Duality Bounds on the Cutoff Rate With Applications to Ricean Fading

Amos Lapidoth, *Fellow, IEEE*, and Natalia Miliou, *Student Member, IEEE*

***Abstract*—We propose a technique to derive upper bounds on Gallager's cost-constrained random coding exponent function. Applying this technique to the noncoherent peak-power or average-power limited discrete time memoryless Ricean fading channel, we obtain the high signal-to-noise ratio (SNR) expansion of this channel's cutoff rate. At high SNR, the gap between channel capacity and the cutoff rate approaches a finite limit. This limit is approximately 0.26 nats per channel-use for zero specular component (Rayleigh) fading and approaches 0.39 nats per channel-use for very large values of the specular component.**

We also compute the asymptotic cutoff rate of a Rayleigh-fading channel when the receiver has access to some partial side information concerning the fading. It is demonstrated that the cutoff rate does not utilize the side information as efficiently as capacity, and that the high SNR gap between the two increases to infinity as the imperfect side information becomes more and more precise.

***Index Terms*—Asymptotic, channel capacity, cutoff rate, duality, fading, high signal-to-noise ratio (SNR), Lagrange, Ricean fading, Rician fading.**

I. Introduction

THIS paper addresses the computation of a function that is key to the evaluation of both the random coding and the sphere packing error exponents. This function, often denoted $E_0(\varrho)$, is usually expressed as a maximization problem over input distributions. Consequently, it is conceptually easily bounded from below: any feasible input distribution gives rise to such a bound. In this paper, we propose to use a dual expression for $E_0(\varrho)$—an expression that involves a minimization over output distributions—in order to derive *upper* bounds on $E_0(\varrho)$. We shall demonstrate this approach by studying the cutoff rate of noncoherent Ricean (or "Rician") fading channels. To that end, we shall have to study the appropriate modifications to the function $E_0(\varrho)$ that are needed to account for input constraints and for infinite input and output alphabets.

It should be noted that the dual expression we propose to use is not new [1], [2, Ch. 2.5, Problem 23]. We merely extend it here to input constrained channels over infinite alphabets and demonstrate how it can be used to derive *analytic* upper bounds on the random coding and sphere packing error exponents. For *numerical* procedures (for unconstrained finite alphabet channels) see [3] and the Geometric Programming approach surveyed in [4] and references therein.

The rest of this section is dedicated to the introduction of the function $E_0(\varrho)$ for discrete memoryless channels. We first treat unconstrained channels and demonstrate the Lagrange duality between Gallager's and Csiszár and Körner's expressions for the random coding error exponent. We then introduce the modifications that are needed to account for input constraints. We describe both the "method of types" approach and Gallager's approach. We pay special attention to the modification that Gallager introduced to account for cost constraints. This introduction is somewhat lengthy because, while the results are not new, we had difficulty pointing to a publication that introduces the two approaches side by side and that compares the two in the presence of cost constraints.

In Section II, we extend the discussion to infinite alphabets and prove the basic inequality on which our technique for upper bounding $E_0(\varrho)$ is based: Proposition 2. In Section III, we introduce the discrete-time memoryless Ricean fading channel with and without full or partial side information at the receiver, and we describe our asymptotic results on this channel's cutoff rate. These asymptotic results are derived using duality in Section IV, which concludes the paper.

A. Unconstrained Inputs

To motivate the interest in the function $E_0(\varrho)$ we shall begin by addressing the case where there are no input constraints. The reliability function $E(R)$ corresponding to rate-R unconstrained communication over a discrete memoryless channel (DMC) of capacity $C \geq R$ is the best exponential decay in the blocklength n of the average probability of error that one can achieve using rate-R block-length-n codebooks. That is,

$$E(R) \triangleq \overline{\lim}_{n\to\infty} -\frac{1}{n}\log \mathrm{P_e}(n,R) \tag{1}$$

where $\mathrm{P_e}(n,R)$ denotes the average probability of error of the best rate-R block-length-n codebook for the given channel.

The problem of computing the reliability function of a general DMC of law $\mathsf{W}(y|x)$ over the finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ is still open. Various upper and lower bounds are, however, known. To derive lower bounds on the reliability function one must derive upper bounds on the probability of error of the best rate-R block-length-n code. This is typically done by demonstrating the existence of good codes for which the average probability of error is small. One such lower bound on $E(R)$ is the random coding lower bound [5]. By considering an

Manuscript received March 14, 2005; revised December 12, 2005. The material in this paper was presented in part at the 2004 IEEE International Symposium on Information Theory, Chicago, IL, June/July 2004.

The authors are with the Department of Information Technology and Electrical Engineering, the Swiss Federal Institute of Technology (ETH), Zurich, CH-8092 Switzerland.

Communicated by K. Kobayashi, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2006.876349

ensemble of codebooks whose codewords are chosen independently, each according to a product distribution of marginal law Q, Gallager derived the lower bound

$$E(R) \geq E_{\mathrm{G}}(R, \mathsf{Q}) \tag{2}$$

where

$$E_{\mathrm{G}}(R, \mathsf{Q}) \triangleq \max_{0 \leq \varrho \leq 1} \{E_{\mathrm{G},0}(\varrho, \mathsf{Q}) - \varrho R\} \tag{3}$$

and

$$E_{\mathrm{G},0}(\varrho, \mathsf{Q}) \triangleq -\log \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} \mathsf{Q}(x) \mathsf{W}(y|x)^{\frac{1}{1+\varrho}} \right)^{1+\varrho}. \tag{4}$$

Since the law Q from which the ensemble of codebooks is constructed is arbitrary, Gallager obtained the bound

$$E(R) \geq E_{\mathrm{G,r}}(R) \tag{5}$$

where $E_{\mathrm{G,r}}(R)$, Gallager's random coding error exponent, is given by

$$\begin{aligned} E_{\mathrm{G,r}}(R) &\triangleq \max_{\mathsf{Q}} E_{\mathrm{G}}(R, \mathsf{Q}) \\ &= \max_{\mathsf{Q}} \max_{0 \leq \varrho \leq 1} \{E_{\mathrm{G},0}(\varrho, \mathsf{Q}) - \varrho R\}. \end{aligned}$$

A different random coding lower bound on the reliability function can be derived using the ensemble of codebooks where the codewords are still chosen independently, but rather than according to a product distribution, each is now chosen uniformly over a type class [2, Ch. 2.5], [1], [6]. With this approach one obtains [2, Ch. 2.5], [1] the lower bound

$$E(R) \geq E_{\mathrm{CK}}(R, \mathsf{Q}) \tag{6}$$

where

$$E_{\mathrm{CK}}(R, \mathsf{Q}) \triangleq \min_{\mathsf{V}(\cdot|\cdot)} \left\{ D(\mathsf{V} \| \mathsf{W} | \mathsf{Q}) + |I(\mathsf{Q}, \mathsf{V}) - R|^{+} \right\}. \tag{7}$$

Here the minimization is over all conditional laws

$$\begin{aligned} \mathsf{V}(y|x) \geq 0, \quad & \sum_{y \in \mathcal{Y}} \mathsf{V}(y|x) = 1, \quad \forall\, x \in \mathcal{X} \\ D(\mathsf{V} \| \mathsf{W} | \mathsf{Q}) &= \sum_{x \in \mathcal{X}} \mathsf{Q}(x) D\left(\mathsf{V}(\cdot|x) \| \mathsf{W}(\cdot|x)\right) \\ &= \sum_{x \in \mathcal{X}} \mathsf{Q}(x) \sum_{y \in \mathcal{Y}} \mathsf{V}(y|x) \log \frac{\mathsf{V}(y|x)}{\mathsf{W}(y|x)} \end{aligned}$$

the term $I(\mathsf{Q}, \mathsf{V})$ denotes the mutual information corresponding to the channel V and the input distribution Q; and $|\xi|^{+}$ stands for $\max\{\xi, 0\}$. Again, since the type Q according to which the ensemble is generated is arbitrary, one obtains

$$E(R) \geq E_{\mathrm{CK,r}}(R)$$

where

$$\begin{aligned} E_{\mathrm{CK,r}}(R) &\triangleq \max_{\mathsf{Q}} E_{\mathrm{CK}}(R, \mathsf{Q}) \\ &= \max_{\mathsf{Q}} \min_{\mathsf{V}(\cdot|\cdot)} \left\{ D(\mathsf{V} \| \mathsf{W} | \mathsf{Q}) + |I(\mathsf{Q}, \mathsf{V}) - R|^{+} \right\}. \end{aligned}$$

There is an alternative form for $E_{\mathrm{CK}}(R, \mathsf{Q})$ that will be of interest to us [1], [2, Ch. 2.5, Problem 23]. This form is more similar to (3)

$$E_{\mathrm{CK}}(R, \mathsf{Q}) = \max_{0 \leq \varrho \leq 1} \{E_{\mathrm{CK},0}(\varrho, \mathsf{Q}) - \varrho R\} \tag{8}$$

where

$$E_{\mathrm{CK},0}(\varrho, \mathsf{Q}) \triangleq \min_{\mathsf{V}(\cdot|\cdot)} \{D(\mathsf{V} \| \mathsf{W} | \mathsf{Q}) + \varrho I(\mathsf{Q}, \mathsf{V})\} \tag{9}$$

$$= \min_{\mathsf{R}} \left\{ -(1+\varrho) \sum_{x \in \mathcal{X}} \mathsf{Q}(x) \cdot \log \sum_{y \in \mathcal{Y}} \mathsf{W}(y|x)^{\frac{1}{1+\varrho}} \mathsf{R}(y)^{\frac{\varrho}{1+\varrho}} \right\} \tag{10}$$

and where the minimization in the latter is over the set of all distributions R on the output alphabet $\mathcal{Y}$.

In general, for any DMC $\mathsf{W}(y|x)$ and any input distribution Q [1], [2, Ch. 2.5, Problem 23]

$$E_{\mathrm{CK},0}(\varrho, \mathsf{Q}) \geq E_{\mathrm{G},0}(\varrho, \mathsf{Q}), \qquad \varrho \geq 0 \tag{11}$$

and hence,

$$E_{\mathrm{CK}}(R, \mathsf{Q}) \geq E_{\mathrm{G}}(R, \mathsf{Q}) \tag{12}$$

with the inequalities typically being strict. These inequalities are a consequence of the fact that the "average constant composition code" performs better than the "average independently and identically distributed code" [7].

However, when optimized over the input distribution, the inequalities turn into equalities [1], [8], [2, Ch. 2.5, Problem 23]

$$\max_{\mathsf{Q}} E_{\mathrm{CK},0}(\varrho, \mathsf{Q}) = \max_{\mathsf{Q}} E_{\mathrm{G},0}(\varrho, \mathsf{Q}), \qquad \varrho \geq 0 \tag{13}$$

and

$$\max_{\mathsf{Q}} E_{\mathrm{CK}}(R, \mathsf{Q}) = \max_{\mathsf{Q}} E_{\mathrm{G}}(R, \mathsf{Q}) \tag{14}$$

i.e.,

$$E_{\mathrm{CK,r}}(R) = E_{\mathrm{G,r}}(R). \tag{15}$$

In fact, the optimization problems appearing on the left-hand side (LHS) and on the right-hand side (RHS) of (13) are Lagrange duals.

Proposition 1: For any discrete memoryless channel with unconstrained input sequences and any $\varrho > 0$, the two maximization problems in (13) are related by strong Lagrange duality. More specifically, the problem

$$\min_{\mathsf{Q}} \exp\{-E_{\mathrm{CK},0}(\varrho, \mathsf{Q})\} \tag{16}$$

is a strong Lagrange dual of the problem

$$\min_{\mathsf{Q}} \exp\{-E_{\mathrm{G},0}(\varrho, \mathsf{Q})\}. \tag{17}$$

Proof: For the purposes of this proof, we replace the information-theoretic notation we have used so far with a notation that is more standard for optimization problems. Namely, let the row vector $\boldsymbol{q} \in \mathbb{R}^{1\times|\mathcal{X}|}$ denote a probability vector on $\mathcal{X}$, and let $\boldsymbol{w} \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{Y}|}$ be the matrix whose (i,j)th element is given by

$$w_{ij} = \mathsf{W}(y_j|x_i)^{\frac{1}{\varrho+1}}, \quad i \in \{1,\dots,|\mathcal{X}|\}, \ \ j \in \{1,\dots,|\mathcal{Y}|\},$$

x_i, $i \in \{1,\dots,|\mathcal{X}|\}$ and y_j, $j \in \{1,\dots,|\mathcal{Y}|\}$ are the distinct elements of the sets $\mathcal{X}$ and $\mathcal{Y}$, respectively. We reformulate (17) by using (4) and introducing the auxiliary row vector $\boldsymbol{f} \in \mathbb{R}^{1\times|\mathcal{Y}|}$, given by $\boldsymbol{f} \triangleq \boldsymbol{q}\boldsymbol{w}$, as follows:

$$\begin{aligned} \min_{\mathsf{Q}} \exp\{-E_{\mathrm{G},0}(\varrho, \mathsf{Q})\} &= \min_{(\boldsymbol{q},\boldsymbol{f})\in \mathrm{D}} \sum_{j=1}^{|\mathcal{Y}|} f_j^{1+\varrho} \\ \text{s.t.} \quad \boldsymbol{q}\boldsymbol{w} = \boldsymbol{f}&, \ \ \boldsymbol{q}\mathbf{1} = 1 \end{aligned} \tag{18}$$

where the domain D of this optimization problem is $\mathrm{D} = \{(\boldsymbol{q},\boldsymbol{f}) | \boldsymbol{q} \succeq \mathbf{0}\}$, and where $\mathbf{1}$ and $\mathbf{0}$ denote the all-one and all-zero vectors of a size specified by the context. Here we use $\boldsymbol{a} \succeq \boldsymbol{b}$ to indicate that each component of the vector $\boldsymbol{a} - \boldsymbol{b}$ is nonnegative. For any $\varrho > 0$, the objective function is convex in D. Furthermore, all equality and inequality constraints are affine. Hence, the problem is a convex optimization problem.

