Impulsive Noise Estimation and Cancellation in DSL using Orthogonal Clustering

Tareq Y. Al-Naffouri[1] and Ahmed A. Quadeer[2]
Electrical Engineering Department
King Fahd University of Petroleum & Minerals
Email: {naffouri[1], aquadeer[2]}@kfupm.edu.sa

Giuseppe Caire
Electrical Engineering Department
University of Southern California
Email: caire@usc.edu

***Abstract*—Impulsive noise is the bottleneck that limits the distance at which DSL communications can take place. By considering impulsive noise a sparse vector, recently developed sparse reconstruction algorithms can be utilized to combat it. We propose an algorithm that utilizes the guard band null carriers for the impulsive noise estimation and cancellation. Instead of relying on ℓ_1 minimization as done in some popular general-purpose compressive sensing (CS) schemes, the proposed method exploits the structure present in the problem and the available a priori information jointly for sparse signal recovery. The computational complexity of the proposed algorithm is very low as compared to the sparse reconstruction algorithms based on ℓ_1 minimization. A performance comparison of the proposed method with other techniques, including ℓ_1 minimization and another recently developed scheme for sparse signal recovery, is provided in terms of achievable rates for a DSL line with impulse noise estimation and cancellation.**

***Index Terms*—Impulsive noise, DSL, Sparse signal reconstruction, and Compressive sensing.**

I. Introduction

One of the most severe problems encountered in Digital Subscriber Line (DSL) design is impulsive noise. As DSL technology works at extremely high SNR, additive white Gaussian noise (AWGN) is generally not a problem. Impulsive noise is a phenomenon that occurs rarely, but when it arises it may "erase" an entire OFDM symbol if no counter-measure is adopted. Since the impulsive noise in the time domain reflects into a large block of corrupted symbols in the frequency domain, coding able to handle large bursts of errors, or concatenation of conventional random-error oriented codes with a large interleaver are envisaged [1].

As an alternative, impulsive noise can be estimated and actively cancelled. A Gaussian/erasure channel model is considered in [2] to calculate the capacity of this channel. In [3] and [4], precoding and frequency algebraic interpolation techniques using Reed-Solomon coding and decoding are proposed but these approaches are very sensitive to background noise. In [5], impulsive noise is considered a sparse vector and compressive sensing (CS) based on convex relaxation methods using ℓ_1 minimization [6] is used for estimation from a small subset of frequency-domain observations. The drawbacks of using this method are 1) ℓ_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), and 3) It does not utilize structure of the sensing matrix which in our case is a submatrix of a DFT matrix.

Recently several low complex alternatives have been proposed in literature for sparse signal recovery that include algorithms based on belief propagation [7], Bayesian methods applied to compressive sensing [8], and iterative greedy approaches including orthogonal matching pursuit (OMP) [9], and fast Bayesian matching pursuit (FBMP) [10]. In this paper, we make use of the free guard band carriers to estimate and cancel impulsive 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. We compare the performance of the proposed method with other techniques, including ℓ_1 minimization [5] and FBMP [10], in terms of achievable rates for a DSL line with impulse noise estimation and cancellation.

II. Transmission Model

The time-domain complex baseband equivalent DSL channel is given by

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

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), $\mathbf{H}$ is a circulant matrix constructed from the channel impulse response, and $\mathbf{z}$ is the complex white Gaussian additive noise $\mathcal{CN}(\mathbf{0}, N_0\mathbf{I})$. We assume the impulsive noise process $\mathbf{e}$ to be Bernoulli-Gaussian, i.e. $e_k = \lambda_k g_k$, where λ_k are i.i.d. Bernoulli random variables, with $P(\lambda_k = 1) = p$, and g_k are i.i.d. Gaussian random variables $\sim \mathcal{CN}(0, I_0)$. We define the channel SNR as $\mathcal{E}_x/N_0$ and the impulse to noise ratio (INR) as I_0/N_0. We can represent (1) in the frequency domain as

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

where $\check{\mathbf{y}}$, $\check{\mathbf{x}}$, and $\check{\mathbf{z}}$ are the DFTs of $\mathbf{y}$, $\mathbf{x}$, and $\mathbf{z}$ respectively[1] and where $\mathbf{D} = \text{diag}(\check{\mathbf{h}})$ with $\check{\mathbf{h}} = \sqrt{n}\mathbf{F}_n\mathbf{h}$. Here $\mathbf{F}_n$ is the size n unitary DFT matrix with $\mathbf{F}_n(k, \ell) = \frac{1}{\sqrt{n}} e^{-j2\pi k\ell/n}$ with $k, \ell \in \{0, \ldots, n-1\}$.

