

Energy Analysis of Decoders for Rakeness-Based Compressed Sensing of ECG Signals

*Original*

Energy Analysis of Decoders for Rakeness-Based Compressed Sensing of ECG Signals / Pareschi, Fabio; Mangia, Mauro; Bortolotti, Daniele; Bartolini, Andrea; Benini, Luca; Rovatti, Riccardo; Setti, Gianluca. - In: IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS. - ISSN 1932-4545. - STAMPA. - 11:6(2017), pp. 1278-1289. [10.1109/TBCAS.2017.2740059]

*Availability:*

This version is available at: 11583/2701986 since: 2020-02-05T22:57:02Z

*Publisher:*

Institute of Electrical and Electronics Engineers Inc.

*Published*

DOI:10.1109/TBCAS.2017.2740059

*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

©2017 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)

# Energy Analysis of Decoders for Rakeness-based Compressed Sensing of ECG signals

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

**Abstract**—In recent years, Compressed Sensing (CS) has proved to be effective in lowering the power consumption of sensing nodes in biomedical signal processing devices. This is due to the fact the CS is capable of reducing the amount of data to be transmitted to ensure correct reconstruction of the acquired waveforms. Rakeness-based CS has been introduced to further reduce the amount of transmitted data by exploiting the uneven distribution to the sensed signal energy. Yet, so far no thorough analysis exists on the impact of its adoption on CS decoder performance. The latter point is of great importance, since body-area sensor network architectures may include intermediate gateway nodes that receive and reconstruct signals to provide local services before relaying data to a remote server. In this paper we fill this gap by showing that rakeness-based design also improves reconstruction performance. We quantify these findings in the case of ECG signals and when a variety of reconstruction algorithms are used either in a low-power microcontroller or a heterogeneous mobile computing platform.

## I. INTRODUCTION

**P**ERSONAL biometric monitoring systems is foreseen to be one of the key technologies which will lead to a major breakthrough in improving life quality in coming years. The applications of this technology range from continuous patient monitoring or elders caring, to athletes' training improvement, to stress detection during safety critical tasks. In all cases what is needed is a wireless body sensor network (WBSN), which consists of a set of low-power miniaturized bio-sensing nodes connected to a gateway collecting the signals and processing them [1], [2]. Commonly, the gateway is assumed to be a remote powerful server or a desktop machine. However, with the increasing computing capabilities present in today's smartphones and wearable devices (watches, goggles, etc.), WBSN architectures may entail local gateways that provide a first level of processing with a possible immediate feedback to the user.

F. Pareschi and G. Setti are with the Department of Engineering, University of Ferrara, 44122 Ferrara, Italy, and also with the Advanced Research Center on Electronic Systems, University of Bologna, 40125 Bologna, Italy (e-mail: fabio.pareschi@unife.it; gianluca.setti@unife.it).

M. Mangia and R. Rovatti are with the Department of Electrical, Electronic, and Information Engineering, University of Bologna, 40136 Bologna, Italy, and also with the Advanced Research Center on Electronic Systems, University of Bologna, 40125 Bologna, Italy (e-mail: mauro.mangia2@unibo.it; riccardo.rovatti@unibo.it).

D. Bortolotti is with Brainiac SAS, Paris, France

A. Bartolini and L. Benini are with the Department of Electrical, Electronic and Information Engineering, University of Bologna, 40123, Bologna, Italy, and with the Integrated Systems Laboratory, ETH Zurich 8092, Zurich, Switzerland. (e-mail: barandrea@iis.ee.ethz.ch; lbenini@iis.ee.ethz.ch).

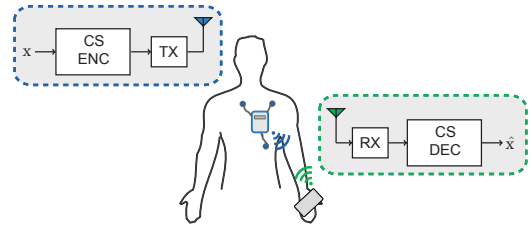


Fig. 1. CS in the link between a sensing node and the local gateway.

As a result, the reception/decoding stages must be accounted for in addition to the encoding/transmission ones in designing the system. This becomes more prominent when the Compressed Sensing (CS) approach is used, as in the typical WBSN scenario of Fig. 1 that highlights the link between a sensing node and the local gateway. In fact, from a power consumption viewpoint, CS is an intrinsically asymmetric method that can dramatically reduce the resources at the sensing node while potentially making reconstruction at the local gateway more expensive [3], [4].

Mathematically, CS is a dimensionality reduction sensing technique that uses a linear, usually random, transformation to map vectors of Nyquist rate samples into smaller vectors of *measurements* that are enough to reconstruct the original signal. The transmitter (the most power hungry stage in the sensing node) benefits from treating a reduced amount of data, while the amount of additional processing (a linear transformation) is small and decreases as the compression ratio increases [5]. At the receiver side, reconstruction is achieved by adopting non-linear, typically iterative, procedures with a computational complexity much higher than the one needed at the encoding stage.

Recently, a CS system improvement has been proposed exploiting the common property of real-world signals to be non-white, i.e., of not distributing their energy uniformly in the signal space [6]. CS can, in fact, be optimized by adapting the statistics of the random linear mapping to such a distribution. The driving concept here is *rakeness*, i.e., the ability of the linear transformation to collect the energy of the signal to acquire. With respect to this, the concept resembles what happens in rake receivers in (chaos-based) DS-CDMA communication, where spreading sequences, the corresponding waveforms, and rake receiver taps, are jointly selected to collect as much energy as possible at the received side [7] [8]. As such, by adopting a rakeness-based design flow, one

increases the amount of information that each measurement carries about the original signal, thus reducing (in some cases drastically) the number of measurements to be transmitted and therefore the power required by the transmitter.

Several works have appeared in the recent literature proposing low-power CS encoders with particular interest in biomedical applications [9]. Commonly, Electroencephalographic (EEG) [10], [11] or Electrocardiographic (ECG) signals [12], [13] are considered. However, a very limited number of works can be found proposing energy considerations on the decoder. In [14] a reconstruction algorithm designed ad-hoc for ECG signals has been optimized and hard-coded in a properly designed 90 nm application-specific circuit.

In both [15] and [16] few considerations on decoder trade-offs are presented for some common reconstruction algorithms when the rakesness approach is used. In particular, [15] focuses on the impact of the rakesness-based CS for three greedy algorithms and for three convex optimization based algorithms. In addition, authors evaluate the power consumption of one of the tested greedy algorithms, namely FOCUSS, in different design limit configurations. The main topic discussed in [16] is the impact of the time window duration on the computational cost of some greedy decoding algorithms for certain target qualities of services. In [16] an analysis of the power cost and conditions for satisfying the real-time constraints for OMP and FOCUSS is also provided, thus confirming how the window length affects the decoder performance. Both [15] and [16] use a reference mobile platform for power consumption estimation.

This paper is an extension of [15] and [16] and proposes an analysis of the energy necessary for decoding ECG signals in a CS system assuming a general-purpose computing architecture. Differently from both our previous works, we consider here only iterative decoding approaches with low computational requirements, since this represents the most natural choice when energy or computational power at the decoder side is an issue. In particular, we investigate the power profile of each decoder to obtain power models that pave the way for new trade-offs. In detail, we compare three algorithms, namely OMP [17], CoSaMP [18] and IHT [19] where we develop:

- develop a complete trade-off analysis with respect to number of measurements, number of iterations, standard or rakesness-based CS approach and reconstruction quality. Such analysis could be used as design flow for different biosignals;
- introduce an innovative decoding approach merging low-complexity requirements and prior information on the acquired class of signals properly specialized for ECGs;
- define power models with respect to number of measurements, number of iterations, standard or rakesness-based CS approach and reconstruction quality for a low-cost ARM architecture (Cortex-M4F);
- validate the proposed power models on a high-end heterogeneous multicore big/LITTLE platform designed for mobile applications.

By combining data on reconstruction quality and energy requirement we conclude that the choice of OMP is the optimum

trade-off between quality and energy, and that the rakesness approach is beneficial not only at the encoder but also at the decoder, allowing to reconstruct a signal with a chosen target quality by using a lower amount of energy.

The rest of the paper is organized as follows. Section II quickly recaps the CS mathematical background including details on the rakesness approach and on some decoding algorithms. Section III illustrates both the standard and the rakesness approaches impact when synthetic ECGs and the considered algorithms are used. In Section IV, real ECG signals and more advanced decoding approaches are taken into account. Power models in terms of energy requirements when decoding algorithms are executed on two different platforms are presented in Section V. Finally, we draw the conclusion.