Since the objective of the problem is nondecreasing in f_j, $j \in \{1,\dots,|\mathcal{Y}|\}$ and constant in $q_i, i \in \{1,\dots,|\mathcal{X}|\}$, we replace the equality constraints $\boldsymbol{q}\boldsymbol{w} = \boldsymbol{f}$ with $\boldsymbol{f} \succeq \boldsymbol{q}\boldsymbol{w}$ without decreasing the value of the solution. Therefore, the problem is restated as follows:

$$\begin{aligned} \min_{\mathsf{Q}} \exp\{-E_{\mathrm{G},0}(\varrho, \mathsf{Q})\} &= \min_{(\boldsymbol{q},\boldsymbol{f})\in \mathrm{D}} \sum_{j=1}^{|\mathcal{Y}|} f_j^{1+\varrho} \\ \text{s.t.} \quad \boldsymbol{f} \succeq \boldsymbol{q}\boldsymbol{w}&, \ \ \boldsymbol{q}\mathbf{1} = 1. \end{aligned}$$

The Lagrangian function L of this problem is [9, Ch. 5]

$$L(\boldsymbol{q},\boldsymbol{f},\boldsymbol{\nu},\mu) = \sum_{j=1}^{|\mathcal{Y}|} f_j^{1+\varrho} + (\boldsymbol{q}\boldsymbol{w} - \boldsymbol{f})\boldsymbol{\nu} + (1 - \boldsymbol{q}\mathbf{1})\mu$$

where $\boldsymbol{\nu} \succeq \mathbf{0} \in \mathbb{R}^{|\mathcal{Y}|\times 1}$ and $\mu \in \mathbb{R}$ are the Lagrange dual variables, and $(\boldsymbol{q},\boldsymbol{f}) \in \mathrm{D}$. The Lagrange dual function [9]

$$g(\boldsymbol{\nu},\mu) \triangleq \inf_{(\boldsymbol{q},\boldsymbol{f})\in\mathrm{D}} L(\boldsymbol{q},\boldsymbol{f},\boldsymbol{\nu},\mu)$$

is, by calculation, given by

$$g(\boldsymbol{\nu},\mu) = \begin{cases} -\varrho \sum_{j=1}^{|\mathcal{Y}|} \left(\frac{\nu_j}{1+\varrho}\right)^{\frac{1+\varrho}{\varrho}} + \mu, & \text{if } \mu\mathbf{1} \preceq \boldsymbol{w}\boldsymbol{\nu} \\ -\infty, & \text{otherwise.} \end{cases}$$

This is best seen by writing

$$L(\boldsymbol{q},\boldsymbol{f},\boldsymbol{\nu},\mu) = \sum_{j=1}^{|\mathcal{Y}|} f_j^{1+\varrho} - \boldsymbol{f}\boldsymbol{\nu} + \mu + \boldsymbol{q}(\boldsymbol{w}\boldsymbol{\nu} - \mathbf{1}\mu)$$

and by then noting that if $\boldsymbol{w}\boldsymbol{\nu} - \mathbf{1}\mu \succeq \mathbf{0}$ then an optimal choice for $\boldsymbol{q}$ is $\mathbf{0}$ and that otherwise $\inf_{\boldsymbol{q}\succeq\mathbf{0}}\{\boldsymbol{q}(\boldsymbol{w}\boldsymbol{\nu} - \mathbf{1}\mu)\}$ is $-\infty$.

The Lagrange dual problem is then

$$\max_{\boldsymbol{\nu}\succeq\mathbf{0},\mu\in\mathbb{R}} g(\boldsymbol{\nu},\mu)$$

i.e.,

$$\begin{aligned} &\max_{\boldsymbol{\nu},\mu} \quad \left\{-\varrho \sum_{j=1}^{|\mathcal{Y}|} \left(\frac{\nu_j}{1+\varrho}\right)^{\frac{1+\varrho}{\varrho}} + \mu\right\} \\ &\text{s.t.} \ \ \mu\mathbf{1} \preceq \boldsymbol{w}\boldsymbol{\nu}, \ \ \boldsymbol{\nu} \succeq \mathbf{0}. \end{aligned}$$

Note that in the above Lagrange dual problem the constraint

$$\mu\mathbf{1} \preceq \boldsymbol{w}\boldsymbol{\nu} \tag{19}$$

is equivalent to the constraint

$$\mu \leq \min_{i\in\{1,\dots,|\mathcal{X}|\}} \sum_{j=1}^{|\mathcal{Y}|} w_{ij}\nu_j. \tag{20}$$

In fact, there is no loss in optimality by insisting that

$$\mu = \min_{i\in\{1,\dots,|\mathcal{X}|\}} \sum_{j=1}^{|\mathcal{Y}|} w_{ij}\nu_j \tag{21}$$

because if (20) is satisfied with a strict inequality, then $\boldsymbol{\nu}$ can be scaled by $\frac{1}{1+\epsilon}$, for $\epsilon > 0$ sufficiently small (with μ fixed), and this increases the value of the function g. Consequently, upon replacing (19) with (21), we obtain

$$\max_{\boldsymbol{\nu}\succeq\mathbf{0},\mu\in\mathbb{R}} g(\boldsymbol{\nu},\mu) = \max_{\boldsymbol{\nu}\succeq\mathbf{0}} \left\{-\varrho \sum_{j=1}^{|\mathcal{Y}|} \left(\frac{\nu_j}{1+\varrho}\right)^{\frac{1+\varrho}{\varrho}} + \min_{i\in\{1,\dots,|\mathcal{X}|\}} \left\{\sum_{j=1}^{|\mathcal{Y}|} w_{ij}\nu_j\right\}\right\}.$$

We next perform the transformation of variables

$$\nu_j = (1+\varrho)\alpha r_j^{\frac{\varrho}{1+\varrho}}, \qquad j \in \{1,\dots,|\mathcal{Y}|\}$$

where $\boldsymbol{r} \in \mathbb{R}^{|\mathcal{Y}|\times 1}$ is a probability vector on $\mathcal{Y}$ and $\alpha \in \mathbb{R}^{+}$ is the appropriate normalizing scalar parameter. Optimizing over α yields

$$\max_{\boldsymbol{\nu}\succeq \boldsymbol{0},\mu\in\mathbb{R}} g(\boldsymbol{\nu},\mu) = \max_{\boldsymbol{r}} \left(\min_{i\in\{1,\dots,|\mathcal{X}|\}} \sum_{j=1}^{|\mathcal{Y}|} w_{ij} r_j^{\frac{\varrho}{1+\varrho}} \right)^{1+\varrho} \tag{22}$$

$$\text{s.t.} \quad \boldsymbol{r}\succeq \boldsymbol{0}, \quad \boldsymbol{r}\boldsymbol{1}=1. \tag{23}$$

The primal problem (18) is a convex optimization problem satisfying Slater's condition. Hence, its dual problem given in (22) and (23) is a strong dual problem and provides a solution to (18) with no loss in optimality. We shall now conclude the proof by showing that the problem stated in (22) and (23) is equivalent to (16). Beginning with (16) we have

$$\begin{aligned}
&\min_{\mathsf{Q}} \exp\{-E_{\mathrm{CK},0}(\varrho,\mathsf{Q})\} \\
&= \min_{\boldsymbol{q}} \max_{\boldsymbol{r}} \left\{ \exp\left((1+\varrho) \sum_{i=1}^{|\mathcal{X}|} q_i \log \left(\sum_{j=1}^{|\mathcal{Y}|} w_{ij} r_j^{\frac{\varrho}{1+\varrho}} \right)\right)\right\} \\
&= \max_{\boldsymbol{r}} \min_{\boldsymbol{q}} \left\{ \exp\left((1+\varrho) \sum_{i=1}^{|\mathcal{X}|} q_i \log \left(\sum_{j=1}^{|\mathcal{Y}|} w_{ij} r_j^{\frac{\varrho}{1+\varrho}} \right)\right)\right\} \\
&= \max_{\boldsymbol{r}} \left\{ \exp\left((1+\varrho) \min_{\boldsymbol{q}} \sum_{i=1}^{|\mathcal{X}|} q_i \log \left(\sum_{j=1}^{|\mathcal{Y}|} w_{ij} r_j^{\frac{\varrho}{1+\varrho}} \right)\right)\right\} \\
&= \max_{\boldsymbol{r}} \left\{ \exp\left((1+\varrho) \min_{i\in\{1,\dots,|\mathcal{X}|\}} \log \left(\sum_{j=1}^{|\mathcal{Y}|} w_{ij} r_j^{\frac{\varrho}{1+\varrho}} \right)\right)\right\} \\
&= \max_{\boldsymbol{r}} \left\{ \exp\left((1+\varrho) \log \min_{i\in\{1\dots,|\mathcal{X}|\}} \left\{ \sum_{j=1}^{|\mathcal{Y}|} w_{ij} r_j^{\frac{\varrho}{1+\varrho}} \right\}\right)\right\} \\
&= \max_{\boldsymbol{r}} \left\{ \left(\min_{i\in\{1,\dots,|\mathcal{X}|\}} \left\{ \sum_{j=1}^{|\mathcal{Y}|} w_{ij} r_j^{\frac{\varrho}{1+\varrho}} \right\}\right)^{1+\varrho}\right\}
\end{aligned}$$

where the first equality follows from (10); the second by the minimax theorem; the third by the monotonicity of the exponential function; the fourth by computing the minimum; the fifth by the monotonicity of the logarithm; and the last equality by algebra. □

In view of (13)–(15), we shall henceforth denote $\max_{\mathsf{Q}} E_{\mathrm{CK},0}(\varrho,\mathsf{Q})$ $(= \max_{\mathsf{Q}} E_{\mathrm{G},0}(\varrho,\mathsf{Q}))$ by $E_0(\varrho)$ and refer to $E_{\mathrm{G,r}}(R)$ $(=E_{\mathrm{CK,r}}(R))$ as the random coding error exponent and denote it by $E_{\mathrm{r}}(R)$. In terms of the function $E_0(\varrho)$, the random coding error exponent $E_{\mathrm{r}}(R)$ is thus given by

$$E_{\mathrm{r}}(R) = \max_{0\le\varrho\le 1} \{E_0(\varrho) - \varrho R\}. \tag{24}$$

The cutoff rate R_0 is defined by

$$R_0 = E_0(\varrho)|_{\varrho=1}. \tag{25}$$

The function $E_0(\varrho)$ plays an important role not only in the study of lower bounds on the reliability function $E(R)$ but also in the study of upper bounds. In fact, the sphere packing error exponent $E_{\mathrm{sp}}(R)$ is given by [5]

$$E_{\mathrm{sp}}(R) = \max_{\varrho>0} \{E_0(\varrho) - \varrho R\}. \tag{26}$$

Combining (13) with (10) and (4) we obtain the two equivalent expressions for $E_0(\varrho)$

$$E_0(\varrho) = \max_{\mathsf{Q}} \left\{ -\log \sum_{y\in\mathcal{Y}} \left(\sum_{x\in\mathcal{X}} \mathsf{Q}(x)\mathsf{W}(y|x)^{\frac{1}{1+\varrho}} \right)^{1+\varrho} \right\} \tag{27}$$

$$E_0(\varrho) = \max_{\mathsf{Q}} \min_{\mathsf{R}} \left\{ -(1+\varrho) \sum_{x\in\mathcal{X}} \mathsf{Q}(x) \cdot \log \sum_{y\in\mathcal{Y}} \mathsf{W}(y|x)^{\frac{1}{1+\varrho}} \mathsf{R}(y)^{\frac{\varrho}{1+\varrho}} \right\}. \tag{28}$$

We refer to the former expression as the "primal" expression and to the latter as the "dual" expression. The primal expression is useful for the derivation of lower bounds on $E_0(\varrho)$. Indeed, any distribution Q on the input alphabet $\mathcal{X}$ induces the lower bound

$$E_0(\varrho) \ge -\log \sum_{y\in\mathcal{Y}} \left(\sum_{x\in\mathcal{X}} \mathsf{Q}(x)\mathsf{W}(y|x)^{\frac{1}{1+\varrho}} \right)^{1+\varrho}.$$

On the other hand, the dual expression is useful for the derivation of upper bounds. Any distribution R on the output alphabet $\mathcal{Y}$ yields the upper bound

$$\begin{aligned}
E_0(\varrho) &\le \max_{\mathsf{Q}} \left\{ -(1+\varrho) \sum_{x\in\mathcal{X}} \mathsf{Q}(x) \cdot \log \sum_{y\in\mathcal{Y}} \mathsf{W}(y|x)^{\frac{1}{1+\varrho}} \mathsf{R}(y)^{\frac{\varrho}{1+\varrho}} \right\} \\
&= \max_{x\in\mathcal{X}} \left\{ -(1+\varrho) \log \sum_{y\in\mathcal{Y}} \mathsf{W}(y|x)^{\frac{1}{1+\varrho}} \mathsf{R}(y)^{\frac{\varrho}{1+\varrho}} \right\}.
\end{aligned}$$

B. Constrained Inputs

Before we can use the above bounds for fading channels, we need to extend the discussion to cost-constrained channels and to channels over infinite input and output alphabets, where the method of types cannot be directly used. For now, we continue our assumption of finite alphabets and address the cost constraint. The case of infinite input and output alphabets is treated in Section II.

