Convolutionally Coded Transmission over Markov-Gaussian Channels: Analysis and Decoding Metrics

Jeebak Mitra, *Student Member, IEEE*, and Lutz Lampe, *Senior Member, IEEE*

***Abstract*— It has been widely acknowledged that the aggregate interference at the receiver for various practical communication channels can often deviate markedly from the classical additive white Gaussian noise (AWGN) assumption due to various ambient phenomena. Moreover, the physical nature of the underlying interference generating process in such cases can lead to a bursty behaviour of the interfering signal, implying that it is highly likely that consecutive symbols are affected by similar noise levels. In this paper, we devise and analyze detection techniques, in conjunction with a convolution code, for such interference channels that possess non-negligible memory by considering optimum and sub-optimum decoding metrics. In particular the inherent memory in the noise process is modeled as a first-order Markov chain, whose state selects the variance of the instantaneous Gaussian noise, leading to a Markov-Gaussian channel model. Analytical expressions are obtained for the cut-off rate, which is an ensemble code parameter, and the bit error rate for a convolutionally coded system, that are subsequently employed for an extensive evaluation of the various metrics considered. Furthermore, the interleaving depth is considered as a design parameter and its effect on performance is analyzed over a range of noise scenarios.**

***Index Terms*— Impulse noise, Markov channels, Gilbert-Elliot model, Convolutional codes, Error rate analysis, Cutoff rate.**

I. Introduction

In various wireless and wireline communication systems transmission is affected not only by the omni-present thermal noise, which is faithfully modeled as an additive white Gaussian noise (AWGN) process, but also by impulse noise. For example, measurements and analysis reported in [1]–[6] have shown that the overall ambient noise experienced in wireless, wireline, and power line communication (PLC) systems and interference from co-channel and ultra-wideband (UWB) interferers exhibit a decidedly non-Gaussian behaviour. Likewise, the assumption of a memoryless noise process is not valid for many transmission scenarios, cf. e.g., [5], [7], [8] (and references therein) for PLC, UWB, and wireless transmission under partial-time jamming, when delay constraints prevent the use of deep interleaving. While a deviation from Gaussianity and the presence of memory in the noise process is beneficial in terms of an increase in channel capacity, systems based on conventional matched-filter receivers and decoding that assumes memoryless disturbance experience a considerable performance degradation in environments with bursty impulse noise.

In this paper, we study coded transmission over channels impaired by bursty impulse noise. In particular, we consider the use of convolutional coding at the transmitter and Viterbi decoding at the receiver, which is still the most popular configuration in wireless communications. As often done in literature, e.g., [9]–[14], we model the one-dimensional density of the noise process by a two-term Gaussian mixture distribution, one term of which corresponds to the presence and the other to the absence of the impulse noise component. To account for memory in the noise process, we model the sequence of mixture terms by a two-state Markov chain. We thus apply a Markov-Gaussian channel [15] whose underlying state process is the same as in the Gilbert-Elliot channel model, which has been widely used to model burst noise channels, e.g. in [8] for jamming, in [7] for UWB interference, and in [15] for bursty noise in PLC. We would like to point out that different from many of the works on Gilbert-Elliot type channels, e.g., [16]–[20], we consider non-binary channel outputs.

The related literature studying impulse noise channels often makes simplifying assumptions, such as (i) independent and identically distributed (i.i.d.), i.e., perfectly interleaved noise [10], [12], [21], [22], (ii) perfect knowledge of the noise statistics at the receiver [15], [21], [22], or (iii) use of a specific (optimal or suboptimal) decoding metric [15], [20]–[22], to either simplify the receiver design or to keep the analysis mathematically tractable. We extend the work in the aforementioned papers in that we (i) propose and compare several decoding metrics which are suitable for convolutional coded transmission over Markov-Gaussian channels and (ii) analyze the effect of finite interleaving on overall system performance. The metrics presented include two novel metrics (originally presented in [23]), which are appealing in that they require minimal knowledge about the noise statistics. While we do not attempt to modify the sequence-detection Viterbi decoder to exploit channel memory, our analysis also includes metrics based on noise-state information, which in practice would necessitate the use of a state estimator, cf. e.g. [8], [16], [24]. We derive analytical expressions for cutoff rate and tight approximations for the bit-error rate (BER) performances for the proposed metrics, thus providing a framework for performance comparisons. The obtained expressions are substantially useful in that they reveal quite explicitly the decoding behavior as a function of the decoding metric and the achievable performance gains compared to the conventional Euclidean distance metric. Likewise, the performance degradation due to finite interleaving depth is quantified in

Manuscript received May 19, 2009; revised September 29, 2009. This work was supported by the National Sciences and Engineering Research Council (NSERC) of Canada. The material in this paper was presented in part at the 2006 IEEE Global Telecommunications Conference (Globecom).

The authors are with the Deptartment of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada (e-mail: jeebakm@ece.ubc.ca, Lampe@ece.ubc.ca).

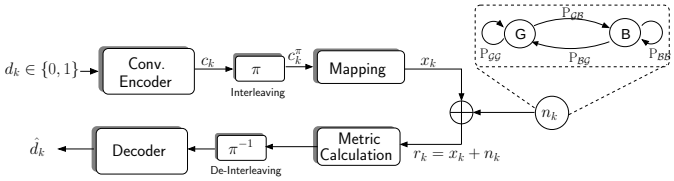

Fig. 1. System model for convolutionally coded transmission over a Markov-Gaussian channel.

terms of both the BER and cutoff rate. Our results indicate that an interleaver depth of about twice the average time spent in the bad noise state is needed to successfully disperse noise bursts. Furthermore, in conformance with previous work in the area [22], we show that considerable performance gains are achievable by considering passband transmission when the statistical dependencies between the quadrature components is exploited.

Organization: The remainder of this paper is organized as follows. Section II introduces the transceiver structure and the noise model. In Section III, various decoding metrics that have the capability to combat impulse noise are presented. Section IV is devoted to the theoretical performance analysis for transmission over Markov-Gaussian channels when applying the decoding metrics from Section III. Numerical and simulations results are presented and discussed in Section V. Finally, concluding remarks are offered in Section VI.

The following notation is used. $\Pr\{\cdot\}$ and $\mathcal{E}\{\cdot\}$ denote the probability of an event and statistical expectation, respectively. $\Re\{\cdot\}$ and $\Im\{\cdot\}$ are the real and imaginary part of a complex number. $Q(x) \triangleq 1/\sqrt{2\pi}\int_x^\infty \mathrm{e}^{-t^2/2}\mathrm{d}t$ is the Gaussian Q-function.

II. System Model

Figure 1 shows the structure and components of the overall coded transmission system. In the following, we provide a brief description of transmitter and receiver operations and the noise model.

A. Transmitter

Information bits $d_k \in \{0,1\}$ ($k \in \mathbb{Z}$ is the discrete time index) are emitted by a source with uniform probability and encoded by a binary rate-k_c/n_c convolutional encoder to produce coded bits c_k. The coded bits are interleaved and then mapped to binary phase-shift keying (BPSK) symbols by the mapper to generate transmit symbols $x_k \in \{-1,1\}$. A regular block interleaver with I rows and I_c columns is assumed, such that the encoder and the interleaver outputs (see Figure 1) are related via

$$c^\pi_{(jI+i)} = c_{(iI_c+j)} \quad 0 \le i < I\ , 0 \le j < I_c\ . \tag{1}$$

B. Channel and Receiver

The channel is assumed to be non-frequency selective and non-fading, i.e., we consider narrowband transmission with stationary or slowly mobile devices (e.g. [8], [10], [13]). The equivalent discrete-time representation of the received symbol after filtering and sampling is given by

$$r_k = x_k + n_k\ , \tag{2}$$

where n_k are samples of the noise process, a detailed description of which is provided in the following section. The received samples r_k are subsequently used to compute decision metrics that are de-interleaved and passed to the Viterbi decoder. The decoding is discussed in detail in Section III.

C. Noise Model

We consider the noise term n_k to be the additive superposition of two terms, w_k and $b_k i_k$, where w_k and i_k are zero-mean Gaussian distributed and b_k is a $\{0,1\}$-random variable. The motivation for such a model stems from the fact that while w_k represents the AWGN, $w_k + i_k$ describes the AWGN plus interference either from other users or from ambient phenomena. The former *state* is referred to as the good $(s_k = \mathcal{G} \Leftrightarrow b_k = 0)$ state, and the latter as the bad $(s_k = \mathcal{B} \Leftrightarrow b_k = 1)$ state. Conditioning on the noise state we can express the noise probability density function (pdf) by

$$p_s(n) = \exp\left(-|n|^2/(2\sigma_s^2)\right) / \left|\sqrt{2\pi}\sigma_s\right|^D\ , \tag{3}$$

where $D = 1$ for real-valued (baseband) transmission and $D = 2$ for complex-valued (passband0 transmission, and $s \in \mathcal{S} \triangleq \{\mathcal{G}, \mathcal{B}\}$. The noise variances are given by $\sigma^2_{\mathcal{G}} = \sigma^2_w$ and $\sigma^2_{\mathcal{B}} = \sigma^2_w + \sigma^2_i$, and σ^2_w and σ^2_i are the variances of w_k and i_k per *real* dimension. For future reference we define the parameter $\kappa \triangleq \sigma^2_{\mathcal{B}}/\sigma^2_{\mathcal{G}} = 1 + \sigma^2_i/\sigma^2_w$, which is indicative of the strength of the interference component compared to the thermal noise. Denoting the probability of being in the bad and good state by $\mathrm{P}_{\mathcal{B}} \triangleq \Pr\{s_k = \mathcal{B}\}$ and $\mathrm{P}_{\mathcal{G}} \triangleq \Pr\{s_k = \mathcal{G}\} = 1 - \mathrm{P}_{\mathcal{B}}$, the noise pdf is given by

$$p(n) = \mathrm{P}_{\mathcal{G}} p_{\mathcal{G}}(n) + \mathrm{P}_{\mathcal{B}} p_{\mathcal{B}}(n)\ . \tag{4}$$

This two-term noise model has been used in e.g. [9]–[14] and it is also a good approximation of Middleton's Class-A noise model [22], [25], [26]. The parallel treatment of real- and complex-valued transmission is (i) practically relevant as BPSK (or BPSK-type) transmission is also often used for carrier-based systems and (ii) allows us to compare the differences with the case of AWGN.

