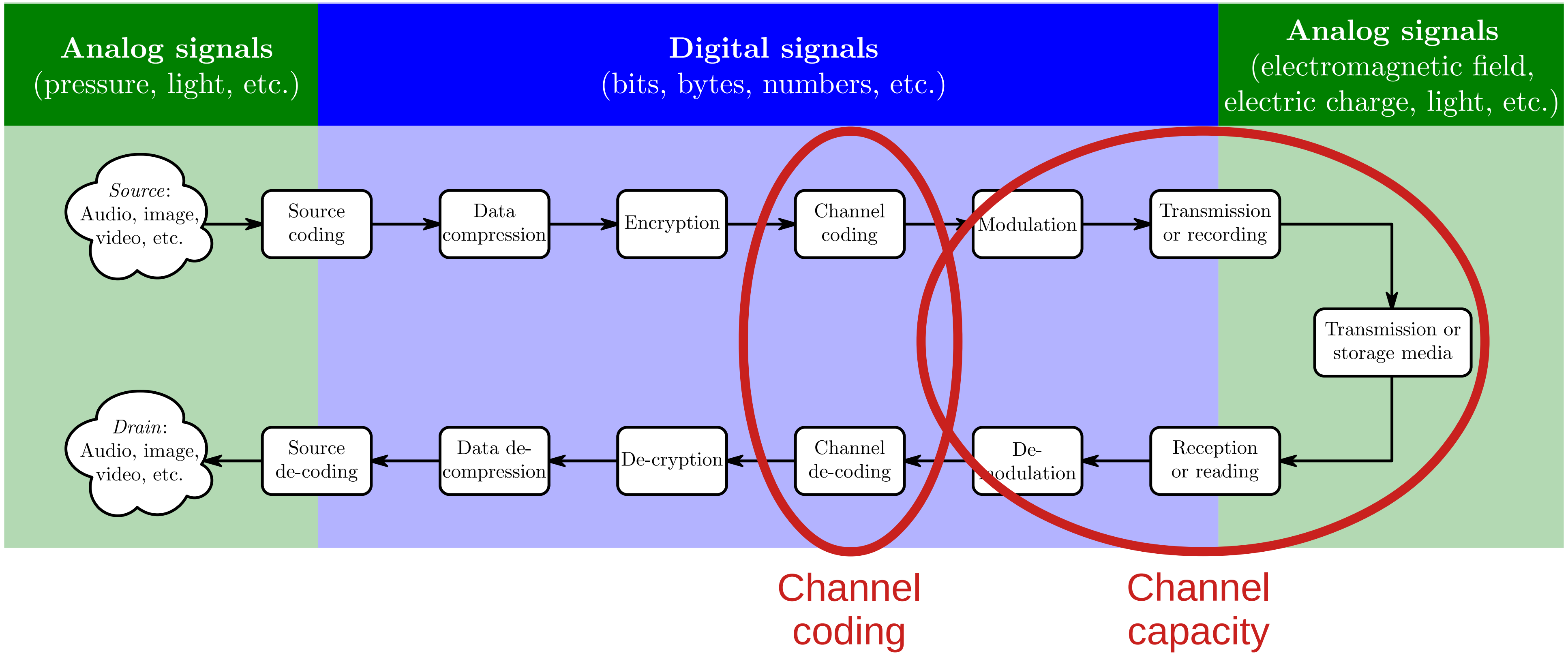
Information Transmission Chapter 5, Channel coding

OVE EDFORS
ELECTRICAL AND INFORMATION TECHNOLOGY

Learning outcomes

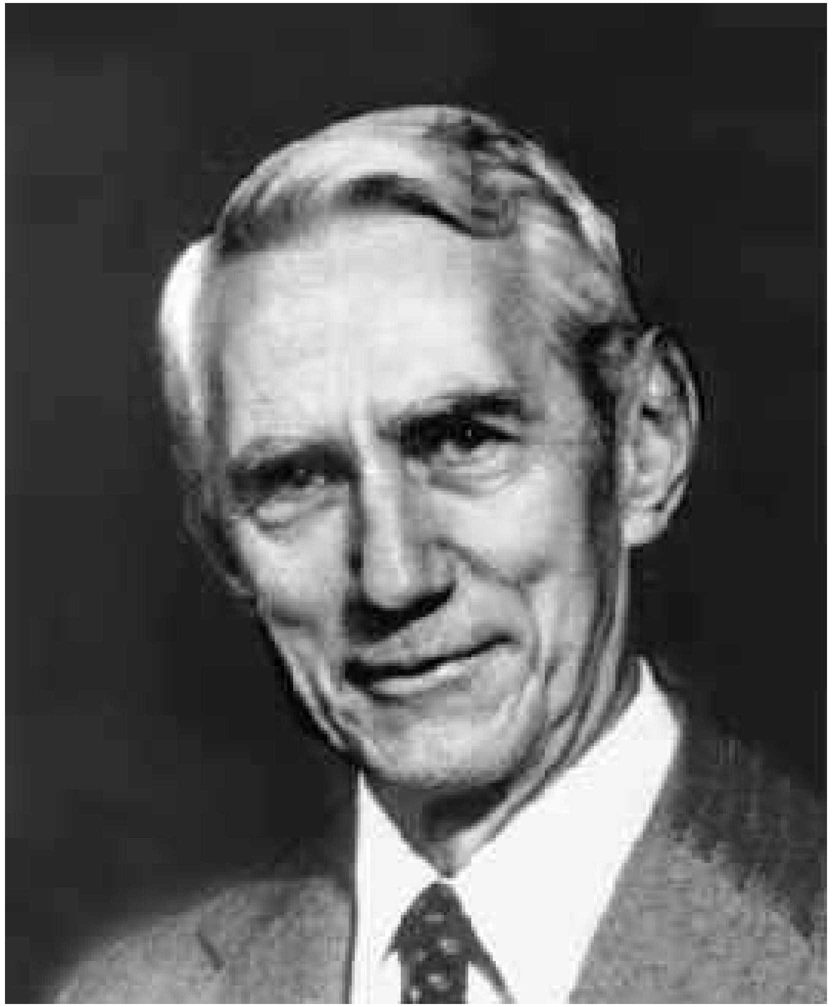
- After this lecture the student should
  - understand the principles of channel coding,
  - understand how typical sequences can be used to find out how "fast" we can send information over a channel,
  - have a basic knowledge about how channel capacity is related to mutual information and its maximization over the channel input distribution
  - know how to calculate the channel capacity for the binary symmetric channel and the additive white Gaussian noise (AWGN) channel

