



Institut MONTEFIORE
Service de Télécommunications et d'Imagerie

ELEN008-0
Principes des télécommunications analogiques et
numériques

Professeur Marc VAN DROOGENBROECK

Février 2013 (version 6.14)

Remerciements

Pour rédiger ces notes, j'ai bénéficié du concours de C. DEMOULIN et de J.-M. WAGNER. Que par ces mots, ils soient remerciés pour leur aide précieuse et pour leur dévouement.

Prof. Marc VAN DROOGENBROECK

Notations

Cadre analytique :

\mathbb{N}	$\mathbb{N} = \{0, 1, 2, \dots\}$
\mathbb{Z}	$\mathbb{Z} = \{\dots - 2, -1, 0, 1, 2, \dots\}$
\mathbb{R}	Ensemble des réels

Constantes :

j	Vecteur unité de l'axe imaginaire $j = \sqrt{-1}$
k_B	Constante de BOLTZMANN = $1,38 \times 10^{-23}$ JOULE / degré KELVIN

Variables et fonctions :

t	Variable de temps continue
f	Variable fréquence
ω	Variable de pulsation $\omega = 2\pi f$
T	Période d'un signal
T_0	Période fondamentale d'un signal
Δx	Pas d'échantillonnage en x
Δy	Pas d'échantillonnage en y
$x(t)$	Fonction à valeurs continues, définie pour tout temps t
$h(t)$	Réponse impulsionnelle d'un système
$\mathcal{H}(f)$	Transmittance d'un système linéaire
$f[n]$	Fonction échantillonnée, fonction discrète

Variables et fonctions de modulation :

$m(t)$	Signal modulant, normalisé à 1 : $ m(t) \leq 1$
$s(t)$	Signal modulé
k_a	Taux de modulation
β	Indice de modulation
E_b	Énergie par bit
D	Débit binaire [b/s]
R	Rapidité de modulation [baud]

Notations fonctionnelles :

$ a $	Valeur absolue
\vec{a}	Vecteur
$\ a\ $	Norme de a
$\vec{a} \cdot \vec{b}$	Produit scalaire de \vec{a} et \vec{b}
$\vec{a} \times \vec{b}$	Produit vectoriel de \vec{a} et \vec{b}
$\nabla \cdot \vec{a}$	Divergence de \vec{a}
$\nabla \times \vec{a}$	Rotationnel de \vec{a}
$\nabla^2 \varphi$	Laplacien de φ
$\nabla \varphi$	Gradient de φ
X^*	Complexe conjugué de X
$Re(a)$	Partie réelle de a
$Im(a)$	Partie imaginaire de a
$\mathcal{X}(f)$	Transformée de FOURIER du signal $x(t)$
\rightleftharpoons	Correspondance entre un signal et sa transformée
\otimes	Convolution
\oplus	OU exclusif ou addition modulo-2
$\frac{\sin(\pi x)}{\pi x} = \text{sinc}(x)$	Sinus cardinal de x
$\delta(\cdot)$	Fonction delta de DIRAC
$J_n(x)$	Fonction de BESSEL d'ordre n
$erfc(\cdot)$	Fonction erreur complémentaire

Fonctions stochastiques :

$p(A)$	Probabilité de A
\bar{a}	Valeur moyenne de a
μ_X	Espérance mathématique du signal X
σ_X^2	Variance de X
$\Gamma_{XX}(\tau)$	Fonction d'autocorrélation du processus aléatoire stationnaire $X(t)$
$\gamma_X(f)$	Densité spectrale de puissance de $X(t)$
N_0	Densité spectrale de puissance de bruit

Table des matières

Remerciements	ii
Notations	iii
1 Introduction	1
1.1 Généralités	1
1.2 Historique	2
1.3 Organismes internationaux de télécommunications	4
1.4 Structure d'une chaîne de télécommunication	6
1.5 Modèles de référence	8
1.5.1 Modèle de référence OSI	8
1.5.2 Internet	9
1.5.3 D'autres modèles	11
2 Signaux et systèmes de télécommunications	13
2.1 Signaux	13
2.1.1 Types de signaux source	14
Signal vocal ou musical	14
Vidéo	14
Signaux numériques	14
Texte	15
2.1.2 Représentations et catégories de signaux	15
Signaux analogiques ou numériques	15
Signaux périodiques ou apériodiques	16
Signaux déterministes ou stochastiques	16
Signaux d'énergie ou de puissance	17
Décibel	18
Rapport signal à bruit	19

2.1.3	Représentation fréquentielle : la transformée de FOURIER	19
	Égalité de PARSEVAL et théorème de RAYLEIGH	20
2.2	Système	21
2.2.1	Systèmes linéaires permanents	21
2.2.2	Réponse fréquentielle	21
2.2.3	Causalité et stabilité	22
2.2.4	Bande passante	22
2.2.5	Système de transmission idéal	23
	Distorsions de linéarité	24
	Délai de groupe et de phase.	24
	Distorsions non linéaires	25
2.2.6	Bruit et interférences	26
3	Modulation d'onde continue	27
3.1	Introduction	27
3.1.1	Hypothèses	27
3.1.2	Modulation d'une porteuse sinusoïdale	28
3.2	Modulation d'amplitude	29
3.2.1	Modulation d'amplitude classique	29
3.2.2	Analyse spectrale	31
3.2.3	Répartition de puissance	32
3.2.4	Réalisation de la modulation d'amplitude classique	34
	Modulation linéaire	34
	Modulation quadratique	35
	Contrôle de la modulation	36
3.2.5	Démodulateurs	36
	Détection d'enveloppe	36
	Démodulation AM synchrone ou cohérente	37
3.2.6	Avantages et limitations	39
3.2.7	Modulations dérivées	40
3.2.8	Modulation d'amplitude à porteuse supprimée	40
	Réalisation de la modulation	41
	Démodulation synchrone ou cohérente	41
	Boucle de COSTAS	42
3.2.9	Modulation en quadrature	42

3.2.10	Modulation à bande latérale unique	44
	Filtrage d'une bande	44
	Démodulation	44
3.2.11	Modulation à bande latérale résiduelle	46
3.3	Modulation angulaire	46
3.3.1	Principes et définitions	46
	Modulation de phase pure	48
	Modulation de fréquence pure	48
	Notion de préaccentuation	49
	Illustration des techniques de modulation	49
3.3.2	Analyse de la modulation de fréquence analogique	49
	Remarque préliminaire	49
	Modulation par une cosinusoïde	51
	Analyse spectrale	51
	Modulation FM à faible indice	53
	Modulation FM à grand indice	55
3.3.3	Bande passante requise	56
	Estimation empirique	56
	Estimation numérique	56
3.3.4	Réalisation de la modulation et de la démodulation en fréquence	57
3.3.5	Résumé des principaux résultats des modulations d'onde continue	57
3.4	Partage du plan de fréquence	57
3.4.1	Multiplexage en fréquence	57
3.4.2	Récepteur super-hétérodyne	59
3.5	Introduction à la modulation numérique	62
4	Variables aléatoires, processus stochastiques et bruit	63
4.1	Introduction	63
4.2	Théorie des probabilités	63
	4.2.1 Axiomatique des probabilités	64
4.3	Variables aléatoires	67
	4.3.1 Fonctions de répartition et densité de probabilité	67
	4.3.2 Moments d'une variable aléatoire	69
	4.3.3 Liens entre moments centrés ou non	70
	Variable aléatoire centrée	71

	Fonction caractéristique	71
4.3.4	Variables aléatoires usuelles	71
4.3.5	Formule de changement de variable	73
	Moyenne	75
4.3.6	Cas de plusieurs variables aléatoires	75
	Fonctions conjointes et densités marginales	75
	Moments de deux variables aléatoires	77
	Somme	77
	Probabilité conditionnelle et indépendance	78
4.4	Processus stochastiques	79
4.4.1	Définition des processus et suites aléatoires	79
4.4.2	Caractéristiques probabilistes	80
	Moments temporels	80
	Moments statistiques	80
	Fonctions d'intercorrélacion	81
4.4.3	Stationnarité	81
	Stationnarité au sens strict	81
	Stationnarité au sens large	82
	Fonction d'autocorrélacion d'un processus aléatoire stationnaire	83
4.4.4	Intérêt de la stationnarité	83
4.4.5	Ergodicité	84
	Ergodicité dans la moyenne	84
	Ergodicité dans la fonction d'autocorrélacion	84
4.4.6	Densité spectrale de puissance	85
4.4.7	Lien entre stationnarité et ergodicité	86
4.5	Processus stochastiques et systèmes	87
4.5.1	Filtrage des processus aléatoires stationnaires au sens large	87
4.5.2	Somme de processus stochastiques	88
4.5.3	Modulation : mélangeur pour signal stochastique	89
4.6	Une classe de processus stochastiques importante : les sources de bruit	89
4.6.1	Bruit thermique	90
	Ordre de grandeur	90
4.6.2	Bruit de grenaille	91
4.6.3	Modélisation du bruit	91
	Processus gaussien	91

Propriétés	91
4.6.4 Bruit blanc	92
Passage du bruit blanc à travers un filtre passe-bande	92
Passage du bruit blanc à travers un filtre passe-bas	93
4.6.5 Décomposition de RICE	94
Décomposition de RICE d'un bruit blanc à bande étroite	94
4.6.6 Bruit total en sortie	95
4.6.7 Bande équivalente de bruit	95
5 Introduction à la numérisation	96
5.1 Introduction	96
5.1.1 Types d'information	97
5.1.2 Transmission	98
5.2 Conversion des signaux	98
5.2.1 Messages, caractères et symboles	98
5.2.2 Traitement des signaux analogiques	99
Échantillonnage instantané : théorème de SHANNON	99
Formule d'interpolation de WHITTAKER	100
Filtrage préalable à l'échantillonnage	102
Formule de NYQUIST pratique.	102
Réalisation de l'échantillonnage	102
Émission.	102
Conversion analogique-numérique	104
5.3 Modulation PAM	105
5.4 Modulation de la position des impulsions	106
5.5 Modulation d'impulsions codées PCM	109
5.5.1 Quantification	109
5.5.2 Bruit de quantification	112
5.5.3 Codage	115
5.5.4 Modulation d'impulsions codées différentielles (DPCM)	116
5.5.5 Modulation delta	116

6	Transmission de signaux numériques en bande de base	117
6.1	Nécessité du codage	117
6.1.1	Bande passante requise	118
6.1.2	Capacité d'un canal en présence de bruit	118
6.1.3	Transmission des données binaires	118
6.2	Spectre des signaux numériques	120
6.2.1	Modèle théorique linéaire	120
6.2.2	Train d'impulsions de DIRAC	121
	Moyenne d'un train d'impulsions	122
	Fonction d'autocorrélation d'un train d'impulsions	122
	Densité spectrale d'un train d'impulsions	123
	Densité spectrale d'un signal numérique mis en forme	123
	Cas particulier 1 : impulsions rectangulaires	124
	Cas particulier 2 : signaux non-corrélés	124
6.3	Transmission d'impulsions en bande de base	125
6.3.1	Codage	125
6.3.2	Codes en ligne d'émission	125
	Codage NRZ unipolaire	125
	Bande passante.	126
	Codage NRZ bipolaire	128
	Codage RZ unipolaire	129
	Codage RZ bipolaire	130
	Codage MANCHESTER	130
	Comparaison des codes	131
	Autres techniques de codage en ligne	131
	Codage en blocs ou complets	133
6.4	Détection de signaux binaires en présence de bruit gaussien	133
6.4.1	Position du problème	133
6.4.2	Problème simplifié : le cas d'un système binaire	135
6.4.3	Modèle de canal et structure du récepteur	135
6.4.4	Première phase : filtrage ou corrélation	136
	Filtre adapté	136
	Propriétés du filtre adapté	139
	Implémentation du filtre adapté	139
6.4.5	Seconde phase : détection par maximum de vraisemblance	140

	Probabilité d'erreur lors de l'envoi du signal NRZ $g_0(t)$	140
	Probabilité d'erreur lors de l'envoi du signal NRZ $g_1(t)$	143
	Probabilité d'erreur moyenne	144
6.5	Effet de la limitation de la bande passante	145
6.5.1	Diagramme de l'œil	145
6.6	Types de transmission	147
6.6.1	Transmission asynchrone	147
6.6.2	Transmission synchrone	148
7	Modulation numérique	150
7.1	Introduction	150
7.1.1	Modulation cohérente ou incohérente	150
7.1.2	Démodulation cohérente ou incohérente	151
7.2	Modulation	151
7.2.1	Hypothèses	151
7.2.2	Modulation d'amplitude numérique cohérente	151
	Occupation spectrale	151
	Efficacité spectrale.	154
	Démodulation cohérente	154
7.2.3	Modulation de phase numérique cohérente	155
	Occupation spectrale	155
	Démodulation cohérente	155
7.2.4	Modulation d'amplitude en quadrature	156
7.2.5	Modulation de fréquence numérique cohérente	157
	Occupation spectrale	157
	Démodulation cohérente	157
7.2.6	Conclusion	159
7.3	Les modems	159
7.3.1	Modes	160
	Mode simplex	160
	Mode half-duplex	160
	Mode duplex simultané	160
7.3.2	Modulation	161
7.3.3	Protocoles de transfert de données	161
7.3.4	Compression de données	161
7.3.5	Normes	161

8	Notions de code	163
8.1	Introduction	163
8.1.1	Exemple	164
8.2	Modèle	165
8.2.1	Modèle de canal	166
	Canal discret sans mémoire	166
	Canal binaire symétrique	166
	Canal gaussien	166
8.2.2	Flux binaire et redondance	167
8.3	Codes linéaires	167
8.3.1	Un exemple simple : code redondant	167
8.3.2	Définitions et notations	168
8.3.3	Matrice génératrice	169
8.3.4	Codes à parité	170
	Codes à détection de parité simple	170
8.3.5	Code systématique	171
8.4	Détection et correction d'erreur	172
8.4.1	Matrice de contrôle de parité	172
8.4.2	Vecteur syndrome d'erreur	173
8.4.3	Correction d'erreur	174
8.5	Efficacité de codage	174
8.5.1	Poids et distance de HAMMING	174
8.5.2	Distance minimale	175
8.5.3	Détection et correction d'erreur	175
8.6	Quelques codes courants	176
8.6.1	Code de HAMMING	176
8.6.2	Code de GOLAY étendu	176
8.6.3	Code BCH	177
8.6.4	Autres codes	177
9	Propagation et systèmes radio	178
9.1	Introduction	178
9.2	Propagation des ondes électromagnétiques	179
9.2.1	Équations de MAXWELL	179
	Énergie, puissance et impédance électromagnétiques	179

9.3	Antennes	181
9.3.1	Propriétés générales à l'émission	181
9.3.2	Détermination d'un potentiel vecteur	182
9.3.3	Propriétés générales à la réception	182
9.3.4	Antennes simples	183
	Doublet de HERTZ	183
	Doublet magnétique	186
	Antennes longues	187
	Réseaux d'antennes	191
9.3.5	Exemples d'antenne	193
9.4	Bilan de puissance	193
9.4.1	Gain	193
9.4.2	Aire effective	196
9.4.3	Propagation en espace libre : équation de FRIIS	198
9.5	Effets de propagation	199
9.5.1	Polarisation	201
9.5.2	Multitrajet	202
9.6	Domaines d'application de la radio	204
9.6.1	Radiodiffusion sonore	204
	Modulation d'amplitude	204
	Modulation de fréquence	204
10	Principes de fonctionnement du réseau GSM	206
10.1	Introduction	206
10.1.1	Historique	206
10.1.2	Évolution technologique	207
10.2	La technologie GSM	207
10.2.1	Le concept cellulaire	207
	Réutilisation des ressources	209
	Estimation du rapport de puissance porteuse à bruit	210
10.2.2	Synthèse des principales caractéristiques du GSM	210
10.3	Architecture du réseau	211
10.3.1	Le sous-système radio	213
	Le mobile	213
	La station de base (BTS)	214

Le contrôleur de station de base (BSC)	214
10.3.2 Le sous-système réseau	215
Le centre de commutation mobile (MSC)	215
L'enregistreur de localisation nominale (HLR)	216
Le centre d'authentification (AuC).	216
L'enregistreur de localisation des visiteurs (VLR)	216
L'enregistreur des identités des équipements (EIR)	217
10.3.3 Le centre d'exploitation et de maintenance	217
10.3.4 L'acheminement des appels entre le réseau fixe et le réseau GSM : deux scénarios en guise d'exemple	217
10.4 Description du canal physique	218
10.4.1 Multiplexage fréquentiel	218
10.4.2 La modulation	218
10.4.3 Multiplexage temporel	219
10.4.4 Mécanismes de protection contre les interférences radio-électriques	221
Le saut de fréquences ou Frequency Hopping	221
Décalage temporel des envois	222
Le codage	222
10.5 Protocoles	222
10.5.1 Pile de protocoles	222
10.5.2 Les interfaces A-bis, A et X25	223
L'interface A-bis	223
L'interface A	224
L'interface X25	224
10.5.3 Typologie des paquets	224
Le burst d'accès	225
Le burst de synchronisation	226
Le burst normal	226
Le burst de correction de fréquence	226
Le burst de bourrage	226
A Analyse de Fourier	227
A.1 Transformée de FOURIER	227
A.1.1 Définition	227
A.1.2 Propriétés	228

A.1.3	Exemples	229
A.2	La fonction Delta de DIRAC	230
A.2.1	Définition	230
A.2.2	Transformée de FOURIER de $\delta(t)$	231
A.2.3	Applications	231
A.3	Quelques signaux fondamentaux	232
A.3.1	Définitions	232
A.3.2	Transformées de FOURIER	233
B	Formules utiles	234
B.1	Constantes physiques	234
B.2	Relations	234
B.2.1	Relations trigonométriques	234
B.2.2	Développement en série	235
B.3	Notation phasorielle	235
B.4	Inégalité de SCHWARZ	235
	Glossaire	240

Chapitre 1

Introduction

1.1 Généralités

De tout temps, l'être humain a éprouvé le besoin d'échanger de l'information, c'est-à-dire de communiquer. Mais c'est véritablement lors de l'industrialisation de nos sociétés que la multiplication du nombre de moyens de communications a pris une ampleur tout à fait impressionnante. On doit ce développement à l'approfondissement

1. des lois physiques qui régissent le fonctionnement des canaux de communication, mais également
2. de la notion d'information. De fait, la difficulté à transmettre un message est proportionnelle au contenu d'information qu'il renferme.

Lois physiques

Les principales lois physiques sont relatives à l'électricité, jusqu'au développement de l'informatique dont l'apparition a permis d'accroître le nombre de canaux de communication. Aujourd'hui, même des machines s'échangent de l'information sans intervention humaine.

Dans ce document, nous envisageons les télécommunications sous la forme restrictive d'échanges d'information entre des systèmes, à l'initiative de l'être humain. L'analyse des signaux et des systèmes de télécommunications est complexe car, pour que deux systèmes communiquent, il faut combiner toute une série d'éléments que l'on retrouve, entre autres, dans le modèle de référence **OSI** (Open Systems Interconnection) : adaptation du signal au support de transmission, détection d'erreurs, synchronisation, etc. Parmi tous ces éléments, l'accent sera mis sur le fonctionnement des couches basses, autrement dit sur l'aspect physique des communications.

Information

L'approche moderne des télécommunications se fonde sur les particularités des signaux à transmettre. Il est dès lors utile de préciser la notion d'information. Un texte, un graphique, un son ou une image traduisent des facettes différentes d'information. Quant il s'agit de quantifier leur "richesse", autrement dit leur contenu en information, il est commode de représenter ces signaux en une série de symboles équivalents. Ainsi, on peut remplacer chaque caractère d'un texte par une suite de huit ou seize symboles 0 ou 1, la correspondance entre le caractère et

l'octet (ou le double octet) ainsi constitué étant prédéfinie. On voit ainsi apparaître la notion de code.

Définition 1 *Un code est une règle de représentation de caractères pris dans un ensemble par des caractères pris dans un autre ensemble.*

Par exemple, les mots sont des suites de sons choisis dans un ensemble d'environ 40 caractères appelés phonèmes.

Il faut bien voir qu'un code établit une règle de correspondance entre deux ensembles de symboles, règle appelée **codage**; la forme matérielle de ces symboles n'est pas nécessairement unique. Illustrons cela par un exemple; soit à transmettre une information codée sous forme binaire en utilisant une ligne de transmission et en consacrant une durée fixée pour la transmission de chaque 0 ou 1 du message. La forme physique que pourraient prendre les signaux représentant ces deux symboles peut être

- l'existence ou l'absence d'un courant continu,
- une tension V_0 ou V_1 ,
- une sinusoïde de fréquence f_0 ou f_1 ,
- ou tout autre choix de deux fonctions temporelles distinctes pendant l'intervalle de temps.

Pour la transmission, l'information se trouve donc représentée par une ou plusieurs grandeurs physiques fonction du temps, et parfois aussi des coordonnées de l'espace, comme en télévision.

L'étude des formes que peut prendre l'information est l'objet de la théorie de l'information. Cette théorie part du postulat que l'information est imprévisible, et considère qu'un message est la réalisation d'une expérience au sens de la théorie des probabilités. Plus la probabilité d'arrivée d'un message déterminé est grande, moins ce message n'apporte d'information. À la limite, le message n'apporte plus aucune information; il est alors déterministe.

1.2 Historique

C'est l'ingénieur et écrivain français ESTAUNIÉ qui utilisa pour la première fois le terme de télécommunications, dans son *Traité pratique de télécommunication électrique*, publié en 1904. Le terme vient du grec télé (au loin, distant) et du latin *communicatio*. Quoique le terme soit relativement récent, le besoin de communication à distance date déjà de l'Antiquité. C'est ainsi que l'on sait que, trois mille ans avant Jésus-Christ, les matelots grecs utilisaient déjà des pigeons voyageurs pour envoyer des messages vers la côte. Dans le Midi de la France, on trouve encore, à Arles et à Nîmes, les ruines des tours de sémaphores construits par les Romains qui émettaient des signaux en agitant des flambeaux [19].

Le système le plus connu de télégraphie optico-mécanique porte le nom de son inventeur, le français CHAPPE (1763–1805). Le système se compose d'un mât surmonté d'une poutre transversale mobile. Un bras de signalisation est fixé aux deux extrémités de cette poutre. À l'aide de leviers, de roues et de cordes, la combinaison poutre et bras est placée dans différentes positions; il y a ainsi $4 \times 8 \times 8 = 256$ combinaisons possibles dont seules 92 étaient utilisées pour des raisons pratiques. On envoyait ainsi, sur une distance moyenne d'une dizaine de kilomètres, des messages visuels. Le réseau CHAPPE se développa en France à partir de 1793 et finit par atteindre 5,000 km rien que pour la France.

C'est au cours de l'hiver 1835–1836 que MORSE s'attaqua à l'invention du télégraphe électrique. Ses principes de base s'appuyaient sur l'ouverture et la fermeture d'un circuit électrique comme

émetteur, un électro-aimant comme récepteur pour convertir le signal électrique en un mouvement, et un code pour traduire ce mouvement en lettres et en chiffres. Fait remarquable, ce code inventé par MORSE s'adaptera à tous les moyens de communication futurs et notamment à la radio, aussi bien en phonie qu'en graphie. Les appareils morse furent mis en service en Belgique en 1853. En 1879, 95% des postes utilisaient ce système.

Cette télégraphie avait donné lieu à la création de puissantes sociétés, précurseurs des opérateurs nationaux actuels. Son importance allait pourtant décroître avec l'invention du téléphone par BELL en 1876. Le développement du téléphone fut extrêmement rapide. Ainsi, les États-Unis comptaient déjà 47.000 téléphones en 1880.

En 1888, HERTZ met le rayonnement électromagnétique en évidence. Cela conduira, au début du vingtième siècle, à la radiotélégraphie. Celle-ci est techniquement bien modeste. On génère une "haute" fréquence, de quelques kHz, au moyen d'alternateurs à pôles multiples, et l'on interrompt cette sinusoïde comme dans le cas du télégraphe à fil.

PUPIN invente et met au point les bobines de charge en 1900. Ces bobines permettront de réduire l'affaiblissement des câbles souterrains aux fréquences vocales. Les câbles pupinisés seront utilisés pendant un tiers de siècle. Notons qu'il n'y a plus de câble de ce type en Belgique mais qu'on en trouve encore aux États-Unis.

HEAVISIDE et KENNELLY prévoient en 1902 que les ondes courtes pourront contourner la terre par réflexion sur l'ionosphère. Pendant un demi-siècle, les ondes courtes seront le seul moyen de communication intercontinentale en téléphonie.

Il faudra attendre 1906 pour que, avec l'invention de la triode par LEE DE FOREST, l'électronique voie le jour et, avec elle, l'utilisation de fréquences de plus en plus élevées. C'est vers 1914 que l'on imaginera de faire varier l'amplitude de la sinusoïde pour y imprimer un message vocal. C'est l'invention de la modulation, d'amplitude d'abord, de fréquence ensuite. C'est aussi l'origine de la radiodiffusion et puis de la radiotéléphonie.

Mais les progrès de la radioélectricité ne laisseront pas les téléphonistes indifférents. Ceux-ci comprendront que la technique de multiplexage par répartition de fréquence, utilisée assez naturellement par les radioélectriciens pour se partager l'espace de propagation, pouvaient tout aussi bien s'appliquer au partage de l'utilisation d'un câble entre de nombreuses communications, pour autant que la bande de fréquences utilisables soit suffisante.

Pour la transmission à grande distance, on trouvera une solution très élégante à partir des années 1960 avec les satellites de communication ; dans une vision très simplifiée des choses, ceux-ci sont des relais radioélectriques placés dans l'espace extra-terrestre et donc visibles à partir d'endroits très éloignés sur la surface du globe. L'orbite géostationnaire est à cet égard particulièrement intéressante. Il s'agit d'une orbite circulaire, situé dans le plan de l'équateur, et de rayon tel qu'un satellite tourne autour de la terre en synchronisme avec la rotation de celle-ci et apparaisse donc immobile. L'altitude de cette orbite est de 36.000 km, ce qui correspond à un rayon de quelques 42.700 km.

Heuristiques à leur début, les télécommunications avaient besoin d'une base théorique solide. Celle-ci leur fut fournie principalement durant la décennie 1930–1940, par des grands noms comme NYQUIST, SHANNON, WIENER, etc. Les théories développées sont nombreuses et comprennent entre autres celles du signal, de l'information et du codage. Elles sont également à la base de l'informatique.

Cette dernière n'aurait pu se développer sans un nouvel apport technologique : le transistor, attribué à SCHOCKLEY (1948), et ensuite des techniques de la microélectronique. Seule

la miniaturisation à outrance permettra de réaliser les opérations complexes requises par l'informatique. Mais elle sera également nécessaire pour mettre en œuvre, dans le domaine des télécommunications, une des découvertes théoriques des années 1930, à savoir la possibilité de numériser des signaux de nature essentiellement analogique, et d'apporter ainsi au domaine des télécommunications toutes les ressources de l'informatique.

Au cours de cette évolution, les besoins et les nouveaux services n'ont fait que croître. Ainsi en est-il de la transmission d'images. Inventée, au niveau des principes, dans les années 1930, la télévision s'est implantée dans les années 1950 sous la forme d'un service de radiodiffusion. Elle s'est alors vu réserver plus de la moitié des fréquences jusqu'à 1 GHz.

On peut considérer que c'est en 1975 qu'apparaissent les premières liaisons expérimentales de télécommunication par fibres optiques. En particulier, à Atlanta, une liaison expérimentale à 450 [Mb/s] est établie sur un trajet de 650 mètres; la perte en ligne est de l'ordre de 8 [dB].

Les dernières inventions qui auront le plus marqué la fin du vingtième siècle sont sans aucun doute le réseau mondial de transmission de données Internet et la mobilophonie (cf. figure 1.1). Il y a fort à parier que ces deux inventions modifient le rapport de la transmission vocale à la transmission de données ainsi que le rapport des appels téléphoniques à partir de téléphones fixes ou mobiles. Fin 2000, on comptait plus de 5.000.000 de téléphones mobiles en Belgique pour 4.000.000 de lignes fixes!

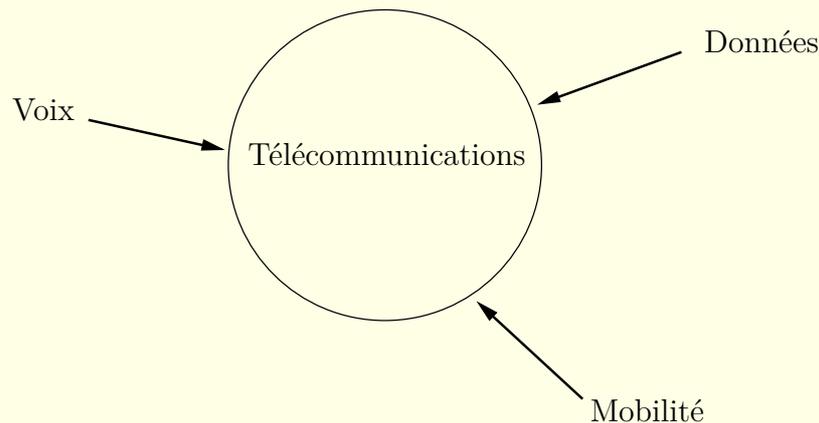


FIGURE 1.1 – Principaux pôles de développement en télécommunications.

Enfin, on ne saurait passer sous silence l'introduction massive de l'informatique dans les systèmes de télécommunications; aujourd'hui, il n'y a pas de nouveau système qui n'intègre un microprocesseur.

1.3 Organismes internationaux de télécommunications

De par leur nature, les télécommunications demandent une exploitation coordonnée; on ne peut communiquer sans une infrastructure offrant un très haut degré de compatibilité à tous les utilisateurs, tant au niveau national qu'international.

Le terme télécommunication n'a été défini officiellement, pour la première fois, qu'à la Conférence internationale de Madrid en 1932. En fait, jusqu'à cette date, il existait deux entités juridiques différentes, la Conférence internationale télégraphique créée en 1865, d'une part, et la Conférence internationale radiotélégraphique, dont la première réunion s'était tenue à Berlin

en 1906. En 1932, ces deux conférences se réunirent à nouveau, mais ensemble à Madrid. Après discussions, on s'entendit pour trouver un nom à cette nouvelle organisation : l'Union Internationale des Télécommunications (UIT) ou International Telecommunications Union (ITU¹). C'est en 1947, à Atlantic City, que se tint la première grande conférence internationale de l'après-guerre. Les relations entre l'Organisation des Nations Unies et l'Union Internationale des Télécommunications furent précisées au cours de conférence et, après bien des discussions, l'ONU donna son accord pour que l'**ITU** soit reconnue comme l'agence spécialisée en matière de télécommunications. Une nouvelle définition du mot télécommunication fut également adoptée lors de cette conférence. Cette version constitue la définition officielle actuelle.

Définition 2 *On entend par télécommunication toute transmission, émission ou réception de signes, de signaux, d'écrits, d'images, de sons ou de renseignements de toute nature par fil, radio-électricité, optiques ou autres systèmes électromagnétiques.*

Les structures, sur le plan national, varient assez fort d'un pays à l'autre. Sur le plan international, le pouvoir de normalisation en matière de télécommunications revient à l'ITU, qui agit de concert avec l'International Standards Organisation (ISO²). L'Union comprend, entre autres, deux branches l'ITU-T et l'ITU-R qui ont la mission d'établir des normes pour les télécommunications et de partager les ressources. L'ITU organise aussi régulièrement des CMAR (Conférence Administrative Mondiale des Radiocommunications) ou WARC (World Administrative Radio Conference) au cours desquelles les pays conviennent de l'attribution des fréquences et du partage des positions sur l'orbite géostationnaire. Par exemple, c'est au cours de la CMAR de 1992 qu'il a été convenu de réserver les bandes de fréquences 1885 – 2025 [MHz] et 2110 – 2200 [MHz] pour les systèmes mobiles de troisième génération.

À côté de cette organisation mondiale, il existe un certain nombre d'organismes internationaux dont la mission est de caractère régional ou spécifique. Parmi ceux-ci, on peut citer

- l'UER³ (Union Européenne de Radio-Télévision, en anglais EBU : European Broadcasting Union), qui organise la coopération entre organismes de radiodiffusion sonore et télévisuelle pour la zone européenne,
- INTELSAT⁴, qui gère le système mondial de communication par satellite,
- EUTELSAT⁵, qui fait de même pour le système régional européen,
- l'ETSI⁶ (European Telecommunications Standards Institute), dont le succès majeur est la création de la **norme GSM**,
- l'IETF⁷ (Internet Engineering Task Force), qui gère les développements techniques d'Internet.

On pourrait encore citer des organisations telle que l'URSI (Union Internationale de Radio-électricité Scientifique), dont la mission est de nature strictement scientifique et consiste à encourager et à coordonner les recherches sur les ondes électromagnétiques, guidées ou non, et sur leur utilisation, notamment en vue des télécommunications, ainsi que de nombreuses autres sociétés scientifiques.

Bien qu'il ne soit pas strictement lié à la normalisation, un nouvel acteur est venu s'ajouter dans le monde des télécommunications. Avec la suppression des monopoles des opérateurs publics

1. <http://www.itu.int>

2. <http://www.iso.ch>

3. <http://www.ebu.ch>

4. <http://www.intelsat.com>

5. <http://www.eutelsat.com>

6. <http://www.etsi.org>

7. <http://www.ietf.org>

en Europe, chaque pays a du créer un organisme régulateur, chargé de veiller au maintien de règles de concurrences saines. En Belgique, il s'agit de l'IBPT⁸, l'Institut Belge des services Postaux et Télécommunications.

1.4 Structure d'une chaîne de télécommunication

La structure conventionnelle d'une chaîne de télécommunications comprend différents éléments repris à la figure 1.2. On part d'un message, d'une information, transmis sous une forme matérielle déterminée. Le système de transmission étant de nature électromagnétique, il convient de convertir le signal physique en un signal électrique par le biais du transducteur d'émission. La chaîne se termine de même par un transducteur de réception, dont le but est de transformer le signal électrique en une grandeur physique adaptée au correspondant destinataire.

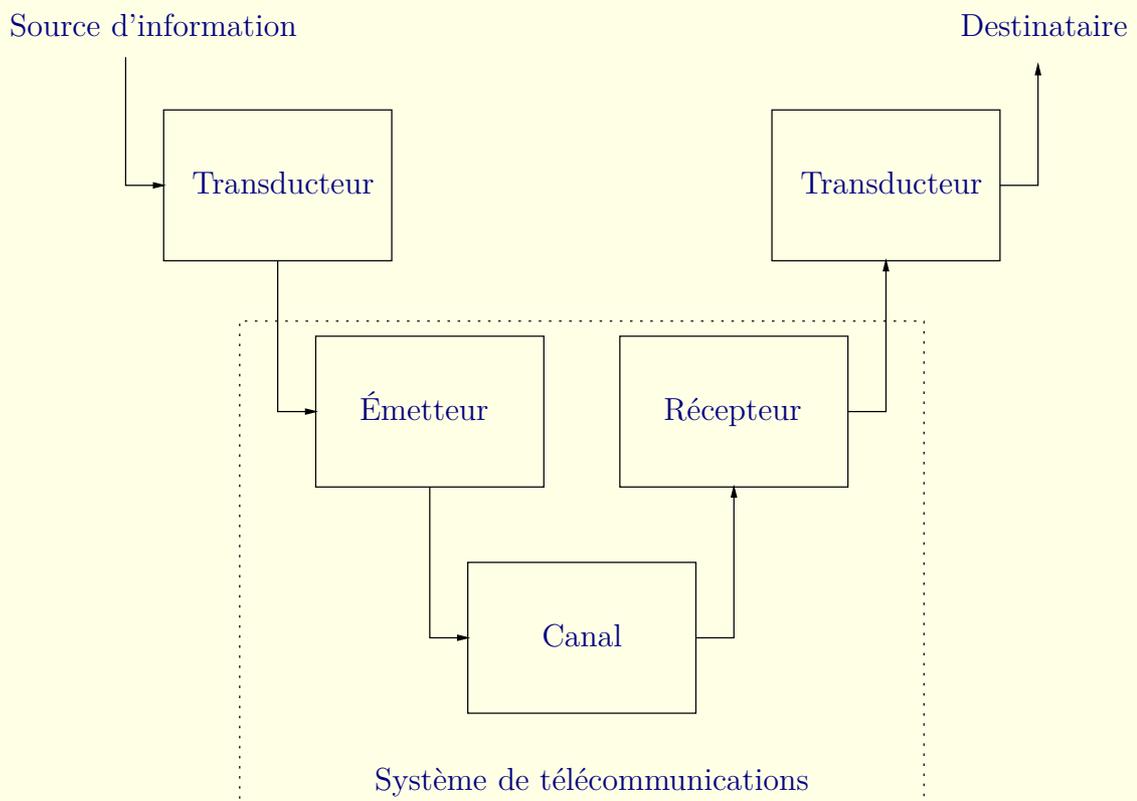


FIGURE 1.2 – Structure d'une chaîne de télécommunications.

Le problème de la transduction est essentiellement un problème de fidélité; ce problème technologique peut être résolu de façon plus ou moins satisfaisante selon le compromis qualité-coût que l'on accepte.

Le signal électromagnétique délivré par le transducteur d'émission doit être adapté au canal de transmission. À cet effet, on lui fait subir certaines transformations auxquelles on peut donner le nom générique de codage. Ces opérations comprennent aussi bien les opérations de filtrage que des mises à niveau électrique ou des convertisseurs d'analogique en numérique. Au terme du codage, que l'on qualifie le plus souvent de codage de source, le signal est injecté à l'entrée d'un système qui effectue le codage de canal ou **modulation**.

8. <http://www.ibpt.be>

Bien entendu, on opère à l'extrémité réceptrice des opérations de décodage inverses des transformations effectuées à l'extrémité émettrice.

Si la structure de la figure 1.2 correspond assez bien à une chaîne de télécommunications analogiques, de nombreux éléments supplémentaires viennent s'ajouter dans le cas d'une communication numérique. Le schéma de la figure 1.3 en dresse un portrait plus typique. On y voit que la transmission est précédée de toute une série d'opérations de traitement sur le signal de départ.

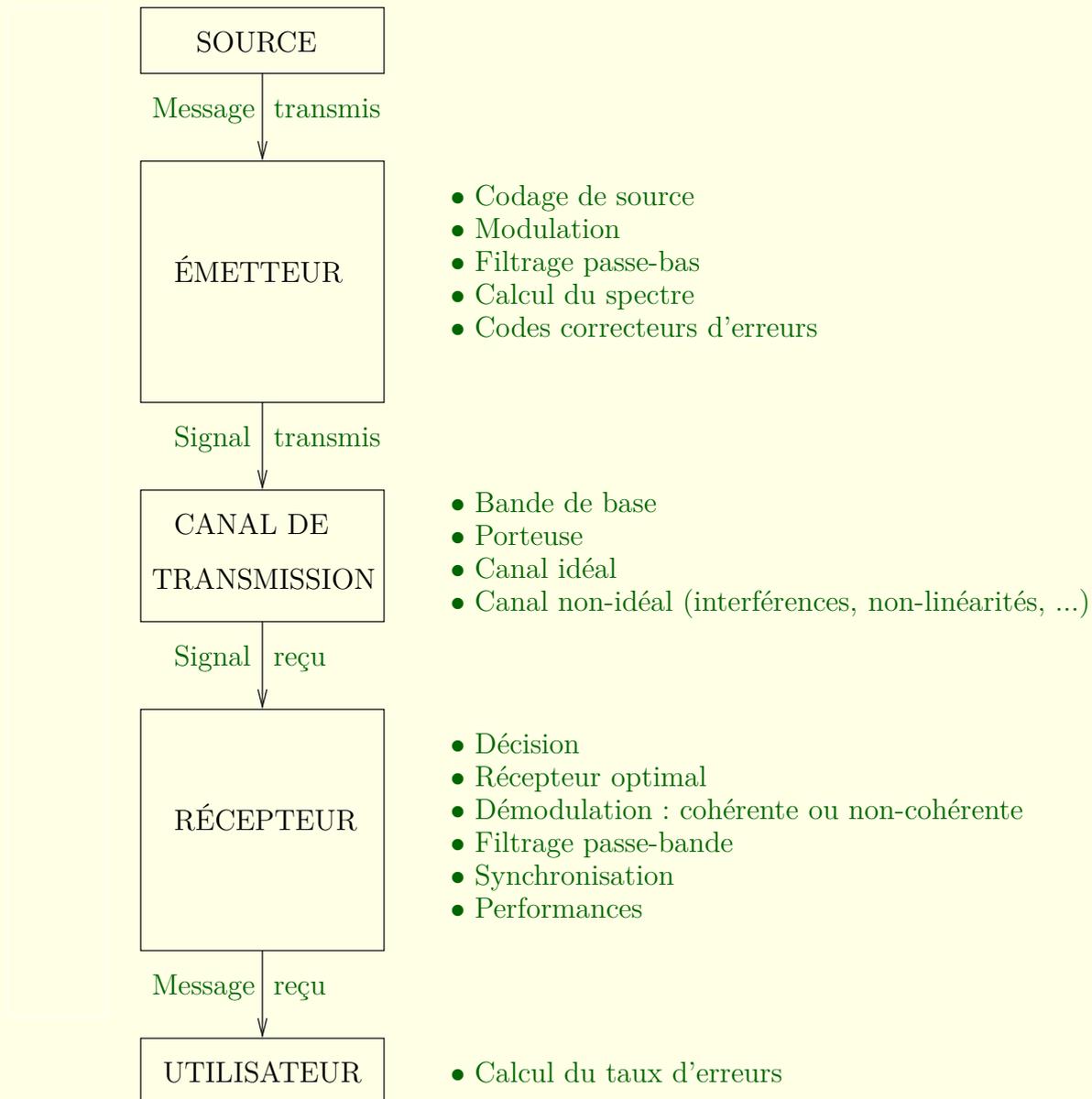


FIGURE 1.3 – Structure d'une chaîne de télécommunications numérique [2].

On peut encore ajouter que le schéma général d'une chaîne de télécommunication donné ne représente pas toutes les possibilités. Il présuppose que l'on crée, au profit des correspondants, une voie de communication, appelé **circuit**, mise à leur disposition exclusive pour la durée de la communication. Et même si l'exclusivité ne signifie pas que le canal de transmission ne puisse pas être partagé avec d'autres utilisateurs grâce à une opération de multiplexage, il n'empêche que le circuit est affecté à la communication pour toute sa durée. On ne retrouve

pas toujours ce concept en transmission de données où certains protocoles prévoient l'envoi de paquets d'information, sans établissement préalable d'un circuit. C'est le cas notamment du protocole réseau utilisé pour l'Internet : le protocole **IP**.

1.5 Modèles de référence

La communication passe obligatoirement par la mise en réseau des terminaux. Cette dernière nécessite alors l'établissement de conventions claires et non ambiguës, d'autant plus qu'on assiste à un accroissement de l'interopérabilité des réseaux informatiques et des réseaux de télécommunications.

En transmission de données, il existe plusieurs architectures dont les principales sont l'architecture provenant de la normalisation de l'**ISO** que l'on appelle **Open System Interconnection (OSI)** et l'architecture utilisée sur Internet baptisée **TCP/IP**, du nom des deux principaux protocoles qui la constituent.

1.5.1 Modèle de référence OSI

Le principe adopté dans la conception des réseaux est d'être le plus indépendant possible des supports physiques et de regrouper les fonctions de communication en catégories. Le modèle de référence développé par l'ISO comporte 7 couches; la figure 1.4 montre la communication entre deux ordinateurs à travers un réseau suivant le modèle **OSI**.

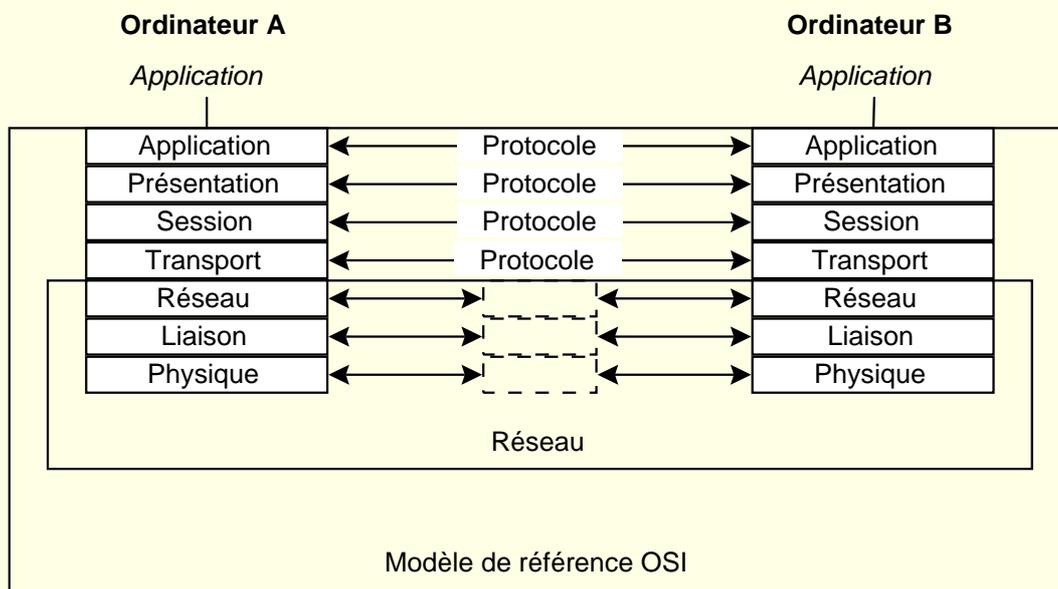


FIGURE 1.4 – Modèle de référence OSI.

Les concepts architecturaux utilisés pour décrire le modèle de référence sont décrits dans la norme 7498-1. Le concept d'architecture en couches demande la définition de trois objets pour chaque niveau N :

- le service. Il correspond aux événements et aux primitives associées, à mettre en place pour rendre un service au niveau supérieur, c'est-à-dire au niveau $N + 1$;

- le **protocole**. Le protocole de niveau N définit un ensemble de règles nécessaires pour que le service de niveau N soit réalisé. Ces règles définissent les mécanismes qui vont permettre de transporter les informations d'un niveau N au niveau N d'une autre machine. En particulier, le protocole N va proposer les règles pour contrôler l'envoi des données ;
- les points d'accès au service N (Service Access Point ou SAP en anglais). Les points d'accès au service N sont situés à la frontière entre les couches $N + 1$ et N . Les services N sont fournis par une entité N à une entité $N + 1$ à ces points d'accès aux services N .

Un service et un protocole sont déterminés par trois attributs :

- la sémantique d'association. Elle peut être de deux types : soit orienté connexion (connection oriented), soit non orienté connexion (connectionless). Dans un dialogue avec connexion, les phases d'établissement de la connexion entre les deux entités communicantes, de transfert de données et de fermeture de connexion se succèdent.
- la sémantique de fonctionnalité. Elle regroupe l'ensemble des procédures qui seront utilisées pendant la phase de transfert de données. Pour une association avec connexion, les principales fonctions sont la fragmentation, la concaténation, le contrôle de flux, le contrôle d'erreurs, etc. La figure 1.5 reprend une série de fonctions pour chaque couche.
- la syntaxe. Il caractérise le codage des primitives de service et des unités de données de protocole, appelées PDU (Protocol Data Unit).

Comme nous l'avons vu, le modèle de référence est structuré en couches. Pour transmettre des données, une couche utilise le service d'une couche inférieure. Une unité de données du service N , ou N-SDU (pour Service Data Unit) est un ensemble de données provenant de l'interface avec la couche N qui doit être transporté sur une connexion N . Les informations de contrôle du protocole N , nécessaires à la coordination, sont rajoutées, le cas échéant, à des SDU sur une connexion $N - 1$. Les règles définies par les protocoles permettent d'offrir une série de fonctionnalités pour une communications entre deux terminaux ; la figure 1.5 énonce quelques fonctionnalités couramment rencontrées dans des systèmes réels.

1.5.2 Internet

Devant le foisonnement de machines utilisant des protocoles de communication différents et incompatibles, la défense américaine a décidé de définir sa propre architecture. Cette architecture est à la base du réseau mondial Internet.

L'architecture Internet, qui se présente sous la forme d'une pile de protocoles, est basée sur le protocole **IP** (Internet Protocol) qui correspond au niveau 3 de l'architecture du modèle de référence OSI ; quelques protocoles sont repris à la figure 1.6.

Le protocole IP a pour but de transporter les paquets, appelés datagrammes, d'une extrémité à l'autre du réseau. Les paquets sont indépendants les uns des autres et sont routés individuellement dans le réseau par chaque commutateur. La sécurisation apportée par ce protocole est très faible : pas de détection de paquets perdus ou de possibilité de reprise sur erreur. Le protocole **TCP** (Transport Control Protocol) regroupe les fonctionnalités de niveau 4 du modèle de référence. C'est un protocole assez complexe qui possède de nombreuses options permettant de résoudre tous les problèmes de perte de niveau inférieur. En particulier, les pertes pourront être récupérées par retransmission sur le flot d'octets. Le protocole TCP est en mode connecté, contrairement au deuxième protocole disponible dans cette architecture qui s'appelle **UDP** (User Datagram Protocol). Ce protocole se positionne aussi au niveau transport mais dans un mode sans connexion et avec pratiquement aucune fonctionnalité.

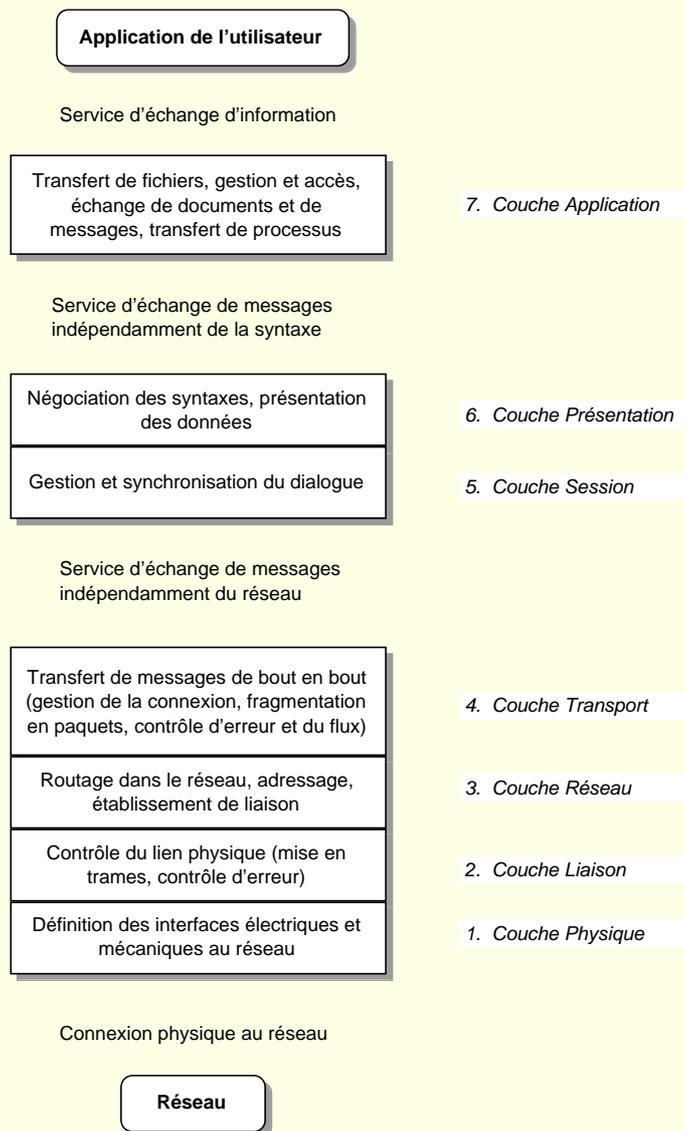


FIGURE 1.5 – Les principales fonctions des couches OSI.

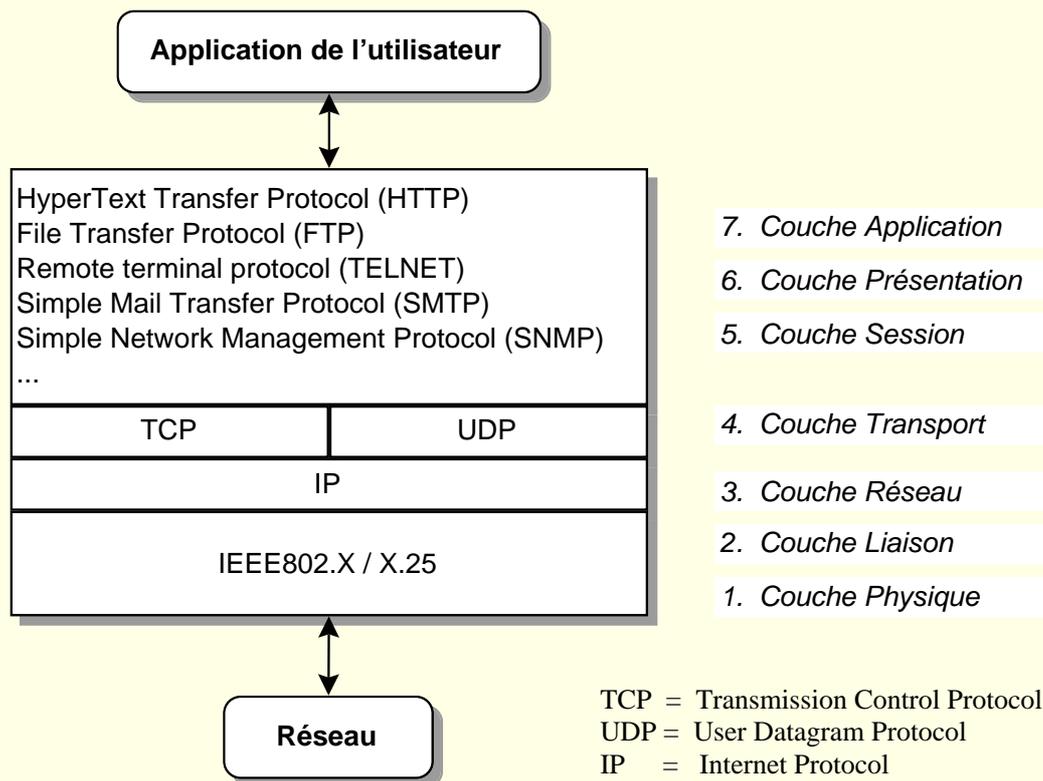


FIGURE 1.6 – Éléments de l'architecture TCP/IP.

En guise d'illustration, la figure 1.7 analyse le contenu d'un paquet TCP/IP généré lors d'une connexion de transfert de fichiers par le protocole FTP (File Transfer Protocol).

Toute la puissance de l'architecture TCP/IP provient de la souplesse de mise en place au-dessus des réseaux existants. Par ailleurs, l'architecture inclut également, sans qu'elle ne soit définie, une interface d'accès au réseau. En effet, de nombreux sous-réseaux peuvent être pris en compte dans l'architecture TCP/IP, aussi bien de type réseaux locaux que réseaux longue distance. Cette généralité peut parfois être un défaut en ce sens que l'optimisation globale du réseau est effectuée par sous-réseau. En revanche, le coût de l'infrastructure est extrêmement bas. Le service rendu par ce réseau est du type "best effort", ce qui signifie que le réseau fait ce qu'il peut pour écouler le trafic.

1.5.3 D'autres modèles

Le modèle de référence OSI a inspiré d'autres modèles. Ainsi, l'ITU-T a développé un modèle de référence en 1993 pour la transmission de données par commutation de cellules, mode appelé Asynchronous Transfer Mode (ATM). La raison en est simple : il faut que les nouveaux réseaux puissent prendre en compte les applications multimédia.

Cette architecture ITU-T peut être dite compatible au modèle de référence de l'ISO. Cependant, les fonctionnalités ne sont pas regroupées aux mêmes niveaux que ceux de l'ISO.

En fait, il existe quantité de modèles de communication, éventuellement partiels, basés sur le modèle de référence. Cela va du réseau de signalisation n^07 , au Réseau téléphonique Numérique à Intégration de Services (RNIS) en passant par le GSM ou le réseau à très haut débit SDH (Synchronous Data Hierarchy).

No.	Source	Destination	Protocol	Info
53	circpc8.montefiore.ul	stoch01.montefiore.ul	TCP	6000 > 42487 [PSH, ACK] Seq=603390095 Ack=3
54	stoch01.montefiore.ul	circpc8.montefiore.ul	TCP	42487 > 6000 [ACK] Seq=3487107646 Ack=60339
55	ring.aist.go.jp	circpc12.montefiore.u	TCP	ftp-data > 1174 [ACK] Seq=1209320986 Ack=39
56	ring.aist.go.jp	circpc12.montefiore.u	FTP	Response: 200 PORT command successful.
57	161.253:2	0.255:2	DDP	AppleTalk Name Binding Protocol packet

[-] Frame (84 on wire, 84 captured)	
[-] Ethernet II	
-	Destination: 00:00:00:00:00:00 (00:00:00:00:00:00)
-	Source: 00:60:83:7c:04:eb (Cisco_7c:04:eb)
-	Type: IP (0x0800)
[-] Internet Protocol	
-	Version: 4
-	Header length: 20 bytes
[-]	Type of service: 0x10 (Minimize delay)
-	Total Length: 70
-	Identification: 0x4d87
[-]	Flags: 0x4
-	Fragment offset: 0
-	Time to live: 231
-	Protocol: TCP
-	Header checksum: 0x0a9c
-	Source: ring.aist.go.jp (150.29.9.6)
-	Destination: circpc12.montefiore.ulg.ac.be (139.165.16.182)
[-] Transmission Control Protocol	
-	Source port: ftp (21)
-	Destination port: 1165 (1165)
-	Sequence number: 1178723011
-	Acknowledgement number: 3911994
-	Header length: 20 bytes
[-]	Flags: 0x18
-	Window size: 8760
-	Checksum: 0xb5f3
[-] File Transfer Protocol	
-	Response: 200
-	Response Arg: PORT command successful.

0000	00 00 00 00 00 00 00 60	83 7c 04 eb 08 00	45 10' E.
0010	00 46 4d 87 40 00 e7 06	0a 9c 96 1d 09 06 8b a5		.FM.@...
0020	10 b6 00 15 04 8d 46 41	e2 c3 00 3b b1 3a 50 18	FA:P.
0030	22 38 b5 f3 00 00 32 30	30 20 50 4f 52 54 20 63		"8....20 0.PORT.c
0040	6f 6d 6d 61 6e 64 20 73	75 63 63 65 73 73 66 75		ommand.s uccessfu
0050	6c 2e 0d 0a			l...

FIGURE 1.7 – Analyse du contenu d'un paquet IP.

Chapitre 2

Signaux et systèmes de télécommunications

Dans le contexte actuel, on peut définir les télécommunications comme le moyen de transporter des signaux d'un point à un autre le plus efficacement et le plus économiquement possible. Il va sans dire que la réalisation de cet objectif dépendra de la nature du signal à transporter. En téléphonie, un délai de 50 [ms] ne perturbe pas le cours naturel de la communication mais, au-delà, les correspondants perçoivent le délai. Cette contrainte n'existe pas pour la consultation d'informations par Internet via le protocole HTTP. Enfin pour le signal de télévision, qui combine son et images, le système de transmission doit permettre le transfert d'un flot de deux informations synchronisées ; on parle de signal composite.

Pour des raisons historiques, la majorité des systèmes de télécommunications ont été développés pour un signal spécifique. Cela pose les problèmes qu'on devine lorsqu'il s'agit de transmettre un signal d'un autre type. Ainsi, le réseau Internet, développé initialement pour la transmission de données, s'accommode mal de la transmission de la parole. Il est dès lors important de comprendre en quoi consistent les sources d'information et les contraintes qui leur sont propres.

À la difficulté de traiter des signaux de nature fort variée s'ajoute la question de la modélisation des composantes d'une chaîne de transmission. En effet, construire un système optimal exigerait la connaissance exacte des signaux à transmettre. Or, le contenu en information étant imprévisible, on ne pourra jamais qu'émettre des hypothèses sur la nature de la source d'information et s'en servir pour la constitution d'un modèle.

Ce chapitre vise à décrire les signaux rencontrés en télécommunications et à introduire les principaux modèles de signaux et de systèmes utilisés en pratique ; il n'a pas la prétention d'être exhaustif en la matière.

2.1 Signaux

En général, le message à transmettre est représenté physiquement par une fonction continue du temps, $x(t)$ par exemple. L'information contenue dans le message réside dans la forme de cette fonction $x(t)$. On en déduit que le rôle du récepteur consiste à analyser la forme de $x(t)$ pour en déduire le signal d'information émis par la source.

2.1.1 Types de signaux source

Signal vocal ou musical

Le son consiste en une onde de pression se propageant dans l'air. Ce phénomène physique peut être étudié de manière quantitative et les propriétés de l'onde acoustique peuvent être exprimées sous la forme de grandeurs objectives. L'une de celles-ci est l'intensité acoustique I qui est la puissance transportée par unité de surface du front d'onde et s'exprime, par exemple, en $[W/cm^2]$. De telles grandeurs sont certes fort utiles, mais lorsqu'il s'agit de déterminer les conditions de bonne transmission de signaux sonores, il apparaît nécessaire de tenir compte des propriétés perceptives très particulières de l'ouïe. On fait alors appel à des grandeurs subjectives qui expriment ce que ressent l'être humain.

Notre oreille perçoit théoriquement des vibrations acoustiques allant de $20 [Hz]$ (son grave) à $20.000 [Hz]$ (son aigu). C'est la bande de fréquence qu'il faut restituer convenablement dans les systèmes haute fidélité, comme la radio **FM**, le canal son de la télévision ou le CD-audio. En radio **AM** cependant, on ne reproduit les fréquences que jusqu'à $4500 [Hz]$ et cela procure une qualité relativement bonne, même pour la musique. En téléphonie, où l'objectif est limité à l'intelligibilité du message, la bande de fréquence normalisée est $300 - 3400 [Hz]$. Cette bande de fréquences est garantie de bout en bout du réseau. Un modem devra donc utiliser cette même gamme de fréquences pour permettre l'information numérique sous peine de perte d'information.

Vidéo

Le signal vidéo est nettement plus complexe qu'un son. Il comporte des informations visuelles, exprimées pour une image généralement sous la forme de trois couleurs fondamentales pour tout point de l'espace, et une information temporelle née du mouvement des objets. De plus, le signal de télévision est un mélange du signal vidéo et de la partie audio ; il s'agit d'un signal composite. Le signal composite occupe une bande de fréquences d'approximativement $5 [MHz]$. En pratique, on considère une bande de $8 [MHz]$ par signal vidéo en Europe. La figure 2.1 donne l'occupation fréquentielle d'un signal vidéo **PAL**.

Signaux numériques

Définition 3 *Le **bit** est l'information élémentaire en informatique. Il ne peut prendre que deux valeurs, 0 ou 1. En électronique, il est facilement représenté par des tensions différentes. Un **octet**, ou **byte** en anglais, est un ensemble de 8 bits.*

Les signaux numériques peuvent prendre différentes formes et il faut s'entendre sur la définition du terme de signal numérique. La définition que nous avons choisie est celle d'un signal dont la valeur est constante sur un certain intervalle de temps. Ainsi, pour deux signaux à représenter 0 et 1, on peut transmettre respectivement un cosinus ou un sinus pour l'intervalle de temps correspondant au symbole. Cette forme analogique du signal qui en résulte est l'expression d'un signal fondamentalement numérique. Bien sûr, on peut avoir un signal analogique de valeur constante durant l'intervalle de temps.

Définition 4 *Le nombre de symboles transmis pendant une seconde est mesuré en **baud**.*

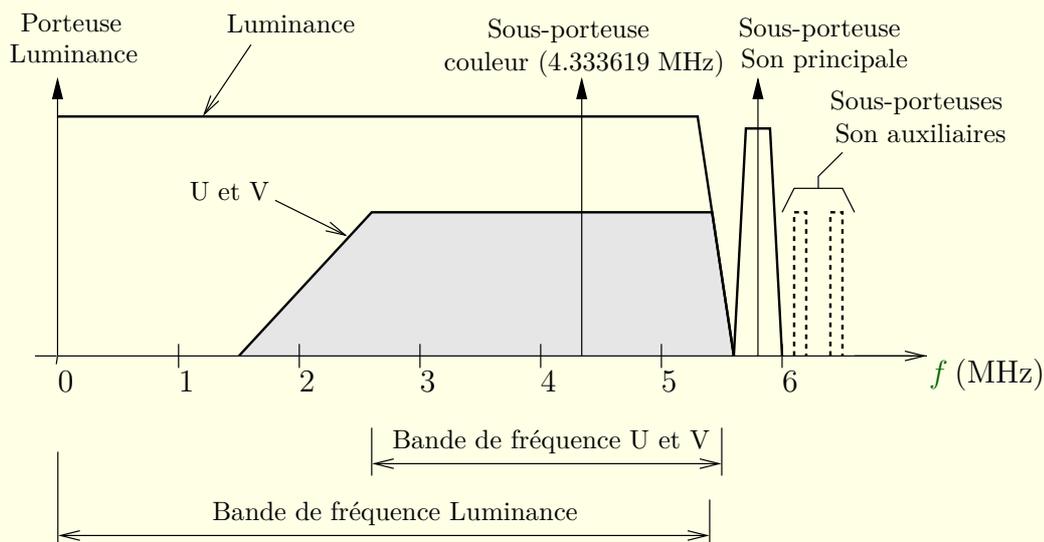


FIGURE 2.1 – Spectre d'un signal vidéo PAL.

Si l'on transmet la séquence 2, 3, 1, 0, 3, 1, 1, 0 en une seconde, le débit vaut 8 [baud] . On parlera de la mesure du **bit**, noté $[b]$, pour caractériser une source binaire de symboles équiprobables. Au sens de la théorie de l'information, et à condition que tous les symboles de la séquence soient équiprobables, chaque symbole vaut l'équivalent de 2 bits. Le débit est en définitive de 16 [b] .

Texte

On peut voir le texte comme une série de lettres appartenant à un alphabet spécifique. Dès lors qu'il existe une représentation numérique de ces lettres, on se ramène au cas de signaux numériques, étant entendu que plusieurs symboles binaires sont nécessaires pour désigner une lettre.

2.1.2 Représentations et catégories de signaux

Pour l'étude de systèmes de communication, on peut distinguer plusieurs classes de signaux. Ces classes établissent les distinctions suivantes entre signaux, détaillées plus loin :

- analogiques ou numériques,
- périodiques ou apériodiques,
- déterministes ou stochastiques,
- d'énergie ou de puissance.

Signaux analogiques ou numériques

Un signal $x(t)$ analogique est une fonction continue pour tout temps t . Un signal numérique est un signal temporel discontinu; on le notera $x[n]$ où n est l'indice d'un élément pris dans l'ensemble d'instant $\{t_0, t_1, \dots\}$. On parle encore de signaux à temps discret.

La nature de l'information, analogique ou numérique, est intrinsèquement distincte de sa représentation dans un canal de transmission. Ainsi, un signal numérique peut très bien être

représenté par une onde continue en vue de la transmission. La figure 2.2 aide à clarifier la distinction entre signal d'information et représentation.

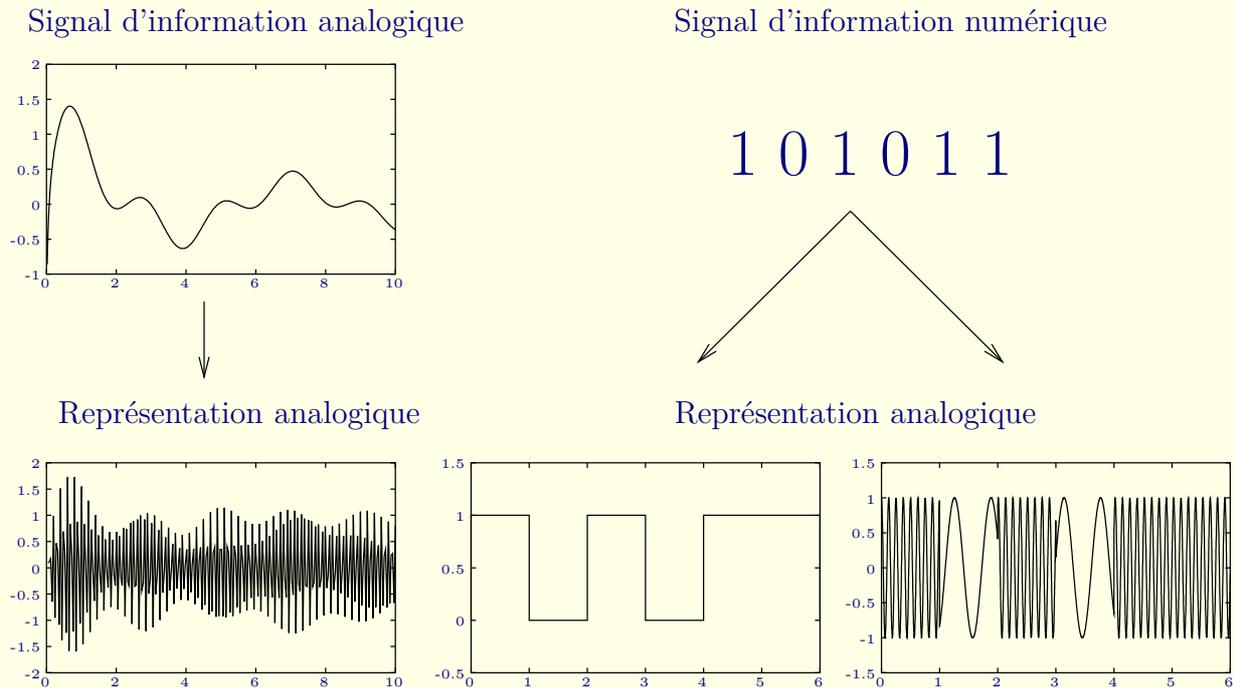


FIGURE 2.2 – Représentation d'un signal analogique ou numérique.

Signaux périodiques ou apériodiques

Un signal $x(t)$ est périodique s'il satisfait la relation suivante

$$x(t) = x(t + T_0) \quad \forall t \quad (2.1)$$

où t est la variable de temps et T_0 une constante. La plus petite valeur T_0 pour laquelle cette relation est vérifiée est appelée période fondamentale de $x(t)$. Un intervalle de temps d'une durée T_0 couvre donc un cycle complet du signal $x(t)$. S'il n'existe pas de constante pour laquelle la relation 2.1 est respectée, on dit que le signal $x(t)$ est apériodique ou non-périodique.

Signaux déterministes ou stochastiques

Un signal déterministe a une évolution connue et prévisible, contrairement aux signaux aléatoires ou stochastiques. Si un signal source est en grande partie déterministe à l'émetteur, le bruit qui l'affecte durant la transmission est inconnu. Le tableau 2.1 reprend les caractéristiques des signaux à l'émetteur et au récepteur.

	Émetteur	Récepteur
Signal utile	déterministe	aléatoire
Bruit et interférences	aléatoire	aléatoire

TABLE 2.1 – Nature des signaux dans une chaîne de télécommunications.

Au vu de la nature des signaux, l'analyse des systèmes de télécommunications nécessitera le recours à des outils stochastiques au moment d'établir les performances. Il en va de même pour l'utilisation de signaux numériques pour lesquels les performances s'exprimeront par des probabilités d'erreur durant la transmission.

Signaux d'énergie ou de puissance

Tout au long de la chaîne de télécommunications, on traite des signaux électriques caractérisés par une tension ou un courant. Soit une tension $v(t)$ qui, à travers une résistance R , produit un courant $i(t)$. La puissance instantanée dissipée dans cette résistance est définie par

$$p(t) = \frac{|v(t)|^2}{R} \quad (2.2)$$

ou encore

$$p(t) = R|i(t)|^2 \quad (2.3)$$

Quelle qu'en soit l'expression, la puissance instantanée est une fonction quadratique du signal caractéristique. À travers une charge unitaire de 1 *Ohm*, noté $[\Omega]$, les expressions sont même égales si bien qu'en définitive, il est de coutume de normaliser l'expression pour une résistance de 1 $[\Omega]$. Pour un signal de tension ou de courant, on obtient alors

Définition 5 [*Puissance instantanée normalisée*]

$$p(t) = |x(t)|^2 \quad (2.4)$$

Définition 6 [*Énergie*] Sur base de cette convention, l'énergie totale du signal $x(t)$ est définie par

$$E = \lim_{T \rightarrow +\infty} \int_{-T}^T |x(t)|^2 dt \quad (2.5)$$

$$= \int_{-\infty}^{+\infty} |x(t)|^2 dt \quad (2.6)$$

Certains signaux possèdent une énergie infinie. On utilise alors la notion de puissance moyenne qui est la moyenne temporelle de l'énergie.

Définition 7 [*Puissance moyenne d'un signal*] Il en découle une puissance moyenne du signal $x(t)$ s'exprimant

$$P = \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^{+T} |x(t)|^2 dt \quad (2.7)$$

Dans le cas d'un signal périodique de période T_0 , l'expression de la puissance moyenne devient

$$P = \frac{1}{T_0} \int_t^{t+T_0} |x(t)|^2 dt \quad (2.8)$$

Les définitions d'énergie et de puissance amènent à distinguer deux types de signaux :

- les signaux à énergie finie, pour lesquels $0 < E < +\infty$. Un signal physiquement réalisable est à énergie finie.

– les signaux à puissance finie. Dans ce cas, la puissance moyenne est bornée, à savoir $0 < P < +\infty$.

Ces deux contraintes sont mutuellement exclusives. En particulier, un signal à énergie finie a une puissance moyenne nulle alors qu'un signal à puissance finie possède une énergie infinie.

Les signaux déterministes et apériodiques sont à énergie finie alors que les signaux périodiques ou aléatoires ont généralement une puissance finie non nulle. Signalons qu'il s'agit de modélisation et qu'en conséquence, certains signaux n'ont pas de réelle signification physique pour des temps infinis, ce qui n'empêche pas qu'ils puissent être d'une grande aide!

Plus qu'une description plutôt qualitative, il est possible de classer la majorité des signaux utilisables en fonction d'une grandeur du second ordre S_x , homogène soit à une énergie E_x , soit à une puissance P_x . Le tableau 2.2 reprend cette caractérisation. Dans la dernière ligne du tableau 2.2, l'indice c signifie que l'on a centré les signaux.

$S_x < +\infty$	Temps continu	Temps discret
Énergie finie $E_x < +\infty$	$E_x \triangleq \int_{-\infty}^{+\infty} x(t) ^2 dt$	$E_x \triangleq \sum_{n=-\infty}^{+\infty} x[n] ^2$
Puissance moyenne finie, signaux périodiques $P_x < +\infty$	$P_x \triangleq \frac{1}{T_0} \int_0^{T_0} x(t) ^2 dt$	$P_x \triangleq \frac{1}{N_0} \sum_{n=0}^{N_0-1} x[n] ^2$
Puissance moyenne finie, signaux apériodiques $P_x < +\infty$	$P_x \triangleq \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^{+T} x(t) ^2 dt$	$P_x \triangleq \lim_{N \rightarrow +\infty} \frac{1}{2N+1} \sum_{n=-N}^{+N} x[n] ^2$
Puissance moyenne finie, signaux aléatoires $P_x < +\infty$	$P_x \triangleq E \{ x_c(t, \omega) ^2\}$	$P_x \triangleq E \{ x_c[n, \omega] ^2\}$

TABLE 2.2 – Caractérisation des signaux en fonction d'une grandeur du second ordre.

Décibel

Pour les calculs de puissance, on utilise fréquemment une unité basée sur le logarithme. Cette unité est le décibel, noté $[dB]$. L'introduction de la notion de décibel est destinée à pouvoir décrire un signal de puissance en termes de décades, car les niveaux de puissance tout au long d'une chaîne de transmission varient dans des proportions considérables; c'est donc un changement d'échelle.

Pour un signal d'énergie ou de puissance x , la relation entre unité décimale et décibel est la suivante

$$x \leftrightarrow 10 \log_{10}(x) \quad (2.9)$$

L'usage des décibels peut aussi s'exprimer relativement à une puissance de référence. C'est ainsi qu'on définit le dBW et le dBm ¹ comme l'écart par rapport à, respectivement, 1 $[W]$ et 1 $[mW]$. La puissance P vaut, en $[dBm]$,

$$P [dBm] = 10 \log_{10} \frac{P [mW]}{1 [mW]} \quad (2.10)$$

Le tableau 2.3 reprend quelques valeurs essentielles.

1. En toute rigueur, on devrait noter $dBmW$.

$x [W]$	$10 \log_{10}(x) [dBW]$
1 [W]	0 [dBW]
2 [W]	3 [dBW]
0,5 [W]	-3 [dBW]
5 [W]	7 [dBW]
$10^n [W]$	$10n [dBW]$

TABLE 2.3 – Valeurs de puissance exprimées en décibels (dBW).

Un calcul simple montre que $50 [W]$ équivaut à $17 [dBW]$ ou à $47 [dBm]$.

Pour le calcul de transmission radio, on parle d'intensité de champ électrique en $[dB\mu V/m]$. Or, le volt ne représente pas une mesure de puissance, contrairement à l'unité du volt au carré. En définitive, pour une tension U exprimée en $[V]$, les décibels s'expriment par

$$10 \log_{10} \left(\frac{U}{1 [V]} \right)^2 = 20 \log_{10} \frac{U}{1 [V]} \quad (2.11)$$

En conclusion, pour des grandeurs dont le carré représente une puissance,

$$x \leftrightarrow 20 \log_{10}(x) \quad (2.12)$$

On notera la présence d'un facteur 20 au lieu de 10 comme pour les puissances.

Exemple. Le confort d'écoute en radiodiffusion **FM** stéréo est défini par un seuil inférieur valant $1 [mV/m]$. En $[dB\mu V/m]$, ce seuil s'exprime comme suit

$$20 \log_{10} \frac{1 [mV/m]}{1 [\mu V/m]} = 20 \log_{10} \frac{1000 [\mu V/m]}{1 [\mu V/m]} = 60 [dB\mu V/m] \quad (2.13)$$

Rapport signal à bruit

Le rapport de la puissance du signal utile à celle du bruit, notées respectivement P_S et P_N , permet souvent de qualifier la qualité de la transmission.

Définition 8 [*Rapport signal à bruit ou Signal to Noise Ratio (SNR)*] *Le rapport signal à bruit, exprimé en décibel, vaut le quotient de puissance suivant*

$$\left(\frac{S}{N} \right) = 10 \log_{10} \left(\frac{P_s}{P_N} \right) [dB] \quad (2.14)$$

Il s'agit d'un critère abondamment utilisé pour la description des performances d'un système.

2.1.3 Représentation fréquentielle : la transformée de Fourier

Les signaux déterministes ont une évolution temporelle connue de leur valeur. On peut dès lors leur trouver un équivalent dans le domaine spectral par le biais de la transformée de FOURIER.

Soit un signal déterministe non périodique $x(t)$, la transformée de FOURIER de $x(t)$ est l'intégrale

$$\mathcal{X}(f) = \int_{-\infty}^{+\infty} x(t) e^{-2\pi jft} dt \quad (2.15)$$

où f est la variable exprimant la fréquence. La transformation inverse fournit le signal original ; elle est définie par

$$x(t) = \int_{-\infty}^{+\infty} \mathcal{X}(f) e^{2\pi jft} dt \quad (2.16)$$

La fréquence se mesure en Hertz, noté **Hz** ou $[Hz]$. Elle est liée à la pulsation angulaire ω par la relation

$$\omega = 2\pi f \quad (2.17)$$

La transformée de FOURIER est parfois définie en termes de pulsation angulaire plutôt qu'en terme de fréquence. Sans que cela ne modifie le principe de l'analyse spectrale, l'usage d'une autre variable introduit un coefficient correcteur dans certaines expressions. Notre choix est guidé par deux considérations : primo, l'usage de la fréquence introduit une symétrie intéressante dans la définition de la transformée et de son inverse, secundo, le contenu spectral des signaux et le comportement des systèmes sont exprimés majoritairement en terme de fréquences dans la littérature des télécommunications.

Il existe des conditions suffisantes d'existence de la transformée de FOURIER ; elles portent sur la nature de la fonction $x(t)$. Toutefois, la question de l'existence est un faux problème pour des signaux décrivant des phénomènes physiques réels. On peut ainsi montrer que tous les signaux à énergie finie, c'est-à-dire que $\int_{-\infty}^{+\infty} |x(t)|^2 dt < \infty$, possèdent une transformée de FOURIER.

Égalité de Parseval et théorème de Rayleigh

Soit $x(t)$ un signal, réel ou complexe, défini de $-\infty$ à $+\infty$. Supposons que sa transformée de FOURIER $\mathcal{X}(f)$ existe. L'énergie totale du signal est définie par l'expression habituelle

$$E = \int_{-\infty}^{+\infty} |x(t)|^2 dt \quad (2.18)$$

Cette énergie peut aussi trouver expression dans le domaine fréquentiel. En effet, le produit scalaire $|x(t)|^2$ est égal au produit de la fonction $x(t)$ par son complexe conjugué $x^*(t)$. Or, la transformée de FOURIER de cette fonction est égale à $\mathcal{X}^*(-f)$. À fréquence nulle, on établit alors le théorème suivant, dû à PARSEVAL, mais auquel on associe abusivement le nom de RAYLEIGH :

Théorème 9 [Rayleigh]

$$E = \int_{-\infty}^{+\infty} |x(t)|^2 dt = \int_{-\infty}^{+\infty} \|\mathcal{X}(f)\|^2 df \quad (2.19)$$

Toute fonction $\mathcal{X}(f)$ qui représente le carré de la transformée de FOURIER d'un signal fournit, après intégration de $-\infty$ à $+\infty$, l'énergie totale dudit signal, ce qui lui vaut parfois le nom de densité spectrale d'énergie. C'est en effet une mesure de la localisation de l'énergie dans le domaine fréquentiel.

2.2 Système

2.2.1 Systèmes linéaires permanents

L'analyse de FOURIER sert également d'outil pour l'étude des systèmes. Il est bien connu qu'un système n'altère un signal sinusoïdal mis à son entrée qu'en amplitude et en phase, sans ajout de nouvelles fréquences supplémentaires, s'il satisfait aux deux conditions suivantes :

1. Le système est linéaire. Autrement dit, si $y_1(t)$ et $y_2(t)$ sont les sorties des signaux respectifs $x_1(t)$ et $x_2(t)$, toute combinaison linéaire $ax_1(t) + bx_2(t)$ produit en sortie le signal $ay_1(t) + by_2(t)$.
2. Le système est invariant en temps ou permanent. Décaler l'entrée revient à décaler le signal de sortie d'une valeur de temps.

Un canal téléphonique est raisonnablement linéaire et permanent pour la durée de la conversation. Par contre, l'invariance est une hypothèse fautive pour des communications mobiles.

2.2.2 Réponse fréquentielle

Un système linéaire se caractérise par sa réponse impulsionnelle, c'est-à-dire pour une impulsion en entrée. Prenons, en toute généralité, un signal complexe à une fréquence précise

$$x(t) = e^{2\pi jft} \quad (2.20)$$

La réponse du système résulte du produit de convolution de sa réponse impulsionnelle $h(t)$ par la fonction d'entrée. Dès lors,

$$y(t) = \int_{-\infty}^{+\infty} h(\tau) e^{2\pi jf(t-\tau)} d\tau \quad (2.21)$$

$$= e^{2\pi jft} \int_{-\infty}^{+\infty} h(\tau) e^{-2\pi jf\tau} d\tau = x(t) \int_{-\infty}^{+\infty} h(\tau) e^{-2\pi jf\tau} d\tau \quad (2.22)$$

ce qui signifie que le signal de sortie est le produit du signal d'entrée par la transformée de FOURIER de la réponse impulsionnelle, évaluée à la fréquence de $x(t)$.

Définition 10 *Définissons la transmittance du système comme étant la transformée de FOURIER de sa réponse impulsionnelle*

$$\mathcal{H}(f) = \int_{-\infty}^{+\infty} h(t) e^{-2\pi jft} dt \quad (2.23)$$

On peut alors écrire que

$$y(t) = e^{2\pi jft} \mathcal{H}(f) \quad (2.24)$$

soit encore, en prenant la transformée de FOURIER des deux membres,

$$\mathcal{Y}(f) = \mathcal{X}(f) \mathcal{H}(f) \quad (2.25)$$

pour conclure que le rapport des transformées des signaux de sortie et d'entrée est égal à la transmittance du système.

2.2.3 Causalité et stabilité

Un système est causal si sa sortie est nulle en l'absence d'un signal d'entrée. Pour un système linéaire permanent, la réponse impulsionnelle $h(t)$ doit être nulle avant qu'un signal ne lui soit soumis. Imaginons que le signal d'entrée soit appliqué au temps $t = 0$. La condition nécessaire et suffisante de causalité est dès lors

$$h(t) = 0, \quad t < 0 \quad (2.26)$$

Un système temps réel réalisable doit être causal, ce qui n'est pas toujours le cas en général car on peut avoir mis en mémoire des valeurs prédéfinies.

La condition de stabilité usuelle est le critère de BIBO (Bounded Input-Bounded Output) : à une entrée bornée correspond toujours une sortie bornée. Soit le signal $x(t)$ borné par la valeur réelle positive finie M , $|x(t)| < M$.

$$|y(t)| \leq M \int_{-\infty}^{+\infty} |h(\tau)| d\tau \quad (2.27)$$

ce qui équivaut à imposer que

$$\int_{-\infty}^{+\infty} |h(t)| dt < \infty \quad (2.28)$$

2.2.4 Bande passante

Plusieurs théorèmes fondamentaux pour le domaine des télécommunications émettent l'hypothèse de signaux ou de canaux à **bande passante** limitée, c'est-à-dire que le contenu spectral est nul partout en dehors de cette bande. Une telle hypothèse est malheureusement irréaliste car elle impose une durée infinie dans le domaine temporel. Une hypothèse plus simple consiste à dire que le contenu spectral doit être majoritairement contenu dans la bande passante. On devine toute l'ambiguïté de cette démarche qui a conduit à plusieurs définitions de la notion de bande passante.

Toutes les définitions ont une caractéristique commune², à savoir qu'elles tentent d'établir une borne W pour tout spectre le long de l'axe des fréquences. Voici les critères de définition (cf. figure 2.3)

- (a) Bande passante à $3 [dB]$. Elle représente l'ensemble des fréquences à moins de $3 [dB]$ du maximum de la puissance.
- (b) Bande passante équivalente. Elle se définit comme la largeur d'une fonction rectangulaire telle que multipliée par la valeur de la puissance en son centre soit égale à la totalité de la puissance initiale.
- (c) Lobe principal. Les premiers passages à zéro autour de la fréquence porteuse définissent le lobe principal et donc un type de bande passante. Cette définition convient bien pour des signaux où l'on distingue effectivement un lobe principal.
- (d) Densité spectrale bornée. Il s'agit de l'ensemble des fréquences dont la puissance est comprise entre la valeur de la puissance au centre et une atténuation de par exemple 35 ou $50 [dB]$.

2. Remarquons que toutes ces définitions ne sont pas interchangeables.

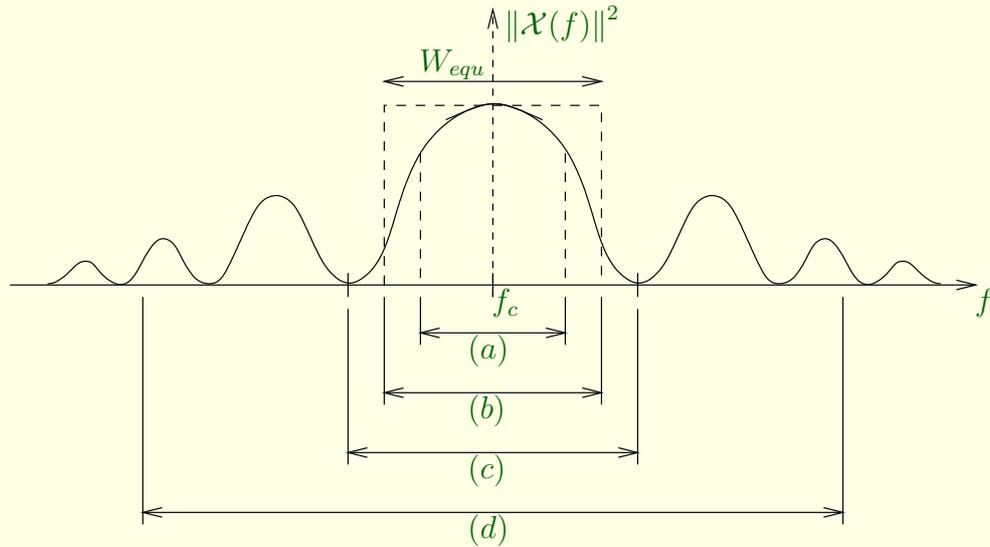


FIGURE 2.3 – Comparaison de définitions de bande passante.

2.2.5 Système de transmission idéal

Soit un signal à transmettre $x(t)$. Comme l'information réside habituellement dans la forme de ce signal, on admet que l'information n'est pas affectée si, au cours de la transmission, le signal ne subit, à travers le **canal**, que l'effet d'une multiplication de l'amplitude par une constante (amplification ou atténuation) ou l'effet d'un retard de transmission raisonnable. Ainsi, lors d'une communication téléphonique, le signal subira des amplifications et un retard de l'ordre de la dizaine de milli-secondes.

Définition 11 [*Système de transmission idéal*] Si le signal porteur d'information est représenté à l'entrée par une fonction $x(t)$, le système de transmission peut être considéré comme idéal s'il fournit en sortie un signal donné par $Kx(t - \tau)$.

Dans ce cas, K est un facteur d'amplification ou d'atténuation de l'amplitude et τ est un retard; ces deux grandeurs peuvent prendre des valeurs arbitraires mais raisonnables pour l'application considérée. L'information étant, par nature, imprévisible, ce comportement doit être valable non seulement pour le signal $x(t)$ mais aussi pour toute la classe des fonctions représentant les signaux possibles. La conclusion est immédiate : le système de transmission idéal doit être linéaire, permanent et posséder une transmittance de la forme

$$\mathcal{H}(f) = Ke^{-2\pi jf\tau} \quad (2.29)$$

On a donc la double condition suivante, outre que le système doit être linéaire et permanent :

1. Le module de la transmittance doit être constant.
2. Le déphasage doit être proportionnel à la fréquence, le coefficient (négatif) de proportionnalité étant le retard.

Il est clair que ces conditions ne doivent être satisfaites que sur la bande de fréquences occupée par l'ensemble des signaux possibles.

