Sparse sensing matrix based compressed sensing in low-power ECG sensor nodes

(Article begins on next page)

This is the author's version of the article that has been presented at IEEE BioCAS2017
The editorial version of the paper is available at http://dx.doi.org/10.1109/BIOCAS.2017.8325155

Sparse Sensing Matrix based Compressed Sensing in Low-power ECG Sensor Nodes

Alex Marchioni[†], Mauro Mangia[‡], Fabio Pareschi[*‡], Riccardo Rovatti[†‡] and Gianluca Setti[*‡]
[*]ENDIF, University of Ferrara, Italy - Email: {fabio.pareschi, gianluca.setti}@unife.it
[†]DEI, [‡]ARCES, University of Bologna, Italy - Email: {alex.marchioni2, mauro.mangia2, riccardo.rovatti}@unibo.it

***Abstract*—Compressed Sensing (CS) is an acquisition technique able to reduce the operating cost (e.g., energy requirements) of a signal processing system thanks to its capability of simultaneously sampling and compressing an input waveform. Here we focus on Electrocardiogram (ECG) signals acquired by means of a custom designed acquisition board that exploits CS as early-digital compression stage. We show that when CS acquisition sequences are sparse ternary, i.e., with symbols $\{-1, 0, +1\}$ and designed to maximize their *rakeness*, it is possible to achieve a reduction in the energy required for ECG signal compression by a factor between 25 and 30 with respect to the standard acquisition with independent and identically distributed random sequences.**

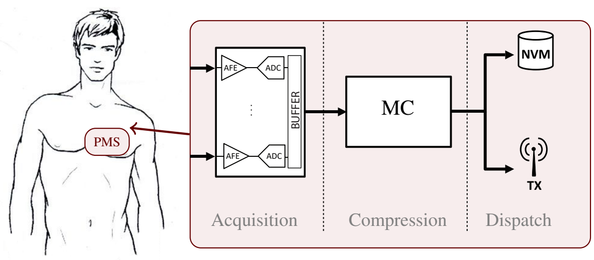

Fig. 1. Block scheme of a typical biomedical monitor equipment.

I. Introduction

The introduction of the Compressed Sensing (CS) paradigm [1] paved the way for adopting resource-efficient Analog-to-Information Converters (AICs) to replace standard Nyquist-rate Analog-to-Digital converters (ADCs). Very interesting applications are related to the acquisition of biomedical signals [2], [3], [4], whose features allows CS to exploit its capability to simultaneously sample and compress signals, thus opening multiple energy saving possibilities. In particular, there is a recent interest in applying CS to sensing nodes in ultra-low power personal monitoring systems (PMS), in the effort of reducing power consumption, extending battery life, and enabling ubiquitous and long-term monitoring policies for the future healthcare system.

This paper is focused on the application of the CS as an early-digital compression stage for Electrocardiogram (ECG) signals. More precisely, the scenario we consider is shown in Fig. 1, where a battery-powered system collects data from biosensors, compresses them via a microcontroller (MC), and either stores them in a non-volatile memory (NVM) or transmits them using a wireless protocol. In this context, CS has been shown to be a very good candidate for lowering power requirements with respect to state-of-the-art compression algorithms [5], [6].

Our work focuses on the impact of the CS-based compression algorithm running on the MC. Its main task is to reduce as much as possible the rate of the output data, thus increasing the energy saved when storing or transmitting it. At the same time it must respect the constraint of preserving the quality of the acquired signal and to be itself efficient in terms of energy requirements and algorithm complexity. With these aims in mind, among the several CS-based approaches proposed in the literature we concentrate on that relying on *rakeness*. The driving concept is to exploit the the statistical features of input signals (i.e., the fact that their energy is *localized* in a certain bandwidth) either to boost the achievable compression rate or to increase signal reconstruction quality. Roughly speaking, this is not dissimilar to what we employed in (chaos-based) DS-CDMA communication, where chip waveforms, spreading sequence statistics and rake receivers taps were jointly selected to collect (rake) as much energy as possible at the received side [7][8]. CS rakeness-based approaches have also recently been improved by allowing the design of sensing stages based on a *sparse* matrix [9], which has shown to considerably improve energy saving.

