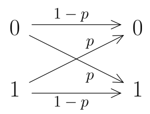
Chapter 3 Error-Correcting Codes

3.1. Coding for a noisy channel. We shall assume that the channel is BS(p), the *binary symmetric channel with crossover probability* p: that is, the only symbols that can be transmitted are 0 and 1, and if $P(a|b)$ denotes the probability that a is received given that b is sent, then

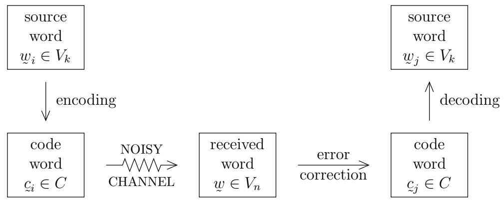
$$P(1|0) = P(0|1) = p,$$
$$P(0|0) = P(1|1) = 1 - p.$$

Diagrammatically:



We shall assume $p < \frac{1}{2}$.

Let S be a memoryless source emitting m source words *with equal probability* $1/m$, and let C be a binary code with m codewords. We assume all codewords have the same length n, that is, $C \subset V_n$, where V_n denotes the set of all 2^n binary words of length n; such a code is a (binary) *block code* of *length* n. The *rate* of C is $r(C) := \frac{\log_2 m}{n} < 1$. Most commonly, the set of source words is V_k for some $k < n$, so that $m = 2^k$ and $r(C) = k/n$. If the source message is broken up into blocks of k digits, each of which is encoded into a codeword of length n, then the rate k/n is the factor by which the message is slowed down by the encoding. We then have the following situation.



We *hope* $\underset{\sim}{c}_j = \underset{\sim}{c}_i$, so that $\underset{\sim}{w}_j = \underset{\sim}{w}_i$.

If the received word $\underset{\sim}{w}$ is not a codeword, we wish to interpret it as the codeword *most likely* to have been sent. Intuitively, this is the codeword *closest* to $\underset{\sim}{w}$. If $\underset{\sim}{x}, \underset{\sim}{y} \in V_n$, their *(Hamming) distance* $d(\underset{\sim}{x}, \underset{\sim}{y})$ is the number of positions in which they differ. The *weight* $w(\underset{\sim}{x})$ of $\underset{\sim}{x}$ is the number of 1's in $\underset{\sim}{x}$; so $w(\underset{\sim}{x}) = d(\underset{\sim}{x}, \underset{\sim}{0})$, where $\underset{\sim}{0} = 0\,0 \ldots 0$ is the word with n 0's. For example,

$$\begin{array}{ll} \text{if} & \underset{\sim}{x} = 0\ 0\ 1\ 0\ 1\ 0 \\ \text{and} & \underset{\sim}{y} = 1\ 0\ 1\ 1\ 0\ 0 \\ & \quad\ \uparrow \quad\ \ \uparrow\uparrow \end{array}$$

then $w(\underset{\sim}{x}) = 2$, $w(\underset{\sim}{y}) = 3$ and $d(\underset{\sim}{x}, \underset{\sim}{y}) = 3$. For all $\underset{\sim}{x}$, $\underset{\sim}{y}$ and $\underset{\sim}{z}$,

(i) $d(\underset{\sim}{x}, \underset{\sim}{y}) = 0$ iff $\underset{\sim}{x} = \underset{\sim}{y}$,

(ii) $d(\underset{\sim}{x}, \underset{\sim}{y}) = d(\underset{\sim}{y}, \underset{\sim}{x})$,

(iii) $d(\underset{\sim}{x}, \underset{\sim}{y}) \leqslant d(\underset{\sim}{x}, \underset{\sim}{z}) + d(\underset{\sim}{z}, \underset{\sim}{y})$ (the triangle inequality).

($d(\underset{\sim}{x}, \underset{\sim}{y})$ is the smallest number of changes of digits that can convert $\underset{\sim}{x}$ to $\underset{\sim}{y}$. But we can change $\underset{\sim}{x}$ to $\underset{\sim}{y}$ by first changing $\underset{\sim}{x}$ to $\underset{\sim}{z}$ and then changing $\underset{\sim}{z}$ to $\underset{\sim}{y}$.) [Metric space axioms.]

Theorem 3.1. When a word $\underset{\sim}{w}$ is received, the codewords $\underset{\sim}{c}$ that are most likely to have been sent are those for which $d(\underset{\sim}{c}, \underset{\sim}{w})$ is smallest.

Proof. (Not for examination.) By Bayes's theorem (twice),

$$\begin{aligned} P(\underset{\sim}{c}\text{ sent} \mid \underset{\sim}{w}\text{ received}) &= \frac{P(\underset{\sim}{c}\text{ sent, } \underset{\sim}{w}\text{ received})}{P(\underset{\sim}{w}\text{ received})} \\ &= \frac{P(\underset{\sim}{c}\text{ sent})P(\underset{\sim}{w}\text{ received} \mid \underset{\sim}{c}\text{ sent})}{P(\underset{\sim}{w}\text{ received})}. \end{aligned}$$

Let $h := \dfrac{P(\underset{\sim}{c}\text{ sent})}{P(\underset{\sim}{w}\text{ received})}$, which is independent of $\underset{\sim}{c}$, since we are assuming that *all codewords are equally likely*. Thus

$$\begin{aligned} P(\underset{\sim}{c}\text{ sent} \mid \underset{\sim}{w}\text{ received}) &= hP(\underset{\sim}{w}\text{ received} \mid \underset{\sim}{c}\text{ sent}) \\ &= hp^{d(\underset{\sim}{c},\underset{\sim}{w})}(1-p)^{n-d(\underset{\sim}{c},\underset{\sim}{w})} \\ &= h(1-p)^n \left(\frac{p}{1-p}\right)^{d(\underset{\sim}{c},\underset{\sim}{w})}, \end{aligned}$$

which is largest when $d(\underset{\sim}{c}, \underset{\sim}{w})$ is smallest, since we are assuming that $p < \frac{1}{2}$ and so $\frac{p}{1-p} < 1$. //

The *ideal observer* always chooses $\underset{\sim}{c}$ so that $P(\underset{\sim}{c}\text{ sent} \mid \underset{\sim}{w}\text{ received})$ is maximal. Theorem 3.1 says that, for a binary symmetric channel with *equally probable* codewords, the ideal observer uses the *minimum-distance* or *nearest-neighbour* decoding scheme.

Example. (See handout.)

A code is *e-error-detecting* if it can *detect* that there is a mistake whenever e or fewer errors are made in any one codeword, and *e-error-correcting* if it can *correct* these errors. The *minimum distance* of a code C is $d(C) := \min\limits_{\substack{\underset{\sim}{c}_i, \underset{\sim}{c}_j \in C \\ \underset{\sim}{c}_i \neq \underset{\sim}{c}_j}} d(\underset{\sim}{c}_i, \underset{\sim}{c}_j)$. It is easy to see that

$$\begin{aligned} C \text{ is } e\text{-error-detecting} &\iff d(C) \geqslant e+1, \\ C \text{ is } e\text{-error-correcting} &\iff d(C) \geqslant 2e+1. \end{aligned}$$

An *(n, m, d)-code* is a code of length n with m codewords and minimum distance d.

Let P_{err} be the average probability that a codeword is incorrectly identified, and let P_{symb} be the average probability of error per symbol after decoding the message. Then $\frac{1}{k}P_{\text{err}} \leqslant P_{\text{symb}} \leqslant P_{\text{err}}$ (where k is the length of the source words), since the proportion of source words that are wrong is also P_{err}, and the proportion of wrong symbols in a wrong source word lies between $\frac{1}{k}$ and $\frac{k}{k} = 1$.

