Rakeness-Based Design of Low-Complexity Compressed Sensing

Mauro Mangia, Fabio Pareschi, Valerio Cambareri, Riccardo Rovatti, Gianluca Setti

***Abstract*—Compressed Sensing (CS) can be introduced in the processing chain of a sensor node as a mean to globally reduce its operating cost, while maximizing the quality of the acquired signal. We exploit CS as a simple early-digital compression stage that performs a multiplication of the signal by a matrix.**

The operating costs (e.g., the consumed power) of such an encoding stage depend on the number of rows of the matrix, but also on the value and position of the rows' coefficients. Our novel design flow yields optimized sparse matrices with very few rows. It is a non-trivial extension of the rakeness-based approach to CS and yields an extremely lightweight stage implemented by a very small number of possibly signed sums with an excellent compression performance.

By means of a general signal model we explore different corners of the design space and show that, for example, our method is capable of compressing the signal by a factor larger than 2.5 while not considering 30% of the original samples (so that they may not be acquired at all, leaving the analog front-end and ADC stages inactive) and by processing each of the considered samples with not more than three signed sums.

I. Introduction

Compressed Sensing (CS) is a technique that allows the reconstruction of a signal starting from a number of linear *measurements* that is potentially much smaller than its number of Nyquist-rate samples. The possibility of such a parsimonious representation hinges on a change of basis: many natural signals are such that, if expressed in a suitable domain, exhibit a large number of zero or almost-zero components, i.e., are *sparse*.

For this reason, CS has been proposed [1], [2] as an appealing, resource-efficient method to substitute Analog-to-Digital Converters (ADC) with Analog-to-Information Converters (AIC) [3], [4], which exploit the possibility of matching the resources needed for the acquisition of an analog signal to its true information rate by computing measurements directly in the analog domain. On the other hand, recent evidence showed that CS is at least as efficient when implemented in a purely digital architecture [5], [6], [7], i.e., as a way to provide digital signal compression. In this context, the investigation of Chen *et al.* [5] showed that there may be a clear practical advantage in using CS as an early-digital compression stage over its analog implementation in terms of power consumption.

The detailed modeling in each of the above papers is beyond our scope. Yet, the guiding idea is that data compression is needed to reduce the energy required by the transmitter, but the complexity of compression stages must be extremely limited not to balance the resources saved in transmission. CS with its simple random projections is an ideal candidate. This is especially true if, as in this paper, the implementation of the linear mapping can be kept to an almost negligible level of complexity by proper design.

M. Mangia and R. Rovatti are with DEI and ARCES, University of Bologna, Italy. Email: {*mauro.mangia2, riccardo.rovatti*}*@unibo.it*. V. Cambareri is with ICTEAM-ELEN, University of Louvain, Belgium. Email: *valerio.cambareri@uclouvain.be*. F. Pareschi and G. Setti are with ENDIF, University of Ferrara, and with ARCES, University of Bologna Italy. Email: {*fabio.pareschi, gianluca.setti*}*@unife.it*

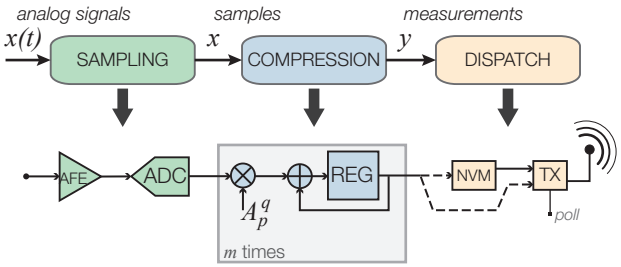

Fig. 1. The three stages of a sensor node that acquires and transmits information depending on a locally acquired signal to a final, possibly remote "consumer" (top); its specialization entailing CS, and either immediate or delayed transfer of its measurements (bottom).

We focus on the part of the acquisition system that computes and dispatches the measurements, being agnostic on the specific implementation of signal reconstruction. This complies with two largely accepted assumptions: i) many deployments of sensor networks have a hierarchy of available resources and sensing nodes sit at the bottom, while data are collected and processed by increasingly more powerful nodes; ii) the amount of resources spent in processing the measurements depends on the application and on the target level of quality, the best achievable reconstruction being an index of the quality of the extracted measurements.

Motivated by this, we focus on methods and algorithms to realize a CS-based, low-complexity digital signal compression stage in a sensor node, whose general signal chain is summarized in Figure 1. The *acquisition stage* provides an interface with the physical, analog and continuous-time world. Right after it, we introduce CS to provide a *compression stage* that encodes samples into digital words, i.e., *compressive measurements*. Once computed, these measurements are transferred to their final consumer by the *measurements' dispatch stage*.

We show how the degrees of freedom offered by CS can be exploited to: *i*) decrease the amount of sampling, *ii*) decrease the number of elementary operations in the compression stage, and *iii*) reduce the number of measurements transmitted/stored in the last stage.

This reduction of the computational burden can im-

pact power consumption that is often the key operating cost. Although an exact quantification of these savings is implementation-dependent, preliminary evidence exists that, for example, even naive and non-optimized zeroing of entries in the projection matrix results in a resource-efficient and effective compression [6][20].

The paper is organized as follows. Section II summarizes the main concepts of CS including an extension from the author's prior work, i.e., *rakeness* [8]. Roughly speaking, choosing CS acquisition sequences maximizing rakeness allows to increase the energy collected by each sample in a way similar to 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 [9][10]. Section III identifies the three parameters controlling the amount of processing performed by each of the stages in Figure 1.

In Sections IV and V, a novel design flow for CS systems is defined in which all the three merit factors affecting computational complexity are taken into consideration along with implementation constraints. This is different from classical rakeness-based design, that is concerned only with the compression ratio and does not consider implementation constraints. A method to generate ternary or binary vectors with a prescribed correlation is also given, which is a non-trivial extension of one of the techniques available for antipodal vectors.

A final Section VI shows the impact of the proposed design flows on some example cases and on two specific tasks in ECG acquisition that allow comparisons with state-of-the art proposals.

II. Overview of Compressed Sensing

We here adopt a discrete-time formulation of CS, in which the input waveform is represented by a set of n samples collected in a *signal* $x = (x_0, \ldots, x_{n-1})^\top \in \mathbb{R}^n$.

The key assumption of CS is κ-*sparsity*, that is the existence of a n-dimensional *sparsity basis* $S \in \mathbb{R}^{n\times n}$ in which any instance of the signal $x = S\xi$ is represented by a vector $\xi \in \mathbb{R}^n$ with not more than $\kappa \ll n$ non-zero components.

The number of true degrees of freedom in x is therefore considerably smaller than n. Leveraging this property, fundamental results [11] have shown that the signal can be captured by a set of $m < n$ properly designed 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., $y = Ax = AS\xi$.

The elements of y are also indicated as *projections*, and their minimum amount m essentially depends on (n, κ). In fact, formal results [11], [12] guarantee that ξ (and thus x) can be recovered from y despite the fact that A (and thus AS) yields a dimensionality reduction, provided that $m = \mathcal{O}(\kappa \log n)$. Roughly speaking, the rationale of these guarantees is that generic, κ-sparse vectors are mapped *almost isometrically* [1] into the measurements; if this is true, the recovery of the original signal x from $y = Ax$ is possible by enforcing the *a priori* knowledge that its representation ξ is sparse.

Algorithmically, methods to perform this *sparse signal recovery* have been explored and improved in recent years [11], [13], [14]. Many recovery algorithms solve convex optimization problems such as

$$\min_{\hat{\xi}\in\mathbb{R}^n} \ \|\hat{\xi}\|_1 \tag{1}$$

$$\text{s.t.} \ \|AS\hat{\xi} - y\|_2^2 \leqslant \epsilon^2 \tag{2}$$

where the 1-norm $\|\hat{\xi}\|_1$ is used to promote sparsity and $\|AS\hat{\xi} - y\|_2^2$ is the usual Euclidean norm that measures the accuracy with which the measurements y are matched by the solution, as $\epsilon \geqslant 0$ is chosen proportionally to the amount of noise affecting y.

With these premises, it is fair to say that most of the practical interest in CS comes from two key facts:

- although theoretical upper bounds exist on the error committed by signal recovery algorithms depending on the features of A, S and the amount of noise, their effective performance largely exceeds that predicted by formal guarantees, yielding signal recovery from a small number of measurements y [15];
- the mathematical conditions that allow signal recovery algorithms to succeed can be matched (with *very high* probability) by simply drawing A as a suitably chosen random matrix. Although theoretical guarantees depend on the choice of specific distributions [16], in practice a wide class of random matrices allows for effective signal recovery [17].

Starting from these facts, one may attempt to optimize acquisition performance by adapting the statistical distribution of A to that of the signal x. Once the distribution of A is chosen, a single instance can be drawn and used for sensing with a substantial chance of achieving the desired result.

This is what rakeness-based design of A does [8], [15]. To formalize it, indicate with $a = (a_0, \ldots, a_{n-1})^\top$ the random column vector corresponding to a generic row of A so that $a^\top x$ is the generic measurement.

Assume also that x is not only sparse, but also *localized* in the sense that its energy is anisotropically distributed in the signal space, or, more formally, that its correlation matrix $\mathcal{X} = \mathbf{E}[xx^\top]$ is not a multiple of the identity. Such a localization can be quantified by computing how much the eigenvalues of $\mathcal{X}$ deviate from the isotropic case, i.e.,

$$\mathfrak{L}_x = \sum_{j=0}^{n-1} \left(\frac{\lambda_j(\mathcal{X})}{\mathrm{tr}(\mathcal{X})} - \frac{1}{n} \right)^2 = \frac{\mathrm{tr}\left(\mathcal{X}^2\right)}{\mathrm{tr}^2(\mathcal{X})} - \frac{1}{n} \tag{3}$$

where $\mathrm{tr}(\cdot)$ stands for matrix trace, $\lambda_j(\cdot)$ is the j-th eigenvalue of its matrix argument, and one can show that $0 \leqslant \mathfrak{L}_x \leqslant 1 - \frac{1}{n}$.

Localization and sparsity are different priors since the subspace along which energy concentrates does not need to be a κ-dimensional canonical subspace in the sparsity reference system. If x is localized (as it almost always happens when dealing with real world signals [15]) one may identify preferred directions along which the signal is most likely lying. These are the directions along which projections may want to focus, though they should remain able to explore other

directions that on the average are less energetic but may be important to reconstruct x.