The paper is organized as follows. In Section II basic concepts of CS, rakeness and sparse sensing are introduced. In Section III shows the hardware architecture. Reconstruction results, along with some energetic trade-off evaluations, are provided in Section IV. Finally, we draw the conclusion.

II. CS Fundamentals

Let $x \in \mathbb{R}^n$ be an instance of a discrete-time signal, defined by its n Nyquist-rate samples. Let also x be κ-sparse, i.e., a proper orthonormal n-dimensional *sparsity basis* $S \in \mathbb{R}^{n\times n}$ exists, in which any instance $x = S\xi$ is represented by a vector $\xi \in \mathbb{R}^n$ with no more than $\kappa \ll n$ non-zero components. Under these assumptions, CS theory states [1] that it is possible to overcome the limit imposed by Shannon sampling and represent x using $m < n$ measurements only obtained via the linear projection of x over a sensing matrix $A \in \mathbb{R}^{m\times n}$, that is

$$y = Ax + \nu = AS\xi + \nu \tag{1}$$

where $y \in \mathbb{R}^m$ is a vector collecting the m measurements, and ν is an additive disturbance term modeling non-idealities and noise. According to (1), the j-th measurements is $y_j = \langle A_{j,\cdot}, x\rangle$, i.e., the scalar product between the j-th row $A_{j,\cdot}$ of A and x. The possibility of retrieving x from y relies on some properties of A [10] and the reconstructed signal $\hat{x} = S\hat{\xi}$ is obtained by solving the optimization problem

$$\hat{\xi} = \arg\min_{\xi} \|\xi\|_1 \;\; \text{s.t.} \;\; \|AS\xi - y\|_2 < \varepsilon \tag{2}$$

where $\|\cdot\|_1$ ($\|\cdot\|_2$) are the standard $\ell_{1(2)}$ norms, and ε bounds the effects of ν. In other words, retrieving the input signal is an ill-posed inverse problem, and $\hat{\xi}$ is determined by looking for the sparsest vector ξ that solves (1). According to [10], the easiest way to ensure reconstruction is to draw elements of A as independent and identically distributed (i.i.d.) random variables. In this case, convergence is guaranteed [1] if $m \geq \mathcal{O}(\kappa \log(n/\kappa))$. Interestingly, a common hardware-friendly choice [3], [4] is to draw elements of A by means of an antipodal Bernoulli distribution, i.e., $A \in \{-1,+1\}^{m\times n}$ where -1 and $+1$ occur with the same probability. No performance reductions are observed with respect to using more complex (e.g., Gaussian) distributions [11].

When x is also localized, i.e., its energy is not distributed uniformly in the signal space (as it is the case for many biomedical signals), one can exploit *rakeness*-based CS [12][9] to enhance signal reconstruction quality or reducing the number of measurements m.

More specifically, by defining the *rakeness* ρ as the expected value of the energy measurement for any possible x and any possible generic row $A_{j,\cdot}$, i.e. $\rho = \mathbf{E}\left[|\langle A_{j,\cdot}, x\rangle|^2\right]$, one improves CS by maximizing ρ under some constraints necessary to preserve randomness of the rows of A. The solution of the optimization problem is a relation between the correlation matrix $\mathcal{X}$ of x and the correlation matrix $\mathcal{A}$ to be used in the generation of the generic row $A_{j,\cdot}$.

Note that this concept is perfectly compatible with the design of an antipodal sensing matrix A and has also recently been extended [9] also to the case of *sparse ternary* sensing matrix, i.e. when $A \in \{-1,0,+1\}^{m\times n}$. The energy cost of computing (1) depends on $m \times n$, but also on the value of the elements of A. By setting zero elements in A, the corresponding multiply-and-accumulate operations can be skipped. In this sense, as shown in [9], the adoption of a rakeness-based sparse A ensures at the same time extremely lightweight sensing with excellent compression performance[1].

III. Hardware Architecture Considered

A custom designed ECG acquisition system has been used to evaluate performance of the sparse ternary CS approach described in the previous section.

