© Hirzel-Verlag Stuttgart

AEÜ, Vol. 47 (1993), No. 4

Information Rates and Error Exponents of Compound Channels with Application to Antipodal Signaling in a Fading Environment

Gideon Kaplan and Shlomo Shamai (Shitz)

Information Rates and Error Exponents of Compound Channels with Application to Antipodal Signaling in a Fading Environment

A compound channel is characterized by a set of permissible transition probabilities (which depend on the state of the channel). We focus on a class of compound channels for which the state sequence has a stochastic characterization and with mismatched decoding. Three levels of receiver side information on the state sequence are addressed, namely known state, known statistics, and mismatch. First, we develop some new relations of error exponents and achievable rates in this general context, and for completeness, review relevant known results. We then apply some of these general results to the specific case of binary PSK modulation operating in a fading channel with ideal interleaving under the abovementioned different levels of side information. We address two variants of the Rician model, as well as the Nakagami-m model, and compare their similarities and differences.

Transinformation und Fehlerexponent zeitvarianter Kanäle und ihre Anwendung auf 2–PSK bei Fading

Zusammengesetzte (zeitvariante) Kanäle werden durch eine Menge zulässiger Übergangswahrscheinlichkeiten beschrieben, die vom momentanen Zustand des Kanals abhängen. Wir konzentrieren uns auf Kanäle mit stochastisch beschreibbaren Zustandsabfolgen und Empfänger mit nicht angepaßter Decodierung. Dabei werden drei Fälle unterschiedlicher Kanalinformationenen, die dem Empfänger zur Verfügung stehen, untersucht: bekannter Zustand des Kanals, bekannte Statistik und völlige Fehlanpassung an den Kanal. Zunächst werden, neben einer Zusammenstellung wichtiger bekannter Ergebnisse, einige neue Beziehungen für den Fehlerexponenten und ereichbare Raten allgemein abgeleitet. Danach werden diese Ergebnisse speziell auf die Übertragung mit binärer PSK-Modulation über einen Fading-Kanal angewandt, wobei die vorstehend beschriebenen Fälle verfügbarer Kanalinformationen zugrundegelegt werden. Dabei werden zwei Varianten des Rice-Modells sowie das Nagakami-m-Modell untersucht und hinsichtlich ihrer Gemeinsamkeiten und Unterschiede verglichen.

1. Introduction

Compound channels [1] may be defined as channels for which the transition probability is actually a set of transition probabilities. The set is composed of $p(y|x,z)$, where y designates the channel output, x designates the channel input, and $z \in Z$ is the state of the channel. Here Z designates the state ensemble (finite or infinite). We assume throughout that z and x are independent, and that z has a stochastic description given by $f(Z)$ (albeit not necessarily known to the user). This well-known model is used to describe many practical channels, such as fading channels [2]–[6], bursty channels [7]–[8], and some particular jammed channels [4]. This class of channels is also termed time-varying channels (e.g. [2]). In many practical cases, such channels possess inherent memory; a simple example is the slowly fading channel. The inherent memory allows the receiver to extract (or estimate) the channel state. Therefore, often the coded performance of such channels is evaluated with and without side information (SI) on the state of the channel (e.g. [2]–[3], [5]–[8]). In this work, we first focus on the random-coding error exponent and the achievable information rates of such channels under different levels of side information. New properties of these quantities are proved, and few known properties are reviewed for completeness. We then apply some of the general tools to investigate a hypothetical practical case: antipodal signaling under several closely-connected flat fading models. The analysis is applicable, for example, to land-mobile satellite channels, and to other schemes in which coherent detection is possible (with a pilot tone or other means). The analysis of three fading models under a unified framework is novel and sheds light on their similarities and differences. The relative performance under various SI level and E/N_o conditions is presented.

Three (principal) grades of SI can be defined:

I) **Known State** The receiver supplies the decoder with the instantaneous channel state, thus the decoder is assumed to possess the exact transition probability (per symbol). This is commonly denoted as a decoder with Channel State Information (CSI), e.g. [6], [7]. Hagenauer [3] and Dorsch [2] investigated coded bit error rate with hard or soft quantized SI for a Rayleigh fading channel, and showed that even 1 bit SI (i.e. "good state" or "bad state") achieves a significant improvement over no use of SI. In Section 3.3, non ideal state estimation with errors is investigated for a fading channel. Imperfect state estimation for a jammed MFSK–FH system is addressed in [14].

II) **Known Statistics** The receiver has only the ex-

Received November 7, 1991.

Department of Electrical Engineering, Israel Institute of Technology, Technion City, Haifa 32000, Israel.

act statistical characterization of the channel (including the statistical characterization of the state) but not the realizations of the channel state. This could be the case, for example, for a phase shift keying (PSK) modulation subject to phase noise with known statistical description (e.g. [15]). Note that this case sometimes appears under the term "no side information" analysis.

III) **Mismatch** The receiver has only an approximate channel statistics and therefore employs a mismatched (non adapted) metric. For example, a receiver with no fading information, may use the optimum metric of a non-faded AWGN channel, in an actual fading environment. In this case we denote the "assumed" channel probabilities by $q(y|x_k)$ — see, e.g., [4], [5], [13].

The channel state is assumed constant over a single symbol duration throughout this work. If one assumes further that ideal interleaving-deinterleaving is employed, then z can be assumed to be independent for each symbol. This will be denoted as the memoryless case. As noted by several researchers, (e.g. [3], [7], [8]) the interleaving process by itself *does not* destroy the inherent memory in the system. If z could be (perfectly) extracted, then there is no loss in interleaving-deinterleaving. It is apparent that the advantage of employing interleaving-deinterleaving is that random error correction codes can be used efficiently for channels with memory. We will focus on the memoryless case but also present a few results for the general case.

In Section 2, we review several known properties of the random coding error exponent and the cut-off rate, under the various levels of side information, and derive some novel results. In particular we focus on relations between the error exponents for the three distinct cases of SI. The data processing theorem for the random coding error exponent is proved. The generalized mutual information under mismatched conditions is defined in parallel to the matched case and compared to the mutual information of a mismatched Discrete Memoryless Channel (DMC) of Hui [16]. The results in this section thus enhance the tools for performance evaluation of such compound channels.

In Section 3, some of these ideas are applied to several useful models of a slowly fading channel with binary PSK modulation, coherent detection, and ideal interleaving. Emphasis is placed on the cut-off rate, conjectured to be a practical measure for achievable reliable information rates (e.g. [9]). The mutual information is also examined. The fade models include the familiar Rician channel with two variants (differentiated by the assumption on the signal that the coherent receiver can perfectly track in phase, i.e. either the composite received signal, or only the direct-path received energy), and also the Nakagami-m fading model. Comparisons with related results are included. In Section 3.4, we apply the power-moment bound theory to the Bhattacharyya distance, for a case where the partial knowledge of the fade includes only its dynamic range and some statistical moments. Concluding remarks, presented in Section 4, terminate the paper.

2. Random Coding Exponent, Cut-Off Rate and Generalized Mutual Information

2.1 Error Exponents – Memoryless Case

To derive the standard random coding upper bound on the average error probability in the general possibly mismatched case, we follow Gallager [10], Theorem 5.6.1. The receiver assumes a sequence transition probability denoted by $q(\underline{Y}|\underline{X}_m)$, while the actual one is $p(\underline{Y}|\underline{X}_m)$, thus the receiver sample space is divided into decision regions Λ_m specified by

$$\bar{\Lambda}_m = \{\underline{Y} : q(\underline{Y}|\underline{X}_{m'}) \geq q(\underline{Y}|\underline{X}_m),\ \text{some } m' \neq m\} \tag{1}$$

where $\bar{\Lambda}_m$ is the complement of Λ_m. Throughout, we use $\underline{X}, \underline{Y}$ or $\underline{Z}$ to denote codewords or sequences. Continuing similarly to [10], one upper bounds the ensemble-average error probability over all the possible signal sets with M codewords (with length N) randomly chosen with $Q_N(\underline{X})$, to get

$$\overline{P}_e \leq (M-1)^\rho \sum_Y \cdot \left[\sum_X Q_N(\underline{X})\, p(\underline{Y}|\underline{X}) q(\underline{Y}|\underline{X})^{-s\rho}\right] \cdot \left[\sum_X Q_N(\underline{X}) q(\underline{Y}|\underline{X})^s\right]^\rho \tag{2}$$

where $\rho \in [0,1]$, $s \geq 0$. For a compound channel condition the r.h.s. on the state sequence $\underline{Z}$ and then average over Z. For the *memoryless case* we assume the following,

$$\begin{aligned} &(i)\ \ p(\underline{Y}|\underline{X},\underline{Z}) = \textstyle\prod_{i=1}^N p(y_i|x_i,z_i) \\ &(ii)\ \ f_N(\underline{Z}) = \textstyle\prod_{i=1}^N f_z(z_i) \end{aligned} \tag{3}$$

The last assumption is applicable, for example, to coding schemes for which (long) interleaving-deinterleaving is employed.

