Coded Modulation in the Block-Fading Channel: Coding Theorems and Code Construction

Albert Guillén i Fàbregas and Giuseppe Caire [*] [†] [‡]

September 22, 2005

Abstract

We consider coded modulation schemes for the block-fading channel. In the setting where a codeword spans a finite number N of fading degrees of freedom, we show that coded modulations of rate R bit/complex dimension, over a finite signal set $\mathcal{X} \subset \mathbb{C}$ of size 2^M, achieve the optimal rate-diversity tradeoff given by the Singleton bound $\delta(N, M, R) = 1 + \lfloor N(1 - R/M) \rfloor$, for $R \in (0, M]$. Furthermore, we show also that the popular bit-interleaved coded modulation achieve the same optimal rate-diversity tradeoff. We present a novel coded modulation construction based on blockwise concatenation that systematically yields Singleton-bound achieving turbo-like codes defined over an arbitrary signal set $\mathcal{X} \subset \mathbb{C}$. The proposed blockwise concatenation significantly outperforms conventional serial and parallel turbo codes in the block-fading channel. We analyze the *ensemble* average performance under Maximum-Likelihood (ML) decoding of the proposed codes by means of upper bounds and tight approximations. We show that, differently from the AWGN and fully-interleaved fading cases, Belief-Propagation iterative decoding performs very close to ML on the block-fading channel for any signal-to-noise ratio and even for relatively short block lengths. We also show that, at constant decoding complexity per information bit, the proposed codes perform close to the information outage probability for any block length, while standard block codes (e.g., obtained by trellis-termination of convolutional codes) have a gap from outage that increases with the block length: this is a different and more subtle manifestation of the so-called "interleaving gain" of turbo codes.

Index Terms: Block-fading channels, outage probability, diversity, MDS codes, concatenated codes, ML decoding, distance spectrum, iterative decoding, bit-interleaved coded modulation.

1 Introduction

The block-fading channel was introduced in [1] (see also [2]) in order to model slowly-varying fading, where codewords span only a fixed number N of fading degrees of freedom, irrespectively of the code block length. This model is particularly relevant in wireless communications situations involving slow time-frequency hopping (e.g., GSM, EDGE) or multicarrier modulation using orthogonal frequency division multiplexing (OFDM). More in general, despite its extreme simplification, it serves as a useful model to develop coding design criteria which turn out to be useful in more general settings of correlated slowly-varying fading.

Coding for the block-fading channel has been considered in a number of recent works (e.g., [3, 4, 5, 6] and references therein). The design criteria for codes over the block-fading channel differ significantly with respect to the standard design criteria for codes over the AWGN channel or over the fully-interleaved fading channel. The key difference is that the block-fading channel is not information stable [7, 8]. Under mild conditions on the fading distribution, the reliability function of the block-fading channel is zero for any finite Signal-to-Noise Ratio (SNR).

Using union bound arguments [3, 4, 5, 6] and error exponent calculations [9], it was shown that in Rayleigh fading the error probability behaves like $O(\text{SNR}^{-d_B})$ for large SNR. The exponent d_B, an integer in $[0, N]$, is referred to as the code *block diversity* and is given by the minimum number of blocks on which any two distinct codewords differ (block-wise Hamming distance). If the code is constructed over a finite alphabet (signal set), there exists a tradeoff between the achievable block diversity and the coding rate. More precisely, a code over an alphabet $\mathcal{X}$ of cardinality $|\mathcal{X}|$, partitioned into N blocks of length L, can be seen as a code over the alphabet $\mathcal{X}^L$ of cardinality $|\mathcal{X}|^L$ with block length N. Hence, we have trivially that any upper bound on the minimum Hamming distance of $|\mathcal{X}|^L$-ary codes of length N and size A yields a corresponding upper bound on the achievable block diversity d_B for codes over $\mathcal{X}$ and rate $R = \frac{1}{NL} \log_2 A$. In [9, Th. 1], it is shown that for binary codes the Singleton bound is tight for any $R \in (0, 1)$. The achievability proof in [9, Th. 1] is based on the existence of maximum distance separable (MDS) codes over $\mathbb{F}_{2^L}$ (e.g., Reed-Solomon codes).

In general, we define the SNR exponent of error probability for a given family of codes as

$$d^\star \triangleq \sup_{\mathcal{C} \in \mathcal{F}} \lim_{\rho \to \infty} \frac{-\log P_e(\rho, \mathcal{C})}{\log \rho} \tag{1}$$

where ρ denotes the channel SNR, $P_e(\rho, \mathcal{C})$ is the error probability of code $\mathcal{C}$, and the supremum

is taken over all codes in the family $\mathcal{F}$.

In [10], a block-fading multiple-input multiple-output (MIMO) channel with $N = 1$ fading blocks is considered and no restriction is imposed on the code family other than the standard average input power constraint. For every $r > 0$, codes of rate $R(\rho) = r \log \rho$ are considered and the optimal SNR exponent is found as a function of r. It is also shown that the optimal exponent coincides with the random coding exponent for an ensemble of Gaussian i.i.d. codes of fixed block length, provided that the block length is larger than a certain integer that depends on the number of transmit and receive antennas.

In this work, we consider a single-input single-output (SISO) block-fading channel with arbitrary (but fixed) number N of fading blocks. We are interested in the ensemble of coded modulations, i.e., codes over a given finite signal set $\mathcal{X} \subset \mathbb{C}$ with fixed rate R that, obviously, cannot be larger than $M = \log_2 |\mathcal{X}|$ bit/complex dimension. We study the SNR exponent (1) as a function of the coding rate, denoted by $d_{\mathcal{X}}^{\star}(R)$. This "SNR reliability function" represents the optimal rate-diversity tradeoff for the given family of codes. We prove that $d_{\mathcal{X}}^{\star}(R)$ is indeed given by the Singleton bound, and we find an explicit expression for the random-coding SNR error exponent, denoted by $d_{\mathcal{X}}^{(r)}(R)$, which lower bounds $d_{\mathcal{X}}^{\star}(R)$ and is tight for all R provided that the code block length grows rapidly enough with respect to $\log(\rho)$: namely, the code block length must be superlinear in the channel SNR expressed in dB. Furthermore, we show that the popular pragmatic Bit-Interleaved Coded Modulation (BICM) scheme [11] achieves the same $d_{\mathcal{X}}^{(r)}(R)$ (and hence $d_{\mathcal{X}}^{\star}(R)$), subject to the same condition on the block length growth with respect to SNR).

Then, we focus on the systematic construction of codes achieving the optimal SNR exponent and we introduce a turbo-like code construction suited to the block-fading channel. Notice that standard code ensemble analysis and optimization techniques based on Density Evolution [12] and on various approximations thereof, such as the ubiquitous EXtrinsic Information Transfer (EXIT) functions [13], are useless over the block-fading channel. In fact, these techniques aim at finding the iterative decoding threshold, defined as the minimum SNR at which the bit error rate (BER) vanishes after infinitely many iterations of the Belief-Propagation (BP) iterative decoder, for a given code ensemble in the limit of infinite block length. In our case, since the block-fading channel is affected by a finite number N of fading coefficients that do not average out as the block length grows to infinity, the iterative decoding threshold is a random variable that depends on the channel realization. Hence, one should optimize the distribution of the fixed points of the Density

3

Evolution with respect to the code ensemble: clearly, a very difficult and mostly impractical task.

For our codes we provide upper bounds and tight approximations to the error probability under maximum-likelihood (ML) decoding. While ML decoding is generally infeasible because of complexity, we show by simulation that the iterative Belief-Propagation (BP) "turbo" decoder performs very close to the ML error probability. This fact stands in stark contrast with the typical behavior of turbo and LDPC codes on the AWGN and fully interleaved fading channels [14, 15, 16, 17, 18], where ML bounds are able to predict accurately the "error floor region" but are quite inaccurate in the "waterfall region" of the BER curve. Hence, our bounds and approximations are relevant, in the sense that they indeed provide very accurate performance evaluation of turbo-like coded modulation in the block-fading channel under BP iterative decoding.

The proposed coded modulation schemes outperform standard turbo-coded or LDPC-coded modulation and outperform also previously proposed trellis codes for the block-fading channel [3, 5, 6]. In particular, by using asymptotic weight enumerator techniques, we show that the word-error rate (WER) of our codes is almost independent of the block length, while the component encoders are fixed, i.e., the decoding complexity of the BP decoder is linear with the block length. On the contrary, in the case of block codes obtained by trellis termination of trellis codes, the WER increases (roughly linearly) with the block length for linear decoding complexity. We interpret this fact as another manifestation of the so-called "interleaving gain" typical of turbo codes, even though, in block-fading, no "waterfall" behavior of the error curve is visible, even for very large block length.

The paper is organized as follows. Section 2 defines the system model. Section 3 presents the coding theorems for the rate-diversity tradeoff of coded modulation and BICM. In Section 4 we present our novel turbo-like coded modulation scheme, we provide useful upper bounds and approximations of its error probability under ML decoding and we show that the error probability is (asymptotically) independent of the block length. Also, several examples of code construction and performance comparisons are provided. Section 5 summarizes the conclusions of this work. Proofs and computation details of the error bounds and approximations are reported in the appendices.

2 System model

We consider the block-fading channel model [1] with N fading blocks, where each block has length L complex dimensions. Fading is flat, constant on each block, and i.i.d. on different blocks. The discrete-time complex baseband equivalent channel model is given by

$$\mathbf{y}_n = \sqrt{\rho}\, h_n\, \mathbf{x}_n + \mathbf{z}_n \quad, \; n = 1, \ldots, N \tag{2}$$

where $\mathbf{y}_n, \mathbf{x}_n, \mathbf{z}_n \in \mathbb{C}^L$, h_n denotes the n-th block fading coefficient and the noise $\mathbf{z}_n$ is i.i.d. complex circularly-symmetric Gaussian, with components $\sim \mathcal{N}_{\mathbb{C}}(0, 1)$.

We consider codes constructed over a complex signal-set $\mathcal{X}$ (e.g., QAM/PSK) of cardinality 2^M, i.e., the components of the vectors $\mathbf{x}_n$ are points in the constellation $\mathcal{X}$. The overall codeword block length is NL (complex dimensions). Therefore, each codeword spans at most N independent fading coefficients. Without loss of generality, we assume normalized fading, such that $\mathbb{E}[|h_n|^2] = 1$ and unit-energy signal set $\mathcal{X}$ (i.e., $2^{-M} \sum_{x \in \mathcal{X}} |x|^2 = 1$). Therefore, ρ denotes the *average* received SNR and the *instantaneous* SNR on block n is given by $\gamma_n \rho$, where $\gamma_n \triangleq |h_n|^2$ denotes the fading *power gain*.

The channel (2) can be expressed in the concise matrix form

$$\mathbf{Y} = \sqrt{\rho}\, \mathbf{H}\mathbf{X} + \mathbf{Z} \tag{3}$$

where $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]^T \in \mathbb{C}^{N \times L}$, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]^T \in \mathbb{C}^{N \times L}$, $\mathbf{H} = \mathrm{diag}(h_1, \ldots, h_N) \in \mathbb{C}^{N \times N}$ and $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_N]^T \in \mathbb{C}^{N \times L}$.

The collection of all possible transmitted codewords $\mathbf{X}$ forms a coded modulation scheme over $\mathcal{X}$. We are interested in schemes $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ obtained by concatenating a binary linear code $\mathcal{C}$ of length NLM and rate r bit/symbol with a memoryless one-to-one symbol mapper $\mu : \mathbb{F}_2^M \to \mathcal{X}$. The resulting coding rate (in bit/complex dimension) is given by $R = rM$.

In this work we assume that the vector of fading coefficients $\mathbf{h} = (h_1, \ldots, h_N)$ is perfectly known at the receiver and not known at the transmitter. It is worthwhile to notice that in the limit of $L \to \infty$ and fixed N, the capacity and, more generally, the outage capacity, of the block-fading channel does not depend on the assumption of perfect channel knowledge at the receiver [2]. Therefore, in this limit such assumption is not optimistic.

Let $w \in \{1, \ldots, |\mathcal{M}|\}$ denote the information message and $\mathbf{X}(w)$ denote the codeword corresponding to w. We shall consider the following decoders:

1. The ML decoder, defined by

$$\widehat{w} = \arg \min_{w=1,\dots,|\mathcal{M}|} \|\mathbf{Y} - \sqrt{\rho}\,\mathbf{H}\mathbf{X}(w)\|_F^2 \tag{4}$$

($\|\cdots\|_F$ denotes the Frobenius norm).

2. A suboptimal decoder that consists of producing, for each received symbol, the posterior probabilities of the binary coded symbols in its label (defined by the symbol mapper μ), and then feeding these probabilities to a ML decoder for the binary code $\mathcal{C}$ over the resulting binary-input continuous-output channel. Since this scheme is particularly effective if used in conjunction with BICM [11], we shall refer to it as the BICM-ML decoder (even though it can also be used without an explicit bit-interleaver between $\mathcal{C}$ and μ). It follows from the definition of the ensemble $\{\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})\}$ that the coded bits output by the binary linear encoder for $\mathcal{C}$ are partitioned into N blocks of length LM, each of which is further partitioned into L binary labels of length M bits, which are eventually mapped into modulation symbols by the mapping μ. Let again w denote the information message and let $\mathbf{C}(w) \in \mathcal{C}$ denote the codeword of $\mathcal{C}$ corresponding to w. The components of of $\mathbf{C}(w)$ are indicated by $c_{n,k,m}(w)$ where the triple of indices (n, k, m) indicates the fading block, the modulation symbol, and the label position. The corresponding "bit-wise" posterior log-probability ratio is given by

$$\mathcal{L}_{n,k,m} = \log \frac{\sum_{x \in \mathcal{X}_0^m} \exp\left(-|y_{n,k} - \sqrt{\rho}\,h_n x|^2\right)}{\sum_{x \in \mathcal{X}_1^m} \exp\left(-|y_{n,k} - \sqrt{\rho}\,h_n x|^2\right)} \tag{5}$$

where $\mathcal{X}_a^m$ denotes the signal subset of all points in $\mathcal{X}$ whose label has value $a \in \{0, 1\}$ in position m. Then, the BICM-ML decoding rule is given by

$$\widehat{w} = \arg \max_{w=1,\dots,|\mathcal{M}|} \sum_{n=1}^{N} \sum_{k=1}^{L} \sum_{m=1}^{M} (1 - 2c_{n,k,m}(w))\mathcal{L}_{n,k,m} \tag{6}$$

In all cases, the average word-error rate (WER) as a function of SNR, averaged over the fading, is defined as $P_e(\rho) = \Pr(w \neq \widehat{w})$ where a uniform distribution of the messages is assumed.

As it will be clear in the following, both the ML and the BICM-ML decoders are practically infeasible for the class of coded modulation schemes proposed in this paper. Hence, the suboptimal turbo decoder based on Belief-Propagation (BP) will be used instead. Nevertheless, the two

decoders defined above are easier to analyze and provide a benchmark to compare the performance of the BP decoder. Since BP iterative decoding is standard and well-known, for the sake of space limitation we shall omit the detailed BP decoder description. The reader is referred to e.g. [19] for details.

3   Optimal rate-diversity tradeoff

Let $I(P_X, \mathbf{h}) \triangleq \frac{1}{NL} I(\mathbf{X}; \mathbf{Y}|\mathbf{h})$ denote the mutual information (per complex dimension) between input and output, for given fading coefficients $\mathbf{h}$ and NL-dimensional input probability assignment P_X, satisfying the input power constraint $\frac{1}{NL} \mathbb{E}[\|\mathbf{X}\|_F^2] = 1$. Since $\mathbf{h}$ is random, $I(P_X, \mathbf{h})$ is generally a random variable with given cumulative distribution function $F_I(z) \triangleq \Pr(I(P_X, \mathbf{h}) \le z)$. The channel ϵ-capacity (as a function of SNR ρ) is given by [7]

$$C_\epsilon(\rho) = \sup_{P_X} \ \sup \left\{ z \in \mathbb{R} \ : \ F_I(z) \le \epsilon \right\} \tag{7}$$

The channel capacity is given by $C(\rho) = \lim_{\epsilon \downarrow 0} C_\epsilon(\rho)$. For fading distributions such that $P(|h| < \delta) > 0$ for any $\delta > 0$ (e.g., Rayleigh or Rice fading), we have $C(\rho) = 0$ for all $\rho \in \mathbb{R}_+$, meaning that no positive rate is achievable. Hence, the relevant measure of performance on this channel is the optimal WER [1] given by

$$\epsilon(\rho) = \inf_{P_X} F_I(R) \tag{8}$$

In many cases, the input distribution is fixed by some system constraint. Hence, it is customary to define the *information outage probability* [2, 1] as $P_{\text{out}}(\rho, R) \triangleq F_I(R)$ for given P_X, ρ and R. The goodness of a coding scheme for the block-fading channel is measured by the SNR gap from outage probability for large block length L.

For the ensemble $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ where $\mathcal{C}$ is a random binary linear code, P_X is the uniform i.i.d. distribution over $\mathcal{X}$. Under this probability assignment, we have that

$$I(P_X, \mathbf{h}) = \frac{1}{N} \sum_{n=1}^{N} J_{\mathcal{X}}(\gamma_n \rho) \tag{9}$$

---

[1]Notice that for short block length L it is possible to find codes with WER smaller than $\epsilon(\rho)$ given in (8). However, in the limit of large L and fixed coding rate R, no code has error probability smaller than $\epsilon(\rho)$. A lower bound to the WER of any code for any finite length L is provided by Fano Inequality and reads [10]:

$$P_e(\rho) \ge \inf_{P_X} \mathbb{E} \left[\max \left\{ 1 - \frac{1}{R} I(P_X, \mathbf{h}) - \frac{1}{RNL}, 0 \right\} \right]$$

that converges to $\epsilon(\rho)$ as $L \to \infty$.

where

$$J_{\mathcal{X}}(s) \triangleq M - 2^{-M} \sum_{x \in \mathcal{X}} \mathbb{E} \left[\log_2 \sum_{x' \in \mathcal{X}} e^{-|\sqrt{s}(x-x')+Z|^2+|Z|^2} \right] \tag{10}$$

is the mutual information of an AWGN channel with input $X \sim \text{Uniform}(\mathcal{X})$ and SNR s (expectation in (10) is with respect to $Z \sim \mathcal{N}_{\mathbb{C}}(0,1)$).

We define the BICM channel associated to the original block-fading channel by including the mapper μ, the modulator $\mathcal{X}$ and the BICM-ML posterior log-probability ratio computer (5) as part of the channel and not as a part of a (suboptimal) encoder and decoder. Following [11], the associated BICM channel can be modeled as set of M binary-input symmetric-output channels, where the input and output of the m-th channel over the n-th fading block are given by $\{c_{n,k,m} : k = 1, \ldots, L\}$ and $\{\mathcal{L}_{n,k,m} : k = 1, \ldots, L\}$, respectively. The resulting mutual information is given by

$$J_{\mathcal{X},\text{BICM}}(s) \triangleq M - 2^{-M} \sum_{m=1}^{M} \sum_{a=0}^{1} \sum_{x \in \mathcal{X}_a^m} \mathbb{E} \left[\log_2 \frac{\sum\limits_{x' \in \mathcal{X}} e^{-|\sqrt{s}(x-x')+Z|^2}}{\sum\limits_{x' \in \mathcal{X}_a^m} e^{-|\sqrt{s}(x-x')+Z|^2}} \right] \tag{11}$$

Notice that the expectation over $Z \sim \mathcal{N}_{\mathbb{C}}(0,1)$ in (10) and (11) can be easily evaluated by using the Gauss-Hermite quadrature rules which are tabulated in [20] and can be computed using for example the algorithms described in [21].

