# An Inequality on Guessing and its Application to Sequential Decoding

Erdal Arikan, *Senior Member, IEEE*

***Abstract*—Let $(X,Y)$ be a pair of discrete random variables with $X$ taking one of $M$ possible values. Suppose the value of $X$ is to be determined, given the value of $Y$, by asking questions of the form "Is $X$ equal to $x$?" until the answer is "Yes." Let $G(x \mid y)$ denote the number of guesses in any such guessing scheme when $X = x$, $Y = y$. We prove that**

$$E[G(X \mid Y)^\rho] \geq (1 + \ln M)^{-\rho} \sum_y \left[ \sum_x P_{X,Y}(x,y)^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

**for any $\rho \geq 0$. This provides an operational characterization of Rényi's entropy. Next we apply this inequality to the estimation of the computational complexity of sequential decoding. For this, we regard $X$ as the input, $Y$ as the output of a communication channel. Given $Y$, the sequential decoding algorithm works essentially by guessing $X$, one value at a time, until the guess is correct. Thus the computational complexity of sequential decoding, which is a random variable, is given by a guessing function $G(X \mid Y)$ that is defined by the order in which nodes in the tree code are hypothesized by the decoder. This observation, combined with the above lower bound on moments of $G(X \mid Y)$, yields lower bounds on moments of computation in sequential decoding. The present approach enables the determination of the (previously known) cutoff rate of sequential decoding in a simple manner; it also yields the (previously unknown) cutoff rate region of sequential decoding for multiaccess channels. These results hold for memoryless channels with finite input alphabets.**

***Index Terms*—Guessing, Hölder's inequality, sequential decoding, Rényi's entropy.**

## I. Introduction

MASSEY [1] considered the problem of guessing the value of a realization of a random variable $X$ by asking questions of the form "Is $X$ equal to $x$?" until the answer is "Yes." Let $G(x)$ denote the number of guesses required by a particular guessing strategy when $X = x$. Massey observed that $\boldsymbol{E}[G(X)]$, the average number of guesses, is minimized by a guessing strategy that guesses the possible values of $X$ in decreasing order of probability. The primary concern in [1] was to discover a relationship between the minimum possible value of $\boldsymbol{E}[G(X)]$ and the Shannon entropy of $X$. The aim in this paper is to give a tight lower bound on $\boldsymbol{E}[G(X)^\rho]$ for $\rho \geq 0$ and apply this bound to the estimation of the computational complexity of sequential decoding. This paper extends and improves the results of [2].

We begin by giving a formal and generalized statement of the above problem. Let $(X,Y)$ be a pair of random variables with $X$ taking values in a finite set $\mathcal{X}$ of size $M$, $Y$ taking values in a countable set $\mathcal{Y}$. Call a function $G(X)$ of the random variable $X$ a *guessing function for $X$* if $G: \mathcal{X} \to \{1, \cdots, M\}$ is one-to-one. Call a function $G(X \mid Y)$ a *guessing function for $X$ given $Y$* if, for any fixed value $Y = y$, $G(X \mid y)$ is a guessing function for $X$. $G(X \mid Y)$ will be thought of as the number of guesses required to determine $X$ when the value of $Y$ is given. The following inequalities on the moments of $G(X)$ and $G(X \mid Y)$, proved in Section II, are the main results of this paper.

*Theorem 1:* For arbitrary guessing functions $G(X)$ and $G(X \mid Y)$, and any $\rho \geq 0$

$$\boldsymbol{E}[G(X)^\rho] \geq (1 + \ln M)^{-\rho} \left[ \sum_{x \in \mathcal{X}} P_X(x)^{\frac{1}{1+\rho}} \right]^{1+\rho} \tag{1}$$

and

$$\boldsymbol{E}[G(X \mid Y)^\rho] \geq (1 + \ln M)^{-\rho} \sum_{y \in \mathcal{Y}} \left[ \sum_{x \in \mathcal{X}} P_{X,Y}(x,y)^{\frac{1}{1+\rho}} \right]^{1+\rho} \tag{2}$$

where $P_{X,Y}$, $P_X$ are the probability distributions of $(X,Y)$ and $X$, respectively.

In Section II we define optimal guessing functions and show that Theorem 1 estimates their $\rho$th moment correctly to within a factor of $(1 + \ln M)^\rho$ for any $\rho \geq 0$. There, we also point out a connection between Rényi's entropy and moments of guessing functions.

For information-theoretic applications of Theorem 1, we think of $(X,Y)$ as the input and output of a communication system. In this context, $X$ represents the transmitted message, $Y$ the observation using which the receiver estimates $X$. $G(X \mid Y)$ is then the number of guesses that a hypothetical decision device would make until determining $X$ given $Y$. For example, if the decision device is allowed to make only one guess, as ordinarily is the case, then the event $G(X \mid Y) > 1$ signifies a decision error. For list-$\ell$ decoding an error occurs if $G(X \mid Y) > \ell$.

In this paper we shall be interested only in the type of decision devices known as *sequential decoders* which, in effect, keep guessing the value of $X$, one at a time, until the guess is correct. The computational complexity of

Manuscript received September 8, 1994; revised August 2, 1995. The material in this paper was presented in part at the 12th Prague Conference on Information Theory Statistical Decision Functions and Random Processes, Prague, the Czech Republic, August 1994.

The author is with the Electrical–Electronics Engineering Department, Bilkent University, 06533 Ankara, Turkey.

