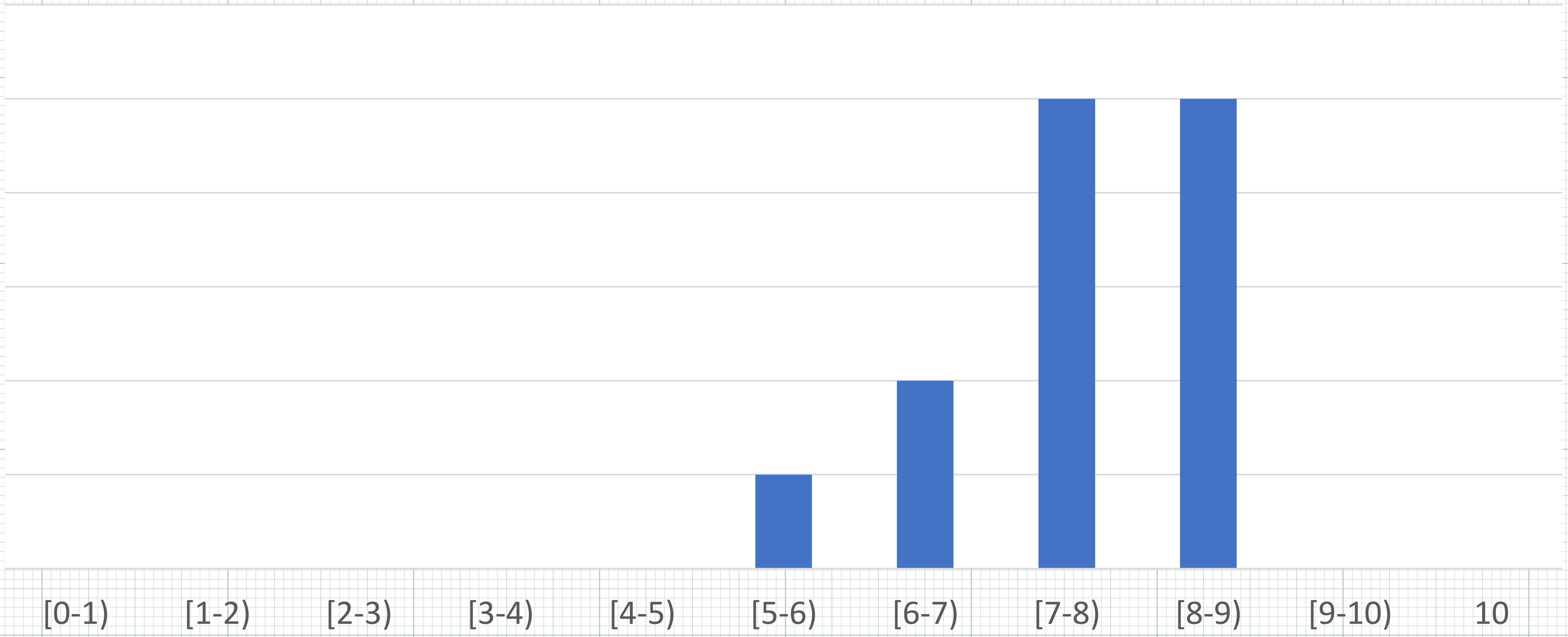
# TN3125
# Information and Computation

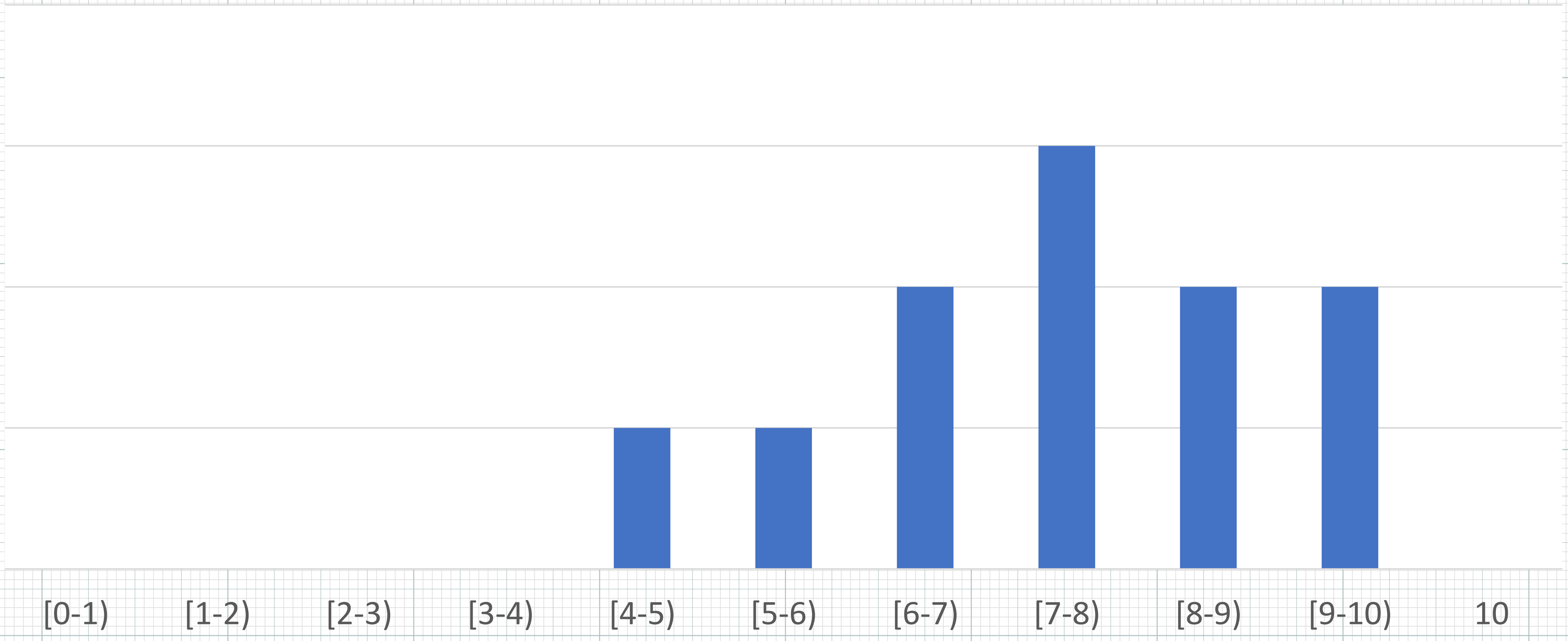
Lecture 4

1- *Introduction*

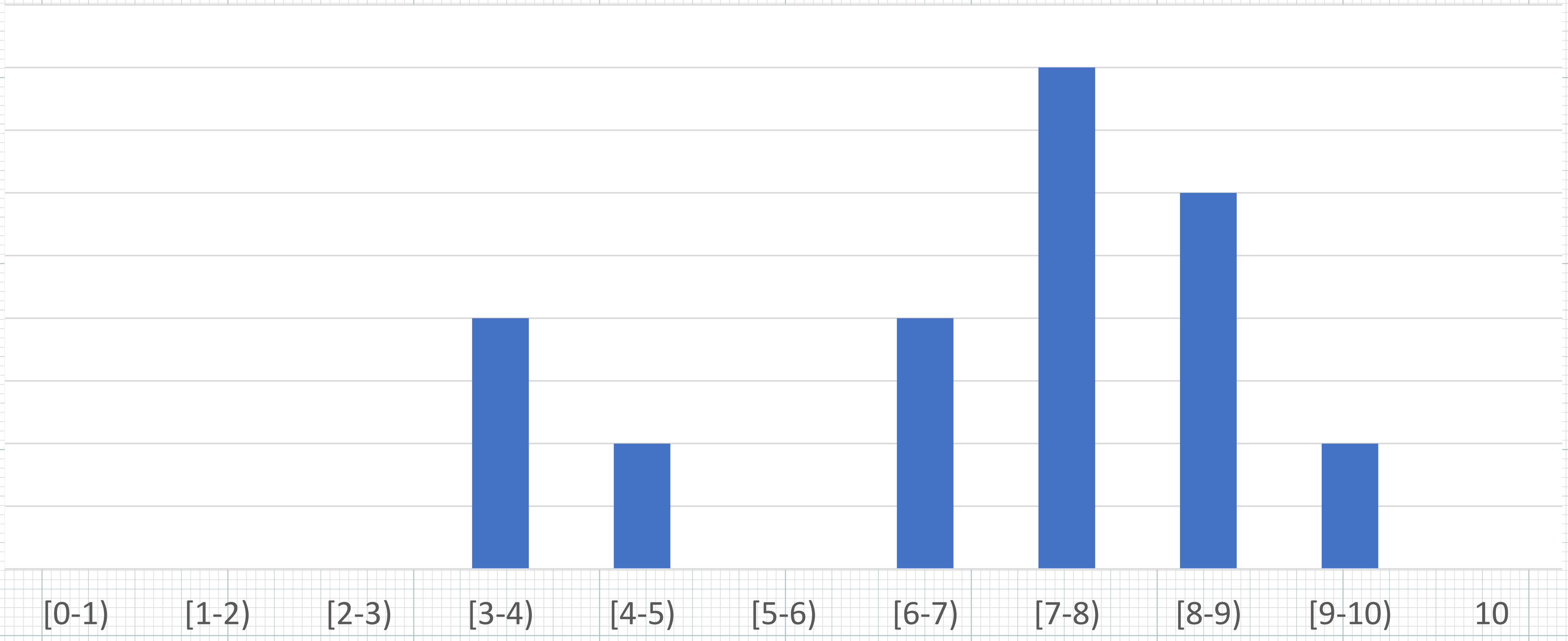
# Minitest 1 results



# Minitest 2 results



# Average



# Minitest 3 – 1a

Consider the code $C = \{00000, 11111\}$.

(a) (5 points) Find the minimum distance of the code.

# Minitest 3 – 1a

Consider the code $C = \{00000, 11111\}$.

(a) (5 points) Find the minimum distance of the code.

$$d_{min}(C) = \min_{x,y} d(x,y) = \min\{5\} = 5$$

# Minitest 3 – 1b

(b) (10 points) Find the maximum number of errors and erasures that a minimum distance decoder will correctly correct. If you were unable to solve the previous exercise, assume the minimum distance $d = 7$.

# Minitest 3 – 1b

(b) (10 points) Find the maximum number of errors and erasures that a minimum distance decoder will correctly correct. If you were unable to solve the previous exercise, assume the minimum distance $d = 7$.

$$t \leq \frac{d-1}{2}, \quad \max t = 3$$

$$e, s \leq d-1, \quad \max e, s = 6$$

# Minitest 3 – 1c

(c) (5 points) Find the words in $S_1(11111)$, the sphere of radius 1 around 11111.

# Minitest 3 – 1c

(c) (5 points) Find the words in $S_1(11111)$, the sphere of radius 1 around 11111.

| $e$ | 11111 |
|---|---|
| 00000 | 11111 |
| 00001 | 11110 |
| 00010 | ⋮ |
