IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 14, 2019, accepted May 19, 2019, date of publication June 4, 2019, date of current version June 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2920724

Joint Channel Estimation and Impulsive Noise Mitigation Method for OFDM Systems Using Sparse Bayesian Learning

XINRONG LV[1,2], YOUMING LI[1], YONGQING WU[3], XIAOLI WANG[1], AND HUI LIANG[4]

[1]Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China
[2]School of Intelligent Electronics, Zhejiang Business Technology Institute, Ningbo 315211, China
[3]Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China
[4]Ningbo Sanxing Electric Co., Ltd., Ningbo 315211, China

Corresponding author: Youming Li (liyouming@nbu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61571250 and Grant 61531018, in part by the Key Laboratory of Mobile Internet Application Technology of Zhejiang Province, in part by the Ningbo Natural Science Foundation under Grant 2015 A610121, in part by the Foundation of Ningbo Jiangbei District Science and Technology Bureau under Grant 011-201801B20013, in part by the K. C. Wong Magna Fund of Ningbo University, and in part by the ZBTI Science Foundation under Grant 2017Z01.

ABSTRACT The impulsive noise can deteriorate sharply the performance of orthogonal frequency division multiplexing (OFDM) systems. In this paper, we propose a novel joint channel impulse response estimation and impulsive noise mitigation algorithm based on compressed sensing theory. In this algorithm, both the channel impulse response and the impulsive noise are treated as a joint sparse vector. Then, the sparse Bayesian learning framework is adopted to jointly estimate the channel impulse response, the impulsive noise, and the data symbols, in which the data symbols are regarded as unknown parameters. The Cramér–Rao Bound is derived for the benchmark. Unlike the previous impulsive noise mitigation methods, the proposed algorithm utilizes all subcarriers without any *a priori* information of the channel and impulsive noise. The simulation results show that the proposed algorithm achieves significant performance improvement on the channel estimation and bit error rate performance.

INDEX TERMS Orthogonal frequency division multiplexing (OFDM), channel estimation, impulsive noise mitigation, sparse Bayesian learning (SBL), compressed sensing.

I. INTRODUCTION

In several applications of wireless communication technology(e.g., vehicular networks [1], smart grid [2], and shallow sea underwater networks [3]), the transmission of data signals will be severely deteriorated by the impulsive noise (IN). The sources of impulsive noise are diverse, such as ignition noise in automobiles [4], switches for electrical equipments [5], various maritime operations [6], and so on. Compared to additive white Gaussian noise (AWGN), the impulsive noise arises randomly with short duration and high power impulses.

Orthogonal frequency division multiplexing (OFDM) technology has been widely adopted in most modern wireless communication standards [7]. In conventional OFDM receivers, the time-domain received signal is converted into the frequency domain through a discrete Fourier transform (DFT), after which each subcarrier is demodulated independently [8]. Such tone-by-tone demodulation achieves optimal maximum likelihood detection in AWGN and perfect channel state information [9]. When the impulsive noise is present, however, the corresponding frequency-domain noise samples will be highly dependent, and tone-by-tone demodulation is no longer feasible since the complexity of performing joint-detection at the receiver increases exponentially with the number of subcarriers [10].

Efficient impulsive noise suppression method plays an important role in promoting the performance of OFDM communication systems in the presence of additive impulsive noise. Since the amplitude of the impulsive noise is usually much higher than the background noise, it is possible to determine the presence of impulsive noise by setting a threshold and then to design a memoryless nonlinear preprocessor (e.g., clipping, blanking, or a combination thereof) to

The associate editor coordinating the review of this manuscript and approving it for publication was Wen Chen.

eliminate the effect of impulse noise [11]–[14]. By setting multiple thresholds, the nonlinear estimator for impulsive noise can improve the signal-to-noise power ratio (SNR) at the receiver [15]. However, these methods require the noise priori statistics to obtain the optimal threshold but suffer from performance degradation when the priori information mismatchs the time-varying noise statistics, which is not easy to acquire in reality as well. Moreover, these nonlinear preprocessor may destroy orthogonality among OFDM subcarriers, thus resulting in intercarrier interference in the frequency domain [16].

Recently, there has been growing interest in developing compressed sensing (CS) based impulsive noise mitigation methods that exploit the time-domain sparsity of impulsive noise [4], [6], [17]–[21]. These methods all make use of the information of null tones (i.e., tones that do not carry data or pilots) of the received OFDM symbol to estimate the IN sample and then subtract it from the received signal. Furthermore, some of them have been extended for detecting bursty (i.e., block sparse) impulsive noise [20], [21] by using structured compressed sensing theory [22]. Although these methods show obvious advantages over those based on nonlinear preprocessor, the common drawback of these algorithms is that their performances are mostly limited by the number of null tones. It is worth pointing out that these approaches also assume that the channel state information is already estimated perfectly before the impulsive noise removal and do not consider the severe impact of impulsive noise on the channel estimation [23].

The performance of the IN estimator can be improved by increasing the number of null tones. However having more null tones means reduced throughput. When the number of null tones is limited, it is desirable to exploit information available in all tones to improve the estimation performance of the impulsive noise. The difficulty for exploiting all tones, however, is how to simultaneously estimate the channel and impulsive noise. An approach for jointly estimating channel and IN is proposed in [24], but it requires that there is no overlap between the support of impulsive noise and channel impulse response. In [25], an iterative channel estimation and impulsive noise mitigation algorithm is proposed on the assumption that the length of channel impulse response is known in advance and that the channel is static for several OFDM symbols. In [26], generalized approximate message passing (GAMP) [27] has been used to jointly estimate the channel taps, the impulse noise samples, symbols, and the unknown bits. This method requires the acquisition of a priori information of the channel and impulsive noise and does not offer rigorous convergence although it is lower in computational complexity [28], [29]. By assuming that the impulsive noise parameter distributions are known at the receiver, a joint channel estimation and data decoding algorithm is developed [30]. By exploiting the sparsity of both them, the orthogonal matching pursuit(OMP) is adopted for joint channel and impulsive noise estimation in underwater acoustic OFDM systems [31]. This algorithm needs to collect the number and position of IN samples by applying a blanking operation.

In this paper, we propose two novel algorithms based on Sparse Bayesian Learning (SBL) framework [32], [33]to jointly estimate both the channel impulse response and impulsive noise by exploiting the sparsity of both them. Our algorithms can also be categorized as an extension of the method proposed in [34]. The first proposed method uses the pilot subcarriers to jointly estimate the channel impulse response and impulsive noise. Once the channel and IN are estimated, the IN is then removed from the received signal and the channel is transformed into the frequency domain followed with the channel equalization. In the second proposed algorithm, we utilize both the data and pilot subcarriers to promote the joint estimation performance of the channel impulse response and impulsive noise. Compared with the algorithms which treat the channel estimation and IN mitigation independently, our proposed joint estimation algorithms can lead to a significant improvement in the Mean Square Error (MSE) of channel estimation. For impulsive noise mitigation, our method using all subcarriers has a smaller Mean Square Error (MSE) of IN estimation than existing impulsive noise mitigation algorithms using only the null subcarriers.

The contributions of this paper are as follows:

- We treat the unknown data symbols as the hyperparameters and develop an iterative technique based on the Expectation Maximization (EM) algorithm for joint channel estimation, IN estimation, and data detection. Our algorithm can efficiently recover a sparse vector even when the measurement matrix is partially unknown due to the presence of unknown data symbols.
- Being different from many CS based IN estimation methods which use only the null subcarriers, our proposed method can exploit all subcarriers to improve the IN estimation performance. Our methods need less null subcarriers and can promote the spectrum efficiency. Apart from the assumption that the channel impulse response and impulsive noise samples are all sparse, our proposed methods do not require other priori information.
- We derive the closed form Bayesian Cramér-Rao Bound(BCRB) of channel and impulsive noise estimation.

The rest of the paper is organized as follows. The system model and IN model are presented in Section II. In Section III, the proposed joint channel estimation and impulsive noise mitigation algorithms are discussed. The Cramér-Rao Lower Bound(CRLB) is derived in section IV. Numerical simulation results are shown in Section V to verify the performance of the proposed algorithms. Finally we draw conclusions in Section VI.

II. SYSTEM MODEL

We consider an OFDM system with N subcarriers where M subcarriers are used for carrying pilot and $N-M$ subcarriers are used for sending data. In some case, we also consider the system which has null subcarriers and then these subcarriers can also be considered as pilot subcarriers with setting pilot symbols as zero. At the transmitter, data symbols which are mapped from the source bits and known pilot symbols are joined as the frequency-domain OFDM symbol $\boldsymbol{x} = (x_0, x_1, \ldots, x_{N-1})^T$. The OFDM modulator, as an inverse discrete Fourier transformation (IDFT), converts the frequency-domain OFDM symbols into the time-domain OFDM signals to which the cyclic prefix (CP) is prepended before feeding into the wireless channel. Assuming that inter-symbol interference is avoided by simply discarding the cyclic prefix at the receiver, the received time-domain signal is expressed as

$$\boldsymbol{r} = \boldsymbol{H}\boldsymbol{F}^*\boldsymbol{x} + \boldsymbol{u} \tag{1}$$

where $\boldsymbol{H}$ is a $N \times N$ circulant matrix whose first column is formed by the zero-padded channel impulse response vector $\boldsymbol{h} = (h_0, h_1, \ldots, h_{L-1})^T$, L is the length of channel impulse response. $\boldsymbol{F}$ is the unitary N-point discrete Fourier transform (DFT) matrix with (m, n) element $[\boldsymbol{F}]_{m,n} = \frac{1}{\sqrt{N}} e^{-j2\pi mn/N}$ with $m, n \in \{0, 1, \ldots, N-1\}$ and $\boldsymbol{F}^*$is its conjugate transpose. $\boldsymbol{u} = \boldsymbol{i} + \boldsymbol{g}$ is the additive noise term which includes $\boldsymbol{i}$, denoting impulsive noise component, and $\boldsymbol{g}$, denoting AWGN component.

