POLITECNICO DI TORINO
Repository ISTITUZIONALE

Energy-Aware Bio-Signal Compressed Sensing Reconstruction on the WBSN-Gateway

Original
Energy-Aware Bio-Signal Compressed Sensing Reconstruction on the WBSN-Gateway / Bortolotti, Daniele; Mangia, Mauro; Bartolini, Andrea; Rovatti, Riccardo; Setti, Gianluca; Benini, Luca. - In: IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTING. - ISSN 2168-6750. - STAMPA. - 6:3(2018), pp. 370-381. [10.1109/TETC.2016.2564361]

Availability:
This version is available at: 11583/2728419 since: 2021-04-07T23:01:19Z

Publisher:
Institute of Electrical and Electronics Engineers Inc.

Published
DOI:10.1109/TETC.2016.2564361

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

12 October 2024

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TETC.2016.2564361

Energy-Aware Bio-signal Compressed Sensing Reconstruction on the WBSN-gateway

Daniele Bortolotti, Mauro Mangia, *Member, IEEE,* Andrea Bartolini, *Member, IEEE,* Riccardo Rovatti, *Fellow, IEEE,* Gianluca Setti, *Fellow, IEEE,* Luca Benini, *Fellow, IEEE*

Abstract—Technology scaling enables today the design of ultra-low power wearable bio-sensors for continuous vital signs monitoring or wellness applications. Such bio-sensing nodes are typically integrated in Wireless Body Sensor Network (WBSN) to acquire and process biomedical signals, e.g. Electrocardiogram (ECG), and transmit them to the WBSN gateway, e.g. smartphone, for online reconstruction or features extraction. Both bio-sensing node and gateway are battery powered devices, although they show very different autonomy requirements (weeks vs. days). The *rakeness*-based Compressed Sensing (CS) proved to outperform standard CS, achieving a higher compression for the same quality level, therefore reducing the transmission costs in the node. However, most of the research focus has been on the efficiency of the node, neglecting the energy cost of the CS decoder. In this work, we evaluate the energy cost and real-time reconstruction feasibility on the gateway, considering different signal reconstruction algorithms running on a heterogeneous mobile SoC based on the ARM big.LITTLE ™ architecture. The experimental results show that it is not always possible to obtain the theoretical QoS under real-time constraints. Moreover, the standard CS does not satisfy real-time constraints, while the rakeness enables different QoS-energy trade-offs. Finally, we show that in the optimal setup (OMP, $n = 128$) heterogeneous architectures make the CS decoding task suitable for wearable devices oriented to long-term ECG monitoring.

Index Terms—Compressed Sensing, Reconstruction, ECG monitoring, WBSN, Real-time decoding, Energy-efficiency.

1 Introduction

Heart activity monitoring is of primary interest in a wide variety of human habits. At the present time, cardiovascular or lifestyle-induced diseases are affecting a growing portion of the world population. In general, human behavior-related illnesses require accurate and long-term medical supervision which, considering the figures, is unsustainable for the traditional healthcare system due to the high costs and medical management resources needed [1]. To address this problem, personal vital signs monitoring systems are able to offer a cost-effective solution on a large scale. As a matter of fact, emerging and future healthcare policies are fueling a shift towards ubiquitous long-term monitoring of bio-signals by means of wearable ultra-low power devices [2].

Wearable miniaturized bio-sensing nodes, integrated in a Wireless Body Sensors Network (WBSN) to continuously measure and remotely report biomedical signals, can indeed provide the ubiquitous, long-term and real-time monitoring required by the patients. In the WBSN scenario the bio-sensing nodes collect sensors data and transmit them to the gateway which in turn can forward the acquired data to the base station through a WAN network, or perform real-time reconstruction for data visualization or bio-signal analysis. Recently, beside medical grade monitors, such as Holter monitors and Electrocardiogram (ECG) loop recorder devices, a growing market segment is represented by reduced-leads ECG monitors used for lifestyle heart activity monitoring. The typical applications include, for instance, wellness and sport activity trackers as well as obesity and stress detectors. In the WBSN context, such applications imply that the information sensed by the WBSN node is displayed in real-time on the personal gateway, e.g. a smartphone, to offer live information to the user. This translates into extra processing requirements with real-time constraints for already limited battery-powered WBSN gateways.

In this direction, the Compressed Sensing (CS) paradigm for signal acquisition and compression [3] has proved to be an energy efficient approach suitable for WBSN based embedded biomedical monitors [4], [5], [6]. The aim of CS is to represent the information content of the input signal using fewer digital words with respect to Nyquist-rate sampling. Embedded signals are acquired from a set of distributed sensors and compressed with very low energy consumption to alleviate the communication constraints (bandwidth and energy) between the sensors and a base station, typically through the gateway. Moreover, the general CS theory was recently extended with the introduction of the concept of *rakeness* [7]. The underlying idea behind this approach is to exploit the localization of signals, i.e. the assumption that the information content of the signal is not equally distributed in its domain. In [7] authors showed that the specialization of the CS approach over classes of localized signals allows to achieve either a higher quality of

- *D. Bortolotti is with the Department of Electrical, Electronic and Information Engineering (DEI), University of Bologna, Bologna, Italy. E-mail: daniele.bortolotti@unibo.it*
- *M. Mangia and R. Rovatti are with the Advanced Research Center on Electronic Systems (ARCES) and the Department of Electrical, Electronic and Information Engineering (DEI), University of Bologna, Bologna, Italy. E-mail: {mauro.mangia2, riccardo.rovatti}@unibo.it*
- *G. Setti is with the Engineering Department of the University of Ferrara (ENDIF), University of Ferrara, Ferrara, Italy. E-mail: gianluca.setti@unife.it*
- *A. Bartolini and L. Benini are with the Integrated Systems Laboratory (ISL), ETH Zurich, Zurich, Switzerland. E-mail: {barandre, lbenini}@isl.ee.ethz.ch*

Manuscript received December 1, 2015; revised Month Day, Year.

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TETC.2016.2564361

the reconstructed signal, or an increased data compression necessary to target a given quality. The authors in [8] further validated the benefits of the rakeness-based CS, showing the improved energy-efficiency for the CS encoder on an ultra-low power bio-medical processor [8].

On the other hand, modern CPU architectures currently permeating in the mobile market, enable to exploit heterogeneity as an additional knob on top of standard power management techniques, such as DVFS and power gating, to increase the energy efficiency on a wider workload spectrum [9], [10], [11]. The big.LITTLE architectural paradigm, introduced by ARM [9], exploits the presence of an additional cluster of processors beside a high-performance Cortex-A15 cluster which embeds simpler, yet more energy efficient, cores with the same ISA. This enables to achieve further energy savings by dynamically migrating and executing less demanding tasks on the more energy efficient cores. Such heterogeneous architectures, clearly represent optimal candidates for energy efficient WBSN gateways in charge of reconstructing in real-time the bio-signals.

In this work we investigate key design parameters of a CS-based system for ECG monitoring, with a focus on the WBSN gateway in charge of reconstructing the compressed bio-signal. At first, we analyze the effects of both the input window length and the sensing procedure and evaluate the performance of different iterative reconstruction algorithms in terms of Quality of Service (QoS) and computational time. We further extend the analysis investigating the energy consumption and the real-time achievable QoS on the gateway, by profiling the reconstruction algorithms on a modern heterogeneous System-on-Chip (SoC). The main contributions of this paper are the following:

- we present an extensive study of the trade-offs in terms of reconstruction quality and energy efficiency, considering CS-based WBSN architectures for bio-signal monitoring, with a focus on the CS decoding stage;
- we highlight that it is not always possible to obtain the theoretical QoS under real-time constraints. Moreover, we show that the common belief that the lowest compression ratio always leads to the highest reconstruction quality, does not hold when considering performance- or energy-constrained HW resources;
- we highlight how CS parameters affect the different reconstruction algorithms, further exploring the performance/energy trade-offs when running on a SoC based on ARM big.LITTLE as WBSN gateway. The ARM big.LITTLE SoC allows to stress this analysis on two corner case processors for embedded systems in the same device;
- we show that in the optimal setup (OMP, $n = 128$) heterogeneous architectures based on the big.LITTLE paradigm make the CS decoding task suitable for wearable devices oriented to long-term ECG monitoring.

The rest of the paper is organized as follows. In Section 2 the related works are presented, while Section 3 focuses on the target WBSN architecture for ECG monitoring. Section 4 introduces the standard CS theory, its rakeness-based extension and the considered reconstruction algorithms. Next, in Section 5 and 6 we describe the experimental setup and the results of the evaluation in terms of reconstruction quality, real-time feasibility and energy efficiency. Finally, the conclusions are presented in Section 7.

