# Impulse Noise Estimation and Removal for OFDM Systems

Tareq Y. Al-Naffouri, *Member, IEEE,* Ahmed A. Quadeer, and Giuseppe Caire, *Fellow, IEEE*

***Abstract*—Orthogonal Frequency Division Multiplexing (OFDM) is a modulation scheme widely used in wired systems, including Digital Subscriber Lines (DSL), Powerline Communications (PLC), and wireless standards (e.g., IEEE 802.11a/g/n/ac, WiMax (IEEE 802.16) and 3GPP LTE). While OFDM is ideally suited to deal with frequency selective channels and AWGN, its performance may be dramatically impacted by the presence of impulse noise. In fact, very strong noise impulses in the time domain might result in the erasure of whole OFDM blocks of symbols at the receiver. Impulse noise can be mitigated by considering it as a sparse vector in time, and using recently developed algorithms for sparse signal reconstruction. We propose an algorithm that utilizes the guard band null subcarriers for the impulse noise estimation and cancellation. Instead of relying on $\ell_1$ minimization as done in some popular general-purpose compressive sensing schemes, the proposed method jointly exploits the specific structure of this problem and the available a priori information for sparse signal recovery. The computational complexity of the proposed algorithm is very competitive with respect to sparse signal reconstruction schemes based on $\ell_1$ minimization. The proposed method is compared with respect to other state-of-the art methods in terms of achievable rates for an OFDM system with impulse noise and AWGN.**

***Index Terms*—OFDM, discrete multitone, sparse signal reconstruction, impulse noise, estimation, compressive sensing.**

## I. Introduction

Orthogonal Frequency Division Multiplexing (OFDM), also referred to as Discrete Multi-Tone (DMT), is a modulation scheme widely used in wired systems, including Digital Subscriber Lines (DSL) [2], [3], Powerline Communications (PLC) [4], [5], and wireless standards (e.g., IEEE 802.11a/g/n/ac, WiMax (IEEE 802.16) and 3GPP LTE) [6]. Focusing on the single-link aspect of such systems, the physical layer must cope with linear distortion due to inter-symbol-interference (ISI), additive white Gaussian noise (AWGN) and impulse noise, which may be generated by spurious sources such as switching of electrical AC devices. While OFDM is ideally suited to handle ISI by frequency domain transmission and equalization using the Inverse Discrete Fourier Transform (IDFT)/Discrete Fourier Transform (DFT) and cyclic prefix approach, and the effect of AWGN is eliminated using an appropriate level of coded modulation [2], impulse noise remains as an important limiting factor. In this work we focus on a scheme for impulse noise estimation and cancellation at the receiver.[1]

The work of T. Y. Al-Naffouri and A. A. Quadeer was supported by SABIC through an internally funded project from DSR, KFUPM (Project No. SB101006). The work of T. Y. Al-Naffouri was also partially supported by the Fulbright Scholar Program. The work of G. Caire was partially supported by NSF Grant CCF 0729162. Part of this work has been presented in the IEEE International Symposium on Information Theory, Russia, 2011 [1].

T. Y. Al-Naffouri is with the Department of Electrical Engineering, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia, and also with the Department of Electrical Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia (e-mail: tareq.alnaffouri@kaust.edu.sa).

A. A. Quadeer is with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (e-mail: aaquadeer@ust.hk).

G. Caire is with the Ming-Hsieh Department of Electrical Engineering, University of Southern California, CA 90089 USA (e-mail: caire@usc.edu).

Manuscript received April 4, 2013; revised October 7, 2013; accepted January 5, 2014. The editor coordinating the review of this paper and approving it for publication was Stefano Galli.

### A. Impulse Noise Types, Models and Approaches for its Removal

Impulse noise can be broadly divided into two types. *Aperiodic impulse noise* (also commonly known as *asynchronous impulse noise* in the context of PLC systems) is characterized by impulses occurring at random times, with short duration and high power (as high as 50 dB above the background noise level [7]). In contrast, *periodic impulse noise* consists of impulses of longer duration and occurring periodically in time. In this paper, we focus on aperiodic impulse noise. It is worthwhile to point out, though, that periodic impulse noise or block-sparse impulse noise (i.e., bursty impulse noise [8]) can be converted into the model treated in this paper by using time domain interleaving of the OFDM samples *after* the modulator IDFT [9]. This technique, referred to as TDI-OFDM [10], is analogous to what is done in coding in order to convert a bursty channel into a random error channel. Clearly, if the burstiness or the periodicity of the impulse noise is explicitly taken into account by the receiver, better performance can be achieved. However, if the goal is to design a general purpose robust algorithm that works for all impulse noise statistics provided that the average number of impulses per OFDM symbol is not larger than some target threshold, it is meaningful to focus on aperiodic random impulses of short duration, and use

[1]In general, all mentioned systems are multiuser networks where multiuser interference and multi-access coordination represent the major problems to be addressed in a comprehensive system design. Nevertheless, existing multi-access protocols (e.g., OFDMA/TDMA in cellular or CSMA/CA in WiFi) and signal processing schemes (e.g., for cancellation of cross-talk in DSL bundles) have been developed and standardized in order to reduce such multiuser networks to a set of non-interfering (or weakly interfering) links. This paper focuses on a single link and addresses impulse noise from external sources of disturbance, not taken into account explicitly by existing standardized protocols. Of course, in a comprehensive system design one should evaluate the proposed single-link methods in the context of the whole network, including resource sharing and multi-access interference. However, including these aspects here would be completely out of the scope of this study, which focuses exclusively on impulse noise estimation and cancellation techniques.

interleaving to convert any a priori unknown impulse statistics to this one, which can be handled by the receiver.

There are different statistical models present in literature for modeling aperiodic impulse noise in different applications. The three widely used models are the Gaussian mixture [12], the Middleton's Class A [13], and the symmetric alpha-stable [14] models. In this paper, we focus on the Gaussian mixture model and assume the impulse noise to be Bernoulli-Gaussian [15]-[17]. Specifically, we model impulse noise as a Gaussian process modulated by very narrow and randomly placed squared pulses, of duration approximately equal to the inverse of the signal bandwidth, i.e., the time domain sampling interval of OFDM. Thus, in the absence of estimation and cancellation, these impulses in the time domain corrupt all the subcarriers of an OFDM block, yielding a burst of very noisy frequency domain symbols that may significantly degrade the performance of the coded modulation scheme, which is targeted to AWGN and to the nominal channel SNR.

The effect of impulse noise on OFDM as well as communication engineering solutions based on impulse estimation and cancellation has been widely studied and represents an area of active research (see for example [16], [18], [19]). For the conventional concatenated error control coding based on inner convolutional and outer Reed-Solomon codes, a time-frequency interleaving is used in order to avoid that blocks of noisy symbols cause the Viterbi decoder to introduce long bursts of decoding errors, which may be beyond the correction capacity of the outer Reed-Solomon code. Alternatively, some techniques try to detect the presence of impulses and possibly their location and use that to enhance the performance. A popular method consists of detecting the presence of an impulse using some thresholding scheme, and erasing the whole OFDM block in order not to exceed the error correction capability of the channel coding ([20] and the references therein). Such schemes require the use of channel codes able to handle errors and erasures [21]. When the physical layer is not able to deal with erasures through forward error correction, it tags the uncertain OFDM symbols and let them to be handled by higher protocol payers. Some recent DSL standard contributions have proposed retransmissions as a way to deal with these cases [22]-[24]. However, it is apparent that handling efficiently the presence of impulse noise at the receiver, instead of requesting retransmissions, can be very desirable from the viewpoint of spectral efficiency, delay and simplicity of protocol operations.

Precoding techniques and frequency domain algebraic interpolation techniques inspired by Reed-Solomon coding and decoding over the complex numbers were proposed in [25], [26] to cope with the very same problem. The drawbacks of such techniques are that: 1) they require a certain structure of the null frequencies or pilots; 2) they are very sensitive to background noise and rounding/quantization errors due to finite precision arithmetic (in fact, these schemes require a number of non-trivial intermediate steps to ensure that the impulse noise decoding algorithm does not malfunction in the presence of AWGN and/or rounding errors [25], [26]).

In [28], [29], impulse noise is modeled as a sparse signal in the time domain and compressive sensing (CS) based on convex relaxation methods using $\ell_1$ minimization [33] is used for estimation from a small subset of frequency domain observations. The drawbacks of using this method are: 1) $\ell_1$ minimization requires high complexity (polynomial average complexity in the problem dimension); 2) It does not make use of any a priori statistical information (apart from the sparsity information); 3) It does not exploit the structure of the sensing matrix. This method has been extended for detecting bursty impulse noise using block CS in [34]. Recently several low complex alternatives have been proposed for sparse signal recovery, including algorithms based on belief propagation [35], Bayesian methods applied to CS [36], and iterative greedy approaches such as orthogonal matching pursuit (OMP) [37], [38], and fast Bayesian matching pursuit (FBMP) [39]. In this paper, similar to [28], [29], we also make use of the free guard band subcarriers present by default in any OFDM system to estimate and cancel impulse noise. This work differs from [28], [29] as instead of employing CS based on convex relaxation methods to estimate impulse noise, we make a collective use of the a priori statistical and sparsity information together with the structure of the problem to obtain nearly optimal estimates at low complexity.

### *B. Our Approach: Utilizing Null carriers for Impulse Noise Detection and Cancelation*

In practical OFDM systems, several subcarriers are not used to send modulation symbols. For example, in IEEE 802.11 the subcarriers at the edges of the channel band are not used in order to avoid to spill inter-channel interference to adjacent channels [30]. Also, in DSL the channel attenuation near the lower and upper edges of the spectrum is typically very strong, such that the bit-loading algorithm, reminiscent of the information-theoretic "waterfilling" power allocation ([31], [32]), allocates zero rate and zero power to these side subcarriers.

We shall exploit these unused/null frequencies in order to obtain a signal-free subspace onto which the impulse noise can be projected. Although this projection is rank-deficient (dimension of the null subcarrier subspace is less than the dimension of the total space), exploiting the fact that the impulse noise is sparse in the time domain, we can still jointly estimate the locations of the impulses and their amplitudes and eventually subtract these estimates from the received signal. Estimation is obtained with a new low complexity scheme, by exploiting the structure of the projection matrix which, in our case, is a submatrix of a unitary DFT matrix obtained by extracting a block of adjacent subcarriers. Furthermore, we also exploit the a priori probability distribution of the impulse noise, which is assumed to be Bernoulli Gaussian with known parameters.

## II. TRANSMISSION MODEL

The discrete-time complex baseband equivalent channel model for the OFDM signal under consideration in this work can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e} + \mathbf{z}, \tag{1}$$

where $\mathbf{y} \in \mathbb{C}^n$ and $\mathbf{x} \in \mathbb{C}^n$ are the time domain OFDM receive and transmit signal blocks (after cyclic prefix removal [11], $\mathbf{H}$ is an $n \times n$ circulant matrix induced by cyclic prefix precoding, with first column $\mathbf{h}$ is formed by the zero-padded channel impulse response,[2] $\mathbf{z}$ is the complex white Gaussian additive noise $\mathcal{CN}(\mathbf{0}, N_0\mathbf{I})$, and $\mathbf{e}$ is the impulse noise vector with i.i.d. Bernoulli-Gaussian entries, i.e., the components $e_i$ are statistically independent random variables distributed as a Gaussian mixture whose distribution has a mass at $0$ with probability $1-p$ and a Gaussian pdf $\mathcal{CN}(0, I_0)$ with probability $p$. Thus, the support (set of non-zero components) $\mathcal{I}$ has Binomially distributed cardinality with mean $pn$. We define the channel SNR as $\mathcal{E}_x/N_0$ and the impulse to noise ratio (INR) as $I_0/N_0$. Letting $\mathbf{F}$ denote the $n \times n$ unitary DFT matrix with $(k, \ell)$ element $[\mathbf{F}]_{k,\ell} = \frac{1}{\sqrt{n}} e^{-j2\pi k\ell/n}$ with $k, \ell \in \{0, \ldots, n-1\}$, the time domain signal is related to the frequency domain signal by $\mathbf{x} = \mathbf{F}^{\mathsf{H}}\check{\mathbf{x}}$. The circulant convolution matrix $\mathbf{H}$ can be decomposed as $\mathbf{H} = \mathbf{F}^{\mathsf{H}}\mathbf{D}\mathbf{F}$, where $\mathbf{D} = \mathrm{diag}(\check{\mathbf{h}})$ and $\check{\mathbf{h}} = \sqrt{n}\mathbf{F}\mathbf{h}$ is the DFT of the channel impulse response. Demodulation amounts to computing the DFT

$$\check{\mathbf{y}} \;\; = \;\; \mathbf{F}\mathbf{y} = \mathbf{D}\check{\mathbf{x}} + \check{\mathbf{z}} + \mathbf{F}\mathbf{e}, \tag{2}$$

