Cours Codage de l'information

Mendy Fatnassi

10 décembre 2020

Table des matières

Chapitre 1 : Codage Physique / Codage sans perte

1.1 Introduction

C'est en 1948 grace aux travaux de shannon , que la theorie de l'information a pris sa forme actuelle.
Le traitementdu contenue d'une source d'information peux etre envisage sous deux forme :
-Sans perte d'information.
-Avec perte d'information.

1.2 Transmissions Avec/Sans perte

Avec perte :
Exemple : Signal analogique stereo , bande de frequence 0 a 20 Khz , echantillonnage 44,1Khz , quantification de 16b.
Donc $44.1 \times 10 + 3 \times 16 \times 2 = 1.411$ MBits/sec
Ex MP3 : reduction a 128 Kbits/sec avec un taux de compression de 11.

Taux de compression :
-Elimination des composantes spectrales.
-Utilisation du codage de Huffman.

La transmission numerique passe par un support physique qui interprete la communication sous forme de signaux numerique.Ainsi des donnees analogique devront prealablement erte numerise.Le codage du signaux peux se faire sur 2 (-X,X) ou 3 niveaux (-X,0,X).

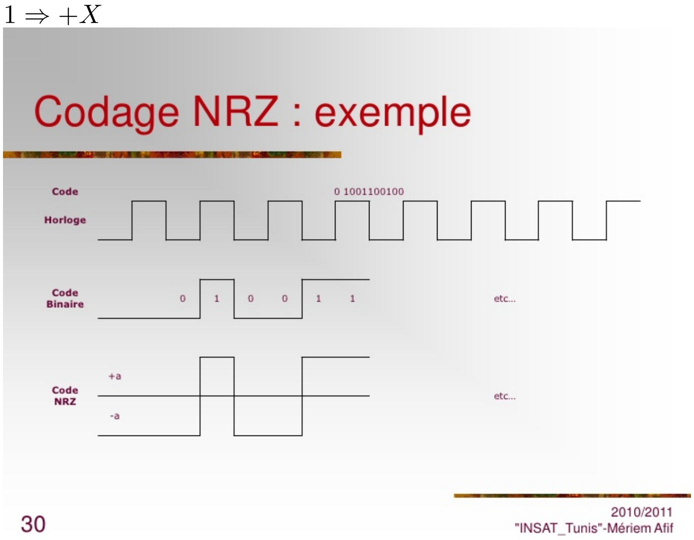
1.3 Mode de transmission/Codage

Voici quelque mode de transmission :

1.3.1 NRZ (No Return to Zero)

Avantage : Le recepteur peux determiner la presence ou non d'un signal.
Inconvenient : Difficulté de synchronisation.

$0 \Rightarrow -X$
$1 \Rightarrow +X$



Quelque derivée du NRZ : RZ(Return to Zero) & NRZI (Inverted).

RZ

:
pareil que pour NRZ sauf que entre deux bits 1 consecutif il y a un changement d'etat au lieu de reste sur +X on feras +X-X+X-X.

NRZI

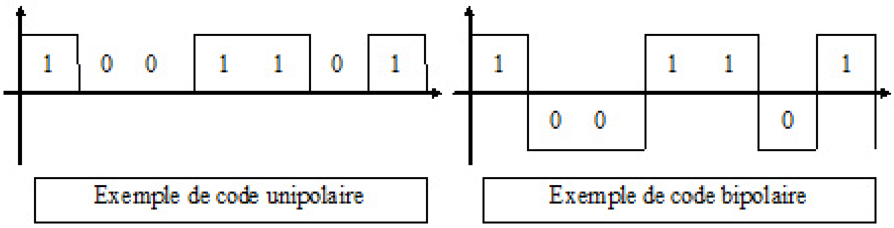
:
1 $\Rightarrow$ le signal change d'etat , 0 $\Rightarrow$ aucun changement d'etat.