## II. BASICS OF COMPRESSED SENSING

In this paper CS is considered in its discrete-time formulation. As such, signals are represented by samples at Nyquist-rate  $1/T$ . CS theory is developed referring to a chunk of  $n$  consecutive samples, i.e., over a time window of length  $nT$  of the original continuous time signal. Longer waveforms need to be split into chunks of  $n$  samples defined over different time windows. Without loss of generality, we consider  $0 \leq t \leq nT$ , and the input signal  $x = (x_0, \dots, x_{n-1})^T \in \mathbb{R}^n$ , where  $^T$  denotes vector transpose.

The key assumption behind the application of the CS paradigm is that  $x$  is *sparse*. Mathematically,  $x$  is  $\kappa$ -sparse if given a proper  $n$ -dimensional sparsity basis  $\Psi \in \mathbb{R}^{n \times n}$ , then  $x = \Psi\alpha$  where  $\alpha \in \mathbb{R}^n$  have at most  $\kappa \ll n$  non-zero components. The  $\kappa$  non-zero components of  $\alpha$  are referred to the signal support. As one may expect, CS can also be applied when  $\kappa \ll n$  components of  $\alpha$  are significant, while the other are negligible, i.e., when the sparsity condition only roughly holds, as for many classes of real-world signals. In this case we will indicate the signal  $x$  as compressible.

It is clear that for both sparse and compressible signals, the amount of information carried by  $x$  is better estimated by  $\kappa$  than by  $n$ . Fundamental results [3] show that the main information content of  $x$  can be captured with of  $m < n$  *measurements* achieved by a *linear projection* of  $x$  by means of a sensing matrix  $A \in \mathbb{R}^{m \times n}$ , i.e.,

$$y = Ax + \nu = B\alpha + \nu \quad (1)$$

where the  $m$ -dimensional vector  $y = (y_0, \dots, y_{m-1})^T \in \mathbb{R}^m$  is introduced to collect all  $m$  scalar measurements and  $B = A\Psi \in \mathbb{R}^{m \times n}$  is the operator that maps the collected measurements with the sparse representation. The term  $\nu \in \mathbb{R}^m$  is used to take into account all possible nonidealities of this process such as noise or quantization.

It is possible to show that, despite the fact that  $A$  (and thus  $B$ ) is a dimensionality reduction operator,  $\alpha$  (and thus  $x$ ) can be fully recovered from  $y$  when  $x$  is sparse (and of course approximately recovered with very high accuracy when  $x$  is compressible) [3], [4]. Roughly speaking, the rationale behind this is that generic,  $\kappa$ -sparse vectors are mapped *almost isometrically* [20] into the measurements; if this is true, the recovery of the original signal  $x$  from  $y$  is possible by enforcing the *a priori* knowledge that its representation is sparse.

In detail,  $x$  is reconstructed as  $\hat{x} = \Psi\hat{\alpha}$ , where  $\hat{\alpha}$  is the sparsest coefficient vector subject to the constraint that the corresponding measurements are as close as possible to the observed ones. In other words, the reconstruction task is equivalent to the solution of a linear ill-posed problem, and CS theory suggests to consider the sparsest vector  $\alpha$  mapped into  $y$  by the operator  $B$  with a proper tolerance due to measurements nonidealities [3]. Sparsity is generally promoted by the  $\ell_1$  norm instead of the computationally intractable count of non-zero components given by  $\ell_0$  norm. Such approach is called basis pursuit with denoising (BPDn) and solves

$$\min_{\alpha} \|\alpha\|_1 \quad \text{s.t.} \quad \|B\alpha - y\|_2 \leq \epsilon \quad (2)$$

where  $\|\cdot\|_p$  indicates the usual  $\ell_p$  norm and  $\epsilon$  is tuned on the characteristics of the disturbance term  $\nu$ .

One of the most interesting properties of CS is that reconstruction is guaranteed when  $A$  is composed by instances of Gaussian (or Sub-Gaussian) random variables when a minimum number of measurement is ensured, i.e.,  $m = O(\kappa \log n)$ .

Intriguingly,  $A$  can also be made only of *antipodal* symbols, i.e.,  $A \in \{-1, +1\}^{m \times n}$ . This constraint is of paramount importance as it allows hardware-friendly architectures, where expensive and cumbersome full multipliers are not required anymore, and represents a key point in the design of effective and parsimonious CS stages for biomedical sensing nodes [13]. For this reason, in this manuscript we assume that the projection matrix  $A$  is antipodal.

#### A. Improving CS Performance: The Rakeness Approach

CS performance depends on  $m$ : the higher  $m$ , the larger the amount of information available, the easier the task of retrieving  $x$  from  $y$ . Yet,  $m$  is also related to the *compression ratio*  $CR = n/m$  and thus to the saving (in terms of energy, bandwidth, etc.) that one may achieve when considering the CS system performing the compression/encoding stage, i.e., computing and transmitting the measurements vector  $y$ . According to this point of view, the minimization of  $m$  (given a target reconstruction quality) is a key factor.

This is what a rakeness-based design [6] does: it improves sensing performance by generating each row of  $A$  independently of each other, but with entries whose correlation is adapted to the second-order statistic of the input signal  $x$  with the aim of increasing the expected energy of the generic  $j$ -th entry of  $y$ .

In other word, the proposed approach uses a set of  $m$  rows  $a_j$ ,  $j = 1 \dots m$  of  $A$  that increases the average energy collected (“raked”) by  $y$ . Such property is measured by the *rakeness*  $\rho$  defined as

$$\rho(a, x) = \mathbf{E}_{a,x} \left[ (a^\top x)^2 \right] = \text{tr}(C^a C^x)$$

where  $\text{tr}(\cdot)$  stands for matrix trace while  $C^a = \mathbf{E}[aa^\top]$  and  $C^x = \mathbf{E}[xx^\top]$  are two  $n \times n$  correlation matrices for the generic row of the sensing matrix and for the input signal. Moreover,  $\mathbf{E}_{a,x}$  stands for the expectations with respect to both vectors.

The approach, introduced and described in [6], maximizes  $\rho$  preserving at same time the isometric property of the projection under the assumption that  $x$  is not only sparse but also *localized*, i.e. its energy content is not uniformly distributed across the whole signal domain. This case is the most common one when dealing with real world signals [13]. The maximization of  $\rho$  has as output the correlation matrix  $C^a$  that identifies the stochastic process to be used for generating sensing vectors  $a_j$ .

Interestingly, the rakeness-based design is compatible with the hardware-friendly constraint of having  $A$  made only of antipodal symbols [21]<sup>1</sup>.

Many different approaches are possible to generate antipodal sequences with a proper correlation profile. The simplest one relies on thresholding of Gaussian random vectors [22], [23]. In fact, given a  $n \times n$  matrix  $C^G$ , used to generate  $m$  Gaussian vectors with zero mean and correlation matrix  $C^G$ , and defined as

$$C^G = \sin\left(\frac{\pi}{2} C^a\right)$$

where the equality is meant componentwise, then the  $a_j$  are simply obtained by computing the sign of the elements of these Gaussian vectors. Though not completely general<sup>2</sup> this approach is very simple, and suitable for offline generation of the  $A$  (i.e., generated and stored into a local memory). This is the approach used in this paper.

Conversely, in cases where online generation of a stream of antipodal symbols with a given correlation profile is necessary, the so-called linear probability feedback generator [24] can be used.

Note that the sensing matrix design based on the rakeness approach is completely different from others presented in the literature [25]–[27], where optimized sensing matrices are obtained by a proper specialization of  $A$  aiming to reduce as much as possible the *mutual coherence* between the columns of  $A\Psi$  without any hypothesis on input signal statistics.

As discussed, the rakeness approach imposes randomness to rows of  $A$  to guarantee an acceptable *mutual coherence* with every possible  $\Psi$  and, with this constraint, to increase the raked signal energy as much as possible with a non negligible improvement in the fidelity between original and reconstructed signals. Hence, methods in [25]–[27] adopts sensing to the sparsity basis but not to the statistics of the signal to acquire. This is the main reason why we may expect higher performance with rakeness-based CS with respect to other signal agnostic techniques, at least when localization is strong enough. For the ECG signals, in [21] it is also proved that rakeness-based CS outperforms the approach discussed in [27] which, similar to our, is suitable for binary sensing matrices. Others techniques (like [25], [26]) have a further limitation, both approaches are not compatible with the antipodal symbol constraints, thus requiring full multipliers at the encoder side.

