercise. However, combining both methods in an adaptive design is as accurate as BayeSlope (up to 1.7% of difference in F_1 score).

Rarely, the adaptive design could perform better (less than 1% difference in score) as it is shown in the sensitivity values. This is due to the initialization process of BayeSlope, which requires the signal to be stable as it does not use any prior information within this initial stage. Therefore, it happens rarely that the signal is more stable later in the segment where BayeSlope is triggered and will be initialized, compared to the initialization at the beginning of the segment (when always running BayeSlope). This can also cause a delay in the adaptation and very few peaks missed and result instead in a slightly worse accuracy. Another reason for a lower performance in the adaptive design compared to always running BayeSlope, specifically for more intense exercise (before and after VO $_2$ max) and during recovery, as shown in Table I, is due to an edge case in the error detection. In fact, the RR ratio distribution used to compute the tail thresholds is performed on the full dataset and accounting for different exercise intensities. Within more intense exercises, as the RR intervals get smaller, it can happen that even if REWARD misses one peak, the RR ratio is still within the distribution. This is shown in Fig. 11, where the RR ratio computed on the small peaks not detected by REWARD is close to the $P_{99.5}$ of the distribution but not enough to trigger an error. This results in a lower accuracy for the adaptive design. One way to fix this problem is to compute different distributions for different exercise intensities. In the case of this dataset, it could be five distributions or two groups of low and high intensities. Another way is to adapt the distribution online by detecting the intensity type and choose the correct tail thresholds. Finally, with the more advanced capabilities of modern heterogeneous platforms, the distribution can be computed directly on the signal acquired through a small training process on BayeSlope and then adapting the tail thresholds.

In conclusion, the accuracy results show that always running BayeSlope is the most accurate and robust of the three designs. At the same time, REWARD's performance highly varies with the intensity of the exercise. However, BayeSlope is approximately $100\times$ more complex than the R peak detection step of REWARD. Therefore, we propose the adaptive design that combines both algorithms and has a similar accuracy compared to BayeSlope. In the next section, we will show the advantages in terms of energy consumption of the adaptive design on the PULP platform.

B. Energy Consumption of Test Benches in PULP

Fig. 12 shows the energy consumption of the platform for the three subjects described in Section V-A. In Subject 7 (Fig. 12a), the worst case scenario, the fully adaptive design consumes the same amount of energy in almost all the windows. In segment 3, the adaptive design achieves 6.5% of energy savings compared to always running BayeSlope, with a 3.7% difference in F_1 score. However, the overall accuracy is far from the required medical standard, and even the best-performing state-of-the-art algorithms only reach 75%. In Subject 3 (Fig. 12b), for all the exercise intensities except segment 4, during exhaustion, the fully adaptive wearable design we propose has energy savings up to 48% compared to BayeSlope with a loss in accuracy of only up to 2% (c.f. Fig. 9b). For segment 3, even if the energy savings are one of the lowest at approximately 3.3%, the fully adaptive design is as accurate as BayeSlope and 18.8% more accurate compared to REWARD. Therefore, on average cases such as Subject 3, in most exercise intensities, choosing the fully adaptive design can improve the energy-accuracy trade-off. Subject 16, representing one of the best case scenarios in Fig. 12c, highlights the adaptivity of the full design and its error detection through the segments, starting with a minimum energy consumption, since only REWARD is running, and maximum attainable accuracy. Then, when the exercise intensity increases, REWARD fails more frequently, and BayeSlope takes over the R peak detection. Finally, during recovery, when the ECG stabilizes and REWARD fails less compared to exhaustion, the energy consumption drops to a lower level. Our fully adaptive design maintains a high level of accuracy (approximately 99%), while limiting the energy consumption compared to executing BayeSlope for the full segment, with energy savings from 31.8% up to 58.6%.

Fig. 13 shows how many times BayeSlope runs in the adaptive design in terms of percentage of windows over the full segment for the three cases analyzed. Considering the windows where an error occurs and triggers BayeSlope, the previous window also counts as triggered since BayeSlope needs an additional window for the initialization process (c.f., Section III-C). As expected, the trend is similar to the energy reduction compared to always running BayeSlope shown in Fig. 12. The large differences between the three subjects show how the proposed design can adapt to the subject and different exercise intensities to reduce energy consumption instead of constantly falling in the worst case scenario. This personalized and adaptive reduction

