

Error Probability Analysis of Bit-Interleaved Coded Modulation

Alfonso Martinez, *Member, IEEE*,
 Albert Guillén i Fàbregas, *Member, IEEE*, and
 Giuseppe Caire, *Fellow, IEEE*

Abstract—This correspondence presents a simple method to accurately compute the error probability of bit-interleaved coded modulation (BICM). Thanks to the binary-input output-symmetric (BIOS) nature of the channel, the pairwise error probability (PEP) is equal to the tail probability of a sum of random variables with a particular distribution. This probability is in turn computed with a saddlepoint approximation. Its precision is numerically validated for coded transmission over standard Gaussian noise and fully interleaved fading channels for both convolutional and turbo-like codes.

Index Terms—Additive white Gaussian noise (AWGN) channel, bit-interleaved coded modulation (BICM), error probability, saddlepoint approximation, Gaussian approximation, fading channel.

I. INTRODUCTION

Bit-interleaved coded modulation (BICM) was introduced by Zehavi [1] as a pragmatic coding scheme for spectrally efficient modulations. Under the assumption of sufficient bit interleaving at the encoder output, it was later extensively studied by Caire *et al.* [2], who suggested that the system essentially behaves as a memoryless binary-input output-symmetric (BIOS) channel. This consideration allows for an easy calculation of channel capacity (average mutual information) and cutoff rate for arbitrary modulation alphabets and symbol labelings. However, the analysis of error probabilities in [2] was either not tight or exceedingly complex to compute. In this correspondence, we elaborate on their methods and obtain a simple and very accurate method to estimate the error probability.

II. ERROR PROBABILITY ANALYSIS

A. Channel Model

We study coded modulation over Gaussian noise channels. The discrete-time received signal can be expressed as

$$y_k = \sqrt{\text{SNR}} h_k x_k + z_k, \quad k = 1, \dots, L \quad (1)$$

where y_k is the (complex-valued, i.e., $y_k \in \mathbb{C}$) k th received sample, $h_k \in \mathbb{C}$ is the k th fading attenuation, $x_k \in \mathbb{C}$ is the transmitted signal

Manuscript received November 30, 2004; revised August 1, 2005. This work was supported in part by the ANTIPODE project of the French Telecommunications Research Council RNRT, and by Institut Eurécom's industrial partners: Bouygues Télécom, Fondation d'Entreprise Groupe Cégétel, Fondation Hasler, France Télécom, Hitachi, STMicroelectronics, Swisscom, Texas Instruments, and Thales. The material in this correspondence was presented in part at the 2004 Conference on Information Sciences and Systems, Princeton University, Princeton, NJ, March 2004, and at the 2004 International Symposium on Information Theory and Its Applications, Parma, Italy, October 2004.

A. Martinez is with the Department of Electrical Engineering, Technische Universiteit Eindhoven, 5600 MB Eindhoven, The Netherlands (e-mail: alfonso.martinez@ieee.org).

A. Guillén i Fàbregas is with the Institute for Telecommunications Research, University of South Australia, Mawson Lakes SA 5095, Australia (e-mail: albert.guillen@unisa.edu.au).

G. Caire was with Institut Eurécom, Sophia-Antipolis, France. He is now with the Electrical Engineering Department, University of Southern California, Los Angeles, CA 90080 USA (e-mail: caire@usc.edu).

Communicated by Ø. Ytrehus, Associate Editor for Coding Techniques.
 Digital Object Identifier 10.1109/TIT.2005.860450

at time k , and $z_k \in \mathbb{C}$ is the k th noise sample, assumed to be complex Gaussian independent and identically distributed (i.i.d.) $\sim \mathcal{N}_{\mathbb{C}}(0, 1)$. BICM codewords $\mathbf{x} = (x_1, \dots, x_L)$ are obtained by bit interleaving the codewords $\mathbf{c} = (c_1, \dots, c_N)$ of the code \mathcal{C} , each of dimension K information bits and length N , and mapping over the signal constellation \mathcal{X} with the labeling rule $\mu : \{0, 1\}^M \rightarrow \mathcal{X}$, $M = \log_2 |\mathcal{X}|$. The corresponding transmission rate is $R = \frac{KM}{N}$ bits per channel use. The average received signal-to-noise ratio is SNR. We denote the vector of received symbols by $\mathbf{y} = (y_1, \dots, y_L)$. The standard additive white Gaussian noise (AWGN) and fully interleaved Rayleigh-fading channels are obtained from (1) by simply letting $h_k = 1$ and $h_k \sim \mathcal{N}_{\mathbb{C}}(0, 1)$, respectively.¹ The operation is depicted in Fig. 1.

B. Error Probability Under ML Decoding

For maximum-likelihood (ML) decoding, the error probability of linear binary codes over BIOS channels is accurately given by the union bound in the region above the cutoff rate [3]. Let A_d denote the number of codewords in \mathcal{C} with Hamming weight d . In the region above the cutoff rate, the codeword error probability is very closely upper-bounded by

$$P_e \leq \sum_d A_d \text{PEP}(d, \mu, \mathcal{X}, \text{SNR}) \quad (2)$$

where $\text{PEP}(d, \mu, \mathcal{X}, \text{SNR})$ is the pairwise error probability (PEP) for two codewords differing in d bits.² Estimating the error probability reduces therefore to computing the PEP. Assuming that codeword \mathbf{c} was transmitted, the probability of choosing a candidate codeword \mathbf{c}' at Hamming distance d from \mathbf{c} is given by

$$\begin{aligned} \text{PEP}(d, \mu, \mathcal{X}, \text{SNR}) &= \Pr(\Pr(\mathbf{c}' | \mathbf{y}) > \Pr(\mathbf{c} | \mathbf{y}) | \mathbf{c}) \\ &= \Pr\left(\left(\sum_i \log \frac{\Pr(c'_i | y_{k(i)})}{\Pr(c_i | y_{k(i)})}\right) > 0 | \mathbf{c}\right) \end{aligned}$$

where we have used that the i th bit depends only its corresponding channel output $y_{k(i)}$. The random elements in the channel output include the noise and fading realizations z and h , respectively, the particular modulation symbol x , and the bit position in the binary label m . In order to avoid cumbersome notation, we group them in a vector $\mathcal{Y} \triangleq (z, h, x, m)$; \mathcal{Y} depends on the modulation alphabet \mathcal{X} , the labeling μ , and SNR. Taking into account that only the bit positions for which $c'_i \neq c_i$ must be considered, the PEP is given by

$$\text{PEP}(d, \mu, \mathcal{X}, \text{SNR}) = \Pr\left(\sum_{j=1}^d \Lambda_j > 0\right) \quad (3)$$

where we have defined a new random variable, denoted by Λ , the *a posteriori* log-likelihood ratio, as

$$\Lambda = \log \frac{\Pr(\hat{c} = \bar{c} | \mathcal{Y})}{\Pr(\hat{c} = c | \mathcal{Y})}. \quad (4)$$

Thanks to the presence of the interleaver [2], the variables Λ can be considered, to a practical extent, i.i.d. Furthermore, due to the sym-

¹We assume perfect channel state information (CSI) at the receiver. However, the extension of technique described here to the nonperfect CSI case is straightforward.

²Similarly, the bit-error probability P_b is given by the right-hand side of (2) with A_d replaced by $\bar{A}_d = \sum_i \frac{i}{K} A_{i,d}$, $A_{i,d}$ being the number of codewords in \mathcal{C} with output Hamming weight d and input weight i .

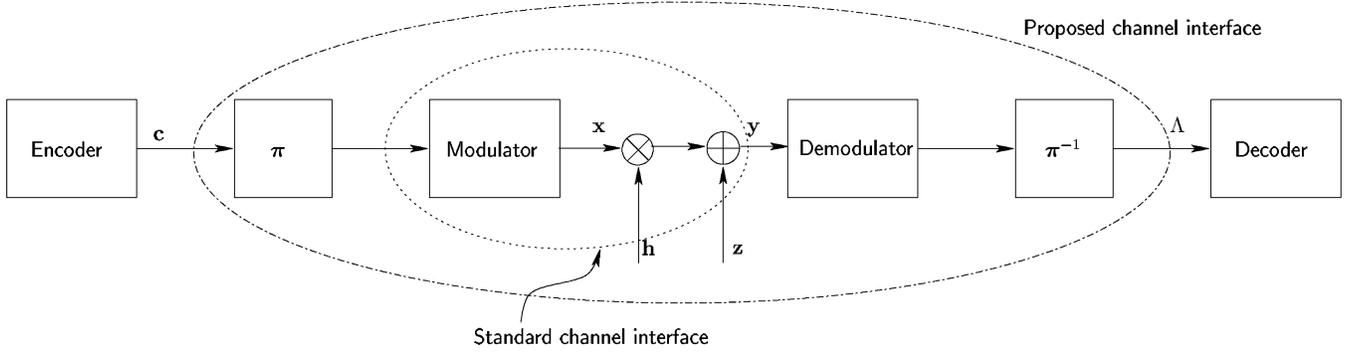


Fig. 1. Channel interfaces: standard nonbinary symbols at channel level, or at demodulator level, with binary symbols.

metry of the channel output,³ their distribution does not depend on the value of c , and we can safely assume that the all-zero codeword has been transmitted.