[1] $\mathbf{x} = \mathbf{F}_n^{\mathsf{H}}\check{\mathbf{x}}$

The work of Tareq Y. Al-Naffouri and Ahmed A. Quadeer was supported by SABIC through an internally funded project from DSR, KFUPM (Project No. SB101006) and the work of Tareq Y. Al-Naffouri was also supported by the Fullbright Scholar Program.

978-1-4577-0595-3/11/$26.00 ©2011 IEEE

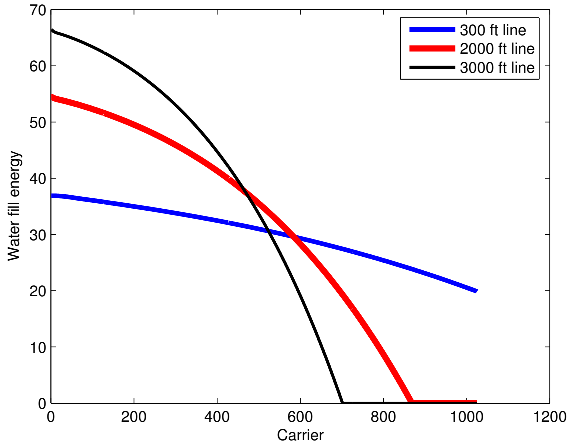


Fig. 1: Waterfill level at SNR = 30 dB for DSL transmission over cat5 coaxial cable having lengths (300ft, 2000ft, and 3000ft)

III. Problem Formulation

Consider the OFDM frequency domain channel model (2). We will use the sparse nature of $\mathbf{e}$ to estimate it and then remove it from the received signal. As in [3], [4], we will use carriers free of modulation symbols to estimate $\mathbf{e}$. Specifically, we will assume that these carriers belong to a continuous band. This is not unusual for in practical DSL transmission, many high frequency carriers are left unutilized due to high attenuation. Figure 1 shows that many of the carriers in the DSL become useless for longer distances.[2]

We construct the time domain transmit signal as $\mathbf{x} = \mathbf{F}_n^{\mathsf{H}}\mathbf{S}_x\check{\mathbf{d}}$, 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 columns of $\mathbf{S}_x$ index the subcarriers that are used for data transmission in the OFDM system. The remaining subcarriers are not used. Let $\mathbf{S}$ be a matrix with a single element equal to 1 per column, that span the orthogonal complement of the columns of $\mathbf{S}_x$. The frequency domain vector is thus given by

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

We shall estimate $\mathbf{e}$ from the projection into the orthogonal complement of the signal subspace. This is given by

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

where the observation vector $\mathbf{y}'$ is a projection of the n-dimensional impulsive noise onto a basis of dimension $n - m < n$ corrupted by the AWGN $\mathbf{z}'$ which is an i.i.d. Gaussian vector with variance N_0 per component, of length m. For later use, we shall denote the $m \times n$ projection matrix obtained by a row selection of $\mathbf{F}_n$ (according to $\mathbf{S}$) by $\mathbf{\Psi} = \mathbf{S}^{\mathsf{T}}\mathbf{F}_n$. When the support $\mathcal{I}$ of $\mathbf{e}$ is known, we can equivalently write (4) as

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

where $\mathbf{\Psi} = [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_n]$, and where $\mathbf{\Psi}_{\mathcal{I}}$ denotes the submatrix formed by columns $\{\boldsymbol{\psi}_j : j \in \mathcal{I}\}$, indexed by the support $\mathcal{I}$.

[2]In our illustrations and simulations, we assume that the last quarter band of carriers is free of data transmission. However, our approach applies for any continuous band of carriers.

IV. Optimum Estimation of $\mathbf{e}$

The MMSE estimate of $\mathbf{e}$ given the observation $\mathbf{y}'$ guarantees to minimize the covariance of the residual noise

$$\hat{\mathbf{e}}_{\text{MMSE}} = \mathbb{E}[\mathbf{e}|\mathbf{y}'] = \sum_{\mathcal{I}} p(\mathcal{I}|\mathbf{y}')\mathbb{E}[\mathbf{e}|\mathbf{y}', \mathcal{I}] \tag{6}$$

where the sum is carried out over the 2^n supports set $\mathcal{I}$ of $\mathbf{e}$ which could be computationally very complex for large n. We show in the following how to calculate the various terms in (6).