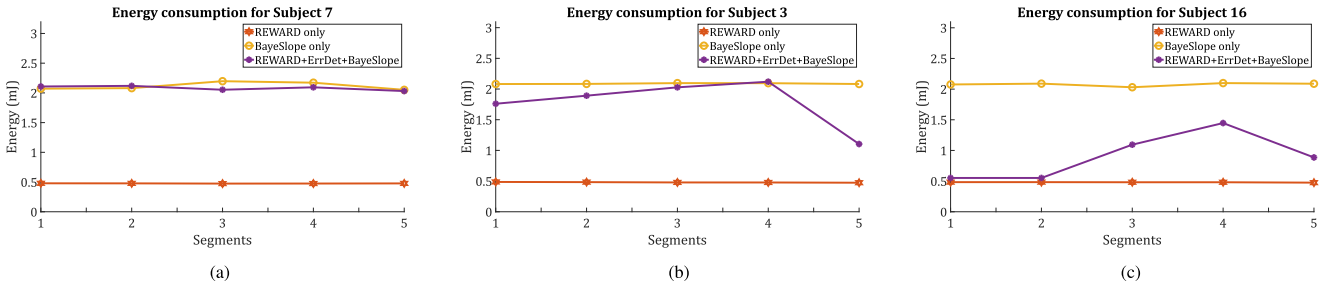

Fig. 12. Energy consumption of the test benches described in Section IV-B for the worst, average, and best case subjects along five segments corresponding to increasing exercise intensities. (a) Worst case. (b) Average case. (c) Best case.

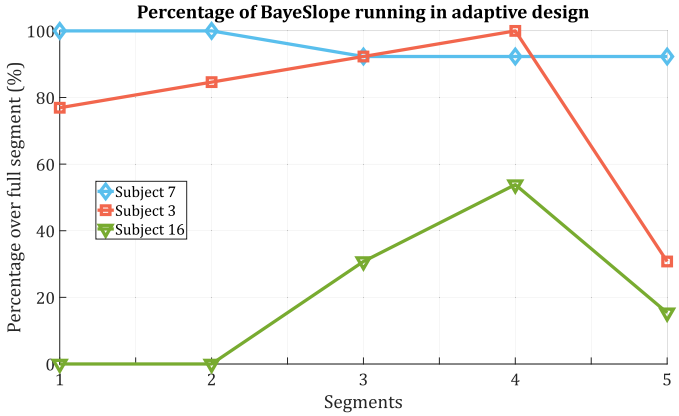

Fig. 13. Percentage of windows over the full segment where BayeSlope is triggered during the adaptive design for three worst, average, and best case scenarios. Comparing these trends with the ones shown in Fig. 12, it is evident that the adaptive design reduces energy consumption by reducing the number of times BayeSlope runs on the CL.

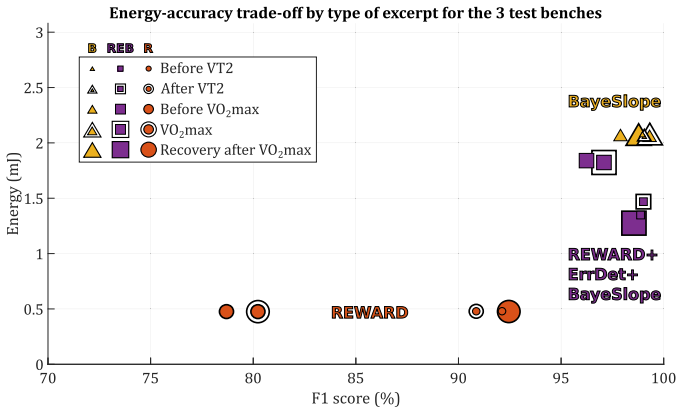

Fig. 14. Energy-accuracy analysis of the three test benches and different exercise intensities.

TABLE II
ENERGY CONSUMPTION IN MJ FOR THE THREE TEST BENCHES AND THE FIVE EXERCISE INTENSITIES COMPUTED ACROSS THE SUBJECTS

		REWARD (RW)	BayeSlope (BS)	RW + ErrDet + BS
Energy (mJ)	Before VT2	0.479±0.004	2.078±0.016	1.348±0.573
	After VT2	0.479±0.003	2.070±0.032	1.469±0.556
	Before VO_2max	0.476±0.004	2.071±0.037	1.840±0.299
	VO_2max	0.476±0.003	2.075±0.032	1.820±0.409
	Recovery	0.477±0.002	2.080±0.020	1.275±0.562
	Total	**0.477±0.004**	**2.075±0.028**	**1.553±0.536**

in energy consumption can lead to a longer battery lifetime for WSNs and better usability.

