Université Claude Bernard Lyon 1

Master 1 Informatique
MIF11 - Réseau
Codage et éléments de théorie de l'information

Florent Dupont

Université Claude Bernard Lyon1 / Laboratoire LIRIS
Florent.Dupont@liris.cnrs.fr
http://liris.cnrs.fr/florent.dupont

Objectifs du cours

- Théorie des codes…liée aux transmissions numériques
  - compression des signaux transmis (son, image, vidéo, 3D…)
  - protection contre les erreurs de transmission
  - cryptage, authentification des messages
  - …

Objectifs du cours

- Rappeler quelques notions de base en transmission de données :
  - Bande passante, capacité d'un canal ;
  - Modulation ;
  - Codage en bande de base ;
  - Multiplexage.

- Avoir un avis éclairé pour répondre, par exemple aux questions suivantes :
  - La transmission sans erreur sur un canal bruité est-elle possible?
  - Est-ce que transmettre sans erreur, c'est avoir un rendement tendant vers 0 ou bien existe-t-il un rendement optimal, pour un niveau de bruit donné ?

Plan du cours

1. Notions de base en transmission de données
2. Théorie de l'information
3. Source discrète / Codage source
4. Canal discret / Codage canal
5. Codes détecteurs et codes correcteurs

9h de Cours Magistral
6h de Travaux Dirigés

Pour en savoir plus…

- Théorie des codes
  Compression, cryptage, correction

  (Ed. Dunod)

  Jean-Guillaume Dumas, Jean-Louis Roch, Eric Tannier, Sébastien Varrette

1. Notions de base en transmission de données

- **Canal** = dispositif permettant d'acheminer un message entre deux points distants
- Sur de courtes distances :

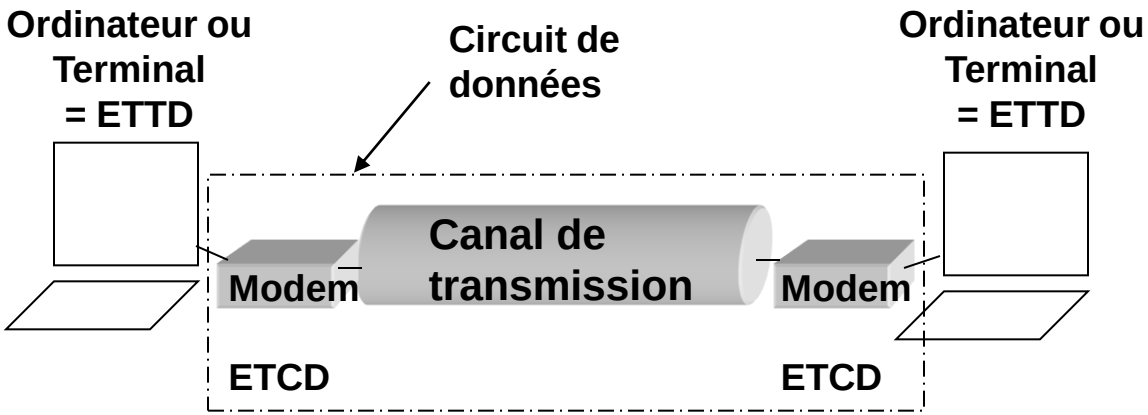
- **Canal de transmission =**

  support physique (câble coaxial, fibre optique, air…)

- **ETTD : Équipement Terminal de Transmission de Données**

Canal discret

- Sur de plus longues distances



- Ex : canal de transmission = Ligne téléphonique
- Modem : Modulateur / démodulateur
- ETCD: Équipement Terminal de Circuit de Données
- Les informations sont transmises sur des supports de transmission en faisant varier un ou plusieurs paramètres physiques des signaux.

Parenthèse

Traitement du signal fréquence / discret

Fréquences : Fourier

- Toute **fonction périodique g(t)** ayant pour période **T=1/f** peut se décomposer en une somme de fonctions périodiques sinusoïdales et cosinusoïdales :

$$g(t) = c + \sum_{n=1}^{\infty} a_n \sin(2\pi nft) + \sum_{n=1}^{\infty} b_n \cos(2\pi nft)$$

- Les coefficients a_n et b_n sont les amplitudes respectives des sinus et cosinus (harmoniques) et c est égal à la valeur moyenne du signal :

$$a_n = \frac{2}{T}\int_0^T g(t)\sin(2\pi nft)dt, \quad b_n = \frac{2}{T}\int_0^T g(t)\cos(2\pi nft)dt, \quad c = \frac{1}{T}\int_0^T g(t)dt$$

- Cette décomposition est appelée ***série de Fourier***.

- Exemples : fréquences dans un signal, une image…

Fréquences : Fourier

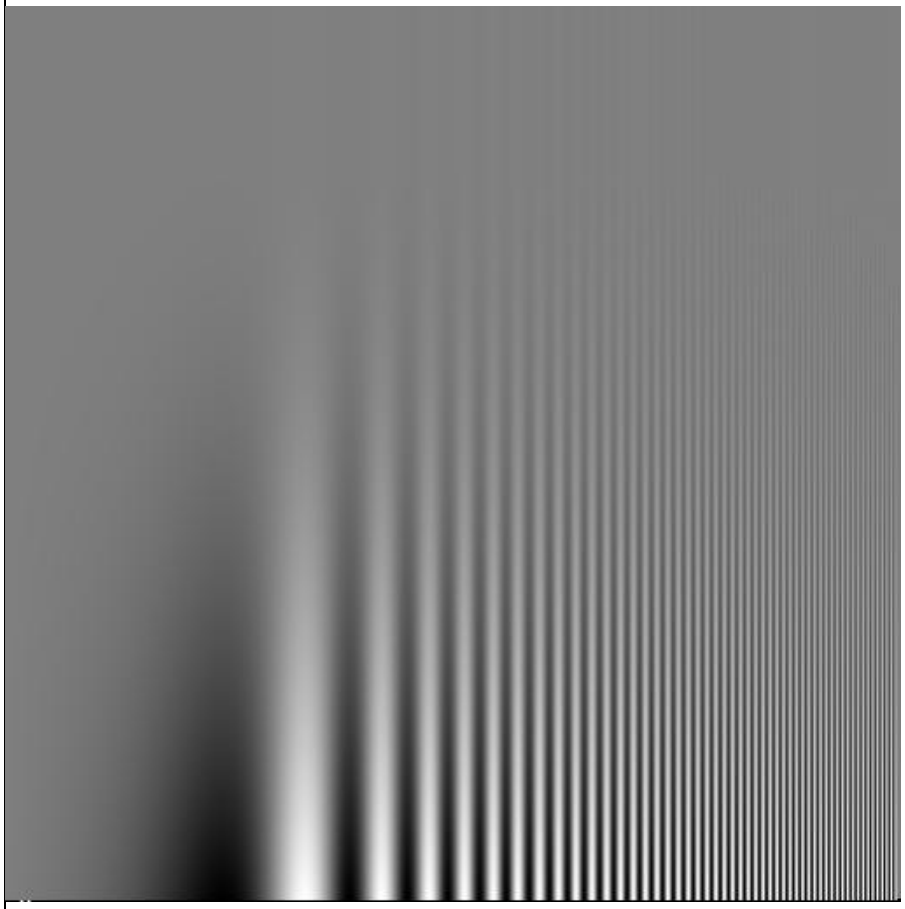
- **Transformée de Fourier**

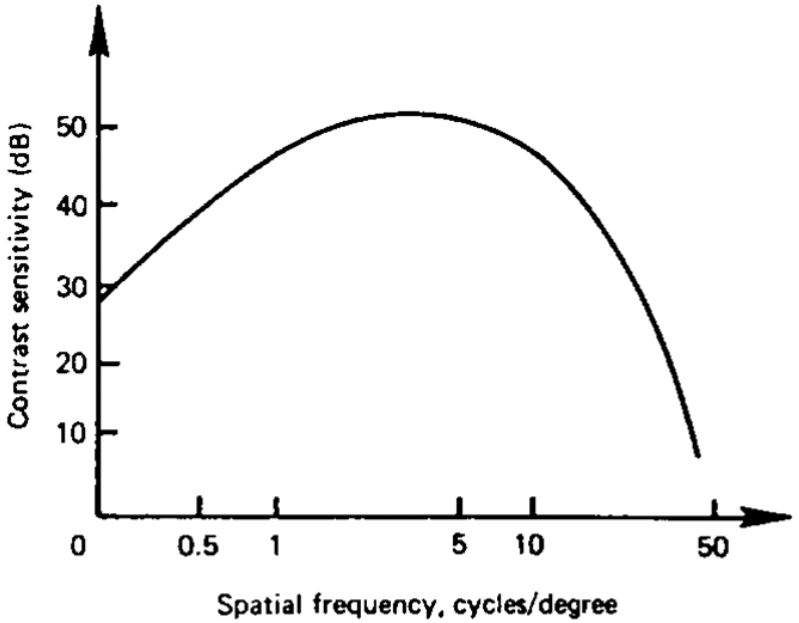
  → Représentation d'un signal sur une base de fonctions exponentielles complexes

  – Cas mono-dimensionnel

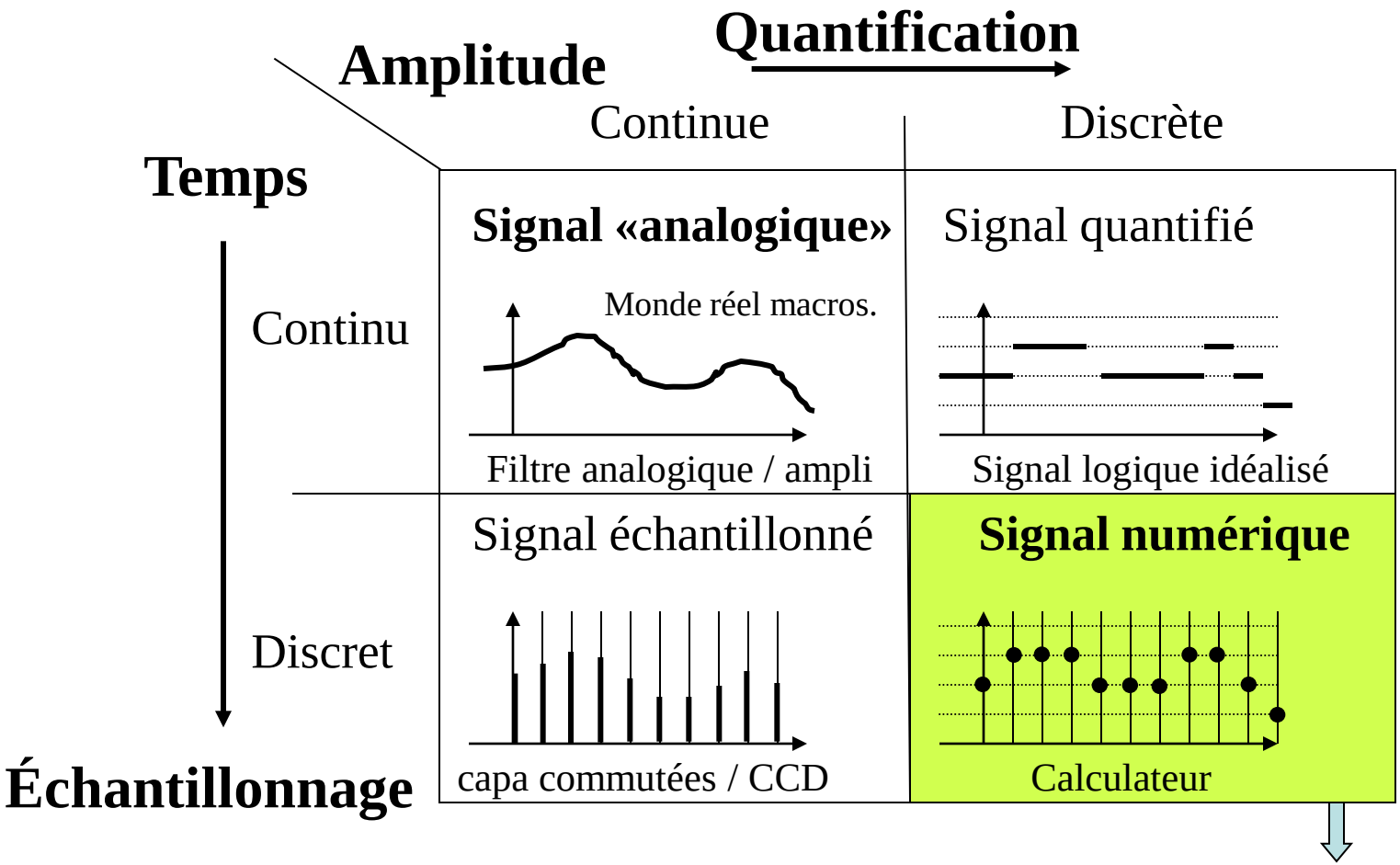
$$f(x) = \int_{-\infty}^{+\infty} F(u).e^{2\pi jux}\,du \qquad \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftarrows}} \qquad F(u) = \int_{-\infty}^{+\infty} f(x).e^{-2\pi jux}\,dx$$

Exemple : réponse en fréquence de l'oeil





Numérisation / discrétisation

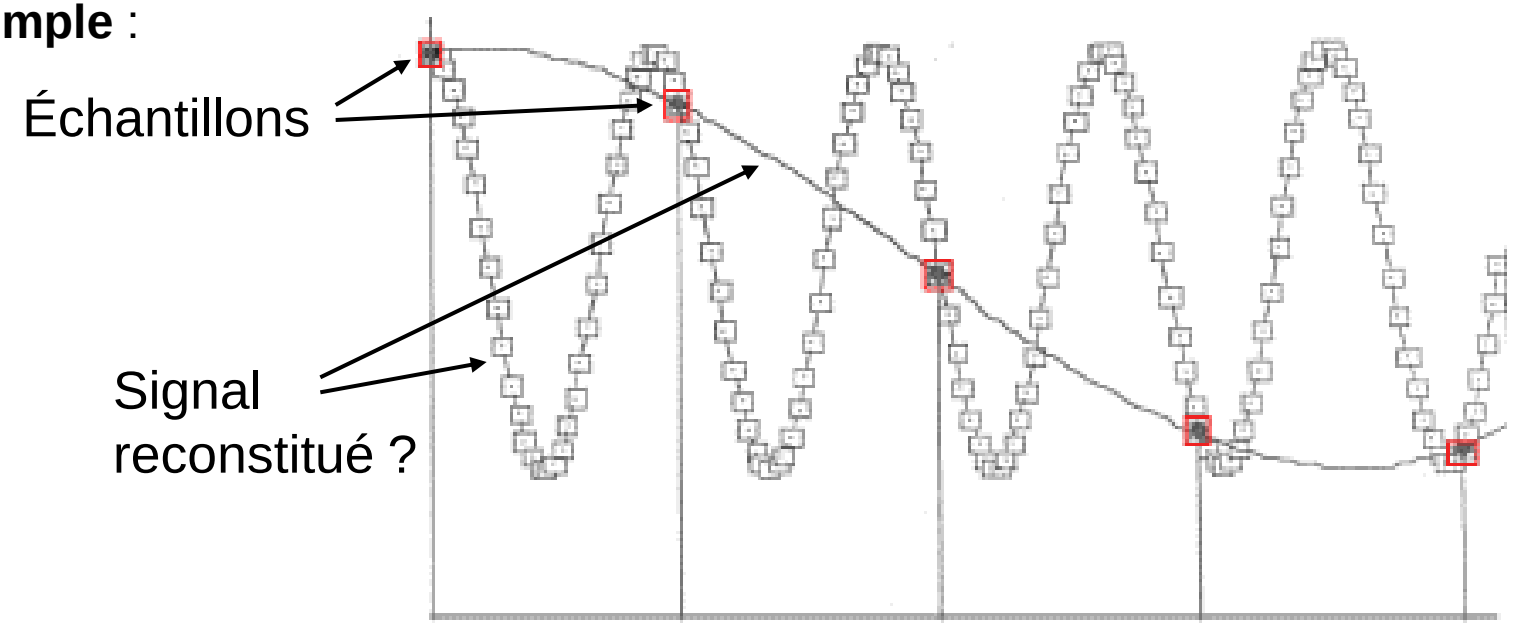


Il n'y a que dans ce cas que l'on peut associer un nombre entier au signal

Échantillonnage : Théorème de Shannon

Théorème De Shannon: Fe > 2 x Fmax(Signal)

Exemple :



Un signal incorrectement échantillonné ne pourra pas être reconstitué

Fin de la parenthèse

Bande passante

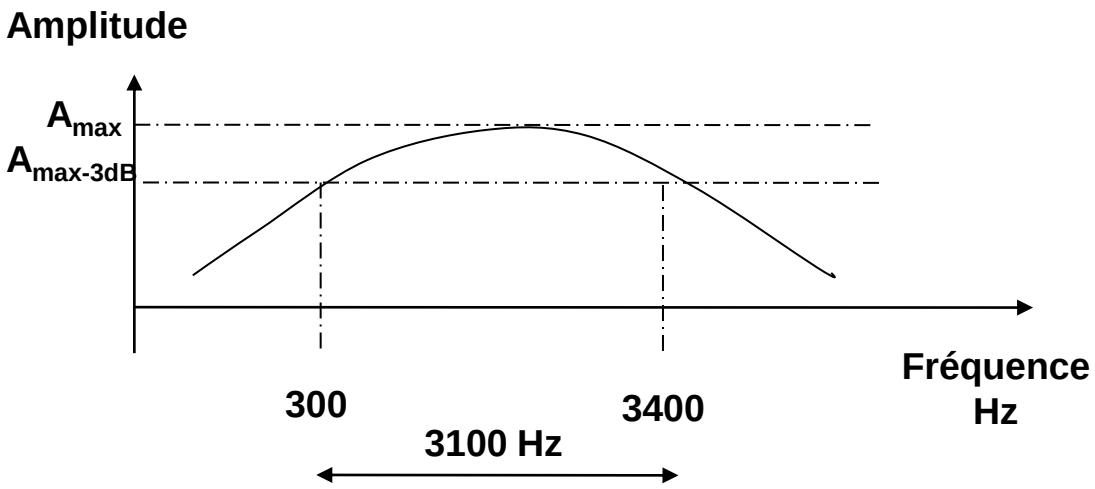
La bande passante caractérise tout support de transmission, c'est la bande de fréquences dans laquelle les signaux sont correctement reçus :

W = Fmax – Fmin (en Hz)

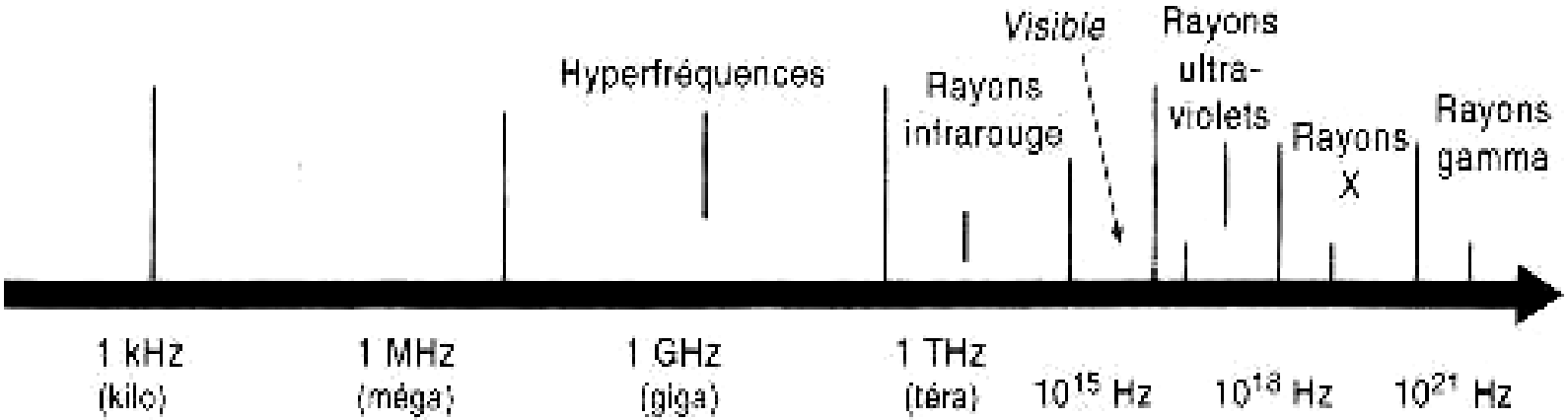
- Le spectre du signal à transmettre (éventuellement modulé) doit être compris dans la bande passante du support physique.

