IX. PROCESSING AND TRANSMISSION OF INFORMATION*

Prof. D. N. Arden	Prof. J. M. Wozencraft	T. S. Huang
Prof. E. Arthurs	S. Asano	F. Jelinek
Prof. J. B. Dennis	G. Cheadle	T. Kailath
Prof. M. Eden	J. E. Cunningham	L. Kleinrock
Prof. P. Elias	P. M. Ebert	H. H. Loomis, Jr.
Prof. R. M. Fano	H. A. Ernst	J. L. Massey
Prof. R. G. Gallager	E. F. Ferretti	J. W. Pan
Prof. E. M. Hofstetter	T. J. Goblick, Jr.	D. L. Reich
Prof. D. A. Huffman	U. F. Gronemann	L. G. Roberts
Prof. W. F. Schreiber	F. C. Hennie III	O. J. Tretiak
Prof. C. E. Shannon		J. Ziv

A. PICTURE PROCESSING RESEARCH

We have now made computer tapes from original pictures and played the tapes back for photography. Operations on the IBM 709 computer will begin during the next reporting period.

An example of a 1024 × 1024 element picture is shown in Fig. IX-1.

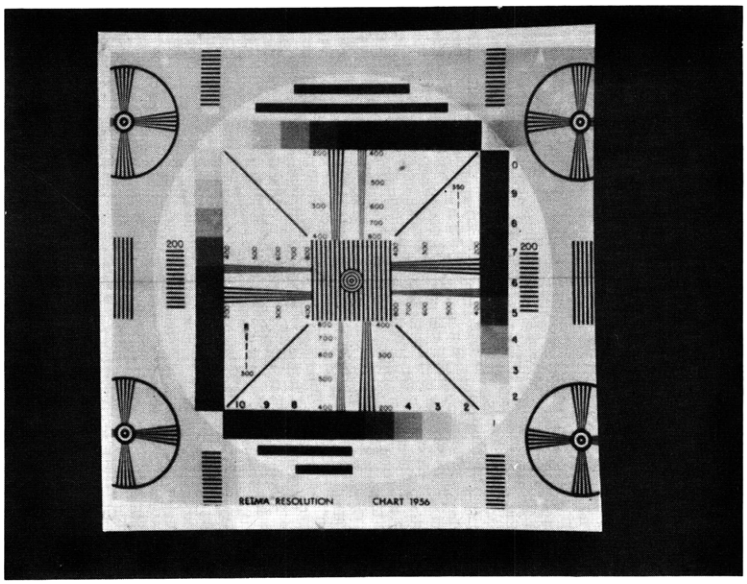

Fig. IX-1.

J. E. Cunningham, U. F. Gronemann, T. S. Huang, J. W. Pan, O. J. Tretiak, W. F. Schreiber

*This research was supported in part by Purchase Order DDL B-00306 with Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology, with the joint support of the U.S. Army, Navy, and Air Force under Air Force Contract AF19(604)-7400.

B. NEW SYSTEMATIC DECODING FOR MEMORYLESS CHANNELS

In this report a sequential decoding scheme for random convolutional codes, which is to be used for memoryless channels, is described. The average number of computations does not grow exponentially with n; it is upper-bounded by a quantity proportional to n^2, for all rates below some cutoff rate R_{comp} (n is the block length).

When this decoding scheme is averaged over a suitably defined ensemble of code words it has an average probability of error with an upper bound whose logarithm is $-nE(R)$. $E(R)$ is dependent on the data rate. ($E(R) > 0$ for rates below channel capacity.) The decoding scheme is different from other effective decoding schemes such as sequential decoding[1] and low-density parity-check codes.[2]

The lower bound on R_{comp} of the systematic decoding scheme that is presented in this report is the same as the R_{comp} of sequential decoding for asymmetric channels.

However, in the case of sequential decoding, R_{comp} is valid only for the incorrect subset of code words. The existence of R_{comp} for the correct subset has not yet been established.

Thus, the systematic decoding scheme yields a better bound on the average number of computations for asymmetric channels. (This is not the case when the channel is symmetric, since the modified sequential decoding scheme after Gallager[3] may be used.)

