

ELEC 519A

Selected Topics in Digital Communications:  
Information Theory

Hamming Codes and Bounds on Codes

# Single Error Correcting Codes

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

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

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

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

(6, 3, 3) code    rate  $1/2$      $n - k = 3$

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

# Hamming Codes

- (7,4,3) Hamming code

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

- (7,3,4) dual code

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

- The minimum distance of a code  $C$  is equal to the minimum number of columns of  $\mathbf{H}$  which sum to zero
- For any codeword  $c$

$$c\mathbf{H}^T = [c_0, c_1, \dots, c_{n-1}] \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_{n-1} \end{bmatrix} = c_0 h_0 + c_1 h_1 + \dots + c_{n-1} h_{n-1} = 0$$

where  $h_0, h_1, \dots, h_{n-1}$  are the column vectors of  $\mathbf{H}$

- $c\mathbf{H}^T$  is a linear combination of the 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 that  $d_{\min} = w$

# Distance 1 and 2 Codes

- (5,2,1) code

$$\mathbf{H} = \begin{bmatrix} 10100 \\ 10010 \\ 00001 \end{bmatrix} \quad \mathbf{G} = \begin{bmatrix} 10110 \\ 01000 \end{bmatrix}$$

- (5,2,2) codes

$$\mathbf{H} = \begin{bmatrix} 11100 \\ 11010 \\ 00001 \end{bmatrix} \quad \mathbf{G} = \begin{bmatrix} 10110 \\ 01110 \end{bmatrix}$$

# (7,4,3) Code

- A codeword with weight  $d_{\min} = 3$  is given by the first row of **G**:  $c = 1000011$
- The linear combination of the first and last 2 columns in **H** gives

$$(011)^T + (010)^T + (001)^T = (000)^T$$

- Thus a minimum of 3 columns ( $= d_{\min}$ ) are required to get a zero value for  $c\mathbf{H}^T$

# Binary Hamming Codes

Let  $m$  be an integer and  $\mathbf{H}$  be an  $m \times (2^m - 1)$  matrix with columns which are the non-zero distinct elements 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  
 $m = n - k$



# Hamming Code Parameters

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

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

$$d = 3$$

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

$$k = m$$

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

# 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  $\leq 1$
- 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  $r$  compute the syndrome  $s = r\mathbf{H}^T$
- **Step 2** If  $s = 0$ , then  $r$  is assumed to be the codeword sent
- **Step 3** If  $s \neq 0$ , then assuming a single error,  $s$  gives the position of the error in binary

# Example

The (7,4,3) Hamming code has 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

$$r = 1101011$$

has syndrome

$$s = 110$$

and therefore the error is in the sixth position.

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

- 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  $n-k = 4$
- What is the limit on how many errors a code can correct?

# The Main Coding Theory Problem

A good  $(n, M, d)$  code has small  $n$ , large  $M$  and large  $d$ .

The main coding theory problem is to optimize one of the parameters  $n, M, d$  for given values of the other two.

For linear codes,  $M = q^k$ , so a good  $(n, k, d)$  code has small  $n$ , large  $k$  and large  $d$ .

The main coding theory problem for linear codes is to optimize one of the parameters  $n, k, d$  for given values of the other two.

# Optimal Binary 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  $2^{m-1}-1 < n \leq 2^m-1$ ,  $k = n-m$ , use **shortening**
- To get a (6,3,3) code, delete one column say  $(1 \ 1 \ 1)^T$  from **H**

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

$n-k$  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  $(0\ 1\ 1)^T$  to get 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  $(1\ 0\ 1)^T$  to get 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  $d_{\min}$
- Question: Does a  $(4,2,3)$  binary code exist?

# Self-Dual (4,2,3) Ternary Code Example

- $C = C^\perp \geq n-k=k \rightarrow k=n/2$
- $\mathbf{G} = [\mathbf{I} \ \mathbf{P}] \quad \mathbf{G}\mathbf{G}^T = \mathbf{0} \Leftrightarrow \mathbf{I} + \mathbf{P}\mathbf{P}^T = \mathbf{0} \rightarrow \mathbf{P}\mathbf{P}^T = -\mathbf{I}$
- Ternary self-dual code with  $n=4, k=2, d=3$

$$\mathbf{G} = \begin{bmatrix} 1011 \\ 0112 \end{bmatrix} \quad \begin{array}{l} (1011) \cdot (1011) = 0 \\ (1011) \cdot (0112) = 0 \\ (0112) \cdot (0112) = 0 \end{array}$$

- 9 Codewords

0000 1011 2022 0112 0221

1120 2201 1202 2101

# 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:

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

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

# Extended Binary Codes

- 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 Binary 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 $t$