This highlights a trade-off analogous to the exploitation/exploration dilemma typical of complex optimization problems: too general methods do not leverage the additional knowledge on typical signals, while too specific ones may fail to perform acceptably in slightly atypical cases. Such a trade-off can be tackled defining the concept of *rakeness* between two vector processes $a, x \in \mathbb{R}^n$ as $\rho(a,x) = \mathbf{E}_{a,x}\left[\left(a^\top x\right)^2\right]$, i.e., the average energy of the projections of instances of x over independently drawn instances of a. If $\mathcal{A} = \mathbf{E}_a[aa^\top]$ (note that the calligraphic typefaces indicate correlations so that $\mathcal{A}$ is a correlation matrix, while A is the sensing matrix), then rakeness can be expressed as $\rho(a,x) = \operatorname{tr}(\mathcal{A}\mathcal{X})$.

With this, we may formalize the intuitive design criteria of concentrating projections on most energetic directions while not neglecting the others, as the *maximization of the rakeness subject to a cap on the localization of the sensing process*, i.e., if $N = \{0, \dots, n-1\}$,

$$\max_{\mathcal{A}} \ \operatorname{tr}(\mathcal{A}\mathcal{X}) \tag{4}$$

$$\text{s.t. } \mathcal{A}_{j,k} = \mathcal{A}_{k,j} \qquad j,k \in N \tag{5}$$

$$\text{s.t. } \mathcal{A} \succeq 0 \tag{6}$$

$$\text{s.t. } \mathfrak{L}_a \leqslant \ell \mathfrak{L}_x \tag{7}$$

where (5) and (6) ensure that $\mathcal{A}$ is a correlation matrix (symmetric and positive semidefinite), and the coefficient $\ell \geqslant 0$ in (7) sets how much the projection waveforms should be localized with respect to the process to sense and thus controls the trade-off between exploitation (high rakeness) and exploration (low localization). For this reason, we will call the optimization problem defined by (4)-(7) Rakeness-Localization Trade-off (RLT).

Though many other methods to optimize sensing matrices A have been proposed (see [18] for an overview of some methods for real matrices and [19] for an example of optimized binary matrices), rakeness-based design is unique in that it tunes sensing depending on a broadly verified property of the input signal, instead of trying to optimize some generic measure of good behavior such as coherence, mutual coherence, or restricted isometry. Thanks to this adaptation to the signal, rakeness-based design largely outperforms the other methods. For example, experiments in Section VI show this for binary matrices comparing our method with that in [19].

Clearly, adapting measurements to increase average energy has effects on the mutual coherence between the projection matrix and the sparsity base, weighted by the statistics of the signal. Moreover, one may reasonably expect that this increases the mutual entropy between x and y. Though out of the scope of this paper, these directions are clearly worth exploring to give a even sounder ground to rakeness-based design.

III. Operating Costs of Compressed Sensing

The general three-stages structure at the top of Figure 1 can be specialized within the CS framework. Since it was suggested [5], [6] that early digitization may be more efficient than analog-domain computation of projections, the sampling stage entails both an Analog Front End (AFE) interfacing our system with the sensor and an Analog-to-Digital Converter (ADC) performing discretization in time and magnitude. Projections are therefore performed by digital Multiply and ACcumulate operations (MAC). The last stage has either the option of immediately transmitting (TX) the measurements or storing them in a non-volatile memory (NVM) waiting for external signal and resources to trigger transfer at a later time. As anticipated, the operating cost of the three stages can be related to features of the CS mechanism.

Clearly, the cost of measurement dispatch depends on the number of measurements that must be transmitted/stored in order to guarantee a correct input signal reconstruction.

As far as measurement computation is concerned, applying a dense A entails a potentially large number ($n \times m$) of MAC operations. A straightforward way of reducing this computational effort is to assume that A is populated either by antipodal ($A_{p,q} \in \{-1, +1\}$), binary ($A_{p,q} \in \{0, +1\}$) or ternary ($A_{p,q} \in \{-1, 0, +1\}$) symbols so that multipliers reduce to simple signed adders, leaving the general reconstruction performance substantially unaltered [3], [6]. Since when $A_{p,q} = 0$ the MAC unit can be left idle, the computational effort of the projection stage is controlled by the number of non-zero entries in A.

More specifically, let P_q be the number of non-null entries in the q-th column of A. Looking at the projection mechanism from the point of view of each sample, P_q quantifies the number of MAC operations triggered by the availability of x_q.

Implementations may want to constrain $P_q \leqslant \Theta < m$ for a certain Θ and for all $q = 0, \dots, n-1$. This requirement can be thought of as a *throttling* since it ensures that a limited computational effort is associated to each incoming sample: for example, if the projection stage is hardwired in a parallel structure, not more than Θ parallel paths must be deployed; alternatively, if Ax is computed by a programmable architecture, at most Θ MAC iterations are needed in each sample time. Hence, from the computational point of view, the product by the original $m \times n$ matrix A is equivalent to that of a $\Theta \times n$ throttled matrix.

Pushing this further, note that if $P_q = 0$ then x_q does not enter the computation of any of the measurements. Hence, one may think of not acquiring that sample and make also the AFE and ADC idle for the corresponding sample time, thus reducing operating cost. We will indicate such an option as *puncturing*, a sample skipping technique that has already been proposed, for example, for compressed estimation [20].

Furthermore, throttling and puncturing affect the subsystem providing the matrix A to the signal chain, let it be a memory or a run-time generator. For example, if A is stored in a non-volatile memory as in [6], the number of non-zeros in A directly impacts the size of the memory dedicated to it.

The saving in the computational effort of the different stages together with the memory footprint for the sensing matrix can be quantified by defining the *compression ratio* (CR), the

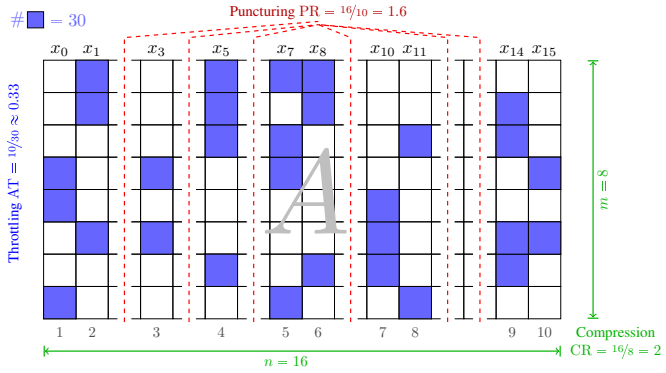

Fig. 2. Visual representation of the zeroing policies and corresponding saving figures. The whole rectangle corresponds to the matrix A in which boxes correspond to single entries. White boxes contain zero coefficients, colored boxes contain non-zero coefficients. Some columns contain only zero entries and are highlighted by a red vertical dashed line.

average throttling (AT) and the *puncturing ratio* (PR) as

$$\begin{aligned} \mathrm{CR} &= \frac{n}{m} \geqslant 1 \\ \mathrm{AT} &= \frac{|\{q|P_q > 0\}|}{\sum_{q=0}^{n-1} P_q} \geqslant \frac{1}{m} \\ \mathrm{PR} &= \frac{n}{|\{q|P_q > 0\}|} \geqslant 1 \end{aligned}$$

where $|\{q|P_q > 0\}| \leqslant n$ is the number of columns with at least one non-zero element. Hence, columns zeroed by puncturing are not counted by AT, which therefore accounts only for the average computational effort spent on samples that are effectively processed.

Note that CR is not a bitwise ratio, so it can be used only as a proxy of the true gain of the compression stage. Yet, it is typically a good proxy, since it is known [3][7] that, though measurements are the results of accumulation, a satisfying reconstruction quality can be obtained by encoding the entries of y with a number of bits that i- is quite close to the number of bits encoding the entries of x; ii- does not significantly depend on m when the latter is reduced by optimization.

Figure 2 offers a visual representation of the zeroing policies we discuss and the corresponding saving parameters for a sparse, $m \times n$ matrix A with $n = 16$ and $m = 8$, such that $CR = 2$. The non-zero elements of A are represented by filled boxes, while empty boxes correspond to zero elements. Puncturing leaves only 10 non-null columns out of $n = 16$ and thus $\mathrm{PR} = 1.6$. The total amount of non-zero elements is $\sum_{q=0}^{n-1} P_q = 30$ over 10 non-punctured columns and thus $\mathrm{AT} = 1/3$ which corresponds to an average of 3 non-zero elements for each column.

From the computational point of view, if no MAC is performed for zero entries in A, the product by such a matrix is equivalent to a product by a matrix throttled to have only 3 rows and punctured to have only 10 columns, i.e., 30 MACs instead of the potential $16 \times 8 = 128$ MACs entailed by a non-sparse matrix.

The definitions of PR, AT, CR are given so that they are merit figures: the larger they are, the lower the computational effort. In fact, an abstract modeling of the operating costs of each stage characterizing the sensor node gives:

- sampling stage:
  $\mathcal{O}(\#\text{ of non-zero columns in } A) = \mathcal{O}\left(\frac{n}{\mathrm{PR}}\right)$;
- compression stage:
  $\mathcal{O}(\#\text{ of non-zero entries in } A) = \mathcal{O}\left(\frac{n}{\mathrm{AT\ PR}}\right)$;
- dispatch stage:
  $\mathcal{O}(\#\text{ of measurements to transfer}) = \mathcal{O}\left(\frac{n}{\mathrm{CR}}\right)$.

Actual implementations will weight differently the contribution of each stage to the overall computational burden.

To give a rough insight on how the proposed merit figures may affect the resources needed by actual implementation we may resort to high-level power and memory modeling of some typical realization options.

As an example, the power models for CS in sensor nodes developed in [20] and [6] may be used to show the different impact of PR, AT and CR.

