Large-SNR Error Probability Analysis of BICM with Uniform Interleaving in Fading Channels

Alfonso Martinez and Albert Guillén i Fàbregas

Abstract

This paper studies the average error probability of bit-interleaved coded modulation with uniform interleaving in fully-interleaved fading channels. At large signal-to-noise ratio, the dominant pairwise error events are mapped into symbols with Hamming weight larger than one, causing a flattening of the error probability. Closed-form expressions for the error probability with general modulations are provided. For interleavers of practical length, the flattening is noticeable only at very low values of the error probability.

I. Motivation and Summary

Whilst QPSK is equivalent to two parallel independent BPSK channels in the Gaussian channel, the equivalence fails in fading channels because of the statistical dependence between the quadrature components introduced by the fading coefficients. This dependence ensures that the error probability is dominated by the number of QPSK symbols with Hamming weight two. The asymptotic slope of the error probability is reduced and an error floor results, a phenomenon similar in nature to the floor appearing in turbo codes [1], [2].

More generally, bit-interleaved coded modulation (BICM) [3] is affected by the same phenomenon. For low signal-to-noise ratio (SNR), the error probability is determined by error events with a high diversity, so that the minimum Hamming weight of the code is distributed across the largest possible number of modulation symbols. Although "worse" error events with smaller diversity are present, they are weighted by a low error probability, and thus remain hidden for most practical purposes. We analyze this behaviour by first studying the union bound to the error probability for QPSK modulation with Gray labeling and fully-interleaved fading, and then extending the results to general constellations. The saddlepoint approximation allows us to derive closed-form expressions for the pairwise error probability, which highlight the aforementioned floor effect. Finally, we estimate of the threshold SNR at which the error probability changes slope.

II. Channel Model

At the transmitter, a linear binary code of rate R_c is used to generate codewords $\boldsymbol{c} = (c_1, c_2, \ldots, c_\ell)$ of length ℓ, which are interleaved before modulation. Then, consecutive groups of m interleaved bits $(c_{m(k-1)+1}, \ldots, c_{mk})$, for $k = 1, \ldots, n$, are mapped onto a modulation symbol x_k using some mapping rule, such as binary reflected Gray labeling. We denote the modulation signal set by $\mathcal{X}$, its cardinality by $|\mathcal{X}|$, and the number of bits per symbol $m = \log_2 |\mathcal{X}|$. The signal set is normalized to unit average energy. Subindices r and i respectively

A. Martinez is with Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands, e-mail: alfonso.martinez@ieee.org. A. Guillén i Fàbregas is with the Department of Engineering, University of Cambridge, Cambridge, UK, e-mail: guillen@ieee.org.
This work was supported by the International Incoming Short Visits Scheme 2007/R2 of the Royal Society.

denote real and imaginary parts. We also assume that $n \triangleq \frac{\ell}{m}$ is an integer. We denote the inverse mapping function for labeling position j as $c_j : \mathcal{X} \to \{0,1\}$, namely $c_j(x)$ is the j-th bit of symbol x. The sets $\mathcal{X}_{c_1,\ldots,c_v}^{j_1,\ldots,j_v}$ contain the symbols with bit labels in positions $j_1,\ldots,j_v$ equal to $c_1,\ldots,c_v$.

For each input symbol the corresponding channel output y_k is given by

$$y_k = h_k\sqrt{\mathrm{SNR}}x_k + z_k, \quad k = 1,\ldots,n \tag{1}$$

where SNR is the average received signal-to-noise ratio, z_k are independent samples of circularly-symmetric Gaussian noise of variance 1, and h_k are i. i. d. fading coefficients. We assume that the fading coefficient is known at the receiver, which implies that the phase of h_k is irrelevant thanks to the circular symmetry of the noise. We also assume that the fading coefficients h_k are drawn from a Nakagami distribution of parameter $m_f > 0$ [4]. This distribution encompasses Rayleigh and AWGN channels, and approximates Rician fading.

III. Union Bound and Average Pairwise Error Probability for BICM

At the receiver, we consider the maximum-metric BICM decoder which selects the codeword with largest metric $\prod_{k=1}^{n} q(x_k, y_k)$ [3]. The real-valued metric function $q(x_k, y_k)$ is computed by the demodulator for each symbol according to the formula

$$q(x_k, y_k) = \prod_{j=1}^{m} q_j\big(c_j(x_k), y_k\big), \tag{2}$$

where the bit metric $q_j\big(c_j(x) = c, y\big)$ is given by

$$q_j\big(c_j(x) = c, y\big) = \sum_{x' \in \mathcal{X}_c^j} p(y|x', h), \tag{3}$$

where $p(y|x,h) = \frac{1}{\pi}e^{-|y-h\sqrt{\mathrm{SNR}}x|^2}$ is the channel transition probability density.

The word error rate, denoted by P_e, is the probability of selecting at the decoder a codeword different from the transmitted one. Similarly, the bit error rate P_b is the average number of input bits in error out of the possible ℓR_c corresponding to a codeword of length ℓ. Exact expressions for P_e or P_b are difficult to obtain and one often resorts to bounding, such as the union bound [4]. In the union bound, the probability of an error event is bounded by the sum of the probabilities of all possible pairwise error events, where a codeword $\boldsymbol{c}'$ other than the transmitted $\boldsymbol{c}$ has a larger metric. We define the *pairwise score* as

$$\xi_{\mathrm{pw}} \triangleq \sum_{k=1}^{n}\sum_{j=1}^{m} \log \frac{q_j(c'_{m(k-1)+j}, y_k)}{q_j(c_{m(k-1)+j}, y_k)}. \tag{4}$$

The pairwise error probability $\mathrm{PEP}(\boldsymbol{c}', \boldsymbol{c})$ between a reference codeword $\boldsymbol{c}$ and the competitor codeword $\boldsymbol{c}'$ is given by $\mathrm{PEP}(\boldsymbol{c}', \boldsymbol{c}) = \Pr\{\xi_{\mathrm{pw}} > 0\}$. By construction, the pairwise score is given by the sum of n symbol scores,

$$\xi_k^{\mathrm{s}} \triangleq \sum_{j=1}^{m} \log \frac{q_j(c'_{m(k-1)+j}, y_k)}{q_j(c_{m(k-1)+j}, y_k)}. \tag{5}$$

If the codewords have Hamming distance d, at most d symbol scores are non zero. Further, each symbol score is in turn given by the sum of m bit scores

$$\xi_{k,j}^{\mathrm{b}} \triangleq \log \frac{q_j(c'_{m(k-1)+j}, y_k)}{q_j(c_{m(k-1)+j}, y_k)}. \tag{6}$$

Clearly, we only need to consider the non-zero bit and symbol scores. These scores are random variables whose density function depends on all the random elements in the channel, as well as the transmitted bits, their position in the symbol and the bit pattern. In order to avoid this dependence, and as done by Yeh *et al.* [5], a uniform interleaver [2] is added between the binary code and the mapper, so that the pairwise error probability is averaged over all possible ways of placing the d bits in the n modulation symbols. We distinguish these alternative placements by counting the number of symbols with weight v, where $0 \leq v \leq w^\star$ and $w^\star = \min(m, d)$. Denoting the number of symbols of weight v by n_v, the symbol pattern ρ_n is given by $\rho_n = (n_0, \ldots, n_{w^\star})$. We also have that $\sum_{v=0}^{w^\star} n_v = n$ and $d = \sum_{v=1}^{w^\star} v n_v$.

