Proof: $P_m^- = 2^{-m}\sum_{j\in I_m^-}2^{\log p_j+m} < 2^{-m}\sum_{j\in I_m^-}2^{-l_j} \le 2^{-m}$. The last inequality follows from the Kraft inequality. □

Lemma: Any node with property X_m^+ has probability $p < 2^{-c(m-1)}$ where $c = (1-\log g)^{-1} - 1 \approx 2.27$ with $g = (\sqrt{5}+1)/2$.

Proof: Property X_m^+ implies $l > \lfloor -\log p + m\rfloor$, where $\lfloor x\rfloor$ denotes the largest integer less than or equal to x. It is shown in [2] that if p and l are the probability and level of a given node, $p \ge 1/F_n$ implies $l \le n-2$ for $n \ge 3$, where $F_n = [g^n - (-g)^{-n}]/\sqrt{5} \ge g^{n-2}$ is the nth Fibonacci number ($n \ge 1$). Therefore, if $\lfloor -\log p + m\rfloor \ge 1$, the inequality $l > \lfloor -\log p + m\rfloor$ implies $p < (F_{\lfloor -\log p+m\rfloor+2})^{-1} \le g^{-\lfloor -\log p+m\rfloor} \le g^{\log p-m+1}$. For $\lfloor -\log p + m\rfloor < 1$, $p < g^{\log p-m+1}$ holds trivially. Solving for p proves the lemma. □

Theorem 2: $P_m^+ = \sum_{j\in I_m^+}p_j < 2^{-c(m-2)+2}$ where $I_m^+ = \{i | l_i > -\log p_i + m\}$, i.e., the probability that a letter has property X_m^+ is smaller than $2^{-c(m-2)+2}$.

Proof: Suppose there is at least one letter—and hence a corresponding leaf—having the property X_m^+. Then, among all nodes having the property X_m^+, there is a nonempty subset with minimum level $n_0 > 0$. In this subset, there is a node having maximum probability p_0. In other words, there is no node having property X_m^+ on a level $n < n_0$, and on level n_0, there is no node with probability $p > p_0$. Thus property X_m^+ implies

$$p_0 > 2^{-n_0+m}.$$

Now, let k_0 be the number of nodes on level $n_0 - 1$, and define the integer $l_0 < n_0$ such that $2^{l_0} \le k_0 < 2^{l_0+1}$. Then the number of level-n_0 nodes is less than 2^{l_0+2}. Since all nodes having property X_m^+ are on levels $n \ge n_0$, it follows that

$$P_m^+ < 2^{l_0+2}p_0.$$

In order to turn this into a useful bound, note the following. The sibling property or, more directly, the optimality of a Huffman code implies that all level-(n_0-1) nodes have probability $p \ge p_0$. Since there are at least 2^{l_0} level-(n_0-1) nodes, it is again a consequence of the sibling property that there exists a level-$(n_0 - 1 - l_0)$ node with probability $p_1 \ge 2^{l_0}p_0 > 2^{-n_0+m+l_0}$ and thus having property X_{m-1}^+. Using the lemma, one finds $p_1 < 2^{-c(m-2)}$, and therefore,

$$P_m^+ < 2^{l_0+2}p_0 \le 2^2 p_1 < 2^{-c(m-2)+2}. \qquad \square$$

The following examples illustrate the theorems. Examples 2 and 3 contain slightly modified versions of the sources introduced in [2] and further discussed in [6].

Example 1: Consider a two-letter source with probabilities $(p_1, p_2) = (0.9, 0.1)$. The (unique) Huffman codeword lengths are $(l_1, l_2) = (1, 1)$. One finds $P_1^- = P_2^- = 0.1$ and $P_3^- = 0$.

Example 2: Here the source consists of a seven-letter alphabet with probabilities $(\frac{5}{13}+\epsilon, \frac{3}{13}, \frac{2}{13}, \frac{1}{13}, \frac{1}{13}, \frac{1}{13}-2\epsilon, \epsilon)$ with $0 < \epsilon < 2^{-8}$. Huffman coding leads to codeword lengths $(1,2,3,4,5,6,6)$. It can easily be seen that these codeword lengths are unique—the only freedom in the coding being a trivial relabeling of equiprobable letters. One finds $P_1^+ = \frac{2}{13} - 2\epsilon$, $P_2^+ = \frac{1}{13} - 2\epsilon$, and $P_3^+ = 0$.

