Teoretične osnove računalništva 2
Teorija informacij

Introduction: Course presentation with examples

Enes Pasalic UP FAMNIT študijsko leto 25/26

Basic course information

- Lecturer:
  - Prof. dr. Enes Pasalic
  - contact: enes.pasalic6@gmail.com
  - consultation: contact TA Dilawar Abbas Khan dilawarabbasm@gmail.com

- Technical info about the course:
  - Credits: 6
    - lectures 30 h, exercises 45 h,
      90 h individual work
    - Mandatory project, implementation of compression alg. 5-10 pts.

Goals and requirements

- Background (desirable):
  - A bit of analysis,
  - Basics on probability theory.

- Goals:
  - Understand the concepts of entropy and information,
  - Become familiar with two capital results on source compression and channel coding,
  - To acquire knowledge about compression schemes, error correcting codes and transmission over noisy channels

Course literature

- Literature:
  - Thomas M. Cover, J. A. Thomas: **Elements of Information Theory**, 2nd Edition, Wiley Interscience, 2006

- Auxilary literature:
  - David J. C. MacKay: Information Theory, Inference and Learning Algorithms, Cambridge University Press,
  - Nikola Pavešić: Informacija in kodi, Fakulteta za elektrotehniko, Univerza v Ljubljani

- Course staff:
  - Lectures: online (zoom)/white board Enes Pasalic
  - Exercise: asistent Dilawar Abbas Khan

Examination

Written exam	**100%**

- Exam details:
  - Written exam:
    - Mostly problems covering the manndatory parts in the course, rarely theoretical questions but you can expect some proofs from the textbook.
  - Colloquium ? Two parts, requirement at least 25-30% for the first one to attend the second

Lectures

- Course highlights:
  - What is information, coding
  - Examples:
  - Morse alphabet
  - Compression of data: lossless, lossy, example gzip,
  - Noisy communication channels: example BSK, channel capacity BSK, repetition coding, Hamming coding

- Mathematical background of information theory
  - Mathematical analysis:
    - Eksponent and logarithmic function, limits, convex/concav
  - Basics on probability theory:
    - Definition of probability, random variables, distribution, joint and conndiitional distributions, mathmatical expectation, law of large numbers
  - Inequalities:
    - Jensen and Gibbs inequality

Lectures II

- Entropy and information (discrete random var.)
  - Definition of entropy: entropy of joint, conditional variables, chain rules, properties of entropy
  - Mutual information: connection between entropy and mutual information
  - Kullback-Leibler distance (relative entropy)
  - Conditional mutual information: chain rules
  - Stochastic processes and mutual information: Markov chains
  - AEP statement: examples, typical sets, coding with AEP, expected length with AEP coding

- Coding of information
  - Scale example
  - Coding: basics, generalization, decoding
  - Instantenous codes (Prefix-free codes)
  - Kraft inequality
  - Kraft-McMillan statement
  - Codewords lengths and probability
  - Shannon‘s result on compression coding
  - Shannon‘s coding
  - Shannon-Fano coding
  - Huffman coding
  - Problems with fixed length coding
  - Arithnmetic coding

Lectures III

- Image compression example

- Communication channels
  - Cahnnel capacity
  - Examples of channels: eror-free BSK, noisy typewriter, binary symmetric channel, binary channel with erasures
  - Shannon‘s result about channel coding
  - Hamming coding
  - Linear block codes

- Entropy of joint continuous random variables (RV)
  - Definition of Entropy
  - AEP statement
  - Connections with entropy of discrete slučajnih RV
  - Entropy of joint and conditional continuous RVs in pogojnih
  - Properties of entropy
  - Gauss model of communication channel: definition, channnel capacity, frequency limited channels: Nyquist result

Example 1

- Standard HDTV assumes a picture of size 1080 columns.
  - What is the size of digital picture, if required, that the ratio width : height = 16:9?
  - What is the size of one movie with duration of 2 hours, if there are 30 pictures per second and the format is HD?

Express everything in 0 and 1

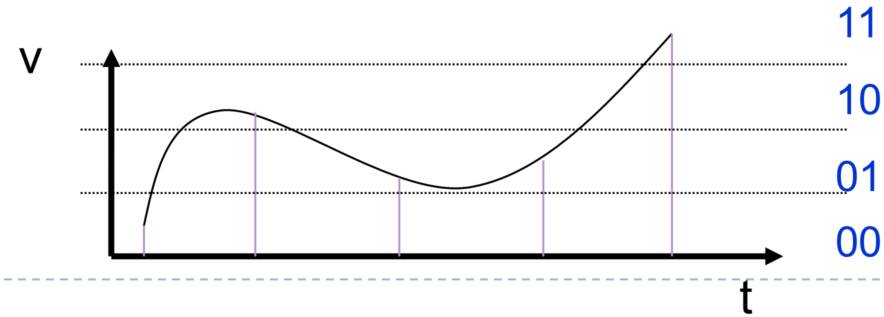
Discrete ensemble:

a,b,c,d $\Rightarrow$ 00, 01, 10, 11

in general: k binary digits specify 2^k messages

Analogue signal:

1) sample and 2) represent sample value binary



Output

00, 10, 01, 01, 11

Efficient: general problem statement

remove redundancy **exact, no errors !!**

remove irrelevance **distortion !!**

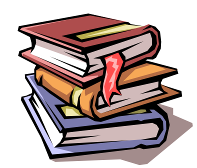
Questions: how ? how good ?

how fast ? how complex ? ...

Efficient: text

- represent every symbol with 8 bit

→ ***1 book: 8 * (500 pages) * 1000 symbols = 4 Mbit ≡ 1 book***



→ compression possible to 1 Mbit (1:4)

Efficient: speech

sampling speed 8000 samples/sec; accuracy 8 bits/sample;

speed 64 kBit/s;

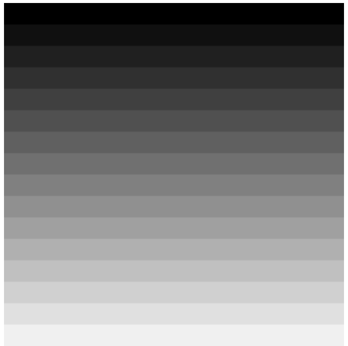
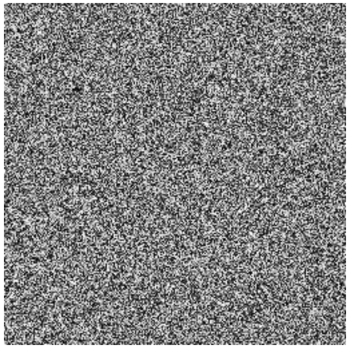
→ ***45 minutes lecture = 45*60*64k =180Mbit ≡ 45 books***

→ compression possible to 4.8 kBit/s (1:10)

Example 2

- Storing digital pictures



How to represent the colors to get a shortest description?