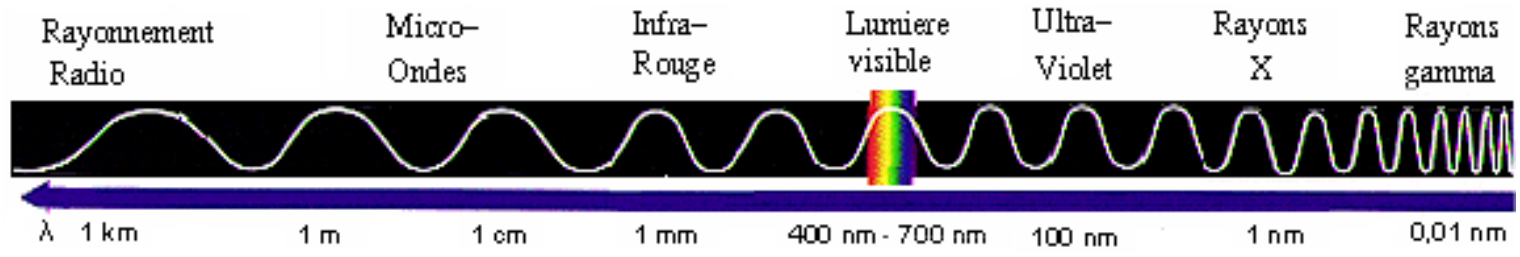
Bande passante

- Exemples:
- l'atmosphère élimine les U.V.
- l'oreille humaine est sensible dans la bande 20 Hz-20 KHz
- Réseau téléphonique commuté (RTC)

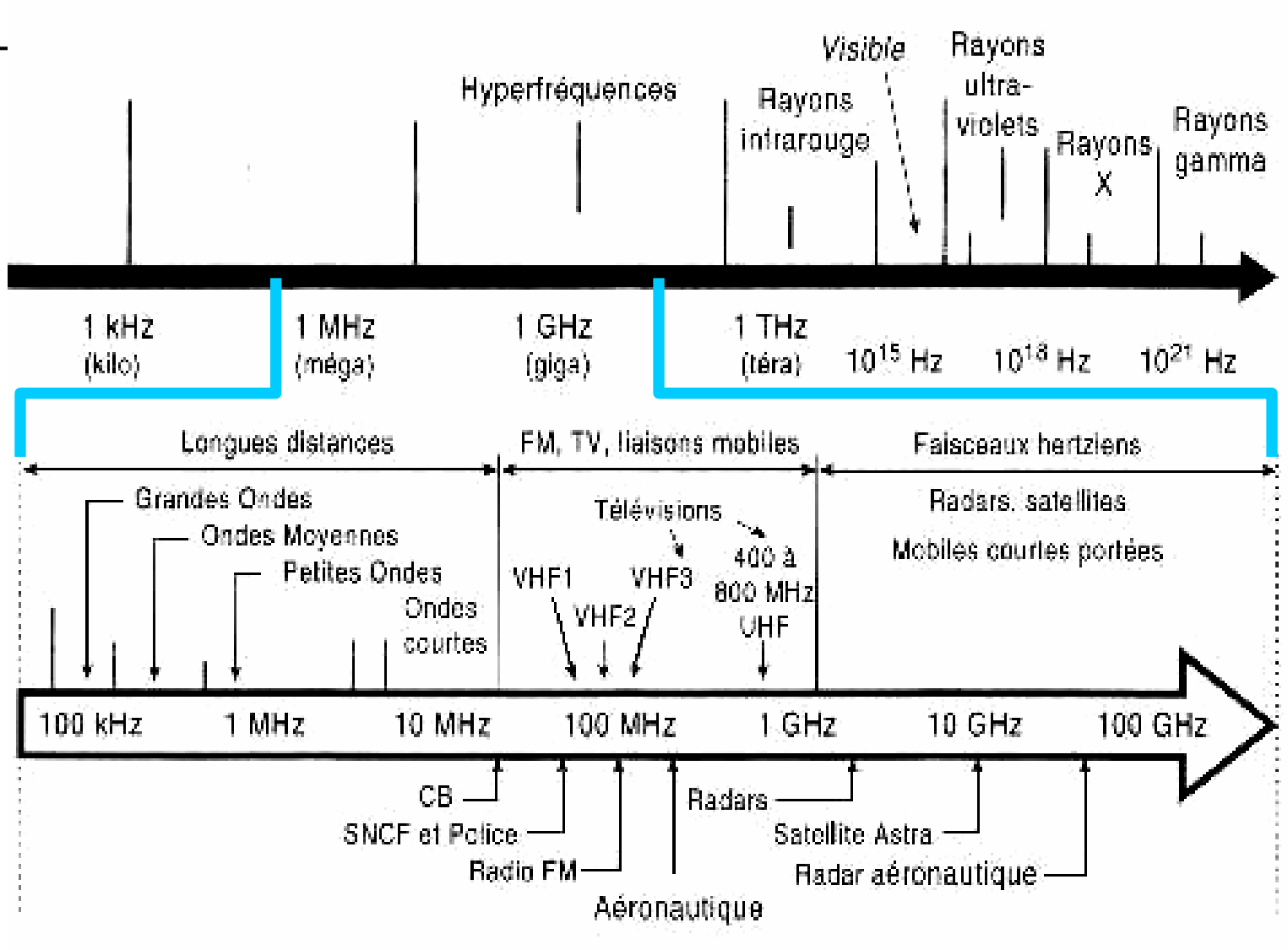


Utilisation des bandes de fréquences





Utilisation des bandes de fréquences



Agence nationale des fréquences (www.anfr.fr)

- **Bandes de fréquences** : attribuées aux différents services de radiocommunication par le ***Règlement des radiocommunications*** de **l'Union internationale des télécommunications**, élaboré par les conférences mondiales des radiocommunications.

- **En France, les bandes ainsi attribuées sont réparties entre 9 affectataires (7 administrations et 2 autorités indépendantes)**
  - **AC** Administration de l'aviation civile
  - **DEF** Ministère de la défense
  - **ESP** Espace
  - **INT** Ministère de l'intérieur
  - **MTO** Administration de la météorologie
  - **PNM** Administration des ports et de la navigation maritime (ex phares et balises)
  - **RST** Ministère de l'éducation nationale, de la recherche et de la technologie
  - **CSA** Conseil supérieur de l'audiovisuel
  - **ART** Autorité de régulation des Télécommunications

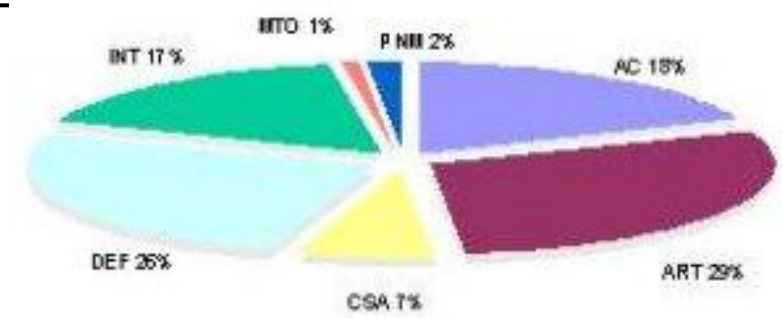
Agence nationale des fréquences (www.afnr.fr)

- + des fréquences utilisables pour certains matériels de faible puissance et de faible portée
- Exemple :

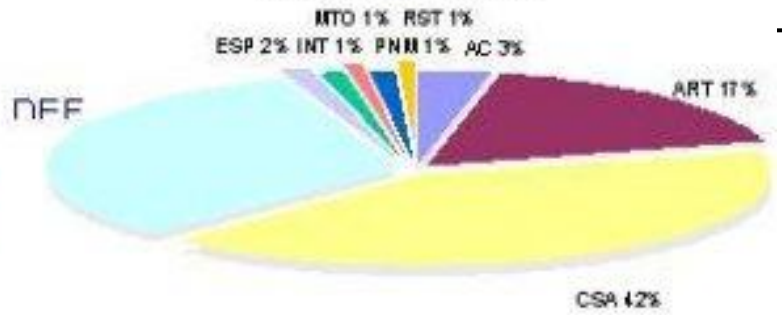
Bande des fréquences	2400 à 2454 MHz
Puissance max.	100 mW
Largeur canal	non imposée
Références	Décisions ART N°xxx

Répartition nationale des bandes de fréquences

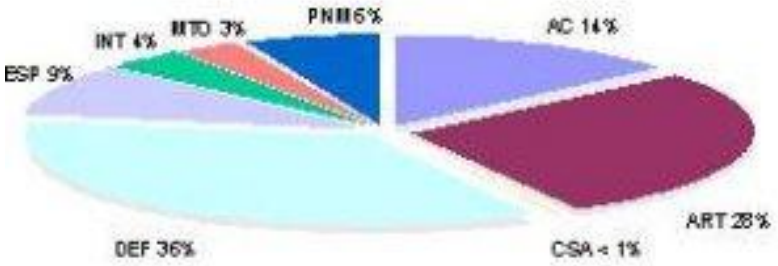
Bande 9 kHz – 29.7 MHz



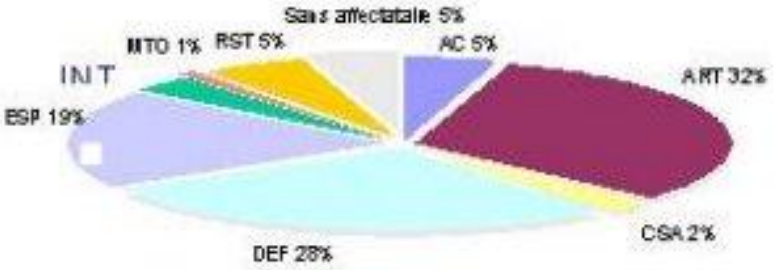
Bande 29.7 - 960 MHz



Bande 960 MHz -



Bande 10 GHz - 65 GHz



- > Aviation civile (AC)
- > Autorité de régulation des télécommunications (ART)
- > Conseil supérieur de l'audiovisuel (CSA)
- > Ministère de la défense (DEF)
- > Espace (ESP)
- > Ministère de l'intérieur (INT)
- > Météorologie (MTO)
- > Ports et navigation maritime (PNM)
- > Radioastronomie (RST)

Débit maximum d'un canal de transmission

- Si un signal quelconque est appliqué à l'entrée d'un filtre passe-bas ayant une bande passante W, le signal ainsi filtré peut être reconstitué avec un échantillonnage à 2W/s (Nyquist, Shannon)

$$\mathbf{D_{max} = 2\ W \log_2 V \quad \text{en bit/s}}$$

si le signal comporte V niveaux significatifs (Valence).

➢ La bande passante limite la **rapidité de modulation**.

Exemple: Pour un canal sans bruit dont la bande passante est de 3000 Hz qui ne peut transmettre qu'un signal binaire, D max = 6000 bit/s.

Bruit, capacité

- Bruits aléatoires $\Rightarrow$ dégradation de la transmission
- Quantité de bruit = rapport de la puissance du signal transmis à la puissance du bruit

  = rapport signal sur bruit, (SNR en anglais signal to noise ratio ou **S/N**).
- Pour un canal de transmission de bande passante W perturbé par du bruit dont le rapport signal sur bruit est S/N, la **capacité** de transmission maximale C en bit/s vaut :

$$\mathbf{C = W \log_2 (1+P_S/P_N) \quad en\ bit/s}$$

S/N est exprimé en dB en général, mais pas dans la formule !

$$\mathbf{(S/N)_{dB} = 10 \log_{10} (P_S/P_N) \Leftrightarrow P_S/P_N = 10^{(S/N)_{dB}/10}}$$

- **<u>Exemple</u>**: Pour un canal dont la bande passante est de 3000 Hz et un rapport S/N=30dB, (valeur typique du réseau téléphonique analogique), P_S/P_N=1000 => C = 30 000 bit/s.

Perturbations

- Perturbations $\Rightarrow$ l'information extraite du signal reçu peut conduire à des erreurs.
- Causes multiples, principale préoccupation dans les systèmes de télécommunication.
- **Affaiblissement ou atténuation = perte d'énergie du signal pendant sa propagation**

  $$\textbf{Atténuation (dB)} = 10 \log_{10} (P_1/P_2)$$

  (-3 dB correspond à une perte de la moitié de la puissance)
- Affaiblissements différents suivant les harmoniques $\Rightarrow$ **distorsions**

  En pratique affaiblissements d'amplitude négligeable jusqu'à f_c appelée ***fréquence de coupure***.

  Pour compenser cet affaiblissement et pour permettre des transmissions sur de longues distances $\Rightarrow$ amplificateurs ou répéteurs
- L'atténuation augmente avec la fréquence (passe-bas).

Perturbations

- La **distorsion temporelle** = toutes les composantes harmoniques d'un signal ne se propagent pas à la même vitesse.

- Un **déphasage** du signal (distorsion de phase) constitue une perturbation. $\Phi = \Phi(f)$. Le déphasage dépend de la fréquence. Le temps de groupe est donné par :

$$\mathbf{T(f)} = \frac{\mathbf{1}}{\mathbf{2\pi}} \times \frac{\mathbf{d(\Phi(f))}}{\mathbf{df}}$$

Bruit

- Tout signal indésirable interprété par le récepteur et délivrant une information incohérente.
- **Sources de bruit** :
  - émetteur du signal ;
  - media de transmission ;
  - perturbation atmosphérique.
- **Bruit thermique** = agitation thermique des électrons (source de bruit la plus courante)
- **Diaphonie** = influence mutuelle entre deux signaux utiles mais sur des conducteurs voisins.

Modulation / Démodulation

- Transmission d'un signal à spectre étroit sur un support à large bande passante $\Rightarrow$ mauvaise utilisation du support

$\Rightarrow$ techniques de **modulation** et de **multiplexage**

- Soit un **signal périodique** : $y(t) = A \sin(2\pi ft + \Phi)$
- Signal transporté sous forme d'une onde faisant varier une des caractéristiques physiques du support:
  - différence de potentiel électrique;
  - onde radioélectrique
  - intensité lumineuse

Porteuse: $p(t) = A_p \cos(2\pi f_p t + \Phi_p)$

- On fait ensuite subir des déformations ou modulations à cette porteuse pour distinguer les éléments du message.

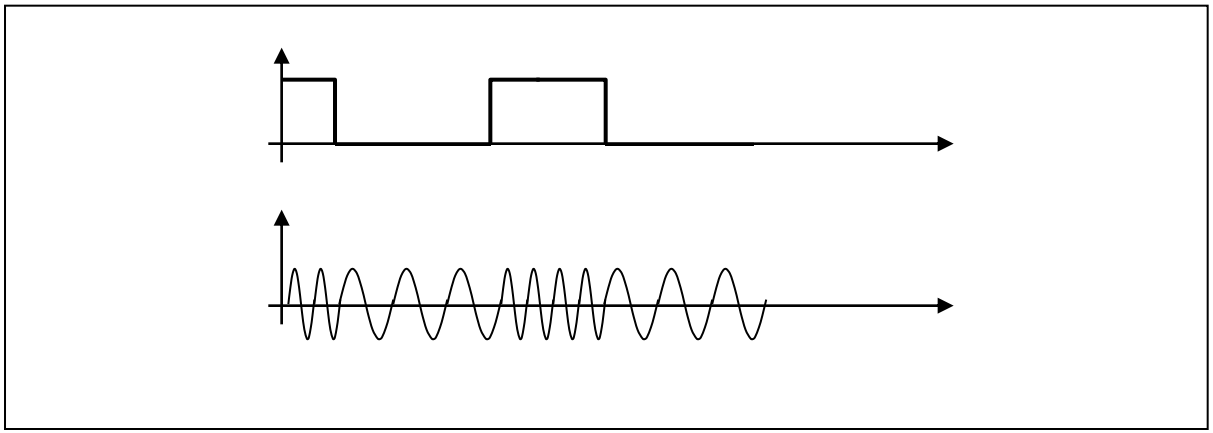
Modulation

- ***La modulation est la transformation d'un message à transmettre en un signal adapté à la transmission sur un support physique.***
- Les objectifs de la modulation sont:
  - une transposition dans un domaine de fréquences adapté au support de transmission;
  - une meilleure protection du signal contre le bruit;
  - une transmission simultanée de messages dans les bandes de fréquences adjacentes, pour une meilleure utilisation du support.
- Trois types de modulation de base existent, en faisant varier les trois paramètres de l'onde porteuse: A_p, f_p, Φ_p.

Modulation de fréquence

(FSK: Frequency Shift Keying)

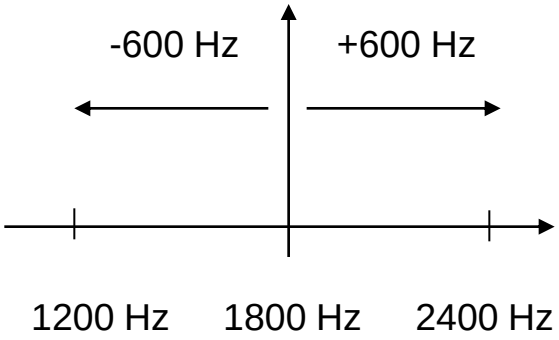
- une valeur de fréquence $\leftrightarrow$ une valeur du signal



Modulation de fréquence

- Porteuse sinusoïdale de fréquence F_0 modulée par deux fréquences opposées $+f_0$ et $-f_0$

$\Rightarrow$ une fréquence est associée à chaque niveau logique.