The information outage probabilities of the block-fading channel with i.i.d. input $X \sim \mathcal{N}_{\mathbb{C}}(0,1)$, $X \sim \text{Uniform}(\mathcal{X})$ and that of the associated BICM channel are denoted by[2] $P_{\text{out}}^{G}(\rho, R)$, $P_{\text{out}}^{\mathcal{X}}(\rho, R)$ and by $P_{\text{out}}^{\mathcal{X},\text{BICM}}(\rho, R)$, respectively. From the data processing inequality and the fact that the proper complex Gaussian distribution maximizes differential entropy [22], we obtain that

$$P_{\text{out}}^{G}(\rho, R) \leq P_{\text{out}}^{\mathcal{X}}(\rho, R) \leq P_{\text{out}}^{\mathcal{X},\text{BICM}}(\rho, R) \tag{12}$$

for all R and ρ.

By evaluating the outage probability for a given signal set $\mathcal{X}$ we can assess the performance loss incurred by the suboptimal coded modulation ensemble $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$. Furthermore, by evaluating the outage probability of the BICM channel, we can assess the performance loss incurred by the suboptimal BICM-ML decoder with respect to the ML decoder.

---

[2]It is straightforward to show that with i.i.d. input $X \sim \mathcal{N}_{\mathbb{C}}(0,1)$, $I(P_X, \mathbf{h}) = \frac{1}{N} \sum_{n=1}^{N} \log(1 + \gamma_n \rho)$.

For the sake of simplicity, we consider independent Rayleigh fading, i.e., the fading coefficients h_n are i.i.d., $\sim \mathcal{N}_{\mathbb{C}}(0,1)$ and the fading power gains γ_n are Chi-squared with two degrees of freedom, i.e., $\gamma_n \sim f_\gamma(z) = e^{-z}\mathbb{1}\{z \geq 0\}$, where $\mathbb{1}\{\mathcal{E}\}$ denotes the indicator function of the event $\mathcal{E}$. This assumption will be discussed and relaxed at the end of this section.

We are interested in the SNR reliability function (1) of the block-fading channel. Lemma 1 below, that follows as a corollary of the analysis in [10], yields the SNR reliability function subject to the average input power constraint.

Lemma 1 *Consider the block-fading channel (2) with i.i.d. Rayleigh fading, under the average input power constraint $\frac{1}{NL}\mathbb{E}[\|\mathbf{X}\|_F^2] \leq 1$. The SNR reliability function for any block length $L \geq 1$ and fixed rate R is given by $d^\star(R) = N$ and it is achieved by Gaussian random codes, i.e., the random coding SNR exponent $d_G^{(r)}(R)$ of the Gaussian i.i.d. ensemble for any $L \geq 1$ is also equal to N.*

Proof. Although Lemma 1 follows as a corollary of [10, Th. 2], we provide its proof explicitly for the sake of completeness and because it is instructive to illustrate the proof technique used for the following Theorem 1.

In passing, we notice that the proof of Lemma 1 deals with the more general case of coding schemes with rate increasing with SNR as $R(\rho) = r \log \rho$, where $r \in [0,1]$, and shows that[3] $d^\star(r) = N(1-r)$ and this optimal SNR exponent can be achieved by coding schemes of any block length $L \geq 1$. The details are given in Appendix A.   □

For the considered coded modulation ensemble, we have the following result:

Theorem 1 *Consider the block-fading channel (2) with i.i.d. Rayleigh fading and input signal set $\mathcal{X}$ of cardinality 2^M. The SNR reliability function of the channel is upperbounded by the Singleton bound*

$$d_{\mathcal{X}}^\star(R) \leq \delta(N,M,R) \triangleq 1 + \left\lfloor N\left(1 - \frac{R}{M}\right)\right\rfloor \tag{13}$$

The random coding SNR exponent of the random coded modulation ensemble $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ defined

---

[3]The exponential equality and inequalities notation $\doteq$, $\dotgeq$ and $\dotleq$ were introduced in [10]. We write $f(z) \doteq z^d$ to indicate that $\lim_{z\to\infty} \frac{\log f(z)}{\log z} = d$. $\dotgeq$ and $\dotleq$ are used similarly.

previously, with block length $L(\rho)$ satisfying $\lim_{\rho \to \infty} \frac{L(\rho)}{\log \rho} = \beta$ and rate R, is lowerbounded by

$$d_{\mathcal{X}}^{(r)}(R) \geq \begin{cases} \beta N M \log(2) \left(1 - \frac{R}{M}\right), & 0 \leq \beta < \frac{1}{M \log(2)} \\ \delta(N, M, R) - 1 + \min\left\{1, \beta M \log(2) \left[N\left(1 - \frac{R}{M}\right) - \delta(N, M, R) + 1\right]\right\}, & \frac{1}{M \log(2)} \leq \beta < \infty. \end{cases}$$
$$(14)$$

Furthermore, the SNR random coding exponent of the associated BICM channel satisfies the same lower bounds (14).

Proof. See Appendix B. □

An immediate consequence of Theorem 1 is the following

Corollary 1 *The SNR reliability function of the block-fading channel with input $\mathcal{X}$ and of the associated BICM channel is given by $d_{\mathcal{X}}^{\star}(R) = \delta(N, M, R)$ for all $R \in (0, M]$, except for the N discontinuity points of $\delta(N, M, R)$, i.e., for the values of R for which $N(1 - R/M)$ is an integer.*

Proof. We let $\beta \to \infty$ in the random coding lower bound (14) and we obtain

$$\delta(N, M, R) \geq d_{\mathcal{X}}^{\star}(R) \geq d_{\mathcal{X}}^{(r)}(R) \geq \left\lceil N\left(1 - \frac{R}{M}\right)\right\rceil$$

where the rightmost term coincides with $\delta(N, M, R)$ for all points $R \in (0, M]$ where $\delta(N, M, R)$ is continuous. □

The following remarks are in order:

1. The codes achieving the optimal diversity order $d_{\mathcal{X}}^{\star}(R)$ in Theorem 1 are found in the ensemble $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ with block length that increases with SNR faster than $\log(\rho)$. This is due to the fact that, differently from the Gaussian ensemble (Lemma 1), for a given discrete signal set $\mathcal{X}$ there is a non-zero probability that two codewords are identical, for any finite length L. Hence, we have to make L increase with ρ rapidly enough such that this probability does not dominate the overall probability of error. Nevertheless, it is easy to find explicit constructions achieving the optimal Singleton bound block-diversity $\delta(N, M, R)$ for several cases of N and finite L [3, 5]. Typically, the WER of diversity-wise optimal codes behaves like $\mathcal{K}\rho^{-\delta(N,M,R)}$ for large ρ. The coefficient $\mathcal{K}$ yields a horizontal shift of the WER vs. SNR curve (in a log-log chart) with respect to the outage probability curve $P_{\text{out}}^{\mathcal{X}}(\rho, R)$ that we refer to as "gap from outage".

Codes found in previous works [3, 4, 5, 6] have a gap from outage that increases with the block length L. On the contrary, the gap from outage of the class of codes proposed in this paper is asymptotically independent of the block length. We say that a code ensemble is *good* if it achieves *vanishing* gap from outage as $L \to \infty$. We say that a code ensemble is *weakly good* if it achieves *constant* gap from outage as $L \to \infty$. In Section 4.3 we give a sufficient condition for weak goodness and argue that the proposed codes are weakly good.

2. For any given coding rate R, we can achieve "full diversity" $\delta(N, M, R) = N$ by considering a signal set large enough. In fact, by letting $M \geq NR$ we have $\delta(N, M, R) = N$ for any desired rate $R < M$. This corresponds to the intuitive argument that larger and larger signal sets approach better and better Gaussian codes.[4]

3. We can relax the assumption of Rayleigh fading by noticing that in the proofs of Lemma 1 and Theorem 1 only the near-zero behavior of the fading power gain distribution has a role. For Rayleigh fading, we have $\Pr(\gamma_n \leq \epsilon) \approx \epsilon$, for small $\epsilon > 0$. Hence, the above results hold for all block-fading channels with i.i.d. fading with power gain distribution with this behavior. More in general, as argued in [10], for a fading distribution with near-zero behavior $\Pr(\gamma_n \leq \epsilon) \approx \epsilon^D$, the SNR reliability function is given by $D\delta(N, M, R)$. For example, this is the case of independent Rayleigh fading with a D antenna receiver using D-fold maximal-ratio combining [23].

Fig. 1 shows $\delta(N, M, R)$ (Singleton bound) and the random coding lower bounds for the two cases $\beta M \log(2) = 1/2$ and $\beta M \log(2) = 2$, in the case $N = 8$ and $M = 4$ ($\mathcal{X}$ is a 16-ary signal set). It can be observed that as β increases (for fixed M), the random coding lower bound coincides over a larger and larger support with the Singleton upper bound. However, in the discontinuity points it will never coincide.

In order to illustrate the operational meaning of the above results and motivate the code construction in the following section, we show in Fig. 2 the outage probability versus SNR of the block-fading channel with i.i.d. Rayleigh fading with $N = 8$ blocks, for Gaussian inputs,

---

[4]For finite SNR, expanding the signal set without proper shaping incurs shaping loss. However, in terms of SNR exponent this effect is not seen as shaping involves only a fixed gap from outage. Using the definition introduced above, we might say that codes found in our ensemble of coded modulation schemes over larger and larger QAM complex constellations can be weakly good, but cannot be good due to the inherent shaping loss.
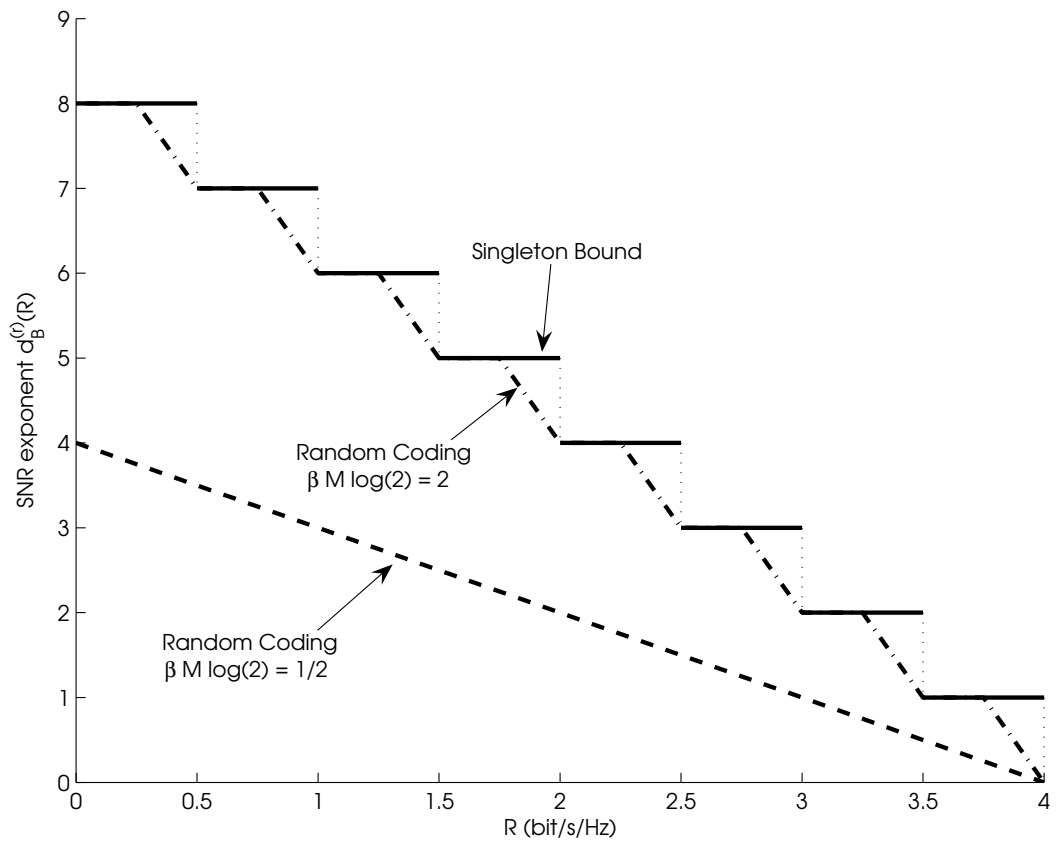
Figure 1: SNR reliability function and random coding exponents $d_{\mathcal{X}}^{(r)}(R)$ for $N = 8$ and $M = 4$.

8-PSK and 16-QAM constellations and for the associated BICM channels with Gray mapping[5] [11], with spectral efficiencies $R = 1, 1.5, 2$ bit/complex dimension. In these log-log charts, the SNR exponent determines the slope of the outage probability curve at high SNR (small outage probability). We notice that Gaussian inputs always show the steepest slope and that this is independent of R for high SNR (in agreement with Lemma 1). For $R = 1$ we observe a slight slope variation since have that $\delta(8, 3, 1) = 6$ (for 8-PSK) and that $\delta(8, 3, 1) = 7$ (for 16-QAM). The slope difference will be more apparent for larger SNR values. For $R = 1.5$, the curves also show different slopes since $\delta(8, 3, 1.5) = 5$ (for 8-PSK) while $\delta(8, 4, 1.5) = 6$ (for 16-QAM). This effect is even more evident for $R = 2$, where $\delta(8, 3, 2) = 4$ (for 8-PSK) and $\delta(8, 4, 2) = 5$ (for 16-QAM). Notice also that, in all cases, the SNR loss incurred by BICM-ML decoding is very small.

4 Blockwise concatenated coded modulation

In this section we introduce a general construction for MDS coded modulation schemes for the block-fading channel and we provide bounds and approximations to their error probability under ML and BICM-ML decoding.

So far, we have considered the ensemble $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ where $\mathcal{C}$ is a random binary linear code. In this section we consider specific ensembles where $\mathcal{C}$ has some structure. In particular, $\mathcal{C}$ belongs to the well-known and vast family of turbo-like codes (parallel and serially concatenated codes, repeat-accumulate codes, etc..) and it is obtained by concatenating linear binary *encoders* through interleavers. Hence, we shall considered the structured random coding ensemble where the component encoders for $\mathcal{C}$ are fixed and the interleavers are randomly selected with uniform probability over all possible permutations of given length. For the sake of notation simplicity, we keep using the notation $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ for any of such ensembles with given component encoders, where now the symbol $\mathcal{C}$ is a placeholder indicating the *set of component encoders* defining the concatenated code.

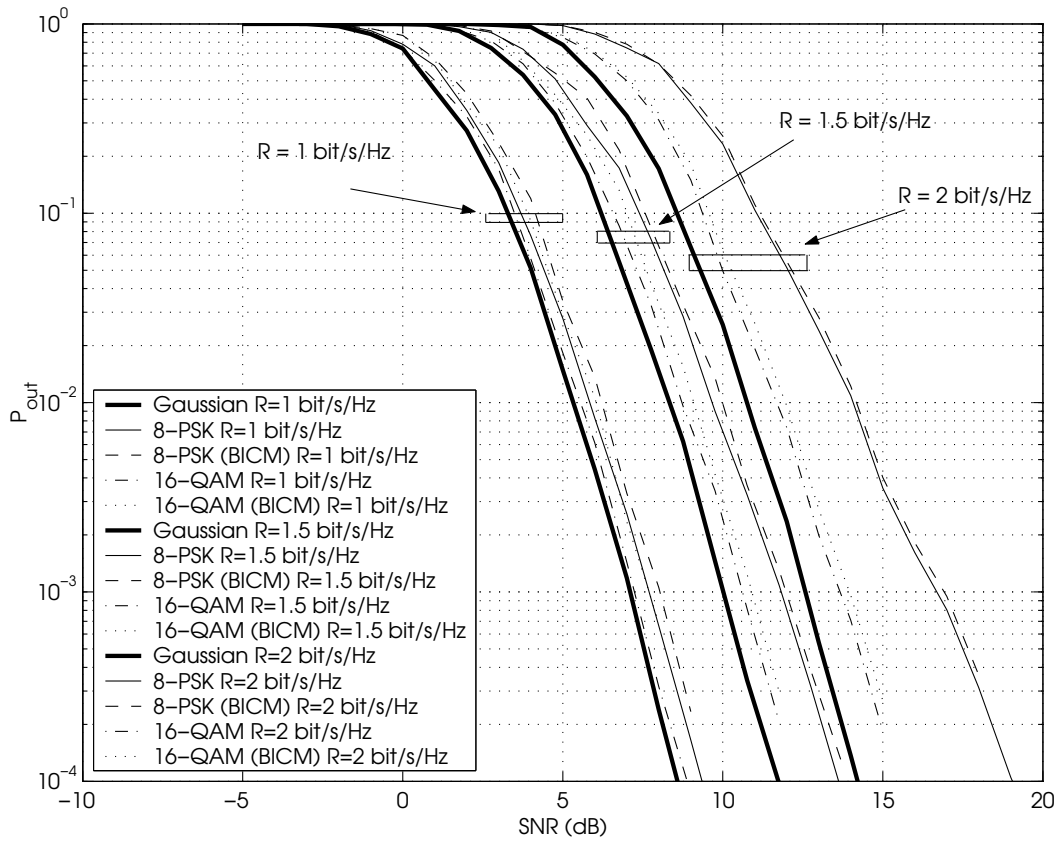[5]All BICM schemes considered in this work make use of Gray mapping.

Figure 2: Outage probability for $N = 8$, $R = 1, 1.5, 2$ bit/complex dimension, Gaussian inputs, 8-PSK and 16-QAM modulations. Thick solid lines correspond to Gaussian inputs, thin solid lines to 8-PSK, dashed lines to 8-PSK with BICM, dashed-dotted lines to 16-QAM and dotted lines to 16-QAM with BICM.

4.1 Code construction

Fig. 3 shows the proposed encoder structure that we refer to as *Blockwise Concatenated Coding* (BCC). The binary linear code is formed by the concatenation of a binary linear encoder $\mathcal{C}^O$ of rate r_O, whose output is partitioned into N blocks. The blocks are separately interleaved by the permutations $(\pi_1, \ldots, \pi_N)$ and the result is fed into N inner encoders $\mathcal{C}^I$ of rate r_I. Finally, the output of each inner encoder is mapped onto a sequence of signals in $\mathcal{X}$ by the one-to-one symbol mapping μ so that the rate of the resulting blockwise concatenated code is $R = r_O r_I M$.

We denote by K the information block length, i.e., K information bits enter the outer encoder. Correspondingly, the length of each outer output block is $L_\pi = K/(N r_O)$ and the length of the inner-encoded blocks is $L_B = L_\pi/r_I$ binary symbols. Eventually, the length of the blocks sent to the channel is $L = L_B/M$ modulation symbols (complex dimensions). Without loss of essential generality, we assume that L_π and L_B defined above are integers.

The codes considered in this work make use of bit-interleaving between the inner encoder and the mapper [11], denoted in Fig. 3 by the permutations $(\pi_1^\mu, \ldots, \pi_N^\mu)$. However, we hasten to say that mapping through interleavers is not necessary for the construction and more general mappings could be envisaged. In any case, since interleavers and inner encoding are performed on a blockwise basis, the block diversity of the concatenated code coincides with the block diversity of the outer code.