| 00100 | |
| 01000 | |
| 10000 | |

# Minitest 3 – 1d

(d) (10 points) How many binary words of length 5 have two ones?

# Minitest 3 – 1d

(d) (10 points) How many binary words of length 5 have two ones?

$$\binom{5}{2} = \frac{5!}{2!\,3!} = \frac{5 \cdot 4 \cdot \cancel{3!}}{2!\,\cancel{3!}} = 10$$

11000 01100
10100 01010
10010 01001
10001

00110 00011
00101

# Minitest 3 – 1e

(e) (10 points) What is the error probability of a minimum distance decoder if we send the word 00000 through a binary symmetric channel with crossover probability $p$? You can leave your answer as a function of binomial coefficients.

# Minitest 3 – 1e

(e) (10 points) What is the error probability of a minimum distance decoder if we send the word 00000 through a binary symmetric channel with crossover probability $p$? You can leave your answer as a function of binomial coefficients.

Wrong if #errors $\geq 3$

$$P_e = P_r(3 \text{ errors}) + P_r(4 \text{ errors}) + P_r(5 \text{ errors})$$

$$= \binom{5}{3}(1-p)^2 p^3 + \binom{5}{4}(1-p)^1 p^4 + \binom{5}{5} p^5$$

# Minitest 3 – 1f

(f) (5 points) Is the following matrix in standard form? Indicate why yes or no.

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

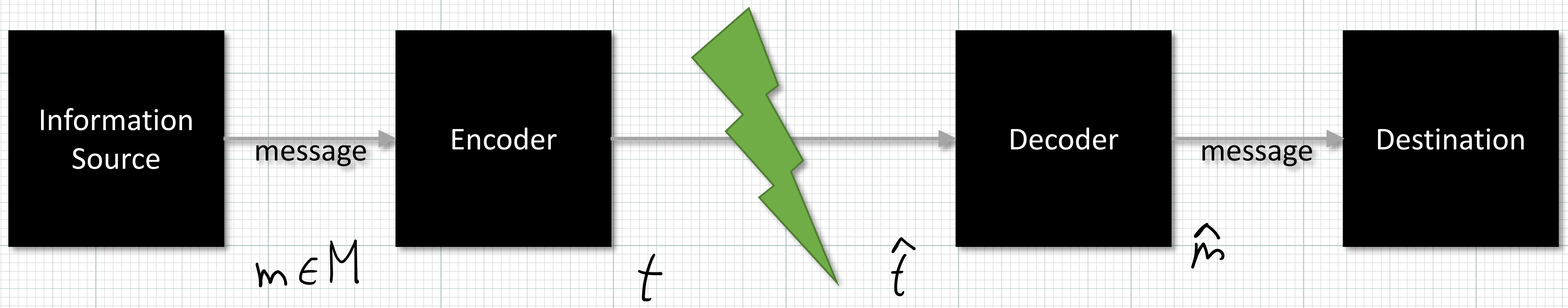
# Minitest 3 – 1f

(f) (5 points) Is the following matrix in standard form? Indicate why yes or no.

