Class no 20 (?) [22nd March]:

Cover & Thomas [Chap 8]: "Channels & Channel Coding":

- Quick Recap:

$X, Y \sim P_{X,Y}$ → $\sum_x P_{X,Y}(x,y) = P_Y(y)$; → $\sum_y P_{X,Y}(x,y) = P_X(x)$

$H(X) \triangleq -\sum_{x \in \text{supp}(P_X)} p(x) \log p(x)$

$X \in \mathcal{X}$, $Y \in \mathcal{Y}$

($\mathcal{X}$: set of all values that X can take.)

$\text{supp}(P_X) = \{x \in \mathcal{X} : P_X(x) \neq 0\}$

$H(X,Y) \triangleq -\sum_{(x,y) \in \text{supp}(P_{X,Y})} p(x,y) \log p(x,y)$

($p(x,y)$: prob that $X = x$ & $Y = y$)

$= H(X) + H(Y|X)$

$p(x,y) = p_X(x)\, p_{Y|X}(y|x)$

$= H(Y) + H(X|Y)$

$H(X_1, \ldots, X_n) = H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + \ldots + H(X_n|X_1, \ldots, X_{n-1})$

$H(Y|X) \triangleq$

Fix x such that $P_X(x) \neq 0$: $\{P_{Y|X}(y|x) : y \in \mathcal{Y}\}$

$\sum_y p_{Y|X}(y|x) = 1$

Thus: $P_{Y|X}(\cdot|x)$ is a prob dist for Y. (Cond dist'n of Y given $X = x$)

$P_{Y|X}$ → This alone refers to a collection of cond. distri. on Y gn X (for each $x \in \mathcal{X}$ s.t $p(x) \neq 0$, there is a distri $p_{Y|X}(-|x)$.)

$$H(Y|X) \triangleq \sum_x p(x) \underbrace{H(Y|X=x)}_{\text{Cond entropy of } Y \text{ gn } X=x}$$

$$= -\sum_y p(y|x) \log p(y|x)$$

$E[g(x)] = \sum p(x) \cdot g(x)$

$g(x) \triangleq H(Y|X=x)$

$$E[g(X)] = \sum_x p_X(x)\, g(x)$$

$$H(X_1, X_2 | Z_1, Y) = \sum_{z_1, y} p_{Z_1, Y}(z_1, y) \underbrace{H(X_1, X_2 | Z_1 = z_1, Y = y)}$$

Where

$$H(X_1, X_2 | Z_1 = z_1, Y = y) = -\sum_{x_1, x_2} p(x_1, x_2 | z_1, y) \log p(x_1, x_2 | z_1, y)$$

$I(X;Y)$ = info abt X having obs Y = $H(X) - H(X|Y)$ → "Uncertainty abt X − "Uncertainty abt X having observed Y"

$= H(Y) - H(Y|X)$

$I(Y;X)$ = info abt Y having observed X

$$I(X_1, Y_1 ; Z_1, X_2) = H(X_1, Y_1) - H(X_1, Y_1 | Z_1, X_2)$$

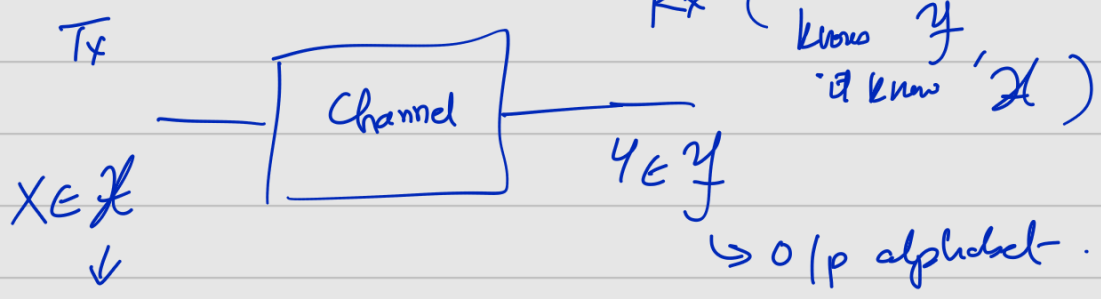
$$= H(Z_1, X_2) - H(Z_1, X_2 | X_1, Y_1)$$

$$\neq H(Y_1, Z_1) - H(Y_1, Z_1 \mid X_1, X_2)$$

$$I(X; Z_1 Z_2 Z_3 Z_4) = H(X) - H(X \mid Z_1, Z_2 Z_3 Z_4)$$

$$= H(Z_1 Z_2 Z_3 Z_4) - H(Z_1 Z_2 Z_3 Z_4 \mid X)$$

Channels:



→ Rx wants info abt X

Rx also knows $\{ P_{Y|X}(\cdot|x) : \forall x \}$

"The chance of getting $Y = y$, on $X = x$, is known to the Rx ; $\forall x, y \in \mathcal{X} \times \mathcal{Y}$"

