Joint Impulsive Noise Estimation and Data Detection Conceived for LDPC-Coded DMT-Based DSL Systems

Tong Bai, *Student Member, IEEE*, Chao Xu, *Member, IEEE*, Rong Zhang, *Senior Member, IEEE*, Anas F. Al Rawi, *Member, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

***Abstract*—Impulsive noise is one of the most challenging issues in digital subscriber lines (DSL). In order to mitigate the deleterious effects of impulsive noise, the conventional automatic repeat request (ARQ) invokes cyclic redundancy checking (CRC) in order to estimate the existence of impulsive noise and then triggers retransmission, which degrades the spectral efficiency attained. More straightforward techniques of mitigating impulsive noise, such as blanking and clipping, require specific design, which increases the implementation complexity. Against the background, we propose a novel two-stage joint impulsive noise estimation and data detection scheme conceived for low-density parity-check (LDPC) coded discrete multitone (DMT)-based DSL systems. More explicitly, first of all, we propose a semi-blind estimation method, which is capable of estimating the arrival of impulsive noise without using CRC and additionally evaluating the power of impulsive noise with an adequate accuracy. Secondly, in order to improve the accuracy of impulsive noise estimation in more advanced LDPC-coded DMT-based DSL systems, we propose a decision-directed (DD) method for the second stage of channel decoding and data detection with the aid of *extrinsic* information transfer (EXIT) charts. Our proposed two-stage scheme is capable of approaching the performance of the idealistic scenario of perfectly knowing both the arrival time and the instantaneous power of impulsive noise. Moreover, we analyse the mean square error (MSE) of the proposed schemes in order to quantify the estimation accuracy and to reduce the estimation complexity. Our simulation results demonstrate that our proposed scheme is capable of achieving a near-capacity performance to using our LDPC coded DMT-based DSL system in the presence of impulsive noise.**

***Index Terms*—digital subscriber line, joint impulsive noise estimation and data detection, near-capacity performance.**

I. Introduction

Telephone lines date back to Alexander Graham Bell's invention of the telephone in 1875, when a voice signal was transmitted within a 3.4kHz-bandwidth channel [1]. With the ever-increasing demand for data services, the feasibility of twisted copper pairs to convey data has drawn substantial attention, resulting in a series of digital subscriber line (DSL) standards, such as the integrated services digital networks (ISDN) [2], asymmetric DSL (ADSL) [3], very high data-rate DSL (VDSL) [4] and the G.fast [5], which aims for supporting Giga-bit data-rate transmission at a low cost [6]–[8].

The bandwidth exploited has been gradually expanded from the 30 MHz of VDSL [4] to 106 MHz for G.fast [5], which is expected to significantly increase the overall throughput, albeit it imposes strong crosstalk at high operating frequencies. The current DSL system performance is limited both by crosstalk as well as by stationary and impulsive noise. The crosstalk can be substantially mitigated by duplex transmission [1] and vectoring [9]–[11] as well as by dynamic spectrum management [12] [13]. In this case, the nonstationary noise will have a more substantial impact on the system performance. Specifically, for the sake of reducing the latency, the duration of a discrete multitone symbol denoted by T_{DMT} is shortened from the 125μs duration of VDSL [4] to $20.83\mu s$ for G.fast [5]. This increases the system's vulnerability to bursty impulsive noise, because a single noise-impulse has a higher probability to contaminate several consecutive DMT symbols, hence resulting in bursty errors at the receiver. If only a small fraction of a DMT symbol is affected by impulsive noise, it is still possible to correct the errors with the aid of forward error correction (FEC) techniques. However, the FEC code is often overloaded in the presence of long bursty errors, which severely degrades the system's performance.

The reduction of impulsive noise effects is necessary for DSL and it has indeed drawn a large amount of research attention [14]–[19]. The combination of Reed-Solomon (RS) codes and trellis coded modulation (TCM) as well as interleaving is adopted in the current DSL standard at the time of writing [5]. Toumpakaris *et al*. [14] proposed an erasure decoding aided RS scheme for reducing the interleaving delay in the face of non-stationary disturbances, while Zhang *et al*. [15] analysed the performance of automatic repeat request (ARQ) schemes based on RS codes. Considering the inability of RS codes to benefit from soft-decisions, the trend is to adopt more advanced FEC codes [20], such as low-density parity-check (LDPC) codes, in DSL. Neckebroek *et al*. [16] proposed a retransmission-aided LDPC scheme for DSL, which is capable of improving the goodput of the standard. However, if the impulsive noise occurs with a relatively high probability, the retransmission will be triggered frequently, which degrades the system's transmission efficiency. Furthermore, the retransmissions are usually checked by a cyclic redundancy check (CRC) codes, which further degrades the spectral efficiency.

This work was supported in part by the EPSRC projects (EP/N004558/1 and EP/N023862/1), in part by the European Research Councils Advanced Fellow Grant under the Beam-Me-Up Project, in part by the Royal Societys Wolfson Research Merit Award and in part by the RAEng Industrial Fellow Grant.

Tong Bai, Chao Xu, Rong Zhang and Lajos Hanzo are with the School of Eletronics and Computer science, University of Southampton, SO17 1BJ, UK.

Anas F. Al Rawi is with Research and Technology, British Telecom (BT), Adastral Park, Martlesham Heath, IP5 3RE, UK.

More straightforward techniques of mitigating the impulsive noise, such as blanking, clipping as well as reconstruction and mitigation, are proposed in [17]–[19]. Blanking and clipping [17] requires threshold optimization and unfortunately it forms an error floor at high signal power levels. Reconstruction and mitigation requires complex component design [18] or requires pilot insertion [19] and hence it increases the implementation complexity.

Against this background, in order to avoid the ARQ-induced spectral efficiency degradation and to improve the accuracy of impulsive noise estimation for advanced LDPC-coded DMT-based DSL systems in this paper, we propose a joint two-stage impulsive noise estimation and data detection scheme. The main contributions of the paper are listed as follows.

- Firstly, we propose a semi-blind impulsive noise estimation as the first-stage estimation method, which is capable of estimating both the arrival instant of impulsive noise and the power of impulsive noise at an adequate accuracy without using CRC. We note that the proposed method effectively eliminates the potential spectral efficiency degradation imposed by ARQ retransmissions;
- Secondly, in order to improve the accuracy of the impulsive noise estimation, we propose an iterative decision-directed method for LDPC-coded DMT-based DSL systems as the second-stage estimation for high-order modulation constellations, where a beneficial iteration gain may be achieved. The estimation is integrated into the detection and decoding process, for the sake of mitigating both the complexity and the delay;
- Furthermore, in order to quantify the estimation accuracy and designate the appropriate samples for the impulsive noise estimation, we mathematically analyse the mean square error (MSE) of the two-stage algorithm and confirm its accuracy by simulation. Our results are also compared to the theoretical Cramer-Rao Lower Bound (CRLB).
- We optimize the number of iterations both within the LDPC decoder as well as between the soft-demapper and the decoder, with the aid of *extrinsic* information transfer (EXIT) charts.
- Finally, our simulation results demonstrate that our proposed scheme is capable of achieving a near-capacity performance for LDPC coded DMT-based DSL systems in the presence of impulsive noise. For example, the SNR-difference of the BER curve of our proposed schemes and the idealistic scenarios of perfectly knowing both the arrival instant and the instantaneous power of the impulsive noise is less 0.2dB for 16QAM and 4096QAM.

The rest of the paper is organized as follows. In Section II, we introduce the transmission system and noise model considered. The proposed joint impulsive noise estimation and data detection scheme is detailed in Section III. In Section IV, two design examples of the novel scheme are assessed both with the aid of EXIT charts and BER evaluations. Finally, the paper is concluded in Section V.

II. System and Noise Model

A. Coded DMT Transceiver

We consider a discrete multitone (DMT) system having N tones as illustrated in Fig. 1. The sequence of information bits denoted by $\boldsymbol{b} \in \{0,1\}$, which are encoded, hence resulting in the bits denoted by $\boldsymbol{c} \in \{0,1\}$. For Q-ary QAM having the alphabet $\boldsymbol{\mathcal{X}}$, the interleaved bits denoted by $\boldsymbol{d} \in \{0,1\}$ are grouped into a set of BPS $= \log_2 Q$ bits and then mapped with index $X^q \in \boldsymbol{\mathcal{X}}$ to Q-ary symbols denoted by $\boldsymbol{X}$, where q is the decimal value of the binary set of the BPS number of bits. In order to ensure having a real-valued output after IDFT, $\boldsymbol{X}$ is Hermitian transposed to $\overline{\boldsymbol{X}}$, subject to the following constraint [21]:

$$\begin{cases} \overline{X}_0 = \overline{X}_{N/2} = 0 \\ \overline{X}_n = \overline{X}^*_{N-n} \end{cases}. \tag{1}$$

Since the DSL channel is slowly time-variant, channel information can be estimated periodically and then compared to opt the results without the influence of impulsive noise. Therefore, it is reasonable to assume that the channel impulse response (IR) is perfectly known at the transmitter and $\overline{\boldsymbol{X}}$ is pre-equalized by the precoder, hence resulting in $\tilde{\boldsymbol{X}}$. Our Transmit Precoder (TPC) design is detailed later in this section. After the IDFT stage of Fig. 1, the time-domain signal denoted by $\tilde{\boldsymbol{x}} = [\tilde{x}_0, \tilde{x}_1, \ldots, \tilde{x}_{N-1}]^T$ is expressed as

$$\tilde{\boldsymbol{x}} = \boldsymbol{\mathcal{F}}^H \tilde{\boldsymbol{X}}, \tag{2}$$

where $\boldsymbol{\mathcal{F}}$ is the normalized discrete Fourier transform matrix, yielding $\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}^H = \boldsymbol{\mathcal{F}}^H\boldsymbol{\mathcal{F}} = \boldsymbol{I_N}$, where the elements of $\boldsymbol{\mathcal{F}}$ are defined as $\mathcal{F}_{i,k} = [1/\sqrt{N}\exp(j2\pi ik/N)]$.

