ECE 405/ECE 511
Error Control Coding

Binary Linear Block Codes

Basic Concept

- The key idea is to encode a message by adding redundant data or parity to the message.
- The parity bits added at the encoder can be used for error detection and/or correction at the decoder.

- Linear codes are the most important class of error correcting codes
  - simple description
  - nice properties
  - easy encoding
  - conceptually easy decoding

Modular Arithmetic

- With binary codes, modulo 2 arithmetic is used.
- A number mod 2 is obtained by dividing it by 2 and taking the remainder.
- For example, 3 ≡ 1 mod 2 and 4 ≡ 0 mod 2.

mod 2 addition

+	0	1
0	0	1
1	1	0

same as logical XOR

mod 2 multiplication

•	0	1
0	0	0
1	0	1

same as logical AND

Vector Space

- Set of *n*-tuples over an alphabet A
  - *n*-dimensional vector space V_n
- Example: binary *n*-tuples of length 5 – V_5
  - 5-dimensional vector space A = {0,1}

$$\left.\begin{matrix} 00000 \\ 00001 \\ 00010 \\ 00011 \\ 00100 \\ \vdots \\ 11111 \end{matrix}\right\} 32 \quad \text{5-tuples}$$

Vector Space Operations

vector addition

$$
\begin{array}{r} 11001 \\ +\underline{10011} \\ 01010 \end{array}
$$

scalar multiplication

$$a \bullet \mathbf{v}, \quad a \in \mathrm{A}$$

$$0 \bullet (11001) = 00000$$

$$1 \bullet (11001) = 11001$$

The space is closed under vector addition and scalar multiplication

Inner Product

$$\mathbf{x} \circ \mathbf{y} = \sum_{i=0}^{n-1} x_i \bullet y_i$$

$$(11001) \circ (10011) = 1 \bullet 1 + 1 \bullet 0 + 0 \bullet 0 + 0 \bullet 1 + 1 \bullet 1$$

$$= 2$$

$$= 0 \bmod 2$$

11001 and 10011 are orthogonal

Vector Subspace

- A subset of a vector space that is closed under vector addition and scalar multiplication
- Example: subspace of V_5

$$S = \begin{matrix} 00000 \\ 00111 \\ 11100 \\ 11011 \end{matrix}$$

$$\begin{array}{r} 00111 \\ +\underline{11011} \\ 11100 \end{array}$$

Basis

- A minimal number of linearly independent vectors (k) from the vector space that span the space

$$\begin{bmatrix} 00111 \\ 11100 \end{bmatrix}$$

$$0 \bullet (00111) + 0 \bullet (11100) = 00000$$
$$0 \bullet (00111) + 1 \bullet (11100) = 11100$$
$$1 \bullet (00111) + 0 \bullet (11100) = 00111$$
$$1 \bullet (00111) + 1 \bullet (11100) = 11011$$

- Any vector in the space is a linear combination of basis vectors

Dual Space

- Set of vectors orthogonal to a vector space

S	$S^{\perp}$
0000	0000
0111	0011
1100	1110
1011	1101

$$|S| \times |S^{\perp}| = |V|$$

Dual Space

- Set of vectors orthogonal to a vector space

$$S = \begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \end{matrix}$$

$$S^{\perp} = \begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{matrix}$$