It should be noted that this formulation is simply a restatement of the results in [2] with a different notation. In particular, the exact dependence of the error probability on the modulation symbol or the bit index is dropped, or rather considered another random variable similar to the noise or fading realizations. Fig. 1 shows the location of Λ in the communication channel, after the demodulator.

The *a posteriori* probabilities used in the computation of Λ are given by

$$\begin{aligned} \Pr(\hat{c} = c | \mathcal{V}) &\triangleq \Pr(\hat{c} = c | z, h, x, m) \\ &\propto \sum_{x \in \mathcal{X}_c^m} \exp(-|y - \sqrt{\text{SNR}}hx|^2) \end{aligned} \quad (5)$$

where \mathcal{X}_c^m is the subset of signal constellation points with m th binary label position equal to c .

In [2], three alternative methods were given to compute $\text{PEP}(d, \mu, \mathcal{X}, \text{SNR})$: the Bhattacharyya-union bound (B-UB), the BICM bound, and the expurgated BICM union bound (ex-UB). Of these, the B-UB will be analyzed later. The BICM bound was used as a means to derive the tighter expurgated bound and, therefore, we do not analyze it further. It is interesting to note that a careful examination of the expression for the expurgated bound in [2] reveals that it is equal to (4) restricting the sum in (5) to one single term, the nearest neighbor. Proceeding directly from the assumption of a memoryless BIOS channel, their derivation can be significantly shortened. Furthermore, for non-Gray labeling, the effect of the other neighbors is not negligible, and thus the ex-UB may not be accurate [2].

C. Log-Likelihood Ratio Distribution

For some BIOS channels, the ratio Λ has a known and easily manageable distribution. For example, for the binary-symmetric channel (BSC) Λ is a binomial random variable. For the binary-input AWGN channel with signal-to-noise ratio SNR, Λ is normally distributed $\mathcal{N}(-4\text{SNR}, 8\text{SNR})$. A little algebra shows that for binary-input Rayleigh-fading channels, the density is two-sided exponential

$$\begin{aligned} f_{\Lambda}(\Lambda) &= \frac{1}{4\sqrt{\text{SNR}(1+\text{SNR})}} \\ &\times \exp\left(-\frac{\Lambda}{2} \left(1 + \text{sign}(\Lambda) \sqrt{\frac{1+\text{SNR}}{\text{SNR}}}\right)\right). \end{aligned} \quad (6)$$

Even though a closed-form expression for the density of Λ for BICM seems difficult to obtain, it is nevertheless simple to evaluate it by computer simulation if required.

³For signal constellations \mathcal{X} that lead to a BICM channel which is not symmetric, the channel can be rendered BIOS by using the mapping μ and its complement $\bar{\mu}$ with probability $1/2$ [2].

In estimates of tail probabilities, the cumulant transform $\kappa(s)$ (or cumulant generating function) of a random variable Λ is a more convenient representation than the density. The transform is given by

$$\kappa(s) \triangleq \log \mathbb{E}[e^{s\Lambda}] \quad (7)$$

with $s \in \mathbb{C}$ [4]. Using the definition of Λ , we rewrite $\kappa(s)$ as

$$\kappa(s) = \log \mathbb{E}_{\mathcal{V}} \left[\left(\frac{\Pr(\hat{c} = 1 | \mathcal{V})}{\Pr(\hat{c} = 0 | \mathcal{V})} \right)^{s\gamma} \right] \quad (8)$$

where the subscript \mathcal{V} indicates that the expectation is taken with respect to all nuisance parameters $\mathcal{V} = (z, x, h, m)$. This expectation can be easily evaluated by numerical integration using the Gauss–Hermite (for the AWGN channel) and a combination of the Gauss–Hermite and Gauss–Laguerre (for the fading channel) quadrature rules, which are tabulated in [5].

It will also prove convenient to define the saddlepoint \hat{s} as the value for which $\kappa'(\hat{s}) = 0$. It can be shown that this point exists and is unique [6]. For BIOS channels, symmetry dictates that the saddlepoint is placed at $\hat{s} = 1/2$, with no need to carry an explicit numerical minimization step [7].

Fig. 2 shows the computer-simulated density of Λ for 16-QAM over an AWGN channel and 8-PSK over a Rayleigh-fading channel with $\text{SNR} = 12$ dB and $\text{SNR} = 7$ dB, respectively. In both cases, the labeling is Gray. For the sake of comparison, Fig. 2 also shows the distribution of a Gaussian random variable with distribution $\mathcal{N}(-4\gamma, 8\gamma)$, with $\gamma = -\kappa(\hat{s})$. It should be noted that this Gaussian approximation is valid in the tail of the distribution, rather at the mean as would be the case for the standard Gaussian approximation $\mathcal{N}(\mathbb{E}[\Lambda], \mathbb{E}[\Lambda^2] - \mathbb{E}[\Lambda]^2)$. It is remarkable how close the tails are to the tail of a Gaussian random variable for the case of AWGN. For the Rayleigh fading, the density inherits the exponential behavior of the binary-input case, and the Gaussian approximation to the tail is somewhat less accurate.

D. Gaussian Approximation

The preceding discussion suggests approximating the PEP by

$$\text{PEP}(d, \mu, \mathcal{X}, \text{SNR}) \simeq Q(\sqrt{-2d\kappa(\hat{s})}) \quad (9)$$

a result which was heuristically introduced in [8]. The approximation in (9) corresponds as well to the zeroth-order term in the Lugannani–Rice formula [9] (see also [10]).

E. Bhattacharyya Union Bound

The Bhattacharyya bound [7] can be used to upperbound the PEP as

$$\begin{aligned} \text{PEP}(d, \mu, \mathcal{X}, \text{SNR}) &\leq e^{d\kappa(\hat{s})} \quad (10) \\ &= \left(\mathbb{E}_{\mathcal{V}} \left[\sqrt{\frac{\Pr(\hat{c} = 1 | \mathcal{V})}{\Pr(\hat{c} = 0 | \mathcal{V})}} \right] \right)^d. \end{aligned} \quad (11)$$

Notice that this coincides with the Chernoff bound as $\hat{s} = 1/2$. Using this in (2) we obtain the B-UB proposed in [2].

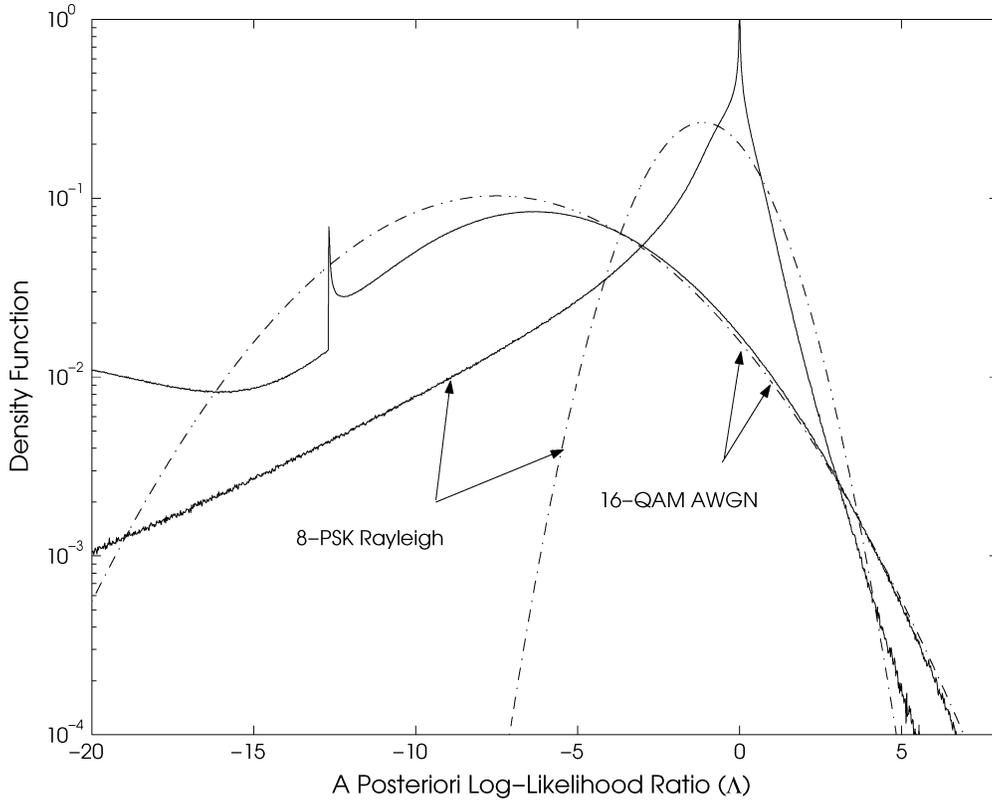


Fig. 2. Density of the a posteriori log-likelihood ratio Λ : empirical distribution (solid line, computer simulated) and Gaussian approximation to the tail (dash-dotted) for 16-QAM/8-PSK, Gray mapping, and AWGN/Rayleigh fading.

