Sequential Decoding for Multiple Access Channels

ERDAL ARIKAN, MEMBER, IEEE

***Abstract* —The use of sequential decoding in multiple access channels is considered. The Fano metric, which achieves all achievable rates in the one-user case, fails to do so in the multiuser case. A new metric is introduced and an inner bound is given to its achievable rate region. This inner bound region is large enough to encourage the use of sequential decoding in practice. The new metric is optimal, in the sense of achieving all achievable rates, in the case of one-user and pairwise-reversible channels. Whether the metric is optimal for all multiple access channels remains an open problem. It is worth noting that even in the one-user case, the new metric differs from the Fano metric in a nontrivial way, showing that the Fano metric is not uniquely optimal for such channels. A new and stricter criterion of achievability in sequential decoding is also introduced and examined.**

I. Introduction

We CONSIDER the application of sequential decoding to multiple access channels (MAC's). Sequential decoding is a decoding algorithm for tree codes, originally developed for channels with one transmitter and one receiver [1], [2]. MAC's are models of communication systems where a number of transmitters share a common transmission medium to transmit statistically independent messages to a common receiver. A typical example of a MAC is the up-link of a satellite channel with multiple ground stations.

The MAC model that we use is the standard information-theoretic one [3] shown in Fig. 1. The sources here generate statistically independent sequences of letters from their respective finite alphabets; source sequences are encoded independently of each other and sent over the channel. The decoder observes the channel output sequence and generates an estimate for each source sequence. The channels that we consider are discrete-time memoryless stationary channels with finite input and output alphabets. A channel with these properties is completely characterized by specifying its input alphabets $X_1,\cdots,X_n$, output alphabet Y, and transition probabilities $P(y|x_1,\cdots,x_n)$ for each $y\in Y$, $x_1\in X_1,\cdots,x_n\in X_n$. The quantity $P(y|x_1,\cdots,x_n)$ denotes the probability that y is observed at the channel output given that x_i is transmitted at input i, $i=1,\cdots,n$. We shall use the notation $(P;X_1,\cdots,X_n;Y)$ to denote such a channel.

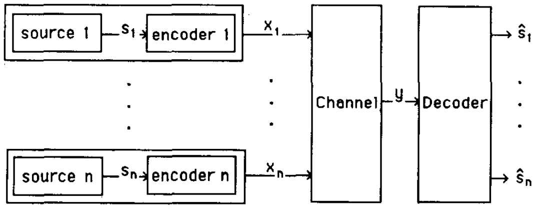

Fig. 1. Multiple access channel model.

A. General Background

To provide a framework and motivation for studying sequential decoding in MAC's, we first summarize some known results on MAC's. For a more detailed discussion of all the issues discussed in this section and a comparison of various approaches to multiple access communications, refer to the excellent survey article by Gallager [4].

The capacity region of a MAC is defined as the region of source rates at which communication with arbitrarily small probability of decoding error is possible. (The probability of decoding error here is the average decoding error assuming that all messages are equally likely.) This region was determined by Ahlswede [5] and Liao [6].

The capacity region of a MAC is typically larger than the set of rates achievable through conventional ways of using such channels, such as time-division multiplexing (TDM) and, if applicable, frequency-division multiplexing (FDM). The desire to find *practical* ways of achieving these theoretically possible higher rates is the main motivation for studying coding for MAC's. In this respect, the following results are significant.

Slepian and Wolf [7] proved that for block codes and for any rate in the capacity region, it is possible to make the probability of maximum-likelihood decoding error go to zero exponentially in the block length. This result also holds [4] for linear block codes for which the encoding complexity grows approximately linearly with the code-length. Thus the probability of decoding error can be made to approach zero exponentially in the encoding complexity. A similar result was proved by Peterson and Costello [8] for tree codes.

Manuscript received February 27, 1986; revised November 6, 1986. This work was supported in part by the Defence Advanced Research Products Agency under Contract N000 14-84-K-0357. This paper was presented in part at the IEEE International Symposium on Information Theory, Ann Arbor, MI, October 6–9, 1986.

The author is with the Department of Electrical Engineering, Bilkent University, P.K. 8, Maltepe, Ankara, 06572, Turkey.

IEEE Log Number 8820329.

0018-9448/88/0300-0246$01.00 ©1988 IEEE

The previous results indicate that a favorable trade-off is possible between the probability of decoding error and the encoding complexity. Unfortunately, these results are all based on maximum likelihood decoding, for which the decoding complexity grows exponentially in the block or constraint length. Therefore, just as in the case of single-user channels, the main obstacle to the use of coding in MAC's appears to be the complexity of decoding. We shall investigate whether sequential decoding, which is a practical decoding algorithm for single-user channels at low enough rates, can be used also in MAC's.

B. Multiuser Tree Codes and Sequential Decoding

Henceforth, we consider a system as in Fig. 1, where the encoders are tree encoders and the decoder is a sequential decoder. We make two assumptions about the system. First, we assume that the number of channel input symbols per branch is the same for each user's tree code. Second, we assume that the users are synchronized so that they start transmitting at a common point in time.[1]

Under these assumptions the tree codes in the system can be collectively represented as a single tree code, which we call a *multiuser* (or *n-user*) tree code. An example of a two-user tree code corresponding to two single-user tree codes is shown in Fig. 2. A more complete description of multiuser tree codes will be given in Section II.

Sequential decoding for MAC's consists of applying the ordinary sequential decoding algorithm to a multiuser tree code. Recall that sequential decoding is essentially a search algorithm for finding the transmitted (correct) path in a tree code. The search is guided by a metric,[2] which is a measure of correlation between paths in the code tree and the received sequence. If the code and the metric are properly chosen, the metric value tends to decrease on all paths except for the transmitted path. Thus sequential decoding is simply a search for that path on which the metric has a nonnegative drift.

The search effort in sequential decoding is a random quantity as it depends on the severity of channel noise (transmission errors). The remarkable point about sequential decoding is the possibility of making the expected value of the search effort per correctly decoded source digit independent of the codelength and hence of the desired level of reliability. This is possible, of course, only if the desired rate is low enough.

The rate of a single-user tree code with k channel input symbols per branch and degree d (i.e., d immediate descendants for each node) is defined as $(1/k)\ln(d)$ nats or $(1/k)\log_2(d)$ bits per channel use. Throughout we shall use the natural units unless otherwise stated. The rate of an n-user tree code is defined as an n-tuple $(R_1,\cdots,R_n)$, where R_i is the rate of the ith user's tree code, $i=1,\cdots,n$.

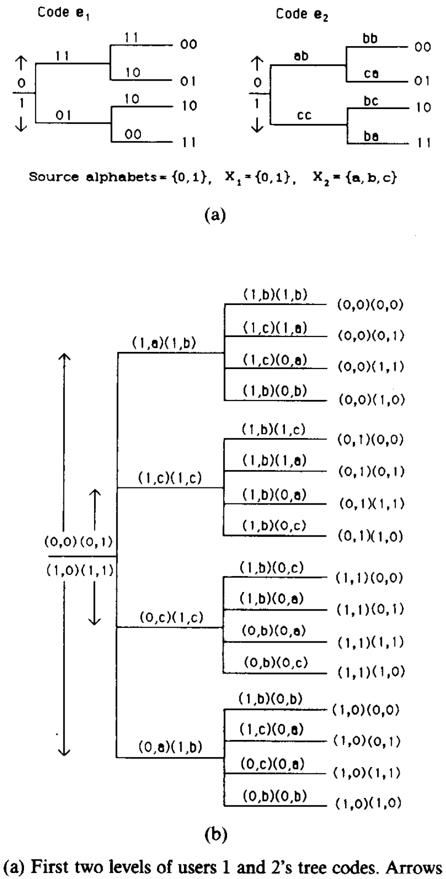

Fig. 2. (a) First two levels of users 1 and 2's tree codes. Arrows indicate mapping from source sequences to paths; e.g., on input of 0, encoder always takes upper branch from current node. (b) First two levels of two-user tree code corresponding to codes in part (a).

A point $(R_1,\cdots,R_n)$ is said to be *achievable* by sequential decoding if an infinite n-user tree code exists with rate $\geq (R_1,\cdots,R_n)$ (componentwise) and a metric such that the expected search effort per correctly decoded source digit is finite. The set of all achievable points is called the *achievable rate region* (of sequential decoding) and is denoted by $\boldsymbol{R}_{\text{comp}}$. Our goal is to find out if $\boldsymbol{R}_{\text{comp}}$ is large enough to make sequential decoding worthwhile.

Searching for the correct path in a multiuser tree code is more difficult than in a single-user tree code. In a multiuser tree code partially correct paths exist (i.e., paths that agree with the correct path in certain components) which are correlated with the correct path and hence with the channel output. Consequently, compared to a totally incorrect path, partially correct paths can be more readily mistaken for the correct path during sequential decoding, causing complications that are nonexistent in the single-user case.

As a result, the well-known Fano metric [2], which is optimal (in the sense of achieving all achievable rates) in the single-user case, fails to work satisfactorily in the multiuser case. The main contribution of this paper is to

[1] It is possible to prove the main results of this paper in a model where one allows arbitrary but bounded time shifts among the users. The first assumption is less crucial and can be dropped easily. These questions will not be addressed in this paper, however.

[2] This is not a metric in the customary mathematical sense of the word.

introduce a new metric and to prove that it works satisfactorily in a sense to be quantified later. We do not know, however, whether this metric achieves all of $\boldsymbol{R}_{\text{comp}}$. Further discussion of problems relating to multiuser sequential decoding can be found in [4].

C. The New Metric

To keep the notation simple in this introduction, we state the results only for the two-user case. These results will be restated and proved for an arbitrary number of users in the following sections.

Consider a two-user channel $(P; X_1, X_2; Y)$ and a two-user tree code for this channel. Let k be the number of channel symbols per branch in each user's tree code. The metric μ that we propose for sequential decoding in this situation is as follows.

For any $\boldsymbol{\xi}_1 \in X_1^k$, $\boldsymbol{\xi}_2 \in X_2^k$, and $\boldsymbol{\eta} \in Y^k$,

$$\mu(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\eta}) := \min\{\mu_1(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\eta}), \mu_2(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\eta}), \mu_3(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\eta})\} \quad (1.1)$$

where

$$\mu_1(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\eta}) := \ln \frac{\sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)}}{\sum\limits_{\boldsymbol{\zeta} \in X_1^k} \boldsymbol{Q}_1(\boldsymbol{\zeta})\sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\zeta}, \boldsymbol{\xi}_2)}} - kB_1$$

$$\mu_2(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\eta}) := \ln \frac{\sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)}}{\sum\limits_{\boldsymbol{\zeta} \in X_2^k} \boldsymbol{Q}_2(\boldsymbol{\zeta})\sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}_1, \boldsymbol{\zeta})}} - kB_2$$

$$\mu_3(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\eta}) := \ln \frac{\sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)}}{\sum\limits_{\boldsymbol{\zeta}_1 \in X_1^k} \boldsymbol{Q}_1(\boldsymbol{\zeta}_1) \sum\limits_{\boldsymbol{\zeta}_2 \in X_2^k} \boldsymbol{Q}_2(\boldsymbol{\zeta}_2)\sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2)}} - kB_3.$$

In the foregoing expressions, $\boldsymbol{P}$ denotes transition probabilities over blocks of length k. (We shall use boldface characters to indicate quantities relating to blocks.) For example, $\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$ is the probability that $\boldsymbol{\eta}$ is received given that user 1 transmits $\boldsymbol{\xi}_1$ and user 2 transmits $\boldsymbol{\xi}_2$.