Vector Space Dimensions

- If a basis has *k* vectors then the vector space is said to have dimension *k*

$$S \qquad\qquad S^{\perp}$$

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

$$\dim(S) + \dim(S^{\perp}) = \dim(V)$$

Example

- For the subspace generated by the basis

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

find a basis of the dual space

- In this case dim(V) = 5 and k = 2 so

$$\dim(S^{\perp}) = 5 - 2 = 3 \qquad \left|S^{\perp}\right| = 2^3 = 8$$

Example

- For the subspace generated by the basis

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

a basis of the dual space is

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

Self-Dual Spaces

$$S = S^{\perp}$$

- Example

S	$S^{\perp}$
0000	0000
1010	1010
0101	0101
1111	1111

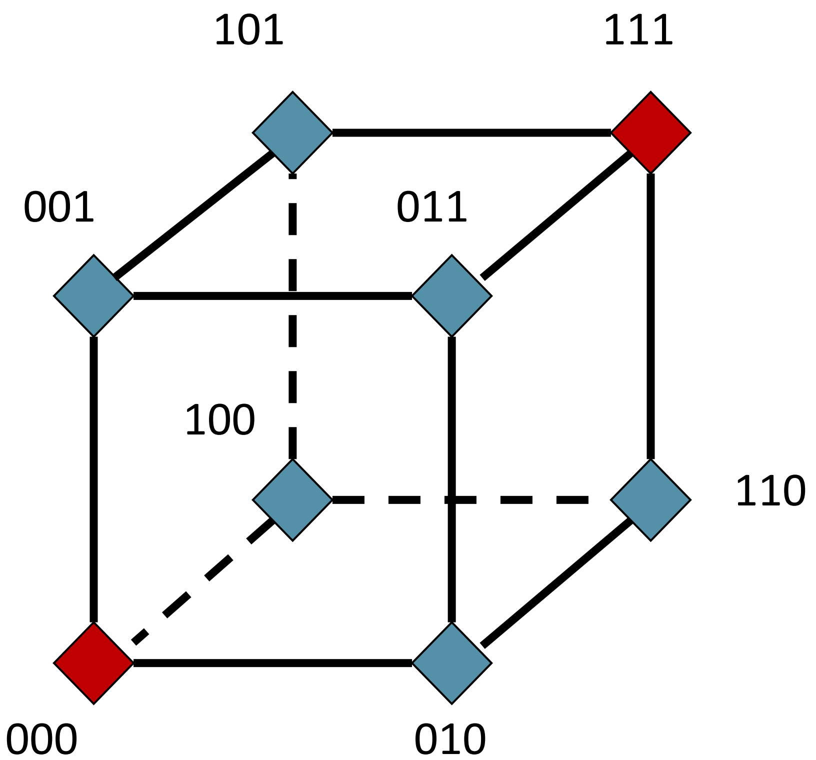
Binary Codes in Vector Spaces

Codewords can be considered as vectors in the vector space V_n of binary vectors of length n.

Definition A subset $C \subseteq V_n$ is a binary linear block code if $\mathbf{u} + \mathbf{v} \in C$ for all $\mathbf{u}, \mathbf{v} \in C$.

C is a k dimensional subspace of V_n.

Triple Repetition Code



V_3

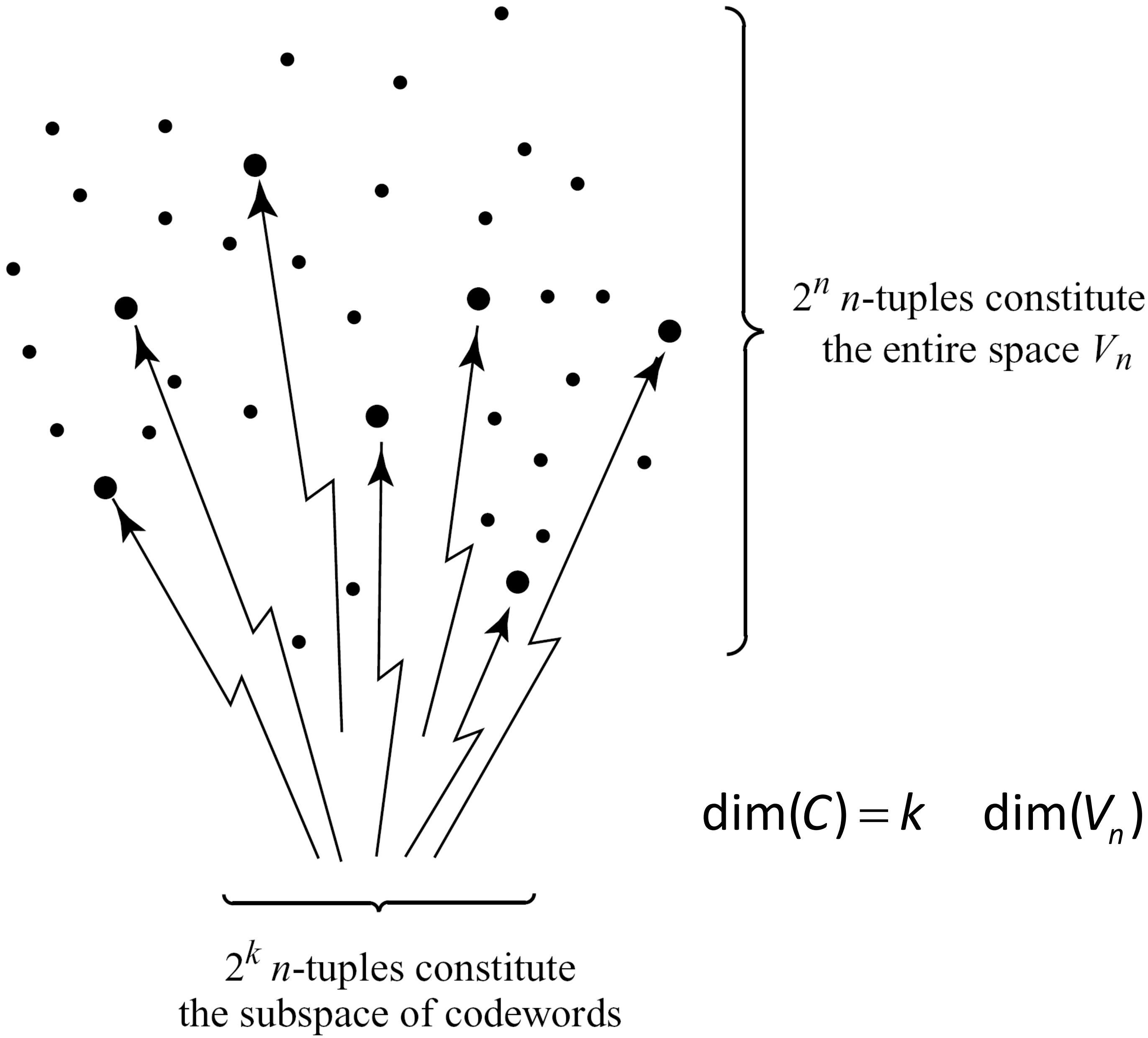
- The vector subspace is 000,111
- Basis [111]
- Dimension $k = 1$
- Length $n = 3$

