Publication information

Title	Robust Decoding for Convolutionally Coded Systems Impaired by Memoryless Impulsive Noise
Author(s)	Tseng, Der-Feng; Han, Yungshiang Sam; Mow, Wai Ho; Chen, Po-Ning; Deng, Jing; Vinck, Adrianus Johannes Han
Source	IEEE Transactions on Communications , v. 61, (11), November 2013, p. 4640-4652
Version	Pre-published version
DOI	https://doi.org/10.1109/TCOMM.2013.101813.130122
Publisher	IEEE

Copyright information

Copyright © 2013 IEEE. This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of The Hong Kong University of Science and Technology's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

Notice

This version is available at HKUST Institutional Repository via

<http://hdl.handle.net/1783.1/61901>

If it is the author's pre-published version, changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published version.

Robust Decoding for Convolutionally Coded Systems Impaired by Memoryless Impulsive Noise

Der-Feng Tseng, *Member, IEEE*, Yunghsiang S. Han, *Fellow, IEEE*, Wai Ho Mow, *Senior Member, IEEE*, Po-Ning Chen, *Senior Member, IEEE*, and Jing Deng, *Member, IEEE*, A. J. Han Vinck, *Fellow, IEEE*

Abstract

It has been known that communication systems are susceptible to strong impulsive noises. To combat this, convolutional coding has long served as a cost-effective tool in the context of moderately frequent occurrence of memoryless impulses with given statistics. Nevertheless, the impulsive noise statistics is hard to be accurately modeled and is generally not time-invariant, making the respective system design challenging. In this article, in the absence of full knowledge of the probability density function (PDF) of impulsive noises, we devise an efficient decoding scheme for single-carrier narrowband communication systems by incorporating a design parameter into the recently introduced joint erasure marking and

Tseng and Han are with the Dept. of Electrical Engineering, National Taiwan Univ. of Sci. and Tech., Taipei, Taiwan. (e-mail: dtseng@mail.ntust.edu.tw). Mow is with the Dept. of Electronic and Computer Engineering, Hong Kong Univ. of Sci. and Tech., Hong Kong, Chen is with the Dept. of Electrical and Computer Engineering, National Chiao Tung Univ., Hsinchu, Taiwan, Deng is with the Dept. of Computer Science, Univ. of North Carolina, Greensboro, North Carolina and Vinck is with the Institute for Experimental Mathematics, University of Duisburg-Essen, Essen, Germany.

Viterbi decoding algorithm, dubbed the *metric erasure Viterbi algorithm* (MEVA). The proposed scheme is equivalent to incorporating a well-designed clipping operation into the Viterbi algorithm, of which the clipping threshold has to be appropriately set. In contrast with existing publications that often resort to extensive simulations, we characterize the bit error probability performance associated with the clipping threshold by deriving its Chernoff bound. Simulation results reveal that with a judicious selection of the clipping threshold, the MEVA can be on par with its optimal maximum-likelihood decoding counterpart under fairly general circumstances.

I. INTRODUCTION

The challenging problem of coded transmission over communication systems impaired by non-Gaussian noises has received renewed interests in the communication community. Some of the primary impetuses are that the capacity-achieving coding is already relatively well understood in the context of the additive white Gaussian noise (AWGN) model, and that nuisances, like man-made electromagnetic interferences and atmospheric noises, are actually the limiting factors in many realistic situations such as power line communications (PLC) [1], digital subscriber line (DSL) loops [2], and wireless communication systems [3]. Communication system designs and their performance investigations for non-Gaussian noise models are thus much needed to be reviewed. The ensuing models for man-made noises, widely adopted for performance analysis in the literature, can be found in [4] and [5]. Particularly, the impact on the performance of binary signaling and quadrature amplitude modulation under the impulsive noise model of [6] was studied in both [7] and [8], provided that the probability density function (PDF) of the impulsive noise is available at the receiver.

In the literature, channel coding [1], [9] has been employed and validated to be able to effectively withstand the obstruction caused by additive white impulsive noise. In addition to using channel coding to compensate for the loss due to the impulsive noise, orthogonal frequency-division multiplexing (OFDM) technique, as a multi-carrier modulation scheme, was shown to be resistant against impulses under certain circumstances [5]. Noteworthy, as also revealed in [5],

an OFDM system is outperformed by its single-carrier counterpart in terms of uncoded symbol error probability in the context of significant impulsive noise energy coupled with moderately or severely high probability of occurrence of impulses. Another study [10] has shown that the information rate of OFDM systems is lower than that of their single-carrier counterparts except for systems operating at a very high spectral efficiency. All these facts motivate us to investigate the coded single-carrier narrowband communication system subject to frequent occurrence of impulses, of which the average power is much greater than that of the background noise.

There are existing systems that fall in the category of communications system that our work considers. An example belongs to the so-called Low Data Rate (LDR) Narrowband Power Line Communications (NB-PLC), which are typically single carrier based and transmit data at rates of few kilobits per second (kbps). The NB-PLC include devices conforming to the following standards: ISO/IEC 14908-3 (LonWorks), ISO/IEC, 14543-3-5 (KNX), CEA-600.31 (CEBus), IEC 61334 (FSK and Spread-FSK), etc. Additional non-SDO-based examples are Insteon, X10, HomePlug C&C, SITRED, Ariane, Controls, BacNet, etc. Moreover, the above transceivers (mainly LonWorks and IEC 61334) are the most deployed ones in the world, perhaps in the order of hundreds of millions. As opposed to the LDR NB-PLC, multi-carrier (OFDM) based standards, referred to as High Data Rate (HDR) NB-PLC, allow much higher data rates (typically up to few hundred kbps). Examples of the HDR NB-PLC include G3-PLC (ITU-T Recommendation G.9903) and PRIME (ITU-T Recommendation G.9904), which are both ratified in 2012.

A. Related work

As mentioned previously, channel coding has been widely used for systems transceiving in an impulsive noise environment. For instance, convolutional coding is applied in [11], turbo coding in [12], and low density parity-check coding in [13]. For more details, readers can refer to Chapter 5 in [14] and the references therein. Performance limits for communication systems corrupted by the impulsive noise are meticulously derived in [10], [15], which provides insights

into the capacity loss incurred for any robust detection schemes such as the ones in [16], [17]. For a Discrete Multitone (DMT) system under an impulsive noise channel model conceived as the concatenation of an AWGN channel with an erasure channel, channel capacity is derived in [18], where a capacity-approaching LDPC code is also pursued. Under the assumption that the PDF of the impulsive noise is known at the receiver, the Chernoff bounds on the pairwise error probability (PEP) for optimal and suboptimal receivers in real and complex Middleton Class-A channels were investigated in [1] and [9], based on which a coding scheme was also devised in [1]. Further, a general expression of the exact PEP in certain impulsive noise channels was later given in [19]. The Chernoff bound on the bit error probability (BEP) for ideally interleaved coded OFDM system was calculated in [20] under the premise that the Central Limit Theorem can be applied to the Fourier transform outputs of time-domain noise samples, of which the PDF is also assumed known.

In contrast with the aforementioned publications that acquire coding gain by taking into account the exact knowledge of the PDF of impulsive noise at the receiver, there are quite a number of published papers [16], [17], [21], [22] investigating the use of a certain clipping-based decoding metric to not only reduce the performance degradation induced from strong impulsive noise but also forgo the computationally demanding estimation of the statistics of impulsive noise. In situations where code concatenation is considered, it is shown in [23] that the latency resulting from interleaving can be greatly reduced with the aid of an outer code. Notably, in [23], the decoding metric that accommodates the impulsive noise is not provided for the inner code decoder, of which the anticipating functionality is to precisely erase the corrupted bytes prior to outer code decoding. The results in [23] enhances the data rate of a Discrete Multitone Very-high-bit-rate Digital Subscriber Loop (DMT-VDSL) system. It is worth noting that in the references mentioned above, the clipping thresholds are mostly devised based on extensive computer simulations, which may be time-consuming, if not computationally prohibitive, especially for low bit-error-rate applications. Relatively recently, the joint erasure

marking and Viterbi decoding algorithm (JEVA) introduced in [19] is shown to outperform its traditional counterpart, where the erasure marking to received signals and the erasure Viterbi decoding are operated sequentially. Under certain circumstances, simulations in [19] show that the JEVA can even achieve the BEP performance of the optimal maximum-likelihood decoding (MLD).