The metric μ is parametrized by the probability distributions (p.d.'s) $\boldsymbol{Q}_1$ on X_1^k and $\boldsymbol{Q}_2$ on X_2^k, and the real numbers B_1, B_2, and B_3, which are referred to as the *bias* terms. For the metric to work satisfactorily, the bias terms must be chosen properly. One possibility, as suggested by the analysis in the following sections, is to take them as follows:

$$B_1 = (R_1 + R_0(\boldsymbol{Q}, \{1\}))/2$$

$$B_2 = (R_2 + R_0(\boldsymbol{Q}, \{2\}))/2$$

$$B_3 = (R_1 + R_2 + R_0(\boldsymbol{Q}, \{1,2\}))/2 \quad (1.2)$$

where (R_1, R_2) is the code rate and[3]

$$R_0(\boldsymbol{Q}, \{1\}) := -(1/k)\ln \sum_{\boldsymbol{\xi}_2 \in X_2^k} \boldsymbol{Q}_2(\boldsymbol{\xi}_2) \sum_{\boldsymbol{\eta} \in Y^k} \cdot \left\{ \sum_{\boldsymbol{\xi}_1 \in X_1^k} \boldsymbol{Q}_1(\boldsymbol{\xi}_1)\sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)} \right\}^2$$

$$R_0(\boldsymbol{Q}, \{2\}) := -(1/k)\ln \sum_{\boldsymbol{\xi}_1 \in X_1^k} \boldsymbol{Q}_1(\boldsymbol{\xi}_1) \sum_{\boldsymbol{\eta} \in Y^k} \cdot \left\{ \sum_{\boldsymbol{\xi}_2 \in X_2^k} \boldsymbol{Q}_2(\boldsymbol{\xi}_2)\sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)} \right\}^2$$

$$R_0(\boldsymbol{Q}, \{1,2\}) := -(1/k)\ln \sum_{\boldsymbol{\eta} \in Y^k} \left\{ \sum_{\boldsymbol{\xi}_2 \in X_2^k} \boldsymbol{Q}_2(\boldsymbol{\xi}_2) \cdot \sum_{\boldsymbol{\xi}_1 \in X_1^k} \boldsymbol{Q}_1(\boldsymbol{\xi}_1)\sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)} \right\}^2.$$

In Section IV we prove that a rate (R_1, R_2) is achievable by μ if the bias terms are chosen as in (1.2) and (R_1, R_2) satisfies

$$0 \le R_1 < R_0(\boldsymbol{Q}, \{1\}) - (2\ln 3)/k$$
$$0 \le R_2 < R_0(\boldsymbol{Q}, \{2\}) - (2\ln 3)/k$$
$$R_1 + R_2 < R_0(\boldsymbol{Q}, \{1,2\}) - (2\ln 3)/k.$$

The integer k appearing in these expressions represents the block length for $\boldsymbol{Q} = (\boldsymbol{Q}_1, \boldsymbol{Q}_2)$. Thus by choosing $\boldsymbol{Q}$ over sufficiently long blocks, we can make the term $(2\ln 3)/k$ as small as desired. It then follows, by a straightforward argument given in Section IV-C, that the achievable rate region of μ, over all $\boldsymbol{Q}$, is inner-bounded by the region $\boldsymbol{R}_0$, which is defined as follows:

$$\boldsymbol{R}_0 := \bigcup_{k=1}^{\infty} \boldsymbol{R}_0(k)$$

$$\boldsymbol{R}_0(k) := \bigcup_{\boldsymbol{Q}} \boldsymbol{R}_0(\boldsymbol{Q}). \quad (1.3)$$

The union in (1.3) is over all $\boldsymbol{Q} = (\boldsymbol{Q}_1, \boldsymbol{Q}_2)$ such that $\boldsymbol{Q}_1$ is a p.d. on X_1^k and $\boldsymbol{Q}_2$ is a p.d. on X_2^k. The region $\boldsymbol{R}_0(\boldsymbol{Q})$ is defined as the set of all (R_1, R_2) such that

$$0 \le R_1 < R_0(\boldsymbol{Q}, \{1\})$$
$$0 \le R_2 < R_0(\boldsymbol{Q}, \{2\})$$
$$R_1 + R_2 < R_0(\boldsymbol{Q}, \{1,2\}).$$

Since the achievable rate region of any metric is by definition a subset of $\boldsymbol{R}_{\text{comp}}$, we have $\boldsymbol{R}_0 \subset \boldsymbol{R}_{\text{comp}}$, the main result of this paper.

To see that $\boldsymbol{R}_0$ can be significantly larger than the set of rates achievable by time-division multiplexing (TDM), consider the two-user erasure channel of Fig. 3. The shaded region represents $\boldsymbol{R}_0((Q_1, Q_2))$ for $Q_1 = Q_2 =$ uniform distribution on $\{0,1\}$. Also included in $\boldsymbol{R}_0$, though not shown

[3] Note that the numbers $R_0(\boldsymbol{Q}, \cdot)$ are closely related to reliability exponents for MAC's [7], [4].

in the figure, are the points (1,0) and (0,1), and their convex hull with the shaded region. Thus sequential decoding achieves a sum rate as large as 1.42 bits/channel use, whereas TDM is limited to sum rates of 1 bit/channel use.

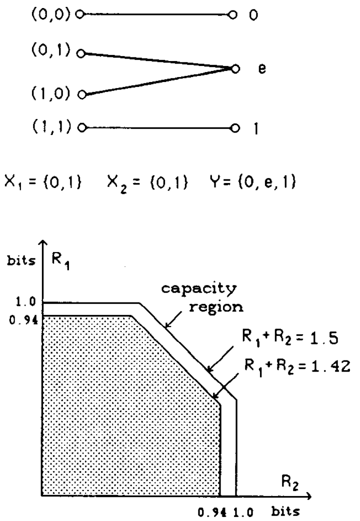

Fig. 3. Two-user erasure channel and its achievable rate regions.

Remarks:

1) It is not known whether a MAC exists for which the following statement is false:

$$\boldsymbol{R}_{\text{comp}} \subset \text{closure of } \boldsymbol{R}_0. \tag{1.4}$$

The set

$$\text{closure of } \boldsymbol{R}_0 \backslash \boldsymbol{R}_0$$

consists of only those points that lie on the "outer surface" of $\boldsymbol{R}_0$, and for all practical purposes $\boldsymbol{R}_0$ and its closure are the same. Therefore, if (1.4) holds for a channel, its $\boldsymbol{R}_{\text{comp}}$ is essentially determined.

Two classes of channels exist for which (1.4) holds. These are the classes of single-user channels, discussed next, and pairwise-reversible channels, discussed in Section IV-D.

2) In the case of a single-user channel $(P; X; Y)$, the metric μ is defined as follows. For each $\boldsymbol{\xi} \in X^k$, $\boldsymbol{\eta} \in Y^k$,

$$\mu(\boldsymbol{\xi}, \boldsymbol{\eta}) = \ln \frac{\sqrt{P(\boldsymbol{\eta}|\boldsymbol{\xi})}}{\sum_{\boldsymbol{\zeta} \in X^k} Q(\boldsymbol{\zeta})\sqrt{P(\boldsymbol{\eta}|\boldsymbol{\zeta})}} - kB$$

where Q is a p.d. on X^k, B is the bias term, and the integer k represents the number of symbols on each branch of the tree code.

The results of Section IV imply that, as Q and B vary over their possible values, this metric achieves all rates in the interval $\boldsymbol{R}_0 := [0, R_0)$, where

$$R_0 := \max_Q -\ln \sum_{\eta \in Y} \left\{ \sum_{\xi \in X} Q(\xi)\sqrt{P(\eta|\xi)} \right\}^2. \tag{1.5}$$

The maximum is over all p.d.'s Q on X. The parameter R_0 is usually referred to as the computational cutoff rate of sequential decoding.

In fact, one does not need to consider all parameter values to show that μ achieves all rates below R_0. To achieve a rate $R \in [0, R_0)$, it suffices to take

$$Q(\boldsymbol{\xi}) = \hat{Q}(\xi_1)\hat{Q}(\xi_2) \cdots \hat{Q}(\xi_k), \qquad \text{all } \boldsymbol{\xi} = (\xi_1, \cdots, \xi_k) \in X^k$$

where $\hat{Q}$ is a p.d. that achieves the maximum in (1.5) and

$$B = (R_0 + R)/2.$$

This choice of parameters also simplifies the form of the metric for it can now be expressed as a sum of metrics over single letters. Unfortunately, no such simplification is possible in the multiuser case.

Note that in the single-user case the metric μ is not equivalent to (more precisely, not an affine function of) the Fano metric. We thus have perhaps the surprising conclusion that the Fano metric is not unique in achieving all of $\boldsymbol{R}_0$ for single-user channels.

For single-user channels it is known [9], [10] that $\boldsymbol{R}_{\text{comp}} \subset [0, R_0]$. Therefore, with the possible exception of R_0, the regions $\boldsymbol{R}_0$ and $\boldsymbol{R}_{\text{comp}}$ coincide. We may have $R_0 \in \boldsymbol{R}_{\text{comp}}$, as in the case of a binary symmetric channel with zero crossover probability. On the other hand, it is not known nor is it of any practical importance if a channel exists for which $R_0 \notin \boldsymbol{R}_{\text{comp}}$.[4]

D. Organization

The rest of this paper is organized as follows. Section II contains the notation and gives a precise definition of $\boldsymbol{R}_{\text{comp}}$. In Section III sufficient conditions are found on the achievability of a rate by a given metric. In Section IV we introduce the new metric in its general form and prove that it achieves $\boldsymbol{R}_0$. Section V contains a discussion of some properties, such as convexity, of $\boldsymbol{R}_0$, and a listing of several open problems. In Section VI we examine an alternative definition of achievability in sequential decoding. Section VII is a summary of the results.

II. Multi-User Tree Coding and Sequential Decoding

A. Description of the System

The discussion henceforth is, unless otherwise stated, in the context of a fixed but arbitrary multiple access communication system. We use the following notation for labeling various parts of this system.

The channel will be denoted by $(P; X_1, \cdots, X_n; Y)$. Thus, n will be the number of users, X_i the channel input alphabet for user i, and Y the channel output alphabet. We shall let $\{0, \cdots, M_i - 1\}$ be the source alphabet of user

[4] We conjecture that, for all channels whose zero-error capacity is 0, either $R_0 = 0$ or $R_0 \notin \boldsymbol{R}_{\text{comp}}$.

i. A generic output sequence of source i will be denoted by

$$\boldsymbol{u}_i = \boldsymbol{u}_i(1), \boldsymbol{u}_i(2), \cdots$$

where $\boldsymbol{u}_i(m) \in \{0, \cdots, M_i - 1\}$ for each $m = 1, 2, \cdots$. The first m letters in $\boldsymbol{u}_i$ will be denoted by $\boldsymbol{u}_i(..m)$, i.e.,

$$\boldsymbol{u}_i(..m) = \boldsymbol{u}_i(1), \cdots, \boldsymbol{u}_i(m).$$

User i's encoder will be denoted by $\boldsymbol{e}_i$, and its function will be to generate a block of k letters from alphabet X_i in response to each letter it accepts from source i. The integer k is assumed the same for each $\boldsymbol{e}_i$, $i = 1, \cdots, n$, but otherwise it is arbitrary. The output sequence of $\boldsymbol{e}_i$, in response to a source sequence $\boldsymbol{u}_i$, will be denoted by $\boldsymbol{e}_i\boldsymbol{u}_i$, the mth block in the output sequence by $\boldsymbol{e}_i\boldsymbol{u}_i(m)$, and the first m blocks by $\boldsymbol{e}_i\boldsymbol{u}_i(..m)$.

The collection of encoders $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n$ will be said to have parameter $(M_1, \cdots, M_n; k)$ and rate $(R_1, \cdots, R_n)$ where $R_i = (1/k)\ln(M_i)$ nats/channel use. We shall also view the encoder $\boldsymbol{e}_i$ as a mapping from sequences over $\{0, \cdots, M_i - 1\}$ to sequences over X_i, represent this mapping as a tree diagram as described later, and refer to $\boldsymbol{e}_i$ as *user i's tree code*.

The tree representing $\boldsymbol{e}_i$ will have one node at level m for each initial segment $\boldsymbol{u}_i(..m)$, and conversely. Because of this one-to-one correspondence, we shall refer to $\boldsymbol{u}_i(..m)$ as a *node* in $\boldsymbol{e}_i$. By convention, $\boldsymbol{u}_i(..0)$ will denote the origin node in $\boldsymbol{e}_i$. The branch $\boldsymbol{e}_i$ connecting node $\boldsymbol{u}_i(..m)$ to node $\boldsymbol{u}_i(..m+1)$ will be labeled by the block $\boldsymbol{e}_i\boldsymbol{u}_i(m+1)$ of k letters from X_i. Because of the resulting one-to-one correspondence between paths in $\boldsymbol{e}_i$ and the set of all source sequences for user i, we shall refer to a source sequence $\boldsymbol{u}_i$ also as a *path* in $\boldsymbol{e}_i$.

We shall assume that the users start their transmissions at a common point in time. This assumption and the one that k, the number of channel input symbols per branch, is the same for all users ensure that for each $m = 1, 2, \cdots$, each user transmits the mth branch of its tree code in synchronism with each other.

Branch synchronization among users allows us to view the collection of encoders (tree codes) $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n$ as one joint encoder (tree code) with a source alphabet of size $M_1 \cdots M_n$ and a channel input alphabet $X_1 \times \cdots \times X_n$, the Cartesian product of the channel input alphabets. The joint encoder (tree code) will be denoted by $\boldsymbol{e}$.

To emphasize that $\boldsymbol{e}$ is not an arbitrary tree code, but the *product* of *n component* tree codes, we will refer to it as an *n-user* tree code. The way a joint tree code is related to its component tree codes is illustrated in Fig. 2.

