Application of compressed sensing to ECG signals: Decoder-side benefits of the rakeness approach

(Article begins on next page)

This is the author's version of the article that has been presented at IEEE BIOCAS2016
The editorial version of the paper is available at http://dx.doi.org/10.1109/BioCAS.2016.7833804

Application of Compressed Sensing to ECG signals: Decoder-side Benefits of the Rakeness Approach

Mauro Mangia[‡], Daniele Bortolotti[†], Andrea Bartolini[†§], Fabio Pareschi[*‡],
Luca Benini[†§], Riccardo Rovatti[†‡] and Gianluca Setti[*‡]
[†]DEI, [‡]ARCES, University of Bologna, Italy - Email: {mauro.mangia2, daniele.bortolotti, riccardo.rovatti}@unibo.it
[§]Integrated Systems Laboratory, ETH Zurich, Switzerland - Email: {barandre, lbenini}@iis.ee.ethz.ch
[*]ENDIF, University of Ferrara, Italy - Email: {fabio.pareschi, gianluca.setti}@unife.it

***Abstract*—Compressed sensing has recently been actively investigated as a mean of lowering the power consumption of sensing nodes in biomedical signal devices due to its capability to reduce the amount of data to be transmitted for the correct reconstruction of the acquired waveforms. Rakeness-based design of compressed sensing stages exploits the uneven distribution of energy in the sensed signal and has proved to be extremely effective in maximizing the energy saving. Yet, many body-area sensor network architectures include intermediate gateway nodes that receive and reconstruct signals to provide local services before relaying data to a remote server. In this case, decoder-side power consumption is also an issue. In this paper, with particular reference to electrocardiographic signals, we show that rakeness-based design is also capable to reduce resources required at the decoder side for reconstruction. This happens across a variety of reconstruction algorithms that see their running time substantially reduced. Actual savings are then experimentally quantified by measuring the energy requirements of one of the algorithms on a common mobile computing platform.**

I. Introduction

The availability of personal biometric monitoring systems is commonly addressed as one of the key enabling technologies capable of a major breakthrough in improving life quality in coming years. Even if a typical application is identified in continuous patient monitoring or elders caring, many other situations may take advantage from this technology, ranging from athletes' training improvement to stress detection during safety critical tasks. In all cases what is needed is a number of miniaturized bio-sensing nodes integrated in a so-called wireless body sensor network (WBSN).

The most useful WBSN architectures entail local gateways that aggregate data coming from multiple sources and provide a first level of processing, with a possible immediate feedback to the user before routing data (either in the original or a processed form) to a remote server. Note that assuming this architecture implies that both encoding/transmission and reception/decoding issues have to be considered in the design of the system.

This is particularly important when the compressed sensing (CS) approach is an option for lowering power requirements of the sensing nodes. CS is a dimensionality reduction technique [1], [2] that by means of a linear transformation (usually a random one) maps vectors of Nyquist rate samples into smaller vectors of so called *measurements* that are enough to reconstruct the original signal. The dimensionality reduction allows a potentially large saving in the resources needed at the sensing node (mainly power) since i- the amount of additional processing (a linear transformation) is intrinsically small and can be effectively reduced; ii- the transmitter (the most power hungry stage) significantly benefits from a reduced load [3].

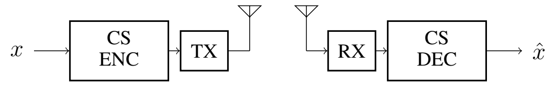


Fig. 1. CS in the link between a sensing node and the local gateway.

At the receiver side, signal reconstruction leverages the possibility of giving a sparse representation to the target signal and adopts non-linear, typically iterative, procedures with a computational complexity much higher than what is sustained by the encoding. Hence, with reference to Fig. 1 schematising the link between a sensing node and the local gateway, from a power-wise point of view CS is an intrinsically asymmetric method that reduces the resources at the sensing node while potentially making reception and reconstruction at the local gateway more expensive.

Recently, a design flow for CS systems has been proposed exploiting the common property of biomedical signals to be non-white, i.e., they do not distribute their energy uniformly in the signal space [4]. CS can be optimized by adapting the statistics with which the random linear mapping is chosen to the distribution of such energy. The driving concept here is *rakeness*, i.e., the ability of the linear transformation to capture the energy of the signal to acquire. By adopting a rakeness-based design flow[1] 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 so the power required by the transmitter.

