Communication systems

Prof. Marco Moretti

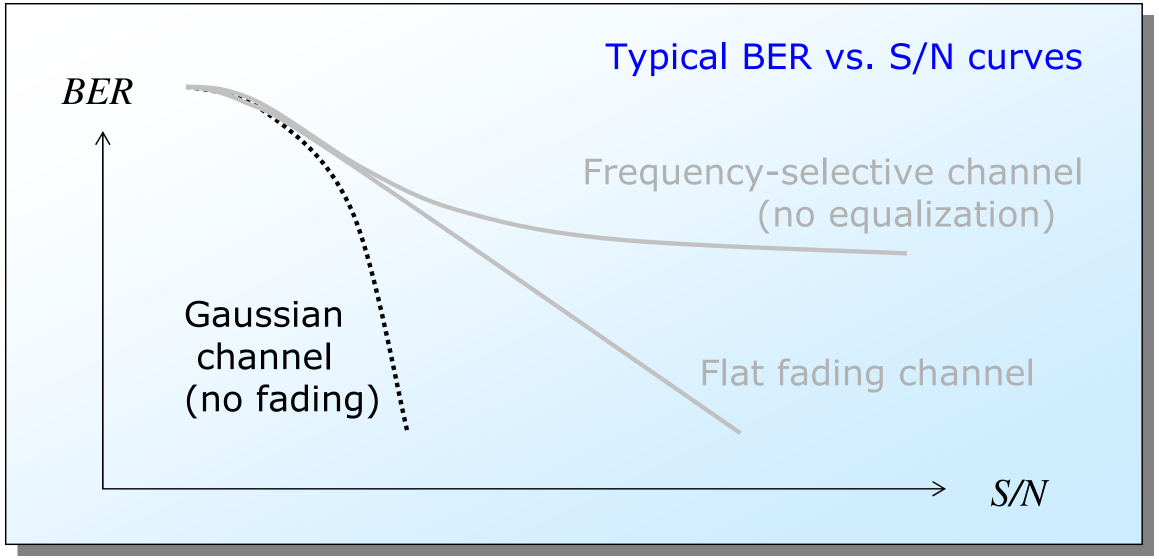
marco.moretti@unipi.it

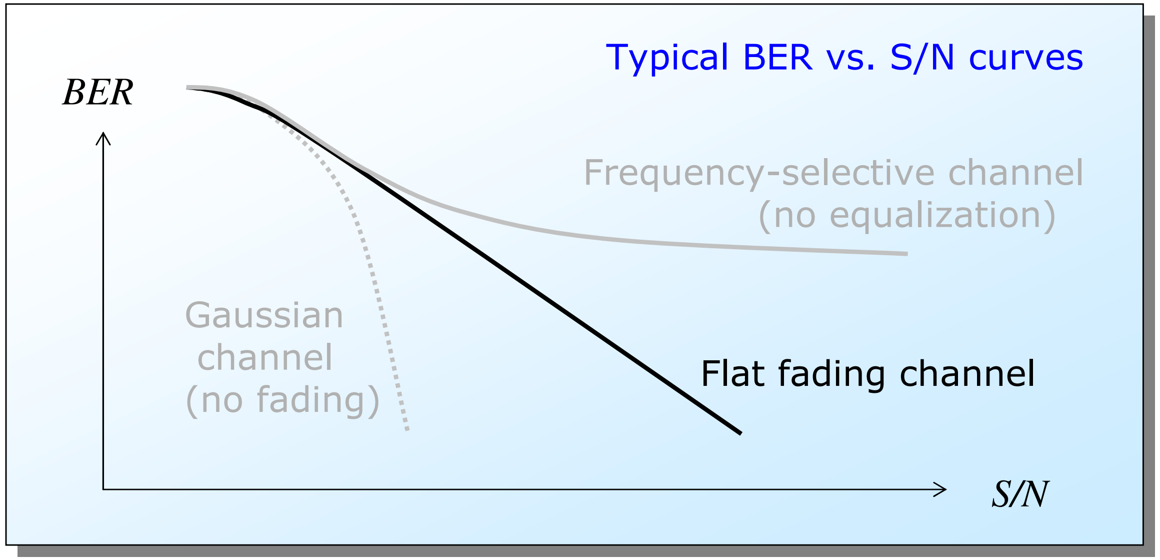
ELECTRONICS AND COMMUNICATIONS SYSTEMS

COMPUTER ENGINEERING

Transmissions over fading channels

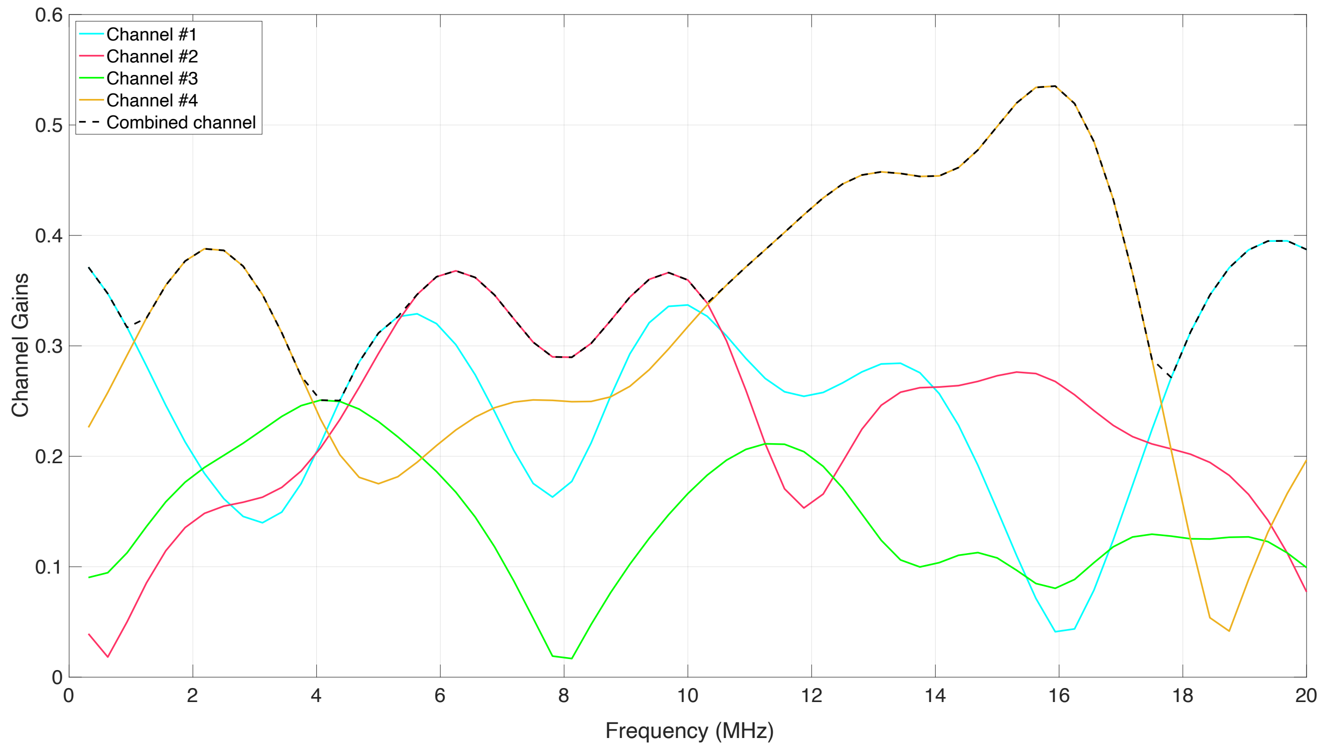
- Signal fading is the main problem in wireless communications.
- OFDM is a technique designed to combat the destructive effects of *multipath* fading.
- *Slow flat* Rayleigh fading is still a big problem.
- One of the most effective resources against the effects of channel fading is *diversity.*





Diversity in wireless communications

- *Diversity* refers to the possibility of improving the reliability of a message by transmitting it over two or more communication channels with different characteristics.
- Diversity is a common technique for combatting fading and co-channel interference and avoiding error bursts.

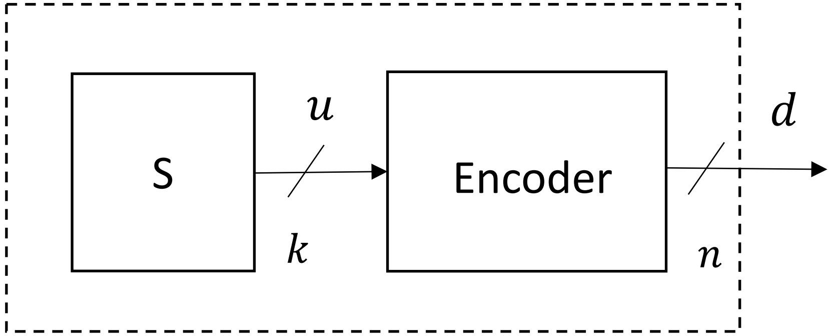


Diversity in wireless communication systems

- *Time diversity*: relates to the *coherence time*
  - Transmission over multiple time slots by channel coding plus interleaving
  - The amount of diversity is small over very slow fading channels
- *Frequency diversity*: relates to the *coherence bandwidth*
  - Transmission over multiple frequency bands
  - The amount of diversity is small over very flat fading channels
- *Spatial diversity*: relates to the *coherence distance*
  - Transmission and reception employing multiple antennas.

Time diversity: interleaving and coding

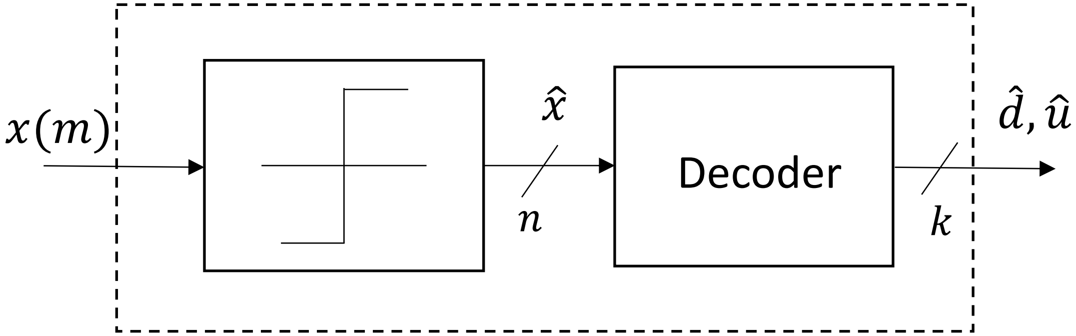
- The main idea was formalized by Claude Shannon in 1948.
- Channel coding introduces some redundancy in the transmitted bits to either
  - Detect errors at the receiver
  - Improve the bit error probability at the receiver.



- The redundancy is measured by the code rate $R = \frac{k}{n} < 1$, the ratio between k, the number of bits at the input of the encoder and n, the number of bits at the output of the encoder.

