

Fixed-Energy Random Coding with Rescaled Codewords at the Transmitter

Daniel Fehr*, Jonathan Scarlett† and Alfonso Martinez*

* Universitat Pompeu Fabra, Barcelona

† École Polytechnique Fédérale de Lausanne

e-mail: daniel.co@bluewin.ch, jmscarlett@gmail.com, alfonso.martinez@ieee.org

Abstract—This paper proposes a new method to reduce the error rate of channel codes over an AWGN channel by renormalizing the codewords to a constant energy before transmission and decoding with the original codebook. Evaluation of the random-coding error exponent reveals that this normalization technique approaches the constant-composition error exponent for certain pairs of rate and signal-to-noise ratio.

I. INTRODUCTION

Given a general coded modulation scheme over an AWGN channel, we investigate the effect on the error rate of rescaling the transmitted codewords so that their energy is constant. We compare the proposed technique with the standard coded modulation scheme and with constant-composition codes [1], [2] by means of the respective random-coding error exponents.

The comparison reveals that, for low to moderate SNRs, codeword rescaling significantly improves the error exponent of random codes and nearly matches the performance of constant-composition codes. While constant-composition codes, which inherently have fixed codeword energy and are known to achieve the optimum exponent, are difficult to design in practice, codeword rescaling can be applied to existing practical codes without increasing their complexity.

In Sect. II we present the general channel coding model. In Sect. III we add codeword rescaling to the model and derive an achievable random-coding exponent. In Sect. IV we introduce two other well known random-coding exponents and compare them with the new one by numerical evaluation.

II. MODEL

The channel input sequence $\mathbf{x} = (x_1, \dots, x_n)$ consists of n symbols $x \in \mathcal{X}$, where \mathcal{X} is the symbol constellation. We denote the channel output sequence by \mathbf{y} and the channel law, that is, the conditional probability density of receiving sequence \mathbf{y} when the sequence \mathbf{x} has been sent, by $W^n(\mathbf{y} | \mathbf{x})$. We represent random variables by capital letters and their realizations by lowercase letters, e.g. $\mathbf{X} = (X_1, \dots, X_n)$ and $\mathbf{Y} = (Y_1, \dots, Y_n)$ denote random input and output vectors. The channel is memoryless and $W^n(\mathbf{y} | \mathbf{x}) = \prod_{i=1}^n W(y_i | x_i)$, where $W(y | x)$ is the single-letter channel law.

This work has been funded in part by the European Research Council under ERC grant agreement 259663, by the European Union's 7th Framework Programme (PEOPLE-2011-CIG) under grant agreement 303633 and by the Spanish Ministry of Economy and Competitiveness under grants RYC-2011-08150 and TEC2012-38800-C03-03.

We focus on a complex-valued additive white Gaussian noise (AWGN) channel. The input alphabet \mathcal{X} is a finite subset of the complex numbers and the channel output is given by

$$y_i = \sqrt{\text{SNR}}x_i + z_i, \quad i = 1, \dots, n \quad (1)$$

where x_i are the symbols, y_i the channel output values and SNR is the signal-to-noise ratio. The noise values z_i are drawn from a circularly-symmetric complex-valued Gaussian random variable with zero mean and unit variance. Therefore, the symbol channel transition probability is given by

$$W(y | x) = \frac{1}{\pi} e^{-|y - \sqrt{\text{SNR}}x|^2}. \quad (2)$$

The empirical average symbol energy of channel input sequence \mathbf{x} is $\mathcal{E}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n |x_i|^2$. Let Q be a distribution on the symbols in \mathcal{X} . We require that the constellation \mathcal{X} is chosen such that the identities $\mathbf{E}[X] = 0$ and $\mathbf{E}[|X|^2] = 1$ hold, where the expectations are with respect to Q .

We use $\mathbf{P}[\mathcal{A}]$ to denote the probability of an event \mathcal{A} , and $\mathbf{E}[\cdot]$ is the expectation operator. We denote the Kullback-Leibler Divergence as $D(P||Q)$, the set of all compositions on length- n sequences drawn from \mathcal{X} as \mathcal{P}_n , the set of all distributions on \mathcal{X} as \mathcal{P} and the image of set \mathcal{A} under the function f as $f[\mathcal{A}]$.