Aim of this contribution is to show that rakeness-based design is beneficial also for the receiver side. In a well-defined and reproducible setting, we analyze a number of different reconstruction algorithms establishing the amount of computation needed by each of them to reconstruct the signal with a prescribed quality. For one of these algorithms we will also be able to translate the needed computational effort into a power consumption figure of merit, under the assumption of an almost standard mobile platform implementation. Results demonstrate that rakeness-based design is useful for a non-negligibly power consumption reduction not only at the transmitter node but also at the receiving local gateway.

The paper is organized as follows. Section II quickly recaps the CS mathematical background including details on the rakeness approach, while Section III illustrates the rakeness impact in the decoding of synthetic ECGs, when different approaches are employed. Results in decoding ECGs on a mobile platform in Section IV. Finally, we draw the conclusion.

[1]see, for example, that described at http://cs.signalprocessing.it.

II. Basics of Compressed Sensing

In this paper we adopt the discrete-time formulation of CS, where the waveform to be acquired in a given time window is represented by a set of n Nyquist-rate samples collected in a *signal* $x = (x_0, \ldots, x_{n-1})^\top \in \mathbb{R}^n$. The key assumption of CS is *sparsity*, that is the existence of a n-dimensional *sparsity basis* or *dictionary*, $\Psi \in \mathbb{R}^{n\times N}, N \geq n$ (where $n = N$ holds only for the basis case) in which any instance of the signal $x = \Psi\alpha$ is represented by $\alpha \in \mathbb{R}^N$. For the sparsity prior to hold, the coefficient vector α must be *κ-sparse*, i.e., have at most $\kappa \ll n$ non-zero components in its support.

In this signal model the actual number of degrees of freedom in x is considerably smaller than n. Leveraging this property, fundamental results [1] have shown that its salient information content can be captured in a set of $m < n$ linear *measurements*. These measurements are gathered in the m-dimensional vector $y = (y_0, \ldots, y_{m-1})^\top \in \mathbb{R}^m$, as obtained by applying a *projection matrix* $A \in \mathbb{R}^{m\times n}$ to x, i.e.,

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

where ν takes into account all possible nonidealities.

Formal results [1], [2] guarantee that α (and thus x) can be recovered from y despite the fact that A (and thus $A\Psi$) is a dimensionality reduction, provided that $m = \mathcal{O}(\kappa \log n)$ and A obeys some requirements that are most likely satisfied when it is drawn at random. Interestingly, these requirements can be satisfied by considering A 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 cumbersum full multipliers are not required anymore, and represents a key point in the design of effective and parsimoniuos CS stages for biomedical sensing nodes [5]. In the following, we always implicitly assume that A is antipodal.

Roughly speaking, the rationale behind all these guarantees is that generic, κ-sparse vectors are mapped *almost isometrically* [6] 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.

From a mathematical point of view [1], [7], signal reconstruction happens by solving dedicated optimization problems looking for the sparsest coefficient vector consistent with measurement. In detail, input signal x is reconstructed as $\hat{x} = \Psi\hat{\alpha}$, where $\hat{\alpha}$ is the sparsest α subject to constraints forcing the corresponding measurements to be as close as possible to the observed y. Sparsity is generally promoted by the ℓ_1 norm instead of the computationally intractable count of non-zero components given by ℓ_0 norm. Along this path at least three methods are commonly employed.

The first method is called basis pursuit (BP) and simply computes $\hat{\alpha}$ by neglecting disturbances to solve

$$\min_\alpha \|\alpha\|_1 \ \text{ s.t. } \ A\Psi\alpha = y$$

where $\|\cdot\|_p$ indicates the usual ℓ_p norm. The main appeal of BP is that it can be recast into a fully linear optimization problem for which standard methods exists though, ad hoc techniques have been developed. The second method is called basis pursuit with denoising (BPDn) and takes into account disturbances solving

$$\min_\alpha \|\alpha\|_1 \ \text{ s.t. } \ \|A\Psi\alpha - y\|_2^2 \leq \epsilon^2$$

where ϵ^2 is tuned on the characteristics of the disturbance term ν. A last method keeps the *denoising* formulation of BPDn but focuses on directly on the target vector x instead of α. This can yield definite advantages in the quality of reconstruction when Ψ is a dictionary, i.e., $N > n$. In that case one solves

$$\min_x \|\Psi^* x\|_1 \ \text{ s.t. } \ \|Ax - y\|_2^2 \leq \epsilon^2$$

where Ψ^* is a *analysis transform* operator [8] that for every x chooses one of the possible representations with respect to Ψ. This problem is called analysis BPDn (ABPDn) and is of interest here since the change of point of view from α to x may imply a different computational burden even if Ψ is a basis and thus $\Psi^* = \Psi^{-1}$.

