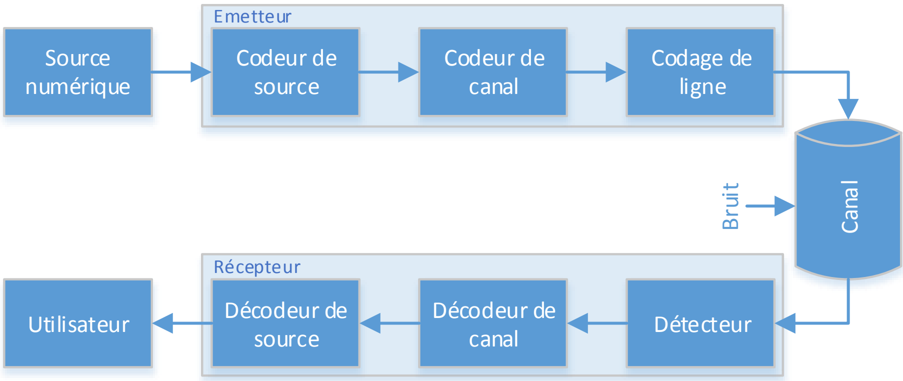
RESUME DE THEORIE DE L'INFORMATION

TI – Résumé

I. Définitions générales

- **Source X :** Système capable d'émettre des séquences de signes parmi un alphabet de taille N
- **Entropie H :** Degré de surprise d'un message, quantité d'information contenue dans un signe, nombre moyen de bits nécessaires pour coder un signe.
- **Information mutuelle I :** Info contenue dans un message
- **Codage de source :** compression des données d'une source (éliminer redondance inutile)
- **Codage de canal :** augmenter la fiabilité du canal (ajout de redondance utile)
- **Codage de ligne :** signal physique envoyé sur le canal
- **Modulation :** Adaptation du spectre du signal au canal sur lequel il est émis



II. Entropie et information

1. Notation des probabilités

- $P_i = \mathbb{P}(X = x_i)$
- $P_{ij} = \mathbb{P}\big(X = x_i, Y = y_j\big)$
- $P_{i|j} = \mathbb{P}\big(X = x_i \,|Y = y_j\big)$ et $P_{j|i} = \mathbb{P}\big(Y = y_j \,|X = x_i\big)$
- $P_{i|seq} = \mathbb{P}(X_t = x_i \,|X_{t-1} \dots X_{t-m} = seq)$
- $P(x_t x_{t+1}) = P(x_{t+1}|x_t)P(x_t) \qquad P(011) = P(01|11)P(11)$

2. Unités selon la base du log

Base 10 : Harley Base e : Nit Base 2 : Shannon

3. Définitions et propriétés

- Propriétés de l'entropie d'une source composée :
  - $H(XY) = H(YX)$
  - $H(XY) = H(X) + H(Y|X)$
- Propriétés de l'entropie conditionnelle :
  - $H(Y|X) = H(Y)$ si X et Y indépendants
  - $H(X) + H(Y|X) = H(Y) + H(X|Y)$
- Entropie d'une extension : $H(X^q) = H\big(X_1 \dots X_q\big)$
- Bornes de l'entropie : $H(X) \in [0; \log N]$
- Entropie limite : $H_\infty(X) = \lim_{q\to\infty} \frac{1}{q} H(X^q) = \lim_{q\to\infty} H\big(X_q\big|X_1 \dots X_{q-1}\big)$
- Entropie de la qième extension : $H\big(S_q\big) = H(X_1) + H(X_2|X_1) + \dots + H\big(X_q\big|X_1 \dots X_{q-1}\big)$

Attention, l'extension d'une source veut dire qu'on a $S_2 = S_t S_{t-1}$, ce qui fait qu'on ne produit à chaque fois qu'un seul signe nouveau, et que S_{t-1} prend la valeur précédente de S_t.

