Correlation tuning in compressive sensing based on rakeness: A case study

This is the author's version of the article that has been presented at IEEE ICECS2013
The editorial version of the paper is available at http://dx.doi.org/10.1109/ICECS.2013.6815403

Correlation Tuning in Compressive Sensing based on Rakeness: a case study

Nicola Bertoni†, Mauro Mangia*, Fabio Pareschi†,* Riccardo Rovatti‡,*, Gianluca Setti†,*
* ARCES – University of Bologna, via Toffano 2/2, Bologna, Italy. mmangia@arces.unibo.it
† ENDIF – University of Ferrara, via Saragat 1, Ferrara, Italy. nicola.bertoni@student.unife.it,{fabio.pareschi, gianluca.setti}@unife.it
‡ DEI – University of Bologna, viale Risorgimento 2, Bologna, Italy. riccardo.rovatti@unibo.it

***Abstract*—In this paper we take into account the *rakeness* approach in the design of Compressed Sensing (CS) based system, which allows, by means of the matching of some statistical properties of the CS sampling functions with statistical features of the input signal, to greatly increase system performance in terms of either a reduction of resources (hardware, energy, etc) required for the signal acquisition or an increase in the acquisition quality. In particular, with respect to the general formulation, we make two additional and non-restrictive hypotheses to ensure a good behavior of the system. With these, we can compute an upper and a lower bound for the parameter r used to control the statistical matching level, and we show with some numerical examples that the choice of r is not critical. In particular, any r value taken from the computed interval ensures almost optimal performance, making the rakeness approach robust and worthwhile.**

I. Introduction

In recent years, along with standard and general purpose Nyquist-rate Analog-to-Digital (A/D) converters, some application-specific Analog-to-Information (A2I) converters have been proposed, aiming at reducing the amount of required resources (hardware, time, energy, etc.) per conversion. Their design and implementation is based on the idea of the *Compressive Sensing* (CS) [1]–[3], [8], that is a paradigm which links the possibility of acquiring a signal by a number of measurements that depends on the actual information of the signals rather than its bandwidth. Many different CS techniques, as well as many hardware implementations and algorithms for signal reconstruction have been so far proposed [3]; in all of them the fundamental concept is to have an *a priori* knowledge on some properties of the signal [2].

We contributed to the general CS theory in some recent works [4]–[6] with the introduction of the concept of *rakeness*. The basic idea behind this is to exploit *localization* of signals, i.e., the assumption that the information of the signal is not equally distributed on its domain. Roughly speaking, localization implies that some realizations of the input process have a higher probability with respect to other ones. Note that, even if this is not in accordance to the common sense, the assumption of localization is not a limiting one. In fact, the only class of signals where all possible realizations have the same probability is white noise [6].

In more detail, one of the CS assumptions is that measurements are obtained trough the projection of the input signal on random *independent and identically distributed* sampling functions. In [4], [6] we showed that this is just a sufficient condition, and using random sampling functions which are not independent but which have a limited correlation (depending on statistical properties of the input signal) let us increase system performance, thus allowing us either to get a higher quality of the reconstructed signal, or a reduced number of measurement necessary to get a given reconstruction quality. The core idea of this approach is to solve an optimization problem to maximize the "raked" energy by the encoder where the randomness level of the sampling function is tuned by a parameter r.

The aim of this paper is to show that the tuning of r is not a critical issue in the design of a rakeness-based CS system. In particular, the performance of the system is almost unchanged when the value of r is chosen from a given set which can be computed in accordance to analytical considerations. This is fundamental since it ensures that a mismatch between the designed value of r and its actual value due, for example, to some parameters drift, does not alter the expected system performance.

The paper is organized as follows. In Section II we recall the basic concepts and notation for a CS system. In Section III we introduce the concept of correlation tuning, as well as the parameter r and its range of interest. Then, in Section IV we present our case study involving CS on small images. Finally, we draw the conclusion.

II. CS System Model

In this section we want to recall some basic concepts of CS systems, mainly to introduce the notation used in the paper. A more comprehensive analysis can be found in [1] or in [6]. Even if many parts of this section (and of the following one) have a more general validity, we focus on the Random Modulation Pre-Integration (RMPI) approach, first proposed in [7]. This approach, when compared to other ones, has the great advantage of being able to cope with any class of signal without performance degradation [3]. The RMPI basic schematic is depicted in Fig. 1 and can be briefly described as follows. Given an input signal $x(t)$ defined for $0 \leq t \leq T$, RMPI collects a set of measurement y_j, $j = 1, 2, \dots m$, by projecting x on a set of sampling functions $\phi_j(t)$, i.e., $y_j = \langle \phi_j(t), x(t) \rangle$.

