Elements of Coding Theory (Based on Grimaldi's Discrete Mathematics)

1 Introduction

Coding theory explores how to transmit data reliably across noisy communication channels. This theory was inspired by the fundamental paper of Claude Shannon (1948), Marcel Golay (1949), and Richard Hamming (1950). It uses concepts from algebraic structures, probability, and combinatorics.

1.1 Message

A **message** is defined as a finite sequence of characters from a finite alphabet.

1.2 Word

In binary coding theory, the alphabet is usually taken to be the binary set:

$$B = \{0, 1\}.$$

Every character or symbol that we want to transmit is now represented as a sequence of m elements from B. This binary string is called a **word**. Thus, our basic unit of information, called a word, is a sequence of m $0's$ and $1's$.

1.3 Group Structure of Binary Words

The set $B = \{0, 1\}$ is a group under addition modulo 2, denoted by $\oplus_2$. This group is denoted by $(\mathbb{Z}_2, \oplus_2)$.

For a fixed integer m, the Cartesian product

$$\mathbb{Z}_2^m = \mathbb{Z}_2 \times \mathbb{Z}_2 \times \cdots \times \mathbb{Z}_2 \quad (m \text{ times})$$

forms a group under component-wise addition modulo 2, denoted by $\oplus$:

$$(x_1, x_2, \ldots, x_m) \oplus (y_1, y_2, \ldots, y_m) = (x_1 \oplus_2 y_1, x_2 \oplus_2 y_2, \ldots, x_m \oplus_2 y_m).$$

1.4 Notation

A element $(b_1, b_2, b_3, \ldots, b_m)$ in $\mathbb{Z}_2^m$ is denoted by $b_1 b_2 b_3 \ldots b_m$. For example, $(1, 0, 1, 1, 0) \in \mathbb{Z}_2^5$ is simply denoted by 10110.

1.5 Properties of $\mathbb{Z}_2^m$:

- Identity element: $e = (0, 0, \ldots, 0)$
- Every element is its own inverse.
- Number of elements: 2^m

For instance, inverse of 10110 in $\mathbb{Z}_2^5$ is 10110. Since $10110 \oplus 10110 = 00000$

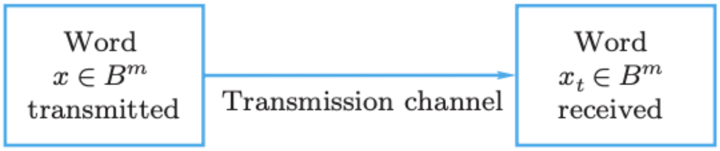
1.6 Transmission and Noise

In digital communication, information is transmitted from a sender to a receiver through a transmission channel. This channel is essentially the medium or pathway that carries the signals. However, real-world channels are imperfect and introduce noise, which can corrupt transmitted data.

When a word $x \in \mathbb{Z}_2^m$ is transmitted over a channel, it may be received as $x_t \in \mathbb{Z}_2^m$. Due to **noise** (e.g., electrical faults, interference), some bits may change, resulting in:

$$x_t \neq x.$$

This situation is illustrated below:

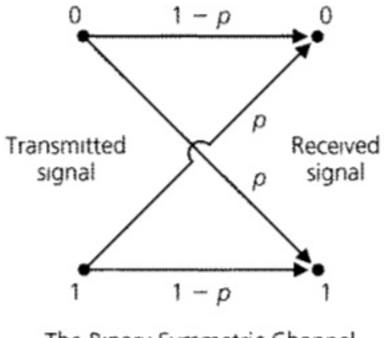


Hence, we want to develop a technique to help us detect and perhaps even correct transmission errors. However, we can only improve the chance of correct transmission; there are no guarantees.

2 Binary Symmetric Channel (BSC)

The Binary Symmetric Channel (BSC) is a fundamental and widely used theoretical model that captures the essential characteristic of errors in a simple yet powerful way. A **Binary Symmetric Channel** is a model where:

- Each bit transmitted (0 or 1) has a fixed probability p of being received incorrectly.
- Probability of correct transmission is $1 - p$.
- The channel is **symmetric** because the error probability is the same for 0 and 1.



The Binary Symmetric Channel

2.1 Probability of Errors

Consider the following example:
Let

$$c = 10110 \in \mathbb{Z}_2^5$$

be the **transmitted code**, and

$$r = 00110 \in \mathbb{Z}_2^5$$

be the **received code**.

Suppose

$$e = 10000$$

We observe that:

$$c \oplus e = r \quad \text{or equivalently} \quad r \oplus c = e$$

Here, e is called the **error pattern**.

- The positions of 1's in e indicate the locations of errors in the received code.
- The number of 1's in e corresponds to the **number of bit errors** during transmission.

General Case

Let

$$c \in \mathbb{Z}_2^n$$

be a transmitted codeword, and

$$r = c \oplus e$$

be the received codeword, where:

- e is the error vector.
- 1's in e represent bit positions where errors occurred in r.
- Number of bit errors in transmission is number of 1's in e.

Theorem 2.1. *Let $c \in \mathbb{Z}_2^n$ be the transmitted codeword through BSC with probability p of incorrect transmission.*

- *The probability of receiving $r = c + e$ (with e having k ones):*

$$P(r = c \oplus e) = p^k(1-p)^{n-k}$$

- *The probability that **exactly** k errors are made in the transmission:*

$$P_k = \binom{n}{k} p^k(1-p)^{n-k}$$

Example 2.2. Let $c = 10110$, $p = 0.05$

$$\begin{aligned} \text{Probability of no error: } & (0.95)^5 \approx 0.77 \\ \text{Probability of 1-bit error: } & \binom{5}{1}(0.05)(0.95)^4 \approx 0.204 \\ \text{Probability of 2-bit error: } & \binom{5}{2}(0.05)^2(0.95)^3 \approx 0.021 \end{aligned}$$

Example 2.3. Let C be a set of codewords, where $C \subseteq \mathbb{Z}_2^n$. In each of the following cases, two of the three values are given: the codeword c, the received word r, and the error pattern e. Using the relation

$$r = c \oplus e,$$

determine the missing term.

