

ECE 405/511

Error Control Coding

Hamming Codes and
Bounds on Codes

Single Error Correcting Codes

- (3,1,3) code $R = 1/3$ $n-k = 2$

$$\mathbf{G} = [\mathbf{I}|\mathbf{P}] = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$$

- (5,2,3) code $R = 2/5$ $n-k = 3$

$$\mathbf{G} = [\mathbf{I}|\mathbf{P}] = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

- (6,3,3) code $R = 3/6$ $n-k = 3$

$$\mathbf{G} = [\mathbf{I}|\mathbf{P}] = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

Hamming Codes

- (7,4,3) Hamming code rate $R = 4/7$ $n-k = 3$

$$\mathbf{G} = [\mathbf{I} | \mathbf{P}] = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\mathbf{H} = [\mathbf{P}^T | \mathbf{I}] = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Theorem The minimum distance of a code is equal to the minimum number of columns of \mathbf{H} which sum to zero.

Let $\mathbf{h}_0, \mathbf{h}_1, \dots, \mathbf{h}_{n-1}$ be the column vectors of \mathbf{H} .

For any codeword \mathbf{c} , \mathbf{cH}^T can be expressed as

$$\mathbf{cH}^T = (c_0, c_1, \dots, c_{n-1}) \begin{bmatrix} \mathbf{h}_0^T \\ \mathbf{h}_1^T \\ \vdots \\ \mathbf{h}_{n-1}^T \end{bmatrix} = c_0 \mathbf{h}_0 + c_1 \mathbf{h}_1 + \dots + c_{n-1} \mathbf{h}_{n-1} = \mathbf{0}$$

\mathbf{cH}^T is a linear combination of columns of \mathbf{H} .

Significance of \mathbf{H}

- For a codeword of weight w (w ones), \mathbf{cH}^T is a linear combination of w columns of \mathbf{H} .
- Thus there is a one-to-one mapping between weight w codewords and linear combinations of w columns of \mathbf{H} that sum to 0.
- The minimum value of w which results in $\mathbf{cH}^T=0$, i.e. codeword \mathbf{c} with weight w , determines d_{\min} .

Example

- For the (7,4,3) code, a codeword with weight $d_{\min} = 3$ is given by the first row of \mathbf{G}
 $\mathbf{c} = 1000011$

- The linear combination of the first and last 2 columns in \mathbf{H} gives

$$0 \quad 0 \quad 0 \quad 0$$

$$1 + 1 + 0 = 0$$

$$1 \quad 0 \quad 1 \quad 0$$

- Thus a minimum of 3 columns ($= d_{\min}$) are required to obtain $\mathbf{cH}^T=0$

Hamming Codes

Definition Let $m > 1$ be an integer and \mathbf{H} be an $m \times (2^m - 1)$ matrix with columns which are the non-zero distinct vectors from V_m . The code having \mathbf{H} as its parity-check matrix is a **binary Hamming code** of length $2^m - 1$.

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \Rightarrow \mathbf{G} = [1 \quad 1 \quad 1]$$

$$\mathbf{H} = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \Rightarrow \mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

The Hamming codes are $(2^m - 1, 2^m - 1 - m, 3)$ codes with $m = n - k$ the dimension of \mathbf{H}

Binary Hamming Code Parameters

$$C: \quad n = 2^m - 1$$

$$k = 2^m - 1 - m$$

$$d = 3$$

$$C^\perp: \quad n = 2^m - 1$$

$$k = m$$

$$d = 2^{m-1}$$

Coset Leaders for the Hamming Codes

- There are $2^{n-k} = 2^m$ coset leaders or correctable error patterns
- The number of single error patterns is $n = 2^m - 1$
- Thus the coset leaders are precisely the words of weight ≤ 1
- The syndrome of the word $0 \dots 010 \dots 0$ with 1 in the j th position and 0 otherwise is the transpose of the j th column of **H**

Decoding Hamming Codes

For the case that the columns of \mathbf{H} are arranged in order of increasing binary numbers that represent the column numbers 1 to 2^m-1

Step 1 Given \mathbf{r} compute the syndrome $\mathbf{s} = \mathbf{r}\mathbf{H}^T$

Step 2 If $\mathbf{s} = \mathbf{0}$, then \mathbf{r} is assumed to be the codeword sent

Step 3 If $\mathbf{s} \neq \mathbf{0}$, then assuming a single error, \mathbf{s} gives the binary position of the error

Example

For the Hamming code with parity-check matrix

$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

the received word

$$\mathbf{r} = 1101011$$

has syndrome

$$\mathbf{s} = 110$$

and therefore the error is in the sixth position.