Table II shows a summary of the average energy consumption for the five exercise intensities. As shown before in our accuracy analysis, higher exercise intensities require to run BayeSlope more often in the adaptive design. However, the algorithm achieves significant energy savings compared to always running BayeSlope. The reason is that as long as BayeSlope is not triggered, the adaptive design uses only the FC. In that situation, the power of the platform corresponds to the power of the FC and the leakage power of the CL, which is significantly lower (approximately $5\times$) than the power of the CL executing BayeSlope on one of its cores, with the other ones clock-gated. Therefore, the energy consumption over the 25-second segment is reduced. As a result, the adaptive design achieves energy savings up to 38.7%, considering the average for the five exercise intensities. Moreover, it reaches up to 74.2% energy savings for the overall dataset analyzed, compared to the scenario where the CL is always active and executes BayeSlope.

C. Energy-Accuracy Trade-Off on Test Benches

Fig. 14 shows the energy-accuracy comparison between the three test benches and an analysis on the different exercise intensities. We use once again the F_1 score as a measure of algorithm detection accuracy. For the three segments before and after VT2, and during the recovery after VO $_2$ max, REWARD is accurate within the medical acceptability, and consumes the minimum energy for this application. However, the fully adaptive design (in purple) is always more advantageous in terms of accuracy, with a performance increase of up to 8.2%. Moreover, it is comparable in F_1 score to BayeSlope although more energy-efficient, with energy savings up to 38.7%.

However, when the exercise intensity increases, the number of peaks within a window increases as well. In this condition, the hysteresis thresholds of REWARD do not adapt to the high

amplitude variability of the peaks within a window of analysis (1.75 s), as described in Section III-A and Fig. 3. In fact, before VO_2 max the exercise intensity is about to reach its maximum, and more sudden changes in the ECG occur, which explains the decreased accuracy of REWARD. The segment extracted during exhaustion (i.e., when reaching VO_2 max) represents the highest intensity and, hence, disruption of the ECG morphology, specifically in the amplitude of the R peak and the RR intervals (HRV reaches its minimum). Therefore, it is the reason for a decreased performance in REWARD. On the contrary, the F_1 score of the fully adaptive design is only up to 1.7% lower than BayeSlope, which is the most accurate. The energy savings for these two segments are lower than the other three, though still significant (up to 12.2%).

Our experimental results show how the proposed BayeSlope algorithm is highly accurate and more robust than the lightweight REWARD when sudden changes in the ECG morphology occur. Moreover, in these conditions, BayeSlope and, consequently, the adaptive design are also more robust than state-of-the-art methods, such as the GQRS or SWT detectors. However, if we consider the design where BayeSlope is mapped on a PULP-based platform and running on the CL (with the preprocessing modules running on the FC), the device consumes on average 4.6 × more than the mapping of REWARD (and the preprocessing) in the FC. In contrast, the adaptive design enhances the energy-accuracy trade-off, maximizing accuracy while limiting energy consumption on modern ULP platforms. This adaptive design is not limited to applications where intense physical exercise is involved, but it can also be applied to pathologies where the ECG morphology changes. Moreover, if BayeSlope is parallelized in the 8-core CL, more computing resources can be assigned to HRV analysis and pathology detection for fully on-node processing to ensure low-rate transmission and data privacy according to the latest remote monitoring healthcare requirements.

VI. Conclusion

In health and wellness monitoring, specifically on the cardiovascular context using wearable systems, there exist multiple pathologies and physical conditions where sudden changes in the measured biosignals occur. In particular, during intense physical exercise, sudden changes in the ECG heart beats amplitude and rhythm cause errors in state-of-the-art standard R peak detection algorithms and, therefore, on any further analysis based on the HR. Moreover, more accurate algorithms often require a higher amount of computing resources leading to a need for more capable wearable platforms with flexible resource management approaches.