In the literature, often ideal (i.e., infinite) interleaving and thus independent noise samples are assumed, whereby (4) fully characterizes the noise model. However, the independence assumption may be invalid for typical finite-size interleavers. We therefore employ a first-order two-state Markov model to describe the sequence of noise states s_k, which leads to a Markov-Gaussian channel model [7], [8], [15]. The Markov chain is assumed to be irreducible, aperiodic, and stationary with transition matrix

$$\begin{aligned}\boldsymbol{T} &= \begin{bmatrix} \Pr\{s_k{=}\mathcal{G}|s_{k-1}{=}\mathcal{G}\} & \Pr\{s_k{=}\mathcal{B}|s_{k-1}{=}\mathcal{G}\} \\ \Pr\{s_k{=}\mathcal{G}|s_{k-1}{=}\mathcal{B}\} & \Pr\{s_k{=}\mathcal{B}|s_{k-1}{=}\mathcal{B}\} \end{bmatrix} \\ &\triangleq \begin{bmatrix} \mathrm{P}_{\mathcal{G}\mathcal{G}} & \mathrm{P}_{\mathcal{G}\mathcal{B}} \\ \mathrm{P}_{\mathcal{B}\mathcal{G}} & \mathrm{P}_{\mathcal{B}\mathcal{B}} \end{bmatrix}\ .\end{aligned} \tag{5}$$

Since $\boldsymbol{T}$ is row-stochastic, two parameters, e.g., $\mathrm{P}_{\mathcal{GG}}$ and $\mathrm{P}_{\mathcal{BB}}$ fully describe the state process. Furthermore, we have for the stationary distribution $\mathrm{P}_{\mathcal{G}} = (1 - \mathrm{P}_{\mathcal{BB}})/(2 - \mathrm{P}_{\mathcal{GG}} - \mathrm{P}_{\mathcal{BB}})$.

III. Decoding Metrics

We consider conventional Viterbi decoding [27] in the log-likelihood domain performing add-compare-select (ACS) operations and thus the path metric is the sum of branch (bit) metrics. Since the Euclidean-distance metric is not optimal anymore, in the following we present a number of different bit-metric formulations which are apt for decoding in channels with impulse noise. As mentioned earlier, we do not attempt to modify the widely implemented ACS Viterbi decoder architecture, but only the computation of bit metrics. Hence, we do not pursue the explicit use of the memory of the noise process for noise state estimation. However, our framework allows us to include bit metrics assuming knowledge of the instantaneous noise state, which enables a more comprehensive comparison and provides performance limits.

1) Euclidean Distance Decoder (EDD): We start with the classical Euclidean-distance metric for the trial symbol $\tilde{x}_k$ at time k given the received sample r_k, which reads

$$\lambda(\tilde{x}_k|r_k) = -|r_k - \tilde{x}_k|^2 \tag{6}$$

and is employed by a Viterbi decoder designed for AWGN at the receiver. This metric formulation is oblivious to the presence of impulse noise and will mainly serve to benchmark the performance of other decoding metrics that are described in the following.

2) Known State Maximum-likelihood (ML) Decoder (KSMLD): At the other end of the performance-complexity spectrum is the decoder that has perfect knowledge of the instantaneous noise state s_k and the pdf parameters, i.e., the pdf given by Eq. (3) is applied. Practical methods to obtain an estimate for s_k include the decision-feedback-aided state estimator [16] and the expectation-maximization (EM) algorithm as described in [8]. The KSMLD branch metric is given by

$$\lambda(\tilde{x}_k|r_k, s_k, \sigma^2_{s_k}) = -|r_k - \tilde{x}_k|^2/(2\sigma^2_{s_k}) \,. \tag{7}$$

The KSMLD marks the best performance as it operates on a channel with full side information.

3) Memoryless Maximum-likelihood Decoder (MSMLD): When the receiver is only aware of an impulsive component in the noise but is oblivious to any correlations thereof and thus only considers the one-dimensional noise density, the decoder uses the log-likelihood function

$$\lambda(\tilde{x}_k|r_k, \boldsymbol{\theta}) = \log(p(r_k - \tilde{x}_k)) \tag{8}$$

as branch metric, where $p(n)$ is given by Eq. (4). As can be seen from (8) and (4), the MSMLD requires knowledge of the noise parameters $\boldsymbol{\theta} = [\mathrm{P}_{\mathcal{G}}, \sigma^2_{\mathcal{G}}, \sigma^2_{\mathcal{B}}]$. Such a decoder has often been considered for i.i.d. non-Gaussian noise, cf. e.g. [10], [21], [22]. A useful simplification of this metric is obtained from the "max-log" approximation

$$\lambda(\tilde{x}_k|r_k, \boldsymbol{\theta}) = \max_{s \in \mathcal{S}} \left\{ \log \left[\mathrm{P}_s \, p_s(r_k - \tilde{x}_k) \right] \right\} \,, \tag{9}$$

which will be considered in the following.

4) Erasure Marking Decoder (EMD): A popular and perhaps more practical alternative to combat impulse noise is Viterbi decoding with erasure marking [24], [28]. Erasure marking could be performed either before decoding (based on the instantaneous value of r_k) or by the joint erasure marking and decoding technique developed in [24]. As an approximation of these types of decoders, we consider an ideal erasure decoder whose decoding metric is given by

$$\lambda(\tilde{x}_k|r_k, s_k) = \begin{cases} -|r_k - \tilde{x}_k|^2 \,, & \text{if } s_k = \mathcal{G}, \\ 0 & \text{if } s_k = \mathcal{B}. \end{cases} \tag{10}$$

As we see later, such an approach leads to an error floor for the considered transceiver.

5) Huber Penalty Function Decoder (HPFD): We now proceed with two novel bit metrics for Viterbi decoding in impulse noise. The first metric is adopted from robust multiuser-detector design in [12] and applies Huber's penalty function (cf. [12, Eq. (33)]):

$$\lambda(\tilde{x}_k|r_k, \sigma^2_n, \xi) = \begin{cases} \frac{-|r_k - \tilde{x}_k|^2}{2\sigma^2_n}, & \text{if } |r_k - \tilde{x}_k| \le \xi\sigma^2_n, \\ \frac{\xi^2\sigma^2_n}{2} - \xi|r_k - \tilde{x}_k|, & \text{if } |r_k - \tilde{x}_k| > \xi\sigma^2_n, \end{cases} \tag{11}$$

where $\sigma^2_n \triangleq \mathrm{P}_{\mathcal{G}}\sigma^2_{\mathcal{G}} + \mathrm{P}_{\mathcal{B}}\sigma^2_{\mathcal{B}}$ is the average noise variance and ξ is the metric parameter.

6) α-Penalty Function Decoder (α-PFD): The second new metric is based on the so-called α-detector devised in [29], again for multiuser detection. The corresponding branch metric reads

$$\lambda(\tilde{x}_k|r_k, \alpha) = \frac{1}{2\alpha} \exp\left(-\alpha|r_k - \tilde{x}_k|^2\right) \,, \quad \alpha > 0 \,. \tag{12}$$

We note that the α-PFD metric tends to the Euclidean-distance metric for $\alpha \to 0$ [29, Eq. (9)].

The bit metrics for the α-PFD are determined by adjusting a single parameter, α. Similarly, the HPFD does not require knowledge of the mixture noise parameters, but only of the average noise variance σ^2_n, and ξ needs to be adjusted. As we shall see later when considering numerical results in Section V, α and the normalized parameter $\xi\sigma_n$ can be optimized offline to provide good performance over a range of channel noise scenarios. Hence, (11) and (12) are directly applicable to the standard ACS Viterbi decoder architecture, which is a distinct implementation advantage over KSMLD- or EMD-type approaches, which require decoder modifications if the corresponding performance limits ought to be approached, cf. [8], [24]. We therefore regard these two metrics as attractive alternatives to the extreme cases of (i) not exploiting the impulse noise characteristic at all (EDD) and (ii) relying on the knowledge of the noise pdf (MSMLD) or the noise state (KSMLD, EMD).

Finally, we note that while we focus on novel metrics inspired from multiuser detection literature, there are other possible choices for single-parameter metrics. For example, metrics based on heavy-tail distributions such as the generalized Gaussian pdf and the Cauchy pdf [30] or the soft-limiting detector and improved variants of it [31], [32] follow the same rationale applied for the HPFD and α-PFD metrics.

IV. Theoretical Analysis

In this section, we derive expressions for the BER and cutoff rate achievable with the decoding metrics introduced above in Markov-Gaussian noise. Our analysis draws significant practical relevance from the fact that we explicitly take the effect of finite interleaving into account. To this end, we first specify the effective noise process including interleaving in Section IV-A. Then, the general approach to the analytical evaluation is presented in Section IV-B, while the specific expressions for the different decoding metrics are derived in Section IV-C.

A. Analysis for Markov Noise and Finite Interleaving

The size $I \times I_c$ of the interleaver (see Section II-A) is typically dependent on the maximal transmission delay acceptable for the communication system. We make the usual assumption that I_c is much larger than the decoder constraint length measured in terms of number of code symbols. Therefore, it suffices to consider the number of rows I, also referred to as the interleaver depth, and we can conveniently incorporate the interleaver-deinterleaver operation into the noise process by replacing the state transition matrix $\boldsymbol{T}$ from (5) with the I-step transition matrix $\boldsymbol{T}^I$. From the eigenvalue decomposition of $\boldsymbol{T}$ we have that

$$\boldsymbol{T}^I = \begin{bmatrix} \mathrm{P}_{\mathcal{G}} & \mathrm{P}_{\mathcal{B}} \\ \mathrm{P}_{\mathcal{G}} & \mathrm{P}_{\mathcal{B}} \end{bmatrix} + \mu^I \begin{bmatrix} \mathrm{P}_{\mathcal{B}} & -\mathrm{P}_{\mathcal{B}} \\ -\mathrm{P}_{\mathcal{G}} & \mathrm{P}_{\mathcal{G}} \end{bmatrix}, \tag{13}$$

where $\mu = (1 - \mathrm{P}_{\mathcal{BG}} - \mathrm{P}_{\mathcal{GB}}) = (1 - \mathrm{P}_{\mathcal{BG}}/\mathrm{P}_{\mathcal{G}})$ ($|\mu| < 1$) is the second eigenvalue of $\boldsymbol{T}$. Clearly, μ determines the performance as function of I, and it has been referred to as channel memory in [16]. Furthermore, μ^I can be approximated by

$$\mu^I \approx 1 - (\mathrm{P}_{\mathcal{BG}} I/\mathrm{P}_{\mathcal{G}}) \triangleq 1 - I/(\bar{\mathcal{D}}_{\mathcal{B}}\mathrm{P}_{\mathcal{G}}) \tag{14}$$

if $\mathrm{P}_{\mathcal{BG}}/\mathrm{P}_{\mathcal{G}} \ll 1$, where $\bar{\mathcal{D}}_{\mathcal{B}} = 1/\mathrm{P}_{\mathcal{BG}}$ is the average time spent in the bad state (average burst length). Hence, for given stationary probabilities, a first-order approximation is to choose the interleaver depth proportional to $\bar{\mathcal{D}}_{\mathcal{B}}$ to sufficiently disperse error bursts.