The DMT symbols pass through the channel denoted by $\tilde{\boldsymbol{h}}$ and they are contaminated by the stationary noise denoted by $\tilde{\boldsymbol{n}}_{\boldsymbol{S}}$ and by impulsive noise denoted by $\tilde{\boldsymbol{n}}_{\boldsymbol{I}}$. Assuming that synchronization is perfectly established, the received symbols denoted by $\tilde{\boldsymbol{y}}$ are expressed as

$$\tilde{\boldsymbol{y}} = \sqrt{\rho}\tilde{\boldsymbol{h}}\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{n}}_{\boldsymbol{S}} + \tilde{\boldsymbol{n}}_{\boldsymbol{I}}, \tag{3}$$

where $\tilde{\boldsymbol{y}}$ is an N-element vector, $\tilde{\boldsymbol{h}}$ is an $(N \times N)$-element matrix and ρ is the received signal power per symbol.

At the receiver of Fig. 1, $\overline{\boldsymbol{Y}}$ can be recovered with the aid of the DFT as follows:

$$\begin{aligned} \overline{\boldsymbol{Y}} &= \boldsymbol{\mathcal{F}}\tilde{\boldsymbol{y}} = \boldsymbol{\mathcal{F}}(\sqrt{\rho}\tilde{\boldsymbol{h}}\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{n}}_{\boldsymbol{S}} + \tilde{\boldsymbol{n}}_{\boldsymbol{I}}) \\ &= \boldsymbol{\mathcal{F}}(\sqrt{\rho}\tilde{\boldsymbol{h}}\boldsymbol{\mathcal{F}}^H\tilde{\boldsymbol{X}} + \tilde{\boldsymbol{n}}_{\mathrm{S}} + \tilde{\boldsymbol{n}}_{\boldsymbol{I}}) \\ &= \sqrt{\rho}\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}^H\tilde{\boldsymbol{H}}\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}^H\tilde{\boldsymbol{X}} + \boldsymbol{\mathcal{F}}\tilde{\boldsymbol{n}}_{\boldsymbol{S}} + \boldsymbol{\mathcal{F}}\tilde{\boldsymbol{n}}_{\boldsymbol{I}} \\ &= \sqrt{\rho}\tilde{\boldsymbol{H}}\tilde{\boldsymbol{X}} + \tilde{\boldsymbol{N}}_{\boldsymbol{S}} + \tilde{\boldsymbol{N}}_{\boldsymbol{I}}, \end{aligned} \tag{4}$$

where $\tilde{\boldsymbol{h}} = \boldsymbol{\mathcal{F}}^H\tilde{\boldsymbol{H}}\boldsymbol{\mathcal{F}}$ and $\tilde{\boldsymbol{H}}$ is a diagonal matrix having the diagonal element of $[\tilde{H}_0, \tilde{H}_1, \cdots, \tilde{H}_{N-1}]$. As mentioned above, $\tilde{\boldsymbol{X}}$ is pre-equalized for each tone $n \in [0, N-1]$ as

$$\tilde{X}_n = \frac{\tilde{H}^*_n}{|\tilde{H}_n|^2}\overline{X}_n. \tag{5}$$

Then, (4) becomes

$$\overline{\boldsymbol{Y}} = \sqrt{\rho}\overline{\boldsymbol{X}} + \tilde{\boldsymbol{N}}_{\boldsymbol{S}} + \tilde{\boldsymbol{N}}_{\boldsymbol{I}}. \tag{6}$$

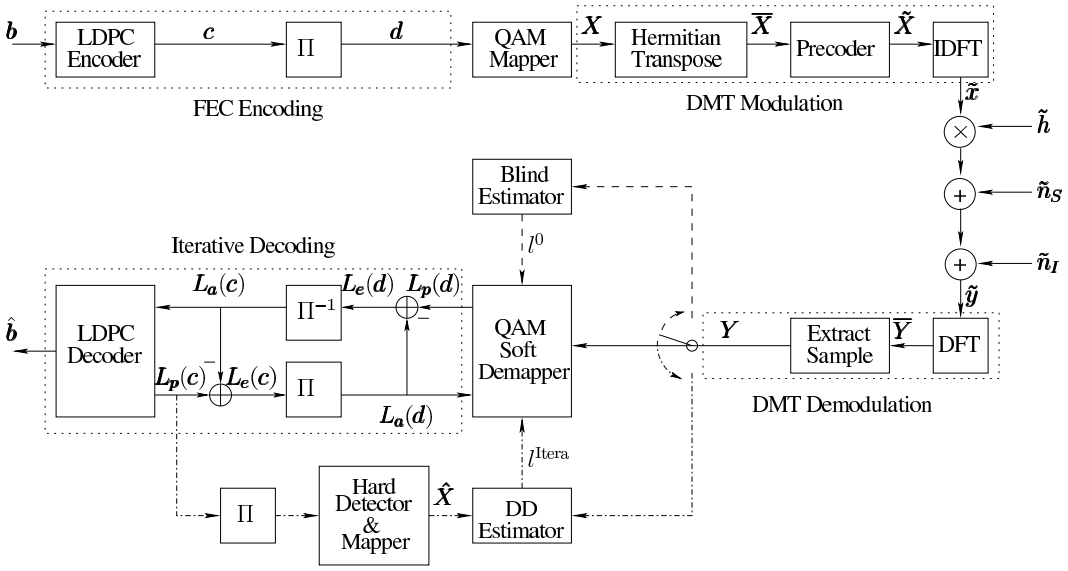


Fig. 1: The structure of our joint impulsive noise estimation and data detection scheme conceived for LDPC coded DMT-based DSL systems. "Blind" and "DD" refer to semi-blind and decision-directed, respectively.

After extracting the samples from $\overline{\boldsymbol{Y}}$, we have

$$\boldsymbol{Y} = \sqrt{\rho}\boldsymbol{X} + \boldsymbol{N_S} + \boldsymbol{N_I}, \tag{7}$$

where $\boldsymbol{Y}$ is a $\frac{N-2}{2}$-element vector, while $\boldsymbol{N_S}$ and $\boldsymbol{N_I}$ are the stationary noise and impulsive noise in the frequency domain, respectively. Then $\boldsymbol{Y}$ is processed in further stages, as discussed in Section III.

B. Noise Model

The noise process includes both stationary and impulsive noise, which are characterized both in the time-domain and the frequency-domain in this subsection.

1) Stationary Noise: As discussed in [22], the power of the stationary noise is flat over all frequencies and its amplitude is assumed to be Gaussian-distributed. Its probability density function (PDF) in the time domain can be expressed as

$$\tilde{\boldsymbol{n}}_{\boldsymbol{S}}[k] \sim \mathcal{N}(0, \sigma_S^2), \text{ for all } k, \tag{8}$$

where we denote the time-domain index by k and $\mathcal{N}(\mu, \sigma^2)$ represents the real-valued Gaussian distribution with the mean of μ and the variance of σ^2. The real-valued noise is transformed to the complex-valued variable after DFT and its PDF in the frequency domain is given by

$$\tilde{\boldsymbol{N}}_{\boldsymbol{S}}[n] \sim \mathcal{CN}(0, \sigma_S^2), \text{ for all } n, \tag{9}$$

where we denote the frequency-domain index by n and $\mathcal{CN}(\mu, \sigma^2)$ represents the complex-valued Gaussian distribution with a mean of μ and the variance of the real and imaginary parts given by $0.5\sigma^2$.

2) Impulsive Noise: The impulsive noise (IN) in DSL is usually characterized in terms of four aspects: amplitude, duration, inter-arrival time (IAT) and spectral characteristics. The main model used for the amplitude is the Weibull distribution [22] [23]. By contrast, the commonly used distribution for the duration is the twin-term log-normal approach of [22] [23] [24]. The IAT is usually modelled using the exponential [22], the Poisson [24] and the Markov Renewable Process (MRP) [23]. The auto-correlation function (ACF) is modelled as a negative logarithmic function [22] or as an exponentially decaying cosine function [23]. Therefore, it can be inferred that the impulsive noise in DSL is a coloured Weibull-distributed variable. When considering the DFT-based demodulation at the receiver, the behaviour of the coloured Weibull-distributed variable after DFT has to be discussed.

Here let us denote a correlated Weibull-distributed variable by $w[k]$, which is processed by DFT having a size of N. Then, we have

$$W[n] = \text{DFT}\big[w[k]\big] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} w[k] \cdot e^{-j\frac{2\pi kn}{N}}. \tag{11}$$

The following convergence has been formally proved in [25]

$$\Big[\Re\big(W[n]\big), \Im\big(W[n]\big)\Big] \Rightarrow \Big[G[n], G[n]\Big], \tag{12}$$

where $G[n]$ is an independent identically distributed (i.i.d.) normal random variable with zero-mean and a variance of $S[n]/2$, while $S[n]$ is the spectral density associated with the autocorrelation of the process. The convergence formulated in (12) can be approached with a sufficiently large DFT size, which is readily satisfied by the DMT size of G.fast [5] having $N = 2048$ or 4096. As a result, the IN amplitude distribution in the frequency domain can be approximated by a Gaussian distribution. Then, the PDF of $\tilde{\boldsymbol{N}}_{\boldsymbol{I}}$ is given by

$$P\big(\tilde{\boldsymbol{N}}_{\boldsymbol{I}}[n]\big) = \sum_{\ell} p(\ell) \cdot \mathcal{CN}\big[0, \sigma^2_{I,n}(\ell)\big], \tag{13}$$

where ℓ is the number of samples in a DMT symbol influenced by the impulsive noise in the time domain, while $p(\ell)$ (whose expression is shown in Section B.3) is the probability that the number of samples in a DMT symbol influenced by IN in the time domain happens to be ℓ. Furthermore, let us denote the IN variance at the tone index n by $\sigma^2_{\text{IN},n}$, when the whole

$$p(\ell) = \begin{cases} \frac{\mathcal{E}[D_{S_0,k}]}{\mathcal{E}[D_{S_0,k}]+\mathcal{E}[D_{S_1,k}]}, & \text{if } \ell = 0 \\ \mathcal{D}_1(\ell, B, v_1, t_1) + \mathcal{D}_1(\ell, 1-B, v_2, t_2), & \text{if } \ell = 1, 2, \cdots, N-1 \\ \mathcal{D}_2(B, v_1, t_1) + \mathcal{D}_2(1-B, v_2, t_2), & \text{if } \ell = N \end{cases} \tag{10}$$

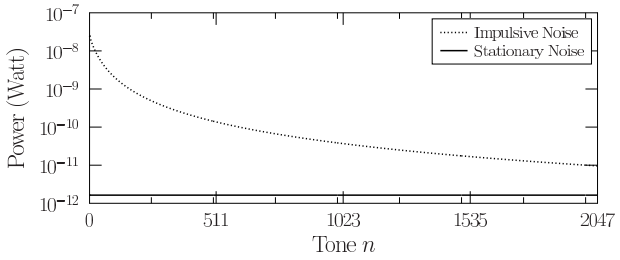


Fig. 2: Spectrum of impulsive noise and stationary noise in DSL, for the bandwidth of 106 MHz and tone spacing of 51.75 kHz. The impulsive noise is plotted according to (16). For the stationary noise, the power spectral density is set to -135 dBm/Hz.