F. Saddlepoint Approximation

In the Appendix I, we present the derivation of the saddlepoint approximation and of an estimate of the approximation error to the PEP. Even though the derivation in the Appendix is uniformly valid for all values of the saddlepoint \hat{s} , including small values of \hat{s} , in our case this is not required as $\hat{s} = 1/2$. Keeping only the first-order term in the asymptotic series, the PEP can be approximated by

$$\text{PEP}(d, \mu, \mathcal{X}, \text{SNR}) = \frac{1}{\sqrt{2\pi d\kappa''(\hat{s})}} e^{d\kappa(\hat{s})} \left\{ 1 + O(d\kappa''(\hat{s}))^{-1} \right\} \quad (12)$$

where the term $O(d\kappa''(\hat{s}))^{-1}$ decays fast as a power of $(d\kappa''(\hat{s}))^{-1}$. The effect of the correction is found to be negligible in practical calculations, which implies that we need not sum over any more terms in the asymptotic series and we may then drop the $O(\cdot)$ term.

The exponent is the same as for the Bhattacharyya bound, in accordance to the asymptotic optimality of the latter, and coincides as well with the exponential decay of the Gaussian approximation. Note that efficient computation of the second derivative $\kappa''(\hat{s})$

$$\begin{aligned} \kappa''(\hat{s}) &= \frac{\mathbb{E}[\Lambda^2 e^{\hat{s}\Lambda}]}{\mathbb{E}[e^{\hat{s}\Lambda}]} \\ &= \frac{1}{\mathbb{E}[e^{\hat{s}\Lambda}]} \mathbb{E}_{\mathcal{V}} \left[\left(\log \frac{\Pr(\hat{c} = 1 | \mathcal{V})}{\Pr(\hat{c} = 0 | \mathcal{V})} \right)^2 \sqrt{\frac{\Pr(\hat{c} = 1 | \mathcal{V})}{\Pr(\hat{c} = 0 | \mathcal{V})}} \right] \end{aligned} \quad (13)$$

can again be performed using Gaussian quadrature rules.

It is worthwhile remarking that the method advocated in [2] to compute this probability for the expurgated union bound (UB) was the use of integration in the complex plane. It can be seen⁴ that the saddlepoint

method is an alternative to the complex-plane integration. Instead of directly computing the integral, its value is very accurately approximated with a method of significantly lower complexity.

III. NUMERICAL RESULTS AND DISCUSSION

In this section, we show some numerical results that illustrate the accuracy of the proposed methods as well as its asymptotic behavior. In particular, we show the following: the B-UB, the saddlepoint approximation (12) union bound (SP-UB), the Gaussian approximation tangential-sphere bound (GA-TSB) [8],⁵ and the simulation of the bit-error rate (BER sim). For every block of information bits a different bit interleaver is randomly generated.

A. AWGN Channel

Figs. 3 and 4 show the bit-error probability as a function of $E_b/N_0 = \text{SNR}/R$ for the aforementioned methods and for convolutional and repeat-accumulate (RA) codes with 16-QAM in the AWGN channel with no fading. In Fig. 3, we use the optimum 64-state and rate-1/2 convolutional code with Gray and set partitioning mappings and in Fig. 4 an RA code [12] of rate 1/4 with Gray mapping.

The performance at medium-to-high signal-to-noise ratio is very well approximated by both the Gaussian and the saddlepoint approximations, for all considered labelings and codes. Note that the performance estimate in the case of set-partitioning labeling remarkably improves the bound presented in [2]. In essence, this can be traced back to the accuracy of the Gaussian approximation to the tail of the log-likelihood ratios Λ , already discussed in Section II-C. The B-UB yields the correct decay of the bit error curve but it remains at a fixed gap from the true bit error probability. The accuracy of

⁵This is the standard tangential sphere bound [11] for a binary-input AWGN channel with $\text{SNR} = -\kappa(\hat{s})$.

⁴With the caveat indicated at the end of Section II-B on the metrics (5).

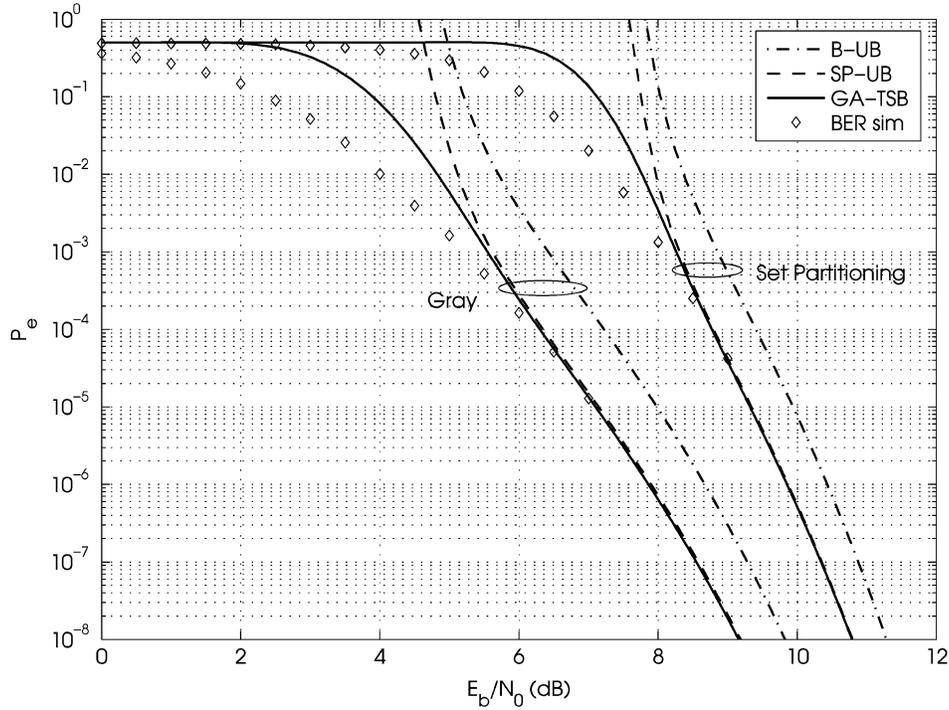


Fig. 3. Comparison of simulation and saddlepoint and Gaussian approximations on the BER of BICM with a 64-state, rate-1/2 convolutional code with 16-QAM modulation with Gray and set partitioning mapping.

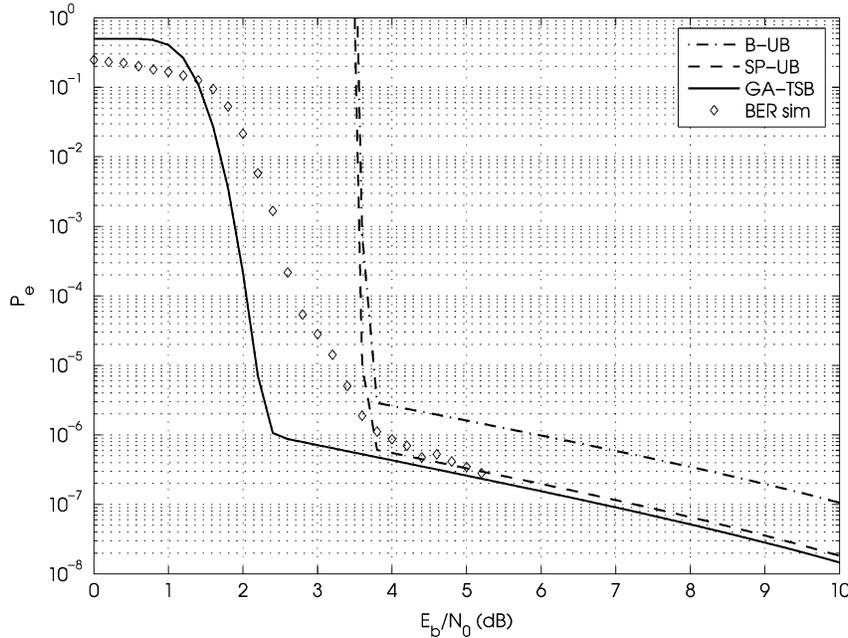


Fig. 4. Comparison of simulation and saddlepoint and Gaussian approximations on the BER of BICM with a RA code of rate 1/4 with 16-QAM modulation and Gray mapping, $K = 1024$ information bits, 20 iterations of belief propagation decoding the AWGN channel.

the union bound-based approximations for the RA code ensemble appears only in the error floor region, since the union bound is not tight for random-like codes for SNR below the corresponding cutoff rate. Nevertheless, the GA-TSB yields a fairly good estimate of the waterfall behavior of the error curve also for low SNR.