Example 3: The probabilities in Example 2 are slightly changed to $(\frac{5}{13}-\epsilon, \frac{3}{13}+\epsilon, \frac{2}{13}-\epsilon, \frac{1}{13}, \frac{1}{13}, \frac{1}{13}, \epsilon)$ with $0 < \epsilon < 2^{-8}$. Huffman coding leads to the (unique) codeword lengths $(2,2,3,3,3,4,4)$. One finds $P_1^+ = P_2^+ = 0$.

One might ask the question if there is a connection between the quantities $P_m^\pm$ and the redundancy $r = \sum p_i(l_i + \log p_i)$. Examples 2 and 3 show that there is at least no simple connection. In Example 2, $P_1^+ = \frac{2}{13} - 2\epsilon$, and in Example 3, $P_1^+ = 0$. On the other hand, the redundancy in both these examples approaches in the limit of small ϵ the value $r \approx 0.174$ bits.

Acknowledgment

The author wishes to acknowledge having profited much from frank discussions with C. M. Caves and C. Fuchs.

References

[1] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proc. IRE*, vol. 40, pp. 1098–1101, Sept. 1952.

[2] G. O. H. Katona and T. O. H. Nemetz, "Huffman codes and self-information," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 337–340, May 1976.

[3] W. H. Zurek, "Algorithmic randomness and physical entropy," *Phys. Rev. A*, vol. 40, pp. 4731–4751, Oct. 1989.

[4] R. Schack and C. M. Caves, "Information and entropy in the baker's map," *Phys. Rev. Lett.*, vol. 69, pp. 3413–3416, Dec. 1992.

[5] R. G. Gallager, "Variations on a theme by Huffman," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 668–674, Nov. 1978.

[6] S. W. Golomb, "Sources which maximize the choice of a Huffmann coding tree," *Inform. Contr.*, vol. 45, pp. 263–272, June 1980.

Generalizing the Fano Inequality

Te Sun Han, *Fellow, IEEE,* and Sergio Verdú, *Fellow, IEEE*

***Abstract*—The Fano inequality gives a lower bound on the mutual information between two random variables that take values on an M-element set, provided at least one of the random variables is equiprobable. We show several simple lower bounds on mutual information which do not assume such a restriction. In particular, this can be accomplished by replacing $\log M$ with the infinite-order Rényi entropy in the Fano inequality. Applications to hypothesis testing are exhibited along with bounds on mutual information in terms of the *a priori* and *a posteriori* error probabilities.**

***Index Terms*—Shannon theory, Fano inequality, mutual information, hypothesis testing.**

I. The Fano Inequality

One of the most useful results in the Shannon theory is the following lower bound on mutual information, which, in the last forty years, has proven to be the key tool in the proof of converse results in information theory.

Theorem 1: Suppose that X and Y are random variables that satisfy the following.

a) X and Y take values on the same finite set with cardinality M;

b) either X or Y is equiprobable.

Manuscript received February 15, 1993; revised December 6, 1993. This research was partially supported by the National Science Foundation under Grant ECSE-8857689. This paper was presented in part at the 1993 IEEE International Symposium on Information Theory, San Antonio, TX, January 17–22, 1993.

T. S. Han is with the Program in Information Sciences University of Electro-Communications, Chofu, Tokyo, Japan.

S. Verdú is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544.

IEEE Log Number 9402577.

0018-9448/94$04.00 © 1994 IEEE

Authorized licensed use limited to: Deakin University. Downloaded on November 19,2024 at 00:34:22 UTC from IEEE Xplore. Restrictions apply.

Then, the mutual information between X and Y satisfies

$$I(X;Y) \geq P[X=Y]\log M - h(P[X=Y]), \quad (1)$$

where h is the binary entropy function, *i.e.*, the continuous extension on $[0,1]$ of

$$h(x) = x\log\frac{1}{x} + (1-x)\log\frac{1}{1-x}.$$

Proof: If X is equiprobable, then $H(X) = \log M$ and

$$I(X;Y) = \log M - H(X|Y),$$

which is then lower bounded, using the Fano inequality [1]:

$$H(X|Y) \leq P[X \neq Y]\log(M-1) + h(P[X=Y]). \quad (2)$$

If Y (instead of X) is equiprobable, then the bound must still be true because of the symmetry of both sides of (1). □

The power of Theorem 1 stems from its ability to lower bound the mutual information between two random variables in terms of a single quantity easily computable from their joint distribution: the probability that the random variables take the same value.