Suppose we restrict ourselves to block code transmissions where we only allow codewords $(x_1,\dots,x_n)$ that satisfy

$$\sum_{\ell=1}^{n} g(x_\ell) \le n\Upsilon \tag{29}$$

where $g : \mathcal{X} \to \mathbb{R}^+$ is a cost function on the input alphabet $\mathcal{X}$; the allowed average cost $\Upsilon \geq 0$ is some prespecified number; and n, as before, is the block length. The reliability function, denoted now $E(R, \Upsilon)$, is defined as in (1) with the modification that $\mathrm{P_e}(n, R)$ should be now understood as the lowest average probability of error that can be achieved using a rate-R block-length-n codebook all of whose codewords satisfy the cost constraint.

To obtain lower bounds on $E(R, \Upsilon)$, Gallager [5], [10] modified his random coding argument in two ways. He introduced a new ensemble of codebooks and introduced an improved technique to analyze the average probability of error over this ensemble. For any probability law Q on the input alphabet satisfying

$$\mathsf{E}_{\mathsf{Q}}[g(X)] \leq \Upsilon \tag{30}$$

where $\mathsf{E}_{\mathsf{Q}}[\cdot]$ denotes expectation with respect to the law Q, so that

$$\mathsf{E}_{\mathsf{Q}}[g(X)] \triangleq \sum_{x \in \mathcal{X}} \mathsf{Q}(x) g(x) \tag{31}$$

define

$$E^{\mathrm{m}}_{\mathrm{G},0}(\varrho, \mathsf{Q}) \triangleq \begin{cases} E_{\mathrm{G},0}(\varrho, \mathsf{Q}), & \text{if } \mathsf{E}_{\mathsf{Q}}[g(X)] < \Upsilon \\ \max_{r \geq 0} E_0(\varrho, \mathsf{Q}, r), & \text{if } \mathsf{E}_{\mathsf{Q}}[g(X)] = \Upsilon \end{cases} \tag{32}$$

where

$$E_0(\varrho, \mathsf{Q}, r) \triangleq -\log \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} \mathsf{Q}(x) e^{r(g(x) - \Upsilon)} \mathsf{W}(y|x)^{\frac{1}{1+\varrho}} \right)^{1+\varrho}. \tag{33}$$

Note that

$$E_0(\varrho, \mathsf{Q}, r)|_{r=0} = E_{\mathrm{G},0}(\varrho, \mathsf{Q}) \tag{34}$$

and hence,

$$\max_{r \geq 0} E_0(\varrho, \mathsf{Q}, r) \geq E^{\mathrm{m}}_{\mathrm{G},0}(\varrho, \mathsf{Q}) \geq E_{\mathrm{G},0}(\varrho, \mathsf{Q}). \tag{35}$$

Thus, Gallager's "modification" can only tighten the bound.

Gallager then showed that for any $0 \leq \varrho \leq 1$, the exponent

$$E^{\mathrm{m}}_{\mathrm{G},0}(\varrho, \mathsf{Q}) - \varrho R$$

is achievable using block codes that satisfy the constraint. (To prove this result when $\mathsf{E}_{\mathsf{Q}}[g(X)] < \Upsilon$, he considered an ensemble of codebooks where the codewords are chosen independently of each other, each according to the *a posteriori* law of a sequence $X_1, \ldots, X_n$ drawn independent and identically distributed according to Q conditional on $\sum_{\ell=1}^{n} g(X_\ell) \leq n\Upsilon$. To prove the result when $\mathsf{E}_{\mathsf{Q}}[g(X)] = \Upsilon$, he considered an ensemble similarly constructed but with the distribution being conditional on $n\Upsilon - \delta \leq \sum_{\ell=1}^{n} g(X_\ell) \leq n\Upsilon$.)

Consequently, the error exponent

$$E^{\mathrm{m}}_{\mathrm{G,r}}(R, \Upsilon) \triangleq \max_{0 \leq \varrho \leq 1} \left\{ E^{\mathrm{m}}_{\mathrm{G},0}(\varrho, \Upsilon) - \varrho R \right\} \tag{36}$$

where

$$E^{\mathrm{m}}_{\mathrm{G},0}(\varrho, \Upsilon) \triangleq \max_{\mathsf{Q}: \mathsf{E}_{\mathsf{Q}}[g(X)] \leq \Upsilon} E^{\mathrm{m}}_{\mathrm{G},0}(\varrho, \mathsf{Q}) \tag{37}$$

is achievable.

It is instructive to distinguish between two types of constraints. We say that the cost constraint is *inactive* if there exists some input distribution Q^* satisfying the constraint that achieves the global unconstrained maximum of $E_{\mathrm{G},0}(\varrho, \mathsf{Q})$. That is, the cost is inactive if

$$\exists \mathsf{Q}^* : \mathsf{E}_{\mathsf{Q}^*}[g(X)] \leq \Upsilon \text{ and } E_{\mathrm{G},0}(\varrho, \mathsf{Q}^*) = \max_{\mathsf{Q}} E_{\mathrm{G},0}(\varrho, \mathsf{Q})$$

(where the maximization in the above is over *all* input distributions Q on $\mathcal{X}$) or, equivalently, if

$$\max_{\mathsf{Q}: \mathsf{E}_{\mathsf{Q}}[g(X)] \leq \Upsilon} E_{\mathrm{G},0}(\varrho, \mathsf{Q}) = \max_{\mathsf{Q}} E_{\mathrm{G},0}(\varrho, \mathsf{Q}). \tag{38}$$

Otherwise, we say that the cost constraint is *active*. With these definitions it can be shown that (37) simplifies to

$$E^{\mathrm{m}}_{\mathrm{G},0}(\varrho, \Upsilon) = \begin{cases} \max_{\mathsf{Q}: \mathsf{E}_{\mathsf{Q}}[g(X)] = \Upsilon} \max_{r \geq 0} E_0(\varrho, \mathsf{Q}, r), & \text{cost active} \\ \max_{\mathsf{Q}} E_{\mathrm{G},0}(\varrho, \mathsf{Q}), & \text{cost inactive.} \end{cases} \tag{39}$$

(The case where the cost constraint is active follows from Gallager's observation that when the cost constraint is active, the maximum of $E_0(\varrho, \mathsf{Q}, r)$ over all $r \geq 0$ and over all laws Q satisfying (30) is achieved by an input distribution Q satisfying the constraint with equality. The case where the cost constraint is inactive follows by noting that by starting from (35) we have for inactive cost constraints

$$\begin{aligned} \max_{\mathsf{Q}: \mathsf{E}_{\mathsf{Q}}[g(X)] \leq \Upsilon} E^{\mathrm{m}}_{\mathrm{G},0}(\varrho, \mathsf{Q}) &\geq \max_{\mathsf{Q}: \mathsf{E}_{\mathsf{Q}}[g(X)] \leq \Upsilon} E_{\mathrm{G},0}(\varrho, \mathsf{Q}) \\ &= \max_{\mathsf{Q}} E_{\mathrm{G},0}(\varrho, \mathsf{Q}) \\ &= \max_{\mathsf{Q}} E_{\mathrm{CK},0}(\varrho, \mathsf{Q}) \\ &\geq \max_{\mathsf{Q}: \mathsf{E}_{\mathsf{Q}}[g(X)] \leq \Upsilon} E^{\mathrm{m}}_{\mathrm{G},0}(\varrho, \mathsf{Q}) \end{aligned}$$

so that all inequalities must hold with equalities. Here, the first inequality follows from (35); the subsequent equality because the cost constraint is assumed inactive (38); the subsequent equality from (13); and the final inequality from (42) ahead.)

An achievable error exponent can also be established using constant composition codes. This yields that the error exponent

$$E_{\mathrm{CK,r}}(R,\Upsilon) \triangleq \max_{0\le\varrho\le 1}\{E_{\mathrm{CK},0}(\varrho,\Upsilon)-\varrho R\} \tag{40}$$

is achievable where

$$E_{\mathrm{CK},0}(\varrho,\Upsilon) \triangleq \max_{\mathsf{Q}:\mathsf{E}_{\mathsf{Q}}[g(X)]\le\Upsilon} E_{\mathrm{CK},0}(\varrho,\mathsf{Q}). \tag{41}$$

The relation (35) notwithstanding, it can be shown that for any law Q satisfying (30) and any $\varrho \ge 0$

$$E_{\mathrm{CK},0}(\varrho,\mathsf{Q}) \ge E^{\mathrm{m}}_{\mathrm{G},0}(\varrho,\mathsf{Q}) \tag{42}$$

with the inequality being, in general, strict.[1] Consequently, by (41) and (37), we have $E_{\mathrm{CK},0}(\varrho,\Upsilon) \ge E^{\mathrm{m}}_{\mathrm{G},0}(\varrho,\Upsilon)$. However, as shown in Appendix I, this holds with equality

$$E_{\mathrm{CK},0}(\varrho,\Upsilon) = E^{\mathrm{m}}_{\mathrm{G},0}(\varrho,\Upsilon). \tag{43}$$

Thus, denoting the two identical functions $E^{\mathrm{m}}_{\mathrm{G},0}(\varrho,\Upsilon)$ and $E_{\mathrm{CK},0}(\varrho,\Upsilon)$ by $E_0(\varrho,\Upsilon)$ and the two identical functions $E_{\mathrm{CK,r}}(R,\Upsilon)$ and $E^{\mathrm{m}}_{\mathrm{G,r}}(R,\Upsilon)$ by $E_{\mathrm{r}}(R,\Upsilon)$, we have

$$E_{\mathrm{r}}(R,\Upsilon) = \max_{0\le\varrho\le 1}\{E_0(\varrho,\Upsilon)-\varrho R\} \tag{44}$$

where $E_0(\varrho,\Upsilon)$ can be expressed either, using (39), as

$$E_0(\varrho,\Upsilon) = \begin{cases} \max\limits_{\mathsf{Q}:\mathsf{E}_{\mathsf{Q}}[g(X)]=\Upsilon} \max\limits_{r\ge 0} E_0(\varrho,\mathsf{Q},r), & \text{cost active} \\ \max\limits_{\mathsf{Q}} E_{\mathrm{G},0}(\varrho,\mathsf{Q}), & \text{cost inactive} \end{cases} \tag{45}$$

or, using (10), as

$$E_0(\varrho,\Upsilon) = \max_{\mathsf{Q}:\mathsf{E}_{\mathsf{Q}}[g(X)]\le\Upsilon} \min_{\mathsf{R}} \left\{ -(1+\varrho)\sum_{x\in\mathcal{X}} \mathsf{Q}(x) \cdot \log \sum_{y\in\mathcal{Y}} \mathsf{W}(y|x)^{\frac{1}{1+\varrho}} \mathsf{R}(y)^{\frac{\varrho}{1+\varrho}} \right\}. \tag{46}$$

[1]In the case $\mathsf{E}_{\mathsf{Q}}[g(X)] < \Upsilon$, this follows directly from (12). For a proof in the case $\mathsf{E}_{\mathsf{Q}}[g(X)] = \Upsilon$ see Proposition 2 ahead, which proves that the RHS of (10) is greater or equal to $E^{\mathrm{m}}_{\mathrm{G},0}(\varrho,\mathsf{Q})$.

The former expression, to which we refer as the "primal" expression, is useful for the derivation of lower bounds on $E_0(\varrho,\Upsilon)$, whereas the latter, the "dual," is useful for upper bounds.

II. Continuous Alphabets

We next extend the discussion to channels with infinite input and output alphabets. Consider a channel $W(\cdot|\cdot)$ whose input and output take value in the separable metric spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively. Thus, for any input $x\in\mathcal{X}$ and any Borel set $\mathcal{B}\subset\mathcal{Y}$, the probability that in response to the input x the channel will produce an output Y that lies in the set $\mathcal{B}$ is $W(\mathcal{B}|x)$. We assume that the mapping $x\mapsto W(\mathcal{B}|x)$ from $\mathcal{X}$ to the interval $[0,1]$ is Borel measurable. Finally, assume the existence of an underlying positive measure μ on $\mathcal{Y}$ with respect to which all the probability measures $\{W(\cdot|x), x\in\mathcal{X}\}$ are absolutely continuous. Denote the Radon-Nykodim derivative of $W(\cdot|x)$ with respect to μ by

$$w(\cdot|x) = \frac{\mathrm{d}W(\cdot|x)}{\mathrm{d}\mu}, \qquad x\in\mathcal{X}.$$

Thus, $w(y|x)$ is the density at y of the channel output corresponding to the input $x\in\mathcal{X}$. For any input $x\in\mathcal{X}$ and any Borel set $\mathcal{B}\subset\mathcal{Y}$

$$W(\mathcal{B}|x) = \int_{\mathcal{B}} w(y|x)\mathrm{d}\mu(y).$$

As to the cost, we shall assume that the function $g:\mathcal{X}\to\mathbb{R}^+$ is measurable and consider block codes that satisfy (29). We extend the definition (31) to infinite alphabets as

$$\mathsf{E}_{\mathsf{Q}}[g(X)] \triangleq \int_{\mathcal{X}} g(x)\mathrm{d}\mathsf{Q}(x). \tag{47}$$

Definition (33) is extended for any probability law Q on $\mathcal{X}$ as shown in the first equation at the bottom of the page. For any input distribution Q satisfying the constraint $\mathsf{E}_{\mathsf{Q}}[g(X)]\le\Upsilon$ we extend (32) as shown in (48) at the bottom of the page. (Note that following Gallager [5], [10] we allow for the optimization over r only when under the law Q the random variable $g(X)$ has a finite third moment.)

