TR-86-7

ROBUST CODING FOR MULTIPLE-ACCESS CHANNELS

by

Evaggelos Geraniotis

ROBUST CODING FOR MULTIPLE-ACCESS CHANNELS

Evaggelos Geraniotis, Member IEEE

Electrical Engineering Department
and Systems Research Center
University of Maryland
College Park, MD 20742

ABSTRACT

The problem of minimax robust coding for classes of multiple-access channels with uncertainty in their statistical description is addressed. We consider: (i) discrete memoryless multiple-access channels with uncertainty in the probability transition matrices and (ii) discrete-time stationary additive Gaussian multiple-access channels with spectral uncertainty. The uncertainty is modeled using classes determined by 2- alternating Choquet capacities. Both block codes and tree codes are considered. A robust maximum-likelihood decoding rule is derived which guarantees that, for all two-user channels in the uncertainty class and all pairs of code rates in a critical rate region, the average probability of decoding error for the ensemble of pairs of random block codes and the ensemble of pairs of random tree codes converges to zero exponentially with increasing block length or constraint length, respectively. The channel capacity and cut-off rate regions of the class are then evaluated.

This research was supported in part by the Office of Naval Research under contract N00014-84-K-0023 and in part by the Systems Research Center at the University of Maryland, College Park, through National Science Foundation CDR-85-00108.

I. INTRODUCTION

For two-user discrete memoryless multiple-access channels whose statistical description (i.e., the probability transition matrix which determines the channel) is known the coding theorems of [1] - [3] guarantee that, if the pair of coding rates lies in a critical region (termed achievable rate region), there exists a pair of block codes such that the error probability of the decoder approaches zero exponentially with increasing block length. Similar results for two-user tree codes were established in [4].

For channels whose statistical description is not perfectly known but the determining quantity (e.g., the transition probability matrix) belongs to a class, the achievable region was derived in [5] for arbitrarily varying MAC's. In [6] a universal coding approach was applied to discrete-memoryless MAC's. According to this approach a finite number of representative channels exists so that, if we code for these channels, all the other channels in the class have asymptotically optimal coded performance. Two possible disadvantages are: (i) a large numer of representative channels may be necessary and (ii) the construction of the representative channels for a given class can be very complicated.

In [7] another method of universal coding which does not use the notion of representative channels was introduced. According to it a "packing lemma" investigates positions of codewords independently of the channel and is used t upperbound the decoding error. The decoding rule employed is termed "maximum mutual information decoding" and is equally independent of the channel statistics.

Here we consider another approach termed minimax robust coding which is based on a worst-case design. The least-favorable channel is singled out and we use its probability transition matrix for maximum-likelihood decoding. Then the probability of error for the ensemble of two-user random block codes approaches zero exponentially with increasing block length for all channels in the class. The disadvantage is that the asymptotic performance for all but the least-favorable channel in the class is not optimal. However, this approach requires only one representative channel for the class (the least-favorable one) which can be explicitly found in several interesting cases. For single user channels this approach was first considered in [8] and for specific uncertainty classes in [9], the companion to this paper. By restricting attention to specific uncertainty classes of channels we can obtain an explicit characterization of the capacity region and of the maximum-likelihood decoding rule which will ensure the asymptotic convergence of the probability of decoding error to zero for all channnels in the class. Therefore this paper is to multiple access channels as the work of [9] is to ordinary Shannon channel. In contrast the more general (and thus less explicit) characterization of capacity regions in [5] is to multiple access channels as the compound channel work of [10] is to the ordinary Shannon channel.

In this paper we apply the minimax robust coding approach for block and tree codes to two-user discrete-memoryless (DM) MAC's and discrete-time stationary additive Gaussian (SG) MAC's which belong to uncertainty classes determined by 2-alternating Choquet capacities [11]. Our choice of these uncertainty models is justified in two ways. First, important uncertainty models like contaminated mixtures [12], total-variation neighborhoods [12], band models [13] and extended p-point models [14] are capacity classes and have played an important role in hypothesis testing [15] and filtering [16]. Second, the least-favorable channels can be explicitly found for the uncertainty classes described by any of the above models. Although in this paper we restrict attention to DM-MAC's and discrete-time stationary Gaussian channels (SGC's) (continuous-time SGC's are also discussed), our results can be extended to other classes of MAC's; e.g., first-order Markov MAC's. As it is common in multi-user information theory the results are established for two-user MAC's, the extension to the multi-user case is then quite straightforward.

The paper is organized as follows. Minimax robust coding for discrete-memoryless MAC's with uncertainty in the probability transition matrices is discussed in Section II and minimax robust coding for discrete-time stationary Gaussian MAC's with uncertainty in the spectral density of the additive Gaussian noise is discussed in Section III. In each of these Sections we first formulate the problem and introduce the necessary concepts and notation. Next, we present channel coding theorems for both block codes and tree codes for the case of mismatch, i.e., when the decoder employs a maximum-likelihood rule which is based on inaccurate knowledge of the channel statistics. Finally, we derive minimax robust coding theorems for the ensemble of two-user random block codes and the ensemble of two-user random tree codes and channels with statistical uncertainty determined by Choquet capacities and evaluate the channel capacity region and the cut-off rate (actually the general error exponent) region for the class of channels. Then, in Section IV a brief summary of this paper and some conclusions are presented.

II. ROBUST CODING FOR DISCRETE MEMORYLESS MULTIPLE-ACCESS CHANNELS

A. Channel Uncertainty Determined by 2-alternating Capacities

Suppose that for a two-user channel X_1 and X_2 are the input alphabets, Y is the output alphabet, and $\mathbf{F} = \sigma(Y)$ is the σ-algebra generated of subsets of Y. A discrete memoryless two-user MAC is characterized by its transition probability matrix $p(y|x_1,x_2)$, $x_1 \in X_1$, $x_2 \in X_2$, $y \in Y$. For each $\underline{x} = (x_1,x_2) \in X_1 \times X_2$ consider the conditional probability measure $P_{\underline{x}}(A) = \int_A dP(y|x_1,x_2)$ where $A \in \mathbf{F}$. Let $p(y|x_1,x_2)$ denote the Radon-Nikodym derivative of $P_{\underline{x}}$ with respect to a measure λ. The reference measure λ is chosen according to the particular case of interest. Thus, if the alphabet Y is a continuum, λ is the Lebesgue measure on Y. If Y is discrete (e.g., a finite set), then λ is the measure which assigns equal mass to all the elements of Y. Finally, if Y has both discrete and continuous components, then λ turns out to be a convex combination of the Lebesgue measure on the continuous part of Y and the measure that assigns equal mass to all the elements of the discrete part of Y.

We assume that for each $\underline{x} \in X_1 \times X_2$ the probability measures $P_{\underline{x}}$ are only known to lie in a convex class generated by a Choquet 2-alternating capacity [11]

$$\mathbf{P}_{v_{\underline{x}}} = \{P_{\underline{x}} \in \mathbf{P} \mid P_{\underline{x}}(A) \leq v_{\underline{x}}(A),\ \forall A \in \mathbf{F}\} \tag{1}$$

where $\mathbf{P}$ denotes the class of all probability measures on $(Y,\mathbf{F})$, and $v_{\underline{x}}$ is

2-alternating capacity on $(Y,\mathbf{F})$ with $v_{\underline{x}}(Y) = 1$. For notational convenience, in the sequel we will drop the dependence of $v_{\underline{x}}$ and $\mathbf{P}_{v_{\underline{x}}}$ on $\underline{x}$.

A Choquet 2-alternating capacity [11] on $(U,\mathbf{F})$ is a finite set function, which is increasing, continuous from below, continuous from above on closed sets, and satisfies $v(\emptyset) = 0$ and $v(A \cup B) + v(A \cap B) \leq v(A) + v(B)$ for all $A, B \in \mathbf{F}$. Notice that any finite measure v is a 2-alternating capacity; in this case the uncertainty class generated by (1) reduces to $\mathbf{P}_v = \{v\}$. If we further assume that U is compact then all the uncertainty models mentioned in Section I are capacity classes. If U is not compact [e.g., $U = (-\infty, \infty)$] only the band model can be defined in terms of a capacity.

An example of a 2-alternating capacity class is the total-variation neighborhood model [12] defined by

$$\mathbf{P}_v = \{P \; |P_0(A) - P(A)| < \varepsilon \quad , \quad \forall \in \mathbf{F}\} \tag{2}$$

where P_0 is a known measure (not necessarily a probability measure) and ε in $[0,1]$ is the degree of uncertainty in the model. Then (2) can be expressed in the form (1) if we set

$$v(A) = \min\{P_0(A) + \varepsilon \quad , 1\} \tag{3}$$

which is a 2-alternating capacity. See [12], [13] and [14] for a description of other capacity classes.

In the sequel we will need the following fundamental result which is due to Huber and Strassen [14]:

Lemma 1: If v is a 2-alternating capacity on $(Y,\mathbf{F})$ and $\mathbf{P}_v$ is a convex class of probability measures determined by it as in (1), then there exists a unique λ measurable function π_v: $Y \to [0,\infty]$ with the defining property that for each $\theta \varepsilon [0, \infty]$ and A_θ defined by $A_\theta = \{\pi_v > \theta\}$

$$\theta\lambda(A) + v(A^c) \leqq \theta\lambda(A_\theta) + v(A_\theta^c) \tag{4}$$

for all A ε **F**. Furthermore there exists a measure $\hat{P}$ in $\mathbf{P_v}$ such that for all $\theta \varepsilon [0,\infty]$

$$\hat{P}(\{\pi_v \leqq \theta\}) = v(\{\pi_v \leqq \theta\}) \tag{5}$$

which means that $\hat{P}$ makes π_v <u>stochastically smallest</u> over all P in $\mathbf{P_v}$, and π_v is a version of $d\hat{P}/d\lambda$, the generalized Radon-Nikodym (R-N) derivative of $\hat{P}$ with respect to λ; that is $d\hat{P}/d\lambda$ may be infinite on a set of λ measure 0.

