Linear Error-Correcting Codes- Concatenated Codes

Outline

1. Maximum Likelihood Decoding ;
2. linear codes ; Reed-Solomon codes ;
3. bounds ;
4. concatenated codes.

1. Maximum Likelihood Decoder

Consider a memoryless channel $(\mathcal{A}, \mathcal{B}, \Pi)$.

Definition Let C be a code of length n over $\mathcal{A}$. A *decoding algorithm for* C is a procedure that maps any element of $\mathcal{B}^n$ to a codeword of C or that fails ($\mapsto$ symbol ∞).

$$\begin{array}{rccc} \varphi: & \mathcal{B}^n & \to & C \cup \{\infty\} \\ & y & \mapsto & \varphi(y) \end{array}$$

Definition A decoding algorithm φ for C is a *maximum likelihod decoder* if for all $y \in \mathcal{B}^n$, the codeword $x = \varphi(y)$ is in C and maximizes the probability $\mathbf{P}(\ x \text{ sent} \mid\ y \text{ received}\ )$.

Maximum likelihod decoder

- Needs to know the channel and the input distribution of x.

$$\mathbf{P}(\ x \text{ sent} \ \mid \ y \text{ received}) = \mathbf{P}(\ y \text{ received} \ \mid \ x \text{ sent}\ ) \cdot \frac{\mathbf{P}(\ x \text{ sent})}{\mathbf{P}(\ y \text{ received})}$$

The x's which attain these maxima coincide when $\mathbf{P}(\ x \text{ sent}) = \text{constant} = 1/|C|$ (commonly made assumption). This is the uniform codeword distribution assumption.

q-ary symmetric channel

Symmetric channel $(\mathcal{A}, \mathcal{B}, \Pi)$ with $\mathcal{A} = \mathcal{B}$, $|\mathcal{A}| = q$ and

$$\mathbf{P}_{\mathcal{B}|\mathcal{A}}(b|a) = \begin{cases} 1-p & \text{if } a = b \\ \frac{p}{q-1} & \text{otherwise} \end{cases} \qquad \begin{cases} p & \text{error probability} \\ \frac{p}{q-1} & \text{transition/crossover probability} \end{cases}$$

Proposition 1. *In a q-ary symmetric memoryless channel with transition probability $< 1/q$, under the uniform codeword distribution assumption, the most likely codeword $x \in C$ given the received word $y \in \mathcal{B}^n$ is a word minimizing $d_H(x, y)$, where $d_H(x, y)$ is the Hamming distance between x and y : $d_H(x, y) \stackrel{\text{def}}{=} \#\{i | x_i \neq y_i\}$.*

Proof

$$\mathbf{P}(y \mid x) = \left(\frac{p}{q-1}\right)^{d_H(x,y)} (1-p)^{n-d_H(x,y)} = (1-p)^n \left(\frac{p}{(q-1)(1-p)}\right)^{d_H(x,y)}$$

$$\frac{p}{q-1} < \frac{1}{q} \implies p < 1 - \frac{1}{q} \implies 1 - p > \frac{1}{q}$$

$$\text{therefore } \frac{p}{(q-1)(1-p)} < \frac{p}{1-\frac{1}{q}} < 1$$

$\Rightarrow$ seek for the closest codeword in terms of the Hamming distance.

Other decoders

NCP (Nearest Codeword Problem, Maximum Likelihood Decoding)

LD (List Decoding) A bound e is given. The problem is to find *all* (there might be none) codewords at distance $\leq e$ from the received word.

BDD (Bounded Distance Decoding) A bound e is given. The problem to find *one* (there might be none) codeword at distance $\leq e$ from the received word.

UD (Unambiguous Decoding) Here $e = (d-1)/2$, where d is the minimum distance of the code, and we look for the *unique* codeword at distance $\leq e$ from the received word (when it exists).

Minimum distance – Decoding

Let $\mathcal{C}$ be a code of minimum distance d.