Modulation de fréquence

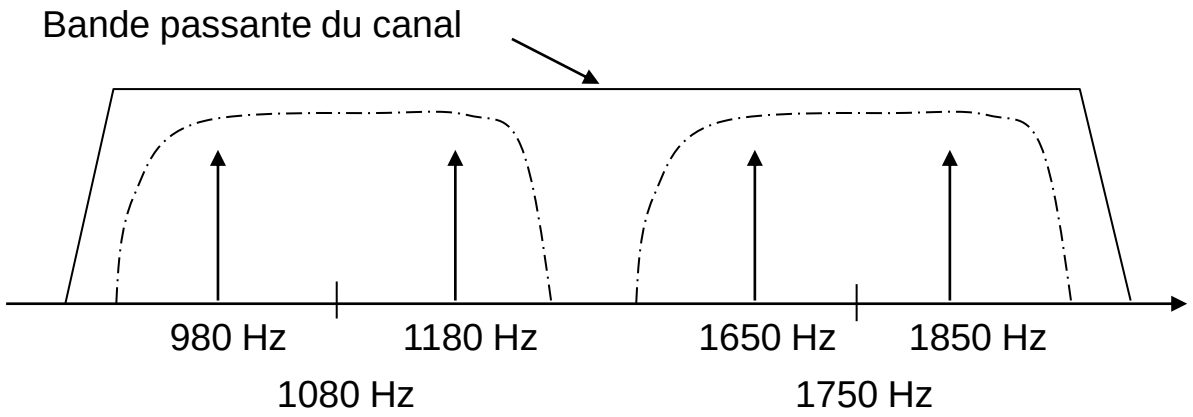
- Liaison "full-duplex":

  Émission / Réception simultanée

⇒ on partage la bande passante du canal

une voie à l'émission F_1 +/- f_1

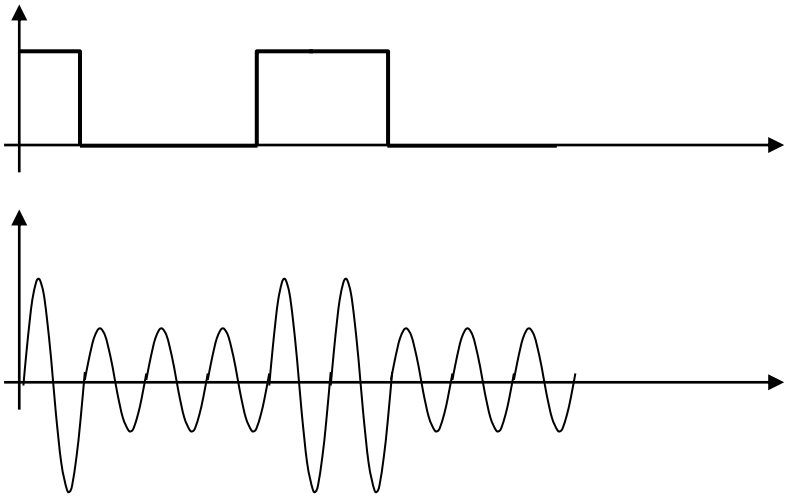
+ une voie à la réception F_2 +/- f_2



Modulation d'amplitude

(ASK: Amplitude Shift Keying)

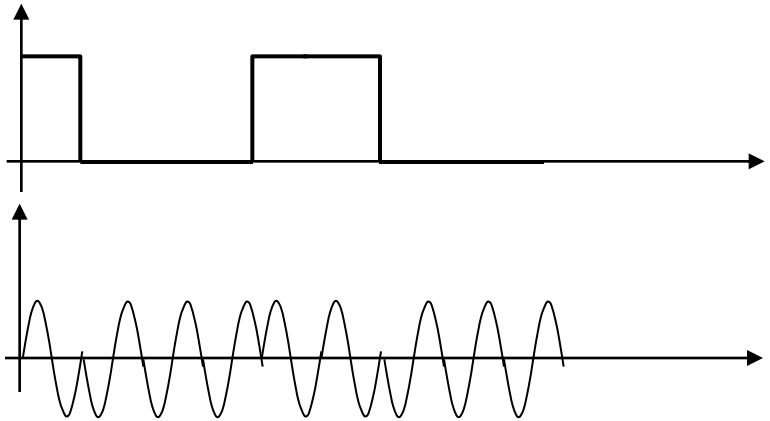
- une valeur d'amplitude ↔ une valeur du signal



Modulation de phase

(PSK: Phase Shift Keying)

- un déphasage ↔ une valeur du signal

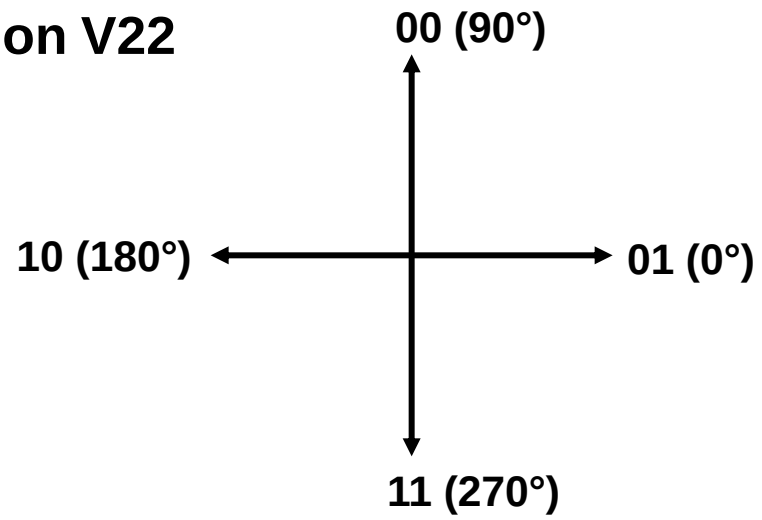


- Avec des codes à plusieurs bits, on peut augmenter le débit sans changer la fréquence de modulation.
- Les vitesses de transmission sont plus élevées qu'en modulation FSK pour la même bande passante

Modulation de phase

- Exemple : avis V22 du CCITT (1200 bauds) - phase codée sur 2 bits

Constellation V22



➢ Nombre de déphasages limité par le bruit pour retrouver le bon signal

Modulation combinée

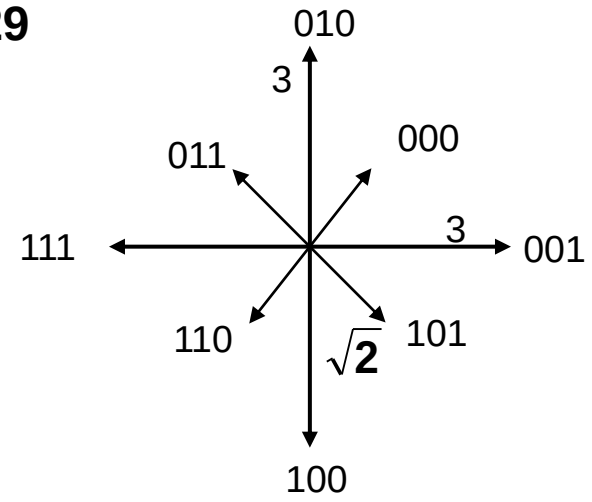
- Combiner plusieurs types de modulation parmi les trois types de modulation décrits auparavant.

➢ Les normes actuelles utilisent des combinaisons des modulations de phase et d'amplitude.

Exemple : Modulation V29 à 7200 bits/s

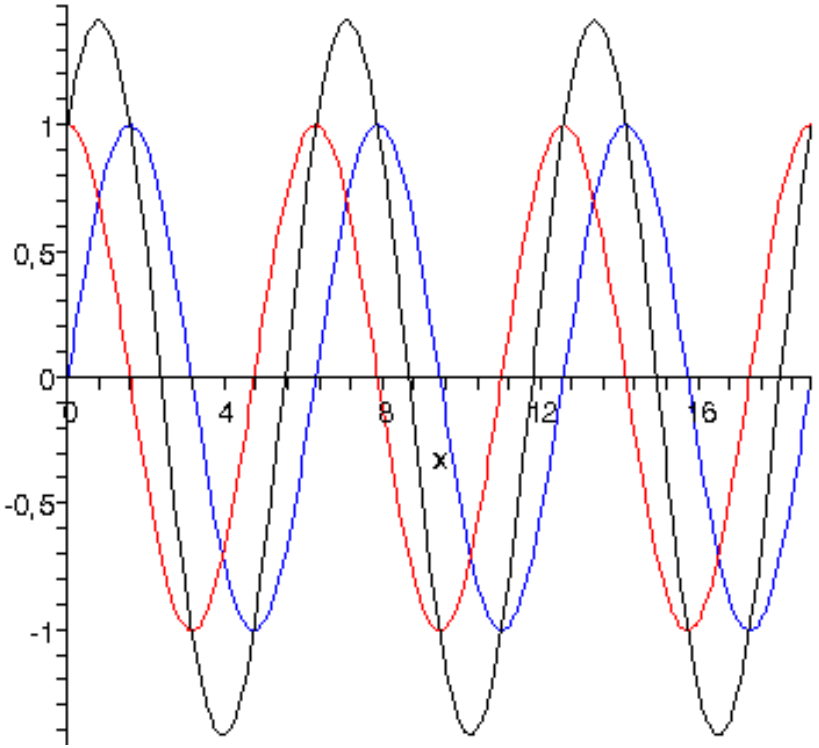
- 8 états de phase et 2 valeurs d'amplitude

Constellation V29



Modulation combinée en quadrature

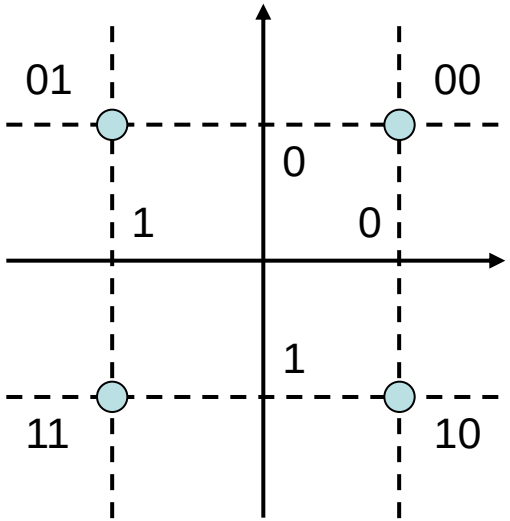
- Porteuses en quadrature : addition de deux porteuses de fréquence f_0 en quadrature, on obtient une seule porteuse, toujours de fréquence f_0



Modulation combinée en quadrature

Modulation de phase

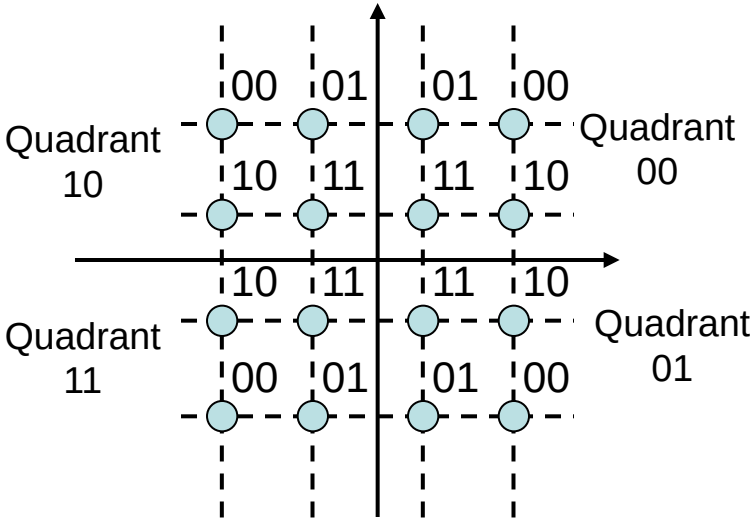
4 états (2 bits)



Quadrature Amplitude Modulation

QAM 16

16 états (4 bits)



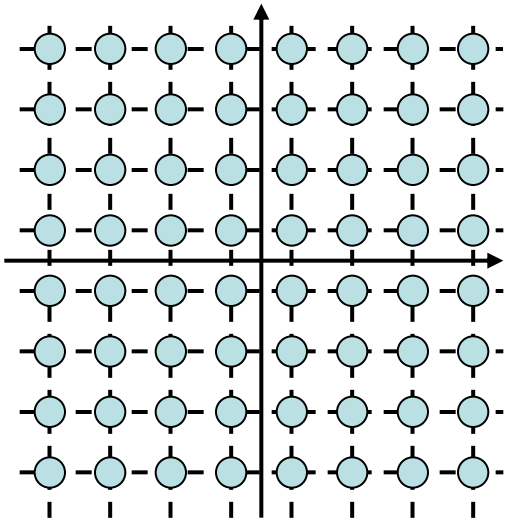
Modulation des 2 porteuses

Modulation combinée en quadrature

Quadrature Amplitude Modulation

QAM 64

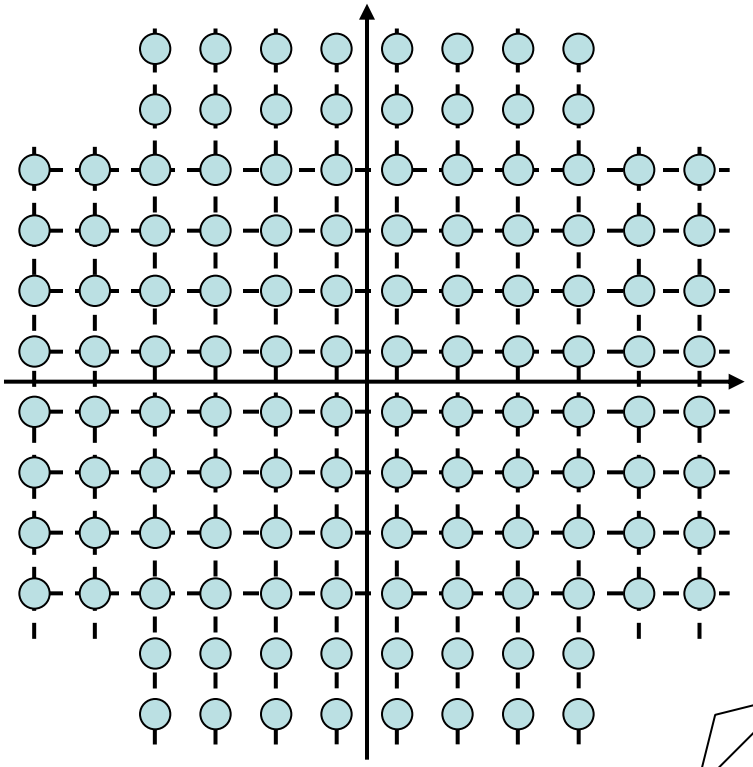
64 états (6 bits)



Quadrature Amplitude Modulation

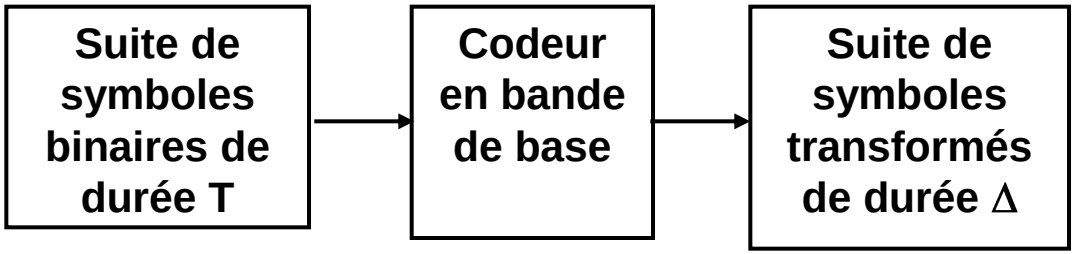
QAM 128

128 états (7 bits)



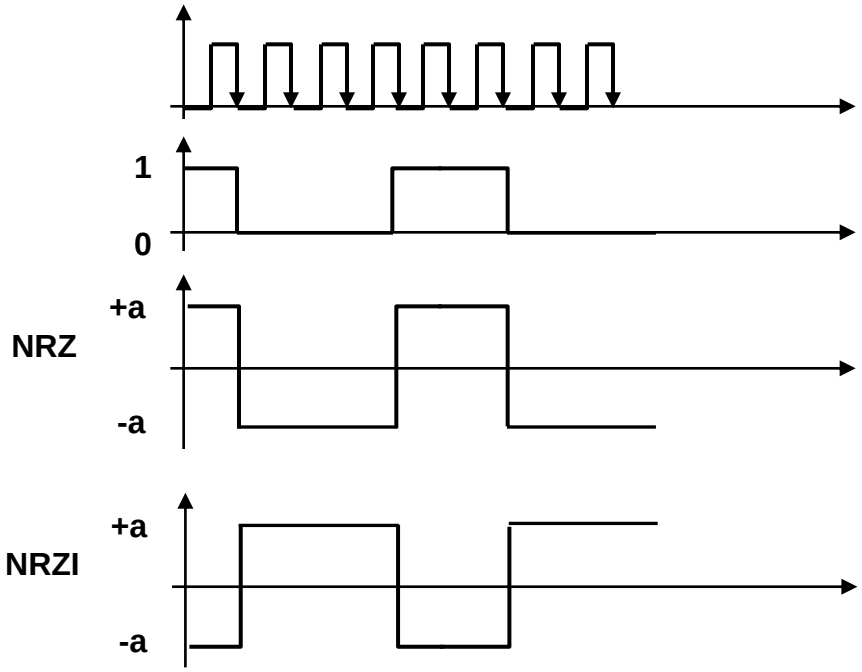
Transmission en bande de base

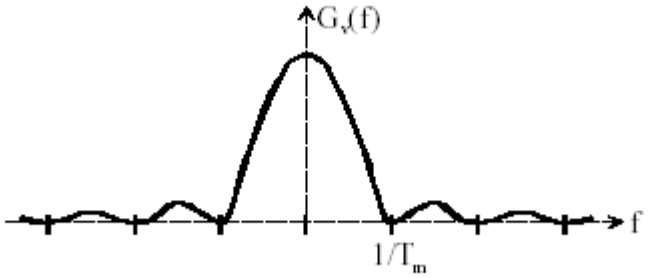
- La transmission directe de la suite des symboles binaires n'est pas possible $\Rightarrow$ le codage permet **d'adapter le signal au support de transmission**.
- Un signal en bande de base ne subit pas de transposition en fréquence : **ETCD = simple codeur**
- Le signal occupe alors toute la bande passante disponible. Les principaux avantages sont la simplicité et le coût (pas de phase de modulation/démodulation).
- La suite des symboles transformés appartient à un alphabet fini $\Delta = n \times T$, $(n \in N, n>0)$.