4. Entropie de sources

a. Source sans mémoire : signes indépendants (modèle de Bernoulli)

Entropie de la source équiprobable	$H_0(X) \;=\; \log N$	[sh/signe]
Entropie de la source	$\boxed{H_0(X) \;= -\sum_{i=1}^{N} P_i \log P_i}$	[sh/signe]
Entropie de la source composée	$\boxed{H_0(XY) \;= -\sum_{i=1}^{N}\sum_{j=1}^{P} P_{ij} \log(P_{ij})}$	[sh/paire]
Entropie conditionnelle moyenne de Y par rapport à X	$H_0(Y\|X) = \sum_{i=1}^{N} P_i\, H(Y\|x_i)$	[sh/signe]
Entropie conditionnelle de Y quand $X = x_i$	$H_0(Y\|x_i) = -\sum_{j=1}^{P} P_{j\|i} \log(P_{j\|i})$	[sh/signe]
Entropie d'un ensemble de q sources (Entropie de la q-ième extension d'une source X)	$H_0(S_q) \;=\; \sum_{i=1}^{q} H_0(X_i)$	[sh/mot]
Entropie limite	$H_\infty(X) \;=\; H_0(X)$	[sh/signe]

b. Source homogène avec mémoire d'ordre m : signes dépendant des m précédents (modèle de Markov)

Si stationnaire, les probabilités conditionnelles sont constantes dans le temps

Entropie conditionnelle à la séquence seq de m signes	$H_m(X\|seq) = -\sum_{i=1}^{N} P_{i\|seq} \log(P_{i\|seq})$	[sh/signe]
Entropie de la source	$H_m(X) \;=\; \sum_{j=1}^{N^m} P_{seq_j}\, H(X\|seq_j)$ $H_m(X) \;= -\sum_{j=1}^{N^m}\sum_{i=1}^{N} P_{seq_j} P_{i\|seq_j} \log\left(P_{i\|seq_j}\right)$	[sh/signe]
Entropie de la q-ième extension d'une source X $\quad H(X^q)$	Si $m \geq q : H(S_q) = H_\mu(S_q) = qH_m(X)$ X^q source de Markov d'ordre $\mu = {}^m/_q$ Sinon : $H(S_q) = H_0(S_q)$ X^q source sans mémoire	[sh/mot]

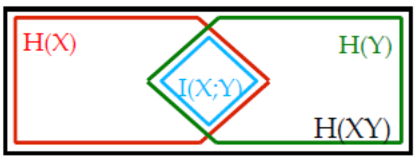
5. Information mutuelle

Information mutuelle de X (source à étudier) et Y (source observable)

$$\boxed{\begin{array}{ccccc} I(Y \to X) & = & I(X \to Y) & = & I(X;Y) \\ \| & & \| & & \| \\ H(X) - H(X|Y) & & H(Y) - H(Y|X) & & H(X) + H(Y) - H(XY) \end{array}}$$

Propriétés :

- $I(X;Y) = 0 \qquad\qquad\qquad \Leftrightarrow X, Y$ independantes
- $I(X;Y) = H(X) = H(Y) \Leftrightarrow X, Y$ équivalentes



6. Autres formules

$\boxed{\eta_X = \frac{H_\infty(X)}{\max H(X)}}$	$\boxed{r_X = 1 - \mu_X}$	$\boxed{\dot{H} = \frac{H_0(S)}{Te}}$	$\boxed{D_S = \frac{1}{T_e}}$
Efficacité informationnelle	Redondance relative	Débit d'information moyen [sh/s]	Cadence d'émission [bits/s]

III. Codage de source : compression de l'information

Codage d'une source S de N signes en une source C de Q mots-codes de longueurs l_i symboles

1. Classification

- Code réversible
  - Longueur fixe $L \geq \log N$
  - Longueur variable
    - Code instantané / à préfixe : décodage dès réception (aucun code n'est préfixe d'un autre)
    - Code avec séparateur.
- Code irréversible avec critère de fidélité

Code instantané $\Leftrightarrow$ Inégalité de Kraft respecté (CNS) $\sum_{i=1}^{N} Q^{-l_i} \leq 1$