For finite-length interleaving, the conditional pairwise error probability $\mathrm{PEP}(d, \rho_n)$ varies for every possible pattern. As done by Yeh *et al.* [5], averaging over all possible ways of choosing d locations in a codeword we have the following union bounds [1]

$$\bar{P}_e \leq \sum_d A_d \sum_{\rho_n} P(\rho_n)\, \mathrm{PEP}(d, \rho_n), \qquad \bar{P}_b \leq \sum_d A'_d \sum_{\rho_n} P(\rho_n)\, \mathrm{PEP}(d, \rho_n) \tag{7}$$

where A_d is the number of binary codewords of Hamming weight d, $A'_d = \sum_j \frac{j}{R_c \ell} A_{j,d}$, with $A_{j,d}$ being the number of codewords of Hamming weight d generated with an input message of weight j, and $P(\rho_n)$ is the probability of a particular pattern ρ_n. A counting argument [5] gives the probability of the pattern ρ_n, $P(\rho_n) \triangleq \Pr\big(\rho_n = (n_0, \ldots, n_{w^\star})\big)$, as

$$P(\rho_n) = \frac{\binom{m}{1}^{n_1} \binom{m}{2}^{n_2} \cdots \binom{m}{w^\star}^{n_{w^\star}}}{\binom{mn}{d}} \frac{n!}{n_0! n_1! n_2! \ldots n_{w^\star}!} \tag{8}$$

We remove the dependence of the pairwise error probability on the specific choice of modulation symbols by averaging over all possible such choices. This method consists of adding to every transmitted codeword $\boldsymbol{c} \in \mathcal{C}$ a random binary word $\boldsymbol{d} \in \{0,1\}^n$ known by the receiver. This is equivalent to scrambling the output of the encoder by a sequence known at the receiver. Scrambling guarantees the symbols corresponding to two m-bit sequences $(c_1, \ldots, c_m)$ and $(c'_1, \ldots, c'_m)$ are mapped to all possible pairs of modulation symbols differing in a given Hamming weight, hence making the channel symmetric. In [3], [5], the scrambler role was played by a random choice between a mapping rule μ and its complement $\bar{\mu}$ with probability 1/2 at every channel use. Scrambling is the natural extension of this random choice to weights larger than 1.

The cumulant transform [6] is an equivalent representation of the probability distribution of a random variable; the distribution can be recovered by an inverse Fourier transform. Consider a non-zero symbol score Ξ^{s} of Hamming weight $1 \leq v \leq m$, i.e., the Hamming weight between the binary labels of the reference and competitor symbols is v. The cumulant transform of Ξ^{s} is

$$\kappa_v(s) \triangleq \log \mathrm{E}\left[e^{s\,\Xi^{\mathrm{s}}}\right] = \log\left(\frac{1}{\binom{m}{v}} \sum_{\boldsymbol{j}=(j_1,\ldots,j_v)} \frac{1}{2^v} \sum_{\boldsymbol{c}\in\{0,1\}^v} \mathrm{E}\left[\frac{\prod_{i=1}^v q_{j_i}(\bar{c}_{j_i}, Y)^s}{\prod_{i=1}^v q_{j_i}(c_{j_i}, Y)^s}\right]\right), \tag{9}$$

where $\boldsymbol{j} = (j_1, \ldots, j_v)$ is a sequence of v bit indices, the bit $\bar{c}$ is the binary complement of c, and y are the channel outputs with bit v-tuple $\boldsymbol{c}$ transmitted at positions in $\boldsymbol{j}$. The cumulant transform of the pairwise score

[1] If a more detailed code spectrum were known, namely A_{d,ρ_n} and A_{j,d,ρ_n}, respectively denoting the number of codewords with Hamming weight d mapped onto the pattern ρ_n and the number of such codewords with input weight j, similar expressions to those in Eq. (7) could be written without the averaging operation.

$\Xi_{\rm pw}(\rho_n)$ is given by

$$\kappa_{\rm pw}(s,\rho_n) \triangleq \log \mathrm{E}[e^{s\Xi_{\rm pw}(\rho_n)}] = \sum_{v=1}^{w^\star} n_v \kappa_v(s). \tag{10}$$

The expectation in (9) is done according to $p_{\boldsymbol{j}}(y|\boldsymbol{c}) = \frac{1}{2^{m-v}} \sum_{x \in \mathcal{X}_{c_1,\ldots,c_v}^{j_1,\ldots,j_v}} p(y|x,h)$.

An important of the cumulant transform is that the tail probability is to great extent determined by the cumulant transform around the *saddlepoint* $\hat{s}$, defined as the value of s for which $\kappa'(\hat{s}) \triangleq \frac{d\kappa(s)}{ds} = 0$. Indeed, in our notation the Chernoff bound is given by $\mathrm{PEP}(d,\rho_n) \leq e^{\kappa_{\rm pw}(\hat{s},\rho_n)}$. Then the saddlepoint approximation to $\mathrm{PEP}(d,\rho_n)$ is given by

$$\mathrm{PEP}(d,\rho_n) \simeq \frac{1}{\hat{s}\sqrt{2\pi\kappa''_{\rm pw}(\hat{s},\rho_n)}} e^{\kappa_{\rm pw}(\hat{s},\rho_n)}, \tag{11}$$

where the saddlepoint $\hat{s}$ is the root of the equation $\kappa'_{\rm pw}(s,\rho_n) = 0$.

