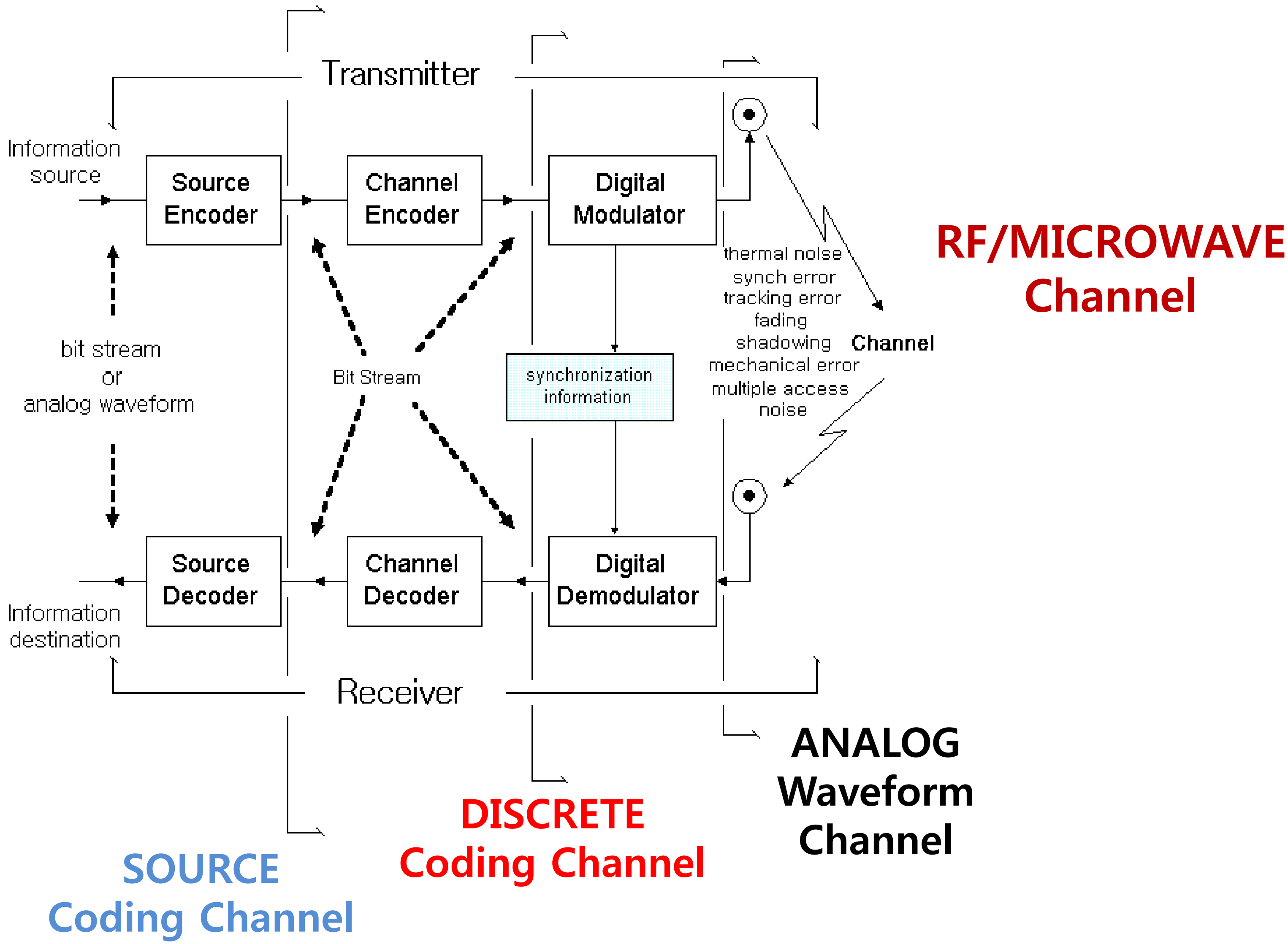
오류정정부호 기본 개념 1

How to detect and then correct errors ?