We denote the channel code by \mathcal{C}_n . The code consists of M codewords \mathbf{x} , i.e. $\mathcal{C}_n = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(M)}\}$, and the corresponding code is said to be an (n, M) block code of block length n . The encoder assigns to each message $m \in \{1, \dots, M\}$ a codeword $\mathbf{x}^{(m)}$ from the codebook \mathcal{C}_n . We assume that the message m is drawn according to a uniform distribution. The rate of a code is defined as $R \triangleq \frac{\log_2 M}{n}$.

The decoder outputs an estimated message \hat{m} according to a maximum-metric rule

$$\hat{m} = \arg \max_{i \in \mathcal{M}} q^n(\mathbf{x}^{(i)}, \mathbf{y}), \quad (3)$$

where q^n denotes the metric that the decoder uses to estimate which message m has been sent. We focus on metrics q^n that can be expressed in terms of the letter metric q as in $q^n(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^n q(x_i, y_i)$. Further, we require q to be positive.

An error occurs if the decoder's estimate differs from the sent message, i.e. $\hat{m} \neq m$. The error probability of a code \mathcal{C}_n is $p_e(\mathcal{C}_n) = \frac{1}{M} \sum_{m \in \mathcal{M}} \mathbf{P}[\hat{m} \neq m | m]$, and we equivalently write $p_e(n, M) = p_e(\mathcal{C}_n)$. Finally, an error exponent $E(Q, R)$ is

said to be achievable if there exists a sequence of (n, M) -codes \mathcal{C}_n such that

$$\liminf_{n \rightarrow \infty} -\frac{1}{n} \log p_e(\mathcal{C}_n) \geq E(Q, R). \quad (4)$$

III. CODEWORD RESCALING

A. Rescaling setup

Let a rescaler η be a block that performs the operation

$$\eta(\mathbf{x}) = \mathcal{E}(\mathbf{x})^{-\frac{1}{2}} \mathbf{x} \quad (5)$$

on a codeword \mathbf{x} ; that is, it renormalizes the codeword \mathbf{x} such that the empirical codeword energy is $\mathcal{E}(\eta(\mathbf{x})) = 1$.

We consider a coded modulation scheme that consists of an encoder and a rescaler η . The encoder maps message m into codeword $\mathbf{x}^{(m)} = \phi(m)$ using codebook \mathcal{C} , the rescaler outputs an energy-normalized version $\tilde{\mathbf{x}}^{(m)}$ of the codeword $\mathbf{x}^{(m)}$. The rescaling operation can be thought as part of the encoder's codebook, so that we use a code $\tilde{\mathcal{C}}$ with rescaled codewords $\tilde{\mathbf{x}}$ that consist of symbols \tilde{x} from an expanded constellation $\tilde{\mathcal{X}}$.

The decoder outputs the estimate \hat{m} under the original codebook \mathcal{C} by maximizing the metric q^n . With the choice of $q^n(\mathbf{x}, \mathbf{y}) = W^n(\mathbf{y} | \mathbf{x})$, we have an instance of mismatched decoding, since the decoder does not account for the rescaling operation neither in the codebook \mathcal{C} nor in the decoding metric. We consider a slightly more general choice given by $q^n(\mathbf{x}, \mathbf{y}) \triangleq W^n(\mathbf{y} | \beta \mathbf{x})$, where β may be optimized to minimize the error probability; however, it cannot depend on the codeword, and hence it cannot be used to undo the rescaling. Note that for a practical code, β is fixed before deployment and such a decoding metric can be implemented without additional computational complexity.

We can build an equivalent model for the rescaling setup by removing the rescaling block from the transmitter and reinterpreting it as a channel property. With this model, the scaling function leads to a new channel law

$$\tilde{W}^n(\mathbf{y} | \mathbf{x}) = W^n(\mathbf{y} | \eta(\mathbf{x})). \quad (6)$$

Note that \tilde{W}^n does not represent a memoryless channel.