Note that many other different CS optimization techniques properly designed on a class of signals have been recently

<sup>1</sup>Matlab code to implement rakeness-based CS is online available at <http://cs.signalprocessing.it/>

<sup>2</sup>The method does not guarantee that a process can be generated for each feasible correlation matrix  $C^a$ .

proposed. All of them work at the decoder side and are based on a proper tuning of  $\Psi$  on the specific input signal. Even if they can increase CS performance [28], they require a noteworthy hardware complexity and are substantially out of scope of our contribution. Nevertheless a comparison with one of them with the introduction of a new low-computational cost decoder paradigm is presented in Section IV.

### B. Reducing Reconstruction Costs: the Greedy Approach

As already mentioned in the introduction, one of the most important properties of the CS is the possibility to reduce costs (particularly, in term of energy) at the encoder side with respect to standard Nyquist-rate acquisition. This is however counterbalanced by an increase of the reconstruction/decoding cost, as solving the BPDn problem in (2) is a computationally hard problem.

Therefore, when the energy consumption of the decoder is an issue, instead of adopting general methods relying on sparsity promotion by means of the  $\ell_1$  norm, one relies on other greedy approaches that iteratively promote sparsity by observing intermediate and approximate solutions. The latter in fact, despite being less rigorous than general methods, ensure a much lower complexity and hence lower computational costs.

Note that, since all iterative algorithms try to refine the approximate solution at each iteration, the reconstruction quality is expected to be dependent also on the total number of iterations  $\bar{i}$ . The higher  $\bar{i}$ , the closer the intermediate solution to the (real) asymptotic one, but also the higher time and energy required by the algorithm.

In this paper we analyze three among the most common iterative approaches and evaluate their performance in terms of energy required to reconstruct an ECG signal at a given quality, when acquisition is performed both with the standard and with the rakesness-based CS approach. More Specifically, the algorithms we focus on are:

- *Orthogonal Matching Pursuit* (OMP), whose application is described in [17];
- *Compressive Sampling Matching Pursuit* (CoSaMP) [18], which is a variant of OMP with increased convergence guarantees;
- *Iterative Hard Thresholding* (IHT), which is originally proposed and discussed in [19]; we here consider its normalized and more stable version described in [29].

The working principle of all algorithms is similar, and can be described as follows. At the step  $i - 1$ , let us indicate with  $\hat{\alpha}_{i-1}$  the approximate intermediate solution, and with  $r_{i-1} = B\hat{\alpha}_{i-1} - y$  the residual error vector between actual and estimated measurements. The following step generates  $\hat{\alpha}_i$  by modifying some entries of  $\hat{\alpha}_{i-1}$  in a way that minimizes  $r_i$ , assuming that  $r_i \rightarrow 0$  as  $i$  increases. In all considered cases,  $\hat{\alpha}_0 = 0$ , so that  $r_0 = -y$ .

The main difference among the three algorithms is how  $\hat{\alpha}_i$  is computed from  $\hat{\alpha}_{i-1}$ . In more details, OMP identifies at each iteration a new element in the support of  $\hat{\alpha}$ . At the  $i$ -th iteration the  $j$ -th column of  $B$  that is most strongly correlated with  $r_{i-1}$  is added as the  $i$ -th column of an auxiliary matrix

$\Phi_i$ , and the index  $j$  is included into the support of  $\hat{\alpha}_{i-1}$  to generate the support of  $\hat{\alpha}_i$ . The values of the  $i$  non-zero elements of  $\hat{\alpha}_i$  are computed to minimize  $r_i$  as a solution of a least squares problem. Since  $r_i$  is always orthogonal to  $\Phi_i$ , this minimization step can be solved with marginal cost by using the modified Gram-Schmidt algorithm exploiting a companion system regulated by the orthonormalized matrix  $\hat{\Phi}_i$  that is constructed step by step along with the  $\Phi_i$ . With this approach, however, the solution of a full least squares minimization involving the non-orthonormalized  $\Phi_i$  is required for the final step in order to find the actual  $\hat{\alpha}_{\bar{i}}$ . Note that, in order to ensure the existence of the solution of the minimization problem, it is necessary that  $i < m$  at each step, and thus that  $\bar{i} < m$ . A more detailed description of the algorithm can be found in [17].

The CoSaMP and IHT algorithms work in a slightly different way. Under the assumption that the signal to reconstruct is  $K$ -sparse, with  $K$  approximating  $\kappa$ , they compel  $\hat{\alpha}_i$  to have only  $K$  non-zero elements. To this aim, in CoSaMP, an auxiliary matrix  $\Phi_i$  is built at the  $i$ -th step by collecting the  $K$  columns of  $B$  corresponding to the support of  $\hat{\alpha}_{i-1}$ , and by the  $\Delta K$  columns that are more strongly correlated to  $r_{i-1}$ , with typically  $\Delta K = K$  or  $\Delta K = 2K$ . The number of columns of  $\Phi_i$  at each step is signal dependent, and ranges from  $\max(K, \Delta K)$  to  $K + \Delta K$ . The support of  $\hat{\alpha}$  is computed as the elements corresponding to the columns of  $\Phi_i$ . The values of non-zero elements of  $\hat{\alpha}_i$  are computed to minimize  $r_i$  as a solution of a least squares problem, and only the most significant  $K$  are retained. Note that, similarly as in OMP, the number of columns of  $\Phi_i$  has to be smaller than the number  $m$  of rows to allow the least squares problem to have a solution. To ensure this in the worst case, it is mandatory that  $K + \Delta K < m$ . No final step is required.

The working principle of IHT is based on the iterative function  $\hat{\alpha}_i = H_K(\alpha_{i-1} + \mu B^T(y - B\hat{\alpha}_{i-1}))$ , as described in [19]. The non-linear function  $H_K(\cdot)$  is a *hard thresholding* function that returns a vector where only the  $K$  larger elements are preserved, and all others are zeroed. This algorithm is very easy from the computational complexity point of view. The choice of the step size  $\mu > 0$  is, however, critical for the convergence. We consider the version proposed in [29], also known as *normalized* IHT, that despite presenting a higher complexity, is capable of computing  $\mu$  at each step, thus being able to guarantee the convergence of the method and also to increase convergence speed.

In the next section we investigate how efficiently these three algorithms can be used as reconstruction strategies in a CS-based ECG acquisition system. In particular, we want to relate performance with the values of  $m$  and  $\bar{i}$  both when standard and in a rakesness-based CS are employed.

## III. EXPERIMENTAL SETTING AND RESULTS

To compare the performance of the reconstruction algorithms MATLAB Montecarlo simulations have been performed. The simulation setup is as follows.

A synthetic ECG generated as in [30] with average beat-rate equal to 60 bpm is considered as input signal. ECGs are

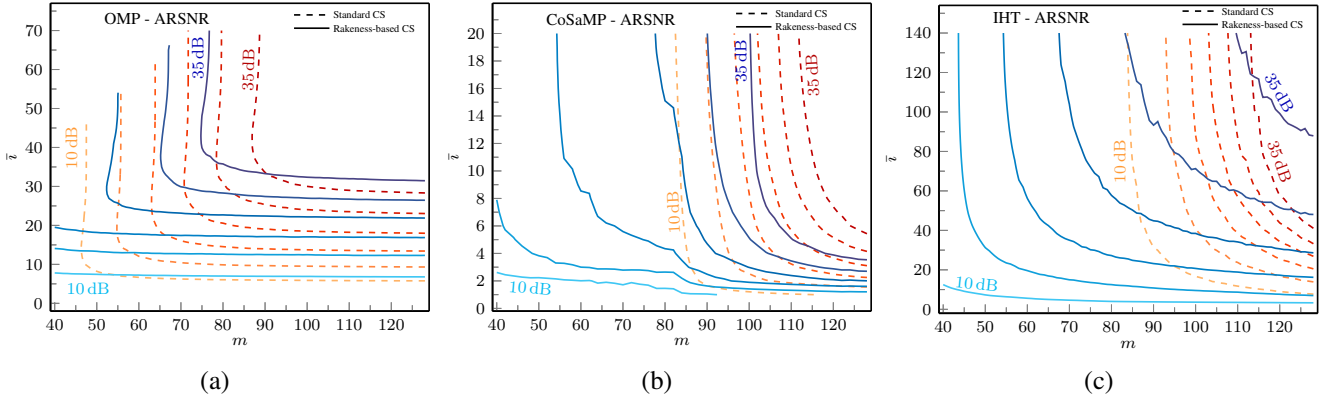


Fig. 2. Performance in terms of ARSNR of the considered greedy algorithms in reconstructing the synthetic ECG signal for different values of number of measurements  $m$  and iterations  $\bar{l}$ . (a): OMP; (b): CoSaMP; (c): IHT. In all contour plots we have highlighted reconstruction qualities ranging from 10 dB to 35 dB with 5 dB steps.

sampled at a rate equal to  $1/T = 360$  Hz, and quantized by using 11 bits (i.e., 2048 levels), as this appears to be a common setup adopted in many ECG online repositories [31]. The reason to use a synthetic ECG instead of a real one is to start from a noiseless signal so that reconstruction performance can be accurately assessed. In fact, the unknown and unavoidable noise present in any real signal will limit reconstruction performance to a level that is strongly dependent on the particular signal. In any case, in order to verify the correctness of the approach, real signals from [31] will be considered in Sections IV and V.

