Notes on Digital Coding*

The consideration of message coding as a means for approaching the theoretical capacity of a communication channel, while reducing the probability of errors, has suggested the interesting number theoretical problem of devising lossless binary (or other) coding schemes serving to insure the reception of a correct, but reduced, message when an upper limit to the number of transmission errors is postulated.

An example of lossless binary coding is treated by Shannon[1] who considers the case of blocks of seven symbols, one or none of which can be in error. The solution of this case can be extended to blocks of 2^n-1-binary symbols, and, more generally, when coding schemes based on the prime number p are employed, to blocks of $p^n-1/p-1$ symbols which are transmitted, and received with complete equivocation of one or no symbol, each block comprising n redundant symbols designed to remove the equivocation. When encoding the message, the n redundant symbols x_m are determined in terms of the message symbols Y_k from the congruent relations

$$E_m \equiv X_m + \sum_{k=1}^{k=(p^n-1)/\ p-1)-n} a_{mk}\, Y_k \equiv 0 \ (\mathrm{mod}\ p).$$

In the decoding process, the E's are recalculated with the received symbols, and their ensemble forms a number on the base p which determines univocally the mistransmitted symbol and its correction.

In passing from n to $n+1$, the matrix with n rows and $p^n-1/p-1$ columns formed with the coefficients of the X's and Y's in the expression above is repeated p times horizontally, while an $(n+1)$ st row added, consisting of $p^n-1/p-1$ zeroes, followed by as many one's etc. up to $p-1$; an added column of n zeroes with a one for the lowest term completes the new matrix for $n+1$.

If we except the trivial case of blocks of $2S+1$ binary symbols, of which any group comprising up to S symbols can be received in error which equal probability, it does not appear that a search for lossless coding schemes, in which the number of errors is limited but larger than one, can be systematized so as to yield a family of solutions. A necessary but not sufficient condition for the existence of such a lossless coding scheme in the binary system is the existence of three or more first numbers of a line of Pascal's triangle which add up to an exact power of 2. A limited search has revealed two such cases; namely, that of the first three numbers of the 90th line, which add up to 2^{12} and that of the first four numbers of the 23rd line, which add up to 2^{11}. The first case does not correspond to a lossless coding scheme, for, were such a scheme to exist, we could designate by r the number of E_m ensembles corresponding to one error and having an odd number of 1's and by $90-r$ the remaining (even) ensembles. The odd ensembles corresponding to two transmission errors could be formed by re-entering term by term all the conbinations of one even and one odd ensemble corresponding each to one error, and would number $r(90-r)$. We should have $r+r(90-r)=2^{11}$, which is impossible for integral values of r.

On the other side, the second case can be coded so as to yield 12 sure symbols, and the a_{mk} matrix of this case is given in Table I. A second matrix is also given, which is that of the only other lossless coding scheme encountered (in addition to the general class mentioned above) in which blocks of eleven ternary symbols are transmitted with no more than 2 errors, and out of which six sure symbols can be obtained.

It must be mentioned that the use of the ternary coding scheme just mentioned will always result in a power loss, whereas the coding scheme for 23 binary symbols and a maximum of three transmission errors yields a power saving of $1\frac{1}{2}$ db for vanishing probabilities of errors. The saving realized with the coding scheme for blocks of 2^n-1 binary symbols approaches 3 db for increasing n's and decreasing probabilities of error, but a loss is always encountered when $n=3$.

Marcel J. E. Golay
Signal Corps Engineering Laboratories
Fort Monmouth, N. J

TABLE I

	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}
X_1	1	0	0	1	1	1	0	0	0	1	1	1
X_2	1	0	1	0	1	1	0	1	1	0	0	1
X_3	1	0	1	1	0	1	1	0	1	0	1	0
X_4	1	0	1	1	1	0	1	1	0	1	0	0
X_5	1	1	0	0	1	1	1	0	1	1	0	0
X_6	1	1	0	1	0	1	1	1	0	0	0	1
X_7	1	1	0	1	1	0	0	1	1	0	1	0
X_8	1	1	1	0	0	1	0	1	0	1	1	0
X_9	1	1	1	0	1	0	1	0	0	0	1	1
X_{10}	1	1	1	1	0	0	0	0	1	1	0	1
X_{11}	0	1	1	1	1	1	1	1	1	1	1	1

	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
X_1	1	1	1	2	2	0
X_2	1	1	2	1	0	2
X_3	1	2	1	0	1	2
X_4	1	2	0	1	2	1
X_5	1	0	2	2	1	1

* Received by the Institute, February 23, 1949.

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. Jour.*, vol. 27, p. 418; July, 1948.

Reprinted from *Proc. IRE,* vol. 37, p. 657, June 1949.