(a) $c = 1010110$, $r = 1011111$.

(b) $c = 1010110$, $e = 0101101$.

(c) Given: $e = 0101111$, $r = 0000111$.

Answer (a):

$$e = r \oplus c = 1011111 \oplus 1010110 = \boxed{0001001}$$

Answer (b):

$$r = c \oplus e = 1010110 \oplus 0101101 = \boxed{1111011}$$

Answer (c):

$$c = r \oplus e = 0000111 \oplus 0101111 = \boxed{0101000}$$

Example 2.4. A binary symmetric channel has probability $p = 0.05$ of incorrect transmission per bit. Let the codeword

$$c = 011011101$$

be transmitted. What is the probability of the following events?

(a) We receive $r = 011111101$.

(b) We receive $r = 111011100$.

(c) A single error occurs.

(d) A double error occurs.

(e) A triple error occurs.

Answer Let $n = 9$, so the probability of correct transmission per bit is $1 - p = 0.95$.
(a): r differs from c in exactly one bit.

$$P = (0.05)^1 \cdot (0.95)^8 \approx \boxed{0.315}$$

(b): This differs from c in 3 bits (positions 1, 2, and 9).

$$P = (0.05)^3 \cdot (0.95)^6 \approx \boxed{0.00029}$$

(c): There are $\binom{9}{1} = 9$ such patterns.

$$P = 9 \cdot (0.05)^1 \cdot (0.95)^8 \approx \boxed{0.315}$$

(d):

$$P = \binom{9}{2} \cdot (0.05)^2 \cdot (0.95)^7 = 36 \cdot 0.0025 \cdot 0.698 \approx \boxed{0.0629}$$

(e):

$$P = \binom{9}{3} \cdot (0.05)^3 \cdot (0.95)^6 = 84 \cdot 0.000125 \cdot 0.735 \approx \boxed{0.0077}$$

3 Encoding and Decoding

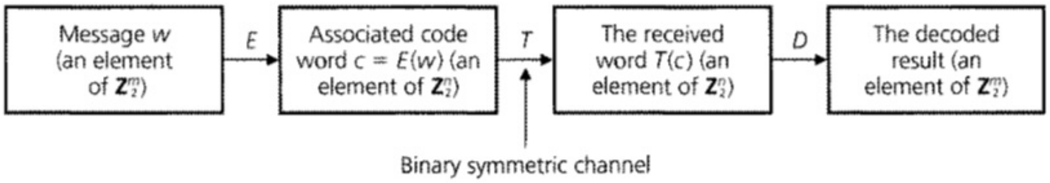
The basic task in the transmission of information is to reduce the probability of receiving a word that differs from the word that was sent. This is done as follows:

- **Encoding function:** Let $m, n \in \mathbb{Z}^+$ with $n > m$. Let $W \subseteq \mathbb{Z}_2^m$ be the set of messages, and $C \subseteq \mathbb{Z}_2^n$ be the set of code words. Consider and define a one-to-one encoding function:
$$E : W \to \mathbb{Z}_2^n \text{ with } E(w) = c \in C.$$
This function maps each m-bit word in W into an n-bit codeword (obtained by appending $n - m$ extra bits), introducing redundancy to detect or correct errors.

- Upon transmission, c is received as $T(c)$ may be different at different transmission times.

- **Decoding function:** Upon receiving $T(c)$, we apply decoding function
$$D : \mathbb{Z}_2^n \to, W$$
and hope to receive original message w.

- Ideally, $D \circ T \circ E$ should be the identity on W, but this is not always possible due to errors. The goal is to design E and D so that the probability of correctly decoding is high.


Message w (an element of $\mathbf{Z}_2^m$)
E
Associated code word $c = E(w)$ (an element of $\mathbf{Z}_2^n$)
T
The received word $T(c)$ (an element of $\mathbf{Z}_2^n$)
D
The decoded result (an element of $\mathbf{Z}_2^m$)
Binary symmetric channel

- **Efficiency (Rate):** The ratio
$$\frac{m}{n}$$
measures the efficiency of the scheme and is called the rate of the code.

- **Block code** The ordered pair (n, m) is called block code of E and D.

We illustrate the above idea using the following codes:

- Parity-Check code
- Triple repetition code

3.1 Parity-Check Code

Definition 3.1 (Weight of $x \in \mathbb{Z}_2^m$)**.** For any $x \in \mathbb{Z}_2^m$, the number of $1's$ is x is called the weight of x and is denoted by $\mathrm{wt}(x)$. For example, if $x = 0111011$, then $\mathrm{wt}(x) = 5$.

3.1.1 Even parity check:

The $(m+1, m)$ block code is called even parity check code if for any $w = w_1w_2w_3 \ldots w_m \in \mathbb{Z}_2^m$, define
$$E : \mathbb{Z}_2^m \to \mathbb{Z}_2^{m+1} \text{ by } E(w) = w_1w_2w_3 \ldots w_m w_{m+1}$$
where
$$w_{m+1} = \begin{cases} 0 & \text{if } \mathrm{wt}(w) \text{ is even,} \\ 1 & \text{if } \mathrm{wt}(w) \text{ is odd.} \end{cases}$$

- Observe that w_{m+1} is zero if and only if the number of 1's in w is even. Hence, every code word $E(w)$ has even weight.

- A single-bit error in the transmission of a code word will change its weight to odd, which can be detected.

- Similarly, any odd number of bit errors can be detected by this code.
- The scheme does **not** detect errors if an even number of bits are altered
- If an error is detected, the codeword is retransmitted and continue the process until the correct code is received or received code has an even number of $1's$.