After the OFDM demodulator implemented by DFT, the resulting frequency-domain symbol becomes

$$\boldsymbol{y} = \boldsymbol{F}\boldsymbol{r} = \boldsymbol{F}\boldsymbol{H}\boldsymbol{F}^*\boldsymbol{x} + \boldsymbol{F}\boldsymbol{i} + \boldsymbol{F}\boldsymbol{g} = \boldsymbol{\Lambda}\boldsymbol{x} + \boldsymbol{F}\boldsymbol{i} + \boldsymbol{n} \tag{2}$$

where $\boldsymbol{\Lambda} \triangleq diag(\check{\boldsymbol{h}})$ is a diagonal matrix with the channel frequency response $\check{\boldsymbol{h}}$ as its diagonal elements. The channel frequency response $\check{\boldsymbol{h}}$ is the DFT of the channel impulse response $\boldsymbol{h}$, namely $\check{\boldsymbol{h}} = \sqrt{N}\boldsymbol{F}_L\boldsymbol{h}$, where $\boldsymbol{F}_L \in \mathbb{C}^{N\times L}$ is the submatrix selected from the first L columns of matrix $\boldsymbol{F}$. $\boldsymbol{n}$ is the frequency-domain background noise vector which is still AWGN since $\boldsymbol{F}$ is unitary matrix.

Gaussian-Mixture(GM) [35], Middleton Class A (MCA) [36], and Symmetric alpha stable [37] are the three IN models most widely adopted in the literature. In this work, we use GM model to simulate the impulsive noise. In particular, a K-component GM is accepted for performance analysis. The probability density function(pdf) of a time-domain impulsive noise sample $\boldsymbol{u}$ is expressed as

$$p(\boldsymbol{u}) = \prod_{i=1}^{N} \sum_{k=1}^{K} \pi_k f_k(u_i) \tag{3}$$

where $f_k(u_i) \sim \mathcal{CN}(u_i; 0, \gamma_k)$denotes a complex Gaussian pdf with zero mean and variance γ_k, π_k is the mixing probability with$\sum_{k=1}^{K} \pi_k = 1$. The impulsive noise samples are assumed to be independent and identically distributed.

III. PROPOSED APPROACHES

From the equation (2), the received OFDM symbol can also be mathematically represented as

$$\begin{aligned} \boldsymbol{y} &= diag(\check{\boldsymbol{h}})\boldsymbol{x} + \boldsymbol{F}\boldsymbol{i} + \boldsymbol{n} \\ &= diag(\boldsymbol{x})\check{\boldsymbol{h}} + \boldsymbol{F}\boldsymbol{i} + \boldsymbol{n} \\ &= \sqrt{N}\boldsymbol{X}\boldsymbol{F}_L\boldsymbol{h} + \boldsymbol{F}\boldsymbol{i} + \boldsymbol{n} \end{aligned} \tag{4}$$

where $\boldsymbol{X} \triangleq diag(\boldsymbol{x})$ is a diagonal matrix with the frequency-domain symbol $\boldsymbol{x}$ as its diagonal elements.

In wireless communications, the discrete-time channel impulse response(CIR)$\boldsymbol{h}$ of length L comprising S resolvable propagation paths can be modeled as [34], [38]:

$$h_l = \sum_{s=1}^{S} \alpha_s \delta[l - \tau_s], \; 0 \le l \le L-1 \tag{5}$$

where α_s and τ_s denote the path gain and the normalized path delay of the s-th path, respectively. Without loss of generality, we assume $0 \le \tau_0 < \tau_1 < \cdots < \tau_{s-1} \le L-1$. Fortunately, numerous theoretical analysis and experimental results have verified that wireless channels are sparse in nature, i.e., in the CIR model(5), the dimension of the CIR L may be large, but the number of the active paths S with significant gains is usually small, i.e., $S \ll L$, especially in the wireless wideband communications [34], [38].

Defining a new vector $\boldsymbol{w} \triangleq [\boldsymbol{h}^T \; \boldsymbol{i}^T]^T \in \mathbb{C}^{L+N}$ and a new matrix $\boldsymbol{\Phi} \triangleq [\sqrt{N}\boldsymbol{X}\boldsymbol{F}_L \; \boldsymbol{F}] \in \mathbb{C}^{N\times(L+N)}$, we can rephrase (4) as an "augmented" model

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{w} + \boldsymbol{n} \tag{6}$$

where the matrix $\boldsymbol{\Phi}$ is obviously an underdetermined matrix. Meanwhile noting that both IN vector $\boldsymbol{i}$ and CIR vector $\boldsymbol{h}$ are sparse, the new constructed vector $\boldsymbol{w}$ is also viewed as a sparse vector reasonably. So the estimation of $\boldsymbol{w}$ in (6) can be considered as a typical compressed sensing problem [39]. Moreover, note that the data symbols, namely some element of the matrix $\boldsymbol{\Phi}$, is unknown at the receiver and need to be estimated, which necessitates the development of techniques that are capable of handling partially unknown dictionary matrices.

A. JOINT CHANNEL AND IN ESTIMATION ALGORITHM USING PILOT SUBCARRIERS

Let P and D denote the index set corresponding to the pilot subcarriers and data subcarriers, respectively. The part pertaining to the pilot subcarriers in (6) can be written as

$$\boldsymbol{y}_P = \boldsymbol{\Phi}_P\boldsymbol{w} + \boldsymbol{n}_P \tag{7}$$

where $\boldsymbol{y}_P$ is a $M \times 1$ vector containing the elements of $\boldsymbol{y}$ sampled at pilot locations, $\boldsymbol{\Phi}_P$ is a $M \times (L+N)$ submatrix of $\boldsymbol{\Phi}$ consisting of the rows corresponding to the pilot locations, and $\boldsymbol{n}_P$ is also a $M \times 1$ vector consisting of the components of $\boldsymbol{n}$ sampled at pilot locations. Since $\boldsymbol{\Phi}_P$ is a known flat matrix and $\boldsymbol{w}$ is a sparse vector, we may use CS theory to estimate $\boldsymbol{w}$ directly.

Many CS algorithms have been proposed in the literatures. In practice, different CS algorithms will have different requirements on the matrix and the sparsity for a reliable recovery. In this paper, we adopt SBL to solve the problem. Compared with the greedy CS algorithm, the SBL algorithms have shown super recovery performance when the recovering signal is less sparsity or the measurement matrix is higher coherence, which could well fit our problem [40], [41]. Another advantage of SBL is that it is capable of handling partially unknown dictionary matrices by virtue of the EM framework, which leads to the solution for (6). In the following, we derive the algorithm for joint channel and IN estimation using pilots based on SBL framework (*JCI*).

For observation model (7), SBL imposes firstly a parameterized Gaussian prior on the vector $\boldsymbol{w}$, given by

$$p(\boldsymbol{w}; 0, \boldsymbol{\Gamma}) = \prod_{i=0}^{N+L-1} (\pi\gamma_i)^{-1} \exp\left(-\frac{|w_i|^2}{\gamma_i}\right) \tag{8}$$

where the covariance matrix $\boldsymbol{\Gamma} \triangleq diag(\gamma_0, \gamma_1, \ldots, \gamma_{N+L-1})$.

Given the observation model (7) and prior (8) of $\boldsymbol{w}$, the posterior of $\boldsymbol{w}$ is also a Gaussian distribution

$$p(\boldsymbol{w}|\boldsymbol{y}_P; \lambda, \boldsymbol{\Gamma}) \sim \mathcal{CN}(\boldsymbol{w}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{9}$$

where λ is a scalar corresponding to the background noise variance. The posterior mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ are respectively given by

$$\boldsymbol{\mu} = \frac{1}{\lambda}\boldsymbol{\Sigma}\boldsymbol{\Phi}_P^*\boldsymbol{y}_P \tag{10}$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\boldsymbol{\Phi}_P^*(\lambda\mathbf{I} + \boldsymbol{\Phi}_P\boldsymbol{\Gamma}\boldsymbol{\Phi}_P^*)^{-1}\boldsymbol{\Phi}_P\boldsymbol{\Gamma} \tag{11}$$

The posterior mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ cannot be obtained directly from (10) and (11) since there are unknown hyperparameters $\boldsymbol{\Gamma}$ and λ. The well-known type-II maximum likelihood (ML) estimator is often used to estimate them which are solved by maximizing the marginalized pdf of $\boldsymbol{y}_P$. Since the ML estimation problem cannot be solved in closed form, iterative algorithm such as EM is employed. Upon termination of the EM algorithm, the maximum a posteriori (MAP) estimate $\hat{\boldsymbol{w}}$ of $\boldsymbol{w}$ is the posterior mean $\boldsymbol{\mu}$, i.e., $\hat{\boldsymbol{w}} = \boldsymbol{\mu}$.