1.3.2 Unipolaire

Avantage : Reduction encore plus significative du spectre.
Inconvenient : Impossibilite de distinguer une suite de 0 et l'absence d'information.
0 $\Rightarrow$ 0 (tension nulle)
1 $\Rightarrow$ +X

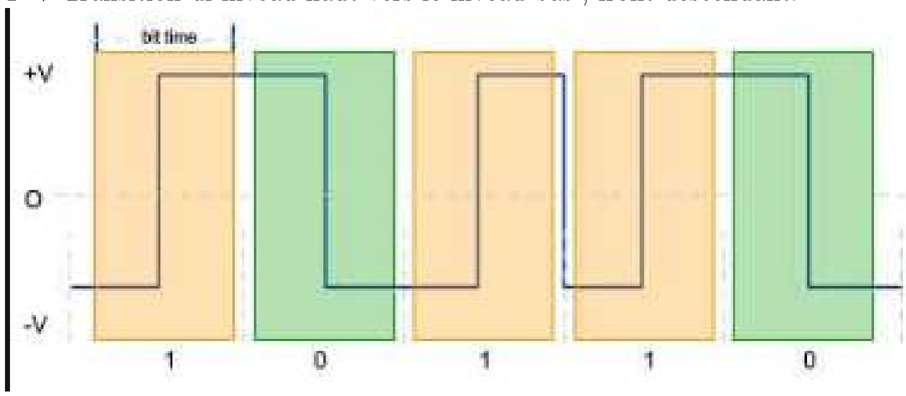
1.3.3 Bipolaire

:
0 ⇒ lorsque le bit est a 0
1 ⇒ alternativement +X et -X



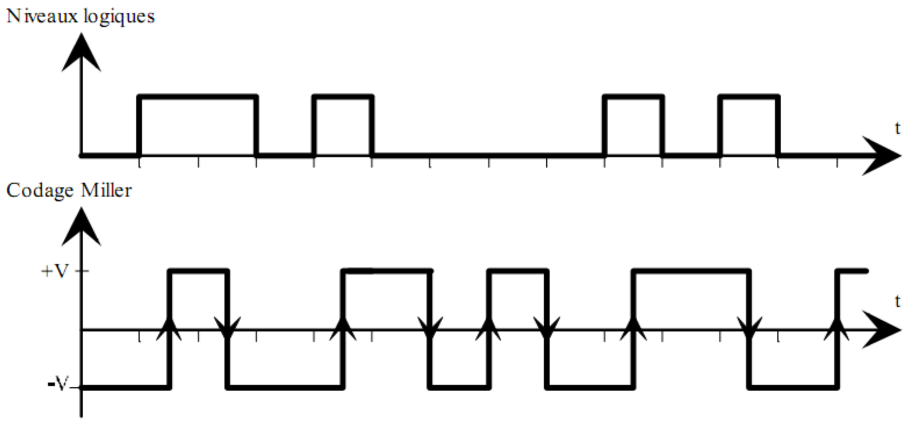
1.3.4 Manchester

Une transition est introduite au milieu de l'intervalle significatif.
0 ⇒ Transition du niveau bas vers le niveau haut ,front motant.
1 ⇒ Transition di niveau haut vers le niveau bas , front descendant.



1.3.5 Delay Mode (Miller)

Signal intermediaire identique au coodage manchester , puis suppression d'une transition sur deux.
Transition (front montant|descendant) ⇒ 1
Pas de transition au milieu du bit 0
Transition en de fin de bit 0 si suivi d'un autre 0



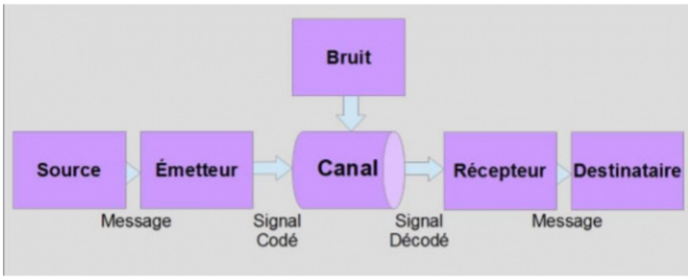
1.3.6 HBDN

:

1.4 Entropie

La theorie de l'information mené par Shannon , qui est une théorie probabiliste permettant de quantifier le contenu moyen en information d'un ensemble de messages, dont le codage informatique satisfait une distribution statistique précise. Ce domaine trouve son origine scientifique avec Claude Shannon qui en est le père fondateur avec son article "A Mathematical Theory of Communication" publié en 1949.
Parmi les branches importantes de la théorie de l'information de Shannon, on peut citer :
-Le codage de l'information
-La mesure quantitative de redondance d'un texte
-La compression de données
-La cryptographie