The function π_v is termed the <u>Huber-Strassen derivative</u> of v with respect to λ (v may not be a measure). The probability measure $\hat{P}$ singled out by Lemma 1 is termed the <u>least-favorable measure</u> of the class $\mathbf{P_v}$. Let $\hat{P} = \hat{P}' + \hat{P}''$ be the Lebesgue decomposition of $\hat{P}$, where $\hat{P}'$ is absolutely continuous with respect to λ and $\hat{P}''$ is singular with respect to λ (that is, it concetrates all its mass on sets of λ measure 0). Then,

$$\hat{P}'(A) = \int_A \pi_v \, d\lambda \tag{6a}$$

and

$$P''(A) = v(A \wedge \{\pi_v = \infty\}), \tag{6b}$$

for all A ε **F**. For example for the total-variation model of (2) the Huber-Strassen derivative $\pi_v = \hat{p}$ is defined as

$$\hat{p}(y) = \max\{c', \min\{c'',\pi_0(u)\}\} \tag{7}$$

where $\pi_0 = dP_0/d\lambda$ is the R-N derivative of P_0 of (2) and c', c'' are chosen so that $\hat{P}(Y) = 1$. See [12] - [14] for the definition of $\hat{p}$ for the other capacity classes.

We emphasize that for the case treated in this section of the paper the probability measure P_x and the Choquet capacity $v_{\underline{x}}$ actually depend on $\underline{x}$ (e.g., $v_{\underline{x}}$ of (3) actually depends on $\underline{x}$ through $P_{0,\underline{x}}$ and $P_{1,\underline{x}}$ which vary with $\underline{x}$) and so does $\hat{p}_{\underline{x}} = \pi_{v_{\underline{x}}}$.

It should be noted that Huber-Strassen derivatives of generalized capacities [a generalized capacity is defined in the same way as a 2-alternating capacity except that it is required to be continuous from above on compact (and not just closed) sets] with respect to σ-finite (and not just finite) measures can be constructed [20, Chapter IV]. Then, Lemma 1 still holds provided that it is properly modified. One of the implications of this extension is that several of the most useful examples of capacity classes (e.g., ε-mixtures,variation neighborhoods) are generalized capacities when U is σ-compact (and not just compact).

B. Mismatch Coding Theorems for Two-User Block and Tree Codes

Suppose that in the presence of uncertainty about $P_{\underline{x}}(A)$, $\underline{x} \in X_1 \times X_2$, $A \in \mathbf{F}$, the decoder mistakenly assumes that (or attempts to estimate $P_{\underline{x}}$ and comes with an estimate that) $\tilde{P}_{\underline{x}}$ is the probability distribution governing the statistics of the DM-MAC. Therefore it uses a maximum-likelihood (ML) decoding rule based on $\tilde{p}(y|\underline{x}) = (d\tilde{P}_{\underline{x}}/d\lambda)(y)$ instead of the true $p(y|\underline{x})=(dP_{\underline{x}}/d\lambda)(y)$. This situation is characterized by <u>mismatch</u>. Then the following result holds for block codes used on a two-user MAC [6]:

<u>Theorem 1</u> : Consider a DM-MAC characterized by $P_{\underline{x}}(A)$, $A \in \mathbf{F}=\sigma(Y)$. Let $Q_j, j = 1,2$ be an arbitrary probability assignment on the user j channel input symbols. Suppose the decoder employs inaccurate ML decoding based on $\tilde{p}(\cdot|\cdot,\cdot)$ instead of the true probability transition matrix $p(\cdot|\cdot,\cdot)$. Then, for $\underline{R} = (R_1,R_2,R_3)$ satisfying

$$R_j < I_j(\underline{Q},\tilde{p};P) \tag{8}$$

where $R_3 = R_1 + R_2$, $\underline{Q} = (Q_1,Q_2)$,

$$I_1(\underline{Q},\tilde{p};P) = \int_{X_1}\int_{X_2}\int_Y \ln \frac{\tilde{p}(y|x_1,x_2)}{\int_{X_1}\tilde{p}(y|x_1',x_2)dQ(x_1')} dP_{\underline{x}}(y)dQ(x_1)dQ(x_2), \tag{9a}$$

$$I_2(\underline{Q},\tilde{p};P) = \int_{X_1}\int_{X_2}\int_Y \ln \frac{\tilde{p}(y|x_1,x_2)}{\int_{X_2}\tilde{p}(y|x_1,x_2')dQ_2(x_2')} dP_{\underline{x}}(y)dQ_1(x_1)dQ_2(x_2), \tag{9b}$$

$$I_3(\underline{Q},\tilde{p};P) = \int_{X_1}\int_{X_2}\int_Y \ln \frac{\tilde{p}(y|x_1,x_2)}{\int_{X_1}\int_{X_2}\tilde{p}(y|x_1',x_2')dQ_1(x_1')dQ_2(x_2')} dQ_1(x_1)dQ_2(x_2), \tag{9c}$$

the average probability of decoding error P_E over the ensemble of pairs of random block codes of rates (R_1, R_2) and length n (for which the n letters of each codeword are chosen from the input alphabets X_1 and X_2 independently and according to Q_1 and Q_2, respectively, while the $[e^{nR_1}]$ and $[e^{nR_2}]$ codewords are mutually independent and equiprobable) is upperbounded by $P_n(\underline{\rho}, \underline{Q}, \tilde{p}; P)$ given by

$$P_n(\underline{\rho}, \underline{Q}, \tilde{p}; P) = \sum_{j=1}^{3} \exp\{-n[E_j(\rho_j, \underline{Q}, \tilde{p}; P) - \rho_j R_j]\} \tag{10}$$

where $\underline{\rho} = (\rho_1, \rho_2, \rho_3)$ and for ρ in [0,1]

$$E_1(\rho, \underline{Q}, \tilde{p}; P) = -\ln\{\int_{X_1}\int_{X_2}\int_Y \tilde{p}(y|x_1, x_2)^{-\frac{\rho}{1+\rho}}$$

$$[\int_{X_1} \tilde{p}(y|x_1', x_2)^{\frac{1}{1+\rho}} dQ_1(x_1')]^{\rho}\} dP_{\underline{X}}(y) dQ_1(x_1) dQ_2(x_2), \tag{11a}$$

$$E_2(\rho, \underline{Q}, \tilde{p}; P) = -\ln\{\int_{X_1}\int_{X_2}\int_Y \tilde{p}(y|x_1, x_2)^{-\frac{\rho}{1+\rho}}$$

$$[\int_{X_2} \tilde{p}(y|x_1, x_2')^{\frac{1}{1+\rho}} dQ_2(x_2')]^{\rho} dP_{\underline{X}}(y) dQ_1(x_1) dQ_2(x_2)\}, \tag{11b}$$

and

$$E_3(\rho,\underline{Q},\hat{p};P) = -\ln\{ \int_{X_1}\int_{X_2}\int_Y \hat{p}(y|x_1,x_2)^{\frac{\rho}{1+\rho}}$$

$$[\int_{X_1}\int_{X_2}\hat{p}(y|x_1',x_2)^{\frac{1}{1+\rho}}dQ_1(x_1')dQ_2(x_2')]^\rho dP_{\underline{x}}(y)dQ_1(x_1)dQ_2(x_2)\}. \tag{11c}$$

For this theorem to be valid it is required that the mismatch mutual information $I_j(\underline{Q},\tilde{p};P)$ $j = 1,2,3$ of (9a) - (9c) should be strictly positive and the exponents $E_j(\rho,\underline{Q},\tilde{p};P)$ $j = 1,2,3$ of (11a) - (11c) should be strictly positive for all ρ in [0,1]. These positivity requirements are satisfied for the choice of $\tilde{p}$ in Theorem 3 below.

The achievable region for the two-user MAC and inaccurate ML decoding is then defined as the closure of the convex hull of the union of the sets $\mathbf{R}(\underline{Q})$ of rate pairs (R_1,R_2) which satisfy (8) as $\underline{Q} = (Q_1,Q_2)$ ranges over all possible probability measures on $X_1 \times X_2$.

Remark 1. We used the notation of $I_j(\underline{Q},\tilde{p};P)$ for $j = 1,2,3$ instead of $I(X_1;Y|X_2)$, $I(X_2;Y|X_1)$, and $I(X_1,X_2;Y)$, respectively, to emphasize the dependence of the mismatch mutual information functions on $\underline{Q}$ and both $\tilde{p}$ and P; the notation $I(X_1;Y|X_2)$ is usually reserved for the matched case ($\hat{p}=p$). Also notice that for notational convenience we have dropped the dependence of $\tilde{p}_x$, $\tilde{P}_x$, p_x, and P_x on x.