The system is based on on a TI ADS1292 low-power analog front-end for ECG connected to a TI EK-TM4C123GXL evaluation board, embedding a low-power low-cost ARM Cortex-M4F MC, that controls the analog front-end and performs signal compression.

Measurements are obtained according to the standard ECG lead system. A fully differential input channel of the AD1292 is used to acquire the electric potential between the left arm and the right arm (bipolar lead I, LA–RA). The differential signal is internally amplified (by a factor of 6) and converted to digital by mean of a 24-bit sigma-delta ADC with a sampling frequency $f_s = 250\,\mathrm{Hz}$. The right leg drive circuit is configured to sense the input and drive the body with the inverted common-mode signal through the right leg electrode with the aim of reducing the common-mode noise. Resolution is then reduced to 11 bits by simply discarding unused high-significant bits and noisy low-significant ones.

This paper focuses on the compression aspects of the system, and its main contribution is to show how a properly designed sensing matrix A could reduce power requirements of generic MC-based ECG compression system, considering the Cortex-M4F processor as reference case.

To this aim, a proper optimized CS-based early-digital compression stage is implemented on the ARM processor, taking $n = 256$ successive ECG samples from the ADS1292 as input signal and a variable number m of measurement. A simple but effective power consumption model has also been devised, based on the observation that the ARM processor consumes almost constant power which mainly depends on the working frequency and the enabled internal peripherals. Experimentally, with a system clock $f_{\mathrm{clk}} = 16\,\mathrm{MHz}$ and all peripherals disabled, the observed current consumption is $I_{\mathrm{avg}} \approx 11.3\,\mathrm{mA}$. With this, we can estimate the energy required by the ARM for executing a code as $E = V_{dd} I_{\mathrm{avg}} N_c / f_{\mathrm{clk}}$, where $V_{dd} = 3.3\,\mathrm{V}$ and N_c is the number of clock cycles. Even if this approach has some drawback (i.e., we are measuring the whole ARM consumption, including, f.i., also the clock generation), we can still use it to identify the saved energy with respect to a standard case.

Three different sensing strategies, all based on either an antipodal or sparse ternary sensing matrix, were implemented to evaluate the achievable gain. The first reference approach is the standard CS characterized by a non-sparse antipodal sensing matrix. Given the Cortex architecture, the most convenient implementation has been identified in storing the generic elements $A_{j,k}$ of the sensing matrix as 32-bit integer values, and computing with a simple loop

$$y_j = \langle A_{j,\cdot}, x\rangle = \sum_{k=0}^{n-1} A_{j,k} x_k \tag{3}$$

The memory footprint of this approach is inefficient (especially since $A_{j,k} \in \{-1,1\}$) and dominated by the storing of A ($4nm$ bytes required). However, the execution of the loop in (3) is pretty fast, taking advantage both from the multiply-and-accumulate ARM hardware instruction and from the 32 bit data alignment. The value of N_c in the evaluation of y depends on m and on n.

The second approach is the non-A-sparse antipodal rakeness-based sensing. From a computational point of view, there is no difference with respect to the standard CS case. Also in this case N_c depends on m and n. The correlation matrix $\mathcal{A}$ used to generate the generic row $A_{j,\cdot}$ maximizing rakeness has been computed by estimating $\mathcal{X}$ as in [4].

The last considered case is the A-sparse rakeness-based approach.

Since A is a sparse matrix with $A_{j,k} \in \{-1,0,1\}$, (3) can be reduced to

$$y_j = \langle A_{j,\cdot}, x\rangle = \sum_{k\in K_j^+} x_k - \sum_{k\in K_j^-} x_k \tag{4}$$

where $K_j^+ = \{k \mid A_{j,k} = 1\}$ and $K_j^- = \{k \mid A_{j,k} = -1\}$ are the index sets of the non-null elements of $A_{j,\cdot}$. Let us indicate with d the number of non-null elements for each

[1] The MATLAB© framework developed in [9] is online available at http://cs.signalprocessing.it/download.html along with a few demo examples.

