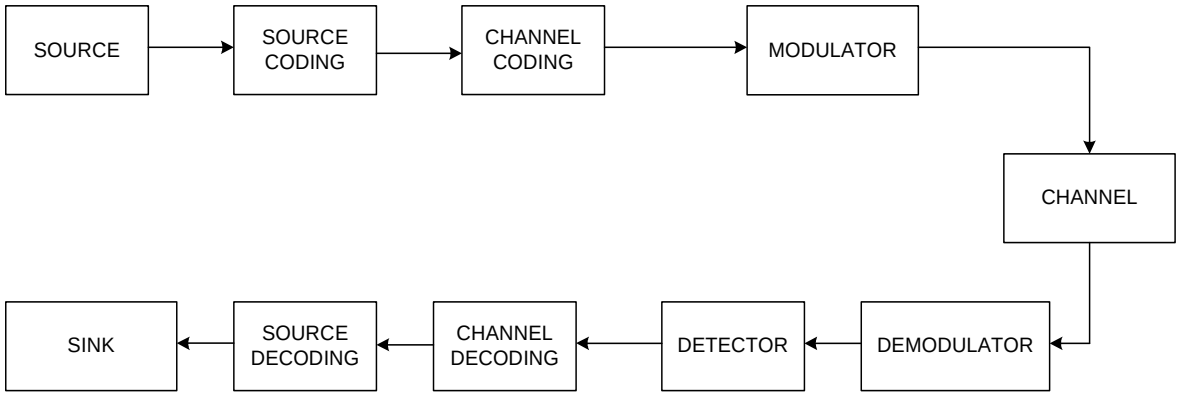
REVIEW OF ERROR CORRECTING CODES

Didier Le Ruyet†

† Electronique et Communications, CNAM, 292 rue Saint Martin, 75141 Paris Cedex 3, France

Email: leruyet@cnam.fr

THE SHANNON PARADIGM



- The source encoder try to eliminate the redundancy available in the source.
- The aim of the channel encoder is to protect the message against the channel perturbations by adding redundancy to the compressed message
- The modulator performs a mapping into the euclidean space.

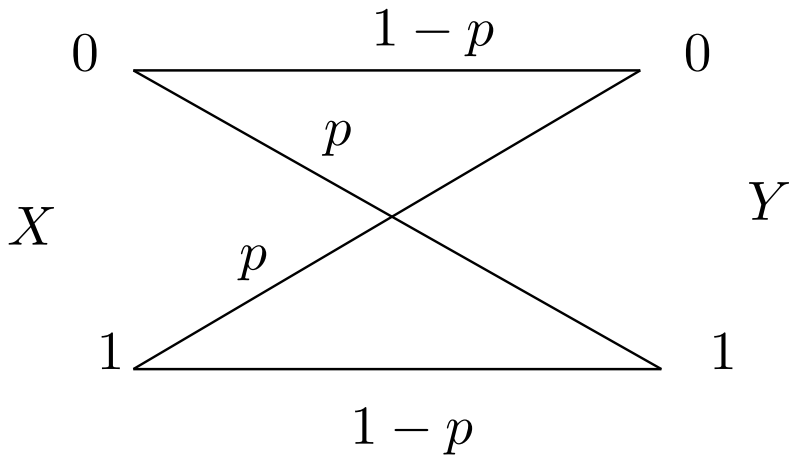
Some Citations

- All codes are good, except for the ones we can think of.

- Never discard information prematurely that may be useful in making a decision until all decisions related to that information have been completed. (Andrew Viterbi)

- It is a capital mistake to theorize before you have all the evidence. It biases the judgement. (Sir Arthur Conan Doyle)

BINARY SYMETRIC CHANNEL



- **This memoryless channel is defined by the transition probability**:

$$P(Y=0|X=1) = P(Y=1|X=0) = p$$
$$P(Y=0|X=0) = P(Y=1|X=1) = 1-p \qquad (1)$$

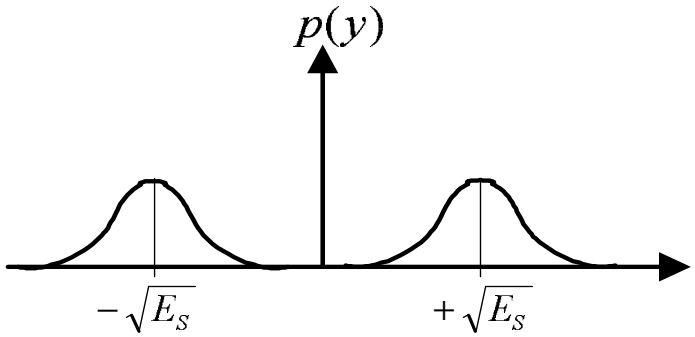
AWGN CHANNEL

- Equivalent model (after matched filter and sampling) :

$$y_i = x_i + n_i \qquad \text{where} \qquad x_i = \pm\sqrt{E_s} \qquad \text{(BPSK modulation)}$$

n_i is a centered random gaussian variable with variance $\sigma^2 = \frac{N_0}{2}$

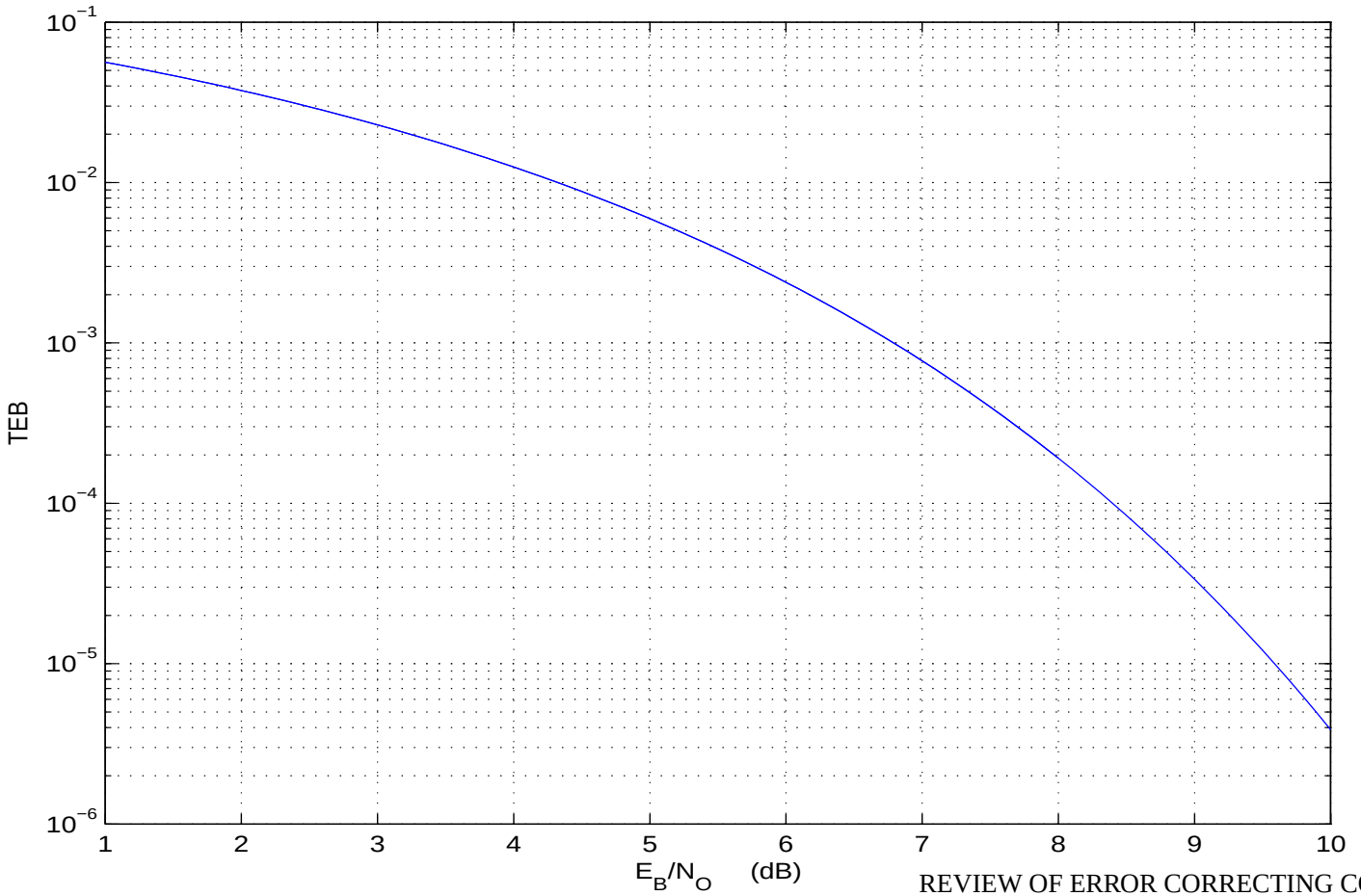
- The ML detector performs a simple threshold.



