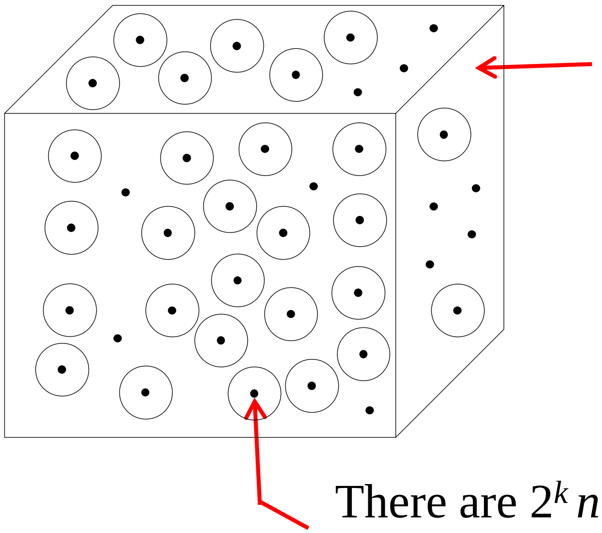
Chapter 4 Channel Coding

- 4.1 An Introduction of Channel Coding
- 4.2 Shannon's Channel Coding Theorem
- 4.3 Block Codes
- 4.4 Cyclic Codes
- 4.5 A Course Towards Decoding

§ 4.1 An Introduction of Channel Coding

- Channel Coding: map a k-dimensional message vector to an n-dimensional codeword vector, and $k < n$.
- If it is a binary channel code, there are at most 2^k n-dimensional codewords. The redundancy of $2^n - 2^k$ enables the error-correction capability of the code.



The n-dimensional binary space that can accommodate at most 2^n binary vectors.

There are 2^k n-dimensional codeword vectors filling the space.

- Codebook $\mathcal{C}$ collects all codewords. It has a cardinality of $|\mathcal{C}| = 2^k$.

§ 4.1 An Introduction of Channel Coding

- Code rate (r): A ratio of code dimension k to codeword length n, i.e., $r = \frac{k}{n}$. The redundancy is $n - k$. It underpins the efficiency in error-correction.
- Decoding:

$$\xrightarrow{\bar{c}} \boxed{\text{Channel}} \xrightarrow{\bar{y}}$$

Aim: with the received vector $\bar{y}$, we try to estimate $\bar{c}$. Let $\hat{\bar{c}}$ denote the estimation produced by the decoder. The decoding can be categorized into three cases:

Case I: $\hat{\bar{c}} = \bar{c}$, correct decoding;

Case II: $\hat{\bar{c}} \in \mathcal{C}$, but $\hat{\bar{c}} \neq \bar{c}$, decoding error;

Case III: Decoder does not produce any outcome, decoding failure.

§ 4.1 An Introduction of Channel Coding

- A channel code is a specific capacity approaching operational strategy.
- Based on the encoder structure, channel codes can be categorized into block codes and convolutional codes.

1. Block codes:

$$k\text{-symbol message} \xrightarrow{\text{Enc.}} n\text{-symbol codeword.}$$

- Encoder is memoryless and can be implemented with a combinatorial logic circuit.
- **Linear Block Code:** If $\bar{c}_i$ and $\bar{c}_j$ belong to a block code, $\bar{c}' = a \cdot \bar{c}_i + b \cdot \bar{c}_j$ also belongs to the block code. $(a, b) \in \mathbb{F}_q$ in which the block code is defined.
- Examples: **Reed-Solomon code**, algebraic-geometric code, **Hamming code**, low-density parity-check (LDPC) code.

§ 4.1 An Introduction of Channel Coding

2. Convolutional codes:

$$k\text{-symbol message} \xrightarrow{\text{Enc.}} n\text{-symbol codeword.}$$

$$m \text{ out of } k \text{ symbols}$$

$$k\text{-symbol message} \longrightarrow n\text{-symbol codeword.}$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$m \text{ out of } k \text{ symbols}$$

$$k\text{-symbol message} \longrightarrow n\text{-symbol codeword.}$$

- Encoder has a memory of order *m*. It can be implemented with a sequential logic circuit.
- Examples: **Convolutional code**, **Trellis coded modulation**, **Turbo code**, Spatially-coupled LDPC code.

§ 4.1* An Introduction of Channel Coding

- Encoding of a k-dimensional message vector over finite field of size q, i.e., $\mathbb{F}_q$.

$$\bar{u} = (u_1, u_2, \cdots, u_k) \in \mathbb{F}_q^k$$

$$\xrightarrow{\text{Enc.}}$$

$$\bar{c} = (c_1, c_2, \cdots, c_k, c_{k+1}, \cdots, c_n) \in \mathbb{F}_q^n$$

- This can be seen as adding $n-k$ parity-check symbols into $\bar{u}$, which are known as the redundancy.
- Sometimes, the parity-check symbols can be denoted as $p_1, p_2, \cdots, p_{n-k}$.

§ 4.1* An Introduction of Channel Coding

- **Hamming Sphere**
- Given a length n code defined over $\mathbb{F}_q$.
- Any codeword $\bar{c}$ can define a Hamming sphere of radius t. The sphere contains

$$V_q(n,t) = \sum_{j=0}^{t} \binom{n}{j} (q-1)^j$$

length-n vectors.

- **Hamming Bound**
- With a codebook $\mathcal{C}$, the number of message and symbols are $\log_q |\mathcal{C}|$.
- Redundancy

$$R = n - \log_q |\mathcal{C}|$$

For linear codes, $|\mathcal{C}| = q^k$, and

$$R = n - k$$

§ 4.1* An Introduction of Channel Coding

- In an n-dimensioned vector space, there are q^n vectors. A t-error-correcting code does not allow the Hamming sphere of the codewords overlap. Hence,

$$|\mathcal{C}|\, V_q(n,t) \leq q^n$$

and

$$\frac{q^n}{|\mathcal{C}|} \geq V_q(n,t)$$

Hence,

$$\log_q \frac{q^n}{|\mathcal{C}|} \geq \log_q V_q(n,t)$$

$$n - \log_q|\mathcal{C}| \geq \log_q V_q(n,t)$$

$$R \geq \log_q V_q(n,t)$$

§ 4.1* An Introduction of Channel Coding

- If $q^R = V_q(n, t)$, the code is a **perfect code**.