- sampling stage:
  Power is required by both the AFE and by the ADC. The AFE contribution is typically dominated by the low noise amplifier whose power requirement depends on the operating temperature, the amplifier gain and its 3 dB bandwidth, the noise efficiency factor, the power supply voltage and the full scale voltage. By using reasonable values for these parameters [20] one may estimate
  $$W_{\mathrm{AFE}} \approx 2^{2\mathrm{b}_x} f_{\mathrm{AFE}} \cdot 36\,\mathrm{fJ}$$
  where f_{AFE} is the average activity of the AFE. Under the assumption [20] that this stage can be put in an extremely low-power state whenever a sample is not needed, i.e., when the corresponding column of A is null, we have $f_{\mathrm{AFE}} = f_s/\mathrm{PR}$ so that
  $$W_{\mathrm{AFE}} \approx 2^{\mathrm{b}_x} \frac{f_s}{\mathrm{PR}} \cdot 36\,\mathrm{fJ} \tag{8}$$
  Similarly, we can estimate the power consumption of the ADC with
  $$W_{\mathrm{ADC}} \approx 2^{\mathrm{b}_x} \frac{f_s}{\mathrm{PR}} \cdot 10\,\mathrm{fJ} \tag{9}$$
  where we have set the Walden figure of merit equal to 10 fJ/conversion-step [21].
- compression stage:
  Matrix multiplication can be realized in a number of ways. Here we focus on two corner cases: an ultra low-power general purpose DSP [6] and a full custom architecture [20].
  For the ultra low power DSP the power model has been obtained from experimental results when $n = 512$, $f_s = 360\,\mathrm{Hz}$, $\mathrm{b}_x = 11$, and leakage is reduced by 65% with a proper reverse body bias technique [6].
  $$W_{\mathrm{MAC}} \approx \frac{1}{\mathrm{PR\ AT}} \cdot 708\,\mathrm{nW} + 7.1\,\mu\mathrm{W} \tag{10}$$
  In the full custom approach all MAC operations are implemented as signed sums and accumulations. The corresponding power model is obtained [20] by designing and

TABLE I
ENERGY PER BIT FOR SOME EMERGING STORAGE/TRANSMISSION TECHNOLOGIES

		Energy per bit [nJ/bit]	Source
NVM	FLASH	0.01	[22]
	CBRAM	0.001	[23]
	STT-MRAM	0.0001	[24]
TX	NB	0.1	[25]
	BLE	1	[26]
	NFC	10	[27]

simulating a dedicated compression stage implemented with a low-leakage 28 nm SOI CMOS technology. Different input and accumulator wordwidths are considered and read operations from a suitable integrate memory with 2 bits per sample (needed to encode ternary coefficients) are taken into account. Power consumptions were obtained from simulating gate-level netlists produced by Synopsis Design Compiler with Mentor Graphics Modelsim and Monte Carlo input data. As a result

$$W_{\mathrm{MAC}} \approx \frac{n}{\mathrm{PR\ AT}}(24\mathrm{b}_x + 10\log_2 n + 19) f_s \cdot 1\,\mathrm{fJ} + mW_{\mathrm{leak}}(\mathrm{b}_x, n) \quad (11)$$

where $W_{\mathrm{leak}}(\mathrm{b}_x, n)$ is the leakage of each parallel MAC branch and

$$W_{\mathrm{leak}}(\mathrm{b}_x, n) \approx (30\mathrm{b}_x + 18\log_2 n + 82) \cdot 1\,\mathrm{nW}$$

where a scaling corresponding to a body bias techniques similar to the previous case has also been applied.

- dispatch stage:

  Every time window, the $m = n/\mathrm{CR}$ measurements must be either stored locally or transmitted. Hence, the required power is estimated in

  $$W_{\mathrm{NVM|TX}} \approx \frac{n}{\mathrm{CR}} \mathrm{b}_y E_{\mathrm{NVM|TX}} f_s \quad (12)$$

  where $\mathrm{b}_y \leqslant \mathrm{b}_x + \log_2 n$ is the number of bits used to store each element of y, E_{NVM} is the energy needed to store a single bit in the NVM, while E_{TX} is the energy needed to transmit a single bit. Values for E_{NVM} and E_{TX} are reported in Table I for some consolidated and emerging technologies.

The above models can be used to target different applications and show the different weights that PR, AT, and CR may have on the overall power consumption.

As a first case consider a system similar to the one modeled in [6] where in (8) and (9) we have $n = 512$, $f_s = 360\,\mathrm{Hz}$ and $\mathrm{b}_x = 11$; the compression stage refers to (10) and for the dispatch stage with $\mathrm{b}_y = 11$ we consider (12) where data are stored in a FLASH memory [22]. The overall power consumption W is

$$W[\mu\mathrm{W}] \approx 7.1 + \frac{54}{\mathrm{PR}} + \frac{0.7}{\mathrm{PR\ AT}} + \frac{20}{\mathrm{CR}} \quad (13)$$

If, instead of storing the measurements, one transmits them using a low energy Bluetooth module (BLE) [26] W becomes

$$W[\mu\mathrm{W}] \approx 7.1 + \frac{54}{\mathrm{PR}} + \frac{0.7}{\mathrm{PR\ AT}} + \frac{2000}{\mathrm{CR}} \quad (14)$$

As a second case considers an architecture closer to what is modeled in [20] with a sampling stage identical to the previous one, a NVM in the dispatch stage and a custom implementation of the compression stage obeying (11). In this case set $n = 64$, $f_s = 1\,\mathrm{MHz}$, $\mathrm{b}_x = \mathrm{b}_y = 5$ and assume that measurements are stored in a STT-MRAM [24] thus yielding

$$W[\mu\mathrm{W}] \approx \frac{9.4}{\mathrm{PR}} + \frac{11.2}{\mathrm{PR\ AT}} + \frac{45.5}{\mathrm{CR}} \quad (15)$$

where the custom layout of the compression stage makes the leakage dependent on CR.

The fact that $1 \leqslant \mathrm{CR} \leqslant n$ with typical values in the units, $1 \leqslant \mathrm{PR} \leqslant n$ with typical values close to 1, while $\frac{1}{m} \leqslant \mathrm{AT} \leqslant 1$, makes the three above cases very different as far as sensitivity to our parameters is concerned.

Further to power consumption, an issue that may be important in implementations exploiting general purpose programmable hardware (possibly involved also in other tasks) certain implementation is the memory footprint, mainly due to the storage of A.

Usually, the addressing granualarity is at byte level and a straighforward deployment would use $n\,m\,\mathrm{byte}$ of memory for A. When PR AT increases, A can be stored exploiting one of the many techniques used for sparse matrices. A common option is to store A column-wise and memorize the position and the values of the the non-null entries in A. Assuming that position and value entries use 1 byte each, we come up with $2n/(\mathrm{PR\ AT})\,\mathrm{byte}$ for the whole A. Hence a reduction in the memory footprint arises whenever

$$\mathrm{PR\ AT} > \frac{2}{m}$$

As an example with $n = 256$ and $m = 128$ it is enough that $\mathrm{PR\ AT} \geqslant 0.016$ to yield an improvement while, in the same setting, we will show that values up to $\mathrm{PR\ AT} > 0.5$ can be obtained (allowing to store A in less than 512 byte instead of 32 Kbyte).

Though only by examples, the above discussion shows that, depending on implementation choices or constraints the role of CR, AT, and PR may change.

This is why our model keeps CR, AT, and PR as distinguished metrics giving the tools to increase them as much as possible while keeping the quality of the reconstruction above a minimum application-dependent accuracy level.

The main point is that as CR, AT and PR increase the number of measurements and the number of samples entering each measurement decreases, and we have less chances to look at the signal and capture information. Hence these chances must be optimally exploited to avoid performance degradation.

This is where the extension of rakeness-based design we present here comes into play. It finds, within the matrices featuring certain PR, AT, and CR, those that promise the best acquisition performance. Conversely, once that a target reconstruction quality is chosen, the novel rakeness-based

design helps increasing PR, AT, and CR thus reducing the complexity of the acquisition stages.

IV. Rakeness-based ternary CS

As noted before, adopting $A_{p,q} \in \{-1, 0, +1\}$ helps significantly to reduce the complexity and the number of MACs entailed by the product Ax. Regrettably, not all the correlation matrices in the feasibility space of the RLT problem in (4)-(7) correspond to a ternary vector process that may be used to fill the rows of A and the full characterization of the true set of matrices in which one should search is unavailable (see [28] for a partial discussion of this issue in the case of antipodal processes).

To cope with this, we add constraints to the RLT, whose role is cutting out of the feasibility space matrices that are trivially unachievable. This produces a version of RLT problem that is still a relaxation of the one we should solve, but hopefully a tight one.

Note first that $a_j \in \{-1, 0, +1\}$ and $\mathcal{A} = \mathbf{E}[aa^\top]$ imply $\mathcal{A}_{j,j} = \mathbf{E}[a_j^2] = \Pr\{a_j \neq 0\} \leqslant 1$. Moreover $|\mathcal{A}_{j,k}| = |\mathbf{E}[a_j a_k]| \leqslant \min\{\Pr\{a_j \neq 0\}, \Pr\{a_k \neq 0\}\}$. In fact, two ternary random variables are maximally positively (negatively) correlated when the chance that they have the same (opposite) sign is maximized, a chance that cannot exceed the probability that each of them is non-zero, i.e., the smallest of the probabilities of being non-zero. Hence, it must be $|\mathcal{A}_{j,k}| \leqslant \min\{\mathcal{A}_{j,j}, \mathcal{A}_{k,k}\}$, that, thanks to the symmetry of $\mathcal{A}$ can be reduced to $|\mathcal{A}_{j,k}| \leqslant \mathcal{A}_{j,j}$.

We use these properties as constraints to restrict optimization to ternary-friendly correlation and to control the average number of non-zeros at each position of a and thus ultimately AT and PR. Formally speaking, we set $\Pr\{a_j \neq 0\} = \mathcal{A}_{j,j} = \eta_j$ for some given $\eta_j \in [0, 1]$ and $j = 0, \ldots, n-1$. Therefore, the parameters η_j will control the density of the resulting acquisition matrix: the lower the η_j the lower the probability that the j-th entry of each row of A is non-zero.

Altogether, the general optimization problem addressing the Ternary Rakeness-Localization Tradeoff (T-RLT) is

$$\begin{aligned}
&\max_{\mathcal{A}} \ \operatorname{tr}(\mathcal{A}\mathcal{X}) \\
&\text{s.t. } (5), (6), (7) \\
&\text{s.t. } \mathcal{A}_{j,j} = \eta_j \qquad\qquad j \in N \qquad (16) \\
&\text{s.t. } |\mathcal{A}_{j,k}| \leqslant \eta_j \qquad j \neq k \quad j, k \in N \qquad (17)
\end{aligned}$$

In such a problem, given a certain index $\bar{\jmath}$, we may force $a_{\bar{\jmath}}$ to be purely antipodal (or, on the contrary, null) by setting $\eta_{\bar{\jmath}} = 1$ (or $\eta_{\bar{\jmath}} = 0$).