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

$$G \neq ( I_3 \mid A )$$

# The abstract communications model




$$m = \hat{m}$$

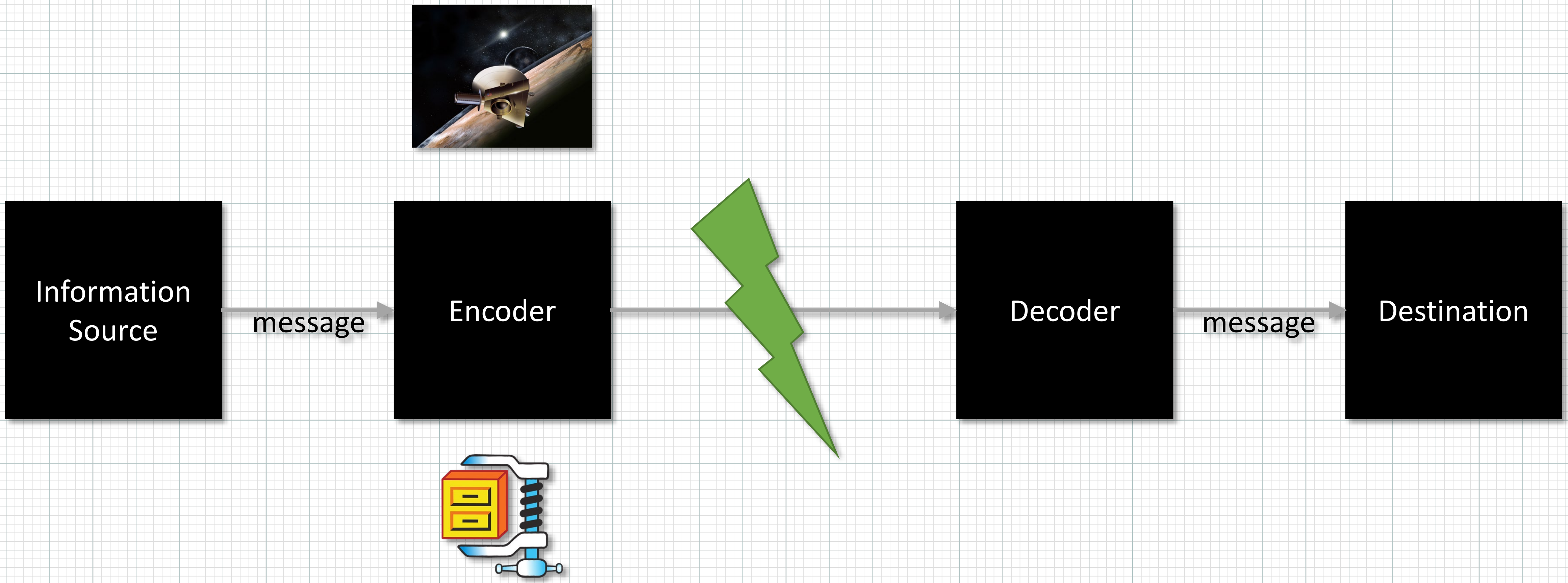
# Summary of weeks 1 to 3

- We derived an information measure from basic axioms
- We proved basic properties of entropy

- We defined several families of codes for data compression
- We proved that average length is bounded by entropy

- We presented linear codes and defined their properties for correction, detection and erasure as well as bounds on code paremeters
- We introduced the minimum distance decoder
- We introduced the binary erasure channel and the binary symmetric channel

# The abstract communications model



# Learning goals for week 4

- Encode information with linear codes
- Decode using a standard array
- Give the general form of Hamming codes and prove basic properties
- Sketch the noisy coding theorem both achievability and converse
- Find the capacity of simple channels

# Binary linear codes

- A code $C$ is a binary linear code if it is a subspace of $V_n$

- The set of codewords is generated by linear combinations of a basis set of vectors $w^1, \dots, w^k$

- A generator matrix is a matrix with rows consisting of the vectors of a basis:
$$G = \begin{pmatrix} w^1 \\ \dots \\ w^k \end{pmatrix}$$

- A generator matrix is in standard form if it is written in the form $G = \left(I_k | A_{k,n-k}\right)$

# Encoding information

- Let $x \in \{0,1\}^k$ how do we encode it into a codeword of $C$ an $[n, k, d]$ code
- Given a generator matrix $G$ for the code:

$$x \mapsto xG = \sum_{i=1}^{k} x_i w^i$$

- Example

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

$$x = (0\ 1)$$

$$x \mapsto xG = (0\,1\,1)$$

# Systematic encoding

- Suppose that $G$ is in standard form
- We call the first $k$ bits of $xG$ the information bits
- We call the last $\mathrm{n} - k$ bits of $xG$ the redundancy bits

# The parity check code

- Consider a code that takes $x = (x_1, \ldots, x_k) \in \{0,1\}^k$ and encodes it with codeword $c = (x_1, \ldots, x_k, \sum_{i=1}^{k} x_i)$

- What is the generator matrix of this code?

- What is the minimum distance of the code?

# The parity check code

- Consider a code that takes $x \in \{0,1\}^k$ and encodes it with codeword $c = (x_1, \dots, x_n, \sum_{i=1}^{n} x_i)$

- What is the generator matrix of this code?

- What is the minimum distance of the code?

# Decoding

- We transmit codeword $x = (x_1, \dots, x_n)$ and receive $y = (y_1, \dots, y_n)$ and let $e = y + x$
- Decoder goal: find $e$ output $y + e = x$

- Definition. Let $C$ be an $[n, k, d]$ code and $v \in \{0,1\}^n$ not necessarily a codeword, we call

$$v + C = \{x + v : x \in C\}$$

a coset of $C$

# Exercise

- Show that if $y \in x + C$ then $y + C = x + C$

$$v + C = \{x + v : x \in C\}$$

# Solution

- From the statement $y \in x + C$ , we have that there exist some codeword $t$ such that $x + t = y$

- Now for all codewords $c$, $c + y \in y + C$. But $c + y = c + x + t = x + (c + t)$, which means that $c + y \in x + C$, and that $c + y \subseteq c + x$

- Similarly, for all codewords $c$, $c + x \in x + C$. And we can run the same argument to conclude $c + x \subseteq c + y$

# Lagrange theorem for codes

- Theorem. Let $C$ be an $[n, k, d]$ code. Then
  - $v \in \{0,1\}^n$ is in some coset of C
  - Each coset has $2^k$ words
  - Two cosets either have no overlap, either they completely coincide
  - There are exactly $2^{n-k}$ cosets

# Proof

- Since $0$ is always a codeword $v \in v + C$
- Since all codewords are different, there is one element in $v + C$ per codeword
- Imagine there is some $v \in x + C$ but also $v \in y + C$ for $x \neq y$ . Then we have $v + C = x + C$ and we also have $v + C = y + C$ hence $x + C = y + C$.
- There are $2^n$ words, each coset is disjoint and has $2^k$ words

# Exercise

- Find the cosets of the code with generator matrix

$$\begin{pmatrix} 1001 \\ 0111 \end{pmatrix}$$

# Solution

- Let us first find all the codewords:
- $(00)\begin{pmatrix}1001\\0111\end{pmatrix} = (0000), (01)\begin{pmatrix}1001\\0111\end{pmatrix} = (0111), (10)\begin{pmatrix}1001\\0111\end{pmatrix} = (1001), (11)\begin{pmatrix}1001\\0111\end{pmatrix} = (1110)$
- $C = \{0000,0111,1001,1110\}$
- $0001 + C = \{0001,0110,1000,1111\}$
- $0010 + C = \{0010,0101,1001,1100\}$
- $0100 + C = \{0100,0011,1101,1010\}$