Paradigme de Shannon :



L'entropie permet de mesurer la quantité d'information moyenne d'un ensemble d'v́enements (en particulier de messages) et de mesurer son incertitude. Supposons maintenant que les boîtes soient de diverses couleurs : n1 boîtes de

couleur C1, n2 boîtes de couleur C2. . . , nk boîtes de couleurs Ck, avec n1 + n2 + . . . + nk = N. La personne C sait de quelle couleur est la boîte recherchée. Quel est le prix de cette information ? .
On la note H(I) : $\sum_{i \in I} pilog_2pi = -(p_ilogp_i + ... + p_{i+n}logp_{i+n})$
avec $pi = \frac{ni}{N}$ la probabilité associée à l'apparition de l'évènement i.
Donc l'information « la boîte est de couleur C1 » vaut log N/n1, et cette éventualité a une probabilité n1/N.

Entropie d'une source discret (numerique :0,1) , si on a un message aleatoire avec un tirage de 27 lettres (alphabet + espace) , on aurait comme entropie H=log27=4,75 bits.
Entropie relative , elle dependra du contexte de son message (limité par le corpus de la langue) et auras un pourcentage de liberté

1.5 Codage Source

Il existe des code a longueur fixe ou tous les mots ont la meme longueur et possede le meme nombre de symbole et les code a longueur variables ou la longueur varie en fonction de leur frequence d'apparition .Un mot sera d'autant plus long que sa probabilite d'apparition est petite.

1.5.1 Code de Shannon-Fano

Deroulement de l'algo :
1-Trier les probabilité par ordre descroissant
2-On separe l'ensemble en 2 groupes (de somme a peux pres egale)
3-$\sum 1 = \sum 2 \Rightarrow \sum 1- > 0 \sum 2- > 1$
4-On divise en 2 les deux sous-groupe $\sum 1/2 = \sum 11- > 0, \sum 21- > 1$ et $\sum 2/2 = \sum 21- > 0, \sum 22- > 1$
5-Bouclez de l'etape 4 a 5 jusqu'a tant de que le tableau soit remplie

Exemple :

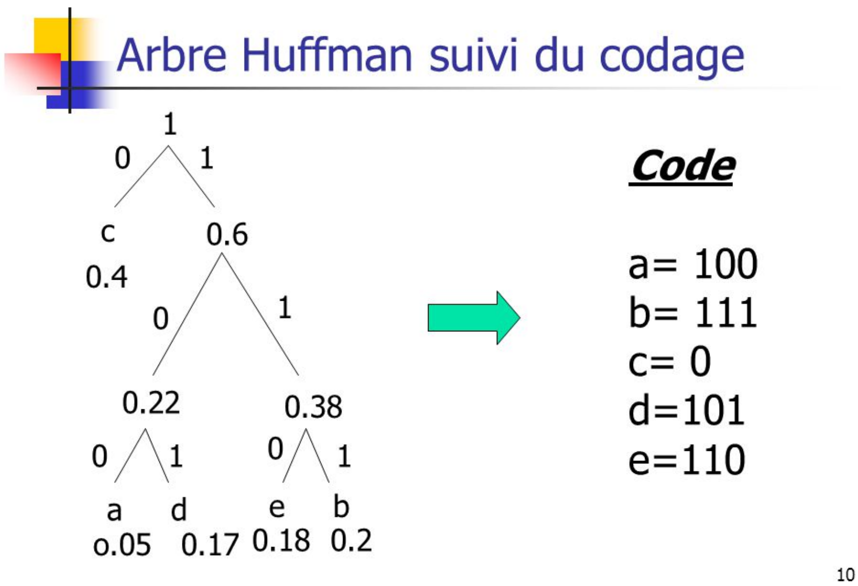
<table>
<tr><th>a_i</th><th>$p(a_i)$</th><th>1</th><th>2</th><th>3</th><th>4</th><th>Code</th></tr>
<tr><td>a_1</td><td>0.36</td><td rowspan="2">0</td><td colspan="3">00</td><td>00</td></tr>
<tr><td>a_2</td><td>0.18</td><td colspan="3">01</td><td>01</td></tr>
<tr><td>a_3</td><td>0.18</td><td rowspan="4">1</td><td colspan="3">10</td><td>10</td></tr>
<tr><td>a_4</td><td>0.12</td><td rowspan="3">11</td><td colspan="2">110</td><td>110</td></tr>
<tr><td>a_5</td><td>0.09</td><td rowspan="2">111</td><td>1110</td><td>1110</td></tr>
<tr><td>a_6</td><td>0.07</td><td>1111</td><td>1111</td></tr>
</table>

On a bien 2groupe qui ont des somme a peux pres egale avec a1,a2 > a3,a4,a5,a6 (0.54>0.46), la somme des a vaux bien 1.