- **Properties of a perfect code**:
  (1). The number of redundancy patterns = The number of vectors in the Hamming sphere of radius t;
  (2). Bounded distance decoding is optimal, i.e., maximum likelihood (ML) decoding.

- Perfect codes include: Hamming codes, Repetition codes, Golay codes.

§ 4.2 Shannon's Channel Coding Theorem

Shannon's Channel Coding Theorem: All rates below capacity C are achievable. For every rate $r < C$, there exists channel codes of length n and dimension nr, such that the maximum error probability $P_e \to 0$. Inversely, any such codes that realize $P_e \to 0$ must have rate $r < C$.

- Shannon's Channel Coding Theorem demonstrates error free transmission is possible by manipulating the code rate according to the channel capacity. It is defined in the mindset of binary transmission, e.g., BPSK.
- Its proof involves the justification of achievability, i.e., if $r < C$, $P_e \to 0$, and its converse, i.e., if $P_e \to 0$, $r < C$. They require the assistance of Jointly Typical Sequences and Fano's Inequality, respectively.

§ 4.2 Shannon's Channel Coding Theorem

- **<u>Empirical Entropy</u>**: Given an X sequence $X^n(x^n: x_1, x_2, \dots, x_n)$, its empirical entropy is

$$H^*(X) = -\frac{1}{n}\log_2 P(x^n)$$

- Similarly, given two sequences $X^n(x^n: x_1, x_2, \dots, x_n)$ and $Y^n(y^n: y_1, y_2, \dots, y_n)$, their joint empirical entropy is

$$H^*(X,Y) = -\frac{1}{n}\log_2 P(x^n, y^n)$$

- If sequences X^n and Y^n have the i.i.d. property, i.e.

$$P(x^n) = \prod_{i=1}^{n} P(x_i) \qquad P(x^n, y^n) = \prod_{i=1}^{n} P(x_i, y_i)$$

the above empirical entropies become

$$H^*(X) = -\frac{1}{n}\sum_{i=1}^{n}\log_2 P(x_i) \qquad H^*(X,Y) = -\frac{1}{n}\sum_{i=1}^{n}\log_2 P(x_i, y_i)$$

§ 4.2 Shannon's Channel Coding Theorem

- **<u>Jointly Typical Sequences</u>:** Given $\epsilon \to 0$, x^n and y^n are jointly typical sequences if

$$|H^*(X) - H(X)| < \epsilon$$
$$|H^*(Y) - H(Y)| < \epsilon$$
$$|H^*(X,Y) - H(X,Y)| < \epsilon.$$

- ① If x^n and y^n are drawn i.i.d. as

$$P(x^n, y^n) = \prod_{i=1}^{n} P(x_i, y_i),$$

when $n \to \infty$,

$$\Pr(x^n \text{ and } y^n \text{ are jointly typical}) \to 1.$$

② If z^n and y^n are independent, as $P(z^n, y^n) = P(z^n)\ P(y^n)$,

$$\Pr(z^n \text{ and } y^n \text{ are jointly typical}) \leq 2^{-n(I(Z;Y)-3\epsilon)}.$$

§ 4.2 Shannon's Channel Coding Theorem

- **<u>Modelling and Assumptions of the Proof</u>**

$$u^k(w) \longrightarrow \boxed{\text{Encoder}} \xrightarrow{c^n(w)} \boxed{\text{Channel}} \xrightarrow{y^n} \boxed{\text{Decoder}} \longrightarrow u^k(\hat{w})/c^n(\hat{w})$$

- Codeword length n, dimension $k = nr$, message/codeword index w
- Decoding error probability $P(\epsilon) = \Pr(\hat{w} \neq w)$

- Assumptions (A):

A-I: A random binary code is generated as

$$P(\mathcal{C}) = \prod_{w=1}^{2^{nr}} P(c^n(w))$$

$$= \prod_{w=1}^{2^{nr}} \prod_{i=1}^{n} P(c_i(w)).$$

§ 4.2 Shannon's Channel Coding Theorem

A-II: Both the transmitter and receiver know the channel, i.e., $P(y_i|c_i(w))$, $\forall i$.

A-III: Messages (codewords of $\mathcal{C}$) are uniformly chosen for transmission as

$$P\left(u^k(w)\right) = P\left(c^n(w)\right) = \frac{1}{2^{nr}} \ .$$

A-IV: The channel is discrete memoryless, i.e.,

$$P(y^n|c^n(w)) = \prod_{i=1}^{n} P(y_i|c_i(w)) \ .$$

Therefore,

$$\begin{aligned} P(c^n(w), y^n) &= P(y^n|c^n(w))\, P\left(c^n(w)\right) \\ &= \prod_{i=1}^{n} P(y_i|c_i(w)) \cdot \prod_{i=1}^{n} P(c_i(w)) \\ &= \prod_{i=1}^{n} P(y_i, c_i(w)). \end{aligned}$$

§ 4.2 Shannon's Channel Coding Theorem

Achievability Proof

- Generate a random binary code of length n rate r as A-I.
  The codebook $\mathcal{C}$ is

$$\mathcal{C} = \begin{pmatrix} c_1(1) & c_2(1) & \cdots & c_n(1) \\ \vdots & \vdots & \cdots & \vdots \\ c_1(w) & c_2(w) & \cdots & c_n(w) \\ \vdots & \vdots & \cdots & \vdots \\ c_1(2^{nr}) & c_2(2^{nr}) & \cdots & c_n(2^{nr}) \end{pmatrix}$$

They are codewords

$$P(\mathcal{C}) = \prod_{w=1}^{2^{nr}} \prod_{i=1}^{n} P(c_i(w))$$

- Based on A-III,

$$P\big(c^n(w)\big) = \prod_{i=1}^{n} P(c_i(w)) = \frac{1}{2^{nr}}.$$

- With received vector y^n , the decoder estimates codeword $c^n(\hat{w})$ such that
  - $c^n(\hat{w})$ and y^n are jointly typical sequences.
  - There is no other codeword $c^n(v)$ such that $c^n(v)$ and y^n are jointly typical sequences.