In all cases, the decay of the bit error for increasing signal-to-noise ratio seems to be of exponential nature. Appendix III proves the asymptotic validity of this conjecture and shows that

$$\lim_{\text{SNR} \rightarrow \infty} \frac{\kappa(\hat{s})}{\text{SNR}} = -\frac{d_{\min}^2}{4} \quad (14)$$

where d_{\min} is the minimum Euclidean distance of the constellation. As outlined in the proof, at large SNR, BICM behaves as a binary modulation with distance d_{\min} , regardless of the mapping. This result confirms that BICM preserves the properties of the underlying binary code and that for large SNR the error probability decays exponentially with SNR as $e^{-\frac{1}{4}d_{\min}^2 \text{SNR}}$. In this line, Fig. 5 shows $-\frac{\kappa(\hat{s})}{\text{SNR}}$ for 16-QAM with Gray and set partitioning mappings in the AWGN channel. The asymptotic value is $\frac{d_{\min}^2}{4} = 0.1$, as established by the preceding result. In the Gaussian approximation, the quantity $-\frac{\kappa(\hat{s})}{\text{SNR}}$ can be interpreted as the SNR scaling with respect to SNR when using BICM [8] and

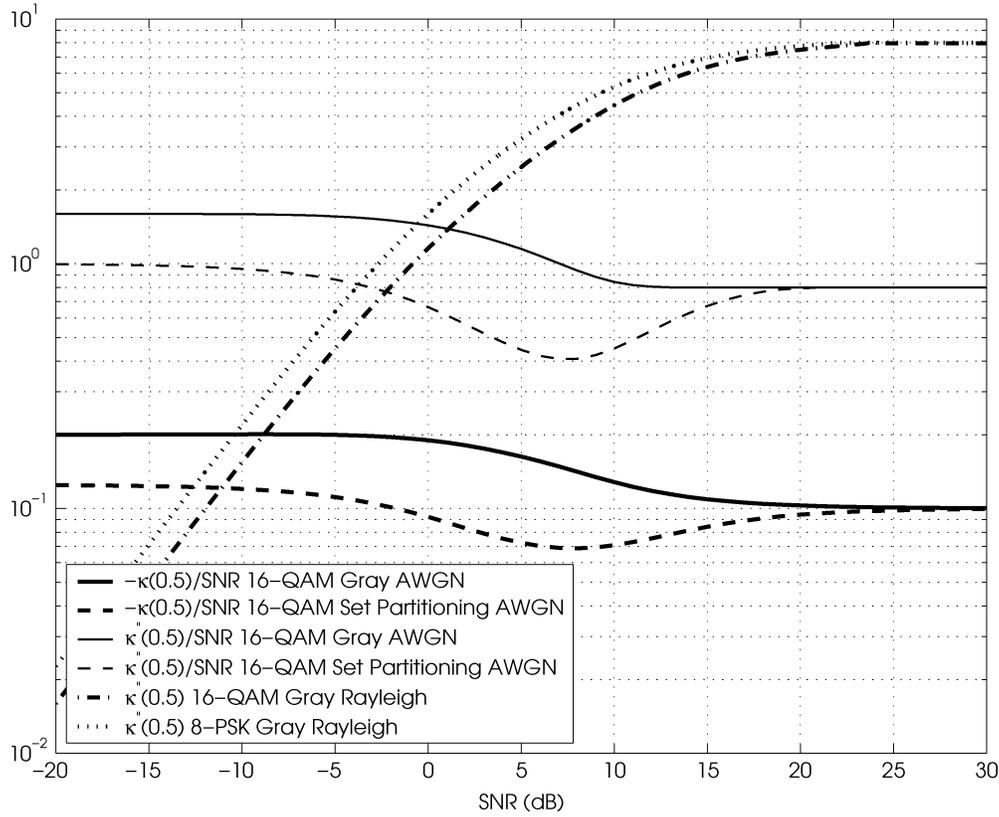


Fig. 5. Cumulant limits: $-\frac{\kappa(\hat{s})}{\text{SNR}}$ and $\frac{\kappa''(\hat{s})}{\text{SNR}}$ for 16-QAM modulation with Gray and set partitioning mapping in the AWGN channel and $\kappa''(\hat{s})$ for 16-QAM and 8-PSK with Gray mapping in the Rayleigh-fading channel.

thus, the asymptotic scaling depends only on the signal constellation \mathcal{X} (through its minimum distance) and not on the labeling μ .

As shown in Appendix I, the second-order cumulant evaluated at the saddlepoint $\kappa''(\hat{s})$ plays an important role in assessing the error of the approximation. Appendix III also shows that

$$\lim_{\text{SNR} \rightarrow \infty} \frac{\kappa''(\hat{s})}{\text{SNR}} = 2d_{\min}^2. \quad (15)$$

Fig. 5 also shows $\frac{\kappa''(\hat{s})}{\text{SNR}}$ for 16-QAM with Gray and set partitioning mappings in the AWGN channel. The limit coincides with the above result, and implies that, in the AWGN channel, the saddlepoint approximation becomes more and more accurate as SNR grows.

B. Fully Interleaved Rayleigh Fading Channel

Figs. 6 and 7 show the estimates of the bit error probability for convolutional and RA codes, respectively, in a fully interleaved AWGN channel with Rayleigh fading. Fig. 6 shows two cases, a rate-2/3, 8-state optimum code over 8-PSK, and the rate-1/2, 64-state optimum code over 16-QAM both with Gray mapping. Fig. 7 shows the performance of an RA code of rate 1/4 with Gray mapping and 16-QAM modulation.

Similarly to the AWGN case, the three approximations to the error rate give the correct slope of the decay with SNR at medium-to-high signal-to-noise ratio, while the horizontal shift of the curves is different. All approximations are close to the simulated value, but now only the saddlepoint approximation gives an accurate estimate. As we saw in Section II-C, the tail of the log-likelihood ratio Λ in the fading channel is approximately exponential, rather than Gaussian, and this shape is not correctly tracked by the Gaussian approximation. On the contrary,

the saddlepoint approximation is able to “learn” the shape of the variable. As evidenced by the results of 16-QAM with the 64-state convolutional code, this effect becomes less apparent for more powerful codes with large minimum distance, since the sum in (3) contains more terms and its tail is closer to a Gaussian. Again, the GA-TSB yields the most accurate estimate of the error probability in the low-SNR region. The accuracy of the UB-based approximations for the RA code ensemble is accurate in the error floor region.

Note also that BICM preserves the properties of the underlying binary code for fully interleaved Rayleigh-fading channels as well, as the error probability decays as an inverse power of SNR. Appendix III shows that in the limit for large SNR the rate of decay varies as

$$\lim_{\text{SNR} \rightarrow \infty} \frac{\kappa(\hat{s})}{\log \text{SNR}} = -1 \quad (16)$$

and that

$$\lim_{\text{SNR} \rightarrow \infty} \kappa''(\hat{s}) = 8 \quad (17)$$

confirming that BICM indeed behaves as a binary modulation and thus, the asymptotic performance depends on the Hamming distance of the code rather than on the Euclidean distance. Fig. 5 also shows $\kappa''(\hat{s})$ as a function of SNR for 16-QAM and 8-PSK with Gray mapping in the fully interleaved Rayleigh-fading channel. As expected, the limit value is 8, and does not depend on the modulation.

IV. CONCLUSION

In this correspondence, we have presented a simple method to compute a tight approximation to the error probability of BICM. This probability is found to correspond in a natural way to the tail probability of a sum of independent random variables, which is calculated using the saddlepoint approximation. The exact form of the approximation

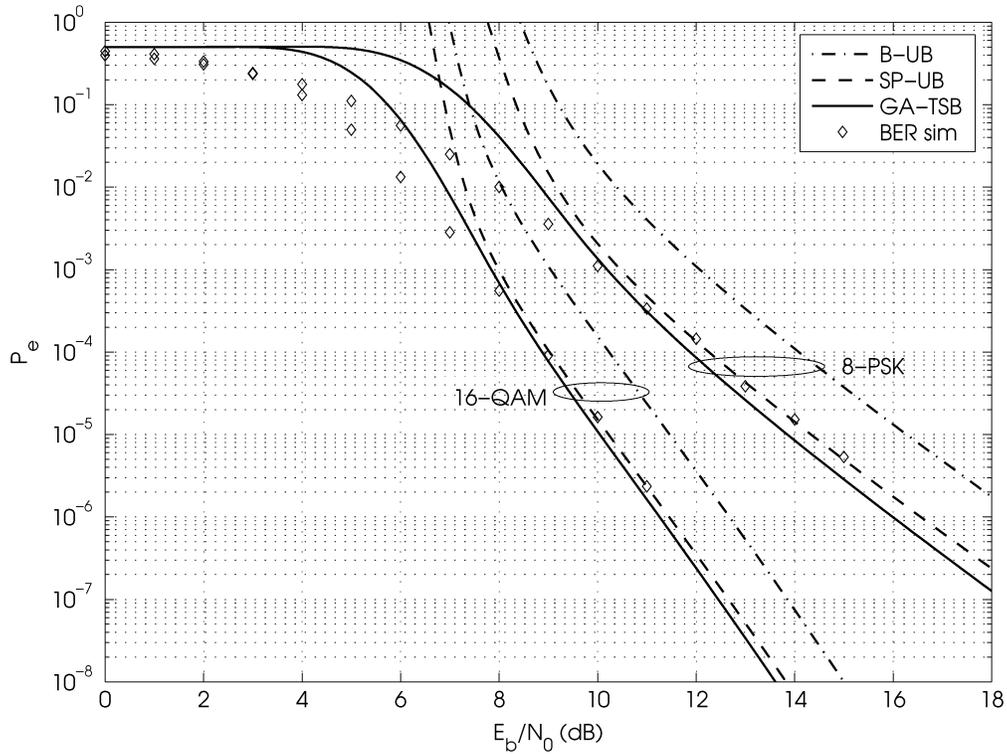


Fig. 6. Comparison of simulation and saddlepoint and Gaussian approximations on the bit error rate of BICM with a 8-state, rate-2/3 convolutional code with 8-PSK modulation and a 64-state, rate-1/2 convolutional code with 16-QAM, both with Gray mapping.