Example: Suppose $w = 11010110$. Then the codeword is

$$c = E(w) = 110101101.$$

If the transmission channel sends this as $T(c) = 100101101$, we compute the weight $\mathrm{wt}(T(c)) = 5$, which is odd. Therefore, we conclude that an error (or an odd number of errors) has occurred.

3.1.2 odd parity check:

The $(m+1, m)$ block code is called odd parity check code if for any $w = w_1w_2w_3\dots w_m \in \mathbb{Z}_2^m$, define

$$E : \mathbb{Z}_2^m \to \mathbb{Z}_2^{m+1} \text{ by } E(w) = w_1w_2w_3\dots w_mw_{m+1}$$

where

$$w_{m+1} = \begin{cases} 0 & \text{if wt}(w) \text{ is odd,} \\ 1 & \text{if wt}(w) \text{ is even.} \end{cases}$$

Example 3.2. For the given codes, apply (i) the even parity check, (ii) the odd parity check codes.

(a) $w = 1011$
(b) $w = 1010$
(c) $w = 101101$
(d) $w = 110111$

Answer:

Label	Original Code (w)	Even Parity	Odd Parity
(a)	1011	10111	10110
(b)	1010	10100	10101
(c)	101101	1011010	1011011
(d)	110111	1101111	1101110

Example 3.3. Apply the **odd parity check** to the code 101101. Determine if the received code is correct and whether the error is detected.

(a) $r = 1011011$
(b) $r = 1111011$
(c) $r = 1001111$

Answer:

$$\text{Original code: } w = 101101 \quad \Rightarrow \quad \text{Transmitted (odd parity): } c = 1011011$$

Case	**Received Code (r)**	**Error in r**	**Error Detected?**
(i)	1011011	No	—
(ii)	1111011	Yes	Yes
(iii)	1001111	Yes	No (even number of 1's)

Example 3.4. Apply the **even parity check** to the code 110111. Determine if the received code is correct and whether the error is detected.

(a) $r = 1101111$
(b) $r = 1100111$
(c) $r = 1001011$

Answer:

$$\text{Original code: } w = 110111 \quad \Rightarrow \quad \text{Transmitted (even parity): } c = 1101111$$

Case	**Received Code (r)**	**Error in r**	**Error Detected?**
(i)	1101111	No	—
(ii)	1100111	Yes	Yes
(iii)	1001011	Yes	No (even number of 1's)

Probability of Correct Detection

Let p be the bit error probability. The probability of sending the code 110101101 and making at most one error is:

$$(1-p)^9 + \binom{9}{1} p(1-p)^8$$

For $p = 0.001$, this gives:

$$(0.999)^9 + \binom{9}{1}(0.001)(0.999)^8 \approx 0.99996417$$

Improved Reliability: If an error is detected, the codeword is retransmitted and continue the process until the received word has an even number of $1's$.

Trade-off:

Adding extra bits reduces the chance of undetected error but also reduces the rate of the code

Example 3.5. Suppose 160 bits are sent in successive strings of length 8. Find the probability of receiving the correct message:

(i) without any coding scheme,

(ii) with the parity-check method.

Assume $p = 0.001$.

Solution:

(i) Since there are 160 bits and each string is of length 8, there are $160/8 = 20$ strings. Therefore, Probability of receiving the correct message without any coding scheme

$$P(\text{correct message}) = (0.999)^{160} \approx 0.852076$$

(ii) Probability with the parity-check method: In parity-check, each 8-bit string becomes a 9-bit code word with error detection.

A 9-bit word is correct if either:

- all 9 bits are received correctly: $(0.999)^9$,
- exactly 1 bit is incorrect and it is detected and corrected: $9 \cdot (0.001) \cdot (0.999)^8$.

So the total probability per word:

$$P(\text{correct}) = (0.999)^9 + 9(0.001)(0.999)^8 \approx 0.999964$$

For 20 such 9-bit strings:

$$P(\text{correct message}) = (0.999964)^{20} \approx 0.9999280$$

Example 3.6. Suppose the message $w = 110101$ is sent. Find the probability of receiving the correct message:

(i) without any coding scheme,

(ii) with the parity-check method.

Assume $p = 0.05$.

Solution:

(i) Probability of receiving the correct code without any coding method:

$$P(\text{correct}) = (0.95)^6 = 0.7351$$

(ii) In the parity-check method, we append a bit to make the number of 1's even. For $w = 110101$, number of 1's is 4 (even), so we append a 0:

$$\text{New code} = 1101010$$

The code is 7 bits long. The probability of correct message:

$$P = (0.95)^7 + 7 \cdot (0.05)(0.95)^6 = 0.9552$$

3.2 Triple repetition code

The $(3m, m)$ block code is called triple repetition code if for any $w = w_1 w_2 w_3 \dots w_m \in \mathbb{Z}_2^m$,

- Define
$$E : \mathbb{Z}_2^m \to \mathbb{Z}_2^{3m} \text{ by } E(w) = w_1 w_2 \dots w_m w_1 w_2 \dots w_m w_1 w_2 \dots w_m$$
- The decoding function $D : \mathbb{Z}_2^{3m} \to \mathbb{Z}_2^m$ uses the **majority rule**.