§ 4.2 Shannon's Channel Coding Theorem

- The decoding error probability is

$$P(\epsilon) = \sum_{\mathcal{C}} \underbrace{P(\mathcal{C})}_{\text{Prob. of a particular code } \mathcal{C}} \; \underbrace{P_e(\mathcal{C})}_{\text{Error prob. of the code } \mathcal{C}}$$

$$P_e(\mathcal{C}) = \frac{1}{2^{nr}} \sum_{w=1}^{2^{nr}} \underbrace{P_{e,w}(\mathcal{C})}_{\text{Error prob. of a particular codeword } c^n(w) \in \mathcal{C}}$$

$$P(\epsilon) = \frac{1}{2^{nr}} \sum_{\mathcal{C}} \sum_{w=1}^{2^{nr}} P(\mathcal{C}) P_{e,w}(\mathcal{C})$$

- Due to symmetry of code construction, we know

$$\frac{1}{2^{nr}} \sum_{w=1}^{2^{nr}} P_{e,w}(\mathcal{C}) = P_{e,1}(\mathcal{C})$$

- Hence,

$$\begin{aligned} P(\epsilon) &= \sum_{\mathcal{C}} P(\mathcal{C})\, P_{e,1}(\mathcal{C}) \\ &= \underbrace{P_{e,1}}_{\text{Average (over all codebooks) error prob. of codeword } c^n(1)} \end{aligned}$$

§ 4.2 Shannon's Channel Coding Theorem

- Let E_w denote the event that codeword $c^n(w)$ (X^n) and y^n (Y^n) are jointly typical sequences.

$$
\begin{aligned}
P(\epsilon) &= P_{e,1} \\
&= \Pr(E_1^C \cup E_2 \cup E_3 \cup \cdots \cup E_{2^{nr}}) \\
&\le \Pr\left(E_1^C\right) + \sum_{w=2}^{2^{nr}} \Pr(E_w)
\end{aligned}
$$

Based on ①, where $n \to \infty$, $\Pr\left(E_1^C\right) \le \epsilon$.

Based on ②, $\Pr(E_w) \le 2^{-n(I(X;Y)-3\epsilon)}$.

- Therefore,

$$
\begin{aligned}
P(\epsilon) &\le \epsilon + \sum_{w=2}^{2^{nr}} 2^{-n(I(X;Y)-3\epsilon)} \\
&= \epsilon + (2^{nr} - 1) \cdot 2^{-n(I(X;Y)-3\epsilon)} \\
&< \epsilon + 2^{3n\epsilon} 2^{-n(I(X;Y)-r)} \\
&= \epsilon + 2^{-n(I(X;Y)-3\epsilon-r)}
\end{aligned}
$$

§ 4.2 Shannon's Channel Coding Theorem

- If n is sufficiently large and $r < I(X;Y) - 3\epsilon$,

$$P(\epsilon) \le 2\epsilon,$$

the decoding error probability can be arbitrarily small.

- Choose $P(c_i(w))$ to be the distribution that maximizes $I(X;Y)$ as

$$C = \max_{P(c_i(w))} \{I(X;Y)\},$$

the above conclusion implies if $r < C$, the decoding error probability $P(\epsilon)$ can be arbitrarily small. <u>Achievability Proof Ends</u>

Remark: The achievability proof is founded on <u>random code construction</u>, <u>large codeword length</u> and <u>ideal codeword symbol distributions</u>. They become the features of capacity approaching (achieving) codes, i.e. Turbo codes, LDPC codes and Polar codes.

§ 4.2 Shannon's Channel Coding Theorem

- **<u>Converse of Shannon's Channel Coding Theorem</u>**

  If $P(\epsilon) \to 0$, $r < C$.

$$\xrightarrow{u^k(w)} \boxed{\text{Encoder}} \xrightarrow{c^n(w)} \boxed{\text{Channel}} \xrightarrow{y^n} \boxed{\text{Decoder}} \xrightarrow{u^k(\widehat{w})/c^n(\widehat{w})}$$

- **<u>Fano's inequality</u>**

  Over a DMC, given a code of rate r with the input message uniformly distributed, let $P(\epsilon) = \Pr(\widehat{w} \neq w)$,

$$H(c^n|y^n) < 1 + P(\epsilon) \cdot nr.$$

Proof: Extending the Fano's inequality into vector domain,

$$\begin{aligned} H(c^n|y^n) &\leq H\big(P(\epsilon)\big) + P(\epsilon)\log_2(2^{nr} - 1) \\ &< 1 + P(\epsilon) \cdot nr. \end{aligned}$$

Note: The 2nd inequality is realized with $n \to \infty$.

§ 4.2 Shannon's Channel Coding Theorem

Converse Proof

- Based on A-III, input message rate are uniformly distributed.

$$H\left(u^k(w)\right) = \log_2 2^{nr} = nr.$$

- Since

$$H\left(u^k(w)\right) = H\left(u^k(w)|y^n\right) + I(u^k(w); y^n)$$

where

$$H\left(u^k(w)\middle|y^n\right) = H(c^n(w)|y^n)$$

and based on Data Processing Inequality,

$$I\left(u^k(w); y^n\right) \leq I(c^n(w); y^n).$$

we have

$$nr = H\left(u^k(w)\right) \leq H(c^n(w)|y^n) + I(c^n(w); y^n).$$

§ 4.2 Shannon's Channel Coding Theorem

- Applying Fano's Inequality

$$H(c^n(w)|y^n) < 1 + P(\varepsilon) \cdot nr.$$

- Over DMC and input being independent

$$I(c^n(w); y^n) = \sum_{i=1}^{n} I(c_i(w); y_i)$$
$$\leq n \cdot C.$$

Therefore,

$$nr < 1 + P(\varepsilon)nr + nC$$
$$r < P(\varepsilon)r + \frac{1}{n} + C$$

With $n \to \infty$ and $P(\varepsilon) \to 0$, $r < C$.

<u>Converse Proof Ends</u>

§ 4.3 Block Codes

- All block codes are defined by their codeword length *n*, dimension *k* and the minimum Hamming distance *d*. <u>A block code is often denoted as an (*n*, *k*, *d*) code.</u>
- Code rate: $r = \frac{k}{n}$.
- Encoding of a linear block code can be written as:

$$\boxed{\bar{c} = \bar{u} \cdot \mathbf{G}}$$

$\bar{u}$ — *k*-dimensional message vector.

G — a generator matrix of size $k \times n$. It defines the legal space among all *n*-dimensional vectors.

$\bar{c}$ — *n*-dimensional codeword vector.