2 Related Work

Most of the research related to CS-based WBSN architectures for bio-signal monitoring has focused on the encoder stage, i.e. the WBSN node, with the goal of achieving a target quality of service in reconstruction and reduce as much as possible the amount of digital words to be transmitted to the gateway [12], [5], [6], [8], [13]. In general, considering streaming architectures for WBSNs, where the data is not stored in the node but continuously transmitted to the gateway, the focus has always been in extending the battery life of the node, neglecting that the gateway is a battery-powered device as well, with limited computational resources. In fact, the CS reconstruction problem is potentially an NP-hard problem, especially in its formulation with l_0 norm minimization. Nevertheless, there are iterative algorithms that find sub-optimal solutions in its l_2 norm version, such as [14], [15].

The authors in [16] validate the concept of real-time CS reconstruction, deploying a Shimmer™ wearable sensor node to acquire an ECG signal and an iPhone as gateway for real-time reconstruction. However, the trade-offs in terms of energy, time and quality are not explored. In general, the link between the compression ratio in the CS encoder (node) and the real-time reconstruction feasibility and achievable target QoS in the CS decoder (gateway), is missing in literature. Moreover, the fact that both the WBSN node and the gateway are battery-powered devices, though with different autonomy requirements (weeks vs. days), requires energy efficiency on both encoding and decoding stages.

Authors in [17], [18] compare the execution time and the maximum achievable SNR of different reconstruction algorithm for ECG signals. The authors consider both L1-norm convex algorithm (basis pursuit and basis pursuit de-noising) and greedy implementations such as Orthogonal Matching Pursuit (OMP) and the Compressed Sensing Matching Pursuit (CoSaMP) algorithms. They show that OMP requires the least computational time. Convex optimization is penalized by its higher computational complexity more than greedy solutions but it is notably independent of sparsity. CoSaMP has the larger computational time as effect of the costly pseudo-inverse calculation required at each iteration. In this work we start from the results of this analysis to evaluate the impact of the different algorithms on the energy consumption of an embedded platform subject to real-time constraints.

A first study of the energy efficiency and achievable CS reconstruction quality in a battery-powered decoder has been carried out in [19]. The authors show that, when performance constrained, the real-time SNR achieved by the decoder is below the theoretical one and that increasing the compression ratio has benefits for the battery life-time of both the WBSN node as well as WBSN gateway. However, the study is conducted for a specific reconstruction algorithm (FOCUSS) which, as shown in this work, is not the

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TETC.2016.2564361

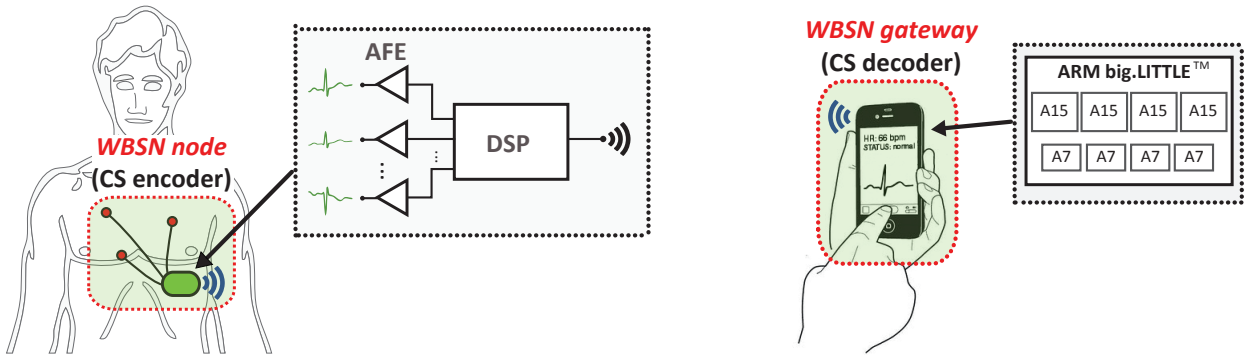

Fig. 1: WBSN architecture for ECG monitoring: wearable bio-sensing *WBSN node*, composed of AFE and DSP and in charge of the *CS encoding* stage, and heterogeneous multi-core mobile *WBSN gateway* in charge of the *CS decoding* stage.

most suitable in different configurations, leading to a reconstruction time overhead. In this work we generalize [19] by considering two additional top-performing reconstruction algorithms and exploring the impact that the input window size has on the performance of the WBSN gateway.

3 WBSN ARCHITECTURE FOR ECG MONITORING

As previously introduced, WBSN architectures to continuously measure and remotely report the ECG signal, can offer the ubiquitous and real-time monitoring required by the patients at a sustainable cost. In such WBSN scenario, the bio-sensing nodes collect bio-signal data from the sensors and wirelessly transmit it to the gateway, which in turn can forward it to a base station or perform real-time reconstruction for data usage directly on the gateway.

A graphical representation of the WBSN architecture for ECG monitoring considered in this work is presented in Figure 1. It mainly consists of two devices: the *WBSN node*, a bio-sensing device worn by the patient, where the bio-signal is collected and the encoding/compression occurs and the *WBSN gateway*, depicted as an heterogeneous multi-core smartphone, where the bio-signal decoding/reconstruction takes place.

Focusing on the WBSN node, the input multi-lead ECG signal is sampled by the analog front-end (AFE) in a data collection phase, with a given sampling frequency according to the properties of the signal to analyze and the accuracy needed. Similarly to [8], once all the samples are collected, the DSP executes the compressed sensing encoding stage (described in Section 4). The considered system performs a burst of computation on the available data for future transmission. In the last stage, within a given time window, the compressed data are transmitted to the gateway to perform real-time reconstruction for further usage.

In traditional CS-based WBSN architectures, the personal server acted as a gateway to a WAN network to make the sensed data remotely available. Therefore, the compressed data was simply forwarded to a remote workstation for offline analysis, neglecting any constraint on real-time decoding and energy consumption. However, modern smartphones offer an impressive amount of GFLOPS, enabling real-time decoding and analysis of the bio-signals. Moreover, with the goal of energy efficiency, a consolidating trend in mobile SoC architectures couples a high-performance cluster of cores with a less powerful, yet more energy efficient, cluster. The paradigm introduced by ARM with the big.LITTLE technology [9] found its way into several current SoCs [11], [10]. This scenario opens up a new design space for an energy-aware real-time reconstruction of the sensed data, and easily extends the single-lead/single-core reconstruction scheme to a multi-core/multi-lead case.

4 COMPRESSED SENSING PARADIGM

The main characteristic of the CS paradigm is to represent the information content of long slices of a signal in a compressed way. with respect to a vector $x \in \mathbb{R}^n$ containing the corresponding n Nyquist samples of the original analog signal, CS aims at reducing the minimum amount of needed digital words for each signal slice with respect to the limit imposed by the Shannon-Nyquist theorem. To overcome this limit, CS works with particular classes of signals, namely *sparse* signals.

A class of signals is sparse if a proper basis (or dictionary) Ψ exists, such that each possible signal instance x holds a sparse representation when expressed over Ψ. More formally, we have $x = \Psi\alpha$ where $\Psi \in \mathbb{R}^{n\times n}$ is a matrix and the corresponding coefficients vector $\alpha \in \mathbb{R}^n$ contains only K non null elements (with $K \ll n$). For all possible classes of sparse signals, the CS theory proposes an innovative way for both encoding and decoding stages [20].

In the *encoder stage*, the signal information content to be transmitted is collected in the measurements vector y, which is evaluated by a set of $m < n$ projections of x (the bio-signal) on a set of properly designed sensing sequences ϕ_j, $j = 1, \ldots, m$, arranged as rows of the *sensing matrix* $\Phi \in \mathbb{R}^{m,n}$ such that:

$$y = \Phi x + \nu = \Phi\Psi\alpha + \nu = A\alpha + \nu \tag{1}$$

where $A = \Phi\Psi$ is a $m \times n$ is a matrix linking the sparse representation α with the measurements vector y and $\nu \in \mathbb{R}^m$ takes into account all system non-idealities, such as quantization error or other noise sources. The main characteristic of a standard CS encoder is the achieved *compression ratio* defined as $CR = n/m$. The CS encoding stage con be made directly in the analog-domain, as proposed in [21], [5], [22]

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TETC.2016.2564361

or in the digital domain after a standard digitalization of the input signal at Nyquist rate, as in [8], [12], [23], [4].

Finally, as prescribed by standard CS theory [3], the sensing sequences design must follow two important guidelines, incoherence and the Restricted Isometry Property (RIP) [3], [24]. Incoherence guarantees that each sparse vector α is mapped in a dense vector y, i.e. each sensing sequence is able to collect part of the signal information content, while a matrix A verifies RIP if it is able to preserves the l_2 norm of x and it guarantees that y is not in the null space of A.