Remark 2. We consider Theorem 1 important in two ways: as being a fundamental intermediate result necessary for the proof of Theorem 2a below, and as an interesting independent result which completely characterizes the achievable rate region for the case of mismatch (i.e., when the actual channel probability transition matrix p is different that the estimate $\tilde{p}$ employed in the ML decoding).

For two-user tree codes and a decoder which employs a ML test based on $\tilde{p}(\cdot|\cdot,\cdot)$ instead of the true $p(\cdot|\cdot,\cdot)$ (about which there is uncertainty) the following result holds:

Theorem 2 : Under the assumptions of Theorem 1 suppose that user j ($j = 1,2$) is assigned a tree code of rate $R_j = \frac{1}{N} \ln M_j$ nats per channel symbol satisfying (8) and constraint length K, and consider the ensemble of random two-user tree codes generated by assigning N channel input letters independently and according to the probability distribution Q_j to the branches of the trees. Then the average probability of decoding error P_E over the above ensemble of pairs of tree code is upperbounded by $P_K(\underline{\rho},\underline{Q},\tilde{p};P)$ given by

$$P_K(\underline{\rho},\underline{Q},\tilde{p};P) = \sum_{j=1}^{2} e^{-KNE_j(\rho_j,\underline{Q},\tilde{p};P)} f(N[E_j(\rho_j,\underline{Q},\hat{p};P)-\rho_j R_j])$$

$$+ e^{-KNE_3(\rho_3,\underline{Q},\tilde{p};P)} f(N[E_3(\rho_3,\underline{Q},\tilde{p};P)-\rho_3(R_1+R_2)])$$

$$\cdot\{1+f(N[E_1(\rho_3,\underline{Q},\tilde{p};P)-\rho_3R_1])+f(N[E_2(\rho_3,\underline{Q},\tilde{p},P)-\rho_3R_2)\} \tag{12}$$

where $f(x) = e^{-x}/(1-e^{-x})$, $\underline{\rho} = (\rho_1,\rho_2,\rho_3)$, $0 \leq \rho_j \leq \min\{E_j(\rho_j,\underline{Q},\tilde{p};P)/R_j,1\}$ for $j=1,2$, and $0 \leq \rho_3 \leq \min\{E_j(\rho_3,\underline{Q},\tilde{p};P)/R_1, E_2(\rho_3,\underline{Q},\tilde{p};P)/R_2, E_3(\rho_3,\underline{Q},\tilde{p};P)/(R_1+R_2),1\}$. The exponents $E_j(\rho,\underline{Q},\tilde{p};P)$ are defined by (11a) - (11c). For this theorem to be valid it is required that $I_j(\underline{Q},\tilde{p};P) > 0$ and $E_j(\rho,\underline{Q},\tilde{p};P) > 0$ for ρ in $[0,1]$ and $j = 1,2,3$; conditions which are satisfied for the choice of $\tilde{p}$ in Theorem 4 below.

The proof of Theorem 2 is based on a straightforward modification of the proof for the case with accurate ML decoding (i.e., $\tilde{p}=p$) given in [4]. The same arguments as in [4] may be used the only difference being that $E_j(\rho,\underline{Q},\tilde{p};P)$ instead of the usual Liao error exponents $E_j(\rho,\underline{Q}) = E_j(\rho,\underline{Q},p;P)$ are involved in the equations, since the decoder now employs $\tilde{p}$ and not p for the ML decision rule.

C. Minimax Robust Coding Theorems for Two-User Block and Tree Codes

In this subsection we assume that the probability measure $P_{\underline{x}}$ which governs the statistics of the channel is only known to lie in a class of the form (1) desribed in Section II.A. The channel encoder employs a ML decoding rule based on $\tilde{p}$ in a way desribed in Theorems 1 and 2, The goal is to choose $\tilde{p}$ so that for all code rates larger than a critical rate the probabiity of decoder error approaches zero with increasing blocklength (or constraint length) for <u>all</u> channels in the class.

Equipped with Theorems 1 and 2 and the Huber-Strassen theory of least-favorability (as condensed in Lemma 1) we now prove the main results of this section.

<u>Theorem 3</u> : Suppose the probability measure $P_{\underline{x}}$ on Y belongs to a class of the form (1) and $\hat{P}_{\underline{x}}$ is the element of the class singled out by Lemma 1. Suppose further that the decoder's ML decoding rule is based on $\hat{p} = d\hat{P}_{\underline{x}}/d\lambda$. Then the following inequalities are true for all pairs of probability measures (Q_1,Q_2) on $X_1 \times X_2$ and ρ in $[0,1]$

$$I_j(\underline{Q},\hat{p};P) \geq I_j(\underline{Q},\hat{p};\hat{P}) \geq I_j(\underline{Q},p;\hat{P}),\ j = 1,2,3 \tag{13}$$

and

$$E_j(\rho,\underline{Q},\hat{p};P) \geq E_j(\rho,\underline{Q},\hat{p};\hat{P}) \geq E_j(\rho,\underline{Q},p;\hat{P}),\ j = 1,2,3. \tag{14}$$

Furthermore, the operating point $(\tilde{\underline{\rho}},\tilde{\underline{Q}},\hat{p})$ where $(\tilde{\underline{\rho}},\tilde{\underline{Q}}) = \arg\min_{(\underline{\rho},\underline{Q})} P_n(\underline{\rho},\underline{Q},\hat{p};\hat{P})$ and the channel determind by $\hat{P}$ form a saddle point for

$\min\limits_{(\underline{\rho},\underline{Q},p')} \max\limits_{P} P_n(\underline{\rho},\underline{Q},p';P)$, i.e.,

$$P_n(\tilde{\underline{\rho}},\tilde{\underline{Q}},\hat{p};P) \leq P_n(\tilde{\underline{\rho}},\tilde{\underline{Q}},\hat{p};\hat{P}) \leq P_n(\underline{\rho},\underline{Q},p;\hat{P}) \tag{15}$$

Finally, for any pair of rates (R_1,R_2) it is necessary and sufficient to lie inside the region determined by the conditions

$$R_j < I_j(\underline{Q},\hat{p};\hat{P}), \; j = 1,2,3 \tag{16}$$

where $\underline{Q} = (Q_1,Q_2)$ ranges over all pairs of probability measures on $X_1 \times X_2$, in order to guarantee that the average probability of decoding error for the ensemble of pairs of random block codes of length n and rates (R_1,R_2) converges to zero exponentially with increasing n for all channels in the class.

Remark 3. The rate region determined by (16) represents the channel capacity region of the class desribed by (1). Similarly, the rate region determined by $R_j < E_j(\rho,\underline{Q},\hat{p};\hat{P})$ where $\underline{Q} = (Q_1,Q_2)$ ranges over all pairs of probability measures on $X_1 \times X_2$ represents for $\rho = .5$ the cutoff rate region.

Remark 4. Notice that equations (13) and (15) indicate that the measure $\hat{P}_{\underline{X}}$ (singled out by Lemma 1) characterizes the worst case (or least-favorable) channel in terms of both the information rate and the error probability among all the channels in the class $P_{\mathbf{v}_{\underline{X}}}$ defined by (1).

Proof: We first prove the inequalities in (13) - (15). In particular, the right-hand inequalities in (13) for j = 1,2,3 are results of Jensen's inequality and the concavity of $\ln(\cdot)$. Similarly, the right-hand inequality in (14) for j = 1 is a result of Holder's inequality

$$\int fgd\mu \leq [\int f^{\alpha} d\mu]^{1/\alpha} \, [\int g^{\beta} d\mu]^{1/\beta} \tag{17}$$

where $1 < \alpha < \infty$, $1 < \beta < \infty$, and $\alpha^{-1} + \beta^{-1} = 1$, when applied for

$$f = p(y|x_1,x_2)^{\frac{\rho}{(1+\rho)^2}}, \; \alpha = (1+\rho)/\rho, \; g = \hat{p}(y|x_1,x_2)^{\frac{1}{1+\rho}} p(y|x_1,x_2)^{-\frac{\rho}{(1+\rho)^2}},$$

$\beta = 1+\rho$, and $d\mu = dQ_1(x_1)$. For j=2 we only need to set $d\mu = dQ_2(x_2)$, whereas for j=3 we should set $d\mu = dQ_1(x_1)dQ_2(x_2)$ and use double integrals instead of single integrals in (17). Finally, the right-hand inequality in (15) is true since

$$P_n(\tilde{\underline{\rho}},\tilde{\underline{Q}},\hat{p};\hat{P}) \leqq P_n(\underline{\rho},\underline{Q},\hat{p};\hat{P}) \leqq P_n(\underline{\rho},\underline{Q},p;\hat{P})$$

where the first inequality holds because of the definition of $\tilde{\underline{\rho}}$ and $\tilde{\underline{Q}}$, while the second inequality follows from the right-hand inequalities of (14) and the fact that P_n [see (10)] is a decreasing function of E_j for j = 1,2,3.

