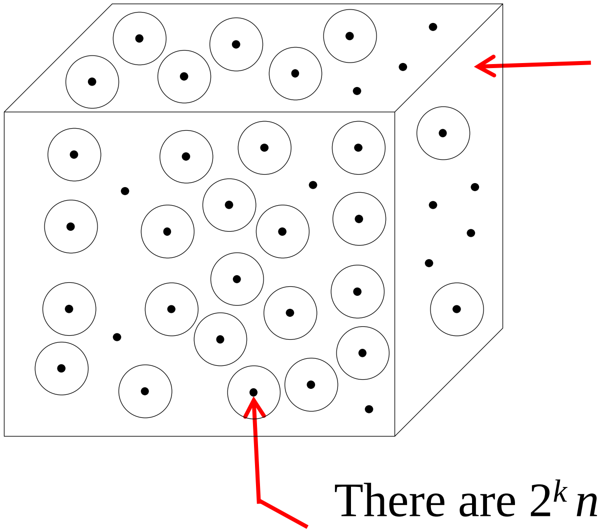
Chapter 4 Channel Coding

- 4.1 An Introduction of Channel Coding
- 4.2 Block Codes
- 4.3 Cyclic Codes
- 4.4 The Parity-Check Matrix

§ 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.



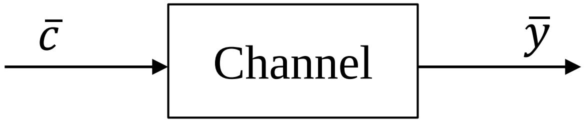
The n-dimensional binary space that can accommodate at most 2^n binary vectors.

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

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

§ 4.1 An Introduction of Channel Coding

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



Aim: with 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

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 $P_e \rightarrow 0$. Inversely, any such codes that realize $P_e \rightarrow 0$ must have rate $r < C$.

- Shannon's channel coding theorem demonstrates error free transmission is possible by manipulating rate of the code according to the channel capacity. It is defined in the mindset of binary transmission, e.g., BPSK.

§ 4.1 An Introduction of Channel Coding

- The proof of **Shannon's Channel Coding Theorem** can be enlightened by the use of **Jointly Typical Sequences**.
- **Jointly Typical Sequences:** Given two sequences *X* (x^n: x_1, x_2, ..., x_n) and *Y* (y^n: y_1, y_2, ..., y_n), they are jointly typical if their empirical entropy is ϵ-closed to the true entropy as

$$\left|-\frac{1}{n}\sum_{i=1}^{n}\log P(x_i,y_i)-H(X,Y)\right|<\epsilon.$$

Note, it is assumed that $P(x_i,y_i)=\frac{1}{n},\forall i.$

- If *X* and *Y* are drawn i.i.d. according to

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

then

- With n → ∞, Pr(*X* and *Y* are jointly typical) → 1.
- If *Z* (z^n: z_1, z_2, ..., z_n) and *Y* are independent, i.e., $P(z^n)$ $P(y^n)$ = $P(z^n, y^n)$, Pr(Z and *Y* are jointly typical) $\leq 2^{-n(I(Z,Y)-3\epsilon)}$.

§ 4.1 An Introduction of Channel Coding

Proof of Shannon's Channel Coding Theorem

- Generate a code of length *n* rate *r* that follows $P(c^n) = \prod_{i=1}^{n} P(c_i)$.
- 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 particular codewords.

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

- It is assumed that codewords are uniformly chosen for transmission, i.e.,

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

- With received vector y^n, the decoder estimates codeword $c^n(\widehat{w})$ such that
  - $c^n(\widehat{w})$ 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.

§ 4.1 An Introduction of Channel Coding

Continue The Proof

- The error probability is

$$P(\varepsilon) = \sum_{\mathcal{C}} \underline{P(\mathcal{C})}\ \underline{P_e(\mathcal{C})}$$

Prob. of a particular code $\mathcal{C}$

Error prob. of the code $\mathcal{C}$

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

Error prob. of a particular codeword $c^n(w) \in \mathcal{C}$

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

- 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,

$$\begin{aligned} P(\varepsilon) &= \sum_{\mathcal{C}} P(\mathcal{C})\, P_{e,1}(\mathcal{C}) \\ &= \underline{P_{e,1}} \end{aligned}$$

Average (over all codebooks) error prob. of codeword $c^n(1)$

§ 4.1 An Introduction of Channel Coding

Continue The Proof

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

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

When n is sufficiently large, $\Pr\left(E_1^C\right) \leq \epsilon$.