Code NRZ et NRZI

- Pour le codage NRZ (No Return to Zero), le signal binaire est transposé en tension pour éviter les valeurs nulles: 0 ⇒ -a et 1⇒ +a





Répartition de la puissance en fonction de la fréquence

- Le spectre de puissance du signal NRZ est concentré au voisinage des basses fréquences

⇒ **mauvaise transmission par le support**

Utilisé dans les normes **V24, RS232, RS421, RS422, RS485**

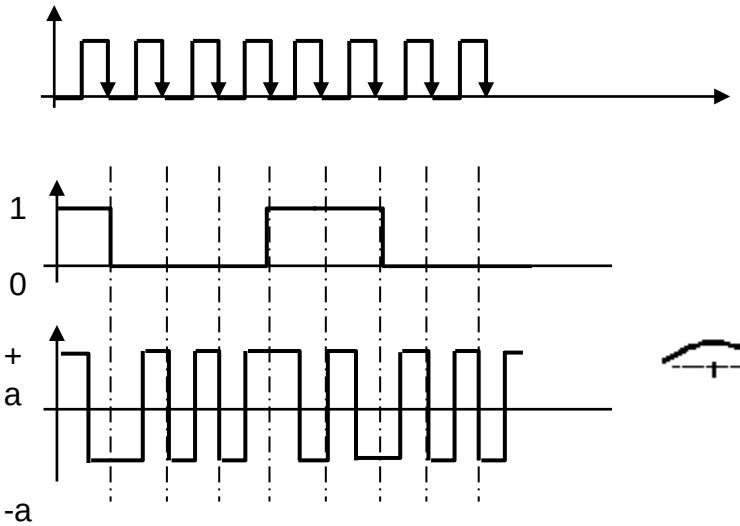
Code NRZI avec “bit stuffing”

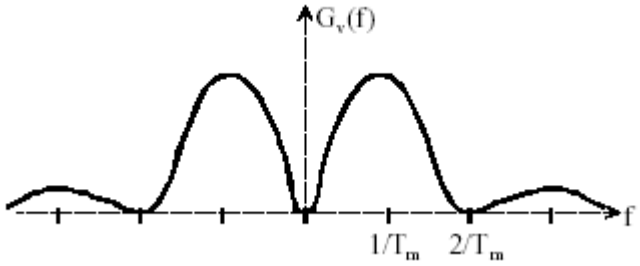
(NRZ Inverted with bit stuffing)

- Code du bus **USB**.
- Le bus commence en ‘idle state’ (état haut à +A). Chaque fois qu’un bit est "1", il n'y a pas de changement de l'état de la ligne. Chaque fois qu'un bit est "0", la ligne change d'état (toggle). Lorsque six "1" consécutifs sont transmis, un "0" artificiel est inséré afin de garantir la récupération d'horloge.

Code biphasé (Manchester)

- Introduction de transition au milieu de chaque intervalle, par exemple: 0 ⇒ front montant et 1⇒ front descendant
- **Signal transmis = signal binaire ⊕ horloge**





Rappel : ⊕ = XOR = OU Exclusif

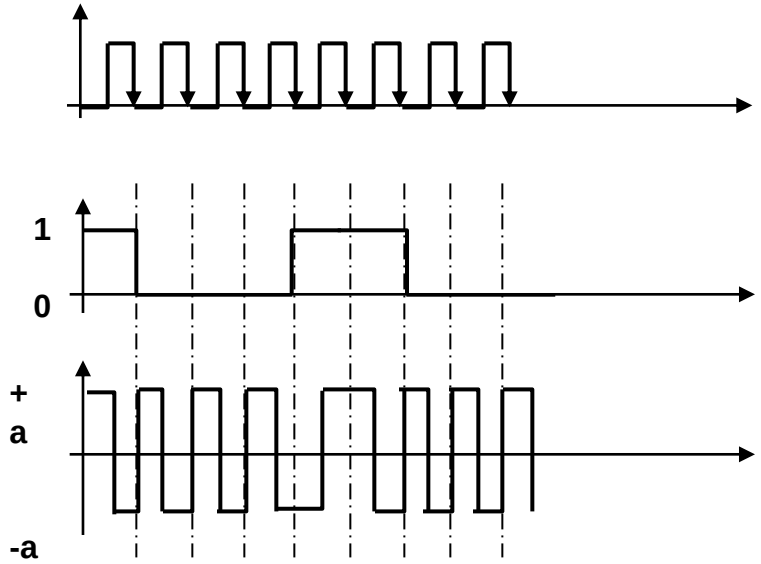
a	b	a ⊕ b
0	0	0
0	1	1
1	0	1
1	1	0

Une transition à chaque bit ⇒ transposition hautes fréquences, transmission de l'horloge (embrouillage), synchronisation.

- **Spectre de puissance étalé** sur la bande de fréquence $[0;2/T_m]$.
- **Ethernet** sur câble coaxial, signal **RDS** (Radio Data System)

Code biphasé différentiel

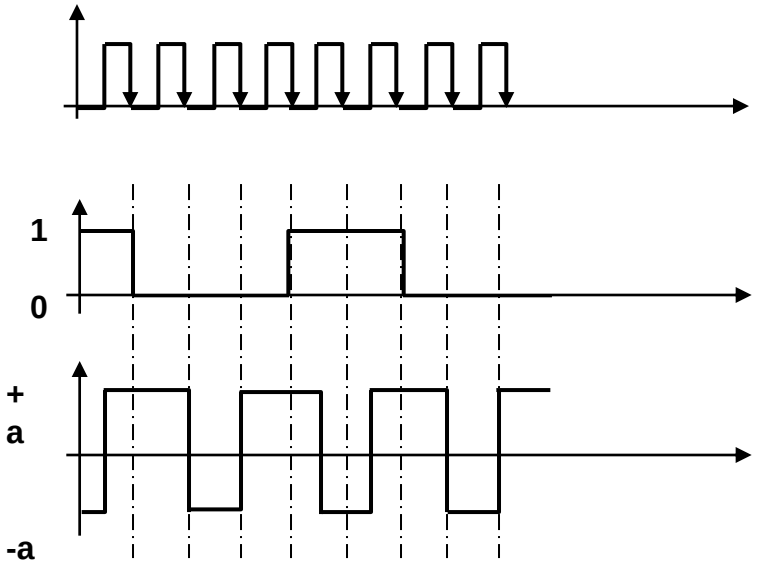
- On applique une transition systématique au milieu de chaque bit, pas de transition pour "1", une transition pour "0".

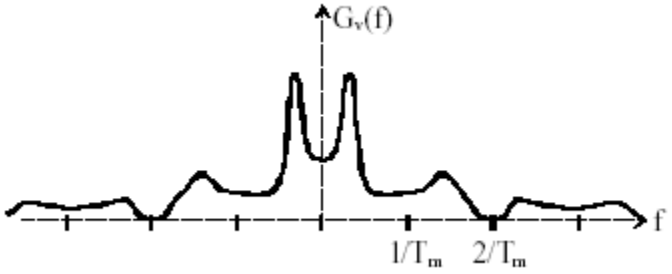


- Une transition à chaque bit $\Rightarrow$ transposition hautes fréquences, transmission de l'horloge (embrouillage), synchronisation.

Code Miller

- On applique une transition au milieu du bit "1", pas de transition au milieu du bit "0", une transition en fin de bit "0" si celui-ci est suivi d'un autre "0".





- Jamais 2 bits sans une transition $\Rightarrow$ transposition hautes fréquences mais moins que code biphasé (Manchester ou différentiel)

Code 4B/5B

- Chaque groupe de 4 bits est transformé en un groupe de 5 bits avec pas plus de deux 0 de suite

1111	**11101**
1110	**11100**
1101	**11011**
1100	**11010**
1011	**10111**
1010	**10110**
1001	**10011**
1000	**10010**
0111	**01111**
0110	**01110**
0000	**11110**
0001	**01001**
0101	**01011**
0100	**01010**
0011	**10101**
0010	**10100**

Autres codes

- Il existe bien d'autres variantes de codage en bande de base notamment avec des codages en symboles ternaires ou quaternaires
- Propriétés des codes en bande de base :
  - Évitent les fréquences nulles (séquences de valeur constante)
  - Ajoutent des transitions (embrouillage de l'horloge)
  - S'adaptent au support de transmission

Multiplexage

- **Objectif** : optimiser l'usage des canaux de transmission pour un transit simultané du maximum d'informations $\Rightarrow$ **partage** (**multiplexage**) du support physique de transmission entre plusieurs signaux.

- Ces techniques peuvent se classer en deux grandes catégories:
  - **multiplexage fréquentiel** :

    MRF (Multiplexage par Répartition de Fréquence) ou

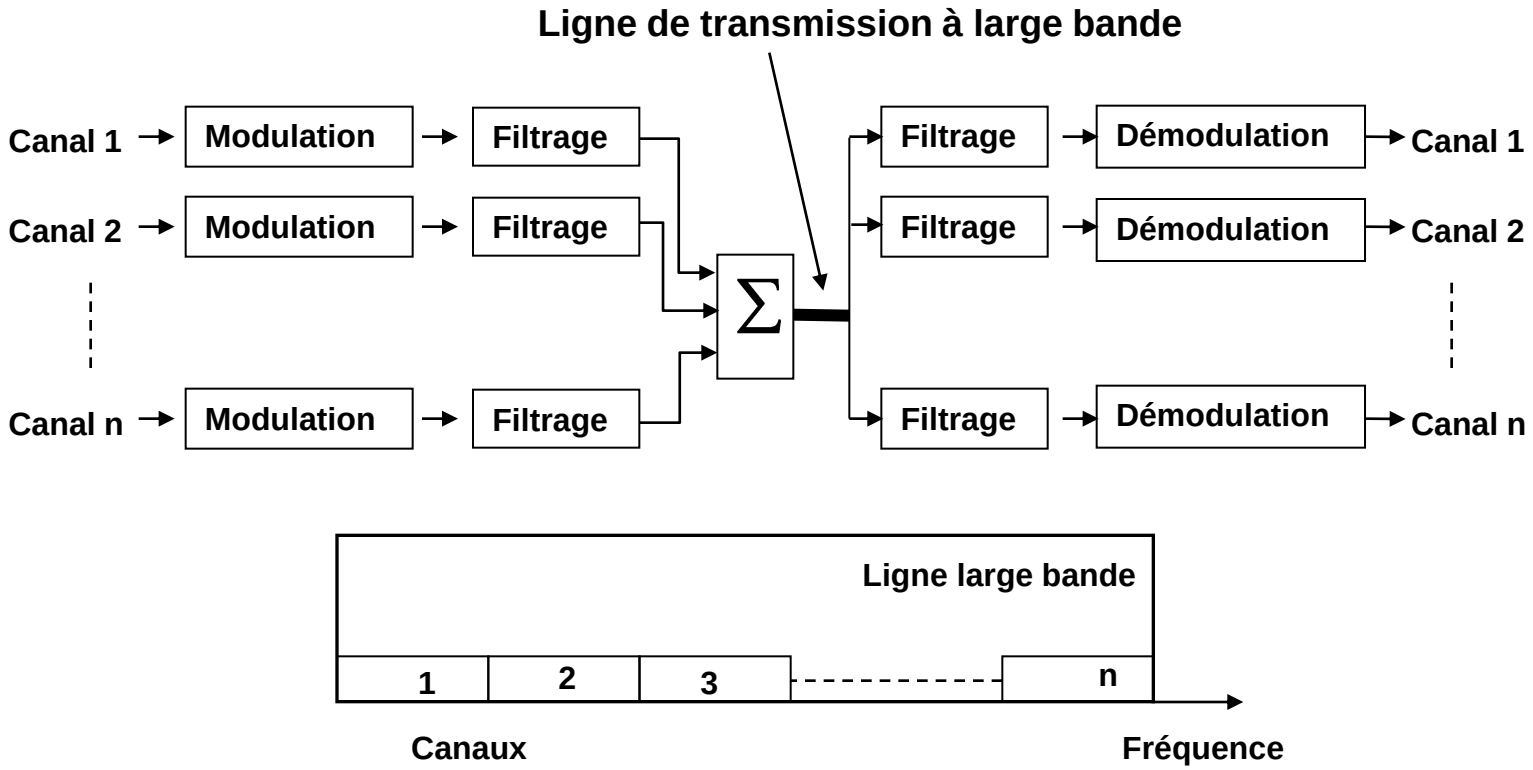
    FDM (Frequency Division Multiplexing)
  - **multiplexage temporel** :

    MRT (Multiplexage à Répartition dans le Temps) ou

    TDM (Time Division Multiplexing)

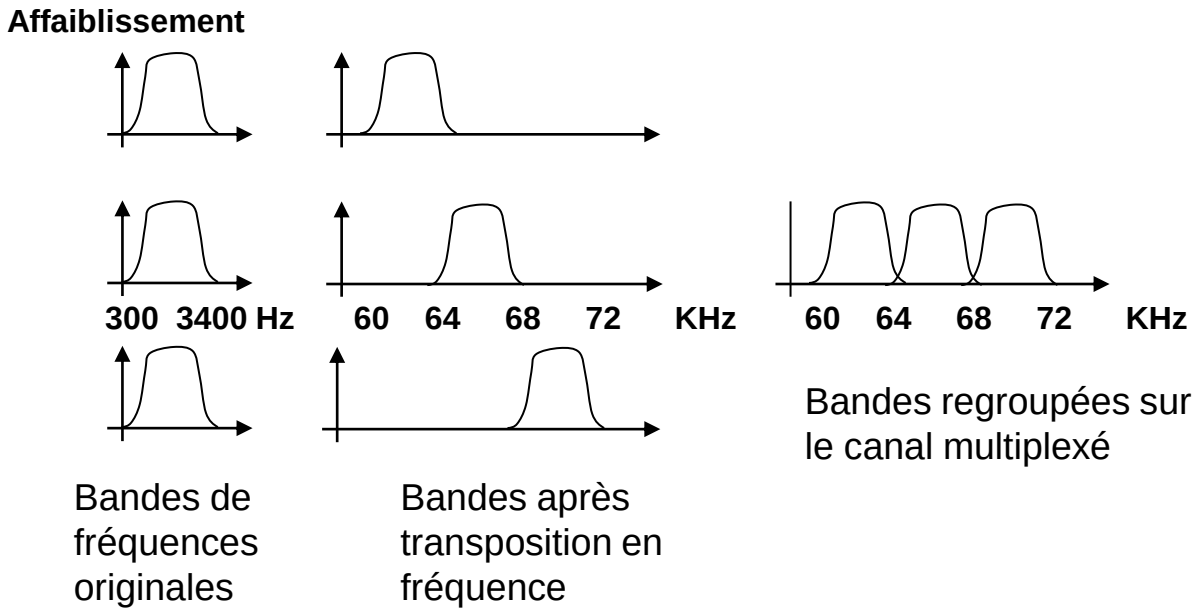
Multiplexage en fréquences

- **Partage de la bande de fréquences** disponible en plusieurs canaux (ou sous-bandes) plus étroits : en permanence chacun de ces canaux est affecté à un "utilisateur" exclusif



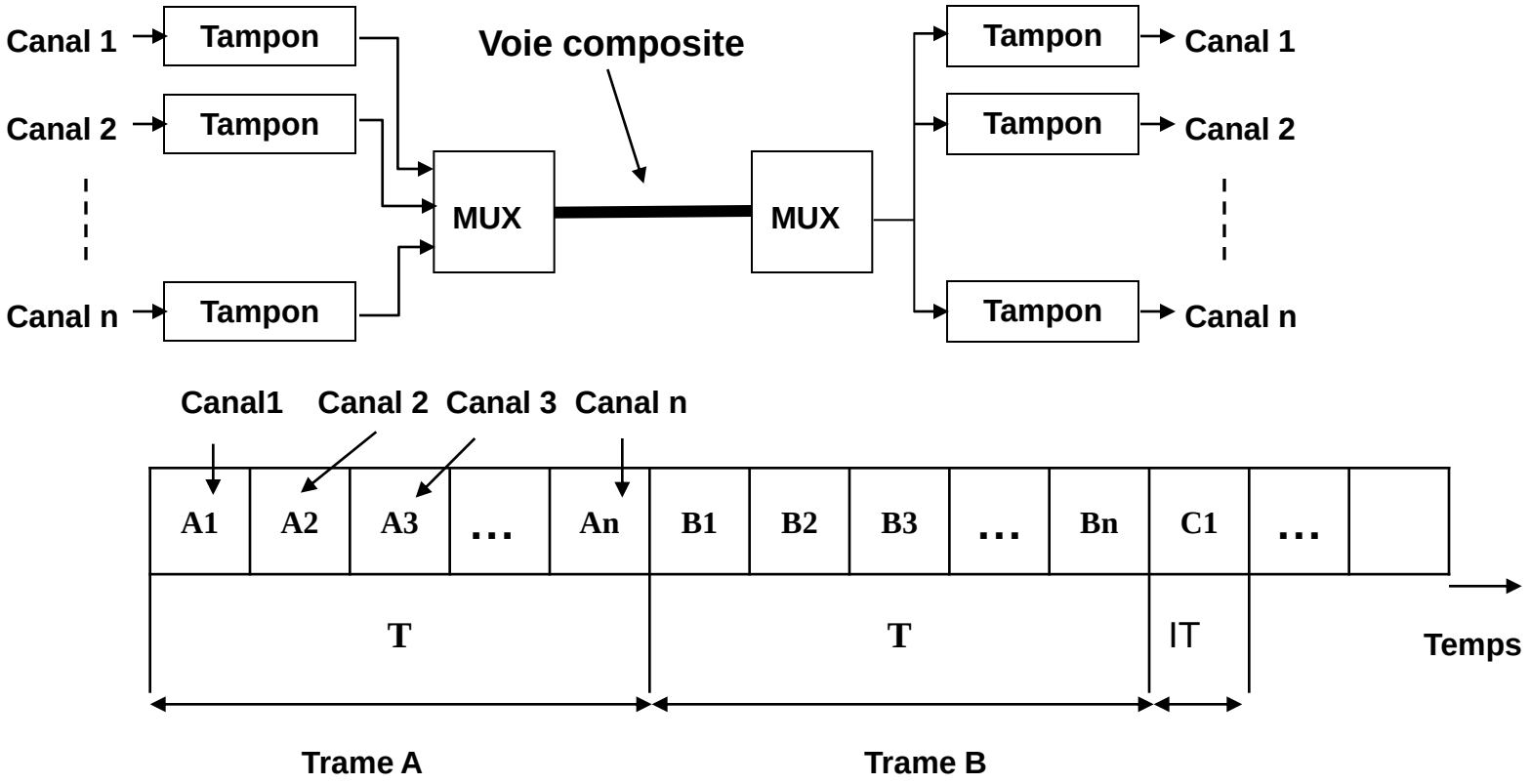
Multiplexage fréquentiel de trois canaux téléphoniques

- **3 liaisons téléphoniques multiplexées avec technique FDM.**
- Des filtres appropriés limitent la bande passante à 3100 Hz par canal téléphonique.
- Pour assurer un multiplexage correct, une bande de fréquences de 4000 Hz est attribuée à chaque canal afin de bien les séparer les uns des autres.



Multiplexage temporel

⇒ chaque "utilisateur" a pendant un court instant et à tour de rôle, la totalité de la bande passante disponible (généralement réservé aux signaux numériques).