Binary Linear Block Codes

- Binary linear code: mod 2 sum of any two codewords is a codeword
- Block code: codewords have a finite length *n*
- The number of codewords in a binary linear block code *C* is

$$|C| = M = 2^k$$

- Each codeword of length *n* represents *k* data bits
- The code rate is $R = \frac{\log_2 M}{n} = \frac{k}{n}$



$$\dim(C) = k \quad \dim(V_n) = n$$

Which of the following binary codes is linear?

C_1 = {00, 01, 10, 11}
C_2 = {000, 011, 101, 110}
C_3 = {00000, 11110, 10011, 01101}
C_4 = {101, 111, 011}
C_5 = {000, 001, 010, 011}
C_6 = {0000, 1001, 0110, 1110}

Answer: C_1, C_2 , C_3 and C_5

Generator (Basis) Matrices

- (3,1) repetition code
  - $n = 3$, $k = 1$

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

- **c = mG**

m = 0 **c** = 000

m = 1 **c** = 111

Generator (Basis) Matrices

- (8,7) single parity check code
  - $n = 8$, $k = 7$

ASCII

E = 1000101 **c** = 10001011

G = 1000111 **c** = 10001110

$$\mathbf{G} = \begin{bmatrix} & & 1 \\ & & 1 \\ & & 1 \\ & I_7 & 1 \\ & & 1 \\ & & 1 \\ & & 1 \end{bmatrix}$$

