Walsh Codes, PN Sequences and their role in CDMA Technology

Term Paper - EEL 201

Kunal Singhal, *2012CS10231*
Student, CSE Department, IIT Delhi

***Abstract*—Walsh codes are error correcting orthogonal codes and PN sequences are deterministically generated sequences which appear to be random noises. Both of these concepts are used in error free communication. This paper outlines properties of Walsh codes and PN sequences and discuss the practical implementation in software and hardware. The paper explores and model Walsh Codes and Hadamard matrices as a finite orthogonal vector space. It further describe their role in communication technologies in particular CDMA.**

***Index Terms*—Walsh code, Pseudo-Noise sequences, PN sequence, generator matrix, recursion, hamming weight, hamming distance, locally decodable, Hadamard matrices, linear codes, pseudorandom sequences, shift registers, spreading spectrum technique, CDMA, Cryptographic Secure PN codes**

I. Introduction

WALSH codes are mutually orthogonal error correcting codes. They have many interesting mathematical properties and vital applications in communication systems. In this paper, apart from the standard linear code model, we shall explore Walsh Codes from view point of a orthogonal vector space over $\mathbb{F}_2$. Pseudo random sequences play an important role in encoding of messages for efficient transmission of messages. Further, many encryption schemes uses pseudo random sequences. They are easily implemented in hardware as well as software, we give both the implementations in this paper. Then we shall in detail discuss the working of CDMA technology specific to Walsh Codes and PN Sequences.

II. Walsh Code

A. Definition

The *Walsh code* is a linear code[1] which maps binary strings of length n to binary codewords codewords of length 2^n. Further these codes are mutually orthogonal.

B. Encoding Walsh Code and Hadamard Matrices

A Hadamard matrix H of order n is an $n \times n$ matrix of 1s and -1s in which $HH^T = nI_n$. (I_n is the $n \times n$ identity matrix.) For Walsh codes, we use an Hadamard matrix of the order 2^N. Hadamard matrices are conjectured to exist for all orders which are multiple of 4. For, powers of 2, there is a constructive proof. *J.J. Sylvester* gave the following recursive construction in 1867:

$$H_1 = \begin{pmatrix}1\end{pmatrix} \quad H_2 = \begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}$$

$$H_{2N} = \begin{pmatrix}H_N & H_N\\ H_N & -H_N\end{pmatrix}$$

For correctness, we see that,

$$\begin{aligned} H_{2N}H_{2N}^T &= \begin{pmatrix}H_N & H_N\\ H_N & -H_N\end{pmatrix}\begin{pmatrix}H_N^T & H_N^T\\ H_N^T & -H_N^T\end{pmatrix}\\ &= \begin{pmatrix}H_NH_N^T + H_NH_N^T & H_NH_N^T - H_NH_N^T\\ H_NH_N^T - H_NH_N^T & H_NH_N^T + H_NH_N^T\end{pmatrix}\\ &= \begin{pmatrix}2I_N & 0\\ 0 & 2I_N\end{pmatrix} = 2I_{2N}\end{aligned}$$

Walsh codes can be generated from Hadamard matrices of orders which are a power of 2. The rows of the matrix of order 2^N constitutes the Walsh codes which encodes N bit sequences. Now, instead of 1 and -1 consider 1 and 0. That is consider the matrix and the codes over the field $\mathbb{F}_2$ or modulo 2.

Since Walsh codes is a linear code, there exist a generator or a transformation matrix for the same. Consider the following $n \times 2^n$ generator matrix for encoding bit strings of length n :

$$G_n = \begin{pmatrix}\uparrow & \uparrow & & \uparrow\\ g_0 & g_1 & \cdots & g_{2^n-1}\\ \downarrow & \downarrow & & \downarrow\end{pmatrix}$$

where $g_i \in \{0,1\}^n$ is the binary representation of i. Walsh Code $WC(x) = x \cdot G_n$ For instance, Generator matrix for $n = 2$ is given by

$$G_2 = \begin{pmatrix}0 & 0 & 1 & 1\\ 0 & 1 & 0 & 1\end{pmatrix}$$

Notice that. G_n is in fact a *basis*[2] for H_{2^n}.