With this definition we can now define

$$E_0(\varrho,\Upsilon) \triangleq \sup_{\mathsf{Q}:\mathsf{E}_{\mathsf{Q}}[g(X)]\le\Upsilon} E^{\mathrm{m}}_{\mathrm{G},0}(\varrho,\mathsf{Q}) \tag{49}$$

$$E_0(\varrho,\mathsf{Q},r) \triangleq -\log\int_{y\in\mathcal{Y}} \left(\int_{x\in\mathcal{X}} e^{r(g(x)-\Upsilon)} w(y|x)^{\frac{1}{1+\varrho}} \mathrm{d}\mathsf{Q}(x)\right)^{1+\varrho} \mathrm{d}\mu(y).$$

$$E^{\mathrm{m}}_{\mathrm{G},0}(\varrho,\mathsf{Q}) \triangleq \begin{cases} \sup\limits_{r\ge 0} E_0(\varrho,\mathsf{Q},r), & \text{if } \mathsf{E}_{\mathsf{Q}}[g(X)]=\Upsilon \text{ and } \mathsf{E}_{\mathsf{Q}}[g^3(X)]<\infty \\ E_0(\varrho,\mathsf{Q},r)|_{r=0}, & \text{otherwise.} \end{cases} \tag{48}$$

and the cutoff rate as

$$R_0(\Upsilon) \triangleq E_0(\varrho, \Upsilon)|_{\varrho=1}. \tag{50}$$

The random coding error exponent

$$\sup_{0\leq\varrho\leq 1} \{E_0(\varrho, \Upsilon) - \varrho R\}$$

is achievable with block codes satisfying the constraint (29) [5], [10].

The following proposition extends (42) to the more general case where the alphabets may be continuous. It is particularly useful for the derivation of upper bounds on $E_0(\varrho, \Upsilon)$ and, in particular, on the cutoff rate $R_0(\Upsilon)$ $(= E_0(1, \Upsilon))$.

Proposition 2: Consider as above a discrete-time memoryless infinite alphabet channel $w(y|x)$, an output measure μ, a measurable cost function $g : \mathcal{X} \to \mathbb{R}^+$, and some arbitrary allowed cost $\Upsilon \geq 0$. Let R be an arbitrary probability distribution on $\mathcal{Y}$ of density f_{R} with respect to μ. Then for any distribution Q on $\mathcal{X}$ satisfying the cost constraint $\mathsf{E}_{\mathsf{Q}}[g(X)] \leq \Upsilon$ we have

$$E^{\mathrm{m}}_{\mathrm{G},0}(\varrho, \mathsf{Q}) \leq -(1+\varrho) \cdot \int_{x\in\mathcal{X}} \log \int_{y\in\mathcal{Y}} w(y|x)^{\frac{1}{1+\varrho}} f_{\mathsf{R}}(y)^{\frac{\varrho}{1+\varrho}} \mathrm{d}\mu(y)\mathrm{d}\mathsf{Q}(x). \tag{51}$$

Proof: Distinguish between the case where $\mathsf{E}_{\mathsf{Q}}[g(X)] < \Upsilon$ and the case where $\mathsf{E}_{\mathsf{Q}}[g(X)] = \Upsilon$ and $\mathsf{E}_{\mathsf{Q}}[g^3(X)] < \infty$. In the former case, by (48), $E^{\mathrm{m}}_{\mathrm{G},0}(\varrho, \mathsf{Q}) = E_0(\varrho, \mathsf{Q}, 0)$ and the result follows by an application of Jensen's inequality and Hölder's inequality as in [2, Ch. 2.5, Problem 23]

$$\begin{aligned}
&-(1+\varrho)\int_{x\in\mathcal{X}} \log \int_{y\in\mathcal{Y}} w(y|x)^{\frac{1}{1+\varrho}} f_{\mathsf{R}}(y)^{\frac{\varrho}{1+\varrho}} \mathrm{d}\mu(y)\mathrm{d}\mathsf{Q}(x)\\
&\quad\geq -(1+\varrho)\log \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} w(y|x)^{\frac{1}{1+\varrho}} \cdot f_{\mathsf{R}}(y)^{\frac{\varrho}{1+\varrho}} \mathrm{d}\mu(y)\mathrm{d}\mathsf{Q}(x)\\
&\quad= -(1+\varrho)\log \int_{y\in\mathcal{Y}}\int_{x\in\mathcal{X}} w(y|x)^{\frac{1}{1+\varrho}} \mathrm{d}\mathsf{Q}(x) \cdot f_{\mathsf{R}}(y)^{\frac{\varrho}{1+\varrho}} \mathrm{d}\mu(y)\\
&\quad\geq -\log \int_{y\in\mathcal{Y}} \left(\int_{x\in\mathcal{X}} w(y|x)^{\frac{1}{1+\varrho}} \mathrm{d}\mathsf{Q}(x)\right)^{1+\varrho} \mathrm{d}\mu(y)\\
&\quad= E_0(\varrho, \mathsf{Q}, 0).
\end{aligned}$$

As for the case where $\mathsf{E}_{\mathsf{Q}}[g(X)] = \Upsilon$ (and $\mathsf{E}_{\mathsf{Q}}[g^3(X)] < \infty$) we have for any $r \geq 0$

$$\begin{aligned}
&-(1+\varrho)\int_{x\in\mathcal{X}} \log \int_{y\in\mathcal{Y}} w(y|x)^{\frac{1}{1+\varrho}} f_{\mathsf{R}}(y)^{\frac{\varrho}{1+\varrho}} \mathrm{d}\mu(y)\mathrm{d}\mathsf{Q}(x)\\
&\quad= r(1+\varrho)\left(\mathsf{E}_{\mathsf{Q}}[g(X)] - \Upsilon\right) - (1+\varrho) \cdot \int_{x\in\mathcal{X}} \log \int_{y\in\mathcal{Y}} e^{r(g(x)-\Upsilon)} w(y|x)^{\frac{1}{1+\varrho}} \cdot f_{\mathsf{R}}(y)^{\frac{\varrho}{1+\varrho}} \mathrm{d}\mu(y)\mathrm{d}\mathsf{Q}(x)\\
&\quad= -(1+\varrho)\int_{x\in\mathcal{X}} \log \int_{y\in\mathcal{Y}} e^{r(g(x)-\Upsilon)} w(y|x)^{\frac{1}{1+\varrho}} \cdot f_{\mathsf{R}}(y)^{\frac{\varrho}{1+\varrho}} \mathrm{d}\mu(y)\mathrm{d}\mathsf{Q}(x)\\
&\quad\geq -(1+\varrho) \cdot \log \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} e^{r(g(x)-\Upsilon)} w(y|x)^{\frac{1}{1+\varrho}} \cdot f_{\mathsf{R}}(y)^{\frac{\varrho}{1+\varrho}} \mathrm{d}\mu(y)\mathrm{d}\mathsf{Q}(x)\\
&\quad\geq -\log \int_{y\in\mathcal{Y}} \left(\int_{x\in\mathcal{X}} e^{r(g(x)-\Upsilon)} w(y|x)^{\frac{1}{1+\varrho}} \mathrm{d}\mathsf{Q}(x)\right)^{1+\varrho} \cdot \mathrm{d}\mu(y)\\
&\quad= E_0(\varrho, \mathsf{Q}, r).
\end{aligned}$$

where the second equality follows because in the case we are considering now $\mathsf{E}_{\mathsf{Q}}[g(X)] = \Upsilon$; the first inequality by Jensen's inequality, and the subsequent by Hölder's inequality. The result for this case now follows because $r \geq 0$ in the above is arbitrary. □

To conclude, to derive lower bounds on $E_0(\varrho, \Upsilon)$ we can choose any input distribution Q satisfying the constraint $\mathsf{E}_{\mathsf{Q}}[g(X)] \leq \Upsilon$ to obtain the lower bound

$$E_0(\varrho, \Upsilon) \geq E^{\mathrm{m}}_{\mathrm{G},0}(\varrho, \mathsf{Q}) \tag{52}$$

where $E^{\mathrm{m}}_{\mathrm{G},0}(\varrho, \mathsf{Q})$ is defined in (48).

To derive upper bounds on $E_0(\varrho, \Upsilon)$ we can use the above proposition by choosing some arbitrary output density $f_{\mathsf{R}}(y)$ to obtain

$$E_0(\varrho, \Upsilon) \leq \sup_{\mathsf{Q}:\mathsf{E}_{\mathsf{Q}}[g(X)]\leq\Upsilon} -(1+\varrho) \cdot \int_{x\in\mathcal{X}} \log \int_{y\in\mathcal{Y}} w(y|x)^{\frac{1}{1+\varrho}} f_{\mathsf{R}}(y)^{\frac{\varrho}{1+\varrho}} \mathrm{d}\mu(y)\mathrm{d}\mathsf{Q}(x). \tag{53}$$

III. Ricean Fading Channels and Their Cutoff Rates

The discrete-time memoryless Ricean fading channel with partial receiver side information is a channel whose input x takes value in the complex field $\mathbb{C}$ and whose corresponding output constitutes of a pair of complex random variables Y and S. We shall refer to Y as "the received signal" and to S as the "side information (at the receiver)." The joint distribution of Y, S corresponding to the input $x \in \mathbb{C}$ is best described using the fading complex random variable H and the additive noise complex random variable Z. The received signal Y corresponding to the input $x \in \mathbb{C}$ is given by

$$Y = Hx + Z \tag{54}$$

where the joint distribution of H, S, and Z does not depend on the input x. The additive noise Z is independent of the pair (H, S) and has a circularly symmetric complex Gaussian distribution of positive variance σ^2. The fading H is of mean $d \in \mathbb{C}$—the "specular component"—and it is assumed that $H - d$ is a unit-variance circularly symmetric complex Gaussian random variable.[2] (The specular component d is a measure of the "extent" to which the channel is fading. An extremely large value of $|d|$ corresponds to a channel that is almost an additive noise channel without fading. The case $d = 0$ corresponds to a severely fading channel, a case that is also called "Rayleigh" fading.) The pair S and $H - d$ are jointly circularly symmetric Gaussian random variables. We denote the conditional variance of H given S by ϵ^2. The case where $\epsilon^2 = 1$ corresponds to the case where H and S are independent, in which case the receiver can discard S without loss in information rates. This case corresponds to "noncoherent" fading. In the case $\epsilon^2 = 0$, the receiver can precisely determine the realization of H from S. This corresponds to "coherent detection." Finally, the case $0 < \epsilon^2 < 1$ corresponds to "partially coherent" communication. In this case, S carries some information about H, but it does not fully determine H. In this paper, we shall only consider the case where $\epsilon^2 > 0$. The case $\epsilon^2 = 0$ is much easier to analyze and has already received considerable attention in the literature. See for example, [11]–[13], and the references in the latter.

The special case of Ricean fading with zero specular component $(d = 0)$ is called "Rayleigh fading." The noncoherent $(\epsilon^2 = 1)$ capacity of this channel was studied in [14]–[16]. The coherent case $(\epsilon^2 = 0)$ was studied in [11]. The capacity of the noncoherent Ricean channel ($\epsilon^2 = 1$ and $d \neq 0$) was studied in [16]–[18].

Unless some restrictions are imposed on the input x, the capacity and cutoff rate of this channel are infinite. Two kinds of restrictions are typically considered. The first corresponds to an average power constraint. Here, only block codes where each codeword satisfies (29) with

$$g(x) = |x|^2 \tag{55}$$

are allowed. In this context, rather than denoting the allowed cost by Υ we shall use the more common symbol $\mathcal{E}$, which stands here for the average energy per symbol. That is, we only allow block-length-n codes in which every codeword $x_1, \ldots, x_n$ satisfies

$$\frac{1}{n}\sum_{\ell=1}^{n} |x_\ell|^2 \leq \mathcal{E}. \tag{56}$$

The second type of constraint is a peak power constraint. Here we only allow channel inputs that satisfy

$$|x|^2 \leq \mathcal{E} \tag{57}$$

where $\mathcal{E}$ now stands for the allowed peak power. Such a constraint is best treated by considering the channel as being free of constraints but with the input alphabet now being the set $\{z \in \mathbb{C} : |z|^2 \leq \mathcal{E}\}$.

[2]We shall sometimes refer to such Ricean fading as "normalized Ricean fading" to make it explicit that the fading is of unit variance. "Un-normalized" Ricean fading need not have unit variance. Those can be normalized by scaling the fading and absorbing the scaling into the input power. Note also that there is no loss in generality in assuming that d is real and nonnegative. The more general complex case can be treated by rotating the output.

For both the average and peak power constraints we define the signal-to-noise ratio (SNR) as

$$\text{SNR} \triangleq \frac{\mathcal{E}}{\sigma^2}. \tag{58}$$

Any codebook satisfying the peak power constraint (57) also satisfies the average power constraint (56). Consequently, the capacity and reliability function under the peak constraint cannot exceed those under the average constraint.