A particularly important case arises when $v = 1$, i.e., when the binary labels of the reference and competitor symbols in the symbol pairwise score differ only by a single bit and all d different bits are mapped onto different constellation symbols. This is the case for interleavers of practical length. As noticed in [3], [7], this simplifies significantly the analysis. The cumulant transform of the symbol score with $v = 1$, denoted by $\Xi_1^{\rm b}$, is given by

$$\kappa_1(s) \triangleq \log \mathrm{E}\big[e^{s\,\Xi_1^{\rm b}}\big] = \log\left(\frac{1}{m}\sum_{j=1}^{m}\frac{1}{2}\sum_{c\in\{0,1\}} \mathrm{E}\left[\frac{q_j(\bar{c},Y)^s}{q_j(c,Y)^s}\right].\right), \tag{12}$$

We denote the corresponding pairwise error probability by $\mathrm{PEP}_1(d)$. As noticed in [7], This expression is related, but not identical to, the ones appearing in [3], [5]. These authors consider an approximation to the bit decoding metric $q_j(c,y)$ whereby only one of the summands in Eq. (3) is kept. Even though this approximation slightly simplifies the analysis, it leads to large inaccuracies for mappings other than Gray mapping. In this paper, we consider the full bit metric.

As shown in Appendix A, the probability that all bit scores are independent approaches 1 as $1-\frac{d(d-1)}{2\ell}(m-1)$ for large interleaver lengths. In general, the effect of the dependence across the bits belonging to the same symbol must be taken into account.

IV. Inequivalence Between BPSK and QPSK with Gray Labeling

Obtaining closed-form expressions for the cumulant transforms can be difficult, and numerical methods are needed. Notable exceptions are BPSK and QPSK with Gray labeling, where the pattern ρ_n is uniquely determined by the number of QPSK symbols whose bits are at Hamming distance 1 and 2, respectively denoted by n_1 and n_2. Then $n_1 + 2n_2 = d$. It is also clear that $\max\{0, d-n\} \leq n_2 \leq \lfloor \frac{d}{2} \rfloor$. We denote the pairwise error probability by $\mathrm{PEP}(d,n_2)$. Reference [8] gave the union bounds to the average error probability for block-fading channels, of which QPSK can be seen as a particular case. In any case, having $n_2 \geq 0$ implies that we have n_2 symbols with Hamming weight two that fade with the same fading coefficient, and therefore, QPSK behaves differently from two independent BPSK channels.

A non-zero symbol score $\xi^{\rm s}$ is the result of having a Hamming distance of either 1 or 2 bits. In the first case, it is easy to see that it has a Gaussian distribution of mean $-2\mathrm{SNR}|h_k|^2$ and variance $4\mathrm{SNR}|h_k|^2$. This score is equal to that of BPSK modulation with effective signal-to-noise ratio $\frac{1}{2}\mathrm{SNR}|h_k|^2$. When both bits are

different, the score is a real-valued random variable with a Gaussian distribution of mean $-4\mathrm{SNR}|h_k|^2$ and variance $8\mathrm{SNR}|h_k|^2$. This score coincides with that of a BPSK modulation with signal-to-noise ratio $\mathrm{SNR}|h_k|^2$. Using the expression for $\kappa_1(s)$ from [7], we conclude that

$$\kappa_{\mathrm{pw}}(s,n_2) = \log\Big(\mathrm{E}\left[e^{s\Xi_1}\right]^{d-2n_2}\mathrm{E}\left[e^{s\Xi_2}\right]^{n_2}\Big) \tag{13}$$

$$= \log\left(\left(1+\frac{2s\mathrm{SNR}}{m_f}-\frac{2s^2\mathrm{SNR}}{m_f}\right)^{-m_f(d-2n_2)}\left(1+\frac{4s\mathrm{SNR}}{m_f}-\frac{4s^2\mathrm{SNR}}{m_f}\right)^{-m_f n_2}\right). \tag{14}$$

Direct calculation shows that the saddlepoint is located at $\hat{s}=+\frac{1}{2}$, and we finally approximate the conditional pairwise error probability $\mathrm{PEP}(d,n_2)$ by the saddlepoint approximation

$$\mathrm{PEP}(d,n_2) \simeq \frac{1}{2\sqrt{\pi m_f \mathrm{SNR}\frac{dm_f+d\mathrm{SNR}-n_2\mathrm{SNR}}{(2m_f+\mathrm{SNR})(m_f+\mathrm{SNR})}}}\left(1+\frac{\mathrm{SNR}}{2m_f}\right)^{-m_f(d-2n_2)}\left(1+\frac{\mathrm{SNR}}{m_f}\right)^{-m_f n_2}. \tag{15}$$

In the limit $m_f\to\infty$, i.e., for the AWGN channel, this formula respectively becomes

$$\mathrm{PEP}(d,n_2) \simeq \frac{1}{\sqrt{2\pi d\mathrm{SNR}}}e^{-\frac{1}{2}d\mathrm{SNR}}, \tag{16}$$

which is independent of n_2, as expected. The error performance approaches that of two codewords at distance d transmitted over BPSK with signal-to-noise ratio $\frac{1}{2}\mathrm{SNR}$.

Zummo *et al.* gave a similar analysis in their study of block-fading channels [8] and of BICM [5]. Our use of the saddlepoint approximation gives a simple and tight closed-form approximation to the pairwise error probability. Figure 1 depicts the bit-error probability of QPSK with the $(5,7)_8$ convolutional code for $\ell=40,200$ and $m_f=0.5$. In all cases, the saddlepoint approximation to the union bound (solid lines) is very accurate for moderate-to-large SNR. In particular, we observe the change in slope with respect to the standard union bound, which assumes that all bits in which the codewords differ are mapped onto different symbols (independent binary channels). The approximated threshold (19) computed with the minimum distance is $\mathrm{SNR}_{\mathrm{th}}=22$ dB for $\ell=40$ and $\mathrm{SNR}_{\mathrm{th}}=36$ dB for $\ell=200$. This floor is absent for independent binary parallel channels. In the next section, we prove that the floor appears at very low error rates (i. e. at very high SNR) for practical codes.