송홍엽교수

전기전자공학부 주니어세미나 수업자료

Digital Communication Modem



Types of Errors
오류의 종류

- Random error:
  - DMC+AWGN, wireless channel
  - Communication channel **without fading**
- Burst error
  - Most recording channel (Hard Disk, DVD)
  - Communication channel with **deep fading**

→ 발생 상관에 따라

000 → 001

- Substitution error
- Insertion error
- Deletion error

→ 발생 형태에 따라

0001 → 00
000 → 0010

오류의 종류는 채널의 특성에 따른다.

Methods of Error Control

- **Forward error correction(FEC)**
  - most powerful and complex (mostly in wireless communication)
  - one-way transmission systems: **deep space** or storage
- **Error detection with ARQ**
  - less powerful and simpler than FEC (mostly in wireline network)
  - **two-way transmission systems**
  - stop-and-wait ARQ, continuous ARQ, selective-repeat ARQ
- **Error concealment**
  - essentially a **signal processing** technique
  - process the received noisy information so that the effect of errors becomes negligible: **Audio/Video Coding**

How is it possible
to **detect** errors
and then sometimes
correct those errors?

Is this a **magic**?
No, it is not.

What's going on?

Space of Natural Language

- 서운시 서대문구 연세로 50
  - 오류검출가능 **그리고** 정정가능
- 연세대학교 전가 공학과
  - 오류검출**가능 그러나** 정정불가능 (전자 ? 전기 ?)
- 오류정정부호 과목 수강생은 20명이다.
  - 오류검출 **불가능**(??) **그러므로** 정정불가

What's going on?

Space of Binary Language

사용하는 코드의 길이는 3,　　모든 가능한 단어는 총 8개

이 중에서 "일부"만 사용하기로 하자.

예를 들어 다음의 4개만 사용: 000, 110, 101, 011.

1. 수신한 단어가 **001**:
   - 오류가 검출되는가?
   - 어느 비트가 오류인가?
2. 수신한 단어가 **110**:
   - 오류가 검출되는가?
   - 어느 비트가 오류인가?

(Hamming) Distance and Weight

- **Hamming Distance**
  - codeword u = 000111
  - codeword v = 011100 ⇒ distance = 4
  - **Minimum (Hamming) Distance** of a code is the minimum of distances between all possible two distinct codewords

- **Hamming Weight**
  - codeword u+v = 011011 ⇒ weight = 4
  - **Minimum (Hamming) Weight** of a code is the minimum weight of all the non-zero codewords