→ Assume $\mathcal{X}, \mathcal{Y}$ are discrete RVs, & finite ($|\mathcal{X}| < \infty$, $|\mathcal{Y}| < \infty$)

$$P_{Y|X} = \begin{bmatrix} & (x,y)^{th} \text{ entry} \\ & P_{Y|X}(y|x) \\ & \downarrow \\ & y \in \mathcal{Y} \end{bmatrix} \quad \downarrow x \in \mathcal{X}$$

→ each row sums to 1

$$\sum_y p(y|x) = 1$$

Prob. transition matrix of the Channel

Channel: This 3-tuple $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$ → Channel. Discrete Memoryless (DMC)

$\underline{Y}^n = (Y_1, \ldots, Y_n)$, $\underline{X}^n = (X_1, \ldots, X_n)$

$(y_1, y_2, \ldots, y_n)$ ← $(x_1, \ldots x_n)$

$$P_{\underline{Y}^n|\underline{X}^n}(\underline{y}|\underline{x}) = \prod_{i=1}^{n} P(y_i|x_i)$$

↓

When this equ is true for all $\underline{x}, \underline{y}$

then we say that the channel is "MEMORYLESS"

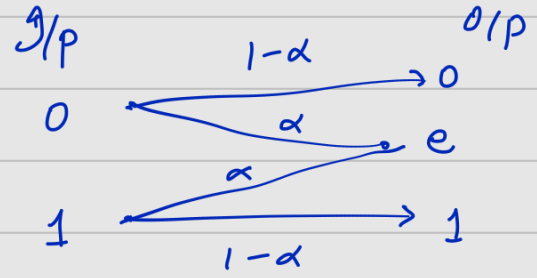
Channel capacity:

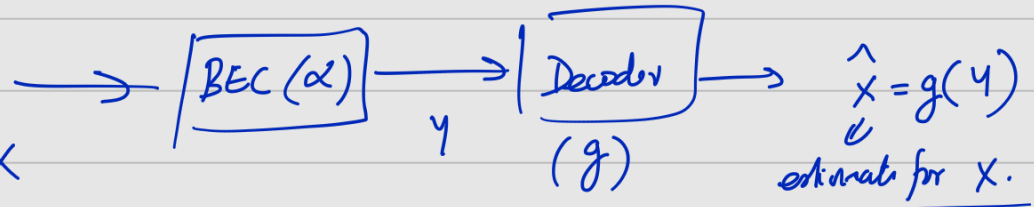
DMC

Example of channel:

(a) Binary Erasure channel.

$\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{0, 1, e\}$ (↓ erasure symbol.) , $P_{Y|X} = \begin{bmatrix} P(0|0) = 1-\alpha & 0 & P(e|0) = \alpha \\ 0 & 1-\alpha & \alpha \end{bmatrix}$, 2×3





$P_r(\text{error})$ (input $\in \{0,1\}$)

$= \sum_{x \in \{0,1\}} P_X(x)\, P_r(\text{error} \mid X = x)$

Expected / Avg Prob of error.

$g(0) = 0$

$g(1) = 1$

$g(e) =$ "Unable to decode".

$$P_r(\text{error} \mid X=0) = p(\hat{X} \neq X \mid X=0)$$

$$= p(\hat{X} \neq 0 \mid X=0)$$

$$= P(Y = e \mid X=0) = \alpha$$

$$P(\text{error} \mid X=1) = \alpha.$$

$$\Rightarrow P_r(\text{error}) = \sum_{x \in \{0,1\}} P_X(x)\, \alpha = \alpha \underbrace{\sum_{x \in \{0,1\}} P_X(x)}_{(=1)} = \alpha.$$

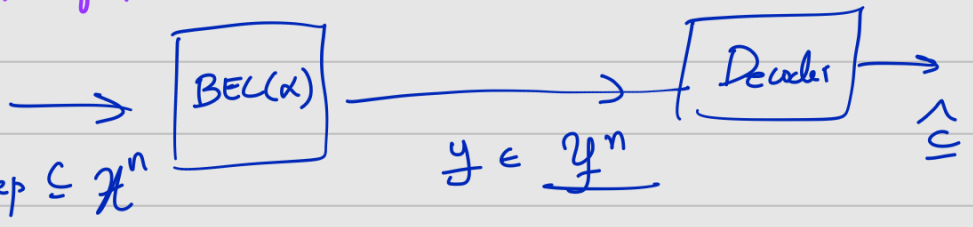
Repetition Coding:

Consider the following "channel code".

$$\mathcal{C}_{\text{repetition}} \triangleq \{\underbrace{(0,\ldots,0)}_{n\text{-length tuple}}, (1,\ldots,1)\} \subseteq \{0,1\}^n$$

The tuples $(0,\ldots,0)$ & $(1,\ldots,1)$ are called "codewords" of $\mathcal{C}_{\text{repetition}}$. Rate of the code $= \frac{\log_2 |\mathcal{C}_{\text{rep}}|}{n} = \frac{1}{n}$

$n \rightarrow$ blocklength of the code used.