# TN3125
# Information and Computation

Lecture 3

2- *Decoding and Hamming codes*

# Coset leader

- We call the vector with minimum hamming weight its leader, if there is more than one vector with minimum weight, any of them can be the coset leader.
- Example:

$$
\begin{aligned}
C &= \{0000,0111,1001,1110\} \\
0001 + C &= \{0001,0110,1000,1111\} \\
0010 + C &= \{0010,0101,1011,1100\} \\
0100 + C &= \{0100,0011,1101,1010\}
\end{aligned}
$$

# Standard array for code $C$

- Table with $2^k$ columns and $2^{n-k}$ rows
- In the top row, we place the elements of $C$, beginning with the zero codeword
- In each other row we place the elements of a coset of $C$, beginning with the coset leader

# Example

- The previous exercise gave us almost the standard array

$$\begin{array}{cccc} 0000 & 0111 & 1001 & 1110 \\ 0001 & 0110 & 1000 & 1111 \\ 0010 & 0101 & 1011 & 1100 \\ 0100 & 0011 & 1101 & 1010 \end{array}$$

# Decoding with the standard array

- When we receive $y$, we look for it in the array
- We find the leader of the coset and add it to $y$, we output $y + \text{coset leader}(y)$

# Explanation

- If we receive, $y \in x + C$, where $x$ is the coset leader, we know it equals to $x + c$, for some codeword $C$

- Now, we know if we output $y + x$, this is the same as having as output $c$ which is a codeword

- Finally, the distance between $c$ and $y$ is $x$, which is the minimum possible as it has the minimum Hamming weight in the coset

# Example

- If we receive $y = 1011$

$$\begin{array}{l} 0000,0111,1001,1110 \\ 0001,0110,1000,1111 \\ 0010,0101,1011,1100 \\ 0100,0011,1101,1010 \end{array}$$

- We look for the coset leader, which is $0010$ and output $1001$ which is at distance one of $y$

# Parity check matrix

- A matrix $H$ is a parity check matrix for a code $C$ if $Hx^T = 0$, if and only if $x \in C$.

- Lemma. $H$ is an $(n-k) \times n$ matrix, and it verifies that $GH^T = 0$

- Exercise. Find the parity check matrix of the length 3 repetition code.

# Hamming codes

- Hamming codes are a family of codes defined for lengths $n = 2^r - 1$, $r \in \mathbb{N}$. The parity check matrix of $H_n$ has as columns all the non-zero elements of $V_r$.

- Exercise. How many bit do they encode? (what is the value of k?)

- Exercise. The parity check matrix of $H_3$ is given by

# Hamming codes

- Hamming codes are a family of codes defined for lengths $n = 2^r - 1$, $r \in \mathbb{N}$. The parity check matrix of $H_n$ has as columns all the non-zero elements of $V_r$.

- Exercise. How many bit do they encode? (what is the value of k?)

$n - k = r \quad , \quad k = n - r$

- Exercise. The parity check matrix of $H_3$ is given by

# Hamming codes

- Hamming codes are a family of codes defined for lengths $n = 2^r - 1$, $r \in \mathbb{N}$. The parity check matrix of $H_n$ has as columns all the non-zero elements of $V_r$

- Exercise. The parity check matrix of $H_3$ is given by

$$3 = 2^2 - 1, \qquad r = 2$$

$$H_3 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

# Properties of Hamming codes 1

- Hamming codes have distance 3

# Properties of Hamming codes 1

- Hamming codes have distance 3

- We need to show that there are no codewords of weight 1 and 2, and that there exist words of weight 3
  - For weight one, this follows because for $x$ of weight one to be a codeword, the associated column of H would need to be zero
  - For weight two,

$$\begin{pmatrix} | & | & & \\ | & | & \cdots & \\ | & | & & \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 0 \iff c_1^1 + c_1^2 = 0, \; \vdots$$

  -

# Properties of Hamming codes 1

- Hamming codes have distance 3

$$\begin{pmatrix} 0 & 0 & 0 & & & \\ 0 & 0 & \cdot & & & \\ 0 & 0 & & & & \\ \vdots & & & \cdots & & \\ 0 & 0 & 0 & & & \\ 0 & 1 & 1 & & & \\ 1 & 0 & 1 & & & \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 0$$

# Properties of Hamming codes 2

- Hamming codes are perfect codes, (i.e they meet the Hamming bound)

# Properties of Hamming codes 2

- Hamming codes are perfect codes, (i.e they meet the Hamming bound)

$$2^k \cdot \sum_{i=0}^{1} \binom{n}{i} = 2^k (1 + n) = 2^k (1 + 2^r - 1)$$

$$= 2^{k+r} = 2^{n-r+r} = 2^n$$

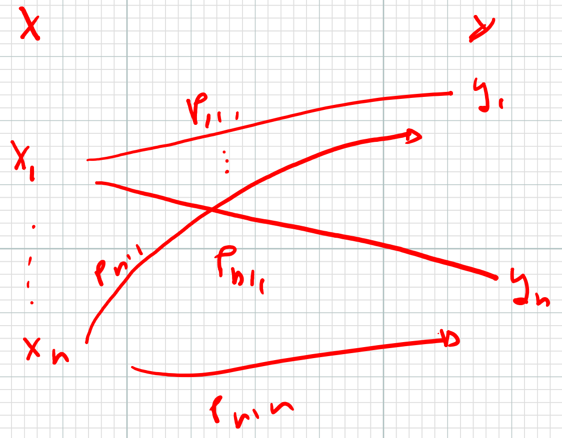
# TN3125
# Information and Computation

Lecture 3

3- *Channel capacity*

# Discrete memoryless channels

- Definition. A discrete memoryless channel takes symbols from a discrete alphabet $X$ to symbols of a discrete alphabet $Y$. It is characterized by a set of probability distributions over alphabet $Y$ one for each element of $X$.



# Transition matrix

- A discrete memoryless channel can be represented by the transition matrix

$$\begin{pmatrix} p(y_1|x_1) & \cdots & p(y_{|Y|}|x_1) \\ \vdots & \ddots & \vdots \\ p(y_1|x_{|X|}) & \cdots & p(y_{|Y|}|x_{|X|}) \end{pmatrix}$$

- Question. Do each of the rows add up to one?
- Question. Do each of the columns add up to one?