A. Asymptotic Analysis

In this section, we concentrate on the asymptotic study of the pairwise error probability averaged over all possible interleavers of length n, using Eq. (8). It seems clear that the best (i. e. steepest) conditional pairwise error probabilty $\mathrm{PEP}(d,n_2)$ is attained for $n_2=0$, which corresponds to all bits in the pairwise score belonging to different symbols. Similarly, the worst (i. e. flattest) conditional pairwise error probability $\mathrm{PEP}(d,n_2)$ is attained for $n_2^*=\lfloor\frac{d}{2}\rfloor$, which corresponds to the bits having maximal concentration in the minimum possible number of QPSK symbols. For these two cases, we show in Appendix B that the respective probabilities $P(n_2)$ admit the following asymptotic (in n) approximations

$$P(n_2=0)\simeq 1-\frac{d\left(d-\frac{1}{2}\right)}{\ell} \quad\text{and}\quad P(n_2=n_2^*)\simeq\begin{cases}2\left(\frac{d}{\ell e}\right)^{\frac{1}{2}d} & d \text{ even},\\ \frac{\sqrt{2}d^{d+\frac{1}{2}}}{e(\ell e)^{\frac{1}{2}(d-1)}(d-1)^{\frac{1}{2}d}} & d \text{ odd}.\end{cases} \tag{17}$$

We determine the position at which the respective pairwise probabilities cross by solving

$$P(n_2 = 0)\,\mathrm{PEP}(d, 0) = P(n_2 = n_2^*)\,\mathrm{PEP}(d, n_2^*). \tag{18}$$

Under the assumptions that SNR and ℓ are large, with the aid of Eq. (17), and keeping only the term at minimum distance $d_{\min}$ (which we denote without the subscript, to remove clutter from the equations), Eq. (18) is easily solved to obtain

$$\mathrm{SNR}_{\mathrm{th}} = \begin{cases} 4m_f \left(\frac{\ell e}{8^{\frac{1}{d}} d}\right)^{\frac{1}{m_f}} & d \text{ even} \\ 4m_f (\ell e)^{\frac{1}{m_f}} \left(\frac{e}{2}\right)^{\frac{2}{m_f(d-1)}} \frac{(d-1)^{\frac{d}{m_f(d-1)}} (d+1)^{\frac{1}{m_f(d-1)}}}{d^{\frac{2(d+1)}{m_f(d-1)}}} & d \text{ odd.} \end{cases} \tag{19}$$

For large d, the threshold is approximately given by $\mathrm{SNR}_{\mathrm{th}} \simeq 4m_f \left(\frac{\ell e}{d}\right)^{\frac{1}{m_f}}$ in both cases. For the special case of Rayleigh fading, $m_f = 1$, $\mathrm{SNR}_{\mathrm{th}}$ grows linearly with the interleaver length ℓ and is inversely proportional to the Hamming distance d for very large d.

Figure 2 depicts the approximate value of P_b, coarsely approximated as $2\,\mathrm{PEP}(d)$, at the threshold signal-to-noise ratio for several values of m_f and ℓ as a function of the minimum Hamming distance d. As expected from Figure 2, the error probability quickly becomes very small, at values typically below the operating point of common communication systems. For short packets using weak codes transmitted over channels with significant fading, the threshold may be of importance.

V. High-order Modulations: Asymptotic Analysis

In this section, we extend the analysis presented in the previous section for QPSK to general constellations and mappings, and estimate the signal-to-noise ratio at which the slope of the error probability changes. Since we are interested in large values of SNR, we shall be working with an asymptotic approximation to the error probability. In particular, our analysis is based on an extension of the results in [3], [7] for the asymptotic behaviour of the error probability in the Rayleigh-fading channel. We will provide such asymptotic expressions for the symbols scores of varying weight and use the approach presented for QPSK in the previous section to determine the value of SNR at which the various approximations to the error probability cross.

In [3], Caire *et al.* considered the Rayleigh fading at large SNR, and derived the following approximation to the cumulant transform of the bit score,

$$\kappa_1(s) \simeq -\log\left(\frac{d_{\mathrm{h}}^2}{4}\,\mathrm{SNR}\right), \tag{20}$$

where d_{h}^2 is a harmonic distance given by

$$d_{\mathrm{h}}^2 = \left(\frac{1}{m2^m} \sum_{c=0}^{1} \sum_{j=1}^{m} \sum_{x \in \mathcal{X}_c^j} \frac{1}{|x - x'|^2}\right)^{-1}, \tag{21}$$

where x' is the closest symbol in the constellation $\mathcal{X}_{\bar{c}}^j$ to x. Eq. (20) may be used in the Chernoff bound to give an approximation to the error probability at large SNR. It was found in [3] that this gives a good approximation. Using that $\hat{s} = \frac{1}{2}$, that $\lim_{\mathrm{SNR}\to\infty} \kappa_1(\hat{s}) = 8m_f$ [7] and the saddlepoint approximation in Eq. (11), we obtain the large SNR heuristic approximation

$$\mathrm{PEP}_1(d) \simeq \frac{1}{2\sqrt{\pi d}} \left(\frac{4}{d_{\mathrm{h}}^2 \mathrm{SNR}}\right)^d. \tag{22}$$

In Appendix C we extend the analysis in [3] to the cumulant transform of the BICM symbol score of weight v for Nakagami-m_f fading, and obtain the following limit

$$\kappa_v(\hat{s}) \simeq -m_f \log\left(\frac{d_{\rm h}^2(v)}{4m_f}\,\mathrm{SNR}\right), \tag{23}$$

where $d_{\rm h}^2(v)$ is a generalization of the harmonic distance given by

$$d_{\rm h}^2(v) = \left(\frac{1}{\binom{m}{v}2^m}\sum_{\boldsymbol{c}}\sum_{\boldsymbol{j}}\sum_{x\in\mathcal{X}_{\boldsymbol{j}}^{\boldsymbol{c}}}\left(\frac{\left|\sum_{i=1}^{v}(x-x_i')\right|}{\sum_{i=1}^{v}|x-x_i'|^2}\right)^{2m_f}\right)^{-\frac{1}{m_f}}. \tag{24}$$

For a given x, x_i' is the i-th symbol in the sequence of v symbols $(x_1',\ldots,x_v')$ which have binary label $\bar{c}_{j_i}$ at position j_i and for which the ratio $\frac{\left|\sum_{i=1}^{v}(x-x_i')\right|}{\sum_{i=1}^{v}|x-x_i'|^2}$ is minimum among all possible such sequences. For $m_f=1$ and $v=1$ we recover the harmonic distance $d_{\rm h}^2$ above.

As it happened with the bit score and $\mathrm{PEP}_1(d)$, Eq. (23) may be used in the Chernoff bound or in the saddlepoint approximation in Eq. (11), to obtain a heuristic approximation to the pairwise error probability for large SNR, namely

$$\mathrm{PEP}_{\rm H}(d,\rho_n) \simeq \frac{1}{2\sqrt{\pi m_f\sum_{v\geq 1}n_v}}\prod_{v=1}^{m}\left(\frac{4m_f}{d_{\rm h}^2(v)}\frac{1}{\mathrm{SNR}}\right)^{n_v m_f}. \tag{25}$$

For the sake of simplicity, we disregard the effect of the coefficient $\frac{1}{\hat{s}\sqrt{2\pi\kappa_{\rm pw}(\hat{s})}}$ in our analysis of the threshold SNR.