Publisher Item Identifier S 0018-9448(96)00033-8.

sequential decoding, which is a random variable, is given by the guessing function $G(X \mid Y)$ defined by the decoding process. Thus Theorem 1 yields lower bounds on the moments of computation in sequential decoding. In Section III, we use this approach and determine the cutoff rate (respectively, cutoff rate region) of sequential decoding for single-user (respectively, two-user multiaccess) memoryless channels with finite input alphabets. The present derivations simplify proofs of some known results on cutoff rates and in certain cases establish new results. A full discussion of the contribution of the present paper in this regard will be given in Section III.

## II. Bounds on Moments of the Number of Guesses

We shall use the notation $P_{X,Y}(x,y)$, $P_X(x)$, $P_Y(y)$, $P_{X|Y}(x \mid y)$, and $P_{Y|X}(y \mid x)$ to denote, respectively, the joint, marginal, and conditional probability distributions for the pair $(X,Y)$. When no confusion can arise, we shall omit the subscripts.

### A. Proof of Theorem 1

Let $Q$ be an arbitrary probability distribution on $\mathcal{X}$. We have

$$\begin{aligned} \boldsymbol{E}[G(X)^\rho] &= \sum_x P(x)G(x)^\rho \\ &= \sum_x Q(x)\exp\left[-\ln\frac{Q(x)}{P(x)G(x)^\rho}\right] \\ &\geq \exp\left[-D(Q \parallel P) + \rho\sum_x Q(x)\ln G(x)\right] \quad (3) \end{aligned}$$

where

$$D(Q \parallel P) = \sum_x Q(x)\ln Q(x)/P(x)$$

is the relative entropy function, and Jensen's inequality is used to obtain (3). Now

$$\begin{aligned} \sum_x Q(x)\ln G(x) &= H(Q) - \sum_x Q(x)\ln\frac{1}{Q(x)G(x)} \\ &\geq H(Q) - \ln\sum_x \frac{1}{G(x)} \quad (4) \\ &= H(Q) - \ln\sum_{i=1}^{M} 1/i \quad (5) \end{aligned}$$

where

$$H(Q) = -\sum_x Q(x)\ln Q(x)$$

is the entropy function, and we have used Jensen's inequality once again to obtain (4). Combining (3) and (5) and noting that

$$\sum_{i=1}^{M} 1/i \leq 1 + \ln M$$

we get

$$\boldsymbol{E}[G^\rho] \geq (1+\ln M)^{-\rho}\exp\left[-D(Q \parallel P) + \rho H(Q)\right] \quad (6)$$

Substitution of

$$Q(x) = \frac{P(x)^{\frac{1}{1+\rho}}}{\sum_{x'} P(x')^{\frac{1}{1+\rho}}} \quad (7)$$

into (6) yields Inequality (1).[1]

Inequality (2) follows readily

$$\begin{aligned} \boldsymbol{E}[G(X \mid Y)^\rho] &= \sum_y P(y)\boldsymbol{E}[G(X \mid Y=y)^\rho] \\ &\geq \sum_y P(y)(1+\ln M)^{-\rho}\left[\sum_x P(x \mid y)^{\frac{1}{1+\rho}}\right]^{1+\rho} \\ &= (1+\ln M)^{-\rho}\sum_y\left[\sum_x P(x,y)^{\frac{1}{1+\rho}}\right]^{1+\rho}. \end{aligned}$$

This completes the proof of Theorem 1. It should be clear from the above proof that the theorem can be generalized to the case where $Y$ is a continuous random variable.

While the above proof has the merit of showing the information-theoretic aspect of the guessing problem, a direct proof can be given using the following variant of Hölder's inequality.

*Lemma 1:* Let $a_i$, $p_i$ be nonnegative numbers indexed over a finite set $1 \leq i \leq M$. For any $0 < \lambda < 1$

$$\sum_{i=1}^{M} a_i p_i \geq \left[\sum_{i=1}^{M} a_i^{\frac{-\lambda}{1-\lambda}}\right]^{\frac{1-\lambda}{-\lambda}}\left[\sum_{i=1}^{M} p_i^\lambda\right]^{\frac{1}{\lambda}}.$$

*Proof:* Put $A_i = a_i^{-\lambda}$, $B_i = a_i^\lambda p_i^\lambda$, in Hölder's inequality

$$\sum_i A_i B_i \leq \left[\sum_i A_i^{\frac{1}{1-\lambda}}\right]^{1-\lambda}\left[\sum_i B_i^{\frac{1}{\lambda}}\right]^{\lambda}.$$

An alternative proof of (1) is obtained by taking $a_i = i^\rho$, $p_i = \Pr[G(X) = i]$, and $\lambda = 1/(1+\rho)$ in the lemma.