Ex: $\underline{c} = (0 \cdots 0)$

$\underline{y} = (0 \cdots 0)$

$P(\underline{y} = (0 \cdots 0) \mid \underline{c} = (0 \cdots 0))$

$\underline{y|c}$

$= \prod_{i=1}^{n} P_{y_i|c_i}(0|0)$

$= (1-\alpha)^n$

$\hookrightarrow (e, 0, \cdots, 0)$

$P_{\underline{y|c}}((e, 0 \cdots 0) \mid \underline{c} = (0, \cdots, 0))$

$= p(e|0) \cdot (p(0|0))^{n-1} = \alpha(1-\alpha)^{n-1}$

$\hookrightarrow (e, \cdots, e) \rightarrow P_{\underline{y|c}}((e \cdots e) \mid (0 \cdots 0))$

$= \alpha^n.$

$P(\underline{y}|\underline{c}) = (1-\alpha)^{(n - n_e(\underline{y}))} \alpha^{n_e(\underline{y})}$

$\underline{y|c}$

($n_e(\underline{y})$ ≜ no. of errors in $\underline{y}$).

(Prob that $\underline{c} \in \mathcal{C}_{rep}$ was the i/p to the channel)

$P_r(\text{error}) = \sum_{\underline{c} \in \mathcal{C}_{rep}} P_{\underline{X}}(\underline{c})\, P(\text{error} \mid \underline{X} = \underline{c})$

$P(\text{error} \mid \underline{X} = (0 \cdots 0))$

$= \alpha^n.$

$P(\text{error} \mid \underline{X} = (1, \cdots, 1)) = \alpha^n$

$\Rightarrow P_r(\text{error}) = \alpha^n \rightarrow$ decreases (goes to 0) exponentially in n.

$g(\underline{y}) = \underline{0}$, when at least one coordinate of $\underline{y}$ is 0.

$g(\underline{y}) = \underline{1}$, when $\cdots$ $\cdots$ of $\underline{y}$ is $\underline{1}$.

$g(\underline{y}) =$ FAIL if $\underline{y} = (e, \cdots, e)$.

More generally: We could use, instead of $\mathcal{C}_{rep}$, some other channel code $\mathcal{C} \subseteq \mathcal{X}^n$.

$\rightarrow$ The rate of this code $\triangleq \dfrac{\log_2 |\mathcal{C}|}{n}$ bits/ch-use

$\rightarrow$ The blocklength of $\mathcal{C} = n$.

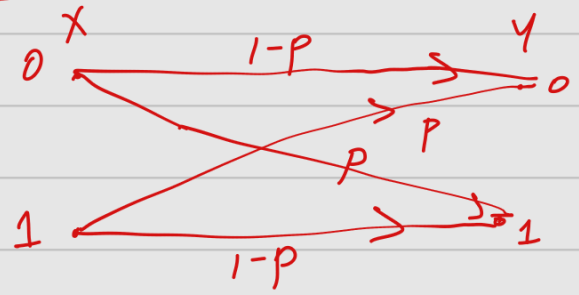
$\rightarrow$ It will have some prob of error.

$$P(\text{error}) = \sum_{\underline{c} \in \mathcal{C}} P_{\underline{X}}(\underline{c})\, P(\text{error} \mid \underline{X} = \underline{c})$$

where $P(\text{error} \mid \underline{X} = \underline{c})$ $\rightarrow$ depends on the decoding fn g, & also the channel transition matrix.

We could assume: Given a particular blocklength n.
By using codes of larger size, we get higher rate codes for communication, but possibly higher error prob.
& vice-versa for smaller size codes.

Binary Symmetric Channel: –



Let $p \in [0,1]$.

$$P_{Y|X}(0|1) = P_{Y|X}(1|0) = p.$$

$$P_{Y|X}(1|1) = P_{Y|X}(0|0) = 1-p.$$

Prob of error calculations: Try as before for uncoded transmissions & for repetition code of length n.

Linear Codes:

In a general code $\mathcal{C} \subseteq \mathcal{X}^n$,

"encoding of information-vector / msg-vector"

Tx maintains this table:

Msg vectors of length $\log_2 \lvert\mathcal{C}\rvert$ over $\{0,1\}$.	Codewords from $\mathcal{C}$
0 0	0 1 0 1
0 1	0 0 0 1
1 0	0 1 1 1
1 1	1 0 1 1

Tx:
→ Has a message vector of length $\log_2 \lvert\mathcal{C}\rvert$.
→ Finds corr codeword from this table
→ Transmit that codeword

Message vector / Info vector

Transmitted codeword

For a binary-channel, $R(\mathcal{C}) \leq 1$

as $\log_2 |\mathcal{C}| \leq n$

(as $\mathcal{C} \subseteq \{0,1\}^n$

$|\mathcal{C}| \leq 2^n$).

A linear code is one in which the encoding of the msg vector into a codeword is a linear operation (i.e. a matrix multiplication).