$$BER = \frac{1}{2}\text{erfc}\left\{\sqrt{\frac{E_B}{N_0}}\right\} \qquad \text{with} \qquad \text{erfc}(a) = \frac{2}{\sqrt{\pi}}\int_a^{+\infty} \exp(-z^2)dz$$

PERFORMANCE

- $BER = f(E_b/N_0)$:



BINARY SYMETRIC CHANNEL

- When using binary modulation, the BSC channel can be seen as an AWGN channel+ decision
- the probability transition is given by (without channel coding)

$$p = \frac{1}{2}\text{erfc}\left\{\sqrt{\frac{E_B}{N_0}}\right\}$$

- the probability transition is given by (with rate R channel coding)

$$p = \frac{1}{2}\text{erfc}\left\{\sqrt{\frac{RE_B}{N_0}}\right\} \quad (2)$$

CHANNEL CAPACITY

definition : The channel capacity is the maximum of the mutual information.

$$C = \max I(X,Y) \quad \text{with} \quad I(X,Y) = H(X) - H(X|Y) \tag{3}$$

- C in Shannon/symbol
- C' capacity per time unit $C' = C \times D_s$

CHANNEL CODING THEOREM

theorem : There exist a channel coding allowing a communication with as small an error probability as desired if and only if :

$$H(U) < C \quad \text{in Sh/symb} \tag{4}$$

$H(U)$ is the entropy at the input of the channel encoder

If we multiply $H(U)$ and C by the symbol rate D_S we have

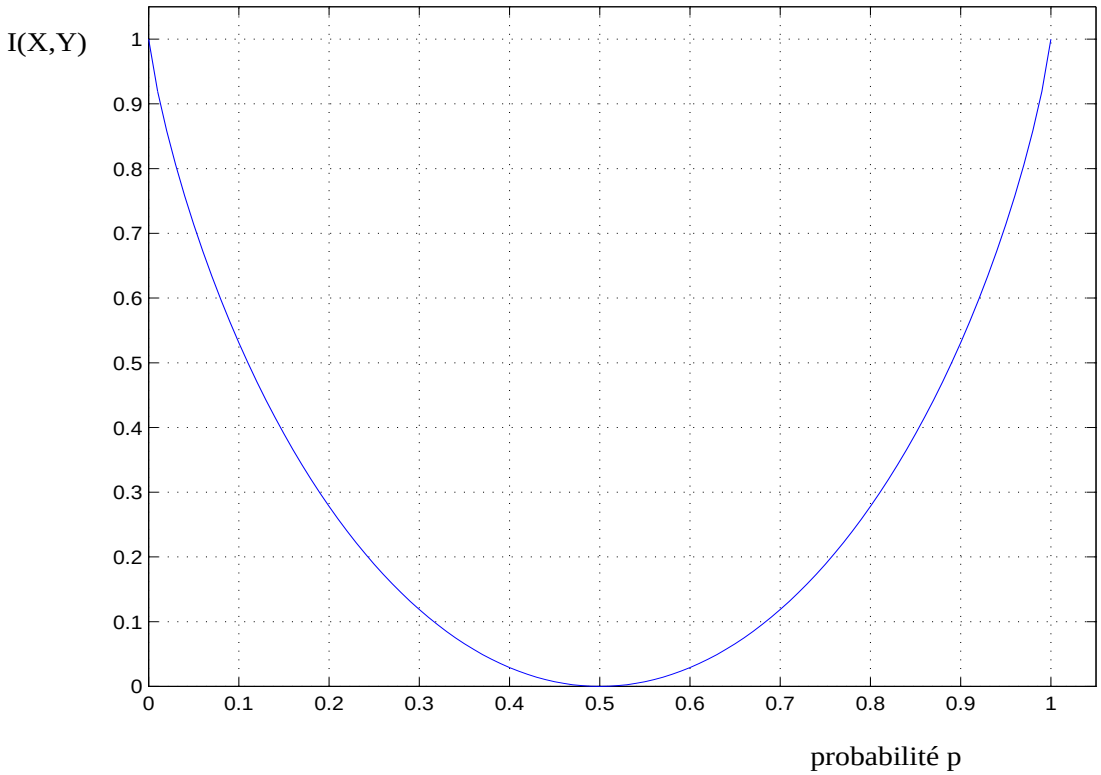
$$H(U) \times D_S < C \times D_S \tag{5}$$

$$D_I < C' \quad \text{Sh/sec} \tag{6}$$

BSC CHANNEL CAPACITY

For $P(X=0) = P(X=1) = 1/2$:

$$I(X,Y) = 1 + p\log_2(p) + (1-p)\log_2(1-p) \qquad (7)$$



AWGN CHANNEL CAPACITY

The relation between the transmitted vector $\mathbf{x}$ and the received vector $\mathbf{y}$ of dimension D is

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \tag{8}$$

Let $\mathbf{n} = (n_1, n_2, \ldots, n_D)$ the noise vector where each element are gaussian, independent with variance σ_n^2.