Space of binary language

Addition

+	0	1
0	**0**	**1**
1	**1**	**0**

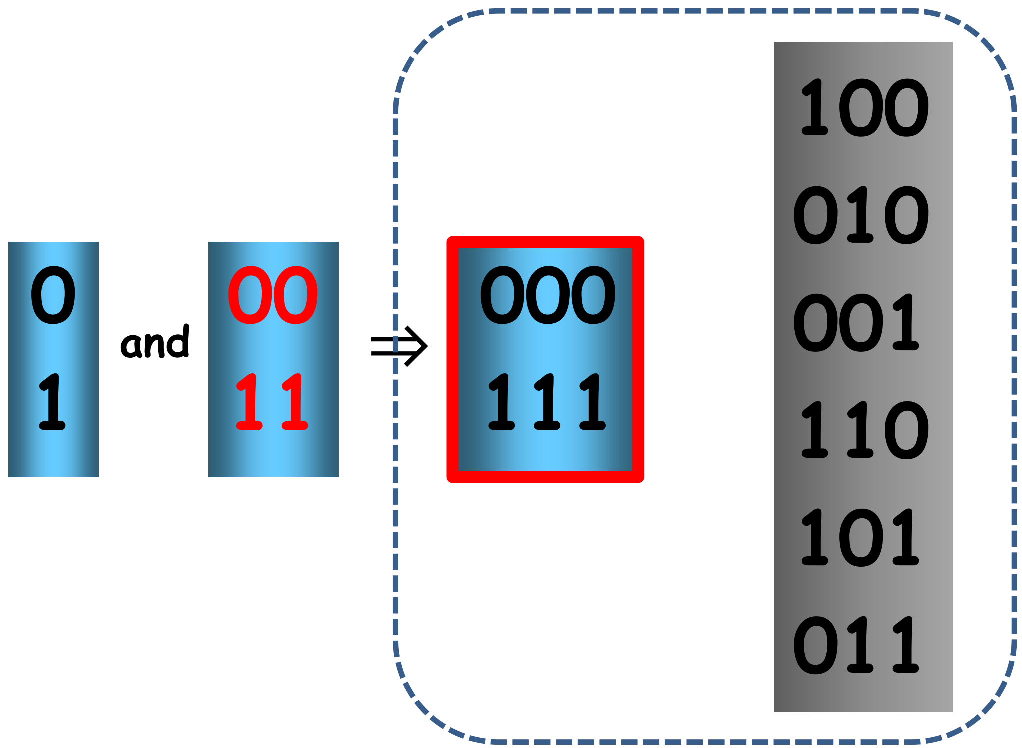
Multiplication

·	0	1
0	**0**	**0**
1	**0**	**1**

- F_2 = Finite Field of size 2
  = GF(2) = binary field
- **V=Set of all the binary n-tuples (size=2^n)**
  - A binary block code C of length n is a subset of V.
- **Error Detection:**
  - **When you receive a word not in C.**
- **Error Correction:**
  - **The most likely (unique) word in C is a decoded word.**
- **Undetected Error:**
  - **Changes a word in C into a different word in C.**

Three-time-repeat code

- Length = n= 3
- **minimum distance = 3** = minimum weight

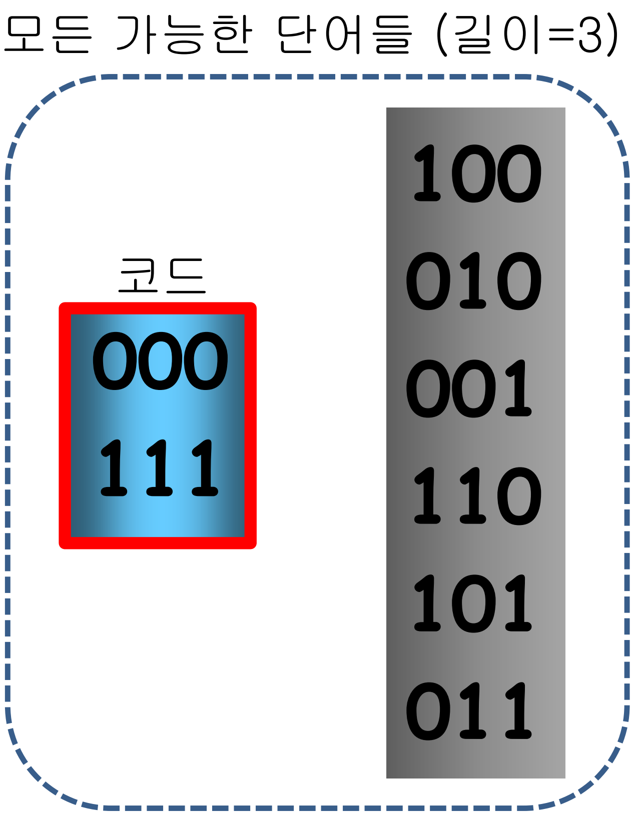


Two Views:

- These are all the possible 1-tuples and **two parity bits** repeating it.
- These are two 3-tuples with distance 3, including 000, out of all possible binary 3-tuples (there are 8 of them).