In the Montecarlo simulation, 1000 different instances of length  $n = 256$  samples (corresponding to a time window length of  $nT = 711.1$  ms) of the synthetic ECG has been generated. The value  $n = 256$  has been selected according to [16], since this appears to be a good trade-off between reconstruction complexity and reconstruction quality. Each input signal instance has been encoded twice. First, 1000 different binary antipodal matrices  $A$  have been used, generated in purely random way (standard CS). Then, the same input signal instances have been coded also by using 1000 different sensing matrices generated by using the Gaussian random vector thresholding technique described in Section II-A (rakeness-based CS). Finally, performance has been evaluated by reconstructing the input signal instances using all three considered algorithms. For the sake of simplicity, a fixed sparsity basis has been used, i.e.,  $\Psi$  has been assumed equal to the orthonormal Symlet-6 wavelet basis [32].

For all considered approaches, the performance is computed as a function of the value of  $m$  and of the number of iterations  $\bar{l}$ . Note that CoSaMP and IHT require additional parameters related to the sparsity of the reconstructed signal, i.e.  $K$  and  $\Delta K$  (the latter used only in CoSaMP). We want to recall here that the ECG actually belongs to the class of compressible signals, so it is not possible to univocally define a value of  $\kappa$ .

Empirically, we found that the value  $K = \Delta K = 40$  ensure optimal reconstruction in all considered cases, therefore in IHT we use  $K = 40$  while in CoSaMP, to satisfy the  $K + \Delta K < m$  requirement, we chose  $K = \Delta K = \min(40, \lfloor m/2 \rfloor - 1)$ .

As a quality indicator, we use the average reconstruction

signal to noise ratio (ARSNR), defined as

$$\text{ARSNR} = \mathbf{E}_{A,x} \left[ \left( \frac{\|x\|_2}{\|x - \hat{x}\|_2} \right)_{\text{dB}} \right]$$

where  $\mathbf{E}_{A,x}$  implies averaging over all considered  $A$  and all considered instances  $x$  of the ECG signal in the Montecarlo runs. Practically speaking, for each couple  $(m, \bar{l})$ , the ARSNR has been computed by averaging results related to 1000 couples  $(x, A)$ .

Results are shown in Fig 2. We have limited the  $m$  range to reasonable values, i.e.,  $40 \leq m \leq 128$  to ensure that compression ratio  $CR \geq 2$  ( $CR = 2$  for  $m = 128$ ) and reconstruction is always guaranteed ( $m \geq 40$ ). The range of  $\bar{l}$  depends on the specific reconstruction algorithm, with the additional constraint  $\bar{l} < m$  for OMP. In all contour plots we have highlighted reconstruction qualities ranging from 10dB to 35dB with 5dB step.

The reconstruction quality is increasing with the number of measurements  $m$  and with number of iterations  $\bar{l}$ . When considering large values of  $\bar{l}$  (so, when the solution has almost reached the final asymptotic one), OMP gives rise to better reconstructions with respects to both CoSaMP and IHT. Given  $m$ , it is always possible to achieve a better ARSNR, or conversely, a target ARSNR is achievable with a smaller  $m$ . This may be intuitively explained by considering that the ECG is actually a *compressible* signal, not a *sparse* one. While CoSaMP and IHT have a constraint given by  $K$ , OMP increases the support of  $\hat{\alpha}_i$  with  $i$ , automatically adapting the solution to the different input signal instance. Convergence is achieved for values of  $\bar{l}$  between 30 and 40, thus confirming the estimation of the average sparsity of the ECG signal. This however leads to a slow convergence rate. CoSaMP requires a much smaller  $\bar{l}$  since its first iteration considers signals whose support size is  $K$ , while OMP needs  $K$  iterations for this. Conversely, the required  $\bar{l}$  for IHT, due to its simplicity, is much higher.

As expected and already observed in previous papers [5], [6], [13], when considering the asymptotic solution, the rakeness approach outperforms the standard one in all considered cases. The lower the  $m$ , the higher the advantage of the rakeness approach.



However, we can observe that in some configuration the standard CS approach has a faster convergence rate, i.e., when  $m$  is large enough, the convergence is faster with respect to rakesness-based CS since  $\bar{\tau}$  is lower.

A few reasonable questions follow from these observations:

- The comparison is made in terms of  $m$  and  $\bar{\tau}$ . However, this does not allow a *fair* comparison of the three algorithms, since each of their basic step has a very different computational cost. Is it possible to add the required energy to the comparison?
- The same performance level can be achieved either with large  $m$  and small  $\bar{\tau}$  (where standard approach is preferable) or with small  $m$  and large  $\bar{\tau}$  (rakesness preferable). When considering the encoder side, typical guidelines suggest to minimize  $m$  (and thus, the amount of information that must be transmitted) for the energy optimization. Is this still true when considering the decoding side and greedy algorithms?
- Only generic greedy reconstructing algorithms have been taken into account, i.e., no particular assumptions on the class of decoded signals are made so that the same design flow is suitable for any class of biosignal. How does performance change when a state-of-the-art reconstruction algorithm specifically tuned for ECG signals is considered?

The last question is considered in Section IV, where approaches specialized in reconstructing ECG signals are introduced. In order to answer the first two questions, Section V develops an energy model for the considered decoding.

#### IV. ADAPTED DECODER FOR ECG

The hypothesis of using a generic greedy algorithm is useful to maintain the computational complexity as low as possible. However, this could not represent the optimum in terms of reconstruction performance. To cope with this, different approaches specialized to a proper class of biosignals were presented so far in the literature. As an example, to increase reconstruction performance in decoding EEG signals, in [28] authors introduced an innovative approach based on dictionary learning and on an additional pre-processing stage that implements a technique, named Sapiros optimization, able to decrease *mutual coherence* between the columns of  $A\Psi$ .

For ECG decoding, in [33] authors use block sparsity hypothesis to develop a decoder, the Block Sparse Bayesian learning (BSBL), able to correctly decode signal details for Fetal ECG applications. The approach presented in [14] uses statistic characterization of the sparse representation of ECGs to define an initial support in the OMP initialization, along with a full custom hardware implementation. In [27] authors propose a decoding algorithm for ECGs, the Weighted  $\ell_1$  Minimization (WLM), based on prior information on the statistic characterization of the wavelet coefficients. Note that the adoption of a decoder properly specialized on the ECG reconstruction does not imply that the rakesness-based CS is useless. As proved in [21] for BSBL and WLM, using an adapted sensing matrix following the rakesness design-flow further increases the performance of a properly specialized decoding stage.

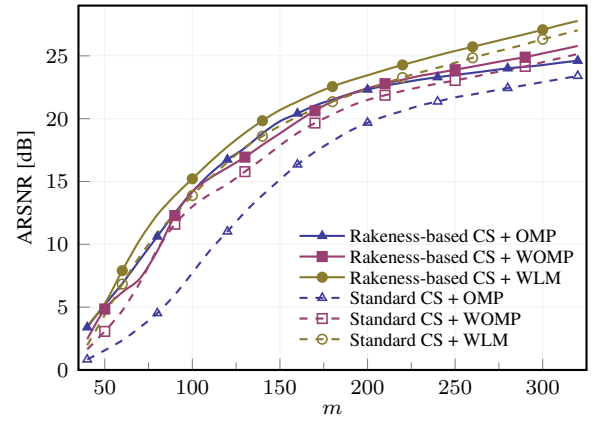


Fig. 3. Performance in terms of ARSNR of the WLM, OMP and of the proposed WOMP in reconstructing record number 100 of the MIT-BIH Arrhythmia on-line database as functions of the number of measurements needed to encode a time window composed by  $n = 512$  samples.