Theorem II.1. *Any basis of the corresponding matrix can be taken as generator matrix.*

Proof: The code space i.e., H_{2^n} of Walsh code is an *n-dimensional*[3] *vector space of order* 2^n *over the field* $\mathbb{F}_2$. Consider B to be a basis of H_{2^n} then by the theory of vector spaces we have, $\forall y \in H_{2^n}\ \exists x \in \{0,1\}^n$ such that

$$y = \left(\langle x_i, z\rangle\right)_{z\in B}$$

Now, putting $WC(x) = y$ we get a Walsh encoding using B as the generator matrix. This proves the theorem. ■

[1] a code for which any linear combination of codewords is a new codeword

[2] basis of a vector space is the set of elements that span the whole vector space by their linear combination

[3] Dimension of a space is defined as the size of it's basis

C. Hamming Weight and Distance

The Walsh code for each string of length n has a hamming distance of 2^{n-1}. Further, the *Distance* or the *Edit Distance* between any two Walsh codes is also 2^{n-1}.

Theorem II.2. *Let $W(x)$ denote the* Hamming Weight *of a binary string/vector x. If x is a non-zero message in $\{0,\,1\}^n$, then $W(x \cdot G_n) = 2^{n-1}$.*

Proof: Let the i^{th} biy of x be non-zero. Pair up two column vectors in the generator matrix G_n if they differ only at i^{th} bit. Notice that this pairing, partitions the matrix into 2^{n-1} pairs. Now, if g and h are paired up then, exactly one of $x\cdot$ and $x \cdot h$ is 1. This is because they differ only in i^{th} bit and $x_i = 1$ thus, $g \cdot x \oplus h \cdot x = 1$. Since there are 2^{n-1} pairs, $W(x \cdot G_n) = 2^{n-1}$. ∎

Theorem II.3. *Let $\Delta(a, b)$ denote hamming distance of a and b. Then, $\Delta(x \cdot G_n, y \cdot G_n) = 2^{n-1}$*

Proof: Being a linear code, the difference of two codewords is also a codeword, implying the *Distance* of two Walsh codes is also 2^{n-1}. ∎

D. Orthogonality

As stated in the *definition*, Walsh codes are mutually orthogonal. And this in fact plays an important role in functioning of communication systems such as CDMA.

Theorem II.4. *Walsh codes are orthogonal*

Proof: If a and b are two Walsh Codes over the alphabet {-1,1} of length m then by **Theorem II.3**, $\Delta(a, b) = m/2$. This implies, $a \cdot b = m/2 - m/2 = 0$. Thus a and b are orthogonal. ∎

E. Decoding Walsh Sequence

Theorem II.5. *Any codeword of length m with up to $\frac{m}{4} - 1$ corrupted can be decoded.*

Proof: Let the given word be c Less than $\frac{m}{4}$ corruption
$\Rightarrow \exists a \in H_m,\ \Delta(a, c) < \frac{m}{4}$
$\Rightarrow \forall b \in H_n \neq a$
$\Delta(b, c) \geq \Delta(a, b) - \Delta(a, c) > \frac{m}{2} - \frac{m}{4} = \frac{m}{4}$ Thus, a is the only candidate for less than $\frac{m}{4}$ corruption and thus c can be decoded to inverse of a. ∎

1) Algorithm: The following is a *python* implementation of the standard algorithm [3]:

```
import random
y = raw_input() # the given codeword
n = int(raw_input()) #length of message
x = ""
for i in xrange(n):
    j = random.randint(0,pow(2,n)-1)
    k = j ^ pow(2,i)
    x += str(int(y[j])^int(y[k]))
print x
```

The above algorithm requires two bit uncorrupted for correctness namely, j and k. Thus, this Walsh code is $(2, \delta, \frac{1}{2} - 2\delta)$-*locally deodable* with this algorithm.

III. PN Sequences

A. Definition

A Pseudo random Noise Sequence is a binary sequence which though deterministically generated by a circuit or an algorithm appears to be statistically random like in the case of a fair coin flipping.