For the sake of simplicity in the following we will make the common and not limiting assumption that $x \in \mathbb{R}^n$. This may stand, for example, for a discrete-time input signal where n corresponds to the amount of Nyquist samples in a T-length time window, but also allow us to extend our model to other different kinds of signals such as n-bit images or even generic data. Following this, the $\phi_j \in \mathbb{R}^n$ assume the role of *sampling vectors*, and can be rearranged as the rows of a matrix $\Phi \in \mathbb{R}^{m \times n}$ called *sampling matrix*. Under this notation, all measurements y_j can be joined in a single measurement vector $y = \Phi \cdot x$, with $y \in \mathbb{R}^m$. According to this model, we have *compression* when $m < n$.

The fundamental hypothesis for achieving compression is to deal with a *sparse* signal. Mathematically, given a proper basis $\Psi \in \mathbb{R}^{n \times n}$, each realization $x = \Psi \cdot \alpha$ of the input signal is such that the coefficients vector α contains only a limited number of non-null entries. More precisely, all realizations are represented by a maximum number $K \ll n$ of coefficients. In this case the input signal is called K-sparse.

If we we also consider, for a more realistic model, that input signal has additive noise ν, we get

$$y = \Phi \cdot (x + \nu) = \Phi \cdot \Psi \cdot \alpha + \Phi \cdot \nu = A \cdot \alpha + \Phi \cdot \nu$$

where $A \in \mathbb{R}^{m \times n}$ represents the link between the input signal sparse representation and the measurement vector.

From the knowledge of y, one can recover α (and so x) by the A operator inversion. This however is an ill-posed problem, that can be overcome by solving the optimization problem based on the sparsity hypothesis [1], [2], [8]–[10]

$$\hat{\alpha} = \min_{\alpha} \|\alpha\|_1 \quad \text{s.t.} \quad y = A \cdot \alpha \tag{1}$$

This is the author's version of the article that has been presented at IEEE ICECS2013
The editorial version of the paper is available at http://dx.doi.org/10.1109/ICECS.2013.6815403

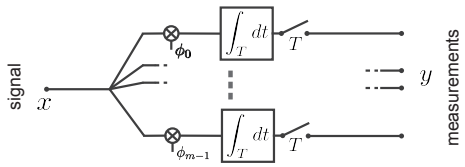


Fig. 1. Basic architecture of a RMPI-based A2I system.

where $\|\cdot\|_p$ stands for the standard p norm.

The reconstructed signal is $\hat{x} = \Psi \cdot \hat{\alpha}$ [11]. CS theory guarantees an accurate reconstruction (i.e., $\hat{x} \approx x$) by choosing $m \geq 4K \log_{10}(n/K)$ [3].

The robustness of this procedure is ensured by two guidelines. These are *incoherence*, that guarantees information extraction for any realization of the input signal, and *restricted isometry*, that ensures robustness with respect to noise, more precisely that the signal-to-noise ratio on y is approximatively that of x. A complete description of both guidelines is in [1], [8]. As we have already introduced, the standard CS approach takes all the ϕ_j as a random vectors made by independent and identically distributed elements from a normal distribution [8], since this is the most simple way to ensure both incoherence and restricted isometry independently of Ψ.

III. Correlation Tuning

In our previous works [4]–[6] we investigated a different way for Φ generation. By defining

$$\rho(\underline{a}, \underline{b}) = \mathbf{E}_{a,b}\left[|\langle a, b\rangle|^2\right] \tag{2}$$

where $\underline{a}$ and $\underline{b}$ are two stochastic processes, and a and b their realizations, we define *rakeness* the expectation of the energy collected by a sampling vector. By indicating with $\underline{x}$ and $\underline{\phi}$ the processes generating the input signal and the sampling vectors, this energy is given by $\rho(\underline{\phi}, \underline{x})$ The basic idea is to generate all sampling vectors in accordance to a process $\underline{\phi}$ which maximizes the collected energy and at the same time is "random enough" to ensure the collection a minimum amount of energy even from the least common realizations of $\underline{x}$. This intuitive approach correspond to following optimization problem

$$\begin{array}{ll} \max_{\underline{\phi}} & \rho(\underline{\phi}, \underline{x}) \\ \text{s.t.} & \langle \underline{\phi}_j, \phi_j \rangle = e \\ & \rho(\underline{\phi}, \underline{\phi}) \leq re^2 \end{array} \tag{3}$$

where e is the energy of each sampling vector and r is a design parameter that can be tuned to set the randomness of $\underline{\phi}$. More considerations about this approach are discussed in details on [6].

