

Chapitre 2

Communications de données

Code de Huffman

L'algorithme d'encodage de Huffman conduit toujours à un code à décodage instantané optimal, c'est à dire possédant la plus petite longueur L possible. On détermine fréquemment cette limite afin de comparer les performances obtenues avec d'autres méthodes de codage. L'algorithme d'encodage de Huffman peut être énoncé comme suit:

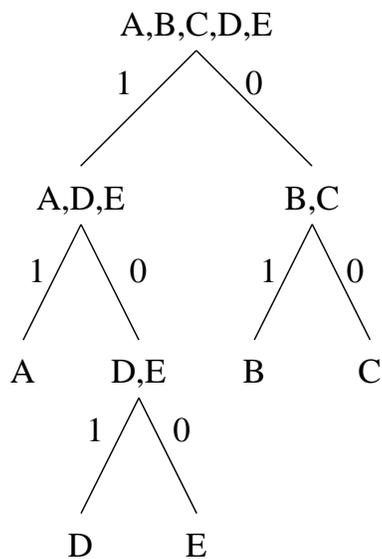
1. Ordonner de façon décroissante les caractères à encoder x_i selon leur probabilité d'apparition p_i ,
2. Construire un arbre d'encodage en partant de la fin. Ceci consiste à attribuer aux deux caractères les moins fréquents les symboles 0 et 1,
3. Remplacer dans la liste ces deux caractères par un caractère conjoint $x_{n-1,n}$ dont la probabilité d'apparition est $p_{n-1} + p_n$. Ce nouveau caractère représente un sommet de l'arbre d'encodage,
4. Répéter la procédure en combinant à nouveau les deux caractères de plus faible probabilité, jusqu'à ce que tous les caractères aient une transcription distincte.

Théorème de Shannon Il est toujours possible de diminuer arbitrairement la redondance R_c par la formation de groupes de caractères (si les groupes sont assez longs).

Exemple de code de Huffman

x_i	p_i	x_i	p_i	x_i	p_i	i	p_i
A	0.3	A	0.3	B,C	0.44	A,D,E	0.56
B	0.24	D,E	0.26	A	0.3	B,C	0.44
C	0.2	B	0.24	D,E	0.26		
D	0.15	C	0.2				
E	0.11						

Ce mécanisme produit en moyenne les meilleurs code.



x_i	code	p_i	$p_i \lg \frac{1}{p_i}$	$w_i p_i$
A	11	0.3	0.521	0.6
B	01	0.24	0.494	0.48
C	00	0.2	0.464	0.4
D	101	0.15	0.411	0.45
E	100	0.11	0.350	0.33
			2.240	2.260

Longueur moyenne des mots codés: $L = 2.260$,
 Entropie de la source: $H = 2.240$,
 Redondance après codage: $R_c = 0.02$.

Compression de données

L'algorithme de codage de Huffman est une des techniques utilisées pour la compression de données . Le but de toute technique de compression est de réduire la redondance pour minimiser le temps de transmission.

D'autres techniques de compression sont:

- Packed decimal (BCD code pour encoder des chiffres)
- Codage de plages ou par blocs (télé-copie ou images)
- JPEG (Joint Photographic Experts Group)
- MPEG (Motion Picture Experts Group)

Codage par plages

Le codage par plages (*run length encoding*) est utilisé pour coder de longues chaînes de données binaires qui contiennent beaucoup de zéros:

- un symbole de k bits indique le nombre de 0 présents entre deux 1.
- pour coder des longues chaînes de zéros, le symbole ne contenant que des 1 indique un nombre de 0 de $2^k - 1$ plus la valeur des symboles suivants.

Exemple: 0000001000000000000010001 contient des chaînes de 0 de longueurs 6,14 et 3. En utilisant des symboles de $k = 3$ bits on obtient:

11011111000011

avec 110 : 6; 111111000 : $7 + 7 + 0 = 14$; 011 : 3

On a alors sauvé 38%.

