Universidade de Aveiro

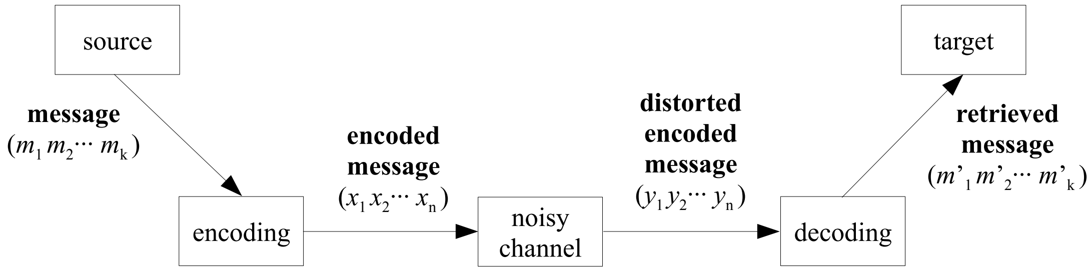
Mestrado Integrado em Engenharia de Computadores e Telemática

Arquitectura de Computadores Avançada

Assignment 1 – Hadamard codes

Academic year 2022/2023

THEORETICAL BACKGROUND



A message $\boldsymbol{m}$, generated by a given *source*, is expressed in k symbols of alphabet Σ and is further encoded, following some definite rule, into an extended message $\boldsymbol{x}$, of n symbols of alphabet Σ. The encoded message $\boldsymbol{x}$ is next transmitted over a *noisy channel*, where the symbols may be changed according to certain probabilities characteristic of the channel. The received message $\boldsymbol{y}$ is finally decoded into the message $\boldsymbol{m}'$, which is retrieved by the *target*.

Bear in mind that the *channel* concept must be taken in a broad sense: it can be a data transmitting medium in the traditional sense, such as copper, optical fiber or air signal propagation, or even a data storage device, such as a hard disk, a flash memory or a dynamic RAM.

One can define the *information rate* R, which measures the slowdown of the effective data transmission, as k/n. On the other hand, given the channel characteristics, one defines the *capacity* C *of the channel* as something which, as Claude Shannon has shown, has the property that, for $R < C$, it is possible to find an *encoding* / *decoding* scheme such that the probability that $\boldsymbol{m}' \neq \boldsymbol{m}$ can be made arbitrarily small. If, however, $R > C$, no such scheme exists.

Claude Shannon, however, did not show how to build such *encoding* / *decoding* schemes. This has been the pioneering work of Richard Hamming. Presently, however, the case for the best codes in terms of the maximal number of errors that one can correct for a given information rate and code length is not clear. Existence theorems are known, but the exact bound is still an open problem.

Without loss of generality, one may assume that the channel uses the same alphabet Σ, of size q, both for input and output. A *code* C *over* Σ is a collection of sequences of symbols from Σ one assigns a special meaning. Each member of the collection is called a *codeword*.

When all codewords have the same length, a so called *block code* is produced. Block codes are popular because they make the decoding process much more straightforward. In general, a block code C over Σ, whose codewords have length n and $|\Sigma| = q$, is called a *q-ary code of length n*.

Let $V_n(\Sigma)$, or if one wants to emphasize the size of alphabet Σ, $V_n(q)$, denote the metric space of all n-sequences of symbols from the alphabet Σ. Its elements are called *words* or *vectors*. The [*Hammimg*] *distance* $d(\boldsymbol{x}, \boldsymbol{y})$ between two words $\boldsymbol{x}$ and $\boldsymbol{y}$ in $V_n(\Sigma)$ is defined as the number of places where $\boldsymbol{x}$ and $\boldsymbol{y}$ differ. In the same sense, the *minimum distance* $d(C)$ of a code C of length n over Σ is defined as $\min d(\boldsymbol{c}_1, \boldsymbol{c}_2)$, where the distance is taken over all pairs $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ of distinct codewords in C.

If the codewords are scattered more or less evenly within the metric space $V_n(\Sigma)$, the case for what it is called a *good* code, the minimum distance of the code C will be maximized and *nearest-neighbor decoding* is then possible, that is, given a distorted codeword $\boldsymbol{y}$, a match is made to the codeword $\boldsymbol{c}$ which is at the minimum distance from $\boldsymbol{y}$ (when there are more than one codewords to match, one is arbitrarily chosen).

It is easy to show that for a code C of length n over Σ, the *nearest-neighbour decoding* scheme will correct up to $1/2\ (d-1)$ errors on a distorted codeword. A code C of length n over Σ that has M codewords and has a minimum distance $d,$ is denoted by (n, M, d). For a fixed n, the parameters M and d act in opposite directions: increasing M tends to decrease d and vice versa.

HADAMARD CODES

The binary case will be considered here where the alphabet Σ is the set $\{0,1\}$, the block code C is a subspace of the linear space F_2^n, F_2 being the 2-element field.

A *Hadamard matrix* of order n is a square matrix H_n whose elements are either 1 or -1, such that $H_n \times H_n^T = n\, I_n$, where H_n^T is the transpose of matrix H_n and I_n is the identity matrix of order n. A *Hadamard matrix* of order n, with $n > 2$, exists only if 4 divides n. Since $H_n \times H_n^T = n\, I_n$, any two distinct rows, or columns, of H_n must be orthogonal, and the matrix obtained from the permutation of rows, or columns, of H_n is also a Hadamard matrix, as is $-H_n$.