For case I (C.S.I) – consider the channel output ensemble as $Y \times Z$ [1], where $\times$ stands for the Cartesian product. The random coding upper bound on the average error probability follows [1]

$$\overline{P}_e \leq \exp\{-N[E_{o_I}(\rho, Q) - \rho R]\} \tag{4}$$

where

$$E_{o_I}(\rho, Q) = -\ln E_z \sum_y \left[\sum_x Q(x) p(y|x,z)^{\frac{1}{1+\rho}}\right]^{1+\rho} \tag{5}$$

For case II, where $f_z(z)$ is known, the decoder uses

$$q(y|x) = \sum_z f_z(z) p(y|x,z) \tag{6}$$

This case is equivalent to a memoryless channel with cross-over probability of $E_z p(y|x, z)$ [1]. The random coding error exponent assumes the form

$$E_{o_{II}}(\rho, Q) = \qquad (7)$$
$$-\ln \sum_y \left\{ \sum_x Q(x)[E_z p(y|x, z)]^{\frac{1}{1+\rho}} \right\}^{1+\rho}$$

Proposition 1. Under the above model,

$$E_{o_I}(\rho, Q) \geq E_{o_{II}}(\rho, Q) \qquad (8)$$

Proof: We examine the argument of the $-\ln\{\ \}$ in $E_{o_{II}}(\rho, Q)$ expression, and use a variant of Minkowski's inequality (on the inner summation on x), given by [10, ch. 5]

$$\left(\sum_j a_j \left[\sum_k b_{jk} \right]^{\frac{1}{r}} \right)^r \geq \sum_k \left[\sum_j a_j b_{jk}^{\frac{1}{r}} \right]^r, \quad r \geq 1$$

where a_j, b_{jk} are nonnegative numbers, and $\sum a_j = 1$. Therefore,

$$\sum_{y_i} \left\{ \sum_{x_j} Q(x_j) \left[\sum_{z_k} f(z_k) p(y_i|x_i, z_k \right]^{\frac{1}{1+\rho}} \right\}^{1+\rho}$$
$$\geq \sum_{y_i} \sum_{z_k} \left\{ \sum_{x_j} Q(x_j) \left[p(y_i|x_j, z_k) f(z_k) \right]^{\frac{1}{1+\rho}} \right\}^{1+\rho}$$
$$= E_z \sum_{y_i} \left[\sum_{x_j} Q(x_j) p(y_i|x_j, z)^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

(with equality iff $p(z_k)p(y_i|x_j, z_k)$ is independent of k). □

2.2 Error Exponent Relations – The General Channel

Here we turn to the case where memory is present.

For case I, with an input $\underline{X} \in X$, and output $\underline{Y}$, $\underline{Z} \in Y \times Z$, one gets

$$F_I \triangleq e^{-E_{oI}(\rho, Q_N)} =$$
$$\sum_{\underline{Z}} f_z(\underline{Z}) \sum_{\underline{Y}} \left[\sum_{\underline{X}} Q_N(\underline{X}) p(\underline{Y}|\underline{X}, \underline{Z})^{\frac{1}{1+\rho}} \right]^{1+\rho} \qquad (9)$$

For case II, where $f_z(\underline{Z})$ is known, it follows that

$$F_{II} \triangleq e^{-E_{oII}(\rho, Q_N)} = \sum_{\underline{Y}} \cdot$$
$$\cdot \left\{ \sum_{\underline{X}} Q_N(\underline{X}) \left[\sum_{\underline{Z}} f_z(\underline{Z}) p(\underline{Y}|\underline{X}, \underline{Z}) \right]^{\frac{1}{1+\rho}} \right\}^{1+\rho} \qquad (10)$$

Again, by using the Minkowski's inequality on the righthand side of (9), the relation $F_{II} \geq F_I$ is proved.

A more general property of the function $E_o(\rho, Q)$ is presented in the following proposition.

Proposition 2. $E_o(\rho, Q)$ satisfies the Data-Processing theorem (which is stated for the mutual information in [10]).

Proof: Let

$$G(\underline{\alpha}) = -\left(\sum_j Q(X_j)\, \alpha_j^{\frac{1}{1+\rho}} \right)^{1+\rho} \qquad (11)$$

where $\underline{\alpha} = \{\alpha_j\}$. Then G is a convex U functional of $\underline{\alpha}$. Define the following nonnegative measures (on finite spaces)

$$\underline{\mu} = \{\mu_j\} = \{p(Y|X = X_j)\}$$
$$P = P(Y) \qquad (12)$$

And note that

$$E_p\, G\left(\frac{\mu}{P}\right) = \qquad (13)$$
$$= -\sum_Y P(Y) \left(\sum_j Q(X_j) \left[\frac{P(Y|X_j)}{P(Y)} \right]^{\frac{1}{1+\rho}} \right)^{1+\rho}$$
$$= -\sum_Y \left(\sum_j Q(X_j)\, P(Y|X_j)^{\frac{1}{1+\rho}} \right)^{1+\rho} = -F(\rho, Q)$$

The functional $E_p G\left(\frac{\mu}{P}\right)$ belongs to a class of functions denoted as "generalized entropy" in [29]. Such functions (in nonnegative measures) are shown in [29] to satisfy the data processing theorem. Thus our claim is proved. (Another proof of this proposition can be shown via straightforward use of the Minkowski inequality). □

It readily follows that the relation $F_{II} \geq F_I$ is a special case of proposition 2.

2.3 Bhattacharyya Distances – Binary Inputs

The generalized cut-off rate [4] is a well-known useful tool to assess the practically achievable information rates for a mismatched coded system. It is expressed, in the memoryless case by [4]

$$\tilde{R}_o = -\ln \sum_x \sum_{\hat{x}} Q(x) \cdot Q(\hat{x}) \cdot D \quad \text{[nats/ch. use]} \qquad (13a)$$

where the Bhattacharyya distance (between the x, $\hat{x}$ symbols) is given by

$$D = \min_{0 \leq \lambda} E_{y,z} \left\{ e^{\lambda[m(\hat{x},y|z) - m(x,y|z)]} \right\} \Big|_x \qquad (13b)$$

where $m(\cdot)$ denotes the decoder metric. Denote the cumulative distribution of z by $F_z(z)$. In the binary-input case, the relation

$$D_I = \int_y \int_z \sqrt{p(y|\text{`0'}, z)p(y|\text{`1'}, z)}\, dF_z\, dy \le$$

$$\int_y \sqrt{\int_z p(y|\text{`0'}, z) dF_z} \sqrt{\int_z p(y|\text{`1'}, z) dF_z} = D_{II}$$

is evident. This relation which follows also by (7) letting $\rho = 1$, again proves the obvious fact that the knowledge of the channel state information yields higher information rates than the case where knowledge of solely the statistics of the state is available.

For the mismatched case, namely no knowledge on the state selection, the decoder uses $q(y|x_k, z) = q(y|x_k)$ (for the binary case we denote $q(y|\text{`0'}) = q_o(y)$, $q(y|\text{`1'}) = q_1(y)$).

Proposition 3. D for case (II) is less or equal to D for case (III).

Here we rely on the method of [13]. Letting $p_0(y) = E_z\, p(y|\text{`0'}, z)$, and $p_1(y) = E_z\, p(y|\text{`1'}, z)$, yields the following expression for D_{II},

$$D_{II} = \int_y p_o(y) \sqrt{\frac{p_1(y)}{p_o(y)}}\, dy\,.$$

Assuming that the statistics used by the decoder are symmetric (in the sense that $q_o(y) = q_1(-y)$), results in

$$D_{III} = \tag{14}$$

$$= \int_y \left[p_o(y) \left(\frac{q_o(-y)}{q_o(y)} \right)^{\lambda} + p_o(-y) \left(\frac{q_o(y)}{q_o(-y)} \right)^{\lambda} \right] dy\,.$$

Now differentiating with respect to $\alpha = \frac{q_o(-y)}{q_o(y)}$ ([13]), and equating the result to zero, shows that the minimizing α is $\left(\frac{p_o(-y)}{p_o(y)}\right)^{-2\lambda}$, and it readily follows that

$$D_{III} \ge \int_y p_o(y) \sqrt{\frac{p_1(y)}{p_o(y)}}\, dy = D_{II}\,, \tag{15}$$

i.e., the third case gives the worst cut-off rate (as expected).