Multiplexage temporel

- La vitesse de transmission des voies bas débit (*d*) est fonction de la vitesse de transmission de la ligne (*D*) et du nombre de voies *n*

  d=D/n

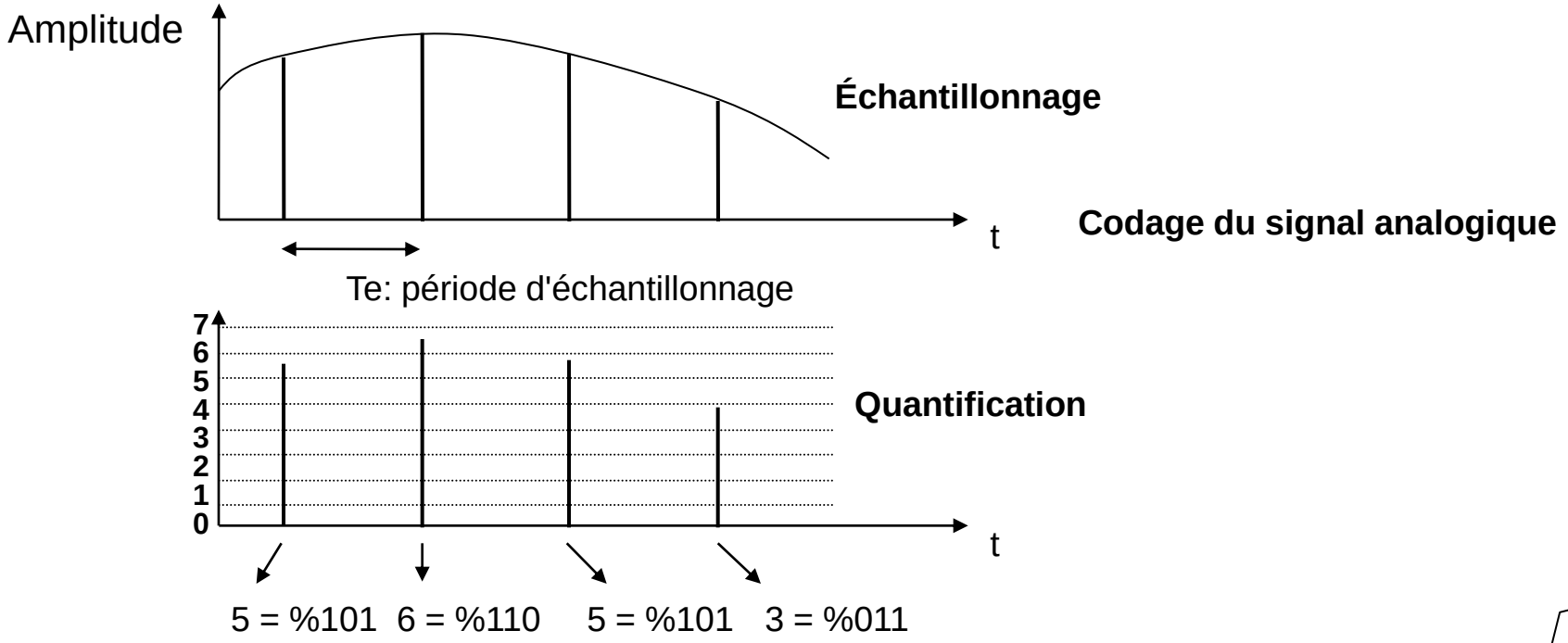
- La période *T* des trames est fonction du nombre de voies et de l'intervalle de temps élémentaire IT.

  T= n x IT

Modulation par impulsions codées (MIC)

Multiplexage temporel pour les transmissions téléphoniques.

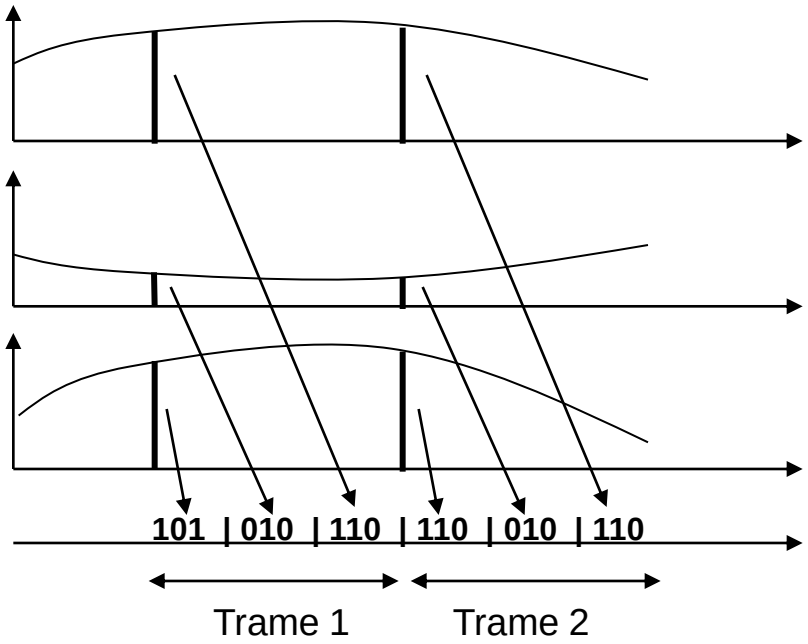
- échantillonnage des signaux analogiques de chacune des voies;
- quantification et codage des échantillons multiplexés pour obtenir un signal numérique.
- multiplexage temporel des échantillons des différentes voies;



Modulation par impulsions codées (MIC)

- Les échantillons sont ensuite multiplexés pour former un ensemble de trames.

Multiplexage temporel des échantillons de trois voies



Multiplexage temporel statistique

- Multiplexage temporel simple : tranches de temps pas toujours utilisées ⇒ des bits ou des caractères de remplissage sont insérés.
- Multiplexage temporel statistique ou asynchrone
  (ATDM: Asynchronous Time Division Multiplexing)
- **Allocation dynamique des tranches de temps aux seules voies qui ont des données à transmettre à un instant donné.**

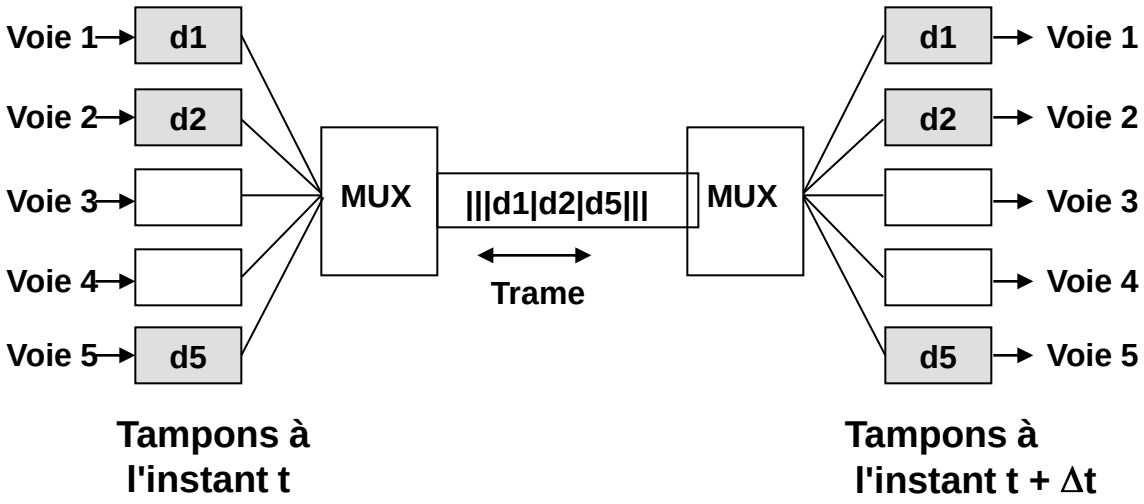
⇒ permet de raccorder plusieurs équipements sur une seule ligne, même si le débit cumulé de chaque voie est supérieur au débit maximum de la ligne.

⇒ Le multiplexeur intègre un microprocesseur et des mémoires tampon:

il permet **des débits et des paramètres de transmission différents sur chaque voie** ou sous-canal et à chaque extrémité.

Multiplexage temporel statistique

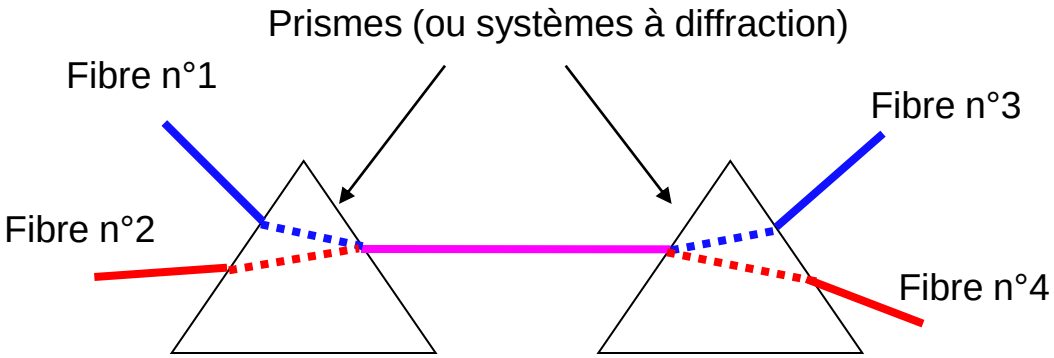
- Le multiplexeur :
  - détecte les tampons non vides,
  - prélève les données mémorisées,
  - supprime les bits non significatifs dans le cas d'une transmission asynchrone (start, stop, parité),
  - comprime éventuellement les données et les insère dans les trames de la voie composite.



Multiplexage en longueur d'onde

WDM (Wavelength Division Multiplexing)

⇒ proche du multiplexage fréquentiel



- Entrée : 2 fibres : flux lumineux d'énergie et de bande de fréquences différentes

⇒ Multiplexage WDM complètement passif ⇒ très haute fiabilité.

- Fibre W ~ 25000 GHZ
- un signal: qq GHz (limite = pb de conversion lumière/électricité)

2. Théorie de l'information

- Premières tentatives de définition de mesure de l'information 1920
- Théorie de l'information : à partir de 1948, travaux de Shannon
- Théorie de l'information : discipline fondamentale qui s'applique dans le domaine des communications.
  - déterminer les limites imposées par les lois de la nature lorsqu'on doit stocker ou transmettre le contenu d'une source (d'information),
  - proposer des dispositifs permettant d'atteindre ou d'approcher ces limites.
  - La théorie de l'information ne cesse de se développer car les exigences actuelles s'orientent vers une augmentation constante de l'information à stocker ou à transmettre.

Théorie de l'information

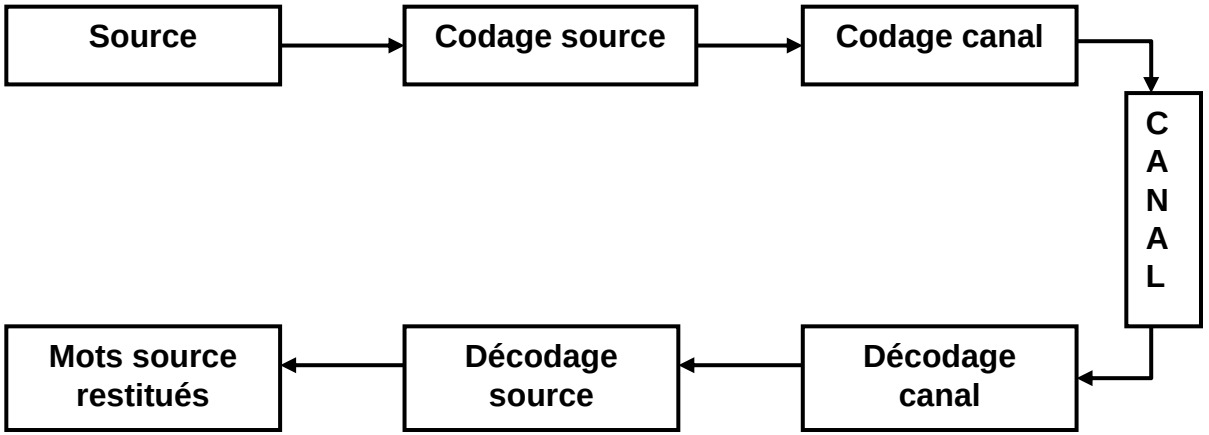
- Compression du contenu de la source d'information nécessaire.

  Peut s'envisager sous deux formes :

  – sans perte d'information;

  – avec perte d'information.

Représentation du schéma d'une communication



- Le codage de source consiste à éliminer les redondances de la source afin d'en réduire le débit binaire.
- Le codage de canal a un rôle de protection contre les erreurs (dues à la transmission sur le canal) qui est assuré en ajoutant de la redondance (codes correcteurs d'erreurs).
- Les points de vue codage de source et codage de canal sont donc fondamentalement différents.

Définition de l'«information»

- **Incertitude d’un événement ou self-information ou quantité d’information.**
- Une **information** est un élément de connaissance qui peut être conservé (mis en mémoire), traité (traitement de l’information) ou transmis.
- La difficulté rencontrée pour définir la quantité d’information relative à un événement est liée au **caractère subjectif** de l'information effectivement apportée par la réalisation de cet événement.
- La **quantité d’information** est relative à un **contenu statistique**, plus grande si on ne s'attend pas à ce que l'évènement se réalise.

Définition

- Soit un événement *E*, une probabilité d'apparition *p(E)*, la **quantité d'information *I(E)*** liée à l'apparition de *E* s'exprime :

$$I(E) = -\log [P(E)]$$

- **Propriétés** :
  - *I(E)* est d'autant plus grande que *P(E)* est petite ;
  - Si *E* est toujours réalisé *I(E)=0* ;
  - Si *E* est très rare, *P(E)* est proche de 0, *I(E)* tend vers l'infini ;
  - $P(E) \leq 1$ donc $I(E) \geq 0$.
- Logarithme
  - népérien pour les signaux continus :
  - à base 2 pour les signaux discrets
    (l'unité est le **bit**, abréviation de *binary unit*).

Exemples

- Un événement qui a une chance sur 2 de se produire conduit à une quantité d'information de
  $-\log_2 (1/2) = 1$ bit
- Une source peut délivrer 8 événements équiprobables : **$I_{total} = -8*\log_2 (1/8) = 24$ bits**
- *I(E)* peut être interprétée :
  - *a priori*,
    par l'incertitude qui règne sur la réalisation de E;
  - *a posteriori*,
    par l'information apportée par la réalisation de E.

Information mutuelle de 2 évènements

- L'information apportée par *F* sur *E* est la diminution de l'incertitude sur *E* lorsque *F* est réalisé.
- **Information mutuelle**

  I(E;F) = $I_{F \to E} = I(E) - I(E / F) = I_{E \to F}$

  car E et F jouent le même rôle.

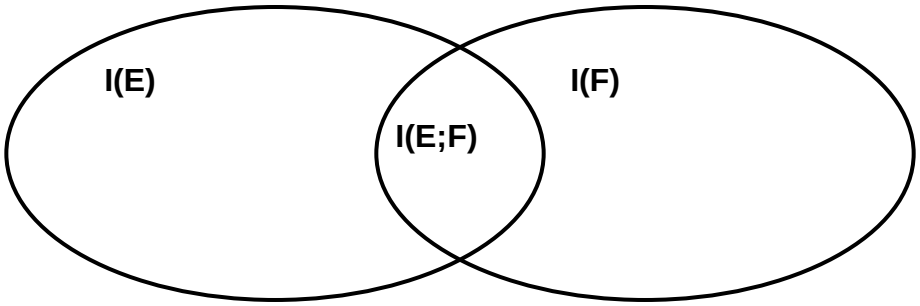
Démonstration :

$$I_{F \to E} = -\log[P(E)] + \log[P(E/F)]$$
$$= -\log[P(E)] + \log[P(E \cap F)/P(F)]$$
$$= -\log[P(E \cap F)/ P(E)P(F)] = I_{E \to F}$$

Information mutuelle de 2 évènements

- Si *E* et *F* sont **indépendants**,

  alors *P (F/E) = P(F) et* ***I(E;F)= 0***.
- On obtient alors : $I(E \cap F) = I(E) + I(F) - I(E;F)$
- On peut résumer les relations précédentes sur un diagramme de Venn :



Entropie

- **Entropie d'une variable aléatoire discrète**
  - Soit X une variable aléatoire à valeurs dans $\{x_1, x_2, ..., x_n\}$ telle que $p_i = P(X = x_i)$ *pour i* de 1 à *n*.
  - L'entropie de X notée H(X) est la moyenne des incertitudes calculée sur les événements *{X = xi}*

$$H(X) = -\sum_{i=1}^{i=n} p_i \log p_i$$

- **L'entropie d'une source discrète** est donc la quantité moyenne d'information par symbole et est exprimée en **bits par symbole**.

Entropie, Débit

Remarques :

- H(X) dépend de la loi de probabilité de X mais n'est pas fonction des valeurs prises par X;
- H(X) correspond à l'espérance mathématique de la variable aléatoire incertitude I(X) définie par $I(X) = -\log P(X)$;
- Exprimée en Shannons, H(X) représente le nombre moyen de bits nécessaires à la codification binaire des différentes réalisations de X.

- **Le taux (débit) d'information** est le nombre de symboles par unité de temps :

  T = n H(X) en **bits par seconde**.

Exemple

- On extrait au hasard une carte d'un jeu de 32 cartes.
- On a $H(X) = -32 \times 1/32 \times \log_2 1/32 = \log_2 32 = 5$ Sh.
- Pour savoir quelle carte a été extraite, on peut demander si sa couleur est rouge ou noire, s'il s'agit d'un cœur ou d'un carreau, etc.
  Les réponses à cinq questions peuvent être résumées par cinq bits ('1' pour oui et '0' pour non).
- Une autre façon de modéliser le problème consiste à attribuer un numéro (de 0 à 31) à chaque carte. L'écriture de ces numéros en base deux requiert
  $\log_2 32 = \log_2 2^5 = 5$ *bits.*

Propriétés de l'entropie

- L'entropie d'une variable aléatoire X à n valeurs possibles est **maximum** et vaut **log(n)** lorsque la **loi de X est uniforme**.

  En effet, l'incertitude sur X est la plus grande si toutes les valeurs possibles ont la même probabilité de se réaliser;

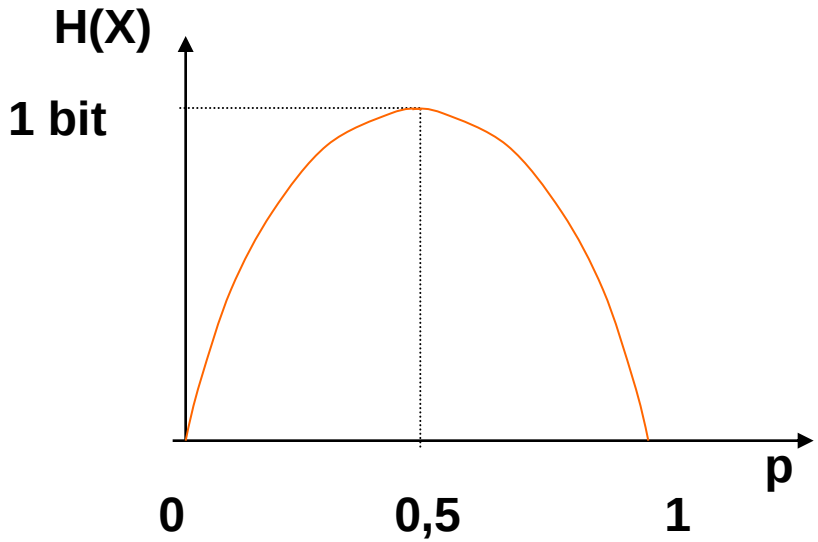
- L'entropie augmente lorsque le nombre de valeurs possibles augmente.

Exemple

- Soit une source binaire à 2 symboles : $\begin{vmatrix} X \\ P \end{vmatrix} = \begin{vmatrix} x_1 & x_2 \\ p & 1-p \end{vmatrix}$

  $H(X)= -p \log[p] – (1-p) \log(1-p)$

  - p=0 (x_2 toujours réalisé) → H(x)=0
  - p=1 (x_1 toujours réalisé) → H(x)=0
  - p=0,5 (x_1 et x_2 équiprobables) → H(x)=1



Entropie et information liées à un couple de variables

- Soient X et Y deux variables aléatoires discrètes à valeurs dans $\{x_1, x_2, ..., x_n\}$ et $\{y_1, y_2, ..., y_m\}$.