For example:

Consider $G = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}_{2 \times 4}$

$\log_2 |\mathcal{C}| \times n$

Suppose msg vector is $\underline{m} = (0,1)$

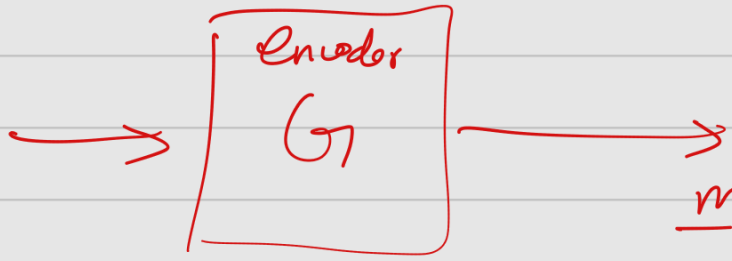
Then corresponding transmitted codeword is

$$\underline{m}\, G = (0,1) \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} = (0\,1\,0\,0)$$

$\underline{m}$: $1 + \log_2 |\mathcal{C}| \to \log_2 |\mathcal{C}| \times n \longrightarrow 1 \times n$.

Source → [Source coding] → [Ch. coding] → [Ch]

Encoding of linear code



$\underline{m}$

msg vector

info vector

length $\log_2 |\mathcal{C}|$

$\underline{m}\, G_1$.

length n

(codeword)

G_1 is called the generator matrix of the linear code.

The linear code generated by G_1 = $\{ \underline{m} G_1 : \forall \underline{m} \in \{0,1\}^{\log_2 |\mathcal{C}|} \}$.

If $G_1 = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$

$(0,0)\, G_1 \xrightarrow{(\text{mod } 2)}$ $(0,1)\, G_1 \xrightarrow{(\text{mod } 2)}$ $(1,0)\, G_1 \xrightarrow{(\text{mod } 2)}$ $(1,1)\, G_1$

$\{ (0,0,0,0), (0,1,1,0), (1,0,1,1), (1,1,0,1) \}$.

When we talk abt linear codes,

$\log_2 |\mathcal{C}|$ represents no. of rows in G_1.

hence it must be an integer always.

$k \triangleq \log_2 |\mathcal{C}|$ for linear codes.

$\Rightarrow$ Gen matrix is a matrix of size $k \times n$.

$\underline{m} \in \{0,1\}^k$

$\underline{c} \in \{0,1\}^n$

$\underline{c} = \underline{m}\, G$ is the code word for msg $\underline{m}$.

Observation:

$$\underline{m}\, G = (m_1, \ldots, m_k) \begin{bmatrix} \underline{g_1} \\ \vdots \\ \underline{g_k} \end{bmatrix}$$

$$= \sum_{i=1}^{k} m_i \underline{g_i} \rightarrow$$

$$\Rightarrow \boxed{\mathcal{C} = \text{rowspace}(G)} \quad (\text{with } \{0,1\} \text{ operations mod 2}).$$

Now: suppose G had rank k.

The rank of G is the no: of linearly independent rows of G.

$$\begin{bmatrix} \underline{g_1} \\ \vdots \\ \underline{g_k} \end{bmatrix}$$

$\underline{g_1}, \ldots, \underline{g_k}$ are l.i if the following condition holds.

① $\sum_{i=1}^{k} \alpha_i \underline{g_i} = \underline{0}$ ONLY IF

$\alpha_i = 0$, $\forall i$

NOTE THAT $\alpha_i \in \{0,1\}$ here $\forall i$

& additions are mod 2.

Now:

What guarantees that no two distinct msg vectors get same codeword?

(Otherwise, encoding will not be one-one).

Ans: Because G_1 is assumed to have rank k (or full row-rank).

Proof: Suppose for $\underline{m} \neq \underline{m}'$, $\underline{m}, \underline{m}' \in \{0,1\}^k$,

$$\underline{m} G_1 = \underline{m}' G_1$$

$$\Rightarrow (\underline{m} - \underline{m}') G_1 = \underline{0} \in \{0,1\}^n$$

As $(\underline{m} - \underline{m}') \neq \underline{0} \in \{0,1\}^k$,

$\Rightarrow$ rows of G_1 are NOT L.I.

Thus, if $\text{rank}(G_{k \times n}) = k$, then encoding is one-one operation.

Ex: Consider code $\mathcal{C}$ gen. by

$$G_1 = \left[\begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right]_{4\times 7}$$

(0 1 1 0)

→ Full row rank because $\underline{m}G = \underline{0}$ happens only if $\underline{m} = \underline{0}$.

Note: $R(\mathcal{C}) = 4/7 = \frac{\log_2|\mathcal{C}|}{n} = \boxed{k/n}$.

Find: codeword for

$\underline{m}$ =	$\underline{m}G$
$(0,1,1,0)$	$(0110\,110)$
$(1,1,0,1)$	$(1101\,010)$

msg → tr^{x} codeword

Find: $\mathcal{C}$ (16 codewords coming from $\forall\, \underline{m} \in \{0,1\}^4$)

$|\mathcal{C}| = 16$. (Rate: $\frac{\log_2|\mathcal{C}|}{n}$ bits/ch-use $= 4/7$).