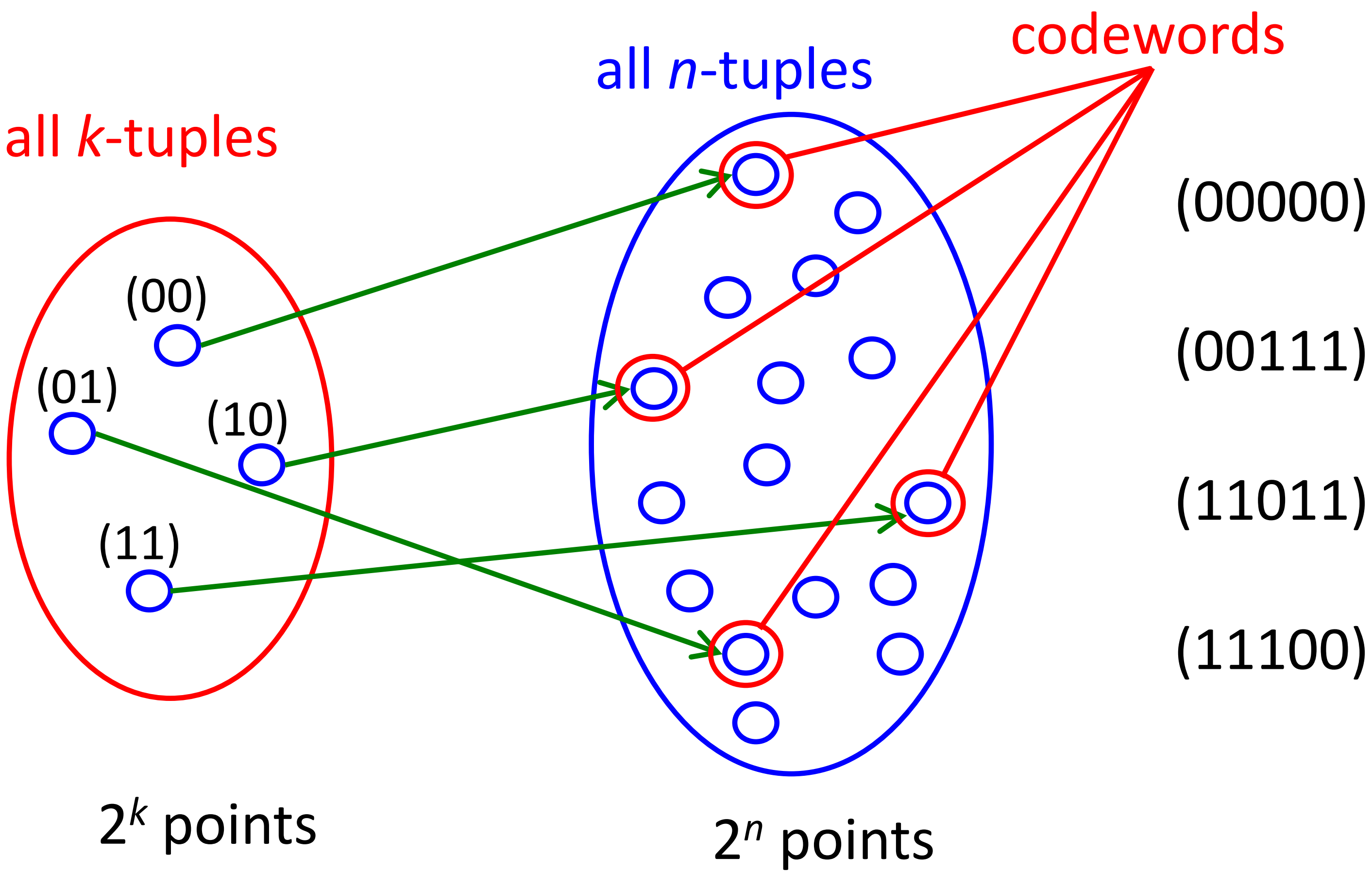
(5,2) Binary Linear Code

- k x n Generator Matrix $\mathbf{G} = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix}$ ($n = 5$, $k = 2$)
- $\mathbf{c} = \mathbf{mG}$

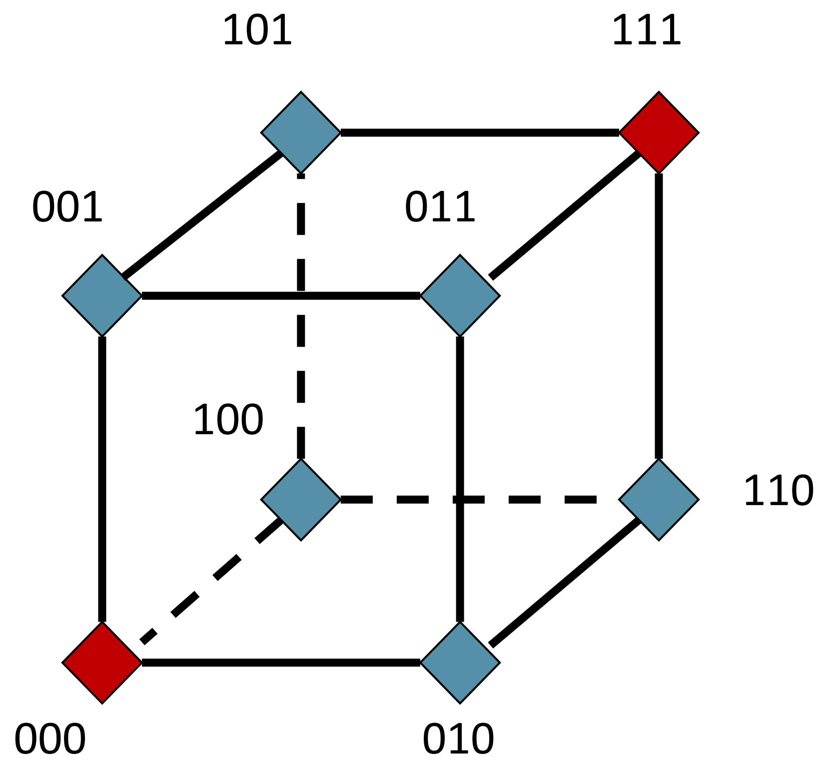
m	**c**
00	00000
01	11100
10	00111
11	11011