Using EM algorithm to estimate the hyperparameter $\boldsymbol{\Gamma}$ and λ, the k-th iteration procedure is as follows:

(1)**E-step**: the sparse vector $\boldsymbol{w}$ is treated as latent variable and the expectation $Q(\boldsymbol{\Gamma}, \lambda)$ of joint pdf $p(\boldsymbol{y}_P, \boldsymbol{w}; \boldsymbol{\Gamma}, \lambda)$ under the posterior pdf $p(\boldsymbol{w}|\boldsymbol{y}_P)$ is given by

$$Q(\boldsymbol{\Gamma}, \lambda|\boldsymbol{\Gamma}^{(k)}, \lambda^{(k)}) = E_{\boldsymbol{w}|\boldsymbol{y}_P}\left\{\log p(\boldsymbol{y}_P, \boldsymbol{w}; \boldsymbol{\Gamma}^{(k)}, \lambda^{(k)})\right\} \tag{12}$$

(2)**M-step**: the updated hyperparameter $\boldsymbol{\Gamma}^{(k+1)}$ and $\lambda^{(k+1)}$ can be obtained by maximizing the expectation $Q(\boldsymbol{\Gamma}, \lambda)$, which is showed in (13):

$$(\boldsymbol{\Gamma}^{(k+1)}, \lambda^{(k+1)}) = \arg\max_{\boldsymbol{\Gamma},\lambda} Q(\boldsymbol{\Gamma}, \lambda|\boldsymbol{\Gamma}^{(k)}, \lambda^{(k)}) \tag{13}$$

The solution of $\boldsymbol{\Gamma}^{(k+1)}$ and $\lambda^{(k+1)}$ are expressed as:

$$\gamma_i^{(k+1)} = \boldsymbol{\Sigma}_{i,i}^{(k)} + (\boldsymbol{\mu}_{i,i}^{(k)})^2 \tag{14}$$

$$\lambda^{(k+1)} = \frac{1}{M}\left\{\begin{array}{l}\left\|\boldsymbol{y}_P - \boldsymbol{\Phi}_P\boldsymbol{\mu}^{(k)}\right\|_2^2 \\ + \lambda^{(k)}\sum_{i=0}^{N+L-1}\left[1 - (\boldsymbol{\gamma}_i^{(k)})^{-1}\boldsymbol{\Sigma}_{i,i}^{(k)}\right]\end{array}\right\} \tag{15}$$

where M is the number of pilot subcarriers.

Upon termination criteria $\left\|\boldsymbol{\mu}^{(k+1)} - \boldsymbol{\mu}^{(k)}\right\|_2^2 \leqslant \varepsilon$ of the EM algorithm, we obtain the MAP estimator $\hat{\boldsymbol{w}}$. According the definition of $\boldsymbol{w}$, the MAP estimator of channel and IN are $\hat{\boldsymbol{h}} = \hat{\boldsymbol{w}}[1 : L]$ and $\hat{\boldsymbol{i}} = \hat{\boldsymbol{w}}[L+1 : N+L]$ respectively.

We then transform the estimation of IN into the frequency domain and subtract it from the received signal in the data tones according to (16):

$$\hat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{F}\hat{\boldsymbol{i}} = \hat{\boldsymbol{\Lambda}}\boldsymbol{x} + \boldsymbol{F}(\boldsymbol{i} - \hat{\boldsymbol{i}}) + \boldsymbol{n} \tag{16}$$

the residual IN is treated as background noise. The main diagonal elements of matrix $\hat{\boldsymbol{\Lambda}}$ is the frequency response of the channel estimation $\hat{\boldsymbol{h}}$. Then equalization with zero-force is used to compensate the channel gain as $\hat{\boldsymbol{x}} = \hat{\boldsymbol{\Lambda}}^{-1}\hat{\boldsymbol{y}}$. After that the conventional detection and decoding algorithms will be applied. The entire algorithm of JCI is summarized in Algorithm 1.

Algorithm 1 JCI

Input: $\boldsymbol{y}_P$, $\boldsymbol{\Phi}_P$, r_{max}, and ε
Output: $\hat{\boldsymbol{w}}$
  Given the initial value: $\boldsymbol{\Gamma}^{(0)} = \mathbf{I}$ and $\lambda^{(0)} = 1$
  while $\left\|\boldsymbol{\mu}^{(r+1)} - \boldsymbol{\mu}^{(r)}\right\|_2^2 \geqslant \varepsilon$ and $r \leqslant r_{max}$ **do**
    E-step:
      Update $\boldsymbol{\mu}$ based on (10)
      Update $\boldsymbol{\Sigma}$ based on (11)
    M-step:
      Update $\boldsymbol{\Gamma}$ based on (14)
      Update λ based on (15)
  end while
  return $\hat{\boldsymbol{w}} = \boldsymbol{\mu}$

B. JOINT CHANNEL AND IN ESTIMATION ALGORITHM WITH SYMBOL DETECTION

The performance of above proposed JCI is limited by the number of pilot subcarriers. However increasing the number of pilot subcarriers will lead to reduced spectra efficiency and system throughput. If the information in data subcarriers of OFDM symbol can be exploited, it is desirable to improve the estimation performance of channel and IN with system throughput guarantee.

Using all subcarriers, the observation model is the same as (6). So the posterior mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ of $\boldsymbol{w}$ are respectively expressed as

$$\boldsymbol{\mu} \triangleq \begin{pmatrix}\boldsymbol{\mu}_h \\ \boldsymbol{\mu}_i\end{pmatrix} = \frac{1}{\lambda}\boldsymbol{\Sigma}\boldsymbol{\Phi}^*\boldsymbol{y} \tag{17}$$

$$\boldsymbol{\Sigma} \triangleq \begin{pmatrix}\boldsymbol{\Sigma}_{hh} & \boldsymbol{\Sigma}_{hi} \\ \boldsymbol{\Sigma}_{ih} & \boldsymbol{\Sigma}_{ii}\end{pmatrix} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\boldsymbol{\Phi}^*(\lambda\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{\Phi}\boldsymbol{\Gamma} \tag{18}$$

where $\boldsymbol{\mu}_h$ and $\boldsymbol{\mu}_i$ are the subsets corresponding to channel and IN partitions in the mean vector $\boldsymbol{\mu}$ respectively and the same principle is applied to the covariance matrix $\boldsymbol{\Sigma}$. Note that (10),(11) and (17),(18) are different in that the former uses only the known pilot symbols to construct the sensing matrix $\boldsymbol{\Phi}_P$, whereas the latter uses the pilot symbols along with the estimated transmit symbols to construct the sensing matrix $\mathbf{\Phi}$. In order to determine the matrix $\mathbf{\Phi}$, we need to estimate the unknown data symbol in data subcarriers. Subsequently we derive the algorithm for joint channel estimation, IN estimation, and symbol detection using pilot and data subcarriers based on SBL framework, which is denoted as JCIS.

Define $\boldsymbol{\theta} \triangleq \{\boldsymbol{X}, \mathbf{\Gamma}, \lambda\}$ for simplicity. Considering the sparse vector $\boldsymbol{w}$ as latent variable and $\boldsymbol{\theta}$ as the parameters to be estimated, the E-step and M-step of the JCIS-SBL algorithm can be given as

$$\text{E-step}: \quad Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = \mathrm{E}_{\boldsymbol{w}|\boldsymbol{y}}\left\{\log p(\boldsymbol{y},\boldsymbol{w};\boldsymbol{\theta}^{(k)})\right\} \tag{19}$$

$$\text{M-step}: \quad \boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) \tag{20}$$

Similar to the (14) and (15), the solutions of $\mathbf{\Gamma}^{(k+1)}$ and $\lambda^{(k+1)}$ are given as

$$\boldsymbol{\gamma}_i^{(k+1)} = \mathbf{\Sigma}_{i,i}^{(k)} + (\boldsymbol{\mu}_{i,i}^{(k)})^2 \tag{21}$$

$$\lambda^{(k+1)} = \frac{1}{N}\left\{\begin{aligned} &\left\|\boldsymbol{y} - \mathbf{\Phi}^{(k)}\boldsymbol{\mu}^{(k)}\right\|_2^2 \\ &+ \lambda^{(k)}\sum_{i=0}^{N+L-1}\left[1 - (\boldsymbol{\gamma}_i^{(k)})^{-1}\mathbf{\Sigma}_{i,i}^{(k)}\right]\end{aligned}\right\} \tag{22}$$

where N is the number of OFDM symbol subcarriers.

Next we derive the update equation for the data symbol matrix $\boldsymbol{X}^{(k+1)}$ with the $\mathbf{\Gamma}^{(k+1)}$ and $\lambda^{(k+1)}$ already obtained by (21) and (22). We notice that $\log p(\boldsymbol{y},\boldsymbol{w};\boldsymbol{\theta}) = \log p(\boldsymbol{y}\,|\boldsymbol{w}\,; \boldsymbol{X}, \lambda) + \log p(\boldsymbol{w}; \mathbf{\Gamma})$ and the second term is independent on $\boldsymbol{X}$, so the objective function in the $\mathbf{M}$-Step to maximize over $\boldsymbol{X}^{(k+1)}$ can be written as

$$\begin{aligned} \boldsymbol{X}^{(k+1)} &= \arg\max_{\boldsymbol{X}} Q(\boldsymbol{X}^{(k+1)}; \mathbf{\Phi}^{(k)}, \mathbf{\Gamma}^{(k+1)}, \lambda^{(k+1)}) \\ &= \arg\max_{\boldsymbol{X}} \left\{c - \mathrm{E}_{\boldsymbol{w}|\boldsymbol{y}}\left\{\frac{\|\boldsymbol{y}-\mathbf{\Phi}\boldsymbol{w}\|_2^2}{\lambda}\right\}\right\} \\ &= \arg\max_{\boldsymbol{X}} \left\{c - \lambda^{-1}\left[\|\boldsymbol{y} - \mathbf{\Phi}\boldsymbol{\mu}\|_2^2 + \mathrm{Tr}(\mathbf{\Phi}^*\mathbf{\Phi}\mathbf{\Sigma})\right]\right\} \\ &= \arg\min_{\boldsymbol{X}} \|\boldsymbol{y} - \mathbf{\Phi}\boldsymbol{\mu}\|_2^2 + \mathrm{Tr}(\mathbf{\Phi}^*\mathbf{\Phi}\mathbf{\Sigma}) \end{aligned} \tag{23}$$