Designing a sensing matrix which verifies both guidelines is not a trivial task. However, it is known that, for an arbitrary Ψ, A holds both RIP and incoherence by generating the ϕ_j as instances of independent and identically distributed (i.i.d.) random variables, such as a set of random antipodal symbols, i.e. where $+1$ and -1 occur with the same probability. In this case the required amount of projection must satisfy $m \geq m_{\min} = 4K\log(n/K)$ [20], [3].

After the transmission of y, the *decoder stage* aims at recovering the original signal, i.e., the reconstruction of α from the knowledge of the measurement collected as in (1). This translates into the inversion of an ill-posed operator A, which means that infinite vectors α are mapped into y by A. Without any knowledge on the acquired class of signal, a trivial solution can be obtained by the Moore-Penrose inverse of A which returns the vector $\hat{\alpha}$ with the lowest l_2 norm. Furthermore, we know that α is sparse so we can look at the sparsest vector $\hat{\alpha}$ that is mapped into y by A. From an algorithmic point of view, this translates into solving the following optimization problem:

$$\begin{aligned} \hat{\alpha} &= \min \|\alpha\|_1 \\ \text{s.t.} \quad & \|\Phi\Psi\alpha - y\|_2 < \epsilon \end{aligned} \tag{2}$$

where $\|\cdot\|_1 = \sum|\cdot|$ and $\|\cdot\|_2 = \sqrt{\sum \cdot^2}$ define the standard l_1 and l_2 norms, ϵ bounds noise (ν) effects and the reconstructed signal can be written as $\hat{x} = \Psi\hat{\alpha}$ [3].

4.1 rakeness-based CS

The standard CS approach works if the sparsity assumption holds without any tuning on the proper class of acquired signals. A new CS approach was introduced in [7], [25] presenting an innovative way to adapt the encoding procedure on the statistical characterization of the acquired class of signals[1]. More specifically, beside sparsity it exploits the *localization* of the input instances, i.e. their energy content is not uniformly distributed in the whole signal domain.

This is based on a novel guideline in the sensing sequences design, named *rakeness*. The rakeness-based CS relies on the tuning of the statistical properties of each sensing sequences ϕ_j to the statistical characterization of the acquired class of signals. Its aim is to minimize the amount of linear projections needed for a correct input signal reconstruction by maximizing the energy collected in the encoder stage, preserving at the same time a minimum level of randomness of the corresponding operator A. In other words, the Φ matrix is still composed row by row by independent random sequences, however each sequence has a correlation profile depending on the correlation profile of the acquired class of signals.

1. [ONLINE] http://cs.signalprocessing.it

Fig. 2: An antipodal sensing matrix following standard CS (top) coupled with another antipodal sensing matrix generated as prescribed by the rakeness-based CS for ECG signals (bottom). Both matrices were generated with CR=4. Note that white means -1 while black corresponds to $+1$.

The final step is to generate antipodal random sequences, which is a hardware-friendly choice, with a prescribed correlation profile. While the generation of a random sequence with an advised correlation is quite easy, the generation of antipodal sequences given its statistical characterization is a more complex task. In this field, different approaches have been proposed in literature [26], [27], [28], of which the simplest one relies on thresholding of Gaussian random vectors [26], [29]. Although not completely general, this approach is very simple and can be easily specialized to our case as in [21]. A trivial example of the output of this process is reported in Fig. 2 (bottom) where ECG signals are taken as acquired class of signals. It is known that this particular class of signals presents a low-pass filter profile in the power spectral density [7] so that, according to the rakeness-based CS, even the sensing sequences must hold a similar spectral profile. Fig. 2 (top) shows also a completely random sensing matrix as prescribed by the standard CS approach. In both cases each window is composed of $n = 128$ samples and $CR = 4$.

The main advantage of the rakeness-based CS with respect to the standard approach is in the reduction of the amount of projections needed for a fixed quality of service, without any change in the adopted family of architectures for the encoder stage [12], [8], [21].

4.2 CS Decoder

In the last two decades a large amount of contributions was presented to solve the optimization problem in (2), with applications in different areas such as spectral estimation, image restoration and denoising. This massive research effort has produced different algorithms and tools that solve (2). Having in mind the power reduction of the decoding stage, in this work we consider three different iterative decoding algorithms within the family of greedy methods, namely (i) Orthogonal Matching Pursuit (*OMP*), (ii) Compressive Sampling Matching Pursuit (*CoSaMP*) and (iii) FOCal Underdetermined System Solver (*FOCUSS*).

The OMP mechanism iteratively chooses Ψ columns and its coefficients by solving a least-squares problem to approximate the signal on the measurements vector y and

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TETC.2016.2564361

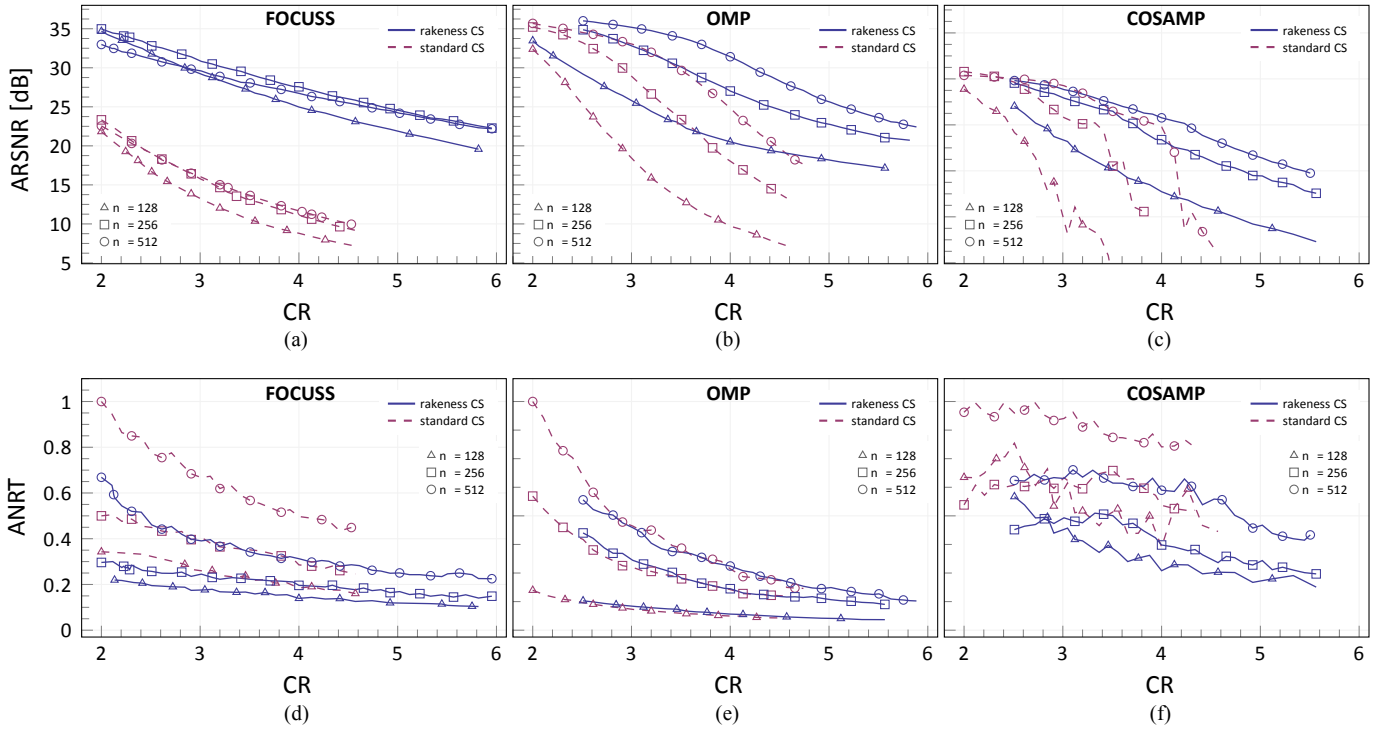

Fig. 3: ARSNR vs. CR for both standard CS and rakeness-based CS, varying the input bio-signal window length (n) for each reconstruction algorithm: FOCUSS (a), OMP (b) and COSAMP (c). ANRT vs. CR for the same configurations and algorithms is reported in plots (d), (e) and (f).

then produces a residual vector that is the signal to be approximated at the next iteration. Its main characteristic is that after every step, all the coefficients extracted so far are updated, by computing the orthogonal projection of the signal onto the set of selected Ψ columns [30].