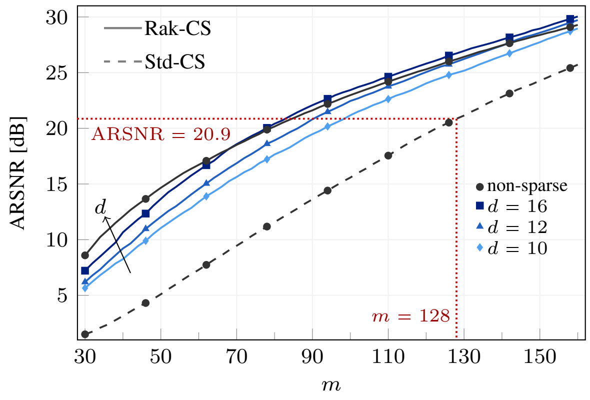

Fig. 2. Performance in terms of ARSN as a function of m for the standard CS, the non-sparse antipodal rakeness-based CS and the sparse ternary rakeness-based CS for different values of d.

y_j, i.e., $|K_j^+| + |K_j^-| = d$, $\forall j$. Clearly, the complexity of computing (4) is much lower than (3) since only d operations (either additions or subtractions) instead of n are required. The value of N_c in this case depends on m and d instead of m and n. This case presents a considerable saving also from the memory footprint point of view. The storing of a row of A is dominated by memorizing both K_j^+ and K_j^+, that can share d memory cells with at least $\log_2 n$ bits (i.e., d bytes in our case where $n = 256$). Therefore, memorizing the entire A requires in this case only md bytes.

IV. Results

Performance in terms of reconstruction quality for the different encoding methods were evaluated by reconstructing the signal with the SPGL1 toolbox[2].

For comparison, we use the Average Reconstruction SNR (ARSNR), defined as the average observed value of the reconstruction SNR

$$\mathrm{RSNR} = 20 \log_{10} \left(\frac{\|x\|_2}{\|x - \hat{x}\|_2} \right)$$

Provided values of ARSNR have been computed by averaging over 500 different combinations of x and A.

Results are shown in Figure 2 as a function of m. The standard CS (Std-CS) and the binary antipodal non-A-sparse rakeness-based CS are considered along with many ternary sparse cases with different values of d (Rak-CS). As expected from [9], all rakeness-based sparse approaches achieve performance between that of the standard CS and of the non-sparse rakeness-based CS. Worth stressing that performance for $d = 16$ is already almost superimposed to the upper limit given by the non-sparse antipodal rakeness case. This highlights how the introduction of the sparse rakeness approach can have an almost negligible cost in terms of performance reduction.

In order to highlight the advantage in terms of energy saving, we introduce two additional figures of merit related to the amount of information and of energy required to ensure a given quality of service. Let us focus on ARSNR $= 20.9$ dB, that is the value we obtain in the Std-CS when using $m = 128$, i.e., half of the measurements with respect to the Nyquist approach. This is typically considered a good working point in CS-based ECG encoding [5]. The number of measurements m required to ensure this ARSNR depends on the encoding strategy, and can be found in Table I.

[2] online available at https://www.math.ucdavis.edu/~mpf/spgl1/

TABLE I
Value of m required for achieving the target service quality ARSNR $= 20.9$ dB for Rak-CS, with the corresponding value of bit saving rate (BSR) and energy saving rate (ESR).

	d	$m_{\min}$	BSR	ESR
sparse	4	148	1.13	41.27
sparse	8	129	1.20	30.85
sparse	10	98	1.48	33.71
sparse	12	90	1.61	31.38
sparse	16	83	1.75	27.28
sparse	20	85	1.60	22.23
non-sparse	256	85	1.51	1.51

Let us consider the number of bits needed by each encoding strategy for achieving the desired ARSNR. This is simply given by mb_y, where, however, also b_y depends on the coding as detailed in the following.

In the sparse approach regulated by (4), only d terms are present in the sum. In order to cope with the worst case we need to use $b_y = b_x + \lceil \log_2 d \rceil$, where $\lceil \cdot \rceil$ is the smallest integer not smaller than its argument, and where $b_x = 11$ bit is the precision used to encode input signal samples.