where c is a constant independent of $\boldsymbol{X}$ and $\mathrm{Tr}(\cdot)$ denotes the trace of matrix. Using the definition of $\mathbf{\Phi}$, $\boldsymbol{\mu}$, $\mathbf{\Sigma}$, we can get the following equation:

$$\begin{aligned} \|\boldsymbol{y} - \mathbf{\Phi}\boldsymbol{\mu}\|_2^2 + \mathrm{Tr}(\mathbf{\Phi}^*\mathbf{\Phi}\mathbf{\Sigma}) = &\left\|\boldsymbol{y} - \sqrt{N}\boldsymbol{X}\boldsymbol{F}_L\boldsymbol{\mu}_h - \boldsymbol{F}\boldsymbol{\mu}_i\right\|_2^2 \\ &+ \mathrm{Tr}(N\boldsymbol{X}\boldsymbol{F}_L\mathbf{\Sigma}_{\mathbf{hh}}\boldsymbol{F}_L^*\boldsymbol{X}^* \\ &+ \sqrt{N}\boldsymbol{F}\mathbf{\Sigma}_{\mathbf{ih}}\boldsymbol{F}_L^*\boldsymbol{X}^* \\ &+ \sqrt{N}\boldsymbol{X}\boldsymbol{F}_L\mathbf{\Sigma}_{\mathbf{hi}}\boldsymbol{F}^* \\ &+ \boldsymbol{F}\mathbf{\Sigma}_{\mathbf{ii}}\boldsymbol{F}^*) \end{aligned} \tag{24}$$

Since the channel and IN are independent and uncorrelated in reality, $\mathbf{\Sigma}_{\mathbf{ih}}$ and $\mathbf{\Sigma}_{\mathbf{hi}}$ in (24) can be set to zero matrix.

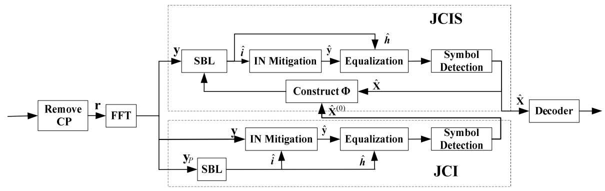


FIGURE 1. Block diagram of our proposed receivers.

Substituting (24) into (23), we can get the update rule of $\boldsymbol{X}$:

$$x_j^{(k+1)} = \arg\min_{x_j\in\Omega}\left\{\begin{aligned} &\left|\boldsymbol{y}\,(j) - x_j\sqrt{N}\boldsymbol{F}_L\,(j,:)\,\hat{\boldsymbol{h}} - \boldsymbol{F}\,(j,:)\,\hat{\boldsymbol{i}}\right|^2 \\ &+ \left|x_j\right|^2\boldsymbol{C}_b\,(j,j)\end{aligned}\right\} \tag{25}$$

where $j \in D$, Ω denotes M-QAM constellation points from which the transmitted symbol is selected, $\boldsymbol{C}_b = N\boldsymbol{F}_L\mathbf{\Sigma}_{\mathbf{hh}}\boldsymbol{F}_L^*$, $\boldsymbol{F}\,(j,:)$ is the j^{th} row of the matrix $\boldsymbol{F}$.

The JCIS requires initial estimate of the unknown data symbol $\boldsymbol{X}$ to construct the initial measurement matrix $\mathbf{\Phi}^{(0)}$. To ensure the convergence of JCIS, we firstly adopt the previous proposed JCI algorithm to estimate the channel and IN. Then we obtain the initial estimate $\boldsymbol{X}^{(0)}$ of $\boldsymbol{X}$ through the equalization and detection according to (16). Hence, the initialization of measurement matrix $\mathbf{\Phi}^{(0)}$ can be constructed by using $\boldsymbol{X}^{(0)}$. The entire algorithm of JCIS is summarized in Algorithm 2 and the receiver structure is depicted in Fig.1.

Algorithm 2 JCIS

Input: $\boldsymbol{y}$, $\mathbf{\Phi}_p$, r_{max}, and ε

Output: $\hat{\boldsymbol{w}}$, $\hat{\boldsymbol{X}}$

Given the initial value: $\mathbf{\Gamma}^{(0)} = \mathbf{I}$, $\lambda^{(0)} = 1$ and $\boldsymbol{X}^{(0)}$ obtained by using JCI

while $\left\|\boldsymbol{\mu}^{(r+1)} - \boldsymbol{\mu}^{(r)}\right\|_2^2 \geqslant \varepsilon$ and $r \leqslant r_{max}$ **do**

Construct $\mathbf{\Phi}^{(r)} = [\sqrt{N}\boldsymbol{X}^{(r)}\boldsymbol{F}_L\ \boldsymbol{F}]$

E-step:

Update $\boldsymbol{\mu}$ based on (17)

Update $\mathbf{\Sigma}$ based on (18)

M-step:

Update $\mathbf{\Gamma}$ based on (21)

Update λ based on (22)

Update $\boldsymbol{X}$ based on (25)

end while

return $\hat{\boldsymbol{w}}$, $\hat{\boldsymbol{X}}$

C. COMPLEXITY ANALYSIS

The computational complexity of the JCI is dominated by the matrix multiplication and inversion operations in (10) and (11), which has a complexity of $O(M(N+L)^2)$ per iteration. After given the initial values of data symbols, the *JCIS* using all tones to estimate channel and IN has a complexity of $O(N(N+L)^2)$ per iteration. So the total complexity of JCIS is $O((N+M)(N+L)^2)$. Compared with JCI, the complexity of JCIS is higher. This means JCIS improves the performance at

TABLE 1. Parameters for simulations.

Parameter	Notation	Value
Subcarrier Interval	Δf	15kHz
Sampling Frequency	f_s	3.84MHz
OFDM symbol Duration	T_s	$83.3\mu s$
Guard Interval Duration	T_g	$16.6\mu s$
Total Number of Subcarriers	N	256
Cyclic Prefix Length	N_G	64
Channel Length	L	64

the expense of increasing complexity, which will be verified by simulation results in the next Section.

IV. CRLB ANALYSIS

The CRLB provides a fundamental lower limit on the MSE performance of unbiased estimators. For random parameter estimation with availability of a *priori* information, the Bayesian Cramér-Rao Bound(BCRB) is already used to obtain lower bounds of dynamical Rayleigh channel complex gains estimation in OFDM system [42]. CRLB for SBL algorithm was derived in [43]. In this section, we present a closed-form BCRB expression for our proposed *JCIS* algorithm.

Let $\hat{\boldsymbol{w}}(\boldsymbol{y})$ denotes an unbiased estimator of $\boldsymbol{w}$ using the observations $\boldsymbol{y}$. The MSE matrix of $\hat{\boldsymbol{w}}(\boldsymbol{y})$ is defined as

$$\boldsymbol{M}^{\boldsymbol{w}} \triangleq \mathrm{E}_{\boldsymbol{y},\boldsymbol{w}}[(\hat{\boldsymbol{w}}(\boldsymbol{y}) - \boldsymbol{w})(\hat{\boldsymbol{w}}(\boldsymbol{y}) - \boldsymbol{w})^H] \tag{26}$$

and the Bayesian Information Matrix(BIM) is defined as

$$\begin{aligned}\boldsymbol{J}^{\boldsymbol{w}} &\triangleq \mathrm{E}_{\boldsymbol{y},\boldsymbol{w}}[-\nabla^2_{\boldsymbol{w}} \log(p(\boldsymbol{y}, \boldsymbol{w}))] \\ &= \mathrm{E}_{\boldsymbol{y},\boldsymbol{w}}[-\nabla^2_{\boldsymbol{w}} \log(p(\boldsymbol{y}|\boldsymbol{w}))] + \mathrm{E}_{\boldsymbol{y},\boldsymbol{w}}[-\nabla^2_{\boldsymbol{w}} \log(p(\boldsymbol{w}))]\end{aligned} \tag{27}$$

where $p(\boldsymbol{y}|\boldsymbol{w})$ is the conditional probability density function of $\boldsymbol{y}$ given $\boldsymbol{w}$ and $p(\boldsymbol{w})$ is the priori probability distribution of $\boldsymbol{w}$. Assuming that the MSE matrix exists and the BIM is non-singular, they should satisfy $\boldsymbol{M}^{\boldsymbol{w}} \succeq (\boldsymbol{J}^{\boldsymbol{w}})^{-1}$.