Such algorithm was further adapted to the recovery of sparse signals, i.e. the solution of (2), by the CoSaMP decoding algorithm. The additional assumption with respect to OMP is the knowledge of the signal sparsity level K. As for the previous case, the core of each iteration is the solution of a least-squares problem in order to approximate as best as possible either the sparse signal or on its residual coming from the previous iteration on the measurement vector y. Furthermore, in every iteration the solution of the least-squares problem is followed by a selection of a subset of Ψ columns corresponding to the higher coefficients with cardinality $2K$ [31].

FOCUSS is one of the most stable and studied iterative approaches [14], [15], [32]. It is composed of a low-resolution initial estimation of the sparse vector, followed by iterations that refine the solution to the final reconstructed signal, i.e., the sparse vector $\hat{\alpha}$. The core of each iteration is a weighted least-squares minimization based on the Singular Value Decomposition (SVD) of the operator A (cf. eq. (2)), which is updated by a proper set of sparsity promoting weights [14], [15]. From a computational point of view, the complexity of each iteration is mainly represented by the heavy SVD operation.

In general, every decoding algorithm performance depends on the choice of n, i.e. the amount of samples taken in the encoding stage. If n is small, the complexity of the solution of (2) scales with it in a non linear way, leading to smaller computational problems in the decoding stage, therefore saving time and energy on the gateway. Intuitively this leads to the choice to take n as small as possible. However, for bio-signals such as ECG, the choice of n too small could lead to an increase in the sparsity of the input bio-signal slices, leading to a reduced quality in reconstruction.

5 CS Reconstruction Quality and Time

One of the points of focus of this work is on the effects of n (number of samples in a slice of signal to be compressed) and the sensing procedure (either standard of rakeness-based CS) on the performance of the different reconstruction algorithms, in terms of quality of service and computational time. To achieve a fair comparison of the proposed scenarios, we firstly adopt MATLAB Montecarlo simulations over a set of 5000 synthetic ECG instances, generated as in [7], with a sampling rate equal to 360Hz and we considered as sparsity basis the orthonormal Symlet-6 wavelet basis. To emulate all possible nonidealities, signals are corrupted by an additive Gaussian noise such that the corresponding signal to noise ratio is 40 dB. To measure the quality of service, we consider the average reconstruction signal to noise ratio (*ARSNR*) defined as follows:

$$\mathrm{ARSNR} = \mathrm{E}_{x,\Phi}\left[20\log_{10}\left(\frac{\|x\|_2}{\|x-\hat{x}\|_2}\right)\right]$$

where $20\log_{10}\left(\frac{\|x\|_2}{\|x-\hat{x}\|_2}\right)$ is the RSNR with $\hat{x} = \Psi\hat{\alpha}$, $\mathrm{E}[\cdot]$ is the expectation and averaging over Φ and x means over a set of sensing matrices and over a set of ECG signal instances.

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TETC.2016.2564361

		FOCUSS			OMP			CoSaMP		
		$n = 128$	$n = 256$	$n = 512$	$n = 128$	$n = 256$	$n = 512$	$n = 128$	$n = 256$	$n = 512$
LQ	standard CS	2.61	3.46	4.12	2.56	3.66	4.61	2.67	3.56	4.16
	rakeness CS	**6.40**	**9.21**	**10.89**	4.13	6.92	8.53	3.37	5.45	6.75
HQ	standard CS	-	-	-	1.91	2.56	3.2	-	-	2.91
	rakeness CS	1.97	2.23	2.09	**2.37**	**2.84**	**3.76**	-	-	2.50

TABLE 1: Comparison between standard CS and rakeness CS in terms of compression ratios for two different operation points (HQ and LQ) considering three different window lengths and three different decoding algorithms.

For the choice of Φ, we can either adopt completely random antipodal sensing matrices, as prescribed by standard CS approach, or the case where the correlation profile of each antipodal Φ row depend on the correlation profile previously estimated for ECGs as prescribed by the rakeness-based CS approach.

The figure of merit we propose to compare the computational efficiency is the Average Normalized Reconstruction Time (*ANRT*). In details, for each reconstruction algorithm we measure the CPU time required to reconstruct $N = 512$ successive signal samples, when considering a target CR (n/m) and a given CS approach (i.e. either standard CS or rakeness-based CS). Having in mind that n is a parameter under investigation, the compression of N samples translates to one window for $n = 512$, two windows for $n = 256$ and four windows for $n = 128$. We limit our exploration to $n = 128$ to avoid a high variance in the sparsity of ECG instances. As an example, considering as window length $n = 64$ means 0.1778 seconds of signal for each time window, which is approximately the time duration of the QRS complex. Instances containing this pieces of signal are not sparse and CS fails the reconstruction. In this work we set $n = 128$ as lower bound in order to prevents these undesired events.

Next, to compute the ANRT, the reconstruction time is averaged over a dataset of ECG signal windows and normalized with respect to the slowest observed for that particular decoding algorithm.

For what concerns the configuration adopted for the reconstruction algorithms herein considered, FOCUSS follows the same setup as the one presented in [19], OMP was consifigured with a stop criteria based on the residual error associated to each iteration, i.e. $\|y - \Phi\Psi\alpha\|_2 \leq 10^{-4}$, while CoSaMP uses the same stop criteria as OMP but requires as input an estimation of the sparsity level K.

The obtained results in terms of ARSNR and ANRT are shown in Figure 3 for all the aforementioned reconstruction algorithms and input window size. The results clearly indicate that the rakeness-based CS outperforms the standard approach for every tested algorithm and for every value of n and CR. This is primarily due to the main property of the rakeness-based CS, i.e. it enables a higher ARSNR for a fixed CR with respect to the standard CS. We show here that the reduced amount of information to be processed is directly connected to the computational time that is, in general, representative of the power consumption. More details on the power figures of these operative points on the gateway will be discussed in the following sections.

Moreover, we exploit the trade-off between data compression and the goodness of the reconstruction, considering two different reconstruction quality standards:

- *High Quality* (HQ): ECG instances are correctly reconstructed for healthcare applications (medical grade)
- *Low Quality* (LQ): the quality of service targets wellness applications (e.g. heart rate detection).

In our setup these operative points were defined by setting the probability to have a RSNR greater than a fixed threshold (97%). Here the operative definition of both standards.

- HQ is such that: *Prob*[RSNR $\geq$ 25 dB] ≥ 0.97
- LQ is such that: *Prob*[RSNR $\geq$ 7 dB] ≥ 0.97

The first threshold is fixed by a visual inspection of the reconstructed signals while for the LQ standard the associated quality of service guarantees at least a good estimation of the heart rate [12].

The compression ratios for all the decoding algorithms and for $n = \{128, 256, 512\}$ corresponding to the two quality standards are reported in Table 1. From the encoder point of view (increase CR as much as possible), these results say that the rakeness-based CS outperforms the standard approach as in the previous case. Furthermore, these data suggest to use FOCUSS for wellness oriented applications (LQ), while in case a higher reconstruction quality is needed (HQ) the OMP decoding procedure guarantees a higher performance. CoSaMP suffers from the high variability of the sparsity estimation K, leading to discard it and focus only on FOCUSS and OMP for embedded bio-signal reconstruction.

With the aim of a robust characterization, in our simulation setting we consider also 50 successive time windows of the MIT-BIH arrhythmia database [33], [34], which is the most commonly used database for the study of ECG compression algorithms. The recordings were digitized at 360 samples per second per channel with a 11-bit resolution over a 10 mV range. We further defined an observation window as composed of $N = 512$ samples, therefore setting the constraint for real-time reconstruction within a time window of $T_W = 1.4122$ seconds. Two chunks of ECG signal from the MIT-BIH arrhythmia database, coupled with their LQ and HQ reconstructed counterparts are reported in Figure 4.

The latter ECG dataset will be ported on the WBSN gateway platform to evaluate embedded power consumption

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TETC.2016.2564361

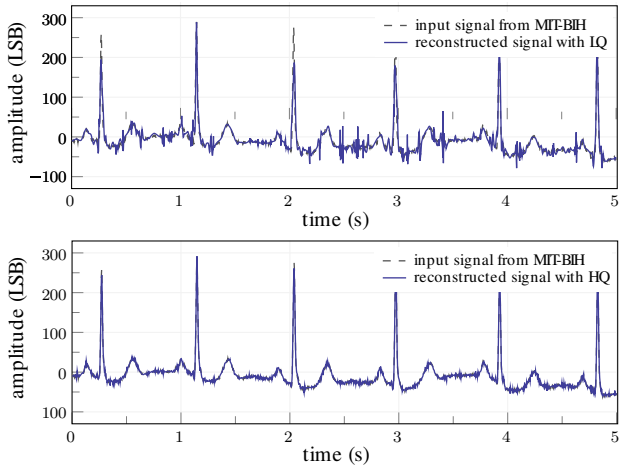

Fig. 4: A 5 seconds chunk of ECG signal from the MIT-BIH database coupled with the reconstructed signals for both LQ (top) and HQ (bottom) operative points, with $n = 512$ and reconstruction performed by the FOCUSS algorithm.

and reconstruction time for the two more interesting configurations, i.e. FOCUSS targeting the LQ quality standard and OMP for the HQ operative point.