1) Traditional definiton: In 1967, Golomb gave the following three properties:

- **P1** Relative frequencies of 0_s and 1_s are each $\frac{1}{2}$.
- **P2** Run[4] lengths are as expected in a coin flipping, i.e., $1/2^n$ of all the runs would be of length n.
- **P3** If the sequence is shifted by any non zero numbers of bits, then the *relative hamming distance*[5] between the two sequences would be half.

(**P1** is known as *Balanace* and **P2** is known as *Run*.) Any sequence which follow the above three properties within extremely small discrepancies can be called PN sequence.

2) Formal Definition: PN sequences are defined with the help of a polynomial of degree n:

$$P(x) = \sum_{i=0}^{n} a_i x^i$$

with $a_i \in \mathbb{F}_2$ and $a_n = a_0 = 1$. The PN Sequence corresponding to this will satisfy the following recursion:

$$p_{n+k} = \sum_{i=0}^{n-1} a_i p_{k+i}$$

over the field $\mathbb{F}_2$. Note that the $p(x) = 0$ is the characteristic equation of the mentioned recursion.

B. Correlation

Similarity between two sequences is denoted by correlation. Qualitatively, when two sequences compare different, they are said to have *cross correlation* and when same then *autocorrelation*. Quantitatively, correlation of two sequence x and y as a function of time delay i is:

$$r(i) = \sum_{k=0}^{L-1} x_k \cdot y_{k+i}$$

As stated in [2], the correlation function r(i) of any PN sequence of length N is given by

$$r(i) = \begin{cases} 1 & i = 0 \\ -\frac{1}{N} & 1 \leq |i| \leq N-1 \end{cases}$$

C. Generation and Different PN sequences

Most of the PN sequences are easily implemented in hardware as well as software. In fact, they are usually categorised on the basis of generation method.

[4] A run is a contiguos sequence of same bit i.e., 0 or 1 like 0000 or 111

[5] relative hamming distance refers to the ratio of hamming disance of the two vectors over the length of vectors

1) Using shift registers with feedback: **Maximal Length Sequences or m-sequence** is the class of PN sequences which can be generated by this method. A shift register for length n can generate a maximal length sequence of $2^n - 1$ bits. Consider the following circuit characteristic polynomial:

$$P(x) = 1 + x + x^4$$

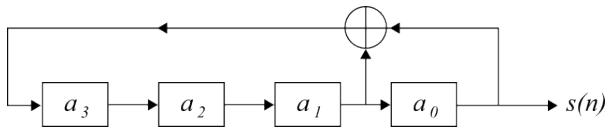

Fig 1 [1]

Following is the software implementation for the same PN Sequence in python:

```
A = input() # first 4 bits input
LEN = 15
for i in xrange(LEN):
    print A[0],
    A = [A[1], A[2], A[3], A[0]^A[1]]
```

2) Gold Sequences: A gold sequence is composed with two Maximal Length sequence generator. The composition is taken using a mod 2 addition or a xor operation. Following is the circuit for the same:

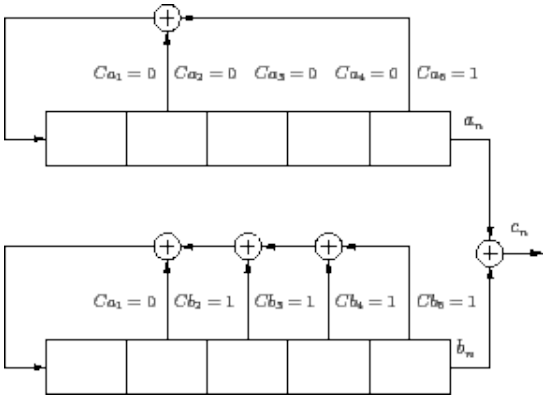

Fig 2 [1]

The python software implementation for the same is as follows:

```
A = input() # first 5 bits input
B = input() # first 5 bits input
LEN = 31
for i in xrange(LEN):
    print A[0]^B[0],
    A = [A[1],A[2],A[3],A[4],A[0]^A[1]]
    B = [B[1],B[2],B[3],B[4],B[0]^B[1]]
```

3) Baker Sequences: Baker sequences are relatively short length codes and have good correlation properties i.e, the autocoreolation function takes very small value. Apart from uniformly low autocorrelation values, these sequences have *balance* and *run* properties. The one drawback that they face is the small period of the sequence.

