Reprinted from "Encyclopaedia of Linguistics, Information and Control"
PERGAMON PRESS
OXFORD · LONDON · EDINBURGH · NEW YORK
TORONTO · SYDNEY · PARIS · BRAUNSCHWEIG

ERROR CORRECTING CODES. A simple idealization of a noisy communication channel is the memoryless *binary symmetric channel* (BSC) of Fig. 1. When a binit

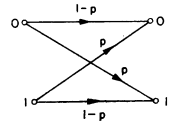

FIG. 1. *The Memoryless Binary Symmetric Channel (or BSC).*

(*bin*ary dig*it*) is transmitted, it is erroneously received with a probability p, $p \leqq \frac{1}{2}$. The information-theoretic capacity of this channel is

$$C = 1 - H(p) \qquad \text{bits/use} \tag{1}$$

where $H(p) = -(p)\log_2(p) - (1-p)\log_2(1-p)$ is the entropy function of a binary ensemble. The development of error correcting codes is largely a product of the study of means for transmitting information reliably over the BSC.

In *block coding*, the sender identifies each of M equiprobable messages with a distinct sequence of n binits which is used to transmit the message over the channel. The *code*, X_n, is just the set of M possible transmitted n-tuples of binits. The message ensemble has an entropy of $\log_2(M)$ bits, hence the *code rate*, R, is just $\log_2(M)/n$ bits per use of the channel. Equivalently, $M = 2^{nR}$.

Let $x = (x_1, x_2, \ldots x_n)$ be a code word in X_n. Each x can be considered as a vertex or point of the unit n-cube. The $R = \frac{1}{3}$ code, $X_3 = \{(0, 0, 0), (1, 1, 1)\}$ is shown in Fig. 2 by shading, the $M = 2$ points of the unit 3-cube

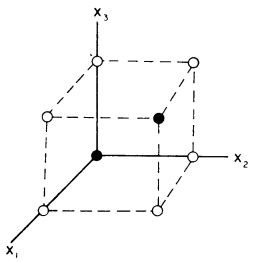

FIG. 2. *The code* $X_3 = \{(0, 0, 0), (1, 1, 1,)\}$ *shown as shaded vertices of the unit 3-cube.*

which are *code points*. When the point x is transmitted over the BSC, the point $y = (y_1, y_2, \ldots, y_n)$ is received where $y \neq x$ if errors have occurred. When $R < 1$, not all 2^n vertices of the unit n-cube are code points. In this case more than one vertex y can be *decoded* into the same vertex x at the receiver, i.e. the code has the ability to correct certain errors.

The *Hamming distance*, $d(v, v')$ between vertices v and v' of the n-cube is the number of coordinates in which v and v' differ. The minimum distance, $d_{\min}$, of a code X_n is the minimum Hamming distance between distinct code points in X_n. For X_3 in Fig. 2, $d_{\min} = 3$. A pattern of t errors perturbs a code point x to a point y such that $d(x, y) = t$. Provided that $t \leqq t_{\max} = [(d_{\min} - 1)/2]$, where the bracket denotes the integral part of the enclosed expression, y will be closer to x than to any other code point. Thus all patterns of $t_{\max}$ or fewer errors can be corrected by the code X_n. Moreover, not all patterns of $t_{\max} + 1$ or fewer errors can be corrected so that $t_{\max}$ is the guaranteed error-correcting radius of the code. The code of Fig. 2 is single-error correcting since $t_{\max} = 1$.

A major problem in coding theory is, for given n and R, finding a code X_n which (nearly) maximizes $d_{\min}$ or $t_{\max}$. Limits on the solutions of this problem can be readily established by geometric arguments. A *sphere* of radius r about a vertex v of the n-cube is the set of all vertices v' such that $d(v, v') \leqq r$. The volume, $V(r)$, of the sphere is the number of vertices which it contains, viz.

$$V(r) = \sum_{i=0}^{r} \binom{n}{i} = 2^{n[H(r/n)+\varepsilon]} \tag{2}$$

where ε approaches zero as n approaches infinity for $(r/n) \leqq \frac{1}{2}$.