The purpose of Section II is to generalize Theorem 1 so that mutual-information lower bounds can be given without the assumptions therein, i.e., that the random variables are finitely valued, and more important, that at least one of them is equiprobable. As a bonus, the proofs of the new bounds of Section II are particularly simple and intuitive. Since it is possible to construct independent (nonequiprobable) random variables (X,Y) for any arbitrarily specified $P[X=Y]$, it is apparent that dropping the assumptions of Theorem 1 will require a lower bound that depends on the distribution of X and Y not only through $P[X=Y]$, but through some other, hopefully simple, quantity.

Several of the mutual-information lower bounds found in Section II involve the Rényi entropy, which is defined as [2]

$$R_\alpha(X) = \frac{1}{1-\alpha}\log\sum_{\omega\in\Omega} P_X^\alpha(\omega)$$

for $\alpha > 0$ and $\alpha \neq 1$. $R_\alpha(X)$ is monotone decreasing in α. $R_0(X)$ is equal to the log of the number of nonzero atoms, and the infinite-order Rényi entropy is equal to

$$R_\infty(X) = \log\frac{1}{\max_{\omega\in\Omega} P_X(\omega)}.$$

It is shown in Section II that one way to generalize Theorem 1 is to replace the zero-order Rényi entropy ($\log M$) that appears in (1) by the infinite-order Rényi entropy $R_\infty(X)$ (or $R_\infty(Y)$). Then, the resulting general bound reduces to (1) when the assumptions of Theorem 1 are satisfied. This new bound finds applications in the proof of a generalized source-channel separation theorem in a nonstandard setting [3]. Other applications of the mutual-information lower bounds of Section II are illustrated in Section III, where we explore their relationship with minimum error probability in hypothesis testing.

II. New Mutual-Information Lower Bounds

First, we observe a simple inequality between information divergences.

Theorem 2: Suppose that the random variables $X, Y, \bar{X}, \bar{Y}$ satisfy the following:

a) they take values on the same set Ω (which need not be finite);

b) $\bar{X}$ and $\bar{Y}$ are independent.

Then,

$$I(X;Y) \geq d(P[X=Y]\|P[\bar{X}=\bar{Y}]) - D(P_X\|P_{\bar{X}}) - D(P_Y\|P_{\bar{Y}}), \quad (3)$$

where P_X denotes the distribution of the random variable X, $D(P\|Q)$ denotes the information divergence, and the binary divergence function $d(x\|y)$ is defined as the continuous extension on $[0,1]^2$ of

$$d(x\|y) = x\log\frac{x}{y} + (1-x)\log\frac{1-x}{1-y},$$

i.e., $d(x\|y) = D([x,1-x]\|[y,1-y])$.

Proof: Under the assumption that $\bar{X}$ and $\bar{Y}$ are independent,

$$D(P_{XY}\|P_{\bar{X}\bar{Y}}) = D(P_{XY}\|P_{\bar{X}}P_{\bar{Y}}) = I(X;Y) + D(P_X\|P_{\bar{X}}) + D(P_Y\|P_{\bar{Y}}).$$

Now, inequality (3) follows by applying the data processing theorem for divergence ("processing reduces divergence") to a processor whose input is (x,y) and whose output is $1\{x=y\}$ under the different input distributions P_{XY} and $P_{\bar{X}\bar{Y}}$. □

Various useful lower bounds can be derived from Theorem 2 depending on the choice of the auxiliary random variables $\bar{X}$ and $\bar{Y}$. We will consider three different choices:

1) • Ω is a finite set with cardinality M.
   • $\bar{X}$ is equiprobable.
   • $P_{\bar{Y}} = P_Y$.
   It is easy to check that with this choice, the inequality in (3) becomes the Fano inequality (2). This way of deriving the Fano inequality is due to Blahut [4].
2) • $P_{\bar{X}} = P_X$.
   • $P_{\bar{Y}} = P_X$.
   Then, (3) becomes the general mutual information lower bound

$$I(X;Y) \geq P[X=Y]R_2(X) - h(P[X=Y]) - D(P_Y\|P_X).$$

3) • $P_{\bar{X}} = P_X$.
   • $P_{\bar{Y}} = P_Y$.
   This leads to the following result.