La donnée

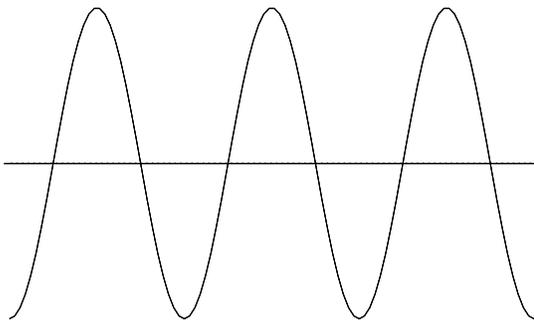
La donnée est la représentation de l'information à transmettre. On distingue deux formes:

analogique : valeurs continues (e.g. voix ou image représentée par différents niveaux d'intensité).

numérique : valeurs discrètes (e.g. texte représenté par des caractères ASCII où chaque caractère correspond à une séquence de 7 ou 8 bits).

Transmission de la donnée

La donnée est encodée sur un signal électromagnétique ou lumineux qui la transporte sur un médium adéquate.



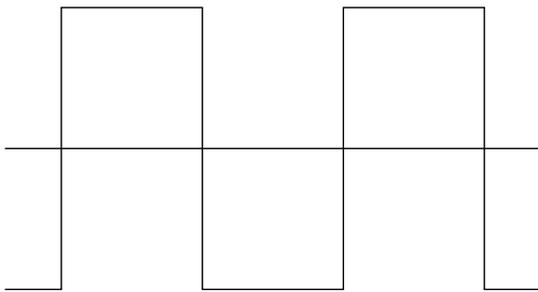
analogique : $A \sin(2\pi ft + \Phi)$

A: amplitude

F: fréquence

Φ : phase

t: temps



numérique: impulsion de voltage

Analyse de Fourier

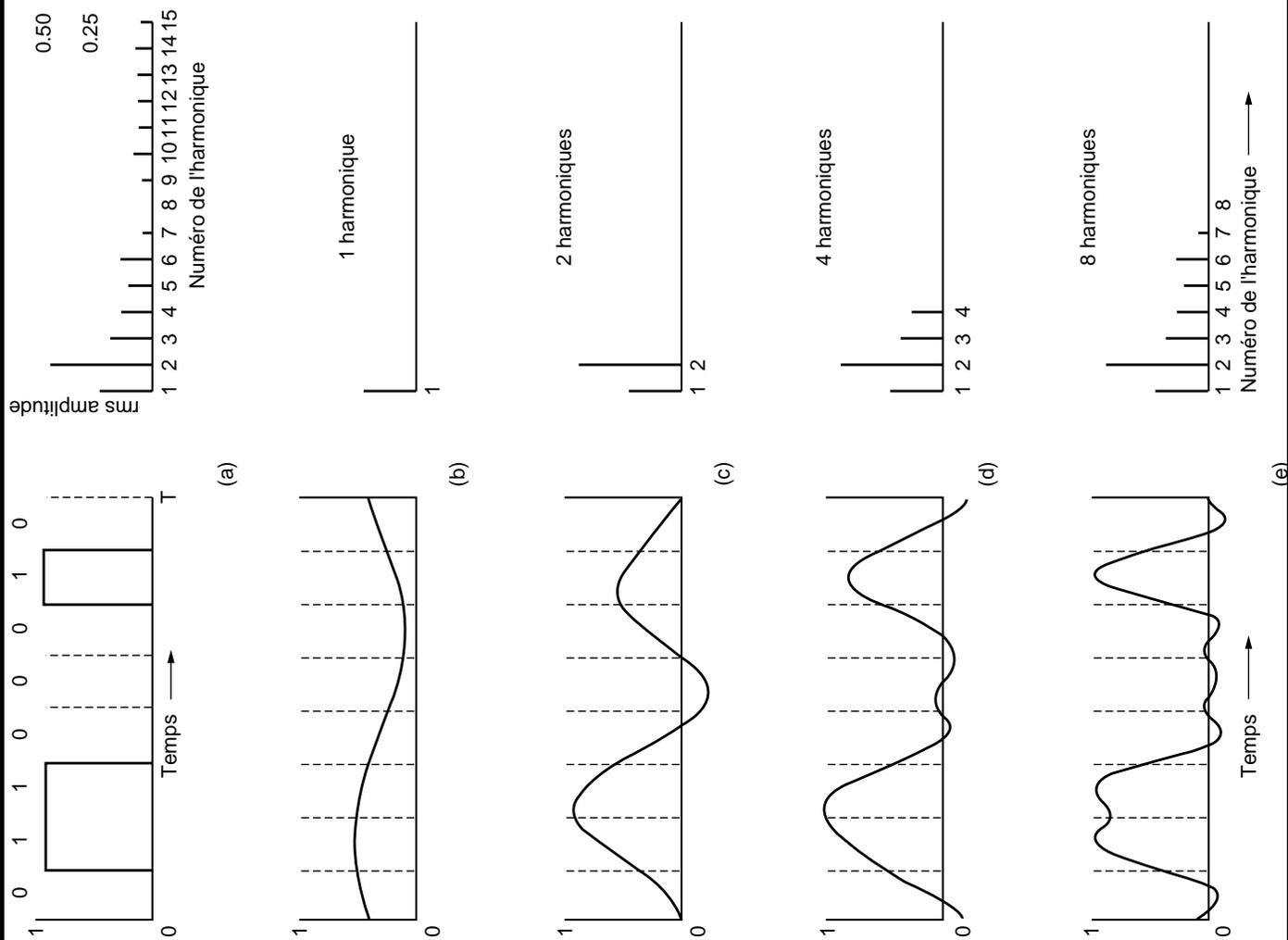
L'information analogique est représentée par des signaux. Selon la théorie de Fourier, un signal périodique quelconque peut être décomposé en une suite de fonctions périodiques sinusoidales: les *composantes harmoniques*.