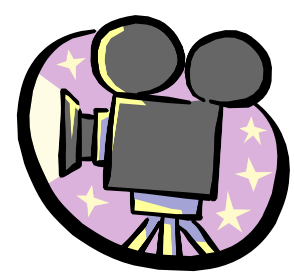
Efficient: digital pictures

300 x 400 pixels x 3 colors x 8 bit/sample (pixel)

→ 2.9 Mbit/picture; for 25 images/second we need 75 Mb/s

2 hour pictures need 540 Gbit ≡ 130.000 books

→ compression needed (1:100)



Efficient: general idea

- represent **likely** symbols with short length binary words
  where likely is derived from

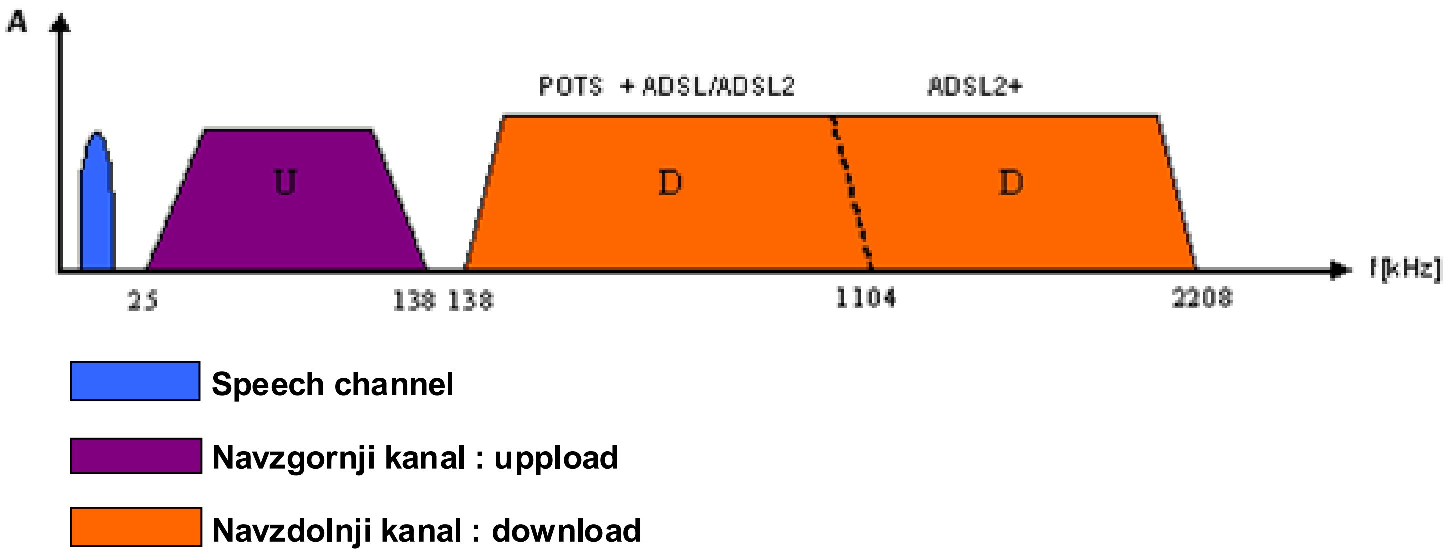
- **prediction** of next symbol in source output

q   qu   q-ue, q-ua, q-ui, q-uo

- **context** between the source symbols, letters in *words*
  context in pictures

Example: wired connnections and data trannsmission

- Frequency bandwidths for xDSL technology



povzeto po B. Batagelj, ULJ, FE

Today‘s lecture

- What is information and information theory.

- Examples:
  - Morse codes
  - Data compression:
    - lossless, lossy,
    - example gzip,
  - Transmitting data over noisy communication channel:
    - example BSK,
    - channel capacity BSK,
    - repetition coding,
    - Hamming coding

- Coding of information over noisy channel (no theory)

Introduction: topic presentation with examples

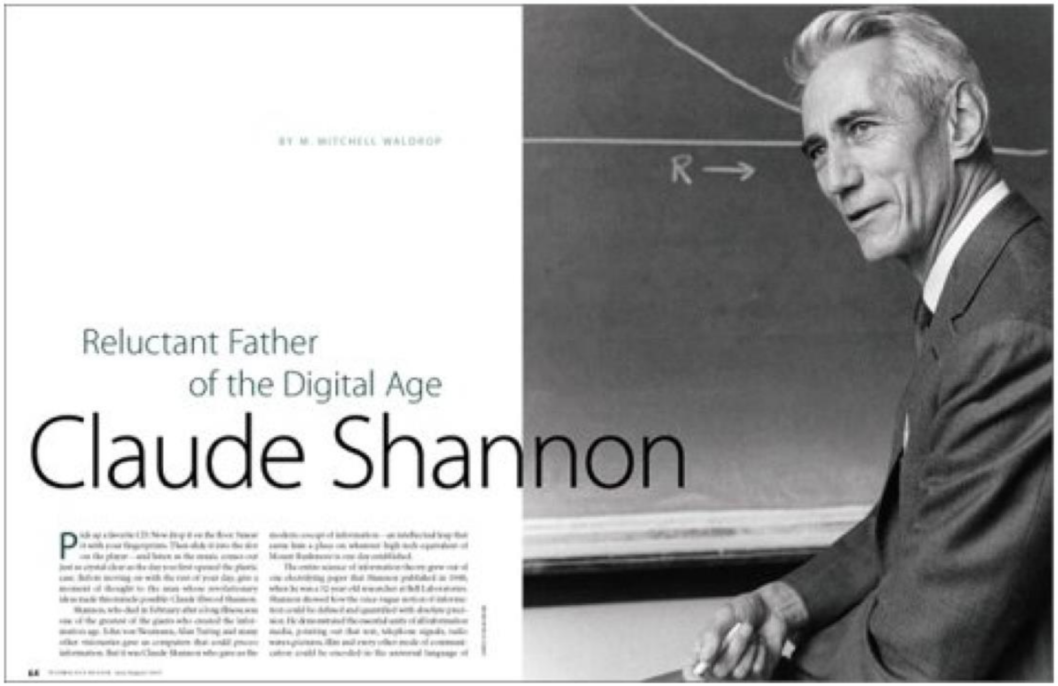
What is information ?