1.5.2 Code de Huffman

Deroulement de l'algo :
1-Trier les probabilite par ordre croissant
2-creer un noeud parent a partir de 2 lettres sources avec les probabilite les plus faible
3-Affecté au noeud parent la somme des 2 noeud fils
4-Supprimer les noeud enfant
5-Bouclez de l'etape 2 a 5 jusqu'a ce que l'arbre soir remplie

Exemple :



Bonus : Code LZW (Lempel-Ziv-Welch) utilisé pour la compression des données.

1.5.3 Code Arithmetique

Chapitre 2 : Code Correcteur d'erreur / Code d'etalement

2.1 Introduction

Un code correcteur d'erreur peux etre representer sous forme matriciel.Lors de la transmission numerique d'un signal il peux y avoir des causes d'erreur due a du bruit(thermique,impulsif,intermodulation,...),de la distortion,de l'echo ou du canal hertzien etc... .En general on ne peux agir sur le canal de transmission ou preferera se preoccuper sur le signal num. pour des contraintes tehchnologique et economique.

On peux distinguer 2 types d'erreur :
-Les erreurs aleatoire due aux bruits blanc.
-Les erreurs par paquets (impulsif,parasite).

<u>Definition</u>
bruit blanc : On appelle la lumière blanche la superposition des ondes visibles du spectre. Et bien lorsqu'on superpose les fréquences audibles on obtient le bruit blanc. C'est un assemblage de plusieurs sons qui donnent un seul son uniforme.

2.2 Code Correcteur

Il existe 2 type de code :
-Code en bloc
-Code convolutionnel

Un code est appele $C(n, m)$, la redondance est definie par : $\frac{m}{n} = 1 - \frac{k}{n} < 1$

2.2.1 Code en Bloc

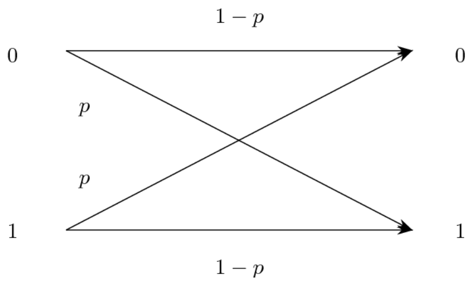
Le message est decouper en bloc de $meb(elementbinaire)$ de longueur fixe (m constant).A chaque bloc de meb, le codeur ajoute keb de controle (appele checksum dans le cas du reseau) .On Vas ecrire $n = m+k$, n : $m+k$ (n composé de m+k) $\Rightarrow$ code séparable ou systématique.

2.2.2 Code Convolutionnel

Des éb (echantillon binaire) de controle sont introduit de maniere continue dans le message utile .éb bloc : ajouter k eb de controle , depend du bloc de controle et anterieure $\Rightarrow$ code recurrent.

2.3 Detection et Correction d'erreur

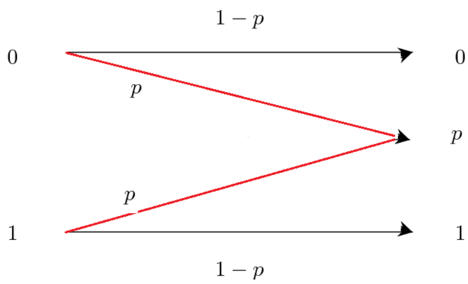
Cas d'un canal sans symbole d'effacement.