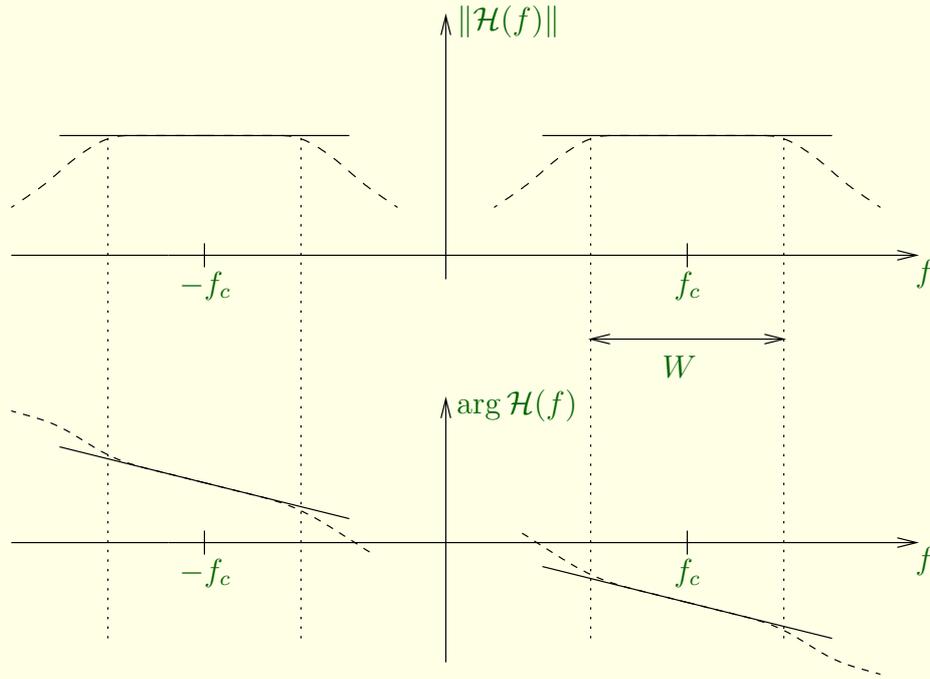


FIGURE 2.4 – Transmittance d'un système idéal.

Distorsions de linéarité

Si, le système de transmission étant linéaire et permanent, l'une ou l'autre des deux conditions précédentes n'est pas remplie, le signal est affecté d'une distorsion. Ces deux types de distorsion sont dites linéaires.

On peut donner de ces types de distorsion une interprétation mathématique. Si le module de la transmittance $\mathcal{H}(f)$ n'est pas constant, cela signifie que les différentes composantes du signal d'entrée sont transmises avec un coefficient d'amplification ou d'atténuation différents.

Si le déphasage

$$\phi(f) = \arg \mathcal{H}(f) \quad (2.30)$$

n'est pas proportionnel à la fréquence, cela signifie que les différentes composantes spectrales ont des temps de propagation inégaux à travers le système et arrivent donc en ordre dispersé; c'est pourquoi un système présentant de la distorsion de phase est dit dispersif.

Délai de groupe et de phase. Il est commode d'identifier la distorsion de phase d'un système linéaire par la dérivée de la phase. Plus précisément, on utilise la grandeur suivante

$$\tau_g = -\frac{1}{2\pi} \frac{d\phi(f)}{df} \quad (2.31)$$

appelée délai de groupe. Pour un système idéal,

$$\phi(f) = -2\pi f \tau_0 \quad (2.32)$$

ce qui fournit un délai de groupe

$$\tau_g = -\frac{1}{2\pi} (-2\pi \tau_0) \quad (2.33)$$

$$= \tau_0 \quad (2.34)$$

Le délai de groupe est donc constant pour un système idéal.

Deux composantes fréquentielles d'un groupe de fréquences subissent un délai fourni par τ_g . Un système à l'entrée duquel on injecte un signal mono-fréquentiel est toujours idéal. Le délai de cette composante unique, appelée délai de phase, vaut

$$\tau_p = -\frac{\phi(f)}{2\pi f} \quad (2.35)$$

Dans le cas d'un système idéal,

$$\tau_p = -\frac{(-2\pi f \tau_0)}{2\pi f} = \tau_0 \quad (2.36)$$

Autrement dit, les délais de groupe et de phase d'un tel système sont égaux.

Distorsions non linéaires

L'étude de systèmes non linéaires n'est guère aisée par manque d'outil adéquat. Nous l'aborderons par le biais d'un exemple. Considérons un système qui, à tout signal d'entrée $x(t)$, associe la sortie $y(t)$ suivante

$$y(t) = ax(t) + bx^2(t) \quad (2.37)$$

Soit le signal d'entrée suivant

$$x(t) = A_0 \cos(2\pi f_0 t) + A_1 \cos(2\pi f_1 t) \quad (2.38)$$

La sortie vaut

$$\begin{aligned} y(t) &= aA_0 \cos(2\pi f_0 t) + aA_1 \cos(2\pi f_1 t) \\ &+ bA_0^2 \cos^2(2\pi f_0 t) + bA_1^2 \cos^2(2\pi f_1 t) \\ &+ 2bA_0A_1 \cos(2\pi f_0 t) \cos(2\pi f_1 t) \end{aligned} \quad (2.39)$$

$$\begin{aligned} &= aA_0 \cos(2\pi f_0 t) + aA_1 \cos(2\pi f_1 t) \\ &+ \frac{bA_0^2}{2} (1 + \cos(4\pi f_0 t)) + \frac{bA_1^2}{2} (1 + \cos(4\pi f_1 t)) \\ &+ bA_0A_1 (\cos(2\pi(f_0 + f_1)t) + \cos(2\pi(f_0 - f_1)t)) \end{aligned} \quad (2.40)$$

La non linéarité se traduit par la présence en sortie de fréquences nouvelles. Ces signaux introduisent une distorsion harmonique, appelée *intermodulation*. C'est une règle générale de retrouver les fréquences doubles pour une relation entrée-sortie en $x^2(t)$, les fréquences triples pour $x^3(t)$, etc.

En pratique, les non linéarités sont très fréquentes ; l'utilisation de circuits actifs tels que des amplificateurs ou répéteurs, ayant des gains non linéaires ou pouvant saturer, entraîne des non linéarités. Ces distorsions sont difficilement modélisées et compensées mais, fort heureusement, elles constituent généralement un phénomène de faible ampleur.

2.2.6 Bruit et interférences

Le système ou canal de transmission peut encore être le siège d'autres perturbations que les effets de non-linéarité. On donne parfois à ces perturbations la dénomination générale de **bruit**. Il arrive que la transmittance d'un canal soit affectée par un facteur d'amplitude variant dans le temps; on parle alors d'un bruit multiplicatif. Dans un tel cas, il arrive fréquemment que ce facteur d'amplitude prenne, à certains moments, des valeurs très faibles : on parle alors d'évanouissements ou fading. Cet effet résulte généralement de l'existence dans le canal de transmission, de deux ou plusieurs chemins de propagation (multitrajets ou multipath) dont les délais de propagation varient dans le temps. À certains moments, les contributions de ces trajets se renforcent, tandis qu'à d'autres moments elles s'opposent, de sorte que le signal de sortie subit des évanouissements plus ou moins profonds. Le canal de transmission ne satisfait alors pas à la condition de linéarité et d'invariance temporelle.

On peut encore avoir affaire à un autre type de perturbation. Il s'agit de signaux d'origine tout à fait extérieure et indépendants du signal véhiculé par le canal de transmission, et qui viennent se superposer à celui-ci. On parle alors de bruit additif. En l'occurrence, il peut s'agir de

- bruits ou parasites naturels,
- autres signaux de télécommunications transmis dans la bande de fréquence du signal utile,
- produits de non-linéarité provenant d'autres signaux de télécommunications situés dans une autre bande de fréquence que le signal utile.

On voit que ceci fixe des limites sur la gamme dynamique (échelle d'amplitude) des signaux que l'on peut utiliser. On ne peut travailler avec des signaux trop faibles, parce qu'ils se retrouveraient noyés dans les bruits additifs. On ne peut pas non plus utiliser des signaux trop puissants sous peine de les distordre ou de perturber d'autres utilisateurs partageant ou non la même bande de fréquence.

Le système de transmission consiste souvent en une longue cascade de sous-systèmes. Il n'est pas interdit de procéder, dans ces sous-systèmes, à des opérations non linéaires ou non permanentes, pour autant que l'on procède ensuite aux opérations inverses. C'est en effet la totalité de la chaîne qui doit s'approcher autant que possible des conditions idéales.

Chapitre 3

Modulation d'onde continue

3.1 Introduction

On se trouve devant le problème de transmettre une information représentée par un signal $x(t)$, qui sera appelé signal modulant. Il peut s'agir d'un signal analogique ou numérique, avec au départ un contenu physique de type parole, texte ou vidéo. Mais nous apprendrons bien vite que les techniques de **modulation** ou de **multiplexage** permettent de construire des signaux qui mélangent plusieurs types d'information –ces signaux sont appelés signaux composites–, ou plusieurs signaux d'un même type. Le signal modulant peut dès lors être lui-même un signal composite, ce qui signifie qu'en amont on a déjà procédé, pour le construire, à des opérations de modulation ou de multiplexage.

Le signal modulant $x(t)$ est souvent représenté par une tension dont l'ensemble des valeurs possibles est soit pris dans un intervalle continu, soit qu'il se compose d'une série de valeurs discrètes, deux dans le cas binaire.

3.1.1 Hypothèses

Quel que soit le cas de figure, le signal modulant est supposé satisfaire aux deux exigences suivantes :

- Il est borné. D'où il existe une borne supérieure de l'amplitude du signal modulant

$$\forall t, |x(t)| \leq x_{\max} \quad (3.1)$$

À partir de cette hypothèse, il est commode de normaliser l'amplitude, en divisant la tension initiale par la valeur maximale, ce qui, somme toute, n'affecte pas l'information contenue dans le signal. Nous utiliserons donc un signal modulant normalisé $m(t)$ défini par l'expression

$$m(t) = \frac{x(t)}{x_{\max}} \quad (3.2)$$

Cela étant, on peut encore distinguer

- les signaux modulants non négatifs, comme les signaux vidéo ou les signaux binaires pour lesquels la tension est toujours positive.

- les signaux modulateurs alternatifs, pour lesquels $-1 \leq m(t) \leq 1$. Appartiennent à cette catégorie, les signaux sonores ou les composantes de différence de couleur d'un signal vidéo.

Le fait qu'un signal soit alternatif ne signifie pas pour autant qu'il ne contienne pas de composante continue. Pour s'en convaincre, il suffit de prendre un signal alternatif binaire pour lequel les deux états 0 et 1 ne seraient pas équiprobables. Mais comme la composante continue ne constitue pas un signal d'information intéressant et que sa présence entraîne un surcoût de puissance, on fait en sorte de la supprimer dans bon nombre d'applications. Les signaux sonores, les composantes de différence de couleur, et aussi la quasi-totalité des signaux modulateurs de nature composite sont dépourvus de composante continue.

- Le signal modulateur $m(t)$ est à spectre limité, c'est-à-dire qu'il existe une fréquence telle que la transformée de FOURIER du signal modulateur, supposée existante, satisfasse à la condition

$$\mathcal{M}(f) = 0 \quad \text{si } |f| > W \quad (3.3)$$

On ne peut jamais imposer strictement cette condition puisqu'elle revient à exiger que le signal soit à durée infinie. C'est néanmoins une hypothèse raisonnable car les signaux réels ont des composantes spectrales négligeables au-delà d'une certaine fréquence. Du reste, sous peine d'effets indésirables d'intermodulation ou de saturation, on est presque toujours amené à traiter les signaux modulateurs par un filtrage passe-bas limitant leur spectre en entrée.

Définition 12 Dès lors que l'intervalle de fréquences est borné par la fréquence W , on appelle bande de base l'intervalle de fréquences $[0, W]$.

Le fait que nous ayons à traiter des signaux réels n'est pas sans importance. En effet, leur transformée de FOURIER est telle que la partie réelle est paire et la partie imaginaire impaire; on dit alors que la transformée de FOURIER est paire, étant entendu qu'on considère le complexe conjugué. Cette propriété sera conservée tant qu'il s'agira d'un processus réalisable sur des signaux réels.

3.1.2 Modulation d'une porteuse sinusoïdale

Historiquement, le problème de la modulation s'est posé dès les débuts de l'ère de la radio; on doit à MARCONI la réalisation de premières transmissions radio. La "haute" fréquence était un signal à quelques kHz auquel on voulait imprimer des variations définies par un signal modulateur. Cette sinusoïde de référence¹ est appelée porteuse et notée $c(t)$

$$c(t) = A_c \cos(2\pi f_c t + \phi_c) \quad (3.4)$$

Pour superposer l'information, on peut agir sur différents paramètres de la porteuse : l'amplitude A_c , la fréquence f_c ou la phase ϕ_c . D'une manière générale, la modulation consiste à remplacer une de ces caractéristiques par une fonction linéaire de $m(t)$. Le signal résultant est appelé signal modulé (cf. figure 3.1). Si le signal modulateur $m(t)$ est continu, on parle de modulation d'onde continue. On pourrait attribuer des valeurs constantes à $m(t)$ pendant un certain intervalle de temps, ce qui reviendrait alors à effectuer une modulation numérique.

Les modulations de porteuses sinusoïdales sont de loin les plus utilisées. Cela est dû à l'importance des systèmes linéaires. En effet, un système linéaire n'introduit pas de fréquences

1. Par abus de langage, on parle de sinusoïde de référence alors qu'il s'agit d'une cosinusoïde.

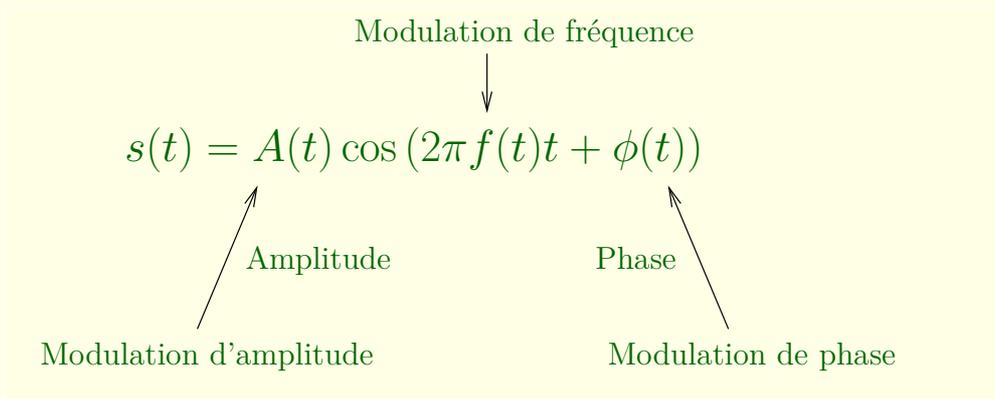


FIGURE 3.1 – Paramètres d'un signal modulé.

en dehors de la bande de fréquences originale. Remarquons toutefois que modifier, même linéairement, la fréquence ou la phase de la porteuse ne signifie pas pour autant que l'étage de modulation soit linéaire. Par ailleurs, la majorité des circuits de modulation ou de démodulation incluent des opérations non linéaires.

Il faut encore préciser qu'en plus de la contrainte d'un signal modulant à spectre limité, on impose que le contenu fréquentiel du signal modulant soit raisonnablement petit par rapport à la fréquence porteuse f_c , à savoir $W \ll f_c$. Il va de soi que, sans cette dernière contrainte, on aurait des phénomènes de battement qui résulteraient en une forte interaction entre le signal modulant et la porteuse.

3.2 Modulation d'amplitude

La modulation d'amplitude classique, dite A3 dans le domaine des radiocommunications, fut historiquement la première. Toutefois, en raison de sa faible efficacité, elle fut progressivement délaissée au profit d'autres techniques de modulation d'amplitude ou de modulations angulaires.

3.2.1 Modulation d'amplitude classique

Considérons la porteuse sinusoïdale $c(t)$ définie par

$$c(t) = A_c \cos(2\pi f_c t) \tag{3.5}$$

pour laquelle nous avons pris le choix d'une phase de référence arbitrairement nulle ($\phi_c = 0$). Soit $m(t)$ le signal en bande de base représentant le message, c'est-à-dire l'information que l'on désire transmettre. On suppose, a priori, que le système générant la porteuse $c(t)$ est totalement indépendant du message $m(t)$.

Définition 13 *La modulation d'amplitude, dite modulation **AM** pour Amplitude Modulation, est le processus par lequel l'amplitude de la porteuse $c(t)$ varie linéairement avec le signal modulant $m(t)$.*

Après modulation, le signal modulé $s(t)$ est décrit par la fonction

$$s(t) = A_c (1 + k_a m(t)) \cos(2\pi f_c t) \quad (3.6)$$

$$= A_c \cos(2\pi f_c t) + k_a A_c m(t) \cos(2\pi f_c t) \quad (3.7)$$

L'amplitude instantanée est donc rendue proportionnelle au signal modulant, et vaut

$$A(t) = A_c (1 + k_a m(t)) \quad (3.8)$$

L'amplitude instantanée $A(t)$ fluctue entre les limites $A_c(1 \pm k_a)$ pour un signal modulant alternatif. La figure 3.2 illustre le procédé et montre respectivement la porteuse, le signal modulant normalisé et le signal modulé.

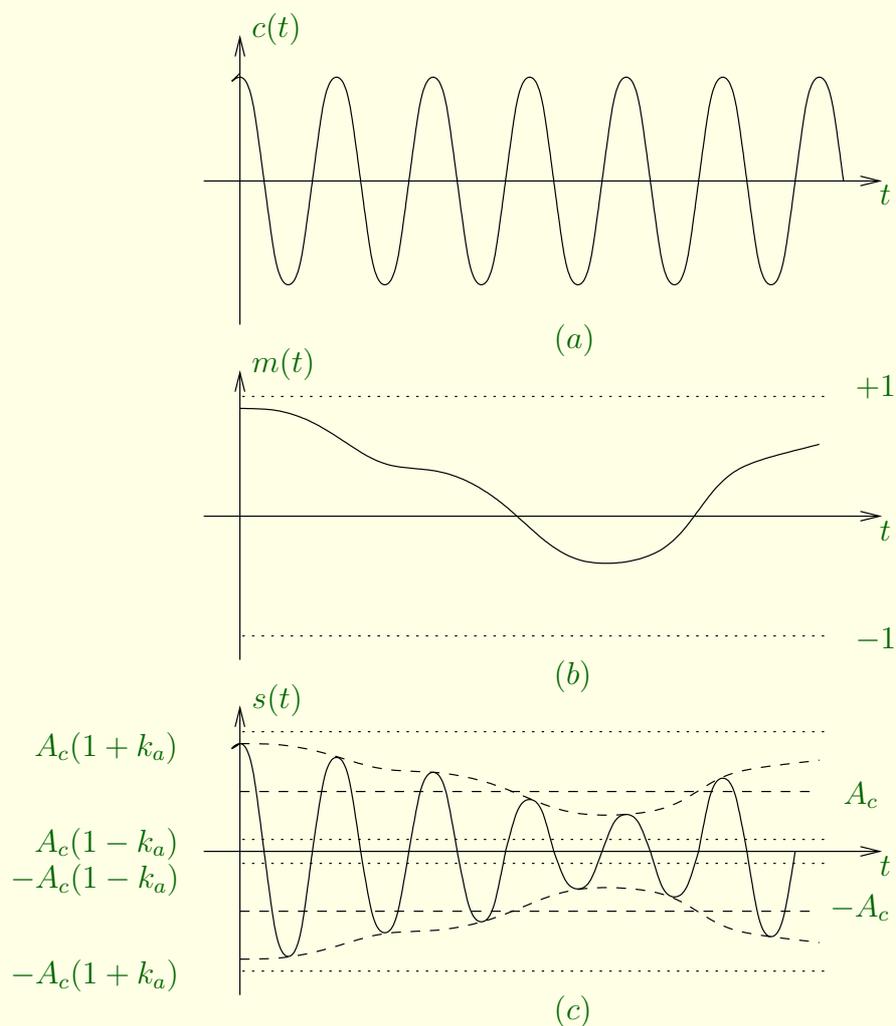


FIGURE 3.2 – Illustration de la modulation d'amplitude classique.

Dans la situation normale de la modulation d'amplitude, la fréquence porteuse est beaucoup plus grande que la fréquence maximale de la bande de base W ; la modulation correspond donc à une variation lente de l'amplitude instantanée. La constante k_a est appelée taux de modulation ; elle doit rester inférieure à l'unité sous peine d'engendrer un effet de surmodulation tel qu'illustré à la figure 3.3.

Surmodulation. L'exemple d'un fonctionnement normal (figure 3.2) suppose que l'amplitude de $k_a m(t)$ soit inférieure à l'unité, c'est-à-dire

$$\forall t, |k_a m(t)| < 1 \quad (3.9)$$

ce qui se traduit par une amplitude instantanée de $s(t)$ positive.

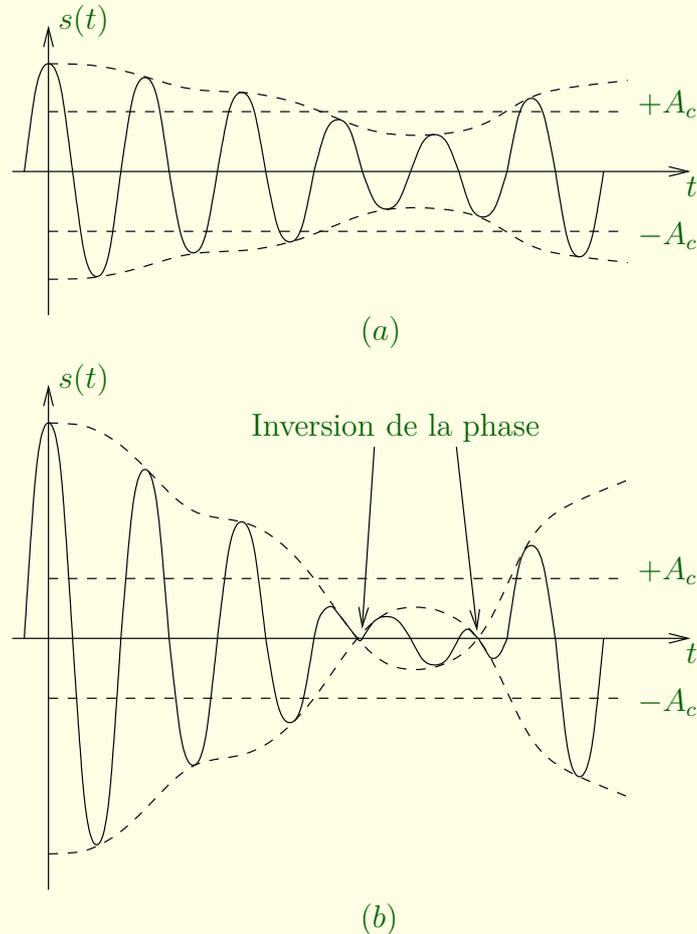


FIGURE 3.3 – Surmodulation : (a) $|k_a m(t)| < 1$, (b) $|k_a m(t)| > 1$.

La figure 3.3 illustre l'effet de surmodulation qui intervient lorsque le taux de modulation est trop grand, si bien que $|k_a m(t)| > 1$ pour certaines valeurs de t . On voit naître un phénomène d'inversion de la phase de $s(t)$ pour tout passage à zéro de la fonction $1 + k_a m(t)$. Le signal modulé $s(t)$ présente alors ce qu'on appelle de la distorsion d'enveloppe.

3.2.2 Analyse spectrale

Soient $m(t)$ un signal en bande de base et $\mathcal{M}(f)$ le spectre du signal, autre dénomination pour sa transformée de FOURIER. Selon les propriétés de la transformée de FOURIER, le spectre du signal modulé vaut

$$\mathcal{S}(f) = \frac{A_c}{2} [\delta(f - f_c) + \delta(f + f_c)] + \frac{k_a A_c}{2} [\mathcal{M}(f - f_c) + \mathcal{M}(f + f_c)] \quad (3.10)$$

Les signaux modulant et modulé étant réels, leur transformée de FOURIER est paire. Le spectre du signal modulé comprend deux raies de DIRAC flanquées de deux bandes latérales obtenues

par translation du spectre du signal modulant. Les bandes latérales correspondent bien évidemment à la transformée de FOURIER du terme $k_a A_c m(t) \cos(2\pi f_c t)$, c'est-à-dire au produit du signal modulant par la porteuse.

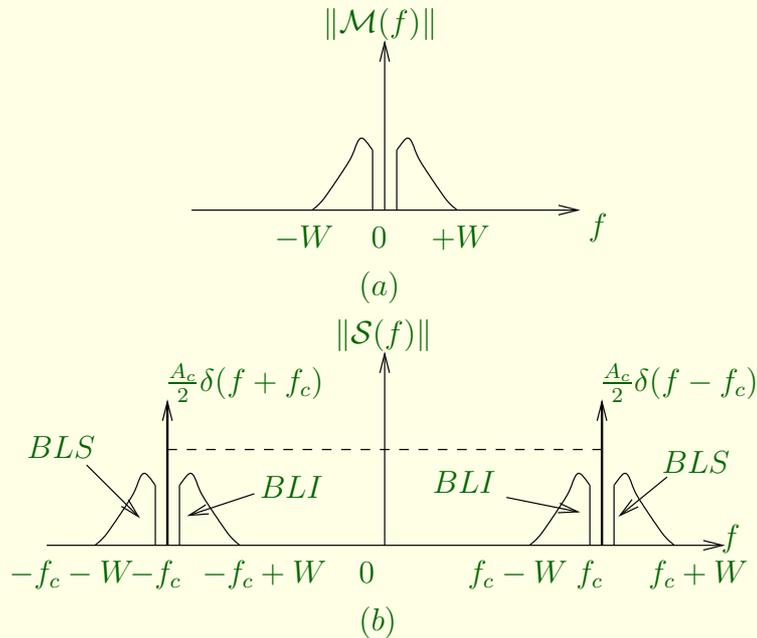


FIGURE 3.4 – Spectres de fréquence : (a) signal en bande de base, (b) signal modulé.

Soit W la bande de base du spectre du signal modulant $m(t)$. La figure 3.4 représente à la fois le spectre de $m(t)$ et le spectre du signal modulé en amplitude correspondant, où l'on a supposé que $f_c \gg W$. On remarque que la moitié du spectre du signal modulant $m(t)$, correspondant aux fréquences négatives (de $-W$ à 0), apparaît telle quelle dans le spectre du signal modulé $s(t)$. La bande passante requise pour le signal modulé est donc égale à deux fois la bande de base.

Il est clair que ce sont les bandes latérales qui portent l'information contenue dans le signal modulant. La raie porteuse, quant à elle, ne porte aucune information spécifique.

Pour les fréquences positives de $\mathcal{S}(f)$, la portion du spectre située au-dessus de la fréquence porteuse f_c est appelée bande latérale supérieure (BLS) et la portion du spectre située en-dessous de f_c est appelée bande latérale inférieure (BLI). Pour les fréquences négatives, la bande latérale supérieure est représentée par la portion du spectre située en-dessous de $-f_c$ et la bande latérale inférieure par la portion située au-dessus de $-f_c$. Il n'y a pas de recouvrement entre les bandes latérales à condition que $f_c > W$.

La radiodiffusion en grandes ondes (150-285 [kHz]), en ondes moyennes (525-1605 [kHz]) et en ondes courtes (jusqu'à 30 [MHz]) utilise la modulation d'amplitude A3. La bande de base est limitée à 4,5 [kHz] et la largeur d'un canal radio est donc de 9 [kHz].

3.2.3 Répartition de puissance

À un facteur de normalisation près, la puissance de la porteuse non modulée, qui est aussi la puissance de la raie porteuse dans le spectre du signal modulé, vaut $P_c = A_c^2/2$. Quant à la

puissance instantanée, par définition elle vaut

$$p(t) = s^2(t) \quad (3.11)$$

$$= A_c^2 (1 + k_a m(t))^2 \cos^2(2\pi f_c t) \quad (3.12)$$

$$= \frac{A_c^2}{2} (1 + k_a m(t))^2 (1 + \cos(4\pi f_c t)) \quad (3.13)$$

Dans la mesure où la fréquence porteuse f_c est beaucoup plus grande que la fréquence maximale de la bande de base W , la modulation d'amplitude est lente, et l'on peut parler d'une puissance instantanée du signal modulé calculée sur la durée d'une période de porteuse. En supposant constante l'amplitude la porteuse sur une période T_c . La puissance instantanée, moyennée sur une période, vaut

$$P_{T_c}(t) = \frac{1}{T_c} \int_t^{t+T_c} s^2(t') dt' \quad (3.14)$$

$$\approx \frac{A_c^2}{2} (1 + k_a m(t))^2 \frac{1}{T_c} \int_t^{t+T_c} (1 + \cos(4\pi f_c t')) dt' \quad (3.15)$$

$$\approx \frac{A_c^2}{2} (1 + k_a m(t))^2 \quad (3.16)$$

$$\approx P_c (1 + k_a m(t))^2 \quad (3.17)$$

Cette puissance fluctue donc entre $P_c(1 - k_a)^2$ pour les creux de modulation et $P_c(1 + k_a)^2$ pour les crêtes de modulation. On voit que la puissance instantanée peut atteindre quatre fois la puissance de la porteuse, pour une modulation à cent pour cent. Les étages de puissance de l'émetteur doivent donc être conçus pour supporter de telles puissances sans distordre l'enveloppe du signal.

Prise sur un temps très long, la puissance moyenne vaut

$$P = P_c \left(1 + 2k_a \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^{+T} m(t) dt + k_a^2 \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^{+T} m^2(t) dt \right) \quad (3.18)$$

$$= P_c \left(1 + 2k_a \overline{m} + k_a^2 \overline{m^2} \right) \quad (3.19)$$

où les barres indiquent qu'il s'agit d'une moyenne temporelle. Pour un signal modulant dépourvu de composante continue, l'expression de la puissance moyenne se réduit à

$$P = P_c \left(1 + k_a^2 \overline{m^2} \right) \quad (3.20)$$

On retrouve un premier terme de puissance relatif à la porteuse; le second terme exprime la puissance contenue dans les bandes latérales. Étant donné que la valeur de crête du signal modulant a été normalisée à l'unité, $P_c k_a^2 \overline{m^2} \leq P_c$. Pour la plupart des signaux modulateurs réels, elle est nettement inférieure à l'unité, et la puissance contenue dans les bandes latérales est beaucoup plus faible que celle de la puissance porteuse.

L'efficacité en terme d'utilisation de la puissance est définie par le rapport suivant

$$\eta = \frac{\text{puissance dans les bandes latérales}}{\text{puissance totale}} = \frac{k_a^2 \overline{m^2}}{1 + k_a^2 \overline{m^2}} \quad (3.21)$$

Exemple. Prenons un taux de modulation de 30% ($k_a = 0,3$), courant dans la pratique. Le signal $m(t)$ étant supposé centré et normalisé à 1, $\frac{k_a}{m^2} \leq 1$. La puissance contenue dans les bandes latérales représente alors une puissance relative maximale de 9% et $\eta = 0,083$

Comme ce sont les bandes latérales qui contiennent l'information utile et que, de plus, une seule bande latérale pourrait suffire, il est évident que la modulation d'amplitude classique est extrêmement inefficace dans l'utilisation qu'elle fait de la puissance. La modulation d'amplitude classique ne trouve donc de justification qu'historique, et au niveau de la très grande simplicité du démodulateur ; ce dernier consiste en un simple détecteur de crête. Aussi a-t-on inventé de nombreuses modulations d'amplitude dérivées de la modulation d'amplitude classique. Ces modulations suppriment la porteuse ou une bande latérale.

3.2.4 Réalisation de la modulation d'amplitude classique

Il existe de nombreux circuits de réalisation de la modulation d'amplitude classique. Il n'entre pas dans notre intention d'étudier les technologies, par ailleurs dépendantes de la fréquence, ou de détailler l'ensemble des schémas possibles. Seuls seront présentés les principes de réalisation avec quelques circuits en guise d'illustration. L'organe de base pour réaliser une modulation d'amplitude est le mélangeur ; il combine deux fréquences pour en créer la somme et la différence.

Deux approches sont possibles. Elles sont connues sous les noms de modulations linéaire ou quadratique.

Modulation linéaire

Le principe de la modulation linéaire est illustré à la figure 3.5. La porteuse est amplifiée dans un dispositif électronique dont le gain $g(t)$ est réglable et est une fonction linéaire du signal modulant.

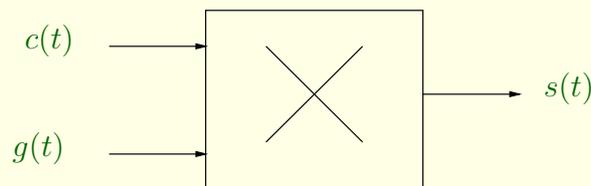


FIGURE 3.5 – Principe de la modulation linéaire.

Le multiplieur est l'élément conceptuellement le plus immédiat pour implémenter la modulation d'amplitude. Comme son nom l'indique, le multiplieur effectue le produit des signaux fournis à l'entrée.

Il existe de nombreux types de multiplieurs analogiques sous forme intégrée, qui peuvent multiplier des tensions alternatives (multiplieur quatre quadrants) ou seulement des tensions positives (multiplieur deux quadrants). Le multiplieur est un premier type de mélangeur. La sortie d'un multiplieur peut s'écrire

$$s(t) = kc(t)g(t) + k'c(t) + k''g(t) + k''' \quad (3.22)$$

Après compensation des distorsions et décalages éventuels, on obtient la relation linéaire d'un multiplieur parfait

$$s(t) = kc(t)g(t) \quad (3.23)$$

La dénomination de modulation linéaire vient de ce que la loi régissant le gain du dispositif en fonction du signal modulant doit être absolument linéaire ; en effet, une loi non-linéaire reviendrait à causer de la distorsion non-linéaire sur le signal d'information, ce qui est généralement inacceptable.

Modulation quadratique

Le principe de la modulation quadratique est le suivant : la porteuse $c(t) = A_c \cos(2\pi f_c t)$ et le signal modulant $x(t) = x_{\max} m(t)$ sont d'abord additionnés, et le résultat est appliqué à un organe non-linéaire. Un filtre passe-bande est utilisé pour éliminer les composantes spectrales indésirables à la sortie de celui-ci (figure 3.6).

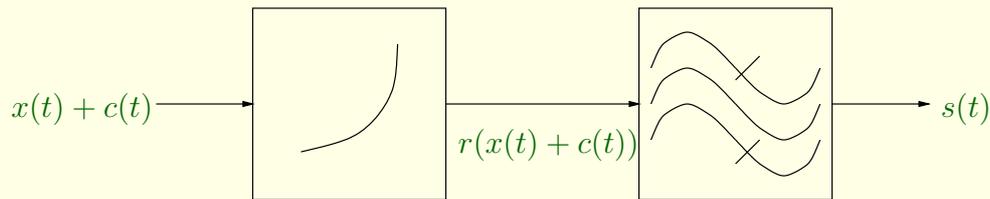


FIGURE 3.6 – Principe de la modulation quadratique.

Supposons que la fonction de réponse de l'organe non-linéaire soit $r(t)$ et possède un développement en série de TAYLOR. On peut alors écrire

$$r(x(t) + c(t)) = r_0 + r_1(x(t) + c(t)) + r_2(x(t) + c(t))^2 + r_3(x(t) + c(t))^3 + \dots \quad (3.24)$$

soit après réarrangement des premiers termes

$$r(x(t) + c(t)) = r_0 + r_1 x(t) + r_1 \left(1 + \frac{2r_2}{r_1} x_{\max} m(t)\right) c(t) + r_2 x^2(t) + r_2 c^2(t) + \dots \quad (3.25)$$

Dans les conditions normales, f_c est beaucoup plus grande que la bande de base, et les termes de la somme sont très nettement disjoints sur l'axe des fréquences. Le troisième terme donne les composantes spectrales correspondant à la modulation d'amplitude classique et peut donc être aisément sélectionné par le filtre passe-bande représenté à la figure 3.6. Le terme donnant les bandes latérales provient du terme quadratique dans le développement en série de TAYLOR de la non-linéarité. D'où le nom de modulation quadratique. Les termes non écrits donnent lieu à des composantes spectrales à $a f_c \pm b W$ qui peuvent toutes être éliminées par filtrage, à une exception près : il s'agit des termes à $f_c \pm b W$ qui recouvrent la bande utile. Ces termes correspondent à une distorsion affectant l'amplitude instantanée du signal résultant, c'est-à-dire à une distorsion non-linéaire sur le signal modulant. Cela est donc hautement indésirable si ce signal est du type analogique, et acceptable dans certaines proportions pour des signaux numériques. Une analyse plus fine montre que les composantes indésirables proviennent des termes d'ordre impair et supérieur ou égal à trois. La modulation quadratique exige donc que ces termes soient absents du développement ; idéalement la caractéristique de la non-linéarité ne devrait comporter que le terme linéaire et quadratique. La suppression des termes d'ordre impair s'obtient généralement par une architecture symétrique et balancée du modulateur non-linéaire.

Contrôle de la modulation

Pour contrôler la linéarité de la caractéristique de modulation et l'absence de surmodulation, on peut appliquer le signal modulé $s(t)$ à l'entrée Y d'un oscilloscope, tandis que la tension modulante $x(t)$ est appliquée en X. On obtient un figure trapézoïdale (figure 3.7) dont les côtés donnent la caractéristique de modulation. Cette méthode s'applique tout aussi bien à la modulation à porteuse supprimée.

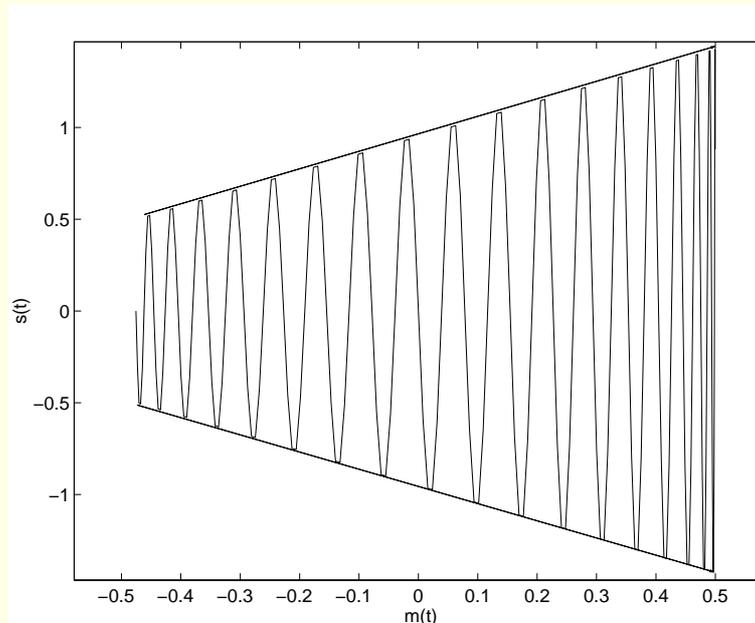


FIGURE 3.7 – Trapèze de modulation.

3.2.5 Démodulateurs

De même que l'on pouvait mettre en œuvre deux principes (linéaire et quadratique) pour réaliser la modulation, il existe plusieurs manières de faire la démodulation, aussi appelée détection depuis les débuts de la radio ; à l'époque, on travaillait en MORSE (A1) et démoduler le signal revenait à détecter la présence ou l'absence de la porteuse.

Détection d'enveloppe

La détection d'enveloppe ou détection de crête est le principe de réalisation de démodulateur le plus courant et le plus simple pour les signaux modulés en amplitude avec porteuse. Son schéma est donné à la figure 3.8.

Le détecteur d'enveloppe est composé d'une diode et d'un filtre RC. Son fonctionnement est le suivant. Lors du demi-cycle positif du signal d'entrée $s(t)$, la diode est polarisée en direct et la capacité C se charge rapidement à la valeur de crête du signal d'entrée. Quand le signal d'entrée chute en-dessous de cette valeur, la diode est polarisée en inverse et la capacité se décharge lentement dans la résistance R_i . La décharge se poursuit jusqu'au demi-cycle positif suivant du signal d'entrée. Quand celui-ci devient plus grand que la tension aux bornes de la capacité, la diode conduit à nouveau et le processus recommence.

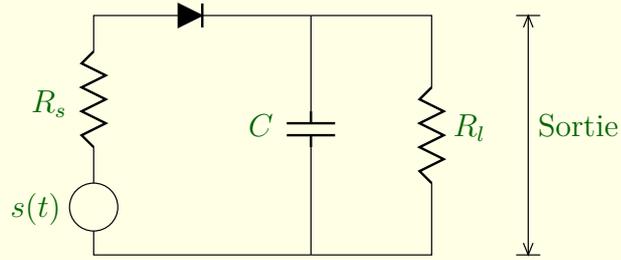


FIGURE 3.8 – Détecteur d'enveloppe.

Nous supposons que la diode est idéale, présentant une résistance r_f au passage du courant lorsqu'elle est polarisée en direct et une résistance infinie lorsqu'elle est polarisée en inverse. Nous supposerons également que le système générant le signal modulé $s(t)$ présente une impédance interne R_s . La constante de temps de charge (lors du demi-cycle positif du signal d'entrée) $(r_f + R_s)C$ doit être petite par rapport à la période de la porteuse $1/f_c$, c'est-à-dire

$$(r_f + R_s)C \ll \frac{1}{f_c}$$

de telle sorte que la capacité C se charge rapidement et suive, lorsque la diode conduit, la tension d'entrée lorsque celle-ci croît rapidement.

D'un autre côté, la constante de temps de décharge $R_l C$ (lorsque la diode est bloquée) doit être assez grande pour assurer une décharge lente de la capacité au travers de R_l entre deux demi-cycles positifs du signal d'entrée. Mais la tension aux bornes de la capacité doit suivre l'enveloppe du signal d'entrée même quand celle-ci diminue. Donc la constante de temps de décharge $R_l C$ doit être suffisamment petite pour pouvoir suivre la descente de l'enveloppe et donc de $m(t)$. En conclusion, nous devons avoir

$$\frac{1}{f_c} \ll R_l C \ll \frac{1}{W}$$

où W est la bande passante du message.

Démodulation AM synchrone ou cohérente

On peut retrouver le signal modulant à partir du signal modulé avec porteuse en multipliant $s(t)$ par une cosinusoïde identique en fréquence et en phase à celle de la porteuse. En effet,

$$s(t) \cos(2\pi f_c t) = \frac{A_c}{2} (1 + k_a m(t)) (1 + \cos(4\pi f_c t)) \quad (3.26)$$

$$= \frac{A_c}{2} (1 + k_a m(t)) + \frac{A_c}{2} (\cos(4\pi f_c t) + k_a m(t) \cos(4\pi f_c t)) \quad (3.27)$$

On peut donc réaliser la démodulation en faisant le produit du signal modulé par la porteuse, ce qui peut se faire avec le même type de circuit que la modulation, par exemple un modulateur en anneau dont les diodes seraient commandées par un générateur de porteuse local. Dans l'expression 3.27, le premier terme restitue en effet le signal modulant (avec une composante

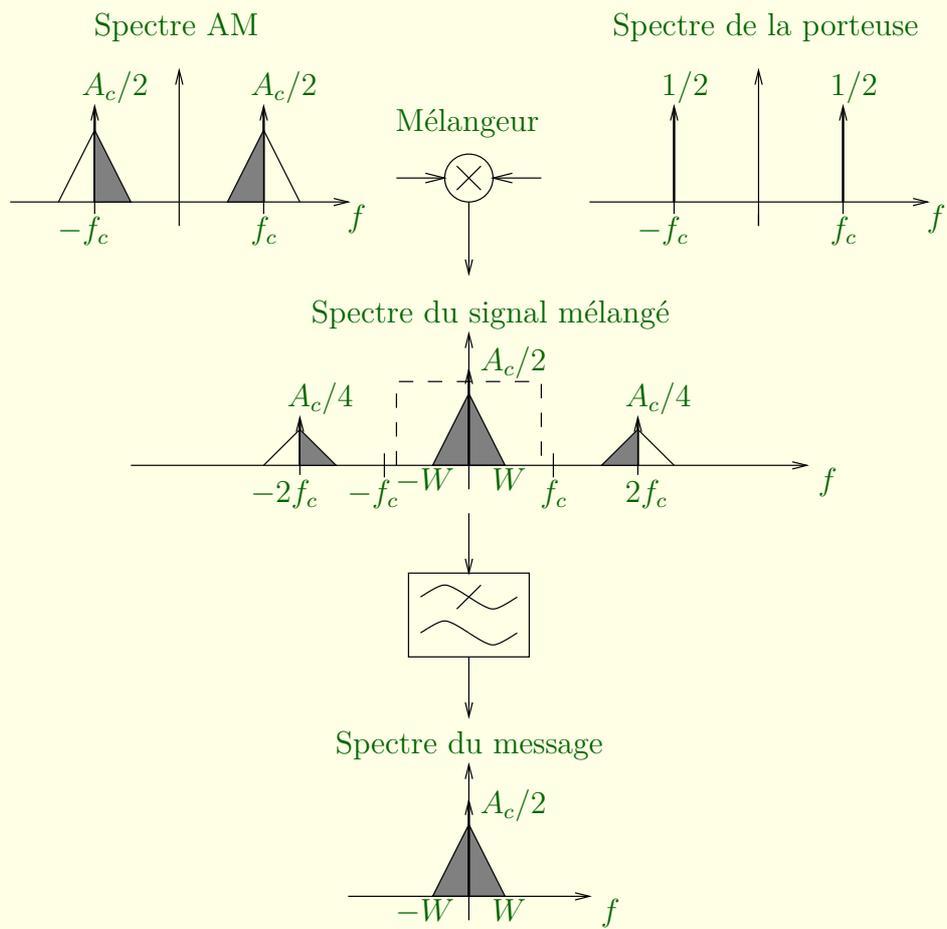


FIGURE 3.9 – Démodulation AM synchrone.

continue indésirable), tandis que le deuxième correspond à une raie flanquée de bandes latérales de part et d'autre de la fréquence double $2f_c$ et peut aisément être éliminé par un filtre passe-bas. Le schéma correspondant est montré à la figure 3.9.

Si le circuit de démodulation proprement dit est d'une simplicité élémentaire, il faut néanmoins résoudre le délicat problème de la création locale de la porteuse servant à la démodulation. Si l'on produit la porteuse locale à l'aide d'un oscillateur, il y a fort à craindre que ce signal n'ait pas exactement la même fréquence et la même phase que la porteuse utilisée pour le signal modulé. Pour clarifier la situation, supposons que cet oscillateur produise un signal $A' \cos(2\pi f_l t + \phi)$. Le produit de ce signal par le signal réceptionné, (on a choisi un signal modulé $s(t)$ à porteuse supprimée pour la simplicité des propos), vaut

$$s(t)A' \cos(2\pi f_l t + \phi) = \frac{A_c A'}{2} m(t) (\cos(2\pi(f_l - f_c)t + \phi) + \cos(2\pi(f_c + f_l)t + \phi)) \quad (3.28)$$

Tout écart entre la fréquence de l'oscillateur local et la fréquence de porteuse entraîne une dérive dans le calcul du signal démodulé. De plus, à supposer que les fréquences f_l et f_c concordent parfaitement, l'écart de phase joue l'effet d'un atténuateur car, après filtrage,

$$s(t)A' \cos(2\pi f_l t + \phi) \rightarrow \frac{A_c A'}{2} m(t) \cos \phi \quad (3.29)$$

Il faut impérativement un écart de phase petit pour une démodulation de qualité.

3.2.6 Avantages et limitations

La modulation d'amplitude est la plus ancienne technique de modulation, mais néanmoins ayant fait ses preuves. Son plus gros avantage est la simplicité de réalisation de la modulation et de la démodulation. Pour s'en convaincre, il suffit de regarder les circuits de modulation et de démodulation décrits ci-dessus. Il en résulte qu'un système de modulation d'amplitude est relativement bon marché, ce qui explique la raison pour laquelle la diffusion de programmes radio AM est populaire depuis si longtemps.

Cependant, la modulation d'amplitude souffre de deux limitations majeures :

1. La modulation d'amplitude implique un gaspillage de puissance. L'onde porteuse est totalement indépendante du message que l'on désire transmettre, c'est-à-dire le signal en bande de base $m(t)$. La transmission de la porteuse représente donc une perte de puissance. De plus, comme nous l'avons vu précédemment, la puissance de la porteuse représente plus de 2/3 de la puissance totale du signal modulé.
2. La modulation d'amplitude implique un gaspillage de bande passante. D'un point de vue fréquentiel, les bandes latérales supérieures et inférieures sont liées entre elles par une symétrie autour de la fréquence f_c . Connaissant une de ces deux bandes, on peut en déduire l'autre. Cela implique qu'il suffirait d'envoyer une de ces deux bandes pour obtenir toute l'information nécessaire à la reconstruction de $m(t)$. Cependant, la modulation AM utilise une bande de transmission égale à deux fois la bande passante du message, ce qui constitue un gaspillage de bande passante.

3.2.7 Modulations dérivées

Afin de surmonter les limitations de la modulation d'amplitude décrites à la section précédente, certaines modifications du processus de modulation et de démodulation doivent être réalisées, ce qui peut augmenter la complexité du système. En fait, un compromis a été recherché entre complexité du système et meilleure utilisation des ressources de communication. Partant de la modulation d'amplitude standard, nous pouvons distinguer quatre formes modifiées de la modulation d'amplitude :

1. Modulation à double bande latérale et porteuse supprimée (appelée en anglais “Double sideband-suppressed carrier” ou DSB-SC). Dans ce type de modulation, l'onde modulée est seulement constituée des bandes latérales supérieure et inférieure. On fait donc l'économie de la puissance de porteuse. Cependant, la bande de transmission est inchangée.
2. Modulation en quadrature de phase (appelée “Quadrature Amplitude Modulation” ou **QAM**). Elle consiste à transmettre un signal sur une porteuse sinusoïdale ainsi qu'un autre signal² sur une porteuse de même fréquence mais décalée de $\frac{\pi}{2}$. Sous sa forme usuelle, elle permet une économie de bande passante d'un facteur 2.
3. Modulation à bande unique (appelée en anglais “Single Sideband modulation” ou **SSB**). Ici, l'onde modulée est seulement constituée de la bande latérale supérieure ou inférieure. Ce type de modulation est particulièrement utilisée pour la transmission de signaux vocaux du fait qu'il existe un creux dans le spectre de la voix entre zéro et quelques centaines de [Hz]. Elle constitue de plus une forme optimale de modulation d'amplitude car elle minimise la puissance du signal modulé et la bande de transmission. Son principal désavantage est sa complexité.
4. Modulation à bande latérale résiduelle (appelée en anglais “Vestigial sideband modulation” ou **VS**B). Cette technique de modulation consiste à ne transmettre entièrement qu'une seule des deux bandes latérales tandis qu'une partie seulement de l'autre est transmise. La bande de transmission est alors réduite. Ce type de modulation est utilisé pour la radiodiffusion ou la transmission sur câble de signaux à large bande comme les signaux de télévision qui contiennent des composantes à des fréquences extrêmement basses.

3.2.8 Modulation d'amplitude à porteuse supprimée

Le signal modulé à porteuse supprimée est, comme l'indique son nom, la modulation d'amplitude classique amputée de la porteuse. Il s'exprime de la manière suivante

$$s(t) = m(t)c(t) \quad (3.30)$$

$$= A_c m(t) \cos(2\pi f_c t) \quad (3.31)$$

L'élimination de la porteuse se remarque également dans le domaine fréquentiel puisque le spectre de $s(t)$ vaut

$$\mathcal{S}(f) = \frac{A_c}{2} (\mathcal{M}(f - f_c) + \mathcal{M}(f + f_c)) \quad (3.32)$$

2. Pour renforcer la qualité de certaines transmissions numériques, il arrive qu'on transmette le même signal modulant sur les deux branches. De la sorte, on gagne en qualité sans que ne soit réalisée la moindre économie en termes de bande passante.

Réalisation de la modulation

La réalisation de la modulation ou de la démodulation d'un signal modulé en amplitude à porteuse supprimée est plus complexe que la modulation classique. L'observation du signal modulé (cf. figure 3.10) montre d'ailleurs que le signal croise l'axe temporel. Il n'est donc plus possible d'utiliser un détecteur de crête pour la démodulation.

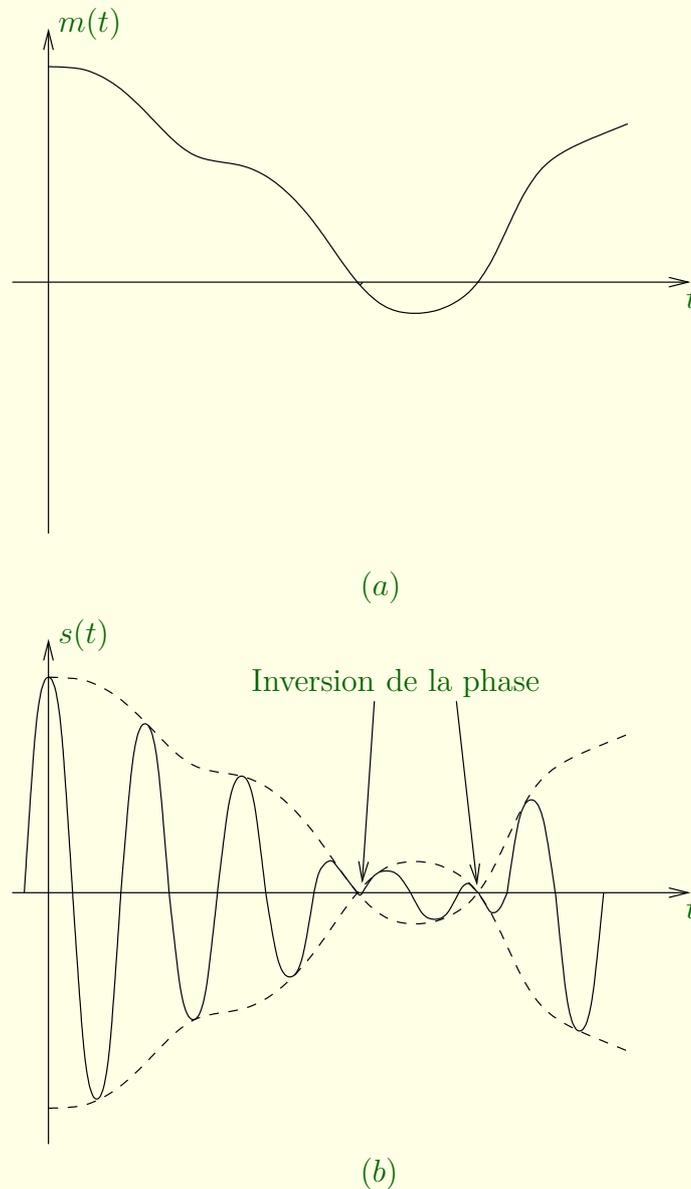


FIGURE 3.10 – Signal modulé par modulation à porteuse supprimée.

Pour la modulation, on fait par exemple usage d'un modulateur en anneau. Il s'agit d'un circuit à 4 diodes composé de deux transformateurs à point milieu.

Démodulation synchrone ou cohérente

Tout comme pour la modulation d'amplitude classique, on peut retrouver le signal modulant à partir du signal modulé à porteuse supprimée en multipliant $s(t)$ par une cosinusoïde identique en fréquence et en phase à celle de la porteuse. En effet,

$$s(t) \cos(2\pi f_c t) = \frac{A_c}{2} m(t) (1 + \cos(4\pi f_c t)) \quad (3.33)$$

On notera l'absence de composante continue.

Boucle de Costas

L'acquisition du rythme de porteuse est une tâche délicate. C'est tant la phase que la fréquence qui posent problème car il ne faut pas oublier qu'en raison des imperfections des équipements, il peut y avoir des fluctuations à la réception et à l'émission. La boucle de COSTAS, dont le schéma est repris à la figure 3.11, utilise un mécanisme astucieux pour réguler l'acquisition du rythme. Elle génère localement le cosinus et le sinus de la porteuse grâce à un oscillateur commandé en tension (VCO), et elle se sert des produits de ces deux signaux par le signal réceptionné pour réguler la fréquence et la phase de la porteuse locale. Ce type de circuit est largement répandu de nos jours.

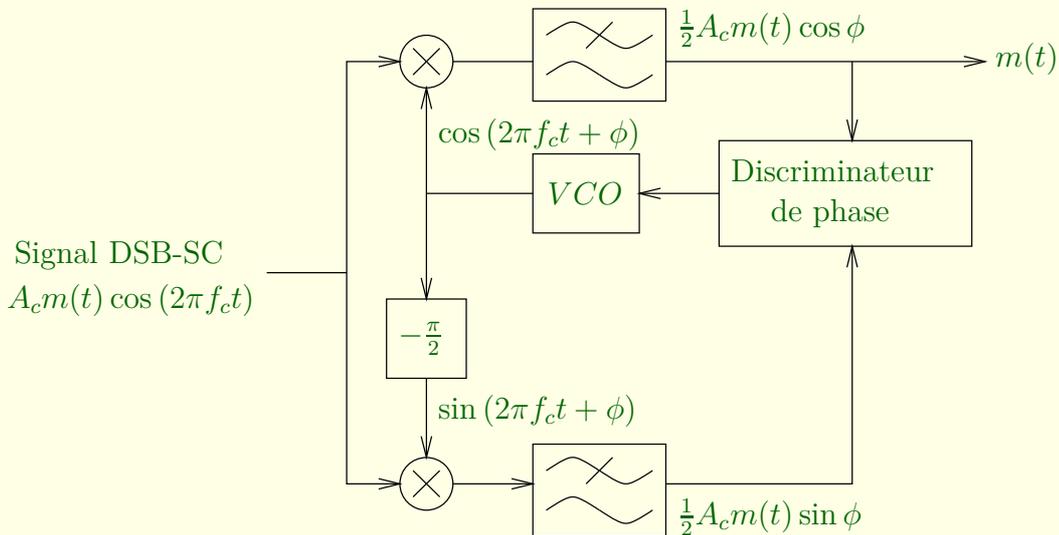


FIGURE 3.11 – Récepteur de COSTAS.

3.2.9 Modulation en quadrature

L'idée d'utiliser le cosinus et le sinus d'une porteuse peut également servir à l'émission. Soient deux signaux modulateurs ayant une même bande de base, $m_1(t)$ et $m_2(t)$, à transmettre. On construit le signal suivant, appelé signal modulé en quadrature,

$$s(t) = A_c m_1(t) \cos(2\pi f_c t) + A_c m_2(t) \sin(2\pi f_c t) \quad (3.34)$$

La figure 3.12 représente les schémas du modulateur et du démodulateur. On montre aisément, qu'avec une bonne acquisition de rythme, il est toujours possible de retrouver les deux signaux modulateurs. Il y a un réel intérêt à utiliser la modulation en quadrature. En effet, les deux signaux modulés occupent la même bande de fréquence, ce qui signifie une réduction de la bande passante d'un facteur deux.

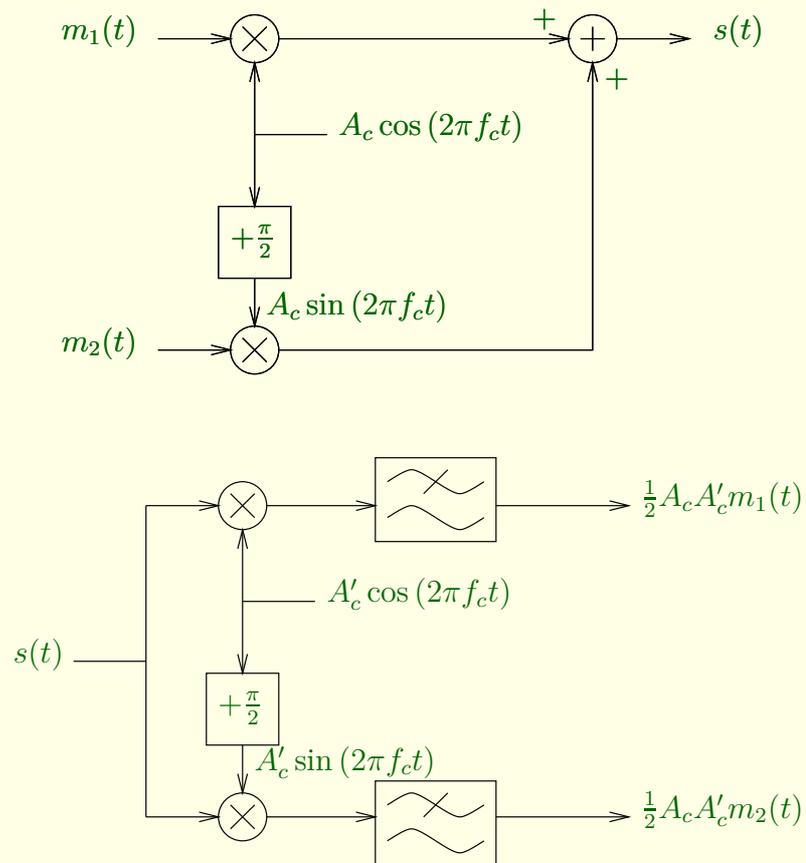


FIGURE 3.12 – Schéma de modulation et de démodulation d'une modulation d'amplitude en quadrature.

3.2.10 Modulation à bande latérale unique

Si la modulation en quadrature économise la moitié de la bande passante, elle a l'inconvénient de nécessiter la création des deux signaux, $m_1(t)$ et $m_2(t)$, à partir d'un signal à transmettre $m(t)$. Une alternative consiste à supprimer une bande latérale; on obtient alors une modulation à bande latérale unique (Single Sideband, **SSB**). Cette technique est abondamment utilisée pour la transmission de signaux de télévision, gros consommateurs de bande passante (un signal de qualité **PAL** nécessite de l'ordre de 5 à 8 [MHz]).

Filtrage d'une bande

Soit à filtrer le signal $u(t) = A_c m(t) \cos(2\pi f_c t)$ dont on ne désire garder qu'une bande. On fait passer le signal à travers un filtre de transmittance $\mathcal{H}(f)$. En sortie, le signal vaut

$$\mathcal{S}(f) = \mathcal{U}(f)\mathcal{H}(f) \quad (3.35)$$

$$= \frac{A_c}{2} [\mathcal{M}(f - f_c) + \mathcal{M}(f + f_c)]\mathcal{H}(f) \quad (3.36)$$

Pour déterminer l'allure du filtre, il faut assurer que la démodulation soit en mesure de reconstituer le signal de départ. Considérons la détection cohérente qui consiste en la multiplication par la porteuse $A'_c \cos(2\pi f_c t)$.

$$v(t) = A'_c \cos(2\pi f_c t) s(t) \quad (3.37)$$

Soit, dans le domaine de FOURIER,

$$\mathcal{V}(f) = \frac{A'_c}{2} [\mathcal{S}(f - f_c) + \mathcal{S}(f + f_c)] \quad (3.38)$$

Après substitution,

$$\begin{aligned} \mathcal{V}(f) &= \frac{A_c A'_c}{4} \mathcal{M}(f) [\mathcal{H}(f - f_c) + \mathcal{H}(f + f_c)] \\ &+ \frac{A_c A'_c}{4} [\mathcal{M}(f - 2f_c) \mathcal{H}(f - f_c) + \mathcal{M}(f + 2f_c) \mathcal{H}(f + f_c)] \end{aligned} \quad (3.39)$$

Le second terme est éliminé à la réception par simple filtrage. Il reste alors un signal en bande de base n'ayant subi aucune distorsion à la condition que, $\forall f$,

$$\mathcal{H}(f - f_c) + \mathcal{H}(f + f_c) = 1 \quad (3.40)$$

Autrement dit, le filtre est symétrique en amplitude autour de la fréquence f_c ; par ailleurs $\mathcal{H}(f_c) = 1/2$. Le filtre idéal le plus simple est le filtre passe-haut ou passe-bas parfait, au droit de f_c , mais l'expression 3.40 permet d'autres choix de filtres, conduisant à la modulation à bande latérale résiduelle.

Démodulation

Le schéma de démodulation cohérente s'applique également au cas d'un signal à bande latérale unique; il est représenté à la figure 3.13.

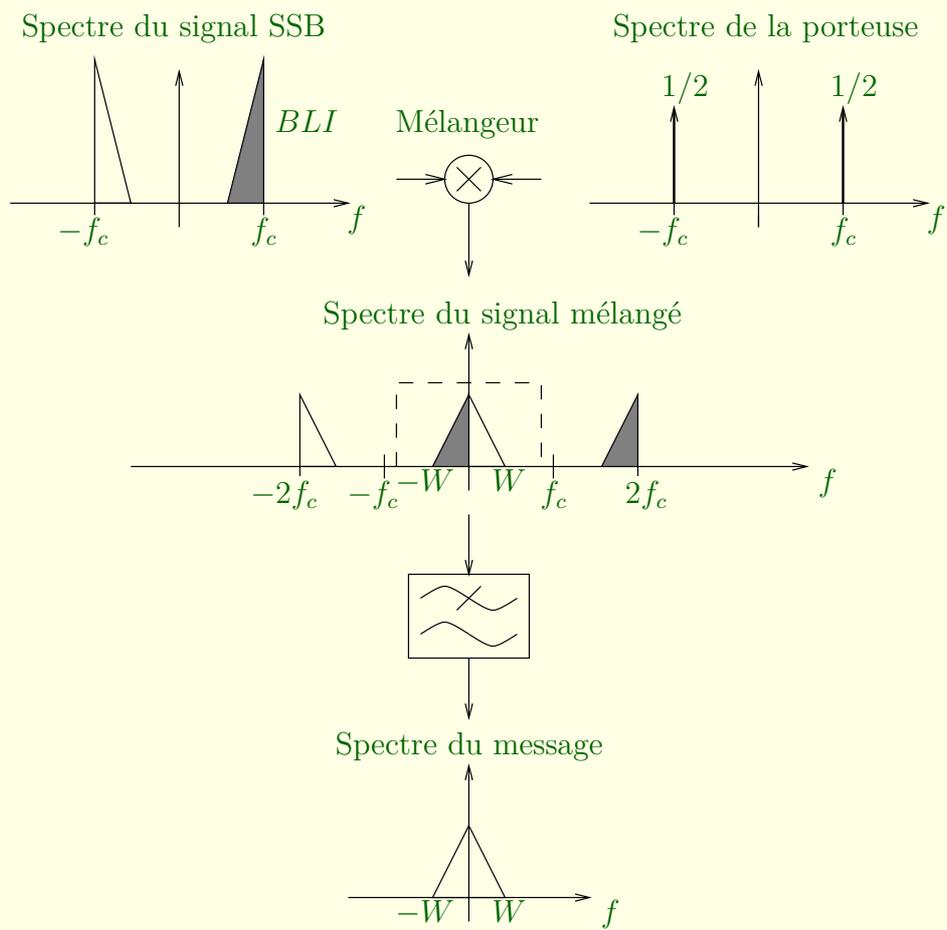


FIGURE 3.13 – Schéma de démodulation SSB.

3.2.11 Modulation à bande latérale résiduelle

Lorsque la bande de base s'étend jusqu'à la fréquence nulle, il n'est pas toujours possible de faire une modulation à bande latérale unique, au sens strict du terme. On peut toutefois obtenir une réduction sensible de la largeur de bande en éliminant partiellement une des deux bandes latérales. C'est la modulation à bande latérale résiduelle (Vestigial Sideband, **VSB**).

Le problème essentiel est de trouver une manière d'opérer le filtrage de telle façon que l'on puisse retrouver le signal modulant sans déformation. L'étude de ce problème a montré que le filtre d'élimination partielle d'une des deux bandes doit avoir une transmittance symétrique par rapport à la fréquence porteuse. En d'autres termes, à supposer que l'on veuille éliminer partiellement la bande latérale inférieure, la transmittance doit être nulle en-dessous d'une fréquence donnée, puis elle croît en passant par $\mathcal{H}(f_c) = 1/2$.

On montre qu'une démodulation cohérente restitue le signal modulant sans aucune déformation. La modulation VSB est utilisée en radiodiffusion télévisuelle, et également dans certaines normes relatives à des modems pour la transmission de données numériques dans l'infrastructure de téléphonie analogique.

3.3 Modulation angulaire

Dans la modulation d'amplitude, partant d'une porteuse $c(t) = A_c \cos(2\pi f_c t + \phi_c)$, on a rendu l'amplitude A_c fonction linéaire du signal modulant $m(t)$. Dans le cas de la modulation angulaire, on introduit une telle dépendance pour l'argument de la fonction cosinus. Cela est moins évident qu'il n'y paraît à première vue. Quel paramètre doit devenir une fonction linéaire du signal modulant? On peut en effet choisir de prendre l'argument de la fonction cosinus ou seulement la fréquence de la porteuse. Dans le premier cas, il s'agit de modulation de phase, tandis que dans le second cas on parlerait plutôt de modulation de fréquence.

Dans les deux cas, l'argument de la fonction cosinus n'est plus une fonction linéaire du temps. Mais dans la mesure où l'on définirait la phase comme l'argument de la fonction cosinus, une modulation de fréquence s'accompagnera nécessairement d'une modulation de phase. Comme on le devine, les concepts de modulation de phase et de fréquence sont strictement indissociables. C'est pourquoi on utilisera le terme général de modulation angulaire.

3.3.1 Principes et définitions

Dans la modulation angulaire, le signal modulé prend la forme

$$s(t) = A_c \cos \phi_i(t) \quad (3.41)$$

où $\phi_i(t)$, appelée phase instantanée du signal modulé, est une fonction du signal modulant. En l'absence de modulation, on aurait évidemment $\phi_i(t) = 2\pi f_c t + \phi_c$ où ϕ_c est la phase au temps $t = 0$. Remarquons que la modulation angulaire n'affecte pas l'amplitude la porteuse.

Définition 14 *La déviation instantanée de phase est la grandeur définie par*

$$\Delta\phi_i(t) = \phi_i(t) - (2\pi f_c t + \phi_c) \quad (3.42)$$

Définition 15 *L'amplitude de la déviation instantanée de phase*

$$\beta = \max |\Delta\phi_i(t)| \quad (3.43)$$

est appelée indice de modulation.

Elle joue un rôle important, comme nous le verrons ultérieurement. La déviation instantanée de phase peut être interprétée comme une variation de la fréquence.

Définition 16 *Par définition, la quantité*

$$f_i(t) = \frac{1}{2\pi} \frac{d\phi_i(t)}{dt} \quad (3.44)$$

est la fréquence instantanée.

Bien entendu, en l'absence de modulation angulaire, on retrouve la fréquence de la porteuse f_c . Ce n'est pas tant la fréquence qui importe mais la différence entre la fréquence instantanée et la fréquence porteuse. Tout comme pour la phase, on définit la déviation instantanée de fréquence.

Définition 17 *La déviation instantanée de fréquence $\Delta f_i(t)$ est l'écart entre la fréquence de la porteuse et la fréquence instantanée*

$$\Delta f_i(t) = f_i(t) - f_c \quad (3.45)$$

Définition 18 *Le maximum de la déviation instantanée de fréquence $\Delta f_i(t)$ fournit l'excursion de fréquence Δf définie par*

$$\Delta f = \max |\Delta f_i(t)| \quad (3.46)$$

Des définitions précises des divers types de modulation angulaire seront données plus loin. On peut déjà dire que la modulation angulaire consiste à faire varier, selon une loi linéaire bien précise, une des quantités $\Delta\phi_i(t)$ ou $\Delta f_i(t)$. Comme suite des définitions qui précèdent, il apparaît que l'on ne peut faire varier l'une sans l'autre ; une modulation de phase entraîne donc une modulation de fréquence et inversement.

Dans le jargon technique, on utilise aussi le terme de taux de modulation : il s'agit de la quantité $\Delta f / \Delta f_{\max}$, où Δf_{\max} est la valeur maximale de l'excursion de fréquence autorisée par les règlements régissant le partage des fréquences, par exemple par le Règlement des radiocommunications. Ainsi, en radiodiffusion sonore à modulation de fréquence, on impose $\Delta f_{\max} = 75 [kHz]$.

À condition de faire varier un paramètre de la phase comme une fonction linéaire du signal modulant, on obtient une modulation appelée modulation angulaire. Dans cette modulation, le signal modulant $m(t)$ est généralement un signal alternatif, tel qu'un signal sonore, oscillant entre -1 et $+1$. Pour l'instant, il s'agira d'un signal continu, auquel cas on parle de modulation angulaire analogique de type F3.

Parmi toutes les possibilités de modulation angulaire, on distingue la modulation de phase pure et la modulation de fréquence pure.

Modulation de phase pure

Définition 19 La modulation de phase (Phase Modulation, **PM**) consiste à faire varier la phase $\phi_i(t)$ en fonction du signal modulant³ $m(t)$, à savoir (on prend $\phi_c = 0$)

$$\phi_i(t) = 2\pi f_c t + k_p m(t) \quad (3.47)$$

Le terme $2\pi f_c t$ représente la phase de la porteuse en l'absence de modulation. La présence d'une tension modulante affecte cette phase en fonction d'un coefficient k_p qui représente la sensibilité du modulateur ; elle s'exprime en radians par volt. Le signal modulé vaut donc

$$s(t) = A_c \cos(2\pi f_c t + k_p m(t)) \quad (3.48)$$

La fréquence instantanée de la cosinusoïde est la dérivée de la phase divisée par 2π

$$f_i(t) = \frac{1}{2\pi} \frac{d\phi_i(t)}{dt} = f_c + \frac{k_p}{2\pi} \frac{dm(t)}{dt} \quad (3.49)$$

Il s'ensuit que la modulation de phase revient à modifier la fréquence de la porteuse. La déviation de fréquence instantanée vaut

$$\Delta f_i(t) = \frac{k_p}{2\pi} \frac{dm(t)}{dt} \quad (3.50)$$

Modulation de fréquence pure

Définition 20 Par définition de la modulation de fréquence (Frequency Modulation, **FM**), la déviation instantanée $f_i(t)$ est proportionnelle au signal modulant

$$f_i(t) = f_c + k_f m(t) \quad (3.51)$$

La fréquence résultante est donc liée, via la sensibilité du modulateur k_f exprimée en $[Hz/V]$, au signal modulant. La phase du signal modulé se calcule par l'intégrale de la fréquence instantanée (on prend $\phi_c = 0$)

$$\phi_i(t) = 2\pi f_c t + 2\pi k_f \int_0^t m(t') dt' \quad (3.52)$$

Dès lors le signal modulé vaut

$$s(t) = A_c \cos(2\pi f_c t + 2\pi k_f \int_0^t m(t') dt') \quad (3.53)$$

Les relations 3.49 et 3.52 mettent bien en évidence qu'une modulation de phase entraîne une modulation de fréquence, et vice versa. Plus précisément, on peut affirmer que la modulation de phase se réduit à une modulation de fréquence par le signal modulant préalablement dérivé. Inversement, une modulation de fréquence est une modulation de phase par l'intégrale du signal modulant. Ces schémas sont représentés à la figure 3.14.

3. Nous n'exigeons pas $|m(t)| \leq 1$ en modulation angulaire.

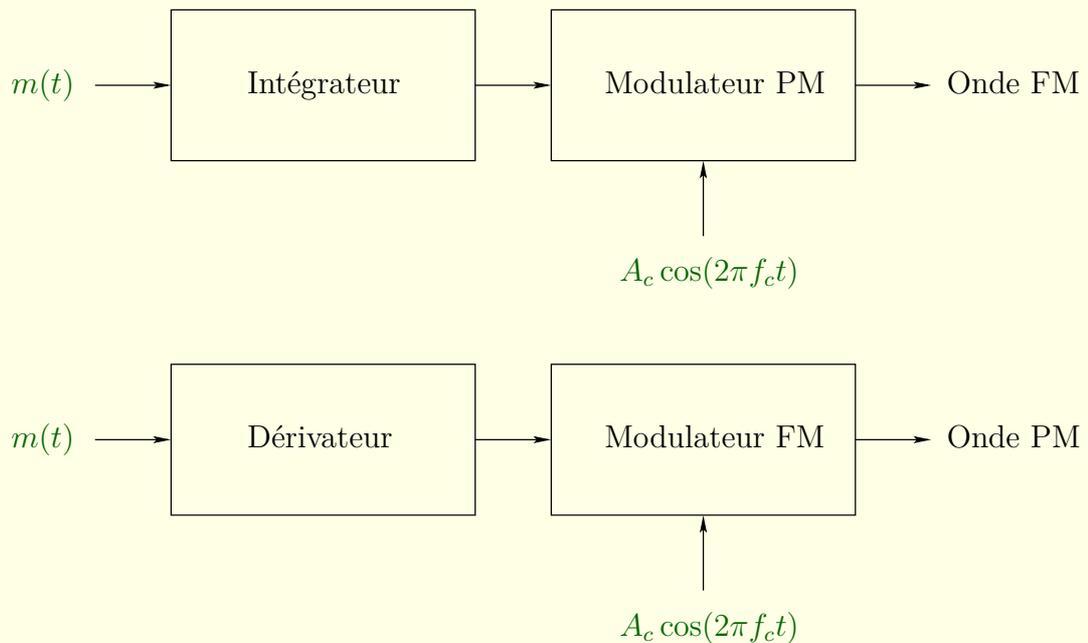


FIGURE 3.14 – Liens entre modulation de phase et modulation de fréquence.

Notion de préaccentuation

Les opérations d'intégrale et de dérivée établissent un lien entre deux types de modulation angulaire. Plus généralement, la modulation angulaire regroupe tous les cas où $\Delta\phi_i(t)$ et $\Delta f_i(t)$ sont des fonctions linéaires du temps. Pour la commodité, on parle de modulation de fréquence avec préaccentuation; ce vocable désigne une modulation de fréquence pure par un signal obtenu par passage du signal modulant proprement dit $m(t)$ au travers d'un système linéaire appelé filtre de préaccentuation. Si l'on note $\mathcal{H}(f)$ la transmittance de ce filtre, l'ensemble de la chaîne de transmission peut être vu comme une paire constituée d'un modulateur et d'un démodulateur insérée entre un filtre de préaccentuation et un filtre de désaccentuation de transmittance $1/\mathcal{H}(f)$.

La préaccentuation présente deux avantages : une meilleure résistance au bruit et une réalisation systématique à l'aide d'un modulateur en fréquence, techniquement plus aisée qu'une modulation de phase.

Illustration des techniques de modulation

En conséquence de la modulation de la phase, les passages par 0 de la fonction ne sont plus équidistants; par contre, l'enveloppe reste constante. La figure 3.15 montre un signal modulant original et les signaux modulés respectivement en amplitude, en phase et en fréquence.

3.3.2 Analyse de la modulation de fréquence analogique

Remarque préliminaire

L'analyse spectrale d'un signal modulé est évidemment très importante, car elle permet de déterminer la bande passante requise. Cette analyse était simple dans le cas de la modulation

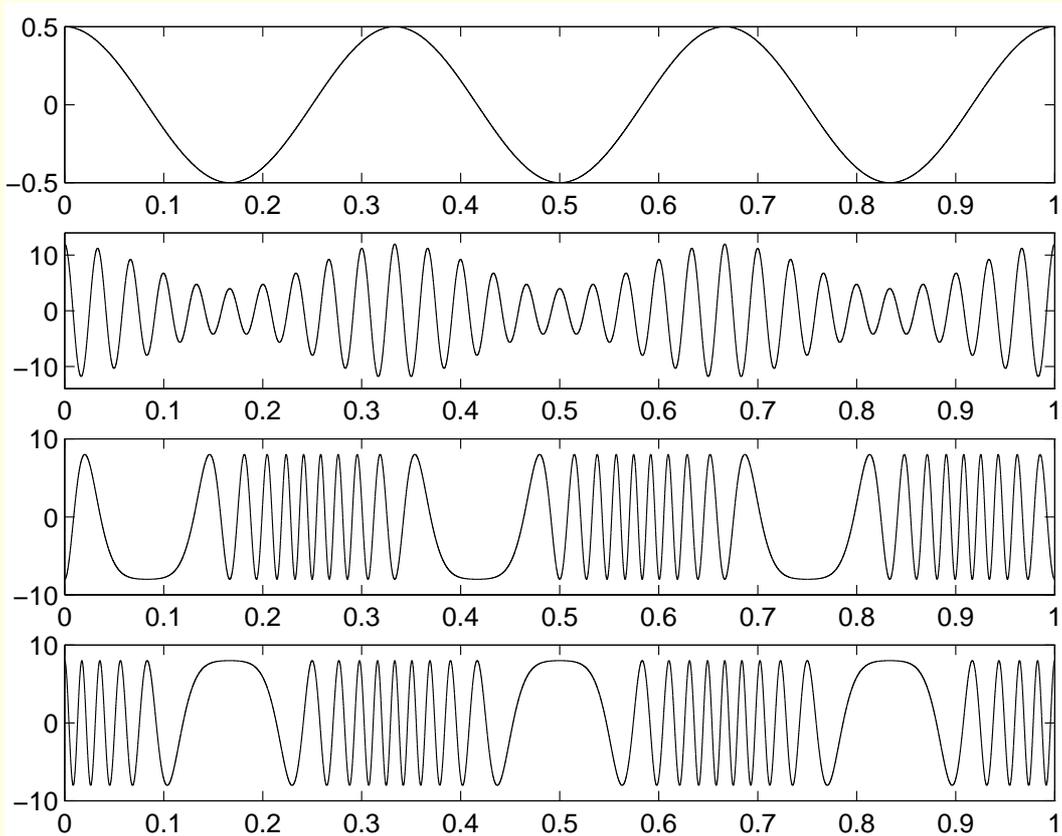


FIGURE 3.15 – Signal modulant et signaux modulés respectivement en AM, PM et FM.

d'amplitude en raison de la propriété de linéarité dont elle jouit. En effet, soient $s_1(t)$ et $s_2(t)$ deux signaux modulés en amplitude créés à partir des signaux modulants $m_1(t)$ et $m_2(t)$ sur une même porteuse. Une combinaison linéaire des signaux modulés s'exprime comme

$$\begin{aligned}
 a s_1(t) + b s_2(t) &= a A_c (1 + k_a m_1(t)) \cos(2\pi f_c t) + b A_c (1 + k_a m_2(t)) \cos(2\pi f_c t) \\
 &= A_c (a + b + k_a (a m_1(t) + b m_2(t))) \cos(2\pi f_c t)
 \end{aligned}
 \tag{3.54}$$

Il s'agit donc bien d'une modulation en amplitude par la même combinaison linéaire des signaux modulants. C'est de là que vient la simplicité de l'analyse spectrale pour la modulation d'amplitude; celle-ci se ramène à des translations du spectre sur l'axe des fréquences.

La modulation angulaire ne possède pas une telle simplicité. En particulier, il est impossible de faire l'analyse spectrale en général. Seule est théoriquement modélisable l'application de la modulation angulaire à des signaux modulants typiques et simples, dont le cas de la cosinussoïde pure. Il faudra donc extrapoler, avec les réserves de circonstance quant à l'interprétation subséquente, le résultat acquis pour des signaux simples.

Le signal simple choisi pour cette étude est le signal modulant

$$m(t) = A_m \cos(2\pi f_m t)
 \tag{3.55}$$

correspondant à un ton simple dans le cas d'un signal sonore, pour lequel on dispose des outils mathématiques. Tant l'amplitude que la fréquence de ce signal sont des paramètres dans l'étude.

Modulation par une cosinusoïde

Il convient d'abord d'établir quelques équations valables pour les divers cas de modulation angulaire analogique et de bien saisir comment évoluent les caractéristiques du signal modulé en fonction des paramètres A_m et f_m du signal modulant $m(t) = A_m \cos(2\pi f_m t)$. Par définition, la fréquence instantanée du signal modulé vaut

$$f_i(t) = f_c + k_f A_m \cos(2\pi f_m t) \quad (3.56)$$

$$= f_c + \Delta f \cos(2\pi f_m t) \quad (3.57)$$

où $\Delta f = k_f A_m$ est l'excursion ou déviation de fréquence. On voit que l'excursion de fréquence est proportionnelle à l'amplitude du signal modulant mais qu'elle ne dépend pas de sa fréquence.

Quant à la phase instantanée, l'application de l'équation 3.52 fournit

$$\phi_i(t) = 2\pi f_c t + \frac{\Delta f}{f_m} \sin(2\pi f_m t) \quad (3.58)$$

Cela entraîne une valeur d'indice de modulation qui est

$$\beta = \frac{\Delta f}{f_m} \quad (3.59)$$

D'un point de vue physique, l'indice de modulation β représente l'écart angulaire maximum par rapport à l'angle de la porteuse ; β s'exprime en radians.

La phase instantanée d'une porteuse modulée en fréquence par un signal modulant cosinusoidal s'exprime dès lors sous la forme

$$\phi_i(t) = 2\pi f_c t + \beta \sin(2\pi f_m t) \quad (3.60)$$

On peut voir le cosinus de la porteuse comme la projection d'un rayon qui tourne le long du cercle trigonométrique (voir figure 3.16). La modulation implique des oscillations de la position du rayon autour d'une position déterminée par la vitesse de rotation angulaire $2\pi f_c$; l'amplitude des oscillations est déterminée par l'indice de modulation.

Analyse spectrale

Examinons à présent le spectre du signal FM tel qu'il résulte de l'application d'un signal modulant cosinusoidal

$$s(t) = A_c \cos(2\pi f_c t + \beta \sin(2\pi f_m t)) \quad (3.61)$$

On peut toujours développer l'expression du signal modulé

$$s(t) = A_c \cos(2\pi f_c t) \cos(\beta \sin(2\pi f_m t)) - A_c \sin(2\pi f_c t) \sin(\beta \sin(2\pi f_m t)) \quad (3.62)$$

Pour calculer le spectre du signal $s(t)$, on recourt à la formule suivante [8, page 987]

$$e^{j\beta \sin \psi} = J_0(\beta) + 2 \sum_{k=1}^{+\infty} j^k J_k(\beta) \cos \left(k \left(\psi - \frac{\pi}{2} \right) \right) \quad (3.63)$$

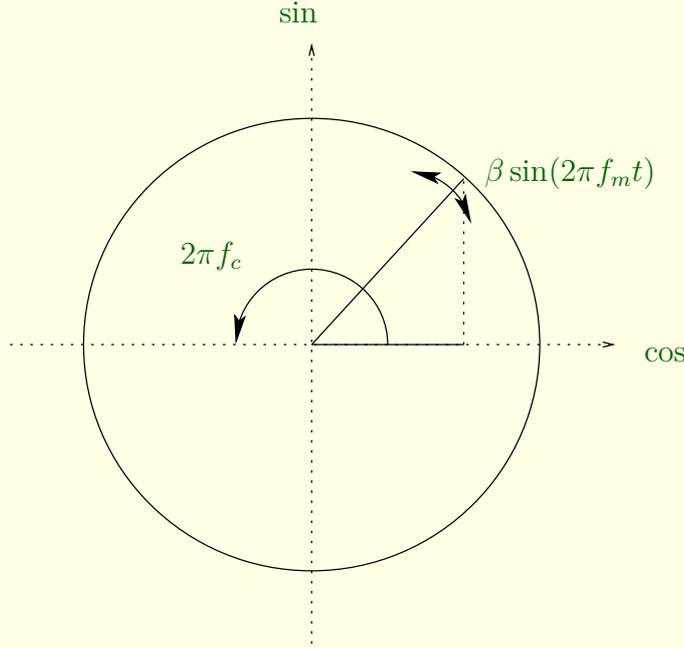


FIGURE 3.16 – Interprétation de la modulation de fréquence.

ce qui peut encore s'écrire

$$\cos(\beta \sin \psi) = J_0(\beta) + 2J_2(\beta) \cos(2\psi) + 2J_4(\beta) \cos(4\psi) + \dots \quad (3.64)$$

$$\sin(\beta \sin \psi) = 2J_1(\beta) \sin \psi + 2J_3(\beta) \sin(3\psi) + \dots \quad (3.65)$$

où $J_n(\beta)$ sont les fonctions de BESSEL de première espèce. La figure 3.17 représente la valeur des fonctions de BESSEL de première espèce.

Sachant en plus que les fonctions de BESSEL possèdent les propriétés $J_{2i}(\beta) = J_{-2i}(\beta)$ et $J_{2i+1}(\beta) = -J_{-(2i+1)}(\beta)$, on peut réécrire le signal modulé sous la forme

$$s(t) = A_c \sum_{n=-\infty}^{+\infty} J_n(\beta) \cos(2\pi(f_c + nf_m)t) \quad (3.66)$$

La transformée de FOURIER de ce signal fait apparaître un nombre infini de paires de raies aux fréquences $f_c + nf_m$ de part et d'autre de la porteuse et ayant des amplitudes $A_c J_n(\beta)$.

$$\mathcal{S}(f) = \frac{A_c}{2} \sum_{n=-\infty}^{+\infty} J_n(\beta) [\delta(f - f_c - nf_m) + \delta(f + f_c + nf_m)] \quad (3.67)$$

Alors que la position des raies latérales est régie par la fréquence du signal modulant, l'amplitude dépend de l'indice de modulation qui lui fait intervenir la fréquence f_m et l'excursion maximale de fréquence Δf .

On comprend bien l'expression du spectre en analysant les figures 3.18 et 3.19.

La première montre l'évolution des valeurs absolues des amplitudes des 8 premières fréquences autour de la porteuse (cette dernière est indiquée par 0); les amplitudes sont normalisées à 1 pour la facilité de la comparaison.

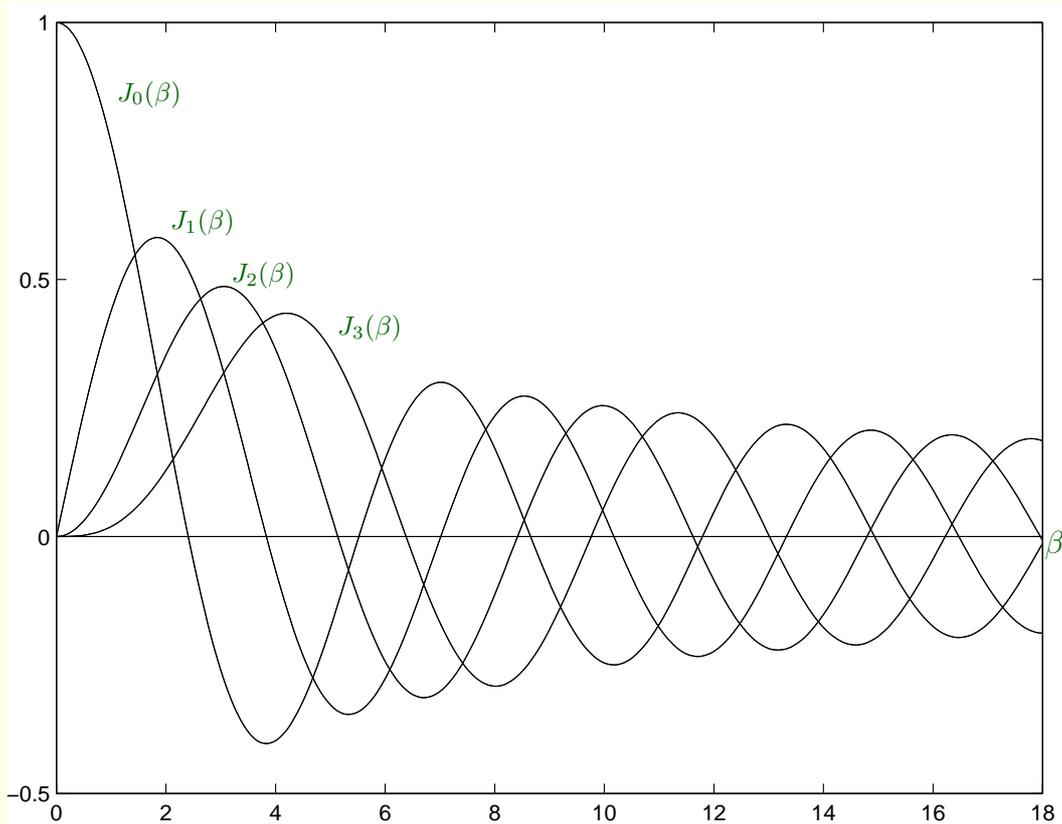


FIGURE 3.17 – Fonctions de BESSEL du premier ordre.