오류 검출의 원리

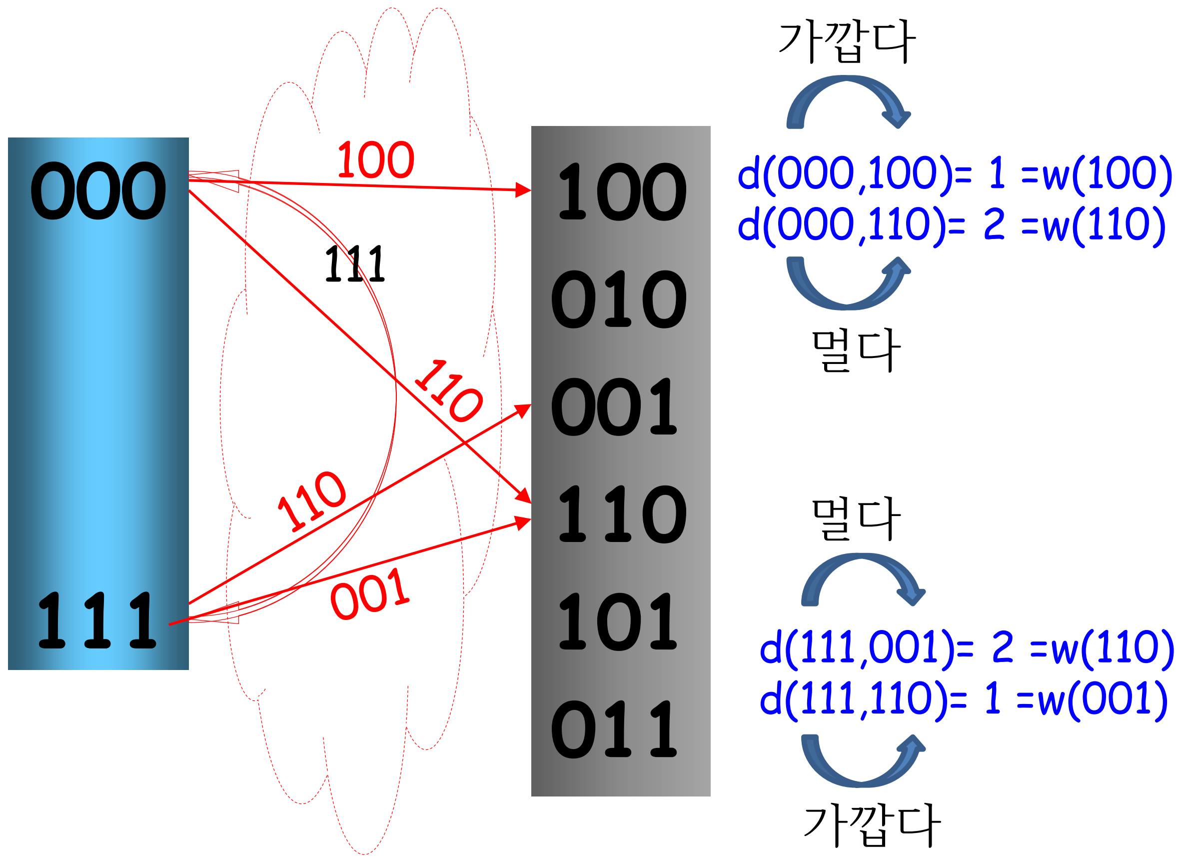
모든 가능한 단어들 중
일부만 사용하기 때문에
오류 검출이 가능하다.



오류의 해석

오류는
한 (송신)단어를 다른 (수신)단어로 바꾼다.

오류가 크다는 뜻은
더 멀리 있는 단어로 바꾼다는 뜻이다.



오류가 하는 짓

멀리 떨어진 단어들

서로 멀리 떨어진 단어들을 사용한다면
좀 더 큰 오류까지도 검출 가능하다.

C

000
111

100
010
001
110
101
011

모든 1비트 오류와
모든 2비트 오류를 검출.
그러나
3비트 오류는 검출 불가능

모든 1비트 오류를 검출함.
그러나
어떤 종류의 2비트 오류는
검출불가능

B

000
101

100
010
001
110
111
011

A

000
010

100
111
001
110
101
011

어떤 종류의 1비트 오류는
검출불가능

Single Parity Code of Length 4

000		0		0000	1000
110		0		1100	0100
011		0		0110	0010
001		1		0011	0001
100	and	1	⇒	1001	1110
101		0		1010	1101
010		1		0101	1011
111		1		1111	0111

OBSERVE:
minimum distance = 2
= minimum weight

Two Views:

- These are all the 3-tuples with some 1 bit (called **parity bit**) adjoined so that the number of 1's is even.
- These are the eight 4-tuples of even weight out of all the 4-tuples (16 of them).