Theorem 3.2. If C is an e-error-correcting code of length n, and $n \leqslant \frac{e}{p} + 1$, then $P_{\text{err}} < \frac{n!}{e!(n-e-1)!}\, p^{e+1}(1-p)^{n-e-1}$.

Proof. (Not for examination.) Let $f(p)$ be the probability that more than e errors are made in a codeword, so that $P_{\text{err}} \leqslant f(p)$. Then

$$\begin{aligned} f(p) &= 1 - (1-p)^n - np(1-p)^{n-1} - \binom{n}{2}p^2(1-p)^{n-2} - \ldots \\ &\qquad \ldots - \binom{n}{e-1}p^{e-1}(1-p)^{n-e+1} - \binom{n}{e}p^e(1-p)^{n-e}. \end{aligned}$$

Note $f(0) = 0$. Also,

$$\begin{aligned} f'(p) &= n(1-p)^{n-1} - n(1-p)^{n-1} + n(n-1)p(1-p)^{n-2} - n(n-1)p(1-p)^{n-2} + \ldots \\ &\qquad \ldots + \frac{n!}{(e-1)!(n-e+1)!}(n-e+1)p^{e-1}(1-p)^{n-e} - \frac{n!}{e!(n-e)!}ep^{e-1}(1-p)^{n-e} \\ &\qquad + \frac{n!}{e!(n-e)!}(n-e)p^e(1-p)^{n-e-1} \\ &= \frac{n!}{e!(n-e-1)!}\,p^e(1-p)^{n-e-1}. \qquad \text{(So we want to prove } f(p) < pf'(p).) \\ f''(p) &= \frac{n!}{e!(n-e)!}\,p^{e-1}(1-p)^{n-e-2}[e(1-p) - (n-e-1)p] \geqslant 0 \end{aligned}$$

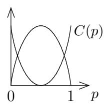
since $n - e - 1 \leqslant \frac{e(1-p)}{p} = \frac{e}{p} - e$ by hypothesis. Moreover, $f''(x) > 0$ if $0 < x < p$ since then $n - e - 1 \leqslant \frac{e}{p} - e < \frac{e}{x} - e$. By the mean-value theorem,

$$\begin{aligned} f(p) &= f(0) + pf'(q) \quad \text{for some } q,\ 0 < q < p, \\ &= 0 + pf'(q) \\ &< pf'(p) \quad \text{since } f''(x) > 0 \text{ if } 0 < x < p. \end{aligned}$$

Hence $P_{\text{err}} \leqslant f(p) < pf'(p)$, as required. //

The upper bound for $f(p)$ in Theorem 3.2 is probably not too bad, but P_{err} is often a lot smaller than $f(p)$ because we can often correct codewords successfully when *more* than e errors have been made.

Since error-correcting ability is achieved by building redundancy into the message, one might expect that the rate k/n would have to $\to 0$ in order for P_{err} to $\to 0$. This is not so. The *capacity* of BS(p) is



$$C(p) := 1 - H_2(p, 1-p) = 1 + p\log_2 p + (1-p)\log_2(1-p),$$

which (by Exercise 1.11) is the information that a received symbol gives about the symbol sent.

Theorem 3.3. *The noisy coding theorem*, or *the fundamental theorem of information theory* (C. E. Shannon, 1948). For any real numbers $R < C(p)$ and $\epsilon > 0$, there exist codes with rate $\geqslant R$ and $P_{\text{err}} < \epsilon$. But if $R > C(p)$ then, for codes with rate $\geqslant R$ and length n, P_{err} is bounded away from 0 by a function of p, R and n that $\to 1$ as $n \to \infty$. //

However, nobody knows of any useful practical codes that realize Shannon's theorem.

3.2. Some linear algebra. The *Galois field* GF(2) or $\mathbb{Z}_2$ is $\{0, 1; +, \times\}$ with $+, \times$ given by

$$\begin{array}{c|cc} + & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array} \qquad\qquad \begin{array}{c|cc} \times & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}$$

(as in $\mathbb{R}$, except that $1+1=0$). Note that $a+a=0\ \forall\, a$, so addition and subtraction are the same: $a = -a$, and $a+b = a-b$. And if $a+b=c$ then $a+c=b$ and $b+c=a$.

We add words in V_n coordinate-wise mod 2 like vectors:

$$\begin{array}{rl} \text{if} & \utilde{x} = 0\,1\,1\,0\,0\,1 \\ \text{and} & \utilde{y} = \underline{1\,1\,0\,1\,0\,1} \\ \text{then } \utilde{x}+\utilde{y} = \utilde{x}-\utilde{y} = & 1\,0\,1\,1\,0\,0 \end{array}$$

Note that $w(\utilde{x}-\utilde{y}) = d(\utilde{x},\utilde{y})$ ($=3$ in this case).

If $\utilde{x} \in V_n$, define $0\utilde{x} := \utilde{0}$ and $1\utilde{x} := \utilde{x}$. With these definitions, V_n becomes a vector space of dimension n over GF(2). Note that if $X \subseteq V_n$ then X is a subspace of V_n if and only if X is nonempty and closed under addition: $\utilde{a}, \utilde{b} \in X \implies \utilde{a}+\utilde{b} \in X$; for then $\utilde{a} \in X \implies \utilde{a}+\utilde{a} = \utilde{0} \in X \implies 0\utilde{a} = \utilde{0} \in X$ and $1\utilde{a} = \utilde{a} \in X$, so X is closed under scalar multiplication as well.

If $\utilde{x} = x_1x_2\ldots x_n$ and $\utilde{y} = y_1y_2\ldots y_n \in V_n$, let

$$\utilde{x}\cdot\utilde{y} := x_1y_1 + x_2y_2 + \ldots + x_ny_n,$$

their dot product mod 2. For example,

$$0\,1\,1\,0\,1\,0\,1 \cdot 1\,1\,0\,1\,1\,0\,1 = 0+1+0+0+1+0+1 = 1.$$

Then $\utilde{x}\cdot\utilde{y} = \utilde{y}\cdot\utilde{x}$, $\utilde{x}\cdot(\utilde{y}_1+\utilde{y}_2) = \utilde{x}\cdot\utilde{y}_1 + \utilde{x}\cdot\utilde{y}_2$ and $(\utilde{x}_1+\utilde{x}_2)\cdot\utilde{y} = \utilde{x}_1\cdot\utilde{y} + \utilde{x}_2\cdot\utilde{y}$. If $X \subseteq V_n$, its *orthogonal complement* is

$$X^\perp := \{\utilde{y} \in V_n : \utilde{x}\cdot\utilde{y} = 0 \text{ for every } \utilde{x} \text{ in } X\}.$$

Theorem 3.4. If X is a subspace of V_n with dimension k, then $X^\perp$ is a subspace with dimension $n-k$.

Proof. Clearly $\utilde{x}\cdot\utilde{0} = 0$ for all $\utilde{x}$, so $\utilde{0} \in X^\perp$, and

$$\begin{aligned} \utilde{y}_1, \utilde{y}_2 \in X^\perp &\implies \utilde{x}\cdot\utilde{y}_1 = \utilde{x}\cdot\utilde{y}_2 = 0 \qquad \forall\, \utilde{x} \in X \\ &\implies \utilde{x}\cdot(\utilde{y}_1+\utilde{y}_2) = 0+0 = 0 \qquad \forall\, \utilde{x} \in X \\ &\implies \utilde{y}_1+\utilde{y}_2 \in X^\perp. \end{aligned}$$

So $X^\perp$ is a subspace of V_n. Now let

$$\begin{aligned} &\utilde{x}_1 = x_{11}x_{12}\ldots x_{1n} \\ &\ \vdots \\ &\utilde{x}_k = x_{k1}x_{k2}\ldots x_{kn} \end{aligned}$$

form a basis for X, and note that $\utilde{y} = y_1y_2\ldots y_n \in X^\perp$ if and only if

$$\begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ \vdots & \vdots & & \vdots \\ x_{k1} & x_{k2} & \ldots & x_{kn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}.$$

The coefficient matrix here has rank k, and so $X^\perp$, the space of solutions, has dimension $n-k$, as required. //

Corollary 3.4.1. $(X^\perp)^\perp = X$.