where $\check{\mathbf{z}} = \mathbf{F}\mathbf{z}$ has the same statistics of $\mathbf{z}$. Without impulse noise, it is well-known that (2) reduces to a set of $n$ parallel Gaussian channels $\check{y}_i = \check{h}_i \check{x}_i + \check{z}_i$, for $i = 0, \ldots, n-1$. In the presence of the impulse noise, the performance of a standard OFDM demodulator may dramatically degrade since even a single impulse in the OFDM block (1) may cause significant degradation to all frequency domain symbols in the block, after DFT demodulation. This can be explained by considering the fact that $\mathbf{e}$ has $|\mathcal{I}|$ impulses with probability $\binom{n}{|\mathcal{I}|} p^{|\mathcal{I}|}(1-p)^{n-|\mathcal{I}|}$. Conditioned on $|\mathcal{I}|$, the variance per frequency domain component is $\frac{|\mathcal{I}|}{n} I_0$. Thus, for $I_0/n \gg N_0$, even a single impulse may cause significant degradation in all the frequency domain symbols. The probability that the OFDM block contains at least one impulse is $1-(1-p)^n$. For example, let us consider the case of $p = 10^{-4}$, $n = 1000$, $I_0 = 40$ dB, $N_0 = 0$ dB, and $\mathcal{E}_x = 20$ dB. In this case, the probability that an OFDM block is corrupted by at least one impulse is equal to $0.0952$. Thus about $10\%$ of the time, a block of $n$ OFDM frequency domain symbols is corrupted by one or more impulses and the nominal SNR of 20 dB is reduced by at least $I_0/n = 10$ dB. As a consequence, if the underlying coded-modulation scheme is designed for the nominal SNR of 20 dB (i.e., taking only AWGN into account), the link post-decoding frame error rate will incur a dramatic degradation, since about 10% of the symbols will be significantly noisier than the target SNR for the given coded modulation scheme. This will be particularly evident for systems employing powerful "modern" codes, such as Turbo Codes or LDPC codes [27], which perform close to capacity but have a very sharp *waterfall* behavior, such that even a small SNR decrease yields a dramatic increase of the post-decoding error probability. In Section VIII, in order to evaluate our scheme and compare with other approaches proposed in the literature in a way independent of the specific coded modulation scheme employed, we will express the link performance in terms of achievable rates with a Gaussian random coding ensemble, under different assumptions on how the impulse noise is handled by the receiver. Expressing the system performance in terms of information theoretic achievable rates is more meaningful than the classical BER of some uncoded modulation (e.g., [9],[20],[21],[25]), since virtually any modern wired or wireless system make use of power channel coding at the physical layer.

[2] Although the channel impulse response is of length $\upsilon+1$ where $\upsilon$ denotes the length of the cyclic prefix, typically significantly less than $n$, here $\mathbf{h}$ denotes its zero-padded version obtained by appending $n-\upsilon-1$ zeros, in order to form the first column of $\mathbf{H}$.

## III. Problem Formulation

Consider the OFDM frequency domain channel model (2). We use the sparse nature of $\mathbf{e}$ to estimate it and then remove it from the received signal. As in [25], [26], we use the subcarriers free of modulation symbols to estimate $\mathbf{e}$. Specifically, we assume that these form a block of consecutive subcarrier indices (with subcarrier index taken modulo $n$, as usually done in periodic DFT). We construct the time domain transmit signal as $\mathbf{x} = \mathbf{F}^{\mathsf{H}}\mathbf{S}_x\check{\mathbf{d}}$, where $\check{\mathbf{d}}$ is frequency domain data symbol vector of dimension $k \leq n$ and where $\mathbf{S}_x$ is an $n \times k$ "selection matrix" containing only one element equal to 1 per column, and with $m = n-k$ zero rows. The positions of the single 1s in the columns of $\mathbf{S}_x$ indicate the subcarrier used for data transmission in the OFDM system. The remaining subcarriers are either not used, or used for transmitting known pilot symbols in the frequency domain, which are not shown here since we do not deal with channel estimation. These known pilot symbols can be subtracted from the received signal at the receiver, such that for the purpose of this paper pilot symbols are equivalently treated as zero symbols (i.e., unused subcarriers). We denote by $\mathbf{S}$ the matrix with a single element equal to 1 per column, spanning the orthogonal complement of the columns of $\mathbf{S}_x$. The frequency domain vector is thus given by

$$\check{\mathbf{y}} = \mathbf{F}\mathbf{y} = \mathbf{D}\mathbf{S}_x\check{\mathbf{d}} + \mathbf{F}\mathbf{e} + \check{\mathbf{z}}. \tag{3}$$

Projecting onto the orthogonal complement of the signal subspace, we obtain

$$\mathbf{y}' = \mathbf{S}^{\mathsf{T}}\check{\mathbf{y}} = \mathbf{S}^{\mathsf{T}}\mathbf{F}\mathbf{e} + \mathbf{z}'. \tag{4}$$

Notice that $\mathbf{y}'$ is a noisy rank-deficient projection of the $n$-dimensional impulse noise vector onto a subspace of dimension $m \ll n$. The AWGN $\mathbf{z}'$ is a subsampling of $\check{\mathbf{z}}$, and therefore it is an AWGN vector of length $m$ and variance $N_0$ per component. For later use, we shall denote the $m \times n$ projection matrix by $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_n] = \mathbf{S}^{\mathsf{T}}\mathbf{F}$, where $\boldsymbol{\psi}_i$ denotes the $i$-th column of $\boldsymbol{\Psi}$. When the support $\mathcal{I}$ of $\mathbf{e}$ is known, we can equivalently write (4) as

$$\mathbf{y}' = \boldsymbol{\Psi}_{\mathcal{I}}\mathbf{e}_{\mathcal{I}} + \mathbf{z}', \tag{5}$$

where $\boldsymbol{\Psi}_{\mathcal{I}}$ denotes the submatrix formed by columns $\{\boldsymbol{\psi}_j : j \in \mathcal{I}\}$ of $\boldsymbol{\Psi}$, and $\mathbf{e}_{\mathcal{I}}$ denotes the vector of dimension $|\mathcal{I}|$ containing only the Gaussian not identically zero components of $\mathbf{e}$. Let $\widehat{\mathbf{e}}$ denote the resulting estimate of $\mathbf{e}$ produced some reconstruction algorithm with input $\mathbf{y}'$. Then, we can subtract the estimated impulse noise from the received signal such that the signal actually fed to the channel decoder is given by

$$\widehat{\mathbf{y}} = \mathbf{D}\mathbf{S}_x\check{\mathbf{d}} + \mathbf{F}(\mathbf{e} - \widehat{\mathbf{e}}) + \check{\mathbf{z}}. \tag{6}$$

The decoder treats this signal as if it was the output of a standard OFDM system without impulse noise. Note that a naive OFDM receiver that simply ignores the presence of the impulse noise, would treat (3) as the output of an OFDM system with Gaussian noise. It is apparent that the gain of the proposed scheme is significant if the variance per component of the residual noise,

$$\mathbf{v} = \mathbf{F}(\mathbf{e} - \widehat{\mathbf{e}}) + \check{\mathbf{z}}, \tag{7}$$

is significantly less than the variance per component of the corresponding frequency domain Gaussian plus impulse noise vector $\mathbf{F}\mathbf{e} + \check{\mathbf{z}}$.

## IV. Optimal Impulse Noise Estimation

The MMSE estimate of $\mathbf{e}$ given the observation $\mathbf{y}'$ minimizes the covariance of the residual noise, $\mathbf{e} - \hat{\mathbf{e}}$, and is given by

$$\hat{\mathbf{e}} \triangleq \mathbb{E}[\mathbf{e}|\mathbf{y}'] = \sum_{\mathcal{I}} p(\mathcal{I}|\mathbf{y}')\mathbb{E}[\mathbf{e}|\mathbf{y}', \mathcal{I}], \tag{8}$$

where $\mathcal{I}$ ranges over the set of possible support sets of $\mathbf{e}$. The above equation can be solved if we can evaluate the sum over all $\mathcal{I}$. However, it is apparent that for large $n$, the computational complexity required by the brute-force evaluation of the MMSE estimator is prohibitive (as there will be $2^n$ such support sets). The idea is to approximate equation (8) by expiating the special structure of the problem in order to reduce computational complexity.

### A. Calculating $\mathbb{E}[\mathbf{e}|\mathbf{y}', \mathcal{I}]$

Conditioned on the support $\mathcal{I}$ (that is assumed to follow Gaussian distribution), $\mathbf{y}'$ and $\mathbf{e}_{\mathcal{I}}$ are jointly Gaussian (see (5)) and therefore the conditional MMSE estimator coincides with the linear MMSE estimator, given by

$$\hat{\mathbf{e}}_{\mathcal{I}} \triangleq \mathbb{E}[\mathbf{e}_{\mathcal{I}}|\mathbf{y}', \mathcal{I}] = \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}}^{\mathsf{H}}\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}}}^{-1}\mathbf{y}', \tag{9}$$

where

$$\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}}} \triangleq \frac{1}{N_0}\mathbb{E}[\mathbf{y}'(\mathbf{y}')^{\mathsf{H}}|\mathcal{I}] = \mathbf{I} + \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}}\mathbf{\Psi}_{\mathcal{I}}^{\mathsf{H}}. \tag{10}$$

Then, letting $\mathbf{S}_{\mathcal{I}}$ denote the selection matrix of dimension $n \times |\mathcal{I}|$ such that each column $i$ is all zero with a single element equal to 1 in the position of the $i$-th component of the support $\mathcal{I}$, we have

$$\mathbb{E}[\mathbf{e}|\mathbf{y}', \mathcal{I}] = \mathbf{S}_{\mathcal{I}}\hat{\mathbf{e}}_{\mathcal{I}} = \frac{I_0}{N_0}\mathbf{S}_{\mathcal{I}}\mathbf{\Psi}_{\mathcal{I}}^{\mathsf{H}}\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}}}^{-1}\mathbf{y}'. \tag{11}$$

### B. Calculating $p(\mathcal{I}|\mathbf{y}')$

Using Bayes' rule, we can write

$$p(\mathcal{I}|\mathbf{y}') = \frac{p(\mathbf{y}'|\mathcal{I})p(\mathcal{I})}{p(\mathbf{y}')} = \frac{p(\mathbf{y}'|\mathcal{I})p(\mathcal{I})}{\sum_{\mathcal{I}'} p(\mathbf{y}'|\mathcal{I}')p(\mathcal{I}')}. \tag{12}$$

Given our model, we have $p(\mathcal{I}) = p^J(1-p)^{n-J}$ with $J = |\mathcal{I}|$. It remains to calculate $p(\mathbf{y}'|\mathcal{I})$. From (5), $\mathbf{y}'$ is conditionally Gaussian given $\mathcal{I}$, therefore

$$p(\mathbf{y}'|\mathcal{I}) = \frac{\exp\left(-\frac{1}{N_0}\mathbf{y}'^{\mathsf{H}}\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}}}^{-1}\mathbf{y}'\right)}{\det\left(\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}}}\right)}, \tag{13}$$

up to a constant multiplicative factor. The denominator in (12) involves the summation over all the possible supports which is computationally prohibitive. We discuss different approaches in the following subsection that intelligently avoid this complex computation exploiting the sparsity of the signal.

### C. Estimation of the support $\mathcal{I}$

Instead of summing over all $2^n$ possible supports, we consider only the set of the most probable (or "dominant") supports. This is obtained by considering the problem of support estimation from the observation model (4). Support recovery (i.e., finding the non-identically zero elements) of a sparse signal observed through a rank-deficient projection and possibly in additive noise is the central problem in the CS literature. Next, we review some of these techniques and compare them with our new proposed algorithm that exploits the particular structure of the problem at hand.

*1) CS based on convex relaxation:* Starting from (4), we can use the standard convex relaxation tools [40] - [45] to obtain an estimate of the sparse vector $\mathbf{e}$. Three alternatives widely proposed in the literature to implement sparse signal estimation in the presence of noise are reviewed in the following.

*a) Candes-Romberg-Tao SOCP estimator:* The original Second-Order Cone Programming (SOCP) problem proposed by [40], [45] is formulated for the reals. In our notation, this is given by

$$\mathbf{min}\ \|\widetilde{\mathbf{e}}\|_1\ \mathbf{s.t.}\ \|\mathbf{y}' - \mathbf{\Psi}\widetilde{\mathbf{e}}\|_2 \le \epsilon, \tag{14}$$

for some small enough $\epsilon$. This is a convex problem for both real and complex vectors, but in this form belongs to the SOCP class only for real vectors. In order to obtain a convenient formulation for complex vectors, we use "sum of norms" SOCP approach [46] and define the additional variables $t_0, \ldots, t_{n-1}$. The resulting SOCP problem becomes

$$\begin{aligned} &\mathbf{min}\ \sum_{i=0}^{n-1} t_i \\ &\mathbf{s.t.}\ |\widetilde{e}_i| \le t_i,\ \forall i = 0, \ldots, n-1\ \wedge\ \|\mathbf{y}' - \mathbf{\Psi}\widetilde{\mathbf{e}}\|_2 \le \epsilon. \end{aligned} \tag{15}$$

*b) Dantzig selector:* An LP estimator was also proposed for real vectors which is as follows [41]

$$\mathbf{min}\ \|\widetilde{\mathbf{e}}\|_1\ \mathbf{s.t.}\ \|\mathbf{y}' - \mathbf{\Psi}\widetilde{\mathbf{e}}\|_\infty \le \lambda. \tag{16}$$

For complex vectors, we arrive at a different SOCP given by

$$\begin{aligned} &\mathbf{min}\ \sum_{i=0}^{n-1} t_i \\ &\mathbf{s.t.}\ |\widetilde{e}_i| \le t_i,\ \forall i = 0, \ldots, n-1\ \wedge\ |y_i' - \mathbf{\Psi}_i\widetilde{\mathbf{e}}| \le \lambda. \end{aligned} \tag{17}$$

where $\mathbf{\Psi}_i$ is the $i^{\text{th}}$ row of $\mathbf{\Psi}$.