TABLE I  
VALUES OF  $m$  TO RECONSTRUCT ECGs BY EITHER OMP OR WOMP WITH SOME TARGET ARSNRS AND FOR BOTH STANDARD AND RAKENESS-BASED CS, INCLUDING THE % SAVING CAUSED BY RAKENESS-BASED CS FOR BOTH DECODING ALGORITHMS.

	target ARSNR		
	15 dB	20 dB	25 dB
Standard CS + WOMP	123	174	316
Rakesness-based CS + WOMP	108	162	293
Standard CS + OMP	150	207	—
Rakesness-based CS + OMP	106	153	—
% saving (WOMP)	12.2	6.9	7.3
% saving (OMP)	29.3	26.1	—

Among the mentioned approach, we focus here on WLM. As shown in [27], it is capable of outperforming IHT, OMP, BSBL and other decoding approaches. In more details, the WLM decoder aims at reconstructing ECG signals by solving the following optimization problem.

$$\hat{\alpha} = \arg \min_{\alpha} \frac{1}{2} \|B\alpha - y\|_2^2 + \lambda \|W\alpha\|_1$$

where  $W$  is an  $n \times n$  diagonal matrix whose entries are related to the probability of each wavelet function to contribute to the reconstruction of an ECG signal,  $\lambda$  is a normalization value (set to 0.1 according to authors' suggestion), while the signal is reconstructed as  $\hat{x} = \Psi\hat{\alpha}$ .

Using the settings proposed in [27] as reference, i.e., taking signals instances from the record 100 of the MIT-BIH Arrhythmia database [31] and  $n = 512$ , we show in Figure 3 performance of the WLM in terms of ARSN for different value of  $m$  compared with OMP performance in the same setting. Results are obtained with montecarlo simulations averaged over 1000 signal instances. For each signals two randomly generated sensing matrices were adopted, one according to standard CS and the other following rakesness-based CS, to take into account both approaches.

As expected, WLM guarantees a non negligible improvement in signal reconstruction with respect to OMP. Furthermore, the gain is maximized by simultaneously adopting the

rakeness-based sensing at the encoder stage. Nevertheless, the approach requires the solution of an optimization problem that does not match our need to keep the decoding complexity as low as possible. This motivates the introduction of a new optimization procedure that merges the low complexity property of OMP with the adoption of prior information on the considered class of signals as in WLM. We refer to this new decoding algorithm as Weighted OMP (WOMP).

The main difference between WOMP and OMP is how a new element in the support is identified at each iteration. As described before, at the  $i$ -th iteration of OMP the  $j$ -th column of  $B$  that is most correlated with the residual vector  $r_{i-1}$  is added in the support of the decoded signal, i.e.,  $j$  is the position of the maximum absolute value of the vector  $B^\top r_{i-1}$ . In the WOMP approach the decision is weighted by the aforementioned matrix  $W$  so that  $j$  is the index of the element of  $(BW)^\top r_{i-1}$  with the maximum absolute value. The performance in terms of ARSNR as a function of  $m$  is shown in Figure 3. With respect to OMP, WOMP increases the ARSNR for all considered  $m$  values when the standard CS is considered, while in case of the rakeness-based CS, WOMP outperforms OMP only for the highest considered  $m$  values, so that it could be taken into account when requirements in terms of reconstructed signals quality are very strict. Same results can be observed in Table I that reports for both greedy algorithms the minimum  $m$  needed to reach some target ARSNRs. The table also shows the percentages of saving in the number of measurements. As it can be noted, significant reduction of  $m$  are possible when Rakeness-based CS replaces standard-CS.

As final remark, from the computational complexity point of view, WOMP introduces negligible additional cost with respect to OMP. The difference is only in the selection of the most correlated column of  $B$ . This comes by looking at the  $n \times 1$  vector given by the matrix multiplication  $B^\top r_{i-1}$  for OMP, and by  $(BW)^\top r_{i-1} = W^\top B^\top r_{i-1}$  for WOMP. Therefore, in a clever implementation of WOMP, the only additional cost with respect to OPM is the multiplication by the  $n \times n$  diagonal matrix  $W^\top$ , that is quantifiable in the storage of  $n$  values, and  $n$  multiplications at each iteration step. For this reason we can assume that any power model developed for OMP fits also for the WOMP approach.

## V. RECONSTRUCTION COST EVALUATION

A version of OMP, CoSaMP and IHT has been implemented in C language and run in two different low power platforms to evaluate reconstruction energetic costs.

More specifically, the algorithms have been written using single precision (32-bit) floating point variables. For all considered algorithms, almost all operations are simple vector/vector or vector/matrix multiplications. The only non-standard operation required is the solution of a least squares problem in OMP (final step) and CoSaMP (every step), that has been implemented by computing a pseudoinverse. In detail, assuming the system to be minimized in the canonical notation  $Z\zeta = b$ , we solve  $Z^\top Z\zeta = Z^\top b$  by Gauss-Jordan elimination relying on precomputed matrix  $Z^\top Z$  and vector  $Z^\top b$ .

The three algorithms have been considered both from a computational point of view and from a memory occupation point of view.

Since  $\hat{\Phi}_i$  increase of size at each step, computational complexity of the  $i$ -th step of OMP is increasing with  $m$  and  $i$ . These iterations are expected to be simple, while the final step, requiring a least squares problem solution, is expected to be computationally much more complex. The complexity of a CoSaMP step depends on  $m$  and also on  $K$  and  $\Delta K$ , but it is clearly independent of  $i$ . Furthermore, a weak signal dependency is now present due to the unpredictability of the number of columns of  $\Phi_i$ . Compared to OMP, it is reasonable to assume that the complexity of a single iteration is higher (at least, for reasonable values of  $i$ ), as it requires the solution of a least squares problem. Yet, it has been also already observed that CoSaMP has a faster convergence. Finally, IHT is a very simple algorithm whose step complexity is expected to depend only on  $m$ . However, in the implementation proposed in [29], a weak signal dependence in the computation of  $\mu$  is present.

We can also consider some memory occupation aspects of the algorithms, since in low-cost devices the amount of memory is usually limited. Neglecting  $B$  that can be reasonably assumed constant and stored into a non-volatile memory, the random access memory allocation in OMP is dominated by  $\hat{\Phi}_i$ , whose size  $m \cdot i$  increases each iteration up to  $m \cdot \bar{i}$ . Note that it is not necessary to reserve a full memory space for  $\Phi_i$  as it is simply composed of some columns of  $B$ , and can be retrieved by means of a column index array. Memory occupation in CoSaMP is dominated by the least squares problem that requires at least a temporary matrix whose size is, assuming a worst case scenario,  $(K + \Delta K) \times (K + \Delta K)$ , while as in OMP the  $\Phi_i$  does not require memory allocation. IHT is a very interesting algorithm from the memory requirements point of view, as no intermediate matrices are required during the computation, making this approach particularly suitable in a low-resources environment.

### A. ARM Cortex M4F

The first considered platform is an EK-TM4C1294XL evaluation board developed by Texas Instruments [34]. This board is designed to evaluate the low-power low-cost TM4C1294NCPDT microcontroller from Texas Instruments, which embeds an ARM Cortex-M4F CPU, a single-precision floating point unit, 256 kB RAM and 1 MB Flash ROM, and can work with a clock up to 120 MHz. The microcontroller requires a single 3.3 V power supply, and the board has a probe point to be used for the direct measurement of the current consumption of the microcontroller core.

The main limitation of this platform is the available memory size. In a 256 kB RAM, in fact, it is possible to store just a  $256 \times 256$  matrix in single precision floating point representation. By comparing parameters used in Section III (i.e.,  $n = 256$ ,  $m \leq 128$  and  $\bar{i} \leq 72$ ) and the memory requirements of the three algorithms accordingly to Section II-B, the embedded RAM is barely enough to store temporary matrices and vectors, and the  $m \times n$ -size  $B$  matrix. Despite being not necessary, moving  $B$  to RAM is convenient to speed-up the



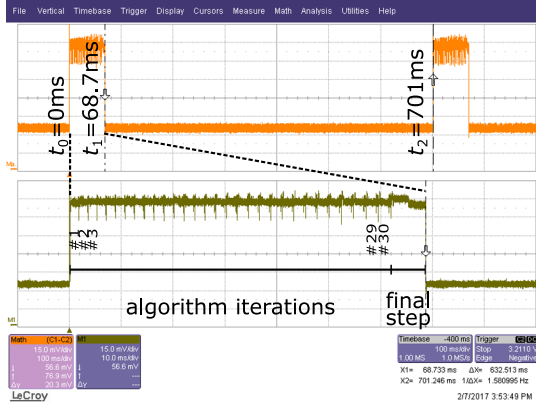


Fig. 4. Current profile of the TM4C1294NCPDT microcontroller while decoding ECG signal using OMP algorithm, with  $n = 256$ ,  $m = 68$  and  $\bar{i} = 30$ . Vertical scale is 7.5 mA/div. The bottom trace (10 ms/div) is a zoom of the top trace (100 ms/div), and allows a clear identification of the 30 OMP iterations and of the final step.