$$s(t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \sin(2\pi nft) + \sum_{n=1}^{\infty} b_n \cos(2\pi nft). \quad (1)$$

$f = \frac{1}{T}$ fréquence fondamentale

a_n, b_n amplitudes des composantes harmoniques

Un signal rectangle est composé d'un nombre infini de composantes sinusoidales avec une fréquence fondamentale f et les multiples $3f, 5f, \dots$



Signal numérique et ses composantes harmoniques (b à e). Approximations successives du même signal. (Tanenbaum Fig. 2-1)

Spectre et bande passante

Caractéristiques fréquentielles des signaux transmis aux débits normalisés selon la recommandation V.1 du CCITT, sur une voie téléphonique avec une largeur de bande (*bande passante*) 300-3400 Hz:

Bps	T(msec)	Première harmonique (Hz)	# harmoniques
300	26.76	37.5	80
600	11.22	75	40
1200	6.67	150	20
2400	3.33	300	10
4800	1.67	600	5
9600	0.83	1200	2
19200	0.42	2400	1
38400	0.21	4800	0

Relation entre débit binaire et harmoniques. (Tanenbaum Fig. 2-2)

Transmission: bruit

Il y a en réalité du bruit parasite sur les lignes de transmission. Ce bruit s'exprime par rapport à la puissance utile du signal transmis:

rapport signal sur bruit: $\frac{P_S}{P_N}$

P_S énergie du signal, P_N énergie du bruit.

La quantité de bruit est exprimée en **décibel**: $10 \log \frac{P_S}{P_N}$ dB.

Le nombre de niveaux significatifs dépend de P_S et P_N :

$$m \leq \sqrt{1 + \frac{P_S}{P_N}} \quad (2)$$

Capacité d'un canal

La capacité de transmission d'un canal est:

$$C = F \cdot \lg\left(1 + \frac{P_S}{P_N}\right) \quad (3)$$

Exemple: bande passante: 3000Hz, bruit: 20dB

$$C = 3000 \lg(1 + 100) = 19974 \text{ bit/sec.}$$

Si l'on transmet sur une voie (canal) perturbée de capacité C une information à un débit D , il existe en général une possibilité de codage de voie permettant de réduire la probabilité d'erreur de transmission à une valeur arbitrairement faible, pour autant que $D \leq C$. (Théorème *Noisy channel coding* de Shannon).