*c) LASSO:* In [42], the following convex relaxation method was considered which is known as LASSO [43]

$$\min \frac{1}{2}\|\mathbf{y}' - \mathbf{\Psi}\widetilde{\mathbf{e}}\|_2^2 + \gamma\|\widetilde{\mathbf{e}}\|_1. \tag{18}$$

The parameters $\epsilon, \lambda$ and $\gamma$ are related to the AWGN variance $N_0$ and are discussed, for example, in [41], [42], [45]. Either one or a combination of the three above algorithms can be employed to estimate the support $\mathcal{I}$.

The drawback of the convex relaxation approaches reviewed above is that they do not make use of any a priori statistical information on the signal to be estimated, other than it is some arbitrary vector with a certain sparsity. Moreover, given the highly structured nature of the sensing matrix $\mathbf{\Psi}$ in the problem at hand,[3] these methods do not perform as well as in the case of random selection of $m$ subcarriers over the whole signal bandwidth.

*2) BCH-type Error Correction over the Reals:* In the language of coding theory, the support recovery problem is analogous to solving the "locator polynomial" to find the location of the errors. Building on this analogy, [25], [26] replicates the BCH/Reed-Solomon framework over the real field and proposed a method to find impulse noise support (i.e., the location of the "errors") by polynomial extrapolation. This method, however, is not numerically stable in the presence of a non-negligible AWGN component. Furthermore, the selection of the free carriers is constrained by the required algebraic properties, and it is akin of choosing the zeros of the code generator polynomial in a BCH/Reed-Solomon construction. For example, it would be impossible to adaptively take advantage of unused frequencies which may depend on the channel frequency response and on the corresponding waterfilling power allocation.

*3) Fast Bayesian Matching Pursuit (FBMP):* A fast Bayesian recursive algorithm, presented in [39], finds the dominant support and the MMSE estimate of the sparse vector jointly, based on the Bernoulli-Gaussian priors. This scheme uses a greedy tree search over all the combinations in pursuit of the dominant support. The algorithm starts with empty active elements set. At each step, an active element is added to the current set that maximizes the MAP-Gaussian metric (13). This procedure is repeated till we reach $s$ active elements in a branch where $s$ is selected such that $P(J > s)$ is very small. The number of branches in the tree search is controlled by a parameter $D$ which governs the tradeoff between performance and complexity. Contrary to standard convex relaxation techniques, this algorithm makes use of the a priori statistical information and reduces complexity by employing a recursive implementation.

## V. Finding Dominant Support using Structure

None of the methods mentioned above make use of the structure of the sensing matrix $\mathbf{\Psi}$. It turns out that using this structure is very useful in reducing the complexity involved in calculating (8). To simplify the exposition, assume that $\frac{n}{m} = \ell$ is an integer. Assume also that we have $m$ consecutive carriers which we assume to lie at the edge of the band.[4] In this case, considering the projected observation defined in (4), we write $\mathbf{y}' = \mathbf{\Psi}\mathbf{e} + \mathbf{z}'$, where $\mathbf{\Psi}$ is an $m \times n$ submatrix of the DFT matrix $\mathbf{F}$, corresponding to a block of $m$ consecutive rows. Consider the columns of $\mathbf{\Psi}$ with indices in the set $\mathcal{F}_m = \{0, \ell, 2\ell, ..., (m-1)\ell\}$. The collection of such rows of $\mathbf{\Psi}$ forms a scaled version of the $m \times m$ DFT matrix $\mathbf{F}_m$ and thus span the column space of $\mathbf{\Psi}$. In fact, spanning happens in a special way. Note that it is not difficult to show that the magnitude correlation $\boldsymbol{\psi}_j$ and $\boldsymbol{\psi}_{j'}$ of $\mathbf{\Psi}$ is given by

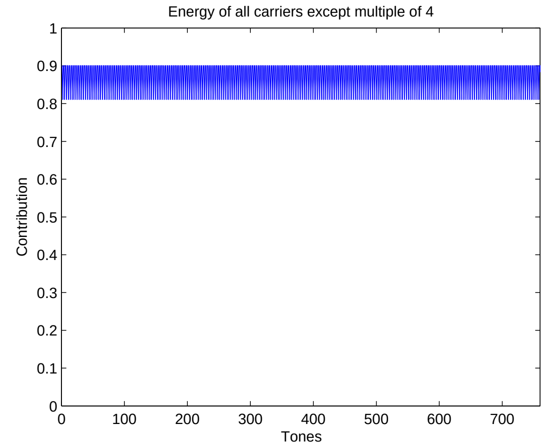


Fig. 1: Energy of the columns (Magnitude square of the inner product) of $\mathbf{\Psi}$ captured by the left and right basis vectors for $n = 1024$, $m = 256$, and $\ell = \frac{n}{m} = 4$.

$$\left|\boldsymbol{\psi}_j^{\mathsf{H}}\boldsymbol{\psi}_{j'}\right| = \begin{cases} 1, & (j = j') \\ \left|\frac{\sin\left(\pi(j-j')m/n\right)}{m\sin(\pi(j-j')/n)}\right|, & (j \neq j') \end{cases}. \tag{19}$$

Based on (19) we see that columns $j$ and $j'$ of $\mathbf{\Psi}$ are nearly orthogonal for large index distance $|j - j'|$. As a consequence, this means that the columns of indices $j \in [r\ell, (r+1)\ell]$, with $r = \lceil j\ell \rceil$, are spanned by the columns of indices $r\ell$ and $(r+1)\ell$ (referred to as the left and right basis vectors) in the orthonormal basis $\mathcal{F}_m$ defined above. This can be evidenced by Fig. 1 which shows that most of the energy of all columns (except $\mathcal{F}_m$) is captured by the left and right basis vectors.

This *semi-orthogonal structure* of $\mathbf{\Psi}$ leads to two observations. First, we can get an initial guess at the impulse noise location as follows: first, project $\mathbf{y}'$ on $\mathbf{F}_m$ to obtain $\mathbf{y}'' = \mathbf{F}_m\mathbf{y}'$; then, the location of the elements of $\mathbf{y}''$ with largest magnitude indicate (with high probability) the neighborhood of the position of the impulses. We construct clusters of size $L$ around these locations (details are discussed in Section VI). Thus, after this preliminary estimation phase, we can identify a set of clusters that contain the support of the impulse noise, with high probability.

The second observation is that the semi-orthogonal structure allows us to calculate the MMSE estimate (8) in a divide-and-conquer manner, assuming that effectively the support is

[3]This is due to the fact that the carriers belong to a continuous set of frequencies in the guard band.

[4]The consecutive selection of carriers is motivated by the structure of the guard band. The carriers could also be split between the edges of the band for the carriers are still consecutive (mod $n$).

contained in the set of clusters identified as described above. We start by illustrating an example. Suppose that the support $\mathcal{I}$ of $\mathbf{e}$ is contained in the union of two disjoint clusters, $\Omega_1$ and $\Omega_2$, which are known to the estimator. In particular, we let $\mathcal{I}_1 \triangleq \mathcal{I} \cap \Omega_1$ and $\mathcal{I}_2 \triangleq \mathcal{I} \cap \Omega_2$, such that $\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2$. As a consequence, we have $\mathbf{\Psi}_{\mathcal{I}} = [\mathbf{\Psi}_{\mathcal{I}_1}\ \mathbf{\Psi}_{\mathcal{I}_2}]$. Any pair of columns $\boldsymbol{\psi}_j$ and $\boldsymbol{\psi}_{j'}$ with $j \in \Omega_1$ and $j' \in \Omega_2$ are mutually near-orthogonal (if two clusters are very close to each other, they can be lumped into a single cluster). Therefore, $\mathbf{\Psi}_{\mathcal{I}_1}^{\mathsf{H}} \mathbf{\Psi}_{\mathcal{I}_2} \approx \mathbf{0}$.

In order to calculate the likelihood function $p(\mathbf{y}'|\mathcal{I})$ in (13), we need to calculate the inverse and determinant of $\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}}}$ defined in (10). Using the matrix inversion lemma, we can write $\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}}}^{-1}$ as shown in (20) at the bottom of the page where, consistently with (10), we let $\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_i}} = \mathbf{I} + \frac{I_0}{N_0} \mathbf{\Psi}_{\mathcal{I}_i} \mathbf{\Psi}_{\mathcal{I}_i}^{\mathsf{H}}$, for $i = 1, 2$. Using semi-orthogonality, (20) can be approximated as shown in (21) at the bottom of the page. By utilizing the matrix inversion lemma and some simplification, (21) leads to (23) which is valid up to an error term of order $\mathcal{O}(\frac{1}{m})$. It follows that

$$\frac{1}{N_0}\mathbf{y}'^{\mathsf{H}}\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}}}^{-1}\mathbf{y}' \approx -\frac{1}{N_0}\|\mathbf{y}'\|^2 + \frac{1}{N_0}\mathbf{y}'^{\mathsf{H}}\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_1}}^{-1}\mathbf{y}' + \frac{1}{N_0}\mathbf{y}'^{\mathsf{H}}\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_2}}^{-1}\mathbf{y}'. \tag{24}$$

Similarly, the determinant of $\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}}}$ is given by

$$\begin{aligned}
&\det\left(\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}}}\right) \\
&= \det\left(\mathbf{I} + \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_1}\mathbf{\Psi}_{\mathcal{I}_1}^{\mathsf{H}} + \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_2}\mathbf{\Psi}_{\mathcal{I}_2}^{\mathsf{H}}\right) \\
&= \det\left(\mathbf{I} + \tfrac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_1}\mathbf{\Psi}_{\mathcal{I}_1}^{\mathsf{H}}\right)\det\left(\mathbf{I} + \tfrac{I_0}{N_0}\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_1}}^{-\frac{1}{2}}\mathbf{\Psi}_{\mathcal{I}_2}\mathbf{\Psi}_{\mathcal{I}_2}^{\mathsf{H}}\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_1}}^{-\frac{1}{2}}\right) \\
&= \det\left(\mathbf{I} + \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_1}\mathbf{\Psi}_{\mathcal{I}_1}^{\mathsf{H}}\right)\det\left(\mathbf{I} + \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_2}^{\mathsf{H}}\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_1}}^{-1}\mathbf{\Psi}_{\mathcal{I}_2}\right) \\
&\approx \det\left(\mathbf{I} + \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_1}\mathbf{\Psi}_{\mathcal{I}_1}^{\mathsf{H}}\right)\det\left(\mathbf{I} + \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_2}\mathbf{\Psi}_{\mathcal{I}_2}^{\mathsf{H}}\right) \qquad (25) \\
&= \det\left(\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_1}}\right)\det\left(\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_2}}\right), \qquad (26)
\end{aligned}$$

where to obtain (25) we again used the fact that $\mathbf{\Psi}_{\mathcal{I}_1}$ and $\mathbf{\Psi}_{\mathcal{I}_2}$ are close to orthogonal. Denoting by $\mathcal{L}_{\mathcal{I}} \triangleq p(\mathbf{y}'|\mathcal{I})p(\mathcal{I})$ the *unnormalized a posteriori support distribution* and using (24) and (26) into (12), we obtain

$$\begin{aligned}
\mathcal{L}_{\mathcal{I}} \approx& p^J(1-p)^{n-J}\exp\left(\frac{1}{N_0}\|\mathbf{y}'\|^2\right)\cdot \\
&\frac{\exp\left(-\frac{1}{N_0}\mathbf{y}'^{\mathsf{H}}\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_1}}^{-1}\mathbf{y}'\right)}{\det(\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_1}})}\frac{\exp\left(-\frac{1}{N_0}\mathbf{y}'^{\mathsf{H}}\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_2}}^{-1}\mathbf{y}'\right)}{\det(\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_2}})} \\
=& p^J(1-p)^{n-J}\exp\left(\frac{1}{N_0}\|\mathbf{y}'\|^2\right)p(\mathbf{y}'|\mathcal{I}_1)p(\mathbf{y}'|\mathcal{I}_2),
\end{aligned}$$

where, consistently with (13), we define the support Likelihood Function

$$p(\mathbf{y}'|\mathcal{I}_i) = \frac{\exp\left(-\frac{1}{N_0}\mathbf{y}'^{\mathsf{H}}\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_i}}^{-1}\mathbf{y}'\right)}{\det\left(\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_i}}\right)}. \tag{27}$$

Generalizing this derivation in the case of $c$ disjoint clusters $\Omega_1, \ldots, \Omega_c$, whose union contains the support $\mathcal{I}$, letting $\mathcal{I}_i = \mathcal{I} \cap \Omega_i$, $J_i = |\mathcal{I}_i|$, and $\mathbf{\Psi}_{\mathcal{I}_i}$ denote the submatrix of $\mathbf{\Psi}$ obtained by taking the columns $\boldsymbol{\psi}_j$, with $j \in \mathcal{I}_i$, we obtain an approximation of the unnormalized a posteriori support distribution in the form

$$\mathcal{L}_{\mathcal{I}} \approx (1-p)^n \exp\left(\frac{c-1}{N_0}\|\mathbf{y}'\|^2\right)\prod_{i=1}^{c}\left(\frac{p}{1-p}\right)^{J_i} p(\mathbf{y}'|\mathcal{I}_i). \tag{28}$$