Given a code X_n with error-correcting radius $t_{\max}$, the spheres of radius $t_{\max}$ about each of the M code points must be disjoint. Since the total volume of these spheres cannot exceed the number of vertices of the n-cube, $(M)V(t_{\max}) \leqq 2^n$ or

$$V(t_{\max}) \leqq 2^{n(1-R)} \tag{3}$$

which is the *Hamming bound*. A code X_n for which (3) holds with equality is called a *sphere-packed code*, a trivial example of which is the code X_3 of Fig. 2. It is known that sphere-packed codes must be exceedingly rare. The only known ($m > 2$) sphere-packed codes are the Hamming codes and the remarkable Golay code. For any m, there is a Hamming code with $n = 2^m - 1$, $t_{\max} = 1$, and $R = (2^m - m - 1)/(2^m - 1)$. The *Golay code* has $n = 23$, $t_{\max} = 3$, and $R = 12/23$. For large n, the Hamming bound becomes $H(t_{\max}/n) \leqq 1 - R$, or equivalently $H(d_{\min}/2n) \leqq 1 - R$. The corresponding upper bound on the attainable $d_{\min}/2n$ ratio is plotted in Fig. 3.

Given any $d \leqq n$, a code X_n with $d_{\min} \geqq d$ can be constructed by the following procedure. Choose the first code point arbitrarily and surround it with a sphere of radius $d-1$. Choose a second code point not in this sphere and surround it with a sphere of radius $d-1$. Choose a third code point not in the *union* of the preceding spheres, etc. At least M code points can be chosen provided that $(M-1)V(d-1) < 2^n$ and this code must have $d_{\min} \geqq d$. Thus $V(d_{\min}) \leqq 2^{n(1-R)}$ is a sufficient condition for the construction of a code X_n with rate R and minimum distance $d_{\min}$. This is the *Gilbert bound*. This bound states that some $d_{\min}/n$ ratio is attainable for which $H(d_{\min}/n) \geqq 1 - R$. This lower bound on the attainable ratio is plotted in Fig. 3.

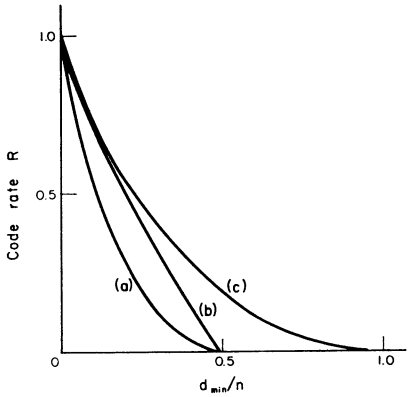

FIG. 3. *Bounds on the ratio of minimum distance $d_{\min}$ to code length n attainable with a block code with rate R:* (a) *Gilbert lowerbound;* (b) *Elias–Shannon–Gallager asymptotic upper bound;* (c) *Hamming asymptotic upper bound.*

An improved asymptotic upper bound on the $d_{\min}/n$ ratio can be found using more subtle arguments. This *Elias–Shannon–Gallager bound* states that $H(\frac{1}{2}-\frac{1}{2}\sqrt{1-2d_{\min}/n}) \leqq 1-R$ and is shown in Fig. 3.

The Gilbert lower bound shows that for any fixed R there exists a sequence of codes X_n such that $d_{\min}/n$ (or $t_{\max}/n$) approaches a non-zero constant as n becomes infinite, but no closed construction procedure is known which realizes this possibility. The most powerful constructive class of codes now known, the *Bose–Chaudhuri–Hoquenghem codes*, attain the Gilbert lower bound out to code lengths n of about 1000 binits.

The search for good codes is greatly simplified by the restriction to *linear codes* (also called *group codes* and *parity-check codes.*) A linear code with rate $R = k/n$ is specified by an $(n-k)\times n$ matrix H with linearly independent rows. This binary matrix H is called the parity-check matrix of the code. All operations with such matrices are assumed to be carried out over the binary number field, i.e. using modulo-two arithmetic. The code X_n defined by H is the set of 2^k binary row vectors $\mathbf{x} = (x_1, x_2, \ldots x_n)$ such that $Hx^T = \mathbf{0}$ where T denotes 'transpose' and $\mathbf{0}$ denotes the $n-k$ dimensional vector of zeroes. Each row of the matrix H specifies a *parity check* for the code since the ones in that row designate a set of coordinates whose values must sum to zero for all code words. The minimum distance of a linear code is always equal to the Hamming weight (i.e. number of non-zero positions) of the non-zero code word in X_n with minimum Hamming weight. It can be shown that there exist linear codes which attain the Gilbert lower bound on the $d_{\min}/n$ ratio for all n and R.

