

# Recognition of a code in a noisy environment

Christophe Chabot

Université de Limoges - XLIM - DMI  
INRIA Rocquencourt - Projet CODES

ISIT Nice - June 29, 2007

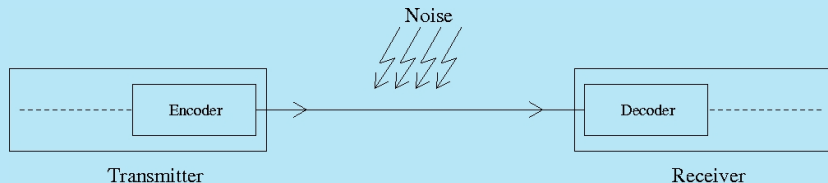
# Outline

- 1 Introduction
- 2 Recognition of a code
  - Problem
  - Theoretical results
  - Experimental results
- 3 Conclusion

# Introduction

## Context:

- Data transmission.
- Several transformations applied to the data.
- Transmitted data is corrupted by noise.



**General Problem:** Finding some information on the transformations used only from an intercepted noisy sequence.

# Reconstruction of code

Recovering the first layer i.e. the error correcting code used.

→ Problem of reconstruction of code.

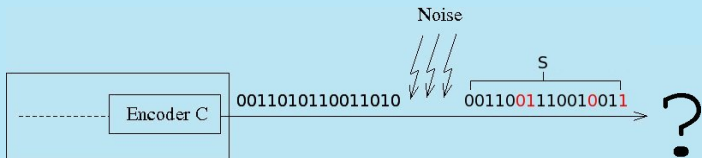
## State of art:

- convolutional codes, turbo codes: [Fil],[Bar],[Din]
- linear block codes: [Val],[Clu],[Bar]  
This problem is **NP-Hard** [Val].  
Complexity exponential in the error rate.

→ Easier problem : recognition of a code.

# Problem

Let be given a binary sequence  $S$  and a binary linear code  $C$ .



Is this sequence  $S$  composed of noisy codewords of  $C$ ?

## Binary linear code

Let  $C$  be a binary linear code of length  $n$  and dimension  $k$ .  
i.e.  $C$  is a vector subspace of  $\mathbb{F}_2^n$  of dimension  $k$ .

It can be represented by:

- its generator matrix (whose lines span the vector subspace  $C$ ).
- its parity check matrix (whose lines span  $C^\perp$ ).

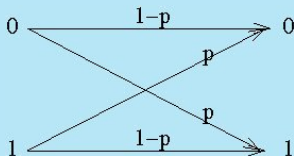
We will use a basis of the dual code of  $C$ ,  $H = (h_1, h_2, \dots, h_{n-k})$ ,  
then, for all  $j \in \{1, \dots, n-k\}$ , for all  $c \in C$ ,

$$\langle h_j, c \rangle = \sum_{i=1}^n h_{j,i} * c_i = 0.$$

We will note, for  $x \in \mathbb{F}_2^n$ ,  $\text{supp}(x) = \{i \in \{1, \dots, n\} \mid x_i \neq 0\}$ ,  
and  $w(x) = \#\text{supp}(x)$  its Hamming weight.

# Binary channel

We will consider a binary symmetric channel with error rate  $p$ .



$S = (s_j)_{1 \leq j \leq M}$  is a sequence taken at the output of the channel,  $s_j \in \mathbb{F}_2^n$ .

**Note:** if  $p = 1/2$ ,  $S$  is random (one-time pad).

# Main idea

- If  $S$  comes from  $C$ :
  - if  $p = 0$ ,  $\langle h_i, s_j \rangle = 0$  for all  $i, j$ .
  - if  $0 < p \ll 1/2$ ,  $\langle h_i, s_j \rangle = 1$  for a few number of  $i, j$ .
- Else:  $\langle h_i, s_j \rangle = 1$  half of the time.

## Different cases

Let  $h$  be a non-zero word of  $C^\perp$ .

- If  $S$  comes from a random sequence:

$$\dim(h^\perp) = n - 1, \quad \frac{\#(h^\perp)}{\#(\mathbb{F}_2^n)} = \frac{1}{2},$$

$$\rightarrow P[\langle h, s_j \rangle = 1] = 1/2.$$

- If  $S$  comes from words of  $C$ :

For all  $j$ ,  $s_j = c_j + e_j$ , with  $c_j \in C$  and  $e_j$  the error vector.

$\langle h, s_j \rangle = 1 \Leftrightarrow \#(\text{supp}(h) \cap \text{supp}(e_j))$  is odd.

$$\rightarrow P[\langle h, s_j \rangle = 1] = \frac{1 - (1 - 2p)^{w(h)}}{2}.$$

## Different cases

- If  $S$  comes from a sequence of words of  $C' \not\subseteq C$ :  
There exists at least one  $h_0$  in a basis of  $C^\perp$  such that  $h_0 \notin C'^\perp$ ,  
→ for this  $h_0$ ,  $P[\langle h_0, s_j \rangle = 1] = 1/2$ .
- Other cases:  
heuristic result but confirmed by experiments.  
→  $P[\langle h, s_j \rangle = 1] = 1/2$ .