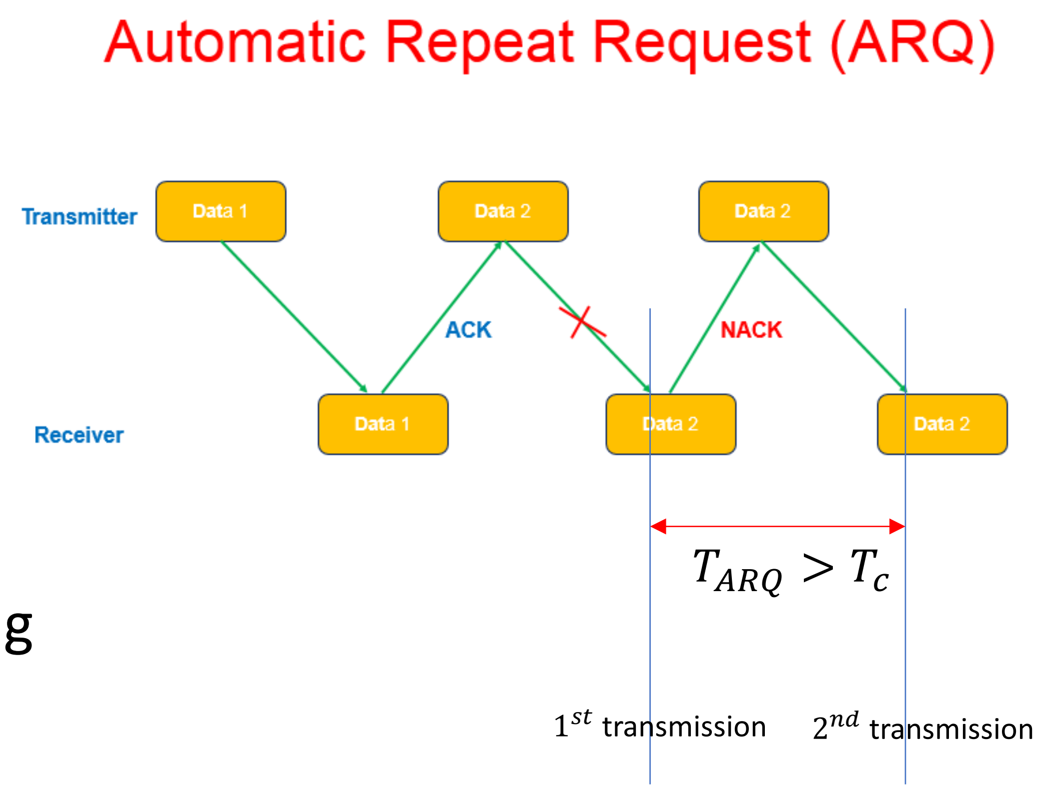
Error detection coding

- Very simple technique: one or more bits of parity are added at the end of a word.
- The receiver computes the parity check applying the same algorithm implemented at the receiver:
  - If the result computed at the receiver matches the parity check bits, the parity bits are discarded and the received bits are considered error-free.
  - If the result does not match the parity check bits, there is one or more errors in the string of received bits and the receiver requests a re-transmission.



Data retransmission

- The receiver feeds back an ACK for a correct reception and a NACK for a faulty reception.
- After receiving a NACK, the transmitter resends the data packet
- ARQ exploits the *time diversity* of the channel by retransmitting the data after T_{ARQ}, a time interval, longer than the channel coherence time T_c.
- The new transmission will experience a different and hopefully better channel
- Advanced receivers are capable of combining the two received messages to further improve the chances of a successful reception.



Error correction coding and channel capacity

- Channel codes can be employed to *correct* the errors introduced by the channel.
- Given a communication channel of bandwidth B, Shannon proved that the *channel capacity* can be computed as

$$C = B\log_2(1 + SNR)\ \mathrm{b/s}$$

  - For any transmission with rate $R < C$ and an arbitrarily small ϵ, it is possible to find an error correction code such that the error probability is $P_e < \epsilon$.
  - On the contrary, if $R > C$ it is not possible to find any code that can make the probability of error of the transmission over the channel arbitrarily small.

Error correction codes

- Linear algebraic codes are the most used type of codes:
  - Block codes
  - Convolutional codes
- All algebraic operations are performed in the GF(2), the Galois field of two elements $\{0,1\}$.

XOR

+	0	1
0	0	1
1	1	0

AND

x	0	1
0	0	0
1	0	1

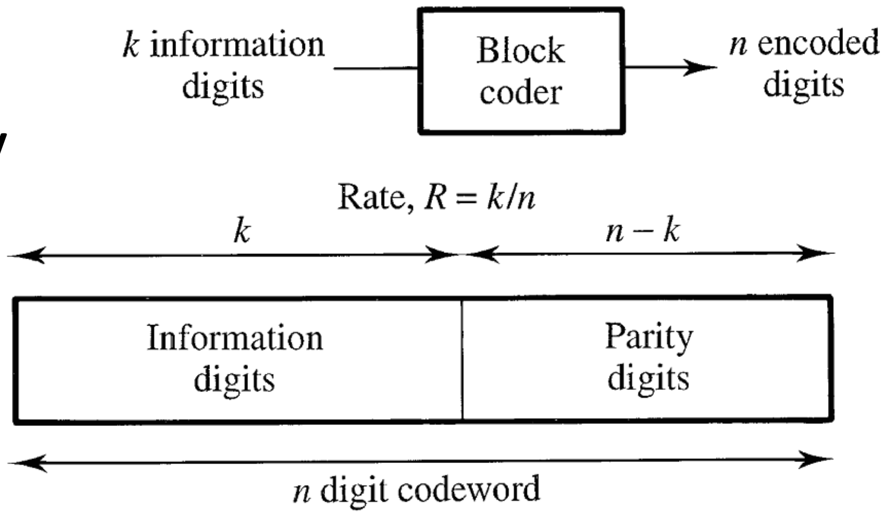
- They have been initially studied for AWGN transmissions but can be employed over fading channels.

Block codes: encoder

- Block codes are most of the times in *systematic* form: the coded word is formed by k information bits and $n-k$ parity bits.
- The encoder can be represented as the code *generator matrix* $\boldsymbol{G}$ the encoder.
- The word $\boldsymbol{u}$ of k bits is encoded in the coded word $\boldsymbol{d}$ of n bits

$$\boldsymbol{d} = \boldsymbol{u}\boldsymbol{G}$$

- The encoder add redundancy so that all coded words differ of as many bits as possible.



Block codes: repetition code

- The simplest block code is the $k = 1, n = 3$ *repetition code.*
- It is composed by $2^k = 2$ codewords:

$$u = 0 \Longrightarrow \mathrm{d} = [0\ 0\ 0], \qquad u = 1 \Longrightarrow \mathrm{d} = [1\ 1\ 1]$$

- The generator matrix is

$$\boldsymbol{G} = [1\ 1\ 1]$$

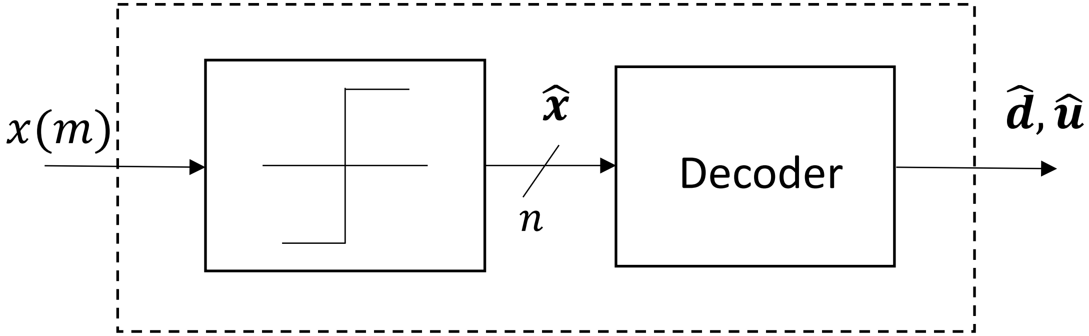
- The decoder takes a majority decision
- $\hat{x} = \left.\begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{matrix}\right\} \rightarrow \hat{d} = [0\ 0\ 0] \rightarrow \hat{u} = 0, \hat{x} = \left.\begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{matrix}\right\} \rightarrow \hat{d} = [1\ 1\ 1] \rightarrow \hat{u} = 1$
-

Block codes: decoder

- After having received the string of n bits $\widehat{\boldsymbol{x}}$, the decoder selects the codeword $\widehat{\boldsymbol{d}}$, as the one that has minimal distance from $\widehat{\boldsymbol{x}}$

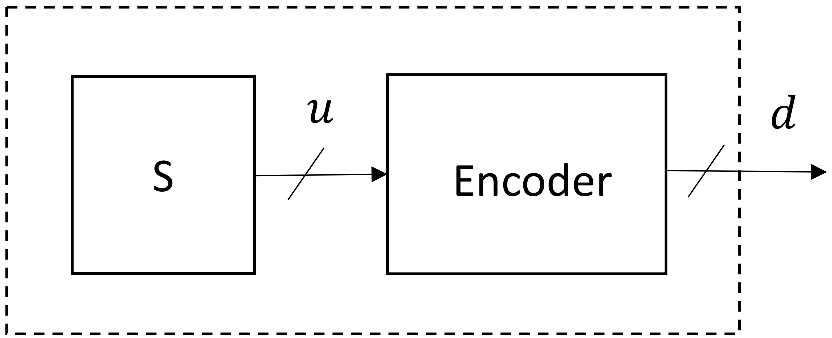
$$\widehat{\boldsymbol{d}} = \arg\min d(\boldsymbol{d}, \widehat{\boldsymbol{x}})$$

- The distance is computed as the number of bits that are different in the two string of bits.
- An error event occurs when, due to the noise, the received vector $\widehat{\boldsymbol{x}}$ is closer to a codeword different from the transmitted one.
- Codes where the distance between words is large are more robust against noise and fading than codes where the distance is small.