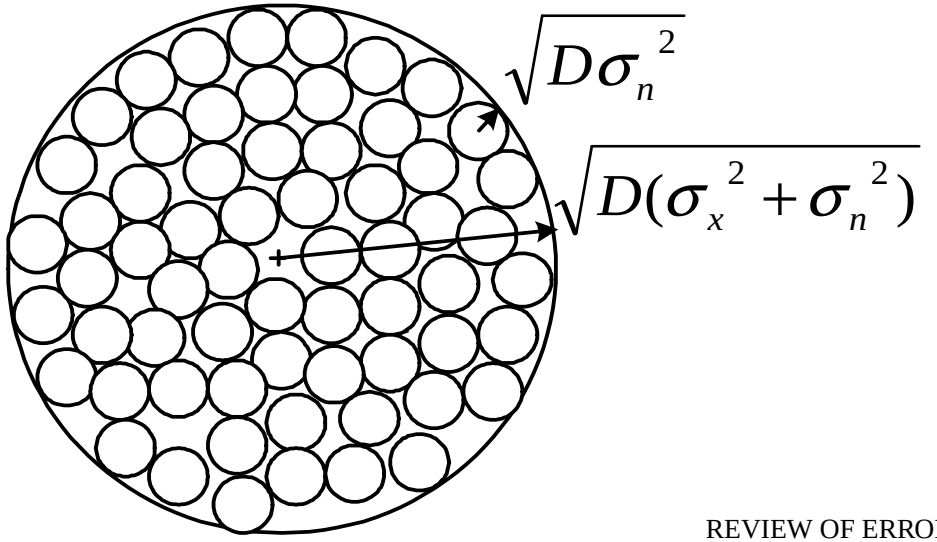
Let $\mathbf{x} = (x_1, x_2, \ldots, x_D)$ the transmitted vector where each element are gaussian, independent with variance σ_x^2 (in order to maximize the mutual information).

AWGN CHANNEL CAPACITY

For $D \to \infty$, we can show that the norm of the noise vector is concentrated on the surface of the D dimension sphere with radius $\sqrt{D\sigma_n^2}$

The norm of the vector $\mathbf{x}$ is concentrated on the surface of the D dimension sphere with radius $\sqrt{D\sigma_x^2}$

The norm of the vector $\mathbf{y}$ is concentrated on the surface of the D dimension sphere with radius $\sqrt{D(\sigma_x^2 + \sigma_n^2)}$.



AWGN CHANNEL CAPACITY

Let M the number of distinguishable vectors $\mathbf{x}$.

In order to guaranty a communication without error, the total volume of the M noise spheres should be smaller than the volume of the sphere with radius $\sqrt{D(\sigma_x^2+\sigma_n^2)}$:

$$\begin{aligned} M &\leq \frac{V(\sqrt{D(\sigma_x^2+\sigma_n^2)}, D)}{V(\sqrt{D.\sigma_n^2}, D)} \\ &\leq \frac{(D(\sigma_x^2+\sigma_n^2))^{D/2}}{(D.\sigma_n^2)^{D/2}} \\ &\leq \left(1+\frac{\sigma_x^2}{\sigma_n^2}\right)^{D/2} \end{aligned} \tag{9}$$

AWGN CHANNEL CAPACITY

$$H(U) = \frac{1}{D} \log_2 M \leq C \tag{10}$$

Consequently :

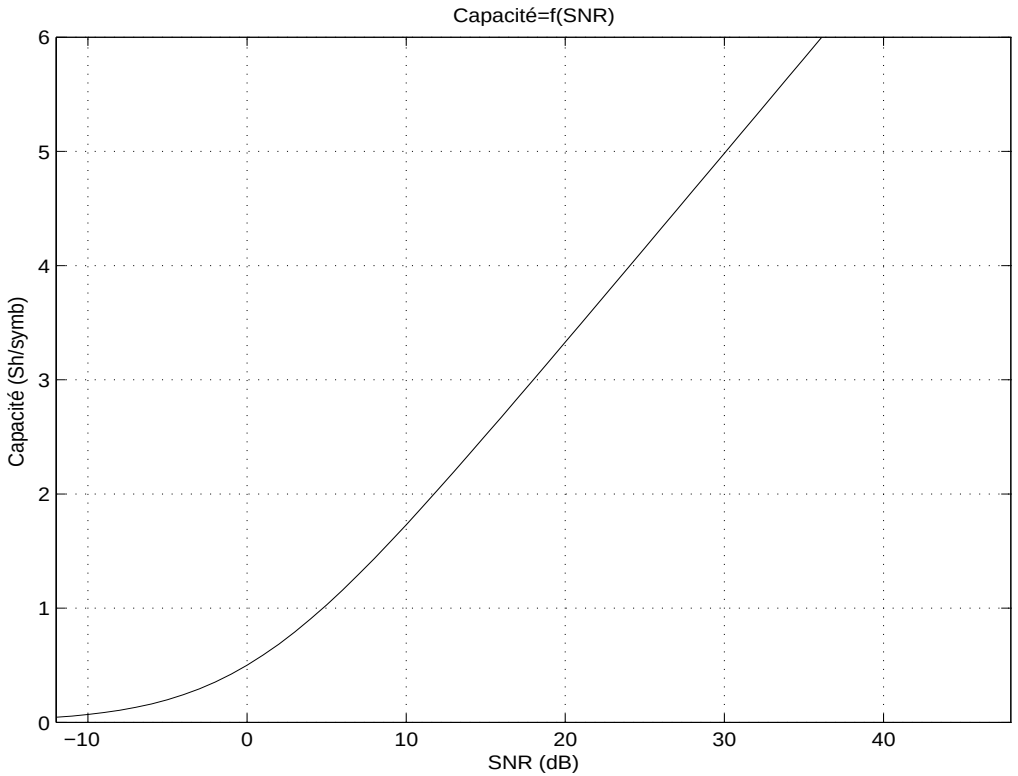
$$C = \frac{1}{2} \log_2 \left(1 + \frac{\sigma_x^2}{\sigma_n^2} \right) \tag{11}$$

For a bandwith B, $D = 2BT$ (T is the transmission duration). The noise power is $N = 2B\sigma_n^2$ and the signal power is $P = 2B\sigma_x^2$.