Let us write $G(X_1,\cdots,X_k \mid Y_1,\cdots,Y_n)$ to denote a function for guessing the value of a joint realization of a number of random variables $X_1,\cdots,X_k$ when the values of $Y_1,\cdots,Y_n$ are known. The above framework covers such cases by taking $X$ and $Y$ as random vectors, $X = (X_1,\cdots,X_k)$, $Y = (Y_1,\cdots,Y_n)$. Theorem 1, stated explicitly, now gives

$$\begin{aligned} \boldsymbol{E}[G(X_1,\cdots,X_k \mid Y_1,\cdots,Y_n)^\rho] &\geq [1+\ln(M_1\cdots M_k)]^{-\rho} \\ &\quad \cdot \exp E_\rho(X_1,\cdots,X_k \mid Y_1,\cdots,Y_n) \end{aligned}$$

where we have defined $M_i$ as the number of possible values of $X_i$, $i = 1,\cdots,k$, and

$$\begin{aligned} &E_\rho(X_1,\cdots,X_k \mid Y_1,\cdots,Y_n) \\ &\quad = \ln\sum_{y_1,\cdots,y_n}\left[\sum_{x_1,\cdots,x_k} P(x_1,\cdots,x_k,y_1,\cdots,y_n)^{\frac{1}{1+\rho}}\right]^{1+\rho}. \end{aligned}$$

The function $E_\rho$ will be useful in expressing the bound in a compact form. As discussed later in this section, $E_\rho/\rho$ equals Rényi's entropy of order $1/(1+\rho)$; so, $E_\rho$ has the properties expected of information measures. We shall state only two such properties that will be used later in the paper.

[1] This choice of $Q$ actually maximizes $\rho H(Q) - D(Q\|P)$ but this need not be proved here.

*Proposition 1:* If $X_1, \cdots, X_n$ are independent, identically distributed (i.i.d.), then

$$E_\rho(X_1, \cdots, X_n) = nE_\rho(X_1).$$

More generally, if $(X_1, Y_1), \cdots, (X_n, Y_n)$ are i.i.d., then

$$E_\rho(X_1, \cdots, X_n \mid Y_1, \cdots, Y_n) = nE_\rho(X_1 \mid Y_1).$$

The proof is straightforward and will be omitted.

*Proposition 2:* For any $k \geq 1$, $n \geq 1$, $\rho > 0$

$$E_\rho(X_1, \cdots, X_{k-1} | Y_1, \cdots, Y_n) \leq E_\rho(X_1 \cdots, X_k | Y_1, \cdots, Y_n) \leq E_\rho(X_1, \cdots, X_k | Y_1, \cdots, Y_{n-1}). \quad (8)$$

*Proof:* For the left inequality in (8), we give the proof of only the special case $E_\rho(X_1) \leq E_\rho(X_1, X_2)$. The general proof follows in the same manner.

$$\begin{aligned} E_\rho(X_1, X_2) &= \ln \left[ \sum_{x_1, x_2} P(x_1, x_2)^{\frac{1}{1+\rho}} \right]^{1+\rho} \\ &= \ln \left[ \sum_{x_1} P(x_1)^{\frac{1}{1+\rho}} \sum_{x_2} P(x_2 \mid x_1)^{\frac{1}{1+\rho}} \right]^{1+\rho} \\ &\geq \ln \left[ \sum_{x_1} P(x_1)^{\frac{1}{1+\rho}} \right]^{1+\rho} \quad (9) \\ &= E_\rho(X_1) \end{aligned}$$

where (9) follows by noting that

$$\sum_{x_2} P(x_2 \mid x_1)^{\frac{1}{1+\rho}} \geq 1, \quad \text{for } \rho \geq 0.$$

For the right inequality in (8), we only prove $E_\rho(X \mid Y) \leq E_\rho(X)$; the general proof is similar. (This inequality was proved earlier by Arimoto [3] in his work on Rényi's entropy.)

$$\begin{aligned} E_\rho(X|Y) &= \ln \sum_y \left[ \sum_x P(x, y)^{\frac{1}{1+\rho}} \right]^{1+\rho} \\ &= E_\rho(X) + \ln \sum_y \left[ \sum_x Q(x) P(y \mid x)^{\frac{1}{1+\rho}} \right]^{1+\rho} \\ &\leq E_\rho(X) + \ln \left\{ \sum_x Q(x) \left[ \sum_y P(y \mid x) \right]^{\frac{1}{1+\rho}} \right\}^{1+\rho} \quad (10) \\ &= E_\rho(X) \end{aligned}$$

where $Q$ is the distribution in (7), and (10) follows by Minkowsky's inequality (specifically, by [4, p. 524, inequality (h)]).

In the remainder of this section we define optimal guessing functions and give an upper bound which complements Theorem 1. We also point out a connection between moments of guessing functions and Rényi's entropy. Section III can be read independently of the rest of this section.

### B. Optimal Guessing

We begin by observing that, for any $\rho \geq 0$

$$\boldsymbol{E}[G(X \mid Y)^\rho] = \sum_y P(y) \sum_x P(x \mid y) G(x \mid y)^\rho$$

is minimized by a guessing function $G(X \mid Y)$ for which $G(x \mid y) < G(x' \mid y)$ implies $P(x \mid y) \geq P(x' \mid y)$, for all possible $x, x', y$. (Otherwise, interchanging the order in which $x$ and $x'$ are guessed when $Y = y$ would decrease the value of $\boldsymbol{E}[G(X \mid Y)^\rho]$.) Thus all nonnegative moments of $G(X \mid Y)$ are minimized simultaneously by a guessing function which guesses the possible values of $X$, when $Y = y$, in decreasing order of *a posteriori* probabilities $P(x \mid y)$. Such guessing functions will be called *optimal.*

It is easy to see that there exists a unique optimal guessing function $G(X \mid Y)$ if and only if, for any possible value $Y = y$, the probability distribution $P_{X|Y}(\cdot \mid y)$ assigns distinct probabilities to the possible values of $X$. It is also easy to see that, even if uniqueness does not hold, all optimal $G(X \mid Y)$ are equal in distribution. Hence, references to statistical properties of optimal guessing functions will be unambiguous.