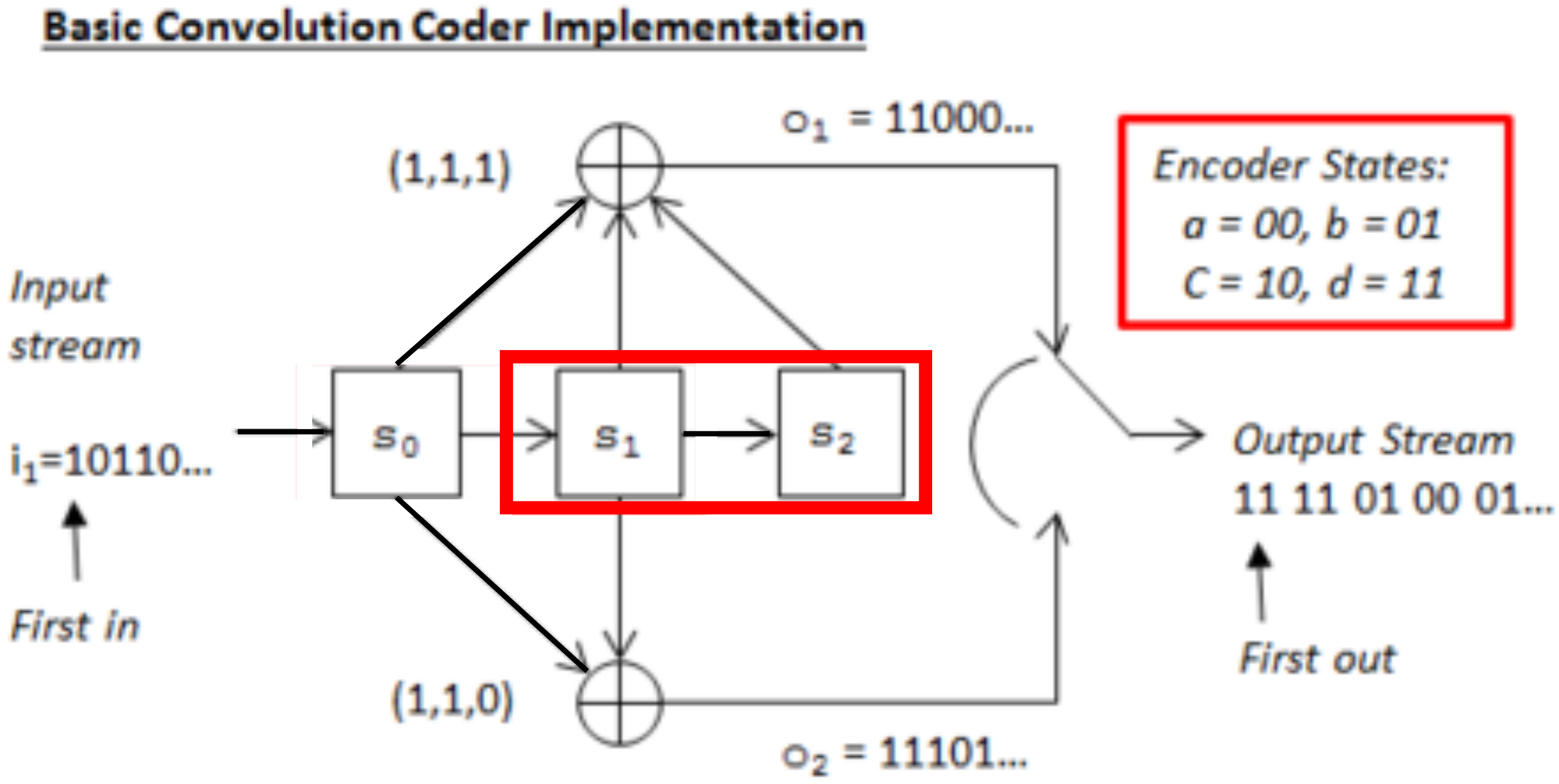


Convolutional codes: encoder

- The encoder of a k, n convolutional code works as n linear filters in GF(2). Each of the n outputs of the encoder is a linear combination of the input
- The filter impulse response for the j-th output bit, the j-th code generator, is a series of 0 and 1.

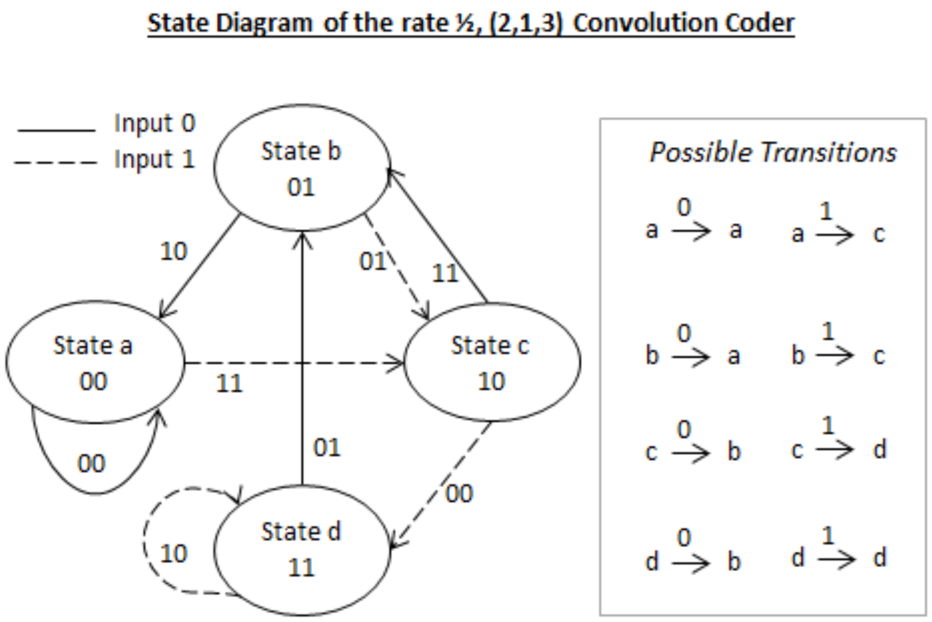


Convolutional codes: encoder



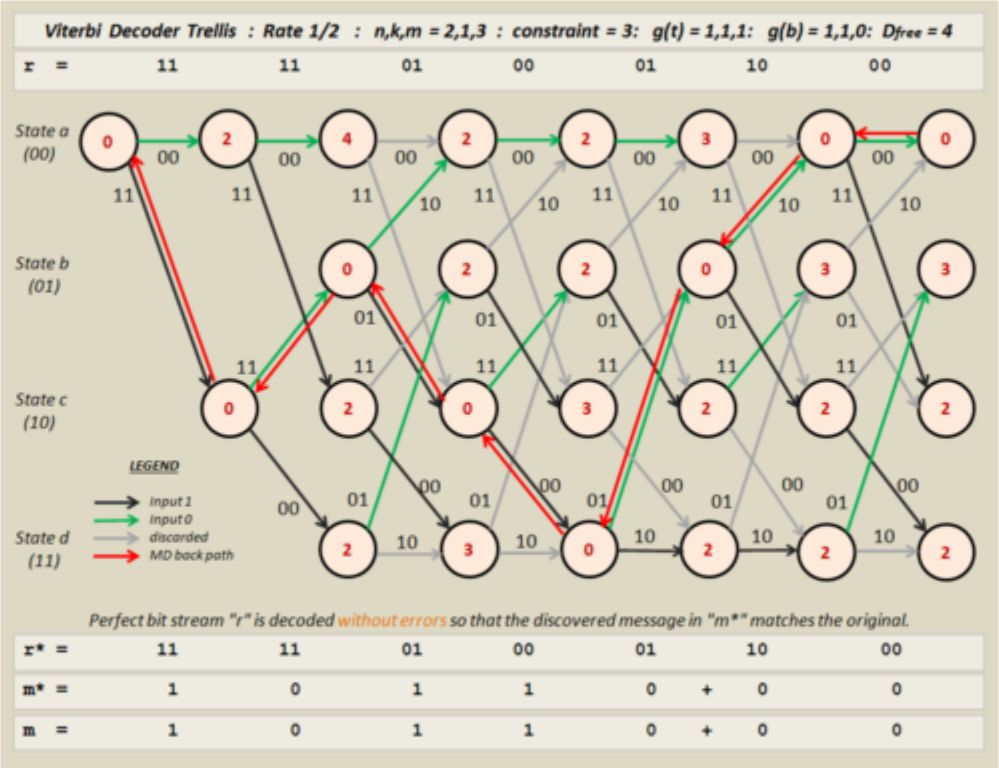
Coder Rate ½, and configuration (n,k,m) = (2,1,3) where:
n: *is the number of output bits* ***k:*** *is the number of input bits*
m: *is the number of shift register stages*

Convolutional codes: state diagram



- The encoder can be represented as a finite-state machine.
- The output of the encoder depends on two elements:
  - The input bit;
  - The state of the encoder: the content of the memory cells of the shift register.
- Each incoming bit determines
  - A new output sequence;
  - A new state
- The state diagram captures the transitions in the encoder.

Convolutional codes: trellis diagram



- The time evolution of the encoder is captured by a *trellis* diagram.
- At each signaling time, the trellis represents all achievable states.
  - At time $t = 0$, the encoder starts from state 00
  - At time $t = 1$ the encoder can be either in state 00 or state 10
  - At time $t = 2$ the encoder can be in any of the four states.
  - At any time $t > 2$ the trellis has four states and for each state there are two paths in and two paths out.
- Any given input sequence and the corresponding encoded word can be represented as a path on the trellis.

Convolutional codes: decoder

- Unlike in block codes, the coded bits are not organized in blocks but they are a continuous flow of data.
- A transmission of N codewords implies that the sequence $\boldsymbol{u}$ of kN bits has been encoded into a sequence $\boldsymbol{d}$ of nN bits.
- By using L shift registers at the encoder, each input bit impacts on Ln coded bits so that the whole sequence needs to be *jointly* decoded.
- The decoder's task is to select among all the possible 2^{kN} possible convolutionally encoded sequences the one that minimize the distance from the received sequence $\widehat{\boldsymbol{x}}$ of nN bits

$$\widehat{\boldsymbol{d}} = \arg\min d(\boldsymbol{d}, \widehat{\boldsymbol{x}})$$

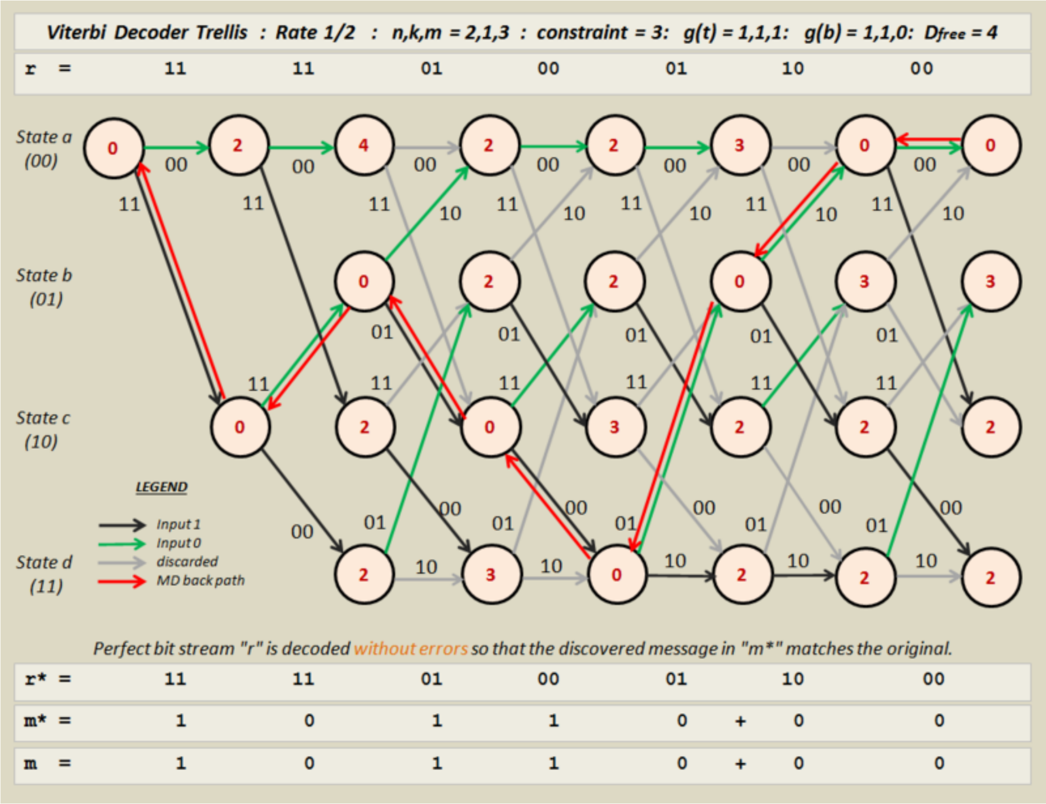
Convolutional codes: the Viterbi algorithm

- Until the discovery of the Viterbi algorithm in 1967, the use of powerful convolutional codes was limited by the exponential complexity of the decoder.
- The Viterbi algorithm is an iterative algorithm that scales down the complexity from exponential to linear in N.
- The main idea is that of all the 2^{kN} possible paths on the trellis, a very large number of them can be discarded because non relevant.
- At each step of the algorithm, the algorithm selects a number of surviving paths on the trellis equal to the number of states of the encoder.