2.4 Mutual Information

One notes that the mutual information between the channel output and input for case I, is given by

$$I\,(\underline{X}\,;\, \underline{Y},\, \underline{Z}) = E_Z\,[I\,(\underline{X}\,;\, \underline{Y}|\underline{Z} = Z)]\,. \tag{16}$$

For an imperfect state estimation (designated by $\hat{Z}$, instead of Z) the following relations hold:

$$I(\underline{Y}\,;\, \underline{X}) \le I(\underline{Y}, \hat{\underline{Z}}\,;\, \underline{X}) \le I(\underline{Y}, \underline{Z}\,;\, \underline{X}) \tag{17}$$

These inequalities are evident by the Data-Processing theorem [10] and generalize [31, Theorem 1]. Similar relations hold for the $E_o(\cdot)$ functional (as can be proved by proposition 2).

The generalized average mutual information (for the mismatched case) is interpreted here as a property of the generalized error exponent, in an analogous way to the standard error exponent. For the general mismatched memoryless case, the average error probability for a code with M messages of length N symbols is upper bounded by (from (2))

$$\overline{P}_e \le M^\rho\, [\xi]^N \tag{18}$$

where

$$\xi \triangleq e^{-E_o(\rho,\, s)} =$$

$$= \sum_{y_n} \left\{ \sum_{x_n} Q(x_n) p(y_n|x_n) q(y_n|x_n)^{-s\rho} \right\} \cdot$$

$$\cdot \left\{ \sum_{x_n} Q(x_n) q(y_n|x_n)^s \right\}^{\rho}.$$

$E_o(\rho, s)$ stands for the generalized (random coding) error exponent, satisfying:

(a) $E_o(\rho = 0, s) = 0$

(b) $\max_{s \ge 0} E_o(\rho = 1, s)$ is the well-known generalized cut-off rate, which we assume is positive (more on this in Proposition 4).

Define the generalized mutual information between Y and X in parallel to the matched case [10] by

$$I_s(X\,;\, Y(p,\, q)) \triangleq \left. \frac{\partial E_o(\rho,\, s)}{\partial \rho} \right|_{\rho=0} = - \left. \frac{\partial \xi}{\partial \rho} \right|_{\rho=0} \tag{19}$$

and $I\,(X\,;\, Y(p,q)) = \max_{s \ge 0} I_s\,(X\,;\, Y(p,q))$, which results in

$$I(X\,;\, Y(p,\, q)) =$$

$$\max_{s \ge 0} \sum_{y_n} \sum_{x_n} Q(x_n) p(y_n|x_n) \cdot \tag{20}$$

$$\ln \frac{q(y_n|x_n)^s}{\sum_{x_n} Q(x_n) q(y_n \mid x_n)^s} \;\text{[nats/ch. use]}$$

In [11], a different approach led to a similar expression but without the parameter s (namely $s \equiv 1$). The above expression leads to better results and is nonnegative, in contrast to the result of [11]. It is readily shown that $I(X; Y(p,q)) \le I(X; Y(p,p))$. The above definition is easily extended to the general not necessarily memoryless case.

Strict positivity of $I(X\,;\, Y(p,\, q))$ is guaranteed by the following proposition, regardless of the choice of $Q(x)$.

Proposition 4. A necessary and sufficient condition for the generalized mutual information, $I(X\,;\, Y(p,q))$ to be positive, is that

$$\sum_{Y} p(\underline{Y} \mid \underline{X}_i) \ln \left[\frac{q(\underline{Y}|\underline{X}_i)}{q(\underline{Y}|\underline{X}_j)} \right] > 0, \quad \begin{array}{l} \forall i = 1 \cdots M \\ \forall j \ne i \end{array} \tag{21}$$

where $\underline{X}_i$, $\underline{X}_j$ are two codewords.

Proof : One has

$$(i) \quad I_s(X\,;\,Y(p,q))|_{s=0} = 0$$

$$(ii) \quad \left.\frac{\partial I_s(X\,;\,Y(p,q))}{\partial s}\right|_{s=0} =$$

$$\sum_X \sum_{X'} Q(\underline{X})Q(\underline{X}') \sum_Y p(\underline{Y}|\underline{X}) \ln \frac{q(\underline{Y}|\underline{X})}{q(\underline{Y}|\underline{X}')}.$$

It is therefore evident that if condition (21) is fulfilled, then the generalized mutual information is positive for some $s \in [0, s_o]$. Moreover, assuming codeword $\underline{X}_1$ is transmitted, we have

$$I(Y\,;\,X_1(p,q))) = -\sum_Y p(\underline{Y}|\underline{X}_1) \ln \sum_{X'} Q(\underline{X}') \frac{q(\underline{Y}|\underline{X}')^s}{q(\underline{Y}|\underline{X}_1)^s} \le s \sum_{X'} Q(\underline{X}') \sum_Y p(\underline{Y}|\underline{X}_1) \ln \frac{q(\underline{Y}|\underline{X}_1)}{q(\underline{Y}|\underline{X}')}$$

where Jensen's inequality has been incorporated. If condition (21) is *not* fulfilled for two specific codewords, there may be an input probability function which makes I negative or zero. Thus, to ensure that I is positive regardless of the input distribution, condition (21) is indeed necessary. This is also the necessary and sufficient condition for positivity of the generalized cut-off rate, as has been proved in [12]. Q.E.D.

Hui [16] addresses the mutual information for a mismatched DMC. By applying the method of types [17], it was proved [16] that reliable communication is possible up to $I'(X;Y)$, where

$$I'(X;Y) = H(X) + H(Y) - H'(X,Y)\,. \tag{22a}$$

Here, $H(\cdot)$ denotes the entropy functional, and

$$H'(X,Y) = \max_{\{f_{xy}\}} \sum_x \sum_y -f_{xy} \cdot \ln f_{xy}$$

where the maximization is performed over all joint density functions $f_{xy} \ge 0$, satisfying

$$\sum_x f_{xy}(x,y) = p(y) = \sum_x p(y|x)Q(x) \quad \text{for all } y$$
$$\sum_y f_{xy}(x,y) = Q(x) \quad \text{for all } x$$
$$\sum_x \sum_y f_{xy}(x,y) \ln q(y|x) \ge \sum_x \sum_y p_{xy}(x,y) \ln q(y|x) \tag{22b}$$

($p_{xy}(x,y)$ is the actual joint density of the DMC).

Proposition 5. The following relations hold for a (mismatched) DMC:

$$\tilde{R}_o \le I(X;Y(p,q)) \le I'(X;Y) \tag{23}$$

Proof : The l.h.s. inequality is directly proved by applying Jensen's inequality to the expression of $\tilde{R}_o$, eq. (13a). The r.h.s. inequality is proved in Appendix A.

The generalized mutual information and $I'(X;Y)$ are indeed additional tools (to the well-known R_o) for the evaluation of the achievable rates of mismatched coded systems. Proposition 5 indicates that $I'(X;Y)$ is preferable; however, its calculation is in general more complex.

3. Antipodal Signaling over Fading Channels

3.1 Slow-Fading Rician Channel

We investigate antipodal (BPSK) modulation over the AWGN channel with Rician fading. The (nonselective) fading is slow as compared to the symbol rate, and a sufficiently long interleaver-deinterleaver is used to render the channel memoryless. The interleaving prevents the decoder from being "hit" by a long burst, and also prevents the need to design different codes to different fade rates. We first review the well-known channel model [5], [18].

The received signal is composed of the Specular (or direct) component plus a Diffuse (or scattered) component. The underlying assumption is that the receiver can perfectly track the phase of the *composite* signal (e.g. [18]): this is an interesting problem beyond the scope of our treatment. We therefore model each received symbol by [5]

$$y_j = \rho_j x_j + n_j \tag{24}$$

where x_j is the j-th transmitted symbol (either '-1' or '1'), n_j stands for the j-th additive white noise sample, and ρ_j designates the instantaneous value of the fade for that symbol (not to be confused with the parameter of the error exponent in the previous section).

For the Rician channel, the probability density of ρ is given by [5]

$$P(\rho) = \begin{cases} 2\rho(1+k)\exp[-k-\rho^2(1+k)] \\ \quad I_0[2\rho\sqrt{k(1+k)}] & , \rho \ge 0 \\ = 0 & , \rho \le 0 \end{cases}, \tag{25}$$

where k is the ratio of specular energy (denoted E_s) to the diffuse energy (denoted E_r), and where $I_o(\cdot)$ is the modified Bessel function of order zero. $E = E_s + E_r$ denotes the symbol energy. The special case $k = 0$ gives a Rayleigh channel, while $k \to \infty$ corresponds to the non-faded AWGN case, see also [18]. The optimal decoding rule weights the "ordinary" inner product $(y_j x_j)$ by the known value of ρ_j [19].