Next we prove the left-hand side inequalities in (13), (14), and (15). We start with the left-hand side inequality in (13). We may use the following sequence of arguments. First we define the functions $G_j(\hat{p}_x,P_x)$ to be

$$G_1(\hat{p}_{\underline{x}},P_{\underline{x}}) = \int_Y \ln \frac{\hat{p}(y|x_1,x_2)}{\int_{X_1}\hat{p}(y|x_1',x_2)dQ(x_1')}\, dP_{\underline{x}}(y) \tag{18a}$$

$$G_2(\hat{p}_{\underline{x}},P_{\underline{x}}) = \int_Y \ln \frac{\hat{p}(y|x_1,x_2)}{\int_{X_2}\hat{p}(y|x_1,x_2')dQ(x_2')}\, dP_{\underline{x}}(y), \tag{18b}$$

$$G_3(\hat{p}_{\underline{x}},P_{\underline{x}}) = \int_Y \ln \frac{\hat{p}(y|x_1,x_2)}{\int_{X_1}\int_{X_2}\hat{p}(y|x_1',x_2')dQ(x_1')dQ(x_2')}\, dP_{\underline{x}}(y), \tag{18c}$$

and observe that we can write $I_j(\underline{Q},\hat{p};P) = \int_{X_1}\int_{X_2} G_j(\hat{p}_x,P_x)dQ_1(x_1)dQ_2(x_2)$ for j = 1,2,3. Here we will make use of the dependence of $\hat{p}$ and P on x and thus we employ the unabbreviated notation $\hat{p}_x$, P_x. Notice that, if we show that

$$G_j(\hat{p}_x,P_x) \geq G_j(\hat{p}_x,\hat{P}_x), \; j = 1,2,3 \tag{19}$$

for all $\underline{x} \in X_1 \times X_2$, then the left-hand side of (13) follows. Equation (19) holds because we can write $G_j(\hat{p}_{\underline{x}},P_{\underline{x}}) = \int_Y g_j(\hat{p}_{\underline{x}})dP_{\underline{x}}$ where for $j = 1,2,3$ g_j is an increasing function of $\hat{p}_{\underline{x}}$ and according to Lemma 1 $\hat{P}_{\underline{x}}$ makes $\hat{p}_{\underline{x}}$ stochastically smallest over all $P_{\underline{x}}$ in $\mathbf{P}_{\mathbf{v}_{\underline{\mathbf{x}}}}$.

The left-hand inequality in (14) can be proved in a similar way. We now define the functions $H_j(\rho,\hat{p}_{\underline{x}},P_{\underline{x}})$ as

$$H_1(\rho,\hat{p}_{\underline{x}},P_{\underline{x}}) = \int_Y \hat{p}(y|x_1,x_2)^{-\frac{\rho}{1+\rho}}\left[\int_{X_1} \hat{p}(y|x_1',x_2)^{\frac{1}{1+\rho}}dQ(x_1')\right]^{\rho}dP_{\underline{x}}(y), \tag{20a}$$

$$H_2(\rho,\hat{p}_{\underline{x}},P_{\underline{x}}) = \int_Y \hat{p}(y|x_1,x_2)^{-\frac{\rho}{1+\rho}}\left[\int_{X_2} \hat{p}(y|x_1,x_2')^{\frac{1}{1+\rho}}dQ(x_2')\right]^{\rho}dP_{\underline{x}}(y), \tag{20b}$$

and

$$H_3(\rho,\hat{p}_{\underline{x}},P_{\underline{x}}) = \int_Y \hat{p}(y|x_1,x_2)^{-\frac{\rho}{1+\rho}}\left[\int_{X_1}\int_{X_2} \hat{p}(y|x_1',x_2')^{\frac{1}{1+\rho}}dQ_1(x_1')dQ_2(x_2')\right]^{\rho}dP_{\underline{x}}(y) \tag{20c}$$

Since $E_j(\rho,\underline{Q},\hat{p};P) = \exp[-\int_{X_1}\int_{X_2} H_j(\rho,\hat{p}_{\underline{x}},P_{\underline{x}})dQ_1(x_1)dQ(x_2)]$ the left-hand side inequality in (14) is satisfied if

$$H_j(\rho,\hat{p}_{\underline{x}},\hat{P}_{\underline{x}}) \geq H_j(\rho,\hat{p}_{\underline{x}},P_{\underline{x}}) \tag{21}$$

is valid for all $\underline{x} \in X_1 \times X_2$ and ρ in [0,1]. Eq. (21) can be proved similarly to (19), that is, by defining an appropriate decreasing function of $\hat{p}_{\underline{x}}$ and applying Lemma 1. The left-hand side inequality in (15) is then a straightforward application of the left-hand side inequality in (14) for $\underline{\rho} = \underline{\tilde{\rho}}$ and $\underline{Q} = \underline{\tilde{Q}}$ and the fact that P_n is a decreasing function of the E_j' s.

Next we prove the positivity of $I_j(\underline{Q},\hat{p};P)$ and $E_j(\rho,\underline{Q},\hat{p};P)$ $j = 1,2,3$ for all $\hat{p} = d\hat{P}/d\lambda$ with P in $\mathbf{P_v}$ all ρ in [0,1], all probability measures $\underline{Q} = (Q_1,Q_2)$ on $X_1 \times X_2$, and $\hat{p} = d\hat{P}/d\lambda$ as singled out by Lemma 1. We first show that $I_j(\underline{Q},\hat{p};P) > 0$ $j = 1,2,3$. We use the fact that $I_j(\underline{Q},\hat{p};\hat{P}) > 0$ [the usual Liao functions are strictly positive unless the channel output are independent in which case they are zero; we exclude this case by requiring that all measures $P_{\underline{x}}$ which belong to the uncertainty class described by (1) are not (for fixed y) constant functions of (x_1,x_2)]; the proof is based on Jensen's inequality and the concavity of $\ln(\cdot)$. Then we use the left-hand side inequality in (13) to prove the desired result. Similarly to prove that $E_j(\rho,\underline{Q},\hat{p};P) > 0$ $j = 1,2,3$ we first need to show that $E_j(\rho,\underline{Q},\hat{p};\hat{P}) > 0$. The proof of this inequality for $j = 1$ is based on applying Holder's inequality [see (17)] for $f = \hat{p}(y|x_1,x_2)^{\frac{1}{1+\rho}}$, $\alpha = 1+\rho$, $g = 1$, $\beta = (1+\rho)/\rho$ and $d\mu = dQ_1(x_1)$. For $j = 2$ we only need to change $d\mu$ to $d\mu = dQ_2(x_2)$, whereas for $j = 3$ we need to set $d\mu = dQ_1(x_1)dQ_2(x_2)$ and use double integrals instead of single integrals in (17). Again the inequalities are strict unless the channel inputs and the channel output are independent. Finally, we use the left-hand side inequality in (14) to prove the desired result.

We can now proceed to the final stage of the proof of Theorem

First, because of (13) $R_j < I_j(\underline{Q},\hat{p};\hat{P})$ implies that $R_j < I_j(\underline{Q},\hat{p};P)$ for all P in $\mathbf{P}_v$. Then Theorem 1 applied for $\tilde{p} = \hat{p}$, implies that, for the ensemble of random block codes of rates (R_1,R_2) and length n described there

the average probability of decoding error converges to zero exponentially with increasing n. Since this is true for all P in the class under consideration, the <u>sufficiency</u> of condition (16) is established. To prove its <u>necessity</u>, notice that according to the converse of the usual capacity theorem for DM-MAC's: if a pair of rates (R_1,R_2) lies outside the region determined by (16) as $\underline{Q} = (Q_1,Q_2)$ ranges over all possible probability measures on $X_1 \times X_2$, then the asymptotically good performance is violated for the channel determined by $\hat{P}$, which is a member of the aforementioned class. This completes the proof of Theorem 3.

At this point we discuss the choice of the <u>operating point</u>, that is of a triple of the form $(\underline{\rho},\underline{Q},\tilde{p})$, where $\underline{\rho}$ is vector parameter in $[0,1]^3$ involved in the minimization of the error probability, $\underline{Q}$ is the probability measure

on the input alphabet $X_1 \times X_2$, and $\tilde{p}$ is involved in the ML decoding at the receiver. This choice depends on the main objective of our optimization. If our main objective is to operate at the maximum transmission rates [near the boundary of the region determined by (16)], then the operating point should be $(\hat{\underline{\rho}},\hat{\underline{Q}},\hat{p})$ where $\hat{\underline{Q}} = (\hat{Q}_1,\hat{Q}_2)$ is the pair of pdf's which achieves a particular point $(\hat{R}_1,\hat{R}_2) = (I_1(\hat{\underline{Q}},\hat{p};\hat{P}),\ I_2(\hat{\underline{Q}},\hat{p};\hat{P}))$ on the boundary of the achievable region and $\hat{\underline{\rho}} = \arg\min_{\rho} P_n(\underline{\rho},\hat{\underline{Q}},\hat{p};\hat{P})$. However, if our main objective is to minimize the error probability, then $(\tilde{\underline{\rho}},\tilde{\underline{Q}},\hat{p})$ ($\tilde{\underline{\rho}}$ and $\tilde{\underline{q}}$ as defined in Theorem 3) should be the operating point and the rates of transmission (R_1,R_2) should lie inside the region determined by

$R_j < I_j(\tilde{\underline{Q}},\hat{p};\hat{P})$ $j = 1,2,3$ $(R_3 = R_1 + R_2)$ instead of that determined by (16).

As a final comment for Theorem 3, notice that, under mild continuity requirements on the convex functions $P_n(\underline{\cdot},\hat{\underline{Q}},\hat{p};\hat{P})$ and $P_n(\underline{\cdot},\underline{\cdot},\hat{p};\hat{P})$ and their derivatives, the minima involved [the minimizing arguments are $\hat{\underline{\rho}}$ and $(\tilde{\underline{\rho}},\tilde{\underline{q}})$, respectively] exist.