Délai de propagation d'un message

Il y a toujours un délai entre l'émission d'un signal (optique, électrique ou radio) à un bout d'un médium et sa réception à l'autre bout. Ceci est appelé le *délai de propagation d'une transmission*, T_p , d'un médium.

$$T_p = \frac{\text{distance physique en mètres}}{\text{vitesse de propagation en mètres par seconde}}$$

Dans le meilleur des cas, le signal se propage dans le vide à la vitesse de la lumière ($3 \times 10^8 \text{ ms}^{-1}$). La vitesse de propagation pour des fils à paires torsadées ou des câbles coaxiaux est de l'ordre de $2 \times 10^8 \text{ ms}^{-1}$.

Normalement, les données sont transmises par blocs de bits et, à la réception d'un bloc, un accusé de réception est envoyé à l'expéditeur. Un paramètre important pour un lien de transmission de données est son *round-trip delay* qui est défini comme la durée entre l'émission du premier bit d'un bloc et la réception du dernier bit de l'accusé de réception. Ceci dépend du temps pris pour transmettre le bloc — le *délai de transmission* T_x — et du délai de propagation T_p . Le poids relatif de ces deux temps varie selon les différents types de liens. On exprime souvent ces deux temps comme le rapport a :

$$a = \frac{T_p}{T_x}$$

où

$$T_x = \frac{\text{nombre de bits à transmettre}}{\text{débit du lien en bits par seconde}}.$$

Exemple: temps de transmission

50 m de câble paires torsadées, 10 kbps, messages en bloc de 500 bits.

$$T_p = \frac{50}{2 \times 10^{-8}} = 2.5 \times 10^{-7} s$$

$$T_x = \frac{5 \times 10^2}{10^4} = 0.05 s$$

$$a = \frac{25 \times 10^{-6}}{5 \times 10^{-2}} = 5 \times 10^{-5} s$$

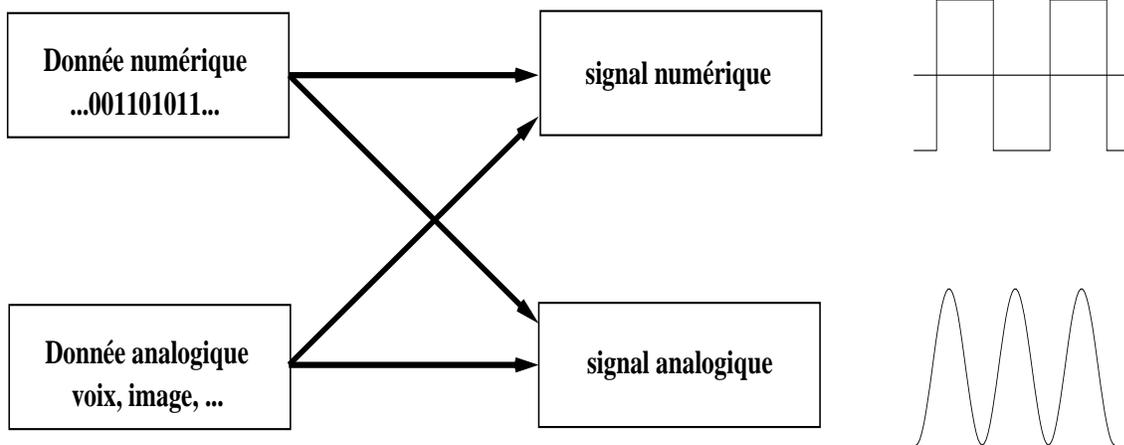
$a \ll 1$: T_x dominant (puce)

$a \simeq 1$: T_x, T_p influence égale (Ethernet)

$a \gg 1$: T_p dominant (satellite)

Encodage

Les données numériques sont encodées soit sur un signal numérique, soit sur un signal analogique.