Linear Codes as Vector Subspaces

$$(m_0 m_1 \ldots m_{k-1}) \rightarrow (c_0 c_1 \ldots c_{n-1})$$

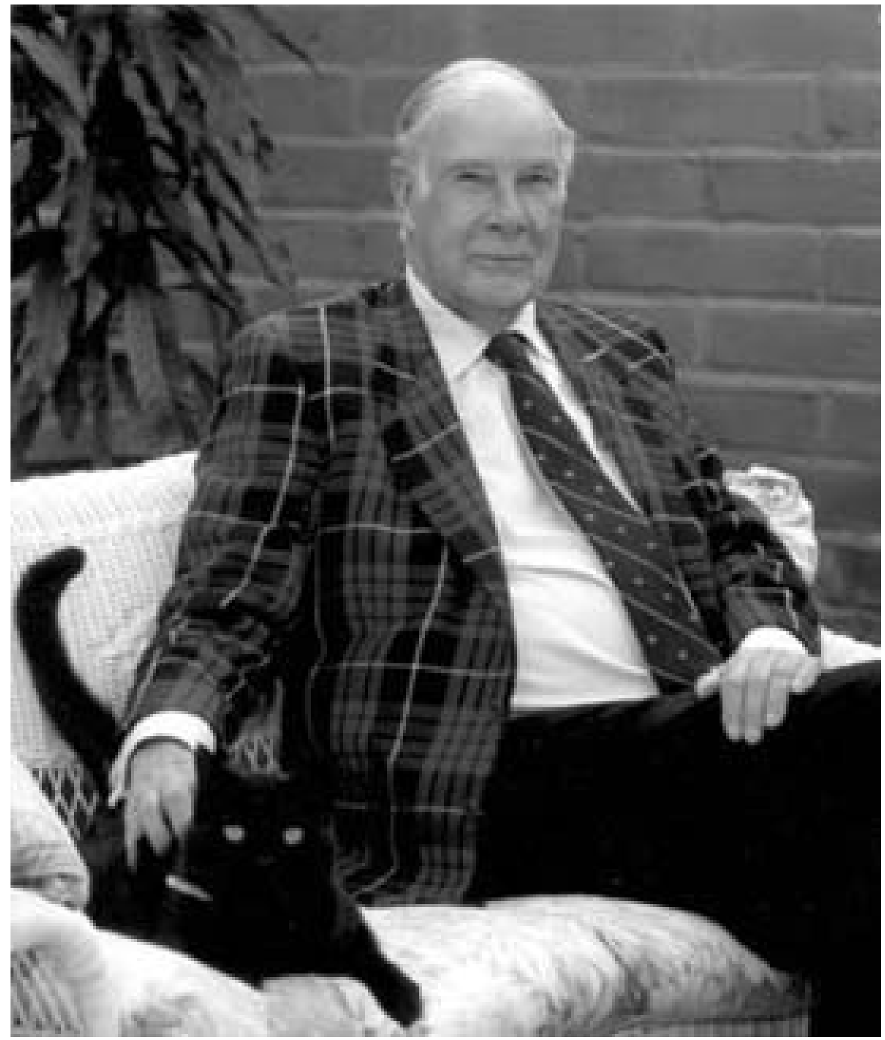


Triple Repetition Code



V_3

Richard W. Hamming (1915-1998)



Hamming at Bell Labs

- The development of error correcting codes began in 1947 at Bell Laboratories
- Hamming had access to a mechanical relay computer on some weekends
- The computer employed an error detecting code, but with no operator on duty during weekends, the computer simply stopped or went on to the next problem when an error occurred

“Two weekends in a row I came in and found that all my stuff had been dumped and nothing was done.” And so I said, “Damn it, if the machine can detect an error, why can't it locate the position of the error and correct it?”

Hamming Weight and Distance

- The concept of closeness of two codewords is formalized through the Hamming distance.
- Let **x** and **y** be two codewords in *C*

  $\mathbf{x} = 00111 \quad \mathbf{y} = 11100$

- The Hamming weight of a codeword is defined as the number of nonzero elements in the codeword

  $w(\mathbf{x}) = w(00111) = 3 \quad w(\mathbf{y}) = w(11100) = 3$

- The Hamming distance between two codewords is defined as the number of places in which they differ

  $d(\mathbf{x},\mathbf{y}) = d(00111,11100) = 4$

Hamming Distances for Linear Codes

- For a binary linear code, the addition of any two codewords is another codeword

  $\mathbf{x} + \mathbf{y} = \mathbf{z}$ 00111 + 11100 = 11011

- Thus

  $d(\mathbf{x},\mathbf{y}) = w(\mathbf{x}+\mathbf{y}) = w(\mathbf{z}) = w(11011) = 4$