B. Scaling exponent

We study the i.i.d. random-coding error probability. We consider an ensemble of codebooks with block length n and $M = 2^{nR}$ codewords. The ensemble consists of codebooks whose codewords $\mathbf{x}^{(i)}$, $i = 1, \dots, M$ are randomly generated. A codeword $\mathbf{x}^{(i)} = (x_1, \dots, x_n)$ at entry i in the random codebook is generated by drawing its n symbols according to the distribution $Q(x)$. We are interested in the achievable random-coding exponent $E_r^{\text{scl}}(Q, R)$ of the ensemble average of the error probability $\bar{p}_e(n, M) = \sum_{\mathcal{C}_n} \mathbf{P}[\mathcal{C}_n] p_e(\mathcal{C}_n)$.

Theorem 1 (Scaling random-coding exponent): The random-coding error probability in a rescaling setup satisfies

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log_2 \bar{p}_e(n, 2^{nR}) \geq E_r^{\text{scl}, \beta}(Q, R), \quad (7)$$

where the scaling exponent is defined as

$$E_r^{\text{scl}, \beta}(Q, R) \triangleq \sup_{\beta \geq 0} \min_{P \in \mathcal{P}} \sup_{\substack{\rho \in [0, 1] \\ s \geq 0}} \{E_0^{\text{scl}, \beta}(Q, \rho, s, P) - \rho R\}, \quad (8)$$

and the corresponding $E_0^{\text{scl}, \beta}$ is defined as

$$E_0^{\text{scl}, \beta}(Q, \rho, s, P) \triangleq D(P \| Q) - \mathbf{E} \left[\log_2 \mathbf{E} \left[\frac{\mathbf{E}[q(\bar{X}, Y)^s | Y]^\rho}{q(X, Y)^{\rho s}} \middle| X \right] \right] \quad (9)$$

and the expectations are with respect to

$$(X, Y, \bar{X}) \sim P(X)W(y | \mathcal{E}(P)^{-\frac{1}{2}} X)Q(\bar{x}). \quad (10)$$

Proof: For the ensemble of codebooks with M codewords of length n , chosen according to random-coding distribution Q^n , transmitted over a channel described by the arbitrary channel law W^n and decoded according to the metric q^n , the average ensemble error probability is bounded by the random-coding union (RCU) bound [4], [5]

$$\begin{aligned} \text{rcu}(n, M) &= \mathbf{E} \left[\min \left\{ 1, (M-1) \mathbf{P} \left[\frac{q^n(\bar{X}, Y)}{q^n(X, Y)} \geq 1 \middle| X, Y \right] \right\} \right], \end{aligned} \quad (11)$$

where $(X, Y, \bar{X}) \sim Q^n(x)W^n(y | x)Q^n(\bar{x})$.

Weakening (11) by replacing the function $z \mapsto \min\{1, z\}$ with the function $z \mapsto z^{\rho(\mathbf{x})}$ and using Markov's inequality with parameter $s(\mathbf{x})$, leads to the definition of a parametrized upper bound on the the RCU bound

$$\text{rcu}_{\rho, s}(n, M) \triangleq \mathbf{E} \left[\left(\frac{\mathbf{E}[q^n(\bar{X}, Y)^{s(\mathbf{X})} | X, Y]}{q^n(X, Y)^{s(\mathbf{X})} (M-1)^{-1}} \right)^{\rho(\mathbf{X})} \right] \quad (12)$$

that holds for all pairs of functions $(\rho(\mathbf{x}), s(\mathbf{x}))$ such that $\rho[\mathcal{X}^n] \subseteq [0, 1]$ and $s[\mathcal{X}^n] \subseteq [0, \infty)$.