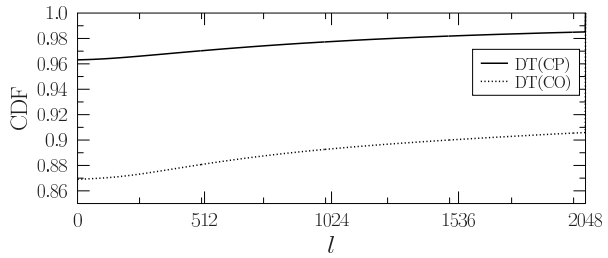


Fig. 3: Cumulative density function of ℓ at the Deutsche Telekom (DT) central offices (CO) and DT customer premises (CP) when $T_{\text{DMT}} = 20.83\mu s$. DT central office (CO) is parametrized as $B = 0.25$, $t_1 = 8\mu$s, $t_2 = 125\mu$s, $v_1 = 0.75$, $v_2 = 1.0$, $\lambda = 0.16\text{s}^{-1}$, $\theta = 1.5$ and $t_s = 1$ms. DT(CP) is parametrized as $B = 1$, $t_1 = 18\mu$s, $v_1 = 1.15$, $\lambda = 0.16\text{s}^{-1}$, $\theta = 1.5$ and $t_s = 1$ms.

OFDM symbol is influenced by IN in the time domain and then $\sigma^2_{I,n}(\ell)$ is a fraction of $\sigma^2_{\text{IN},n}$, yielding

$$\sigma^2_{I,n}(\ell) = \frac{\ell}{N}\sigma^2_{\text{IN},n}. \tag{14}$$

In this paper, the exponentially decaying cosine function of [23] is adopted to represent the ACF of IN:

$$R(t) = cos(2\pi\alpha t)\exp\big(-\beta|t|\big). \tag{15}$$

Let us denote the power spectral density of impulsive noise by $S(f) = \mathcal{FT}\big[R(t)\big]$, where $\mathcal{FT}$ represents the continuous Fourier transform. Then the IN variance can be expressed as

$$\begin{aligned} \sigma^2_{\text{IN},n} &= \frac{1}{Z}\int_{n\Delta f}^{(n+1)\Delta f} S(f)\,\mathrm{d}f \\ &= \frac{1}{2\pi Z}\Bigg\{\tan^{-1}\Big(\frac{2\pi\big[(n+1)\Delta f+\alpha\big]}{\beta}\Big) \\ &\quad - \tan^{-1}\Big(\frac{2\pi\big[n\Delta f+\alpha\big]}{\beta}\Big) \\ &\quad + \tan^{-1}\Big(\frac{2\pi\big[(n+1)\Delta f-\alpha\big]}{\beta}\Big) \\ &\quad - \tan^{-1}\Big(\frac{2\pi\big[n\Delta f-\alpha\big]}{\beta}\Big)\Bigg\}, \end{aligned} \tag{16}$$

where Δf is the bandwidth of each tone in the DMT, while Z is the impedance. The spectrum of both impulsive noise and stationary noise is shown in Fig 2.

3) Temporal Characteristics: The duration of IN can be modelled by the two-term log-normal form [23],

$$\begin{aligned} f_{\text{D}}(t) = B\frac{1}{\sqrt{2\pi}v_1 t}\exp\Big(-\frac{\ln^2(t/t_1)}{2v_1^2}\Big) \\ + (1-B)\frac{1}{\sqrt{2\pi}v_2 t}\exp\Big(-\frac{\ln^2(t/t_2)}{2v_2^2}\Big), \end{aligned} \tag{17}$$

and the IAT of IN can be modelled by the following two-state MRP [23],

$$f_{\text{IAT}}(t) = \begin{cases} \frac{1}{1-\exp(-\lambda t_s)}\lambda\exp(-\lambda t), & \text{if } t < t_s \\ \theta t_s^{\theta}/t^{\theta+1}, & \text{if } t \geq t_s, \end{cases} \tag{18}$$

where the state-transitions occur according to the matrix $\boldsymbol{P}$. The parameters of the noise model, such as $\alpha, \beta, t_1, t_2, v_1, v_2, \lambda, \theta, B$ and $\boldsymbol{P}$ vary according to the geographic location and the time of day according to [23] and to the realistic measurement-based document [26] by British Telecom (BT).

Based on the above temporal model, $p(\ell)$ is given as (10) in [27]. The expression of $p(\ell)$ is complex and long and hence readers are recommended to read the original reference [27] for more details. The cumulative density function (CDF) of ℓ is plotted in Fig 3.

III. Impulsive Noise Estimation and Data Detection

A. Impulsive Noise Estimation

The power spectrum of impulsive noise depends α and β in (16), which were found to be Gaussian distributed variables with a small variance [23] [26]. In this paper, we assume that the mean values of α and β are known. Furthermore, with the aid of (14), assuming that the power level of stationary noise is known at the receiver, the impulsive noise estimation can be simplified to estimate ℓ, namely the number of samples influenced in the time domain representation of a DMT symbol. The structure of our proposed joint impulsive noise estimation and data detection scheme is shown in Fig. 1. At the receiver side, the extracted samples $\boldsymbol{Y}$ are firstly forwarded to the proposed blind estimator, which estimates ℓ as it will be detailed in Section III-A.1. With knowledge of both $\sigma^2_{\text{IN},n}$ and estimated ℓ, the instantaneous impulsive noise variance level can be obtained according to (14). Given the estimate of the instantaneous noise variance, the QAM soft demapper becomes capable of producing more reliable log-likelihood ratios (LLRs) for the channel decoder. For the bit-to-symbol mapping schemes of high-order modulation, where substantial iteration gains can be achieved, a further

decision-directed (DD) estimator is proposed for improving the estimation accuracy. The interleaved *a posteriori* LLRs gleaned from the channel decoder are passed to the hard detector and the results then enter the proposed DD estimator as one of its inputs. Its second input is given by the extracted samples $\boldsymbol{Y}$. The DD estimator outputs the estimated value of ℓ for the QAM soft demapper, as detailed in Section III-A.2. Again, utilizing the estimated results and $\sigma^2_{\text{IN},n}$, the QAM soft demapper produces the soft LLRs for further iterations, where the DD estimation is activated, until an error-free status or a pre-set outer iteration limit is achieved. The proposed blind-DD estimators are detailed as follows.

1) Proposed Semi-Blind Method: According to (7), the received symbol of tone n is expressed as

$$Y_n = \sqrt{\rho}X_n + N_{S,n} + N_{I,n}. \tag{19}$$

Then, we have

$$\begin{aligned} Y_n^* Y_n = \rho X_n^* X_n + N_{S,n}^* N_{S,n} + N_{I,n}^* N_{I,n} + 2\Re(N_{I,n}^* N_{S,n}) \\ + 2\sqrt{\rho}\Re(X_n^* N_{S,n}) + 2\sqrt{\rho}\Re(X_n^* N_{I,n}), \end{aligned} \tag{20}$$

and

$$\begin{aligned} \boldsymbol{Y}_{sel}^H \boldsymbol{Y}_{sel} = \rho \sum X_n^* X_n + \sum N_{S,n}^* N_{S,n} + \sum N_{I,n}^* N_{I,n} \\ + 2\sqrt{\rho}\sum \Re(X_n^* N_{S,n}) + 2\sqrt{\rho}\sum \Re(X_n^* N_{I,n}) \\ + 2\sum \Re(N_{I,n}^* N_{S,n}), \end{aligned} \tag{21}$$

where $\boldsymbol{Y}_{sel}$ is the received vector associated with the elements selected for estimation and the selection of $\boldsymbol{Y}_{sel}$ is discussed later in Section III-A.5. Taking the expectation of (21), we have

$$\mathbb{E}[\boldsymbol{Y}_{sel}^H \boldsymbol{Y}_{sel}] = \rho K + K\sigma_S^2 + \frac{\ell}{N}\sum \sigma_{I,n}^2, \tag{22}$$

where K is the number of elements in $\boldsymbol{Y}_{sel}$. The estimated value $\hat{\ell}$ can be obtained as:

$$\begin{aligned} \hat{\ell} &= \frac{\mathbb{E}[\boldsymbol{Y}_{sel}^H \boldsymbol{Y}_{sel}] - \rho K - \sigma_S^2 K}{\frac{1}{N}\sum \sigma_{I,n}^2} \\ &\asymp \frac{\boldsymbol{Y}_{sel}^H \boldsymbol{Y}_{sel} - \rho K - \sigma_S^2 K}{\frac{1}{N}\sum \sigma_{I,n}^2} \\ &= \frac{\boldsymbol{Y}_{sel}^H \boldsymbol{Y}_{sel}}{\Omega} - \frac{\Psi}{\Omega}, \end{aligned} \tag{23}$$

where we define $\Psi = \rho K + \sigma_S^2 K$ and $\Omega = \frac{1}{N}\sum \sigma_{I,n}^2$.