Since $c^n(w), w = 2,3,\cdots,2^{nr}$ and y^n are independent, $\Pr(E_w) \leq 2^{-n(I(c^n(w),y^n)-3\epsilon)}$.

- $$\begin{aligned} P(\varepsilon) &\leq \epsilon + \sum_{w=2}^{2^{nr}} 2^{-n(I(c^n(w),y^n)-3\epsilon)} \\ &= \epsilon + (2^{nr} - 1) \cdot 2^{-n(I(c^n(w),y^n)-3\epsilon)} \\ &= \epsilon + 2^{3n\epsilon} 2^{-n(I(c^n(w),y^n)-r)} \end{aligned}$$

§ 4.1 An Introduction of Channel Coding

Continue The Proof

- If n is sufficiently large and $r < I(c^n(w), y^n) - 3\epsilon$,

$$P(\varepsilon) \leq 2\epsilon,$$

  the error probability can be arbitrarily small.

- Choose $P(c_i)$ to be the distribution that maximizes $I(c^n(w), y^n)$ as

  $C = \max\limits_{P(c_i)}\{I(c^n(w), y^n)\}$, the above conclusion implies if $r < C$, the error probability can be arbitrarily small. The proof is completed.

§ 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.

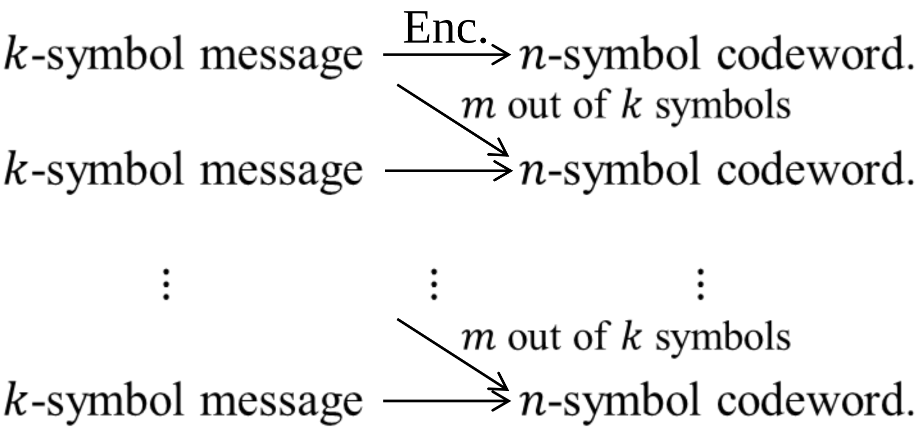
1. Block codes:

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

- 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. $(a, b) \in \mathrm{F}_q$ 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-symbol message $\xrightarrow{\text{Enc.}}$ n-symbol codeword.

m out of k symbols

k-symbol message $\longrightarrow$ n-symbol codeword.

$\vdots$ $\vdots$ $\vdots$

m out of k symbols

k-symbol message $\longrightarrow$ n-symbol codeword.

- Encoder has a memory of order *m* and 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

Start with Error-Dectection:

The simplest class of block code is the **parity-check code**, which cannot correct errors but can **detect** a single error.

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

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

§ 4.2 Block Codes

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

$$\bar{c} = \bar{m} \cdot \mathbf{G}.$$

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

G — a generator matrix of size *k* × *n*. It defines the legal space among all *n*-dimensional vector.

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

Linear block code:

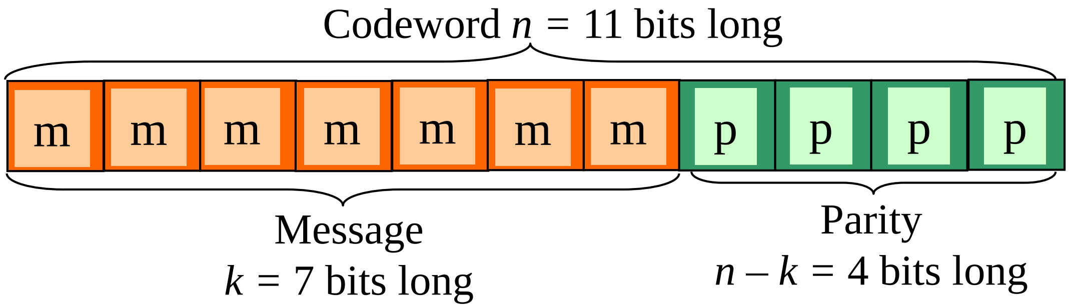
$$\bar{c}_1 = \bar{m}_1 \cdot \mathbf{G}$$