- Two balls of radius $(d-1)/2$ centered around two distinct codewords are disjoint.
  $\Rightarrow$ a code of minimum distance d can correct $\lfloor (d-1)/2 \rfloor$ errors
- A ball of radius $d-1$ centered around a codeword does not contain another codeword.
  $\Rightarrow$ a code of minimum distance d can detect $d-1$ errors.

Performance

Definition A decoding algorithm φ for a code C is *bounded by t* if for all $x \in \mathcal{C}$, $d_H(x, y) \leq t \Rightarrow \varphi(y) = x$

If the converse is true and $\phi(y) \neq \infty$ for all y, the algorithm is a *perfect bounded decoder*. Every code of minimum distance d has a bounded decoder with $t = \lfloor (d-1)/2 \rfloor$.

Proposition 2. *The probability of error after decoding on a binary symmetric channel of crossover probability p for a perfect bounded decoder bounded by t is equal to*

$$\sum_{i=t+1}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

2. Reminder : finite field

A finite field $\mathbb{F}_q$ is a set of cardinality q, with $(+, -, \times, /)$ satisfying the appropriate Abelian group equations and distributive law.

- We necessarily have $q = p^m$, p prime.
- Structure : $\mathbb{F}_{p^m} = \mathbb{F}_p[X]/P_m(X)$ where P is an irreducible polynomial in $\mathbb{F}_p[X]$ of degree m.

Example :

$$\begin{aligned} \mathbb{F}_4 &= \mathbb{F}_2[X]/(1+X+X^2) = \{0, 1, X, 1+X\} \\ X(1+X) &= X^2 + X \equiv 1 \bmod (1+X+X^2) \end{aligned}$$

Linear codes

When the alphabet is a finite field (for example $\mathcal{A} = \mathbf{F}_2 = \{0, 1\}$) the Hamming space $\mathcal{A}^n$ is a vector space.

Definition A linear block code of length n over $\mathbb{F}_q$ (the finite field with q elements) is a subspace of $\mathbb{F}_q^n$.

We say that this is an $[n, k]_q$-code if the code is of dimension k and we say it is an $[n, k, d]_q$-code if its minimum distance is d.

Such a code has q^k elements, and its (q-ary) rate is equal to

$$\frac{\log_q q^k}{n} = \frac{k}{n}$$

The two matrices

A linear code $\mathcal{C}[n,k]_q$ is characterized by

– a generator matrix G (of size $k \times n$ over $\mathbb{F}_q$) :

$$\mathcal{C} = \left\{(u_1, \ldots, u_k)G \mid (u_1, \ldots, u_k) \in \mathbb{F}_q^k\right\}$$

The rows of G form a basis for $\mathcal{C}$.

– or a parity-check matrix H (of size $(n-k) \times n$ over $\mathbb{F}_q$) :

$$\mathcal{C} = \left\{(x_1, \ldots, x_n) \in \mathbb{F}_q^n \mid H(x_1, \ldots, x_n)^T = 0\right\}$$

The rows of H form a basis of the dual code $\mathcal{C}^\perp$ of $\mathcal{C}$:

$$\mathcal{C}^\perp \stackrel{\text{def}}{=} \{(v_1, \ldots, v_n) | \forall (c_1, \ldots, c_n) \in \mathcal{C}, \sum_{i=1}^{n} v_i c_i = 0\}.$$

Linear codes – Properties

Proposition 3. *For any linear block code $\mathcal{C}$*

$$\min_{x \neq y | x,y \in \mathcal{C}} d_H(x,y) = \min_{x \neq 0 | x \in \mathcal{C}} w_H(x)$$

(the minimum distance is equal to the minimum nonzero weight of a codeword)

Proposition 4. *Let $\mathcal{C}$ be a code of parity-check matrix H*

$$\left(\begin{array}{c} \mathcal{C} \text{ of minimum} \\ \text{distance} \geq d \end{array} \right) \Leftrightarrow \left(\begin{array}{c} \text{any set of } d-1 \text{ columns of} \\ H \text{ are linearly independent} \end{array} \right)$$