The Chernoff parameter for case I (CSI) is given by

$$D_I = E_\rho \left\{ \int_{-\infty}^{\infty} \left[\frac{1}{\pi N_o} e^{-(y+\rho\sqrt{E})^2/N_o} e^{-(y-\rho\sqrt{E})^2/N_o} \right]^{1/2} dy \right\} = E_\rho \left\{ e^{-\rho^2 E/N_o} \right\} \tag{26}$$

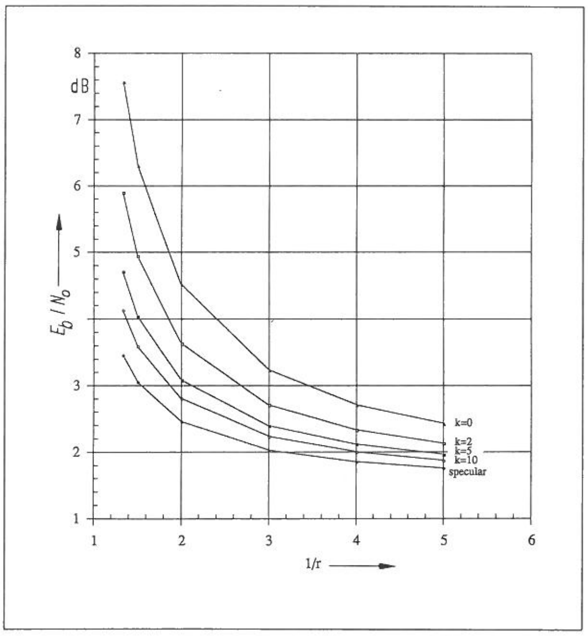

Fig. 1. E_b/N_o required to achieve $r = R_o$ (bits/symbol), Rician channel with state information.

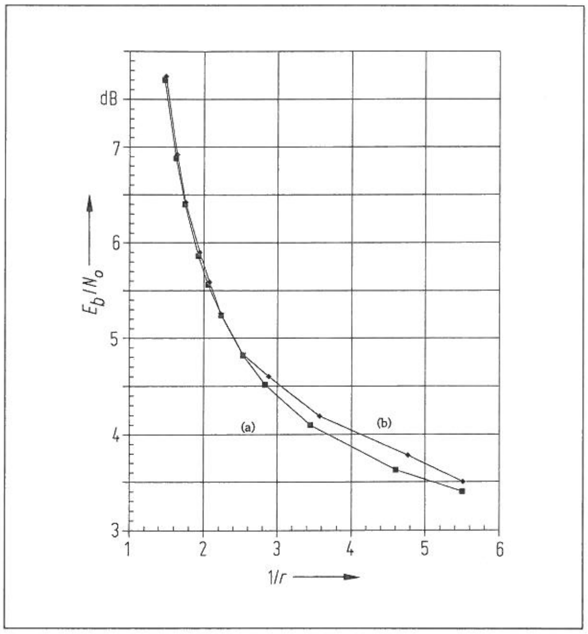

Fig. 2. E_b/N_o required to achieve $r = R_o$ (bits/symbol), Rayleigh channel (a) with statistics information, (b) no side information.

This integral expression (using (25)) is reduced ([20], eq. 3 in 6.614) to

$$D_I = (1+k)e^{-k}\frac{\exp\left(\frac{1}{2}\frac{k(1+k)}{1+k+E/N_o}\right)}{\sqrt{k(1+k)(1+k+E/N_o)}} M_{-1/2;0}\left(\frac{k(1+k)}{1+k+E/N_o}\right),$$

and then using ([20], eq. 2 in 9.22) for the Whittaker function M, results in

$$D_I = \frac{1}{1+E_r/N_o}\exp\left[-\frac{E_s/N_o}{1+E_r/N_o}\right] \qquad (27)$$

in agreement with the result (given without proof) of [19]. Fig. 1 shows the E_b/N_o (bit energy to noise density) needed for reliable communication (i.e. operation at $r = R_o$) vs. the code rate r, for several values of k (throughout, the graphs present r in bits/symbol for convenience). Note that $(E_r + E_s)/N_o = rE_b/N_o$, or $E_s/N_o = \frac{rE_b/N_o}{1+1/k}$. The coding gain is increased as the code rate is lowered, as expected.

For cases II and and III, we specialize here to the *Rayleigh-fading channel* for which $E_r = E$, the transmitted symbol energy. The motivation for this is the evident expectation, that the worse the fading is, the more degradation is incurred by departing from the optimal matched metric.

For case II, i.e. when only the statistics of ρ are given, the following expression

$$E_\rho p(y|\text{'0'},\rho) = \frac{1}{\sqrt{\pi N_o}}\int_0^\infty 2\rho e^{-\rho^2} e^{-(y-\rho\sqrt{E})^2/N_o} d\rho =$$

$$\frac{N_o}{(E+N_o)\sqrt{\pi N_o}}\exp\left[-y^2\left(1-\frac{E}{2(E+N_o)}\right)/N_o\right] D_{-2}\left(-y\frac{\sqrt{2E/N_o}}{\sqrt{E+N_o}}\right) \qquad (28)$$

is calculated (using Eq. 3.462, no. 1 from [20]). Here $D_p(\cdot)$ is the parabolic cylinder function. The value of D_{II} is evaluated using Eq. 19.3.7 and 19.14 from [27], to yield

$$D_{II} = \frac{N_o}{(E+N_o)\sqrt{\pi N_o}} 2\int_0^\infty \exp\left[-y^2/(E+N_o)\right] \cdot [erfc(\beta y)\sqrt{2\pi}\beta y + \exp(-\beta^2 y^2/2)]^{1/2} \cdot \left[\exp(-\beta^2 y^2/2) - erfc(-\beta y)\sqrt{2\pi}\beta y\right]^{1/2} dy, \qquad (29)$$

where

$$\beta = \frac{\sqrt{2E/N_o}}{\sqrt{E+N_o}}$$

$$erfc(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-z^2/2}dz \quad .$$

This expression can be calculated numerically. Fig.2 shows the E_b/N_o needed to achieve $r = R_o$, in this case. We note a degradation of about 1.1 dB at $r = 1/2$ and $r = 1/4$, with respect to the optimal processing E_b/N_o (Fig. 1, $k = 0$ curve). The results for case II, $r = 1/2$ and for case I, rate 1/2, agree with those reported in [2].

For case III, we assume that the receiver uses the common additive white Gaussian channel metric (i.e. it is matched to antipodal signaling over AWGN channel) and investigate its performance in a Rayleigh fading environment. We therefore use

$$\frac{q(y|\text{'0'},z)}{q(y|\text{'1'},z)}=\frac{q(y|\text{'0'})}{q(y|\text{'1'})}=e^{4y\sqrt{E}/N_o}$$

and D_{III} is given by (see (13), using 2λ for convenience)

$$D_{III}=\inf_{\lambda\geq 0}\int_{-\infty}^{\infty}E_\rho p(y|\text{'1'},\rho)e^{4\lambda y\sqrt{E}/N_o}dy\ . \quad (30)$$

Here we first integrate over y, then invoke standard mathematical treatment to obtain

$$D_{III}=\inf_{\lambda\geq 0}e^{2\lambda^2/\alpha}\left[1-2\sqrt{\pi}\lambda e^{\lambda^2}erfc(-\sqrt{2}\lambda)\right]\ . \quad (31)$$

Fig. 2 shows the E_b/N_o required to work at $r=R_o$ versus the code rate r, where the optimal λ was found numerically. We note that the performance is very close to that of case II. For $r=1/2$, there is a 1.3 dB degradation versus case I. (In [3], a degradation of about 2 dB is shown for rate $=\frac{1}{2}$, $k=7$ convolutional-coded BPSK – with soft-decision decoding – for this case vs. the CSI case).

The former model assumes that the receiver tracks in phase the composite signal of specular and diffuse components. Jacobs [21] used *another* description for the Rician channel. According to this model, for a transmitted waveform x_j, the receiver observes the quadrature outputs:

$$\begin{aligned}r_1&=w_1x_j+n_1\\ r_2&=w_2x_j+n_2\end{aligned} \quad (32)$$

where n_1,n_2 are statistically independent, zero mean Gaussian random variables (r.v.) with variance of $N_o/2$. Here x_j represents the transmitted signal, normalized to unit energy; w_1,w_2 are statistically independent Gaussian r.v. and account for both the Specular and Diffuse components of the received signal. The means and variances of w_1,w_2 are:

$$\overline{w}_j=\begin{cases}\sqrt{E_s}\cos\theta_o, & j=1\\ \sqrt{E_s}\sin\theta_o, & j=2\end{cases} \quad (33)$$

$$\sigma^2(w_1)=\sigma^2(w_2)=E_r/2\ ,$$

where E_s represents the specular energy, E_r the diffuse component, and θ_o is the phase of the specular component, assumed constant and known (or ideally tracked) by the receiver[1]. Note the difference between this model and the one introduced in the beginning of this paragraph; this model relates the mean received signal to the specular energy, namely the receiver *tracks in phase only the Specular component.* Had the receiver been given the values of w_1 and w_2, the value of D_I for $E_s>0$ equals to the value found in (27). In this case, one obtains

$$p(r_1,r_2|\text{'0'})=\frac{1}{\pi(E_r+N_o)} \quad (34)$$
$$e^{-\left[(r_1-\sqrt{E_s}\cos\theta_o)^2+(r_2-\sqrt{E_s}\sin\theta_o)^2\right]/(E_r+N_o)}\ .$$

[1] For an antipodal signaling, the receiver could deal with the inphase component only, using this phase shift; for generality, we choose to keep the two quadrature components.