2. Longueur optimale

a. Encodage d'un signe d'une source sans mémoire

$$H_0(C) = \frac{H_0(S)}{L}$$

Entropie après codage source sans mémoire

$$\eta_C = \frac{H_0(C)}{\log Q} = \frac{H_0(S)}{L \log Q}$$

Efficacité informationnelle

$$L = L_{moy} \geq \frac{H_0(S)}{\log Q}$$
$$L \geq H_0(S) \text{ si } Q = 2$$

Longueur moyenne [symbole/signe]

b. Bloc de q signes

$$\frac{H(S)}{\log Q} \leq \frac{L_q}{q} < \frac{H(S)}{\log Q} + \frac{1}{q}$$

Longueur moyenne *[symbole/q signes]*

$$\eta_C = \frac{H_m(C)}{\frac{L_q}{q}}$$

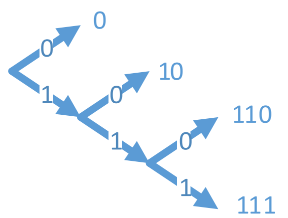
Efficacité Info.

Propriétés :

- **1er théorème de Shannon :** Un signe peut être codé au minimum avec $L = H_0(S)$ bits
- Optimum atteint sur $l_i = -\log p_i = I(s_i)$
- Il y a toujours au moins un codage tel que $\frac{H_0(S)}{\log Q} \leq L < \frac{H_0(S)}{\log Q} + 1$

3. Codages binaires sans distorsion

a. Codage par arbre d'encodage



b. Codage de Shannon-Fano

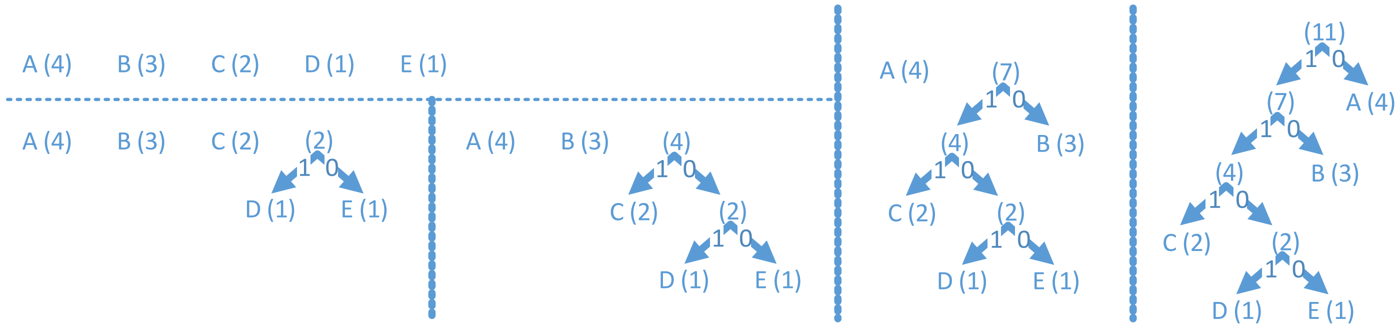
- Ordonner les signes par p_i croissants
- Diviser l'ensemble en 2 sous-ensembles équiprobables
- Attribuer un symbole à chaque ensemble
- Répéter la procédure pour chaque sous ensemble

A	5	0	00	00
B	2	0	01	01
C	2	1	10	10
D	1	1	11	110
E	1	1	11	111

Note : ne garantit pas un L minimum

c. Codage de Huffman

- Chaque signe constitue une des feuilles de l'arbre de poids p_i.
- On associe à chaque fois les deux nœuds de plus faibles poids pour donner un nœud dont le poids équivaut à la somme des poids de ses fils.
- Répéter jusqu'à n'avoir qu'un arbre.
- On associe ensuite le code 0 à la branche de gauche et le code 1 à la branche de droite.



d. Codage par plage

On code des paires (lgPlage$_i$, signe$_i$). Ex : 001110 → (2,0) (3,1) (1,0)

On peut ensuite coder les paires avec 1 ou 2 mots codes à longueur variable avec Huffman ou SF.

e. Codage de Lempel-Ziv

Principe : repérer les séquences par MAJ d'une table de traduction. Une séquence est codée par son adresse dans la table.

