

Bounds on the Second-Order Coding Rate of the MIMO Rayleigh Block-Fading Channel

Jakob Hoydis
Bell Laboratories, Alcatel-Lucent
Stuttgart, Germany
jakob.hoydis@alcatel-lucent.com

Romain Couillet
Dept. of Telecommunications
SUPELEC, France
romain.couillet@supelec.fr

Pablo Piantanida
Dept. of Telecommunications
SUPELEC, France
pablo.piantanida@supelec.fr

Abstract—We study the second-order coding rate of the multiple-input multiple-output (MIMO) Rayleigh block-fading channel via statistical bounds from information spectrum methods and random matrix theory. Based on an asymptotic analysis of the mutual information density which considers the simultaneous growth of the block length n and the number of transmit and receive antennas K and N , we derive closed-form upper and lower bounds on the optimal average error probability when the code rate is within $\mathcal{O}(1/\sqrt{nK})$ of the asymptotic capacity. A Gaussian approximation is then used to establish an upper bound on the error probability for arbitrary code rates which is shown by simulations to be accurate for small N , K , and n .

I. INTRODUCTION

The channel capacity describes the maximal rate at which data exchange with vanishing error probability is possible, provided that the length of each codeword is allowed to grow without limit. By focusing on this asymptotic limit, the theory of information ignores the role of delay as an important parameter. Although the channel capacity and the related Shannon theory constitute a scientific success story [1], in many real-world applications, relatively short block lengths are required due to either delay and/or complexity constraints. Thus, it is important to analyze the required back-off from capacity to guarantee a certain error probability for a given block length. Unfortunately, in the finite block-length regime, there are no exact *tractable* formulas to facilitate the analysis. This is in particular the case for practical quasi-static fading multiple-input multiple-output (MIMO) channels.

Among the first to investigate bounds on the error probability for a given coding rate were Feinstein and Shannon [2], [1] who established the convergence of the optimal rate to the capacity with growing block length. These bounds are based on the fundamental relation between the *information density* [3], the coding rate, and the error probability. The optimal exponential rate of decrease of the error probability was derived by Gallager [4]. However, his result does not provide the best upper bound on the average error probability in channel coding when the transmission rate is larger than the capacity [5]. Focusing on the characterization of channel capacity via this statistical approach, Strassen [6] derived a general expression for the discrete memoryless channel with unconstrained inputs, where the notion of normal approximation of the second-order coding rate was introduced for the first time. Unfortunately, Strassen's approach cannot

be generalized to channels with input constraints (e.g., the Gaussian and the fading MIMO channels). Further work on the asymptotic block-length regime via information-spectrum methods established the most general capacity formula [7] which required a novel statistical bound [8], [9] in the converse proof. The authors of [10], [5] investigated the finite block-length regime and the second-order coding rate of several channel models in presence of cost (or input) constraints. Along the same lines, the scalar additive white Gaussian noise (AWGN) block-fading channel was addressed in the coherent and non-coherent settings in [11] and [12], respectively.

In this paper, we investigate the finite block-length regime of the MIMO Rayleigh block-fading channel. The case of study is made difficult because the channel does not satisfy the ergodic requirements to apply the usual tools for the analysis of the second-order coding rate. Following our first contribution in [13], we therefore study the asymptotic behavior of the error probability when the coding rate is a perturbation of order $\mathcal{O}(1/\sqrt{nK})$ of the asymptotic capacity while the block-length n , and the number of transmit and receive antennas K and N , respectively, grow infinitely large at the same rate. In this asymptotic regime, we establish a new lower and upper bound on the optimal average error probability and apply them to obtain an upper bound approximation for finite n .

II. SYSTEM MODEL AND PROBLEM STATEMENT

Consider the following MIMO fading channel:

$$\mathbf{y}_t = \frac{1}{\sqrt{K}} \mathbf{H}^n \mathbf{x}_t + \sigma \mathbf{w}_t, \quad t = \{1, \dots, n\} \quad (1)$$