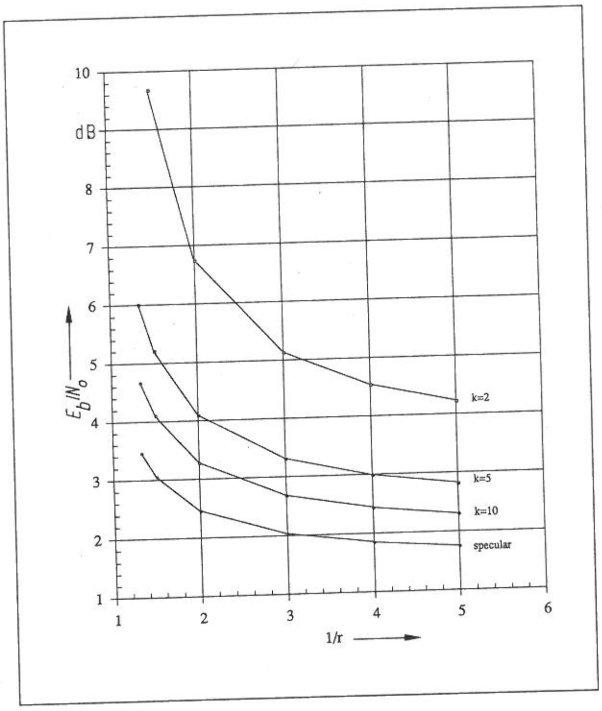

Fig. 3. E_b/N_o required to achieve $r=R_o$ (bits/symbol), Jacobs' Rician channel, statistics information.

D_{II} is given by

$$D_{II}=\exp\left[-\frac{E_s/N_o}{1+E_r/N_o}\right]\ , \quad (35)$$

which is the same expression as can be deduced from Jacobs' Chernoff upper bound result [21].

The channel probabilities here are Gaussian, and therefore using the "normal" metric as for an AWGN channel, does not incur any degradation (namely one gets $D_{III}=D_{II}$); in other words no additional gain is evidenced by incorporating the Rayleigh statistics information. Fig. 3 shows the required E_b/N_o to achieve $r=R_o$, for several values of $k=E_s/E_r$, vs. code rate r. D_{II} of Jacobs' model happens to be close to the value of D_I when the fading is not severe (say, a k of 10). The amount by which E_b/N_o has to be increased in case II, to get the same D as in case I for the first model, is only about 0.5 dB, for $k=10$. Thus, under moderate fading conditions, the degradation due to mismatch is also moderate. Note that this model treats the diffuse energy as Gaussian noise, while the former model did not. Thus, in this model, no information transfer by coherent PSK on a Rayleigh channel is possible.

Mutual Information Calculations

Here we specialize to calculation of the (generalized) mutual information per symbol between channel output and input, for the binary PSK, Rayleigh fading, and cases I and III. A symmetric input distribution achieves capacity, as the channel is symmetric. In case I,

$$I(X;Y)=E_\rho I(X;Y|\rho)=$$

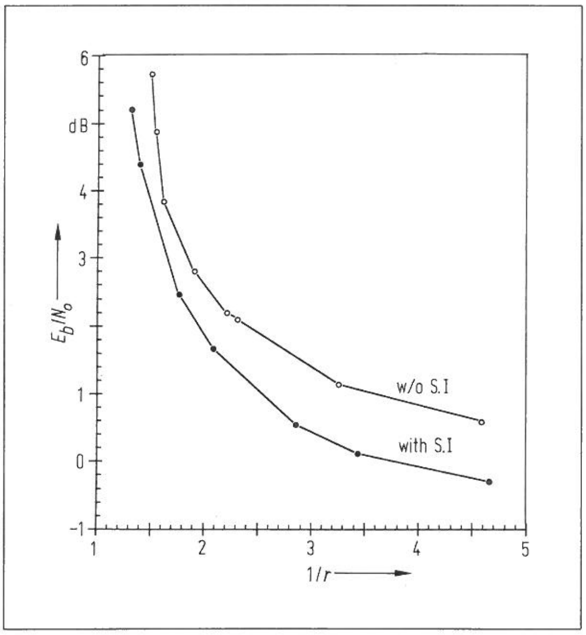

Fig. 4. E_b/N_o required to achieve $r = I(X;Y)$, Rayleigh channel with/without side information.

$$\ln 2 - \int_y dy e^{-y^2 E/N_o} \sqrt{\frac{E}{\pi N_o}} \int_{\rho=0}^{\infty} 2\rho \quad (36)$$
$$e^{-\rho^2(E/N_o+1)+2y\rho E/N_o} \ln(1+e^{-4y\rho E/N_o}) dy \,.$$

This expression is calculated numerically. Fig. 4 shows the required E_b/N_o to operate at $r = C$.

In case III, the decoder uses

$$q(y|x) = \frac{1}{\sqrt{\pi N_o}} e^{-(y-\sqrt{E})^2/N_o} \qquad ,$$

and the effective channel transition probabilities are $E_\rho p(y|x,\rho)$. Thus, the generalized mutual information (20) is

$$I(X;Y(p,q))$$
$$= \max_{s\geq 0} \int_y E_\rho p(y|\text{'0'},\rho) \ln \frac{q(y|\text{'0'})^s}{\frac{1}{2}q(y|\text{'0'})^s + \frac{1}{2}q(y|\text{'1'})^s} dy$$
$$= \max_{s\geq 0} E_\rho \int_y dy \frac{1}{\sqrt{\pi N_o}} e^{-(y-\rho\sqrt{E})^2/N_o}$$
$$\ln \frac{2}{1+e^{-4sy\sqrt{E}/N_o}} \,. \quad (37)$$

The required E_b/N_o to work at $r = I(X;Y)$ for this case, namely no side information, are depicted in Fig. 4. One notes that the behavior of the mutual information is similar to the cut-off rate behavior, shown in Fig. 2, in the sense that lower E_b/N_o values are required for lower code rates. The generalized mutual information is very close to the capacity for case II, rate $\frac{1}{2}$, reported in [2].

3.2 Nakagami-m Fading Channel

The Nakagami-m distribution is also a commonly used statistical model for the overall amplitude of a slowly fading channel (e.g. [22]–[24]). It has been suggested by several authors that this distribution fits experimental data better than the Rician model — in particular for urban radio multipath channels and scintillation channels ([23], [25]). According to this model, the amplitude ρ of the received signal has the p.d.f.:

$$P(\rho) = \frac{2m^m \rho^{2m-1}}{\Gamma(m)\Omega^m} e^{-\frac{m\rho^2}{\Omega}} \,, \quad (38)$$
$$m = \frac{\Omega^2}{var(\rho^2)} \geq \frac{1}{2}$$
$$\Omega = E(\rho^2) \triangleq \overline{\rho^2} \,,$$

where $\Gamma(\cdot)$ is the Gamma function.

With no loss in generality, one can take $\Omega = 1$, so that the average received energy is $\overline{\rho^2 E} = E$. This distribution was introduced by Nakagami, who noted that there is a close fit between the m-distribution and the Rician distribution when the following relationship holds:

$$\frac{1}{k} = \frac{m - \sqrt{m^2-m}}{\sqrt{m^2-m}} \,, \quad \text{for } m \geq 1 \quad (39)$$

where k is defined in Section 3.1 for the Rician model. Note that $m = 1$ gives exactly the Rayleigh model, while $m \to \infty$ corresponds to the nonfaded channel. Similarly to the first Rician model, we assume that the receiver tracks the phase of the composite signal. Most published work dealt with the uncoded performance of several modulation schemes over this channel. Uncoded and coded performance of M-ary FSK (with hard-decision decoding) was recently reported [23]. The results show that with large E_b/N_o, this model predicts an m-th order diversity; i.e. the bit error rate of the uncoded scheme varies as the inverse m-th power of E_b/N_o, and in the coded case we get the inverse m-th power of $(t+1)\,E_b/N_o$, where t is the error correction capability of the code used.