6 Experimental Results

In this section we first present a characterization in terms of power consumption of the big.LITTLE architecture when executing the CS decoding algorithms. Then we analyze the feasibile QoS for real-time CS reconstruction on the WBSN-gateway. Finally we quantify the energy costs considering the different target QoS identified in Section 5.

6.1 Evaluation Board

To analyze the energy consumption and the quality of service on the WBSN-gateway, we considered the Hardkernel Odroid-XU3 board [35] based on Samsung Exynos 5422, the same CPU found in modern high-end smartphones. The Exynos 5422 implements ARM's big.LITTLE heterogeneous multiprocessing solution with a cluster of four Cortex-A15, "big" out-of-order processors, and a cluster of four "LITTLE" in order Cortex-A7 processors. Since both CPUs are architecturally compatible (ARMv7-A architecture), workloads can be allocated on demand to each CPU, to suit performance needs. However, the two clusters have very different floating point performance. To measure the consumption we deploy the on-board voltage/current sensors and split power rails, which allow to measure separately the power consumption of A15 cores, A7 cores, GPU and DRAM. To not affect power measurements, the readout of the sensors was implemented in a low-priority thread, with a sampling interval of 25ms and an average CPU consumption below $\approx 3\%$.

Following the considerations in the previous section, FOCUSS and OMP algorithms have been implemented in C++ to run on the ARM cores in order to profile the energy consumption on the embedded device. On top of the Odroid-XU3 runs Ubuntu 14.04.1 LTS (GNU/Linux 3.10.51+ armv7l) with `gcc` version 4.8.2. To achieve an efficient implementation and ease the porting of the algorithms, the C++ Armadillo library (v. 4.2) [36] was used for linear algebra. The single-threaded code to reconstruct one compressed ECG lead can be easily adapted to perform multi-threaded multi-lead reconstruction by mapping the multiple measurements version of the reconstruction algorithm on the eight ARM cores. The difference in terms of RSNR between the Matlab implementation (workstation, 64-bit floating point precision) and the embedded C++ (on Odroid-XU3, 32-bit floating point precision) is less than 10^{-15} averaged over the dataset for both reconstruction algorithms.

6.2 Power Profiling

As a first test, we profiled the execution of the CS reconstruction algorithms (OMP and FOCUSS) on both big and LITTLE cores of our WBSN gateway. The frequency range considered is $[0.8, 1.9]$ GHz for the Cortex-A15 and $[0.8, 1.3]$ GHz for the Cortex-A7. Before running one instance of the decoding algorithms, a sequence of scripts sets the frequency (bounded by setting $f_{min} = f_{max} = f_{target}$), triggers the sensors readout loop and finally sets the CPU affinity to run the reconstruction task on either a big or LITTLE core. The results are shown in Table 2, where we report the active and idle power of the running core for both OMP and FOCUSS algorithms, averaged over the MIT-BIH dataset of 50 windows. Our setup does not take into account the CR level, since the power consumption is not significantly affected by it. Table 2 reports the results of this exploration. We can first notice that FOCUSS and OMP have pretty similar power consumption for the different frequency levels and both of

			FOCUSS				OMP			
	P_{idle} (W)		P_{active} (W)		savings (%)		P_{active} (W)		savings (%)	
frequency [MHz]	A15	A7	A15	A7	A15	A7	A15	A7	A15	A7
800	0.185	0.040	0.469	0.112	78%	95%	0.463	0.099	79%	95%
1100	0.264	0.092	0.768	0.195	64%	91%	0.743	0.189	66%	90%
1300	0.334	0.141	0.960	0.288	54%	86%	0.980	0.255	56%	88%
1500	0.404	-	1.212	-	43%	-	1.120	-	45%	-
1700	0.538	-	1.618	-	23%	-	1.635	-	26%	-
1900	0.757	-	2.107	-	0%	-	2.205	-	0%	-

TABLE 2: Profiling of the idle and active power of FOCUSS (LQ standard) and OMP (HQ standard) algorithms running on top of the ODROID-XU3 board, considering either A15 or A7 over the frequency range.

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TETC.2016.2564361

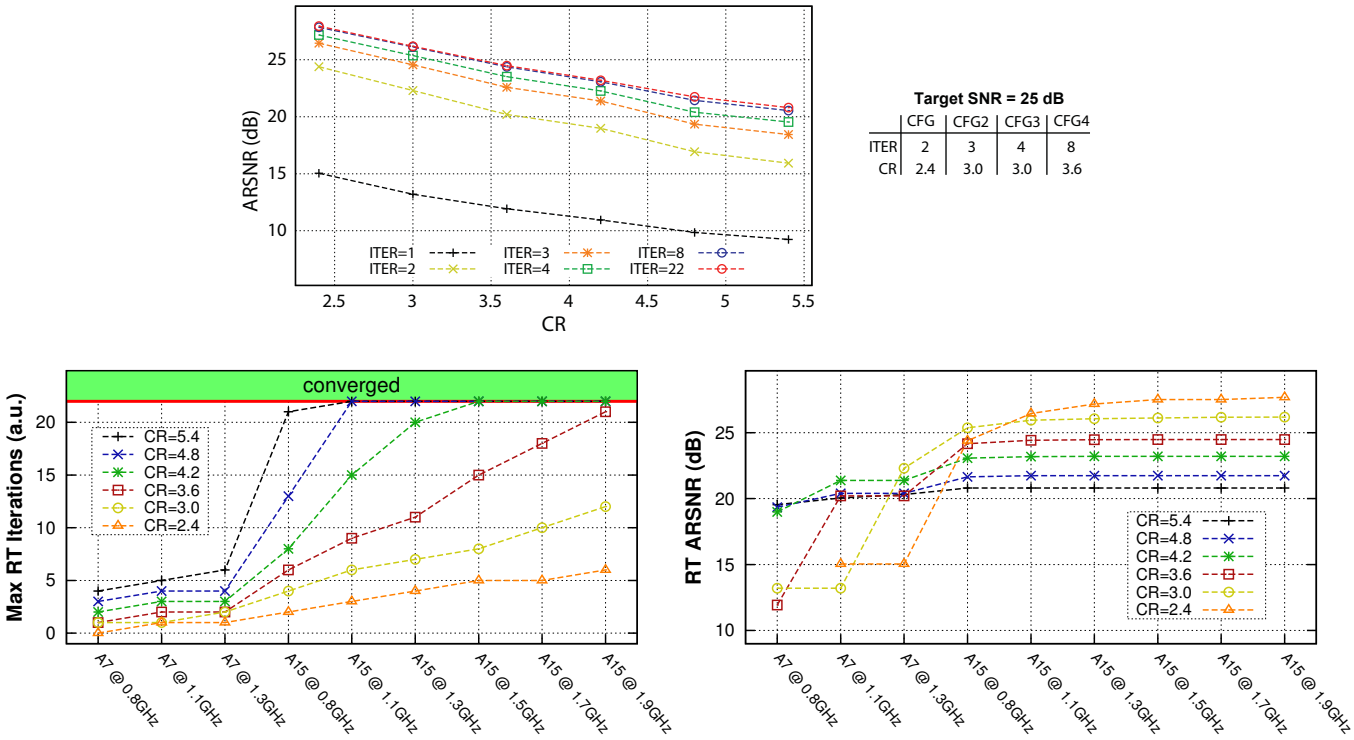

Fig. 5: Real-time analysis for FOCUSS on the big.LITTLE: ARSNR vs. CR as a function of CR and iterations (top), maximum number of iterations achieved in real-time (left) and relative ARSNR (right) for different compression ratios.

them increase the power consumption w.r.t. idle power. This suggests that the CS reconstruction task is a power intensive tasks and its energy cost depends on the operating point and on the decoding time. We can also notice that the power consumed by FOCUSS for CS reconstruction decreases of the 78%, when scaling the A15 frequency from 1.9GHz to 800MHz. Similar benefits are achieved with OMP, gaining a 79%. In addition, the big.LITTLE architecture can gain extra power reduction by migrating the reconstruction task to the A7 core. Specifically, for both OMP and COSAMP the maximum gain at the lowest operating point (A7 @ 0.8GHz) can consume as little as 5% with respect to the A15 @ 1.9GHz.