In order to exploit (3) into a useful form, we can consider the correlation matrices C_x and C_ϕ of, respectively, the two stochastic processes $\underline{x}$ and $\underline{\phi}$. Under some non-restrictive assumptions on these two matrices [6] (including that they are Hermitian and positive semidefinite) we can write their elements $\{C_x\}_{t,s}$ and $\{C_\phi\}_{t,s}$ as

$$\{C_x\}_{t,s} = \sum_{j=0}^{n-1} \mu_j \theta_j^*[t]\theta_j[s], \quad \{C_\phi\}_{t,s} = \sum_{j=0}^{n-1} \lambda_j \gamma_j^*[t]\gamma_j[s] \tag{4}$$

where $\theta_j[\cdot]$, $\gamma_j[\cdot]$ are the elements of the eigenvectors θ_j, γ_j, and the sequences of real non-negative numbers $\mu_0 \geq \mu_1 \geq \mu_{n-1} > 0$ and $\lambda_0 \geq \lambda_1 \geq \lambda_{n-1}$ are the corresponding eigenvalues. Note that the assumption $\mu_{n-1} > 0$, which is an additional one with respect to [6], ensures that the signal dimensionality is actually n.

Now, solving (3) for a given r value means finding the optimum λ_j and γ_j given the knowledge of $\underline{x}$, i.e. of μ_j and θ_j. The analytical solution (see [6] for details) is given by

$$\gamma_j = \theta_j \tag{5}$$

$$\lambda_j = \frac{e}{n}\left[1 + \frac{n\mu_j - \Sigma_1}{\sqrt{\frac{\Sigma_2 - \frac{1}{n}\Sigma_1^2}{r - \frac{1}{n}}}}\right] \tag{6}$$

where $\Sigma_1 = \sum_{j=0}^{n-1} \mu_j$ and $\Sigma_2 = \sum_{j=0}^{n-1} \mu_j^2$. With the obtained λ_j and γ_j we can use (4) to design the optimum $\underline{\phi}$ process. Note that this is a simplified representation with respect to the more general case considered in [6] when we consider only solutions of (3) that ensure that all λ_j are greater than zero. This assumption preserves the ability of sampling vectors to span the whole input signal domain.

The main contribution of this paper is to show with numerical evidences that, under the two above additional assumptions with respect to [6], the effect of the rakeness is always positive and almost independent on the value of r.

The first results we show is that the above additional assumptions imply a new upper bound for r. Using the constraint $\lambda_{n-1} > 0$ in (6) gives rise to

$$r < \frac{\Sigma_2 - 2\Sigma_1\mu_{n-1} + n\mu_{n-1}^2}{(\Sigma_1 - n\mu_{n-1})^2} \tag{7}$$

On the opposite side, the lower bound for r corresponds to the case where all ϕ_j are composed by random independent and identically distributed symbols, as suggested in the classical CS theory following both incoherence and restricted isometry. Here all λ_j have the same value $\lambda_j = \frac{e}{n}$, so that the r lower bound is given by.

$$r > \frac{1}{n} \tag{8}$$

In the next section, supported by a case study, we will show system performance in terms of average reconstructed SNR (ARSNR) when r is swept in the above computed range.

IV. Results

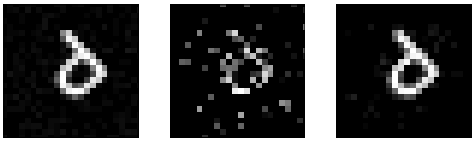
The class of signals we consider in this section is composed by black and white, 24×24-pixel images. Each instance represents a small white printed number or letter on black background (gray-level dithering is allowed), where numbers and letters are randomly rotated and displaced from the center of the image. The color of each pixel has been coded with a value ranging from 0 (black) to 1 (white). All pixel values of an image have been unrolled in a single vector with size $n = 24 \cdot 24 = 756$ pixel, which has been considered as input signal in a CS system based on RMPI. This signal class is actually sparse, since each instance image has no more than about 85 non-black pixels. The considered sparsity basis is the canonical one. Furthermore, to make the system more realistic, we added a white additive Gaussian noise ν to all images to get a 30 dB SNR.