Calculating $\mathbb{E}[\mathbf{e}|\mathbf{y}', \mathcal{I}]$*:* Since $\mathbf{e}_{\mathcal{I}}$ is Gaussian, we can easily deduce from (4) that the MMSE estimate of the active elements of $\mathbf{e}$ is

$$\mathbf{e}_{\mathcal{I}} = I_0\mathbf{\Psi}_{\mathcal{I}}^{\mathsf{H}}\mathbf{\Sigma}_{\mathcal{I}}^{-1}\mathbf{y}' \tag{7}$$

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

Calculating $p(\mathcal{I}|\mathbf{y}')$*:* Using Bayes rule, we can write

$$p(\mathcal{I}|\mathbf{y}') = \frac{p(\mathbf{y}'|\mathcal{I})p(\mathcal{I})}{\sum_{\mathcal{I}} p(\mathbf{y}'|\mathcal{I})p(\mathcal{I})} \tag{8}$$

where $p(\mathcal{I}) = p^{|\mathcal{I}|}(1-p)^{n-|\mathcal{I}|}$. Moreover since $\mathbf{e}|\mathcal{I}$ is Gaussian, $\mathbf{y}'|\mathcal{I}$ is Gaussian with zero mean and covariance $\mathbf{\Sigma}_{\mathcal{I}}$ and we can write[3]

$$p(\mathbf{y}'|\mathcal{I}) = \frac{\exp\left(-\frac{1}{N_0}\|\mathbf{y}'\|^2_{\mathbf{\Sigma}_{\mathcal{I}}^{-1}}\right)}{\det(\mathbf{\Sigma}_{\mathcal{I}})} \tag{9}$$

up to an irrelevant constant multiplicative factor.

Evaluation over $\mathcal{I}$*:* Note that the expressions (6)-(9) are different for different values of $\mathcal{I}$ and the two summations in (6) and (8) need to be evaluated over all possible 2^n sets. Instead of this exhaustive search approach, we can limit the MMSE evaluation over the most probable support of $\mathbf{e}$. There are two techniques to limit the search space.

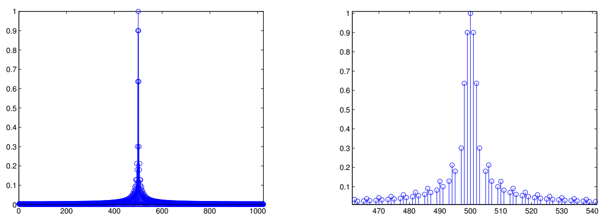
CS based on convex relaxation: Starting from (4), we can use the standard convex relaxation tools [6] to find the most probable support of the sparse vector $\mathbf{e}$. This approach does not make use of any a priori statistical information. Moreover, given the highly structured nature of Ψ,[4] this method does not perform as well as when the free carriers are chosen randomly from the whole band.

FBMP: A fast Bayesian recursive algorithm is presented in [10] that finds the dominant support and the MMSE estimate of the sparse vector jointly. It uses a greedy tree search over all the combinations in pursuit of the dominant supports. The algorithm starts with zero active elements support. At each step, an active element is added that maximizes the MAP-Gaussian metric. This procedure is repeated till we reach P active elements in a branch. The number of branches in the tree search is controlled by a parameter D which presents a tradeoff between performance and complexity.[5]

[3]$\|\mathbf{b}\|^2_{\mathbf{A}} \triangleq \mathbf{b}^{\mathsf{H}}\mathbf{A}\mathbf{b}$.

[4]This is due to the fact that the carriers are chosen from the guard band.

[5]Contrary to standard convex relaxation techniques, this algorithm is able to make use of the a priori statistical information and reduce complexity by employing a recursive implementation.



(a) Normal (b) Zoomed

Fig. 2: The 500^{th} column has high correlation with its neighbours only

V. Finding Dominant Support using Structure

Neither of the two 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 (6). Specifically, let $m = \frac{n}{\ell}$ be the fraction of unused carriers. In this case, the observation vector $\mathbf{y}'$ is given by (see (4))

$$\mathbf{y}' = \mathbf{\Psi}\mathbf{e} + \mathbf{z}' \tag{10}$$

where the $\mathbf{\Psi}$ is the $m \times n$ projection matrix. Note that columns $j \in 0, \ell, 2\ell, ..., m\ell$, of $\mathbf{\Psi}$ correspond to (a scaled version of) the DFT matrix $\mathbf{F}_m$. As such, the columns of this matrix span the column space of $\mathbf{\Psi}$. In fact, spanning happens in a special way in that any column of $\mathbf{\Psi}$ that does not belong to $\mathbf{F}_m$ has most of its energy along its left and right basis vectors. Thus, the column vector $\mathbf{\Psi}_{j(1)+i}$ $(i = 1, 2, ..., \ell - 1)$, has its dominant components along the basis vectors $\mathbf{\Psi}_{j(1)}$ and $\mathbf{\Psi}_{j(2)}$

Mathematically, the correlation between two columns can be shown to be

$$\mathbf{\Psi}_k^{\mathsf{H}}\mathbf{\Psi}_{k'} = \begin{cases} 1, & (k = k') \\ \left|\frac{\sin\left(\pi(k-k')m/n\right)}{m\sin\left(\pi(k-k')/n\right)}\right|, & (k \neq k') \end{cases} \tag{11}$$

which is a function of the difference $(k - k') \mod n$, so it suffices to consider correlation of one column with the remaining ones. Figures 2a and 2b demonstrate this correlation for the case when $n = 1024$, $\ell = 4$, and $m = \frac{n}{\ell} = 256$.