- Mettre les signes de l'alphabet dans la table
- Trouver la plus longue chaine du message dans la table et la replacer par son code
- Ajouter cette chaine suivie du caractère suivant à la table
- Répéter jusqu'à fin d'encodage

1	**A**
2	B
3	AB
4	AA
5	ABB
6	BA
7	AAA

A	B	A	A	B	B	A	A	A	A
1	2	1	3		2	4		4	

f. Codage arithmétique

Règle de partition : Découper l'intervalle en N partition de tailles $\propto p_i$ $\;M_i$: position du signe i ; N_i : norme du signe i

- Découper l'intervalle disponible (initialement [0,1])
- Choisir comme nouvel intervalle la portion qui correspond au signe à coder
  - $M = M + N \times M_i$
  - $N = N \times N_i$
- Répéter jusqu'à fin

Un message est codé par l'intervalle final. On envoie les bits après la virgule de M.

Exemple :

s_i	p_i	M_i	N_i	s	M	N
a	1/2	0.0	0.1	**b**	0.1	0.01
b	1/4	0.1	0.01	**ba**	0.100	0.001
c	1/8	0.01	0.001	**bac**	0.10011	0.000001
d	1/8	0.001	0.001	**baca**	0.**1001100**	

4. Codages binaires avec distorsion

a. Théorie

Pour s'adapter au débit du canal, on peut accepter une distorsion du message.

- **Distorsion admissible δ :** $R_{min}(\delta) \le L_{dist} < H(S) \le L_{\overline{dist}}$
- **Taux de compression :** $T_c = \frac{L_{\overline{dist}}}{L_{dist}}$
- **Distances :** $d(x_i, \widehat{x_i}) = \begin{cases} 0 \text{ si } x_i = \widehat{x_i} \\ 1 \text{ si } x_i \neq \widehat{x_i} \end{cases}$ $\quad d(x_i, \widehat{x_i}) = (x_i - \widehat{x_i})^2 \quad \dots$
- **Mesure de distorsion :**

$$EQM = \sum p_i (x_i - \widehat{x_i})^2 \qquad RSB = 10 \log \frac{P_{signal\ max}}{P_{moy\ err}} = 10 \log \frac{E(X^2)}{E(X - \hat{X})^2} = \frac{\sigma_X^2}{EQM} \qquad RSBC = 10 \log \frac{(niveau\ de\ gris\ max)^2}{EQM}$$

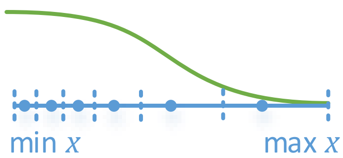
- **Information mutuelle :** $I(X; \hat{X}) = H(X) - \underbrace{H(X|\hat{X})}_{\text{compression}}$
- **Fonction du taux de distorsion :** $R_{min}(\delta) = \min_{p(x\,|\,\hat{x})\bar{d}} I(X; \hat{X})$
- **Taux de compression max :** $\max T_C = \frac{H}{R_{min}}$

b. Quantification scalaire : quantification d'une seule valeur

Uniforme

Niveaux et seuils équirépartis.
Pour EQM fixé, on minimise le nombre de niveaux.

Non-uniforme



Niveaux et seuils non équirépartis déterminés d'après la répartition des valeurs.
En pratique on transforme le signal puis on utilise une QS uniforme.

c. Quantification vectorielle : quantification par groupe de valeurs (ex : bloc de px)