Linear codes – Syndrome decoding

The following syndrome mapping is associated to any parity-check matrix H of $\mathcal{C}$

$$\begin{array}{rccc} \sigma: & \mathbb{F}_q^n & \to & \mathbb{F}_q^{n-k} \\ & y & \mapsto & Hy^T \end{array}$$

Consider $\sigma^{-1}(s) = \{y \in \mathbb{F}_q^n \mid \sigma(y) = s\}$. We obtain

$$\sigma^{-1}(Hy^T) = y + \mathcal{C} = \{y + c \mid c \in \mathcal{C}\}$$

For all $s \in \mathbb{F}_q^{n-k}$, denote by $L_H(s)$ the word of minimal weight in $\sigma^{-1}(s)$ (if there are several of them, one of them is just chosen arbitrarily).

Proposition 5. *The decoder $y \mapsto y - L_H(Hy^T)$ is a maximum likelihood decoder over the q-ary symmetric channel.*

Syndrome decoding

Table lookup decoder : Put $L_H(s)$ in a lookup table.

Algebraic decoding : Find in an algebraic fashion $L_H(s)$ for certain values of s.

Example

$[7, 4, 3]$ Hamming code.

- Can be obtained by solving the following question : find the longest binary linear code of type $[n, k, 3]$ such that $n - k = 3$.

The Hamming code

$[7,4]_2$ Hamming code. Parity-check matrix :

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

For every received word $y \in \mathbb{F}_2^7$, there are 8 syndromes which are possible

$$Hy^T \in \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\},$$

$\Rightarrow$ every word in the Hamming space $\{0,1\}^n$ can be written as $x+e$ with x in the code and e being of weight at most 1 (perfect code = code of minimum distance d with balls of radius $\lfloor \frac{d-1}{2} \rfloor$ centered around the codewords which partition the ambient space).

Hamming Code of length $2^m - 1$

It is a code of length $2^m - 1$ and dimension $2^m - m - 1$, whose parity-check matrix columns are vectors in $\mathbb{F}_2^m \setminus \{0\}$. It is an $[n = 2^m - 1, k = 2^m - m - 1, d = 3]$ perfect code

$$\begin{aligned} 2^{2^m-m-1}\left(\binom{n}{1}+1\right) &= 2^{2^m-m-1}2^m \\ &= 2^n \end{aligned}$$

Theorem 1. *The parameters of perfect codes are known : they are those of the repetition code in odd length, those of the Hamming code, those of the* $[23, 12, 7]_2$ *binary Golay code and those of the* $[11, 6, 5]_3$ *ternary Golay code.*

3. Bounds : Singleton bound

Proposition 6. (Singleton bound)
For all $[n, k, d]$-codes we have $d \leq n - k + 1$.

Proof

The parity-check matrix has $n-k$ rows. There exists therefore a set of $n-k+1$ columns of H which are linearly dependent (actually any set of $n-k+1$ columns has this property) :

$$\Longrightarrow \; d \leq n-k+1.$$

A code such that $k+d=n+1$ is MDS (Maximum Distance Separable).

Reed-Solomon Codes

These are codes defined over large alphabets $\mathbb{F}_q$. We choose n distinct elements $x_1, \ldots, x_n \in \mathbb{F}_q$.

Let ev be the evaluation function :

$$\begin{array}{rccl} \mathrm{ev}: & \mathbb{F}_q[X] & \to & \mathbb{F}_q^n \\ & f & \mapsto & \mathrm{ev}(f) = (f(x_1), \ldots, f(x_n)) \end{array}$$

and

$$L = \{f \in \mathbb{F}_q[X] \mid \deg f < k\}.$$

The Reed-Solomon code of dimension k is given by

$$C \stackrel{\text{def}}{=} \mathrm{ev}(L).$$

Parameters

Proposition 7. *If $k \leq n$, this is a code of dimension k and minimum distance $d = n - k + 1$, correcting $t = \lfloor \frac{n-k}{2} \rfloor$ errors.*

Proof :

- If $k \leq n$, then ev is one-to-one.
- a polynomial of degree $< k$ has at most $k - 1$ zeros. There are therefore at least $n - k + 1$ non-zero coordinates in a non-zero codeword.

