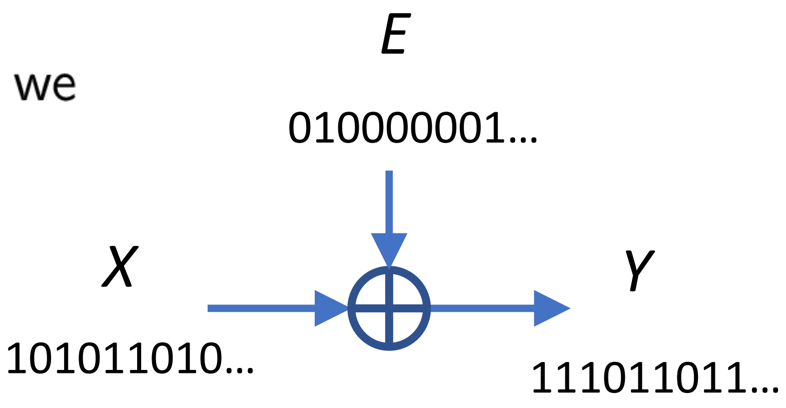
Capacity of communication channels

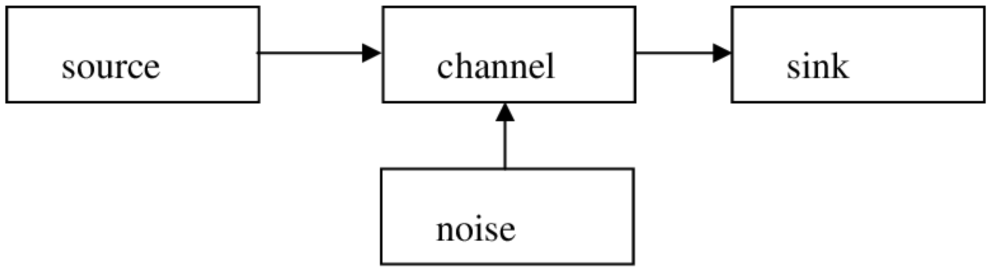
Capacity of communications channels

- In information theory a communications channel is a conditional probability distribution that describes the channel output Y given the channel input X.
- Discrete time channels are defined by the conditional probability distributions $p(y^n|x^n)$, which denotes $p(y_1, y_2, \ldots, y_n|x_1, x_2, \ldots, x_n)$ (in "Matlab" notation $p(y_{1:n}|x_{1:n})$).
- For example if the channel has binary inputs and outputs, we can think of the output being a result of XOR addition of input and binary noise.



Communications channels

$$p\left(y_1, y_2, \ldots, y_n \,|\, x_1, x_2, \ldots, x_n\right)$$



- The noise may be neither independent nor time invariant.
- If the noise is independent i.e. has no memory, the channel has no memory as a result.

Discrete memoryless channel

Definition

A discrete memoryless channel (DMC) is defined by:

$$p\left(y_{1:n} \,|x_{1:n}\right) = \prod_{i=1}^{n} p\left(y_i \,|x_i\right)$$

This definition implies that past outputs do not affect the future outputs.

Channel coding theorem

- In 1948 Claude Shannon published a paper in which he established that for any transmission rate less or equal to the channel capacity, there exist a coding scheme that achieves an arbitrary small probability of error.
- Conversely, transmission with rates higher than the capacity will always incur errors.
- In his proof Shannon used a new concept of random codes – as a consequence the proof is non-constructive (there was no practical encoding or decoding with infinite code lengths).