Summing over all possible supports $\mathcal{I}$ included in the union of the clusters $\Omega_i$, we find the posterior probability normalizing term as shown in (29) at the top of the next page, where we wrote the sum of products as a product of sums. Eventually, using (12), the (normalized) posterior support distribution under the assumption that $\mathcal{I} \subseteq \bigcup_{i=1}^{c}\Omega_i$ is obtained in the form

$$p(\mathcal{I}|\mathbf{y}') = \prod_{i=1}^{c}\frac{\left(\frac{p}{1-p}\right)^{J_i} p(\mathbf{y}'|\mathcal{I}_i)}{\sum_{\mathcal{I}_i'}\left(\frac{p}{1-p}\right)^{J_i'} p(\mathbf{y}'|\mathcal{I}_i')}. \tag{30}$$

Notice that the denominator in the above expression involves the sum over all the possible supports $\mathcal{I}_i' \subseteq \Omega_i$, i.e., the subsets of the support included in the cluster $\Omega_i$. This sum contains $2^L$ terms and calculating all of them can also be computationally expensive. Fortunately, for practical applications of impulse noise, $p \ll 1$, and thus the weighting coefficient $\left(\frac{p}{1-p}\right)^{J_i'}$ becomes negligible for large values of $J_i'$. For example, in the

$$\begin{aligned}
\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}}}^{-1} &= \left(\mathbf{I} + \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}}\mathbf{\Psi}_{\mathcal{I}}^{\mathsf{H}}\right)^{-1} = \left(\mathbf{I} + \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_1}\mathbf{\Psi}_{\mathcal{I}_1}^{\mathsf{H}} + \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_2}\mathbf{\Psi}_{\mathcal{I}_2}^{\mathsf{H}}\right)^{-1} \\
&= \mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_1}}^{-1} - \tfrac{I_0}{N_0}\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_1}}^{-1}\mathbf{\Psi}_{\mathcal{I}_2}\left(\mathbf{I} + \tfrac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_2}^{\mathsf{H}}\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_1}}^{-1}\mathbf{\Psi}_{\mathcal{I}_2}\right)^{-1}\mathbf{\Psi}_{\mathcal{I}_2}^{\mathsf{H}}\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_1}}^{-1} \qquad (20)
\end{aligned}$$

$$\begin{aligned}
\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}}}^{-1} &\approx \mathbf{I} - \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_1}\left(\mathbf{I} + \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_1}^{\mathsf{H}}\mathbf{\Psi}_{\mathcal{I}_1}\right)^{-1}\mathbf{\Psi}_{\mathcal{I}_1}^{\mathsf{H}} - \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_2}\left(\mathbf{I} + \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_2}^{\mathsf{H}}\mathbf{\Psi}_{\mathcal{I}_2}\right)^{-1}\mathbf{\Psi}_{\mathcal{I}_2}^{\mathsf{H}} \qquad (21) \\
&= -\mathbf{I} + \left(\mathbf{I} - \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_1}\left(\mathbf{I} + \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_1}^{\mathsf{H}}\mathbf{\Psi}_{\mathcal{I}_1}\right)^{-1}\mathbf{\Psi}_{\mathcal{I}_1}^{\mathsf{H}}\right) + \left(\mathbf{I} - \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_2}\left(\mathbf{I} + \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_2}^{\mathsf{H}}\mathbf{\Psi}_{\mathcal{I}_2}\right)^{-1}\mathbf{\Psi}_{\mathcal{I}_2}^{\mathsf{H}}\right) \qquad (22) \\
&= -\mathbf{I} + \left(\mathbf{I} + \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_1}\mathbf{\Psi}_{\mathcal{I}_1}^{\mathsf{H}}\right)^{-1} + \left(\mathbf{I} + \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_2}\mathbf{\Psi}_{\mathcal{I}_2}^{\mathsf{H}}\right)^{-1} \qquad (23)
\end{aligned}$$

$$
\begin{aligned}
p(\mathbf{y}') = \sum_{\mathcal{I}} \mathcal{L}_{\mathcal{I}} &= (1-p)^n \exp\left(\frac{c-1}{N_0}\|\mathbf{y}'\|^2\right) \sum_{\mathcal{I}} \prod_{i=1}^{c} \left(\frac{p}{1-p}\right)^{J_i} p(\mathbf{y}'|\mathcal{I}_i) \\
&= (1-p)^n \exp\left(\frac{c-1}{N_0}\|\mathbf{y}'\|^2\right) \sum_{\mathcal{I}_1} \cdots \sum_{\mathcal{I}_c} \prod_{i=1}^{c} \left(\frac{p}{1-p}\right)^{J_i} p(\mathbf{y}'|\mathcal{I}_i) \\
&= (1-p)^n \exp\left(\frac{c-1}{N_0}\|\mathbf{y}'\|^2\right) \prod_{i=1}^{c} \left(\sum_{\mathcal{I}_i} \left(\frac{p}{1-p}\right)^{J_i} p(\mathbf{y}'|\mathcal{I}_i)\right) \qquad (29)
\end{aligned}
$$

case when $p = 10^{-4}$, the value of the term is of the order $10^{-12}$ for $J_i' = 3$. Thus, we can truncate the sum to include supports of limited size such that the sum in the denominator contains a polynomial number of terms in the cluster size $L$ (further details are discussed in Section VI).

Using again quasi-orthogonality, we can easily show that

$$
\hat{\mathbf{e}}_{\mathcal{I}} = \frac{I_0}{N_0} \mathbf{\Psi}_{\mathcal{I}}^{\mathsf{H}} \mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}}}^{-1} \mathbf{y}' \approx \frac{I_0}{N_0} \begin{bmatrix} \mathbf{\Psi}_{\mathcal{I}_1}^{\mathsf{H}} \mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_1}}^{-1} \\ \mathbf{\Psi}_{\mathcal{I}_2}^{\mathsf{H}} \mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_2}}^{-1} \\ \vdots \\ \mathbf{\Psi}_{\mathcal{I}_c}^{\mathsf{H}} \mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_c}}^{-1} \end{bmatrix} \mathbf{y}'. \qquad (31)
$$

Letting $\mathbf{S}_{\mathcal{I}} = [\mathbf{S}_{\mathcal{I}_1}, \mathbf{S}_{\mathcal{I}_2}, \ldots, \mathbf{S}_{\mathcal{I}_c}]$, where the selection matrices $\mathbf{S}_{\mathcal{I}_i}$ are the vertical slices of columns of $\mathcal{S}_{\mathcal{I}}$ corresponding to the support positions in $\mathcal{I}_i$, and using (31) into (11), we can write

$$
\mathbb{E}[\mathbf{e}|\mathbf{y}', \mathcal{I}] = \frac{I_0}{N_0} \sum_{i=1}^{c} \mathbf{S}_{\mathcal{I}_i} \mathbf{\Psi}_{\mathcal{I}_i}^{\mathsf{H}} \mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_i}}^{-1} \mathbf{y}'. \qquad (32)
$$

Finally, combining (30) and (32), we obtain

$$
\begin{aligned}
\hat{\mathbf{e}} &= \sum_{\mathcal{I}} p(\mathcal{I}|\mathbf{y}') \mathbb{E}[\mathbf{e}|\mathbf{y}', \mathcal{I}] \\
&\approx \frac{I_0}{N_0} \sum_{i=1}^{c} \frac{\sum_{\mathcal{I}_i} \left(\frac{p}{1-p}\right)^{J_i} p(\mathbf{y}'|\mathcal{I}_i)\, \mathbf{S}_{\mathcal{I}_i} \mathbf{\Psi}_{\mathcal{I}_i}^{\mathsf{H}} \mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_i}}^{-1} \mathbf{y}'}{\sum_{\mathcal{I}_i} \left(\frac{p}{1-p}\right)^{J_i} p(\mathbf{y}'|\mathcal{I}_i)}. \qquad (33)
\end{aligned}
$$

The argument of truncating the sum to include supports of limited size for the practical small values of $p$ is valid here as well.

To summarize, we have obtained an approximated expression for the MMSE estimate of the impulse noise vector that can be calculated in a divide-and-conquer manner, by treating each cluster separately. This is possible provided that we can identify the set of clusters $\Omega_i$ that contains the support $\mathcal{I}$. The above discussion motivates the development of orthogonal clustering algorithm for impulse noise estimation which is described in the following.

## VI. ALGORITHM

### A. Initial Guess

First of all, an initial guess of the impulse noise locations is obtained. It consists of the following steps.

$(i)$ Project $\mathbf{y}'$ on $\mathbf{F}_m$ to obtain $\mathbf{y}'' = \mathbf{F}_m \mathbf{y}'$.

$(ii)$ The elements of $\mathbf{y}''$ with large magnitude determine (with high probability) the neighborhood of the position of the impulses.

### B. Cluster Formation

The clusters are constructed using the indices obtained from $\mathbf{y}''$ in the above step.

$(i)$ Let $\beta$ denote the index of the largest value of $\mathbf{y}''$. As it is very likely that an impulse is located in the neighborhood of the column of $\mathbf{\Psi}$ indexed by $\beta\ell$, a cluster $\Omega$ is formed around with $\Omega = \{\beta\ell, \beta\ell \pm 1, \beta\ell \pm 2, \ldots, \beta\ell \pm (\frac{L-1}{2})\}$, where $L = 2\ell - 1$ is the length of the cluster. The effect of choosing different lengths of the cluster on the performance of the algorithm are discussed in Section IX.

$(ii)$ If two constructed clusters are overlapping or close to each other (i.e., the difference between the last index and the first index of two clusters is less than $\frac{L-1}{2}$), they are joined into one big cluster. Thus, this results in formation of clusters with variable lengths.

$(iii)$ The above two steps are repeated till $c$ clusters are formed. Notice that $J$ follows a binomial distribution. In practical applications, the mean $pn$ is very small. Hence, we can use the Poisson approximation of the binomial probability mass function, i.e., $P(J = c) \approx \frac{(pn)^c}{c!} e^{-pn}$. For example, for $n = 1000$ and $p = 10^{-4}$ we have $pn = 0.1$, yielding $P(J > 1) = 0.0952$, $P(J > 2) = 0.0045$, $P(J > 3) = 1.5 \times 10^{-4}$, $P(J > 4) = 3.8 \times 10^{-6}$, We select the value of $c$ for which $P(J = c) > 10^{-6}$. Hence, in this case, $c$ can be safely limited to supports of cardinality up t o 4.

### C. Evaluating the impulse noise estimate for each cluster

For cluster $i$ with indices $\Omega_i$ and of length $L_i \geq L$ (As the clusters are semi-orthogonal, these calculations can be done *in parallel*),

$(i)$ Calculate the support likelihood function $p(\mathbf{y}'|\mathcal{I})$ for support of cardinality $J_i = 0, 1, ..., J_i^{\max}$ using (27). As $J_i$ also follows binomial distribution, $J_i^{\max}$ is obtained using the Poisson approximation, i.e., $P(J_i = J_i^{\max}) \approx \frac{(pL_i)^{J_i^{\max}}}{J_i^{\max}!} e^{-pn}$. Similar to $c$, $J_i^{\max}$ is also selected for which $P(J_i = J_i^{\max}) > 10^{-6}$. Note that this calculation will be performed for a total of $N_i = \binom{L_i}{0} + \binom{L_i}{1} + \cdots + \binom{L_i}{J_i^{\max}}$ combinations.

$(ii)$ Evaluate the estimate of the impulse noise using the following expression (similar to (33))

$$
\hat{\mathbf{e}}_i = \frac{\sum_{\mathcal{I}_i} \left(\frac{p}{1-p}\right)^{J_i} p(\mathbf{y}'|\mathcal{I}_i)\, \mathbf{S}_{\mathcal{I}_i} \mathbf{\Psi}_{\mathcal{I}_i}^{\mathsf{H}} \mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_i}}^{-1} \mathbf{y}'}{\sum_{\mathcal{I}_i} \left(\frac{p}{1-p}\right)^{J_i} p(\mathbf{y}'|\mathcal{I}_i)},
$$

where $\hat{\mathbf{e}}_i$ consists of 0s at all positions except for $\Omega_i$ and the summation is performed over $N_i$ terms.

### D. Evaluating the complete MMSE estimate $\hat{\mathbf{e}}$

Obtain $\hat{\mathbf{e}}$ by adding the $\hat{\mathbf{e}}_i$s calculated in the above step as follows (33)

$$\hat{\mathbf{e}} = \frac{I_0}{N_0} \sum_{i=1}^{c} \hat{\mathbf{e}}_i.$$

In the following section, we discuss how the complexity of the proposed algorithm can be reduced using the inherent structure of the partial DFT sensing matrix $\mathbf{\Psi}$.