Proposition 1: Using the signal model in (6), the **BCRB** of the variable $\boldsymbol{w}$ is given by

$$\begin{aligned}\mathbf{BCRB}(\boldsymbol{w}) &= (\boldsymbol{J}^{\boldsymbol{w}})^{-1} \\ &= \left(\frac{\boldsymbol{\Phi}^*\boldsymbol{\Phi}}{\lambda} + \boldsymbol{\Gamma}^{-1}\right)^{-1} \\ &= \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\boldsymbol{\Phi}^*(\lambda\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{\Phi}\boldsymbol{\Gamma}\end{aligned} \tag{28}$$

Proof: See Appendix A. □

The closed form expressions of BCRBs associated to the estimation of $\boldsymbol{h}$ and $\boldsymbol{i}$ are separately given by

$$\mathrm{BCRB}(\boldsymbol{h}) = \mathrm{Tr}(\mathbf{BCRB}(\boldsymbol{w})_{[1:L,1:L]}). \tag{29}$$

and

$$\mathrm{BCRB}(\boldsymbol{i}) = \mathrm{Tr}(\mathbf{BCRB}(\boldsymbol{w})_{[L+1:N+L,L+1:N+L]}). \tag{30}$$

V. SIMULATION RESULTS

In this section, we demonstrate the performance of the proposed joint channel estimation and IN estimation algorithms through Monte Carlo simulations. We consider a 3 MHz OFDM system with 256 subcarriers, with a sampling frequency of $f_s = 3.84$MHz, resulting in an OFDM symbol duration of $83.3\mu s$ with Cyclic Prefix (CP) of $16.67\mu s$. The Rayleigh-fading uncorrelated-scattering model with sparse impulse response [44] is adopted and the length of channel length(L) equals the length of CP. It is assumed that the channel taps remain constant during the entire duration of one OFDM symbol as well. For IN and background noise simulation, we use the publicly available software [45], which adopt the GM model with $K = 3$, $p_k = \{0.9, 0.07, 0.03\}$, and $\gamma_k = \{1, 100, 1000\}$. The choice of the noise model parameters is such that the impulsive-to-background noise power ratio is up to 20dB(about 7% in all noise samples) and 30dB(about 3% in all noise samples). The modulation schemes is 4-QAM. The signal-to-noise ratio (SNR) is defined as SNR $= P_s/P_u$, where P_s and P_u are the power of the transmitted signal and total noise respectively. Without loss of generality, the pilot and data symbol power are normalized as one.

A. COMPARISON OF CHANNEL ESTIMATION PERFORMANCE

In this subsection, we compare the channel estimation performance of our proposed algorithms with the following algorithms.

- Ideal LS [9]: assumed that the Multipath Intensity Profile (MIP) of channel is known and IN is completely removed, the least squares(LS) method is used for channel estimation.
- SBL-LS [18]:assumed that the Multipath Intensity Profile (MIP) of channel is known and IN is mitigated by SBL only using null tones, the least square(LS) method is used for channel estimation.
- SBL-JCS: the IN is firstly mitigated by SBL using null tones and then the algorithm proposed in [34] is used to estimate the channel.
- JCSwIN: the algorithm proposed in [34] is used to estimate the channel but the IN is not mitigated.
- CS: the ℓ^1-norm minimization algorithm [39] is used to jointly estimate the channel and IN based on (7).

The Mean Square Error (MSE) of channel estimation is defined as $MSE \triangleq E\left\{\left\|\boldsymbol{h} - \hat{\boldsymbol{h}}\right\|_2^2\right\}$. Fig. 2 and Fig. 3 show the MSE of all algorithms versus SNR when the number of pilot subcarriers is set to 44 and 64 respectively. The BCRB of JCIS is also plotted as benchmark. In order to use the conventional CS-based IN mitigation algorithm, we set the number of null subcarriers as 50. But it is important to note that our proposed algorithms need not rely on null subcarriers. It can be seen that the performance of JCIS is almost not declined with the reduction of pilot subcarriers but the performance of the Ideal LS, SBL-JCS, and SBL-LS are degraded obviously. It shows that the channel estimation methods with joint data detection need less pilot subcarriers. The JCIS achieves about 15dB SNR gain over SBL-LS in that the former estimates jointly the channel, IN, and data symbols but the latter does not. JCSwIN has the worst performance at lower SNRs because it do not

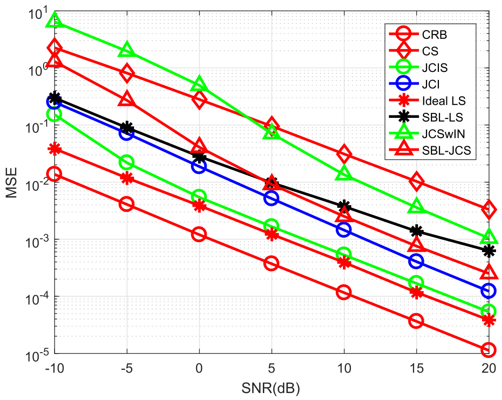


FIGURE 2. The MSE of channel estimation versus SNR. The number of pilot subcarriers is 44. The number of null subcarriers is 50.

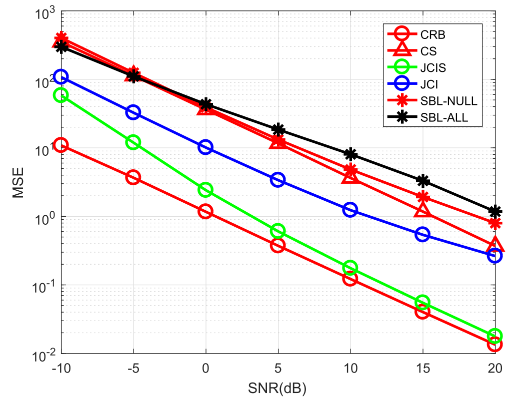


FIGURE 4. The MSE of IN estimation versus SNR. The number of null subcarriers is 50. The number of pilot subcarriers is 44.

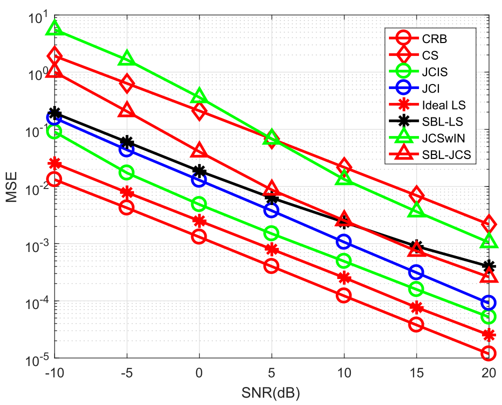


FIGURE 3. The MSE of channel estimation versus SNR. The number of pilot subcarriers is 64. The number of null subcarriers is 50.

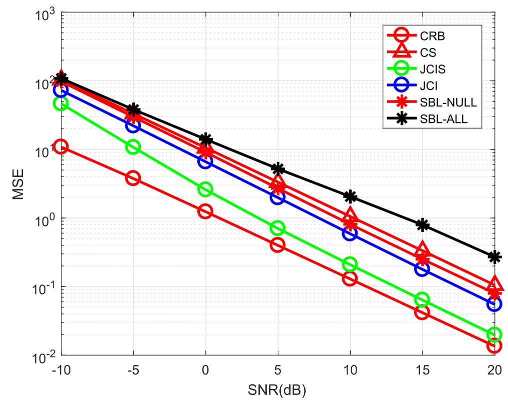


FIGURE 5. The MSE of IN estimation versus SNR. The number of null subcarriers is 100. The number of pilot subcarriers is 44.

remove the IN and treat the IN as background noise, which shows that those optimal channel estimation algorithms under AWGN will deteriorate rapidly in the presence of IN.

B. COMPARISON OF IN ESTIMATION PERFORMANCE

In this subsection, the IN estimation performance is demonstrated. The MSE of IN estimation is defined as $MSE \triangleq E\left\{\left\|i-\hat{i}\right\|_2^2\right\}$. The number of pilot subcarriers is set to 44 and the number of null subcarriers is set to 50 and 100 respectively. Fig. 4 and Fig. 5 show the IN estimation MSE performance of our proposed algorithms and others versus SNR. SBL-NULL and SBL-ALL are proposed in [18] which exploit the null subcarriers and all subcarriers respectively. The BCRB of JCIS is also plotted as benchmark. It can be seen that JCIS maintains stable estimation performance under two circumstance. It manifests that JCIS needs less the number of null subcarriers by jointly estimating the channel, IN, and data symbols. As a conventional IN estimation method only exploiting the null subcarriers, the performance of SBL-NULL is degraded along with the decrease of the number of null subcarriers. SBL-ALL demonstrates poor performance in that it simply views the received signal at

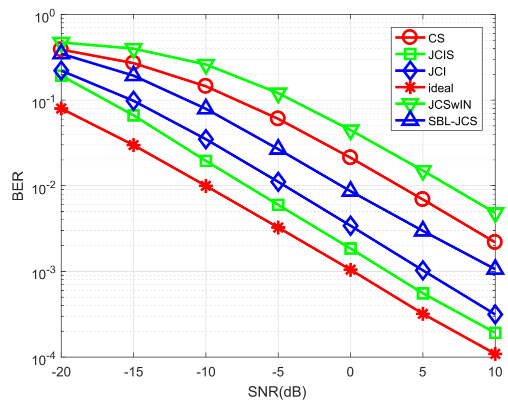


FIGURE 6. BER versus SNR in uncoded OFDM system. The number of pilot subcarriers is 44. The number of null subcarriers is 50.

data subcarriers as background noise and cannot carry out the channel estimation and data detection.