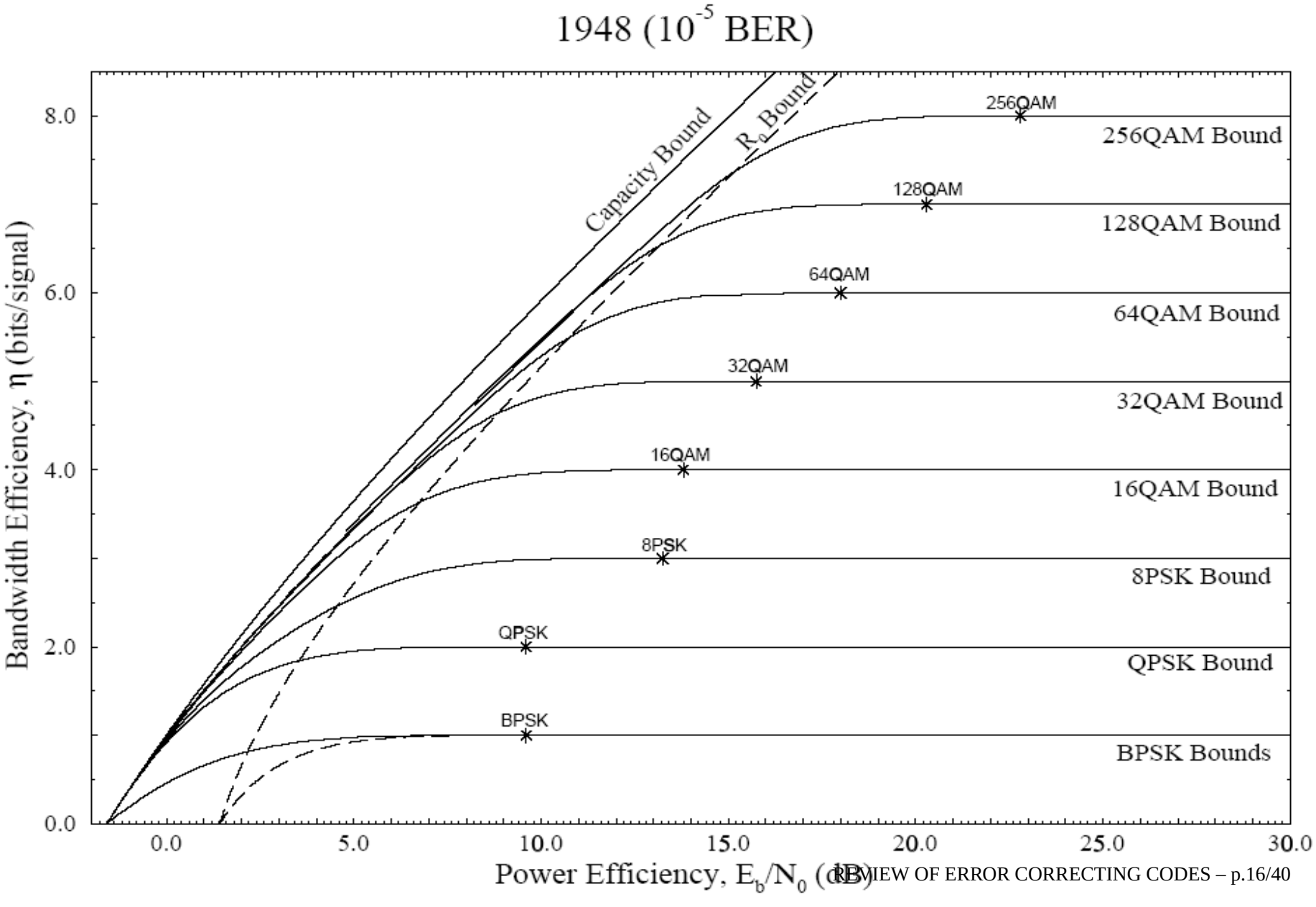
$$C = \frac{1}{2} \log_2 \left(1 + \frac{P}{N} \right) \quad \text{Sh/dim} \tag{12}$$

$$C' = B \log_2 \left(1 + \frac{P}{N} \right) \quad \text{Sh/s} \tag{13}$$

AWGN CHANNEL CAPACITY

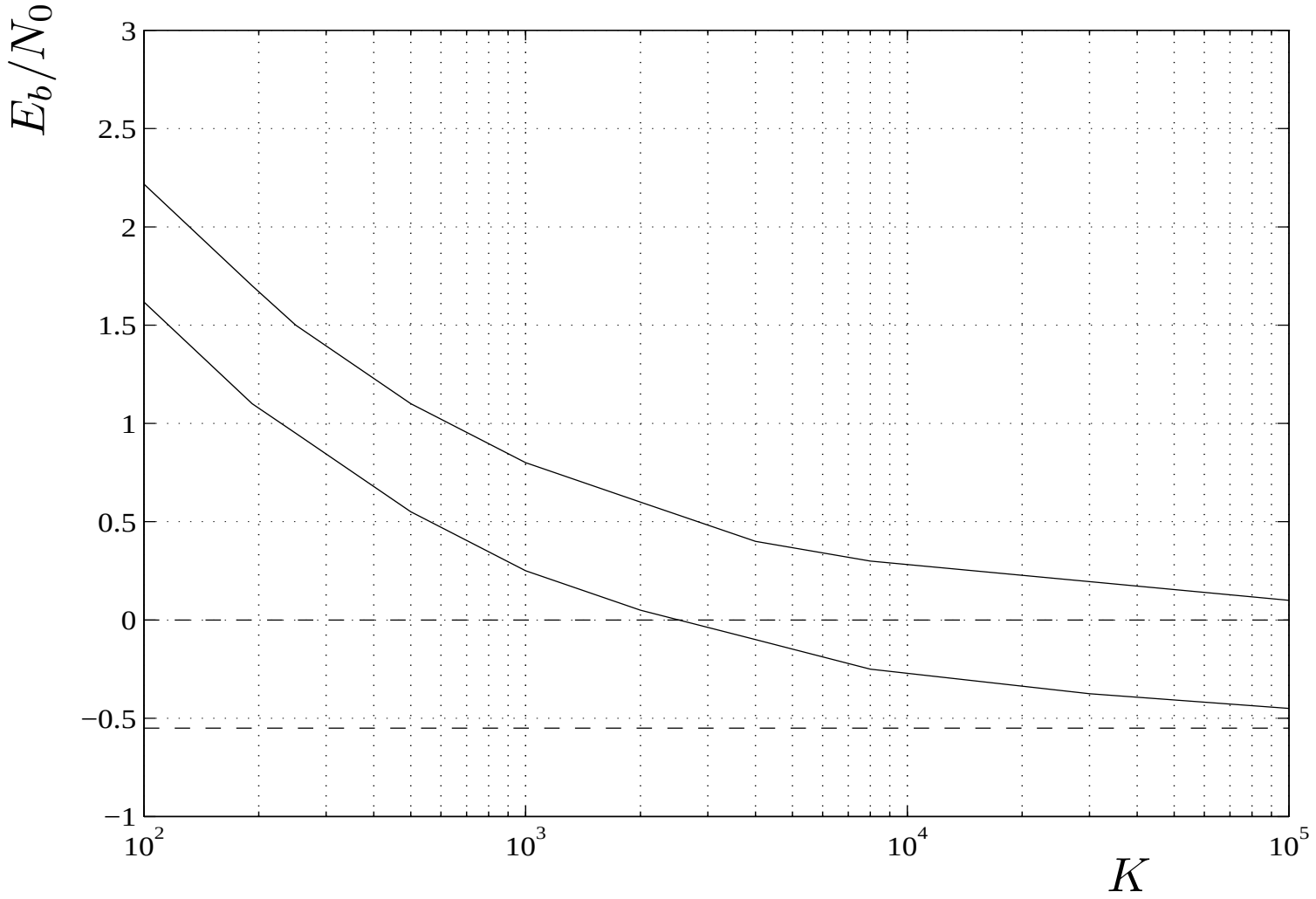


SPECTRAL EFFICIENCY



SPHERE PACKING BOUND

- E_b/N_0 versus K for rate $R = 1/2$ and $R = 1/3$



CHANNEL CODING

• The aim of the channel coding is to protect the message against the channel perturbations by adding redundancy.

• Instead of using a random coding, we will use codes with an algebraic structure such as the linearity to simplify the encoding and also the decoding.