The correlation matrix of this class of signals has been estimated by the statistical analysis of a wide collection of different instances. The eigenvalues and eigenvectors μ_j and θ_j has been computed and used in (7) and (8) to get the following upper and lower bound for r

$$r_{min} = 0.001736, \quad r_{max} = 0.01911 \tag{9}$$

This test signal has been used to compare performance of a standard CS system and a rakeness-based CS system with different r values. In the first case, measurement vectors y has been obtained by projecting the unrolled input vector on normalized (i.e., $e = 1$) independent Gaussian sampling vectors ϕ_j with zero mean and variance $\frac{1}{n}$. In the second one the ϕ_j has been achieved as realizations of multivariance Gaussian process with correlation matrix in accordance to the

This is the author's version of the article that has been presented at IEEE ICECS2013
The editorial version of the paper is available at http://dx.doi.org/10.1109/ICECS.2013.6815403



(a) (b) (c)

Fig. 2. Example instance of the considered class of signals (a) and its reconstruction made with $m = 106$ measurements using both standard CS approach (b) and rakeness based approach with $r = 0.008$ (c).

solutions (5) and (6) of rakeness optimization problem, obtained for a given value of r. An example image, along with the reconstructed ones using both standard CS and rakeness-based CS is depicted in Fig. 2.

According to the optimization problem (3), the aim of the rakeness is to increase the energy collected by the measurements at the cost of introducing correlation in the ϕ_j vectors. Since an increase in r relaxes the constraints in the optimization problem (3), we can expect a non-decreasing relation between the collected energy and r. This is confirmed by Fig. 3, where we have plotted the average energy of the measurement vector, computed as $\|y\|_2$, for r ranging in set defined by (9), compared to what we get in the standard CS approach based on independent measuring vectors and corresponding to the case $r = r_{min}$. The number of measurements in this example has been fixed to $m = 180$, which has been chosen since it ensures a correct signal reconstruction for the rakeness-based approach.

While increasing r implies increasing $\|y\|_2$, and this has certainly a positive influence for the overall system, we have also to cope with the negative effect given by the violation of the incoherence and restricted isometry requirements. For this reason, the connection between the r value and system performance is not clear and intuitive due to the strong non-linearity of the reconstruction process.

However, if we limit ourselves to values of r in the above computed range, the effect of the rakeness is always positive and furthermore almost independent on the chosen value of r. In Fig. 4 we plotted the ARSNR for the rakeness-based CS system, for r ranging in its definition set, compared to what we get in the standard CS approach. As in the previous case, we have $m = 180$. Excluding cases when r is similar to r_{min} (where the system behavior is very similar that of a standard CS system as expected) and to r_{max}, system performance is always improved by almost 7 dB. Furthermore, this increment has a very weak dependence on r. This is an extremely important result, since guarantees that the choice of r is *not critical* in the system design.

This effect can be intuitively explained by observing Fig. 5, where we have plotted the probability distribution function (PDF) obtained for $\|y\|_2$ with some different values of r. In the figure it is clear that, when r assumes values in the right neighborhood of the minimum value (compare, in the figure, the PDF associated to $r_{min} \approx 0.0017$ and $r = 0.002$) there is a remarkable difference between the two distributions. However, for larger values of r (compare the cases $r = 0.008$ and $r = 0.014$), despite the fact that the mean value is actually increasing as observed in Fig. 4, the difference between the two PDFs is not so evident.

In conclusion, since the choice of r has been shown to be not critical, any value taken from the mid-range of the above computed set ensures near-optimum system performance. For example, in the design of the system considered in this paper, a good choice could be $r = 0.008$. Assuming this value, we can show some system performance to confirm the advantages of the rakeness-based CS approach with respect to the standard one.