Encodage numérique – numérique

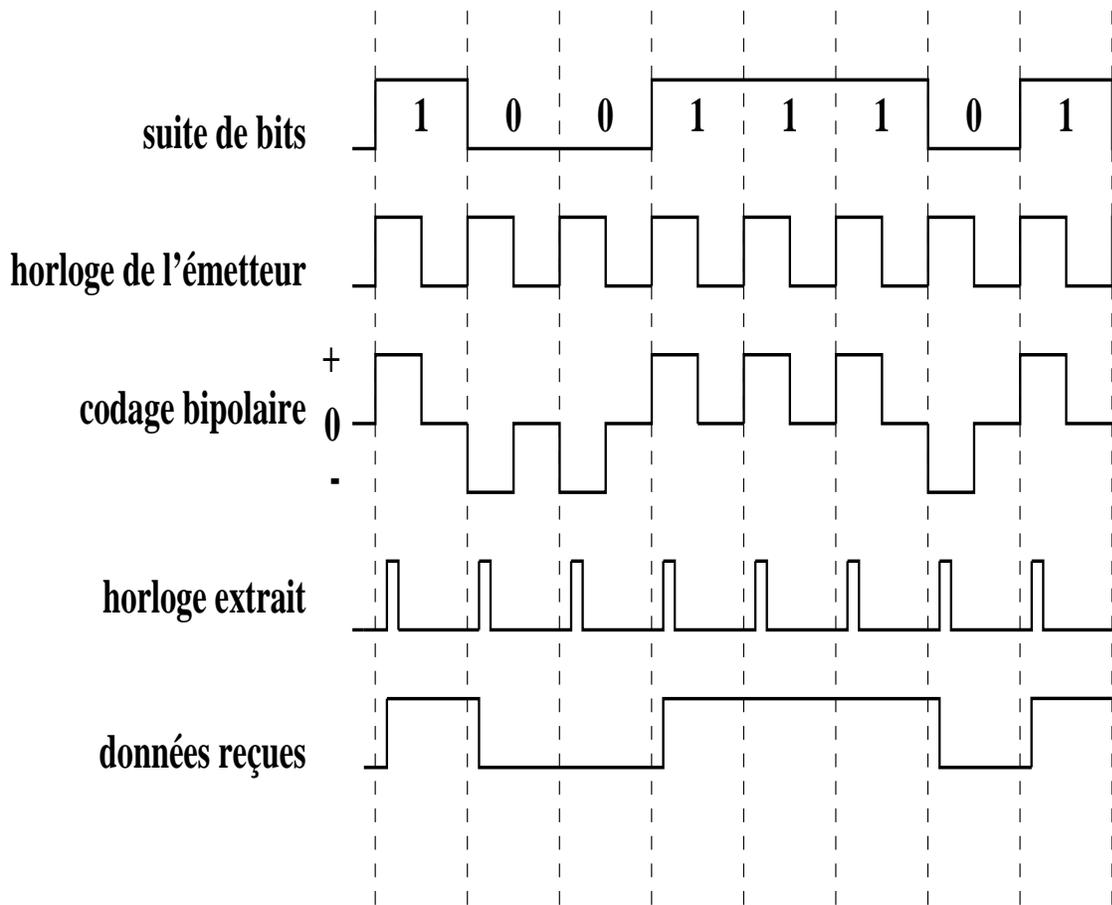
Les données numériques sont habituellement représentées par une suite de bits $b(t)$. Cette suite est encodée sur un signal numérique. Les méthodes d'encodage (numérique - numérique) principaux sont:

- Codage non-retour à zéro
- Codage bipolaire
- Codage biphasé
- Codage Manchester différentiel