Proof. Since $\utilde{x}\cdot\utilde{y} = 0$ for each $\utilde{x}$ in X and $\utilde{y}$ in $X^\perp$, clearly $X \subseteq X^{\perp\perp}$. But, by Theorem 3.4,

$$\dim X^{\perp\perp} = n - \dim X^\perp = n - (n-k) = k = \dim X,$$

and so $X^{\perp\perp} = X$. //

3.3. Linear codes. A *linear code* or *group code* (of dimension k and length n) is a code C that is a vector space (a k-dimensional subspace of V_n). If $|C| = m$ then $m = 2^k$ and the rate $r(C) = \frac{\log_2 m}{n} = \frac{k}{n}$. If C has minimum distance d then it is an $[n,k]$*-code* or $[n,k,d]$*-code*; thus an $[n,k,d]$-code is simply a linear $(n, 2^k, d)$-code.

Almost all codes used in practice are linear, because:

(1) encoding and decoding are easy (by matrix multiplication);
(2) error-correction *may* be easy, and is certainly *relatively* straightforward when $n-k$ is small;
(3) the error-correcting ability of linear codes is easier to determine.

Associated with every linear code C are four types of matrix: generator, parity-check, encoding and decoding matrices. The first two depend only on C; the last two depend also on the encoding map. The encoding matrix is uniquely determined by the encoding map; the other matrices are not unique. However, by permuting bits positionwise in the codewords if necessary, we can choose a generator matrix in 'standard form' that can be used as an encoding matrix and that uniquely determines standard forms for the other two matrices by a simple construction.

A matrix G is a *generator matrix* for C if C is the *row space* of G; i.e., $\utilde{x} \in C \iff \utilde{x}$ is a linear combination of rows of G. A matrix H is a *parity-check matrix* for C if $C = \text{Ker}\ H$, the *kernel* of H; i.e., $\utilde{x} \in C \iff H\utilde{x}^\top = \utilde{0}^\top \iff \utilde{x} \in D^\perp$, where D is the row space of H. Clearly rank $G = \dim C = k$ and rank $H = \dim D = n-k$ by Theorem 3.4, since $C = D^\perp$ and hence

(by Corollary 3.4.1) $D = D^{\perp\perp} = C^\perp$. $C^\perp$ is the *dual code* of C: H is a generator matrix for $C^\perp$ and G is a parity-check matrix for it.

Example.

$C = D^\perp$, $\dim k$:

$$k \begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix} = G \qquad (n \text{ columns})$$

$$\begin{array}{ccccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \end{array}$$

$D = C^\perp$, $\dim n-k$:

$$n-k \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix} = H \qquad (n \text{ columns})$$

Two linear codes are *equivalent* if one can be obtained by applying the same positionwise permutation of bits to all the codewords in the other. Let I_j denote the $j \times j$ identity matrix.

Theorem 3.5. By replacing C by an equivalent code if necessary, we can choose

$$G = [\ I_k \mid A\ ] \quad \text{and} \quad H = [\ -A^\top \mid I_{n-k}\ ]$$

for some $k \times (n-k)$ matrix A. G is then said to be in *standard form.*

Proof. Let G be a $k \times n$ generator matrix for C. Note that elementary row operations on G do not change C, while permuting the columns corresponds to permuting the bits so as to give a code equivalent to C. Also, no row of G is all zero. So permute the columns to ensure $g_{11} = 1$, then add row 1 to other rows as necessary to clear the first column apart from g_{11}. Now permute columns $2, \ldots, n$ to ensure $g_{22} = 1$, then operate with row 2 to clear the second column apart from g_{22}. And so on. (In practice it is preferable to permute rows rather than columns, since this does not change C; but it will not always work.)

Given G in standard form, form H as above. It clearly has rank $n-k$, and so it suffices to show that every row of G is orthogonal to every row of H. But the dot product of the ith row of G with the jth row of H is

$$0+0+\ldots+0+(-a_{ij})+0+\ldots+0+a_{ij}+0+\ldots+0=0. \quad //$$

Example. Let $C_0=\{0\,0\,0\,0\,0,\ 0\,0\,1\,1\,1,\ 1\,1\,1\,0\,0,\ 1\,1\,0\,1\,1\}$, which is a single-error-correcting code with generator matrix (say)

$$G_0=\begin{bmatrix}0&0&1&1&1\\1&1&1&0&0\end{bmatrix}.$$

By permuting just the rows, we can only get the first column right:

$$\begin{bmatrix}1&1&1&0&0\\0&0&1&1&1\end{bmatrix}.$$

Now interchange columns 2 and 3 to get $\begin{bmatrix}1&1&1&0&0\\0&1&0&1&1\end{bmatrix}$ and add row 2 to row 1 to get $\begin{bmatrix}1&0&1&1&1\\0&1&0&1&1\end{bmatrix}$. Thus we have a new code

$$C=\{0\,0\,0\,0\,0,\ \ 0\,1\,0\,1\,1,\ \ 1\,1\,1\,0\,0,\ \ 1\,0\,1\,1\,1\},$$

equivalent to C_0, with generator and parity-check matrices

$$G=\left[\begin{array}{cc|ccc}1&0&1&1&1\\0&1&0&1&1\end{array}\right] \quad\text{and}\quad H=\left[\begin{array}{cc|ccc}1&0&1&0&0\\1&1&0&1&0\\1&1&0&0&1\end{array}\right].$$

We can use G as the encoding matrix (on the right): the map

$$(x,y)\mapsto(x,y)G=(x,y)\begin{bmatrix}1&0&1&1&1\\0&1&0&1&1\end{bmatrix}=(x,y,x,x+y,x+y)$$

describes the encoding map

$$\begin{array}{l}0\,0\ \to\ 0\,0\,0\,0\,0\\0\,1\ \to\ 0\,1\,0\,1\,1\\1\,0\ \to\ 1\,0\,1\,1\,1\\1\,1\ \to\ 1\,1\,1\,0\,0\end{array}$$

Since, with this encoding map, the source words appear at the start of the corresponding codewords, we can use the matrix $\left[\dfrac{I_k}{Z}\right]$ as the decoding matrix:

$$(x,y,x,x+y,x+y)\begin{bmatrix}1&0\\0&1\\0&0\\0&0\\0&0\end{bmatrix}=(x,y).$$

The *error-pattern vector* of a received word $\underset{\sim}{x}_{\mathrm{R}}$ is the vector (word) with 1's where errors have been made. The *corrector* $\underset{\sim}{x}_{\mathrm{C}}$ is what we add to $\underset{\sim}{x}_{\mathrm{R}}$ to get a codeword — hopefully the right one. For example, suppose that in the above code C we receive $\underset{\sim}{x}_{\mathrm{R}}=1\,1\,0\,0\,0$. The possible codewords and error-pattern vectors are:

codeword	error-pattern vector
0 0 0 0 0	1 1 0 0 0
0 1 0 1 1	1 0 0 1 1
1 1 1 0 0	0 0 1 0 0
1 0 1 1 1	0 1 1 1 1