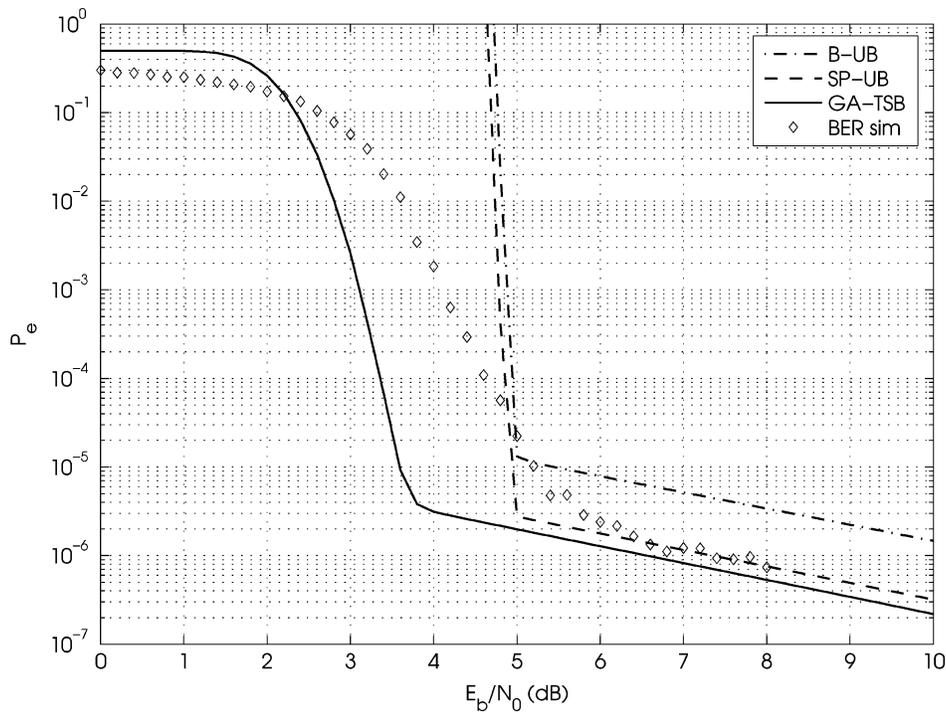


Fig. 7. Comparison of simulation and saddlepoint and Gaussian approximations on the bit error rate of BICM with a Repeat-and-Accumulate code of rate 1/4 with 16-QAM modulation and Gray mapping, $K = 512$ information bits, 20 iterations of belief propagation decoding the fully-interleaved Rayleigh fading channel.

is new since, as opposed to the usual formulas, it is uniformly valid for all values of the saddlepoint. The proposed method benefits from simple numerical integration using Gaussian quadratures for noise and fading averaging. We have verified the validity of the approximation for both, convolutional and turbo-like code ensembles with BICM, over

AWGN and fully interleaved Rayleigh-fading channels. In both cases, the asymptotic behavior of BICM mimics that of binary modulation. This simple technique constitutes a powerful tool to the analysis of finite-length BICM. Furthermore, being simpler and tighter than the original bounds in [2], it shows a wide range of practical applications.

APPENDIX I
DERIVATION OF THE SADDLEPOINT APPROXIMATION

We wish to estimate the tail probability of Z , a continuous random variable with density $f_Z(z)$. To Z we associate its cumulant transform (or cumulant generating function) $\kappa(s)$, defined as $\kappa(s) = \log E[e^{sZ}]$, $s \in \mathbb{C}$. We shall be concerned with the case of Z being the sum of M random variables X_i , $Z = \sum_{i=1}^M X_i$. For independent X_i , it is immediate that the total cumulant transform is the sum of the transforms for each component. In this case, the density of Z and its tail probability (or equivalently its distribution) can be recovered from $\kappa(s)$ by Fourier inversion [6]

$$f_Z(z) = \frac{1}{2\pi j} \int_{s=-j\infty}^{j\infty} e^{\kappa(s)-sz} ds \quad (18)$$

$$\Pr(Z > z) = \frac{1}{2\pi j} \int_{s=-j\infty}^{j\infty} \frac{e^{\kappa(s)-sz}}{s} ds. \quad (19)$$

In the following, we study the tail probability only and assume, without loss of generality, that $z > E[Z]$.

An application of Cauchy's integral theorem allows us to move the integration path to the right, from the imaginary axis to a line $\mathcal{L} = (\hat{s} - j\infty, \hat{s} + j\infty)$ that crosses the real axis at another point \hat{s} [13]. It is most convenient to choose \hat{s} so that $\kappa'(\hat{s}) = z$; this point is called a saddlepoint, as complex-variable analytic functions do not reach extreme points in their domain of analyticity [13]. This point exists and is unique due to the convexity of $\kappa(s)$ [6].

Along the integration path $s = \hat{s} + j\sigma$, $-\infty < \sigma < \infty$ and $(s - \hat{s}) = j\sigma$. Using this new variable of integration, we now expand the argument of the exponential term in a Taylor series around \hat{s}

$$\kappa(s) - sz = \kappa(\hat{s}) - \hat{s}z + \frac{\kappa''(\hat{s})}{2!}(j\sigma)^2 + R_2(\sigma) \quad (20)$$

where we have used that the first derivative is zero and $R_2(\sigma)$ is a shorthand for the remaining terms in the expansion around \hat{s}

$$R_2(\sigma) = \sum_{\ell=3}^{\infty} \frac{\kappa^{(\ell)}(\hat{s})}{\ell!} (j\sigma)^\ell. \quad (21)$$

In the following, we shall indistinctly refer to the ℓ th-order derivative as the ℓ th-order cumulant. Equation (19) can be rewritten as

$$\begin{aligned} \Pr(Z > z) &= \frac{1}{2\pi} e^{\kappa(\hat{s}) - \hat{s}z} \int_{-\infty}^{+\infty} e^{-\frac{\kappa''(\hat{s})}{2}\sigma^2} e^{R_2(\sigma)} \frac{d\sigma}{\hat{s} + j\sigma} \\ &= \frac{1}{2\pi} e^{\kappa(\hat{s}) - \hat{s}z} \int_{-\infty}^{+\infty} e^{-\frac{\kappa''(\hat{s})}{2}\sigma^2} e^{R_2(\sigma)} \frac{\hat{s} - j\sigma}{\hat{s}^2 + \sigma^2} d\sigma \end{aligned} \quad (22)$$

where we have multiplied numerator and denominator times a factor $\hat{s} - j\sigma$.

Using the Taylor expansion for the exponential, $e^z = \sum_{m=0}^{\infty} \frac{1}{m!} z^m$, we have now

$$\begin{aligned} (\hat{s} - j\sigma) e^{R_2(\sigma)} &= (\hat{s} - j\sigma) \sum_{m=0}^{\infty} \frac{1}{m!} \left(\sum_{\ell=3}^{\infty} \frac{\kappa^{(\ell)}(\hat{s})}{\ell!} (j\sigma)^\ell \right)^m \\ &= \sum_{m=0}^{\infty} \tilde{\eta}_m (j\sigma)^m = \sum_{m=0}^{\infty} \eta_m \sigma^m \end{aligned}$$

where we have grouped the terms with common factor $(j\sigma)^m$, and called the corresponding coefficient $\tilde{\eta}_m$; similarly for η_m , which incorporates the power of j into the coefficient. The symmetry of the integrand (see (22)) implies that the integral of the terms with odd m is zero; we need thus consider only the even values of m . The first few terms are

$$\begin{aligned} \eta_0 &= \hat{s}, & \eta_2 &= 0 \\ \eta_4 &= -\frac{\kappa^{(3)}(\hat{s})}{3!} + \hat{s} \frac{\kappa^{(4)}(\hat{s})}{4!} \\ \eta_6 &= \frac{\kappa^{(5)}(\hat{s})}{5!} - \hat{s} \frac{\kappa^{(6)}(\hat{s})}{6!} - \hat{s} \frac{1}{2!} \left(\frac{\kappa^{(3)}(\hat{s})}{3!} \right)^2. \end{aligned}$$

At this point, we normalize the cumulants. As the cumulants are all linear terms in m , the number of random variables contributing to Z , we get rid of this dependence on m by dividing all cumulants by $\kappa''(\hat{s})$ and denote the normalized cumulant by $\tilde{\kappa}^{(\ell)}(\hat{s})$. The coefficients η_m become now a polynomial of $\kappa''(\hat{s})$

$$\begin{aligned} \eta_4 &= -\frac{\tilde{\kappa}^{(3)}(\hat{s})}{3!} \kappa''(\hat{s}) + \hat{s} \frac{\tilde{\kappa}^{(4)}(\hat{s})}{4!} \kappa''(\hat{s}) \\ \eta_6 &= \frac{\tilde{\kappa}^{(5)}(\hat{s})}{5!} \kappa''(\hat{s}) - \hat{s} \frac{\tilde{\kappa}^{(6)}(\hat{s})}{6!} \kappa''(\hat{s}) - \hat{s} \frac{1}{2!} \left(\frac{\tilde{\kappa}^{(3)}(\hat{s})}{3!} \right)^2 (\kappa''(\hat{s}))^2. \end{aligned}$$

The degree of the polynomials will prove useful when tracking the various terms in the final expansion.