Irrespective of whether an average power or a peak power constraint is imposed, at high SNR the capacity $C(\text{SNR}; d, \epsilon^2)$ of this channel is given asymptotically as [16]

$$C(\text{SNR}; d, \epsilon^2) = \log\log \text{SNR} + \log |d|^2 - \text{Ei}\left(-|d|^2\right) - 1 + \log\frac{1}{\epsilon^2} + o(1) \tag{59}$$

where the correction term $o(1)$ depends on the SNR and tends to zero as the SNR tends to infinity. Here, $\text{Ei}(\cdot)$ denotes the Exponential Integral function

$$\text{Ei}(-\xi) = -\int_{\xi}^{\infty} \frac{e^{-t}}{t}\text{d}t, \qquad \xi > 0 \tag{60}$$

and we define the value of the function $\log(\xi) - \text{Ei}(-\xi)$ at $\xi = 0$ as $-\gamma$, where $\gamma \approx 0.577$ denotes Euler's constant. (With this definition the function $\log(\xi) - \text{Ei}(-\xi)$ is continuous from the right at $\xi = 0$.)

Here we shall study the cutoff rate in two cases: in the absence of side information $(\epsilon^2 = 1)$ but where the specular component d is arbitrary, and in the presence of (imperfect) side information $(0 < \epsilon^2 < 1)$ but where the specular component is zero. For the former case we have the following.

Proposition 3: Consider a discrete-time memoryless Ricean fading channel of unit fading variance and of specular component d. Assume that neither transmitter nor receiver have any side information about the fading realization but that both know its law. Then, irrespective of whether a peak or an average power constraint is imposed, the channel's cutoff rate $R_0(\text{SNR}; d, 1)$ is given by

$$\boxed{R_0(\text{SNR}; d, 1) = \log\log \text{SNR} + \frac{|d|^2}{2} - 2\log \text{I}_0\left(\frac{|d|^2}{4}\right) - \log(2\pi) + o(1).} \tag{61}$$

Here $\text{I}_0(\cdot)$ denotes the zeroth-order modified Bessel function of the first kind, which is given by

$$\text{I}_0(\xi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{\xi\cos\theta}\text{d}\theta, \qquad \xi \in \mathbb{R} \tag{62}$$

and the $o(1)$ term is a correction term that depends on the SNR and that approaches zero as the SNR tends to infinity.

Proof: See Section IV-A. □

Fig. 1 depicts the SNR-independent term (i.e., the second-order term) in the high-SNR expansion of channel capacity (59)

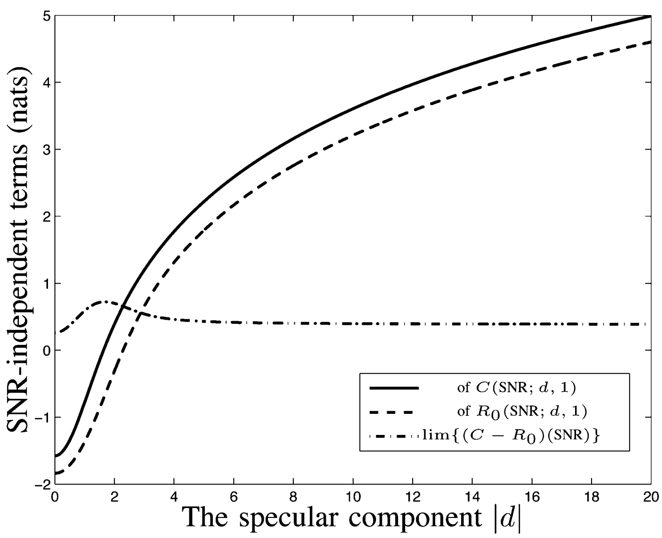

Fig. 1. The SNR-independent terms (i.e., the second-order terms in the high-SNR expansion) of $C(\text{SNR}; d, 1)$ and $R_0(\text{SNR}; d, 1)$ and their difference as a function of the specular component $|d|$ for $\epsilon^2 = 1$, i.e., in the absence of side information.

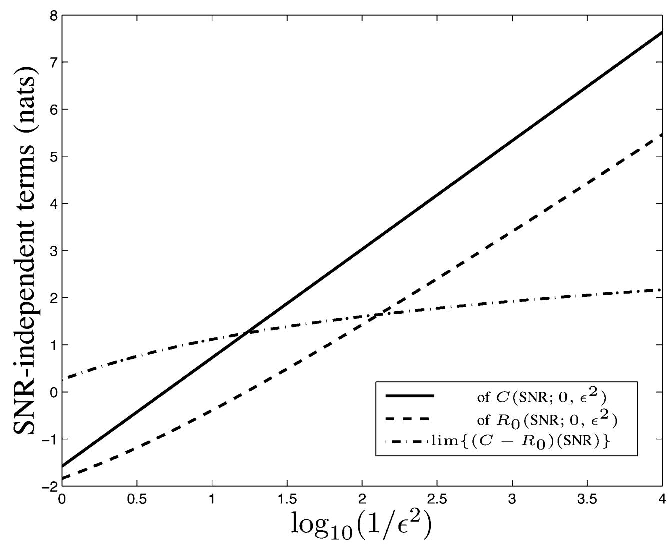

Fig. 2. The SNR-independent terms of $C(\text{SNR}; 0, \epsilon^2)$, of $R_0(\text{SNR}; 0, \epsilon^2)$, and of their difference $C(\text{SNR}; 0, \epsilon^2) - R_0(\text{SNR}; 0, \epsilon^2)$ as functions of $\log_{10} 1/\epsilon^2$, where ϵ^2 is the minimum mean squared error in estimating the fading from the side information. Rayleigh fading ($d = 0$) is assumed.

and of the cutoff rate (61) as functions of the specular component d in the absence of side information. For zero specular component the difference between the two SNR-independent terms is $\log(2\pi) - 1 - \gamma \approx 0.26$ nats; for very large specular components ($|d| \to \infty$) this difference approaches $\log(4/e) \approx 0.39$ nats.[3]

For the case where the side information is present but is not perfect ($0 < \epsilon^2 < 1$) we only treat the case of zero specular component ($d = 0$, i.e., Rayleigh fading).

Proposition 4: Consider a discrete-time memoryless Ricean-fading channel of unit fading variance and of zero specular component $d = 0$. Assume that available to the receiver, but not to the transmitter, is some side information S that is jointly Gaussian with the fading H such that (S, H) is independent of the additive noise Z and such that the joint law of (S, H, Z) does not depend on the channel input. Let $0 < \epsilon^2 < 1$ denote the least estimation variance in estimating the fading H based on the side information S. Let $R_0(\text{SNR}; 0, \epsilon^2)$ denote the channel's cutoff rate. Then, irrespective of whether a peak or an average power constraint is imposed

$$\boxed{\begin{aligned} R_0(\text{SNR}; 0, \epsilon^2) = \;& \log\log \text{SNR} - \log 4 \\ & + \log \tfrac{1}{\epsilon^2} - \log K\left(\sqrt{1-\epsilon^4}\right) + o(1) \end{aligned}} \tag{63}$$

where $\mathrm{K}(\cdot)$ is the complete elliptic integral of the first kind

$$\mathrm{K}(\xi) = \int_0^1 \frac{1}{\sqrt{1-t^2}\sqrt{1-\xi^2 t^2}} \mathrm{d}t, \quad \xi^2 < 1. \tag{64}$$

Proof: See Section IV-B. □

[3]All logarithms in this paper are natural logarithms.

For the case of Rayleigh fading with perfect side information ($\epsilon^2 = 0$) see [12]. For the case of "almost perfect side information" ($0 < \epsilon^2 \ll 1$) we note the expansion

$$\begin{aligned} &\log \frac{1}{\epsilon^2} - \log \mathrm{K}\left(\sqrt{1-\epsilon^4}\right) - \log 4 \\ &\quad = \log \frac{1}{\epsilon^2} - \log\log \frac{4}{\epsilon^2} - \log 4 + o(\epsilon^4), \qquad 0 < \epsilon^2 \ll 1 \end{aligned}$$

which follows from the approximation [19]

$$\mathrm{K}(k) = \frac{1}{1-\theta} \log \frac{4}{\sqrt{1-k^2}}, \qquad 0 \le k < 1 \tag{65}$$

for some

$$0 < \theta < \frac{1-k^2}{4}. \tag{66}$$

Fig. 2 depicts the SNR-independent terms of the channel capacity (59) when $d = 0$, i.e.,

$$C(\text{SNR}; 0, \epsilon^2) = \log\log \text{SNR} + \log \frac{1}{\epsilon^2} - 1 - \gamma + o(1)$$

and the cutoff rate (63) as a function of the estimation error ϵ^2 in estimating the fading from the side information for Rayleigh-fading channels ($d = 0$).

We were unable to find an explicit expansion for the cutoff rate in the general case where $d \neq 0$ and $0 < \epsilon^2 < 1$ because we encountered some complicated integrals that did not seem tractable.

IV. Derivations of the Ricean Cutoff Rate Expansions

In this section, we shall present the derivation of the high-SNR asymptotic expansion of the cutoff rate in the case where side information is absent ($\epsilon^2 = 1$) and the specular component d is arbitrary (61) and in the case where side information is present $(0 < \epsilon^2 < 1)$ and the specular component d is zero (63). We begin with the former.

A. The Cutoff Rate in the Absence of Side Information

The derivation of (61) is performed by deriving upper and lower bounds that asymptotically coincide. We begin with the upper bound.

1) Upper Bound: To derive an upper bound on the cutoff rate of the Ricean channel in the absence of side information we use Proposition 2 with the density (with respect to the Lebesgue measure μ on $\mathbb{C}$)

$$f_{\mathsf{R}}(y) = \frac{(|y|^2+\delta)^{\alpha-1} e^{-\frac{|y|^2+\delta}{\beta}}}{\pi \beta^{\alpha} \Gamma(\alpha, \delta/\beta)}, \qquad y \in \mathbb{C}. \tag{67}$$

Here the parameters $\delta \geq 0$, $\alpha > 0$, and $\beta > 0$ can be chosen freely in order to obtain the tightest bound, and $\Gamma(\alpha, \xi)$ denotes the incomplete Gamma function

$$\Gamma(\alpha, \xi) = \int_{\xi}^{\infty} t^{\alpha-1} e^{-t} \mathrm{d}t, \qquad \alpha > 0, \ \xi \geq 0. \tag{68}$$

This family of densities was introduced in [16] for the purpose of studying the channel capacity and the fading number. Since any choice of f_{R} in Proposition 2 leads to an upper bound on the cutoff rate, this choice requires no mathematical justification. The intuition for this choice is that this family of distributions was shown in [16] to be rich enough to approximate the capacity-achieving output distribution of the Ricean-fading channel at high SNR. It is thus also natural to hope that it may also be rich enough to yield asymptotically optimal upper bounds on the cutoff rate. Not less importantly, for the case at hand where side information is absent, it also leads to tractable analytic calculations.

By Proposition 2 applied with $\varrho = 1$ we obtain for any law Q under which

$$\mathsf{E}_{\mathsf{Q}}[|X|^2] \leq \mathcal{E} \tag{69}$$

the upper bound

$$E_{\mathrm{G},0}^{\mathrm{m}}(1, \mathsf{Q}) \leq -2 \int_{x \in \mathbb{C}} \log \psi(x) \mathrm{d}\mathsf{Q}(x) \tag{70}$$

where

$$\psi(x) \triangleq \int_{y \in \mathbb{C}} \sqrt{w(y|x) \cdot f_{\mathsf{R}}(y)} \mathrm{d}\mu(y) \tag{71}$$

$$= \frac{2 \exp\left(\frac{-\delta}{2\beta}\right) \cdot \exp\left(-\frac{|d|^2|x|^2}{2(|x|^2+\sigma^2)}\right)}{\sqrt{\Gamma(\alpha, \frac{\delta}{\beta})} \beta^{\frac{\alpha}{2}} \sqrt{|x|^2+\sigma^2}} \ell(x; \alpha, \beta, \delta) \tag{72}$$

and where [20, Sec. 3.338]

$$\ell(x; \alpha, \beta, \delta) = \int_0^{\infty} \exp\left(-\frac{\rho^2(\beta+|x|^2+\sigma^2)}{2\beta(|x|^2+\sigma^2)}\right) \cdot \rho(\rho^2+\delta)^{\frac{\alpha-1}{2}} \mathrm{I}_0\left(\frac{|d| \cdot |x| \cdot \rho}{|x|^2+\sigma^2}\right) \mathrm{d}\rho. \tag{73}$$

For our high-SNR analysis it will suffice to consider (for sufficiently large powers $\mathcal{E}$) the possibly suboptimal choice of the parameters

$$\beta = \mathcal{E} \log \mathcal{E}, \quad \alpha = \frac{\delta}{\log \beta} \tag{74}$$

and to consider the limiting behavior of the bound as $\mathcal{E} \to \infty$. After taking this limit with $\delta > 0$ held fixed we shall consider the additional limit of $\delta \to 0$.

The analytic computation of $\ell(x; \alpha, \beta, \delta)$ is difficult. Note, however, that any lower bound to this quantity will yield an upper bound on $E_{\mathrm{G},0}^{\mathrm{m}}(1, \mathsf{Q})$. Also, the integral is computable when both α and δ are formally set to zero.[4] We can thus use a limiting argument to study $\ell(x; \alpha, \beta, \delta)$ for α, δ very small. Indeed, in Appendix II it is shown that for any $m_1 > 0$

$$\ell(x; \alpha, \beta, \delta) \geq a(\alpha, \beta, \delta, m_1) \cdot \ell(x; 0, \beta, 0) \tag{75}$$

where

$$a(\alpha, \beta, \delta, m_1) = \delta^{\alpha/2} \sqrt{\frac{m_1}{m_1+1}} \cdot \left(1 - \frac{\sqrt{m_1 \delta} \cdot \mathrm{I}_0\left(\frac{|d|\sqrt{m_1\delta}}{2\sigma}\right)}{\sqrt{\frac{\pi\beta\sigma^2}{2(\beta+\sigma^2)}}}\right). \tag{76}$$

As we shall see, the term $a(\alpha, \beta, \delta, m_1)$ will have a negligible asymptotic contribution to our bound.

