XV. PROCESSING AND TRANSMISSION OF INFORMATION*

Academic and Research Staff

Prof. R. M. Gallager
Prof. E. V. Hoversten
Prof. I. M. Jacobs
Prof. R. E. Kahn
Prof. R. S. Kennedy

Graduate Students

D. S. Arnstein
J. D. Bridwell
E. A. Bucher
D. Chase
D. D. Falconer
R. L. Greenspan
D. Haccoun
H. M. Heggestad
J. A. Heller
M. Khanna
J. Max
J. H. Meyn
J. C. Molden
G. Q. McDowell
G. C. O'Leary
R. Pilc
J. T. Pinkston III
E. M. Portner, Jr.
J. S. Richters
J. E. Roberson
M. G. Taylor
D. A. Wright
R. Yusek

A. CODING FOR SOURCE-CHANNEL PAIRS

In many communication problems the source output is not simply one of M equally likely messages and the user is not merely interested in the probability that the received information is right or wrong. More generally, a source may have any probability distribution P(w) defined over its possible outputs, and the system performance may be measured by the average value of a distortion function $d(w_i, w_j)$ which gives the distortion to the user when w_i is transmitted but decoded as w_j. The problem of communicating the output of such a source over a given channel with minimum distortion is being studied in this research. Unlike previous work, which separated the coding operation into two parts, a source representation problem[1,2] and a channel coding problem, we shall consider the coding operation as one process.

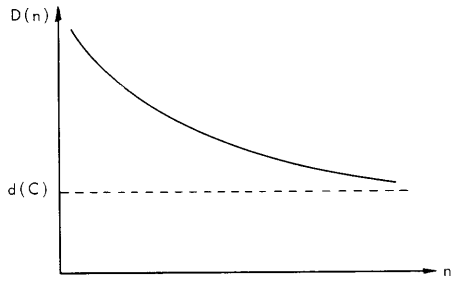

Fig. XV-1. Performance curve for source $\mathcal{S}$ transmitted over channel $\mathcal{C}$.

The performance curve is defined for a given source-channel pair as the minimum obtainable distortion, using a direct source to channel encoder which operates on a block of source outputs of length n. A typical curve is shown in Fig. XV-1. If the capacity of the channel $\mathcal{C}$ is C, it is known from Shannon's earlier results[1] that d(C), the value of distortion at the rate C on the rate-distortion curve for the source $\mathcal{S}$, provides a lower bound to the performance curve. It is also shown by Shannon to be obtainable, therefore it must be the limit of the performance curve as n becomes large.

A stronger lower bound has been derived which, unlike Shannon's, is a function of n

*This work was supported by the National Aeronautics and Space Administration (Grant NsG-334).

and therefore provides information about the rate of approach of the performance curve to its limiting value as n increases. The derivation is summarized below.

1. Lower Bound to D(n)

The first step in the derivation is an application of sphere-packing ideas used many times before in Information Theory. If a source word $\underline{w}$ is transmitted, using a channel input word $\underline{x}$, and two lists are made – one a list of possible decoded words $\underline{w}'$ ordered in decreasing distortion $d(\underline{w}, \underline{w}')$ from w, and the other a list of channel output words $\underline{y}$ ordered in decreasing conditional probability $p(\underline{y}/\underline{x})$ – a lower bound to any obtainable distortion can be found by evaluating the completely ideal, and unrealizable, situation wherein channel output words that have higher conditional probabilities are always in decoding regions that result in lower transmission distortions.