Pour poursuivre l'étude, distinguons deux cas en fonction de la valeur choisie pour l'indice de modulation β :

1. Modulation FM à faible indice ou modulation à bande étroite, pour laquelle β est petit par rapport à 1 radian.
2. Modulation FM à grand indice ou modulation à large bande. Dans ce cas, β est largement supérieur à 1 radian.

Modulation FM à faible indice

La connaissance de l'allure des fonctions de BESSEL permet de simplifier l'expression du signal modulé. En effet, on peut montrer que, pour de faibles valeurs d'indice de modulation, les fonctions de BESSEL prennent les valeurs suivantes

$$\begin{cases} J_0(\beta) \simeq 1 \\ J_1(\beta) \simeq \frac{\beta}{2} \\ J_n(\beta) \simeq 0, n > 2 \end{cases} \quad (3.68)$$

Autrement dit, seules les deux premières fonctions ont une valeur significative pour une modulation à faible indice. On en déduit que le signal modulé vaut

$$s(t) \simeq A_c \cos(2\pi f_c t) + \frac{1}{2} \beta A_c [\cos(2\pi(f_c + f_m)t) - \cos(2\pi(f_c - f_m)t)] \quad (3.69)$$

Ce résultat est quelque peu semblable à ce que fournirait le développement d'une modulation d'amplitude par un signal cosinusoidal, à savoir

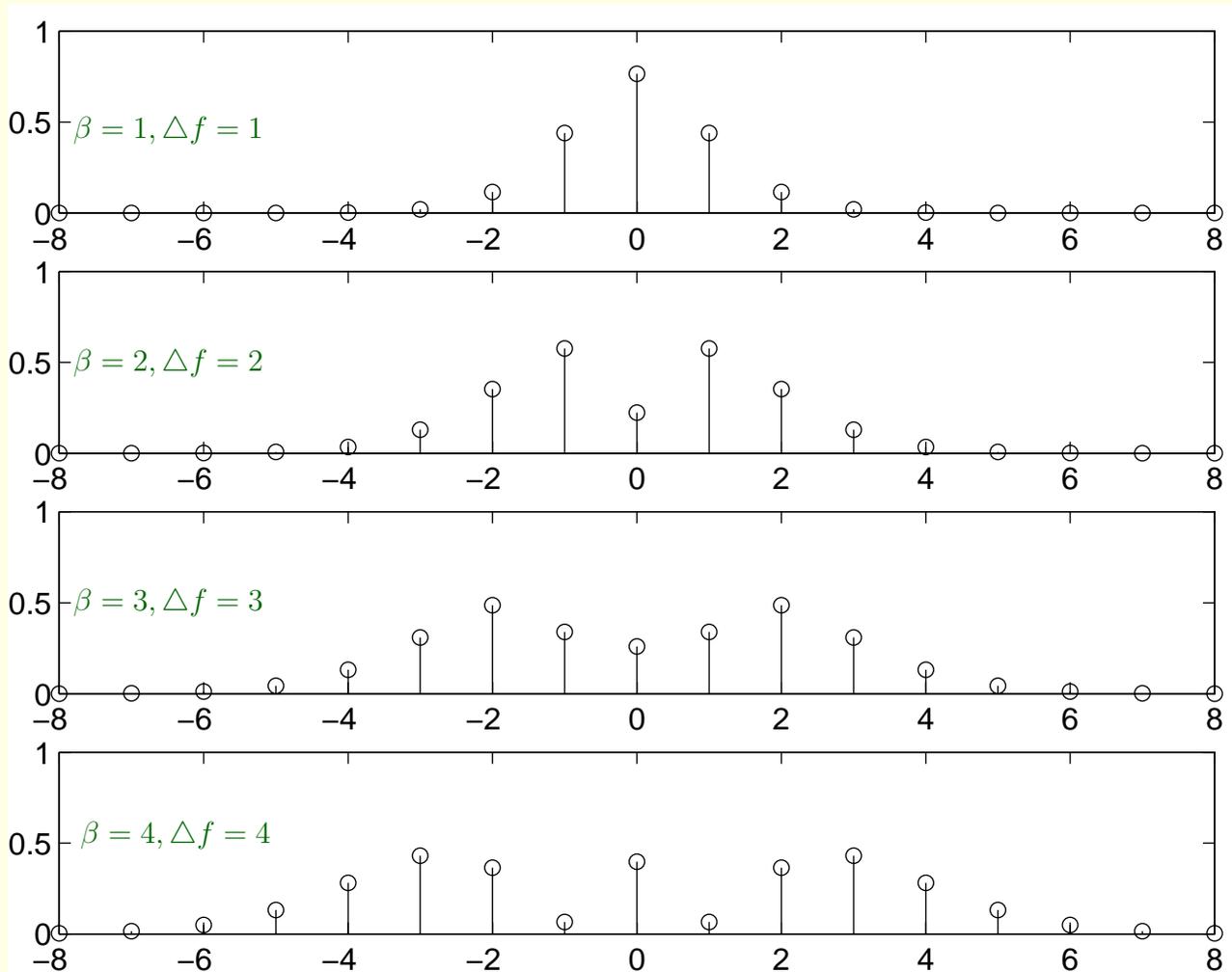


FIGURE 3.18 – Spectre d'un signal FM ($f_m = 1$, β variable).

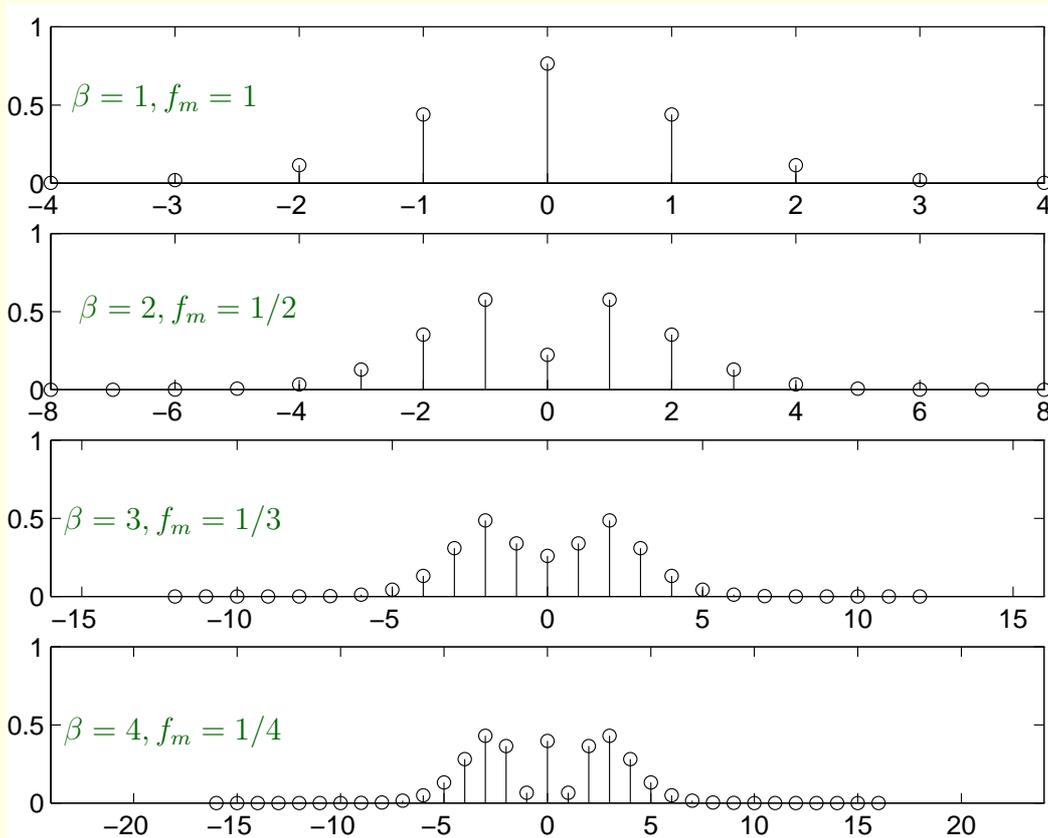


FIGURE 3.19 – Spectre d'un signal FM ($\Delta f = 1$, β variable).

$$s(t)_{AM} = A_c \cos(2\pi f_c t) + \frac{1}{2} k_a A_c [\cos(2\pi(f_c + f_m)t) + \cos(2\pi(f_c - f_m)t)] \quad (3.70)$$

La différence entre une modulation de fréquence à faible indice et une modulation d'amplitude par un signal cosinusoidal se remarque au niveau du signe du dernier terme. La bande passante nécessaire est, par contre, identique.

Modulation FM à grand indice

Dans le cas d'une modulation de fréquence à grand indice, la détermination de la bande passante est plus délicate. Il existe néanmoins une propriété intéressante relative à la puissance et la distribution des valeurs relatives des termes du spectre. En effet, quelle que soit la valeur de l'indice de modulation, la puissance du signal modulé vaut invariablement $P = \frac{1}{2} A_c^2$. On remarquera d'autre part que $\sum_{n=-\infty}^{+\infty} J_n^2(\beta) = 1$, d'où l'on peut déduire

$$P = \frac{1}{2} A_c^2 \sum_{n=-\infty}^{+\infty} J_n^2(\beta) \quad (3.71)$$

La somme des énergies des raies est donc constante et égale à la puissance du signal. Un étalement du spectre implique dès lors une diminution globale des amplitudes des raies. Parfois même, la porteuse disparaîtra pour certains choix de valeur pour l'indice de modulation⁴.

4. Supprimer totalement la porteuse est généralement déconseillé car cela complique l'acquisition de la porteuse à la réception; mais c'est néanmoins possible.

3.3.3 Bande passante requise

La bande passante d'une modulation FM à grand indice est théoriquement infinie. Il apparaît néanmoins que la puissance est principalement véhiculée par la porteuse et quelques harmoniques autour de cette fréquence.

Estimation empirique

On peut alors estimer la bande passante au moyen de la formule empirique suivante

Proposition 21 [Règle de Carson] *La bande passante requise est*

$$B \simeq 2(\Delta f + f_m) = 2\Delta f \left(1 + \frac{1}{\beta}\right) \quad \beta > 100 \quad (3.72)$$

où f_m correspond à la plus haute composante fréquentielle non nulle du signal modulant.

Pour une modulation à faible indice (par exemple $\beta < 0,5$), la bande passante est approximativement

$$B \simeq 2f_m \quad \beta < 0,5 \quad (3.73)$$

Pour des valeurs d'indice β intermédiaires ($0,5 \leq \beta \leq 100$), la règle de CARSON sous-estime la largeur de bande. On prend alors la valeur de la règle modifiée suivante

$$B \simeq 2(2\Delta f + f_m) = 2\Delta f \left(2 + \frac{1}{\beta}\right) \quad 0,5 \leq \beta \leq 100 \quad (3.74)$$

Exemple. La déviation maximale de fréquence en radio FM est limitée à $\Delta f = 75 [kHz]$. Comme le signal composite d'une émission audio stéréo s'étend, jusqu'à $f_m = 60 [kHz]$, on obtient une bande passante de l'ordre de $270 [kHz]$ par la formule de CARSON.

Estimation numérique

Pour une estimation plus rigoureuse de la bande passante, on recourt à un calcul numérique. Ce calcul s'appuie sur le développement suivant (cf. [1]).

Soit un signal modulant $m(t)$. Le signal modulé vaut

$$s(t) = A_c \cos(2\pi f_c t + 2\pi k_f \int_0^t m(t') dt') \quad (3.75)$$

Supposons que la phase $2\pi k_f \int_0^t m(t') dt' = \phi(t)$ soit un signal périodique de période T_0 . Dès lors,

$$s(t) = A_c \cos(2\pi f_c t + \phi(t)) = \text{Re} \left(A_c e^{2\pi j f_c t + j\phi(t)} \right) \quad (3.76)$$

S'agissant d'un signal $\phi(t)$ périodique, il est possible de l'exprimer par une série de FOURIER

$$e^{j\phi(t)} = \sum_{k=-\infty}^{+\infty} \mathcal{X}[k] e^{2k\pi j f_0 t} \quad \text{où} \quad \mathcal{X}[k] = \frac{1}{T_0} \int_{T_0} e^{j\phi(t) - 2k\pi j f_0 t} dt \quad (3.77)$$

Par substitution dans l'équation 3.76,

$$s(t) = \text{Re} \left(A_c \sum_{k=-\infty}^{+\infty} \mathcal{X}[k] e^{2\pi j(f_c + kf_0)t} \right) \quad (3.78)$$

La relation de PARSEVAL nous permet d'affirmer que

$$P = \frac{1}{T_0} \int_{T_0} \|e\|^{j\phi(t)^2} dt = \sum_{k=-\infty}^{+\infty} \|\mathcal{X}[k]\|^2 = 1 \quad (3.79)$$

Comme $\|e\|^{j\phi(t)^2} = 1$, la puissance totale est normalisée à 1. On peut alors calculer le nombre N de raies latérales autour de la fréquence porteuse tel que $\sum_{k=-N}^{+N} \|\mathcal{X}[k]\|^2$ soit égal à un certain pourcentage de la puissance totale, par exemple 98%. La bande passante vaut alors $2Nf_0$.

3.3.4 Réalisation de la modulation et de la démodulation en fréquence

Sans entrer dans le détail, la modulation ou démodulation doit être abordée en toute généralité par le biais de la modulation angulaire. D'un point de vue pratique néanmoins, les modulations de phase et de fréquence posent des problèmes différents. En effet, la modulation de phase devient difficile si l'indice de modulation est grand. Par contre, en modulation de fréquence, il y a une contradiction apparente entre la stabilité d'un oscillateur et la modification de la fréquence.

On distingue trois méthodes pour la démodulation de fréquence :

1. Démodulation d'amplitude suivie d'une interprétation en termes de la phase. C'est le principe du discriminateur.
2. Utilisation d'un corrélateur. Cette technique est surtout utilisée pour les modulations angulaires numériques.
3. Utilisation d'un circuit appelé boucle à verrouillage de phase dont l'objectif consiste à suivre l'évolution de la phase et de la fréquence de la porteuse.

3.3.5 Résumé des principaux résultats des modulations d'onde continue

Le tableau 3.20 reprend l'expression analytique de certains signaux modulés ainsi que leur spectre respectif.

3.4 Partage du plan de fréquence

3.4.1 Multiplexage en fréquence

L'utilisation de certains supports de transmission exige un partage adéquat des ressources fréquentielles. La technique réalisant ce partage est appelée multiplexage en fréquence ou Frequency Division Multiplexing (FDM). La figure 3.21 en illustre le principe.

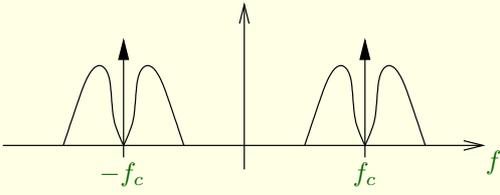
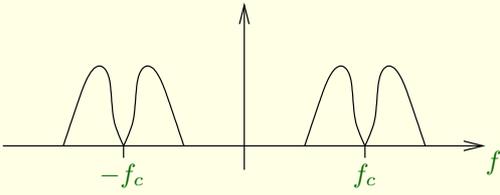
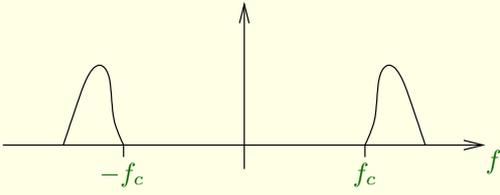
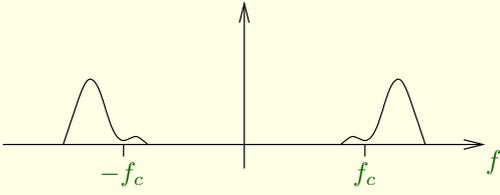
	$s(t)$	$\ S(f)\ $	BP
AM classique	$A_c(1 + k_a m(t)) \cos(2\pi f_c t)$		$2W$
DSB-SC	$A_c m(t) \cos(2\pi f_c t)$		$2W$
QAM	$A_c m_1(t) \cos(2\pi f_c t) + A_c m_2(t) \sin(2\pi f_c t)$		$2W$
SSB			W
VSB			$(1 + \alpha)W$
FM	$A_c \cos(2\pi f_c t + 2\pi k_f \int_0^t m(\tau) d\tau)$		$2(\Delta f + W)$

FIGURE 3.20 – Tableau récapitulatif des modulations d'onde continue.

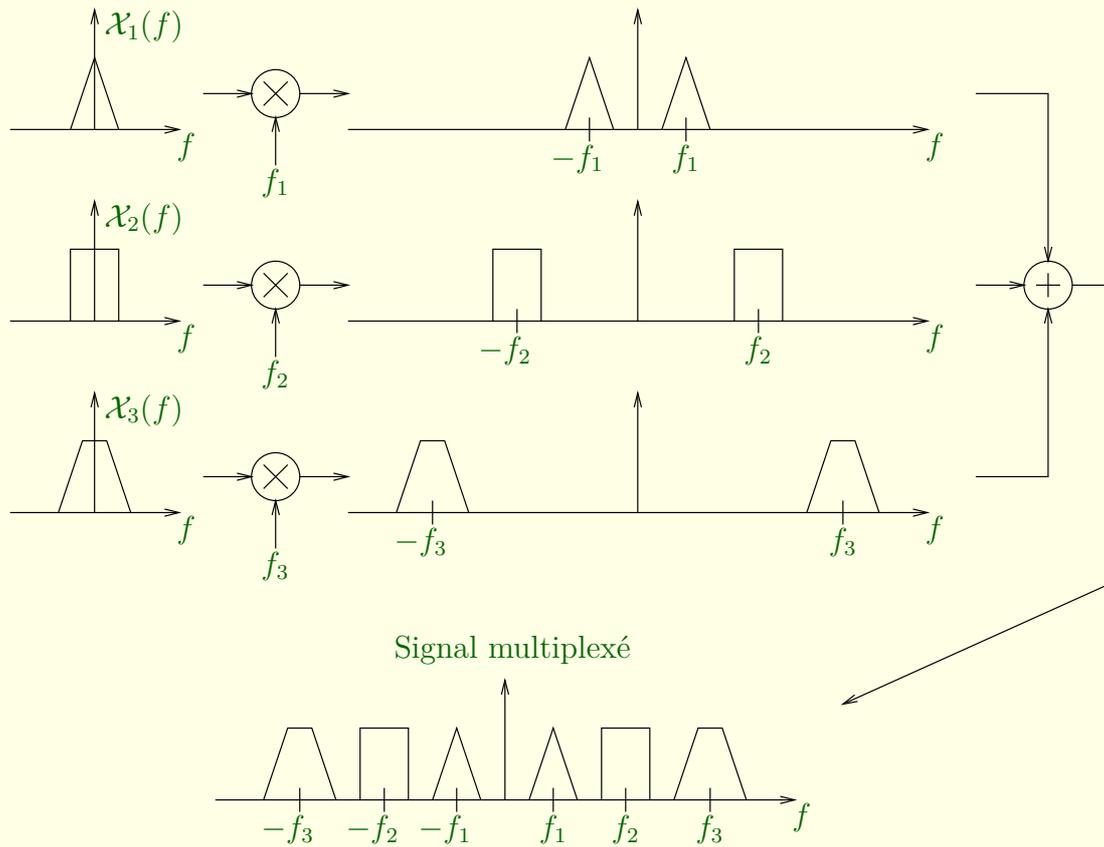


FIGURE 3.21 – Principe du multiplexage en fréquence.

On dispose d'une série de signaux $\mathcal{X}_i(f)$ en bande de base à transmettre simultanément. Au moyen de mélangeurs accordés à des fréquences spécifiques, le spectre de chaque signal est déplacé le long de l'axe des fréquences et ajouté au signal multiplex de manière à couvrir une certaine plage fréquentielle, tout en évitant un chevauchement en ménageant des bandes de garde entre les signaux.

Le signal multiplexé est transmis au récepteur qui doit extraire un à un tous les signaux au moyen de mélangeurs accordés aux mêmes fréquences qu'à l'émission. Le principe est identique à celui d'une démodulation cohérente (voir figure 3.22).

3.4.2 Récepteur super-hétérodyne

La transposition d'un signal situé à très haute fréquence vers la bande de base n'est pas réalisable en une étape unique en raison de la difficulté qu'il y a à construire des filtres à bande très étroites à haute fréquence. On procède alors par étapes intermédiaires ; c'est le principe du récepteur super-hétérodyne, tel qu'illustré à la figure 3.23.

Le premier étage du récepteur ramène les signaux à une fréquence intermédiaire f_{IF} . Puis le signal est transposé en bande de base grâce à un deuxième étage. La fréquence intermédiaire est la même pour tous les signaux, ce qui facilite la conception du système ainsi que la production.

Le principe du récepteur super-hétérodyne est largement utilisé en pratique : radiodiffusion FM, communications par satellite, etc.

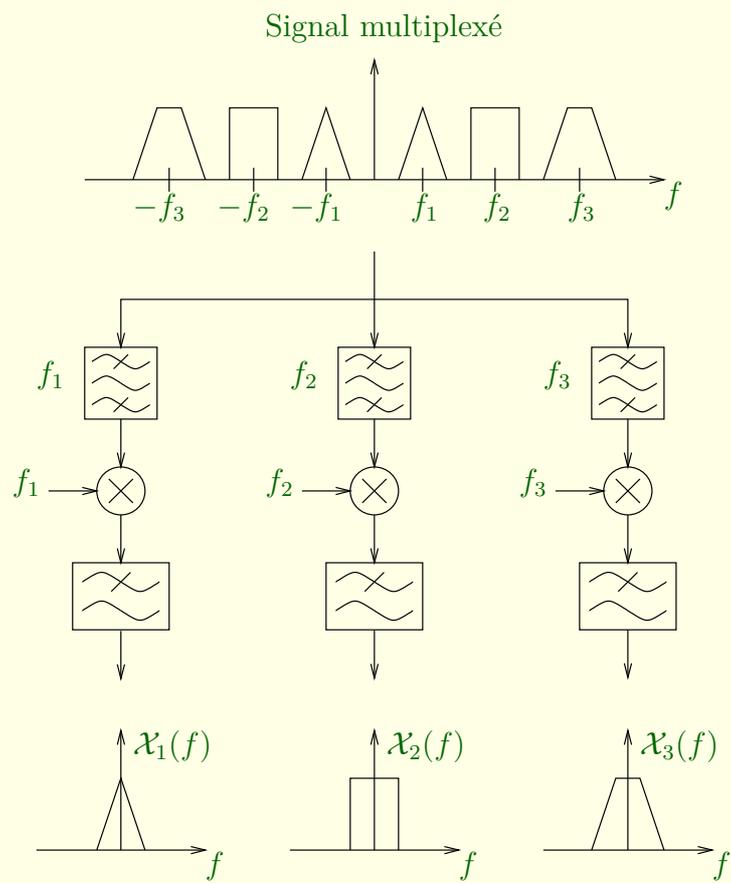


FIGURE 3.22 – Principe du démultiplexage en fréquence.

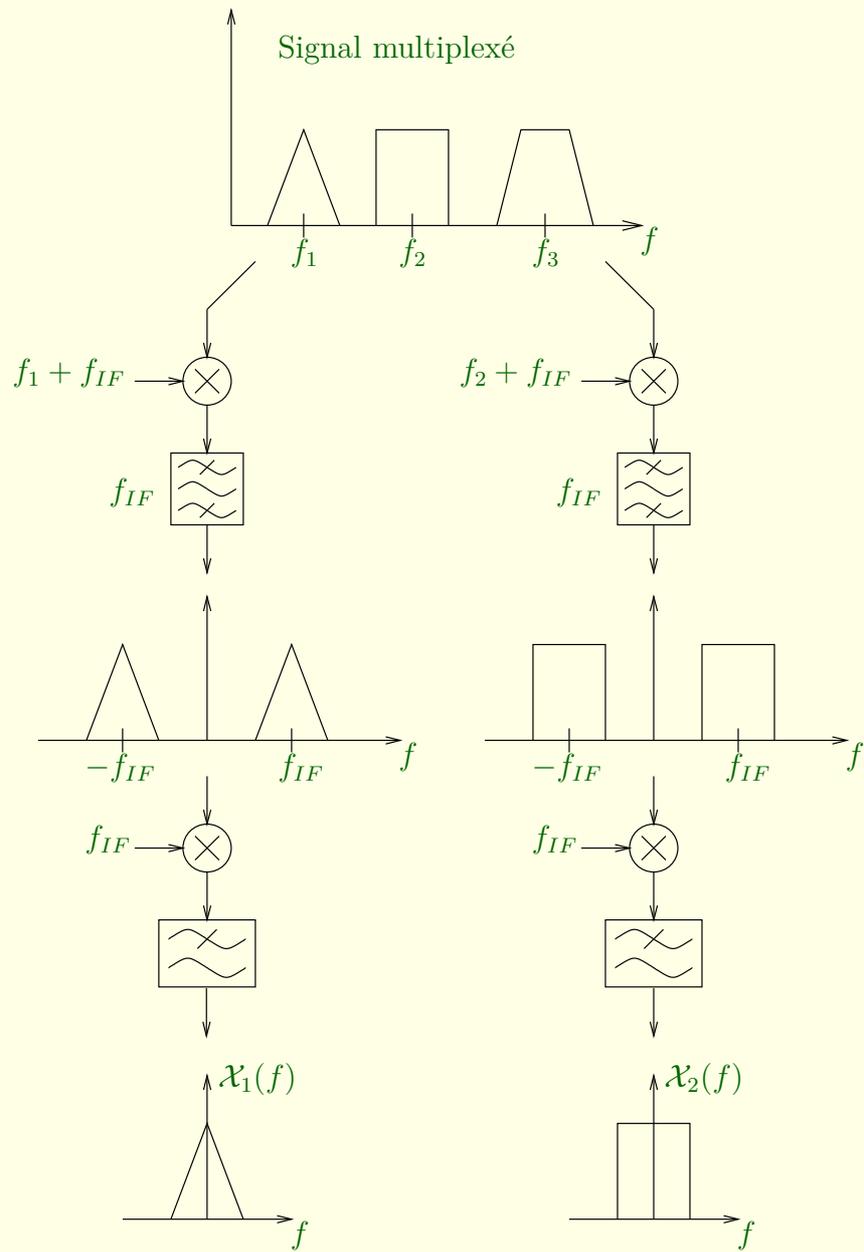


FIGURE 3.23 – Principe du récepteur super-hétérodyne.

3.5 Introduction à la modulation numérique

Jusqu'à présent, nous avons examiné la modulation par des signaux continus. Ce n'est pas la seule manière de concevoir un système de télécommunications. Dans un système numérique, l'amplitude, la phase ou la fréquence prennent des valeurs tirées d'un ensemble fini de valeurs possibles ; cette valeur est alors maintenue constante sur la durée d'un bit. La figure 3.24 montre le cas de signaux choisis à chaque fois parmi deux valeurs possibles.

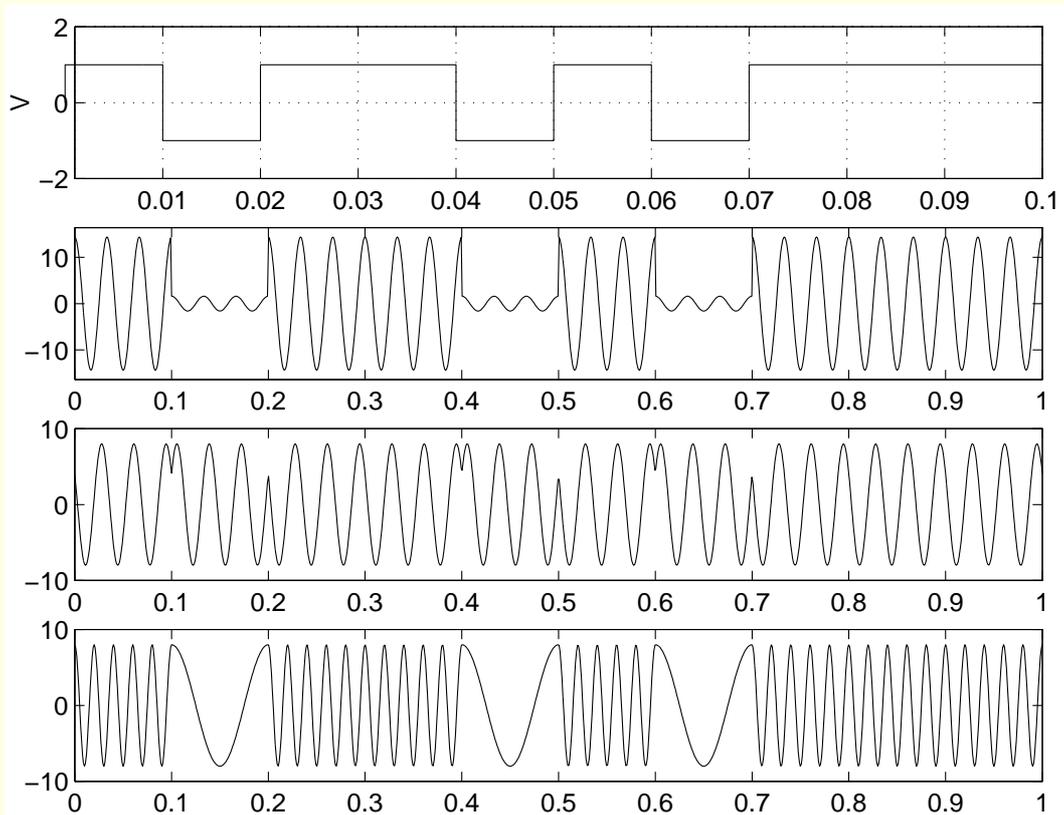


FIGURE 3.24 – Signal modulant numérique et signaux modulés respectivement en AM, PM et FM.

Ce type de modulation offre une meilleure résistance au bruit. Le prix à payer est un accroissement de la bande passante, auquel on pourra remédier par la compression du signal à transmettre.

Chapitre 4

Variables aléatoires, processus stochastiques et bruit

4.1 Introduction

La modulation d'une porteuse par un signal déterministe résulte en un signal déterministe. En effet, si l'on considère le cas de la modulation de fréquence, le signal modulé vaut $s(t) = A_c \cos(2\pi f_c t + 2\pi k_f \int_0^t m(t') dt')$, soit un signal prévisible pour chaque instant t si $m(t)$ est connu.

La maîtrise de la connaissance déterministe s'arrête malheureusement là car, dès l'instant où le signal est injecté dans un système de télécommunications, de nombreux bruits affectent la transmission. Or ces bruits ne suivent pas une évolution déterministe, d'où le besoin d'adopter une autre démarche analytique. La théorie des variables aléatoires et des processus stochastiques vient à notre aide. L'approche de cette théorie consiste à traiter les phénomènes comme des processus dont les caractéristiques statistiques sont connues. C'est donc en termes de probabilité qu'il conviendra de raisonner.

On peut tirer un parallèle avec les problèmes rencontrés pour modéliser le réseau Internet. Si la modélisation déterministe de quelques connexions simultanées est de l'ordre du possible, il est illusoire d'espérer représenter un nombre important de connections par un modèle déterministe.

4.2 Théorie des probabilités

Considérons une expérience dont le résultat n'est pas connu mais bien l'ensemble des résultats possibles. Un tel type d'expérience est appelé expérience aléatoire. Par exemple, l'expérience peut être l'observation du résultat du lancé d'une pièce de monnaie. Dans cette expérience, les deux seuls résultats possibles sont pile ou face.

Supposons que l'événement A représente un des résultats possible d'une expérience aléatoire. Par exemple, dans l'expérience de la pièce de monnaie, A peut représenter pile. Appelons $N_n(A)$ le nombre de fois que l'événement A se produit lorsque l'expérience est réalisée n fois (dans les mêmes conditions). $N_n(A)$ peut donc prendre n'importe quelle valeur entière comprise entre 0 et n . Nous pouvons donc écrire

$$0 \leq \frac{N_n(A)}{n} \leq 1 \quad (4.1)$$

Le rapport $\frac{N_n(A)}{n}$ est appelé fréquence relative de l'événement A . Nous dirons que l'expérience présente une régularité statistique si, pour n'importe quelle séquence de n réalisations de l'expérience, la fréquence relative $\frac{N_n(A)}{n}$ converge vers la même limite lorsque n devient très grand. Il semble alors naturel de définir la probabilité de l'événement A comme

$$p(A) = \lim_{n \rightarrow \infty} \left(\frac{N_n(A)}{n} \right) \quad (4.2)$$

Il est important de remarquer que la limite dans l'équation 4.2 ne doit pas être vue d'un aspect mathématique mais qu'il s'agit plutôt d'une définition empirique.

Par exemple, dans l'expérience du lancé de la pièce, nous pouvons supposer que sur un million de lancers, il y ait autant de pile que de face, à condition bien sûr que la pièce ne soit pas truquée! Nous dirons que la probabilité de l'événement pile est égale à $1/2$,

$$p(\text{pile}) = p(\text{face}) = \frac{1}{2}$$

4.2.1 Axiomatique des probabilités

On considère lors d'une expérience aléatoire, l'ensemble Ω de toutes les descriptions possibles du résultat de cette expérience. Un élément de Ω , noté ω dans la suite, décrit une réalisation particulière de l'expérience en cours et sera appelé épreuve. Un événement A est décrit par une propriété caractéristique susceptible d'être vérifiée lors de différentes épreuves. Par conséquent, un événement est un sous-ensemble de Ω , $A \subseteq \Omega$. Il faut bien noter que la description des résultats d'une expérience aléatoire est totalement subjective et laissée à l'appréciation de l'observateur.

Historiquement, la définition de la probabilité d'un événement a été basée sur la fréquence d'apparition de cet événement : à savoir le nombre de cas favorables sur le nombre de cas possibles. On sait que cette approche bien que très parlante à l'intuition, reste floue et conduit à des ambiguïtés. Il faut prendre conscience que les traitements de plus en plus raffinés que l'on fait subir aujourd'hui aux signaux, ne peuvent se contenter d'un point de vue probabiliste aussi élémentaire.

Il est intéressant d'associer à une expérience aléatoire et ses résultats possibles, un espace et ses points. Nous notons alors ω_k le point associé au k -ième résultat possible de l'expérience. L'ensemble de tous les points résultant de l'association de tous les résultats possibles de l'expérience est appelé espace témoin, que nous noterons Ω . Un événement peut correspondre à un point ou à un ensemble de points de Ω . En particulier, l'ensemble de tous les points de Ω est appelé événement certain et l'ensemble vide est appelé événement impossible. Un point seul est finalement appelé événement élémentaire.

Considérons, à titre d'exemple, l'expérience du lancer de dé. Dans cette expérience, 6 résultats sont possibles : 1, 2, 3, 4, 5 ou 6 sur la face supérieure du dé. En assignant un point à chacun de ces résultats, on obtient un espace témoin à une dimension contenant 6 points, comme le montre la figure 4.1. L'événement élémentaire tirer un 5 correspond au point $\{5\}$ de l'espace tandis que l'événement tirer un chiffre pair correspond au sous-ensemble $\{2, 4, 6\}$.

En théorie des probabilités, on introduit un ensemble noté Ω dont les éléments ω symbolisent les différentes épreuves d'une expérience aléatoire. Par épreuve il faut entendre une réalisation possible de l'expérience.

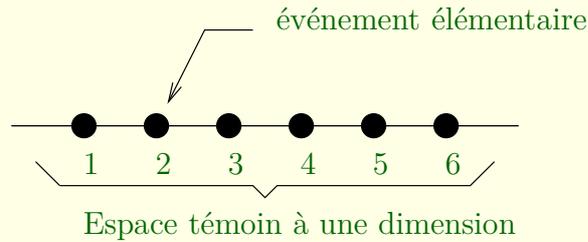


FIGURE 4.1 – Espace témoin de l’expérience du lancer de dé.

La théorie des probabilités ne s’intéresse qu’aux ensembles d’événements qui possèdent une structure particulière appelée tribu. Une tribu ou σ -algèbre \mathcal{F} sur Ω est une famille d’événements (sous-ensembles) de Ω telle que :

- Ω et \emptyset sont dans \mathcal{F} ,
- si $A \in \mathcal{F}$ alors le complémentaire de A , noté \bar{A} , appartient à \mathcal{F} ,
- si la suite $A_{n \in \mathbb{N}} \in \mathcal{F}$ alors $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}$ (union dénombrable permise).

Le couple (Ω, \mathcal{F}) est dit espace mesurable.

En pratique, on est intéressé par la probabilité d’un événement. Par définition, une mesure de probabilité est une application P de la tribu des événements \mathcal{F} vers \mathbb{R} composée comme suit :

1. Un espace témoin Ω d’événements élémentaires.
2. Une classe \mathcal{L} d’événements qui sont des sous-ensembles de Ω .
3. Une mesure de la probabilité $p(\cdot)$ associée à chaque événement A de la classe \mathcal{L} et qui a les propriétés suivantes :
 - (a) $p(\Omega) = 1$
 - (b) $0 \leq p(A) \leq 1$
 - (c) Si la suite $A_{n \in \mathbb{N}} \in \mathcal{F}$ avec $A_i \cap A_j = \emptyset$ pour $i \neq j$ alors $p(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} p(A_n)$.

Un espace probabilisé est un triplet (Ω, \mathcal{F}, P) où Ω est un ensemble quelconque, \mathcal{F} est une tribu de Ω , \mathcal{F} étant un sous-ensemble de Ω , et P est une mesure positive définie sur \mathcal{F} , appelée probabilité, telle que $P(\Omega) = 1$.

Les propriétés (a), (b) et (c) sont connues comme axiomes des probabilités. L’axiome (a) indique que la probabilité de l’événement certain vaut 1. L’axiome (b) indique que la probabilité d’un événement est un nombre réel non négatif, inférieur ou égal à 1. Finalement, étant donné que A et B sont deux événements mutuellement exclusifs, l’axiome (c) indique que la probabilité que l’événement A se produise ou que l’événement B se produise est égale à la somme des probabilités respectives de ces deux événements.

Propriétés.

1. $p(A^c) = 1 - p(A)$ où A^c (appelé “non A ”) est le complément de A .
2. Si M événements mutuellement exclusifs A_1, A_2, \dots, A_M ont la propriété suivante :

$$A_1 + A_2 + \dots + A_M = \Omega \tag{4.3}$$

alors

$$p(A_1) + p(A_2) + \dots + p(A_M) = 1 \tag{4.4}$$

3. Si les événements A et B ne sont pas mutuellement exclusifs, la probabilité de l'événement union “ A ou B ” égale

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) \quad (4.5)$$

où $p(A \cup B)$ est la probabilité de l'événement joint “ A et B ”.

Définition 22 $p(A \cap B)$ est appelée probabilité conjointe. Nous avons l'interprétation suivante :

$$p(A \cap B) = \lim_{n \rightarrow \infty} \left(\frac{N_n(A \cap B)}{n} \right) \quad (4.6)$$

où $N_n(A \cap B)$ est le nombre de fois que les événements A et B se réalisent simultanément, n étant le nombre de fois que l'expérience a été répétée.

L'axiome (c) est un cas particulier de la relation 4.5. Quand A et B sont mutuellement exclusifs, $p(A \cap B) = 0$.

Probabilité conditionnelle

Définition 23 $p(B|A)$ est appelée probabilité conditionnelle. Elle représente la probabilité de l'événement B , étant donné que l'événement A s'est réalisé. En supposant que $p(A) \neq 0$, la probabilité conditionnelle $p(B|A)$ est définie par

$$p(B|A) = \frac{p(A \cap B)}{p(A)} \quad (4.7)$$

où $p(A \cap B)$ est la probabilité conjointe de A et B .

Nous pouvons réécrire l'équation 4.7 comme

$$p(A \cap B) = p(B|A)p(A) \quad (4.8)$$

Il est évident que nous pouvons écrire également

$$p(A \cap B) = p(A|B)p(B) \quad (4.9)$$

Il existe des situations pour lesquelles la probabilité conditionnelle $p(B|A)$ et les probabilités $p(A)$ et $p(B)$ peuvent être déterminées aisément alors que la probabilité conditionnelle $p(A|B)$ est inconnue. À partir de 4.8 et 4.9, il est possible de déterminer $p(A|B)$ en utilisant la relation

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} \quad (4.10)$$

qui constitue une forme spéciale de la formule de BAYES.

Proposition 24 Formule de BAYES

$$p(A_i|B) = \frac{p(B|A_i)p(A_i)}{\sum_{j=1}^N p(B|A_j)p(A_j)} \quad (4.11)$$

Elle permet le calcul des probabilités a posteriori $p(A_i|B)$ en terme de probabilités a priori $p(B|A_i)$.

Supposons maintenant que la probabilité conditionnelle $p(B|A)$ soit simplement égale à la probabilité d'occurrence de l'événement B , c'est-à-dire

$$p(B|A) = p(B) \quad (4.12)$$

Sous cette condition, la probabilité de l'événement joint $A \cap B$ est égale à

$$p(A \cap B) = p(A)p(B) \quad (4.13)$$

Dès lors,

$$p(A|B) = p(A) \quad (4.14)$$

Nous voyons que dans ce cas, la connaissance de l'occurrence d'un des deux événements ne nous renseigne pas plus sur la probabilité d'occurrence de l'autre événement. Des événements A et B qui satisfont cette condition sont dits statistiquement indépendants.

4.3 Variables aléatoires

Définition 25 Une variable aléatoire est une fonction dont le domaine est l'espace témoin d'une expérience aléatoire et dont la valeur est un nombre réel¹.

Quand le résultat d'une expérience aléatoire est ω , la variable aléatoire est dénotée $X(\omega)$ ou tout simplement X . Par exemple, dans l'expérience du lancé de dé, l'espace témoin est composé de six points que l'on peut assigner aux six entiers 1, 2, 3, 4, 5 et 6. Dès lors, on peut définir la variable aléatoire $X(k) = k$, où k est un de ces six nombres entiers. La variable aléatoire $X(k)$ représente donc le chiffre situé sur la face supérieure du dé. Dans cet exemple, la variable aléatoire peut seulement prendre des valeurs discrètes. Nous parlerons de variable aléatoire discrète. Plus précisément, une variable aléatoire discrète ne peut prendre qu'un nombre fini de valeurs dans un intervalle donné. Par contre, si la variable aléatoire peut prendre n'importe quelle valeur dans un intervalle donné, on parlera de variable aléatoire continue. Par exemple, la variable représentant la tension électrique de bruit aux bornes d'une résistance à un moment donné est une variable aléatoire continue car elle peut prendre n'importe quelle valeur entre moins et plus l'infini.

Il serait intéressant d'obtenir une description probabiliste des variables aléatoires qui s'adapte aussi bien aux variables discrètes que continues. C'est le propos des sections suivantes.

4.3.1 Fonctions de répartition et densité de probabilité

Soit X une variable aléatoire et considérons la probabilité de l'événement² $X \leq x$. Nous noterons $p(X \leq x)$ cette probabilité. Il est évident que celle-ci est une fonction de la variable x .

1. Sauf exception, nous traiterons des signaux réels. L'alternative consiste à définir toutes les notions à partir de variables aléatoires à valeurs complexes.

2. Ceci présuppose évidemment que l'on ait défini une notion d'ordre.

Définition 26 La fonction de distribution cumulative, encore appelée fonction de distribution ou fonction de répartition, d'une variable aléatoire X est définie par

$$F_X(x) = p(X \leq x) \quad (4.15)$$

Il est important de remarquer que $F_X(x)$ est une fonction de x et non de la variable aléatoire X . Pour chaque valeur de x , $F_X(x)$ représente une probabilité cumulée.

Propriétés de la fonction de répartition

1. $F_X(x)$ est comprise entre 0 et 1.
2. $F_X(x)$ est une fonction monotone non-décroissante de x , c'est-à-dire

$$x_1 \leq x_2 \Rightarrow F_X(x_1) \leq F_X(x_2) \quad (4.16)$$

Une description alternative de la probabilité d'une variable aléatoire est souvent utilisée, il s'agit de la dérivée de la fonction de répartition.

Définition 27 La fonction de densité de probabilité d'une variable aléatoire X est définie par

$$f_X(x) = \frac{dF_X(x)}{dx} \quad (4.17)$$

Il est important de remarquer que la dérivation dans la relation 4.17 se fait par rapport à la variable x . Le nom de densité provient du fait que la probabilité de l'événement $x_1 < X \leq x_2$ est égale à

$$p(x_1 < X \leq x_2) = p(X \leq x_2) - p(X \leq x_1) \quad (4.18)$$

$$= F_X(x_2) - F_X(x_1) \quad (4.19)$$

$$= \int_{x_1}^{x_2} f_X(x) dx \quad (4.20)$$

La probabilité d'un intervalle est donc égale à l'aire sous la fonction de densité de probabilité dans cet intervalle. En posant $x_1 = -\infty$ dans l'égalité 4.20, il est possible de lier la fonction de répartition à la fonction de densité de probabilité :

$$F_X(x) = \int_{-\infty}^x f_X(\xi) d\xi \quad (4.21)$$

Étant donné que $F_X(+\infty) = 1$ correspond à la probabilité de l'événement certain, nous pouvons déduire de 4.21

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1 \quad (4.22)$$

Nous avons mentionné précédemment que la fonction de répartition était toujours monotone croissante. Cela signifie que la fonction de densité de probabilité doit toujours être positive.

Propriétés de la fonction de densité de probabilité

1. $f_X(x)$ est toujours positive.
2. $f_X(x)$ possède sous sa courbe une aire totale égale à 1.

4.3.2 Moments d'une variable aléatoire

Le résultat d'une expérience aléatoire n'étant pas déterministe, il convient de caractériser le comportement moyen des résultats d'une expérience aléatoire par des outils appropriés. Pour ce faire, nous introduisons quelques définitions.

Définition 28 *Le n -ième moment de la variable aléatoire X est défini par*

$$\mu_{X^n} = E \{X^n\} = \int_{-\infty}^{+\infty} x^n f_X(x) dx \quad (4.23)$$

où $E \{ \}$ désigne l'opérateur d'espérance statistique.

Les deux premiers moments de la variable aléatoire X sont les plus importants. Considérons tout d'abord l'équation 4.23 pour $n = 1$.

Définition 29 [Moyenne]

$$\mu_X = E \{X\} = \int_{-\infty}^{+\infty} x f_X(x) dx \quad (4.24)$$

μ_X est appelée espérance ou moyenne de la variable aléatoire X .

La moyenne μ_X est située au centre de gravité de l'aire située sous la courbe de la fonction de densité de probabilité de X . Considérons maintenant l'équation 4.23 pour $n = 2$.

$$E \{X^2\} = \int_{-\infty}^{+\infty} x^2 f_X(x) dx \quad (4.25)$$

$E \{X^2\}$ représente la moyenne du carré de la variable X , à ne pas confondre avec le carré de la moyenne $(E \{X\})^2 = \mu_X^2$.

Définition 30 *Le n -ième moment centré de la variable aléatoire X est défini par*

$$E \{(X - \mu_X)^n\} = \int_{-\infty}^{+\infty} (x - \mu_X)^n f_X(x) dx \quad (4.26)$$

Pour $n = 1$, le moment centré est évidemment égal à zéro, tandis que pour $n = 2$, nous obtenons le moment centré d'ordre 2 aussi appelé variance de la variable aléatoire X .

Définition 31 [Variance]

$$\sigma_X^2 = var \{X\} = E \{(X - \mu_X)^2\} = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x) dx \quad (4.27)$$

σ_X , la racine carrée de la variance est appelée écart-type de la variable aléatoire X . La variance σ_X^2 de la variable aléatoire X est en quelque sorte une mesure de l'aspect aléatoire de la variable. En spécifiant la variance σ_X^2 , nous limitons la largeur effective de la fonction de densité de probabilité $f_X(x)$ autour de la moyenne μ_X .

Une description précise de cette limite a été donnée par CHEBYSHEV.

Théorème 32 [Chebyshev [3, page 177]] *Si Y est une variable aléatoire positive, alors pour tout $\epsilon > 0$ et $0 < p < +\infty$,*

$$P(Y \geq \epsilon) \leq \frac{E(Y^p)}{\epsilon^p} \quad (4.28)$$

Sous une forme particulière, on a

$$\forall \epsilon > 0 \quad p(|X - \mu_X| \geq \epsilon) \leq \frac{\sigma_X^2}{\epsilon^2} \quad (4.29)$$

Nous voyons donc que la moyenne et la variance d'une variable aléatoire donnent une description partielle de sa distribution de probabilité.

Finalement, nous tirons de 4.27 que la variance et la moyenne du carré de la variable sont liées par la relation

$$\begin{aligned} \sigma_X^2 &= E\{X^2 - 2\mu_X X + \mu_X^2\} \\ &= E\{X^2\} - 2\mu_X E\{X\} + \mu_X^2 \\ &= E\{X^2\} - \mu_X^2 \end{aligned}$$

dans laquelle nous avons utilisé le fait que l'opérateur d'espérance statistique E est linéaire.

4.3.3 Liens entre moments centrés ou non

Proposition 33 [3, page 135] *Les moments centrés d'ordre n sont entièrement déterminés par les moments d'ordre k tels que $k \leq n$.*

Démonstration

L'extension binomiale permet d'écrire la relation

$$(X - \mu_X)^n = \sum_{i=0}^n C_i^n X^i (-\mu_X)^{n-i} \quad (4.30)$$

Dès lors, par linéarité de l'espérance,

$$E\{(X - \mu_X)^n\} = \sum_{i=0}^n C_i^n (-\mu_X)^{n-i} E\{X^i\} \quad (4.31)$$

■

Proposition 34 *Les moments d'ordre n sont entièrement déterminés par les moments centrés d'ordre k tels que $k \leq n$.*

Démonstration

Par la proposition précédente et en prenant $X^n = (X - \mu_X + \mu_X)^n$

$$E\{X^n\} = E\{(X - \mu_X + \mu_X)^n\} = \sum_{i=0}^n C_i^n E\{(X - \mu_X)^i\} \mu_X^{n-i} \quad (4.32)$$

■

Variable aléatoire centrée

En rapport avec la moyenne et la variance, on associe à la variable aléatoire X , la variable aléatoire centrée X_c et la variable aléatoire réduite X_r définies respectivement par :

$$X_c = X - E\{X\} \quad (4.33)$$

$$X_r = \frac{X - E\{X\}}{\sigma} \quad (4.34)$$

Fonction caractéristique

Il est parfois plus facile de calculer les moments par le biais d'une représentation spectrale de la densité de probabilité; cela conduit à la définition de la fonction caractéristique.

Définition 35 La fonction caractéristique $\phi_X(\omega)$ d'une variable aléatoire X est définie par

$$\phi_X(\omega) = \int_{-\infty}^{+\infty} f_X(x) e^{j\omega x} dx \quad (4.35)$$

Par transformée inverse, il est possible de retrouver la fonction de densité de probabilité de X à partir de sa fonction caractéristique :

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \phi_X(\omega) e^{-j\omega x} d\omega \quad (4.36)$$

4.3.4 Variables aléatoires usuelles

1. Variable aléatoire continue et uniforme

Une variable aléatoire X présente une densité de probabilité uniforme sur l'intervalle (a, b) si

$$f_X(x) = \begin{cases} 0 & x \leq a \\ \frac{1}{b-a} & a < x \leq b \\ 0 & x > b \end{cases} \quad (4.37)$$

La fonction de répartition de X est alors donnée par

$$F_X(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x \leq b \\ 1 & x > b \end{cases} \quad (4.38)$$

La figure 4.2 montre le graphe de ces deux fonctions.

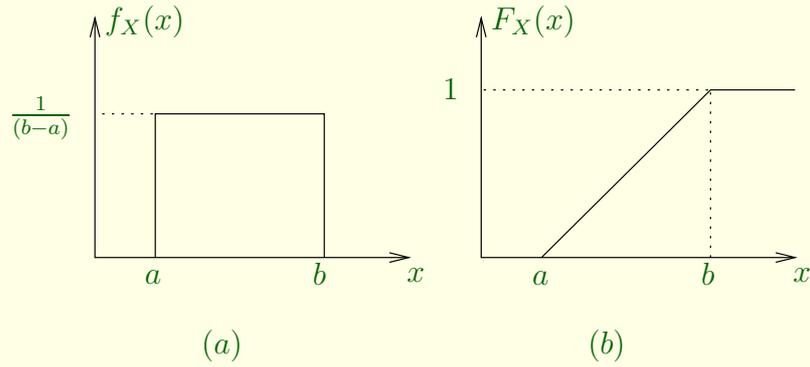


FIGURE 4.2 – Variable uniforme : (a) fonction de densité de probabilité, (b) fonction de répartition.

2. Variable binomiale ou variable de Bernoulli

Soit X une variable aléatoire discrète. La variable X présente une densité binomiale d'ordre n si elle peut prendre les valeurs $0, 1, \dots, n$ de telle sorte que

$$p(X = k) = C_n^k p^k q^{n-k} \quad \text{avec} \quad p + q = 1 \quad (4.39)$$

où

$$C_n^k = \frac{n!}{k!(n-k)!} \quad (4.40)$$

La fonction de densité de probabilité de X est donnée par

$$f_X(x) = \sum_{k=0}^n C_n^k p^k q^{n-k} \delta(x - k) \quad (4.41)$$

La fonction de répartition de X est une fonction en escaliers donnée par

$$F_X(x) = \sum_{k=0}^m C_n^k p^k q^{n-k} \quad \text{avec} \quad m \leq x < m + 1 \quad (4.42)$$

La loi binomiale est utilisée pour le dénombrement d'événements.

3. Variable aléatoire de Poisson

Soit X une variable aléatoire discrète. La variable X présente une densité de POISSON, avec un paramètre λ , si elle peut prendre les valeurs $0, 1, \dots, n, \dots$ de telle sorte que

$$p(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{avec} \quad k = 0, 1, \dots \quad (4.43)$$

La fonction de densité de probabilité de X est donnée par

$$f_X(x) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \delta(x - k) \quad (4.44)$$

Comme pour la variable binomiale, la fonction de répartition est une fonction en escalier.

4. Variable aléatoire gaussienne ou normale

Soit X une variable aléatoire de moyenne μ_X et de variance σ_X^2 . Cette variable présente une densité de probabilité gaussienne ou normale si, pour $-\infty < x < +\infty$, elle a la forme

$$f_X(x) = \frac{1}{\sigma_X \sqrt{2\pi}} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} \quad (4.45)$$

Deux résultats classiques permettent de rattacher la probabilité d'un événement et la fréquence d'apparition obtenue lors d'une séquence supposée finie d'essais répétés. Le premier porte le nom de loi des grands nombres (c'est en fait un théorème) et le second porte le nom de théorème de la limite centrale.

Ainsi, la loi de GAUSS joue un rôle privilégié à cause de l'existence du théorème de la limite centrale dont voici l'une des expressions.

Théorème 36 [Limite centrale] Si $\{X_i\}$ est une suite de variables aléatoires indépendantes, de même loi, centrées et de variance égale à 1, alors la fonction de répartition $F_Y(y)$ de la variable aléatoire

$$Y = \frac{1}{N} \sum_{i=1}^N X_i \quad (4.46)$$

vérifie

$$\lim_{N \rightarrow +\infty} F_Y(y) = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \quad (4.47)$$

Son application sur le plan pratique conduit à considérer comme gaussien un phénomène qui est la superposition d'un grand nombre d'effets élémentaires ayant la même origine : erreurs de mesure, bruit de fond des récepteurs.

4.3.5 Formule de changement de variable

On peut voir cette expression comme une généralisation des différents moments décrits plus hauts.

Un problème qui se pose souvent est de déterminer la fonction de densité de probabilité d'une variable aléatoire Y lorsque celle-ci est liée à une variable aléatoire X par la relation

$$Y = g(X) \quad (4.48)$$

La fonction de densité de probabilité de X étant connue, notre but est de déterminer $f_Y(y)$. Deux cas doivent être considérés : les transformations monotones et les transformations non-monotones dites "many-to-one".

Transformations monotones

Soit X une variable aléatoire continue avec une fonction de densité de probabilité $f_X(x)$. Supposons que $Y = g(X)$ soit une fonction monotone dérivable de X , comme montré à la figure 4.3.

Considérons maintenant l'événement $(y < Y \leq y + dy)$ où dy est un incrément infinitésimal. D'après la figure 4.3, on peut noter que cet événement correspond biunivoquement à l'événement $(x < X \leq x + dx)$ où dx est l'incrément infinitésimal correspondant à dy . Nous pouvons donc déduire que

$$p(y < Y \leq y + dy) = p(x < X \leq x + dx) \quad (4.49)$$

En termes de fonction de densité de probabilité, nous pouvons réécrire cette égalité sous la forme

$$f_Y(y)dy = f_X(x)dx \quad (4.50)$$

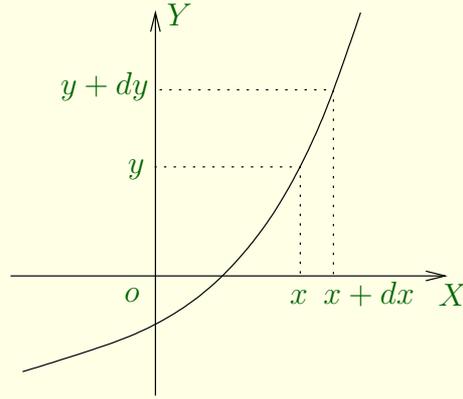


FIGURE 4.3 – Transformation monotone d’une variable aléatoire.

où nous avons supposé que $g(X)$ est une fonction monotone croissante, comme indiqué à la figure 4.3. Si maintenant $g(X)$ est une fonction monotone décroissante, nous avons

$$f_Y(y)dy = -f_X(x)dx \quad (4.51)$$

Nous pouvons combiner les deux résultats précédents en écrivant

$$f_Y(y) |dy| = f_X(x) |dx| \quad (4.52)$$

La relation 4.52 peut être vue comme une expression de la conservation de la probabilité, dans le cas d’une transformation monotone d’une variable aléatoire X vers une autre variable aléatoire Y . Il reste maintenant deux opérations à effectuer pour obtenir la fonction $f_Y(y)$:

- diviser les 2 membres de l’égalité 4.52 par $|dy|$.
- substituer x par $g^{-1}(y)$ dans l’expression résultante.

Nous écrivons donc

$$f_Y(y) = \frac{f_X(x)}{|dy/dx|} \quad (4.53)$$

$$= \frac{f_X(x)}{|dg/dx|} \Big|_{x=g^{-1}(y)} \quad (4.54)$$

qui constitue la formule de changement de variable dans le cas d’une transformation monotone.

Transformations non monotones

Considérons le cas plus général d’une fonction $y = g(x)$ pour laquelle plusieurs valeurs de l’argument x correspondent à une même valeur y de la fonction. Dans ce cas, la formule donnant la fonction de densité de probabilité de Y en fonction de celle de X est donnée par

$$f_Y(y) = \sum_k \frac{f_X(x_k)}{|dg/dx|} \Big|_{x_k=g^{-1}(y)} \quad (4.55)$$

où les x_k sont les racines de l’équation $y = g(x)$ et où la valeur absolue est utilisée parce qu’une probabilité est toujours positive. Si, pour une valeur particulière de y , l’équation $g(x) = y$ possède n racines, la somme 4.55 contiendra n termes.

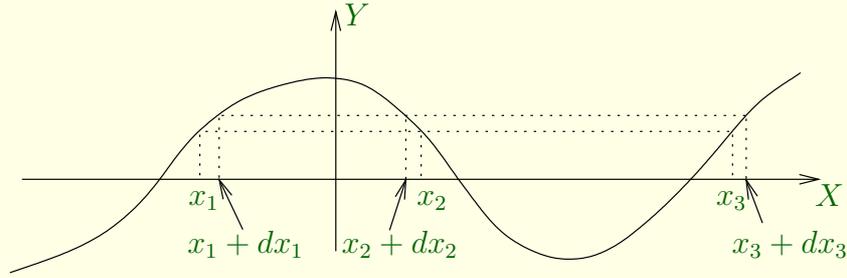


FIGURE 4.4 – Transformation non monotone.

Pour illustrer la validité de 4.55, considérons la situation décrite à la figure 4.4. Dans ce cas, l'équation $g(x) = y$ possède trois racines x_1 , x_2 et x_3 . L'événement $(y < Y \leq y + dy)$ se réalise lorsqu'un des trois événements $(x_1 < X \leq x_1 + dx_1)$, $(x_2 < X \leq x_2 + dx_2)$ ou $(x_3 < X \leq x_3 + dx_3)$ se réalise. En supposant que dy est suffisamment petit, ces trois événements sont mutuellement exclusifs et nous pouvons écrire

$$\begin{aligned}
 p(y < Y \leq y + dy) &= p(x_1 < X \leq x_1 + dx_1) \\
 &\quad + p(x_2 < X \leq x_2 + dx_2) \\
 &\quad + p(x_3 < X \leq x_3 + dx_3)
 \end{aligned}
 \tag{4.56}$$

En termes de fonctions de densité de probabilité,

$$f_Y(y)dy = f_X(x_1)dx_1 + f_X(x_2)|dx_2| + f_X(x_3)dx_3
 \tag{4.57}$$

où, pour le second terme, nous avons utilisé une valeur absolue car la dérivée dg/dx est négative en x_2 . En divisant les deux membres de cette équation par dy , nous obtenons la formule de changement de variables 4.55, dans le cas d'une transformation non monotone.

Moyenne

Définition 37 La moyenne d'une variable aléatoire Y , égale à une fonction $g(X)$ d'une autre variable aléatoire X , est définie par

$$E\{Y\} = E\{g(X)\} = \int_{-\infty}^{+\infty} g(x)f_X(x)dx
 \tag{4.58}$$

4.3.6 Cas de plusieurs variables aléatoires

Jusqu'à maintenant, nous avons examiné des situations comportant une seule variable aléatoire. Nous allons maintenant considérer le cas de deux variables aléatoires. Le développement réalisé ci-dessous pourra aisément être transposé au cas de plusieurs variables.

Fonctions conjointes et densités marginales

Considérons deux variables aléatoires X et Y . Ces variables peuvent être deux variables séparées à une dimension ou alors une seule variable aléatoire à deux dimensions. Dans les deux cas, l'espace témoin est le plan xy .

Définition 38 La fonction de répartition conjointe $F_{X,Y}(x, y)$ est définie par

$$F_{X,Y}(x, y) = p(X \leq x, Y \leq y) \quad (4.59)$$

c'est-à-dire que la fonction de répartition conjointe de X et Y est égale à la probabilité que la variable X et la variable Y soient respectivement inférieures aux valeurs x et y .

En d'autres mots, la fonction de répartition conjointe est égale à la probabilité que le résultat d'une expérience aléatoire corresponde à un point situé dans le quadrant $(-\infty < X \leq x, -\infty < Y \leq y)$ de l'espace témoin, qui n'est autre que le plan xy .

Propriétés de la fonction de répartition conjointe

1. $F_{X,Y}(x, y)$ est comprise entre 0 et 1.
2. $F_{X,Y}(x, y)$ est une fonction monotone croissante de x et de y .

Définition 39 La fonction de densité de probabilité conjointe $f_{X,Y}(x, y)$ est définie par

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y} \quad (4.60)$$

Propriétés de la fonction de densité de probabilité conjointe

1. $f_{X,Y}(x, y)$ est toujours positive.
2. $f_{X,Y}(x, y)$ possède sous son graphe un volume total égal à 1, c'est-à-dire

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(\xi, \eta) d\xi d\eta = 1 \quad (4.61)$$

La fonction de densité de probabilité d'une des deux variables, par exemple $f_X(x)$, peut être obtenue à partir de la fonction de densité de probabilité conjointe. En effet, il faut tout d'abord noter que

$$F_X(x) = \int_{-\infty}^{+\infty} \int_{-\infty}^x f_{X,Y}(\xi, \eta) d\xi d\eta \quad (4.62)$$

Maintenant, en dérivant des deux côtés l'équation 4.62, nous obtenons la relation désirée

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, \eta) d\eta \quad (4.63)$$

Par le même raisonnement, il serait aisé de déterminer $f_Y(y)$ à partir de $f_{X,Y}(x, y)$. Les fonctions de densité de probabilité $f_X(x)$ et $f_Y(y)$ sont appelées densités marginales. Il vient finalement que la fonction $f_{X,Y}(x, y)$ contient toutes les informations relatives aux variables aléatoires conjointes X et Y .

Moments de deux variables aléatoires

Considérons une paire de variables aléatoires X et Y . Les moments joints des variables X et Y sont donnés par

$$E \{X^i Y^k\} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^i y^k f_{X,Y}(x, y) dx dy \quad (4.64)$$

où i et k sont deux entiers positifs. Le moment le plus important correspond à $i = k = 1$.

Définition 40 [Corrélation] La corrélation de deux variables aléatoires X et Y est définie par

$$E \{XY\} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_{X,Y}(x, y) dx dy \quad (4.65)$$

On peut également évaluer la corrélation des deux variables aléatoires centrées $X - \mu_X$ et $Y - \mu_Y$. On parle alors de covariance.

Définition 41 [Covariance] La covariance de deux variables aléatoires X et Y est définie par

$$C_{XY}(x, y) = E \{(X - \mu_X)(Y - \mu_Y)\} \quad (4.66)$$

$$= E \{XY\} - \mu_X \mu_Y \quad (4.67)$$

Définition 42 Le coefficient de corrélation de deux variables aléatoires est défini par

$$\rho_{X,Y} = \frac{C_{XY}(x, y)}{\sqrt{\sigma_X^2 \sigma_Y^2}} \quad (4.68)$$

Somme

Théorème 43 [22, page 178] L'espérance d'une somme de variables aléatoires est égale à la somme des espérances

$$\mu_{X+Y} = E \{X + Y\} = E \{X\} + E \{Y\} \quad (4.69)$$

Par contre, la variance d'une somme n'est pas égale à la somme des variances en toute généralité comme en atteste le théorème suivant.

Théorème 44 [22, page 179]

$$\sigma_{X+Y}^2 = E \{(X + Y - \mu_{X+Y})^2\} = \sigma_X^2 + \sigma_Y^2 + 2C_{XY}(x, y) \quad (4.70)$$

Probabilité conditionnelle et indépendance

Définition 45 La fonction de densité de probabilité conditionnelle de la variable Y , étant donné que $X = x$, est définie par

$$f_Y(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \quad (4.71)$$

où nous avons supposé que X et Y sont deux variables aléatoires continues et $f_X(x)$ est la densité marginale de X .

$f_Y(y|x)$ peut être vue comme une fonction de la variable y , avec x comme paramètre. De plus, elle satisfait toutes les conditions pour être une fonction de densité de probabilité, c'est-à-dire

$$f_Y(y|x) \geq 0 \quad (4.72)$$

et

$$\int_{-\infty}^{+\infty} f_Y(y|x) dy = 1 \quad (4.73)$$

Propriété 46 Si les variables X et Y sont statistiquement indépendantes, la connaissance du résultat de X n'affecte pas la densité de probabilité de Y et nous pouvons écrire

$$f_Y(y|x) = f_Y(y) \quad (4.74)$$

Et dès lors,

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad (4.75)$$

De même, si la fonction de densité de probabilité conjointe de deux variables aléatoires X et Y est égale au produit de leur densités marginales, alors X et Y sont statistiquement indépendantes. C'est la relation fondamentale pour les calculs dans le cas de variables aléatoires indépendantes.

Définition 47 Par définition, les variables aléatoires X et Y sont dites

$$\text{non - corrélées} \iff C_{XY}(x,y) = 0 \quad (4.76)$$

$$\text{orthogonales} \iff E\{XY\} = 0 \quad (4.77)$$

De l'équation 4.67, nous observons que si X et Y ont des moyennes nulles et si elles sont orthogonales, alors elles sont non corrélées, et vice versa. Il faut noter également que si X et Y sont statistiquement indépendantes, alors elles sont non-corrélées mais la réciproque n'est pas nécessairement vraie. On peut aussi montrer les résultats suivants dans le cas de l'indépendance.

Théorème 48 [22, page 185] Soient deux variables aléatoires X et Y indépendantes,

$$E\{g(X)h(Y)\} = E\{g(X)\}E\{h(Y)\} \quad (4.78)$$

$$C_{XY}(x,y) = 0 \quad (4.79)$$

$$E\{(X+Y)^2\} = \sigma_X^2 + \sigma_Y^2 \quad (4.80)$$

4.4 Processus stochastiques

Jusqu'ici, nous avons parlé d'événements, de réalisations d'événements, de variables aléatoires essentiellement continues. Nous avons vu que le résultat d'une expérience aléatoire pouvait être représenté par un point dans une espace témoin, celui-ci représentant l'ensemble des résultats de l'expérience en question. A chacun de ces points ou ensemble de points (donc à chaque événement), on associait un nom, comme par exemple "pile" qui correspond à l'événement "obtenir pile au jeu du lancer de la pièce", ou un nombre, comme par exemple 6, celui-ci représentant le chiffre situé sur la face supérieure du dé dans le jeu du lancer de dé.

Nous allons maintenant étendre le concept de point de l'espace témoin, en lui associant, non plus un nombre ou un nom, mais une fonction du temps t . Une réalisation d'une expérience aléatoire sera, dans ce cas, une fonction du temps. Nous parlerons alors de processus aléatoires ou stochastiques.

4.4.1 Définition des processus et suites aléatoires

Considérons une expérience aléatoire caractérisée par ses résultats possibles ω de l'espace témoin Ω , ses événements définis sur Ω et les probabilités associées à ces événements. Supposons que, à chaque point ω de l'espace témoin, nous associons une fonction du temps dénotée

$$X(t, \omega) \quad -T \leq t \leq T \quad (4.81)$$

où $2T$ est l'intervalle d'observation total. Pour chaque point fixé ω_j , le graphe de $X(t, \omega_j)$ en fonction du temps t est appelée réalisation du processus aléatoire.

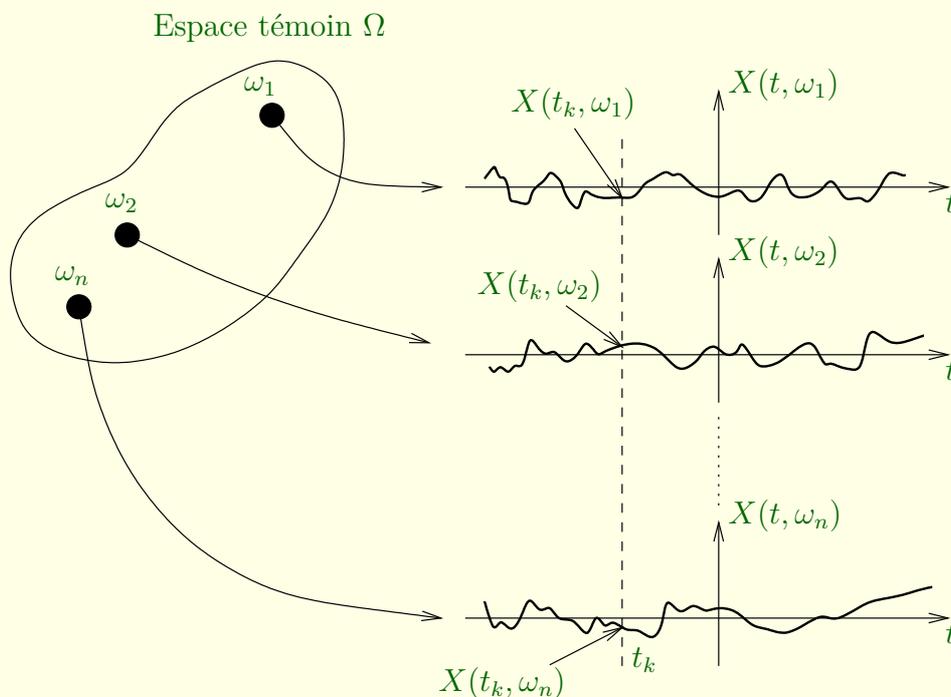


FIGURE 4.5 – Un ensemble de réalisations du processus aléatoire $X(t)$.

La figure 4.5 donne une illustration d'un ensemble de réalisations. De cette figure, nous pouvons noter que, pour un temps t_k fixé dans l'intervalle d'observation, l'ensemble des nombres

$$\{X(t_k, \omega_1), X(t_k, \omega_2), \dots, X(t_k, \omega_n)\} \quad (4.82)$$

constitue une variable aléatoire de dimension n , ou ce qu'on appelle également une suite aléatoire. Il faut remarquer qu'une suite aléatoire peut également être définie par un ensemble quelconque de n variables aléatoires.

Pour simplifier les notations, il est commode de supprimer le ω et d'utiliser $X(t)$ pour dénoter un processus aléatoire. Nous noterons également $x(t)$ une réalisation du processus aléatoire $X(t)$.

4.4.2 Caractéristiques probabilistes

Moments temporels

Considérons un processus aléatoire $X(t)$ et une de ses réalisations $x(t)$, avec un intervalle d'observation défini par $-T \leq t \leq T$. Les deux moments temporels les plus importants d'un processus aléatoire sont la moyenne temporelle et la fonction d'autocorrélation temporelle.

Définition 49 *La moyenne temporelle d'un processus aléatoire $X(t)$ est définie par*

$$\mu_X(T) = \frac{1}{2T} \int_{-T}^T x(t) dt \quad (4.83)$$

où $x(t)$ est une réalisation du processus aléatoire.

Clairement, $\mu_X(T)$ est une variable aléatoire, avec une moyenne $E\{\mu_X(T)\}$ et une variance $var\{\mu_X(T)\}$, car sa valeur dépend de l'intervalle d'observation et de la réalisation particulière $x(t)$ du processus aléatoire $X(t)$ utilisée dans l'équation 4.83.

Définition 50 *La fonction d'autocorrélation temporelle d'un processus aléatoire $X(t)$ est définie par*

$$\Gamma_{XX}(\tau, T) = \frac{1}{2T} \int_{-T}^T x(t + \tau)x(t) dt \quad (4.84)$$

où $x(t)$ est une réalisation du processus aléatoire.

À nouveau, il est clair que, pour τ fixé, $\Gamma_{XX}(\tau, T)$ peut être vue comme une variable aléatoire, avec une moyenne et une variance.

Moments statistiques

Les moments statistiques les plus importants d'un processus aléatoire $X(t)$ sont la moyenne et la fonction d'autocorrélation.

Définition 51 *La moyenne d'un processus aléatoire $X(t)$ observé au temps t est définie par*

$$\mu_X(t) = E\{X(t)\} \quad (4.85)$$

$$= \int_{-\infty}^{+\infty} x f_{X(t)}(x) dx \quad (4.86)$$

où $f_{X(t)}(x)$ est la fonction de densité de probabilité du premier ordre du processus aléatoire $X(t)$.

Pour t fixé, $X(t)$ est une variable aléatoire qui est caractérisée par sa fonction de densité de probabilité $f_{X(t)}(x)$.

Définition 52 La fonction d'autocorrélation d'un processus aléatoire $X(t)$ est définie par

$$\Gamma_{XX}(t_1, t_2) = E \{X(t_1)X(t_2)\} \quad (4.87)$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_1 x_2 f_{X(t_1), X(t_2)}(x_1, x_2) dx_1 dx_2 \quad (4.88)$$

où $f_{X(t_1), X(t_2)}(x_1, x_2)$ est la fonction de densité de probabilité du second ordre du processus aléatoire $X(t)$.

Pour t_1 et t_2 fixés, $X(t_1)$ et $X(t_2)$ sont deux variables aléatoires caractérisées par leur fonction de densité de probabilité conjointe $f_{X(t_1), X(t_2)}(x_1, x_2)$.

Fonctions d'intercorrélacion

Considérons deux processus aléatoires $X(t)$ et $Y(t)$, avec leur propres fonctions d'autocorrélation $\Gamma_{XX}(t, u)$ et $\Gamma_{YY}(t, u)$.

Définition 53 Les deux fonctions d'intercorrélacion de deux processus aléatoires $X(t)$ et $Y(t)$ sont définies par

$$\Gamma_{XY}(t, u) = E \{X(t)Y(u)\} \quad (4.89)$$

$$\Gamma_{YX}(t, u) = E \{Y(t)X(u)\} \quad (4.90)$$

où t et u sont deux temps auxquels les processus ont été observés.

4.4.3 Stationnarité

Stationnarité au sens strict

Considérons un processus aléatoire $X(t)$ qui est initié en $t = -\infty$. Soient $X(t_1), X(t_2), \dots, X(t_k)$ les variables aléatoires obtenues en observant le processus aléatoire $X(t)$ aux temps t_1, t_2, \dots, t_k respectivement. La fonction de répartition de probabilité conjointe de cet ensemble de variables aléatoires est notée

$$F_{X(t_1), X(t_2), \dots, X(t_k)}(x_1, x_2, \dots, x_k) \quad (4.91)$$

Supposons maintenant que nous décalions tous les instants d'observation d'un temps τ . Nous obtenons alors les variables aléatoires $X(t_1 + \tau), X(t_2 + \tau), \dots, X(t_k + \tau)$. La fonction de répartition conjointe de ces nouvelles variables aléatoires est notée

$$F_{X(t_1+\tau), X(t_2+\tau), \dots, X(t_k+\tau)}(x_1, x_2, \dots, x_k) \quad (4.92)$$

Le processus aléatoire $X(t)$ sera dit stationnaire au sens strict si

$$F_{X(t_1), X(t_2), \dots, X(t_k)}(x_1, x_2, \dots, x_k) = F_{X(t_1+\tau), X(t_2+\tau), \dots, X(t_k+\tau)}(x_1, x_2, \dots, x_k) \quad (4.93)$$

pour tout τ , tout k et tous les choix possibles des temps d'observation t_1, t_2, \dots, t_k .

Définition 54 *Un processus aléatoire est stationnaire au sens strict si ses caractéristiques probabilistes sont invariantes pour tout changement de l'origine des temps.*

De l'équation 4.93, nous pouvons tirer deux propriétés particulièrement intéressantes :

1. Pour $k = 1$, nous avons

$$F_{X(t)}(x) = F_{X(t+\tau)}(x) = F_X(x) \quad \forall t, \tau \quad (4.94)$$

c'est-à-dire que la fonction de répartition de probabilité du premier ordre $F_{X(t)}(x)$ d'un processus aléatoire $X(t)$, stationnaire au sens strict, est indépendante du temps.

2. Pour $k = 2$ et $\tau = -t_1$, nous avons

$$F_{X(t_1), X(t_2)}(x_1, x_2) = F_{X(0), X(t_2-t_1)}(x_1, x_2) \quad \forall t_1, t_2 \quad (4.95)$$

c'est-à-dire que la fonction de répartition du second ordre $F_{X(t_1), X(t_2)}(x_1, x_2)$ d'un processus aléatoire $X(t)$, stationnaire au sens strict, dépend uniquement de la différence entre les temps d'observation et non des temps particuliers où le processus aléatoire est observé.

Stationnarité au sens large

Considérons un processus aléatoire $X(t)$, stationnaire au sens strict. Dans ce cas, nous savons que sa fonction de densité de probabilité du premier ordre $f_{X(t)}(x)$ est indépendante du temps. Dès lors, sa moyenne $\mu_X(t)$, définie en 4.86, se réduit à

Proposition 55 [*Moyenne d'un processus stationnaire au sens large*]

$$\mu_X(t) = \mu_X \quad \forall t \quad (4.96)$$

Évaluons maintenant la fonction d'autocorrélation de $X(t)$. Nous savons que sa fonction de densité de probabilité du second ordre $f_{X(t_1), X(t_2)}(x_1, x_2)$ ne dépend que de la différence entre les temps d'observation t_1 et t_2 . Dès lors, sa fonction d'autocorrélation $\Gamma_{XX}(t_1, t_2)$ ne dépend que de la différence $t_2 - t_1$, comme le montre

Proposition 56 [*Fonction d'autocorrélation d'un processus stationnaire au sens large*]

$$\Gamma_{XX}(t_1, t_2) = \Gamma_{XX}(t_2 - t_1) \quad \forall t_1, t_2 \quad (4.97)$$

Il est très important de noter que les conditions 4.96 et 4.97 ne suffisent pas pour garantir que le processus aléatoire $X(t)$ soit stationnaire au sens strict. Cependant, des considérations pratiques nous dictent que nous pouvons simplement nous contenter de la moyenne et de la fonction d'autocorrélation pour décrire, même partiellement, le processus aléatoire. Dès lors, nous définissons une classe de stationnarité moins large que la stationnarité stricte.

Définition 57 *Un processus aléatoire est stationnaire au sens large s'il vérifie les deux conditions 4.96 et 4.97, c'est-à-dire si*

- sa moyenne est indépendante du temps, et
- sa fonction d'autocorrélation ne dépend que de la différence entre les temps d'observation.

Il évident que, si un processus aléatoire est stationnaire au sens strict, il est automatiquement stationnaire au sens large. Par contre, la réciproque n'est pas vraie. Si rien n'est précisé lorsque l'on parle de processus aléatoire stationnaire, il faut considérer qu'il s'agit d'un processus aléatoire stationnaire au sens large.

Fonction d'autocorrélation d'un processus aléatoire stationnaire

Pour simplifier les notations, nous allons redéfinir la fonction d'autocorrélation d'un processus stationnaire $X(t)$:

$$\Gamma_{XX}(\tau) = E\{X(t+\tau)X(t)\} \quad \forall t \quad (4.98)$$

où τ représente la différence entre les deux temps d'observation du processus aléatoire. Cette fonction d'autocorrélation possède plusieurs propriétés importantes :

– la moyenne du carré du processus aléatoire peut être obtenue simplement en posant $\tau = 0$ dans 4.98 :

$$\Gamma_{XX}(0) = E\{X^2(t)\} \quad (4.99)$$

– la fonction d'autocorrélation $\Gamma_{XX}(\tau)$ est une fonction paire de τ :

$$\Gamma_{XX}(\tau) = \Gamma_{XX}(-\tau) \quad (4.100)$$

– la fonction d'autocorrélation $\Gamma_{XX}(\tau)$ présente son amplitude maximum en $\tau = 0$:

$$|\Gamma_{XX}(\tau)| \leq \Gamma_{XX}(0) \quad (4.101)$$

Donnons à présent une interprétation physique de $\Gamma_{XX}(\tau)$: la fonction d'autocorrélation donne un moyen de décrire l'interdépendance deux variables aléatoires obtenues en observant le processus aléatoire $X(t)$ à des instants séparés de τ secondes. Il est clair que, plus le processus aléatoire $X(t)$ fluctue rapidement avec le temps, plus la fonction d'autocorrélation $\Gamma_{XX}(\tau)$ décroît rapidement de sa valeur maximum $\Gamma_{XX}(0)$ lorsque τ augmente, comme le montre la figure 4.6.

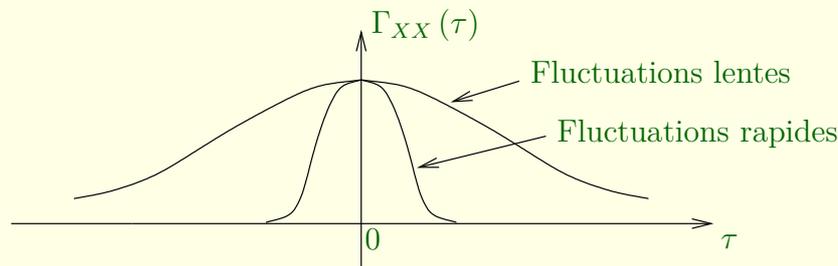


FIGURE 4.6 – Illustration de la fonction d'autocorrélation d'un processus aléatoire avec fluctuations lentes et rapides.

4.4.4 Intérêt de la stationnarité

La stationnarité est un outil commode puisque permettant d'ignorer le temps d'échantillonnage. C'est la raison pour laquelle bon nombre de développements partent de cette hypothèse. Par exemple, une source productrice d'images est modélisée comme génératrice d'un processus stationnaire. Or, rien n'est moins sûr. Dans le cas de la modélisation de canaux de transmission pour des communications mobiles, l'hypothèse de la stationnarité des caractéristiques du canal ne peut pas être retenue car l'expérience la contredit.

4.4.5 Ergodicité

Nous sommes maintenant intéressés à relier les moments statistiques d'un processus aléatoire stationnaire à ses moments temporels. En effet, les moments temporels nous donneraient un moyen aisé de déterminer les moments statistiques du processus. La question qui se pose est donc : quand pouvons-nous remplacer un moment temporel par un moment statistique ?

Ergodicité dans la moyenne

Considérons un processus aléatoire stationnaire $X(t)$, avec un intervalle d'observation défini par $-T \leq t \leq T$. Comme nous l'avons vu plus haut, la moyenne temporelle $\mu_X(T)$ de $X(t)$ peut être vue comme une variable aléatoire. Évaluons-en l'espérance statistique

$$E \{ \mu_X(T) \} = \frac{1}{2T} \int_{-T}^T E \{ x(t) \} dt \quad (4.102)$$

$$= \frac{1}{2T} \int_{-T}^T \mu_X dt \quad (4.103)$$

$$= \mu_X \quad (4.104)$$

où nous avons utilisé le fait que le processus $X(t)$ est stationnaire au sens large. Nous voyons donc que la moyenne temporelle de $\mu_X(T)$ représente une estimation de la moyenne statistique μ_X .

Définition 58 *Un processus aléatoire stationnaire $X(t)$ est dit ergodique dans la moyenne si :*
 (1) *la moyenne temporelle $\mu_X(T)$ tend vers la moyenne statistique μ_X lorsque l'intervalle d'observation $2T$ tend vers l'infini, c'est-à-dire*

$$\lim_{T \rightarrow \infty} \mu_X(T) = \mu_X \quad (4.105)$$

(2) *la variance de la variable aléatoire $\mu_X(T)$ tend vers zéro lorsque l'intervalle d'observation $2T$ tend vers l'infini, c'est-à-dire*

$$\lim_{T \rightarrow \infty} \text{var} \{ \mu_X(T) \} = 0 \quad (4.106)$$

Ergodicité dans la fonction d'autocorrélation

À nouveau, considérons un processus aléatoire stationnaire $X(t)$, avec un intervalle d'observation défini par $-T \leq t \leq T$. Pour τ fixé, la fonction d'autocorrélation temporelle $\Gamma_{XX}(\tau, T)$ peut être vue comme une variable aléatoire dont nous allons évaluer l'espérance statistique :

$$E \{ \Gamma_{XX}(\tau, T) \} = \frac{1}{2T} \int_{-T}^T E \{ x(t+\tau)x(t) \} dt \quad (4.107)$$

$$= \frac{1}{2T} \int_{-T}^T \Gamma_{XX}(\tau) dt \quad (4.108)$$

$$= \Gamma_{XX}(\tau) \quad (4.109)$$

où nous avons utilisé le fait que le processus $X(t)$ est stationnaire au sens large. De nouveau, comme dans le cas de la moyenne, nous voyons que la fonction d'autocorrélation temporelle $\Gamma_{XX}(\tau, T)$ représente une estimation de la fonction d'autocorrélation statistique $\Gamma_{XX}(\tau)$.

Définition 59 Un processus aléatoire stationnaire $X(t)$ sera dit ergodique dans la fonction d'autocorrélation si les deux conditions suivantes sont remplies :

$$\lim_{T \rightarrow \infty} \Gamma_{XX}(\tau, T) = \Gamma_{XX}(\tau) \quad (4.110)$$

$$\lim_{T \rightarrow \infty} \text{var} \{ \Gamma_{XX}(\tau, T) \} = 0 \quad (4.111)$$

4.4.6 Densité spectrale de puissance

Définition 60 La densité spectrale de puissance $\gamma_X(f)$ d'un processus aléatoire stationnaire $X(t)$ est la transformée de FOURIER de sa fonction d'autocorrélation $\Gamma_{XX}(\tau)$:

$$\gamma_X(f) = \int_{-\infty}^{+\infty} \Gamma_{XX}(\tau) e^{-2\pi j f \tau} d\tau \quad (4.112)$$

$$\Gamma_{XX}(\tau) = \int_{-\infty}^{+\infty} \gamma_X(f) e^{2\pi j f \tau} df \quad (4.113)$$

Ces relations sont appelées relations d'EINSTEIN-WIENER-KINTCHINE. Voici quelques propriétés importantes de la densité spectrale de puissance :

- la valeur à l'origine de la densité spectrale de puissance d'un processus aléatoire stationnaire est égale à l'aire sous le graphe de la fonction d'autocorrélation :

$$\gamma_X(0) = \int_{-\infty}^{+\infty} \Gamma_{XX}(\tau) d\tau \quad (4.114)$$

- la valeur moyenne du carré d'un processus aléatoire stationnaire est égale à l'aire sous le graphe de la densité spectrale de puissance :

$$E \{ X^2(t) \} = \Gamma_{XX}(0) = \int_{-\infty}^{+\infty} \gamma_X(f) df \quad (4.115)$$

Cette relation est importante car elle indique que la puissance d'un signal aléatoire, qui n'est autre que $E \{ X^2(t) \}$, peut se calculer par l'intégrale de la densité spectrale de puissance. C'est bien entendu à cette propriété fondamentale que la fonction $\gamma_X(f)$ doit le nom de densité spectrale de puissance.

- la densité spectrale de puissance d'un processus aléatoire stationnaire à valeurs réelles est une fonction paire de la fréquence :

$$\gamma_X(-f) = \gamma_X(f) \quad (4.116)$$

Remarquons que la transformée de FOURIER d'un signal aléatoire est elle-même aléatoire, et de ce fait, on ne peut pas la calculer. On peut par contre déterminer sa densité spectrale de puissance car cette fonction est déterministe –on suppose bien sûr la stationnarité au sens large ainsi que l'ergodicité de $X(t)$.