For example, if $w = 10110111$, then the encoded word is:

$$c = E(w) = 101101111011101110111011.$$

If we receive:

$$T(c) = 1010011100110111110110110,$$

we decode each group of three bits by majority:

- First triplet: positions 1, 9, 17 $= 1, 1, 0 \Rightarrow 1$
- Second triplet: positions 2, 10, 18 $= 0, 0, 0 \Rightarrow 0$
- Third triplet: positions 3, 11, 19 $= 1, 1, 1 \Rightarrow 1$
- Fourth triplet: $0, 0, 0 \Rightarrow 0$
- Fifth triplet: $0, 1, 1 \Rightarrow 1$
- Sixth triplet: $1, 1, 1 \Rightarrow 1$
- Seventh triplet: $1, 1, 1 \Rightarrow 1$
- Eighth triplet: $0, 1, 0 \Rightarrow 0$

So we recover the correct message:

$$10110111$$

Error Analysis

Even though multiple bit errors occurred, as long as at most one error happens in each group of three, the original bit can be correctly recovered. However, if two or more errors occur in the same group (e.g., at positions 2, 10, and 18), decoding may fail.

Probability of Correct Decoding

Let $p = 0.001$ be the probability of a bit error. Then, the probability of correctly decoding a single bit using triple repetition is:

$$\begin{aligned} P_{\text{correct}} &= (1-p)^3 + \binom{3}{1}(p)(1-p)^2 \\ &= (0.999)^3 + \binom{3}{1}(0.001)(0.999)^2 \approx 0.99999700. \end{aligned}$$

Therefore, the probability of correctly decoding an 8-bit message is:

$$(0.99999700)^8 \approx 0.99997600.$$

This is slightly better than the parity-check code, but note that it requires transmitting 24 bits instead of 8, so the transmission rate is $\frac{1}{3}$. While transmission time increases, we avoid retransmissions and gain the ability to *correct* errors—not just detect them.

Conclusion

- The triple repetition code is effective at correcting single-bit errors with high probability.
- The rate of code is $\frac{1}{3}$. That is, lower transmission efficiency.
- Retransmission is avoided.

Example 3.7. Let $E : \mathbb{Z}_2^3 \to \mathbb{Z}_2^9$ be the encoding function for the $(9, 3)$ triple repetition code.

(a) Let $D : \mathbb{Z}_2^9 \to \mathbb{Z}_2^3$ be the corresponding decoding function. Apply D to decode the received words:

$$\begin{aligned} &\text{(i)} \quad r = 111101100 \\ &\text{(ii)} \quad r = 010011111 \end{aligned}$$

(b) Find three different received words r for which $D(r) = 000$.

(c) For each $w \in \mathbb{Z}_2^3$, what is $|D^{-1}(w)|$?

Solution (a):

(i)

$$\begin{aligned} &\text{1st, 4th, and 7th positions: } 1, 1, 1 \Rightarrow \text{decoded bit: } 1 \\ &\text{2nd, 5th, and 8th positions: } 1, 0, 0 \Rightarrow \text{decoded bit: } 0 \\ &\text{3rd, 6th, and 9th positions: } 1, 1, 0 \Rightarrow \text{decoded bit: } 1 \\ &\Rightarrow \text{Final decoded word: } 101 \end{aligned}$$

(ii)

$$\begin{aligned}&\text{1st, 4th, and 7th positions: } 0,0,1 \Rightarrow \text{decoded bit: } 0\\&\text{2nd, 5th, and 8th positions: } 1,1,1 \Rightarrow \text{decoded bit: } 1\\&\text{3rd, 6th, and 9th positions: } 0,1,1 \Rightarrow \text{decoded bit: } 1\\&\Rightarrow \text{Final decoded word: } 011\end{aligned}$$

Solution (b):

For $D(r) = 000$, in each of the positions:

- 1st, 4th, and 7th: majority should be 0
- 2nd, 5th, and 8th: majority should be 0
- 3rd, 6th, and 9th: majority should be 0

Possible received words:

$$000000000, \quad 100000000, \quad 010000000$$

Solution (c):

$$\begin{aligned}&D : \mathbb{Z}_2^9 \to \mathbb{Z}_2^3\\&|\mathbb{Z}_2^9| = 512, \quad |\mathbb{Z}_2^3| = 8\\&\Rightarrow \left|D^{-1}(w)\right| = \frac{|\mathbb{Z}_2^9|}{|\mathbb{Z}_2^3|} = \frac{512}{8} = 64\end{aligned}$$

Example 3.8. The $(5m, m)$ five-times repetition code has encoding function $E : \mathbb{Z}_2^m \to \mathbb{Z}_2^{5m}$, where $E(w) = wwwww$. Decoding with $D : \mathbb{Z}_2^{5m} \to \mathbb{Z}_2^m$ is accomplished by the majority rule (i.e., we are able to correct single and double errors made in transmission).

(a) With $p = 0.05$, what is the probability for the transmission and correct decoding of the signal 0?

(b) Answer part (a) for the message 110 in place of the signal 0.

(c) For $m = 2$, decode the received word $r = 0111001001$.

(d) If $m = 2$, find three received words r where $D(r) = 00$.

(e) For $m = 2$ and $D : \mathbb{Z}_2^{10} \to \mathbb{Z}_2^2$, what is $|D^{-1}(w)|$ for each $w \in \mathbb{Z}_2^2$?

Solution:

(a) Here $w = 0, \quad \Rightarrow E : \mathbb{Z}_2^1 \to \mathbb{Z}_2^5, \quad \Rightarrow E(w) = 00000$
then $D : \mathbb{Z}_2^5 \to \mathbb{Z}_2^1$.

With $p = 0.05$, the probability of correctly decoding a single bit is (by majority rule we can make at most 2 errors out of 5 positions):

$$(0.95)^5 + \binom{5}{1}(0.05)(0.95)^4 + \binom{5}{2}(0.05)^2(0.95)^3 = 0.998842$$

So the probability of receiving and correctly decoding the one-bit message is:

$$(0.998842)^1 = 0.998842$$

(b) Here $w = 110, \quad E : \mathbb{Z}_2^3 \to \mathbb{Z}_2^{15}, \quad \Rightarrow T(w) = 111001110111000$
then $D : \mathbb{Z}_2^{15} \to \mathbb{Z}_2^3$

With $p = 0.05$, the probability of correctly decoding a single bit is:

$$(0.95)^5 + \binom{5}{1}(0.05)(0.95)^4 + \binom{5}{2}(0.05)^2(0.95)^3 = 0.998842$$

So the probability of receiving and correctly decoding the three-bit message is:

$$(0.998842)^3 = 0.996530$$

(c) With $m = 2$, received word $r = 0111001001$.