6.3 Real-Time Analysis

The previously defined power savings alone are not enough to evaluate the final energy-efficiency of the decoding algorithms when subjected to real-time constraint and limited computational performance typical of embedded platforms. We conducted an in depth analysis of the FOCUSS algorithm with $n = 512$, which is the most computationally demanding configuration. This section summarizes this analysis, initially carried out in [19], investigating the maximum achievable RSNR under a real-time (RT) constraint. Specifically, we analyzed the impact of the number of FOCUSS iterations on the ARSNR running on our evaluation board.

The plot in Figure 5 (top) shows the average ARSNR as a function of the CR, stopping the FOCUSS algorithm after a different number of iterations. Whereas the algorithm converges within the target tolerance for every CR value in maximum 22 iterations, "good" quality levels are reached with significantly less iterations. While after 1 iteration the ARSNR is significantly degraded of $\approx$ 15dB, after 2 iterations the degradation accounts for $\approx$ 5dB on the whole CR range. Such degradation becomes negligible only after 8 iterations and this holds for different compression ratios.

Figure 5 (left) shows that by increasing the CR level the computational complexity and the time required by each iteration decreases, allowing to compute more iterations within the same time budget. Moreover, by scaling the operating point the number of RT iterations decreases significantly. In the fastest and most power demanding operating point (A15@1.9GHz), the reconstruction algorithm converges to the optimal solution, within the RT constraint, only for CR ≥ 4.2. Instead, when considering the A7 cores, none of the CR reaches the optimal quality. There is a clear trade-off on the ARSNR since lower CR levels lead to better solutions however, at the same time, limit the number of RT iterations.

To complete the analysis, the plot in Figure 5 (right) shows the RT ARSNR achieved at the different operating points for different CR levels. We can see that the common belief that the lowest CR always leads to the highest reconstruction quality, does not hold when considering performance/energy- constrained HW resources. In fact, the plot shows that in the slowest, yet most energy efficient, operating point (A7@800MHz) the maximum ARSNR is achieved with a larger compression ratio (CR=5.4). In the same configuration the smallest compression ratio (CR=3.0) leads to an ARSNR reduction of $\approx$ 8dB, due to the reduced number of iterations imposed by the real-time constraint. Conversely, when we consider the A15@1.9GHz the highest ARSNR is achieved with the lowest compression ratio.

With the goal of an energy efficient design in mind, once

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TETC.2016.2564361

		CFG1	CFG2	CFG3	CFG4
CPU usage (%)	A7 @ 0.8 GHz	237.7%	219.5%	291.5%	396.3%
	A15 @ 1.9 GHz	30.7%	30.4%	40.4%	54.5%
Energy (J)	A7 @ 0.8 GHz	0.378	0.349	0.464	0.631
	A15 @ 1.9 GHz	1.100	1.089	1.447	1.952

TABLE 3: CPU usage, total energy for decoding one window considering different FOCUSS setups (CR and iterations) that target a RSNR of 25 dB. A CPU usage $> 100\%$ means that RT reconstruction can not be achieved.

we define a given RSNR as target quality, there are different couples of CR and FOCUSS iterations that can achieve it. With reference to Figure 5 (top), aiming at a RSNR of 25 dB[2], there are different possible configurations eligible for a design exploration. Table 3 shows the results in terms of energy and CPU utilization per window, when running the different FOCUSS configurations in the two corners of minimum and maximum performance. The results highlight that for the selected RSNR target level, a lower CR and less iterations improve both the performance metrics (lower CPU utilization and higher energy-efficiency) w.r.t. larger CRs and a higher iterations count. To meet the real-time constraint even with the most energy-efficient and less powerful core (A7@0.8GHz), in future works we will investigate the feasibility of a multi-core CS reconstruction.

As a final remark, note that the presented trade-off does not consider the energy consumption of the sensor node that is, in general, strongly related to CR, i.e. the amount of measurement transmitted from the sensor node to the gateway. In some scenarios, as for the case were the measurement vector is locally stored in an efficient non volatile memory [8], the low power requirement on the sensor node can be partially relaxed to pave the way for an optimization of the algorithm performing signal reconstruction. This is the case of [19], where an increased CR in order to improve the node lifetime, does not translate into an extended operation for the whole CS system due to the limitation imposed by the gateway.

6.4 Embedded CS Reconstruction

After analyzing the power consumption in the different operating points to perform the embedded CS reconstruction, we focused on the two algorithms that best performed at the different LQ and HQ quality standards previously identified in Section 5.

The plot in Figure 6 shows the average time in seconds to decode an ECG signal window of length $T_W = 1.4122$ seconds (cf. Section 5), using OMP for the HQ compression and FOCUSS for the LQ compression. For each case we evaluated the decoding time using a different n for the input window size (and therefore in the sensing matrix) and the relative CR is reported in Table 1. The dashed red line at $t = T_W$ in Figure 6 represents the deadline to have

2. Please note that the RSNR value in this use case is different from the HQ definition as in Section 5, where the 25dB are a statistical threshold ($Prob[\text{RSNR} \geq 25 \text{ dB}] \geq 0.97$).

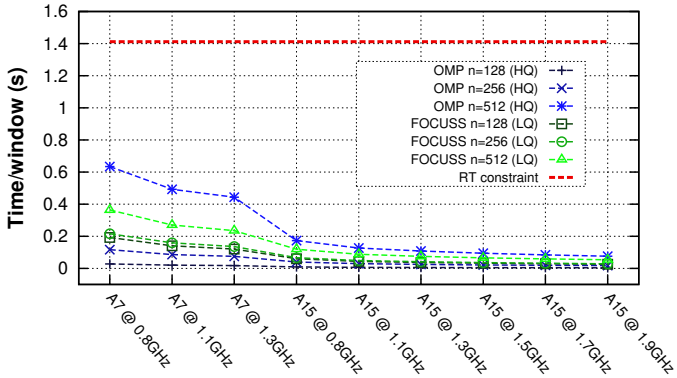

Fig. 6: Time per window for OMP and FOCUSS, respectively targeting HQ and LQ standards, for different input n values.

real-time reconstruction[3]. On the x-axis the plot reports the different operating conditions at which the decoding task was performed. From the plot we can notice that all the algorithms, configurations and operating point respect the real-time constraint. However, the time taken for reconstruction scales up as the frequency decreases and by moving from the A15 core to the A7 core. In addition we can notice that $n = 128$ for both FOCUSS and OMP leads to a lower decoding time compared to $n = 256$ and $n = 512$, thus it represents a preferable setting. Moreover, we can notice that the LQ compression does not always guarantee to have better performance than the HQ compression. Indeed

3. To avoid confusion in the reader, the feasibility of reconstruction with FOCUSS in real-time as in Figure 6 is now due to the LQ standard (7dB), compared to the target 25dB as in Section 6.3

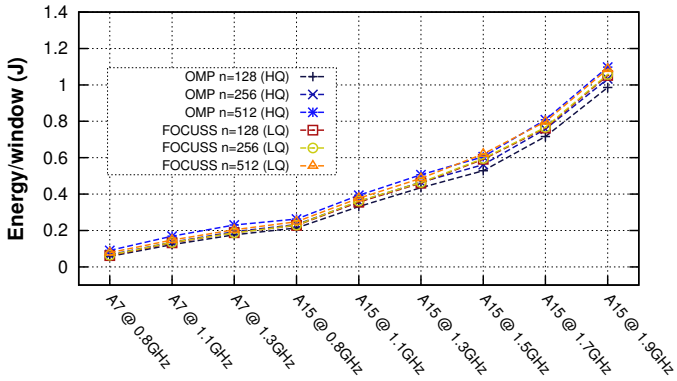

Fig. 7: Energy/window for OMP and FOCUSS, respectively targeting HQ and LQ standards, for different input n values.

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TETC.2016.2564361

		FOCUSS (LQ)			OMP (HQ)		
		$n = 128$	$n = 256$	$n = 512$	$n = 128$	$n = 256$	$n = 512$
CPU usage (%)	A7 @ 0.8 GHz	13.7 %	15.3 %	25.7 %	7.8 %	16.5 %	44.9 %
	A15 @ 1.9 GHz	1.8 %	2.1 %	3.8 %	1.1 %	2.4 %	5.4 %
Energy (J)	A7 @ 0.8 GHz	0.062	0.064	0.078	0.057	0.067	0.090
	A15 @ 1.9 GHz	1.052	1.058	1.087	0.986	1.033	1.098
Battery (h)	A7 @ 0.8 GHz	173.8	168.4	138.2	189.1	160.9	119.8
	A15 @ 1.9 GHz	10.3	10.2	9.9	10.9	10.4	9.8

TABLE 4: CPU usage, total energy for decoding one window and battery lifetime of a WBSN gateway. The CS decoder uses FOCUSS for LQ quality standard and OMP in HQ quality standard for different input window sizes.

while this is true for $n = 512$, the performance gain in using $n = 128$ is larger for OMP than for FOCUSS which suffers more from the lower sparsity, making the HQ case with $n = 128$ most performing than the respective LQ configuration. This suggests that one should not focus only on the CR when trying to optimize the RSNR under a performance constraint.