It has been shown by Fano[3] that an improvement in this bound can be made if a probability function f(y), defined over the channel output space Y, is included in the ordering of channel output words and subsequently varied to obtain the tightest possible bound. We shall use this idea and order the channel output words according to increasing values of information difference

$$I(\underline{x}, \underline{y}) = \ln \frac{f(\underline{y})}{p(\underline{y}|\underline{x})},$$

where

$$f(\underline{y}) = \prod_{i=1}^{n} f(y_i).$$

To obtain a lower bound, it is necessary to relate members on the list of all possible decoded words with members on the list of all possible received words (now ordered in distortion and information difference respectively). This is the "idealized decoder function" and is defined as that function which maps y_j into the w_i' for which

$$\sum_{\underline{w}' \in W_i'} g(\underline{w}') \leq \sum_{\underline{y} \in Y_j} f(\underline{y}) < \sum_{\underline{w}' \in W_{i+}'} g(\underline{w}'), \tag{1}$$

where

$$Y_j = \{\underline{y} : I(\underline{x}, \underline{y}) \leq I(\underline{x}, \underline{y}_j)\}, \tag{2}$$

$$W_i' = \{\underline{w}' : d(\underline{w}, \underline{w}') \leq d(\underline{w}, \underline{w}_i')\}, \tag{3}$$

$$W_{i+}' = W_i' \cup (\text{the next } \underline{w}' \text{ on the list}), \tag{4}$$

and where g(w') is a probability function, defined over the decoding space W', that will be determined later. The function value $g(\underline{w}_i')$ can be interpreted as the total "size", as measured by $f(\underline{y})$, of the subset of Y^n that is decoded into $\underline{w}_i'$. That is,

$$g(\underline{w}_i') = \sum_{\underline{y} \in Y(\underline{w}_i')} f(\underline{y}),$$

where

$$Y(\underline{w}_i') = \{\underline{y}: \text{decoded into } \underline{w}_i'\}.$$

The lower bound to the distortion that results when the source word $\underline{w}$ is transmitted using a channel input word $\underline{x}$ can now be shown to be

$$D(\underline{w}) \geq \int d(I)\, dF_2(I) \tag{5}$$

in which $F_2(I)$ is the cumulative distribution function of the information difference $I(\underline{x}, \underline{y})$ when the probability distribution $p(\underline{y}/\underline{x})$ is in effect, and d(I) is the distortion function implicitly defined by the first inequality of Equation 1 which essentially equates two distribution functions; one G(d) where

$$G(d) = \Pr_{g(\underline{w}')} (d(\underline{w}, \underline{w}') \leq d)$$

and the other $F_1(I)$ where

$$F_1(I) = \Pr_{f(\underline{y})} (I(x, \underline{y}) \leq I).$$

Since G(d) and $F_1(I)$ can only be approximated,[4] an upper bound to G(d) is equated to a lower bound to $F_1(I)$ to define a function $d_L(I)$ satisfying $d_L(I) \leq d(I)$. This is consistant with the inequality in Equation 5. Finally expanding $d_L(I)$ in a Taylor series about E(I) with respect to the cumulative distribution $F_2(I)$ yields

$$D(\underline{w}) \geq \int d_L(I)\, dF_2(I) = \sum_{i=0} \frac{d_L^{(i)}(I)}{i!} \int (I-\bar{I})\, dF_2(I).$$

After successive derivatives and central moments are evaluated we obtain the result

$$D(\underline{w}) \geq \mu'(s_o) - \frac{1}{2ns_o} \left[\frac{\gamma''(-1)}{s_o^2 \mu''(s_o)} - 1 \right] + 0\left(\frac{1}{n}\right) \tag{6}$$

where s_o satisfies

$$\mu(s_o) - s_o\mu'(s_o) = \gamma'(-1) - \frac{1}{2n} \ln \frac{\gamma''(-1)}{s_o^2\mu''(s_o)} + 0\left(\frac{1}{n}\right), \tag{7}$$

and in which

$$\mu(s) = \sum_i q_i\mu_i(s), \qquad \text{comp } (\underline{w}) = \bar{q}, \tag{8}$$

$$\gamma(t) = \sum_i c_i\gamma_i(t), \qquad \text{comp } (\underline{x}) = \bar{c}. \tag{9}$$