where $\mathbf{y}_t \in \mathbb{C}^N$ is the channel output at time t , \mathbf{H}^n is a realization of the random channel matrix $H^n \in \mathbb{C}^{N \times K}$ whose entries are independent and identically distributed (i.i.d.) $\mathcal{CN}(0, 1)$ and the index n is used to remind that \mathbf{H}^n is constant for the duration of n channel uses, \mathbf{x}_t is the realization of the random channel input $x_t \in \mathbb{C}^{K \times 1}$ at time t , and $\sigma \mathbf{w}_t$ is the realization of the random noise vector $\sigma w_t \in \mathbb{C}^N$ at time t whose entries are i.i.d. $\mathcal{CN}(0, \sigma^2)$. The transmitter has only statistical knowledge about H^n while the receiver knows H^n perfectly. In particular, we will assume H^n , x_t , and w_t to be independent for each t . We define $\mathbf{X}^n = (\mathbf{x}_1 \cdots \mathbf{x}_n) \in \mathbb{C}^{K \times n}$, $\mathbf{W}^n = (\mathbf{w}_1 \cdots \mathbf{w}_n) \in \mathbb{C}^{N \times n}$, and $\mathbf{Y}^n = (\mathbf{y}_1 \cdots \mathbf{y}_n) \in \mathbb{C}^{N \times n}$, and their random counterparts $X^n = (x_1 \cdots x_n) \in \mathbb{C}^{K \times n}$,

$W^n = (w_1 \cdots w_n) \in \mathbb{C}^{N \times n}$, and $Y^n = (y_1 \cdots y_n) \in \mathbb{C}^{N \times n}$. For $\Gamma > 0$, we further define

$$\mathcal{S}_\Gamma^n \triangleq \left\{ \mathbf{X}^n \in \mathbb{C}^{K \times n} \mid \frac{1}{nK} \text{tr} \mathbf{X}^n (\mathbf{X}^n)^H \leq \Gamma \right\} \quad (2)$$

i.e., the set of inputs \mathbf{X}^n with energy constraint Γ .

The *mutual information density* of $\mathbb{P}_{Y^n|X^n}$, i.e., the probability measure of Y^n conditioned on X^n , is defined by [3]

$$I_{N,K}^{(n)} \triangleq \frac{1}{nK} \log \frac{\mathbb{P}_{Y^n|X^n}(dY^n|X^n)}{\mathbb{P}_{Y^n}(dY^n)} \quad (3)$$

where, for a given \mathbf{X}^n , $\mathbb{P}_{Y^n|X^n}(\cdot|\mathbf{X}^n)/\mathbb{P}_{Y^n}$ is the Radon–Nykodym derivative of $\mathbb{P}_{Y^n|X^n}(\cdot, \mathbf{X}^n)$ with respect to \mathbb{P}_{Y^n} if $\mathbb{P}_{Y^n|X^n}(\cdot, \mathbf{X}^n) \ll \mathbb{P}_{Y^n}$ and is set to ∞ otherwise.

Of particular importance is the case of independent Gaussian inputs, i.e., $\mathbf{x}_t \sim \mathcal{CN}(\mathbf{0}, \Gamma \mathbf{I}_K)$, for which the mutual information density takes the form

$$I_{N,K}^{(n)} = C_{N,K} + R_{N,K}^{(n)} \quad (4)$$

where

$$C_{N,K} \triangleq \frac{1}{K} \log \det \left(\mathbf{I}_N + \frac{\Gamma}{\sigma^2 K} H^n (H^n)^H \right) \quad (5)$$

$$R_{N,K}^{(n)} \triangleq \frac{1}{nK} \text{tr} \left[\left(\frac{\Gamma}{K} H^n (H^n)^H + \sigma^2 \mathbf{I}_N \right)^{-1} Y^n (Y^n)^H - W^n (W^n)^H \right]. \quad (6)$$

Note that such inputs do not respect the energy constraint (2).

Definition 1: A $(P_e^{(n)}, M_n, \Gamma)$ -code \mathcal{C}_n for the channel model (1) with power constraint (2) consists of:

- An encoder mapping: $\varphi: \mathcal{M}_n \mapsto \mathbb{C}^{K \times n}$.
The transmitted symbols are $\mathbf{X}_m^n = \varphi(m) \in \mathcal{S}_\Gamma^n$ for every message m uniformly distributed over the set $\mathcal{M}_n = \{1, \dots, M_n\}$ of messages.
- A decoder mapping: $\phi_{\mathbf{H}^n}: \mathbb{C}^{N \times n} \mapsto \mathcal{M}_n \cup \{e\}$
which produces the decoder's decision $\hat{m} = \phi_{\mathbf{H}^n}(\mathbf{Y}_m^n)$, $\mathbf{Y}_m^n = \frac{1}{\sqrt{K}} \mathbf{H}^n \mathbf{X}_m^n + \sigma \mathbf{W}^n$, on the transmitted message m , or the error event e .

