Full length article

Compressive sampling based differential detection for UWB impulse radio signals

Shahzad Gishkori [a],[*], Geert Leus [a], Vincenzo Lottici [b]

[a] *Faculty of EEMCS, Delft University of Technology, Delft, The Netherlands*
[b] *Department of Information Engineering, University of Pisa, Pisa, Italy*

ABSTRACT

Noncoherent detectors significantly contribute to the practical realization of the ultra-wideband (UWB) impulse-radio (IR) concept, in that they allow avoiding channel estimation and provide highly efficient reception capabilities. Complexity can be reduced even further by resorting to an all-digital implementation, but Nyquist-rate sampling of the received signal is still required. The current paper addresses this issue by proposing a novel differential detection (DD) scheme, which exploits the compressive sampling (CS) framework to reduce the sampling rate much below the Nyquist-rate. The optimization problem is formulated to jointly recover the sparse received signal as well as the differentially encoded data symbols, and is compared with both the separate approach and the scheme using the compressed received signal directly, i.e., without reconstruction. Finally, a maximum *a posteriori* based detector using the compressed symbols is developed for a Laplacian distributed channel, as a reference to compare the performance of the proposed approaches. Simulation results show that the proposed joint CS-based DD brings the considerable advantage of reducing the sampling rate without degrading the performance, compared with the optimal MAP detector.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Ultra-wideband (UWB) impulse radio (IR) is a promising signaling scheme, particularly suitable for low-power short-range communications, in virtue of many appealing features, such as high user capacity, fine timing resolution, frequency-overlay based coexistence with existing services, low probability of interception and detection [1,2]. Rich multipath propagation, however, makes each transmitted pulse appear at the receiver as hundreds of echoes [3]. Although Rake receivers allow to collect most of the energy conveyed by the multipath components [4], they require a large number of fingers together with an intensive computational load and a high sampling rate

to perform channel estimation [5], thus contradicting the main requirement of simple transceiver devices. As a sub-optimal yet effective alternative, noncoherent receivers have been proposed in order to skip the difficult channel estimation task, in the form of autocorrelation based receivers (AcRs) [6]. We can refer to transmitted reference (TR), where a reference pulse is transmitted together with the data pulse [7,8], and differential detection (DD), which employs differential encoding [9]. The detection performance of DD schemes can be further improved with the multi-symbol DD approach (MSDD) [10,11], and its variant based on symbol-level synchronization only [12], even though for an all-digital implementation they are all affected by the basic issue of still requiring high rate analog-to-digital converters (ADCs).

Relations with prior work. The compressive sampling (CS) concept has been recently pursued as a powerful way to reduce the sampling rate of sparse signals much below the Nyquist-rate without incurring large performance

* Corresponding author. Tel.: +31 152782845; fax: +31 152786190.
E-mail addresses: s.s.gishkori@tudelft.nl (S. Gishkori), g.j.t.leus@tudelft.nl (G. Leus), vincenzo.lottici@iet.unipi.it (V. Lottici).

degradations [13,14]. The key idea relies on representing a sparse signal with a few measurements only, obtained via random projections in the analog domain [15,16], and then, reconstructing it through a sparse recovery method. Now, exploiting the fact that the received UWB signal can be considered to be sparse in the time domain [17], we can argue that the CS-based approach can be useful for data detection. Toward this direction, a few works have recently appeared, such as [18] for coherent receivers, [19] for joint time of arrival (ToA) estimation and data decoding, and [20] for a generalized likelihood ratio test (GLRT) detector based on the transmission of pilot symbols.

Purpose and contributions. In this paper, we focus on CS-based noncoherent receivers for differentially encoded UWB signals, as preliminarily discussed in [21]. A few important features are gained which differentiate our contributions from previous works.

(1) The key to our method is the formulation of a cost function, as the composition of the sparse regularized mean squared errors for the two compressed-rate consecutively received signal waveforms, combined with the squared DD error, which is minimized using an efficient iterative method derived from the elastic net optimization framework [22]. Thus, reconstruction from the compressed signal samples and detection of encoded information is performed jointly.

(2) The proposed CS-based DD does not require any channel estimation as in [19] nor the transmission of pilot symbols as in [20].

(3) A simpler two-step approach is formulated wherein first the sparse regularized mean squared error is minimized, and then the recovered symbol waveforms are used to perform conventional DD.

(4) A direct detection method working directly on the compressed samples is considered as well, which avoids signal reconstruction, and its limitations are clarified.

(5) A compressed-rate MAP DD is derived as a performance benchmark for the proposed detectors, assuming a Laplacian distributed channel response which emulates sparsity.

Organization. The rest of the paper is organized as follows. Section 2 describes the signal model, Section 3 introduces the CS-based separate and joint reconstruction and detection methods, while Section 4 derives the MAP based DDs at both Nyquist- and compressed-rate. Simulation results are discussed in Section 5, and finally concluding remarks are drawn in Section 6.

Notations. Matrices are in upper case bold while column vectors are in lower case bold, $[\mathbf{A}]_{i,j}$ is the (i,j)th entry of the matrix $\mathbf{A}$, $[\mathbf{a}]_i$ is the ith entry of the vector $\mathbf{a}$, $\mathbf{I}_N$ is the identity matrix of size $N \times N$, $|\mathbf{A}|$ is the determinant of the matrix $\mathbf{A}$, $(\cdot)^T$ is transpose, $(\cdot)^+$ is pseudo-inverse, $\otimes$ stands for the Kronecker product, diag$\{\cdot\}$ gives a block matrix having the arguments along its main diagonal, $\hat{x}$ is the estimate of x, $\mathrm{E}\{\cdot\}$ denotes expectation, $P(\cdot)$ and $p(\cdot)$ represent probability distribution and probability density function (pdf), respectively, $\triangleq$ defines an entity, $\|\mathbf{a}\|_p = \left(\sum_{i=0}^{N-1} |[\mathbf{a}]_i|^p\right)^{1/p}$ is the ℓ_p norm of $\mathbf{a}$, sign(x) is the sign function which takes values -1 and 1 depending on the polarity of the element x, whereas the function $(x)_+ = x$ if and only if $x > 0$ otherwise $(x)_+ = 0$.

2. Signal model

In the adopted IR-UWB signal model, each symbol is conveyed by a pulse $q(t)$ of duration T_q much less than the symbol interval T_s, i.e., $T_q \ll T_s$.[1] The transmitted signal composed of a block of Q symbols takes the form

$$s(t) = \sum_{k=0}^{Q-1} b_k q(t - kT_s) \tag{1}$$

where $b_k \in \{\pm 1\}$ are the differentially encoded transmitted symbols, i.e., $b_k = b_{k-1}a_k$, $a_k \in \{\pm 1\}$ being the information symbols. As a reference transmitted symbol, without loss of generality we take $b_{-1} = 1$.

The signal travels through a slow-fading multipath channel, assumed to be time-invariant within the interval of Q consecutive symbols, and with delay spread smaller than T_s, so that inter symbol interference (ISI) is avoided. Let $g(t) \triangleq \sum_{l=0}^{L-1} \alpha_l \delta(t - \tau_l)$ represent the channel impulse response (CIR) with L paths, where α_l and τ_l are the gain and path delay of the lth path, respectively.

The received signal $r(t)$ can then be written as

$$r(t) = \underbrace{\sum_{k=0}^{Q-1} b_k h(t - kT_s)}_{\triangleq x(t)} + v(t) \tag{2}$$

where $h(t) \triangleq \sum_{l=0}^{L-1} \alpha_l q(t - \tau_l)$ is the received pulse, and $v(t)$ is the zero mean additive white Gaussian noise component with variance σ_v^2. Denoting the Nyquist sampling rate with $1/T = N/T_s$, the received signal in its sampled version can be written as $\mathbf{r} \triangleq [\mathbf{r}_0^T, \mathbf{r}_1^T, \dots, \mathbf{r}_{Q-1}^T]^T$ where $\mathbf{r}_k \triangleq [r(kT_s), r(kT_s + T), \dots, r(kT_s + NT - T)]^T$ collects the N Nyquist-rate samples corresponding to the kth symbol. In view of (2), it can be written

$$\mathbf{r}_k = \mathbf{x}_k + \mathbf{v}_k = b_k \mathbf{h} + \mathbf{v}_k \tag{3}$$

where $\mathbf{h} \triangleq [h(0), h(T), \dots, h(NT - T)]^T$ is the sampled CIR whose entries are modeled as independent and identically distributed (i.i.d.) Laplacian random variables (owing to the sparse nature of the UWB channel), and $\mathbf{v}_k \triangleq [v(kT_s), v(kT_s + T), \dots, v(kT_s + NT - T)]^T$ is a zero mean Gaussian random vector with covariance matrix $\mathrm{E}\{\mathbf{v}_k \mathbf{v}_k^T\} = \sigma_v^2 \mathbf{I}_N$.