In this work, we have proposed a new online machine learning-based design to detect R peaks in a single-lead ECG signal, which adapts at run time to the changes in its morphology. Furthermore, this adaptive design exploits the core heterogeneity of modern ULP wearable platforms, which can run efficiently more complex algorithms using different types of cores. Our new online adaptive design uses a standard lightweight algorithm, REWARD, and an error detection method to measure the algorithm's accuracy. When REWARD fails, a novel algorithm called BayeSlope, which focuses on robustness to sudden variations in the signal properties though more complex, is triggered and runs in a more capable core. In the context of a maximal exercise test, and, in particular, during high intensity exercise, our proposed BayeSlope outperforms state-of-the-art standard algorithms. Similarly, our online adaptive design achieves a high F_1 score, up to 99.0% across five different exercise intensities, which is comparable to always running BayeSlope, and up to 17.5% more accurate compared to running only REWARD. By implementing the newly proposed adaptive method in the heterogeneous PULP SoC wearable architecture, it can reach energy savings up to 38.7% compared to always running the more complex BayeSlope. Therefore, the newly proposed online adaptive design maximizes the accuracy while minimizing the energy consumption for an optimal energy-accuracy trade-off when used in latest SoC architectures of wearable systems.

Acknowledgment

The authors acknowledge the help of Dr. Fabio Montagna for the initial implementation of BayeSlope in the PULP platform, and Dr. Miguel Péon-Quirós for the help in the design of the experiments in PULP and the insightful comments on the paper. Additionally, the authors acknowledge Dr. Nicolas Bourdillon and Leandre Tschanz for the help in the preparation of the ethical protocol and the collection of the data.

References

[1] K. Xu et al., *Public Spending on Health: A Closer Look at Global Trends*. Geneva, Switzerland: WHO, 2018.
[2] M. Al-Khafajiy et al., "Remote health monitoring of elderly through wearable sensors," *Multimedia Tools Appl.*, vol. 78, pp. 24681–24706, 2019.
[3] A. Kumar, R. Komaragiri, and M. Kumar, "From pacemaker to wearable: Techniques for ECG detection systems," *J. Med. Syst.*, vol. 42, Feb. 2018, Art. no. 34.
[4] V. Ilkka, A. B. Lars, and C. Nick, "Physical activity policies for cardiovascular health," *Eur. Heart Netw.*, Dec. 2019. [Online]. Available: http://www.ehnheart.org/publications-and-papers/publications/1243:physical-activity-policies-for-cardiovascular-health.html
[5] E. Stamatakis et al., "Untapping the health enhancing potential of vigorous intermittent lifestyle physical activity (VILPA): Rationale, scoping review, and a 4-Pillar research framework," *Sports Med.*, vol. 51, no. 1, pp. 1–10, Jan. 2021.
[6] S. Sharma and G. Whyte, *Practical ECG for Exercise Sci. and Sports Med.*. Champaign, IL, USA: Human Kinetics, 2010.
[7] A. P. Chandrakasan, N. Verma, and D. C. Daly, "Ultralow-power electronics for biomedical applications," *Annu. Rev. Biomed. Eng.*, vol. 10, no. 1, pp. 247–274, Aug. 2008.
[8] D. Zoni, A. Galimberti, and W. Fornaciari, "An FPU design template to optimize the accuracy-efficiency-area trade-off," *Sustain. Comput. Inform. Syst.*, vol. 29, 2021, Art. no. 100450.
[9] A. Pullini et al., "Mr.Wolf: An energy-precision scalable parallel ultra low power SoC for IoT edge processing," *IEEE J. Solid-State Circuits*, vol. 54, no. 7, pp. 1970–1981, Jul. 2019.
[10] S. Benatti et al., "Online learning and classification of EMG-based gestures on a parallel ultra-low power platform using hyperdimensional computing," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 3, pp. 516–528, Jun. 2019.
[11] E. De Giovanni, et al., "Real-time personalized atrial fibrillation prediction on multi-core wearable sensors," *IEEE Trans. Emerg. Topics Comput.*, vol. 9, no. 4, pp. 1654–1666, Oct.–Dec. 2021.
[12] P. Kirchhof et al., "2016 ESC guidelines for the management of atrial fibrillation developed in collaboration with EACTS," *Eur. Heart J.*, vol. 37, no. 38, pp. 2893–2962, Oct. 2016.