# Exercises

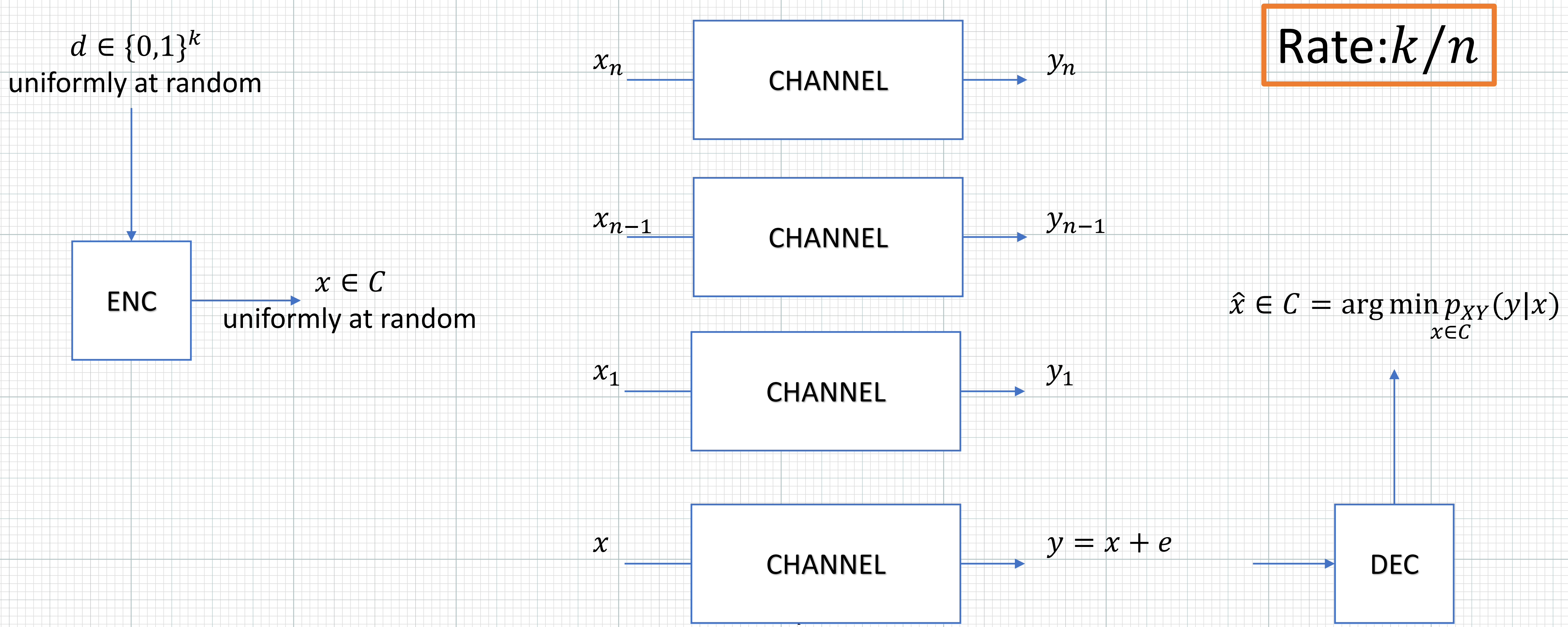
- Write the transition matrix of the binary symmetric and binary erasure channels.

# Exercises

- Write the transition matrix of the binary symmetric and binary erasure channels.

$$\begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} \qquad \begin{pmatrix} 1-e & e & 0 \\ 0 & e & 1-e \end{pmatrix}$$

# Communications setup



# Channel capacity

- Given some discrete and memoryless channel $N$ what is the maximum rate at which it is possible to make the error probability as small as desired by coding over blocks of large enough length?

- Answer:

$$C(N) = \max_{p_X} I(X;Y) \textbf{ in bits}$$

# Fano’s inequality

- We want to guess the value of random variable $X$ from the outcomes of correlated random variable $Y$
- $\hat{X} = g(Y)$ represents our guess
- $E$ takes value $0$ if $\hat{X}$ =X and $1$ when $\hat{X} \neq X$.
- **Exercise**. Show that

$$H(X|Y) \leq H(E) + p_E(E = 1)\log |X|$$

$$H(X|Y) \leq H(E) + p_E(E = 1)\log|X|$$

We will expand $H(E, X|Y)$ in two different ways. First:

$$\begin{aligned}
H(E, X|Y) &= H(E, X, Y) - H(Y) \\
&= H(E, X, Y) - H(Y) + H(E, Y) - H(E, Y) \\
&= H(E|Y) + H(X|E, Y) \\
&\leq H(E) + H(X|E, Y) \\
&= H(E) + p_E(E = 0)H(X|E = 0, Y) + p_E(E = 1)H(X|E = 1, Y) \\
&= H(E) + p_E(E = 1)H(X|E = 1, Y) \\
&\leq H(E) + p_E(E = 1)\log|X|
\end{aligned}$$

$$H(X|Y) \leq H(E) + p_E(E = 1)\log|X|$$

We will expand $H(E, X|Y)$ in two different ways. Second:

$$\begin{aligned}
H(E, X|Y) &= H(E, X, Y) - H(Y) \\
&= H(E, X, Y) - H(Y) + H(X, Y) - H(X, Y) \\
&= H(X|Y) + H(E|X, Y) \\
&= H(X|Y)
\end{aligned}$$

# Rephrasing Fano's inequality

- Lemma. Given a channel, a code $C$ and the message uniformly chosen over the $2^{nR}$ words

$$H(X^n|\mathrm{dec}(Y^n)) \leq 1 + p_e nR$$

- Proof. Follows from direct application of Fano's lemma:

$$H(E) \leq 1$$

$$p_e = P\left(X^n \neq \mathrm{dec}(Y^n)\right)$$

The alphabet of $X^n$ has size $2^{nR}$

# Converse to channel capacity

- Theorem. Given a code for which we choose the codewords uniformly at random, the probability of error over a discrete memoryless channel $N$ is bounded from below by

$$p_e \geq 1 - \frac{1}{nR} - \frac{C(N)}{R}$$

# Proof sketch

- Since codewords are choosing uniformly at random $H(X^n) = nR$
- We can also expand
$$H(X^n) = H\left(X^n \hat{X}^n\right) - H\left(\hat{X}^n\right) + H\left(\hat{X}^n\right) + H(X^n) - H\left(X^n \hat{X}\right)$$

That is $H(X^n) = H\left(X^n \middle| \hat{X}^n\right) + I\left(X^n; \hat{X}^n\right)$

- By the data processing inequality $I\left(X^n; \hat{X}^n\right) \leq I(X^n; Y^n)$ and $I(X^n; Y^n) \leq nC(N)$
- From Fano's inequality $H\left(X^n \middle| \hat{X}^n\right) \leq 1 + p_e nR$
- Putting all together $nR \leq 1 + p_e nR + nC(N)$

# Markov inequality

- Theorem. Given a non-negative random variable $X$

$$\Pr[X \geq x] \leq \frac{\mathbb{E}[X]}{x}$$

# Markov inequality

- Theorem. Given a non-negative random variable $X$

$$\Pr[X \geq x] \leq \frac{\mathbb{E}[X]}{x}$$

- Proof.

$$\mathbb{E}[X] = \sum_{t} t p_X(t) = \sum_{t \geq x} t p_X(t) + \sum_{t < x} t p_X(t)$$

hence: $\mathbb{E}[X] \geq \sum_{t \geq x} t p_X(t)$, moreover

$$\sum_{t \geq x} t p_X(t) \geq x \sum_{t \geq x} p_X(t) = x \Pr[X \geq x]$$

# Union bound

- Given a set of events $A_1, A_2, \dots, A_n$

$$\Pr[A_1 \mathrm{or} A_2 \mathrm{or} \dots \mathrm{or} A_n] \leq \Pr[A_1] + \Pr[A_1] + \cdots + \Pr[A_n]$$