B. Performance Measures

As widely accepted performance yardsticks for convolutionally coded transmission we consider (i) the BER for given codes and (ii) the computational cutoff rate for ensembles of codes [27].

1) Bit-error Rate (BER): Since we have a linear coding and modulation scheme and an output symmetric channel, it suffices to consider the all-zero word as the transmitted code word. As commonly done for convolutional coded systems [27], we invoke the union bound to approximate the BER. There are two significant differences here with respect to the analysis for memoryless noise. First, the pairwise error probability is a function of not only the Hamming weight d of the error event, but also of the number of bad, $n_{\mathcal{B}}$, and good, $n_{\mathcal{G}} = (d - n_{\mathcal{B}})$, noise states occurring during the event. Secondly, the probability of an error event $\boldsymbol{e} = [e_1, e_2, \ldots]$ with Hamming weight $d_{\mathrm{H}}(\boldsymbol{e})$ depends on the error positions $\boldsymbol{p}(\boldsymbol{e}) = [p_1, \ldots, p_{d_{\mathrm{H}}(\boldsymbol{e})}]$, where $e_{p_i} = 1$, through the probability of noise-state sequences $\Pr\{\boldsymbol{s}|\boldsymbol{p}(\boldsymbol{e})\}$ with $\boldsymbol{s} = [s_{p_1}, \ldots, s_{p_{d_{\mathrm{H}}(\boldsymbol{e})}}]$. This leads us to the union bound on the BER as

$$P_b \leq (1/k_c) \sum_{d \geq d_{\mathrm{free}}} \sum_{n_{\mathcal{B}}=0}^{d} \mathrm{PEP}(d, n_{\mathcal{B}}) \Upsilon(d, n_{\mathcal{B}}) \,, \tag{15}$$

where d_{free} denotes the free distance of the code, $\mathrm{PEP}(d, n_{\mathcal{B}})$ is the pairwise error probability (PEP) between the all-zero word and a code word with Hamming weight d given $n_{\mathcal{B}}$ bad and $(d - n_{\mathcal{B}})$ good noise states, respectively, and

$$\Upsilon(d, n_{\mathcal{B}}) = \sum_{\{\boldsymbol{e} \in \mathcal{V} | d_{\mathrm{H}}(\boldsymbol{e}) = d\}} W(\boldsymbol{e}) \sum_{\boldsymbol{s} \in \mathcal{S}^d_{n_{\mathcal{B}}}} \Pr\{\boldsymbol{s}|\boldsymbol{p}(\boldsymbol{e})\} \tag{16}$$

In (16), $\mathcal{V}$ denotes the set of first-event error vectors [27, Section 4.4], $W(\boldsymbol{e})$ denotes the input weight for the error event $\boldsymbol{e}$, and $\mathcal{S}^d_{n_{\mathcal{B}}}$ is the set of noise-state vectors of length d with $n_{\mathcal{B}}$ bad states. The probability of the state sequence $\boldsymbol{s}$ is given by

$$\Pr\{\boldsymbol{s}|\boldsymbol{p}(\boldsymbol{e})\} = \Pr\{s_{p_1}\} \prod_{i=2}^{d_{\mathrm{H}}(\boldsymbol{e})} \mathrm{P}_{s_{p_{i-1}} s_{p_i}}^{((p_i - p_{i-1})I)} \,, \tag{17}$$

in which $\mathrm{P}_{s_{p_{i-1}} s_{p_i}}^{((p_i - p_{i-1})I)}$ are the transition probabilities according to the state transition matrix $\boldsymbol{T}^{(p_i - p_{i-1})I}$. In the evaluation of $\Upsilon(d, n_{\mathcal{B}})$ we take advantage of the generating series approach developed in [19] for finite state channels. The method of [19] however, requires adaption for our case as [19] uses a binary-output channel model. More specifically, we decompose $\boldsymbol{T}^I$ into $\boldsymbol{P}(0) + \boldsymbol{P}(1) = \boldsymbol{T}^I$, where the first column of $\boldsymbol{P}(0)$ and the second column of $\boldsymbol{P}(1)$ are zero, which corresponds to the analogous definitions in [19, Eqs. (2), (3)]. We then have that $\Upsilon(d, n_{\mathcal{B}})$ is the coefficient of $y^d \omega^{n_{\mathcal{B}}}$ in the power series $\frac{\partial \mathcal{T}(\omega, x, y)}{\partial x}\big|_{x=1}$, where $\mathcal{T}(\omega, x, y)$ is referred to as the generating series for the probability of error patterns and expressed in [19, Eq. (18)] with explicit dependence on $\boldsymbol{P}(0)$ and $\boldsymbol{P}(1)$.

A commonly used BER approximation is obtained when only the dominant error events in the union bound, i.e., the error events whose Hamming weight d does not exceed an upper limit $d_{\max}$, are considered. The computation of the corresponding truncated series $\mathcal{T}(\omega, x, y)$ can be done as described in [19, Section III. A]. Furthermore, for the case of ideal interleaving, since

$$\begin{aligned} \Upsilon(d, n_{\mathcal{B}}) &\overset{I \to \infty}{=} \binom{d}{n_{\mathcal{B}}} P_{\mathcal{B}}^{n_{\mathcal{B}}} P_{\mathcal{G}}^{(d - n_{\mathcal{B}})} \sum_{\{\boldsymbol{e} \in \mathcal{V} | d_{\mathrm{H}}(\boldsymbol{e}) = d\}} W(\boldsymbol{e}) \\ &\triangleq \binom{d}{n_{\mathcal{B}}} P_{\mathcal{B}}^{n_{\mathcal{B}}} P_{\mathcal{G}}^{(d - n_{\mathcal{B}})} W(d) \,, \end{aligned} \tag{18}$$

we only need the distance spectrum $W(d)$ of the code [27, p. 239].

The PEP in (15) can be written as

$$\mathrm{PEP}(d, n_{\mathcal{B}}) = \Pr\{\Delta(d, n_{\mathcal{B}}) < 0\}, \tag{19}$$

where

$$\begin{aligned} \Delta(d, n_{\mathcal{B}}) &\triangleq \sum_{i=1}^{n_{\mathcal{B}}} \delta_{i|\mathcal{B}} + \sum_{i=n_{\mathcal{B}}+1}^{d} \delta_{i|\mathcal{G}} \,, \\ \delta_{i|s} &\triangleq [\lambda(\tilde{x}_i = +1) - \lambda(\tilde{x}_i = -1) | s_i = s] \,, \end{aligned} \tag{20}$$

and $\lambda(\tilde{x}_i)$ are the bit-metrics presented in Section III. In case of EDD, KSMLD and EMD metrics, (20) and thus the PEP can be obtained in closed form, as will be shown in Section IV-C. For the rest of the decoding metrics, it is advantageous to proceed in the Laplace domain. Introducing the Laplace transform $\Phi_\delta(\zeta|s) \triangleq \mathcal{E}\{\mathrm{e}^{-\zeta\delta_{i|s}}\}$, and noting that conditioned on the noise state the metric differences are statistically independent, the PEP can be evaluated through the inverse Laplace transform

$$\mathrm{PEP}(d, n_{\mathcal{B}}) = \frac{1}{2\pi\mathrm{j}} \int_{\chi-\mathrm{j}\infty}^{\chi+\mathrm{j}\infty} [\Phi_\delta(\zeta|\mathcal{G})]^{(d-n_{\mathcal{B}})} [\Phi_\delta(\zeta|\mathcal{B})]^{n_{\mathcal{B}}} \frac{d\zeta}{\zeta} \,, \tag{21}$$

where $\chi > 0$ lies in the region of convergence of the integral. This integral lends itself to efficient numerical integration using Gauss-Chebyshev quadratures with N nodes [33, p. 889], [34, Eq. (10)] as follows

$$\mathrm{PEP}(d, n_{\mathcal{B}}) = \frac{1}{N} \sum_{i=1}^{N/2} \left(\Re\{\Phi(\chi+\mathrm{j}\chi\tau_i)\} + \tau_i \Im\{\Phi(\chi+\mathrm{j}\chi\tau_i)\}\right), \tag{22}$$

where $\Phi(\zeta) \triangleq [\Phi_\delta(\zeta|\mathcal{G})]^{d-n_{\mathcal{B}}} [\Phi_\delta(\zeta|\mathcal{B})]^{n_{\mathcal{B}}}$ and $\tau_i \triangleq \tan((2i-1)\pi/(2N))$. The expressions for the Laplace transforms $\Phi_\delta(\zeta|s)$ of the respective metrics are presented in Section IV-C. These expressions also play a pivotal role in computation of the cutoff rate, as explained in the next section.

2) Cutoff Rate: Since except for the KSMLD and MSMLD metrics that assume, respectively, the instantaneous and statistical knowledge about the noise process, the decision metrics considered in this work are non-ML, we employ the notion of generalized cutoff rate as an information-theoretic performance measure. The generalized cutoff rate has widely been used in the context of fading channels, e.g. [35], [36] with mismatched decoding. To this end, denoting the transmitted and received signal vectors of length L by $\boldsymbol{x} = [x_1, \ldots, x_L]$ and $\boldsymbol{r} = [r_1, \ldots, r_L]$, and introducing the decoding path metric for $\boldsymbol{x}$ given $\boldsymbol{r}$ by $\Lambda(\boldsymbol{x}|\boldsymbol{r})$, we upper bound the PEP between $\boldsymbol{x}$ and an alternative vector $\tilde{\boldsymbol{x}}$, for a given noise state sequence $\boldsymbol{s} = [s_1, \ldots, s_L]$, using the Chernoff bound

$$\mathrm{PEP}(\boldsymbol{x} \to \tilde{\boldsymbol{x}}|\boldsymbol{s}) \leq \min_{\rho \geq 0} \mathcal{E}_{\boldsymbol{r}|\boldsymbol{x},\boldsymbol{s}} \left\{\exp\left[-\rho(\Lambda(\boldsymbol{x}|\boldsymbol{r}) - \Lambda(\tilde{\boldsymbol{x}}|\boldsymbol{r}))\right]\right\}. \tag{23}$$