Single Parity Code (Overall Parity)

Given a message (information) sequence of length *k*

$$\underline{m} = (m_1, m_2, m_3, \cdots, m_k)$$

we calculate

$$p = m_1 + m_2 + \cdots m_k \pmod 2$$

and form a codeword of length *k*+1 as

$$\underline{c} = (m_1, m_2, m_3, \cdots, m_k, p)$$

- Note that *p*=1 if $\underline{m}$ has odd number of 1's, and *p*=0 if $\underline{m}$ has even number of 1's.
- A single error can be detected if the number of 1's in the received word is not even. In fact, any odd number of error bits can be detected.
- Note that, since we compute **mod 2**, we have

$$p = m_1 + m_2 + \cdots m_k \Leftrightarrow p + m_1 + m_2 + \cdots m_k = 0$$

ML decoding의 개념

- Suppose we have **rx=1000.**
- Which one is **more likely** to happen in the channel?

tx		error		rx
0000		1000		1000
1100		0100		1000
0110		1110		1000
0011	+	1011	=	1000
1001		0001		1000
1010		0010		1000
0101		1101		1000
1111		0111		1000

These 4 possibilities have the same **likelihood**!!

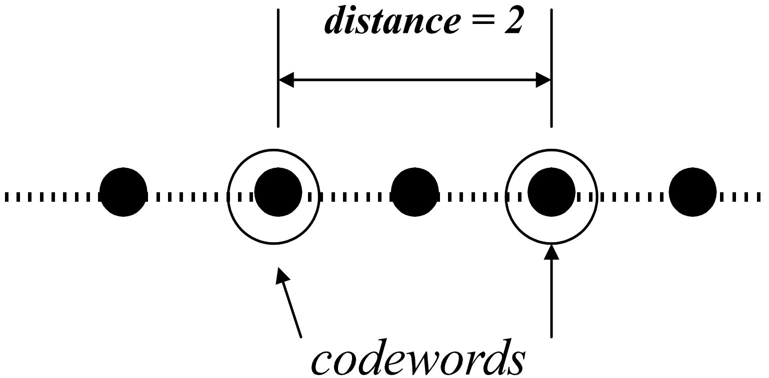
Thus, errors detected, but the ML choice is not unique!!

What is the main difference with three-time-repeat code?

Geometric View

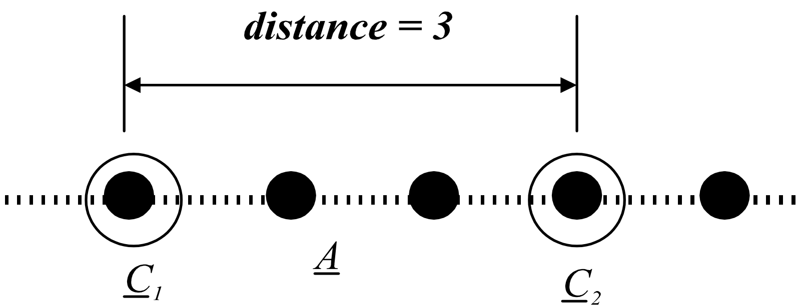
- A word is a binary n-tuple.
- V has a total of 2^n words. (16 when n=4)
- Each pair of these words has a distance, which can be 1, 2, 3, or 4.
- C is a subset of V.
  - Single parity code of length 4 has 8 codewords.
  - Each pair of codewords has a distance at least 2, which is the minimum distance of the code.

Error-Detecting and Error-Correcting



No single error can change (transform) a codeword into another codeword.

However, certain (but not all) double errors can change a codeword into another codeword.



In order to correct a single error, we must have $d \geq 3$.

Under the minimum distance decoding, the received word $\underline{A}$ is decoded as $\underline{c}_1$ since it is closer to $\underline{c}_1$ than to $\underline{c}_2$.

오류정정의 기본법칙

- **detect (검출)하지 못한 error는 correction(수정/정정)하지 못한다.**
- detect (검출)한 error는 (암튼?) 수정할 수 있다.
  - 올바른 수정인가?
  - 잘못된 수정인가?

- 올바른 수정이 가능한 경우는
  - **① most likely codeword가 unique할 때** (코드 설계로 가능하게 만들 수 있다)
  - **② 채널에서 발생할 오류비트 수가 제한되어 있을 때** (???)

- 한번의 **decoding**에서 실제로 일어나는 상황은 아무도 모른다.
  - 올바른 수정이 될 확률을 구할 수 있을 뿐이다.
  - 매우 작은 확률이지만 많은 오류가 발생할 수 있다.
  - **최적의 decoding**수단을 **반복**한다면 궁극적으로 **decoding error를 최소화**할 수 있다.