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

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

§ 4.2 Block Codes

Hamming Distance



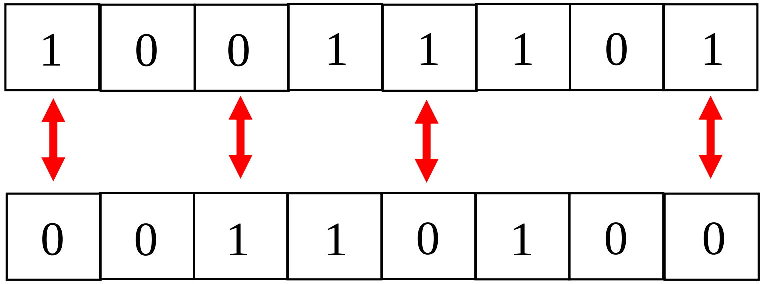
Codeword n = 11 bits long

Message
k = 7 bits long

Parity
$n - k$ = 4 bits long

m = 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.

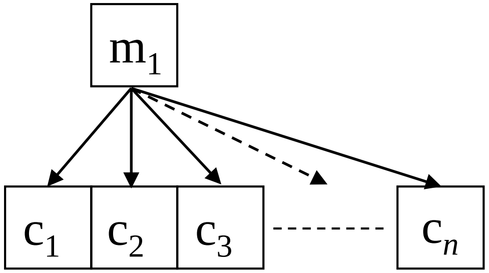
§ 4.2 Block Codes

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 $d = n$
- Hence it is an $(n, 1, n)$ block code



The generator matrix of the code is simply

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

§ 4.2 Block Codes

Majority Decoding

To recover the transmitted codeword of a repetition code, a simple decoder known as a **Majority Decoder** is 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: Say our message bit was a 1 and it was encoded by the (5, 1, 5) repetition code then the codeword will be **c** = 11111.

- If after transmission we receive the word **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 **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.

§ 4.2 Block Codes

The Minimum Hamming Distance and Error Correction of a Block Code

Take the (3, 1, 3) repetition code with codewords 000 and 111

If we add **one** error, the possible received words are

Codewords	
000	111
001	**110**
010	**101**
100	**011**

A majority decoder will be able to recover the correct message.

If we add **two** errors, the possible received words are

Codewords	
000	111
011	**100**
110	**001**
101	**010**

Message recovered by the majority decoder will not be correct.