In Fig. 6 we compare, for different values of m, the ARSNR of the standard CS approach with respect to the one achieved

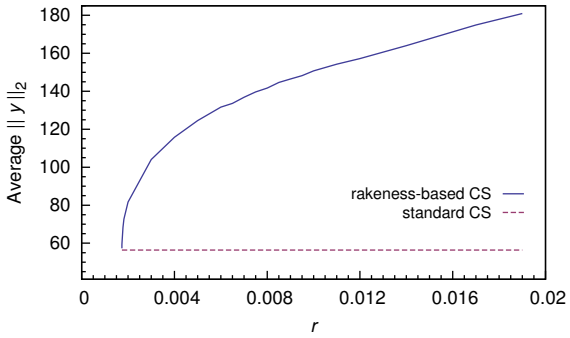


Fig. 3. Average energy $\|y\|_2$ of the measurement vector y for different values of r and for $m = 180$, compared with the energy obtained according to the standard CS approach.

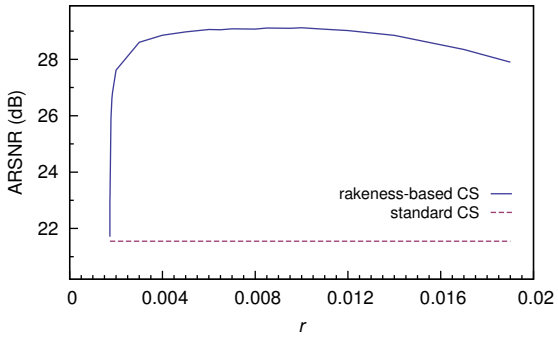


Fig. 4. Average reconstruction SNR achieved by the rakeness-based CS system for different values of r and for $m = 180$, compared with the energy obtained according to the standard CS approach.

by the rakeness-based approach with the above value of r. This comparison were done by considering two different encoding strategies, optimization based reconstruction and iterative support-guessing reconstruction [3]. Where the second is an iterative approach that suppose knowledge about the sparsity level of the acquired signal. We report both decoding strategies to highlight that advantage introduced by rakenees approach are not decoding algorithm depending. In the rest we will consider only results obtained by optimization based reconstruction. It is clear that, for optimum performance (i.e., for ARSNR closed to the 30 dB upper bound given by the input additive noise) we need m greater than approximatively 160 for the rakeness-based system, while we need m greater than approximatively 300 for the standard approach. This corresponds to a gain in the compression rate $\frac{m}{n}$ from 1.9 (standard approach) to 3.6 (rakeness-based approach).

As an additional figure of merit, we can compare performance of the two CS systems (standard and rakeness-based) not only in terms of average reconstruction SNR, but in terms of the reconstruction SNR distribution. Results are shown in Fig. 7 and can be commented as follows. In case (a) we have $m = 106$, which is shown to be insufficient to ensure that the input signal is correctly reconstructed. In the standard approach only a limited number of instances are correctly reconstructed, while in the rakeness-based approach the number of correctly reconstructed instances is much higher, but still insufficient. On the contrary, in case (b) the value $m = 160$ is enough to ensure correct reconstruction, but only in the rakeness-based approach. In the case (c) the m value is raised to 213, which still does not ensure correct reconstruction for the standard approach, since we can notice the presence of many input instance with a very low reconstruction SNR.

This is the author's version of the article that has been presented at IEEE ICECS2013
The editorial version of the paper is available at http://dx.doi.org/10.1109/ICECS.2013.6815403

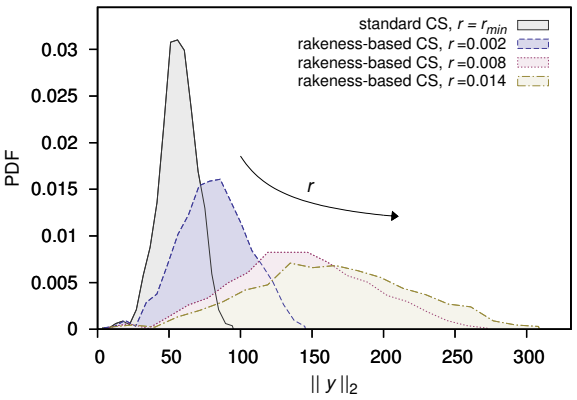


Fig. 5. PDF of the energy $\|y\|_2$ captured in the sampling for a standard CS system (solid contour plot) and for rakeness-based CS with $r = 0.002$ (dashed contour plot), $r = 0.008$ (dotted contour plot) and $r = 0.014$ (dot-dashed contour plot).