In Eqs. 8 and 9, $\mu_i(s)$ and $\gamma_i(t)$ are, respectively, the semi-invariant moment-generating functions of the random variables d_i and I_i, which have the distribution functions

$$\Pr_{d_i}(d_{ij}) = g(w_j')$$

and

$$\Pr_{I_i}(I_{ij}) = f(y_j).$$

The transmission distortion for the source can be obtained by averaging $D(\underline{w})$ over the entire source space W^n. If the code is restricted to be a "fixed-composition" code, that is, all channel input words have composition $\bar{c}$, the averaging can be completed, and it results in the lower bound

$$D \geq \mu'(s_o) - \frac{1}{2ns_o}\left[\frac{\gamma''(-1)}{s_o^2\mu''(s_o)} - 1 + \frac{\sigma^2(s_o)}{s_o^2\mu''(s_o)}\right] + 0\left(\frac{1}{n}\right) \tag{10}$$

with

$$\sigma^2(s_o) = \underset{p(w_i)}{\text{Variance}} \left[\mu_i(s_o) - s_o\mu'(s_o)\right],$$

$$\bar{q} = \bar{p} = (p(w_1), p(w_2), \ldots, p(w_J))$$

and with s_o satisfying Eq. 7.

The lower bound in Eq. 10 is in terms of the vectors $\bar{g}$, $\bar{c}$, and $\bar{f}$ which have not yet been specified. The vectors $\bar{g}$ and $\bar{c}$ must be picked to minimize the right side of Eq. 10, abbreviated $D(g, \bar{c}, \bar{f}, s_o)$, in order to choose the optimum set of decoding set sizes and the best channel-input composition. The vector $\bar{f}$ can be freely chosen, but

the tightest lower bound results when $D(\bar{g}, \bar{c}, \bar{f}, s_o)$ is maximized with respect to $\bar{f}$. Therefore

$$D \geq \min_{\bar{g}} \min_{\bar{c}} \max_{\bar{f}} D(\bar{g}, \bar{c}, \bar{f}, s_o). \tag{11}$$

As n becomes large, the vectors $\bar{c}$ and $\bar{f}$ which provide the bound in Eq. 11 approach the channel input and output probabilities associated with the channel $\mathscr{C}$ when it is used to capacity, and the vector $\bar{g}$ approaches the output probability distribution of the test channel associated with the point (d(c), c) on the rate-distortion curve for $\mathcal{S}$. For finite, but large n, these vectors could be used in Eqs. 7 and 10 to obtain an approximation to the correct lower bound. The limit, as n increases, of the lower bound is

$$D(n=\infty) \geq \mu'(s_o),$$

where

$$\mu(s_o) - s_o\mu'(s_o) = -C$$

which is the correct parametric expression for the distortion at the point (d(c), c) on the rate-distortion curve for $\mathcal{S}$.[2]

The previous results can be applied, with obvious modifications, to a communication system with vector sources and channels and with amplitude continuous sources and channels. If, in particular, for Gaussian sources and channels the channel-input fixed-composition requirement is replaced by an input energy constraint, the lower bound to distortion is the same as that given in Eq. 10, except the term involving $\sigma^2(s_o)$ is not present. The channel-input composition problem, which is believed to affect only this term, remains one of the problems under present investigation.

At this point it is not known how well the dependence upon n given in the lower bound agrees with that of the actual performance curve. To get this information, an upper bound to the performance curve is also required. Such a bound is now being developed.

R. Pilc

References

1. C. E. Shannon, "Coding Theorems for Discrete Sources with a Fidelity Criterion," Information and Decision Processes, 1960.
2. T. J. Goblick, "Coding for a Discrete Information Source with a Distortion Meas ure," Ph. D. Thesis, Department of Electrical Engineering, M. I. T., October 1962.
3. R. M. Fano, Transmission of Information (The M. I. T. Press, Cambridge, Mass, 1961).
4. R. G. Gallager, "Lower Bounds on the Tails of Probability Distribution," Quarterly Progress Report, No. 65, Research Laboratory of Electronics, M. I. T., April 15, 1962.