Exemple : onde sinusoïdale avec phase aléatoire.

Considérons un signal sinusoïdal avec phase aléatoire, défini par

$$X(t) = A_c \cos(2\pi f_c t + \Theta) \quad (4.117)$$

où A_c et f_c sont constants et Θ est une variable aléatoire uniformément distribuée sur l'intervalle $[-\pi, +\pi]$, c'est-à-dire

$$f_{\Theta}(\theta) = \begin{cases} \frac{1}{2\pi} & -\pi \leq \theta \leq +\pi \\ 0 & \text{sinon} \end{cases} \quad (4.118)$$

Calculons tout d'abord la moyenne statistique du processus aléatoire $X(t)$:

$$\mu_X(t) = E\{X(t)\} \quad (4.119)$$

$$= \int_{-\pi}^{+\pi} A_c \cos(2\pi f_c t + \theta) \frac{1}{2\pi} d\theta \quad (4.120)$$

$$= 0 \quad (4.121)$$

La moyenne statistique du processus aléatoire $X(t)$ est donc indépendante du temps. Évaluons maintenant sa fonction d'autocorrélation :

$$\Gamma_{XX}(t_1, t_2) = E\{X(t_1)X(t_2)\} \quad (4.122)$$

$$= E\{A_c^2 \cos(2\pi f_c t_1 + \Theta) \cos(2\pi f_c t_2 + \Theta)\} \quad (4.123)$$

$$= \frac{A_c^2}{2} E\{\cos(2\pi f_c(t_2 - t_1))\} + \frac{A_c^2}{2} E\{\cos(2\pi f_c(t_2 + t_1) + 2\Theta)\}$$

$$= \frac{A_c^2}{2} \cos(2\pi f_c(t_2 - t_1)) \quad (4.124)$$

Il apparaît que la fonction d'autocorrélation du processus aléatoire $X(t)$ dépend exclusivement de la différence entre les temps d'observation $\tau = t_2 - t_1$. Nous pouvons donc écrire

$$\Gamma_{XX}(\tau) = \frac{A_c^2}{2} \cos(2\pi f_c \tau) \quad (4.125)$$

Étant donné que la moyenne statistique de $X(t)$ ne dépend pas du temps et que sa fonction d'autocorrélation ne dépend que de la différence entre les temps d'observation, nous pouvons conclure que le processus aléatoire $X(t)$ est stationnaire au sens large. Évaluons à présent la densité spectrale de puissance de $X(t)$:

$$\gamma_X(f) = \frac{A_c^2}{4} [\delta(f - f_c) + \delta(f + f_c)] \quad (4.126)$$

qui est constituées de deux raies situées en $f = \pm f_c$ et pondérées par le facteur $A_c^2/4$. Notons que l'aire totale sous le graphe de $\gamma_X(f)$ est égale à $A_c^2/2$, ce qui correspond bien à $E\{X^2(t)\} = \Gamma_{XX}(0)$.

4.4.7 Lien entre stationnarité et ergodicité

Les notions de stationnarité et d'ergodicité sont distinctes. Ainsi un processus aléatoire peut être stationnaire et non ergodique, et inversement. Toutefois, les processus qui sont à la fois stationnaires et ergodiques ont une propriété remarquable ; on montre alors qu'il y a presque sûrement identité entre moyennes temporelles et moyennes statistiques. On en déduit une conséquence expérimentale très importante : l'ensemble des propriétés statistiques d'un processus aléatoire stationnaire et ergodique peut être obtenu à partir de l'observation d'une seule et unique trajectoire allant de $t = -\infty$ à $t = +\infty$.

4.5 Processus stochastiques et systèmes

4.5.1 Filtrage des processus aléatoires stationnaires au sens large

Soit $X(t)$ un processus aléatoire stationnaire du second ordre, de moyenne μ_X , de fonction d'autocorrélation $\Gamma_{XX}(\tau)$ et de densité spectrale de puissance $\gamma_X(f)$.

Soit le filtre linéaire de réponse impulsionnelle $h(t)$ et de gain complexe $\mathcal{H}(f)$

$$\mathcal{H}(f) = \int_{-\infty}^{+\infty} h(t)e^{-2\pi jft} dt \quad (4.127)$$

En sortie, on a

$$y(t) = h(t) \otimes x(t) = \int_{-\infty}^{+\infty} h(u)x(t-u)du \quad (4.128)$$

Il est intéressant de déterminer la moyenne ainsi que la fonction d'autocorrélation du processus aléatoire $Y(t)$. Intéressons-nous tout d'abord à la moyenne.

$$\mu_Y = E\{Y(t)\} \quad (4.129)$$

$$= E\left\{\int_{-\infty}^{+\infty} h(u)X(t-u)du\right\} \quad (4.130)$$

$$= \int_{-\infty}^{+\infty} h(u)E\{X(t-u)\}du \quad (4.131)$$

$$= \mu_X \int_{-\infty}^{+\infty} h(u)du \quad (4.132)$$

$$= \mu_X \mathcal{H}(0) \quad (4.133)$$

où $\mathcal{H}(0)$ est la réponse DC ($f=0$) du système. Ce résultat est résumé comme suit :

Proposition 61 Soient $X(t)$ un processus aléatoire stationnaire du second ordre de moyenne μ_X , $\mathcal{H}(f)$ la transmittance d'un filtre linéaire et $Y(t)$ le processus à la sortie du système linéaire

$$\mu_Y = \mu_X \mathcal{H}(0) \quad (4.134)$$

Calculons à présent la fonction d'autocorrélation de $Y(t)$:

$$\Gamma_{YY}(t_1, t_2) = E\{Y(t_1)Y(t_2)\} \quad (4.135)$$

$$= E\left\{\int_{-\infty}^{+\infty} h(u_1)X(t_1-u_1)du_1 \int_{-\infty}^{+\infty} h(u_2)X(t_2-u_2)du_2\right\} \quad (4.136)$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(u_1)h(u_2)E\{X(t_1-u_1)X(t_2-u_2)\}du_1du_2 \quad (4.137)$$

$$= \int_{-\infty}^{+\infty} h(u_1) \int_{-\infty}^{+\infty} h(u_2)\Gamma_{XX}(t_1-u_1, t_2-u_2)du_1du_2 \quad (4.138)$$

Comme le processus $X(t)$ est stationnaire, en posant $\tau = t_1 - t_2$, la fonction d'autocorrélation du processus de sortie vaut

$$\Gamma_{YY}(\tau) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(u_1)h(u_2)\Gamma_{XX}(\tau - u_1 + u_2)du_1du_2 \quad (4.139)$$

Clairement, le processus aléatoire $Y(t)$ est stationnaire au sens large vu que sa moyenne est indépendante du temps et que sa fonction d'autocorrélation ne dépend que de l'intervalle de temps entre les instants d'observation t_1 et t_2 . Il reste maintenant à examiner la densité spectrale de puissance de $Y(t)$.

$$\gamma_Y(f) = \int_{-\infty}^{+\infty} \Gamma_{YY}(\tau) e^{-2\pi j f \tau} d\tau \quad (4.140)$$

$$= \int \int \int_{\mathbb{R}^3} h(u_1)h(u_2)\Gamma_{XX}(\tau - u_1 + u_2) e^{-2\pi j f \tau} d\tau du_1 du_2 \quad (4.141)$$

On réalise ensuite le changement de variable $\tau = u_0 + u_1 - u_2$ et finalement il vient

$$\gamma_Y(f) = \mathcal{H}(f)\mathcal{H}^*(f)\gamma_X(f) \quad (4.142)$$

$$= \|\mathcal{H}(f)\|^2 \gamma_X(f) \quad (4.143)$$

Cette dernière relation indique que la densité spectrale de puissance du processus aléatoire $Y(t)$ est égale à la densité spectrale de $X(t)$ multipliée par le module de la fonction de transfert du filtre au carré.

Ce résultat est fourni sous la forme de la proposition suivante.

Proposition 62 [Wiener-Kintchine] *Soient $X(t)$ un processus stochastique stationnaire au sens large, $\mathcal{H}(f)$ la transmittance d'un filtre linéaire et $Y(t)$ le processus à la sortie du système linéaire*

$$\gamma_Y(f) = \|\mathcal{H}(f)\|^2 \gamma_X(f) \quad (4.144)$$

4.5.2 Somme de processus stochastiques

Pour être en mesure de déterminer la puissance du signal modulé, il nous faut examiner le comportement d'une somme de processus stochastiques. Soient $X(t)$ et $N(t)$ deux processus stochastiques stationnaires au sens large, de moyenne nulle. Considérons le processus obtenu par sommation

$$Y(t) = X(t) + N(t) \quad (4.145)$$

La densité spectrale de puissance de $Y(t)$ nécessite le calcul de sa fonction d'autocorrélation.

$$\Gamma_{YY}(t_1, t_2) = E\{Y(t_1)Y(t_2)\} \quad (4.146)$$

$$= E\{(X(t_1) + N(t_1))(X(t_2) + N(t_2))\} \quad (4.147)$$

$$= E\{X(t_1)X(t_2)\} + E\{X(t_1)N(t_2)\} \\ + E\{N(t_1)X(t_2)\} + E\{N(t_1)N(t_2)\} \quad (4.148)$$

$$= \Gamma_{XX}(t_1, t_2) + \Gamma_{XN}(t_1, t_2) + \Gamma_{NX}(t_1, t_2) + \Gamma_{NN}(t_1, t_2) \quad (4.149)$$

On effectue ensuite le changement de variables $\tau = t_1 - t_2$, sachant que les processus sont stationnaires au sens large. Et donc

$$\Gamma_{YY}(\tau) = \Gamma_{XX}(\tau) + \Gamma_{XN}(\tau) + \Gamma_{NX}(\tau) + \Gamma_{NN}(\tau) \quad (4.150)$$

Finalement, la transformée de FOURIER des membres de cette relation fournit la correspondance entre les densités spectrales de puissance

$$\gamma_{YY}(f) = \gamma_{XX}(f) + \gamma_{XN}(f) + \gamma_{NX}(f) + \gamma_{NN}(f) \quad (4.151)$$

Si les processus sont non corrélés,

$$\gamma_{YY}(f) = \gamma_{XX}(f) + \gamma_{NN}(f) \quad (4.152)$$

4.5.3 Modulation : mélangeur pour signal stochastique

Soit un processus stochastique $M(t)$, stationnaire au sens large, que l'on fait passer à travers un mélangeur $\cos(2\pi f_c t + \Phi)$, Φ étant une phase uniformément distribuée sur l'intervalle $[0, 2\pi]$.

On cherche à déterminer la densité spectrale de puissance du processus aléatoire défini par

$$S(t) = M(t) \cos(2\pi f_c t + \Phi) \quad (4.153)$$

La fonction d'autocorrélation vaut

$$\Gamma_{SS}(\tau) = E\{S(t+\tau)S(t)\} \quad (4.154)$$

$$= E\{M(t+\tau) \cos(2\pi f_c t + 2\pi f_c \tau + \Phi) M(t) \cos(2\pi f_c t + \Phi)\} \quad (4.155)$$

$$= E\{M(t+\tau)M(t)\} E\{\cos(2\pi f_c t + 2\pi f_c \tau + \Phi) \cos(2\pi f_c t + \Phi)\} \quad (4.156)$$

$$= \frac{1}{2} \Gamma_{MM}(\tau) E\{\cos(2\pi f_c \tau) + \cos(4\pi f_c t + 4\pi f_c \tau + 2\Phi)\} \quad (4.157)$$

$$= \frac{1}{2} \Gamma_{MM}(\tau) \cos(2\pi f_c \tau) \quad (4.158)$$

D'où

$$\gamma_S(f) = \frac{1}{4} [\gamma_M(f - f_c) + \gamma_M(f + f_c)] \quad (4.159)$$

4.6 Une classe de processus stochastiques importante : les sources de bruit

En toute généralité, on appelle bruit tout signal qui se superpose au signal utile. De par sa fonction, un système de télécommunications est particulièrement sensible au bruit et nombreuses sont les sources influençant la qualité de la transmission. Cependant, lors d'une transmission, beaucoup de phénomènes peuvent détériorer le signal, et un signal pour un observateur peut être un bruit pour un autre. C'est le cas pour des techniques de spectre étalé.

Beaucoup de dispositifs sont générateurs de bruit : tout d'abord, il y a les organes électroniques de transmission et de réception (amplificateur non-linéaire), puis le canal de transmission (bande passante limitée, linéarité, ...), ensuite il y a les bruits artificiels qui s'ajoutent au signal (bruits dû au 50 [Hz], bruit diaphonique, ...) et enfin il y a les bruits d'origine microscopique (bruit thermique, bruit de grenaille, ...). On comprend dès lors que le bruit joue un rôle essentiel dans la conception d'un système de télécommunications.

C'est la catégorie des bruits d'origine microscopique qui nous intéressera ici.

4.6.1 Bruit thermique

Ce bruit, encore appelé JOHNSON noise, est dû à l'agitation des électrons dans un conducteur, sous l'effet de l'activation thermique. On sait que cette agitation croît avec la température ; il en résultera un terme de bruit proportionnel à la température.

Pour des raisons pratiques, on modélise le bruit thermique par la valeur quadratique moyenne de la tension V_{BT} de bruit thermique qui apparaît aux bornes d'une résistance, mesurée dans la bande de fréquences $\Delta f [Hz]$, telle que fournie par

$$E\{V_{BT}^2\} = 4k_B T R \Delta f [V^2] \quad (4.160)$$

où k_B est la constante de BOLTZMANN, égale à $1,38 \times 10^{-23}$ joules par degré KELVIN, T la température exprimée en KELVIN et R la valeur de la résistance. L'équivalent de THÉVENIN de la résistance (cf. figure 4.7) soumise au bruit thermique consiste à mettre en série la résistance et une force électromotrice de valeur quadratique moyenne $E\{V_{BT}^2\}$.

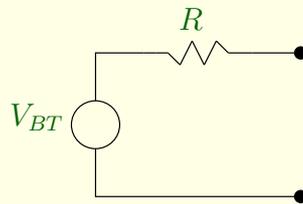


FIGURE 4.7 – Équivalent de Thévenin d'une source de bruit thermique.

En raison de la présence d'un très grand nombre d'électrons dont le mouvement est indépendant, le théorème de la limite centrale conduit à considérer le bruit thermique comme gaussien de moyenne nulle ; d'où une densité de probabilité de la forme

$$p(x) = \frac{1}{\sqrt{2\pi}V_{BT}} e^{-\frac{x^2}{V_{BT}^2}} \quad (4.161)$$

Pour déterminer la puissance de bruit produite par la résistance, il faut considérer l'adaptation de la source de THÉVENIN, flanquée de sa résistance interne, par rapport à la résistance de la charge connectée à ses bornes. Le transfert maximum de puissance a lieu pour une résistance de charge égale à la résistance interne. Dans ce cas, on détermine une puissance de bruit disponible de $k_B T \Delta f [W]$.

Ordre de grandeur

À la température ambiante ($T = 300 [K]$), la puissance de bruit disponible vaut $4,14 \times 10^{-21} \Delta f [W]$. Aux bornes d'une résistance de $1 [k\Omega]$ et pour une bande de fréquences de $1 [GHz]$, elle correspond à une tension de $\sqrt{4,14 \times 10^{-21} R \Delta f} = 64 [\mu V]$. Cette tension est négligeable pour des systèmes numériques en raison d'autres effets de l'ordre du millivolt. Il en va autrement pour des récepteurs radio qui fonctionnent avec des niveaux de tension de cet ordre. Il en résulte qu'on ne peut pas négliger les effets des bruits thermiques lors de la conception des systèmes radio ou de l'analyse des performances des techniques de modulation continue.

4.6.2 Bruit de grenaille

Ce bruit, encore appelé shot noise, est dû à la fluctuation du courant lors de la traversée d'une barrière de potentiel par les électrons. On le retrouve donc dans tous les semi-conducteurs. Soit un courant I , le nombre de porteurs de charge arrivant au temps t est

$$N = \frac{It}{q} \quad (4.162)$$

où $q = 1,6 \times 10^{-19} [C]$ est la charge d'un électron. L'entier N fluctue légèrement en fonction du temps, d'où il résulte une variation de la valeur instantanée du courant. L'examen analytique (complexe) de ce problème conduit à considérer en pratique que la fluctuation de courant due au bruit a l'allure d'un bruit blanc gaussien pour lequel

$$E\{I_{Gr}^2\} = 2qI\Delta f [A^2] \quad (4.163)$$

Par exemple, pour un courant valant $1 [mA]$ et pour une bande de fréquences de $1 [GHz]$, $I_{Gr} = 566 [nA]$ ce qui équivaut à $28 [\mu V]$ à travers une résistance de $50 [\Omega]$, sans considérer le taux de transfert maximum d'énergie.

4.6.3 Modélisation du bruit

La plupart des processus physiques réels, bruits ou autre, sont réellement aléatoires ou sont considérés comme tels. C'est donc par le biais des probabilités qu'il convient d'analyser les bruits.

Comme la puissance utilisée dans un système de télécommunications et sa gamme de fréquences le caractérisent, c'est sans surprise en terme de puissance de bruit que s'est développée l'étude théorique, d'autant que la densité spectrale d'un processus stochastique représentant le bruit, stationnaire au sens large et ergodique, permet d'en calculer la puissance. La formule pratique de la puissance de bruit thermique ne fait que confirmer ce fait.

Processus gaussien

Les processus aléatoires gaussiens jouent un rôle fondamental en pratique. La loi de GAUSS s'introduit en effet naturellement dans de nombreux phénomènes physiques, qui se présentent macroscopiquement comme la somme de nombreux phénomènes microscopiques, via le théorème de la limite centrale.

Définition 63 *Un processus aléatoire est dit gaussien si quelle que soit la suite des instants t_1, t_2, \dots, t_n et quel que soit n , la variable aléatoire à n dimensions X_1, X_2, \dots, X_n est gaussienne.*

Propriétés

Les propriétés dont jouit la gaussienne permet également d'établir une série de résultats utiles pour les simplifications qu'ils apportent. En particulier, les propriétés suivantes sont établies [4] :

- Un processus aléatoire gaussien est entièrement défini par la donnée des moments du premier et du second ordre.

- Un processus aléatoire gaussien qui serait stationnaire au sens large est aussi stationnaire au sens strict, puisque la loi temporelle est entièrement caractérisée par ses deux premiers moments.
- Si les variables aléatoires X_1, X_2, \dots, X_n , obtenues par échantillonnage d'un processus gaussien aux temps t_1, t_2, \dots, t_n sont non corrélées, c'est-à-dire que

$$E\{(X_k - \mu_{X_k})(X_i - \mu_{X_i})\} = 0, \quad i \neq k \quad (4.164)$$

alors ces variables sont statistiquement indépendantes. L'implication de cette propriété est que la fonction de densité de probabilité jointe des variables aléatoires X_1, X_2, \dots, X_n , s'exprime comme le produit des densités de probabilité respectives.

- Il est aisé de montrer que toute combinaison linéaire finie de variables aléatoires gaussiennes est gaussienne. De façon plus générale, on montre que le caractère gaussien se conserve dans toute transformation linéaire et donc en particulier par filtrage linéaire.

4.6.4 Bruit blanc

Parmi les processus aléatoires, le bruit blanc revêt une importance particulière parce qu'il représente l'archétype de beaucoup de phénomènes physiques.

Définition 64 *Le bruit blanc est un processus aléatoire stationnaire du second ordre centré et dont la densité spectrale de puissance est constante sur tout l'axe des fréquences³ :*

$$\gamma_W(f) = \frac{N_0}{2} \quad (4.165)$$

N_0 s'exprime en Watt par Hertz. La fonction d'autocorrélation s'écrit donc

$$\Gamma_{WW}(\tau) = \frac{N_0}{2} \delta(\tau) \quad (4.166)$$

Elle présente une raie à l'origine et est nulle partout ailleurs. Dès lors, deux variables aléatoires distinctes résultant de l'échantillonnage d'un processus stochastique blanc, quelle que soit l'écart temporel, sont toujours non corrélées.

Par analogie avec le bruit thermique, il arrive que la densité spectrale soit décrite par une température de bruit équivalente T_e , comme si le bruit affectait l'étage d'entrée d'un récepteur. Cette température se déduit de

$$N_0 = k_B T_e \quad (4.167)$$

Passage du bruit blanc à travers un filtre passe-bande

L'expression 4.165 montre que le processus est de puissance infinie. Toutefois, comme les systèmes rencontrés en pratique sont à bande de fréquence limitée, on modélise souvent le bruit par un processus réel, dont la densité spectrale de puissance est égale à $N_0/2$ dans cette bande et nulle partout ailleurs; pour une largeur de bande B , sa puissance est alors égale à $N_0 B$.

3. Le terme blanc prend son origine dans l'analogie avec la lumière blanche, dont la puissance est répartie uniformément sur l'axe des fréquences.

Cette forme bien commode explique pourquoi on définit une densité spectrale bilatérale faisant intervenir un facteur 1/2.

Prenons le cas d'un bruit blanc de type passe-bande centré autour de la fréquence f_c . Sa description est

$$\gamma_W(f) = \begin{cases} \frac{N_0}{2}, & f \in [f_c - \frac{B}{2}, f_c + \frac{B}{2},] \text{ ou } f \in [-f_c - \frac{B}{2}, -f_c + \frac{B}{2},] \\ 0, & \text{sinon} \end{cases} \quad (4.168)$$

La fonction d'autocorrélation est la transformée de FOURIER inverse de la densité spectrale d'où, en utilisant le fait que la densité spectrale est paire,

$$\Gamma_{WW}(\tau) = \int_{-\infty}^{+\infty} \gamma_W(f) e^{2\pi j f \tau} df \quad (4.169)$$

$$= 2 \int_0^{+\infty} \gamma_W(f) \cos(2\pi f \tau) df \quad (4.170)$$

$$= \frac{N_0}{2\pi\tau} [\sin(2\pi f \tau)]_{f_c - \frac{B}{2}}^{f_c + \frac{B}{2}} \quad (4.171)$$

$$= BN_0 \frac{\sin(\pi B \tau)}{\pi B \tau} \cos(2\pi f_c \tau) \quad (4.172)$$

La puissance de bruit vaut donc $\sigma^2 = \Gamma_{WW}(0) = BN_0$.

Passage du bruit blanc à travers un filtre passe-bas

Imaginons à présent le même bruit blanc en bande de base :

$$\gamma_W(f) = \begin{cases} \eta, & -\frac{B}{2} \leq f \leq \frac{B}{2} \\ 0, & |f| > \frac{B}{2} \end{cases} \quad (4.173)$$

La fonction d'autocorrélation vaut

$$\Gamma_{WW}(\tau) = 2\eta \int_0^{+\frac{B}{2}} \cos(2\pi f \tau) df \quad (4.174)$$

$$= \frac{\eta}{\pi\tau} [\sin(2\pi f \tau)]_0^{+\frac{B}{2}} \quad (4.175)$$

$$= B\eta \frac{\sin(\pi B \tau)}{\pi B \tau} \quad (4.176)$$

Il faut donc prendre $\eta = N_0$ pour avoir la même puissance totale qu'un signal passe-bande. Cette forme de fonction d'autocorrélation en relation avec la précédente suggère qu'un bruit en bande étroite puisse être vu comme un bruit en bande de base transposé à une fréquence de porteuse. Cette remarque conduit à la décomposition de RICE.

Dans la modélisation des signaux aléatoires, nous avons défini séparément le caractère blanc et le caractère gaussien. Le premier porte uniquement sur la valeur des deux premiers moments : processus centré et densité spectrale constante. Le second porte sur les distributions statistiques du processus. Il n'y a donc aucune implication entre ces deux propriétés⁴.

4. Pour ne rien arranger à l'affaire, certains auteurs appellent processus blanc, un processus possédant à la fois ces deux caractères

4.6.5 Décomposition de Rice

L'expression

$$s(t) = A(t) \cos(2\pi f_c t + \phi(t)) \quad (4.177)$$

représente un signal modulé⁵.

Ce signal peut aussi être écrit sous la forme de deux composantes dont la première est en phase, composante I , et la seconde en quadrature de phase, composante Q :

$$s_I(t) = A(t) \cos \phi(t) \quad (4.178)$$

$$s_Q(t) = A(t) \sin \phi(t) \quad (4.179)$$

telles que

$$s(t) = A(t) \cos(2\pi f_c t + \phi(t)) = s_I(t) \cos(2\pi f_c t) - s_Q(t) \sin(2\pi f_c t) \quad (4.180)$$

En d'autres termes, l'amplitude des composantes modulées I/Q dépend du temps, même lorsque l'enveloppe du signal modulé, $A(t)$, est constante comme c'est le cas pour le **GSM**.

Une autre interprétation permet de traiter les composantes comme deux signaux modulés en amplitude à porteuse supprimée. Pour un signal $s(t)$ à bande limitée, la décomposition de RICE mène à deux composantes en quadrature indépendantes et en bande de base.

Décomposition de Rice d'un bruit blanc à bande étroite

Il est intéressant de calculer, à titre d'illustration, la fonction d'autocorrélation de la décomposition de RICE d'un bruit à bande étroite, et de la comparer au résultat précédent.

Soient $n_I(t)$ et $n_Q(t)$ des processus de bruit indépendants et de densité spectrale $\gamma(f) = N_0$ pour $|f| < \frac{B}{2}$, tels qu'ils définissent

$$n(t) = n_I(t) \cos(2\pi f_c t) - n_Q(t) \sin(2\pi f_c t) \quad (4.181)$$

La fonction d'autocorrélation des processus $n_I(t)$ et $n_Q(t)$ est la même; elle vaut

$$\Gamma_{N_I N_I}(\tau) = \Gamma_{N_Q N_Q}(\tau) = BN_0 \frac{\sin(\pi B\tau)}{\pi B\tau} \quad (4.182)$$

Quant à celle de $n(t)$, elle se calcule comme suit

$$\Gamma_{NN}(\tau) = E\{n(t)n(t-\tau)\} \quad (4.183)$$

$$\begin{aligned} &= E\{n_I(t)n_I(t-\tau)\} \cos(2\pi f_c t) \cos(2\pi f_c t - 2\pi f_c \tau) \\ &\quad + E\{n_Q(t)n_Q(t-\tau)\} \sin(2\pi f_c t) \sin(2\pi f_c t - 2\pi f_c \tau) \\ &\quad - E\{n_I(t)n_Q(t-\tau)\} \cos(2\pi f_c t) \sin(2\pi f_c t - 2\pi f_c \tau) \\ &\quad - E\{n_Q(t)n_I(t-\tau)\} \sin(2\pi f_c t) \cos(2\pi f_c t - 2\pi f_c \tau) \end{aligned} \quad (4.184)$$

$$\begin{aligned} &= BN_0 \frac{\sin(\pi B\tau)}{\pi B\tau} [\cos(2\pi f_c t) \cos(2\pi f_c t - 2\pi f_c \tau) \\ &\quad + \sin(2\pi f_c t) \sin(2\pi f_c t - 2\pi f_c \tau)] \end{aligned} \quad (4.185)$$

$$= BN_0 \frac{\sin(\pi B\tau)}{\pi B\tau} \cos(2\pi f_c \tau) \quad (4.186)$$

5. Pour l'étude des modulations numériques, il est courant de définir $s(t) = \sqrt{\frac{E_b}{T_b}} a(t) \cos(2\pi f_c t + \phi(t))$ où E_b et T_b désignent respectivement l'énergie transmise par bit et la durée d'un bit.

Cette dernière expression correspond bien à la fonction d'autocorrélation d'un bruit blanc à bande étroite. Remarquons également que chaque composante en bande de base contient autant de puissance que la décomposition de RICE de $n(t)$. En effet,

$$\Gamma_{NN}(0) = BN_0 \quad (4.187)$$

Propriété 65 Les composantes $n_I(t)$ et $n_Q(t)$ d'un bruit blanc à bande étroite $n(t)$ sont de moyenne nulle.

Propriété 66 Si le bruit $n(t)$ à bande étroite est gaussien, ses composantes $n_I(t)$ et $n_Q(t)$ sont conjointement gaussiennes.

4.6.6 Bruit total en sortie

Un dispositif électronique est souvent caractérisé par le rapport signal à bruit en sortie. Il faut donc connaître la puissance de bruit totale résultante de toutes les sources élémentaires de bruit (indépendantes) du dispositif. Certaines sources ont généralement un effet négligeable par rapport à d'autres.

4.6.7 Bande équivalente de bruit

Les bruits étant souvent blancs, le calcul de la puissance de bruit en sortie d'un filtre linéaire fait intervenir l'intégrale du carré du module du gain complexe du filtre. Aussi est-il commode de définir la bande de bruit équivalente B . Comme

$$N_{out} = \frac{N_0}{2} \int_{-\infty}^{+\infty} \|\mathcal{H}(f)\|^2 df \quad (4.188)$$

$$= N_0 \int_0^{+\infty} \|\mathcal{H}(f)\|^2 df \quad (4.189)$$

et

$$N_{out} = N_0 B \mathcal{H}^2(0) \quad (4.190)$$

on aboutit à la définition suivante.

Définition 67 La bande de bruit équivalente est définie comme

$$B = \frac{\int_0^{+\infty} \|\mathcal{H}(f)\|^2 df}{\mathcal{H}^2(0)} \quad (4.191)$$

Chapitre 5

Introduction à la numérisation

5.1 Introduction

L'analyse de la modulation analogique et de ses performances donnent l'impression que l'objectif poursuivi est de transmettre à tout instant la valeur exacte du signal porteur d'information. C'est en effet à quoi tendent les techniques de modulation classique par une porteuse sinusoïdale, telles qu'utilisées dans la majorité des systèmes de télécommunications actuels. Pourtant, dès les années 1940, des savants comme NYQUIST, SHANNON ou WIENER établirent le cadre mathématique théorique des télécommunications. C'est grâce à eux que l'on pu prendre conscience du caractère illusoire de cette prétention.

Les signaux porteurs d'information sont généralement à spectre limité –c'est à tout le moins l'hypothèse que nous avons formulée précédemment. Dans les systèmes de télécommunications analogiques, on tire usage de cette propriété en effectuant un multiplexage en fréquence. Mais il est possible d'adopter une autre approche. En effet, le signal étant à spectre limité, sa valeur ne peut varier infiniment vite et il suffit sans doute de sélectionner des échantillons représentatifs du signal à une condition appropriée. Le théorème de SHANNON montre que si la fréquence d'échantillonnage est supérieure à une certaine limite, cela suffit pour que le récepteur soit à même de reconstituer parfaitement le signal de départ.

D'autre part, que l'on transmette le message de façon continue ou échantillonnée, il est illusoire d'en donner la valeur exacte. Comme les transducteurs d'émission et de réception n'ont pas un comportement parfait et que le canal de transmission lui-même introduit du bruit et des distorsions, il est superflu de vouloir restituer la valeur instantanée exacte du message. Aussi, si du moins cela apporte un avantage, peut-on coder cette valeur en un nombre comportant une quantité limitée de chiffres significatifs : c'est une quantification de la valeur instantanée. Ce procédé se marie parfaitement avec une technique d'échantillonnage.

On retrouve les opérations d'échantillonnage et de quantification déjà évoquées à la figure 5.1 qui montre une chaîne de numérisation complète.

La partie gauche couvre les étapes de conversion de l'analogique en signaux numériques. Ces derniers se prêtent tout naturellement à une opération de **compression** qui, dans les systèmes

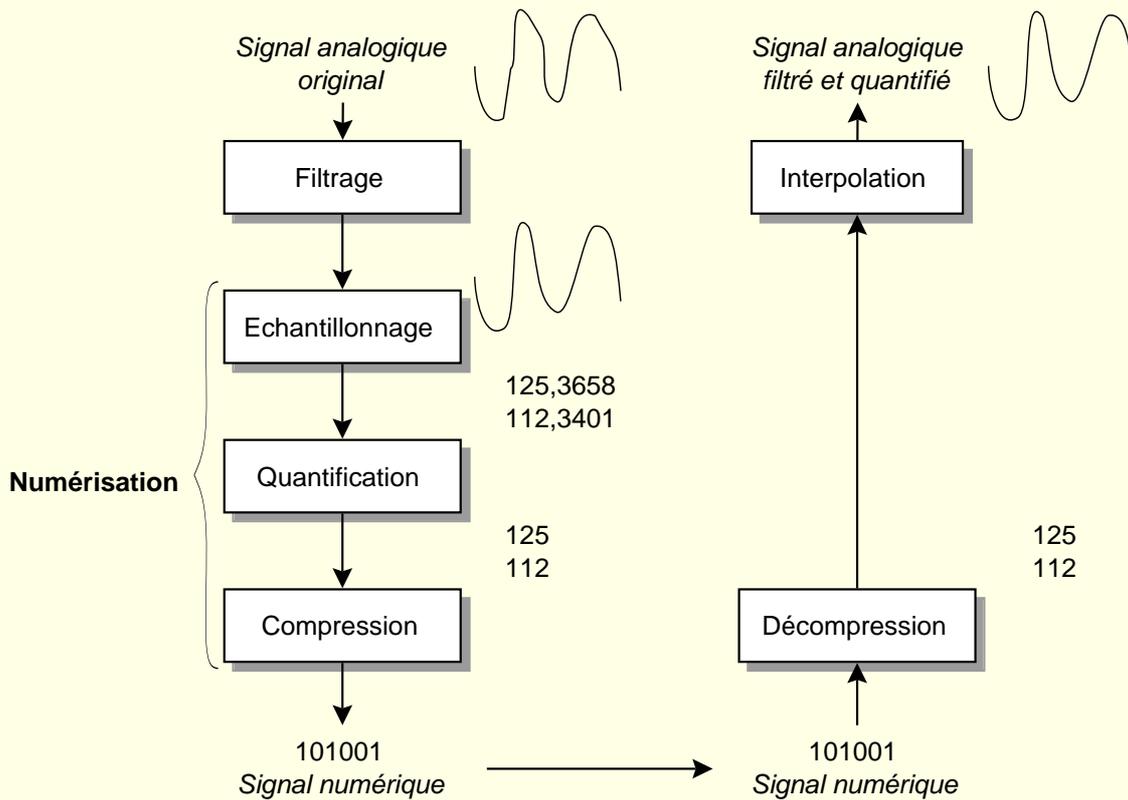


FIGURE 5.1 – Passage de l’analogique au numérique et conversion inverse.

modernes de télécommunications, est toujours présente. Pour reconstituer le signal analogique, les signaux numériques sont traités par un interpolateur représenté à droite.

La numérisation du signal ne garantit pas la correspondance entre le signal original et le signal issu de l’interpolateur. En effet, certaines étapes de la chaîne de numérisation introduisent une erreur sur le signal. La maîtrise de cette erreur est alors la clé d’une numérisation réussie.

5.1.1 Types d’information

Le terme de numérisation englobe différentes opérations dont la nature dépend fondamentalement des signaux originaux. Ainsi, comme l’illustre la figure 5.2, distingue-t-on les informations initialement numériques, les informations textuelles et les informations analogiques. Il est évident que seules les informations analogiques doivent être échantillonnées ; quant aux informations textuelles, elles requièrent un codage qui établit la correspondance entre l’alphabet et un code définissant, pour chaque caractère, une série de bits unique. C’est essentiellement le rôle du code ASCII (American Standard Code for Information Interchange), du code de BAUDOT ou de l’Unicode, une version internationale du code ASCII.

La mise en forme des informations est appelée codage de source, ce qui engendre parfois une certaine confusion tant les opérations de codage sont nombreuses dans une chaîne de transmission numérique. En effet, on parle de codage de source, de codage de canal, et de codage en lieu et place de compression.

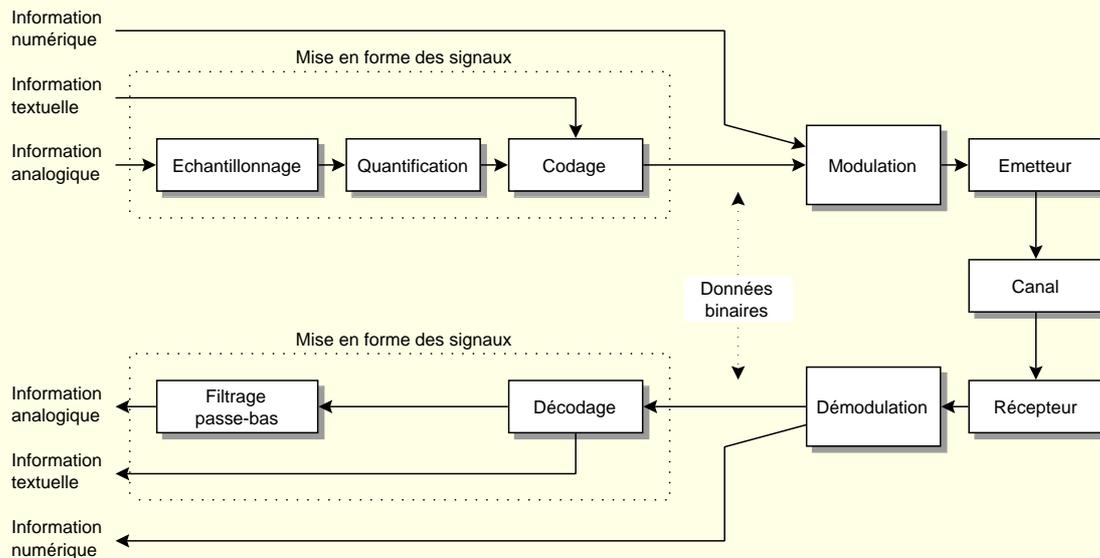


FIGURE 5.2 – Mise en forme et transmission [17].

5.1.2 Transmission

Les données binaires seront ensuite transmises au travers d'un canal, comme une paire de cuivre ou un coaxe. Mais avant d'y parvenir, il faut adapter le signal aux caractéristiques du canal. Dans le cas le plus simple d'une adaptation du signal au canal, on parle de transmission en bande de base. Ce procédé consiste à convertir les données binaires en signaux en forme d'impulsions ou de créneaux de tension. Contrairement à la modulation par une porteuse sinusoïdale, le spectre n'est pas translaté sur l'axe des fréquences. C'est la caractéristique essentielle d'une transmission en bande de base. Cela ne signifie nullement que cette forme de transmission soit la seule utilisable pour les signaux numériques. En effet, il est tout à fait possible de recourir aux mécanismes de modulation par porteuse au départ de signaux numériques ; il s'agit alors de modulation numérique. La différence importante est le fait que l'amplitude, la fréquence ou la phase, suivant le type de modulation choisi, sera altérée par sauts et non de manière continue. En conséquence, les critères de performance diffèrent de ceux d'une modulation analogique puisqu'il ne s'agit pas de reconstruire une valeur analogique mais de reconnaître un état discret parmi d'autres. On parle alors de taux d'erreur.

5.2 Conversion des signaux

5.2.1 Messages, caractères et symboles

Les informations textuelles se composent de caractères alphanumériques. Pour une transmission, les caractères sont tout d'abord convertis en une séquence de bits. En regroupant k bits, on forme de nouveaux nombres, appelés symboles, appartenant à un ensemble fini de caractères, ou alphabet, composés de $M = 2^k$ symboles. Un tel système est appelé système M-aire. Le choix de la valeur de k est un choix important pour la synthèse d'un schéma de télécommunications. Si $k = 1$, le système est dit binaire, le cardinal de l'ensemble M est 2 ; de plus, le modulateur doit représenter deux signaux. Dans ce cas particulier, il n'y a pas de différence entre le bit et

le symbole. Avec $M = 4$, on obtient un système quaternaire ou 4-aire. Le modulateur effectue, pour chaque symbole, le choix parmi quatre formes d'onde.

5.2.2 Traitement des signaux analogiques

L'information analogique ne se prête pas telle quelle à un codage; il faut, au préalable, la transformer en numérique. La première étape d'une conversion analogique-numérique est le filtrage ou l'échantillonnage. Comme nous le verrons, ces deux opérations sont intimement liées.

Échantillonnage instantané : théorème de Shannon

On décrit généralement l'échantillonnage dans le domaine temporel. Il s'agit du processus par lequel un signal analogique est converti en une série d'échantillons également espacés. L'opération n'a d'intérêt que s'il s'avère possible de reconstituer le signal analogique à partir des échantillons, ce qui dépendra principalement de la cadence d'échantillonnage. L'étude de cette cadence optimale fait l'objet du théorème de SHANNON.

Considérons un signal $g(t)$ d'énergie finie, connu pour tout temps t . Supposons que l'on prenne un échantillon de ce signal à une cadence régulière, soit un échantillon toutes les T_s secondes. La séquence ainsi obtenue est notée $\{g[nT_s]\}$, avec $n \in \mathbb{N}$ où T_s est la période d'échantillonnage et son inverse, $f_s = 1/T_s$ est la fréquence d'échantillonnage. Cette forme d'échantillonnage est appelée échantillonnage instantané. L'important théorème suivant lie la fréquence d'échantillonnage à la valeur de la bande de base.

Théorème 68 [Shannon] *Une fonction $g(t)$ à énergie finie et à spectre limité, c'est-à-dire dont la transformée de FOURIER $\mathcal{G}(f)$ est nulle pour $|f| > W$, est entièrement déterminée par ses échantillons $g[nT_s]$, $n \in \{-\infty, +\infty\}$ pour autant que la fréquence d'échantillonnage f_s soit strictement supérieure au double de la borne supérieure du spectre*

$$f_s > 2W \quad (5.1)$$

Démonstration

Pour démontrer cet important théorème, il suffit de noter que la fonction échantillonnée $g_s(t)$, définie par

$$g_s(t) = \sum_{n=-\infty}^{+\infty} g[nT_s] \delta(t - nT_s) \quad (5.2)$$

n'est rien d'autre que le produit de la fonction $g(t)$ par le train d'impulsions $\sum_{n=-\infty}^{+\infty} \delta(t - nT_s)$. En fait, $g_s(t)$ correspond à une modulation d'amplitude de ce train par $g(t)$. La transformée de FOURIER de $g_s(t)$ est donc la convolution des transformées des fonctions intervenant dans le produit temporel. Or, on montre que la transformée de FOURIER d'un train d'impulsions est un autre train d'impulsions [16]

$$\sum_{n=-\infty}^{+\infty} \delta(t - nT_s) \Leftrightarrow \frac{1}{T_s} \sum_{k=-\infty}^{+\infty} \delta(f - \frac{k}{T_s}) \quad (5.3)$$

Il s'ensuit que

$$g_s(t) \Leftrightarrow f_s \sum_{k=-\infty}^{+\infty} \mathcal{G}(f - kf_s) \quad (5.4)$$

Le spectre du signal échantillonné s'obtient donc en répétant le spectre de $g(t)$, centré sur les multiples de kf_s . La figure 5.3 illustre les effets de l'échantillonnage. Dès lors, on ne peut reconstituer $g(t)$, c'est-à-dire $\mathcal{G}(f)$, à partir de $g_s(t)$, soit $\mathcal{G}_s(f)$, que si ces spectres ne se recouvrent pas. Cela implique $f_s > 2W$, qui est la fréquence de NYQUIST. De plus, l'opération de reconstruction de $g(t)$ à partir de sa version échantillonnée $g_s(t)$ est très simple ; il suffit d'utiliser un filtre passe-bas de fréquence de coupure comprise entre W et $f_s - W$ et ayant un gain T_s dans la bande passante. ■

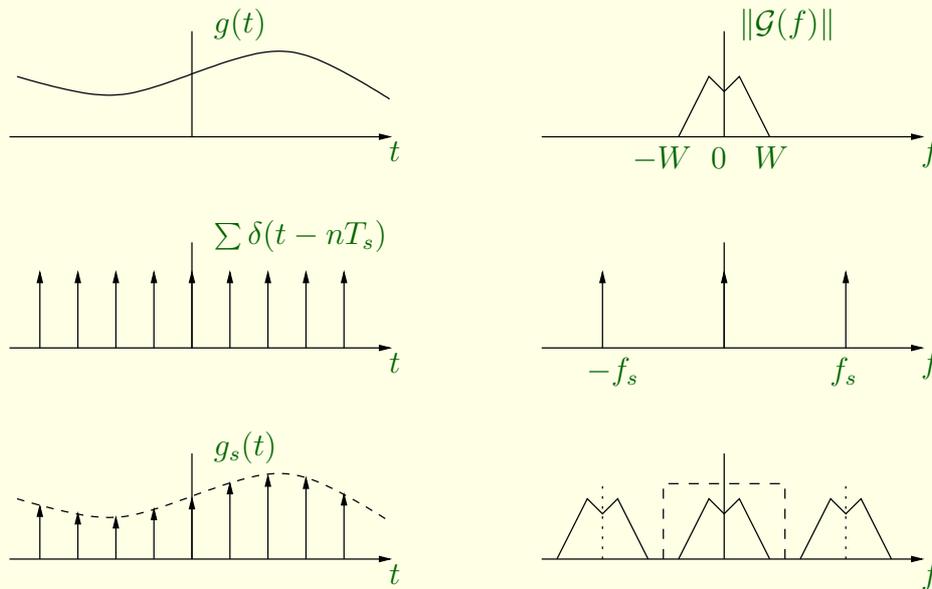


FIGURE 5.3 – Échantillonnage instantané.

Sur le plan technique, l'importance du théorème tient à ceci : l'infinité continue des valeurs du signal analogique contient la même information que l'infinité dénombrable de ses échantillons. La fréquence minimale d'échantillonnage, soit $2W$, est appelée fréquence de NYQUIST.

Il convient de remarquer que la contrainte $f_s > 2W$ est souvent ramenée à $f_s \geq 2W$. Il s'agit d'une erreur dans le cas de signaux sinusoïdaux. Pour s'en convaincre, il suffit d'échantillonner une sinusoïde de fréquence W à une fréquence $2W$ au droit de ses passages par zéro. Tous les échantillons de cette séquence sont nuls et il n'est dès lors pas possible de reconstituer la sinusoïde.

Formule d'interpolation de Whittaker

Le spectre de la fonction échantillonnée reproduit certes le signal initial dans la bande de base mais il lui adjoint une infinité de copie centrée sur les multiples de la fréquence d'échantillonnage ; c'est l'effet du train d'impulsions dans le domaine temporel. Pour reconstruire le signal original, il suffit d'utiliser un filtre passe-bas idéal de transmittance égale à T_s jusqu'à la fréquence de coupure f_{co} , choisie supérieure à W .

Prenons donc un filtre passe-bas idéal de transmittance égale à 1 jusqu'à la fréquence de coupure $f = f_{co}$. Ce filtre passe-bas idéal possède la réponse impulsionnelle suivante

$$2f_{co} \frac{\sin(2\pi f_{co}t)}{2\pi f_{co}t} = 2f_{co} \operatorname{sinc} \left(\frac{2t}{T_{co}} \right) \quad (5.5)$$

Le signal reconstruit $g_r(t)$ résulte du passage de la fonction échantillonnée au travers du filtre passe-bas idéal auquel, pour l'instant, on attribue un gain A :

$$g_r(t) = g_s(t) \otimes 2Af_{co} \text{sinc}\left(\frac{2t}{T_{co}}\right) \quad (5.6)$$

$$= \sum_{n=-\infty}^{+\infty} g[nT_s] \delta(t - nT_s) \otimes 2Af_{co} \text{sinc}\left(\frac{2t}{T_{co}}\right) \quad (5.7)$$

$$= 2Af_{co} \sum_{n=-\infty}^{+\infty} g[nT_s] \text{sinc}\left(\frac{2(t - nT_s)}{T_{co}}\right) \quad (5.8)$$

Il s'agit d'une formule d'interpolation permettant de reconstituer la fonction $g(t)$ à une constante près. Le théorème de WHITTAKER résulte du cas précis où le gain A et la fréquence de coupure f_{co} valent respectivement T_s et $f_s/2$.

Théorème 69 [Formule d'interpolation de Whittaker] Soit $g(t)$ un signal analogique intégrable de spectre borné $[-W, W]$. Soit $\{g[nT_s]\}$ l'ensemble de ses échantillons de pas $T_s = 1/f_s$. La fonction $g(t)$ s'écrit comme la série de fonctions

$$g(t) = \sum_{n=-\infty}^{+\infty} g[nT_s] \text{sinc}\left(\frac{t - nT_s}{T_s}\right) \quad (5.9)$$

Ce théorème est connu comme théorème de SHANNON par les uns, comme théorème de KOTELNIKOV par les autres. En fait, WHITTAKER l'a mentionné en 1918 et on attribue sa paternité à CAUCHY.

L'interprétation suivante est immédiate (cf. figure 5.4) : $\text{sinc}\left(\frac{t - nT_s}{T_s}\right)$ est égal à 1 pour $t = nT_s$ et égal à 0 pour tous les $t = kT_s$ tels que $k \neq n$. La somme de la série 5.9 se réduit donc à $g[nT_s]$ pour $t = nT_s$. Le théorème signifie qu'entre ces valeurs cette série interpole exactement $g(t)$.

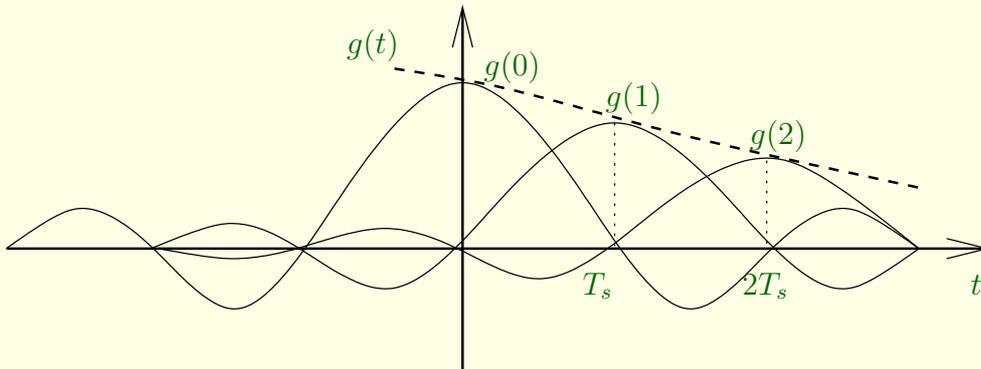


FIGURE 5.4 – Interprétation des termes de la formule d'interpolation de WHITTAKER.

Dans la pratique, on opère sur des sommes finies telles

$$g_N(t) = \sum_{n=-N}^{+N} g[nT_s] \text{sinc}\left(\frac{t - nT_s}{T_s}\right) \quad (5.10)$$

$g_N(t)$ a les mêmes échantillons que $g(t)$ pour $|n| \leq N$, mais $g_N[nT_s] = 0$ pour $|n| \geq N$.

Filtrage préalable à l'échantillonnage

L'établissement d'un cadre pour l'échantillonnage repose sur la propriété de largeur de bande limitée du signal $g(t)$. En pratique, l'information n'est naturellement pas à bande limitée. La convolution par le train d'impulsions introduit alors un recouvrement entre bandes voisines dans le spectre; ce phénomène est appelé repli de spectre ou aliasing. L'opération de filtrage passe-bas s'avère incapable de supprimer ce recouvrement et on risque de voir apparaître, en raison du repli de spectre, une série de fréquences initialement absentes dans $g(t)$. Deux approches suppriment ces effets indésirables :

1. Un filtrage passe-bas préalable à l'échantillonnage réduit la largeur de la bande de base dans le respect du critère de SHANNON;
2. Le signal est échantillonné à une fréquence légèrement supérieure à la fréquence de NYQUIST.

Formule de Nyquist pratique. En raison de la difficulté qu'il y a à réaliser un filtre ayant un flanc raide au droit de la fréquence de coupure, il est d'usage de définir une bande de garde dans laquelle la transition est plus douce. La bande de garde, typiquement de l'ordre de 10 à 20%, entraîne donc une augmentation de la fréquence d'échantillonnage. On parle parfois de la forme pratique du critère de NYQUIST, tenant compte de cette bande de garde

$$f_s = 2,2W \quad (5.11)$$

Exemple. Soit à produire un signal numérique à partir d'un signal musical s'étendant jusqu'à 20 [kHz]. En toute rigueur, une fréquence d'échantillonnage supérieure à 40 [kHz] suffit. Pour des questions de réalisation de filtre, on utilise plutôt 44,1 [kHz] dans le standard du CD-Audio et même 48 [kHz] pour des signaux de qualité studio.

Réalisation de l'échantillonnage

Émission. L'échantillonnage se fait en pratique avec une porte analogique qui est un interrupteur fermé pendant une durée τ . Celle-ci doit évidemment être très petite par rapport à T_s . En effet, le signal échantillonné ainsi obtenu $g_\tau(t)$ est le produit, dans le domaine temporel, de $g(t)$ par un train $p_s(t)$ de rectangles de largeur τ

$$p_s(t) = p(t) \otimes \sum_{n=-\infty}^{+\infty} \delta(t - nT_s) = \sum_{n=-\infty}^{+\infty} p(t - nT_s) \quad (5.12)$$

avec

$$p(t) = \begin{cases} 1 & \text{si } 0 \leq t \leq \tau \\ 0 & \text{sinon} \end{cases} \quad (5.13)$$

L'impulsion et son spectre sont dessinés à la figure 5.5.

La transformée de FOURIER de $p_s(t)$ étant

$$\mathcal{P}_s(f) = \frac{e^{-\pi j f \tau} \sin(\pi f \tau)}{T_s} \sum_{n=-\infty}^{+\infty} \delta(f - n f_s) \quad (5.14)$$

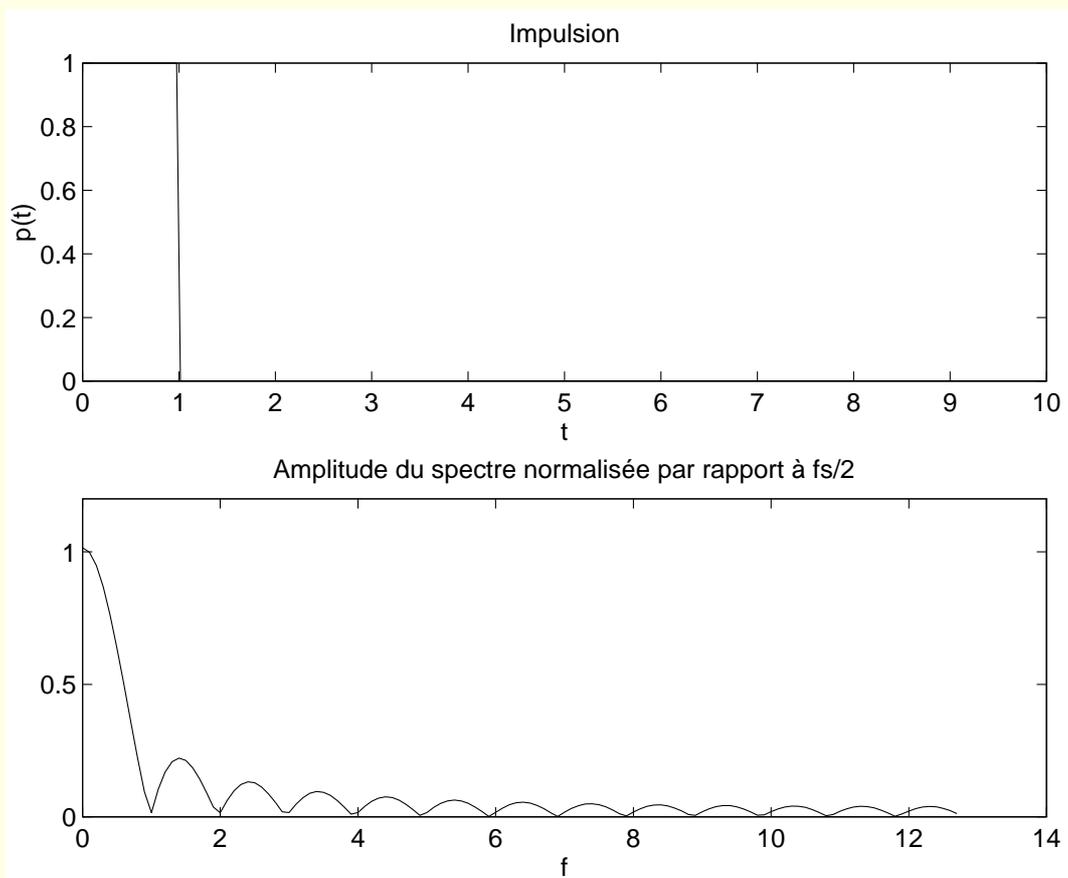


FIGURE 5.5 – Une impulsion et son spectre.

Celle de $g_\tau(t)$, parfois appelé signal échantillonné naturel, vaut

$$\mathcal{G}_\tau(f) = \mathcal{G}(f) \otimes \mathcal{P}_s(f) \quad (5.15)$$

$$= \frac{\tau}{T_s} \sum_{n=-\infty}^{+\infty} e^{-n\pi j f_s \tau} \frac{\sin(n\pi f_s \tau)}{n\pi f_s \tau} \mathcal{G}(f - n f_s) \quad (5.16)$$

De ceci, on peut conclure que si la durée de fermeture de l'interrupteur n'est pas infinitésimale, les spectres répétés sont affectés de coefficients différents, bien que chacun d'eux reste une copie conforme de $\mathcal{G}(f)$. La reconstruction à l'aide d'un filtre passe-bas reste donc correcte. Bien entendu, il faut pour cela que l'interrupteur soit parfait, comme on l'a supposé, c'est-à-dire qu'il ait une impédance nulle ou infinie selon qu'il est fermé ou ouvert.

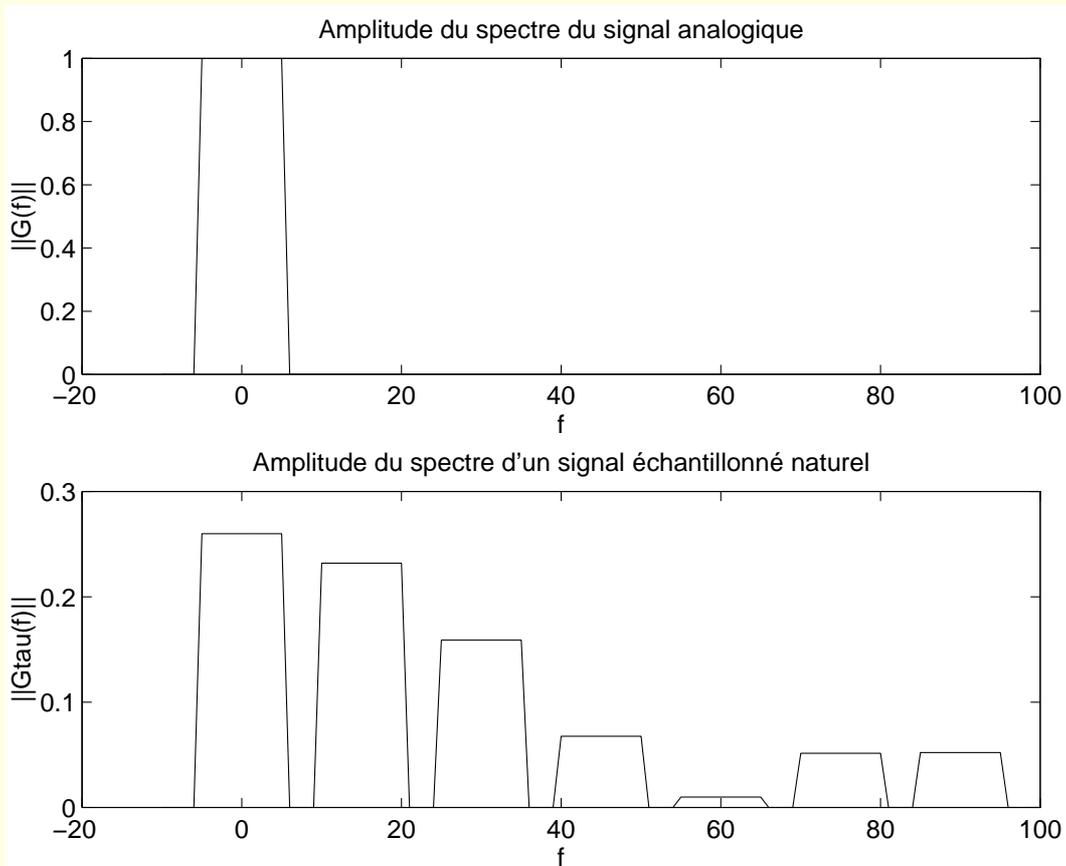


FIGURE 5.6 – Spectre d'un signal analogique et du signal échantillonné naturel.

En pratique, il faut souvent saisir le signal à l'instant d'échantillonnage et le maintenir dans une mémoire pendant un temps τ_m nécessaire à l'exécution de certaines opérations de calcul. Le plus souvent, cette mémoire est un condensateur.

Conversion analogique-numérique

Il existe une assez grande variété de convertisseurs analogique-numérique (A/D, pour Analog-to-Digital).

Dans un système de télécommunications où l'information fait l'objet d'un échantillonnage, le signal transmis a la forme d'un train d'impulsions. Il existe différentes façons de présenter

l'information sous la forme d'une impulsion. Par exemple, on peut moduler l'amplitude d'un train d'impulsions (PAM, Pulse Amplitude Modulation). On pourrait de même conserver une amplitude constante à l'impulsion, mais rendre sa largeur τ proportionnelle à l'échantillon (PDM, Pulse Duration Modulation) ou encore, garder à l'impulsion une amplitude et une largeur constante mais la décaler par rapport à son instant d'arrivée théorique kT_s d'un temps variable, proportionnel à $g[nT_s]$, ce qui constitue une modulation de phase ou de position (PPM, Pulse Position Modulation). Dans toutes ces méthodes, la modulation d'impulsions reste analogique, en ce sens qu'un paramètre continu est rendu proportionnel à l'échantillon. Comme d'après le théorème de SHANNON, l'échantillonnage n'introduit aucune erreur, ces méthodes permettent en théorie une transmission rigoureusement fidèle de l'information. Elles ont en principe un avantage par rapport à la transmission continue du signal d'information $g(t)$: à puissance moyenne égale, les impulsions ont une amplitude d'autant plus grande que le rapport τ/T_s est petit et elles peuvent d'autant mieux être isolées du bruit. La contrepartie est cependant sévère car la bande passante augmente dans des proportions inquiétantes.

Par exemple, pour un signal téléphonique limité à 3400 [Hz], on pourrait prendre $f_s = 8$ [kHz], soit $T_s = 125$ [μ s]. Un rapport $\tau/T_s = 0,1$ donne une durée de $\tau = 12,5$ [μ s], ce qui exige une bande passante minimale de l'ordre de 50 [kHz] si l'on veut plus ou moins respecter la forme des impulsions¹. Cela ne constituerait pas nécessairement un inconvénient majeur, car c'est une règle générale que l'élargissement du spectre apporte une amélioration du rapport signal à bruit.

Aussi, les modulations analogiques d'un train d'impulsions n'ont-elles pas connu un très grand succès. Ce n'est pas le cas des modulations numériques comme on va le voir. En plus, une transmission d'impulsions sans en déformer la forme n'est pas une chose simple.

Dans un système de télécommunications où l'information fait l'objet d'un échantillonnage, on dispose d'un temps inférieur ou égal à la période d'échantillonnage T_s pour transmettre l'échantillon. Le signal transmis a donc la forme d'un train d'impulsions. Il existe diverses façons de présenter l'information sous la forme d'une impulsion.

5.3 Modulation PAM

La forme la plus simple de modulation d'impulsions analogiques est la modulation de l'amplitude des impulsions (Pulse Amplitude Modulation, PAM). Comme son nom l'indique, elle consiste à moduler l'amplitude d'un train d'impulsions équidistantes, en fonction des valeurs des échantillons d'un signal analogique (cf. figure 5.7). Remarquons que l'usage du mot modulation prend ici un sens particulier car il n'est nullement question de moduler les paramètres d'une porteuse mais bien d'affecter le signal de départ.

Que les impulsions soient de forme rectangulaire ou autre importe peu pour le principe de la méthode, ce qui ne signifie nullement que la forme n'ait pas de d'influence sur l'occupation spectrale ou sur les performances, que du contraire!

On construit le signal PAM $m_{\tau'}(t)$ en maintenant le signal à une valeur constante. Ce signal diffère de $g_{\tau}(t)$ car $m_{\tau'}(t)$ reste constant et égal à la valeur de l'échantillon pendant un certain temps τ' .

1. Ce calcul se base sur le fait que la bande passante d'un signal de largeur périodique de largeur T vaut plus de $\frac{0,6}{T}$.

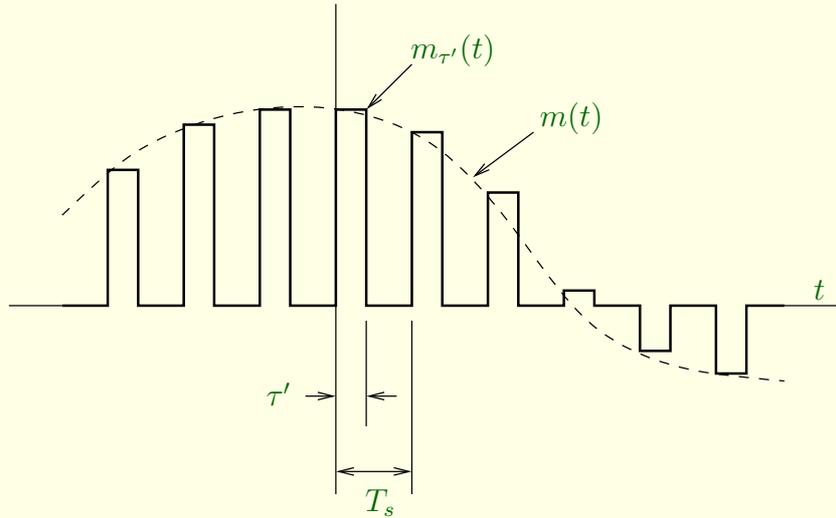


FIGURE 5.7 – Modulation PAM au moyen d'impulsions rectangulaires.

On peut considérer que $m_{\tau'}(t)$ résulte non plus de la multiplication de la fonction échantillonnée par un rectangle de largeur τ' mais de la convolution :

$$m_{\tau'}(t) = p(t) \otimes \left[g(t) \sum_{n=-\infty}^{+\infty} \delta(t - nT_s) \right] \quad (5.17)$$

Sa transformée de FOURIER vaut

$$\mathcal{M}_{\tau'}(f) = \mathcal{P}(f) \left[\mathcal{G}(f) \otimes \left(\frac{1}{T_s} \sum_{n=-\infty}^{+\infty} \delta(f - nf_s) \right) \right] \quad (5.18)$$

$$= \mathcal{P}(f) \frac{1}{T_s} \sum_{n=-\infty}^{+\infty} \mathcal{G}(f - nf_s) \quad (5.19)$$

$$= \frac{e^{-\pi j f \tau'} \sin(\pi f \tau')}{T_s \pi f} \sum_{n=-\infty}^{+\infty} \mathcal{G}(f - nf_s) \quad (5.20)$$

Son module est représenté à la figure 5.8.

Cette expression diffère de 5.16 en ce que chaque spectre répété de la fonction $m(t)$ subit une déformation due au facteur du sinus cardinal. Un filtrage passe-bas du signal restitué donc $m(t)$ avec une déformation importante si τ' est appréciable par rapport à T_s , mais ceci peut se corriger par un filtre correcteur.

Une chaîne de conditionnement qui se limite à la mise en forme de signaux en vue d'un traitement effectif comprend, dans une version relativement complète, une conversion analogique-numérique suivie, selon les besoins, d'un filtre numérique destiné à adapter partiellement les propriétés fréquentielles au traitement ultérieur. Comme nous l'avons vu, la chaîne de conversion se voit précédée d'un filtre analogique passe-bas qui ajuste la bande passante du signal en fonction de la fréquence caractéristique de l'échantillonnage.

5.4 Modulation de la position des impulsions

Dans un système de modulation des impulsions, on peut représenter la valeur des échantillons par un autre paramètre que l'amplitude d'un train d'impulsions, tout comme dans le cas de

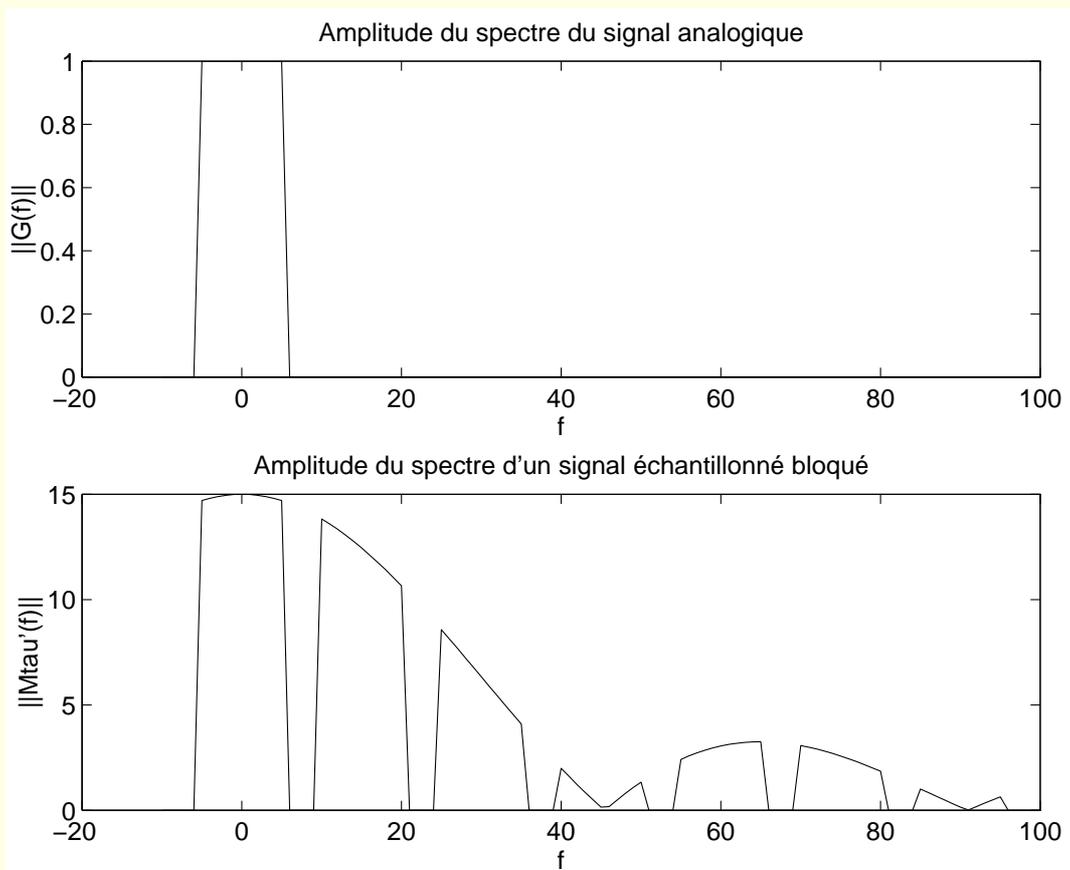


FIGURE 5.8 – Spectre d'un signal analogique et du signal échantillonné bloqué.

modulation analogique il était possible de moduler, outre l'amplitude, la phase ou la fréquence. Si l'on conserve une amplitude constante à l'impulsion mais que sa largeur est rendue proportionnelle à l'échantillon, on obtient une modulation appelé Pulse-Duration Modulation (PDM) ; on parle aussi de pulse-width modulation ou encore de pulse-length modulation. La modulation affectera soit le début de l'impulsion, soit la fin, voire les deux extrémités.

Le déplacement d'une extrémité de l'impulsion n'est pas gênant en soi, par contre la modulation PDM consomme une certaine puissance à transmettre un signal constant, non porteur d'information. En ne transmettant qu'une impulsion de courte durée au droit d'une transition du signal PDM, on fait l'économie de cette puissance; il s'agit d'une technique appelée pulse-position modulation (PPM). La figure 5.9 illustre les techniques de modulation précitées.

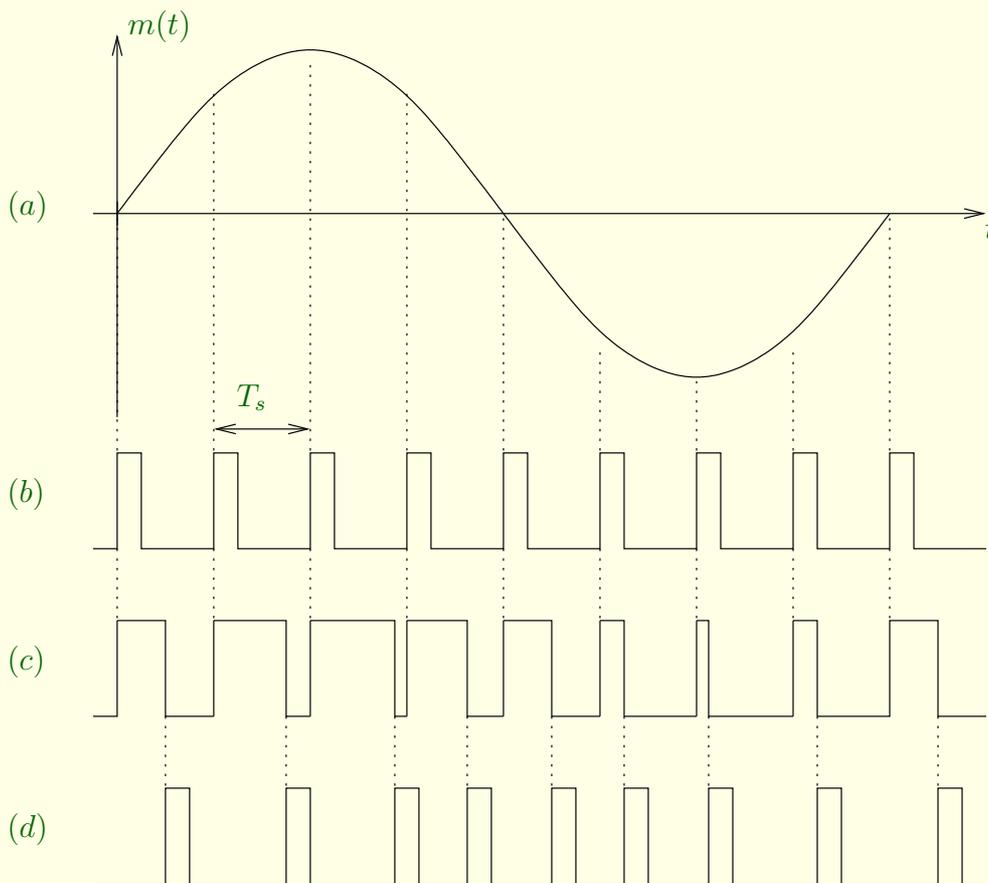


FIGURE 5.9 – Un signal modulant, train d'impulsions à moduler, modulation PDM et modulation PPM.

Dans le cas d'une modulation PDM utilisant un signal $p(t)$ de largeur τ variable, les impulsions voisines ne peuvent se chevaucher, ce que propose la condition suffisante suivante, à partir d'un signal modulant $m(t)$,

$$p(t) \neq 0, \quad \tau = k_p |m(t)|_{\max} < \frac{T_s}{2} \quad (5.21)$$

où k_p représente la sensibilité du modulateur.

Un signal PPM s'écrit

$$s(t) = \sum_{n=-\infty}^{+\infty} p(t - nT_s - k_p m[nT_s]) \quad (5.22)$$

où $p(t)$ est une impulsion centrée de largeur fixe τ . Pour éviter l'interférence entre symboles successifs, il faut respecter la condition suivante qui fait intervenir la largeur de l'impulsions

$$p(t) \neq 0, \quad k_p |m(t)|_{\max} < \frac{T_s}{2} - \frac{\tau}{2} \quad (5.23)$$

Plus $k_p |m(t)|_{\max}$ approche $T_s/2$, plus la forme de l'impulsion $p(t)$ doit être étroite pour éviter l'interférence entre impulsions successives, et plus la bande occupée par le signal PPM sera large. À condition d'avoir une modulation sans interférence entre impulsions successives et un signal modulant $m(t)$ à spectre limité, le respect du théorème de SHANNON garantit que l'on soit en mesure de reconstituer parfaitement le signal modulant $m(t)$ à partir de $s(t)$; la modulation PPM n'est pas source de distorsion.

5.5 Modulation d'impulsions codées PCM

Dans la modulation d'impulsions codées (MIC), en anglais Pulse Code Modulation (**PCM**), on estime que la transmission exacte des échantillons est superflue, car illusoire en raison des imperfections inéluctables de tout canal de transmission, et l'on ne craint pas d'introduire délibérément une erreur, pour autant que celle-ci reste dans des limites acceptables. Le fait qu'un être humain ne détecte qu'un nombre fini de différences visuelles ou auditives renforce l'intérêt qu'il peut y avoir à approximer un signal par un ensemble de valeurs discrètes. À cet effet, le domaine de variation du signal d'information $m(t)$ est divisé en un certain nombre, fini, de niveaux discrets v_k , $k = 1, 2, \dots, L$.

5.5.1 Quantification

Définition 70 *Le processus consistant à transformer un échantillon d'amplitude $m[nT_s]$ d'un message $m(t)$ pris au temps $t = nT_s$ en une amplitude $v[nT_s]$, choisie dans un ensemble fini de valeurs possibles, est appelé quantification.*

On supposera l'instantanéité et l'absence de mémoire durant le processus de quantification, c'est-à-dire que la quantification d'un échantillon ne dépend pas de l'échantillon quantifié précédent ou suivant. C'est de loin la forme de quantification la plus répandue, bien que sous-optimale pour certains aspects.

La modulation PCM consiste en ceci : à chaque instant d'échantillonnage, au lieu de transmettre la valeur exacte $m[nT_s]$, on transmet le numéro d'ordre du niveau de quantification le plus proche. La figure 5.10 schématise le fonctionnement d'un quantificateur sans mémoire. Comme le montre le dessin de droite, les bornes m_k délimitent une série d'intervalles $I_k =]m_k, m_{k+1}]$.

Dans un premier temps, un signal d'amplitude m appartenant à l'intervalle $I_k =]m_k, m_{k+1}]$ est indicé par k ($k = 1, 2, \dots, L$). En sortie du quantificateur, l'indice k est transformé par une amplitude v_k ou par une suite de nombre, généralement en base binaire, significatives pour les valeurs contenues dans l'intervalle. Les valeurs représentatives v_k sont appelés niveaux représentatifs et la différence entre deux niveaux représentatifs successifs définit les intervalles ou pas de quantification. Les niveaux peuvent être équidistants ou non, la quantification étant dite linéaire ou non. En guise d'illustration, la figure 5.11 reprend l'ensemble des étapes d'une quantification de type PCM.

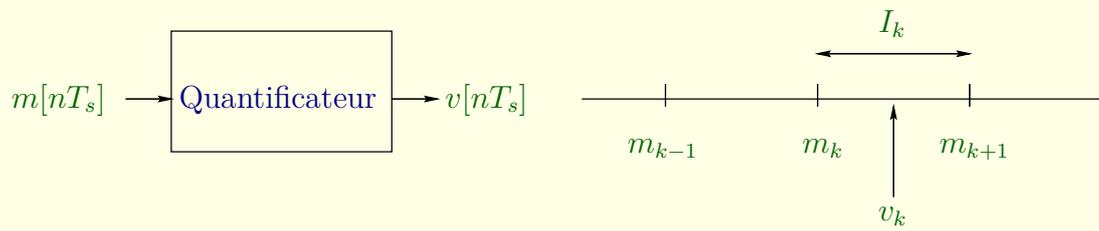


FIGURE 5.10 – Description d'un quantificateur sans mémoire.

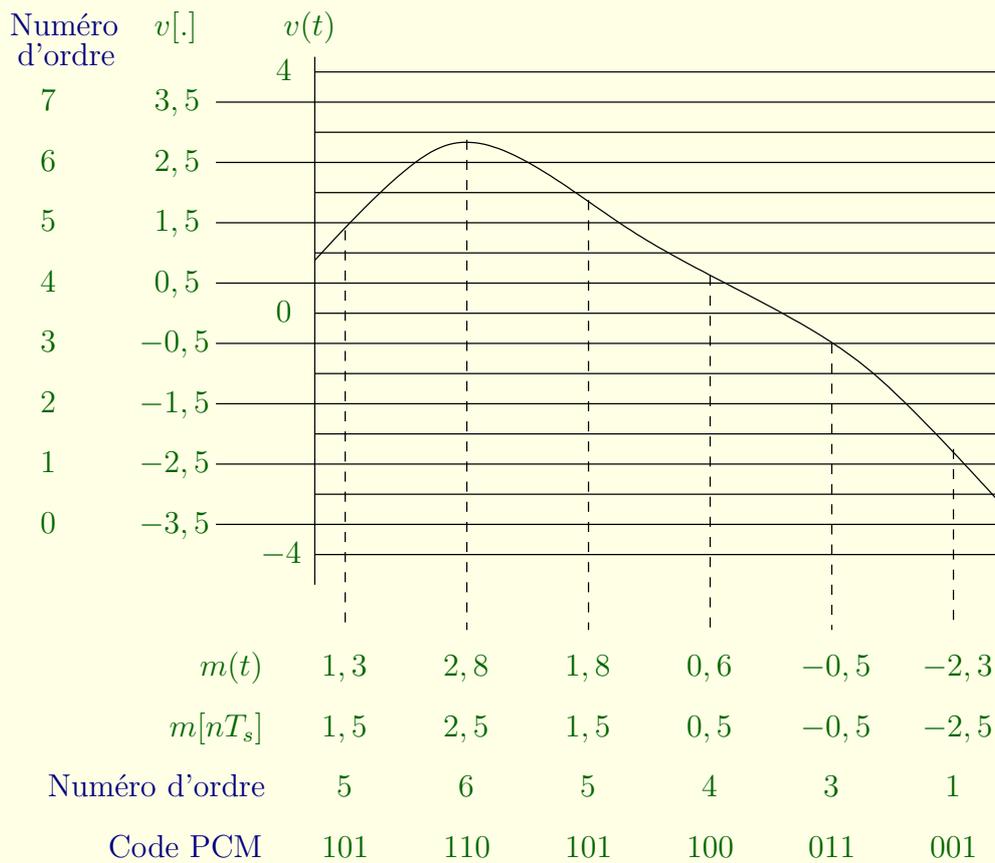


FIGURE 5.11 – Échantillons instantanés, quantifiés et codes PCM.

La fonction caractéristique d'un organe de quantification, $v = g(m)$, prend la forme d'une fonction en escaliers. Les deux alternatives, suivant que l'on transmette systématiquement le niveau représentatif immédiatement inférieur ou le plus proche, sont présentées à la figure 5.12.

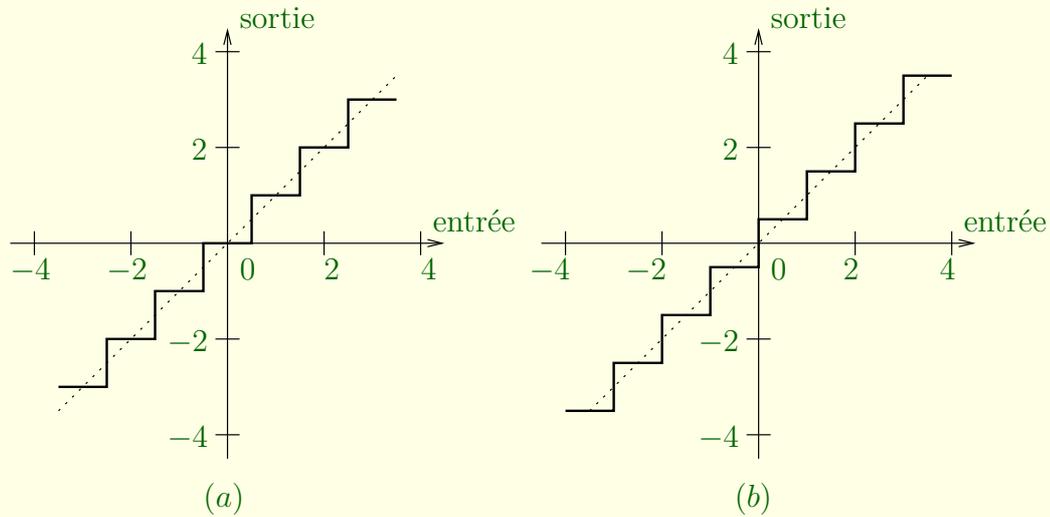


FIGURE 5.12 – Formes alternatives pour la fonction de quantification.

Il est clair qu'au travers d'une quantification, l'on commet une erreur d'arrondi. Le récepteur ne pourra jamais que reconstruire un signal $v(t)$ à spectre limité et passant, aux instants d'échantillonnage, par les niveaux quantifiés.

L'erreur $q(t) = v(t) - m(t)$ apparaît comme un bruit superposé au signal exact et appelé bruit de quantification². On ne peut le réduire qu'en choisissant un nombre de niveaux suffisants.

La bande passante requise pour la transmission des nombres issus des niveaux est évidemment beaucoup plus grande que celle qu'exigeait le signal initial. Par exemple, si l'on échantillonne exactement à la fréquence de NYQUIST $f_s = 2W$ et si l'on utilise toute la durée de l'intervalle T_s pour transmettre les M bits du code, la bande requise est

$$B = M \frac{f_s}{2} \quad (5.24)$$

Mais, par rapport aux modulateurs analogiques d'impulsions, on a ici l'avantage que les impulsions du code peuvent être largement distordues, entachées de bruit ou d'interférence aussi longtemps que ces aléas n'entraînent pas d'erreurs sur les mots binaires. En plus, on bénéficie de toute la technologie des circuits intégrés et de l'informatique.

Outre les questions de réalisation, un certain nombre de problèmes doivent cependant être examinés :

- Calcul du rapport signal à bruit de quantification.
- Influence de la quantification non-linéaire.
- Problème de codage : faut-il utiliser un code binaire simple ou un autre code ? En plus, il faut étudier la représentation des nombres binaires par des signaux électriques adaptés aux caractéristiques du canal de transmission (par exemple, si ce canal ne transmet pas le courant continu, le signal codé ne peut pas contenir de très basses fréquences).

2. Le bruit gênant est le bruit qui se situe dans la bande de base ; les bruits haute fréquence ne posent pas de problème.

- Problème de synchronisation. Il faut que le récepteur puisse repérer l'origine des mots binaires. Comment peut-on lui fournir cette information de cadrage ? Ce problème se résout en ajoutant des bits supplémentaires (par exemple un bit valant toujours 1 en tête de chaque mot).
- Influence des erreurs de transmission.

Nous n'examinerons pas l'ensemble des questions précitées mais la question du bruit de quantification mérite d'être abordée.

5.5.2 Bruit de quantification

L'étude du bruit de quantification est très importante non seulement pour le choix du nombre de niveaux de quantification, mais aussi de la représentation de ceux-ci. Ce qui importe en effet n'est pas le niveau de bruit de quantification mais plutôt le rapport des puissances moyennes à bruit de quantification.

Or, on sait que le puissance moyenne d'un signal de télécommunications peut varier dans une large mesure. Par exemple, pour un signal téléphonique, la puissance moyenne peut varier de quelques 20 à 30 [dB] selon la distance de l'abonné, l'état de son appareil et selon qu'il parle plus ou moins fort. Avec une quantification linéaire, le niveau de bruit reste constant, quelle que soit la puissance du signal ; par conséquent, les signaux faibles auront un rapport signal à bruit de quantification défavorable. Si l'on veut uniformiser le rapport signal à bruit de quantification, il y a donc lieu de diminuer le pas de quantification aux faibles niveaux, c'est-à-dire d'adopter une quantification non linéaire, comme illustrée à la figure 5.13. En fait, il faut proportionner la densité des niveaux de quantification à la densité de probabilité $f_M(m)$ du signal.

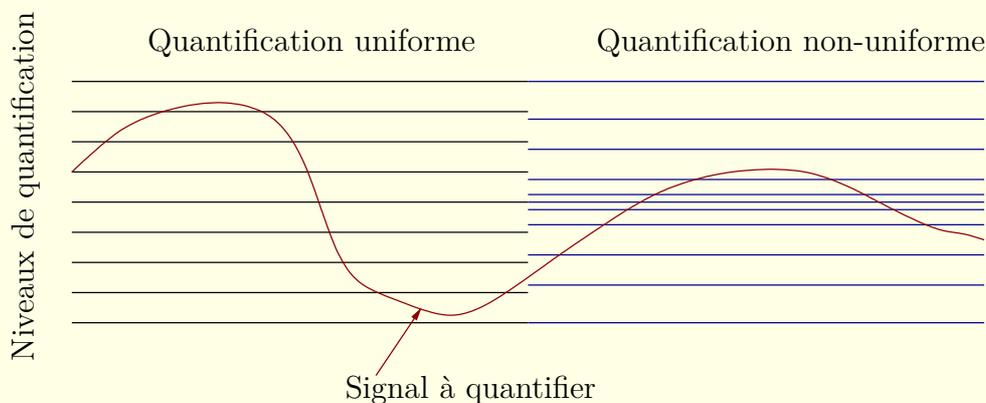


FIGURE 5.13 – Quantification uniforme et non uniforme.

Ceci peut se faire de deux manières : soit en utilisant un codeur à niveaux non uniformément répartis, soit en utilisant un codeur linéaire précédé d'un circuit non linéaire ayant une caractéristique en forme de S , et dont le but est d'uniformiser la densité de probabilité. Un tel circuit non linéaire est appelé compresseur. Le circuit complémentaire utilisé à la réception s'appelle un extenseur.

Pour une répartition donnée des niveaux de quantification, le rapport signal à bruit de quantification dépend de la densité de probabilité du signal m ainsi que de sa répartition de puissance dans le spectre. Soit le signal modulant m la réalisation d'une variable aléatoire M de moyenne nulle. Si l'entrée n'a pas une moyenne nulle, rien n'empêche de soustraire la valeur moyenne avant quantification et de rajouter cette valeur au moment de reconstruire les échantillons. Un

quantificateur $\phi(\cdot)$ est une application qui envoie la variable aléatoire M d'amplitude continue sur une variable aléatoire discrète V

$$\phi : M \rightarrow V \quad (5.25)$$

Soit une variable aléatoire Q représentant l'erreur de quantification de valeur q :

$$q = v - m \quad (5.26)$$

ou encore

$$Q = V - M \quad (5.27)$$

Lorsque le signal d'entrée M est de moyenne nulle et que le quantificateur est symétrique, il advient que la sortie du quantificateur est également de moyenne nulle. Il suffit dès lors de déterminer la valeur quadratique moyenne de l'erreur de quantification pour en déduire le rapport signal à bruit.