B. Main Contributions of this paper

- Despite lacking of the statistical knowledge of impulses, a robust decoding scheme is proposed by incorporating a design parameter p into the new approximate decoding metric derived from the MAP decoding rule. Owing to the memoryless nature of the impulsive noise channels, the proposed decoding scheme is analogous to perform clipping on the basis of the decoder metric [15]–[17], [22], of which the clipping threshold is induced by the parameter p . Note that a similar metric-clipping based decoding algorithm has already been reported in [15]–[17],¹ where a fixed robust clipping threshold $\Delta = 10^{-3}$ was chosen from extensive simulations therein. Aside from the similarity on conception of clipping, the definition of the clipping threshold Δ is however different from what devised in this work.
- By invoking the Chernoff bound on the BEP, we propose an analytical method to choose the value of the clipping threshold (or equivalently, the design parameter p). Our exclusively derived method can not only be applied to the scenarios, in which performing extensive computer simulation becomes unaffordable (e.g., for low-BEP applications), but also shed light on how the design parameter should be adjusted as the system parameters vary. Under a similar system setting as in [15], [16], our simulation results indicate that a fixed robust

¹The proposed robust decoding scheme in this paper was originally inspired by the joint erasure marking and Viterbi decoding algorithm [19]. We became aware of the very relevant work reported in [15]–[17] after completing an early version of this paper. We acknowledge the anonymous reviewers, who brought our attention to [15]–[17] and suggested the necessary comparison presented in this paper.

choice of p vindicates the choice of $\Delta = 10^{-3}$ in [15]–[17] in an analytical manner, and can even result in a system performance close to that of a maximum-likelihood decoder, which however requires the full statistical knowledge of impulsive noises. Nevertheless, simulation results reveal that at higher probability of occurrence of impulses and in various levels of Impulse-Gaussian power Ratio (IGR), both of which are unknown to the receiver, the clipping threshold suggested by the derived Chernoff bound on the BEP induces a substantial performance gain over the schemes given in [15], [16].

- To delve into the feasibility of the proposed decoder, the impact of the analog-to-digital converter (ADC) on the BEP performance is also investigated via simulations. It is shown that with a properly adjusted dynamic range of the ADC, the performance loss due to quantization error is only about a small constant in terms of signal-to-noise ratio in dB, similar to the conventional impulse-free case.

The rest of the paper is organized as follows. Prevalently used impulsive noise models along with the system framework are first reviewed in Section II, followed by the details of the proposed efficient decoding algorithm in Section III. The performance evaluation, based on the Chernoff bound on the BEP for the proposed decoding scheme, is elaborately carried out in Section IV. Section V presents the simulation results, and Section VI draws conclusions afterwards.

II. SYSTEM MODEL

Let $\mathbf{C}$ represent an (n, k, m) convolutional code (CC) with information sequence of length kL bits, where n code bits are blockwisely produced in response to each block input of k information bits, and m is the maximum length of shift registers in the CC encoder. Since m blocks of k -bit zeros are appended to the end of information sequence to clear the contents of shift registers, the length of each codeword is thus $N = n(L + m)$. During its transmission, the convolutioally coded sequence is subject to memoryless impulsive noises characterized by, but not restricted to, the Bernoulli-Gaussian model [5] or the Middleton Class-A model [6], where

each noise sample n_j at any time instant j is the sum of an AWGN noise g_j and an impulsive noise i_j , i.e., $n_j = g_j + i_j$. As a convention, the AWGN noise $\{g_j\}_{j=0}^{N-1}$ has a flat single-sided power spectral density (PSD) of height N_0 .

In the Bernoulli-Gaussian model [5], the impulsive noise i_j can be further represented as a product of two independent variables, i.e., $i_j = b_j \omega_j$, where $b_j \in \{0, 1\}$ is a Bernoulli random variable with the probability of occurrence of impulses $\Pr(b_j = 1) = p_b$, and $\{\omega_j\}_{j=0}^{N-1}$ is an independent and identically distributed (*i.i.d.*) Gaussian random process with mean zero and variance $\frac{N_0}{2}\Gamma$. Here and throughout the paper, $\Pr(\cdot)$ denotes probability of the event inside parentheses. Notably, the constant Γ is actually the mean power ratio of impulsive noise ω_j against AWGN noise g_j , and thus is exactly the IGR. The PDF of n_j in the Bernoulli-Gaussian model is thus given by

$$f_{\text{BG}}(x) = (1 - p_b) \cdot \frac{1}{\sqrt{N_0/2}} \varphi\left(\frac{x}{\sqrt{N_0/2}}\right) + p_b \cdot \frac{1}{\sqrt{(N_0/2)(1 + \Gamma)}} \varphi\left(\frac{x}{\sqrt{(N_0/2)(1 + \Gamma)}}\right), \quad (1)$$

where $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\{-x^2/2\}$ is the Gaussian PDF with zero mean and unit variance.

Also popularly adopted in the literature for characterizing channel noises with impulsive characteristic is the Middleton Class-A model [6], for which the PDF of *i.i.d.* additive noise samples $\{n_j = g_j + b_j \omega_j\}_{j=0}^{N-1}$ is described as [14]:

$$f_{\text{M}}(x) = \sum_{\ell=0}^{\infty} \alpha_{\ell} \cdot \frac{1}{\sigma_{\ell}} \varphi\left(\frac{x}{\sigma_{\ell}}\right), \quad (2)$$

where $\alpha_{\ell} = e^{-A} \frac{A^{\ell}}{\ell!}$ and $\sigma_{\ell}^2 = \frac{N_0}{2} \left(1 + \frac{\ell}{\Lambda A}\right)$. Notably, in specifying α_{ℓ} and σ_{ℓ}^2 , the parameter A is a constant that is usually referred to as the *impulsive index*, and Λ can be regarded as a mean power ratio of the AWGN component against the impulsive noise component [14] since $\{\omega_j\}_{j=0}^{N-1}$ is an *i.i.d.* Gaussian random process with mean zero and variance $(\frac{N_0}{2})/(\Lambda A)$. It is interesting to note that b_j is a Poisson variable with mean A such that $\Pr(b_j = \ell) = \alpha_{\ell}$ for $\ell \geq 0$; hence, f_{M} actually exhibits the same form as f_{BG} if $\alpha_0 = 1 - p_b$, $\alpha_1 = p_b$, and $\alpha_{\ell} = 0$ for $\ell \geq 2$.

III. METRIC ERASURE VITERBI ALGORITHM (MEVA)

As mentioned previously, it may be hard to precisely obtain the PDF of impulsive noises in reality. In order to address the adverse impact on decoding in the context of unknown statistics of impulses, we attempt to preform clipping on the decoding metrics, of which the clipping threshold will be further investigated in Section IV. The evolution that leads to clipping from erasure marking [19] will be briefly presented below.

By assuming the modulation format is Binary Phase-Shift Keying (BPSK), the received symbol sequence $\mathbf{r} = (r_0, r_1, \dots, r_{N-1})$ can be represented by

$$\mathbf{r} = (-1)^{\mathbf{v}} \sqrt{E} + \mathbf{n}, \quad (3)$$

where $\mathbf{v} = (v_0, v_1, \dots, v_{N-1}) \in \{0, 1\}^N$ is the transmitted codeword and E is the energy of modulated symbol. From the perspective of the decoder, the impulsive noise sequence $\mathbf{n} = (n_0, n_1, \dots, n_{N-1})$ is only known to be *i.i.d.* with n_j being an impulse plus AWGN with probability p , and AWGN only with probability $(1 - p)$. It should be stressed that the parameter p is likely to be different from the true probability of occurrence of impulses p_b in the Bernoulli-Gaussian model (or $(1 - \alpha_0)$ in the Middleton Class-A model) as the knowledge of impulsive noise statistics is unknown to or extremely difficult to be estimated at the receiver.

Further, we conveniently define an *indicator* e_j for each received symbol to indicate whether or not this symbol is corrupted by an impulsive noise, where $e_j = 1$ means the occurrence of an impulse arrival, and $e_j = 0$ means otherwise. With the aid of the notation e_j , we denote the jointly maximum-likelihood decision by $(\hat{\mathbf{v}}, \hat{\mathbf{e}})$ over all possible pairs $(\tilde{\mathbf{v}}, \tilde{\mathbf{e}}) \in \mathbf{E}_N$, where $\mathbf{E}_N$ is the collection of all $(\tilde{\mathbf{v}}, \tilde{\mathbf{e}})$ pairs, each of which corresponds to a codeword $\tilde{\mathbf{v}}$ paired with an indicator sequence $\tilde{\mathbf{e}} = (\tilde{e}_0, \tilde{e}_1, \dots, \tilde{e}_{N-1})$. This is in contrast with finding the maximum-likelihood codeword simply over the code book $\mathbf{C} = \{\mathbf{v}_0, \dots, \mathbf{v}_{2^{kL}-1}\}$. To clarify the notations in the following, we will reserve v_j to denote the code bit that was truly transmitted, $\tilde{v}_j$ the code bit associated with the branch of the trellis currently decoded over, and $\hat{v}_j$ the decoding

decision. The same convention will be applied for the notations of indicator sequences.