While the Chernoff factor ρ could be optimized for each $\boldsymbol{s}$, the simpler (and looser) upper bound

$$\begin{aligned} \mathrm{PEP}(\boldsymbol{x} \to \tilde{\boldsymbol{x}}) &\leq \min_{\rho \geq 0} \mathcal{E}_{\boldsymbol{s}} \left\{\mathcal{E}_{\boldsymbol{r}|\boldsymbol{x},\boldsymbol{s}} \{\exp\left[-\rho(\Lambda(\boldsymbol{x}|\boldsymbol{r}) - \Lambda(\tilde{\boldsymbol{x}}|\boldsymbol{r}))\right]\}\right\} \\ &\triangleq \min_{\rho \geq 0} C(\tilde{\boldsymbol{x}}, \boldsymbol{x}, \rho) \end{aligned} \tag{24}$$

for the average PEP is obtained when choosing an optimized ρ independent of $\boldsymbol{s}$ (cf. [37] for a similar approach to obtain an upper bound for block fading channels based on random coding arguments). Employing (24) allows us to express the generalized cutoff rate as

$$R_0 = \lim_{L\to\infty} \max_{\rho \geq 0} -\frac{1}{L} \log_2 \left[\mathcal{E}_{\boldsymbol{x},\tilde{\boldsymbol{x}}} \left\{C(\tilde{\boldsymbol{x}}, \boldsymbol{x}, \rho)\right\}\right] \,, \tag{25}$$

where R_0 is in bit/symbol. Exploiting the fact that additive metrics are used, i.e.,

$$\Lambda(\tilde{\boldsymbol{x}}|\boldsymbol{r}) = \sum_{k=1}^{L} \lambda(\tilde{x}_k|n_k) \,, \tag{26}$$

and that the transmitted symbols are chosen independently and uniformly distributed, the expression in (25) can be simplified to

$$\begin{aligned} R_0 = \lim_{L\to\infty} \max_{\rho\in\mathcal{R}} -\frac{1}{L} \log_2 \Big[&\sum_{s_0 \in \{\mathcal{G},\mathcal{B}\}} \Pr\{s_0\} \\ &\times \sum_{\boldsymbol{s}\in\mathcal{S}^L} \prod_{k=1}^{L} \frac{1}{2} \Pr\{s_k|s_{k-1}\} \left(\Phi_\delta(\rho|s_k) + 1\right) \Big], \end{aligned} \tag{27}$$

where ρ lies in the intersection $\mathcal{R}$ of the convergence regions of the Laplace transforms and we use the stationary distribution for the initial state s_0. With some thought [38, p. 184] and by defining

$$\begin{aligned} \boldsymbol{\Phi}(\rho) &\triangleq \frac{1}{2} \begin{bmatrix} \Phi_\delta(\rho|\mathcal{G}) + 1 & 0 \\ 0 & \Phi_\delta(\rho|\mathcal{B}) + 1 \end{bmatrix}, \\ \boldsymbol{\Pi} &\triangleq \begin{bmatrix} P_{\mathcal{G}} \\ P_{\mathcal{B}} \end{bmatrix}, \quad \mathbf{1} \triangleq \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \end{aligned}$$

R_0 can be written in a matrix form as

$$R_0 = \lim_{L\to\infty} \max_{\rho\in\mathcal{R}} -\frac{1}{L} \log_2 \left[\boldsymbol{\Pi}^T (\boldsymbol{T}^I \boldsymbol{\Phi}(\rho))^L \mathbf{1}\right] \,. \tag{28}$$

Since $\boldsymbol{T}^I \boldsymbol{\Phi}(\rho)$ is irreducible, we have that $[e_{\max}(\rho)]^L/q \leq \boldsymbol{\Pi}^T (\boldsymbol{T}^I \boldsymbol{\Phi}(\rho))^L \mathbf{1} \leq [e_{\max}(\rho)]^L q$, where $q > 1$ is the ratio of the components of the real eigenvector corresponding to the real, positive, largest eigenvalue $e_{\max}(\rho)$ of $\boldsymbol{T}^I \boldsymbol{\Phi}(\rho)$ [38, p. 184]. Therefore, we finally obtain the expression

$$R_0 = -\log_2 \left[\min_{\rho\in\mathcal{R}} e_{\max}(\rho)\right] \,. \tag{29}$$

From an optimization standpoint, it is convenient to express R_0 in terms of $e^2_{\max}(\rho)$, i.e., $R_0 = -\frac{1}{2}\log_2\left[\min_{\rho\in\mathcal{R}} e^2_{\max}(\rho)\right]$. We note that $e^2_{\max}(\rho)$ being the largest eigenvalue of the symmetric matrix $\boldsymbol{A} \triangleq \boldsymbol{T}^I \boldsymbol{\Phi}^2(\rho) (\boldsymbol{T}^I)^T$, is a convex function of the elements a_{ij}, $i, j = 1, 2$, of $\boldsymbol{A}$ [39, Example 3.10]. Furthermore, since

$$e^2_{\max}(\rho) = \frac{1}{2}\left[a_{11} + a_{22} + \sqrt{4a_{12}a_{21} + (a_{11} - a_{22})^2}\right] \,, \tag{30}$$

and $a_{ij} > 0$, $i, j = 1, 2$, it is easy to show that $e^2_{\max}(\rho)$ is also monotonically increasing in a_{ij}. Hence, from the convexity of the Laplace transform and consulting the composition rules [39, p. 86], we conclude that $e^2_{\max}(\rho)$ is a convex function of ρ, which greatly facilitates the minimization problem (29) with the assurance of a unique inflexion point. Moreover, note that in case of ideal interleaving ($I \to \infty$),

$$\begin{aligned} e_{\max}(\rho) &= \frac{1}{2}\left[\mathrm{P}_{\mathcal{G}}(\Phi_\delta(\rho|\mathcal{G}) + 1) + \mathrm{P}_{\mathcal{B}}(\Phi_\delta(\rho|\mathcal{B}) + 1)\right] \\ &\triangleq \frac{1}{2}\left[\Phi_\delta(\rho) + 1\right] \,, \end{aligned} \tag{31}$$

and the familiar expression

$$R_0 = 1 - \log_2 \left[1 + \min_{\rho \in \mathcal{R}} \Phi_\delta(\rho) \right] \tag{32}$$

is recovered [34].

C. Expressions for Different Metrics

We now present the expressions required to evaluate the PEP and cutoff rate for the different metrics introduced in Section III.

1) EDD: Substitution of (6) into (20) yields that $\Delta(d, n_\mathcal{B})$ is Gaussian distributed for EDD and the corresponding PEP can be expressed as

$$\mathrm{PEP}(d, n_\mathcal{B}) = Q\left(d / \sqrt{n_\mathcal{B}\sigma_\mathcal{B}^2 + (d - n_\mathcal{B})\sigma_\mathcal{G}^2} \right) . \tag{33}$$

Likewise, the Laplace transform can also be expressed in closed form as

$$\Phi_\delta(\zeta|s) = \exp\left(\frac{2}{\sigma_\mathcal{G}^2} \zeta \left[\frac{\sigma_s^2}{\sigma_\mathcal{G}^2} \zeta - 1 \right] \right) . \tag{34}$$

2) KSMLD: Also in the case of the KSMLD, $\Delta(d, n_\mathcal{B})$ is Gaussian distributed. The PEP and Laplace transform are obtained as

$$\mathrm{PEP}(d, n_\mathcal{B}) = Q\left(\sqrt{n_B \sigma_\mathcal{B}^{-2} + (d - n_\mathcal{B})\sigma_\mathcal{G}^{-2}} \right) \tag{35}$$

and

$$\Phi_\delta(\zeta|s) = \exp\left(2\sigma_s^{-2} \zeta(\zeta - 1) \right) , \tag{36}$$

respectively. Since the Chernoff factor $\rho = 1/2$ uniformly minimizes the eigenvalue $e_{\max}(\rho)$ for this case, the cutoff rate R_0 in (29) is also obtained in closed form.

An interesting observation here is that the ratio of the arguments of the Q-functions in (35) and (33) is the ratio of the arithmetic and the harmonic mean of the variances $\sigma_\mathcal{G}^2$ and $\sigma_\mathcal{B}^2$. Hence KSMLD is strictly superior to the EDD in terms of BER unless $\sigma_\mathcal{G}^2 = \sigma_\mathcal{B}^2$.

3) EMD: The expressions for the EMD immediately follow from those for the KSMLD by letting $\sigma_\mathcal{B}^2 \to \infty$. We note that for asymptotically large signal-to-noise ratio (SNR) the average PEP is given by

$$\sum_{n_\mathcal{B}=0}^{d} \mathrm{PEP}(d, n_\mathcal{B}) \Upsilon(d, n_\mathcal{B}) \overset{\sigma_\mathcal{G}^2 \to 0}{\longrightarrow} \frac{1}{2} \Upsilon(d, d) \tag{37}$$

which implies that the BER curve will floor out with increasing SNR when decoding with this metric. The Laplace transform for EMD is obtained as

$$\Phi_\delta(\zeta|s) = \begin{cases} \exp\left(2\sigma_s^{-2}\zeta(\zeta-1)\right), & \text{if } s = \mathcal{G} \\ 1, & \text{if } s = \mathcal{B} \end{cases} \tag{38}$$

Again, $\rho = 1/2$ uniformly minimizes $e_{\max}(\rho)$ and a closed-form expression for R_0 results from (29). Furthermore, for asymptotically large SNRs we find that $e_{\max}(\rho) \overset{\sigma_\mathcal{G} \to 0}{\to} 1 + \mathrm{P}_{\mathcal{B}\mathcal{B}}^{(I)}$, where $\mathrm{P}_{\mathcal{B}\mathcal{B}}^{(I)}$ is the transition probability according to $\boldsymbol{T}^I$ (13). Thus,

$$R_0 = 1 - \log_2(1 + \mathrm{P}_{\mathcal{B}\mathcal{B}}^{(I)}) , \tag{39}$$

which is strictly smaller than 1.

4) MSMLD: For the max-log MSMLD metric we need to evaluate the PEP based on (21) and thus are interested in the Laplace transform $\Phi_\delta(\zeta|s)$ for computation of BER as well. The transform is given by

$$\Phi_\delta(\zeta|s) = \int_n \left(\max_{a \in \mathcal{S}} \left[p_a(n+2) \right] / \max_{a \in \mathcal{S}} \left[p_a(n) \right] \right)^\zeta p_s(n) \, \mathrm{d}n \, . \tag{40}$$

This integral can be written as the sum of four integrals, where the integrand is a Gaussian pdf and the domains of integration are given by $\{n : |n|^2 > t \wedge |n+2|^2 > t\}$, $\{n : |n|^2 > t \wedge |n+2|^2 < t\}$, $\{n : |n|^2 < t \wedge |n+2|^2 > t\}$, $\{n : |n|^2 < t \wedge |n+2|^2 < t\}$ and

$$t = \frac{2\sigma_\mathcal{G}^2}{1 - \sigma_\mathcal{G}^2/\sigma_\mathcal{B}^2} \log\left(\frac{\mathrm{P}_\mathcal{G}/\mathrm{P}_\mathcal{B}}{\sigma_\mathcal{G}^D/\sigma_\mathcal{B}^D} \right) \tag{41}$$

is the threshold at which the two terms of the Gaussian mixture pdf attain the same value. A closed-form solution as sum of Gaussian Q-functions results in the real-valued channel case, while using the alternative representation of the Q-function [40] simple one-dimensional integrals need to be computed for complex-valued channels.