Linear block code:

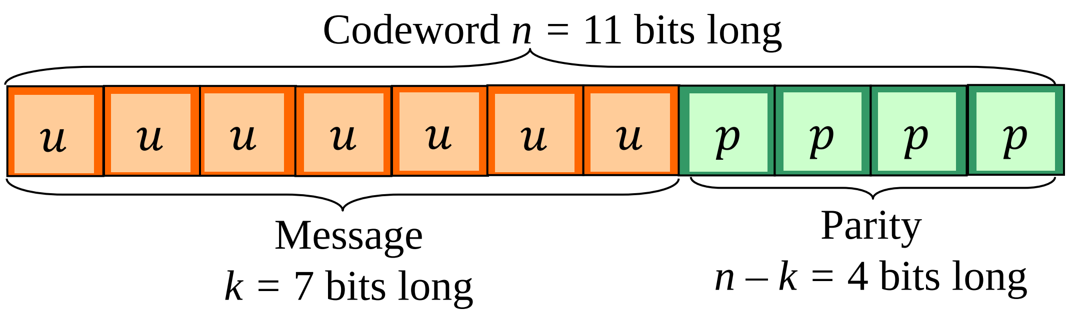
$$\bar{c}_1 = \bar{u}_1 \cdot \mathbf{G}$$

$$\bar{c}_2 = \bar{u}_2 \cdot \mathbf{G}$$

$$(\bar{u}_1 + \bar{u}_2) \cdot \mathbf{G} = (\bar{c}_1 + \bar{c}_2) \in \mathcal{C}$$

§ 4.3 Block Codes

Hamming Distance



u = message bits
p = parity-check bits

The Hamming Distance between any two codewords is the total number of positions where the two codewords differ.

1	0	0	1	1	1	0	1

0	0	1	1	0	1	0	0

The total number of positions where these two codewords differ is 4. Therefore, the Hamming distance is 4.

Weight: Given a vector, its weight is the number of nonzero positions.

1	0	0	1	1	1	0	1

The weight of the vector is 5.

§ 4.3 Block Codes

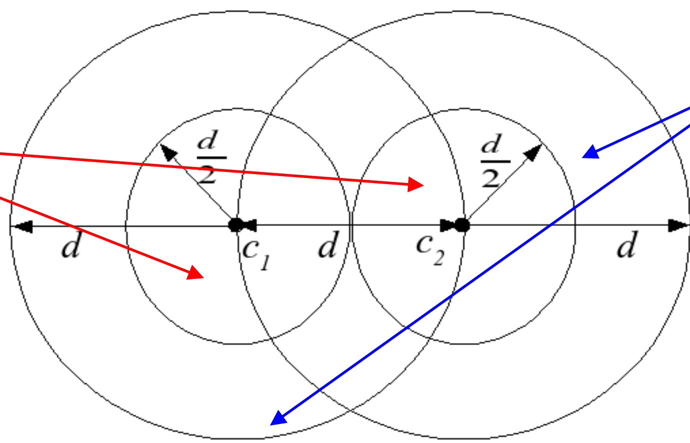
The Minimum Hamming Distance and Error-Correction Capability

The minimum Hamming distance: for any two codewords $\bar{c}_i$ and $\bar{c}_j$ picked up from the codebook $\mathscr{C}$, the minimum Hamming distance d is defined as:

$$d = \min_{(\bar{c}_i, \bar{c}_j) \in \mathscr{C}} \{d_{\text{Ham}}(\bar{c}_i, \bar{c}_j)\}.$$

- In general, a block code can correct up to $\left\lfloor \frac{d-1}{2} \right\rfloor$ errors, where $\lfloor x \rfloor$ means that x is rounded down to the nearest integer, e.g., $\lfloor 2.5 \rfloor = 2$.
- A block code can **detect** d – 1 errors.

A block code can ***correct*** received words with up to $\left\lfloor \frac{d-1}{2} \right\rfloor$ errors.



A block code can ***detect*** up to d – 1 errors

- For a linear block code, $d = \min\{\text{weight}\left(\bar{c}_j\right), \bar{c}_j \neq 0\}$.

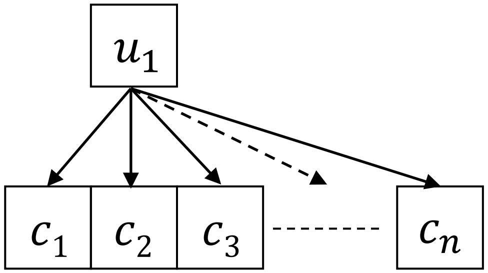
§ 4.3 Block Codes

Repetition Codes

A repetition encoder takes a **single** message bit and gives a codeword that is the message bit repeated n times, where n is an **odd** number

A message bit **0** will be encoded to give the codeword **0000...000**
A message bit **1** will be encoded to give the codeword **1111...111**

- This is the simplest type of error-correcting code as it only has **two codewords**
- We can easily see that it has a minimum Hamming distance $d = n$
- It is an $(n, 1, n)$ block code



The generator matrix of the code is simply

$$\mathbf{G} = [1\ 1\ 1\ 1\ \ldots\ 1]$$

§ 4.3 Block Codes

Repetition Codes

To recover the transmitted codeword of a repetition code, a simple decoder known as a **Majority Decoder** can be used

1. The number of 0s and 1s in the received word are counted.
2. If the number of 0s > number of 1s (i.e., a majority) , then the message bit was a 0. Else if the number of 1s > number of 0s, then the message bit was a 1.

Example 4.1**:** Say our message bit was a 1 and it was encoded by the (5, 1, 5) repetition code. The codeword will be $\bar{c}$ = (11111).

- If after transmission we receive the word $\bar{r}$ = (10011), then the number of 1s > number of 0s and so the majority decoder decides that the original message was 1.
- However, if we receive the word $\bar{r}$ = (00011) then the number of 0s > number of 1s and the majority decoder **incorrectly** decides that the original message was 0.

In general, a (n, 1, n) repetition code can correct up to $\frac{n-1}{2}$ errors.

§ 4.3 Block Codes

Repetition Codes

The Great Wall

§ 4.3 Block Codes

Hamming Codes

- Single-error-correcting codes.
- Given any positive integer $m \geq 3$, its

  $n = 2^m - 1$

  $k = 2^m - m - 1$

  $d = 3$

- ***Example 4.2*** **:** Given $m = 3$, the generator matrix of the (7, 4, 3) Hamming code is

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

The codewords can be generated by $\bar{c} = \bar{u} \cdot \mathbf{G}$.

This code can correct 1 error.