Find: Rank (G_1) .

≤ 4

Observe that: $G_1 = [I_4 \mid P]$ where $P = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

I_4: 4×4, $[I_4 \mid P]$: 4×7, P: 4×3

Now, $\underline{m} = (m_1 \cdots m_4)$ → $(c_1, \ldots, c_7)$

$$(m_1\, m_2\, m_3\, m_4)\,[I_4 \mid P] = (m_1, m_2, m_3, m_4, \underline{m_1+m_3+m_4}, m_1+m_2+m_4, m_2+m_3+m_4)$$

Observe that, in any arbitrary codeword: $c_5 + c_1 + c_3 + c_4 = 0 \rightarrow (1)$

$$c_6 + c_1 + c_2 + c_4 = 0 \quad \rightarrow (2)$$

$$c_7 + c_2 + c_3 + c_4 = 0 \quad \rightarrow (3)$$

(1), (2), (3) $\rightarrow$ "Parity check" equations [satisfied by all codewords in the code gen by G_1]

Hence:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \end{array}\right] \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \\ c_7 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

P^T | I_3 $\quad 3 \times 7$

$$\underset{3\times 7}{\left[P^T \mid I_3\right]} \underset{7\times 1}{\underline{c}^T} = \underset{3\times 1}{\underline{0}} \quad \rightarrow$$ Parity check eqns in matrix form.

(Whenever $G_1 = [I_k \mid P]_{k \times n}$,

we get these eqns of the form $\underbrace{[P^T \mid I_{n-k}]}_{(n-k)\times n} \underline{c}^T = \underline{0}$

This matrix $[P^T \mid I_{n-k}]$ is called the parity check matrix for code $\mathcal{C}$ generated by $G_1 = [I_k \mid P]$.

H

$$H \underline{c}^T = \underline{0} \rightarrow \text{P.C. eqns.}$$

Remark: The set $\{\underline{x} : H \underline{x}^T = \underline{0}\}$ is exactly the code $\mathcal{C}$.

Proof is by Rank-Nullity theorem.

(Surely $\mathcal{C} \subseteq \{\underline{x} : H\underline{x}^T = \underline{0}\}$ ($\to NS(H)$). Now $|\mathcal{C}| = 2^k$. But rank$(H) = n-k$.
$\Rightarrow$ Nullity$(H) = (n-(n-k)) = k \Rightarrow |NS(H)| = 2^k \Rightarrow \mathcal{C} = NS(H)$)

Thus: $\underline{x} \in \{0,1\}^n$ is a codeword **if and only if**

$$H\underline{x}^T = \underline{0}.$$

Now: consider erasures.

holds for every codeword.

$$\begin{cases} c_5 + c_1 + c_3 + c_4 = 0 \to (1) \\ c_6 + c_1 + c_2 + c_4 = 0 \to (2) \\ c_7 + c_2 + c_3 + c_4 = 0 \to (3) \end{cases}$$

If decoder is able to 'reconstruct' the erased coordinates correctly, then decoder gets the true codeword, & then the true message.

By eqn (1),

$$c_4 = \underbrace{c_5 + c_1 + c_3}_{\text{known}}$$

$\downarrow$

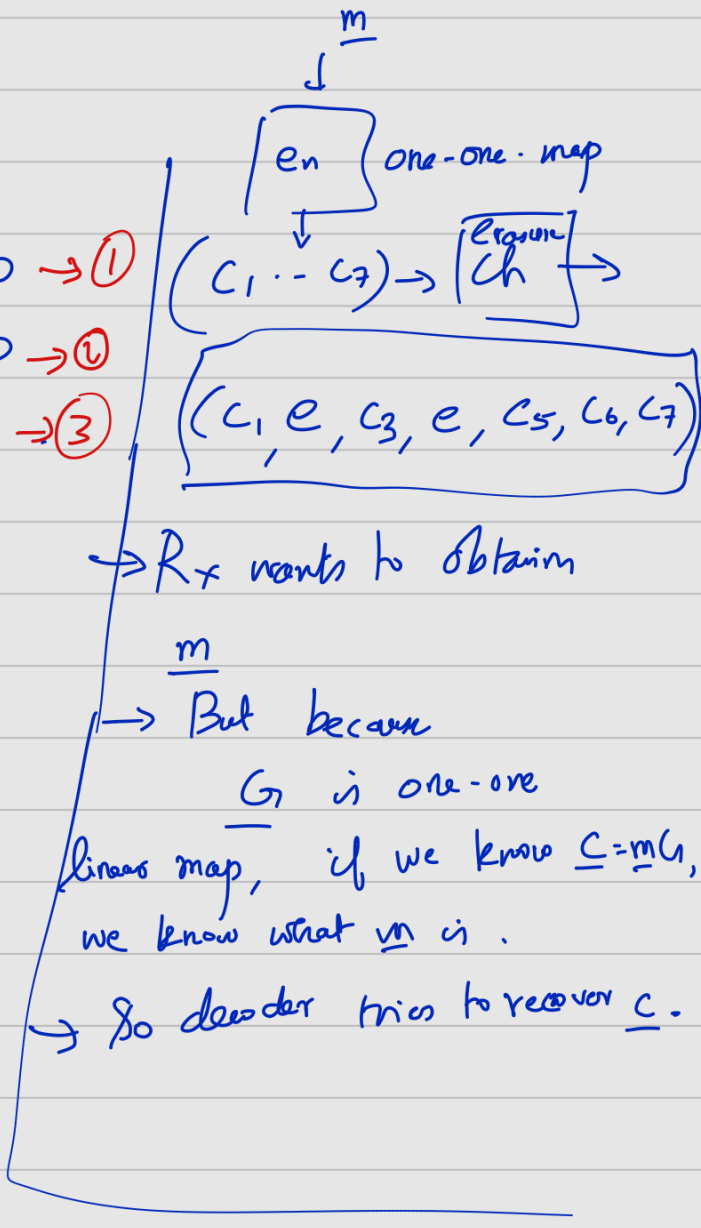
can be decoded.