The term $\ell(x; 0, \beta, 0)$ can be computed analytically [20, Sec. 6.618]

$$\ell(x; 0, \beta, 0) = \sqrt{\frac{\pi}{2}} \sqrt{\frac{\beta(|x|^2+\sigma^2)}{\beta+|x|^2+\sigma^2}} \cdot \exp\left(\frac{\beta|d|^2|x|^2}{4(|x|^2+\sigma^2)(\beta+|x|^2+\sigma^2)}\right) \cdot \mathrm{I}_0\left(\frac{\beta|d|^2|x|^2}{4(|x|^2+\sigma^2)(\beta+|x|^2+\sigma^2)}\right). \tag{77}$$

We thus conclude from (70), (72), (75), and (77)

$$\begin{aligned} &E_{\mathrm{G},0}^{\mathrm{m}}(1, \mathsf{Q}) \\ &\leq \frac{\delta}{\beta} - 2\log a(\alpha, \beta, \delta, m_1) + \alpha \log \beta \\ &\quad + \log \Gamma\left(\alpha, \frac{\delta}{\beta}\right) - \log(2\pi) \\ &\quad + \mathsf{E}_{\mathsf{Q}}\left[\log\left(1 + \frac{|X|^2+\sigma^2}{\beta}\right)\right] \end{aligned}$$

[4]In fact, it suffices that δ be set to zero.

$$+ |d|^2 \cdot \mathsf{E}_{\mathsf{Q}}\left[\frac{|X|^2}{|X|^2+\sigma^2}\cdot\left(1-\frac{\beta}{\beta+|X|^2+\sigma^2}\right)\right] + \mathsf{E}_{\mathsf{Q}}\left[\frac{|d|^2}{2}\frac{|X|^2}{|X|^2+\sigma^2}\frac{\beta}{\beta+|X|^2+\sigma^2} -2\log \mathrm{I}_0\left(\frac{|d|^2}{4}\frac{|X|^2}{|X|^2+\sigma^2}\frac{\beta}{\beta+|X|^2+\sigma^2}\right)\right].$$

The expectations in the above cannot be computed without knowledge of the law Q. We thus proceed to upper-bound the expectations using the average power constraint (69). The expectation of the logarithm is upper-bounded using Jensen's inequality and the power constraint (69); the following expectation is upper-bounded using the point-wise upper bound $|x|^2/(|x|^2+\sigma^2) < 1$, Jensen's inequality, and the power constraint (69); and the final expectation by noting that the function $\xi \mapsto \xi - 2\log \mathrm{I}_0(\xi/2)$ is monotonically increasing and by noting that

$$\frac{|d|^2}{2}\frac{|X|^2}{|X|^2+\sigma^2}\frac{\beta}{\beta+|X|^2+\sigma^2} < \frac{|d|^2}{2}.$$

We thus conclude that with the allowed average power $\mathcal{E}$ the cutoff rate satisfies

$$\begin{aligned} R_0\left(\frac{\mathcal{E}}{\sigma^2};d,1\right) - \log\log\frac{\mathcal{E}}{\sigma^2} &\le \frac{\delta}{\beta} - 2\log a(\alpha,\beta,\delta,m_1) + \alpha\log\beta \\ &+ \log\Gamma\left(\alpha,\frac{\delta}{\beta}\right) - \log\log\frac{\mathcal{E}}{\sigma^2} \\ &+ \log\left(1+\frac{\mathcal{E}+\sigma^2}{\beta}\right) \\ &+ \left(1-\frac{\beta}{\beta+\mathcal{E}+\sigma^2}\right)|d|^2 + \frac{|d|^2}{2} \\ &- 2\log \mathrm{I}_0\left(\frac{|d|^2}{4}\right) - \log(2\pi). \end{aligned}$$

Holding $\delta > 0$ (small) and $m_1 > 0$ (large) fixed, and letting $\mathcal{E} \to \infty$ with $\alpha = \alpha(\mathcal{E})$ and $\beta = \beta(\mathcal{E})$ as in (74), we obtain from the above and (76)

$$\begin{aligned} \varlimsup_{\mathcal{E}\to\infty}\left\{R_0\left(\frac{\mathcal{E}}{\sigma^2};d,1\right) - \log\log\frac{\mathcal{E}}{\sigma^2}\right\} &\le \log\left(\frac{m_1+1}{m_1}\right) - 2\log\left(1-\frac{\sqrt{m_1\delta}\cdot \mathrm{I}_0\left(\frac{|d|\sqrt{m_1\delta}}{2\sigma}\right)}{\sqrt{\frac{\pi\sigma^2}{2}}}\right) \\ &+ \log\frac{1-e^{-\delta}}{\delta} + \frac{|d|^2}{2} - 2\log \mathrm{I}_0\left(\frac{|d|^2}{4}\right) - \log(2\pi) \end{aligned}$$

where in computing the limiting difference between the Incomplete Gamma function and $\log\log\mathcal{E}$ we used [16, Appendix XI]. Holding m_1 fixed and letting $\delta \to 0$ we obtain

$$\varlimsup_{\mathcal{E}\to\infty}\left\{R_0\left(\frac{\mathcal{E}}{\sigma^2};d,1\right) - \log\log\frac{\mathcal{E}}{\sigma^2}\right\} \le \log\left(\frac{m_1+1}{m_1}\right) + \frac{|d|^2}{2} - 2\log \mathrm{I}_0\left(\frac{|d|^2}{4}\right) - \log(2\pi).$$

Letting now m_1 tend to infinity we obtain the desired asymptotic upper bound

$$\varlimsup_{\mathcal{E}\to\infty}\left\{R_0\left(\frac{\mathcal{E}}{\sigma^2};d,1\right) - \log\log\frac{\mathcal{E}}{\sigma^2}\right\} \le \frac{|d|^2}{2} - 2\log \mathrm{I}_0\left(\frac{|d|^2}{4}\right) - \log(2\pi). \quad (78)$$

2) Lower Bound: Any input distribution satisfying the cost constraint (possibly strictly) induces a lower bound on the cutoff rate (50). Indeed, for any input distribution $\tilde{\mathsf{Q}}$ satisfying the cost constraint

$$\begin{aligned} R_0(\Upsilon) &\ge E^{\mathrm{m}}_{\mathrm{G},0}(\varrho,\tilde{\mathsf{Q}})|_{\varrho=1} \\ &\ge E_0(\varrho,\tilde{\mathsf{Q}},r)|_{\varrho=1,r=0} \end{aligned}$$

where the first inequality follows by the definition of the cutoff rate (50) (and holds with equality if $\tilde{\mathsf{Q}}$ achieves the cutoff rate) and where the second inequality follows from (48) (and holds with equality if $\tilde{\mathsf{Q}}$ satisfies the cost constraint with strict inequality).

We thus proceed to lower-bound $E_0(1,\tilde{\mathsf{Q}},0)$ for a law $\tilde{\mathsf{Q}}$ of our choice. Under this law, X is a circularly symmetric random variable with

$$\log|X|^2 \sim \mathrm{Uniform}\left(\log\log\mathcal{E},\log\mathcal{E}\right). \quad (79)$$

The motivation for using this law is that it is known to asymptotically achieve capacity [16]. Moreover, this law also satisfies the peak power constraint $|X|^2 \le \mathcal{E}$, so that the lower bound on the cutoff rate we compute will also be valid as a lower bound for the cutoff rate under a peak constraint. Finally, as the next proposition shows, the fact that under $\tilde{\mathsf{Q}}$ the input X satisfies, with probability one, $|X| \ge x_{\min}$, where $x_{\min} \to \infty$ greatly simplifies our analysis. It allows us to asymptotically ignore the additive noise.

Proposition 5: Let $E_0(1,\mathsf{Q},0)$ denote the function $E_0(\rho,\mathsf{Q},r)$ evaluated at $\rho = 1$, $r = 0$ for the input law Q to the Ricean channel of specular component d and additive noise variance σ^2. Let $E_0^{\sigma=0}(1,\mathsf{Q},0)$ be similarly defined for the Ricean channel with the same specular component but without any additive noise. If under the law Q the input $X \in \mathbb{C}$ satisfies with probability one

$$|X| \ge x_{\min}$$

for some $x_{\min} > 0$ then

$$E_0(1,\mathsf{Q},0) \ge E_0^{\sigma=0}(1,\mathsf{Q},0) - O\left(\frac{|d|^2+1}{x_{\min}^2}\right), \quad (80)$$

where the $O(\cdot)$ term depends on d and $x_{\min}^2$ and tends to zero as $x_{\min}$ tends to infinity with d held fixed.

Proof: For any input probability distribution Q, the term $E_0(1, \mathsf{Q}, 0)$ can be expressed as

$$\begin{aligned} E_0(1, \mathsf{Q}, 0) &= -\log \int_x \int_{x'} \int_y \sqrt{w(y|x)w(y|x')}\mathrm{d}\mu(y)\mathrm{d}\mathsf{Q}(x')\mathrm{d}\mathsf{Q}(x) \\ &= -\log \int_x \int_{x'} B(x, x'; \sigma)\mathrm{d}\mathsf{Q}(x')\mathrm{d}\mathsf{Q}(x) \end{aligned} \tag{81}$$

where

$$B(x, x'; \sigma) \triangleq \int_y \sqrt{w(y|x)w(y|x')}\mathrm{d}\mu(y)$$

and where for the Ricean-fading channel with additive noise of variance σ^2

$$B(x, x'; \sigma) = \frac{2\sqrt{|x|^2+\sigma^2}\sqrt{|x'|^2+\sigma^2}}{|x'|^2+|x|^2+2\sigma^2} \cdot \exp\left(\frac{-|d|^2 \cdot |x-x'|^2}{2(|x|^2+|x'|^2+2\sigma^2)}\right). \tag{82}$$

Comparing $B(x, x'; \sigma)$ with the corresponding term in the absence of noise $B(x, x'; 0)$ we obtain

$$\begin{aligned} &B(x, x'; \sigma) \\ &\le B(x, x'; 0)\sqrt{1+\sigma^2/|x|^2}\sqrt{1+\sigma^2/|x'|^2} \\ &\quad \cdot \exp\left(|d|^2\sigma^2 \frac{|x-x'|^2}{(|x|^2+|x'|^2+2\sigma^2)(|x|^2+|x'|^2)}\right) \\ &\le B(x, x'; 0)\sqrt{1+\sigma^2/|x|^2}\sqrt{1+\sigma^2/|x'|^2} \\ &\quad \cdot \exp\left(|d|^2\sigma^2 \frac{(|x|+|x'|)^2}{(|x|^2+|x'|^2+2\sigma^2)(|x|^2+|x'|^2)}\right) \end{aligned} \tag{83}$$

where the last inequality follows by the triangle inequality. It thus follows from (81) and (83) that if under the law Q the random variable X satisfies with probability one $|X| \ge x_{\min}$ then

$$\begin{aligned} E_0(1, \mathsf{Q}, 0) \ge E_0^{\sigma=0}(1, \mathsf{Q}, 0) - \sup_{|x|,|x'| \ge x_{\min}} \Big\{ &\log\sqrt{1+\sigma^2/|x|^2} \\ &+ \log\sqrt{1+\sigma^2/|x'|^2} \\ &+ |d|^2\sigma^2 \frac{(|x|+|x'|)^2}{(|x|^2+|x'|^2+2\sigma^2)(|x|^2+|x'|^2)} \Big\} \\ = E_0^{\sigma=0}(1, \mathsf{Q}, 0) - O\left((|d|^2+1)/x_{\min}^2\right). \end{aligned}$$

□

Using this proposition with the law $\tilde{\mathsf{Q}}$ under which X is distributed according to (79) we obtain that

$$\varliminf_{\mathcal{E}\to\infty} \left\{ R_0\left(\frac{\mathcal{E}}{\sigma^2}; d, 1\right) - E_0^{\sigma=0}(1, \tilde{\mathsf{Q}}, 0) \right\} \ge 0. \tag{84}$$

Computing $E_0^{\sigma=0}(1, \tilde{\mathsf{Q}}, 0)$ from (81) and (82) we obtain

$$\begin{aligned} E_0^{\sigma=0}(1, \tilde{\mathsf{Q}}, 0) &= -\log 8 + \frac{|d|^2}{2} + 2\log\log\frac{\mathcal{E}}{\log\mathcal{E}} \\ &- \log \int_{\sqrt{\log\mathcal{E}}}^{\sqrt{\mathcal{E}}} \int_{\sqrt{\log\mathcal{E}}}^{\sqrt{\mathcal{E}}} \frac{1}{\rho^2+\rho'^2} \mathrm{I}_0\left(\frac{|d|^2\rho\rho'}{\rho^2+\rho'^2}\right) \mathrm{d}\rho\mathrm{d}\rho'. \end{aligned} \tag{85}$$

The last term on the RHS of the above is difficult to evaluate precisely. However, since the integrand is positive, the double integral can be upper-bounded by inflating the region of integration to the region

$$\{\rho, \rho' \ge 0 : 2\log\mathcal{E} \le \rho^2 + \rho'^2 \le 2\mathcal{E}\}.$$