2) Proposed Decision-Directed Method: (19) can be reformulated as

$$N_{S,n} + N_{I,n} = Y_n - \sqrt{\rho}X_n. \tag{24}$$

Taking the square of both sides, we have

$$|N_{S,n} + N_{I,n}|^2 = |Y_n - \sqrt{\rho}X_n|^2. \tag{25}$$

Here we denote the samples estimated on the basis of the previous decision by $\hat{X}_n = X_n + \epsilon_n$, where ϵ_n is the error between the transmitted sample X_n and the estimated sample $\hat{X}_n$ on tone n. Taking the expectation of both sides of (25), we have

$$\begin{aligned} \sigma_S^2 + \frac{\ell_n}{N}\sigma_{I,n}^2 &= \mathbb{E}\{|Y_n - \sqrt{\rho}X_n|^2\} \\ &= \mathbb{E}\{|Y_n - \sqrt{\rho}(\hat{X}_n - \epsilon_n)|^2\} \\ &\approx \mathbb{E}\{|Y_n - \sqrt{\rho}\hat{X}_n|^2\} \\ &\asymp |Y_n - \sqrt{\rho}\hat{X}_n|^2, \end{aligned} \tag{26}$$

where we denote the instantaneous estimate of ℓ at tone n by $\hat{\ell}_n$. The approximation in the third line of (26) is reasonable because the semi-blind estimation method results in a small value ϵ_n. Then we have

$$\hat{\ell}_n = \frac{N\big(|Y_n - \sqrt{\rho}\hat{X}_n|^2 - K\sigma_S^2\big)}{\sigma_{I,n}^2}. \tag{27}$$

Since ℓ may be deemed to be identical over all tones in a DMT symbol, the number of error-infested samples in a DMT symbol can be expressed as

$$\hat{\ell} = \frac{1}{K}\sum_{n=M}^{M+K-1} \frac{N\big(|Y_n - \sqrt{\rho}\hat{X}_n|^2 - K\sigma_S^2\big)}{\sigma_{I,n}^2}, \tag{28}$$

where M and K represent the initial tone index and the number of tones involved for variance estimation, respectively.

3) Numerical Results: In this subsection, we will investigate the performance of the proposed scheme.

Lemma 1: The MSE is one of the most popular parameters used for evaluating the estimation accuracy. According to Appendix A, the MSE of the blind estimation methods may be expressed as

$$\begin{aligned} \text{MSE}(\hat{\ell}) &= \mathbb{E}\{(\ell - \hat{\ell})^2\} \\ &= \frac{\sum \sigma_{I,n}^4}{N^2\Omega^2}\ell^2 + \frac{2(\rho + \sigma_S^2)}{\Omega}\ell + \frac{K\sigma_S^4}{\Omega^2}. \end{aligned} \tag{29}$$

Lemma 2: It should be noted that the results of the proposed estimation method are unbiased. According to Appendix B, the Cramer-Rao Lower Bound (CRLB) is expressed as

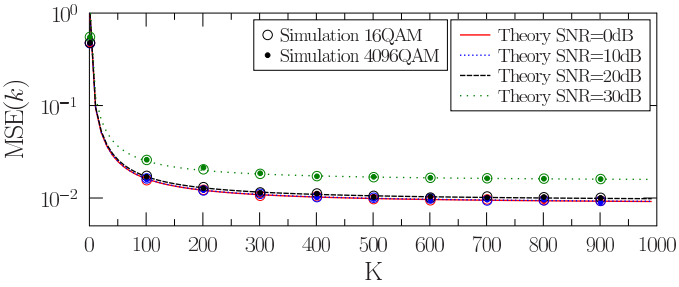
$$\text{CRLB}(\hat{\ell}) = \frac{1}{\sum_{n=M}^{M+K-1} \frac{(\sigma_{I,n}^2)^2}{(\ell\sigma_{I,n}^2 + N\sigma_S^2)^2}}. \tag{30}$$

The normalized MSE and CRLB are defined as follows:

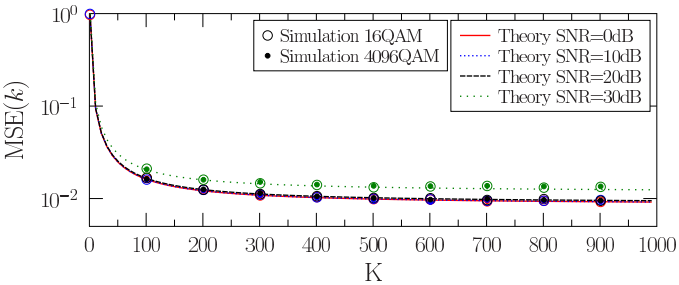
$$\overline{\text{MSE}}(\hat{\ell}) = \frac{\text{MSE}(\hat{\ell})}{\ell}, \tag{31}$$

$$\overline{\text{CRLB}}(\hat{\ell}) = \frac{\text{CRLB}(\hat{\ell})}{\ell}. \tag{32}$$

Fig. 4 shows the normalized MSE of the blind estimation method at various SNR levels defined as ρ/N_S, when employing 16QAM and 4096QAM. Since it takes an extremely long time to evaluate the overall MSE for all possible ℓ values, we opted for the scenarios, where $\ell = 1000$ and $\ell = 2048$, respectively, while the impulsive noise occurrence probability is 0.1. Our observations are as follows. Firstly, the proposed blind estimation method has the same performance regardless of the QAM order, because the expectation of $Y_n^* Y_n$ in (23) is 1 for any QAM order and the results are expected to be identical for a sufficiently large number of samples. Secondly,

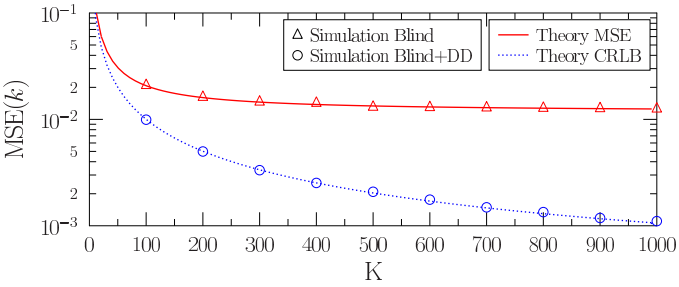


(a)

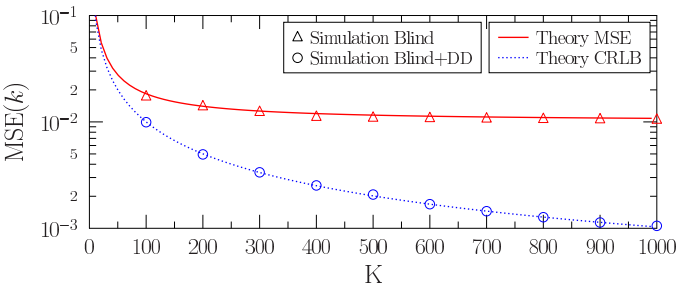


(b)

Fig. 4: The MSE of the blind estimation method when the probability of impulsive noise occurrence is 0.1. (a): $\ell = 1000$; (b): $\ell = 2048$. The lines representing the theory calculated according to (29).



(a)



(b)

Fig. 5: The CRLB and the MSE of the proposed estimation method at SNR = 27dB in 4096QAM. The probability of impulsive noise occurrence is 0.1 (a): $\ell = 1000$; (b): $\ell = 2048$. 'Blind' refers to the blind estimation method and 'Blind+DD' refers to the blind estimation method associated with 3 more iterations in the DD estimation method. The indeces of the samples used for blind estimation span from 1 to 300. The 'Theory MSE' and 'Theory CRLB' were calculated according to (31) and (32), respectively.

the MSE value does not decrease significantly for $K > 300$, which is explained as follows. It can be inferred from (29) that the MSE is partially dependant on the power of impulsive noise, while the power of impulsive noise becomes weaker, when the tone index increases, especially when $K > 300$ as shown in Fig 2. Based on this observation, the samples with indeces spanning from 1 to 300 are selected as $\boldsymbol{Y}_{sel}$ for blind estimation. Thirdly, any increase in the SNR degrades the estimation performance and the normalized MSE associated with $\ell = 2048$ is lower than that of $\ell = 1000$, which can be explained as follows. The estimation process is similar to the maximum-likelihood detection, where the detection accuracy is dependant on the power ratio between the desired variable and the undesired variable. The increase of SNR decreases the ratio of impulsive noise power over the transmitted signal power. Similarly, the impulsive noise has a higher power for $\ell = 2048$ than for $\ell = 1000$. These observations can also be directly inferred from (29).

Fig. 5 presents CRLB and the achievable normalized MSE of the proposed methods, when employing 4096QAM. Again, the impulsive noise occurrence probability is set to 0.1. The label 'Blind' refers to the blind estimation method and 'Blind+DD' refers to the blind estimation method associated with 3 further iterations of the DD estimation method. Note that 4096QAM is considered according to the maximum number of bits (as 12) defined in [5]. Low-order QAM, such as 16QAM, is not considered for comparison with the CRLB, because as shown in Fig. 6, the EXIT chart of 16QAM is almost horizontal, which means that its iterations gain is negligible. Hence 16QAM cannot benefit from the proposed DD estimation. Our observations ares discussed as follows. Firstly, the proposed 'Blind+DD' algorithm outperforms the 'Blind' method and it is capable of achieving the CRLB. Secondly, the normalized MSE of the 'Blind+DD' method becomes lower when we increase the number of samples used for estimation in a single DMT symbol, because the accuracy of the estimation can be improved upon increasing the number of selected samples. Therefore, we opt for 1000 samples for the DD estimation method. Thirdly, the CRLB is not dependent on the value of ℓ, which can also be observed from (30).

B. Data Detection

The corresponding data detection of the receiver is also shown in Fig. 1, which consists of the QAM soft demapper and LDPC decoder. More explicitly, the associated *a priori* information and *extrinsic* information are interleaved and exchanged between the QAM soft demapper and the LDPC decoder L_{itera} times, while the information gleaned from the QAM soft demapper is first deinterleaved and then exchanged L_{LPDC} times within the LDPC decoder. The final hard decision is carried out by the LDPC decoder in order to produce the estimated $\hat{\boldsymbol{b}}$ of the transmitted bits $\boldsymbol{b}$.

For the ease of explanation, the information exchanged between the components during data detection, in terms of the LLRs, as shown in Fig. 1, is defined as follows:

- $\boldsymbol{L_a(d)}$, $\boldsymbol{L_p(d)}$, $\boldsymbol{L_e(d)}$: the *a priori*, *a posteriori* and *extrinsic* LLRs, respectively, associated with the QAM soft demapper.
- $\boldsymbol{L_a(c)}$, $\boldsymbol{L_p(c)}$, $\boldsymbol{L_e(c)}$: the *a priori*, *a posteriori* and *extrinsic* LLRs, respectively, associated with the LDPC decoder.