We can observe that the signal vector $\mathbf{x}_k$ is generally sparse due to the fact that the channel $\mathbf{h}$ is sparse, i.e., most of its components are zero or negligible [3]. Thus, according to the CS framework theory [13,14], it can be represented by M linear measurements, with $M \ll N$. This is generally obtained through analog processing of $r(t)$, as illustrated in [15,16]. For the sake of convenience, however, the model we will adopt here is based on an operation that is performed on the Nyquist rate samples of $r(t)$. Hence,

---

[1] Generalizations of the proposed framework to signaling based on multiple frames to comply with the FCC power spectral density requirements [23] can be easily performed, and so for the sake of simplicity, it will not be addressed.

the compressed received signal within one symbol can be expressed as

$$\mathbf{y}_k = \boldsymbol{\Phi}_k \mathbf{r}_k = \boldsymbol{\Phi}_k \mathbf{x}_k + \boldsymbol{\xi}_k \qquad (4)$$

where the $M \times N$ matrix $\boldsymbol{\Phi}_k$ is the measurement matrix at time instant k and $\boldsymbol{\xi}_k \triangleq \boldsymbol{\Phi}_k \mathbf{v}_k$ is the noise component. It is worth recalling that $\boldsymbol{\Phi}_k$ satisfies the restricted isometry property (RIP) [14], thus allowing the recovery of the received signal from its CS version in the asymptotic sense as a function of the number of measurements M, with $M \ll N$ [24]. A wide range of both random (Gaussian or Bernoulli) and structured (Fourier or identity) measurement matrices satisfy the RIP, although the latter have been proved to be the better choice for smaller N. An important parameter that has a direct influence on the performance of CS-based systems is the compression ratio defined as $\mu \triangleq M/N$, with $\mu \in (0, 1]$. A higher value of μ implies a higher value of M and hence a better performance, whereas on the other side, a lower M is desirable to keep the sampling rate at affordable levels, although this is usually achieved at the price of a given performance degradation.

3. Compressed-sensing based detection

Several methods are available to recover differentially encoded information from the samples of the received signal. Considering that each received symbol waveform is obtained in compressed form, data decoding may optionally require prior signal reconstruction followed by differential detection, or alternatively, a joint reconstruction and detection process, as illustrated in the sequel.

3.1. Conventional differential detection

Differential detection involves the correlation between consecutive symbols within a received block. In the case of Nyquist-rate differential detection (NDD), the estimate of the information symbol can be expressed as

$$\hat{a}_{k+1}^{(\mathrm{NDD})} = \mathrm{sign}\left(\arg\min_a \left\{ \| \mathbf{r}_k - a\mathbf{r}_{k+1} \|_2^2 \right\} \right). \qquad (5)$$

Hence from (5), it can be seen that one possible yet coarse way of decoding information from the compressed received signal consists of performing correlation directly on the compressed samples. We will designate this approach as direct compressed differential detection (DC-DD), which can be described as

$$\hat{a}_{k+1}^{(\mathrm{DC\text{-}DD})} = \mathrm{sign}\left(\arg\min_a \left\{ \| \mathbf{y}_k - a\mathbf{y}_{k+1} \|_2^2 \right\} \right). \qquad (6)$$

This method does not involve sparse reconstruction of the actual received signal, but exploits only the compressed waveform $\mathbf{y}_k$ given by (4). We note however that the DC-DD works under the condition that every compressed symbol waveform is the result of the same linear transformation of the received signal, otherwise it may exhibit strong limitations. We will come back to this aspect in the following subsections. Nevertheless, direct compressed detection can be favorably applied when synchronization requirements may be relaxed (and accordingly, signal reconstruction can be avoided), such as for instance in [12].

3.2. Overview of reconstruction techniques

Focusing on the reconstruction of $\mathbf{x}_k$, a naive way is to adopt the ordinary least squares (OLS) optimization method, thus obtaining from (4)

$$\hat{\mathbf{x}}_k^{(\mathrm{OLS})} = \arg\min_{\mathbf{x}_k} \left\{ \| \mathbf{y}_k - \boldsymbol{\Phi}_k \mathbf{x}_k \|_2^2 \right\}. \qquad (7)$$

Due to the fact that the $M \times N$ measurement matrix $\boldsymbol{\Phi}_k$ is fat ($M \ll N$), and so not full column rank, the solution to the OLS problem in (7) is not unique. One way to circumvent this drawback is to use Tikhonov regularization based on the ℓ_2 norm, which penalizes the OLS cost function with a quadratic penalty, also known as ridge regression (RR), leading to

$$\hat{\mathbf{x}}_k^{(\mathrm{RR})} = \arg\min_{\mathbf{x}_k} \left\{ \| \mathbf{y}_k - \boldsymbol{\Phi}_k \mathbf{x}_k \|_2^2 + \lambda \| \mathbf{x}_k \|_2^2 \right\} \qquad (8)$$

where λ is the Lagrangian constant. Although the RR solution is unique, it does not care about the sparsity of $\mathbf{x}_k$. A specific solution to this problem is the least absolute shrinkage and selection operator (LASSO) [25], which adopts a regularization term based on the ℓ_1 norm, as

$$\hat{\mathbf{x}}_k^{(\mathrm{LASSO})} = \arg\min_{\mathbf{x}_k} \left\{ \| \mathbf{y}_k - \boldsymbol{\Phi}_k \mathbf{x}_k \|_2^2 + \lambda \| \mathbf{x}_k \|_1 \right\} \qquad (9)$$

where λ is again the Lagrangian constant. Due to the ℓ_1 regularization that induces sparsity, part of the entries of $\hat{\mathbf{x}}_k^{(\mathrm{LASSO})}$ will be switched off (hopefully the noisy or the non-significant ones), under the condition that the value of λ is properly chosen. This appealing feature explains why the interest in the LASSO technique is growing more and more whenever a sparse signal has to be reconstructed. The above fully motivates the adoption of LASSO, or its modified versions, to address the CS-based detection problem we are dealing with, as will be illustrated in the rest of this section.

3.3. Separate reconstruction and detection

According to the separate compressed differential detection approach (SC-DD), the sparse received signal is first reconstructed from the compressed samples applying the LASSO algorithm, and is subsequently used to decode the information symbols through correlation of consecutive symbol waveforms. Among the various algorithms to solve the LASSO problem, we mention the LARS scheme [26], which has a low complexity but requires $M > N$, and the one proposed in [27], which is applicable for $M < N$ but is computationally intensive. On the other side, the pathwise coordinate descent (PCD) optimization idea is proposed in [28] as a way to solve the LASSO problem, and turns out to be particularly competitive as far as the computational complexity aspects are concerned. The PCD is based on optimizing one coordinate of $\mathbf{x}_k$ at-a-time, while all the others are kept at the values evaluated at the previous iteration, so that each update works as a warm start for the next step. Hence, the PCD solution to (9)

for the $(n + 1)$th iteration, $n \geq 0$, and the jth coordinate, $1 \leq j \leq N$, of $\hat{\mathbf{x}}_{k+l}$, $l = 0, 1$ can be proved to be [28]

$$
[\hat{\mathbf{x}}_{k+l}]_j(n + 1)
$$

$$
= \text{shrink} \left(\sum_{i=1}^{M} [\boldsymbol{\Phi}_{k+l}]_{i,j} \{ [\mathbf{y}_{k+l}]_i - [\hat{\mathbf{y}}_{k+l}^{(j)}]_i(n + 1) \}, \lambda \right) (10)
$$

where the "shrink" operator is defined as shrink $(z, \lambda) \triangleq \text{sign}(z)(|z| - \lambda)_+$, with the parameter λ optimized through a cross-validation (CV) approach (Section 5.1), and $[\hat{\mathbf{y}}_{k+l}^{(j)}]_i(n + 1)$ is evaluated as

$$
[\hat{\mathbf{y}}_{k+l}^{(j)}]_i(n + 1) = \sum_{m<j} [\boldsymbol{\Phi}_{k+l}]_{i,m} [\hat{\mathbf{x}}_{k+l}]_m(n + 1)
$$

$$
+ \sum_{m>j} [\boldsymbol{\Phi}_{k+l}]_{i,m} [\hat{\mathbf{x}}_{k+l}]_m(n), \quad (11)
$$

i.e., excluding the effect of the jth coordinate $[\hat{\mathbf{x}}_{k+l}]_j(n)$, and using for the earlier $(j - 1)$ entries the values updated at the current $(n + 1)$th iteration, namely $[\hat{\mathbf{x}}_{k+l}]_1(n + 1), \ldots, [\hat{\mathbf{x}}_{k+l}]_{j-1}(n + 1)$, and for the remaining ones, namely $[\hat{\mathbf{x}}_{k+l}]_{j+1}(n), \ldots, [\hat{\mathbf{x}}_{k+l}]_N(n)$, those values updated at the previous iteration. The PCD iterations (10) and (11) continue till convergence, i.e., when a predefined tolerance level has been reached for each coordinate. Next, from the symbol waveform estimates $\hat{\mathbf{x}}_k(P)$ and $\hat{\mathbf{x}}_{k+1}(P)$ reconstructed after P iterations, we can obtain the detected symbol as

$$
\hat{a}_{k+1}^{(\text{SC-DD})} = \text{sign} \left(\hat{\mathbf{x}}_{k+1}(P)^T \hat{\mathbf{x}}_k(P) \right). \quad (12)
$$

The computational complexity required by the PCD algorithm for each reconstruction iteration can be shown to be $\mathcal{O}(NM)$ [29], while that for the detection step is simply equal to $\mathcal{O}(N)$. Therefore, the overall complexity of the SC-DD for P iterations amounts to $\mathcal{O}(PNM)$.

3.4. Joint reconstruction and detection

An alternative to the SC-DD approach is to perform joint reconstruction and detection, which will be referred to as the joint compressed differential detection (JC-DD) approach. Formally, the corresponding cost function of the JC-DD optimization problem to be minimized over $\mathbf{x}_k$, $\mathbf{x}_{k+1}$ and a_{k+1} can be formulated as

$$
\mathcal{C}_{k+1}^{(\text{JC-DD})} (\mathbf{x}_k, \mathbf{x}_{k+1}, a_{k+1})
$$

$$
\triangleq \sum_{l=0}^{1} \left[\|\mathbf{y}_{k+l} - \boldsymbol{\Phi}_{k+l}\mathbf{x}_{k+l}\|_2^2 + \lambda \|\mathbf{x}_{k+l}\|_1 \right]
$$

$$
+ \alpha \|\mathbf{x}_k - a_{k+1}\mathbf{x}_{k+1}\|_2^2 \quad (13)
$$

where λ is the Lagrangian constant and α is a weight constant. The following remarks about the JC-DD are now of interest.