We further weaken (12) by applying $M-1 \leq 2^{nR}$, use the random-coding distribution $Q_{\text{id}}^n = \prod_{i=1}^n Q(x_i)$ and the equivalent scaling channel model (6), to obtain

$$\bar{p}_e(n, M) \leq \sum_{\mathbf{x} \in \mathcal{X}^n} 2^{n\rho(\mathbf{x})R} Q_{\text{id}}^n(\mathbf{x}) f^n(\rho(\mathbf{x}), s(\mathbf{x}), \mathbf{x}), \quad (13)$$

where

$$\begin{aligned} f^n(\rho, s, \mathbf{x}) &\triangleq \int_{\mathbf{y}} W(\mathbf{y} | \mathcal{E}(\mathbf{x})^{-\frac{1}{2}} \mathbf{x}) \left(\sum_{\bar{\mathbf{x}} \in \mathcal{X}^n} Q_{\text{id}}^n(\bar{\mathbf{x}}) \frac{q^n(\bar{\mathbf{x}}, \mathbf{y})^s}{q^n(\mathbf{x}, \mathbf{y})^s} \right)^\rho d\mathbf{y}. \end{aligned} \quad (14)$$

We split the outer summation over the channel-input sequences in (13) into summations over sequences \mathbf{x} of composition $P = \hat{P}_{\mathbf{x}}$ and obtain

$$\bar{p}_e(n, M) \leq \sum_{P \in \mathcal{P}_n} 2^{n\rho(P)R} \sum_{\mathbf{x} \in T(P)} Q_{\text{id}}^n(\mathbf{x}) f^n(\rho(P), s(P), \mathbf{x}). \quad (15)$$

We also reduced the degrees of freedom for the parameters $\rho(\mathbf{x})$ and $s(\mathbf{x})$ such that they only depend on the composition $P = \hat{P}_{\mathbf{x}}$. This simplifies the analysis.

The codeword average symbol energy only depends on the codewords composition. Hence we write $\mathcal{E}(\mathbf{x}) = \mathcal{E}(\hat{P}_{\mathbf{x}})$ and use it in (14) to obtain $f^n(\rho, s, P, \mathbf{x})$. This, together with the product nature of decoding metric, channel law and Q_{iid}^n , allows us to factor f^n as $f^n(\rho, s, P, \mathbf{x}) = \prod_{i=1}^n f(\rho, s, P, x_i)$, where

$$f(\rho, s, P, x) \triangleq \int_y W(y | \mathcal{E}(P)^{-\frac{1}{2}}x) \left(\sum_{\bar{x} \in \mathcal{X}} Q(\bar{x}) \frac{q(\bar{x}, y)^s}{q(x, y)^s} \right)^\rho dy, \quad (16)$$

by invoking the distributive law. For the product in f^n , only the composition of \mathbf{x} is relevant, that is, it can be expressed independent of \mathbf{x} as $f^n(\rho, s, P) = \prod_{x \in \mathcal{X}} f(\rho, s, P, x)^{nP(x)}$. We use this form in (15) to obtain

$$\begin{aligned} \bar{p}_e(n, M) &\leq \sum_{P \in \mathcal{P}_n} \sum_{\mathbf{x} \in T(P)} \frac{2^{n\rho(P)R} \prod_{x \in \mathcal{X}} f(\rho(P), s(P), P, x)^{nP(x)}}{2^{n(D(P||Q)+H(P))}}, \end{aligned} \quad (17)$$

where we expressed the codeword probability in terms of its composition as in $Q_{\text{iid}}^n(\mathbf{x}) = 2^{-n(D(\hat{P}_{\mathbf{x}}||Q)+H(\hat{P}_{\mathbf{x}}))}$ [6].