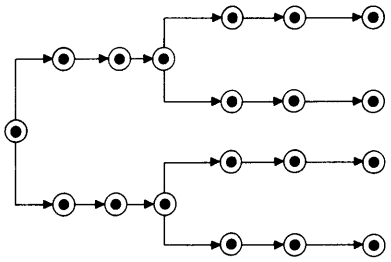

Fig. IX-2. A convolutional tree code.

A convolutional tree code may be viewed topologically as shown in Fig. IX-2. The words are all the directed paths from the input node to the output nodes of the tree (there are no closed paths). From all nontrivial intermediate nodes there emerge ℓ directed links, one for each of ℓ nodes. Let the number of input symbols per code word be n. Let the number of input symbols per link be d. Then the number of links per word is $m = n/d$ (m is the number of information digits per word). In Fig. IX-2, $n = 6$; $m = 2$; $d = 3$; $\ell = 2$.

Reiffen[4] has shown that the convolutional codes may be generated sequentially.

1. The Decoding Procedure

The decoding procedure consists of the following successive operations:

Step 1: The a posteriori probability of each one of the ℓ links of length d that

emerge from the input node to the first ℓ nodes in the tree is computed. The one link that yields the largest a posteriori probability is chosen to represent the corresponding part of the transmitted code word. This detected link connects the input node with one of the ℓ nodes of the next set of nodes (set I in Fig. IX-3).

The same procedure is then repeated with the detected node of set I as a starting point. Thus, the a posteriori probability of each one of the ℓ links emerging from the

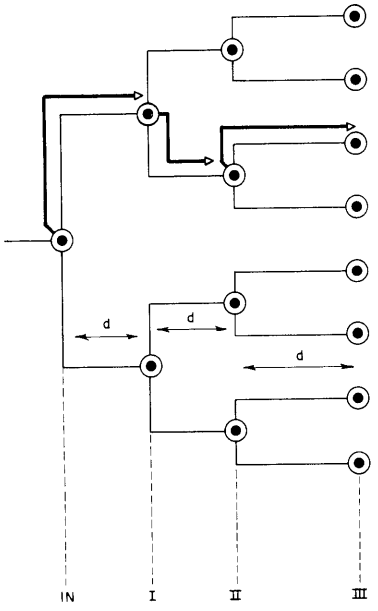

Fig. IX-3. The decoding procedure of step 1.

node that was previously detected, is now computed, and a decision is made. This procedure is then repeated again and again until termination (i.e., until the detected path reaches one of the output nodes). A metric of the form $D(u, v) = \ln\left[\frac{P(v/u)}{P(v)}\right]$ is then computed. Here, $P(v/u)$ is the a posteriori probability of the complete detected word, u, and $P(v)$ is the probability of the output symbol v.

If $D(u, v)$ is larger than some preset threshold D_o, a final decision is made and the detection of the first information digit is completed. If $D(u, v) \leq D_o$, the computation procedure is then to go to step 2.

Step 2: The a posteriori probability of each one of the ℓ^2 links of length 2d that emerge from the input node to set II (that consist of ℓ^2 nodes) is computed (Fig. IX-4). The one link that yields the largest a posteriori probability is chosen to represent the corresponding part of the transmitted code word.

The same procedure is then repeated with the detected node of set I as a starting point, and so on. This procedure is continued until termination (i.e., until the detected path reaches one of the output nodes).

D(u, v) is then computed for the detected path and compared with the threshold D_o. If $D(u, v) > D_o$, a final decision is made with respect to the first information digit. If

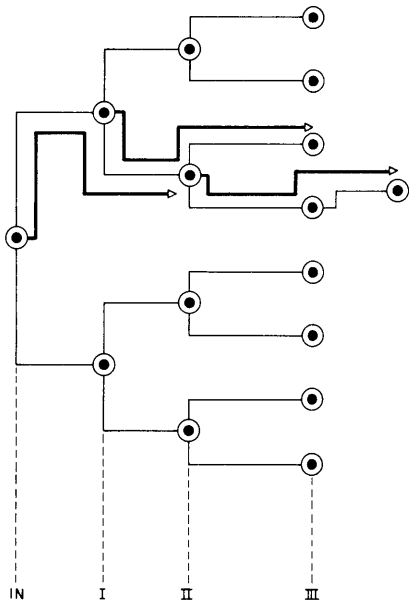

Fig. IX-4. The decoding procedure of step 2.