(1) The parameter α has to be chosen by trading off the performance of the reconstruction against the detection steps. A higher value may result in a wrong correlation estimate due to excess noise on $\hat{\mathbf{x}}_k$ and $\hat{\mathbf{x}}_{k+1}$. Conversely, a lower value may be detrimental as well due to an accuracy loss in the estimate $\hat{a}_{k+1}$. Indeed, in

that case the JC-DD collapses into the SC-DD approach, where we first reconstruct independent of detection, and then detect optimizing only with respect to a_{k+1}.

(2) In view of the joint optimization, the reconstruction and detection steps reinforce each other during iterations. Therefore, improved performance over both the DC-DD and SC-DD is expected.

(3) Several regression methods are available to minimize the cost function (13), even though we will show in a while that none of them exhibits the regularization features that properly match the JC-DD problem. Generally speaking, denoting with $\mathbf{u}$ and $\mathbf{z}$ the vectors with size M and N, collecting the received compressed samples and the samples to be optimally reconstructed, respectively, and with $\boldsymbol{\Lambda}$ an $M \times N$ measurement matrix, we can basically enumerate the following three methods.

- *Standard LASSO*. Taking into account (9), the standard LASSO can be put into the form

$$
\begin{cases}
\hat{\mathbf{z}}^{(\text{LASSO})} = \arg \min_{\mathbf{z}} \left\{ \|\mathbf{u} - \boldsymbol{\Lambda}\mathbf{z}\|_2^2 \right\} \\
\text{s.t.} \quad \|\mathbf{z}\|_1 \leq \gamma
\end{cases} \quad (14)
$$

where γ is a given threshold. We note that the main effect of the constraint based on the ℓ_1 norm is to induce parsimony in the solution, in the sense that among all the feasible solutions (14) takes specific care of those solutions with higher sparsity. However, no quadratic constraint on the optimization variables is involved as required by the JC-DD cost function (13). Therefore, it can be concluded that the standard LASSO is of scarce utility for our purpose and some alternatives have to be searched for.

- *Elastic net*. Elastic net (EN) is a modified version of LASSO where a quadratic constraint is considered as well [22], according to the form

$$
\begin{cases}
\hat{\mathbf{z}}^{(\text{EN})} = \arg \min_{\mathbf{z}} \left\{ \|\mathbf{u} - \boldsymbol{\Lambda}\mathbf{z}\|_2^2 \right\} \\
\text{s.t.} \quad \|\mathbf{z}\|_1 \leq \gamma_1 \\
\quad\quad \|\mathbf{z}\|_2^2 \leq \gamma_2,
\end{cases} \quad (15)
$$

with γ_1 and γ_2 being predefined thresholds. The added constraint has the effect of grouping the elements of the optimization vector $\mathbf{z}$, which adds to the action of favoring sparse solutions played by the ℓ_1-based constraint. The actual result is that parts of $\mathbf{z}$ will be different from zero and others will be negligible, thus matching the cluster-based propagation encountered in typical UWB environments [3], but again what is now lacking is the differential aspect related to the JC-DD cost function (13).

- *Fused LASSO*. An additional variant of LASSO is represented by the fused LASSO (F-LASSO), which is proposed in [30] as

$$
\begin{cases}
\hat{\mathbf{z}}^{(\text{F-LASSO})} = \arg \min_{\mathbf{z}} \left\{ \|\mathbf{u} - \boldsymbol{\Lambda}\mathbf{z}\|_2^2 \right\} \\
\text{s.t.} \quad \|\mathbf{z}\|_1 \leq \gamma_1 \\
\quad\quad \sum_{j=2}^{N} \left| [\mathbf{z}]_j - [\mathbf{z}]_{j-1} \right| \leq \gamma_2.
\end{cases} \quad (16)
$$

The F-LASSO method penalizes the cost function with not only the sum of the absolute values of the coefficients of the optimization variable, i.e., $\|\mathbf{z}\|_1$, but also their differences. That way, sparsity is induced while "fusing" successive coefficients to each other, but again, these features are not exactly what is required.

(4) From the regularization methods (14)–(16), it is apparent that none of them satisfies the requirements for the optimization of the JC-DD cost function, including both an ℓ_1-based as well as a squared differential penalty on two sets of optimization variables and not just one. Hence, this need fully motivates the development of a different method that we will focus on in the next subsection.

3.5. Differential elastic net

We propose here a novel regularization method, which we will designate as differential elastic net (DEN), and which can be formulated as

$$\begin{cases} (\hat{\mathbf{z}}_1, \hat{\mathbf{z}}_2, \hat{a})^{(\text{DEN})} = \underset{\mathbf{z}_1, \mathbf{z}_2, a}{\arg\min} \left\{ \sum_{l=1}^{2} \left[\|\mathbf{u}_l - \boldsymbol{\Lambda}_l \mathbf{z}_l\|_2^2 \right] \right\} \\ \text{s.t.} \quad \|\mathbf{z}_1\|_1 \leq \gamma_1 \\ \qquad \|\mathbf{z}_2\|_1 \leq \gamma_1 \\ \qquad \|\mathbf{z}_1 - a\mathbf{z}_2\|_2^2 \leq \gamma_2 \end{cases} \tag{17}$$

where $\mathbf{z}_1$ and $\mathbf{z}_2$ are the two sets of variables to be optimally reconstructed, each with size N, $\mathbf{u}_1$ and $\mathbf{u}_2$ are the two sets of compressed samples, each with size M, and $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ are the corresponding $M \times N$ measurement matrices. The rationale of the DEN method relies on searching the sparse solutions $\hat{\mathbf{z}}_1$ and $\hat{\mathbf{z}}_2$ while imposing at the same time fusion between their respective elements, together with deriving the optimal estimate $\hat{a}$ of the transmitted information symbol.

As an effective way to solve (17), we resort to the PCD algorithm illustrated in Section 3.3. Due to its iterative nature, convergence to a unique solution may be an issue. Indeed, convergence of the PCD is typically not ensured for non-differentiable cost functions. It has been proved, however, that an exception occurs whenever the non-differentiable part is separable in its variables [31]. Interestingly the ℓ_1 part in the cost function (13) just satisfies that condition, and accordingly, this proves the uniqueness of the PCD solution to (17). Now, the DEN solutions to (17) can be derived, as stated in the following proposition.

Proposition 1. *The jth entries of the solutions $\hat{\mathbf{z}}_1$ and $\hat{\mathbf{z}}_2$ to (17) at the $(n + 1)$th iteration, $n \geq 0$, can be written as given in Box I.*

Proof. The Lagrangian of the cost function in (17) is

$$\mathcal{L}(\mathbf{z}_1, \mathbf{z}_2, a) = \sum_{l=1}^{2} \Big[\|\mathbf{u}_l - \boldsymbol{\Lambda}_l \mathbf{z}_l\|_2^2$$
$$+ \lambda \|\mathbf{z}_l\|_1 \Big] + \alpha \|\mathbf{z}_1 - a\mathbf{z}_2\|_2^2 \tag{22}$$

where λ and α are the Lagrangian constants, depending on the thresholds γ_1 and γ_2. Upon differentiating (22) with respect to the jth element of $\mathbf{z}_1$ and $\mathbf{z}_2$ and equating them to zero, it is easy to obtain (18) and (19), respectively. Then, (20) follows. $\square$

Hence, in view of Proposition 1 and the structure of the cost function (13), the optimal solutions to the JC-DD problem can be readily derived by directly replacing, respectively: $\hat{\mathbf{z}}_1(n)$ and $\hat{\mathbf{z}}_2(n)$ with $\hat{\mathbf{x}}_k(n)$ and $\hat{\mathbf{x}}_{k+1}(n)$, $\mathbf{u}_1$ and $\mathbf{u}_2$ with $\mathbf{y}_k$ and $\mathbf{y}_{k+1}$, $\hat{\mathbf{u}}_1^{(j)}(n)$ and $\hat{\mathbf{u}}_2^{(j)}(n)$ with $\hat{\mathbf{y}}_k^{(j)}(n)$ and $\hat{\mathbf{y}}_{k+1}^{(j)}(n)$, and finally $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ with $\boldsymbol{\Phi}_k$ and $\boldsymbol{\Phi}_{k+1}$. To conclude, it is worth noting that the computational complexity of the JC-DD approach based on the PCD iterative algorithm for a total of P iterations results in $\mathcal{O}(PNM)$, and therefore, it is comparable with that of the SC-DD.

4. MAP detectors

In this section, MAP detectors will be derived as performance benchmarks assuming that the received signal is sampled at the Nyquist rate or at the compressed rate. Differently from [32], the channel response is Laplacian distributed so as to take into account its inherent sparsity.

4.1. Nyquist-rate MAP detector

The Nyquist-rate sampled waveform corresponding to two consecutive symbols can be written as

$$\mathbf{r} = (\mathbf{b} \otimes \mathbf{I}_N)\mathbf{h} + \mathbf{v} \tag{23}$$

where $\mathbf{r} \triangleq [\mathbf{r}_k^T, \mathbf{r}_{k+1}^T]^T$, with $\mathbf{r}_k$ being expressed by (3), $\mathbf{b} \triangleq [b_k, b_{k+1}]^T$ includes two consecutive differentially-encoded symbols, and $\mathbf{v} \triangleq [\mathbf{v}_k^T, \mathbf{v}_{k+1}^T]^T$ is the noise component. Hence, the Nyquist-rate MAP differential detector (N-MAP-DD) can be expressed as

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\arg\max} \left\{ p(\mathbf{r}|\mathbf{b})P(\mathbf{b}) \right\} \tag{24}$$

where $P(\mathbf{b})$ is the *a priori* distribution of the transmitted symbols $\mathbf{b}$. Under some assumptions, it can be proved that the N-MAP-DD (24) takes a simple form, as illustrated in the following proposition.