T-RLT is different from the RLT that is classically used for rakeness-based design [8], [15] since the implementation requirement $a_j \in \{-1, 0, +1\}$ is implicitly encoded in constraints (16) and (17). This prevents the spectral decomposition of $\mathcal{A}$ to mimic that of $\mathcal{X}$, thus impairing the eigenvector-eigenvalue solution technique that is commonly pursued.

Yet, the objective function is linear, the average (16) and ternary implementation constraints (17) are linear, and the semidefiniteness constraint (6) is convex. Moreover, since $\operatorname{tr}(\mathcal{A}) = \sum_{j=0}^{n-1} \eta_j$, ℓ, and $\mathfrak{L}_x$ are fixed, (3) implies that the localization constraint (7) is quadratic and convex. Hence, T-RLT is a convex programming problem that can be numerically attacked by using established and general-purpose modeling strategies and convex solvers [29], [30].

Those tools are quite effective in finding provable optima in reasonable time for n up to few tens (and thus for the number of degrees of freedom in the hundreds). For larger values of the signal dimensionality a specialized approach may be used, as described in subsection VIII-A of the Appendix.

However solved, T-RLT allows us to obtain a correlation matrix $\mathcal{A}$ for ternary-valued random vectors so that, once they are generated accordingly and collected in the rows of the projection matrix A, their application to a localized signal x preserves most of its information in the measurements y.

What we address in the two following Subsections is the generation of such vectors considering the architectural options anticipated in the previous Sections.

A. Unstructured zeroing

In this case no constraint is put on the positions of the zeros in A. What matters is only that A is sparse (with a sparsity controlled by PR and AT) to avoid accumulations corresponding to zero entries.

It is then sensible to proceed by solving T-RLT with proper values $0 < \eta_j < 1$ to obtain the matrix $\mathcal{A}$ that should regulate the correlation of the rows of A. For example, if AT $= 1/4$ and $m = 50$ we want, on average, 4 non-null elements out of 50 in each column. This is mapped to $\eta_j = (50 \cdot 1/4)^{-1} = 0.08$ for $j = 0, \ldots, n-1$.

The generation of those rows, i.e., of ternary vectors with a prescribed correlation matrix, can be pursued by generalizing the methods proposed for antipodal vectors (see, e.g., [31], [32], [33], [28]).

In particular, we focus on thresholding of Gaussian random vectors [31], a method based on the early result in [34]. Though not completely general (not every feasible correlation matrix can be obtained) such a method is very simple and allows an almost equally simple generalization to the ternary case.

For antipodal vectors generation, i.e., $a \in \{-1, +1\}^n$, a random Gaussian vector g is generated with zero mean, unit variance and a correlation matrix $\mathcal{G}$ computed as

$$\mathcal{G} = \sin\left(\frac{\pi}{2} \frac{n\mathcal{A}}{\operatorname{tr}(\mathcal{A})}\right) \qquad (18)$$

where $\sin$ is applied componentwise. Then, the antipodal vector is simply obtained by computing the sign of the elements of g.

To generate a ternary vector a, the procedure starts again from a Gaussian random vector g whose entries have zero mean, unit variances and a new correlation matrix $\mathcal{G}$ properly designed to produce $a = (a_0, \ldots, a_{n-1})^\top$ componentwise as $a_j = \tau^{\mathrm{t}}_{\theta_j}(g_j)$ where

$$\tau^{\mathrm{t}}_{\theta_j}(g_j) = \begin{cases} -1 & \text{if } g_j \leqslant -\theta_j \\ 0 & \text{if } -\theta_j < g_j \leqslant \theta_j \\ +1 & \text{if } \theta_j < g_j \end{cases} \qquad (19)$$

is a three level quantization function whose symmetric threshold θ_j can be chosen differently for each entry of a. The calculations generalizing (18) to go from the desired $\mathcal{A}$ to $\mathcal{G}$ and $\theta = (\theta_0, \dots \theta_{n-1})^\top$ for the ternary case is reported in subsection VIII-B of the Appendix.

B. Structured zeroing

In this case, the implementation seeks further optimization by requiring that the positions of the non-zeros in A satisfy additional constraints that help reducing resources dedicated to sampling and accumulation.

The most obvious structured zeroing is puncturing, that zeroes whole columns, though other policies may be devised depending on the features of the underlying implementation.

In any case, once the positions of the non-zeros in A are given, one also knows the support $\Gamma^p = \{q | A_{p,q} \neq 0\}$ of the p-th row for $p = 0, \dots, m-1$.

Hence, we may proceed row-by-row and solve T-RLT with $\eta_j = 0$ for any $j \notin \Gamma^p$. Constraints on the other diagonal entries can be put depending whether further zeros are allowed or not. If no other zeros are admitted further to what is set either by puncturing or throttling, then we may set $\eta_j = 1$ for $j \in \Gamma^p$.

Once average constraints are decided, T-RLT gives us an optimal correlation matrix $\mathcal{A}$ whose submatrix $\bar{\mathcal{A}}$ collecting rows and columns with indexes in Γ^p regulates the entries in the random subvector $\bar{a}$ containing the free components of A_p.

V. Rakeness-based Binary CS

An even more aggressive approach to resource saving is to consider only binary projections, i.e., assume $a \in \{0,1\}^n$. As an example, if A is sparse and binary, one needs to store only the positions of the non-null entries instead of their positions and values thus potentially halving the memory footprint.

In the binary case the rakeness maximization problem can still be formulated if the bounds on the elements of $\mathcal{A}$ are suitably adjusted to cut out of the feasibility space some matrices that could not be achieved by a binary process. In fact, though we have $\mathcal{A}_{j,k} \leqslant \min\{\eta_j, \eta_k\}$ the lower bound on the same quantity is different.

More specifically, for binary variables $\mathcal{A}_{j,j} = \mathbf{E}[a_j] = \Pr\{a_j = 1\}$ and $\mathcal{A}_{j,k} = \mathbf{E}[a_j a_k] = \Pr\{a_j = 1 \wedge a_k = 1\}$. Yet, the probability that the two binary random variables a_j and a_k are simultaneously non-zero cannot be less than $\Pr\{a_j = 1\} + \Pr\{a_k = 1\} - 1$. Hence, $\mathcal{A}_{j,k} \geqslant (\eta_j + \eta_k - 1)^+$ where $(\cdot)^+ = \max\{0, \cdot\}$.

The optimization problems addressing the Binary Rakeness-Localization Tradeoff (B-RLT) is therefore

$$\begin{aligned} &\max_{\mathcal{A}} \ \operatorname{tr}(\mathcal{A}\mathcal{X}) \\ &\text{s.t.} \ \ (5) - (16) \\ &\text{s.t.} \ \ (\eta_j + \eta_k - 1)^+ \leqslant \mathcal{A}_{j,k} \leqslant \eta_j \quad j \neq k \quad j,k \in N \end{aligned} \tag{20}$$

where (20) replaces the ternary constraint (17).

The optimization problem B-RLT can be solved, either by general convex programming tool or with the same projected-gradient approach used for T-RLT that is specialized in subsection VIII-C of the Appendix. This yields the optimal correlation matrix $\mathcal{A}$ that should regulate the binary sensing waveforms.

In the two following Subsections we detail how such binary sensing vectors with the desired correlation matrix $\mathcal{A}$ can be generated also considering the architectural options anticipated in the previous Sections.

A. Unstructured zeroing

Binary random vectors with a prescribed correlation matrix can be generated following the same path used for the ternary case by defining

$$\tau^{\mathrm{b}}_{\theta_j}(g_j) = \begin{cases} 0 & \text{if } g_j < \theta_j \\ 1 & \text{if } g_j \geqslant \theta_j \end{cases} \tag{21}$$

and by setting $a_j = \tau^{\mathrm{B}}_{\theta_j}(g_j)$ starting from a proper choice of the thresholds $\theta = (\theta_0, \dots, \theta_{n-1})^\top$ and of the correlation matrix $\mathcal{G}$ regulating the Gaussian, unit-variance vector $g = (g_0, \dots, g_{n-1})^\top$.

How to go from the desired $\mathcal{A}$ to $\mathcal{G}$ and θ is reported in subsection VIII-D of the Appendix.

B. Structured zeroing

Structured zeroing can be applied in the binary case only if components that may be non-zero still have a probability to be null. Otherwise no randomness survives the enforcement of the structure and there is no statistics to optimize.

Beyond that, given the above procedure to generate binary vectors with a prescribed correlation, the steps taken in the case of structured zeroing of antipodal matrices can be repeated here to obtain optimized rows for A.

VI. Examples

The design flow described in Sections IV and V, concretizes in a set of MATLAB© functions available online[1] along with few demo examples. The two functions `TRak` and `BRak` solve, respectively, T-RLT and B-RLT problems, and take as input the signal correlation $\mathcal{X}$, the average constraints η_j for $j = 0, \dots, n-1$, and the normalized localization constraint ℓ. They give the optimal correlation matrix $\mathcal{A}$ for ternary (`TRak`) or binary (`BRak`) sensing waveform. In both functions, a zero in the j-th position may be forced by setting $\eta_j = 0$ thus allowing structured zeroing. Conversely, a non-zero may be forced by setting $\eta_j = 1$.

The two functions `TGau` and `BGau` take as input the desired correlation matrix $\mathcal{A}$. They use the methods in, respectively, Subsection IV-A and Subsection V-A to yield the correlation $\mathcal{G}$ and threshold vector θ that allow to generate ternary or binary vectors with a correlation as close as possible to $\mathcal{A}$ by quantizing Gaussian vectors.

Such functions allow us to explore different options as far as the computation of projections is concerned, thus sampling the CR, AT and PR design space.