Under the assumption of equal *a priori* probabilities of information sequences, the maximum *a posteriori* probability (MAP) decoding rule suggests that the optimal decision $(\hat{\mathbf{v}}, \hat{\mathbf{e}})$ should be given by

$$\begin{aligned} (\hat{\mathbf{v}}, \hat{\mathbf{e}}) &= \arg \max_{(\tilde{\mathbf{v}}, \tilde{\mathbf{e}}) \in \mathbf{E}_N} \ln f(\mathbf{r}|\tilde{\mathbf{v}}, \tilde{\mathbf{e}}) \Pr(\tilde{\mathbf{e}}) \\ &= \arg \max_{(\tilde{\mathbf{v}}, \tilde{\mathbf{e}}) \in \mathbf{E}_N} \sum_{j=0}^{N-1} \left((1 - \tilde{e}_j) [\ln f(r_j|\tilde{v}_j, \tilde{e}_j = 0) + \ln \Pr(\tilde{e}_j = 0)] \right. \\ &\quad \left. + \tilde{e}_j [\ln f(r_j|\tilde{v}_j, \tilde{e}_j = 1) + \ln \Pr(\tilde{e}_j = 1)] \right) \end{aligned} \quad (4)$$

where f is used to denote a PDF. It is obvious that $f(r_j|\tilde{v}_j, \tilde{e}_j = 0) = \frac{1}{\sqrt{N_0/2}} \varphi\left(\frac{r_j - (-1)^{\tilde{v}_j} \sqrt{E}}{\sqrt{N_0/2}}\right)$ when the received sample r_j is impulse-free. In case $\tilde{e}_j = 1$, the contribution of r_j to the quantity $\ln f(\mathbf{r}|\tilde{\mathbf{v}}, \tilde{\mathbf{e}}) \Pr(\tilde{\mathbf{e}})$ in (4) is $\ln f(r_j|\tilde{v}_j, \tilde{e}_j = 1) + \ln(p)$, which without the statistical knowledge of impulsive noise is untrustworthy and may dictate the decoding decision if it is still used as a part of the decision metric. From the perspective of a robust decoder design, a simple way to avert this problem is to neutralize this untrustworthy contribution by erasing the metric $\ln f(r_j|\tilde{v}_j, \tilde{e}_j = 1)$ in (4). For this reason, we refine (4) to

$$\begin{aligned} (\hat{\mathbf{v}}, \hat{\mathbf{e}}) &= \arg \max_{(\tilde{\mathbf{v}}, \tilde{\mathbf{e}}) \in \mathbf{E}_N} \sum_{j=0}^{N-1} \left((1 - \tilde{e}_j) \ln f(r_j|\tilde{v}_j, \tilde{e}_j = 0) + (1 - \tilde{e}_j) \ln(1 - p) + \tilde{e}_j \ln(p) \right) \\ &= \arg \min_{(\tilde{\mathbf{v}}, \tilde{\mathbf{e}}) \in \mathbf{E}_N} \sum_{j=0}^{N-1} \left[(1 - \tilde{e}_j) \left(\frac{(r_j - (-1)^{\tilde{v}_j} \sqrt{E})^2}{N_0} - \ln \frac{(1 - p)}{\sqrt{\pi N_0}} \right) - \tilde{e}_j \ln(p) \right]. \end{aligned} \quad (5)$$

From (5), it can be seen that when erasure of $\ln f(r_j|\tilde{v}_j, \tilde{e}_j = 1)$ is adopted, keeping the term $\ln(p)$ is indispensable; otherwise, the decoder will always opt for choosing a trivial decision, i.e., $\hat{\mathbf{e}} = \mathbf{1}$, if, e.g., $N_0 \geq \frac{1}{\pi}$.

We then propose to use the right-hand side of (5) as the path metric of $(\tilde{\mathbf{v}}, \tilde{\mathbf{e}})$, denoted by $M(\mathbf{r}|\tilde{\mathbf{v}}, \tilde{\mathbf{e}})$. The bit metric of $(\tilde{v}_j, \tilde{e}_j)$ is accordingly defined as

$$M(r_j|\tilde{v}_j, \tilde{e}_j) = (1 - \tilde{e}_j) \left(\frac{(r_j - (-1)^{\tilde{v}_j} \sqrt{E})^2}{N_0} - \ln \frac{(1 - p)}{\sqrt{\pi N_0}} \right) - \tilde{e}_j \ln(p). \quad (6)$$

Note that the decoding scheme that uses (6) as the decoding metric was first named the *metric erasure Viterbi algorithm* (MEVA) in [24]. Depending on whether the received sample r_j is corrupted by an impulse or not, there are two probable bit metrics resulted from (6), between which the smaller one is retained for updating the path metric. Specifically, at time instant j , if the following inequality holds

$$\underbrace{\frac{(r_j - (-1)^{\tilde{v}_j} \sqrt{E})^2}{N_0} - \ln \frac{(1-p)}{\sqrt{\pi N_0}}}_{M(r_j|\tilde{v}_j, \tilde{e}_j=0)} > \underbrace{-\ln(p)}_{M(r_j|\tilde{v}_j, \tilde{e}_j=1)}, \quad (7)$$

the MEVA will deem the associated sample as impulse-corrupted. In this regard, the decoding behavior of the MEVA is equivalent to perform clipping on the Euclidean metric $(r_j - (-1)^{\tilde{v}_j} \sqrt{E})^2$, where the clipping threshold is measured by

$$T = N_0 \ln \left(\frac{1-p}{p\sqrt{\pi N_0}} \right), \quad (8)$$

and (7) can be equivalently written as

$$(r_j - (-1)^{\tilde{v}_j} \sqrt{E})^2 > T. \quad (9)$$

In light of (9), we can transform the path metric to:

$$\begin{aligned} \sum_{j=0}^{N-1} M(r_j|\tilde{v}_j, \tilde{e}_j) &= \sum_{j=0}^{N-1} \min \{M(r_j|\tilde{v}_j, \tilde{e}_j=0), M(r_j|\tilde{v}_j, \tilde{e}_j=1)\} \\ &= \frac{1}{N_0} \sum_{j=0}^{N-1} \min \left\{ (r_j - (-1)^{\tilde{v}_j} \sqrt{E})^2, T \right\} - N \ln \frac{(1-p)}{\sqrt{\pi N_0}}. \end{aligned} \quad (10)$$

By removing the last inconsequential term for decision-making from (10) and multiplying the resultant quantity by N_0 , the MEVA is equivalent to perform the VA on the original (i.e., non-erasure-expanded) trellis of the CC with the clipped Euclidean bit metric

$$\tilde{M}(r_j|\tilde{v}_j) \triangleq \min \left\{ (r_j - (-1)^{\tilde{v}_j} \sqrt{E})^2, T \right\}. \quad (11)$$

By abusing the notation, we further denote the corresponding path metric by

$$\tilde{M}(\mathbf{r}|\tilde{\mathbf{v}}) = \sum_{j=0}^{N-1} \tilde{M}(r_j|\tilde{v}_j). \quad (12)$$

It should be mentioned that when p is larger than $\frac{1}{1+\sqrt{\pi N_0}}$, we will have a negative clipping threshold $T < 0$. In this case, all code bits are erased and the MEVA can only output a random guess on the transmitted codeword with the probability of correct decoding equal to 2^{-kL} .

As opposed to applying the VA on the erasure-expanded trellis in [24], which elicits an increase of decoding complexity by 2^n times relative to the VA on the original trellis, the clipping mechanism (12) applied for implementing the MEVA only induces an additional clipping operation (11) for each branch metric computation and hence increases the decoding complexity only by a constant multiplicative factor (< 2).

We end this section by emphasizing again that in the absence of the knowledge of impulsive noise statistics, the parameter p assumed at the decoder is very likely to be different from the true probability of occurrence of impulses p_b . In particular, after erasing the untrustworthy metric $\ln f(r_j|\tilde{v}_j, \tilde{e}_j = 1)$ in (4), this parameter becomes an offset provider $-\tilde{e}_j \ln(p)$ in the bit metric in (6). Hence, it is somehow transformed into a *design parameter* rather than an estimate on the probability of occurrence of impulses p_b . How to select a suitable p for the MEVA will be presented in Subsection V-A. Moreover, in spite of the binary signaling assumption, the above derivation can be directly applied to single-carrier narrowband communication systems using Quadrature Phase-Shift Keying (QPSK) with Gray mapping and real and imaginary clipping respectively for in-phase and quadrature branches.

IV. PERFORMANCE BOUNDS FOR GIVEN STATISTICS OF IMPULSIVE NOISES

After introducing the concept of clipping inherently realized in the MEVA in Section III, we hereby derive the PEP bound for the MEVA by resorting to the technique of the Chernoff bound for the two statistical models of impulsive noises in Section II.

For ease of understanding, only the analysis for the Bernoulli-Gaussian noise model is given. An analogous bound for the Middleton Class-A model is simply provided without detailed derivation. These bounds are then used to select a suitable value of p when it is treated as a design

parameter. The way to select p according to these bounds will be presented in Subsection V-A.