Channel coding theorem

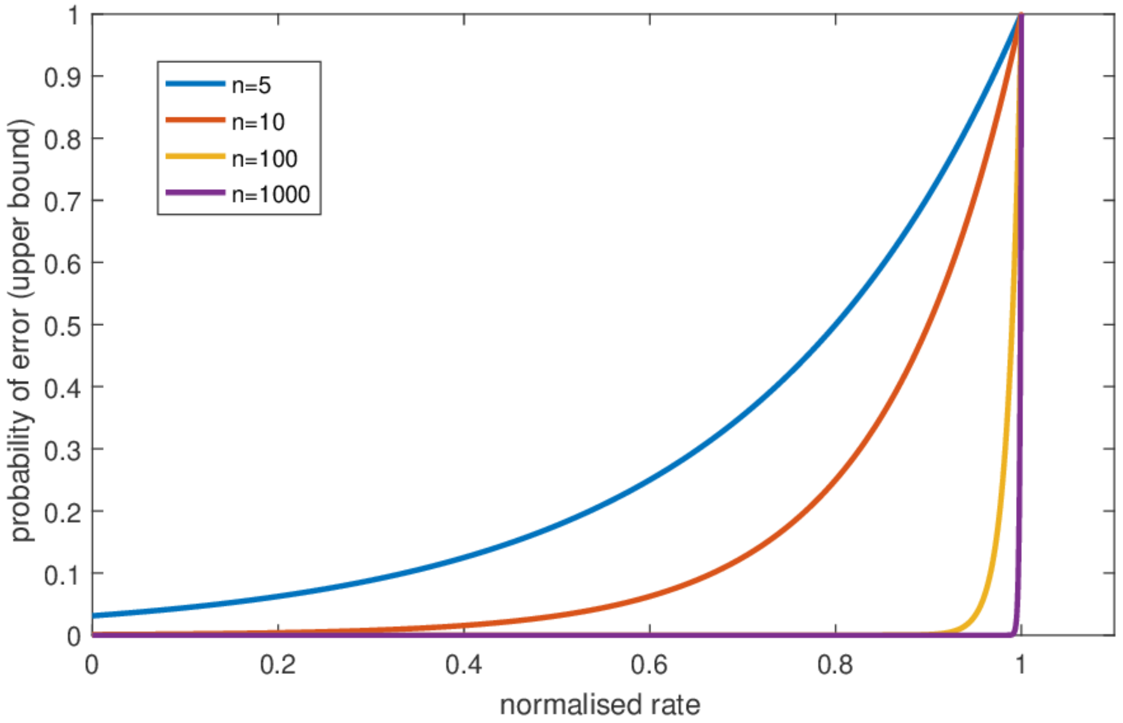
Shannon proved that:

$$P_e \leq 2^{-n(I(X;Y)-R)}$$

The probability of error can be made arbitrarily small (by making n sufficiently large), as long as $I(X;Y) > R$.

- Proof of Shannon Channel coding theorem can be found on "Blackboard"

Channel coding theorem



In this figure normalised rate is $\frac{I(X;Y)-R}{I(X;Y)}$.

- Rates below the curves are achievable with vanishingly small probability of error (as the code length increases $n \uparrow$)

DMC channel capacity

Definition

The information capacity of a DMC channel is

$$C = \max_{p(x)} I(X;Y)$$

that is:

$$C = \max_{p(x)} I(X;Y) = \max_{p(x)} \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)}$$

Channel capacity is the maximum number of bits per use of the channel that can be reliably transmitted.

Capacity calculations

We can calculate the capacity using any definition of $I(X;Y)$:

- $\max_{p(x)} I(X;Y) = \max_{p(x)} [H(X) - H(X|Y)]$
- $\max_{p(x)} I(X;Y) = \max_{p(x)} [H(Y) - H(Y|X)]$
- $\max_{p(x)} I(X;Y) = \max_{p(x)} [H(X) + H(Y) - H(X,Y)]$

The strategy is to find which of the (conditional) entropies is easiest to calculate.

Important properties of capacity

- $C \geq 0$, this is because $I(X;Y) \geq 0$
- if $C = 0$, the channels is called useless since the output "says" nothing about the input.
- $C \leq \min\{\log|\mathcal{X}|, \log|\mathcal{Y}|\}$, this is because $I(X;Y) \leq H(X) \leq \log|\mathcal{X}|$ and $I(X;Y) \leq H(Y) \leq \log|\mathcal{Y}|$

Examples: Noiseless binary channel

$\mathcal{X}, \mathcal{Y} \in \{0, 1\}$, $p(y|x) = 1$ iff $x = y$

$$0 \longrightarrow 0$$

$$1 \longrightarrow 1$$

The conditional entropy is given by
$H(X|Y) = p_Y(0)\,H(X|Y = 0) + p_Y(1)\,H(X|Y = 1) =$
$p_Y(0)\,H_2(1) + p_Y(1)\,H_2(0) = 0$
where $H_2(p)$ is the binary entropy function:
$H_2(p) = H(\{p, 1 - p\}) = -p \log p - (1 - p) \log (1 - p)$