For sufficiently large signal-to-noise ratio, the error probability is determined by the worst possible distribution pattern ρ_n of the d bits onto the n symbols, that with the largest tail probability for the pairwise score. Then, since $d=\sum_{v=1}^{w^\star}n_v v$ by construction, we can view the pattern $\rho_n=(n_0,n_1,\ldots,n_{w^\star})$ as a (non-unique) representation of the integer d as a weighted sum of the integers $\{0,1,2,\ldots,w^\star\}$. By construction, the sum $\sum_{v\geq 1}n_v$ is the number of non-zero Hamming weight symbols in the candidate codeword. Clearly, the lowest $\sum_{v\geq 1}n_v$ gives the worst (flattest) pairwise error probability in the presence of fading. We obtain an equation for $\mathrm{SNR}_{\rm th}$ similar to Eq. (18) for QPSK for the fully-interleaved fading channel,

$$\left(P(\rho_0)\left(\frac{4m_f}{d_{\rm h}^2(1)}\frac{1}{\mathrm{SNR}_{\rm th}}\right)^{m_f}\right)^{d} = \frac{1}{\mathrm{SNR}_{\rm th}^{m_f\sum_{v=1}^{m}n_v}}\sum_{\rho_n:\min\sum_v n_v}P(\rho_n)\prod_{v=1}^{m}\left(\frac{4m_f}{d_{\rm h}^2(v)}\right)^{n_v m_f}. \tag{26}$$

The left-hand side corresponds to the steepest pairwise error probability, namely $\mathrm{PEP}_1(d)$, weighted by the probability that all bit scores are independent, denoted by $P(\rho_0)$. The right-hand side corresponds to the largest pairwise error probability with smallest number of non-zero symbol scores, that is among all possible patterns ρ_n with minimum $\sum_v n_v$. Note that the exponent of SNR is $-m_f\sum_v n_v$, and thus has the lowest possible diversity, as it should.

From Eq. (26), we can extract the value of $\mathrm{SNR}_{\rm th}$ as

$$\mathrm{SNR}_{\rm th} \simeq 4m_f\left(\sum_{\rho_n:\min\sum_v n_v}\frac{P(\rho_n)\left(d_{\rm h}^2(1)\right)^d}{P(\rho_0)\prod_v\left(d_{\rm h}^2(v)\right)^{n_v}}\right)^{-\frac{1}{m_f(d-\sum_v n_v)}}. \tag{27}$$

As expected, for the specific case of QPSK with Gray mapping, computation of this value of $\mathrm{SNR}_{\rm th}$ ($d_{\rm h}^2(1)=2$, $d_{\rm h}^2(2)=4$) gives a result which is consistent with the result derived in Section IV-A, namely Eq. (19), with the minor difference that we use now the Chernoff bound whereas the saddlepoint approximation was used in the QPSK case.

As we observe in Figure 3, the slope change present for QPSK is also present for 8-PSK with Rayleigh fading and Gray labeling; the code is the optimum 8-state rate-2/3 convolutional code, with $d_{\min} = 4$. Again, this effect is due to the probability of having symbol scores of Hamming weight larger than 1. The figure depicts simulation results (for interleaver sizes $\ell = 90, 3000$) together with the saddlepoint approximations for finite ℓ (for $\ell = 90, 300, 3000$), infinite interleaving (with $\mathrm{PEP}_1(d)$), and with the heuristic approximation $\mathrm{PEP}_{\mathrm{H}}(d\,\rho_n)$ (only for $\ell = 90$ and $d = 4$). For 8-PSK with Gray mapping, evaluation of Eq. (24) gives $d_{\mathrm{h}}^2(1) = 0.7664$, $d_{\mathrm{h}}^2(2) = 1.7175$, and $d_{\mathrm{h}}^2(3) = 2.4278$. Table I gives the values of $P(\rho_n)$ for the various patterns ρ_n. Table I also gives the threshold $\mathrm{SNR}_{\mathrm{th}}$ given in Eq. (27) for all possible values of $\sum_{v\geq 1} n_v$, not only for the worst case. We observe that the main flattening of the error probability takes place at high SNR. This effect essentially disappears for interleavers of practical length: for $\ell = 300$ (resp. $\ell = 3000$) the error probability at the first threshold is about 10^{-8} (resp. 10^{-12}). The saddlepoint approximation is remarkably precise; the heuristic approximation $\mathrm{PEP}_{\mathrm{H}}(d_{\mathrm{dmin}}, \rho_n)$ also gives very good results.

VI. Conclusions

We have studied the large-SNR behavior of the error probability of BICM over fully-interleaved fading channels. Our analysis reveals that the pairwise error probability is asymptotically dominated by the number pairwise error symbols with Hamming weight larger than one, yielding an error floor. We have derived closed-form approximations to this error probability. For practical code lengths, the error floor appears at very low error rates.

Appendix A
Probability of All-One Sequence

We use Stirling's approximation to the factorial, $n! \simeq n^n e^{-n}\sqrt{2\pi n}$, to Eq. (8) to obtain

$$P_{\text{ind}} \triangleq P(n_1 = d, n_2 = 0, \ldots, n_{w^\star} = 0) = \frac{\ell^n}{(\ell - md)^{n-d+\frac{1}{2}}} \frac{(\ell-d)^{\ell-d+\frac{1}{2}}}{\ell^\ell}, \tag{28}$$

with the obvious simplifications and combinations. Extracting a factor ℓ in $(\ell - d)$ and $(\ell - md)$, and cancelling common powers of ℓ in numerator and denominator, we get

$$P_{\text{ind}} \simeq \left(1 - \frac{d}{n}\right)^{-n+d-\frac{1}{2}} \left(1 - \frac{d}{\ell}\right)^{\ell-d+\frac{1}{2}}. \tag{29}$$

We now take logarithms, and use Taylor's expansion of the logarithm, $\log(1+t) \simeq t - \frac{1}{2}t^2$, in the right-hand side of Eq. (29). Discarding all powers of ℓ higher than ℓ^{-2}, and combining common terms, we obtain

$$\log P_{\text{ind}} \simeq -\frac{md^2}{2\ell} + \frac{d}{2n} + \frac{d^2}{2\ell} - \frac{d}{2\ell} = -\frac{d(d-1)}{2\ell}(m-1). \tag{30}$$

Finally, recovering the exponential, $P_{\text{ind}} \simeq e^{-\frac{d(d-1)}{2\ell}(m-1)}$.