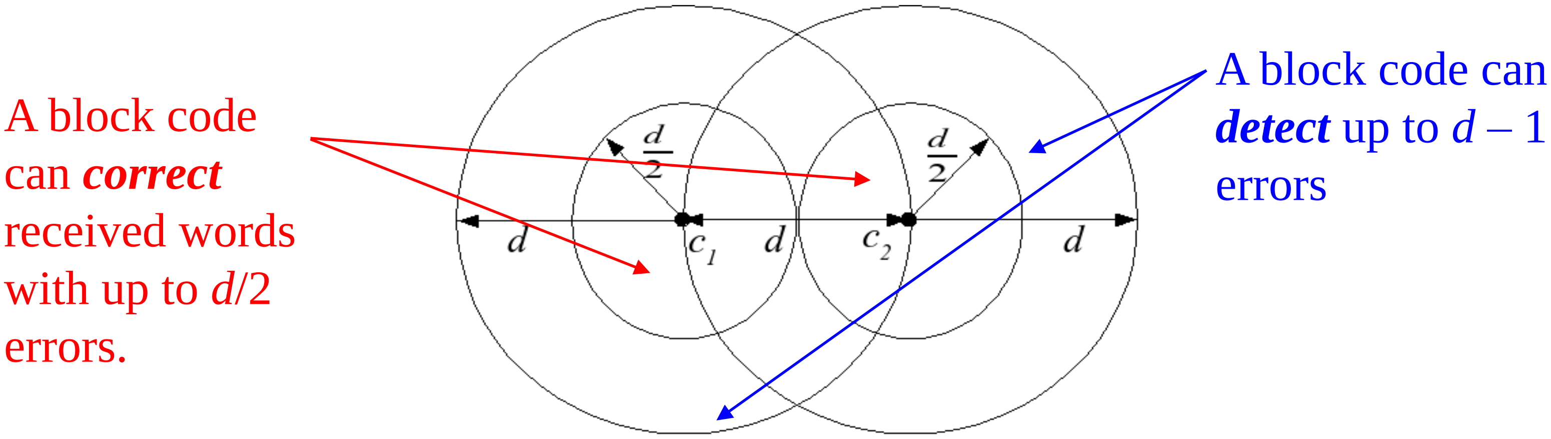
§ 4.2 Block Codes

The Minimum Hamming Distance and Error Correction Capability

The minimum Hamming distance: for any two codewords c_i and c_j picked up from the codebook $\mathcal{C}$, the minimum Hamming distance d is defined as:

$$d = \min_{(c_i, c_j) \in \mathcal{C}} \{d_{\text{Ham}}(c_i, c_j)\}.$$

- 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 *d*/2 errors.
$\frac{d}{2}$
$\frac{d}{2}$
d
c_1
d
c_2
d
A block code can ***detect*** up to *d* – 1 errors

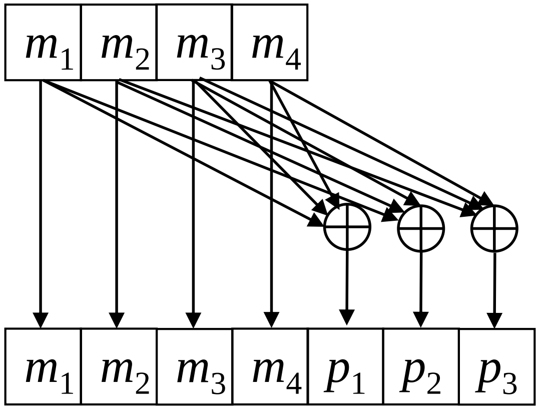
- For a linear block code, $d = \min\{\text{weight}\,(\bar{c}_j), \bar{c}_j \neq 0\}$.

§ 4.2 Block Codes

The (7, 4, 3) Hamming Code

This code can correct 1 error.

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



(7, 4, 3) Hamming Code

The parity bits are calculated by

$$p_1 = m_1 \oplus m_3 \oplus m_4$$
$$p_2 = m_1 \oplus m_2 \oplus m_3$$
$$p_3 = m_2 \oplus m_3 \oplus m_4$$

The encoding can be written as

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

and

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

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

Message	Codeword
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

§ 4.3 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

- A 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

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

$$m(x) = m_0 + m_1 x + \cdots + m_{k-1} x^{k-1}$$

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

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

§ 4.3 Cyclic Codes

Cyclic Hamming Code

- The (7, 4, 3) Hamming code is actually a cyclic code and can be constructed using the generator polynomial $g(x) = x^3 + x^2 + 1$.

- For example, to encode the binary message 1010 we first write it as the message polynomial $m(x) = x^3 + x$ and then multiply it with $g(x)$ modulo-2