[1] http://cs.signalprocessing.it/download.html

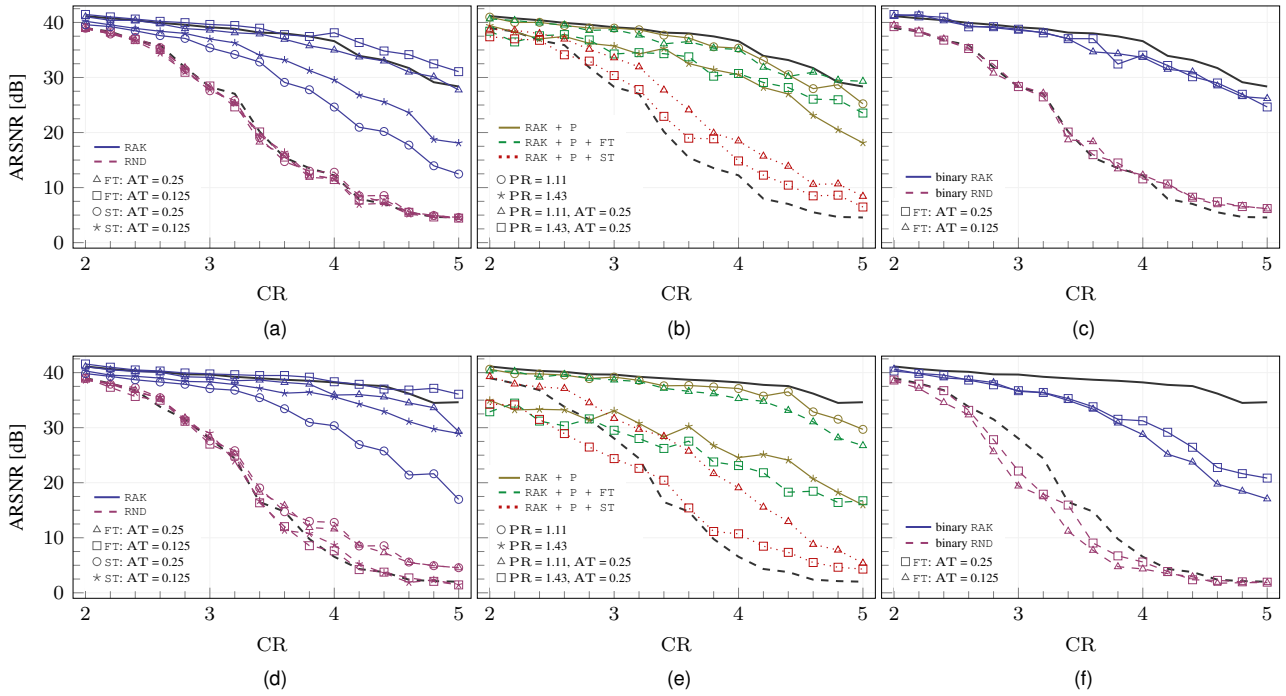

Fig. 3. ARSNR as a function of CR for different choices of projections and target signals. Plots (a),(b), and (c) are for LP target signals while plots (d),(e), and (f) are for BP target signals. Note that in all plots the solid black track without any markers account for the antipodal projection rakeness-based CS, while the dashed black track without any markers correspond to antipodal i.i.d. projection as for the standard CS case.

For throttling, we examine two options. The first is *free throttling* (labelled FT), in which each column may have a different number of non-zeros entries and the global parameter AT is a true average of what happens over the whole matrix. The second is *strict throttling* (labelled ST), in which we ensure that the number of non-zeros in each column is the same, and AT characterizes non only the average behavior but also the behavior of each column. Clearly, free throttling amounts to unstructured zeroing of the projection matrix while strict throttling translates into a structured zeroing of the projection matrix.

The reconstruction performance of each configuration is assessed by Montecarlo simulation. Once the optimal statistic for the projection waveform is established, 100 instances of the matrix A are drawn and each of them is used to encode a random instance of the target signal class. For all examples (1) is solved by SPGL$_1$ [35] with ϵ adapted to an intrinsic signal to noise ratio (ISNR) equal to 40 dB.

Once (1) is solved, then reconstruction quality is assessed as the ARSNR, i.e., the empirical average over the Montecarlo trials of the reconstruction signal-to-noise ratio $\|\xi\|^2/\|\xi-\hat{\xi}\|^2$.

A. Synthetic low-pass and band-pass sparse signals

A synthetic localized vector that is κ-sparse with respect to the base S is generated starting from an n-dimensional random Gaussian vector x' with zero mean and correlation matrix $\mathcal{X}'$ with $n = 128$. Then we compute $S^{-1}x'$ and keep in ξ the $\kappa = 12$ largest modulus entries of the result, setting all the others to zero. Finally, we set $x = S\xi + \nu$, where the noise vector ν is made of independent, zero mean Gaussian entries whose variance is set to have a known intrinsic signal-to-noise ratio ISNR=40 dB ensuring reproducibility and uniform trial conditions.

If the eigenvalues of $\mathcal{X}'$ are not identical, x' is localized and this property is approximately propagated through sparsification since only the $n-\kappa$ smallest components of $S^{-1}x'$ are discarded. Empirical evidence suggests that a good match between the localization set by $\mathcal{X}'$ and the one of x is maintained for sparsity ratios n/κ up to 15.

Results are reported in Figure 3 where subfigures (a), (b), and (c) deal with the LP signals while (d), (e), and (f) deal with BP signals. All figures plot ARSNR against CR for different values of AT and PR.

In all plots in Figure 3, the solid black track without any markers (RAK) accounts for the performance of full, rakeness-based, antipodal projections (i.e., with a correlation matrix that is the solution of T-RLT with $\eta_j = 1$ for $j = 0, \ldots, n-1$). This is the reference case in which no additional degree of freedom like zeroing is introduced but no saving is sought. The dashed black track without any markers (RND) corresponds to what would be the standard choice for classical CS sensing, i.e., projection by means of a matrix of independent and uniformly distributed antipodal symbols.

Figures 3-(a)(c) show what happens for ternary rakeness-based projections (solid tracks labelled RAK) when AT changes and throttling is either free (FT) or it is strict (ST).

As a comparison and to assess the effectiveness of statistical adaptation in each configuration, the purely random case is also modified by introducing zeros with the same strategies used for the rakeness-based case. The corresponding plots are in dashed lines, are labelled RND and feature the same markers

TABLE II
THE SAVING FIGURES OF THE SAMPLING, PROJECTION, AND DISPATCHING STAGES WHEN DIFFERENT STRATEGIES ARE EMPLOYED TO ACHIEVE A TARGET ARSNR OF 34 dB FOR LP AND BP SIGNALS. LABELLED CONFIGURATIONS ARE RECALLED IN THE TEXT.

	Antipodal		Ternary: Puncturing					Ternary: Throttling				Ternary: Punct. + Throt.					Binary						
	RND	RAK	RAK+P					RAK+FT		RAK+ST		RAK+P+FT		RAK+P+ST			RAK+FT		RAK+P+FT				
LP signals																							
PR	1	1	1.11	1.25	1.43	1.67	A	1	1	1	1	1.1	1.43	1.1	1.43	C	1	1	1.1	1.1	1.25	1.25	D
PR·AT	0.021	0.033	0.035	0.036	0.038	0.035		0.25	0.5	0.25	0.5	0.37	0.476	0.37	0.476		0.25	0.5	0.275	0.55	0.312	0.625	
CR	2.7	4.29	4	3.76	3.37	2.66		4.12	3.65	3.2	2.98	4	3.37	2.98	2.72		3.76	3.88	3.72	3.96	3.54	3.64	
BP signals																							
PR	1	1	1.11	1.25	1.43	1.67	B	1	1	1	1	1.1	1.43	1.1	1.43		1	1	1.1	1.1	1.25	1.25	
PR·AT	0.021	0.038	0.04	0.036	0.023	n/a		0.25	0.5	0.25	0.5	0.37	n/a	0.37	n/a		0.25	0.5	0.275	0.55	0.312	0.625	
CR	2.66	4.92	4.57	3.76	2.06	n/a		4.74	4.26	3.55	3.05	4.27	n/a	2.84	n/a		3.55	3.65	3.21	3.56	2.44	2.48	

as the RAK tracks to distinguish different options.

The gap between the random cases and the rakeness-based cases shows that statistical adaptation has non-negligible advantages. The gains are maintained even when AT increases, though strict throttling causes a larger performance loss especially at extreme throttling values (note that $AT = 0.25$ implies that only 4 entries are non-zero in each column of A).

Figures 3-(b)(e) focus on ternary rakeness-based design and consider both throttling (either free FT or strict ST) and puncturing (P). Though performance remains consistently better than classical purely random sensing it is clear that increasing the number of zeros (i.e., adding puncturing to throttling) causes an increasing loss that is larger when throttling is strict.

Figures 3-(c)(f) focus on binary sensing. As in Figures 3-(a)(b) both rakeness-based and purely random cases are reported. Rakeness-based design consistently yields gains, though how much the maximum performance is affected by savings depend on the signal.

Note that the availability of only two symbols yields, in general, slightly poorer performance with respect to the ternary case.

Table II shows how the machinery presented above can be used to address the problem of saving on operating costs along different directions. All figures refer to an ARSNR of 34 dB (6 dB less than the ISNR, roughly corresponding to the loss of one bit per sample in a standard acquisition scheme).

There are two groups of rows, one for LP signals and one for BP signals. In each group there is a line for each stage in the sensing chain reporting the corresponding figure of saving, namely, PR for the sensing stage, $\mathrm{AT} \cdot \mathrm{PR}$ for the projection stage, and CR for the transfer stage. Both LP and BP target signals are considered. Note that BP signals are more critical since in some of the configurations the target ARSNR cannot be reached (marked as n/a in Table II).

The first two columns report the saving figures for the two reference cases of full rakeness-based projections and full purely random projections. Other columns refer only to rakeness-based design that always dominates the purely random approach.

The second group of columns deals with the effect of puncturing alone. Columns are listed for increasing values PR that correspond to skipping, respectively, 10% ($\mathrm{PR} = 1.11$), 20% ($\mathrm{PR} = 1.25$), 30% ($\mathrm{PR} = 1.43$), and 40% ($\mathrm{PR} = 1.67$) of the samples while $\mathrm{PR} = 1$ means no puncturing. Note that there is a clear trade-off between PR and CR so that the optimum design point depends, in this case, on the relative operating cost of the sampling and transfer stages. In the configuration labeled as A, rakeness-based design allows to discard 40% of the samples while still achieving a CR almost equivalent to what can be obtained by classical random ternary acquisition.