Theorem 3: If X and Y take values on the same set Ω, then

$$I(X;Y) \geq d(P[X=Y]\|P[\bar{X}=\bar{Y}]), \quad (4)$$

where $\bar{X}$ and $\bar{Y}$ are independent, and have the same marginal distributions as X and Y, respectively. Furthermore, equality holds in (4) if and only if for some constants α and β,

$$P_{XY}(x,y) = \begin{cases} \alpha P_X(x)P_Y(y), & x=y, \\ \beta P_X(x)P_Y(y), & x\neq y. \end{cases} \quad (5)$$

Proof: The bound follows from Theorem 2 as noted. The necessary and sufficient condition (5) for equality in (4) follows from the identity

$$D(P_U\|Q_U) = D(P_V\|Q_V) + D(P_{U|V}\|Q_{U|V}|P_V),$$

applied to the case where $U=(X,Y)$, $V = 1\{X=Y\}$, $P_U = P_{XY}$, and $Q_U = P_XP_Y$. □

Authorized licensed use limited to: Deakin University. Downloaded on November 19,2024 at 00:34:22 UTC from IEEE Xplore. Restrictions apply.

Regarding the lower bound in (4), note that

$$P[\bar{X} = \bar{Y}] = \sum_{\omega \in \Omega} P_X(\omega) P_Y(\omega),$$

i.e., the inner product between the marginals of X and Y, which is often easy to obtain from the description of X and Y.

It can be checked that except in the trivial case where X and Y are independent, condition (5) implies that the marginals are either nonoverlapping or both equiprobable (on a subset of Ω).

We will now loosen (4) so that we can obtain bounds with the same structure as (1). We do so simply by lower bounding binary divergence.

Theorem 4: If X and Y take values on the same set, then

$$I(X;Y) \geq P[X = Y] \log \frac{1}{P[\bar{X} = \bar{Y}]} - h(P[X = Y]). \quad (6)$$

Proof: The desired inequality follows from Theorem 3 and the lower bound on binary divergence

$$d(x \| y) \geq x \log \frac{1}{y} - h(x). \qquad \square$$

In some cases, the marginal distribution of Y may not be immediately available, in which case it is convenient to replace $P[\bar{X} = \bar{Y}]$ in (6) by a quantity which is a function of the marginal distribution of X only. This is done in the next result.

Theorem 5: If X and Y take values on the same set, then

$$I(X;Y) \geq P[X = Y] R_\infty(X) - h(P[X = Y]), \quad (7)$$

where by symmetry we can replace $R_\infty(X)$ by $R_\infty(Y)$.

Proof: The result follows from (6) and

$$P[\bar{X} = \bar{Y}] \leq \min\left\{ \max_{\omega \in \Omega} P_X(\omega), \max_{\omega \in \Omega} P_Y(\omega) \right\}. \qquad \square$$

Note that Theorem 5 takes the same form as Theorem 1, replacing the zero-order Rényi entropy by the infinite-order Rényi entropy. If the conditions of Theorem 1 are satisfied, then both bounds are identical. However, Theorem 5 holds in full generality; neither X nor Y need be equiprobable or even finitely valued.

The infinite-order Rényi entropy $R_\infty(X)$ is a measure of the randomness of X, which quantifies how hard it is to guess the value of X knowing only its distribution. The probability of error with no information on X, ε_X (prior Bayes risk with a Hamming loss function), is the monotonic transformation

$$\epsilon_X = 1 - (\exp(-R_\infty X)).$$

The infinite-order Rényi entropy satisfies the properties

$$0 \leq R_\infty(X) \leq H(X) \quad (8)$$

and

$$R_\infty(X_1, \cdots X_n) = R_\infty(X_1) + \cdots + R_\infty(X_n),$$

if $X_1, \cdots X_n$ are independent. If X is restricted to take M values, then the bounds in (8) can be improved. The region of allowable $(R_\infty(X), H(X))$ pairs as a function of M has been determined in [5]. If the cardinality of X is not bounded, then the bounds in (8) cannot be improved.