Hamming codes were first used to deal with errors in long-distance telephone calls.

Optimal Codes

- The $(7,4,3)$ code is an optimal single error correcting code for $n-k = 3$
- An $(8,5,3)$ code does not exist
- The $(15,11,3)$ code is an optimal single error correcting code for $n-k = 4$
- A $(16,12,3)$ code does not exist

- What is the limit on the dimension of a code of length n and minimum distance d_{\min} ?

Optimal Codes

$d_{\min} = 1$ $(n, n, 1)$ entire vector space

$d_{\min} = 2$ $(n, n-1, 2)$ single parity check codes

$d_{\min} = 3$ $n = 2^m - 1$ Hamming codes

what about other values of n ?

Shortening

- For $2^{m-1} \leq n < 2^m - 1$, $k = n - m$, use **shortening**
- To get a (6,3,3) code, delete one column say $(1 \ 1 \ 1)^T$ from \mathbf{H} for the (7,4,3) code

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

$n - k$ is constant

so both n and k are changed

$$\mathbf{H}^1 = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{G}^1 = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

- Next delete $(0\ 1\ 1)^T$ which gives a $(5,2,3)$ code

$$\mathbf{H}^2 = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \quad \mathbf{G}^2 = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

- Next delete $(1\ 0\ 1)^T$ which gives a $(4,1,3)$ code

$$\mathbf{H}^3 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \mathbf{G}^3 = [1\ 1\ 1\ 0]$$

- The $(4,1,4)$ repetition code has larger d_{\min} , but a code with $n=4$ and $d_{\min}=3$ cannot have $k>1$

Extending

- The process of deleting a message coordinate from a code is called **shortening**
 $(n, k) \rightarrow (n-1, k-1)$
- Adding an overall parity check to a code is called **extending**
 $(n, k) \rightarrow (n+1, k)$
- Example:

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\mathbf{G}' = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

- If $d(C)$ is odd, $d(C')$ is even
 - In this case, $d(C') = d(C) + 1$
- Example $(7,4,3) \rightarrow (8,4,4)$
- The extended Hamming codes are optimal
 $d_{\min} = 4$ codes

Optimal Codes

$d_{\min} = 1$ $(n, n, 1)$ entire vector space

$d_{\min} = 2$ $(n, n-1, 2)$ single parity check codes

$d_{\min} = 3$ Hamming and shortened Hamming codes

$d_{\min} = 4$ extended $d_{\min} = 3$ codes

Binary Spheres of Radius t

- Let \mathbf{c} be a word of length n . The number of binary words (vectors) of length n and distance i from \mathbf{c} is

$$\binom{n}{i} = \frac{n!}{i!(n-i)!}$$

- For $0 \leq t \leq n$, the number of words of length n a distance at most t from \mathbf{c} is

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{t} = Vol(n, t)$$

Hamming or Sphere Packing Bound

- Spheres of radius t around the M codewords must be disjoint
- The volume of a sphere with radius t is the number of vectors in the sphere: $Vol(n,t)$
- The total volume of the spheres is: $M \times Vol(n,t)$
- This volume must be less than the volume of the vector space: $M \times Vol(n,t) \leq 2^n$

$$M \leq \left\lfloor \frac{2^n}{Vol(n,t)} \right\rfloor$$

Hamming Bound Example

- Give an upper bound on the size of a code C of length $n=11$ and distance $d=3$

$$M \leq \left\lfloor \frac{2^{11}}{\binom{11}{0} + \binom{11}{1}} \right\rfloor = \left\lfloor \frac{2048}{12} \right\rfloor = 170$$