Since the summand is independent of the codeword \mathbf{x} in (17), we can upper-bound the summation over the codewords by using the bound on the number of sequences in a composition class $|T(P)| \leq 2^{nH(P)}$ [6]. Doing some rearrangements and bringing all terms on a common exponent base, we obtain

$$\bar{p}_e(n, M) \leq \sum_{P \in \mathcal{P}_n} 2^{-n\xi(\rho(P), s(P), P, R)}, \quad (18)$$

where

$$\xi(\rho, s, P, R) \triangleq D(P||Q) - \sum_{x \in \mathcal{X}} P(x) \log_2 f(\rho, s, P, x) - \rho R. \quad (19)$$

A simple upper bound on a sum is obtained by fixing its summands to the largest one, that is $\sum_{a \in \mathcal{A}} a \leq |\mathcal{A}|(\max_{a \in \mathcal{A}} a)$. We weaken (18) with this bound and the fact that the number of compositions is bounded by $|\mathcal{P}_n| \leq (n+1)^{|\mathcal{X}|}$ [6] and get

$$\bar{p}_e(n, M) \leq (n+1)^{|\mathcal{X}|} 2^{-n[\min_{P \in \mathcal{P}_n} \xi(\rho(P), s(P), P, R)]}. \quad (20)$$

The bound (20) is achievable for any pair of parameters $(\rho(P), s(P))$, which we exploit by choosing them such that the bound gets as tight as possible. By definition of the min and sup operators, this is the case when we replace ξ with

$$\xi^*(P, R) = \sup_{0 \leq \rho \leq 1, s \geq 0} \xi(\rho, s, P, R). \quad (21)$$

That is, place the supremum inside the minimum operator.

Finally, we observe the sub-exponential factor in (20) which suggests to transform the inequality as

$$-\frac{1}{n} \log_2 \bar{p}_e(n, 2^{nR}) \geq \min_{P \in \mathcal{P}_n} \xi^*(P, R) - |\mathcal{X}| \frac{\log_2(n+1)}{n} \quad (22)$$

and take the limit

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log_2 \bar{p}_e(n, 2^{nR}) \geq \min_{P \in \mathcal{P}} \xi^*(P, R). \quad (23)$$

C. Swapped exponent

An inspection by example of the optimization parameters in (8) shows that $E_0^{\text{scl}, \beta}$ is not convex in P for fixed values of ρ, s and R , which complicates its numerical computation. Further, note that it is crucial to find the true P^* to guarantee the achievability of $E_r^{\text{scl}, \beta}$, since P^* is the result of a minimization over a variable that is not an arbitrary parameter. For these reasons, we introduce a lower bound on (8) that is computationally tractable.

Theorem 2 (Swapped scaling random-coding exponent): The swapped random-coding error exponent, defined as

$$E_r^{\text{swp}, \beta}(Q, R) \triangleq \sup_{\beta \geq 0} \min_{\epsilon \in \mathcal{S}} \sup_{\substack{\rho \in [0, 1] \\ s \geq 0}} \{E_0^{\text{swp}, \beta}(Q, \rho, s, \epsilon) - \rho R\} \quad (24)$$

with

$$\begin{aligned} E_0^{\text{swp}, \beta}(Q, \rho, s, \epsilon) &\triangleq \min_{P \in \mathcal{P}_\epsilon} \left\{ D(P||Q) - \mathbf{E} \left[\log_2 \mathbf{E} \left[\frac{\mathbf{E}[q(\bar{X}, Y)^s | Y]^\rho}{q(X, Y)^{\rho s}} \middle| X \right] \right] \right\}, \end{aligned} \quad (25)$$

is a lower bound on $E_r^{\text{scl}, \beta}$, where ϵ is optimized over the interval $\mathcal{S} = [\min_{x \in \mathcal{X}} |x|^2, \max_{x \in \mathcal{X}} |x|^2]$, the set \mathcal{P}_ϵ is $\{P \in \mathcal{P} \mid \mathcal{E}(P) = \epsilon\}$ and the expectations are according to $(X, Y, \bar{X}) \sim P(x)W(y | \epsilon^{-\frac{1}{2}}x)Q(\bar{x})$.