In this analysis, we only focus on (n, k, m) CCs with $k = 1$. It can be similarly extended to CCs with $k > 1$. With length- L information sequences, an $(n, 1, m)$ CC can be regarded as an (N, L) linear block code of length $N = n(L + m)$. Without loss of generality, due to the linearity of the code and the memoryless property of the additive noise, the all-zero codeword $\mathbf{v}_0$ can be assumed to be transmitted when calculating the decoding error.

By the system setting in the previous paragraph, a well-known BEP bound P_B is given by

$$P_B \leq \frac{1}{k} \sum_{w=d_{\min}}^{d_{\max}} B_w \cdot \Pr \left(\tilde{M}(\mathbf{r}|\mathbf{v}^{(w)}) \leq \tilde{M}(\mathbf{r}|\mathbf{v}_0) \middle| \mathcal{A}_0 \right), \quad (13)$$

where $d_{\min}$ and $d_{\max}$ are respectively the minimum and maximum pairwise Hamming distances of the (N, L) block code, $\mathbf{v}^{(w)}$ is any codeword with Hamming weight w , B_w is the total number of nonzero information bits in all Hamming-weight- w codewords, and $\mathcal{A}_0$ is the event that the all-zero codeword is chosen for transmission. It remains to derive $\Pr \left(\tilde{M}(\mathbf{r}|\hat{\mathbf{v}}) \leq \tilde{M}(\mathbf{r}|\mathbf{v}_0) \middle| \mathcal{A}_0 \right)$ as a function of the Hamming weight of $\hat{\mathbf{v}}$.

Continue the derivation of the PEP in (13):

$$\begin{aligned} \Pr \left(\tilde{M}(\mathbf{r}|\hat{\mathbf{v}}) \leq \tilde{M}(\mathbf{r}|\mathbf{v}_0) \middle| \mathcal{A}_0 \right) &= \Pr \left(\sum_{j: \hat{v}_j=1} \phi_j \leq 0 \middle| \mathcal{A}_0 \right) \\ &\leq \min_{s: s>0} E \left[e^{-s \sum_{j: \hat{v}_j=1} \phi_j} \middle| \mathcal{A}_0 \right], \end{aligned} \quad (14)$$

where for those j 's with $\hat{v}_j = 1$, we can readily have

$$\begin{aligned} \phi_j &= \min \left((r_j + \sqrt{E})^2, T \right) - \min \left((r_j - \sqrt{E})^2, T \right) \\ &= \begin{cases} 4\sqrt{E}r_j, & (r_j + \sqrt{E})^2 < T \wedge (r_j - \sqrt{E})^2 < T; \\ T - (r_j - \sqrt{E})^2, & (r_j + \sqrt{E})^2 \geq T \wedge (r_j - \sqrt{E})^2 < T; \\ (r_j + \sqrt{E})^2 - T, & (r_j + \sqrt{E})^2 < T \wedge (r_j - \sqrt{E})^2 \geq T; \\ 0, & (r_j + \sqrt{E})^2 \geq T \wedge (r_j - \sqrt{E})^2 \geq T. \end{cases} \end{aligned}$$

Depending on the gap between the clipping threshold T and the symbol energy E , we then distinguish between two cases: $E \geq T$ and $E < T$, where we assume $T > 0$ and $t = \sqrt{T}$.

If $E \geq T$,

$$\phi_j = \begin{cases} T - (r_j - \sqrt{E})^2, & \sqrt{E} - t < r_j < \sqrt{E} + t \\ (r_j + \sqrt{E})^2 - T, & -\sqrt{E} - t < r_j < -\sqrt{E} + t \\ 0, & \text{otherwise.} \end{cases}$$

Else if $E < T$, then

$$\phi_j = \begin{cases} 4\sqrt{E}r_j, & \sqrt{E} - t < r_j < -\sqrt{E} + t \\ T - (r_j - \sqrt{E})^2, & -\sqrt{E} + t \leq r_j < \sqrt{E} + t \\ (r_j + \sqrt{E})^2 - T, & -\sqrt{E} - t \leq r_j \leq \sqrt{E} - t \\ 0, & \text{otherwise.} \end{cases}$$

Note that $r_j = \sqrt{E} + n_j$ under $\mathcal{A}_0$, and hence the above equations can be equivalently written as: If $E \geq T$,

$$\phi_j = \begin{cases} T - n_j^2, & -t < n_j < t \\ (2\sqrt{E} + n_j)^2 - T, & -t - 2\sqrt{E} < n_j < t - 2\sqrt{E} \\ 0, & \text{otherwise,} \end{cases}$$

else

$$\phi_j = \begin{cases} 4\sqrt{E}(\sqrt{E} + n_j), & -t < n_j < t - 2\sqrt{E} \\ T - n_j^2, & t - 2\sqrt{E} \leq n_j < t \\ (2\sqrt{E} + n_j)^2 - T, & -t - 2\sqrt{E} \leq n_j \leq -t \\ 0, & \text{otherwise,} \end{cases}$$

where by following (1), the PDF of n_j is equal to

$$f_{n_j}(x) = (1 - p_b) \frac{1}{\sqrt{N_0/2}} \varphi\left(\frac{x}{\sqrt{N_0/2}}\right) + p_b \frac{1}{\sqrt{N_0/2(1 + \Gamma)}} \varphi\left(\frac{x}{\sqrt{(N_0/2)(1 + \Gamma)}}\right).$$

As a result, when $E \geq T$, and by letting $\sigma_0 = \sqrt{N_0/2}$, $\sigma_1 = \sqrt{(N_0/2)(1+\Gamma)}$ and $\mu = \sqrt{E}$, we obtain

$$\begin{aligned}
& E [e^{-s\phi_j} | \mathcal{A}_0] \\
&= \int_{-t}^t e^{-s(T-x^2)} f_{n_j}(x) dx + \int_{-t-2\mu}^{t-2\mu} e^{-s((2\mu+x)^2-T)} f_{n_j}(x) dx \\
&\quad + \left(1 - \int_{-t}^t f_{n_j}(x) dx - \int_{-t-2\mu}^{t-2\mu} f_{n_j}(x) dx \right) \\
&= e^{-sT} [(1-p_b)A_1(s, t, \sigma_0) + p_b A_1(s, t, \sigma_1)] + e^{sT} [(1-p_b)B_1(\mu, s, t, \sigma_0) + p_b B_1(\mu, s, t, \sigma_1)] \\
&\quad + [(1-p_b)C_1(t, \sigma_0, \mu) + p_b C_1(t, \sigma_1, \mu)] , \tag{15}
\end{aligned}$$

where $Q(t) = \int_t^\infty \varphi(x) dx$ is the tail probability of the standard normal distribution, and the detailed derivations as well as definitions of functions A_1 , B_1 and C_1 are placed in Appendix for better readability. Else if $T > E$, we obtain

$$\begin{aligned}
& E [e^{-s\phi_j} | \mathcal{A}_0] \\
&= \int_{-t}^{t-2\mu} e^{-4s\mu(\mu+x)} f_{n_j}(x) dx + \int_{t-2\mu}^t e^{-s(T-x^2)} f_{n_j}(x) dx + \int_{-t-2\mu}^{-t} e^{-s((2\mu+x)^2-T)} f_{n_j}(x) dx \\
&\quad + \left(1 - \int_{-t-2\mu}^t f_{n_j}(x) dx \right) \\
&= e^{-4s\mu^2} [(1-p_b)D(s, t, \sigma_0, \mu) + p_b D(s, t, \sigma_1, \mu)] + e^{-sT} [(1-p_b)A_2(s, t, \sigma_0, \mu) + p_b A_2(s, t, \sigma_1, \mu)] \\
&\quad + e^{sT} [(1-p_b)B_2(s, t, \sigma_0, \mu) + p_b B_2(s, t, \sigma_1, \mu)] + [(1-p_b)C_2(t, \sigma_0, \mu) + p_b C_2(t, \sigma_1, \mu)], \tag{16}
\end{aligned}$$

where the detailed derivations as well as definitions of functions A_2 , B_2 , C_2 and D are also placed in Appendix.