- SSKJ : informacija -e ž (a)
  - kar se o določeni stvari pove, sporoči; *obvestilo, pojasnilo: dati, dobiti informacijo; iskati informacije; imeti dobre, zanesljive informacije; napačna informacija; zahtevali so natančne informacije o bolnikovem zdravstvenem stanju; vir informacij / informacija o dogodku je bila nepotrebna* **informiranje** // mn. celota vednosti o določeni dejavnosti ali področju, namenjena javnosti, podatki: turistične, železniške informacije; izmenjava informacij; oddelek za informacije / radijske, televizijske informacije poročila
  - elektr. množica vrednosti, ki jo (elektronski) računalnik sprejme ali po obdelavi izda: brati, hraniti informacijo; informacijo sestavlja šestdeset bitov / izhodna, vhodna informacija
  - mat. mera za ugotavljanje negotovosti o izidu poskusa: nastanek, uporaba informacije / **teorija informacije, teorija, ki proučuje količinske zakonitosti v zvezi z zbiranjem, prenašanjem in kodiranjem informacij**

Why bother with information theory ?

- Information is almost ‘everything‘ (internet)
- Need for efficient storing, transmission and processing of information.
- Retreiving of information from big data, describing concepts/phenomena, simple or complex.
- Process automatization:
  - (aritificial) inteligence for machines.