- Since we are concerned with the error correcting capability of a code *C*, an important criteria is the minimum Hamming distance d(*C*) or d_{min} between all pairs of codewords

Minimum Hamming Distance

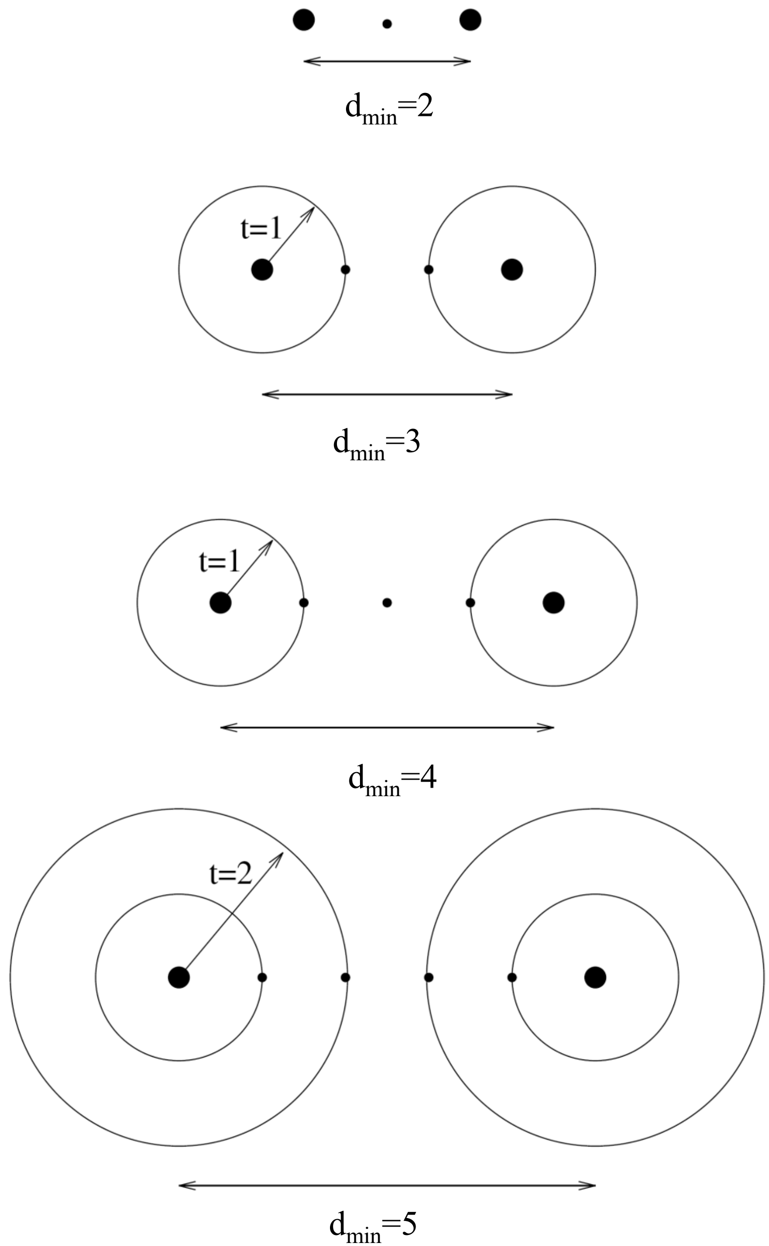
- The minimum Hamming distance of a code *C* is

  $d(C) = \min \{d(\mathbf{x},\mathbf{y}) \mid \mathbf{x},\mathbf{y} \in C, \mathbf{x} \neq \mathbf{y}\}$

  (also called d_{min})