We can similarly extend the above derivation to the Middleton Class-A model, where (15) and (16) respectively become

$$E [e^{-s\phi_j} | \mathcal{A}_0] = e^{-sT} \sum_{\ell=0}^{\infty} \alpha_\ell A_1(s, t, \sigma_\ell) + e^{sT} \sum_{\ell=0}^{\infty} \alpha_\ell B_1(\mu, s, t, \sigma_\ell) + \sum_{\ell=0}^{\infty} \alpha_\ell C_1(t, \sigma_\ell, \mu)$$

Fig. 1. Bit error probability bounds for the MEVA over Bernoulli-Gaussian channels.

and

$$\begin{aligned}
 E \left[e^{-s\phi_j} \middle| \mathcal{A}_0 \right] &= e^{-4s\mu^2} \sum_{\ell=0}^{\infty} \alpha_{\ell} D(s, t, \sigma_{\ell}, \mu) + e^{-sT} \sum_{\ell=0}^{\infty} \alpha_{\ell} A_2(s, t, \sigma_{\ell}, \mu) \\
 &\quad + e^{sT} \sum_{\ell=0}^{\infty} \alpha_{\ell} B_2(s, t, \sigma_{\ell}, \mu) + \sum_{\ell=0}^{\infty} \alpha_{\ell} C_2(t, \sigma_{\ell}, \mu).
 \end{aligned}$$

With the availability of the expression of $E \left[e^{-s\phi_j} \middle| \mathcal{A}_0 \right]$, the desired PEP bound follows after substituting it into (14).

V. SIMULATION RESULTS

We hereby justify the proposed MEVA by conducting computer simulations in the configuration of transmitting the $(2, 1, 6)$ CC with generators 147,135 (octal) over the Bernoulli-Gaussian and Middleton Class-A noise channel models. The number of information bits per data frame for the CC is fixed at $L = 500$. To simplify the exposition, the BPSK modulation is exclusively used in the following simulations.

Fig. 2. Bit error probability bounds versus the design parameter p for the MEVA over Bernoulli-Gaussian channels with $\Gamma = 100$ at $E_b/N_0 = 5$ dB. Here the diamond markers correspond to the choice of clipping threshold suggested in [17].

A. The Chernoff bounds on the BEP of the MEVA

To attest our derived Chernoff bound on the BEP for the MEVA, we first measure it with respect to a wide range of probability of occurrence of impulses p_b , and plot the results for different IGRs (i.e., parameter Γ) in Fig. 1. The design parameter p is set to be equal to p_b in this figure even though the value of its induced BEP bound is unfavorable, where a favored choice of p will be illuminated later in Fig. 2. It can be observed from Fig. 1 that the curves for BEP bounds are fairly close to each other for each illustrated probability p_b , indicating the robustness of the MEVA against impulses, of which the power is unknown to the receiver. Besides, it is noteworthy that the value of the BEP performance bound is worse as the probability p_b is larger, reflecting indirectly that the effectiveness of the MEVA in couple with the CC is likely to be weakened in the context of significantly frequent occurrence of impulses.

Subsequently, to measure the impact of the choice of p on the performance of the MEVA,

we resort to the derived Chernoff bound to expedite the assessment prior to laborious computer simulations in the sequel. Here the values of p_b relevant to the target application scenario are assumed to be as large as up to 0.2. In particular, we measure the Chernoff bound on the BEP of the MEVA at $E_b/N_0 = 5$ dB under the system setting that the design parameter p , which is used to determine the clipping threshold, is very likely to be deviated from the unknown p_b . The entailing result is shown in Fig. 2. Here the IGR is fixed at $\Gamma = 100$. It is remarkable to see that regardless of the probability of occurrence of impulses p_b , the incurred BEP bounds are all of U-shape, prompting us to avoid certain range of p such as the region of $p \geq 10^{-2}$. On the other hand, it is possible to set a fixed clipping threshold T (see (8)) derived from a properly chosen parameter p that is independent of the probability p_b . This allows the practical use of the proposed decoding scheme without resorting to a sophisticated estimation algorithm for p_b , while the resultant coding gain is likely to be least compromised. In this regard, parameter $p = 10^{-3}$ is selected for the following simulations, unless otherwise specified.

As mentioned earlier, the clipping threshold in [17] was empirically sought out, which, after being converted to conform with our clipping threshold formula in (8), is also marked in Fig. 2 (see the diamond markers in each curve). One can clearly see that the value associated with diamonds is fairly close to our chosen $p = 0.001$ in this context. In this perspective, we can expect our simulation results shown below to be on par with [17].

B. The impact of the design parameter p on the simulated BEPs

To further reflect the effectiveness of the derived analytical bound in interpreting the behavior of the performance of the MEVA, we hereby investigate the BEPs at $E_b/N_0 = 5$ dB for a much wider ranges of p and p_b over the Bernoulli-Gaussian model as well as A over the Middleton Class-A model. As one can observe from Fig. 3, the BEP increases in the regime of exaggerated p (such as $p \geq p_b$). A similar conclusion has been drawn in [16]. As the value of p is reduced gradually from $p = p_b$, the BEP improves at first but would increase again when an excessively

Fig. 3. Bit error probabilities versus the design parameter p with respect to various p_b and A at $E_b/N_0 = 5$ dB.

small p is reached. Such a U -shape trend has already been attested in our analytical results in Fig. 2. Particularly for lower p_b like 0.02, there exists a more significant performance change within the presented range of design parameter p , reassuring that it is useful to derive the Chernoff bound for the BEP of a clipping-featured decoder to arrive at a suitable clipping threshold to avert underperformance regardless of the impulsive noise model encountered.

C. Comparison between the MEVA and the works of [16], [17], [19]

To compare the MEVA with the existing works such as [16], [17], [19], we conduct computer simulations over the Bernoulli-Gaussian noise model and the Middleton Class-A noise model, and summarize the results in Figs. 4(b) and 4(a), respectively. The BEP curve corresponding to the ideal situation that the receiver has perfect statistical knowledge on the noise model is served as a benchmark performance in these subfigures, and is labeled as MLD. The curve with label VA shows the performance of the Viterbi decoding using the conventional (non-clipped)

(a) Middleton Class-A channel with $A = 0.02$ and $\Lambda = 1/2$ (b) Bernoulli-Gaussian channel with $p_b = 0.02$ and $\Gamma = 15$

(c) Bernoulli-Gaussian channel with $\Gamma = 100$

(d) Bernoulli-Gaussian channel with $\Gamma = 400$

Fig. 4. Comparisons of bit error probabilities for various schemes.

Euclidean bit metric $(r_j - (-1)^{\tilde{v}_j} \sqrt{E})^2$, thereby corresponding to the situation that the impulsive effect is completely neglected at the receiver.

It is noteworthy that in Fig. 4(a), the BEP curve derived from the MEVA with the clipping threshold dictated by a fixed value of $p = 0.001$ (see the dash-dot line marked with “ $\square$ ”) is almost overlapped with the curve induced by the threshold in [16], [17] (see the dash-dot line

marked with “+”) as well as that of the MLD (see the solid line marked with “ ∇ ”), indicating that both clipping methods are robust against impulses and are as effective as the benchmark MLD. Moreover, for the MEVA, there is a 0.5 dB gain from setting $p = 0.001$ to setting p equal to A (see the dash-dot line marked with “ $\diamond$ ”). Noteworthy, the Chernoff bound for the BEP of the MEVA (see the solid line marked with “+”) exhibits similar trend to the simulated BEPs, justifying the efficacy of the derived bound.

Similar observations to those drawn from Fig. 4(a) can be made from Fig. 4(b), in which we additionally compare the MEVA with the iterative JEVA in [19]. We then note from Fig. 4(b) that with $p_b = 0.02$ and $\Gamma = 15$, the JEVA has comparable performance to the MEVA and the scheme in [16], [17] over the Bernoulli-Gaussian noise channel, whereas a multi-fold complexity increase due to iterations is a major hurdle that prevents the JEVA from applying to practical systems.

When the probability of occurrence of impulses p_b reaches a very high value such as 0.1 or 0.2 as attested in Figs. 4(c) and 4(d), the MEVA and the MLD perform fairly the same at low E_b/N_0 . However, the gap between the BEP curve induced by the MEVA and that of the MLD becomes visible at high E_b/N_0 . It can also be noted from this two subfigures that this gap is further enlarged when p_b increases from 0.1 to 0.2, especially for smaller IGR (i.e., Γ). The main reason behind is that when IGR is small in the context of a large p_b , a much higher E_b/N_0 value is inevitable for the BEP line to decline; yet at this occasion, the model mismatch between the assumed $f(r_j|v_j, e_j = 1) \approx 1$ (cf., the paragraph immediately after Eq. (5)) and the conditional PDF of the impulsive noise $f_{\text{BG}}(r_j - (-1)^{v_j}\sqrt{E}|b_j = 1)$ starts to exacerbate. Even so, the MEVA based on the induced clipping threshold can still outperform the scheme proposed in [16], [17] in the region of high E_b/N_0 values.

D. Examination of the impact of the ADC at the front end

We hereby investigate the saturation effect of impulsive noises on the MEVA preceded by an ADC over the Bernoulli-Gaussian noise model.