Now use (2) to get $c_2 = \underbrace{c_6 + c_1 + c_4}_{\text{known + decoded}}$

$\downarrow$

can be decoded.

Thus Rx can find c_2 & c_4.

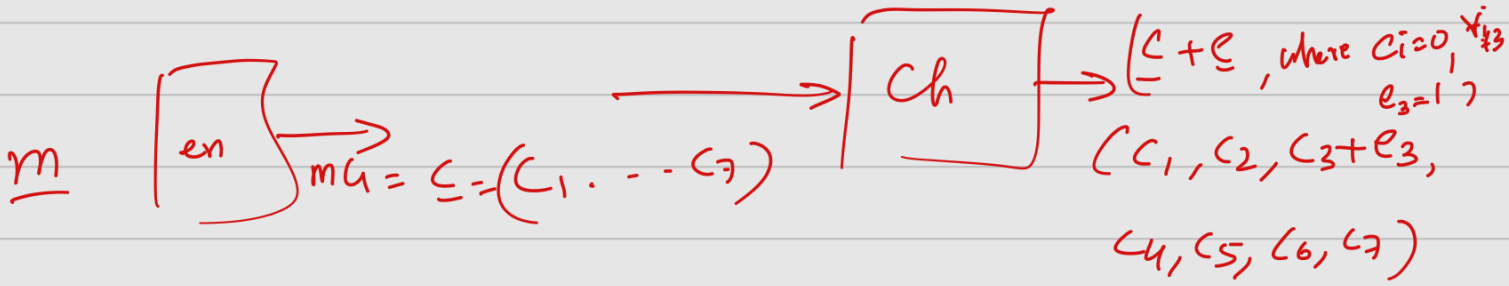


Similar process for any two or lesser erasures

(any 2 or fewer symbols can be erased).

(More than 2 erasures → sometimes correctable sometimes not).

Errors:-



Rx does not know e (where errors happened). ← (Assume at 3rd position the channel flipped the bit, i.e $e_3 = 1$).

Now: Rx wants to find c (true codeword) & hence true message.

Recall, for any $\underline{c} \in \mathcal{C}$,

$$H\,c^T = \underline{0}$$

$$= \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} c^T = \underline{0}.$$

Assume Rx does

$$H \left[\text{received vector } \underline{y} \right]^T = H \left(\begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \cdot \\ c_7 \end{bmatrix}_{\underline{c}} + \begin{bmatrix} 0 \\ 0 \\ e_3 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}_{\underline{e}} \right)^T$$

$$= H\underline{c}^T + H\underline{e}^T = \underline{0} + H\underline{e}^T = H\underline{e}^T$$

Now Rx's job is to find $\underline{e}$.

$$\begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} e_1 \\ \vdots \\ e_7 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

↑ thus obtained by Rx

↳ Assuming the scenario that only one error happened (i.e. only one of the elements of $\underline{e}$ is non-zero), then Rx can correctly identify that 3rd bit of $\underline{e}$ was the error bit.

⇒ Estimate of the true codeword is $\underline{y} - \underline{e} =$ (flip only 3rd bit of $\underline{r}$)

Similarly, as long as only single error happened, decoding (the above) will always be correct in finding the true codeword.

↓

Since $G_{k \times n}$ is $[I \mid P]$ form, the first k bits of the decoded codeword estimate is precisely the msg vector of length k.

$$\underline{m} G = \underline{m} [I \mid P] = [\underline{m} \mid \underline{m} P] = \underline{c}$$

(Remark: Observe that this particular example has higher rate ($4/7$) than a repetition code of length 3, but probably a lower error capability)

Intuitively it seems higher rate codes have larger P(error)

& Lower rate codes have lower P(error)

However, Shannon's channel coding theorem proved a counterintuitive result.