Where are we in the BIG PICTURE?



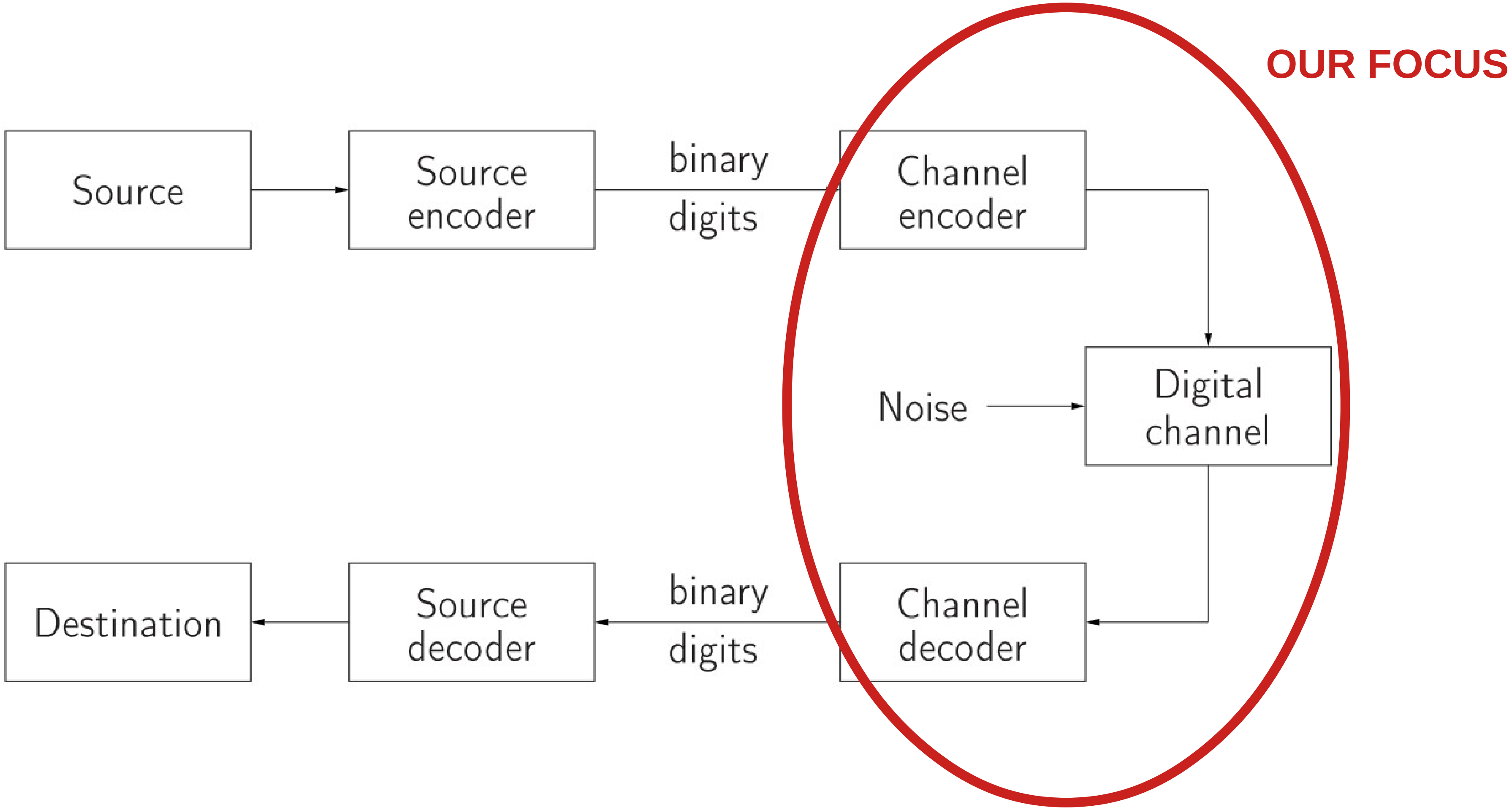
Lecture relates to pages 166-179 in textbook.