Conversely, when considering a non-sparse approach each measurement is evaluated as in (3), i.e., as the sum of n terms. Since $n = 256$ is large enough to assume that the central limit theorem holds for (3), y_j can be approximated as a normal random variable with standard deviation $\sigma_y = \sqrt{n}\sigma_x$, where σ_x is the standard deviation of the input signal [11]. Thanks to this, we limit ourselves to ensure conversion in the range $[-4\sigma_y, 4\sigma_y]$. If this assumption holds we can set $b_y = \log_2(4\sqrt{n}\sigma_x) \approx b_x + \log_2 \sqrt{n} + \log_2 4 = 17$ bits.

Consider the standard CS as reference case. The corresponding total number of bits used for encoding a signal with the target quality mentioned above is $m(b_x + \log_2 \sqrt{n} + 2) = 2176$ bits, and the energy required for the evaluation of y, computed as $V_{dd} I_{\mathrm{avg}} N_c / f_{\mathrm{clk}}$ is $\approx 614\,\mu$J. We can then introduce as figures of merit the Bit Saving Ratio (BSR) and the Energy Saving Ratio (ESR) as the reduction achieved by the (sparse) rakeness approach with respect to the reference case in terms, respectively, of number of bits and energy required for encoding. All values are shown in Table I.

Looking at Table I, it is interesting to note that, while the lower d, the larger the advantage in terms of ESR, it is possible to identify an optimum point for BSR in $d = 16$. Here, the number of bit required for the coding is even smaller than that required by the non-sparse rakeness approach. An example of an ECG signal acquired by the designed architecture, encoded with $d = 16$ and reconstructed by means of the SPGL1 toolbox is depicted in Figure 3.

As a final comment, despite being computed on a particular hardware, the proposed dimensionless values of BSR and ESR can be in principle extended to have general validity for CS-based ECG sensing system relying on a similar architecture.

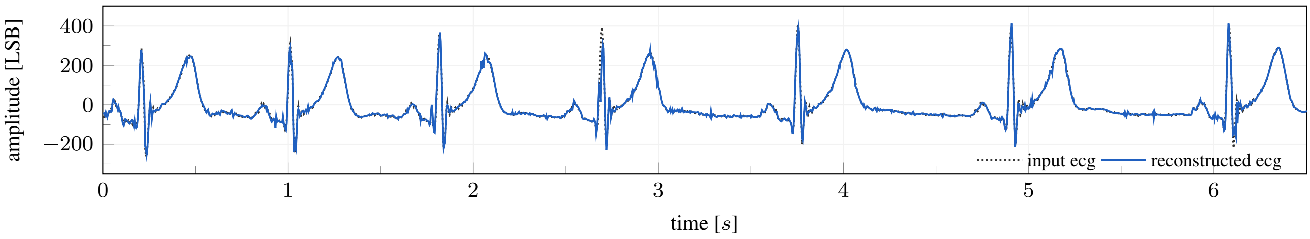

Fig. 3. Short example (about 6.5 s) of an ECG signal reconstructed after being encoded with the sparse rakeness approach with $d = 16$.

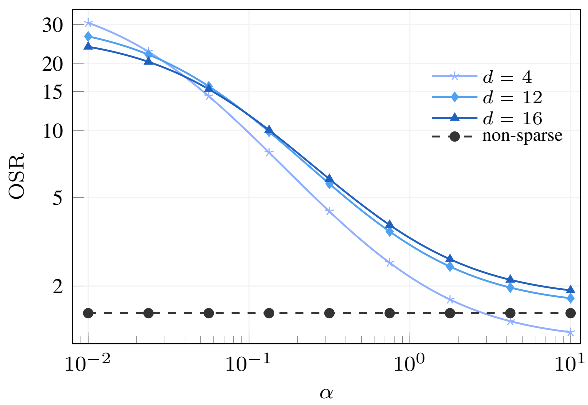

Fig. 4. Overall saving rate (OSR) for a generic system as a function of α.

Let us assume that E_{std}^{CS} is the energy required for computing measurements in a standard CS approach. Given the ESR, in a rakeness approach energy is reduced to $E_{std}^{\mathrm{CS}}/\mathrm{ESR}$. Indicating with E_{std}^{D} the energy for the measurements dispatch (either transmission or storage) in the standard CS approach, and assuming that this energy is proportional to the encoded bits, when considering a rakeness approach we need $E_{std}^{\mathrm{D}}/\mathrm{BSR}$.