5) HPFD: The HPFD also requires evaluation of (21) with the corresponding metric difference expression to obtain the PEP. As in the case of MSMLD, we can express the Laplace transforms $\Phi_\delta(\zeta|s)$ in closed form as sums of Q-functions for real-valued transmission. In the complex-valued case we need to resort to numerical integration. Denoting the unit-step function as $u(x)$ and, for convenience, defining the variables $\tau_{k,\ell} \triangleq \tan((2k-1)\pi/(2N)) + \mathrm{j}\tan((2\ell-1)\pi/(2N))$, $\omega_{k,\ell} \triangleq \cos((2k-1)\pi/(2N))\cos((2\ell-1)\pi/(2N))$, $a_{k,\ell} \triangleq \xi\sigma_n^2 - |\tau_{k,\ell}|$, $b_{k,\ell} \triangleq \xi^2\sigma_n^2/2 - \xi|\tau_{k,\ell}|$, $c_{k,\ell} \triangleq \xi\sigma_n^2 - |\tau_{k,\ell}+2|$, and $d_{k,\ell} \triangleq \xi^2\sigma_n^2/2 - \xi|\tau_{k,\ell}+2|$, we can well approximate $\Phi_\delta(\zeta|s)$ using Gauss-Chebyshev quadratures [33, p. 889]

$$\begin{aligned} \Phi_\delta(\zeta|s) \approx \frac{\pi^2}{N^2} \sum_{k=1}^{N} \sum_{\ell=1}^{N} \exp\Big(-\zeta \Big[& -\frac{|\tau_{k,\ell}|^2}{2\sigma_n^2} u(a_{k,\ell}) + b_{k,\ell} u(-a_{k,\ell}) \\ & + \frac{|\tau_{k,\ell}+2|^2}{2\sigma_n^2} u(c_{k,\ell}) - d_{k,\ell} u(-c_{k,\ell}) \Big] \Big) \frac{p_s(\tau_{k,\ell})}{\omega_{k,\ell}^2} , \end{aligned} \tag{42}$$

which we found to converge well for $N = 100$ nodes.

6) α-PFD: Due to the form of the metric for the α-PFD (12) there is no closed-form expression for the PEP or the Laplace transform $\Phi_\delta(\zeta|s)$. The latter can however, be numerically computed with sufficient accuracy using Gauss-Chebyshev quadratures as above by expressing $\Phi_\delta(\zeta|s)$ as

$$\begin{aligned} \Phi_\delta(\zeta|s) \quad \approx & \\ \frac{\pi^2}{N^2} \sum_{k=1}^{N} \sum_{\ell=1}^{N} & \exp\left(-\frac{\zeta}{2\alpha} \left[\mathrm{e}^{-\alpha|\tau_{k,\ell}|^2} - \mathrm{e}^{-\alpha|\tau_{k,\ell}+2|^2} \right] \right) \frac{p_s(\tau_{k,\ell})}{\omega_{k,\ell}^2} \end{aligned} \tag{43}$$

in the complex-valued channel case, while a single summation is sufficient for real-valued channels. The PEP then follows from (21).

V. Numerical Results and Discussion

In this section, we put the analytical and semi-analytical expressions obtained in the previous section to use to (i)

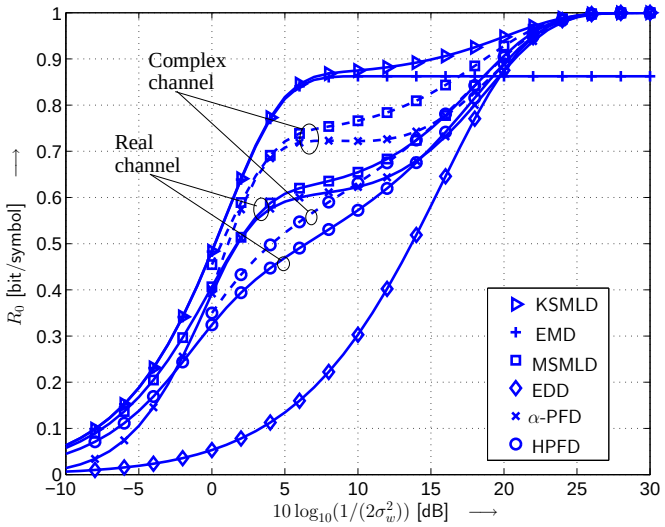

Fig. 2. Cutoff rate for decoding with different metrics (proposed in Section III) for infinite interleaving. Noise parameters: $\kappa = 100$, $\mathrm{P}_{\mathcal{B}} = 0.1$. HPFD metric with $\xi\sigma_n = 0.1$, α-PFD metric with $\alpha = 0.5$. Solid lines: baseband transmission. Dashed lines: passband transmission.

gauge the different metrics for their effectiveness in Markov-Gaussian channels when used with Viterbi decoding, (ii) obtain suitable values for the single parameters of the HPFD-metric and the α-PFD metric, (iii) study the interplay of channel memory and interleaving and their effect on performance, and (iv) substantiate the benefit of using both quadrature components in complex-valued channels.

A. Cutoff Rate

In order to clearly separate the effects of decoding metrics and interleaving, we first present cutoff rate results assuming infinite interleaving and thereafter proceed to discuss the performance degradation incurred due to finite I, using a rate loss criterion defined later.

1) Infinite Interleaving ($I\to\infty$): Figure 2 presents the R_0 results, as a function of SNR $(1/(2\sigma_w^2))$, for decoding with the different metrics for both real- and complex-valued transmission. The exemplarily considered channel noise parameters are $\kappa = 100$ and $\mathrm{P}_{\mathcal{B}} = 0.1$, which represents a channel with a strong and frequent impulse noise component. The parameters for the HPFD and α-PFD metrics are $\xi = 0.1/\sigma_n$ and $\alpha = 0.5$, respectively (see below for the optimization of these parameters).

We note that for the cases where the noise state is assumed known, i.e., KSMLD and EMD, no additional information can be drawn from the quadrature component of the received signal, and thus the R_0 curves for real and complex transmission are identical. Furthermore, since the EDD treats in-phase and quadrature components independently, it is not able to exploit the statistical dependencies between the two signal components (see noise pdf (3)), and hence only one R_0 curve is observed for the EDD in Figure 2. In contrast to this, the MSMLD, HPFD, and α-PFD utilize these dependencies and achieve notably higher rates in the complex-valued channel scenario.

The cutoff-rate curve for the EMD saturates at $R_0 = 1 - \log_2(1 + \mathrm{P}_{\mathcal{B}}) = 0.86$ bit/(channel use) [cf. (13) with $I \to \infty$ and (39)]. The cutoff rate for the KSMLD steadily approaches 1 with increasing SNR, by extracting information from noisy (bad state) received samples as well. In fact, the R_0 curve consists of two parts, which, as can be inferred from the representation for $e_{\max}(\rho)$ in (31), correspond to the good and bad noise states. This two-part characteristic of the R_0 curves also manifests for the other decoding metrics with the exception of the conventional Euclidean-distance metric, which is evidently ill-suited for the two-term mixture noise.

In terms of absolute performance the KSMLD can be considered as an idealized benchmark. Clearly, the acquisition of instantaneous state information requires additional bandwidth and computational resources. For example, the EM algorithm proposed in [8] consists of two forward-backward algorithms and pilot symbols are needed for its initialization. It is interesting to observe that, until its saturation point, the EMD performs almost as good as the KSMLD, which suggests that decoding of bad-state received signals can be omitted with negligible performance loss. We note that the EMD considered here also relies on instantaneous noise-state information. Barring KSMLD and EMD, all the other decoders thus have a significant computational advantage as no state estimation is done, whereby MSMLD is the clearly the optimal choice. However, the α-PFD, which has the distinct advantage of requiring the selection of only a single parameter, approaches the MSMLD performance closely. This is particularly remarkable considering α-PFD works without any statistical knowledge of the noise process while MSMLD perfectly knows $\boldsymbol{\theta}$. Furthermore, while the HPFD seems to suffer more when compared to the ideal cases, the gains exhibited by both α-PFD and HPFD over the conventional Euclidean-distance based decoder are significant for all but almost uncoded transmission for the noise scenario considered in Figure 2.

2) Effect of Finite Interleaving: We now turn to the case of finite interleaving depth I. We choose the exemplary state transition parameters $\mathrm{P}_{\mathcal{GB}} = 0.003$ and $\mathrm{P}_{\mathcal{BG}} = 0.025$, such that the average burst length is $\bar{\mathcal{D}}_{\mathcal{B}} = 40$ symbols and the stationary probabilities are $\mathrm{P}_{\mathcal{G}} = 0.9$ and $\mathrm{P}_{\mathcal{B}} = 0.1$ as in the previous section. Again, $\kappa = 100$ is chosen. As an indicator of the effect of finite interleaving we define the relative rate loss

$$\Delta R_0 \triangleq \frac{R_0(\infty) - R_0(I)}{R_0(\infty)}\,, \tag{44}$$

where $R_0(I)$ denotes the cutoff rate for given interleaver depth I. The rate loss ΔR_0 is plotted in Figure 3 as a function of the SNR for the decoding metrics which do not rely on knowledge of the instantaneous channel state. As discussed in Section IV-A (cf. Eq. (14)), we consider different interleaver depths parameterized by $\bar{\mathcal{D}}_{\mathcal{B}}$, namely $I = [0, \bar{\mathcal{D}}_{\mathcal{B}}/2, 2\bar{\mathcal{D}}_{\mathcal{B}}]$.