The error-pattern vector of minimum weight is $0\,0\,1\,0\,0$, so we adopt this as $\underset{\sim}{x}_{\mathrm{C}}$ and add it to $\underset{\sim}{x}_{\mathrm{R}}$ to get $1\,1\,1\,0\,0$, which is the codeword closest to $\underset{\sim}{x}_{\mathrm{R}}$ and hopefully the one that was sent. The set of possible error-pattern vectors is the coset $\underset{\sim}{x}_{\mathrm{R}}+C=\{\underset{\sim}{x}_{\mathrm{R}}+\underset{\sim}{c}:\underset{\sim}{c}\in C\}$, and $\underset{\sim}{x}_{\mathrm{C}}$ is any word of minimum weight in this coset, called the *coset leader*. We can locate the coset by calculating $\underset{\sim}{x}_{\mathrm{R}}H^{\top}$, the (*error*) *syndrome* of $\underset{\sim}{x}_{\mathrm{R}}$, since

$$\begin{aligned}\underset{\sim}{x}H^{\top}=\underset{\sim}{y}H^{\top}\iff H\underset{\sim}{x}^{\top}=H\underset{\sim}{y}^{\top}&\iff H(\underset{\sim}{x}-\underset{\sim}{y})^{\top}=\underset{\sim}{0}^{\top}\\&\iff \underset{\sim}{x}-\underset{\sim}{y}\in C\\&\iff \underset{\sim}{x}\text{ and }\underset{\sim}{y}\text{ are in the same coset.}\end{aligned}$$

For the above code C with $H=\begin{bmatrix}1&0&1&0&0\\1&1&0&1&0\\1&1&0&0&1\end{bmatrix}$ we have the table:

coset	syndrome $\underset{\sim}{x}_{\mathrm{R}}H^{\top}$	coset leader $\underset{\sim}{x}_{\mathrm{C}}$
0 0 0 0 0, 0 1 0 1 1, 1 1 1 0 0, 1 0 1 1 1	0 0 0	0 0 0 0 0
0 0 0 0 1, 0 1 0 1 0, 1 1 1 0 1, 1 0 1 1 0	0 0 1	0 0 0 0 1
0 0 0 1 0, 0 1 0 0 1, 1 1 1 1 0, 1 0 1 0 1	0 1 0	0 0 0 1 0
0 1 0 0 0, 0 0 0 1 1, 1 0 1 0 0, 1 1 1 1 1	0 1 1	0 1 0 0 0
0 0 1 0 0, 0 1 1 1 1, 1 1 0 0 0, 1 0 0 1 1	1 0 0	0 0 1 0 0
0 0 1 0 1, 0 1 1 1 0, 1 1 0 0 1, 1 0 0 1 0	1 0 1	0 0 1 0 1 (say)
0 0 1 1 0, 0 1 1 0 1, 1 1 0 1 0, 1 0 0 0 1	1 1 0	1 0 0 0 1 (say)
1 0 0 0 0, 1 1 0 1 1, 0 1 1 0 0, 0 0 1 1 1	1 1 1	1 0 0 0 0

Note that we do not need to store the cosets. We need only store the table of 2^{n-k} coset leaders indexed by the syndromes, known as the *syndrome look-up table*. The error-correcting routine is then: for each received word $\underset{\sim}{x}_{\mathrm{R}}$, calculate the syndrome $\underset{\sim}{x}_{\mathrm{R}}H^{\top}$, use it to read off the corrector $\underset{\sim}{x}_{\mathrm{C}}$ from the table, and return the codeword $\underset{\sim}{x}_{\mathrm{R}}+\underset{\sim}{x}_{\mathrm{C}}$. (*And then decode it!*)

For a linear code we have an exact expression for P_{err}:

Theorem 3.6. Let the number of coset leaders of weight i be a_i. Then $P_{\mathrm{err}}=1-\sum_{i=0}^{n}a_i\,p^i(1-p)^{n-i}$.

Proof. A received word is corrected successfully iff the error-pattern vector is one of the coset leaders. The probability that it is a *particular* coset leader of weight i is $p^i(1-p)^{n-i}$, whence the result. //

Note that, if C is e-error-correcting, then $a_i=\binom{n}{i}$ for $0\leqslant i\leqslant e$, since all words with weight $\leqslant e$ are coset leaders, and so $P_{\mathrm{err}}\leqslant 1-\sum_{i=0}^{e}\binom{n}{i}p^i(1-p)^{n-i}$. [This can be seen directly.]

Example. For the code above, $n=5$, $a_0=1$, $a_1=5$ and $a_2=2$, and so

$$\begin{aligned}P_{\mathrm{err}}&=1-[(1-p)^5+5p(1-p)^4+2p^2(1-p)^3]\\&=1-[\ 1-5p+10p^2-10p^3+\ 5p^4-\ p^5\\&\qquad\quad +5p-20p^2+30p^3-20p^4+5p^5\\&\qquad\qquad\quad +\ 2p^2-\ 6p^3+\ 6p^4-2p^5\ ]\\&=8p^2-14p^3+9p^4-2p^5.\end{aligned}$$

We also have a criterion for error-correcting ability:

Theorem 3.7. Let C be a linear code with parity-check matrix H. Then the following three statements are equivalent.

(a) C is e-error-correcting.
(b) Each nonzero word in C has weight at least $2e+1$.
(c) Each set of $2e$ or fewer columns of H is linearly independent.

Proof. Note that $\{\underset{\sim}{x}-\underset{\sim}{y}:\underset{\sim}{x},\underset{\sim}{y}\in C\}=\{\underset{\sim}{x}:\underset{\sim}{x}\in C\}=C$. Also, a set $\underset{\sim}{c}_1,\ldots,\underset{\sim}{c}_r$ of columns of H is linearly dependent iff there exist $a_1,\ldots,a_r$, each of which is 0 or 1 and not all of which are 0, such that $a_1\underset{\sim}{c}_1+\ldots+a_r\underset{\sim}{c}_r=\underset{\sim}{0}$, which is the same as saying that some subset of $\underset{\sim}{c}_1,\ldots,\underset{\sim}{c}_r$ add up to $\underset{\sim}{0}$. Thus

$$\begin{aligned}\text{(a)}&\iff d(\underset{\sim}{x},\underset{\sim}{y})\geqslant 2e+1 \quad \forall\underset{\sim}{x},\underset{\sim}{y}\in C\quad(\underset{\sim}{x}\neq\underset{\sim}{y})\\&\iff w(\underset{\sim}{x}-\underset{\sim}{y})\geqslant 2e+1\quad\forall\underset{\sim}{x},\underset{\sim}{y}\in C\quad(\underset{\sim}{x}\neq\underset{\sim}{y})\\&\iff w(\underset{\sim}{x})\geqslant 2e+1\quad\forall\underset{\sim}{x}\in C\quad(\underset{\sim}{x}\neq\underset{\sim}{0})\qquad\iff\text{(b)}\\&\iff\text{no sum of }2e\text{ or fewer columns of }H\text{ is equal to }\underset{\sim}{0}\\&\iff\text{no set of }2e\text{ or fewer columns of }H\text{ is linearly dependent}\\&\qquad\qquad\qquad\qquad\qquad\qquad\iff\text{(c)}.\quad //\end{aligned}$$