The 1st, 3rd, 5th, 7th and 9th positions have $0, 1, 0, 1, 0$
By majority rule the correct message is 0.

The 2nd, 4th, 6th, 8th and 10th positions have $1, 1, 0, 0, 1$
By majority rule the correct message is 1.

The decoded message is 01.

(d) $m = 2$, and $D(r) = 00$, the received word r is:

$$0000000000, \quad 0000000011, \quad 1100000000$$

(e) $m = 2$, $D : \mathbb{Z}_2^{10} \to \mathbb{Z}_2^2$

For each $w \in \mathbb{Z}_2^2$,

$$|D^{-1}(w)| = \frac{|\mathbb{Z}_2^{10}|}{|\mathbb{Z}_2^2|} = \frac{2^{10}}{2^2} = 256$$

4 Hamming metric

Definition 4.1 (**Weight**)**.** For any $x = x_1x_2 \dots x_n \in \mathbb{Z}_2^n$, the weight of x, denoted by $\mathrm{wt}(x)$, is the number of 1's in x:

$$\mathrm{wt}(x) = \sum_{i=1}^{n} x_i.$$

Lemma 4.2. *For all $x, y \in \mathbb{Z}_2^n$, we have:*

$$\mathrm{wt}(x \oplus y) \leq \mathrm{wt}(x) \oplus \mathrm{wt}(y).$$

Proof. This follows because bitwise addition modulo 2 does not introduce any new 1's. The only case where $x_i \oplus_2 y_i = 1$ is when exactly one of x_i or y_i is 1. □

Example 4.3. We have

- $\mathrm{wt}(01001) = 2$
- $\mathrm{wt}(11101) = 4$
- $\mathrm{wt}(01001 + 11101) = \mathrm{wt}(10100) = 2$

Definition 4.4 (**Distance**)**.** The **Hamming distance** between $x, y \in \mathbb{Z}_2^n$ is defined as:

$$d(x, y) = \sum_{i=1}^{n} d(x_i, y_i), \quad \text{where} \;\; d(x_i, y_i) = \begin{cases} 0 & \text{if } x_i = y_i, \\ 1 & \text{if } x_i \neq y_i. \end{cases}$$

It counts the number of positions where $x_i \neq y_i$ for $1 \leq i \leq n$..

Example 4.5. We have

- $x = 01001$, $y = 11101 \Rightarrow d(x, y) = 2$
- $x = 10101$, $y = 10101 \Rightarrow d(x, y) = 0$

Note:

$$d(x, y) = \mathrm{wt}(x \oplus y).$$

This is because in $\mathbb{Z}_2$, $x_i \oplus_2 y_i = 1$ if and only if $x_i \neq y_i$.

Properties of $d(x, y)$:

Let $d(x, y)$ be the Hamming distance. Then for all $x, y, z \in \mathbb{Z}_2^n$:

1. $d(x, y) \geq 0$
2. $d(x, y) = 0 \Leftrightarrow x = y$
3. $d(x, y) = d(y, x)$
4. $d(x, z) \leq d(x, y) + d(y, z)$ (Triangle Inequality)

4.1 Hamming Metric

The function $d(x, y)$ defined on $\mathbb{Z}_2^n$ is a **distance function** or **metric**, and the space $(\mathbb{Z}_2^n, d)$ is called a **metric space**. It captures the idea of closeness between binary vectors.

Definition 4.6 (Sphere of Radius k). For a vector $x \in \mathbb{Z}_2^n$, the **sphere** of radius k centered at x is:

$$S(x, k) = \{y \in \mathbb{Z}_2^n : d(x, y) \leq k\}$$

Example 4.7. For $n = 3$, $x = 110 \in \mathbb{Z}_2^3$, we have:

- $S(x, 1) = \{110, 010, 100, 111\}$
- $S(x, 2) = \{110, 010, 100, 111, 000, 101, 011\}$

4.2 Major Theorem

Theorem 4.8. *Let $E : W \to C$ be an encoding function with the set of all messages $W \subseteq \mathbb{Z}_2^m$ and the set of code words $E(W) = C \subseteq \mathbb{Z}_2^n$, where $m < n$.*

1. *We can detect all transmission errors of weight $\leq k$ iff the minimum distance between code words is at least $k + 1$.*
2. *We can correct all transmission errors of weight $\leq k$ iff the minimum distance between code words is at least $2k + 1$.*

Example 4.9. Let $W = \mathbb{Z}_2^2$, and let $E : W \to \mathbb{Z}_2^6$ be defined by:

$$\begin{aligned} E(00) &= 000000, \\ E(10) &= 101010, \\ E(01) &= 010101, \\ E(11) &= 111111. \end{aligned}$$

1. Compute Hamming distances between code words and hence find minimum distance between code words.
2. Discuss error detecting and correcting capabilities.

Solution 1: Let $c_1 = E(00),\ c_2 = E(10),\ c_3 = E(01),\ c_4 = E(11)$.

We compute Hamming distances:

$$\begin{aligned} d(c_1, c_2) &= 3, \\ d(c_1, c_3) &= 3, \\ d(c_1, c_4) &= 6, \\ d(c_2, c_3) &= 6, \\ d(c_2, c_4) &= 3, \\ d(c_3, c_4) &= 3. \end{aligned}$$

The **minimum distance** between the codewords is 3.