En tenant compte des moyennes nulles, les valeurs quadratiques moyennes du signal et de bruit de quantification s'obtiennent par les formules

$$\sigma_M^2 = \int_{-\infty}^{+\infty} m^2 f_M(m) dm \quad (5.28)$$

$$\sigma_Q^2 = \int_{-\infty}^{+\infty} q^2 f_Q(q) dq \quad (5.29)$$

Prenons le cas d'un signal m défini sur l'intervalle $[-m_{\max}, m_{\max}]$. Supposons qu'il s'agisse d'un quantificateur linéaire de la forme représentée à la figure 5.12(b). Le pas de ce quantificateur, noté Δ , vaut

$$\Delta = \frac{2m_{\max}}{L} \quad (5.30)$$

s'il y a L niveaux possibles. La linéarité du quantificateur permet de définir les bornes de l'erreur : $-\Delta/2 \leq q \leq \Delta/2$. Pour un pas de quantification suffisamment petit, ce qui signifie que le nombre de niveaux de quantification est grand, on peut admettre que l'erreur de quantification Q a une densité de probabilité uniforme. Moyennant cette hypothèse, la densité de probabilité de la variable Q s'écrit

$$f_Q(q) = \begin{cases} \frac{1}{\Delta}, & -\frac{\Delta}{2} < q \leq \frac{\Delta}{2} \\ 0, & \text{sinon} \end{cases} \quad (5.31)$$

Ajoutons que pour valider cette hypothèse, il est impératif que le signal m n'amène pas le quantificateur en saturation. On calcule alors la valeur de l'erreur quadratique moyenne

$$\sigma_Q^2 = \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} q^2 f_Q(q) dq = E \{Q^2\} \quad (5.32)$$

$$\sigma_Q^2 = \frac{1}{\Delta} \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} q^2 dq = \frac{\Delta^2}{12} \quad (5.33)$$

On peut traduire cette valeur en nombre de bits car c'est en effet sous une forme binaire que l'on transmet généralement les valeurs des indices k . Soit R le nombre de bits utilisés par échantillon. Le nombre de niveaux s'exprime comme

$$L = 2^R \quad (5.34)$$

ou encore

$$R = \log_2 L \quad (5.35)$$

La substitution de cette valeur dans l'équation 5.30 mène à

$$\Delta = \frac{2m_{\max}}{2^R} \quad (5.36)$$

et donc une valeur de σ_Q^2 valant

$$\sigma_Q^2 = \frac{1}{3} m_{\max}^2 2^{-2R} \quad (5.37)$$

Soit σ_M^2 la puissance moyenne du message $m(t)$. Le rapport signal à bruit³ de quantification σ_M^2/σ_Q^2 est égal à

$$\left(\frac{S}{N}\right)_q = \frac{\sigma_M^2}{\sigma_Q^2} \quad (5.38)$$

$$= \left(\frac{3\sigma_M^2}{m_{\max}^2}\right) 2^{2R} \quad (5.39)$$

L'ajout d'un bit supplémentaire augmente le rapport signal à bruit de 6 [dB] puisqu'en décibels, le rapport signal à bruit vaut

$$10 \log_{10} \left(\frac{S}{N}\right)_q = 10 \log_{10} 3 + 20 \log_{10} \left(\frac{\sigma_M}{m_{\max}}\right) + 20R \log_{10} 2 \quad (5.40)$$

$$= 4,77 + 20 \log_{10} \left(\frac{\sigma_M}{m_{\max}}\right) + 6R \quad (5.41)$$

L'ajout d'un bit élargit malheureusement aussi la bande nécessaire au signal. Toutefois, le compromis entre le rapport signal à bruit et la bande passante est plus avantageux que dans le cas de la modulation PPM, à condition de présumer qu'un système PPM soit jugé sur base du rapport signal à bruit et que le PCM le soit sur base du bruit de quantification.

Exemple : signal modulant sinusoïdal.

Prenons le cas d'un signal modulant sinusoïdal d'amplitude A_m que l'on veut quantifier sur toute sa dynamique. Pour une résistance valant $1 [\Omega]$, la puissance vaut $A_m^2/2$. La plage de quantification s'étend de $-A_m$ à A_m . Dès lors,

$$10 \log_{10} \left(\frac{S}{N}\right)_q = 1,8 + 6R \quad (5.42)$$

Pour des signaux modulateurs dont la densité de probabilité n'est pas aussi uniforme mais prend la forme d'une gaussienne par exemple, la quantification linéaire ne permet plus d'obtenir un bon rapport signal à bruit ; celui-ci peut être amélioré en optant pour une quantification non linéaire. En téléphonie, on a adopté pour l'Europe une loi de codage non linéaire dite loi A ; elle correspond à la courbe de compression de la figure 5.14(b).

3. La notion de rapport signal à bruit, appelé Signal to Noise Ratio (SNR) en anglais, est couramment utilisée en télécommunications. Elle qualifie la qualité d'un canal par le biais d'un rapport de puissance.

La loi de codage normalisée par l'ITU est la loi $A = 87,6$ approchée par 13 segments linéaire 8 bits, dans lequel le premier bit donne le signe, les trois suivants déterminent le segment et les quatre derniers fixent la position sur le segment. Cette courbe est particulièrement adaptée au signal téléphonique. Pour le codage d'autres signaux, dans une application bien particulière, il y aura lieu d'étudier la courbe de compression optimale.

Quant à la loi A , elle est définie comme

$$|v| = \begin{cases} \frac{A|m|}{1+\log A}, & 0 \leq |m| \leq \frac{1}{A} \\ \frac{1+\log(A|m|)}{1+\log A}, & \frac{1}{A} \leq |m| \leq 1 \end{cases} \quad (5.43)$$

$$\frac{d|m|}{d|v|} = \begin{cases} \frac{1+\log A}{A}, & 0 \leq |m| \leq \frac{1}{A} \\ (1 + \log A)|m|, & \frac{1}{A} \leq |m| \leq 1 \end{cases} \quad (5.44)$$

Aux États-Unis, on utilise la loi μ dont l'expression analytique est

$$|v| = \frac{\log(1 + \mu|m|)}{\log(1 + \mu)} \quad (5.45)$$

$$\frac{d|m|}{d|v|} = \frac{\log(1 + \mu)}{\mu}(1 + \mu|m|) \quad (5.46)$$

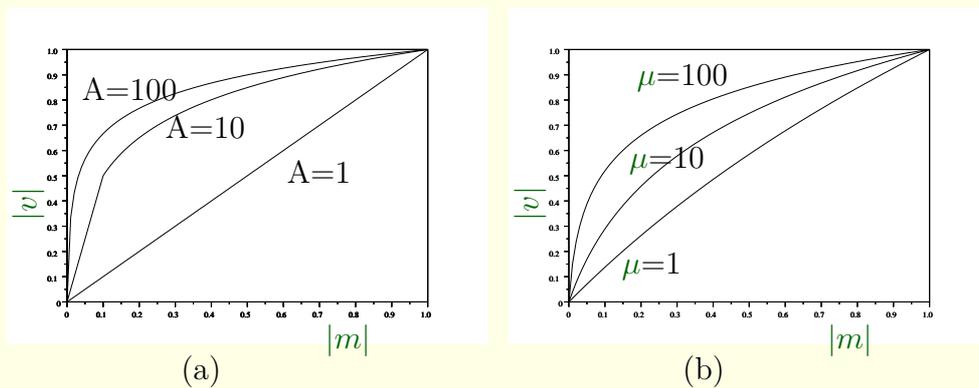


FIGURE 5.14 – Loïs de compression : (a) loi A , (b) loi μ .

5.5.3 Codage

Le travail de modulation PCM ne s'arrête pas avec l'attribution d'un intervalle car la valeur de l'indice n'est pas utilisable telle quelle. D'une manière ou d'une autre, il convient de transformer cette valeur en une forme d'onde que l'on puisse transmettre à travers le canal. Cette opération, appelée codage en ligne, revient à décrire la série de bits comme une série de signaux électriques.

En fait, les performances d'un système PCM sont principalement affectées par deux sources de bruit :

1. le bruit dû au canal de transmission.
2. le bruit de quantification, introduit par l'émetteur et dépendant du signal.

Les échantillons étant transmis sous forme binaire, un canal de transmission imparfait donne lieu à des erreurs sur bit qui se traduisent par une dégradation du rapport signal à bruit lors de la reconstruction du signal analogique. Les erreurs peuvent être dues à plusieurs causes : par exemple, si le canal est dispersif (délai de groupe non constant), le signal représentant un bit se disperse dans les intervalles de temps réservés aux bits voisins et il en résulte des interférences entre symboles pouvant donner lieu à des erreurs ; dans ce cas, la probabilité d'erreur dépend de la valeur du bit et de celle des voisins. Le canal peut aussi être le siège d'un bruit impulsionnel provoquant des erreurs par paquets ; c'est le cas du bruit de commutation sur les lignes téléphoniques. L'étude du codage sera abordée plus loin.

5.5.4 Modulation d'impulsions codées différentielles (DPCM)

Dans la modulation d'impulsions codées habituelles à L bits, la valeur du signal analogique est susceptible de changer du tout au tout d'un échantillon au suivant, puisque ces L bits sont transmis. Or il est bien peu probable que cela arrive ; en effet, pour passer sans cesse du niveau de quantification maximal au niveau minimal et inversement, il faudrait que le signal à coder soit une sinusoïde d'amplitude maximale et ayant une fréquence égale à la moitié de la fréquence d'échantillonnage. On transmet donc des bits superflus.

En modulation différentielle, on exploite ce fait pour réduire la vitesse de transmission en gardant en mémoire le niveau quantifié de l'échantillon précédent et en ne codant que l'incrément. Après codage, cette mémoire est remise à jour par addition binaire de l'incrément codé à l'ancienne valeur. Dans le décodeur, le code de l'incrément est ajouté à celui de l'échantillon précédent, qui a été gardé en mémoire, et la somme est décodée.

L'intérêt de cette modulation différentielle est d'autant plus grand qu'il y a peu de variation d'un échantillon au suivant, ce qui arrive lorsque la puissance du signal modulant est surtout concentrée dans les basses fréquences du spectre, ce qui est souvent le cas. Le nombre de bits à utiliser pour coder les incréments doit faire l'objet d'une étude statistique.

5.5.5 Modulation delta

La modulation delta est la version extrême d'une modulation différentielle. La fréquence d'échantillonnage est bien supérieure à la fréquence de NYQUIST et prise telle que l'accroissement entre deux échantillons successifs ne dépasse pas le pas de quantification.

La figure 5.15 montre le principe de la modulation delta et l'erreur qui résulte d'une variation trop rapide du signal par rapport à la cadence des bits.

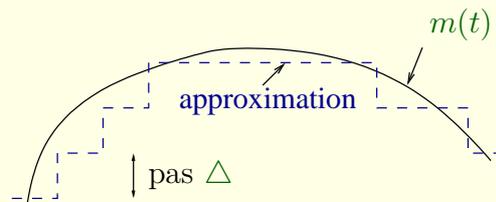


FIGURE 5.15 – L'erreur de quantification dans une modulation delta.

Chapitre 6

Transmission de signaux numériques en bande de base

6.1 Nécessité du codage

Le problème central des transmissions numériques est le suivant : on désire transmettre une information numérique sous une forme analogique optimale. À cette fin, on peut tout de suite différencier la notion de rapidité de modulation R , exprimée en *baud* de celle de débit binaire D . La rapidité de modulation R correspond au nombre de symboles transmis sur le canal par seconde ; elle est différente du débit de transmission dès qu'on associe plusieurs bits au même symbole transmis.

Définition 71 [Valence] La relation entre la rapidité de modulation R et le débit binaire D met en jeu la valence V ; elle est donnée par l'équation

$$D = R \log_2(V) \quad (6.1)$$

La valence est le nombre d'états distincts présents sur le canal de transmission.

Exemple. On peut décider d'associer à chaque groupe de 2 bits à transmettre une tension définie par le tableau 6.1.

00	$-5[V]$
01	$-2,5[V]$
11	$2,5[V]$
10	$5[V]$

TABLE 6.1 – Représentation physique de 4 états.

Ici, sur la ligne, on va trouver 4 états différents : la valence est donc $V = 4$. La relation entre le débit binaire D et la rapidité de modulation est

$$D = R \log_2(4) = 2R \quad (6.2)$$

La vitesse de transmission de symboles est donc deux fois plus faible que le débit binaire.

6.1.1 Bande passante requise

Soit à transmettre des données binaires au rythme f_b [b/s]. Chaque bit a donc une durée $T_b = 1/f_b$. La bande passante minimale pour transmettre ces données, appelée bande de NYQUIST est

$$\frac{1}{2T_b} = \frac{f_b}{2} \quad (6.3)$$

En effet, le canal se comporte comme un filtre passe-bas qui filtre le signal carré composé de signaux binaires de durée T_b . Ce signal est composé de plusieurs raies et il faut au minimum que le fondamental de ce signal soit transmis. Or la fréquence de ce fondamental est

$$\frac{1}{2T_b} = \frac{f_b}{2} \quad (6.4)$$

6.1.2 Capacité d'un canal en présence de bruit

En présence de bruit sur le canal, la transmission n'est pas parfaite et il convient dès lors de déterminer la proportion des symboles qui risquent d'être entachés d'erreur.

Pour chaque milieu de transmission, on peut chiffrer la capacité théorique du canal C en [b/s], c'est-à-dire le nombre maximum de bits par seconde qu'il est possible de transmettre avec une erreur aussi faible que voulue, exprimée à partir de la bande passante du canal $[0, W]$ et du rapport signal à bruit S/N à l'intérieur du canal. Cette relation est dite de SHANNON-HARTLEY.

Théorème 72 [Shannon-Hartley]

$$C = W \log_2 \left(1 + \frac{S}{N} \right) \quad (6.5)$$

Cette relation est illustrée à la figure 6.1.

Il faut veiller à ne pas confondre cette limite correspondant à la transmission d'un signal codé ou modulé en présence de bruit avec celle correspondant à la bande passante propre à une technique particulière de codage. En pratique, toutes les transmissions sont bruitées et c'est donc la capacité C qui fixe le maximum qu'il est possible d'atteindre.

6.1.3 Transmission des données binaires

Pour transmettre des informations (données, texte, image, son), on dispose de deux méthodes :

- la transmission en bande de base, méthode correspondant à l'émission directe sur le canal de transmission, et
- la transmission par modulation d'une porteuse, méthode permettant d'adapter le signal au canal de transmission.

Le signal numérique transmis possède un certain nombre de caractéristiques qu'il convient d'examiner :

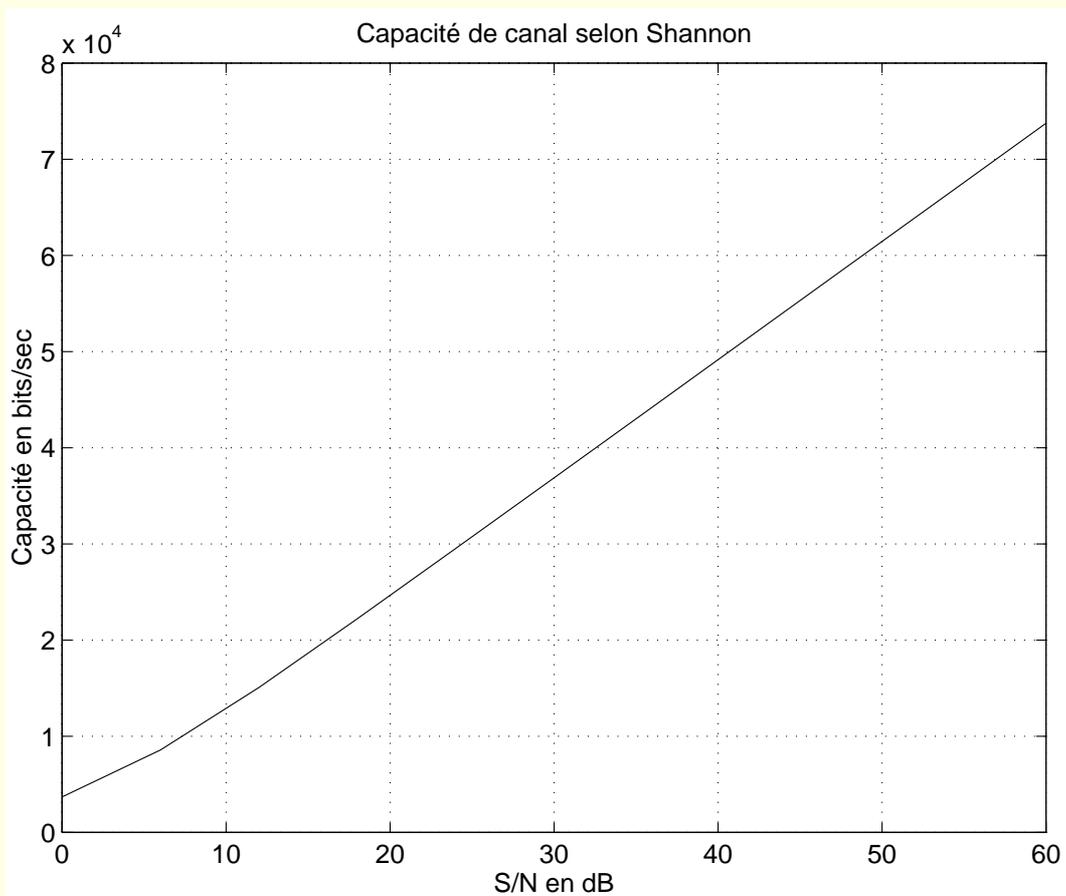


FIGURE 6.1 – Capacité d'un canal téléphonique ($W = 3,7 [kHz]$).

- l’encombrement spectral. La présence d’une raie à la fréquence nulle ou non, la présence d’une raie de synchronisation ou la valeur de la fréquence maximale. L’absence d’une raie à la fréquence nulle permet la transmission du signal sur un canal ne laissant pas passer la valeur moyenne. Par exemple, les lignes téléphoniques sont terminées par des transformateurs qui ne laissent pas passer le continu. La présence d’une raie à une fréquence multiple ou sous multiple de la fréquence d’émission des bits permet au récepteur de se synchroniser sur l’horloge d’émission.
- le débit, exprimé en $[b/s]$.
- les niveaux physiques associés à chaque bit (0 ou 1) ou groupes de bits. On peut transmettre les bits directement par des tensions ou alors faire un codage.

La notion d’efficacité spectrale, définie ci-après, fait le lien entre le débit binaire d’un signal d’information numérique et la bande occupée par la représentation analogique de ce signal.

Définition 73 *L’efficacité spectrale est définie comme le flux binaire par Hz. Il s’agit du rapport entre le débit binaire et la bande nécessaire à la transmission du signal d’information numérique.*

Le récepteur a pour tâche de déterminer la présence ou l’absence d’une impulsion pour chaque période. On verra que la détection d’une impulsion est une fonction de l’énergie contenue dans l’impulsion. Il y a donc intérêt à avoir une impulsion aussi large que possible, à la limite l’impulsion est de forme rectangulaire et elle couvre tout l’intervalle de temps T . On va donc non plus décrire s’il y a ou non une impulsion mais on va chercher à détecter des transitions dans le train de signaux rectangulaires.

6.2 Spectre des signaux numériques

Le principe d’une communication numérique consiste à maintenir une certaine valeur pendant un intervalle de temps fini, à l’inverse de l’analogique où le signal prend des valeurs continue. On devine que la discrétisation de l’axe temporel offre une meilleure résistance au bruit du signal à transmettre, car c’est après avoir analysé le signal sur toute la durée entre les échantillons que le récepteur prend une décision. Par contre, le numérique est plus coûteux en termes de bande passante.

6.2.1 Modèle théorique linéaire

En télécommunications numériques, on désire transmettre une série de valeurs d’information i_k , de probabilités respectives p_k . Pour ce faire, il faut mettre cette information sous la forme d’un signal électromagnétique analogique continu. On fait un usage fréquent du signal

$$g(t) = \sum_{k=-\infty}^{+\infty} A_k \phi_k(t - kT) \quad (6.6)$$

où $\phi_k(t)$ est une forme d’onde, nulle en dehors de l’intervalle $[0, T]$, et A_k une séquence de variables aléatoires éventuellement complexes ; la forme d’onde $\phi_k(t)$ et l’amplitude A_k sont associées à un symbole d’information i_k . Comme on peut le voir, il s’agit d’un modèle linéaire de la mise en forme de signaux PCM. Nous utiliserons ce modèle pour décrire un sous-ensemble

des techniques de transmission en bande de base, la modulation numérique ainsi que les codes linéaires.

Pour la simplicité, prenons une onde de mise en forme unique¹ $\phi_k(t) = \phi(t), \forall k$. La transformée de FOURIER de $g(t)$ est

$$\mathcal{G}(f) = \sum_{k=-\infty}^{+\infty} A_k \mathcal{F}\{\phi(t - kT)\} = \mathcal{F}\{\phi(t)\} \sum_{k=-\infty}^{+\infty} A_k e^{-2\pi j f k T} \quad (6.7)$$

$$= \Phi(f) \mathcal{F}\left\{ \sum_{k=-\infty}^{+\infty} A_k \delta(t - kT) \right\} \quad (6.8)$$

où $\Phi(f) = \mathcal{F}\{\phi(t)\}$.

Il s'agit donc du produit du spectre de l'onde de forme d'une part et du spectre d'une séquence de raies modulées en amplitude d'autre part. En d'autres termes, la transformée de FOURIER du signal issu de la mise en forme de signaux numériques résulte du passage d'un train d'impulsions de DIRAC à travers un filtre définissant la forme d'onde.

Par la suite, nous allons nous intéresser à la répartition fréquentielle de la puissance du signal –cette notion est appelée densité spectrale de puissance.

6.2.2 Train d'impulsions de Dirac

Intéressons-nous à la densité spectrale de puissance d'un train d'impulsions de DIRAC modulé en amplitude par une séquence de variables aléatoires

$$\sum_{k=-\infty}^{+\infty} A_k \delta(t - kT) \quad (6.9)$$

Tel quel ce signal n'est pas stationnaire. La stationnarité s'obtient en introduisant un délai T_0 uniformément réparti sur l'intervalle $[0, T]$, indépendant de la séquence A_k . Cela revient en pratique à considérer que la synchronisation n'est pas parfaite. Soit donc à analyser le signal aléatoire $X(t)$

$$X(t) = \sum_{k=-\infty}^{+\infty} A_k \delta(t - kT - T_0) \quad (6.10)$$

Hypothèse : on suppose en outre que la séquence A_k est stationnaire au sens large, de moyenne

$$\mu_A = E\{A_k\} \quad (6.11)$$

pour laquelle la fonction d'autocorrélation vaut²

$$\Gamma_{AA}(k, k-l) = E\{A_k A_{k-l}\} = \Gamma_{AA}(l) \quad (6.12)$$

1. Dès lors que la forme d'onde n'est plus unique, il n'est plus possible de traiter l'ensemble des modulations numériques par ce modèle. Ainsi, l'onde de mise en forme est différente pour chaque symbole dans le cas d'une modulation numérique non cohérente.

2. En toute rigueur, on définit la fonction d'autocorrélation en faisant intervenir le complexe conjugué, c'est-à-dire que $\Gamma_{AA}(k, k-l) = E\{A_k A_{k-l}^*\} = \Gamma_{AA}(l)$. Comme nous ne considérons que des signaux réels, la notation a été volontairement simplifiée.

Moyenne d'un train d'impulsions

Dans un premier temps, on fait l'hypothèse statistique que T_0 prend une valeur déterminée t_0 . La moyenne conditionnelle de $X(t)$ vaut

$$\mu_{X|T_0}(t|t_0) = \sum_{k=-\infty}^{+\infty} \mu_A \delta(t - kT - t_0) \quad (6.13)$$

Dans un deuxième temps, on calcule la moyenne non conditionnelle. La densité de probabilité de la variable T_0 étant donnée par

$$f_{T_0}(t_0) = \begin{cases} \frac{1}{T} & \text{si } 0 \leq t_0 \leq T \\ 0 & \text{sinon} \end{cases} \quad (6.14)$$

on a

$$\mu_X = \int_{-\infty}^{+\infty} \mu_{X|T_0}(t|t_0) f_{T_0}(t_0) dt_0 \quad (6.15)$$

$$= \frac{\mu_A}{T} \sum_{k=-\infty}^{+\infty} \int_0^T \delta(t - kT - t_0) dt_0 \quad (6.16)$$

En effectuant le changement de variable $z = t - kT - t_0$,

$$\mu_X = \frac{\mu_A}{T} \sum_{k=-\infty}^{+\infty} \int_{t-kT}^{t-(k+1)T} \delta(z) (-dz) \quad (6.17)$$

$$= \frac{\mu_A}{T} \sum_{k=-\infty}^{+\infty} \int_{t-(k+1)T}^{t-kT} \delta(z) dz \quad (6.18)$$

$$= \frac{\mu_A}{T} \int_{-\infty}^{+\infty} \delta(z) dz \quad (6.19)$$

$$= \frac{\mu_A}{T} \quad (6.20)$$

La moyenne d'un train d'impulsions modulé en amplitude par une séquence aléatoire stationnaire au sens large est donc liée à la moyenne de la séquence et est inversement proportionnelle à la durée de l'onde.

Fonction d'autocorrélation d'un train d'impulsions

Comme précédemment, on ajoute un délai aléatoire T_0 et on procède d'abord sous l'hypothèse statistique que $T_0 = t_0$ pour calculer la fonction d'autocorrélation. Dès lors

$$E \{X(t)X(t - \tau)|t_0\} = \sum_{k=-\infty}^{+\infty} \sum_{k'=-\infty}^{+\infty} \Gamma_{AA}(k - k') \delta(t - kT - t_0) \delta(t - \tau - k'T - t_0) \quad (6.21)$$

Puis on calcule l'espérance mathématique non conditionnelle

$$\begin{aligned} \Gamma_{XX}(\tau) &= \sum_{k=-\infty}^{+\infty} \sum_{k'=-\infty}^{+\infty} \Gamma_{AA}(k - k') \int_0^T \frac{1}{T} \delta(t - kT - t_0) \delta(t - \tau - k'T - t_0) dt_0 \\ &= \sum_{k=-\infty}^{+\infty} \sum_{k'=-\infty}^{+\infty} \frac{\Gamma_{AA}(k - k')}{T} \int_{t-(k+1)T}^{t-kT} \delta(z) \delta(z - \tau + (k - k')T) dz \end{aligned} \quad (6.22)$$

On pose ensuite $k - k' = l$ et, dès lors,

$$\Gamma_{XX}(\tau) = \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} \frac{\Gamma_{AA}(l)}{T} \int_{t-(k+1)T}^{t-kT} \delta(z)\delta(z - \tau + lT)dz \quad (6.23)$$

$$= \sum_{l=-\infty}^{+\infty} \frac{\Gamma_{AA}(l)}{T} \sum_{k=-\infty}^{+\infty} \int_{t-(k+1)T}^{t-kT} \delta(z)\delta(z - \tau + lT)dz \quad (6.24)$$

$$= \sum_{l=-\infty}^{+\infty} \frac{\Gamma_{AA}(l)}{T} \int_{-\infty}^{+\infty} \delta(z)\delta(z - \tau + lT)dz \quad (6.25)$$

$$= \sum_{l=-\infty}^{+\infty} \frac{\Gamma_{AA}(l)}{T} \delta(\tau - lT) \quad (6.26)$$

Cette expression présente des similitudes avec le théorème de l'échantillonnage. C'est assez logiquement qu'on trouve que la fonction d'autocorrélation d'un signal aléatoire échantillonné $X(t)$ est elle-même échantillonnée. On pourrait parler d'un théorème sur l'échantillonnage des fonctions aléatoires.

Densité spectrale d'un train d'impulsions

La densité spectrale de puissance est donnée par la transformée de FOURIER de l'expression (6.26) :

$$\gamma_X(f) = \sum_{l=-\infty}^{+\infty} \frac{\Gamma_{AA}(l)}{T} e^{-2\pi j f l T} \quad (6.27)$$

qui n'est autre que la densité spectrale de la séquence A_k à un facteur $1/T$ près.

Densité spectrale d'un signal numérique mis en forme

Nous avons vu que le signal $g(t)$

$$g(t) = \sum_{k=-\infty}^{+\infty} A_k \phi(t - kT) \quad (6.28)$$

n'était ni plus ni moins que le résultat d'un train d'impulsions modulés en amplitude par la séquence A_k au travers d'un filtre de mise en forme. Par application des résultats relatifs au passage d'un signal aléatoire à travers un filtre linéaire, la moyenne du signal aléatoire $g(t)$ vaut

$$\mu_g = \frac{\mu_A}{T} \Phi(0) \quad (6.29)$$

De même, par application du théorème de WIENER-KINTCHINE,

$$\gamma_g(f) = \|\Phi(f)\|^2 \sum_{l=-\infty}^{+\infty} \frac{\Gamma_{AA}(l)}{T} e^{-2\pi j f l T} \quad (6.30)$$

$$= \|\Phi(f)\|^2 \frac{\gamma_A(fT)}{T} \quad (6.31)$$

où

$$\gamma_A(fT) = \sum_{l=-\infty}^{+\infty} \Gamma_{AA}(l) e^{-2\pi j f l T} \quad (6.32)$$

est la densité spectrale de puissance de la séquence aléatoire A_k . On voit donc que $\gamma_g(f)$ est déterminée par la forme d'impulsion et par la corrélation existant entre symboles A_k successifs.

Cas particulier 1 : impulsions rectangulaires

Dans le cas particulier d'impulsions rectangulaires de durée T , la fonction $\phi(t)$ est

$$\phi(t) = \begin{cases} 1 & \text{si } t \in [0, T] \\ 0 & \text{sinon} \end{cases} \quad (6.33)$$

on a

$$\|\Phi(f)\|^2 = T^2 \left(\frac{\sin(\pi f T)}{\pi f T} \right)^2 \quad (6.34)$$

Cas particulier 2 : signaux non-corrélés

Un autre cas important est celui d'une séquence A_k de variables aléatoires non-corrélées³ : $C\{A_k A_{k-l}\} = E\{(A_k - \mu_A)(A_{k-l} - \mu_A)\} = \Gamma_{AA}(l) - \mu_A^2 = 0, \forall l \neq 0$. Autrement dit, $\Gamma_{AA}(l \neq 0) = \mu_A^2$.

D'autre part, par définition, $\sigma_A^2 = E\{(A_k - \mu_A)(A_k - \mu_A)\} = E\{A_k A_k\} - \mu_A^2$. D'où, $\Gamma_{AA}(0) = E\{A_k A_k\} = \sigma_A^2 + \mu_A^2$.

On en déduit que

$$\gamma_A(fT) = \sum_{l=-\infty}^{+\infty} \Gamma_{AA}(l) e^{-2\pi j l f T} \quad (6.35)$$

$$= \sigma_A^2 + \mu_A^2 \sum_{l=-\infty}^{+\infty} e^{-2\pi j l f T} \quad (6.36)$$

$$= \sigma_A^2 + \mu_A^2 \sum_{m=-\infty}^{+\infty} \frac{1}{T} \delta(f - \frac{m}{T}) \quad (6.37)$$

La dernière relation a été obtenue en vertu de la relation suivante : $\sum_{l=-\infty}^{+\infty} e^{-2\pi j l f T} = \sum_{m=-\infty}^{+\infty} \frac{1}{T} \delta(f - \frac{m}{T})$. En effet, $\sum_{l=-\infty}^{+\infty} e^{-2\pi j l f T} = 1 + \sum_{m=1}^{+\infty} 2 \cos(2\pi f m T)$. On prend ensuite la transformée de FOURIER de ce signal $\delta(f') + \sum_{m=1}^{+\infty} [\delta(f' - mT) + \delta(f' + mT)] = \sum_{m=-\infty}^{+\infty} \delta(f' - mT)$. Ce qui par transformée inverse fournit $\sum_{m=-\infty}^{+\infty} \frac{1}{T} \delta(f - \frac{m}{T})$.

La densité spectrale vaut donc finalement

$$\gamma_g(f) = \|\Phi(f)\|^2 \frac{1}{T} \left[\sigma_A^2 + \mu_A^2 \sum_{m=-\infty}^{+\infty} \frac{1}{T} \delta(f - \frac{m}{T}) \right] \quad (6.38)$$

Le second terme du membre de droite fait intervenir une série de raies. On peut le voir comme la composante périodique de la densité spectrale de puissance du signal mis en forme.

Moyennant l'hypothèse de variables à moyenne nulle ($\mu_A = 0$),

$$\gamma_g(f) = \|\Phi(f)\|^2 \frac{\sigma_A^2}{T} \quad (6.39)$$

c'est-à-dire que la densité spectrale est proportionnelle au carré de la transformée de FOURIER de la forme d'onde élémentaire. En particulier, pour une forme d'onde rectangulaire,

$$\gamma_g(f) = T \sigma_A^2 \left(\frac{\sin(\pi f T)}{\pi f T} \right)^2 \quad (6.40)$$

3. C'est bien évidemment dans le cas d'une séquence de A_k non-corrélées que l'information apportée est la plus importante mais elle est moins robuste en présence de bruit.

6.3 Transmission d'impulsions en bande de base

6.3.1 Codage

Deux technologies sont mises en œuvre pour coder les signaux numériques :

- les codages en ligne, qui impliquent la conversion en temps réel des bits d'information (par exemple, NRZ-L, NRZI, biphasé, ...). On parle de codes linéaires.
- les codages complets, qui se réfèrent à des tables de conversion (par exemple 5B/4T, 4B/3T, 2B1Q).

6.3.2 Codes en ligne d'émission

On peut distinguer les principales catégories suivantes pour le codage linéaire de signaux PCM :

1. Non-return-to-zero (NRZ),
2. Return-to-zero (RZ),
3. Codage de la phase, et
4. Codage multi-niveaux.

La comparaison de ces codes porte tant sur leurs performances que sur leur occupation spectrale. Dans un premier temps, nous allons déterminer l'occupation spectrale.

Codage NRZ unipolaire

La manière la plus simple de transmettre un signal est d'utiliser deux niveaux de tension pour représenter les deux états binaires 0 et 1. La présence d'une tension peut, par exemple, représenter un 1 tandis que l'absence d'une tension représente un 0. C'est le cas du codage NRZ-L (Non Return to Zero Level) unipolaire. Cette technique, ainsi que d'autres, est illustrée à la figure 6.2.

La modélisation complète du codage NRZ unipolaire est résumée dans le tableau suivant :

Symbole	Probabilité	A_k	Onde
0	$1 - p$	0	...
1	p	V	$1, 0 \leq t < T$

Pour calculer la densité spectrale de puissance du signal mis en forme, on calcule la moyenne, la variance ainsi que la transformée de FOURIER du signal de mise en forme.

Moyenne

$$\mu_A = Vp \quad (6.41)$$

Variance

$$\sigma_A^2 = E \left\{ (A_k - \mu_A)^2 \right\} = E \left\{ A_k^2 \right\} - \mu_A^2 = pV^2 - (pV)^2 = p(1 - p)V^2 \quad (6.42)$$

Filtre

$$\|\Phi(f)\|^2 = T^2 \left(\frac{\sin(\pi fT)}{\pi fT} \right)^2 \quad (6.43)$$

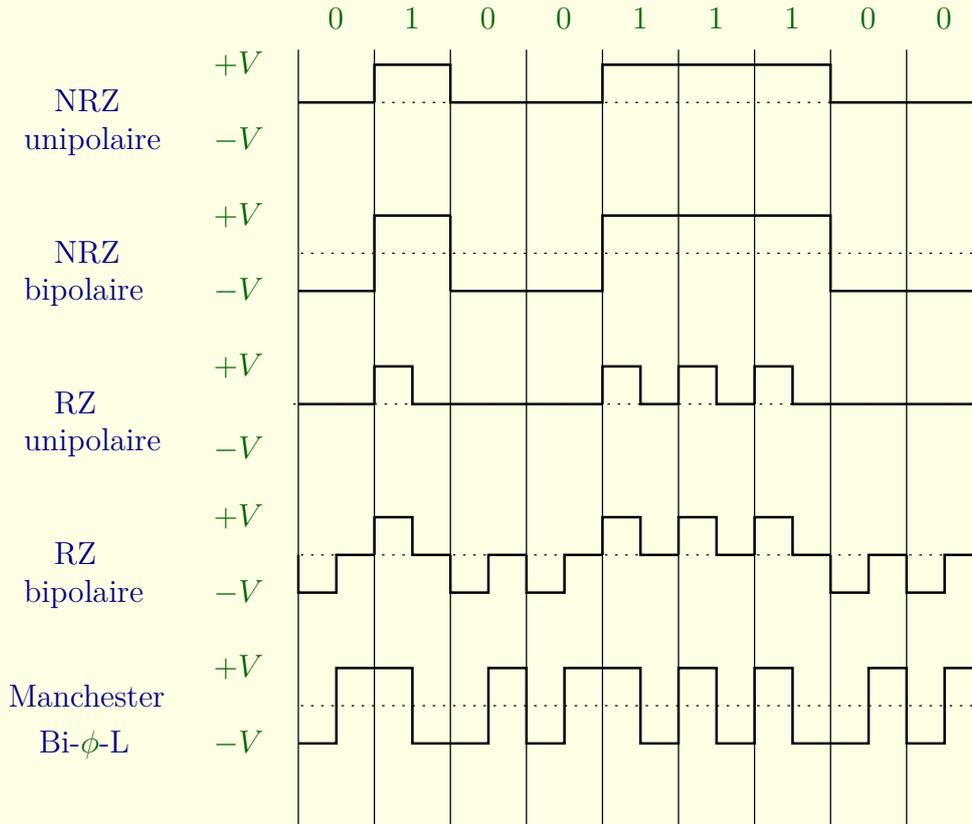


FIGURE 6.2 – Variantes de codage en ligne PCM.

Comme les symboles sont non-corrélés, il advient

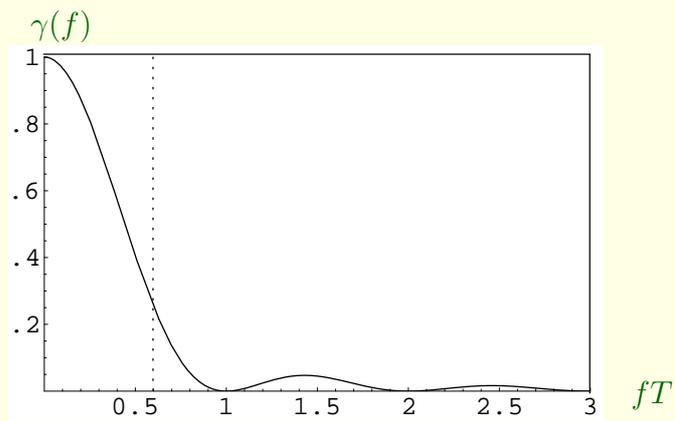
$$\gamma(f) = p(1-p)V^2T \left(\frac{\sin(\pi fT)}{\pi fT} \right)^2 + p^2V^2\delta(f) \quad (6.44)$$

La densité spectrale est donc en forme de sinus cardinal avec une composante continue importante⁴. Le fait que $\gamma(f)$ soit nulle pour $f = k/T$ rend la synchronisation difficile car il n'y a pas de composante rythmée à la cadence d'émission binaire.

Bande passante. Le spectre du signal mis en forme est infini. En théorie donc, il faudrait une bande passante infinie, ce qui n'est pas acceptable. Afin de définir une largeur pratique de bande, examinons les graphes de la figure 6.3.

Le premier graphe montre la densité spectrale $\gamma(f)$. Le second reprend la valeur de la puissance cumulée comprise dans la bande $[0, fT]$. Cette puissance vaut 84%, 91,3% et 95% respectivement pour $fT = 0,6$, $fT = 1$ et $fT = 2$. Un choix pratique consiste à prendre une largeur de bande égale à $\frac{0,6}{T}$.

4. Il faut distinguer la composante continue de la densité spectrale de puissance avec la moyenne temporelle. Ces deux notions ne sont pas équivalentes!



Pourcentage de puissance comprise dans la bande

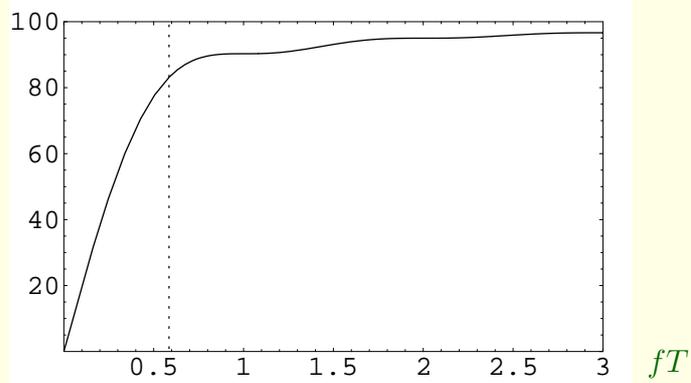


FIGURE 6.3 – Analyse de la répartition de puissance.

Codage NRZ bipolaire

Le codage NRZ bipolaire s'étudie suivant la même démarche à partir de la représentation suivante :

Symbole	Probabilité	A_k	Onde
0	$1 - p$	$-V$	$1, 0 \leq t < T$
1	p	V	$1, 0 \leq t < T$

Moyenne

$$\mu_A = Vp - V(1 - p) = (2p - 1)V \quad (6.45)$$

Variance

$$\sigma_A^2 = E\{A_k^2\} - \mu_A^2 \quad (6.46)$$

$$= V^2p + V^2(1 - p) - V^2(2p - 1)^2 \quad (6.47)$$

$$= 4p(1 - p)V^2 \quad (6.48)$$

Filtre

$$\|\Phi(f)\|^2 = T^2 \left(\frac{\sin(\pi fT)}{\pi fT} \right)^2 \quad (6.49)$$

Comme les symboles sont non-corrélés, il advient

$$\begin{aligned} \gamma(f) &= 4p(1 - p)V^2T \left(\frac{\sin(\pi fT)}{\pi fT} \right)^2 + (2p - 1)^2V^2 \left(\frac{\sin(\pi fT)}{\pi fT} \right)^2 \sum_{m=-\infty}^{+\infty} \delta(f - \frac{m}{T}) \\ &= 4p(1 - p)V^2T \left(\frac{\sin(\pi fT)}{\pi fT} \right)^2 + (2p - 1)^2V^2 \sum_{m=-\infty}^{+\infty} \left(\frac{\sin(\pi m)}{\pi m} \right)^2 \delta(f - \frac{m}{T}) \\ &= 4p(1 - p)V^2T \left(\frac{\sin(\pi fT)}{\pi fT} \right)^2 + (2p - 1)^2V^2\delta(f) \end{aligned} \quad (6.50)$$

Dans le cas fort théorique de deux symboles totalement équiprobables, il n'y a pas de composante continue dans le spectre de puissance due à la présence du $\delta(f)$. De plus, on observe que la densité spectrale se s'annule pas pour $f = 0$, ce qui traduit la présence d'une composante continue systématique. Ce n'est donc pas parce la moyenne du niveau de tension vaut 0 (lorsque $p = \frac{1}{2}$) que la densité spectrale n'a pas de composante continue! Remarquons aussi qu'en toute logique, le niveau moyen de puissance d'un signal NRZ bipolaire est supérieur de 6 [dB] à celui d'un codage NRZ unipolaire.

Le codage NRZ a la particularité de représenter une longue série de 0 par un signal continu, ce qui le rend sensible aux bruits parasites. De plus, la synchronisation sur les variations peut facilement être perdue lorsqu'il n'y a pas de variation pendant un certain laps de temps.

Le codage NRZ-M (Non Return to Zero Mark) est une variante du codage NRZ-L qui consiste à inverser la polarité à chaque valeur binaire 1. Il permet d'éviter l'absence de transitions, à condition d'assurer également l'absence de rafales de 0.

Codage RZ unipolaire

Comme son nom l'indique, le codage RZ ramène le signal à une tension nulle pendant tout intervalle T . On a ainsi la certitude d'avoir régulièrement des transitions, mais il faut en payer le prix car la présence de deux transitions par période T entraîne le doublement de la bande passante ce que confirme le développement théorique suivant.

Soit le signal de mise en forme que voici :

Symbole	Probabilité	A_k	Onde
0	$1 - p$	0	...
1	p	V	pour $t \in [0, \alpha T]$ ($\alpha \leq 1$)

Moyenne

$$\mu_A = Vp \quad (6.51)$$

Variance

$$\sigma_A^2 = E\{A_k^2\} - \mu_A^2 = pV^2 - (pV)^2 = p(1-p)V^2 \quad (6.52)$$

Filtre

$$\|\Phi(f)\|^2 = \alpha^2 T^2 \left(\frac{\sin(\pi f \alpha T)}{\pi f \alpha T} \right)^2 \quad (6.53)$$

Comme les symboles sont non-corrélés, il advient

$$\gamma(f) = p(1-p)V^2\alpha^2 T \left(\frac{\sin(\pi f \alpha T)}{\pi f \alpha T} \right)^2 + p^2 V^2 \alpha^2 \sum_{m=-\infty}^{+\infty} \left(\frac{\sin(\pi m \alpha)}{\pi m \alpha} \right)^2 \delta\left(f - \frac{m}{T}\right) \quad (6.54)$$

Si on prend $\alpha = \frac{1}{2}$, ce qui est le choix habituel,

$$\begin{aligned} \gamma(f) &= \frac{p(1-p)}{4} V^2 T \left(\frac{\sin(\pi f T/2)}{\pi f T/2} \right)^2 + \frac{p^2}{4} V^2 \delta(f) \\ &\quad + \frac{p^2}{4} V^2 \sum_{n=-\infty}^{+\infty} \left(\frac{\sin((2n+1)\pi/2)}{(2n+1)\pi/2} \right)^2 \delta\left(f - \frac{(2n+1)}{T}\right) \end{aligned} \quad (6.55)$$

$$\begin{aligned} &= \frac{p(1-p)}{4} V^2 T \left(\frac{\sin(\pi f T/2)}{\pi f T/2} \right)^2 + \frac{p^2}{4} V^2 \delta(f) \\ &\quad + p^2 V^2 \sum_{n=-\infty}^{+\infty} \frac{1}{(2n+1)^2 \pi^2} \delta\left(f - \frac{(2n+1)}{T}\right) \end{aligned} \quad (6.56)$$

On constate l'existence de raies à des multiples impairs de $1/T$. La première de ces raies peut servir à la synchronisation. D'autre part, il faut une bande passante 2 fois plus large.

Codage RZ bipolaire

Symbole	Probabilité	A_k	Onde
0	$1 - p$	$-V$	pour $t \in [0, \alpha T]$ ($\alpha \leq 1$)
1	p	V	pour $t \in [0, \alpha T]$ ($\alpha \leq 1$)

Moyenne

$$\mu_A = (2p - 1)V \quad (6.57)$$

Variance

$$\sigma_A^2 = 4p(1 - p)V^2 \quad (6.58)$$

Filtre

$$\|\Phi(f)\|^2 = \alpha^2 T^2 \left(\frac{\sin(\pi f \alpha T)}{\pi f \alpha T} \right)^2 \quad (6.59)$$

$$\begin{aligned} \gamma(f) = & 4p(1 - p)V^2 \alpha^2 T \left(\frac{\sin(\pi f \alpha T)}{\pi f \alpha T} \right)^2 + \frac{(2p - 1)^2}{4} V^2 \delta(f) + \\ & + (2p - 1)^2 V^2 \alpha^2 \sum_{m=-\infty}^{+\infty} \left(\frac{\sin(\pi m \alpha)}{\pi m \alpha} \right)^2 \delta\left(f - \frac{m}{T}\right) \end{aligned} \quad (6.60)$$

À nouveau, pour $\alpha = \frac{1}{2}$,

$$\gamma(f) = p(1 - p)V^2 T \left(\frac{\sin(\pi f T/2)}{\pi f T/2} \right)^2 + \frac{(2p - 1)^2}{4} V^2 \delta(f) \quad (6.61)$$

$$+ (2p - 1)^2 V^2 \sum_{n=-\infty}^{+\infty} \frac{1}{(2n + 1)^2 \pi^2} \delta\left(f - \frac{(2n + 1)}{T}\right) \quad (6.62)$$

Codage Manchester

Le code MANCHESTER est de type biphasé, c'est-à-dire qu'un bit d'information est représenté par une transition de tension en milieu de période. La valeur 0 est représentée par un front montant, et la valeur 1 par un front descendant. Cette transition sert également de référence pour l'horloge.

Symbole	Probabilité	A_k	Onde
0	$1 - p$	$-V$	$-1, 0 \leq t < T/2$ $+1, T/2 \leq t < T$
1	p	V	$+1, 0 \leq t < T/2$ $-1, T/2 \leq t < T$

Moyenne

$$\mu_A = (2p - 1)V \quad (6.63)$$

Variance

$$\sigma_A^2 = 4p(1 - p)V^2 \quad (6.64)$$

Filtre

$$\Phi(f) = \frac{T}{2} \left(\frac{\sin(\pi f T/2)}{\pi f T/2} \right) (-e^{-2\pi j f T/4} + e^{+2\pi j f T/4}) = T \left(\frac{\sin(\pi f T/2)}{\pi f T/2} \right) j \sin(\pi f T/2) \quad (6.65)$$

$$\|\Phi(f)\|^2 = T^2 \left(\frac{\sin(\pi f T/2)}{\pi f T/2} \right)^2 \sin^2(\pi f T/2) \quad (6.66)$$

$$\begin{aligned} \gamma(f) &= 4p(1 - p)V^2 T \left(\frac{\sin^4(\pi f T/2)}{(\pi f T/2)^2} \right) + (2p - 1)^2 V^2 \sum_{m=-\infty}^{+\infty} \left(\frac{\sin^4(m\pi/2)}{(m\pi/2)^2} \right) \delta\left(f - \frac{m}{T}\right) \\ &= 4p(1 - p)V^2 T \left(\frac{\sin^4(\pi f T/2)}{(\pi f T/2)^2} \right) + 4(2p - 1)^2 V^2 \sum_{n=-\infty}^{+\infty} \frac{1}{(2n + 1)^2 \pi^2} \delta\left(f - \frac{(2n + 1)}{T}\right) \end{aligned}$$

Les codages biphasés ont la particularité de nécessiter deux transitions pour un bit d'information (1B/2T). La fréquence du signal doit donc être le double de la bande passante offerte au réseau local. Par exemple, les composants d'un réseau **Ethernet**, qui utilisent un codage MANCHESTER, sont cadencés à 20 [MHz] afin d'offrir le 10 [Mb/s] de débit.

Comparaison des codes

Le tableau 6.2 reprend les spectres des signaux précités. Il permet de comparer l'utilisation de la bande passante, la présence ou non de raies à la fréquence d'émission et l'amplitude absolue des spectres.

Autres techniques de codage en ligne

Le codage de MANCHESTER différentiel est une variante du précédent. Il consiste à assurer deux transitions pour un symbole 0 et une seule pour un symbole 1. Le standard Token-Ring 802.5 y recourt.

On peut également citer le codage HDB3 (High-Density Bi-polar modulus 3) qui est utilisé sur les liaisons de modulation par impulsions codées des réseaux téléphoniques. Il a pour caractéristique de limiter à trois le nombre de zéros successifs.

De nouveaux codages liés aux technologies xDSL (x Digital Subscriber Loop) sont apparues récemment ; il s'agit notamment de CAP (Carrierless Amplitude Phase) dans les versions CAP12, CAP16 et CAP64, de QAM25 (Quadrature Amplitude Modulation) et PAM5.

En fait, il existe de très nombreuses variantes de codage en ligne d'émission.

Modulation	Domaine temporel	Domaine spectral
NRZ unipolaire		
NRZ bipolaire		
RZ unipolaire		
MANCHESTER		

TABLE 6.2 – Comparaison des techniques de modulation en bande de base.

Codage en blocs ou complets

De par leur nature, ces codages supposent une meilleure qualité de transmission, notamment obtenue par fibre optique, car ils sont sensibles aux parasites. Les signaux utilisés sont à trois états (ternaires) ou à quatre états (quaternaires). Le codage 4B/3T permet ainsi de coder 4 bits d'information en 3 signaux ternaires. Il est, par exemple, utilisé sur le bus S0 du RNIS.

La codage 2B/1Q permet de coder 2 bits d'information en un signal quaternaire. Il est utilisé sur le bus S0 du RNIS aux États-Unis. Le tableau 6.3 présente le système employé.

Valeur binaire	Code quaternaire
10	+3
11	+1
01	-1
00	-3

TABLE 6.3 – Codage 2B/1Q.

Le codage 4B/5T, plus performant, est utilisé dans les réseaux FDDI. Un débit de 100 [Mb/s] peut donc être véhiculé par un signal à 125 [MHz]. Cette technique consiste à représenter 4 bits consécutifs par un signal composé de 5 transitions.

Les codages complets entraînent une implémentation plus complexe que les codages en ligne car ils sont plus sensibles à la diaphonie et offrent un rapport signal à bruit moins élevé. Par contre, ils permettent de diminuer la fréquence du signal de transmission grâce à une plus grande redondance d'information. Le codage 1B/1Q permet ainsi d'offrir une bande passante double de la fréquence du signal, et donc de diminuer le coût des composants.

La qualité d'un codage se mesure entre autres par son taux d'interférence entre symboles transmis. L'interférence provient du fait que la forme de l'onde d'un symbole affecte la suivante. Ainsi le 2B/1Q a un plus grand taux d'interférence que le 4B/3T car la forme des signaux est moins riche.

Les codages complets sont généralement associés à des codages en ligne, notamment sur des supports en cuivre qui sont sujets aux perturbations électromagnétiques. Par exemple, les réseaux ATM utilisent une association 4B/5T+NRZI.

6.4 Détection de signaux binaires en présence de bruit gaussien

6.4.1 Position du problème

Après avoir été transformé sous la forme d'une onde électrique, le signal d'information $g(t)$ est transmis via le canal. À la réception, il s'agit donc de retrouver le signal d'émission

$$g(t) = \sum_{k=-\infty}^{+\infty} A_k \phi_k(t - kT) \quad (6.67)$$

auquel est venu se superposer un bruit de transmission. Dans le cas d'une forme d'onde unique, le signal $g(t)$ s'exprime comme la mise en forme par un filtre de transmittance $\Phi(f)$ d'un train

d'impulsions qui n'est autre qu'une série de valeurs aléatoires échantillonnées à une cadence T . C'est donc cette information que l'on cherche à récupérer.

La figure 6.4 montre l'ensemble d'une chaîne de transmission numérique. C'est sur base de ce modèle que nous travaillerons par la suite.

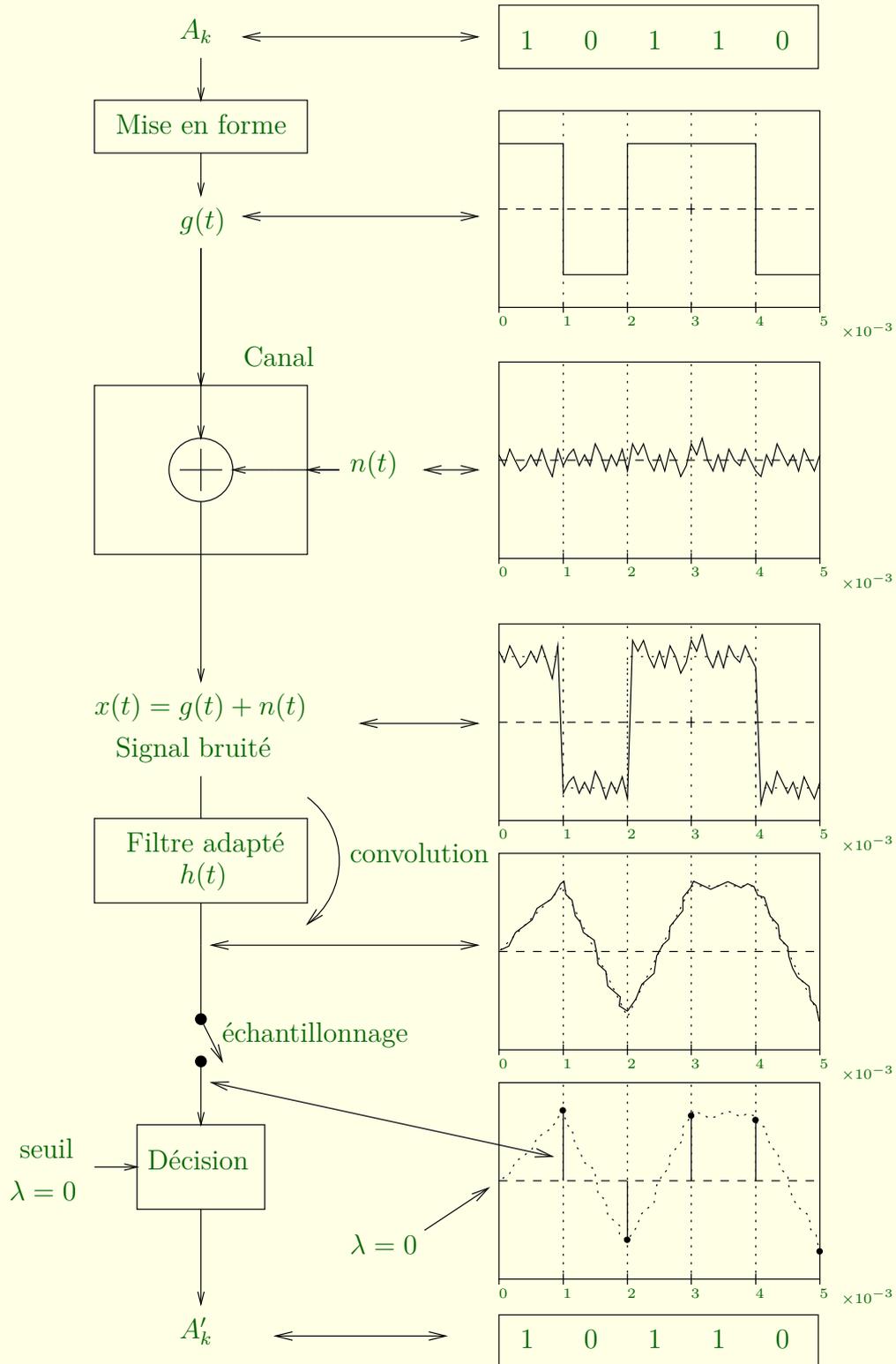


FIGURE 6.4 – Structure d'une chaîne de transmission numérique complète.

6.4.2 Problème simplifié : le cas d'un système binaire

Pour la facilité, nous nous limiterons à un système linéaire binaire. Un tel système considère deux ondes sur la durée T , notées $g_0(t)$ et $g_1(t)$. Le signal transmis durant l'intervalle de temps T est dès lors

$$g(t) = \begin{cases} g_0(t), & 0 \leq t \leq T & \text{pour un signal binaire 0} \\ g_1(t), & 0 \leq t \leq T & \text{pour un signal binaire 1} \end{cases} \quad (6.68)$$

$g_0(t)$ et $g_1(t)$ peuvent bien évidemment résulter de l'application d'une forme d'onde unique, mais ce n'est pas obligatoire.

6.4.3 Modèle de canal et structure du récepteur

Le modèle du canal retenu dans cette étude considère que le signal $x(t)$ reçu par le détecteur est

$$x(t) = g(t) + n(t) \quad (6.69)$$

où $n(t)$ représente un bruit additif gaussien. La figure 6.5 montre un signal de type MANCHESTER affecté par un bruit. Ce modèle est réaliste pour toute une série de phénomènes dont ceux de propagation en espace libre.

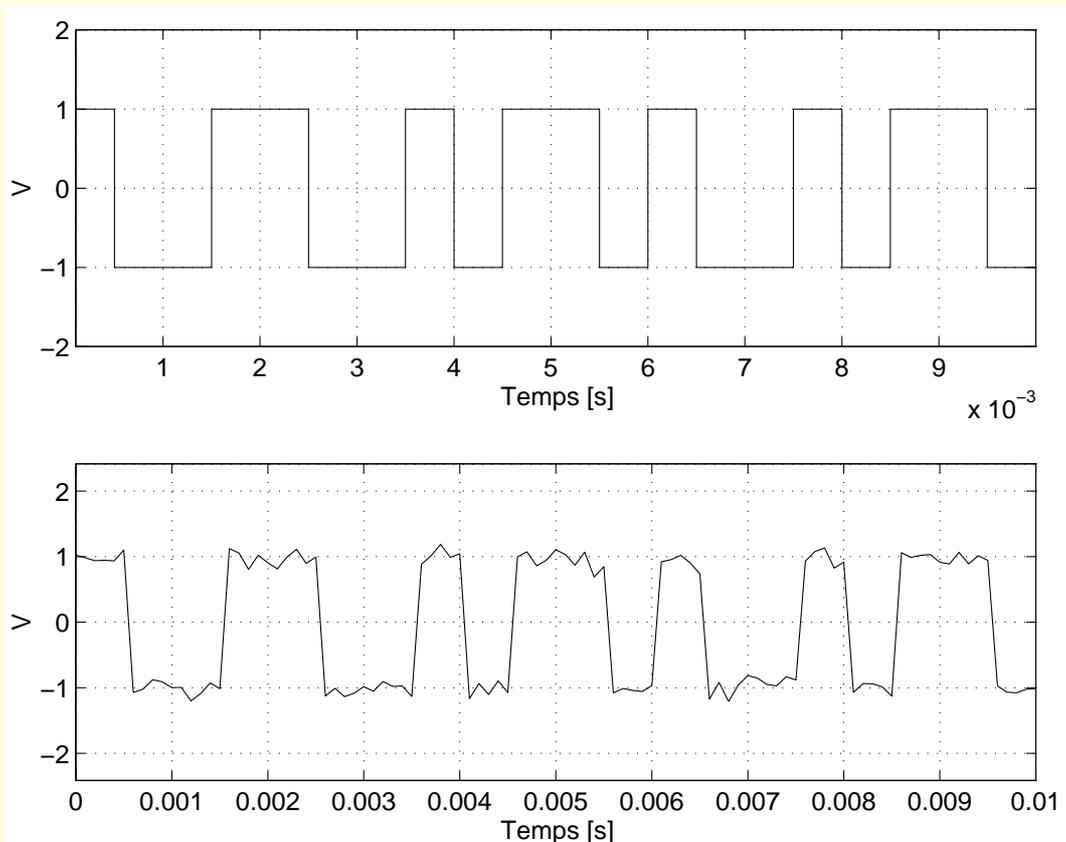


FIGURE 6.5 – Effet du bruit sur un signal MANCHESTER : en haut le signal original, en bas le signal bruité.

Le récepteur doit retrouver, à partir de l'observation du signal bruité, la valeur des variables aléatoires échantillonnées. S'il se trompe, l'erreur ne pourra être corrigée sans artifice supplémentaire comme par l'ajout de redondance. C'est donc en termes de probabilité d'erreur qu'un récepteur numérique s'analyse. Le taux d'erreur sera fonction de la puissance relative du signal utile et du bruit, ainsi que de la forme de l'onde constitutive.

La figure 6.6 illustre les deux phases qui interviennent dans le processus de détection. La première phase consiste à échantillonner le signal à une cadence T , après avoir filtré le signal reçu ; on obtient alors une suite aléatoire $y[T]$ à partir du signal continu $y(t)$. Pendant la période T précédant l'échantillonnage, le récepteur compare le signal à toutes les valeurs d'émission possibles, ce qui peut se faire de plusieurs manières : un filtre linéaire suivi d'un échantillonneur, par un filtre adapté ou un corrélateur. Les conditions initiales sont remises à zéro avant l'arrivée de chacun des symboles.

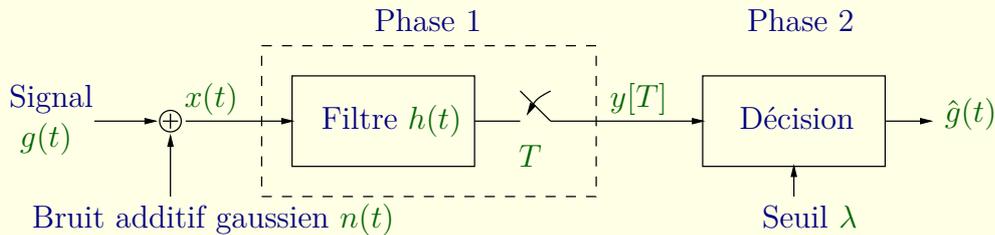


FIGURE 6.6 – Structure d'un détecteur linéaire.

6.4.4 Première phase : filtrage ou corrélation

Le signal de bruit $n(t)$ est un processus de bruit blanc à moyenne nulle et de densité spectrale de puissance $N_0/2$. On suppose que le récepteur connaît la forme des ondes reçues de sorte que son comportement a pu être optimisé en fonction de ces ondes. L'incertitude résulte de l'addition du bruit $n(t)$. La fonction du récepteur consiste à détecter la nature de l'impulsion $g(t)$ sur base du signal $x(t)$.

Le filtre étant supposé linéaire, la sortie peut s'exprimer sous la forme

$$y(t) = g_h(t) + n_h(t) \quad (6.70)$$

où $g_h(t)$ et $n_h(t)$ représentent les contributions dues respectivement à $g(t)$ et $n(t)$. Le signal $y(t)$ est illustré à la figure 6.7.

Filtre adapté

Une vue réaliste du problème consiste à exiger qu'à l'instant d'échantillonnage, la composante $g_h(t)$ soit la plus grande possible par rapport au bruit introduit par le canal, comme résultat du filtrage. Cela revient à choisir un filtre qui maximisera le rapport de puissance suivant

$$\eta = \frac{|g_h(T)|^2}{E\{n_h^2(t)\}} \quad (6.71)$$

où $|g_h(T)|^2$ est la puissance instantanée du signal filtré à l'instant d'échantillonnage T et $E\{n_h^2(t)\} = \sigma_N^2$ mesure la puissance moyenne due au bruit à la sortie du filtre. Cherchons donc à effectuer un choix judicieux du filtre $h(t)$.

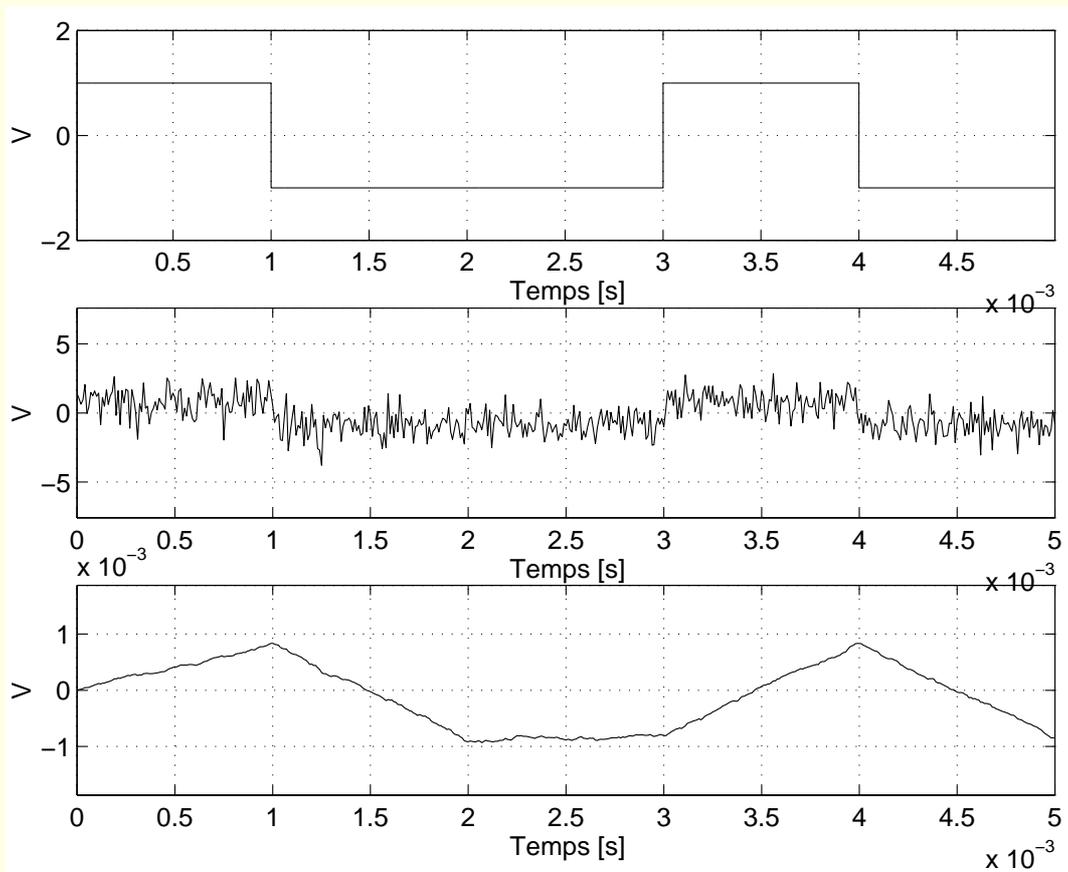


FIGURE 6.7 – Signaux intervenant au cours de la démodulation d'un signal numérique en bande de base : le signal original, le signal à l'entrée du récepteur et le signal à la sortie du filtre.

Comme

$$g_h(t) = \int_{-\infty}^{+\infty} \mathcal{H}(f)\mathcal{G}(f)e^{2\pi jft} df \quad (6.72)$$

le numérateur de η vaut donc

$$|g_h(T)|^2 = \left\| \int_{-\infty}^{+\infty} \mathcal{H}(f)\mathcal{G}(f)e^{2\pi jfT} df \right\|^2 \quad (6.73)$$

Il reste à déterminer le dénominateur de η : la puissance de bruit. On sait que

$$\gamma_{N_h}(f) = \frac{N_0}{2} \|\mathcal{H}(f)\|^2 \quad (6.74)$$

par application du théorème de WIENER-KINTCHINE. Dès lors

$$E \{n_h^2(t)\} = \int_{-\infty}^{+\infty} \gamma_{N_h}(f) df \quad (6.75)$$

$$= \frac{N_0}{2} \int_{-\infty}^{+\infty} \|\mathcal{H}(f)\|^2 df \quad (6.76)$$

L'inégalité de SCHWARZ établit que si

$$\int_{-\infty}^{+\infty} \|\phi_1(x)\|^2 dx < +\infty \quad (6.77)$$

et

$$\int_{-\infty}^{+\infty} \|\phi_2(x)\|^2 dx < +\infty \quad (6.78)$$

alors

$$\left\| \int_{-\infty}^{+\infty} \phi_1(x)\phi_2(x) dx \right\|^2 \leq \int_{-\infty}^{+\infty} \|\phi_1(x)\|^2 dx \int_{-\infty}^{+\infty} \|\phi_2(x)\|^2 dx \quad (6.79)$$

Par ailleurs, l'égalité tient à la condition que

$$\phi_1(x) = k\phi_2^*(x) \quad (6.80)$$

Dans le cas présent,

$$\left\| \int_{-\infty}^{+\infty} \mathcal{H}(f)\mathcal{G}(f)e^{2\pi jfT} df \right\|^2 \leq \int_{-\infty}^{+\infty} \|\mathcal{H}(f)\|^2 df \int_{-\infty}^{+\infty} \|\mathcal{G}(f)\|^2 df \quad (6.81)$$

Le rapport à maximiser

$$\eta = \frac{\left\| \int_{-\infty}^{+\infty} \mathcal{H}(f)\mathcal{G}(f)e^{2\pi jfT} df \right\|^2}{\frac{N_0}{2} \int_{-\infty}^{+\infty} \|\mathcal{H}(f)\|^2 df} \quad (6.82)$$

devient donc

$$\eta \leq \frac{2}{N_0} \int_{-\infty}^{+\infty} \|\mathcal{G}(f)\|^2 df \quad (6.83)$$

D'où la valeur maximale

$$\eta_{\max} = \frac{2}{N_0} \int_{-\infty}^{+\infty} \|\mathcal{G}(f)\|^2 df = \frac{2E_b}{N_0} \quad (6.84)$$

où l'on a défini l'énergie du signal d'entrée par

Définition 74 [Énergie du signal]

$$E_b = \int_{-\infty}^{+\infty} \|\mathcal{G}(f)\|^2 df = \int_0^T |g(t)|^2 dt \quad (6.85)$$

cette dernière relation résultant du théorème de RAYLEIGH.

En d'autres termes, le rapport signal à bruit au moment de l'échantillonnage est fonction de l'énergie du signal d'entrée et de la puissance du bruit additif; la forme d'onde n'intervient pas directement mais bien son énergie.

Propriétés du filtre adapté

Par contre, pour avoir le maximum correspondant à l'égalité, on a supposé que (cf. relation 6.80)

$$\mathcal{H}_{opt}(f) = k\mathcal{G}^*(f)e^{-2\pi jfT} \quad (6.86)$$

et donc

$$h_{opt}(t) = k \int_{-\infty}^{+\infty} \mathcal{G}^*(f)e^{-2\pi jf(T-t)} df \quad (6.87)$$

Comme pour un signal réel $g(t)$, $\mathcal{G}^*(f) = \mathcal{G}(-f)$, on a

$$h_{opt}(t) = k \int_{-\infty}^{+\infty} \mathcal{G}(-f)e^{-2\pi jf(T-t)} df \quad (6.88)$$

$$= kg(T-t) \quad (6.89)$$

On peut donc écrire le filtre optimal par

$$h_{opt}(t) = \begin{cases} kg(T-t) & 0 \leq t \leq T \\ 0 & \text{ailleurs} \end{cases} \quad (6.90)$$

À une amplitude près, le filtre optimal, appelé filtre adapté, correspond à l'onde de mise en forme retournée sur la période T .

Implémentation du filtre adapté

En pratique, on dispose de plusieurs moyens de réaliser le filtre adapté :

1. Par convolution. C'est l'implémentation directe de la formule de filtrage

$$y(t) = \int_{-\infty}^{+\infty} x(\tau)h(t-\tau)d\tau \quad (6.91)$$

On échantillonne ce signal à l'instant $t = T$ pour obtenir la valeur $y[T]$.

2. Par corrélation. Considérons l'expression de $y(t)$

$$y(t) = \int_{-\infty}^{+\infty} x(\tau)g(T-t+\tau)d\tau \quad (6.92)$$

à partir de quoi

$$y[T] = \int_0^T x(\tau)g(\tau)d\tau \quad (6.93)$$

Par faciliter les calculs, on préfère le signal

$$z(t) = \int_0^t x(\tau)g(\tau)d\tau \quad (6.94)$$

En $t = T$, les deux signaux sont égaux même si ailleurs $y(t) \neq z(t)$. Cela importe peu car c'est la valeur au droit de l'échantillonnage qui nous intéresse.

3. Par intégration. Dans le cas particulier d'une fonction $g(t) = 1$ sur $[0, T]$, la formule de $z(t)$ se réduit à

$$z(t) = \int_0^t x(\tau)d\tau \quad (6.95)$$

Il s'agit donc d'un simple intégrateur qu'on échantillonnera en $t = T$.

Quand l'allure de l'onde de mise en forme s'y prête, l'intégrateur est souvent choisi pour l'implémentation. Il convient néanmoins de remarquer que l'intégrateur doit être remis à 0 après chaque échantillonnage sous peine d'une dérive de la valeur de $z(t)$.

6.4.5 Seconde phase : détection par maximum de vraisemblance

Le signal reçu durant l'intervalle de temps T est

$$x(t) = \begin{cases} g_0(t) + n(t), & 0 \leq t \leq T & \text{pour un signal binaire 0} \\ g_1(t) + n(t), & 0 \leq t \leq T & \text{pour un signal binaire 1} \end{cases} \quad (6.96)$$

Pour notre étude, considérons un signal NRZ, $g_0(t) = -V$ et $g_1(t) = +V$.

Comme in fine le récepteur doit estimer quel signal a été reçu, il peut y avoir deux types d'erreur au moment de prendre une décision

1. Sélectionner le symbole 1 alors qu'on a transmis le symbole 0 ; c'est l'erreur de type 1.
2. Sélectionner le symbole 0 alors qu'on a transmis le symbole 1 ; c'est l'erreur de type 2.

Pour déterminer le taux d'erreur durant la transmission, il faut examiner distinctement le cas de chaque symbole transmis. Dans le cas binaire, l'analyse revient à considérer l'émission du symbole $g_0(t)$ ou $g_1(t)$.

Probabilité d'erreur lors de l'envoi du signal NRZ $g_0(t)$

Supposons que l'on ait transmis un symbole 0, autrement dit le signal $g_0(t) = -V$ pour $0 \leq t \leq T_b$. Le signal reçu au récepteur est alors

$$x(t) = -V + n(t) \quad 0 \leq t \leq T_b \quad (6.97)$$

où T_b est la durée d'un bit.

Dès lors, la sortie du filtre adapté ou corrélateur est

$$y(t) = \int_{-\infty}^{+\infty} x(\tau)h_0(t - \tau)d\tau \quad (6.98)$$

$$= \int_{-\infty}^{+\infty} x(\tau)kg_0(T_b - t + \tau)d\tau \quad (6.99)$$

$$= \int_0^{T_b} (-V)kg_0(T_b - t + \tau)d\tau + \int_0^{T_b} n(\tau)kg_0(T_b - t + \tau)d\tau \quad (6.100)$$

En $t = T_b$,

$$y[T_b] = \int_0^{T_b} (-V)k g_0(\tau) d\tau + \int_0^{T_b} n(\tau)k g_0(\tau) d\tau \quad (6.101)$$

$$= kV^2 T_b - kV \int_0^{T_b} n(\tau) d\tau \quad (6.102)$$

Pour un gain $k = -1/VT_b$,

$$y[T_b] = -V + \frac{1}{T_b} \int_0^{T_b} n(\tau) d\tau \quad (6.103)$$

Remarquons qu'il s'agit de la réalisation d'une variable aléatoire Y échantillonnée au temps $t = T_b$. Moyennant l'hypothèse d'un bruit additif blanc et gaussien $n(t)$, la variable aléatoire Y jouit des caractéristiques suivantes :

- Y est une variable aléatoire gaussienne de moyenne égale à $-V$
- la variance de la variable aléatoire Y vaut

$$\sigma_Y^2 = E\{(Y + V)^2\} \quad (6.104)$$

$$= \frac{1}{T_b^2} E\left\{ \int_0^{T_b} \int_0^{T_b} n(t)n(u) dt du \right\} \quad (6.105)$$

$$= \frac{1}{T_b^2} \int_0^{T_b} \int_0^{T_b} E\{n(t)n(u)\} dt du \quad (6.106)$$

$$= \frac{1}{T_b^2} \int_0^{T_b} \int_0^{T_b} \Gamma_{NN}(t, u) dt du \quad (6.107)$$

La fonction d'autocorrélation $\Gamma_{NN}(t, u)$ du bruit blanc gaussien est par hypothèse sur la nature de ce bruit une impulsion de DIRAC

$$\Gamma_{NN}(t, u) = \frac{N_0}{2} \delta(t - u) \quad (6.108)$$

Après substitution de cette valeur dans l'équation 6.107,

$$\sigma_Y^2 = \frac{1}{T_b^2} \int_0^{T_b} \int_0^{T_b} \frac{N_0}{2} \delta(t - u) dt du \quad (6.109)$$

$$= \frac{N_0}{2T_b} \quad (6.110)$$

Et donc, la fonction de densité de probabilité de la variable aléatoire Y après transmission du symbole 0 vaut

$$f_Y(y|0) = \frac{1}{\sqrt{\pi N_0/T_b}} e^{-\frac{(y+V)^2}{N_0/T_b}} \quad (6.111)$$

Soit P_{e0} la probabilité d'erreur conditionnelle tenant compte de l'émission du symbole 0. En l'absence de bruit, la sortie du filtre adapté fourni $-V$ au temps d'échantillonnage $t = T_b$. La situation est tout autre en présence de bruit. En effet, la réalisation $y(t)$ prend des valeurs tantôt supérieure, tantôt inférieure à $-V$. À la limite, le bruit est tel que le signal se rapproche

de la valeur $+V$ relative au symbole 1 , ce qu'il faut éviter. Si l'on définit un seuil λ de décision situé entre $-V$ et $+V$, il y aura erreur sur le symbole 0 si la variable aléatoire dépasse la valeur de ce seuil. Formellement, on écrit

$$P_{e0} = p(y > \lambda | 0) \quad (6.112)$$

$$= \int_{\lambda}^{+\infty} f_Y(y|0) dy \quad (6.113)$$

$$= \frac{1}{\sqrt{\pi N_0/T_b}} \int_{\lambda}^{+\infty} e^{-\frac{(y+V)^2}{N_0/T_b}} dy \quad (6.114)$$

Pour poursuivre, il faut fixer la valeur du seuil λ . Une telle détermination se fonde sur la probabilité a priori des symboles 0 et 1 , notées respectivement p_0 et p_1 . En prenant des probabilités égales telles que $p_0 + p_1 = 1$,

$$p_0 = p_1 = \frac{1}{2} \quad (6.115)$$

Dans le cas que nous traitons, les valeurs de crête des deux symboles sont équidistantes par rapport à l'origine le long de l'axe y . Il est dès lors logique de prendre

$$\lambda = 0 \quad (6.116)$$

Cela nous permet de poursuivre le calcul de la probabilité conditionnelle; P_{e0} devient

$$P_{e0} = \frac{1}{\sqrt{\pi N_0/T_b}} \int_0^{+\infty} e^{-\frac{(y+V)^2}{N_0/T_b}} dy \quad (6.117)$$

Après le changement de variable $z = \frac{y+V}{\sqrt{N_0/T_b}}$, on obtient

$$P_{e0} = \frac{1}{\sqrt{\pi}} \int_{\sqrt{E_b/N_0}}^{+\infty} e^{-z^2} dz \quad (6.118)$$

où E_b est l'énergie transmise par bit telle que calculée par (cf. définition 74, page 139)

$$E_b = V^2 T_b \quad (6.119)$$

Il est pratique d'introduire la fonction d'erreur complémentaire; elle est définie par

$$\text{erfc}(u) = \frac{2}{\sqrt{\pi}} \int_u^{+\infty} e^{-z^2} dz \quad (6.120)$$

Grâce à cette fonction, on obtient

$$P_{e0} = \frac{1}{2} \text{erfc} \left(\sqrt{\frac{E_b}{N_0}} \right) \quad (6.121)$$

La densité de probabilité $f_Y(y|0)$ et la probabilité d'erreur P_{e0} sont représentées à la figure 6.8(a).

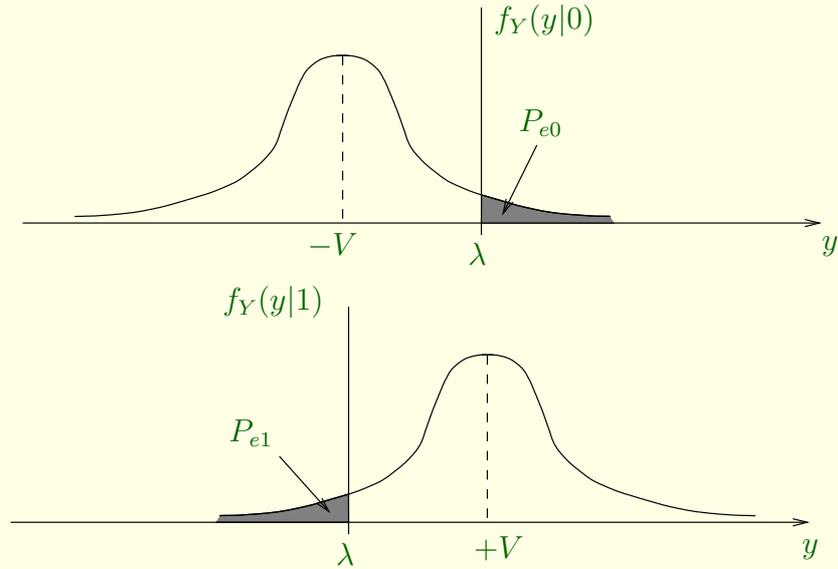


FIGURE 6.8 – Forme des densités de probabilités $f_Y(y|0)$, $f_Y(y|1)$ et probabilités d'erreur.

Probabilité d'erreur lors de l'envoi du signal NRZ $g_1(t)$

Supposons à présent que ce soit le signal $g_1(t)$ qui ait été envoyé. Un raisonnement similaire au précédent conduit rapidement à

$$f_Y(y|1) = \frac{1}{\sqrt{\pi N_0/T_b}} e^{-\frac{(y-V)^2}{N_0/T_b}} \quad (6.122)$$

On note par P_{e1} la probabilité d'erreur conditionnelle après émission du symbole 1. Son expression est

$$P_{e1} = p(y < \lambda|1) \quad (6.123)$$

$$= \int_{-\infty}^{\lambda} f_Y(y|1) dy \quad (6.124)$$

$$= \frac{1}{\sqrt{\pi N_0/T_b}} \int_{-\infty}^{\lambda} e^{-\frac{(y-V)^2}{N_0/T_b}} dy \quad (6.125)$$

En opérant le changement de variable $-z = \frac{y-V}{\sqrt{N_0/T_b}}$, on remarque que $P_{e0} = P_{e1}$.

Les deux densités de probabilité sont représentées à la figure 6.9.

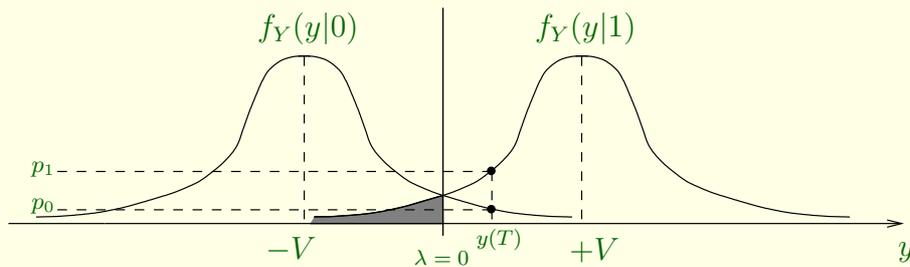


FIGURE 6.9 – Densités de probabilité conditionnelles.

Probabilité d'erreur moyenne

Pour déterminer la probabilité d'erreur au récepteur, on note que les deux types d'erreur énoncés précédemment sont exclusifs. On définit une probabilité d'erreur moyenne au récepteur P_e en pondérant les erreurs respectives par les probabilités a priori des signaux à l'émetteur

$$P_e = p_0 P_{e0} + p_1 P_{e1} \quad (6.126)$$

Dans notre cas, on en déduit la valeur suivante

$$P_e = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_b}{N_0}} \right) \quad (6.127)$$

Cette grandeur est généralement appelée taux d'erreur sur bit ou Bit Error Rate (**BER**).

Le taux d'erreur moyen d'un système PCM à deux états opposés dépend donc exclusivement du rapport entre l'énergie transmise par bit et la densité spectrale de bruit ; la courbe est dessinée à la figure 6.10.

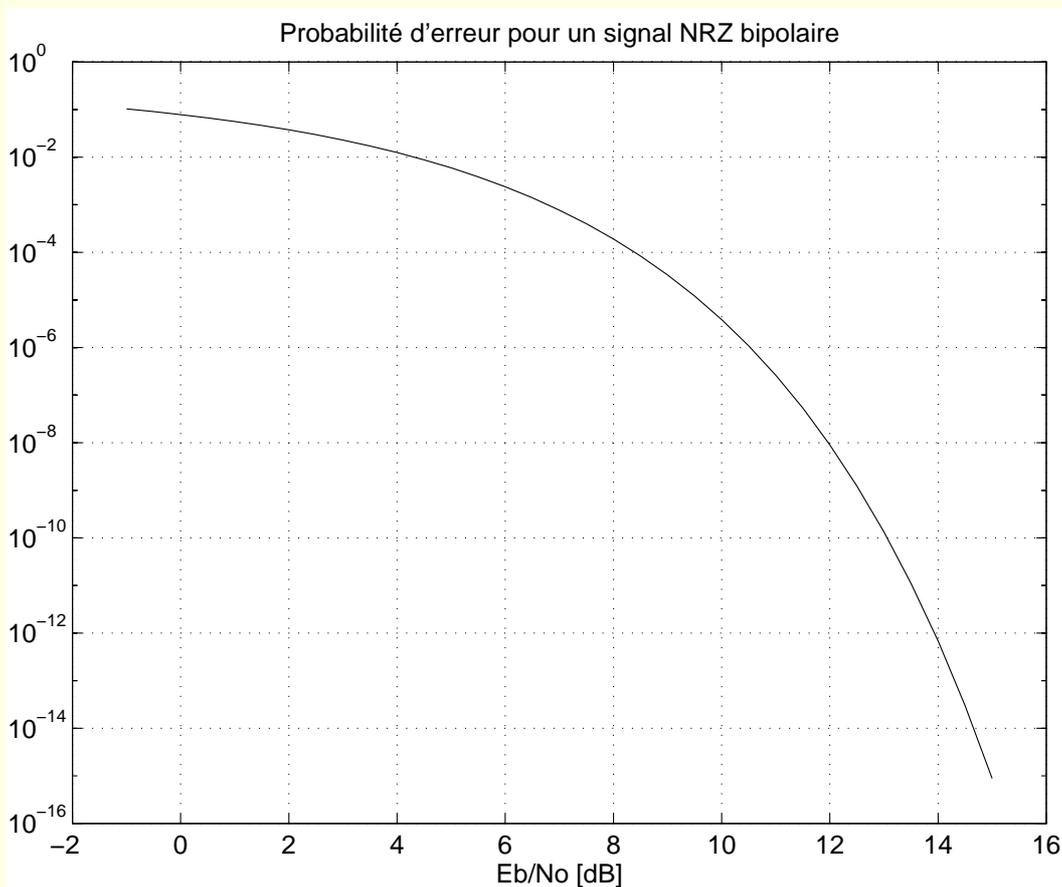


FIGURE 6.10 – Taux d'erreur pour un codage en ligne de type NRZ bipolaire.

Comme il n'est pas possible d'agir sur la valeur de la puissance de bruit N_0 , diminuer la probabilité d'erreur ne pourra se faire que par l'augmentation de E_b . Puisque $E_b = V^2 T_b$, il faudrait donc augmenter V –du moins quand ce degré de liberté existe– ou ralentir la cadence d'émission.

En pratique, on recourt à une autre technique, en l'occurrence à celle des codes détecteurs ou correcteurs d'erreur qui diminue le taux d'erreur mais qui présente l'inconvénient d'élargir la bande de base à même débit utile d'information.

6.5 Effet de la limitation de la bande passante

Jusqu'à présent, nous avons considéré que le système ne souffrait d'aucune limitation de bande passante, hormis celle que pourrait introduire le filtre adapté. Or, comme l'onde d'émission est réelle, il faudrait théoriquement une largeur de bande infinie pour ne pas déformer les ondes. En pratique malheureusement, le canal de transmission est toujours à bande limitée. Il convient donc d'examiner l'effet de la limitation de la largeur de bande sur la qualité de la réception.

Nous nous limiterons à une analyse intuitive ici. À ces fins, les figures 6.11 et 6.12 illustrent l'effet de la limitation de bande passante pour différentes fréquences de coupure.