In this experiment, we assume that the ADC preceding the MEVA decoder uses 8-level quantization with uniform quantization spacing [25]. In the case of employing an impulse-unsaturated (IU) ADC, of which the dynamic range is determined by the magnitude of signal plus the impulsive background noise, we refer to it as “IU-ADC + MEVA.” The case corresponding to the usual 8-level uniform quantization [25] is referred to as “IS-ADC + MEVA,” where the dynamic range of the impulse-saturated (IS) ADC is the signal level plus the background noise level but not including the impulsive noise level. The results are illustrated in Fig. 5, based on which four observations are made in the following.

First, we point out that without a preceding ADC, the dash-dot lines marked with squares respectively in Figs. 5(a) and (b) are nearly overlapped, reinforcing the immune property of the MEVA against strong impulses, as $\text{IGR} = 100$ and 400 respectively in Figs. 5(a) and (b). Secondly, we observe that regardless of the value of the IGR, adding an IS-ADC to the MEVA front end induces roughly 1 dB performance loss at $\text{BEP} = 10^{-5}$ due to the unavoidable quantization noise; nevertheless, the “IS-ADC + MEVA” decoder can still approach the benchmark performance of the “IS-ADC + MLD” decoder. Thirdly, because of the excess quantization noise that scales with the IGR, a much more pronounced performance loss will occur when an IU-ADC is used. As an example, the “IU-ADC + MEVA” BEP curves corresponding to $\text{IGR} = 400$ (see the solid line marked with diamonds) never descends down to 10^{-3} until $E_b/N_0 = 6$ dB; yet, the same impairing factor also leads to a serious performance degradation for the MLD (see the curves corresponding to “IU-ADC + MLD”). Fourthly, we note that the coding gain of the “IS-ADC + MEVA” decoder over the hard-decision Viterbi algorithm, termed as “HD-VA” in Fig. 5, is between 1 and 1.5 dB, confirming again the superiority of the MEVA

Fig. 5. Bit error probabilities of the MEVA decoders with a preceding impulse-saturated (IS) and impulse-unsaturated (IU) ADC with $p = 0.001$ over the Bernoulli-Gaussian channel with $p_b = 0.02$ and two different IGRs.

over the conventional VA even with a preceding ADC.

VI. CONCLUSION

A robust decoding scheme is indispensable for a convolutionally coded data stream suffering impulsive noises because the realization of the maximum-likelihood decoding performance relies heavily on an accurate estimate of the statistics of impulsive noises, of which the time-varying nature takes a heavy toll on the receiver's complexity. By incorporating a design parameter p that emulates the unknown probability of occurrence of impulses into the joint erasure marking and Viterbi decoding algorithm, we devised the *metric erasure Viterbi algorithm* (MEVA) that is applicable without detailed statistical knowledge of impulsive noises. With the memoryless attribute of impulsive noises (e.g., resulted from ideal interleaving), the MEVA can be equivalently implemented using the conventional VA with a well-designed clipping operation on the Euclidean metric, where the clipping threshold is set according to the chosen p .

To seek out a feasible clipping threshold, we resort to the Chernoff bound for the bit er-

ror probability (BEP) of the MEVA over the Bernoulli-Gaussian and the Middleton Class-A impulsive noise models. Remarkably, by examining the analytical error bounds for a practical range of noise model parameters, choosing a fixed value of p in a robust manner is likely to be leveraged for a satisfactory BEP performance within the specified range. Particularly, in the context where the mean power ratio between impulses and background noises is large and the probability of occurrence of impulses is moderate, simulation results show that the MEVA with a fixed robust choice of the clipping threshold is not only commensurate with the traditional VA in terms of the order of complexity but also on par with the ideal MLD in BEP performance. As a consequence of its efficiency in decoding and robustness in performance, the MEVA provides an attractive solution for practical applications over impulsive noise channels such as power line communications. It is worthwhile to note that the BEP simulation results corroborate the efficacy of the fixed robust choice of design parameter p induced from the Chernoff bound. This justifies the use of the Chernoff bound on the BEP as an expeditious tool for the analytical study of the performance of the MEVA.

Finally, we suggest based on our extensive simulations not shown in the paper that a more powerful channel code than the convolutional codes may need to be invoked if the probability of occurrence of impulses is fairly high. For instance, the LDPC code [17] has been proposed when this probability is in the order of 10^{-1} . Therefore, to adapt our proposed MEVA approach to a more sophisticated channel code and verify its effectiveness in such a harsh scenario could be a future topic of practical interest.

REFERENCES

- [1] J. Haring and A. J. H. Vinck, "Coding for impulsive noise channels," in *IEEE Proc. ISPLC*, April 2001, pp. 103–108.
- [2] J. Modestino, D. H. Sargard, and R. E. Bollen, "Use of coding to combat impulse noise on digital subscriber loops," *IEEE Trans. Commun.*, pp. 529–537, May 1988.
- [3] K. L. Blackard and T. S. Rappaport, "Measurements and models of radio frequency impulsive noise for indoor wireless communications," *IEEE J. Sel. Areas Commun.*, pp. 991–1001, Sept. 1993.
- [4] D. Middleton, "Statistical-physical models of electromagnetic interference," *IEEE Trans. Electromagn. Compat.*, pp. 106–127, Aug. 1977.
- [5] M. Ghosh, "Analysis of the effect of impulse noise on multicarrier and single carrier QAM systems," *IEEE Trans. Commun.*, pp. 145–147, Feb 1996.

- [6] D. Middleton, "Canonical and quasi-canonical probability models of class A interference," *IEEE Trans. Electromagn. Compat.*, pp. 76–106, May 1983.
- [7] A. D. Spaulding and D. Middleton, "Optimum reception in the impulsive interference environment-part I: coherent detection," *IEEE Trans. Commun.*, pp. 910–923, Sept. 1977.
- [8] M. K. S. Miyamoto and N. Morinaga, "Performance analysis of QAM systems under class A impulse noise environment," *IEEE Trans. Electromagn. Compat.*, pp. 260–267, May 1995.
- [9] J. Haring and A. J. H. Vinck, "Performance bounds for optimum and suboptimum reception under Class-A impulsive noise," *IEEE Trans. Commun.*, pp. 1130–1136, July 2002.
- [10] R. Pighi, M. Franceschini, G. Ferrari, and R. Raheli, "Fundamental performance limits of communications systems impaired by impulse noise," *IEEE Trans. Commun.*, pp. 171–182, Jan 2009.
- [11] D. H. Sargrad and J. W. Modestino, "Error-and-erasures coding to combat impulse noise on digital subscriber loops," *IEEE Trans. Commun.*, pp. 1145–1155, Aug. 1990.
- [12] D. Umehara, H. Yamaguchi, and Y. Morihoro, "Turbo decoding in impulsive noise environment," in *IEEE Global Telecom. Conf.*, Nov 2004, pp. 194–198.
- [13] H. Nakagawa, D. Umehara, S. Denno, and Y. Morihoro, "A decoding for low density parity check codes over impulsive noise channels," in *Proc. IEEE ISPLC*, May 2005, pp. 85–89.
- [14] H. C. Ferreira, L. Lampe, J. Newbury, and T. G. Swart, *Power Line Communications: Theory and Applications for Narrowband and Broadband Communications over Power Lines*. John Wiley and Sons, 2010.
- [15] D. Fertoni and G. Colavolpe, "Theoretical limits and practical detection schemes for channels affected by Class-A impulse noise," in *Proc. IEEE Global Telecom. Conf.*, Nov. 2007, pp. 146–150.
- [16] —, "A simplified metric for soft-output detection in the presence of impulse noise," in *Proc. IEEE ISPLC*, Mar. 2007, pp. 121–126.
- [17] —, "A robust metric for soft-output detection in the presence of class-A noise," *IEEE Trans. Commun.*, pp. 36–40, Jan 2009.
- [18] M. Ardakani, F. R. Kschischang, and W. Yu, "Near-capacity coding for discrete multitone systems with impulse noise," *EURASIP Journal on Applied Signal Processing*, pp. 1–10, Vol. 2006, Article ID 98738.
- [19] T. Li, W. H. Mow, and M. Siu, "Joint erasure marking and Viterbi decoding algorithm for unknown impulsive noise channels," *IEEE Trans. Wireless Commun.*, pp. 3407–3416, Sept. 2008.
- [20] P. Amirshahi, S. M. Navidpour, and M. Kavehrad, "Performance analysis of uncoded and coded OFDM broadband transmission over low voltage power-line channels with impulsive noise," *IEEE Trans. Power Delivery*, pp. 1927–1934, Oct 2006.
- [21] T. C. Chuah, "Robust iterative decoding of turbo codes in heavy-tailed noise," *IEE-Proc.*, pp. 20–38, Feb. 2005.
- [22] J. Mitra and L. Lampe, "Robust decoding for channels with impulse noise," in *Proc. IEEE Global Telecom. Conf.*, Nov./Dec. 2006, pp. 1–6.
- [23] D. Toumpakaris, W. Yu, J. Cioffi, D. Gardan, and M. Ouzzif, "A byte-erasure method for improved impulse immunity in DSL systems using soft information from an inner code," in *IEEE Int. Conf. on Commun.*, May 2003, pp. 2431–2435.
- [24] D.-F. Tseng, Y. S. Han, W. H. Mow, and J. Deng, "Efficient decoding over power-line channels," in *Proc. The Fifth Int'l Workshop on Signal Design and its Applications in Communications (IWSDA'11)*, October 2011, pp. 138–141.
- [25] J. A. Heller and I. M. Jacobs, "Viterbi decoding for satellite and space communication," *IEEE Trans. Commun. Technol.*, pp. 835–848, Oct. 1971.