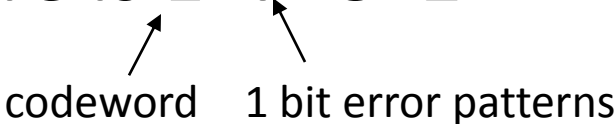
- The number of binary words (vectors) of length  $n$  and distance  $i$  from a word  $c$  is

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

- Let  $c$  be a word of length  $n$ . For  $0 \leq t \leq n$ , the number of words of length  $n$  a distance at most  $t$  from  $c$  is

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{t}$$

# Hamming or Sphere Packing Bound

- Consider an  $(n,k,d)$  binary code
- Spheres of radius  $t$  around the  $2^k$  codewords must be disjoint
- Volume of a sphere with radius  $t$  is the number of vectors in the sphere
- Example:  $(7,4,3)$  Hamming code  $t=1$
- Volume of each sphere is  $1+7=8=2^3$   


codeword    1 bit error patterns

- Number of spheres (codewords) is  $2^k = 16$
- Volume of all spheres is  $2^k \cdot 2^3 = 2^7 = 2^n$
- The spheres completely fill the  $n$ -dimensional space

- The Hamming bound (binary)

$$2^k \left[ 1 + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{t} \right] \leq 2^n \quad \text{or} \quad \sum_{i=0}^t \binom{n}{i} \leq 2^{n-k}$$

- A binary code is called **perfect** if it meets this bound with equality



# Hamming Bound Example

- Give an upper bound on the size of a linear code  $C$  of length  $n=6$  and distance  $d=3$

$$|C| = 2^k \leq \frac{2^6}{\binom{6}{0} + \binom{6}{1}} = \frac{64}{7}$$

- This gives  $|C| \leq 9$  but the size of a linear code must be a power of 2 so  $|C| \leq 8$  and  $k \leq 3$

# Codes that meet the Hamming Bound

- Binary Hamming codes

$$\binom{n}{0} + \binom{n}{1} = 1 + 2^m - 1 = 2^m = 2^{n-k}$$

- 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)

# Hamming Bound for Nonbinary Codes

- For  $\text{GF}(q)$

$$\sum_{i=0}^t \binom{n}{i} (q-1)^i \leq q^{n-k}$$

- Size of the vector space is  $q^n$
- The number of codewords is  $q^k$
- Each error location has  $q-1$  possible error values
- Two of the three classes of perfect binary linear codes also exist for nonbinary alphabets

- Vector space codes

$$\sum_{i=0}^0 \binom{n}{i} (q-1)^i = q^0 = q^{n-k}$$

- Nonbinary Hamming codes

- $\mathbf{H}$  has  $m$  rows
- There are  $q^m - 1$  possible nonzero  $q$ -ary  $m$ -tuples
- For each  $q$ -ary  $m$ -tuple, there are  $q - 1$  distinct nonzero  $m$ -tuples that are a multiple of that  $m$ -tuple
- Example: GF(5)
 

1	2	3	4
0	0	0	0
0	0	0	0

- **H** has dimension  $m \times \frac{q^m - 1}{q - 1}$

$$n = \frac{q^m - 1}{q - 1} \quad k = n - m \quad d_{\min} = 3$$

- Example:  $m=3, q=3$ , 26 possible nonzero  $m$ -tuples

only  $\frac{1}{2}$  are usable

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

$$\sum_{i=0}^1 \binom{n}{i} (q-1)^i = 1 + n(q-1) = q^m = q^{n-k}$$

# Marcel J. E. Golay (1902-1989)



# Golay Codes

- Marcel Golay considered the problem of perfect codes in 1949
- He found three 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 [Van Lint and Tietäväinen, 1973]

# 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 Shannon<sup>1</sup> 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

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

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	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	0	1							
$X_9$	1	1	1	1	0	1	0	1	0	0	0	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							

\* Received by the Institute, February 23, 1949.