Proposition 2. *Assuming a uniform distribution of the transmitted symbols $\mathbf{b}$ and Laplacian distribution of the channel response $\mathbf{h}$, the N-MAP-DD coincides with the conventional Nyquist-rate DD (5)*

$$\hat{a}_{k+1}^{(\text{N-MAP-DD})} = \text{sign}\left(\mathbf{r}_{k+1}^T \mathbf{r}_k \right). \tag{25}$$

Proof. Upon representing the channel response as the product $\mathbf{h} = \rho\mathbf{n}$ between a Rayleigh random variable ρ and a joint normal random vector $\mathbf{n}$, the expression of $p(\mathbf{r}|\mathbf{b})$, as derived in Appendix A, is

$$p(\mathbf{r}|\mathbf{b}) = \int_0^{\infty} p(\mathbf{r}|\mathbf{b}, \rho)p(\rho)d\rho \tag{26}$$

where $p(\mathbf{r}|\mathbf{b}, \rho)$ is the zero-mean joint normal distribution

$$[\hat{\mathbf{z}}_1]_j(n+1) = \frac{\text{shrink}\left(\sum_{i=1}^{M}[\boldsymbol{\Lambda}_1]_{i,j}\{[\mathbf{u}_1]_i - [\hat{\mathbf{u}}_1^{(j)}(n+1)]_i\} + \alpha\hat{a}(n)[\hat{\mathbf{z}}_2]_j(n), \lambda\right)}{1+\alpha} \tag{18}$$

$$[\hat{\mathbf{z}}_2]_j(n+1) = \frac{\text{shrink}\left(\sum_{i=1}^{M}[\boldsymbol{\Lambda}_2]_{i,j}\{[\mathbf{u}_2]_i - [\hat{\mathbf{u}}_2^{(j)}(n+1)]_i\} + \alpha\hat{a}(n)[\hat{\mathbf{z}}_1]_j(n), \lambda\right)}{1+\alpha\hat{a}^2(n)} \tag{19}$$

$$\hat{a}(n+1) = \hat{\mathbf{z}}_2^T(n+1)\hat{\mathbf{z}}_1(n+1) \tag{20}$$

where

$$[\hat{\mathbf{u}}_l^{(j)}(n+1)]_i = \sum_{m<j}[\boldsymbol{\Lambda}_l]_{i,m}[\hat{\mathbf{z}}_l]_m(n+1) + \sum_{m>j}[\boldsymbol{\Lambda}_l]_{i,m}[\hat{\mathbf{z}}_l]_m(n), \quad l = 1, 2. \tag{21}$$

Box I.

$$p(\mathbf{r}|\mathbf{b}, \rho) = \frac{1}{\pi^{2N}\sigma_v^2(\sigma_v^2 + 2\rho^2)}\exp\left\{-\frac{1}{\sigma_v^2}\mathbf{r}^T\left[\mathbf{I}_{2N}\right.\right.$$
$$\left.\left.-\frac{\rho^2}{\sigma_v^2 + 2\rho^2}(\mathbf{b}\mathbf{b}^T \otimes \mathbf{I}_N)\right]\mathbf{r}\right\}. \tag{27}$$

Since **b** is assumed to be uniformly distributed, from (24) it can be argued that maximizing the product $p(\mathbf{r}|\mathbf{b})P(\mathbf{b})$ is equivalent to maximizing $p(\mathbf{r}|\mathbf{b})$ over **b**. From (26), we can say that if the maximum of $p(\mathbf{r}|\mathbf{b}, \rho)$ over **b** is independent of each value of ρ, then that is also the maximum of $p(\mathbf{r}|\mathbf{b})$. Now from (27), maximizing $p(\mathbf{r}|\mathbf{b}, \rho)$ means that for a given **r** and ρ, finding the value of **b** that maximizes

$$\Gamma_N(\mathbf{r}|\mathbf{b}, \rho) \triangleq -\frac{1}{\sigma_v^2}\mathbf{r}^T\left[\mathbf{I}_{2N} - \frac{\rho^2}{\sigma_v^2 + 2\rho^2}(\mathbf{b}\mathbf{b}^T \otimes \mathbf{I}_N)\right]\mathbf{r}. \tag{28}$$

Dropping immaterial addends independent of **b**, from (28) it can be obtained that the MAP estimate is the value of **b** maximizing the function (independent of ρ) defined as

$$\Psi_N(\mathbf{r}|\mathbf{b}) \triangleq \mathbf{r}^T(\mathbf{b}\mathbf{b}^T \otimes \mathbf{I}_N)\mathbf{r}$$
$$= b_{k+1}b_k\mathbf{r}_{k+1}^T\mathbf{r}_k. \tag{29}$$

Thus, in view of the differential encoding rule $a_{k+1} = b_{k+1}b_k$, (29) turns equivalently into

$$\Psi_N(\mathbf{r}|\mathbf{b}) = a_{k+1}\mathbf{r}_{k+1}^T\mathbf{r}_k \tag{30}$$

which provides the desired result (25). □

4.2. Compressed-rate MAP detector

The signal model for two consecutive received symbol waveforms sampled at compressed rate can be formulated as

$$\mathbf{y} = \boldsymbol{\Phi}(\mathbf{b} \otimes \mathbf{I}_N)\mathbf{h} + \boldsymbol{\Phi}\mathbf{v} \tag{31}$$

where $\mathbf{y} \triangleq [\mathbf{y}_k^T, \mathbf{y}_{k+1}^T]^T$, with $\mathbf{y}_k$ being expressed by (4), **b** and **v** are defined as in (23), and $\boldsymbol{\Phi} \triangleq \text{diag}\{\boldsymbol{\Phi}_k, \boldsymbol{\Phi}_{k+1}\}$, with $\boldsymbol{\Phi}_{k+l}$, $l = 0, 1$ being the $M \times N$ measurement matrices for which we assume $\boldsymbol{\Phi}_{k+l}\boldsymbol{\Phi}_{k+l}^T = \mathbf{I}_M$, $l = 0, 1$. The compressed-rate MAP differential detector (C-MAP-DD) is given by

$$\hat{\mathbf{b}} = \arg\max_{\mathbf{b}} \{p(\mathbf{y}|\mathbf{b})P(\mathbf{b})\} \tag{32}$$

where $P(\mathbf{b})$ is the *a priori* distribution of the transmitted symbols **b**. The structure of the C-MAP-DD scheme can be derived as illustrated in the sequel.

Proposition 3. *Assuming a uniform distribution of the transmitted symbols* **b** *and Laplacian distribution of the channel response* **h**, *the C-MAP-DD rule results approximately in*

$$\hat{a}_{k+1}^{(\text{C-MAP-DD})} = \text{sign}\left(\mathbf{y}_{k+1}^T\boldsymbol{\Phi}_{k+1}\boldsymbol{\Phi}_k^T\mathbf{y}_k\right). \tag{33}$$

Proof. Following the approach pursued in Proposition 2, in Appendix B it is shown that

$$p(\mathbf{y}|\mathbf{b}) = \int_0^\infty p(\mathbf{y}|\mathbf{b}, \rho)p(\rho)d\rho \tag{34}$$

where ρ is a Rayleigh distributed random variable and $p(\mathbf{y}|\mathbf{b}, \rho)$ is the zero-mean joint normal distribution

$$p(\mathbf{y}|\mathbf{b}, \rho) = \frac{1}{\pi^{2M}\sigma_v^{4M}|\boldsymbol{\Sigma}|}\exp\left\{-\frac{1}{\sigma_v^2}\mathbf{y}^T\left[\mathbf{I}_{2M}\right.\right.$$
$$\left.\left.-\frac{\rho^2}{\sigma_v^2}\boldsymbol{\Phi}(\mathbf{b} \otimes \mathbf{I}_N)\boldsymbol{\Sigma}^{-1}(\mathbf{b} \otimes \mathbf{I}_N)^T\boldsymbol{\Phi}^T\right]\mathbf{y}\right\} \tag{35}$$

with the $N \times N$ positive definite matrix $\boldsymbol{\Sigma}$ being defined as

$$\boldsymbol{\Sigma} \triangleq \mathbf{I}_N + \frac{\rho^2}{\sigma_v^2}\left(\boldsymbol{\Phi}_k^T\boldsymbol{\Phi}_k + \boldsymbol{\Phi}_{k+1}^T\boldsymbol{\Phi}_{k+1}\right). \tag{36}$$