A similar result holds for two-user tree codes:

<u>Theorem 4</u> : Under the assumptions of Theorem 3 and for any pair of rates (R_1,R_2) which lies inside the region determined by (16), the probability of decoding error for the ensemble of two-user random tree codes of rates (R_1,R_2) and constraint length K(which is described in Theorem 2 when applied for $p=\hat{p}$) converges to zero exponentially with increasing K for all channels in the class. Furthermore, if we define $(\underline{\rho}'\underline{Q}') = \arg\min_{(\underline{\rho},\underline{Q})} P_K(\underline{\rho},\underline{Q},\hat{p};\hat{P})$ where

$0 \leq \rho_j \leq \min\{E_j(\rho_j,\underline{Q},\hat{p};\hat{P})/R_j,1\}$ for $j = 1,2$

$$0 \leq \rho_3 \leq \min\{E_1(\rho_3,\underline{Q},\hat{p};\hat{P})/R_1, E_2(\rho_3,\underline{Q},\hat{p};\hat{P})/R_2, E_3(\rho_3,\underline{Q},\hat{p};\hat{P})/(R_1+R_2), 1\}, \tag{22}$$

then the operating point $(\underline{\rho}',\underline{Q}',\hat{p})$ and the channel determined by $\hat{P}$ form a saddle point for $\min_{(\underline{\rho},\underline{Q},\tilde{p})} \max_{P} P_K(\underline{\rho},\underline{Q},\tilde{p};P)$; i.e., the following inequalities hold for all P in $\mathbf{P}_V$:

$$P_K(\underline{\rho}',\underline{Q}',\hat{p};P) \leq P_K(\underline{\rho}',\underline{Q}',\hat{p};\hat{P}) \leq P_K(\underline{\rho},\underline{Q},p;\hat{P}) \tag{23}$$

Proof: We first prove the inequalities in (23). The left-hand side inequality in (23) is a result of the left-hand side inequality in (14) applied for $\underline{\rho} = \underline{\rho}'$ and the fact that P_K is a decreasing function of the E_j's for j = 1,2,3. Then the right-hand side inequality in (23) is true because of the definition of $(\underline{\rho}',\underline{Q}')$ above, the right-hand side inequality in (14) and the fact that P_K is a decreasing function of the E_j's, j=1,2,3.

The positivity requirements on $I_j(\underline{Q},\hat{p};P)$ and $E_j(\rho,\underline{Q},\hat{p};P)$ j = 1,2,3 which are necessary for the validity of Theorem 2 are the same as those for Theorem 1 and are satisfied as shown during the proof of Theorem 3. To complete the proof of Theorem 4 notice that because of the left-hand inequality in (13) the rate region determined by (16) lies inside the rate region determined by $R_j < I_j(\underline{Q},\hat{p};P)$ j = 1,2,3 for all P in the uncertainty class considered. Furthermore, because of (14), any $\underline{\rho} = (\rho_1,\rho_2,\rho_3)$ which satisfies the conditions (22) also satisfies these conditions when $E_j(\rho,\underline{Q},\hat{p};\hat{P})$ is replaced by $E_j(\rho,\underline{Q},\hat{p};P)$ for j = 1,2,3. Consequently all the assumptions of Theorem 2 are satisfied and Theorem 2 applied for $\tilde{p} = \hat{p}$ implies that for the ensemble of two-user random tree codes of rates (R_1,R_2)

in the region described by (16) and constraint length K the average probability of decoding error converges to zero exponentially with increasing K. Since this is true for all $\mathbf{P}$ in the uncertainty class, the proof is complete.

As discussed at the end of the proof of Theorem 3 the choice of the operating point depends on the objective of optimization. For tree codes a similar choice should be made and we omit the details.

III. ROBUST CODING FOR STATIONARY GAUSSIAN MULTIPLE-ACCESS CHANNELS

A. Spectral Uncertainty Classes Generated by Choquet Capacities

Suppose that $X_1 = X_2 = Y = (-\infty, \infty)$ for the input and output alphabets and the discrete-time stationary Gaussian multiple-access channel (SG-MAC) is characterized by the probability transition matrix $p^{(n)}(\underline{y}|\underline{x}_1, \underline{x}_2)$ for $\underline{x}_1 \in X_1^n$ $, \underline{x}_2 \in X_2^n$, $\underline{y} \in Y^n$ given by

$$p^{(n)}(\underline{y}|\underline{x}_1,\underline{x}_2) = (2\pi)^{-n/2}|\underline{R}^{(n)}|^{-1/2}\exp\{-1/2(\underline{y}-\underline{x}_1-\underline{x}_2)^T[\underline{\Sigma}^{(n)}]^{-1}(\underline{y}-\underline{x}_1-\underline{x}_2)\}. \quad (24)$$

In (24) $|\underline{A}|$ denotes the determinant of the matrix $\underline{A}$ and the matrix $\underline{\Sigma}^{(n)}$ is a correlatiom matrix of order n (which because of the stationarity is a symmetric Toeplitz matrix) associated with the spectral density $\phi(\omega)$, $\omega \in [-\pi, \pi] = \Omega$

Suppose the spectral density ϕ is the R-N derivative of a spectral measure Φ defined on sets $A \in \mathbf{B}$ (where $\mathbf{B}$ is the σ-algebra generated by subsets of $\Omega = [-\pi, \pi]$) with respect to the Lebesgue measure on Ω. The spectral measure Φ is only known to lie in the convex class $\mathbf{\Phi}_v$ defined by

$$\mathbf{\Phi}_v = \{\Phi \in \mathbf{\Phi} \mid \Phi(A) \leq v(A),\ \forall\ A \in \mathbf{B},\ \Phi(\Omega) = v(\Omega)\}. \quad (25)$$

In (25) $\mathbf{\Phi}$ is the class of all spectral measures on $(\Omega, \mathbf{B})$. We impose on the spectral measures Φ the additional constraint $\Phi([-\pi,\pi]) = v([-\pi,\pi]) = 2\pi\sigma^2$, which is a fixed noise power constraint and transforms the normalized spectral measures $\Phi(A)/(2\pi\sigma^2)$ into probability measures; this is necessary for the validity of the Huber-Strassen theory of least favorability.

All the results about Choquet capacities and uncertainty classes of probability measures presented in Section II.A are also valid for the spectral uncertainty classes. Let $\hat{\phi}$ and $\hat{\Phi}$ denote the Huber-Strassen derivative and the least-favorable spectral measure in Φ_V.

We will also assume that $\hat{\Phi}$ is absolutely continuous with respect to λ (i.e., $\hat{\Phi} << \lambda$). This is not so restrictive because as we can show for the total-variation spectral class defined by

$$\phi_V = \{\Phi \mid |\Phi_0 - \Phi| < \varepsilon\}$$

where ε in [0,1] is known, assuming the known nominal spectral measure Φ_0 to satisfy $\Phi_0 << \lambda$, implies that $\hat{\Phi} << \lambda$, as well. Similar conditions on the nominal spectral measures of the contaminated mixture class [12] and the band class [13] result in $\hat{\Phi}$ being absolutely continuous with respect to λ.

B. Mismatch Coding Theorems for Two-User Block and Tree Codes

It is assumed that the channel inputs satisfy average input energy constraints of the form

$$E\{||\underline{X}_j||^2\} = nE_j, \ j = 1,2 \tag{26}$$

where $||\ ||$ is the Euclidean norm of the n-dimensional random vector $\underline{X}_j$ and E_j is the input energy per channel use for user j.

Suppose that in the presence of uncertainty about $p^{(n)}(\underline{y}|\underline{x}_1,\underline{x}_2)$ (induced by the spectral density ϕ) the user mistankenly assumes that $\tilde{p}^{(n)}(\underline{y}|\underline{x}_1,\underline{x}_2)$ (induced by $\tilde{\phi}$) is the n-th order probability transition matrix governing the statistics of the SG-MAC. Let Φ and $\tilde{\Phi}$ denote spectral measures for which $\phi = d\Phi/d\lambda$ and $\tilde{\phi} = d\tilde{\Phi}/d\lambda$, respectively.

The above situation is characterized by mismatch as in the case desribed in Theorem 1. Therefore we can apply Theorem 1 to this special case. We will start with the evaluation of the mismatch mutual information and the mismatch error exponent functions for the new case.

In the case of discrete-time SG-MAC we have to deal with triplets of n-tuples $(\underline{x}_1,\underline{x}_2,\underline{y})$ whose components (x_{1i},x_{2i},y_i) and (x_{1j},x_{2j},y_j) may be strongly correlated. It is advantageous to follow the technique of [17, Section 4.5.2] and make the problem equivalent to that of n parallel independent additive Gaussian noise (AGN) channels. This involves a unitary transformation of $\underline{x}_1$, $\underline{x}_2$, and $\underline{y}$ associated with $\underline{\Sigma}^{(n)}$ which preserves the mutual information relationships and the average input energy constraints.

The variance σ_i^2 of the Gaussian noise of the i-th channel is the i-th eigenvalue of the Toeplitz matrix $\underset{\sim}{\Sigma}^{(n)}$. Furthermore the initial average input constraints (26) become [18, Section 7.5]

$$\frac{1}{n}\sum_{i=1}^{n} E_{ji} = E_j \qquad\qquad j = 1,2 \tag{27}$$

where the j-th user's input to the i-th channel is a zero-mean Gaussian random variable with variance E_{ji}. Let $E_{3i} = E_{1i} + E_{2i}$ for $i = 1,2,\ldots,n$.