- A code with $M=144$ is known to exist

Hamming Bound for Linear Codes

- For a binary linear code $M = 2^k$

$$2^k \times Vol(n, t) \leq 2^n$$

$$Vol(n, t) \leq 2^{n-k}$$

$$\log_2(Vol(n, t)) \leq n - k$$

$$\lceil \log_2 Vol(n, t) \rceil \leq n - k$$

$$k \leq n - \lceil \log_2 Vol(n, t) \rceil$$

Hamming Bound Example

- Give an upper bound on the size of a **linear** code C of length $n=11$ and distance $d=3$

$$k \leq n - \left\lceil \log_2 \left(\binom{11}{0} + \binom{11}{1} \right) \right\rceil = 11 - \lceil 3.585 \rceil = 7$$

- An $(11,7,3)$ code is known to exist

Hamming Codes

- Consider an (n,k,d) binary Hamming code
- A sphere of radius 1 has volume

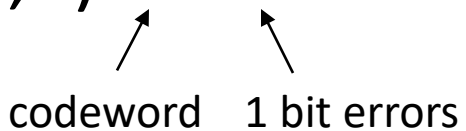
$$Vol(n,1) = 2^m - 1 + 1 = 2^m$$

- The number of codewords is 2^k
- The total volume of the spheres is

$$M \times Vol(n,t) = 2^k \times 2^m = 2^k \times 2^{n-k} = 2^n$$

- The spheres completely fill the n -dimensional vector space V_n

Hamming Code Example

- (7,4,3) Hamming code
- Volume of each sphere is $Vol(7,1)=1+7=8=2^3$


codeword 1 bit errors

- Number of spheres (codewords) is $2^k = 16$
- Volume of all spheres is

$$2^k \times 2^3 = 2^4 \times 2^3 = 2^7 = 2^n$$

Perfect Codes

- A binary linear code is called **perfect** if it meets the Hamming bound with equality

$$2^k \left[1 + \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{t} \right] = 2^n$$

$$\sum_{i=0}^t \binom{n}{i} = Vol(n, t) = 2^{n-k}$$

Perfect Codes

- Binary Hamming codes $t=1$

$$Vol(n,1) = \binom{n}{0} + \binom{n}{1} = 1 + 2^m - 1 = 2^m = 2^{n-k}$$

- Odd binary repetition codes $(2m+1,1,2m+1)$

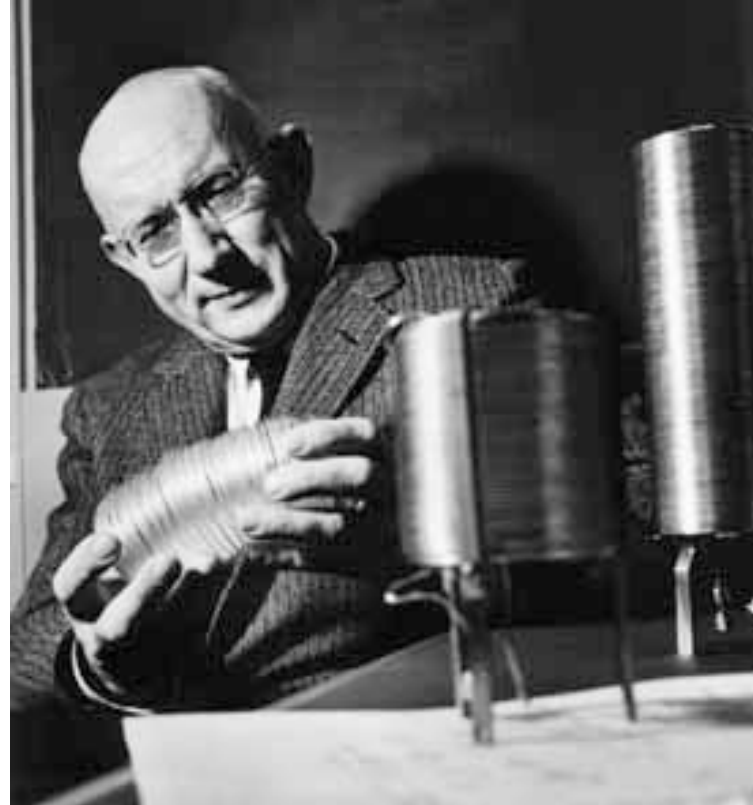
$t=m$

$$\text{sphere volume} = \sum_{i=0}^m \binom{2m+1}{i} = 2^{2m} = 2^{n-k}$$

- $(n,n,1)$ codes (all vectors in V_n are codewords)

$$t=0 \quad Vol(n,0) = 1 \quad 2^n \times 1 = 2^n$$

Marcel Golay (1902-1989)