Exploiting the assumption that $P(\mathbf{b})$ is independent of **b**, from (32) the result is that maximizing the product $p(\mathbf{y}|\mathbf{b})P(\mathbf{b})$ over **b** equals doing the same with $p(\mathbf{y}|\mathbf{b})$ over **b**. Now from (34), if the maximum of $p(\mathbf{y}|\mathbf{b}, \rho)$ over **b** for each value of ρ is independent of ρ then that is also the maximum of $p(\mathbf{y}|\mathbf{b})$. From (35), finding the maximum of $p(\mathbf{y}|\mathbf{b}, \rho)$ over **b** means maximizing

$$\Gamma_C(\mathbf{y}|\mathbf{b}, \rho) \triangleq -\frac{1}{\sigma_v^2}\mathbf{y}^T\left[\mathbf{I}_{2M}\right.$$
$$\left.-\frac{\rho^2}{\sigma_v^2}\boldsymbol{\Phi}(\mathbf{b} \otimes \mathbf{I}_N)\boldsymbol{\Sigma}^{-1}(\mathbf{b} \otimes \mathbf{I}_N)^T\boldsymbol{\Phi}^T\right]\mathbf{y}, \tag{37}$$

or equivalently, the function obtained after dropping immaterial addends independent of $\mathbf{b}$ as

$$\Psi_{\mathrm{C}}(\mathbf{y}|\mathbf{b}, \rho) \triangleq \mathbf{y}^T \boldsymbol{\Phi}(\mathbf{b} \otimes \mathbf{I}_N) \boldsymbol{\Sigma}^{-1}(\mathbf{b} \otimes \mathbf{I}_N)^T \boldsymbol{\Phi}^T \mathbf{y} \qquad (38)$$

which, however, is still dependent on ρ due to the presence of $\boldsymbol{\Sigma}^{-1}$. Such a matrix inverse can be computed by exploiting the eigenvalue decomposition (EVD) of the $N \times N$ semi-positive definite matrix $\boldsymbol{\Phi}_k^T \boldsymbol{\Phi}_k + \boldsymbol{\Phi}_{k+1}^T \boldsymbol{\Phi}_{k+1}$ given by $\mathbf{Q} \boldsymbol{\Omega} \mathbf{Q}^T$, with $\boldsymbol{\Omega}$ having non-negative elements along its main diagonal and $\mathbf{Q} \mathbf{Q}^T = \mathbf{I}_N$. Thus, plugging the EVD into (38) yields

$$\Psi_{\mathrm{C}}(\mathbf{y}|\mathbf{b}, \rho) \triangleq \mathbf{y}^T \boldsymbol{\Phi}(\mathbf{b} \otimes \mathbf{I}_N) \mathbf{Q} \left(\mathbf{I}_N + \frac{\rho^2}{\sigma_v^2} \boldsymbol{\Omega} \right)^{-1} \mathbf{Q}^T$$
$$\times (\mathbf{b} \otimes \mathbf{I}_N)^T \boldsymbol{\Phi}^T \mathbf{y}. \qquad (39)$$

Now, considering the fact that the diagonal matrix $\left(\mathbf{I}_N + \frac{\rho^2}{\sigma_v^2} \boldsymbol{\Omega} \right)^{-1}$ has entries which are strictly positive and less than unity, (39) can be approximated by its upper bound (independent of ρ)

$$\Upsilon_{\mathrm{C}}(\mathbf{y}|\mathbf{b}) \triangleq \mathbf{y}^T \boldsymbol{\Phi}(\mathbf{b} \otimes \mathbf{I}_N)(\mathbf{b} \otimes \mathbf{I}_N)^T \boldsymbol{\Phi}^T \mathbf{y}, \qquad (40)$$

that can be properly rearranged as

$$\Upsilon_{\mathrm{C}}(\mathbf{y}|\mathbf{b}) = b_{k+1} b_k \mathbf{y}_{k+1}^T \boldsymbol{\Phi}_{k+1} \boldsymbol{\Phi}_k^T \mathbf{y}_k. \qquad (41)$$

Thus, in view of the differential encoding rule $a_{k+1} = b_{k+1} b_k$, we end up with the desired result (33). $\quad \square$

Some remarks about the C-MAP-DD scheme can be of interest.

(1) The OLS-DD estimate of the information symbol a_{k+1} is obtained from (7) as

$$\hat{a}_{k+1}^{\mathrm{OLS\text{-}DD}} = \mathrm{sign}\left((\boldsymbol{\Phi}_{k+1}^+ \mathbf{y}_{k+1})^T (\boldsymbol{\Phi}_k^+ \mathbf{y}_k) \right) \qquad (42)$$

where $\boldsymbol{\Phi}_{k+l}^+$ is the pseudo-inverse of $\boldsymbol{\Phi}_{k+l}$, $l = 0, 1$. Since the measurement matrices have orthonormal rows, it can be shown that $\boldsymbol{\Phi}_{k+l}^+ = \boldsymbol{\Phi}_{k+l}^T$, $l = 0, 1$. Therefore, we get

$$(\boldsymbol{\Phi}_{k+1}^+ \mathbf{y}_{k+1})^T (\boldsymbol{\Phi}_k^+ \mathbf{y}_k) = \mathbf{y}_{k+1}^T \boldsymbol{\Phi}_{k+1} \boldsymbol{\Phi}_k^T \mathbf{y}_k, \qquad (43)$$

from which we argue that the OLS-DD coincides with the C-MAP-DD.

(2) Assuming $\boldsymbol{\Phi}_{k+1} = \boldsymbol{\Phi}_k$ and exploiting $\boldsymbol{\Phi}_k \boldsymbol{\Phi}_k^T = \mathbf{I}_M$, we obtain from (33)

$$\mathbf{y}_{k+1}^T \boldsymbol{\Phi}_{k+1} \boldsymbol{\Phi}_k^T \mathbf{y}_k = \mathbf{y}_{k+1}^T \mathbf{y}_k \qquad (44)$$

which means that, whenever the measurement matrices are invariant, the C-MAP-DD coincides with the DC-DD.

5. Simulation results

The detectors we discussed in the previous sections are verified here by means of numerical simulations taking as performance quality the bit error rate (BER) metric as a function of both the ratio of the mean-received bit-energy and the noise spectral density ratio defined as $E_b/N_0 \triangleq \|\mathbf{h}\|_2^2 / \sigma_v^2$, and the compression

ratio μ. The conventional DD at Nyquist-rate (NDD) is compared with the compressed DD schemes based on the approaches of the direct type DC-DD in (6), the separate type SD-DD in (10)–(12) and the joint type JD-DD solved through the iterative method outlined in Proposition 1. The performance results of the compressed MAP DD derived in Proposition 3 and the least squares DD defined by (42), labeled as C-MAP-DD and OLS-DD, respectively, are also plotted as performance benchmarks.

5.1. Simulation setup

The transmitted signal consists of differentially encoded symbols, each conveyed by an ultra short pulse traveling through a Laplacian distributed propagation channel. For the sake of simplicity, we assume that the channel response, identified as $\mathbf{h}$ in (3), includes the effects of the shaping filters at both the transmitter and receiver sides. The received symbol waveform sampled at Nyquist rate contains $N = 32$ samples, or alternatively, is compressed with a compression ratio μ, thus resulting in $M < N$ samples. The measurement matrix $\boldsymbol{\Phi}_k$ has zero-mean unit-variance i.i.d. normal entries with orthonormalized rows, and can be chosen within consecutive symbols to be the same ($\boldsymbol{\Phi}_k = \boldsymbol{\Phi}_{k+1}$) or different from each other ($\boldsymbol{\Phi}_k \neq \boldsymbol{\Phi}_{k+1}$). The methods PCD in (10)–(11) and DEN in (18)–(20) are iterated for a maximum of 200 iterations or if a tolerance level of 10^{-5} is reached.

The optimal value of the parameter λ is selected for the SC-DD case by applying a $\mathcal{K}$-fold cross validation (CV) approach [33, Chapter 17]. For a given λ, the received samples $\mathbf{y}$ are subdivided into the sequence $\mathbf{y}_m$, $1 \leq m \leq \mathcal{K}$, each including $M/\mathcal{K}$ samples. Then, $\mathbf{y}_m$ is predicted as $\hat{\mathbf{y}}_m$ using the samples obtained by removing $\mathbf{y}_m$ itself from $\mathbf{y}$. The optimal λ is thus evaluated as the value minimizing the prediction error

$$\lambda^{(\mathrm{opt})} = \arg \min_\lambda \left\{ \frac{1}{M} \sum_{m=1}^{\mathcal{K}} \left\| \mathbf{y}_m - \hat{\mathbf{y}}_m(\lambda) \right\|_2^2 \right\} \qquad (45)$$

where $\mathcal{K} = 8$ and the trial values of λ are $1, 0.1, 0.01, 0.001$. Conversely, for the JC-DD the optimal λ is chosen as

$$\lambda^{(\mathrm{opt})} = \arg \max_\lambda \left\{ \left| \hat{a}_{k+1}(\lambda) \right| \right\} \qquad (46)$$

where $\hat{a}_{k+1}$ is the DEN symbol soft estimate given by the correlation (20) of Proposition 3.