The plot in Figure 7 shows the energy consumed by the CS decoder to reconstruct a window of T_W seconds of ECG signal in both LQ and HQ configurations. Even if the decoding time increases while scaling the frequency and moving from the A15 core to the A7 core, the energy consumption reduces as well. Moreover, from the same plot we can notice a similar trend as in the previous one when scaling n. Indeed, by moving from $n = 512$ to $n = 128$ the energy decreases as well. As before, for $n = 128$ the setting OMP and HQ configuration achieves the lowest energy consumption even if is linked to a higher QoS than the LQ configuration.

Table 4 summarizes these findings by reporting for all the configurations the overall CPU usage (computed as ratio of the active time over T_W), the total energy for decoding one window and the overall battery lifetime of the CS decoder when considering a 10.78Wh battery for the WBSN gateway (as in the Samsung Galaxy S5). As design guidelines, we reported only the highest and lowest performing operating point. We can notice that the LQ decoding task ranges from almost 27% of CPU usage to 1.8%, while the HQ decoding task ranges from 45% to 1.1%. However, when selecting the optimal setting of $n = 128$ both LQ and HQ configurations are below 2% when running on the A15 at the maximum frequency. This percentage increases to almost 14% for the LQ case and to 8% when running the decoding task on the A7 core at the lower frequency. When looking at the energy consumption the most efficient configuration is with the A7 at the lowest frequency, which represents only the 6% of the cost of decoding on the A15 at the maximum frequency.

This is more visible when observing the battery duration. When the A15 at the maximum frequency is used the CS decoding task by itself leads to a battery duration of only 10h, which is almost constant for all the configurations. Such short duration would not be even enough for a 24h monitoring device. On the other side, if we look at the A7 case, we can notice that the battery duration is increased of more than an order of magnitude, enabling long-time ECG monitoring as well as a low overhead on energy consumption, especially if the gateway is shared among other resource-hungry tasks (i.e. smartphone OS). Moreover, we can notice that if the A7 core is chosen, the impact of the target quality and the choice of n largely affect the battery lifetime compared to the A15 core. As already shown by previous results, OMP (HQ) with $n = 128$ leads to the best configuration extending the battery duration of almost 1.6x w.r.t. to LQ compression with FOCUSS and $n = 512$.

As a matter of fact, the big.LITTLE architecture shows interesting features for ECG personalized gateway devices, allowing the devices to be shared for other functions compatible with modern smarthphones. Indeed, when the A7 is used for the ECG reconstruction task the core is used only at less than the 8% of its computational bandwidth for both HQ and LQ ECG signal reconstruction. It must be noted that this impressive results is directly connected with the rakeness algorithm which significantly reduces the encoding and decoding cost at fixed QoS when compared to standard CS.

7 Conclusions

Modern ultra-low power bio-sensors for continuous vital signs monitoring or lifestyle applications are typically integrated in CS-based WBSN to acquire and process biomedical signals and transmit them to the WBSN gateway for further usage. The recently proposed rakeness-based CS expands the standard CS paradigm deploying the localization of input signal energy to further increase data compression without sensible RSNR degradation. In this paper we focus on the CS decoder, analyzing the real-time capabilities, the energetic cost and the achievable QoS on a ARM big.LITTLE architecture. The experimental results show the effectiveness of the rakeness approach, in terms of reconstruction time and quality against compression. Moreover, we explore the design space on the evaluation board to achieve an energy-aware real-time decoding with the OMP and FOCUSS reconstruction algorithms.

Acknowledgment

This work was supported by the ICYSoC RTD project (no. 20NA21 150939), evaluated by the Swiss NSF and funded by Nano-Tera.ch with Swiss Confederation financing.

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TETC.2016.2564361

REFERENCES

[1] World Health Organization (WHO), "Cardiovascular diseases." http://www.who.int/mediacentre/factsheets/fs317/en, 2015. [Online].

[2] M. Kay, J. Santos, and M. Takane, "mhealth: New horizons for health through mobile technologies," World Health Organization, pp. 66–71, 2011.

[3] D. L. Donoho, "Compressed sensing," Information Theory, IEEE Transactions on, vol. 52, no. 4, pp. 1289–1306, 2006.

[4] F. Chen, A. P. Chandrakasan, and V. M. Stojanović, "Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors," Solid-State Circuits, IEEE Journal of, vol. 47, no. 3, pp. 744–756, 2012.

[5] M. Shoaib, K. H. Lee, N. K. Jha, and N. Verma, "A 0.6–107 μw energy-scalable processor for directly analyzing compressively-sensed eeg," Circuits and Systems I: Regular Papers, IEEE Transactions on, vol. 61, no. 4, pp. 1105–1118, 2014.

[6] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vandergheynst, "Compressed sensing for real-time energy-efficient ecg compression on wireless body sensor nodes," Biomedical Engineering, IEEE Transactions on, vol. 58, no. 9, pp. 2456–2466, 2011.

[7] M. Mangia, R. Rovatti, and G. Setti, "Rakeness in the design of analog-to-information conversion of sparse and localized signals," Circuits and Systems I: Regular Papers, IEEE Transactions on, vol. 59, no. 5, pp. 1001–1014, 2012.

[8] D. Bortolotti, M. Mangia, A. Bartolini, R. Rovatti, G. Setti, and L. Benini, "Rakeness-based compressed sensing on ultra-low power multi-core biomedical processors," in Design and Architectures for Signal and Image Processing (DASIP), 2014 Conference on, pp. 1–8, Oct 2014.

[9] ARM Ltd., "big.little white paper." http://www.arm.com/files/downloads/big_LITTLE_Final_Final.pdf, 2015. [Online].

[10] NVIDIA Corporation, "Tegra white paper." http://www.nvidia.com/content/PDF/tegra_white_papers/tegra-whitepaper-0911b.pdf, 2015. [Online].

[11] Qualcomm Inc., "Snapdragon processors." https://www.qualcomm.com/products/snapdragon/processors/810, 2015. [Online].

[12] D. Bortolotti, M. Mangia, A. Bartolini, R. Rovatti, G. Setti, and L. Benini, "An ultra-low power dual-mode ecg monitor for healthcare and wellness," in Proceedings of the 2015 Design, Automation & Test in Europe Conference & Exhibition, pp. 1611–1616, EDA Consortium, 2015.

[13] D. Bortolotti, H. Mamaghanian, A. Bartolini, M. Ashouei, J. Stuijt, D. Atienza, P. Vandergheynst, and L. Benini, "Approximate compressed sensing: ultra-low power biosignal processing via aggressive voltage scaling on a hybrid memory multi-core processor," in Proceedings of the 2014 international symposium on Low power electronics and design, pp. 45–50, ACM, 2014.

[14] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using focuss: A re-weighted minimum norm algorithm," Signal Processing, IEEE Transactions on, vol. 45, no. 3, pp. 600–616, 1997.

[15] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," Signal Processing, IEEE Transactions on, vol. 53, no. 7, pp. 2477–2488, 2005.

[16] K. Kanoun, H. Mamaghanian, N. Khaled, and D. Atienza, "A real-time compressed sensing-based personal electrocardiogram monitoring system," in Design, Automation & Test in Europe Conference & Exhibition (DATE), 2011, pp. 1–6, IEEE, 2011.

[17] A. M. Dixon, E. G. Allstot, A. Y. Chen, D. Gangopadhyay, and D. J. Allstot, "Compressed sensing reconstruction: Comparative study with applications to ecg bio-signals," in Circuits and Systems (ISCAS), 2011 IEEE International Symposium on, pp. 805–808, IEEE, 2011.

[18] D. Craven, B. McGinley, L. Kilmartin, M. Glavin, and E. Jones, "Compressed sensing for bioelectric signals: a review," Biomedical and Health Informatics, IEEE Journal of, vol. 19, no. 2, pp. 529–540, 2015.