For Q-ary QAM associated with the alphabet $\boldsymbol{\mathcal{X}}$ as detailed in Section II-A, the *a posteriori* LLRs of the received signal

can be calculated by the log-Maximum A Posterior (Log-MAP) algorithm [28] with the aid of the estimated $\hat{\ell}$, as:

$$\boldsymbol{L_p}(\boldsymbol{d}(i)) = \log\left(\frac{\sum_{\forall X^q \in \boldsymbol{\mathcal{X}}_{\boldsymbol{d}(i)=1}} p(X^q|Y_n)}{\sum_{\forall X^q \in \boldsymbol{\mathcal{X}}_{\boldsymbol{d}(i)=0}} p(X^q|Y_n)}\right) = \log\left(\frac{\sum_{\forall X^q \in \boldsymbol{\mathcal{X}}_{\boldsymbol{d}(i)=1}} p(Y_n|X^q)p(X^q)}{\sum_{\forall X^q \in \boldsymbol{\mathcal{X}}_{\boldsymbol{d}(i)=0}} p(Y_n|X^q)p(X^q)}\right), \quad (33)$$

where $\boldsymbol{\mathcal{X}}_{\boldsymbol{d}(i)=1}$ and $\boldsymbol{\mathcal{X}}_{\boldsymbol{d}(i)=0}$ denote the Q-ary QAM subsets, when the specific bit $\boldsymbol{d}(i)$ is fixed to 1 and 0, respectively. Then $p(Y_n|X^q)$ can be calculated as

$$p(Y_n|X^q) = \frac{1}{\pi(\sigma_S^2 + \frac{\hat{\ell}}{N}\sigma_{I,n}^2)} \exp\Big(-\frac{|Y_n - \sqrt{\rho}X^q|^2}{\sigma_S^2 + \frac{\hat{\ell}}{N}\sigma_{I,n}^2}\Big). \quad (34)$$

During the first inner iteration, we have the *a priori* probability of $p(X^q) = \frac{1}{Q}$ upon assuming that the transmitted symbols are with equal probability. When the soft-demapper receives the *a priori* LLRs $\boldsymbol{L_a}(\boldsymbol{d})$ from the channel decoder of Fig. 1, $p(X^q)$ can be calculated as

$$p(X^q) = \prod_{i=1}^{\text{BPS}} \frac{\exp\left[\boldsymbol{d}^q(i)\boldsymbol{L_a}(\boldsymbol{d})\right]}{1 + \exp\left[\boldsymbol{L_a}\big(\boldsymbol{d}(i)\big)\right]}, \quad (35)$$

where $\{\boldsymbol{d}^q\}_{i=1}^{\text{BPS}}$ refers to the bits mapped to the specific constellation point of $X^q \in \boldsymbol{\mathcal{X}}$.

The *extrinsic* LLRs $\boldsymbol{L_e}(\boldsymbol{d}) = \boldsymbol{L_p}(\boldsymbol{d}) - \boldsymbol{L_a}(\boldsymbol{d})$ gleaned from the QAM soft demapper are deinterleaved and fed in form of the *a priori* LLRs $\boldsymbol{L_a}(\boldsymbol{c})$ into the LDPC decoder of Fig. 1. Then the updated *extrinsic* LLRs $\boldsymbol{L_e}(\boldsymbol{c})$ are fed back and interleaved again as the *a priori* LLRs to the QAM soft demapper of Fig. 1 for the next iteration. After the convergence of the iterations both within the LDPC decoder and between the QAM soft-demapper and the LDPC decoder or after the predefined iteration limits specified by L_{itera} and L_{LDPC} are met, the LDPC decoder will finally carry out a hard-decision yielding $\hat{\boldsymbol{b}}$.

IV. EXIT Chart and BER Performance Evaluation

Let us now investigate the achievable performance of our joint impulsive noise estimation and data detection schemes shown in Fig. 1. A $1/2$-rate 61380-bit LDPC code and the min-sum decoding algorithm having the maximum number of decoding iterations given by 25 and 100 were adopted for the channel code. The samples spanning from 1 to 300 were used for the blind estimator and the samples spanning from 1 to 1000 were used for the DD estimator. The SNR of the system is set to ρ/σ_S^2, which represents the received signal power normalized by the stationary noise of each tone. The achievable performance is assessed through two examples in terms of two metrics: the achievable BER and the EXIT charts, as follows.

1) Example 1: A DMT-based system associated with $N = 2048$ using 16QAM under the impulsive noise model of DT central office (CO) was simulated. An interleaver length of 900 DMT symbols was used.

Fig. 6 plots the EXIT chart of the LDPC decoder and of the soft demapper for different types of impulsive noise variance knowledge, namely for perfect knowledge, for the

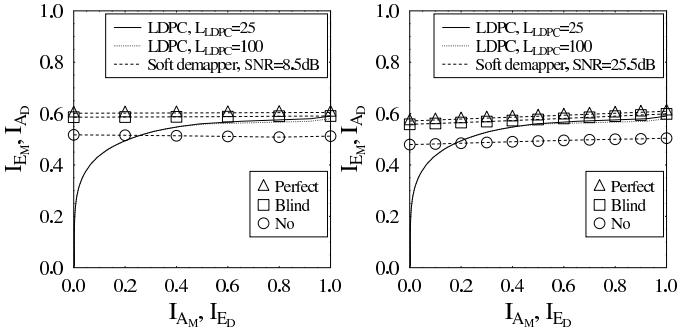


Fig. 6: EXIT chart of the QAM demapper Fig. 1 with different impulsive noise variance knowledge: Perfect (perfect estimation), Blind (blind estimation) and No (no estimation). The impulsive noise is configured according to the statistical model of DT central office (CO). The modulation schemes are (a) 16QAM; (b): 4096QAM. The indeces of the samples used for blind estimation span from 1 to 300. L_{LDPC} refers to the number of iterations within the LDPC decoder.

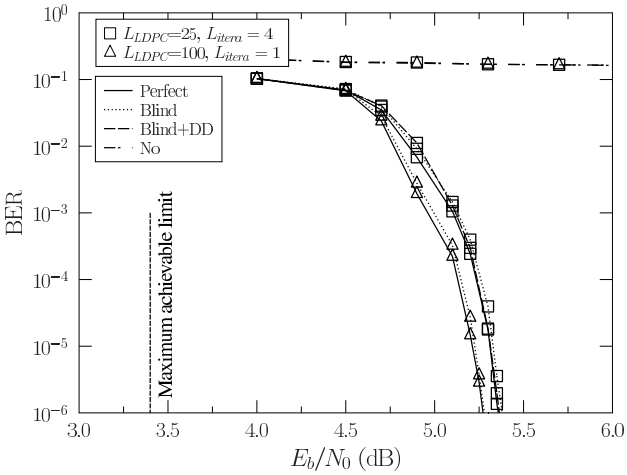


Fig. 7: BER performance of 16QAM for various impulsive noise variance knowledge. The impulsive noise is modelled according to the parameters specified by DT(CO). The interleaver length is 900 DMT symbols. The lines correspond to the different types of impulsive noise variance knowledge: 'Perfect' (perfect knowledge), 'Blind' (Blind estimation without DD estimation, 'Blind+DD' (blind estimation and 3 more DD estimation) and 'No' (no knowledge).

knowledge extracted from the proposed blind estimator and no knowledge. The selected samples used for blind estimation are those spanning from 1 to 300. The impulsive noise model is parametrized by DT(CO) as shown in Fig 3. Our observations are listed as follows. Firstly, the EXIT curve of the soft demapper gleaning its knowledge from the blind estimator is above that of the scenario operating without estimation and it is quite close to the case with perfect knowledge. Hence, as expected the blind estimation method is capable of providing a more reliable LLR for the soft demapper. Secondly, the EXIT curve of the 16QAM soft-demapper is almost horizontal, while the *extrinsic* LLR of 4096QAM soft demapper increases visibly upon the increasing the *apriori* LLRs. This is because 16QAM has no iteration gains, while 4096QAM benefits from increasing the number of the iterations between the soft demapper and the channel decoder. Thirdly, as shown in Fig. 6b, there is still a gap in the EXIT curve of the QAM soft

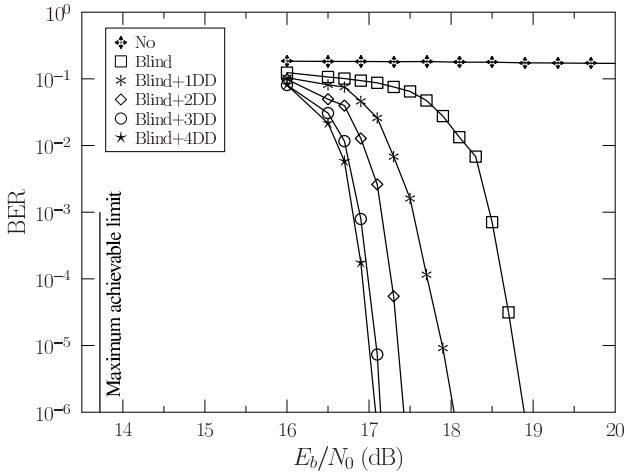


Fig. 8: BER performance of 4096QAM for various impulsive noise variance estimation methods. The impulsive noise is modelled according to the parameters specified by DT(CO). The interleaver length is 900 DMT symbols. The symbols represent the different types of impulsive noise estimations: 'No' (no estimation), 'Blind' (blind estimation), 'Blind+jDD' (blind estimation with j more iterations DD estimation). The outer iterations for each case can be calculated as $L_{itera} = j + 1$.