It is now tempting to strengthen the lower bound in Theorem 4 with

$$I(X;Y) \geq P[X = Y] H(X) - h(P[X = Y]). \quad (!?)$$

However, this bound does not hold in general. For example, if X and Y are independent with identical distribution

$$P[X = i] = \begin{cases} q, & i = 0, \\ (1-q)/N, & i = 1, \cdots N, \end{cases} \quad (9)$$

then the left side of (!?) is 0, whereas the right side is positive for any $0 < q < 1$ provided N is large enough.

Introducing the maximal probability of error in lieu of the average probability of error, it is possible to modify the incorrect bound (!?) and obtain the following result involving the input entropy.

Theorem 6: Assume that X and Y take values on the same set Ω and denote

$$\rho = \inf_{\omega \in \Omega} P[X = Y | X = \omega] = \inf_{\omega \in \Omega} P_{Y|X}(\omega|\omega).$$

Then,

$$I(X;Y) \geq \rho H(X) - h(P[X = Y]).$$

Proof:

$$\begin{aligned} I(X;Y) &= \sum_{a \in \Omega} P_{Y|X}(a|a) P_X(a) \log \frac{1}{P_X(a)} \\ &\quad + \sum_{a \in \Omega} P_{Y|X}(a|a) P_X(a) \log \frac{P_{Y|X}(a|a) P_X(a)}{P_Y(a)} \\ &\quad + \sum_{a \in \Omega} \sum_{b \neq a} P_{Y|X}(b|a) P_X(a) \log \frac{P_{Y|X}(b|a) P_X(a)}{P_Y(b) P_X(a)} \\ &\geq \rho H(X) \\ &\quad + P[X = Y] \log P[X = Y] \\ &\quad + P[X \neq Y] \log \frac{P[X \neq Y]}{P[\bar{X} \neq \bar{Y}]}, \end{aligned}$$

where we have used the definition of ρ, the log-sum inequality (e.g., [2]), and the notation $P[\bar{X} = \bar{Y}] = \sum_{\omega \in \Omega} P_X(\omega) P_Y(\omega)$. $\square$

If, in addition to the sufficient condition in Theorem 6, the following condition holds:

$$\rho \geq P[X \neq Y]$$

(which occurs, for example, when $\rho > \frac{1}{2}$), then the bound in Theorem 6 can be replaced by the weaker bound

$$I(X;Y) \geq \rho H(X) - h(\rho),$$

which was known ([6], Lemma 3.5) to hold in the special case $\rho > 1 - e^{-1}$.

To conclude this section, we note that the restriction that X and Y take values on the same set has been made throughout for convenience in expressing the results. It is easy to see from the mutual-information data processing lemma that the restriction can be lifted in the foregoing results by replacing $P[X = Y]$ and $P[\bar{X} = \bar{Y}]$ by $P[X = \phi(Y)]$ and $P[\bar{X} = \phi(\bar{Y})]$, respectively, where ϕ is an arbitrary function mapping the set of Y values to the set of X values.

III. Mutual Information and Error Probability

In this section we illustrate the use of the results found in Section II in order to lower bound the error probability of M-ary hypothesis testing.

Let X take values on $\{1, \cdots M\}$, and let Z be the observable whose conditional distribution given that $X = j$ is Q_j. Define

$$\epsilon_X = 1 - \max_{1 \leq j \leq M} P_X(j)$$

Authorized licensed use limited to: Deakin University. Downloaded on November 19,2024 at 00:34:22 UTC from IEEE Xplore. Restrictions apply.

and

$$\epsilon_{X|Z} = 1 - E\left[\max_{1 \le j \le M} Q_j(Z)P_X(j)/P_Z(Z)\right],$$

where $P_Z(b) = \sum_{j=1}^{M} P_X(j)Q_j(b)$. Note that ϵ_X and $\epsilon_{X|Z}$ are the *a priori* and *a posteriori* minimum probabilities of error, respectively. In decision theory [7], the following bound on the error probability of equiprobable hypothesis testing is well known.[1]

Theorem 7: The minimum probability of error for any test between equiprobable hypotheses $\{Q_i,\ i = 1, \cdots M\}$ is lower bounded by

$$\epsilon_{X|Z} \ge 1 - \frac{1}{\log M}\left(\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} D(Q_i\|Q_j) + \log 2\right).$$

Proof: Fix a test, and let Y be its output. Theorem 7 follows by applying Theorem 1 to (X, Y) and bounding:

$$\begin{aligned} I(X;Y) &\le I(X;Z) \\ &= \frac{1}{M}\sum_{i=1}^{M} D\left(Q_i\Big\|\frac{1}{M}\sum_{j=1}^{M} Q_j\right) \\ &\le \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} D(Q_i\|Q_j), \end{aligned}$$

where the inequalities follow from the data processing lemma and the convexity of divergence, respectively. □

Following a similar proof, Theorem 7, which holds only for equiprobable hypotheses, can be generalized via Theorem 5 as follows.