Andrew James Viterbi
(born Andrea Giacomo Viterbi in Bergamo, Italy)
is Professor of Electrical Engineering at the University of Southern California's Viterbi School of Engineering, which was named after him in recognition of his **$52 million** gift.

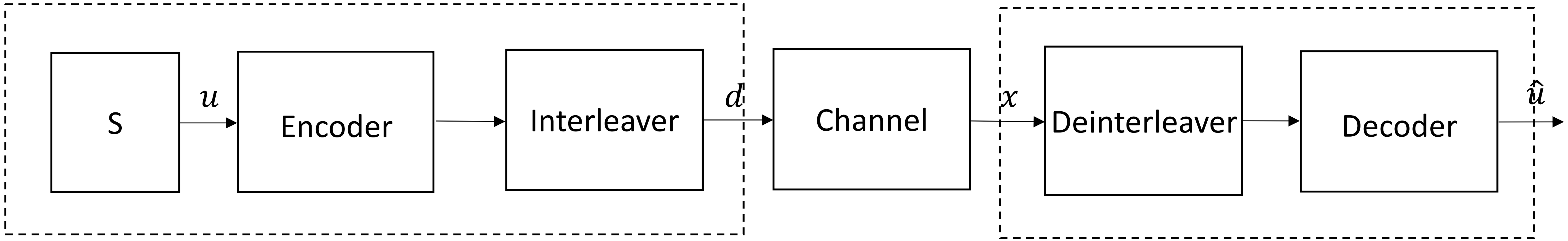
Convolutional codes: the Viterbi algorithm

- The algorithm's objective is to find the path on the trellis that minimizes the distance from the received sequence of bits.
- Starting and finishing in a known state, the algorithm:
  1. Computes all the possible state transitions on the trellis and for each transition the distance from the corresponding received sequence of n bits (*branch metrics*);
  2. Assuming that there is a single path arriving at each state, computes the accumulated distance for each branch out of a given state. The distance is the sum of the cumulated metric of the path arriving at the state and the branch metric;
  3. Discard all the paths leading to one state except the one with minimum accumulated distance (*survivor path*).
- The survivor path leading to the last state is algorithm's output and the final cumulated metric is the number of corrected errors.



Interleaving

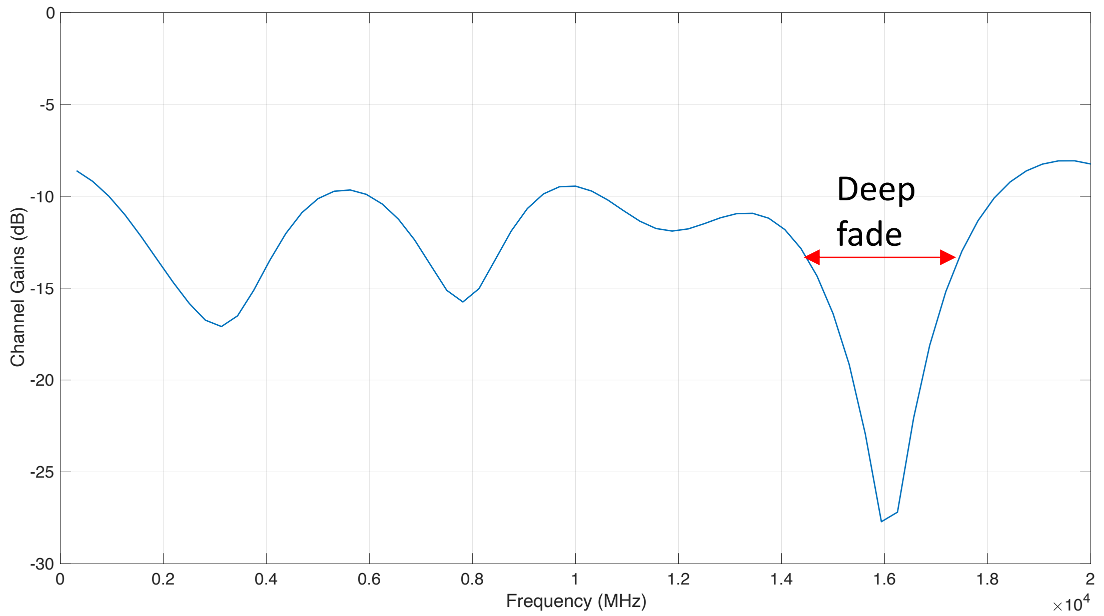
- Convolutional codes are mostly suitable for memoryless channels with random error events.
- Error correcting codes perform well when the errors are uniformly distributed and uncorrelated.
- Fading channels tend to cause bursty errors: when a channel is in a deep fade, there is a statistical dependence among successive error events.
- *Interleaving* makes the channel looks like as a memoryless channel at the decoder and tends to *decorrelate* error events.



Interleaving …

- Interleaving is achieved by spreading the coded symbols in time or frequency before transmission.
- The reverse is done at the receiver by deinterleaving the received sequence.
- Interleaving makes bursty errors look like random, so that convolutional codes perform best.
- The price to pay with interleaving is the large *latency*: both at transmitter and at the receiver it is necessary to have the entire block of data to start the encoding/decoding process.
- There is a trade-off: the larger the interleaver depth K, the more decorrelated are the errors but also the longer is the latency and delay.
- Types of interleaving:
  - Block interleaving
  - Convolutional or cross interleaving

Block interleaver example



$\{A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P\}$

A	B	C	D
E	F	G	H
I	J	K	L
M	N	O	P

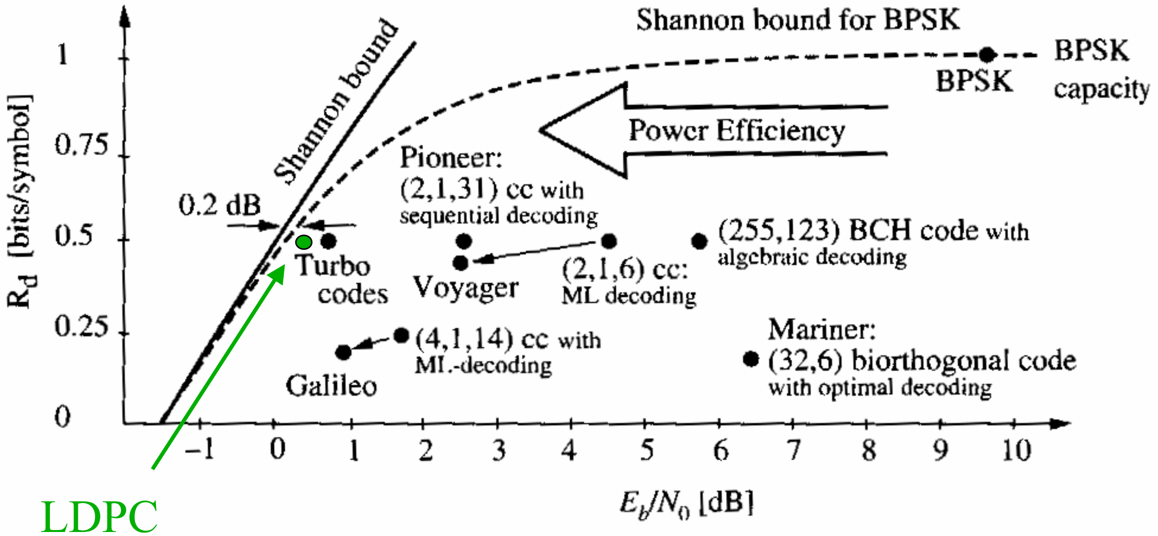
A	E	I	M
B	F	J	N
C	G	K	O
D	H	L	P

$\{A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P\}$ → Interleaver → $\{A, E, I, M, B, F, J, N, C, G, K, O, D, H, L, P\}$

4 positions of distance between two consecutively encoded bits

Turbo codes and LDPC

- In the channel capacity theorem, Shannon uses infinitely long codes, i.e. the code rate $R = k/n$ is fixed but $k \to \infty$ and $n \to \infty$.
- In practical system, the length of the code is limited because of the complexity of decoding: the performance of physical systems are far from Shannon theoretical limits.
- Around the turning of the century there have been two major breakthroughs:
  - Turbo codes (1993)
  - Low-density parity check codes (LDPC, 1999).

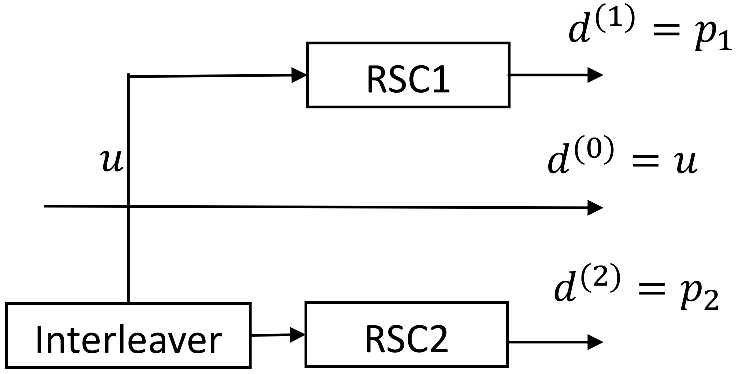


LDPC codes

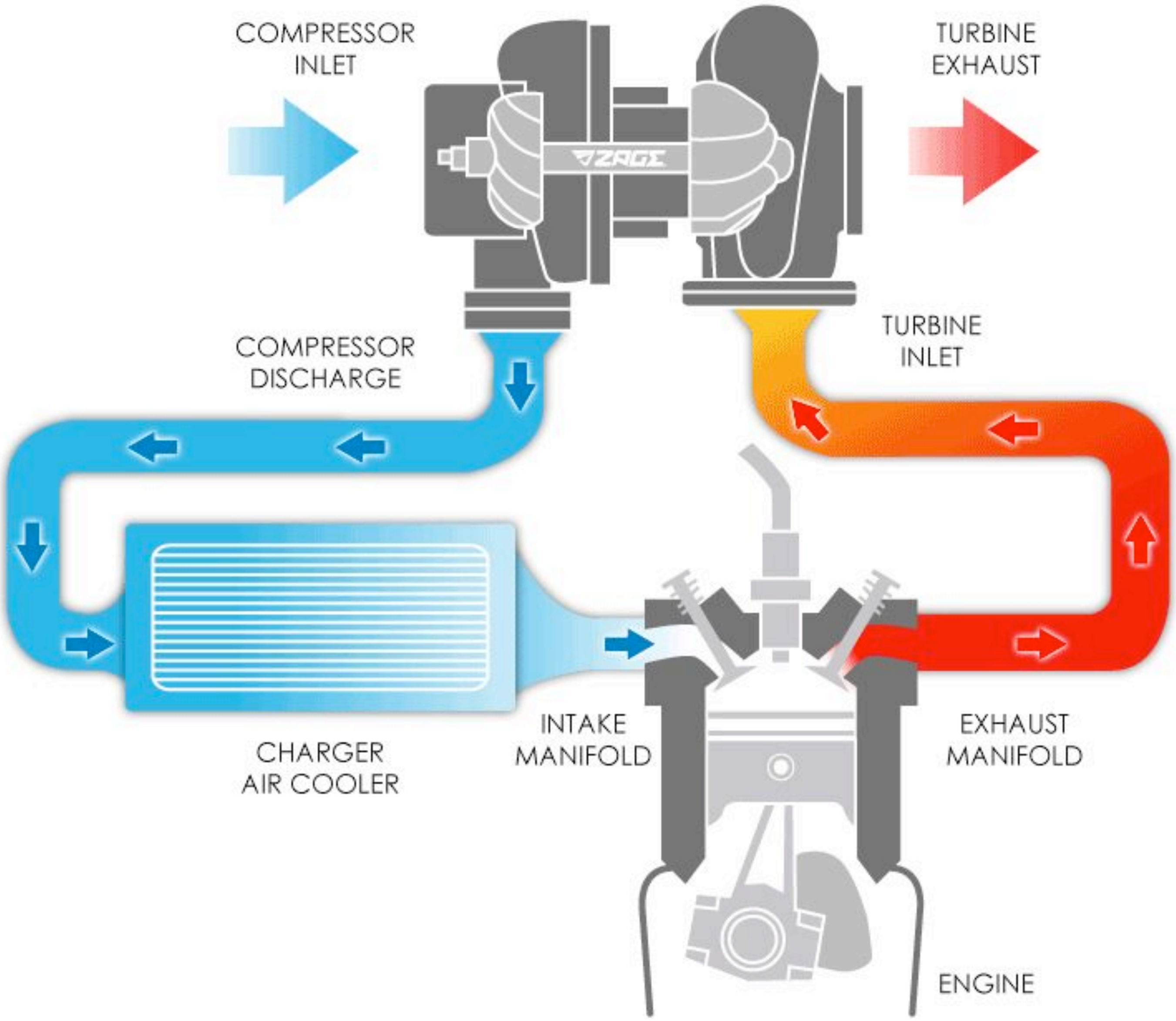
from Trellis and Turbo Coding, Schlegel and Perez, IEEE Press, 2004