The following notation will be useful for further description and analysis of the system. For any collection of r-tuples $\boldsymbol{a} = (a_1, \cdots, a_r) \in A^r$, $\boldsymbol{b} = (b_1, \cdots, b_r) \in B^r, \cdots, \boldsymbol{c} = (c_1, \cdots, c_r) \in C^r$ over arbitrary alphabets $A, B, \cdots, C$, let

$$\boldsymbol{a} \times \boldsymbol{b} \times \cdots \times \boldsymbol{c} := ((a_1, b_1, \cdots, c_1), \cdots, (a_r, b_r, \cdots, c_r)).$$

If $\boldsymbol{a} = a_1, a_2, \cdots$, $\boldsymbol{b} = b_1, b_2, \cdots$, $\boldsymbol{c} = c_1, c_2, \cdots$ are sequences over $A, B, \cdots, C$, respectively, then let

$$\boldsymbol{a} \times \boldsymbol{b} \times \cdots \times \boldsymbol{c} := (a_1, b_1, \cdots, c_1), (a_2, b_2, \cdots, c_2), \cdots.$$

In either case, $\boldsymbol{a} \times \boldsymbol{b} \times \cdots \times \boldsymbol{c}$ is said to be the *product* of $\boldsymbol{a}, \boldsymbol{b}, \cdots, \boldsymbol{c}$.

Using this notation, we define the *joint source sequence* $\boldsymbol{u}$ corresponding to a collection of source sequences $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n$ by $\boldsymbol{u} = \boldsymbol{u}_1 \times \cdots \times \boldsymbol{u}_n$. The mth element in $\boldsymbol{u}$ will be denoted by $\boldsymbol{u}(m)$, and the first m elements by $\boldsymbol{u}(..m)$. Note that these definitions imply the relations

$$\boldsymbol{u}(m) = \boldsymbol{u}_1(m) \times \cdots \times \boldsymbol{u}_n(m) = (\boldsymbol{u}_1(m), \cdots, \boldsymbol{u}_n(m))$$
$$\boldsymbol{u}(..m) = \boldsymbol{u}_1(..m) \times \cdots \times \boldsymbol{u}_n(..m).$$

The output sequence, in response to a joint source sequence $\boldsymbol{u}$, of the joint encoder $\boldsymbol{e}$ will be denoted by $\boldsymbol{eu}$, the mth output block of $\boldsymbol{e}$ by $\boldsymbol{eu}(m)$, and the first m output blocks of $\boldsymbol{e}$ by $\boldsymbol{eu}(..m)$. Note that, for $\boldsymbol{u} = \boldsymbol{u}_1 \times \cdots \times \boldsymbol{u}_n$,

$$\boldsymbol{eu} = \boldsymbol{e}_1\boldsymbol{u}_1 \times \cdots \times \boldsymbol{e}_n\boldsymbol{u}_n$$
$$\boldsymbol{eu}(m) = \boldsymbol{e}_1\boldsymbol{u}_1(m) \times \cdots \times \boldsymbol{e}_n\boldsymbol{u}_n(m)$$
$$\boldsymbol{eu}(..m) = \boldsymbol{e}_1\boldsymbol{u}_1(..m) \times \cdots \times \boldsymbol{e}_n\boldsymbol{u}_n(..m).$$

The *correct path* in $\boldsymbol{e}_i$ is defined as the path that corresponds to the actual output sequence of source i. A node in $\boldsymbol{e}_i$ will be referred to as a *correct node* if it lies on the correct path. The correct path in $\boldsymbol{e}$ is defined as the product of the correct paths in the component trees. A node in $\boldsymbol{e}$ will be referred to as a correct node if it lies on the correct path in $\boldsymbol{e}$.

We shall denote the channel output sequence by $\boldsymbol{y}$, the block in $\boldsymbol{y}$ received in response to the mth channel input block by $\boldsymbol{y}(m)$, and the first m such blocks by $\boldsymbol{y}(..m)$. We shall assume that the sources are memoryless and statistically independent of each other, and that each source generates letters according to the uniform p.d. on its output alphabet. These assumptions are equivalent to assuming that each path in the joint tree code $\boldsymbol{e}$ is equally likely to be the correct (transmitted) path. We shall assume that the decoder is a sequential decoder, which will try to find the correct path in the joint tree code $\boldsymbol{e}$ after observing the channel output sequence $\boldsymbol{y}$.

B. Sequential Decoding

The purpose of this section is to give a precise definition of *achievability* in sequential decoding. We begin with a review of sequential decoding in the context of the system just described.

Generally speaking, a *metric* Γ for sequential decoding is any function

$$\Gamma: \bigcup_{h=1}^{\infty} (X_1 \times \cdots \times X_n)^{hk} \times Y^{hk} \to [-\infty, \infty)$$

which assigns the value $\Gamma(\boldsymbol{eu}(..m), \boldsymbol{y}(..m))$ to node $\boldsymbol{u}(..m)$ given that $\boldsymbol{y}$ is received. Note that $\Gamma(\boldsymbol{eu}(..m), \boldsymbol{y}(..m))$ does not depend on the portion of the received sequence beyond level m, a restriction essential to the idea of sequential decoding. Also note that the metric is allowed to take on the value $-\infty$, thereby enabling the sequential decoder to exclude a node from further consideration if somehow it is determined that that node is incorrect.

A metric Γ is said to be branchwise additive if there exists a function $\gamma, \gamma\colon (X_1 \times \cdots \times X_n)^k \times Y^k \to [-\infty, +\infty)$ such that

$$\Gamma(\boldsymbol{eu}(..m), \boldsymbol{y}(..m)) = \sum_{i=1}^{m} \gamma(\boldsymbol{eu}(i), \boldsymbol{y}(i)),$$

in which case γ is called the *branch metric* for Γ. We shall refer to a branch metric simply as a metric when no confusion can arise. An example of a branch metric is the function μ that was introduced in Section I.

Our analysis of sequential decoding will be given for the *stack algorithm*,[5] which is a version of sequential decoding due to Zigangirov [11] and Jelinek [12]. The following is a brief description of the stack algorithm. The nodes that are referred to in the description are the nodes of the joint tree code $\boldsymbol{e}$.

Each step of the algorithm consists of updating an ordered list of nodes, referred to as the *stack*. Initially, the stack contains only the origin node. At all times the nodes on the stack are ordered with respect to their metric values, the node with the highest metric value being at the *stack top*. To update the stack, one deletes the node at the stack top and inserts its immediate descendants into the stack after computing their metric values. Ties among metric values are broken according to some arbitrary rule.

In a finite tree code the stack algorithm ends as soon as a node at the last level of the tree reaches the stack top, which is then declared as the decoder output. Of course, in an infinite tree code the stack algorithm never ends. We shall consider only infinite tree codes because we are concerned primarily with the average complexity of sequential decoding—a concept most cleanly formalized and conservatively estimated in the framework of infinite tree codes.

For every integer $m \geq 1$ and path $\boldsymbol{u}$ in $\boldsymbol{e}$, the mth *incorrect subtree* of $\boldsymbol{u}$ is defined as follows:

$$\begin{aligned} I_m(\boldsymbol{u}) = \{&\text{nodes } \boldsymbol{v}(..j) \text{ in} \\ &\boldsymbol{e}\colon j \geq m,\ \boldsymbol{v}(..m-1) = \boldsymbol{u}(..m-1), \\ &\text{and } \boldsymbol{v}(..m) \neq \boldsymbol{u}(..m)\}. \end{aligned}$$

In words, $I_m(\boldsymbol{u})$ is the set of nodes, at levels $\geq m$, lying on paths that diverge from $\boldsymbol{u}$ at level $m-1$.

We let $C_j(\boldsymbol{e}, \boldsymbol{s}, \boldsymbol{y}, \Gamma)$ denote the number of nodes in $I_j(\boldsymbol{s})$ that reach the stack top given that $\boldsymbol{e}$ is the code, $\boldsymbol{s}$ the correct path, $\boldsymbol{y}$ the received sequence, and Γ the metric. We let $C_j(\boldsymbol{e}, \Gamma)$ denote the value of $C_j(\boldsymbol{e}, \boldsymbol{s}, \boldsymbol{y}, \Gamma)$ averaged over $\boldsymbol{s}$ and $\boldsymbol{y}$. More precisely, $C_j(\boldsymbol{e}, \Gamma) = E_{\boldsymbol{s}} E_{\boldsymbol{y}|\boldsymbol{es}} C_j(\boldsymbol{e}, \boldsymbol{s}, \boldsymbol{y}, \Gamma)$, where $E_{\boldsymbol{s}}$ denotes expectation with respect to (wrt) the probability that $\boldsymbol{s}$ is the correct path, and $E_{\boldsymbol{y}|\boldsymbol{es}}$ denotes expectation wrt the conditional probability that $\boldsymbol{y}$ is received given that $\boldsymbol{es}$ is transmitted.

For each $L \geq 1$, we let

$$D_L(\boldsymbol{e}, \Gamma) := (C_1(\boldsymbol{e}, \Gamma) + \cdots + C_L(\boldsymbol{e}, \Gamma))/L.$$

Since a node can reach the stack top at most once, we can interpret $D_L(\boldsymbol{e}, \Gamma)$ as a rough measure of the average work required to move forward one step on the correct path. This interpretation is the intuitive basis of the following definitions. A metric Γ is said to *achieve* a point[6] $R = (R_1, \cdots, R_n)$ if, for any given L, a tree code $\boldsymbol{e}$ exists with rate $\geq R$ (componentwise) such that

$$D_L(\boldsymbol{e}, \Gamma) < A$$

where A is a finite number independent of L.[7] A point R is said to be *achievable by sequential decoding* if there exists a metric Γ that achieves R.

The *achievable rate region* of a metric Γ is defined as the set of all points that Γ achieves. The *achievable rate region of sequential decoding* is defined as the set of all points that are achievable by some metric and is denoted by $\boldsymbol{R}_{\text{comp}}$.

III. Sufficient Conditions on Achievability

This section finds sufficient conditions on the achievability of a point by a branchwise-additive metric Γ. Throughout the section, we take Γ as fixed but otherwise arbitrary, and let γ denote the branch metric for Γ. In the single-user case, there are well-known sufficient conditions on achievability (see, e.g., [13, ch. 6]). The following is a straightforward generalization of these results.

Recall that a point R is achievable by Γ iff for each L a code $\boldsymbol{e}$ exists with rate $\geq R$ such that $D_L(\boldsymbol{e}, \Gamma) < A < \infty$, where A is independent of L. The computation of $D_L(\boldsymbol{e}, \Gamma)$ for any nontrivial $\boldsymbol{e}$ and Γ is a hopelessly complicated task, so instead of working directly with any individual code, we shall use *random-coding* techniques. That is, we shall consider the average value of $D_L(\boldsymbol{e}, \Gamma)$ over all $\boldsymbol{e}$ in an ensemble of codes. By an ensemble, we mean a set of codes and a probability measure on this set. The idea here is simple: if we can find a finite number A such that, for all L, the ensemble average of $D_L(\boldsymbol{e}, \Gamma) < A$, then for any given L, a code $\boldsymbol{e}$ exists such that $D_L(\boldsymbol{e}, \Gamma) < A$.

In the following, we shall consider a family of ensembles $\{\text{Ens}(\boldsymbol{M}, \boldsymbol{Q})\}$ parametrized by $(\boldsymbol{M}, \boldsymbol{Q})$. Here, $\boldsymbol{M}$ is of the form $\boldsymbol{M} = (M_1, \cdots, M_n; k)$ and designates the parameter of the codes in the ensemble $\text{Ens}(\boldsymbol{M}, \boldsymbol{Q})$. The second parameter is of the form $\boldsymbol{Q} = (Q_1, \cdots, Q_n)$, where Q_i is a p.d. on X_i^k for each $i = 1, \cdots, n$, and it specifies the probability measure associated with $\text{Ens}(\boldsymbol{M}, \boldsymbol{Q})$.

More precisely, the ensemble $\text{Ens}(\boldsymbol{M}, \boldsymbol{Q})$ has as its code set the set of all n-user tree codes with parameter $\boldsymbol{M}$. The probability measure associated with $\text{Ens}(\boldsymbol{M}, \boldsymbol{Q})$ is such that a code chosen at random from this probability measure has the same statistical properties as the outcome of a random experiment in which, for each $i = 1, \cdots, n$, each branch in user i's tree code is independently assigned the symbol $\xi \in X_i^k$ with probability $Q_i(\xi)$.

[5] The main results of this paper can be easily extended to the Fano algorithm.