What did Shannon promise?



- As long as the SNR is above -1.6 dB in an AWGN channel we can provide reliable communication

A schematic communication system



Typical sequences

REP.

All typical long sequences have approximately the same probability and from the law of large numbers it follows that the set of these typical sequences is overwhelmingly probable.

The probability that a long source output sequence is typical is close to one, and, there are approximately

$$2^{nh(p)}$$

typical long sequences.

Example from textbook (draw from urn)

REP.

sequence	probability	
●●●●●	1/3 1/3 1/3 1/3 1/3 ⇒ 0.0041	
●●●●○	1/3 1/3 1/3 1/3 2/3 ⇒ 0.0082	
●●●○●	1/3 1/3 1/3 2/3 1/3 ⇒ 0.0082	
●●●○○	1/3 1/3 1/3 2/3 2/3 ⇒ 0.0165	
●●○●●	1/3 1/3 2/3 1/3 1/3 ⇒ 0.0082	
●●○●○	1/3 1/3 2/3 1/3 2/3 ⇒ 0.0165	
●●○○●	1/3 1/3 2/3 2/3 1/3 ⇒ 0.0165	
●●○○○	1/3 1/3 2/3 2/3 2/3 ⇒ 0.0329	⋆
●○●●●	1/3 2/3 1/3 1/3 1/3 ⇒ 0.0082	
●○●●○	1/3 2/3 1/3 1/3 2/3 ⇒ 0.0165	
●○●○●	1/3 2/3 1/3 2/3 1/3 ⇒ 0.0165	
●○●○○	1/3 2/3 1/3 2/3 2/3 ⇒ 0.0329	⋆
●○○●●	1/3 2/3 2/3 1/3 1/3 ⇒ 0.0165	
●○○●○	1/3 2/3 2/3 1/3 2/3 ⇒ 0.0329	⋆
●○○○●	1/3 2/3 2/3 2/3 1/3 ⇒ 0.0329	⋆
●○○○○	1/3 2/3 2/3 2/3 2/3 ⇒ 0.0658	⋆
○●●●●	2/3 1/3 1/3 1/3 1/3 ⇒ 0.0082	
○●●●○	2/3 1/3 1/3 1/3 2/3 ⇒ 0.0165	
○●●○●	2/3 1/3 1/3 2/3 1/3 ⇒ 0.0165	
○●●○○	2/3 1/3 1/3 2/3 2/3 ⇒ 0.0329	⋆
○●○●●	2/3 1/3 2/3 1/3 1/3 ⇒ 0.0165	
○●○●○	2/3 1/3 2/3 1/3 2/3 ⇒ 0.0329	⋆
○●○○●	2/3 1/3 2/3 2/3 1/3 ⇒ 0.0329	⋆
○●○○○	2/3 1/3 2/3 2/3 2/3 ⇒ 0.0658	⋆
○○●●●	2/3 2/3 1/3 1/3 1/3 ⇒ 0.0165	
○○●●○	2/3 2/3 1/3 1/3 2/3 ⇒ 0.0329	⋆
○○●○●	2/3 2/3 1/3 2/3 1/3 ⇒ 0.0329	⋆
○○●○○	2/3 2/3 1/3 2/3 2/3 ⇒ 0.0658	⋆
○○○●●	2/3 2/3 2/3 1/3 1/3 ⇒ 0.0329	⋆
○○○●○	2/3 2/3 2/3 1/3 2/3 ⇒ 0.0658	⋆
○○○○●	2/3 2/3 2/3 2/3 1/3 ⇒ 0.0658	⋆
○○○○○	2/3 2/3 2/3 2/3 2/3 ⇒ 0.1317	
	0.9998	

● - probability 1/3
○ - probability 2/3

Number of typical sequences should be about:

$$2^{nh(p)} = 2^{5h(1/3)} = 2^{5\cdot 0.9183} \approx 24$$

Sequences with “observed uncertainty” within 15% of h(1/3) (probability between 0.027 and 0.068):

15 (the ones marked with stars)

Why the large discrepancy?

Only valid for “long” sequences.

… but the 15 sequences are less than 1/2 of all sequences and contain about 2/3 of all probability.

LUND UNIVERSITY

Properties of typical sequences

REP.

Let $\mathcal{T}_\epsilon^{(n)}$ *be the set of sequences* $\boldsymbol{x} = x_1 x_2 \ldots x_n$ *such that*

$$2^{-n(H(X)+\epsilon)} \leq P_{\boldsymbol{X}}(\boldsymbol{x}) \leq 2^{-n(H(X)-\epsilon)} \qquad (5.43)$$

The set $\mathcal{T}_\epsilon^{(n)}$ *is called the typical set and it has the following properties:*

1. *If* $\boldsymbol{x} \in \mathcal{T}_\epsilon^{(n)}$, *then* $P_{\boldsymbol{X}}(\boldsymbol{x}) \approx 2^{-nH(X)}$.
2. $Pr(\mathcal{T}_\epsilon^{(n)}) > 1 - \epsilon$, *for* n *sufficiently large.*
3. $|\mathcal{T}_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon)}$.