3.4. Approaches to perfection; the Hamming codes.

Theorem 3.8. (a) *Hamming's sphere-packing bound* (R. W. Hamming, 1950). If $C\subseteq V_n$ is an e-error-correcting code with $|C|=2^k$ then

$$2^{n-k} \geqslant \sum_{i=0}^{e} \binom{n}{i} \qquad \left(\text{i.e., } |C| = 2^k \leqslant 2^n \Big/ \sum_{i=0}^{e} \binom{n}{i}\right).$$

(b) (E. N. Gilbert, 1952; R. R. Varsharmov, 1957.) There exists a linear e-error-correcting code $C \subseteq V_n$ of dimension k if

$$2^{n-k} > \sum_{i=0}^{2e-1} \binom{n-1}{i} \qquad \left(\text{i.e., } |C| = 2^k < 2^n \Big/ \sum_{i=0}^{2e-1} \binom{n-1}{i}\right).$$

Proof. (a) Let $\Sigma := \sum_{i=0}^{e} \binom{n}{i} = 1 + n + \binom{n}{2} + \ldots + \binom{n}{e}$. If $\underset{\sim}{x} \in V_n$, let $S(\underset{\sim}{x}, e) := \{\underset{\sim}{y} \in V_n : d(\underset{\sim}{x}, \underset{\sim}{y}) \leqslant e\}$, the *sphere* (strictly, ball) centre $\underset{\sim}{x}$ radius e. Then $|S(\underset{\sim}{x}, e)| = \Sigma$. But if C is e-error-correcting, then the spheres $S(\underset{\sim}{x}, e)$ for all $\underset{\sim}{x} \in C$ must be disjoint. Therefore $2^n = |V_n| \geqslant |C|\Sigma = 2^k\Sigma$, and so $2^{n-k} \geqslant \Sigma$.

(b) [If $2^{n-k} \geqslant \sum_{i=0}^{2e} \binom{n}{i}$ then we can construct an e-error-correcting code with 2^k codewords by choosing codewords one at a time so that each new codeword is never within distance $2e$ of any codeword already chosen. However, the resulting code might not be linear, and in any case we have needed a stronger condition than in (b).]

We shall construct an $(n-k) \times n$ matrix H with columns $\underset{\sim}{c}_1, \ldots, \underset{\sim}{c}_n$ in which every set of $2e$ or fewer columns is linearly independent, and the result will then follow from Theorem 3.7. Let $\underset{\sim}{c}_1$ be any nonzero word in V_{n-k}. Given $\underset{\sim}{c}_1, \ldots, \underset{\sim}{c}_r$ $(r < n)$, let $\underset{\sim}{c}_{r+1}$ be any word in V_{n-k} that is not linearly dependent on any $2e-1$ or fewer of $\underset{\sim}{c}_1, \ldots, \underset{\sim}{c}_r$. This is the same as saying that $\underset{\sim}{c}_{r+1}$ must not equal the sum of any $0, 1, 2, \ldots, 2e-1$ of $\underset{\sim}{c}_1, \ldots, \underset{\sim}{c}_r$, and so the number of words we can choose for $\underset{\sim}{c}_{r+1}$ is at least

$$|V_{n-k}| - 1 - r - \binom{r}{2} - \ldots - \binom{r}{2e-1}.$$

Thus provided $r \leqslant n-1$ and

$$|V_{n-k}| = 2^{n-k} > \sum_{i=0}^{2e-1} \binom{n-1}{i} \geqslant \sum_{i=0}^{2e-1} \binom{r}{i},$$

we can choose $\underset{\sim}{c}_{r+1}$ as required. //

Remarks. (1) The above construction proves that such codes exist, but it is not useful in practice.
(2) It would be desirable to close the gap between (a) and (b). Some progress has been made with this.

A code is *perfect* if equality holds in the sphere-packing bound. This means that the spheres $S(\underset{\sim}{x}, e)$ $(\underset{\sim}{x} \in C)$ exactly cover V_n. The following linear codes are perfect.

(0) Any code with just one codeword.
(1) The *binary repetition code of length* n, for odd n. (Every word in V_n lies within distance $\frac{1}{2}(n-1)$ of exactly one of the two codewords $0\,0\ldots0$ and $1\,1\ldots1$.) These are $[n, 1, n]$-codes, usually called *trivial*.
(2) The *Hamming single-error-correcting codes* (see below), which are $[2^r - 1, 2^r - 1 - r, 3]$-codes that are much used in practice.
(3) A 3-error-correcting $[23, 12, 7]$-code, the *binary Golay code*, found by M. J. E. Golay in 1949.

There are also some nonlinear perfect binary codes, but A. Tietäväinen proved in 1973 that all perfect binary codes have the same parameters as these linear ones.

For $r \geqslant 2$, let H_r be the $r \times (2^r - 1)$ matrix whose columns are all the nonzero words in V_r, arranged so that the ith column is i in binary. The *binary Hamming code* $\mathrm{Ham}(r)$ is the code of length $2^r - 1$ with H_r as its parity-check matrix (not in standard form). For example,

$H_2 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$, and $\mathrm{Ham}(2) = \{0\,0\,0,\ 1\,1\,1\}$, the binary repetition code of length 3.

$$H_3 = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

If a single error occurs, $\underset{\sim}{x}_{\mathrm{R}} H_r^{\top}$ gives the *position of the error in binary*. E.g., suppose with $\mathrm{Ham}(3)$ $1\,1\,0\,0\,1\,1\,0$ is sent.

Received	Syndrome	Coset leader	Diagnosis
1 1 0 0 1 1 0	$0\,0\,0 = 0$	0 0 0 0 0 0 0	No error
1 1 0 0 0 1 0	$1\,0\,1 = 5$	0 0 0 0 1 0 0	Error in 5th position
1 1 0 0 0 0 0	$0\,1\,1 = 3$	0 0 1 0 0 0 0	Error in 3rd position

There are 2^r possible syndromes $\underset{\sim}{x}_{\mathrm{R}} H_r^{\top}$, each of which is the binary representation of some number j $(0 \leqslant j \leqslant 2^r - 1)$. So $\underset{\sim}{x}_{\mathrm{R}} H_r^{\top}$ is either $\underset{\sim}{0}$ or the jth column of H_r, and $\underset{\sim}{x}_{\mathrm{R}}$ is in the same coset as either $0\,0\ldots0$ (if $j = 0$) or $0\ldots0\,1\,0\ldots0$ (if $j \geqslant 1$), where the 1 occurs in the jth position. Thus every coset contains exactly one word of weight $\leqslant 1$. Equivalently, every word is within distance 1 of exactly one codeword. Thus $\mathrm{Ham}(r)$ is a *perfect single-error-correcting code*. H_r is of size $r \times (2^r - 1)$, so $\mathrm{Ham}(r)$ has length $n = 2^r - 1$ and dimension $k = n - \operatorname{rank} H_r = n - r = 2^r - r - 1$, whence it has parameters $[2^r - 1, 2^r - r - 1, 3]$. Its rate is $\frac{k}{n} = \frac{2^r - r - 1}{2^r - 1} \to 1$ as $r \to \infty$.

Alternatively:

$\exists$ a codeword of weight 1 $\iff$ H has a zero column.

$\exists$ a codeword of weight 2 $\iff$ H has two equal columns.

H_r has distinct nonzero columns, so (as in Theorem 3.7) $\mathrm{Ham}(r)$ has minimum distance $\geqslant 3$ (in fact, exactly 3, since $1\,1\,1\,0\ldots0 \in \mathrm{Ham}(r)$), and so is single-error-correcting. Since

$$2^{n-k} = 2^r = 1 + n = \sum_{i=0}^{1} \binom{n}{i},$$

$\mathrm{Ham}(r)$ is perfect. [Use in computer memory.]