For a code \mathcal{C}_n with block length n , codebook size M_n , encoder φ , and decoder $\{\phi_{\mathbf{H}^n}\}_{\mathbf{H}^n \in \mathbb{C}^{N \times K}}$, the *average error probability* is defined as

$$P_e^{(n)} = P_e^{(n)}(\mathcal{C}_n) \triangleq \frac{1}{M_n} \sum_{m=1}^{M_n} \Pr[\hat{m} \neq m | m] \quad (7)$$

where the probability is taken over the random variables W^n and H^n .

Let $\text{supp}(\mathcal{C}_n)$ denote the *codebook* $\{\varphi(1), \dots, \varphi(M_n)\}$. The *optimal average error probability for the rate R* with energy constraint Γ is defined as

$$\mathbb{P}_e^{(n)}(R) \triangleq \inf_{\mathcal{C}_n: \text{supp}(\mathcal{C}_n) \subseteq \mathcal{S}_\Gamma^n} \left\{ P_e^{(n)}(\mathcal{C}_n) \mid \frac{\log M_n}{nK} \geq R \right\}. \quad (8)$$

The exact characterization of $\mathbb{P}_e^{(n)}(R)$ for fixed n, K, N is generally intractable. In the case of stationary discrete memoryless single-input single-output (SISO) additive white

Gaussian noise (AWGN) channels, there have been recent efforts [5] (see also [10] and [6]) to establish error probability approximations when the coding rate is within $\mathcal{O}(1/\sqrt{n})$ of the ergodic capacity as n grows large. However, immediate extensions of these results to block-fading channels, as considered in this paper, are not possible because the ergodic capacity for these channels is not defined. To circumvent this issue, [11] assumes coding over a large number of independent realizations of increasingly large block-fading channels, which makes the overall channel ergodic (the article is actually restricted to SISO channels but would easily adapt to the MIMO case). In the present article, we take the approach of inducing ergodicity by growing the channel matrix dimensions. Indeed, letting $K, N \rightarrow \infty$, the i.i.d. structure of H^n makes the channel ergodic in the limit. Precisely, we assume here that $K, N, n \rightarrow \infty$ while $\frac{n}{K} = \beta$ and $\frac{N}{K} = c$ for some constants $\beta, c > 0$. This will be denoted by $n \xrightarrow{(\beta, c)} \infty$. In this limiting regime, the per-antenna capacity of the channel converges for almost every channel realization to an asymptotic limit C . We can then characterize the error probability in the second-order coding rate, i.e., when the coding rate is within $\mathcal{O}(1/\sqrt{nK})$ of the limiting capacity C . With these assumptions, similar to [5], [10], $\mathbb{P}_e^{(n)}(R)$ is replaced by the following tractable limiting error probability:

Definition 2: The *optimal average error probability for the second-order coding rate r* with input energy constraint Γ is

$$\mathbb{P}_e(r|\beta, c, \Gamma) \triangleq \inf_{\substack{\{\mathcal{C}_n\}_{n=1}^{\infty} \\ \text{supp}(\mathcal{C}_n) \subseteq \mathcal{S}_\Gamma^n}} \left\{ \limsup_{n \xrightarrow{(\beta, c)} \infty} P_e^{(n)}(\mathcal{C}_n) \mid \liminf_{n \xrightarrow{(\beta, c)} \infty} \sqrt{nK} \left(\frac{1}{nK} \log M_n - C \right) \geq r \right\} \quad (9)$$

where

$$C = \liminf_{n \xrightarrow{(\beta, c)} \infty} \mathbb{E}[C_{N,K}]. \quad (10)$$

Without loss of generality and for simplicity, we take $\Gamma = 1$ and denote $\mathcal{S}^n \triangleq \mathcal{S}_1^n$ and $\mathbb{P}_e(r|\beta, c) \triangleq \mathbb{P}_e(r|\beta, c, 1)$.

III. MAIN RESULTS

A. Bounds on the optimal average error probability

Theorem 1 (Bounds on $\mathbb{P}_e(r|\beta, c)$): For $x > 0$ and $c > 0$, define

$$\delta_0(x) \triangleq \frac{c-1}{2x} - \frac{1}{2} + \frac{\sqrt{(1-c+x)^2 + 4cx}}{2x}$$

with derivative

$$\delta_0'(x) = -\frac{\delta_0(x)(1+\delta_0(x))}{1-c+x+2x\delta_0(x)}$$

and denote, for $\sigma^2 > 0$,

$$\theta_- \triangleq \left[-\beta \log \left(1 - \frac{1}{c} \frac{\delta_0(\sigma^2)^2}{(1+\delta_0(\sigma^2))^2} \right) + (c + \sigma^4 \delta_0'(\sigma^2)) \right]^{\frac{1}{2}}$$

$$\theta_+ \triangleq \left[-\beta \log \left(1 - \frac{1}{c} \frac{\delta_0(\sigma^2)^2}{(1+\delta_0(\sigma^2))^2} \right) + 2(c - \sigma^2 \delta_0(\sigma^2)) \right]^{\frac{1}{2}}.$$