5.2. Performance comparisons

Figs. 1 and 2 quantify the BER detection performance as a function of the E_b/N_0 ratio, assuming that the measurement matrices are chosen to be the same or different from each other, respectively. While the reference NDD works at Nyquist-rate, all the other schemes adopt a compression ratio of $\mu = 0.5$ or $\mu = 0.75$. Focusing on the case $\mu = 0.5$ of Fig. 1, it can be noted that the JC-DD closely follows the C-MAP-DD, but if compared to the NDD, it degrades approximately by 1.5 dB at a BER level of 10^{-2}. Further, the DC-DD overlaps with the C-MAP-DD, according to what we observed in Remark (2) of Section 4.2, whereas the SC-DD lags behind by 1 dB. Increasing the compression ratio to $\mu = 0.75$, the JC-DD, DC-DD and SC-DD BER degradation from the reference NDD
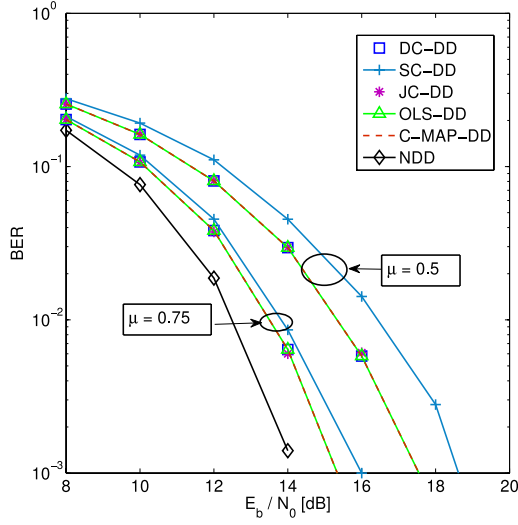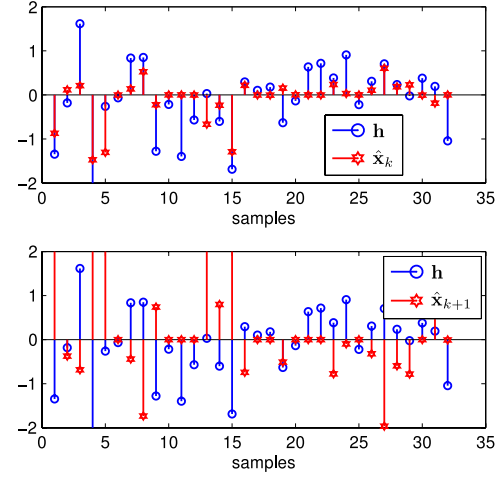
Fig. 1. BER comparison for different detection methods with $\boldsymbol{\Phi}_k = \boldsymbol{\Phi}_{k+1}$ and compression ratio $\mu = 0.5, 0.75$.
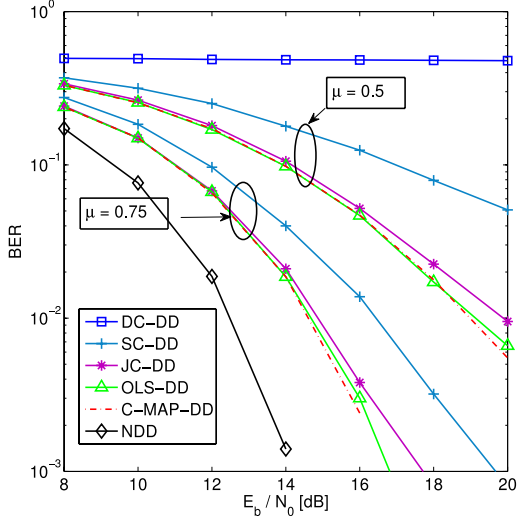
Fig. 2. BER comparison for different detection methods with $\boldsymbol{\Phi}_k \neq \boldsymbol{\Phi}_{k+1}$ and compression ratio $\mu = 0.5, 0.75$.

Fig. 3. Reconstruction results of JC-DD at $10\log_{10}(E_b/N_0) = 20$ dB.
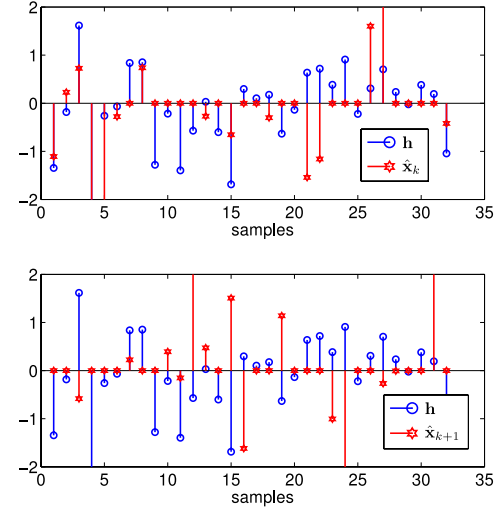
Fig. 4. Reconstruction results of SC-DD at $10\log_{10}(E_b/N_0) = 20$ dB.

reduces to around 1 dB, 1 dB and 1.3 dB, respectively. Thus, we show that the above compressed detectors can trade off performance against complexity in terms of compression ratio.

The results of Fig. 2 confirm that: (i) the scenario with different measurement matrices is more demanding than the one when they are the same as illustrated in Fig. 1, and (ii) an increase of the compression ratio to $\mu = 0.75$ alleviates the performance gap from the conventional NDD at the price of increasing the complexity. For $\mu = 0.5$ and a BER level of 10^{-2}, the JC-DD has a gap of approximately 1 dB from the C-MAP-DD and 7.5 dB from the NDD, but shows a considerable edge over the separate approach SC-DD. Differently from Fig. 1, the direct scheme

DC-DD completely misses detection and so turns out to be useless. This result proves that a reconstruction step, separate or better joint with detection, is clearly mandatory.

Further, from both Figs. 1 and 2 it is apparent that the performance of the OLS-DD equals that offered by the C-MAP-DD, as expected from Remark (1) of Section 4.2. Nevertheless, we remark that the OLS-DD is considerably outperformed by the JC-DD and SC-DD in terms of sparse signal reconstruction, due to the intrinsic lack of inducing sparsity on the solutions exhibited by the least squares method.

Figs. 3 and 4 give a quantitative picture about the reconstruction performance of the JC-DD and SC-DD, respectively. We assume that the received waveforms without noise are $\mathbf{x}_k = \mathbf{h}$ and $\mathbf{x}_{k+1} = -\mathbf{h}$, $E_b/N_0 = 20$ dB, and λ is set to 0.1. In each figure, the upper part shows $\mathbf{h}$ and the reconstructed signal component $\hat{\mathbf{x}}_k$,
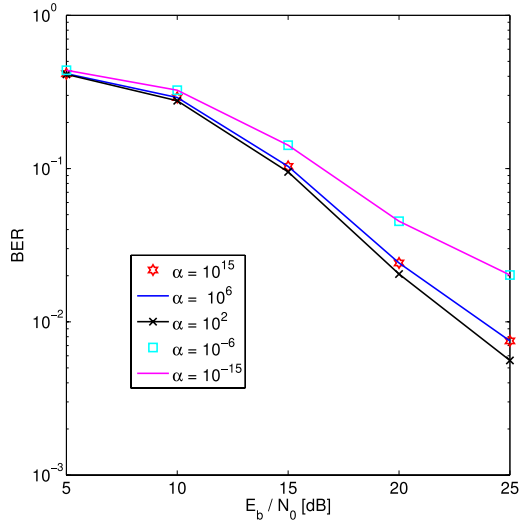
Fig. 5. BER comparison of JC-DD for different weight coefficients α with $\boldsymbol{\Phi}_k \neq \boldsymbol{\Phi}_{k+1}$.

Fig. 6. BER comparison of JC-DD and NDD for different compression ratios μ with $\boldsymbol{\Phi}_k \neq \boldsymbol{\Phi}_{k+1}$.

whereas the lower part does the same for the adjacent symbol, namely $\mathbf{h}$ and $\hat{\mathbf{x}}_{k+1}$. For a given realization of $\mathbf{h}$, we obtain that out of $2N = 64$ signal samples for both symbols, the JC-DD forces $24/64 \approx 37\%$ components to zero and correctly reconstructs $31/64 \approx 48\%$ non-zero components, whereas the above percentages for the SC-DD turn into $36/64 \approx 56\%$ and $19/64 \approx 30\%$, respectively. These results make us argue that the SC-DD has a higher tendency of setting signal components to zero, whereas the JC-DD exploits its inherent fusion capabilities between the two sets of variables, leading not only to joint sparsity but also to a fair amount of reconstructed non-zero components. The different behavior plays a role in taking a correct decision based on correlation, and justifies the detection performance superiority of the joint approach over considering detection as a separate step from reconstruction.

Fig. 5 shows the sensitivity of the JC-DD scheme to the choice of the coefficient α which weighs the differential squared error in (13). It is apparent that better results over E_b/N_0 are obtained for values in the range of around $\alpha = 100$. Finally, the effect of the compression ratio μ over the BER of the JC-DD is evaluated in Fig. 6 for an E_b/N_0 of 12 and 14 dB. As expected, it can be shown that the smaller the μ the worse the BER level, and when μ approaches 1, the performance of the JC-DD tends to that of the NDD.

6. Conclusions

In this paper, the compressive sampling framework has been applied to differentially encoded UWB signals. A joint reconstruction and detection method for the compressed symbol waveforms has been presented, which has been shown to outperform the simpler method based on a separate approach. Direct detection without reconstruction has been evaluated as well, whereas a compressed MAP differential detector has been derived to
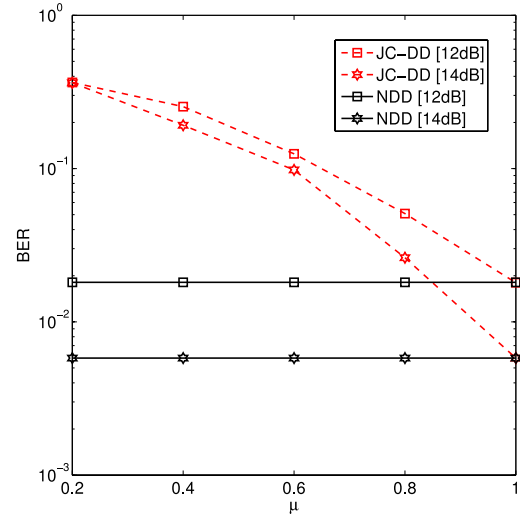
have a performance benchmark for the proposed detectors. Simulation results confirm that the major advantages we gain are (i) the reduced sampling rate, (ii) the ability to carry out the differential detection process in the digital domain, and (iii) the option of a competitive performance in different scenarios where the measurement matrices are the same as well as different.