Moreover, the polynomial $\prod_{i=1}^{k-1}(X - x_i)$ has exactly $k - 1$ zeros.
The Reed-Solomon codes are MDS.

Reed-Solomon decoding by interpolation

Let $y = (y_1, \ldots, y_n)$ be the received word and c be the closest codeword with $c = \mathrm{ev}(f(X))$ where $\deg f(X) < k$.

let I be the set of positions where there is an error :

$$I = \{i \in \{1, \ldots, n\}, \quad f(x_i) \neq y_i\},$$

and construct the polynomial $E(X) = \prod_{i \in I}(X - x_i)$. Then we have

$$E(x_i)y_i = E(x_i)f(x_i), \quad i \in \{1, \ldots, n\}. \tag{1}$$

Decoding (II)

let

$$X^t + \sum_{i=0}^{t-1} e_i X^i \quad \stackrel{\text{def}}{=} \quad E(X)$$

$$\sum_{i=0}^{t+k-1} a_i X^i \quad \stackrel{\text{def}}{=} \quad E(X) f(X)$$

- $2t + k$ unknowns and n affine equations :

$$E(x_i) y_i = E(x_i) f(x_i), \quad i \in \{1, \ldots, n\}. \tag{2}$$

- One can hope to correct in this way $\frac{n-k}{2} = \frac{d-1}{2}$ errors.

Hamming bound

Let C be a code of cardinality M, error-correction capacity $t = \lfloor \frac{d-1}{2} \rfloor$, and length n over the alphabet $\mathbb{F}_q$. Then

$$M \left(\sum_{i=0}^{t} (q-1)^i \binom{n}{i} \right) \leq q^n$$

Asymptotic form

$$\begin{aligned} h_q(\delta/2) &\leq 1 - R \text{ with} \\ \delta &\stackrel{\text{def}}{=} d/n \\ R &\stackrel{\text{def}}{=} \log_q M/n \\ h_q(x) &\stackrel{\text{def}}{=} -x \log_q \frac{x}{q-1} - (1-x) \log_q (1-x) \end{aligned}$$

Existence of good codes – Gilbert-Varshamov bound

Theorem 2. ***(Gilbert-Varshamov bound)***

$$\sum_{i=0}^{d-2}(q-1)^i\binom{n-1}{i} < q^{n-k} \Rightarrow \left(\begin{array}{c} \exists \text{ a code} \\ [n,k,d]_q \end{array} \right)$$

Theorem 3. ***(Asymptotic Gilbert-Varshamov bound)***
Let $0 \leq \delta \leq (q-1)/q$. *For all* $0 \leq R < 1 - h_q(\delta)$ *there exists an infinity of* $[n,k,d]_q$*-codes such that* $d \geq \delta n$ *and* $k \geq Rn$ *where* $h_q(x) \stackrel{\text{def}}{=} -x\log_q\frac{x}{q-1} - (1-x)\log_q(1-x)$ *is the* q*-ary entropy function.*

Proof

We construct the columns of a parity-check matrix of such a code one by one, with the property that any subset of $d-1$ columns is linearly independent.

Assume now that the i first columns are such that any subset of columns of size $d-1$ is linearly independent.

Number N of linear combinations involving at most $d-2$ columns among i columns :

$$1 + \binom{i}{1}(q-1) + \cdots + \binom{i}{d-2}(q-1)^{d-2}$$

If $N < q^{n-k}$, one can add a column which is not a linear combination of at most $d-2$ columns.