[19] D. Bortolotti, A. Bartolini, M. Mangia, R. Rovatti, G. Setti, and L. Benini, "Energy-aware bio-signal compressed sensing reconstruction: Focuss on the wbsn-gateway," in Embedded Multicore/Many-core Systems-on-Chip (MCSoC), 2015 IEEE 9th International Symposium on, pp. 120–126, Sept 2015.

[20] J. Haboba, M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "A pragmatic look at some compressive sensing architectures with saturation and quantization," Emerging and Selected Topics in Circuits and Systems, IEEE Journal on, vol. 2, no. 3, pp. 443–459, 2012.

[21] F. Pareschi, P. Albertini, G. Frattini, M. Mangia, R. Rovatti, and G. Setti, "Hardware-algorithms co-design and implementation of an analog-to-information converter for biosignals based on compressed sensing," Biomedical Circuits and Systems, IEEE Transactions on, vol. PP, no. 99, pp. 1–1, 2015.

[22] D. Gangopadhyay, E. G. Allstot, A. M. Dixon, K. Natarajan, S. Gupta, and D. J. Allstot, "Compressed sensing analog front-end for bio-sensor applications," Solid-State Circuits, IEEE Journal of, vol. 49, no. 2, pp. 426–438, 2014.

[23] J. Zhang, Y. Suo, S. Mitra, S. P. Chin, S. Hsiao, R. F. Yazicioglu, T. D. Tran, and R. Etienne-Cummings, "An efficient and compact compressed sensing microsystem for implantable neural recordings," Biomedical Circuits and Systems, IEEE Transactions on, vol. 8, no. 4, pp. 485–496, 2014.

[24] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," Comptes Rendus Mathematique, vol. 346, no. 9, pp. 589–592, 2008.

[25] V. Cambareri, M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "A case study in low-complexity ecg signal encoding: How compressing is compressed sensing?," Signal Processing Letters, IEEE, vol. 22, pp. 1743–1747, Oct 2015.

[26] G. Jacovitti, A. Neri, and G. Scarano, "Texture synthesis-by-analysis with hard-limited gaussian processes," Image Processing, IEEE Transactions on, vol. 7, pp. 1615–1621, Nov 1998.

[27] R. Rovatti, G. Mazzini, and G. Setti, "Memory- m antipodal processes: Spectral analysis and synthesis," Circuits and Systems I: Regular Papers, IEEE Transactions on, vol. 56, pp. 156–167, Jan 2009.

[28] A. Caprara, F. Furini, A. Lodi, M. Mangia, R. Rovatti, and G. Setti, "Generation of antipodal random vectors with prescribed non-stationary 2-nd order statistics," Signal Processing, IEEE Transactions on, vol. 62, pp. 1603–1612, March 2014.

[29] J. Van Vleck and D. Middleton, "The spectrum of clipped noise," Proceedings of the IEEE, vol. 54, pp. 2–19, Jan 1966.

[30] J. Tropp, A. C. Gilbert, et al., "Signal recovery from random measurements via orthogonal matching pursuit," Information Theory, IEEE Transactions on, vol. 53, no. 12, pp. 4655–4666, 2007.

[31] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," Applied and Computational Harmonic Analysis, vol. 26, no. 3, pp. 301–321, 2009.

[32] Z. Zhang, Z. Pi, and B. Liu, "Troika: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise," Biomedical Engineering, IEEE Transactions on, vol. 62, no. 2, pp. 522–531, 2015.

[33] G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," Engineering in Medicine and Biology Magazine, IEEE, vol. 20, no. 3, pp. 45–50, 2001.

[34] Physionet.org, "Mit-bih arrhythmia database." http://www.physionet.org/physiobank/database/mitdb, 2015. [Online].

[35] Hardkernel Ltd., "Odroid-xu 3." http://www.hardkernel.com, 2015. [Online].

[36] C. Sanderson, "Armadillo: An open source c++ linear algebra library for fast prototyping and computationally intensive experiments," 2010.

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TETC.2016.2564361

Daniele Bortolotti received the M.S. degree in Electronic Engineering and the Ph.D. degree in Electronics, Computer Science, and Telecommunications from the University of Bologna, Italy in 2010 and 2014, respectively. He is currently a Post-Doctoral Researcher in the Department of Electrical, Electronic and Information Engineering Guglielmo Marconi (DEI) at the University of Bologna. The focus of his research has initially been on virtual platforms, architectural aspects for multi-processors systems-on-chip. Recently his focus comprises HW/SW design strategies for ultra-low power biosensors nodes operating in near-threshold for WBSN applications and low-level power management techniques for many-cores HPC nodes.

Mauro Mangia (S'09-M'13) received the B.S. and M.S. degree in Electronic Engineering from the University of Bologna, Italy, in 2004 and 2009 respectively; he received the Ph.D. degree in Information Technology from the University of Bologna in 2013. He is currently a post-doc researcher in the statistical signal processing group of ARCES, University of Bologna, Italy. In 2009 and 2012 he was a visiting Ph.D. student at the École Polytechnique Fédérale de Lausanne (EPFL). His research interests are in nonlinear systems, compressed sensing, ultra-wideband systems and system biology. He was recipient of the 2013 IEEE CAS Society Guillemin-Cauer Award and the best student paper award at IEEE ISCAS 2011.

Andrea Bartolini received a Ph.D. degree in Electrical Engineering from the University of Bologna, Italy, in 2011. He is currently a postdoc researcher in the Department of Electrical, Electronic and Information Engineering Guglielmo Marconi (DEI) at the University of Bologna. He also holds a postdoc position in the Integrated Systems Laboratory at ETH Zrich. His research interests concern dynamic resource management ranging from embedded to large scale HPC systems with special emphasis on software-level thermal and power-aware techniques. His research interest also includes ultra-low power design strategies for biosensors nodes operating in near-threshold.

Riccardo Rovatti (M'99-SM'02-F'12) received the M.S. degree in Electronic Engineering and the Ph.D. degree in Electronics, Computer Science, and Telecommunications both from the University of Bologna, Italy in 1992 and 1996, respectively. He is now a Full Professor of Electronics at the University of Bologna. He is the author of approximately 300 technical contributions to international conferences and journals, and of two volumes. His research focuses on mathematical and applicative aspects of statistical signal processing and on the application of statistics to nonlinear dynamical systems. He received the 2004 IEEE CAS Society Darlington Award, the 2013 IEEE CAS Society Guillemin-Cauer Award, as well as the best paper award at ECCTD 2005, and the best student paper award at EMC Zurich 2005 and ISCAS 2011. He was elected IEEE Fellow in 2012 for contributions to nonlinear and statistical signal processing applied to electronic systems.

Gianluca Setti (S'89-M'91-SM'02-F'06) received the Ph.D. degree in Electronic Engineering and Computer Science from the University of Bologna in 1997. Since 1997 he has been with the School of Engineering at the University of Ferrara, Italy, where he is currently a Professor of Circuit Theory and Analog Electronics and is also a permanent faculty member of ARCES - University of Bologna, Italy. His research interests include nonlinear circuits, implementation and application of chaotic circuits and systems, electromagnetic compatibility, statistical signal processing and biomedical circuits and systems. Dr. Setti received the 2013 IEEE CAS Society Meritorious Service Award and co-recipient of the 2004 IEEE CAS Society Darlington Award, of the 2013 IEEE CAS Society Guillemin-Cauer Award, as well as of the best paper award at ECCTD 2005, and the best student paper award at EMC Zurich 2005 and at ISCAS 2011. He held several editorial positions and served, in particular, as the Editor-in-Chief for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMSPART II (2006-2007) and of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMSPART I (2008-2009). Dr. Setti was the Technical Program Co-Chair at ISCAS 2007, ISCAS 2008, ICECS 2012, BioCAS 2013 as well as the General Co-Chair of NOLTA 2006. He was a member of the Board of Governors of the IEEE CAS Society (2005-2008), served as its 2010 President, and he is a Distinguished Lecturer of CASS (2015-2016). He held several other volunteer positions for the IEEE and in 2013-2014 he was the first non North-American Vice President of the IEEE for Publication Services and Products.

Luca Benini is full professor at the University of Bologna and he is the chair of Digital Circuits and Systems at ETHZ. He has served as chief architect for the Platform2012/STHORM project in STmicroelectronics, Grenoble in the period 2009-2013. He has held visiting and consulting researcher positions at EPFL, IMEC, Hewlett-Packard Laboratories and Stanford University. Dr. Beninis research interests are in energy-efficient system design and multi-core SoC design. He is also active in the area of energy-efficient smart sensors and sensor networks for biomedical and ambient intelligence applications. He has published more than 700 papers in peer-reviewed international journals and conferences, four books and several book chapters. He is a fellow of the IEEE and a member of the Academia Europaea.

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.