D. PN Sequences in Cryptography

From Shannon's theory of Cryptography we know that perfect security cannot be achieved unless we use large random keys. But, we use PN sequences, we can generate a pseudo random sequences of lengths of the order 2^n from a key of b bits. But, we cannot use any PN Sequence, because there are many predicting algorithms which exploit the deterministic aspect of PN Sequence and decode them partially or completely. So, a PN Sequence should Cryptographically secure. One of the factor that contributes is low correlation. Theoretically, a PN is said to be secure if given a set of bits of the sequence, probability of predicting the next bit is less than $\frac{1}{2} + \varepsilon$ where ε is negligible.

IV. Code Division Multiple Acess (CDMA)

In the last few decades there has been an drastic increase in global use of internet and mobile phone services. One of the biggest challenge faced in communication is to cater to large number of demands at the same time with a limited bandwidth of frequency over which transmission lines and devices can actually operate. Traditional solution this problem like Time Division Multiple Access or Frequency Division Multiple Access are quite slow and unstable. CDMA on the other hand uses the same frequency channel at the same time to send data to many users without any hinderance.

A. Spreading Spectrum Technique

In CDMA, data for every user is encoded by multiplying it with a PN Sequence more precisely, a Gold Sequence. And all the data is sent over the same line. Since the frequency of data is much less than the PN sequence used, users are able to decode the data back by multiplying the encoded message by the same PN sequence.

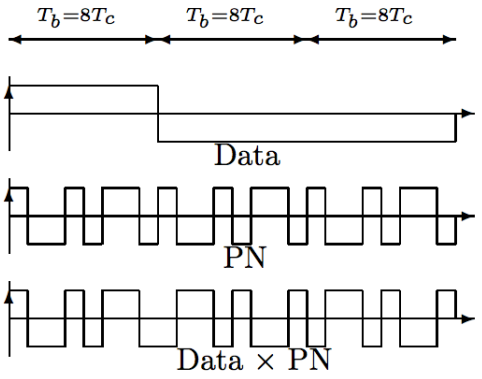

Fig 3 [10]

But, that is not the end of it. Whenever we transmit data over large distances there is bound to some additional noise corrupting the message.

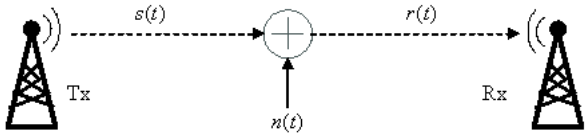

Fig 4 [11]

Let the message as a function of time t be $m(t)$ and PN code be $p(t)$, then we transmit $s(t) = m(t) \cdot p(t)$. But over transmission, a noise $n(t)$ gets added to the signal so the receiver gets

$$r(t) = s(t) + n(t) = m(t) \cdot p(t) + n(t)$$

Thus, decoded message $M(t) = r(t) \cdot p(t)$ is :
$r(t) \cdot p(t) = m(t) \cdot p(t) \cdot p(t) + n(t) \cdot p(t)$
$\Rightarrow M(t) = m(t) + N(t)$ This can be achieved in in two ways, *Forward Link* and *Reverse Link*.

B. Forward Link

Forward link is synchronous and thus sometimes it also known as synchronous CDMA. Messages are also multiplied with Walsh codes which are unique for every mobile user for a particular base station. Every user on receiving a message stream, multiplies it with the PN code as well as it's own unique Walsh Code.

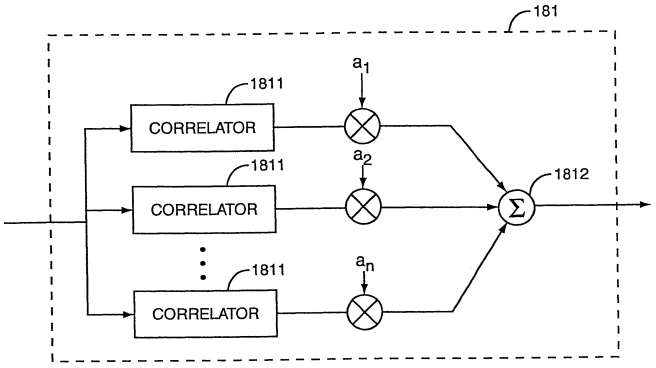

Fig 4 [12]