Proof: First, let the energy level ϵ denote the expected symbol energy with respect to P or equivalently $\epsilon = \mathcal{E}(P)$. We observe that the probability simplex \mathcal{P} can be partitioned into disjoint subsets \mathcal{P}_ϵ , where a subset consists of all distributions P that obtain the energy level ϵ , that is, we have $\mathcal{P} = \bigcup_{\epsilon \in \mathcal{S}} \mathcal{P}_\epsilon$ where $\mathcal{P}_\epsilon = \{P \in \mathcal{P} \mid \mathcal{E}(P) = \epsilon\}$ and $\mathcal{S} = [\min_{x \in \mathcal{X}} |x|^2, \max_{x \in \mathcal{X}} |x|^2]$. The subsets \mathcal{P}_ϵ are compact and convex since they are intersections of two hyperplanes and the closed positive orthant [7]. In correspondence with these observations, we rephrase the scaling exponent (8) in terms of energy levels and subsets as

$$E_r^{\text{scl}, \beta}(Q, R) = \min_{\epsilon \in \mathcal{S}} \min_{P \in \mathcal{P}_\epsilon} \sup_{\substack{\rho \in [0, 1] \\ s \geq 0}} \{E_0^{\text{scl}, \beta}(Q, \rho, s, P) - \rho R\}. \quad (26)$$

We swap the order of the inner two optimization operators and define $E_r^{\text{swp}, \beta}$ and $E_0^{\text{swp}, \beta}$ as in (24) and (25) respectively. Observing that the minimax inequality

$$\sup_x \min_y f(x, y) \leq \min_y \sup_x f(x, y) \quad (27)$$

holds we conclude that swapping the order of optimization results in a lower bound on (8). ■

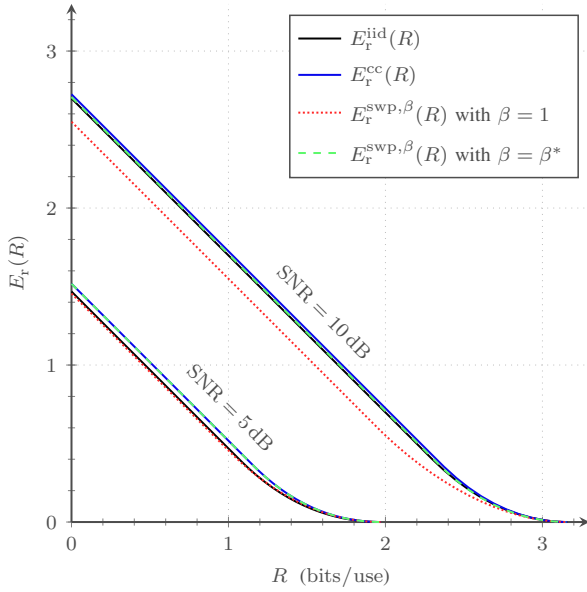


Fig. 1. Random-coding error exponents for 16QAM and fixed SNRs of 5 dB and 10 dB, respectively.

IV. NUMERICAL RESULTS

We are interested in how the error exponent of channel coding with rescaling performs with respect to setups with the iid and constant-composition exponents [1], [2].

For a discrete-input continuous-output memoryless channel with ML decoder and input distribution Q , the iid exponent is given by [1]

$$E_r^{\text{iid}}(Q, R) = \max_{\rho \in [0,1]} E_0^{\text{iid}}(\rho) - \rho R, \quad (28)$$

where

$$E_0^{\text{iid}}(Q, \rho) \triangleq -\log_2 \int_y \mathbf{E} \left[W(y|X)^{\frac{1}{1+\rho}} \right]^{1+\rho} dy \quad (29)$$

is the Gallager function [3]. The expectation is taken with respect to Q .

For a discrete-input continuous-output memoryless channel with ML decoder and input distribution Q , the constant-composition exponent is given by [4]

$$E_r^{\text{cc}}(Q, R) = \max_{\rho \in [0,1]} E_0^{\text{cc}}(\rho) - \rho R, \quad (30)$$

where

$$E_0^{\text{cc}}(Q, \rho) \triangleq \sup_{a(\cdot)} -\log_2 \int_y \mathbf{E}_Q \left[W(y|X)^{\frac{1}{1+\rho}} e^{a(X) - \phi_a} \right]^{1+\rho} dy \quad (31)$$

and $\phi_a = \mathbf{E}_Q[a(X)]$, and the optimization in (31) is over all real-valued functions a . The expectations are taken with respect to $(X, Y, \bar{X}) \sim Q(x)W(y|x)Q(\bar{x})$.