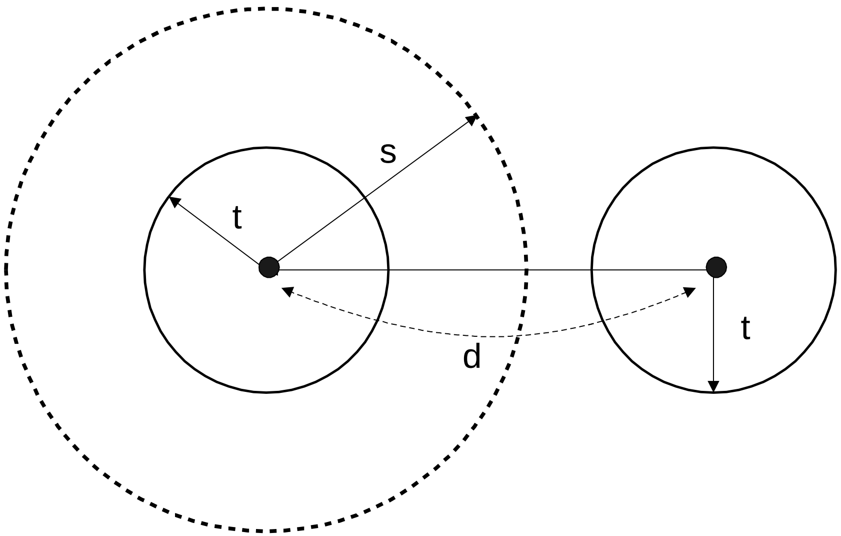
Simultaneous correction and detection

To detect s errors and correct t (≤s) errors **simultaneously**, we must have

$$d_{min} = d \geq t + s + 1$$

For error correction only, thus, we must have

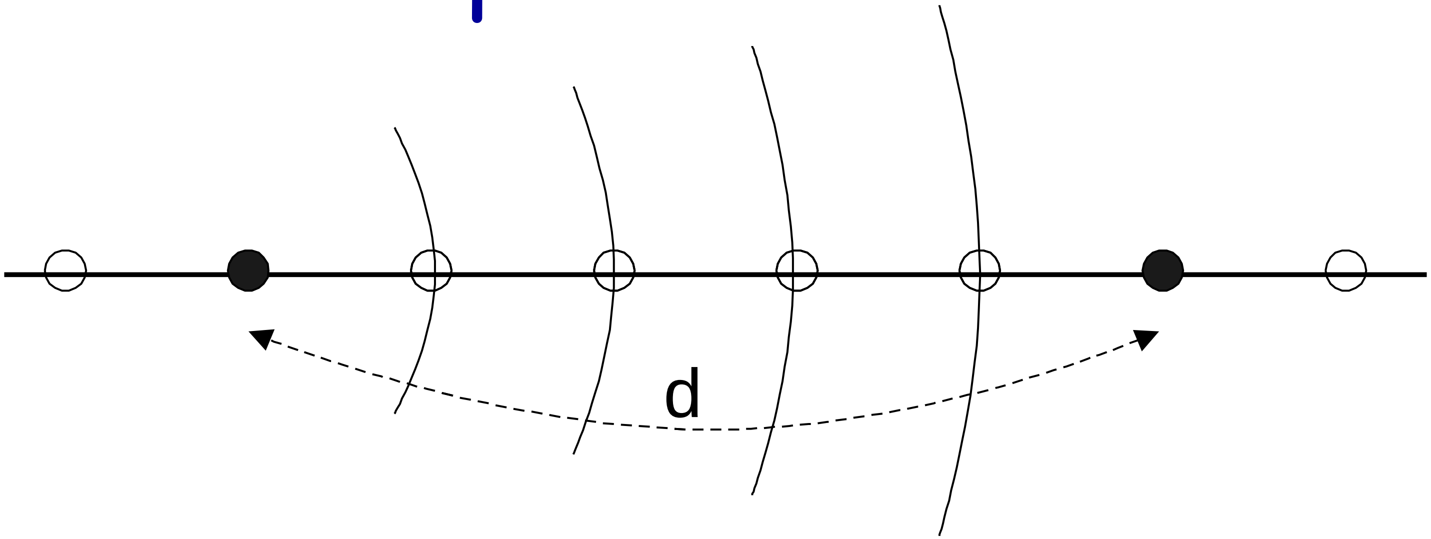
$$d_{min} = d \geq 2t + 1$$



or

$$t \leq \left\lfloor \frac{d-1}{2} \right\rfloor$$

Example: When d=5



When d=5, three schemes are possible:

4-error-detecting code

- ✓ no correction.