Let $\Gamma = \frac{\sin(\pi\varphi m/n)}{m\sin(\pi\varphi/n)}$ where $\varphi = (\ell - \ell')$, then

$$\Gamma = \left|\frac{\sin(mx)}{m\sin(x)}\right| \lesssim \left|\frac{\sin(mx)}{mx}\right| \tag{12}$$

for small values of x where $x = \frac{\pi\varphi}{n}$

This suggests that we can get a first guess on the location of impulsive noise by projecting $\mathbf{y}'$ on $\mathbf{F}_m$ to get $\mathbf{y}'' = \mathbf{F}_m\mathbf{y}'$. The larger values of $\mathbf{y}''$ point to the neighbourhood of where these impulses are located. Thus, if the largest value of $\mathbf{y}''$ is say the $17th$ one, it indicates that one impulse belongs to the 17ℓ column of $\mathbf{\Psi}$ or its neighbours, namely 17ℓ, $17\ell \pm 1$, $17\ell \pm 2$, ... $17\ell \pm (\ell - 1)$. Thus, the largest values of $\mathbf{y}''$ point to the clusters where the support of the impulsive noise might belong.

Adjacent clusters can be grouped to form larger clusters. Based on the dominant support $\mathcal{I}$ of $\mathbf{e}$, we can reduce $\mathbf{\Psi}$ to $\mathbf{\Psi}_{\mathcal{I}}$ which can be written in block form as $\mathbf{\Psi}_{\mathcal{I}} = [\mathbf{\Psi}_1 \ \mathbf{\Psi}_2 \ \cdots \ \mathbf{\Psi}_P]$ where P is the maximum number of clusters to which the support of $\mathbf{e}$ is expected to belong.[6] As is evident from Figure 2, these clusters are almost orthogonal to each other and thus we can deal with them separately. Specifically, it turns out that the overall likelihood is the product of likelihoods of the individual clusters. We will prove this for the two clusters case (i.e. $P = 2$). The proof for the general case can be derived similarly. Now to calculate the likelihood (9), we need to calculate the inverse and determinant of $\mathbf{\Sigma}_{\mathcal{I}}$. Now for simplicity, assume that $\mathbf{y}''$ points us to two clusters so that $\mathbf{\Psi}_{\mathcal{I}} = [\Omega_1 \ \Omega_2]$. Thus, using the matrix inversion lemma, we can write $\mathbf{\Sigma}_{\mathcal{I}}^{-1}$ as

$$\begin{aligned} \mathbf{\Sigma}_{\mathcal{I}}^{-1} &= (\mathbf{I} + \frac{I_0}{N_0}\mathbf{\Psi}_{\mathcal{I}}\mathbf{\Psi}_{\mathcal{I}}^{\mathsf{H}})^{-1} &(13)\\ &= \mathbf{\Theta}^{-1} - \frac{I_0}{N_0}\mathbf{\Theta}^{-1}\Omega_2(\mathbf{I} + \frac{I_0}{N_0}\Omega_2^{\mathsf{H}}\mathbf{\Theta}^{-1}\Omega_2)^{-1}\Omega_2^{\mathsf{H}}\mathbf{\Theta}^{-1} &(14) \end{aligned}$$

where $\mathbf{\Theta} = \mathbf{I} + \frac{I_0}{N_0}\Omega_1\Omega_1^{\mathsf{H}}$. As Ω_1 and Ω_2 are orthogonal (i.e. $\Omega_1\Omega_2^{\mathsf{H}} = \Omega_2\Omega_1^{\mathsf{H}} = \mathcal{O}(\frac{1}{m})$), (14) becomes

$$\begin{aligned} \mathbf{\Sigma}_{\mathcal{I}}^{-1} &= \mathbf{I} - \frac{I_0}{N_0}\Omega_1(\mathbf{I} + \frac{I_0}{N_0}\Omega_1^{\mathsf{H}}\Omega_1)^{-1}\Omega_1^{\mathsf{H}} \\ &\quad - \frac{I_0}{N_0}\Omega_2(\mathbf{I} + \frac{I_0}{N_0}\Omega_2^{\mathsf{H}}\Omega_2)^{-1}\Omega_2^{\mathsf{H}} \\ &= -\mathbf{I} + \left(\mathbf{I} - \frac{I_0}{N_0}\Omega_1(\mathbf{I} + \frac{I_0}{N_0}\Omega_1^{\mathsf{H}}\Omega_1)^{-1}\Omega_1^{\mathsf{H}}\right) \\ &\quad + \left(\mathbf{I} - \frac{I_0}{N_0}\Omega_2(\mathbf{I} + \frac{I_0}{N_0}\Omega_2^{\mathsf{H}}\Omega_2)^{-1}\Omega_2^{\mathsf{H}}\right) \\ &= -\mathbf{I} + \left(\mathbf{I} + \frac{I_0}{N_0}\Omega_1\Omega_1^{\mathsf{H}}\right)^{-1} + \left(\mathbf{I} + \frac{I_0}{N_0}\Omega_2\Omega_2^{\mathsf{H}}\right)^{-1} &(15) \end{aligned}$$