For arbitrary real-valued random variables $U$, $V$, let us write $U \prec V$ if the condition $\Pr[U \geq t] \leq \Pr[V \geq t]$ holds for all $t$.

The following result ranks the difficulty of guessing in various situations.

*Proposition 3:* For any positive integers $k$, $n$, and any choice of random variables $X_1, \cdots, X_k, Y_1, \cdots, Y_n$, optimal guessing functions satisfy

$$G^*(X_1, \cdots, X_{k-1} | Y_1, \cdots, Y_n) \prec G^*(X_1, \cdots, X_k | Y_1, \cdots, Y_n) \prec G^*(X_1, \cdots, X_k | Y_1, \cdots, Y_{n-1}). \quad (11)$$

*Proof:* For the left part of (11), we give the proof of only the special case $G^*(X_1) \prec G^*(X_1, X_2)$ to keep the notation simple. The general proof is similar. Given an optimal guessing function $G^*(X_1, X_2)$, let $G(X_1)$ be the guessing function for $X_1$ defined by the condition that $G(x_1) < G(x_1')$ if and only if $\min_{x_2}\{G^*(x_1, x_2)\} < \min_{x_2}\{G^*(x_1', x_2)\}$. That is, $G(X_1)$ guesses the possible values of $X_1$ in the order in which they are first guessed by $G^*(X_1, X_2)$, disregarding the guess about $X_2$. Then, $G(x_1) \leq G^*(x_1, x_2)$ for all $x_2$ and, hence, $G(X_1) \prec G^*(X_1, X_2)$. Since $G^*(X_1) \prec G(X_1)$, the proof is complete.

The right part of (11) follows by observing that any guessing function $G(X_1, \cdots, X_k \mid Y_1, \cdots, Y_{n-1})$ is a valid guessing function for $X_1, \cdots, X_k$ given $Y_1, \cdots, Y_n$ (we may simply ignore $Y_n$).

*Corollary 1:* Optimal guessing functions satisfy, for all $\rho \geq 0$

$$\begin{aligned} &\boldsymbol{E}[G^*(X_1, \cdots, X_{k-1} \mid Y_1, \cdots, Y_n)^\rho] \\ &\quad \leq \boldsymbol{E}[G^*(X_1, \cdots, X_k \mid Y_1, \cdots, Y_n)^\rho] \\ &\quad \leq \boldsymbol{E}[G^*(X_1, \cdots, X_k \mid Y_1, \cdots, Y_{n-1})^\rho]. \quad (12) \end{aligned}$$

This follows from the following formula (see, e.g., [5]) for the moments of a random variable $U$ taking positive integer

values:

$$\boldsymbol{E}[U^t] = \sum_{k=1}^{\infty} [k^t - (k-1)^t] \Pr[U \geq k].$$

Next we show that Theorem 1 is tight to within a factor of $(1 + \ln M)^\rho$ for optimal guessing functions.

*Proposition 4:* For any optimal guessing function $G^*(X \mid Y)$, and $\rho \geq 0$

$$\boldsymbol{E}[G^*(X \mid Y)^\rho] \leq \exp E_\rho(X \mid Y). \tag{13}$$

*Proof:* For an optimal $G^*(X \mid Y)$, we have

$$\begin{aligned} G^*(x \mid y) &= \sum_{x':G^*(x'|y) \leq G^*(x|y)} 1 \\ &\leq \sum_{x':G^*(x'|y) \leq G^*(x|y)} [P(x' \mid y)/P(x \mid y)]^{\frac{1}{1+\rho}} \\ &\leq \sum_{\text{all } x'} [P(x' \mid y)/P(x \mid y)]^{\frac{1}{1+\rho}}. \end{aligned}$$

The proof is completed as follows:

$$\begin{aligned} &\boldsymbol{E}[G^*(X \mid Y)^\rho] \\ &= \sum_y P(y) \sum_x P(x \mid y) G^*(x \mid y)^\rho \\ &\leq \sum_y P(y) \sum_x P(x \mid y) \left[ \sum_{x'} [P(x' \mid y)/P(x \mid y)]^{\frac{1}{1+\rho}} \right]^\rho \\ &= \sum_y P(y) \left[ \sum_x P(x \mid y)^{\frac{1}{1+\rho}} \right]^{1+\rho} \\ &= \sum_y \left[ \sum_x P(x,y)^{\frac{1}{1+\rho}} \right]^{1+\rho}. \end{aligned}$$

### C. Relation to Rényi's Entropy

Rényi's entropy of order $\alpha$ ($\alpha > 0$, $\alpha \neq 1$) for a discrete random variable $X$ is defined as [6]

$$H_\alpha(X) = \frac{\alpha}{1-\alpha} \ln \left[ \sum_x P(x)^\alpha \right]^{1/\alpha}.$$

Following Arimoto [3], we define Rényi's conditional entropy of order $\alpha$ for $X$ given $Y$ as

$$H_\alpha(X \mid Y) = \frac{\alpha}{1-\alpha} \ln \sum_y \left[ \sum_x P(x,y)^\alpha \right]^{1/\alpha}.$$

Noting the relations

$$E_\rho(X) = \rho H_{\frac{1}{1+\rho}}(X)$$

and

$$E_\rho(X \mid Y) = \rho H_{\frac{1}{1+\rho}}(X \mid Y)$$

the preceding bounds on moments of guessing functions can be written in terms of Rényi's entropy functions. Of particular interest is the following result which gives an operational characterization to Rényi's entropy.