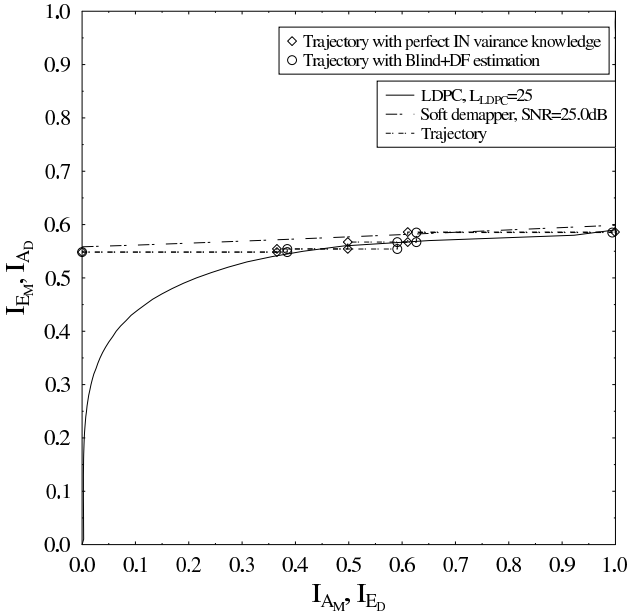


Fig. 9: EXIT chart of 4096QAM for perfect impulsive noise variance knowledge. The impulsive noise is modelled according to the parameter specified by DT(CO). The interleaver length is 900 DMT symbols. 'Blind+DD' refers to blind estimation with 3 additional iterative DD estimations.

demapper having perfect knowledge and that relying on the blind estimation. Hence, the DD estimator can be incorporated into the iterative loop of high-order QAM schemes during further iterations. Fourthly, increasing the number of inner iterations within the LDPC decoder from 25 to 100 improves its EXIT-chart performance albeit the improvement is not very significant.

As shown in Fig. 6a, the EXIT curve of the 16QAM soft-demapper is almost horizontal, which means that the iteration gain of 16QAM is negligible. Therefore, it is desired to opt for a higher number L_{LDPC} of inner iterations and lower number L_{itera} of outer iterations. The BER performance of the proposed joint impulsive noise estimation and data detection scheme is shown in Fig. 7, in comparison to that of the perfect variance knowledge and of no knowledge of the impulsive noise variance. Our observations are discussed as follows. Firstly, the error floor is formed for the system without impulsive noise estimation, which is explained as follows. The channel state information and noise variance level as well as the received signal are fed into the QAM soft-demapper at the receiver of Fig. 1. If we rely on using the stationary noise variance level to detect the samples suffering from impulsive noise, unreliable LLRs will be calculated, which may result in error bursts. Secondly, our proposed joint impulsive noise estimation and data detection aided 16QAM has a similar BER performance to that associated with perfect knowledge of the impulsive noise variance and exhibits a significant improvement over that operating without IN variance estimation. Thirdly, the E_b/N_0 gap between our proposed scheme and the maximum achievable limit is as low as 1.7 dB, hence our scheme is capable of attaining a near-capacity performance.

2) Example 2: The system setup is identical to that of Example 1, except that 4096QAM was employed. As discussed above, 4096QAM has a significant iteration gain and hence it is desired to use a higher number L_{itera} of outer iterations and a lower number L_{LDPC} of the inner iterations.

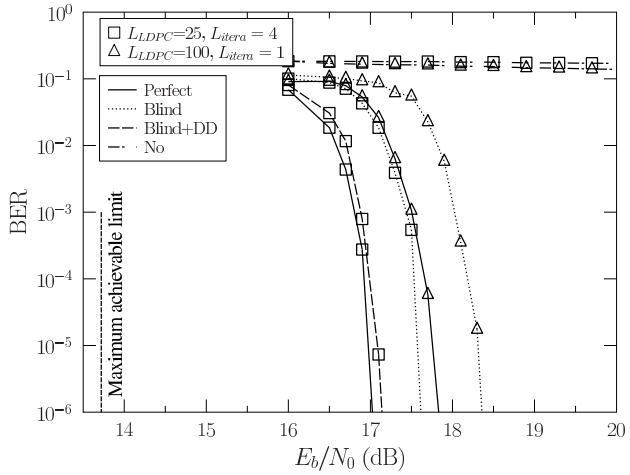


Fig. 10: BER performance of 4096QAM for various impulsive noise variance provision methods. The impulsive noise is modelled according to the parameters specified by DT(CO). The lines in the figure correspond to different types of impulsive noise variance knowledge: 'Perfect' (perfect knowledge), 'Blind' (Blind estimation without DD estimation, 'Blind+DD' (blind estimation and 3 more DD estimation) and 'No' (no knowledge).

We can select the optimal number of outer iterations with the aid of Fig. 8. On one hand, it can be observed that our proposed scheme achieves a better BER performance, when we increase the number L_{itera} of outer iterations. On the other hand, the improvement becomes smaller, when L_{itera} increases. To elaborate a little further, the SNR gap between 'Blind+3DD' and 'Blind+4DD' is small, but 'Blind+4DD' requires a further outer iteration. Considering the trade off between the complexity and the BER performance, the system 'Blind+3DD' is deemed to be the most attractive, hence it is selected for our further investigations.

Fig. 9 plots our EXIT chart analysis, where the soft-demapper is fed by our proposed estimation method, or

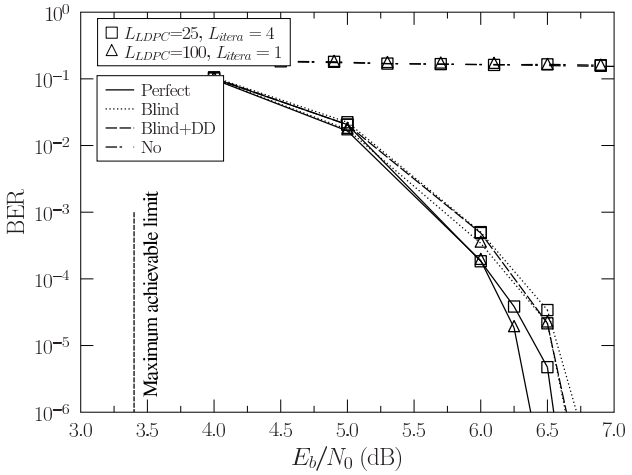


Fig. 11: BER performance of 16QAM for various impulsive noise variance estimation methods. The impulsive noise is modelled according to the parameters specified by DT(CO). The interleaver length is 90 DMT symbols. The lines in the figure correspond to the different types of impulsive noise variance knowledge: 'Perfect' (perfect knowledge), 'Blind' (Blind estimation without DD estimation, 'Blind+DD' (blind estimation and 3 more DD estimation) and 'No' (no knowledge).

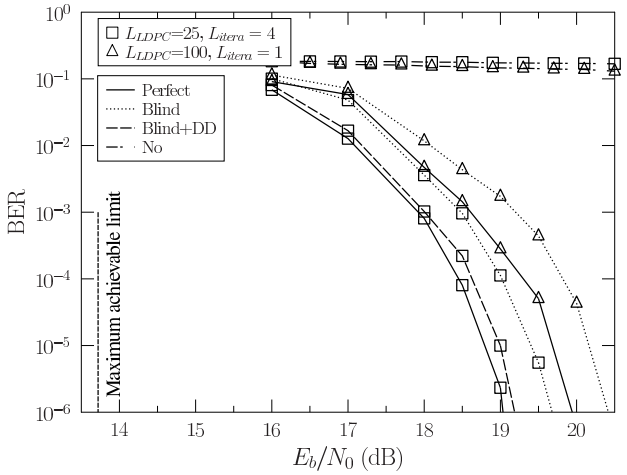


Fig. 12: BER performance of 4096QAM for various impulsive noise variance estimation methods. The impulsive noise is modelled according to the parameters specified by DT(CO). The interleaver length is 90 DMT symbols. The lines in the figure correspond to the different types of impulsive noise variance knowledge: 'Perfect' (perfect knowledge), 'Blind' (Blind estimation without DD estimation, 'Blind+DF' (blind estimation and 3 more DD estimation) and 'No' (no knowledge).

alternatively, it relies on perfect impulsive noise variance knowledge, respectively. Since the EXIT curve of the soft-demapper associated with the DD estimator cannot be plotted, only the EXIT curve of perfect knowledge is plotted. It can be inferred that an open tunnel emerges at SNR $= 25$ dB. The two stair-case shaped decoding trajectories were recorded for the soft-demapper fed with the exact impulsive noise variance information, or by that estimated using our proposed algorithm. The point of convergence around $(1.0, 0.6)$ is reached by both types of variance estimation schemes relying on $L_{\text{itera}} = 4$ iterations, which means that the proposed scheme is capable of approaching the optimal ML detection performance recorded for perfect variance knowledge at the same number of iterations.

Fig. 10 shows the BER performance of different combinations of our proposed joint impulsive noise estimation and data detection scheme under the DT(CO) impulsive noise model, in comparison to both the perfect knowledge and no knowledge scenarios. Our observations are listed as follows. Firstly, the E_b/N_0 gap between the schemes fed with our 'Blind+DD' estimation results and with perfect impulsive noise variance knowledge is less 0.1dB, when we have $L_{\text{itera}} = 4$ and $L_{\text{LDPC}} = 25$. Secondly, it requires 1.2 dB lower SNR for the 'Blind+DD' method to attain BER $= 10^{-6}$ than for the 'Blind' method, which means that our proposed DD estimator beneficially improves the performance. Thirdly, the E_b/N_0 gap between our proposed scheme and the maximum achievable limit is 3.4 dB at BER $= 10^{-6}$. Therefore we conclude that our proposed joint estimation and decoding relying on the 'Blind+DD' scheme is capable of achieving a near-capacity performance.

3) Effect of interleaver length: Since using the interleaver length of 900 DMT symbols is not plausible for some application scenarios, we investigate the effect of the interleaver length by simulations. The system setup is identical to the case of Example 1 and Example 2, except for employing an interleaver length of 90 DMT symbols, as seen in Fig. 11 and Fig. 12 for 16QAM and 4096QAM, respectively. The different schemes have the same trends as in the above examples. Comparing them to the corresponding performance in Fig. 7 an Fig. 10, it can be inferred that the system associated with an interleaver length of 90 DMT symbols requires about 1.3 dB and 2 dB higher E_b/N_0 to achieve the point of BER $= 10^{-6}$, which is explained as follows. Even if the reliable LLRs of the samples suffering from impulsive noise can be calculated by our proposed estimation algorithm, the LLRs still remain relatively small and the corresponding bits have to be recovered with the aid of the neighbouring LLRs. The system having a shorter interleaver length sometimes fails to spread these LLRs uniformly, hence the channel decoder cannot recover all the transmitted bits. However, compared to the BER performance recorded without impulsive noise estimation, our proposed method is still capable of significantly improving the system performance.