LUND UNIVERSITY

Longer typical sequences

REP.

Let us now choose a smaller ϵ namely $\epsilon = 0.046$

(5% of h(1/3)), and increase the length of the sequences.

Then we obtain the following table:

n	$\lvert\mathcal{T}_\epsilon^{(n)}\rvert$	$Pr(\mathcal{T}_\epsilon^{(n)})$
100	$2^{92.6}$	0.660
500	$2^{474.9}$	0.971
1000	$2^{953.4}$	0.998
2000	$2^{1910.3}$	1.000

Note: In the first example with length-five sequences we had a wider tolerance of 15% of h(1/3), and captured 2/3 of the probability in our typical sequences.

With this tighter tolerance we need sequences of length 100 to capture 2/3 of the total probability in the typical sequences.

Typical sequences in text

REP.

If we have L letters in our alphabet, then we can compose L^n different sequences that are n letters long.

Only approximately $2^{nH(X)}$, where $H(X)$ is the uncertainty of the language, of these are "meaningful".

What is meant by "meaningful" is determined by the structure of the language; that is, by its grammar, spelling rules etc.

LUND UNIVERSITY

Typical sequences in text

REP.

Only a fraction

$$\frac{2^{nH(X)}}{L^n} = \frac{2^{nH(X)}}{2^{n\log_2 L}} = 2^{-n(\log_2 L - H(X))},$$

which vanishes when *n* grows provided that $H(X) < \log_2 L$, is ”meaningful” text of length *n* letters.

For the English language *H*(*X*) is typically 1.5 bits/letter and $\log_2 L = \log_2 26 \approx 4.7$ bits/letter.

LUND UNIVERSITY

Structure in text

REP.

Shannon illustrated how increasing structure between letters will give better approximations of the English language.

Assuming an alphabet with 27 symbols – 26 letters and one space – he started with an approximation of the first order.

The symbols are chosen *independently* of each other but with the actual probability distribution (12 % E, 2 % W, etc.):

OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA

TH EEI ALHENHTTPA OOBTTVA NAH BRL

Structure in text

REP.

Then Shannon continued with the approximation of the second order. The symbols are chosen with the actual *bigram* statistics – when a symbol has been chosen, the next symbol is chosen according to the actual conditional probability distribution:

ON IE ANTSOUTINYS ARE T INCTORE ST BE S

DEAMY ACHIN D ILONASIVE TUCOOWE AT

TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE

Structure in text

REP.

The approximation of the third order is based on the *trigram* statistics – when two successive symbols have been chosen, the next symbol is chosen according to the actual conditional probability distribution:

IN NO IST LAT WHEY CRATICT FROURE BIRS

GROCID PONDENOME OF DEMONSTRURES OF THE

REPTAGIN IS REGOACTIONA OF CRE

The principle of source coding

REP.

Consider the set of typical long output sequences of n symbols from a source with uncertainty $H(X)$ bits per source symbol.

Since there are fewer than $2^{n(H(X)+\epsilon)}$ typical long sequences in this set, they can be represented by $n(H(X)+\epsilon)$ binary digits; that is, by $H(X)+\epsilon$ binary digits per source symbol.

Channel coding

Block coding basics

Divide the information-sequence to be transmitted into blocks $\boldsymbol{u} = [\, u_1\ u_1\ \ldots\ u_K\,]$ of K bits.

… 1001|1110|1010|0011|1010|1111|1110 …

Divided into blocks of 4 bits here

There are 2^K different blocks $\boldsymbol{u}$ of K information bits (here 16).

For each unique block of information bits, assign a unique code word $\boldsymbol{x} = [\, x_1\ x_2\ \ldots\ x_N\,]$ of length $N > K$ bits. Let's use $N = 7$. Note that this is a subset of all possible sequences of length N.

Encode your information sequence by replacing each information block $\boldsymbol{u}$ with the corresponding code word $\boldsymbol{x}$.

… 0011001|0010110|0100101|1000011|1011010|1111111|0010110 …

7 bit code words here

$\mathbf{u}$	$\mathbf{x}$
0000	0000000
0001	1101001
0010	0101010
0011	1000011
0100	1001100
0101	0100101
0110	1100110
0111	0001111
1000	1110000
1001	0011001
1010	1011010
1011	0110011
1100	0111100
1101	1010101
1110	0010110
1111	1111111