- For a linear code

  $d(C) = \min \{w(\mathbf{x}) \mid \mathbf{x} \in C, \mathbf{x} \neq 0\}$



d_min=2
t=1
d_min=3
t=1
d_min=4
t=2
d_min=5

Minimum Hamming Distance

- A code *C* can detect up to *v* errors where

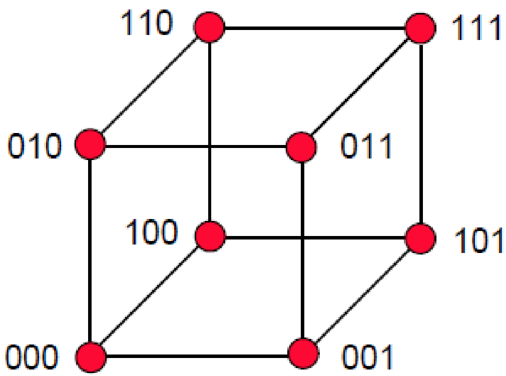
$$v = d(C)-1$$

- A code *C* can correct up to *t* errors where

$$t = \left\lfloor \frac{d(C)-1}{2} \right\rfloor$$

Linear Codes of Length 3

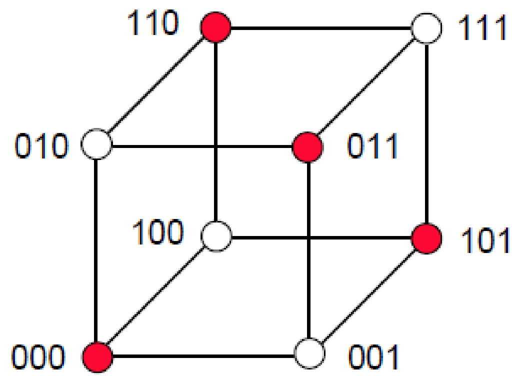
C_1 all 8 vectors used



$d_{min}=1$

Code rate R = 1
No error correction
No error detection

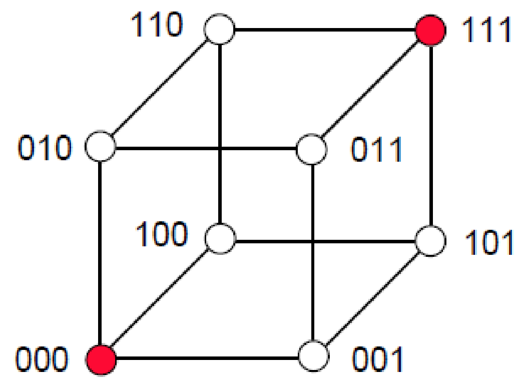
C_2 only 4 vectors used



$d_{min}=2$

Code rate R = 2/3
No error correction
Single error detection

C_3 only 2 vectors used



$d_{min}=3$

Code rate R = 1/3
Single error correction
Double error detection

Notation and Examples

An (n,k,d) code C is a linear code where

- n is the length of the codewords
- k is the number of data bits represented by a codeword
- d is the minimum distance of C

$$d = \mathrm{d}(C) = d_{\mathrm{min}}$$

Examples

C_1 = {000, 100, 010, 001, 011, 101, 110, 111} is a (3,3,1) code

C_2 = {000, 011, 101, 110} is a (3,2,2) code

C_3 = {000, 111} is a (3,1,3) code

A good code has small n-k and large d.

(5,2,3) Binary Linear Code

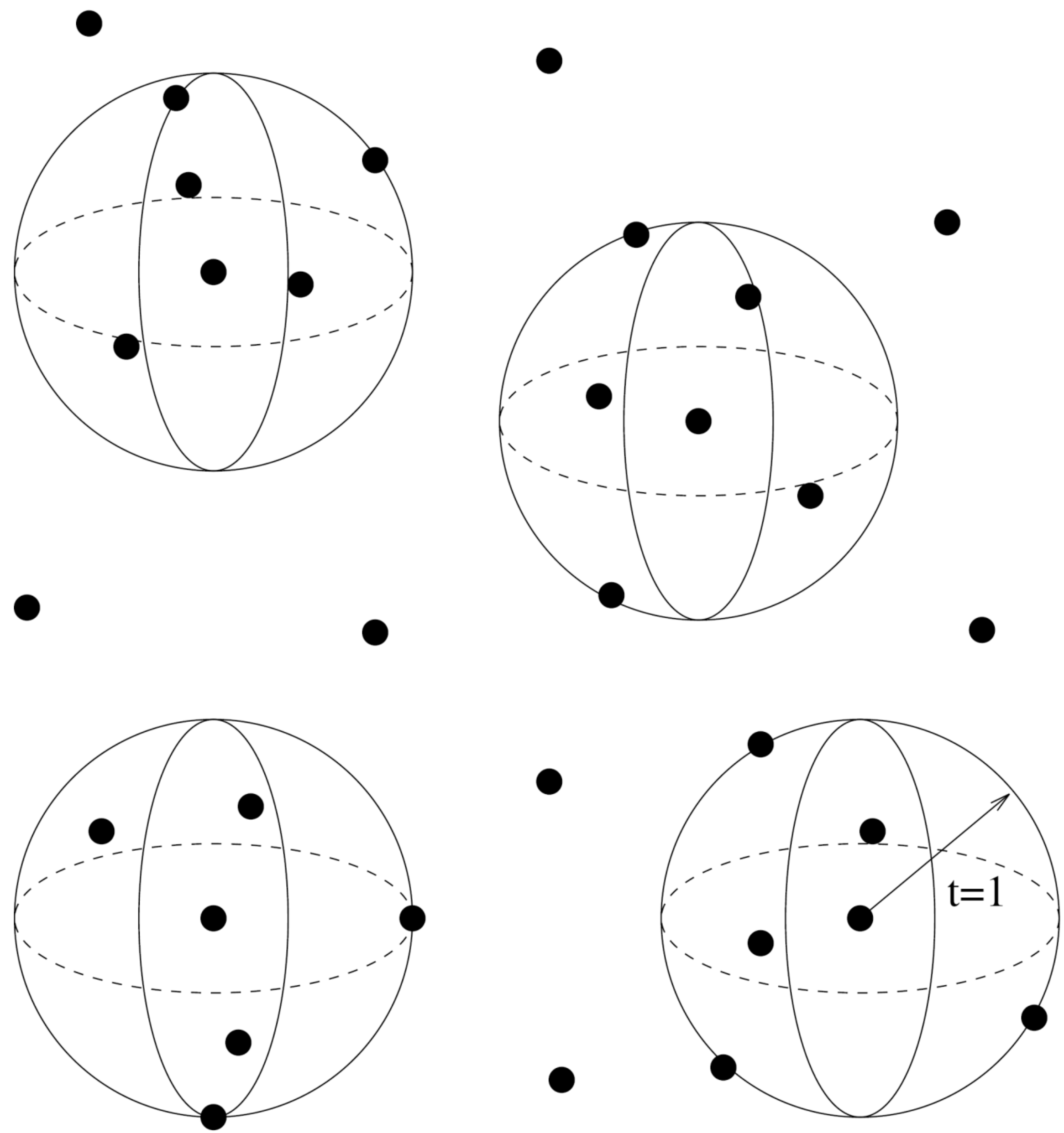
- k x n Generator Matrix
- **c** = **mG**