C. THE COMPARISON OF SYSTEM PERFORMANCE FOR UNCODED AND CODED SYSTEM

In this subsection, the BER performance of all algorithms in uncoded and coded systems are plotted in

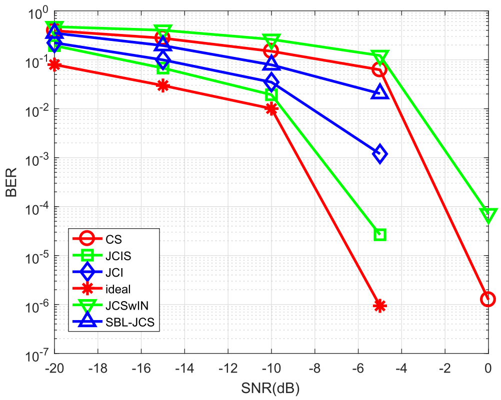


FIGURE 7. **BER versus SNR in coded OFDM system. The number of pilot subcarriers is 44. The number of null subcarriers is 50.**

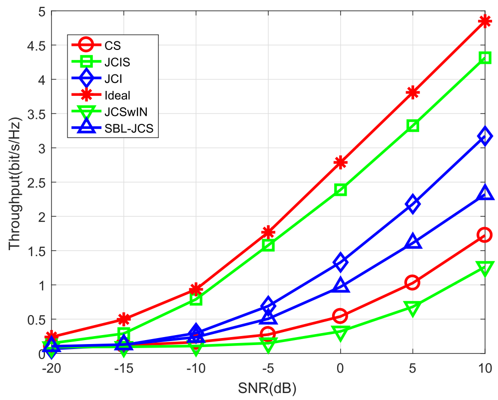


FIGURE 8. **System throughput versus SNR. The number of pilot subcarriers is 44. The number of null subcarriers is 50.**

Fig. 6 and Fig. 7 respectively. The number of pilot and null subcarriers are 44 and 50 respectively. The total number of transmitted symbols is 10^4 in uncoded systems. In coded systems, a rate $1/3$ Turbo code is used and the total number of transmitted frames is 10^4. For both uncoded and coded systems, the 4-QAM modulation scheme is adopted. The Ideal case is also depicted as benchmarks that the channel state information is perfectly obtained and the impulsive noise is completely mitigated at the receiver side.

In the uncoded system, JCI achieves 5dB SNR gain over SBL-JCS, 7dB SNR gain over CS, and 12-15dB SNR gain over JCSwIN respectively. JCIS obtains additional 2-3dB gain by using all tones. In the coded system, JCI achieves 2dB SNR gain over CS and 3 dB SNR gain over JCSwIN at a target BER of 10^{-3} respectively. JCIS also obtains additional 3dB SNR gain over JCI. It is also noted that, in the $-10 \sim -5$dB SNR region, the BER of JCIS declines from 2×10^{-2} to 3×10^{-5} and the BER of JCI declines from 3×10^{-2} to about 10^{-3}. Yet the BER of SBL-JCS, CS, and JCSwIN decline quite slowly.

We also notice that the performance of JCSwIN is the worst among all algorithms in all experiments. It manifests

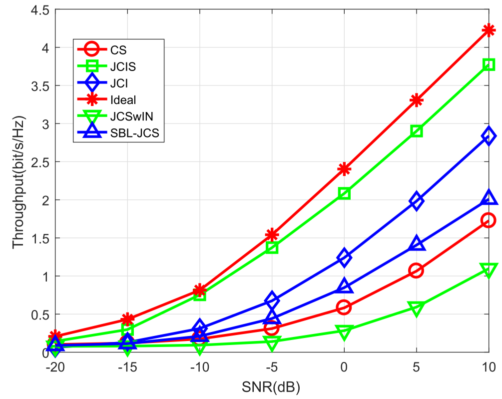


FIGURE 9. **System throughput versus SNR. The number of pilot subcarriers is 64. The number of null subcarriers is 50.**

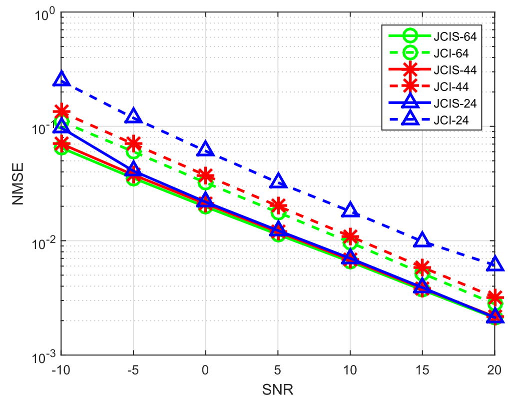


FIGURE 10. **The MSE performance comparison between JCI and JCIS with various number of pilot tones.**

that the IN will deteriorate rapidly the performance of those algorithms which are optimal under AWGN.

Subsequently, the system throughput performance of all algorithms are plotted in Fig. 8 and Fig. 9 respectively. The throughput of our proposed algorithms is derived in Appendix B. It can be observed that JCIS and JCI can obtain higher throughput than the other algorithms. By exploiting all tones to estimate the channel and impulsive noise, JCIS can achieve higher throughput performance.

D. COMPARISON BETWEEN JCI AND JCIS

For further comparison, we study the performance of JCI and JCIS with a different number of pilot tone and null tone in this subsection. Firstly we compare the performance of JCI and JCIS as the number of pilot tone is set to 64, 44, and 24 respectively, which are showed in Fig. 10 and Fig. 11 respectively. From Fig.10, we observe that the channel estimation performance of JCIS is not deteriorated with the reduction of pilot tones and the performance of JCI is deteriorated conversely. Fig.11 shows that the BER performance of JCIS is slightly degraded and that of JCI is degraded significantly. Next we compare the performance of JCI and JCIS as the number of null tone is set to 80, 50, and 20 respectively, which

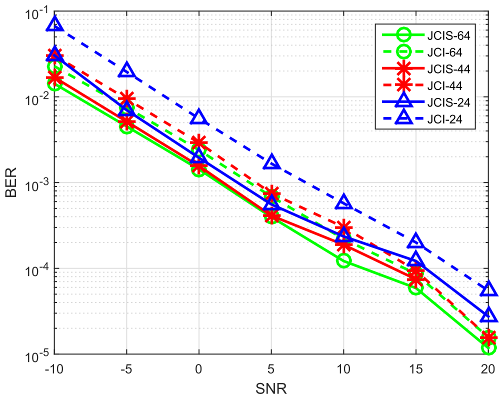


FIGURE 11. The uncoded BER performance comparison between JCI and JCIS with various number of pilot tones.

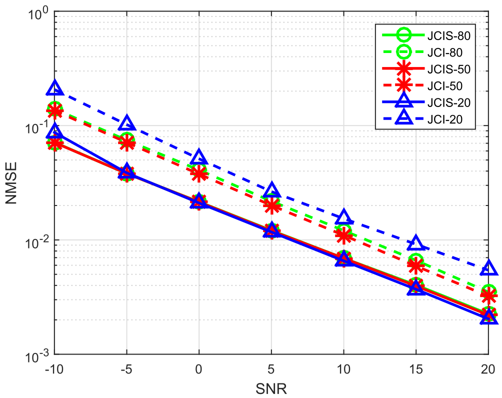


FIGURE 12. The MSE performance comparison between JCI and JCIS with various number of null tones.

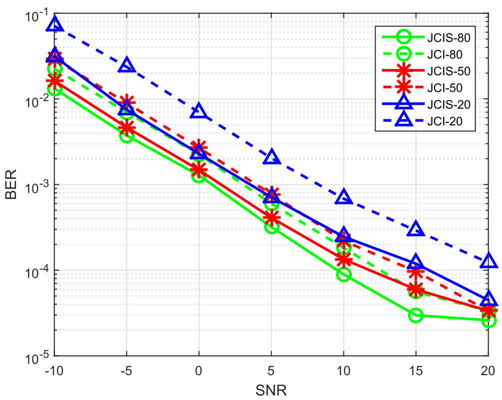


FIGURE 13. The uncoded BER performance comparison between JCI and JCIS with various number of null tones.

are showed in Fig.12 and Fig.13 respectively. From Fig.12 we observe that the channel estimation performance of JCIS is not deteriorated with the reduction of null tones and the performance of JCI is deteriorated conversely. Fig.13 shows that the BER performance of JCIS is slightly degraded and that of JCI is degraded significantly. The reason why JCIS can obtain better and more stable performance than JCI is that JCIS can utilize all tones of received OFDM symbol for joint channel, impulsive noise, and symbol estimation.

VI. CONCLUSION

In this paper, we consider the joint sparse channel estimation, impulsive noise mitigation, and data detection for OFDM systems. By observing the sparsity of channel and impulsive noise in the time domain, we construct an expanded sparse vector to represent the channel and impulsive noise together. To estimate the augmented vector, JCI algorithm is proposed which uses only the pilot and null subcarriers. Furthermore JCIS algorithm is developed to improve the performance of channel estimation and impulsive noise cancelation, which apply the data detection simultaneously. We derive the analytical expression of BCRB for JCIS algorithm as well. The MSE performance of our proposed scheme outperforms the conventional methods and is close to the lower bound. Moreover, simulation results show our methods can have a good BER performance with fewer pilot and null subcarriers and obtain better spectral efficiency.