The third group of columns deals with throttling alone, either free or strict. Both can be increased up to 0.5 (meaning only 2 non-zero entries per column) but free throttling yields a larger CR. In general, there is a trade-off between throttling and compression. In the case labeled as B $\mathrm{CR} = 4.74$ is obtained by only $1/\mathrm{AT} = 4$ signed sums per incoming sample.

The fourth group of columns deals with joint puncturing and throttling, either free or strict. A global trade-off is present since the increase in any of the saving figures implies a reduction in the others. Yet, as shown by the case labeled C, compression ratios close to 3 can be obtained by skipping 30% of the samples ($\mathrm{PR} = 1.43$) and performing not more than 3 signed sums for each of the remaining samples ($\mathrm{AT} = 0.476/1.43 = 0.333$).

The fifth group of columns deals with binary projections with either free throttling or puncturing and free-throttling (strict throttling does not apply to the binary case).

Note that binary projections do not exhibit the same trade off between the number of zeros and CR, actually they seem more tolerant to zeroing if compared to ternary projections though not always offering the same absolute level of performance.

Yet, in the case labelled as D, rakeness-based design allows to discard 20% of the samples ($\mathrm{PR} = 1.25$) and treat each of the others with only two sums ($\mathrm{AT} = 0.625/125 = 0.5$) to still achieve $\mathrm{CR} = 3.64$.

B. Comparison with other methods

In [19] the authors proposed both a method to construct binary matrices A with promising properties and a reconstruction mechanism tuned to ECG acquisition.

The matrix A is built by placing a certain number d of non-zeros in each column so that the mutual coherence of

TABLE III
PERFORMANCE COMPARISON BETWEEN THE PROPOSED APPROACH AND [19] IN THE CONDITIONS OF ITS SECTION III

m	d	CR $= n/m$	AT $= 1/d$	average PRD [19]	average PRD RAK	RAK is better
96	4	5.33	0.25	13.3%	10.7%	85%
128	5	4	0.2	8.85%	7.42%	80%
160	6	3.2	0.167	6.37%	5.81%	75%
256	8	2	0.125	3.40%	3.01%	83%
256	12	2	0.083	3.44%	3.09%	81%

the columns is kept as low as possible resulting in Minimal Mutual Coherence (MMC).

The reconstruction algorithm substitutes the general sparsity criterion with a priori knowledge on the decay of the coefficients of typical ECG signals when expressed on a Daubechies-6 wavelet basis. The resulting algorithm is indicated as Weighted ℓ_1 Minimization (WLM).

The overall approach is completely different from what we propose here since A is given good but general purpose properties, while adaptation to the signal is exploited at the reconstruction and is strictly linked to ECG-like signals.

Yet, performance comparison is possible and we refer to the setting of [19, Section III]. In particular, we consider the *MIT-BIH arrhythmia database* [40] and acquire those signals by means of binary matrices with $n = 512$ and either $m = 96$ with $d = 4$, or $m = 128$ with $d = 5$, or $m = 160$ with $d = 6$, or $m = 256$ with $d = 8$, or $m = 256$ with $d = 12$. For each configuration, MMC matrices are constructed by means of [19, Algorithm 1] [2] and Rakeness-based matrices are built as in subsection V.A with CR $= n/m$, AT $= 1/d$ and $\eta_j = d/m$ for $j = 0, \ldots, n-1$, sieving is used to ensure that the resulting matrix has exactly nd non-zeros.

Comparisons are made encoding randomly selected windows from the first tracks of each pair in the database with a MMC matrix and with a rakeness-based matrix, and decoding the two resulting measurement vectors with WLM. Performance is measured by Percentage Root-mean-square Difference (PRD) defined as $100 \times \|\xi - \hat{\xi}\|/\|\xi\|$.

Table III shows the result of such a comparison when 1000 trials are made for each configuration. The average PRD is reported for the two encoding strategies along with percentage of the times in which rakeness-based sensing outperforms MMC sensing. It is evident that, though performance tend to saturate to the same level, rakeness-based design is always convenient with respect to minimum coherence design.

As a second example, in [42] the authors apply their reconstruction algorithm called Block-Sparse Bayesian Learning (BSBL) to the difficult problem of acquiring a signal coming from ECG sensor in which mothers' and fetal tracks superimpose, so that the latter can be retrieved by Independent Component Analysis (ICA). To compare that method with the one propose here, we consider the setting in [42, sub-Section III.B], that is the most challenging one.

[2]The algorithm produces deterministic and highly structured matrices yielding poor performance. Though not declared in [19] best performance is obtained by shuffling columns to disrupt the unwanted structure while keeping the low value of mutual coherence.

TABLE IV
PERFORMANCE COMPARISON BETWEEN THE PROPOSED APPROACH AND [42] IN THE CONDITIONS OF ITS SECTION III

m	d	CR $= n/m$	AT $= 1/d$	average PCC [42]	average PCC RAK	RAK is better
256	12	2	0.083	0.876	0.936	96%
205	10	2.5	0.1	0.793	0.858	97%

The ground truth is the fetal ECG extracted by ICA directly from the non-compressed track.

We consider both matrices with d non-zeros per column whose positions are drawn at random as in [42], and rakeness-based matrices built as in the previous case with AT $= 1/d$. Signal windows of length $n = 512$ are encoded by means of either $m = 256$ with $d = 12$, or $m = 205$ with $d = 10$ (higher compression ratios do not allow a high quality recovery of the fetal track). The resulting measurement vectors are decoded by means of BSBL, and ICA is applied on the resulting tracks to find the fetal ECG. The quality of the acquisition is quantified by the Pearson's Correlation Coefficient (PCC) between the ground truth and the extracted fetal ECG.

Table IV shows the result of such a comparison when 100 trials are made for each configuration. The average PCC is reported for the two encoding strategies along with the percentage of the times in which rakeness-based sensing outperforms random sensing. Rakeness-based design is always convenient.

VII. CONCLUSION

Early-digital compressed sensing can be used to limit the resources needed by the acquisition and dispatch of sensor information in nodes that aim at being autonomous.

Along this direction, we develop a systematic procedure employing the rakeness concept and producing sparse sensing matrices whose non-zero elements are only ± 1 and with various structures addressing the resources needed by different stages in the signal chain from the ADC to the transmitter. The approach is extremely effective.

As an example, while the adoption of a rakeness-based design reduces the number of bits to be finally stored/transmitted to 21% of the conventional number, throttling reduces the amount of computation so that only 4 signed sums have to be performed for each incoming sample (see the case labeled B in Table II).

As a further example, puncturing can be called into play to allow the switch-off of the ADC and its analog preprocessing for entire sample times. Again, the use of rakeness-based design allows to keep reconstruction performance almost unaltered when, for example, up to 30% of the signal samples are simply skipped (see the case labeled C in Table II).

Beyond these cases, the overall design flow has been tested widely both on synthetic signals and on real-world ECG signals coming both from healthy and from pathological patients

yielding non-negligible improvements over both conventional general purpose techniques and over state-of-the-art methods specialized to ECG acquisition.

Besides this empirical evidence, future work will have to address the effect of our method on mutual coherence and, especially, on the mutual entropy between the signal x and the measurement y to formally quantify the amount of additional information that rakeness-based design is able to put into measurements.

VIII. Appendix

A. Solving T-RLT

To normalize the form of the T-RLT problem define $L = \left(\sum_{j=0}^{n-1}\eta_j\right)^2\left(\ell\mathfrak{L}_x+\frac{1}{n}\right)$ and rewrite the rakeness problem as

$$\begin{aligned}\max_{\mathcal{A}}\ & \langle\mathcal{A},\mathcal{X}\rangle\\ \text{s.t. }& (5)-(17)\\ \text{s.t. }& \|\mathcal{A}\|^2\leqslant L\end{aligned}\tag{22}$$

where we have recognized that for any two $n\times n$ real matrices X and Y, $\mathrm{tr}(XY)=\langle X,Y\rangle$ is a scalar product with a corresponding norm $\|X\|=\sqrt{\langle X,X\rangle}=\sqrt{\mathrm{tr}(X^2)}$, and (7) is translated into (22) by means of (3) and the definition of L.

Then, define few convex subsets of the space of $n\times n$ matrices, namely $\mathbf{C}_{\mathrm{PSD}}$ containing matrices satisfying (5) and (6), $\mathbf{C}_{\mathrm{Ave}}$ containing matrices satisfying (5) as well as (16), $\mathbf{C}_{\mathrm{Ter}}$ containing matrices satisfying (5) and (17), and $\mathbf{C}_{\mathrm{Loc}}$ containing matrices satisfying (5) and (22).

With this, the rakeness problem further translates into

$$\begin{aligned}\max_{\mathcal{A}}\ & \langle\mathcal{A},\mathcal{X}\rangle\\ \text{s.t. }& \mathcal{A}\in\mathbf{C}_{\mathrm{PSD}}\cap\mathbf{C}_{\mathrm{Ave}}\cap\mathbf{C}_{\mathrm{Ter}}\cap\mathbf{C}_{\mathrm{Loc}}\end{aligned}$$

that is a classical case of a bounded-gradient objective function ($\nabla_{\mathcal{A}}\langle\mathcal{A},\mathcal{X}\rangle=\mathcal{X}$) subject to convex constraint, i.e., something that can be effectively tackled by a *projected-gradient-ascent* [43].

To specialize such a method to our purpose indicate with $\Pi_{\mathbf{C}}$ the projection operator on a set of matrices $\mathbf{C}$, i.e., the operator that takes any $n\times n$ matrix X and maps it into the matrix $Y=\Pi_{\mathbf{C}}(X)$ such that $Y\in\mathbf{C}$ and $\|X-Y\|$ is minimum.

Then start from $\mathcal{A}^{(0)}=\mathrm{diag}(\eta_0,\ldots,\eta_{n-1})$ and compute iteratively

$$\mathcal{A}^{(t+1)}=\Pi_{\mathbf{C}_{\mathrm{PSD}}\cap\mathbf{C}_{\mathrm{Ave}}\cap\mathbf{C}_{\mathrm{Ter}}\cap\mathbf{C}_{\mathrm{Loc}}}\left(\mathcal{A}^{(t)}+\delta_t\mathcal{X}\right)\tag{23}$$

for suitably non-increasing scalars $\delta_0\geqslant\delta_1\geqslant\ldots$. It can be proved that the sequence of $\mathcal{A}^{(t)}$ converges to the optimal solution as $t\to\infty$.