§ 4.3 Block Codes

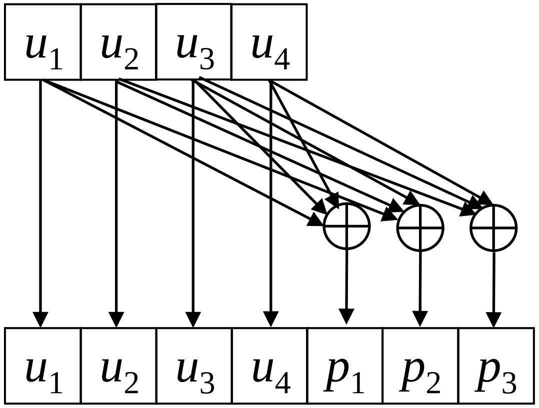
Notice that only 16 of 128 possible sequences of length 7 bits are used for transmission.

The parity bits are calculated by

$$p_1 = u_1 \oplus u_3 \oplus u_4$$

$$p_2 = u_1 \oplus u_2 \oplus u_3$$

$$p_3 = u_2 \oplus u_3 \oplus u_4$$



The encoding can be written as

$$\bar{c} = \bar{u} \cdot \mathbf{G},$$

and

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

Remark: This is a **systematic encoding** as the message symbols appear in the codeword.

$\bar{u}$	$\bar{c}$
0000	0000 000
0001	0001 101
0010	0010 111
0011	0011 010
0100	0100 011
0101	0101 110
0110	0110 100
0111	0111 001
1000	1000 110
1001	1001 011
1010	1010 001
1011	1011 100
1100	1100 101
1101	1101 000
1110	1110 010
1111	1111 111

§ 4.4 Cyclic Codes

- A cyclic code is a block code which has the property that cyclically shifting a codeword results in another codeword

- Cyclic codes have the advantage that simple encoders can be constructed using shift registers and low complexity decoding algorithms exist to decode them

- An (n, k) cyclic code is constructed by first choosing a generator polynomial $g(x)$ and multiplying this by a message polynomial $m(x)$ to generate a codeword polynomial $c(x)$ as

$$c(x) = u(x) \cdot g(x)$$

$$u(x) = u_0 + u_1 x + \cdots + u_{k-1} x^{k-1}$$

$$g(x) = g_0 + g_1 x + \cdots + g_{n-k} x^{n-k}$$

$$c(x) = c_0 + c_1 x + \cdots + c_{n-1} x^{n-1}$$

§ 4.4 Cyclic Codes

Cyclic Hamming Code

- The (7, 4, 3) Hamming code is also a cyclic code that can be constructed using the generator polynomial $g(x) = x^3 + x^2 + 1$.
- ***Example 4.3***: To encode the binary message 1010, we first write it as the message polynomial $u(x) = x^3 + x$ and then multiply it with $g(x)$ modulo-2

$$\begin{aligned} c(x) &= u(x)g(x) \\ &= (x^3 + x)(x^3 + x^2 + 1) \\ &= x^6 + x^5 + x^4 + x^3 + x^3 + x \qquad [(x^3 + x^3) \bmod 2 = 2x^3 \bmod 2 = 0] \\ &= x^6 + x^5 + x^4 + x \end{aligned}$$

This codeword polynomial corresponds to 1 1 1 0 0 1 0. However, notice that the first 4 bits of this codeword are not the same as the original message 1010.

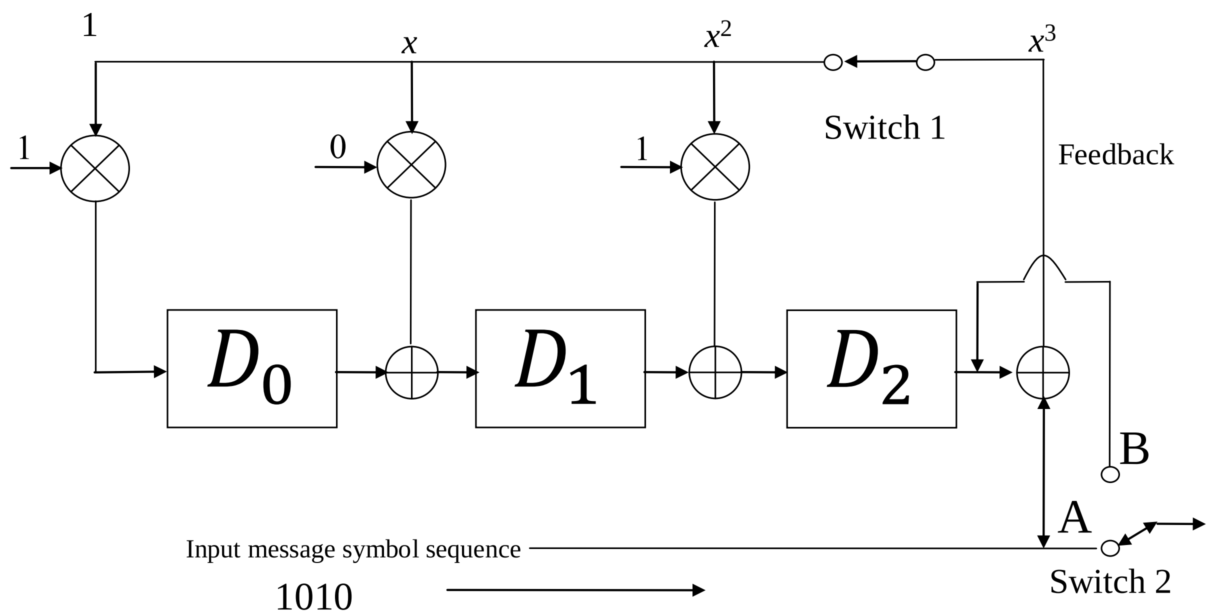
- This is an example of a **non-systematic code**.

Remark: Systematic encoding and non-systematic encoding only change the mapping between message and codeword, not the codebook.