From Figure 3 we observe significant losses in the absence of interleaving ($I = 0$), which are mitigated with increasing I and virtually disappear for $I = 2\bar{\mathcal{D}}_{\mathcal{B}}$. Since $0.9 \lesssim \mathrm{P}_{\mathcal{G}} < 1$ for typical mixture noise scenarios, we conclude that configuring the interleaver depth according to double the average burst length is sufficient for most practical purposes. Note that our

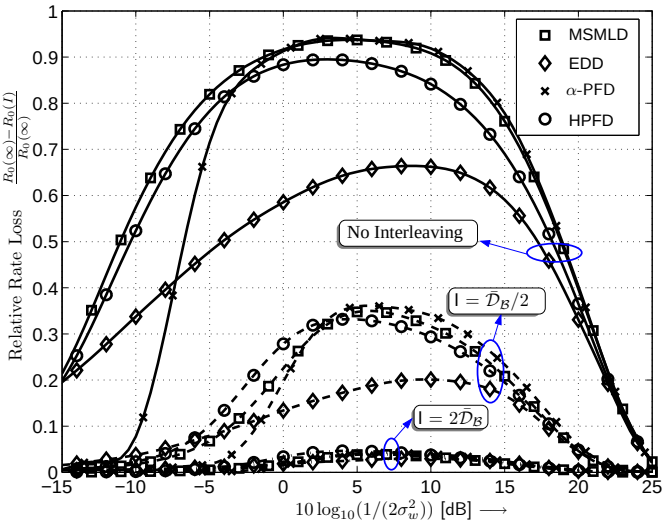

Fig. 3. Loss in cutoff rate (R_0) in the case of finite interleaving with Markov-Gaussian noise compared to memoryless noise through various levels of interleaving. Noise parameters $\kappa = 100$, $\mathrm{P}_{\mathcal{B}} = 0.1$, mean occupation time of bad state $\bar{\mathcal{D}}_{\mathcal{B}} = 40$ symbols.

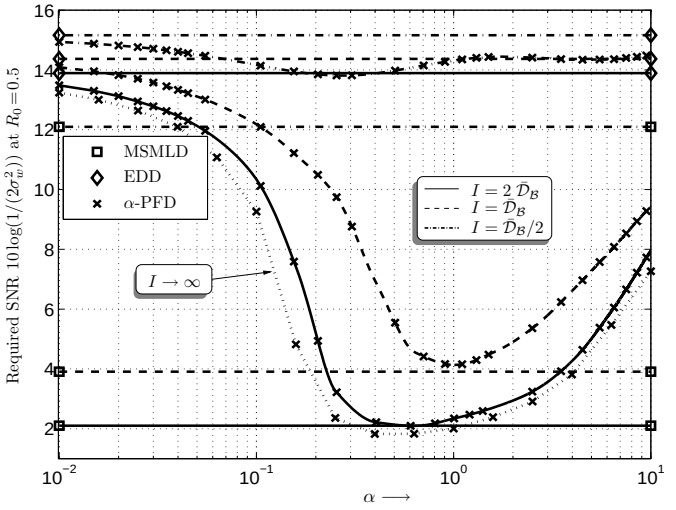

Fig. 4. Optimization of α for the α-PFD based on required SNR for a desired cutoff rate of $R_0 = 0.5$ bit/symbol. The corresponding values for MSMLD and EDD are also shown for comparison. Noise parameters: $\kappa = 100$, $\bar{\mathcal{D}}_{\mathcal{B}} = 40$, $\mathrm{P}_{\mathcal{B}} = 0.1$.

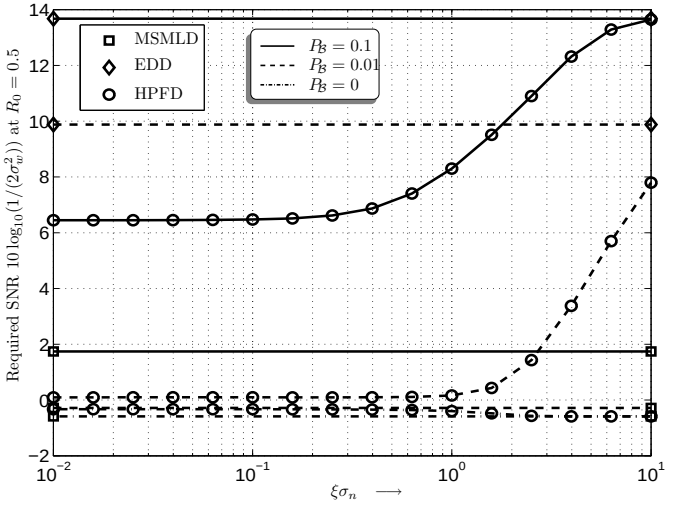

Fig. 5. Optimization of ξ for the HPFD based on required SNR for a desired cutoff rate of $R_0 = 0.5$ bit/symbol. The corresponding values for MSMLD and EDD are also shown for comparison. Infinite interleaving and $\kappa = 100$.

results are not a contradiction to the paradigm that memory increases capacity [16], since (i) the considered decoders do not attempt to make use of the channel memory and (ii) it is known that cutoff rate deteriorates with increasing channel memory even if the channel state is known [41].

3) Parameter Optimization based on Cutoff Rate: While the uni-parameter definition of the proposed α-PFD metric makes it particularly attractive, a better understanding of the metric is obtained by obtaining the optimal values for the parameter α for the various noise scenarios. In particular, we consider the SNR required to achieve a cutoff rate $R_0 = 0.5$ bit/symbol, i.e., transmission with code rate 1/2, as function of α as the optimization criterion. To this end, Figure 4 depicts the variation with increasing values of α for baseband transmission with multiple interleaver depths. As reference, curves for MSMLD and EDD are also plotted. We observe that $\alpha \in [0.5, 2]$ provides close-to-optimal performance for different interleaver depths. Furthermore, at $I = 2\bar{\mathcal{D}}_{\mathcal{B}}$ the performance for the memoryless channel is well approached, which corroborates our previous conclusions from Fig. 3. We note that the α-PFD converges to the EDD for $\alpha \to 0$, cf. [29, Eq. (9)]. This also indicates that for decreasing $\mathrm{P}_{\mathcal{B}}$ the optimum value of α will decrease. Nonetheless, the results for $\mathrm{P}_{\mathcal{B}} = 0.01$ (not shown here) reveal that $\alpha \in [0.5, 2]$ is a good choice for this case also.

HPFD being the other detector in the league of α-PFD in terms of minimality of information required, we optimize its operating parameter ξ on the same criterion as above. The optimal value(s) for ξ can be inferred from Figure 5, which shows the required SNR (as in Figure 4) as function of $\xi\sigma_n$. This time we assume $I \to \infty$ and plot results for $\mathrm{P}_{\mathcal{B}} = [0.1, 0.01, 0]$ to analyze the range from frequent impulses to AWGN channels. Evidently, relatively small values of $\xi\sigma_n$ are advantageous in impulse noise channels, whereas larger values achieve a slightly better performance in the Gaussian noise case. We note that the HPFD approaches the EDD for large values of ξ (see (11)). Similarly flat optima (as in Figure 5) were found for finite interleaver depth (not shown here). Using, e.g., $\xi = 0.1/\sigma_n$ appears to be a good compromise for all scenarios.

B. Bit-error Rate

We now present BER results obtained from the analytic expressions derived in Section IV and simulations. As a relevant example, we consider the maximum free-distance, rate-1/2, memory-4 convolutional code with generator polynomials $(23)_8$ and $(35)_8$, for which $d_{\mathrm{free}} = 7$. We apply a truncated union bound with $d_{\max} = 21$ for the infinite interleaving case and $d_{\max} = d_{\mathrm{free}}$ for the case of finite interleaving, which requires $\Upsilon(d, n_{\mathcal{B}})$ to be generated according to (16). Hence

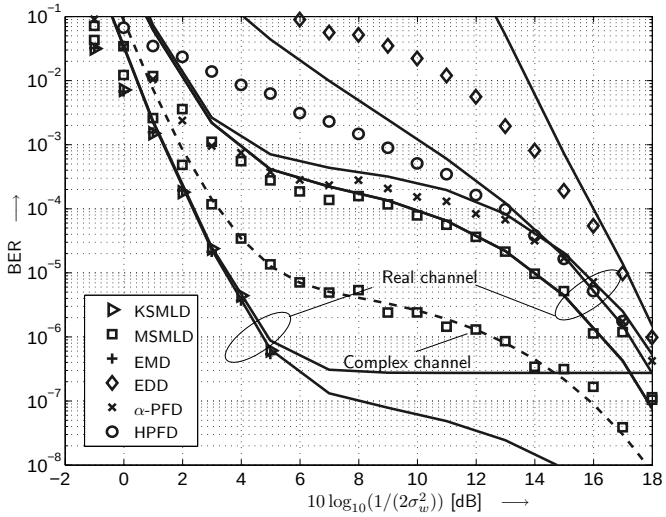

Fig. 6. BER performance of the various metrics proposed in Section III with infinite interleaving and for noise parameters $\kappa = 100$ and $\mathrm{P}_{\mathcal{B}} = 0.1$. Lines: Analytical results. Markers: Simulation.

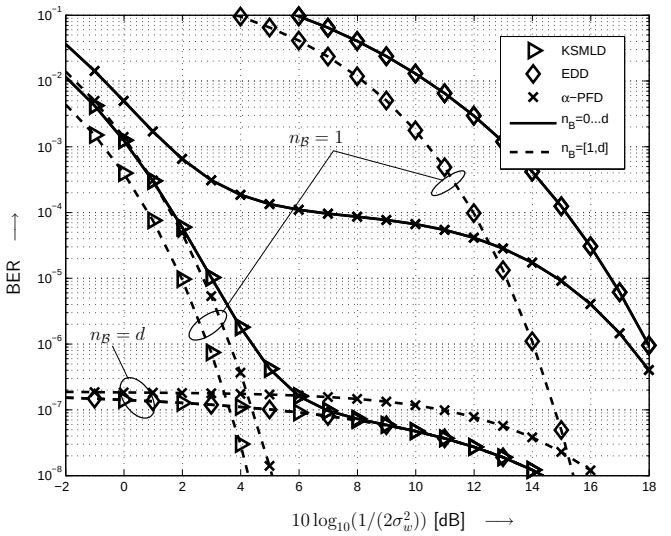

Fig. 7. Analytical BER results for different metrics proposed in Section III using infinite interleaving with noise parameters $\kappa = 100$ and $\mathrm{P}_{\mathcal{B}} = 0.1$. Only events with $d = d_{\mathrm{free}}$ are considered. Solid lines show (15) for $d = d_{\mathrm{free}}$ and $n_{\mathcal{B}} = [0 \dots d]$. Dashed lines show (15) for $d = d_{\mathrm{free}}$ and $n_{\mathcal{B}} = 1$ and $n_{\mathcal{B}} = d$, respectively.

the presented analytical BER curves are approximations, rather than a bound. The noise parameters are $\mathrm{P}_{\mathcal{B}} = 0.1$ and $\kappa = 100$.