Further to these methods relying on sparsity promotion by means of the ℓ_1 norm, other greedy approaches exists that iteratively promote sparsity by observing intermediate and approximate solutions. Implementations of CS decoding on embedded, low-resources platforms usually look into this set of methods rather than feed a solver with a suitably defined optimization problem.

In all cases, quality of reconstruction depends on m, i.e., on the amount of information that is passed from the encoder to the decoder. Since the same m is also related to the *compression ratio* n/m and thus to the saving that one may experience at the transmitter when using CS, its minimization is of paramount importance.

This is what rakeness-based design [4] does: it improves sensing performance by generating each row of A independently of the others, but with entries whose correlation is adapted to the second-order statistic of x. Interestingly, rakeness-based design is compatible [9], [10], [11] with the hardware-friendly constraint of having A made only of antipodal symbols.

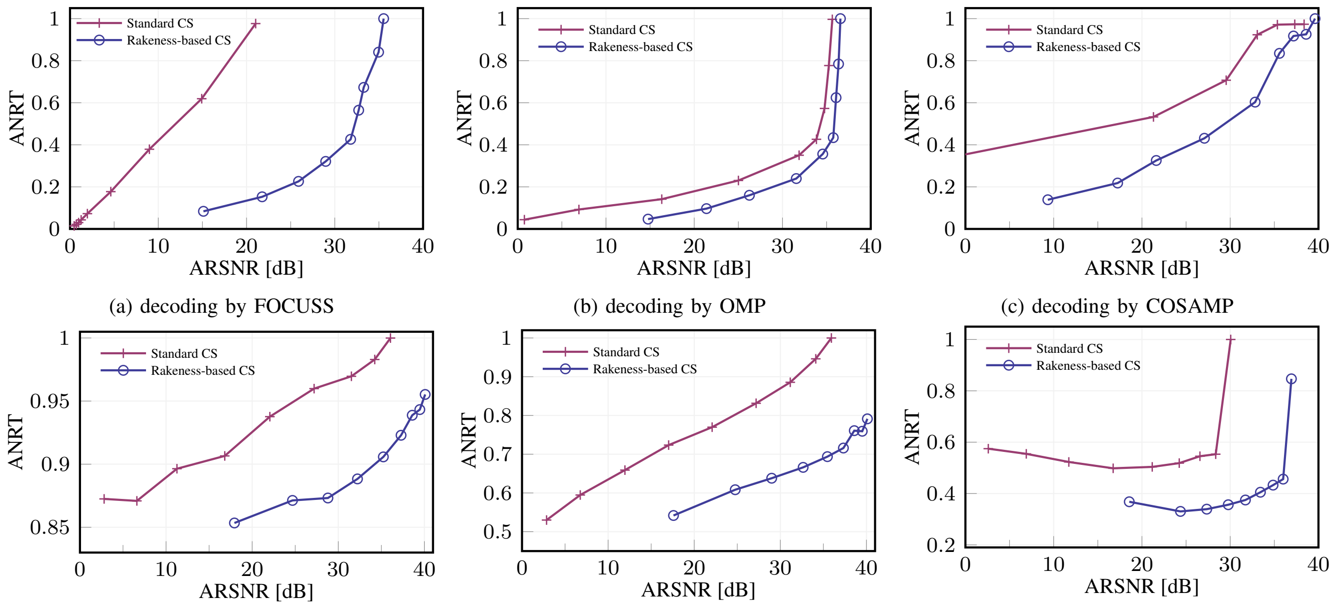
What we do here is to change the point of view and consider the effect of rakeness-based design at the decoder: measurements have been computed, sent and received and the local gateway must reconstruct the original signals complying with a power budget that is larger than that of sensing nodes, but still limited. To see that rakeness-based design positively affects also this component of the overall system consumption, we consider implementations of BP, BPDn, ABPDn and greedy algorithms, experimentally verifying that the reduction in m, and thus in the number of rows of A is always beneficial.