APPENDIX

A. Detailed derivation of (15)

For $s > 0$, $E [e^{-s\phi_j} | \mathcal{A}_0] =$

$$\begin{aligned}
& \int_{-t}^t e^{-s(T-x^2)} f_{n_j}(x) dx + \int_{-t}^t e^{-s(x^2-T)} f_{n_j}(x-2\mu) dx + \left(1 - \int_{-t}^t f_{n_j}(x) dx - \int_{-t}^t f_{n_j}(x-2\mu) dx\right) \\
= & e^{-sT} \int_{-t}^t e^{sx^2} \left[(1-p_b) \frac{1}{\sigma_0} \varphi\left(\frac{x}{\sigma_0}\right) + p_b \frac{1}{\sigma_1} \varphi\left(\frac{x}{\sigma_1}\right) \right] dx \\
& + e^{sT} \int_{-t}^t e^{-sx^2} \left[(1-p_b) \frac{1}{\sigma_0} \varphi\left(\frac{x-2\mu}{\sigma_0}\right) + p_b \frac{1}{\sigma_1} \varphi\left(\frac{x-2\mu}{\sigma_1}\right) \right] dx \\
& + 1 - \int_{-t}^t \left[(1-p_b) \frac{1}{\sigma_0} \varphi\left(\frac{x}{\sigma_0}\right) + p_b \frac{1}{\sigma_1} \varphi\left(\frac{x}{\sigma_1}\right) + (1-p_b) \frac{1}{\sigma_0} \varphi\left(\frac{x-2\mu}{\sigma_0}\right) + p_b \frac{1}{\sigma_1} \varphi\left(\frac{x-2\mu}{\sigma_1}\right) \right] dx \\
= & e^{-sT} [(1-p_b)A_1(s, t, \sigma_0) + p_b A_1(s, t, \sigma_1)] + e^{sT} [(1-p_b)B_1(\mu, s, t, \sigma_0) + p_b B_1(\mu, s, t, \sigma_1)] \\
& + [(1-p_b)C_1(\mu, t, \sigma_0) + p_b C_1(\mu, t, \sigma_1)] ,
\end{aligned}$$

where by letting

$$a = \begin{cases} \frac{1}{\sqrt{1-2s\sigma^2}}, & 2s\sigma^2 < 1 \\ \infty, & 2s\sigma^2 = 1 \\ \frac{1}{\sqrt{2s\sigma^2-1}}, & 2s\sigma^2 > 1 , \end{cases}$$

$b = \frac{1}{\sqrt{1+2s\sigma^2}}$, $\tilde{\varphi}(x) = \frac{1}{\sqrt{2\pi}}e^{\frac{1}{2}x^2}$, and $\tilde{Q}(y) = \int_0^y \tilde{\varphi}(x)dx$, we derive that

$$\begin{aligned}
 A_1(s, t, \sigma) &= \int_{-t}^t e^{sx^2} \frac{1}{\sigma} \varphi\left(\frac{x}{\sigma}\right) dx = \int_{-t/\sigma}^{t/\sigma} e^{s\sigma^2 y^2} \varphi(y) dy = \int_{-t/\sigma}^{t/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(1-2s\sigma^2)y^2} dy \\
 &= \begin{cases} \int_{-t/(\sigma a)}^{t/(\sigma a)} a \cdot \varphi(x) dx, & 2s\sigma^2 < 1 \\ \frac{2t}{\sqrt{2\pi}\sigma}, & 2s\sigma^2 = 1 \\ \int_{-t/(\sigma a)}^{t/(\sigma a)} a \cdot \tilde{\varphi}(x) dx, & 2s\sigma^2 > 1 \end{cases} = \begin{cases} a \left(1 - 2Q\left(\frac{t}{\sigma a}\right)\right), & 2s\sigma^2 < 1 \\ \frac{2t}{\sqrt{2\pi}\sigma}, & 2s\sigma^2 = 1 \\ 2a\tilde{Q}\left(\frac{t}{\sigma a}\right), & 2s\sigma^2 > 1 \end{cases} \\
 &= \begin{cases} \frac{1}{\sqrt{1-2s\sigma^2}} \left(1 - 2Q\left(\frac{\sqrt{1-2s\sigma^2}t}{\sigma}\right)\right), & 2s\sigma^2 < 1 \\ \frac{2t}{\sqrt{2\pi}\sigma}, & 2s\sigma^2 = 1 \\ \frac{2}{\sqrt{2s\sigma^2-1}} \tilde{Q}\left(\frac{\sqrt{2s\sigma^2-1}t}{\sigma}\right), & 2s\sigma^2 > 1 \end{cases}
 \end{aligned}$$

and

$$\begin{aligned}
 B_1(\mu, s, t, \sigma) &= \int_{-t}^t e^{-sx^2} \frac{1}{\sigma} \varphi\left(\frac{x-2\mu}{\sigma}\right) dx \\
 &= e^{-2\mu^2(1-b^2)/\sigma^2} \int_{-t}^t \frac{1}{\sigma} \varphi\left(\frac{x-2\mu b^2}{\sigma b}\right) dx \\
 &= e^{-2\mu^2(1-b^2)/\sigma^2} \int_{(-t-2\mu b^2)/(\sigma b)}^{(t-2\mu b^2)/(\sigma b)} b \varphi(y) dy \\
 &= b e^{-2\mu^2(1-b^2)/\sigma^2} \left(Q\left(\frac{2\mu b^2 - t}{\sigma b}\right) - Q\left(\frac{2\mu b^2 + t}{\sigma b}\right) \right) \\
 &= \frac{1}{\sqrt{1+2s\sigma^2}} e^{-\frac{4s\mu^2}{1+2s\sigma^2}} \left(Q\left(\frac{2\mu - t(1+2s\sigma^2)}{\sigma\sqrt{1+2s\sigma^2}}\right) - Q\left(\frac{2\mu + t(1+2s\sigma^2)}{\sigma\sqrt{1+2s\sigma^2}}\right) \right),
 \end{aligned}$$

and

$$C_1(t, \sigma, \mu) = 2Q\left(\frac{t}{\sigma}\right) - Q\left(\frac{2\mu - t}{\sigma}\right) + Q\left(\frac{2\mu + t}{\sigma}\right).$$

Notably, since $t \leq \sqrt{E} = \mu$, it is always valid that for $2s\sigma^2 < 1$,

$$2\mu b^2 - t \geq \frac{2\mu}{1+2s\sigma^2} - \mu = \mu \frac{(1-2s\sigma^2)}{(1+2s\sigma^2)} > 0.$$

B. Detailed derivation of (16)

For $s > 0$,

$$\begin{aligned}
E[e^{-s\phi_j} | \mathcal{A}_0] &= e^{-4s\mu^2} \int_{-t}^{t-2\mu} e^{-4s\mu x} f_{n_j}(x) dx + e^{-sT} \int_{t-2\mu}^t e^{sx^2} f_{n_j}(x) dx \\
&\quad + e^{sT} \int_{-t}^{-t+2\mu} e^{-sx^2} f_{n_j}(x-2\mu) dx + \left(1 - \int_{-t-2\mu}^t f_{n_j}(x) dx\right) \\
&= e^{-4s\mu^2} \int_{-t}^{t-2\mu} e^{-4s\mu x} \left[(1-p_b) \frac{1}{\sigma_0} \varphi\left(\frac{x}{\sigma_0}\right) + p_b \frac{1}{\sigma_1} \varphi\left(\frac{x}{\sigma_1}\right) \right] dx \\
&\quad + e^{-sT} \int_{t-2\mu}^t e^{sx^2} \left[(1-p_b) \frac{1}{\sigma_0} \varphi\left(\frac{x}{\sigma_0}\right) + p_b \frac{1}{\sigma_1} \varphi\left(\frac{x}{\sigma_1}\right) \right] dx \\
&\quad + e^{sT} \int_{-t}^{-t+2\mu} e^{-sx^2} \left[(1-p_b) b \frac{1}{\sigma_0} \varphi\left(\frac{x-2\mu}{\sigma_0}\right) + p_b \frac{1}{\sigma_1} \varphi\left(\frac{x-2\mu}{\sigma_1}\right) \right] dx \\
&\quad + \left(1 - \int_{-t-2\mu}^t \left[(1-p_b) \frac{1}{\sigma_0} \varphi\left(\frac{x}{\sigma_0}\right) + p_b \frac{1}{\sigma_1} \varphi\left(\frac{x}{\sigma_1}\right) \right] dx\right) \\
&= e^{-4s\mu^2} [(1-p_b)D(s, t, \sigma_0, \mu) + p_b D(s, t, \sigma_1, \mu)] \\
&\quad + e^{-sT} [(1-p_b)A_2(s, t, \sigma_0, \mu) + p_b A_2(s, t, \sigma_1, \mu)] \\
&\quad + e^{sT} [(1-p_b)B_2(s, t, \sigma_0, \mu) + p_b B_2(s, t, \sigma_1, \mu)] \\
&\quad + [(1-p_b)C_2(\mu, t, \sigma_0) + p_b C_2(\mu, t, \sigma_1)]
\end{aligned}$$

where by letting

$$a = \begin{cases} \frac{1}{\sqrt{1-2s\sigma^2}}, & 2s\sigma^2 < 1 \\ \infty, & 2s\sigma^2 = 1 \\ \frac{1}{\sqrt{2s\sigma^2-1}}, & 2s\sigma^2 > 1 \end{cases}$$