There are three families of error correcting codes

- **The linear block codes**
- **The convolutional codes**
- **The concatenated codes**

BINARY LINEAR BLOCK CODES

Let $\mathbf{u} = [u_1, u_2, \ldots, u_K]$ an information vector composed of K information bits

Let $\mathbf{c} = [c_1, c_2, \ldots, c_N]$ the associated codeword composed of N bits.

We have the matrix relation between $\mathbf{u}$ and $\mathbf{c}$:

$$\mathbf{c} = \mathbf{uG} \tag{14}$$

where $\mathbf{G}$ is the generator matrix of the encoder of dimension $K \times N$.

$$\mathbf{G} = \begin{pmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \vdots \\ \mathbf{g}_K \end{pmatrix} = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1N} \\ g_{21} & g_{22} & \cdots & g_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ g_{K1} & g_{K2} & \cdots & g_{KN} \end{pmatrix} \tag{15}$$

PROPERTIES AND DEFINITIONS

- **Rate** : the rate R of a block code (N, K) is $R = \frac{K}{N}$

- **Hamming distance** : let $\mathbf{c}_1$ and $\mathbf{c}_2$ be two codewords of the binary code $\mathcal{C}$, the Hamming distance $d_H(\mathbf{c}_1, \mathbf{c}_2)$ is the number of different bits between the two codewords.

  Example : $\mathbf{c}_1 = [001100]$ et $\mathbf{c}_2 = [001111]$, $d_H(\mathbf{c}_1, \mathbf{c}_2) = 2$

- **Hamming weight** : the Hamming weight $w(\mathbf{c})$ of a binary block code $\mathbf{c}$ is the number of non zero bits of this codeword.

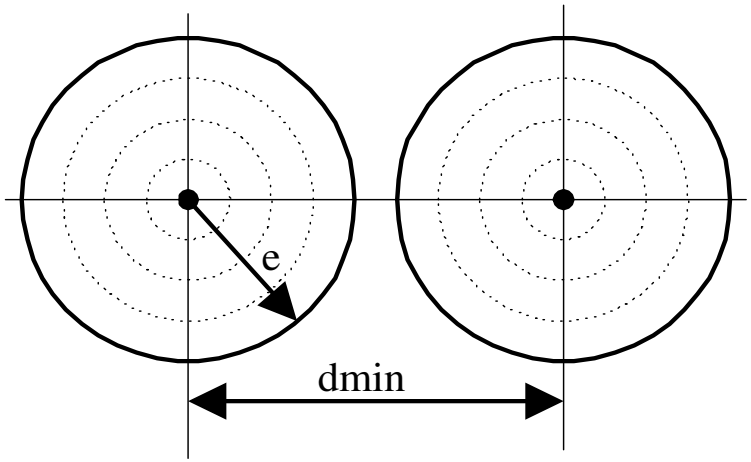
- **Minimum distance** : The minimum distance d_{min} of the code $\mathcal{C}$ is the number of different bits between the two closest codewords :

$$d_{min} = \min_{i,j,i \neq j} d_H(\mathbf{c}_i, \mathbf{c}_j) = \min_{i,i \neq 0} w(\mathbf{c}_i) \tag{16}$$

ERROR CORRECTION CAPACITY

A hard input decoder can decode until e bit errors with :

$$e = \left\lfloor \frac{d_{min} - 1}{2} \right\rfloor \tag{17}$$



PARITY CHECK MATRIX

Each codeword $\mathbf{x}$ of $\mathcal{C}$ is orthogonal to the parity check matrix $\mathbf{H}$:

$$\mathbf{cH}^T = \mathbf{0}$$

Since this relation is true for all the codewords, we have

$$\mathbf{GH}^T = \mathbf{0}$$

Each line of the parity check matrix is associated to a parity check equation

HARD DECODING OF BLOCK CODES

- The received word **r** is the modulo 2 summation between the transmitted codeword **x** and the error vector **e**

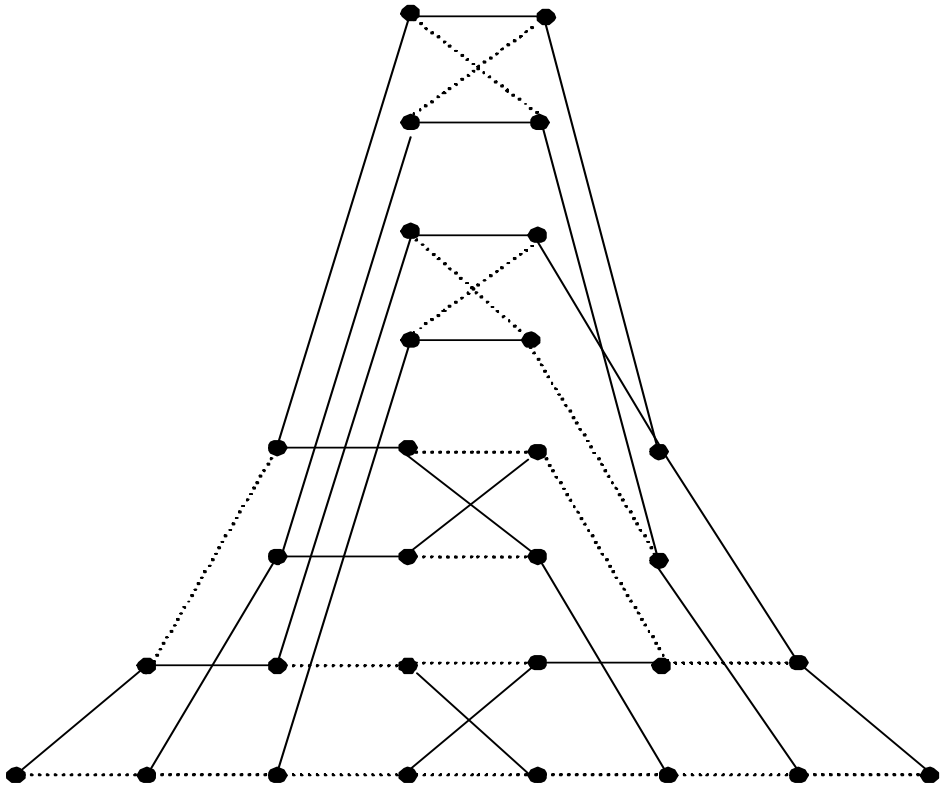
$$\mathbf{r} = \mathbf{c} + \mathbf{e}$$

syndrome decoding

$$\begin{aligned} \mathbf{s} &= \mathbf{r}\mathbf{H}^T \\ &= \mathbf{c}\mathbf{H}^T + \mathbf{e}\mathbf{H}^T \\ &= \mathbf{e}\mathbf{H}^T \qquad \text{since} \quad \mathbf{c}\mathbf{H}^T = 0 \end{aligned} \tag{18}$$

SOFT DECODING OF BLOCK CODES

- A binary block code can be represented graphically using a trellis.
- Exemple : Hamming code (7,4)



- To perform the soft decoding we can use the Viterbi algorithm
- Another soft decoding algorithm : Chase algorithm

WEIGHT ENUMERATOR FUNCTION

Definition 1 : the weight enumerator function (WEF) of a binary block code (N, K) is given by :

$$A(D) = \sum_{d=0}^{N} A_d D^d \tag{19}$$

where A_d is the number of codewords of weight d.