## VII. Reducing Computational Complexity

We have seen before that, in order to compute the MMSE estimate $\hat{\mathbf{e}}$ knowing that the support $\mathcal{I}$ is contained in the union of disjoint clusters $\Omega_i$, is it sufficient to compute the terms $p(\mathbf{y}'|\mathcal{I}_i)$ for support subsets $\mathcal{I}_i \in \Omega_i$. This, in turn, requires calculating the inverse and determinant of $\mathbf{\Sigma}_{\Psi_{\mathcal{I}_i}}$. In this section we will show that it is enough to calculate these quantities for one cluster (say, for $i = 0$) and the corresponding quantities for the other clusters can be easily obtained with little extra computation. To this end, let $\boldsymbol{\psi}_i, \ \boldsymbol{\psi}_{i+1}, \ \cdots, \ \boldsymbol{\psi}_{i+L-1}$ and $\boldsymbol{\psi}_j, \ \boldsymbol{\psi}_{j+1}, \ \cdots, \ \boldsymbol{\psi}_{j+L-1}$ denote the columns of $\mathbf{\Psi}$ corresponding to the $i$-th and $j$-th clusters, respectively. Then, it is immediate to see that

$$\boldsymbol{\psi}_{j+k} = \boldsymbol{\psi}_{i+k} \odot \boldsymbol{\psi}_{\Delta_{ji}}, \quad k = 0, 1, ..., L-1, \tag{34}$$

where $\odot$ denotes element-wise product and $\boldsymbol{\psi}_{\Delta_{ji}}$ is a vector that depends only on the difference between the indices $j$ and $i$, i.e., $\Delta_{ji} \triangleq |(j-i) \bmod n|$. In particular, for the case where $\mathbf{\Psi}$ is formed by $m$ adjacent subcarriers placed at the edge of the transmission band, we have

$$\boldsymbol{\psi}_{\Delta_{ji}} = \left[\exp\left(-\tfrac{\jmath 2\pi}{n}(n-m-1)\Delta_{ji}\right) \cdots \exp\left(-\tfrac{\jmath 2\pi}{n}(n-1)\Delta_{ji}\right)\right]^{\mathsf{T}}.$$

Now, assume that we calculate the inverse $\mathbf{\Sigma}^{-1}_{\mathbf{\Psi}_{\mathcal{I}_i}}$ and determinant $\det(\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_i}})$ for a set of columns $\mathcal{I}_i \subseteq \Omega_i$, and let $\mathcal{I}_j \subseteq \Omega_j$ denote the same set of columns (i.e., with the same relative positions in the set) chosen in the $j$-th cluster $\Omega_j$. Then, In Appendix A we show that

$$\mathbf{y}'^{\mathsf{H}} \mathbf{\Sigma}^{-1}_{\mathbf{\Psi}_{\mathcal{I}_j}} \mathbf{y}' = \mathbf{y}'^{\mathsf{H}}_{\Delta_{ji}} \mathbf{\Sigma}^{-1}_{\mathbf{\Psi}_{\mathcal{I}_i}} \mathbf{y}'_{\Delta_{ji}}, \tag{35}$$

where $\mathbf{y}'_{\Delta_{ji}} = \mathbf{y}' \odot \boldsymbol{\psi}^*_{\Delta_{ji}}$, and

$$\det(\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_j}}) = \det(\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_i}}). \tag{36}$$

In other words, to calculate the terms appearing at the numerator and denominator of the $i$-th term in the sum with respect to $i$ in (33), it is sufficient to pre-calculate the determinant values $\det(\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_0}})$ and the inverse matrices $\mathbf{\Sigma}^{-1}_{\mathbf{\Psi}_{\mathcal{I}_0}}$ for a set of supports $\mathcal{I}_0$ of limited size[5] contained in a suitable defined reference cluster $\Omega_0$, and then use (35) and (36) to obtain the terms for arbitrary clusters $\Omega_i$.

[5]See discussion on the fact that the sums can be truncated to supports of small size, for practical values of $p \ll 1$.

## VIII. Performance Analysis

In this section we compare the performance of the proposed scheme with respect to other competing schemes for impulse noise estimation/cancellation, as well as with a "naive" receiver that does not try to actively cancel the impulse noise. Recall that the received frequency domain OFDM symbol after estimation/compensation is given by

$$\check{\mathbf{y}} = \mathbf{D}\mathbf{S}_x \check{\mathbf{d}} + \underbrace{\mathbf{F}(\mathbf{e} - \hat{\mathbf{e}}) + \mathbf{z}}_{\mathbf{v}}. \tag{37}$$

For simplicity, we assume that the signal is restricted to span the *same* set of subcarriers $i \in \mathcal{S}_x$, corresponding to the positions of the 1s in the columns of the selection matrix $\mathbf{S}_x$, for all systems under consideration. As discussed in Section I-B, such subcarriers may be determined by other system constraints.[6] The relevant channel output for a naive receiver that does not explicitly compensate for the impulse noise is obtained from (37) by setting $\hat{\mathbf{e}} = \mathbf{0}$.

Using known results on the achievable rate with Gaussian random coding ensembles and minimum distance decoding [47], the achievable rate (expressed in bit per symbol, or bit/s/Hz) is given by

$$R = \frac{1}{n} \sum_{i \in \mathcal{S}_x} \log\left(1 + \frac{|\check{h}_i|^2 \mathcal{E}_i}{\sigma_i^2}\right), \tag{38}$$

where $\mathcal{E}_i$ is the energy per symbol allocated to the $i$-th used subcarrier, and $\sigma_i^2 = \mathbb{E}[|v_i|^2]$ is the variance of the $i$-th frequency domain noise component. We assume that the transmitter has knowledge of the channel frequency response coefficients $\{|\check{h}_i|^2\}$ and of the noise variance at each subcarrier $\sigma_i^2$. Hence, it can maximizes $R$ in (38) with respect to the power allocation $\{\mathcal{E}_i\}$. This yields the classical waterfilling solution

$$R = \frac{1}{n} \sum_{i \in \mathcal{S}_x} \left[\log\left(\frac{\mu |\check{h}_i|^2}{\sigma_i^2}\right)\right]_+, \tag{39}$$

where the Lagrange multiplier (water level) $\mu$ is the solution of

$$\frac{1}{n} \sum_{i \in \mathcal{S}_x} \left[\mu - \frac{\sigma_i^2}{|\check{h}_i|^2}\right]_+ = \mathcal{E}_x,$$

where $[\cdot]_+$ denotes the positive part and $\mathcal{E}_x$ is the average symbol energy. Our numerical results will be given in terms of value of SNR= $\mathcal{E}_x/N_0$ in dB.

For the naive receiver, we have

$$\mathrm{cov}(\mathbf{v}) = \mathbf{F}\mathbb{E}[\mathbf{e}\mathbf{e}^{\mathsf{H}}]\mathbf{F}^{\mathsf{H}} + N_0\mathbf{I} = (pI_0 + N_0)\mathbf{I}, \tag{40}$$

such that $\sigma_i^2 = pI_0 + N_0$ for all subcarriers. Similarly, for *any* receiver that estimates the impulse noise and subtracts its estimate from the received signal, we need to calculate

$$\mathrm{cov}(\mathbf{v}) = \mathbf{F}\mathbb{E}[(\mathbf{e} - \hat{\mathbf{e}})(\mathbf{e} - \hat{\mathbf{e}})^{\mathsf{H}}]\mathbf{F}^{\mathsf{H}} + N_0\mathbf{I}. \tag{41}$$

Here, the challenge is to calculate the error covariance matrix $\mathbb{E}[(\mathbf{e} - \hat{\mathbf{e}})(\mathbf{e} - \hat{\mathbf{e}})^{\mathsf{H}}]$, since $\hat{\mathbf{e}}$ is, in general, the result of a

[6]If the set of unused subcarriers can be set freely, then it makes sense to optimize such set for each specific impulse noise estimation scheme. However, this goes beyond the scope of this paper and it is left for future investigation.

complicated sparse signal estimation scheme. In general, this can be obtained by Monte Carlo simulation. More details on the error covariance matrix for the scheme proposed in this paper are given in Section VIII-A.

It has been observed that the naive receiver performance can be improved by estimating the level of impulse noise power present in any given OFDM symbol by using the power on the unused subcarrier, i.e., from the observation of $\mathbf{y}' = \mathbf{S}^{\mathsf{T}}\check{\mathbf{y}}$, and by incorporating this information in the decoder.

Notice also that this "informed receiver" that makes use of the knowledge of the impulse noise level in each OFDM block is akin a soft version of the receiver that erases the noisy OFDM blocks [20], [21] (in this case, the receiver is told only "Good" or "Bad", and the bad symbols are treated as erasures). Here, for simplicity, we assume that the receiver is given the exact knowledge of the number of impulses $J = |\mathcal{I}|$ present in the OFDM symbol. The resulting conditional noise covariance matrix is given by

$$\mathrm{cov}(\mathbf{v}||\mathcal{I}| = J) = \frac{1}{\binom{n}{J}} \sum_{\mathcal{I}:|\mathcal{I}|=J} \left( \sum_{\ell\in\mathcal{I}} \mathbf{f}_\ell \mathbf{f}_\ell^{\mathsf{H}} \right) I_0 + N_0 \mathbf{I}, \quad (42)$$

with diagonal elements $\sigma_i^2 = JI_0/n + N_0$, independent of $i$. Since $J$ is a binomial random variable, the achievable rate in this case is given by

$$R_{\text{naive}}^{\text{genie}} = \frac{1}{n} \sum_{i\in\mathcal{S}_x} \sum_{s=0}^{n} \binom{n}{s} p^s (1-p)^{n-s} \log\left(1 + \frac{|\check{h}_i|^2 \mathcal{E}_i}{sI_0/n + N_0}\right). \quad (43)$$

The above rate cannot be maximized by straightforward water-filling, since the number of impulses occurring in any OFDM symbol, i.e., the random variable $J$, is unknown a priori to the transmitter. On the contrary, the transmitter can optimize the input power allocation by solving the convex optimization problem:

$$\mathbf{max} \ \ R_{\text{naive}}^{\text{genie}}, \quad \mathbf{s.t.} \ \ \frac{1}{n} \sum_{i\in\mathcal{S}_x} \mathcal{E}_i \le \mathcal{E}_x, \quad \mathcal{E}_i \ge 0 \ \ \forall\, i. \quad (44)$$

This can be solved by the standard method of Lagrangian multipliers and KKT conditions [48]. Details are omitted for the sake of space limitation. Furthermore, for the practically relevant case of large $n$ and small $p$, the already mentioned Poisson approximation of the binomial distribution allows to easily calculate $P(J = s) = \binom{n}{s} p^s (1-p)^{n-s} \approx \frac{(pn)^s}{s!} e^{-pn}$, and truncate the sum with respect to $s$ to the first dominant terms.

As mentioned, several current proposals to deal with impulse noise in OFDM (e.g., for DSL, and Powerline Communications), consist of interleaving long codewords over many OFDM blocks and introducing erasure of the symbols corresponding to OFDM blocks corrupted by the impulse noise. Such methods yield rates not larger than (43). In fact, the erasure technique can yield at most the rate

$$R_{\text{naive}}^{\text{erasure}} = \frac{(1-p)^n}{n} \sum_{i\in\mathcal{S}_x} \log\left(1 + \frac{|\check{h}_i|^2 \mathcal{E}_i}{N_0}\right), \quad (45)$$

which is strictly less than $R_{\text{naive}}^{\text{genie}}$ in (43).

For a smart receiver that explicitly estimates and subtracts the impulse noise (as the proposed scheme), we introduce also a "genie-aided" upper bound, beyond the rate in (39) that requires no information on the residual interference level at any given OFDM symbol at the decoder input. This bound assumes perfect knowledge of the number of impulses $J$ affecting each OFDM block. Hence, we have

$$R_{\text{smart}}^{\text{genie}} = \frac{1}{n} \sum_{i\in\mathcal{S}_x} \sum_{s=0}^{n} \binom{n}{s} p^s (1-p)^{n-s} \log\left(1 + \frac{|\check{h}_i|^2 \mathcal{E}_i}{\sigma_i^2(s)}\right), \quad (46)$$

where $\sigma_i^2(J)$ is the $i$-th diagonal element of the conditional covariance matrix $\mathrm{cov}(\mathbf{v}||\mathcal{I}| = J)$. The rate $R_{\text{smart}}^{\text{genie}}$ can be optimized with respect to the power allocation by solving convex optimization a problem formally identical to (44), once the coefficients $\sigma_i^2(J)$ are known. Next, we focus on the calculation of the residual noise variances $\sigma_i^2$ and $\sigma_i^2(J)$ for the system at hand.