The integral over this larger set can be now computed analytically by changing to polar coordinates to obtain

$$\begin{aligned} &\int_{\sqrt{\log\mathcal{E}}}^{\sqrt{\mathcal{E}}} \int_{\sqrt{\log\mathcal{E}}}^{\sqrt{\mathcal{E}}} \frac{1}{\rho^2+\rho'^2} \mathrm{I}_0\left(\frac{|d|^2\rho\rho'}{\rho^2+\rho'^2}\right) \mathrm{d}\rho\mathrm{d}\rho' \\ &\qquad \le \frac{\pi}{2} \cdot \mathrm{I}_0^2\left(\frac{|d|^2}{4}\right) \cdot \log\sqrt{\frac{\mathcal{E}}{\log\mathcal{E}}} \end{aligned} \tag{86}$$

where we have used the identity

$$\frac{2}{\pi}\int_0^{\frac{\pi}{2}} \mathrm{I}_0(\xi\sin\varphi)\mathrm{d}\varphi = \mathrm{I}_0^2(\xi/2), \qquad \xi \in \mathbb{R} \tag{87}$$

which follows from [20, Sec. 6.567]. Consequently, by (85) and (86)

$$\begin{aligned} &E_0^{\sigma=0}(1, \tilde{\mathsf{Q}}, 0) \\ &\ge \log\log\frac{\mathcal{E}}{\log\mathcal{E}} + \frac{|d|^2}{2} - \log(2\pi) - 2\log\mathrm{I}_0\left(\frac{|d|^2}{4}\right) \end{aligned} \tag{88}$$

so that by (84)

$$\begin{aligned} &\varliminf_{\mathcal{E}\to\infty} \left\{ R_0\left(\frac{\mathcal{E}}{\sigma^2}; d, 1\right) - \log\log\frac{\mathcal{E}}{\sigma^2} \right\} \\ &\qquad \ge \frac{|d|^2}{2} - \log(2\pi) - 2\log\mathrm{I}_0\left(\frac{|d|^2}{4}\right). \end{aligned} \tag{89}$$

B. The Cutoff Rate in the Presence of Receiver Side Information

We next consider the case where the specular component d is zero (Rayleigh fading) and where the receiver has access to some side information S that is jointly Gaussian with H. We assume that the pair (H, S) is independent of the additive noise Z and that the joint law of (H, S) and Z does not depend on the channel input $x \in \mathbb{C}$. We denote the conditional mean of H given $S = s$ by

$$\hat{d}_s \triangleq \mathsf{E}[H|S = s] \tag{90}$$

and the estimation error by

$$\epsilon^2 \triangleq \mathsf{E}\left[|H-\hat{d}_s|^2|S=s\right]. \tag{91}$$

Note that unconditionally, $\hat{d}_s$ is a zero-mean circularly symmetric Gaussian random variable of variance $1-\epsilon^2$

$$\hat{d}_s \sim \mathcal{N}_{\mathbb{C}}(0, 1-\epsilon^2). \tag{92}$$

Recall also that we only treat here the case $\epsilon^2 > 0$. Denoting the conditional density of (Y,S) corresponding to the input $x \in \mathbb{C}$ by $w(y,s|x)$, we have by the independence of the side information S and the input that

$$w(y,s|x) = f_S(s)w(y|x,s) \tag{93}$$

where f_S is the density of the side information and where $w(y|x,s)$ is the conditional law of Y given the input x and the side information s. Note that, because (H,S) are jointly Gaussian, the density $w(y|x,s)$ is the Gaussian density of mean $\hat{d}_s \cdot x$ and variance $\epsilon^2 \cdot |x|^2 + \sigma^2$. Consequently

$$\begin{aligned} E_0(1,\mathsf{Q},r) &= -\log \int_s \int_y \left(\int_x e^{r(|x|^2-\mathcal{E})} \sqrt{w(y,s|x)} \mathrm{d}\mathsf{Q}(x) \right)^2 \\ &\quad \cdot \mathrm{d}\mu(y)\mathrm{d}\mu(s) \\ &= -\log \int_s f_S(s) \\ &\quad \cdot \int_y \left(\int_x e^{r(|x|^2-\mathcal{E})} \sqrt{w(y|x,s)} \mathrm{d}\mathsf{Q}(x) \right)^2 \\ &\quad \cdot \mathrm{d}\mu(y)\mathrm{d}\mu(s) \qquad (94) \\ &= -\log \int_s f_S(s) \cdot \exp\left(-E_0(1,\mathsf{Q},r|s)\right) \mathrm{d}s \qquad (95) \end{aligned}$$

where (94) follows from (93) and where (95) follows by defining

$$\begin{aligned} &E_0(1,\mathsf{Q},r|s) \\ &\triangleq -\log \int_y \left(\int_x e^{r(|x|^2-\mathcal{E})} \sqrt{w(y|x,s)} \mathrm{d}\mathsf{Q}(x) \right)^2 \mathrm{d}\mu(y) \qquad (96) \end{aligned}$$

as the E_0 function corresponding to the channel $w(y|x,s)$, s being fixed. (This channel is a Ricean-fading channel, except that the fading is not normalized to have unit variance.)

1) Upper Bound: Using (95) and (96), we can upper-bound the cutoff rate $R_0\left(\frac{\mathcal{E}}{\sigma^2}; 0, \epsilon^2\right)$ in the presence of the side information S by (97) at the bottom of the page, where we define

$$R_0(\mathcal{E}|S=s) \triangleq \sup_{\mathsf{Q}: \mathsf{E}_\mathsf{Q}[|X|^2] \le \mathcal{E}} \sup_{r \ge 0} E_0(1,\mathsf{Q},r|s).$$

Note that $R_0(\mathcal{E}|S=s)$ is the cutoff rate corresponding to power $\mathcal{E}$ communication over the channel $w(y|x,s)$ for fixed s.[5]

It now follows from (97) that

$$\begin{aligned} &R_0\left(\frac{\mathcal{E}}{\sigma^2}; 0, \epsilon^2\right) - \log\log\frac{\mathcal{E}}{\sigma^2} \\ &\le -\log \int_s f_S(s) \cdot \exp\left(-\left(R_0(\mathcal{E}|S=s) - \log\log\frac{\mathcal{E}}{\sigma^2}\right)\right) \mathrm{d}s \end{aligned}$$

and consequently see equation (98) at the top of the following page. Here the swapping of the limit and the expectation (second inequality) is justified using Fatou's lemma and we use the result

$$\begin{aligned} &\lim_{\mathcal{E}\to\infty} \left\{ R_0(\mathcal{E}|S=s) - \log\log\frac{\mathcal{E}}{\sigma^2} \right\} \\ &\qquad = \frac{|\hat{d}_s|^2}{2\epsilon^2} - \log(2\pi) - 2\log \mathrm{I}_0\left(\frac{|\hat{d}_s|^2}{4\epsilon^2}\right) \end{aligned}$$

which follows from (61) applied to the un-normalized Ricean-fading channel whose specular component is $\hat{d}_s$ and whose granular component is of variance ϵ^2. The evaluation of the last integral is based on an identity combining [20, Sec. 6.612] and [21, Sec. 160.02]

$$\int_0^\infty e^{-\alpha x}(\mathrm{I}_0(\beta x))^2 \mathrm{d}x = \frac{2}{\pi\alpha} K\left(\frac{2\beta}{\alpha}\right), \qquad \alpha,\beta > 0$$

and the identity for the elliptic function [22, eq. (3.2.4)]

$$K(k) = \frac{2}{1+k'} K\left(\frac{1-k'}{1+k'}\right), \quad \forall\, k^2 + k'^2 = 1,\ 0 < k, k' < 1. \tag{99}$$

[5]This definition is consistent with (49) since the cost constraint on the cutoff rate is always active for the Ricean-fading channel.

$$\begin{aligned} R_0\left(\frac{\mathcal{E}}{\sigma^2}; 0, \epsilon^2\right) &\le \sup_{\mathsf{Q}:\ \mathsf{E}_\mathsf{Q}[|X|^2] \le \mathcal{E}} \sup_{r \ge 0} \left\{ -\log \int_s f_S(s) \cdot \exp\left(-E_0(1,\mathsf{Q},r|s)\right) \mathrm{d}s \right\} \\ &\le -\log \int_s f_S(s) \cdot \exp\left(-\sup_{\mathsf{Q}: \mathsf{E}_\mathsf{Q}[|X|^2] \le \mathcal{E}} \sup_{r \ge 0} E_0(1,\mathsf{Q},r|s) \right) \mathrm{d}s \\ &= -\log \int_s f_S(s) \cdot \exp\left(-R_0(\mathcal{E}|S=s)\right) \mathrm{d}s \end{aligned} \tag{97}$$

$$\begin{aligned}\overline{\lim_{\mathcal{E}\to\infty}}\left\{R_0\left(\frac{\mathcal{E}}{\sigma^2};0,\epsilon^2\right)-\log\log\frac{\mathcal{E}}{\sigma^2}\right\} &\le \overline{\lim_{\mathcal{E}\to\infty}}\left\{-\log\int_s f_S(s)\exp\left(-\left(R_0(\mathcal{E}|S=s)-\log\log\frac{\mathcal{E}}{\sigma^2}\right)\right)\mathrm{d}s\right\}\\
&= -\log\varliminf_{\mathcal{E}\to\infty}\left\{\int_s f_S(s)\cdot\exp\left(-\left(R_0(\mathcal{E}|S=s)-\log\log\frac{\mathcal{E}}{\sigma^2}\right)\right)\mathrm{d}s\right\}\\
&\le -\log\int_s f_S(s)\varliminf_{\mathcal{E}\to\infty}\exp\left(-\left(R_0(\mathcal{E}|S=s)-\log\log\frac{\mathcal{E}}{\sigma^2}\right)\right)\mathrm{d}s\\
&= -\log\int_s f_S(s)\exp\left(-\lim_{\mathcal{E}\to\infty}\left\{R_0(\mathcal{E}|S=s)-\log\log\frac{\mathcal{E}}{\sigma^2}\right\}\right)\mathrm{d}s\\
&= -\log\int_s f_S(s)\exp\left(-\frac{|\hat{d}_s|^2}{2\epsilon^2}+\log(2\pi)+2\log\mathrm{I}_0\left(\frac{|\hat{d}_s|^2}{4\epsilon^2}\right)\right)\mathrm{d}s\\
&= \log\frac{1}{\epsilon^2}-\log\mathrm{K}\left(\sqrt{1-\epsilon^4}\right)-\log 4. \qquad (98)\end{aligned}$$

2) Lower Bound: In view of (98), to establish (63) it now suffices to show

$$\varliminf_{\mathcal{E}\to\infty}\left\{R_0\left(\frac{\mathcal{E}}{\sigma^2};0,\epsilon^2\right)-\log\log\frac{\mathcal{E}}{\sigma^2}\right\} \ge \log\frac{1}{\epsilon^2}-\log\mathrm{K}\left(\sqrt{1-\epsilon^4}\right)-\log 4. \quad (100)$$

To this end, we note that by (95) and (96) evaluated at $r=0$

$$R_0\left(\frac{\mathcal{E}}{\sigma^2};0,\epsilon^2\right)\ge -\log\int_s f_S(s)\cdot\exp\left(-E_0(1,\tilde{\mathsf{Q}},0|s)\right)\mathrm{d}s \quad (101)$$

for any law $\tilde{\mathsf{Q}}$ satisfying $\mathsf{E}_{\tilde{\mathsf{Q}}}[|X|^2]\le\mathcal{E}$. We next choose, as before, $\tilde{\mathsf{Q}}$ to be a law under which X is circularly symmetric with

$$\log|X|^2\sim\mathrm{Uniform}\left(\log\log\mathcal{E},\log\mathcal{E}\right)$$

whence, by Proposition 5 and (88) applied to the Ricean channel of fading mean $\hat{d}_s$ and granular component ϵ^2 and the tightness of the lower bound

$$\varliminf_{\mathcal{E}\to\infty}\left\{E_0(1,\tilde{\mathsf{Q}},0|s)-\log\log\frac{\mathcal{E}}{\sigma^2}\right\} = \frac{|d_s|^2}{2\epsilon^2}-\log(2\pi)-2\log\mathrm{I}_0\left(\frac{|d_s|^2}{4\epsilon^2}\right) \quad (102)$$

for every s. The desired result (100) now follows from (101) and (102) using the Dominated Convergence Theorem and (92).