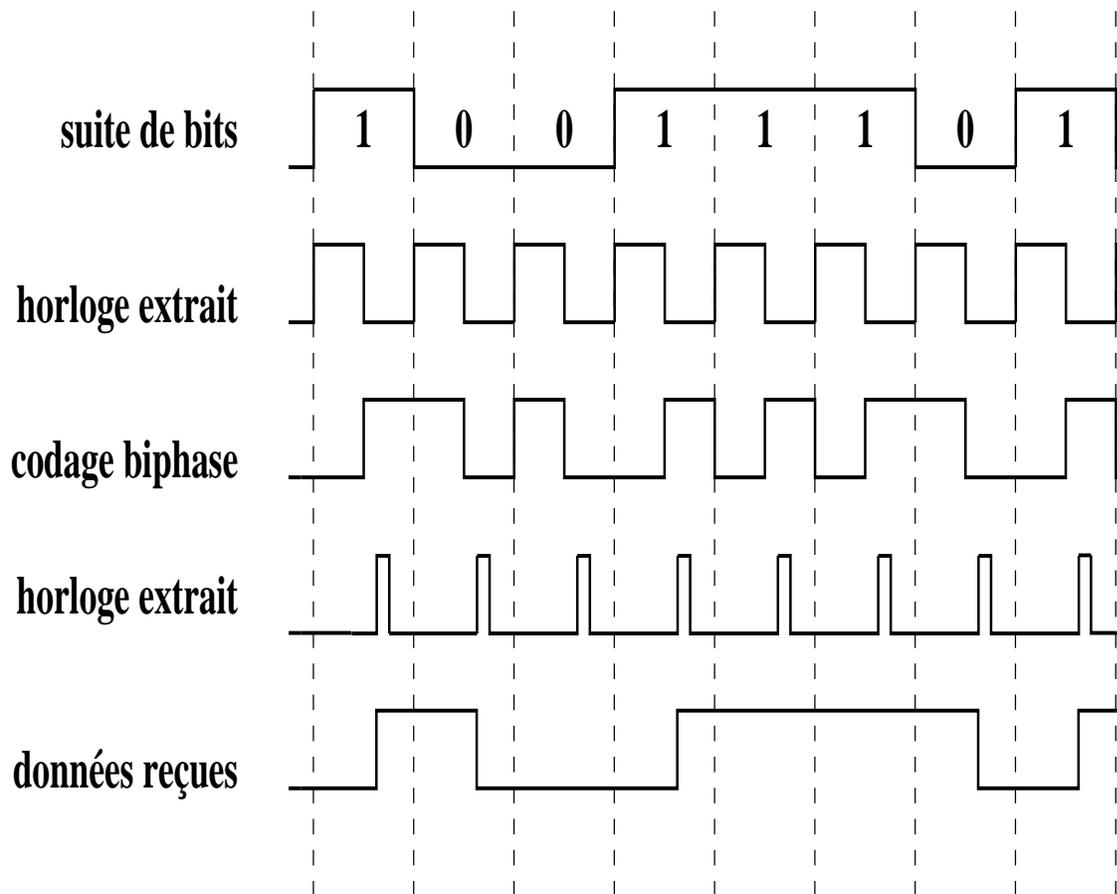
Encodage numérique – numérique

Méthode de codage	Algorithme de codage
non-retour à zéro	1 = haut niveau 0 = bas niveau
biphase	1 = transition haut – bas au milieu de l'intervalle 0 = transition bas – haut au milieu de l'intervalle
bipolaire	1 = transition au début de l'intervalle en alternant les polarités 0 = pas de transition
Manchester différentiel	toujours une transition au milieu de l'intervalle 1 = pas de transition au début de l'intervalle 0 = transition au début de l'intervalle