Theorem 8: The minimum probability of error for any test between hypotheses $\{Q_i,\ i = 1, \cdots M\}$ is lower bounded by

$$\epsilon_{X|Z} \ge 1 + \frac{1}{\log \max_j q_j}\left(\sum_{i=1}^{M}\sum_{j=1}^{M} q_i q_j D(Q_i\|Q_j) + \log 2\right),$$

where q_j is the a priori probability of the j-th hypothesis. (Note here that M need not be finite.)

Rather than using Theorem 7, it is more common in information theory to use (in converse proofs) the tighter result

$$\epsilon_{X|Z} \ge 1 - \frac{I(X;Y) + \log 2}{\log M}, \tag{10}$$

which follows directly from Theorem 1. However, (10) holds only for equiprobable hypotheses. In the general nonequiprobable case, Theorem 5 results in

$$\epsilon_{X|Z} \ge 1 + \frac{I(X;Z) + \log 2}{\log(1 - \epsilon_X)},$$

which can be viewed as a lower bound on mutual information as a function of $\epsilon_{X|Z}$ and ϵ_X. A tighter such bound is given by the following result.

Theorem 9: If X is finitely valued (or countably infinite),

$$I(X;Z) \ge d(\epsilon_{X|Z}\|\epsilon_X).$$

Proof: Let $\hat{X}(Z)$ be the maximum *a posteriori* estimate of X given Z. The mutual-information data processing theorem and Theorem 3 yield

$$\begin{aligned} I(X;Z) &\ge I(X;\hat{X}(Z)) \\ &\ge d(P[X \ne \hat{X}(Z)]\|P[\bar{X} \ne \hat{X}(Z)]) \qquad (11) \\ &\ge d(\epsilon_{X|Z}\|\epsilon_X), \qquad (12) \end{aligned}$$

where $\bar{X}$ is independent of Z and has the same distribution as X. In order to check (12), note that

$$\epsilon_{X|Z} = P[X \ne \hat{X}(Z)]$$

and

$$\epsilon_X \le P[\bar{X} \ne \hat{X}(Z)],$$

because when the maximum *a posteriori* estimator is driven by observations that are independent of X, it cannot achieve better error probability than the minimum *a priori* error probability ϵ_X. Finally, it is easy to check that $d(a\|b) \le d(a\|c)$ if $0 \le a \le b \le c \le 1$. □

In Theorem 3 we derived a necessary and sufficient condition for equality, which leads us to conclude that (11) and, thus, the bound in Theorem 9, will not be tight unless X is equiprobable. If X is indeed equiprobable over the M elements, then, for every value $\delta \le 1 - 1/M$, it is possible to find Z such that $\delta = \epsilon_{X|Z}$ and $I(X;Z) = d(\epsilon_{X|Z}\|\epsilon_X)$ because $\epsilon_X = 1 - 1/M$. (For example, given X, let $Z = X$ with probability $1 - \delta$ and let it be equidistributed on the other $M - 1$ values, with probability δ.)

Conversely to Theorem 9, we can find upper bounds on mutual information as a function of the *a priori* and *a posteriori* error probabilities:

$$\begin{aligned} I(X;Z) &= H(X) - H(X|Z) \\ &\le \log M - \epsilon_{X|Z}\log 4, \end{aligned} \tag{13}$$

where the last inequality follows from ([9], p. 520). It is possible to tighten (13) using the sharp bounds

$$H(X) \le \epsilon_X \log(M - 1) + h(\epsilon_X)$$

and

$$H(X|Z) \ge \phi^*(\epsilon_{X|Z}),$$

where $\phi^*(\epsilon_{X|Z})$ is the piecewise linear convex function defined in [5] and shown to be the tightest lower bound on the conditional entropy $H(X|Z)$ as a function of $\epsilon_{X|Z}$. However, it does not follow that

$$I(X;Z) \le \epsilon_X \log(M - 1) + h(\epsilon_X) - \phi^*(\epsilon_{X|Z})$$

is the tightest possible bound in terms of ϵ_X and $\epsilon_{X|Z}$ because the foregoing bounds have not been shown to be simultaneously tight.