This is called an (N,K) block code, with **code rate**

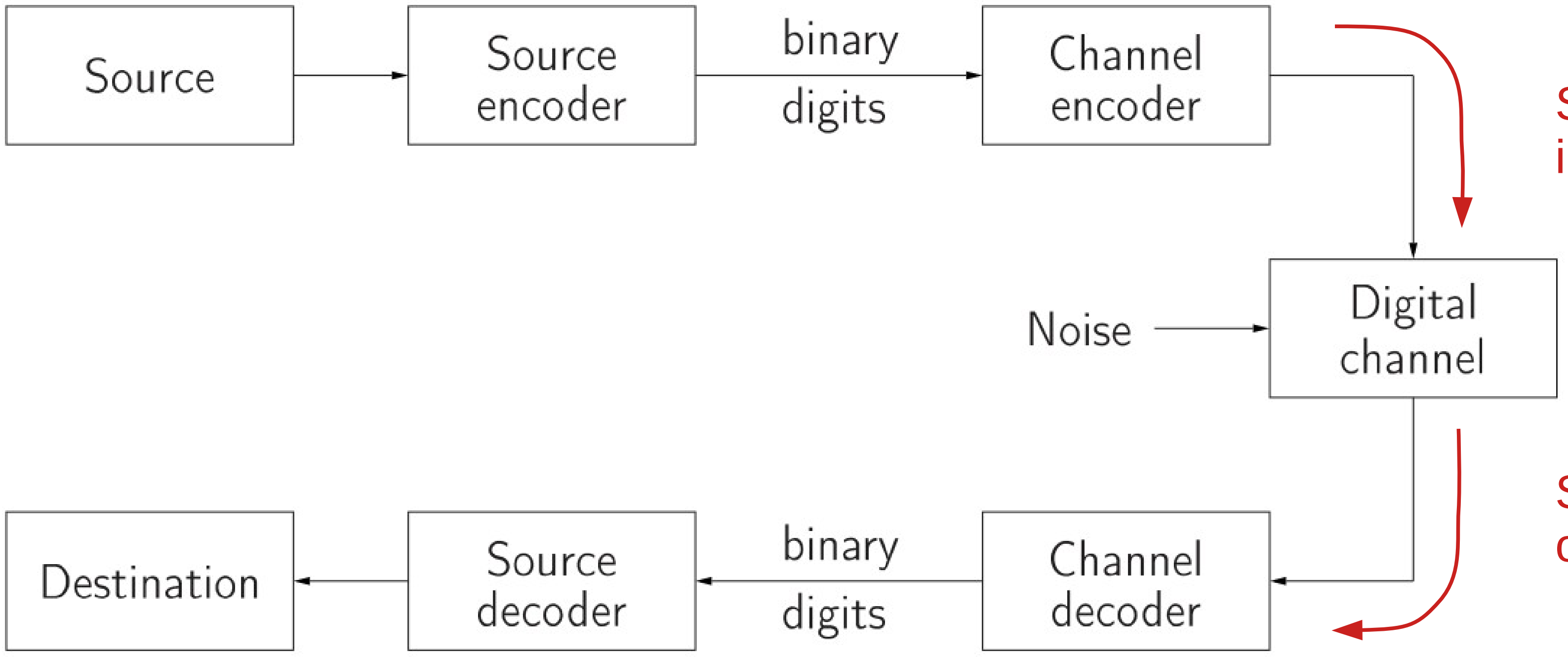
$$R = \frac{K}{N}$$

and in this case it is a (7,4) code with rate

$$R = \frac{4}{7}$$

LUND UNIVERSITY

Digital channel – symbols in and out



Our code words

Symbols X as input to channel

Symbols Y as output from channel

Our code words, corrupted by the noisy channel

Fans (of a typical input sequence and its typical output sequences)

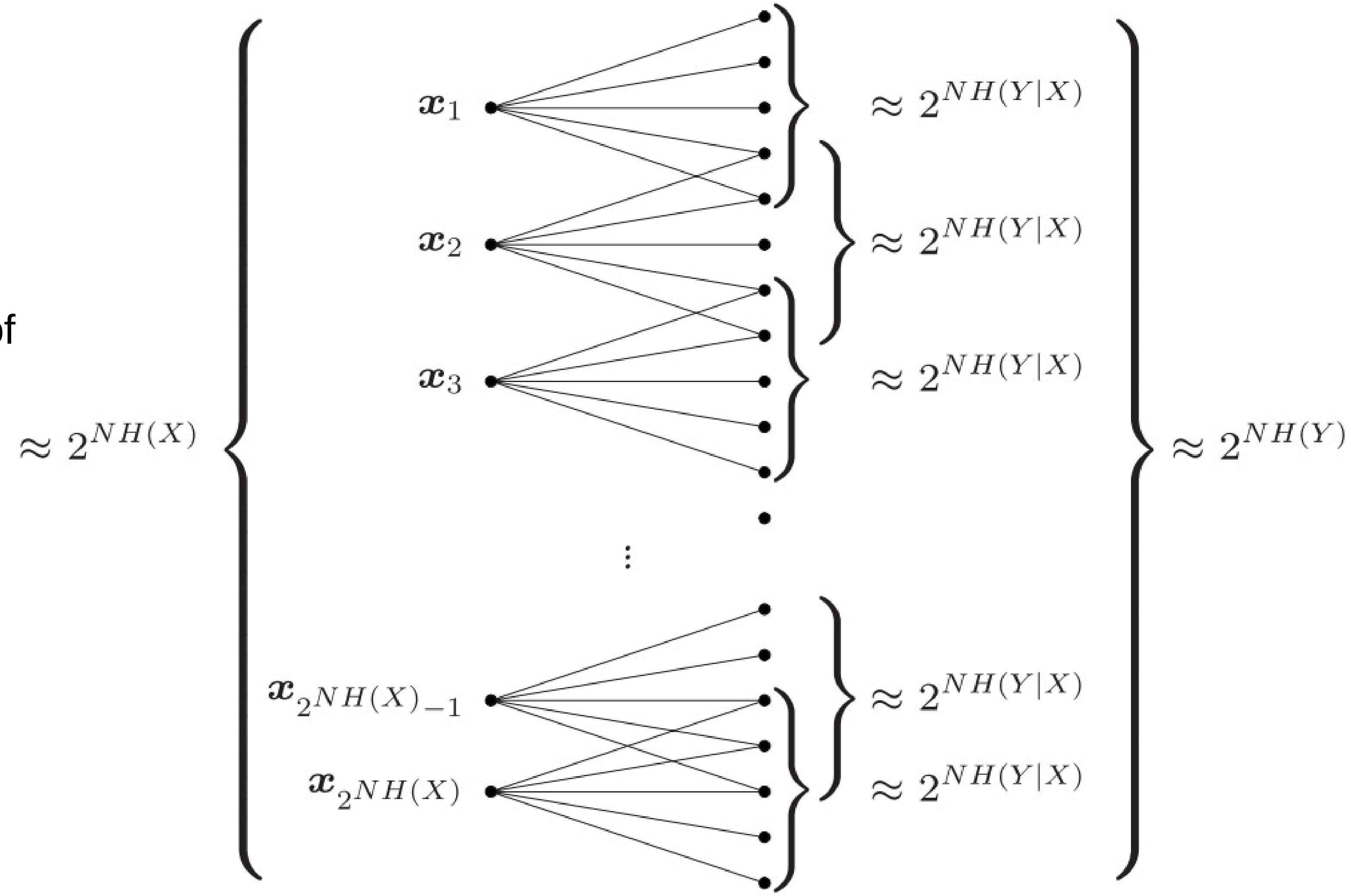
Consider a channel with input X and output Y.