[6] In all of these definitions, it is implicitly assumed that the coordinates of the points are nonnegative.

[7] This definition, which is the one commonly used in the literature, allows $\boldsymbol{e}$ to depend on L. For an alternative definition, where $\boldsymbol{e}$ has to be independent of L, see Section VI.

Now we shall consider the expected value of $D_L(\boldsymbol{e},\Gamma)$ over $\mathrm{Ens}(\boldsymbol{M},\boldsymbol{Q})$ for a fixed but otherwise arbitrary $(\boldsymbol{M},\boldsymbol{Q})$. To simplify the notation, we shall suppress arguments that are held fixed through the following development. The notation $E_{\boldsymbol{e}}$ will denote the expectation operation wrt the probability measure associated with $\mathrm{Ens}(\boldsymbol{M},\boldsymbol{Q})$.

$$\begin{aligned} E_{\boldsymbol{e}}D_L(\boldsymbol{e}) &= (1/L)E_{\boldsymbol{e}}\{C_1(\boldsymbol{e})+\cdots+C_L(\boldsymbol{e})\}\\ &= (1/L)\{E_{\boldsymbol{e}}C_1(\boldsymbol{e})+\cdots+E_{\boldsymbol{e}}C_L(\boldsymbol{e})\}. \quad (3.1)\end{aligned}$$

Therefore, $E_{\boldsymbol{e}}D_L(\boldsymbol{e})$ can be upper-bounded by upper-bounding $E_{\boldsymbol{e}}C_i(\boldsymbol{e})$ for each i.

$$\begin{aligned} E_{\boldsymbol{e}}C_i(\boldsymbol{e}) &= E_{\boldsymbol{e}}E_{\boldsymbol{s}}E_{\boldsymbol{y}|\boldsymbol{es}}C_i(\boldsymbol{e},\boldsymbol{s},\boldsymbol{y})\\ &= E_{\boldsymbol{s}}E_{\boldsymbol{e}}E_{\boldsymbol{y}|\boldsymbol{es}}C_i(\boldsymbol{e},\boldsymbol{s},\boldsymbol{y}). \quad (3.2)\end{aligned}$$

Reversing the order of expectations is justified by the nonnegativity of the terms involved. $E_{\boldsymbol{e}}C_i(\boldsymbol{e})$ will be upper-bounded with the help of the following lemma (c.f., [14, lemma 6.2.1]).[8]

Lemma 1: For any $t \geq 0$,

$$C_i(\boldsymbol{e},\boldsymbol{s},\boldsymbol{y}) \leq \sum_{\boldsymbol{u}(..j)\in I_i(\boldsymbol{s})}\ \sum_{m\geq i} \exp\{t(\Gamma(\boldsymbol{eu}(..j),\boldsymbol{y}(..j)) - \Gamma(\boldsymbol{es}(..m),\boldsymbol{y}(..m)))\}. \quad (3.3)$$

Proof: A node $\boldsymbol{u}(..j)\in I_i(\boldsymbol{s})$ reaches the stack top only if

$$\Gamma(\boldsymbol{eu}(..j),\boldsymbol{y}(..j)) \geq \Gamma(\boldsymbol{es}(..m),\boldsymbol{y}(..m)), \quad \text{for some } m\geq i. \quad (3.4)$$

Therefore, for $t\geq 0$, $\Sigma_{m\geq i}\exp\{t(\Gamma(\boldsymbol{eu}(..j),\boldsymbol{y}(..j)) - \Gamma(\boldsymbol{es}(..m),\boldsymbol{y}(..m)))\}$ upper-bounds the indicator function of the event that $\boldsymbol{u}(..j)\in I_i(\boldsymbol{s})$ reaches the stack top. The lemma follows by summing over all nodes in $I_i(\boldsymbol{s})$.

Let

$$\Lambda(\boldsymbol{s},m,\boldsymbol{u}(..j),t) := E_{\boldsymbol{e}}E_{\boldsymbol{y}|\boldsymbol{es}}\exp\{t(\Gamma(\boldsymbol{eu}(..j),\boldsymbol{y}(..j)) - \Gamma(\boldsymbol{es}(..m),\boldsymbol{y}(..m)))\}.$$

By Lemma 1,

$$E_{\boldsymbol{e}}C_i(\boldsymbol{e}) \leq E_{\boldsymbol{s}} \sum_{\boldsymbol{u}(..j)\in I_i(\boldsymbol{s})}\ \sum_{m\geq i}\Lambda(\boldsymbol{s},m,\boldsymbol{u}(..j),t), \quad \text{for all } t\geq 0. \quad (3.5)$$

We now seek an upper bound on $\Lambda(\boldsymbol{s},m,\boldsymbol{u}(..j),t)$. Note that, for any $\boldsymbol{u}(..j)\in I_i(\boldsymbol{s})$, if $j>m$,

$$\begin{aligned}\Lambda&(\boldsymbol{s},m,\boldsymbol{u}(..j),t)\\ &= E_{\boldsymbol{e}}E_{\boldsymbol{y}|\boldsymbol{es}}\exp\Bigg\{t\Bigg(\sum_{i\leq h\leq m}\{\gamma(\boldsymbol{eu}(h),\boldsymbol{y}(h))\\ &\quad -\gamma(\boldsymbol{es}(h),\boldsymbol{y}(h))\} + \sum_{m<h\leq j}\gamma(\boldsymbol{eu}(h),\boldsymbol{y}(h))\Bigg)\Bigg\}, \quad (3.6)\end{aligned}$$

and if $m\geq j$,

$$\begin{aligned}\Lambda&(\boldsymbol{s},m,\boldsymbol{u}(..j),t)\\ &= E_{\boldsymbol{e}}E_{\boldsymbol{y}|\boldsymbol{es}}\exp\Bigg\{t\Bigg(\sum_{i\leq h\leq j}\{\gamma(\boldsymbol{eu}(h),\boldsymbol{y}(h))\\ &\quad -\gamma(\boldsymbol{es}(h),\boldsymbol{y}(h))\} - \sum_{j<h\leq m}\gamma(\boldsymbol{es}(h),\boldsymbol{y}(h))\Bigg)\Bigg\}. \quad (3.7)\end{aligned}$$

Since the branch labels are statistically independent and the channel is memoryless, (3.6) and (3.7) can be rewritten as follows. For any $\boldsymbol{u}(..j)\in I_i(\boldsymbol{s})$, if $j>m$,

$$\begin{aligned}\Lambda&(\boldsymbol{s},m,\boldsymbol{u}(..j),t)\\ &= \prod_{i\leq h\leq m} E_{\boldsymbol{e}}E_{\boldsymbol{y}(h)|\boldsymbol{es}(h)}\\ &\quad \cdot\exp\{t(\gamma(\boldsymbol{eu}(h),\boldsymbol{y}(h))-\gamma(\boldsymbol{es}(h),\boldsymbol{y}(h)))\}\\ &\quad \times\prod_{m<h\leq j}E_{\boldsymbol{e}}E_{\boldsymbol{y}(h)|\boldsymbol{es}(h)}\exp\{t\gamma(\boldsymbol{eu}(h),\boldsymbol{y}(h))\}, \quad (3.8)\end{aligned}$$

and if $m\geq j$,

$$\begin{aligned}\Lambda&(\boldsymbol{s},m,\boldsymbol{u}(..j),t)\\ &= \prod_{i\leq h\leq j} E_{\boldsymbol{e}}E_{\boldsymbol{y}(h)|\boldsymbol{es}(h)}\\ &\quad \cdot\exp\{t(\gamma(\boldsymbol{eu}(h),\boldsymbol{y}(h))-\gamma(\boldsymbol{es}(h),\boldsymbol{y}(h)))\}\\ &\quad \times\prod_{j<h\leq m}E_{\boldsymbol{e}}E_{\boldsymbol{y}(h)|\boldsymbol{es}(h)}\exp\{-t\gamma(\boldsymbol{es}(h),\boldsymbol{y}(h))\} \quad (3.9)\end{aligned}$$

where $E_{\boldsymbol{y}(h)|\boldsymbol{es}(h)}$ denotes expectation wrt the conditional probability that $\boldsymbol{y}(h)$ is received at the channel output given that $\boldsymbol{es}(h)$ is transmitted.

To simplify the expression for $\Lambda(\boldsymbol{s},m,\boldsymbol{u}(..j),t)$, we need some definitions. The *type* of a node $\boldsymbol{u}(..m) = \boldsymbol{u}_1(..m)\times\cdots\times\boldsymbol{u}_n(..m)$ wrt a path $\boldsymbol{v}=\boldsymbol{v}_1\times\cdots\times\boldsymbol{v}_n$ is defined as an m-tuple $(T_1,\cdots,T_m)$, where T_j, $1\leq j\leq m$, is the set of users i, $1\leq i\leq n$, such that $\boldsymbol{u}_i(..j)\neq\boldsymbol{v}_i(..j)$.

For later use, we wish to point out here that the number of types for level-m nodes wrt any fixed path equals $(m+1)^n$. This can be seen by observing that if $(T_1,\cdots,T_m)$ is a type, then T_j must be a subset of T_h for all $h>j$. Thus $m+1$ ways exist in which a user i first appears (one possibility being that it does not appear at all) in the sequence of sets $T_1,\cdots,T_m$. By a similar argument it can be seen that the number of types for level-m nodes in $I_i(\boldsymbol{u})$, where $\boldsymbol{u}$ is an arbitrary fixed path, equals $(m-\mathrm{i}+2)^n$ because now $T_1=\cdots=T_{i-1}=\varnothing$.

The underlying probability measure, incorporating the randomness of the code and the channel transitions, with respect to which the expectations must be computed can now be stated explicitly as follows. Let $\boldsymbol{u}(..j)$ be a node in $I_i(\boldsymbol{s})$ with type $(T_1,\cdots,T_j)$ wrt $\boldsymbol{s}$, where necessarily $T_1=\cdots=T_{i-1}=\varnothing$. Then for any $h\in\{1,\cdots,j\}$, $\eta\in Y^k$, $\xi=\xi_1\times\cdots\times\xi_n$, and $\zeta=\zeta_1\times\cdots\times\zeta_n$, where $\xi_r\in X_r^k$, $\zeta_r\in$

[8] We could simplify (3.2) by observing that $E_{\boldsymbol{e}}E_{\boldsymbol{y}|\boldsymbol{es}}C_i(\boldsymbol{e},\boldsymbol{s},\boldsymbol{y})$ has the same value for all $\boldsymbol{s}$; hence in (3.2) $E_{\boldsymbol{s}}$ can be dropped and $\boldsymbol{s}$ replaced by any fixed path.

X_r^k, $r=1,\cdots,n$,

$$\Pr\{\boldsymbol{es}(h)=\boldsymbol{\xi},\boldsymbol{eu}(h)=\boldsymbol{\zeta},\boldsymbol{y}(h)=\boldsymbol{\eta}\}$$
$$=\prod_{1\le r\le n}Q_r(\xi_r)\prod_{q\in T_h}Q_q(\zeta_q)\prod_{p\in\overline{T}_h}\mathbf{1}\{\zeta_p=\xi_p\}\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}) \tag{3.10}$$

where $\mathbf{1}$ is the indicator function and $\overline{T}_h$ is the complement of T_h in $\{1,\cdots,n\}$.

The notation is simplified by writing

$$\boldsymbol{Q}(\boldsymbol{\xi})\text{ in place of }\prod_{1\le r\le n}Q_r(\xi_r)$$
$$\boldsymbol{Q}(\boldsymbol{\xi}_{T_h})\text{ in place of }\prod_{q\in T_h}Q_q(\xi_q)$$
$$\mathbf{1}\{\boldsymbol{\zeta}_{\overline{T}_h}=\boldsymbol{\xi}_{\overline{T}_h}\}\text{ in place of }\prod_{p\in\overline{T}_h}\mathbf{1}\{\zeta_p=\xi_p\}$$

for every nonempty $T_h\subset\{1,\cdots,n\}$. For $T_h=\varnothing$, we shall adopt the convention that $\mathbf{1}\{\boldsymbol{\zeta}_{\overline{T}_h}=\boldsymbol{\xi}_{\overline{T}_h}\}=1$ and $\boldsymbol{Q}(\boldsymbol{\xi}_{T_h})=1$.