Solution 2: By Theorem 4.8 (1), $k+1=3 \Rightarrow k=2$; and by Theorem 4.8 (2), $2k+1=3 \Rightarrow k=1$.

Hence, we can:

- Detect up to 2 errors,
- Correct up to 1 error.

Example 4.10. Let $W=\mathbb{Z}_2^2$ and $E: W \to \mathbb{Z}_2^6$ be defined by:

$$\begin{aligned} E(00) &= 000000, \\ E(10) &= 101010, \\ E(01) &= 010101, \\ E(11) &= 111111. \end{aligned}$$

1. List the elements in $S(101010,1)$ and $S(111111,1)$.
2. Decode each of the following received words using majority rule.
   (i) 110101
   (ii) 101011
   (iii) 001111
   (iv) 110000

Solution 1:

$S(101010,1)=\{x \in \mathbb{Z}_2^6 \mid d(x,101010) \leq 1\}$:

$$\{101010,\ 001010,\ 111010,\ 100010,\ 101110,\ 101000,\ 101011\}$$

$S(111111,1)=\{x \in \mathbb{Z}_2^6 \mid d(x,111111) \leq 1\}$:

$$\{111111,\ 011111,\ 101111,\ 110111,\ 111011,\ 111101,\ 111110\}$$

Solution 2:

(i) 110101

Majority at 1st, 3rd, 5th positions: 0
Majority at 2nd, 4th, 6th positions: 1
$\Rightarrow$ Decoded word is: $\boxed{01}$

(ii) 101011

Majority at 1st, 3rd, 5th positions: 1
Majority at 2nd, 4th, 6th positions: 0
$\Rightarrow$ Decoded word is: $\boxed{10}$

(iii) 001111

Majority at 1st, 3rd, 5th positions: 1
Majority at 2nd, 4th, 6th positions: 1
$\Rightarrow$ Decoded word is: $\boxed{11}$

(iv) 110000

Majority at 1st, 3rd, 5th positions: 0
Majority at 2nd, 4th, 6th positions: 0
$\Rightarrow$ Decoded word is: $\boxed{00}$

Note:

Suppose $c = 010101$ and $T(c) = r = 111101$. Then:

- r is not a codeword, so a double error can be detected.

However, if $T(c) = r_1 = 111111$ (a triple error), then:

- We incorrectly decode r_1 as the codeword 111111, assuming $c = 11$ instead of the correct message 01.

Example 4.11. Answer the following:

1. If $x \in \mathbb{Z}_2^{10}$, determine $|S(x,1)|$, $|S(x,2)|$, $|S(x,3)|$.
2. For $n, k \in \mathbb{Z}^+$ with $1 \le k \le n$, if $x \in \mathbb{Z}_2^n$, what is $|S(x,k)|$?

Solution 1:

$$\begin{aligned} S(x,1) &= \big\{y \in \mathbb{Z}_2^{10} \;\big|\; d(x,y) \le 1\big\} \\ &= \big\{y \in \mathbb{Z}_2^{10} \;\big|\; d(x,y) = 0 \text{ or } d(x,y) = 1\big\} \end{aligned}$$

For $d(x,y) = 0$, y should be same as x, there is only one such possibility.
For $d(x,y) = 1$, y should differ from x at only one position, and there are $C(10,1) = 10$ such possibilities. Therefore,

$$|S(x,1)| = 1 + 10 = 11$$

$$\begin{aligned} S(x,2) &= \big\{y \in \mathbb{Z}_2^{10} \;\big|\; d(x,y) \le 2\big\} \\ &= \big\{y \in \mathbb{Z}_2^{10} \;\big|\; d(x,y) = 0 \text{ or } d(x,y) = 1 \text{ or } d(x,y) = 2\big\} \end{aligned}$$

For $d(x,y) = 2$, y should differ from x at only two positions, and there are $C(10,2) = 45$ possibilities.

$$|S(x,2)| = 1 + 10 + 45 = 56$$

Similarly,

$$|S(x,3)| = 1 + C(10,1) + C(10,2) + C(10,3) = 176$$

Solution 2:

With $1 \leq k \leq n$, for $x \in \mathbb{Z}_2^n$,

$$|S(x,k)| = 1 + C(n,1) + C(n,2) + \cdots + C(n,k) = \sum_{r=0}^{k} C(n,r)$$

Example 4.12. Let $E : \mathbb{Z}_2^5 \to \mathbb{Z}_2^{25}$ be an encoding function where the minimum distance between code words is 9. What is the largest value of k such that we can detect errors of weight $\leq k$? If we wish to correct errors of weight $\leq t$, what is the maximum value of t?

Soln:

We can use Theorem 4.8 (1) to find k: $k + 1 = 9 \Rightarrow k = 8$

We can use Theorem 4.8 (2) to find t: $2t + 1 = 9 \Rightarrow t = 4$

Example 4.13. For each of the following encoding functions, find the minimum distance between the code words. Discuss the error-detecting and error-correcting capabilities of each code.

1. $E : \mathbb{Z}_2^2 \to \mathbb{Z}_2^5$

$$\begin{aligned} 00 &\to 00001 \\ 01 &\to 01010 \\ 10 &\to 10100 \\ 11 &\to 11111 \end{aligned}$$

2. $E : \mathbb{Z}_2^2 \to \mathbb{Z}_2^{10}$

$$\begin{aligned} 00 &\to 0000000000 \\ 01 &\to 0000011111 \\ 10 &\to 1111000000 \\ 11 &\to 1111111111 \end{aligned}$$

3. $E : \mathbb{Z}_2^3 \to \mathbb{Z}_2^6$

$$\begin{aligned} 000 &\to 000011 \\ 001 &\to 001001 \\ 010 &\to 010010 \\ 011 &\to 011100 \\ 100 &\to 100100 \\ 101 &\to 101010 \\ 110 &\to 110001 \\ 111 &\to 111000 \end{aligned}$$

4. $E : \mathbb{Z}_2^3 \to \mathbb{Z}_2^8$

$$\begin{aligned}
000 &\to 00011111\\
001 &\to 00111010\\
010 &\to 01010101\\
011 &\to 01110000\\
100 &\to 10001101\\
101 &\to 10101000\\
110 &\to 11000100\\
111 &\to 11100011
\end{aligned}$$

Solution 1:

$$d(00001, 01010) = 3, \quad d(00001, 10100) = 3, \quad d(00001, 11111) = 4$$
$$d(01010, 10100) = 4, \quad d(01010, 11111) = 3, \quad d(10100, 11111) = 3$$

Minimum distance between code words is 3.
Applying Theorem 4.8 (1): $k + 1 = 3 \Rightarrow k = 2$

The code can detect all errors of weight ≤ 2.