Code lineaire C de distance minimum dm permet de :
-Detecter au plus $dm-1$ erreur
-Corriger au plus $t=\frac{dm-1}{2}$ erreur (entier)

On dit qu'on code est lineaire systematique .(lineaire : informtion envoyee et messages recus , systematique : envoie bits de controle)

Cas d'un canal avec symbole d'effacement :



-Detecte au plus $\rho+1\leq dm$
-Corrige $2t+\rho+1\leq dm$

Code Parfait :
Une condition necessaire est suffisante pour qu'un code lineaire Ct correcteur et de distance minimum dm soit parfait :
$\sum_{i=0}^{t} C_n^j(q-1)^j=q^k$

2.4 Generation et Detection

La **matrice generatrice** est utiliser dans le cas des code lineaire ,exemple : Avec n=4 et m=2 ou (n :nombre de mot , m :la longueur du mot)

On a g1= $\begin{pmatrix}1\\0\\0\\0\end{pmatrix}$ et g2 = $\begin{pmatrix}0\\1\\0\\0\end{pmatrix}$

U= $\begin{bmatrix}1 & 0\\0 & 1\\0 & 0\\0 & 0\end{bmatrix} \times \begin{pmatrix}x\\y\end{pmatrix}$

Le mot a coder est de longueur 2 : 00,01,10,11
On vas ensuite prendre chaque mot et effectuer le produit matriciel par U :

$$\begin{bmatrix}1 & 0\\0 & 1\\0 & 0\\0 & 0\end{bmatrix} \times \begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}1\\0\\0\\0\end{bmatrix}$$

Ducoup le mode coder est : 0000 , 0100 , 1000 ,1100
On peux ecrire le poid dm (nb de 1) de chaque mot coder : 0 , 1 , 1 ,2 ici dm=1 (poid le plus faible)
On a donc dm-1 = 0 Detection
Et $\frac{dm-1}{2}$ Correction

Un code lineaire systematique peux etre representer sous forme d'un systeme (m :message) :
c1 = m1
c2 = m2
$-------$
c3 = m1+m2
c4 = m1
c5 = m1+m2

Selon le systeme du dessus on peux representer la matrice generatrice de cette facons :
G =

$$\begin{array}{c} \\ c1 \\ c2 \\ --- \\ c3 \\ c4 \\ c5 \end{array} \begin{array}{c} m1\ m2 \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \hline 1 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ m1 \\ m2 \\ --- \\ m1+m2 \\ m1 \\ m1+m2 \end{array}$$

Matrice de controle :

Elle se note : H = $\begin{bmatrix} -P^T \\ I \end{bmatrix} \Rightarrow H^T = \begin{bmatrix} P \\ I \end{bmatrix} = \begin{bmatrix} 1 & 1 & \| & 1 & 0 & 0 \\ 1 & 0 & \| & 0 & 1 & 0 \\ 1 & 1 & \| & 0 & 0 & 1 \end{bmatrix}$

G$^T H = 0$

Syndrome :
C est un mot codé , alors $H^T C = 0$
y est un mot recu , syndrome de y : $s(y) = H^T y$
$y = C + e$ avec e :erreur [0010] (on met un 1 sur le bit errone)
Donc $s(y) = s(C+e) = H^T C + H^T e = H^T e$

Procedure :
-Calculer $s(y) = H^T y$
-Si $s(y) = 0$ alors y est le mot code
-Sinon on cherche une sequence Z de longueur n telle que $H^T Z = s(y)$, $C = y + Z$

2.5 Sequence d'hadamard

Elle se definit comme suit et est de taille minimal 8 : $\begin{bmatrix} H & H \\ H & -H \end{bmatrix}$

On commence a H0= $\begin{bmatrix} 1 \end{bmatrix}$
H1 = $\begin{bmatrix} [1] & 1 \\ 1 & -1 \end{bmatrix}$

$$H2 = \begin{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \end{bmatrix}$$

$$H3 = \begin{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \\ \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} & \begin{bmatrix} -1 & -1 & -1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 \end{bmatrix} \end{bmatrix}$$

Chaque ligne de la matrice correspond a une sequence , il y en a donc 8 par defaut .
Si un utilisateur U1 veux transmettre la sequence S1 "101" en utilisant la sequence 2 (2ieme ligne de la matrice) , on auras donc :
S1 : "1" seras code par la 2ieme lignes donc "1 -1 1 -1 1 -1 1 -1" et "0" par son inverse "-1 1 -1 1 -1 1 -1 1"
S1 = "101" $\Rightarrow$ $"1-11-11-11-1"" -11-11-11-11""1-11-11-11-1"$