Blaise Pascal (1623-1662)

- French religious philosopher, writer, physicist, inventor, and mathematician
- First mechanical calculator (1642-1644)
- Developed the modern theory of probability with Pierre de Fermat (1654)
- Pascal's triangle was discovered by Chinese mathematician Yanghui, 500 years before Pascal and in the Eleventh century by Persian mathematician and poet Omar Khayam



Pascal's Triangle

```

      1
     1 1
    1 2 1
   1 3 3 1
  1 4 6 4 1
 1 5 10 10 5 1
1 6 15 20 15 6 1
 1 7 21 35 35 21 7 1
   1 8 28 56 70 56 28 8 1
    1 9 36 84 126 126 84 36 9 1
     1 10 45 120 210 252 210 120 45 10 1
      1 11 55 165 330 462 462 330 165 55 11 1
       1 12 66 220 495 792 924 792 495 220 66 12 1
        1 13 78 286 715 1287 1716 1716 1287 715 286 78 13 1
         1 14 91 364 1001 2002 3003 3432 3003 2002 1001 364 91 14 1
          1 15 105 455 1365 3003 5005 6435 6435 5005 3003 1365 455 105 15 1

```

Rows 87 to 94

1	87	3741	105995	2225895	36949857	504981379
1	88	3828	109736	2331890	39175752	541931236
1	89	3916	113564	2441626	41507642	581106988
1	90	4005	117480	2555190	43949268	622614630
1	91	4095	121485	2672670	46504458	666563898
1	92	4186	125580	2794155	49177128	713068356
1	93	4278	129766	2919735	51971283	762245484
1	94	4371	134044	3049501	54891018	814216767

Rows 19 to 26

1	19	171	969	3876	11628	27132	50388
1	20	190	1140	4845	15504	38760	77520
1	21	210	1330	5985	20349	54264	116280
1	22	231	1540	7315	26334	74613	170544
1	23	253	1771	8855	33649	100947	245157
1	24	276	2024	10626	42504	134596	346104
1	25	300	2300	12650	53130	177100	480700
1	26	325	2600	14950	65780	230230	657800

Golay Codes

- Marcel Golay considered the problem of perfect codes in 1949
- He found three more possible solutions to equality for the Hamming bound
 - $q = 2, n = 23, t = 3$
 - $q = 2, n = 90, t = 2$
 - $q = 3, n = 11, t = 2$
- Only the first and third codes exist

Correspondence

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 Shannon¹ 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

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

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 combinations 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}		Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
X_1	1	0	0	1	1	1	0	0	0	1	1	1	X_1	1	1	1	2	2	0
X_2	1	0	1	0	1	1	0	1	1	0	0	1	X_2	1	1	2	1	0	2
X_3	1	0	1	1	0	1	1	0	1	0	1	0	X_3	1	2	1	0	1	2
X_4	1	0	1	1	1	0	1	1	0	1	0	0	X_4	1	2	0	1	2	1
X_5	1	1	0	0	1	1	1	0	1	1	0	0	X_5	1	0	2	2	1	1
X_6	1	1	0	1	1	1	1	0	0	1	0	0							
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	1	0	0	0	1	1	0	1							
X_{11}	0	1	1	1	1	1	1	1	1	1	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.

Gilbert Bound

- There exists a code of length n , distance d , and M codewords with

$$M \geq \frac{2^n}{\sum_{j=0}^{d-1} \binom{n}{j}} = \frac{2^n}{\text{Vol}(n, d-1)}$$