$D(u, v) \leq D_o$ the computation procedure is then to go to step 3.

Step 3: This step is similar to step 2, except that the computation length in this case is 3d, and the number of links involved at each stage is ℓ^3. If no termination occurs at step 3, the computation procedure then reaches step 4, and so on. Following the detection of the first information digit, the whole procedure is repeated for the next digit, and so forth.

a. The Average Number of Computations per Information Digit

The number of computations that are involved in step 1 is equal to $m \cdot \ell$. The number of computations in step 2 is equal to $(m \cdot \ell^2)$. The number of computations in step 3 is equal to $(m \cdot \ell^3)$.

In general, the number of computations in step k is

$$N_k = m\ell^k$$

Let C_k be the condition that no termination occur at step k. Step k will be used only if there were no terminations at all in the previous k - 1 steps. Thus the probability of step k being used is

$$P(k) = P_r(C_1, C_2, C_3, C_4 \ldots C_{k-1}). \tag{1}$$

The average number of computations is given by

$$N = N_1 P(1) + N_2 P(2) + \ldots + N_k P(k) + \ldots N_m P(m)$$

$$= \sum_{k=1}^{m} N_k P(k) = \sum_{k=1}^{m} m \ell^k P(k). \tag{2}$$

b. An Upper Bound on the Average Number of Computations

P(k) may be bounded by

$$P(k) = P_r(C_1, C_2, C_3, C_4 \ldots C_{k-1})$$

$$\leq P_r(C_{k-1}). \tag{3}$$

Thus

$$N = \sum_{k=1}^{m} N_i P_r(C_{k-1}) = \sum_{k=1} (m-k+1)\, \ell^k P_r(C_{k-1}). \tag{4}$$

Now let u_k be the code word detected at step k, and let u be the transmitted code word. Then

$$P_r(C_{k-1}) = P_r(D(u_k, v) \leq D_o)$$

$$= P_r(D(u_k, v) \leq D_o; u_k = u) + P_r(D(u_k, v) \leq D_o; u_k \neq u)$$

$$= P_r(D(u, v) \leq D_o; u_k = u) + P_r(D(u_k, v) \leq D_o; u_k \neq u)$$

Thus

$$P_r(C_{k-1}) \leq P_r[D(u, v) \leq D_o] + P_r[u_k \neq u] \tag{5}$$

The threshold D_o is set so as to make

$$P_r[D(u, v) \leq D_o] \leq e^{-\epsilon(R) \cdot n}, \tag{6}$$

where $-\epsilon(R)$ is a function of the rate R.

The number of detected links needed to construct u_k is, as we have shown, (m-k+1). Now, u is equal to u_k only if all of the (m-k+1) links of u_k are equal to the corresponding links of u.

Let e_i be the condition that the a posteriori probability of one of the $(\ell-1)\ell^{k-1}$ links emerging from the i^{th} node of u and not including the $(i+1)^{th}$ node of u is greater or equal to that of the corresponding link of u. Then

$$P_r(u_k=u) = P_r(\text{not } e_1;\ \text{not } e_2;\ \text{not } e_3;\ \ldots;\ \text{not } e_i \ldots;\ \text{not } e_{m-k+1})$$

or

$$P_r(u_k \neq u) = 1 - P_r(u_k=u) = P_r\left[\bigcup_{i=1}^{m-k+1} \{e_i\}\right].$$

The probability of the union of events is upper-bounded by the sum of the probabilities of the individual events. Thus

$$P_r(u_k \neq u) \leq \sum_{i=1}^{m-k+1} P_r(e_i) \tag{7}$$

The rate per symbol is defined as

$$R = \frac{1}{n} \ln M = \frac{1}{n} \ln \ell^m = \frac{1}{n} \ln \ell^{n/d} = \frac{1}{d} \ln \ell,$$

where M is the number of code words of length n. Thus

$$R = \frac{1}{d} \ln \ell \tag{8}$$

Fano has shown that for a given rate R and a given word length n the probability of error is bounded by