Appendix B
Asympotics of $P(n_2)$

Assume that d is even, so $n_2 = \frac{1}{2}d$ and $n_1 = 0$. Then

$$P(n_2) = \frac{d!(2n-d)!n!}{(2n)!(n-\frac{1}{2}d)!(\frac{1}{2}d)!}. \tag{31}$$

Using Stirling's approximation to the factorial, and after some simplifications, we have that

$$P(n_2) \simeq 2\left(\frac{2n-d}{2n}\right)^n \left(\frac{d}{2n-d}\right)^{\frac{1}{2}d}. \tag{32}$$

In the limit of large n, using that $\lim_{n\to\infty}\left(1+\frac{a}{n}\right)^n = e^a$, and $2n \gg d$, we have

$$P(n_2) \simeq 2e^{-\frac{1}{2}d}\left(\frac{d}{2n}\right)^{\frac{1}{2}d} = 2\left(\frac{d}{2ne}\right)^{\frac{1}{2}d}. \tag{33}$$

If d is odd, then $n_2 = \frac{1}{2}(d-1)$ and $n_1 = 1$. Then

$$P(n_2) = \frac{2d!(2n-d)!n!}{(2n)!\left(n-\frac{1}{2}(d+1)\right)!\left(\frac{1}{2}(d-1)\right)!}. \tag{34}$$

Using again Stirling's approximation, and after some simplifications, we have

$$P(n_2) \simeq \frac{\sqrt{2}}{e}\left(\frac{2n-d}{2n}\right)^{2n}\left(\frac{n}{n-\frac{1}{2}(d+1)}\right)^n \frac{d^{d+\frac{1}{2}}\left(2n-(d+1)\right)^{\frac{1}{2}d}}{(2n-d)^{d-\frac{1}{2}}(d-1)^{\frac{1}{2}d}}. \tag{35}$$

Again for large n, using that $\lim_{n\to\infty}\left(1+\frac{a}{n}\right)^n = e^a$, and $2n \gg d$, we have

$$P(n_2) \simeq \frac{\sqrt{2}}{e}e^{-d}e^{\frac{1}{2}(d+1)}\frac{d^{d+\frac{1}{2}}}{(2n)^{\frac{1}{2}(d-1)}(d-1)^{\frac{1}{2}d}} = \frac{\sqrt{2}d^{d+\frac{1}{2}}}{e(2ne)^{\frac{1}{2}(d-1)}(d-1)^{\frac{1}{2}d}}. \tag{36}$$

APPENDIX C
ASYMPTOTIC ANALYSIS WITH FULLY-INTERLEAVED NAKAGAMI FADING

We wish to compute the limit $\ell_v(s) \triangleq \lim_{\mathrm{SNR}\to\infty}\frac{e^{\kappa_v(s)}}{\mathrm{SNR}^{-m_f}}$, given by

$$\ell_v(s) = \lim_{\mathrm{SNR}\to\infty}\frac{1}{\mathrm{SNR}^{-m_f}}\left(\frac{1}{2^v\binom{m}{v}}\sum_{\boldsymbol{j},\boldsymbol{c}}\mathrm{E}\left[\frac{\prod_{i=1}^v q_{j_i}(\bar{c}_{j_i},Y)^s}{\prod_{i=1}^v q_{j_i}(c_{j_i},Y)^s}\right]\right). \tag{37}$$

We can rewrite the denominator as

$$\prod_{i=1}^v q_{j_i}(\bar{c}_{j_i},y) = \prod_{i=1}^v\left(\sum_{x'\in\mathcal{X}_{\bar{c}_{j_i}}^{j_i}} e^{-|H\sqrt{\mathrm{SNR}}(X-x')+Z|^2}\right) = \sum_{\boldsymbol{x}'}\prod_{i=1}^v e^{-|H\sqrt{\mathrm{SNR}}(X-x_i')+Z|^2}, \tag{38}$$

where $\boldsymbol{x}'$ is one of all possible sequences of v modulation symbols, with symbol at index j_i drawn from the set $\mathcal{X}_{\bar{c}_{j_i}}^{j_i}$. A similar formula holds for the denominator, now with symbols drawn from the set $\mathcal{X}_{c_{j_i}}^{j_i}$. Expanding the exponent in Eq. (38), we obtain

$$\prod_{i=1}^v q_{j_i}(\bar{c}_{j_i},Y) = \sum_{\boldsymbol{x}'} e^{-|H|\mathrm{SNR}\sum_{i=1}^v|X-x_i'|^2+2\sqrt{\mathrm{SNR}}\,\mathrm{Re}\left(\sum_{i=1}^v H(X-x_i')Z^*\right)+v|Z|^2}. \tag{39}$$

As done in [3], [7], we keep only the dominant summand in the bit scores $q_{j_i}(\cdot,y)$ appearing in numerator and denominator. For a given x, this summand corresponds to the sequence $\boldsymbol{x}'$ having the smallest possible value of the ratio $\frac{\left|\sum_{i=1}^v(x-x_i')\right|}{\sum_{i=1}^v|x-x_i'|^2}$. In particular, in the denominator all the symbols in the sequence coincide with x. We now carry out the expectation over Z. Completing squares, and using that the formula for the density of Gaussian noise, we have that

$$\begin{aligned}\int\frac{1}{\pi}e^{-|z|^2}e^{-\sum_{i=1}^v s\left(\mathrm{SNR}|H|^2|X-X_i'|^2+2\sqrt{\mathrm{SNR}}\,\mathrm{Re}(H(X-X_i')z^*)\right)}dz \\ = e^{-\mathrm{SNR}|H|^2\left(s\sum_{i=1}^v|X-X_i'|^2-s^2\left|\sum_{i=1}^v(X-X_i')\right|^2\right)}.\end{aligned} \tag{40}$$

In turn, the expectation over h of this quantity yields [9]

$$\left(1+\left(s\sum_{i=1}^{v}|X-X_i'|^2-s^2\left|\sum_{i=1}^{v}(X-X_i')\right|^2\right)\frac{\mathrm{SNR}}{m_f}\right)^{-m_f}. \tag{41}$$

We next turn back to the limit of large SNR. We have that

$$\ell_v(s)=\frac{m_f^{m_f}}{2^m\binom{m}{v}}\sum_{\boldsymbol{j},\boldsymbol{c}}\sum_{x\in\mathcal{X}_{\boldsymbol{c}}^{\boldsymbol{j}}}\left(s\sum_{i=1}^{v}|x-x_i'|^2-s^2\left|\sum_{i=1}^{v}(x-x_i')\right|^2\right)^{-m_f}. \tag{42}$$

For each summand, the optimizing s is readily computed to be $\hat{s}=\frac{\sum_{i=1}^{v}|x-x_i'|^2}{2\left|\sum_{i=1}^{v}(x-x_i')\right|^2}$, which gives

$$\ell_v(\hat{s})=\frac{1}{2^m\binom{m}{v}}\sum_{\boldsymbol{j},\boldsymbol{c}}\sum_{x\in\mathcal{X}_{\boldsymbol{c}}^{\boldsymbol{j}}}\left(\frac{4m_f\left|\sum_{i=1}^{v}(x-x_i')\right|^2}{\left(\sum_{i=1}^{v}|x-x_i'|^2\right)^2}\right)^{m_f}. \tag{43}$$

REFERENCES

[1] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes." in *IEEE Int. Conf. Commun., Geneva*, vol. 2, 1993.