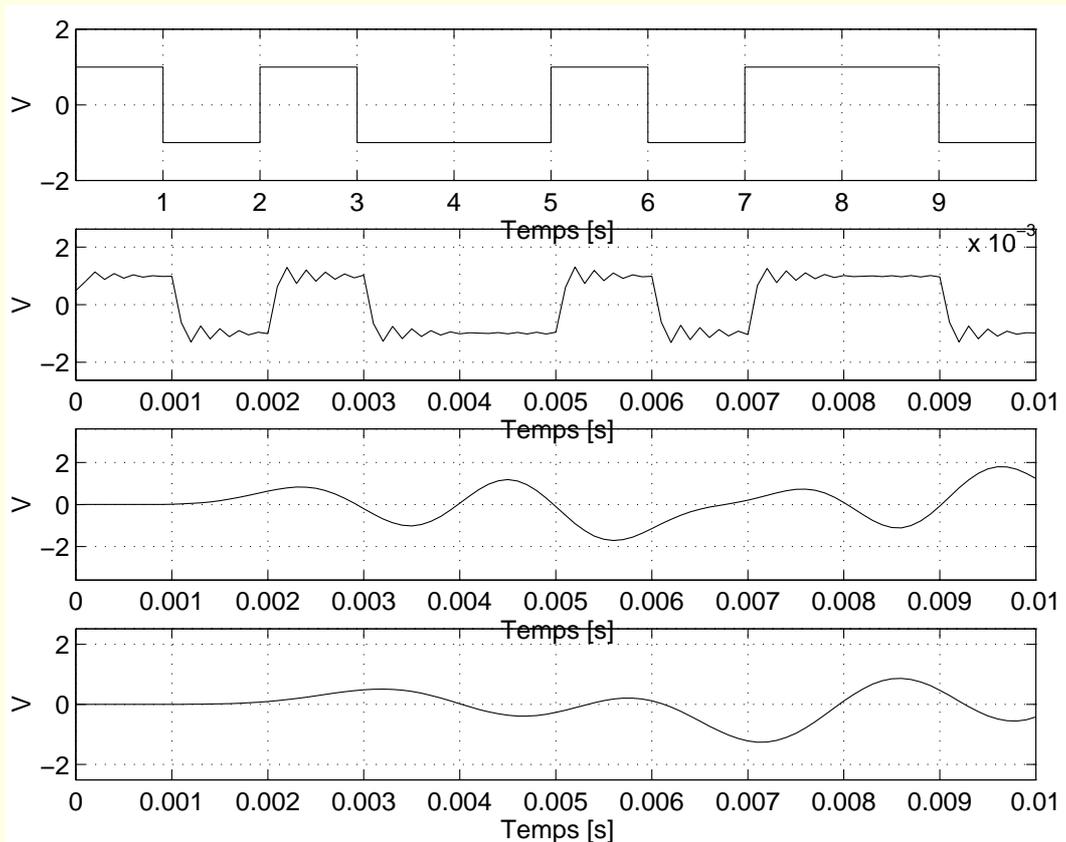


FIGURE 6.11 – Effet de la limitation de la bande de fréquence sur un signal NRZ bipolaire : (a) signal original, (b) signal filtré à $5f_b$, (c) signal filtré à $0,6f_b$ et (d) signal filtré à $0,4f_b$.

D'une manière générale, on admet qu'il est encore possible de discerner les signaux successifs en l'absence de bruit pour une fréquence de coupure supérieure ou égale à la valeur du premier zéro de la densité spectrale de puissance multipliée par 0,6. Concrètement, pour un débit binaire de $1000 [b/s]$ d'un signal NRZ, la fréquence de coupure minimale vaut $600 [Hz]$.

6.5.1 Diagramme de l'œil

Si l'on superpose les symboles successifs par paires sur un oscilloscope, on obtient le diagramme de l'œil. Ce diagramme permet de vérifier expérimentalement l'effet d'une limitation de la bande passante (cf. figure 6.13). Mais il permet aussi de mesurer plus généralement la qualité du signal reçu.

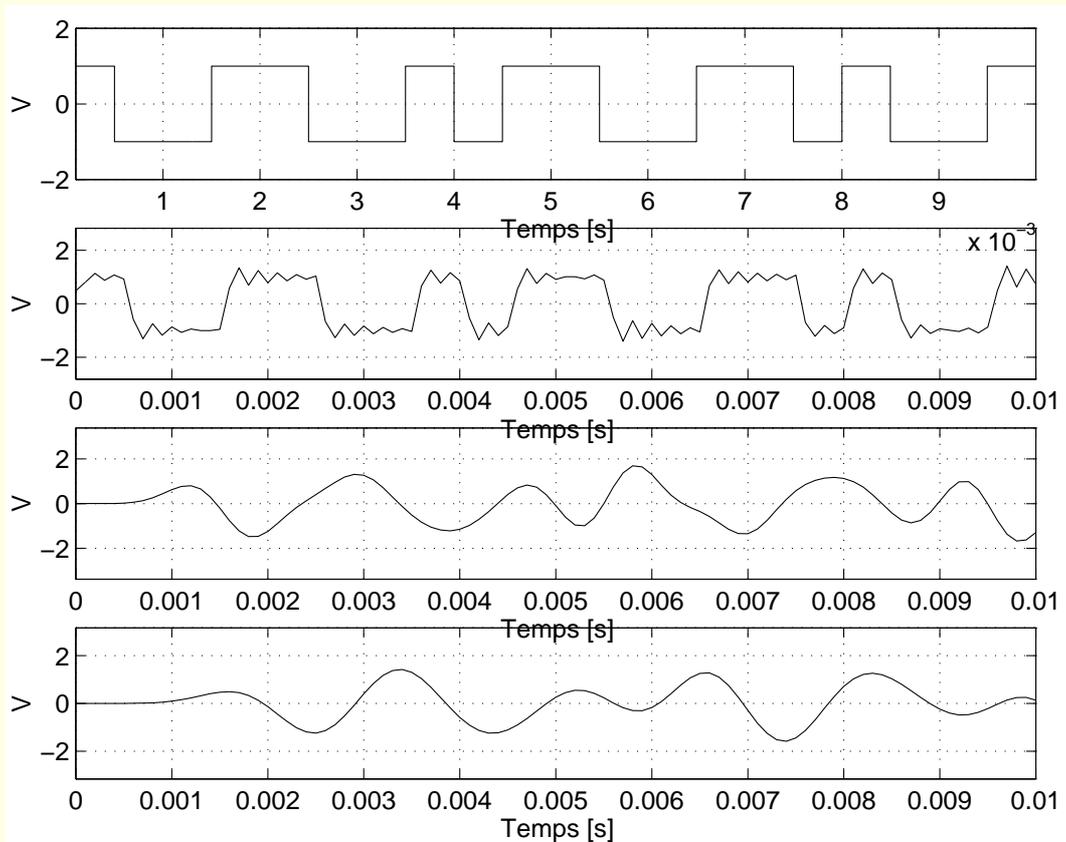


FIGURE 6.12 – Effet de la limitation de la bande de fréquence sur un signal de type MANCHESTER (a) signal original, (b) signal filtré à $5f_b$, (c) signal filtré à $0,6f_b$ et (d) signal filtré à $0,4f_b$.

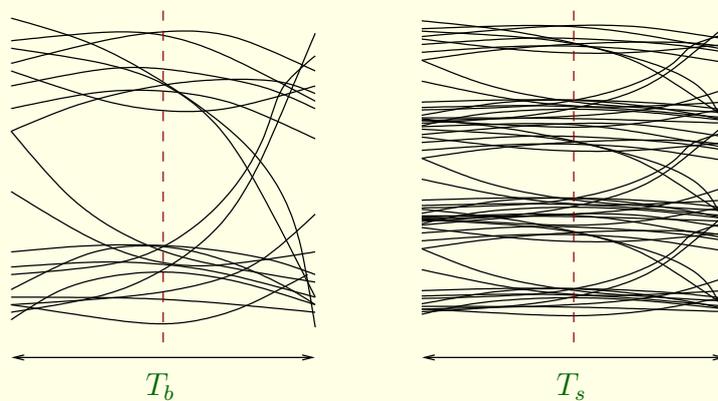


FIGURE 6.13 – Diagrammes de l’œil obtenus à travers un canal à réponse impulsionnelle finie pour un système à 2 et 4 états.

La figure 6.14 montre toute une série de paramètres observables. Le principe général est que plus l'aire centrale est grande, plus la qualité du signal reçu est bonne. La largeur est liée à la facilité à synchroniser et différencier les échantillons successifs alors que la hauteur du lobe central traduit le rapport d'énergie entre le signal original et le bruit de canal.

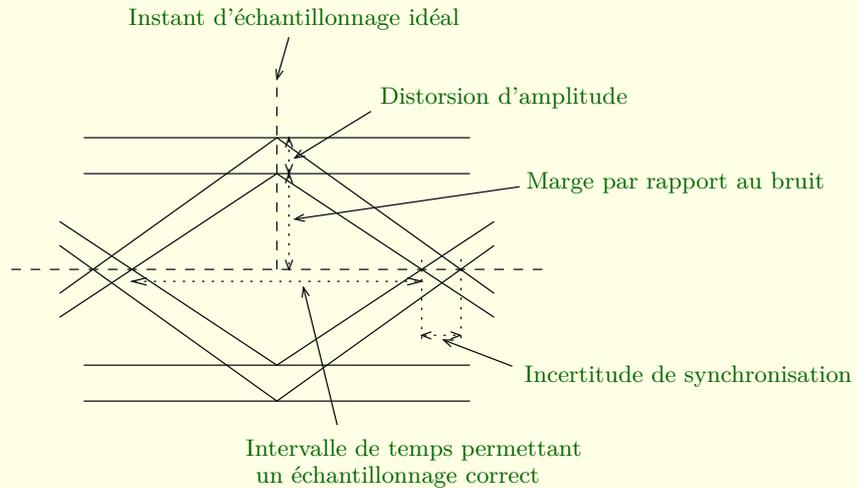


FIGURE 6.14 – Interprétation du diagramme de l'œil.

6.6 Types de transmission

Les données sont généralement transmises en série, c'est-à-dire que les informations numériques (voix, données, ...) sont envoyées bit par bit, à l'inverse de la transmission en parallèle dans laquelle plusieurs bits sont envoyés en même temps. Cette dernière technique requiert au moins une paire de fils pour chaque bit transmis en parallèle et une logique plus complexe. Les coûts de tels câbles et des composants associés sont trop importants pour envisager de la mettre en œuvre dans le cadre des réseaux longues distances.

Les communications numériques reposent sur la capacité des émetteurs et des récepteurs à connaître le début et la durée de chaque bit transmis afin de déterminer la valeur binaire de l'information. De plus, les signaux peuvent être altérés par affaiblissement ou consécutivement à des perturbations électromagnétiques. Pour permettre la synchronisation de l'horloge du récepteur sur celle de l'émetteur, deux approches sont utilisées : la transmission asynchrone et la transmission synchrone.

6.6.1 Transmission asynchrone

Cette première méthode consiste à transmettre les bits par groupes, appelés caractères, de 5 à 8 bits, ce qui permet au récepteur de se synchroniser au début de chaque caractère. Quand aucun caractère n'est transmis, la ligne est en état d'attente. Cet état est représenté par un signal continu de valeur 1, c'est-à-dire une tension négative.

L'intervalle de temps qui sépare l'envoi de chaque mot est aléatoire et dépend de l'émetteur (figure 6.15). Le début d'un caractère est signalé par une valeur binaire 0 (bit start) suivie de 5 à 8 bits composant le caractère. Un bit de parité suit le caractère et permet de détecter une erreur de transmission sur le principe suivant : l'émetteur positionne ce bit à 1 si le nombre total

de valeurs binaires 1 présentes dans le caractère (et y compris le bit de parité lui-même) est pair. Une autre convention positionne ce bit si le même nombre total est impair. On distingue alors la parité paire et la parité impaire.

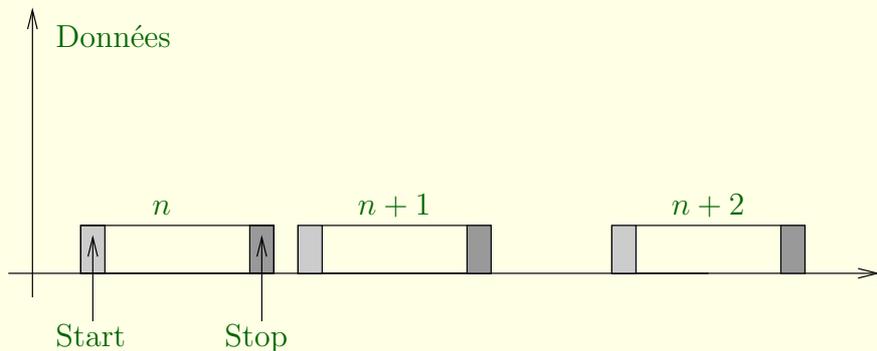


FIGURE 6.15 – Transmission asynchrone.

Un dernier bit, appelé bit stop et de valeur binaire 1, clôt le groupe de bits transmis. Ce bit est en réalité du même niveau de tension que le signal représentant l'état d'attente. L'émetteur transmettra donc ce signal jusqu'au prochain caractère à émettre, la longueur du signal étant d'au moins 1 à 2 équivalents temps bits.

Cette méthode autorise des divergences d'horloge de l'ordre de 5%. Par exemple, si le récepteur est 5% plus lent ou plus rapide que l'émetteur, l'échantillonnage de 8 bits d'information sera décalé de 45% mais sera toujours correct. Les composants sont donc simples à fabriquer et donc bon marché.

Par contre, l'inconvénient de cette méthode est dû au surcoût, appelé overhead, c'est-à-dire au nombre de bits de contrôle (de 2 à 4) nécessaires pour transporter un caractère. Par exemple, pour transporter un caractère de 8 bits il faut ajouter un bit start, un bit de parité (conseillé, bien qu'optionnel), un bit stop de longueur 2, soit en tout 4 bits supplémentaires, ce qui représente un surcoût de près de 33%. Au mieux, l'overhead est de 20% pour 8 bits de données transmis sans parité.

6.6.2 Transmission synchrone

La transmission synchrone permet de transmettre un bloc de bits d'information, appelé trame, sous la forme d'un flot continu de bits sans bit de synchronisation start et stop. Les données sont rythmées par une horloge qui assure un temps constant entre chaque bit envoyé, mais aussi entre chaque mot (cf. figure 6.16). La synchronisation des horloges est obtenue soit en générant un signal sur une ligne séparée, soit en utilisant les signaux de données comme référence d'horloge. La dernière méthode est la plus utilisée car elle permet d'utiliser moins de fils.

Ce type de transmission implique un autre niveau de synchronisation pour déterminer le début et la fin d'une trame. Un préambule, et subsidiairement un postambule, doivent être générés au début et à la fin d'une trame. De plus, le bit de parité ne suffit plus pour détecter les erreurs, car le nombre de bits est trop important. Ces techniques font partie de protocoles comme le protocole HDLC (High Data Level Control).

Cette méthode présente les avantages et les inconvénients inverses de la transmission asynchrone : le surcoût généré est d'environ 1% pour des trames de 1000 bits. Par contre, la logique de contrôle est plus complexe et fait appel à des logiciels de niveau supérieur.

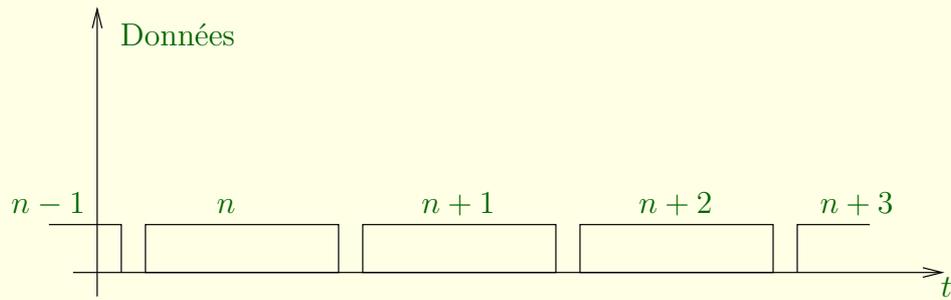


FIGURE 6.16 – Transmission synchrone.

Il est à noter que la mise en œuvre d'un protocole asynchrone est possible sur une ligne synchrone, mais le débit de cette dernière doit être quatre fois plus élevé que celui désiré pour l'asynchrone.

En règle générale, le récepteur doit reconstituer l'horloge si elle n'est pas transmise. Il utilise pour cela des techniques de récupération d'horloge ou récupération de rythme qui permettent d'extraire celle-ci à partir du signal reçu.

Chapitre 7

Modulation numérique

7.1 Introduction

L'utilisation des modulations numériques répond à deux impératifs :

- lorsque la transmission des signaux numériques en bande de base n'est pas possible (dans l'air ou sur les lignes téléphoniques), on est obligé de translater le spectre de la bande de base vers les hautes fréquences : il faut réaliser une modulation,
- la demande en plages de fréquences augmente considérablement, alors que le spectre en espace libre est une ressource partagée. Le numérique utilise plus efficacement ces plages car son encombrement est réduit grâce à des techniques de compression de données et de modulations adaptées.

Le principe de la réalisation de la modulation numérique est le même que dans le cas des modulations analogiques : le mélangeur équilibré en est la pièce centrale. La qualité de la liaison se chiffre aussi à l'aide du rapport signal à bruit, mais ces dégradations entraînent des erreurs de transmission, ce qui entraîne la réception de bits erronés. Aussi, la quantité utilisée pour chiffrer la qualité de la transmission est la probabilité d'erreur, qui est une fonction du rapport signal à bruit.

7.1.1 Modulation cohérente ou incohérente

On peut distinguer deux grandes classes de modulation numérique, qui possèdent chacune des propriétés particulières :

- la modulation cohérente : la fréquence de la porteuse est un multiple entier du rythme d'émission $1/T_b$,
- la modulation incohérente : la fréquence de la porteuse n'est pas un multiple entier du rythme d'émission $1/T_b$.

Dans le premier cas, la modulation peut être vue comme un cas particulier de codage en ligne à l'émission en bande de base ; l'onde de mise en forme est tout simplement un signal sinusoïdal périodique. Toutefois, la modulation cohérente exige une synchronisation parfaite entre la source d'information et la porteuse, ce qui n'est guère aisé à réaliser. Il faut alors recourir à une modulation incohérente. On devine que la conception du récepteur sera guidée par d'autres principes et qu'elle demandera un soin tout particulier.

7.1.2 Démodulation cohérente ou incohérente

Il ne faudrait pas confondre la nomenclature du type de modulation avec le principe utilisé dans certains démodulateurs utilisant une reconstitution de la porteuse pour démoduler le signal reçu. Ainsi, on définit également une démodulation cohérente, pour laquelle le récepteur dispose de la fréquence et de la phase instantanées en permanence, et une démodulation incohérente, pour laquelle cette information n'est pas disponible.

7.2 Modulation

7.2.1 Hypothèses

Le signal modulant $m(t)$ sera, sauf précision contraire, un signal aléatoire NRZ bipolaire à deux états d'amplitude. La période bit est $T_b = 1/f_b$. L'énergie par bit est E_b . Le bruit est considéré comme un bruit blanc de densité spectrale $N_0/2$.

Dans ce qui suit, nous n'analyserons que les techniques de modulation cohérente. Pour ces modulations, chaque bit modulé contient le même nombre de cycles de la porteuse, c'est-à-dire que fréquence de porteuse et d'émission sont liées par la relation

$$f_c = \frac{n_c}{T_b} \quad (7.1)$$

avec n_c entier.

7.2.2 Modulation d'amplitude numérique cohérente

La modulation numérique d'amplitude porte le nom anglais d'Amplitude-Shift Keying (ASK) qui, dans le cas où un signal est nul, s'appelle on/off keying. Son principe est illustré, en même temps que celui de la modulation de phase et de fréquence, à la figure 7.1. Un signal modulant NRZ à deux états d'amplitude, A_0 et A_1 , multiplie une porteuse. À supposer que les symboles successifs de la source soient indépendants entre eux, la séquence modulée peut être vue comme la somme de deux signaux modulés de type on/off keying. Pour construire chacun d'eux, il suffit de prendre un signal nul au droit des symboles de l'autre séquence.

Soit $s_0(t)$, le premier de ces signaux. Il s'exprime comme

$$s_0(t) = \sum_{k=-\infty}^{+\infty} A_k \text{rect}_{[0, T_b]}(t - kT_b) \cos(2\pi f_c t) \quad (7.2)$$

où A_k est une séquence de valeurs prises dans l'ensemble $\{0, A_0\}$.

Occupation spectrale

Tout signal fini ayant un spectre infini et cette situation étant inacceptable pour les systèmes de transmission, le signal d'émission subira un filtrage préalable à la transmission. Encore faut-il être en mesure de déterminer les fréquences les plus significatives du signal à émettre.

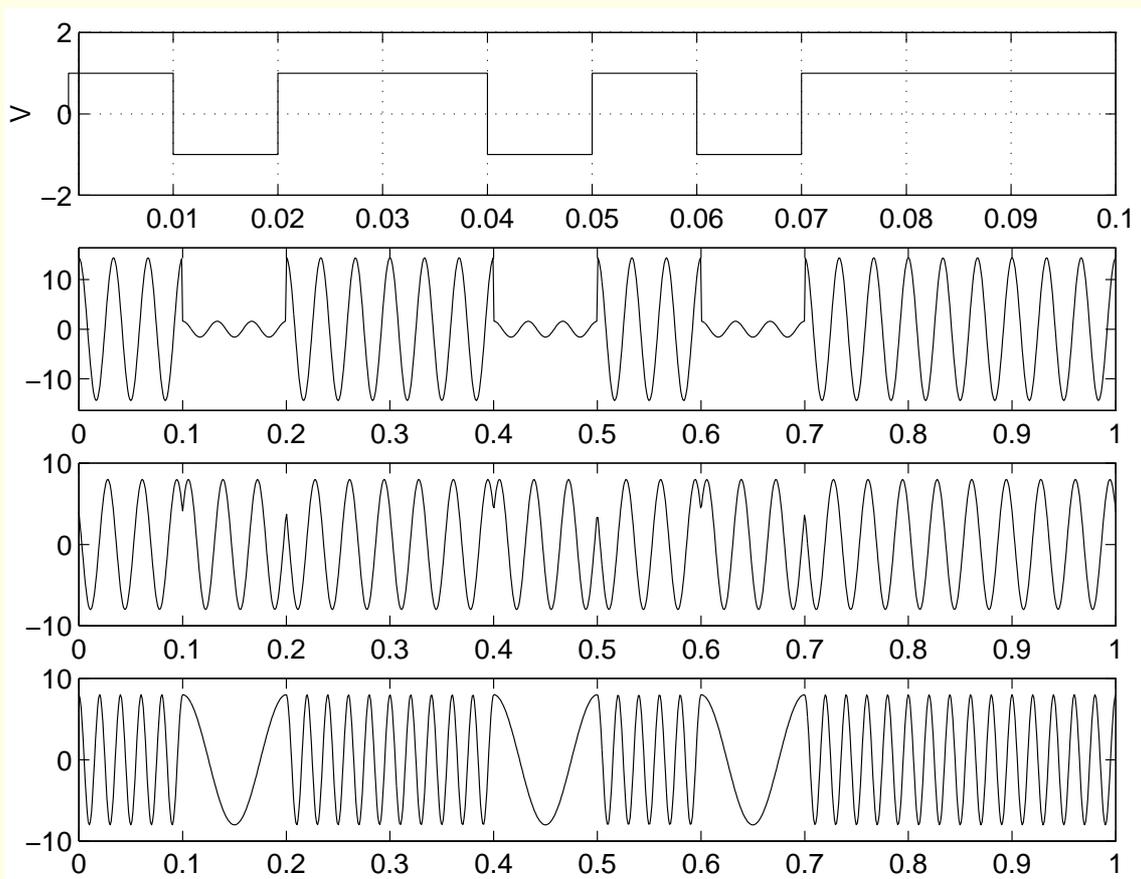


FIGURE 7.1 – Signal modulant numérique et signaux modulés respectivement en AM, PM et FM.

Dans un premier temps, prenons une approche déterministe. Le cas le plus défavorable, c'est-à-dire celui qui fournira la fréquence significative la plus élevée, est celui où il y a alternance continue entre les deux symboles dans le signal modulant $m_0(t)$. Cela mène à un développement en série de FOURIER de la forme

$$m_0(t) = \frac{A_0}{2} \left[1 + \frac{4}{\pi} \cos(2\pi(\frac{1}{2}f_b)t) - \frac{4}{3\pi} \cos(2\pi(\frac{3}{2}f_b)t) + \dots \right] \quad (7.3)$$

La composante continue vaut $A_0/2$ et on remarque la présence de raies spectrales aux multiples impairs de la moitié de la fréquence des bits, c'est-à-dire $(2n + 1)f_b/2$. Il n'y a pas de raie spectrale pour les multiples entiers de la cadence du flux binaire. Le spectre du signal modulé $s_0(t)$ est obtenu par transposition du spectre initial à la fréquence de porteuse

$$s_0(t) = \frac{A_0}{2} \left[\cos(2\pi f_c t) + \frac{2}{\pi} \cos(2\pi(f_c + \frac{f_b}{2})t) + \frac{2}{\pi} \cos(2\pi(f_c - \frac{f_b}{2})t) - \frac{2}{3\pi} \cos(2\pi(f_c + \frac{3f_b}{2})t) - \frac{2}{3\pi} \cos(2\pi(f_c - \frac{3f_b}{2})t) + \dots \right] \quad (7.4)$$

Si, au lieu d'être régulière, la séquence d'entrée est aléatoire et les probabilités d'un 0 ou d'un 1 équiprobables, on peut raisonnablement penser que le spectre réel prolongera continûment le spectre discret obtenu pour la séquence régulière. D'un point de vue formel, la modélisation ne diffère en rien de la modélisation faite pour la transmission numérique en bande de base à condition de respecter la condition de cohérence 7.1. Il faut néanmoins tenir compte du cosinus lors de la stationnarisation du signal.

Le résultat est la multiplication de la fonction d'autocorrélation par un facteur de modulation d'amplitude¹ $(1/2)\cos(2\pi f_c \tau)$. La densité spectrale de puissance s'exprime donc comme la densité spectrale du signal modulant convoluée par la transformée de FOURIER de la porteuse (avec un facteur $\frac{1}{2}$ supplémentaire) ce qui, dans le cas présent, donne

$$\gamma(f) = \left[p_0(1 - p_0)A_0^2 T_b \left(\frac{\sin(\pi f T_b)}{\pi f T_b} \right)^2 + p_0^2 A_0^2 \delta(f) \right] \otimes \frac{\delta(f - f_c) + \delta(f + f_c)}{4} \quad (7.5)$$

$$= \frac{p_0(1 - p_0)A_0^2 T_b}{4} \left(\frac{\sin(\pi(f - f_c)T_b)}{\pi(f - f_c)T_b} \right)^2 + \frac{p_0(1 - p_0)A_0^2 T_b}{4} \left(\frac{\sin(\pi(f + f_c)T_b)}{\pi(f + f_c)T_b} \right)^2 + \frac{p_0^2 A_0^2}{4} \delta(f - f_c) + \frac{p_0^2 A_0^2}{4} \delta(f + f_c) \quad (7.6)$$

Comme on pouvait le pressentir, le spectre se compose d'une raie à la fréquence porteuse flanquée, de part et d'autre, du spectre du signal modulant. Le spectre du signal ASK s'obtient en additionnant les spectres respectifs de $s_0(t)$ et de $s_1(t)$ en tenant compte du fait que $p_0 + p_1 = 1$, ce qui ne change rien à l'occupation spectrale.

L'expression rigoureuse du spectre du signal ASK (relation 7.6) confirme l'étendue infinie du spectre. L'usage veut que le spectre filtré contienne au moins la fréquence critique telle que déterminée précédemment, à savoir $f_b/2$. On choisira typiquement une fréquence de coupure de $0,6f_b$ à $0,75f_b$, ce qui conduit à une bande passante du signal modulé de $1,2f_b$ à $1,5f_b$.

1. C'est le résultat établi pour un mélangeur stochastique.

Efficacité spectrale. La limite supérieure de l'efficacité spectrale pour l'ASK est $1 [b/s/Hz]$. En pratique, elle est plutôt entre $0,65 [b/s/Hz]$ et $0,8 [b/s/Hz]$.

Démodulation cohérente

Le schéma d'un démodulateur ASK cohérent ne diffère pas du schéma utilisé pour la bande de base ; il comporte un filtre adapté, suivi d'un organe de décision. Quelques éléments relatifs à la chaîne de modulation et de démodulation sont illustrés aux figures 7.2 et 7.3.

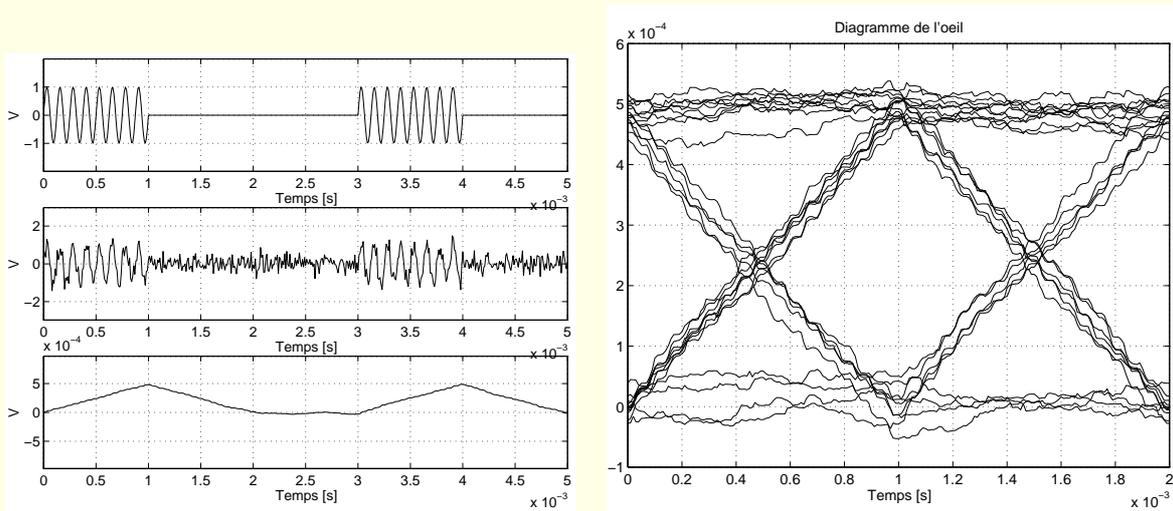


FIGURE 7.2 – Signal ASK en présence d'un faible bruit.

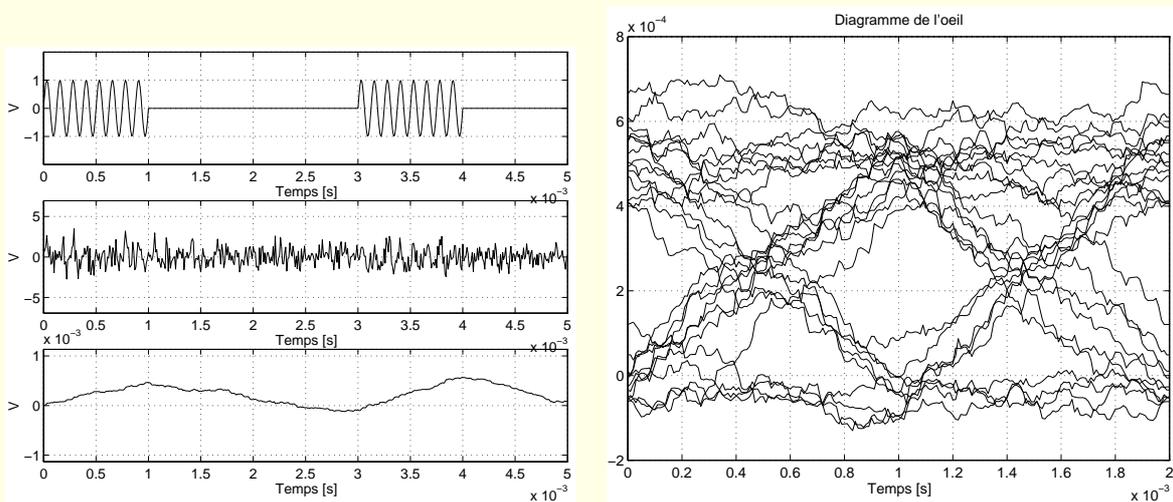


FIGURE 7.3 – Signal ASK en présence d'un bruit important.

On y voit respectivement le signal modulé, le signal modulé bruité, la sortie du filtre adapté et le diagramme de l'œil.

La forte sensibilité de la modulation ASK à des phénomènes de distorsion d'amplitude la rend inadéquate pour des environnements fortement bruités ; on lui préfère alors une modulation angulaire.

7.2.3 Modulation de phase numérique cohérente

En modulation de phase numérique cohérente, appelée Phase-Shift Keying (PSK), on agit sur la phase de la porteuse. Il y a autant d'états de phase différents qu'il y a de niveaux distincts dans l'alphabet d'émission. Considérons le cas d'une PSK binaire. Cette modulation définit les deux états suivants

$$s_0(t) = -\sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t) \quad (7.7)$$

$$s_1(t) = -\sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t + \pi) = +\sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t) \quad (7.8)$$

Il s'agit donc, dans ce cas, d'une modulation équivalente à une modulation d'amplitude à deux états de valeurs opposées. Rappelons qu'on choisit une fréquence de porteuse multiple de $1/T_b$ de manière à avoir une série de cycles de porteuse complets sur la durée d'un bit.

Occupation spectrale

L'occupation spectrale d'une modulation de phase cohérente binaire est identique à celle d'une modulation d'amplitude à deux états. Pour s'en convaincre, il suffit de choisir une modulation d'amplitude telle que $A_0 = -A_1$. Les conclusions obtenues précédemment s'appliquent donc également au cas de la PSK à deux états.

Démodulation cohérente

La démodulation d'un signal PSK peut se faire de différentes manières. Examinons le cas de la démodulation cohérente. On peut voir les deux états comme deux réalisations d'un signal d'énergie unitaire

$$\psi(t) = \sqrt{\frac{2}{T_b}} \cos(2\pi f_c t), \quad 0 \leq t \leq T_b \quad (7.9)$$

Dès lors,

$$s_0(t) = -\sqrt{E_b} \psi(t), \quad 0 \leq t \leq T_b \quad (7.10)$$

$$s_1(t) = +\sqrt{E_b} \psi(t), \quad 0 \leq t \leq T_b \quad (7.11)$$

À supposer que la réception soit cohérente, on disposera du signal $\psi(t)$ au récepteur². Autrement dit, le corrélateur fournira l'un des deux résultats suivants

$$\rho_0 = \int_0^{T_b} s_0(t) \psi(t) dt = -\sqrt{E_b} \quad (7.12)$$

$$\rho_1 = \int_0^{T_b} s_1(t) \psi(t) dt = +\sqrt{E_b} \quad (7.13)$$

2. Une solution consiste à émettre de temps à autre le signal de porteuse non modulée.

La règle de décision est simple : on choisit le symbole correspondant à l'état de sortie le plus proche du filtre adapté. Dans le cas présent, une valeur négative conduira à préférer 0, et 1 pour toute valeur positive. Formellement, pour calculer la probabilité d'une erreur de type 1, on associe la variable aléatoire Z_0 au signal 0. Cette variable aléatoire définie comme

$$Z_0 : 0 < x_1 < +\infty \quad (7.14)$$

est issue du corrélateur à partir du signal observé $x_1(t)$ au temps $t = T_b$

$$x_1 = \int_0^{T_b} x(t)\psi(t)dt \quad (7.15)$$

On montre que, dans le cas d'un bruit blanc additif gaussien, la densité de probabilité de Z_0 vaut

$$f_{Z_0}(x_1|0) = \frac{1}{\sqrt{\pi N_0}} e^{-\frac{(x_1 + \sqrt{E_b})^2}{N_0}} \quad (7.16)$$

La probabilité conditionnelle consistant à retenir le symbole 1 au récepteur alors que l'émetteur a envoyé un 0 est

$$p_{e0} = \int_0^{+\infty} f_{Z_0}(x_1|0) dx_1 \quad (7.17)$$

$$= \int_0^{+\infty} \frac{1}{\sqrt{\pi N_0}} e^{-\frac{(x_1 + \sqrt{E_b})^2}{N_0}} dx_1 \quad (7.18)$$

Par le changement de variable

$$z = \frac{x_1 + \sqrt{E_b}}{\sqrt{N_0}} \quad (7.19)$$

on obtient

$$p_{e0} = \frac{1}{\sqrt{\pi}} \int_{\sqrt{E_b/N_0}}^{+\infty} e^{-z^2} dz \quad (7.20)$$

$$= \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_b}{N_0}} \right) \quad (7.21)$$

où $\operatorname{erfc}(\cdot)$ est la fonction d'erreur complémentaire. Sans surprise, il s'agit de la même probabilité que dans le cas d'une transmission en bande de base d'un signal bipolaire NRZ. Un calcul similaire conduit à une valeur identique pour P_{e1} et à

$$P_e = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_b}{N_0}} \right) \quad (7.22)$$

7.2.4 Modulation d'amplitude en quadrature

Dès que la connaissance exacte de la phase de la porteuse est acquise au récepteur, il n'y a plus aucune raison de coder exclusivement le cosinus de la porteuse. En combinant cosinus et sinus, on peut finalement transmettre deux fois plus d'information dans la même bande passante. L'efficacité spectrale est donc le double de celle d'une ASK ou PSK.

7.2.5 Modulation de fréquence numérique cohérente

La modulation de fréquence numérique cohérente se définit par la série de signaux $s_i(t)$ suivants

$$s_i(t) = \begin{cases} \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_i t), & 0 \leq t \leq T_b \\ 0, & \text{ailleurs} \end{cases} \quad (7.23)$$

tels que les fréquences instantanées respectent

$$f_i = \frac{n_c + i}{T_b} \quad (7.24)$$

n_c et i étant des entiers.

Occupation spectrale

En raison de la non-linéarité de la modulation de fréquence, la détermination de la bande de fréquences d'une FSK est compliquée. Elle dépend de l'indice de modulation, du nombre de fréquences instantanées possibles et de leur écart.

Démodulation cohérente

Dans le cas d'une modulation de fréquence binaire, les signaux sont

$$s_0(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_0 t) \quad (7.25)$$

$$s_1(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_1 t) \quad (7.26)$$

Très clairement, les deux signaux sont complètement différents sur l'intervalle $0 \leq t \leq T_b$. Le démodulateur devra donc réaliser deux structures de filtre adapté et tenir compte de deux sorties pour prendre une décision. Prenons les valeurs des fonctions de corrélation normalisées comme suit

$$\psi_j(t) = \sqrt{\frac{2}{T_b}} \cos(2\pi f_j t), \quad 0 \leq t \leq T_b, \quad j \in \{0, 1\} \quad (7.27)$$

Les états de sortie des filtres adaptés en l'absence de tout bruit valent

$$\rho_{ij} = \int_0^{T_b} s_i(t) \psi_j(t) dt \quad (7.28)$$

$$= \int_0^{T_b} \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_i t) \sqrt{\frac{2}{T_b}} \cos(2\pi f_j t) dt \quad (7.29)$$

$$= \begin{cases} \sqrt{E_b}, & i = j \\ 0, & i \neq j \end{cases} \quad (7.30)$$

Le vecteur d'observation \vec{x} est défini par

$$\begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} \int_0^{T_b} x(t) \sqrt{\frac{2}{T_b}} \cos(2\pi f_0 t) dt \\ \int_0^{T_b} x(t) \sqrt{\frac{2}{T_b}} \cos(2\pi f_1 t) dt \end{bmatrix} \quad (7.31)$$

L'espace des valeurs est représenté à la figure 7.4. La règle de décision est déterminée par la grande diagonale ; le choix se porte toujours sur l'état idéal possible le plus proche. Sur la figure, ces états sont représentés par deux disques noirs.

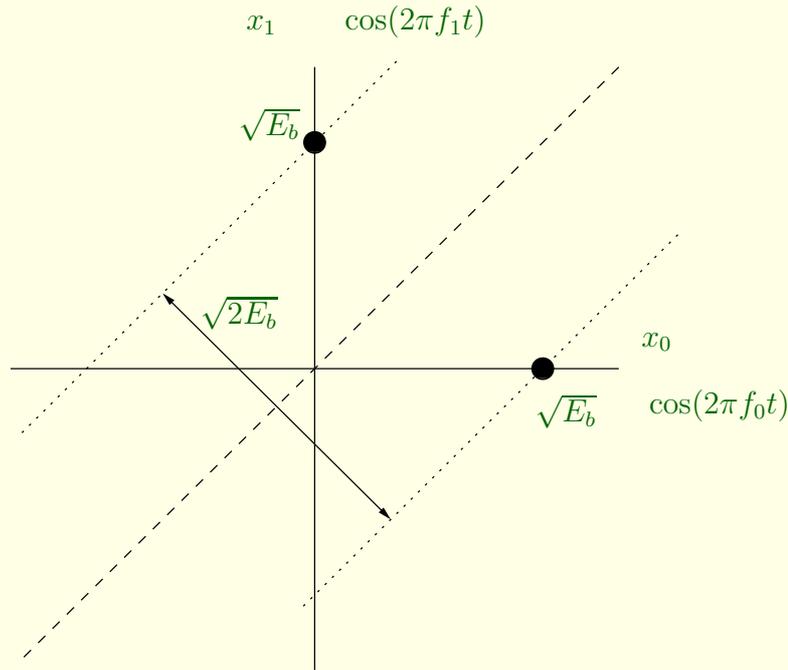


FIGURE 7.4 – Espace de décision pour la démodulation cohérente d'une FSK binaire.

Pour déterminer la probabilité d'erreur, on définit une nouvelle variable aléatoire gaussienne L telle que ses valeurs l soient la différence des résultats des filtres adaptés

$$l = x_0 - x_1 \quad (7.32)$$

La moyenne de cette variable aléatoire, conditionnellement à 0, vaut

$$\mu_{L|0} = \mu_{X_0|0} - \mu_{X_1|0} = \sqrt{E_b} \quad (7.33)$$

La variance de L ne dépend pas du symbole. Comme les variables X_0 et X_1 sont statistiquement indépendantes et de variance $N_0/2$,

$$\sigma_L^2 = \sigma_{X_0}^2 + \sigma_{X_1}^2 = N_0 \quad (7.34)$$

Supposons que le symbole 0 ait été transmis, la densité de probabilité conditionnelle de L vaut

$$f_L(l|0) = \frac{1}{\sqrt{2\pi N_0}} e^{-\frac{(l - \sqrt{E_b})^2}{2N_0}} \quad (7.35)$$

Soit P_{e0} la probabilité d'erreur conditionnelle tenant compte de l'émission du symbole 0. L'organe de décision va préférer le symbole 1 lorsque $l < 0$. Dès lors,

$$P_{e0} = p(l < 0|0) \quad (7.36)$$

$$= \int_{-\infty}^0 f_L(l|0) dl \quad (7.37)$$

$$= \frac{1}{\sqrt{2\pi N_0}} \int_{-\infty}^0 e^{-\frac{(l-\sqrt{E_b})^2}{2N_0}} dl \quad (7.38)$$

Après le changement de variable $z = \frac{l-\sqrt{E_b}}{\sqrt{2N_0}}$, on obtient

$$P_{e0} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\sqrt{E_b/N_0}} e^{-z^2} dz \quad (7.39)$$

$$= \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_b}{2N_0}} \right) \quad (7.40)$$

Par un raisonnement similaire, on obtient P_{e1} et P_e toutes deux égales à

$$\frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_b}{2N_0}} \right) \quad (7.41)$$

En comparant les expressions 7.22 et 7.41, on constate qu'il faut doubler le rapport d'énergie à bruit d'un système FSK binaire cohérent pour maintenir un taux d'erreur identique à celui d'un système PSK binaire. Ce résultat s'interprète aisément car il est apparu que la distance entre points de décision d'une FSK était $\sqrt{2}$ plus petite. En fait, la démodulation cohérente binaire d'un signal affecté d'un bruit aléatoire blanc gaussien dépend exclusivement de la distance entre les états idéaux.

7.2.6 Conclusion

La caractérisation des techniques de modulation numérique est d'une grande complexité. Pour les modulations angulaires, il est en outre difficile de déterminer l'occupation spectrale, et donc de calculer leur efficacité spectrale.

Il se fait que, si les avantages des modulations angulaires sont tels qu'on les préfère à des techniques de modulation d'amplitude, les modulations angulaires présentent des sauts de phase et des déviations de fréquence instantanée très importants. Pour y remédier, on a développé des techniques qui minimisent à tout moment la valeur des sauts de phase. On parle alors de techniques du type Minimum-shift keying (MSK). C'est une variante de type MSK qui est utilisée pour la transmission par GSM; cette variante porte le nom de Gaussian Minimum-shift keying (GMSK).

7.3 Les modems

Les modems (abréviation de modulateur - démodulateur) permettent principalement le transfert de données numériques à travers le réseau téléphonique commuté ainsi que les lignes téléphoniques louées (à deux et quatre fils). La liaison entre un terminal (ETTD) et un modem (ETCD) se fait à travers une liaison V24.

On utilise également des modems dans le transfert de données entre radio-amateurs (connexion du modem à l'entrée/sortie audio des émetteurs/récepteurs, par exemple), dans les liaisons numériques utilisant les réseaux de distribution câblée, etc.

Les modems sont des équipements devenus très courant depuis que la connexion individuelle au réseau Internet s'est développée. Comme tous les produits liés à l'informatique grand public, ils ont connu une évolution technique rapide. De nombreuses fonctionnalités implémentées il y a quelques années sous forme logicielle au niveau du terminal sont maintenant réalisées sous forme matérielle dans les modems (compression, détection d'erreurs, ...).

Le terme de "modem bande de base" est souvent utilisé pour désigner un "émetteur récepteur en bande de base" (ERBDB). On préférera ce dernier terme, moins ambigu. Le terme de modulation est en effet généralement associé à l'utilisation d'une onde porteuse, ce qui n'est pas le cas des ERBDB.

7.3.1 Modes

Mode simplex

Le mode simplex désigne un fonctionnement unidirectionnel. Un modem est chargé d'émettre les données, l'autre de les recevoir.

Mode half-duplex

Le mode semi-duplex (ou half-duplex) permet tour à tour à chaque modem d'émettre et de recevoir des données. Les transferts de données sont autorisés dans les deux sens mais jamais en même temps.

Mode duplex simultané

Le mode duplex simultané (ou full-duplex) désigne un fonctionnement bidirectionnel simultané. À tout moment, les deux modems peuvent recevoir et émettre des données. Cela est simple à réaliser sur une liaison spécialisée à quatre fils mais pose quelques problèmes spécifiques sur une liaison à deux fils.

Le mode duplex simultané peut être symétrique. Dans ce cas, les deux modems émettent au même rythme. Il arrive cependant que la liaison soit dissymétrique et offre des débits différents dans chaque sens. C'est typiquement le cas des liaisons entre un terminal et un serveur distant. Le débit nécessaire à la transmission de données issues d'un clavier est faible si on le compare au débit nécessaire au rafraîchissement de l'écran.

Une liaison bidirectionnelle simultanée peut être réalisée à l'aide d'un multiplexage en fréquence. L'établissement de la communication permet de désigner un modem appelant et un modem appelé. La répartition des fréquences se fait généralement ainsi : le modem appelant émet dans la bande de fréquences la plus basse et reçoit dans la bande la plus haute. Pour le modem appelé, c'est évidemment le contraire. L'inconvénient majeur de ce procédé est de diviser par deux la bande passante disponible pour chaque sens de transmission.

Il est également possible de simuler une liaison duplex simultanée à l'aide de modems semi-duplex équipés de mémoires tampon (buffers). Des procédures de contrôle de flux se chargent

alors de commuter le sens de transmission des données selon les besoins. Le débit global de la liaison est, là encore, partagé entre chaque sens de transmission. Cette technique est peu performante dans le cas de transmissions sur le réseau téléphonique commuté à cause des temps de propagation qui ralentissent les changements de sens de transmission.

La méthode la plus efficace et la plus utilisée dans les modems à haut débit utilise l'annulation d'écho. Cette méthode permet de superposer sur une même paire de fils, au même instant et dans la même bande de fréquence un signal émis et un signal reçu.

7.3.2 Modulation

Les modems les plus rapides utilisent une modulation de type **QAM** avec codage en treillis des différents symboles.

Les taux d'erreurs de transmission sont notamment fonction de la qualité de la ligne et de la densité de états numériques distincts. Certains modems ajustent automatiquement, en fonction des évolutions momentanées de la qualité de la ligne, le nombre de symboles et donc le débit binaire fourni.

7.3.3 Protocoles de transfert de données

Plusieurs protocoles de transfert de données peuvent être implémentée dans le modem lui-même. Ces protocoles complètent ou remplacent avantageusement les protocoles de transfert de fichier de type Xmodem.

Nous citerons ici les protocoles MNP (pour Microcom Networking Protocol) et LAP-M (Link Access Procedure for Modems). Ces protocoles effectuent la détection des erreurs grâce au calcul d'un code à redondance cyclique (**CRC**). En cas d'apparition d'erreurs, ils gèrent la retransmission des trames erronées. Les protocoles MNP et LAP-M sont spécifiés dans les recommandations V42 et V42bis de l'ITU-T.

7.3.4 Compression de données

De nombreux modems intègrent un algorithme de compression de données. Il s'agit, le plus souvent, de l'algorithme LZ (pour LEMPEL et ZIV qui ont inventé cette technique).

Contrairement à certains algorithmes statistiques utilisés pour compresser des fichiers stockés, l'algorithme LZ n'a pas besoin de connaître l'ensemble des données à transmettre avant de commencer à fonctionner. Il effectue une compression des données "au vol", c'est-à-dire au fur et à mesure de leur apparition. De plus, il s'agit d'un algorithme efficace et rapide, aussi bien en compression qu'en décompression. Un algorithme de ce type est spécifié dans la recommandation V42bis de l'ITU-T. Il permet alors des taux de compression de **3 : 1** à **4 : 1** pour des fichiers de texte.

7.3.5 Normes

Les modems permettent l'accès au réseau téléphonique commuté (RTC) ou à des liaisons spécialisées (LS). La série V de l'ITU-T regroupe l'ensemble des recommandations sur le sujet. Seules les principales sont regroupées dans la table **7.1**.

Rec.	Mode	Débit [b/s]	Modulation	Réalisation du mode multiplex
V21	duplex	300	FSK	multiplexage en fréquence
V22	duplex	1200	PSK	multiplexage en fréquence
V22bis	duplex	2400	QAM	multiplexage en fréquence
V23	semi-duplex	1200/75	FSK	voie de retour optionnelle
V29	-	9600	QAM	multiplexeur optionnel
V32	duplex	9600	QAM	annulation d'écho
V32bis	duplex	14400	QAM	annulation d'écho
V33	-	14400	QAM	annulation d'écho
V34	duplex	33600	QAM	annulation d'écho
V90		56000	PCM	

TABLE 7.1 – Recommandations ITU-T.

Chapitre 8

Notions de code

8.1 Introduction

Transmettre de l'information est utile en soi ; la transmettre sans erreur c'est encore mieux. On pourrait ainsi résumer la problématique des télécommunications.

Dans le cas de communications numériques, le théorème de SHANNON-HARTLEY met en lumière l'impossibilité qu'il y a à transmettre de l'information numérique sans erreur dans un canal bruité lorsque le rapport signal à bruit est insuffisant pour le débit souhaité. En effet, alors qu'on est maître de la puissance du signal utile, il n'en est pas de même pour la puissance de bruit. Cette dernière dépend de facteurs non contrôlés par l'émetteur.

Formes de contrôle d'erreur

Pourtant, il existe différents procédés visant à atteindre l'objectif d'une transmission sans erreur à travers un canal bruité. La solution consiste à agir dans les couches supérieures du modèle OSI, au-delà de la couche physique qui elle est régie par les contraintes définies en termes de probabilité d'erreurs en fonction d'un rapport énergie par bit sur bruit.

Ces solutions sont principalement de deux types :

1. Adjoindre une série de bits supplémentaires au message afin de détecter et de corriger les erreurs détectées. On peut tout simplement songer à dupliquer l'information ; une erreur sur 1 bit devient alors aisément détectable.
2. Commander la retransmission de messages erronés. Le mécanisme est alors celui d'une transmission avec accusé de réception. Le récepteur qui reçoit un message sans erreur renvoie un accusé de réception positif (Acknowledgment, ACK). Dans le cas contraire, le récepteur renvoie un accusé de réception négatif (Negative Acknowledgment, NACK) qui équivaut à une demande de retransmission. Cette procédure automatique de demande de retransmission est appelée, en anglais, Automatic Repeat Request (ARQ). On la retrouve dans bon nombre de protocoles du modèle OSI ou de ses équivalents.

L'étude menée ici porte sur les codes détecteurs et correcteurs d'erreur ; les techniques de retransmission automatique font partie de l'étude des protocoles.

8.1.1 Exemple

Les systèmes de télévision offrent des exemples simples et instructifs de codage de mots par blocs. En télévision numérique, le codage de source, c'est-à-dire le codage numérique des informations de couleur et de son, préalable à tout système d'émission, est régi par la norme MPEG-2.

Cette norme définit plusieurs niveaux de qualité par l'entremise, entre autres, de la résolution des échantillonnages de la luminance et des deux signaux de chrominance. Le format le plus courant, format 4 : 2 : 2, consiste à garder tous les échantillons de luminance mais seulement un échantillon sur deux pour les chrominances.

Les deux composantes de chrominance ont des valeurs analogiques de tension comprises entre $-0,5 [V]$ et $+0,5 [V]$. Ces niveaux sont quantifiés par codage PCM pour aboutir à des valeurs comprises entre 16 et 240, 128 étant le niveau 0 $[V]$. Le tableau 8.1 reprend ces 3 niveaux de chrominance et leur adjoint leur équivalent quantifié.

Tension	Niveau quantifié	Équivalent binaire
$-0,5 [V]$	16	00010000
$0 [V]$	128	10000000
$+0,5 [V]$	240	11110000

TABLE 8.1 – Liens entre 3 valeurs analogiques de chrominance et les niveaux quantifiés.

Le passage entre la deuxième colonne et la troisième colonne revient à générer implicitement un code affecté aux mots de 8 bits, appelés octets, représentant les niveaux. Grâce à cette écriture, on remarque que les octets qui codent des amplitudes supérieures ou égales à 0 débutent par un bit de poids fort (MSB, Most Significant Bit) égal à 1, tandis que les octets qui codent les niveaux inférieurs à 0 ont 0 comme MSB.

En télévision numérique de couleur au standard 4 : 2 : 2, chaque ligne active de la composante de luminance est encadrée d'un délimiteur de début de ligne (start) et de fin de ligne (stop). Ces délimiteurs comportent un octet XY formé de deux demi-octets X et Y . Le quartet X est de la forme $X = (1, F, V, H)$. Le bit F définit le numéro de la trame en cours; le bit V précise le type de ligne (active ou de suppression). Le bit H précise si le délimiteur est start ou stop. Le quartet Y a pour but de sécuriser la transmission du quartet X en permettant de corriger les erreurs simples et de détecter les erreurs doubles sur le triplet FVH . Pour ce faire, $Y = P_1P_2P_3P_4$ est défini comme suit

$$\begin{aligned}
 P_1 &= V \oplus F \oplus H \\
 P_2 &= V \oplus F \\
 P_3 &= F \oplus H \\
 P_4 &= V \oplus H
 \end{aligned}$$

où le OU exclusif (XOR), noté \oplus , correspond à une addition modulo 2, comme indiqué dans la table ci-après :

V	H	$V \oplus H$
0	0	0
0	1	1
1	0	1
1	1	0

À partir de cette table de vérité et sachant que l'addition modulo 2 est distributive, on peut établir la valeur du bit P_1 du quartet Y . Les résultats sont repris dans le tableau 8.2.

F	V	H	$P_1 = (F \oplus V) \oplus H = F \oplus (V \oplus H) = (F \oplus H) \oplus V$
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

TABLE 8.2 – Détermination du bit P_1 d'un délimiteur MPEG-2.

En scrutant ces deux tableaux, on remarque que les quatre bits de Y sont à 1 lorsqu'un nombre impair de variables est à 1. À la réception, en rajoutant les bits de Y aux variables qui les engendrent, on doit obtenir un nombre pair de 1.

Application au codage détecteur et correcteur d'erreurs

En conséquence, les bits apparaissent comme des bits de parité, ce qui permet de définir le quartet $Y = P_1P_2P_3P_4$, comme un mot de code de parité. Dans ces conditions, l'assemblage du mot message $m = (F, V, H)$, longueur $k = 3$ bits, et du mot parité $p = (P_1, P_2, P_3, P_4)$, de longueur $r = 4$ bits, forme un ensemble dit mot de code par blocs ou mot codé par bloc $c = (P_1, P_2, P_3, P_4|F, V, H)$, de longueur $n = 7$ bits.

Détection des erreurs en réception

Si une erreur se glisse lors de la transmission, donnant lieu à l'émission d'un bit erroné sur F , V ou H , elle sera facilement détectée en réception en recalculant les bits de parité sur les valeurs de F , V , H reçues et en les comparant aux bits de parité reçus.

L'utilisation d'un seul bit de parité permettrait de constater la présence d'une erreur en transmission, mais ne permettrait pas de localiser le bit erroné. C'est pourquoi on est amené à transmettre un nombre relativement élevé de bits de parité (4 bits ici pour un message de 3 bits!).

Dans un tel type de code, les bits du message sont transmis tels quels, précédés ou suivis du bloc des bits de parité, si bien qu'on parle de code systématique.

8.2 Modèle

L'exemple de signaux MPEG-2 est généralisable sous un formalisme adéquat. Comme préalable, il faut définir un modèle de canal qui sous-tend la conception de codes.

8.2.1 Modèle de canal

Canal discret sans mémoire

Définition 75 Un *canal* discret sans mémoire est caractérisé par un alphabet d'entrée, un alphabet de sortie et un jeu de probabilités conditionnelles, $p(j|i)$, où $1 \leq i \leq M$ représente l'indice du caractère d'entrée, $1 \leq j \leq Q$ représente l'indice du caractère de sortie, et $p(j|i)$ la probabilité d'avoir j en réception alors que i a été émis.

La probabilité d'un symbole de réception dépend donc exclusivement du caractère correspondant émis. Pour un modèle de canal avec mémoire, on rompt cette limitation en autorisant qu'un caractère dépende de plusieurs symboles d'entrée pris à des temps différents.

Canal binaire symétrique

Le canal binaire symétrique est un cas particulier de canal discret sans mémoire. Les alphabets d'entrée et de sortie se limitent aux deux caractères 0 et 1. De plus, les probabilités conditionnelles sont symétriques

$$\begin{aligned} p(0|1) &= p(1|0) = p \\ p(1|1) &= p(0|0) = 1 - p \end{aligned}$$

Sur base d'un modèle de ce type, on définit aisément la probabilité d'erreur d'un démodulateur. Dans le cas d'une modulation de phase à deux états, on établit que la probabilité d'erreur P_e vaut

$$P_e = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_b}{N_0}} \right) \quad (8.1)$$

La démodulation comprenant le filtre adapté suivi d'un organe de décision est relativement contraignante car elle s'effectue statiquement sur base d'un seuil.

Canal gaussien

On peut généraliser la notion de canal discret sans mémoire en prenant le même alphabet d'entrée discret à M niveaux, mais en considérant que la sortie est continue sur l'intervalle $[-\infty, +\infty]$. Cette modélisation permet de considérer que le canal ajoute du bruit au signal. Soit donc un bruit gaussien de moyenne nulle et de variance σ^2 , la densité de probabilité de la variable aléatoire de sortie z , liée à la valeur estimée u_k , est

$$p(z|u_k) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-u_k)^2}{2\sigma^2}} \quad (8.2)$$

Quand l'organe de décision dispose des densités de probabilité pour toutes les valeurs d'entrée, il est en mesure de prendre une décision plus souple sur le symbole transmis mais ce type de décision est plus complexe.

Dans notre cas, nous prendrons le schéma usuel de décision, à savoir celui qui consiste à établir les décisions sur base de seuils fixés une fois pour toutes.

8.2.2 Flux binaire et redondance

Dans le cas de messages codés en blocs, les données de la source sont découpées en blocs de k bits. Chacun de ces blocs représente une possibilité parmi 2^k combinaisons possibles. Ce bloc sert à déterminer r bits supplémentaires, appelés bits redondants, bits de parité ou encore bits de contrôle pour former un mot de n bits. L'ajout de bits se traduit par un surcoût quantifié par le taux de redondance.

Définition 76 *Le taux de redondance d'un code est défini par le rapport*

$$\frac{n - k}{n} \quad (8.3)$$

En raison de la redondance qu'ils introduisent, la majorité des codes entraînent un accroissement de la bande de fréquences. Ainsi, un code dont le taux est de $1/2$ (soit 100% de redondance) nécessite le doublement de la bande de fréquences.

8.3 Codes linéaires

Revenons un instant au signal MPEG-2. Soient $m_1 = F$, $m_2 = V$ et $m_3 = H$, les trois bits du message que l'on veut coder. On écrit pour les bits du message codé : $c_1 = P_1$, $c_2 = P_2$, $c_3 = P_3$, $c_4 = P_4$, $c_5 = m_1$, $c_6 = m_2$ et $c_7 = m_3$. Les expressions explicites de chacun des bits de parité s'écrivent alors de manière formelle

$$\begin{aligned} c_1 &= \alpha_{11}m_1 \oplus \alpha_{21}m_2 \oplus \alpha_{31}m_3 \\ c_2 &= \alpha_{12}m_1 \oplus \alpha_{22}m_2 \oplus \alpha_{32}m_3 \\ c_3 &= \alpha_{13}m_1 \oplus \alpha_{23}m_2 \oplus \alpha_{33}m_3 \\ c_4 &= \alpha_{14}m_1 \oplus \alpha_{24}m_2 \oplus \alpha_{34}m_3 \\ c_5 &= m_1 \\ c_6 &= m_2 \\ c_7 &= m_3 \end{aligned}$$

avec des valeurs α_{ij} adéquates.

Les bits de parité apparaissent donc comme des combinaisons linéaires du message proprement dit. Il en va de même, plus généralement, des bits c_i ($1 \leq i \leq n$) du mot de code par blocs. On se trouve donc dans le cas d'une algèbre linéaire. Il est tout naturel d'appeler code linéaire le code étudié et tout code de même nature. On peut alors écrire ce type de codage à l'aide de représentations matricielles qui en simplifient grandement l'exposé.

8.3.1 Un exemple simple : code redondant

Prenons le cas d'une source binaire. On peut, à titre d'exemple, décider de tripler les symboles d'émission. Ainsi, 0 et 1 deviennent respectivement 000 et 111. Comme le montre la figure 8.1, l'effet du triplement peut s'interpréter comme le passage d'un espace à une dimension à un espace à trois dimensions. Néanmoins, en l'absence de toute erreur, seuls les états (0, 0, 0) et (1, 1, 1) sont possibles.

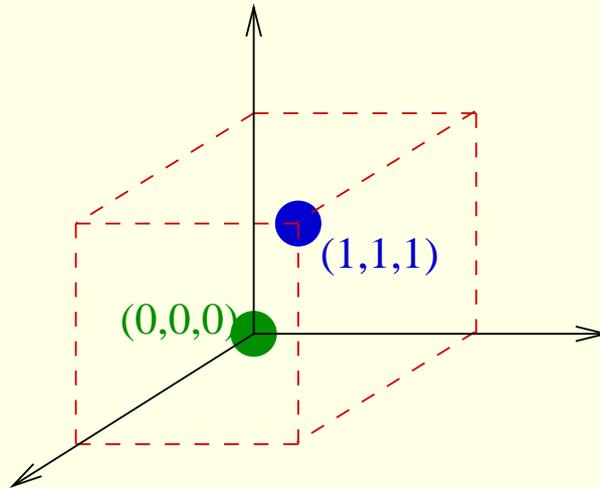


FIGURE 8.1 – Interprétation d’un code redondant dans un espace à 3 dimensions.

Après transmission, il se peut fort bien que le récepteur obtienne un triplet erroné, comme $(0,0,1)$. Dans ce cas d’erreur simple¹, le décodeur peut opérer une correction par vote majoritaire. Ce type de correction est illustré à la figure 8.2.

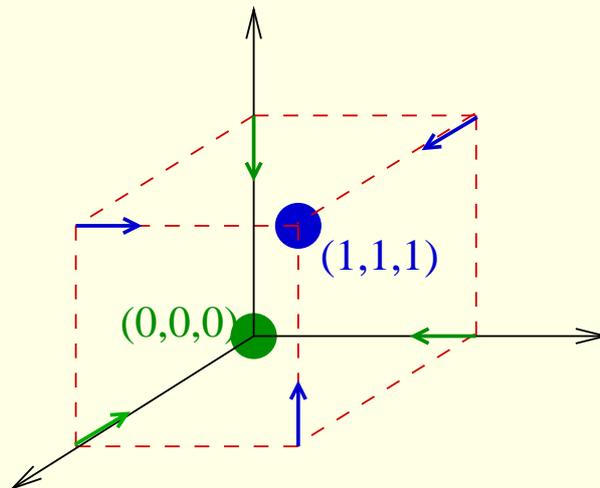


FIGURE 8.2 – Possibilité de corrections d’erreurs simples par vote majoritaire.

Dans le cas d’erreurs doubles ou triples, la correction par vote majoritaire ne permet pas d’obtenir la valeur correcte du bit d’information généré par la source. En conséquence, on observe que le codage par triplement permet seulement certaines corrections ; ces corrections se traduisent pas une diminution de la probabilité d’erreur sur bit d’information de la source. Remarquons au passage que le codage par doublement n’aurait permis aucune correction d’erreur.

8.3.2 Définitions et notations

On convient d’appeler vecteur message \vec{m} la matrice ligne dont les k éléments sont les bits ordonnés du message

$$\vec{m} = (m_1, m_2, \dots, m_k) \quad (8.4)$$

1. Une erreur est dite simple lorsque seul un bit est incorrect.

On appelle de même vecteur de parité \vec{p} la matrice ligne dont les r éléments sont les bits de parité ordonnés

$$\vec{p} = (p_1, p_2, \dots, p_r) \quad (8.5)$$

et on note \vec{c} le vecteur constitué des n bits constituant le mot codé

$$\vec{c} = (c_1, c_2, \dots, c_n) \quad (8.6)$$

8.3.3 Matrice génératrice

Au moyen de cette notation, on écrit la forme matricielle d'un mot codé

$$\vec{c} = \vec{m}G \quad (8.7)$$

La matrice G est appelée matrice génératrice. Elle a pour expression générale

$$G = \begin{bmatrix} \vec{v}_1 \\ \vec{v}_2 \\ \vdots \\ \vec{v}_k \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1n} \\ v_{21} & v_{22} & \dots & v_{2n} \\ \vdots & & & \\ v_{k1} & v_{k2} & \dots & v_{kn} \end{bmatrix} \quad (8.8)$$

Exemple. Dans le cas des signaux MPEG, la matrice génératrice se ramène à

$$G = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \quad (8.9)$$

On peut voir les 3 vecteurs $\vec{v}_1, \vec{v}_2, \vec{v}_3$ ainsi déterminés comme les vecteurs de base d'un sous-espace vectoriel d'un espace vectoriel à 7 dimensions défini à partir du champ binaire $(\{0, 1\}, \oplus, \times)$ où \oplus et \times représentent respectivement l'addition modulo 2 et la multiplication. Par exemple, le mot codé pour le message (110) est

$$\vec{c} = (110) \begin{bmatrix} \vec{v}_1 \\ \vec{v}_2 \\ \vec{v}_3 \end{bmatrix} = 1 \times \vec{v}_1 \oplus 1 \times \vec{v}_2 \oplus 0 \times \vec{v}_3 \quad (8.10)$$

$$= (1101100) \oplus (1110010) \oplus (0000000) \quad (8.11)$$

$$= (0011110) \quad (8.12)$$

En prenant toutes les combinaisons possibles pour les éléments du message (il y en a 2^k), on génère 2^k vecteurs dispersés dans un espace à n dimensions. En d'autres termes, l'effet du code est de répartir 2^k messages dans un espace comprenant 2^n messages. Plus ces messages seront éloignés entre eux, plus la probabilité de les confondre diminuera et plus la résistance au bruit augmentera. C'est là tout l'enjeu des codes.

8.3.4 Codes à parité

Codes à détection de parité simple

Les codes à parité simple se contentent d'ajouter un bit de parité pour la détection ou correction d'erreur. La valeur de la parité est calculée de manière à assurer que la somme des bits du message plus le bit de parité soit égale à 1, on parle alors de parité impaire, ou égale à 0, il s'agit dans ce cas de parité paire. La figure 8.3 montre le principe des codes de parité paire pour des transmissions série et parallèle.

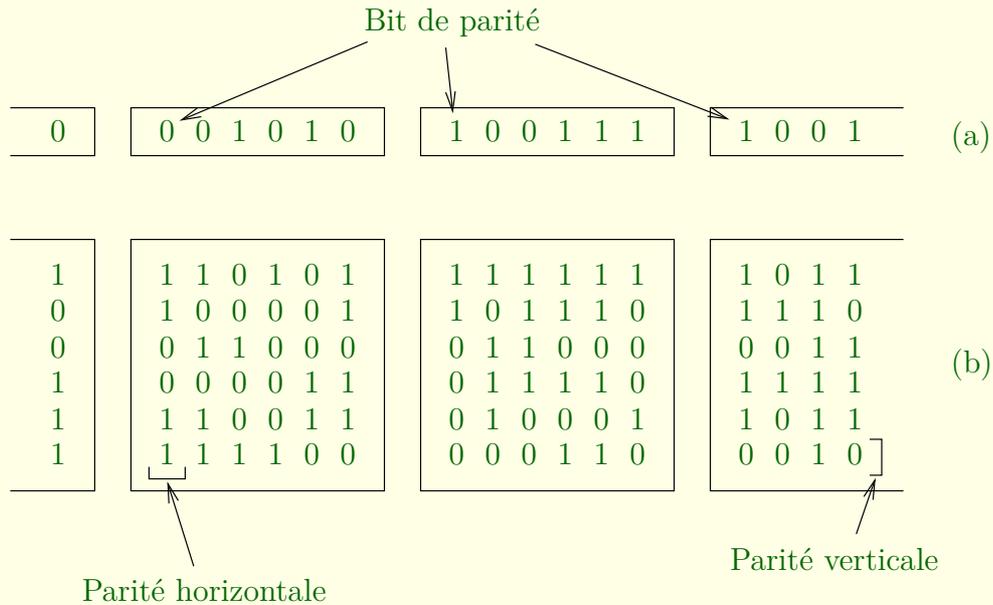


FIGURE 8.3 – Codes de parité paire pour (a) connexion série ou (b) parallèle.

Le récepteur se charge de vérifier si la parité est respectée. Si la parité n'est pas respectée, le récepteur peut en déduire la présence d'une erreur. L'erreur peut être de plusieurs types :

- bit de parité erroné,
- bit du message erroné, ou
- nombre impair de bits erronés dans le mot codé.

Le récepteur n'est pas pour autant en mesure de corriger systématiquement l'erreur. Il devra, pour bien faire, analyser toute l'information dont il dispose, c'est-à-dire l'ensemble des bits de parité, et en déduire un résultat aussi proche que possible du mot codé réel.

Supposons que les erreurs soient de probabilité égale sur tous les bits de \vec{c} . La probabilité d'avoir j erreurs dans un bloc de taille n est

$$p(j, n) = \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} \quad (8.13)$$

où p représente la probabilité d'avoir un bit erroné. Dans le cas d'un bit de parité unique, la probabilité de ne pas détecter un bit erroné vaut donc

$$\sum_{j=1}^{\begin{matrix} n/2 & n \text{ pair} \\ (n-1)/2 & n \text{ impair} \end{matrix}} \frac{n!}{(2j)!(n-2j)!} p^{2j} (1-p)^{n-2j} \quad (8.14)$$

La figure 8.4 montre la probabilité d'erreur P_e en fonction du rapport E_b/N_0 d'un démodulateur dont les messages ont été préalablement encodés par des codes (n, k) .

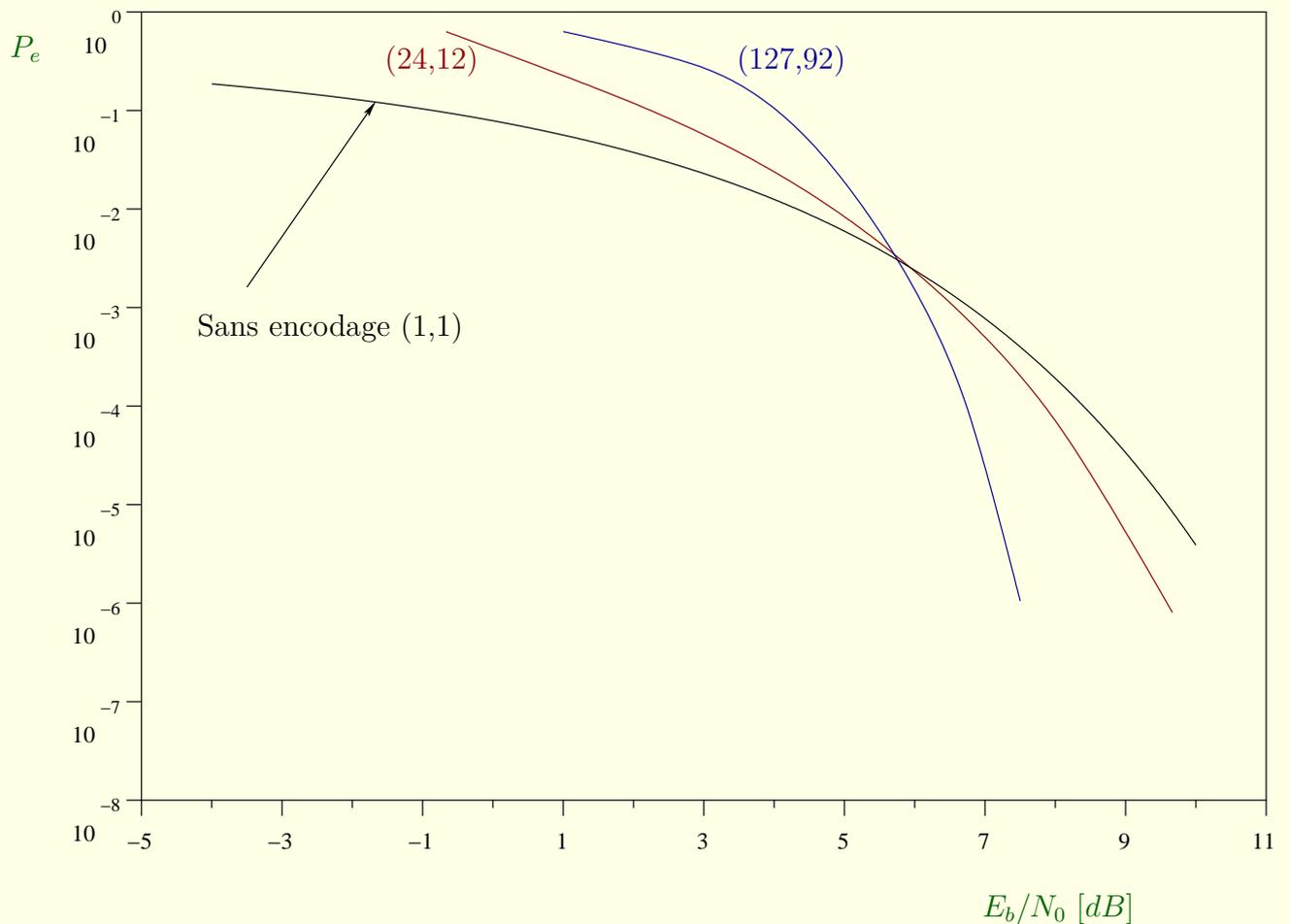


FIGURE 8.4 – Performance d'une détection PSK après codage.

La courbe $(1, 1)$ représente la PSK à deux états (BPSK) sans encodage. Dans cet exemple, l'intérêt de l'encodage se manifeste au-delà d'une valeur de E_b/N_0 de $5,5$ [dB]. En-dessous de cette valeur, le codage introduit une perte de l'énergie E_b qui n'est pas compensée par une diminution de l'effet de bruit.

8.3.5 Code systématique

Définition 77 *Un code est dit systématique si une partie du mot codé coïncide avec le message.*

Dans le cas d'un code systématique, il est d'usage de placer le message original dans la seconde

partie du mot codé. Dès lors, la matrice G s'exprime comme

$$G = [P \mid I_k] \quad (8.15)$$

$$= \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1(n-k)} & 1 & 0 & \dots & 0 \\ p_{21} & p_{22} & \dots & p_{2(n-k)} & 0 & 1 & \dots & 0 \\ \vdots & & & & & & \vdots & 0 \\ p_{k1} & p_{k2} & \dots & p_{k(n-k)} & 0 & 0 & \dots & 1 \end{bmatrix} \quad (8.16)$$

Et donc

$$\vec{c} = (c_1, c_2, \dots, c_n) = (m_1, m_2, \dots, m_k) \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1(n-k)} & 1 & 0 & \dots & 0 \\ p_{21} & p_{22} & \dots & p_{2(n-k)} & 0 & 1 & \dots & 0 \\ \vdots & & & & & & \vdots & 0 \\ p_{k1} & p_{k2} & \dots & p_{k(n-k)} & 0 & 0 & \dots & 1 \end{bmatrix} \quad (8.17)$$

tel que

$$\begin{aligned} c_i &= m_1 p_{1i} + m_2 p_{2i} + \dots + m_k p_{ki} & i = 1, \dots, (n-k) \\ &= m_{i-n+k} & i = (n-k+1), \dots, n \end{aligned} \quad (8.18)$$

Finalement, le mot codé est constitué de bits de parité précédant le message

$$\vec{c} = \underbrace{p_1, p_2, \dots, p_{n-k}}_{\text{parité}} \underbrace{m_1, m_2, \dots, m_k}_{\text{message}} \quad (8.19)$$

Exemple. Dans le cas de MPEG,

$$\vec{c} = (m_1 \oplus m_2 \oplus m_3, m_1 \oplus m_2, m_1 \oplus m_3, m_2 \oplus m_3, m_1, m_2, m_3) \quad (8.20)$$

Cette expression permet de deviner les raisons qui ont conduit au choix des bits de parité. Les bits c_2, c_3, c_4 sont tels qu'on essaie d'isoler une erreur sur un bit du message. Le bit c_1 est lui plutôt destiné à indiquer des erreurs multiples.

8.4 Détection et correction d'erreur

L'introduction de bits de parité diminue la probabilité d'erreur, à condition de travailler avec un rapport signal à bruit suffisant. Il reste néanmoins des erreurs qu'il convient de détecter et de corriger dans la mesure du possible. La détection d'erreurs se fait par le biais de la matrice de contrôle de parité et du syndrome d'erreur.

8.4.1 Matrice de contrôle de parité

Soit H la matrice appelée matrice de contrôle de parité, destinée à permettre le contrôle d'erreur. Pour toute matrice G de taille $(k \times n)$, il existe une matrice H , de taille $(n-k, n)$, telle que les lignes de H soient orthogonales aux lignes de la matrice G ; autrement dit, $GH^T = \underline{0}$, où H^T est la transposée de la matrice H et $\underline{0}$ une matrice $(k, n-k)$ dont tous les éléments sont nuls. Pour respecter la contrainte d'orthogonalité, la matrice H a la forme

$$H = [I_{n-k} \mid P^T] \quad (8.21)$$

ce qui conduit à la forme transposée

$$H^T = \left[\begin{array}{c} \frac{I_{n-k}}{P} \end{array} \right] \quad (8.22)$$

$$\left[\begin{array}{cccc} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & 1 \\ p_{11} & p_{12} & \dots & p_{1(n-k)} \\ p_{21} & p_{22} & \dots & p_{2(n-k)} \\ \vdots & & & \\ p_{k1} & p_{k2} & \dots & p_{k(n-k)} \end{array} \right] \quad (8.23)$$

On vérifie aisément que tout produit $\vec{c}H^T$ pour tout \vec{m} généré au moyen de la matrice génératrice G fournit un vecteur nul

$$\vec{c}H^T = (p_1 + p_1, p_2 + p_2, \dots, p_{n-k} + p_{n-k}) = 0 \quad (8.24)$$

Ce système d'équations fournit un outil précieux pour la détection des erreurs de transmission lors du décodage en réception. En effet, un mot codé \vec{c} a été généré par G si et seulement si $\vec{c}H^T = 0$.

8.4.2 Vecteur syndrome d'erreur

En réception, les équations 8.24 de contrôle de parité ne peuvent être vérifiées qu'en l'absence d'erreur de transmission.

Le vecteur \vec{e} , dit vecteur d'erreur, correspond à l'erreur de transmission qui se superpose au mot \vec{c} transmis. En réception, on observe donc le vecteur

$$\vec{r} = \vec{c} + \vec{e} \quad (8.25)$$

Pour savoir s'il y a erreur de transmission ou non, il est naturel de former le vecteur

$$\vec{s} = \vec{r}H^T \quad (8.26)$$

Définition 78 *Le vecteur $\vec{s} = \vec{r}H^T$ est appelé vecteur syndrome d'erreur ou plus simplement syndrome.*

En développant l'expression du syndrome,

$$\vec{s} = (\vec{c} + \vec{e})H^T \quad (8.27)$$

$$= \vec{c}H^T + \vec{e}H^T \quad (8.28)$$

$$= \vec{e}H^T \quad (8.29)$$

Dès lors, si le vecteur \vec{s} est le vecteur nul, il n'y a pas eu d'erreur de transmission. Sinon la transmission a été affectée par du bruit.

8.4.3 Correction d'erreur

La détection d'erreur peut parfois s'accompagner d'une correction d'erreur. Pour arriver à une expression analytique des erreurs qui peuvent être corrigées, il convient de disposer d'une table reliant la valeur du syndrome à l'erreur \vec{e} correspondante.

On obtient cette table en construisant une table à double entrée comprenant tous les mots codés possibles ainsi que toutes les vecteurs d'erreur (voir le tableau 8.3). Chaque ligne contient un vecteur d'erreur ainsi que le résultat de la superposition de ce message d'erreur sur l'ensemble des mots de code.

	\vec{c}_1	...	\vec{c}_i	...	\vec{c}_{2^k}
\vec{e}_1	$\vec{c}_1 + \vec{e}_1$...	$\vec{c}_i + \vec{e}_1$...	$\vec{c}_{2^k} + \vec{e}_1$
\vec{e}_2	$\vec{c}_1 + \vec{e}_2$...	$\vec{c}_i + \vec{e}_2$...	$\vec{c}_{2^k} + \vec{e}_2$
\vdots	\vdots		\vdots		\vdots
\vec{e}_j	$\vec{c}_1 + \vec{e}_j$...	$\vec{c}_i + \vec{e}_j$...	$\vec{c}_{2^k} + \vec{e}_j$
\vdots	\vdots		\vdots		\vdots
$\vec{e}_{2^{n-k}}$	$\vec{c}_1 + \vec{e}_{2^{n-k}}$...	$\vec{c}_i + \vec{e}_{2^{n-k}}$...	$\vec{c}_{2^k} + \vec{e}_{2^{n-k}}$

TABLE 8.3 – Ensemble des combinaisons de mots codés et de vecteurs d'erreur.

Au total, le tableau contient 2^n n -uplets de l'espace vectoriel à n dimensions. Un calcul simple permet d'obtenir les syndromes de toutes ces combinaisons vectorielles. En effet, soit le vecteur $\vec{r} = (\vec{c}_i + \vec{e}_j)$ obtenu à la réception. Le syndrome de ce vecteur vaut

$$\vec{s} = (\vec{c}_i + \vec{e}_j) H^T = \vec{c}_i H^T + \vec{e}_j H^T \quad (8.30)$$

Mais comme \vec{c}_i est un mot codé, $\vec{c}_i H^T = \underline{0}$. Dès lors,

$$\vec{s} = (\vec{c}_i + \vec{e}_j) H^T = \vec{e}_j H^T \quad (8.31)$$

Autrement dit, le syndrome dépend de l'erreur et non pas de la valeur de message. De plus, cette relation montre que tous les éléments d'une même ligne du tableau 8.3 ont un syndrome identique. De cette manière, il y a une correspondance entre un syndrome et un vecteur d'erreur. On en déduit l'algorithme de correction d'erreur suivant :

1. Calcul du syndrome $\vec{s} = \vec{r} H^T$ sur base du signal reçu.
2. Détermination du vecteur d'erreur \vec{e}_j correspondant.
3. Estimation du mot codé réel au moyen de $\vec{c} = \vec{r} \oplus \vec{e}_j$.

8.5 Efficacité de codage

8.5.1 Poids et distance de Hamming

La discussion précédente a clairement montré que toutes les erreurs ne pouvaient pas être corrigées. Avant de déterminer la capacité de correction d'un code, il faut pouvoir le caractériser.

Définition 79 Le poids de HAMMING $w(\vec{c})$ du vecteur \vec{c} est le nombre de 1 qu'il contient.

Exemple. Voici le poids de deux vecteurs

$$\begin{aligned} w(\vec{c}_1) &= w(100101101) = 5 \\ w(\vec{c}_2) &= w(011110100) = 5 \end{aligned} \quad (8.32)$$

On peut aussi définir la distance entre deux vecteurs.

Définition 80 Soient deux vecteurs binaires \vec{c}_1, \vec{c}_2 , la distance de HAMMING $d(\vec{c}_1, \vec{c}_2)$ est le nombre de bits qui diffèrent.

Reprenons l'exemple précédent. $d(\vec{c}_1, \vec{c}_2) = 6$. On montre aisément que poids et distance sont liés par la relation

$$d(\vec{c}_1, \vec{c}_2) = w(\vec{c}_1 \oplus \vec{c}_2) \quad (8.33)$$

8.5.2 Distance minimale

Considérons à présent l'ensemble de paires de vecteurs d'un espace vectoriel. Le code est caractérisé par la plus petite distance entre ses vecteurs, d_{\min} . Cette distance définit le maillon le plus faible du code car elle détermine le cas le plus contraignant de résistance au bruit.

Par définition, l'espace linéaire est tel que toute combinaison linéaire de mots codés est également un mot codé. Par exemple $d(\vec{c}_1, \vec{c}_2) = w(\vec{c}_1 \oplus \vec{c}_2) = w(\vec{c}_3)$. Dès lors, pour déterminer la distance entre deux mots codés, on peut se contenter de déterminer le poids de l'ensemble des mots codés²; le poids minimum correspond à la distance minimale.

8.5.3 Détection et correction d'erreur

Le récepteur a pour mission de déterminer le mot émis \vec{c}_i sur base du mot reçu \vec{r} . La stratégie consiste à recourir au maximum de vraisemblance. On choisit le vecteur \vec{c}_i qui vérifie la relation

$$p(\vec{r} | \vec{c}_i) = \max_{\vec{c}_j} p(\vec{r} | \vec{c}_j) \quad (8.34)$$

Dans le cas le plus simple, le vecteur \vec{c}_i est choisi tel que

$$d(\vec{r}, \vec{c}_i) = \min_{\vec{c}_j} d(\vec{r}, \vec{c}_j) \quad (8.35)$$

En d'autres termes, le décodeur détermine la distance entre le mot reçu et tous les mots, avant de sélectionner le mot de code le plus proche.

Le processus de décision est conceptuellement similaire à celui utilisé lors de la démodulation de signaux ASK par exemple. On détermine une frontière entre les signaux possibles et on choisit un signal suivant la zone à laquelle il appartient. Il en va de même ici. Le seuil est défini par

$$t = \text{arrondi}_- \frac{d_{\min} - 1}{2} \quad (8.36)$$

où arrondi_- est l'arrondi à la valeur entière inférieure. Pour t , on parle de capacité de correction d'erreur. La grandeur t représente en fait le nombre maximum d'erreurs que l'on peut corriger.

2. On ne considère évidemment pas le mot codé nul.

8.6 Quelques codes courants

8.6.1 Code de Hamming

Les codes de HAMMING constituent un sous-ensemble des codes en blocs pour lesquels (n, k) valent

$$(n, k) = (2^m - 1, 2^m - 1 - m) \quad (8.37)$$

pour $m = 2, 3, \dots$. Ces codes ont une distance minimale de 3. En conséquence, ils permettent la correction d'une erreur simple ou la détection de toute combinaison d'au plus deux erreurs.

SKLAR [17, page 298] montre que la probabilité d'erreur s'écrit

$$P_B \simeq p - p(1 - p)^{n-1} \quad (8.38)$$

où p est la probabilité d'erreur du canal.

Il est utile de relier cette relation au taux d'erreur de détection classique. Prenons le cas d'une démodulation BPSK cohérente. Le taux d'erreur due à un bruit additif gaussien dans le canal est

$$P_e = \frac{1}{2} \left(\sqrt{\frac{E_c}{N_0}} \right) \quad (8.39)$$

E_c représente l'énergie par symbole (binaire) du message source. On relie aisément l'énergie par bit réellement transmis E_b à E_c

$$E_b = \frac{n}{k} E_c \quad (8.40)$$

Ce qui, dans le cas d'un code de HAMMING, conduit à déterminer la probabilité d'erreur par bit transmis sur base du changement de variable suivant

$$\frac{E_c}{N_0} = \frac{2^m - 1 - m}{2^m - 1} \frac{E_b}{N_0} \quad (8.41)$$

dans l'expression 8.39. Cette relation est utile si l'on veut comparer les performances d'un code comprenant ou non un précodage de redondance car elle permet de normaliser l'énergie par bit lors du tracé des courbes d'erreur. C'est ainsi qu'on calcule l'expression des courbes d'erreur correspondantes à différents couples de précodage.

Le code a pour effet de modifier l'inclinaison de la courbe de probabilité d'erreur. Au-delà d'une certaine valeur du rapport énergie par bit à puissance de bruit (E_b/N_0), il est préférable. Au droit du point d'intersection d'une courbe de précodage avec la courbe sans précodage, un accroissement d'un dB signifie une différence de taux de probabilités d'erreur qui peut atteindre plusieurs ordres de grandeur.

Mais que l'on ne s'y trompe pas, l'utilisation d'un précodage n'a pas que des avantages. En effet, un taux de redondance de $1/2$ entraîne le doublement de bande passante. Or, près de $0 [dB]$, l'absence de précodage garantit encore la meilleure performance. Utiliser un précodage dans cette plage conduirait à accroître la bande de fréquences tout en diminuant les performances.