Then, for the channel model (1) with unit input energy constraint, the optimal average error probability $\mathbb{P}_e(r|\beta, c)$ satisfies:

- If $r \leq 0$, $\Phi\left(\frac{r}{\theta_-}\right) \leq \mathbb{P}_e(r|\beta, c) \leq \Phi\left(\frac{r}{\theta_+}\right)$
- If $r > 0$, $\frac{1}{2} \leq \mathbb{P}_e(r|\beta, c) \leq \Phi\left(\frac{r}{\theta_+}\right)$.

Proof: The full proof is provided in [14]. Although the theorem provides *closed-form bounds* on $\mathbb{P}_e(r|\beta, c)$, it must not hide the fact that its proof is quite involved. Our approach follows closely Hayashi's method [5]. The major difficulty and technical contribution lie in the thorough analysis of the asymptotic statistics of $I_{N,K}^{(n)}$ under different assumptions on the distribution \mathbb{P}_{X^n} of X^n . We make extensive use of tools from random matrix theory, especially the characteristic function approach due to Pastur, see, e.g., [15], along with the integration by parts formula for Gaussian vectors and the Poincaré–Nash inequality. In contrast to the usual setting of large random matrix theory, because of (2), $X^n (X^n)^H$ is only bounded in trace rather than in spectral norm. This complicates the analysis at many occasions. In a nutshell, for the lower bound, we prove that $I_{N,K}^{X^n}$ (where $I_{N,K}^{X^n}$ is $I_{N,K}^{(n)}$ conditioned on a particular \mathbf{X}^n) has a variance which scales as $\mathcal{O}(1 + \frac{1}{K} \text{tr}(\mathbf{A}^n)^2)$ with $\mathbf{A}^n = \mathbf{I} - \frac{1}{n} \mathbf{X}^n (\mathbf{X}^n)^H$, which can then grow infinitely large or not, depending on $\mathbf{X}^n \in \mathcal{S}^n$. We then show that if $1 - \varepsilon \leq \frac{1}{K} \text{tr} \mathbf{X}^n (\mathbf{X}^n)^H \leq 1$, for some ε small, $I_{N,K}^{(n)}$, when properly centered and scaled, satisfies a central limit theorem (CLT). The minimization of the corresponding asymptotic error probability then brings the limiting mean C and variance θ_-^2 for $r < 0$ or ∞ for $r \geq 0$. For the upper bound, we use a sequence of Gaussian inputs X^n with variance less than but arbitrarily close to 1. In this case, we prove that the random variable $I_{N,K}^{(n)}$ satisfies a CLT with asymptotic mean C and variance θ_+^2 . ■

Theorem 1 indicates that, for sufficiently large channel dimensions and block length, the optimal error probability when coding close to the ergodic capacity is contained within two explicit bounds which depend only on c , β , and σ^2 . In the AWGN scenario of [5], [10], the corresponding bounds were found to depend only on σ^2 . Note that, for code rates above the ergodic capacity limit (i.e., for $r > 0$), the lower bound is very pessimistic and can be far from the upper bound. In contrast, for $r < 0$, both bounds are generally very close to one-another. This can be seen from Fig. 1 which depicts the bounds on the optimal average error probability for varying second-order coding rates r and different SNR values (defined as $\text{SNR} = \sigma^{-2}$), including also the extreme high- and low-SNR cases. We choose $c = 2$ and $\beta = 16$. For negative second-order coding rates, the gap between the upper and lower bounds is rather small and decreases with either growing r or decreasing SNR.

Remark 1: One can show that for every $c, \beta, \sigma^2 > 0$,

$$\begin{cases} \Phi\left(\frac{r}{\theta_+}\right) > \Phi\left(\frac{r}{\theta_-}\right) & , r < 0 \\ \Phi\left(\frac{r}{\theta_+}\right) > \frac{1}{2} & , r > 0. \end{cases} \quad (11)$$