*Proposition 5:* Let $X_1, \cdots, X_n$ be a sequence of i.i.d. random variables over a finite set. Let $G^*(X_1, \cdots, X_n)$ be an optimal guessing function. Then, for any $\rho > 0$

$$\lim_{n\to\infty} \frac{1}{n} \ln \left( \boldsymbol{E}[G^*(X_1, \cdots, X_n)^\rho] \right)^{1/\rho} = H_{\frac{1}{1+\rho}}(X_1).$$

More generally, let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be i.i.d., and $G^*(X_1, \cdots, X_n \mid Y_1, \cdots, Y_n)$ be an optimal guessing function. Then, for any $\rho > 0$

$$\lim_{n\to\infty} \frac{1}{n} \ln \left( \boldsymbol{E}[G^*(X_1, \cdots, X_n \mid Y_1, \cdots, Y_n)^\rho] \right)^{1/\rho} = H_{\frac{1}{1+\rho}}(X_1 \mid Y_1).$$

The proof follows directly from Theorem 1, Proposition 1, and Proposition 4.

In light of the above result, the quantity

$$H_{\frac{1}{1+\rho}}(X) - H_{\frac{1}{1+\rho}}(X \mid Y),$$

which Arimoto [3] called the mutual information of order $1/(1+\rho)$, can be interpreted as a kind of complexity reduction, provided by the knowledge of $Y$, in guessing the value of $X$. Note that, by Proposition 2, this quantity is nonnegative. (In fact, it equals zero if and only if $X$, $Y$ are independent.)

Alternative operational characterizations of Rényi's entropy were given by Arimoto [3] and Csiszár [7].

## III. Application to Sequential Decoding

### A. Single-User Channels

Sequential decoding is a search algorithm invented by Wozencraft [8] for finding the transmitted path through a tree code. Well-known versions of sequential decoding are due to Fano [9], Zigangirov [10], and Jelinek [11].

The computational effort in sequential decoding is a random variable, depending on the transmitted sequence, the received sequence, and the exact search algorithm. The following connection between guessing and sequential decoding, due to Jacobs and Berlekamp [5], makes it possible to lowerbound the moments of computation in sequential decoding by applying the lower bound of Theorem 1.

Consider an arbitrary tree code and let $\mathcal{X}$ denote the set of nodes at some fixed but arbitrary level, $N$ channel symbols into the tree from the origin. Let $X$ be a random variable uniformly distributed on $\mathcal{X}$. We think of $X$ as the node in $\mathcal{X}$ which lies on the transmitted path. Abusing the notation, we also let $X$ denote the channel input sequence of length $N$ from the origin to node $X$. We let $Y$ denote the channel output sequence that is received when $X$ is transmitted.

Any sequential decoder, applied to this code, begins its search at the origin and extends it branch by branch eventually to examine a node $x'$ in $\mathcal{X}$, possibly going on to explore nodes beyond $x'$. We assume that if $X \neq x'$, i.e., if $x'$ does not lie on the transmitted path, the decoder, with the aid of its metric, will eventually retrace its steps back to below level $N$ and proceed to examine a second node $x''$ in $\mathcal{X}$. If $X \neq x''$, eventually a third node in $\mathcal{X}$ will be examined, and so on. We assume that with probability one the sequential decoder sooner

or later examines the correct node $X$. (Though this is never the case in practice, the probability of decoding error can be made arbitrarily small by using tree codes with sufficiently large constraint lengths.) If $X$ is not among the first $M-1$ nodes examined (not counting multiple visits to a node more than once[2]), the decoder will examine all $M$ nodes at level $N$. Thus for any given $Y = y$, we have an ordering of the nodes in $\mathcal{X}$, namely, that in which they are examined by the decoder. We let $G(x \mid y)$ denote the position of $x \in \mathcal{X}$ in this ordering when $Y = y$. (By definition of sequential decoding, the value $G(x \mid y)$ is well-defined in the sense that, for any fixed sequential decoder and fixed tree code, the order in which node $x \in \mathcal{X}$ is examined does not depend on the portion of the received sequence beyond level $N$; it depends only on $y$.)

Clearly, $G(\cdot \mid \cdot)$ is a guessing function and $G(x \mid y)$ equals the number of nodes in $\mathcal{X}$ examined before and including the correct node $X = x$ when $Y = y$ is received. Thus $G(X \mid Y)$ is a lower bound to the computation performed by the decoder in decoding the first $N$ symbols of the transmitted sequence. Lower bounds to moments of $G(X \mid Y)$ serve as lower bounds to moments of computation in sequential decoding.

In the remainder of this section, we assume that $X$ and $Y$ are connected by a discrete memoryless channel. The channel has a finite input alphabet $\mathcal{I}$, a countable output alphabet $\mathcal{J}$, and transition probability matrix $V(j \mid i)$, $j \in \mathcal{J}$, $i \in \mathcal{I}$. The conditional probability of $Y$ given $X$ is then $P_{Y|X}(y \mid x) = V_N(y \mid x)$ where $V_N$ denotes channel transition probability assignment for sequences of length $N$. Since the channel is memoryless

$$V_N(y \mid x) = \prod_{n=1}^{N} V(y_n \mid x_n)$$

where $y_n$, $x_n$ are the $n$th coordinates of the sequences $x$ and $y$, respectively. As stated above, we assume that $X$ is uniformly distributed over $\mathcal{X}$, the set of possible values of $X$; i.e., $P(x) = 1/M$ for $x \in \mathcal{X}$ where $M$ denotes the size of $\mathcal{X}$. Letting $R$ denote the rate, in nats per channel symbol, of the underlying tree code, the size of $\mathcal{X}$ is given by $M = \exp NR$.