$SiU2veuxtransmettre"011"onluiattribueraslasequence4parexemple(4iemelignedeH), onaurasdonc:$

$S2 = "011" \Rightarrow " -111-1-111-1""1-1-111-1-11""1-1-111-1-11"$

$MaintenantsiU1etU2veulentemettreenmemetemps, ilfaudrafaireuneoperationd'etalementS:$

$S = S1 + S2$
$S = "002-2002-2""00-2200""2-2002-200"$

Chapitre 3 : Code Cyclique/ Pseudo-Aleatoire

3.1 Code Pseudo Aleatoire

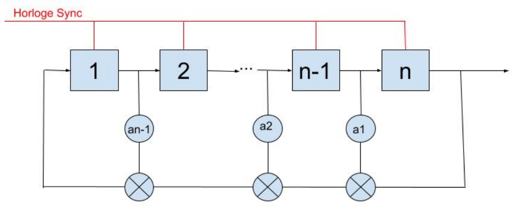
Un nombre n'etant pas aléatoire , on vas piocher une partie d'une tres grande sequence binaire .Cette sous-sequence seras genere de tel sorte qu'il ne seras presque impossible de trouver la meme sous-sequence.
La Sequence doit etre :
1) Generer facilement (seq. binaire 0,1)
2)Suffisament longue
3)Difficilement reconstituable par petit segment
4)Distribution des eb qui apparait de facons aleatoire (evite les meme suite de 0 et 1 ex : 001 01 001 1 001 001)

Solution : utiliser code a longueur maximal (LM) , code Gold ou JPL.

3.1.1 Code a Longueur Maximal LM

La fonction c , rand() utilise un code LM.

Principe : Il s'agit de registre a decalage en reaction lineaire composant n etages.

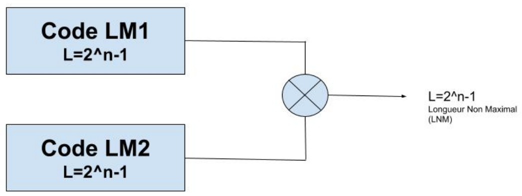


Le schema peux representer un polynome de second degre , ce systeme permet d'avoir autant de 0 que de 1 (Distributivite) , $L = (2^n - 1)$ bits (L :taille)

L'addition mod(%) 2 d'une sequence LM avec une sequence décaler/repliqué sur elle meme donne $\Rightarrow unesequenceLMrepliqu$

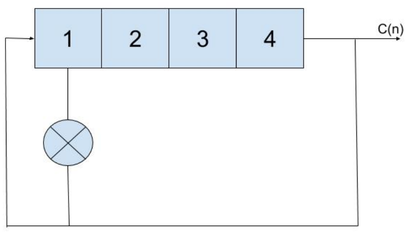
3.1.2 Code Gold

Implemente le code LM mais permet de generer des sequence encore plus longue . Utilise dans le systeme CDMA (AMRC en anglais).



Inutile si on prend 2 fois le meme sequence car LM1 et LM2 sont demarrer et initialiser en meme temps .

Exemple : **Registre a decalage** de longueur 4



On preleve sur le 1er et 4ieme element [4,1].

-Initialisation : cela peux etre fait avec des valeurs quelconque de la sequence "0000" , en generale on initialise a "1111".

Principe :
1)Decalage de la sequence de 1 rang vers la droite
2)Xor [4,1] : On fait un ou-exclusif sur le 1er et le 4ieme element jusqu'a temps de retomber sur la sequence initial "1111"

$$\begin{array}{ccccc}
Cn & 1 & 1 & 1 & 1 \\
C(n+1) & 0 & 1 & 1 & 1 \\
C(n+2) & 1 & 0 & 1 & 1 \\
C(n+3) & 0 & 1 & 0 & 1 \\
C(n+4) & 1 & 0 & 1 & 0 \\
C(n+5) & 1 & 1 & 0 & 1 \\
\\
C(n+6) & 0 & 1 & 1 & 0 \\
C(n+7) & 0 & 0 & 1 & 1 \\
C(n+8) & 1 & 0 & 0 & 1 \\
C(n+9) & 0 & 1 & 0 & 0 \\
C(n+10) & 0 & 0 & 1 & 0 \\
C(n+11) & 0 & 0 & 0 & 1 \\
C(n+12) & 1 & 0 & 0 & 0 \\
C(n+13) & 1 & 1 & 0 & 0 \\
C(n+14) & 1 & 1 & 1 & 0 \\
C(n+15) & 1 & 1 & 1 & 1
\end{array}$$