# Statistical test

## Summary:

- If  $S$  does not come from  $C$ :

There exists at least one  $h \in \{h_1, \dots, h_{n-k}\}$ , s.t.

$$P[\langle h, s_j \rangle = 1] = 1/2.$$

- If  $S$  comes from  $C$ :

For all  $h \in \{h_1, \dots, h_{n-k}\}$ ,

$$P[\langle h, s_j \rangle = 1] = \frac{1}{2} - \frac{1}{2}(1 - 2p)^{w(h)}.$$

## Idea:

Is  $\sum_{j=1}^M \langle h, s_j \rangle$  (sum in  $\mathbb{Z}$ ) “close” to  $\frac{M}{2} - \frac{M}{2}(1 - 2p)^{w(h)}$ ?

## Theorem

**Theorem:** The statistical test consisting in deciding that  $S = (s_j)_{1 \leq j \leq M}$  comes from a sequence of  $M$  words of  $h^\perp$  if and

only if  $\sum_{j=1}^M \langle h, s_j \rangle \leq T$  with

$$M = \left( \frac{b\sqrt{1-(1-2p)^{2w(h)}-a}}{(1-2p)^{w(h)}} \right)^2, \quad T = \frac{1}{2}(M + a\sqrt{M})$$

and  $a = \phi^{-1}(\alpha)$ ,  $b = \phi^{-1}(1 - \beta)$  verifies:

$$P\left[\left(\sum_{j=1}^M \langle h, s_j \rangle\right) \leq T \mid S \text{ is random}\right] = \alpha \text{ (false alarm),}$$

$$P\left[\left(\sum_{j=1}^M \langle h, s_j \rangle\right) \geq T \mid S \text{ comes from } h^\perp\right] = \beta \text{ (non detection).}$$

**Input:**

$C$  a  $[n, k]$ -linear binary code ,  
 a binary symmetric channel with error rate  $p$ ,  
 $S = (s_j)_{1 \leq j \leq M}$  a sequence taken at the output of the channel,  
 $\alpha, \beta$  false alarm and non detection probabilities.

**Initialization:**

Compute  $(h_1, \dots, h_{n-k})$  a basis of  $C^\perp$  with low weight words.  
 Compute  $M_i$  and  $T_i$  for each  $i \in \{1, \dots, n-k\}$ .

**Algorithm:**

$$N_i = \sum_{j=1}^{M_i} \langle h_i, s_j \rangle \text{ (sum in } \mathbb{Z}) \text{ for } i \in \{1, \dots, n-k\}.$$

**Output:**

If  $N_i \leq T_i$  for all  $i \in \{1, \dots, n-k\}$ ,  
 say that  $S$  comes from a sequence of words of  $C$ .  
 If  $N_i \geq T_i$  for at least one  $i \in \{1, \dots, n-k\}$ ,  
 say that  $S$  does not come from a sequence of words of  $C$ .

# Computing results

| Code used     | $n$  | $w(h)$<br>max | Error<br>rate $p$ | $M$   | $T$   | Time<br>is s |
|---------------|------|---------------|-------------------|-------|-------|--------------|
| <i>BCH</i>    | 511  | 140           | 0.005             | 1462  | 640   | $\leq 1$     |
| <i>BCH</i>    | 511  | 140           | 0.01              | 25823 | 12529 | 17           |
| <i>RM</i>     | 1024 | 136           | 0.005             | 1345  | 585   | 4            |
| <i>RM</i>     | 1024 | 136           | 0.01              | 21963 | 10629 | 28           |
| <i>Random</i> | 2000 | 535           | 0.001             | 724   | 298   | 9            |
| <i>Random</i> | 2000 | 535           | 0.002             | 6540  | 3078  | 62           |

Here,  $\alpha = \beta = 10^{-6}$ .

# Synchronization

## Input:

$C, p, S, \alpha$  and  $\beta$ .

## Initialization:

Compute  $S^{(l)} = (s_j^{(l)})_{1 \leq j \leq M}$ ,  $0 \leq l \leq n-1$ .

Compute  $(h_1, \dots, h_r)$  some words of a basis of  $C^\perp$ .

Compute  $M_i$  and  $T_i$  for each  $i \in \{1, \dots, r\}$ .

## Algorithm:

For  $l$  from 0 to  $n-1$  do

$$N_i^{(l)} = \sum_{j=1}^{M_i} \langle h_i, s_j^{(l)} \rangle \text{ (sum in } \mathbb{Z} \text{) for } i \in \{1, \dots, r\}.$$

## Output:

If  $N_i^{(l)} \leq T_i$  for all  $i \in \{1, \dots, r\}$ ,

say that  $S$  seems to come from a sequence of words of  $C$   
with synchronization  $l \rightarrow$  **check it !**



# Conclusion

- Easy and fast algorithm to recognize a code.
- Can reach high error rate (compared to reconstruction).
- Application to synchronization.

Thanks for your attention.