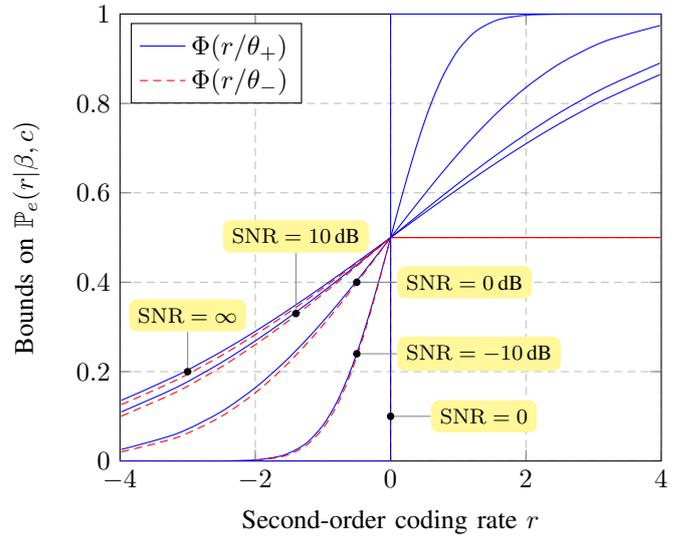


Fig. 1. Bounds on the optimal average error probability as a function of the second-order coding rate r for different SNRs and the parameters $c = 2$ and $\beta = 16$.

Apart for $r = 0$, the lower and upper bounds on the optimal average error probability are therefore never equal. This is in sharp contrast to [5], [10] where, for SISO AWGN channels, the bounds are proved to be equal. The reason for this discrepancy lies in the presence of the random channel H^n which naturally induces a dependence of the second order statistics of $I_{N,K}^{(n)}$ on the “fourth order moment” $\mathbb{E}[K^{-1} \text{tr}(n^{-1} X^n (X^n)^H)^2]$ of \mathbb{P}_{X^n} . The weak lower bound $1/2$ for $r > 0$ is in particular a consequence of the impossibility to bound the fourth order moment of \mathbb{P}_{X^n} from above under the sole constraint (2). In [5], [10], only (scalar) second order moments of \mathbb{P}_{X^n} play a role in the second order statistics of $I_{N,K}^{(n)}$. These are easily controlled by (2).

Remark 2: In [15, Theorem 1], it was shown that

$$\mathbb{E}[C_{N,K}] = C + \mathcal{O}(n^{-2}) \quad (12)$$

where the limiting mutual information C is given as

$$C = \log(1 + \delta_0) + c \log\left(1 + \frac{1}{\sigma^2(1 + \delta_0)}\right) - \frac{\delta_0}{1 + \delta_0} \quad (13)$$

with $\delta_0 = \delta_0(\sigma^2)$ as defined in Theorem 1. Thus, the optimal average error probability may be alternatively written as

$$\mathbb{P}_e(r|\beta, c) = \inf_{\substack{\{C_n\}_{n=1}^{\infty} \\ \text{supp}(C_n) \subseteq \mathcal{S}^n}} \left\{ \limsup_{n \xrightarrow{(\beta, c)} \infty} P_e^{(n)}(C_n) \mid \liminf_{n \xrightarrow{(\beta, c)} \infty} \sqrt{nK} \left(\frac{1}{nK} \log M_n - \mathbb{E}[C_{N,K}] \right) \geq r \right\} \quad (14)$$

since

$$\liminf_{n \xrightarrow{(\beta, c)} \infty} \sqrt{nK} (\mathbb{E}[C_{N,K}] - C) = 0. \quad (15)$$

In the finite N, K, n -regime, we may therefore see the optimal average error probability as an approximation of the optimal achievable error under the rate constraint

$$\frac{1}{nK} \log M_n \geq \mathbb{E}[C_{N,K}] + \frac{r}{\sqrt{nK}}. \quad (16)$$

Note that the relation (15) is fundamentally dependent on the Gaussianity of H^n . Indeed, (12) is a much stronger result than the well-known convergence of the per-antenna mutual information to its asymptotic limit (see, e.g., [16]) which holds for channels composed of arbitrary i.i.d. entries with finite second-order moment. It was precisely shown in [17, Theorem 4.4] that, whenever the entries of H^n have a non-zero fourth-order cumulant $\kappa = \mathbb{E}[|H_{11}^n|^4] - 2$, a bias term \mathcal{B} proportional to κ arises such that (15) must be modified to $\sqrt{nK}(\mathbb{E}[C_{N,K}] - C) \rightarrow \mathcal{B}$ as $n \xrightarrow{(\beta,c)} \infty$. In this case the equivalence of (14) and (9) does not hold. For Gaussian channels (since $\kappa = 0$ and then $\mathcal{B} = 0$), however, the asymptotic mutual information is reached at a sufficiently fast rate of $\mathcal{O}(n^{-2})$ [15].