3.5. The first-order Reed–Muller codes $R(1,m)$ $(m \geqslant 1)$.

Two important classes of code are Hadamard codes and Reed–Muller codes. The codes that belong to both classes simultaneously are the codes $R(1,m)$. $R(1,5)$ was used by NASA from 1969 to 1976, in their Mariner and Viking spacecraft. $R(1,m)$ has generator matrix G_m, which is the $(m+1) \times 2^m$ matrix whose ith column is 1 followed by the binary representation of $i-1$:

$$G_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \qquad G_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \qquad G_3 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

$R(1,m)$ has length 2^m and dimension $m+1$. Every codeword apart from $0\,0\ldots0$ and $1\,1\ldots1$ has weight exactly 2^{m-1}. [In fact, any sum of rows of G_m including row i from the bottom $(1 \leqslant i \leqslant m)$ has exactly half 0's and half 1's in any 2^i consecutive positions ending on a multiple of 2^i.] Thus $R(1,m)$ is a $(2^m, 2^{m+1}, 2^{m-1})$-code which is $(2^{m-2} - 1)$-error-correcting. Its rate $\frac{k}{n}$ is $\frac{m+1}{2^m}$. The Mariner code $R(1,5)$ has parameters $(32, 64, 16)$ and is 7-error-correcting with rate $\frac{6}{32} = 0.1875$. A syndrome look-up table for this code would have $2^{n-k} = 2^{32-6} = 2^{26}$ entries! [$\approx 67\,000\,000$] [Fast Fourier transform.]

If G_m is used as the encoding matrix, then error-correction and decoding can be carried out in a single step which we illustrate with G_3. Suppose $\underset{\sim}{y} = y_1 y_2 y_3 y_4$ is the sourceword and $\underset{\sim}{x} = x_1 \ldots x_8$ is the codeword. Then (mod 2)

$$\begin{aligned} y_4 &= x_1 + x_2 = x_3 + x_4 = x_5 + x_6\ [= x_7 + x_8], \\ y_3 &= x_1 + x_3 = x_2 + x_4 = x_5 + x_7\ [= x_6 + x_8], \\ y_2 &= x_1 + x_5 = x_2 + x_6 = x_3 + x_7\ [= x_4 + x_8]. \end{aligned}$$

On receiving $\underset{\sim}{x}_{\mathrm{R}} = x_1' \ldots x_8'$, evaluate $x_1' + x_2'$, $x_3' + x_4'$, $x_5' + x_6'$, and obtain y_4 by 'majority rule'. Obtain y_3 and y_2 similarly, and finally choose $y_1 = 1$ or 0 according as $\underset{\sim}{x}_{\mathrm{R}} + y_2 \underset{\sim}{r}_2 + y_3 \underset{\sim}{r}_3 + y_4 \underset{\sim}{r}_4$ has more 1's than 0's or fewer, where $\underset{\sim}{r}_1, \ldots, \underset{\sim}{r}_4$ denote the rows of G_3.

This decoding scheme is fast, requires little storage, and corrects up to $2^{m-2}-1$ errors, since each y_i $(2 \leqslant i \leqslant m+1)$ is determined by $2^{m-1}-1$ 'votes', and so at least 2^{m-2} errors must be made in order to get it wrong; and if $y_2, \ldots, y_{m+1}$ are right then at least 2^{m-1} errors must be made to get y_1 wrong. However, it is *not* a minimum-distance decoding scheme. For example, in $R(1,4)$ (which is 3-error-correcting), suppose $\underset{\sim}{x}_{\mathrm{R}} = 1\,0\,0\,0\,1\,0\,0\,0\,1\,0\,0\,0\,1\,1\,1\,1$. Then $y_2 = y_3 = y_4 = y_5 = 0$ (all by 4 0's to 3 1's or 5 0's to 2 1's), and so $y_1 = 0$. Thus we are interpreting $\underset{\sim}{x}_{\mathrm{R}}$ as $\underset{\sim}{0}$. But $d(\underset{\sim}{x}_{\mathrm{R}}, \underset{\sim}{0}) = 7$, whereas $d(\underset{\sim}{x}_{\mathrm{R}}, \underset{\sim}{r}_2) = 5$.

3.6. BCH codes. A code C of length n is *cyclic* if

$$c_1 c_2 \ldots c_{n-1} c_n \in C \Longrightarrow c_n c_1 c_2 \ldots c_{n-1} \in C.$$

Theorem 3.9. If T is an $m \times m$ matrix over GF(2) and $\underset{\sim}{x} \in V_m$ and $T^n \underset{\sim}{x}^\top = \underset{\sim}{x}^\top$, and if H is the $m \times n$ matrix

$$H := [\, \underset{\sim}{x}^\top \;\; T\underset{\sim}{x}^\top \;\; T^2\underset{\sim}{x}^\top \;\; \ldots \;\; T^{n-1}\underset{\sim}{x}^\top \,],$$

then Ker H is a cyclic code of length n.

Proof. Recall Ker $H = \{\underset{\sim}{z} \in V_n : H\underset{\sim}{z}^\top = \underset{\sim}{0}^\top\}$. Clearly Ker H is a code of length n. Now,

$$\begin{aligned}
c_1 c_2 \ldots c_{n-1} c_n \in \text{Ker } H &\iff H \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_{n-1} \\ c_n \end{bmatrix} = \underset{\sim}{0}^\top \\
&\iff c_1 \underset{\sim}{x}^\top + c_2 T \underset{\sim}{x}^\top + \ldots + c_{n-1} T^{n-2} \underset{\sim}{x}^\top + c_n T^{n-1} \underset{\sim}{x}^\top = \underset{\sim}{0}^\top \\
&\Longrightarrow H \begin{bmatrix} c_n \\ c_1 \\ c_2 \\ \vdots \\ c_{n-1} \end{bmatrix} = c_n \underset{\sim}{x}^\top + c_1 T \underset{\sim}{x}^\top + c_2 T^2 \underset{\sim}{x}^\top + \ldots + c_{n-1} T^{n-1} \underset{\sim}{x}^\top \\
&\qquad = T(c_1 \underset{\sim}{x}^\top + c_2 T \underset{\sim}{x}^\top + \ldots + c_{n-1} T^{n-2} \underset{\sim}{x}^\top + c_n T^{n-1} \underset{\sim}{x}^\top) \\
&\qquad = T\underset{\sim}{0}^\top = \underset{\sim}{0}^\top \qquad \text{since } T^n \underset{\sim}{x}^\top = \underset{\sim}{x}^\top \\
&\iff c_n c_1 c_2 \ldots c_{n-1} \in \text{Ker } H.
\end{aligned}$$

So Ker H is cyclic. //

Example. With $m = 4$, let $T = \begin{bmatrix} 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \\ 1&0&1&0 \end{bmatrix}$ and $\underset{\sim}{x}^\top = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$.

Then $H = \begin{bmatrix} 0&0&0&1&0&1 \\ 0&0&1&0&1&0 \\ 0&1&0&1&0&0 \\ 1&0&1&0&0&0 \end{bmatrix}$. $\left(T^6 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}\right.$, so we can take $n = 6$.)

Ker $H = \{0\,0\,0\,0\,0\,0,\; 0\,1\,0\,1\,0\,1,\; 1\,0\,1\,0\,1\,0,\; 1\,1\,1\,1\,1\,1\}$.