It is worthwhile to point out some special cases of the BCC construction. When $\mathcal{C}^O$ is a convolutional encoder and $\mathcal{C}^I$ is the trivial rate-1 identity encoder, we refer to the resulting scheme as a *blockwise partitioned Convolutional Code* (briefly, CC). Interestingly, most previously proposed codes for the block-fading channel (see [3, 4, 5, 6]) belong to this class. When the outer code is a simple repetition code of rate $r_O = 1/N$ and the inner codes are rate-one accumulators (generator $1/(1+D)$) [24], the resulting scheme is referred to as Repeat and *Blockwise* Accumulate (RBA) code. When both outer and inner codes are convolutional codes, we will refer to the resulting scheme as *blockwise* concatenated convolutional codes (BCCC).

As anticipated in the Introduction, practical decoding of BCCs resorts to BP iterative decoding algorithm over the code graph [19]. In particular, when either $\mathcal{C}^O$ or $\mathcal{C}^I$ are convolutional codes, the well-known forward-backward decoding algorithm is used over the subgraph representing the corresponding trellis [25].
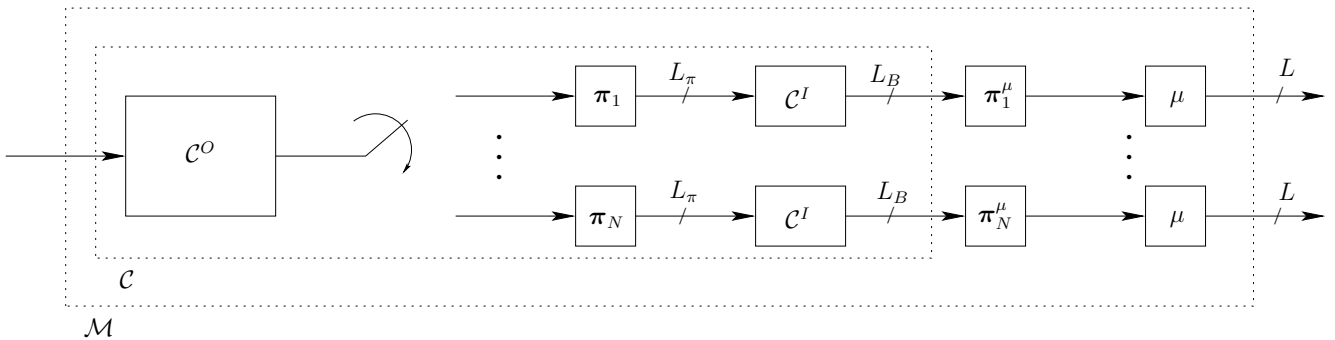
Figure 3: The general encoder for Blockwise Concatenated Coding.

Fig. 4 illustrates the effectiveness of blockwise concatenation with respect to standard turbo-like codes designed for the AWGN. In particular, we compare the WER of a binary $R = 1/2$ RBA and BCCC (with convolutional $(5, 7)_8$ outer code and inner accumulators) with that of their standard counterparts (namely, a Repeat and Accumulate (RA) code and a Serially Concatenated Convolutional Code (SCCC)), mapped over $N = 2$ fading blocks with 10 BP decoder decoding iterations. In all cases, the information block length is $K = 1024$. We observe a significant difference in the slope of the WER curve, due to the fact that blockwise concatenation preserves the block diversity d_B of the outer code while standard concatenation does not.

In order to show the generality of the proposed approach to construct MDS BCCs, Figure 5 illustrates the WER performance obtained by simulation with BP decoding of binary $r = 1/2$ BCCCs $(5, 7)_8$ and $(25, 35)_8$ both with with inner accumulators, the SCCCs with outer $(5, 7)_8$ anf $(25, 35)_8$ and inner accumulators and best known 4 and 64 states CCs [6] mapped over $N = 8$ fading blocks with block length of 1024 information bits. In this case, the Singleton bound is $\delta(N, M, R) = 5$. Notice that since the $(5, 7)_8$ code is not MDS [3, 6], the corresponding BCCC (and of course the CC itself) will show a different slope and performance degradation at high SNR. Indeed, we can appreciate a steeper slope of the BCCC with $(25, 35)_8$ and the 64 states CC since both are MDS codes. We also observe clear advantage of BCCCs over standard CCs at this block length (this point will be further discussed in depth in section 4.3). Finally, as illustrated also in the previous figure, the MDS BCCCs remarkably outperform their SCCC counterparts, which are designed for the ergodic channel.
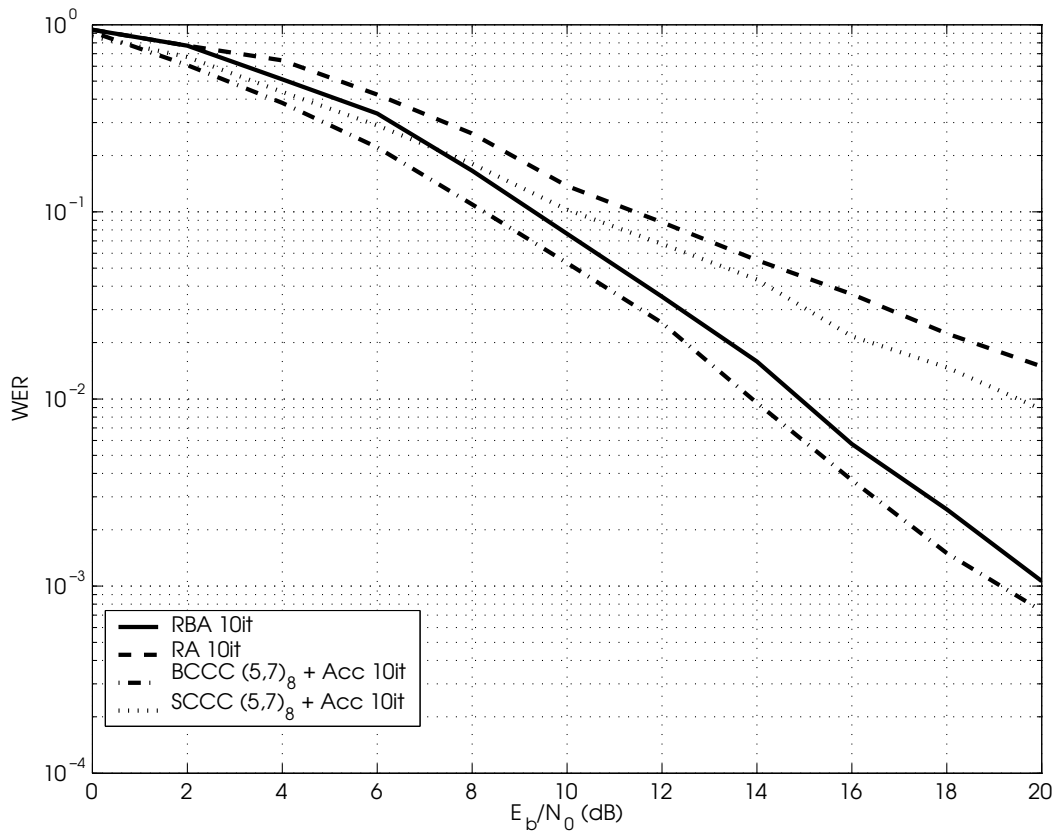
Figure 4: WER obtained by BP decoding (simulation with 10 iterations) of binary RBA, RA, BCCC and SCCC of rate $R = 1/2$ for $N = 2$ and $K = 1024$.
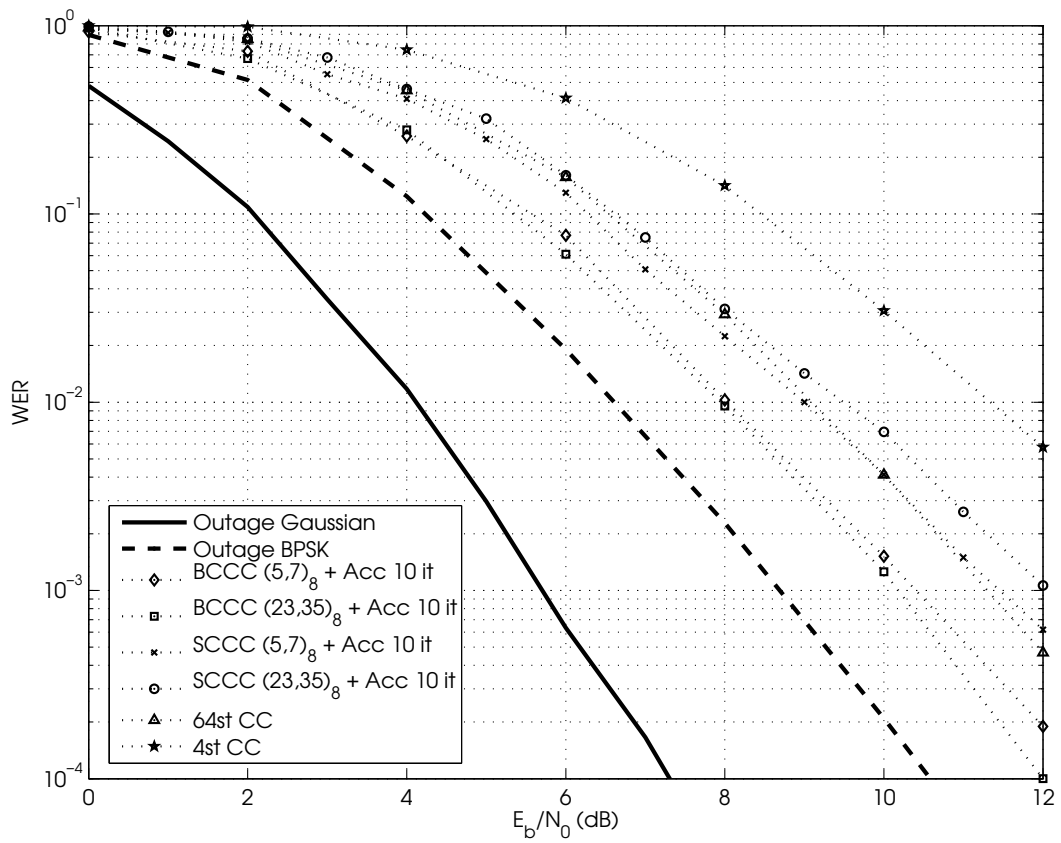
Figure 5: WER $r = 1/2$ BCCCs and CCs mapped over $N = 8$ fading blocks.

4.2 Upper bounds and approximations on ML decoding error probability

For the sake of simplicity we consider first codes over the QPSK with Gray mapping, or, equivalently, over BPSK. This case is particularly simple since the squared Euclidean distance between the constellation points is proportional to the Hamming distance between their binary labels. A tight upper bound on the WER of binary codes mapped over QPSK with Gray mapping and transmitted over N fading blocks, is given by Malkamaki and Leib (M&L) in [5], and reads

$$P_e(\rho) \leq \mathbb{E}\left[\min\left\{1, \sum_{w_1,\ldots,w_N} A_{w_1,\ldots,w_N} Q\left(\sqrt{\kappa\rho\sum_{n=1}^{N}\gamma_n w_n}\right)\right\}\right] \tag{15}$$

where $A_{w_1,\ldots,w_N}$ is the Multivariate Weight Enumeration Function (MWEF) of $\mathcal{C}$ [26] which accounts for the number of pairwise error events with output Hamming weights per block $w_1,\ldots,w_N$, $\kappa = 2$ for BPSK and $\kappa = 1$ for QPSK, and

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{t^2}{2}}dt \tag{16}$$

is the Gaussian tail function. Expectation in (15) is with respect to the fading power gains $(\gamma_1,\ldots,\gamma_N)$. In order to compute (15), we need to compute a multivariate expectation that does not break into the individual expectation of each term in the union bound because of the $\min\{1,\cdot\}$. Hence, in practice, we have to resort to Monte Carlo methods.

In [27], Byun, Park and Lee presented a simpler upper bound to (15) in the context of ML decoding of trellis space-time codes. Unfortunately, the bound in [27] upperbounds (15) only if the sum over $w_1,\ldots,w_N$ contains a single term. Nevertheless, we shall demonstrate through several examples that this technique, referred to as the BPL *approximation*, if applied to full diversity codes (i.e., codes with blockwise Hamming distance $d_B = N$) yields a very good approximation of the WER, with the advantage that it is much easier to compute than the M&L bound.

Assuming $d_B = N$, which implies that $\min w_n > 0$ for all $n = 1,\ldots,N$, the BPL approximation takes on the form

$$P_e(\rho) \lesssim \mathbb{E}\left[\min\left\{1, \sum_{\Delta_p} A_{\Delta_p} Q\left(\sqrt{\kappa\rho\Delta_p^{1/N}\sum_{n=1}^{N}\gamma_n}\right)\right\}\right] \tag{17}$$

where $\Delta_p \triangleq \prod_{n=1}^{N} w_n$ is the product weight and A_{Δ_p} is the Product Weight Enumeration Function (PWEF) of $\mathcal{C}$, i.e., the number of codewords of $\mathcal{C}$ with product weight Δ_p. By noticing that

$\gamma = \sum_{n=1}^{N} \gamma_n$ is central chi-squared with $2N$ degrees of freedom and mean N, (17) becomes

$$P_e(\rho) \lessapprox \int_0^{+\infty} \min\left\{1, \sum_{\Delta_p} A_{\Delta_p} Q\left(\sqrt{\kappa \rho \Delta_p^{1/N} z}\right)\right\} f_\gamma(z) dz \tag{18}$$

where

$$f_\gamma(z) = \frac{z^{N-1}}{(N-1)!} e^{-z} \tag{19}$$

is the pdf of γ. In this way, only product weights have to be enumerated and the computation of (18) requires just a one-dimensional integration, that is easily computed numerically.

Union bound-based techniques are known to be loose for turbo codes and other capacity-approaching code ensembles such as LDPC and RA codes over the AWGN channel. As a matter of fact, improved bounding techniques are needed in order to obtain meaningful upper bounds in the SNR range between the capacity threshold and the cut-off rate threshold [14, 15, 16, 17, 18]. Among those, the tangential-sphere bound (TSB) is known to be the tightest. The TSB can be simply extended to the block-fading channel for each fixed realization of the fading vector $\mathbf{h}$ (for more details see [28, 29]). Then, an outer Monte Carlo average over the fading is required. Since the TSB requires the optimization of certain parameters for each new fading realization, the computation of the TSB is very intensive. A slight simplification is obtained by applying the TSB technique to the PWEF, as in the BPL approximation. The resulting approximation (referred to as BPL-TSB) requires only a single variate expectation.

The following examples illustrate the bounds and the approximations described above for BPSK and QPSK with Gray mapping. The MWEF and PWEF are obtained as described in Appendix C. In particular, Fig. 6 compares the simulation (with 10 BP decoder iterations) with the ML bounds and approximations for RBA codes of $R = 1/2$ with information block length $K = 256$, over $N = 2$ fading blocks. The expectation in the M&L bound and in the TSB are computed by Monte Carlo. We observe an excellent matching between the performance of BP decoding and the bounds on ML decoding, even for such short block lengths, in contrast to the AWGN case. We also notice that the TSB is only marginally tighter than the M&L bound and, due to its high computational complexity, it is useless in this context. The BPL approximation predicts almost exactly the WER of the RBA code for all block lengths. Based on such examples (and on very extensive numerical experiments not reported here for the sake of space limitation) we conclude that the performance of BCCs on block-fading channels can be predicted very accurately
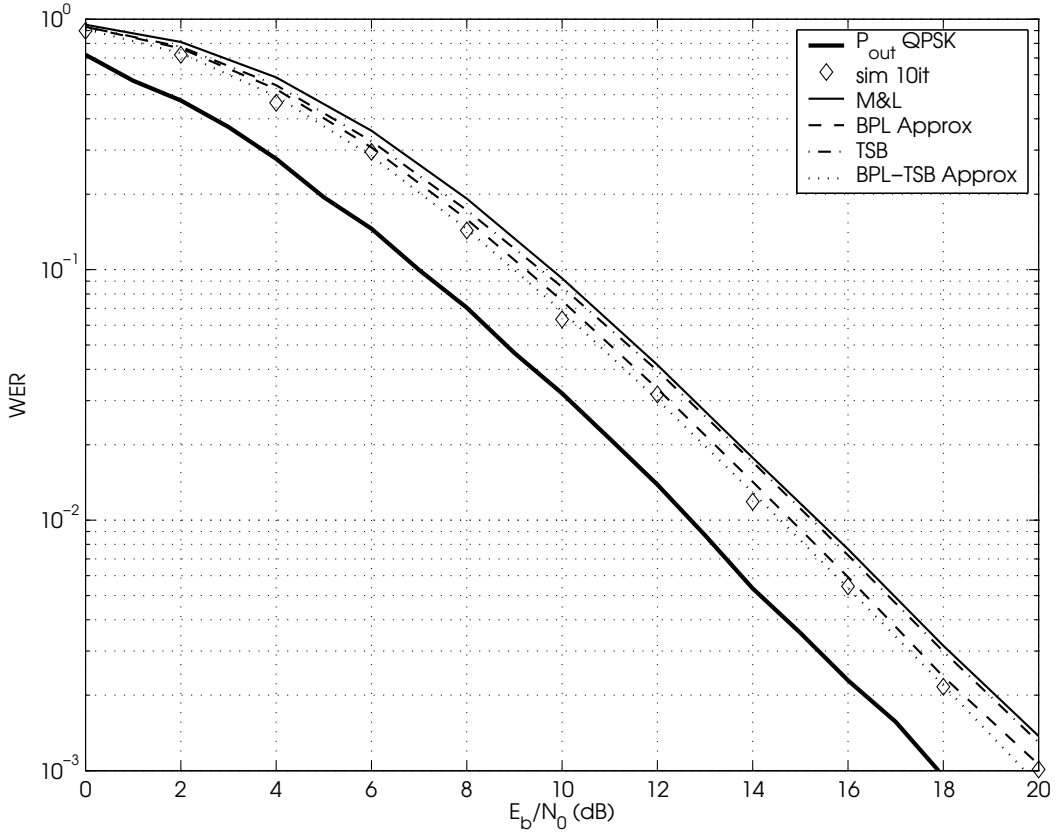
by simple ML analysis techniques.

Figure 6: WER obtained by BP decoding simulation with 10 iterations and ML bounds and approximations for binary RBA of $R = 1/2$ and $K = 256$ over $N = 2$ blocks.

For general signal sets $\mathcal{X}$ and modulator mappings μ the above bounds are no longer valid since the squared Euclidean distance between signals depends, in general, on the individual labels and not only on the labels' Hamming distance. Assuming bit-interleaving between the inner binary codes and the modulator mapping, we can make use of the BICM Bhattacharyya union bound developed in [11], combined with the "limit before average" approach of [5]. We obtain

$$P_e(\rho) \leq \mathbb{E}\left[\min\left\{1, \sum_{w_1,\ldots,w_N} A_{w_1,\ldots,w_N} \frac{1}{2} \prod_{n=1}^{N} B_n(\rho, \mu, \mathcal{X})^{w_n}\right\}\right] \tag{20}$$

where

$$B_n(\rho, \mu, \mathcal{X}) \triangleq \frac{2^{-M}}{M} \sum_{m=1}^{M} \sum_{a=0}^{1} \sum_{x \in \mathcal{X}_a^m} \mathbb{E}\left[\sqrt{\frac{\sum_{x' \in \mathcal{X}_{\bar{a}}^m} e^{-|\sqrt{\rho \gamma_n}(x-x')+Z|^2}}{\sum_{x' \in \mathcal{X}_a^m} e^{-|\sqrt{\rho \gamma_n}(x-x')+Z|^2}}}\right] \tag{21}$$

21

is the Bhattacharyya factor of the BICM channel associated to the n-th fading block, with SNR $\gamma_n \rho$.