# A random code with rate $R = k/n$

- Choose randomly $2^{nR}$ codewords according to some probability distribution on the input alphabet $p_X$

$$p_{X_1 \ldots X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} p_X(x_i)$$

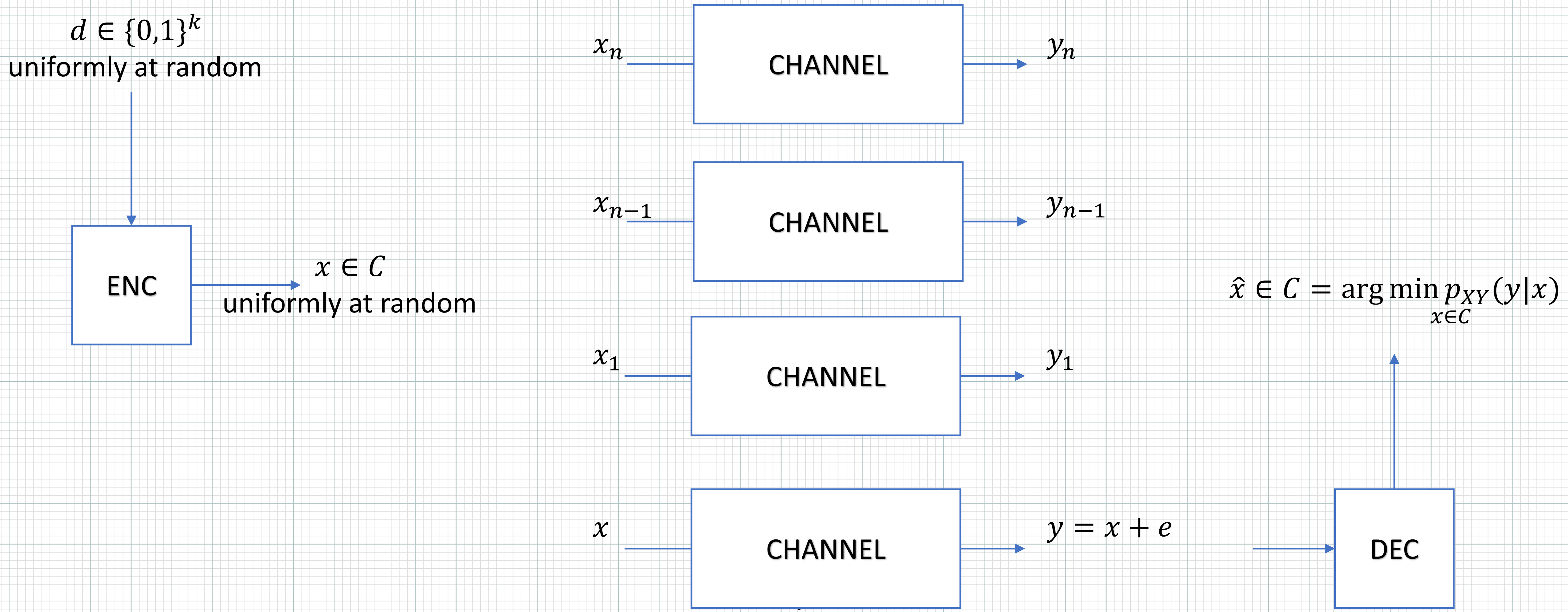
$$C = \begin{pmatrix} x^1 \\ \ldots \\ x^{2^{nR}} \end{pmatrix} = \begin{pmatrix} x_1^1 & \cdots & x_n^1 \\ \vdots & \ddots & \vdots \\ x_1^{2^{nR}} & \cdots & x_n^{2^{nR}} \end{pmatrix}$$

# Remarks

- The code $C$ is an instance of an ensemble of codes induced by $p_X$
- A code $C$ from the ensemble of code $\mathcal{C}$ has probability

$$p_{\mathcal{C}}(C) = \prod_{i=1}^{2^{nR}} \prod_{j=1}^{n} p_X\left(x_j^i\right)$$

# Communication setup



# Proof (1 of 4)

- Fix a codeword $x$ and the output of the channel $y$
- What is the probability over the ensemble of codes that the $i$-th codeword is more likely than $x$?

$$\Pr\left(p_{X^iY}(y|X^i) \leq p_{XY}(y|x)\right) \leq \frac{\mathbb{E}\left(p_{X^iY}(y|X^i)\right)}{p_{XY}(y|x)}$$

$$= \frac{\sum_{x^i \in \{0,1\}^n} p_{X^i}(x^i) p_{X^iY}(y|x^i)}{p_{XY}(y|x)}$$

$$= \frac{p_Y(y)}{p_{XY}(y|x)}$$

# Recap

- Let $e^i$ be the event the $i$-th codeword is more likely than $x$

$$\begin{aligned}\Pr(e^i) &= \Pr\left(p_{X^iY}(y|X^i) \le p_{XY}(y|x)\right)\\ &\le \frac{p_Y(y)}{p_{XY}(y|x)}\\ &= \frac{p_Y(y_n)}{p_{XY}(y_n|x_n)} \cdots \frac{p_Y(y_1)}{p_{XY}(y_1|x_1)}\\ &= \prod_{i=1}^{n} \frac{p_Y(y_i)}{p_{XY}(y_i|x_i)}\end{aligned}$$

# Proof (2 of 4)

- Let's take the log of the previous expression

$$\frac{1}{n}\log\frac{p_Y(y)}{p_{XY}(y|x)} = \frac{1}{n}\sum_{i=1}^{n}\log\frac{p_Y(y_i)}{p_{XY}(y_i|x_i)}$$

- And consider the associated distribution

$$\frac{1}{n}\sum_{i=1}^{n}\log\frac{p_Y(Y_i)}{p_{XY}(Y_i|X_i)}$$

- By the law of large numbers this conveges to

$$\mathbb{E}\left[\log\frac{p_Y(Y_i)}{p_{XY}(Y_i|X_i)}\right] = -I(X;Y)$$

# Proof (3 of 4)

- What is the probability that one of the $2^{nR} - 1$ remaining codewords is more likely than $x$?

$$\Pr\left(e^1 \text{ or } \dots \text{ or } e^{2^{nR}-1}\right) \leq \sum_{i=1}^{2^{nR}-1} \Pr\left(e^i\right)$$

$$\leq \sum_{i=1}^{2^{nR}-1} \frac{p_Y(y)}{p_{XY}(y|x)}$$

$$= 2^{nR} \frac{p_Y(y)}{p_{XY}(y|x)}$$

# Proof (4 of 4)

- From law of large numbers, fix $\epsilon, \delta > 0$, there exists $n$ such that with probability greater than $1 - \epsilon$

$$\frac{1}{n} \log \frac{p_Y(y)}{p_{XY}(y|x)} \leq -I(X;Y) + \delta$$

- Finally, putting all together

$$\Pr(e) \leq 2^{nR} \frac{p_Y(y)}{p_{XY}(y|x)} \leq \epsilon + 2^{nR} 2^{-n(I(X;Y)-\delta)} = \epsilon + 2^{-n(I(X;Y)-R-\delta)}$$