This observation motivated our interest in the Nakagami model, for we know that the Rician channel model does not bear any equivalent "diversity". To this end, we investigate the D-parameter for case I. Equivalently to Eq. (26), we have

$$D_I = \int_0^\infty e^{-\rho^2 E/N_o} \frac{2m^m \rho^{2m-1}}{\Gamma(m)} e^{-m\rho^2} \cdot d\rho$$
$$= \frac{m^m}{\Gamma(m)} \int_0^\infty \rho^{2(m-1)} e^{-\rho^2(m+E/N_o)} 2\rho d\rho$$

Assuming m is an integer, using Eq. 3.351 in [20] yields

$$D_I = \frac{m^m}{\Gamma(m)}(m-1)!\,(m+E/N_o)^{-m} = \left[\frac{m}{m+E/N_o}\right]^m \,. \quad (40)$$

Indeed, this is the same result as (27) for $k = 0$, when $m = 1$. Fig. 5 shows the required Eb/N_o for $r = R_o$ versus r, for the following values of m; $m = 6$ (corresponding to $k = 10.47$), $m = 3$ (corresponding to $k = 4.45$), and $m = 2$ ($k = 2.41$). It is noted that the difference between these results and those shown

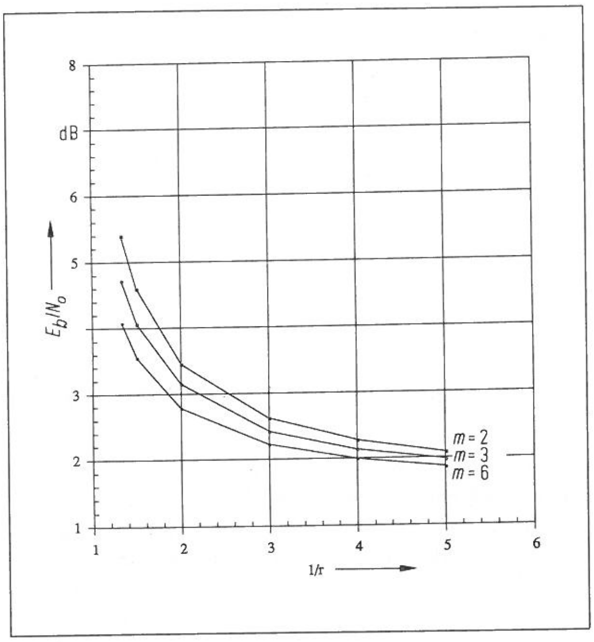

Fig. 5. E_b/N_o required to achieve $r = R_o$ (bits/symbol), Nakagami channel with state information.

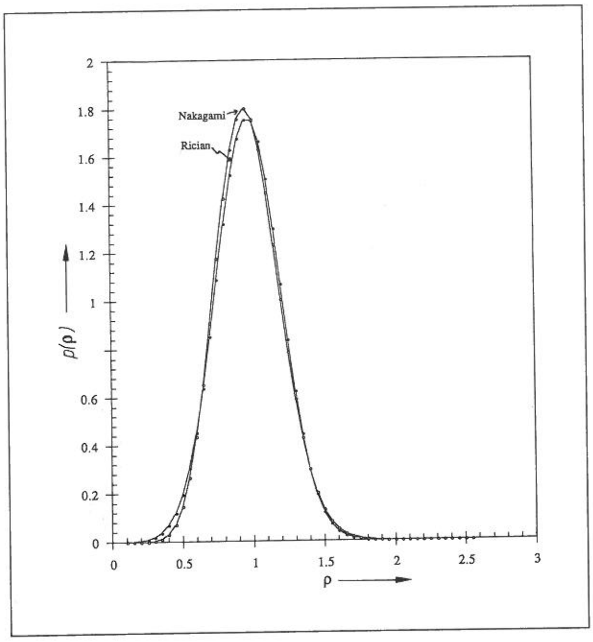

Fig. 6. Rician and Nakagami distributions ($k = 8.4745$ and $m = 5$, respectively).

in Fig. 1 for $k = 10, 5, 2$ are indeed small. Only at $m = 2$ and high code rates which require high E_b/N_o values the difference is noticeable. So it does seem that for some cases the Nakagami model yields equivalent performance results to the Rician model.

Let us check the case of high E/N_o and $m > 1$. For example, taking $m = 5$, eq. (40) shows a behavior like $[E/N_o]^{-5}$ for high E/N_o values, however, eq. (27), for very high E/N_o, behaves like $[E/N_o]^{-1}$. Indeed, the Nakagami model shows an m-th order "diversity" also here. This difference is attributed to the different behavior of the two probability distributions for very small values of ρ (i.e. those events which actually produce most of the transmission errors in a high E/N_o situation). The Nakagami-m distribution behaves like $[\rho^2]^{m-1/2}$, for very small values of ρ, while the Rician distribution for $\rho \to 0$ behaves roughly like ρ. Figs. 6 and 7 show a comparison of the two distributions on a linear and logarithmic scale, respectively, for $m = 5, k = 1/0.118$. It is clear that the two p.d.f. are close, in general, but exhibit a different behavior at small values of ρ.

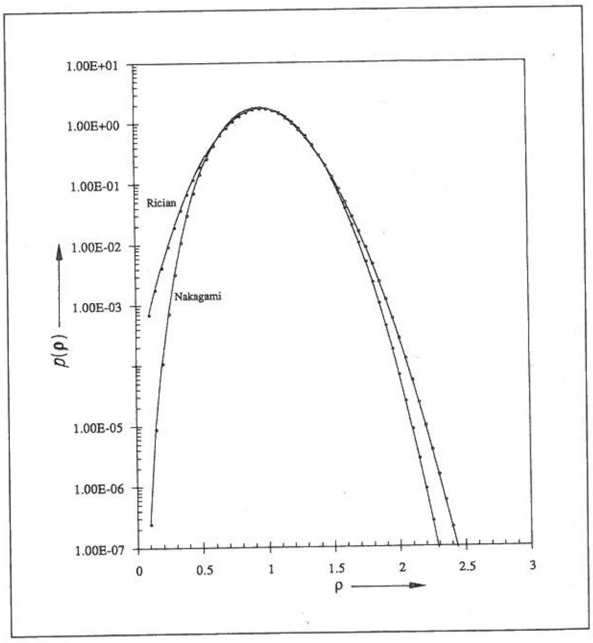

Fig. 7. Rician and Nakagami distributions – semi-log scale ($k = 8.4745$ and $m = 5$, respectively).

3.3 Bhattacharyya Distance – State Estimation with Errors

The Bhattacharyya distance for the mismatched case with side information, is given by (from (13))

$$D = \inf_{\lambda \geq 0} E_z \left\{ E_{y|z} e^{\lambda[m(\hat{x},y|z) - m(x,y|z)]} \right\} \quad (41)$$

where $m(x, y|z)$ is the metric used by the decoder. For a decoder with practical (imperfect) side information $\hat{z}$ assumed for simplicity to be independent of the channel inputs, one has

$$D = \inf_{\lambda \geq 0} E_z E_{\hat{z}|z} E_{y|z,\hat{z}} e^{\lambda[m(\hat{x},y|\hat{z}) - m(x,y|\hat{z})]} . \quad (42)$$

We assume here the Gaussian approximation,

$$p(\hat{z}|z) = \frac{1}{\sqrt{2\pi\sigma_e^2}} e^{-(z-\hat{z})^2/2\sigma_e^2} , \quad (43)$$

that is $\hat{z}$ given z is Gaussian with mean z and variance σ_e^2 standing for the estimation error. However we as-

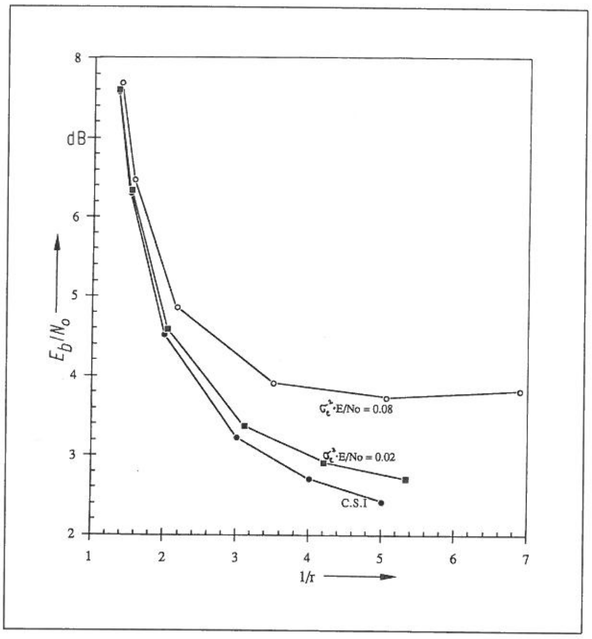

Fig. 8. E_b/N_o required to achieve $r = R_o$ (bits/symbol) (state estimation variance times E_s/N_o as a parameter).

sume that the decoder does not make use of (43), but rather utilizes $\hat{z}$ as though it is perfect. The Gaussian statistics are justified assuming that the channel state is estimated via many outputs.