Remark 3: We may also consider the *second-order outage probability* $\mathbb{P}_{\text{out}}(r|\beta, c)$ for the rate r , which we define as

$$\mathbb{P}_{\text{out}}(r|\beta, c) \triangleq \inf_{\{\mathcal{C}_n: \text{supp}(\mathcal{C}_n) \subseteq \mathcal{S}^n\}_{n=1}^{\infty}} \left\{ \limsup_{n \xrightarrow{(\beta,c)} \infty} P_e^{(n)}(\mathcal{C}_n) \mid \liminf_{n \xrightarrow{(\beta,c)} \infty} K \left(\frac{1}{nK} \log M_n - C \right) \geq r \right\}.$$

Note that $\mathbb{P}_{\text{out}}(r|\beta, c) = \mathbb{P}_e(r\beta^{\frac{1}{2}}|\beta, c)$. This definition allows us to study the behavior of $\mathbb{P}_{\text{out}}(r|\beta, c)$ for growing β . In the finite dimensional setting, this corresponds to increasing the block length while maintaining N and K (and thus the capacity KC) fixed. This cannot be performed on $\mathbb{P}_e(r|\beta, c)$ since, by growing n , $\sqrt{nK}C$ grows as well. From the above definition and Theorem 1, we have

$$\min \left\{ \Phi \left(\frac{r}{\theta_{\text{out}}^-} \right), \frac{1}{2} \right\} \leq \mathbb{P}_{\text{out}}(r|\beta, c) \leq \Phi \left(\frac{r}{\theta_{\text{out}}^+} \right) \quad (17)$$

where $\theta_{\text{out}}^- \triangleq \beta^{-\frac{1}{2}}\theta_-$ and $\theta_{\text{out}}^+ \triangleq \beta^{-\frac{1}{2}}\theta_+$. Interestingly, for $r \leq 0$, as $\beta \rightarrow \infty$, we recover the limiting outage probability of MIMO Gaussian fading channels, e.g., [17],

$$\lim_{\beta \rightarrow \infty} \mathbb{P}_{\text{out}}(r|\beta, c) = \Phi \left(\frac{r}{\theta_{\text{out}}} \right) \quad (18)$$

with

$$\theta_{\text{out}} \triangleq \left[-\log \left(1 - \frac{1}{c} \frac{\delta_0(\sigma^2)^2}{(1 + \delta_0(\sigma^2))^2} \right) \right]^{\frac{1}{2}}. \quad (19)$$

Although both results coincide, there is a fundamental difference in the way they are obtained. In [17], the block length is assumed to be infinitely large from the start and then the limit is taken in N and K . In contrast, we have obtained (18) by changing the order of both limits.

Figure 2 depicts the bounds on $\mathbb{P}_{\text{out}}(r|\beta, c)$ in (17) as a function of β for different values of c , assuming SNR = 10 dB

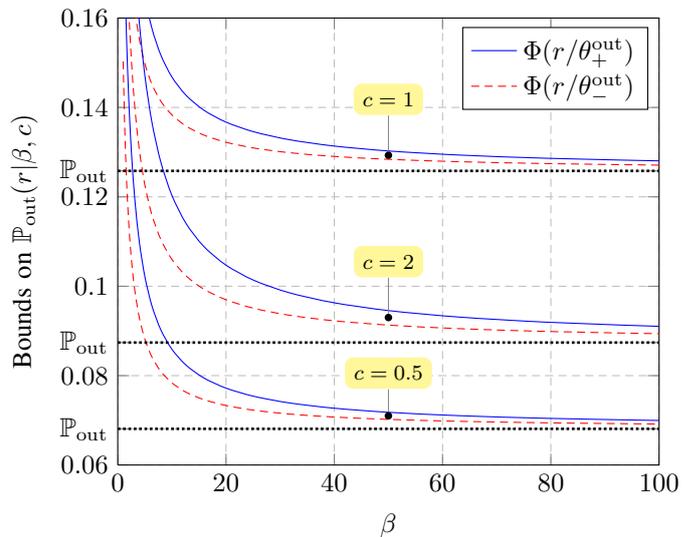


Fig. 2. Bounds on the second-order outage probability as a function of β for different values of c , $r = -1$, and SNR = 10 dB. The limiting outage probability is $\mathbb{P}_{\text{out}} \triangleq \mathbb{P}_{\text{out}}(r|\infty, c)$.

and $r = -1$ fixed. For each value of c we also provide the limiting outage probability as given in (18). The upper and lower bounds are seen to approach the outage probability at a rate $\mathcal{O}(\beta^{-1})$ as β grows, which can be easily proved.