APPENDIX A

According the OFDM system model (4), the conditional probability density $p(\boldsymbol{y}|\boldsymbol{h},\boldsymbol{i})$ is expressed as

$$p(\boldsymbol{y}|\boldsymbol{h},\boldsymbol{i})=(\pi\lambda)^{-N}\exp\left(-\frac{\|\boldsymbol{y}-\sqrt{N}\boldsymbol{X}\boldsymbol{F}_L\boldsymbol{h}-\boldsymbol{F}\boldsymbol{i}\|_2^2}{\lambda}\right). \tag{31}$$

The probability density of $\boldsymbol{h}$ and $\boldsymbol{i}$ are respectively

$$p(\boldsymbol{h}) = (\pi)^{-L}\,|\boldsymbol{\Gamma}_{\boldsymbol{h}}|^{-1}\exp\left(-\boldsymbol{h}^*\boldsymbol{\Gamma}_{\boldsymbol{h}}^{-1}\boldsymbol{h}\right) \tag{32}$$

and

$$p(\boldsymbol{i}) = (\pi)^{-N}\,|\boldsymbol{\Gamma}_{\boldsymbol{i}}|^{-1}\exp\left(-\boldsymbol{i}^*\boldsymbol{\Gamma}_{\boldsymbol{i}}^{-1}\boldsymbol{i}\right). \tag{33}$$

The joint probability of $\boldsymbol{y},\boldsymbol{h},\boldsymbol{i}$ is $p(\boldsymbol{y},\boldsymbol{h},\boldsymbol{i}) = p(\boldsymbol{y}|\boldsymbol{h},\boldsymbol{i})p(\boldsymbol{h})p(\boldsymbol{i})$. The Fisher Information Matrix (FIM) $\boldsymbol{J}_D$ is expressed as

$$\begin{aligned}\boldsymbol{J}_D &= -\mathrm{E}_{\boldsymbol{h},\boldsymbol{i}}[\nabla^2_{\boldsymbol{h},\boldsymbol{i}}\log(p(\boldsymbol{y}|\boldsymbol{h},\boldsymbol{i}))]\\ &= \frac{1}{\lambda}\begin{bmatrix} N\boldsymbol{F}_L^*\boldsymbol{X}^*\boldsymbol{X}\boldsymbol{F}_L & \sqrt{N}\boldsymbol{F}_L^*\boldsymbol{X}^*\boldsymbol{F}\\ \sqrt{N}\boldsymbol{F}^*\boldsymbol{X}\boldsymbol{F}_L & \boldsymbol{F}^*\boldsymbol{F}\end{bmatrix}\end{aligned} \tag{34}$$

The priori information matrix $\boldsymbol{J}_P$ is expressed as

$$\begin{aligned}\boldsymbol{J}_P &= -\mathrm{E}_{\boldsymbol{h},\boldsymbol{i}}[\nabla^2_{\boldsymbol{h},\boldsymbol{i}}(\log(p(\boldsymbol{h}))+\log(p(\boldsymbol{i})))]\\ &= \begin{bmatrix}\boldsymbol{\Gamma}_{\boldsymbol{h}}^{-1} & \\ & \boldsymbol{\Gamma}_{\boldsymbol{i}}^{-1}\end{bmatrix}\end{aligned} \tag{35}$$

The Bayesian Information Matrix(BIM) of variable $\boldsymbol{h},\boldsymbol{i}$ is given by

$$\boldsymbol{J} = \boldsymbol{J}_D + \boldsymbol{J}_P \tag{36}$$

$$= \frac{1}{\lambda}\begin{bmatrix} N\boldsymbol{F}_L^*\boldsymbol{X}^*\boldsymbol{X}\boldsymbol{F}_L & \sqrt{N}\boldsymbol{F}_L^*\boldsymbol{X}^*\boldsymbol{F}\\ \sqrt{N}\boldsymbol{F}^*\boldsymbol{X}\boldsymbol{F}_L & \boldsymbol{F}^*\boldsymbol{F}\end{bmatrix} + \begin{bmatrix}\boldsymbol{\Gamma}_{\boldsymbol{h}}^{-1} & \\ & \boldsymbol{\Gamma}_{\boldsymbol{i}}^{-1}\end{bmatrix} \tag{37}$$

$$= \frac{1}{\lambda}\begin{bmatrix}\sqrt{N}\boldsymbol{X}\boldsymbol{F}_L\\ \boldsymbol{F}\end{bmatrix}^*\begin{bmatrix}\sqrt{N}\boldsymbol{X}\boldsymbol{F}_L & \boldsymbol{F}\end{bmatrix} + \begin{bmatrix}\boldsymbol{\Gamma}_{\boldsymbol{h}} & \\ & \boldsymbol{\Gamma}_{\boldsymbol{i}}\end{bmatrix}^{-1} \tag{38}$$

$$= \frac{\boldsymbol{\Phi}^*\boldsymbol{\Phi}}{\lambda} + \boldsymbol{\Gamma}^{-1} \tag{39}$$

We also can utilize the observation model (6) to derive the BIM. The conditional probability density $p(\boldsymbol{y}|\boldsymbol{w})$ is expressed as

$$p(\boldsymbol{y}|\boldsymbol{w}) = (\pi\lambda)^{-N} \exp\left(-\frac{\|\boldsymbol{y}-\boldsymbol{\Phi w}\|_2^2}{\lambda}\right). \tag{40}$$

The probability density of $\boldsymbol{w}$ is

$$p(\boldsymbol{w}) = (\pi)^{-(N+L)} |\boldsymbol{\Gamma}|^{-1} \exp\left(-\boldsymbol{w}^*\boldsymbol{\Gamma}^{-1}\boldsymbol{w}\right). \tag{41}$$

Thus the log function of joint probability of $\boldsymbol{y}, \boldsymbol{w}$ has the following form

$$\begin{aligned}\log(p(\boldsymbol{y},\boldsymbol{w})) &= \log(p(\boldsymbol{y}|\boldsymbol{w})) + \log(p(\boldsymbol{w}))\\ &= -N\log(\pi\lambda) - (N+L)\log(\pi) - \log(|\boldsymbol{\Gamma}|)\\ &\quad - \frac{\|\boldsymbol{y}-\boldsymbol{\Phi w}\|_2^2}{\lambda} - \boldsymbol{w}^*\boldsymbol{\Gamma}^{-1}\boldsymbol{w}. \end{aligned}\tag{42}$$

The Bayesian Information Matrix(BIM) of variable $\boldsymbol{w}$ is given by

$$\begin{aligned}\boldsymbol{J}^{\boldsymbol{w}} &= -\mathrm{E}_{\boldsymbol{w}}[\nabla^2_{\boldsymbol{w}} \log(p(\boldsymbol{y},\boldsymbol{w}))]\\ &= -\mathrm{E}_{\boldsymbol{w}}\left[\nabla_{\boldsymbol{w}}\left(\frac{\boldsymbol{\Phi}^*(\boldsymbol{y}-\boldsymbol{\Phi w})}{\lambda} - \boldsymbol{\Gamma}^{-1}\boldsymbol{w}\right)\right]\\ &= \frac{\boldsymbol{\Phi}^*\boldsymbol{\Phi}}{\lambda} + \boldsymbol{\Gamma}^{-1} \end{aligned}\tag{43}$$

APPENDIX B

We denote that the signal power in i-th subcarrier as $s_i = E\{|x_i|^2\}$. The total noise power in i-th subcarrier is given by

$$\sigma_i^2 = E\{|(H_i - \hat{H}_i)x_i|^2\} + E\{|\boldsymbol{F}_{i\cdot}(\boldsymbol{i}-\hat{\boldsymbol{i}})|^2\} + E\{|n_i|^2\} \tag{44}$$

where $\boldsymbol{F}_{i\cdot}$ expresses the i-th row of DFT matrix $\boldsymbol{F}$. H_i and $\hat{H}_i$ express the perfect channel frequency response coefficients and its estimation respectively. $\boldsymbol{F}_{i\cdot}(\boldsymbol{i}-\hat{\boldsymbol{i}})$ is the projection of the residual impulsive noise onto each subcarrier. n_i expresses the background noise component of each subcarrier. (44) means that the total noise of each subcarrier consists of the interference due to channel estimation error, the residual impulsive noise, and the background noise.

The SNR on the i-th subcarrier can be expressed as

$$SNR(i) = \frac{|\hat{H}_i|^2 s_i}{\sigma_i^2} \tag{45}$$

The total channel capacity of OFDM system expressed in bit per symbol is given by [19], [46]

$$\begin{aligned} R &= \frac{1}{N}\sum_{i\in D}\log_2\left(1+\frac{SNR(i)}{\Gamma}\right)\\ &= \frac{1}{N}\sum_{i\in D}\log_2\left(1+\frac{|\hat{H}_i|^2 s_i}{\Gamma\sigma_i^2}\right) \text{bit/s/Hz} \end{aligned}\tag{46}$$

where Γ denotes the SNR gap which is dependent on coding gain and targeted bit error rate.

ACKNOWLEDGMENT

The work of K. C. Wong was supported in part by the Magna Fund, Ningbo University, and in part by the Science Foundation of Zhejiang Business Technology Institute under Grant 2017Z01.

REFERENCES

[1] S. F. Hasan, X. Ding, N. H. Siddique, and S. Chakraborty, "Measuring disruption in vehicular communications," *IEEE Trans. Veh. Technol.*, vol. 60, no. 1, pp. 148–159, Jan. 2011.

[2] S. W. Lai and G. G. Messier, "Using the wireless and PLC channels for diversity," *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3865–3875, Dec. 2012.

[3] Z. Wang, S. Zhou, J. Catipovic, and P. Willett, "Asynchronous multiuser reception for OFDM in underwater acoustic communications," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1050–1061, Mar. 2013.

[4] S. Liu, F. Yang, W. Ding, and J. Song, "Double kill: Compressive-sensing-based narrow-band interference and impulsive noise mitigation for vehicular communications," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5099–5109, Jul. 2016.

[5] F. Sacuto, F. Labeau, and B. L. Agba, "Wide band time-correlated model for wireless communications under impulsive noise within power substation," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1449–1461, Mar. 2014.

[6] X. Kuai, H. Sun, S. Zhou, and E. Cheng, "Impulsive noise mitigation in underwater acoustic OFDM systems," *IEEE Trans. Veh. Technol.*, no. 65, no. 10, pp. 8190–8202, Oct. 2016.

[7] *IEEE Standard for Information technology–Local and metropolitan area networks–Specific requirements–Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments*, IEEE Standard 802.11p Std., Jul. 2010.

[8] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[9] Y. Liu, Z. Tan, H. Hu, L. J. Cimini, and G. Y. Li, "Channel estimation for OFDM," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1891–1908, 4th Quart., 2014.