Edgar Nelson Gilbert
1923-2013

- Note that the constructive proof does not result in a linear code

Bound Comparison

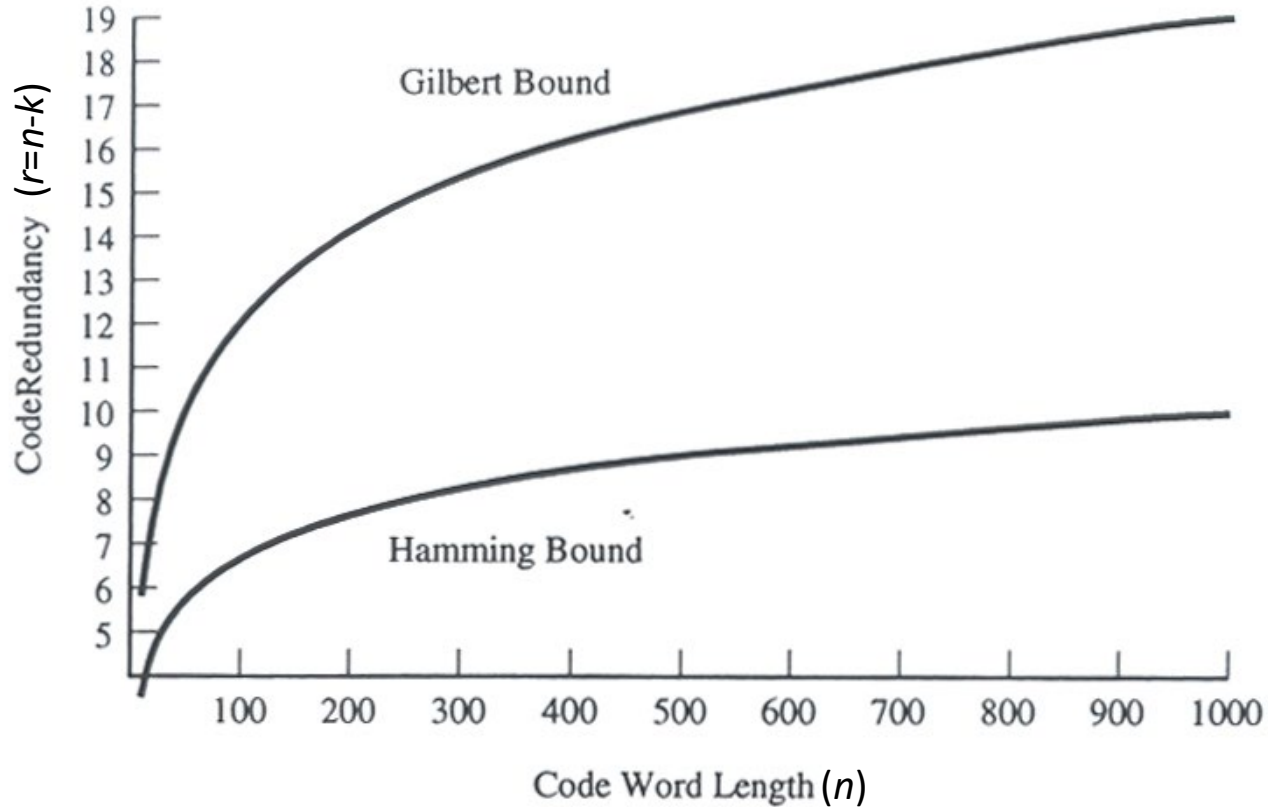


Figure 4-3. A comparison of the Hamming and Gilbert bounds on required redundancy for binary single-error-correcting codes

Gilbert-Varshamov Bound

- The Gilbert bound can be improved by considering linear codes
 - There exists a binary linear code of length n , dimension k and minimum distance d if

$$\binom{n-1}{0} + \binom{n-1}{1} + \dots + \binom{n-1}{d-2} = \text{Vol}(n-1, d-2) < 2^{n-k}$$

- Proof: construct a parity check matrix based on the condition that any combination of up to $d-1$ columns of \mathbf{H} is linearly independent
- Thus a binary (n, k, d) code exists with

$$k \geq n - \left\lfloor \log_2 \left(\sum_{j=0}^{d-2} \binom{n-1}{j} \right) \right\rfloor - 1$$