Applying Theorem 4.8 (2): $2k + 1 = 3 \Rightarrow k = 1$

The code can correct all errors of weight ≤ 1.

Solution 2: Let the code words be denoted as follows:

$$\begin{aligned}
r_1 &= 0000000000,\\
r_2 &= 0000011111,\\
r_3 &= 1111000000,\\
r_4 &= 1111111111
\end{aligned}$$

$$\begin{aligned}
d(r_1, r_2) &= 5,\\
d(r_1, r_3) &= 5,\\
d(r_1, r_4) &= 10,\\
d(r_2, r_3) &= 10,\\
d(r_2, r_4) &= 5,\\
d(r_3, r_4) &= 5
\end{aligned}$$

Minimum distance between code words is 5.

Applying Theorem 1: $k + 1 = 5 \Rightarrow k = 4$

The code can detect all errors of weight ≤ 4.

Applying Theorem 2: $2k+1=5 \Rightarrow k=2$

The code can correct all errors of weight ≤ 2.

Solution 3:

Let the code words be denoted as follows:

$$r_1 = 000111, \quad r_2 = 001001, \quad r_3 = 010010, \quad r_4 = 011100,$$
$$r_5 = 100100, \quad r_6 = 101010, \quad r_7 = 110001, \quad r_8 = 111000$$

	r_2	r_3	r_4	r_5	r_6	r_7	r_8
r_1	3	3	4	4	4	4	6
r_2		4	3	4	3	3	3
r_3			3	4	3	3	3
r_4				3	4	4	2
r_5					3	3	3
r_6						4	2
r_7							2
r_8							

Minimum distance between code words is 2.

Applying Theorem 1: $k+1=2 \Rightarrow k=1$

The code can detect all errors of weight ≤ 1.

Applying Theorem 2: $2k+1=2 \Rightarrow k=\frac{1}{2}$

The code can correct all errors of weight $\leq \dfrac{1}{2} \Rightarrow$ the code cannot correct any error.

Introduction

In coding theory, **generator matrices** and **parity-check matrices** help encode and decode messages using binary linear codes over $\mathbb{Z}_2$. These matrices are useful in locating the *nearest codeword* to a received word, and they play a vital role in *error detection and correction.*

Generator Matrix G

A generator matrix G is used to encode a message vector into a codeword.

Let G be a $k \times n$ matrix over $\mathbb{Z}_2$, and let $w \in \mathbb{Z}_2^k$ be a message (row vector). The encoded codeword is:

$$E(w) = wG \in \mathbb{Z}_2^n$$

Example

Let

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

This is a 3×6 matrix. The first 3 columns form the identity matrix I_3, and the last 3 columns form matrix A. Thus,

$$G = [I_3 \mid A]$$

This structure ensures **systematic encoding**, meaning the original message appears directly in the codeword.

Codewords

Each message $w \in \mathbb{Z}_2^3$ produces a codeword in $\mathbb{Z}_2^6$.

- $E(110) = (110)G = 110101$
- $E(010) = (010)G = 010011$

For

$$W = \{000, 100, 010, 001, 110, 101, 011, 111\}$$

The set of codewords is:

$$C = \{000000, 100110, 010011, 001101, 110101, 101011, 011110, 111000\}$$

Parity-Check Matrix H

The parity-check matrix H is used to verify if a received word is a valid codeword.

Given A as the right part of G, define:

$$H = [A^T \mid I_{n-k}]$$

In our example:

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \quad \Rightarrow \quad H = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Syndrome Decoding

Given a received word $r \in \mathbb{Z}_2^6$, compute the **syndrome**:

$$\text{syndrome}(r) = H \cdot r^T$$

- If $H \cdot r^T = 0$, then r is a valid codeword.
- If $H \cdot r^T \neq 0$, then r has errors.

Example (Continued...)

Suppose $r = 110110$. Then,

$$H \cdot r^T = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

This nonzero result indicates an error. Based on the syndrome, we identify and correct the bit in error.

Error Detection and Correction

Let d be the **minimum Hamming distance** between any two codewords.

- The code can detect all errors of weight $\leq k$ if:

$$d > k$$

- The code can correct all errors of weight $\leq t$ if:

$$d \geq 2t + 1$$

In Our Example

- Minimum distance, $d = 3$
- Can detect all errors of weight ≤ 2
- Can correct all errors of weight ≤ 1
- $H \cdot r^T = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ is same as second column of H. Therefore, it has single error and error is in second component.
- Thus, transmitted code is 100110 and message is 100.

Multiple Errors

Suppose we receive $r = 000111$. Then,

$$H \cdot r^T = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \text{which is not a column of } H.$$

Yet

- $H \cdot r^T$ can be obtained as the sum of two columns of H.
- If $H \cdot r^T$ came from the first and sixth columns of H, correcting these components in r results in the code word 100110.
- If we sum the third and fifth columns of H to get this syndrome, upon changing the third and fifth component of r we get a second code word, 001101.
- So we cannot expect it to correct multiple errors.