Then we have approximately $2^{NH(X)}$ and $2^{NH(Y)}$ typical input and output sequences of length N, respectively.

Furthermore, for each typical long input sequence we have approximately $2^{NH(Y|X)}$ typical long output sequences that are jointly typical with the given input sequence, we call such an input sequence together with its jointly typical output sequences a *fan*.

LUND UNIVERSITY

Input sequences of length *N*



Output sequences of length *N*

We can have at most

$$\frac{2^{NH(Y)}}{2^{NH(Y|X)}} = 2^{N(H(Y)-H(Y|X))} = 2^{NI(X;Y)}$$

non-overlapping fans

LUND UNIVERSITY

Maximum rate

Each fan can represent a message. Hence, the number of distinguishable messages, $M = 2^K = 2^{RN}$, can be at most, $2^{NI(X;Y)}$, that is $2^{RN} = 2^{NI(X;Y)}$

Equivalently, the largest value of the rate *R* for non-overlapping fans is

$$R = I(X;Y) \text{ bits/channel use}$$

Channel capacity

Since we would like to communicate with as high code rate *R* as possible we choose the input symbols according to the probability distribution $P_X(x)$ that maximizes the mutual information *I(X;Y).* This maximum value is defined as the *capacity* of the channel,

$$\boxed{C \stackrel{\text{def}}{=} \max_{P_X(x)} \{I(X;Y)\} \text{ bit/channel use}}$$

Channel capacity

Let the encoder map the messages to the typical long input sequences that represent non-overlappling fans, which requires that the code rate R is at most equal to the capacity of the channel, that is,

$$R \leq C$$

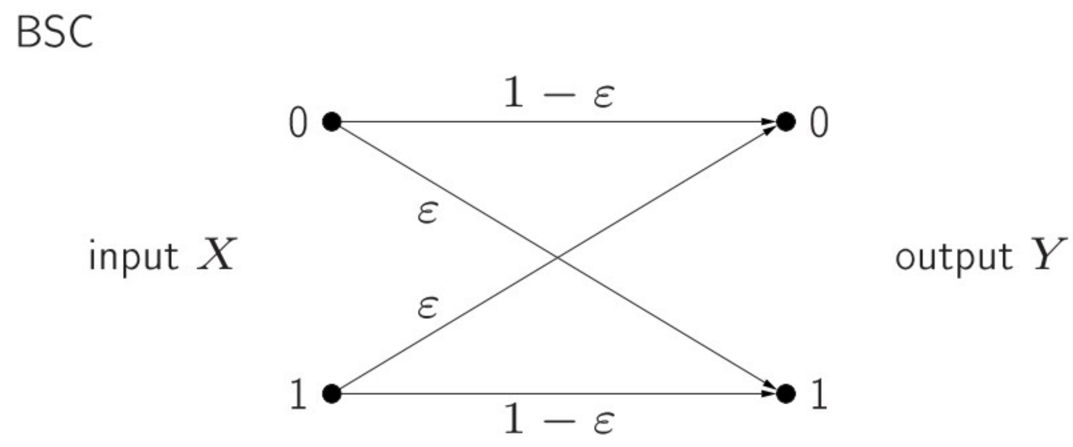
Then the received typical long output sequence is used to identify the corresponding fan and, hence, the corresponding typical long input sequence, or, equivalently, the message, and this can be done correctly with a probability arbitrarily close to 1.