where (15) is valid up to an error term of order $\mathcal{O}(\frac{1}{m})$. As such, we can write $\exp(-\frac{1}{N_0}\|\mathbf{y}'\|^2_{\mathbf{\Sigma}_{\mathcal{I}}^{-1}})$ as in (16) where $\mathbf{\Sigma}_{\Omega_1} = \mathbf{I} + \frac{I_0}{N_0}\Omega_1\Omega_1^{\mathsf{H}}$ and $\mathbf{\Sigma}_{\Omega_2} = \mathbf{I} + \frac{I_0}{N_0}\Omega_2\Omega_2^{\mathsf{H}}$. Using a similar reasoning, we can show that the determinant decomposes in a similar way,

$$\det(\mathbf{\Sigma}_{\mathcal{I}}) = \det(\mathbf{\Sigma}_{\Omega_1})\det(\mathbf{\Sigma}_{\Omega_2}) \tag{17}$$

Thus if we denote the MAP Gaussian metric at a particular support $\mathcal{I}$ by $\mathcal{L}_{\mathcal{I}}$ and incorporate the results in (16) and (17), (9) becomes

$$\begin{aligned} \mathcal{L}_{\mathcal{I}} &= p^{|\mathcal{I}|}(1-p)^{n-|\mathcal{I}|}\exp\left(\frac{1}{N_0}\|\mathbf{y}'\|^2\right) \\ &\quad \frac{\exp\left(-\frac{1}{N_0}\|\mathbf{y}'\|^2_{\mathbf{\Sigma}_{\Omega_1}^{-1}}\right)}{\det(\mathbf{\Sigma}_{\Omega_1})} \frac{\exp\left(-\frac{1}{N_0}\|\mathbf{y}'\|^2_{\mathbf{\Sigma}_{\Omega_1}^{-1}}\right)}{\det(\mathbf{\Sigma}_{\Omega_2})} \\ &= p^{|\mathcal{I}|}(1-p)^{n-|\mathcal{I}|}\exp\left(\frac{1}{N_0}\|\mathbf{y}'\|^2\right)\mathcal{L}_{\Omega_1}\mathcal{L}_{\Omega_2} &(18) \end{aligned}$$

Thus, the general form of (18) can be written as

$$\mathcal{L}_{\mathcal{I}} = p^{|\mathcal{I}|}(1-p)^{n-|\mathcal{I}|}\exp\left(\frac{P-1}{N_0}\|\mathbf{y}'\|^2\right)\prod_{m=0}^{P}\mathcal{L}_m \tag{19}$$

where P is the total number of clusters considered.

A. Orthogonal Clustering Algorithm

From the discussion, we can summarize our algorithm as follows.

Start by taking the m point DFT of $\mathbf{y}'$. The dominant values of the resulting vector $\mathbf{y}''$ point to the most probable support of $\mathbf{e}$. Group the resulting support regions into disjoint and

[6] $P = \lceil \text{erfc}^{-1}(10^{-2})\sqrt{2np(1-p)} + np \rceil$ similar to [10].

$$\exp(-\frac{1}{N_0}\mathbf{y}'^{\mathsf{H}}\mathbf{\Sigma}_{\mathcal{J}}^{-1}\mathbf{y}') = \exp\left(\frac{1}{N_0}\|\mathbf{y}'\|^2 - \frac{1}{N_0}\|\mathbf{y}'\|^2_{\mathbf{\Sigma}_{\Omega_1}^{-1}} - \frac{1}{N_0}\|\mathbf{y}'\|^2_{\mathbf{\Sigma}_{\Omega_2}^{-1}}\right)$$
$$= \exp\left(\frac{1}{N_0}\|\mathbf{y}'\|^2\right)\exp\left(-\frac{1}{N_0}\|\mathbf{y}'\|^2_{\mathbf{\Sigma}_{\Omega_1}^{-1}}\right)\exp\left(-\frac{1}{N_0}\|\mathbf{y}'\|^2_{\mathbf{\Sigma}_{\Omega_2}^{-1}}\right) \quad (16)$$

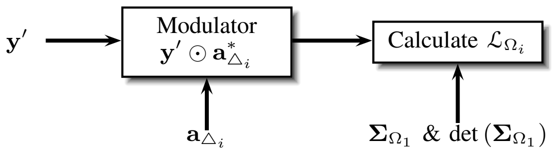


Fig. 3: Block diagram of the reduced complexity algorithm to calculate the likelihood of the ith cluster

separate clusters $\Omega_1, \Omega_2, \cdots, \Omega_P$ and let L be the maximum cluster length. For each cluster, proceed as follows.