III. Experimental setting and Results

In this paper we focus on reconstruction algorithms that do not rely on large-scale, general-purpose solvers but that are all specialized to the task. More in detail, for BP and BPDn families we consider the dedicated functions in the SPGL1 package [14] `spgl_bp` and `spgl_bpdn`. For the ABPDn we consider NESTA [15]. Despite theoretically possible, we do not apply NESTA also to BP and BPDn since in these cases the algorithm requires A to have orthogonal rows, and this goes against the hardware-friendly philosophy of antipodal entries. Within the family of greedy methods, we consider FOCUSS [16], OMP [17], and CoSaMP [7].

In order to compare the performance of the reconstruction algorithms of the standard CS approach with that of the rakeness-based one, some MATLAB Montecarlo simulations have been performed. For all the aforementioned decoding algorithms, and for both CS approaches (standard and rakeness-based) a set of 100 synthetic ECG instances generated as in

This is the author's version of the article that has been presented at IEEE BIOCAS2016
The editorial version of the paper is available at http://dx.doi.org/10.1109/BioCAS.2016.7833804



(a) decoding by FOCUSS (b) decoding by OMP (c) decoding by COSAMP

(d) decoding by SPGL1 (BP) (e) decoding by SPGL1 (BPDn) (f) decoding by NESTA (ABPDn)

Fig. 2. Comparisons between standard and rakeness-based for different decoding implementations where each plot reports the ANRT as a function of the quality of service (counted by the ARSNR) for ECGs reconstruction

[4] with a sampling rate equal to 360 Hz has been considered. Each instance is composed by $n = 512$ consecutive samples and it has been encoded and decoded with a different A. For all the considered reconstruction algorithms Ψ has been assumed equal to the orthonormal Symlet-6 wavelet basis [12] including the ABPDn implementation in NESTA, for which it follows that $\Psi^* = \Psi^{-1}$. Finally, to emulate possible nonidealities of the sampling stage, we inject an additive white Gaussian noise corresponding to a 40 dB signal to noise ratio[2].

Since our interest is focused on the difference between using the standard CS approach and rakeness-based one and not on the different reconstruction algorithms performance, we propose as figure of merit the average normalized reconstruction time (ANRT) defined as follows. For each decoding approach, we compute the average CPU time required for signal reconstruction when using a given m and a given CS approach (i.e., either standard or rakeness-based). This time is normalized with respect to the slower one observed, that is generally the one associated to the higher m. This normalization allows us to assume that this figure of merit is almost independent of the used hardware.

Furthermore, ANRT is not considered as a function of m, but as a function of the quality of service, i.e., of the quality of the reconstructed signal $\hat{x}$ at the given m. More formally, the quality indicator we use is 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 $\mathbf{E}_{A,x}$ stands for averaging over all considered A and all considered instances of the input signal x in the Montecarlo.

The configuration setting adopted for each algorithm is discussed in the following. FOCUSS runs has been made with a setting optimized for ECG and described in [13]. OMP worked with a stop criteria based on the residual error associated to each iteration and counted by $r = \|y - A\Psi\hat{\alpha}\|_2^2$, with $r < 10^{-4}$. The greedy approach CoSaMP takes as input an estimation of the sparsity level κ which was fixed to $\min\{\lfloor m/3 \rfloor, 50\}$[3]. BP does not posses any parameter to be tuned on noise level perturbing the measurement vector, while for both BPDn and ABPDn we set $\epsilon^2 = 10^{-4}$.

The obtained results for all aforementioned decoding approaches are shown in plots composing Fig. 2 and clearly indicate that the rakeness-based CS outperforms the standard approach for every tested algorithm and for all ARSNR values. The CPU time measured when using a rakeness approach is always lower (sometimes up to 90%) than the time required by the standard approach.

IV. Energy Efficient Reconstruction

In the general case, the computational time may be considered as a first order estimation of the energetic requirements. Despite that, with the aim of proving the advantages of the rakeness approach in a real system, we propose in this section a detailed evaluation of the power consumption of one of the reconstruction algorithms presented above on a representative embedded platform. Specifically, we profiled the FOCUSS algorithm on the Hardkernel Odroid-XU3 board[4], an evaluation board based on Samsung Exynos 5422, a representative multi-core CPU found in 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 floating point performance.