The next problem is the evaluation of integrals of the form

$$I(m) = \frac{\eta_m}{2\pi} \int_{-\infty}^{+\infty} \exp\left(-\frac{\kappa''(\hat{s})}{2}\sigma^2\right) \frac{\sigma^m}{\hat{s}^2 + \sigma^2} d\sigma \quad (23)$$

where m is an even number. The value of this integral is given in (34), in Appendix II. Setting $\beta^2 = \frac{1}{2}\kappa''(\hat{s})$ and $\alpha = \hat{s}$ we obtain

$$I(m) = \eta_m \frac{1 \cdot 3 \cdots (m-1)}{\hat{s}^2 \sqrt{2\pi(\kappa''(\hat{s}))^{m+1}}} \{1 + O((\kappa''(\hat{s}))^{-1})\}. \quad (24)$$

In particular, for $m = 0$ and discarding the $O()$ term, we recover the classical saddlepoint approximation

$$\Pr(Z > z) \simeq \frac{1}{\sqrt{2\pi\kappa''(\hat{s})}\hat{s}} e^{\kappa(\hat{s}) - \hat{s}z}. \quad (25)$$

Note that even though this equation loses its validity for small \hat{s} , we may use the original (33) and show that the probability tends to 1/2 for $\hat{s} \rightarrow 0$

$$\begin{aligned} \frac{1}{2\pi} \eta_0 \frac{\pi}{\hat{s}} \operatorname{erfc}\left(\hat{s} \sqrt{\frac{\kappa''(\hat{s})}{2}}\right) \exp\left(\frac{1}{2} \hat{s}^2 \kappa''(\hat{s})\right) \\ = \frac{1}{2} \operatorname{erfc}\left(\hat{s} \sqrt{\frac{\kappa''(\hat{s})}{2}}\right) \exp\left(\frac{1}{2} \hat{s}^2 \kappa''(\hat{s})\right). \end{aligned} \quad (26)$$

This yields an approximation which is uniformly valid for all values of the saddlepoint [6].

If required, higher order terms may be obtained by extending the outlined procedure. The following term is given by $m = 4$, which gives an extra term with leading coefficient $(\kappa''(\hat{s}))^{-\frac{5}{2}}$. As η_4 is a degree-1 polynomial of $\kappa''(\hat{s})$, the term grows rather like $(\kappa''(\hat{s}))^{-\frac{3}{2}}$. A careful analysis of the remaining terms shows that there is only one additional term with the same factor, namely, the one corresponding to the squared

third cumulant in η_6 . Putting all this information back together, the second-order saddlepoint approximation is given by

$$\Pr(Z > z) \simeq \frac{1}{\sqrt{2\pi\kappa''(\hat{s})\hat{s}}} e^{\kappa(\hat{s})-\hat{s}x} \left\{ 1 + \frac{1}{\kappa''(\hat{s})} \left(-\frac{1}{\hat{s}^2} - \frac{\tilde{\kappa}^{(3)}(\hat{s})}{2\hat{s}\kappa''(\hat{s})} + \frac{\kappa^{(4)}(\hat{s})}{8\kappa''(\hat{s})} - \frac{15}{72} \left(\frac{\kappa^{(3)}(\hat{s})}{\kappa''(\hat{s})} \right)^2 \right) + O((\kappa''(\hat{s}))^{-2}) \right\}. \quad (27)$$

This additional term in the expansion also serves as an estimate of the error made by the approximation. In general, the first term of the expansion gives a very good approximation to the real tail probability, with no need of considering extra terms.

APPENDIX II

SOME INTEGRALS AND EXPANSIONS OF INTEREST

The error complementary function is defined as

$$\operatorname{erfc}(x) = \frac{2}{\pi} \int_x^\infty e^{-t^2} dt.$$

Its asymptotic series is derived by integration by parts [13] and gives

$$\begin{aligned} \operatorname{erfc}(x) &= \frac{e^{-x^2}}{x\sqrt{\pi}} \sum_{m=0}^{\infty} (-1)^m \frac{1 \cdot 3 \cdots (2m-1)}{2^m x^{2m}} \quad (28) \\ &= \frac{e^{-x^2}}{x\sqrt{\pi}} \left(1 - \frac{1}{2x^2} + \frac{1 \cdot 3}{2^2 x^4} - \frac{1 \cdot 3 \cdot 5}{2^3 x^6} + \cdots \right). \quad (29) \end{aligned}$$

Here the absolute error committed by truncating may be shown to be smaller than the first neglected term. For large values of the parameter, the approximation $\operatorname{erfc}(x) \simeq \frac{1}{\sqrt{\pi}x} e^{-x^2}$ is valid. More precisely, for values of x larger than 1, the relative error in approximating $\operatorname{erfc}(x) \exp(x^2)$ by $\frac{1}{\sqrt{\pi}x}$ is smaller than 25% (obtained by evaluation of the formula, not with an estimate of the error).

An integral that appears often in our calculations is the following [13]:

$$\int_{-\infty}^{+\infty} \exp(-\beta^2 x^2) \frac{1}{\alpha^2 + x^2} dx = \frac{\pi}{\alpha} \operatorname{erfc}(\alpha\beta) \exp(\alpha^2 \beta^2). \quad (30)$$

Note that we may easily apply the asymptotic expansion for $\operatorname{erfc}(x)$.

We will also evaluate integrals of the more general form

$$\int_{-\infty}^{+\infty} \exp(-\beta^2 x^2) \frac{x^{2n}}{\alpha^2 + x^2} dx \quad (31)$$

where n is an integer. Their value is calculated as follows. First expand the fraction in the integrand

$$\begin{aligned} \frac{x^{2n}}{\alpha^2 + x^2} &= \sum_{m=1}^n (-1)^{m-1} \alpha^{2(m-1)} x^{2(n-m)} + (-1)^n \frac{\alpha^{2n}}{\alpha^2 + x^2}, \\ &= \sum_{m=0}^{n-1} (-1)^{n-m-1} \alpha^{2(n-m-1)} x^{2m} + (-1)^n \frac{\alpha^{2n}}{\alpha^2 + x^2} \quad (32) \end{aligned}$$

and then integrate term by term. Each term is seen to be the 2ν th moment of a normal random variable with zero mean and variance

$(2\beta^2)^{-1}$, whose value is $1 \cdot 3 \cdots (2\nu - 1)(2\beta^2)^{-\nu}$ [4]. Combining these values back into (31) and after some algebraic manipulations we get

$$\begin{aligned} &\sum_{m=0}^{n-1} (-1)^{n-m-1} \alpha^{2(n-m-1)} \int_{-\infty}^{+\infty} \exp(-\beta^2 x^2) x^{2m} dx \\ &+ (-1)^n \alpha^{2n} \int_{-\infty}^{+\infty} \exp(-\beta^2 x^2) \frac{1}{\alpha^2 + x^2} dx \\ &= \sum_{m=0}^{n-1} (-1)^{n-m-1} \alpha^{2(n-m-1)} \frac{\sqrt{\pi}}{\beta} 1 \cdot 3 \cdots (2m-1) (2\beta^2)^{-m} \\ &+ (-1)^n \alpha^{2n} \frac{\pi}{\alpha} \operatorname{erfc}(\alpha\beta) \exp(\alpha^2 \beta^2) \quad (33) \end{aligned}$$

$$\begin{aligned} &= \sum_{m=0}^{n-1} (-1)^{n-m-1} \alpha^{2(n-m-1)} \frac{\sqrt{\pi}}{\beta} 1 \cdot 3 \cdots (2m-1) (2\beta^2)^{-m} \\ &+ (-1)^n \alpha^{2n} \frac{\pi}{\alpha} \frac{e^{-\alpha^2 \beta^2}}{\alpha \beta \sqrt{\pi}} \\ &\times \sum_{m=0}^{\infty} (-1)^m \frac{1 \cdot 3 \cdots (2m-1)}{2^m \alpha^{2m} \beta^{2m}} \exp(\alpha^2 \beta^2) \\ &= \sum_{m=0}^{n-1} (-1)^{n-m-1} \alpha^{2(n-m-1)} \frac{\sqrt{\pi}}{\beta} 1 \cdot 3 \cdots (2m-1) (2\beta^2)^{-m} \\ &+ \sum_{m=0}^{\infty} (-1)^{n+m} \alpha^{2(n-1)} \frac{\sqrt{\pi}}{\beta} \frac{1 \cdot 3 \cdots (2m-1)}{2^m \alpha^{2m} \beta^{2m}} \\ &= \sum_{m=n}^{\infty} (-1)^{n+m} \alpha^{2(n-m-1)} \frac{\sqrt{\pi}}{\beta} \frac{1 \cdot 3 \cdots (2m-1)}{2^m \beta^{2m}}. \quad (34) \end{aligned}$$

In the last step, we exploit that the first $n-1$ terms in both summations exactly cancel each other. As it is derived from the asymptotic expansion of $\operatorname{erfc}(x)$, the formula inherits the former's bound on error, that is, the error by truncating the series is upper-bounded by the absolute value of the following term.