Channel coding theorem

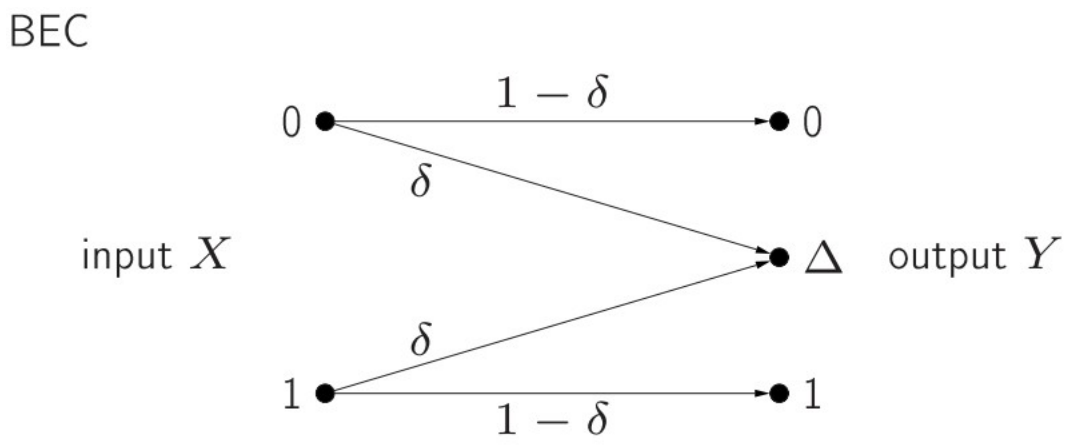
Suppose we transmit information symbols at rate $R=K/N$ bits per channel using a block code via a channel with capacity C.

Provided that $R<C$ we can achieve arbitrary reliability, that is, we can transmit the symbols virtually error-free, by choosing N large enough. Conversely, if $R>C$, then significant distortion must occur and reliable communication is not possible.

Binary symmetric channel (BSC)
Binary erasure channel (BEC)



ε = BSC error probability



δ = BEC erasure probability

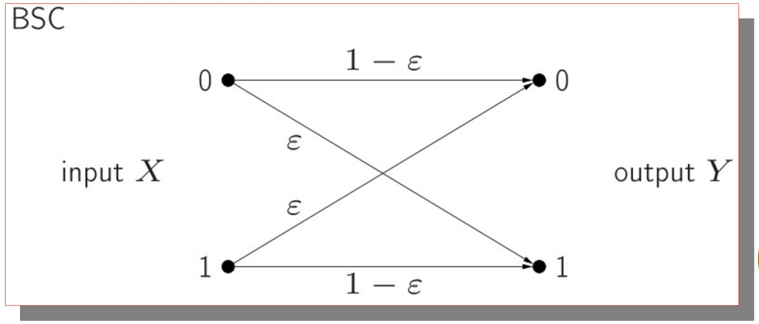
LUND UNIVERSITY

Channel capacity of the BSC

$$
\begin{aligned}
C_{BSC} &= \max\{I(X;Y)\} \\
&= \max\{(H(Y) - H(Y \mid X))\} \\
&= 1 - \left(- \sum_{i=1}^{2} \sum_{j=1}^{2} P_{XY}(x_i, y_j) \log P_{Y|X}(y_j \mid x_i) \right) \\
&= 1 + \sum_{i=1}^{2} \sum_{j=1}^{2} P_{Y|X}(y_j \mid x_i) P_X(x_i) \log P_{Y|X}(y_j \mid x_i) \\
&= 1 + 2(1-\varepsilon)\frac{1}{2}\log(1-\varepsilon) + 2\varepsilon\frac{1}{2}\log\varepsilon \\
&= 1 + (1-\varepsilon)\log(1-\varepsilon) + \varepsilon\log\varepsilon \\
&= 1 - h(\varepsilon) \ \ \text{bits/channel use}
\end{aligned}
$$