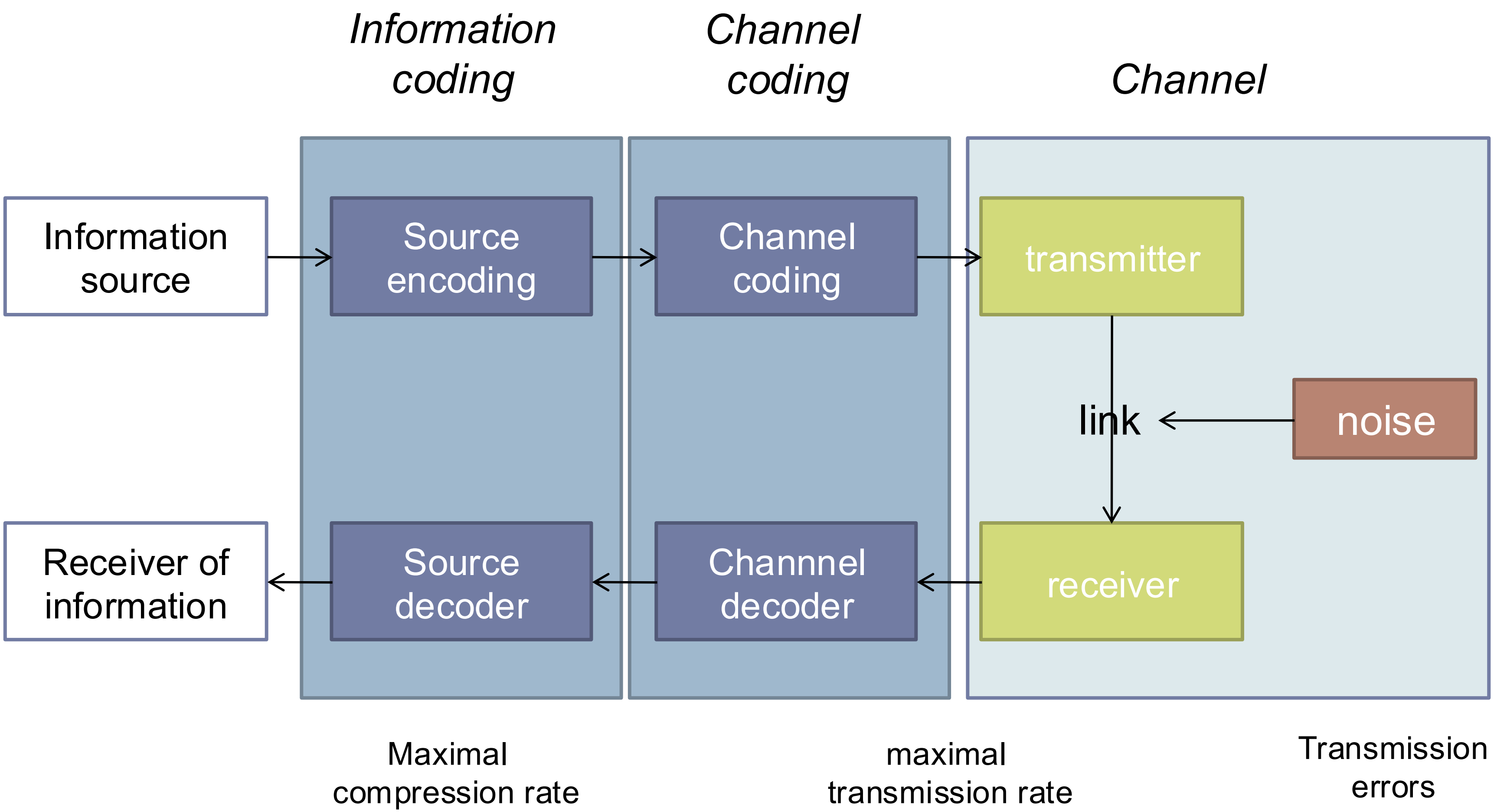
Information theory



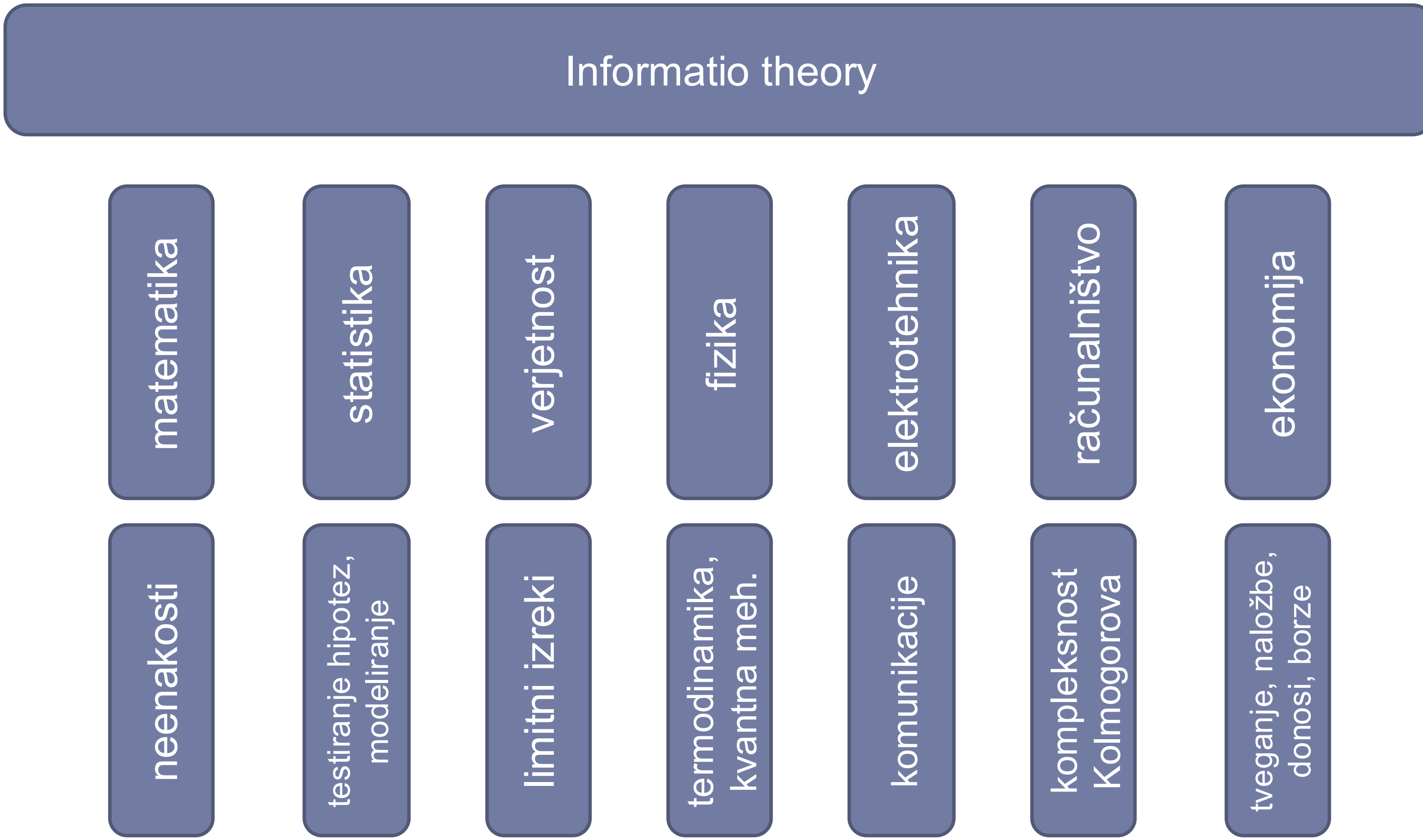
- The father of the modern information theory **Claude Shannon**, who published in 1948 (Bell System Technical Journal) groundreaking article entitled:

  “The Mathematical Theory of Communication”.

Standard communication scheme



Information theory - connections



Example: coding

- Encode letters in English alphabet (A to Z) using dots '.' and underscore '_' , all letters different.

- E.g. A is '.', B is '_ .' etc.

A	________	G	________	M	________	S	________	Y	________
B	________	H	________	N	________	T	________	Z	________
C	________	I	________	O	________	U	________		
D	________	J	________	P	________	V	________		
E	________	K	________	Q	________	W	________		
F	________	L	________	R	________	X	________		

Example: coding

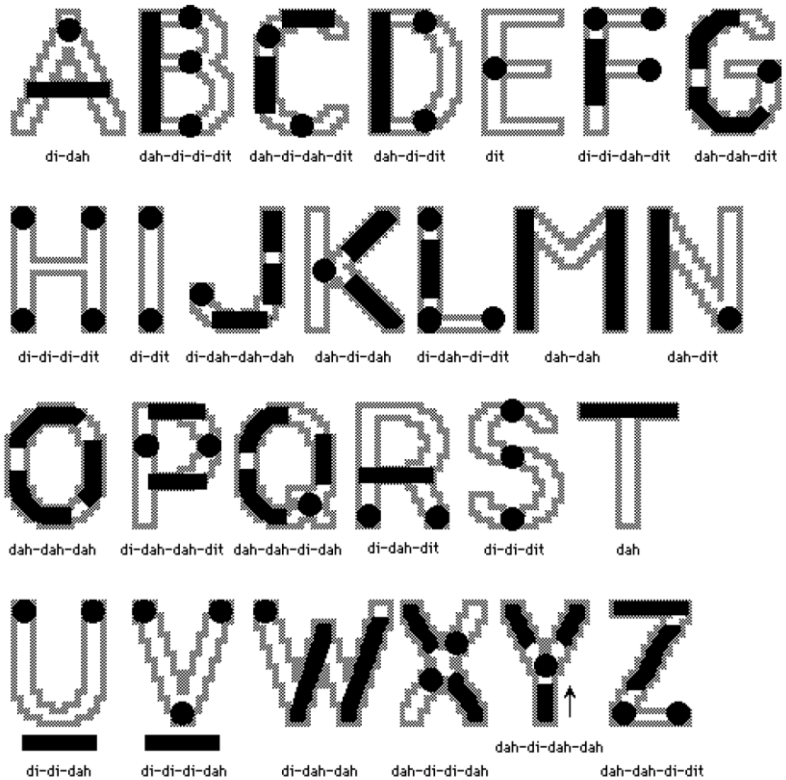
- Now use your coding scheme to encode:

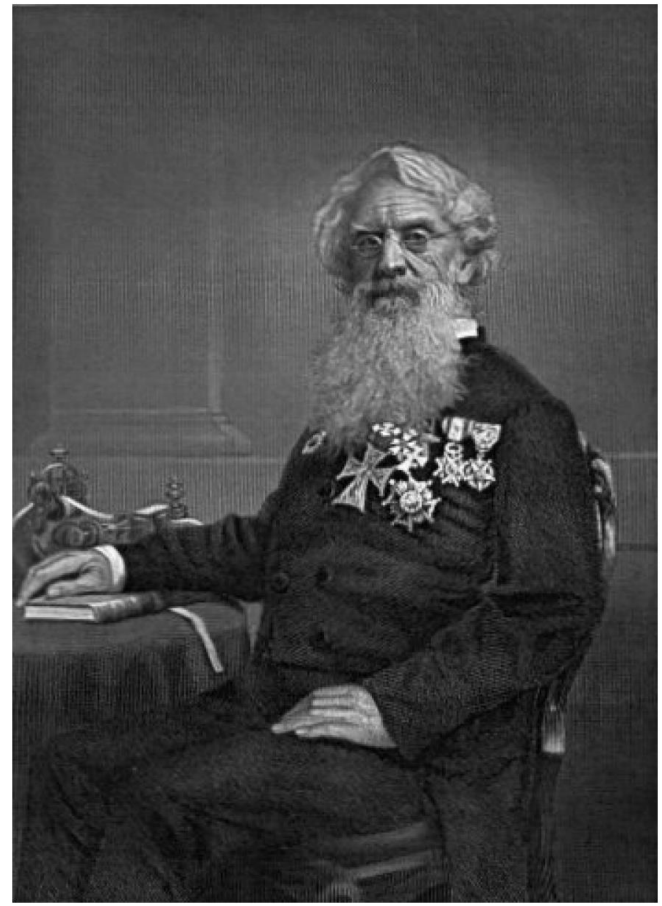
```
TEORIJA INFORMACIJ
```

- How long is your encoded message:
  - dot costs 1 unit
  - undersore costs 2 units
  - Space costs 2 units.
  - Example: . . . - - - . . . :: 1 + 1 + 1 + 2 + 2 + 2 + 1 + 1 + 1 = 12

- **Coding ratio (gain):**
  - Ratio between the length of encoded message and the original one

Morse coding





Samuel F.M. Morse (1791–1872)

Morse coding

TEORIJA INFORMACIJ

- Encoding using the Morse code:

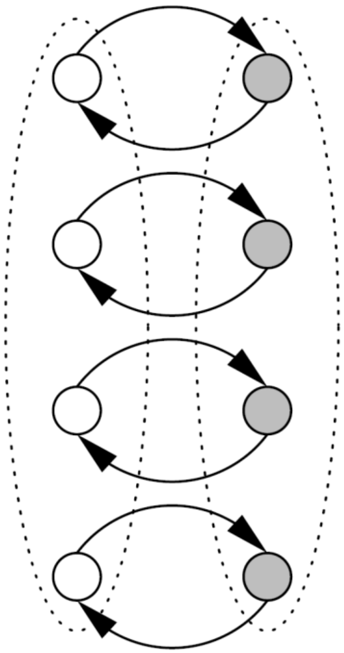
```
- . --- .-. .. .--- .- /
.. -. ..-. --- .-. -- .- -.-. .. .---
```

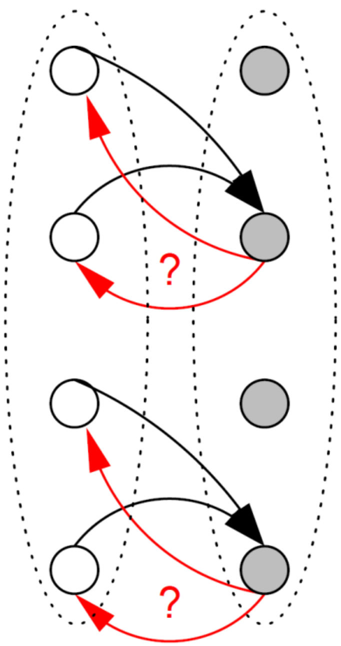
- Coding ratio:
  - 21 dots, 23 unnderscores, 1 space: 21 + 46 + 2 = 69
  - Coding ratio: 69 / 18 = 3.83333…
  - What did you get ?

Mappings between symbols and codewords

lossless codig
injective mapping

lossy coding
not injective mapping





Only lossless codes cann be uniquely decoded.

Example: data compression

		Compression ratio	
general	gzip	~ 1 : 3	lossless compression
	bzip	~ 1 : 3.5	
images	png	~ 1 : 2.5	
	jpeg	~ 1 : 25	Lossy compressionn
music	mp3	~ 1 : 12	
film	mpeg	~ 1 : 30	

Data compression

- Can we indeed compress the data ?

- Statement:
  - **The portion of binary sequnces of length n, that can be compressed by more than k bits, is less than 2^{-k}.**

- Proof: $(1+ 2 + 2^2 + ... + 2^{n-(k+1)})/2^n <= 2^{n-k} / 2^n = 2^{-k}$.

- Consequences?
  - Less than 50% of data can be compressed with more tha k=1 bit.
  - Less than 1% of data can be compressed with more than k=7 bits .
  - Less than 0.000000000000000000000000000001 % of data can be compressed with more than 100 bits.

Data compression

- Then what !?
- Why do we get compression ratio which is much larger than claimed by this result ?
- What data (informatio sources), we can compress (encode)?

Example: data compression

```
echo  <x>  | gzip -  | wc  -c      #multiply with 8, to get bits
```

vhodni niz		dolžina zakodiranega niza	razmerje
aaaaa … a	(10000 x a)	368	27.2 : 1
aabaabbbbbabbbbb …	(10000 a,b randomly)	13456	0.74 : 1
ababababab …	(5000 x ab)	368	27.2 : 1
aa … abb … b	(5000 x a, 5000 x b)	376	26.6 : 1
abbaababba …	(1000 x *abbaababba*)	488	20.5 : 1
aaabbabbabb …	(π, 0-4 is a, 5-9 is b)	13416	0.74 : 1

- Sequences, with regularities/rules, can be easily compressed.
- π also assigns rule for a and b, but gzip cannot compress it. Why?