Once the SG-MAC has been decomposed to n parallel AGN DM-MAC's we can apply the theory of parallel AGN channels (see [18, Section 7.5]), (24), and the definitions (9) and (11) of Theorem 1a. The asymptotic (in the limit of large n) mismatch Liao functions and the asymptotic mismatch error exponents take the form

$$I_j(r_j,\tilde{\phi};\phi) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left[1/2\ln\left(1+\frac{E_{ji}}{\tilde{\sigma}_i^2}\right) + 1/2\frac{E_{ji}}{E_{ji}+\tilde{\sigma}_i^2}\left(1-\frac{\sigma_i^2}{\tilde{\sigma}_i^2}\right)\right] \tag{28}$$

and

$$E_j(\rho,r_j,\tilde{\phi};\phi) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\rho-1}{2}\ln\left[1+\frac{E_{ji}}{(1+\rho)\tilde{\sigma}_i^2}\right]\right.$$

$$\left. + 1/2\ln\left[1+\frac{E_{ji}}{(1+\rho)\tilde{\sigma}_i^2}\left(1+\rho-\rho\frac{\sigma_i^2}{\tilde{\sigma}_i^2}\right)\right]\right\} \tag{29}$$

for $j = 1,2,3$. In (28) - (29) σ_i^2, $\hat{\sigma}_i^2$, and E_{ji} $j = 1,2,3$ for $i = 1,2,\ldots, n$ are the eigenvalues of the n-th order Toeplitz matrices induced by the spectral densities ϕ, $\tilde{\phi}$, and r_j $(j = 1,2,3)$, respectively. The eigenvalues

E_{ji} $j = 1,2$ satisfy the average input energy constraints (27) as discussed above and $E_{3i} = E_{1i} + E_{2i}$. Then $q_j^{(n)}(\underline{x})$ is the n-th order probability density function (pdf) on the input alphabet X_j^n induced by the spectral density r_j $j = 1,2$. Consequently, by taking the limits as $n \to \infty$ in (27), (28), and (29) and using the discrete-time version of the Toeplitz Distribution Theorem [19] we obtain:

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} r_j(\omega)\lambda(d\omega) = E_j \tag{30}$$

$$I(r_j,\tilde{\phi};\phi) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{\ln\left[1+ \frac{r_j(\omega)}{\tilde{\phi}(\omega)}\right] + \frac{r_j(\omega)}{r_j(\omega)+\tilde{\phi}(\omega)}\left[1- \frac{\phi(\omega)}{\tilde{\phi}(\omega)}\right]\right\}\lambda(d\omega), \tag{31}$$

$$E_j(\rho,r_j,\tilde{\phi};\phi) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{(\rho-1)\ln\left[1+ \frac{r_j(\omega)}{(1+\rho)\tilde{\phi}(\omega)}\right]\right.$$

$$\left. + \ln\left[1+ \frac{r_j(\omega)}{(1+\rho)\tilde{\phi}(\omega)}\left(1+\rho-\rho\frac{\phi(\omega)}{\tilde{\phi}(\omega)}\right)\right]\right\}\lambda(d\omega). \tag{32}$$

for $j = 1,2,3$.

Next we consider the input spectral densities $\tilde{r}_j$ which maximize $I_j(r_j,\tilde{\phi};\tilde{\phi})$ for $j = 1,2,3$, respectively, the asymptotic Liao functions, for the matched case ($\phi = \tilde{\phi}$). The spectral density r_j has been shown in [18, Section 7.5] to be defined in terms of a parameter $\tilde{\gamma}_j$ as:

$$\tilde{r}_j(\omega) = \max\{0,\tilde{\gamma}_j - \tilde{\phi}(\omega)\}, \quad \omega \in [-\pi,\pi] \qquad j = 1,2,3 \tag{33a}$$

where the parameter $\tilde{\gamma}_j$ is determined by the condition:

$$E_j = \frac{1}{2\pi} \int_{\{\tilde{\phi}(\omega)<\tilde{\gamma}_j\}} [\tilde{\gamma}_j - \tilde{\phi}(\omega)]\lambda(d\omega). \qquad j=1,2,3 \tag{33b}$$

where $E_3 = E_1 + E_2$.

We can now state the main result of this Section which follows from Theoren 1 when it is applied to SG-MAC's:

Theorem 5 : Consider a two-user discrete-time stationary additive Gaussian channel with independent inputs and n-th order probability transition matrix $p^{(n)}(\underline{y}|\underline{x}_1,\underline{x}_2)$ induced by spectral density $\phi(\omega)$, $\omega \in [-\pi,\pi]$. User j employs the n-th order input pdf $\tilde{q}_j^{(n)}(\underline{x})$ induced by the spectral density $\tilde{r}_j$ satisfying (33a)-(33b) for j = 1,2 and the decoder employs inaccurate ML decoding based on $\tilde{p}^{(n)}(\underline{y}|\underline{x}_1,\underline{x}_2)$ (induced by the spectral density $\tilde{\phi}$) instead of the true $p^{(n)}(\underline{y}|\underline{x}_1,\underline{x}_2)$. Consider the ensemble of pairs of random block codes of length n and rates (R_1,R_2) whose codewords are chosen independently with equal probability and the n letters of each codeword are chosen from the input alphabet X_j according to $\tilde{q}_j^{(n)}(\underline{x})$. Then, if the rates R_1, R_2 satisfy

$$R_j < I_j(\tilde{\gamma}_j,\tilde{\phi};\phi), \ j = 1,2,3 \tag{34}$$

where,

$$I_j(\tilde{\gamma}_j,\tilde{\phi};\phi) = \frac{1}{4\pi} \int_{\{\tilde{\phi}(\omega)<\tilde{\gamma}_j\}} \left\{ \ln\frac{\tilde{\gamma}_j}{\tilde{\phi}(\omega)} + \frac{1}{\tilde{\gamma}_j}\left[\frac{\tilde{\gamma}_j}{\tilde{\phi}(\omega)} - 1\right]\left[\tilde{\phi}(\omega)-\phi(\omega)\right]\right\}\lambda(d\omega), \tag{35}$$

then for large n the average error probability of decoding error P_E is upperbounded by

$$P_E < \sum_{j=1}^{3} \exp\{-n[E_j(\rho_j, \tilde{\gamma}_j, \tilde{\phi};\phi) - \rho_j R_j]\} \tag{36}$$

where for ρ in [0,1]

$$E_j(\rho, \tilde{\gamma}_j, \tilde{\phi};\phi) = \frac{1}{4\pi} \int_{\{\tilde{\phi}(\omega)<\tilde{\gamma}_j\}} \Big\{(\rho-1)\ln\Big[1+\frac{1}{1+\rho}\Big[\frac{\tilde{\gamma}_j}{\tilde{\phi}(\omega)} -1\Big]\Big]$$

$$+ \ln\Big[1+\frac{1}{1+\rho}\Big[\frac{\tilde{\gamma}_j}{\tilde{\phi}(\omega)}-1\Big]\Big[1+\rho-\rho\frac{\phi(\omega)}{\tilde{\phi}(\omega)}\Big]\Big]\Big\}\lambda(d\omega). \tag{37}$$

For the validity of this Theorem it is required that for all ϕ on Ω the mismatch Liao functions satisfy $I_j(\tilde{\gamma}_j, \tilde{\phi};\phi) > 0$ and the mismatch error exponents satisfy $E_j(\rho, \tilde{\gamma}_j, \tilde{\phi};\phi) > 0$ for all ρ in [0.1]. These positivity requirements are satisfied for the choice of $\tilde{\phi}$ and $\tilde{\gamma}_j$ in Theorem 7 below.

Remark 5. The mismatch error exponents in (37) have been evaluated for an input spectral density $\tilde{r}_j$ [given by (33a)-(33b)] which maximize the mutual information functions $I_j(\tilde{r}_j, \tilde{\phi}:\phi)$ of (31). If the objective is to maximize the error exponents $E_j(\rho, \tilde{r}_j, \tilde{\phi};\phi)$ of (32) (and thus minimize the bound on P_E) the appropriate input spectral densities $\tilde{r}_{j\rho}$ is given by

$$\tilde{r}_{j\rho}(\omega) = (1+\rho)\max\{0, \tilde{\gamma}_{j\rho} - \tilde{\phi}(\omega)\}, \quad \omega \in [-\pi, \pi], \tag{38a}$$

$$E = \frac{1+\rho}{4\pi} \int_{\{\tilde{\phi}(\omega)<\tilde{\gamma}_{j\rho}\}} [\tilde{\gamma}_{j\rho} - \tilde{\phi}(\omega)]\lambda(d\omega), \tag{38b}$$

and the mismatch error exponents become:

$$E_j(\rho,\tilde{\gamma}_{j\rho},\tilde{\phi};\phi) = \frac{1}{4\pi}\int_{\{\tilde{\phi}(\omega)<\tilde{\gamma}_{j\rho}\}}\{(\rho-1)\ln\frac{\tilde{\gamma}_{j\rho}}{\tilde{\phi}(\omega)}+\ln[1+[\frac{\tilde{\gamma}_{j\rho}}{\tilde{\phi}(\omega)}-1][1+\rho-\rho\frac{\phi(\omega)}{\tilde{\phi}(\omega)}]]\}\lambda(d\omega) \tag{39}$$

For two-user tree codes of rates (R_1,R_2) where $R_j = \frac{1}{N}\log_2 M_j$ bits per channel symbol with Viterbi decoding and a ML test based on $\tilde{p}^{(n)}(\underline{\cdot}|\underline{\cdot},\underline{\cdot})$ [the overall equivalent block length in channel input symbols is now $n=(L+K-1)N$, which corresponds to $L\log_2 M_j$ input bits of user j, j = 1,2] a similar result holds:

Theorem 6 : Under the assumptions of Theorem 5 suppose a tree code for user j (j = 1,2) has constraint length K and rate $R_j = \frac{1}{N}\log_2 M_j$ bits per channel symbol satisfying (34), and consider the ensemble of pairs of random codes generated as described in Theorem 2. Then the average bit error probability of the Viterbi decoder P_b is for large K upperbounded by $P_K(\underline{\rho},\tilde{\underline{\gamma}},\tilde{\phi};\phi)$ which can be obtained from (12), if we replace $E_j(\rho,\underline{Q},\tilde{p};P)$ with $E_j(\rho,r_j,\phi;\phi)$ for j = 1,2,3. The parameters ρ_j must satisfy the same conditions as for Theorem 2 provided that we replace the error exponents E_j

used there with the exponents defined in (37). The same positivity requirements on I_j and E_j as these for Theorem 5 should be satisfied.