§ 4.4 Cyclic Codes

Systematic Cyclic Hamming Code

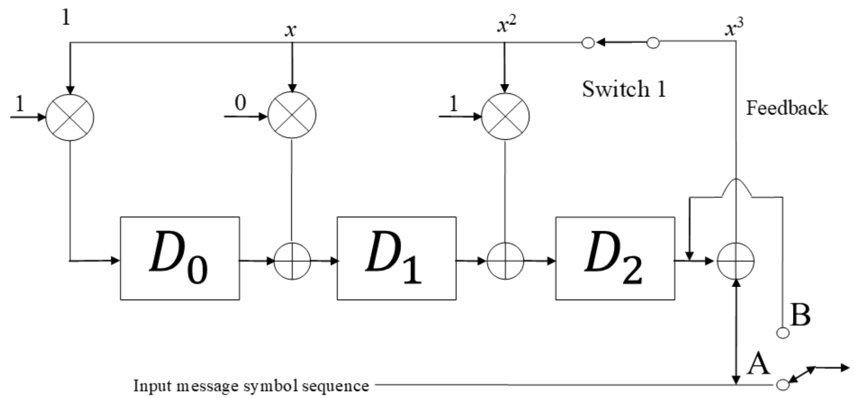
- Encoding of a systematic cyclic Hamming code can be performed by shift-registers.



An encoder for the systematic (7, 4, 3) cyclic Hamming code

1. For the first k = 4 message bits, switch 1 is closed and switch 2 is in position A
2. After the first 4 message bits have entered, switch 1 is open, switch 2 is in position B and the contents of memory elements are read out giving the parity-check bits

§ 4.4 Cyclic Codes



Example 4.4: Let the message be $\bar{u} = (u_1, u_2, u_3, u_4)$, the shift register computes

Input	Registers (left to right)		
u_1	u_1	0	u_1
u_2	$u_1 \oplus u_2$	u_1	$u_1 \oplus u_2$
u_3	$u_1 \oplus u_2 \oplus u_3$	$u_1 \oplus u_2$	$u_2 \oplus u_3$
u_4	$u_2 \oplus u_3 \oplus u_4$	$u_1 \oplus u_2 \oplus u_3$	$u_1 \oplus u_3 \oplus u_4$

Update of the shift registers:

$$Feedback = D_2 \oplus Input$$
$$D_2' = D_1 \oplus 1 \cdot Feedback$$
$$D_1' = D_0 \oplus 0 \cdot Feedback$$
$$D_0' = 1 \cdot Feedback$$

Hence, $p_1 = u_1 \oplus u_3 \oplus u_4$

$p_2 = u_1 \oplus u_2 \oplus u_3$

$p_3 = u_2 \oplus u_3 \oplus u_4$

This is equivalent to the systematic encoding of ***Example 4.2***.

§ 4.5 A Course Towards Decoding

$$u^k(w) \longrightarrow \boxed{\text{Encoder}} \xrightarrow{c^n(w)} \boxed{\text{Channel}} \xrightarrow{y^n} \boxed{\text{Decoder}} \longrightarrow u^k(\widehat{w})/c^n(\widehat{w})$$

- Given a received word y^n, decoding aims to recover codeword $c^n(w)$ (or message $u^k(w)$), yielding its estimation $(c^n(\widehat{w}))$(or $u^k(\widehat{w})$).
- Error-Correction starts from error-detection.
- The **<u>Parity-Check Code</u>**: for each binary message, a parity-check bit is added so that there are an even number of 1s in each codeword.

  If k = 3 then there are 8 possible messages. The eight codewords will be:

000 → 000**0**	100 → 100**1**
001 → 001**1**	101 → 101**0**
010 → 010**1**	110 → 110**0**
011 → 011**0**	111 → 111**1**

When there are odd number of 1, the decoder (detector) knows error has been introduced.

§ 4.5 A Course Towards Decoding

Parity-Check Matrix

- A primitive thought: given a received word $\bar{r}$, we can search the whole codebook and find the codeword (message) that has the smallest Hamming distance to $\bar{r}$. But even for a binary code, this has a complexity of $O(2^k)$. This process is called the maximum likelihood (ML) decoding.
- Alternatively, we can utilize the algebraic structure of the code, which is often told by the parity-check matrix **H**.
- A parity-check matrix **H** is defined as the **null space** of the generator matrix **G**, i.e., the inner product of the two matrices results in an all-zero matrix, $\mathbf{G}\mathbf{H}^T = \mathbf{0}$ (T is the transpose of the matrix)
- When a codeword is multiplied by the parity-check matrix, it should result in an all-zero vector, i.e.,

$$\bar{c} \cdot \mathbf{H}^T = \bar{u} \cdot \mathbf{G} \cdot \mathbf{H}^T = \underset{\uparrow\,\text{Syndrome vector.}}{\underline{0}}.$$

- If $\hat{\bar{c}} \cdot \mathbf{H}^T = 0$, it implies $\hat{\bar{c}}$ is a valid codeword.

§ 4.5 A Course Towards Decoding

- If the generator matrix is of the form $\mathbf{G} = [\mathbf{I}_k \mid \mathbf{P}]$, where $\mathbf{I}_k$ is a $k \times k$ identity matrix and $\mathbf{P}$ is a parity matrix, the parity-check matrix is in the form of $\mathbf{H} = [\mathbf{P}^T \mid \mathbf{I}_{n-k}]$.

Example 4.5: Taking the (7, 4, 3) Hamming code in ***Example 4.2***

$$\mathbf{G} = \left[\begin{array}{cccc:ccc} \overbrace{1 \quad 0 \quad 0 \quad 0}^{\mathbf{I}_4} & & \overbrace{1 \quad 1 \quad 0}^{\mathbf{P}} \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{array}\right]$$

The parity-check matrix is

$$\mathbf{H} = \left[\begin{array}{cccc:ccc} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right]$$

$\mathbf{P}^T$

$\mathbf{I}_{n\text{-}k} = \mathbf{I}_{7\text{-}4} = \mathbf{I}_3$

- Dual code property

	Generator matrix	Parity-check matrix
Code	$\mathbf{G}_{k\times n}$	$\mathbf{H}_{(n-k)\times n}$
Dual code	$\mathbf{G}_{(n-k)\times n}$	$\mathbf{H}_{k\times n}$

§ 4.5 A Course Towards Decoding

- Note that

$$\mathbf{G}\cdot\mathbf{H}^T = \left[\mathbf{I}_k \vdots \mathbf{P}_{k\times(n-k)}\right]\cdot\begin{bmatrix}\mathbf{P}_{k\times(n-k)}\\ \cdots\cdots\cdots\\ \mathbf{I}_k\end{bmatrix}$$
$$= \left[\mathbf{P}_{k\times(n-k)} + \mathbf{P}_{k\times(n-k)}\right]$$
$$= [0]_{k\times(n-k)}.$$

For a pair of dual codes, their codewords are generated by $\bar{c} = \bar{u}\cdot\mathbf{G}$, $\bar{c}^{\perp} = \bar{u}'\cdot\mathbf{H}$, where $\bar{u}\in\mathbb{F}_q^k$, $\bar{u}'\in\mathbb{F}_q^{n-k}$.