The bound (20) holds under the assumption that the mapping μ is symmetrized, as explained in [11], i.e., that a random i.i.d. scrambling sequence, known both to the transmitter and to the receiver, chooses at every symbol with probability $1/2$ either the mapping μ or its complement $\bar{\mu}$, obtained by complementing each bit in the labels of μ.[6] The factor $1/2$ in front of the Bhattacharyya union bound follows from the fact that, under the symmetrized mapping assumption, the associated BICM channel with inputs $c_{n,k,m}$ and outputs $\mathcal{L}_{n,k,m}$ defined in (5) is binary-input output-symmetric (see [30]). The expectation in (21) can be efficiently computed by Gauss-Hermite quadratures.

As shown in [31], the tail of the pdf of the bit-wise posterior log-probability ratio (5) at the output of the associated BICM channel is very close to the corresponding output of a binary-input AWGN channel with fading power gain

$$\zeta_n = -\frac{1}{\rho} \log B_n(\rho, \mu, \mathcal{X}) \tag{22}$$

Moreover, for given fading gain γ_n we have [31]

$$\lim_{\rho \to \infty} \zeta_n = \frac{d_{\min}^2}{4} \gamma_n. \tag{23}$$

independently of the mapping μ. Under this Gaussian approximation, we obtain

$$P_e(\rho) \lessapprox \mathbb{E}\left[\min\left\{ 1, \sum_{w_1,\ldots,w_N} A_{w_1,\ldots,w_N} Q\left(\sqrt{2\rho \sum_{n=1}^{N} w_n \zeta_n} \right) \right\} \right], \tag{24}$$

and the corresponding BPL approximation (for full diversity codes)

$$P_e(\rho) \lessapprox \mathbb{E}\left[\min\left\{ 1, \sum_{\Delta_p} A_{\Delta_p} Q\left(\sqrt{2\rho \Delta_p^{1/N} \sum_{n=1}^{N} \zeta_n} \right) \right\} \right]. \tag{25}$$

Unfortunately, in this case $\sum_{n=1}^{N} \zeta_n$ is no longer chi-squared distributed (from (23) it follows that it is chi-squared in the limit of high SNR). Therefore, (25) has to be computed via a Monte Carlo

---

[6] If the mapping μ and the constellation $\mathcal{X}$ are such that, for all label positions $m = 1,\ldots, M$, the log-probability ratio defined in (5) is symmetrically distributed, that is, $p_{\mathcal{L}_{n,k,m}}(z|c_{n,k,m} = a) = p_{\mathcal{L}_{n,k,m}}(-z|c_{n,k,m} = \bar{a})$, then the scrambling assumption is not needed.

average, reducing only slightly the computational burden with respect to (24). We will refer to (20) as the M&L-Bhattacharyya bound and to (24) as the M&L-GA.

We hasten to say that, although the proposed methods are *just* approximation, they represent so far the only alternative to extensive simulation. Indeed, they might be regarded as the analogous for the block-fading channel to the EXIT chart "analysis" commonly used for fully-interleaved fading channels and AWGN channels: they are both based on approximating a complicated binary-input output-symmetric channel by a binary-input AWGN channel, "matched" in some sense to the former.

In Fig. 7 we show the WER (obtained by simulation with 10 BP decoder iterations) and the various upper bounds and approximations on ML decoding error probability described above, for a RBA code of rate $r = 1/2$ over $N = 2$ fading blocks and information block length $K = 256$, with 8-PSK and 16-QAM (the corresponding spectral efficiencies are $R = 1.5$ and 2 bit/complex dimension). We show the BICM outage probability for 8-PSK and 16-QAM for the sake of comparison. Again, we observe an excellent match between simulation with BP decoding and ML approximations, for all modulations. We also observe that the BICM Bhattacharyya bound is looser than the Gaussian Approximation (24).

4.3 Weak goodness of BCC ensembles

As introduced in Section 3, we say that a code ensemble over $\mathcal{X}$ is good if, for block length $L \to \infty$, its WER converges to the outage probability $P_{\text{out}}^{\mathcal{X}}(\rho, R)$. We say that a code ensemble over $\mathcal{X}$ is weakly good if, for block length $L \to \infty$, its WER shows a fixed SNR gap to outage probability, asymptotically independent of L. In this section we give an explicit sufficient condition for weak goodness in terms of the asymptotic exponential growth rate function [32] of the multivariate weight enumerator of specific ensembles.

The issue of weak goodness is non-trivial, as illustrated by the following argument. A code ensemble $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ such that, for all sufficiently large L, a randomly generated member in the ensemble attains the Singleton bound with probability 1 is a good candidate for weak goodness. However, this condition is neither necessary nor sufficient. For example, the ensemble $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ considered in Theorem 1 has a small but non-zero probability that a randomly selected member is not blockwise MDS, nevertheless it attains the optimal SNR exponent provided that L grows faster
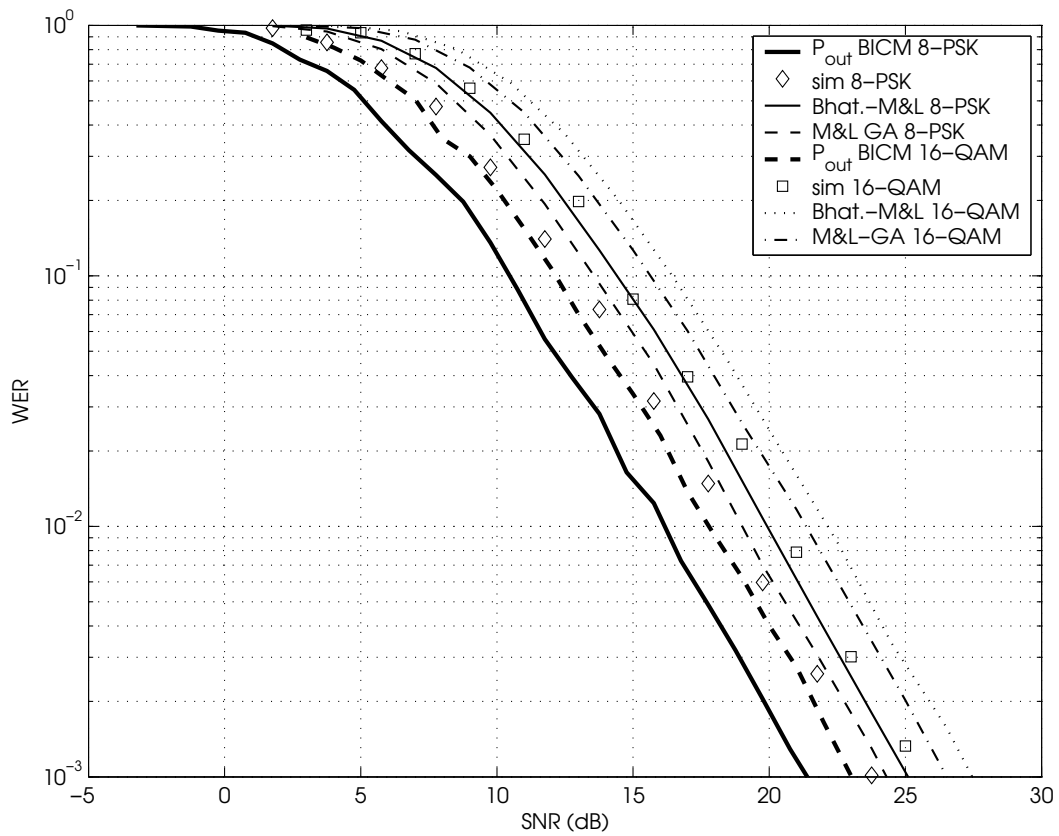
Figure 7: WER obtained by BP decoding simulation with 10 iterations and ML bounds and approximations for RBA with BICM of $r = 1/2$ over $N = 2$ blocks with 8-PSK and 16-QAM.

than $\log \rho$, and hence it is weakly good. On the contrary, the ensemble of random BCCs with given outer and non-trivial inner encoders and the ensemble of blockwise partitioned CCs (i.e., BCCs with convolutional outer encoder and rate-1 identity encoder considered in [3, 4, 5, 6]) that can be seen as BCCs with convolutional outer encoder and trivial (identity) inner encoder, attain the Singleton bound with probability 1 provided that the outer code is blockwise MDS. Nevertheless, simulations show that while the WER of general BCCs with recursive inner encoder is almost independent of the block length, the WER of CCs grows with the block length. For example, Fig.8 shows the WER for fixed SNR versus the information block length K, for the ensemble of $R = 1/4$ RBA codes and the standard 64-states CCs with generators $(135, 135, 147, 163)_8$ mapped over $N = 4$ blocks, and of $r = 1/2$ BCCs (with outer convolutional encoder $(5, 7)_8$ and inner accumulators) and the 64-states CCs mapped over $N = 8$ blocks optimized in [6] with generators $(103, 147)_8$ for the block-fading channel. The different behavior of the WER as a function of the block length for the two ensembles is evident.

We focus first on codes over the BPSK modulation. Therefore, in this case $L = L_B$. Let $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_N) \in [0, 1]^N$ be the vector of normalized Hamming weights per block. The asymptotic exponential growth rate function [32] of the multivariate weight enumerator is defined by

$$a(\boldsymbol{\omega}) \triangleq \lim_{\epsilon \to 0} \lim_{L \to \infty} \frac{1}{L} \log \left| \mathcal{S}_\epsilon^L(\boldsymbol{\omega}) \right| \tag{26}$$

where $\mathcal{S}_\epsilon^L(\boldsymbol{\omega})$ is the set of codewords in the length-L ensemble with Hamming weights per block satisfying

$$|w_n/L - \omega_n| \leq \epsilon, \quad n = 1, \ldots, N \tag{27}$$

We have the following results:

Theorem 2 *Consider an ensemble of codes $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ of rate R, where $\mathcal{X}$ is BPSK, over a block-fading channel with N blocks. Let $a(\boldsymbol{\omega})$ be the asymptotic exponential growth rate function of the ensemble multivariate weight enumerator. For $1 \leq k \leq N$, let $\mathcal{W}(N, k) \in \mathbb{F}_2^N$ denote the set of binary vectors with Hamming weight not smaller than $N - k + 1$ and define $\widehat{s}$ to be the infimum of all $s \geq 0$ such that*

$$\inf_{\mathbf{x} \in \mathcal{W}(N, \delta(N, M, R))} \inf_{\boldsymbol{\omega} \in [0, 1]^N} \left\{ s \sum_{n=1}^N x_n \omega_n - a(\boldsymbol{\omega}) \right\} > 0 \tag{28}$$

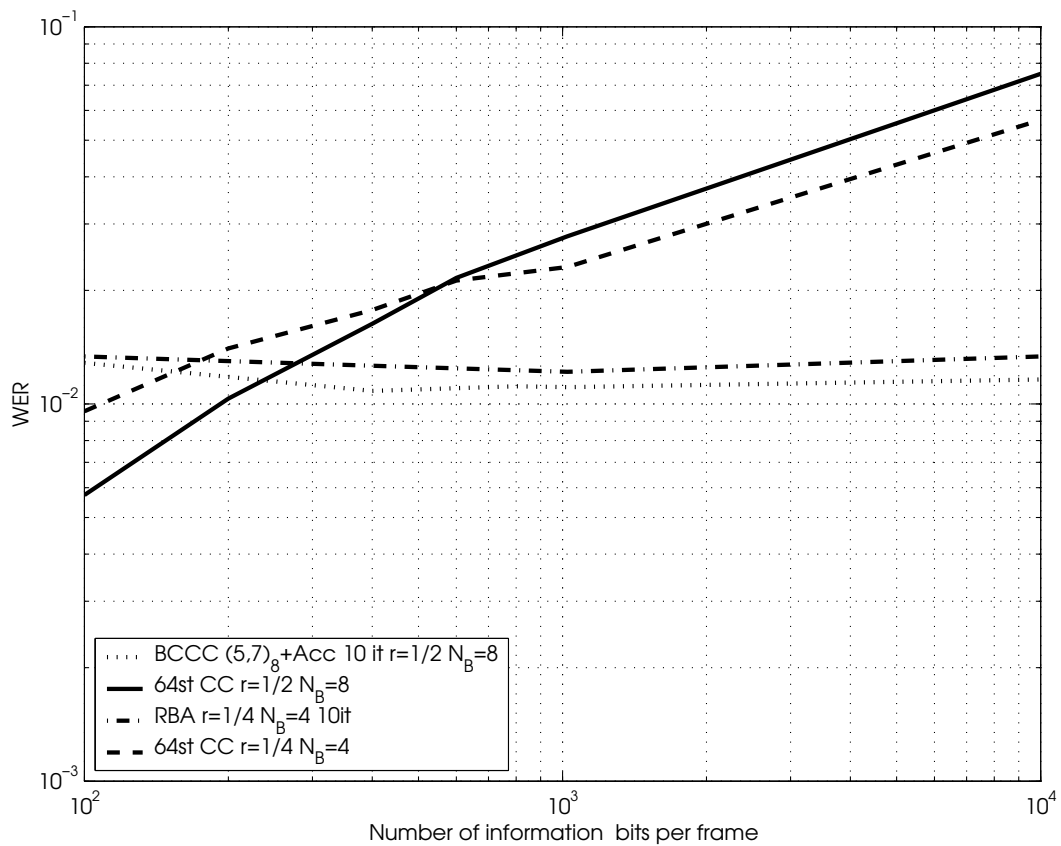*If $\widehat{s} < \infty$, then the code ensemble is weakly good.*

Figure 8: WER vs. information block length at $E_b/N_0 = 8$dB for binary BCC, RBA and trellis terminated CCs obtained by simulation (10 BP decoding iterations for the BCCs and ML Viterbi decoding for the CCs).

Proof.   See Appendix D.   □

As far as higher order coded modulations are concerned, we have the following

Corollary 2 *Consider an ensemble of codes* $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ *of rate R, where* $\mathcal{X}$ *is a complex signal set of size* 2^M, *over a block-fading channel with N blocks, where modulation is obtained by (random) bit-interleaving and decoding by the BICM-ML decoder defined by (6). If the underlying ensemble of binary codes (i.e., mapping the binary symbols of* $\mathcal{C}$ *directly onto BPSK) is weakly good, then the ensemble* $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ *is weakly good.*

Proof.   See Appendix D.   □

The above results (and the proofs of Appendix D) reveal that the error probability of weakly good codes in the regime where both the block length and the SNR are large is dominated by the event that more than $\delta(N, M, R)$ fading components are *small* (in the sense of the proof of Theorem 2). On the contrary, when less than $\delta(N, M, R)$ fading components are small, the code projected over the significant fading components has a finite ML decoding threshold (with probability 1). Therefore, for large SNR, its error probability vanishes for all such fading realizations. Apart from a gap in SNR, this is the same behavior of the information outage probability for rate R and discrete signal set $\mathcal{X}$. This observation provides a partial explanation of the striking fact that, differently from the case of AWGN or fully interleaved fading, in block fading the error probability under BP decoding is closely approximated by the analysis of the ML decoder. In fact, we argue that the two regimes of more or less than $\delta(N, M, R)$ small fading components dominate the average error probability, while the detailed behavior of the decoder in the transition region between these two extremes is not very important, provided that the probability that a channel realization hits the transition region is small, i.e., that the transition is sufficiently sharp. The sharper and sharper transition between the below-threshold and above-threshold regimes of random-like concatenated codes of increasing block length is referred to as *interleaving gain* in [33, 34]. We argue that weak goodness of BCCs in block-fading channels is another manifestation of the interleaving gain, even if for such channel no waterfall behavior is observed.

In Appendix D we show also that the ensemble of trellis terminated CCs of increasing block length considered in [3, 4, 5, 6] does not satisfy the condition of Theorem 2. Numerical verification

of Theorem 2 is needed for a specific code ensemble. In particular, one has to show that

$$\sup_{\mathbf{x} \in \mathcal{W}(N, \delta(N, M, R))} \sup_{\boldsymbol{\omega} \in [0,1]^N} \frac{a(\boldsymbol{\omega})}{\sum_{n=1}^{N} x_n \omega_n} < \infty \tag{29}$$

Supported by the simulations in Figs. 8, 9 and 10 and by the case of RBAs, where explicit calculation of the multivariate weight enumerator is possible (see Appendix C), we conjecture that (29) holds for the family of random BCCs with MDS outer code and inner recursive encoders.

As an example, in Fig. 9 we show the asymptotic WER for the RBA ensemble of rate $1/2$ with BPSK modulation, over a channel with $N = 2$ fading blocks. The asymptotic WER is computed via the asymptotic Bhattacharyya M&L bound given by

$$P_e(\rho) \leq \Pr\left(\max_{\boldsymbol{\omega} \in [0,1]^N} \frac{a(\boldsymbol{\omega})}{\sum_{n=1}^{N} \omega_n \gamma_n} \geq \rho\right) \tag{30}$$

as motivated in Appendix D. Simulations (BP iterative decoder) for information block lengths $K = 100, 1000$ and 10000 are shown for comparison. This figure clearly shows that the WER of these codes becomes quickly independent of the block length and shows fixed gap from the outage probability.

In order to illustrate the weak goodness of BCCs with BICM and high-order modulations, Fig. 10 shows the asymptotic WER of an RBA code of rate $R = 2$ bit/complex dimension with 16-QAM modulation over $N = 2$ fading blocks. The asymptotic WER is computed via the asymptotic Bhattacharyya M&L bound given by

$$P_e(\rho) \leq \Pr\left(\max_{\boldsymbol{\omega} \in [0,1]^N} \frac{a(\boldsymbol{\omega})}{\sum_{n=1}^{N} \omega_n \zeta_n} \geq \rho\right) \tag{31}$$

as motivated in Appendix D, where ζ_n is defined in (22). Simulations (BP iterative decoder) for information block lengths $K = 100, 1000$ and 10000 are shown for comparison.