$b = \frac{1}{\sqrt{1+2s\sigma^2}}$, $\tilde{\varphi}(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}x^2}$, and $\tilde{Q}(y) = \int_0^y \tilde{\varphi}(x) dx$, we derive that

$$D(s, t, \sigma, \mu) = \int_{-t}^{t-2\mu} e^{-4s\mu x} \frac{1}{\sigma} \varphi\left(\frac{x}{\sigma}\right) dx = \int_{-t/\sigma}^{(t-2\mu)/\sigma} e^{-4s\mu\sigma y} \varphi(y) dy$$

$$\begin{aligned}
&= \int_{-t/\sigma}^{(t-2\mu)/\sigma} e^{8s^2\mu^2\sigma^2} \varphi(y + 4s\mu\sigma) dy \\
&= e^{8s^2\mu^2\sigma^2} \int_{-t/\sigma+4s\mu\sigma}^{(t-2\mu)/\sigma+4s\mu\sigma} \varphi(x) dx \\
&= e^{8s^2\mu^2\sigma^2} \int_{(\mu/\sigma)(-t/\mu+4s\sigma^2)}^{(\mu/\sigma)(t/\mu-2+4s\sigma^2)} \varphi(x) dx \\
&= \begin{cases} e^{8s^2\mu^2\sigma^2} \left(Q\left(\frac{2\mu-t}{\sigma} - 4s\mu\sigma\right) - Q\left(\frac{t}{\sigma} - 4s\mu\sigma\right) \right), & 4s\sigma^2 < 2 - t/\mu \\ e^{8s^2\mu^2\sigma^2} \left(1 - Q\left(\frac{t-2\mu}{\sigma} + 4s\mu\sigma\right) - Q\left(\frac{t}{\sigma} - 4s\mu\sigma\right) \right), & 2 - t/\mu \leq 4s\sigma^2 < t/\mu, \\ e^{8s^2\mu^2\sigma^2} \left(Q\left(-\frac{t}{\sigma} + 4s\mu\sigma\right) - Q\left(\frac{t-2\mu}{\sigma} + 4s\mu\sigma\right) \right), & t/\mu \leq 4s\sigma^2 \end{cases} \\
\\
A_2(s, t, \sigma, \mu) &= \int_{t-2\mu}^t e^{sx^2} \frac{1}{\sigma} \varphi\left(\frac{x}{\sigma}\right) dx = \int_{(t-2\mu)/\sigma}^{t/\sigma} e^{s\sigma^2 y^2} \varphi(y) dy = \int_{(t-2\mu)/\sigma}^{t/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(1-2s\sigma^2)y^2} dy \\
&= \begin{cases} \int_{(t-2\mu)/(\sigma a)}^{t/(\sigma a)} a \cdot \varphi(x) dx, & 2s\sigma^2 < 1 \\ \frac{2\mu}{\sqrt{2\pi}\sigma}, & 2s\sigma^2 = 1 \\ \int_{(t-2\mu)/(\sigma a)}^{t/(\sigma a)} a \cdot \tilde{\varphi}(x) dx, & 2s\sigma^2 > 1 \end{cases} \\
&= \begin{cases} a \left(Q\left(\frac{t-2\mu}{\sigma a}\right) - Q\left(\frac{t}{\sigma a}\right) \right), & 2s\sigma^2 < 1 \text{ and } t > 2\mu \\ a \left(1 - Q\left(\frac{t}{\sigma a}\right) - Q\left(\frac{2\mu-t}{\sigma a}\right) \right), & 2s\sigma^2 < 1 \text{ and } \mu < t \leq 2\mu \\ \frac{2\mu}{\sqrt{2\pi}\sigma}, & 2s\sigma^2 = 1 \\ a \left(\tilde{Q}\left(\frac{t}{\sigma a}\right) - \tilde{Q}\left(\frac{t-2\mu}{\sigma a}\right) \right), & 2s\sigma^2 > 1 \text{ and } t > 2\mu \\ a \left(\tilde{Q}\left(\frac{t}{\sigma a}\right) + \tilde{Q}\left(\frac{2\mu-t}{\sigma a}\right) \right), & 2s\sigma^2 > 1 \text{ and } \mu < t \leq 2\mu \end{cases} \\
&= \begin{cases} \frac{1}{\sqrt{1-2s\sigma^2}} \left(Q\left(\frac{\sqrt{1-2s\sigma^2}}{\sigma}(t-2\mu)\right) - Q\left(\frac{\sqrt{1-2s\sigma^2}}{\sigma}t\right) \right), & t > 2\mu \\ \frac{1}{\sqrt{1-2s\sigma^2}} \left(1 - Q\left(\frac{\sqrt{1-2s\sigma^2}}{\sigma}t\right) - Q\left(\frac{\sqrt{1-2s\sigma^2}}{\sigma}(2\mu-t)\right) \right), & \mu < t \leq 2\mu \\ \frac{2\mu}{\sqrt{2\pi}\sigma}, & 2s\sigma^2 = 1 \\ \frac{1}{\sqrt{2s\sigma^2-1}} \left(\tilde{Q}\left(\frac{t}{\sigma a}\right) - \tilde{Q}\left(\frac{t-2\mu}{\sigma a}\right) \right), & 2s\sigma^2 > 1 \text{ and } t > 2\mu \\ \frac{1}{\sqrt{2s\sigma^2-1}} \left(\tilde{Q}\left(\frac{t}{\sigma a}\right) + \tilde{Q}\left(\frac{2\mu-t}{\sigma a}\right) \right), & 2s\sigma^2 > 1 \text{ and } \mu < t \leq 2\mu \end{cases}
\end{aligned}$$

and

$$\begin{aligned}
B_2(s, t, \sigma, \mu) &= \int_{-t}^{-t+2\mu} e^{-sx^2} \frac{1}{\sigma} \varphi\left(\frac{x-2\mu}{\sigma}\right) dx = e^{-2\mu^2(1-b^2)/\sigma^2} \int_{-t}^{-t+2\mu} \frac{1}{\sigma} \varphi\left(\frac{x-2\mu b^2}{\sigma b}\right) dx \\
&= e^{-2\mu^2(1-b^2)/\sigma^2} \int_{(-t-2\mu b^2)/(\sigma b)}^{(-t+2\mu-2\mu b^2)/(\sigma b)} b \varphi(y) dy \\
&= b e^{-2\mu^2(1-b^2)/\sigma^2} \left(Q\left(\frac{2\mu b^2 + t - 2\mu}{\sigma b}\right) - Q\left(\frac{2\mu b^2 + t}{\sigma b}\right) \right), \\
&= \frac{1}{\sqrt{1+2s\sigma^2}} e^{-\frac{4s\mu^2}{1+2s\sigma^2}} \left(Q\left(\frac{t+2ts\sigma^2-4s\mu\sigma^2}{\sigma\sqrt{1+2s\sigma^2}}\right) - Q\left(\frac{t+2st\sigma^2+2\mu}{\sigma\sqrt{1+2s\sigma^2}}\right) \right),
\end{aligned}$$

and

$$C_2(t, \sigma, \mu) = Q\left(\frac{t}{\sigma}\right) + Q\left(\frac{t+2\mu}{\sigma}\right).$$

Notably, as $t > \mu$, it is always valid that

$$2\mu b^2 + t - 2\mu = t - \frac{4\mu s\sigma^2}{1+2s\sigma^2} > \mu - \frac{4\mu s\sigma^2}{1+2s\sigma^2} = \mu \frac{(1-2s\sigma^2)}{(1+2s\sigma^2)} > 0.$$