The projection operator in (23) can be obtained by the alternating projection method [44], [45] using the individual projection operators on $\mathbf{C}_{\mathrm{PSD}}$, $\mathbf{C}_{\mathrm{Ave}}\cap\mathbf{C}_{\mathrm{Ter}}$, and $\mathbf{C}_{\mathrm{Loc}}$.

If X can be spectrally decomposed as $X=E\Lambda E^{\top}$, where E is the matrix of orthornormal eigenvectors and Λ is the diagonal matrix of the corresponding eigenvalues, then

$$\Pi_{\mathbf{C}_{\mathrm{PSD}}}(X)=E\max\{\Lambda,0\}E^{\top}$$

Moreover, if

$$\mathcal{A}^{\max}_{j,k}=\begin{cases}\eta_j & \text{if } j=k\\ \min\{\eta_j,\eta_k\} & \text{if } j\neq k\end{cases}$$

and

$$\mathcal{A}^{\min}_{j,k}=\begin{cases}\eta_j & \text{if } j=k\\ -\min\{\eta_j,\eta_k\} & \text{if } j\neq k\end{cases}$$

then

$$\Pi_{\mathbf{C}_{\mathrm{Ave}}\cap\mathbf{C}_{\mathrm{Ter}}}(X)=\max\{\mathcal{A}^{\min},\min\{\mathcal{A}^{\max},X\}\}$$

Additionally, due to the spherical nature of (22), it is easy to see that

$$\Pi_{\mathbf{C}_{\mathrm{Loc}}}(X)=X\min\left\{1,\frac{\sqrt{L}}{\|X\|}\right\}$$

All this allows an extremely efficient iteration of (23) to rapidly approximate $\mathcal{A}=\lim_{t\to\infty}\mathcal{A}^{(t)}$.

B. Gaussian-based generation of ternary vectors with prescribed correlation

From the fact that each g_j is a unit-variance Gaussian we have that $\Pr\{|g_j|\geqslant\theta_j\}=\mathrm{erfc}\left(\frac{\theta_j}{\sqrt{2}}\right)$ and from (19) we get $\Pr\{|g_j|\geqslant\theta_j\}=\Pr\{a_j^2=1\}=\mathcal{A}_{j,j}$. Hence we must set

$$\theta_j=\sqrt{2}\mathrm{erfc}^{-1}(\mathcal{A}_{j,j})\tag{24}$$

Then, consider that if the correlation γ between two unit-variance jointly Gaussian random variable α and β is known, then their joint probability density is

$$f(\alpha,\beta,\gamma)=\frac{1}{2\pi\sqrt{1-\gamma^2}}e^{-\frac{\alpha^2+\beta^2-2\gamma\alpha\beta}{2(1-\gamma^2)}}$$

so that the correlation between $\tau^{\mathrm{T}}_{\theta'}(\alpha)$ and $\tau^{\mathrm{T}}_{\theta''}(\beta)$, is

$$\begin{aligned}&\mathbf{E}[\tau^{\mathrm{T}}_{\theta'}(\alpha)\tau^{\mathrm{T}}_{\theta''}(\beta)]=\\ &=2\int_{\theta'}^{\infty}\int_{\theta''}^{\infty}f(\alpha,\beta,\gamma)\mathrm{d}\,\alpha\mathrm{d}\,\beta-2\int_{\theta'}^{\infty}\int_{-\infty}^{-\theta''}f(\alpha,\beta,\gamma)\mathrm{d}\,\alpha\mathrm{d}\,\beta\\ &=\frac{1}{\sqrt{2\pi}}\int_{\theta''}^{\infty}e^{-\frac{\beta^2}{2}}\mathrm{erfc}\left(\frac{\theta'-\gamma\beta}{\sqrt{2(1-\gamma^2)}}\right)\mathrm{d}\,\beta\\ &\quad-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\theta''}e^{-\frac{\beta^2}{2}}\mathrm{erfc}\left(\frac{\theta'-\gamma\beta}{\sqrt{2(1-\gamma^2)}}\right)\mathrm{d}\,\beta\end{aligned}\tag{25}$$

where we have exploited the fact that $f(-\alpha,-\beta,\gamma)=f(\alpha,\beta,\gamma)$.

Pairing (25) and (24) we obtain a function $T_{\eta',\eta''}$ such that $\mathcal{A}_{j,k}=T_{\mathcal{A}_{j,j},\mathcal{A}_{k,k}}(\mathcal{G}_{j,k})$.

Such a function cannot be given a fully analytical expression but has some recognizable properties. In particular, $T_{\eta',\eta''}(\gamma) = T_{\eta'',\eta'}(\gamma) = -T_{\eta',\eta''}(-\gamma)$ is continuous and monotonically increasing in γ, and can be extended by continuity in the domain $[-1,1]$ with $T_{\eta',\eta''}(\pm 1) = \pm\min\{\eta',\eta''\}$. Moreover, coherently with [34], we have $T_{1,1}(\gamma) = \frac{2}{\pi}\sin^{-1}(\gamma)$.

The range of $T_{\eta',\eta''}$ is compatible with that of a correlation between two ternary variables. Hence, any desired matrix $\mathcal{A}$ can be transformed into a corresponding $T^{-1}(\mathcal{A})$ defined componentwise as $(T^{-1}(\mathcal{A}))_{j,k} = T^{-1}_{\mathcal{A}_{j,j},\mathcal{A}_{k,k}}(\mathcal{A}_{j,k})$, where $T^{-1}_{\mathcal{A}_{j,j},\mathcal{A}_{j,j}}$ is the inverse of $T_{\mathcal{A}_{j,j},\mathcal{A}_{k,k}}$.

If we indicate with $\mathbf{C}_{\rm One}$ the set of $n \times n$ real symmetric matrices whose diagonal entries are equal to 1, we have that by definition $T^{-1}(\mathcal{A}) \in \mathbf{C}_{\rm One}$.

Yet, the correlation matrix of a jointly-Gaussian, zero-mean and unit-variance vector belongs both to $\mathbf{C}_{\rm One}$ and to $\mathbf{C}_{\rm PSD}$. To cope with both requirements it is sensible to set $\mathcal{G} = \Pi_{\mathbf{C}_{\rm PSD}\cap\mathbf{C}_{\rm One}}(T^{-1}(\mathcal{A}))$, use it to generate the jointly Gaussian vector g, and obtain $a_j = \tau^{\rm T}_{\theta_j}(g_j)$.

Note that, further to the above described $\Pi_{\mathbf{C}_{\rm PSD}}$, if $Y = \Pi_{\mathbf{C}_{\rm One}}(X)$ then $Y_{j,k} = X_{j,k}$ for $j \neq k$ and $Y_{j,j} = 1$ for $j = 0,\ldots,n-1$. Hence, $\Pi_{\mathbf{C}_{\rm PSD}\cap\mathbf{C}_{\rm One}}$ can be obtained by alternating projection methods [44], [45], [46] though even faster Newton-based methods exist [47].

C. Solving B-RLT

To adjust the above approach to the binary case define $\mathbf{C}_{\rm Bin}$ as the subset of the $n \times n$ real matrices containing matrices satisfying (5) and (20) and note that

$$\Pi_{\mathbf{C}_{\rm Ave}\cap\mathbf{C}_{\rm Bin}}(X) = \max\{\mathcal{A}^{\min'}, \min\{\mathcal{A}^{\max}, X\}\}$$

with

$$\mathcal{A}^{\min'}_{j,k} = \begin{cases} \eta_j & \text{if } j = k \\ \max\{0, \eta_j + \eta_k - 1\} & \text{if } j \neq k \end{cases}$$

D. Gaussian-based generation of binary vectors with prescribed correlation

In this case, given (21), we must set

$$\theta_j = \sqrt{2}\,\mathrm{erfc}^{-1}(2\mathcal{A}_{j,j}) \tag{26}$$

and observe that if two jointly-Gaussian zero-mean and unit-variance random variables α and β have correlation γ then

$$\mathbf{E}[\tau^{\rm B}_{\theta'}(\alpha)\tau^{\rm B}_{\theta''}(\beta)] = \int_{\theta'}^{\infty}\int_{\theta''}^{\infty} f(\alpha,\beta,\gamma)d\alpha d\beta = \\ = \frac{1}{2\sqrt{2\pi}}\int_{\theta''}^{\infty} e^{-\frac{\beta^2}{2}}\mathrm{erfc}\left(\frac{\theta' - \gamma\beta}{\sqrt{2(1-\gamma^2)}}\right)d\beta \tag{27}$$

Following the same path that we used for ternary vectors, we obtain a function $B_{\eta',\eta''}$ that transforms the correlation of jointly-Gaussian random variable in the correlation of the corresponding binarized random variable with assigned averages η' and η''.

This function has the same favorable properties as the function $T_{\eta',\eta''}$ of the ternary case. In particular, $B_{\eta',\eta''}(\gamma) = B_{\eta'',\eta'}(\gamma)$ is continuous and monotonically increasing in γ, and can be extended by continuity in the domain $[-1,1]$ with $B_{\eta',\eta''}(-1) = \max\{0,\eta' + \eta'' - 1\}$ and $B_{\eta',\eta''}(1) = \min\{\eta',\eta''\}$.

Hence, any desired matrix $\mathcal{A}$ can be transformed into $B^{-1}(\mathcal{A})$ defined component-wise as $(B^{-1}(\mathcal{A}))_{j,k} = B^{-1}_{\mathcal{A}_{j,j},\mathcal{A}_{k,k}}(\mathcal{A}_{j,k})$, where $B^{-1}_{\eta',\eta''}$ is the inverse of $B_{\eta',\eta''}$.

By setting $\mathcal{G} = \Pi_{\mathbf{C}_{\rm PSD}\cap\mathbf{C}_{\rm One}}(B^{-1}(\mathcal{A}))$, we may generate the jointly Gaussian vector g according to such a correlation matrix, and obtain $a_j = \tau^{\rm B}_{\theta_j}(g_j)$.