OPTIMAL DETECTION

• Let $\mathbf{x}$ be the transmitted vector over a memoryless stationary discrete channel with conditional probability density function $p(y/x)$ and $\mathbf{y}$ be the received vector.

• A *maximum a posteriori* (MAP) search among all the possible messages $\mathbf{x}$, the estimated message $\hat{\mathbf{x}}$ with the highest $Pr(\mathbf{x}|\mathbf{y})$.

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} Pr(\mathbf{x}|\mathbf{y}) \tag{20}$$

• A *maximum likelihood* (ML) search among all the possible messages $\mathbf{x}$, the estimated message $\hat{\mathbf{x}}$ with the highest $p(\mathbf{y}|\mathbf{x})$.

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}) \tag{21}$$

OPTIMAL DETECTION

- Using Bayes rule, we have :

$$Pr(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})Pr(\mathbf{x})}{p(\mathbf{y})} \tag{22}$$

- Equiprobable messages $\Rrightarrow$ MAP detector= ML detector.

OPTIMAL DETECTION

BSC channel case

$$p(\mathbf{y}|\mathbf{x}) = p^{d_H(\mathbf{y},\mathbf{x})}(1-p)^{N-d_H(\mathbf{y},\mathbf{x})} = (1-p)^N \left(\frac{p}{1-p}\right)^{d_H(\mathbf{y},\mathbf{x})} \quad (23)$$

where $d_H(\mathbf{y},\mathbf{x})$ is the Hamming distance between $\mathbf{y}$ and $\mathbf{x}$. Since $0 \leq p \leq 0.5$ we have $0 < \frac{p}{1-p} < 1$.

The maximisation of $p(\mathbf{y}|\mathbf{x})$ is equivalent to the minimization of $d_H(\mathbf{y},\mathbf{x})$.

$$\begin{aligned} WER_{hard} &\leq \sum_{i=e+1}^{N} \binom{N}{i} p^i (1-p)^{N-i} \\ &\leq 1 - \sum_{i=0}^{e} \binom{N}{i} p^i (1-p)^{N-i} \end{aligned}$$

e is the error correction capacity of the code

OPTIMAL DETECTION

AWGN channel case

After matched filter and sampling we have :

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \tag{24}$$

with $x_i = \pm\sqrt{RE_b}$ (bipodal modulation) and n_i gaussian random variable with variance $\sigma^2 = \frac{N_0}{2}$.

$$p(y_i|x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(y_i - x_i)^2}{2\sigma^2} \right\} \tag{25}$$

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \sum_{i=0}^{N-1} (y_i - x_i)^2 \tag{26}$$

PAIRWISE ERROR PROBABILITY

• Let $\mathbf{x}_i$ and $\mathbf{x}_j$ be two codewords. The euclidian distance between them is $d(\mathbf{x}_i, \mathbf{x}_j)$. For the AWGN channel, the probability $Pr(\mathbf{x}_i \rightarrow \mathbf{x}_j)$ that $\mathbf{y}$ be closer to $\mathbf{x}_j$ than $\mathbf{x}_i$ assuming $\mathbf{x}_i$ is transmitted is given by :

$$Pr(\mathbf{x}_i \rightarrow \mathbf{x}_j) = \frac{1}{2}\text{erfc}\left(\frac{d(\mathbf{x}_i, \mathbf{x}_j)}{2\sqrt{N_0}}\right) \tag{27}$$

If the Hamming distance between two codewords $\mathbf{x}_i$ and $\mathbf{x}_j$ is d, their euclidian distance is $2\sqrt{dRE_b}$ where R is the rate of the code.
Then we have :

$$Pr(\mathbf{x}_i \rightarrow \mathbf{x}_j) = \frac{1}{2}\text{erfc}\left(\sqrt{dR\frac{E_b}{N_0}}\right) \tag{28}$$

WORD ERROR PROBABILITY

Using the union bound, we obtain the upper bound on the word error probability (WER) of the ML decoder on AWGN channel associated to the linear block code (N, K) :

$$WER \leq \frac{1}{2} \sum_{d=d_{min}}^{N} A_d \text{erfc}\left(\sqrt{dR\frac{E_b}{N_0}}\right)$$

where A_d is the number of codewords of weight d.

SOFT AND HARD DECODER WER PERFORMANCE

- hard decoding

$$WER_{hard} \leq 1 - \sum_{i=0}^{e} \binom{N}{i} p^i (1-p)^{N-i}$$

e error correction capacity

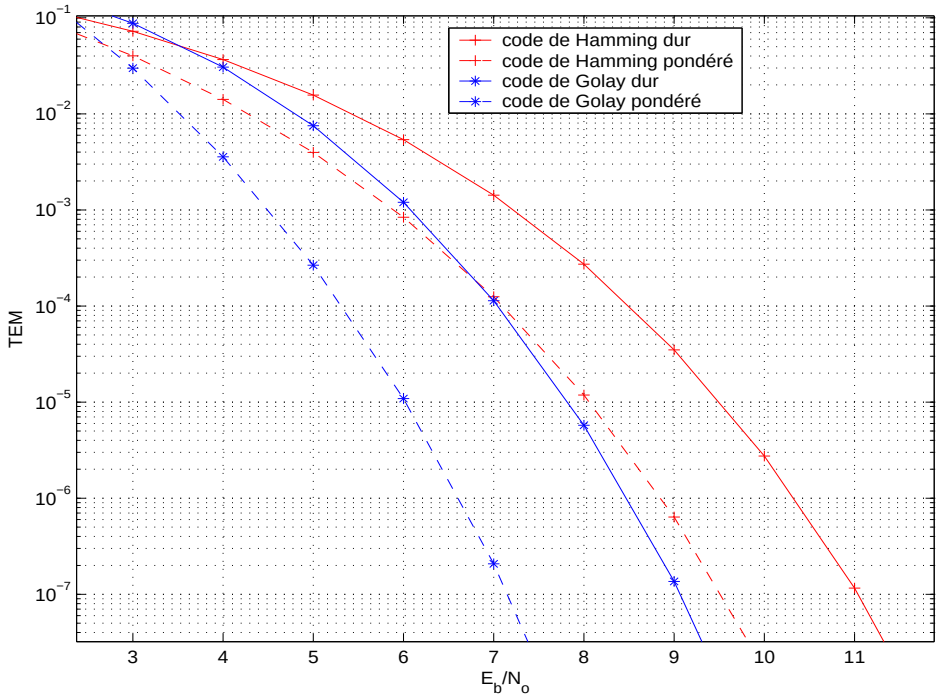
$$p = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{REb}{N_0}}\right) \tag{29}$$

- soft decoding

$$WER_{soft} \leq \frac{1}{2} \sum_{d=d_{min}}^{N} A_d \text{erfc}\left(\sqrt{dR\frac{E_b}{N_0}}\right)$$