B. AN UPPER BOUND ON THE DISTRIBUTION OF COMPUTATION FOR SEQUENTIAL DECODING WITH RATE ABOVE R_{comp}

A previous report[1] has described the simulation of a sequential decoder operating at a rate just above R_{comp}, the computational cutoff rate of a discrete memoryless channel. The tail of the cumulative distribution of the number of computations per search was observed to be Pareto; that is,

$$\text{pr}\,(C \geq X) = AX^{-\alpha} \qquad (X \gg 1), \tag{1}$$

where A is a constant. The measured Pareto exponent, α, was less than one, and satisfied the relation

$$\frac{E_o(\alpha)}{\alpha} = R \tag{2}$$

to a good approximation. In (2), R is the code information rate in nats per channel use, and $E_o(\alpha)$ is a well-known function of α and of the channel input and transition probabilities.[2] In fact it turns out that $R_{comp} = E_o(1)$.

We have obtained by random coding arguments an upper bound on $\text{pr}(C \geq X)$ for rates in the range $R_{comp} \leq R < C$, where C is the channel capacity. Previously Savage[3] and Yudkin[4] have established similar upper bounds for $0 \leq R < R_{comp}$. Jacobs and Berlekamp[5] have obtained a lower bound agreeing asymptotically with (1) and (2) for $0 \leq R < C$. Thus the asymptotic behavior of the distribution of computation for any rate less than capacity is now known to be Pareto with exponent given by (2).

1. Outline of the Derivation

In what follows, we will provide a rough outline of the derivation of the bound. A complete description of tree codes and sequential decoding will not be given here. An up-to-date description has been given by Wozencraft and Jacobs.[2] Suffice it to say that the decoding procedure is a sequential search through a tree in an attempt to find the correct path representing the intended information sequence. Decisions are made by comparing a path metric, which is a function of the received and hypothesized channel symbol sequences, to a running threshold. The path metric along the correct path tends to increase, while incorrect path metrics tend to decrease with increasing penetration into the code tree.

We assume that the decoder employs the Fano algorithm and that the k^{th} path metric increment is

$$Z_k = \ln \frac{p(y_k|x_k)}{f(y_k)} - R.$$

The spacing between adjacent thresholds is Δ.

We are concerned with the total number of computations ever done in the incorrect subset of a reference node on the correct path. The incorrect subset consists of the reference node plus all nodes on incorrect paths stemming from the reference node. One computation is said to be done on a node whenever the decoding algorithm visits that node to examine branches diverging from it or leading to it. This is the usual measure of computation for purposes of establishing upper bounds.[3] Other definitions are more convenient for experimental measurements, but the same asymptotic Pareto distribution is always observed.[1,7]

Each node in the incorrect subset will be labelled by indices (ℓ, n), where $\ell = 0, 1, 2, \ldots$ is its depth in the tree measured from the reference node, and n denotes which node it is at depth ℓ. $[n = 1, 2, \ldots (u-1)u^{\ell-1}]$.

The expression overbounding the number of computations in the incorrect subset of a reference node depends on three properties of the Fano algorithm which we state here without further elucidation. A fuller exposition is found in Savage,[3] and in Wozencraft and Jacobs.[6]

(1). With a given running threshold in effect, no more than (u+1) computations are done on any node.

(2). For at least one computation to be done on some node (ℓ, n) when a given running threshold is in effect, a necessary condition is that the path metric along the path connecting node (ℓ, n) to the reference node be everywhere greater than the running threshold.

(3). The running threshold is not eventually reduced by Δ from its current value unless the path metric at some node along the entire correct path stemming from the reference node is less than the current running threshold.