Using this notation, (3.10) can be rewritten as

$$\Pr\{\boldsymbol{es}(h)=\boldsymbol{\xi},\boldsymbol{eu}(h)=\boldsymbol{\zeta},\boldsymbol{y}(h)=\boldsymbol{\eta}\}$$
$$=\boldsymbol{Q}(\boldsymbol{\xi})\boldsymbol{Q}(\boldsymbol{\zeta}_{T_h})\mathbf{1}\{\boldsymbol{\zeta}_{\overline{T}_h}=\boldsymbol{\xi}_{\overline{T}_h}\}\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}).$$

We can now evaluate the terms appearing in (3.8) and (3.9) as follows:

$$E_eE_{\boldsymbol{y}(h)|\boldsymbol{es}(h)}\exp\{-t\gamma(\boldsymbol{es}(h),\boldsymbol{y}(h))\}$$
$$=\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}\boldsymbol{Q}(\boldsymbol{\xi})\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi})\exp\{-t\gamma(\boldsymbol{\xi},\boldsymbol{\eta})\} \tag{3.11}$$

$$E_eE_{\boldsymbol{y}|\boldsymbol{es}}\exp\{t(\gamma(\boldsymbol{eu}(h),\boldsymbol{y}(h))-\gamma(\boldsymbol{es}(h),\boldsymbol{y}(h)))\}$$
$$=\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}\sum_{\boldsymbol{\zeta}}\boldsymbol{Q}(\boldsymbol{\xi})\boldsymbol{Q}(\boldsymbol{\zeta}_{T_h})\mathbf{1}\{\boldsymbol{\zeta}_{\overline{T}_h}=\boldsymbol{\xi}_{\overline{T}_h}\}$$
$$\cdot\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi})\exp\{t(\gamma(\boldsymbol{\zeta},\boldsymbol{\eta})-\gamma(\boldsymbol{\xi},\boldsymbol{\eta}))\} \tag{3.12}$$

$$E_eE_{\boldsymbol{y}|\boldsymbol{es}}\exp\{t\gamma(\boldsymbol{eu}(h),\boldsymbol{y}(h))\}$$
$$=\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}\sum_{\boldsymbol{\zeta}}\mathbf{Q}(\boldsymbol{\xi})\mathbf{Q}(\boldsymbol{\zeta}_{T_h})\mathbf{1}\{\boldsymbol{\xi}_{\overline{T}_h}=\boldsymbol{\zeta}_{\overline{T}_h}\}$$
$$\cdot\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi})\exp\{t\gamma(\boldsymbol{\zeta},\boldsymbol{\eta})\}. \tag{3.13}$$

In (3.11)–(3.13), $\boldsymbol{\eta}$ runs through Y^k, $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ through $(X_1\times\cdots\times X_n)^k$.

We observe that the right side of (3.11) is independent of $\boldsymbol{s}$ and h, and the right sides of (3.12) and (3.13) depend only on T_h. Therefore, for further simplifications we define $\alpha(t)$, $\sigma(T_h,t)$, and $\beta(T_h,t)$ as the right sides of (3.11), (3.12), and (3.13), respectively. We also define, for any type $\boldsymbol{T}=(T_1,\cdots,T_j)$,

$$\overline{\Lambda}(m,\boldsymbol{T},t):=\begin{cases}\sigma(T_1,t)\cdots\sigma(T_m,t)\beta(T_{m+1},t)\cdots\beta(T_j,t), & \\ & j>m\\ \sigma(T_1,t)\cdots\sigma(T_j,t)\alpha(t)^{m-j}, & m\ge j.\end{cases}$$

Now for any node $\boldsymbol{u}(..j)$ of type $\boldsymbol{T}$ wrt $\boldsymbol{s}$, $\Lambda(\boldsymbol{s},m,\boldsymbol{u}(..j),t)=\overline{\Lambda}(m,\boldsymbol{T},t)$. Thus $\Lambda(\boldsymbol{s},m,\boldsymbol{u}(..j),t)$ depends on $\boldsymbol{u}(..j)$ only through its type wrt $\boldsymbol{s}$. We now proceed to simplify the summation in (3.5).

For all i and j, $1\le i\le j$, let $\mathbb{T}(i,j)$ be the set of types $(T_1,\cdots,T_j)$ such that $T_1=\cdots=T_{i-1}=\varnothing$ and $T_i\ne\varnothing$. Thus $\mathbb{T}(i,j)$ is just the set of all possible types for level-j nodes that are in $I_i(\boldsymbol{s})$, for any arbitrary $\boldsymbol{s}$.

Let $N(\boldsymbol{T})$ be the number of nodes of type T wrt a fixed path. (Clearly, $N(\boldsymbol{T})$ is the same for any fixed path.) We can rewrite (3.5) as follows:

$$E_eC_i(\boldsymbol{e})\le E_s\sum_{j=i}^{\infty}\sum_{\boldsymbol{T}\in\mathbb{T}(i,j)}\sum_{\substack{\boldsymbol{u}(..j):\\ \text{type of }\boldsymbol{u}(..j)\text{ wrt }\boldsymbol{s}=\boldsymbol{T}}}\sum_{m=i}^{\infty}$$
$$\cdot\Lambda(\boldsymbol{s},m,\boldsymbol{u}(..j),t)$$
$$=E_s\sum_{j=i}^{\infty}\sum_{\boldsymbol{T}\in\mathbb{T}(i,j)}N(\boldsymbol{T})\sum_{m=i}^{\infty}\overline{\Lambda}(m,\boldsymbol{T},t),$$
$$\text{for all }t\ge0. \tag{3.14}$$

Now we shall upper-bound $N(\boldsymbol{T})\overline{\Lambda}(m,\boldsymbol{T},t)$.

For $T\subset\{1,\cdots,n\}$ and $\boldsymbol{M}=(M_1,\cdots,M_n;k)$, let

$$M(T):=\begin{cases}1, & \text{if }T=\varnothing;\\ \prod_{i\in T}M_i, & \text{otherwise.}\end{cases}$$

For any type $\boldsymbol{T}=(T_1,\cdots,T_j)$, let $\overline{M}(\boldsymbol{T}):=M(T_1)\cdots M(T_j)$. Also let

$$\Psi(t):=\max_{T\ne\varnothing}\{M(T)\max\{\sigma(T,t),\beta(T,t)\}\}.$$

Now for any $\boldsymbol{T}\in\mathbb{T}(i,j)$, we have the following trivial inequality:

$$N(\boldsymbol{T})\overline{\Lambda}(m,\boldsymbol{T},t)\le\overline{M}(\boldsymbol{T})\overline{\Lambda}(m,\boldsymbol{T},t)$$
$$\le\begin{cases}\Psi(t)^{j-i}, & \text{if }j>m;\\ \Psi(t)^{j-i}\alpha(t)^{m-j}, & \text{if }m\ge j.\end{cases}$$

Substituting this inequality into (3.14) and noting that E_s is superfluous,

$$E_eC_i(e)\le\sum_{j=i}^{\infty}\sum_{\boldsymbol{T}\in\mathbb{T}(i,j)}\Psi(t)^{j-i}\left(j-i+\sum_{h=0}^{\infty}\alpha(t)^h\right),$$
$$\text{all }t\ge0. \tag{3.15}$$

Recalling that the number of types in $\mathbb{T}(i,j)$ equals $(j-i+2)^n$,

$$E_eC_i(e)\le\sum_{j=i}^{\infty}(j-i+2)^n\Psi(t)^{j-i}\left(j-i+\sum_{h=0}^{\infty}\alpha(t)^h\right)$$
$$=\sum_{j=0}^{\infty}(j+2)^n\Psi(t)^j\left(j+\sum_{h=0}^{\infty}\alpha(t)^h\right),$$
$$\text{all }t\ge0. \tag{3.16}$$

The right side of (3.16) will be finite if $\Psi(t)<1$ and $\alpha(t)<1$. We have thus obtained the following sufficient conditions on achievability.[9]

[9] It is possible to obtain a stronger version of Theorem 1 in which $\alpha(t)<1$ is relaxed to $\alpha(t)\le1$ by following Gallager's proof for the single-user case [13, appendix 6B] more closely than we did here.

Theorem 1: A point R is achievable by Γ if, for an ensemble $\mathrm{Ens}(M, \boldsymbol{Q})$ of codes with rate $\geq R$, there exists a $t > 0$ such that for each nonempty $T \subset \{1, \cdots, n\}$,

$$\alpha(t) < 1 \qquad M(T)\beta(T,t) < 1 \qquad M(T)\sigma(T,t) < 1.$$

Proof: Let R be given, and suppose that an ensemble exists as in the statement of the theorem. Then for this ensemble, by (3.1) and (3.16), $E_e D_L(e, \Gamma) < A$, where A denotes the right side of (3.16) and is finite and independent of L. It can be concluded that, for any given L, the ensemble contains a code $\boldsymbol{e}$ for which $D_L(\boldsymbol{e}, \Gamma) < A$.

IV. The New Metric

In this section, we shall consider a parametric family of metrics and find an inner bound to the union of their achievable rate regions. Both the metric and the inner bound region are generalizations of the ones that were introduced in Section I, and the reader may find it helpful to go through the following arguments first by setting $n = 2$.

A. The Metric

The metric parameters are of the form $(\boldsymbol{Q}, B)$ where $\boldsymbol{Q} = (\boldsymbol{Q}_1, \cdots, \boldsymbol{Q}_n)$ and $B = \{B(T)\}$. Here $\boldsymbol{Q}_i$ is a p.d. on X_i^k for each $i = 1, \cdots, n$, and some integer $k \geq 1$ (the same k for each i). The integer k corresponds to the number of symbols per branch in the tree codes to which this metric is applicable; in other words, in decoding a tree code with k symbols per branch, we can only use those metrics for which $\boldsymbol{Q}_i$ is a p.d. on X_i^k. The numbers $B(T)$ are called the *bias terms*, and one $B(T)$ exists for each nonempty subset T of $\{1, \cdots, n\}$.

For any such parameter $(\boldsymbol{Q}, B)$, the branch metric $\mu_{(\boldsymbol{Q}, B)}$ is defined as follows. For each $\boldsymbol{\eta} \in Y^k$ and $\boldsymbol{\xi} = \boldsymbol{\xi}_1 \times \cdots \times \boldsymbol{\xi}_n$, where $\boldsymbol{\xi}_i \in X_i^k$, $i = 1, \cdots, n$,

$$\mu_{(\boldsymbol{Q}, B)}(\boldsymbol{\xi}, \boldsymbol{\eta}) := \min_T \left\{ \mu_{(\boldsymbol{Q}, B), T}(\boldsymbol{\xi}, \boldsymbol{\eta}) \right\} \tag{4.1}$$

where the minimum is taken over all nonempty subsets of $\{1, \cdots, n\}$, and for each T we have

$$\mu_{(\boldsymbol{Q}, B), T}(\boldsymbol{\xi}, \boldsymbol{\eta}) := \ln \frac{\sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi})}}{\sum_{\boldsymbol{\zeta}_T} \boldsymbol{Q}(\boldsymbol{\zeta}_T) \sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}_{\bar{T}}, \boldsymbol{\zeta}_T)}} - kB(T). \tag{4.2}$$

Here, $\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi})$ and $\boldsymbol{Q}(\boldsymbol{\zeta}_T)$ are as defined in the previous sections. The notation $\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}_{\bar{T}}, \boldsymbol{\zeta}_T)$ is new; it denotes the conditional probability that $\boldsymbol{\eta}$ is received at the channel output given that the transmitted block at input i equals $\boldsymbol{\zeta}_i$ if $i \in T$ and $\boldsymbol{\xi}_i$ if $i \in \bar{T}$. The summation in (4.2) is over the Cartesian product of the sets $\{X_i\}_{i \in T}$.

The "minimum" in the definition of $\mu_{(\boldsymbol{Q}, B)}$ serves the following purpose. We want to have a metric that has a negative drift on each incorrect path, irrespective of the number of users that are followed correctly on that path. This is accomplished here by setting $\mu_{(\boldsymbol{Q}, B)} = \min\{\mu_{(\boldsymbol{Q}, B), T}\}$ and choosing $\mu_{(\boldsymbol{Q}, B), T}$ in such a way that it has a negative expectation on all branches on which only the users in set $\bar{T}$ are followed correctly. As will become clear in the following analysis, another major consideration in the selection of the functions $\mu_{(\boldsymbol{Q}, B), T}$ has been the form of the sufficiency conditions of Theorem 1.

B. Inner Bound for a Fixed Parameter

We now fix M, $\boldsymbol{Q}$, and B and consider the ensemble $\mathrm{Ens}(M, \boldsymbol{Q})$ and the metric $\mu_{(\boldsymbol{Q}, B)}$. The goal is to find an inner bound to the achievable rate region of $\mu_{(\boldsymbol{Q}, B)}$. We start the analysis with the following lemma, which is just a special case of Theorem 1 at $t = 1$.

Lemma 2: Suppose that for $\mathrm{Ens}(M, \boldsymbol{Q})$ and $\mu_{(\boldsymbol{Q}, B)}$ the following conditions are satisfied:

1) $\alpha(1) < 1$;
2) $M(T)\beta(T, 1) < 1$ for all nonempty $T \subset \{1, \cdots, n\}$;
3) $M(T)\sigma(T, 1) < 1$ for all nonempty $T \subset \{1, \cdots, n\}$.