[13] D. Sopic, et al., "Real-time event-driven classification technique for early detection and prevention of myocardial infarction on wearable systems," *IEEE Trans. Biomed. Circuits Syst.*, vol. 12, no. 5, pp. 982–992, Oct. 2018.
[14] F. Forooghifar et al., "A self-aware epilepsy monitoring system for real-time epileptic seizure detection," *Mobile Netw. Appl.*, vol. 27, pp. 677–690, 2022.
[15] B.-U. Köhler, C. Hennig, and R. Orglmeister, "The principles of software QRS detection," *IEEE Eng. Med. Biol. Mag.*, vol. 21, no. 1, pp. 42–57, Jan./Feb. 2002.
[16] J. P. Martínez, et al., "A wavelet-based ECG delineator: Evaluation on standard databases," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 570–581, Apr. 2004.
[17] Q. Li, R. G. Mark, and G. D. Clifford, "Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter," *Physiol. Meas.*, vol. 29, no. 1, pp. 15–32, Jan. 2008.
[18] L. Orlandic et al., "REWARD: Design, optimization, and evaluation of a real-time relative-energy wearable R-peak detection algorithm," in *Proc. Eng. Med. Biol. Conf.*, 2019, pp. 3341–3347.
[19] R. Gosselink, T. Troosters, and M. Decramer, "Exercise testing: Why, which and how to interpret," *Breathe*, vol. 1, pp. 120–129, Dec. 2004.
[20] F. Cottin et al., "Assessment of ventilatory thresholds from heart rate variability in well-trained subjects during cycling," *Int. J. Sports Med.*, vol. 27, pp. 959–967, Dec. 2006.
[21] D. J. Bentley, V. E. Vleck, and G. P. Millet, "The isocapnic buffering phase and mechanical efficiency: Relationship to cycle time trial performance of short and long duration," *Can. J. Appl. Physiol.*, vol. 30, no. 1, pp. 46–60, Feb. 2005.
[22] M. Buchheit, R. Solano, and G. P. Millet, "Heart-rate deflection point and the second heart-rate variability threshold during running exercise in trained boys," *Pediatr. Exercise Sci.*, vol. 19, no. 2, pp. 192–204, May 2007.
[23] R. Braojos, et al., "Embedded real-time ECG delineation methods: A comparative evaluation," in *Proc. 12th Int. Conf. Bioinf. Bioeng.*, 2012, pp. 99–104.
[24] M. S. Thaler, *The Only EKG Book You'll Ever Need*, 9th ed.Alphen aan den Rijn, The Netherlands: Wolters Kluwer Law & Business, 2018.
[25] M. P. Tulppo, et al., "Quantitative beat-to-beat analysis of heart rate dynamics during exercise," *Amer. J. Physiol.-Heart Circulatory Physiol.*, vol. 271, pp. H244–H252, Jul. 1996.
[26] F. J. Richards, "A flexible growth function for empirical use," *J. Exp. Botany*, vol. 10, pp. 290–301, Jun. 1959.
[27] PULP SDP - *GitHub - pulp-platform/pulp-sdk*, 2019. [Online]. Available: https://github.com/pulp-platform/pulp-sdk
[28] P. D. Schiavone et al., "Slow and steady wins the race? A comparison of ultra-low-power RISC-V cores for Internet-of-Things applications," in *Proc. 27th Int. Symp. Power Timing Model. Optim. Simul.*, 2017, pp. 1–8.
[29] Research | biopac, 2022. [Online]. Available: https://www.biopac.com/research/
[30] Association for the Advancement of Medical Instrumentation (AAMI), "Testing and Reporting Performance Results of Cardiac Rhythm and ST-Segment Measurement Algorithms," ANSI/AAMI EC57:2012 (R2020) Standard, 2020.
[31] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985.
[32] A. Lourenço, et al., "Real time electrocardiogram segmentation for finger based ECG biometrics," in *Proc. Int. Conf. Bio-Inspired Syst. Signal Process.*, 2012, pp. 49–54.
[33] G. Moody, T. Pollard, and B. Moody, "WFDB software package (version 10.6.2)," *Physionet*, 2021. [Online]. Available: https://doi.org/10.13026/zzpx-h016
[34] V. Kalidas and L. Tamil, "Real-time QRS detector using stationary wavelet transform for automated ECG analysis," in *Proc. IEEE 17th Int. Conf. Bioinf. Bioeng.*, 2017, pp. 457–461.
[35] B. Porr, H. Luis, S. Ioannis, and N. Yoav, "Popular ECG R peak detectors written in python (version 1.3.2)," Zenodo, 2022, doi: 10.5281/zenodo.6812478.
[36] E. De Giovanni et al., "ECG in high intensity exercise dataset," Zenodo, 2021, doi: 10.5281/zenodo.5727800.
[37] E. Flamand et al., "GAP-8: A RISC-V SoC for AI at the edge of the IoT," in *Proc. Int. Conf. Appl.-Specific Syst. Architectures Processors*, 2018, pp. 1–4.
[38] E. De Giovanni et al., "Modular design and optimization of biomedical applications for ultralow power heterogeneous platforms," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 39, no. 11, pp. 3821–3832, Nov. 2020.