This can be done as long as

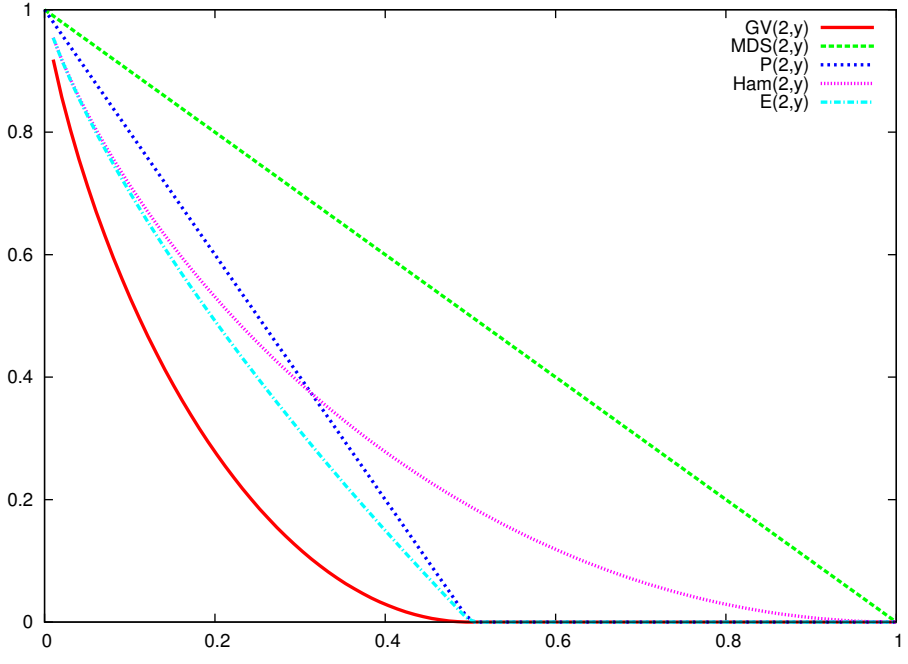
$$1 + \binom{i}{1}(q-1) + \cdots + \binom{i}{d-2}(q-1)^{d-2} < q^{n-k}$$

We finish the proof with the following bounds for $i \leq n-1$:

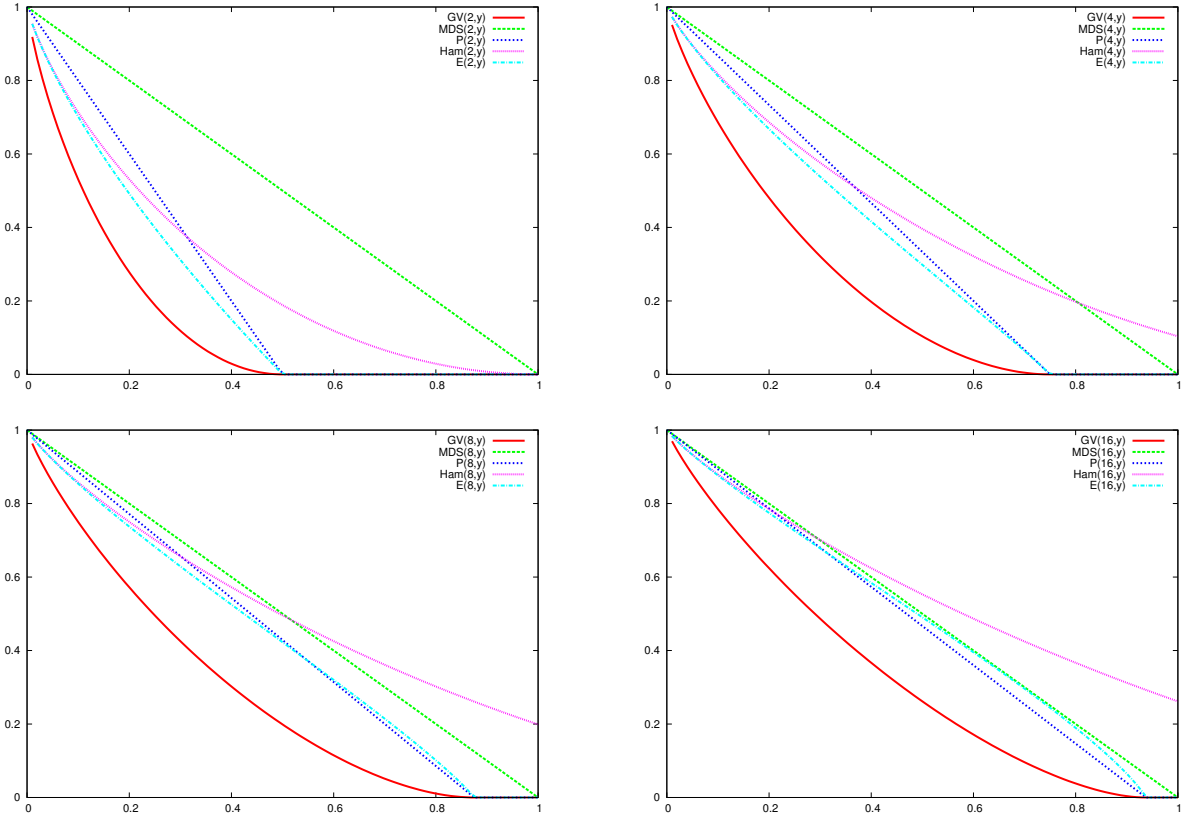
$$\begin{aligned}
& 1 + \binom{i}{1}(q-1) + \cdots + \binom{i}{d-2}(q-1)^{d-2} \\
\leq\ & 1 + \binom{n-1}{1}(q-1) + \cdots + \binom{n-1}{d-2}(q-1)^{d-2} \\
\leq\ & 2^{(n-1)h_q\left(\frac{d-2}{n-1}\right)} \qquad (3) \\
\leq\ & 2^{nh_q\left(\frac{d}{n}\right)} \qquad (4)
\end{aligned}$$