Asymptotic Bounds

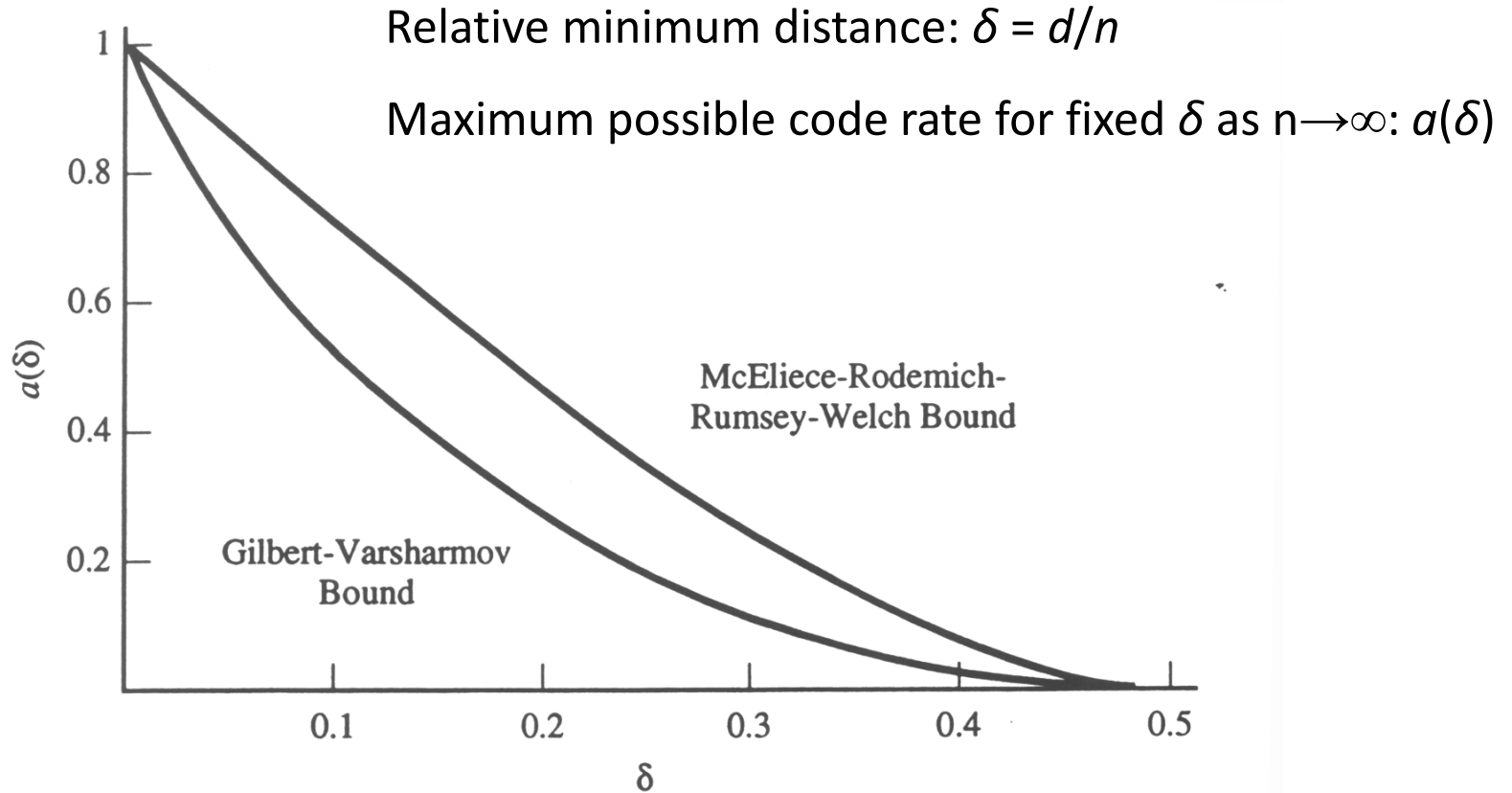


Figure 4-4. Upper and lower bounds for asymptotic binary code performance

Hamming Bound for Linear Codes

$$\sum_{i=0}^t \binom{n}{i} \leq 2^{n-k}$$

$$\left\lceil \log_2 \left(\sum_{i=0}^t \binom{n}{i} \right) \right\rceil \leq n - k$$

$$k \leq n - \left\lceil \log_2 \left(\sum_{i=0}^t \binom{n}{i} \right) \right\rceil$$

Bounds on Binary Linear Codes

- Hamming Bound

$$k \leq n - \left\lceil \log_2 \left(\sum_{i=0}^t \binom{n}{i} \right) \right\rceil$$

- Gilbert-Varshamov Bound

$$k \geq n - \left\lceil \log_2 \left(\sum_{j=0}^{d-2} \binom{n-1}{j} \right) \right\rceil - 1$$