The capacity of this channel is:

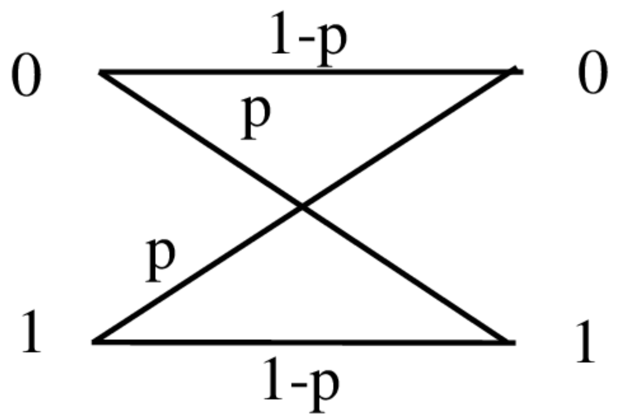
$$I(X;Y) = H(X) - H(X|Y) = H(X)$$

$$C = \max I(X;Y) = \max H(X) = 1$$

This is the communications channel assumed by most digital logic.

Binary symmetric channel (BSC)

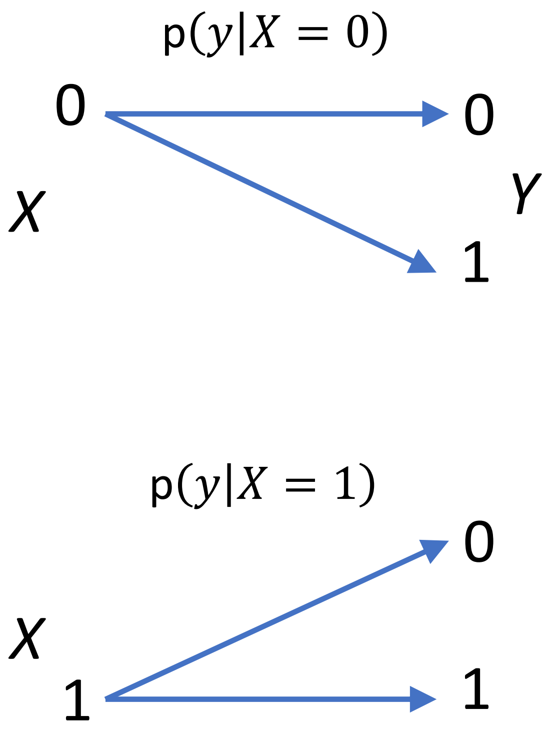
The binary symmetric channel corresponds to a very simple noise model: independent, time invariant, and data independent.



$$p(y|x) = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

BSC capacity

The BSC has a single parameter, p, the bit error probability (BER – bit error rate).
$H(Y|X=0) = H(\{1-p, p\}) = H_2(p)$
$H(Y|X=1) = H(\{p, 1-p\}) = H_2(p)$
$H(Y|X) = p_X(0) H_2(p) + p_X(1) H_2(p) = H_2(p)$
The conditional entropy is independent of the input pdf (because of symmetry).



The mutual information is:
$I(X;Y) = H(Y) - H_2(p)$
The maximum value for $H(Y)$ is obtained when Y is equiprobable, which occurs when X is equiprobable. In this case

$$C = \max I(X;Y) = 1 - H_2(p)$$

BSC capacity - interpretation

- The BSC channel can be modeled as $y = x \oplus e$, where $\oplus$ - is "+" in GF2 (XOR), and e - is error sequence $p_E(e=1) = p$
- Suppose that we send the following data sequence:

$x_{1:n} = 101101000110111$

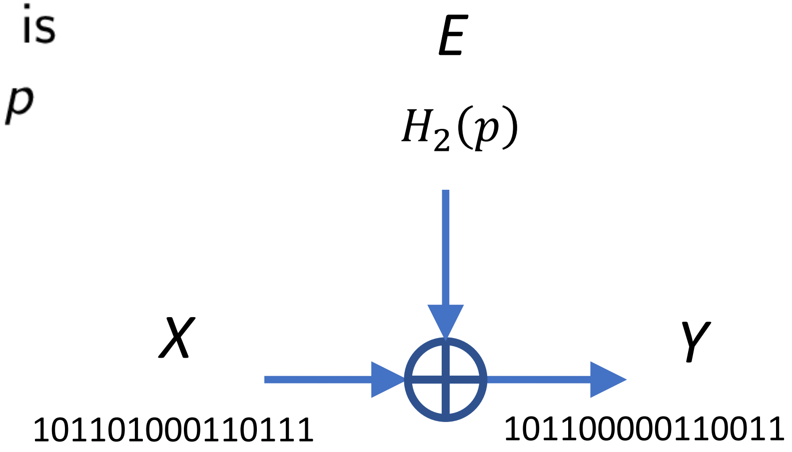
- and the corresponding received sequence is:

$y_{1:n} = 101100000110011$

- the error sequence is therefore

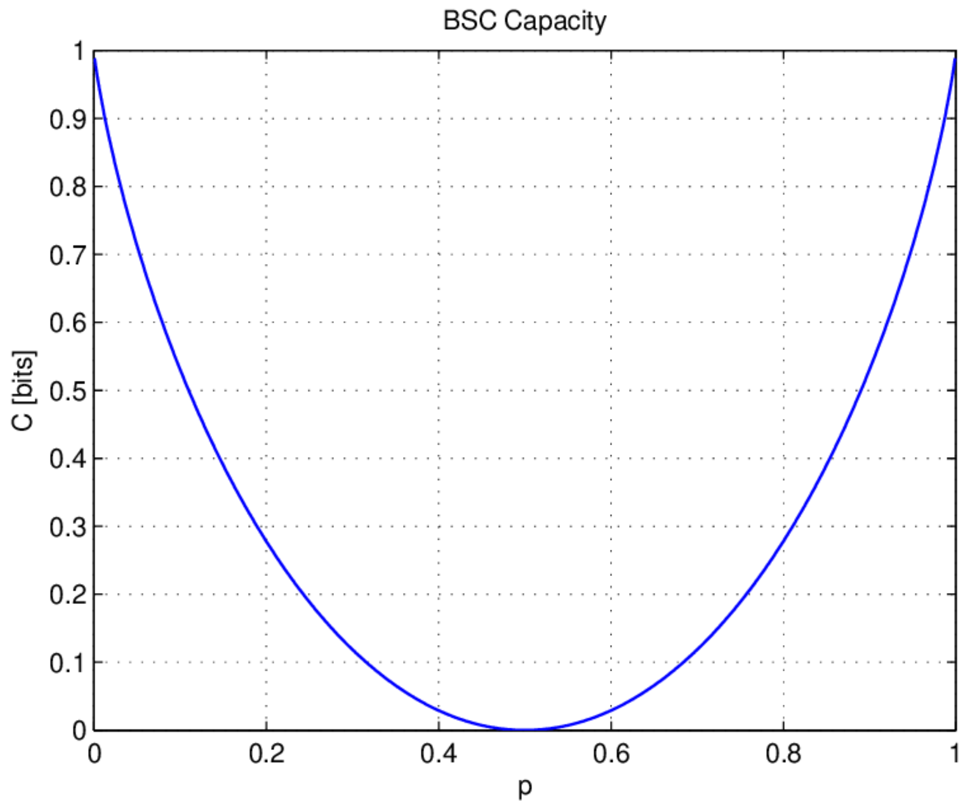
$e_{1:n} = 000001000000100$

If we knew this sequence, we could find $x_{1:n}$ exactly. This sequence contains exactly $H_2(p)$ bits of information per symbol. Since we don't know this sequence, this deficiency represents a loss. Therefore, the remaining capacity available for our data is $C_{BSC} = 1 - H_2(p)$.