Exercise : Show (3) with an information theoretic proof.

Gilbert-Varshamov – Binary case



Curves for $q \in \{2, 4, 8, 16\}$



Codes attaining the bounds

- The Hamming bound : the perfect codes : repetition codes, Hamming codes, binary and ternary Golay codes.
- Singleton bound : MDS codes : Reed-Solomon codes ($n \leq q$),...
- Gilbert-Varshamov bound : almost any linear code...

Error correction on average/worst case

On a binary symmetric channel of probability p there are typically $\approx pn$ errors in a code of length n and rate R.

One can almost always correct these errors as long as $R < 1 - h(p)$, that is

$$p < h^{-1}(1 - R).$$

The minimum distance d of a linear code of length n and rate R is almost always of the form $d \approx nh^{-1}(1 - R)$. One can correct $t = \frac{d-1}{2}$ errors in all cases with such a code. Note that in this case

$$\frac{t}{n} \approx \frac{h^{-1}(1 - R)}{2}$$

Therefore on average we correct twice as many errors as in the worst case.

4. Concatenated Codes

Idea : use a second level of encoding to reduce the probability of error after decoding.

$$M \in \{0,1\}^k \xrightarrow{\text{coding}} \boxed{C \in \{0,1\}^n \xrightarrow{\text{channel}} C' \in A^n} \xrightarrow{\text{decoding}} \hat{M} \in \{0,1\}^k$$

virtual channel

Let $B \stackrel{\text{def}}{=} \{0,1\}^k$, a code of type $[N, ?]$ over B is chosen to protect the binary codewords (now viewed as symbols in B).