Then

$$\bar{c}\cdot(\bar{c}^{\perp})^T = (\bar{u}\cdot\mathbf{G})\cdot(\mathbf{H}^T\cdot(\bar{u}')^T)$$
$$= \bar{u}\cdot\mathbf{G}\cdot\mathbf{H}^T\cdot(\bar{u}')^T$$
$$= 0.$$

G and **H** define two orthogonal vector spaces (of the same length).

- **H** can be constituted by $n-k$ linearly independent codewords of an $(n, n-k)$ code.
- **G** can be constituted by k linearly independent codewords of an (n, k) code.

§ 4.5 A Course Towards Decoding

Example 4.5: Decoding of (7, 4, 3) Hamming code.

$$\mathbf{H} = \left[\begin{array}{cccc:ccc} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right]$$

Assume the transmittal codeword is

$$\bar{c} = (0\ 1\ 0\ 1\ 1\ 1\ 0).$$

The received word is

$$\bar{r} = \bar{c} + \bar{e} = (0\ 1\ 0\ 1\ 0\ 1\ 0).$$

($\bar{e} = (0\ 0\ 0\ 0\ 1\ 0\ 0)$ is the error pattern.)

The syndrome is

$$\bar{r} \cdot \mathbf{H}^T = (\bar{c} + \bar{e}) \cdot \mathbf{H}^T .$$

§ 4.5 A Course Towards Decoding

The syndrome is

$$
\begin{aligned}
\bar{r} \cdot \mathbf{H}^T &= (\bar{c} + \bar{e}) \cdot \mathbf{H}^T \\
&= \bar{c} \cdot \mathbf{H}^T + \bar{e} \cdot \mathbf{H}^T \\
&= \bar{0} + (0\ 0\ 0\ 0\ 1\ 0\ 0) \cdot \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
&= (1\ 0\ 0)
\end{aligned}
$$

=> Column-4 of **H**. (Row-4 of $\mathbf{H}^T$)

=> $c_4 = r_4 + 1 = 1.$

=> $\hat{c} = (\ 0\ 1\ 0\ 1\ 1\ 1\ 0\ ).$

§ 4.5 A Course Towards Decoding

<u>**Singleton Bound**</u>: Given an (n, k) linear block code with minimum Hamming distance d, we have

$$d \leq n - k + 1.$$

Proof:

- For the code, its parity-check matrix $\mathbf{H}_{(n-k)\times n}$ can be written as

  $$\mathbf{H} = [\,\bar{h}_1, \bar{h}_2, \dots, \bar{h}_n].$$

  Given a minimum weight codeword $\bar{c}$, it has a support of $\{i_1, i_2, \dots, i_d\}$. Moreover,

  $$c_{i_1} \cdot \bar{h}_{i_1}^T + c_{i_2} \cdot \bar{h}_{i_2}^T + \cdots + c_{i_d} \cdot \bar{h}_{i_d}^T = \bar{0}$$

  Hence, there are <u>**at least**</u> d column of **H** are linearly dependent.
- For **H**, its row rank equals to its column rank.
  Hence, there are <u>**at most**</u> $n - k$ linearly independent columns in **H**. That says any $n - k + 1$ columns of **H** are linearly dependent.
- Therefore,

  $$d \leq n - k + 1.$$

- Otherwise if $d > n - k + 1$, the minimum Hamming distance of the code will not be d.

Remark: If a code with $d = n - k + 1$, it is a maximum distance separable (MDS) code.

§ 4.6* Maximum Likelihood (ML) Decoding

Maximum Likelihood (ML) Decoding: Given a received word / symbol sequence $\bar{y}$, the codeword $\bar{c}$ (or message $\bar{u}$) that maximizes the channel transition probability $P(\bar{y}|\bar{c})$ is the decoding output which is denoted as $\hat{\bar{c}}$ (or $\hat{\bar{u}}$). That says

$$\hat{\bar{c}} = \operatorname*{argmax}_{\bar{c}\in\mathcal{C}} P(\bar{y}|\bar{c})\,.$$

- Based on Bayes' theorem, the a posteriori probability can be determined as

$$P(\bar{c}|\bar{y}) = \frac{P(\bar{y}|\bar{c})P(\bar{c})}{P(\bar{y})}\,.$$

Maximum A Posteriori (MAP) Decoding: Given $\bar{y}$, the codeword $\bar{c}$ (or message $\bar{u}$) that maximizes the MAP $P(\bar{c}|\bar{y})$ is the decoding output. That says

$$\hat{\bar{c}} = \operatorname*{argmax}_{\bar{c}\in\mathcal{C}} P(\bar{c}|\bar{y})\,.$$

- By assuming equiprobable codeword as $P(\bar{c}) = |\mathcal{C}|^{-1}$, the ML decoding output coincides with the MAP decoding.

§ 4.6* Maximum Likelihood (ML) Decoding

<u>Union Bound</u>

- Union bound can be used to characterize the ML decoding performance of codes, which requires knowledge of the code's <u>weight spectrum (distribution of codewords of different weights)</u>.

- The codeword $c_1^N = \bar{c} = \{c_1, c_2, \cdots, c_N\} \in \mathcal{C}$ of a linear block code has discrete weight values, denoted as $\{d_0, d_1, d_2, \ldots, d_s\}$, where $d_0 = 0$, $d \leq d_i \leq N$ and $i = 1, 2, \ldots, s$. The number of codewords with weight d_i is denoted as A_{d_i}. Hence, weight spectrum is $\{\mathrm{A}_{d_i}, \forall i\}$.