$$\mathbf{G} = \overbrace{\begin{bmatrix} 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix}}^{5} \; 2$$

m	**c**	w(**c**)
00	00000	0
01	11100	3
10	00111	3
11	11011	4

- $d_{\min}$ = d(C) = 3 $\quad t = \left\lfloor \dfrac{3-1}{2} \right\rfloor = 1$

V_5



Advantages of Linear Block Codes

1. The minimum distance $d_{\min}$ is relatively easy to compute.
2. Linear codes can be simply characterized.
   - To specify a non-linear code often requires all codewords to be listed.
   - To specify a linear (n,k) code it is enough to list k linearly independent codewords. These codewords form a basis for the vector subspace and this $k\times n$ matrix is called a generator matrix for C.

Examples

$$C_3 = \begin{matrix} 0 & 0 & 0 \\ 1 & 1 & 1 \end{matrix} \qquad \mathbf{G} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \qquad \text{(3,1,3) code}$$

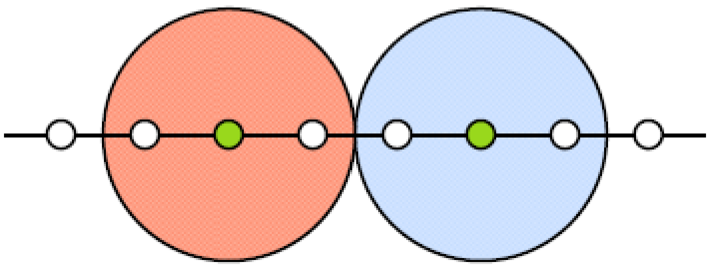
$$C_2 = \begin{matrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{matrix} \qquad \mathbf{G} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \qquad \text{(3,2,2) code}$$

3. There are simple encoding and decoding procedures for linear codes.

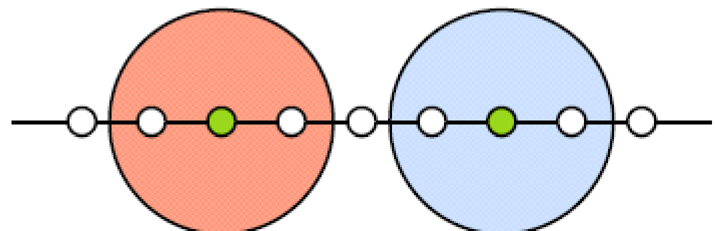
Important Linear Block Codes

There are many classes of practical linear block codes:

- Hamming codes
- Cyclic codes (CRC codes)
- BCH codes
- Reed-Solomon codes
- Reed-Muller codes
- Product codes
- LDPC codes
- Turbo codes
- …



$d_{\min} = 3$



$d_{\min} = 4$