1) Find the support for the case of $1, 2, \cdots, s$ impulses in each cluster by maximizing (9).
2) For each cluster, find the most probable support by selecting the maximum one from the s supports calculated in step 1, save $\mathcal{L}_{\Omega_i}$, and evaluate $\mathbb{E}[\mathbf{e}|\mathbf{y}', \mathcal{J}]$ using (7).
3) Evaluate $\mathcal{L}_{\mathcal{J}}$ (19) for all the 2^{P-1} dominant supports using $\mathcal{L}_{\Omega_i}$ saved in the previous step.
4) Evaluate $p(\mathcal{J}|\mathbf{y}')$ by averaging over all the dominant supports and obtain $\mathbf{e}_{\mathrm{MMSE}}$.

Summarizing, the $\frac{n}{4}$ DFT of $\mathbf{y}'$ points us to the candidate clusters whose likelihoods can be calculated independently. Now, since the clusters have relatively small spans, the likelihood of an impulse or a number of impulses appearing in each cluster can be calculated independently and combined with the likelihoods of the other clusters to calculate the overall likelihood with low complexity.

B. Computational Complexity

As evident from the previous section, calculating the likelihood $\mathcal{L}$ can be done in a divide and conquer manner by calculating the likelihood from each cluster independently. That in turn requires calculating the inverse and determinant of $\mathbf{\Sigma}_{\Omega_i}$ for each cluster. In this subsection we will show that it is enough to calculate the quantities for one cluster (say the first one) and the corresponding quantities can be deduced for other clusters. To this end, let $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_L$ denote the sensing columns associated with the first cluster. Then, it is easy to see that the corresponding columns for the ith cluster of equal length that is $\triangle_i$ columns away are $\mathbf{a}_1 \odot \mathbf{a}_{\triangle_i}, \mathbf{a}_2 \odot \mathbf{a}_{\triangle_i}, \cdots, \mathbf{a}_L \odot \mathbf{a}_{\triangle_i}$, where $\odot$ denotes element-by-element multiplication and

$$\mathbf{a}_{\triangle_i} = \left[\begin{array}{ccc} \exp\left(-\frac{\jmath 2\pi}{n}\frac{3n}{4}\triangle_i\right)^{\mathsf{T}} & \cdots & \exp\left(-\frac{\jmath 2\pi}{n}n\triangle_i\right)^{\mathsf{T}} \end{array}\right]^{\mathsf{T}}$$

Now assume that we calculate the inverse $\mathbf{\Sigma}_{\Omega_1}^{-1}$ and determinant $\det(\mathbf{\Sigma}_{\Omega_1})$ for a set of columns Ω_1 in the first cluster and let Ω_i denote the same set of columns chosen in the ith cluster. Then it is easy to show that

$$\det(\mathbf{\Sigma}_{\Omega_i}) = \det(\mathbf{\Sigma}_{\Omega_1}) \quad \text{and} \quad \|\mathbf{y}'\|^2_{\mathbf{\Sigma}_{\Omega_i}^{-1}} = \|\mathbf{y}' \odot \mathbf{a}^*_{\triangle_i}\|^2_{\mathbf{\Sigma}_{\Omega_1}^{-1}} \quad (20)$$

In other words, to calculate the likelihood $\mathcal{L}_{\Omega_i}$, we can use the same quantities ($\det(\mathbf{\Sigma}_{\Omega_1})$ & $\mathbf{\Sigma}_{\Omega_1}^{-1}$) involved in calculating $\mathcal{L}_{\Omega_1}$ and we only need to modulate $\mathbf{y}'$ by $\mathbf{a}^*_{\triangle_i}$ before calculating the likelihood $\mathcal{L}_{\Omega_i}$. See Figure 4 for a block diagram for this calculation. As such, the complexity of proposed algorithm comes out to be $\mathcal{O}(mPL)$ where P and L correspond to the maximum number of clusters and maximum length of a cluster respectively.[7] The complexity of CS based on convex relaxation using ℓ_1 minimization is $\mathcal{O}(m^2n^{3/2})$ [11] and that of FBMP is $\mathcal{O}(mnPD)$ [10].