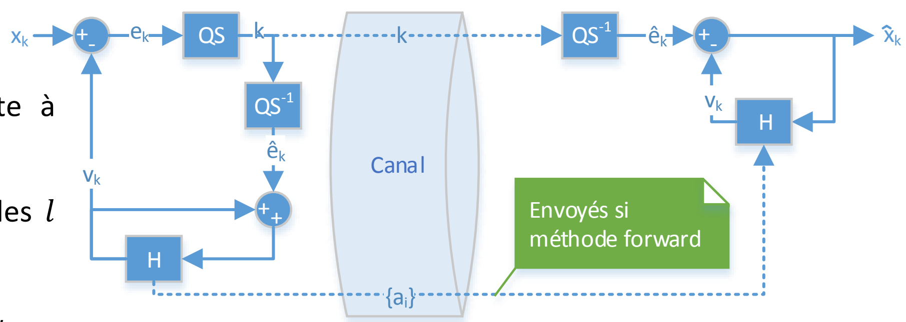
- Même principe que QS avec dictionnaires de vecteurs et non de valeurs.
- Prise en compte de la corrélation interne au vecteur.

d. Codage prédictif

Quantification de la différence entre la valeur réelle et la valeur prédite à partir des précédents.

Prédiction v_k du signe x_k à partir des l précédents : erreur $\varepsilon_k \sim \mathcal{N}(0, \sigma^2)$

Prédiction linéaire : $v_k = \sum_{i=1}^{l} a_i x_{k-i}$



En méthode **forward**, les coefficients a_i sont envoyés dans le canal. En méthode **backward**, ils sont estimés.

Exemple : MICDA (Modulation d'Impulsions Codées Différentielles Adaptatives) : $v_k = x_{k-1}$

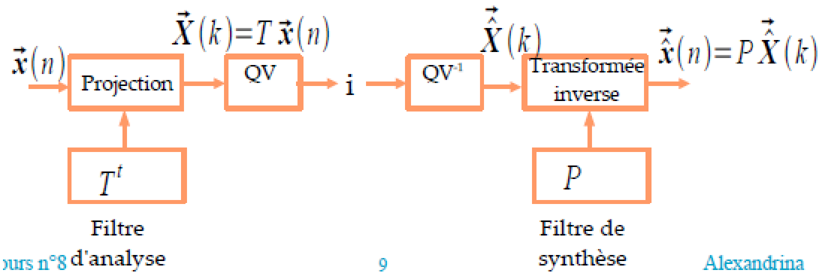
5. Formats

- **GIF :** LZW
- **PNG :** LZ77 + prédiction
- **TIFF :** LWZ
- **JPEG :** QV, codage par plage + Huffman (DTC)
- **MP3 :** QS, Huffman
- **MPEG :** MP3 + JPEG + mouvement en arithmétique
- **H261 :** Différences codées en QV, codage par plage + Huffman
- **T4 / T5 / CCITT :** Plage

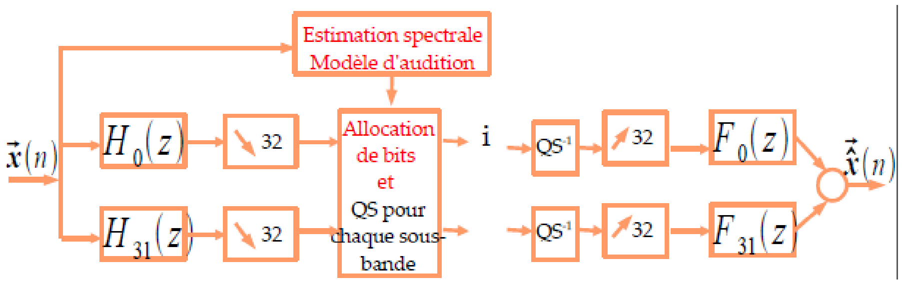
IV. Architectures complexes de compression de l'information

Solutions de quantification d'un signal non-stationnaire

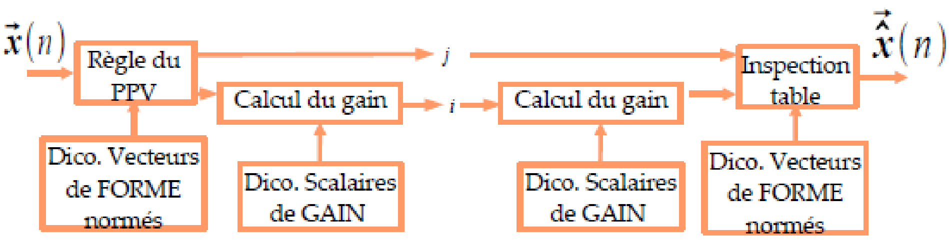
- Modifier le signal pour pouvoir utiliser un seul dictionnaire
  - Application de transformée (DCT, TFD, ondelettes)
    - Elimination des composantes peu énergétique