REFERENCES

[1] E.J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, n. 9, pp. 589–592, 2008

[2] D.L.Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 54, n. 4, pp. 1289–1306, 2006

[3] J. Haboba, M. Mangia, F. Pareschi, R. Rovatti, G. Setti, "A Pragmatic Look at Some Compressive Sensing Architectures With Saturation and Quantization," *IEEE Journal of Emerging and Selected Topics in Circuits and Systems*, vol. 2, n. 3, pp. 443–459, 2012

[4] 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 Transations on Biomedical Circuit and Systems*, vol. 10, no. 1, pp. 149-162, Feb. 2016

[5] F. Chen, A.P. Chandrakasan, V.M. Sojanovic, "Design and analysis of Hardware-Efficient Compressed Sensing Architecture for Data Compression in Wireless Sensors," *IEEE Journal of Solid State Circuits*, vol. 47, no. 3, pp. 744–756, 2012

[6] D. Bortolotti, M. Mangia, A. Bartolini, R. Rovatti, G. Setti, L. Benini, "Rakeness-based Compressed Sensing on Ultra-Low Power Multi-Core Biomedical Processors," proceedings of Design & Architectures for Signal & Image Processing, pp. 1–8, 2014

[7] V. Cambareri, M. Mangia, F. Pareschi, R. Rovatti, G. Setti, "A Case Study in Low-Complexity ECG Signal Encoding: How Compressing is Compressed Sensing?," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1743–1747, Oct. 2015

[8] 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

[9] R. Rovatti, G. Mazzini, G. Setti, "Enhanced rake receivers for chaos-based DS-CDMA", *IEEE Transactions on Circuits and Systems I*, vol. 48, no. 7, pp. 818-829, July 2001

[10] G. Setti, R. Rovatti, G. Mazzini, "Performance of chaos-based asynchronous DS-CDMA with different pulse shapes", *IEEE Communications Letters*, vol. 8, no. 7, pp. 416-418, 2004

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

[12] E.J. Candes, J.K. Romberg, T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on pure and applied mathematics*, vol. 59, n. 8, pp. 1207–1223, 2006

[13] D. Needell, 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

[14] D. L. Donoho, A. Maleki, A. Montanari, "Message passing algorithms for compressed sensing," *Proceeding of the National Academy of Science*, vol. 106, pp. 18914—18919, 2009

[15] V. Cambareri, M. Mangia, F. Pareschi, R. Rovatti, G. Setti, "A rakeness-based design flow for Analog-to-Information conversion by Compressive Sensing," IEEE International Symposium on Circuits and Systems (ISCAS 2013) pp. 1360–1363, 2013

[16] 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

[17] E.J. Candes, T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, no.12, pp. 5406–5425, Dec 2006

[18] N. Cleju, "Optimized projections for compressed sensing via rank-constrained nearest correlation matrix," *Applied and Computational Harmonic Analysis*, vol.36, pp. 495–507, 2014

[19] J. Zhang, Z. Gu, Z. Liang Yu, Y. Li, "Energy-Efficient ECG Compression on Wireless Biosensors via Minimal Coherence Sensing and Weighted ℓ_1 Minimization Reconstruction," *IEEE Journal of Biomededical and Health Informatics*, vol. 19, no. 2, pp. 520–528, 2015

[20] D. Bellasi, R. Rovatti, L. Benini, G. Setti, "A Low-Power Architecture for Punctured Compressed Sensing and Estimation in Wireless Sensor-Nodes," *IEEE Trasaction on Circuits and Systems I*, vol.62, no.5, pp. 1296–1305, May 2015

[21] B. Murmann, "ADC Performance Survey 1997-2016," [Online]. Available: http://web.stanford.edu/ murmann/adcsurvey.html.

[22] D. Shum, J. Power, R. Ullmann, E. Suryaputra, K. Ho, J. Hsiao, C. Tan, W. Langheinrich, C. Bukethal, V. Pissors, et al., "Highly reliable ash memory with self-aligned split-gate cell embedded into high performance 65nm cmos for automotive & smartcard applications," IEEE International Memory Workshop, pp. 1–4, 2012

[23] N. Gilbert, Y. Zhang, J. Dinh, B. Calhoun, S. Hollmer, "A 0.6v 8 pj/write non-volatile cbram macro embedded in a body sensor node for ultra low energy applications," Symposium on VLSI Circuits, pp. C204–C205,2013

[24] S. Yu and P. Y. Chen, "Emerging Memory Technologies: Recent Trends and Prospects," in IEEE Solid-State Circuits Magazine, vol. 8, no. 2, pp. 43–56, 2016

[25] G. Papotto, F. Carrara, A. Finocchiaro, G. Palmisano, "A 90nm cmos 5mb/s crystal-less rf transceiver for rf-powered wsn nodes," IEEE International Solid-State Circuits Conference, pp. 452–454,2012

[26] Y. H. Liu, X. Huang, M. Vidojkovic, A. Ba, P. Harpe, G. Dol- mans, H. de Groot, "A 1.9 nj/b 2.4 ghz multistandard (bluetooth low energy/zigbee/ieee802. 15.6) transceiver for personal/body- area networks," IEEE International Solid-State Circuits Conference, pp. 446–447, 2013

[27] W. L. Lien, T. Y. Choke, Y. C. Tan, M. Kong, E. C. Low, D. P. Li, L. Jin, H. Zhang, C. H. Leow, S. L. Chew, et al., "9.1 a self- calibrating nfc soc with a triple-mode reconfigurable pll and a single-path picc-pcd receiver in 0.11μm cmos," IEEE International Solid-State Circuits Conference, pp. 158–159, 2014

[28] 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

[29] M. Grant, S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.0 beta. `http://cvxr.com/cvx`, September 2013

[30] K.C. Toh, M.J. Todd, R.H. Tutuncu, "SDPT3 — a Matlab software package for semidefinite programming," *Optimization Methods and Software*, no. 11, pp. 545–581, 1999

[31] G. Jacovitti, A. Neri, G. Scarano, "Texture synthesis-by-analysis with hard-limited Gaussian processes," *IEEE Transactions on Image Processing*, vol. 7, no. 11, pp. 1615–1621, 1998

[32] R. Rovatti, G. Mazzini, G. Setti, S. Vitali, "Linear probability feedback processes," IEEE International Symposium on Circuits and Systems 2008, pp 548–551

[33] 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

[34] J.H. Van Vleck, D. Middleton, "The Spectrum of Clipped Noise," *IEEE Proceedings*, vol. 54, n. 1, pp. 2–19, 1966

[35] E. van den Berg, M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, Nov. 2008

[36] P.E. McSharry, G.D. Clifford, L. Tarassenko, L.A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 3, pp. 289—294, Mar. 2003

[37] S. Mallat, *A wavelet tour of signal processing: the sparse way*, Elsevier, 2008

[38] N. Perraudin, D. Shuman, G. Puy, P. Vandergheynst, "UNLocBoX: A matlab convex optimization toolbox using proximal splitting methods," *arXiv preprint*, arXiv:1402.0779, 2014

[39] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.Ch. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C-K. Peng, H.E. Stanley, *PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals*, Circulation vol. 101, n. 23, pp. e215-e220, 2000

[40] G.B. Moody, R.G. Mark, "The impact of the MIT-BIH Arrhythmia Database," *IEEE Engineering in Medicine and Biology*, vol. 20, n. 3, pp. 45–50, 2001

[41] G.B. Moody, W.E. Muldrow, R.G. Mark, "A noise stress test for arrhythmia detectors," *Computers in Cardiology*, vol. 11, pp. 381–384, 1984

[42] Z. Zhang,T.-P. Jung, S. Makeig, 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. 3, pp. 300–309, 2013

[43] A.A. Goldstein, "Convex Programming in Hilbert Space," *Bullettin of the American Mathematical Society*, vol. 70, pp. 709–710, 1964

[44] R.L. Dykstra, "An Algorithm for Restricted Least Squares Regression," *Journal of the American Statistical Association*, vol. 78, n. 384, pp. 837–842, 1983

[45] J.P. Boyle, R.L. Dykstra, "A method for finding projections onto the intersection of convex sets in Hilbert spaces," *Lecture Notes in Statistics*, vol. 37, pp. 28—47, 1986

[46] N.J. Higham, "Computing the Nearest Correlation Matrix — A Problem from Finance," IMA Journal of Numerical Analysis, vol. 22, pp. 329–343, 2002

[47] Q. Hou-Duo,D. Sun, "A Quadratically Convergent Newton Method for Computing the Nearest Correlation Matrix," *SIAM Journal on Matrix Analysis and Applications*, vol. 28, pp. 360–385, 2006

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

Fabio Pareschi (S'05-M'08) received the Ph.D. in Information Technology under the European Doctorate Project (EDITH) from University of Bologna in 2007. He is currently an Assistant Professor in the Department of Engineering (ENDIF), University of Ferrara. He served as Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS - PART II (2010-2013). His research activity focuses on analog and mixed-mode electronic circuit design, statistical signal processing, random number generation and testing, and electromagnetic compatibility.

Valerio Cambareri (S'13) received the Ph.D. in electronic engineering from the University of Bologna in 2015. In 2014 he was a Visiting Ph.D. Student in the Integrated Imagers team at IMEC, Belgium. In 2015 he joined the Image and Signal Processing Group of ICTEAM-ELEN, Université catholique de Louvain, Belgium, as a Postdoctoral Researcher. His research activity focuses on statistical and digital signal processing, compressed sensing, and computational imaging.

Riccardo Rovatti (M'99-SM'02-F'12) received the Ph.D. in Electronics, Computer Science, and Telecommunications from the University of Bologna in 1996. He is now a Professor of Electronics at the University of Bologna. 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.

Gianluca Setti (S'89-M'91-SM'02-F'06) received the Ph.D. in Electronics, Computer Science, and Telecommunications from the University of Bologna in 1997. He is now a Professor of Circuit Theory at the University of Ferrara. His research interests include implementation and application of chaotic circuits and systems, electromagnetic compatibility, statistical signal processing and biomedical circuits and systems. He received the 2013 IEEE CAS Society Meritorious Service Award, the 2004 IEEE CAS Society Darlington Award, and the 2013 IEEE CAS Society Guillemin-Cauer Award. He served as Editor-in-Chief for the IEEE TRANSACTION ON CIRCUITS AND SYSTEMS—PART II and for the IEEE TRANSACTION ON CIRCUITS AND SYSTEMS—PART I. A member of the BoG of the IEEE CAS Society, he served as its 2010 President. In 2013-2014 he was the first non North-American IEEE Vice President for Publication Services and Products.