Appendix I
Proof of (43)

Proof: We begin with the case where the cost constraint is active. Fix some $\varrho\ge 0$ and let Q_* and r_* achieve

$$\max_{\mathsf{Q}:\mathsf{E}_\mathsf{Q}[g(X)]=\Upsilon}\max_{r\ge 0}E_0(\varrho,\mathsf{Q},r)$$

so that

$$E_0(\varrho,\mathsf{Q}_*,r_*)=\max_{\mathsf{Q}:\mathsf{E}_\mathsf{Q}[g(X)]=\Upsilon}\max_{r\ge 0}E_0(\varrho,\mathsf{Q},r). \quad (103)$$

Following [5, eq. (7.3.26)] we define

$$\alpha(y)\triangleq\sum_{x\in\mathcal{X}}\mathsf{Q}_*(x)e^{r_*(g(x)-\Upsilon)}\mathsf{W}(y|x)^{\frac{1}{1+\varrho}},\qquad y\in\mathcal{Y}. \quad (104)$$

With this definition, we have by (103) and (33)

$$\max_{\mathsf{Q}:\mathsf{E}_\mathsf{Q}[g(X)]=\Upsilon}\max_{r\ge 0}E_0(\varrho,\mathsf{Q},r)=E_0(\varrho,\mathsf{Q}_*,r_*) = -\log\sum_{y\in\mathcal{Y}}\alpha^{1+\varrho}(y). \quad (105)$$

Also, by [5, eq. (7.3.28)]

$$\sum_{y\in\mathcal{Y}}\alpha^{\varrho}(y)e^{r_*(g(x)-\Upsilon)}\mathsf{W}(y|x)^{\frac{1}{1+\varrho}}\ge\sum_{y\in\mathcal{Y}}\alpha^{1+\varrho}(y),\quad\forall\,x\in\mathcal{X}. \quad (106)$$

Consider now the distribution R_* on $\mathcal{Y}$ given by

$$\mathsf{R}_*(y)=\frac{\alpha^{1+\varrho}(y)}{\sum_{y'\in\mathcal{Y}}\alpha^{1+\varrho}(y')},\qquad y\in\mathcal{Y}. \quad (107)$$

We now have by (10) that for any distribution Q

$$E_{\mathrm{CK},0}(\varrho,\mathsf{Q}) \le -(1+\varrho)\sum_{x\in\mathcal{X}}\mathsf{Q}(x)\log\sum_{y\in\mathcal{Y}}\mathsf{W}(y|x)^{\frac{1}{1+\varrho}}\mathsf{R}_*(y)^{\frac{\varrho}{1+\varrho}} \quad (108)$$

and if $\mathsf{E}_\mathsf{Q}[g(X)]=\Upsilon$ then

$$E_{\mathrm{CK},0}(\varrho,\mathsf{Q}) \le -(1+\varrho)\sum_{x\in\mathcal{X}}\mathsf{Q}(x)\log\sum_{y\in\mathcal{Y}}e^{r_*(g(x)-\Upsilon)} \cdot\mathsf{W}(y|x)^{\frac{1}{1+\varrho}}\mathsf{R}_*(y)^{\frac{\varrho}{1+\varrho}}$$

$$\begin{aligned}&= -(1+\varrho)\sum_{x\in\mathcal{X}} \mathsf{Q}(x)\log\sum_{y\in\mathcal{Y}} e^{r_*(g(x)-\Upsilon)}\\ &\quad\cdot \mathsf{W}(y|x)^{\frac{1}{1+\varrho}}\alpha^{\varrho}(y) + \varrho\log\sum_{y\in\mathcal{Y}}\alpha^{1+\varrho}(y)\\ &\le -\log\sum_{y\in\mathcal{Y}}\alpha^{1+\varrho}(y)\\ &= \max_{\mathsf{Q}:\mathsf{E_Q}[g(X)]=\Upsilon}\max_{r\ge 0} E_0(\varrho,\mathsf{Q},r).\end{aligned}$$

Here, the first inequality follows from (108) because the condition $\mathsf{E_Q}[g(X)] = \Upsilon$ guarantees that the introduction of the exponential term $\exp\{r_*(g(x)-\Upsilon)\}$ has zero net effect; the subsequent equality by (107); the subsequent inequality by (106); and the final equality by (103). It thus follows upon taking the supremum in the above over all laws Q satisfying $\mathsf{E_Q}[g(X)] = \Upsilon$ that

$$\max_{\mathsf{Q}:\mathsf{E_Q}[g(X)]=\Upsilon} E_{\mathrm{CK},0}(\varrho,\mathsf{Q}) \le \max_{\mathsf{Q}:\mathsf{E_Q}[g(X)]=\Upsilon}\max_{r\ge 0} E_0(\varrho,\mathsf{Q},r). \tag{109}$$

On the other hand, by (42) we obtain

$$\max_{\mathsf{Q}:\mathsf{E_Q}[g(X)]=\Upsilon} E_{\mathrm{CK},0}(\varrho,\mathsf{Q}) \ge \max_{\mathsf{Q}:\mathsf{E_Q}[g(X)]=\Upsilon}\max_{r\ge 0} E_0(\varrho,\mathsf{Q},r) \tag{110}$$

which combines with (109) to prove the claim for active cost constraints.

For the case of inactive cost constraints we have

$$\begin{aligned}\max_{\mathsf{Q}:\mathsf{E_Q}[g(X)]\le\Upsilon} E_{\mathrm{CK},0}(\varrho,\mathsf{Q}) &\le \max_{\mathsf{Q}} E_{\mathrm{CK},0}(\varrho,\mathsf{Q})\\ &= \max_{\mathsf{Q}} E_{\mathrm{G},0}(\varrho,\mathsf{Q})\\ &= \max_{\mathsf{Q}:\mathsf{E_Q}[g(X)]\le\Upsilon} E^{\mathrm{m}}_{\mathrm{G},0}(\varrho,\mathsf{Q}).\end{aligned}$$

Here the first inequality follows by relaxing the constraint; the subsequent equality by (13); and the final equality by (39). This combines with (42) to conclude the proof. □

Appendix II
Derivation of (75)

To derive (75), we begin by noting that for $\rho \ge 1$ the integrand can be lower-bounded by its value when $\alpha = 0$ because

$$\left(\rho^2+\delta\right)^{\frac{\alpha-1}{2}} \ge \left(\rho^2+\delta\right)^{-\frac{1}{2}}, \qquad \alpha,\delta\ge 0,\ \ \rho\ge 1.$$

In the region $0\le\rho\le 1$ we can use the inequality

$$\left(\rho^2+\delta\right)^{\frac{\alpha-1}{2}} \ge \delta^{\frac{\alpha}{2}}\left(\rho^2+\delta\right)^{-\frac{1}{2}}, \qquad \alpha,\delta> 0,\ \ 0\le\rho\le 1.$$

Combining the above two bounds we obtain that throughout the region of integration

$$\left(\rho^2+\delta\right)^{\frac{\alpha-1}{2}} \ge \delta^{\frac{\alpha}{2}}\left(\rho^2+\delta\right)^{-\frac{1}{2}}, \qquad \alpha>0,\ \ 0<\delta<1,\ \ 0\le\rho<\infty$$

and hence,

$$\ell(x;\alpha,\beta,\delta) \ge \delta^{\frac{\alpha}{2}}\cdot\ell(x;0,\beta,\delta). \tag{111}$$

We next relate $\ell(x;0,\beta,\delta)$ to $\ell(x;0,\beta,0)$. To that end, denote the integrand in $\ell(x;0,\beta,\delta)$ by

$$\eta(\rho;x,\beta,\delta,d) = \exp\left(-\rho^2\frac{\beta+|x|^2+\sigma^2}{2\beta(|x|^2+\sigma^2)}\right)\cdot\sqrt{\frac{\rho^2}{\rho^2+\delta}}\mathrm{I}_0\left(\frac{|d|\cdot|x|\cdot\rho}{|x|^2+\sigma^2}\right) \tag{112}$$

and express the integral as

$$\ell(x;0,\beta,\delta) = \int_0^{\sqrt{m_1\delta}} + \int_{\sqrt{m_1\delta}}^{\infty}\eta(\rho;x,\beta,\delta,d)\mathrm{d}\rho.$$

In the region $\rho\ge\sqrt{m_1\delta}$, we have

$$\sqrt{\frac{\rho^2}{\rho^2+\delta}} \ge \sqrt{\frac{m_1}{m_1+1}}$$

and hence,

$$\int_{\sqrt{m_1\delta}}^{\infty}\eta(\rho;x,\beta,\delta,d)\mathrm{d}\rho \ge \sqrt{\frac{m_1}{m_1+1}}\int_{\sqrt{m_1\delta}}^{\infty}\eta(\rho;x,\beta,0,d)\mathrm{d}\rho. \tag{113}$$

We next show that when $\sqrt{m_1\delta}$ is small, the integral over the interval $[0,\sqrt{m_1\delta}]$ is also small. To that end, we upper-bound $\eta(\rho;x,\beta,\delta,d)$ in (112) by upper-bounding the exponential by 1 (its argument is negative); by upper-bounding $\rho^2/(\rho^2+\delta)$ by 1; and by using the monotonicity of $\mathrm{I}_0(\cdot)$ and the inequality

$$\frac{|x|}{|x|^2+\sigma^2} \le \frac{1}{2\sigma}, \qquad x\in\mathbb{C}$$

to obtain

$$0 \le \eta(\rho;x,\beta,0,d) \le \mathrm{I}_0\left(\frac{|d|\cdot\rho}{2\sigma}\right).$$

Consequently, we have for any $m_1>0$

$$0 \le \int_0^{\sqrt{m_1\delta}}\eta(\rho;x,\beta,0,d)\mathrm{d}\rho \le \sqrt{m_1\delta}\cdot\mathrm{I}_0\left(\frac{|d|\sqrt{m_1\delta}}{2\sigma}\right). \tag{114}$$

On the other hand, a straightforward calculation demonstrates that

$$\begin{aligned}\int_0^{\infty}\eta(\rho;x,\beta,0,d)\mathrm{d}\rho &\ge \int_0^{\infty}\eta(\rho;x,\beta,0,0)\mathrm{d}\rho\\ &= \sqrt{\frac{\pi}{2}}\cdot\sqrt{\frac{\beta(|x|^2+\sigma^2)}{\beta+|x|^2+\sigma^2}}\\ &\ge \sqrt{\frac{\pi\beta\sigma^2}{2(\beta+\sigma^2)}}\end{aligned} \tag{115}$$

where the first inequality follows from the monotonicity of $\mathrm{I}_0(\cdot)$ and the final inequality follows from simple algebra. We thus conclude that

$$\begin{aligned}\ell(x;0,\beta,\delta) &= \int_0^\infty \eta(\rho;x,\beta,\delta,d)\mathrm{d}\rho \\ &\geq \int_{\sqrt{m_1\delta}}^\infty \eta(\rho;x,\beta,\delta,d)\mathrm{d}\rho \\ &\geq \sqrt{\frac{m_1}{m_1+1}}\int_{\sqrt{m_1\delta}}^\infty \eta(\rho;x,\beta,0,d)\mathrm{d}\rho \\ &= \sqrt{\frac{m_1}{m_1+1}} \\ &\quad \cdot \left(\int_0^\infty - \int_0^{\sqrt{m_1\delta}} \eta(\rho;x,\beta,0,d)\mathrm{d}\rho\right) \\ &= \sqrt{\frac{m_1}{m_1+1}}\left(1-\frac{\int_0^{\sqrt{m_1\delta}}\eta(\rho;x,\beta,0,d)\mathrm{d}\rho}{\int_0^\infty \eta(\rho;x,\beta,0,d)\mathrm{d}\rho}\right) \\ &\quad \cdot \int_0^\infty \eta(\rho;x,\beta,0,d)\mathrm{d}\rho \\ &\geq \sqrt{\frac{m_1}{m_1+1}}\left(1-\frac{\sqrt{m_1\delta}\cdot \mathrm{I}_0\left(\frac{|d|\sqrt{m_1\delta}}{2\sigma}\right)}{\sqrt{\frac{\pi\beta\sigma^2}{2(\beta+\sigma^2)}}}\right) \\ &\quad \cdot \ell(x;0,\beta,0) \end{aligned} \tag{116}$$

where the first inequality follows from the nonnegativity of the integrand; the subsequent inequality from (113); and the final inequality from (114) and (115). The desired bound (75) now follows from (116) and (111).

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers for their careful reading of the manuscript and for their helpful comments.

REFERENCES

[1] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, vol. 20, no. 4, pp. 405–417, Jul. 1974.
[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
[3] S. Arimoto, "Computation of random coding exponent functions," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 6, pp. 665–671, Nov. 1976.
[4] M. Chiang, "Geometric programming for communication systems," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 1/2, pp. 1–153, 2005.
[5] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[6] G. S. Poltyrev, "Random coding bounds for discrete memoryless channels," *Probl. Inf. Transm.*, vol. 18, no. 1, pp. 12–26, Jan./Mar. 1982.
[7] R. G. Gallager, "The random coding bound is tight for the average code," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 2, pp. 244–246, Mar. 1973.
[8] I. Csiszár, "The method of types," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2505–2523, Oct. 1998.
[9] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge Univ. Press, 2004.
[10] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. IT–11, no. 1, pp. 3–18, Jan. 1965.
[11] T. H. E. Ericson, "A Gaussian channel with slow fading," *IEEE Trans. Inf. Theory*, vol. IT–16, no. 3, pp. 353–355, May 1970.
[12] W. Ahmed and P. McLane, "Random coding error exponents for two-dimensional flat fading channels with complete channel state information," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 1338–1346, May 1999.
[13] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
[14] I. Abou-Faycal, M. Trott, and S. Shamai (Shitz), "The capacity of discrete time Rayleigh fading channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290–1301, May 2001.
[15] G. Taricco and M. Elia, "Capacity of fading channels with no side information," *Electron. Lett.*, vol. 33, no. 16, pp. 1368–1370, Jul. 1997.
[16] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.
[17] M. C. Gursoy, H. V. Poor, and S. Verdú, "The noncoherent Rician fading channel—Part I: Structure of the capacity achieving input," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2193–2206, Sep. 2005.
[18] ——, "The noncoherent Rician fading channel—Part II: Spectral efficiency in the low power regime," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2207–2221, Sep. 2005.
[19] B. C. Carlson and J. L. Gustafson, "Asymptotic expansions of the first elliptic integral," *SIAM J. Math. Anal.*, pp. 1072–1092, 1985.
[20] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 5th ed. San Diego, CA: Academic, 1994.
[21] P. S. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Scientists*, 2nd ed. New York: Springer, 1971.
[22] G. E. Andrews, R. Askey, and R. Ranjan, "Special Functions," in *Encyclopedia of Mathematics and Its Applications*. Cambridge, U.K.: Cambridge Univ. Press, 1999.