NOT recorded:

Remark: A note on the terminology associated with codes & linear codes.

① Code of length n: → a subset of n-length vectors over $\{0,1\}$.
$\mathcal{C} \subseteq \{0,1\}^n$.

② Codeword of a code: Each element of $\mathcal{C}$.
$\underline{c} = (c_1, \ldots, c_n) \in \mathcal{C}$

③ Vector of length n: $\underline{v} = (v_1, \ldots, v_n) \in \{0,1\}^n$. May or may not be a codeword of $\mathcal{C}$.
(binary)

④ Linear code $\mathcal{C}$ generated by $G_{k \times n}$: → rowspace (G) = all possible binary linear combinations of rows of G = $\left\{ \underset{1\times k}{\underline{m}} \; \underset{k\times n}{G} : \forall \underline{m} \in \{0,1\}^k \right\}$

(5) Encoding operation: $\underset{1\times k}{\underline{m}} \times \underset{k\times n}{G} = \underline{m}G = \underline{c}$

($\underline{c}$: codeword associated with message $\underline{m}$.)

(6) Received vector: Output of Channel $= \underline{y}$, when input is some codeword.

(7) Decoding function: Takes $\underline{y}$ as input and returns some codeword from $\mathcal{C}$ as its estimate. The message vector estimate can be calculated from this codeword estimate.

(8) Dimension of the linear code $\mathcal{C}$: = Rank of generator matrix G
= No. of linearly independent rows (or cols.) of G.

(9) Parity check matrix: A matrix H of size $(n - \text{rank}(G)) \times n$ such that $\underset{k\times n-k}{GH^T} = \underset{k\times n-k}{0}$

(10) Parity check equations: $n-k$ homogeneous linear equations satisfied by all codewords of $\mathcal{C}$.
(H matrix can be derived from these equations)

(11) Linear code $\mathcal{C}$ from its H matrix: $\mathcal{C} = \{\underline{x} \in \{0,1\}^n : \underset{n-k\times n}{H}\underline{x}^T = \underline{0}_{n-k\times 1}\}$.

Shannon's channel Coding theorem: (circa 1948). This theorem has 2 parts. called (1) Achievability
(2) Converse:

— Achievability of any rate $R < C$ (the channel capacity):

Let $(\mathcal{X}, \mathcal{Y}, P_{y|x})$ be a given discrete memoryless channel.
($\mathcal{X}$: i/p alphabet, $\mathcal{Y}$: o/p alphabet, $P_{y|x}$: Channel probabilities (Cond. distribution of O/p RV Y given i/p RV X))

Let $C \triangleq \max_{P_X} I(X;Y)$ → (to be explained later) denote its "capacity". (also called Shannon capacity, or simply "Channel" capacity)

Then, for any value $R \in (0, C)$, there exists a code of rate $\geq R$ with some large block length n, such that the probability of decoding error of this code is $2^{-O(n)}$.

(note that, as n increases, this goes to 0).

Converse:

If we use any code with rate $R > C$, then the probability of decoding error of using this code can never be made arbitrarily small, i.e., $P(\text{error}) >$ some constant independent of n.

→ These two statements together form the statement of Shannon's Channel Coding Theorem

& Remarks about the theorem:

(1) The remarkable result of the theorem shows that it is possible to simultaneously achieve rates close to the channel capacity ('arbitrarily close' $\Leftrightarrow$ as close as we want, say $R = C - \delta$ for any small $\delta > 0$), but at the same time have as low probability of error as one wants (any small $P_e > 0$).

(2) However, the caveat is that such a code as above exists for possibly large blocklength n.

(3) For fixed (possibly small) n, the Shannon Channel Capacity theorem does not assure us of similar result, but other modern researchers have results which show that, for fixed n, there is a tradeoff between achievable rate & probability of error.

(4) The achievability part of the SCT guarantees only the "existence" of a 'good' code. The actual construction (i.e. showing

how to obtain the code, do the encoding/decoding etc.) of such good codes have been done in the recent past few decades.

(5) The converse part of the theorem says that achieving rates higher than capacity is not possible if we want arbitrarily low probability of error.

(6) The channel capacity C is defined as $\max_{P_X} I(X;Y)$, and can be seen to be a function of only the channel transition probabilities $P_{Y|X}$, as follows.

Observe that $I(X;Y) = H(Y) - H(Y|X)$

Now $H(Y|X) = \sum_{x \in \mathcal{X}} P_X(x)\, H(Y|X=x) = \sum_{x \in \mathcal{X}} P_X(x) \left(-\sum_{y \in \mathcal{Y}} P_{Y|X}(y|x) \log P_{Y|X}(y|x) \right)$

Observe that $H(Y|X)$ is therefore a function of P_X & $P_{Y|X}$.

Also $H(Y) = -\sum_{y} P_Y(y) \log P_Y(y)$. But to get P_Y, we use $\sum_{x \in \mathcal{X}} p(x,y) = \sum_{x \in \mathcal{X}} P_{Y|X}(y|x)\, P_X(x)$.