1) Infinite Interleaving ($I \to \infty$): Figure 6 shows the BER versus SNR from the union bound (15) (lines) and from simulations (markers) for the convolutional coded system and memoryless noise process (i.e., $I \to \infty$). For the sake of clarity, complex channel results are only included for MSMLD. As in Figure 2, the parameters for the HPFD and α-PFD metrics are $\xi = 0.1/\sigma_n$ and $\alpha = 0.5$, respectively.

We observe that the union bound approximation matches the simulated BER curves very well and is fairly tight in the region of interest, which emphasizes the relevance of the PEP expressions derived in Section IV. With regards to error-rate performance, we see that, using a constant value of α, the α-PFD closely follows the MSMLD performance which is optimal in the absence of state information. Both the uni-parametric detectors, α-PFD and HPFD, clearly outperform the conventional EDD over a wide range of BERs. Furthermore, exploiting the information in the quadrature component of the received signal, if available, provides an order of magnitude improvement in BER. This is decidedly different from the case of AWGN. Finally, the significant performance gains achievable by noise-state estimation are evident from the BER curves for KSMLD and EMD. These detectors exhibit the best achievable performance for the communication framework considered with the EMD suffering from an error floor at about $\frac{1}{2}W(d_{\mathrm{free}})\mathrm{P}_{\mathcal{B}}^{d_{\mathrm{free}}} = 2 \cdot 10^{-7}$ (see Eq. (37)) due to its inherent limitations.

We note that the BER curves for the improved detectors in Figure 6 consist of two segments, most discernible for the KSMLD, which is reminiscent of error-rate curves for Turbo codes. This behaviour is made more explicit in Figure 7, where we plot the analytical BER approximation for $d = d_{\mathrm{free}}$ (solid lines) together with $\mathrm{PEP}(d_{\mathrm{free}}, n_{\mathcal{B}})\Upsilon(d_{\mathrm{free}}, n_{\mathcal{B}})$ with $n_{\mathcal{B}} = 1$ and $n_{\mathcal{B}} = d_{\mathrm{free}}$ (dashed lines) for the EDD, KSMLD, and α-PFD. Clearly, for sufficiently high SNR the BER is eventually determined by the maximal PEPs for which $n_{\mathcal{B}} = d_{\mathrm{free}}$. However, the EDD suffers from contribution of impulsive noise sequences with $n_{\mathcal{B}} < d_{\mathrm{free}}$ at relatively low SNR, e.g., with $n_{\mathcal{B}} = 1$ as shown in Figure 7. The ideal KSMLD successfully suppresses those error events and thus BER drops quickly with increasing SNR to the level of the minimum distance event ($n_{\mathcal{B}} = d_{\mathrm{free}}$), i.e., a waterfall region occurs. The proposed α-PFD approximates this behaviour, as can be seen for the case of $n_{\mathcal{B}} = 1$, which results in the significant gains over the EDD for a certain SNR range.

2) Finite Interleaving: We consider Markov-Gaussian noise with the same parameters as in Section V-A and interleaving with a short block interleaver of $I = \bar{\mathcal{D}}_{\mathcal{B}}/2 = 20$ and $I_c = 50$ (cf. Section IV-A). Figure 8 shows the analytical (lines) and simulated BER (markers) results for the different detectors. For the sake of readability of the figure, only results for real-valued transmission are shown. It can be seen that the BER expressions well approximate the simulation results for all the receivers. Furthermore, we observe that, different from infinite interleaving, the BER curves in Figure 8 tend to bunch up in the low BER region. This is a consequence of the larger multiplicative factors $\Upsilon(d, n_{\mathcal{B}})$ for $n_{\mathcal{B}} > 0$ compared to the ideally interleaved case. This also results in a rapid convergence of the BER curves for all detectors with increasing SNR. This fact is further highlighted in Figure 9, which shows the asymptotic BER approximation $\mathrm{PEP}(d_{\mathrm{free}}, d_{\mathrm{free}})\Upsilon(d_{\mathrm{free}}, d_{\mathrm{free}})$ (lines) for the EDD, MSMLD, and KSMLD, together with the corresponding simulation results (markers). Since this error event is seen to dominate performance for even moderately high SNRs, we conclude that insufficient interleaving limits the benefits of modified decoding metrics over EDD to relatively high BERs.

The necessary interleaver depth can quickly be determined by means of the analytical BER expressions derived in Section IV. To this end, Figure 10 presents the BER approxi-

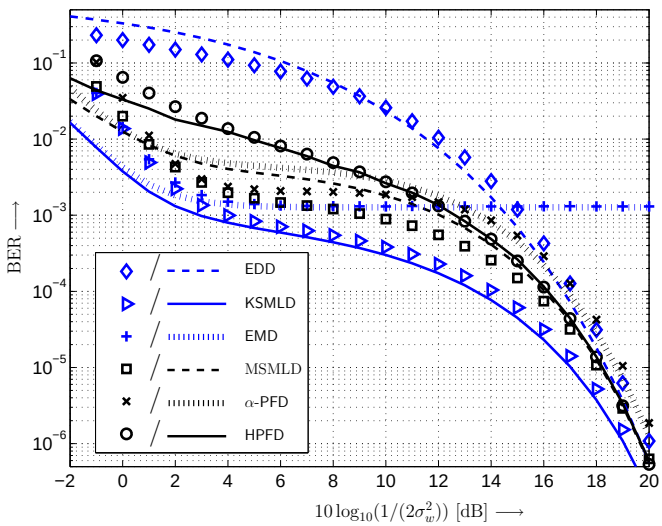

Fig. 8. BER performance of the various metrics proposed in Section III in the presence of Markov-Gaussian noise and a finite block interleaver of depth $I = 20$. Lines: Analytical results. Markers: Simulations.

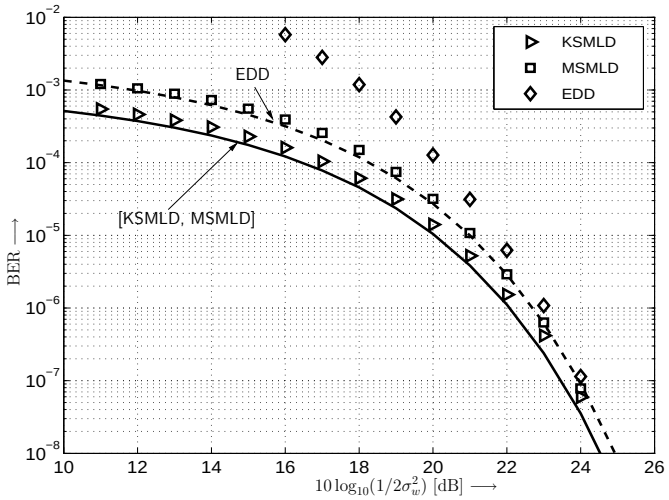

Fig. 9. Asymptotic BER results for different metrics proposed in Section III in the presence of Markov-Gaussian noise and a finite block interleaver of depth $I = 20$. Lines show BER $\approx \mathrm{PEP}(d_{\mathrm{free}}, d_{\mathrm{free}})\Upsilon(d_{\mathrm{free}}, d_{\mathrm{free}})$. Markers: Simulation results.

mations for different effective interleaver depths, specified by the ratio $I/\bar{\mathcal{D}}_{\mathcal{B}}$. We consider the conventional EDD and ideal KSMLD as benchmarks and recommend the α-PFD as an improved practical solution. We observe that $I/\bar{\mathcal{D}}_{\mathcal{B}} = 2$ leads to a BER performance close to that for ideal interleaving, which is consistent with the cutoff-rate results in Figure 3. We also note that the absolute value of $\bar{\mathcal{D}}_{\mathcal{B}}$ has negligible influence on performance, as can be seen from the curves for $I/\bar{\mathcal{D}}_{\mathcal{B}} = 1$ with $\bar{\mathcal{D}}_{\mathcal{B}} = 40$ (solid lines) and $\bar{\mathcal{D}}_{\mathcal{B}} = 20$ (dashed lines). This can be expected from the approximation in (14). Finally, we remark that the results in Figure 10 emphasize the importance of interleaving in order to realize performance gains with improved decoding metrics over the conventional EDD.

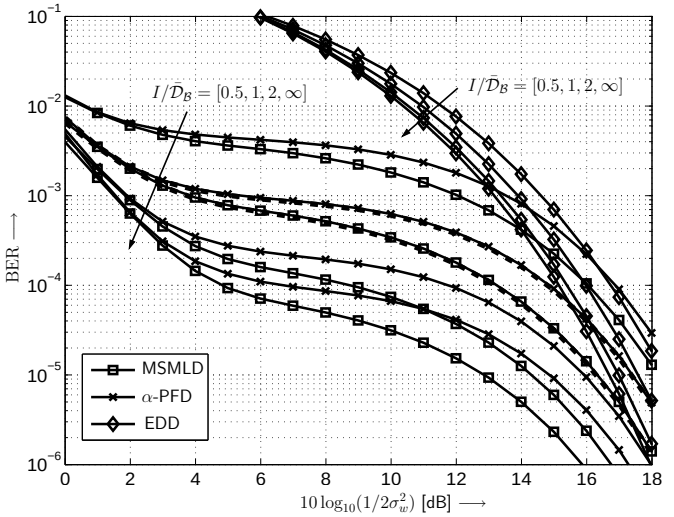

Fig. 10. BER performance for different metrics proposed in Section III in the presence of Markov-Gaussian noise and a finite block interleaver with depths $I = \bar{\mathcal{D}}_{\mathcal{B}} \times [0.5, 1, 2, \infty]$. Solid lines: $\bar{\mathcal{D}}_{\mathcal{B}} = 40$. Dashed lines (only for $I = \bar{\mathcal{D}}_{\mathcal{B}}$): $\bar{\mathcal{D}}_{\mathcal{B}} = 20$.

VI. Conclusion

In this paper, we have studied convolutionally coded transmission over Markov-Gaussian channels. We have considered and proposed several decoding metrics that are shown to be better suited to this impulse noise environment than the conventional Euclidean distance metric and whose applicability is governed by the amount of information about the noise process available at the receiver. We have derived analytical and semi-analytical expressions for the cutoff rates and BERs associated with the decoding metrics. These expressions are shown to be fairly tight in evaluating the decoding performance with finite-depth interleaving such that the effect of residual memory is well incorporated, which as a special case includes the memoryless impulse noise channel as well. Numerical evidence has been presented that confirms the usefulness of the analytical results, shows the efficacy of improved decoding metrics, and also highlights the differences to the case of transmission over AWGN channels.