Decoding of linear codes is facilitated by use of the *syndrome*, s, which is the $n-k$ dimensional vector $\mathbf{s} = Hy^T$ where y is the received point. When x is transmitted, $y = x+e$ where e is the error pattern and has a one in each position of y where an error has occurred in transmission. Hence $s = Hy^T = H(x+e)^T = Hx^T + He^T = He^T$ so that s depends only upon the error pattern. Exactly 2^n error patterns give the same syndrome. Optimum decoding is accomplished by decoding y into $x+e$ (addition and subtraction coincide in modulo-two arithmetic) where e is the most probable of the 2^n error patterns consistent with s. For the BSC, the most probable error pattern is that with the fewest ones. For an example of syndrome decoding, consider the Hamming $R = \frac{4}{7}$ code with the parity check matrix

$$H = \begin{bmatrix} 0&0&0&1&1&1&1 \\ 0&1&1&0&0&1&1 \\ 1&0&1&0&1&0&1 \end{bmatrix}. \qquad (4)$$

As with any linear code, $s = He^T = 0$ when no errors occur. If a single error occurs in the ith position, it follows from (4) that $s = He^T = (s_1, s_2, s_3)$ where $s_1s_2s_3$ is the radix-two representation of the integer i. Since the code has $t_{\max} = 1$ and is sphere-packed, it cannot correct more than single errors. Thus the optimum decoding algorithm is simply: Change the ith binit of y if $i = (s_1s_2s_3)_2 \neq 0$.

Encoding is also greatly simplified for linear codes. Any linear code is equivalent (i.e. differs by at most a permutation of the positions in the code words) to a code X_n for which $H = [P : I]$ where I is the $(n-k)\times (n-k)$ identity matrix. Such a code is called *systematic.* In a systematic code, the binits $x_1, x_2, \ldots x_k$ of the code word x may be chosen arbitrarily and hence are called *information bits.* The parity check defined by the ith row of H is then satisfied by choice of the *parity binit* x_{k+i}. Thus the parity binits are uniquely determined once the information bits have been selected.

Further encoding and decoding simplifications are possible for a *linear code* that is also *cyclic.* A linear code is cyclic if the cyclic shift, $(h_1 h_2 \ldots h_{n-1} h_n) \rightarrow (h_2 h_3 \ldots h_n h_1)$, of any row of H produces another row in the row-space of H. For such a code, the cyclic shift of a code word x produces another code word. Very simple shift-register circuits can be used to encode and to form the syndrome for cyclic codes. Moreover, the decoding algorithm need only decode the first binit, y_1, of the received block. The received block is then cycled and the same algorithm used to decode y_2, etc. The Bose-Chaudhuri-Hoquenghem (BCH) codes are cyclic codes as are many other important classes of codes, but it is not known whether there exist cyclic codes which attain the Gilbert lower bound on the $d_{\min}/n$ ratio for fixed R as n becomes infinite.

Cyclic codes have an interesting algebraic interpretation. The code point $x = (x_1, x_2, \ldots x_n)$ is identified with the polynomial $x(X) = x_1X^{n-1} + x_2X^{n-2} + \ldots + x_{n-1}X + x_n$. For any cyclic code X_n, there exists a polynomial $g(X)$ of degree $n-k$ which divides $X^n - 1$ such that the 2^k distinct multiples of $g(X)$ with degree less than n are the polynomials corresponding to the 2^k code words in

X_n. Conversely, any $g(X)$ which divides X^n+1 generates the code words of a cyclic code. The BCH codes are formed by choosing $g(X)$ to be the minimum degree polynomial having among its roots $d-1$ consecutive powers of some specified element β, $\beta \neq 0$, of $GF(2^m)$. $GF(2^m)$ is the finite field of 2^m elements, all $n = 2^m-1$ non-zero elements of which are roots of X^n+1. If the $d-1$ successive powers of β are distinct, then any multiple of $g(X)$ must have at least d non-zero coefficients so that $d_{\min} \geq d$ in the corresponding cyclic code. The most important BCH codes result when β is taken to be a primitive element of $GF(2^m)$. In this case with $d-1 = 2t$, the resulting cyclic codes have block length $n = 2^m-1$, error-correcting radius $t_{\max} \geq t$, and at most $n-k = mt$ parity binits. The Hamming codes are the special case where $t = 1$.