- Union upper bound on a linear block code's ML decoding frame error rate (FER) under BPSK modulation, AWGN channel and soft decision, is

code rate

It is an expectation over $\mathcal{C}$

energy of each info. bit

$$P_{\mathrm{ML,e}} \leq \sum_{d_i=d}^{d_s} \mathrm{A}_{d_i} Q\left(\frac{\sqrt{d_i}}{\sigma}\right) = \sum_{d_i=d}^{d_s} \mathrm{A}_{d_i} Q\left(\frac{\sqrt{r d_i E_b}}{\sigma}\right).$$

noise standard deviation

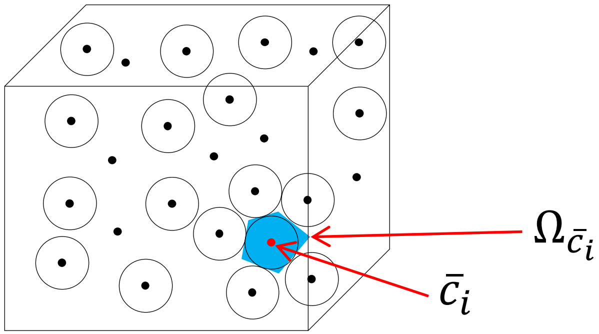
§ 4.6* Maximum Likelihood (ML) Decoding

- Proof:

The function $\mathbb{1}(\cdot)$ denotes the indicator function, where $\mathbb{1}(\text{true}) = 1$ and $\mathbb{1}(\text{false}) = 0$. Then,

$$P_{\text{ML,e}} = \sum_{\bar{c}\in\mathcal{C}} \sum_{\bar{y}\in\mathcal{Y}} \Pr(\bar{c},\bar{y}) \cdot \mathbb{1}(\text{Decoder}_{\text{ML}}(\bar{y}) \neq \bar{c}).$$

The set $\Omega_{\bar{c}_i}$ is defined as $\Omega_{\bar{c}_i} = \{\bar{y} | \Pr(\bar{y}|\bar{c}_i) \geq \Pr(\bar{y}|\bar{c}_{i'}), \forall i \neq i'\}$, which represents the space of all received signals $\bar{y}$ that will be decoded as the codeword $\bar{c}_i$ under the ML decision rule. The set $\mathcal{Y}$ is n-dimensional real vector space, and $\mathcal{Y} = \cup_{\bar{c}_i\in\mathcal{C}} \Omega_{\bar{c}_i}$.



§ 4.6* Maximum Likelihood (ML) Decoding

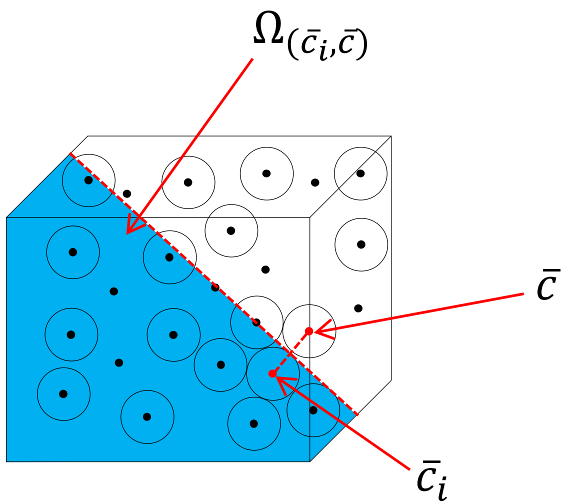
By symmetry, let $\bar{c}$ be $\bar{0}$. We have

$$P_{\mathrm{ML,e}} = \sum_{\bar{c}\in\mathcal{C}} \sum_{\bar{y}\in\mathcal{Y}\setminus\Omega_{\bar{c}}} \Pr(\bar{c},\bar{y}) = \Pr(\bar{c}) \sum_{\bar{c}\in\mathcal{C}} \sum_{\bar{y}\in\cup_{\bar{c}_i\neq\bar{c}}\Omega_{\bar{c}_i}} \Pr(\bar{y}|\bar{c}) = \sum_{\bar{y}\in\cup_{\bar{c}_i\neq\bar{c}}\Omega_{\bar{c}_i}} \Pr(\bar{y}|\bar{c})\,.$$

The set $\Omega_{(\bar{c}_i,\bar{c})}$ is further defined as $\Omega_{(\bar{c}_i,\bar{c})} = \{\bar{y}|\Pr(\bar{y}|\bar{c}_i) \geq \Pr(\bar{y}|\bar{c}),\ \bar{c}_i \neq \bar{c}\} \supset \Omega_{\bar{c}_i}$,

which represents the space of $\bar{y}$ that will be decoded as $\bar{c}_i$ instead of $\bar{c}$. Hence

$$\begin{aligned} P_{\mathrm{ML,e}} &= \sum_{\bar{y}\in\cup_{\bar{c}_i\neq\bar{c}}\Omega_{\bar{c}_i}} \Pr(\bar{y}|\bar{c}) \\ &= \sum_{\bar{y}\in\cup_{\bar{c}_i\neq\bar{c}}\Omega_{(\bar{c}_i,\bar{c})}} \Pr(\bar{y}|\bar{c}) \\ &\leq \sum_{i=1}^{2^K-1} \sum_{\bar{y}\in\Omega_{(\bar{c}_i,\bar{c})}} \Pr(\bar{y}|\bar{c})\,. \end{aligned}$$



§ 4.6* Maximum Likelihood (ML) Decoding

Under AWGN channel and BPSK modulation,

$$\sum_{\bar{y}\in\Omega_{(\bar{c}_i,\bar{c})}} \Pr(\bar{y}|\bar{c}) = Q\left(\sqrt{\frac{2E_c d_{\text{Ham}}(\bar{c}_i,\bar{c})}{N_0}}\right) = Q\left(\frac{\sqrt{d_{\text{Ham}}(\bar{c}_i,\bar{c})}}{\sigma}\right) = Q\left(\frac{\sqrt{d_{\text{Ham}}(\bar{c}_i,\bar{0})}}{\sigma}\right).$$

pairwise error probability

assuming $E_c = 1$

Then,

$$\sum_{i=1}^{2^K-1} \sum_{\bar{y}\in\Omega_{(\bar{c}_i,\bar{c})}} \Pr(\bar{y}|\bar{c}) = \sum_{\bar{y}\in\Omega_{(\bar{c}_1,\bar{c})}} \Pr(\bar{y}|\bar{c}) + \sum_{\bar{y}\in\Omega_{(\bar{c}_2,\bar{c})}} \Pr(\bar{y}|\bar{c}) + \cdots + \sum_{\bar{y}\in\Omega_{(\bar{c}_{2^K-1},\bar{c})}} \Pr(\bar{y}|\bar{c}).$$

$$E_c = 1 = rE_b \quad \Rightarrow \quad P_{\text{ML,e}} \leq \sum_{d_i=d}^{d_s} \text{A}_{d_i} Q\left(\frac{\sqrt{r d_i E_b}}{\sigma}\right).$$

□

References:

[1] Elements of Information Theory, by T. Cover and J. Thomas.