V. Conclusions

The major challenge in coded DSL systems operating in the presence of impulsive noise has been the acquisition of the reliable LLRs from the samples suffering from impulsive noise. We have proposed a two-stage impulsive noise estimation algorithm for determining the arrival instant and the variance of impulsive noise, so that reliable LLRs of the samples inflicting impulsive noise can be calculated. The MSE and CRLB results characterize the accuracy of the blind estimation method and of our DD estimation method. Based on the results, we also select the appropriate samples for two different estimation methods in our system considered. We have also optimized the joint estimation and data detection scheme conceived for our LDPC coded DMT-based DSL systems two different estimation methods in our system considered, with the aid of EXIT chart. Our extensive simulation results have

confirmed that the proposed scheme is capable of providing accurate impulsive noise variance estimation, and exhibiting a near-capacity performance associated with the idealized perfect impulsive noise variance knowledge at a low number of iterations.

APPENDIX A
DERIVATION OF THE MSE

Here let us introduce the short-hand of $\Omega = \frac{1}{N}\sum \sigma_{I,n}^2$ and $\Psi = \rho K + \sigma_S^2 K$. Then MSE of the estimated value $\hat{\ell}$ is expressed as

$$\begin{aligned}\mathrm{MSE}_{\hat{\ell}} &= \mathbb{E}\{(\ell - \hat{\ell})^2\} \\ &= \mathbb{E}\Big\{\Big(\ell - \frac{\boldsymbol{Y}_{sel}^H \boldsymbol{Y}_{sel}}{\Omega} + \frac{\Psi}{\Omega}\Big)^2\Big\} \\ &= \mathbb{E}\Big\{\Big(\ell + \frac{\Psi}{\Omega}\Big)^2\Big\} - 2\mathbb{E}\Big\{\Big(\ell + \frac{\Psi}{\Omega}\Big)\frac{\boldsymbol{Y}_{sel}^H \boldsymbol{Y}_{sel}}{\Omega}\Big\} \\ &\quad + \mathbb{E}\Big\{\Big(\frac{\boldsymbol{Y}_{sel}^H \boldsymbol{Y}_{sel}}{\Omega}\Big)^2\Big\}. \end{aligned} \tag{36}$$

Let us denote the first, second and third term of (36) by T_1, T_2 and T_3, respectively. Then T_2 becomes

$$\begin{aligned}&2\mathbb{E}_{(\ell,\rho,\sigma_S,\sigma_I,\boldsymbol{X})}\Big\{\Big(\ell + \frac{\Psi}{\Omega}\Big)\frac{\boldsymbol{Y}_{sel}^H \boldsymbol{Y}_{sel}}{\Omega}\Big\} \\ &= 2\mathbb{E}_{(\ell,\rho,\sigma_S,\sigma_I)}\Big\{\Big(\ell + \frac{\Psi}{\Omega}\Big)\mathbb{E}_{\boldsymbol{X}}\Big\{\frac{\boldsymbol{Y}_{sel}^H \boldsymbol{Y}_{sel}}{\Omega}\Big\}\Big\} \\ &= 2\mathbb{E}\Big\{\Big(\ell + \frac{\Psi}{\Omega}\Big)^2\Big\}\Big\}. \end{aligned} \tag{37}$$

Furthermore, T_3 can be expanded as

$$\begin{aligned}&\mathbb{E}\Big\{\Big(\frac{\boldsymbol{Y}_{sel}^H \boldsymbol{Y}_{sel}}{\Omega}\Big)^2\Big\} \\ &= \frac{1}{\Omega^2}\mathbb{E}\Big\{\Big(\rho\sum X_n^* X_n\Big)^2 + 2\Big(\rho\sum X_n^* X_n\Big) \\ &\quad \times \Big(N_{S,n}^* N_{S,n} + N_{I,n}^* N_{I,n} + 2\Re(N_{I,n}^* N_{S,n}) \\ &\quad + 2\sqrt{\rho}\Re(X_n^* N_{S,n}) + 2\sqrt{\rho}\Re(X_n^* N_{I,n})\Big) \\ &\quad + \Big(N_{S,n}^* N_{S,n} + N_{I,n}^* N_{I,n} + 2\Re(N_{I,n}^* N_{S,n}) \\ &\quad + 2\sqrt{\rho}\Re(X_n^* N_{S,n}) + 2\sqrt{\rho}\Re(X_n^* N_{I,n})\Big)^2\Big\} \\ &= \frac{\rho^2 K^2 + 2\rho K^2\sigma_S^2 + 2\rho K\Omega\mathbb{E}[\ell]}{\Omega^2} \\ &\quad + \frac{1}{\Omega^2}\mathbb{E}\Big\{\Big(N_{S,n}^* N_{S,n} + N_{I,n}^* N_{I,n} \\ &\quad + 2\Re(N_{I,n}^* N_{S,n}) + 2\sqrt{\rho}\Re(X_n^* N_{S,n}) \\ &\quad + 2\sqrt{\rho}\Re(X_n^* N_{I,n})\Big)^2\Big\}. \end{aligned} \tag{38}$$

The second term of (38), denoted by T_{32} can then be expanded as

$$\begin{aligned}T_{32} &= \frac{1}{\Omega^2}\Big\{\mathbb{E}\Big\{\big(\sum N_{S,n}^* N_{S,n}\big)^2\Big\} + \mathbb{E}\Big\{\big(\sum N_{I,n}^* N_{I,n}\big)^2\Big\} \\ &\quad + \mathbb{E}\Big\{2\big(\sum N_{S,n}^* N_{S,n}\big)\big(\sum N_{I,n}^* N_{I,n}\big)\Big\} \\ &\quad + \mathbb{E}\Big\{4\rho\Big[\sum \Re(X_n^* N_{S,n})\Big]^2\Big\} \\ &\quad + \mathbb{E}\Big\{4\rho\Big[\sum \Re(X_n^* N_{I,n})\Big]^2\Big\} \\ &\quad + \mathbb{E}\Big\{4\Big[\sum \Re(N_{S,n}^* N_{I,n})\Big]\Big\}\Big\} \\ &= \frac{1}{\Omega^2}\Big\{\mathrm{MSE}_1 + \mathrm{MSE}_2 + \mathrm{MSE}_3 + \mathrm{MSE}_4 \\ &\quad + \mathrm{MSE}_5 + \mathrm{MSE}_6\Big\}. \end{aligned} \tag{39}$$

To expound further, MSE_1 can be simplified as

$$\begin{aligned}\mathrm{MSE}_1 &= \mathbb{E}\Big[\sum |N_{S,n}|^4\Big] + \mathbb{E}\Big[\sum |N_{S,n}|^2 \sum_{n\neq m} |N_{S,m}|^2\Big] \\ &= 2K\sigma_S^2 + (K^2 - K)\sigma_S^2 \\ &= (K^2 + K)\sigma_S^2. \end{aligned} \tag{40}$$

Similarly, MSE_2 can be simplified as

$$\begin{aligned}\mathrm{MSE}_2 &= \mathbb{E}\Big[\sum |N_{I,n}|^4\Big] + \mathbb{E}\Big[\sum |N_{I,n}|^2 \sum_{n\neq m} |N_{I,m}|^2\Big] \\ &= \frac{2\mathbb{E}[\ell^2]}{N}\sum \sigma_{I,n}^4 + \frac{\mathbb{E}[\ell^2]}{N^2}\sum_{n=M}^{M+K-1} \sigma_{I,n}^2\big(\sum_{n\neq m}\sigma_{I,j}^2\big) \\ &= \frac{\ell^2}{N}\sum \sigma_{I,n}^4 + \Omega^2\ell^2. \end{aligned} \tag{41}$$

While MSE_3 can be simplified to

$$\mathrm{MSE}_3 = 2K\sigma_S^2\Omega\ell. \tag{42}$$

Substituting $\Re(X_n^* N_{S,n}) = \Re(X_n^*)\Re(N_{S,n}) + \Im(X_n^*)\Im(N_{S,n})$ into MSE_4, we have

$$\begin{aligned}\mathrm{MSE}_4 &= 4\rho\mathbb{E}\Big\{\sum(\Re(X_n^*))^2(\Re(N_{S,n}))^2\Big\} \\ &\quad + 4\rho\mathbb{E}\Big\{\sum(\Im(X_n^*))^2(\Im(N_{S,n}))^2\Big\} \\ &= \sum\frac{1}{2}\frac{\sigma_S^2}{2} + \sum\frac{1}{2}\frac{\sigma_S^2}{2} \\ &= 2\rho K\sigma_S^2. \end{aligned} \tag{43}$$

Similarly, substituting $\Re(X_n^* N_{I,n}) = \Re(X_n^*)\Re(N_{I,n}) + \Im(X_n^*)\Im(N_{I,n})$ into MSE_4, we have

$$\begin{aligned}\mathrm{MSE}_5 &= 4\rho\mathbb{E}\Big\{\sum(\Re(X_n^*))^2(\Re(N_{I,n}))^2\Big\} \\ &\quad + 4\rho\mathbb{E}\Big\{\sum(\Im(X_n^*))^2(\Im(N_{I,n}))^2\Big\} \\ &= 2\rho\Omega\ell. \end{aligned} \tag{44}$$

Similarly,

$$\mathrm{MSE}_6 = 2\sigma_S^2\Omega\ell. \tag{45}$$

In this way, it is easy to get T_{32} from all the MSE_i contributions. Summing up T_1, T_2 and T_3, we arrive at the MSE of the estimated $\hat{\ell}$ as

$$\text{MSE}(\hat{\ell}) = \frac{\sum \sigma_{I,n}^4}{N^2\Omega^2}\ell^2 + \frac{2(\rho+\sigma_S^2)}{\Omega}\ell + \frac{K\sigma_S^4}{\Omega^2}. \quad (46)$$

Since the stationary noise power is much lower than that of the impulsive noise, the third term is negligible. The first term dominates $\text{MSE}_{\hat{\ell}}$, when ρ is relatively small. The second term has a significant impact on the $\text{MSE}_{\hat{\ell}}$, when ρ is increasing. Therefore, the proposed estimation method works well in the relatively low SNR range.