Properties (1), (2), and (3) may be combined to enable us to write a mathematical expression overbounding C, the total number of computations that must eventually be done on all nodes in the incorrect subset:

$$C \leq (u+1) \sum_{m=-1}^{\infty} \sum_{\ell=0}^{\infty} \sum_{n=1}^{M(\ell)} S\left[\left(\sum_{k=1}^{\ell v} Z_k'(n) - \min_{0 \leq h < \infty} \sum_{k=1}^{hv} Z_k^o\right) - m\Delta\right], \tag{3}$$

where $M(\ell) = (u-1)u^{\ell-1}$, and $S[\cdot]$ is the unit step function which is one when its argument is zero or positive, and zero otherwise. Z_k^o and $Z_k'(n)$ are the k^{th} path metric increments on the correct path and on the n^{th} incorrect path, respectively.

The minimum over h in (3) may be removed by the use of a simple union bound:

$$C < (u+1) \sum_{m=-1}^{\infty} \sum_{\ell=0}^{\infty} \sum_{h=0}^{\infty} \sum_{n=1}^{M(\ell)} S\left[\sum_{k=1}^{\ell v} Z_k'(n) - \sum_{k=1}^{hv} Z_k^o - m\Delta\right]. \tag{4}$$

For the upper bound, following Savage, we use a form of the Chebysheff inequality:

If C is a positive random variable,

$$\text{pr}(C \geq L) \leq \frac{\overline{C^{\alpha}}}{L^{\alpha}} \qquad \text{for } \alpha > 0. \tag{5}$$

Thus finding a Pareto upper bound on the cumulative distribution of computation is equivalent to showing that moments of C lower than the α^{th} are bounded for α satisfying (2).

The upper bound on $\overline{C^{\alpha}}$ is established by a random-coding argument. It is assumed that the channel is discrete and memoryless, has an input alphabet of P symbols and an output alphabet of Q symbols and that it is characterized by the set of transition probabilities $\{q_{ij}, i=1,2,\ldots,P; j=1,2,\ldots,Q\}$. The tree characterizing the code is assumed to be infinite, implying a convolution code with infinite constraint length . There are u branches diverging from every node and v symbols of the channel input alphabet for each branch (the code rate, R, is then $\frac{1}{v}$ ln u nats per channel use). Each symbol is picked statistically independently from a probability distribution $\{p_i, i=1,2,\ldots,P\}$. Thus the joint probability is $p\left(x_k^o\right) p\left(y_k \middle| x_k^o\right) p(x_k'(n))$ that during the k^{th} use of the channel a symbol x_k^o is transmitted, a symbol y_k is received and that the k^{th} symbol on the n^{th} incorrect path up to depth k is $x_k'(n)$. Similarly, for the first L uses of the channel, the sequences of symbols may be written as L-component vectors, and the joint probability measure is

$$p\left(\bar{x}_L^o\right) p\left(\bar{y}_L \middle| \bar{x}_L^o\right) p(\bar{x}_L'(n)) = \prod_{k=1}^{L} p\left(x_k^o\right) \prod_{k=1}^{L} p\left(y_k \middle| x_k^o\right) \prod_{k=1}^{L} p(x_k'(n)), \tag{6}$$

where, for example, $\bar{x}_L^o$ represents the first L input symbols to the channel. There is also a probability distribution defined on the channel output symbols $f_j = \sum_{i=1}^{P} p_i q_{ij}$ $j=1,2,\ldots,Q$.

Now since successive uses of the channel are assumed to be statistically independent, we have from the definitions of $Z_k'(n)$ and Z_k^o and our joint probability measure that

$$\sum_{k=1}^{\ell v} Z_k'(n) - \sum_{k=1}^{hv} Z_k^o = \ln \left[\frac{p(\bar{y}_{\ell v} | \bar{x}'_{\ell v}(n))\ f(\bar{y}_{hv})}{f(\bar{y}_{\ell v})\ p\left(\bar{y}_{hv} \middle| \bar{x}_{hv}^o\right)} \right] - (\ell - h)R \tag{7}$$

for the n^{th} incorrect path at depth ℓ, $(n=1,2,\ldots,M(\ell))$.