B. Finite dimensional approximation

We now provide an upper bound approximation on the optimal average error probability for arbitrary coding rates R in the finite dimensional regime. We assume transmissions with an *average* energy constraint rather than a *peak* energy constraint and define $(\bar{P}_e^{(n)}, M_n, 1)$ -codes as the equivalent to $(P_e^{(n)}, M_n, 1)$ -codes with input distribution \mathbb{P}_{X^n} satisfying a unit average energy constraint, i.e., $\mathbb{P}_{X^n} \in \bar{\mathcal{S}}^n$, where

$$\bar{\mathcal{S}}^n \triangleq \left\{ \mathbb{P}_{X^n} \mid \mathbb{E} \left[\frac{1}{nK} \text{tr} X^n (X^n)^H \right] \leq 1 \right\}. \quad (20)$$

We then define the optimal average error probability $\bar{\mathbb{P}}_e^{(n)}(R)$ for rate R under unit average energy constraint as

$$\bar{\mathbb{P}}_e^{(n)}(R) \triangleq \inf_{\mathcal{C}^n: \mathbb{P}_{X^n} \in \bar{\mathcal{S}}^n} \left\{ \bar{P}_e^{(n)}(\mathcal{C}^n) \mid \frac{1}{nK} \log M_n \geq R \right\} \quad (21)$$

where $\bar{P}_e^{(n)}(\mathcal{C}^n)$ is the average error probability for a given $(\bar{P}_e^{(n)}, M_n, 1)$ -code.

Before we continue, we need to introduce an auxiliary lemma which is a simple generalization of Feinstein's lemma [2] to arbitrary input distributions:

Lemma 1 (Variation of Feinstein's lemma): Let $n \geq 1$ be an integer and let $\mathbb{P}_{X^n} \in \mathcal{A}^n$ be an arbitrary probability measure where $\mathcal{A}^n \subseteq \mathcal{P}(\mathbb{C}^{K \times n})$. Denote by Y^n the output from the channel $\mathbb{P}_{Y^n|X^n}$ corresponding to the input X^n . Then, there exists a block length n codebook of size M_n that, together with the maximum a posteriori (MAP) decoder, forms

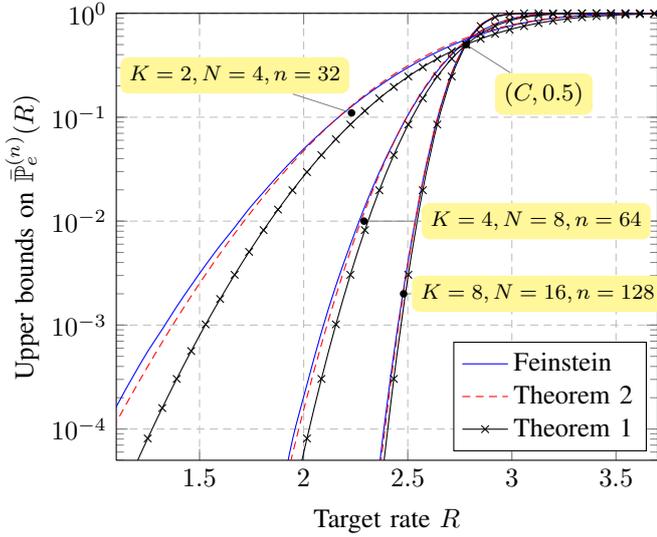


Fig. 3. Upper-bounds on $\bar{\mathbb{P}}_e^{(n)}(R)$ as a function of the coding-rate R for SNR = 10 dB and different values of K , N , and n .

a code \mathcal{C}_n whose average error probability $P_e^{(n)}(\mathcal{C}_n)$ satisfies

$$P_e^{(n)}(\mathcal{C}_n) \leq \inf_{\gamma > 0} \left\{ \Pr \left[\log \frac{\mathbb{P}_{Y^n|X^n}(dY^n|X^n)}{\mathbb{P}_{Y^n}(dY^n)} \leq \log \gamma \right] + \frac{M_n}{\gamma} \right\}. \quad (22)$$

Proof: The full proof is provided in [14]. ■

Since Lemma 1 holds in particular for $\mathcal{A}^n = \bar{\mathcal{S}}^n$ it can be used to prove the following result:

Theorem 2 (Approximation of Feinstein's upperbound):

Let $\{R_n\}_{n=1}^{\infty}$ be a real sequence. Then, there exists a real sequence $\{\ell_n\}_{n=1}^{\infty}$ such that

$$\bar{\mathbb{P}}_e^{(n)}(R_n) \leq \Phi \left(\frac{\sqrt{nK}}{\theta_+} (R_n - C + \delta_n^*) \right) + e^{-nK\delta_n^*} + \ell_n$$

with $\ell_n \downarrow 0$ as $n \xrightarrow{(\beta, c)} \infty$, where

$$\delta_n^* \triangleq (C - R_n + \theta_+^2) \left[1 - \sqrt{1 - \frac{(C - R_n)^2 + (nK)^{-1}\theta_+^2 \log(2\pi nK\theta_+^2)}{(C - R_n + \theta_+^2)^2}} \right]$$

and where C is given by (13) and θ_+ is defined in Theorem 1.