Lemma 3.10.1. Let T be an $m \times m$ matrix over GF(2) of the form

$$T = \begin{bmatrix} 0&1&0&0&\ldots&0 \\ 0&0&1&0&\ldots&0 \\ 0&0&0&1&\ldots&0 \\ &&&&\ddots& \\ 0&0&0&0&\ldots&1 \\ a_0&a_1&a_2&a_3&\ldots&a_{m-1} \end{bmatrix},$$

and let $\phi_T(x) := x^m + a_{m-1}x^{m-1} + \ldots + a_1 x + a_0$. Then the characteristic and minimal polynomials of T are both equal to $\phi_T(x)$.

Proof. The characteristic polynomial is $\det(T - xI) =$

$$\begin{vmatrix} -x&1&0&\ldots&0&0&0 \\ 0&-x&1&\ldots&0&0&0 \\ 0&0&-x&\ddots&&& \\ &&&\ddots&1&0&0 \\ 0&0&0&\ldots&-x&1&0 \\ 0&0&0&\ldots&0&-x&1 \\ a_0&a_1&a_2&\ldots&a_{m-3}&a_{m-2}&a_{m-1}-x \end{vmatrix} =$$

$$= \begin{vmatrix} 0&1&0&&0&0&0 \\ 0&0&1&&0&0&0 \\ 0&0&0&\ddots&&& \\ &&&&1&0&0 \\ 0&0&0&&0&1&0 \\ 0&0&0&&0&0&1 \\ \phi_T(x)&?&?&&a_{m-3}+a_{m-2}x+a_{m-1}x^2-x^3&a_{m-2}+a_{m-1}x-x^2&a_{m-1}-x \end{vmatrix}$$

$= (\pm)\phi_T(x) \det I_{m-1} = \phi_T(x)$. Hence, by the Cayley–Hamilton theorem, $\phi_T(T) = Z$ (the zero matrix), where

$$\phi_T(T) = T^m + a_{m-1}T^{m-1} + \ldots + a_1 T + a_0 I.$$

However, the top rows of $I, T, T^2, \ldots, T^{m-1}$ are the rows of I, so these matrices are linearly independent. Thus there is no nonzero polynomial $\psi(x)$ of degree $m-1$ or less s.t. $\psi(T) = Z$, and so $\phi_T(x)$ is the minimal polynomial of T. //

The above matrix T has determinant a_0 and so is nonsingular iff $a_0 = 1$. If $a_0 = 1$ we call T a *T-matrix*.

Lemma 3.10.2. If T is an $m \times m$ T-matrix then $\exists\, n$ $(0 < n \leqslant 2^m - 1)$ s.t. $T^n = I$.

Proof. For each n, T^n can be written as a polynomial in T of degree $\leqslant m-1$ (using $\phi_T(T) = Z$). Since T is nonsingular, $T^n \neq Z$. There are 2^m-1 distinct nonzero polynomials over GF(2) of degree $\leqslant m-1$, so $\exists\, i, j$ $(0 \leqslant i < j \leqslant 2^m - 1)$ s.t. $T^i = T^j$. But then $T^{j-i} = I$ and $0 < j - i \leqslant 2^m - 1$. //

The smallest $n > 0$ s.t. $T^n = I$ is the *period* of T. T is *primitive* if its period is $2^m - 1$.

Example. Let $T = \begin{bmatrix} 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \\ 1&0&0&1 \end{bmatrix}$. Then $\phi_T(x) = x^4 + x^3 + 1$,

so $T^4 + T^3 + I = Z$, and $T^4 = T^3 + I$. Hence

$$\begin{array}{l|l}
T^5 = T^4 + T = T^3 + T + I & T^{11} = T^4 + T^2 = T^3 + T^2 + I \\
T^6 = T^4 + T^2 + T = T^3 + T^2 + T + I & T^{12} = T^4 + T^3 + T = T + I \\
T^7 = T^4 + T^3 + T^2 + T = T^2 + T + I & T^{13} = T^2 + T \\
T^8 = T^3 + T^2 + T & T^{14} = T^3 + T^2 \\
T^9 = T^4 + T^3 + T^2 = T^2 + I & T^{15} = T^4 + T^3 = I \\
T^{10} = T^3 + T &
\end{array}$$

So T has period $15 = 2^4 - 1$ and hence is primitive.

Lemma 3.10.3. If T is a primitive $m \times m$ T-matrix, $n = 2^m - 1$, and $F = \{Z, I, T, T^2, \ldots, T^{n-1}\}$, then F is a field under matrix addition and multiplication.

Proof. Let $F^* := F \setminus \{Z\}$. Since T is primitive, F^* is precisely the set of $n = 2^m - 1$ nonzero polynomials in T of degree $\leqslant m-1$, and F is the set of *all* polynomials in T of degree $\leqslant m-1$. Thus F is closed under addition. F is also closed under multiplication since $T^n = I$. It is clear that F forms an additive abelian group and F^* forms a multiplicative (cyclic) group, and the distributive laws hold, so F is a field. //

Lemma 3.10.4. If T is a primitive $m \times m$ T-matrix, $p(T)$ is a polynomial in T and $\exists\, \underset{\sim}{x} \in V_m$ $(\underset{\sim}{x} \neq \underset{\sim}{0})$ such that $p(T)\underset{\sim}{x}^\top = \underset{\sim}{0}^\top$, then $p(T) = Z$.

Proof. Since F is closed under addition and multiplication, $p(T) \in F$. Thus either $p(T) = Z$ or $p(T) = T^k$ for some k $(0 \leqslant k \leqslant n-1)$, in which case $p(T)$ is nonsingular. But $p(T)\underset{\sim}{x}^\top = \underset{\sim}{0}^\top$ and $\underset{\sim}{x} \neq \underset{\sim}{0}$, so $p(T)$ is singular, and so $p(T) = Z$. //

Theorem 3.10. Let T be a primitive $m \times m$ T-matrix, $\underset{\sim}{x} \in V_m$ $(\underset{\sim}{x} \neq \underset{\sim}{0})$, $n := 2^m - 1$, and

$$H_T(\underset{\sim}{x}) := \left[\, \underset{\sim}{x}^\top \;\; T\underset{\sim}{x}^\top \;\; T^2\underset{\sim}{x}^\top \;\; \ldots \;\; T^{n-1}\underset{\sim}{x}^\top \,\right].$$

Then Ker $H_T(\underset{\sim}{x})$ is a cyclic code equivalent to Ham(m).

Proof. Ker $H_T(\underset{\sim}{x})$ is a cyclic code by Theorem 3.9. Now, T is nonsingular and $\underset{\sim}{x} \neq \underset{\sim}{0}$, so $T^i\underset{\sim}{x}^\top \neq \underset{\sim}{0}^\top$ for each i. Also,

$$\begin{aligned}
T^i\underset{\sim}{x}^\top = T^j\underset{\sim}{x}^\top &\Longrightarrow (T^j - T^i)\underset{\sim}{x}^\top = \underset{\sim}{0}^\top \\
&\Longrightarrow T^j - T^i = Z \qquad \text{by Lemma 3.10.4} \\
&\Longrightarrow T^j = T^i \\
&\Longrightarrow T^{j-i} = I,
\end{aligned}$$

which is impossible if $0 \leqslant i < j \leqslant n-1$ since then $0 < j-i < n$ and T is primitive. Thus the columns of $H_T(\underset{\sim}{x})$ are precisely the $2^m - 1$ distinct nonzero elements of V_m, and so Ker $H_T(\underset{\sim}{x})$ is equivalent to Ham(m). //

Example. Let $m = 3$, $T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$, $\underset{\sim}{x}^\top = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.

Then $H_T(\underset{\sim}{x}) = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$.