References

[1] K. L. Blackard, T. S. Rappaport, and C. Bostian, "Measurements and Models of Radio Frequency Impulsive Noise for Indoor Wireless Communications," *IEEE J. Select. Areas Commun.*, vol. 11, no. 7, pp. 991–1001, Sept. 1993.

[2] W. Henkel, T. Kessler, and H. Chung, "Coded 64-CAP ADSL in an Impulse-noise Environment – Modeling of Impulse Noise and First Simulation Results," *IEEE J. Select. Areas Commun.*, vol. 13, no. 9, pp. 1611–1621, Dec. 1995.

[3] T. Blankenship, D. M. Krizman, and T. S. Rappaport, "Measurements and Simulation of Radio Frequency Impulsive Noise in Hospitals and Clinics," in *Proc. IEEE Veh. Techn. Conf.*, vol. 3, May 1997, pp. 1942 – 1946.

[4] P. Cardieri and T. Rappaport, "Statistical Analysis of Co-channel Interference in Wireless Communication Systems," *Wireless Communication and Mobile Computing*, vol. 1, no. 1, pp. 111 – 121, Jan-March 2001.

[5] M. Zimmermann and K. Dostert, "Analysis and Modeling of Impulsive Noise in Broadband Powerline Communications," *IEEE Trans. Electromagn. Compat.*, vol. 44, no. 1, pp. 249–258, Feb. 2002.

[6] A. Nasri, R. Schober, and L. Lampe, "Analysis of Narrowband Communication Systems Impaired by MB-OFDM UWB Interference," *IEEE Trans. Wireless Commun.*, vol. 6, no. 11, pp. 4090–4100, Nov. 2007.

[7] M. Flury and J.-Y. Le Boudec, "Interference Mitigation by Statistical Interference Modeling in an Impulse Radio UWB Receiver," in *IEEE Intl. Conf. on Ultra-Wideband*, Waltham, MA, USA, Sept. 2006, pp. 393–398.

[8] J.-W. Moon, T. Wong, and J. Shea, "Pilot-assisted and Blind Joint Data Detection and Channel Estimation in Partial-time Jamming," *IEEE Trans. Commun.*, vol. 54, no. 11, pp. 2092–2102, Nov. 2006.

[9] S. Kosmopoulos, P. Mathiopoulos, and M. Gouta, "Fourier-Bessel Error Performance Analysis and Evaluation of M-ary QAM Schemes in an Impulsive Noise Environment," *IEEE Trans. Commun.*, vol. 39, no. 3, pp. 398–404, Mar. 1991.

[10] M. Ghosh, "Analysis of the Effect of Impulse Noise on Multicarrier and Single Carrier QAM Systems," *IEEE Trans. Commun.*, vol. 44, no. 2, pp. 145–147, Feb. 1996.

[11] R. Blum, R. Kozick, and B. Sadler, "An Adaptive Signal Diversity Receiver for Non-Gaussian Interference and Noise," *IEEE Trans. Signal Processing*, vol. 47, no. 8, pp. 2100–2111, Aug. 1999.

[12] X. Wang and V. Poor, "Robust Multiuser Detection in Non-Gaussian Channels," *IEEE Trans. Signal Processing*, vol. 47, no. 2, pp. 289–305, Feb. 1999.

[13] H. Poor and M. Tanda, "Multiuser Detection in Flat Fading Non-Gaussian Channels," *IEEE Trans. Commun.*, vol. 50, no. 11, pp. 1769–1777, Nov. 2002.

[14] R. Pighi, M. Franceschini, G. Ferrari, and R. Raheli, "Fundamental Performance Limits of Communication Systems Impaired by Impulse Noise," *IEEE Trans. Commun.*, vol. 57, no. 1, pp. 171–182, Jan. 2009.

[15] D. Fertonani and G. Colavolpe, "On Reliable Communications over Channels Impaired by Bursty Impulse Noise," *IEEE Trans. Commun.*, vol. 57, no. 7, pp. 2024–2030, July 2009.

[16] M. Mushkin and I. Bar-David, "Capacity and Coding for Gilbert-Elliot Channels ," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1277 – 1290, Nov. 1989.

[17] L. Wilhelmsson and L. Milstein, "On the Effect of Imperfect Intleaving for the Gilbert-Elliott channel," *IEEE Trans. Commun.*, vol. 47, no. 5, pp. 681–688, May 1999.

[18] J. García-Frías and J. D. Villasenor, "Turbo Decoding of Gilbert-Elliot Channels," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 357 – 363, March 2002.

[19] C. Pimentel, "Generating Series and Performance Bounds for Convolutional Codes Over Burst-Error Channels," *IEEE Trans. Veh. Technol.*, vol. 51, no. 5, pp. 1011–1017, Sept. 2002.

[20] A. W. Eckford, F. R. Kschischang, and S. Pasupathy, "Analysis of Low-density Parity-check Codes for the Gilbert-Elliott Channel," *IEEE Trans. Inform. Theory*, vol. 51, no. 11, pp. 3872 – 3889, November 2005.

[21] S. Miyamoto, M. Katayama, and N. Morinaga, "Performance Analysis of QAM Systems under Class A Impulsive Noise Environment," *IEEE Trans. Electromagn. Compat.*, vol. 37, no. 2, pp. 260–267, May 1995.

[22] J. Häring and A. Vinck, "Performance Bounds for Optimum and Suboptimum Reception under Class-A Impulsive Noise," *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1130–1136, July 2002.

[23] J. Mitra and L. Lampe, "Robust Decoding for Impulsive Noise Channels," in *Proc. IEEE Global Telecom. Conf.*, San Francisco, CA, USA, Nov.-Dec. 2006.

[24] T. Li, W. Mow, and M. Siu, "Joint Erasure Marking and Viterbi decoding Algorithm for Unknown Impulsive Noise Channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 9, pp. 3407–3416, September 2008.

[25] D. Middleton, "Canonical and Quasicanocial Probability Models of Class-A Interference," *IEEE Trans. Electromagn. Compat.*, vol. 25, pp. 76–106, May 1983.

[26] P. Delaney, "Signal Detection in Multivariate Class-A interference," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 365–373, Feb./Mar./Apr. 1995.

[27] A. Viterbi and J. Omura, *Principles of Digital Communication and Coding*. New York: McGraw–Hill, 1979.

[28] D. Sargrad and J. Modestino, "Errors-and-erasures Coding to Combat Impulse Noise on Digital Subscriber Loops," *IEEE Trans. Commun.*, vol. 38, no. 8, pp. 1145–1155, Aug. 1990.

[29] B. Seyfe and S. Valaee, "A New Choice of Penalty Function for Robust Multiuser detection based on M-estimation," *IEEE Trans. Commun.*, vol. 53, no. 2, pp. 224–227, Feb. 2005.

[30] S. Kassam, *Signal Detection in Non-Gaussian Noise*. Berlin: Springer Verlag, 1988.

[31] N. Beaulieu and B. Hu, "An Adaptive Threshold Soft-Limiting UWB Receiver with Improved Performance in Multiuser Interference," in *IEEE Intl. Conf. on Ultra-Wideband*, Waltham, MA, USA, Sept. 2006, pp. 405–410.

[32] D. Fertonani and G. Colavolpe, "A Robust Metric for Soft-Output Detetction in the Presence of Class-A Noise," *IEEE Trans. Commun.*, vol. 57, no. 1, pp. 36–40, Jan. 2009.

[33] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*. New York: Dover Publications, 1972.

[34] E. Biglieri, G. Caire, G. Taricco, and J. Ventura-Traveset, "Computing Error Probabilities over Fading Channels: A Unified Approach," *Europ. Trans. Telecom. (ETT)*, vol. 9, pp. 15–25, Jan./Feb. 1998.

[35] D. Rainish and J. M. Perl, "Generalized Cutoff Rate of Time and Frequency-Selective Fading Channels," *IEEE Trans. Commun.*, vol. 37, no. 5, pp. 449 – 467, May 1989.

[36] C. Schlegel and D. Costello, Jr., "Bandwidth Efficient Coding for Fading Channels: Code Construction and Performance Analysis," *IEEE J. Select. Areas Commun.*, vol. 7, no. 9, pp. 1356–1368, Dec. 1989.

[37] E. Malkamäki and H. Leib, "Coded Diversity on Block Fading Channels," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 771–782, Mar. 1999.

[38] R. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, Inc., 1968.

[39] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2005.

[40] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels*. Wiley-IEEE Press, 2000.

[41] R. McEliece and W. Stark, "Channels with Block Interference," *IEEE Trans. Inform. Theory*, vol. 30, no. 1, pp. 44 – 53, January 1984.

Jeebak Mitra (S'03) received a B.Eng. degree in Electronics and Communication Engineering from Birla Institute of Technology, Ranchi, India in 2002 and an M.A.Sc. in Electrical Engineering from University of British Columbia (UBC), Vancouver, Canada in 2005. From July 2002 to July 2003 he worked as an Expert Team member in the Contact Center Solutions Group of AVAYA GlobalConnect (Tata Telecom) India. He is currently a Ph.D. candidate and Research Assistant with the Communication Theory Group at UBC.

He received the Best Student Paper Award in 2009 at the IEEE Canadian Conference in Electrical and Computer Engineering (CCECE) and won the second place in the The Ind-US Enterpreneurs (TiE) National Paper Presentation Competition in 2002. His research interests lie broadly in the area of physical layer communication for both wireless and wireline networks with an emphasis on design and analysis of receivers in interference limited environments. He also serves regularly as a reviewer for several conferences and journals in related areas.

Lutz Lampe (M'02, SM'08) received the Diplom (Univ.) and the Ph.D. degrees in electrical engineering from the University of Erlangen, Germany, in 1998 and 2002, respectively. Since 2003 he has been with the Department of Electrical and Computer Engineering at the University of British Columbia, where he is currently an Associate Professor.

He is co-recipient of the Eurasip Signal Processing Journal Best Paper Award 2005 and the Best Paper Award at the 2006 IEEE International Conference on Ultra-Wideband (ICUWB). In 2003, he received the Dissertation Award of the German Society of Information Techniques (ITG). He was awarded the UBC Killam Research Prize in 2008 and the Friedrich Wilhelm Bessel Research Award by the Alexander von Humboldt Foundation in 2009.

He is an Editor for the IEEE Transactions on Wireless Communications and the International Journal on Electronics and Communications (AEUE), and he has served as Associate Editor for the IEEE Transactions on Vehicular Technology from 2004 to 2008. He is the Vice-Chair of the IEEE Communications Society Technical Committee on Power Line Communications. He was General Chair of the 2005 International Symposium on Power Line Communications and the 2009 IEEE International Conference on Ultra-Wideband.