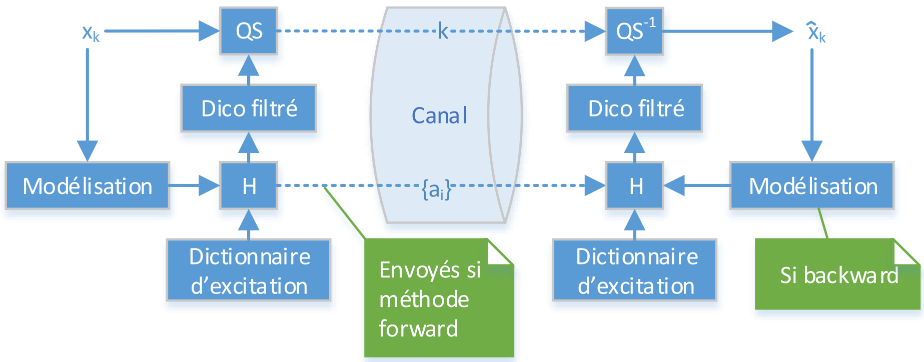
- Décomposition du signal en plusieurs signaux
  - Sous-bandes
    - Quantif : On découpe le signal en plusieurs bandes, on sous-échantillonne chaque bande et on utilise 1 dico par bande.
    - Reconstit : On interpole les bandes fréquentielles et on recrée le signal temporel.



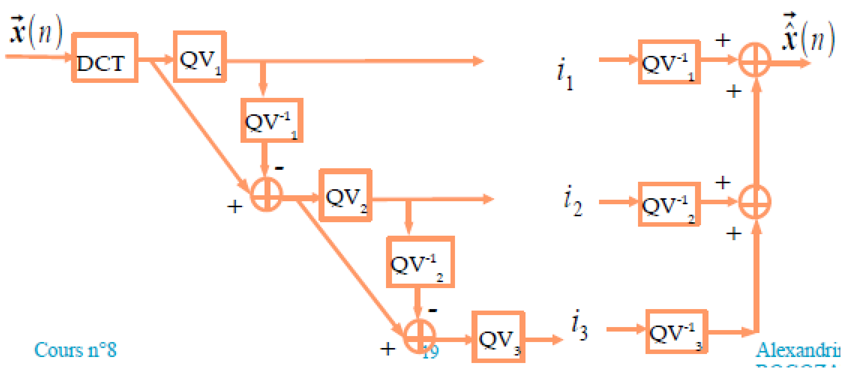
  - Représentation Forme-Gain



- Modifier régulièrement le dictionnaire pour s'adapter au signal
  - Utiliser une batterie de dictionnaire définis à priori
  - Utiliser un modèle simple de production (quantification probabiliste)



Reconstitution à différentes résolutions



V. Codage de canal

1. Canal bruité

Source en entrée du canal X, source en sortie du canal

$$P_{ij} = \mathbb{P}(Y = y_j | X = x_i)$$

Matrice de perturbation
Sans bruit, $P = I$
Symétrique : $P_{ij,i\neq j} = \alpha$

$$\boxed{V_m = \frac{1}{T_{plus\ court\ signal}}}$$

Vitesse de modulation de la voie [bauds]

$$I_{max}(C) = \max_{\mathbb{P}(x_i)} I(X;Y)$$

Max d'entropie [sh]
Canal binaire sans bruit
$I_{max} = 1$

$$\boxed{C = I_{max}(C) \times D_C}$$

Capacité d'un canal [sh/s]
N nb de signes

2. Décodeur à taux d'erreur minimal

Principe : Associer à y_j le symbole correct x_j^c, minimise statistiquement le taux d'erreur.