8.6.2 Code de Golay étendu

Un des codes les plus utiles est le code binaire $(24, 12)$, appelé code de GOLAY étendu, formé en ajoutant un bit de parité couvrant la totalité d'un code $(23, 12)$. L'ajout de ce bit fait basculer la distance minimale de 7 à 8. De plus, le taux de codage d'un demi est facilement réalisable.

Le taux d'erreur est significativement plus faible que celui d'un code de HAMMING.

8.6.3 Code BCH

Les codes de BOSE-CHADHURI-HOCQUENGHEM (BCH) constituent une généralisation des codes de HAMMING. Il s'agit de codes cycliques permettant, entre autres, la correction d'erreurs multiples.

8.6.4 Autres codes

Pour être complet, mentionnons encore l'existence d'autres codes :

- Codes de REED-SOLOMON. Ces codes font partie des codes BCH non binaires ; ils sont antérieurs à la généralisation effectuée lors de la découverte des codes BCH.
- Turbo-codes. Il s'agit de codes, relativement récents, ayant des propriétés remarquables pour des canaux particulièrement bruités, là où les codes plus anciens ne conviennent pas. Ces codes combinent des bits de parité horizontaux avec des bits de parité verticaux.

Chapitre 9

Propagation et systèmes radio

9.1 Introduction

Dans un canal radio, le canal de transmission consiste dans la liaison entre deux antennes par une onde électromagnétique ; celle-ci n'est généralement pas guidée, mais se propage librement en subissant cependant certains effets dus à la présence de la terre et de son atmosphère. Ces effets déterminent la qualité de toute transmission.

L'organe de régulation radio de l'ITU définit le terme d'ondes radio pour des fréquences arbitrairement inférieures à 3000 GHz, limite que certains auteurs n'hésitent pas à franchir quand ils parlent d'ondes radio. La table 9.1 reprend les principales bandes de fréquences, par décades, telles que définies par l'ITU secteur des radiocommunications.

Fréquences	λ [m]	Dénomination	
< 3 [kHz]	> 100 [km]		ELF
3 – 30 [kHz]	10 – 100 [km]	Ondes myriamétriques	VLF
30 – 300 [kHz]	1 – 10 [km]	Ondes kilométriques	LF
300 – 3000 [kHz]	100 – 1000 [m]	Ondes hectométriques	MF
3 – 30 [MHz]	10 – 100 [m]	Ondes décamétriques	HF
30 – 300 [MHz]	1 – 10 [m]	Ondes métriques	VHF
300 – 3000 [MHz]	10 – 100 [cm]	Ondes décimétriques	UHF
3 – 30 [GHz]	1 – 10 [cm]	Ondes centimétriques	SHF
30 – 300 [GHz]	1 – 10 [mm]	Ondes millimétriques	EHF
300 – 3000 [GHz]	0,1 – 1 [mm]		
3 – 30 [THz]	10 – 100 [μ m]		
30 – 430 [THz]	0,7 – 10 [μ m]		
430 – 860 [THz]	0,35 – 0,7 [μ m]	Ondes optiques	

TABLE 9.1 – Nomenclature de l'ITU-R pour les bandes du spectre électromagnétique.

Dans ce chapitre, nous examinerons la propagation en espace libre plutôt que les ondes guidées, les lignes ou les fibres optiques. Il convient toutefois de préciser qu'à certaines fréquences, l'ionosphère, les couches d'indices de réfraction distinctes et la terre jouent le rôle d'un guide d'onde.

9.2 Propagation des ondes électromagnétiques

9.2.1 Équations de Maxwell

La propagation des ondes électromagnétiques est régie par les équations de MAXWELL. Ces équations établissent un lien entre les sources de rayonnement et les champs électromagnétiques. On peut les écrire sous plusieurs formes : les formes locale ou intégrale, d'une part, décrites dans le domaine temporel ou fréquentiel, d'autre part.

Ces équations constituent l'outil de base de tous ceux qui ont à appliquer ou développer l'électricité : génération et transmission de puissance, composants électroniques, radio, hyperfréquences, radar, fibres optiques, etc. Dans le domaine temporel et sous forme locale, les équations de MAXWELL s'écrivent [14]

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad (9.1)$$

$$\nabla \times \vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t} \quad (9.2)$$

$$\nabla \cdot \vec{D} = \rho \quad (9.3)$$

$$\nabla \cdot \vec{B} = 0 \quad (9.4)$$

À l'exception de la densité de charge ρ , toutes les grandeurs sont des champs de \mathbb{R}^3 , variables dans le temps. Le champ électrique \vec{E} [V/m] est généré par la variation temporelle de l'induction magnétique, parfois appelé champ magnétique, \vec{B} [tesla] ou [Vs/m²]. Quant au champ magnétique \vec{H} [A/m], il peut résulter de la présence d'une densité de courant \vec{J} [C/m²] ou d'une variation de la densité de flux électrique \vec{D} .

Énergie, puissance et impédance électromagnétiques

Une onde électromagnétique véhicule une certaine puissance. L'essence des transmissions sans fil consiste à transmettre un signal dans une certaine gamme de fréquence de manière à assurer un flux de puissance suffisant à la réception. Étudions donc tout d'abord l'expression de la puissance d'une onde électromagnétique.

Le produit des amplitudes du champ électrique et du champ magnétique $|\vec{E}| |\vec{H}|$ a pour unité [VA/m²], soit encore [W/m²]. C'est donc l'expression d'une puissance sur une certaine surface.

Définition 81 [Vecteur de Poynting] Le vecteur de POYNTING \vec{S} est défini comme le produit vectoriel

$$\vec{S} = \vec{E} \times \vec{H} \quad (9.5)$$

Le vecteur \vec{S} est perpendiculaire au plan formé par \vec{E} et \vec{H} . De par sa nature, il représente un flux de puissance, d'où l'intérêt d'analyser sa divergence. Comme $\nabla \cdot (\vec{A} \times \vec{B}) = \vec{B} \cdot (\nabla \times \vec{A}) - \vec{A} \cdot (\nabla \times \vec{B})$, il vient

$$\nabla \cdot \vec{S} = \vec{H} \cdot (\nabla \times \vec{E}) - \vec{E} \cdot (\nabla \times \vec{H}) \quad (9.6)$$

Par substitution de 9.1 et de 9.2 dans 9.6,

$$\nabla \cdot \vec{S} = -\vec{H} \cdot \frac{\partial \vec{B}}{\partial t} - \vec{E} \cdot \vec{J} - \vec{E} \cdot \frac{\partial \vec{D}}{\partial t} \quad (9.7)$$

ou, après regroupement,

$$-\nabla \cdot \vec{S} = \vec{E} \cdot \vec{J} + \left(\vec{E} \cdot \frac{\partial \vec{D}}{\partial t} + \vec{H} \cdot \frac{\partial \vec{B}}{\partial t} \right) \quad (9.8)$$

équation connue sous le nom de forme locale du théorème de POYNTING, valable en tout point à tout instant. Si on applique le théorème de la divergence sur l'intégrale volumique des membres de l'égalité précédente, on obtient l'égalité entre puissances

$$-\oint_s \vec{S} \cdot d\vec{s} = \int_v \vec{E} \cdot \vec{J} dv + \int_v \left(\vec{E} \cdot \frac{\partial \vec{D}}{\partial t} + \vec{H} \cdot \frac{\partial \vec{B}}{\partial t} \right) dv \quad (9.9)$$

Deux composantes contribuent donc au flux net entrant, $-\oint_s \vec{S} \cdot d\vec{s}$, dans le volume délimité par la surface S :

– $\int_v \vec{E} \cdot \vec{J} dv$, la puissance dissipée, P_R , par le champ électrique pour déplacer des charges.

Pour un milieu où seule intervient la conduction, $\vec{J} = \sigma \vec{E}$, d'où

$$P_R = \sigma \int_v |\vec{E}|^2 dv \quad (9.10)$$

ce qui indique qu'il s'agit de la puissance dissipée dans une résistance.

– le terme $\int_v \left(\vec{E} \cdot \frac{\partial \vec{D}}{\partial t} + \vec{H} \cdot \frac{\partial \vec{B}}{\partial t} \right) dv$, qui a une signification thermodynamique précise dans les milieux non dispersifs : il est la variation de la différence d'énergie interne emmagasinée dans le milieu par suite de l'application du champ électromagnétique, à densité et entropie constantes. Dans ce cas d'ailleurs, il est souvent mis sous une autre forme. Ainsi, dans un milieu pour lequel $\vec{D} = \epsilon \vec{E}$ et $\vec{B} = \mu \vec{H}$,

$$\vec{E} \cdot \frac{\partial \vec{D}}{\partial t} = \epsilon \vec{E} \cdot \frac{\partial \vec{E}}{\partial t} \quad (9.11)$$

$$= \frac{1}{2} \epsilon \frac{\partial |\vec{E}|^2}{\partial t} \quad (9.12)$$

De même,

$$\vec{H} \cdot \frac{\partial \vec{B}}{\partial t} = \frac{1}{2} \mu \frac{\partial |\vec{H}|^2}{\partial t} \quad (9.13)$$

Le deuxième terme est donc l'expression de la variation temporelle d'énergie électrique et magnétique à l'intérieur du volume v .

Dès lors, sous la forme intégrale,

$$-\oint_s \vec{S} \cdot d\vec{s} = \sigma \int_v |\vec{E}|^2 dv + \frac{\partial}{\partial t} \int_v \left(\frac{\epsilon}{2} |\vec{E}|^2 + \frac{\mu}{2} |\vec{H}|^2 \right) dv \quad (9.14)$$

Moyennant ces hypothèses, le théorème de POYNTING établit que le flux net entrant du vecteur de POYNTING à l'intérieur d'une surface fermée est la somme de la puissance dissipée sous

l'effet de la conductivité du milieu et du taux de changement d'énergie électromagnétique emmagasinée dans le volume ainsi délimité. Remarquons que c'est la divergence $\nabla \cdot \vec{S}$ qui est un flux de puissance électromagnétique et non le vecteur \vec{S} lui-même. En particulier, un système statique caractérisé par des lignes de champ électrique et de champ magnétique qui se croisent, a un vecteur de POYNTING non nul alors que la divergence de ce vecteur peut être nulle en tout point. Dans ce cas, il n'y a pas de flux d'énergie.

La théorie des lignes montre que le concept de puissance est souvent plus commode à utiliser que ceux de tension et de courant. Il importe dès lors de définir un certain nombre de termes.

Définition 82 [Adaptation conjuguée] C'est la condition d'absorption maximum de puissance par une charge. Dans cette condition, l'impédance vue vers la charge en un point de la ligne est égale à l'impédance complexe conjuguée de celle vue vers la source.

Définition 83 [Adaptation de ligne] C'est la condition sous laquelle l'impédance d'entrée de la ligne est égale à son impédance caractéristique.

9.3 Antennes

Une antenne d'émission est un dispositif qui émet des ondes électromagnétiques. Quant à l'antenne de réception, elle se charge de capter les ondes électromagnétiques et elle les traduit en un signal électrique disponible aux extrémités du dipôle électrique qu'elle représente en sortie. Ce sont les caractéristiques du signal électrique de sortie en regard d'un signal électrique d'entrée, à l'émission, qui quantifient la qualité de la transmission. Pour arriver à comprendre ces caractéristiques, il nous faut étudier les solutions aux équations de MAXWELL pour des structures d'antennes typiques. Cette étude conduit à établir la formulation d'un bilan de puissance entre antennes, bilan qui se révèle être le principal outil pratique de dimensionnement d'un réseau de radiocommunications.

9.3.1 Propriétés générales à l'émission

Si l'on écrit les équations de MAXWELL en notation phasorielle, ce qui présuppose un fonctionnement en régime sinusoïdal, on a

$$\nabla \times \hat{E} = -j\omega \hat{B} \quad (9.15)$$

$$\nabla \times \hat{H} = \hat{J} + j\omega \hat{D} \quad (9.16)$$

$$\nabla \cdot \hat{D} = \hat{\rho} \quad (9.17)$$

$$\nabla \cdot \hat{B} = 0 \quad (9.18)$$

Remarquons que les équations de MAXWELL n'introduisent pas les caractéristiques propres aux matériaux dans lesquels l'onde est émise. Expérimentalement, on est parvenu à relier induction et champ, tant électrique que magnétique. Mais ces paramètres (la perméabilité, la permittivité et la conductivité) peuvent évidemment être fonctions de la fréquence de l'onde.

Une antenne est un corps muni d'un accès électrique. Si l'on raccorde un doublet générateur à l'accès de l'antenne, l'antenne est parcourue par un courant. En espace libre, les rotationnels

s'écrivent

$$\nabla \times \hat{E} = -j\omega\mu_0\hat{H} \quad (9.19)$$

$$\nabla \times \hat{H} = \hat{J}_s + j\omega\epsilon_0\hat{E} \quad (9.20)$$

où \hat{J}_s représente le phaseur de la densité de courant de surface.

9.3.2 Détermination d'un potentiel vecteur

Sachant que $\nabla \cdot \nabla \times \hat{A} = 0$ pour tout champ vectoriel, et que $\nabla \times \nabla V = 0$ pour toute fonction scalaire, on peut choisir de définir un potentiel vecteur magnétique

$$\hat{B} = \nabla \times \hat{A} \quad (9.21)$$

et comme

$$\nabla \times \hat{E} = -j\omega\hat{B} = -j\omega\nabla \times \hat{A} \Rightarrow \nabla \times (\hat{E} + j\omega\hat{A}) = 0 \quad (9.22)$$

on peut définir \hat{V} , une fonction potentiel, telle que

$$\hat{E} = -j\omega\hat{A} - \nabla\hat{V} \quad (9.23)$$

Il suffit dès lors de connaître les potentiels \hat{A} , \hat{V} pour déterminer les champs électriques et magnétiques.

Après calcul et grâce à la condition de LORENTZ¹ qui stipule que $\nabla \cdot \hat{A} = -j\omega\mu_0\epsilon_0\hat{V}$,

$$\nabla^2\hat{A} + \omega^2\mu_0\epsilon_0\hat{A} = -\mu_0\hat{J}_s \quad (9.24)$$

Cette relation est typique de l'équation d'onde avec un terme de source dans le membre de droite. Définissant $\beta_0 = \omega\sqrt{\mu_0\epsilon_0}$, la solution est

$$\hat{A} = \frac{\mu_0}{4\pi} \int_v \frac{\hat{J}_s e^{-j\beta_0 r}}{r} dv \quad (9.25)$$

où v représente le volume contenant entièrement \hat{J}_s et r est la distance qui sépare l'élément infinitésimal dv du droit du champ \hat{A} .

En principe, connaître la distribution de courant de surface le long d'une antenne permet de déduire le potentiel vecteur magnétique et donc, par le biais de calculs vectoriels, les champs électriques et magnétiques. Pour des formes d'antenne simple, il est possible de calculer \hat{J}_s . Déterminer la densité de courant pour des antennes complexes s'avère analytiquement difficile ; on recourt alors plutôt à un bilan de puissance tel que celui de l'équation de FRIIS (cf. page 198).

9.3.3 Propriétés générales à la réception

Les propriétés générales à la réception sont identiques aux propriétés générales à l'émission en vertu du théorème de réciprocité. Dès lors, il suffit d'examiner le comportement à l'émission pour en déduire les propriétés à la réception. Les notions de directivité, de résistance de rayonnement et autres auront ainsi la même valeur à l'émission et à la réception.

1. On peut montrer que la condition de LORENTZ est équivalente à la loi de la conservation de la charge.

9.3.4 Antennes simples

L'étude des doublets est intéressante car elle éclaire le fonctionnement des mécanismes de calculs des champs.

En plus de leur fonction de guide d'onde, une ligne non blindée ou un micro-ruban rayonnent une partie de leur énergie dans l'air. Il ne s'agira pas pour autant de s'en servir comme antenne car leurs propriétés au rayonnement ne le justifient pas. On peut usiner des structures métalliques ou diélectriques de manière à rayonner efficacement dans l'air avec, en plus, la possibilité de définir des directions privilégiées.

Doublet de Hertz

Le doublet électrique ou doublet de HERTZ constitue une structure d'antenne simplifiée : un élément de courant infinitésimal de longueur dl est parcouru par un courant décrit par le phaseur \hat{I} supposé constant (en amplitude et en phase) sur toute sa longueur. La figure 9.1 montre un doublet placé dans un système d'axes cartésien et sphérique.

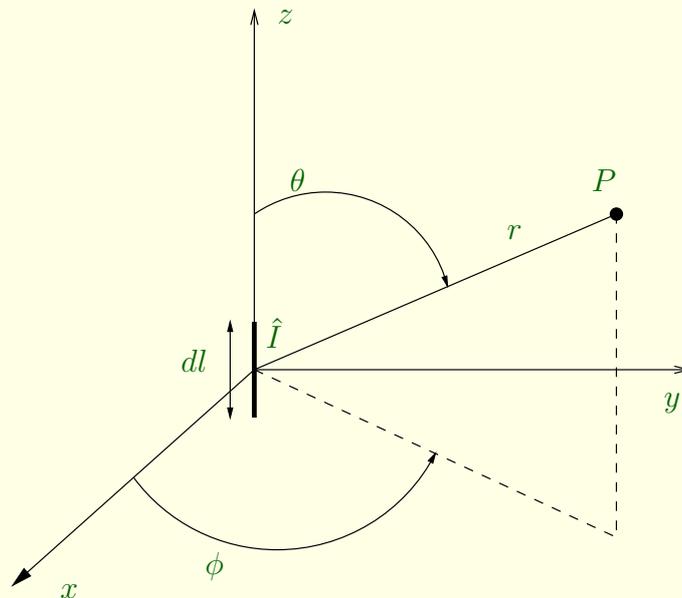


FIGURE 9.1 – Doublet électrique.

L'antenne raccordée à un doublet générateur crée des ondes électromagnétiques en tout point d'observation P repéré par une distance r définie par rapport à l'origine. Si le doublet est aligné sur l'axe z , le potentiel vecteur se réduit à

$$\hat{A}_z = \frac{\mu_0}{4\pi} \hat{I} dl \frac{e^{-j\beta_0 r}}{r} \quad (9.26)$$

En coordonnées sphériques, on en déduit

$$\hat{H}_r = 0 \quad (9.27)$$

$$\hat{H}_\theta = 0 \quad (9.28)$$

$$\hat{H}_\phi = \frac{\hat{I} dl}{4\pi} \beta_0^2 \sin \theta \left(j \frac{1}{\beta_0 r} + \frac{1}{\beta_0^2 r^2} \right) e^{-j\beta_0 r} \quad (9.29)$$

Quant au champ électrique, il vaut

$$\hat{E}_r = 2 \frac{\hat{I} dl}{4\pi} \eta_0 \beta_0^2 \cos \theta \left(\frac{1}{\beta_0^2 r^2} - j \frac{1}{\beta_0^3 r^3} \right) e^{-j\beta_0 r} \quad (9.30)$$

$$\hat{E}_\theta = \frac{\hat{I} dl}{4\pi} \eta_0 \beta_0^2 \sin \theta \left(j \frac{1}{\beta_0 r} + \frac{1}{\beta_0^2 r^2} - j \frac{1}{\beta_0^3 r^3} \right) e^{-j\beta_0 r} \quad (9.31)$$

$$\hat{E}_\phi = 0 \quad (9.32)$$

où $\eta_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} = 120\pi = 377 \text{ } [\Omega]$ est l'impédance intrinsèque du vide. Remarquons que $\beta_0 r = 2\pi(r/\lambda_0)$, ce qui définit une sorte de longueur électromagnétique de rayonnement normalisée.

Dans un système de coordonnées sphériques dont l'axe $\theta = 0$ est orienté parallèlement au doublet, le doublet rayonne un champ magnétique avec une polarisation orientée selon les petits cercles. L'amplitude du champ rayonné varie en fonction de la direction comme $\sin \theta$. Elle se compose d'un terme en $1/r$ et d'un autre en $1/r^2$. Leurs valeurs en amplitude sont égales lorsque $r = 1/\beta_0 \simeq \lambda_0/6$. À grande distance, notée R , le terme en $1/r$ contribue principalement au champ magnétique –c'est l'hypothèse dite du champ éloigné– ; on obtient alors

$$\hat{H} = \frac{j\beta_0 \hat{I} dl}{4\pi R} \sin \theta e^{-j\beta_0 R} \vec{a}_\phi \quad (9.33)$$

Sous la même hypothèse de champ éloigné, le champ électrique vaut

$$\hat{E}_r \approx 0 \quad (9.34)$$

$$\hat{E}_\theta = \frac{j\eta_0 \beta_0 \hat{I} dl}{4\pi R} \sin \theta e^{-j\beta_0 R} \quad (9.35)$$

Ce qui montre que les champs magnétiques et électriques sont en phase et orthogonaux. De plus, ils sont égaux en norme au facteur d'impédance près. Aussi,

$$\|\hat{S}\| = \frac{\|\hat{E}_\theta\|^2}{120\pi} \quad (9.36)$$

La puissance rayonnée (phasorielle) est fournie par le vecteur de POYNTING :

$$\hat{S}_{av} = \text{Re}(\hat{E} \times \hat{H}^*) \quad (9.37)$$

Après calcul, on obtient

$$\hat{S}_{av} = 30\pi \left(\frac{dl}{\lambda_0} \right)^2 \|\hat{I}\|^2 \frac{\sin^2 \theta}{R^2} \vec{a}_r \quad (9.38)$$

Ainsi, la puissance émise par unité de surface est dirigée le long de l'axe reliant le doublet au point d'observation.

Directivité et gain. On utilise souvent des représentations graphiques pour exprimer la directivité d'une antenne par un diagramme de rayonnement. Ce diagramme peut être représenté sous diverses formes.

Ce diagramme représente l'intensité du champ électrique ou la densité de puissance pour différents angles d'observation par rapport à l'antenne. On parle de diagramme de rayonnement absolu s'il est exprimé en termes d'intensité électrique \hat{E} ou de densité de puissance \hat{S} . S'il est défini par rapport à une puissance de référence, on obtient le diagramme de rayonnement relatif².

Une forme commune du diagramme de rayonnement est la représentation polaire qui donne la forme de la zone de portée de l'antenne. Le diagramme peut être composé d'un lobe principal et de lobes secondaires ; dans ce cas, un paramètre définissant assez bien la directivité de l'antenne est l'angle d'ouverture du lobe principal à 3 [dB], θ_{3dB} . Il est illustré à la figure 9.2.

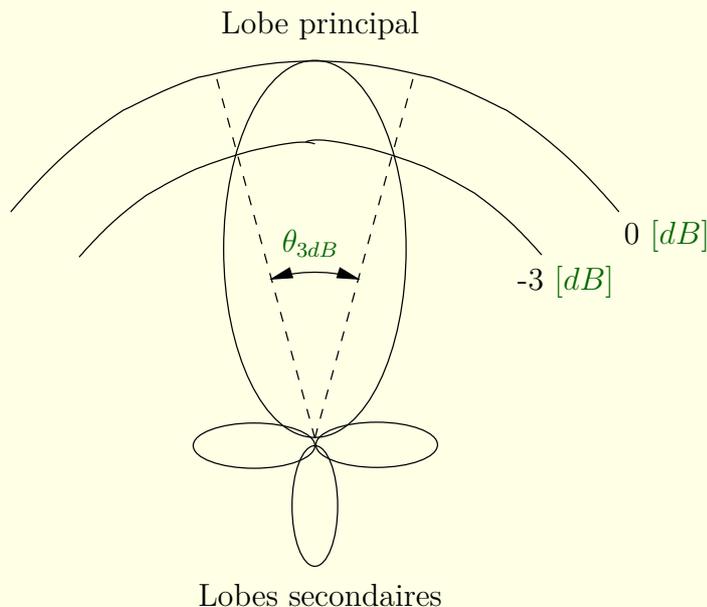


FIGURE 9.2 – Angle d'ouverture à 3 [dB].

Il est évident qu'une antenne directive économise la puissance de l'émetteur en concentrant celle-ci dans une direction préférentielle et apporte un gain. Intuitivement, on définit le gain dans une direction comme le rapport du vecteur de POYNTING dans cette direction à celui d'une antenne à rayonnement isotrope rayonnant la même puissance.

Revenons au doublet de HERTZ. Pour une distance R fixée, on peut dessiner une figure tridimensionnelle de l'amplitude de la puissance moyenne rayonnée ; il s'agit du diagramme de rayonnement qui caractérise les propriétés de l'antenne au rayonnement et ses capacités à émettre dans une direction donnée. La figure 9.3 reprend le diagramme de rayonnement du doublet électrique. On remarquera la symétrie dans le plan de l'équateur et la dépendance en $\sin \theta$.

La puissance totale rayonnée est l'intégrale de \hat{S} sur une surface incluant le doublet. En choi-

². Si l'on exprime les diagrammes en [dB], il n'y a plus de différence d'allure entre un diagramme qui représente \hat{E} ou \hat{S} .

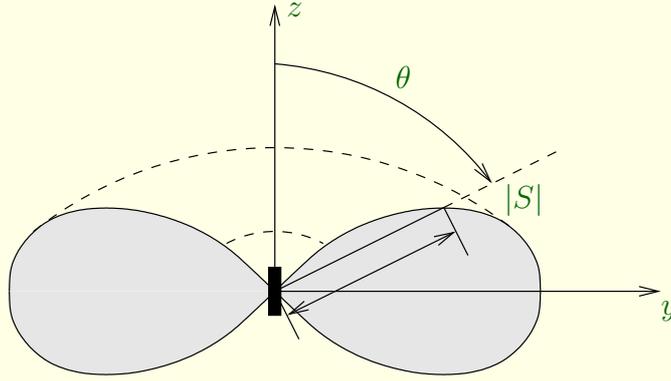


FIGURE 9.3 – Diagramme de rayonnement d'un doublet électrique.

En intégrant sur une sphère de rayon R , on obtient³

$$P_{rad} = \frac{1}{2} \oint_S \hat{S} \cdot ds \quad (9.39)$$

$$= 80\pi^2 \left(\frac{dl}{\lambda_0} \right)^2 \frac{|I|^2}{2} \quad (9.40)$$

Résistance de rayonnement. La résistance fictive de rayonnement est le rapport de la puissance rayonnée par l'antenne et du carré de l'intensité du courant (exprimé en valeur efficace si le courant est représenté par un phaseur) alimentant l'antenne

$$R_r = \frac{P_{rad}}{i^2} \quad (9.41)$$

C'est la résistance qui, en lieu et place de l'antenne, dissiperait la même puissance que la puissance rayonnée par l'antenne. Par l'équation 9.40, on calcule une résistance équivalente au rayonnement valant $80\pi^2 \left(\frac{dl}{\lambda_0} \right)^2$. Cette résistance représente une résistance fictive qui dissipe la même puissance qu'émet un doublet électrique. Les performances du doublet sont cependant médiocres. Par exemple, pour un doublet de longueur $dl = 1 [cm]$ et une fréquence de $300 [MHz]$ ($\lambda_0 = 1 [m]$), la résistance équivalente vaut $79 [m\Omega]$. Pour rayonner une puissance de $1 [W]$, il faut donc un courant de $3,6 [A]$!

Doublet magnétique

Un calcul équivalent à celui du doublet de HERTZ peut être mené pour une boucle parcourue par un courant. La structure ainsi obtenue porte le nom de doublet magnétique (cf. figure 9.4).

Son rayonnement est également médiocre. On peut ainsi montrer que le vecteur de POYNTING vaut, en champ éloigné,

$$\hat{S} = 2 \times 1860 \left(\frac{A}{\lambda_0^2} \right)^2 \|\hat{I}\|^2 \frac{\sin^2 \theta}{R^2} \vec{a}_r \quad (9.42)$$

3. Lorsqu'on utilise des phaseurs, il faut tenir compte des valeurs efficaces (c'est-à-dire la valeur de crête divisée par $\sqrt{2}$) pour le calcul des puissances. En effet, la puissance réelle d'un dipôle vaut $\frac{V}{\sqrt{2}} \frac{I}{\sqrt{2}} = \frac{VI}{2}$ où V et I sont des valeurs de crête, autrement dit l'amplitude des sinusoides. Pour éviter toute confusion, nous noterons $|E|$ pour la valeur de crête du phaseur \hat{E} .

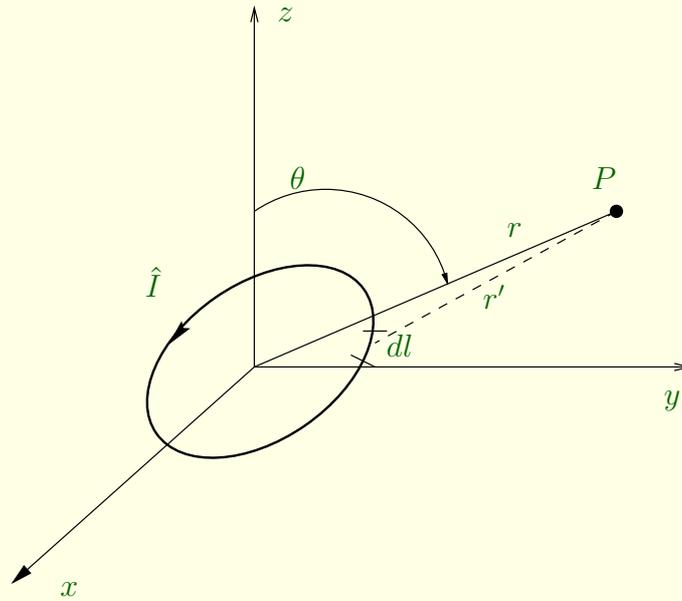


FIGURE 9.4 – Doublet magnétique.

où A est l'aire de la boucle de courant [14, page 509-511]. Il en résulte une puissance de rayonnement de

$$P_{rad} = \frac{1}{2} \oint_s \hat{S}_{av} \cdot ds \quad (9.43)$$

$$= 15,585 |I|^2 \left(\frac{A}{\lambda_0^2} \right)^2 \quad (9.44)$$

À nouveau, pour une boucle de courant de rayon égal à $1 [cm]$ et une fréquence de $300 [MHz]$ ($\lambda_0 = 1 [m]$), la résistance équivalente vaut $3,08 [m\Omega]$. Pour rayonner une puissance de $1 [W]$, il faut à cette fois un courant de $18 [A]$!

Antennes longues

La raison pour laquelle le doublet de HERTZ rayonne mal tient en partie à sa petite longueur. Le remède imaginé a consisté à allonger le dipôle.

Le principe de fonctionnement intuitif est celui d'une ligne parallèle dont on écarterait progressivement les conducteurs, tel que représenté à la figure 9.5.

Des ondes de courant et de tension, prises sinusoïdales, parcourent la ligne et contribuent à un champ électromagnétique dont la phase respective des grandeurs électriques d'excitation est fonction de la position du doublet infinitésimal le long de l'antenne. La figure 9.6 reprend la forme de l'onde de courant d'excitation le long de l'antenne.

L'interprétation telle qu'envisagée, à savoir l'ouverture d'une ligne parallèle et addition des champs rayonnés par une succession de segments infinitésimaux, conduit à un résultat analytiquement correct. Après calculs, le champ électrique à grande distance vaut

$$\hat{E}_\theta = \frac{60 \hat{I}_m e^{-j\beta_0 R} \cos(\pi l / \lambda_0 \cos \theta) - \cos(\pi l / \lambda_0)}{R \sin \theta} \vec{a}_\theta \quad (9.45)$$

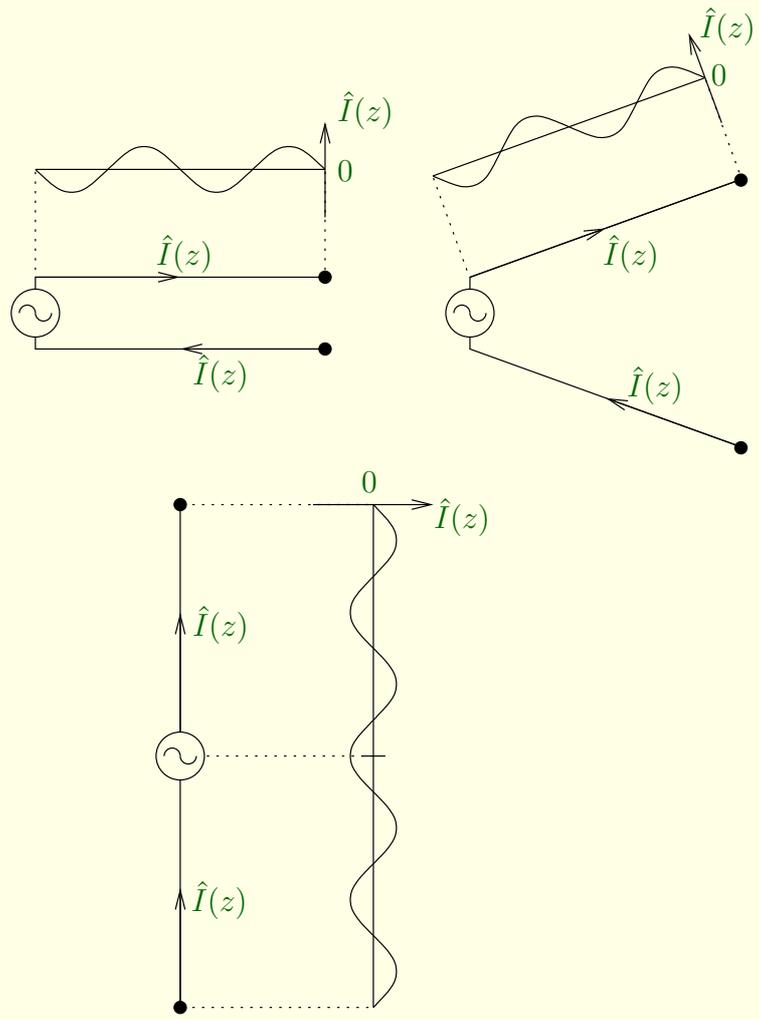


FIGURE 9.5 – Distributions de courant le long d’une antenne longue.

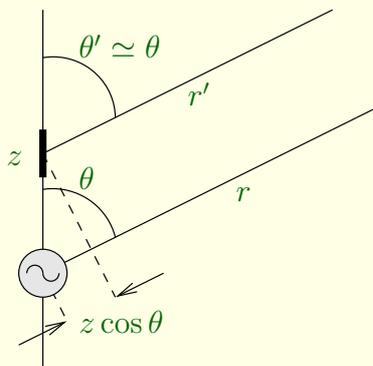
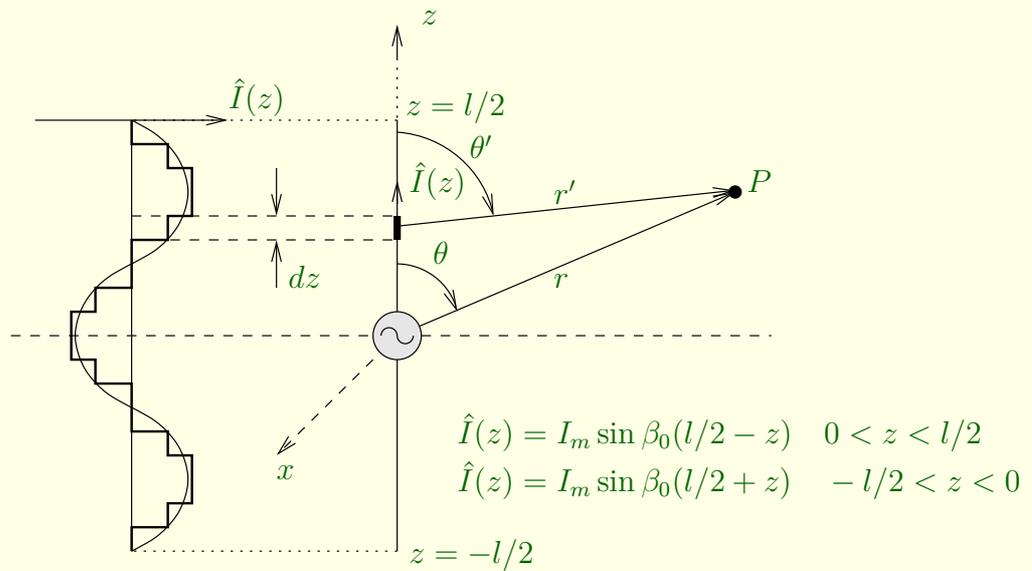


FIGURE 9.6 – Configuration géométrique pour le dipôle allongé.

On remarque que la forme du diagramme de rayonnement est fonction de la longueur d'onde alors que, précédemment, λ_0 n'intervenait qu'au niveau de l'amplitude. Le tableau 9.2 illustre ce phénomène pour une fréquence fixe mais pour une longueur d'antenne variable.

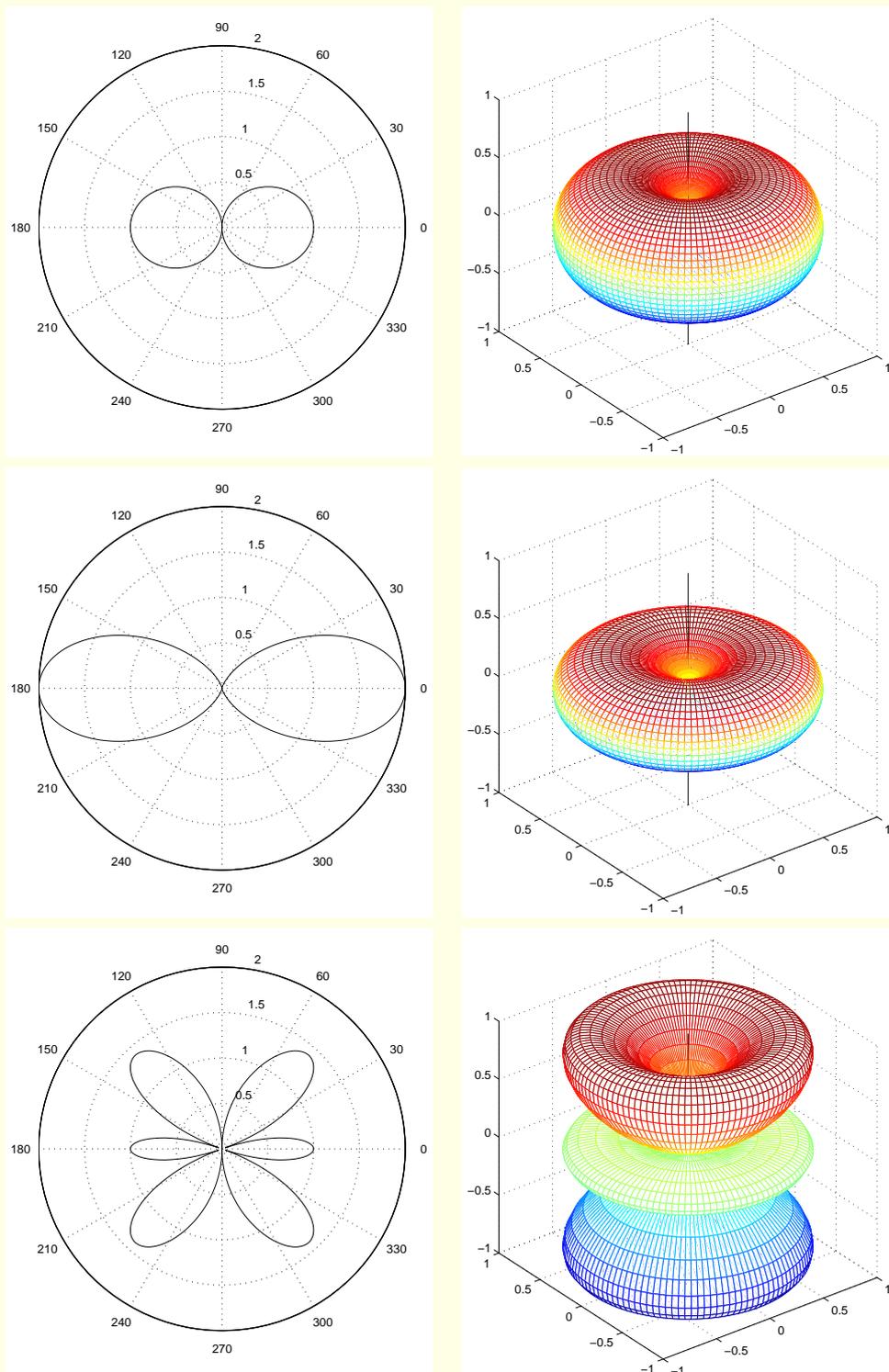


TABLE 9.2 – Diagrammes de rayonnement (du champ électrique) bidimensionnels et tridimensionnels pour le dipôle allongé avec, respectivement, $l = \frac{\lambda_0}{2}$, $l = \lambda_0$ et $l = \frac{3\lambda_0}{2}$.

L'antenne longue la plus utilisée est l'antenne demi longueur d'onde, $l = \frac{\lambda_0}{2}$, ce qui correspond

à un champ électrique

$$\hat{E}_\theta = \frac{60\hat{I}_m e^{-j\beta_0 R}}{R} \frac{\cos(\pi/2 \cos \theta)}{\sin \theta} \vec{a}_\theta \quad (9.46)$$

maximum dans le plan horizontal ($\theta = 90^\circ$).

En pratique, l'antenne demi longueur d'onde se réduit à une antenne quart d'onde au-dessus d'un plan de masse conducteur. Pour cette dernière, la résistance de rayonnement vaut

$$R_{rad} = 36,5 [\Omega] \quad (9.47)$$

Pour obtenir la puissance réellement dissipée par l'antenne, il faudra encore considérer la résistance interne de la force électromotrice d'alimentation et la résistance des conducteurs d'alimentation, car c'est à l'adaptation conjuguée que le transfert de puissance sera maximum.

Réseaux d'antennes

L'antenne longue seule permet d'assurer une directivité en θ . Par contre, le diagramme de rayonnement est isotrope dans le plan de la terre. Comme il est souvent nécessaire d'orienter le rayonnement dans une certaine direction, on recourt à des réseaux d'antennes, formés d'une série d'antenne longues reliées entre elles avec un courant d'alimentation déphasé. Ce déphasage (noté α) déterminé en fonction de la longueur inter-antennes et montré à la figure 9.7, est à l'origine de la modification de l'allure du diagramme de rayonnement dans le plan de terre.

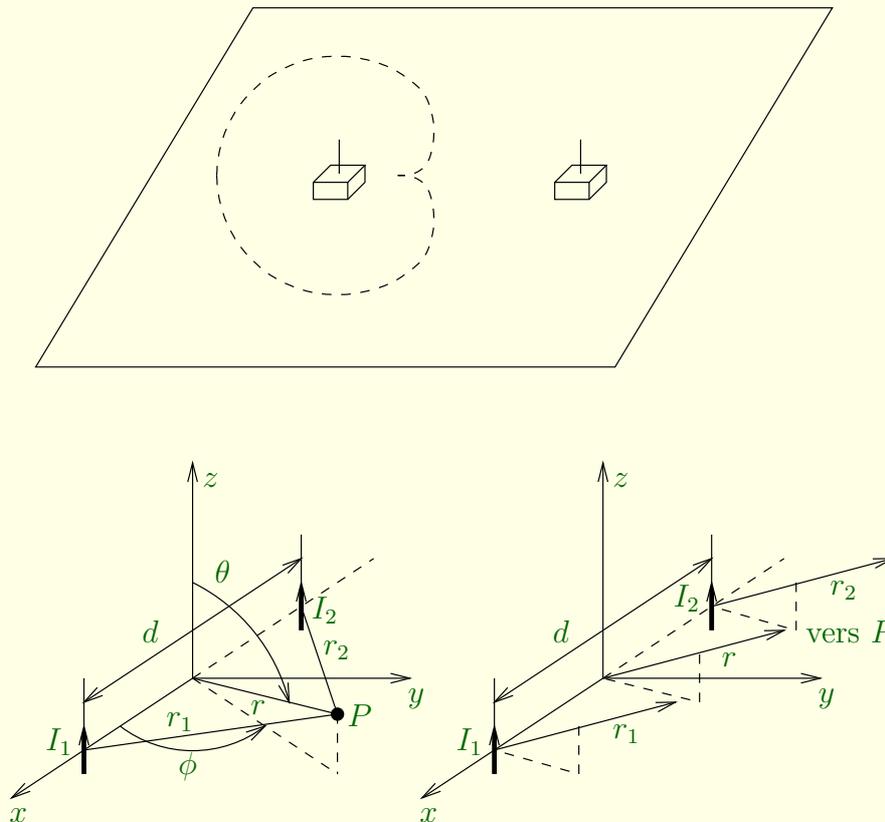


FIGURE 9.7 – Réseau d'antennes.

La figure 9.8 montre les diagrammes de rayonnement dans le plan de terre pour différents rapports λ_0/d .

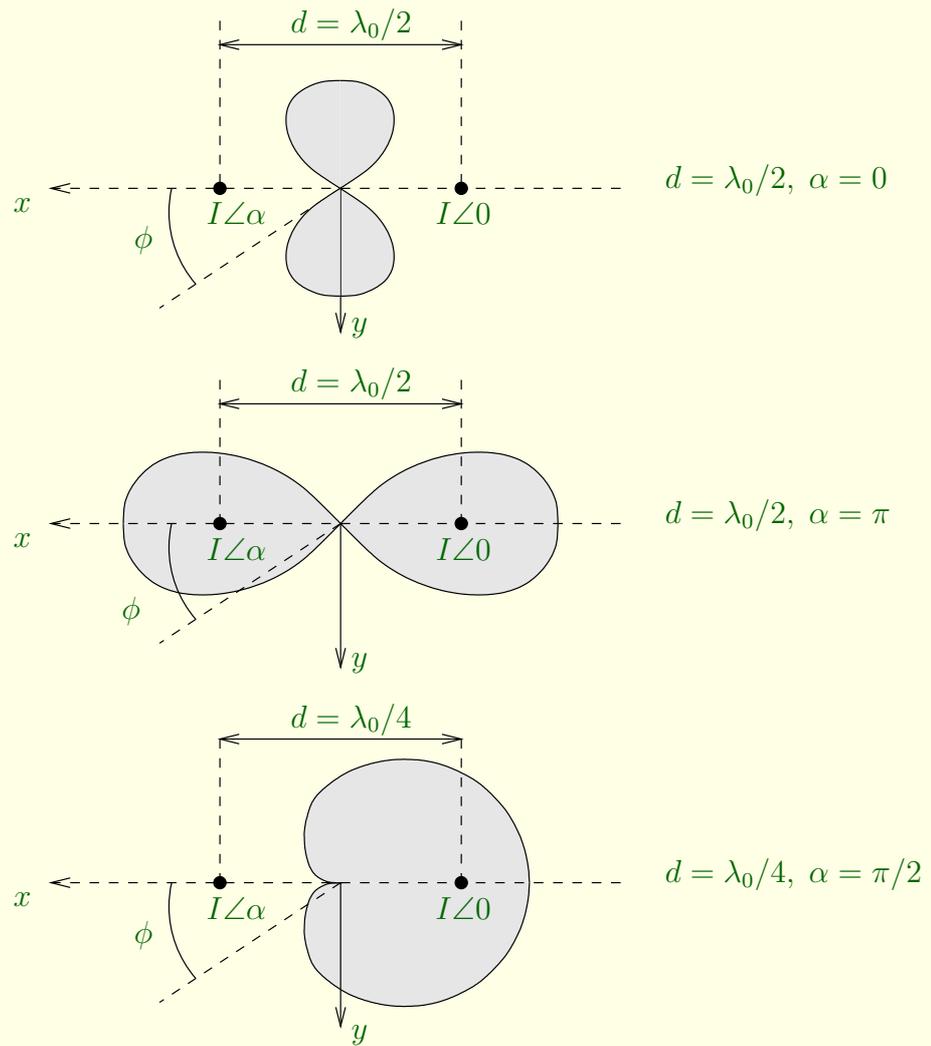


FIGURE 9.8 – Diagrammes de rayonnement.

9.3.5 Exemples d'antenne

Les figures 9.9 et 9.10 montrent plusieurs antennes dont des antennes paraboliques et des antennes GSM. Les antennes de la figure 9.11 sont d'un autre type. Elles ressemblent à une boucle de courant. D'autre part, seule la partie avant est raccordée; les structures métalliques situées à l'arrière servent de réflecteur de manière à offrir une directivité adéquate vers la mer.



FIGURE 9.9 – Exemple d'antennes paraboliques (ROCKHAMPTON, QUEENSLAND, AUSTRALIE).

9.4 Bilan de puissance

9.4.1 Gain

L'étude de la directivité d'une antenne telle qu'elle apparaît au travers de son diagramme de rayonnement est essentielle à la détermination d'un bilan de liaison entre antennes. En effet,



FIGURE 9.10 – Exemple d’antennes GSM et paraboliques (ROCKHAMPTON, QUEENSLAND, AUSTRALIE).



FIGURE 9.11 – Antennes d'aide à la navigation maritime (VILLERVILLE, FRANCE).

de toute l'onde rayonnée, seule une infime partie parviendra au droit de l'antenne de réception. Ce phénomène s'amplifie même à mesure que l'antenne de réception s'éloigne de l'antenne d'émission.

Prenons tout d'abord le cas d'une antenne isotrope; il s'agit d'une antenne dont le vecteur de POYNTING est identique dans toutes les directions. Soit P_E , la puissance totale rayonnée par l'antenne, l'isotropie fait en sorte que $|S|$ vaut, à une distance d'observation d ,

$$|S| = \frac{P_E}{4\pi d^2} \quad (9.48)$$

En espace libre, le champ électrique en découle immédiatement en raison de l'équation 9.36

$$|E| = \frac{\sqrt{30P_E}}{d} \quad (9.49)$$

Exemple. En radiodiffusion FM, on dispose typiquement d'une source de 100 [W]. À une distance de 10 [km], le champ électrique vaut alors 5,5 [mV/m], ce qui est au-delà du seuil d'écoute.

Si en radiodiffusion, il n'y a aucune raison d'assurer une très forte directivité, hormis celle dans le plan de terre, il en va autrement pour des liaisons point à point telles que les faisceaux hertziens ou les communications par satellites. La solution consiste à privilégier des directions particulières en modifiant la forme du diagramme de rayonnement.

Définition 84 L'ITU définit le gain d'une antenne comme le "ratio, généralement exprimé en décibels, entre la puissance nécessaire à l'entrée d'une antenne de référence sans perte à la puissance effective fournie à l'antenne considérée de manière à ce qu'elle fournisse le même champ électrique ou la même puissance dans une direction donnée".

En général, on prend une antenne de référence isotrope en espace libre. On parle alors de Puissance Isotrope Rayonnée Équivalente (PIRE).

Définition 85 La Puissance Isotrope Rayonnée Équivalente (PIRE) est le produit de la puissance d'émission d'une antenne par le gain dans la direction d'observation.

En effet, un récepteur ne fait pas la distinction entre la puissance d'une antenne P_E de gain G_E et la puissance fournie par une antenne isotrope de puissance $P_E G_E$.

La figure 9.12 compare une antenne isotrope et une antenne directive de gain G_E dans une direction donnée.

9.4.2 Aire effective

Un autre concept utile est celui d'aire effective de l'antenne; elle exprime la capacité d'une antenne à capturer une onde électromagnétique.

Définition 86 L'aire effective d'une antenne est définie comme le rapport entre la puissance disponible à ses bornes au vecteur de POYNTING incident (exprimé en valeur efficace⁴)

$$A_{eff} = \frac{P_R}{S_{eff}} \quad (9.50)$$

4. Nous avons donc $S_{eff} = \frac{|S|}{2}$.

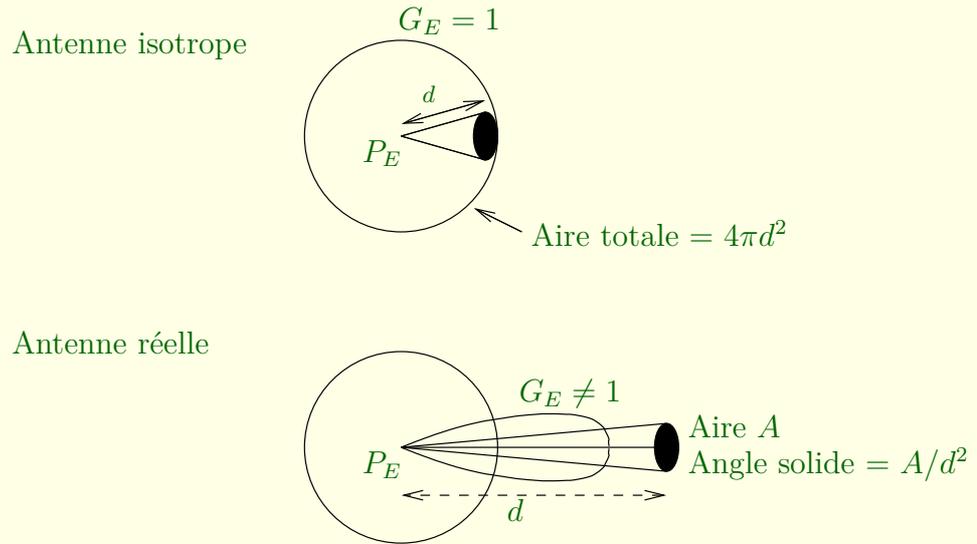


FIGURE 9.12 – Gain d’antenne.

Exemple : calcul de l’aire effective du dipôle de Hertz en réception. Prenons le cas d’une onde arrivant dans l’axe $\theta = 90^\circ$, ce qui correspond à un champ électrique maximum, parallèle au doublet. La densité de puissance de l’onde incidente vaut

$$S_{eff} = \frac{1}{2} \frac{|E|^2}{\eta_0} \quad (9.51)$$

La tension disponible aux bornes de l’antenne en circuit ouvert est

$$|V_{co}| = |E| dl \quad (9.52)$$

À l’adaptation conjuguée, la puissance disponible en découle

$$P_R = \frac{|V_{co}|^2}{8R_{rad}} \quad (9.53)$$

$$= \frac{|E|^2 dl^2}{8R_{rad}} \quad (9.54)$$

Or, nous avons vu que (cf. équation 9.40)

$$P_{rad} = 80\pi^2 \left(\frac{dl}{\lambda_0} \right)^2 \frac{|I|^2}{2} \quad (9.55)$$

dans le cas du doublet électrique, ce qui fournissait la valeur de R_{rad}

$$R_{rad} = 80\pi^2 \left(\frac{dl}{\lambda_0} \right)^2 \quad (9.56)$$

et finalement

$$P_R = \frac{|E|^2 \lambda_0^2}{640\pi^2} \quad (9.57)$$

L’aire effective se calcule sur base des résultats 9.51 et 9.57 :

$$A_{eff} = \frac{|E|^2 \lambda_0^2}{640\pi^2} \frac{2\eta_0}{|E|^2} = 1,5 \frac{\lambda_0^2}{4\pi} = G_E \frac{\lambda_0^2}{4\pi} \quad (9.58)$$

où G_E est le gain de l'antenne dans la direction considérée ($\theta = 90^\circ$).

La dernière expression est en fait tout à fait générale. Elle établit que la surface effective d'une antenne est liée à son gain par la relation

$$A_{eff} = G_E \frac{\lambda_0^2}{4\pi} \quad (9.59)$$

En pratique, on tente de lier la surface effective à la surface réelle de l'antenne $A_{eff} = \eta A$ pour une gamme de fréquence données; le coefficient η doit tenir compte de plusieurs paramètres tels que la finition de la surface, etc. Pour une parabole, η vaut typiquement 0,6.

9.4.3 Propagation en espace libre : équation de Friis

Dans ce paragraphe, on établira la formule de FRIIS donnant l'affaiblissement d'une liaison entre deux antennes séparées par une distance d en espace libre. Cette distance est supposée suffisamment grande pour permettre l'utilisation de l'expression à champ éloigné. De plus, les polarisations sont supposées correspondantes. Les gains des antennes sont notés G_E et G_R (cf. figure 9.13 pour un schéma de principe).

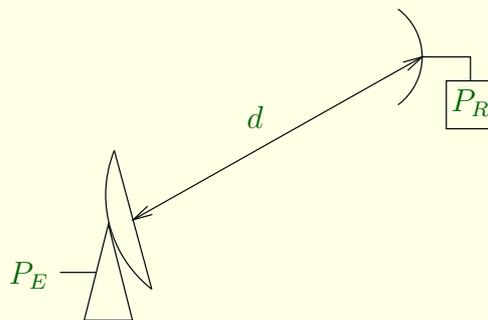


FIGURE 9.13 – Bilan de liaison.

La puissance reçue est le produit de l'aire effective de l'antenne par le vecteur de POYNTING incident

$$P_R = A_{eff} S_{eff} = \left(\frac{P_E G_E}{4\pi d^2} \right) A_{eff} = \left(\frac{P_E G_E}{4\pi d^2} \right) \left(\frac{\lambda^2}{4\pi} \right) G_R = P_E G_E G_R \left(\frac{\lambda}{4\pi d} \right)^2 \quad (9.60)$$

On obtient ainsi la formule de FRIIS qui fournit l'affaiblissement en espace libre

$$\epsilon = \frac{P_E}{P_R} = \left(\frac{4\pi d}{\lambda} \right)^2 \frac{1}{G_E G_R} \quad (9.61)$$

Le premier facteur de cette expression est appelé l'affaiblissement d'espace; le second exprime tout l'avantage d'antennes directives et montre que les deux antennes jouent à cet égard des rôles identiques. Sous une forme pratique, la formule peut s'écrire en unités logarithmiques –c'est la formule de FRIIS

$$\epsilon = 32,5 + 20 \log f_{[MHz]} + 20 \log d_{[km]} - G_{E[dB]} - G_{R[dB]} \quad (9.62)$$

On serait tenté de déduire de 9.61 et de 9.62 que les basses fréquences sont avantageuses : cela est vrai pour des antennes de gain donné. Mais il vaut mieux faire ce raisonnement en supposant données les dimensions de l'antenne. En-dessous d'une certaine fréquence, les dimensions de l'antenne sont inférieures à la longueur d'onde et il n'est pas possible d'obtenir de la directivité; les gains de l'antenne restent de l'ordre de 0 [dB], et le raisonnement précédent est correct. Au-dessus de cette fréquence, le rapport dimension/longueur d'onde croît proportionnellement à la fréquence et le gain augmente; la conclusion doit être inversée. En effet, on peut écrire l'affaiblissement fourni par 9.61 sous la forme

$$\epsilon = \frac{\lambda^2 d^2}{A_E A_R} = \frac{c^2 d^2}{f^2 A_E A_R} \quad (9.63)$$

L'affaiblissement décroît donc de 20 [dB] par décade, ce qui laisse pressentir l'intérêt des hyperfréquences.

9.5 Effets de propagation

Si l'équation de FRIIS a le mérite de conduire à la détermination de l'affaiblissement en espace libre et en l'absence de tout obstacle, la pratique se révèle tout autre. L'effet majeur dont il faut tenir compte à basse fréquence est la présence du sol; elle contribue à la présence d'une onde réfléchie dont les caractéristiques sont déterminées par la nature du sol. Le schéma de la figure 9.14 illustre le principe.

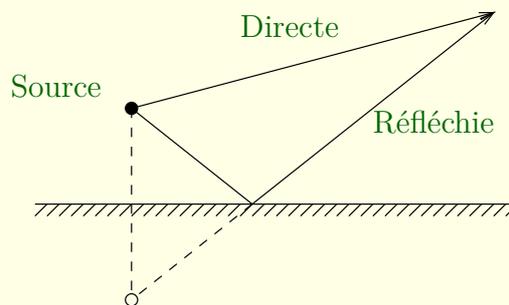


FIGURE 9.14 – Rayon direct et réfléchi.

L'onde reçue au niveau de l'antenne réceptrice se compose donc de l'onde de trajet direct et d'une onde réfléchie au sol. Considérer le sol comme un conducteur électrique parfait permettrait d'appliquer la méthode des images. Mais la terre est loin d'être un conducteur électrique parfait. En conséquence, une onde est réfractée vers l'intérieur de la terre et le rayon d'incidence diffère du rayon réfléchi. De plus, le champ électrique ne s'annule pas totalement à la surface de la terre, d'où le fait que le champ électrique parallèle à la terre ne subira pas une inversion totale de polarité.

Avec les conventions de notation de la figure 9.15, on montre que le potentiel vecteur se compose de trois termes

$$\hat{A} = \frac{\mu_0 \hat{I} dl}{4\pi} \left(\frac{e^{-j\beta_0 d_1}}{d_1} + K_v \frac{e^{-j\beta_0 d_2}}{d_2} + f(d_2, \theta_2) \frac{e^{-j\beta_0 d_2}}{d_2} \right) \quad (9.64)$$

Les deux premiers termes sont relatifs à une onde plane à polarisation verticale; le troisième est appelé onde de surface.

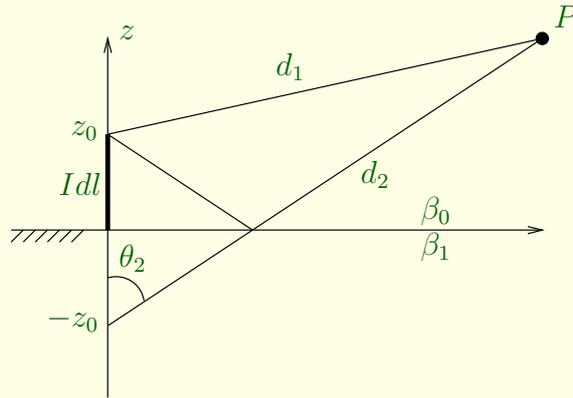


FIGURE 9.15 – Antenne linéaire au-dessus d'un plan conducteur.

Outre les effets du sol, il convient de considérer les effets suivants :

- Réflexion sur certaines couches de l'atmosphère, dont l'ionosphère. Le phénomène n'est pas négligeable à certaines fréquences. C'est précisément en jouant sur l'effet de la réflexion, telle que représentée à la figure 9.16, que l'on parvient à envoyer des signaux en l'absence de visibilité directe entre deux points à très grande distance.

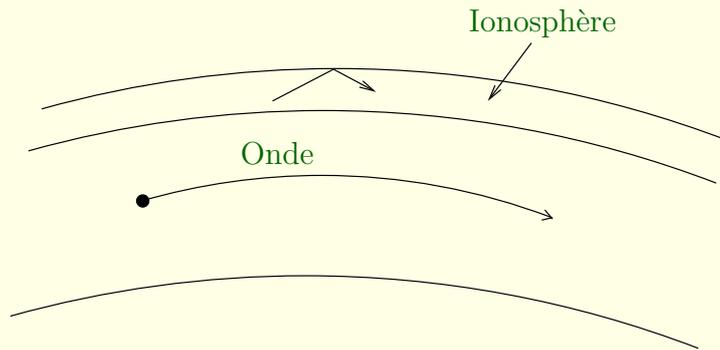


FIGURE 9.16 – Réflexion sur l'ionosphère.

- Diffusion due à la non-planéité des surfaces réfléchissantes. Ce phénomène est illustré à la figure 9.17.

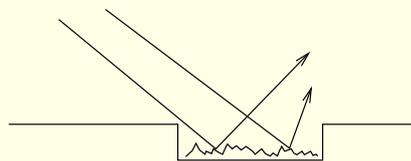


FIGURE 9.17 – Diffusion sur une surface irrégulière.

- Diffraction causée par des obstacles. Elle contribue à augmenter l'affaiblissement de propagation. La figure 9.18 montre un obstacle et le niveau de puissance d'illumination à l'arrière de l'obstacle.

Le phénomène de **diffraction** explique pourquoi il est malgré tout possible de capter un signal électromagnétique en l'absence de visibilité directe avec l'émetteur, notamment en radiodiffusion **FM**.

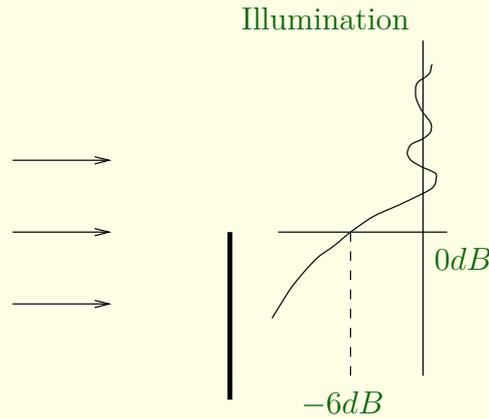


FIGURE 9.18 – Exemple de diffraction.

– Topographie du sol, la hauteur relative des antennes, les précipitations, ...

Les aspects physiques tels que présentés ci-avant peuvent être résumés très grossièrement en disant que l’onde de surface est prépondérante en dessous de $10 [MHz]$, qu’entre $10 [MHz]$ et $30 [MHz]$, la réflexion sur l’ionosphère est le phénomène majeur et qu’au-delà de $30 [MHz]$, l’onde n’est plus guidée mais qu’elle devient sensible aux réflexions multiples dues à toute sorte d’obstacles.

Il existe de très nombreux modèles déterminant le bilan de puissance pour les hautes fréquences. Ceux-ci sont exprimés en fonction de paramètres statistiques permettant de tenir compte de l’évolution de la température par exemple, ou de la présence d’obstacles en mouvement.

9.5.1 Polarisation

L’orientation du phaseur du champ électrique généré par une antenne est appelée **polarisation** ; elle est généralement linéaire (horizontale ou verticale).

La polarisation de l’antenne joue un rôle dans la qualité de transmission. Ainsi, supposons que l’on dispose de 2 antennes à polarisation verticale. Si une antenne est placée dans un sens et l’autre à 90^0 , le signal reçu sera nul. Si les antennes sont alignées, le signal reçu sera maximum (abstraction faite des problèmes de réflexion et de diffraction qui peuvent générer un signal en opposition de phase). Le problème est plus délicat pour des antennes à polarisation circulaire car de telles antennes produisent un champ électrique dont le phaseur tourne dans le sens horlogique ou anti-horlogique. Il faut alors utiliser des antennes pairées pour garantir une transmission de qualité suffisante.

En pratique, on utilise des antennes à ondes polarisées lorsque la transmission se fait en espace libre, c’-à-d. sans obstacle comme pour une liaison par faisceau hertzien, ou pour des propagations à grande distance. On peut aussi utiliser des antennes polarisées pour caractériser l’environnement de propagation et déterminer le taux de polarisation croisée, c’est-à-dire le résidu en puissance d’une polarisation sur l’autre polarisation. Complémentairement, il est possible de “jouer” sur la polarisation pour optimiser certains paramètres de propagation et affecter les effets de multitrajets et de diffraction.

9.5.2 Multitrajet

Pour comprendre comment la présence d'un obstacle réfléchissant affecte le bilan de liaison, considérons l'exemple très simple tel que dessiné à la figure 9.19.

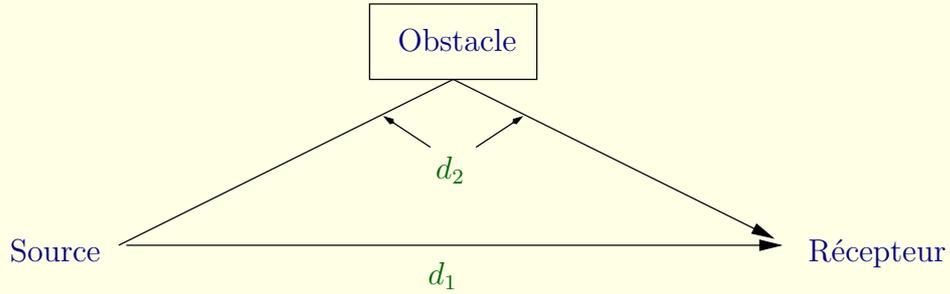


FIGURE 9.19 – Bilan de liaison en présence d'une réflexion.

On émet un signal $x(t)$. Le signal réceptionné est la somme de deux signaux

$$y(t) = a_1x(t - \tau_1) + a_2x(t - \tau_2) \quad (9.65)$$

La somme se compose d'un terme dû au trajet direct ; le second est le résultat de la réflexion. Supposons que l'atténuation de l'onde soit de type $e^{-\alpha d}$, avec d la distance parcourue, dans le milieu considéré. Le déphasage temporel est lui lié à la distance et à la vitesse v de propagation dans le milieu : $\tau = \frac{d}{v}$. En fréquentiel, le signal $y(t)$ s'écrit, en supposant une réflexion totale de l'onde,

$$\mathcal{Y}(f) = \mathcal{X}(f) \left[e^{-\alpha d_1} e^{-2\pi j f \frac{d_1}{v}} + e^{-\alpha d_2} e^{-2\pi j f \frac{d_2}{v}} \right] \quad (9.66)$$

$$= \mathcal{X}(f) \left[e^{-\alpha d_1} e^{-2\pi j f \frac{d_1}{v}} \right] \left[1 + e^{-\alpha(d_2-d_1)} e^{-2\pi j f \frac{d_2-d_1}{v}} \right] \quad (9.67)$$

$$= \mathcal{X}(f) \mathcal{H}(f) \mathcal{R}(f) \quad (9.68)$$

$\mathcal{H}(f)$ représente la fonction de transfert d'un canal idéal. Dans le cas d'une communication en espace libre, c'est l'équation de Friis qui fournit l'amplitude de ce terme. Le facteur multiplicatif $\mathcal{R}(f)$ est quant à lui dû à la réflexion. Son module vaut

$$\begin{aligned} \mathcal{R}(f) &= \sqrt{\left(1 + e^{-\alpha(d_2-d_1)} \cos\left(2\pi f \frac{d_2-d_1}{v}\right)\right)^2 + \left(e^{-\alpha(d_2-d_1)} \sin\left(2\pi f \frac{d_2-d_1}{v}\right)\right)^2} \\ &= \sqrt{1 + 2e^{-\alpha(d_2-d_1)} \cos\left(2\pi f \frac{d_2-d_1}{v}\right) + e^{-2\alpha(d_2-d_1)}} \end{aligned} \quad (9.69)$$

Dans certaines situations, la phase de l'onde réfléchiée concorde avec la phase de l'onde directe, auquel cas le facteur multiplicatif ajoute jusqu'à 6 [dB] au bilan de liaison. Pour certaines valeurs $d_2 - d_1$, l'opposition des phases des ondes réceptionnées se traduit par une perte significative de puissance. Le phénomène est illustré à la figure 9.20.

Les communications mobiles sont particulièrement sensibles au multitrajet. Ainsi, un récepteur qui s'éloigne de la source captera un signal dont le niveau de puissance pourrait être insuffisant à certains moments.

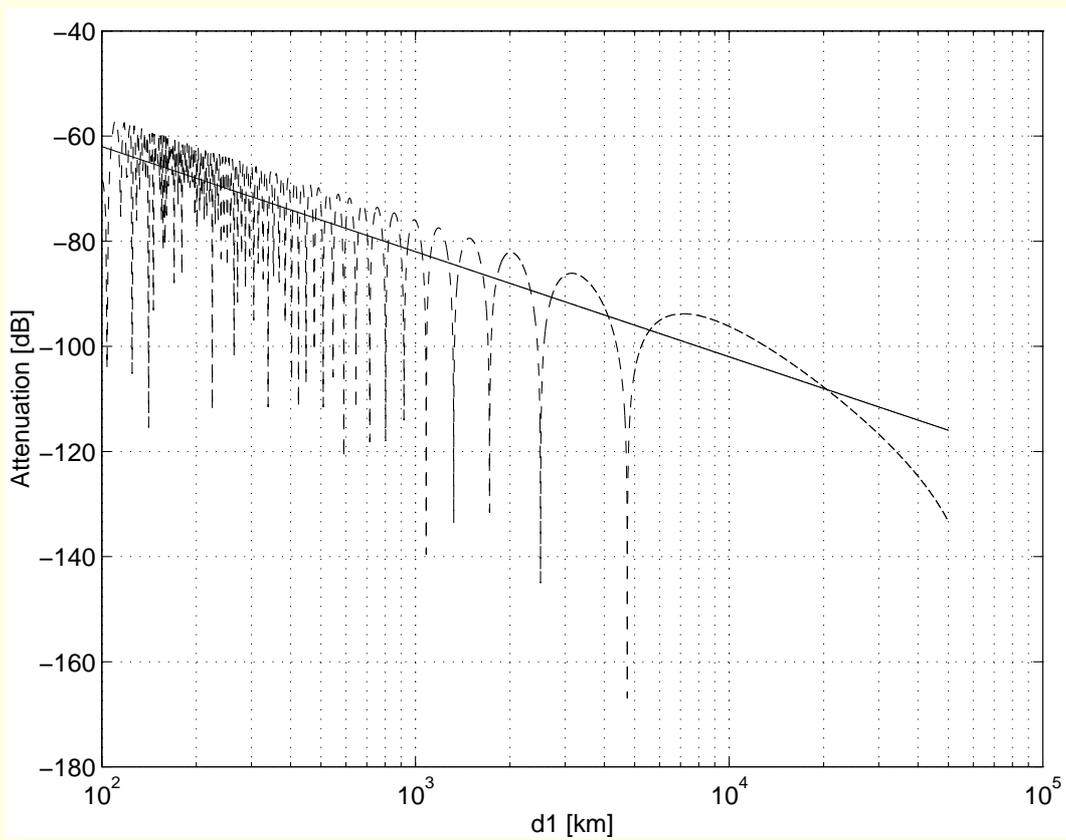


FIGURE 9.20 – Affaiblissement de puissance ϵ en fonction de la distance : (a) en espace libre (trait continu) et (b) en présence d’une réflexion (traits interrompus).

En pratique, on s'assure un confort d'écoute suffisant en garantissant une marge de puissance supplémentaire, appelée parfois marge de RAYLEIGH, valable pour un certain pourcentage de l'éloignement d_1 , en fonction de la distribution des évanouissements rencontrés.

9.6 Domaines d'application de la radio

9.6.1 Radiodiffusion sonore

La radiodiffusion sonore s'effectue soit sous forme analogique par modulation d'amplitude ou par fréquences, soit sous forme numérique selon la norme DAB (Digital Audio Broadcasting).

Modulation d'amplitude

La radiodiffusion sonore à modulation d'amplitude A3 se fait dans trois bandes de fréquences⁵ couramment appelées

- Grandes ondes : 150 à 285 [kHz]
- Ondes moyennes : 535 à 1605 [kHz]
- Ondes courtes : partie de la bande 2,3 à 16,1 [MHz]

La largeur d'un canal est de 9 [kHz], le signal modulé étant limité à 4,5 [kHz].

Modulation de fréquence

Elle se pratique en Europe dans la bande 87,5 à 108 [MHz]. Cette bande est divisée en canaux espacés de 300 [kHz]. La déviation de fréquence maximale autorisée est de 75 [kHz] et l'on applique une préaccentuation de 6 [dB] par octave au-dessus de 2,12 [kHz], correspondant à une constante de temps de 75 [μs].

Les émetteurs utilisent des antennes de type réseau vertical à rayonnement transversal destinés à obtenir un diagramme de rayonnement omnidirectionnel, mais à forte directivité dans le plan vertical. La puissance isotrope rayonnée équivalente (PIRE), produit de la puissance fournie à l'antenne par le gain de celle-ci, va de quelques dizaines de [W] à quelques [kW] selon le type de station.

Le niveau de champ requis en 50% des endroits pendant plus de 50% du temps varie de 50 [μV/m] à 3 [mV/m] selon le type de région et la qualité du service que l'on désire assurer. Les récepteurs ont en général une sensibilité de quelques [μV/m]; ils sont à une seule transposition de fréquence, la fréquence intermédiaire étant universellement prise égale à 10,7 [MHz].

À titre d'exemple de modulation en cascade et de spectre composite, voici comment on procède pour transmettre un signal stéréophonique en modulation FM. Cette transmission de deux signaux acoustiques G (canal de gauche) et D (canal de droite) doit rester compatible avec les récepteurs équipés pour la monophonie et ne pourrait guère en affecter la qualité de réception. La méthode choisie consiste à moduler la porteuse en fréquence par le spectre composite représenté à la figure 9.21.

5. Il s'agit des répartitions de fréquences pour l'Europe. Le règlement des radiocommunications publié par l'ITU reprend l'ensemble des attributions de fréquences.

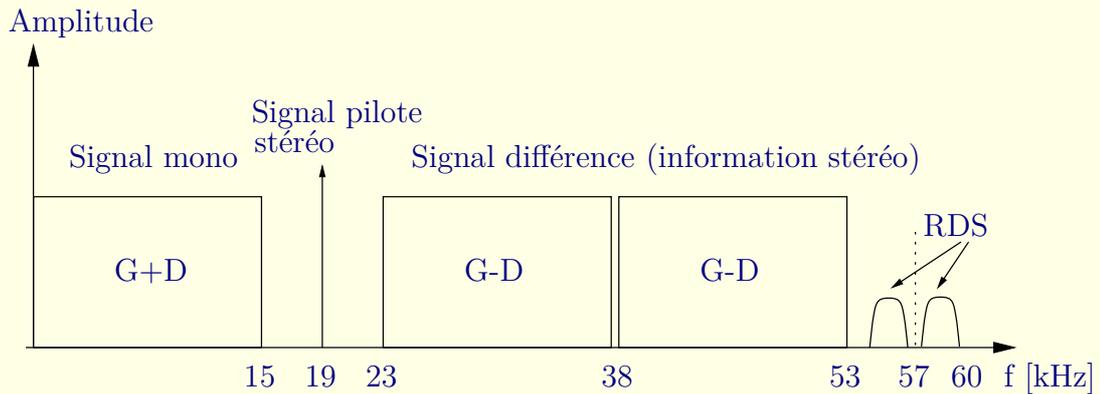


FIGURE 9.21 – Signal modulant en radiodiffusion stéréophonique.

Le signal principal restant dans la bande audible est la somme (G+D) ; la différence (G-D) est transposée dans la bande $23 - 53$ [kHz] par modulation d'amplitude à deux bandes latérales et porteuse supprimée. Au lieu de transmettre cette sous-porteuse de 38 [kHz], on transmet une sous-porteuse pilote de fréquence moitié, soit 19 [kHz], ayant ses passages à zéro en même temps que la sous-porteuse fictive. Comme la sous-porteuse et les deux bandes latérales sont inaudibles, le système est compatible avec les récepteurs monophoniques. Dans les récepteurs stéréophoniques, on doit, après le discriminateur, acquérir la sous-porteuse pilote à l'aide d'un filtre sélectif ou d'une boucle à verrouillage de phase, en doubler la fréquence, démoduler les sous-bandes latérales (G-D) et reconstituer les signaux G et D. La sous-porteuse pilote ne peut, si elle est appliquée seule, provoquer une déviation de fréquence supérieure à $7,5$ [kHz].

L'introduction de la diffusion de données en radio en modulation de fréquence par le système RDS (Radio Data System) a redynamisé la réception FM analogique. Le canal de données utilisé par le RDS permet au récepteur de trouver, entre autres, des fréquences alternatives lorsque l'autoradio sort du domaine de la station sur laquelle il est syntonisé, de recevoir des informations routières ou de synchroniser une montre.

Les données sont transmises par modulation d'amplitude à deux bandes latérales à porteuse supprimée. La sous-porteuse auxiliaire de 57 [kHz] est synchronisée en phase ou en phase inverse avec la troisième harmonique de la fréquence pilote. Les données sont transmises avec un débit de $1187,5$ [b/s] après mise en forme du signal en bande de base.

Chapitre 10

Principes de fonctionnement du réseau GSM

Note : ce chapitre a été écrit en partie par Cédric DEMOULIN.

10.1 Introduction

10.1.1 Historique

L'histoire de la téléphonie mobile (numérique) débute réellement en 1982. En effet, à cette date, le Groupe Spécial Mobile, appelé **GSM**¹, est créé par la Conférence Européenne des administrations des Postes et Télécommunications (CEPT) afin d'élaborer les normes de communications mobiles pour l'Europe dans la bande de fréquences de 890 à 915 [MHz] pour l'émission à partir des stations mobiles² et 935 à 960 [MHz] pour l'émission à partir de stations fixes. Il y eut bien des systèmes de mobilophonie analogique (MOB1 et MOB2, arrêté en 1999), mais le succès de ce réseau ne fut pas au rendez-vous.

Les années 80 voient le développement du numérique tant au niveau de la transmission qu'au niveau du traitement des signaux, avec pour dérivés des techniques de transmission fiables, grâce à un encodage particulier des signaux préalablement à l'envoi dans un canal, et l'obtention de débits de transmission raisonnables pour les signaux (par exemple 9,6 kilobits par seconde, noté [kb/s], pour un signal de parole).

Ainsi, en 1987, le groupe GSM fixe les choix technologiques relatifs à l'usage des télécommunications mobiles : transmission numérique, multiplexage temporel des canaux radio, chiffrement des informations ainsi qu'un nouveau codage de la parole. Il faut attendre 1991 pour que la première communication expérimentale par GSM ait lieu. Au passage, le sigle GSM change de signification et devient Global System for Mobile communications et les spécifications sont adaptées pour des systèmes fonctionnant dans la bande des 1800 [MHz].

En Belgique, c'est en 1994 que le premier réseau GSM (proximus) est déployé ; Mobistar et Orange (rebaptisé Base) viendront plus tard. Aujourd'hui, le nombre de numéros attribués

1. C'est en fait la première définition de l'acronyme GSM.

2. Pour des raisons d'économie de puissance et en vertu de la loi de Friis, on privilégie toujours les basses fréquences pour les émissions omnidirectionnelles.

pour des communications GSM dépasse largement le nombre de numéros dédiés à des lignes fixes et cette tendance se poursuit.

10.1.2 Évolution technologique

Tel quel, le réseau GSM est adéquat pour les communications téléphoniques de parole. En effet, il s'agit principalement d'un réseau commuté, à l'instar des lignes "fixes" et constitués de circuits, c'est-à-dire de ressources allouées pour la totalité de la durée de la conversation. Rien ne fut mis en place pour les services de transmission de données. Or, parallèlement au déploiement du GSM en Belgique, en 1994, la société Netscape allait donner un tour spectaculaire à un réseau de transmission de données, appelé Internet, en diffusant le premier logiciel de navigation grand public, articulé sur le protocole http et communément appelé web.

Comme le réseau GSM ne convenait guère pour la transmission de données, les évolutions récentes ont visé à accroître la capacité des réseaux en termes de débit mais à élargir les fonctionnalités en permettant par exemple l'établissement de communications ne nécessitant pas l'établissement préalable d'un circuit.

Pour dépasser la borne des $14,4 [kb/s]$, débit nominal d'un canal téléphonique basculé en mode de transmission de données, l'ETSI a défini un nouveau service de données en mode paquet : le General Packet Radio Service (**GPRS**) qui permet l'envoi de données à un débit de $115 [kb/s]$ par mise en commun de plusieurs canaux. D'une certaine manière, le GPRS prépare l'arrivée de la téléphonie de troisième génération, appelée Universal Mobile Telecommunications System (**UMTS**), qui permettra d'atteindre un débit de $2 [Mb/s]$. Mais le chemin est long car les applications nécessitant l'UMTS se font attendre, sans perdre de vue que tous les éléments du réseau UMTS sont incompatibles avec ceux du GSM. Pourquoi les investisseurs devraient-ils donc mettre la main au portefeuille ?

10.2 La technologie GSM

10.2.1 Le concept cellulaire

Les réseaux de première génération possédaient des cellules de grande taille ($50 [km]$ de rayon) au centre desquelles se situait une station de base (antenne d'émission). Au tout début, ce système allouait une bande de fréquences de manière statique à chaque utilisateur qui se trouvait dans la cellule qu'il en ait besoin ou non. Ce système ne permettait donc de fournir un service qu'à un nombre d'utilisateurs égal au nombre de bandes de fréquences disponibles. La première amélioration consista à allouer un canal à un utilisateur uniquement à partir du moment où celui-ci en avait besoin permettant ainsi d'augmenter "statistiquement" le nombre d'abonnés, étant entendu que tout le monde ne téléphone pas en même temps. Mais ce système nécessitait toujours des stations mobiles de puissance d'émission importante ($8 [W]$) et donc des appareils mobiles de taille et de poids conséquents. De plus, afin d'éviter les interférences, deux cellules adjacentes ne peuvent pas utiliser les mêmes fréquences. Cette organisation du réseau utilise donc le spectre fréquentiel d'une manière sous-optimale.

C'est pour résoudre ces différents problèmes qu'est apparu le concept de **cellule**. Le principe de ce système est de diviser le territoire en de petites zones, appelées cellules, et de partager les fréquences radio entre celles-ci. Ainsi, chaque cellule est constituée d'une station de base (reliée

au Réseau Téléphonique Commuté, **RTC**) à laquelle on associe un certain nombre de canaux de fréquences à bande étroite, sommairement nommés fréquences. Comme précédemment, ces fréquences ne peuvent pas être utilisées dans les cellules adjacentes afin d'éviter les interférences³. Ainsi, on définit des motifs, aussi appelés clusters, constitués de plusieurs cellules, dans lesquels chaque fréquence est utilisée une seule fois. La figure 10.1 montre un tel motif, en guise d'exemple.

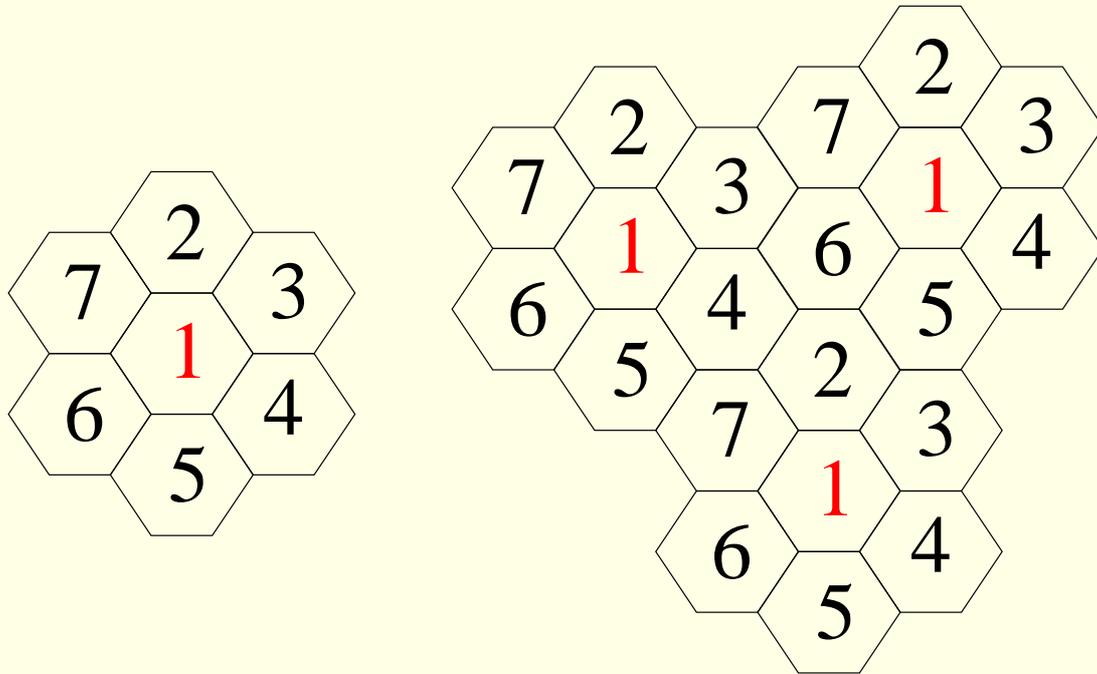


FIGURE 10.1 – Figure représentant un motif élémentaire (à gauche) et un ensemble de motifs dans un réseau (à droite).

Graphiquement, on représente une cellule par un hexagone car cette forme approche celle d'un cercle. Cependant, en fonction de la nature du terrain et des constructions, les cellules n'ont pas une forme circulaire. De plus, afin de permettre à un utilisateur passant d'une cellule à une autre de garder sa communication, il est nécessaire que les zones de couverture se recouvrent de 10 à 15%, ce qui renforce la contrainte de ne pas avoir une même bande de fréquences dans deux cellules voisines.

Pour éviter les interférences à plus grande distance entre cellules utilisant les mêmes fréquences, il est également possible d'asservir la puissance d'émission de la station de base en fonction de la distance qui la sépare de l'utilisateur. Le même processus du contrôle de la puissance d'émission est également appliqué en sens inverse. En effet, pour diminuer la consommation d'énergie des mobiles et ainsi augmenter leur autonomie, leur puissance d'émission est calculée en fonction de leur distance à la station de base. Grâce à des mesures permanentes entre un téléphone mobile et une station de base, les puissances d'émission sont régulées en permanence pour garantir une qualité adéquate pour une puissance minimale.

En résumé, une cellule se caractérise :

- par sa puissance d'émission nominale⁴ –ce qui se traduit par une zone de couverture à l'intérieur de laquelle le niveau du champ électrique est supérieur à un seuil déterminé–,

3. En pratique, une distance minimale de deux cellules sépare deux cellules utilisant la même fréquence.

4. Cette puissance est typiquement de l'ordre de la centaine de Watts.

- par la fréquence de porteuse utilisée pour l'émission radio-électrique et
- par le réseau auquel elle est interconnectée.

Il faut noter que la taille des cellules n'est pas la même sur tout le territoire. En effet, celle-ci dépend :

- du nombre d'utilisateurs potentiels dans la zone,
- de la configuration du terrain (relief géographique, présence d'immeubles, ...),
- de la nature des constructions (maisons, buildings, immeubles en béton, ...) et
- de la localisation (rurale, suburbaine ou urbaine) et donc de la densité des constructions.

Ainsi, dans une zone rurale où le nombre d'abonnés est faible et le terrain relativement plat, les cellules seront plus grandes qu'en ville où le nombre d'utilisateurs est très important sur une petite zone et où l'atténuation due aux bâtiments est forte. Un opérateur devra donc tenir compte des contraintes du relief topographique et des contraintes urbanistiques pour dimensionner les cellules de son réseau. On distingue pour cela quatre services principaux :

1. Le service "Outdoor" qui indique les conditions nécessaires pour le bon déroulement d'une communication en extérieur.
2. Le service "Incar" qui tient compte des utilisateurs se trouvant dans une voiture. On ajoute typiquement une marge supplémentaire de $6 [dB]$, dans le bilan de puissance pour en tenir compte.
3. Le service "Indoor" qui permet le bon déroulement des communications à l'intérieur des bâtiments. Cette catégorie de service se subdivise à son tour en deux :
 - (a) le "Soft Indoor" lorsque l'utilisateur se trouve juste derrière la façade d'un bâtiment et
 - (b) le "Deep Indoor" lorsqu'il se trouve plus à l'intérieur.

Typiquement, on considère que, lors de l'établissement du bilan de puissance, c'est-à-dire de l'analyse du rapport de la puissance émise à la puissance reçue au droit du récepteur, il faut tenir compte de $10 [dB]$ d'atténuation supplémentaire pour le Soft Indoor et de $20 [dB]$ pour Deep Indoor à $900 [MHz]$. Quand on sait que $10 [dB]$ représente un facteur de 10 en puissance, on comprend qu'il est crucial pour un opérateur de dimensionner au mieux son réseau, quitte à effectuer des mesures sur le terrain.