### *A. Approximate Residual Noise Covariance using the Orthogonality of Clusters*

We compute (approximately) the quantities $\sigma_i^2$ and $\sigma_i^2(J)$ using the orthogonality of clusters for the case when the impulse noise support $\mathcal{I}$ is known. Recall that $\mathbf{e} = \mathbf{S}_\mathcal{I} \mathbf{e}_\mathcal{I}$ (see definitions in Section III). We start by calculating the error covariance matrix resulting from the MMSE estimate of $\mathbf{e}_\mathcal{I}$ given $\mathcal{I}$ and $\mathbf{y}'$. After some simple algebra, we obtain

$$\mathrm{cov}(\mathbf{e}_\mathcal{I} - \hat{\mathbf{e}}_\mathcal{I} | \mathcal{I}) = \left( \frac{1}{I_0} \mathbf{I} + \frac{1}{N_0} \mathbf{\Psi}_\mathcal{I}^{\mathsf{H}} \mathbf{\Psi}_\mathcal{I} \right)^{-1}. \quad (47)$$

As before, assume that $\mathcal{I}$ is included in the union of $c$ disjoint clusters $\Omega_1, \dots, \Omega_c$, and let again $\mathcal{I}_i = \mathcal{I} \cap \Omega_i$. Then, we can write $\mathbf{\Psi}_\mathcal{I} = [\mathbf{\Psi}_{\mathcal{I}_1}, \dots, \mathbf{\Psi}_{\mathcal{I}_c}]$. Replacing this expression into (47), we obtain

$$\mathrm{cov}(\mathbf{e}_\mathcal{I} - \hat{\mathbf{e}}_\mathcal{I} | \mathcal{I}) = \mathbb{E} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1c} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2c} \\ \vdots & & \ddots & \vdots \\ \mathbf{A}_{c1} & \mathbf{A}_{c2} & \cdots & \mathbf{A}_{cc} \end{bmatrix}^{-1}, \quad (48)$$

where

$$\mathbf{A}_{ij} = \begin{cases} \frac{1}{I_0}\mathbf{I} + \frac{1}{N_0} \mathbf{\Psi}_{\mathcal{I}_i}^{\mathsf{H}} \mathbf{\Psi}_{\mathcal{I}_1} & i = j \\ \frac{1}{N_0} \mathbf{\Psi}_{\mathcal{I}_i}^{\mathsf{H}} \mathbf{\Psi}_{\mathcal{I}_j} & i \neq j \end{cases}.$$

Since the clusters that are disjoint and therefore semi-orthogonal, we approximate

$$\mathbf{\Psi}_{\mathcal{I}_i}^{\mathsf{H}} \mathbf{\Psi}_{\mathcal{I}_j} \approx \mathbf{0} \quad \text{for } (j-i) \bmod n > Q, \quad (49)$$

for some suitable integer $Q$, which is a parameter that governs the accuracy of the approximation and allows us to significantly reduce the computation complexity. In order to evaluate the covariances of interest, namely, $\mathrm{cov}(\mathbf{e} - \hat{\mathbf{e}})$ and $\mathrm{cov}(\mathbf{e} - \hat{\mathbf{e}}||\mathcal{I}| = J)$, it is sufficient to perform the corresponding expectation with respect to the support $\mathcal{I}$. We have

$$\mathrm{cov}(\mathbf{e} - \hat{\mathbf{e}}) = \sum_{\mathcal{I}} p^J (1-p)^{n-J} \mathbf{S}_\mathcal{I} \mathrm{cov}(\mathbf{e}_\mathcal{I} - \hat{\mathbf{e}}_\mathcal{I} | \mathcal{I}) \mathbf{S}_\mathcal{I}^{\mathsf{H}}, \quad (50)$$

where, as usual, $J = |\mathcal{I}|$. This expression is generally hard to compute, although the sum can be truncated to include only supports of small cardinality, for sufficiently small $p$. Letting $Q = 1$, i.e., including only the diagonal blocks of $\text{cov}(\mathbf{e}_{\mathcal{I}} - \hat{\mathbf{e}}_{\mathcal{I}}|\mathcal{I})$, and letting $\mathbf{S}_{\mathcal{I}} = [\mathbf{S}_{\mathcal{I}_1}, \ldots, \mathbf{S}_{\mathcal{I}_c}]$, we obtain the simplified approximated expression

$$\begin{aligned}\text{cov}(\mathbf{e} - \hat{\mathbf{e}}) &\approx \sum_{\mathcal{I}} p^J (1-p)^{n-J} \mathbf{S}_{\mathcal{I}} \text{diag}\left(\mathbf{A}_{11}^{-1}, \ldots, \mathbf{A}_{cc}^{-1}\right) \mathbf{S}_{\mathcal{I}}^{\mathsf{H}} \\ &= \sum_{i=1}^{c} \sum_{\mathcal{I}_i} p^{J_i} (1-p)^{L-J_i} \mathbf{S}_{\mathcal{I}_i} \mathbf{A}_{ii}^{-1} \mathbf{S}_{\mathcal{I}_i}^{\mathsf{H}}. \qquad (51)\end{aligned}$$

where $L$ denotes the cluster size. Finally, the sought residual noise covariance is given by

$$\text{cov}(\mathbf{v}) = \mathbf{F}\text{cov}(\mathbf{e} - \hat{\mathbf{e}})\mathbf{F}^{\mathsf{H}} + N_0\mathbf{I}, \qquad (52)$$

and $\sigma_i^2$ is the $i$-th diagonal element. For the conditional residual noise covariance given $|\mathcal{I}| = J$, we have just to limit our summation to the supports of given cardinality $J$, i.e.,

$$\text{cov}(\mathbf{e}-\hat{\mathbf{e}}||\mathcal{I}| = J) = \frac{1}{\binom{n}{J}} \sum_{\mathcal{I}:|\mathcal{J}|=J} \mathbf{S}_{\mathcal{I}} \text{cov}(\mathbf{e}_{\mathcal{I}}-\hat{\mathbf{e}}_{\mathcal{I}}|\mathcal{I}) \mathbf{S}_{\mathcal{I}}^{\mathsf{H}}. \quad (53)$$

Again letting $Q = 1$, we obtain the simplified approximated expression

$$\text{cov}(\mathbf{e} - \hat{\mathbf{e}}||\mathcal{I}| = J) \approx \frac{1}{\binom{n}{J}} \sum_{\mathcal{I}:|\mathcal{I}|=J} \sum_{i=1}^{c} \mathbf{S}_{\mathcal{I}_i} \mathbf{A}_{ii}^{-1} \mathbf{S}_{\mathcal{I}_i}^{\mathsf{H}}. \qquad (54)$$

Unfortunately, this expression cannot be simplified further since when we constrain $|\mathcal{I}| = J$, then the cardinalities of the partial supports $|\mathcal{I}_i| = J_i$ are constrained to satisfy $\sum_{i=1}^{c} J_i = J$ with $0 \le J_i \le L$. Hence, we cannot sum independently over the partial supports $\mathcal{I}_i$, unlike in (51). Finally, the sought conditional residual noise covariance is given by

$$\text{cov}(\mathbf{v}||\mathcal{I}| = J) = \mathbf{F}\text{cov}(\mathbf{e} - \hat{\mathbf{e}}||\mathcal{I}| = J)\mathbf{F}^{\mathsf{H}} + N_0\mathbf{I}, \qquad (55)$$

and $\sigma_i^2(J)$ is the $i$-th diagonal element.

## IX. Simulations

We consider a system with $n$ = 1024 subcarriers per OFDM symbol and $m = \frac{n}{4} = 256$ null carriers at the edge of the transmission band. The channel SNR is equal to 20 dB. The Bernoulli-Gaussian impulse noise has probability $p$ ranging from $1 \times 10^{-5}$ to $1 \times 10^{-3}$ (i.e., approximately one impulse per 100 DMT symbols to one impulse in every DMT symbol). We assume that the average power of the impulse noise process, given by $pI_0$, is constant, i.e., $I_0$ is inversely proportional to $p$. This reflects a scenario where more catastrophic events are rare, and less catastrophic events are more frequent, which may be meaningful in practical settings. We set $pI_0 = 10$ for which INR changes from 60 dB to 40 dB for $p$ ranging from $1 \times 10^{-5}$ to $1 \times 10^{-3}$, respectively. A frequency-flat channel response is assumed, i.e., $|\check{h}_i|^2 = 1$ for all $i$ (subcarriers). In this way, the results are independent of the specific channel response and focus on the impact of impulse noise and gains achieved by smart receivers with estimation/cancellation. Thanks to this assumption, the waterfilling power allocation (39) and also

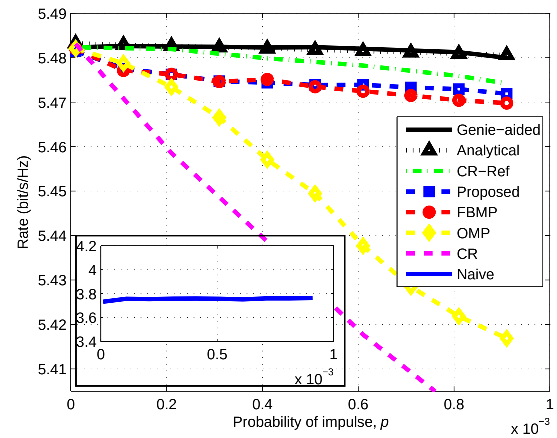


Fig. 2: Comparison of the achievable rate of the uninformed receivers as a function of the probability of impulse $p$.

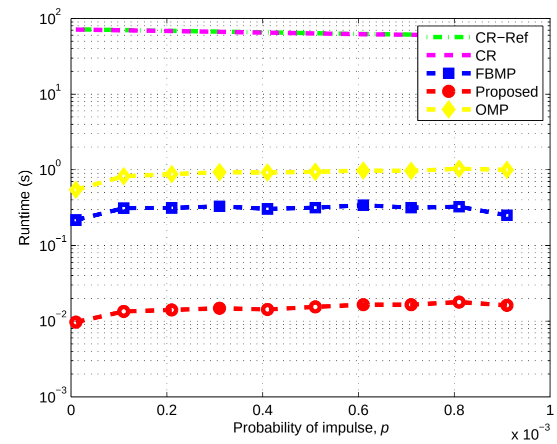


Fig. 3: Comparison of the mean runtime of the uninformed receivers as a function of the probability of impulse $p$.

the solution of the convex optimization in the case of $J$ known (43) depends only on the coefficients $\sigma_i^2$ and $\sigma_i^2(J)$ respectively and not on the channel frequency response.

In the following, we first present a comparison of different algorithms with respect to achievable rates and mean runtime for the uninformed and the informed cases (derived in Section VIII), followed by the effect of different parameters on the performance of the proposed algorithm.

### A. Comparison for the uninformed receiver case

In this subsection, we compare the performance and runtime of the following receivers.

- Naive receiver (no impulse noise estimation),
- Genie-aided receiver (this is the upper bound (benchmark), i.e., the case when $\mathcal{I}$ is perfectly known and MMSE is used for estimation of impulse amplitudes),
- Receiver that calculates the approximate residual noise covariance analytically using the orthogonality of clusters (given in (55)) for the case when $\mathcal{I}$ is known,

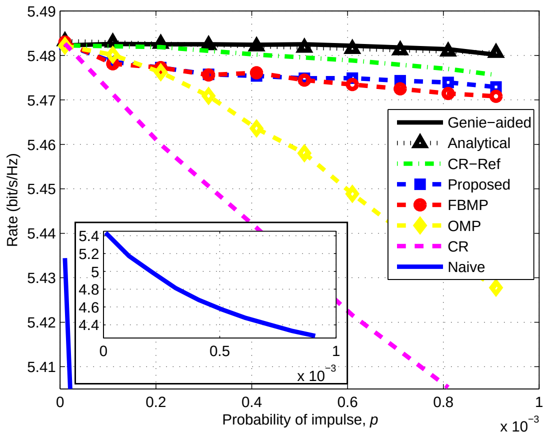


Fig. 4: Comparison of the achievable rate of the informed receivers as a function of the probability of impulse $p$.

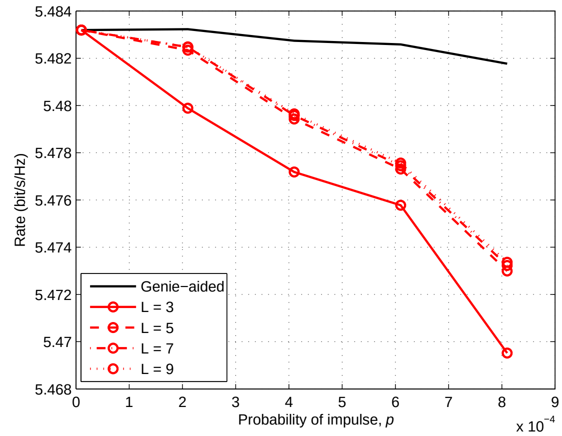


Fig. 5: Effect of the cluster length on the performance of the proposed algorithm.

- Proposed smart receiver (impulse noise estimation/compensation with our orthogonal clustering algorithm),
- Other smart receivers (different impulse noise estimation methods, including, $(i)$ CR [28] that uses CS based on convex relaxation (with $\epsilon = \sqrt{N_0(n + \sqrt{2n})}$) to find the impulse support, $(ii)$ CR-REF [29] that is similar to [28] but introduces a support refinement stage (based on a priori statistical information) followed by impulse amplitudes refinement using MMSE, $(iii)$ OMP [37], and $(iv)$ FBMP [39]. FBMP is implemented with number of greedy searches ($D$) set to 10).

The waterfilling power allocation (function of the coefficients $\sigma_i^2$) is computed for the uninformed receivers (that do not know the number of impulses in each OFDM symbol) using (39). The achievable rate for all the receivers is plotted versus $p$ in Fig. 2. The performance of the receiver that calculates the residual noise covariance analytically is quite close to the genie-aided receiver. This shows that the residual noise covariance derived in Section VIII-A using the orthogonality of clusters is a good approximation of the actual noise. The performance of the proposed algorithm is quite close to FBMP while it outperforms CR and OMP easily. It can be observed from Fig. 3 that the proposed algorithm is faster than OMP and FBMP by more than an order of magnitude while the long mean runtime of CR indicates its high complexity. It is interesting to point out that for fixed $pI_0$, the naive uninformed receiver has constant performance (shown as inset in Fig. 2 due to very low achievable rate as compared to other receivers) since its rate depends on the average impulse noise power ($pI_0$) as shown in (40). This indicates that it is not able to take advantage of the localization (sparsity) of the impulse noise.