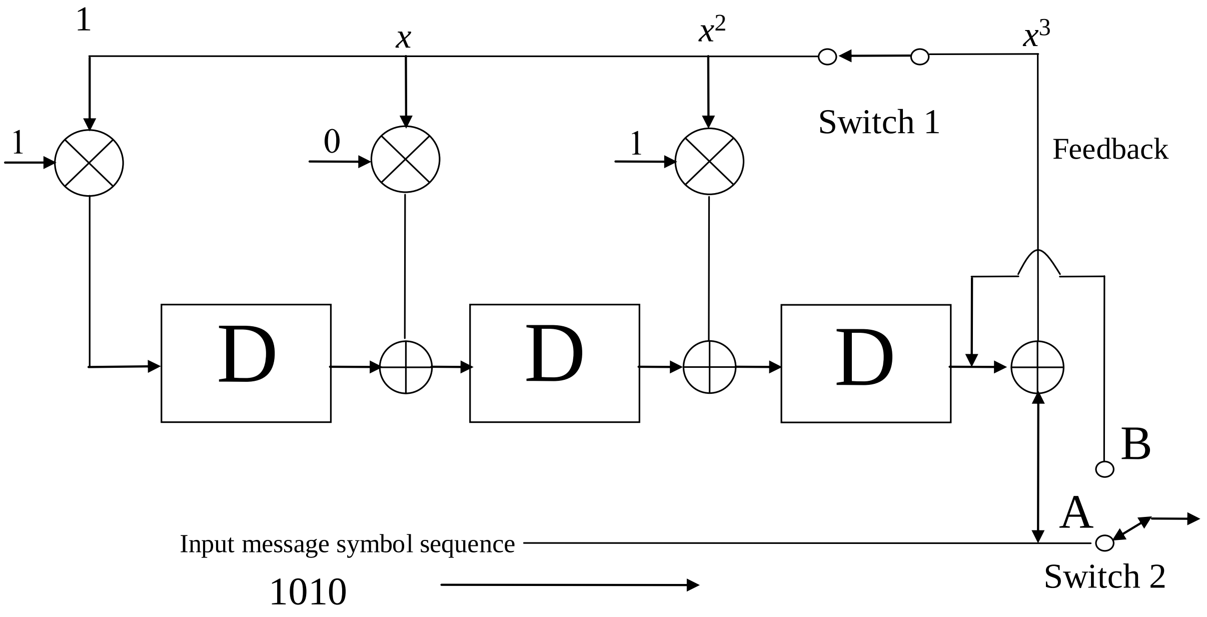
$$\begin{aligned} c(x) &= m(x)g(x) \\ &= (x^3 + x)(x^3 + x^2 + 1) \\ &= x^6 + x^5 + x^3 + x^4 + x^3 + x \qquad [(x^3 + x^3) \bmod 2 = 2x^3 \bmod 2 = 0] \\ &= x^6 + x^5 + x^4 + x \end{aligned}$$

This codeword polynomial corresponds to 1 1 1 0 0 1 0

- However, notice that the first four bits of this codeword are not the same as the original message 1010
- This is an example of a **non-systematic code**

§ 4.3 Cyclic Codes

Systematic Cyclic Hamming Code



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

§ 4.3 Cyclic Codes

- This shift register encoding is equivalent to the systematic block code encoding.

Let the input message be $\bar{m} = (m_1, m_2, m_3, m_4)$

Input	Registers (left to right)		
m_1	m_1	0	m_1
m_2	$m_1 \oplus m_2$	m_1	$m_1 \oplus m_2$
m_3	$m_1 \oplus m_2 \oplus m_3$	$m_1 \oplus m_2$	$m_2 \oplus m_3$
m_4	$m_2 \oplus m_3 \oplus m_4$	$m_1 \oplus m_2 \oplus m_3$	$m_1 \oplus m_3 \oplus m_4$

Hence, $p_1 = m_1 \oplus m_3 \oplus m_4$

$p_2 = m_1 \oplus m_2 \oplus m_3$

$p_3 = m_2 \oplus m_3 \oplus m_4$

§ 4.4 The Parity-Check Matrix

- We need to know when a codeword is valid.
- 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, $\mathbf{GH}^T = \mathbf{0}$ (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.,

$$\bar{c}\cdot\mathbf{H}^T = \bar{m}\cdot\mathbf{G}\cdot\mathbf{H}^T = \underline{0}.$$

Syndrome vector.

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

§ 4.4 The Parity-Check Matrix

- 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 **P** is a parity matrix, then the parity-check matrix is of the form $\mathbf{H} = [\mathbf{P}^T \mid \mathbf{I}_{n-k}]$.
- Dual code property

$$\begin{tikzcd}
 & \text{Generator matrix} & \text{Parity-check matrix} \\
\text{Code} & \mathbf{G}_{k\times n} \arrow[dr] & \mathbf{H}_{(n-k)\times n} \\
\text{Dual code} & \mathbf{G}_{(n-k)\times n} \arrow[ur] & \mathbf{H}_{k\times n}
\end{tikzcd}$$

Taking the (7, 4, 3) Hamming code

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

The parity-check matrix is $\longrightarrow$

$$\mathbf{H} = \begin{array}{c} \begin{array}{cc} \mathbf{P}^T & \mathbf{I}_{n\text{-}k} = \mathbf{I}_{7\text{-}4} = \mathbf{I}_3 \end{array} \\ \left[\begin{array}{cccc:ccc} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right] \end{array}$$

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

References:

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