Fig. 1 compares the error exponents (28), (30) and (24) at SNRs 5 dB and 10 dB. It suggests that $E_r^{\text{swp}, \beta}(R)$ approaches the constant-composition exponent.

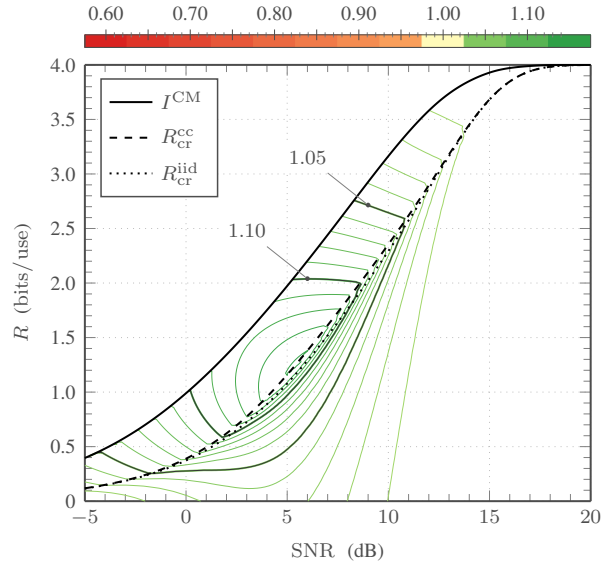


Fig. 2. Relative constant-composition exponent E_r^{cc} for 16QAM.

For a more detailed comparison, we introduce error exponent ratios with the iid exponent as the baseline. We call these the relative constant-composition exponent and the relative swapped scaling exponent, respectively given by

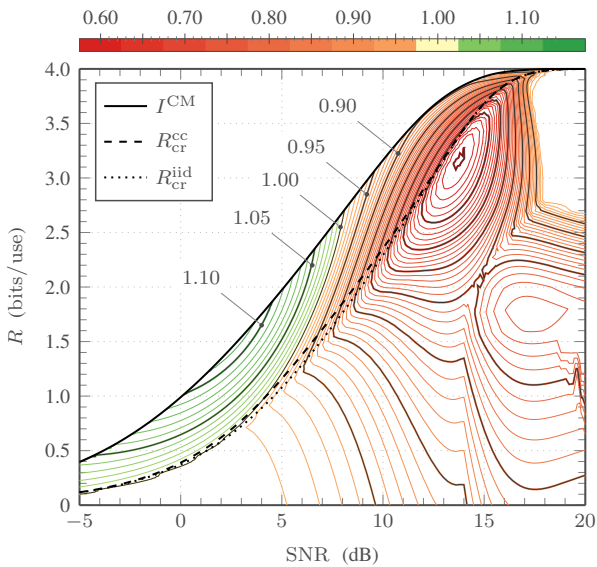
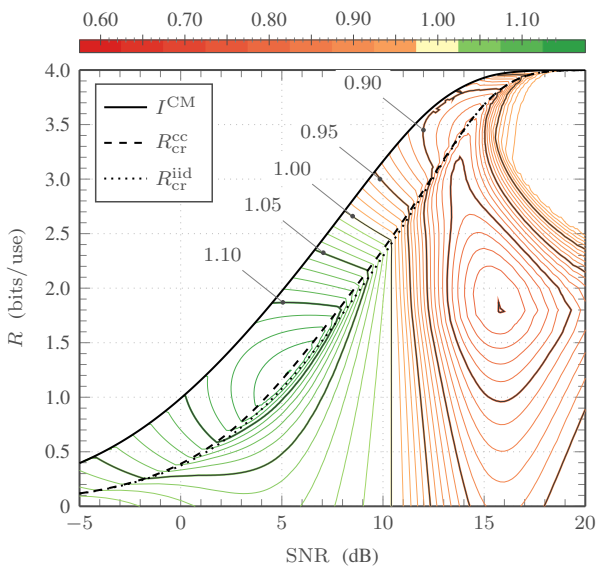
$$E_r^{\text{cc}}(R) \triangleq \frac{E_r^{\text{cc}}(R)}{E_r^{\text{iid}}(R)} \quad \text{and} \quad E_r^{\text{swp}, \beta}(R) \triangleq \frac{E_r^{\text{swp}, \beta}(R)}{E_r^{\text{iid}}(R)}. \quad (32)$$