- Si on désigne par $p_{ij} = P(X = x_i \cap Y = y_j)$ la loi du couple (X, Y), on peut définir les entropies conditionnelles et l'information mutuelle :

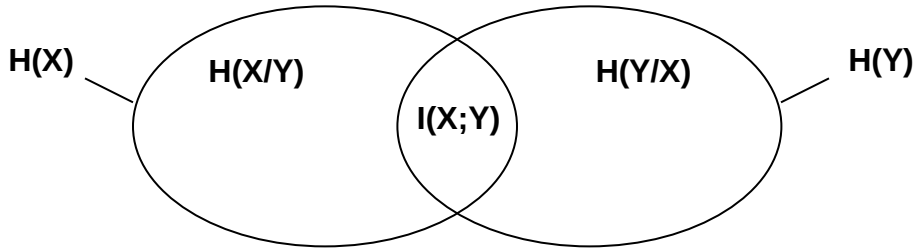
  $H(X/Y)$ représente l'incertitude sur X lorsqu'on connaît Y.

- De même l'information mutuelle moyenne entre X et Y peut s'écrire : $I(X;Y) = H(X) - H(X / Y) = H(Y) - H(Y/X)$

- I(X;Y) correspond à la diminution de l'incertitude sur X (resp. Y) lorsqu'on connaît Y (resp. X).

Propriétés

- **L'information mutuelle moyenne de X et de Y est toujours positive** (ce n'est pas le cas pour l'information mutuelle entre deux événements qui prend des valeurs négatives lorsque la réalisation de l'un des événements rend l'autre moins probable);
- **Le conditionnement diminue l'incertitude.** En d'autres termes cela signifie que $H(X) \geq H(X / Y)$
- $H(X) + H(Y) = H(X,Y) + I(X;Y)$
- $H(X,Y) = H(X) + H(Y/X) = H(Y) + H(X/Y)$
- Diagramme de Venn :



Propriétés

Dans le cas particulier où X et Y sont indépendantes, on obtient :

- $H(X/(Y = y)) = H(X)$
- $H(X/Y) = H(X)$
- $I(X;Y) = 0$
- $H(X) + H(Y) = H(X,Y)$

3. Source discrète

Source discrète	Suite de variables aléatoires, ayant chacune un nombre fini de réalisations.
Symbole – Lettre	Le plus petit élément irréductible contenant une information
Alphabet	Ensemble de lettres que peut délivrer la source
Mot	Suite finie de lettres
Vocabulaire	Ensemble des mots possibles avec l'alphabet de la source
Codage	Correspondance entre deux vocabulaires
Source stationnaire (ou sans mémoire)	Source dans laquelle la probabilité d'apparition d'un caractère ne dépend pas de ce qui est apparu auparavant. La loi de probabilité ne dépend pas de l'instant considéré

Entropie d'une source discrète

- Comme pour une variable aléatoire, on souhaite caractériser une source discrète par son entropie.
- Si S est une source sans mémoire (les valeurs S_i émises par la source sont indépendantes), alors on appelle $H(S_i)$ l'entropie par lettre source S_i

$$H(S_i) = -p(S_i) \log (p(S_i)).$$

Codage de source

- Soit S une source discrète à N valeurs possibles et d'entropie $H(S) < \log_2 N$.
  L'entropie relative $H_r(S) = H(S)/\log_2 N$ étant < 1,
  la redondance $r = 1 - H_r(S)$ est différente de zéro.

  Codage source = élimination de la partie redondante de l'information délivrée par la source = réduction du débit binaire de la source tout en conservant l'information (compression sans perte d'information)

- Deux types de codes peuvent être utilisés :
  - **Codes à longueur fixe**. Tous les mots ont même longueur (même nombre de symboles)
  - **Codes à longueur variable**. La longueur des mots varie en fonction de leur fréquence d'apparition. **Un mot sera d'autant plus long que sa probabilité d'apparition sera petite**.

Généralités sur les codes

- Soit B un alphabet de référence (un ensemble de lettres),
  en général B est l'alphabet binaire $B = \{0,1\}$.

- En concaténant des lettres de l'alphabet B, on forme des mots dont la longueur correspond au nombre de lettres utilisées. Un code C est un ensemble de mots.

Exemples :

- $B = \{0,1\}$ est un code de longueur fixe =1
- $C = \{00,01,10,11\}$ est un code de longueur fixe =2

Généralités sur les codes

On dira qu'un code est **uniquement déchiffrable (ou à décodage unique)** si toute suite de mots code ne peut être interprétée (décodée) que d'une seule manière.

Exemples :

- {0,11,010} est uniquement déchiffrable.
- {0,1,101} n'est pas uniquement déchiffrable car 101 peut être interprété comme 1-0-1 ou 101.

Généralités sur les codes

- Un **code préfixe** est un code pour lequel aucun mot n'est le début d'un autre mot. Ce code est **uniquement déchiffrable** et est dit à **décodage instantané**.
- Construction d'un code préfixe : utiliser un arbre dont chaque branche représente une lettre de l'alphabet élémentaire (alphabet de référence).
  Il faut alors choisir les mots de telle sorte que pour tout couple de mots, le chemin permettant de conduire à l'un des mots ne soit pas inclus dans le chemin menant à l'autre mot.

Exemples :

- {0,10,110,111} est un code préfixe.
- {0,11,110,111} n'est pas un code préfixe car le chemin menant à 11 est inclus dans les chemins menant à 110 et 111.

Codage d'une variable aléatoire discrète

- **Longueur moyenne des mots code :**

$$\bar{n} = \sum_{i=1}^{M} p_i n(i)$$

avec :

- ***M*** **= nombre de mots code;**
- ***pi*** **= fréquence relative d' apparition du mot code n° i;**
- ***n*****(*****i*****) = longueur du mot code n° i.**

Codage d'une variable aléatoire discrète

- **Théorème 1**

Pour tout codage d'une variable aléatoire X par un code uniquement déchiffrable (sur un alphabet de référence à b lettres), la longueur moyenne des mots code vérifie:

$$H(X) \leq \bar{n} \log(b)$$

où la base du logarithme coïncide avec celle utilisée pour la mesure de *H(X)* (entropie de X).

- **Théorème 2**

Pour toute variable aléatoire X, **il existe un code préfixe** tel que :

$$\bar{n} \log(b) < H(X) + \log(b)$$

Codage d'une variable aléatoire discrète

- Finalement, on déduit que pour tout codage d'une variable aléatoire X, il existe un code uniquement déchiffrable qui vérifie:

$$\frac{H(X)}{\log(b)} \leq \bar{n} < \frac{H(X)}{\log(b)} + 1$$

➢ **La limite inférieure ne peut être dépassée par aucun code**.

Exemples de codes à longueur variable utilisés pour compacter une source

Code Morse

= code ternaire : point, trait et pause

- A chaque lettre une succession d'émissions de courants brefs (points) et longs (traits) séparés par des pauses courtes.
- Entre chaque lettre est insérée une pause plus longue
- Les mots sont séparés par une pause deux fois plus longue que celle disposée entre les lettres.

➢ **Aux lettres les plus fréquentes sont affectées les mots code les plus courts**

Code Morse

Lettre	Fréq.	Code	Durée
A	0,0642	. _	9
B	0,0127	_ . . .	13
C	0,0218	_ . _ .	15
D	0,0317	_ . .	11
E	0,1031	.	5
F	0,0208	. . _ .	13
G	0,0152	_ _ .	13
H	0,0467	. . . .	11
I	0,0575	. .	7
J	0,0008	. _ _ _	17
K	0,0049	_ . _	13
L	0,0321	. _ . .	13
M	0,0198	_ _	11
ESP	0,1859		6

Lettre	Fréq.	Code	Durée
N	0,0574	_ .	9
O	0,0632	_ _ _	15
P	0,0152	. _ _ .	15
Q	0,0008	_ _ . _	17
R	0,0484	. _ .	11
S	0,0514	. . .	9
T	0,0796	_	7
U	0,0228	. . _	11
V	0,0083	. . . _	13
W	0,0175	. _ _	13
X	0,0013	_ . . _	15
Y	0,0164	_ . _ _	17
Z	0,0005	_ _ . .	15

Durée du point avec pause courte = 2, durée du trait avec pause courte = 4, durée de la pause longue = 3.

Code de Shannon-Fano

1er code utilisé pour exploiter la **redondance** d'une source (années 50)

Algorithme

1. Classer les différents symboles à coder suivant l'ordre décroissant de leur probabilité
2. Diviser ces symboles en deux sous-groupes de telle sorte que les probabilités cumulées de ces deux sous-groupes soient aussi proches que possible l'une de l'autre.
3. Affecter le chiffre binaire "0" au 1er sous-groupe et "1" au 2ème sous groupe.
   Les mots code du premier sous-groupe commenceront tous par "0" tandis que ceux du second commenceront par "1".
4. Recommencer à 1 jusqu'à ce que les sous-groupes ne contiennent qu'un élément.

➢ Tous les symboles source ont alors un mot code.

Code de Shannon-Fano - Exemple

- Exemple :

Soit une source avec cinq symboles A, B, C, D, E

Symbole	Fréquence d'apparition	Première division	Deuxième division	Troisième division	Quatrième division	Code binaire
A	**15/39**	**0**	**0**			**00**
B	**7/39**	**0**	**1**			**01**
C	**6/39**	**1**		**0**		**10**
D	**6/39**	**1**		**1**	**0**	**110**
E	**5/39**	**1**		**1**	**1**	**111**

La **longueur moyenne** d'un mot code est :

2x15/39+ 2x7/39+ 2x6/39+ 3x6/39+ 3x5/39 = **2,28** *bits*

- Meilleur qu'un code binaire simple à longueur fixe: **3 bits**
- Entropie de la source (limite) $\mathbf{H(X) = -\sum_{i=1}^{i=n} p_i \log_2 p_i = 2{,}18\ Sh}$

Code de Huffman

- La construction de l'arbre permettant d'affecter aux n lettres source un mot code s'effectue de façon ascendante (contrairement au code de Shannon-Fano).

Algorithme

1. Classer les lettres source suivant l'ordre décroissant de leur probabilité.
2. Créer un nouveau symbole à partir des deux lettres source de probabilités les plus faibles. On lui affecte une probabilité égale à la somme des probabilités des deux lettres source.
3. Considérant les n-1 lettres source restantes, revenir à la première étape jusqu'à ce qu'il ne reste plus que 2 lettres source.
4. Affecter "0" au 1er sous-groupe et "1" au 2ème sous groupe à chaque étape.

Code de Huffman- Exemple

- Même source que l'exemple précédent

1		→ 8	2 ←		→ 7	3 ←		→ 6	4 ←		5
A	15/39	1	A	15/39	1	A	15/39	1	BCDE	24/39	0
B	7/39	000	D E	11/39	01	BC	13/39	00	A	15/39	1
C	6/39	001	B	7/39	000	DE	11/39	01			
D	6/39	010	C	6/39	001						
E	5/39	011									

La **longueur moyenne** d'un mot code est :

1x15/39 +3x (1 -15/39) = **2,23** *bits*

- Toujours meilleur que *Shannon-Fano* (***2,28*** *bits)*
- Entropie de la source (***2,18*** *Sh*) = limite

Codage source - remarques

- Le codage de Huffman (comme celui de Shannon-Fano) nécessite la connaissance des probabilités a priori des lettres source.

  Sinon estimation en mesurant les fréquences d'apparition dans le message à transmettre --> transmettre l'arbre des codes pour que le décodage soit possible.

- Pour améliorer les performances, élever **l'ordre d'extension de la source** (**ordre n** = prendre **n symboles** à la fois pour les coder). Inconvénient : l'arbre des codes à transmettre plus long.

  Exemple : si la source émet des symboles du code ASCII étendu, il y a 256 caractères. extension d'ordre 2 = d'une table de probabilités de 256^2 = 65536 nombres!

  gain obtenu par la compression inutile pour des messages courts.

➢ Pour pallier cet inconvénient, on a introduit le **code de Huffman adaptatif** dont le principe repose sur la mise à jour des statistiques sur les symboles au fur et à mesure de la lecture du message.

Codage arithmétique

- Principe : affecter à l'ensemble du message un seul nombre en virgule flottante
- Exemple : BALLE

Symbole	A	B	E	L
P_i	1/5	1/5	1/5	2/5
Intervalle	[0;0,2[	[0,2;0,4[	[0,4;0,6[	[0,6;1[

1. Probabilités
2. Intervalles
3. B --> [0,2;0,4[
4. BA --> [0,2+0,2*0; 0,2+0,2*0,2[= [0,2; 0,24[
5. BAL --> [0,2+0,04*0,6; 0,2+0,04*1[= [0,224; 0,24[
6. … (valeur soulignée = longueur de l'intervalle précédent)

➢ **La borne inférieure code le message : 0,23616**

Décodage arithmétique

Symbole	A	B	E	L
P_i	1/5	1/5	1/5	2/5
Intervalle	[0;0,2[	[0,2;0,4[	[0,4;0,6[	[0,6;1[

1. 0,23616 --> B (0,23616 - 0,2)/0,2 = 0,1808
2. 0,1808 --> A (0,1808 - 0)/0,2 = 0,904
3. 0,904 --> L (0,904 – 0,6)/0,4 = 0,76
4. 0,76 --> L (0,76 – 0,6)/0,4 = 0,4
5. 0,4 --> E

- Le décodeur doit connaître les intervalles associés aux lettres

Code de Lempel-Ziv-Welch (LZW)

- LZW amélioration de LZ77 et LZ78, utilisé dans le format GIF, dans les fichiers .Z (commande Compress) et .gzip sous Unix et sur PC dans les fichiers .zip
- **Principe** : **fabriquer un dictionnaire au fur et à mesure** de la lecture du message à coder
- L'algorithme LZ78 En 1978, Jacob Ziv et Abraham Lempel (IEEE Transactions on Information Theory, "Compression of Individual Sequences via Variable-Rate Coding".

Algorithme :

  - Dictionnaire initialisé avec la chaîne vide placée à l'adresse 0.
  - Lire caractère par caractère le texte à compresser et vérifier si la chaîne ainsi construite se trouve bien dans un dictionnaire.
  - Si la concaténation (⊕) de la chaîne précédente (P) avec le dernier caractère lu (c) se trouve dans le dictionnaire, alors la lecture continue avec le caractère suivant.

    Sinon le couple (adresse de P, caractère c) est émis en sortie, puis la chaîne P ⊕ c est ajoutée au dictionnaire.

Exemple de compression LZ78

- **Texte à compresser:** "si_six_scies_scient_six" …

adresse dans le dictionnaire	chaîne dans le dictionnaire	caractère lu	signes émis
0		S	0S
1	S	I	0I
2	I	∧	0∧
3	∧	S,I	1I
4	SI	X	0X
5	X	∧,S	3S
6	∧S	C	0C
7	C	I,E	2E
8	IE	S,∧	1∧
9	S∧	S,C	1C
10	SC	I,E,N	8N
11	IEN	T	0T
12	T	∧,S,I	6I
13	∧SI	X,∧	5∧
14	X∧	C,Y	7Y

Exemple de compression LZ78

- L'ensemble des couples (indice,caractère) émis est donc le suivant:

  0S, 0I, 0_, 1I, 0X, 3S, 0C, 2E, 1_, 1C, 8N, 0T, 6I, 5_, 7Y

- Il y a 15 couples. Dans un couple, le caractère est codé sur 1 octet.

- Si l'indice de chaîne est codé sur 1 octet, on ne pourra "adresser" que 256 chaînes, ce qui est insuffisant. Si cet indice de chaîne est codé sur 2 octets, c'est 65536 chaînes qui seront accessibles. Le message ci dessus utiliserait alors (2+1)x15 = 45 octets

➢ L'efficacité de la compression est liée à la longueur des fichiers.

L'algorithme LZW

- Terry Welch à publié (1984) une variante de l'algorithme LZ78 dans IEEE Computer "A technique for High-performance Data Compression" : l'amélioration consiste à ne plus émettre un couple (adresse, caractère), mais seulement une adresse.

Algorithme :

  - Dictionnaire initialisé avec l'ensemble des caractères de l'alphabet.
  - Lire caractère par caractère le texte à compresser et vérifier si la chaîne ainsi construite se trouve bien dans un dictionnaire.
  - Si la concaténation (⊕) de la chaîne précédente (P) avec le dernier caractère lu (c) se trouve dans le dictionnaire, alors la lecture continue avec le caractère suivant.

    Sinon, l'adresse de P est émise en sortie, puis la chaîne P ⊕ c est ajoutée au dictionnaire et le caractère c est utilisé pour initialiser la chaîne P suivante.

LZW

- LZW15VC : amélioration de LZW capable de coder les adresses jusqu'à 15 bits.
  - Il commence à coder sur 9 bits et utilise un bit supplémentaire après avoir codé 256 valeurs supplémentaires.
  - Le dictionnaire est vidé lorsqu'il est complètement rempli.
  - Ceci permet de le rendre aussi performant que d'autres codes pour les messages courts.

Codes bloc

- Soit une source, avec un alphabet de taille K.
- Considérons la $n^{ième}$ extension de la source (concaténation de n lettres source).
- Soit un alphabet de référence comportant D lettres code.
- **Code bloc** formé en concaténant r lettres code. Pour ne pas perdre d'information, il faudra que le nombre de mots code soit au moins égal au nombre de mots source :

$$D^r \geq K^n$$

- Le **taux du code** ainsi constitué est le rapport $R = r/n$ qui représente le nombre de lettres code nécessaires pour représenter une lettre source.

Codes bloc

- **1er THÉORÈME DE SHANNON**

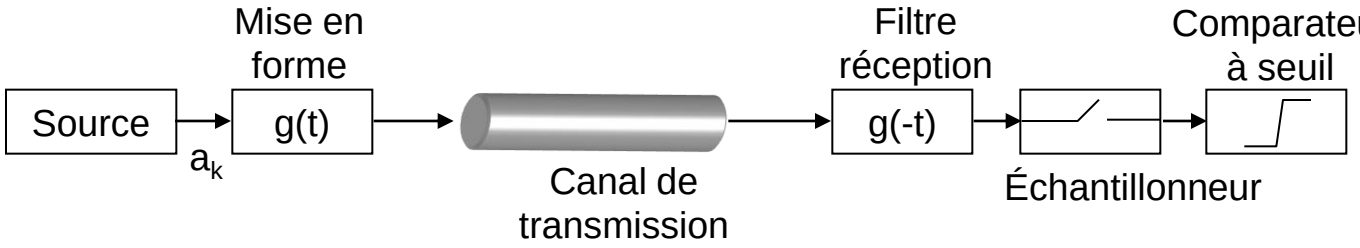
  Soit X une source discrète sans mémoire, d'entropie H(X).

  Alors il est possible de trouver un code bloc pour encoder les mots source de longueur n avec un taux R = r/n tel que la probabilité de ne pouvoir associer avec certitude un mot source à un mot code soit aussi petite que l'on veut. Il suffit pour cela que, d'une part n soit suffisamment grand et d'autre part que R vérifie:

$$R = \frac{r}{n} > \frac{H(X)}{\log D}$$

4. Canal discret

- **Modèle** = mise en cascade du **canal de transmission** et du **récepteur**
- Exemple d'une chaîne de **transmission numérique en bande de base** :



source binaire S, symboles a_k "mis en forme" par un filtre de réponse impulsionnelle *g(t)* de telle sorte que le signal à la sortie de ce filtre a pour expression

$$\sum_{k=-\infty}^{+\infty} a_k g(t - kT)$$

où 1/*T* représente le débit binaire de la source.