Now consider an arbitrary sequential decoder with a guessing function $G(X \mid Y)$ for the above situation. By Theorem 1, for $\rho > 0$

$$\boldsymbol{E}[G(X \mid Y)^{\rho}] \geq (1 + NR)^{-\rho} \exp E_{\rho}(X \mid Y).$$

Since $P_X$ is a uniform distribution, we have the relation

$$E_{\rho}(X \mid Y) = \rho NR - E_0(\rho, P_X)$$

where

$$E_0(\rho, P_X) = -\ln \sum_{y} \left[ \sum_{x} P_X(x) V_N(y \mid x)^{\frac{1}{1+\rho}} \right]^{1+\rho}.$$

The function $E_0(\rho, \cdot)$ was introduced by Gallager [12] in his work on bounding the probability of error in block coding. Gallager examined properties of this function in detail and, in particular, showed that [12, Theorem 5] for any probability distribution $Q_N$ on $\mathcal{I}^N$

$$E_0(\rho, Q_N) \leq N E_0(\rho)$$

where $E_0(\rho)$ is defined as the maximum of

$$E_0(\rho, Q) = -\ln \sum_{j} [\sum_{i} Q(i) V(j \mid i)^{\frac{1}{1+\rho}}]^{1+\rho}$$

over all probability distributions $Q$ on $\mathcal{I}$. Thus

$$E_{\rho}(X \mid Y) \geq \rho NR - N E_0(\rho)$$

and we have proved that, for $\rho > 0$

$$\boldsymbol{E}[G(X \mid Y)^{\rho}] \geq (1 + NR)^{-\rho} \exp N[\rho R - E_0(\rho)]. \quad (14)$$

Thus at rates $R > E_0(\rho)/\rho$, the $\rho$th moment of computation performed at level $N$ of the tree code must go to infinity exponentially as $N$ is increased. The infimum of all real numbers $R'$ such that, at rates $R > R'$, $\boldsymbol{E}[G(X \mid Y)^{\rho}]$ must go to infinity as $N$ is increased is called the cutoff rate (for the $\rho$th moment) and denoted by $R_{\text{cutoff}}(\rho)$. We have thus obtained the following bound.

*Theorem 2:* For any discrete memoryless channel with a finite input alphabet

$$R_{cutoff}(\rho) \leq E_0(\rho)/\rho, \quad \rho > 0. \quad (15)$$

The converse inequality

$$R_{\text{cutoff}}(\rho) \geq E_0(\rho)/\rho, \quad \rho > 0 \quad (16)$$

has been proved in the works of Falconer [13], Savage [14], Jelinek [15], and Hashimoto and Arimoto [16]. We conclude that $R_{\text{cutoff}}(\rho) = E_0(\rho)/\rho$ for all $\rho > 0$.

Previous upper bounds on $R_{\text{cutoff}}(\rho)$ were given by Jacobs and Berlekamp [5], and Arikan [17]–[19]. In [5], it is shown that

$$R_{\text{cutoff}}(\rho) \leq \hat{E}_0(\rho)/\rho, \quad \rho > 0 \quad (17)$$

where $\hat{E}_0(\rho)$ is the concave hull of $E_0(\rho)$. Since there are channels for which $\hat{E}_0(\rho) > E_0(\rho)$ (see, e.g., the example in [18]), in general the bound (17) is loose. Inequality (15) was proved in [18] for $\rho = 1$, and in [19] for all $\rho > 0$.

The result (15) is not new; however, the present proof is much simpler and direct than the previous ones. The approaches in [5], [17]–[19] for upperbounding $R_{\text{cutoff}}(\rho)$ all rely on lower bounds on the probability of error for block codes and are considerably more complicated. Moreover, as the next section shows, the preceding proof easily extends to the case of multiaccess channels, determining their previously unknown cutoff rate region.

Finally, let us note that the restriction in the above discussion that the channel output alphabet $\mathcal{J}$ be countable has been made only for notational convenience; the result can be extended to channels with continuous output alphabets.

[2]Fano's version may examine a node more than once. The stack algorithm version, due to Zigangirov and Jelinek, examines a node at most once.

### B. Multiaccess Channels

We consider a triple of random variables $(X_1, X_2, Y)$ where $X_1$, $X_2$ are the inputs to a two-user multiaccess channel and $Y$ the channel output. Here, $X_1$, $X_2$ stand for the correct nodes at level $N$ of the respective tree codes for users 1 and 2, and $Y$ denotes the received channel output when $(X_1, X_2)$ is transmitted. A sequential decoder in this case carries out a search on the joint tree code (which is the product of the individual tree codes) and is identified by a guessing function $G(X_1, X_2 \mid Y)$ for purposes of lowerbounding its computational complexity. For a detailed description of sequential decoding for multiaccess channels, we refer to [20].