### B. Comparison for the informed receiver case

The achievable rate for all the informed receivers (that are provided with the information of number of impulses in each OFDM symbol) is computed by solving the convex optimization problem (44) for $R_{\text{smart}}^{\text{genie}}$ given in (46). For each value of $p$, the binomial probability is calculated based on the Poisson approximation with $J = 0, 1, \cdots, J^{\max}$ where the value of $J^{\max}$ is computed based on the inequality $P(J = J^{\max}) > 10^{-6}$. Fig. 4 presents the achievable rate of these informed receivers as a function of $p$. Comparing with the uninformed case in Fig. 2, the performance of the naive informed receiver meliorates significantly as expected. This improved performance however is still worse than the smart uninformed receivers. The performance of the informed smart receivers is marginally improved as compared to the uninformed case.

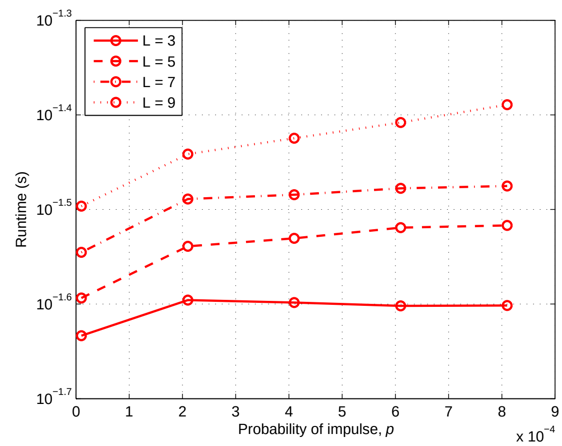


Fig. 6: Effect of the cluster length on the mean runtime of the proposed algorithm.

### C. Effect of the length of cluster $L$

As mentioned in Section VI, the cluster length is fixed at $L = 2\ell - 1 = 7$ for the current setting. In this subsection, we explore the effect of choosing different initial cluster lengths $L \in 3, 5, 7, 9$. Fig. 5 plots the achievable rate of the proposed algorithm in this case. It can be seen that the rate of the algorithm increases with increase in $L$ though there is not

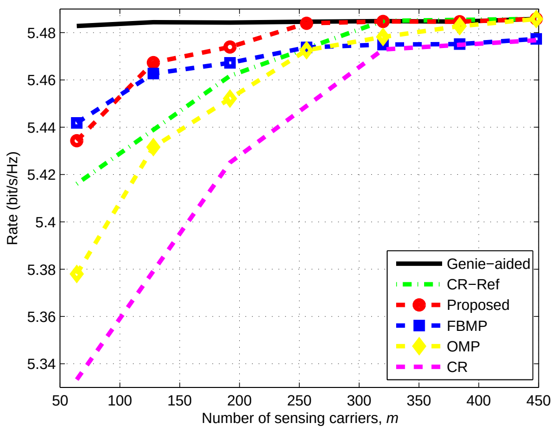


Fig. 7: Comparison of the achievable rate of the algorithms for different number of sensing carriers.

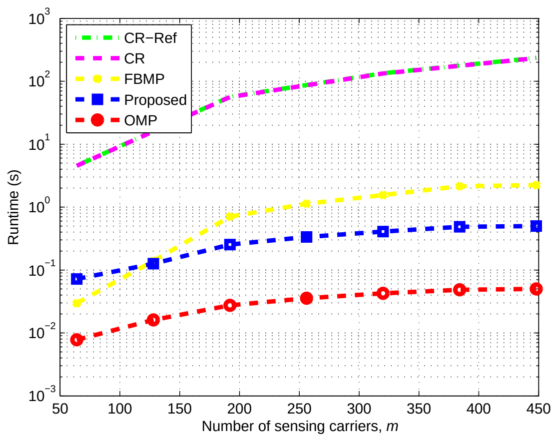


Fig. 8: Comparison of the mean runtime of the algorithms for different number of sensing carriers.

much difference between the rates for cluster lengths $L > 3$. This gain in performance is obviously achieved at the expense of higher complexity as presented in Fig. 6.

### *D. Effect of the number of sensing carriers $m$*

Fig. 7 compares the performance of the proposed algorithm with FBMP and OMP in terms of achievable rate when the number of sensing carriers ($m$) is varied from $64$ to $448$ with increments of $64$. It can be seen that the performance of all the algorithms improves with the increase in the number of sensing carriers with the proposed algorithm almost aligning with the perfect case after $m \geq 256$. Fig. 8 demonstrates the superior speed of the proposed algorithm. It is an order of magnitude faster than FBMP.

## X. Conclusion

In this paper, we propose a smart receiver that estimates and removes impulse noise in OFDM-based communications schemes (like DSL and PLC). Such intelligent receivers have significant advantage with respect to spectral efficiency, speed, and simplicity, when compared to the conventional retransmission techniques proposed in many standards. The impulse noise is assumed to be sparse and thus any sparse reconstruction algorithm can be utilized at the receiver. Unlike convex relaxation methods or matching pursuit algorithms for sparse reconstruction, the proposed approach makes a collective use of the structure of the sensing matrix (partial DFT matrix) in OFDM systems and a priori information of the impulse noise distribution, resulting in a fast and efficient algorithm. Simulation results demonstrate the superior performance of the proposed algorithm.

## Appendix A
## Proof of Equations (35) and (36)

The covariance $\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_j}}$ for a set of columns $\mathcal{I}_j \subseteq \Omega_j$ can be written as

$$\begin{aligned}\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_j}} &= \mathbf{I} + \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}_j}\mathbf{\Psi}_{\mathcal{I}_j}^{\mathsf{H}} \\ &= \mathbf{I} + \frac{I_0}{N_0}(\mathbf{B}_{\Delta_{ji}} \odot \mathbf{\Psi}_{\mathcal{I}_i})(\mathbf{B}_{\Delta_{ji}} \odot \mathbf{\Psi}_{\mathcal{I}_i})^{\mathsf{H}}, \quad (56)\end{aligned}$$

where $\mathbf{B}_{\Delta_{ji}}$ is a matrix consisting of $L$ identical columns $\boldsymbol{\psi}_{\Delta_{ji}}$, i.e. $\mathbf{B}_{\Delta_{ji}} = \left[\boldsymbol{\psi}_{\Delta_{ji}} \; \boldsymbol{\psi}_{\Delta_{ji}} \; \cdots \; \boldsymbol{\psi}_{\Delta_{ji}}\right]$. Its inverse can be evaluated by using the matrix inversion lemma,

$$\begin{aligned}\mathbf{\Sigma}^{-1}_{\mathbf{\Psi}_{\mathcal{I}_j}} &= \mathbf{I} - \frac{I_0}{N_0}\frac{(\mathbf{B}_{\Delta_{ji}} \odot \mathbf{\Psi}_{\mathcal{I}_i})(\mathbf{B}_{\Delta_{ji}} \odot \mathbf{\Psi}_{\mathcal{I}_i})^{\mathsf{H}}}{\mathbf{I} + \frac{I_0}{N_0}(\mathbf{B}_{\Delta_{ji}} \odot \mathbf{\Psi}_{\mathcal{I}_i})^{\mathsf{H}}(\mathbf{B}_{\Delta_{ji}} \odot \mathbf{\Psi}_{\mathcal{I}_i})} \\ &= \mathbf{I} - \frac{I_0}{N_0}\frac{(\mathbf{B}_{\Delta_{ji}} \odot \mathbf{\Psi}_{\mathcal{I}_i})(\mathbf{\Psi}^{\mathsf{H}}_{\mathcal{I}_i} \odot \mathbf{B}^{\mathsf{H}}_{\Delta_{ji}})}{\mathbf{I} + \frac{I_0}{N_0}(\mathbf{\Psi}^{\mathsf{H}}_{\mathcal{I}_i} \odot \mathbf{B}^{\mathsf{H}}_{\Delta_{ji}})(\mathbf{B}_{\Delta_{ji}} \odot \mathbf{\Psi}_{\mathcal{I}_i})} \\ &= \mathbf{I} - \frac{I_0}{N_0}\frac{(\mathbf{B}_{\Delta_{ji}} \odot \mathbf{\Psi}_{\mathcal{I}_i})(\mathbf{\Psi}^{\mathsf{H}}_{\mathcal{I}_i} \odot \mathbf{B}^{\mathsf{H}}_{\Delta_{ji}})}{\mathbf{I} + \frac{I_0}{N_0}(\mathbf{\Psi}^{\mathsf{H}}_{\mathcal{I}_i}\mathbf{\Psi}_{\mathcal{I}_i})},\end{aligned}$$

where in the last line $\mathbf{B}^{\mathsf{H}}_{\Delta_{ji}}$ and $\mathbf{B}_{\Delta_{ji}}$ cancel each other out. As we are actually interested in the quantity $\mathbf{y}'^{\mathsf{H}}\mathbf{\Sigma}^{-1}_{\mathbf{\Psi}_{\mathcal{I}_j}}\mathbf{y}'$, substituting value of $\mathbf{\Sigma}^{-1}_{\mathbf{\Psi}_{\mathcal{I}_j}}$ in it

$$\begin{aligned}&\mathbf{y}'^{\mathsf{H}}\mathbf{\Sigma}^{-1}_{\mathbf{\Psi}_{\mathcal{I}_j}}\mathbf{y}' \\ &= \mathbf{y}'^{\mathsf{H}}\mathbf{y}' - \frac{I_0}{N_0}\frac{\mathbf{y}'^{\mathsf{H}}(\mathbf{B}_{\Delta_{ji}} \odot \mathbf{\Psi}_{\mathcal{I}_i})(\mathbf{\Psi}^{\mathsf{H}}_{\mathcal{I}_i} \odot \mathbf{B}^{\mathsf{H}}_{\Delta_{ji}})\mathbf{y}'}{\mathbf{I} + \frac{I_0}{N_0}(\mathbf{\Psi}^{\mathsf{H}}_{\mathcal{I}_i}\mathbf{\Psi}_{\mathcal{I}_i})} \\ &= \mathbf{y}'^{\mathsf{H}}\mathbf{y}' - \frac{I_0}{N_0}\frac{(\mathbf{y}'^{\mathsf{H}} \odot \boldsymbol{\psi}^{\mathsf{T}}_{\Delta_{ji}})(\mathbf{\Psi}_{\mathcal{I}_i}\mathbf{\Psi}^{\mathsf{H}}_{\mathcal{I}_i})(\boldsymbol{\psi}^{*}_{\Delta_{ji}} \odot \mathbf{y}')}{\mathbf{I} + \frac{I_0}{N_0}(\mathbf{\Psi}^{\mathsf{H}}_{\mathcal{I}_i}\mathbf{\Psi}_{\mathcal{I}_i})} \\ &= (\mathbf{y}' \odot \boldsymbol{\psi}^{*}_{\Delta_{ji}})^{\mathsf{H}}(\mathbf{y}' \odot \boldsymbol{\psi}^{*}_{\Delta_{ji}}) \\ &\qquad - \frac{I_0}{N_0}\frac{(\mathbf{y}' \odot \boldsymbol{\psi}^{*}_{\Delta_{ji}})^{\mathsf{H}}(\mathbf{\Psi}_{\mathcal{I}_i}\mathbf{\Psi}^{\mathsf{H}}_{\mathcal{I}_i})(\mathbf{y}' \odot \boldsymbol{\psi}^{*}_{\Delta_{ji}})}{\mathbf{I} + \frac{I_0}{N_0}(\mathbf{\Psi}^{\mathsf{H}}_{\mathcal{I}_i}\mathbf{\Psi}_{\mathcal{I}_i})} \\ &= (\mathbf{y}' \odot \boldsymbol{\psi}^{*}_{\Delta_{ji}})^{\mathsf{H}}\left[\mathbf{I} - \frac{I_0}{N_0}\frac{\mathbf{\Psi}_{\mathcal{I}_i}\mathbf{\Psi}^{\mathsf{H}}_{\mathcal{I}_i}}{\mathbf{I} + \frac{I_0}{N_0}(\mathbf{\Psi}^{\mathsf{H}}_{\mathcal{I}_i}\mathbf{\Psi}_{\mathcal{I}_i})}\right](\mathbf{y}' \odot \boldsymbol{\psi}^{*}_{\Delta_{ji}}) \\ &= \mathbf{y}'^{\mathsf{H}}_{\Delta_{ji}}\mathbf{\Sigma}^{-1}_{\mathbf{\Psi}_{\mathcal{I}_i}}\mathbf{y}'_{\Delta_{ji}},\end{aligned}$$

where $\mathbf{\Sigma}^{-1}_{\mathbf{\Psi}_{\mathcal{I}_i}} = \mathbf{I} - \frac{I_0}{N_0}\frac{\mathbf{\Psi}_{\mathcal{I}_i}\mathbf{\Psi}^{\mathsf{H}}_{\mathcal{I}_i}}{\mathbf{I} + \frac{I_0}{N_0}(\mathbf{\Psi}^{\mathsf{H}}_{\mathcal{I}_i}\mathbf{\Psi}_{\mathcal{I}_i})}$ (from matrix inversion lemma).