Acknowledgment

Section III benefited from suggestions by Meir Feder.

References

[1] R. M. Fano, "Class notes for Transmission of Information, Course 6.574," MIT, Cambridge, MA, 1952.

[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[3] T. S. Han and S. Verdu, "Nonserial Information Sources," in preparation.

[4] R. E. Blahut, "Information bounds of the Fano-Kullback type," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 410–421, July 1976.

[5] M. Feder and N. Merhav, "Relations between entropy and error probability," *IEEE Trans. Inform. Theory*, vol. 40, pp. 259–266, Jan. 1994.

[6] H. G. Ding, "On Shannon theorem and its converse for sequences of communication schemes in the case of abstract random variables," in *Trans. Third Prague Conf. Inf. Theory, Stat. Decision Functions Random Processes*, (Prague, Czechoslovakia), June 1962, pp. 285–332.

[7] L. LeCam, *Asymptotic Methods in Statistical Decision Theory*. New York: Springer-Verlag, 1986.

[1]Actually, in the minimax decision theory literature (e.g., [7] and [8]) it is a slightly weaker version of Theorem 7 where $\log(M - 1)$ takes the role of $\log M$, which is known as Fano's lemma.

Authorized licensed use limited to: Deakin University. Downloaded on November 19,2024 at 00:34:22 UTC from IEEE Xplore. Restrictions apply.

[8] L. Devroye, *A Course in Density Estimation.* Boston: Birkhauser, 1987.
[9] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

On n-Phase Barker Sequences

Ning Chang and Solomon W. Golomb[1]

***Abstract*—An n-phase Barker sequence can be easily distinguished from the time-shifted versions of itself. This property is important for such applications as radar systems, synchronization systems, and spread-spectrum communications systems. In this paper, we study some transformations on n-phase Barker sequences. Also, we give an efficient algorithm for finding sextic Barker sequences. Through an exhaustive computer search, numerical data for n-phase Barker sequences are given. Specifically, we extend the list of known n-phase Barker sequences to length $L = 19$.**

***Index Terms*—Barker sequence, n-phase Barker sequence, aperiodic autocorrelation, Barker-preserving transformation.**

I. Introduction

In 1954, Barker [1] exhibited binary ± 1 sequences $\{b_i\}_{i=1}^{L}$ of respective lengths $L = 2, 3, 4, 5, 7, 11$, and 13, with the aperiodic autocorrelation property $|C(\tau)| \le 1$ for $\tau = 1, 2, \cdots, L-1$, where $C(\tau) = \sum_{i=1}^{L-\tau} b_i b_{i+\tau}$.

It was shown (Storer and Turyn [2]) that there are no other Barker sequences of odd length. Moreover, Turyn [3] obtained constraints on possible even lengths $L > 4$ (e.g., they must be perfect squares, cannot be powers of 2, etc.), which make it appear quite unlikely that any such examples exist. In 1988, Eliahou *et al.* [4] proved that for any possible even length $L > 4$, there is no prime factor p of L such that $p \equiv 3 \pmod 4$.

An n-phase Barker sequence of length L is defined as an L-term sequence, $b_1, b_2, \cdots, b_L$, where b_i is an nth root of unity and $|C(\tau)| \le 1$ for all i, $1 \le i \le L$ and all τ, $1 \le \tau \le L-1$, where $C(\tau) = \sum_{i=1}^{L-\tau} b_i b_{i+\tau}^*$. (Here, z^* denotes the complex conjugate of z.)

In 1965, Golomb and Scholtz [5] found sextic Barker sequences of all lengths L, $1 \le L \le 13$. They also found a group of Barker-preserving transformations of order $4n^2$. As generators of this group, we may take the two order-2 transformations of time reversal and complex conjugation, and the two order-n transformations of constant multiplication by ρ^a and progressive multiplication by successive powers of ρ^a, where $\rho = e^{2\pi i/n}$. Circa 1974, Scholtz found sextic Barker sequences of lengths 14 and 15. Other examples, not using sixth roots, had been found for lengths $L = 15$ (4-phase) and $L = 16$ (8-phase), using constructions by Price and Carley, respectively.