Then $\mu_{(\boldsymbol{Q}, B)}$ achieves the point $(1/k)(\ln M_1, \cdots, \ln M_n)$.

Next we find upper bounds on $\alpha(1)$, $\beta(T, 1)$, and $\sigma(T, 1)$. These upper bounds will yield weaker but more readily applicable sufficient conditions on achievability than those in Lemma 2:

$$\begin{aligned}
\alpha(1) &= \sum_{\boldsymbol{\eta}} \sum_{\boldsymbol{\xi}} \boldsymbol{Q}(\boldsymbol{\xi}) \boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}) \exp\left\{ -\mu_{(\boldsymbol{Q}, B)}(\boldsymbol{\xi}, \boldsymbol{\eta}) \right\} \\
&\leq \sum_T \sum_{\boldsymbol{\eta}} \sum_{\boldsymbol{\xi}} \boldsymbol{Q}(\boldsymbol{\xi}) \boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}) \exp\left\{ -\mu_{(\boldsymbol{Q}, B), T}(\boldsymbol{\xi}, \boldsymbol{\eta}) \right\} \\
&= \sum_T \sum_{\boldsymbol{\eta}} \sum_{\boldsymbol{\xi}} \boldsymbol{Q}(\boldsymbol{\xi}) \boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}) \frac{\sum_{\boldsymbol{\zeta}_T} \boldsymbol{Q}(\boldsymbol{\zeta}_T) \sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}_{\bar{T}}, \boldsymbol{\zeta}_T)}}{\sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi})}} \\
&\quad \cdot \exp\{kB(T)\} \\
&= \sum_T \sum_{\boldsymbol{\eta}} \sum_{\boldsymbol{\xi}} \boldsymbol{Q}(\boldsymbol{\xi}) \sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi})} \sum_{\boldsymbol{\zeta}_T} \boldsymbol{Q}(\boldsymbol{\zeta}_T) \sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}_{\bar{T}}, \boldsymbol{\zeta}_T)} \\
&\quad \cdot \exp\{kB(T)\} \\
&= \sum_T \sum_{\boldsymbol{\eta}} \sum_{\boldsymbol{\xi}_{\bar{T}}} \boldsymbol{Q}(\boldsymbol{\xi}_{\bar{T}}) \sum_{\boldsymbol{\xi}_T} \boldsymbol{Q}(\boldsymbol{\xi}_T) \sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}_{\bar{T}}, \boldsymbol{\xi}_T)} \sum_{\boldsymbol{\zeta}_T} \boldsymbol{Q}(\boldsymbol{\zeta}_T) \\
&\quad \cdot \sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}_{\bar{T}}, \boldsymbol{\zeta}_T)} \exp\{kB(T)\} \\
&= \sum_T \sum_{\boldsymbol{\eta}} \sum_{\boldsymbol{\xi}_{\bar{T}}} \boldsymbol{Q}(\boldsymbol{\xi}_{\bar{T}}) \left\{ \sum_{\boldsymbol{\xi}_T} \boldsymbol{Q}(\boldsymbol{\xi}_T) \sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi})} \right\}^2 \\
&\quad \cdot \exp\{kB(T)\} \\
&= \sum_T \exp\{-k[R_0(\boldsymbol{Q}, T) - B(T)]\}
\end{aligned} \tag{4.3}$$

where

$$R_0(\boldsymbol{Q}, T) := -(1/k) \ln \sum_{\boldsymbol{\eta}} \sum_{\boldsymbol{\xi}_{\bar{T}}} \boldsymbol{Q}(\boldsymbol{\xi}_{\bar{T}}) \left\{ \sum_{\boldsymbol{\xi}_T} \boldsymbol{Q}(\boldsymbol{\xi}_T) \sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi})} \right\}^2.$$

This is the desired upper bound on $\alpha(1)$.

In the following expressions we will use the notation $\mu_{(\boldsymbol{Q}, B)}(\boldsymbol{\zeta}_T, \boldsymbol{\xi}_{\bar{T}}, \boldsymbol{\eta})$ to denote the value of $\mu_{(\boldsymbol{Q}, B)}$ for the case where the branch label for user i equals $\boldsymbol{\zeta}_i$ if $i \in T$ and $\boldsymbol{\xi}_i$ if

$i \in \bar{T}$.

$$
\begin{aligned}
\sigma(T,1) &= \sum_{\xi} \sum_{\zeta_T} \sum_{\eta} \boldsymbol{Q}(\boldsymbol{\xi})\boldsymbol{Q}(\boldsymbol{\zeta}_T)\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}) \\
&\quad \cdot \exp\left\{\mu_{(Q,B)}(\boldsymbol{\zeta}_T, \boldsymbol{\xi}_{\bar{T}}, \boldsymbol{\eta}) - \mu_{(Q,B)}(\boldsymbol{\xi}, \boldsymbol{\eta})\right\} \\
&\le \sum_{\xi} \sum_{\zeta_T} \sum_{\eta} \boldsymbol{Q}(\boldsymbol{\xi})\boldsymbol{Q}(\boldsymbol{\zeta}_T)\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}) \\
&\quad \cdot \exp\left\{\mu_{(Q,B),T}(\boldsymbol{\zeta}_T, \boldsymbol{\xi}_{\bar{T}}, \boldsymbol{\eta}) - \mu_{(Q,B)}(\boldsymbol{\xi}, \boldsymbol{\eta})\right\} \\
&= \sum_{\xi} \sum_{\zeta_T} \sum_{\eta} \boldsymbol{Q}(\boldsymbol{\xi})\boldsymbol{Q}(\boldsymbol{\zeta}_T)\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}) \\
&\quad \cdot \frac{\sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\zeta}_T, \boldsymbol{\xi}_{\bar{T}})}\exp\{-kB(T)\}}{\sum_{\tau_T} \boldsymbol{Q}(\boldsymbol{\tau}_T)\sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\tau}_T, \boldsymbol{\xi}_{\bar{T}})}} \\
&\quad \cdot \exp\left\{-\mu_{(Q,B)}(\boldsymbol{\xi}, \boldsymbol{\eta})\right\} \\
&= \exp\{-kB(T)\} \sum_{\xi} \sum_{\eta} \boldsymbol{Q}(\boldsymbol{\xi}) \\
&\quad \cdot \boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}) \frac{\sum_{\zeta_T} \boldsymbol{Q}(\boldsymbol{\zeta}_T)\sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\zeta}_T, \boldsymbol{\xi}_{\bar{T}})}}{\sum_{\tau_T} \boldsymbol{Q}(\boldsymbol{\tau}_T)\sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\tau}_T, \boldsymbol{\xi}_{\bar{T}})}} \\
&\quad \cdot \exp\left\{-\mu_{(Q,B)}(\boldsymbol{\xi}, \boldsymbol{\eta})\right\} \\
&= \exp\{-kB(T)\} \sum_{\xi} \sum_{\eta} \boldsymbol{Q}(\boldsymbol{\xi})\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}) \\
&\quad \cdot \exp\left\{-\mu_{(Q,B)}(\boldsymbol{\xi}, \boldsymbol{\eta})\right\} \\
&= \alpha(1)\exp\{-kB(T)\}; \qquad (4.4)
\end{aligned}
$$

$$
\begin{aligned}
\beta(T,1) &= \sum_{\xi} \sum_{\zeta_T} \sum_{\eta} \boldsymbol{Q}(\boldsymbol{\xi})\boldsymbol{Q}(\boldsymbol{\zeta}_T)\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}) \\
&\quad \cdot \exp\left\{\mu_{(Q,B)}(\boldsymbol{\zeta}_T, \boldsymbol{\xi}_{\bar{T}}, \boldsymbol{\eta})\right\} \\
&\le \sum_{\xi} \sum_{\zeta_T} \sum_{\eta} \boldsymbol{Q}(\boldsymbol{\xi})\boldsymbol{Q}(\boldsymbol{\zeta}_T)\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}) \\
&\quad \cdot \exp\left\{\mu_{(Q,B)}(\boldsymbol{\zeta}_T, \boldsymbol{\xi}_{\bar{T}}, \boldsymbol{\eta})\right\} \\
&= \sum_{\xi} \sum_{\zeta_T} \sum_{\eta} \boldsymbol{Q}(\boldsymbol{\xi})\boldsymbol{Q}(\boldsymbol{\zeta}_T)\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}) \\
&\quad \cdot \frac{\sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\zeta}_T, \boldsymbol{\xi}_{\bar{T}})}}{\sum_{\tau_T} \boldsymbol{Q}(\boldsymbol{\tau}_T)\sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\tau}_T, \boldsymbol{\xi}_{\bar{T}})}}\exp\{-kB(T)\} \\
&= \sum_{\xi} \sum_{\eta} \boldsymbol{Q}(\boldsymbol{\xi})\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}) \frac{\sum_{\zeta_T} \boldsymbol{Q}(\boldsymbol{\zeta}_T)\sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\zeta}_T, \boldsymbol{\xi}_{\bar{T}})}}{\sum_{\tau_T} \boldsymbol{Q}(\boldsymbol{\tau}_T)\sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\tau}_T, \boldsymbol{\xi}_{\bar{T}})}} \\
&\quad \cdot \exp\{-kB(T)\} \\
&= \exp\{-kB(T)\}. \qquad (4.5)
\end{aligned}
$$

As a result of (4.3)–(4.5), Lemma 2 can be restated in the following weaker but more useful form.

Lemma 3: Suppose that the following are true:

1) $\Sigma_T \exp\{-k[R_0(Q,T) - B(T)]\} < 1$;
2) $M(T)\exp\{-kB(T)\} < 1$, for each nonempty subset T of $\{1,\cdots,n\}$.

Then $\mu_{(Q,B)}$ achieves the point $(1/k)(\ln M_1,\cdots,\ln M_n)$.

Now we focus on the following particular choice for the bias terms, which turns out to be a satisfactory one. For every nonempty subset T of $\{1,\cdots,n\}$, let

$$B_M(T) := (1/2)\{R_0(Q,T) + (1/k)\ln M(T)\}.$$

We shall denote the metric with parameter (Q, B_M) by $\mu_{(M,Q)}$. Let

$$\delta(M,Q) := \min_T \{R_0(Q,T) - (1/k)\ln M(T)\}$$

where the minimum is over all nonempty subsets of $\{1,\cdots,n\}$.

Lemma 4: If $\delta(M,Q) > (2/k)\ln(2^n - 1)$, then $\mu_{(M,Q)}$ achieves the point $(1/k)(\ln M_1,\cdots,\ln M_n)$.

Proof: Suppose that $\delta(M,Q) > (2/k)\ln(2^n - 1)$. We only need to show that $\mu_{(M,Q)}$ satisfies the conditions of Lemma 3:

$$
\begin{aligned}
&\sum_T \exp\{-k[R_0(Q,T) - B_M(T)]\} \\
&\quad = \sum_T \exp\{-k[R_0(Q,T) - (R_0(Q,T) \\
&\qquad + (1/k)\ln M(T))/2]\} \\
&\quad \le \sum_T \exp\{-k\delta(M,Q)/2\} \\
&\quad < (2^n - 1)\exp\{-k\delta(M,Q)/2\} < 1.
\end{aligned}
$$

Thus the first condition is satisfied, and we have

$$
\begin{aligned}
&M(T)\exp\{-kB_M(T)\} \\
&\quad = M(T)\exp\{-k[R_0(Q,T) + (1/k)\ln M(T)]/2\} \\
&\quad = \exp\{-k[R_0(Q,T) - (1/k)\ln M(T)]/2\} \\
&\quad < \exp\{-k\delta(M,Q)/2\} < 1.
\end{aligned}
$$

The second condition is also satisfied, and the proof is complete.

We have thus shown that the region

$$\bar{\boldsymbol{R}}_0(Q) := \left\{(R_1,\cdots,R_n): \right. \\ \left. 0 \le \sum_{i \in T} R_i < R_0(Q,T) - (2/k)\ln(2^n - 1), \text{ all } T\right\}$$

is an inner bound to the achievable rate region of $\mu_{(M,Q)}$ for all (M,Q) for which $\delta(M,Q) > (2/k)\ln(2^n - 1)$.

C. The Union of $\bar{\boldsymbol{R}}_0(Q)$ Over All Q

We now seek an inner bound to the union, over all possible (M,Q), of the achievable rate region of $\mu_{(M,Q)}$.