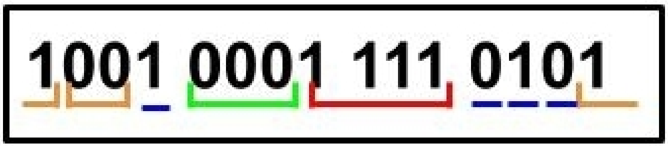
$L = 2^n - 1 = 15(onprendpasla1erligned'initialisation$

La lecture des Cn se lis en colonne , Cn commenceras donc par la premier ligne et C(n+6) respectivement a la 6ieme lignes d'ou l'espacement de le systeme, ce qui donne :

C(n) : 1111 0101 100 1000
C(n+6) : 0110 0100 011 1101
C(n) xor C(n+6) = 1001 0001 111 0101

Propriete : -Il n'y a aucune serie de "0" de longueur R
-Il y a qu'une serie de "1" de longueur R
-Il y a qu'une serie de "0" de longueur R-1
-Il n'y a aucune serie de "1" de longueur R-1
-Il y a 2^{R-P-2} serie de "0" de longueur P
-Il y a 2^{R-P-2} serie de "1" de longueur P

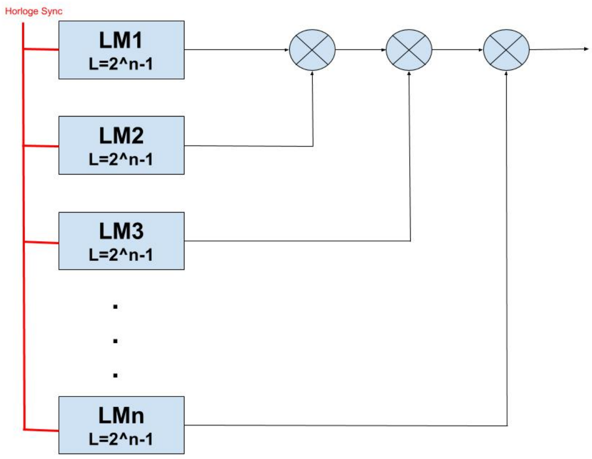
Si on prend :
R=4 : On a 0 serie de "0000" et 1 serie de "1111" (rouge)
R-1=3 : On a 1 serie de "000" et 0 serie de "111" (vert)
P=2 : On a $2^{4-2-2} = 2^0 = 1$ donc 1 serie de "00" et 1 serie de "11" (orange)
P=1 : On a $2^{4-1-2} = 2^1 = 2$ donc 2 serie de "0" et 2 serie de "1" (bleue)



Note : Pour avoir les meilleur resultat , des chercheurs ont trouve un polynome qui possede assez de combinaisons pour etre dit fiable , on parle du polynome 89 [89,6,5,3]

3.1.3 Code JPL

Implementation de plusieur code LM avec une taille differente (nmp) , ce systeme est utilisé en radio-localistion (GPS).



LM1,LM2 et LM3 doivent etre premier entre eux .

3.2 Code Cyclique

Code le plus utilise dans de nombreux systeme.Les code cyclique sont detecteur et correcteur d'erreur .Les bits de controle sont le reste de la division polynomial.

Exemple :
Determinez si le code engendre par la matrice est cyclique ? Avec G= 1 1 0 0
0 1 1 0
0 0 1 1

On vas effectuer le produit matriciel avec une matrice M
M= 0 0 0
0 0 1
0 1 0
0 1 1
1 0 0
1 0 1
1 1 0
1 1 1

On a donc $M \times G = C$
C= 0 0 0 0
0 0 1 1
0 1 1 0
0 1 0 1
1 1 0 0
1 1 1 1
1 0 1 0
1 0 0 1

Si on decale une ligne sur la gauche on retombe sur une autre ligne de la matrice , C est cyclique si tout les lignes sont decalable

Chapitre 4 : Temps Reel