We conclude this section by pointing out an interesting fact that follows as a consequence of weak goodness and allows the accurate WER evaluation of codes with given block length by using weight enumerators of codes in the same ensemble but with much smaller block length. This observation is illustrated by Fig. 11, showing the WER and the BPL approximation for an RBA code of rate $R = 1/4$ mapped over $N = 4$ fading blocks with $K = 100$. We also show the simulation of BP decoding with 10 iterations, the BPL approximation computed by truncating the PWEF to maximum product weight $\Delta_p^{\max} = 10000$, and the PBL approximation computed for
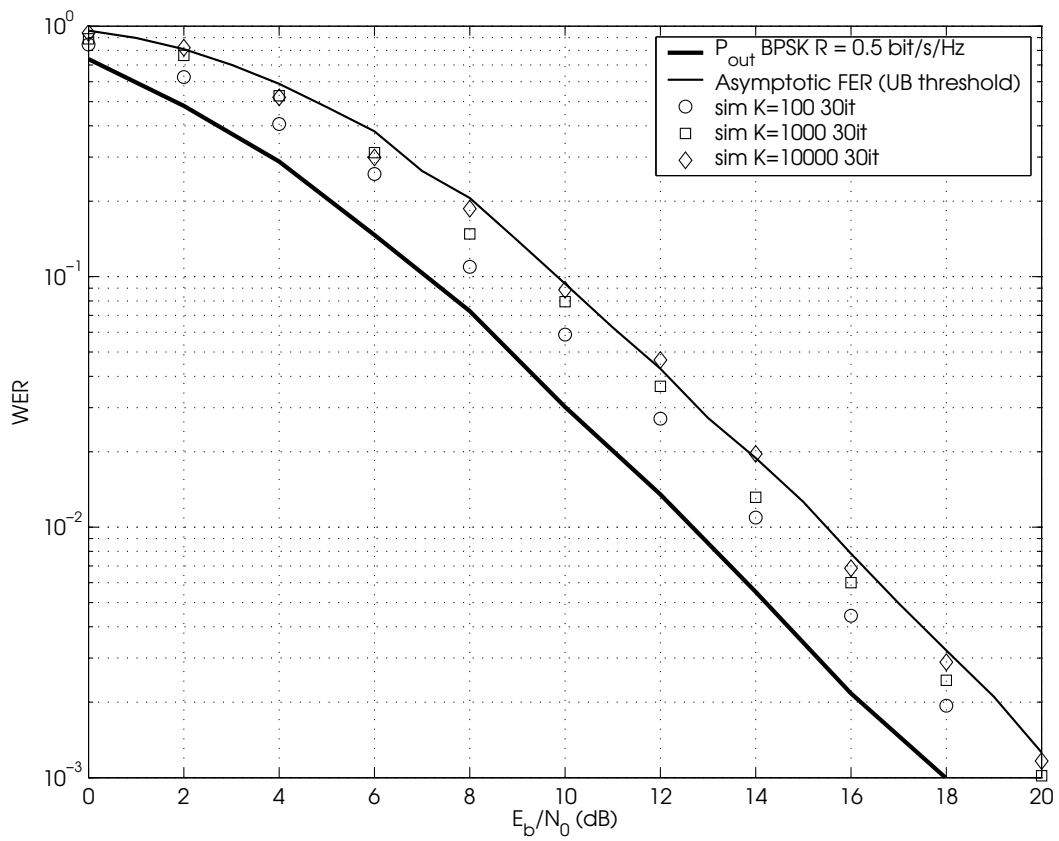
Figure 9: Asymptotic error probability (30) for a binary rate $r = 1/2$ RBA code mapped over $N = 2$ fading blocks and corresponding BP decoding simulation with 30 iterations and $K = 100, 1000$ and 10000.
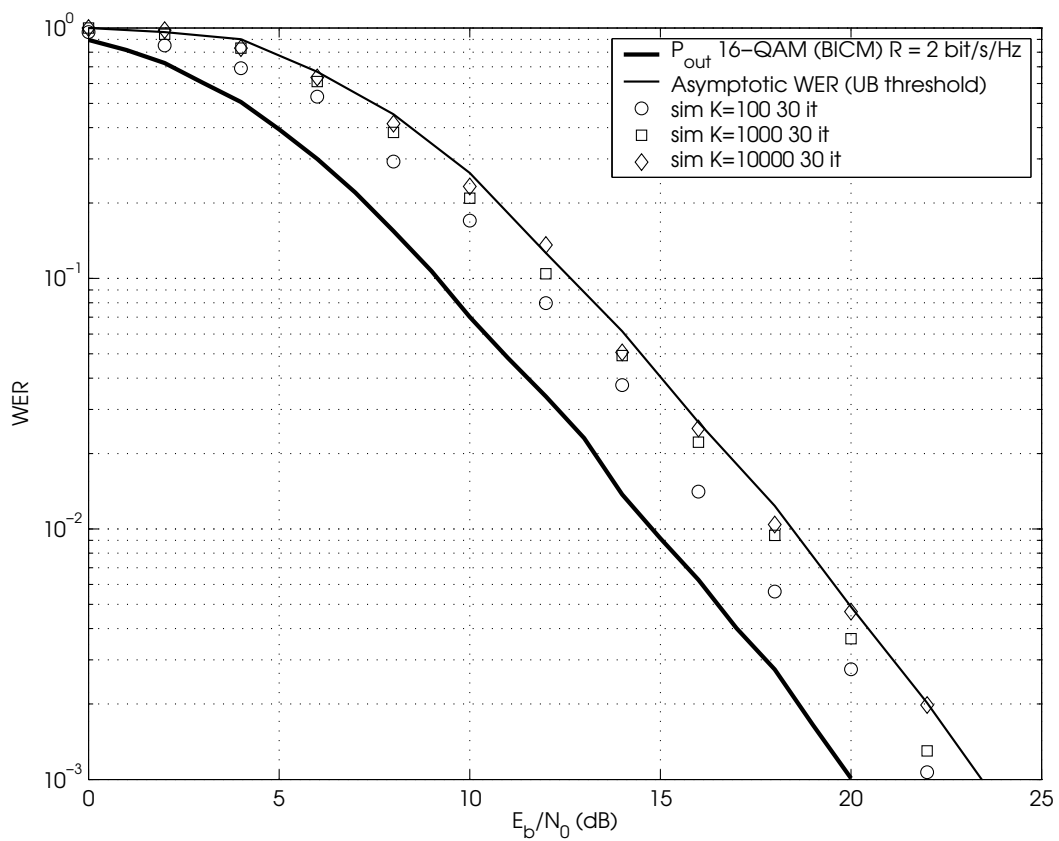
Figure 10: Asymptotic error probability (31) for a rate $R = 2$ RBA code mapped over $N = 2$ fading blocks with 16-QAM (BICM) and corresponding BP decoding simulation with 30 iterations for $K = 100, 1000$ and 10000.

the PWEF of the same code with information block length $K = 20$. Interestingly, the truncation of the PWEF yields too optimistic results, while the approximation based on the complete PWEF of the shorter code still approximates very accurately the WER of the longer code. This has the advantage that, in practice, computing the weight enumerator of shorter codes is in general less computationally intensive.

As a matter of fact, the PWEF of the short code contains much more information on the code behavior than the truncated PWEF of the long code. This is clearly illustrated by the PWEFs in Figs. 12(a) and 12(b), showing the (non-asymptotic) exponential growth rate of the PWEF defined as

$$F(\overline{\Delta}_p) \triangleq \frac{1}{L^N} \log A_{\overline{\Delta}_p} \tag{32}$$

as a function of the normalized product weight $\overline{\Delta}_p = \Delta_p / L_B^N$ for the RBAs of rate 1/4, with 20 and 100 information bits (every mark corresponds to one pairwise error event with normalized product weight $\overline{\Delta}_p$). Truncation at $\Delta_p^{\max} = 10000$ corresponds to maximum normalized product 10^{-4}, which means that only the portion for $0 \leq \overline{\Delta}_p \leq 10^{-4}$ of the distribution of Fig. 12(b) is taken into account in the BPL approximation using the truncated enumerator. This is clearly not sufficient to describe the RBA product weight enumerator, as opposed to the PWEF of the shorter code.

4.4 On code optimization

So far we have seen that the BCC coding structure yields weakly good codes for the block-fading channel. However, most of the shown examples were based on the simple RBA structure. It is then natural to ask whether more general BCCs can reduce significantly the gap from outage. In this section we show some examples of other BCC constructions that in some case improve upon the basic RBA of same rate. Figs. 13 and 14 show the performance of BCCCs with binary rate $r = 1/4$, attaining full diversity, with BPSK and 16-QAM BICM respectively for $N = 4$ fading blocks, for $K = 1024$ and 40 BP decoder iterations. The octal generators are given in the legend. We have also considered the 4 states accumulator given in [35, Ch. 4] with generator $(1/7)_8$. We observe that in both cases the gap from outage is approximately of 1 dB. We notice from Fig. 13 that using more complicated outer or inner codes does not yield a significant gain. Using the 4 states inner accumulator in an RBA scheme yields almost the same performance that the best

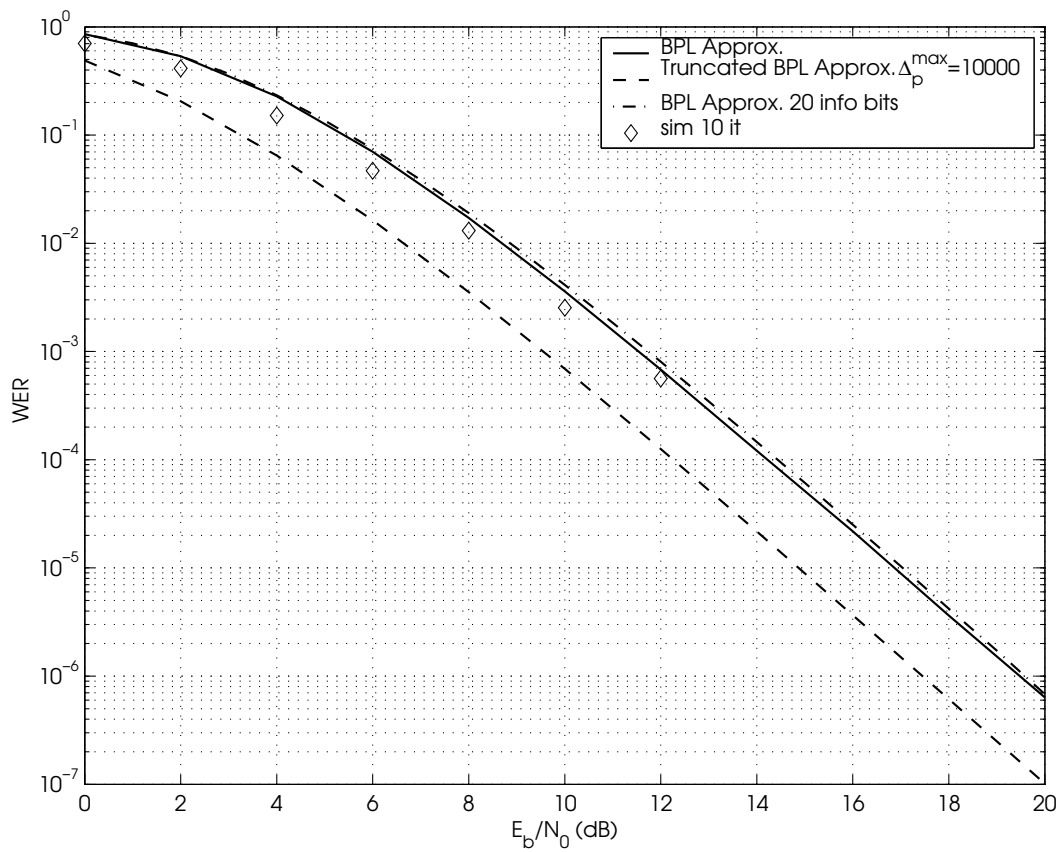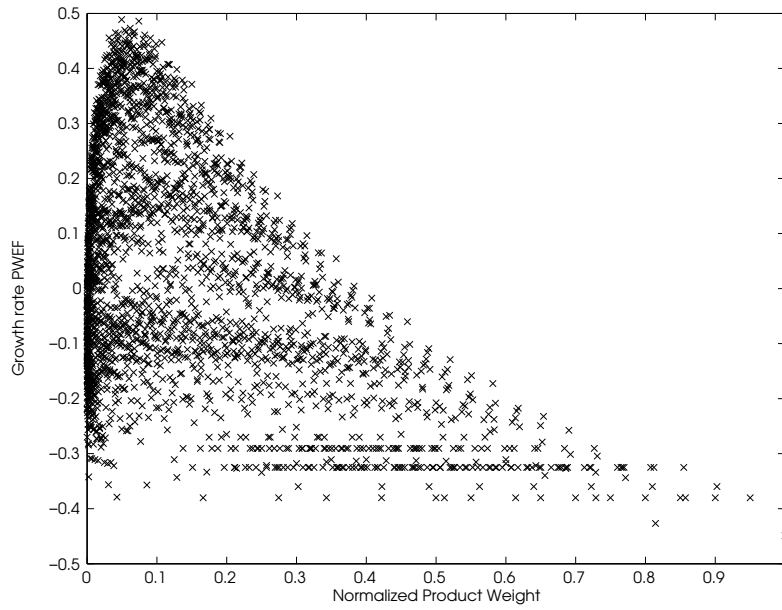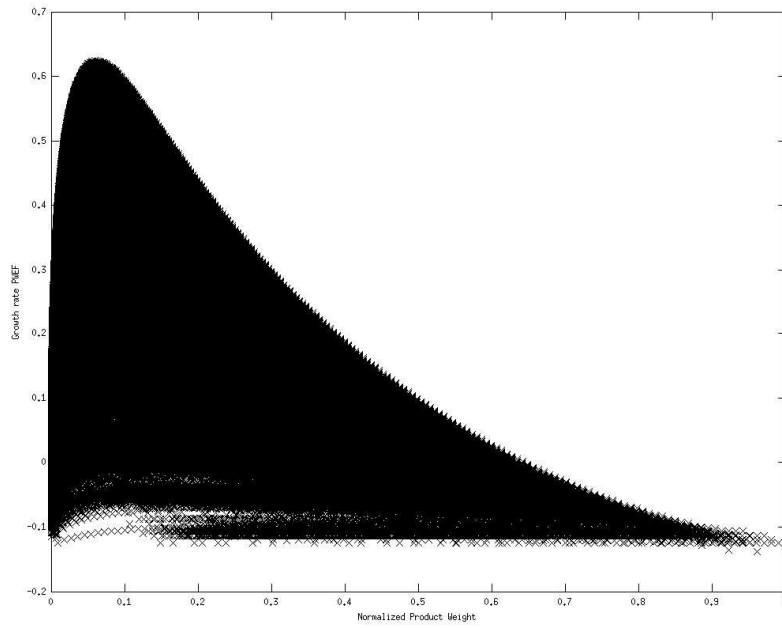Figure 11: WER obtained by BP decoding simulation with 10 iterations and BPL approximations for RBA with rate $R = 1/4$ and 100 information bits per frame, over $N = 4$ fading blocks.

(a) RBA of rate $R = 1/4$ and $K = 20$ information bits

(b) RBA of rate $R = 1/4$ and $K = 100$ information bits

Figure 12: PWEF growth rate for RBA of rate $R = 1/4$ with 20 (a) and 100 (b) information bits per frame, over $N = 4$ blocks.

33

BCCC.

From these examples, and several other numerical experiments not reported here for the sake of space limitation, it seems that, while some room is left for code optimization by searching over the component code generators, the improvements that may be expected are not dramatic and probably do not justify the decoding complexity increase (similar conclusions can be drawn from the results of [3, 4, 5, 6]).
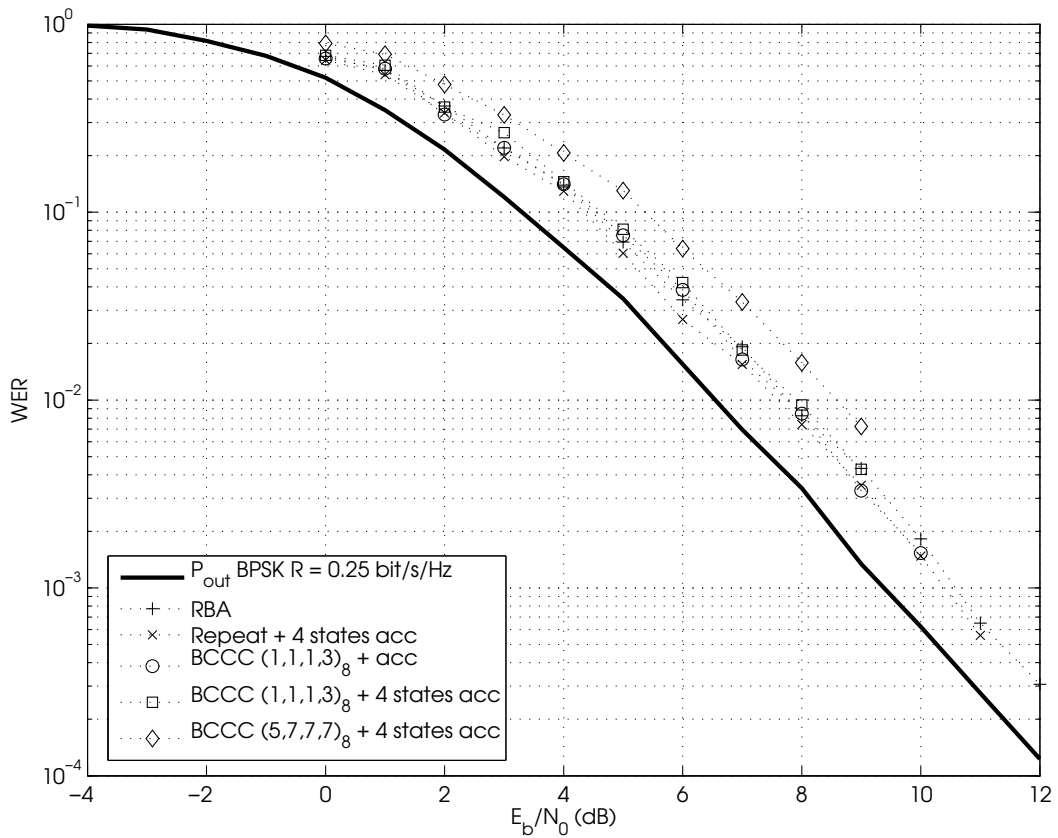

Figure 13: WER (simulation with 40 BP decoding iterations) of several BCCs of rate $R = 1/4$ over BPSK, for $N = 4$ fading blocks and $K = 1024$.
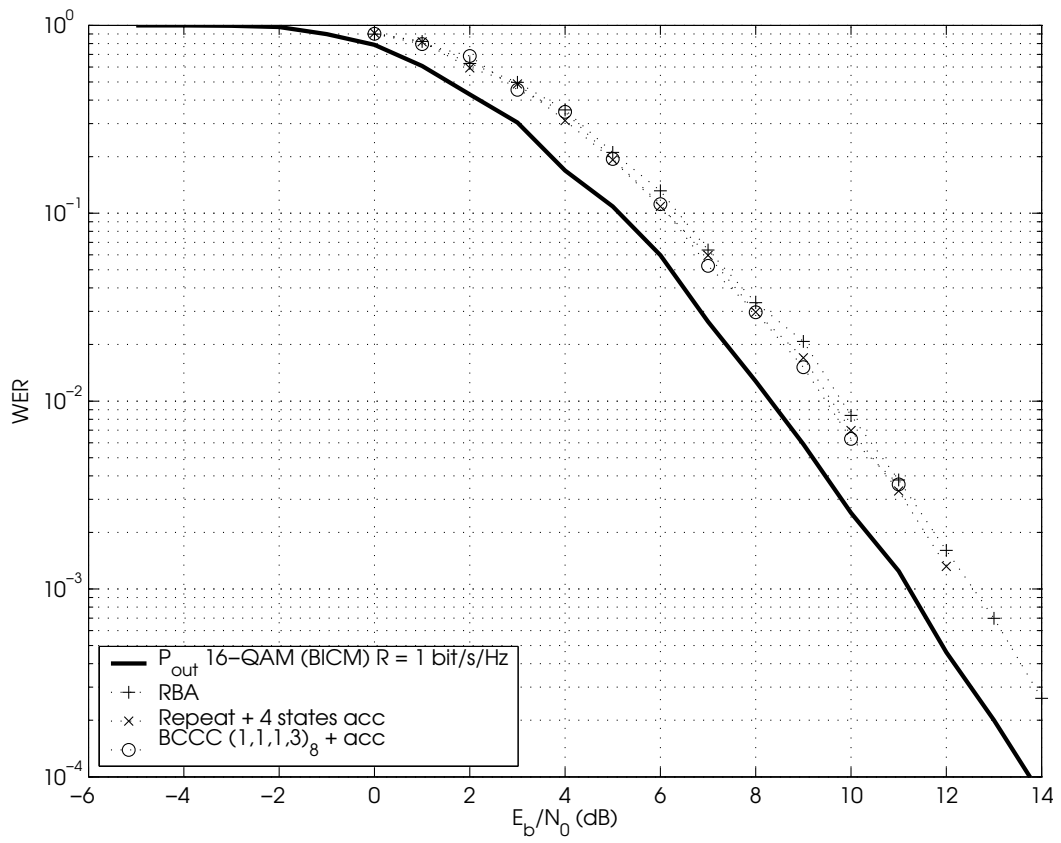
Figure 14: WER (simulation with 40 BP decoding iterations) of several BCCs of rate $R = 1/4$ over 16-QAM (BICM), for $N = 4$ fading blocks and $K = 1024$.