We start with some definitions:

$$\mathbb{P}(k) := \{\boldsymbol{Q} = (Q_1,\cdots,Q_n)\colon Q_i \text{ is a p.d. on } X_i^k, \text{ all } i=1,\cdots,n\}$$

$$\overline{\boldsymbol{R}}_0(k) := \bigcup_{\boldsymbol{Q}\in\mathbb{P}(k)} \overline{\boldsymbol{R}}_0(\boldsymbol{Q}) \qquad \overline{\boldsymbol{R}}_0 := \bigcup_{k\geq 1} \overline{\boldsymbol{R}}_0(k)$$

$$\boldsymbol{R}_0(\boldsymbol{Q}) := \left\{(R_1,\cdots,R_n)\colon 0 \leq \sum_{i\in T} R_i < R_0(\boldsymbol{Q},T), \text{ all } T\right\}$$

$$\boldsymbol{R}_0(k) := \bigcup_{\boldsymbol{Q}\in\mathbb{P}(k)} \boldsymbol{R}_0(\boldsymbol{Q}) \qquad \boldsymbol{R}_0 := \bigcup_{k\geq 1} \boldsymbol{R}_0(k).$$

It is clear that $\overline{\boldsymbol{R}}_0(\boldsymbol{Q}) \subset \boldsymbol{R}_0(\boldsymbol{Q})$, $\overline{\boldsymbol{R}}_0(k) \subset \boldsymbol{R}_0(k)$, $\overline{\boldsymbol{R}}_0 \subset \boldsymbol{R}_0$, and

$$\overline{\boldsymbol{R}}_0 \subset \bigcup_{\text{all }(M,\boldsymbol{Q})} \{\text{achievable rate region of } \mu_{(M,\boldsymbol{Q})}\}. \quad (4.6)$$

The remainder of this section is devoted to proving the following result.

Theorem 2: We have

$$\boldsymbol{R}_0 \subset \bigcup_{\text{all }(M,\boldsymbol{Q})} \{\text{achievable rate region of } \mu_{(M,\boldsymbol{Q})}\}. \quad (4.7)$$

We prove (4.7) by showing that $\boldsymbol{R}_0 = \overline{\boldsymbol{R}}_0$. The arguments involved are trivial corollaries to Lemma 5, stated next.

For any finite set X, any p.d. Q on X, and any integer $m \geq 1$, let $Q^{(m)}$ denote the (product-form) p.d. on X^m such that for any $\xi = (\xi_1,\cdots,\xi_m) \in X^m$,

$$Q^{(m)}(\xi) = Q(\xi_1)\cdots Q(\xi_m).$$

For any k, m, and $\boldsymbol{Q} = (Q_1,\cdots,Q_n) \in \mathbb{P}(k)$, let the parameter $\boldsymbol{Q}^{(m)} \in \mathbb{P}(mk)$ be defined by $\boldsymbol{Q}^{(m)} := (Q_1^{(m)},\cdots,Q_n^{(m)})$.

Lemma 5: We have

$$R_0(\boldsymbol{Q},T) = R_0(\boldsymbol{Q}^{(m)},T), \quad \text{for all possible } \boldsymbol{Q},\ T, \text{ and } m.$$

The proof is straightforward and will be omitted. We will also omit the proof of the following corollary.

Corollary 1: We have the following:

1) $\overline{\boldsymbol{R}}_0(\boldsymbol{Q}) \subset \overline{\boldsymbol{R}}_0(\boldsymbol{Q}^{(m)})$;
2) $\overline{\boldsymbol{R}}_0(k) \subset \overline{\boldsymbol{R}}_0(km)$;
3) $\boldsymbol{R}_0(\boldsymbol{Q}) = \boldsymbol{R}_0(\boldsymbol{Q}^{(m)})$;
4) $\boldsymbol{R}_0(k) \subset \boldsymbol{R}_0(km)$.

These statements hold for all possible values of $\boldsymbol{Q}$, m, and k.

It follows from the corollary that $\overline{\boldsymbol{R}}_0(k!)$ includes all of the sets $\overline{\boldsymbol{R}}_0(h)$, $1 \leq h \leq k$. Also, the difference between $\boldsymbol{R}_0(k)$ and $\overline{\boldsymbol{R}}_0(k)$ vanishes in the limit as k goes to infinity. Therefore, $\overline{\boldsymbol{R}}_0 = \boldsymbol{R}_0$, and the proof of Theorem 2 is complete.

D. Discussion of Results

1) This section has shown that $\boldsymbol{R}_0 \subset \boldsymbol{R}_{\text{comp}}$ for all MAC's. To complement this result, we note that it is not known if a channel exists for which the statement

$$\boldsymbol{R}_{\text{comp}} \subset \text{closure of } \boldsymbol{R}_0 \quad (4.8)$$

is false. We should also note that the set (closure of $\boldsymbol{R}_0 \backslash \boldsymbol{R}_0$) consists of only those points that lie on the "outer surface" of $\boldsymbol{R}_0$. More precisely, for any point $(R_1,\cdots,R_n)$ in this set, and any $\epsilon > 0$, a point exists in $\boldsymbol{R}_0$ which is componentwise $\geq (R_1-\epsilon, R_2-\epsilon,\cdots,R_n-\epsilon)$. Therefore, if (4.8) is satisfied for a channel, then $\boldsymbol{R}_{\text{comp}}$ and $\boldsymbol{R}_0$ for that channel can be regarded as essentially the same.

As pointed out in Section I, (4.8) is true for all single-user channels. Another class of channels for which (4.8) is true is, as proved in [9], the class of pairwise-reversible channels. A MAC $(P; X_1,\cdots,X_n; Y)$ is said to be pairwise reversible [15] if, for each $\xi_i, \zeta_i \in X_i$, $i=1,\cdots,n$,

$$\sum_{\eta\in Y} \sqrt{P(\eta|\xi_1,\cdots,\xi_n)P(\eta|\zeta_1,\cdots,\zeta_n)} \cdot \log\frac{P(\eta|\xi_1,\cdots,\xi_n)}{P(\eta|\zeta_1,\cdots,\zeta_n)} = 0.$$

An example of a pairwise-reversible channel is the two-user erasure channel of Fig. 3. In fact, pairwise-reversible channels include all such noiseless channels, i.e., channels for which the transition probabilities are either 0 or 1.

For noiseless channels it is known that $\boldsymbol{R}_{\text{comp}} = \boldsymbol{R}_0$. Furthermore, all points in $\boldsymbol{R}_{\text{comp}}$ are achieved by a simple branch metric which assigns the value 0 to a branch that is consistent with the channel output and $-\infty$ to one that is inconsistent.

2) A further result relating to the regions $\boldsymbol{R}_0$ and $\boldsymbol{R}_{\text{comp}}$ is the following. For any ensemble $\text{Ens}(M,\boldsymbol{Q})$ with $\delta(M,\boldsymbol{Q}) < 0$,

$$E_{\boldsymbol{e}} D_L(\boldsymbol{e},\Gamma) \to \infty \quad \text{as} \quad L \to \infty,$$

regardless of what the metric Γ is. Here the expectation is over all codes in $\text{Ens}(M,\boldsymbol{Q})$, and Γ need not be a branchwise additive metric.

This is a result about "typical" codes. It does not imply that $\boldsymbol{R}_{\text{comp}} \subset$ closure $\boldsymbol{R}_0$ for any MAC. However, this result is significant for at least one reason. It shows that $\boldsymbol{R}_0$ is essentially the largest region that can be proven to be an inner bound to $\boldsymbol{R}_{\text{comp}}$ by using random-coding arguments over the class of ensembles we have considered in this paper. For a proof of this result and further discussions, refer to [9].

3) The previous results suggest the following procedure for finding, if they exist, a suitable code-metric pair for achieving a desired rate $R = (R_1,\cdots,R_n)$. First try to find a parameter $(M,\boldsymbol{Q})$ such that $\delta(M,\boldsymbol{Q}) > 0$ and the rate of codes in $\text{Ens}(M,\boldsymbol{Q})$ is $\geq R$. (Unfortunately, no practical algorithm is known for finding such an $(M,\boldsymbol{Q})$ or determining that none exists.) Supposing that such an $(M,\boldsymbol{Q})$ has been found, let m be the smallest integer such that $\delta(M^{(m)},\boldsymbol{Q}^{(m)}) > (2/(mk))\ln(2^n-1)$, where k is the block length for $\boldsymbol{Q}$, (i.e., $\boldsymbol{Q} \in \mathbb{P}(k)$), and $M^{(m)} = (M_1^m,\cdots,M_n^m; km)$. Select a code at random from $\text{Ens}(M^{(m)},\boldsymbol{Q}^{(m)})$ and use $\mu_{(M^{(m)},\boldsymbol{Q}^{(m)})}$ as the metric.

The probability that the average computation for a code selected at random from an ensemble is more than twice the average computation over all codes in the ensemble is smaller than $1/2$. Therefore, significant assurance exists that a randomly constructed code will not be far worse than typical.

We have suggested the use of the smallest m satisfying $\delta(M^{(m)}, Q^{(m)}) > (2/(mk))\ln(2^n - 1)$ because the complexity of each step in the stack algorithm (also in other sequential decoding algorithms) is proportional to the degree of the tree code, and the degree increases exponentially with the number of symbols per branch.

4) The metric in sequential decoding can be regarded as a likelihood ratio for testing the hypothesis H_0: "the branch is on the correct path" against the alternative H_1: "the branch is on an incorrect path." H_0 is a simple hypothesis for all n. On the other hand, H_1 is a composite hypothesis consisting of $2^n - 1$ p.d.'s (one p.d. for each possible way the current branch may have diverged from the correct path).

From this point of view, the additional difficulties in multiuser sequential decoding can be attributed to the fact that testing a simple H_0 against a composite H_1 is an inherently more difficult problem than testing a simple H_0 against a simple H_1, which is the case for $n = 1$. This point of view can also be used to explain why we have been able to prove larger inner bounds for metrics over longer branches. The number of symbols per branch of the tree code corresponds to the number of samples in the hypothesis testing framework. Therefore, as the branch length increases, it becomes easier to distinguish H_0 from each p.d. in H_1. Clearly, it is of crucial importance that the number $2^n - 1$ of p.d.'s in H_1 is independent of the branch length.

V. Some Remarks on the Regions $\boldsymbol{R}_0$ and $\boldsymbol{R}_{\text{comp}}$

Here, we list some properties of the regions $\boldsymbol{R}_0$ and $\boldsymbol{R}_{\text{comp}}$ and some open problems relating to them.

1) No algorithm is known for determining whether or not a given point belongs to $\boldsymbol{R}_{\text{comp}}$; neither is one known for $\boldsymbol{R}_0$. In fact, the only known general characterization of $\boldsymbol{R}_{\text{comp}}$ is its definition.

2) It is not known if $\boldsymbol{R}_{\text{comp}}$ is convex. Note that the possibility of time-sharing between two different tree codes and decoding each independently does not imply the convexity of $\boldsymbol{R}_{\text{comp}}$. This is because, in general, the operation of two independent sequential decoders cannot be simulated by a single sequential decoder. Also note that it may be possible to prove the convexity of $\boldsymbol{R}_{\text{comp}}$ (if indeed it is convex) without actually having to find an analytical characterization of $\boldsymbol{R}_{\text{comp}}$.

3) For $n = 1$, we have $\boldsymbol{R}_0 = \boldsymbol{R}_0(1)$ by Gallager's parallel channels theorem [13, pp. 149–150], and $\boldsymbol{R}_0$ is easily determined. However, for $n \geq 2$, no such single letter characterization of $\boldsymbol{R}_0$ is known. In fact, there are multiuser channels for which $\boldsymbol{R}_0 \neq \boldsymbol{R}_0(1)$. An example of such a channel is the two-user m-ary collision channel $(P; X_1, X_2; Y)$, where $X_1 = X_2 = \{0, 1, \cdots, m-1\}$, $Y = \{e, 0, 1, \cdots, m-1\}$, and

$$\begin{aligned} &P(x_1|x_1, 0) = P(x_2|0, x_2) = 1, && \text{all } x_1 \in X_1,\ x_2 \in X_2,\\ &P(e|x_1, x_2) = 1, && \text{all } x_1 \in \{1, \cdots, m-1\},\\ & && x_2 \in \{1, \cdots, m-1\}. \end{aligned}$$

A simple calculation shows that the point (1.5 bits, 1.5 bits) lies outside $\boldsymbol{R}_0(1)$. $\boldsymbol{R}_0(2)$ contains all points (R, R) such that $0 \leq R < (1/2)\log_2(m)$ bits. Therefore, for $m > 8$, $\boldsymbol{R}_0(2)$ is strictly larger than $\boldsymbol{R}_0(1)$. It can also be shown that, for any r, r-user collision channels exist for which $\boldsymbol{R}_0 \neq \boldsymbol{R}_0(1) \cup \cdots \cup \boldsymbol{R}_0(r)$.