We assume the channel is memoryless with finite input alphabets $\mathcal{I}_1, \mathcal{I}_2$, a countable output alphabet $\mathcal{J}$, and transition probability matrix $V(j \mid i_1, i_2)$, $i_1 \in \mathcal{I}_1$, $i_2 \in \mathcal{I}_2$, $j \in \mathcal{J}$. We assume $X_1$, $X_2$, $Y$ are sequences of length $N$ over $\mathcal{I}_1$, $\mathcal{I}_2$, $\mathcal{J}$, respectively. We denote the set of possible values of $X_1$ (respectively, $X_2$) by $\mathcal{X}_1$ (respectively, $\mathcal{X}_2$), and the size of this set by $M_1$ (respectively, $M_2$). Letting $R_1$, $R_2$ denote the rates, in nats per channel symbol, of the tree codes for users 1 and 2, respectively, we have $M_1 = \exp NR_1$ and $M_2 = \exp NR_2$.

We assume the random variables $X_1$, $X_2$ are independent and uniformly distributed over $\mathcal{X}_1$, $\mathcal{X}_2$. (That is, the messages by the two users are independent and equiprobable.) The conditional probability of $Y$ given $X_1$, $X_2$ is given by $P(y \mid x_1, x_2) = V_N(y \mid x_1, x_2)$ where $V_N$ is the transition probability matrix for sequences of length $N$. By the memoryless channel assumption

$$V_N(y \mid x_1, x_2) = \prod_{n=1}^{N} V(y_n \mid x_{1n}, x_{2n})$$

where $y_n$, $x_{1n}$, $x_{2n}$ denote the $n$th coordinates of $y$, $x_1$, $x_2$, respectively.

For $k \geq 1$ and $Q_1$, $Q_2$ arbitrary probability distributions on $\mathcal{I}_1^k$, $\mathcal{I}_2^k$, respectively, define

$$E_0(\rho, Q_1Q_2) = -\ln \sum_y \left[ \sum_{x_1, x_2} Q_1(x_1)Q_2(x_2)V_k(y \mid x_1, x_2)^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

$$E_0(\rho, Q_1 \mid Q_2) = -\ln \sum_y \sum_{x_2} Q_2(x_2) \left[ \sum_{x_1} Q_1(x_1)V_k(y \mid x_1, x_2)^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

$$E_0(\rho, Q_2 \mid Q_1) = -\ln \sum_y \sum_{x_1} Q_1(x_1) \left[ \sum_{x_2} Q_2(x_2)V_k(y \mid x_1, x_2)^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

where the summations are over all possible values of the indices.

Define $\mathcal{R}_0(\rho)$ as the closure of the set of all pairs $(r_1, r_2)$ such that, for some $k \geq 1$ and some pair of probability distributions $Q_1$ on $\mathcal{I}_1^k$, $Q_2$ on $\mathcal{I}_2^k$

$$0 \leq r_1 \leq k^{-1}E_0(\rho, Q_1 \mid Q_2)/\rho$$
$$0 \leq r_2 \leq k^{-1}E_0(\rho, Q_2 \mid Q_1)/\rho$$
$$r_1 + r_2 \leq k^{-1}E_0(\rho, Q_1Q_2)/\rho.$$

(No single-letter characterization of this region is known.)

Now consider an arbitrary sequential decoder with a guessing function $G(X_1, X_2 \mid Y)$ for the above two-user channel. By Theorem 1, we have, for any $\rho > 0$

$$\boldsymbol{E}[G(X_1, X_2 \mid Y)^\rho] \geq [1 + N(R_1 + R_2)]^{-\rho} \cdot \exp E_\rho(X_1, X_2 \mid Y). \quad (18)$$

By Proposition 2, we have[3]

$$E_\rho(X_1, X_2 \mid Y) \geq E_\rho(X_1 \mid X_2, Y) \quad (19)$$
$$E_\rho(X_1, X_2 \mid Y) \geq E_\rho(X_2 \mid X_1, Y). \quad (20)$$

It is easy to verify that (since $P_{X_1}$ and $P_{X_2}$ are uniform)

$$E_\rho(X_1, X_2 \mid Y) = \rho N(R_1 + R_2) - E_0(\rho, P_{X_1}P_{X_2})$$
$$E_\rho(X_1 \mid X_2, Y) = \rho NR_1 - E_0(\rho, P_{X_1} \mid P_{X_2})$$
$$E_\rho(X_2 \mid X_1, Y) = \rho NR_2 - E_0(\rho, P_{X_2} \mid P_{X_1}).$$

Thus if $(R_1, R_2)$ does not belong to $\mathcal{R}_0(\rho)$, then at least one of the terms $E_\rho(X_1, X_2 \mid Y)$, $E_\rho(X_1 \mid X_2, Y)$, $E_\rho(X_2 \mid X_1, Y)$ is greater than $N\epsilon$ where $\epsilon > 0$ is a constant that depends on $(R_1, R_2)$ and $\mathcal{R}_0(\rho)$ but not on $N$. This, combined with (18)–(20), implies that, at rates $(R_1, R_2)$ outside the region $\mathcal{R}_0(\rho)$, $\boldsymbol{E}[G(X_1, X_2 \mid Y)^\rho]$ must go to infinity exponentially as the sequence length $N$ is increased. The infimum (i.e., closure of the intersection) of all sets $\mathcal{R}'$ of pairs of positive real numbers $(r_1, r_2)$ such that, at rates outside $\mathcal{R}'$, $\boldsymbol{E}[G(X_1, X_2 \mid Y)^\rho]$ must go to infinity is called the cutoff rate region (for the $\rho$th moment) and denoted by $\mathcal{R}_{\text{cutoff}}(\rho)$. Summarizing the above discussion, we have

*Theorem 3:* For any memoryless two-user multiaccess channel with finite input alphabets, $\mathcal{R}_{cutoff}(\rho) \subseteq \mathcal{R}_0(\rho)$, for all $\rho > 0$.