[2] S. Benedetto and G. Montorsi, "Unveiling turbo codes: some results on parallel concatenated coding schemes," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 409–428, 1996.

[3] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, 1998.

[4] J. G. Proakis, "Digital Communications," *McGraw-Hill USA*, 4th edition, 2001.

[5] P. Yeh, S. Zummo, and W. Stark, "Error probability of bit-interleaved coded modulation in wireless environments," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 2, pp. 722–728, 2006.

[6] R. W. Butler, *Saddlepoint Approximations with Applications*. Cambridge University Press, 2007.

[7] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "Error probability analysis of bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 262–271, Jan. 2006.

[8] S. Zummo, P. Yeh, and W. Stark, "A Union Bound on the Error Probability of Binary Codes Over Block Fading Channels," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 6, pp. 2085 – 2093, 2005.

[9] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "A closed-form approximation for the error probability of BPSK fading channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2051–2056, Jun. 2007.

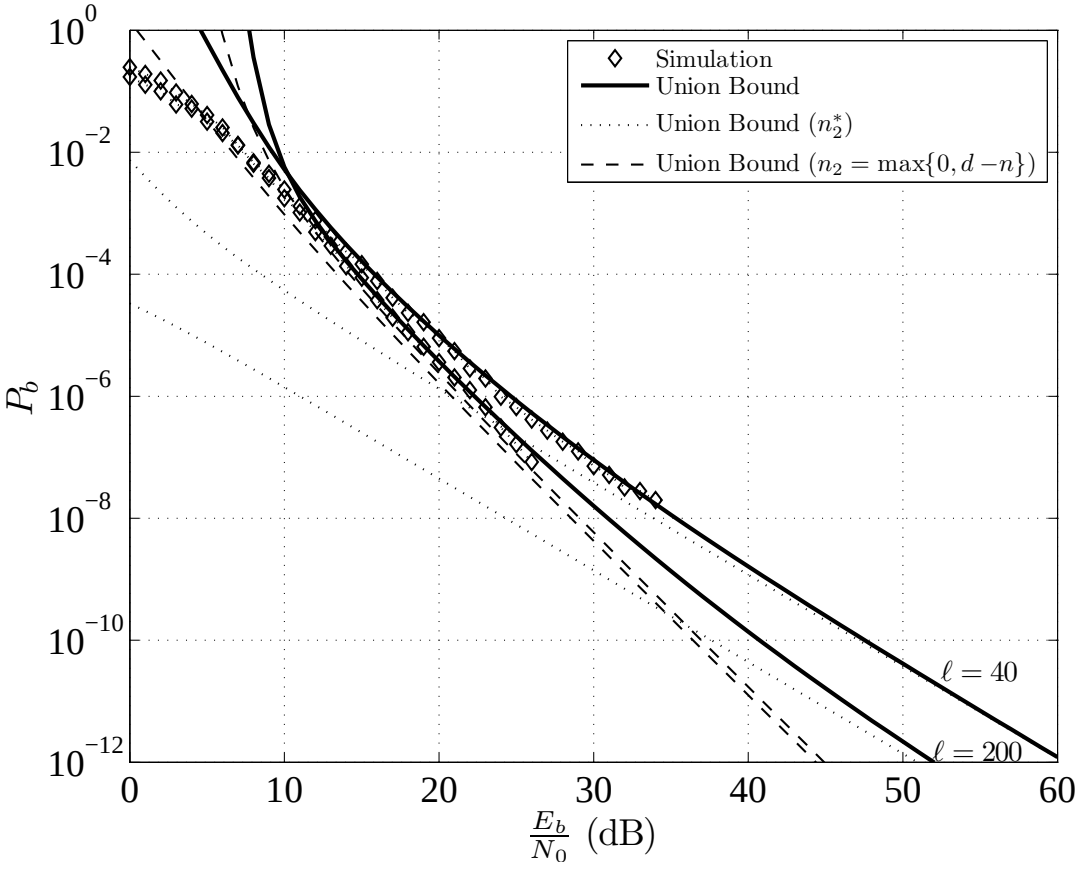

Fig. 1. Error probability of QPSK and the $(5,7)_8$ convolutional code with a uniform interleaver of length $\ell = 40, 200$ in a fully-interleaved fading channel with $m_f = 0.5$. Diamonds correspond to bit-error rate simulation, the solid line corresponds to the union bound, dashed lines correspond to the union bound for $n_2 = \max\{0, d-n\}$ (the upper one corresponds to $\ell = 200$) and dotted lines correspond to the union bound for $n_2 = n_2^*$.