For the binary-input, Rician channel case,

$$D = \inf_{\lambda \geq 0} E_\rho E_{\hat{\rho}|\rho} E_{y|\rho,\hat{\rho}} e^{-4\lambda\hat{\rho}\sqrt{E}y/N_o}$$
$$= \inf_{\lambda \geq 0} E_\rho \frac{1}{\sqrt{2\pi\sigma_e^2}} \int_{\hat{\rho}} e^{-(\rho-\hat{\rho})^2/2\sigma_e^2}$$
$$e^{-(4E)(\lambda\rho\hat{\rho}-\lambda^2\hat{\rho}^2)/(N_o)}\, d\hat{\rho} \qquad .$$

Here we select $\lambda = 1/2$ which is the optimal choice for $\sigma_e^2 \to 0$, to get

$$D \leq E_\rho e^{-\rho^2 E/N_o} \frac{1}{\sqrt{2\pi\sigma_e^2}} \int_{\hat{\rho}'} e^{(\hat{\rho}'-\rho')^2} d\hat{\rho}' \frac{1}{\sqrt{\frac{1}{2\sigma_e^2} - \frac{E}{N_o}}}$$
$$= E_\rho e^{-\rho^2 E/N_o} \frac{1}{\sqrt{1 - 2\sigma_e^2 \frac{E}{N_o}}} \qquad (44)$$

where it is assumed that $\sigma_e^2 2E/N_o < 1$. The degradation of the cut-off rate, from the ideal case I, depends on the ratio of estimation S/N ratio to symbol S/N ratio.

Note also, that for Rayleigh fading D_I with perfect SI is only inversely proportional to $E_r/N_o = E/N_o$, (Eq. 27), thus a multiplicative degradation factor is not necessarily negligible. Fig. 8 shows the required E_b/N_o to communicate at $r = R_o$ with Rayleigh fading for $\sigma_e^2 E/N_o = 0.08$ and 0.02. The ideal CSI case is also shown. With low-rate coding, where low values of E_b/N_o are appropriate, state estimation errors can have a non-negligible effect.

3.4 Worst-Fading: The Moment Bounding Technique

Assume the same model as in eq. (8a), namely $y_i = \rho_i x_i + n_i$, and perfect side information available to the receiver. The fading statistics are not always known; We examine here the case where the M first moments of the fading ρ are known and we further assume that the dynamic range of the fading is $\rho \in [\rho_{min}, \rho_{max}]$. The best and worst resultant R_o (or best and worst $P(\rho)$) are addressed.

As shown the cut-off is determined by D, where $D = E_\rho e^{-\rho^2 E/N_o}$, in this case. The moment bounding technique ([28]) is a very useful tool for solving such problems, if certain conditions are fulfilled. Basically, the expectation of a functional $g(a)$, is bounded by

$$\sum_{i=1}^{L} w_i' g_1(X_i') \leq E[g(a)] \leq \sum_{i=1}^{L} w_i'' g(X_i'') \qquad (45)$$

However, the basic conditions to check for given M moments of ρ, is that the first $(M+1)$-derivatives of $g(a)$ are continuous, and that the $(M+1)$th derivative is non-negative through the range of a. To meet this condition define $a \triangleq \rho^2$, to get

$$D = \int p_a(a) e^{-a\,E/N_o} da \qquad (46)$$

where $a \in [a_1, a_2]$. We use $g(a) = -e^{-aE/N_o}$ for even M or $-g(a)$ for odd values of M. As is well known, the extremizing conditions are met by *discrete* density functions. For a numerical example, suppose $M = 2$, namely $E[a] = E[\rho^2]$ and $E[a^2] = E[\rho^4]$ are prescribed. Then applying the moment bounding technique, one has

$$E[e^{-aE/N_o}] \leq (1-w_2)(e^{-a_1 E/N_o}) + w_2(e^{-X_2' E/N_o}) \qquad (47)$$

where

$$X_2' = \frac{\overline{a^2} - \bar{a}a_1}{\bar{a} - a_1}, \quad w_2 = \frac{(\bar{a}-a_1)^2}{(\bar{a}-a_1)^2 + \sigma_a^2} \qquad (48)$$

[the pdf is composed of two Dirac delta functions, one at a_1 and the other at X_2']. Equation (47) provides a lower bound on R_o.

The upper bound extremizing density has mass at X_1'' and $X_2'' = a_2$, where

$$X_1'' = \frac{a_2\bar{a} - \overline{a^2}}{a_2 - \bar{a}} = \bar{a} - \frac{\sigma_a^2}{a_2 - \bar{a}} \qquad (49)$$

and the weight

$$w_1 = \frac{(a_2 - \bar{a})^2}{(a_2 - \bar{a})^2 + \sigma_a^2}$$

which yields,

$$E[e^{-aE/N_o}] \geq w_1 e^{-X_1'' E/N_o} + (1-w_1) e^{-a_2\,E/N_o} \qquad (50)$$

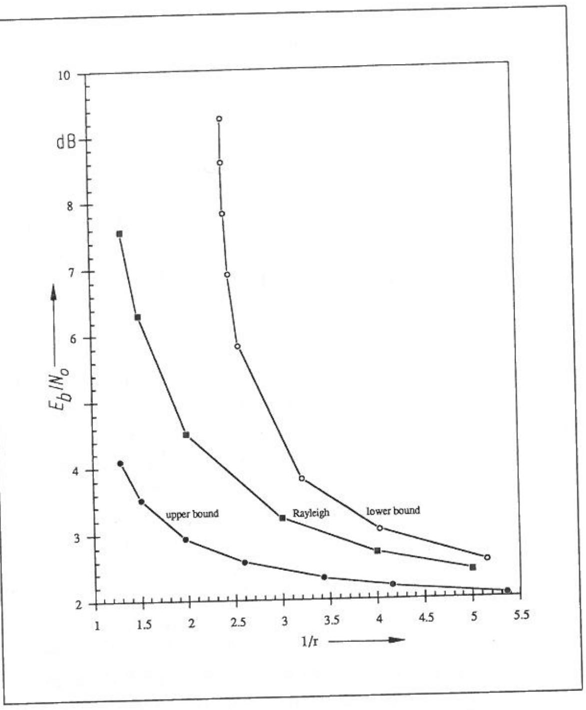

Fig. 9. Lower and upper moment bounds on the required E_b/N_o to achieve $r = R_o$ (bits/symbol), state side information.

To get some insight, we use the value of the moments of the Rayleigh pdf (for ρ). Namely, $E[a] = 1, E[a^2] = 2$. We take $a_1 = 0$, for the lower bound and $a_2 = 10$ for the upper bound. Note that although the Rayleigh pdf has a semi-infinite range, any practical system will have a finite upper limit on signal strength, due to quantization (or limiting). It also turns out, that for $a_2 \to \infty$, the bound converges to D of unfaded BPSK.

Fig. 9 presents the required E_b/N_o to communicate at $r = R_o$ for these bounds, together with the actual result for the Rayleigh pdf. In this example, the lower bound is not very useful at high code rates, due to the large weight at $a_1 = 0$. On the other hand, at low code rates, the actual performance is close *to both bounds.* It is expected that the use of more moments will tighten these bounds at high code rates.

If however *no* side information is available, the receiver uses a metric which does not depend on the instantaneous ρ. Then,

$$D = \inf_{\lambda \geq 0} E_\rho \left\{ \int_y p(y|x,\rho) e^{\lambda[m(y,\hat{x})-m(y,x)]} \, dy \right\}$$

and, for a *given* metric, range of ρ, and M moments of ρ, the best and worst performance can still be checked via the same method. For example, for the binary PSK case, we found that

$$D = \inf_{\lambda \geq 0} e^{4\lambda^2 E/N_o} E_\rho e^{-4\lambda \rho E/N_o} \quad .$$

In this case the bounds involve moments of ρ, (not ρ^2 as in the previous case), and for each bound (and E/N_o) the optimum value of $\lambda \geq 0$ should be found.