This result is new. Although the proof has been given for a two-user channel, it should be clear that it can be generalized to multiaccess channels with an arbitrary number of users. It should also be clear that the proof can be generalized to channels with continuous output alphabets. Such a result was previously proved only for $\rho = 1$ and only for the restricted class of pairwise-reversible channels by Arikan [17], [21].

For $\rho = 1$, the converse result $\mathcal{R}_{\text{cutoff}}(1) \supseteq \mathcal{R}_0(1)$ was first proved by Arikan [17], [20]. Recently, Balakirsky [22] proved that $\mathcal{R}_{\text{cutoff}}(\rho) \supseteq \mathcal{R}_0(\rho)$ for all $\rho > 0$.

Thus for multiaccess channels with finite input alphabets it is established that the cutoff rate region $\mathcal{R}_{\text{cutoff}}(\rho)$ equals $\mathcal{R}_0(\rho)$ for all $\rho > 0$.

## ACKNOWLEDGMENT

The author wishes to thank J. L. Massey and M. Burnashev for discussions on this problem.

## REFERENCES

[1] J. L. Massey, "Guessing and entropy," in *Proc. 1994 IEEE Int. Symp. on Information Theory* (Trondheim, Norway, 1994), p. 204.

[2] E. Arikan, "On the average number of guesses required to determine the value of a random variable," in *Proc. 12th Prague Conf. on Information Theory Statistical Decision Functions and Random Processes* (Prague, the Czech Republic, Aug. 29–Sept. 2, 1994), pp. 20–23.

[3] *Proof:* $E_\rho(X_1, X_2 \mid Y) \geq E_\rho(X_1, X_2 \mid X_2, Y) \geq E_\rho(X_1 \mid X_2, Y)$.

[3] S. Arimoto, "Information measures and capacity of order $\alpha$ for discrete memoryless channels," in *Topics in Information Theory (Colloquia Math. Soc. J. Bolyai)*, vol. 16, I. Csiszár and P. Elias, Eds. Amsterdam, The Netherlands: North Holland, 1977, pp. 41–52.
[4] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[5] I. M. Jacobs and E. R. Berlekamp, "A lowerbound to the distribution of computation for sequential decoding," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 167–174, Apr. 1967.
[6] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. on Math. Statist. Probability* (Berkeley, CA, 1961), vol. 1, pp. 547–561.
[7] I. Csiszár, "Generalized cutoff rates and Rényi's information measures," *IEEE Trans. Inform. Theory*, vol. 41, pp. 26–34, Jan. 1995.
[8] J. M. Wozencraft, "Sequential decoding for reliable communications," Tech. Rep. 325, RLE, MIT, Cambridge, MA, 1957.
[9] R. M. Fano, "A heuristic discussion of sequential decoding," *IEEE Trans. Inform. Theory*, vol. IT-9, pp. 66–74, Jan. 1963.
[10] K. Zigangirov, "Some sequential decoding procedures," *Probl. Pered. Inform.*, vol. 2, pp. 13–25, 1966.
[11] F. Jelinek, "A fast sequential decoding algorithm using a stack," *IBM J. Res. Devel.*, vol. 13, pp. 675–685, 1969.
[12] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 3–18, Jan. 1965.
[13] D. D. Falconer, "A hybrid coding scheme for discrete memoryless channels," *Bell Syst. Tech. J.*, vol. 48, pp. 691–728, Mar. 1969.
[14] J. E. Savage, "Sequential decoding the computation problem," *Bell Syst. Tech. J.*, vol. 45, pp. 149–175, 1966.
[15] F. Jelinek, "An upper bound on moments of sequential decoding effort," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 140–149, Jan. 1969.
[16] T. Hashimoto and S. Arimoto, "Computational moments for sequential decoding of convolutional codes," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 584–591, Sept. 1979.
[17] E. Arikan, "Sequential decoding for multiple access channels," Ph.D. dissertation, MIT, Cambridge, MA, Nov. 1985.
[18] ——, "An upper bound on the cutoff rate of sequential decoding," *IEEE Trans. Inform. Theory*, vol. 34, pp. 55–63, Jan. 1988.
[19] ——,"Lower bounds to moments of list size," in *Abstract of Papers, IEEE Int. Symp. on Information Theory* (San Diego, CA, Jan. 14–19, 1990), pp. 145–146.
[20] ——, "Sequential decoding for multiple access channels," *IEEE Trans. Inform. Theory*, vol. 34, pp. 246–259, Mar. 1988.
[21] ——, "On the achievable rate region of sequential decoding for a class of multiaccess channels," *IEEE Trans. Inform. Theory*, vol. 36, pp. 180–183, Jan. 1990.
[22] V. B. Balakirsky, "An upper bound on the distribution of computation of a sequential decoder for multiple access channels," in *Proc. 6th Swedish-Russian Int. Workshop on Information Theory* (Mölle, Sweden, Aug. 22–27, 1993), pp. 179–189.