Encoding

$$M = (x_1 \dots x_K) \in B^K \xrightarrow{\text{outer encoding}} C = (y_1 \dots y_N) \in B^N$$

$$\xrightarrow{\text{inner encoding } y_i} C' = (c_1 \dots c_{nN}) \in \{0,1\}^{nN}$$

outer code : code of length N and rate $\frac{K}{N}$ over $B = F_{2^k}$,

Inner code : binary code of length n and rate k/n.

$$\text{Rate of the concatenated code} = \frac{kK}{nN}$$

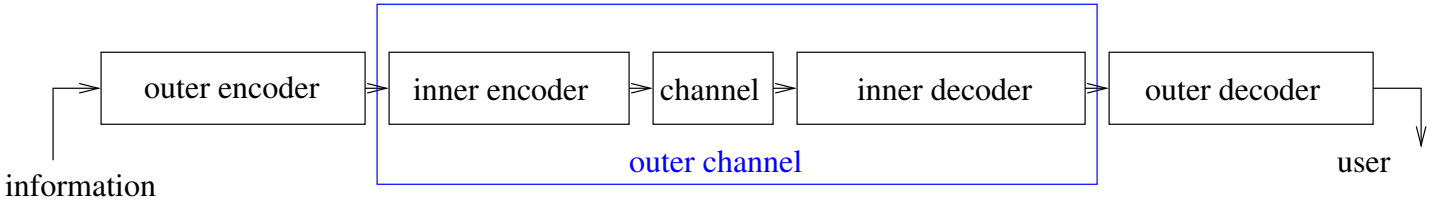
Decoding

$$C' = (c_1 \dots c_{nN}) \in \{0,1\}^{nN} \xrightarrow{\text{channel}} W = (a_1 \dots a_{nN}) \in A^{nN}$$

$$\xrightarrow{\text{inner decoder}} C'' = (y'_1 \dots y'_N) \in B^N$$

$$\xrightarrow{\text{outer decoder}} M' = (x'_1 \dots x'_k) \in B^K.$$

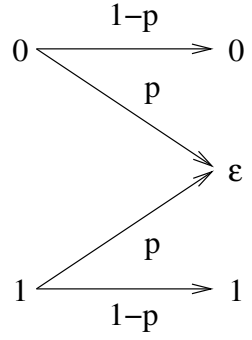
General scheme



1. A word whose symbols are taken over a large alphabet is encoded.
2. each symbol is encoded by the inner encoder
3. the codeword is transmitted
4. each symbol is decoded
5. the decoded word is then decoded with the outer decoder.

Concatenated Codes, tool : decoding erasures

An erasure can be viewed as an error whose location is known. In other words, it is the outcome of the following channel



For any code of minimum distance d, there exists a decoding algorithm correcting $d-1$ erasures.

(There is a single codeword which coincides with the received word on the positions which were not erased)

Error/erasure correction

Proposition 8. *For any code of minimum distance d, there exists a decoding algorithm correcting ν errors and ρ erasures iff*

$$2\nu + \rho < d$$

let J be the set of non-erased positions and

$$C_J = \{c_J; \quad c \in C\}$$

The minimum distance of C_J is $\geq d - \rho$.

One can therefore correct 2ν errors, if $2\nu < d - \rho$.

After this, one recovers the erasures.

Concatenated codes– Decoding

the received word is of the form

$$\underbrace{(y_{1,1},\ldots,y_{1,n})}_{y_1} \quad \| \quad \underbrace{(y_{2,1},\ldots,y_{2,n})}_{y_2} \quad \| \quad \cdots \quad \| \quad \underbrace{(y_{N,1},\ldots,y_{N,n})}_{y_N}$$