In this paper, we consider the effects of the Barker-preserving transformations. Also, we discuss some properties of sixth roots of unity. We observe that the sixth roots of unity (with zero adjoined) have the useful property of being partially closed under addition, as well as closed under multiplication. We give an efficient algorithm for finding sextic Barker sequences.

By a computer search, we have found all sextic Barker sequences for all lengths L, $1 \le L \le 22$. We give some other computer search results: all 5-phase Barker sequences of length $L \le 16$, all 8-phase Barker sequences of length $L \le 16$, and all 12-phase Barker sequences of length $L \le 14$, with partial results for lengths $L = 15$ and 16. Moreover, we have found a 15-phase Barker sequence of length $L = 17$, a 24-phase Barker sequence of length $L = 18$, a 360-phase Barker sequence of length $L = 19$, and a 180-phase Barker sequence of length $L = 19$. (As pointed out to us by Lüke (private communication), the example of a 60 phase Barker sequence of length $L = 19$ in [6] is erroneous. The 180-phase sequence of length $L = 19$ given in this paper is the example with the fewest phases which is currently known.)

II. Transformations

We let $\{b_j = e^{ia_j}\}_{j=1}^{L}$ be an n-phase Barker sequence of length L, where a_j belongs to N_n, for every j, where $N_n = \{2\pi j/n\}_{j=0}^{n-1}$.

Definition 1:

1) The order-2 transformation of time reversal:

$$T : \{e^{ia_j}\}_{j=1}^{L} \to \{e^{ia_{L+1-j}}\}_{j=1}^{L}.$$

2) The order-2 transformation of complex conjugation:

$$C : \{e^{ia_j}\}_{j=1}^{L} \to \{e^{i(-a_j)}\}_{j=1}^{L}.$$

3) The order-n transformation of constant multiplication by e^{ik}:

$$k : \{e^{ia_j}\}_{j=1}^{L} \to \{e^{i(a_j+k)}\}_{j=1}^{L}.$$

4) The order-n transformation of progressive multiplication by successive powers of e^{ik}:

$$p(h) : \{e^{ia_j}\}_{j=1}^{L} \to \{e^{i[a_j+(j-1)h]}\}_{j=1}^{L},$$

where k and h belong to N_n.

5) Any transformation f belonging to the group generated by all of the above transformations is called a *composed transformation* if it is neither an identity transformation nor a generator of the group.

It is easy to verify that any n-phase Barker sequence under these transformations is still an n-phase Barker sequence. In [5], the first four transformations were called Barker-preserving transformations for obvious reasons.

III. Some Properties of Sixth Roots of Unity

Let $\rho = e^{2\pi i/6}$ and $A_6 = \{1, \rho^1, \rho^2, \rho^3, \rho^4, \rho^5\}$. We find that the sixth roots of unity (with zero adjoined) have the useful property of being partially closed under addition, as well as closed under multiplication.

Property 1: The sum of two vectors ρ^x and ρ^y lies within the unit circle if and only if $|x - y| \pmod 6 > 1$, where x and y belong to $\{0, 1, 2, 3\}$.

Property 2: The result of finite addition and multiplication of any sextic roots of unity is of the form "$A\rho^a + B\rho^b$," where a, b, A, and B are four positive integers, and $|a - b| \pmod 6 \le 1$.

It is easy to show the following.

Theorem 1: Given any $A\rho^a + B\rho^b$, where $|a - b| \pmod 6 \le 1$ and a, b, A, B are four positive integers, then $F(x, y) = (\rho^x +$

Manuscript received October 29, 1992; revised April 22, 1993.
N. Chang is with Pacific Bell, 2600 Camino Ramon, 15950K, San Ramon, CA 94583.
S. W. Golomb is with the Department of Electrical Engineering—Systems, University of Southern California, Los Angeles, CA 90089.
IEEE Log Number 9403842.
[1]Dr. Golomb's research was suppported in part by the United States Navy, Office of Naval Research, under Grant No. N00014-90-J-1341.

0018-9448/94$04.00 © 1994 IEEE

Authorized licensed use limited to: Deakin University. Downloaded on November 19,2024 at 00:34:22 UTC from IEEE Xplore. Restrictions apply.