SOFT AND HARD DECODER WER PERFORMANCE

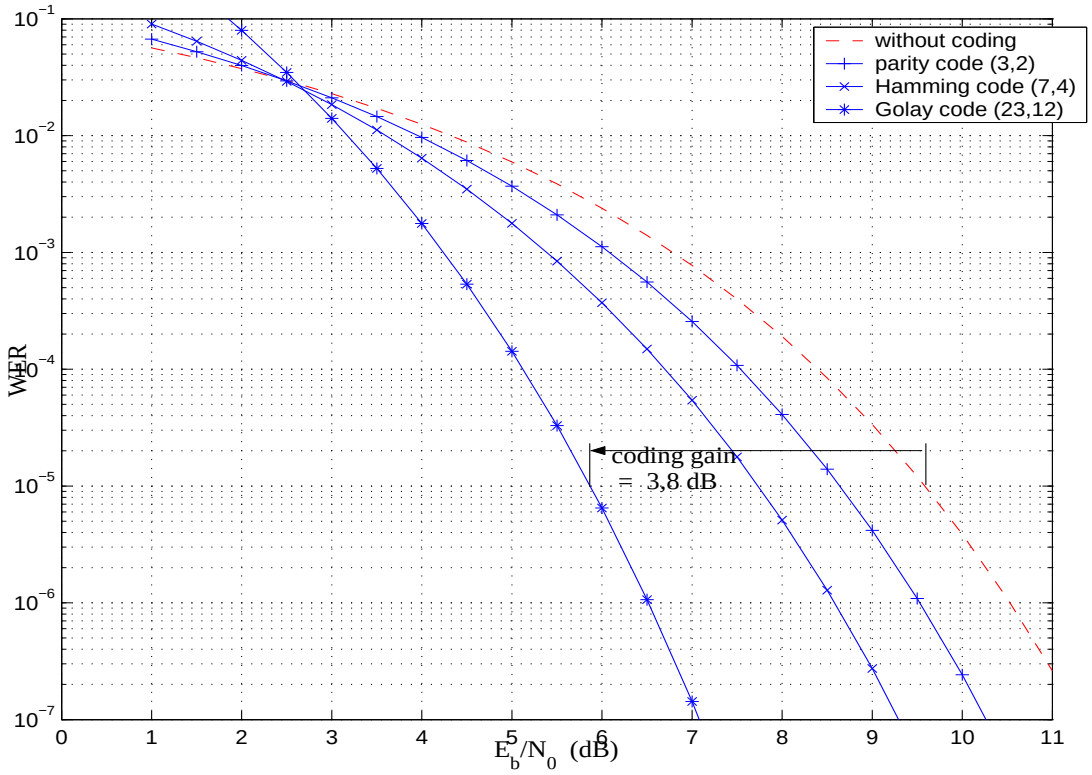
$WER = f(E_b/N_0)$ of a transmission chain using an Hamming code (7,4)and a Golay code (23,12).



• We obtained about 2 dB gain using soft input decoding compared to hard input decoding.

CODING GAIN

- According to the channel capacity, it is theoretically possible to obtain a transmission without error for $E_b/N_0 = 0dB$ using a rate 1/2 code.
- The coding gain is the signal to noise E_B/N_0 difference between a transmission chain with and without channel code.



CONVOLUTIONAL CODES

A convolutional code transforms a semi infinite sequence of information words into a semi infinite sequence of codewords

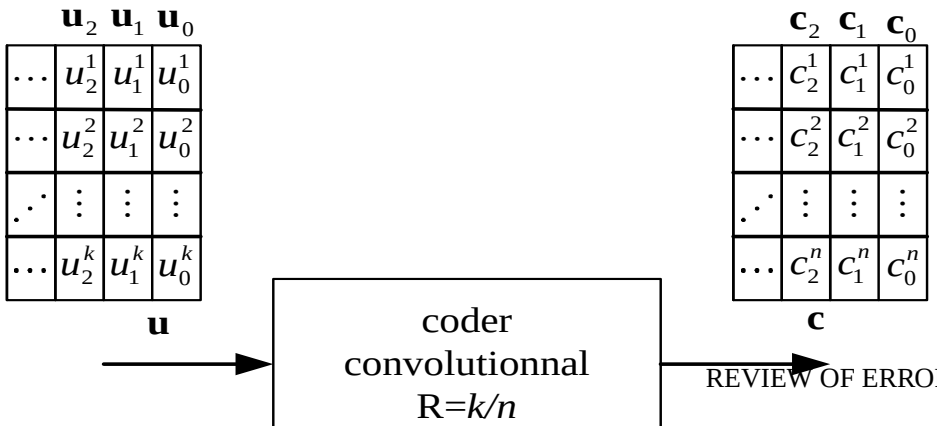
u : information word sequence of dimension k

$$\mathbf{u} = \mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \ldots \qquad \text{with} \qquad \mathbf{u}_i = [u_i^1, u_i^2, ..., u_i^k]$$

x : codeword sequence of dimension n

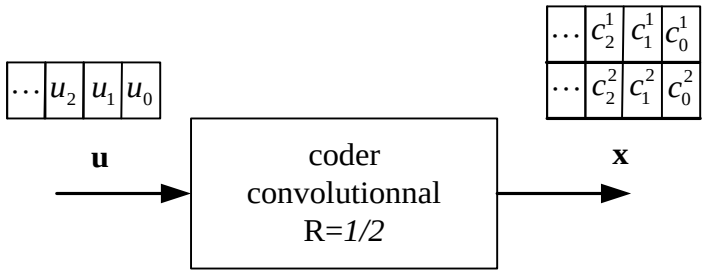
$$\mathbf{c} = \mathbf{c}_0, \mathbf{c}_1, \mathbf{c}_2, \ldots \qquad \text{with} \qquad \mathbf{c}_i = [c_i^1, c_i^2, ..., c_i^n]$$

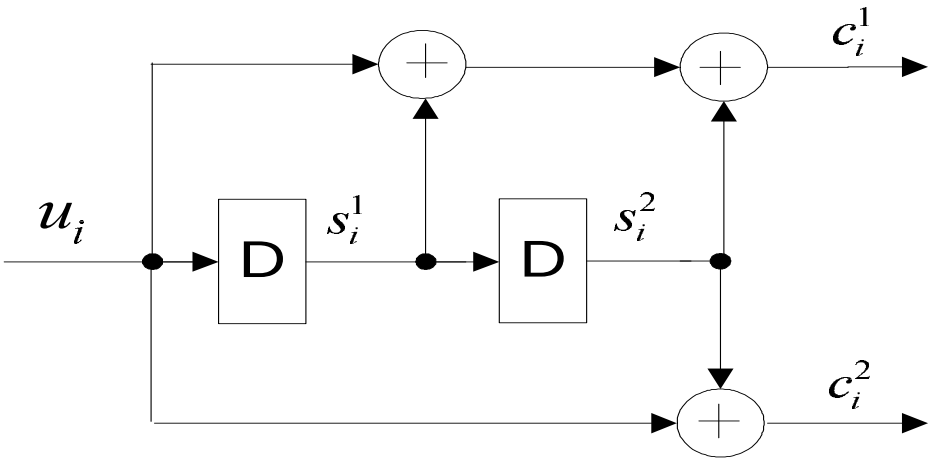
The rate of the convolutional code is $\frac{k}{n}$.