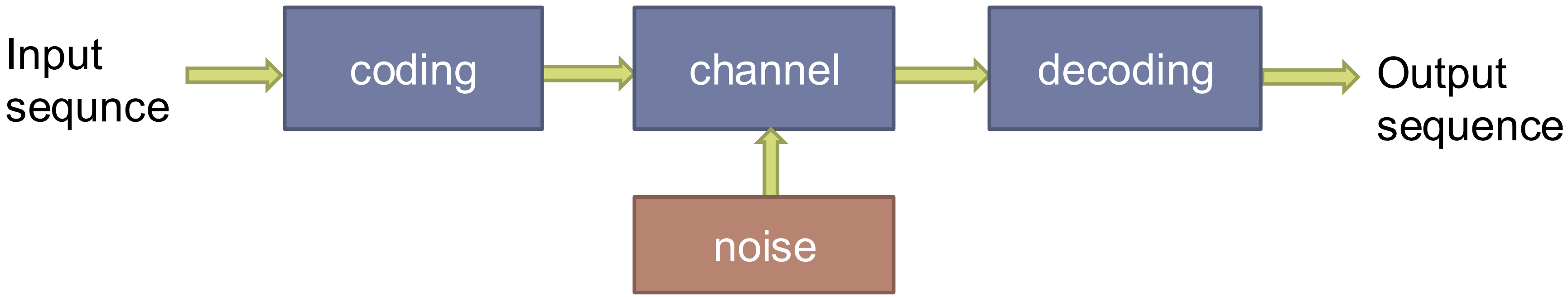
Real-life communication channels

- In practice, communication channels are noisy, inntroduce errors.

- Examples:
  - modem connections,
  - Satelite communication,
  - hard disc, DVD, CD, …

- **Main problem**: How to transmit the information reliably over a noisy channel so that we can reconstruct the information (perfectly) even though some bits are not correct ?

Channel coding



- We want to minimize:
  - Length of the input seqeuce
    (less transmission over the channel)
  - Probability of error of the output sequence
    (match between input and output as much as possible)

Example: repetition coding

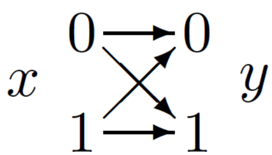
- Repeat input symbols several times.

```
 I  N  F  O  R  M  A  C  I  J  E

IIINNNFFFOOORRRMMMAAACCCIIIJJJEEE

IIIHNNFFFOOBRRRMMMAAACCCJILJJJEEF

 I  N  F  O  R  M  A  C  ?  J  E
```

- **Transmission rate**: 1 : 3 (repeating each symbol 3 times9
- Binary seq.:

$\mathbf{s}$	0	0	1	0	1	1	0
$\mathbf{t}$	000	000	111	000	111	111	000
$\mathbf{n}$	000	001	000	000	101	000	000
$\mathbf{r}$	000	001	111	000	010	111	000
$\hat{\mathbf{s}}$	0	0	1	0	0	1	0

Binary symmetric channel (BSC)



$$
\begin{aligned}
P(y=0 \mid x=0) &= 1-f; & P(y=0 \mid x=1) &= f; \\
P(y=1 \mid x=0) &= f; & P(y=1 \mid x=1) &= 1-f.
\end{aligned}
$$

Shannon result on channel coding

- BSC:

- Assume, we want to transmit the information through BSC channel with **the probability of error *10^{-15}*** through a BSC channel, where *$f = 0.1$*.

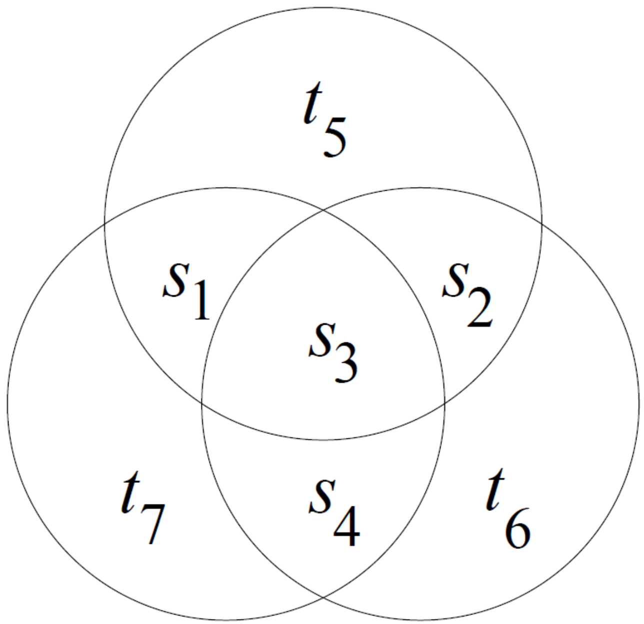
- If we use **repetition coding**, one can show that we need to **repeat each symbol 61 times**.

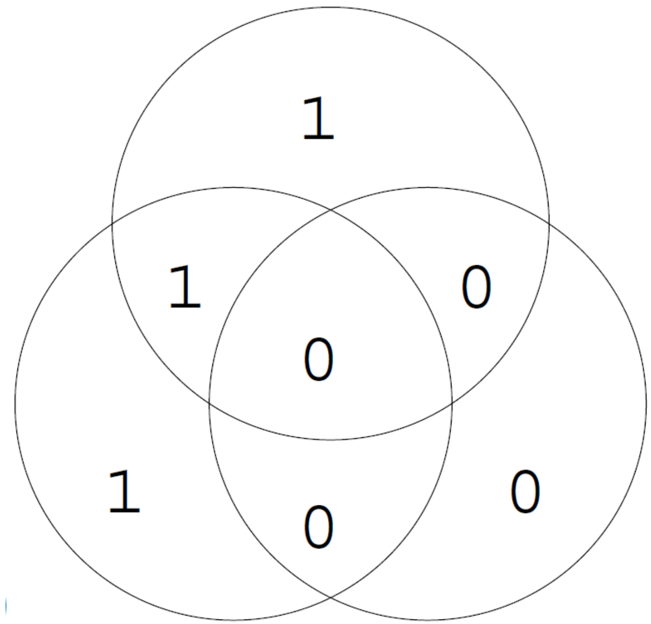
Parity checking

- One method to correct transmission errors is **parity checking** for the codewords.
- If we add a single bit to any codeword we can make any codeword to have an even weight. We can easily detect if a single error has occured.
- We can actually add several bits to each codeword and detect+correct more errors !!
- Example: Hamming coding

Hamming coding: Hamming (7,4)

- Coding:



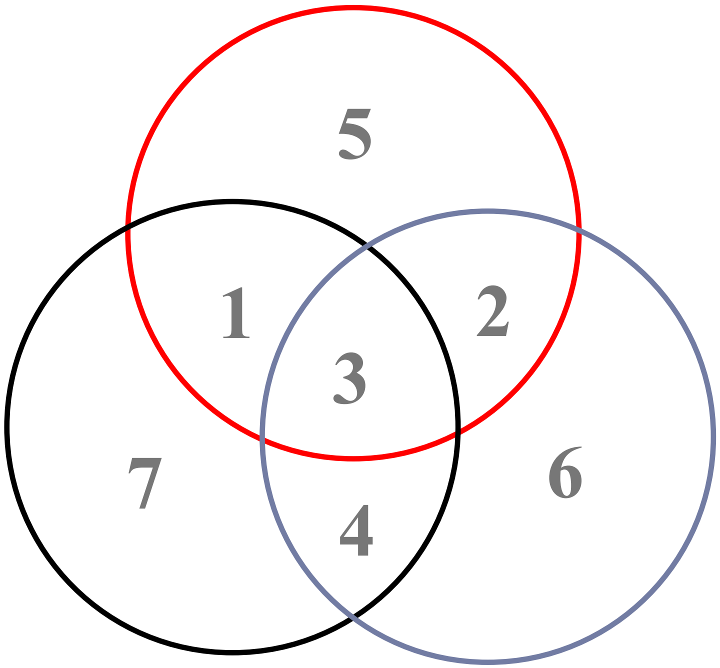


Input sequence : $s_1 s_2 s_3 s_4$
Add 3 bits : $t_5 t_6 t_7$
Coded seq. : $s_1 s_2 s_3 s_4\ t_5 t_6 t_7$

Input sequence : 1000
Add 3 bits : 101
Coded seq. : 1000101

Coding: $t_5 = 1$, if $s_1 + s_2 + s_3 = 1 \mod 2$
$t_5 = 0$, if $s_1 + s_2 + s_3 = 0 \mod 2$

Hamming Code – Parity Checks



1	2	3	4	5	6	7
1	1	1	0	1	0	0
0	1	1	1	0	1	0
1	0	1	1	0	0	1

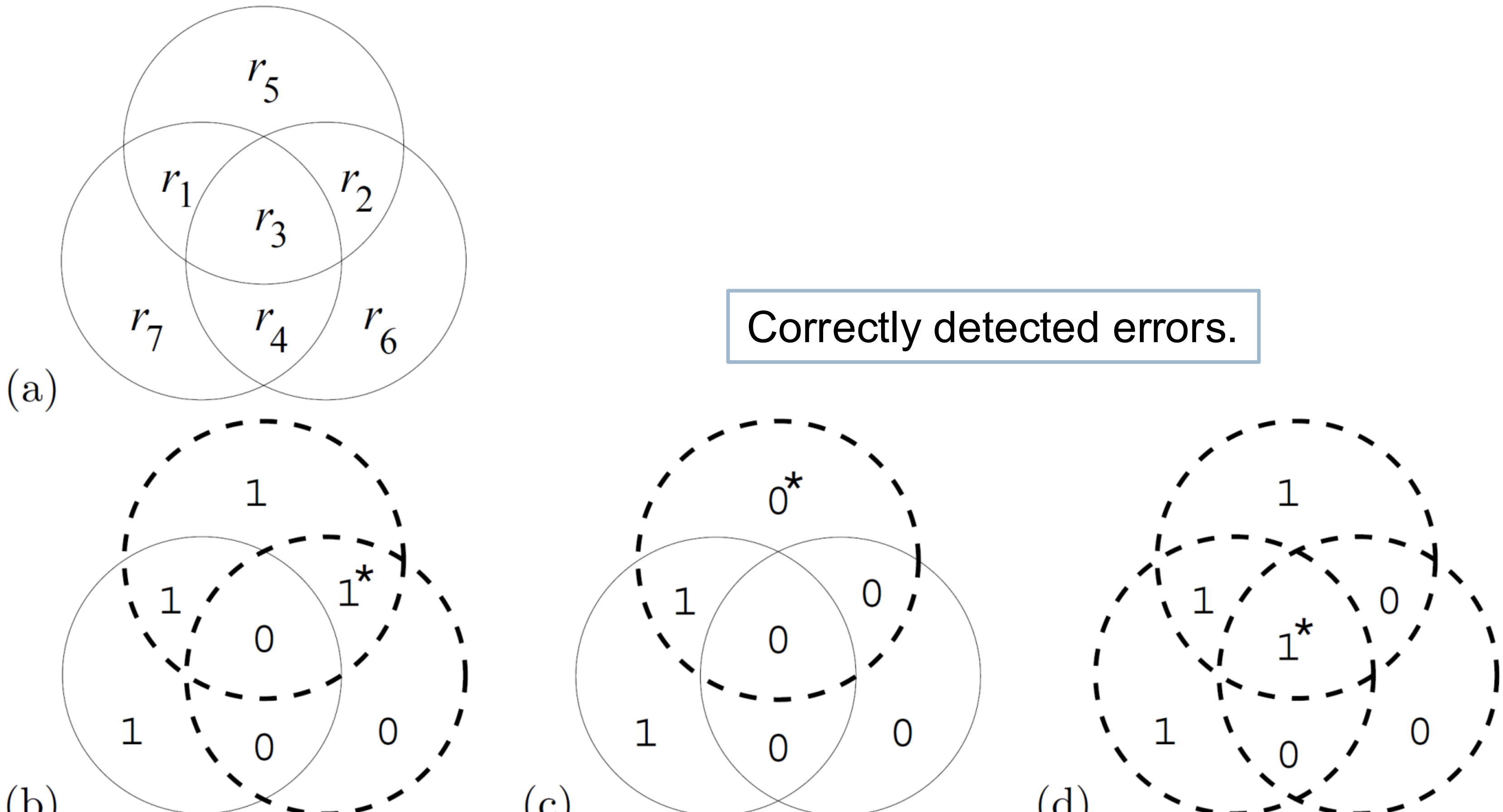
Parity-Check Equations

- Parity bits implies system of linear equations.

$$H = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \qquad \begin{array}{l} c_1 + c_2 + c_3 + c_5 = 0 \\ c_1 + c_3 + c_4 + c_6 = 0 \\ c_1 + c_2 + c_4 + c_7 = 0 \end{array}$$

- Parity-check matrix H is not unique.
- Any set of vectors that span the row space generated by H can serve as the rows of a parity check matrix (including sets with more than 3 vectors).

Hamming decoding: Hamming (7,4)



Correctly detected errors.