In bounding $\overline{C^{\alpha}}$ we exploit the Chernoff bounding technique, in which the unit step function S[t] is overbounded by $e^{\frac{t}{1+\alpha}}$ before averaging. The second principal mathematical artifice is the use (on all but the innermost sum on n in (4)) of the following standard inequality[8]:

For a sequence of positive random variables $\{x_i\}$,

$$\overline{\left(\sum_i x_i\right)^\alpha} \leq \sum_i \overline{x_i^\alpha}, \qquad \text{provided} \quad 0 \leq \alpha \leq 1. \tag{8}$$

Then after some algebraic manipulations, we obtain

$$\overline{C^\alpha} < A_1 \sum_{\ell=0}^{\infty} e^{-\frac{\alpha \ell v R}{\alpha+1}} \sum_{h=0}^{\infty} e^{\frac{\alpha h v R}{\alpha+1}} F(h, \ell), \tag{9}$$

where

$$A_1 = \frac{(u+1)^\alpha e^{\frac{\Delta\alpha}{\alpha+1}}}{\left(1 - e^{-\frac{\Delta}{\alpha+1}}\right)^\alpha}$$

which is a constant, and

$$F(h, \ell) = \left[\sum_{n=1}^{M(\ell)} \left(\frac{p(\bar{y}_{\ell v} | \bar{x}'_{\ell v}(n))\ f(\bar{y}_{hv})}{p\left(\bar{y}_{hv} | \bar{x}^o_{hv}\right) f(\bar{y}_{\ell v})} \right)^{\frac{1}{1+\alpha}} \right]^\alpha .$$

It is straightforward to obtain a simple overbound on $F(h, \ell)$ by using the basic probability measure (6) and the inequality[8] $\bar{x}^o < \overline{x^\rho}$ if $0 < \rho < 1$ and x is a positive random variable. The bound on $F(h, \ell)$ was originally derived by Yudkin.[4] Substitution of this bound in (9), leads to an infinite sum which converges if

$$R < \frac{1}{\alpha} E_o(\alpha)$$

where $E_o(\alpha) = -\ln \sum_{j=1}^{Q} \left(\sum_{i=1}^{P} p_i q_{ij}^{\frac{1}{1+\alpha}} \right)^{1+\alpha}$. Thus $\overline{C^\alpha}$, for $0 < \alpha < 1$, is bounded if

$$R < \frac{1}{\alpha} E_o(\alpha). \tag{10}$$

Condition (10) agrees with the lower-bound condition found by Jacobs and Berlekamp.[5] Thus we have obtained the (asymptotically) tightest possible result. Condition (10) is also identical to Savage's upper bound for integer values of α.[3] Recently, Yudkin has extended this result to all $\alpha \geq 1$.[4]

2. Remarks

We may now employ (5) to obtain an upper bound on the distribution of computation. Over the ensemble of infinite tree codes, $\text{pr}(C > L) < \overline{C^{\rho}} L^{-\rho}$, provided $R < \frac{1}{\rho} E_o(\rho)$ and $\rho > 0$. If $R = \frac{E_o(\alpha)}{\alpha}$, the tightest asymptotic result follows from letting ρ approach α.

$$\text{Thus for } R = \frac{E_o(\alpha)}{\alpha}, \quad \text{pr}(C \geq L) < \overline{C^{\alpha-\epsilon}} L^{-(\alpha-\epsilon)}, \tag{11}$$

where ϵ is positive but arbitrarily small. Comparison of our result with the lower bound of Jacobs and Berlekamp shows that (neglecting ϵ);

$$\text{pr}(C \geq L) = AL^{-\alpha} \qquad (L \gg 1).$$

For any R and α such that $R = \frac{E_o(\alpha)}{\alpha}$, where A is a finite constant, it can be shown that $\frac{E_o(\alpha)}{\alpha}$ approaches C, the channel capacity as α approaches zero.