5 Conclusions

In this paper we determined the SNR reliability function of codes over given finite signal sets over the block-fading channel. Random coding obtained by concatenating a linear binary random code to the modulator via a fixed one-to-one mapping achieve the same optimal SNR reliability function provided that the block length grows rapidly enough with SNR. Pragmatic BICM schemes under suboptimal BICM-ML decoding achieve the same random coding SNR exponent of their non-BICM counterparts (under optimal ML decoding).

Driven by these findings, we have proposed a general structure for random-like codes adapted to the block-fading channel, based on blockwise concatenation and on BICM (to attain large spectral efficiency). We provided some easily computable bounds and approximations to the WER of these codes under ML decoding and BICM-ML decoding. Remarkably, our approximations agree very well with the simulated performance of the iterative BP decoder at any SNR and even for relatively short block length.

The proposed codes have WER almost independent of the block length (for large block length), showing a fixed SNR gap from outage probability. We introduced the concept of "weak goodness" for specific ensembles of codes having this behavior for large block length and large SNR, and we provided a sufficient condition for weak goodness of specific code ensembles in terms of their asymptotic multivariate weight enumerator exponential growth rate function.

Finally, we showed via extensive computer simulation that, while some improvement can be expected by careful optimization of the component codes, weakly good BCC ensembles have very similar behavior and only marginal improvements can be expected from careful optimization of the component encoders.

APPENDIX

A   Proof of Lemma 1

Consider a family of codes for the block-fading channel (2) of given block length L, indexed by their operating SNR ρ, such that the ρ-th code has rate $R(\rho) = r \log \rho$ (in nat), where $r \in [0, 1]$, and WER (averaged over the channel fading) $P_e(\rho)$. Using Fano inequality [10] it is easy to show that $P_{\text{out}}^G(\rho, R(\rho))$ yields an upper bound on the best possible SNR exponent. Recall that the fading power gains are defined as $\gamma_n = |h_n|^2$, for $n = 1, \ldots, N$, and are i.i.d. exponentially distributed. Following in the footsteps of [10] we define the normalized log-fading gains as $\alpha_n = -\log \gamma_n / \log \rho$. Hence, the joint distribution of the vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)$ is given by

$$p(\boldsymbol{\alpha}) = (\log \rho)^N \exp\left(-\sum_{n=1}^N \rho^{-\alpha_n} \right) \rho^{-\sum_{n=1}^N \alpha_n} \tag{33}$$

The information outage event under Gaussian inputs is given by $\{\sum_{n=1}^N \log(1 + \rho \gamma_n) \leq Nr \log \rho\}$. By noticing that $(1 + \rho \gamma_n) \doteq \rho^{[1-\alpha_n]_+}$, we can write the outage event as

$$\mathcal{A} = \left\{ \boldsymbol{\alpha} \in \mathbb{R}^N \ : \ \sum_{n=1}^N [1 - \alpha_n]_+ \leq rN \right\} \tag{34}$$

The probability of outage is easily seen to satisfy the exponential equality

$$P_{\text{out}}^G(\rho, R(\rho)) \doteq \int_{\mathcal{A} \cap \mathbb{R}_+^N} \rho^{-\sum_{n=1}^N \alpha_n} d\boldsymbol{\alpha} \tag{35}$$

Therefore, the SNR exponent of outage probability is given by the following limit

$$d_{\text{out}}(r) = -\lim_{\rho \to \infty} \frac{1}{\log(\rho)} \log \int_{\mathcal{A} \cap \mathbb{R}_+^N} \exp\left(-\log(\rho) \sum_{n=1}^N \alpha_n \right) d\boldsymbol{\alpha} \tag{36}$$

We apply Varadhan's integral lemma [36] and we obtain

$$d_{\text{out}}(r) = \inf_{\boldsymbol{\alpha} \in \mathcal{A} \cap \mathbb{R}_+^N} \left\{ \sum_{n=1}^N \alpha_n \right\} \tag{37}$$

The constraint set is the intersection of the region defined by $\sum_{n=1}^N \alpha_n \geq N(1-r)$ and the region defined by $\alpha_n \in [0, 1]$ for all $n = 1, \ldots, N$. For all $r \in [0, 1]$, the infimum in (37) is given by

$$d_{\text{out}}(r) = N(1-r) \tag{38}$$

In order to show that this exponent is actually the SNR reliability function for any $L \geq 1$, we have to prove achievability. We examine the average WER of a Gaussian random coding ensemble. Fix $L \geq 1$ and, for any SNR ρ, consider the ensemble generated with i.i.d. components with input probability $P_X = \mathcal{N}_{\mathbb{C}}(0,1)$ and rate $R(\rho) = r \log \rho$. The pairwise error probability under ML decoding, for two codewords $\mathbf{X}(0)$ and $\mathbf{X}(1)$ for given fading coefficients is upperbounded by the Chernoff bound

$$P(\mathbf{X}(0) \rightarrow \mathbf{X}(1)|\mathbf{h}) \leq \exp\left(-\frac{\rho}{4} \|\mathbf{H}(\mathbf{X}(0) - \mathbf{X}(1))\|_F^2\right) \tag{39}$$

By averaging over the random coding ensemble and using the fact that the entries of the matrix difference $\mathbf{X}(0) - \mathbf{X}(1)$ are i.i.d. $\sim \mathcal{N}_{\mathbb{C}}(0,2)$, we obtain

$$\overline{P(\mathbf{X}(0) \rightarrow \mathbf{X}(1)|\mathbf{h})} \leq \prod_{n=1}^{N}\left[1 + \frac{1}{2}\rho\gamma_n\right]^{-L} \doteq \rho^{-L\sum_{n=1}^{N}[1-\alpha_n]_+} \tag{40}$$

(in general, the bar denotes quantities averaged over the code ensemble). By summing over all $\rho^{rNL} - 1$ messages $w \neq 0$, we obtain the ensemble average union bound

$$\overline{P_e(\rho|\mathbf{h})} \;\dot{\leq}\; \rho^{-L\sum_{n=1}^{N}[1-\alpha_n]_+ + LNr} \tag{41}$$

Next, we upperbound the ensemble average WER by separating the outage event from the non-outage event (the complement set denoted by $\mathcal{A}^c$) as follows:

$$\overline{P_e(\rho)} \leq \Pr(\mathcal{A}) + \overline{\Pr(\text{error}, \mathcal{A}^c)} \tag{42}$$

Achievability is proved if we can show that $\overline{\Pr(\text{error}, \mathcal{A}^c)} \;\dot{\leq}\; \rho^{-N(1-r)}$ for all $L \geq 1$ and $r \in [0,1]$. We have

$$\overline{\Pr(\text{error}, \mathcal{A}^c)} \;\dot{\leq}\; \int_{\mathcal{A}^c \cap \mathbb{R}_+^N} \exp\left(-\log(\rho)\left(\sum_{n=1}^{N}\alpha_n + L\left(\sum_{n=1}^{N}[1-\alpha_n]_+ - rN\right)\right)\right) \tag{43}$$

By using again Varadhan intergral lemma we obtain the lower bound on the Gaussian random coding exponent

$$d_G^{(r)}(r) \geq \inf_{\boldsymbol{\alpha} \in \mathcal{A}^c \cap \mathbb{R}_+^N}\left\{\sum_{n=1}^{N}\alpha_n + L\left(\sum_{n=1}^{N}[1-\alpha_n]_+ - rN\right)\right\} \tag{44}$$

where $\mathcal{A}^c$ is defined explicitly by

$$\sum_{n=1}^{N}[1-\alpha_n]_+ \geq rN$$

It is easily seen that for all $L \geq 1$ and $r \in [0, 1]$ the infimum is obtained[7] by

$$
\begin{aligned}
\alpha_n &= 1, \quad \text{for } n = 1, \ldots, k-1 \\
\alpha_k &= 1 + \lfloor rN \rfloor - rN \\
\alpha_n &= 0, \quad \text{for } n = k+1, \ldots, N
\end{aligned}
\tag{45}
$$

where $k = N - \lfloor rN \rfloor$, and yields $d_G^{(r)}(r) \geq N(1 - r)$. Since this lower bound coincides with the outage probability upper bound (38), we obtain that $d_G^\star(r) = N(1 - r)$ and it is achieved by Gaussian codes for any $L \geq 1$. Any fixed coding rate R corresponds to the case $r = 0$, from which the statement of Lemma 1 follows.

---

[7]This solution is not unique. Any configuration of the variables α_n having $k-1$ variables equal to 1, $N - k - 1$ variables equal to 0 and one variable equal to $1 + \lfloor rN \rfloor - rN$ yields the same value of the infimum. Moreover, for $L = 1$ also the solution $\alpha_n = 0$ for all n yields the same value.

B   Proof of Theorem 1

We fix the number of fading blocks N, the coding rate R and the (unit energy) modulation signal set $\mathcal{X}$. Using Fano inequality [10] it is easy to show that $P_{\text{out}}^{\mathcal{X}}(\rho, R)$ yields an upper bound on the best possible SNR exponent attained by coded modulations with signal set $\mathcal{X}$. The outage probability with discrete inputs is lower bounded by

$$
\begin{aligned}
P_{\text{out}}^{\mathcal{X}}(\rho, R) \;\overset{\text{a}}{\geq}\; & \Pr\left(\frac{1}{N} \sum_{n=1}^{N} J_{\mathcal{X}}(\gamma_n \rho) < R \right) \\
=\; & \Pr\left(\frac{1}{N} \sum_{n=1}^{N} \left(M - 2^{-M} \sum_{x \in \mathcal{X}} \mathbb{E}\left[\log_2 \sum_{x' \in \mathcal{X}} e^{-|\sqrt{\rho \gamma_n}(x - x') + Z|^2 + |Z|^2} \right] \right) < R \right) \\
=\; & \Pr\left(\frac{1}{N} \sum_{n=1}^{N} \left(M - 2^{-M} \sum_{x \in \mathcal{X}} \mathbb{E}\left[\log_2 \sum_{x' \in \mathcal{X}} e^{-\left|\sqrt{\rho^{1-\alpha_n}}(x - x') + Z\right|^2 + |Z|^2} \right] \right) < R \right) \quad (46)
\end{aligned}
$$

where $Z \sim \mathcal{N}_{\mathbb{C}}(0, 1)$, and the last equality follows just from the definition of the normalized log-fading gains $\alpha_n = -\log \gamma_n / \log \rho$. The inequality (a) is due to the fact that we have a strict inequality in the probability on the right. Since the term,

$$
0 \leq \log_2\left(\sum_{x' \in \mathcal{X}} e^{-\left|\sqrt{\rho^{1-\alpha_n}}(x - x') + z\right|^2 + |z|^2} \right) \leq \log_2\left(|\mathcal{X}| e^{|z|^2} \right) = \frac{\log |\mathcal{X}| + |z|^2}{\log 2} \quad (47)
$$

and

$$
\mathbb{E}\left[\frac{\log |\mathcal{X}| + |z|^2}{\log 2} \right] < \infty \quad (48)
$$

we can apply the dominated convergence theorem [37], for which,

$$
\lim_{\rho \to \infty} \mathbb{E}\left[\log_2\left(\sum_{x' \in \mathcal{X}} e^{-\left|\sqrt{\rho^{1-\alpha_n}}(x - x') + Z\right|^2 + |Z|^2} \right) \right] = \mathbb{E}\left[\lim_{\rho \to \infty} \log_2\left(\sum_{x' \in \mathcal{X}} e^{-\left|\sqrt{\rho^{1-\alpha_n}}(x - x') + Z\right|^2 + |Z|^2} \right) \right].
$$
$$(49)$$

Since for all $z \in \mathbb{C}$ with $|z| \leq \infty$ and $s \neq 0$ we have that

$$
\lim_{\rho \to \infty} e^{-\left|\sqrt{\rho^{1-\alpha_n}}s + z\right|^2 + |z|^2} = \begin{cases} 0 & \text{for} \quad \alpha_n < 1 \\ 1 & \text{for} \quad \alpha_n > 1 \end{cases}
$$

we have that, for large ρ and $\alpha_n < 1$, $J_{\mathcal{X}}(\rho^{1-\alpha_n}) \to M$ and for $\alpha_n > 1$, $J_{\mathcal{X}}(\rho^{1-\alpha_n}) \to 0$. Hence, for every $\epsilon > 0$, we have the lower bound

$$
\begin{aligned}
P_{\text{out}}(\rho, R) \;\overset{\cdot}{\geq}\; & \Pr\left(\frac{1}{N} \sum_{n=1}^{N} \mathbb{1}\{\alpha_n \leq 1 + \epsilon\} < \frac{R}{M} \right) \\
\doteq\; & \int_{\mathcal{A}_\epsilon \cap \mathbb{R}_+^N} \rho^{-\sum_{n=1}^{N} \alpha_n} d\boldsymbol{\alpha} \quad (50)
\end{aligned}
$$

40

where we define the event

$$\mathcal{A}_\epsilon = \left\{ \boldsymbol{\alpha} \in \mathbb{R}^N \ : \ \frac{1}{N} \sum_{n=1}^N \mathbb{1}\{\alpha_n \leq 1+\epsilon\} < \frac{R}{M} \right\} \tag{51}$$

Using Varadhan's lemma, we get the upperbound to the SNR reliability function as

$$d^\star_{\mathcal{X}}(R) \leq \inf_{\epsilon > 0} \inf_{\boldsymbol{\alpha} \in \mathcal{A}_\epsilon \cap \mathbb{R}^N_+} \left\{ \sum_{n=1}^N \alpha_n \right\} \tag{52}$$

It is not difficult to show that the $\boldsymbol{\alpha}$ achieving the inner infimum in (52) is given by

$$\begin{aligned}
\alpha_n &= 1+\epsilon, \quad \text{for} \ \ n=1,\ldots,N-k \\
\alpha_n &= 0, \quad \text{for} \ \ n=N-k+1,\ldots,N
\end{aligned} \tag{53}$$

where $k = 0,\ldots,N-1$ is the unique integer satisfying $\frac{k}{N} < \frac{R}{M} \leq \frac{k+1}{N}$. Since this holds for each $\epsilon > 0$, we can make the bound as tight as possible by letting $\epsilon \downarrow 0$, thus obtaining $\delta(N,M,R)$ defined in (13).

In order to show that this exponent is actually the SNR reliability function, we have to prove achievability. We examine the average WER of the coded modulation ensemble obtained by concatenating a random binary linear code $\mathcal{C}$ with the signal set $\mathcal{X}$ via an arbitrary fixed one-to-one mapping $\mu : \mathbb{F}_2^M \to \mathcal{X}$. The random binary linear code and the one-to-one mapping induce a uniform i.i.d. input distribution over $\mathcal{X}$. The pairwise error probability under ML decoding, for two codewords $\mathbf{X}(0)$ and $\mathbf{X}(1)$ for given fading coefficients is again upperbounded by the Chernoff bound (39). By averaging over the random coding ensemble and using the fact that the entries of each codeword $\mathbf{X}(0)$ and $\mathbf{X}(1)$ are i.i.d. uniformly distributed over $\mathcal{X}$, we obtain

$$\overline{P(\mathbf{X}(0) \to \mathbf{X}(1)|\mathbf{h})} \leq \prod_{n=1}^N B_n^L \tag{54}$$

where we define the Bhattacharyya coefficient [38]

$$B_n = 2^{-2M} \sum_{x \in \mathcal{X}} \sum_{x' \in \mathcal{X}} \exp\left(-\frac{\rho}{4} \gamma_n |x-x'|^2 \right) \tag{55}$$

By summing over all $2^{NLR} - 1$ messages $w \neq 0$, we obtain the union bound

$$\begin{aligned}
\overline{P_e(\rho|\mathbf{h})} &\leq \exp\left(-NLM\log(2) \left[1 - \frac{R}{M} - \frac{1}{NM} \sum_{n=1}^N \log_2 \left(1 + 2^{-M} \sum_{x \neq x'} e^{-\frac{1}{4}|x-x'|^2 \rho^{1-\alpha_n}} \right) \right] \right) \\
&= \exp\left(-NLM\log(2) G(\rho, \boldsymbol{\alpha}) \right)
\end{aligned} \tag{56}$$

41

We notice a fundamental difference between the above union bound for random coding with discrete inputs and the corresponding union bound for random coding with Gaussian inputs given in (41). With Gaussian inputs, we have that the union bound vanishes for any finite block length L as $\rho \to \infty$. On the contrary, with discrete inputs the union bound (56) is bounded away from zero for any finite L as $\rho \to \infty$. This is because with discrete inputs the probability that two codewords coincide (i.e., have zero Euclidean distance) is positive for any finite L. Hence, it is clear that in order to obtain a non-zero random coding SNR exponent we have to consider a code ensemble with block length $L = L(\rho)$, increasing with ρ. To this purpose, we define

$$\beta \triangleq \lim_{\rho \to \infty} \frac{L(\rho)}{\log \rho}. \tag{57}$$

By averaging over the fading and using the fact that error probability cannot be larger than 1, we obtain

$$\overline{P_e(\rho)} \dot{\leq} \int_{\mathbb{R}_+^N} \rho^{-\sum_{n=1}^N \alpha_n} \min \left\{ 1, \exp\left(-NL(\rho)M \log(2) G(\rho, \boldsymbol{\alpha}) \right) \right\} \, d\boldsymbol{\alpha} \tag{58}$$

We notice that

$$\lim_{\rho \to \infty} \log_2 \left(1 + 2^{-M} \sum_{x \neq x'} e^{-\frac{1}{4}|x-x'|^2 \rho^{1-\alpha_n}} \right) = \begin{cases} 0 & \text{for} \quad \alpha_n < 1 \\ M & \text{for} \quad \alpha_n > 1 \end{cases}$$

Hence, for $\epsilon > 0$, a lower bound on the random coding SNR exponent can be obtained by replacing $G(\rho, \boldsymbol{\alpha})$ by

$$\widetilde{G}_\epsilon(\boldsymbol{\alpha}) = 1 - \frac{R}{M} - \frac{1}{N} \sum_{n=1}^N \mathbb{1}\{\alpha_n \geq 1 - \epsilon\} \tag{59}$$

We define the event

$$\mathcal{B}_\epsilon = \left\{ \boldsymbol{\alpha} : \widetilde{G}_\epsilon(\boldsymbol{\alpha}) \leq 0 \right\} \tag{60}$$

Hence, from (58) and what said above we obtain

$$\overline{P_e(\rho)} \quad \dot{\leq} \quad \int_{\mathcal{B}_\epsilon \cap \mathbb{R}_+^N} \rho^{-\sum_{n=1}^N \alpha_n} \, d\boldsymbol{\alpha} +$$
$$+ \int_{\mathcal{B}_\epsilon^c \cap \mathbb{R}_+^N} \exp\left(-\log(\rho) \left[\sum_{n=1}^N \alpha_n + N\beta M \log(2) \widetilde{G}_\epsilon(\boldsymbol{\alpha}) \right] \right) \, d\boldsymbol{\alpha} \tag{61}$$

By applying again Varadhan's lemma we obtain a lower bound to the random coding SNR exponent given by

$$d_\mathcal{X}^{(r)}(R) \geq \sup_{\epsilon > 0} \min \left\{ \inf_{\boldsymbol{\alpha} \in \mathcal{B}_\epsilon \cap \mathbb{R}_+^N} \left\{ \sum_{n=1}^N \alpha_n \right\}, \inf_{\boldsymbol{\alpha} \in \mathcal{B}_\epsilon^c \cap \mathbb{R}_+^N} \left\{ \sum_{n=1}^N \alpha_n + N\beta M \log(2) \widetilde{G}_\epsilon(\boldsymbol{\alpha}) \right\} \right\} \tag{62}$$

It is not difficult to show that the $\boldsymbol{\alpha}$ achieving the first infimum in (62) is given by

$$
\begin{aligned}
\alpha_n &= 1 - \epsilon, \quad \text{for } n = 1, \ldots, N - k \\
\alpha_n &= 0, \quad \text{for } n = N - k + 1, \ldots, N
\end{aligned}
\tag{63}
$$

where $k = 0, \ldots, N - 1$ is the unique integer satisfying $\frac{k}{N} \leq \frac{R}{M} < \frac{k+1}{N}$.