APPENDIX B
DERIVATION OF THE CRLB

For the tone having an index n in the DMT symbol, the received signal is expressed as

$$\begin{aligned} Y_n &= \sqrt{\rho}X_n + N_{I,n} + N_{S,n} \\ &= \sqrt{\rho}X_n + \sqrt{\frac{\ell}{N}}\cdot\sigma_{I,n}\cdot W_{I,n} + \sigma_S\cdot W_{S,n}, \end{aligned} \quad (47)$$

where $W_{I,n}$ and $W_{S,n}$ are zero-mean unit-variance Gaussian variables. Then it is easy to obtain

$$Y_n - \sqrt{\rho}X_n = \sqrt{\frac{\ell}{N}}\sigma_{I,n}W_{I,n} + \sigma_S\cdot W_{S,n}. \quad (48)$$

Then $Y_n - \sqrt{\rho}X_n$ is Gaussian with zero mean and an $(K\times K)$-element covariance matrix. Let us denote the covariance of the left side as $\mathbf{C}(\ell)$, whose $[i,j]$th element is

$$\begin{aligned} [\mathbf{C}(\ell)]_{ij} &= \mathbb{E}\Big[\Big(\sqrt{\frac{\ell}{N}}\sigma_{I,i}W_{I,i} + \sigma_S\cdot W_{S,i}\Big) \\ &\qquad \Big(\sqrt{\frac{\ell}{N}}\sigma_{I,j}W_{I,j} + \sigma_S\cdot W_{S,j}\Big)^*\Big] \\ &= \frac{\ell}{N}\sigma_{I,i}^2\delta_{ij} + \sigma_S^2\delta_{ij}. \end{aligned} \quad (49)$$

Therefore,

$$\begin{aligned} \mathbf{C}(\ell) = \text{diag}\Big\{&\frac{\ell\sigma_{I,M}^2}{N} + \sigma_S^2, \frac{\ell\sigma_{I,M+1}^2}{N} + \sigma_S^2, \\ &\cdots, \frac{\ell\sigma_{I,M+K-1}^2}{N} + \sigma_S^2\Big\}, \end{aligned} \quad (50)$$

where $\text{diag}\{\boldsymbol{a}\}$ refers to the diagonal matrix with elements in $\boldsymbol{a}$ on its diagonal. We have

$$\begin{aligned} \mathbf{C}^{-1}(\ell) = \text{diag}\Big\{&\frac{1}{\frac{\ell}{N}\sigma_{I,M}^2 + \sigma_S^2}, \frac{1}{\frac{\ell}{N}\sigma_{I,M+1}^2 + \sigma_S^2}, \\ &\cdots, \frac{1}{\frac{\ell}{N}\sigma_{I,M+K-1}^2 + \sigma_S^2}\Big\}. \end{aligned} \quad (51)$$

Furthermore, since

$$\begin{aligned} \frac{\partial\mathbf{C}(\ell)}{\partial\ell} = \text{diag}\Big\{&\frac{1}{N}\sigma_{I,M}^2, \frac{1}{N}\sigma_{I,M+1}^2, \\ &\cdots, \frac{1}{N}\sigma_{I,M+K-1}^2\Big\}. \end{aligned} \quad (52)$$

we have

$$\begin{aligned} \mathbf{C}^{-1}(\ell)\frac{\partial\mathbf{C}(\ell)}{\partial\ell} = \text{diag}\Big\{&\frac{\sigma_{I,M}^2}{\ell\sigma_{I,M}^2 + N\sigma_S^2}, \frac{\sigma_{I,M+1}^2}{\ell\sigma_{I,M+1}^2 + N\sigma_S^2}, \\ &\cdots, \frac{\sigma_{I,M+K-1}^2}{\ell\sigma_{I,M+K-1}^2 + N\sigma_S^2}\Big\}. \end{aligned} \quad (53)$$

According to [29], the Fisher information matrix is expressed as

$$\begin{aligned} \boldsymbol{I}(\ell) &= \text{tr}\Big[\Big(\mathbf{C}^{-1}(\ell)\frac{\partial\mathbf{C}(\ell)}{\partial\ell}\Big)^2\Big] \\ &= \sum_{n=M}^{M+K-1}\frac{(\sigma_{I,n}^2)^2}{(\ell\sigma_{I,n}^2 + N\sigma_S^2)^2}. \end{aligned} \quad (54)$$

Therefore, the CRLB becomes:

$$\begin{aligned} \text{CRLB}(\hat{\ell}) &= \frac{1}{\boldsymbol{I}(\ell)} \\ &= \frac{1}{\sum_{n=M}^{M+K-1}\frac{(\sigma_{I,n}^2)^2}{(\ell\sigma_{I,n}^2 + N\sigma_S^2)^2}}. \end{aligned} \quad (55)$$

REFERENCES

[1] T. Starr, J. M. Cioffi, and P. J. Silverman, *Understanding digital subscriber line technology*. Prentice Hall PTR, 1999.

[2] "Digital transmission system on metallic local lines for ISDN basic rate access," *ITU-T G.961*, 1993.

[3] "Asymmetric digital subscirber line (ADSL) transceivers," *ITU-T G.992.1*, 1999.

[4] "Very high speed digital subscirber line (VDSL) transceivers," *ITU-T G.993.1*, 2005.

[5] "Fast access to subscriber termials (G.fast) - Physical Layer Specification," *Recommendation Draft ITU-T G.9701*, 2014.

[6] P. Ödling, T. Magesacher, S. Höst, P. O. Börjesson, M. Berg, and E. Areizaga, "The fourth generation broadband concept," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 62–69, 2009.

[7] M. Timmers, M. Guenach, C. Nuzman, and J. Maes, "G. fast: evolving the copper access network," *IEEE Communications Magazine*, vol. 51, no. 8, pp. 74–79, 2013.

[8] V. Oksman, R. Strobel, X. Wang, D. Wei, R. Verbin, R. Goodson, and M. Sorbara, "The ITU-T's new G.fast standard brings DSL into the gigabit era," *IEEE Communications Magazine*, vol. 54, pp. 118–126, March 2016.

[9] R. Cendrillon, G. Ginis, E. V. D. Bogaert, and M. Moonen, "A near-optimal linear crosstalk canceler for upstream VDSL," *IEEE Transactions on Signal Processing*, vol. 54, pp. 3136–3146, Aug 2006.

[10] G. Ginis and J. M. Cioffi, "Vectored transmission for digital subscriber line systems," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 5, pp. 1085–1104, 2002.

[11] R. Zhang, A. F. A. Rawi, L. D. Humphrey, and L. Hanzo, "Expanded constellation mapping for enhanced far-end-cross-talk cancellation in G.fast," *IEEE Communications Letters*, vol. 21, pp. 56–59, Jan 2017.

[12] W. Yu, G. Ginis, and J. M. Cioffi, "Distributed multiuser power control for digital subscriber lines," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 5, pp. 1105–1115, 2002.

[13] Y. Xu, T. Le-Ngoc, and S. Panigrahi, "Global concave minimization for optimal spectrum balancing in multi-user DSL networks," *IEEE Transactions on Signal Processing*, vol. 56, pp. 2875–2885, July 2008.

[14] D. Toumpakaris, J. M. Cioffi, and D. Gardan, "Reduced-delay protection of DSL systems against nonstationary disturbances," *IEEE Transactions on Communications*, vol. 52, pp. 1927–1938, Nov 2004.

[15] D. Zhang, K. Ho-Van, and T. Le-Ngoc, "Impulse noise detection techniques for retransmission to reduce delay in DSL systems," in *2012 IEEE International Conference on Communications (ICC)*, pp. 3160–3164, IEEE, 2012.

[16] J. Neckebroek, M. Moeneclaey, M. Guenach, M. Timmers, and J. Maes, "Comparison of error-control schemes for high-rate communication over short DSL loops affected by impulsive noise," in *IEEE International Conference on Communications (ICC)*, pp. 4014–4019, June 2013.

[17] S. V. Zhidkov, "Analysis and comparison of several simple impulsive noise mitigation schemes for OFDM receivers," *IEEE Transactions on Communications*, vol. 56, pp. 5–9, January 2008.
[18] S. V. Zhidkov, "Impulsive noise suppression in OFDM-based communication systems," *IEEE Transactions on Consumer Electronics*, vol. 49, pp. 944–948, Nov 2003.
[19] G. Caire, T. Y. Al-Naffouri, and A. K. Narayanan, "Impulse noise cancellation in OFDM: an application of compressed sensing," in *2008 IEEE International Symposium on Information Theory*, pp. 1293–1297, July 2008.
[20] E. Eleftheriou, S. Olcer, and H. Sadjadpour, "Application of capacity approaching coding techniques to digital subscriber lines," *IEEE Communications Magazine*, vol. 42, pp. 88–94, Apr 2004.
[21] J. S. Chow, J. C. Tu, and J. M. Cioffi, "A discrete multitone transceiver system for HDSL applications," *IEEE Journal on Selected Areas in Communications*, vol. 9, pp. 895–908, Aug 1991.
[22] W. Henkel, T. Kessler, and H. Y. Chung, "Coded 64-CAP ADSL in an impulse-noise environment-modeling of impulse noise and first simulation results," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 9, pp. 1611–1621, 1995.
[23] I. Mann, S. McLaughlin, W. Henkel, R. Kirkby, and T. Kessler, "Impulse generation with appropriate amplitude, length, inter-arrival, and spectral characteristics," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 5, pp. 901–912, 2002.
[24] W. Henkel and T. Keßler, "A wideband impulsive noise survey in the german telephone network: statistical description and modeling," *Archiv fur Elektronik und Ubertragungstechnik*, vol. 48, pp. 277–277, 1994.
[25] M. Peligrad and W. B. Wu, "Central limit theorem for Fourier transforms of stationary processes," *The Annals of Probability*, pp. 2009–2022, 2010.
[26] T. Kessler, R. Kirkby, W. Henkel, and S. McLaughlin, "Text for 'realistic impulsive noise model'," *ETSI TM6 011T20*, February 2001.
[27] T. Bai, H. Zhang, R. Zhang, L. L. Yang, A. F. A. Rawi, J. Zhang, and L. Hanzo, "Discrete multi-tone digital subscriber loop performance in the face of impulsive noise," *IEEE Access*, vol. 5, pp. 10478–10495, 2017.
[28] L. Hanzo, O. Alamri, M. El-Hajjar, and N. Wu, *Near-capacity multi-functional MIMO systems: sphere-packing, iterative detection and co-operation*, vol. 4. John Wiley & Sons, 2009.
[29] S. M. Kay, *Fundamentals of Statistical Signal Processing: Practical Algorithm Development*, vol. 3. Pearson Education, 2013.