$$x_j^c = \underset{x_i}{\operatorname{argmax}}\, P_{ij} p_i = \underset{x_i}{\operatorname{argmax}}\, \mathbb{P}(Y = y_j | X = x_i)\mathbb{P}(X = x_i)$$

$$\mathbb{P}_{moy\ erreur} = 1 - \sum_{i=1}^{N_y} \mathbb{P}(Y = y_j | X = x_j^c)\mathbb{P}(X = x_j^c)$$

3. Codage correcteur d'erreur

k *bits d'info,* m *bits de redondance,* $n = m + k$ *bits au total*

a. Classification des erreurs

- Erreurs individuelles : erreurs indépendantes
- Paquets d'erreurs : suite dont le premier et le dernier symbole sont erronés

b. Classification des codes correcteurs

- Codes convolutifs : effectués de manière continue
- Codes en blocs : effectuées sur un bloc
  - Codes q-aires : mots considérés ∈ corps à q éléments
  - Codes binaires : mots considérés ∈ corps de Galois à 2 éléments (opérations $mod\ 2$)
    - Linéaires : sommes de contrôle
    - Systématiques : bits de redondance à la suite des bits d'info
  - Codes groupes : mots considérés vecteurs dans un espace vectoriel
  - Codes cycliques : mots considérés comme polynômes

c. Code groupe linéaire systématique (ex : Hamming)

$d \in \mathbb{B}^k$	$c \in \mathbb{B}^n$	$r \in \mathbb{B}^n$	$e \in \mathbb{B}^n$	$s \in \mathbb{B}^m$	$G \in \mathbb{B}^{k \times n}$	$H \in \mathbb{B}^{m \times n}$
données	mot-code émis	mot-code reçu	erreurs	syndrome	matrice génératrice	matrice de contrôle de parité

$\boxed{G = [I_k \quad P]}$ $\boxed{H = [P^\top \quad I_m]}$ $\boxed{c = dG = [d \quad dP]}$ $GH^\top = 0 \Rightarrow \underline{cH^\top = 0}$

Détection : On vérifie que $r = c + e$ est un mot code en vérifiant $\underline{s = rH^\top} = 0 \Rightarrow \exists\, c' \mid e = 0$

Correction : s prend la valeur de la colonne de H correspondant à la position de l'erreur

Distance minimale de Hamming d_{min} : plus petite distance entre 2 mots-codes = nombre minimum de 1 dans les mots-codes $c \neq 0$.

Puissance : Il est possible de façon exclusive de :

- Détecter $d_{min} - 1$ erreurs
- Corriger $\frac{d_{min}-1}{2}$ erreurs

P **:** Lignes de P différentes entre elles et différentes de 0.

Il faut $\boxed{2^m = 2^{n-k} \geq \sum_{i=0}^{n_c} \binom{n}{i}}$ pour corriger n_c erreurs. (Ex : $n_c = 1 : n + 1 \ /\ n_c = 2 : \frac{n^2}{2} + \frac{n}{2} + 1$)

Détecter une erreur : $m = 1 \Rightarrow d_{min} = 2$

Corriger une erreur ou détecteur 2 erreurs (code de Hamming) : $\forall\, m \in \mathbb{N}, n = 2^m - 1 \Rightarrow d_{min} = 3$

Codage de Hamming : Les colonnes de H sont les représentations des nombres de 1 à n.

d. Code binaire cyclique non-systématique (ex : BCH)

$$d(x) = \sum_{i=0}^{k-1} d_i x^i \qquad g(x) = \sum_{i=0}^{m-1} g_i x^i + x^m = \frac{x^n + 1}{h(x)} \qquad c(x) = \sum_{i=0}^{n-1} c_i x^i$$