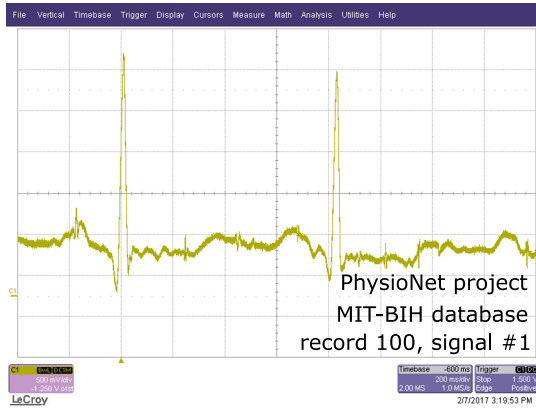


Fig. 5. Example for ECG signal encoded with the rakeness approach and decoded by OMP, with  $n = 256$ ,  $m = 68$  and  $\bar{i} = 30$ . The signal is taken from the PhysioNet project, MIT-BIH database, record 100.

execution of the code, since Flash access time is much longer than RAM access time.

The power consumption of the TM4C1294NCPDT microcontroller, due to the internal architecture, is almost constant and mainly depending on the running frequency and on which peripherals are enabled. Actually, a small increment in the power consumption can be observed when floating point operations are executed. Experimentally, the current consumption while executing the decoding algorithms is observed to be almost constant, and roughly equal to  $I_{\text{avg}} = 45.5$  mA when the system clock is set to  $f_{\text{clk}} = 120$  MHz. With this first order approximation we can model the energy required for decoding a time window as  $E_D = V_{dd} I_{\text{avg}} T_d$ , where  $V_{dd} = 3.3$  V and  $T_d$  is the decoding time.

As a visual example, Fig. 4 show the current profile measured during a signal decoding using OMP with  $n = 256$ ,  $\bar{i} = 30$  and  $m = 68$ . At the reference time  $t_0 = 0$  (trigger event), the processor starts the decoding. Note that 30 iterations can be observed, with increasing duration. A final step, much longer with respect to the iterations and with a different profile, is also identifiable. At the time  $t_1 \approx 68.7$  ms the reconstruction is complete, and the processor enters a low-

power sleep mode, up to the time  $t_2 = nT \approx 701$  ms, when the system is woken up and the decoding of the new time window starts. An example of the reconstruction of an ECG signal using OMP with  $n = 256$ ,  $\bar{i} = 30$  and  $m = 68$  is in Fig. 5. The depicted ECG signal has been taken from the PhysioNet database [31]. More precisely, it is a 2 seconds length signal extracted from signal 1 of record 100 of the MIT-DIB database. The rakeness CS approach has been used to encode the signal.

By means of empirical observations, we developed a simple model for estimating the number of CPU cycles required for decoding a signal.

In OMP, the observed duration of the  $i$ -th iteration is almost linearly increasing with  $i$  and with  $m$ . A good fit is achieved by estimating the duration of the  $i$ -th iteration with

$$N_i^{\text{OMP}} \approx 7.5 \cdot 10^3 + 2.9 \cdot 10^3 \cdot m + 44 \cdot m \cdot i$$

OMP also requires a final step. Its length is dominated by the least squares problem whose size is  $\bar{i}$ . The best fit we get is

$$N_{\text{LS}}^{\text{OMP}} \approx -5.3 \cdot 10^5 + 10^3 \cdot \bar{i}^2 + 1.8 \cdot 10^2 \cdot m \cdot \bar{i}$$

The full decoding (i.e., after  $\bar{i}$  iterations and the final step) requires a number of CPU cycles equal to

$$N^{\text{OMP}} = \sum_{i=1}^{\bar{i}} N_i^{\text{OMP}} + N_{\text{LS}}^{\text{OMP}} \quad (3)$$

As a simple verification, we can consider  $m = 68$  and  $\bar{i} = 30$  as in Fig. 4. With these values, it is  $N^{\text{OMP}} = 8.23 \cdot 10^6$ , corresponding to an execution time of 68.6 ms with a CPU clock  $f_{\text{clk}} = 120$  MHz. The value of 68.7 ms observed in Fig. 4 is almost equal to the estimated one.

When considering CoSaMP, the duration of the  $i$ -th iteration is not determined by  $i$ , but there is an uncertainty due to the data dependency. For this reason, we focus on the *average* number of cycles of the generic iteration given different input signal instances, that has been found to be dependent on  $m$  and on  $K + \Delta K$

$$N_i^{\text{CoSaMP}} \approx 5.6 \cdot 10^5 + 9.2 \cdot m \cdot (K + \Delta K)^2$$

while the total number of cycles after  $\bar{i}$  iterations is

$$N^{\text{CoSaMP}} = \bar{i} \cdot N_i^{\text{CoSaMP}} \quad (4)$$

Finally, also in the considered version of IHT we have to compute an average number of CPU cycles due to the data dependency. Empirically, the dependence on  $K$  is very weak and be neglected, so that

$$N_i^{\text{IHT}} \approx 1.1 \cdot 10^6 + 9.4 \cdot 10^4 \cdot m$$

that leads to a total number of cycles

$$N^{\text{IHT}} = \bar{i} \cdot N_i^{\text{IHT}} \quad (5)$$

Comparing average performance results shown in Fig. 2 in terms of  $m$  and  $\bar{i}$  and limited to the most interesting cases ARSNR = 25 dB, 30 dB and 35 dB, with energy consumption achieved with the proposed model  $E_D = V_{dd} I_{\text{avg}} N / f_{\text{clk}}$  where the number of cycles  $N$  is computed accordingly to

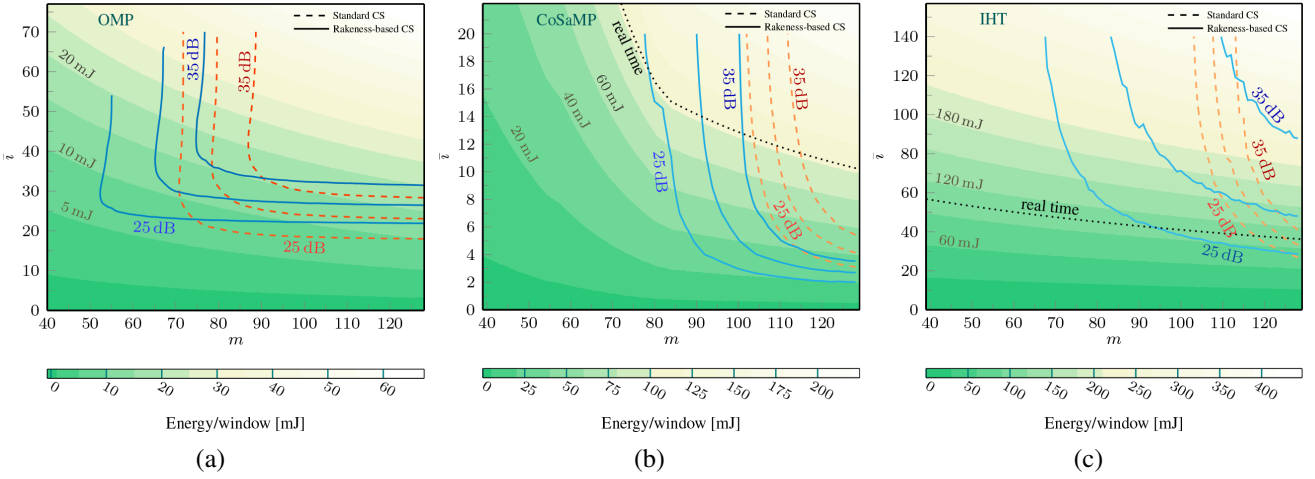


Fig. 6. Energy required for decoding a single time windows for the considered greedy algorithms in the TM4C1294NCPDT microcontroller running at 120 MHz for different values of number of measurements  $m$  and iterations  $\bar{i}$ . (a): OMP; (b): CoSaMP; (c): IHT. The plots also show the real-time constraint curve, defined as the point for which the decoding time is equal to the time window length  $nT = 711$  ms, and some performance curves at given ARSNR taken from Fig. 2.

(3), (4) or (5), and  $f_{\text{clk}} = 120$  MHz, we get the plots shown in Fig. 6 that can be commented as follows.