Turbo codes

- The combination of coding and interleaving can be exploited to boost the performance of convolutional codes in general.
- *Turbo codes* are a particular example of concatenated codes, where two encoders are in parallel and the input data of the second encoder are first interleaved.
- Thanks to the interleaver, the decoder will have two independent replicas of the same data and can use both streams to decode the information sequence.
- Increase the redundancy of the transmitted information but also the *diversity* of the system.



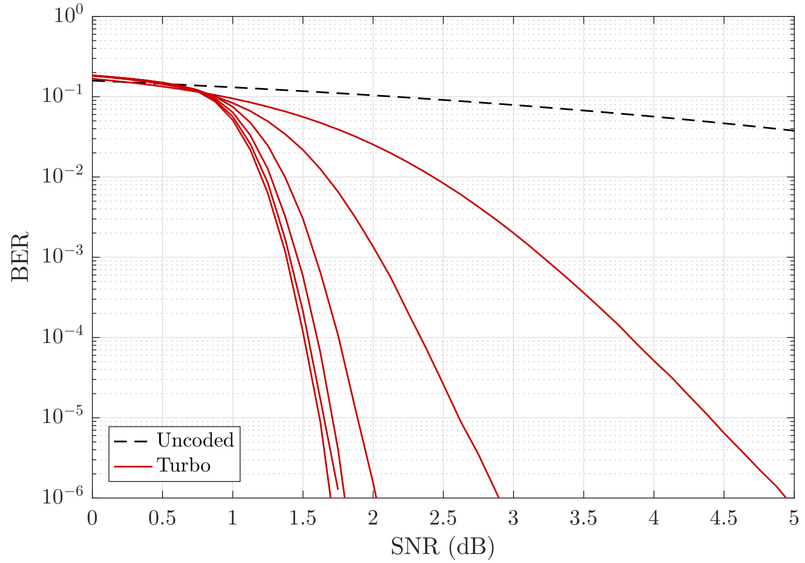
Turbo.....



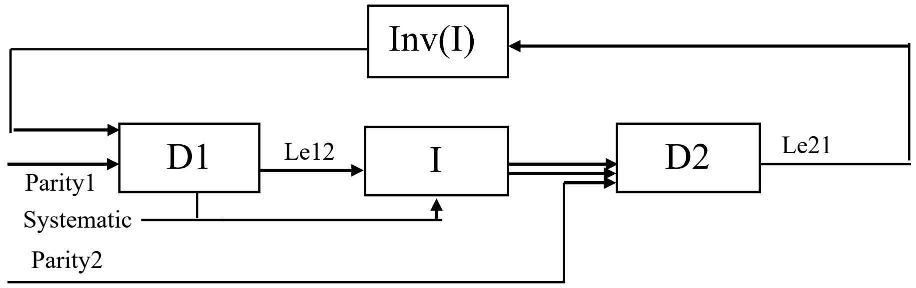
COMPRESSOR INLET
TURBINE EXHAUST
ZAGE
COMPRESSOR DISCHARGE
TURBINE INLET
CHARGER AIR COOLER
INTAKE MANIFOLD
EXHAUST MANIFOLD
ENGINE

Turbo decoding

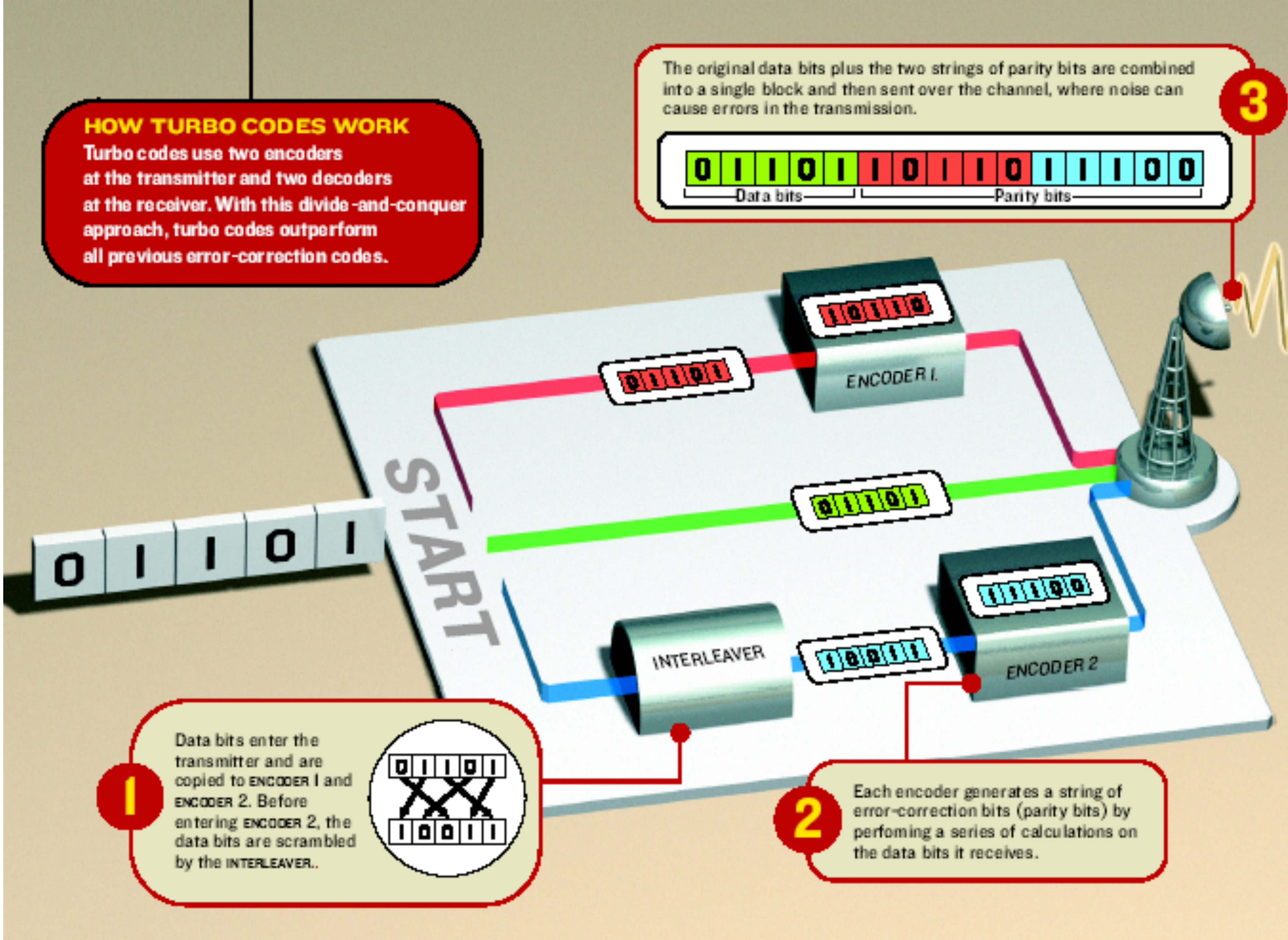
- The turbo decoder employs the cascade of two decoders, with the output of one decoder used as prior information to the next.
- Feedback in decoding circuit allows for multiple iterations and improves bit error performance.
- The process is iterative and at each step the combination of the two decoders corrects some of the errors.



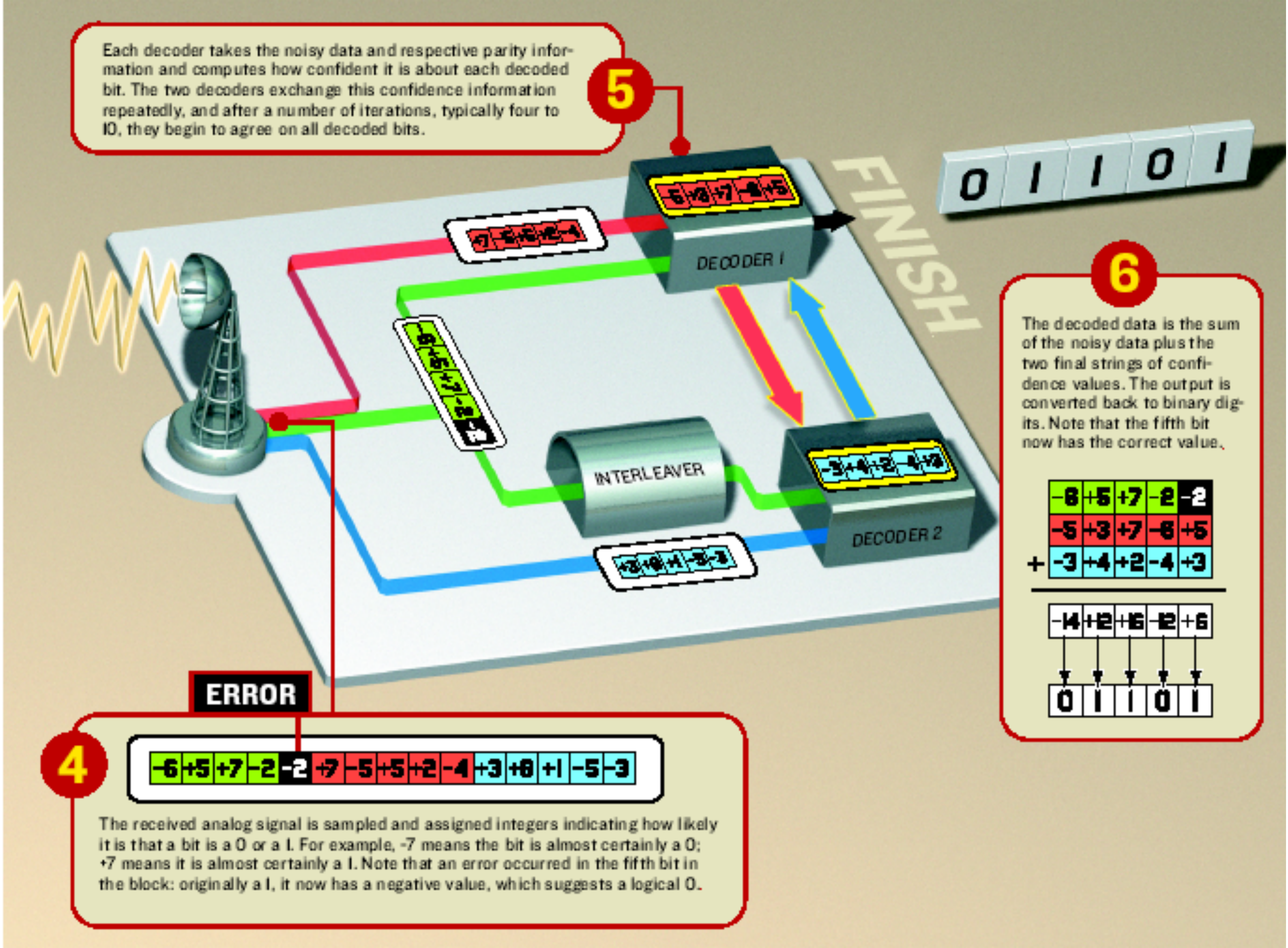
Turbo code convergence, $K = 2048$, $R = 1/2$, Iterations = 32, 16, 8, 4, 2, 1 from left to right.