4. Conclusions

In the work, the achievable information rates for a class of compound channels (often denoted as time-varying channels) with various levels of side-information were investigated. First, some general novel properties of the random-coding error exponent were proved and then the generalized mutual information was presented and discussed in view of other mismatched information theoretic measures. In the sequel, we dealt with antipodal signaling with three models of flat fading, analyzed the performance under various SI levels, and pointed out the similarities and differences between the fading models used. An interesting *conjecture* can be drawn: at high code rates, where large values of E_b/N_o are appropriate, the decoder metric is less dominant, while differences in the statistical model of the fading can be rather important. At low coding rates, on the other hand, the amount of mismatch is indeed a dominant factor.

Appendix

Here the proof of the r.h.s. inequality of (23) is outlined. Rewrite $I(X;\, Y(p,q))$ (omitting the maximization over $s \geq 0$) as

$$\begin{aligned} I(X;Y(p,q)) = & \\ & -\sum_x Q(x) \ln Q(x) + \sum_y p(y) \sum_x p(x|y) \cdot \\ & \cdot \ln \frac{Q(x) q^s(y|x)}{\sum_{x'} Q(x') q^s(y|x)} = \\ & = H(x) + \sum_y p(y) \sum_x p(x|y) \ln q^s(y|x) Q(x) - \\ & - \sum_y p(y) \ln \sum_{x'} Q(x') q^s(y|x') \, . \end{aligned} \quad (51)$$

Use the inequality (by (22b))

$$\begin{aligned} & \textstyle\sum_y p(y) \sum_x p(x|y) \ln q^s(y|x) \leq \\ & \textstyle\sum_y f(y) \sum_x f(x|y) \ln q^s(y|x) \end{aligned}$$

and replace $p(y)$ by $f(y)$, to obtain

$$\begin{aligned} & I(X;Y(p,q)) \leq \\ & , \quad H(X) + \sum_y f(y) \sum_x f(x|y) \ln q^s(y(x)) Q(x) - \\ & - \sum_x f(x|y) \sum_y f(y) \ln \sum_{x'} Q(x') q^s(y|x') \\ & = H(X) + \sum_y f(y) \sum_x f(x|y) \cdot \\ & \cdot \ln \frac{Q(x) q^s(y|x)}{\sum_{x'} Q(x') q^s(y|x')} . \end{aligned}$$

As the argument of the $\ln(\cdot)$ function in the last expression is a probability density function in x, we can use the well-known inequality $\sum p \ln q \leq \sum p \ln p$, to get

$$I(X;Y(p,q)) \leq H(X) + \sum_y f(y) \sum_x f(x|y) \ln f(x|y) . \tag{52}$$

As this holds for any $s \geq 0$ (in the l.h.s.) and $f(x,y)$ satisfying the constraints in (22b), we finally have

$$I(X;Y(p,q)) \leq H(X) - H'(X|Y) = I'(X;Y)$$

□

References

[1] Wolfowitz, J.: Coding theorems of information theory. 3rd ed. Berlin: Springer-Verlag, 1978.
[2] Dorsch, B.: Performance and limits of coding for simple time-varying channels. Proc. Int. Zurich Sem. on Dig. Commun., Zürich, 1980. G1.1–G1.6.
[3] Hagenauer, J.: Viterbi decoding of convolutional codes for fading and burst channels. Proc. Int. Zurich Sem. on Dig. Commun., Zürich, 1980. G2.1–G.2.7.
[4] Simon, M.; Omura, J.; Scholtz, R.; Levitt, B.: Spread spectrum communications. Vol. I, Ch. 4. Computer Science Press, 1985.
[5] Divsalar, D.; Simon, M.: Trellis coded modulation for 4800–9600 bit/s transmission over a fading mobile satellite channel. IEEE Journ. on Sel. Areas in Commun. SAC- **5** (1987), 162–174.
[6] Stark, W.: Capacity and cut-off rate of noncoherent FSK with nonselective Rician fading. IEEE Trans. Commun. COM- **33** (1985), 1153–1159.
[7] McEliece, R.; Stark, W.: Channels with block interference. IEEE Trans. Inform. Theory IT- **30** (1984), 44–53.
[8] Mushkin, M.; Bar-David, I.: Capacity and coding for the Gilbert-Elliott channel. IEEE Trans. Inform. Theory IT- **35** (1989), 1277–1290.
[9] Massey, J.: Coding and modulation in digital communications. Proc. Int. Zurich Sem. in Dig. Commun., Zürich, 1974.
[10] Gallager, R.: Information theory and reliable communication. New York: Wiley, 1968.
[11] Fisher, T.: Some remarks on the role of inaccuracy in Shannon's theory of information transmission. Trans. 8th Prag. Conf. Information. Theory, 1978. 211–226.
[12] Kazakos, D.: Upper and lower bounds for noisy channel coding under mismatch. Proc. 1981 Conf. Information Sciences and Systems, John Hopkins University, 1981. 37–42.
[13] Dunham, J.; Tzou, K.: Performance bounds for convolutional codes with digital Viterbi decoders in Gaussian noise. IEEE Trans. Commun. COM- **31** (1983), 1124–1132.
[14] Kwon, H.: Capacity and cutoff rate of coded FH/MFSK communications with imperfect side information generators. IEEE Journ. on Sel. Areas in Commun. SAC- **8** (1990), 750–761.
[15] Kaplan, G.; Ram, U.: Bounds on performance for the noisy reference PSK channel. IEEE Trans. on Commun. COM- **38** (1990), 1699–1707.
[16] Hui, J.: Fundamental issues of multiple accessing. Dissertation. Cambridge, MA: MIT, 1983.
[17] Cover, T.; Thomas, J.: Elements of information theory. J. Wiley & Sons, 1991.
[18] Modestino, J.; Mui, S.: Convolutional code performance in the Rician fading channel. IEEE Trans. Commun. COM- **24** (1976), 592–606.
[19] Viterbi, A.; Padovani, R.: Optimal performance of coded-interleaved PSK modulation in slowly varying Rician fading and interference. presented at Milcom '87.
[20] Gradshteyn, I.; Ryzhik, I.: Table of integrals, series and products. Academic Press, 1965.
[21] Jacobs, I.: Probability of error bounds for binary transmission on a slowly fading Rician channel. IEEE Trans. Inform. Theory IT- **12** (1966), 431–441.
[22] Wojnar, A.: Unknown bounds on performance in Nakagami channels. IEEE Trans. Commun. COM- **34** (1986), 22–24.
[23] Crepeau, P.: Coding performance on generalized fading channels. Proc. MILCOM '88, 1988. 15.1.1–15.1.7.
[24] Crepeau, P.: Asymptotic performance of M-ary orthogonal modulation in generalized fading channels. IEEE Trans. Commun. COM- **36** (1988), 1246–1248.
[25] Esposito, R.: Error probabilities for the Nakagami channel. IEEE Trans. Inform. Theory IT- **13** (1967), 145–148.
[26] Stiglitz, I.: Coding for a class of unknown channels. IEEE Trans. Inform. Theory IT- **12** (1966), 189–195.
[27] Abramowitz, M.; Stegun, I.: Handbook of mathematical functions. NBS, 1964.
[28] Krein, M.; Nudelman, A.: The Markov moment problem and extermal problems. Vol. 50. American Math. Soc., 1977. – (Translation of Math. Monographs).
[29] Zakai, M.; Ziv, J.: A generalization of the rate-distortion theory and applications. – In: Information Theory – New Trends and Open Problems. New York: Springer-Verlag, 1975.

Gideon Kaplan was born in Haifa, Israel, on November 13, 1955. He received the B.Sc. (Summa Cum Laude) and M.Sc. degrees in Electrical Engineering from the Technion - Israel Institute of Technology, in 1978 and 1983, respectively. Between 1978-1988 he has been with the Department of Research and Development, Israeli Ministry of Defense, as a Communication Engineer and Project Leader. From 1988 until mid-1989 he was with Qualcomm, Inc., San Diego, CA.

He is presently with Gilat Communication Systems, Tel-Aviv, Israel, and pursuing his Ph.D. in Electrical Engineering at the Technion, Israel. His main intrests are digital and satellite communications and information theory.

Shlomo Shamai (Shitz) received the B.Sc., M.Sc. and Ph.D. degrees in electrical engineering from the Technion - Israel Institute of Technology, in 1975, 1981 and 1986 respectively. During 1975-1985 he was with the Signal Corps Research Labs (Israel Defense Forces) in the capacity of a senior research engineer. Since 1986 he is with the Department of Electrical Engineering, Technion, Israel, where he is now an associate professor.

His research intresets include topics in information teory and analog and digital communications. He is especially interested in theoretical limits in communication with practical constraints, digital communication in optical channels, information theoretic models for magnetic recording, channel coding, combined modulation and coding, digital specrally efficient modulation methods, FM threshold extending techniques, and representation of signals with partial information.