The proof of Theorem 6 is a straightforward extension of Theorem 2 to the SG-MAC case and will be omitted. The same technique of decomposing to n parallel AGN MAC's may be applied.

C. Minimax Coding Theorems for Two-User Block and Tree Codes

Next we assume that the spectral density ϕ which induces the transition probability matrix $p^{(n)}(\underline{y}|\underline{x}_1,\underline{x}_2)$ is given by $\phi = d\Phi/d\lambda$ where Φ belongs to a class of the form (25) described in Section III.A. The channel decoder employs a ML decoding rule based on $\tilde{p}^{(n)}(\cdot|\cdot)$ (induced by a spectral density $\tilde{\phi}$) in a way desribed in Theorems 5 and 6 . The goal is to choose $\tilde{\phi}$ so that the asymptotic convergence of the probability of decoding error is guaranteed for all channels in the class. This is accomplished with the following result:

Theorem 7 : Suppose the spectral measure Φ [where $\phi = d\Phi/d\lambda$ induces $p^{(n)}(\underline{y}|\underline{x}_1,\underline{x}_2)$] belongs to a class of the form (25) and $\hat{\Phi}$ is the element of the class singled out by Lemma 1 which also satisfies $\hat{\Phi}<<\lambda$. Suppose further that j-th encoder employs an input pdf $\hat{q}_j^{(n)}(\underline{x})$ induced by $\hat{r}_j$ defined by (33a)-(33b) for $\tilde{\phi} = \hat{\phi}$ and $\tilde{\gamma} = \hat{\gamma}$ and the decoder's ML decoding rule is based on $\hat{p}^{(n)}(\cdot|\cdot,\cdot)$ induced t $\hat{\phi}$ where $\hat{\phi} = d\hat{\Phi}/d\lambda$. Then the following inequalities are satisfied for all ϕ with Φ in Φ_V :

$$I_j(\hat{\gamma}_j,\hat{\phi};\phi) \geq I_j(\hat{\gamma}_j,\hat{\phi};\hat{\phi}) \geq I_j(\gamma_j,\phi;\hat{\phi}),\ j = 1,2,3 \tag{40}$$

[i.e., the operating point $(\hat{\gamma}_j,\hat{\phi})$ and the channel determined by $\hat{\phi}$ form a saddle point for $\max\limits_{(\gamma_j,\tilde{\phi})} \min\limits_{\phi} I_j(\gamma_j,\tilde{\phi};\phi)$], and

$$E_j(\rho,\hat{\gamma}_j,\hat{\phi};\phi) \geq E_j(\rho,\hat{\gamma}_j,\hat{\phi};\hat{\phi}) \geq E_j(\rho,\hat{\gamma}_j,\phi;\hat{\phi}) \tag{41}$$

for all ρ in [0,1]. Furthermore, the conditions

$$R_j < I_j(\hat{\gamma}_j,\hat{\phi};\hat{\phi}),\ j = 1,2,3 \tag{42}$$

are sufficient and necessary to guarantee that for the ensemble of pairs of random block codes of length n and rates (R_1,R_2) desribed in Theorem 5 (when applied for $\tilde{\phi} = \hat{\phi}$) the average probability of decoding error converges to zero exponentially with increasing n for all channels in the class.

Remark 6. The inequalities $R_j < I_j(\hat{\gamma}_j,\hat{\phi};\hat{\phi})$ determine the channel capacity region of the class described by (25) where the Liao functions are given by

$$I_j(\hat{\gamma}_j,\hat{\phi};\hat{\phi}) = \frac{1}{4\pi}\int_{\{\hat{\phi}(\omega)<\hat{\gamma}_j\}} \ln\frac{\hat{\gamma}_j}{\hat{\phi}(\omega)}\lambda(d\omega). \tag{43}$$

Similarly the quantities $E_j(\rho,\hat{\gamma}_{j\rho},\hat{\phi};\hat{\phi})$ obtained from (39) and (38a)-(38b) for $\tilde{\phi} = \hat{\phi}$ and given by

$$E_j(\rho,\hat{\gamma}_{j\rho},\hat{\phi};\hat{\phi}) = \frac{\rho}{4\pi}\int_{\{\hat{\phi}(\omega)<\hat{\gamma}_{j\rho}\}} \ln\frac{\hat{\gamma}_{j\rho}}{\hat{\phi}(\omega)}\ \lambda(d\omega) \tag{45}$$

represent the error exponents of the class; for $\rho = .5$ the inequalities $R_j < E_j(\rho,\hat{\gamma}_{j\rho},\hat{\phi};\hat{\phi})$ determine the cutoff rate region. Notice that the boundaries of both regions are expressed in terms of the Huber-Strassen derivative $\hat{\phi} = d\hat{\Phi}/d\lambda$ which characterizes the worst-case (least-favorable) channel.

Proof: The sequence of steps is similar to that for the proof of Theorem 3 but the individual steps differ. We first prove the inequalities (40) and (41). To prove the right-hand side inequality in (40) we first prove the inequalities $I_j(\hat{r}_j,\hat{\phi};\hat{\phi}) \geq I_j(r_j,\hat{\phi};\hat{\phi}) \geq I_j(r_j,\phi;\hat{\phi})$ for all input spectral densities which satisfy (30) and $j = 1,2,3$. The first part of these inequalities follows from the definition of $\hat{r}_j$ [see (33a)-(33b)]. The second part of the inequalities can be proved by considering the difference $I_j(r_j,\hat{\phi};\hat{\phi}) - I_j(r_j,\phi;\hat{\phi})$, gather the logarithmic terms together, apply the inequality $\ln x \geq 1-x^{-1}$ and show that the above difference is nonnegative. Then, since the inequalities above are valid for all r_j satisfying (30), we can apply them to the case $r_j = \hat{r}_j$ (related to $\hat{\gamma}_j$) to obtain that $I_j(\hat{\gamma}_j,\hat{\phi};\hat{\phi}) \geq I_j(\gamma_j,\hat{\phi};\hat{\phi}) \geq I_j(\gamma_j,\phi;\hat{\phi})$ where γ_j is the parameter satisfying (33b) for $\tilde{\phi} = \phi$.

To prove the left-hand side inequality in (40) we use Lemma 1, a second lemma (Lemma 2 below) and the following fact (See eq. (46) of [9] and the justification which follows). If $g(u) \geq 0$ for all $u \in A$, $u \in \mathbf{B}$, then

$$\int_A g\,d\Phi' << \int_A g\,d\Phi \tag{45}$$

for any spectral measure Φ with Lebesgue decomposition $\Phi = \Phi' + \Phi''$ where $\Phi' << \lambda$ and $\Phi'' \perp \lambda$ (i.e., Φ singular with respect to λ).

Lemma 2: Let g be a continuous decreasing function on the real line, let X be a continuous real random variable, and let P be a probability measure on the σ-algebra generated by the subsets of the real line. Then, the following relationship holds for all a and b with a < b:

$$\int_{[a,b]} g(X)dP = -\int_a^b P\{X \leq t\}g'(t)dt + g(b)P\{X \leq b\} - g(a)P\{X \leq a\}. \quad (46)$$

The proof of Lemma 2 is provided in [?] (equations (47a)-(47b)) and will not be included here.

Then we prove the left-hand side inequalities in (40) as follows. First, we notice that $d\hat{\Phi} = \hat{\phi}d\lambda$. Then we define $g_j(u) = (\hat{\gamma}_j/u-1)/\hat{\gamma}_j$ for $u \leq \hat{\gamma}_j$, j=1,2,3 and apply (45) to obtain that for the desired inequality to hold it suffices that

(continue with next page)

$$\int_{\{\hat{\phi}<\hat{\gamma}_j\}} g_j(\hat{\phi})d\Phi \leq \int_{\{\hat{\phi}<\hat{\gamma}_j\}} g_j(\hat{\phi})d\hat{\Phi}, \quad j = 1,2,3 \tag{47}$$

Since g_j is a decreasing function with $g_j(\hat{\gamma}_j)=0$ and $P\{\hat{\phi}<0\}=0$ we can apply Lemma 2 for $a=0$, $b=\hat{\gamma}_j$ twice, once for $P=\Phi$ and once for $P=\hat{\Phi}$, and then use the fact that $\hat{\Phi}$ makes $\hat{\phi}$ stochastically smallest over all Φ in Φ_V (Lemma 1) to show that (47) is satisfied.