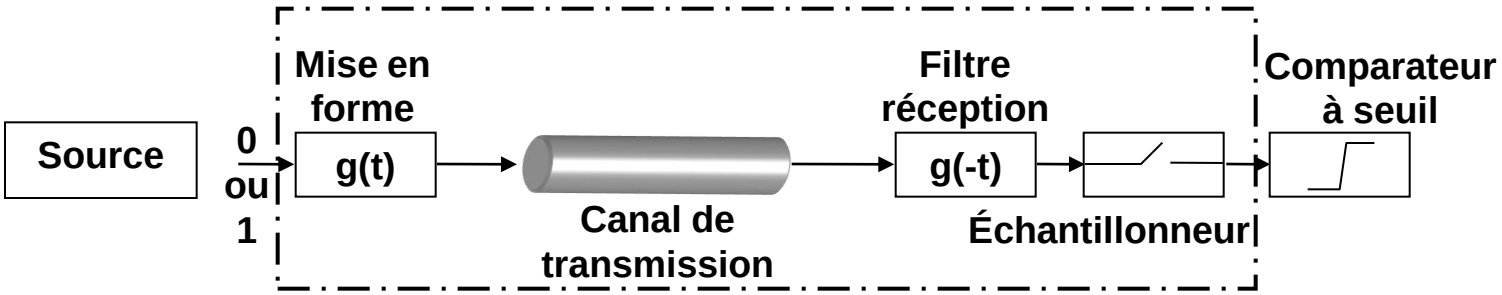
À l'extrémité du canal de transmission sont disposés:

- un filtre adapté de réponse impulsionnelle *g(-t)* (sa présence contribue à minimiser la probabilité d'erreur) ;
- un échantillonneur ;
- un comparateur à seuil.

Modèle 1

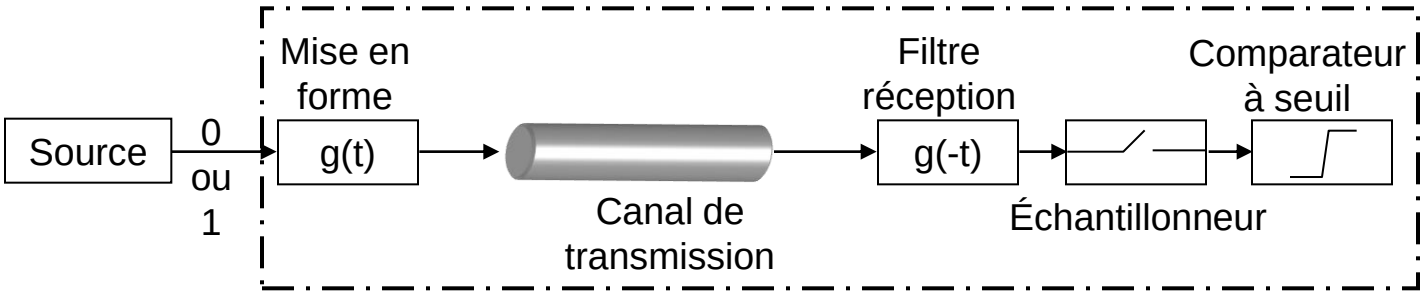
- Plusieurs modèles peuvent être élaborés à partir de la chaîne de transmission :
- **1er modèle** obtenu en englobant le formant, le canal de transmission, le filtre réception et l'échantillonneur.

➢ On obtient un canal à deux entrées ("0" et "1"), la variable de sortie est continue.



Modèle 2

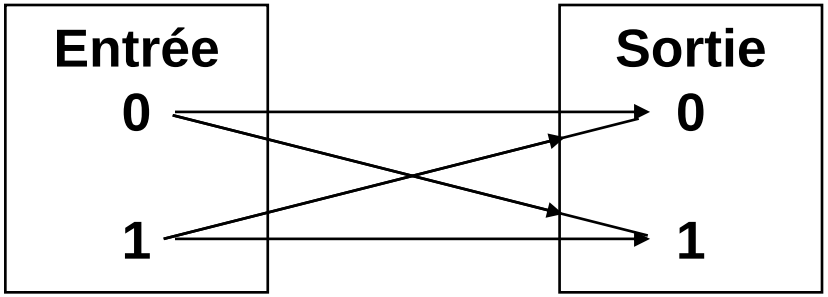
- Si on connecte un comparateur à seuil après l'échantillonneur de telle sorte que la valeur échantillonnée est interprétée en "0" ou "1"), on obtient un canal à deux entrées (les éléments binaires) et deux sorties (les éléments binaires **estimés**);



Cette structure de récepteur est justifiée lorsque les éléments binaires sont codés en des valeurs symétriques -V et +V, et lorsque le canal de transmission est assimilé à un canal à bruit additif gaussien.

Caractérisation du canal

- Le **canal** est caractérisé par les **probabilités** de transition entre les **symboles d'entrée** et les **symboles de sortie**.

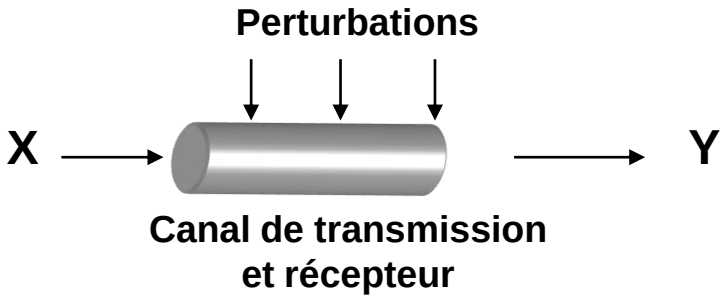


Une **matrice de transition** décrira ces probabilités

Caractérisation du canal

- Rappel : **codage de source** = comment utiliser les redondances d'une source pour diminuer son débit binaire tout en conservant sa quantité d'information

- Ici, la variable Y reçue à la sortie du récepteur comportera des différences avec la variable X initialement émise dues aux perturbations (le bruit) agissant sur le support de transmission.



Caractérisation du canal

- Du point de vue de la théorie de l'information, les imperfections du canal peuvent être traduites en termes d'information qu'apporte la variable de sortie Y sur la variable d'entrée X

$$I(X;Y) = H(X) - H(X/Y)$$

- $H(X/Y)$ = **ambiguïté =** incertitude qui reste sur X lorsque Y est connue (d'autant plus grande que le canal sera perturbé)

- On modélisera un canal par 2 alphabets pour X et pour Y et une matrice de transition Q q_{ij} , ligne i colonne j = probabilité pour que la $i^{ème}$ valeur de l'alphabet d'entrée soit transformée en la $j^{ème}$ valeur de l'alphabet de sortie.

- La quantité $I(X;Y) = H(X) - H(X/Y)$ ne permet pas de caractériser un canal de façon intrinsèque car elle est fonction de la loi de probabilité de X.

Capacité d'un canal

C'est pourquoi on définit la **capacité d'un canal** par le maximum de *I(X;Y)* en prenant en compte toutes les lois de probabilité possibles sur X.

$$\mathbf{C} = \max_{\text{les lois de X}} \mathbf{I(X;Y)}$$ avec C exprimée en bits.

- C correspond au maximum d'information que peut apporter le canal de transmission.

Caractérisation du canal

- Un canal **sans mémoire** est un canal pour lequel la sortie à un instant donné ne dépend statistiquement que de l'entrée correspondante.

- Un canal est **symétrique** si l'ensemble des valeurs constituant l'alphabet de sortie peut être partitionné en sous-ensembles de telle sorte que pour chacun de ces sous-ensembles, la sous-matrice de transition possède les propriétés suivantes :
  - toutes les lignes sont identiques (à des permutations près) ;
  - toutes les colonnes (s'il y en a au moins deux) sont identiques (à des permutations près).

Théorème

Pour un canal symétrique, la capacité est atteinte pour une **loi uniforme** sur l'alphabet d'entrée.

Exemple de canal symétrique

Soient {0,1} (resp. {0,1,2}) l'alphabet d'entrée (resp. de sortie) et Q la matrice de transition.

X\Y	0	1	2
0	0,7	0,2	0,1
1	0,1	0,2	0,7

On peut partitionner l'alphabet de sortie en {0,2}et {1}. Les deux sous-matrices de transition sont alors respectivement

X\Y	0	2
0	0,7	0,1
1	0,1	0,7

X\Y	1
0	0,2
1	0,2

Ces deux matrices possèdent bien les propriétés requises, donc le canal est symétrique.

Caractérisation du canal

Remarques

- Si la capacité C a été calculée en considérant des mots de n symboles, on exprimera la capacité par symbole par le rapport *C/n*.

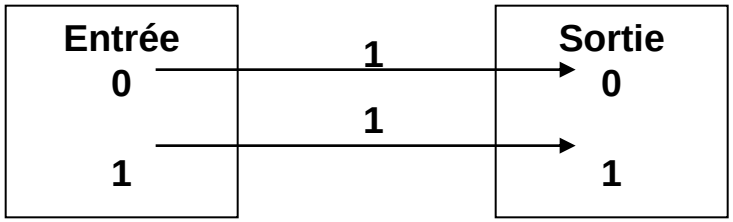
- La capacité d'un canal correspondant à l'aptitude du dispositif à transmettre de l'information, on sera amené à utiliser la capacité par unité de temps (en général la seconde).

  Cette grandeur, exprimée en bits par seconde, est obtenue en divisant la capacité par symbole par l'inverse du débit symbole. Généralement cette quantité est notée C' .

- La remarque précédente conduit naturellement à définir l'entropie d'une source par unité de temps, notée H' , correspondant au rapport de l'entropie par symbole par l'inverse du débit symbole.

- Lorsque se posera le problème de la connexion d'une source à un canal, on aura à comparer H' et C' (on cherchera H' < C')

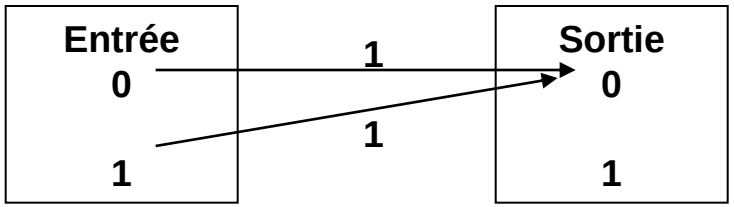
Exemples de calculs de capacités



- La capacité de ce canal est de 1 bit.
  En effet: I(X;Y) = H(X)-H(X/Y) = H(X).
- La capacité est atteinte pour une loi uniforme sur l'entrée. La connaissance de Y entraîne la connaissance de X.
- L'information de Y sur X valant *H(X)*, elle permet de lever l'incertitude sur X.

➢ Cas idéal : transmission sans défaut, canal parfait

Exemples de calculs de capacités



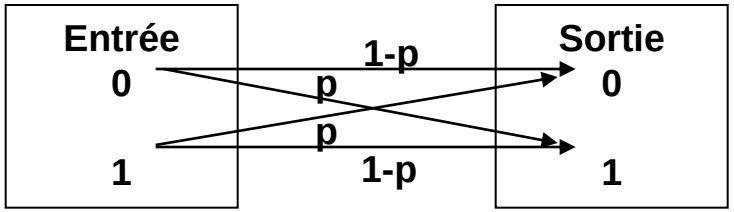
- Dans ce cas, la capacité est nulle et est atteinte quelle que soit la loi de probabilité à l'entrée.

  $I(X;Y)=H(Y)-H(Y/X)=0.$

➢ Pire cas.

Exemples de calculs de capacités

- *Canal symétrique, cas général*



- *Matrice de transition*

	0	1
0	1-p	p
1	p	1-p

- *La capacité est atteinte pour une loi uniforme sur l'alphabet d'entrée.*
  On a donc P{X=0}=P{X=1}=1/2.

Exemples de calculs de capacités

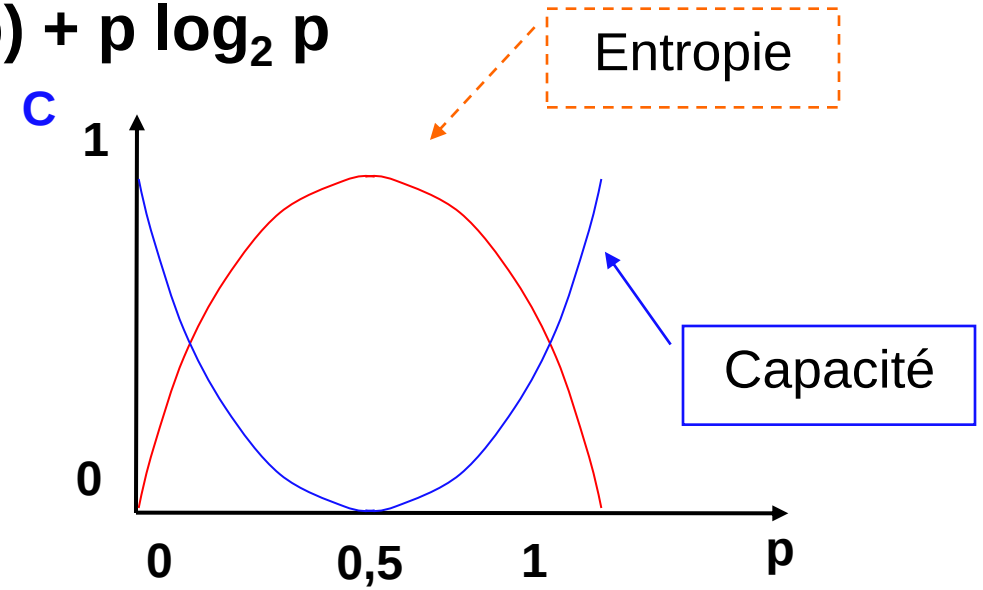
- Pour calculer la capacité C, on va utiliser la relation *I(X;Y)=H(Y)-H(Y/X)* car on connaît la loi de Y sachant X. Calculons la loi de Y.

- P(Y=0) = 0,5*(1-p)+0,5*p=0,5 P(Y=1)=0,5
- H(Y) = -2*0,5$\log_2$(0,5)=1
- H(Y/X) = 0,5*(H(Y/X=0)+H(Y/X=1))
  H(Y/X=0) = - (1-p) $\log_2$ (1-p) - p $\log_2$ p
  H(Y/X=1) = - (1-p) $\log_2$ (1-p) - p $\log_2$ p
- H(Y/X) = - (1-p) $\log_2$ (1-p) - p $\log_2$ p

- D'où **C = 1+ (1-p) $\log_2$ (1-p) + p $\log_2$ p**

Exemples de calculs de capacités

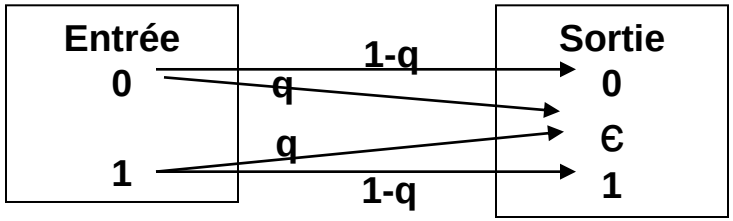
- **$C = 1 + (1-p) \log_2 (1-p) + p \log_2 p$**



- **<u>Commentaires</u> :**
  - **L'entropie est maximum et vaut 1 bit pour *p* = 1/2, donc la capacité est nulle. C'est le cas le plus défavorable car X et Y sont indépendantes. La sortie n'apporte aucune information sur l'entrée.**
  - **Lorsque *p* = 0 , il n'y a jamais d'erreur de transmission : Y coïncide avec X et la capacité est maximum.**
  - **Pour *p* = 1, il y a erreur systématique. On sait qu'à *Y* = 0 (resp. *Y* =1) correspond *X* = 1 (resp. *X* = 0). La connaissance de Y permet de déterminer X. La capacité est maximum.**

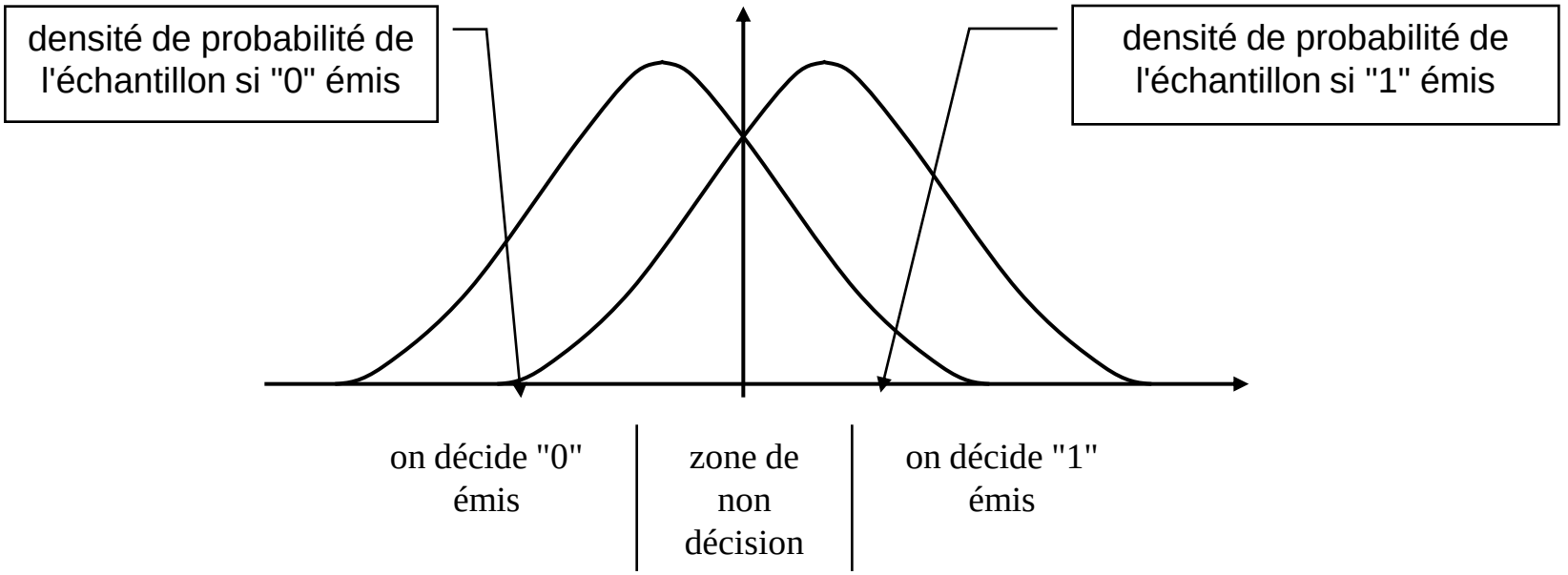
Exemples de calculs de capacités

- Canal binaire à effacement



- l'alphabet d'entrée est binaire {0,1}
- l'alphabet de sortie est ternaire {0,ϵ,1}. Le symbole ϵ est appelé symbole d'effacement
- Si on suppose que les deux symboles 0 et 1 sont codés respectivement en -V et +V avant d'être transmis, les perturbations agissant sur le canal de transmission vont modifier ces valeurs --> **seuil de décision**

Canal binaire à effacement



- 2 façons de gérer cette situation :
  - utiliser une voie de retour pour demander la réémission du symbole,
  - utiliser un code correcteur d'erreurs pour "remplir" les effacements, c'est-à-dire remplacer le symbole d'effacement par l'élément binaire effectivement émis.