APPENDIX III

CUMULANT TRANSFORM ASYMPTOTIC ANALYSIS

In this appendix, we show that in the limit for large SNR, BICM behaves as a binary modulation with squared Euclidean distance

$$d_{\min}^2 \triangleq \min_{x, x' \in \mathcal{X}} d^2(x, x') = \min_{x, x' \in \mathcal{X}} |x - x'|^2.$$

In particular, we have that

$$\lim_{\text{SNR} \rightarrow \infty} \frac{\kappa(\hat{s})}{\text{SNR}} = -\frac{d_{\min}^2}{4} \quad (35)$$

and

$$\lim_{\text{SNR} \rightarrow \infty} \frac{\kappa''(\hat{s})}{\text{SNR}} = 2d_{\min}^2 \quad (36)$$

for the AWGN channel, and

$$\lim_{\text{SNR} \rightarrow \infty} \frac{\kappa(\hat{s})}{\log \text{SNR}} = -1 \quad (37)$$

and

$$\lim_{\text{SNR} \rightarrow \infty} \kappa''(\hat{s}) = 8 \quad (38)$$

for the fully interleaved Rayleigh-fading channel. In this appendix, and without loss of generality, assume that s is real.

A. AWGN Channel

Consider first the AWGN channel without fading. Then

$$\lim_{\text{SNR} \rightarrow \infty} \frac{\kappa(s)}{\text{SNR}} = \lim_{\text{SNR} \rightarrow \infty} \frac{1}{\text{SNR}} \times \log \left(\mathbb{E}_{z,x,m} \left[\left(\frac{\sum_{x' \in \mathcal{X}_1^m} e^{-|\sqrt{\text{SNR}}(x-x') + z|^2}}{\sum_{x' \in \mathcal{X}_0^m} e^{-|\sqrt{\text{SNR}}(x-x') + z|^2}} \right)^s \right] \right).$$

We can upper-bound the term inside the expectation by upper-bounding the sum at the numerator by $\frac{|\mathcal{X}|}{2}$ and lower-bounding the sum at the denominator by $e^{-|z|^2}$. Then

$$\left(\frac{\sum_{x' \in \mathcal{X}_1^m} e^{-|\sqrt{\text{SNR}}(x-x') + z|^2}}{\sum_{x' \in \mathcal{X}_0^m} e^{-|\sqrt{\text{SNR}}(x-x') + z|^2}} \right)^s \leq \left(\frac{|\mathcal{X}|}{2} \right)^s e^{s|z|^2} \quad (39)$$

and since

$$\mathbb{E} \left[e^{s|z|^2} \right] < \infty;$$

if $s < 1$, we can use the dominated convergence theorem [4]. Note that since $\hat{s} = \frac{1}{2}$ for BIOS channels, this restriction poses no practical limitation to the validity of the result. We can then take the dominant terms in the sums and write that

$$\lim_{\text{SNR} \rightarrow \infty} \frac{\kappa(s)}{\text{SNR}} = \lim_{\text{SNR} \rightarrow \infty} \frac{1}{\text{SNR}} \log \left(\mathbb{E}_{z,x,m} \left[\left(\frac{e^{-|\sqrt{\text{SNR}}(x-x') + z|^2}}{e^{-|z|^2}} \right)^s \right] \right)$$

where x' denotes the signal constellation symbol closest to x in the complementary set \mathcal{X}_1^m .

Then we have that

$$\begin{aligned} & \lim_{\text{SNR} \rightarrow \infty} \frac{\kappa(s)}{\text{SNR}} \\ &= \lim_{\text{SNR} \rightarrow \infty} \frac{1}{\text{SNR}} \\ & \quad \times \log \left(\mathbb{E}_{z,x,m} \left[e^{-s|\sqrt{\text{SNR}}(x-x') + z|^2 + s|z|^2} \right] \right) \\ &= \lim_{\text{SNR} \rightarrow \infty} \frac{1}{\text{SNR}} \\ & \quad \times \log \left(\mathbb{E}_{z,x,m} \left[e^{-s\text{SNR}d^2(x,x') - 2s\text{Re}\{ \sqrt{\text{SNR}}(x-x')z^* \}} \right] \right) \\ &= \lim_{\text{SNR} \rightarrow \infty} \frac{1}{\text{SNR}} \log \left(\mathbb{E}_{x,m} \left[e^{-\text{SNR}d^2(x,x')(s-s^2)} \right] \right) \\ &= \lim_{\text{SNR} \rightarrow \infty} \frac{1}{\text{SNR}} \log \left(\mathcal{K} e^{-\text{SNR}d_{\min}^2(s-s^2)} \right) \\ &= -d_{\min}^2(s-s^2) \end{aligned}$$

where \mathcal{K} may depend on the actual mapping rule. Note, however, that the result does not. By letting $s = \hat{s} = \frac{1}{2}$, we then obtain that

$$\lim_{\text{SNR} \rightarrow \infty} \frac{\kappa(\hat{s})}{\text{SNR}} = -\frac{d_{\min}^2}{4}.$$

Furthermore, at large SNR, the second-order cumulant behaves as

$$\lim_{\text{SNR} \rightarrow \infty} \frac{\kappa''(s)}{\text{SNR}} = 2d_{\min}^2$$

which again mimics the behavior of a binary modulation with squared minimum distance d_{\min}^2 .

B. Fully Interleaved Rayleigh-Fading Channel

In the case of the fully interleaved fading channel, we have

$$\lim_{\text{SNR} \rightarrow \infty} \frac{\kappa(s)}{\log \text{SNR}} = \lim_{\text{SNR} \rightarrow \infty} \frac{1}{\log \text{SNR}} \times \log \left(\mathbb{E}_{z,h,x,m} \left[\left(\frac{\sum_{x' \in \mathcal{X}_1^m} e^{-|\sqrt{\text{SNR}}h(x-x') + z|^2}}{\sum_{x' \in \mathcal{X}_0^m} e^{-|\sqrt{\text{SNR}}h(x-x') + z|^2}} \right)^s \right] \right).$$

The upper bound in (39) applies here as well and, then, for $s < 1$, the dominated convergence theorem leads to

$$\begin{aligned} & \lim_{\text{SNR} \rightarrow \infty} \frac{\kappa(s)}{\log \text{SNR}} \\ &= \lim_{\text{SNR} \rightarrow \infty} \frac{1}{\log \text{SNR}} \log \left(\mathbb{E}_{z,h,x,m} \left[\left(\frac{e^{-|\sqrt{\text{SNR}}h(x-x') + z|^2}}{e^{-|z|^2}} \right)^s \right] \right) \\ &= \lim_{\text{SNR} \rightarrow \infty} \frac{1}{\log \text{SNR}} \log \left(\mathbb{E}_{\gamma,x,m} \left[e^{-\text{SNR}\gamma d^2(x,x')(s-s^2)} \right] \right) \end{aligned}$$

where $\gamma \triangleq |h|^2$ is the fading power and x' denotes again the closest point to x in the set \mathcal{X}_1^m . Averaging over the fading we get that

$$\begin{aligned} & \lim_{\text{SNR} \rightarrow \infty} \frac{\kappa(s)}{\log \text{SNR}} \\ &= \lim_{\text{SNR} \rightarrow \infty} \frac{1}{\log \text{SNR}} \log \left(\mathbb{E}_{x,m} \left[\frac{1}{1 + \text{SNR}d^2(x,x')(s-s^2)} \right] \right) \\ &= \lim_{\text{SNR} \rightarrow \infty} \frac{1}{\log \text{SNR}} \log \left(\frac{\mathcal{K}}{1 + \text{SNR}d_{\min}^2(s-s^2)} \right) \\ &= -1. \end{aligned}$$

Therefore, since at large SNR

$$\lim_{\text{SNR} \rightarrow \infty} \kappa(s) = \lim_{\text{SNR} \rightarrow \infty} \log \left(\frac{\mathcal{K}}{1 + \text{SNR}d_{\min}^2(s-s^2)} \right)$$

we obtain that the second-order cumulant behaves as

$$\begin{aligned} & \lim_{\text{SNR} \rightarrow \infty} \kappa''(s) \\ &= \lim_{\text{SNR} \rightarrow \infty} \frac{2\text{SNR}d_{\min}^2}{1 + \text{SNR}d_{\min}^2(s-s^2)} + \left(\frac{\text{SNR}d_{\min}^2(2s-1)}{1 + \text{SNR}d_{\min}^2(s-s^2)} \right)^2. \end{aligned}$$

By letting $s = \hat{s} = \frac{1}{2}$, it is easy to verify that at the saddlepoint

$$\lim_{\text{SNR} \rightarrow \infty} \kappa''(\hat{s}) = 8 \quad (40)$$

as in the binary case.