Closely related to the problem of finding codes with (nearly) maximum $d_{\min}$ is the problem of finding codes for which the probability of a decoding error on the BSC is (nearly) minimum. On the BSC, the probability that y is received when x is transmitted decreases monotonically with $d(x, y)$. Thus the decoding error probability is minimized by a *maximum likelihood decoder* which decodes y into the nearest code point x. The probability of a decoding error for a code X_n when used on a BSC and decoded by a maximum likelihood decoder will be denoted $P(X_n)$.

One form of the noisy coding theorem of information theory states that for any rate R less than capacity C, there exists a sequence of codes X_n with rate at least R such that $P(X_n) \to 0$ as $n \to \infty$; conversely no such sequence exists if $R > C$.

The lower bound on attainable $P(X_n)$ is established from the Hamming bound on $t_{\max}$. For any code, $P(X_n)$ must be at least as great as that for a postulated sphere-packed code of the same rate R. A sphere-packed code, however, corrects a pattern of t errors if and only if $t \leq t_{\max}$. The average number of errors in a received block is just np on the BSC, and the probability that the fraction of errors differs from p by any fixed amount $\varepsilon > 0$ approaches zero as n becomes infinite. Thus $P(X_n)$ for the sequence of sphere-packed codes can approach zero only if $t_{\max}/n$ approaches a limit greater than p. Using (1), (2) and (3) this condition becomes $H(t_{\max}/n) = 1-R \geq H(p) = 1-C$, or simply $R \leq C$ which is the converse part of the coding theorem.

The direct part of the coding theorem was proved by Shannon using a 'random coding' argument, i.e. by showing that the average of $P(X_n)$ over all codes of rate R and length n approaches zero as n becomes infinite provided that $R < C$. Only one constructive class of codes, the Elias iterated codes, has been found with the property that the rates approach a non-zero rate R and $P(X_n)$ approaches zero as n becomes infinite. But these codes require R to be considerably less than C and the approach of $P(X_n)$ to zero is much less rapid than is known to be attainable.

Non-block forms of coding have also been studied for the BSC. *Convolutional coding* (also called recurrent coding) is a non-block form of linear coding. In a convolutional code, the transmitted binits are generated in segments of n_0 binits, k_0 of which are information bits. The parity binits in each segment are a modulo-two sum of information binits in that segment and the preceding m segments. Convolutional codes are found to have essentially the error-correcting capability of block codes of length $n = (m+1)n_0$. These codes have played an important theoretical role in *sequential decoding*, which is a form of decoding applicable to any convolutional code. Over the ensemble of all convolutional codes with rate $R = k_0/n_0$, the probability of a decoding error vanishes exponentially as m is increased while the average number of decoding computations is bounded provided that $R < R_{\text{comp}} < C$ where R_{comp} is the computational cut-off rate.

Codes have also been studied for many noisy communication channels besides the BSC. One such channel is the binary burst channel where errors can occur only in clusters spanning at most b channel binits, such clusters being separated by at least g error-free binits. For this simple channel, essentially optimum codes and simple decoding methods are known. Non-binary input channels have also been considered. Virtually all the theory of linear binary codes can be extended to linear q-ary codes where $q = p^m$ is the power of a prime. The code digits in this case are taken to be elements of $GF(p^m)$, the finite field of p^m elements. Error correcting codes have also been developed for non-communication applications. For instance, a special type of block binary code called a *residue code* has been developed to correct the kinds of errors that commonly arise in digital circuits for the adding of integers in radix-two representation, viz. the inversion of binits in the sum and the loss or spurious generation of carry binits.

See also: Communication theory. Random signals and noise. Reliable computation using unreliable elements.

Bibliography

PETERSON W. W. (1961) *Error-Correcting Codes*, Cambridge, Mass.: M.I.T. Press; New York: Wiley.

SHANNON C. E. (1948) *The Mathematical Theory of Communication, Bell Syst. Tech. J.*, **27**, 379, 623.

WOZENCRAFT J. M. and JACOBS I. M. (1966) *Principles of Communication Engineering*, New York: Wiley.

J. L. MASSEY