Walsh codes of all the users are orthogonal so, one can only decode it's own message and thus avoid any kind of hinderance. For instance, consider a user with Walsh code a receives a message encoded with b then,

$$a \cdot (m \cdot b \cdot PN) = (a \cdot b) \cdot m \cdot PN = 0 \cdot m \cdot PN = 0$$

which is precisely what we desire. Further, since Walsh codes are error correcting codes, noise $N(t)$ can also be remove if the bit corruption is low.

C. Reverse Link

Due to portability of cell phones, synchronisation in not feasible. And without synchronisation, Walsh codes lose their property of orthogonality. But, we still have low correlation property of PN codes.

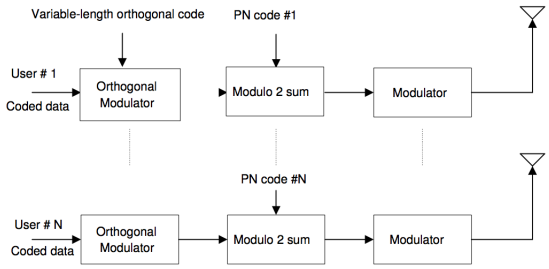

Fig 4 [13]

So, all the users messages are coded with different PN codes. Walsh can still be employed as error corrected codes. But, due to many PN codes, there is an additional noise which requires modulation and hence extra power.

V. Conclusion

Walsh codes encodes n bit messages into 2^n bit orthogonal codewords. Original message can be recovered even after about one-fourth of the bits have been corrupted.

PN Sequence are statistically random sequences with low correlation property. There are simple electronic circuits which can generate these sequences very quickly. These sequences are widely used in communication and cryptography.

There is a subtle interplay of Walsh Codes and PN Sequences in transmission of data via **CDMA** scheme. CDMA is implemented in two ways asynchronous and synchronous. Though, the asynchronous way serves the needs of mobile cell phones, it is comparatively more power consuming.

References

[1] *Analysis of Different Pseudo Noise Sequences*, by Alka Sawlikar, Manisha Sharma (International Journal of Computer Technology and Electronics Engineering)

[2] *On the Properties of Pseudo Noise Sequences with a Simple Proposal of Randomness Test*, by Abhijit Mitra (International Journal of Electrical and Computer Engineering)

[3] *Hadamard Code*, Wikipedia (http://en.wikipedia.org/wiki/Hadamard_code)

[4] *Introduction to Coding Theory - Lecture Notes*, by Yehuda Lindell (Computer Science Department, Bar Ilan University, Israel)

[5] *Hadamarad Codes*, by Massoud Malek (http://www.mcs.csueastbay.edu/∼malek/TeX/Hadamard.pdf)

[6] *Thoughts on inverse orthogonal matrices.* by J.J. Sylvester. (Philosophical Magazine, 34:461475, 1867)

[7] *Hadamard Matrices and Hadamard Codes*, by Leon, University of Illinois at Chicago (http://homepages.math.uic.edu/simleon/mcs425-s08/handouts/Hadamard_codes.pdf)

[8] *CDMA: Principles of Spread Spectrum Communication,* by A. J. Viterbi, Addison-Wesley Publishing Company, 1995.

[9] *Shannon's Theory of Cryptography,* by R. Vijay Shankar (http://www.cse.iitm.ac.in/∼theory/tcslab/cryptpage/report1.pdf)

[10] *Spread Spectrum, Cryptography and InformationHiding,* by Laurent Dubreuil and Thierry P. Berger (Universite de Limoges, Mathematics Department)

[11] *Understanding Spread Spectrum for Communications,* White Pages (http://www.ni.com/white-paper/4450/en/)

[12] *Fast forward link power control for CDMA system* Sourour, Essam (Cary, NC), Atarius, Roozbeh (Morrisville, NC), Khayrallah, Ali (Apex, NC), United States Patent 6768727

[13] *Spreading Codes in CDMA Detection* by Ayse Kortum, Eastern Mediterranean University