Proof: The full proof is provided in [14]. ■

Note that Theorem 2 fully exploits Lemma 1 in the sense that, for all finite n , the optimal choice for γ (whose role is played by δ_n^* here) in (22) is considered. This quantity is known to be zero in the asymptotic limit, so that it does not appear in Theorem 1. Nonetheless, since we cannot obtain the convergence rate of ℓ_n to zero with respect to that of δ_n^* , the potential gains of Theorem 2 cannot be analytically assessed. This is in contrast to [10] where a Berry–Esseen inequality is used to show that, for $R_n = C + (nK)^{-1/2}r$, $\delta_n^* = \mathcal{O}(n^{-1} \log n)$, while $\ell_n = \mathcal{O}(n^{-1})$.

Figure 3 provides the comparative performance of Theorem 1 and Theorem 2 as an approximation of Feinstein's upper bound (Lemma 1) for $\bar{\mathbb{P}}_e^{(n)}(R)$. Precisely, the curves of Figure 3 are associated to the following approximations of the upper bound on $\bar{\mathbb{P}}_e^{(n)}(R)$:

$$\begin{cases} \inf_{\delta > 0} \left\{ \Pr \left[I_{N,K}^{(n)} \leq R + \delta \right] + e^{-nK\delta} \right\} & \text{(Feinstein)} \\ \Phi \left(\frac{\sqrt{nK}}{\theta_+} (R - C + \delta_n^*) \right) + e^{-nK\delta_n^*} & \text{(Theorem 2)} \\ \Phi \left(\frac{\sqrt{nK}}{\theta_+} (R - C) \right) & \text{(Theorem 1)} \end{cases}$$

where $I_{N,K}^{(n)}$ is the mutual information density for Gaussian inputs X^n defined in (4) with $\Gamma = 1$. We consider three different sets of parameters (K, N, n) . As expected, the larger all these parameters, the smaller the gap between the bounds of Theorem 2 and Theorem 1. For small values of these parameters, the approximation by Theorem 2 provides a much better approximation of Lemma 1 due to a non-negligible value of δ_n^* .

REFERENCES

- [1] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, no. 3, pp. 611–656, 1959.
- [2] A. Feinstein, "A new basic theorem of information theory," *IRE Transactions on Information Theory*, pp. 2–20, 1954.
- [3] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer-Verlag, 2003.
- [4] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. 11, no. 1, pp. 3–18, Jan. 1965.
- [5] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.
- [6] V. Strassen, "Asymptotische Abschätzungen in Shannon's Informationstheorie," in *Proc. 3rd Prague Conf. Inf. Theory*, Czechoslovak Academy of Sciences, Prague, Czech Republic, 1962, pp. 689–723.
- [7] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.
- [8] J. Wolfowitz, "Notes on a general strong converse," *Information and Control*, vol. 12, no. 1, pp. 1–4, 1968.
- [9] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.
- [10] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [11] Y. Polyanskiy and S. Verdú, "Scalar coherent fading channel: Dispersion analysis," in *IEEE Int. Symp. Inf. Theory (ISIT)*, Aug. 2011, pp. 2959–2963.
- [12] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Diversity versus channel knowledge at finite block-length." [Online]. Available: <http://arxiv.org/abs/1204.2927>
- [13] J. Hoydis, R. Couillet, P. Piantanida, and M. Debbah, "A random matrix approach to the finite blocklength regime of MIMO fading channels," in *IEEE Int. Symp. Inf. Theory (ISIT)*, Cambridge, MA, US, Jul. 2012.
- [14] J. Hoydis, R. Couillet, and P. Piantanida, "The second-order coding rate of the MIMO Rayleigh block-fading channel," to be submitted. Available upon request.
- [15] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, "A new approach for mutual information analysis of large dimensional multi-antenna channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 3987–4004, 2008.
- [16] S. Verdú and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 622–640, Mar. 1999.
- [17] W. Hachem, P. Loubaton, and J. Najim, "A CLT for information theoretic statistics of Gram random matrices with a given variance profile," *The Annals of Probability*, vol. 18, no. 6, pp. 2071–2130, Dec. 2008.