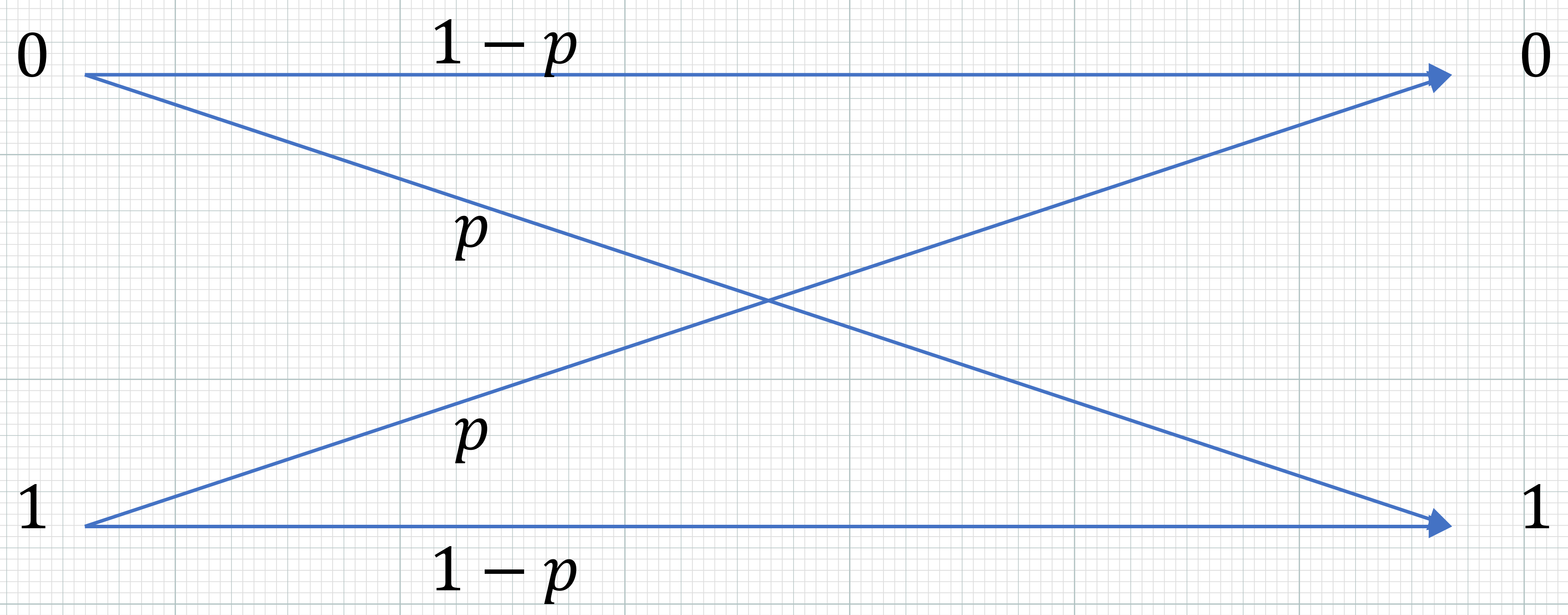
# TN3125
# Information and Computation

Lecture 3

4 – *Computing capacity*

# The binary symmetric channel

- **Exercise.** Find the capacity of the binary symmetric channel.



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

$$\leq 1 - \sum_x p_X(x)\, H(Y|X=x)$$

$$= 1 - H(p, 1-p)$$

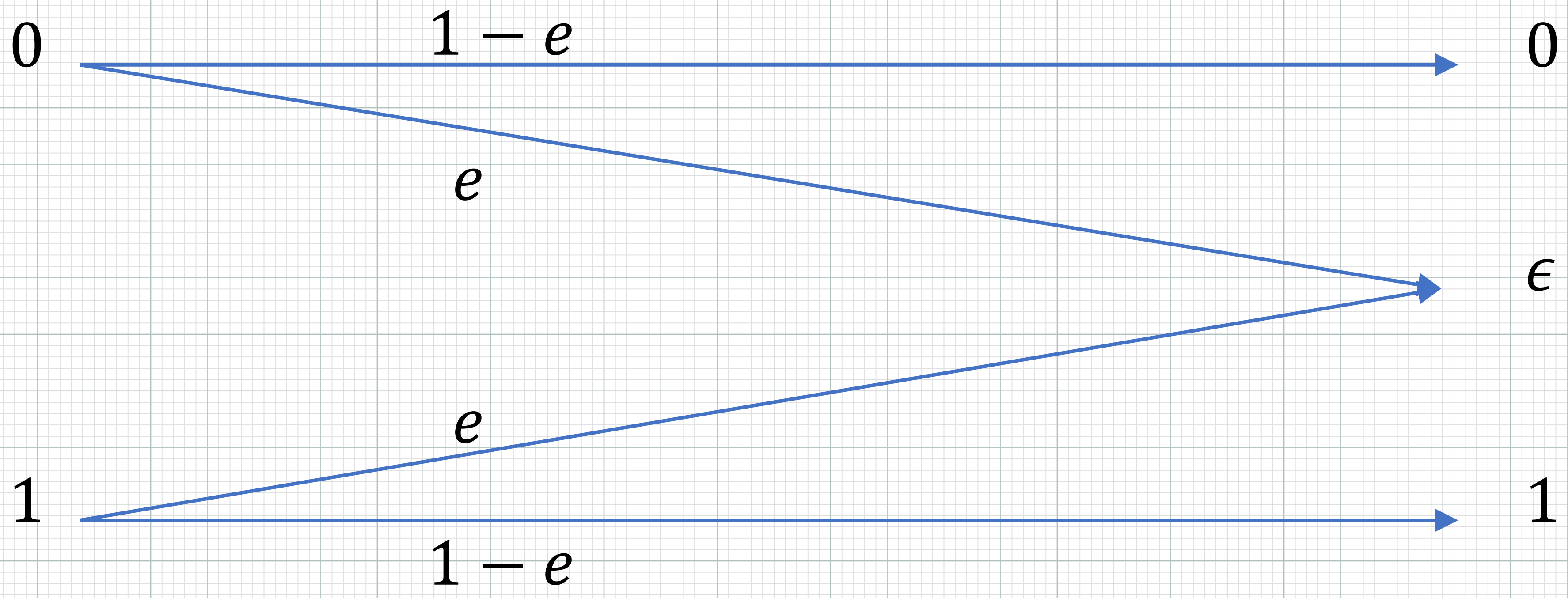
$$p_X(0) = p_X(1) = \frac{1}{2}, \qquad p_Y(0) = p_X(0) \cdot p_{Y|X}(0|0) + p_X(1)\, p_{Y|X}(1|0)$$

$$= \frac{1}{2} \cdot (1-p) + \frac{1}{2} \cdot p = \frac{1}{2}$$

$$I(X:Y) = 1 - H(p, 1-p)$$

# The binary erasure channel

- **Exercise.** Find the capacity of the binary erasure channel.