- OMP ensures not only better performance, but also lower cost. Energy required for most of the configurations ranges from 5 mJ to 20 mJ per reconstruction. CoSaMP ranges from 20 mJ to 50 mJ, while IHT requires at least 70 mJ. Note also that, for many CoSaMP configurations and for almost all IHT configurations, the real time constraint is not satisfied;
- A fair comparison between reconstruction costs for the standard and the rakeness approach is difficult, due to the different operating points. Focusing on OMP, that has been shown to ensure both the highest performance and the lowest cost, it is easy to find a point ensuring the minimum reconstruction energy given a target quality. As an example, considering the rakeness approach, with  $\text{ARSNR} = 30$  dB, we get  $m = 68$  and  $\bar{i} = 30$ , that are the parameters used in the example of Fig. 4 and 5. In this point,  $E_d = 10.4$  mJ. With the standard approach, the minim energy configuration is  $m = 85$  and  $\bar{i} = 26$ , with  $E_d = 10.7$  mJ. Even if the decoding energy is almost identical, the twenty percent smaller  $m$  required by rakeness allow many system optimization such as a reduced corresponding transmission and reception cost and a lower amount of memory required at the decoder side. Interestingly, two points ensuring a minimum  $m$  value can be found not very far from the already considered ones. For the rakeness approach, for  $\text{ARSNR} = 30$  dB, it is given by  $m = 65$  and  $\bar{i} = 37$ , with  $E_d = 13.3$  mJ, while for the standard approach is  $m = 78$  and  $\bar{i} = 34$ , with  $E_d = 14.0$  mJ. As in the previous case, the difference in the decoding energy is small, but the rakeness approach allows a 15% percent reduction in terms of  $m$ .

### B. ARM big.LITTLE

We profiled the OMP, CoSaMP and IHT algorithms on the Hardkernel Odroid-XU3 board, an evaluation board based on Samsung Exynos 5422, a multi-core CPU representative of recent high-end smartphones. The Exynos 5422 implements ARM's big.LITTLE heterogeneous multiprocessing solution with a cluster of four Cortex-A15, out-of-order "big" processors, and a cluster of four, in-order "LITTLE" Cortex-A7 processors. Since both CPUs are architecturally compatible, the reconstruction tasks can be allocated on demand to each CPU, to suit performance needs. Nonetheless, the two clusters have very different performance and power consumption.

The reconstruction algorithms, introduced in Section II.B, were implemented in C to run on the ARM cores. On top of the Odroid-XU3 runs Ubuntu 14.04.1 LTS (GNU/Linux 3.10.51+armv7l) with gcc v. 4.8.2. To measure the energy consumption, we make use of the on-board voltage/current sensors and split power rails, which allow us to measure separately the power consumption of the A15 cores, A7 cores, GPU and DRAM. The readout of the sensors was implemented in a low-priority thread, with a sampling interval of 25 ms and an average CPU consumption below 3%.

Table II shows the results of our evaluation, comparing the energy required by the three algorithms to reconstruct a window of ECG samples with a target ARSNR of 30 dB (there are three different couples of  $m$  and  $\bar{i}$  for each algorithm that achieve 30 dB) when running on two corner operating points of the Odroid-XU3 board as well as on the TI board. As expected OMP shows the best energy efficiency, while IHT is the most energy consuming algorithm. In addition, from Table II we can notice that for each algorithm the impact of the different  $(m, \bar{i})$  configuration on the energy consumption is preserved across the different architectures. The higher energy reported for the TI board comes from two factors: (i) we are measuring the whole embedded system consumption (board measurements) with respect to a CPU-only measurement from the Odroid-XU3 sensors; (ii) the TM4C1294 is manufactured

TABLE II

ENERGY CONSUMPTION [mJ] TO RECONSTRUCT AN ECG WINDOW WITH A TARGET ARSNR=30 dB FOR OMP, CoSaMP, IHT RUNNING ON THE ODROID-XU3 BOARD IN BOTH THE FASTEST (A15 AT 1.9 GHz) AND SLOWEST (A7 AT 0.8 GHz) CORNERS AS WELL AS ON THE TI PLATFORM. FOR EACH ALGORITHM THERE ARE THREE DIFFERENT CONFIGURATIONS ( $m, \bar{r}$ ) THAT LEAD TO 30 dB.

	OMP			CoSaMP			IHT		
	$m = 65$ $\bar{r} = 37$	$m = 80$ $\bar{r} = 28$	$m = 100$ $\bar{r} = 27$	$m = 95$ $\bar{r} = 8$	$m = 92$ $\bar{r} = 12$	$m = 117$ $\bar{r} = 3$	$m = 90$ $\bar{r} = 95$	$m = 100$ $\bar{r} = 70$	$m = 115$ $\bar{r} = 55$
A7 @ 0.8 GHz	1.36	1.13	1.32	6.22	9.28	2.49	29.61	23.91	19.32
A15 @ 1.9 GHz	6.01	5.06	5.89	23.19	35.08	9.55	144.02	127.95	88.33
M4F @ 120 MHz	13.31	11.08	13.07	61.22	89.19	27.78	228.57	176.47	148.15

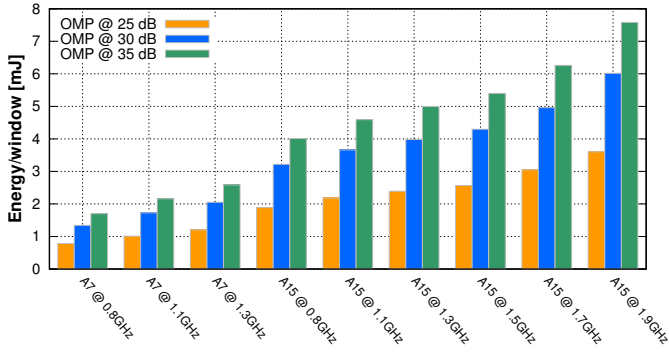


Fig. 7. Energy (mJ) for OMP to target three different quality levels (25, 30, 35 dB) on the different operating points available on the Odroid-XU3.

in 65 nm whereas the Odroid-XU3 in a 28 nm process, which combined with the difference in reconstruction time (almost one order of magnitude), justifies the disparity in consumption.

As a final experiment, we wanted to investigate how the reconstruction quality (considering the minimum  $m$  possible) affects the power consumption. We profiled the more energy-efficient algorithm, i.e. OMP, on all the operating points on the Odroid-XU3 for 3 target reconstruction qualities. Figure 7 shows the results of such analysis. First, as expected for both A7 and A15 the reconstruction energy decreases proportionally with the core frequency, with a linear trend for the A7 and a superlinear trend for the A15. This is primarily due to the voltage scaling associated with the frequency reduction. Second, the energy-saving ratio achieved by reducing the QoS is preserved on the different operating point. Thanks to that Digital Voltage-Frequency Scaling is an effective knob to trade-off reconstruction time, quality of service and energy consumption.

## VI. CONCLUSION

The benefits introduced by rakesness at the decoder side have been discussed in a CS system designed for ECG signals. Three iterative algorithms have been considered, and some trade-offs including the number of measurements, the number of iterations, reconstruction quality, and energy required by the decoding on two different ARM architectures are considered. In all cases, OMP shows to be the best choice in terms of reconstruction algorithm both for the lower energy requirement and for the higher reconstruction quality. Furthermore, the rakesness approach proves to be the a better choice with respect to the standard approach also at the decoder side, as it is capable

to reach a target quality with the same amount of energy, but with many advantages given by the lower required  $m$ , such as the reduction of the encoder/decoder transmission/reception costs or of the decoder memory requirements.