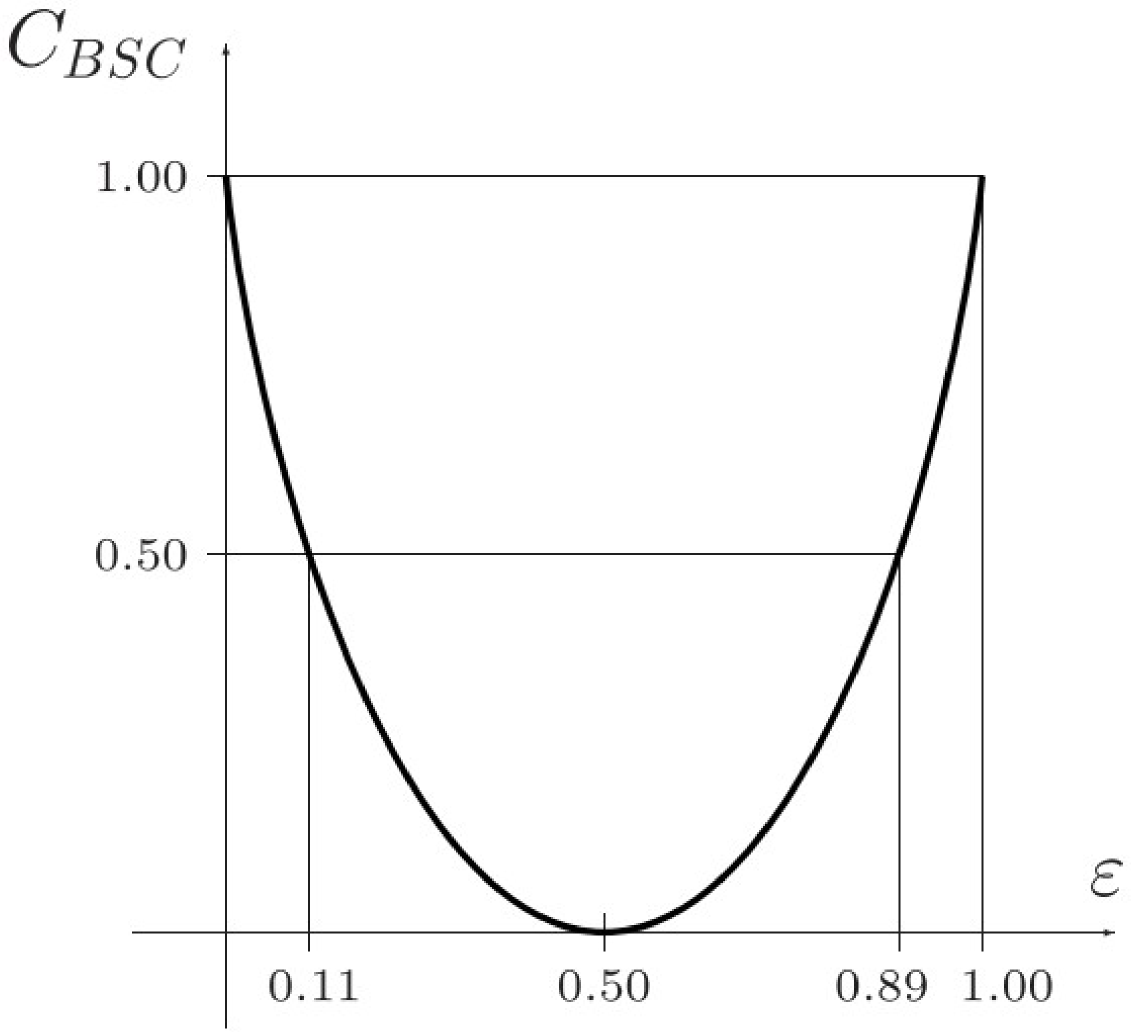
Since the channel is symmetric (behaves the same for 0 and 1) we can assume that the maximizing input distribution is

$$P_X(0) = P_X(1) = \frac{1}{2}$$



LUND UNIVERSITY

Channel capacity for the BSC



The additive white Gaussian noise (AWGN) channel

So far we have considered only channels with binary inputs. Now we shall introduce the time-discrete Gaussian channel whose output Y_i at time i is the sum of the input X_i and the noise Z_i

$$Y_i = X_i + Z_i$$

where X_i and Y_i are real numbers and Z_i is a Gaussian random variable with mean 0 and variance $N_0/2$.

Capacity of the AWGN

A natural limitation on the inputs is an average energy constraint; assuming a codeword of N symbols being transmitted, we require that $x = x_1\, x_2\, \ldots\, x_N$

$$\frac{1}{N}\sum_{i=1}^{N} x_i^2 \leq E$$

where *E* is the signaling energy per symbol.

It can be shown that the capacity of a Gaussian channel with energy constraint *E* and noise variance $N_0/2$ is

$$C = \frac{1}{2}\log\left(1 + \frac{2E}{N_0}\right) \quad \text{bits/channel use}$$

LUND UNIVERSITY

Capacity of band limited AWGN channel

The channel capacity of the bandwidth limited Gaussian channel with two-sided noise spectral density $N_0/2$

$$C_t^W = W \log\left(1 + \frac{P_s}{N_0 W}\right) \text{ bits/s}$$

where W denotes the bandwidth in Hz and P_s is the signaling power in Watts.

Shannon's channel coding theorem

In any system that provides reliable communication over a Gaussian channel the signal-to-noise ratio E_b/N_0 must exceed the Shannon limit, -1.6 dB!

So long as E_b/N_0 > -1.6 dB, Shannon's channel coding theorem guarantees the existence of a system – although it might be very complex – for reliable communication over the channel.

Summary

- Digital channels are characterized by the transition probabilities ($X \rightarrow Y$)
- Typical sequences can help us find out how fast we can communicate on a channel
- Channel capacity is defined as the maximal mutual information between input (X) and output (Y) and it shows how fast we can communicate reliably over the channel
- Capacity of the binary symmetric channel (BSC) is $C_{\mathrm{BSC}} = 1 - h(\varepsilon)$ [bit/channel use]
- Capacity of the additive white Gaussian noise (AWGN) channel is
  - (time-discrete) $C = \frac{1}{2} \log \left(1 + \frac{2E}{N_0}\right)$ [bit/channel use]
  - (continuous band limited) $C_t^W = W \log \left(1 + \frac{P_s}{N_0 W}\right)$ [bit/sec]

LUND UNIVERSITY

LUND
UNIVERSITY