As for (36), starting from (56), we have

$$\begin{aligned}\det(\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_j}}) &= \det\left(\mathbf{I}+\frac{I_0}{N_0}(\mathbf{B}_{\Delta_{ji}}\odot\mathbf{\Psi}_{\mathcal{I}_i})(\mathbf{\Psi}_{\mathcal{I}_i}^{\mathsf{H}}\odot\mathbf{B}_{\Delta_{ji}}^{\mathsf{H}})\right)\\ &= \det\left(\mathbf{I}+\frac{I_0}{N_0}(\mathbf{\Psi}_{\mathcal{I}_i}^{\mathsf{H}}\odot\mathbf{B}_{\Delta_{ji}}^{\mathsf{H}})(\mathbf{B}_{\Delta_{ji}}\odot\mathbf{\Psi}_{\mathcal{I}_i})\right)\\ &= \det\left(\mathbf{I}+\frac{I_0}{N_0}(\mathbf{\Psi}_{\mathcal{I}_i}^{\mathsf{H}}\mathbf{\Psi}_{\mathcal{I}_i})\right)\\ &= \det\left(\mathbf{I}+\frac{I_0}{N_0}(\mathbf{\Psi}_{\mathcal{I}_i}\mathbf{\Psi}_{\mathcal{I}_i}^{\mathsf{H}})\right)=\det(\mathbf{\Sigma}_{\mathbf{\Psi}_{\mathcal{I}_i}}).\end{aligned}$$

## REFERENCES

[1] T. Y. Al-Naffouri, A. A. Quadeer, and G. Caire, "Impulsive noise estimation and cancellation in DSL using orthogonal clustering," *IEEE Int. Symp. on Inform. Theory*, Russia, 2011.

[2] T. Starr, M. Sorbara, J. M. Cioffi and P. J. Silverman, *DSL adcances.* Prentice Hall Professional, 2003.

[3] "Network and Customer Installation Interfaces - Asymmetric Digital Subscriber Line (ADSL) metallic interfaces," ANSI standard T1.413, 1995.

[4] H. Hrasnica, A. Haidine, and R. Lehnert, *Broadband powerline communications.* London, U.K.: Wiley, 2004.

[5] H.C. Ferreira, L. Lampe, J. Newbury, T.G. Swart, "Power line communications: Theory and applications for narrowband and broadband communications over power lines," John Wiley & Sons, ISBN 978-0-470-74030-9, 2010.

[6] A. F. Molisch, *Wireless communications.* New York: IEEE Press/Wiley, 2005.

[7] M. Zimmermann and K. Dostert, "An analysis of the broadband noise scenario in powerline networks." *International Symposium on Powerline Communications and its Applications*, 2000.

[8] D. Fertonani and G. Colavolpe, "On reliable communications over channels impaired by bursty impulse noise," *IEEE Trans. on Comm.*, vol.57, no.7, pp.2024-2030, July 2009.

[9] J. Lin, M. Nassar, and B. L. Evans, "Impulsive noise mitigation in powerline communications using sparse Bayesian learning," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, Jul. 2013, pp. 1172-1183.

[10] A. Al-Dweik, A. Hazmi, B. Sharif, and C. Tsimenidis, "Efficient interleaving technique for OFDM system over impulsive noise channels," *Proc. IEEE Int. Symp. Pers. Indoor and Mobile Radio Comm.*, 2010.

[11] J. M. Cioffi, "Advanced digital communication," *class reader EE379C*, Stanford University. Available from http://www.stanford.edu/class/ee379c/

[12] K. Gulati, B. Evans, J. Andrews, and K. Tinsley, "Statistics of co-channel interference in a field of Poisson and Poisson-Poisson clustered interferers," *IEEE Trans. on Signal Proc.*, vol. 58, no. 12, pp. 62076222, 2010.

[13] D. Middleton, "Non-Gaussian noise models in signal processing for telecommunications: New methods and results for Class A and Class B noise models," *IEEE Trans. on Info. Theory*, vol. 45, no. 4, pp. 11291149, 1999.

[14] G. Tsihrintzis and C. Nikias, "Fast estimation of the parameters of alpha-stable impulsive interference," *IEEE Trans. on Signal Processing*, vol. 44, no. 6, pp. 14921503, 1996.

[15] D. Gianfelici, L. Franchina, and S. A. Kosmopoulos, "Symbol-to-symbol performance evaluation of densely populated asynchronous DS-CDMA in Gaussian and impulsive noise environments: A FourierBessel series approach," vol. 46, no. 7, pp. 866-869, Jul. 1998.

[16] M. Ghosh, "Analysis of the effect of impulse noise on multicarrier and single carrier QAM systems," *IEEE Trans. on Comm.*, vol. 44, no. 2, pp. 145-147, Feb. 1996.

[17] H. Dai and H. V. Poor, "Crosstalk mitigation in DMT VDSL with impulse noise" *IEEE Trans. on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, no. 10, pp. 1205-1213, Oct. 2001.

[18] J. Rinne and A. Hazmi, "Impulse burst position detection and channel estimation schemes for OFDM systems," *IEEE Trans. on Consumer Electronics*, vol. 49, no. 3, 2003.

[19] H. Matsuo, D. Umehara, M. Kawai and Y. Morihim, "An iterative detection scheme for OFDM over impulsive noise channels," *Int. Symp. on Powerline Communications*, Athens, Mar. 2002.

[20] S. V. Zhidkov, "Performance analysis and optimization of OFDM receiver with blanking nonlinearity in impulsive noise environment," *IEEE Trans. Vehicular Technology*, vol. 55, no. 1, pp. 234-242, 2006.

[21] W. Yu, D. Toumpakaris, J. M. Cioffi, D. Gardan and F. Gauthier, "Performance of asymmetric digital subscriber lines in an impulse noise environment," *IEEE Trans. Comm.*, vol. 51, no. 10, Oct. 2003.

[22] V. Mitlin, *Performance optimization of digital communications systems.* Boca Raton: Auerbach Publications, 2006.

[23] R. Boden, "Real-time ARQ protocol for improved impulsive noise robustness of ADSL systems," *Next Generation Internet Networks*, pp. 40-47, Jun. 2007.

[24] ITU-T Recommendation G.998.4: Improved impulse noise protection for DSL transceivers, ITU, Geneva, 2010.

[25] F. Abdelkefi, P. Duhamel, and F. Alberge, "Impulsive noise cancellation in multicarrier transmission," *IEEE Trans. on Comm.* vol. 53, no. 1, Jan. 2005.

[26] F. Abdelkefi, P. Duhamel, and F. Alberge, "A Necessary condition on the location of pilot tones for maximizing the correction capacity in OFDM systems," *IEEE Trans. on Comm.* vol. 55, no. 2, Feb. 2007.

[27] T. Richardson and R. Urbanke, *Modern coding theory.* Cambridge, UK: Cambridge University Press, 2008.

[28] G. Caire, T. Y. Al-Naffouri, and A. Narayanan, "Impulse noise cancellation in OFDM: an application of compressed sensing," *IEEE Int. Symp. on Inform. Theory*, Canada, Jul. 2008.

[29] T. Y. Al-Naffouri, A. A. Quadeer, H. Hmida, and F. F. Al-Shaalan, "Impulsive noise estimation and cancellation in DSL Using compressive sensing," *IEEE Int. Symp. Circuits and Systems*, Brazil, May 2011.

[30] M. Gast, *802.11 ac: A survival guide.* O'Reilly Media, Inc., 2013.

[31] T.M. Cover and J.A. Thomas, *Elements of information theory.* Wiley, New York, 1991.

[32] S. Panigrahi, *Advanced resource allocation for multicarrier DSL Systems: Power, modulation & coding allocation mechanisms for multiuser & multicarrier systems with objective of maximizing capacity or error robustness.* VDM Verlag, Germany, 2008.

[33] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Sig. Proc. Mag.*, vol. 25, no. 2, March 2008.

[34] L. Lampe, "Bursty impulsive noise detection by compressed sensing," *IEEE International Symposium on Power Line Communications and Its Applications (ISPLC)*, April 2011.

[35] X. Tan and J. Li, "Computationally efficient sparse Bayesian learning via belief propagation," *IEEE Trans. on Signal Proc.*, vol. 58, no. 4, pp. 2010-2021, 2010.

[36] Shihao Ji, Ya Xue, and Lawrence Carin, "Bayesian compressive sensing," *IEEE Trans. on Signal Proc.*, vol. 56, no. 6, pp. 2346-2356, Jun. 2008.

[37] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," *Asilomar Conf. on Signals, Systems and Comput.*, Nov. 1993. (software available at http://www.personal.soton.ac.uk/tb1m08/sparsify/sparsify.html.)

[38] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. on Inform. Theory*, vol. 53, pp. 46554666, 2007.

[39] P. Schniter, L. C. Potter, and J. Ziniel, "Fast Bayesian Matching Pursuit," *Inform. Theory and Appl. Workshop*, pp. 326-333, 2008. (software available at http://www2.ece.ohio-state.edu/∼zinielj/fbmp/download.html.)

[40] E. Candes, J. Romberg and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[41] E. Candes and T. Tao, "The Dantzig selector: statistical estimation when $p$ is much larger than $n$," *Annals of Statistics*, vol. 35, no. 6, 2313–2351, 2007.

[42] J. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. on Inform. Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.

[43] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. R. Statist. Soc. Series B*, vol. 58, no. 1, pp. 267288, 1996.

[44] E. J. Candes and T. Tao. "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, pp. 5406-5425, Dec. 2006.

[45] E. J. Candes and P. Randall, "Highly robust error correction by convex programming," *IEEE Trans. on Inform. Theory*, vol. 54, no. 7, pp. 2829–2840, Jun. 2008.

[46] M. Sousa Lobo, L. Vandenberghe, S. Boyd and H. Lebret, "Applications of second order cone programming," unpublished `http://www.stanford.edu/˜boyd/papers/socp.html`

[47] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 42, no. 5, pp. 1520–1529, Sep. 1996.

[48] S. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge, U.K.: Cambridge Univ. Press, 2004.

PLACE
PHOTO
HERE

**Tareq Y. Al-Naffouri** (M'10) received the B.S. degree in mathematics and electrical engineering (with first honors) from King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, in 1994, the M.S. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 1998, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 2004.

He was a Visiting Scholar at the California Institute of Technology, Pasadena, CA, USA, from January to August 2005 and during the summer of 2006. He was a Fulbright Scholar at the University of Southern California from February to September 2008. He held internship positions at NEC Research Labs, Tokyo, Japan, in 1998, Adaptive Systems Lab., University of California, Los Angeles, CA, USA, in 1999, National Semiconductor, Santa Clara, CA, USA, in 2001 and 2002, and Beceem Communications Santa Clara, CA, USA, in 2004. He is currently an Associate Professor in the Electrical Engineering Department, King Fahd University of Petroleum and Minerals, Saudi Arabia, and jointly at the Electrical Engineering Department, King Abdullah University of Science and Technology. He has over 80 publications in journal and conference proceedings, 9 standard contributions, and four issued and four pending patents. His research interests lie in the areas of adaptive and statistical signal processing and their applications to wireless communications, seismic signal processing, and in multiuser information theory. He has recently been interested in compressive sensing and random matrix theory and their applications.

Dr. Al-Naffouri has been serving as an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING since August 2013. He is the recipient of a Best Student Paper Award at the 2001 IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing for his work on adaptive filtering analysis, the IEEE Education Society Chapter Achievement Award in 2008, and the Al-Marai Award for innovative research in communication in 2009.

PLACE
PHOTO
HERE

**Ahmed A. Quadeer** received the B.E. degree in electronic engineering with honors from NED University of Engineering and Technology, Karachi, Pakistan, in 2006 and the M.S. degree in electrical engineering, with majors in communication and signal processing, from King Fahd University of Petroleum & Minerals (KFUPM), Dhahran, Saudi Arabia, in 2008. In 2006, he worked as an Executive Engineer in UFONE, a leading telecommunication company, in Pakistan for six months. He worked as a Lecturer in Electrical Engineering department at KFUPM for four years before joining Hong Kong University of Science and Technology (HKUST) in September 2012 to pursue the Ph.D. degree in Electronic and Computer Engineering. His major area of interest is signal processing for communication that includes channel estimation and multiuser detection in OFDM, compressed sensing and its applications. He is currently working on applications of statistical signal processing in immunology.

Mr. Quadeer is the recipient of Hong Kong Ph.D. Fellowship Scheme (HKPFS) provided by the Research Grants Council (RGC) of Hong Kong.

PLACE
PHOTO
HERE

**Giuseppe Caire** (S'92-M'94-SM'03-F'05) was born in Torino, Italy, in 1965. He received the B.Sc. in Electrical Engineering from Politecnico di Torino (Italy), in 1990, the M.Sc. in Electrical Engineering from Princeton University in 1992 and the Ph.D. from Politecnico di Torino in 1994. He was a recipient of the AEI G.Someda Scholarship in 1991, has been with the European Space Agency (ESTEC, Noordwijk, The Netherlands) from May 1994 to February 1995, was a recipient of the COTRAO Scholarship in 1996 and of a CNR Scholarship in 1997.

He has been visiting Princeton University in Summer 1997 and Sydney University in Summer 2000. He has been Assistant Professor in Telecommunications at the Politecnico di Torino, Associate Professor at the University of Parma, Italy, Professor with the Department of Mobile Communications at the Eurecom Institute, Sophia-Antipolis, France, and he is currently a professor of Electrical Engineering with the Viterbi School of Engineering, University of Southern California, Los Angeles, CA.

He served as Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS in 19982001 and as Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY in 20012003. He received the Jack Neubauer Best System Paper Award from the IEEE Vehicular Technology Society in 2003, and the IEEE Communications Society and Information Theory Society Joint Paper Award in 2004 and in 2011. Giuseppe Caire has served in the Board of Governors of the IEEE Information Theory Society from 2004 to 2007, and as President of the IEEE Information Theory Society in 2011. His main research interests are in the field of communications theory, information theory, channel and source coding with particular focus on wireless communications.