## REFERENCES

- [1] A. Munir, A. Gordon-Ross, and S. Ranka, "Multi-core embedded wireless sensor networks: Architecture and applications," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 6, pp. 1553–1562, June 2014.
- [2] X. Zhang, H. Jiang, L. Zhang, C. Zhang, Z. Wang, and X. Chen, "An energy-efficient ASIC for wireless body sensor networks in medical applications," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 4, no. 1, pp. 11–18, Feb 2010.
- [3] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [4] D. L. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [5] M. Mangia, D. Bortolotti, F. Pareschi, A. Bartolini, L. Benini, R. Rovatti, and G. Setti, "Zeroing for hw-efficient compressed sensing architectures targeting data compression in wireless sensor networks," *Microprocessors and Microsystems*, vol. 48, pp. 69 – 79, 2017.
- [6] M. Mangia, R. Rovatti, and G. Setti, "Rakesness in the design of analog-to-information conversion of sparse and localized signals," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 5, pp. 1001–1014, May 2012.
- [7] R. Rovatti, G. Mazzini, and G. Setti, "Performance of chaos-based asynchronous ds-cdma with different pulse shapes," *IEEE Commun. Letters*, vol. 8, no. 7, pp. 416–418, 2004.
- [8] G. Setti, R. Rovatti, and G. Mazzini, "Enhanced rake receivers for chaos-based ds-cdma," *IEEE Transaction on Circuits and Systems I, Fundam. Theory Appl.*, vol. 48, no. 7, pp. 818–829, 2001.
- [9] F. Chen, A. P. Chandrakasan, and V. M. Stojanovic, "Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 3, pp. 744–756, March 2012.
- [10] M. Shooran, M. H. Kamal, C. Pollo, P. Vandergheynst, and A. Schmid, "Compact low-power cortical recording architecture for compressive multichannel data acquisition," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 8, no. 6, pp. 857–870, Dec 2014.
- [11] X. Liu, M. Zhang, T. Xiong, A. G. Richardson, T. H. Lucas, P. S. Chin, R. Etienne-Cummings, T. D. Tran, and J. V. der Spiegel, "A fully integrated wireless compressed sensing neural signal acquisition system for chronic recording and brain machine interface," *IEEE Trans. on Biomedical Circuits and Systems*, vol. 10, no. 4, pp. 874–883, 2016.
- [12] D. Gangopadhyay, E. G. Allstot, A. M. R. Dixon, K. Natarajan, S. Gupta, and D. J. Allstot, "Compressed sensing analog front-end for bio-sensor applications," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 2, pp. 426–438, Feb 2014.
- [13] 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," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 10, no. 1, pp. 149–162, Feb. 2016.
- [14] Y. C. Cheng, P. Y. Tsai, and M. H. Huang, "Matrix-inversion-free compressed sensing with variable orthogonal multi-matching pursuit based on prior information for ecg signals," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 10, no. 4, pp. 864–873, Aug 2016.

- [15] M. Mangia, D. Bortolotti, A. Bartolini, F. Pareschi, L. Benini, R. Rovatti, and G. Setti, "Application of compressed sensing to ecg signals: Decoder-side benefits of the rakeness approach," in *2016 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Oct 2016, pp. 352–355.
- [16] D. Bortolotti, M. Mangia, A. Bartolini, R. Rovatti, G. Setti, and L. Benini, "Energy-aware bio-signal compressed sensing reconstruction on the wbsn-gateway," *IEEE Transactions on Emerging Topics in Computing*, vol. PP, no. 99, pp. 1–1, 2017.
- [17] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [18] 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.
- [19] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265 – 274, 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1063520309000384>
- [20] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.
- [21] M. Mangia, F. Pareschi, V. Cambareri, R. Rovatti, and G. Setti, "Rakeness-based design of low-complexity compressed sensing," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, no. 5, pp. 1201–1213, May 2017.
- [22] G. Jacovitti, A. Neri, and G. Scarano, "Texture synthesis-by-analysis with hard-limited gaussian processes," *Image Processing, IEEE Transactions on*, vol. 7, no. 11, pp. 1615–1621, 1998.
- [23] J. H. Van Vleck and D. Middleton, "The spectrum of clipped noise," *Proceedings of the IEEE*, vol. 54, no. 1, pp. 2–19, 1966.
- [24] 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, no. 1, 2009.
- [25] M. Elad, "Optimized projections for compressed sensing," *Signal Processing, IEEE Transactions on*, vol. 55, no. 12, pp. 5695–5702, 2007.
- [26] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *Image Processing, IEEE Transactions on*, vol. 18, no. 7, pp. 1395–1408, 2009.
- [27] J. Zhang, Z. Gu, Z. L. Yu, and Y. Li, "Energy-efficient ecg compression on wireless biosensors via minimal coherence sensing and weighted  $\ell_1$  minimization reconstruction," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 2, pp. 520–528, March 2015.
- [28] Y. Suo, J. Zhang, T. Xiong, P. S. Chin, R. Etienne-Cummings, and T. D. Tran, "Energy-efficient multi-mode compressed sensing system for implantable neural recordings," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 8, no. 5, pp. 648–659, Oct 2014.
- [29] T. Blumensath and M. E. Davies, "Normalized iterative hard thresholding: Guaranteed stability and performance," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 298–309, April 2010.
- [30] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 3, pp. 289–294, March 2003.
- [31] A. L. Goldberger *et al.*, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. 215–220, Jun. 2000.
- [32] S. Mallat, *A wavelet tour of signal processing: the sparse way*. Access Online via Elsevier, 2008.
- [33] Z. Zhang, T. P. Jung, S. Makeig, and B. D. Rao, "Compressed sensing for energy-efficient wireless telemonitoring of noninvasive fetal ecg via block sparse bayesian learning," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 2, pp. 300–309, Feb 2013.
- [34] *Tiva™ C Series TM4C1294 Connected LaunchPad Evaluation Kit*, Texas Instruments Inc., Oct. 2016.



His research activity focuses on analog and mixed-mode electronic circuit design, statistical signal processing, random number generation and testing, and electromagnetic compatibility. He was recipient of the best paper award at ECCTD 2005 and the best student paper award at EMC Zurich 2005.



He was the recipient of the 2013 IEEE CAS Society Guillemain-Cauer Award and best student paper award at ISCAS2011. He is also the Web and Social Media Chair for ISCAS2018



ultra-low power bio-sensors nodes operating in near-threshold for WBSN applications and low-level power management techniques for many-cores HPC nodes. Currently, his research domain comprises low-power neuromorphic architectures for bio-inspired algorithms and cognitive applications.



**Fabio Pareschi** (S'05-M'08) received the Dr. Eng. degree (with honours) in Electronic Engineering from University of Ferrara, Italy, in 2001, and the Ph.D. in Information Technology under the European Doctorate Project (EDITH) from University of Bologna, Italy, in 2007. He is currently an Assistant Professor in the Department of Engineering, University of Ferrara. He is also a faculty member of ARCES - University of Bologna, Italy. He served as Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS - PART II (2010-2013).

**Mauro Mangia** (S'09-M'13) received the B.Sc. and M.Sc. in Electronic Engineering and the Ph.D. degree in Information Technology from the University of Bologna (Bologna, Italy), respectively in 2005, 2009 and 2013. He is currently a Postdoctoral Researcher in the statistical signal processing group of ARCES - University of Bologna. In 2009 and 2012, he was a visiting Ph.D. student at the Ecole Polytechnique Federale de Lausanne (EPFL). His research interests are in nonlinear systems, compressed sensing, ultra-wideband systems, and systems biology. He was the recipient of the 2013 IEEE CAS Society Guillemain-Cauer Award and best student paper award at ISCAS2011. He is also the Web and Social Media Chair for ISCAS2018

**Daniele Bortolotti** received the Ph.D. degree in Electronics, Computer Science, and Telecommunications from the University of Bologna, Italy in 2014. After, he held the position Post-Doctoral Researcher in the DEI department at the University of Bologna as well as at the AES group at TU Berlin, Germany. Initially, the focus of his research has been on virtual platforms to study architectural aspects of multi-processors systems-on-chip and technological implications of sub-micrometer technology. Afterwards, he worked on HW/SW design strategies for

**Andrea Bartolini** received a Ph.D. degree in Electrical Engineering from the University of Bologna, Italy, in 2011. He is currently an Assistant Professor in the Department of Electrical, Electronic and Information Engineering Guglielmo Marconi (DEI) at the University of Bologna. Dr. Bartolini's research interests are on the HW/SW co-design of energy, power and thermal efficient computing systems: from low-power embedded to large scale HPC systems.



**Luca Benini** (S94M97SM04F07) holds the chair of digital Circuits and systems at ETHZ and is Full Professor at the Università di Bologna. Dr. Benini's research interests are in energy-efficient system design for embedded and high-performance computing. He is also active in the area of energy-efficient smart sensors and ultra-low power VLSI design. He has published more than 800 papers, five books and several book chapters. He is a Fellow of the IEEE and the ACM and a member of the Academia Europaea. He is the recipient of the 2016

IEEE CAS Mac Van Valkenburg award.



**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 a 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. His research interests include nonlinear circuits, implementation and application of chaotic circuits and systems, electromagnetic compatibility, statisti-

cal 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 ECCTD2005, and the best student paper award at EMCZurich2005 and at ISCAS2011. He held several editorial positions and served, in particular, as the Editor-in-Chief for the IEEE Transactions on Circuits and Systems - Part II (2006-2007) and of the IEEE Transactions on Circuits and Systems - Part I (2008-2009). Dr. Setti was the Technical Program Co-Chair ISCAS2007, ISCAS2008, ICECS2012, BioCAS2013 as well as the General Co-Chair of NOLTA2006 and ISCAS2018. He was Distinguished Lecturer of the IEEE CAS Society (2004-2005 and 2014-2015), a member of its Board of Governors (2005-2008), and he served as the 2010 President of CASS. 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.