BSC capacity

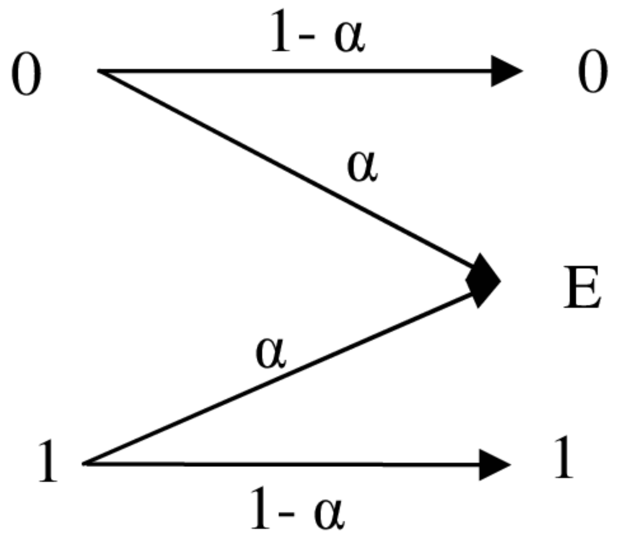
BSC capacity $C = 1 - H_2(p)$



- It is not easy to think of a coding scheme which achieves the capacity

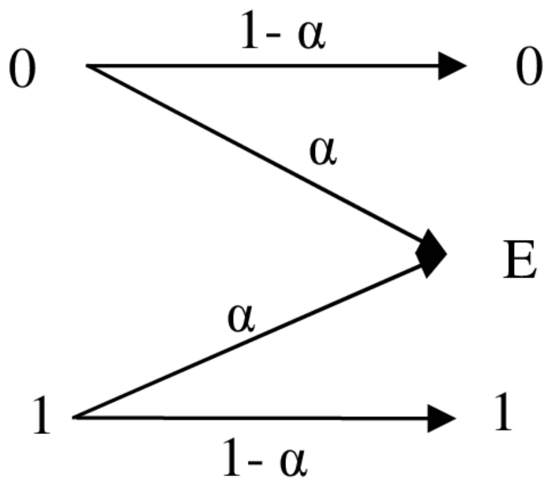
Binary erasure channel (BEC)

Simple channel model where the information bits are either received without errors or lost (we know when the erasure occurs). This channel models the Internet. It is often applicable also to many communications systems to model the behavior at the packet level (e.g. wireless communications).



$$p(y|x) = \begin{bmatrix} 1-\alpha & 0 & \alpha \\ 0 & 1-\alpha & \alpha \end{bmatrix}$$

BEC capacity "$H(X) - H(X|Y)$"

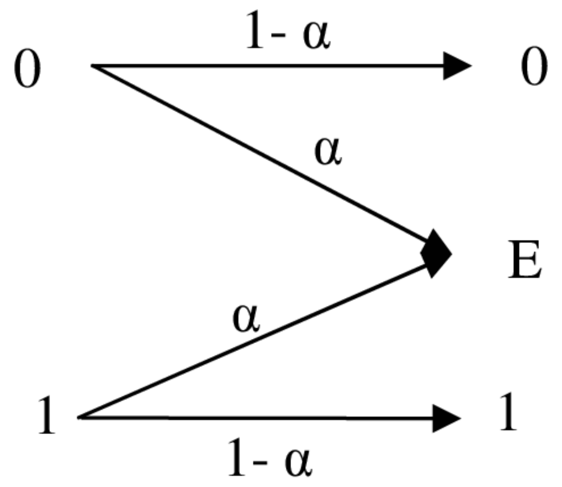


$$C = \max_{p(x)} I(X;Y) = \max_{p(x)} [H(X) - H(X|Y)]$$

$$H(X|Y) = p(y=0)\,H(X|Y=0) + p(y=1)\,H(X|Y=1) + p(y=E)\,H(X|Y=E) = \underbrace{p(y=E)}_{\alpha}\underbrace{H(X|Y=E)}_{H(X)} = \alpha H(X)$$

$$I(X;Y) = \max_{p(x)} H(X) - \alpha H(X) = \max_{p(x)} H(X)(1-\alpha) = 1 - \alpha$$

BEC capacity "$H(Y) - H(Y|X)$"



$H(Y|X) = H_2(\alpha)$

now:

$p_Y(0) = (1-\alpha)\, p_X(0) = (1-\alpha)(1 - p_X(1))$

$p_Y(E) = \alpha p_X(0) + \alpha p_X(1) = \alpha,$

$p_Y(1) = (1-\alpha)\, p_X(1)$

hence:

$H(Y) = H(\{(1-\alpha)(1-p_X(1)), \alpha, (1-\alpha)\, p_X(1)\})$

$= -(1-\alpha)(1-p_X(1)) \log((1-\alpha)(1-p_X(1)))$

$-\alpha \log \alpha - (1-\alpha)\, p_X(1) \log((1-\alpha)\, p_X(1))$

$= H_2(\alpha) + (1-\alpha)\, H(X)$

hence:

$C = \max I(X;Y) = \max(H(Y) - H_2(\alpha)) =$

$\max(H_2(\alpha) + (1-\alpha)\, H(X) - H_2(\alpha)) = \max(1-\alpha)\, H(X) =$

$1 - \alpha$

BEC capacity - interpretation

- Suppose that we send the following data sequence:

$x_{1:n} = 101101000110111$

- and the corresponding received sequence is:

$y_{1:n} = 10110?000110?11$

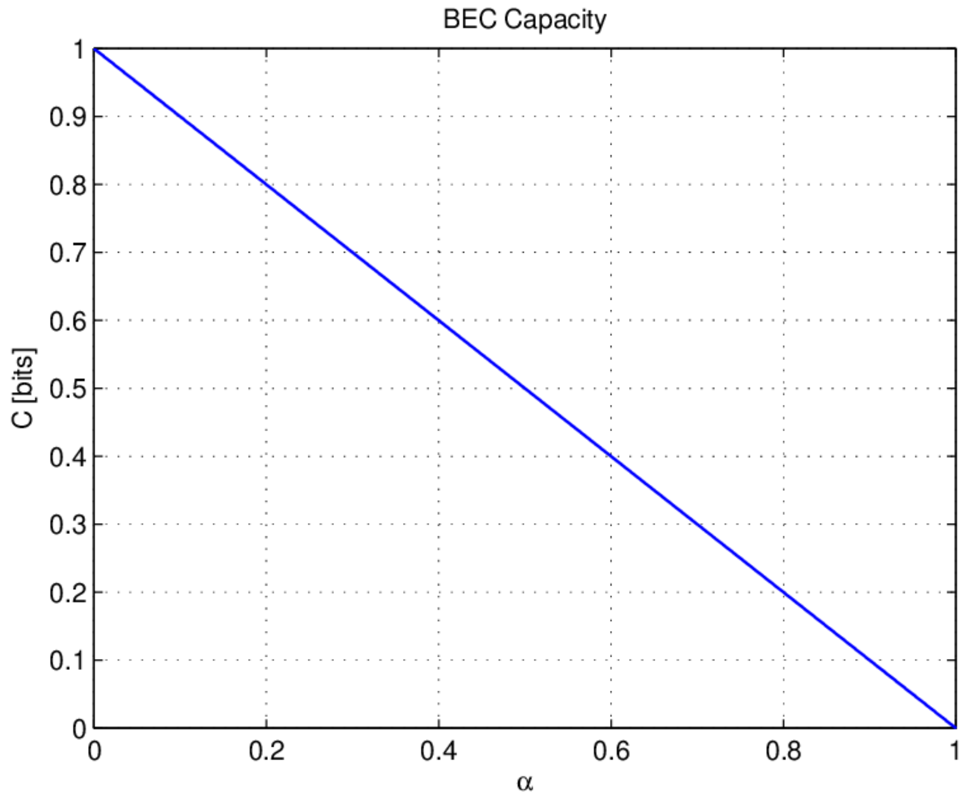
- The channel will erase αn bits on average (out of total n transmitted bits). Therefore, the number of remaining (non erased) bits available is $(1-\alpha)\,n$ (i.e. $1-\alpha$ per channel use)