C. Approximate Residual Noise Covariance using the Orthogonality of Clusters

Let $\mathbf{e} = \mathbf{S}_e\mathbf{u}_e$, where $\mathbf{S}_e$ is a 0-1 selection matrix of dimension $n \times r$, with ones corresponding to the true support $\mathcal{J}$ of $\mathbf{e}$, of cardinality r, and $\mathbf{u}_e$ denoting the vector of dimension r that collects only the non-zero elements of $\mathbf{e}$. Thus the observation $\mathbf{y}'$ is given by

$$\mathbf{y}' = \mathbf{S}^{\mathsf{T}}\mathbf{F}_n\mathbf{e} + \mathbf{z}' \quad (21)$$
$$= \underline{\mathbf{F}}\mathbf{S}_e\mathbf{u}_e + \mathbf{z}' = \mathbf{\Phi}_e\mathbf{u}_e + \mathbf{z}' \quad (22)$$

If the receiver has perfect knowledge of the support $\mathcal{J}$ of $\mathbf{e}$, the covariance of the MMSE estimate of $\mathbf{u}_e$ is given by

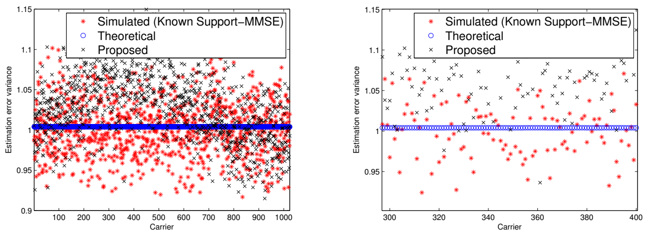
$$\mathbf{C}_{\mathbf{u}}|\mathcal{J} = \mathbb{E}[\tilde{\mathbf{u}}_e\tilde{\mathbf{u}}_e^{\mathsf{H}}|\mathcal{J}] = \left(\frac{1}{I_0}\mathbf{I} + \frac{1}{N_0}\mathbf{\Phi}_e^{\mathsf{H}}\mathbf{\Phi}_e\right)^{-1} \quad (23)$$

To get the overall estimation error, we need to average over $\mathcal{J}$. To do so, we pursue the clustering approach and divide the support into $\frac{n}{\ell}$ clusters of size ℓ each. Thus, we write $\underline{\mathbf{F}} = [\mathbf{\Psi}_1\ \mathbf{\Psi}_2\ \cdots\ \mathbf{\Psi}_{\frac{n}{\ell}}]$. Using our semi-orthogonal approach, only neighbouring clusters will interact, and it is easy to show that $\mathbf{C}_{\mathbf{u}}|\mathcal{J}$ will take the form of a $P \times P$ block matrix with only the blocks (i, j) non-zero, for $|i - j| <= 1$ (when the support $\mathcal{J}$ intersects with the ith block). Moreover, it is easy to show that the expectation $\mathbb{E}\left(\frac{1}{I_0}\mathbf{I} + \frac{1}{N_0}\Omega_i^{\mathsf{H}}\Omega_j\right)^{-1}$ is invariant over clusters (i.e. it is enough to calculate it for $i = 1$). Denote these expectations by $\mathbf{A}_1$ and $\mathbf{A}_{12}$ respectively. Then, we can write the residual estimation error as

$$\mathbb{E}[\tilde{\mathbf{e}}\tilde{\mathbf{e}}^{\mathsf{H}}] = \mathbf{A}_1 \otimes \mathbf{I}_{\frac{n}{\ell}} + \mathbf{A}_{12} \otimes \bar{\mathbf{I}}_{\frac{n}{\ell}} + \mathbf{A}_{12}^{\mathsf{H}} \otimes \bar{\mathbf{I}}^{\mathsf{H}}_{\frac{n}{\ell}} \quad (24)$$

where $\bar{\mathbf{I}}_{\frac{n}{\ell}}$ is a matrix with ones on only the super-diagonal. In Figure 4, we compare the residual noise covariance for probability of impulse $p = 10^{-3}$, the case when support is perfectly known (MMSE used for impulse amplitudes estimation), and the orthogonal clustering algorithm discussed in Section V. A cluster of maximum length $L = 16$ is considered. The expectation involved is calculated by evaluating $p^{\mathcal{J}}(1-p)^{L-\mathcal{J}}\left(\frac{1}{I_0}\mathbf{I} + \frac{1}{N_0}\mathbf{\Psi}_{\mathcal{J}}^{\mathsf{H}}\mathbf{\Psi}_{\mathcal{J}}\right)^{-1}$ over all the possible combinations.

[7]Typical values of P and L are 8 and 32 respectively for $n = 1024$, $m = 256$, and $p = 3 \times 10^{-3}$.