Theorem 3.11. (R. C. Bose and D. K. Ray-Chaudhuri, 1960; A. Hocquenghem, 1959.) Let T, $\underset{\sim}{x}$, m and n be as in Theorem 3.10, let $e \geqslant 1$, and let H be the $me \times (2^m - 1) = me \times n$ matrix

$$H = \begin{array}{l} \left[\begin{array}{ccccc} \underset{\sim}{x}^\top & T\underset{\sim}{x}^\top & T^2\underset{\sim}{x}^\top & \ldots & T^{n-1}\underset{\sim}{x}^\top \\ \underset{\sim}{x}^\top & T^3\underset{\sim}{x}^\top & T^6\underset{\sim}{x}^\top & \ldots & T^{3(n-1)}\underset{\sim}{x}^\top \\ \underset{\sim}{x}^\top & T^5\underset{\sim}{x}^\top & T^{10}\underset{\sim}{x}^\top & \ldots & T^{5(n-1)}\underset{\sim}{x}^\top \\ \vdots & \vdots & \vdots & & \vdots \\ \underset{\sim}{x}^\top & T^{2e-1}\underset{\sim}{x}^\top & T^{(2e-1)2}\underset{\sim}{x}^\top & \ldots & T^{(2e-1)(n-1)}\underset{\sim}{x}^\top \end{array}\right]. \end{array} \begin{array}{l} \leftarrow H_T(\underset{\sim}{x}) \\ \leftarrow H_{T^3}(\underset{\sim}{x}) \\ \leftarrow H_{T^5}(\underset{\sim}{x}) \\ \\ \leftarrow H_{T^{2e-1}}(\underset{\sim}{x}) \end{array}$$

Then Ker H is an e-error-correcting cyclic code. [These are the *BCH-codes*, with rate $\frac{n-\text{rank}\,H}{n} \geqslant \frac{n-me}{n} = 1 - \frac{me}{2^m-1}$.]

Proof. Since $T^n = I$, $(T^i)^n = I$ for $i = 1, 3, 5, \ldots, 2e-1$. Now, Ker H = Ker $H_T(\underset{\sim}{x}) \cap$ Ker $H_{T^3}(\underset{\sim}{x}) \cap \ldots \cap$ Ker $H_{T^{2e-1}}(\underset{\sim}{x})$, and since each of these is cyclic by Theorem 3.9 it follows that Ker H is cyclic.

By Theorem 3.7, to prove that Ker H is e-error-correcting, it suffices to show that no sum of $2e$ or fewer columns of H is zero. So suppose $\exists\, k_1, \ldots, k_r$ $(0 \leqslant k_1 < \ldots < k_r \leqslant n-1,\ r \leqslant 2e)$ s.t.

$$\begin{array}{ccccccc} T^{k_1}\underset{\sim}{x}^\top & + & T^{k_2}\underset{\sim}{x}^\top & + \ldots + & T^{k_r}\underset{\sim}{x}^\top & = & \underset{\sim}{0}^\top, \\ T^{3k_1}\underset{\sim}{x}^\top & + & T^{3k_2}\underset{\sim}{x}^\top & + \ldots + & T^{3k_r}\underset{\sim}{x}^\top & = & \underset{\sim}{0}^\top, \\ \vdots & & & & & & \\ T^{(2e-1)k_1}\underset{\sim}{x}^\top & + & T^{(2e-1)k_2}\underset{\sim}{x}^\top & + \ldots + & T^{(2e-1)k_r}\underset{\sim}{x}^\top & = & \underset{\sim}{0}^\top. \end{array}$$

Let $S_i := T^{k_i}$ $(1 \leqslant i \leqslant r)$. By Lemma 3.10.4,

$$\begin{array}{ccccccc} S_1 & + & S_2 & + \ldots + & S_r & = Z, \\ S_1^3 & + & S_2^3 & + \ldots + & S_r^3 & = Z, \\ \vdots & & & & & \\ S_1^{2e-1} & + & S_2^{2e-1} & + \ldots + & S_r^{2e-1} & = Z. \end{array}$$

Also $S_1^2 + S_2^2 + \ldots + S_r^2 = (S_1 + S_2 + \ldots + S_r)^2 = Z$ (etc.). But each S_i is an element of a *field* F by Lemma 3.10.3. So

$$\begin{bmatrix} 1 & 1 & \ldots & 1 \\ S_1 & S_2 & \ldots & S_r \\ S_1^2 & S_2^2 & \ldots & S_r^2 \\ \vdots & \vdots & & \vdots \\ S_1^{r-1} & S_2^{r-1} & \ldots & S_r^{r-1} \end{bmatrix} \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_r \end{bmatrix} = \begin{bmatrix} Z \\ Z \\ \vdots \\ Z \end{bmatrix}. \qquad (*)$$

The coefficient matrix here — call it M — is a *van der Monde* matrix with determinant

$$\det M = \prod_{1 \leqslant i < j \leqslant r} (S_j - S_i) = \prod_{1 \leqslant i < j \leqslant r} (T^{k_j} - T^{k_i}).$$

But T is primitive and so $T^{k_i} \neq T^{k_j}$ whenever $0 \leqslant k_i < k_j \leqslant n-1$. Thus $\det M \neq 0$, which means that $(*)$ is impossible, and so Ker H is e-error-correcting. //

Example. With $m = 4$, $n = 2^4 - 1 = 15$, take

$T = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$, $\underset{\sim}{x}^\top = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ and $e = 3$.

$$H = \left[\begin{array}{ccc|ccc|ccc|ccc|ccc} 1&0&0&0&1&0&0&1&1&0&1&0&1&1&1 \\ 0&0&0&1&0&0&1&1&0&1&0&1&1&1&1 \\ 0&0&1&0&0&1&1&0&1&0&1&1&1&1&0 \\ 0&1&0&0&1&1&0&1&0&1&1&1&1&0&0 \\ \hline 1&0&0&0&1&1&0&0&0&1&1&0&0&0&1 \\ 0&1&1&1&1&0&1&1&1&1&0&1&1&1&1 \\ 0&0&1&0&1&0&0&1&0&1&0&0&1&0&1 \\ 0&0&0&1&1&0&0&0&1&1&0&0&0&1&1 \\ \hline 1&0&1&1&0&1&1&0&1&1&0&1&1&0&1 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&1&1&0&1&1&0&1&1&0&1&1&0&1&1 \\ 0&1&1&0&1&1&0&1&1&0&1&1&0&1&1 \end{array}\right] \begin{array}{l} \\ \\ \\ \\ \\ \\ \\ \\ \\ \leftarrow \text{ row of 0's} \\ \left.\begin{array}{l} \\ \\ \end{array}\right\} \text{ identical} \end{array}$$

Ker $H_T(\underset{\sim}{x})$ (top 4 rows) is a single-error-correcting Hamming code with rate $\frac{15-4}{15} = \frac{11}{15}$.

Ker (top 8 rows) is a 2-error-correcting code with rate $\frac{15-8}{15} = \frac{7}{15}$.

Ker H is a 3-error-correcting code with rate $\frac{15-\text{rank}\,H}{15} = \frac{15-10}{15} = \frac{1}{3}$.

Contrast Ker H, length 15, dimension 5, rate $\frac{1}{3}$, with 3-error-correcting Reed–Muller code, length 16, dimension 5, rate $\frac{5}{16}$.

Evaluation of van der Monde determinant:

$$\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ 0 & b-a & c-a \\ 0 & b^2-ab & c^2-ac \end{vmatrix} = \begin{vmatrix} b-a & c-a \\ b^2-ab & c^2-ac \end{vmatrix}$$

$$= (b-a)(c-a) \begin{vmatrix} 1 & 1 \\ b & c \end{vmatrix}$$

$$= (b-a)(c-a)(c-b).$$