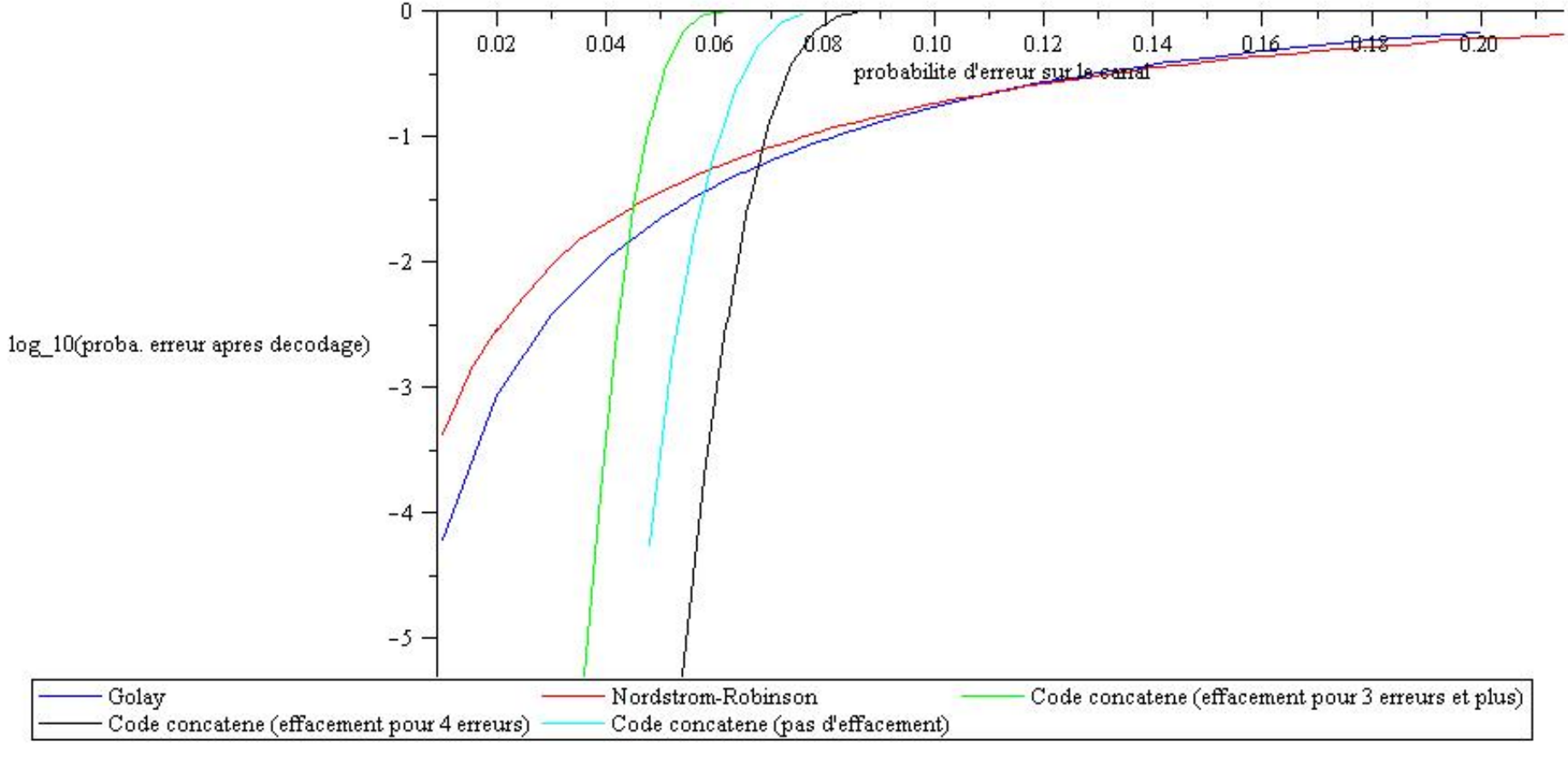
Each of the N blocks is decoded with the inner decoder

$$\begin{array}{ccccc} \varphi & : & \{0,1\} & \to & C_{\text{inner}} \cup \{\infty\} \\ & & y_i & \mapsto & z_i \end{array}$$

Each symbol $(z_1,\ldots,z_N) \leftrightarrow$ a symbol of B or an erasure (symbol ∞). This word is then decoded with the outer code C_{outer}.

$\Rightarrow$ not necessarily optimal to take an optimal inner decoder, $\exists$ optimal value for number of erasures / number of errors.

TD


0
-1
-2
-3
-4
-5
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
0.18
0.20
probabilite d'erreur sur le canal
log_10(proba. erreur apres decodage)
Golay
Nordstrom-Robinson
Code concatene (effacement pour 3 erreurs et plus)
Code concatene (effacement pour 4 erreurs)
Code concatene (pas d'effacement)