These matrices were introduced by the french mathematician Jacques Hadamard in 1893. Yet, despite much attention devoted by numerous mathematicians, the central question of existence of orthogonal $4n \times 4n$ matrices whose elements are either 1 or -1, for $n \in \mathbb{N}$, has not been answered.

There is a class of Hadamard matrices which are specially suited to building codes. They are called the *Hadamard-Sylvester matrices* and they can be built from H_2 by using the recurrent relation $H_{2n} = H_2 \otimes H_n$, for $n \in \mathbb{N}$, where $\otimes$ represents the *Kronecker* or *direct product* of two matrices. Thus, one has

$$H_2 = \begin{Vmatrix} 1 & 1 \\ 1 & -1 \end{Vmatrix}$$

$$H_4 = H_2 \otimes H_2 = \begin{Vmatrix} H_2 & H_2 \\ H_2 & -H_2 \end{Vmatrix} = \begin{Vmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{Vmatrix}$$

$$H_8 = H_2 \otimes H_4 = \begin{Vmatrix} H_4 & H_4 \\ H_4 & -H_4 \end{Vmatrix} = \begin{Vmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{Vmatrix} .$$

Consider next matrices $\widehat{H}_n$ and $\widehat{-H}_n$, which are obtained from Hadamard-Sylvester matrices H_n and $-H_n$ by changing the encoding alphabet from $\{-1,1\}$ to $\{0,1\}$, and form the $2n \times n$ matrix

$$\begin{Vmatrix} \widehat{H}_n \\ \widehat{-H}_n \end{Vmatrix}$$

the rows of which form the codewords of a binary $(n,\ 2n,\ n/2)$ code called a *Hadamard code*. Since n is a power of two, the code is linear (the sum of all codewords is still a codeword).

Hadamard codes of this class are represented by $[2^k, k+1, 2^{k-1}]_2$, where the first parameter is the block length, the second is the message length and the third is the code minimum distance. Hadamard code $[32, 6, 16]_2$ was used in 1971 by the *Mariner* 9 space probe to transmit pictures from Mars to Earth. This code can correct up to 7 errors on a distorted codeword and detect 8 through *nearest-neighbor decoding*.

The *encoding* process is described by

$$\boldsymbol{x} = \boldsymbol{m} \times \boldsymbol{G} = \| m_1, m_2, \cdots, m_{k+1} \| \times \boldsymbol{G} \quad ,$$

where $\boldsymbol{m}$ is the message, $\boldsymbol{x}$ its codeword and $\boldsymbol{G}$ the $(k+1) \times 2^k$ *code generating matrix*, defined by the vectors that form a basis of the code. For a Hadamard code $[8, 4, 4]_2$, $\boldsymbol{G}$ may be given by

$$\boldsymbol{G} = \left\| \begin{matrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{matrix} \right\| \quad .$$

The *decoding* process, on the other hand, is mostly based on the local decodability property of the Hadamard codes, that is, in order to determine the value of some of the message bits (1 up to k) only two codeword bits need to be tested. This stems from the fact that

$$\forall_{i,j \in \{0,1,\cdots,2^k-1\}} \; i \neq j \;\Rightarrow\; x_i \oplus x_j = \boldsymbol{m} \times \boldsymbol{g}_i \oplus \boldsymbol{m} \times \boldsymbol{g}_j = \boldsymbol{m} \times (\boldsymbol{g}_i \oplus \boldsymbol{g}_j) = \boldsymbol{m} \times \boldsymbol{g}_{i \oplus j}$$

due to the way matrix $\boldsymbol{G}$ was constructed. Hence, one gets

$$\forall_{l \in \{0,1,\cdots,2^k-1\}} \; \exists_{k \in \{0,1,\cdots,2^k-1\}} \; l \neq k \;\wedge\; l \oplus k = e_i \;\Rightarrow\; x_l \oplus x_k = m_i \quad , \qquad i \in \{1, 2, \cdots, k\} \; .$$

There are 2^{k-1} independent l, k pairs for each i. If one assigns the probability $1/2^{k-1}$ for each test and performs all the tests, one may assert the value of a particular message bit by a majority rule.

This procedure does not work for message bit m_{k+1} because it enters in the computation of all the codeword bits. Decoding is carried out here by directly performing nearest-neighbor approximation to the two possible values.

Design a digital circuit, called the *encoder*, which performs the message encoding for a $[8, 4, 4]_2$ Hadamard code (either for parallel or serial input). Design also a digital circuit, called the *decoder*, which extracts the received message, corrects a 1-bit error on the received codeword and detects 2-bit errors (either for parallel or serial input).

The assignment entails that some investigation should be made on finding the best possible algorithms for circuit implementation.

GRADING

- full specification of the *encoder* and proof of correctness of its design by VHDL simulation in Quartus – 13 valores
- full specification of the *decoder* and proof of correctness of its design by VHDL simulation in Quartus – 17 valores
- bear in mind that I am expecting both a parallel and a serial input version.

DELIVARABLE

- an archive, named `HAD_T$G#.zip` (where $, equal to 1, … ,4, means the lab number and #, equal to 1, …, 9, means the group number), of a directory `HAD_T$G#` containing two subdirectories, `COD` and `DEC`, with the VHDL files of your solution and a pdf file, up to 5 power point like pages, where the main ideas of the design are presented.

DEADLINE

- November, 13, at midnight.