# The binary erasure channel

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

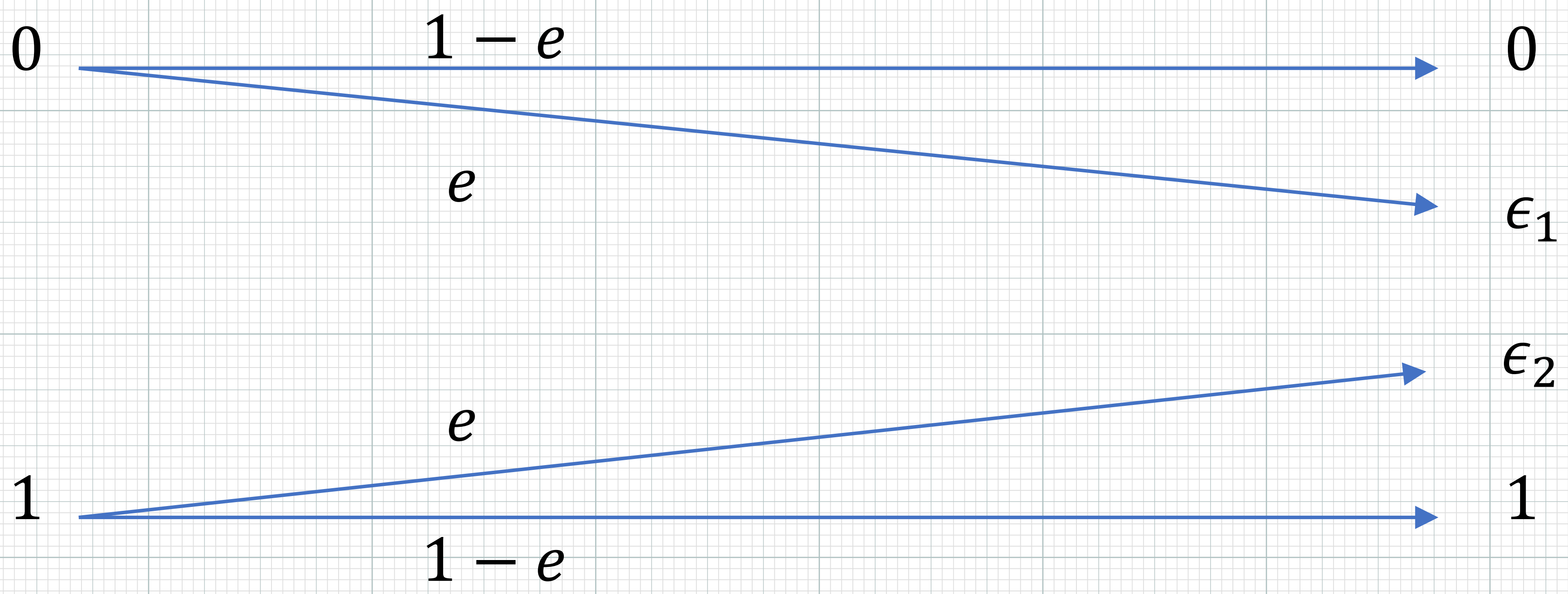
$$= H(p, 1-p) - P_Y(0) \cdot 0 - P_Y(1) \cdot 0 - e \cdot H(p, 1-p)$$

$$= (1-e) \cdot H(p, 1-p)$$

$$\leq (1-e)$$

# The noisy channel with no overlapping output

- **Exercise.** Find the capacity of the following channel.



# Weakly symmetric channels

- **Definition.** A chanel is weakly symmetric if all rows are permutations of each other and all columns have equal sum.

- Example.

$$\begin{pmatrix} 0.2 & 0.4 & 0.4 \\ 0.4 & 0.2 & 0.4 \\ 0.4 & 0.4 & 0.2 \end{pmatrix}$$

# Weakly symmetric channels

- **Exercise.** Show that the capacity of a weakly symmetric channel is

$$C = \log|Y| - H(\mathrm{row})$$

where $r$ is one of the rows of the transition matrix of the channel.

# Solution

- $I(X;Y) = H(Y) - H(Y|X) = H(Y) - H(\mathrm{row}) \leq \log Y - H(\mathrm{row})$

- Can we achieve the upper bound? Let's compute the probability of some value $y$ induced by an uniform distribution on the input

$$p_Y(y) = \sum_x p_{XY}(y|x) p_X(x) = \frac{1}{|X|} \sum_x p_{XY}(y|x)$$

- We are done, why?

# Exercise

- Find the capacity of a channel with transition matrix

$$\begin{pmatrix} 0.2 & 0.4 & 0.4 \\ 0.4 & 0.2 & 0.4 \\ 0.4 & 0.4 & 0.2 \end{pmatrix}$$

# You will do great in the exam if you can

- Encode information using the generator of a linear code, decode using the standard array method
- Deduce properties of a code from the parity check or generator matrix
- Find the capacity of simple channels
- Show basic entropic relations similar to Fano’s inequality
- The proof of the noisy coding theorem (slides 59-72) will not be asked in exam

# Recap course

- We derived an information measure from basic axioms
- We proved basic properties of entropy

- We defined several families of codes for data compression
- We proved that average length is bounded by entropy

- We presented linear codes and defined their properties for correction, detection and erasure as well as bounds on code paremeters. One family of codes we studied in detail were Hamming codes.
- We introduced the minimum distance decoder

- We introduced discrete memoryless channels
- We proved the noisy coding theorem and showed that reliable communication is only possible for rates below capacity

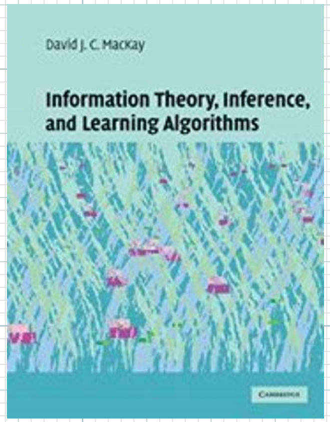
# Outlook

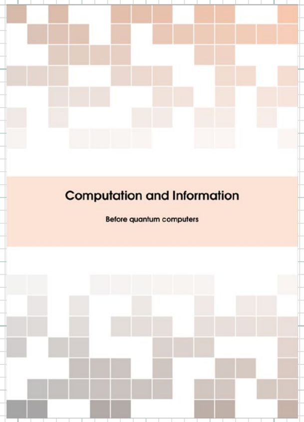
- Similar to the entropy of a random variable we can define entropies for quantum systems, they are at the basis of
  - Quantum information theory
  - Quantum cryptography

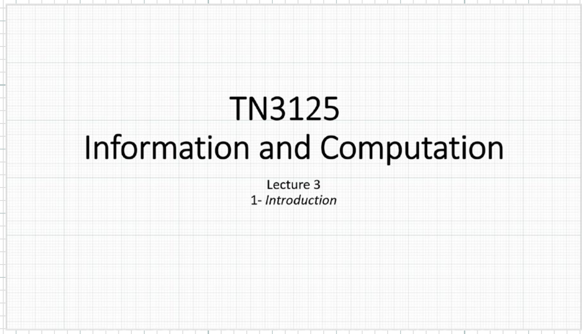
- If you are interested in quantum computation, you will learn about quantum error correcting codes, where the same ideas that you saw here will appear!

# Resources

- Lecture notes
- Slides
- MacKay chapter 9,10



# Ideas for next year

- Content
- Resources
- Pace
- Structure
- Evaluation
- Implementation