For the second infimum in (62) we can rewrite the argument in the form

$$
N\beta M \log(2) \left(1 - \frac{R}{M} \right) + \sum_{n=1}^{N} [\alpha_n - \beta M \log(2) \mathbb{1} \{\alpha_n \geq 1 - \epsilon\}]
\tag{64}
$$

We distinguish two cases. If $0 \leq \beta M \log(2) < 1$, then

$$
\alpha_n - \beta M \log(2) \mathbb{1} \{\alpha_n \geq 1 - \epsilon\}
\tag{65}
$$

attains its minimum at $\alpha_n = 0$. Hence, we obtain that the infimum is given by $N\beta M \log(2) \left(1 - \frac{R}{M} \right)$. If $\beta M \log(2) \geq 1$, then (65) attains its absolute minimum at $\alpha_n = 1 - \epsilon$, and its second smallest minimum at $\alpha_n = 0$. The number of terms α_n that can be made equal to $1 - \epsilon$ subject to the constraint $\boldsymbol{\alpha} \in \mathcal{B}_\epsilon^c \cap \mathbb{R}_+^N$ is given by $\lfloor N(1 - R/M) \rfloor$. Hence, the infimum is given by

$$
N\beta M \log(2) \left(1 - \frac{R}{M} \right) + (1 - \epsilon - \beta M \log(2)) \left\lfloor N \left(1 - \frac{R}{M} \right) \right\rfloor
\tag{66}
$$

Both the first and the second infima are simultaneously maximized by letting $\epsilon \downarrow 0$. By collecting all results together, we obtain that the random coding SNR exponent is lower bounded by (14).

The random coding SNR exponent of the associated BICM channel is immediately obtained by using again the Bhattacharyya union bound [11]. In particular, for two randomly generated codewords $\mathbf{C}(0), \mathbf{C}(1) \in \mathcal{C}$

$$
\overline{P(\mathbf{C}(0) \to \mathbf{C}(1)|\mathbf{h})} \leq 2^{-LMN} \prod_{n=1}^{N} (1 + B_n(\rho, \mu, \mathcal{X}))^{LM}
\tag{67}
$$

where $B_n(\rho, \mu, \mathcal{X})$ is defined in (21). The error probability averaged over the random BICM ensemble can be upperbounded by

$$
\begin{aligned}
\overline{P_e(\rho|\mathbf{h})} &\leq \exp \left(-NLM \log(2) \left[1 - \frac{R}{M} - \frac{1}{N} \sum_{n=1}^{N} \log_2 \left(1 + B_n(\rho, \mu, \mathcal{X}) \right) \right] \right) \\
&= \exp \left(-NLM \log(2) G'(\rho, \boldsymbol{\alpha}) \right)
\end{aligned}
\tag{68}
$$

It is not difficult to see that

$$\lim_{\rho \to \infty} \log_2 \left(1 + B_n(\rho, \mu, \mathcal{X})\right) = \begin{cases} 0 & \text{for} \quad \alpha_n < 1 \\ 1 & \text{for} \quad \alpha_n > 1 \end{cases}$$

Hence, for $\epsilon > 0$, a lower bound on the random coding SNR exponent can be obtained by replacing $G'(\rho, \boldsymbol{\alpha})$ by $\widetilde{G}_\epsilon(\boldsymbol{\alpha})$ defined in (59). It then follows that the random coding SNR exponent of the associated BICM channel satisfies the same lower bound of the original block-fading channel.

C   Weight Enumerators

Throughout this work we used extensively the multivariate weight enumerating function (MWEF) of binary codes partitioned into N blocks. In this Appendix we focus on their calculation.

C.1   MWEF of BCCs

To compute the MWEF, we assume that blockwise concatenation is performed through a set of N uniform interleavers [33, 34]. Then, we can compute the average multivariate weight enumeration function according to the following

Proposition 1 *Let* $\mathcal{C}^{BCC}$ *be a blockwise concatenated code mapped over* N *fading blocks constructed by concatenating an outer code* $\mathcal{C}^O$ *mapped over* N *blocks with input multivariate-output weight enumeration function* $A^O_{i,w_1,\ldots,w_N}$, *and* N *inner codes* $\mathcal{C}^I$ *with input-output weight enumeration functions* $A^I_{i,w}$, *through* N *uniform interleavers of length* L_π. *Then, the average input multivariate-output weight enumeration function of* $\mathcal{C}^{BCC}$, $A^{BCC}_{i,w_1,\ldots,w_N}$, *is given by,*

$$A^{BCC}_{i,w_1,\ldots,w_N} = \sum_{\ell_1,\ldots,\ell_N} \frac{A^O_{i,\ell_1,\ldots,\ell_N} \prod\limits_{n=1}^{N} A^I_{\ell_n,w_n}}{\prod\limits_{n=1}^{N} \binom{L_\pi}{\ell_n}}. \tag{69}$$

Proof.   Define the Cartesian product code $\mathcal{C}_P = \mathcal{C}^I \times \ldots \times \mathcal{C}^I = (\mathcal{C}^I)^N$ with multivariate-input multivariate-output weight enumeration function $A^P_{i_1,\ldots,i_N,w_1,\ldots,w_N}$. Then,

$$\begin{aligned} A^{BCC}_{i,w_1,\ldots,w_N} &= \mathbb{E}\left[A^P_{\ell_1,\ldots,\ell_N,w_1,\ldots,w_N} \Big| i \right] \\ &= \sum_{\ell_1,\ldots,\ell_N} P(\ell_1,\ldots,\ell_N|i) A^P_{\ell_1,\ldots,\ell_N,w_1,\ldots,w_N} \end{aligned} \tag{70}$$

where $\mathbb{E}[.]$ denotes expectation over all uniform interleavers and $P(\ell_1,\ldots,\ell_N|i)$ is the probability of having an error event of input weights $\ell_1,\ldots,\ell_N$ of $\mathcal{C}^P$ given the input weight i of $\mathcal{C}^O$ and it is given by,

$$P(\ell_1,\ldots,\ell_N|i) = \frac{A^O_{i,\ell_1,\ldots,\ell_N}}{\prod\limits_{n=1}^{N} \binom{L_\pi}{\ell_n}}. \tag{71}$$

By construction,

$$A^P_{\ell_1,\ldots,\ell_N,w_1,\ldots,w_N} = \prod_{n=1}^{N} A^I_{\ell_n,w_n} \tag{72}$$

and thus we obtain (69). □

Remarkably, in the case of RBAs the function $A^{RBA}_{i,w_1,\ldots,w_N}$ can be written in closed form in a way similar to [24] for RA codes in the AWGN channel. We obtain

$$A^{RBA}_{i,w_1,\ldots,w_N} = \frac{\displaystyle\prod_{n=1}^{N} \binom{L_\pi - w_n}{\lfloor i/2 \rfloor}\binom{w_n - 1}{\lceil i/2 \rceil - 1}}{\binom{L_\pi}{i}^{N-1}} \tag{73}$$

The MWEF of the BCC is eventually obtained as

$$A_{w_1,\ldots,w_N} = \sum_i A_{i,w_1,\ldots,w_N}.$$

In order to evaluate the approximations (18) and (25), we need the PWEF which can be trivially obtained from the MWEF.

C.2 Asymptotic MWEF of BCCs

Following in the footsteps of [32], we can derive the asymptotic exponential growth rate of the MWEF $A^{BCC}_{i,w_1,\ldots,w_N}$. Define the asymptotic exponential growth rate functions of the input-output MWEFs

$$a^O(\iota,\delta_1,\ldots,\delta_N) \triangleq \lim_{\epsilon\to 0}\lim_{L_\pi\to\infty} \frac{1}{L_\pi}\log\left|\mathcal{S}^{L_\pi,O}_\epsilon(\iota,\delta_1,\ldots,\delta_N)\right| \tag{74}$$

$$a^I(\delta,\omega) \triangleq \lim_{\epsilon\to 0}\lim_{L\to\infty} \frac{1}{L}\log\left|\mathcal{S}^{L,I}_\epsilon(\delta,\omega)\right| \tag{75}$$

$$a^{BCC}(\iota,\omega_1,\ldots,\omega_N) \triangleq \lim_{\epsilon\to 0}\lim_{L\to\infty} \frac{1}{L}\log\left|\mathcal{S}^{L,BCC}_\epsilon(\iota,\omega_1,\ldots,\omega_N)\right| \tag{76}$$

where $\mathcal{S}^{L_\pi,O}_\epsilon(\iota,\delta_1,\ldots,\delta_N)$ is the set of codewords in the (per block) length-L_π outer code with input Hamming weight satisfying $|i/K - \iota| \le \epsilon$ and output Hamming weights per block satisfying $|w_n/L_\pi - \delta_n| \le \epsilon$ for all $n = 1,\ldots,N$, where $\mathcal{S}^{L,I}_\epsilon(\delta,\omega)$ is the set of codewords in the length-L inner code with input Hamming weight satisfying $|i/L_\pi - \delta| \le \epsilon$ and output Hamming weights satisfying $|w/L-\delta| \le \epsilon$, and where $\mathcal{S}^{L,BCC}_\epsilon(\iota,\omega_1,\ldots,\omega_N)$ is the set of codewords in the (per block) length-L BCC code with input Hamming weight satisfying $|i/K - \iota| \le \epsilon$ and output Hamming weights per block satisfying $|w_n/L - \omega_n| \le \epsilon$ for all $n = 1,\ldots,N$.

46

Assuming that we can compute (e.g., extending the methods of [32] to the case of multivariate weight enumerators) the functions $a^O(\iota, \delta_1, \ldots, \delta_N)$ and $a^I(\delta, \omega)$, the desired function $a^{BCC}(\iota, \omega_1, \ldots, \omega_N)$ for the BCC scheme is readily obtained via Proposition 1. By definition, notice that $L_\pi = r_I L$ and $K = r_I r_O N L$. Moreover,

$$\lim_{L_\pi \to \infty} \binom{L_\pi}{L_\pi \delta_n} = e^{L_\pi h(\delta)} = e^{L r_I h(\delta)} \tag{77}$$

where $h(p) = -p \log(p) - (1-p) \log(1-p)$ is the binary entropy function. Therefore, from (69) we find the asymptotic input-output growth rate of the BCC as

$$a^{BCC}(\iota, \omega_1, \ldots, \omega_N) = \max_{\boldsymbol{\delta} \in [0,1]^N} \left\{ r_I a^O(\iota, \delta_1, \ldots, \delta_N) + \sum_{n=1}^{N} a^I(\delta_n, \omega_n) - r_I \sum_{n=1}^{N} h(\delta_n) \right\} \tag{78}$$

Finally, the asymptotic output growth rate defined in (26) is given by

$$a_{\omega_1, \ldots, \omega_N} = \max_{\iota \in [0,1]} \left\{ F^{BCC}_{\iota, \omega_1, \ldots, \omega_N} \right\}. \tag{79}$$

For the case of RBA codes we can write explicitly the asymptotic input-output growth rate as

$$a^{RBA}(\iota, \omega_1, \ldots, \omega_N) = \sum_{n=1}^{N} \left[(1 - \omega_n) h\left(\frac{\iota}{2(1-\omega_n)}\right) + \omega_n h\left(\frac{\iota}{2\omega_n}\right) \right] - (N-1) h(\iota) \tag{80}$$

and thus, the asymptotic output growth rate is given by,

$$a^{RBA}(\omega_1, \ldots, \omega_N) = \max_{0 \le \iota \le \wp} \left\{ a^{(RBA}(\iota, \omega_1, \ldots, \omega_N) \right\} \tag{81}$$

where

$$\wp = \min_{n=1,\ldots,N} \left\{ 2(1 - \omega_n), 2\omega_n \right\}.$$

As an example, in Fig. 15 we plot the (non-asymptotic) growth rate of the MWEF for a rate $r = 1/2$ RBA code of information block length $K = 100$ mapped over $N = 2$ fading blocks, computed using (26) and (73). In Fig. 16 we plot the asymptotic growth rate of the same RBA ensemble. Notice that, already for block length as shows as 100 information bits, the finite-length growth rate and its corresponding asymptotic counterpart are indeed very similar.
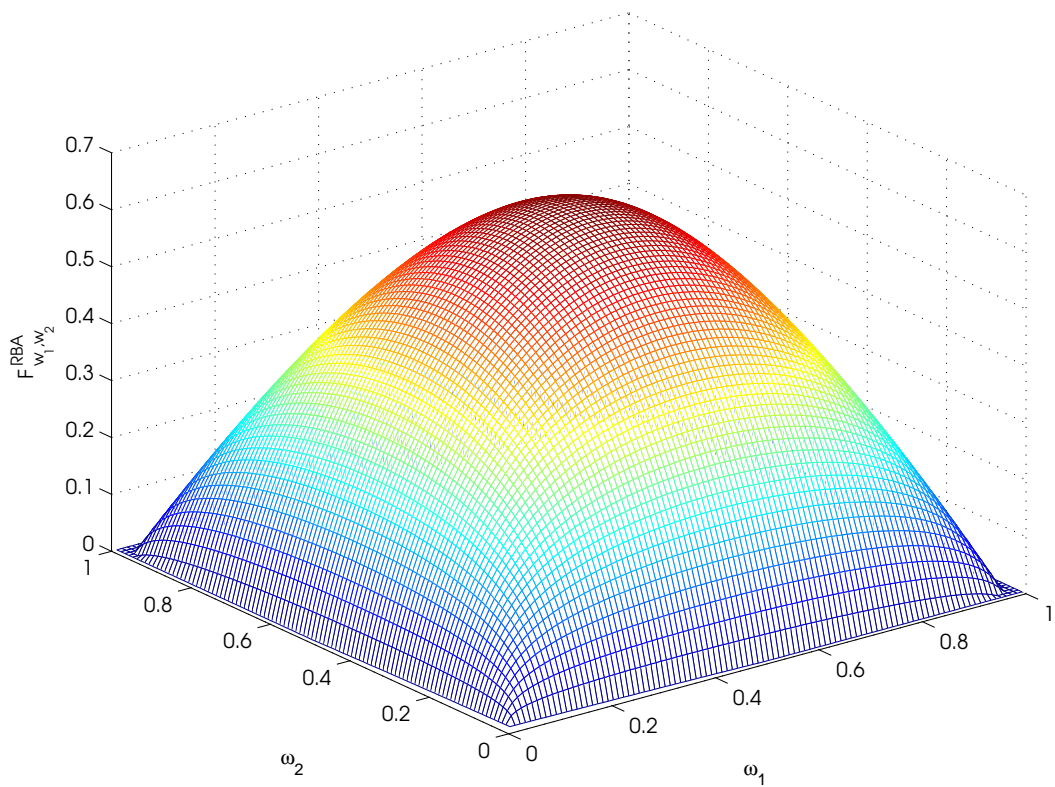
Figure 15: Exponential growth rate of the MWEF for a rate $r = 1/2$ RBA code of information block length $K = 100$ mapped over $N = 2$ fading blocks.
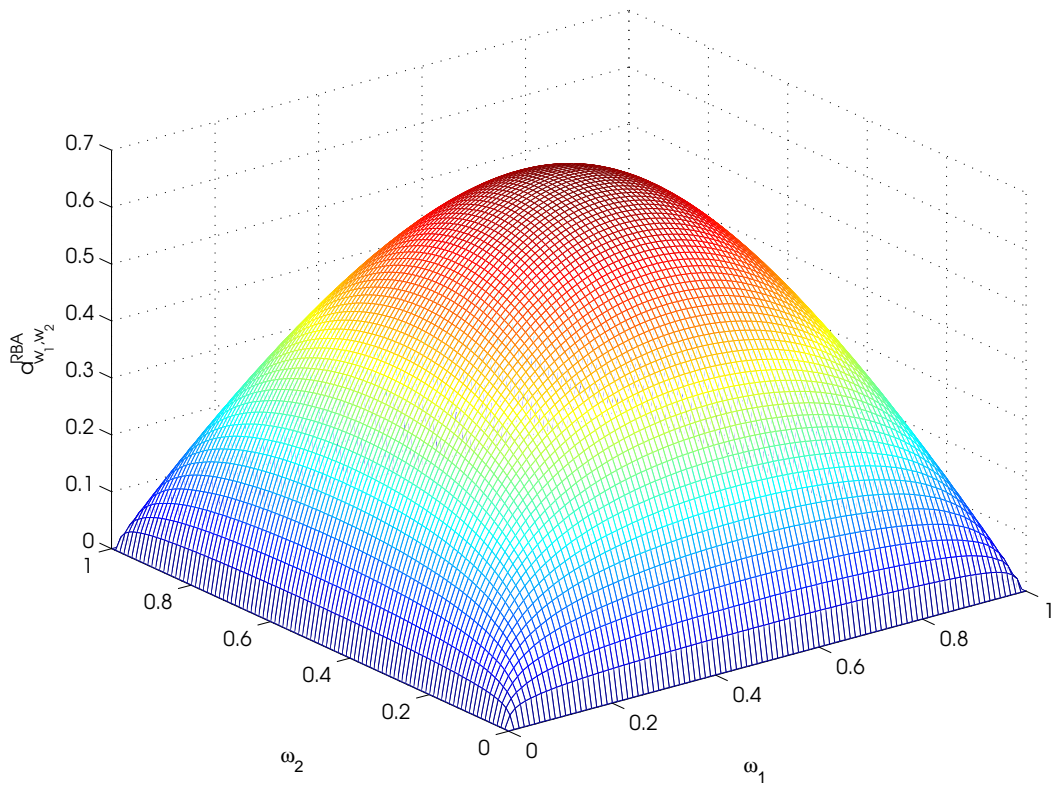
Figure 16: Asymptotic exponential growth rate $a_{\omega_1,\dots,\omega_N}^{RBA}$ for a rate $r = 1/2$ RBA code mapped over $N = 2$ fading blocks.