Repetition coding vs Hamming coding

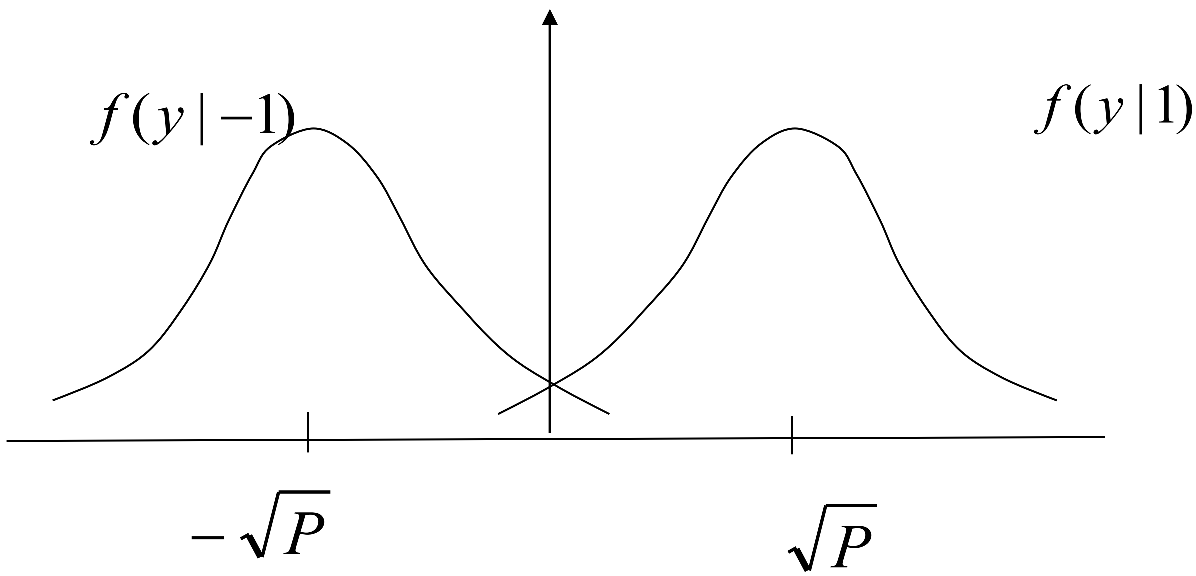
- Repetition coding R3:
  - Information ratio: 1:3
  - Error probability (assuming independence and symmetry of bit errors):

$$3(1-f)f^2 + f^3 \approx 3f^2 \Rightarrow O(f^2)$$

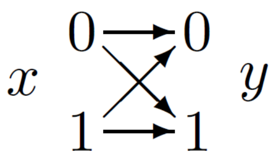
- Hamming coding Hamming(7,4)
  - Information ratio: 4 : 7
  - Error probability (assuming independence and symmetry of bit errors): $O(f^2)$

- Conclusion:
  - Better information ratio for Hamming and similar error probability.

AWGN BSK channel



Binary symetric channel (BSC)



$$P(y=0\,|\,x=0) = 1-f; \quad P(y=0\,|\,x=1) = f;$$
$$P(y=1\,|\,x=0) = f; \quad P(y=1\,|\,x=1) = 1-f.$$

- **Channel capacity (Shannon):**

$$C(f) = 1 - H_2(f) = 1 - \left[f \log_2 \frac{1}{f} + (1-f) \log_2 \frac{1}{1-f} \right]$$

H(f) is entropy. We define and explain it later.

- <u>Example:</u> *C(0.1) ~= 0.53. information ratio c.a.* : 1 : 2

Shannon result on channel coding

- Shannon (informal):
  If the information ratio is SMALLER than the channel capacity, there exist coding technique such that the probability of error is ARBITRARY SMALL !

- Converse is also true:
  If the information ratio is LARGER than the channel capacity, there does not exist coding scheme such that the probability of error is ARBITRARY SMALL !

- What does that mean?

Define information (transmission) rate

$$R = \frac{\text{number of source symbols}}{\text{number of transmitted symbols}}$$

- Repetition coding: Sourse symbols 0 and 1
Transmitted symbols 000 and 111

Can always correct a single error since if 000 becomes 001 (e.g.) then the receiver decides that the sender sent 000 (and decode it as 0)

Rate for this code is $\frac{1}{3}$

Better protection of 0 and 1 if we repeat it say 100 times, thus encode $0 \to \underbrace{000 \ldots 000}_{100\times}$ and $1 \to \underbrace{111 \ldots 111}_{100\times}$

BUT — this in Not for free, thing about bandwidth!

Example: Your source is sending symbols "0" or "1" at rate 1kbit/second.

If you want to protect your symbols over a noisy channel by repeating each "0" or "1" hunderd times you are sending the data at rate 100 × 1kbit/sec!

SHANNON: Exist codes with arbitrary small prob. of error if $R < C$. If $C = 0.53$ then take $R = 0.5$

BUT such codes are defined for extremely large blocks

Take e.g. 1 million bits and add 1 million bits as a protection, implying that $R = 1/2$.

PROBLEM: If you have blocks of 10^6 bits you have $2^{1000000}$ possible blocks (??) that you need to encode and create a list for your encoding!! Impossible of course

Blocks of 30 bits are doable (but it is not SHANNON, decrease R!!)

Shannon result on channel coding

- BSC:
- Assume, we want to transmit the information through BSC channel with the probability of error 10^{-15} skozi kanal BSC, where $f = 0.1$. If we use repetition coding, one can show that we need to repeat each symbol **61 times**.

- Shannon's claim: **2 times** is enough, since the capacity of channel is C=0.53 !!

- BSC (second part):
  Now we want to transmit through BSC with f=0.1 (same) with probability of error 10^{-100}.
- Shannon this time: **2 times** is enough !!!!

Summary

- Infomation theory concerns two main areas:
  - How to compress data, to get the shortest possible footprint, either for transmission or when storing on a hard disc
  - **Shannon's claim:**
    A sequence of length N of independent indentically distributed random variables (iid), we can compress it to N*H(X) bits (with arbitrary small probability oflosing information.
  - How to transfer information over a noisy channel so that we can encode it (add redundancy) so that it can be recovered with arbitrary small probability of error..
    - **Shannon claims:**

    if the information rate is cmaller than the capacity of channel, then there exists acoding scheme such that the probabilityof error is arbitrary.