CONVOLUTIONAL CODES

Example : non recursive convolutional encoder $k = 1, n = 2\ M = 2$



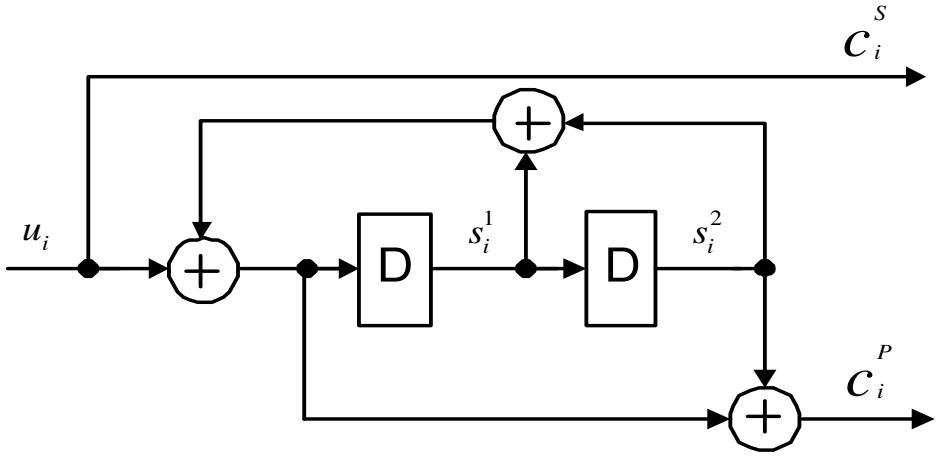


$$c_i^1 = u_i + u_{i-1} + u_{i-2}$$

$$c_i^2 = u_i + u_{i-2}$$

CONVOLUTIONAL CODES

Example : recursive convolutional encoder $k = 1, n = 2\ M = 2$



$$s_{i+1}^1 = u_i + s_i^1 + s_i^2$$

$$s_{i+1}^2 = s_i^1$$

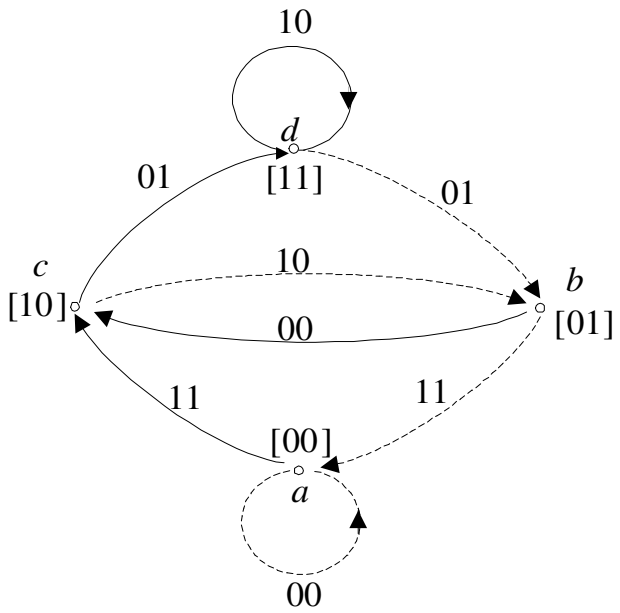
$$c_i^S = u_i$$

$$c_i^P = u_i + s_i^1$$

STATE TRANSITION DIAGRAM

The internal state of the encoder at time i is defined by a vector $\mathbf{s}_i$ of dimension M : $\mathbf{s}_i = [s_{1i}, s_{2i}, ...s_{Mi}]$. s_{ji} is the state at time i of the j-th memory cell.

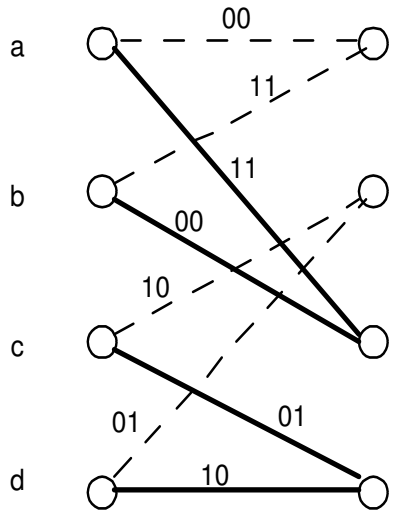
Transition diagram for the non recursive convolutional coder (7,5) of rate 1/2



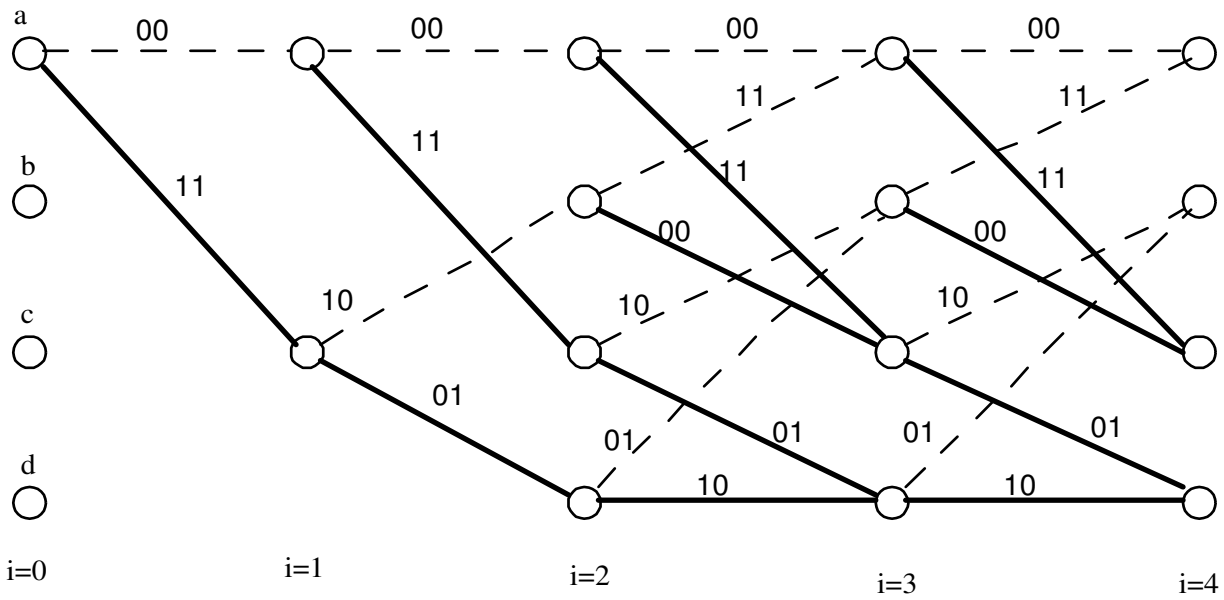
Each branch is labelled with the output bits (here c_{1i} and c_{2i}). The dashed and continuous lines correspond to an input bit 0 and 1 respectively.

ELEMENTARY TRELLIS

From the state transition diagram, it is possible to draw the elementary trellis of the convolutional code. Each branch b links a starting state $s^-(b)$ to an ending state $s^+(b)$.



TRELLIS DIAGRAM



On each branch we labelled the bits c_{1i} and c_{2i}. The continuous and dashed lines correspond to $u_i = 1$ and $u_i = 0$ respectively .