Appendix A. PDF of the Nyquist-rate sampled received signal

In order to evaluate the pdf of the received signal sampled at Nyquist rate, let us start from the result that a Laplacian distributed random variable can be represented as the product between a Rayleigh and a normal one [34]. The extension to the multivariate case allows us to write the channel response as $\mathbf{h} = \rho\mathbf{n}$, where the pdf of ρ is Rayleigh, i.e., $p(\rho) = \rho e^{-\rho^2/2}$, and $\mathbf{n}$ is a zero-mean joint normal random vector with covariance matrix $\mathbf{C_n} = \mathbf{I}_N$. Thus, the signal model for the Nyquist-rate sampled waveform corresponding to two consecutive received symbols can be formulated as

$$\mathbf{r} = (\mathbf{b} \otimes \mathbf{I}_N)\rho\mathbf{n} + \mathbf{v} \tag{47}$$

where $\mathbf{v}$ is the joint normal noise component with zero mean and covariance matrix $\mathbf{C_v} = \sigma_v^2 \mathbf{I}_{2N}$, statistically independent of both ρ and $\mathbf{n}$. From (47), it is apparent that the pdf of $\mathbf{r}$ given $\mathbf{b}$ is expressed by

$$p(\mathbf{r}|\mathbf{b}) = \int_0^\infty p(\mathbf{r}|\mathbf{b}, \rho)p(\rho)d\rho \tag{48}$$

where $p(\mathbf{r}|\mathbf{b}, \rho)$ is the zero-mean joint normal distribution

$$p(\mathbf{r}|\mathbf{b}, \rho) = \frac{1}{\pi^{2N} \left| \mathbf{C}_{\mathbf{r}|\mathbf{b}, \rho} \right|} e^{-\mathbf{r}^T \mathbf{C}_{\mathbf{r}|\mathbf{b}, \rho}^{-1} \mathbf{r}}, \tag{49}$$

with covariance matrix

$$
\mathbf{C}_{\mathbf{r}|\mathbf{b},\rho} = \mathrm{E}\left\{\left[(\mathbf{b}\otimes\mathbf{I}_N)\rho\mathbf{n}+\mathbf{v}\right]\left[(\mathbf{b}\otimes\mathbf{I}_N)\rho\mathbf{n}+\mathbf{v}\right]^T\right\}
$$

$$
= \sigma_v^2\mathbf{I}_{2N}+\rho^2(\mathbf{b}\mathbf{b}^T\otimes\mathbf{I}_N). \tag{50}
$$

From the binomial inverse theorem, it can be obtained

$$
\mathbf{C}_{\mathbf{r}|\mathbf{b},\rho}^{-1} = \left[\sigma_v^2\mathbf{I}_{2N}+\rho^2(\mathbf{b}\mathbf{b}^T\otimes\mathbf{I}_N)\right]^{-1}
$$

$$
= \left[\sigma_v^2\mathbf{I}_{2N}+\rho^2(\mathbf{b}\otimes\mathbf{I}_N)(\mathbf{b}\otimes\mathbf{I}_N)^T\right]^{-1}
$$

$$
= \frac{1}{\sigma_v^2}\Bigg\{\mathbf{I}_{2N}-\frac{\rho^2}{\sigma_v^2}(\mathbf{b}\otimes\mathbf{I}_N)\Bigg[\mathbf{I}_N
$$

$$
+\frac{\rho^2}{\sigma_v^2}(\mathbf{b}\otimes\mathbf{I}_N)^T(\mathbf{b}\otimes\mathbf{I}_N)\Bigg]^{-1}(\mathbf{b}\otimes\mathbf{I}_N)^T\Bigg\}. \tag{51}
$$

By exploiting the result

$$
(\mathbf{b}\otimes\mathbf{I}_N)^T(\mathbf{b}\otimes\mathbf{I}_N) = 2\mathbf{I}_N, \tag{52}
$$

(51) can be simplified into

$$
\mathbf{C}_{\mathbf{r}|\mathbf{b},\rho}^{-1} = \frac{1}{\sigma_v^2}\left[\mathbf{I}_{2N}-\frac{\rho^2}{\sigma_v^2+2\rho^2}(\mathbf{b}\mathbf{b}^T\otimes\mathbf{I}_N)\right]. \tag{53}
$$

Concerning the determinant of $\mathbf{C}_{\mathbf{r}|\mathbf{b},\rho}$, applying the Sylvester theorem yields

$$
\left|\mathbf{C}_{\mathbf{r}|\mathbf{b},\rho}\right| = \left|\sigma_v^2\mathbf{I}_{2N}+\rho^2(\mathbf{b}\mathbf{b}^T\otimes\mathbf{I}_N)\right|
$$

$$
= \sigma_v^{4N}\left|\mathbf{I}_N+\frac{\rho^2}{\sigma_v^2}(\mathbf{b}\otimes\mathbf{I}_N)^T(\mathbf{b}\otimes\mathbf{I}_N)\right|
$$

$$
= \sigma_v^{4N}\left|\mathbf{I}_N+\frac{2\rho^2}{\sigma_v^2}\mathbf{I}_N\right|
$$

$$
= \left(\sigma_v^4+2\sigma_v^2\rho^2\right)^N \tag{54}
$$

which turns out to be independent of $\mathbf{b}$.

Appendix B. PDF of the compressed-rate sampled received signal

Following the approach of Appendix A, let us consider the signal model corresponding to two consecutive received symbols sampled at compressed rate

$$
\mathbf{y} = \boldsymbol{\Phi}(\mathbf{b}\otimes\mathbf{I}_N)\rho\mathbf{n}+\boldsymbol{\Phi}\mathbf{v} \tag{55}
$$

where ρ, $\mathbf{n}$ and $\mathbf{v}$ are defined as in (47), and $\boldsymbol{\Phi}$ as in (31). According to (55), the pdf of $\mathbf{y}$ given $\mathbf{b}$ can be written as

$$
p(\mathbf{y}|\mathbf{b}) = \int_0^\infty p(\mathbf{y}|\mathbf{b},\rho)p(\rho)d\rho \tag{56}
$$

where $p(\mathbf{y}|\mathbf{b},\rho)$ is the zero-mean joint normal distribution

$$
p(\mathbf{y}|\mathbf{b},\rho) = \frac{1}{\pi^{2M}\left|\mathbf{C}_{\mathbf{y}|\mathbf{b},\rho}\right|}e^{-\mathbf{y}^T\mathbf{C}_{\mathbf{y}|\mathbf{b},\rho}^{-1}\mathbf{y}}. \tag{57}
$$

Taking into account that $\boldsymbol{\Phi}\boldsymbol{\Phi}^T=\mathbf{I}_{2M}$, the covariance matrix $\mathbf{C}_{\mathbf{y}|\mathbf{b},\rho}$ results in

$$
\mathbf{C}_{\mathbf{y}|\mathbf{b},\rho} = \mathrm{E}\left\{\left[\boldsymbol{\Phi}(\mathbf{b}\otimes\mathbf{I}_N)\rho\mathbf{n}+\boldsymbol{\Phi}\mathbf{v}\right]\left[\boldsymbol{\Phi}(\mathbf{b}\otimes\mathbf{I}_N)\rho\mathbf{n}+\boldsymbol{\Phi}\mathbf{v}\right]^T\right\}
$$

$$
= \sigma_v^2\mathbf{I}_{2M}+\rho^2\boldsymbol{\Phi}(\mathbf{b}\mathbf{b}^T\otimes\mathbf{I}_N)\boldsymbol{\Phi}^T \tag{58}
$$

whereas its inverse can be computed from the binomial inverse theorem as

$$
\mathbf{C}_{\mathbf{y}|\mathbf{b},\rho}^{-1} = \left[\sigma_v^2\mathbf{I}_{2M}+\rho^2\boldsymbol{\Phi}(\mathbf{b}\mathbf{b}^T\otimes\mathbf{I}_N)\boldsymbol{\Phi}^T\right]^{-1}
$$

$$
= \frac{1}{\sigma_v^2}\left[\mathbf{I}_{2M}-\frac{\rho^2}{\sigma_v^2}\boldsymbol{\Phi}(\mathbf{b}\otimes\mathbf{I}_N)\boldsymbol{\Sigma}^{-1}(\mathbf{b}\otimes\mathbf{I}_N)^T\boldsymbol{\Phi}^T\right] \tag{59}
$$

where

$$
\boldsymbol{\Sigma} \triangleq \mathbf{I}_N+\frac{\rho^2}{\sigma_v^2}(\mathbf{b}\otimes\mathbf{I}_N)^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}(\mathbf{b}\otimes\mathbf{I}_N)
$$

$$
= \mathbf{I}_N+\frac{\rho^2}{\sigma_v^2}\left(\boldsymbol{\Phi}_k^T\boldsymbol{\Phi}_k+\boldsymbol{\Phi}_{k+1}^T\boldsymbol{\Phi}_{k+1}\right). \tag{60}
$$

Finally, using the Sylvester theorem, the determinant of $\mathbf{C}_{\mathbf{y}|\mathbf{b},\rho}$ is given by

$$
\left|\mathbf{C}_{\mathbf{y}|\mathbf{b},\rho}\right| = \left|\sigma_v^2\mathbf{I}_{2M}+\rho^2\boldsymbol{\Phi}(\mathbf{b}\mathbf{b}^T\otimes\mathbf{I}_N)\boldsymbol{\Phi}^T\right|
$$

$$
= \sigma_v^{4M}\left|\mathbf{I}_{2M}+\frac{\rho^2}{\sigma_v^2}\boldsymbol{\Phi}(\mathbf{b}\otimes\mathbf{I}_N)(\mathbf{b}\otimes\mathbf{I}_N)^T\boldsymbol{\Phi}^T\right|
$$

$$
= \sigma_v^{4M}\left|\mathbf{I}_N+\frac{\rho^2}{\sigma_v^2}(\mathbf{b}\otimes\mathbf{I}_N)^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}(\mathbf{b}\otimes\mathbf{I}_N)\right| \tag{61}
$$

or equivalently from (60),

$$
\left|\mathbf{C}_{\mathbf{y}|\mathbf{b},\rho}\right| = \sigma_v^{4M}\left|\mathbf{I}_N+\frac{\rho^2}{\sigma_v^2}\left(\boldsymbol{\Phi}_k^T\boldsymbol{\Phi}_k+\boldsymbol{\Phi}_{k+1}^T\boldsymbol{\Phi}_{k+1}\right)\right| \tag{62}
$$

which is independent of $\mathbf{b}$.