1	2	2	3	4	5	5	6	6	7

- channel erasures

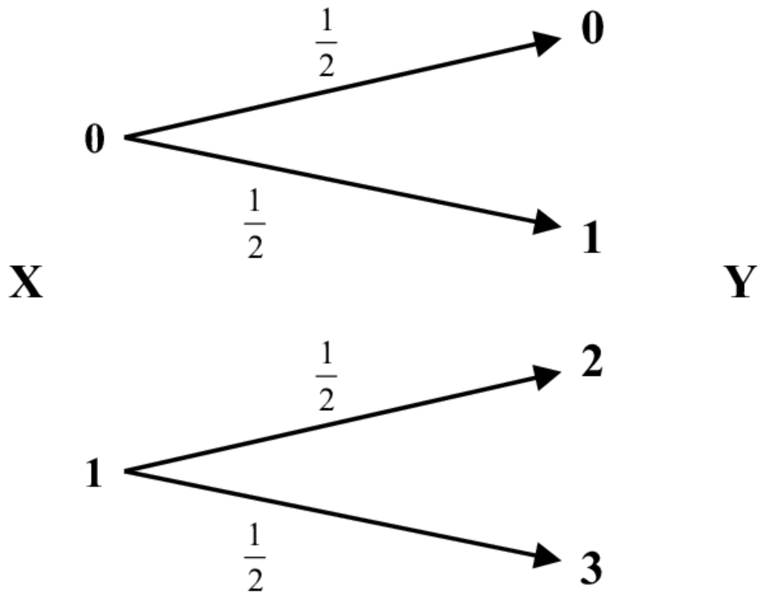
BEC capacity

$$C = 1 - \alpha$$



- It is not easy to think of a coding scheme, which achieves the capacity (what if we had a feedback channel?)

Non-overlapping outputs channel

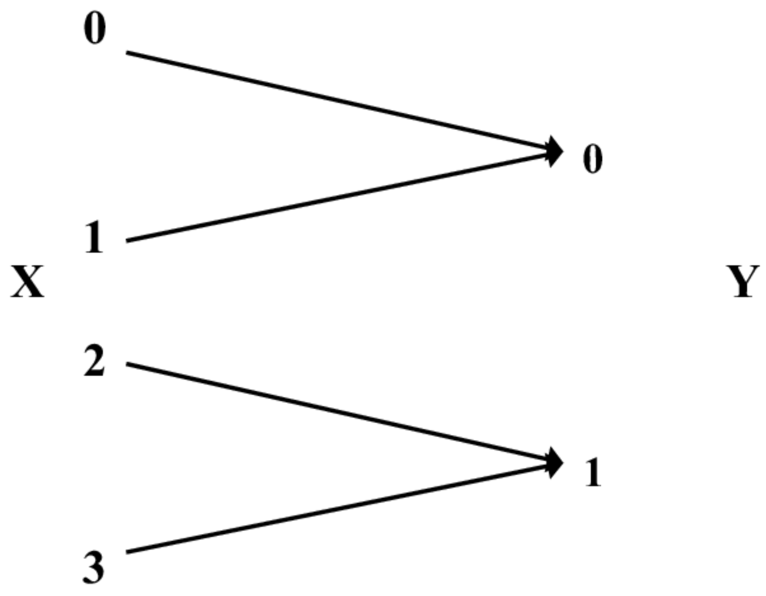


$$p(y|x) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$H(X|Y) = 0$ (however note that $H(Y|X) = 1$)