[2]the level of injected noise has been estimated by averaging quantization noise measured in many databases available online.

[3]the CoSaMP implementation proposed in [7] impose $k > m/3$.

[4][ONLINE] http://www.hardkernel.com.

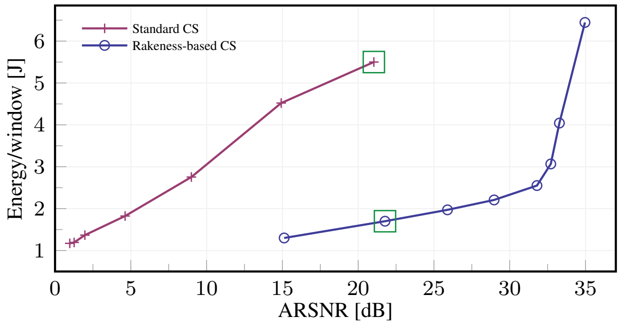


Fig. 3. Energy vs. ARSNR of the FOCUSS algorithm running on ODROID Cortex-A15. Comparison between rakeness-based and standard CS.

The FOCUSS algorithm, previously introduced in Section III, was 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` version 4.8.2 and, to achieve an efficient algorithm implementation, the Armadillo [18] library (v. 4.2) was used for linear algebra. To measure the energy consumption we deploy the on-board voltage/current sensors and split power rails, which allow to measure separately the power consumption of Cortex-A15 cores, Cortex-A7 cores, GPU and DRAM. The readout of the sensors was implemented in a low-priority thread, with a sampling interval of 25ms and an average CPU consumption below 3%.

Fig. 3 shows the results of our evaluation, comparing the energy required by FOCUSS to reconstruct a window of ECG samples considering both standard and rakeness-based CS. The energy measurements, obtained by considering the most performing operating point[5] (Cortex-A15 at 1.9GHz) are coupled to the respective ARSNR. Clearly, the rakeness-based CS outperforms the standard CS in terms of energy efficiency. For instance, when designing the system for a target ARSNR=$22\,\mathrm{dB}$ there is a factor of ≈ 3.7X in terms of energy consumed (as highlighted in Fig. 3). On a battery-powered device, such as a typical WBSN gateway, this translates in a consistent battery life extension.

Finally, for a visual check of the correctness of the developed approach, a short chunk of ECG signal reconstructed at the target ARSNR and compared with the input signal is depicted in Fig. 4. To achieve a $22\,\mathrm{dB}$ ARSNR, it is necessary to set $m = 256$ in the standard CS approach, with a compression ratio equal to 2, while $m = 78$ is enough for the rakeness-based case, introducing a compression ratio equal to 6.56. Note that the gain in terms of compression ratio achieved by the rakeness approach is ≈ 3.2X, that is actually similar to the gain in terms of saved energy.

V. Conclusion

Benefits introduced by rakeness-based CS at the decoder side was discussed with simulation on ECGs decoding showing a reduction in the ANRT across different reconstruction algorithm. By implementing one of them on a real mobile platform, we also have shown a non negligible reduction in terms of energy for the reconstruction of each time window.

[5]such operating point leads to the best performance (i.e., real-time reconstruction guaranteed) but also to highest power consumption[13]. There is a clear room for improvements with respect to energy consumption, by tuning the algorithmic implementation as well as by using less precision.

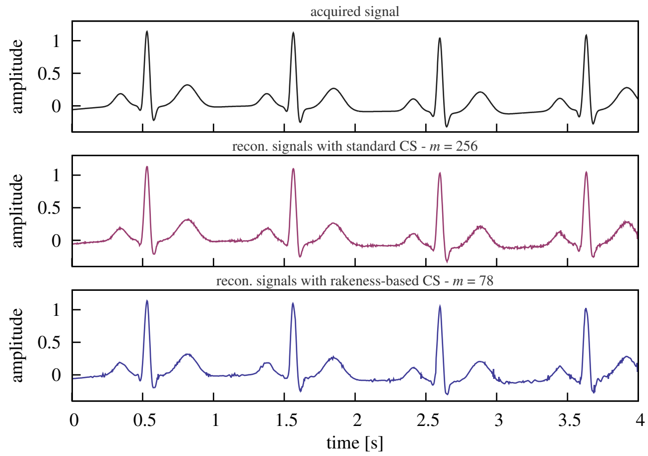


Fig. 4. Chunks of the considered ECG signal with the corresponding reconstructed ones for standard CS ($m = 256$ with compression ratio equal to 2) and for rakeness-based CS ($m = 78$ with comp. ratio equal to 6.56).