$$P(e) \leq 2\, e^{-E_{opt}(R) \cdot n}. \tag{9}$$

$E(R)_{opt}$ is the optimum exponent of the average probability of error and is a function of the rate R. Now, in the case of $P_r(e_i)$, the number of the involved links is $(\ell-1)\,\ell^{k-1}$ and the length of each link is kd; thus

$$R_k < \frac{1}{dk} \ln \ell^k = \frac{1}{d} \ln \ell = R$$

Therefore

$$P_r(e_i) \leq 2\, e^{-E_{opt}(R)kd} \tag{10}$$

Thus, by Eqs. 7 and 10,

$$P(u_k \neq u) \leq 2m\, e^{-E_{opt}(R)kd} \tag{11}$$

Therefore, by Eqs. 5, 6, and 11,

$$P_r(C_k) \leq e^{-\epsilon(R)n} + 2m\, e^{-E(R)_{opt}kd} \tag{12}$$

or

$$P_r(C_k) \leq 2m\, e^{-E(R)kd} \tag{13}$$

because, as we shall show, by Eq. 21, $\epsilon(R) \leq E_{opt}(R)$. Thus

$$P_r(C_{k-1}) \leq 2m\, e^{-E(R)(k-1)d} \tag{14}$$

The average number of computations, by Eqs. 4 and 14, is therefore bounded by

$$N \leq 2m^2 \ell \sum_{k=1}^{m} \ell^k e^{-\epsilon(R)(k-1)d} \tag{15}$$

Now,

$$R = \frac{1}{(k-1)\,d} \ell^{(k-1)d}.$$

Thus $\ell^{k-1} = e^{(k-1)dR}$. Therefore

$$N \leq 2m^2 \ell \sum_{k=1}^{m} e^{(k-1)d[R-\epsilon(R)]}. \tag{16}$$

Let R_{comp} be the solution of

$$R = \epsilon(R) \tag{17}$$

Then, for all rates below R_{comp}, $R - \epsilon(R) < 0$, and

$$N \leq 2m^2 \ell \left[\frac{1}{1 - e^{[R-\epsilon(R)]d}}\right]. \tag{18}$$

The average amount of computations is therefore an algebraic function of m for all rates below R_{comp}.

c. Evaluation of R_{comp}

Fano[5] has shown that

$$E_{opt}(R) = E_{opt}(0) - R; \quad \text{for } R \leq R_{crit} \tag{19}$$

and

$$E_{opt}(R) \geq E_{opt}(0) - R; \quad \text{for } R_{crit} \leq R < C. \tag{20}$$

Let us set D_o so as to make $\epsilon(R)$ of Eq. 6 equal to

$$\epsilon(R) = E_{opt}(0) - R_{comp}. \tag{21}$$

Thus, by Eqs. 12 and 21,

$$R_{comp} = E_{opt}(0) - R_{comp}$$

or

$$R_{comp} = \frac{1}{2} E_{opt}(0). \tag{22}$$

Also,

$$\epsilon(R) = E_{opt}(0) - R_{comp} = \frac{1}{2} E_{opt}(0). \tag{23}$$

Thus

$$N \leq 2m^2 \ell \left[\frac{1}{1 - \exp\left\{ d\left[R - \frac{1}{2} E_{opt}(0) \right] \right\}} \right].$$

$E_{opt}(0)$ is the zero-rate exponent of the upper bound on the average probability of error P_e of the same channel, when the optimal decoding scheme is used.[5]

2. The Average Probability of Error

Let u be the transmitted code word. Let v be the corresponding channel output vector. Let u' be one of the $\frac{M(\ell-1)}{\ell}$ code words which starts with an information letter other than that of u. The probability of error is bounded by

$$P_e \leq P_r(D(u, v) \leq D_o) + \frac{M(\ell-1)}{\ell} P_r(D(u', v) > D_o; D(u, v) > D_o) \tag{24}$$

Thus

$$P_e \leq P_r(D(u, v) \leq D_o) + e^{nR} P_r(D(u', v) > D_o; D(u, v) > D_o) \tag{25}$$

or

$$P_e \leq P_r(D(u, v) \leq D_o) + e^{nR} P_r(D(u', v) > D; D(u, v) > D_o) \tag{26}$$