Our figures show the relative random-coding exponents as a function of both SNR and the rate R in a single contour plot for coded modulation with a 16QAM constellation and a uniform distribution Q . The figures also depict the mutual information (MI) as a solid line and the critical rate¹ $R_{\text{cr}}^{\text{iid}}$ of the iid exponent (resp. $R_{\text{cr}}^{\text{cc}}$ of the constant-composition exponent) as a dashed (resp. dotted) line.

1) *Constant-composition exponent:* The constant-composition exponent's gain with respect to the iid exponent provides a benchmark to assess the performance of the codeword rescaling. The relative exponent is shown in Fig. 2. The figure reveals that the constant-composition exponent exhibits the largest gains for low to moderate SNRs. At high SNRs above 12 dB, it is roughly equal to the iid exponent. A similar tendency is seen at very low SNRs. For our comparison, the most relevant part of the contour plot is located between the critical rates and the MI, since below the critical rate expurgated versions of the exponents lead to better bounds. The maximal gain with respect to the iid exponent is roughly 13% and occurs around SNR = 5.5 dB and $R = 1.25$ bits per channel use.

2) *Swapped exponent with $\beta = 1$:* We discuss first the relative swapped exponent with $\beta = 1$, which corresponds to

¹The critical rate is defined as the largest rate R at which the optimal ρ in the exponent optimization is one.


 Fig. 3. Relative swapped exponent $E_r^{\text{swp},\beta}$ with $\beta = 1$ for 16QAM.

 Fig. 4. Relative swapped exponent $E_r^{\text{swp},\beta}$ for 16QAM.

the standard decoding rule of maximizing W^n . Fig. 3 shows this relative exponent. We observe two regimes in the relevant region between the critical rate and the capacity. Below about 7 dB, there is a gain with respect to the iid exponent, whereas above 7 dB the swapped exponent is worse.

In the low-SNR region and close to the capacity, the swapped exponent achieves similar gains as the constant-composition exponent, namely, about 10% gain at 3 dB and 1.5 bits per channel use. However, it falls short as the rate approaches the critical rate. For example, there is no gain at 5.5 dB and 1.25 bits per channel use, the point where the constant-composition exponent achieves the highest gain. At high SNRs, especially around 14 dB, Fig. 3 unveils a large loss compared to the iid exponent. In this region, we observe not even at capacity a gain and the loss is up to 35% close to the critical rate.

3) *Swapped exponent with optimal β* : Optimizing over β to adapt the mismatched metric leads to respectable gains in some regions of the SNR-rate plane, as we can see in Fig. 4. At high SNRs, the swapped exponent with optimized β exhibits a considerable improvement compared to the swapped exponent with $\beta = 1$, even though it is still below the iid exponent.

Most notably, for low SNRs, the swapped exponent with optimized β achieves more than 90% of the gain that the constant-composition exponent achieves with respect to the iid exponent. Especially in regions farther away from capacity, the fully optimized exponent shows a large improvement with respect to the one with $\beta = 1$.

REFERENCES

- [1] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz), "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.
- [2] A. G. D'yachkov, "Bounds on the average error probability for a code ensemble with fixed composition," *Prob. Inf. Transm.*, vol. 16, no. 4, pp. 3–8, 1980.
- [3] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.
- [4] J. Scarlett, A. Martinez and A. Guillén i Fàbregas, "Mismatched Decoding: Error Exponents, Second-Order Rates and Saddlepoint Approximations," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2647–2666, May 2014.
- [5] Y. Polyanskiy, 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.
- [6] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.
- [7] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.