REFERENCES

- [1] E. Zehavi, "8-PSK trellis codes for a rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 873-884, May 1992.
- [2] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927-946, May 1998.
- [3] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.
- [4] R. Durrett, *Probability: Theory and Examples*. Belmont, CA: Duxbury, 1996.
- [5] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. New York: Dover, 1972.
- [6] J. L. Jensen, *Saddlepoint Approximations*. Oxford, U.K.: Clarendon, 1995.
- [7] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [8] A. Guillén i Fàbregas, A. Martínez, and G. Caire, "Error probability of bit-interleaved coded modulation using the gaussian approximation," in *Proc. Conf. Information Science and Systems*, Princeton, NJ, Mar. 2004.
- [9] R. Lugannani and S. O. Rice, "Saddle point approximation for the distribution of the sum of independent random variables," *Adv. Appl. Probab.*, vol. 12, pp. 475-490, 1980.

- [10] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "New simple evaluation of the error probability of bit-interleaved coded modulation using the saddlepoint approximation," in *Proc. 2004 Int. Symp. Information Theory and its Applications*, Parma, Italy, Oct. 2004, pp. 1491–1496.
- [11] G. Poltyrev, "Bounds on the decoding error probability of linear codes via their spectra," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1284–1292, Jul. 1994.
- [12] D. Divsalar, H. Jin, and R. J. McEliece, "Coding theorems for "turbo-like" codes," in *Proc. 36th Allerton Conf. Communication, Control, and Computing*, Allerton House, Monticello, IL, Sep. 1998, pp. 201–210.
- [13] F. W. J. Olver, *Asymptotics and Special Functions*. New York: Academic, 1974.

On the Distribution of SINR for the MMSE MIMO Receiver and Performance Analysis

Ping Li, Debashis Paul, Ravi Narasimhan, *Member, IEEE*, and John Cioffi, *Fellow, IEEE*

Abstract—This correspondence studies the statistical distribution of the signal-to-interference-plus-noise ratio (SINR) for the minimum mean-square error (MMSE) receiver in multiple-input multiple-output (MIMO) wireless communications. The channel model is assumed to be (transmit) correlated Rayleigh flat-fading with unequal powers. The SINR can be decomposed into two independent random variables: $\text{SINR} = \text{SINR}^{\text{ZF}} + T$, where SINR^{ZF} corresponds to the SINR for a zero-forcing (ZF) receiver and has an exact Gamma distribution. This correspondence focuses on characterizing the statistical properties of T using the results from random matrix theory. First three asymptotic moments of T are derived for uncorrelated channels and channels with equicorrelations. For general correlated channels, some limiting upper bounds for the first three moments are also provided. For uncorrelated channels and correlated channels satisfying certain conditions, it is proved that T converges to a Normal random variable. A Gamma distribution and a generalized Gamma distribution are proposed as approximations to the finite sample distribution of T . Simulations suggest that these approximate distributions can be used to estimate accurately the probability of errors even for very small dimensions (e.g., two transmit antennas).

Index Terms—Asymptotic distributions, channel correlation, error probability, Gamma approximation, minimum mean square error (MMSE) receiver, multiple-input multiple-output (MIMO) system, random matrix, signal-to-interference-plus-noise ratio (SINR).

I. INTRODUCTION

This study considers the following signal and channel model in a multiple-input multiple-output (MIMO) system:

$$y_r = \frac{1}{\sqrt{m}} \mathbf{H} \mathbf{W} \mathbf{R}_t^{\frac{1}{2}} \mathbf{P}^{\frac{1}{2}} x_t + n_c = \frac{1}{\sqrt{m}} \mathbf{H} x_t + n_c \quad (1)$$

where $x_t \in \mathbb{C}^p$ is the (normalized) transmitted signal vector and $y_r \in \mathbb{C}^m$ is the received signal vector. Here p is the number of transmit antennas and m is the number of receive antennas. $\mathbf{H} \mathbf{W} \in \mathbb{C}^{m \times p}$ consists of independent and identically distributed (i.i.d.) standard complex Normal entries. $\mathbf{R}_t \in \mathbb{C}^{p \times p}$ is the transmitter correlation matrix. $\mathbf{P} = \text{diag}[\tilde{c}_1, \tilde{c}_2, \dots, \tilde{c}_p] \in \mathbb{R}^{p \times p}$; $\tilde{c}_k = c_k \frac{m}{p}$, where c_k is the signal-to-noise ratio (SNR) for the k th spatial stream. This definition of SNR is consistent with [1, Sec. 7.4]. $\mathbf{H} = \mathbf{H} \mathbf{W} \mathbf{R}_t^{\frac{1}{2}} \mathbf{P}^{\frac{1}{2}} \in \mathbb{C}^{m \times p}$ is treated as the channel matrix. $n_c \in \mathbb{C}^m$ is the complex noise vector and is assumed to have zero mean and identity covariance. Note that the power matrix \mathbf{P} has terms involving the variance of the noise. The correlation matrix \mathbf{R}_t and power matrix \mathbf{P} are assumed to be nonrandom. Also, we restrict our attention to $p \leq m$.

We consider the popular linear minimum mean-square error (MMSE) receiver. Conditional on the channel matrix \mathbf{H} , the signal-to-interference-plus-noise ratio (SINR) on the k th spatial stream can be expressed as (e.g., [1]–[6])

$$\text{SINR}_k = \frac{1}{\text{MMSE}_k} - 1 = \frac{1}{\left[(\mathbf{I}_p + \frac{1}{m} \mathbf{H}^\dagger \mathbf{H})^{-1} \right]_{kk}} - 1 \quad (2)$$

where \mathbf{I}_p is a $p \times p$ identity matrix, and \mathbf{H}^\dagger is the Hermitian transpose of \mathbf{H} . Note that (2), in the same form as equation (7.49) of [1], is derived based on the second-order statistics of the input signals, not restricted to binary signals.

For binary inputs, Verdú [4, eq. (6.47)] provides the exact formula for computing the bit-error rate (BER) (also see [7]). Conditional on \mathbf{H} , this BER formula requires computing 2^{p-1} Q-functions. To compute BER unconditionally, we need to sample \mathbf{H} enough times (e.g., 10^5) to get a reliable estimate. When $p \geq 32$ (or $p \geq 64$), the computations become intractable [4], [8].

Recently, study of the asymptotic properties of multiuser receivers (e.g., [2]–[4], [6], [8]–[11]) has received a lot of attention. Works that relate directly to the content of this correspondence include Tse and Hanly [11] and Verdú and Shamai [6], who independently derived the asymptotic first moment of SINR for uncorrelated channels. Tse and Zeitouni [3] proved the asymptotic Normality of SINR for the equal power case, and commented on the possibility of extending the result to the unequal powers scenario. Zhang *et al.* [12] proved the asymptotic Normality of the multiple-access interference (MAI), which is closely related to SINR. Guo *et al.* [8] proved the asymptotic Normality of the decision statistics for a variety of linear multiuser receivers. [8] considered a general power distribution and corresponding unconditional asymptotic behavior.

Based on the asymptotic Normality results, Poor and Verdú [2] (also in [4], [8]) proposed using the limiting BER (denoted by BER_∞) for binary modulations, which is a single Q-function

$$\text{BER}_\infty = Q(\sqrt{E(\text{SINR}_k)_\infty}) = \int_{\sqrt{E(\text{SINR}_k)_\infty}}^{\infty} e^{-t^2/2} dt \quad (3)$$

where $E(\text{SINR}_k)_\infty$ denotes the asymptotic first moment of SINR_k .

Equation (3) is convenient and accurate for large dimensions. However, its accuracy for small dimensions is of some concern. For instance, [8] compared the asymptotic BER with simulation results, which showed that even with $p = 64$ there existed significant discrepancies. In general, (3) will underestimate the true BER. For example, in our simulations, when $m = 16$, $p = 8$, $\text{SNR} = 15$ dB, the asymptotic BER given by (3) is roughly $\frac{1}{10000}$ of the exact BER. In current practice, code-division multiple-access (CDMA) channels with m, p between 32 and 64 are typical and in multiple-antenna systems arrays of 4 antennas are typical but arrays with 8 to 16 antennas would be feasible in the near future [9]. Therefore, it would be useful if

Manuscript received January 20, 2005; revised September 9, 2005.
 P. Li is with the Department of Statistics, Stanford University, Stanford, CA 94305 USA (e-mail: pingli@stat.stanford.edu).
 D. Paul is with the Department of Statistics, University of California, Davis, Davis, CA 95616 USA (e-mail: debashis@wald.ucdavis.edu).
 R. Narasimhan is with the Department of Electrical Engineering, University of California, Santa Cruz, Santa Cruz, CA 95064 USA (e-mail: ravi@soe.ucsc.edu).
 J. Cioffi is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: cioffi@stanford.edu).
 Communicated by R. R. Müller, Associate Editor for Communications.
 Digital Object Identifier 10.1109/TIT.2005.860466