Let

$$D(u, v) = \ln \frac{P(v/u)}{P(v)}$$

$$D(u', v) = \ln \frac{P(v/u')}{P(v)}$$

where

$$P(v/u) = \prod_{i=1}^{n} P(y_i/x_i)$$

and

$$P(v) = \prod_{i=1}^{n} P(y_i).$$

$P(y_i/x_i)$ is the probability that the output is y_i, given that the input symbol x_i was transmitted.

$$\sum_{x_i} P(x_i)\, P(y_i/x_i) = P(y_i)$$

Thus

$$D(u, v) = \sum_{i=1}^{n} d_i(x_i, y_i),$$

where

$$d_i(x_i, y_i) = \ln \frac{P(y_i/x_i)}{P(y_i)}.$$

Thus by the use of the Chernoff Bound,[6]

$$P_r(D(u, v) \leq D_o) \leq e^{n(\mu(s) - s\mu'(s))}; \quad s \leq 0$$

$$P_r(D(u', v) \leq D_o) \leq e^{n(\gamma(t) - t\gamma'(t))}; \quad t \geq 0$$

where

$$\mu(s) = \sum_{xy} P(x)\, P(y/x)\, e^{sd(x, y)}$$

$$\gamma(t) = \sum_{xxy} P(x)\, P(x')\, P(y/x)\, e^{sd(x', y)}$$

or

$$\mu(s) = \sum_{xy} P(x)\, P(y/x)^{1+s}\, P(y)^{-s}$$

and

$$\gamma(t) = \sum_{x'y} P(x')\, P(y)^{1-t}\, P(y/x')^{t}$$

$$= \sum_{xy} P(x)\, P(y)^{1-t}\, P(y/x)^{t}.$$

Also,

$$\mu'(s) = \gamma'(t) = \frac{D_o}{n}.$$

If we let $t = 1 + s$, it can be shown that $\gamma'(t) = \mu'(t)$ (also $\gamma(t) = \mu(s)$). Thus

$$P_r(D(u, v) \leq D_o) \leq e^{n(\mu(s)-s\mu'(s))}$$

$$P_r(D(u', v) \geq D_o) \leq e^{n(\mu(s)-(1+s)\mu'(s))},$$

where

$$\mu'(s) = \frac{D_o}{n}$$

$$\mu(s) = \sum_{xy} P(x)\, P(y/x)^{1+s}\, P(y)^{-s}.$$

Thus

$$P_e \leq e^{n(\mu(s)-s\mu'(s))} + e^{n(R+\mu(s)-(1+s)\mu'(s))}. \tag{27}$$

Now, by Eqs. 6 and 23, D_o is set so as to get

$$P_r(D(u, v) \leq D_o) \leq e^{n(\mu(s)-s-'(s))} = e^{-n\epsilon(R)}, \tag{28}$$

where

$$\epsilon(R) = \frac{1}{2} E_{opt}(0); \quad \text{for all } R \leq R_{comp} \tag{28a}$$

and, as shown by Fano,[5]

$$E_{opt}(0) = E_{opt}(R)\Big|_{R=0} = -\ln \sum_y \left[\sum_x P(x)\, P(y/x)^{\frac{1}{2}}\right]^2. \tag{28b}$$

We shall now prove that, once D_o is set so as to make $-\mu(s) + s\mu'(s) = \epsilon(R) = \frac{1}{2} E_{opt}(0)$, as in Eq. 28, then

$$-\mu(s) + (1+s)\, \mu'(s) \geq \frac{1}{2} E_{opt}(0). \tag{29}$$

PROOF: The minimum of $\{(2s+1)\mu'(s)-2\mu(s)\}$ occurs at that s for which

$$\frac{\partial}{\partial s}\left[(2s+1)\mu'(s)-2\mu(s)\right] = 0$$

$$(2s+1)\, \mu''(s) + 2\mu'(s) - 2\mu'(s) = 0$$

Thus $s = -\frac{1}{2}$.