<sup>1</sup> 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

$$M = \left\lceil \frac{2^n}{\sum_{j=0}^{d-1} \binom{n}{j}} \right\rceil$$

- The bound also holds for **linear** codes but  $M$  must be a power of 2
  - Let  $k$  be the largest integer such that

$$2^k \leq \frac{2^n}{\sum_{j=0}^{d-1} \binom{n}{j}}$$

then an  $(n,k,d)$  code exists

# Gilbert-Varshamov Bound

- For linear codes, the Gilbert bound can be improved
  - There exists a linear code of length  $n$ , dimension  $k$  and minimum distance  $d$  if

$$\binom{n-1}{0} + \binom{n-1}{1} + \dots + \binom{n-1}{d-2} < 2^{n-k}$$

- Proof: construct a parity check matrix based on this condition
- Thus if  $k$  is the largest integer such that

$$2^k < \frac{2^n}{\sum_{j=0}^{d-2} \binom{n-1}{j}}$$

then an  $(n,k,d)$  code exists

# Examples

- Does there exist a binary linear code of length  $n=9$ , dimension  $k=2$ , and distance  $d=5$ ? **Yes**, because

$$\binom{8}{0} + \binom{8}{1} + \binom{8}{2} + \binom{8}{3} = 93 < 128 = 2^{9-2}$$

- Give a lower and an upper bound on the dimension,  $k$ , of a linear code with  $n=9$  and  $d=5$
- G-V lower bound:**  $2^k < \frac{2^9}{93} = 5.55$  but  $|C|$  is a power of 2 so  $|C| \geq 4$  and  $k \geq 2$

# Examples (Cont.)

- Hamming upper bound:

$$|C| \leq \frac{2^9}{\binom{9}{0} + \binom{9}{1} + \binom{9}{2}} = \frac{512}{1+9+36} = 11.13$$

but  $|C|$  must be a power of 2 so  $|C| \leq 8$  and  $k \leq 3$

- From the tables, the optimal codes are
  - (9,2,6)
    - the G-V bound is exceeded, so it is sufficient but not necessary for a code to exist
  - (9,3,4)
    - the Hamming bound is necessary but not sufficient for a code to exist

# G-V Bound and Codes

- Does a (15,7,5) binary linear code exist?
  - Check the G-V bound

$$\binom{n-1}{0} + \binom{n-1}{1} + \dots + \binom{n-1}{d-2} = \binom{14}{0} + \binom{14}{1} + \binom{14}{2} + \binom{14}{3} \\ = 1 + 14 + 91 + 364 = 470 > 2^{15-7} = 2^{n-k} = 256$$

- G-V bound does not hold, so it does not tell us whether or not such a code exists.
- Actually such a code does exist - the (15,7,5) BCH code

- Check with the Hamming bound
- A (15,7,5) BCH code has sphere volume

$$1 + 15 + \binom{15}{2} = 121$$

- The total volume of the spheres is

$$121 \times 2^7 = 15488 < 2^{15}$$

# The Nordstrom-Robinson Code

- Adding an overall parity check to the  $(15,7,5)$  code gives a  $(16,7,6)$  linear code
  - This is an optimal linear code
  - The G-V bound says a  $(16,5,6)$  code exists
- A  $(16,256,6)$  **nonlinear** code exists
  - Twice as many codewords as the optimal linear code

# Bounds for Nonbinary Codes

- Gilbert bound for nonlinear codes: There exists a code of length  $n$ , dimension  $k$  and distance  $d$  if

$$M = \left\lceil \frac{q^n}{\sum_{j=0}^{d-1} \binom{n}{j} (q-1)^j} \right\rceil$$

- Gilbert-Varshamov bound for linear codes: Let  $k$  be the largest integer such that

$$q^k < \frac{q^n}{\sum_{j=0}^{d-2} \binom{n-1}{j} (q-1)^j}$$

is satisfied, then an  $(n, k, d)$  code exists



# Singleton Bound

- Singleton bound (upper bound)

For any  $(n,k,d)$  linear code,  $d \leq n-k-1$

$$k \leq n-d+1 \text{ or } |C| \leq q^{n-d+1}$$

Proof: the parity check matrix  $\mathbf{H}$  of an  $(n,k,d)$  linear code is an  $n-k$  by  $n$  matrix such that every  $d-1$  columns of  $\mathbf{H}$  are independent. Since the columns have length  $n-k$ , we can never have more than  $n-k$  independent columns. Hence  $d-1 \leq n-k$  or  $d \leq n-k+1$ .

- For an  $(n,k,d)$  linear code  $C$ , the following are equivalent:
  - $d = n-k+1$
  - Every  $n-k$  columns of the parity check matrix are linearly independent
  - Every  $k$  columns of the generator matrix are linearly independent
  - $C$  is Maximum Distance Separable (MDS) (definition:  $d=n-k+1$ )
  - $C^\perp$  is MDS

# Example

- (255,223,33) RS code over  $\text{GF}(2^8)$

$$\frac{\text{\# of codewords} \times \text{volume}}{\text{size of vector space}} = 2.78 \times 10^{-14}$$

- Singleton bound:

$$d_{\min} \leq n-k+1$$

- The (255,223,33) RS code meets the Singleton bound with equality

# Marcel J. E. Golay