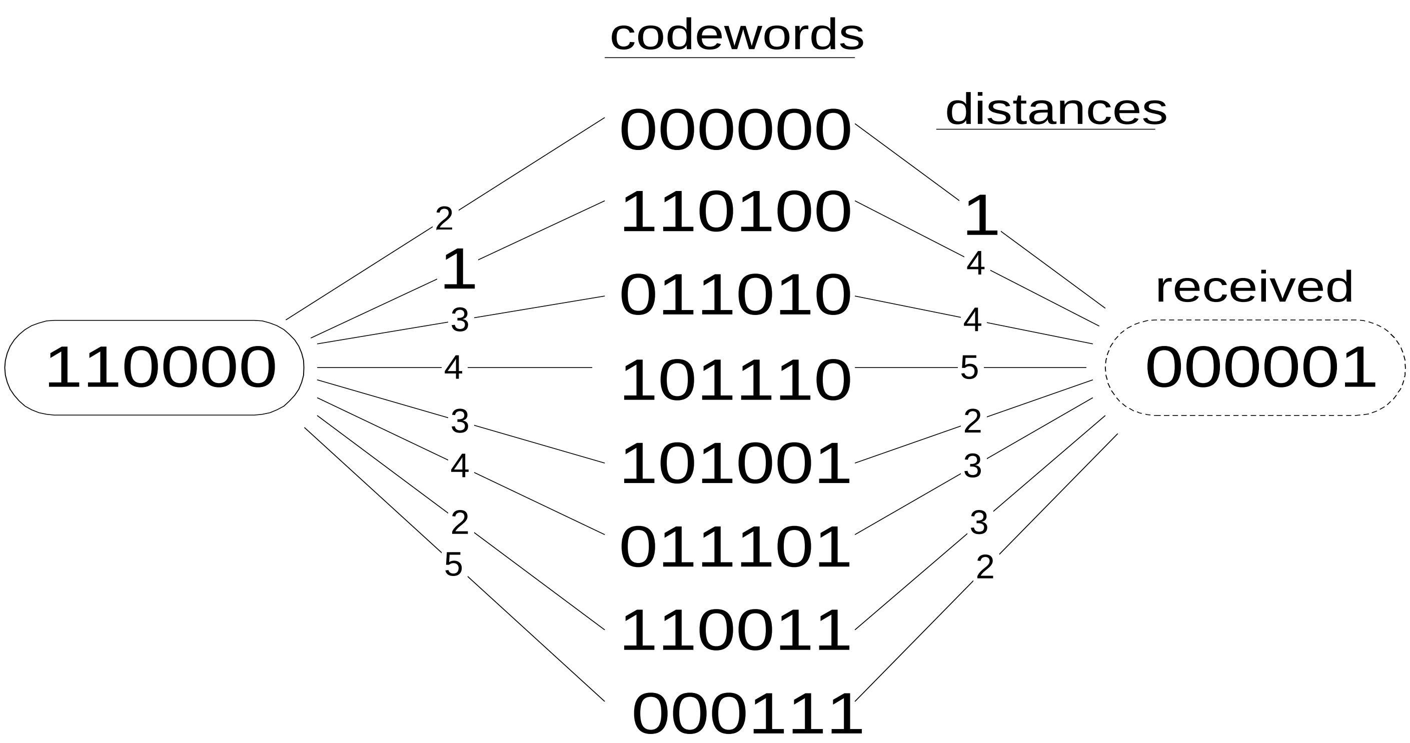
3-error-detecting and 1-error-correcting code

- ✓ Observe that 4-errors are not detected.
- ✓ It is regarded as 1-error and corrected.