Figure XV-2 shows the Pareto exponent α, as a function of R for the communication system described in the previous report,[1] consisting of binary antipodal signalling, white Gaussian noise channel, matched filter and 8-level output quantizer. The only difference is that the outer quantization levels have been changed from ±2T to ±1.7 T.[9] We are indebted to Professor I. Jacobs for providing these curves. Note the high sensitivity of the Pareto exponent to small changes in rate, both above and below R_{comp}.

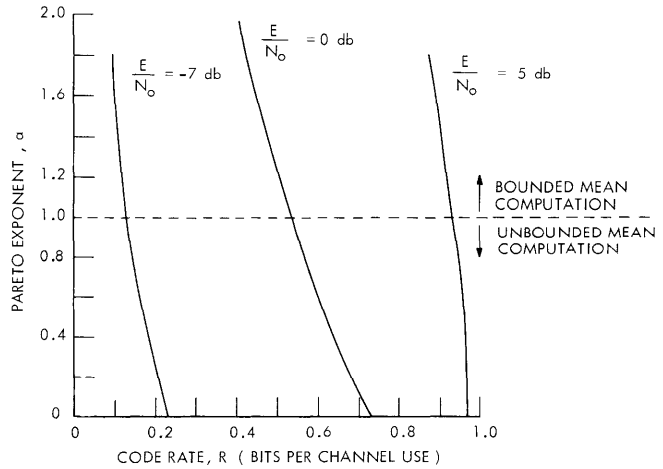

Fig. XV-2. Variation of α with R for a white Gaussian noise channel.

We observe that if $R_{comp} < R < C$, then $0 < \alpha < 1$, and the average computation is unbounded; however, an asymptotically Pareto cumulative distribution still exists. For some applications, operation at a rate above R_{comp} would still be possible (and perhaps feasible). For example, the average number of computations per digit could be made finite simply by imposing an upper limit on the allowable number of computations per digit or group of digits, and passing on as erasures any group of digits for which this limit is exceeded. In such a scheme, periodic resynchronization or a feedback channel would be necessary to allow the decoder to continue past a group of "erased" digits. If no gaps in the incoming data stream can be

tolerated, the erasures may be corrected by an outer level of block encoding and decoding. As a further feature, the upper limit on computation could be variable, being made just large enough to enable the number of erasures in a block of the outer code to be correctable by that code.

Concatenation schemes of this type are being investigated analytically and by simulation.

D. D. Falconer

References

1. D. Falconer and C. Niessen, "Simulation of Sequential Decoding for a Telemetry Channel," Quarterly Progress Report No. 80, Research Laboratory of Electronics, M. I. T., January 15, 1966, pp. 180-193.
2. R. G. Gallager, "A Simple Derivation of the Coding Theorem and Some Applications," IEEE Trans., Vol. IT-11, No. 1, pp. 3-18, January 1965.
3. J. E. Savage, "The Computation Problem with Sequential Decoding," Ph. D. Thesis, Department of Electrical Engineering, M. I. T., February, 1965.
4. H. L. Yudkin (paper submitted for publication).
5. I. M. Jacobs and E. Berlekamp, "A Lower Bound to the Distribution of Computation for Sequential Decoding," JPL SPS 37-34, Vol. 4 (to appear also in IEEE Trans. on Information Theory).
6. J. M. Wozencraft and I. M. Jacobs, Principles of Communication Engineering (John Wiley and Sons, Inc., New York, 1965).
7. G. Blustein and K. L. Jordan, Jr., "An Investigation of the Fano Sequential Decoding Algorithm by Computer Simulation," Group Report 62G-5, Lincoln Laboratory, M. I. T., 1963.
8. G. H. Hardy, J. E. Littlewood, and G. Polya, Inequalities (Cambridge University Press, London, 1959).
9. See Fig. XXII-1, Quarterly Progress Report No. 80, op. cit., p. 184.