To prove the right-hand side inequalities in (41) we follow a procedure similar to that for proving the right-hand side inequalities in (40). To prove the left-hand side inequalities in (41) we define the functions $h_j(\alpha) = E_j(\rho,\hat{\gamma}_j,\hat{\phi};(1-\alpha)\hat{\phi}+\alpha\phi)$ for α in $[0,1]$ and $j = 1,2,3$. Then, since h_j are convex functions of α the desired inequalities which can be written as $h_j(1) \geq h_j(0)$ become equivalent to $\partial h_j(\alpha)/\partial\alpha|_{\alpha=0} \geq 0$. After we evaluate the directional derivative $\partial h_j(\alpha)/\partial\alpha$ at $\alpha=0$ and apply (45) for the function $f_j(u)= \rho(\hat{\gamma}_j/u-1)/(\rho u+\hat{\gamma}_j)$ $u \leq \hat{\gamma}_j$, the desired inequalities hold if

$$\int_{\{\hat{\phi}<\hat{\gamma}_j\}} f_j(\hat{\phi})d\Phi \leq \int_{\{\hat{\phi}<\hat{\gamma}_j\}} f_j(\hat{\phi})d\hat{\Phi}. \tag{48}$$

Finally, as for the proof of (47), we can use Lemma 2 for the decreasing functions f_j [for which $f_j(\hat{\gamma})=0$] twice and Lemma 1 to show that (48) is satisfied for all Φ in Φ_V.

To establish the positivity of $I_j(\hat{\gamma}_j,\hat{\phi};\phi)$ and $E_j(\rho,\hat{\gamma}_j,\hat{\phi};\phi)$ for all $\phi = d\Phi/d\lambda$ with Φ in $\mathbf{\Phi_V}$ and ρ in $[0,1]$ we use thr right-hand side inequalities in (40) and (41), respectively, and the fact that both $I_j(\hat{\gamma}_j,\hat{\phi};\hat{\phi})$ as defined by (43) and $E_j(\rho,\hat{\gamma}_j,\hat{\phi};\hat{\phi})$ defined by

$$E_j(\rho,\hat{\gamma}_j,\hat{\phi};\hat{\phi}) = \int_{\{\hat{\phi}(\omega)<\hat{\gamma}_j\}} \ln[1+\frac{1}{1+\rho}[\frac{\hat{\gamma}_j}{\hat{\phi}(\omega)} -1]]\lambda(d\omega). \tag{49}$$

are strictly positive for $j = 1,2,3$ because of their definitions.

To complete the proof of Theorem 7 we notice that because of (40) $R_j < I_j(\hat{\gamma}_j,\hat{\phi};\hat{\phi})$ implies that $R_j < I_j(\hat{\gamma}_j,\hat{\phi};\phi)$ for all $\phi = d\Phi/d\lambda$ with Φ in $\mathbf{\Phi_V}$. Then from Theorem 5 applied for $\tilde{\phi} = \hat{\phi}$ and $\tilde{\gamma} = \hat{\gamma}$ it follows that for the ensemble of pairs of random block codes of rates (R_1,R_2) and length n described there the average probability of decoding error converges to zero exponentially with increasing n. Since this is true for all Φ in the class under consideration, the sufficiency of condition (42) is established. To prove its necessity, notice that according to the converse of the usual capacity region theorem for DM-MAC the violation of any of the coditions (42) implies that the average probability of decoding error converges to 1 exponentially for the channel determined by $\hat{\phi}$, which is a member of the aforementioned class.

The discussion for the choice of the operating point is similar to that which followed the proof of Theorem 3 and we do not repeat it here.

The corresponding result for two-user tree codes is:

Theorem 8 : Under the assumptions of Theorem 4a condition (42) guarantees that for the ensemble of pairs of random tree codes of constraint length K and rates (R_1, R_2) (described in Theorem 6 as applied for $\tilde{\phi} = \hat{\phi}$ and $\tilde{\Upsilon} = \hat{\Upsilon}$) the average probability of decoding error converges to zero exponentially with increasing K for all channels in the class.

For the proof of this Theorem one can follow the same steps as for the proof of Theorem 4 and use the various definitions involved in Theorems 5, 6, and 7. The proof is therefore omitted.

It should be noted that all the results of this section can be extended to continuous-time stationary additive Gaussian bandlimited (e.g., with spectral densities defined on $\Omega = [-\omega_0, \omega_0]$) channels. Since Huber-Strassen derivatives of capacities with respect to σ-finite (and not finite) measures can be constructed [20], these results can possibly be extended to nonbandlimited [e.g., $\Omega = (-\infty, \infty)$] channels provided that the definition of π_v is appropriately modified. However, several of the most useful examples of capacity classes (e.g., the ε-mixtures and variation neighborhoods) are not capacities when Ω is not compact.

IV. SUMMARY AND CONCLUSIONS

We have addressed the problem of minimax robust coding for multiple-access channels with uncertainty in their statistical description. For uncertainty classes determined by Choquet 2-alternating capacities coding theorems were proved for discrete memoryless channels with uncertainty in the probability transition matrices, and for stationary additive Gaussian channels with spectral uncertainty. It was established that for the ensembles of pairs of random block codes and random tree codes the average error probability of the decoder converges to zero exponentially with increasing block length or constraint length, respectively, for all two-user channels in the class, provided that the decoder employs a suitable robust maximum-likelihood decodidng rule and the code rates lie inside a critical region. The channel capacity region and the cut-off rate region for the class of channels were evaluated. The boundaries of these regions , a well as the aforementioned robust maximum-likelihood decoding rule are characterized in terms of a Radon-Nikodym type derivative between the upper measure of the Choquet capacity class and a Lebesgue-like measure defined on the appropriate set.

REFERENCES

[1] R. Alswede, "Multi-way communication channels," in Proc. 2nd Int'l Symp. Inform Theory Thakadsor, Hungary, 1971, Publishing House of the Hungarian Acad. Sci., pp. 23-52, 1973.

[2] H.H. Liao, "Multiple-Access Channels," Ph.D. dissertation, Dept. Elec. Engr., Univ. Hawaii, Honolulu, 1972.

[3] D. Slepian and J.K. Wolf, "A coding theorem for multiple-access channels," Bell Syst. Tech. J., vol. 52, pp. 1037-1076, 1973.

[4] R.L. Peterson and D.J. Costello Jr., "Error probability and free distance bounds for two-user tree codes on multiple-access channels," IEEE Trans. Inform Theory, vol. IT-26, pp. 658-670, November 1980.

[5] J.H. Jahn, "Coding of arbitrarily varying multiuser channels," IEEE Trans. Inform. Theory, vol. IT-27, pp. 212, March 1981.

[6] M. Ohkubo, Universal coding for multiple access channels," IEEE Trans. Inform Theory, vol. IT-27, pp. 728-734, November 1981.

[7] I. Csiszar, J. Korner, and K. Marton, "A new look at the error exponent of discrete memoryless channels". Preprint. Presented at the IEEE Int'l Symp. Inform Theory, October 1977, Cornell Univ., Ithaca, N.Y. See also Chapter 2 5 of Information Theory: Coding Theorems for Discrete Memoryless Systems by I. Csiszar and J. Korner, Academic Press, 1982.

[8] I.G. Stiglitz, "Coding for a class of unknown channels," IEEE Trans. Inform. Theory, vol. IT-12, pp. 189-195, April 1966.

[9] E.A. Geraniotis, "Minimax robust coding for channels with uncertain statistics," accepted for publication in the IEEE Trans. Inform. Theory, to appear in November 1985.

[10] D. Blackwell, L. Breiman, and A.J. Thomasian, "The capacities of certain channels under random coding," Ann. Math. Statist., vol. 31, pp. 558-567, 1960.

[11] G. Choquet, "Theory of capacities," Ann. Inst. Fourier, vol. 5, pp. 131-292, 1953/54.

[12] P.J. Huber, "A robust version of the probability ratio test," Ann. Math Statistics, vol. 36, pp. 1753-1758, 1965.

[13] S.A. Kassam, "Robust hypothesis testing for bounded classes of probability densities," IEEE Trans. Inform. Theory, vol. IT-27, pp. 242-247, 1981.

[14] K.S. Vastola and H.V. Poor, "On the p-point uncertainty class," IEEE Trans. Inform. Theory, vol. IT-30, pp. 374-376,1984.

[15] P.J. Huber and V. Strassen, "Minimax tests and the Neyman-Pearson lemma for capacities," Ann. Statistics, vol. 1, pp. 251-265, 1973.

[16] H.V. Poor, "Mimimax linear smoothing for capacities", Ann. Probability, vol. 10, 1982.

[17] T. Berger, "Rate Distortion Theory: A Mathematical Basis for Data Compression, Prentice Hall, 1971.

[18] R.G. Gallager, Information Theory and Reliable Communications, Wiley, 1968.

[19] V. Grenader and G. Szego, "Toeplitz Forms and Their Applications, Univ. of California Press, 1958.

[20] K.S. Vastola, "topics in robust statistical signal processing," Report R-965 (also Ph.D. dissertation), Coordinated Science Laboratory, Univ. of Illinois, Urbana, September 1982.