[10] A. Mahmood, M. Chitre, and M. A. Armand, "Detecting OFDM signals in alpha-stable noise," *IEEE Trans. Commun.*, vol. 62, no. 10, pp. 3571–3583, Oct. 2014.

[11] S. V. Zhidkov, "Performance analysis and optimization of OFDM receiver with blanking nonlinearity in impulsive noise environment," *IEEE Trans. Veh. Technol.*, vol. 55, no. 1, pp. 234–242, Jan. 2006.

[12] S. V. Zhidkov, "Analysis and comparison of several simple impulse noise mitigation schemes for OFDM receivers," *IEEE Trans. Commun.*, vol. 56, no. 1, pp. 5–9, Jan. 2008.

[13] U. Epple and M. Schnell, "Advanced blanking nonlinearity for mitigating impulsive interference in OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 146–158, Jan. 2017.

[14] F. H. Juwono, Q. Guo, Y. Chen, L. Xu, D. D. Huang, and K. P. Wong, "Linear combining of nonlinear preprocessors for OFDM-based power-line communications," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 253–260, Jan. 2016.

[15] N. Rožić, P. Banelli, D. Begušić, and J. Radić, "Multiple-threshold estimators for impulsive noise suppression in multicarrier communications," *IEEE Trans. Signal Process.*, vol. 66, no. 6, pp. 1619–1633, Mar. 2018.

[16] D. Darsena, G. Gelli, F. Melito, and F. Verde, "ICI-free equalization in OFDM systems with blanking preprocessing at the receiver for impulsive noise mitigation," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1321–1325, Sep. 2015.

[17] G. Caire, T. Y. Al-Naffouri, and A. K. Narayanan, "Impulse noise cancellation in OFDM: An application of compressed sensing," in *Proc. IEEE Int. Symp. Inf. Theory*, Toronto, ON, Canada, Jul. 2008, pp. 1293–1297.

[18] J. Lin, M. Nassar, and B. L. Evans, "Impulsive noise mitigation in powerline communications using sparse Bayesian learning," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1172–1183, Jul. 2013.

[19] T. Y. Al-Naffouri, A. A. Quadeer, and G. Caire, "Impulse noise estimation and removal for OFDM systems," *IEEE Trans. Commun.*, vol. 62, no. 3, pp. 976–989, Mar. 2014.

[20] L. Lampe, "Bursty impulse noise detection by compressed sensing," in *Proc. IEEE Int. Symp. Power Line Commun. Appl.*, Udine, Italy, Apr. 2011, pp. 29–34.

[21] M. Korki, J. Zhang, C. Zhang, and H. Zayyani, "Block-sparse impulsive noise reduction in OFDM systems—A novel iterative Bayesian approach," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 271–284, Jan. 2016.

[22] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, Sep. 2011.

[23] X. Jiang, T. Kirubarajan, and W.-J. Zeng, "Robust sparse channel estimation and equalization in impulsive noise using linear programming," *Signal Process.*, vol. 93, no. 5, pp. 1095–1105, May 2013.

[24] A. Mehboob, L. Zhang, J. Khangosstar, and K. Suwunnapuk, "Joint channel and impulsive noise estimation for OFDM based power line communication systems using compressed sensing," in *Proc. IEEE 17th Int. Symp. Power Line Commun. Appl.*, Johannesburg, South Africa, Mar. 2013, pp. 203–208.

[25] Y.-R. Chien, "Iterative channel estimation and impulsive noise mitigation algorithm for OFDM-based receivers with application to power-line communications," *IEEE Trans. Power Del.*, vol. 30, no. 6, pp. 2435–2442, Dec. 2015.

[26] M. Nassar, P. Schniter, and B. L. Evans, "A factor graph approach to joint OFDM channel estimation and decoding in impulsive noise environments," *IEEE Trans. Signal Process.*, vol. 62, no. 6, pp. 1576–1589, Mar. 2014.

[27] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul./Aug. 2011, pp. 2174–2178.

[28] J. Ziniel and P. Schniter, "Efficient high-dimensional inference in the multiple measurement vector problem," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 340–354, Jan. 2013.

[29] P. Schniter, "A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1462–1474, Dec. 2011.

[30] Y. Hou, R. Liu, B. Dai, and L. Zhao, "Joint channel estimation and LDPC decoding over time-varying impulsive noise channels," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2376–2383, Jun. 2018.

[31] P. Chen, Y. Rong, S. Nordholm, Z. He, and A. J. Duncan, "Joint channel estimation and impulsive noise mitigation in underwater acoustic OFDM communication systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6165–6178, Sep. 2017.

[32] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–214, Jun. 2001.

[33] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[34] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint approximately sparse channel estimation and data detection in OFDM systems using sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 62, no. 14, pp. 3591–3603, Jul. 2014.

[35] M. Nassar, X. E. Lin, and B. L. Evans, "Stochastic modeling of microwave oven interference in WLANs," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kyoto, Japan, Jun. 2011, pp. 1–6.

[36] D. Middleton, "Statistical-physical models of electromagnetic interference," *IEEE Trans. Electromagn. Compat.*, vol. EMC-19, no. 3, pp. 106–127, Aug. 1977.

[37] H. El Ghannudi, L. Clavier, N. Azzaoui, F. Septier, and P.-A. Rolland, "α-stable interference modeling and cauchy receiver for an IR-UWB ad hoc network," *IEEE Trans. Commun.*, vol. 58, no. 6, pp. 1748–1757, Jun. 2010.

[38] W. Ding, F. Yang, C. Pan, L. Dai, and J. Song, "Compressive sensing based channel estimation for OFDM systems under long delay channels," *IEEE Trans. Broadcast.*, vol. 60, no. 2, pp. 313–321, Jun. 2014.

[39] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[40] D. P. Wipf, "Sparse estimation with structured dictionaries," in *Proc. 24th Int. Conf. Neural Inf. Process. Syst.*, Granada, Spain, Dec. 2011, pp. 2016–2024.

[41] W. Ding, Y. Lu, F. Yang, W. Dai, P. Li, S. Liu and J. Song, "Spectrally efficient CSI acquisition for power line communications: A Bayesian compressive sensing perspective," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 7, pp. 2022–2032, Jul. 2016.

[42] H. Hijazi and L. Ros, "Analytical analysis of Bayesian Cramér-Rao bound for dynamical rayleigh channel complex gains estimation in OFDM system," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1889–1900, May 2009.

[43] R. Prasad and C. R. Murthy, "Cramér-Rao-type bounds for sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 622–632, Mar. 2013.

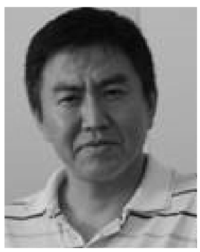
[44] P. Schniter, "Belief-propagation-based joint channel estimation and decoding for spectrally efficient communication over unknown sparse channels," *Phys. Commun.*, vol. 5, no. 2, pp. 91–101, Jun. 2012.

[45] K. Gulati, M. Nassar, A. Chopra, N. B. Okafo, M. DeYoung, N. Aghasadeghi, A. Sujeeth, and B. L. Evans, *UT Austin Interference Modeling and Mitigation Toolbox*. [online], Available:. [Online]. Available: http://users.ece.utexas.edu/~bevans/projects/rfi/index.html

[46] T. N. Vo, K. Amis, T. Chonavel, and P. Siohan, "Achievable throughput optimization in OFDM systems in the presence of interference and its application to power line networks," *IEEE Trans. Commun.*, vol. 62, no. 5, pp. 1704–1715, May 2014.

XINRONG LV received the B.E. degree in electrical engineering from the China University of Petroleum, Dongying, China, in 1997, and the M.S. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2002. He is currently pursuing the Ph.D. degree in communication engineering with Ningbo University, Ningbo, China. His current research interests include sparse signal processing and its applications in the fields of power line communications, smart grids, and underwater acoustic communications.

YOUMING LI received the B.S. degree in computational mathematics from Lanzhou University, Lanzhou, China, in 1985, the M.S. degree in computational mathematics from Xi'an Jiaotong University, Xi'an, China, in 1988, and the Ph.D. degree in electrical engineering from Xidian University, Xi'an, in 1995. From 1988 to 1998, he was with the Department of Applied Mathematics, Xidian University, where he was an Associate Professor. From 1999 to 2000, he was a Research Fellow of the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. From 2001 to 2003, he joined DSO National Laboratories, Singapore. In 2004, he was a Research Fellow of the School of Engineering, Bar-Ilan University, Israel. Since 2005, he has been with the Faculty of Electrical Engineering and Computer Science, Ningbo University, where he is currently a Professor. His research interests include the areas of statistical signal processing and its applications in communications, cognitive radio, and underwater acoustic communications.



YONGQING WU received the B.S. degree in communication engineering from Xidian University, Xi'an, China, in 1990, the M.S. degree in computer application from the Hebei University of Technology, Tianjin, China, in 1998, and the Ph.D. degree in precision instrument from Tianjin University, Tianjin, in 2001. Since 2001, he has been with the Institute of Acoustics, Chinese Academy of Sciences, where he is currently a Professor. His research interest includes the areas of array signal processing and its applications in underwater acoustic communications.

XIAOLI WANG received the M.S. degree in mechanical engineering from the Dalian University of Technology, in 2010. Since 2010, she has been with the Faculty of Electrical Engineering and Computer Science, Ningbo University, China, where she is currently a Lecturer. Her research interest includes signal processing with applications to communications.

HUI LIANG is currently the Project Manager of Ningbo Sanxing Electric Co., Ltd., Ningbo, China. His research interest includes OFDM technology with applications to smart grids.