Of course, the overall energy requirement is mainly due by the sum of the two contributions.

Evaluating E_{std}^{D} is a non-trivial task, and out of the scope of this paper. Typically, E_{std}^{D} is very small in case we save data into a NVM, much larger when considering a short-range wireless protocol, and even larger if a long-range transmission is considered. Here, we simply assume that $E_{std}^{\mathrm{D}} = \alpha E_{std}^{\mathrm{CS}}$, with α a proper constant to be evaluated for each specific case. The overall energy for computing and sending measurements can therefore be expressed as $(1+\alpha)E_{std}^{\mathrm{D}}$ in the standard CS, and as $(1/ESR + \alpha/BSR)E_{std}^{\mathrm{D}}$ in the rakeness case. With this, we can define the Overall energy Saving Rate (OSR) as

$$OSR = \frac{1+\alpha}{1/\mathrm{ESR} + \alpha/\mathrm{BSR}}$$

The OSR value is plotted in Figure 4 as a function of α for some typical cases. For small values of α, the energy is dominated by computing measurements, and the case $d = 4$ due to the high ESR ensures a gain up to 40. As α increases, BSR gains in importance, and other values of d ensure the highest OSR. When α further the increases, overall energy is dominated by the transmission cost, and the case $m = 16$, ensuring the highest BSR, is the optimum one. Note that, for any value of α, the sparse ternary rakeness CS approach outperforms the antipodal non-sparse rakeness CS.

V. Conclusion

In this paper, a practical example of the advantage in terms of energy saving when adopting sparse ternary rakeness approach in a CS-based ECG acquisition system is considered. When real ECG signals are taken into account, the gain in terms of energy saved for signal encoding is more than 20. The gain when considering the overall system including data storing or transmission depends on the energy required by the latter, but with a simple model it is possible to show that there is always an advantage with respect both to the standard CS approach and to the non-sparse rakeness approach.

Acknowledgment

One of the authors would like to thank Giorgia. Without her smile, this paper would not have been made possible.

References

[1] D. L. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, Apr. 2006.
[2] D. Gangopadhyay *et al.*, "Compressed sensing analog front-end for biosensor applications," *IEEE Journal of Solid-State Circuits*, vol. 49, pp. 426–438, Feb 2014.
[3] M. Shoaran *et al.*, "Compact low-power cortical recording architecture for compressive multichannel data acquisition," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 8, pp. 857–870, Dec 2014.
[4] F. Pareschi *et al.*, "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, pp. 149–162, Feb. 2016.
[5] H. Mamaghanian *et al.*, "Compressed sensing for real-time energy-efficient ecg compression on wireless body sensor nodes," *IEEE Transactions on Biomedical Engineering*, vol. 58, pp. 2456–2466, Sept 2011.
[6] Z. Zhang *et al.*, "Compressed sensing for energy-efficient wireless telemonitoring of noninvasive fetal ecg via block sparse bayesian learning," *IEEE Transactions on Biomedical Engineering*, vol. 60, pp. 300–309, Feb 2013.
[7] R. Rovatti, G. Mazzini, and G. Setti, "Enhanced rake receivers for chaos-based ds-cdma," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, pp. 818–829, Jul 2001.
[8] G. Setti, R. Rovatti, and G. Mazzini, "Performance of chaos-based asynchronous ds-cdma with different pulse shapes," *IEEE Communications Letters*, vol. 8, pp. 416–418, July 2004.
[9] M. Mangia *et al.*, "Rakeness-based design of low-complexity compressed sensing," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, pp. 1201–1213, May 2017.
[10] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, pp. 4203–4215, Dec. 2005.
[11] J. Haboba *et al.*, "A pragmatic look at some compressive sensing architectures with saturation and quantization," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, pp. 443–459, Sep. 2012.
[12] M. Mangia, R. Rovatti, and G. Setti, "Rakeness in the design of analog-to-information conversion of sparse and localized signals," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, pp. 1001–1014, May 2012.