Turbo codes (© IEEE spectrum)



Turbo codes (© IEEE spectrum)



Turbo codes: latency

- Convolutional codes in general and turbo codes in particular have a problem with *latency* due to the presence of the interleaver and the iterative decoding process.
- At any given SNR there is a tradeoff between latency due to interleaver and QOS
  - Small block sizes ($K \sim 300$ bits) can be used for real time voice (medium-high BER can be tolerated).
  - Mid range block sizes ($K \sim 4000$ bits) used for video play back (low BER).
  - Large block sizes ($K \sim 16000$ bits), useful for file transfer (very low BER).

Spatial diversity

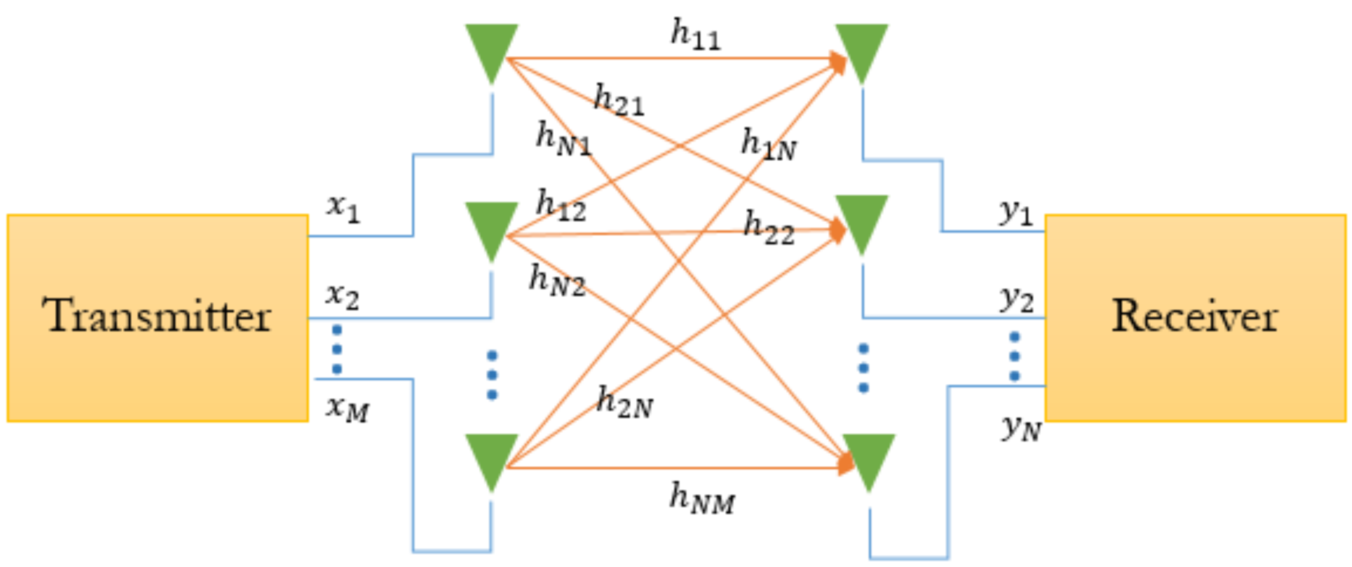
Receive diversity

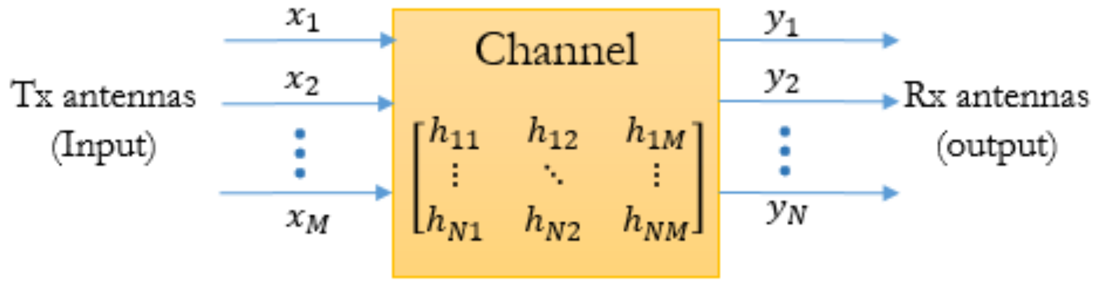
- Among the many ways to obtain diversity:
  - *Frequency* and *time* diversity require expensive resources (bandwidth or time) and do not provide array gain.
  - *Space* diversity by means of multiple antennas does not sacrifice bandwidth or time and may provide array gain.
- *Array gain*: is the power *gain* achieved by using multiple antennas with respect to the single antenna case. The more correlated is the spatial channel the higher the potential array gain.
- *Diversity gain*: is the power gain due to the exploitation of the diversity of the spatial channel. It is maximum when the spatial channel is uncorrelated.

MIMO channel

- Multiple-input multiple-output systems are systems where where both the transmitter and the receiver are equipped with several antennas.
- *Narrowband* assumption: the channel linking the n-th receive antenna with the m-th transmit antenna is the scalar $h_{n,m}$

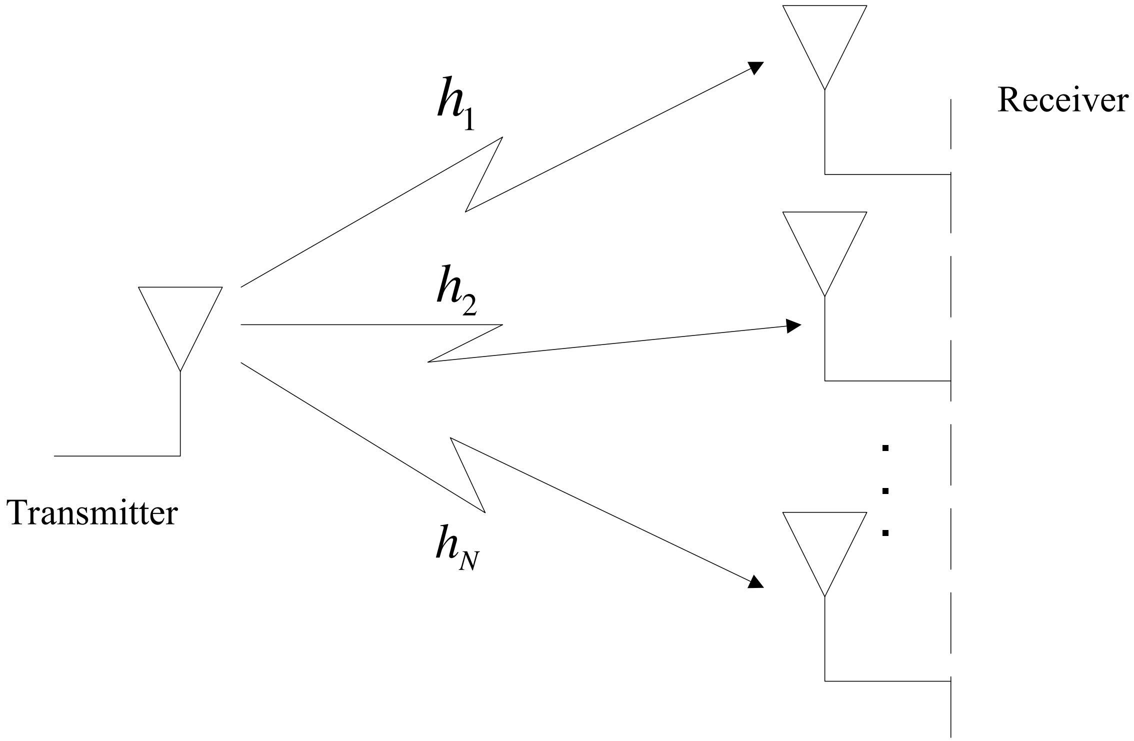
Multiple Input Multiple Output (MIMO) System





MIMO from channel perspective

SIMO channel: receive diversity



- $N > 1$ antennas at the receiver, $M = 1$ at the transmitter.
- The decision variable at the i-th receive antenna is
$$x_i(m) = h_i c_m + n_i(m)$$
- The signals received at the N antennas are combined together and the decision variable is
$$z(m) = w_1 x_1(m) + \cdots + w_N \, x_N(m)$$

Maximal ratio combining (MRC)

- The decision variable is

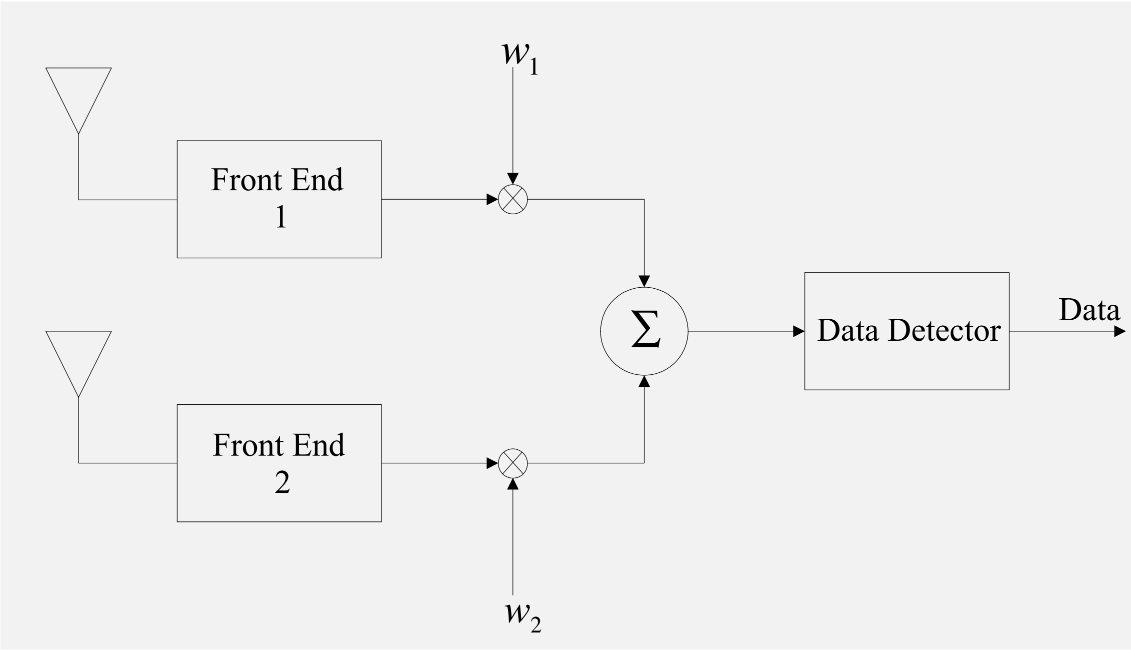
$$z(m) = \sum_{i=1}^{N} w_i h_i \, c_m + \sum_{i=1}^{N} w_i n_i(m)$$