D   Proofs for section 4.3

Proof of Theorem 2.   Let $F(\boldsymbol{\omega})$ and $a(\boldsymbol{\omega})$ denote the exponential growth rate and the asymptotic exponential growth rate of the considered code ensemble over the BPSK modulation. For each length L, the conditional error probability given the fading vector $\mathbf{h}$ can be upperbounded by the Bhattacharyya union bound as

$$
\begin{aligned}
P_e(\rho|\mathbf{h}) &\leq \sum_{\boldsymbol{\omega}\in[0,1]^N} \exp\left(-L\left(\rho\sum_{n=1}^{N}\omega_n\gamma_n - F(\boldsymbol{\omega})\right)\right) \\
&= \exp\left(-LG_L(\rho,\boldsymbol{\gamma})\right)
\end{aligned}
\tag{82}
$$

where we define the fading power gain vector $\boldsymbol{\gamma} = (\gamma_1,\dots,\gamma_N)$ and let

$$
G_L(\rho,\boldsymbol{\gamma}) \triangleq -\frac{1}{L}\log\left\{\sum_{\boldsymbol{\omega}\in[0,1]^N}\exp\left(-L\left(\rho\sum_{n=1}^{N}\omega_n\gamma_n - F(\boldsymbol{\omega})\right)\right)\right\}
$$

The limit of the above exponent for $L\to\infty$ is given by

$$
G_\infty(\rho,\boldsymbol{\gamma}) = \inf_{\boldsymbol{\omega}\in[0,1]^N}\left\{\rho\sum_{n=1}^{N}\omega_n\gamma_n - a(\boldsymbol{\omega})\right\}
\tag{83}
$$

Let $\gamma_0 > 0$ be an arbitrary small quantity. We define the set

$$
\Gamma_{\gamma_0,\delta(N,M,R)} = \left\{\boldsymbol{\gamma}\in\mathbb{R}_+^N \;:\; \sum_{n=1}^{N}\mathbb{1}\{\gamma_n\leq\gamma_0\} \geq \delta(N,M,R)\right\}
\tag{84}
$$

The set $\Gamma_{\gamma_0,\delta(N,M,R)}$ contains the "bad fading vectors", i.e., the fading vectors having $\delta(N,M,R)$ or more small components (smaller than γ_0).

Using fading statistical independence, the property of the fading cdf $\Pr(\gamma_n \leq \gamma_0) \triangleq p_0 = \gamma_0 + o(\gamma_0)$ (e.g., valid for Rayleigh fading) and standard bounds on the tail of the binomial distribution, we can write

$$
\begin{aligned}
\Pr(\Gamma_{\gamma_0,\delta(N,M,R)}) &= \sum_{k=\delta(N,M,R)}^{N}\binom{N}{k}p_0^k(1-p_0)^{N-k} \\
&\leq \mathcal{K}_1\gamma_0^{\delta(N,M,R)} + o(\gamma_0^{\delta(N,M,R)})
\end{aligned}
\tag{85}
$$

for a suitable constant $\mathcal{K}_1 > 0$.

Using (82), (83), (84) and (85), indicating by $f_{\boldsymbol{\gamma}}(\mathbf{z})$ the pdf of $\boldsymbol{\gamma}$ we can write, for sufficiently large ρ and L,

$$
\begin{aligned}
P_e(\rho) &\leq \mathbb{E}\left[\min\left\{1, \exp(-LG_L(\rho, \boldsymbol{\gamma}))\right\}\right] \\
&\leq \int_{\Gamma_{\gamma_0, \delta(N,M,R)}} f_{\boldsymbol{\gamma}}(\mathbf{z})d\mathbf{z} + \int_{\Gamma^c_{\gamma_0, \delta(N,M,R)}} \exp(-LG_L(\rho, \boldsymbol{\gamma}))f_{\boldsymbol{\gamma}}(\mathbf{z})d\mathbf{z} \\
&\leq \mathcal{K}_1 \gamma_0^{\delta(N,M,R)} + \int_{\Gamma^c_{\gamma_0, \delta(N,M,R)}} \exp(-LG_\infty(\rho, \boldsymbol{\gamma}))f_{\boldsymbol{\gamma}}(\mathbf{z})d\mathbf{z} + \epsilon_L + o(\gamma_0^{\delta(N,M,R)}) \quad (86)
\end{aligned}
$$

where $\epsilon_L \to 0$ as $L \to \infty$.

Recall that $\mathcal{W}(N, k) \subset \{0, 1\}^N$ is the set of binary vectors of length N with Hamming weight not smaller than $N - k + 1$. For any $\boldsymbol{\gamma} \in \Gamma^c_{\gamma_0, \delta(N,M,R)}$ there exists $\mathbf{x} \in \mathcal{W}(N, \delta(N, M, R))$ such that $\gamma_0 \mathbf{x} \leq \boldsymbol{\gamma}$ (componentwise). In fact, it is sufficient to replace each $\gamma_n < \gamma_0$ in $\boldsymbol{\gamma}$ by 0, and each $\gamma_n \geq \gamma_0$ in $\boldsymbol{\gamma}$ by γ_0. Since by definition there are at most $\delta(N, M, R) - 1$ components of $\boldsymbol{\gamma}$ smaller than γ_0, the resulting vector has Hamming weight at least $N - \delta(N, M, R) + 1$ and therefore (up to the scaling by γ_0) it belongs to $\mathcal{W}(N, \delta(N, M, R))$.

For $\boldsymbol{\gamma} \in \Gamma^c_{\gamma_0, \delta(N,M,R)}$, it follows from the observation above that $G_\infty(\rho, \boldsymbol{\gamma})$ is lowerbounded by

$$
G_\infty(\rho, \boldsymbol{\gamma}) \geq \inf_{\mathbf{x} \in \mathcal{W}(N, \delta(N,M,R))} \inf_{\boldsymbol{\omega} \in [0,1]^N} \left\{ \rho\gamma_0 \sum_{n=1}^{N} x_n \omega_n - a(\boldsymbol{\omega}) \right\} \triangleq G_{\infty, LB}(\rho, \gamma_0) \quad (87)
$$

Define $\widehat{s}$ as in Theorem 2 (eq. (28)) and suppose that, for the given ensemble, $\widehat{s}$ is finite. Then, we let $\gamma_0 = \widehat{s}/\rho$ and, continuing the chain of inequalities (86),

$$
\begin{aligned}
P_e(\rho) &\leq \mathcal{K}_1 \gamma_0^{\delta(N,M,R)} + \exp(-LG_{\infty,LB}(\rho, \gamma_0)) \int_{\Gamma^c_{\gamma_0, \delta(N,M,R)}} f_{\boldsymbol{\gamma}}(\mathbf{z})d\mathbf{z} + \epsilon_L + o(\gamma_0^{\delta(N,M,R)}) \\
&\leq \mathcal{K}_1 \left(\frac{\widehat{s}}{\rho}\right)^{\delta(N,M,R)} + \epsilon'_L + o(\rho^{-\delta(N,M,R)}) \\
&\leq \mathcal{K}\rho^{-\delta(N,M,R)} + \epsilon'_L + o(\rho^{-\delta(N,M,R)}) \quad (88)
\end{aligned}
$$

where $\epsilon'_L \to 0$ as $L \to \infty$. This shows the weak goodness of the ensemble.

Proof of Corollary 2. For ensembles of BICM codes with random bit interleaving we use the Bhattacharyya union bound with Bhattacharyya factor $B_n(\rho, \mu, \mathcal{X})$ defined in (21). Following the same approach as in the proof of Theorem 2, we see that the upper bound to the error probability takes on the same form of (86) – (88) provided that we replace γ_n by ζ_n defined in (22). For large

ρ, we have that $\zeta_n \to \frac{d_{\min}^2}{4} \gamma_n$ where $d_{\min}$ is the minimum Euclidean distance of the normalized signal constellation $\mathcal{X}$ (see (22) and [31]). The scaling by the constant factor $d_{\min}^2/4$ (independent of ρ and of L) involves at most a scaling of the SNR threshold $\widehat{s}$. Therefore, weak goodness of the underlying binary code implies weak goodness of the corresponding BICM ensemble with random bit interleaving.

Trellis terminated CCs do not satisfy Theorem 2. For simplicity, we consider trellis terminated CCs of rate k/N, such that in each trellis transition the N coded bits are sent to the N fading blocks (this defines the blockwise partitioning of the CC codeword). Then, output blocks of length L correspond exactly to trellises of L trellis sections. Let

$$F_\epsilon^L(\boldsymbol{\omega}) \triangleq \frac{1}{L} \log \left| \mathcal{S}_\epsilon^L(\boldsymbol{\omega}) \right| \tag{89}$$

define the length-L MWEF exponential growth rate, such that $a(\boldsymbol{\omega})$ defined in (26) is given by $\lim_{\epsilon \to 0} \lim_{L \to \infty} F_\epsilon^L(\omega)$.

A *simple* error path in the trellis is a path that leaves state 0 and remerges for the first time to state 0 after a certain number of trellis sections, while it coincides with the all-zero path everywhere else. Take a fixed Hamming weight $\beta < \infty$ such that there exist a simple error path of length $\ell(\beta)$ (independent of L) in the code having Hamming weights per block $|w_n - \beta| \leq \epsilon L/2$, for all $n = 1, \ldots, N$. Such simple error path exists for any trellis terminated CC with given encoder. For example, we can take the minimum-length event corresponding to the free Hamming distance of the code. It follows that, for all $\epsilon > 0$, sufficiently large L and $\boldsymbol{\omega}$ such that $|\omega_n - \frac{\beta}{L}| \leq \epsilon/2$, the lower bound

$$\left| \mathcal{S}_\epsilon^L(\boldsymbol{\omega}) \right| \geq L - \ell(\beta) \tag{90}$$

holds. In fact, $\mathcal{S}_\epsilon^L(\boldsymbol{\omega})$ must contain all the simple error events of length at most $\ell(\beta)$ starting at the $L - \ell(\beta) + 1$ distinct positions in the length-L trellis.

By continuity, we can write the sufficient condition of Theorem 2 as

$$\widehat{s} = \lim_{\epsilon \to 0} \lim_{L \to \infty} \sup_{\mathbf{x} \in \mathcal{W}(N, \delta(N,M,R))} \sup_{\boldsymbol{\omega} \in [0,1]^N} \frac{F_\epsilon^L(\boldsymbol{\omega})}{\sum_{n=1}^N x_n \omega_n} \tag{91}$$

For any $\mathbf{x} \in \mathcal{W}(N, \delta(N, M, R))$, take $\boldsymbol{\omega}$ in a box of side ϵ around the point $\frac{\beta}{L} \mathbf{1} \in [0,1]^N$. The lower bound (90) implies that

$$\widehat{s} \geq \lim_{\epsilon \to 0} \lim_{L \to \infty} \frac{\frac{1}{L} \log L + O(\frac{1}{L})}{\frac{\beta N}{L} + \epsilon} \tag{92}$$

which clearly shows that trellis terminated CCs cannot satisfy Theorem 2.

More in general, any code ensemble such that the MWEF $A_{w_1,\ldots,w_N}$ increases *linearly* with the block length L for some Hamming weight vector $\mathbf{w}$ such that $\|\mathbf{w}\|_\infty \leq \beta < \infty$ cannot satisfy Theorem 2. These code ensembles have an infinite ML decoding threshold [16] in the standard AWGN channel.

Asymptotic union bounds (30) and (31). The M&L Bhattacharyya union bound for a code over the BPSK signal set can be written as

$$P_e(\rho) \leq \mathbb{E}\left[\min\left\{1, \sum_{\boldsymbol{\omega}\in[0,1]^N} e^{-L\left(\rho\sum_{n=1}^N \gamma_n\omega_n - F(\boldsymbol{\omega})\right)}\right\}\right]. \tag{93}$$

Since $\min(1, f(x))$ is continuous in x for continuous f and $\min(1, f(x)) \leq 1$, we can apply the dominated convergence theorem [37] and write,

$$\begin{aligned}
\lim_{L\to\infty} \; & \mathbb{E}\left[\min\left\{1, \sum_{\boldsymbol{\omega}\in[0,1]^N} e^{-L\left(\rho\sum_{n=1}^N \gamma_n\omega_n - F(\boldsymbol{\omega})\right)}\right\}\right] \\
= \;\; & \mathbb{E}\left[\min\left\{1, \lim_{L\to\infty} \sum_{\boldsymbol{\omega}\in[0,1]^N} e^{-L\left(\rho\sum_{n=1}^N \gamma_n\omega_n - F(\boldsymbol{\omega})\right)}\right\}\right] \\
= \;\; & \mathbb{E}\left[\min\left\{1, \lim_{L\to\infty} \sum_{\boldsymbol{\omega}\in[0,1]^N} e^{-L\left(\rho\sum_{n=1}^N \gamma_n\omega_n - a(\boldsymbol{\omega})\right)}\right\}\right]
\end{aligned} \tag{94}$$

The factor multiplying L in the exponent of the RHS of (94) is positive for a given channel realization $\boldsymbol{\gamma}$ if

$$\widehat{\rho} \triangleq \max_{\boldsymbol{\omega}\in[0,1]^N} \frac{a(\boldsymbol{\omega})}{\displaystyle\sum_{n=1}^N \gamma_n\omega_n} < \rho \tag{95}$$

Conditioning with respect to $\boldsymbol{\gamma}$ we have that, in the limit of large L, $P_e(\rho|\boldsymbol{\gamma}) \to 0$ if $\rho > \widehat{\rho}$ while $P_e(\rho|\boldsymbol{\gamma}) \leq 1$ otherwise. It follows that in the limit for $L \to \infty$ the M&L Bhattacharyya bound takes on the form (30).

In the case of BICM codes with random bit interleaving (under the symmetrized mapping assumption), we can use the same M&L Bhattacharyya bound by replacing γ_n with ζ_n defined in (22), and (31) follows.

References

[1] L. H. Ozarow, S. Shamai and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. on Vehicular Tech.*, vol. 43, no. 2, pp. 359–378, May 1994.

[2] E. Biglieri, J. Proakis and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Trans. on Inform. Theory*, vol. 44, no. 6, pp. 2619 –2692, Oct. 1998.

[3] R. Knopp and P. Humblet, "On coding for block fading channels," *IEEE Trans. on Inform. Theory*, vol. 46, no. 1, pp. 1643–1646, July 1999.

[4] R. Knopp, *Coding and Multiple-Access over Fading Channels*, Ph.D. thesis, Ecole Polytechnique Federale de Lausanne (EPFL) and Institut Eurecom, 1997.

[5] E. Malkamaki and H. Leib, "Evaluating the performance of convolutional codes over block fading channels," *IEEE Trans. on Inform. Theory*, vol. 45, no. 5, pp. 1643–1646, Jul. 1999.

[6] M. Chiani, A. Conti and V. Tralli, "Further results on convolutional code search for block fading channels," *to appear IEEE Trans. on Inform. Theory*, 2004.

[7] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. on Inform. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.

[8] G. Caire, G. Taricco and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. on Inform. Theory*, vol. 45, no. 5, pp. 1468–1489, July 1999.

[9] E. Malkamaki and H. Leib, "Coded diversity on block-fading channels," *IEEE Trans. on Inform. Theory*, vol. 45, no. 2, pp. 771–781, March 1999.

[10] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. on Inform. Theory*, vol. 49, no. 5, May 2003.

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

[12] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[13] S. ten Brink, "Designing iterative decoding schemes with the extrinsic information transfer chart," *AEU Int. J. Electron. Commun.*, vol. 54, no. 6, pp. 389–398, Dec. 2000.

[14] I. Sason and S. Shamai, "Improved upper bounds on the ml decoding error probability of parallel and serial concatenated turbo codes via their ensemble distance spectrum," *IEEE Trans. on Inform. Theory*, vol. 46, no. 1, pp. 24–47, Jan. 2000.

[15] I. Sason and S. Shamai, "On improved bounds on the decoding error probability of block codes over interleaved fading channels, with applications to turbo-like codes," *IEEE Trans. on Inform. Theory*, vol. 47, no. 6, pp. 2275–2299, Sept. 2001.

[16] D. Divsalar, "A simple tight bound on error probability of block codes with application to turbo codes," *JPL TMO Progress Report 42-139*, November 1999.

[17] D. Divsalar and E. Biglieri, "Upper bounds to error probabilities of coded systems beyond the cutoff rate," *IEEE Trans. on Commun.*, vol. 51, no. 12, pp. 2011–2018, December 2003.

[18] S. Shamai I. Sason and D. Divsalar, "Tight exponential upper bounds on the ml decoding error probability of block codes over fully interleaved fading channels," *IEEE Trans. on Commun.*, vol. 51, no. 8, pp. 1296–1305, August 2003.

[19] F. R. Kschischang, B. J. Frey, and H-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[20] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, New York: Dover Press, 1972.

[21] W.H.Press, S.A.Teukolsky, W.T.Vetterling and B.P. Flannery, *Numerical Recepies in C*, Cambridge University Press, also available at `http://www.nr.com`, 1992.

[22] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, 1991.

[23] J. Proakis, *Digital Communications*, McGraw-Hill, 1995.

[24] D. Divsalar, H. Jin and R. J. McEliece, "Coding theorems for turbo-like codes," in *Proc. 1998 Allerton Conference*, October 1998.

[25] L. Bahl, J. Cocke, F. Jelinek and J. Raviv, "Optimal decoding of linear codes for minimizing the symbol error rate," *IEEE Trans. on Inform. Theory*, vol. 20, pp. 284–287, Mar. 1974.

[26] S. Wilson Y. Leung and J. Ketchum, "Multifrequency trellis coding with low delay for fading channels," *IEEE Trans. on Commun.*, vol. 41, no. 10, pp. 1450–1458, October 1993.

[27] M-K. Byun, D. Park and B.G Lee, "On the performance analysis of space-time codes in quasi-static fading channels," *IEEE Trans. on Inform. Theory*, vol. 50, no. 11, pp. 2865–2873, Nov. 2004.

[28] C.F. Leanderson, G. Caire and O. Edfors, "Performance of incremental redundancy schemes with turbo codes," in *Proc. Radiovetenskap och Kommunikation 2002, Stockholm, Sweden*, 2002.

[29] A. Guillén i Fàbregas and G. Caire, "Turbo-like codes over the block-fading channel: code design and construction," in *3rd International Symposium on Turbo Codes and Related Topics, Brest, France*, Sept. 2003.

[30] A. Viterbi, *CDMA : Principles of Spread Spectrum Communication*, Prentice Hall, 1995.

[31] A. Guillén i Fàbregas, A. Martinez and G. Caire, "Error probability of Bit-interleaved Coded Modulation using the Gaussian Approximation," *in Proc. 38th Annual Conference on Information Sciences and Systems, CISS 2004, Princeton, NJ, March 2004*.

[32] I. Sason, E. Telatar and R. Urbanke, "On the asymptotic input-output weight distributions and thresholds of convolutional and turbo-like encoders," *IEEE Trans. on Inform. Theory*, vol. 48, no. 12, pp. 3052–3061, December 2002.

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

[34] S. Benedetto, D. Divsalar, G. Montorsi and F. Pollara, "Serial concatenated of interleaved codes: Performance analysis, design and iterative decoding," *IEEE Trans. on Inform. Theory*, vol. 44, no. 3, pp. 909–926, May 1998.

[35] H. Jin, *Analysis and Design of Turbo-like Codes*, Ph.D. thesis, California Institute of Technology, 2001.

[36] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Number 38 in Applications of Mathematics. Springer Verlag, 2nd edition, April 1998.

[37] R. Durrett, *Probability: Theory and Examples*, Duxbury Press, 1996.

[38] A. J. Viterbi and Jim K. Omura, *Principles of Digital Communication and Coding*, McGraw-Hill, 1979.