Canal binaire à effacement

- Matrice de transition

	0	1	ε
0	1-q	0	q
1	0	1-q	q

- 2 sous-matrices de transition

1-q	0
0	1-q

et

q
q

- Le canal est symétrique,
  la capacité est atteinte pour une loi uniforme sur l'entrée.

- P{Y=0} = 0,5 (1-q)
- P{Y=1} = 0,5 (1-q)
- P{Y=e} = q

On a donc :

H(Y) = - ((1-q) log2 (1-q) + q log2 q - (1-q) log2 2)

H(Y/X=0) = - ((1-q) log2 (1-q) + q log2 q)

H(Y/X=1) = - ((1-q) log2 (1-q) + q log2 q)

H(Y/X) = - ((1-q) log2 (1-q) + q log2 q) soit C = 1-q.

➢ Tout se passe comme si la fraction q de l'information correspondant aux symboles effacés était perdue.

Codage de canal

- Après avoir caractérisé un canal du point de vue de la théorie de l'information en introduisant sa capacité, nous allons maintenant nous intéresser à la qualité de la transmission en termes de probabilité d'erreur.
- 2 théorèmes fondamentaux:
  - Le 2ème théorème de Shannon qui énonce une condition d'adéquation entre la source et le canal pour obtenir un taux d'erreur aussi faible que souhaité.
  - Le théorème réciproque du deuxième théorème de Shannon qui fournira un minorant de la probabilité d'erreur lorsque la condition d'adéquation source canal n'est pas satisfaite.

Codage de canal

- L'incertitude qui subsiste sur X lorsque Y est connue peut être divisée en deux termes :
  - un premier terme qui correspond à l'incertitude liée à la question de savoir si oui ou non une erreur a été commise ;
  - un second terme relatif à l'incertitude sur le symbole qui a été effectivement émis lorsque l'on commet une erreur (cette incertitude concerne les m-1 symboles autres que celui reçu et ceci avec la probabilité p_e).
- On obtient donc l'**inégalité de Fano** :

$$H(X/Y) \leq H_2(p_e) + p_e \log_2 (m-1)$$

Codage de canal

- **2ème théorème de Shannon**

  Soient un canal discret sans mémoire de capacité C et une source discrète stationnaire d'entropie **R.**

  Alors si R < C, ∀ **ϵ** > 0 il existe un code de longueur n tel que la probabilité d'erreur p_e après le décodeur soit inférieure à **ϵ** si le code est utilisé sur le canal.

Codage de canal

- Ce théorème donne tout son sens à la notion de **capacité** que l'on peut interpréter comme la **quantité maximum d'information** qui peut être **transmise sans erreur**.
- Le résultat énoncé par ce théorème est surprenant, en ce sens qu'a priori, on ne pouvait envisager d'effectuer une transmission sur un **canal bruité**, avec un **taux d'erreur aussi faible que souhaité**.
- Contrairement au théorème de codage de source, le 2ème théorème de Shannon n'indique pas comment construire le code bloc optimum. Les **codes correcteurs d'erreurs** permettent de réduire la probabilité d'erreur sur un canal. Cette réduction sera obtenue en ajoutant de la redondance aux messages à transmettre.
- En pratique, il faut que la condition Entropie < Capacité soit vérifiée sur un même laps de temps.

  Si le débit source $D_s = 1/T_s$ (resp. le débit canal $D_c = 1/T_c$), on devra vérifier :

$$\frac{\text{entropie}}{\text{unité de temps}} < \frac{\text{capacité}}{\text{unité de temps}} \qquad \frac{H(X)}{T_s} < \frac{C}{T_c}$$

5. Codes détecteurs et codes correcteurs

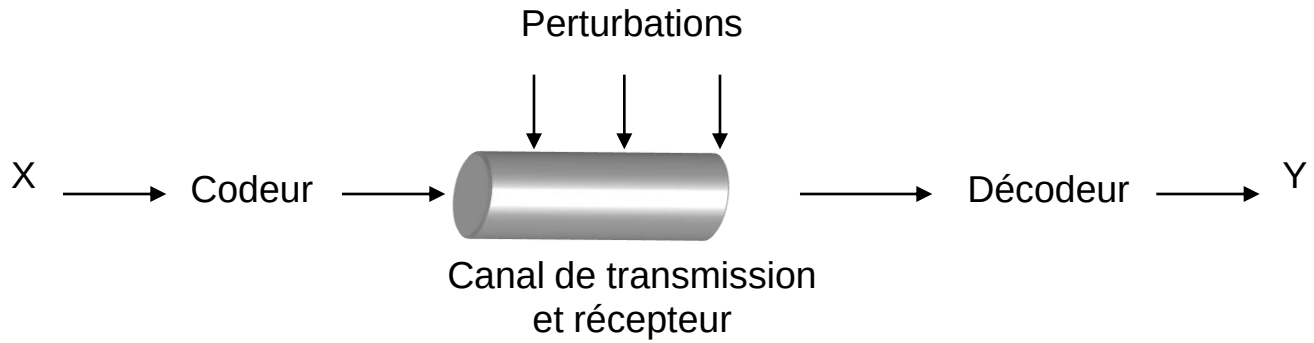
- Le nombre d'erreurs de transmission est très variable : Probabilité d'erreur de 10^{-3} (RTC de mauvaise qualité) à 10^{-12} / 10^{-15}

- 3 opérations peuvent être mises en place :
  - la détection d'une erreur,
  - sa localisation
  - et sa correction.

- La protection peut s'appliquer à deux niveaux :
  - au niveau bit ou caractère (bit de parité) ;
  - au niveau de la trame (CRC) ou du paquet réseau.

Protection

- Un codeur est introduit avant la transmission sur le canal et un décodeur est placé en sortie du canal.



Protection

Codeur : introduit de la redondance : **C(n,k) = r**

n : bits à transmettre, k : redondance

- Il existe différentes classes de codes :
  - En bloc : les r bits rajoutés ne dépendent que des k bits d'information.
    - Détection d'erreur : parité, codes polynomiaux, codes cycliques.
    - Correction d'erreur : codes de Hamming, codes BCH).
  - Conventionnels (ou récurrents).
  - Turbo-codes

Codes détecteurs d'erreur

- **Parité simple** : à chaque caractère, un bit est ajouté.

Exemple :

0011010 0 parité impaire
0011010 1 parité paire

➢ Dans ce cas, on est capable de détecter une erreur de parité, mais pas de la localiser.

- **Parité double** : à chaque bloc de caractère, on ajoute un champ de contrôle supplémentaire (LRC : Longitudinal Redondancy Check, VRC : Vertical Redondancy Check)

Exemple : pour les données 0011010 1100101 0101010

00110101
11001010 parité LRC et VRC paire
01010101
10101010

La suite 00110101 11001010 01010101 10101010 est émise.

➢ La combinaison de LRC et VRC permet de détecter 2 erreurs dans un seul mot ou de corriger 1 erreur.

Codes détecteurs d'erreur

Détection d'erreur par code cyclique

- Un mot de code est représenté sous forme polynomiale dans laquelle la suite de bits à transmettre $M=m_1m_2\ldots m_n$ est représentée par $M(x) = u_1+u_2x+\ldots+u_nx^{n-1}$

- Par exemple 1100101 est représenté par $1x^6+1x^5+0x^4+0x^3+1x^2+0x+1$ c'est à dire $x^6+x^5+x^2+1$

Codes détecteurs d'erreur

Principe de détection :

- On utilise le reste R(x) de la division polynomiale M(x) par un polynôme diviseur G(x) qui donne un quotient Q(x). R(x) est calculé par l'émetteur puis transmis au récepteur. Le récepteur fait le même calcul R'(x) en divisant M(x)+R(x) (message + contrôle).
  - Si R'=0 alors pas d'erreur, si R'≠0 alors erreur.

- M(x) – R(x) est divisible par G(x) et est équivalent à M(x) + R(x) modulo 2.

- Cette méthode fonctionne bien car la table de vérité de l'addition (en modulo 2) équivaut au ou exclusif (XOR).

Codes détecteurs d'erreur

Exemple de polynôme générateur :

- Suivant l'avis V41 du CCITT : $G(x) = x^{16}+x^{12}+x^{5}+1$

- Autre exemple : calcul de LRC en série

Transmission de 2 octets

01001101
01101111
00100010 = LRC

- LRC sur 1 octet est un code cyclique de polynôme générateur $x^{8}+1$.

Codes détecteurs d'erreur

- 2 octets précédents : 01001101 01101111
- $M(x) = x^{14}+x^{11}+x^{10}+x^8+x^6+x^5+x^3+x^2+x+1$
- $x^8 M(x) = x^{22}+x^{19}+x^{18}+x^{16}+x^{14}+x^{13}+x^{11}+x^{10}+x^9+ x^8$ (on multiplie par x^8 pour faciliter la division)
- $x^8 M(x) / (x^8+1) =$

$x^{22}+x^{19}+x^{18}+x^{16}+x^{14}+x^{13}+x^{11}+x^{10}+x^9+ x^8$ | x^8+1

$-x^5-x$ | $x^{14}+x^{11}+x^{10}+x^8+x^5+x$

ou $+x^5+x$ |

- $R(x)=x^5+x$ → **00100010** on trouve comme précédemment ! !
- On transmet donc : 01001101 01101111 **00100010**

- Le récepteur divise le message reçu par x^8+1 pour vérifier qu'il n'y a pas d'erreur.

➢ Pour fabriquer ces codes à la volée, on utilise un registre à décalage ainsi que un ou plusieurs ou exclusif.

Codes correcteurs d’erreur

- La capacité de correction est d'autant meilleure que les "mots" du code diffèrent beaucoup entre eux.
- "distance" minimale entre deux mots, la distance étant le nombre de symboles qui diffèrent ;
- Exemple, la distance dite "de Hamming" – « nombre de différences » entre 10100111 et 10111111 vaut 2, tandis que celle entre 10100111 et 11000001 vaut 4.
- Plus la distance minimale est élevée, plus le code peut corriger d'erreurs.
- On peut augmenter la distance minimale entre mots en rajoutant des bits. Mais rallonger le message accroît la durée et le coût des communications.

  Objectif : trouver des algorithmes de codage performants, qui permettent de détecter et de corriger le maximum d'erreurs, tout en allongeant le moins possible les mots. De plus, les procédures de codage et de décodage doivent être suffisamment simples.

Codes correcteurs d'erreur

Un code de Hamming (R. W. Hamming, années 1950)

0000***000***	0100***111***	1000***101***	1100***010***
0001***011***	0101***10***0	1001***110***	1101***001***
0010***110***	0110***00***1	1010***011***	1110***100***
0011***101***	0111***010***	1011***000***	1111***111***

Dans le code correcteur ci-dessus, les mots de départ $a_1a_2a_3a_4$ ont quatre bits (il y a donc $2^4 = 16$ mots distincts).

On rajoute à chaque mot trois bits de contrôle a_5, a_6, a_7 dont la valeur est déterminée par les quatre premiers bits :

$$a_5 = a_1 + a_2 + a_3,$$
$$a_6 = a_2 + a_3 + a_4,$$
$$a_7 = a_1 + a_2 + a_4$$

Ces relations étant calculées "modulo 2", c'est-à-dire en ne retenant que le reste dans la division par 2 (par exemple, 1 + 1 + 1 = 1, 1 + 0 + 1 = 0). La distance minimale entre deux mots de ce code vaut 3, ce qui permet de détecter et corriger une erreur sur l'un des sept bits d'un mot.

Codes correcteurs d'erreur

Code correcteur à vérification de synchronisation

- **BCH : Bose Chandhuri Hocquengheim**
  + traitement simple après codage
  - **Complémenter à 1 le $i^{ème}$ bit ou bien permutation du $i^{ème}$ bit et du $j^{ème}$ bit.**
  - **A la réception, on effectue l'opération inverse.**
  - **Si la division ne donne pas le bon résultat : erreur.**

Codes correcteurs d'erreur

Exemple de correction d'erreur

- Tout mot de code $X = (x_1, \ldots, x_{15})$

  où $x_1 \ldots x_7$ sont des données

  et $x_8 \ldots x_{15}$ sont des bits de redondance
- on doit vérifier $H X^t = 0$ (X^t : transposé de X) où H est la matrice de contrôle de parité

Codes correcteurs d'erreur

H

Si les bits de données valent 0101100,

X=(0101100$x^8x^9x^{10}x^{11}x^{12}x^{13}x^{14}x^{15}$)

H X^t = 0 s'écrit :

1000100110101111	$1+x^8+x^9+x^{11}+x^{13}+x^{14}+x^{15}=0$
0100110101111100	$x^8+x^{10}+x^{11}+x^{12}+x^{13}=0$
0010011010111110	$x^9+x^{11}+x^{12}+x^{13}+x^{14}=0$
0001001101011111	$1+x^8+x^{10}+x^{12}+x^{13}+x^{14}+x^{15}=0$
1000110001100001	$1+x^{10}+x^{11}+x^{15}=0$
0001100011000011	$x^9+x^{10}+x^{14}+x^{15}=0$
0010100101001011	$1+x^8+x^{10}+x^{13}+x^{15}=0$
0111101111011111	$1+x^8+x^9+x^{10}+x^{12}+x^{13}+x^{14}+x^{15}=0$

Codes correcteurs d'erreur

- Détail du calcul de la 1ère ligne :
  (100010011010111) XOR (0101100 $x^8x^9x^{10}x^{11}x^{12}x^{13}x^{14}x^{15}$)^t

- Après résolution du système d'équations (ex : Ligne 4 + ligne 8 → x^9=0), la valeur du mot code est : 0101100 **00101010**

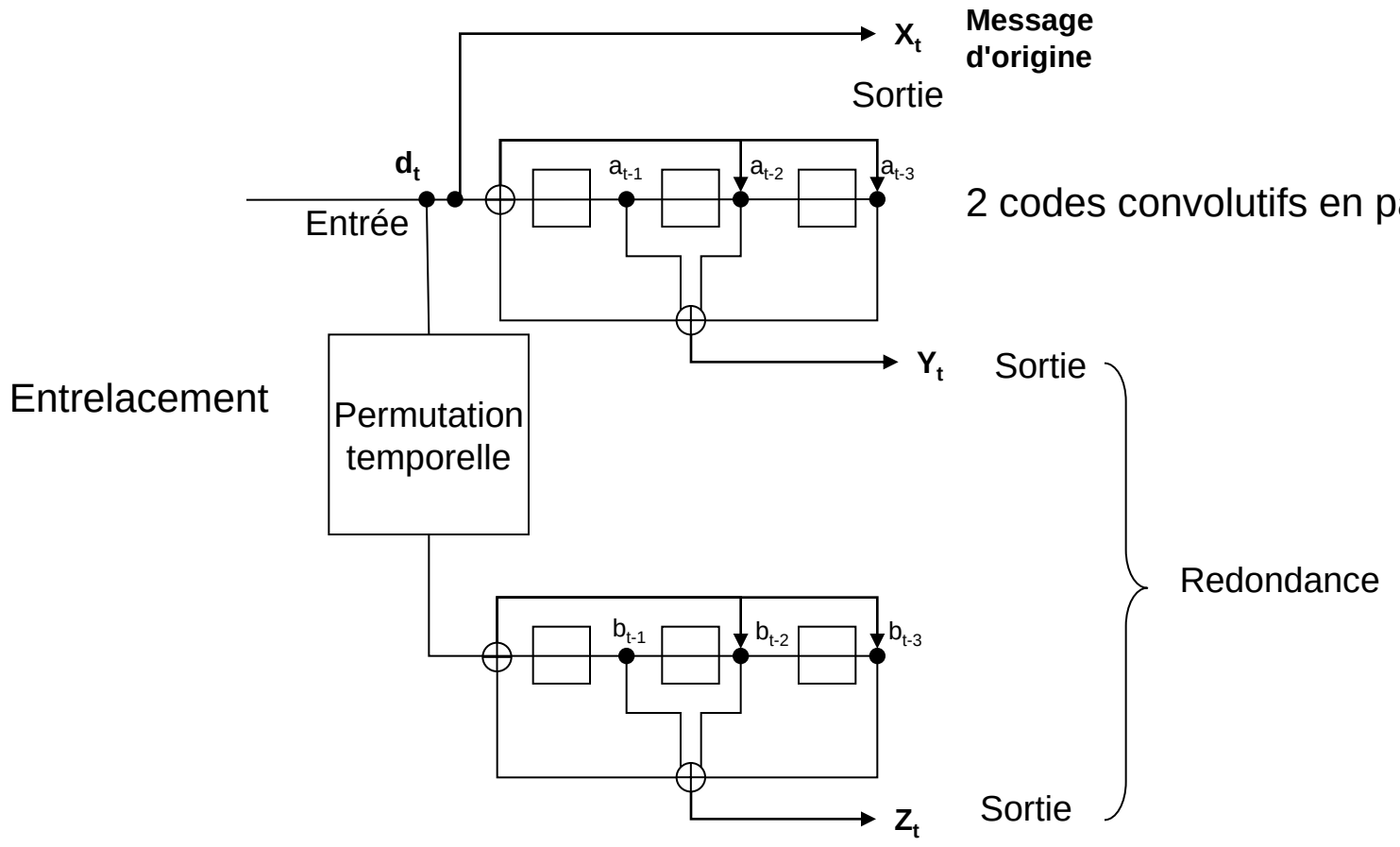
- 1er cas : détection d'erreur
  si erreur sur le 9ème bit Y= 0101100 01101010
  on calcule Y.H^t → on trouve (10100101)^t → correspond à la 9ème colonne de la matrice, on corrige donc le 9ème bit.

- 2ème cas : détection d'erreur de synchronisation
  On décide de complémenter à 1 le 9ème bit pour détecter les erreurs de synchronisation.
  On reçoit 1011000 11010100, le récepteur complémente à 1 le 9ème bit. Z = 1011000 10010100.
  H.Z^t → (01110110)^t → somme de la 8ème et de la 9ème colonne, on corrige donc les 2 bits concernés.

Turbo-codes

- C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: turbo codes", in Proc. of ICC '93, Geneva, pp. 1064-1070, May 1993 *(ENST / France Télécom)*
- Adoptés par :
  - CCSDS (Consultative Committee for Space Data Systems),
  - comité de normalisation pour les agences spatiales mondiales (ESA, NASA, NASDA, ...).
  - la première mission européenne avec turbocode est la sonde SMART-1 qui tournera autour de la lune au courant de l'année 2005.
  - standards de 3ème génération de systèmes de communication : l'UMTS, en Europe, ou le CDMA2000, aux États-unis et en Asie, sont les applications commerciales les plus connues.
  - D'autres systèmes avec turbocode (INMARSAT, EUTELSAT, DVB-RCS, DVB-RCT, BRAN, IEEE 802.16, enregistrement magnétique...) sont d'ores et déjà normalisés ou en cours de spécification.

Turbo-codes : principe



Analogie avec les mots croisés : décodage itératif utilisant les lignes et les colonnes