References

[1] M.Z. Win, R.A. Scholtz, Impulse radio: how it works, IEEE Communications Letters 2 (2) (1998) 36–38.
[2] M. Ghavami, L.B. Michael, R. Kohno, Ultra Wideband: Signals and Systems in Communication Engineering, 2nd ed., John Wiley and Sons, West Sussex, England, 2007.
[3] A.F. Molisch, J.R. Foerster, M. Pendergrass, Channel models for ultra-wideband personal area networks, IEEE Wireless Communications 10 (6) (2003) 14–21.
[4] J.D. Choi, W.E. Stark, Performance of ultra-wideband communications with suboptimal receivers in multipath channels, IEEE Journal on Selected Areas in Communications 20 (9) (2002) 1754–1766.
[5] V. Lottici, A.N. D'Andrea, U. Mengali, Channel estimation for ultra-wideband communications, IEEE Journal on Selected Areas in Communications 20 (2002) 1638–1645.
[6] K. Witrisal, G. Leus, G. Janssen, M. Pausini, F. Troesch, T. Zasowski, J. Romme, Noncoherent ultra-wideband systems, IEEE Signal Processing Magazine 26 (4) (2009) 48–66.
[7] R. Hoctor, H. Tomlinson, Delay-hopped transmitted-reference RF communications, in: Proc. IEEE UWBST2002, 2002, pp. 265–269.
[8] Y.L. Chao, R. Scholtz, Optimal and suboptimal receivers for ultra-wideband transmitted reference systems, in: Proc. IEEE Globecom2003, December 2003.
[9] M. Ho, S. Somayazulu, J. Foerster, S. Roy, A differential detector for UWB communications system, in: IEEE VTC2002, vol. 4, May 2002, pp. 1896–1900.

[10] N. Guo, R.C. Qiu, Improved autocorrelation demodulation receivers based on multiple-symbol detection for UWB communications, IEEE Transactions on Wireless Communications 5 (8) (2006).

[11] V. Lottici, Z. Tian, Multiple symbol differential detection for UWB communications, IEEE Transactions on Wireless Communications 7 (5) (2008) 1656–1666.

[12] V. Lottici, Z. Tian, G. Leus, A novel approach to UWB data detection with symbol-level synchronization, Elsevier Physical Communication 2 (2009) 296–305.

[13] David L. Donoho, Compressed sensing, IEEE Transactions on Information Theory 52 (4) (2006).

[14] E. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Transactions on Information Theory 52 (2) (2006) 489–509.

[15] S. Kirolos, J. Laska, M. Wakin, M. Duarte, D. Baron, T. Ragheb, Y. Massoud, R. Baraniuk, Analog-to-information conversion via random demodulation, in: Proc. IEEE Dallas Circuits and Systems Workshop, DCAS, 2006.

[16] T. Ragheb, J. Laska, H. Nejati, S. Kirolos, R. Baraniuk, Y. Massoud, A prototype hardware for random demodulation based compressive analog-to-digital conversion, in: Proc. MWSCAS 2008, August 2008, pp. 37–40.

[17] M. Vetterli, P. Marziliano, T. Blu, Sampling signals with finite rate of innovation, IEEE Transactions on Signal Processing 50 (6) (2002) 1417–1428.

[18] Y. Vanderperren, W. Dehaene, G. Leus, Performance analysis of a flexible subsampling receiver for pulsed UWB signals, IEEE Transactions on Wireless Communications (2009).

[19] A. Oka, L. Lampe, A compressed sensing receiver for UWB impulse radio in bursty applications like wireless sensor networks, Physical Communication 2 (4) (2009) 248–264.

[20] Z. Wang, G.R. Arce, J.L. Paredes, B.M. Sadler, Compressed detection for ultra-wideband impulse radio, in: IEEE SPAWC, 2007.

[21] S. Gishkori, G. Leus, V. Lottici, Compressive sampling based differential detection of ultra wideband signals, in: Proc. IEEE PIMRC 2010, September 2010.

[22] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, Journal of the Royal Statistical Society: Series B 67 (Part 2) (2005) 301–320.

[23] ET Docket 98–153, FCC, Revision of part 15 of the commissions rules regarding ultra-wideband transmission systems, Technical Report, 2002.

[24] S. Gishkori, G. Leus, H. Deliç, Energy detectors for sparse signals, in: Proc. IEEE SPAWC 2010, June 2010.

[25] R. Tibshirani, Regression shrinkage and selection via the lasso, Technical Report, University of Toronto, 1994.

[26] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, Journal of the Royal Statistical Society, Series B 58 (1) (1996) 267–288.

[27] M. Osborne, B. Presnell, B. Turlach, On the lasso and its dual, Journal of Computational and Graphical Statistics 9 (2000) 319–337.

[28] J. Friedman, T. Hastie, H. Höfling, R. Tibshirani, Pathwise coordinate optimization, The Annals of Applied Statistics 1 (2) (2007) 302–332.

[29] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, Journal of Statistical Software 33 (1) (2010) 1–22.

[30] R. Tibshirani, S. Rosset, J. Zhu, K. Knight, Sparsity and smoothness via the fused lasso, Journal of the Royal Statistical Society: Series B 67 (2005) 91–108.

[31] P. Tseng, Convergence of a block coordinate descent method for nondifferentiable minimization, Journal of Optimization Theory and Applications 109 (3) (2001) 475–494.

[32] P. Stoica, J. Liu, J. Li, M.A. Prasad, The heuristic, GLRT, and MAP detectors for double differential modulation are identical, IEEE Transactions on Information Theory 51 (5) (2005).

[33] B. Efron, R. Tibshirani, An Introduction to the Bootstrap, Chapman and Hall Inc., 1993.

[34] S. Kotz, T.J. Kozubowski, K. Podgrski, The Laplace Distribution and Generalizations, Birkhauser, Boston, 2001.

Shahzad Gishkori received his B.Sc. degree in electrical engineering, in 2002, from the University of Engineering and Technology Lahore, Pakistan and later on, worked in the industry for almost five years. In August 2009, he received his M.Sc. degree (cum Laude) in electrical engineering from the Delft University of Technology, The Netherlands. Since November 2009, he has been with the circuits and systems group at the Faculty of Electrical Engineering, Mathematics and Computer Science of the Delft University of Technology, The Netherlands, in pursuance of his Ph.D. degree. His research interests include compressive sampling (compressed sensing), signal processing for communications and wireless communications.

Geert Leus was born in Leuven, Belgium, in 1973. He received the electrical engineering degree and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Belgium, in June 1996 and May 2000, respectively. He has been a Research Assistant and a Postdoctoral Fellow of the Fund for Scientific Research—Flanders, Belgium, from October 1996 till September 2003. During that period, Geert Leus was affiliated with the Electrical Engineering Department of the Katholieke Universiteit Leuven, Belgium. Currently, Geert Leus is an Associate Professor at the Faculty of Electrical Engineering, Mathematics and Computer Science of the Delft University of Technology, The Netherlands. During the summer of 1998, he visited Stanford University, and from March 2001 till May 2002 he was a Visiting Researcher and Lecturer at the University of Minnesota. His research interests are in the area of signal processing for communications. Geert Leus received a 2002 IEEE Signal Processing Society Young Author Best Paper Award and a 2005 IEEE Signal Processing Society Best Paper Award. He was the Chair of the IEEE Signal Processing for Communications and Networking Technical Committee, and an Associate Editor for the IEEE Transactions on Signal Processing, the IEEE Transactions on Wireless Communications, and the IEEE Signal Processing Letters. Currently, he serves on the Editorial Board of the EURASIP Journal on Applied Signal Processing.

Vincenzo Lottici received the Dr. Ing. degree (cum Laude) and the BTA in electrical engineering both from the University of Pisa in 1985 and 1986, respectively. From 1987 to 1993 he was engaged in the research of sonar digital signal processing algorithms. Since 1993 he has been with the Department of Information Engineering of the University of Pisa, where he is currently an Assistant Professor in Communication Systems. He participated in several international and national research projects, and specifically, from 2005 he has been taking an active role in the standardization activity promoted by ETSI (European Telecommunications Standards Institute) of the physical layer of the mobile radio system TETRA release 2. He frequently served as technical program committee (TPC) member for major IEEE conferences in wireless communications and signal processing, such as WCNC2012, ICUWB2011, ICC2011, WCNC2011, ICASSP2010, PIMRC2010, WCNC2010, CIP2010, Globecom2009, SPAWC2009, EUSIPCO2006, Globecom2006. Dr. Lottici received the Best Paper Award in 2006 for the work "A Theoretical Framework for Soft Information Based Synchronization in Iterative (Turbo) Receivers", EURASIP Journal on Wireless Communications and Networking, April 2005, by the EU-funded project Network of Excellence in Wireless Communications (NEWCOM). His research interests include the broad area of signal processing for communications, with emphasis on synchronization, dynamic resource allocation, cognitive ratio and compressive sensing.