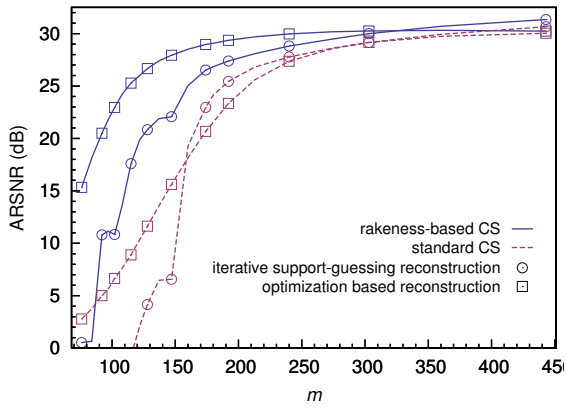
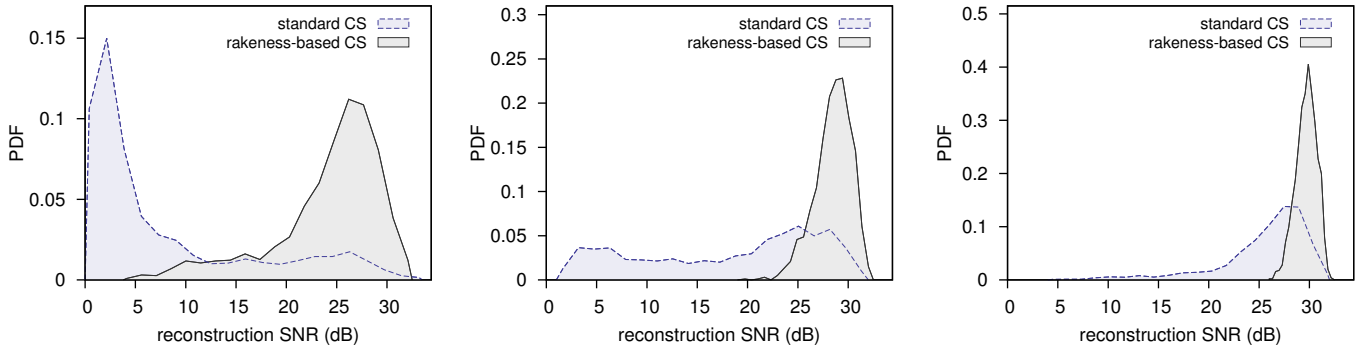


Fig. 6. ARSNR as a function of m for the standard CS system, and for the rakeness-based one for $r = 0.008$ considering both iterative support-guessing reconstruction and optimization based reconstruction.



(a) (b) (c)

Fig. 7. PDF of the reconstruction SNR for the standard CS approach and the rakeness-based CS approach with $r = 0.008$ in the three cases $m = 108$ (a), $m = 160$ (b), and $m = 213$ (c).

V. Conclusion

In this paper we have taken into account the rakeness approach in the design of CS based system and, with respect to the general formulation, we have made two additional and non-restrictive hypotheses to ensure a good behavior of the system. This allowed us to compute an upper and a lower bound for the parameter r, which actually is the only one used in the design of rakeness-based systems and which is used to control the statistical matching level between the input signal x and the sampling functions ϕ_j.

With a numerical analysis of a case study we have been able to prove that the choice of r is not critical. In particular, any r value taken from the computed interval ensures almost optimal performance. In this way, the rakeness approach has been shown to be robust and worthwhile in real systems.

References

[1] D. L. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[2] E. J. Candes, J. K. Romberg, K. Justin, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[3] J. Haboba, M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "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, no. 3, Sep. 2012.

[4] M. Mangia, R. Rovatti, and G. Setti, "Analog-to-information conversion of sparse and non-white signals: Statistical design of sensing waveforms," in *Proceedings of 2011 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2011, pp. 2129–2132.

[5] M. Mangia, J. Haboba, R. Rovatti, and G. Setti, "Rakeness-based approach to compressed sensing of ECGs," in *Biomedical Circuits and Systems Conference (BioCAS), 2011 IEEE*, Nov. 2011, pp. 424 –427.

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

[7] J. N. Laska, S. Kirolos, M. F. Duarte, T. S. Ragheb, R. G. Baraniuk, and Y. Massoud, "Theory and Implementation of an Analog-to-Information Converter using Random Demodulation," in *Proceedings of 2007 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2007, pp. 1959–1962.

[8] E. J. Candes and M. B. Wakin, "An Introduction to Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[9] 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.

[10] Y. Wang and W. Yin, "Sparse signal reconstruction via iterative support detection," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 462–491, 2010.

[11] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.