2-error-detecting and 2-error-correcting code

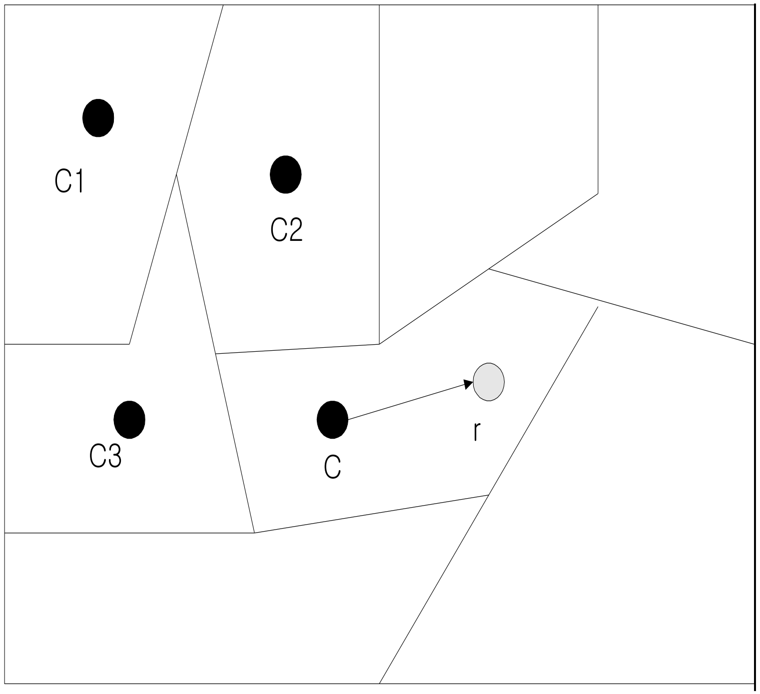
- ✓ It is usually called as "2-error-correcting code".

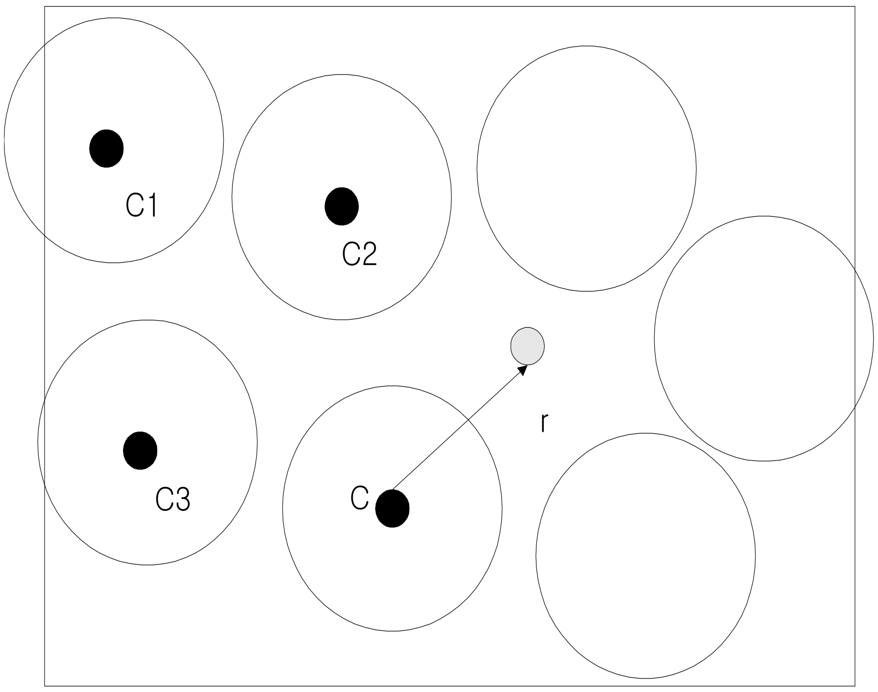
Minimum Distance Decoding



Minimum distance decoding is maximum likelihood (ML) decoding
under some reasonable condition
(We will later prove this over BSC and with hard-decision decoding)

Complete/Incomplete Decoding





연습문제

1. 다음 두 코드워드의 해밍거리는? 001100, 010101.

2. C_1 = { 000, 011, 110, 101 }을 생각하자. 이 코드의 최소거리, 오류검출능력, 오류정정능력은 얼마인가?

3. C_2 = {00000, 11111}을 생각하자. 이 코드의 최소거리, 오류검출능력, 오류정정능력은 얼마인가?