- For the i-th antenna the optimal weight is

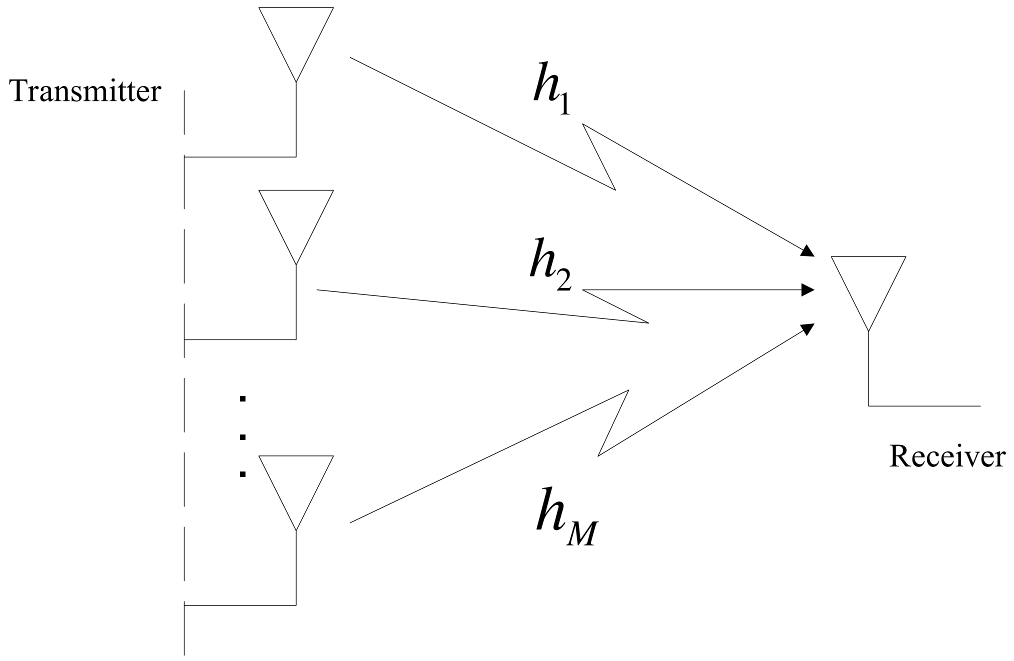
$$w_i = h_i^*$$

- In the 1x2 case, the signal-to-noise ratio is

$$SNR = (|h_1|^2 + |h_2|^2) \frac{A}{\sigma^2}$$



MISO channel: transmit diversity

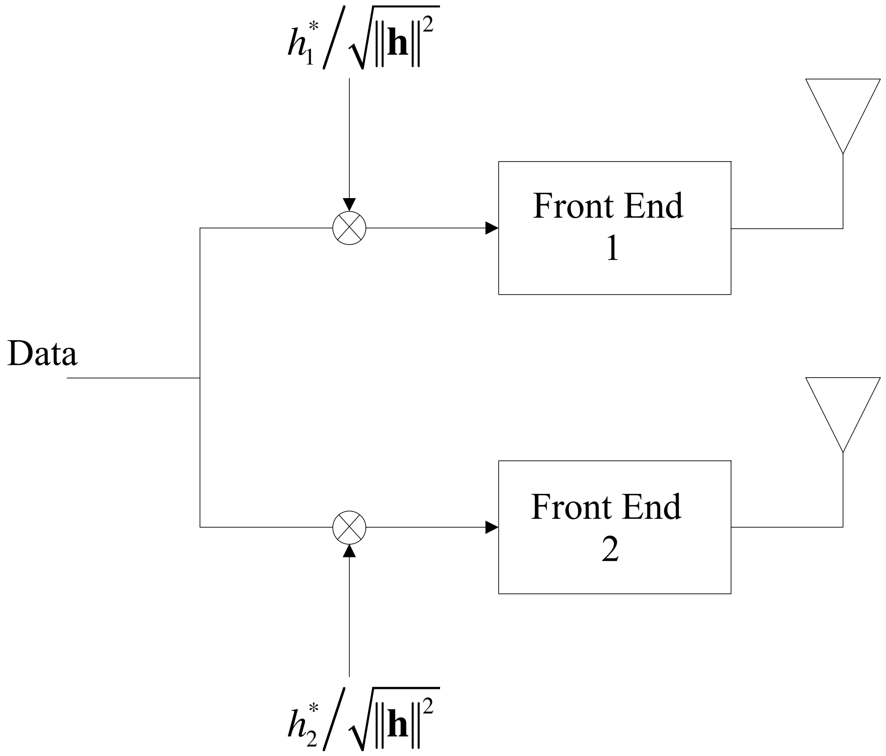


- $M > 1$ antennas at the transmitter and $N = 1$ at the receiver
- Spatial precoding: the signals are precoded at the transmitter.
- The signal at the j-th transmit antenna is $y_j(m) = w_j c_m$
- The transmitted energy depends also on the precoding weights.
- The received signal is

$$x(m) = h_1 y_1 + \cdots + h_M y_M$$

Maximal ratio transmit (MRT) combining

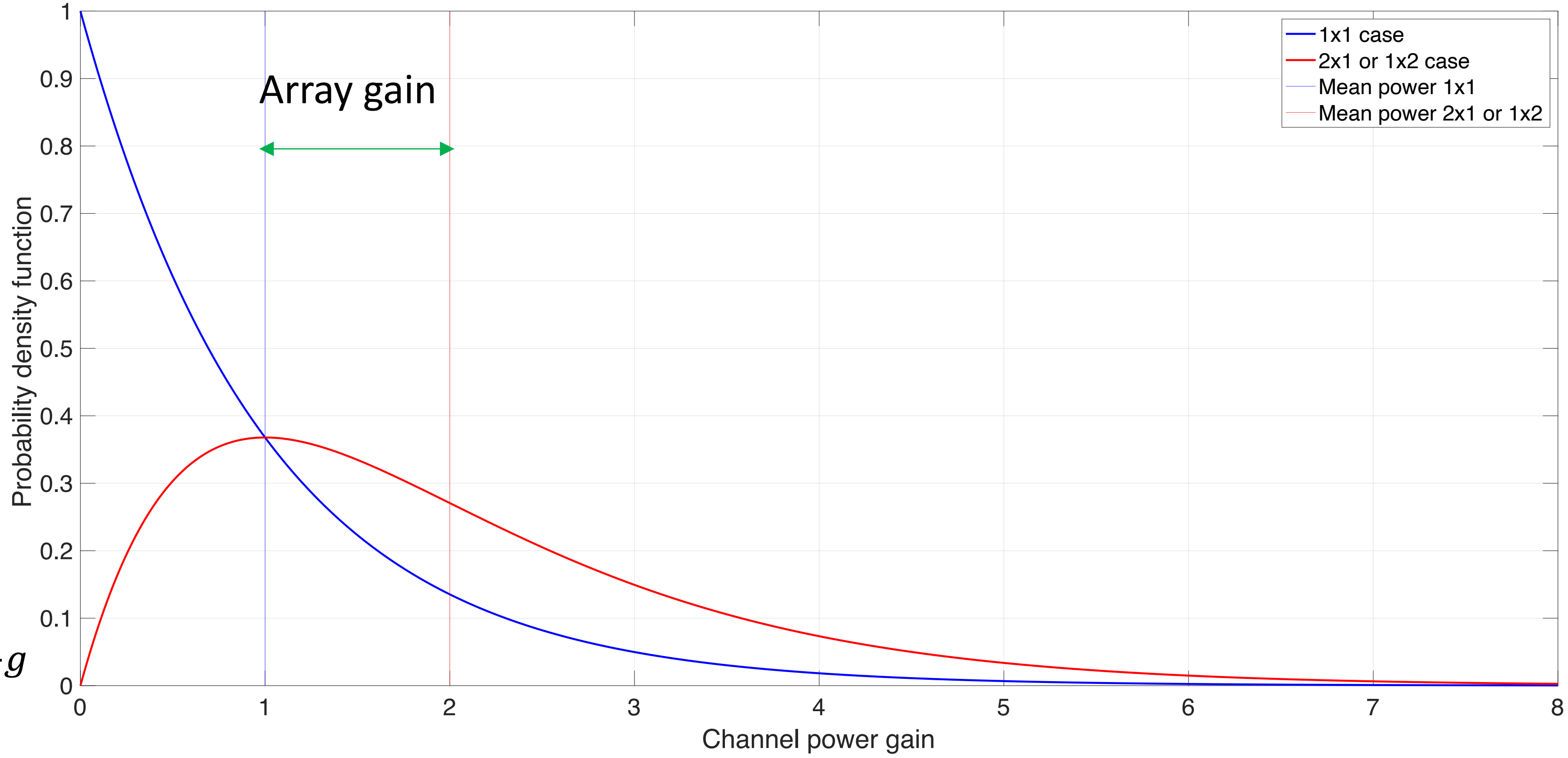
- The optimal precoding weight for the j-th transmit antenna is
$$w_j = h_j / \|\mathbf{h}\|$$
so that the overall transmitted energy per symbol is still $E_s = A/2$.
- At the receiver
$$x(m) = \left(\frac{h_1 h_1^*}{\|\mathbf{h}\|} + \frac{h_2 h_2^*}{\|\mathbf{h}\|} \right) c_m + n(m)$$
- The signal to noise is
$$SNR = (|h_1|^2 + |h_2|^2) \frac{A}{\sigma^2}$$



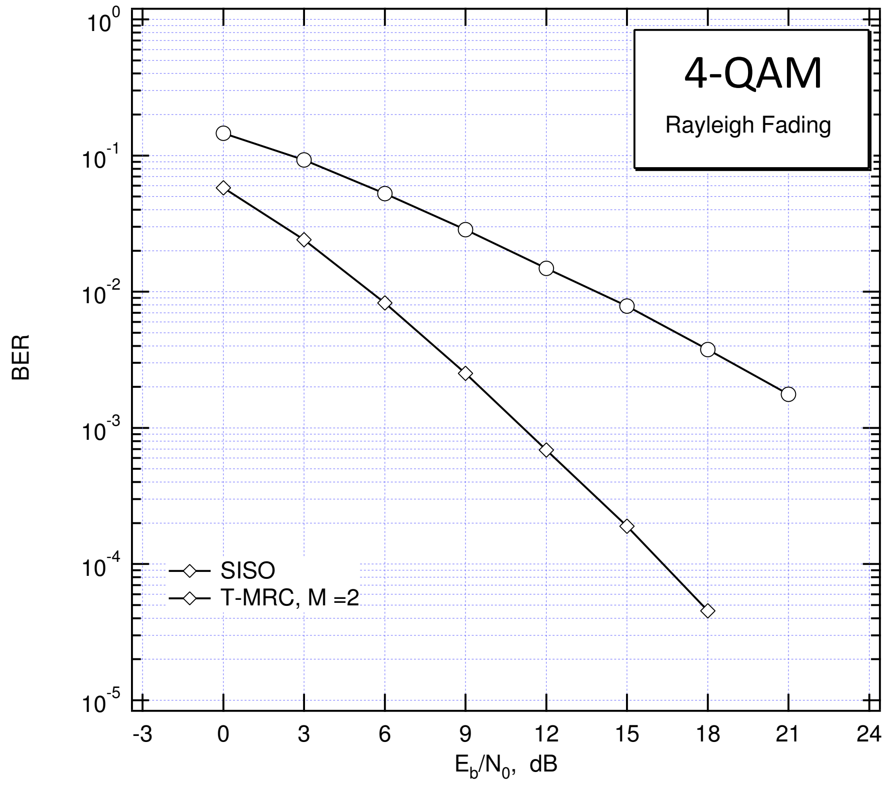
Channel gain $||h||^2$ distribution for D antennas

- With MRT or MRC, the SNR is proportional to the channel power gain $||h||^2$.
- The pdf of $||h||^2$ depends on D, the number of antennas

$$f_{||h||^2}(g) = \frac{1}{(D-1)!} g^{D-1} e^{-g}$$



Maximal ratio combining



- Pros
  - Diversity gain
  - Array gain
  - MRT: No additional processing at the receiver
- Cons
  - MRT: Requires channel knowledge at the transmitter
  - MRC: requires some extra processing at the receiver.