(a) Normal (b) Zoomed

Fig. 4: Comparison of the residual noise covariance for probability of impulse $p = 10^{-3}$

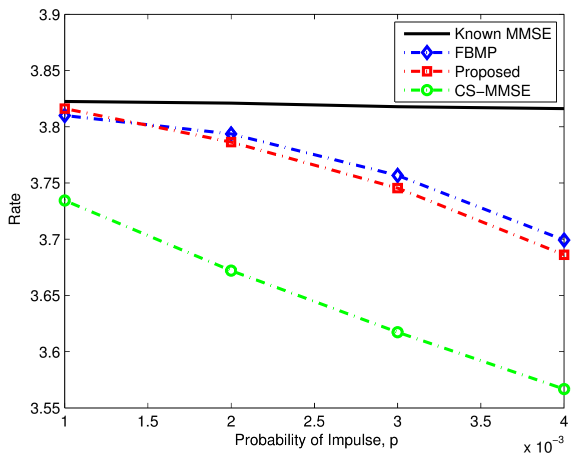


Fig. 5: Comparison of algorithms using complete quarter band

VI. Simulation Results

As we are interested in transmission over DSL, we assume the channel impulse response to be constant and known forever. The parameters used are n = 1024 tones and m = $\frac{n}{4}$ = 256 null carriers at the end of the transmission band, with SNR = 20 dB and INR = 35 dB. The range of number of impulses (K) is $0 \leq K \leq 10$ and all the algorithms are run for 256 Monte Carlo iterations at each value of K. The performance of the proposed algorithm is compared with [5] that uses CS based on convex relaxation to find the impulse support followed by MMSE estimation to recover the impulse amplitudes and FBMP [10]. The upper bound (benchmark) is given by the case when support is perfectly known and MMSE is used for estimation of impulse amplitudes. The algorithms are compared in terms of the achievable rates evaluated similar to [5]. The simulation results are shown in Figure 5. It can be seen that both the proposed and FBMP algorithms perform better than ℓ_1 reconstruction, reported in [5]. The performance of FBMP is slightly better than the proposed algorithm but this gain in performance is due to greater complexity as shown in Figure 6 that compares the mean runtime of both the algorithms.

VII. Conclusion

In this paper, we propose a new method for impulsive noise estimation and cancellation in DSL. Instead of using compressive sensing or matching pursuit algorithms for sparse reconstruction, the proposed paper utilizes structure of the sensing matrix and a priori information of the impulsive noise distribution jointly resulting in a fast and efficient algorithm. Simulation results demonstrate the good performance of the proposed algorithm.

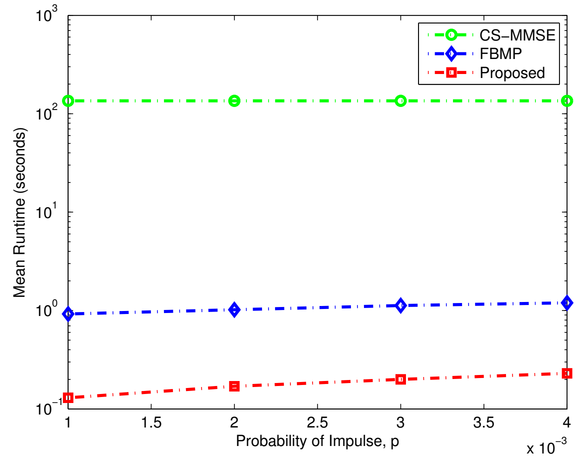


Fig. 6: Comparison of mean runtime of proposed algorithm, FBMP, and CS-MMSE

Acknowledgement

The authors would like to thank King Fahd University of Petroleum & Minerals for supporting this work.

References

[1] D. Toumpakaris, Wei Yu, J. M. Cioffi, D. Gardan, M. Ouzzif, "A simple byte-erasure method for improved impulse immunity in DSL," *IEEE Int. Conf. on Comm.*, vol. 4, pp. 2426- 2430, May 2003.

[2] A. Tulino, S. Verdu, G. Caire and S. Shamai, "The Gaussian Erasure Channel," *IEEE Int. Symp. on Inform. Theory,* France, June 24-29, 2007.

[3] F. Abdelkefi, P. Duhamel, and F. Alberge, "Impulsive Noise Cancellation in Multicarrier Transmission," *IEEE Trans. on Comm.* vol. 53, no. 1, Jan. 2005.

[4] 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.

[5] 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.

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

[7] 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.

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

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

[10] P. Schniter, L. C. Potter, and J. Ziniel, "Fast bayesian matching pursuit," *Inform. Theory and Appl. Workshop*, pp. 326-333, 2008.

[11] I. E. Nesterov, A. Nemirovskii, and Y. Nesterov, *Interior-Point Polynomial Algorithms in Convex Programming.* SIAM, 1994.