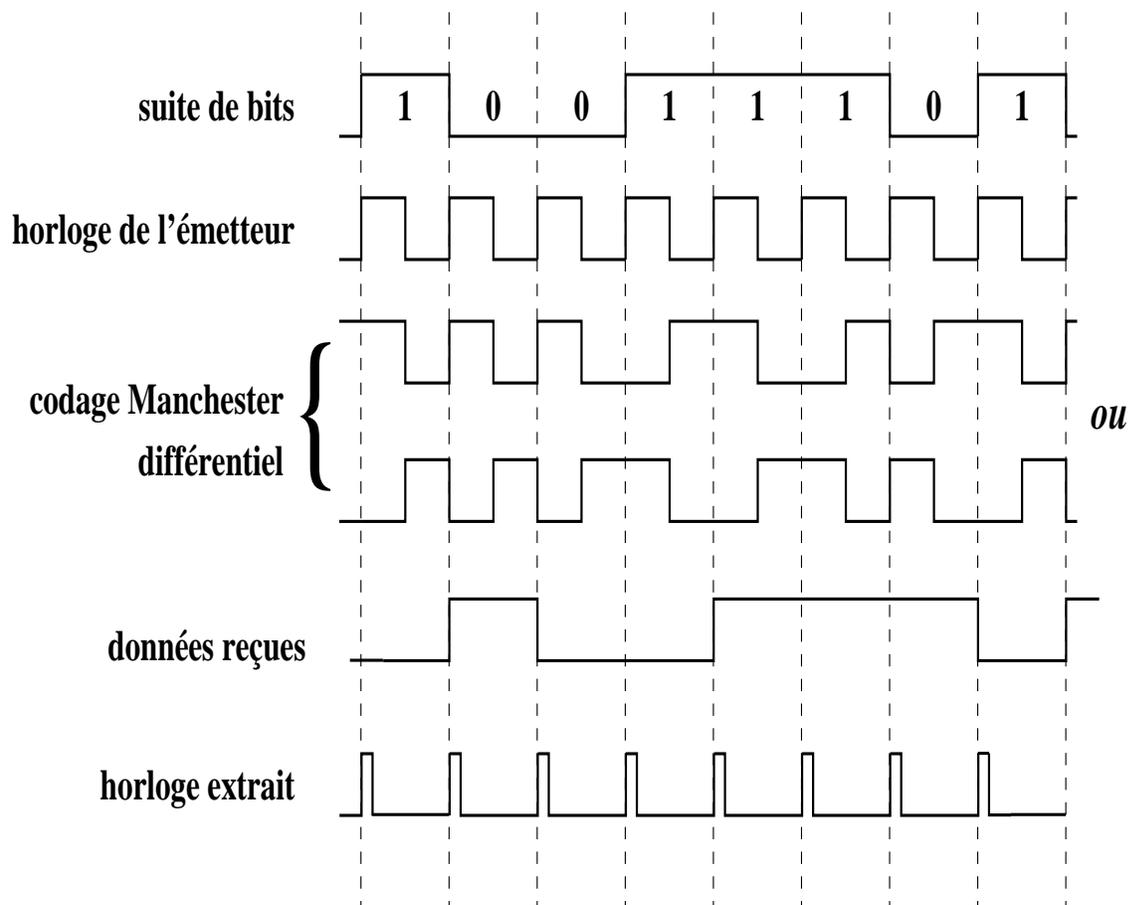
Codage bipolaire



Codage biphasé



Codage Manchester différentiel



Encodage numérique - analogique

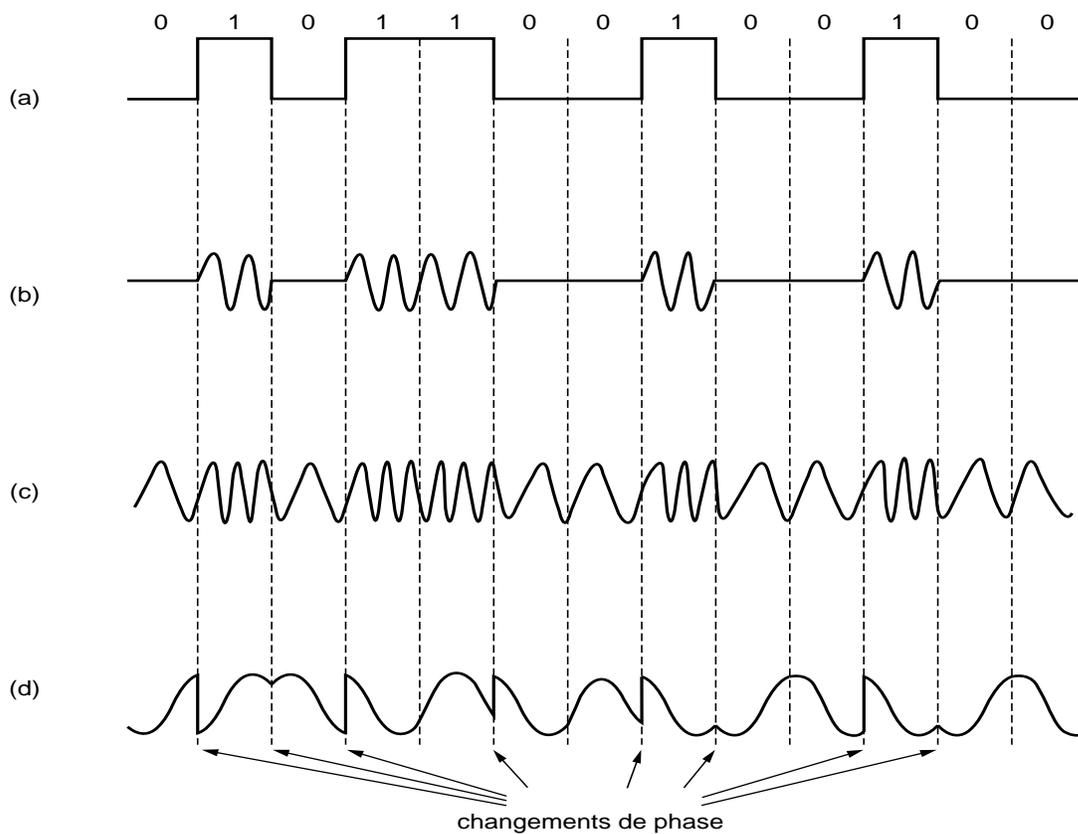
Les données numériques sont encodées en signaux analogiques avec la technique de **modulation**. La conversion inverse est appelée **démodulation**. Ces fonctionnalités sont réalisées par un **modem**. La modulation consiste à modifier un ou plusieurs paramètres de l'onde porteuse, tels que sa phase, son amplitude ou sa fréquence.

Modulation d'amplitude: deux différents niveaux de voltage sont utilisés pour représenter 0 et 1. La fréquence reste constante.

Modulation de fréquence: deux (ou plusieurs) fréquences différentes représentent 0 et 1. L'amplitude reste constante.

Modulation de phase: la phase de l'onde porteuse varie de 45, 135, 225 ou 315 degrés à des instants régulièrement espacés. Chaque changement de phase transmet 2 bits d'information.

Types de modulation



(a) Signal numérique, (b) Modulation d'amplitude, (c) Modulation de fréquence, (d) Modulation de phase. (Tanenbaum Fig. 2-18)

Les supports de transmission

Les supports ou médiums de transmission le plus couramment utilisés pour les applications de télécommunications et de téléinformatique sont:

- Les câbles à paires symétriques ou les câbles à paires torsadées.
- Les câbles coaxiaux (*baseband* ou *broadband*).
- Les fibres optiques.
- les systèmes de transmission par ondes (faisceaux dirigés, transmission par satellites, systèmes de transmission avec les mobiles).
- les systèmes de transmission infrarouge et micro-ondes.

Câbles

- coaxial (épais ou mince)
- fibre optique (monomode ou multimode)
- paire torsadée (blindée ou non-blindée)

Catégorie	paires torsadées	distance maximale
3,4	4 paires UTP 2 paires STP	100 m
5	4 paires UTP	200m

paire torsadée (catégorie 5)	10Mbps - 1Gbps, 200m
50 ohm coax (mince)	10Mbps - 5Gbps, 200m
75 ohm coax (épais)	10Mbps - 5Gbps, 500m
fibre multimode	100Gbps, 2km
fibre monomode	100Gbps, 40km