$C = \max I(X;Y) = \max(H(X) - H(X|Y)) = \max(H(X)) =$ 1bit

Deterministic channel

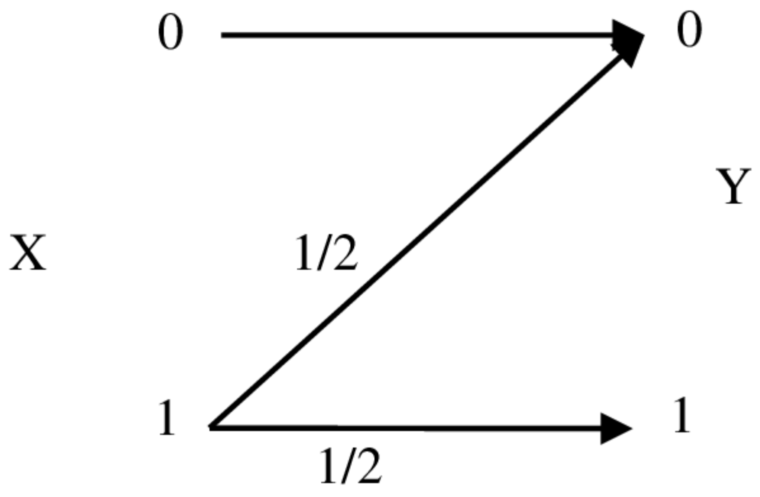


$H(Y|X) = 0$
$C = \max I(X;Y) = \max(H(Y) - H(Y|X)) = \max(H(Y)) = 1$
bit

Z channel

- This is a non-symmetric channel, we don't expect the capacity to be maximised for a uniform distribution.

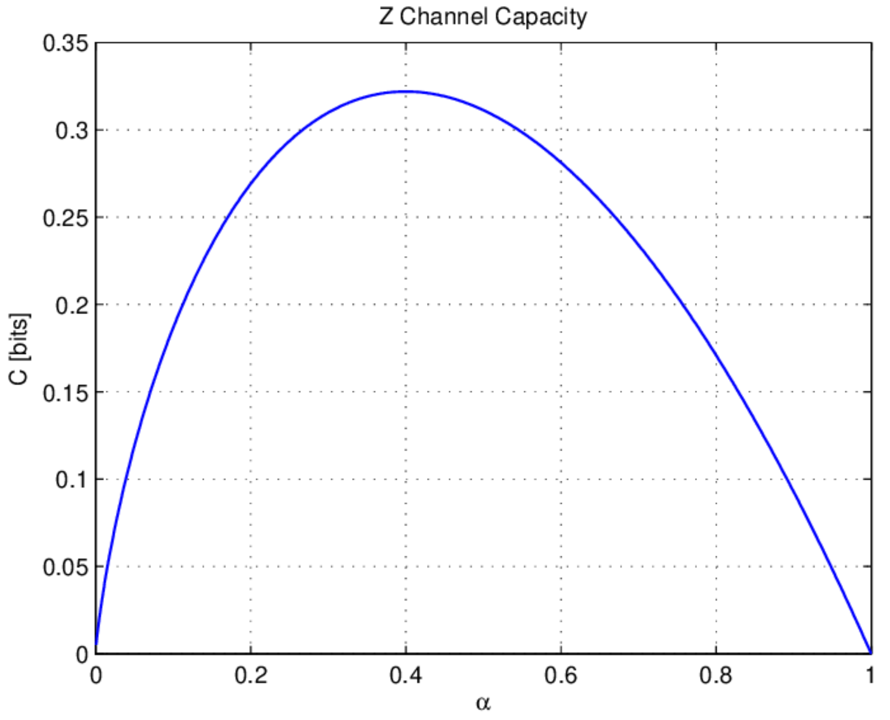
Denote $\alpha = p(X = 1)$
$H(Y|X) = p(X = 0) \cdot 0 + p(X = 1) \cdot 1 = p(X = 1) = \alpha$
$H(Y) = H_2(p(Y = 1)) = H_2\left(\frac{\alpha}{2}\right)$
$I(X; Y) = H(Y) - H(Y|X) = H_2\left(\frac{\alpha}{2}\right) - \alpha$



$\frac{d}{d\alpha} I(X;Y) = \frac{d}{d\alpha}\left[-\frac{\alpha}{2}\log\left(\frac{\alpha}{2}\right) - \left(1-\frac{\alpha}{2}\right)\log\left(1-\frac{\alpha}{2}\right) - \alpha\right] =$
$\frac{1}{2\log_e(2)}\log_e\left(\frac{1-\frac{\alpha}{2}}{\frac{\alpha}{2}}\right) - 1$
this is equal to zero for $\alpha = 2/5$. So the capacity of the Z-channel is:

$$\max I(X;Y) = H_2\left(\frac{1}{5}\right) - \frac{2}{5} = 0.322$$

Z channel capacity as a function of input distribution $p_X(x)$, where $\alpha = p(X = 1)$



- The optimal distribution $p_X^*(x)$ is non-symmetric. It favours symbol "0" i.e. $p^*(X = 1) = 0.4 \Rightarrow p^*(X = 0) = 0.6$