Also,

$$[(1+2s)\mu'(s)-2\mu(s)]''\Big|_{s=-1/2} = 2\mu''\left(-\tfrac{1}{2}\right) \geq 0,$$

since $\mu''\left(-\frac{1}{2}\right)$ is the variance (see Fano[5]) of a random variable. Thus $s = -\frac{1}{2}$ is indeed a minimum point. Thus

$$[(2s+1)\mu'(s)-2\mu(s)] \geq -2\mu\left(-\tfrac{1}{2}\right). \tag{30}$$

Now

$$\mu\left(-\tfrac{1}{2}\right) = \ln \sum_{xy} P(x)\, P(y)^{1/2}\, P(y/x)^{1/2}$$

and therefore

$$2\mu\left(-\tfrac{1}{2}\right) = \ln\left[\sum_{xy} P(x)\, P(y)^{1/2}\, P(y/x)^{1/2}\right]^2.$$

Thus

$$2\mu\left(-\tfrac{1}{2}\right) = \ln\left\{\sum_{y} f(y)\, P(y)^{1/2}\right\}^2$$

where $f(y) = \sum_{x} P(x)\, P(y/x)^{1/2}$. By the Schwarz inequality,

$$\left\{\sum_{y} f(y)\, P(y)^{1/2}\right\}^2 \leq \sum_{y} f(y)^2 \sum_{y} P(y)$$

Thus

$$2\mu\left(-\tfrac{1}{2}\right) \leq \sum_{y} P(y)^2 \sum_{y} P(y) = \sum_{y} P(y)^2,$$

since $\sum_{y} P(y) = 1$. Therefore, by Eq. 28b,

$$2\mu\left(-\tfrac{1}{2}\right) \leq \ln \sum_{y}\left[\sum_{x} P(x)\, P(y/x)^{1/2}\right]^2 = -E_{opt}(0).$$

Thus $-2\mu\left(-\frac{1}{2}\right) \geq E_{opt}(0)$, and therefore, by Eq. 30,

$$-[\mu(s)-s\mu'(s)+\mu(s)-(1+s)\mu'(s)] \geq E_{opt}(0). \tag{31}$$

But, by Eq. 28,

$$-E(R) = \mu(s) - s\mu'(s) = -\frac{1}{2}E_{opt}(0).$$

Therefore, by Eq. 31,

$$\mu(s) - (1+s)\,\mu'(s) \geq -\frac{1}{2}\,E_{opt}(0) \qquad \text{Q.E.D.} \tag{32}$$

Thus, by Eqs. 27, 28, and 32,

$$P_e \leq e^{-\frac{1}{2}E_{opt}(0)n} + e^{n\left(R-\frac{1}{2}E_{opt}(0)\right)}$$

$$P_e \leq 2\,e^{n\left(R-\frac{1}{2}E_{opt}(0)\right)} = 2\,e^{n(R-R_{comp})}$$

$$P_e \leq 2\,e^{n(R-R_{comp})}.$$

If D_o is set so as to minimize the probability of error by making $R + \mu(s) - (1+s)\mu'(s) = \mu(s) - s\mu'(s) = E(R)$, Shannon[7] has shown that

$$P_e \leq 2\,e^{-nE(R)},$$

where $E(R) > 0$ for $R < C$, and $E(0) > \frac{1}{2}E_{opt}(0)$. However, R_{comp} is then lower-bounded by $R_{comp} \geq \frac{1}{2}E(0) > \frac{1}{4}E_{opt}(0)$.

This research was supported in part by a fellowship from the Government of Israel.

J. Ziv

References

1. J. M. Wozencraft and B. Reiffen, Sequential Decoding (The M.I.T. Press, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1961).

2. R. G. Gallager, Low Density Parity Check Codes, Sc.D. Thesis, Department of Electrical Engineering, M.I.T., September 1960.

3. B. Reiffen, Sequential Encoding and Decoding for the Discrete Memoryless Channel, Technical Report 374, Research Laboratory of Electronics, M.I.T.; Technical Report 231, Lincoln Laboratory, M.I.T., August 12, 1960, p. 29.

4. Ibid., see Chapter 2.

5. R. M. Fano, Transmission of Information (The M.I.T. Press, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1961), see Chapter 9.

6. Ibid., see Chapter 8; B. Reiffen, op. cit., see Appendix B.

7. C. E. Shannon, Error Probability Bounds for Noisy Channels (unpublished report, 1959).