Thus $H(Y)$ is also a function of P_X & $P_{Y|X}$.

$\Rightarrow I(X;Y)$ is a function of P_X & $P_{Y|X}$.

Now, to compute $C = \max_{P_X} I(X;Y)$, we maximize the value $I(X;Y)$ over all possible choices of the channel input distribution, P_X. [It is best to view this purely as a mathematical operation, without worrying about what such 'maximization' means, in the engineering sense].

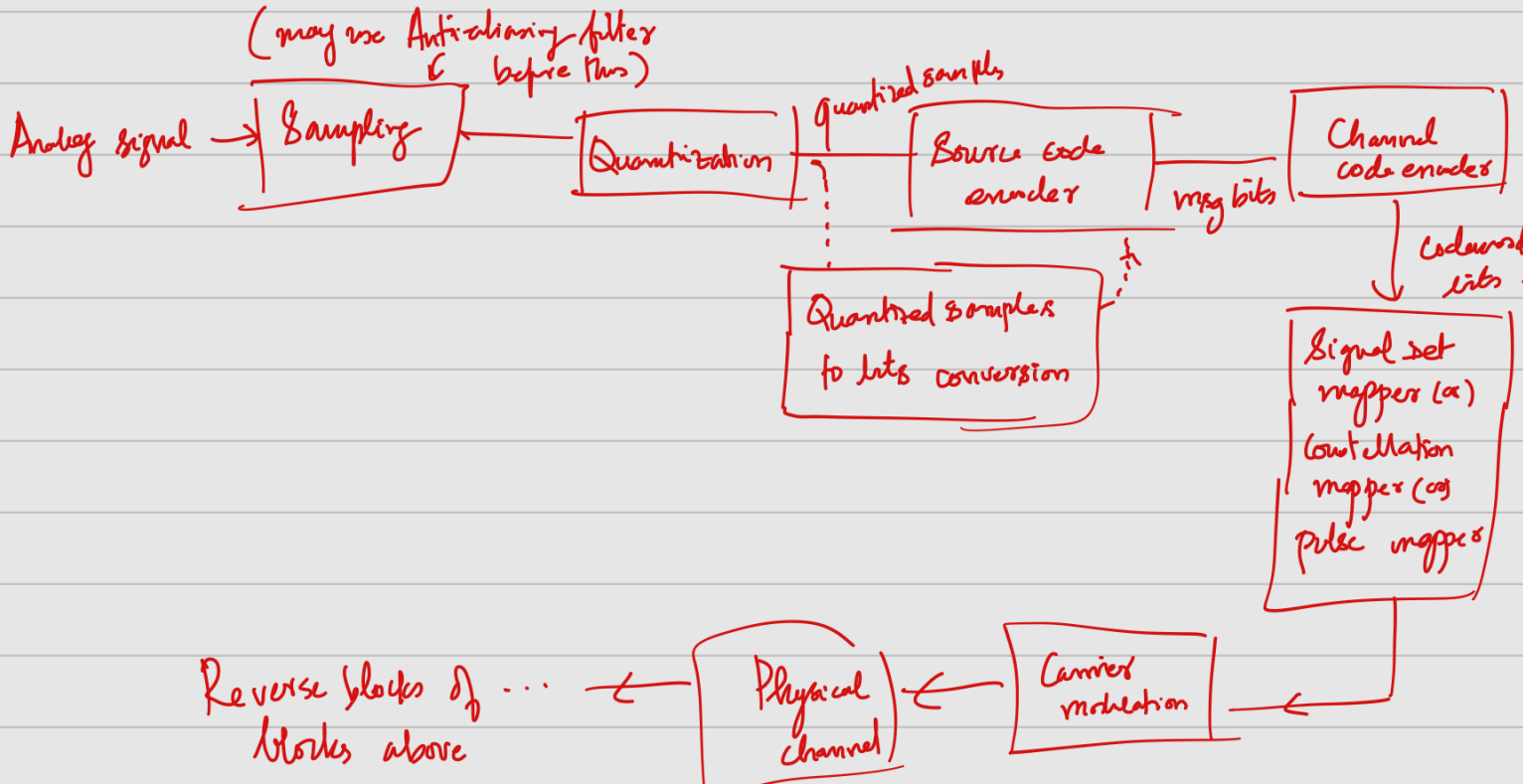
Note that P_X essentially consists of a tuple of $|\mathcal{X}|$ non-negative real numbers which sum to 1. The maximization is over all such tuples.

As an example, the capacity of BEC(α) (binary erasure channel, with erasure probability α) was derived in the class. It was shown that the capacity of BEC(α) is $1-\alpha$. The capacity of BSC(p) is left as an exercise. Students are advised to see these derivations from the "Elements of Information Theory" book.

— x —

Through this course, we have studied basics of signals used in communication, FT analysis, analog & digital communication basics.

Digital Comm. Conceptual block diagram



— x —

This course is connected to various other courses done in ECE programme