These results propose the rakeness approach as a good candidate for power saving in all scenarios presenting a battery powered device running a proper decoding algorithm.

References

[1] E.J. Candes, T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005
[2] D.L.Donoho, "Compressed Sensing," IEEE Transactions on Information Theory, vol. 54, n. 4, pp. 1289–1306, 2006
[3] M. Mangia, D. Bortolotti, A. Bartolini, F. Pareschi, L. Benini, R. Rovatti and G. Setti, "Long-Term ECG monitoring with zeroing Compressed Sensing approach," Nordic Circuits and Systems Conference (NORCAS), pp. 1-4, 2015
[4] M. Mangia, R. Rovatti, G. Setti, "Rakeness in the Design of Analog-to-Information Conversion of Sparse and Localized Signals," IEEE Transactions on Circuits and Systems I, vol. 59, no. 5, pp. 1001-1014, May 2012
[5] F. Pareschi, P. Albertini, G. Frattini, M. Mangia, R. Rovatti, G. Setti, "Hardware-Algorithms Co-design and Implementation of an Analog-to-Information Converter for Biosignals based on Compressed Sensing," *IEEE Trans. on Biomed. Circuits Syst.*, vol. 10, no. 1, pp. 149-161. February 2016.
[6] R. Baraniuk, M. Davenport, R. DeVore, M. Wakin, "A Simple Proof of the Restricted Isometry Property for Random Matrices," Constructive Approximation, vol. 28, no. 3, pp. 253-263, 2008
[7] D. Needell and J.A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples", Applied and Computational Harmonic Analysis, v. 26, no. 3, pp. 301–321, 2009
[8] Candes, E. J., Eldar, Y. C., Needell, D., Randall, P. "Compressed sensing with coherent and redundant dictionaries," Applied and Computational Harmonic Analysis, 31(1), 59-73, 2011
[9] G. Jacovitti, A. Neri and G Scarano, "Texture synthesis-by-analysis with hard-limited Gaussian processes", IEEE Transactions on Image Processing, v. 7, no. 11, pp. 1615–1621, 1998
[10] A. Caprara, F. Furini, A. Lodi, M. Mangia, R. Rovatti, G. Setti, "Generation of Antipodal Random Vectors With Prescribed Non-Stationary 2-nd Order Statistics," IEEE Transactions on Signal Processing, vol. 62, n. 6, pp 1603-1612, 2014
[11] R. Rovatti, G. Mazzini, G. Setti, "Memory-m Antipodal Processes: Spectral Analysis and Synthesis", IEEE Transactions on Circuits and Systems - I, vol. 56, n. 1, pp. 156-167, 2009
[12] S. Mallat, *A wavelet tour of signal processing: the sparse way*, Elsevier
[13] D. Bortolotti, M. Mangia, A. Bartolini, R. Rovatti, G. Setti and L. Benini, "Rakeness-based compressed sensing on ultra-low power multi-core biomedicai processors", International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC-15), pp. 1-6, 2015
[14] E. Van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions", SIAM Journal on Scientific Computing, v. 31, no. 2, pp. 890–912, 2008
[15] S. Becker, J. Bobin and E. J. Candes, "NESTA: A Fast and Accurate First-Order Method for Sparse Recovery", SIAM Journal on Imaging Sciences, v.4, no. 1, pp. 1–39, 2011
[16] I. F. Gorodnitsky and D. R. Bhaskar, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm", IEEE Tran. on Signal Processing, v. 45, no. 3, pp. 600–616, 1997.
[17] J. A. Tropp and A. C. Gilbert, "Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit", IEEE Tran. on in Information Theory, v. 53, no. 12, pp. 4655–4666, 2007
[18] Sanderson, C. "Armadillo: An open source C++ linear algebra library for fast prototyping and computationally intensive experiments", NICTA Technical Report, 2010