(This is no surprise, since the minimum distance between code words is 3:

$$2k + 1 = 3 \Rightarrow k = 1,$$

the code can correct errors of at most 1.)

Example 4.14. Let $E : \mathbb{Z}_2^3 \to \mathbb{Z}_2^6$.

(a) Use the parity-check matrix H:

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

To decode the following received words:

(i) 111101, (ii) 110101, (iii) 001111, (iv) 100100, (v) 110001, (vi) 111111, (vii) 111100, (viii) 010100.

(b) Are all the results in part (a) uniquely determined?

Solution a):

Let $r_1 = 111101$ Consider $H \cdot r_1^T$:

$$\begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \Rightarrow H \cdot r_1^T = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

Similarly, we can obtain all other decoded words.
The decoded words are:

	Received Word	**Correction at Position**	**Decoded Word**	**Considering First 3 Bits (Decoded Word)**
(i)	111101	3rd	110101	110
(ii)	110101	nil	110101	110
(iii)	001111	5th	001101	001
(iv)	100100	5th	100110	100
(v)	110001	4th	110101	110
(vi)	111111	1st & 6th	011110 (other possibilities)	011
(vii)	111100	4th	111000 (other possibilities)	111
(viii)	010100	1st & 6th	110101 (other possibilities)	110

Solution b): No, not all results are uniquely determined, (vi) and (viii) received words have other possibilities.

Example 4.15. The encoding function $E : \mathbb{Z}_2^2 \to \mathbb{Z}_2^5$ is given by the generator matrix

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

(a) Determine all code words. What can we say about the error-detection capability of this code? What about its error-correction capability?

(b) Find the associated parity-check matrix H.

(c) Use H to decode each of the following received words:

i) 11011

ii) 10101

iii) 11010

iv) 00111

v) 11101

vi) 00110

Solution a):

Let $W = \{00, 10, 01, 11\}$. Then

$$c = wG = \begin{bmatrix} 00 \\ 10 \\ 01 \\ 11 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 00000 \\ 10110 \\ 01011 \\ 11101 \end{bmatrix}$$

So the code is: $C = \{00000, 10110, 01001, 11101\}$.
Let $c_1 = 00000,\ c_2 = 10110,\ c_3 = 01001,\ c_4 = 11101$

The Hamming distance between code words are:

$$\begin{aligned} d(c_1, c_2) = 3, \quad d(c_1, c_3) = 3, \quad d(c_1, c_4) = 4 \\ d(c_2, c_3) = 4, \quad d(c_2, c_4) = 3, \quad d(c_3, c_4) = 3 \end{aligned}$$

The minimum distance between the code words is 3. Therefore:

- The code can detect all errors of weight ≤ 2.
- The code can correct all single errors. since $2t + 1 \leq 3 \Rightarrow t \leq 1$.

Solution b):

Now find H. Since $E : \mathbb{Z}_2^2 \to \mathbb{Z}_2^5$, $m = 2$, $n = 5$:

$$G = [I_2 \mid A], \quad \text{where } A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \Rightarrow H = \begin{bmatrix} A^T \mid I_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Solution c): Use H to decode the received words:

Compute syndrome: $s = Hr^T$.

i) : $r = 11011$, $s = Hr^T = 110 \Rightarrow$ error at position 1 $\Rightarrow$ decode to 01011

ii) : $r = 10101$, $s = 011 \Rightarrow$ error at position 2 $\Rightarrow$ decode to 11101

iii) : $r = 11010$, $s = 111 \Rightarrow$ error at position 1 and 5 $\Rightarrow$ decode to 01011

iv) : $r = 00111$, $s = 111 \Rightarrow$ error at position 1 and 5 $\Rightarrow$ decode to 10110

v) : $r = 11101$, $s = 000 \Rightarrow$ valid codeword

vi : $r = 00110$, $s = 110 \Rightarrow$ error at position 1 $\Rightarrow$ decode to 10110

Considering the first 2 bits (message), the decoded words are:

$$01,\ 11,\ 01,\ 10,\ 11,\ 10$$

Example 4.16. Define the encoding function $E : \mathbb{Z}_2^3 \to \mathbb{Z}_2^6$ by means of the parity-check matrix

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

(a) Determine all code words.

(b) Does this code correct all single errors in transmission?

Solution a):

Any $x = x_1x_2x_3x_4x_5x_6 \in \mathbb{Z}_2^6$ is a codeword if and only if $Hx^T = 0$.
Consider

$$Hx^T = H = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

This implies the equations:

$$x_1 + x_3 = x_4$$
$$x_1 + x_2 = x_5$$
$$x_1 + x_3 = x_6$$

Possible combinations:

x_1	x_2	x_3	x_4	x_5	x_6	Codeword	Label
0	0	0	0	0	0	000000	c_1
1	0	0	1	1	1	100111	c_2
0	1	0	0	1	0	010010	c_3
0	0	1	1	0	1	001101	c_4
1	1	0	1	0	1	110101	c_5
1	0	1	0	1	0	101010	c_6
0	1	1	1	1	1	011111	c_7
1	1	1	0	0	0	111000	c_8

Solution b):

Hamming distance table $d(c_i, c_j)$:

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
c_1	$-$	4	2	3	4	3	5	3
c_2	4	$-$	4	3	2	3	3	5
c_3	2	4	$-$	5	4	3	3	3
c_4	3	3	5	$-$	3	4	2	4
c_5	4	2	4	3	$-$	5	3	3
c_6	3	3	3	4	5	$-$	4	2
c_7	5	3	3	2	3	4	$-$	4
c_8	3	5	3	4	3	2	4	$-$

The minimum distance $= 2$. Therefore:

$2t + 1 \leq 2 \Rightarrow t \leq 0 \Rightarrow$ This code **cannot** correct single errors in transmission.