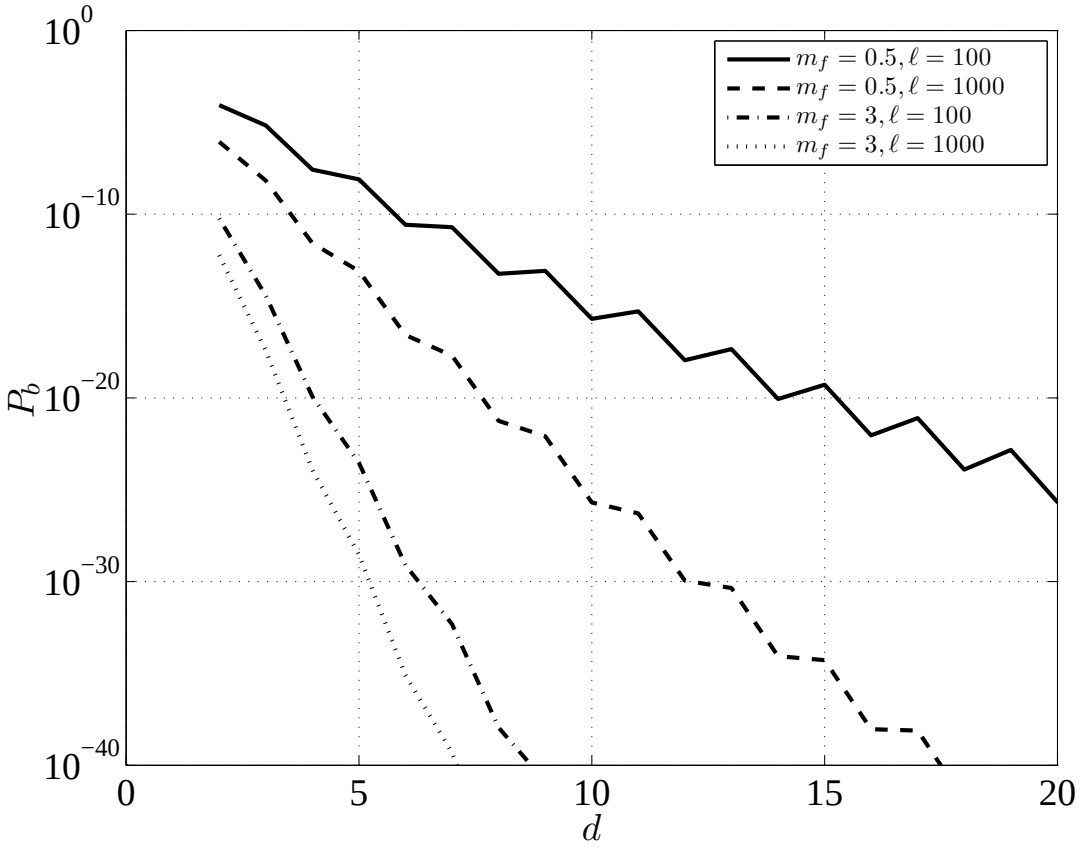

Fig. 2. Approximation to the bit error probability of QPSK at the threshold signal-to-noise ratio $\mathrm{SNR}_{\mathrm{th}}$, $P_b = 2\,\mathrm{PEP}_1(d)$, as a function of the minimum Hamming distance d for several values of m_f and ℓ. In solid line, $m_f = 0.5, \ell = 100$; in dashed line, $m_f = 0.5, \ell = 1000$; in dash-dotted line, $m_f = 3, \ell = 100$; and in dotted line $m_f = 3, \ell = 1000$.

TABLE I

ASYMPTOTIC ANALYSIS FOR 8-PSK WITH VARYING INTERLEAVER LENGTH $\ell = 3n$ AND MINIMUM DISTANCE $d = 4$.

Pattern ρ_n	ℓ	$P(\rho_n)$	Threshold $\frac{E_b}{N_0}$ (dB)
$(n-4, 4, 0, 0)$	$\ell = 90$	0.8688	N/A
	$\ell = 300$	0.9602	N/A
	$\ell = 3000$	0.9960	N/A
$(n-3, 2, 1, 0)$	$\ell = 90$	0.1287	16.0
	$\ell = 300$	0.0396	21.5
	$\ell = 3000$	0.0040	31.6
$(n-2, 0, 2, 0)$,	$\ell = 90$	0.0015	20.5
$(n-1, 1, 0, 1)$	$\ell = 300$	0.0002	26.0
	$\ell = 3000$	$2 \cdot 10^{-6}$	39.1

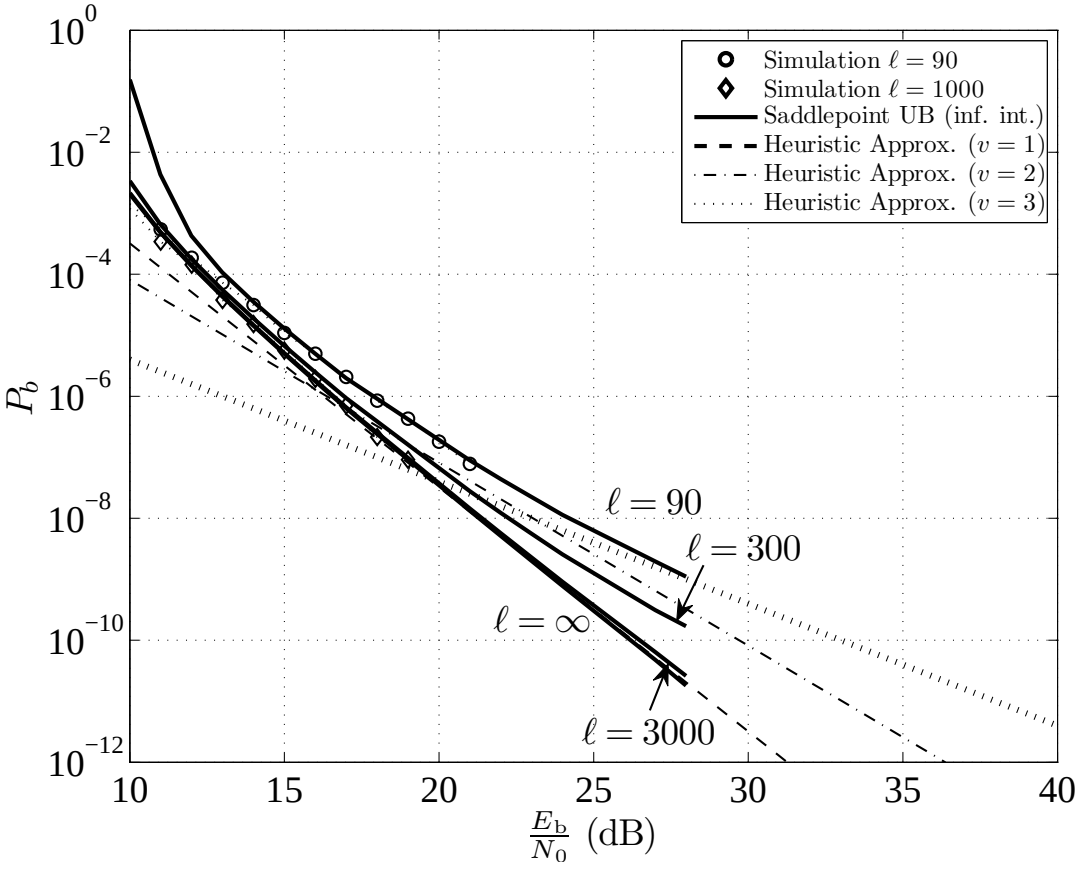

Fig. 3. Bit error probability union bounds and bit-error rate simulations of 8-PSK with the 8-state rate-2/3 convolutional code in a fully-interleaved Rayleigh fading channel. Interleaver length $\ell = 90$ (circles) and $\ell = 3000$ (diamonds). In solid lines, the saddlepoint approximation union bounds for $\ell = 90$, $\ell = 300$, $\ell = 3000$ and for infinite interleaving, with $\mathrm{PEP}_1(d)$. In dashed, dashed-dotted, and dotted lines, the heuristic approximations with weight $v = 1, 2, 3$ respectively.