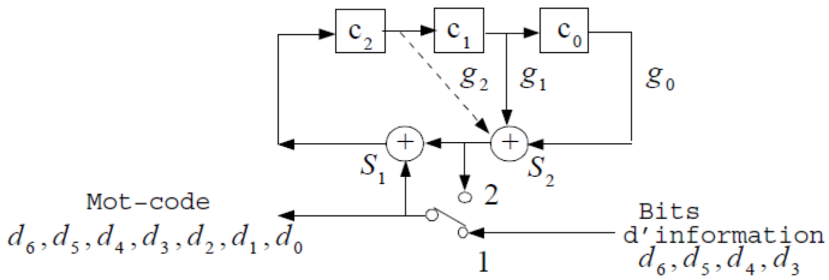
i. Non systématique

$$c(x) = d(x)g(x)$$

ii. Systématique

$$c(x) = reste\left(\frac{d(x)x^m}{g(x)}\right) + d(x)x^m$$

$$s(x) = reste\left(\frac{r(x)}{g(x)}\right)$$



Interrupteur sur 1 pendant k instants puis sur 2 pendant m instants

iii. Code BCH

Correction de t erreurs.

$$g(x) = x^n + 1 \qquad n = 2^z - 1 \qquad k \geq 2^k - zt - 1 \qquad d_{min} \geq 2t + 1$$

Racines α^i de $g(x)$ générées par un polynôme primitif $q(x)$ de degré z diviseur de $g(x)$. *Ex :* $q(x) = 1 + x + x^4$

On exprime les α^i modulo $q(x)$ et les coefficients modulo 2. *Ex :* $\alpha^4 = 1 + \alpha$

On choisit les racines r_j parmi les α^i : $\{\alpha, \alpha^3, \dots, \alpha^{2t-1}\}$ *Ex :* $\{\alpha, \alpha^3, \alpha^5\}$ *pour* $t = 3$

On détermine les polynômes minimaux distincts correspondants $pm_j(x)$. Ces polynômes ont pour racine $\alpha^j, \alpha^{2j}, \alpha^{2^2 j}, \dots$ On écrit donc $pm_j(x) = (x - \alpha^j)(x - \alpha^{2j})(\dots)$ jusqu'à avoir tous les $\alpha^{2^k j\ [n]}$ possibles. On réduit ensuite $pm_j(x)\ [q(x)]\ [2]$. *Ex :* $pm_2 = (1 - \alpha^3)(1 - \alpha^6)(1 - \alpha^{12})(1 - \alpha^{9 \equiv 24})$ *pour* $n = 15$

$$pm_2 = x^3 + x^2 + x + 1\ [q]\ [2]$$

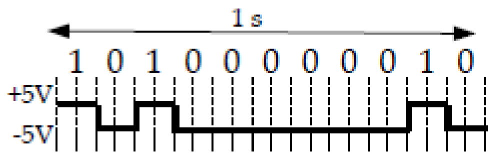
On groupe ensuite $g(x) = \prod pm_j$ de degré m (nombre de bits de contrôle).

VI. Codage de ligne

Associe un support physique adéquat aux éléments binaires émis par la source

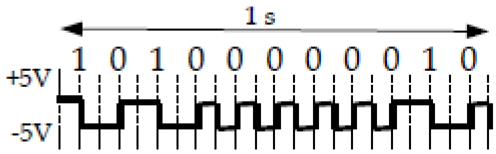
a. En bande de base

- BF atténuées
- La densité de transitions permet de reconstituer l'horloge
- Immunité contre le bruit variable



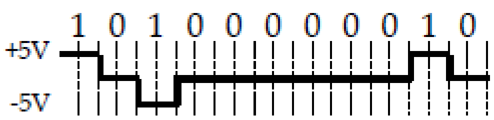
Codage en ligne NRZ

$V_m = 11$ bauds $D_e = 11$ bits/s



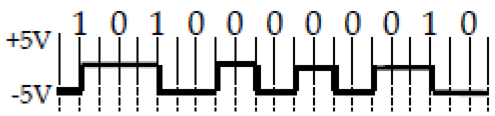
Codage en ligne de Manchester

$V_m = 22$ bauds $D_e = 11$ bits/s



Codage en ligne bipolaire

$V_m = 11$ bauds $D_e = 11$ bits/s



Codage en ligne de Miller

$V_m = 11$ bauds $D_e = 11$ bits/s

b. Modulation

- Modulation de phase, de fréquence, d'amplitude