MIMO: spatial multiplexing

- The channel is a (N, M)-dimensional matrix
- The optimal technique is called *spatial multiplexing* based on the *singular value decomposition* (SVD) of the channel matrix $\boldsymbol{H}$.
- By employing SVD and coordinating the precoding weights at the transmitter with the combiner weights at the receiver, it is possible to create a certain number of independent orthogonal channels.
- Assuming that $M = N$, spatial multiplexing creates N *independent spatial channels*.

LTE physical layer

LTE

- Long Term Evolution (LTE) is the 4th wireless communication standard
  - 1G: various national analog FDMA-based systems
  - 2G: GSM. First digital standard. Maximum data rate supported $R = 9.6$ kb/s
  - 3G: UMTS (CDMA)
  - 4G: LTE (OFDM-based as WiFi)
  - 5G: NR (currently deployed now)
- LTE is a very flexible communication system mainly used for data transmissions

Pyhysical layer LTE numerology

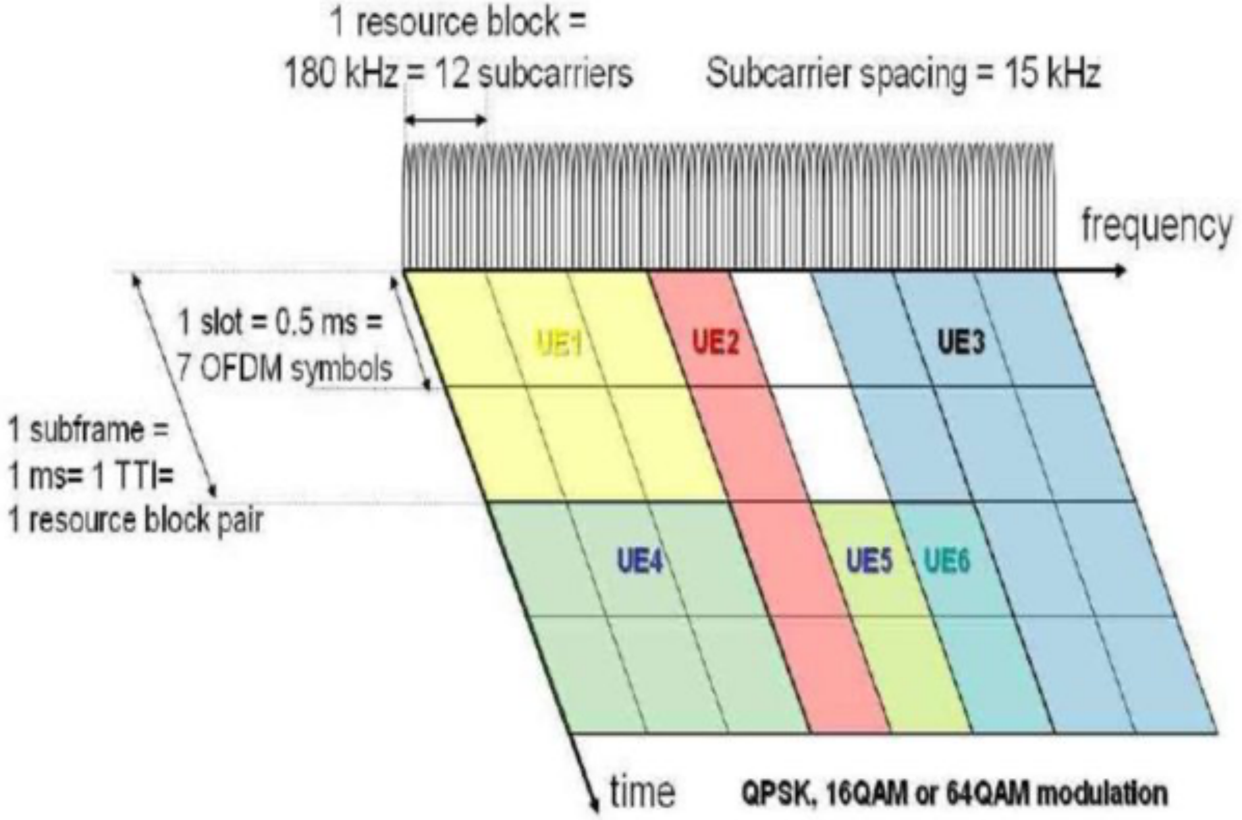
- LTE is based on OFDM
- The sampling time is

$$f_s = 30.72 \text{ MHz}$$

- FFT size is $N = 2048$
- Subcarrier's bandwidth is

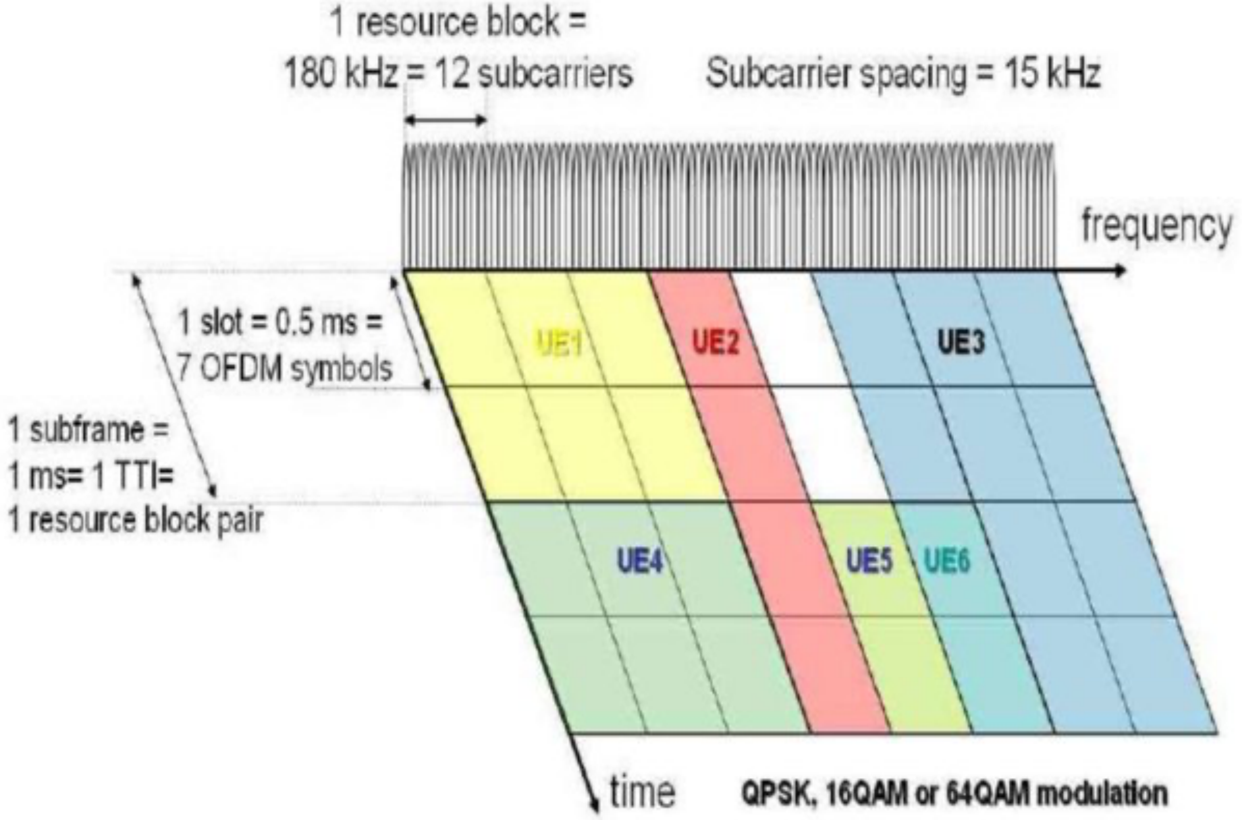
$$\Delta f = \frac{30.72}{2048} \mathrm{MHz} = 15 \text{ kHz}$$

- A single *slot* groups 7 OFDM blocks and has a duration of 0.5 ms



Pyhysical layer LTE numerology

- The multiple access technique is OFDMA: groups of subcarriers are allocated to the user.
- The minimum allocation unit is a resource block (RB). A RB contains 12 subcarriers and spans 180 kHz for the duration of a slot.



Pyhysical layer LTE: peak data rate

- Recent high-end mobile phones fall in the 19-20 category list.
- Raw Bandwidth:

  Of the 2048 available subcarriers only 1200 (100 RB) are used the remaining 848 are 0-level guard subcarrier. The maximum available raw bandwidth for data transmission is

$$B_{raw} = 1200x15 \text{ kHz} = 18 \text{ MHz}$$

- Available bandwidth

  Approximately 4 MHz are spent for sending *control* and *synchronization information*

$$B_{av} = B_{raw} - 4 \text{ MHz} = 14 \text{ MHz}$$

Pyhysical layer LTE: peak data rate

- Modulation and coding

In LTE *coding rate* ranges in the interval 0.0762 - 0.9258. The modulation used are 4-QAM to 256-QAM. Modulation order and coding rate are based on the radio link quality.

Radio link quality is estimated based on CQI (Channel Quality Indicator), which is measured by the terminal and fed back to the base station. The highest coding rate (including CRC) is $R = 0.9258$ and the largest modulation order is $M = 256$ (8 bits per symbol).

The maximum bit rate per SISO channel is is

$$R_{1x1} = 0.9258x8x14 \text{ MHz} \approx 100 \text{ Mb/s}$$

Pyhysical layer LTE: peak data rate

- MIMO

  In LTE various configurations are supported (1x1,2x1,1x2,2x2,4x4) depending on
    - user equipment capabilities (number of antennas);
    - eNode-B capabilities: cell traffic, hardware limits;
    - channel conditions.
  - Most performing scheme is spatial multiplexing 4x4 , which creates 4 parallel channels. With 4x4 MIMO spatial multiplexing the peak data rate is

$$R_{4x4} = 4xR_{1x1} = 400 \text{ Mb/s}$$

- Carrier aggregation
  - Most advanced terminals can aggregate several 20 MHz bandwidth together
  - Iphone 11 pro: 3 with $4x4$ MIMO and 2 with 2x2 MIMO

$$R_I = 3x400 + 2x200 = 1.6 \text{ Gb/s}$$

  - Samsung S20+

$$R_s = 5x400 = 2 \text{ Gb/s}$$