4) The region $\boldsymbol{R}_0$ is convex. To see this, let $\boldsymbol{Q} = (\boldsymbol{Q}_1, \cdots, \boldsymbol{Q}_n) \in \mathbb{P}(k)$ and $\overline{\boldsymbol{Q}} = (\overline{\boldsymbol{Q}}_1, \cdots, \overline{\boldsymbol{Q}}_n) \in \mathbb{P}(j)$ be arbitrary parameters. Let m_1 and m_2 be arbitrary integers. Consider the parameter $\boldsymbol{H} = (\boldsymbol{H}_1, \cdots, \boldsymbol{H}_n) \in \mathbb{P}(m_1 k + m_2 j)$ such that

$$\boldsymbol{H}_i = \boldsymbol{Q}_i^{(m_1 k)} \times \overline{\boldsymbol{Q}}_i^{(m_2 j)}, \qquad \text{all } i.$$

In other words, $\boldsymbol{H}_i$ is a product-form distribution with $m_1 k$ "copies" of $\boldsymbol{Q}_i$ and $m_2 j$ "copies" of $\overline{\boldsymbol{Q}}_j$. Convexity of $\boldsymbol{R}_0$ is an immediate consequence of the following relationship:

$$(m_1 + m_2) R_0(\boldsymbol{H}, T) = m_1 R_0(\boldsymbol{Q}, T) + m_2 R_0(\overline{\boldsymbol{Q}}, T), \qquad \text{all } T.$$

While $\boldsymbol{R}_0$ is convex, given any r we can find a channel (e.g., an r-user collision channel) for which $\boldsymbol{R}_0(r)$ is not convex.

It is not known if $\boldsymbol{R}_0 =$ convex-hull $\boldsymbol{R}_0(1)$ for all MAC's. If this were true, we would then have a characterization of $\boldsymbol{R}_0$ similar to that for the capacity region [5], [6].

5) If $T \subset S$, then $R_0(\boldsymbol{Q}, T) \leq R_0(\boldsymbol{Q}, S)$.

Proof: Let $\boldsymbol{Q} \in \mathbb{P}(k)$. Then

$$\begin{aligned} &kR_0(\boldsymbol{Q}, S)\\ &\quad = -\ln \sum_{\eta} \sum_{\xi_{\bar{S}}} Q(\xi_{\bar{S}}) \Big\{ \sum_{\xi_S} Q(\xi_S) \sqrt{\boldsymbol{P}(\eta|\xi)} \Big\}^2\\ &\quad = -\ln \sum_{\eta} \sum_{\xi_{\bar{S}}} Q(\xi_{\bar{S}}) \Big\{ \sum_{\xi_{S\setminus T}} Q(\xi_{S\setminus T}) \sum_{\xi_T} Q(\xi_T) \sqrt{\boldsymbol{P}(\eta|\xi)} \Big\}^2\\ &\quad \geq -\ln \sum_{\eta} \sum_{\xi_{\bar{S}}} Q(\xi_{\bar{S}}) \sum_{\xi_{S\setminus T}} Q(\xi_{S\setminus T}) \Big\{ \sum_{\xi_T} Q(\xi_T) \sqrt{\boldsymbol{P}(\eta|\xi)} \Big\}^2\\ &\quad = -\ln \sum_{\eta} \sum_{\xi_{\bar{T}}} Q(\xi_{\bar{S}}) \Big\{ \sum_{\xi_T} Q(\xi_T) \sqrt{\boldsymbol{P}(\eta|\xi)} \Big\}^2\\ &\quad = kR_0(\boldsymbol{Q}, T) \end{aligned}$$

where the third step follows by Jensen's inequality.

Note that the proof works for $T = \varnothing$ as well. Hence $R_0(\boldsymbol{Q}, S) \geq 0$.

6) For any subset T of users, let

$$\boldsymbol{P}(\eta|\xi_T) := \sum_{\xi_{\bar{T}}} Q(\xi_{\bar{T}}) \boldsymbol{P}(\eta|\xi).$$

The quantity $\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}_T)$ can be interpreted as the transition probability between the users in the set T and the receiver, supposing that the users in $\overline{T}$ transmit $\boldsymbol{\xi}_{\overline{T}}$ with probability $\boldsymbol{Q}(\boldsymbol{\xi}_{\overline{T}})$. If one is only interested in decoding the messages of the users in T, then one may model the remaining users as noise sources and thus obtain a reduced channel with these transition probabilities. Such schemes have been studied in [9]. The following inequality is of interest in comparing the achievable rates for the reduced channel with those for the original one. For any $\boldsymbol{Q} \in \mathbb{P}(k)$,

$$-\ln \sum_{\boldsymbol{\eta}} \left\{ \sum_{\boldsymbol{\xi}_T} \boldsymbol{Q}(\boldsymbol{\xi}_T) \sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}_T)} \right\}^2 \le kR_0(\boldsymbol{Q}, T).$$

Proof: We have

$$\begin{aligned} kR_0(\boldsymbol{Q},T) &= -\ln \sum_{\boldsymbol{\eta}} \sum_{\boldsymbol{\xi}_{\overline{T}}} \boldsymbol{Q}(\boldsymbol{\xi}_{\overline{T}}) \left\{ \sum_{\boldsymbol{\xi}_T} \boldsymbol{Q}(\boldsymbol{\xi}_T) \sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}_T)} \right\}^2 \\ &= -\ln \sum_{\boldsymbol{\eta}} \sum_{\boldsymbol{\xi}_{\overline{T}}} \left\{ \sum_{\boldsymbol{\xi}_T} \boldsymbol{Q}(\boldsymbol{\xi}_T) \sqrt{\boldsymbol{Q}(\boldsymbol{\xi}_{\overline{T}}) \boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi})} \right\}^2 \\ &\ge -\ln \sum_{\boldsymbol{\eta}} \left\{ \sum_{\boldsymbol{\xi}_T} \boldsymbol{Q}(\boldsymbol{\xi}_T) \sqrt{\sum_{\boldsymbol{\xi}_{\overline{T}}} \boldsymbol{Q}(\boldsymbol{\xi}_{\overline{T}}) \boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi})} \right\}^2 \\ &= -\ln \sum_{\boldsymbol{\eta}} \left\{ \sum_{\boldsymbol{\xi}_T} \boldsymbol{Q}(\boldsymbol{\xi}_T) \sqrt{\boldsymbol{P}(\boldsymbol{\eta}|\boldsymbol{\xi}_T)} \right\}^2 \end{aligned}$$

where the third step follows by Minkowsky's inequality.

VI. Strong Achievability

The achievability concept used in the sequential decoding literature coincides with the one given in Section II. There is a disturbing point about this definition, however: the code $\boldsymbol{e}$ is allowed to depend on L. Instead, we would like to see achievability defined as follows. A point $R = (R_1, \cdots, R_n)$ is said to be *strongly achievable* by a metric Γ if a code $\boldsymbol{e}$ exists with rate $\ge R$ such that for all L,

$$D_L(\boldsymbol{e}, \Gamma) < A$$

where A is a finite number independent of L.

Thus for achievability in this stronger sense, we need a code $\boldsymbol{e}$ that is uniformly good for all L. Obviously, any point that is strongly achievable is also achievable. It is not known, however, if the converse is true. The following theorem provides a partial answer.

Theorem 3: All points in $\boldsymbol{R}_0$ are strongly achievable.

Proof: For any $R = (R_1, \cdots, R_n) \in \boldsymbol{R}_0$, by the results of Section IV, an ensemble $\mathrm{Ens}(M, \boldsymbol{Q})$ of codes exists with rate $\ge R$ such that

$$E_{\boldsymbol{e}} C_i(\boldsymbol{e}, \Gamma) < A, \qquad \text{all } i \ge 1,$$

where Γ denotes the metric of Section IV and A is a finite number independent of i.

If we regard $\boldsymbol{e}$ as a random code from $\mathrm{Ens}(M, \boldsymbol{Q})$, then $C_1(\boldsymbol{e}, \Gamma), C_2(\boldsymbol{e}, \Gamma), \cdots$ is a stationary sequence of random variables. By Birkhoff's ergodic theorem [16, p. 113], as $L \to \infty$, $D_L(\boldsymbol{e}, \Gamma)$ converges (almost surely) to a random variable $D(\boldsymbol{e}, \Gamma)$ such that $E_{\boldsymbol{e}} D(\boldsymbol{e}, \Gamma) = E_{\boldsymbol{e}} C_1(\boldsymbol{e}, \Gamma)$. Thus one has

$$E_{\boldsymbol{e}} \lim_{L \to \infty} D_L(\boldsymbol{e}, \Gamma) = E_{\boldsymbol{e}} D(\boldsymbol{e}, \Gamma) = E_{\boldsymbol{e}} C_1(\boldsymbol{e}, \Gamma).$$

Therefore a code $\bar{\boldsymbol{e}}$ must exist in $\mathrm{Ens}(M, \boldsymbol{Q})$ such that $\limsup D_L(\bar{\boldsymbol{e}}, \Gamma) \le E_{\boldsymbol{e}} C_1(\boldsymbol{e}, \Gamma)$. Now we show that this implies $\sup\{D_L(\bar{\boldsymbol{e}}, \Gamma) \colon L \ge 1\} < \infty$.

Suppose to the contrary that $\sup\{D_L(\bar{\boldsymbol{e}}, \Gamma) \colon L \ge 1\} = \infty$. There must then be some finite L such that $D_L(\bar{\boldsymbol{e}}, \Gamma) = \infty$, implying that $C_i(\bar{\boldsymbol{e}}, \Gamma) = \infty$ for some i, $1 \le i \le L$. This implies in turn that $D_j(\bar{\boldsymbol{e}}, \Gamma) = \infty$ for all $j \ge i$, a contradiction.

VII. Summary

A new metric has been introduced for sequential decoding on MAC's, and an inner bound $\boldsymbol{R}_0$ to the achievable rate region of this metric has been found. The region $\boldsymbol{R}_0$ is large enough to suggest that sequential decoding is a worthwhile decoding algorithm for MAC's.

Several open problems regarding the regions $\boldsymbol{R}_{\text{comp}}$ and $\boldsymbol{R}_0$ have been presented. The main open problem is whether $\boldsymbol{R}_{\text{comp}}$ is strictly larger than the closure of $\boldsymbol{R}_0$ for any MAC. Two other interesting open problems are whether $\boldsymbol{R}_{\text{comp}}$ is convex and whether $\boldsymbol{R}_0 = \text{convex-hull } \boldsymbol{R}_0(1)$. A shortcoming of the common concept of achievability in sequential decoding has been pointed out, and it has been proven that points in $\boldsymbol{R}_0$ are also achievable in a stronger sense.

Acknowledgment

I am greatly indebted to Prof. Robert G. Gallager for his help and support that made this work possible.

References

[1] J. M. Wozencraft, "Sequential decoding for reliable communications," Mass. Inst. Technol., Cambridge, Tech. Rep. 325, RLE, 1957.

[2] R. M. Fano, "A heuristic discussion of probabilistic decoding," *IEEE Trans. Inform. Theory*, vol. IT-9, pp. 64–74, 1963.

[3] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math. Stat. Prob.*, 1961, vol. 1, pp. 611–644.

[4] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 124–142, Mar. 1985.

[5] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Inform. Theory*, Tsahkadsor, USSR, 1971.

[6] H. Liao, "A coding theorem for multiple access communications," in *Proc. Int. Symp. Information Theory*, Asilomar, CA, 1972.

[7] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell Syst. Tech. J.*, vol. 52, pp. 1037–1076, Sept. 1973.

[8] R. L. Peterson and D. J. Costello, "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, Nov. 1980.

[9] E. Arikan, "Sequential decoding for multiple access channels," Ph.D. dissertation, Electrical Engineering and Computer Science Dept., Mass. Inst. Technol., Cambridge, Nov. 1985.

[10] ____, "An upper bound on the cutoff rate of sequential decoding," *IEEE Trans. Inform. Theory*, vol. 34, pp. 55–63, Jan. 1988.

[11] K. Zigangirov, "Some sequential decoding procedures," *Probl. Peredach. Inform.*, vol. 2, pp. 13–25, 1966.

[12] F. Jelinek, "A fast sequential decoding algorithm using a stack," *IBM J. Res. Develop.*, vol. 13, pp. 675–685, 1969.

[13] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[14] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.

[15] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels, Parts I and II," *Inform. Contr.*, vol. 10, pp. 65–103 and 522–552, 1967.

[16] L. Breiman, *Probability*. Reading, MA: Addison-Wesley, 1968.