Réutilisation des ressources

Par rapport au système de première génération, les cellules étant de taille plus petite, la puissance d'émission est plus faible et le nombre d'utilisateurs peut être augmenté pour une même zone géographique. C'est grâce au principe de ré-utilisation des fréquences qu'un opérateur peut augmenter la capacité de son réseau. En effet, il lui suffit de découper une cellule en plusieurs cellules plus petites et de gérer son plan de fréquences pour éviter toute interférence. Il y a ainsi toute une nomenclature spécifique pour classer les cellules en fonction de leur taille (macro, micro, pico, etc).

Définition 87 [Capacité] *La capacité est le trafic maximum que peut écouler une cellule en fonction du nombre de fréquences qui lui sont attribuées, le trafic étant fonction du nombre moyen de personnes qui communiquent et de la durée moyenne d'une communication.*

Estimation du rapport de puissance porteuse à bruit

Étant donné que, dans un réseau, une même fréquence est réutilisée plusieurs fois, il est nécessaire d'évaluer la distance minimum qui doit séparer deux cellules utilisant la même fréquence pour qu'aucun phénomène perturbateur n'intervienne. En calculant le rapport entre la puissance de la porteuse et celle du bruit, il est possible d'estimer cette distance.

Pratiquement, dans une cellule, un mobile reçoit à la fois le message utile (dont la puissance vaut C) qui lui est destiné et un certain nombre de signaux perturbateurs. La connaissance du rapport entre ces puissances, nous permettra de connaître la qualité de la communication.

Pour commencer, il est nécessaire d'identifier les différents signaux perturbateurs. On peut les subdiviser en deux classes :

1. Les interférences de puissance totale I qui sont dues aux signaux émis par les autres stations. On peut distinguer :
 - (a) Les interférences co-canal qui sont dues aux signaux émis par les autres stations de base utilisant la même fréquence.
 - (b) Les interférences de canaux adjacents dues aux signaux émis par les stations de base utilisant des fréquences voisines.
2. Le bruit, de puissance N , provenant principalement du bruit de fond du récepteur.

Dès lors, c'est le rapport

$$\frac{C}{N + I} \quad (10.1)$$

qui permet d'évaluer la qualité de la communication ainsi que la distance de réutilisation des fréquences.

10.2.2 Synthèse des principales caractéristiques du GSM

La norme **GSM** prévoit que la téléphonie mobile par GSM occupe deux bandes de fréquences aux alentours des $900 [MHz]$:

1. la bande de fréquence $890 - 915 [MHz]$ pour les communications montantes (du mobile vers la station de base) et
2. la bande de fréquence $935 - 960 [MHz]$ pour les communications descendantes (de la station de base vers le mobile).

Comme chaque canal fréquentiel utilisé pour une communication a une largeur de bande de $200 [kHz]$, cela laisse la place pour 124 canaux fréquentiels à répartir entre les différents opérateurs. Mais, le nombre d'utilisateurs augmentant, il s'est avéré nécessaire d'attribuer une bande supplémentaire aux alentours des $1800 [MHz]$. On a donc porté la technologie GSM $900 [MHz]$ vers une bande ouverte à plus haute fréquence. C'est le système **DCS-1800** (Digital Communication System) dont les caractéristiques sont quasi identiques au GSM en termes de protocoles et de service. Les communications montantes se faisant alors entre 1710 et $1785 [MHz]$ et les communications descendantes entre 1805 et $1880 [MHz]$.

Connaissant les différents canaux disponibles, il est alors possible d'effectuer un multiplexage fréquentiel, appelé Frequency Division Multiple Access (**FDMA**), en attribuant un certain nombre de fréquences porteuses par station de base. Un opérateur ne dédie pas pour autant une bande de fréquences par utilisateur, car cela conduirait à un gaspillage de ressources radio étant

donné qu'un utilisateur émet par intermittence. De plus, avec un tel système, si une source parasite émet un bruit à une fréquence bien déterminée, le signal qui se trouve dans la bande de fréquence contenant le parasite sera perturbé. Pour résoudre ces problèmes, on combine le multiplexage en fréquence à un multiplexage temporel (appelé Time Division Multiple Access ou **TDMA**) consistant à diviser chaque canal de communication en trames de 8 intervalles de temps (dans le cas du GSM). Pour être complet, signalons qu'il existe encore une autre technique de multiplexage appelé Code Division Multiple Access (**CDMA**), utilisée dans la norme américaine **IS-95** ou promue pour l'**UMTS**.

Ainsi, avec le TDMA, il est par exemple possible de faire parler huit utilisateurs l'un après l'autre dans le même canal. On multiplie donc le nombre de canaux disponibles par unité de temps par huit.

Le tableau 10.1 montre les caractéristiques des réseaux à technologie GSM et il compare les normes.

	GSM	DCS-1800
Bande de fréquences (\uparrow)	890, 2 – 915 [MHz]	1710 – 1785 [MHz]
Bande de fréquences (\downarrow)	935, 2 – 960 [MHz]	1805 – 1880 [MHz]
Nombre d'intervalles de temps par trame TDMA	8	8
Débit total par canal	271 [kb/s]	271 [kb/s]
Débit de la parole	13 [kb/s]	13 [kb/s]
Débit maximal de données	12 [kb/s]	12 [kb/s]
Technique de multiplexage	Multiplexage fréquentiel et temporel	Multiplexage fréquentiel et temporel
Rayon de cellules	0, 3 à 30 [km]	0, 1 à 4 [km]
Puissance des terminaux	2 à 8 [W]	0, 25 et 1 [W]
Sensibilité des terminaux	-102 [dB]	
Sensibilité de la station de base	-104 [dB]	

TABLE 10.1 – Comparaison des systèmes GSM et DCS-1800.

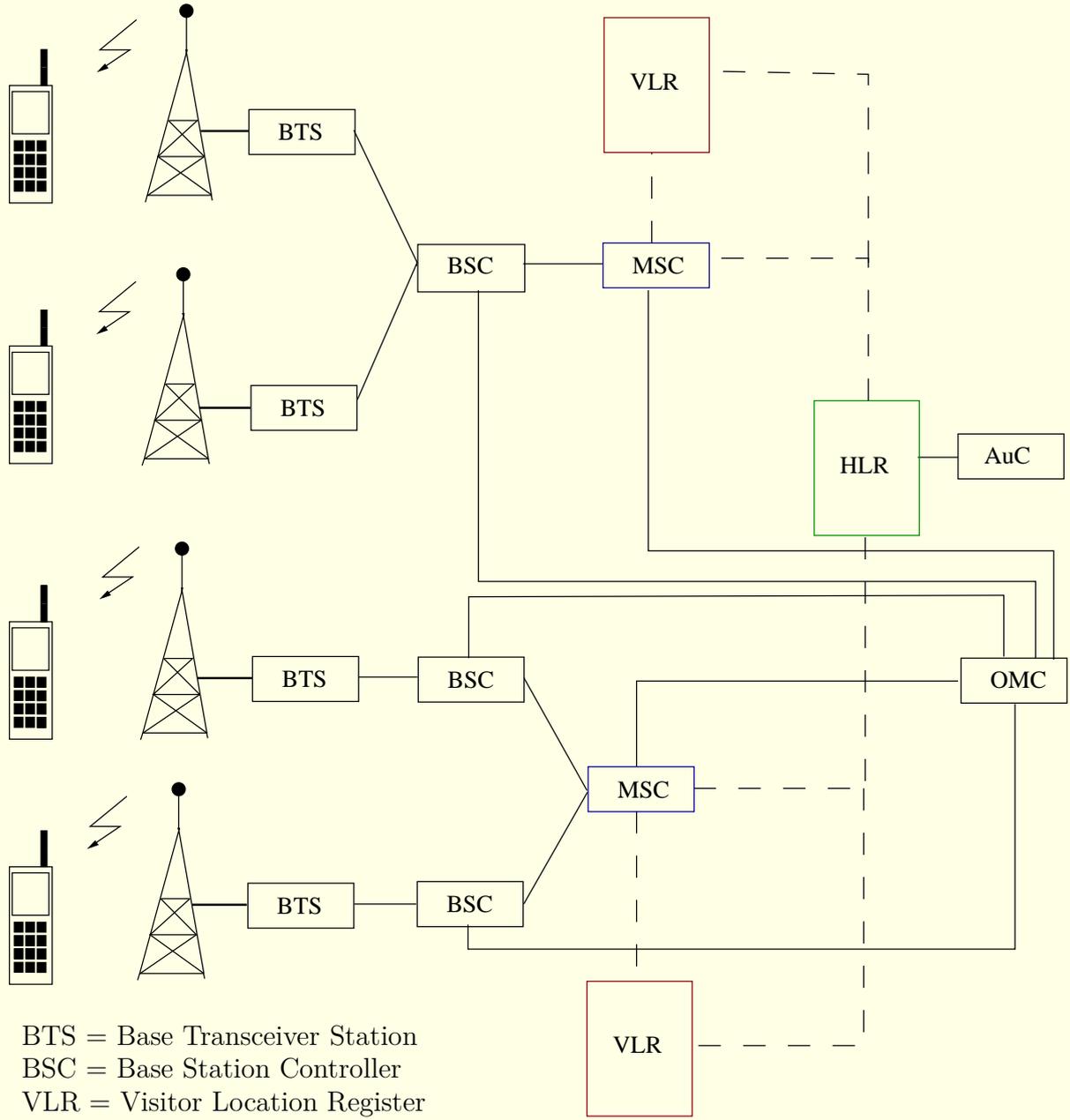
Tous les terminaux mobiles fabriqués actuellement sont compatibles avec les 2 normes ; ces terminaux sont appelés bi-bandes ou dual-band. Sur le territoire des États-Unis, aucune des bandes de fréquences pré-citées n'étaient encore disponibles. C'est pourquoi le réseau à technologie GSM américain utilise des bandes autour des 1900 [MHz]. Des terminaux capables d'opérer dans les trois bandes sont appelés tri-bandes.

10.3 Architecture du réseau

L'architecture d'un réseau GSM peut être divisée en trois sous-systèmes :

1. Le sous-système radio contenant la station mobile, la station de base et son contrôleur.
2. Le sous-système réseau ou d'acheminement.
3. Le sous-système opérationnel ou d'exploitation et de maintenance.

Les éléments de l'architecture d'un réseau GSM sont repris sur le schéma de la figure 10.2.



BTS = Base Transceiver Station
 BSC = Base Station Controller
 VLR = Visitor Location Register
 MSC = Mobile Switching Center
 HLR = Home Location Register
 AuC = Authentication Center
 OMC = Operation and Maintenance Center

FIGURE 10.2 – Architecture du réseau GSM.

10.3.1 Le sous-système radio

Le sous-système radio gère la transmission radio. Il est constitué de plusieurs entités dont le mobile, la station de base (**BTS**, Base Transceiver Station) et un contrôleur de station de base (**BSC**, Base Station Controller).

Le mobile

Le téléphone et la carte **SIM** (Subscriber Identity Module) sont les deux seuls éléments auxquels un utilisateur a directement accès. Ces deux éléments suffisent à réaliser l'ensemble des fonctionnalités nécessaires à la transmission et à la gestion des déplacements.

La principale fonction de la carte SIM est de contenir et de gérer une série d'informations. Elle se comporte donc comme une mini-base de données dont les principaux champs sont fournis dans le tableau 10.2.

Paramètres	Commentaires
Données administratives	
PIN /PIN2	Mot de passe demandé à chaque connexion
PUK /PUK2	Code pour débloquer une carte
Language	Langue choisie par l'utilisateur
Données liées à la sécurité	
Clé K_i	Valeur unique, connue de la seule carte SIM et du HLR
CKSN	Séquence de chiffrement
Données relatives à l'utilisateur	
IMSI	Numéro international de l'abonné
MSISDN	Numéro d'appel d'un téléphone GSM
Données de "roaming"	
TMSI	Numéro attribué temporairement par le réseau à un abonné
Location updating status	Indique si une mise à jour de la localisation est nécessaire
Données relatives au réseau	
Mobile Country Code (MCC), Mobile Network Code (MNC), etc	Identifiants du réseau mobile de l'abonné
Numéros de fréquence absolus	Fréquences utilisées par le PLMN

TABLE 10.2 – Liste partielle des informations contenues dans une carte SIM (voir [11, page 13] pour la liste complète).

L'identification d'un mobile s'effectue exclusivement au moyen de la carte SIM. En effet, elle contient des données spécifiques comme le code **PIN** (Personal Identification Number) et d'autres caractéristiques de l'abonné, de l'environnement radio et de l'environnement de l'utilisateur.

L'identification d'un utilisateur est réalisée par un numéro unique (**IMSI**, International Mobile Subscriber Identity) différent du numéro de téléphone connu de l'utilisateur (**MSISDN**, Mobile Station ISDN Number), tous deux étant incrustés dans la carte SIM.

La station de base (BTS)

La station de base est l'élément central, que l'on pourrait définir comme un ensemble émetteur/récepteur pilotant une ou plusieurs cellules. Dans le réseau GSM, chaque cellule principale au centre de laquelle se situe une station base peut-être divisée, grâce à des antennes directionnelles, en plus petites cellules qui sont des portions de celle de départ et qui utilisent des fréquences porteuses différentes. En Belgique, il est fréquent d'avoir des antennes tri-sectorielles, qui couvrent un peu plus de 120 degrés. Ces antennes ont l'allure de paires de segments verticaux, disposées en triangle. On trouve aussi des antennes à l'intérieur des bâtiments (voir figure 10.3).



FIGURE 10.3 – Antenne GSM (station de métro Rogier, Bruxelles).

C'est la station de base qui fait le relais entre le mobile et le sous-système réseau. Comme le multiplexage temporel est limité à 8 intervalles de temps, une station de base peut gérer tout au plus huit connections simultanées par cellule. Elle réalise les fonctions de la couche physique et de la couche liaison de données.

En cas de besoin, on peut exploiter une station de base localement ou par télécommande à travers son contrôleur de station de base.

Le contrôleur de station de base (BSC)

Le contrôleur de station de base gère une ou plusieurs stations de base et communique avec elles par le biais de l'interface A-bis. Ce contrôleur remplit différentes fonctions tant au niveau de la communication qu'au niveau de l'exploitation.

Pour les fonctions des communications des signaux en provenance des stations de base, le BSC agit comme un **concentrateur** puisqu'il transfère les communications provenant des différentes

stations de base vers une sortie unique. Dans l'autre sens, le contrôleur commute les données en les dirigeant vers la bonne station de base.

Dans le même temps, le BSC remplit le rôle de relais pour les différents signaux d'alarme destinés au centre d'exploitation et de maintenance. Il alimente aussi la base de données des stations de base. Enfin, une dernière fonctionnalité importante est la gestion des ressources radio pour la zone couverte par les différentes stations de base qui y sont connectées. En effet, le contrôleur gère les transferts inter-cellulaires des utilisateurs dans sa zone de couverture, c'est-à-dire quand une station mobile passe d'une cellule dans une autre. Il doit alors communiquer avec la station de base qui va prendre en charge l'abonné et lui communiquer les informations nécessaires tout en avertissant la base de données locale **VLR** (Visitor Location Register) de la nouvelle localisation de l'abonné.

C'est donc un maillon très important de la chaîne de communication et il est, de plus, le seul équipement de ce sous système à être directement gérable (via l'interface **X25** qui le relie au sous-système d'exploitation et de maintenance).

10.3.2 Le sous-système réseau

Le sous-système réseau, appelé Network Switching Center (**NSS**), joue un rôle essentiel dans un réseau mobile. Alors que le sous-réseau radio gère l'accès radio, les éléments du NSS prennent en charge toutes les fonctions de contrôle et d'analyse d'informations contenues dans des bases de données nécessaires à l'établissement de connexions utilisant une ou plusieurs des fonctions suivantes : **chiffrement**, **authentification** ou **roaming**.

Le NSS est constitué de :

- Mobile Switching Center (**MSC**)
- Home Location Register (**HLR**) / Authentication Center (**AuC**)
- Visitor Location Register (**VLR**)
- Equipment Identity Register (**EIR**)

Le centre de commutation mobile (MSC)

Le centre de commutation mobile est relié au sous-système radio via l'interface A. Son rôle principal est d'assurer la commutation entre les abonnés du réseau mobile et ceux du réseau commuté public (RTC) ou de son équivalent numérique, le réseau **RNIS** (**ISDN** en anglais). D'un point de vue fonctionnel, il est semblable à un commutateur de réseau ISDN, mis à part quelques modifications nécessaires pour un réseau mobile.

De plus, il participe à la fourniture des différents services aux abonnés tels que la téléphonie, les services supplémentaires et les services de messagerie. Il permet encore de mettre à jour les différentes bases de données (HLR et VLR) qui donnent toutes les informations concernant les abonnés et leur localisation dans le réseau.

Les commutateurs MSC d'un opérateur sont reliés entre eux pour la commutation interne des informations. Des MSC servant de passerelle (Gateway Mobile Switching Center, **GMSC**) sont placées en périphérie du réseau d'un opérateur de manière à assurer une inter-opérabilité entre réseaux d'opérateurs.

L'enregistreur de localisation nominale (HLR)

Il existe au moins un enregistreur de localisation (HLR) par réseau (PLMN). Il s'agit d'une base de données avec des informations essentielles pour les services de téléphonie mobile et avec un accès rapide de manière à garantir un temps d'établissement de connexion aussi court que possible.

Le HLR contient :

- toutes les informations relatives aux abonnés : le type d'abonnement, la clé d'authentification K_i – cette clé est connue d'un seul HLR et d'une seule carte SIM –, les services souscrits, le numéro de l'abonné (IMSI), etc
- ainsi qu'un certain nombre de données dynamiques telles que la position de l'abonné dans le réseau – en fait, son VLR – et l'état de son terminal (allumé, éteint, en communication, libre, ...).

Les données dynamiques sont mises à jour par le MSC. Cette base de données est souvent unique pour un réseau GSM et seules quelques personnes y ont accès directement.

Le centre d'authentification (AuC). Lorsqu'un abonné passe une communication, l'opérateur doit pouvoir s'assurer qu'il ne s'agit pas d'un usurpateur. Le centre d'authentification remplit cette fonction de protection des communications. Pour ce faire, les normes GSM prévoient deux mécanismes :

1. Le chiffrement des transmissions radio. Remarquons qu'il s'agit d'un chiffrement faible, qui ne résiste pas longtemps à la crypto-analyse ! Ceci explique sans doute pourquoi, en Belgique, de nombreux toits de bâtiments de puissance étrangère sont équipés d'antennes servant exclusivement à la réception de signaux GSM...
2. L'authentification des utilisateurs du réseau au moyen d'une clé K_i , qui est à la fois présente dans la station mobile et dans le centre d'authentification.

L'authentification s'effectue par résolution d'un défi sur base d'un nombre M généré aléatoirement et envoyé au mobile. À partir de ce nombre, un algorithme identique (algorithme A_3) qui se trouve à la fois dans la carte SIM et dans l'AuC produit un résultat sur base de la clé K_i et du nombre M .

Dès lors, lorsqu'un VLR obtient l'identifiant d'un abonné, il demande, au HLR du réseau de l'abonné, le nombre M servant au défi et le résultat du calcul afin de le comparer à celui qui sera produit et envoyé par le mobile. Si les résultats concordent, l'utilisateur est reconnu et accepté par le réseau.

Grâce à ce mécanisme d'authentification, un VLR peut accueillir un mobile appartenant à un autre réseau (moyennant un accord préalable entre opérateurs de réseau !) sans qu'il ne soit nécessaire de divulguer la clé de chiffrement du mobile.

On peut dès lors distinguer trois niveaux de protection :

1. La carte SIM qui interdit à un utilisateur non enregistré d'avoir accès au réseau.
2. Le chiffrement des communications destiné à empêcher l'écoute de celles-ci.
3. La protection de l'identité de l'abonné.

L'enregistreur de localisation des visiteurs (VLR)

Cette base de données ne contient que des informations dynamiques et est liée à un MSC. Il y en a donc plusieurs dans un réseau GSM. Elle contient des données dynamiques qui lui sont

transmises par le HLR avec lequel elle communique lorsqu'un abonné entre dans la zone de couverture du centre de commutation mobile auquel elle est rattaché. Lorsque l'abonné quitte cette zone de couverture, ses données sont transmises à un autre VLR ; les données suivent l'abonné en quelque sorte.

L'enregistreur des identités des équipements (EIR)

Malgré les mécanismes introduits pour sécuriser l'accès au réseau et le contenu des communications, le téléphone mobile doit potentiellement pouvoir accueillir n'importe quelle carte SIM de n'importe quel réseau. Il est donc imaginable qu'un terminal puisse être utilisé par un voleur sans qu'il ne puisse être repéré.

Pour combattre ce risque, chaque terminal reçoit un identifiant unique (International Mobile station Equipment Identity, **IMEI**) qui ne peut pas être modifié sans altérer le terminal. En fonction de données au sujet d'un terminal, un opérateur peut décider de refuser l'accès au réseau. Tous les opérateurs n'implémentent pas une telle base de données.

10.3.3 Le centre d'exploitation et de maintenance

Cette partie du réseau regroupe trois activités principales de gestion : la gestion administrative, la gestion commerciale et la gestion technique.

Le réseau de maintenance technique s'intéresse au fonctionnement des éléments du réseau. Il gère notamment les alarmes, les pannes, la sécurité, ... Ce réseau s'appuie sur un réseau de transfert de données, totalement dissocié du réseau de communication GSM.

10.3.4 L'acheminement des appels entre le réseau fixe et le réseau GSM : deux scénarios en guise d'exemple

Illustrons brièvement le fonctionnement des entités d'un réseau en traitant deux scénarios typiques simplifiés entre un réseau mobile et un réseau fixe :

1. Un abonné GSM compose le numéro de téléphone d'un abonné du réseau fixe.
Sa demande arrive d'abord au BTS de la cellule puis passe à travers le BSC et arrive enfin au MSC qui vérifie les droits de l'abonné (autorisation d'accéder à ce service, état de l'abonnement, ...). Si l'abonné remplit les conditions, le MSC transmet l'appel au réseau public et demande au BSC de réserver un canal pour la communication. Il ne reste alors plus qu'à attendre que le poste fixe soit décroché pour que la communication soit établie.
2. Supposons au contraire qu'un abonné du réseau fixe veuille joindre un abonné du réseau GSM.

Le fonctionnement est un peu plus complexe car l'opérateur GSM n'alloue des ressources à un abonné que lorsque celui-ci reçoit ou émet un appel.

Le numéro composé sur le poste fixe est tout d'abord aiguillé vers le réseau de l'abonné GSM. La demande de connexion est interprétée par un commutateur passerelle entrant du réseau GSM –il s'agit d'un GMSC. Le numéro formé par l'abonné du réseau fixe n'est pas utilisé tel quel pour commuter la communication. À l'instar des numéros verts ou des numéros d'urgence, il y a un mécanisme qui, au droit du GMSC, va convertir

le numéro de l'abonné en un autre numéro attribué dynamiquement en fonction de la position de l'utilisateur. C'est sur base de ce numéro dynamique que l'appel sera re-dirigé dans le réseau GSM. Concrètement, le HLR est interrogé afin de connaître la position de l'utilisateur du réseau mobile ainsi que son état (libre, occupé, éteint). Si le mobile est dans l'état "libre", le réseau interroge alors le VLR de la zone pour savoir dans quelle cellule le mobile se situe. Ainsi, le BSC de la zone demande aux différentes stations de base de sa zone de diffuser un avis d'appel. Comme le mobile est "libre", le destinataire écoute le réseau et s'aperçoit qu'on tente de le joindre et la sonnerie du terminal est activée. Une fois que l'utilisateur a décroché, un canal de communication est alloué à l'appel et les bases de données VLR et HLR sont mises à jour.

10.4 Description du canal physique

Dans un réseau GSM, deux techniques de multiplexage sont mises en œuvre : le multiplexage fréquentiel (FDMA) et le multiplexage temporel (TDMA).

10.4.1 Multiplexage fréquentiel

Dans sa version à 900 [MHz], la norme GSM occupe deux bandes de 25 [MHz]; l'une est utilisée pour la voie montante (890,2 – 915 [MHz]), l'autre pour la voie descendante (935,2 – 960 [MHz]). Il est également défini que chaque porteuse de cellule possède une densité spectrale confinée dans une bande de 200 [kHz] ce qui signifie que, théoriquement, on peut disposer de 124 canaux. Notons au passage que la bande de fréquences du DCS-1800 étant plus large, elle peut contenir 374 canaux.

Aussi, si on indique par F_u les fréquences porteuses montantes et par F_d les fréquences porteuses descendantes, les valeurs de fréquence porteuse valent

$$F_u(n) = 890,2 + 0,2 \times (n - 1) \text{ [MHz]} \quad (10.2)$$

$$F_d(n) = 935,2 + 0,2 \times (n - 1) \text{ [MHz]} \quad (10.3)$$

où $1 \leq n \leq 124$. Connaissant les canaux disponibles, il est alors possible d'effectuer un multiplexage fréquentiel en attribuant un certain ensemble de fréquences porteuses par opérateur GSM –c'est le rôle du régulateur, l'IBPT⁵ en Belgique– et l'opérateur choisit son plan de fréquences en allouant une ou plusieurs fréquences par station de base. Proximus utilise les canaux 1 à 30 et 61 à 90, Mobistar s'est vu attribuer les canaux 31 à 60 et 91 à 120; quant aux canaux restants (121 à 124), ils ont été réservés pour les téléphones sans fil. Base (nommé Orange auparavant) n'opère pas dans la bande des 900 [MHz]. Par contre, la bande DCS-1800 est utilisée par les 3 opérateurs (Base, Mobistar et Proximus).

10.4.2 La modulation

En raison de la forte variabilité de l'amplitude des signaux dans un environnement mobile, on préfère recourir à une technique de modulation angulaire pour ce type d'environnement. La technique de modulation utilisée pour porter le signal à haute fréquence est la modulation GMSK

5. <http://www.ibpt.be>

(Gaussian Minimum Shift Keying). Comme le suggère son nom, il s'agit d'une variante d'une modulation **MSK** appartenant à la famille des modulations de fréquence (**FM**) numériques. On utilise la GMSK car, en raison de la transition rapide entre 2 fréquences ($f_c - \Delta f$ et $f_c + \Delta f$), la modulation par MSK aurait nécessité une trop large bande de fréquences.

La modulation GMSK consiste en une modulation de fréquence à deux états portant non pas sur la séquence originale mais sur une nouvelle séquence dont le bit n est produit comme le résultat de la fonction du OU exclusif (**XOR**) entre le bit courant et le bit précédent. Après application du XOR, le signal est filtré. La figure 10.4 montre la création d'un signal modulé par GMSK.

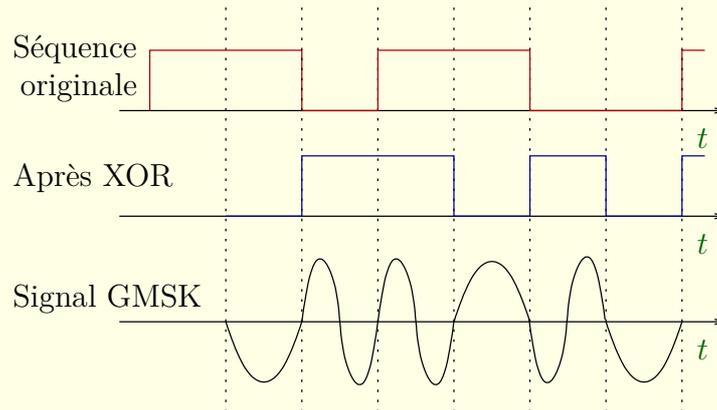


FIGURE 10.4 – Création d'un signal modulé par GMSK au départ d'un train binaire.

Au bout du compte, il faut une largeur de $200 [kHz]$ par fréquence porteuse. Sachant que le débit atteint $270 [kb/s]$, on atteint un rapport du débit à largeur de bande, appelé efficacité spectrale, proche de 1. Cette valeur est typique pour des environnement mobiles, ce qui signifie que, pour doubler le débit, il n'y a d'autre solution que de doubler la largeur de bande.

10.4.3 Multiplexage temporel

Tant pour des questions d'interférences électromagnétiques que pour des raisons d'augmentation de capacité, le multiplexage fréquentiel se double d'une multiplexage temporel. Le multiplexage temporel consiste à diviser chaque canal de communication en 8 intervalles de temps de $0,577 [ms]$ chacun.

Définition 88 [Trame] On définit dès lors une **trame** élémentaire de 8 intervalles pour une durée de $8 \times 0,577 = 4,615 [ms]$.

Comme il est exclus de transmettre toutes les informations en une fois, il faut découper l'information et la transmettre au moyen de plusieurs trames consécutives. La norme GSM prévoit une organisation spécifique de structure hiérarchique de trames. Cette hiérarchie est dessinée à la figure 10.5. Les trames sont regroupées comme suit :

- 1 multitraine de type 26 = 26 trames TDMA élémentaires et 1 multitraine de type 51 = 51 trames TDMA élémentaires,
- 1 supertraine de type 26 = 26 multitraines et 1 supertraine de type 51 = 51 multitraines
- 1 **hypertraine** = 2048 supertraines = 2.715.648 trames.

La structure en trames est à mettre en relation avec la typologie des informations véhiculées, dont il est question dans cet article à la section 10.5.3.

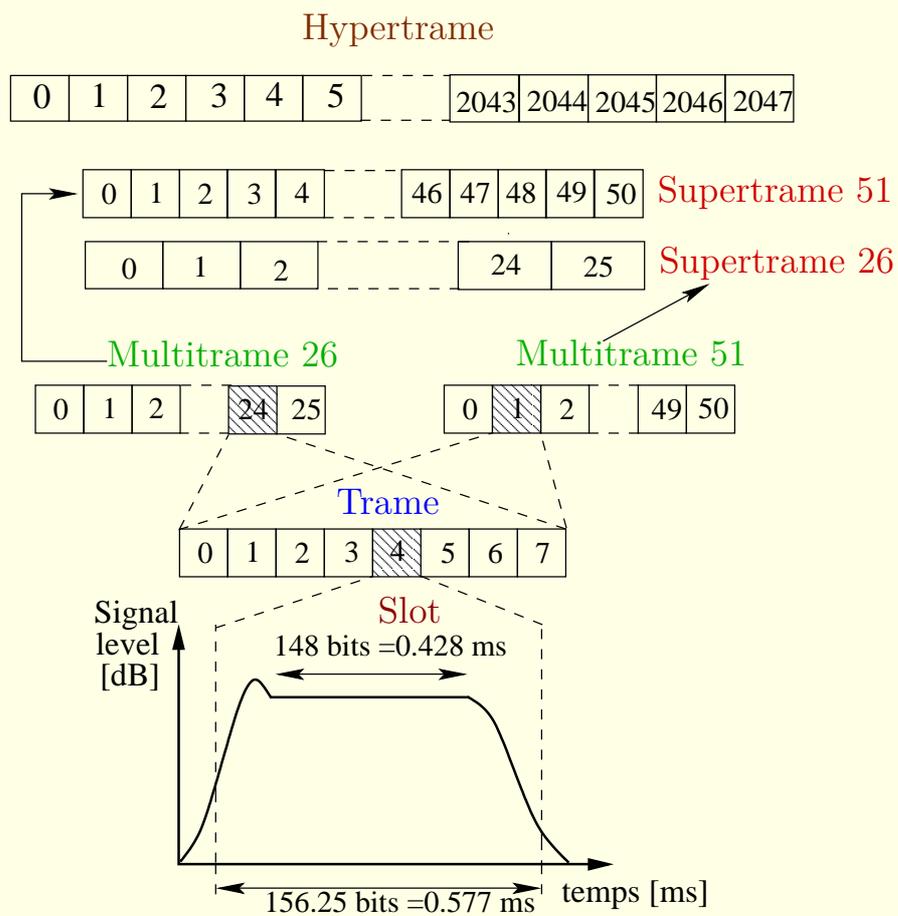


FIGURE 10.5 – Organisation des multiples de trames.

10.4.4 Mécanismes de protection contre les interférences radio-électriques

Le canal d'une communication avec un mobile est changeant et sujet à perturbations. Plusieurs mécanismes sont donc mis en œuvre pour réduire l'impact des interférences.

Le saut de fréquences ou Frequency Hopping

Pour protéger les canaux d'une source radio parasite, la fréquence porteuse utilisée pour transmettre une salve de données fluctue au cours du temps. C'est le principe du saut de fréquence ou du Frequency Hopping (FH) ; il est illustré à la figure 10.6. Ainsi, si à un moment déterminé, une bande de fréquences est sujette à fortes perturbations, seule une petite quantité de données sera perturbée.

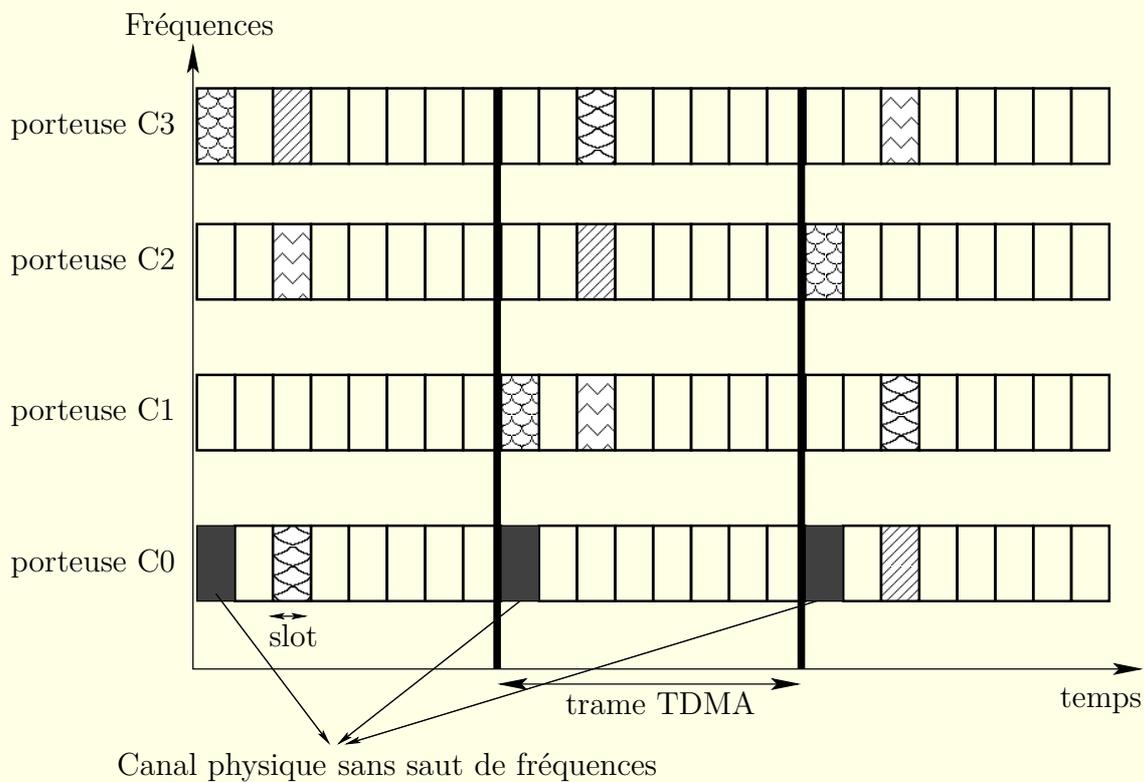


FIGURE 10.6 – Principe du saut de fréquence (d'après [13, page 169]).

La norme GSM définit un parcours de fréquence cyclique ou pseudo-aléatoire, comprenant au plus 64 fréquences porteuses. Habituellement, un algorithme standardisé génère une suite pseudo-aléatoire de nombres s_i compris dans une liste de N fréquences disponibles pour les sauts.

La configuration des sauts se fait au moyen de paramètres tels que :

- le Cell Allocation (CA), la liste des numéros des fréquences utilisées dans une cellule,
- le Mobile Allocation (MA), la liste des numéros des fréquences disponibles pour les sauts⁶,
- le Hopping Sequence Number (HSN), une valeur comprise entre 0 et 63, servant à initialiser le générateur pseudo-aléatoire, et

6. Il y en a au maximum 64.

- le Mobile Allocation Index Offset (**MAIO**), une valeur comprise entre 0 et 63 qui indique quel décalage doit être utilisé. Cette valeur de décalage est convenue à l’initialisation de l’appel et elle diffère d’un mobile à l’autre.

L’opérateur choisit le nombre de fréquences à prendre pour chaque numéro de trame élémentaire (appelée slot). Néanmoins, il doit tenir compte de la nécessité d’un mobile entrant dans le réseau de pouvoir communiquer. Pour cela, on fixe la fréquence porteuse de certains slots dans le sens de la station de base vers le mobile. Ce canal a pour fonction essentielle de permettre au mobile de se rattacher à la station de base la plus favorable. Il s’agit par exemple du slot 0 pour lequel on fixe une fréquence et on n’autorise le saut que sur 3 fréquences (pour les appels en cours sur le slot 0) au lieu de 4 pour les autres slots.

Signalons qu’il n’est pas possible de passer de la bande 900 [MHz] à la bande 1800 [MHz] pour effectuer le saut de fréquence. Autrement dit, les ressources radio des bandes GSM et DCS-1800 ne sont jamais mélangées.

Décalage temporel des envois

Pour permettre le basculement d’un mobile du mode réception en mode émission, la norme GSM prévoit un décalage de 3 slots. Plus précisément, le mobile émet des informations 3 slots après réception des signaux envoyés par la station de base.

Malgré tout, les informations envoyées par les différents mobiles autour d’une même fréquence porteuse entre en collision au droit de la station de base si la distance entre les mobiles et l’antenne est fort différente d’un mobile à l’autre. Pour éviter des collisions, certains mobiles (les plus distants) doivent avancer le moment de l’envoi. La durée de l’avance temporelle de l’envoi est appelée Timing Advance (**TA**). Elle est fournie dynamiquement par la station de base.

Le codage

Pour protéger les informations des erreurs qui peuvent se produire lors des communications radio et ainsi réduire le taux d’erreur sur bit, trois techniques de codage sont appliqués :

1. Un codage en bloc qui ajoute un bit de parité au bloc et qui permet la détection d’un nombre impair d’erreurs.
2. Un codage récurrent (algorithme de VITERBI). L’information n’est plus sectionnée en parties indépendantes mais certains bits de redondance sont placés régulièrement dans le code.
3. On effectue également un entrelacement par blocs de 464 bits. Cet entrelacement consiste à répartir les blocs selon une méthode définie. Ainsi, si le canal perturbe une suite de bits consécutifs, l’altération sera diffusée sur un grand nombre de blocs plutôt que sur la totalité d’un bloc ; les blocs affectés pourront alors être corrigés grâce aux bits redondants.

10.5 Protocoles

10.5.1 Pile de protocoles

La figure 10.7 représente l’architecture des protocoles GSM des différents éléments du réseau. Au niveau applicatif, on distingue les protocoles suivants qui, au travers de différents éléments du réseau, relie un mobile à un centre de communication (**MSC**) :

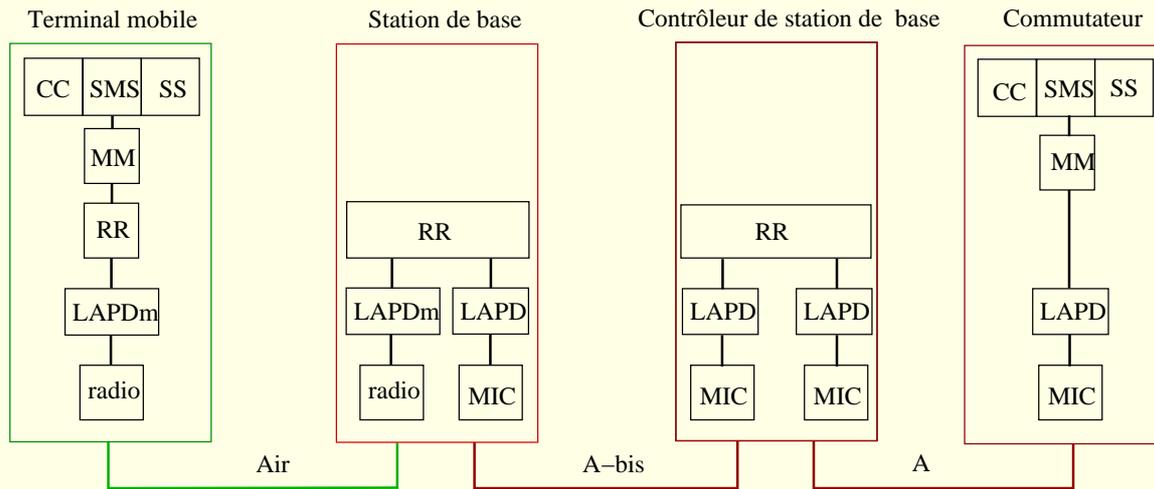


FIGURE 10.7 – Piles de protocoles de différents sous-systèmes du réseau GSM (d’après [18, page 58]).

1. Le protocole Call Control (CC) prend en charge le traitement des appels tels que l’établissement, la terminaison et la supervision.
2. Le protocole Short Message Service (SMS) qui permet l’envoi de courts messages au départ d’un mobile. La longueur d’un SMS est limité à 160 caractères de 7 bits, soit 140 bytes.
3. Le protocole Supplementary Services (SS) prend en charge les compléments de services. La liste de ces services est longue mais, à titre d’exemple, citons le Calling Line Identification Presentation (CLIP), le Calling Line Identification Restriction (CLIR) et le Call Forwarding Unconditional (CFU).
4. Le protocole Mobility Management (MM) gère l’identification, l’authentification sur le réseau et la localisation d’un terminal. Cette application se trouve dans le sous-réseau de commutation (NSS) et dans le mobile car ils doivent tous deux connaître la position du mobile dans le réseau.
5. Le protocole Radio Ressource management (RR) s’occupe de la liaison radio. Il interconnecte une BTS et un BSC car ce dernier gère l’attribution des fréquences radio dans une zone.

Les trois premiers protocoles applicatifs pré-cités (CC, SMS et SS) ne sont implémentés que dans les terminaux mobiles et les commutateurs ; leurs messages voyagent de façon transparente à travers le BSC et le BTS.

10.5.2 Les interfaces A-bis, A et X25

Présentons brièvement les trois types d’interface qui relient le BSC respectivement à la station de base (interface A-bis), au commutateur (interface A) et au centre d’exploitation et de maintenance (interface X25).

L’interface A-bis

La couche physique est définie par une liaison PCM à 2 [Mb/s] ⁷ (recommandation de la série G de l’ITU) et la couche liaison de données est composée du protocole Link Access Protocol

7. Une telle liaison offre 32 canaux à 64 [kb/s] .

D-channel (LAPD).

Comme le canal de liaison PCM a un débit unitaire de $64 [kb/s]$ et que le débit par canal radio GSM est de $13 [kb/s]$, il faut donc adapter le débit. Cette fonction est appelée **transcodage** et elle est réalisée dans une unité appelée Transcoding Rate and Adaptation Unit (TRAU). Deux solutions sont techniquement possibles et rencontrées dans les réseaux GSM :

1. Mutliplexer quatre canaux à $13 [kb/s]$ pour produire un canal à $64 [kb/s]$.
2. Faire passer le débit de chaque canal à $64 [kb/s]$.

Tout est affaire de compromis et de choix. L'avantage de la première solution est de diminuer le débit entre la station de base et le BSC où le trafic est fortement concentré. La seconde solution offre par contre l'avantage de banaliser les équipements du système en ramenant tous les équipements à $64 [kb/s]$. Souvent, la deuxième solution est utilisée au niveau des commutateurs et la première au niveau du BSC afin de garder l'avantage du faible débit de parole.

L'interface A

La couche physique est toujours définie par une liaison à PCM à $2 [Mb/s]$ mais c'est le protocole CCITT numéro 7 qui est utilisé pour la couche liaison de données.

L'interface X25

Cette interface relie le BSC au centre d'exploitation et de maintenance (OMC). Elle possède la structure en 7 couches du modèle OSI.

10.5.3 Typologie des paquets

Chaque trame consiste en un certain nombre de bits. Ces bits sont organisés suivant une structure qui diffère en fonction du protocole applicatif mis en œuvre pour chaque slot mais aussi de l'état intermédiaire du protocole considéré.

La durée d'un paquet ($0,577 [ms]$) correspond à l'émission de $156,25$ bits, dont 114 bits de message "net". En admettant que les slots se suivent sans interruption, un simple calcul ($\frac{156,25}{0,577*10^{-3}}$) montre que le débit maximum vaut $270 [kb/s]$. En pratique, le débit maximum utile (en mode full-rate) ne dépasse pas $13 [kb/s]$ en raison des bits nécessaires à la correction d'erreurs. Pour la transmission des données, cette limite descend même à $9,6 [kb/s]$ en raison de la sur-protection nécessaire à la garantie d'un taux d'erreur acceptable.

La norme définit 5 types de paquets fonctionnels, appelés bursts dans la terminologie GSM :

1. Les bursts d'accès qui sont envoyés par les mobiles lorsqu'ils veulent entrer en contact avec le réseau.
2. Les bursts de synchronisation qui contiennent les informations sur la localisation et les fréquences utilisées.
3. Les bursts normaux qui transportent les messages.
4. Les bursts de correction de fréquence.
5. Les bursts de bourrage (dummy packet) qui sont placés dans les espaces vides si aucune donnée ne doit être envoyée. Pour être précis, ce burst est composé de 2 salves de 58 bits préfixés interrompus par une séquence d'entraînement de 26 bits.

- Tous les types de burst ont une forme semblable. Ils sont composés, dans l'ordre, de :
- bits d'en-tête (tail bit, TB), nécessaires à la synchronisation. Ils correspondent toujours au code 000 sauf pour les bursts d'accès.
 - 148 bits utiles dont le format dépend du type de burst.
 - bits de fin, aussi appelés tail bit, terminés par une période temporelle de garde requise pour permettre à l'émetteur de réduire sa puissance de 70 [dB]. Elle sert aussi à compenser la durée de transmission qui est variable pour la réception d'un paquet au suivant si le mobile a bougé.

La structure des 5 types de burst est représentée à la figure 10.8.

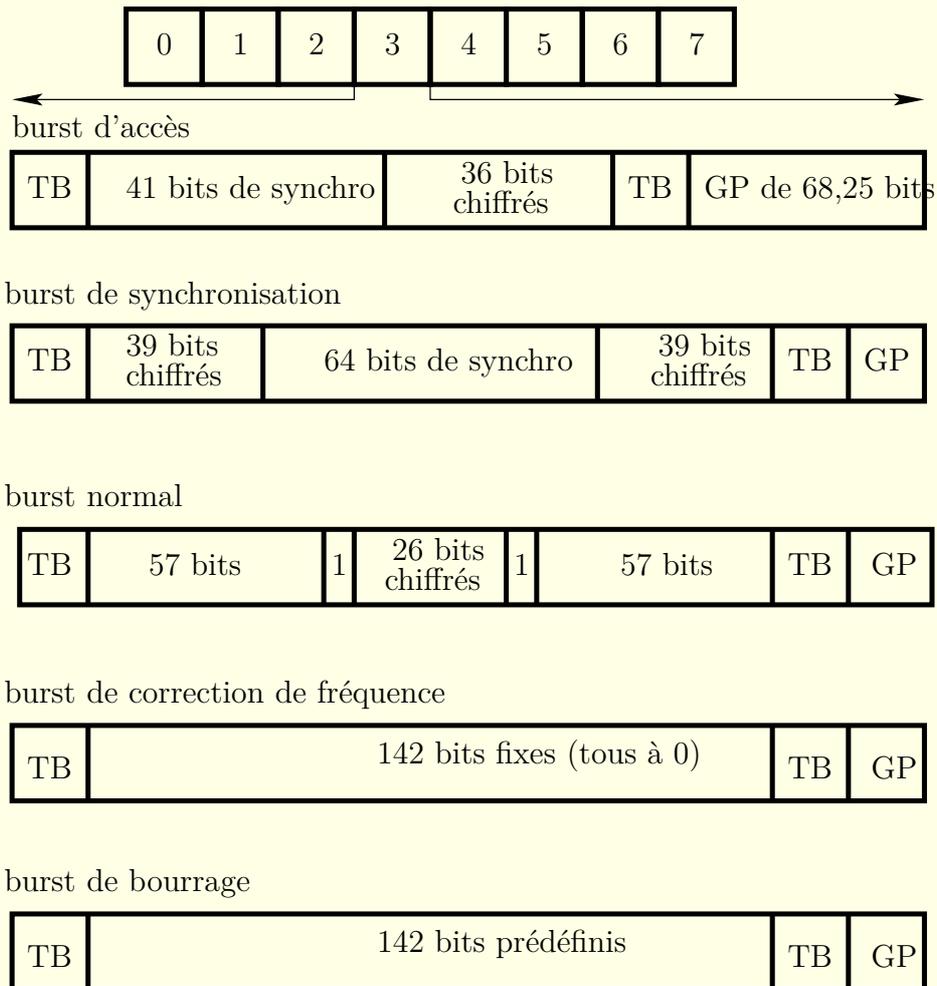


FIGURE 10.8 – Structures des 5 types de burst définis par la norme GSM (d'après la norme et [20, page 140]).

Le burst d'accès

Ce burst est émis, sur un canal dédié, par la station mobile lorsqu'elle cherche à entrer en contact avec le réseau soit pour l'établissement d'une communication, soit pour un **handover**. Il est le plus court des quatre types car il ne contient que 77 bits (41 bits de synchronisation et 36 bits d'information). Son temps de garde est de 68,25 bits, soit 0,252 [ms]. Ce temps de garde permet de tenir compte de grandes cellules et d'établir ainsi une communication avec un mobile distant jusqu'à 35 [km].

En calculant la durée de voyage d'un burst, la station peut asservir l'instant du début d'émission pour compenser le retard entraîné par la propagation des ondes. En effet, l'horloge interne des récepteurs est synchronisée grâce à un top de synchronisation envoyé par la station de base.

Le burst de synchronisation

Pour ce type de burst, 78 bits d'informations sont véhiculés pour les stations mobiles. Ces bits contiennent les renseignements concernant les fréquences à utiliser et la localisation (identité de la station de base, de la zone et de la cellule).

Le burst normal

Ce burst transporte $2 \times 57 = 114$ bits d'information séparées par 26 bits qui sont une séquence d'apprentissage destinée à régler les paramètres de réception. De plus, la zone TB correspond à 8,25 bits. Enfin, il faut ajouter à cela 2 bits qui indique s'il s'agit d'un canal de données ou d'un canal de signalisation et 6 bits pour marquer la montée ou la descente en amplitude.

Le burst de correction de fréquence

Le type de burst au format le plus simple. La station de base envoie 142 bits de données servant à prévenir des interférences possibles avec des fréquences voisines.

Le burst de bourrage

Lorsqu'un mobile est allumé, le terminal teste le niveau de puissance des fréquences des cellules proches pour déterminer la station de base à laquelle il doit s'asservir. Le burst de bourrage (dummy burst) est une séquence prédéfinie qui sert donc d'étalon de puissance. Il est aussi utilisé pour forcer une décision de handover.

Annexe A

Analyse de Fourier

A.1 Transformée de Fourier

A.1.1 Définition

Soit $g(t)$ un signal non périodique. Par définition, la transformée de FOURIER du signal $g(t)$ est donnée par

$$\mathcal{G}(f) = \int_{-\infty}^{+\infty} g(t)e^{-2\pi jft} dt \quad (\text{A.1})$$

À partir de la transformée de FOURIER, il est possible de reconstituer exactement le signal original au moyen de la transformée inverse :

$$g(t) = \int_{-\infty}^{+\infty} \mathcal{G}(f)e^{2\pi jft} df \quad (\text{A.2})$$

Pour que la transformée d'un signal $g(t)$ existe, il suffit (condition suffisante mais pas nécessaire), que $g(t)$ satisfasse les trois conditions de DIRICHLET :

1. La fonction $g(t)$ est à valeur unique avec un nombre fini de maxima et de minima pour tout intervalle fini.
2. La fonction $g(t)$ a un nombre fini de discontinuités pour tout intervalle fini.
3. La fonction est “absolument” intégrable, c'est-à-dire

$$\int_{-\infty}^{+\infty} |g(t)| dt < \infty \quad (\text{A.3})$$

La transformée de FOURIER nous fournit une interprétation intéressante puisqu'elle décompose le signal en composantes fréquentielles définies sur $[-\infty, +\infty]$.

Nous dirons que $g(t)$ et $\mathcal{G}(f)$ forment une paire de transformées de FOURIER représentée par

$$g(t) \rightleftharpoons \mathcal{G}(f)$$

En général, $\mathcal{G}(f)$ est une fonction de f à valeurs complexes. Nous pouvons donc l'exprimer sous la forme

$$\mathcal{G}(f) = \|\mathcal{G}(f)\| e^{j\theta(f)} \quad (\text{A.4})$$

où $\|\mathcal{G}(f)\|$ est appelé module de $\mathcal{G}(f)$ et $\theta(f)$ est appelée phase de $\mathcal{G}(f)$. Dans le cas particulier important où $g(t)$ est une fonction à valeurs réelles, nous avons

$$\mathcal{G}(-f) = \mathcal{G}^*(f) \quad (\text{A.5})$$

où $*$ représente le complexe conjugué. Il vient

$$|\mathcal{G}(-f)| = |\mathcal{G}(f)| \quad (\text{A.6})$$

$$\theta(-f) = -\theta(f) \quad (\text{A.7})$$

Dès lors, nous pouvons déduire deux propriétés importantes d'un signal à valeurs réelles :

1. Le spectre d'amplitude du signal est une fonction paire de la fréquence, c'est-à-dire que le spectre d'amplitude est symétrique par rapport à l'axe vertical.
2. Le spectre de phase du signal est une fonction impaire de la fréquence, c'est-à-dire que le spectre de phase est antisymétrique par rapport à l'axe vertical.

A.1.2 Propriétés

1. Linéarité

Soient $g_1(t) \rightleftharpoons \mathcal{G}_1(f)$ et $g_2(t) \rightleftharpoons \mathcal{G}_2(f)$. Alors, pour toutes constantes c_1 et c_2 , nous avons

$$c_1 g_1(t) + c_2 g_2(t) \rightleftharpoons c_1 \mathcal{G}_1(f) + c_2 \mathcal{G}_2(f) \quad (\text{A.8})$$

2. Dilatation temporelle

$$g(at) \rightleftharpoons \frac{1}{|a|} \mathcal{G}\left(\frac{f}{a}\right) \quad (\text{A.9})$$

Il est à noter que la fonction $g(at)$ représente $g(t)$ compressée dans le temps par un facteur a tandis que la fonction $\mathcal{G}(f/a)$ représente $\mathcal{G}(f)$ dilatée en fréquence par le même facteur a . Dès lors, il vient qu'une compression dans le domaine temporel équivaut à une dilatation dans le domaine fréquentiel et vice versa.

3. Dualité

Si $g(t) \rightleftharpoons \mathcal{G}(f)$, alors

$$\mathcal{G}(t) \rightleftharpoons g(-f) \quad (\text{A.10})$$

4. Translation temporelle

Si $g(t) \rightleftharpoons \mathcal{G}(f)$, alors

$$g(t - t_0) \rightleftharpoons \mathcal{G}(f) e^{-2\pi j f t_0} \quad (\text{A.11})$$

Il en résulte que le fait de traduire la fonction $g(t)$ de t_0 ne modifie pas l'amplitude de sa transformée de FOURIER, par contre sa phase est modifiée d'un facteur linéaire $-2\pi j f t_0$.

5. Translation fréquentielle

Si $g(t) \rightleftharpoons \mathcal{G}(f)$, alors

$$g(t) e^{2\pi j f_c t} \rightleftharpoons \mathcal{G}(f - f_c) \quad (\text{A.12})$$

La multiplication de la fonction $g(t)$ par le facteur $e^{2\pi j f_c t}$ est équivalente à une translation de la transformée de FOURIER $\mathcal{G}(f)$ dans le domaine fréquentiel. Cette propriété est appelée théorème de modulation.

6. Aire

Si $g(t) \rightleftharpoons \mathcal{G}(f)$, alors

$$g(0) = \int_{-\infty}^{+\infty} \mathcal{G}(f) df \quad (\text{A.13})$$

7. Dérivée temporelle

Si $g(t) \rightleftharpoons \mathcal{G}(f)$, alors

$$\frac{d}{dt}g(t) \rightleftharpoons 2\pi j f \mathcal{G}(f) \quad (\text{A.14})$$

8. Intégration dans le domaine temporel

Si $g(t) \rightleftharpoons \mathcal{G}(f)$ et $\mathcal{G}(0) = 0$, alors

$$\int_{-\infty}^t g(\tau) d\tau \rightleftharpoons \frac{1}{2\pi j f} \mathcal{G}(f) \quad (\text{A.15})$$

9. Multiplication de deux fonctions dans le domaine temporel

Soient $g_1(t) \rightleftharpoons \mathcal{G}_1(f)$ et $g_2(t) \rightleftharpoons \mathcal{G}_2(f)$. Alors

$$g_1(t)g_2(t) \rightleftharpoons \int_{-\infty}^{+\infty} \mathcal{G}_1(\lambda)\mathcal{G}_2(f - \lambda)d\lambda \quad (\text{A.16})$$

L'intégrale du membre de droite est connue sous le nom d'intégrale de convolution exprimée dans le domaine fréquentiel et la fonction de f résultante est appelée convolution de $\mathcal{G}_1(f)$ et $\mathcal{G}_2(f)$. Dès lors, la multiplication de deux signaux dans le domaine temporel est transformée dans le domaine fréquentiel en la convolution de leur transformée de FOURIER respective. Cette propriété est aussi appelée théorème de multiplication.

10. Multiplication dans le domaine fréquentiel

Soient $g_1(t) \rightleftharpoons \mathcal{G}_1(f)$ et $g_2(t) \rightleftharpoons \mathcal{G}_2(f)$. Alors

$$\int_{-\infty}^{+\infty} g_1(\tau)g_2(t - \tau)d\tau \rightleftharpoons \mathcal{G}_1(f)\mathcal{G}_2(f) \quad (\text{A.17})$$

La convolution de deux signaux dans le domaine temporel est transformée dans le domaine fréquentiel en la multiplication de leur transformée de FOURIER respective. Cette propriété est appelée théorème de convolution.

L'équation (A.17) peut être réécrite sous la forme

$$g_1(t) \otimes g_2(t) \rightleftharpoons \mathcal{G}_1(f)\mathcal{G}_2(f) \quad (\text{A.18})$$

11. Égalité de Parseval

Si $g(t) \rightleftharpoons \mathcal{G}(f)$, alors

$$\int_{-\infty}^{+\infty} \|g(t)\|^2 dt = \int_{-\infty}^{+\infty} \|\mathcal{G}(f)\|^2 df \quad (\text{A.19})$$

Cette propriété est également appelée théorème d'énergie de RAYLEIGH. La quantité $\|\mathcal{G}(f)\|^2$ est appelée densité spectrale d'énergie du signal $g(t)$.

A.1.3 Exemples

1. Considérons un signal rectangulaire $g(t)$ de durée T et d'amplitude A . Afin de définir mathématiquement ce signal, nous introduisons la fonction

$$rect(t) = Rect_{[-\frac{1}{2}, \frac{1}{2}]}(t) = \begin{cases} 1 & -\frac{1}{2} < t < \frac{1}{2} \\ 0 & |t| > \frac{1}{2} \end{cases} \quad (\text{A.20})$$

appelée en général fonction rectangle. Dès lors, nous pouvons écrire

$$g(t) = A \operatorname{rect}\left(\frac{t}{T}\right) = A \operatorname{Rect}_{[-\frac{T}{2}, \frac{T}{2}]}(t) \quad (\text{A.21})$$

La transformée de FOURIER du signal $g(t)$ est donnée par

$$\mathcal{G}(f) = \int_{-T/2}^{T/2} A e^{-2\pi jft} dt \quad (\text{A.22})$$

$$= AT \left(\frac{\sin(\pi ft)}{\pi ft} \right) \quad (\text{A.23})$$

Afin de simplifier la notation précédente, nous introduisons la fonction sinus cardinal, *sinc*, définie par

$$\operatorname{sinc}(\lambda) = \frac{\sin(\pi\lambda)}{\pi\lambda} \quad (\text{A.24})$$

Finalement, il vient

$$\mathcal{G}(f) = AT \operatorname{sinc}(fT) \quad (\text{A.25})$$

Nous avons donc la paire de transformées de FOURIER :

$$A \operatorname{rect}\left(\frac{t}{T}\right) \Leftrightarrow AT \operatorname{sinc}(ft) \quad (\text{A.26})$$

2. Considérons le signal $g(t)$ suivant :

$$g(t) = A \operatorname{sinc}(2Wt) \quad (\text{A.27})$$

En appliquant la propriété de dualité de la transformée de FOURIER à la relation A.26 et étant donné que la fonction rectangle est une fonction paire du temps, il vient

$$AT \operatorname{sinc}(Tt) \Leftrightarrow A \operatorname{rect}\left(\frac{f}{T}\right) \quad (\text{A.28})$$

En posant $T = 2W$, nous obtenons la paire de transformées de FOURIER suivante

$$A \operatorname{sinc}(2Wt) \Leftrightarrow \frac{A}{2W} \operatorname{rect}\left(\frac{f}{2W}\right) \quad (\text{A.29})$$

A.2 La fonction Delta de Dirac

A.2.1 Définition

La fonction Delta de DIRAC, notée $\delta(t)$, est définie par

$$\delta(t) = 0, \quad t \neq 0 \quad (\text{A.30})$$

et

$$\int_{-\infty}^{+\infty} \delta(t) dt = 1 \quad (\text{A.31})$$

Il est possible de donner une autre définition de la fonction $\delta(t)$ qui incorpore les équations A.30 et A.31 :

$$\int_{-\infty}^{+\infty} g(t) \delta(t - t_0) dt = g(t_0) \quad (\text{A.32})$$

où $g(t)$ est une fonction continue du temps.

A.2.2 Transformée de Fourier de $\delta(t)$

Par définition, la transformée de FOURIER de $\delta(t)$ est donnée par

$$\int_{-\infty}^{+\infty} \delta(t) e^{-2\pi jft} dt \quad (\text{A.33})$$

Étant donné que la fonction $e^{-2\pi jft}$ évaluée en $t = 0$ vaut 1, il vient finalement

$$\delta(t) \Leftrightarrow 1 \quad (\text{A.34})$$

Donc, le spectre de la fonction Delta de DIRAC s'étend uniformément sur tout l'intervalle de fréquence $([-\infty, +\infty])$, comme le montre la figure A.1.

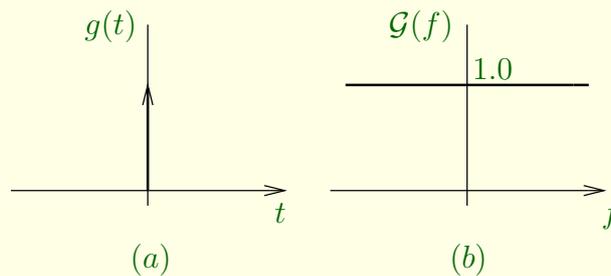


FIGURE A.1 – (a) La fonction Delta de DIRAC. (b) Spectre de $\delta(t)$.

A.2.3 Applications

1. Signal DC

En utilisant la propriété de dualité de la transformée de FOURIER et étant donné que la fonction Delta est une fonction paire, on peut écrire

$$1 \Leftrightarrow \delta(f) \quad (\text{A.35})$$

Le spectre d'un signal DC est donc discret et comporte une seule raie située en $f = 0$ (figure A.2).

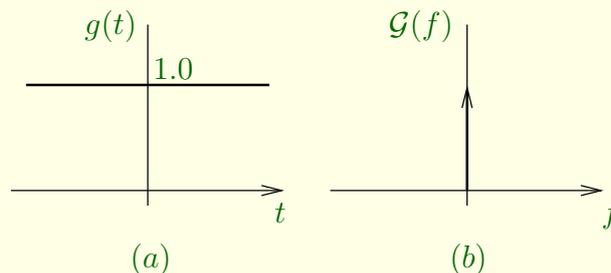


FIGURE A.2 – (a) Signal DC. (b) Spectre.

2. Signal complexe exponentiel

En appliquant la propriété de translation fréquentielle de la transformée de FOURIER à A.35, il vient

$$e^{2\pi jfct} \Leftrightarrow \delta(f - f_c) \quad (\text{A.36})$$

Le spectre d'un signal complexe exponentiel de fréquence f_c se limite donc à une raie située en $f = f_c$.

3. Signal sinusoïdal

Soit le signal $\cos(2\pi f_c t)$. Utilisons tout d'abord la formule bien connue

$$\cos(2\pi f_c t) = \frac{e^{2\pi j f_c t} + e^{-2\pi j f_c t}}{2} \quad (\text{A.37})$$

En utilisant (A.36) et la propriété de linéarité de la transformée de FOURIER, il vient

$$\cos(2\pi f_c t) \Leftrightarrow \frac{\delta(f - f_c) + \delta(f + f_c)}{2} \quad (\text{A.38})$$

Le spectre d'un signal cosinusoidal comporte donc 2 raies pondérées par le facteur $1/2$ et situées en $f = \pm f_c$.

Il est également aisé de montrer que

$$\sin(2\pi f_c t) \Leftrightarrow \frac{\delta(f - f_c) - \delta(f + f_c)}{2j} \quad (\text{A.39})$$

4. Convolution d'une fonction par une raie de Dirac.

Soit un signal $g(t)$ dont la transformée de FOURIER est $\mathcal{G}(f)$. De la relation $\delta(t) \Leftrightarrow 1$, on déduit que

$$g(t) \otimes \delta(t) \Leftrightarrow \mathcal{G}(f) \quad (\text{A.40})$$

Dès lors,

$$g(t) \otimes \delta(t) = g(t) \quad (\text{A.41})$$

ou encore

$$\mathcal{G}(f) \otimes \delta(f) = \mathcal{G}(f) \quad (\text{A.42})$$

La convolution d'une fonction par une impulsions de DIRAC fournit donc la fonction originale. Une extension simple de cette propriété conduit à

$$\mathcal{G}(f) \otimes \delta(f - f_c) = \mathcal{G}(f - f_c) \quad (\text{A.43})$$

A.3 Quelques signaux fondamentaux

A.3.1 Définitions

1. Fonction rectangle

$$\text{rect}(t) = \begin{cases} 1 & -\frac{1}{2} < t < \frac{1}{2} \\ 0 & |t| > \frac{1}{2} \end{cases} \quad (\text{A.44})$$

2. Fonction échelon

$$u(t) = \begin{cases} 1 & t > 0 \\ 0 & t < 0 \end{cases} \quad (\text{A.45})$$

3. Fonction signe

$$\text{sign}(t) = \begin{cases} 1 & t > 0 \\ -1 & t < 0 \end{cases} \quad (\text{A.46})$$

4. Fonction sinc

$$\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t} \quad (\text{A.47})$$

5. Fonction Si

$$\text{Si}(u) = \int_0^u \frac{\sin(x)}{x} dx \quad (\text{A.48})$$

A.3.2 Transformées de Fourier

$$\operatorname{rect}\left(\frac{t}{T}\right) \quad T \operatorname{sinc}(fT) \quad (\text{A.49})$$

$$\operatorname{sinc}(2Wt) \quad \frac{1}{2W} \operatorname{rect}\left(\frac{f}{2W}\right) \quad (\text{A.50})$$

$$e^{-at}u(t), \quad a > 0 \quad \frac{1}{a+2\pi jf} \quad (\text{A.51})$$

$$e^{-a|t|}, \quad a > 0 \quad \frac{2a}{a^2+(2\pi f)^2} \quad (\text{A.52})$$

$$e^{-\pi t^2} \quad e^{-\pi f^2} \quad (\text{A.53})$$

$$\delta(t) \quad 1 \quad (\text{A.54})$$

$$1 \quad \delta(f) \quad (\text{A.55})$$

$$\delta(t-t_0) \quad e^{-2\pi jft_0} \quad (\text{A.56})$$

$$e^{2\pi jfct} \quad \delta(f-f_c) \quad (\text{A.57})$$

$$\cos(2\pi fct) \quad \frac{1}{2}[\delta(f-f_c) + \delta(f+f_c)] \quad (\text{A.58})$$

$$\sin(2\pi fct) \quad \frac{1}{2j}[\delta(f-f_c) - \delta(f+f_c)] \quad (\text{A.59})$$

$$\operatorname{sgn}(t) \quad \frac{1}{\pi jf} \quad (\text{A.60})$$

$$\frac{1}{\pi t} \quad -j \operatorname{sgn}(f) \quad (\text{A.61})$$

$$u(t) \quad \frac{1}{2}\delta(f) + \frac{1}{2\pi jf} \quad (\text{A.62})$$

$$\sum_{i=-\infty}^{+\infty} \delta(t-iT_0) \quad \frac{1}{T_0} \sum_{n=-\infty}^{+\infty} \delta\left(f - \frac{n}{T_0}\right) \quad (\text{A.63})$$

Annexe B

Formules utiles

B.1 Constantes physiques

Nom	Symbole	Valeur	Unités
Vitesse de la lumière dans le vide	c	3×10^8	$[m/s]$
Constante de BOLTZMAN	k_B	$1,38 \times 10^{-23}$	$[J/K]$
Charge d'un électron	q	$1,6 \times 10^{-19}$	$[C]$
Permittivité du vide	ϵ_0	$8,85 \times 10^{-12}$	$[F/m]$
Perméabilité du vide	μ_0	$4\pi \times 10^{-7}$	$[H/m]$

B.2 Relations

B.2.1 Relations trigonométriques

$$e^{\pm j\theta} = \cos \theta \pm j \sin \theta \quad (\text{B.1})$$

$$\cos(A \pm B) = \cos A \cos B \mp \sin A \sin B \quad (\text{B.2})$$

$$\sin(A \pm B) = \sin A \cos B \pm \cos A \sin B \quad (\text{B.3})$$

$$\cos A \cos B = \frac{1}{2}(\cos(A - B) + \cos(A + B)) \quad (\text{B.4})$$

$$\sin A \cos B = \frac{1}{2}(\sin(A - B) + \sin(A + B)) \quad (\text{B.5})$$

$$\sin A \sin B = \frac{1}{2}(\cos(A - B) - \cos(A + B)) \quad (\text{B.6})$$

B.2.2 Développement en série

Série de TAYLOR

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \dots \quad (\text{B.7})$$

B.3 Notation phasorielle

Dans tout circuit, sans source d'énergie, les courants sont amortis et s'annulent au cours du temps. L'application de sources de tension ou de courant sinusoïdales engendre des courants qui, après disparition des régimes transitoires, sont sinusoïdaux et de même fréquence que celle des sources si tous les éléments du circuit ont des caractéristiques linéaires. C'est le régime sinusoïdal permanent. Dès lors, la réponse d'un circuit à l'excitation sinusoïdale sera décrite, en fonction de la fréquence, par sa fonction de transfert. La linéarité des équations joue ici un rôle essentiel.

À chaque grandeur physique évoluant sinusoïdalement, on associe une grandeur complexe, dont la partie réelle sera identifiée avec la grandeur physique à décrire, et dont l'amplitude complexe, qui en regroupe l'amplitude et la phase, sera la grandeur utile.

Définition 89 *Considérons une grandeur sinusoïdale $x(t) = X \cos(\omega t + \theta)$. On lui associe le phaseur*

$$\hat{X} = X e^{j\theta} \quad (\text{B.8})$$

L'expression temporelle est la partie réelle du phaseur préalablement multiplié par l'exponentielle imaginaire $e^{j\omega t}$. En effet,

$$x(t) = \text{Re}(\hat{X} e^{j\omega t}) \quad (\text{B.9})$$

$$= \text{Re}(X e^{j(\omega t + \theta)}) \quad (\text{B.10})$$

$$= X \cos(\omega t + \theta) \quad (\text{B.11})$$

Le phaseur est un concept purement mathématique ; il n'a pas de signification physique mais il permet de simplifier l'expression et le calcul de certaines expressions. Ainsi, par exemple

$$\frac{\partial(\hat{X} e^{j\omega t})}{\partial t} = j\omega \hat{X} e^{j\omega t} \quad (\text{B.12})$$

et

$$\int \hat{X} e^{j\omega t} dt = \frac{\hat{X} e^{j\omega t}}{j\omega} \quad (\text{B.13})$$

B.4 Inégalité de Schwarz

Étant donnés deux signaux f et g associés, chacun à une grandeur du second ordre finie, telle que

$$\|f\|^2 \triangleq S_f < +\infty \quad (\text{B.14})$$

$$\|g\|^2 \triangleq S_g < +\infty \quad (\text{B.15})$$

On considère le signal de norme finie, par hypothèse

$$h \triangleq f + k \langle f, g \rangle g \quad (\text{B.16})$$

où k est un réel. Écrire que le carré de la norme de h est positif entraîne, en tenant compte des propriétés de symétrie hermitienne du produit scalaire, que le trinôme du second degré suivant est toujours positif

$$k^2 |\langle f, g \rangle|^2 \langle g, g \rangle + 2k |\langle f, g \rangle|^2 + |\langle f, f \rangle| \quad (\text{B.17})$$

En écrivant alors que le discriminant réduit est toujours négatif ou nul il vient l'inégalité de SCHWARZ

$$|\langle f, g \rangle|^2 \leq \langle f, f \rangle \langle g, g \rangle \quad (\text{B.18})$$

L'égalité a lieu lorsque f et g sont proportionnels, c'est-à-dire s'il existe α tel que

$$f = \alpha g \quad (\text{B.19})$$

Index

A

ACK, 163
AM, 29
ARQ, 163
authentification, 215

B

bande
 de base, 28, 117
 de garde, 102
 passante, 22
baud, 117
bit de parité, 165
bruit, 26
BSC, 213
BTS, 213
burst, 224

C

CA, 221
canal
 binaire, 166
 discret sans mémoire, 166
 gaussien, 166
 modèle, 166
capacité, 118
CDMA, 211
cellule, 207
chiffrement, 215
CLIP, 223
CLIR, 223
code, 2
 à parité, 170
 de Golay, 176
 de Hamming, 176
 de Reed-Solomon, 177
 linéaire, 167
 systématique, 165, 171
concentrateur, 214
correction d'erreur, 172, 174
Costas, 42
crête

détecteur, 36

D

décibel, 18
délai, 24
détection d'erreur, 172
dB, 18
dBm, 18
dBW, 18
DCS-1800, 210
densité spectrale de puissance, 121
diagramme de l'œil, 145
distance
 de Hamming, 175
 minimale, 175
distorsion, 24
DPCM, 116

E

échantillonnage, 99
efficacité spectrale, 120
EIR, 215, 217
énergie, 17, 138
ETSI, 5
Eutelsat, 5

F

FDMA, 210
filtre adapté, 139
FM, 48
Fourier
 transformée, 19
Frequency Hopping, 221
FTP, 11
full-duplex, 160

G

GMSC, 215
GMSK, 218
GPRS, 207
GSM, 206

H

half-duplex, 160

Hamming

code, 176

distance, 175

poids, 175

handover, 225

Hertz, 20

HLR, 216

HSN, 221

hypertrame, 219

I

IBPT, 6

IETF, 5

IMEI, 217

IMSI, 213

indice

de modulation, 47

Intelsat, 5

Internet, 9

IS-95, 211

ITU, 5

L

LAPD, 224

Lempel, 161

M

MA, 221

MAIO, 222

Manchester, 130

matrice de contrôle de parité, 172

matrice génératrice, 169

MCC, 213

MIC, 109

modem, 159

modulation

AM, 29

angulaire, 46

de phase, 48

delta, 116

indice, 47

quadratique, 35

taux, 47

MSB, 164

MSC, 215

MSISDN, 213

MSK, 219

multiplieur, 34

multitrajet, 26

multitrame, 219

N

NACK, 163

NRZ, 125

bipolaire, 128

unipolaire, 125

NSS, 215

Nyquist, 100, 118

O

octet, 14, 164

OSI, 163

P

période, 16

PAL, 15

PAM, 105

parité

bit, 165

impair, 170

pair, 170

Parseval, 20

PCM, 223

PDM, 108

PIN, 213

PLMN, 213, 216

PM, 48

poids de Hamming, 175

porteuse, 28

PPM, 108

probabilité d'erreur, 144

protocole, 9

puissance, 17

densité spectrale, 121

PUK, 213

Q

QAM, 40

quantification, 109

bruit, 111, 112

loi, 115

pas, 109

R

Rayleigh, 20

RDS, 205

Reed-Solomon, 177

roaming, 213, 215
RTC, 208
RZ, 125
 bipolaire, 130
 unipolaire, 129

S

série V, 161
SAP, 9
SDU, 9
service, 8
Shannon, 99, 118
signal
 modulant, 27
SIM, 213
simplex, 160
SMS, 223
SSB, 44
surmodulation, 31
symbole, 98
syndrome, 173

T

TA, 222
taux de modulation, 47
taux de redondance, 167
TDMA, 211
TMSI, 213
trame, 219
transmission
 asynchrone, 147
 synchrone, 148
TRAU, 224
turbo-code, 177

U

UER, 5
UMTS, 207, 211
URSI, 5

V

V24, 159
V34, 162
valence, 117
VLR, 215

W

Whittaker, 101

X

xDSL, 131

Z

Ziv, 161

Glossaire

- AM** Amplitude Modulation Technique de modulation qui consiste à modifier l'amplitude d'une porteuse sinusoïdale en fonction de l'évolution temporelle du signal à transmettre (appelé signal modulant). 14, 29
- ATM** Asynchronous Transfer Mode. Une technologie de transmission de données à haut débit, qui repose sur la commutation de cellules longues de 53 bytes. Cette technologie a un rôle similaire à celui du protocole IP. 11
- AuC** Authentication Center. Centre d'authentification (lié à un HLR) utilisé dans les réseaux GSM. 215
- authentification** Fonction cryptographique qui consiste à identifier une personne. Cette fonction peut être assurée par différentes implémentations dont PGP par exemple. . . 215
- bande passante** Plage de fréquences qui peuvent être transmises correctement à travers un canal de transmission. Elle s'exprime en Hertz (Hz). Par exemple, la bande passante d'un signal téléphonique analogique va de 300 à 3400 Hz. 22
- baud** du nom de BAUDOT, inventeur du code télégraphique. Unité exprimant le nombre de symboles par seconde. Elle équivaut au bits par seconde si un symbole représente une information binaire. 14
- BER** Bit Error Rate. Taux d'erreur sur bit. Il s'agit d'une mesure pour déterminer la qualité d'une transmission numérique. Le taux d'erreur sur bit est le nombre d'erreurs, exprimé par une probabilité, dues à des imperfections du canal (bruit, interférences, ...). . . 144
- bit** Terme résultant de la contraction de binary digit (chiffre binaire). Unité fondamentale de mesure de l'information (numérique). En informatique, un bit vaut 0 ou 1. . . 14, 15
- bruit** Perturbation d'une transmission susceptible de dégrader le signal utile transmis. Le bruit peut par exemple s'ajouter au signal utile (bruit additif) ou perturber la phase d'un signal modulé (bruit de phase). 26
- BSC** Base Station Controller. Station qui contrôle les communications d'un groupe de cellules dans un réseau de communications GSM. Elle concentre le trafic de plusieurs BTS. . . 213
- BTS** Base Transceiver Station. Station de base d'un réseau GSM. Elle permet notamment d'émettre et de recevoir un signal radio. 213
- byte** Le byte, ou octet, est une mesure de l'information binaire. 1 byte vaut 8 bits. . . . 14
- CA** Certification Authority ou Cell Allocation. L'autorité de certification est une entité d'un système transactionnel électronique sécurisé. Généralement, cette autorité délivre et vérifie des certificats. Dans la terminologie GSM, il s'agit de la liste des numéros de fréquences utilisées dans une cellule. 221
- canal** En théorie de la communication, partie d'un système de communication qui assure le transfert de l'information entre une source et un destinataire. Il existe plusieurs modèles mathématiques de canal destinés à prédire les performances. 23, 166
- CDMA** Code Division Multiple Access. Technologie de transmission numérique permettant la transmission de plusieurs flux simultanés par répartition de code. Cette technologie permet une utilisation permanente de la totalité de la bande de fréquences allouée à l'ensemble des utilisateurs. La technologie prévoit un mécanisme d'accès aux ressources. 211
- cellule** En radiocommunications, zone géographique élémentaire d'un réseau radiocellulaire à laquelle on affecte un ensemble de fréquences non réutilisables dans les zones contiguës. C'est également le nom donné à un paquet ATM qui a une taille de 53 bytes dont 48 sont destinées à recevoir les données d'un utilisateur. 207

CFU	Call Forwarding Unconditional. Numéro de téléphone vers lequel tout appel est redirigé à la demande de l'abonné appelé.	223
chiffrement	Terme qui désigne l'action de chiffrer un texte, des informations ou des données. Le chiffrement consiste à transformer un texte de sorte qu'il faille une clé pour comprendre le message.	215
circuit	Voie de transmission physique entre deux points ou plus. Un circuit est établi pour la totalité de la durée de transmission. L'établissement et le relâchement du circuit s'effectuent par signalisation.	7
CLIP	Calling Line Identification Presentation. Service complémentaire de téléphonie qui consiste à afficher le numéro du correspondant sur le terminal.	223
CLIR	Calling Line Identification Restriction. Service complémentaire de téléphonie qui empêche que le numéro du correspondant n'apparaisse sur le terminal d'un utilisateur.	223
codage	Technique de numérisation consistant à affecter une valeur numérique aux échantillons obtenus après échantillonnage du signal analogique. Le codage peut également désigner un changement de représentation de l'information numérique.	2
compression	Procédé permettant de réduire le volume (en bits) ou le débit (en bits par seconde) des données numérisées (parole, image, texte, etc). Un signal numérique non comprimé représente généralement un volume trop important pour les applications courantes.	96
concentrateur	Organe permettant de concentrer le trafic et pouvant posséder une intelligence capable de gérer diverses commutations et divers protocoles.	214
CRC	Cyclic Redundancy Check. Code à redondance cyclique. Ce code est ajouté à un message numérique pour permettre la détection ou la correction d'erreurs de transmission.	161
DCS	Digital Communication System. Un système GSM porté de la bande de fréquences des 900 [MHz] vers 1800 [MHz]. Le système DCS-1800 a plus de canaux (374) mais les protocoles et services sont quasi identiques.	210
diffraction	Terme associé à l'effet d'interaction entre une onde et un obstacle. Un obstacle diffractant se comporte comme une source d'émission secondaire.	200
EIR	Equipment Identity Register. Identifiant destiné à permettre de désactiver un téléphone mobile (GSM) qui aurait été volé.	215
Ethernet	Protocole de réseau local (LAN) qui utilise une topologie de bus ou d'étoile et permet des transferts de données à 10 [Mb/s]. Il existe des versions plus rapides : Fast Ethernet (100 [Mb/s]) et Gigabit Ethernet (1000 [Mb/s]).	131
FDM	Frequency Division Multiplexing. Mécanisme de répartition de ressources par multiplexage fréquentiel.	57
FDMA	Frequency Division Multiple Access. Technique de répartition de ressources par multiplexage fréquentiel. Cette technique prévoit un mécanisme d'accès aux ressources.	210, 218
FH	Frequency Hopping. Technique du saut de fréquences qui consiste à modifier la fréquence porteuse d'un signal modulé en suivant une liste pré-déterminée.	221
FM	Frequency Modulation. Modulation de fréquences. Technique par laquelle on module la fréquence instantanée d'une porteuse au moyen du signal modulant à transmettre.	14, 19, 48, 196, 200, 219
FTP	File Transfer Protocol. Protocole permettant de transférer des fichiers, via l'internet, d'un ordinateur à un autre. FTP est un système client/serveur.	11

GMSC	Gateway Mobile Switching Center. Centre de commutation pour mobile semblable à un MSC. Il est placé en bordure de réseau d'un opérateur GSM de manière à permettre l'interconnexion avec d'autres réseaux.	215
GMSK	Gaussian Minimum Shift Keying. Nom de la technique de modulation numérique utilisée pour la transmission radio des mobiles GSM.	159, 218
GPRS	General Packet Radio Service. Technologie de transmission par paquets facilitant l'accès à Internet à haut débit par GSM. Le débit peut varier de 56 jusqu'à 115 [kb/s]. Il est également possible d'établir des connexions permanentes.	207
GSM	Global System for Mobile Communications. Standard de téléphonie mobile adopté en Europe, en Asie et en Australie.	5, 11, 94, 193, 206, 210
handover	Terme désignant le mécanisme par lequel un mobile peut transférer sa connexion d'une station de base vers une autre ou, sur la même station, d'un canal radio vers un autre.	225
HLR	Home Location Register. Base de données centrale d'un réseau GSM contenant toutes les informations relatives aux abonnés du réseau (profil, position actuelle, ...).	215
HSN	Hopping Sequence Number. Une classe de paramètres, définis dans la norme GSM, pour configurer la séquence de porteuses utilisées pour des sauts de fréquences.	221
HTTP	Hypertext Transfer Protocol. Protocole définissant comment les pages web sont transférées entre un serveur et un navigateur internet.	13
hypertrame	L'unité temporelle la plus longue de la hiérarchie GSM. Elle totalise 3 heures, 28 minutes, 53 secondes et 760 millisecondes. Elle est composée de 2048 supertrames, composées elles-mêmes de 1326 multitrames.	219
Hz	Hertz. Unité de mesures de la fréquence. 5 Hz signifie une fréquence de 5 cycles par seconde.	20
IMEI	International Mobile station Equipment Identity. Numéro unique identifiant un terminal GSM; il est indépendant du numéro d'abonné et il permet de désactiver un équipement volé.	217
IMSI	International Mobile Subscriber Identity. Numéro international unique d'un abonné GSM.	213, 216
IP	Internet Protocol. Protocole de transmission de données de niveau 3 (couche réseau). Il s'agit du principal protocole (avec TCP) utilisé pour les communications par Internet. Les données à envoyer sont réparties en "paquets" individuels et totalement indépendants. Chaque ordinateur (ou "hôte") sur Internet possède au moins une adresse permettant de l'identifier de façon unique par rapport à tous les autres; chaque paquet de données reprend à la fois l'adresse de l'expéditeur et celle du destinataire. Le rôle du protocole Internet est de veiller à ce que chaque paquet de données arrive bien à l'adresse appropriée. Le protocole Internet est un protocole dit "sans connexion" : il n'y a pas de liaison entre les points terminaux de la communication, les paquets pouvant être transmis par différentes routes et ne devant pas forcément arriver à destination dans l'ordre correct. Une fois les paquets de données arrivés à destination, c'est en effet un autre protocole, le protocole TCP (Transmission Control Protocol), qui se charge de les mettre dans le bon ordre.	8, 9
IS-95	Norme américaine de réseau cellulaire (dit de seconde génération ou 2G) basée sur la méthode d'accès CDMA.	211
ISDN	Integrated Services Digital Network Désigne le réseau téléphonique numérique RNIS.	215
ISO	International Organization for Standardization. Fédération mondiale d'organismes nationaux de normalisation de quelque 130 pays, à raison d'un organisme par pays. Elle a	

	pour mission de favoriser le développement de la normalisation et des activités connexes dans le monde, en vue de faciliter entre les nations les échanges de biens et de services et de développer la coopération dans les domaines intellectuel, scientifique, technique et économique. Les travaux de l'ISO aboutissent à des accords internationaux qui sont publiés sous la forme de normes internationales.	8
ITU	International Telecommunications Union. Groupe de normalisation spécifique aux télécommunications, rattaché aux Nations-Unies comme agence spécialisée. C'est la section ITU-R qui alloue les bandes de fréquences.	5, 115
LAPD	Link Access Protocol D-channel. Protocole de liaison de données utilisée dans le réseau GSM. Il est défini dans la famille des recommandations X25 de l'ITU.	224
MA	Mobile Allocation. Liste des numéros de fréquences utilisables pour des sauts de fréquences dans un réseau GSM.	221
MAIO	Mobile Allocation Index Offset. Décalage permettant à chaque terminal GSM d'utiliser une série de fréquences différentes d'un mobile à l'autre pour les sauts de fréquence.	222
MCC	Mobile Country Code. Nombre à 3 chiffres identifiant un pays (Belgique = 206, France = 208).	213
MNC	Mobile Network Code. Un nombre à 2 chiffres utilisé par identifier un PLMN.	213
modem	modulateur démodulateur. Appareil transmettant des signaux numériques sur le réseau téléphonique analogique. Offre les fonctions de numérotation, de connexion, et éventuellement de compression et de correction d'erreur.	159
modulation	Technique consistant à modifier l'amplitude, la phase ou la fréquence d'une onde sinusoïdale, appelée porteuse, au moyen d'un signal à transmettre, appelé signal modulant. Grâce à la modulation, on peut translater le contenu fréquentiel d'un signal modulant ; ce procédé permet de partager le spectre de fréquences entre plusieurs utilisateurs. Pour retrouver le signal modulant original, il faut procéder à une démodulation.	6, 27
MSC	Mobile Switching Center. Centre de commutation pour mobile. Cet équipement réalise la commutation des appels d'une ou plusieurs cellules.	215, 222
MSISDN	Mobile Subscriber ISDN. Numéro d'abonné au réseau GSM. Il est possible d'avoir plusieurs numéros (pour des services différents) au sein d'une seule carte SIM.	213
MSK	Minimum Shift Keying. Technique de modulation numérique consistant à effectuer une fonction XOR entre 2 bits successifs préalablement à une modulation de fréquence à 2 états.	219
multiplexage	Terme technique utilisé en télécommunications pour désigner un procédé qui consiste à partager des ressources entre plusieurs utilisateurs. Dans un autre contexte et lorsqu'on parle de signaux, il s'agit d'une manière de combiner plusieurs signaux.	27
norme	Document établi par consensus et approuvé par un organisme de normalisation reconnu (ISO, ETSI, ...). On distingue parfois une norme d'un standard, qui lui fait plutôt référence à une solution propriétaire s'étant imposée sur le marché.	5
NSS	Network Switching Center. Sous-système d'un réseau de téléphonie mobile. C'est la partie qui prend principalement en charge la commutation des appels, la signalisation et l'identification.	215, 223
octet	L'octet, ou byte en anglais, est une mesure de l'information binaire. 1 octet vaut 8 bits.	14
OSI	Open System Interconnection. Standard de référence d'interconnexion de réseaux développé par l'OSI. Ce système est différent du modèle Internet.	1, 8, 163

PAL	Phase Alternate Line. Format de télévision analogique utilisé dans la majorité des pays européens (hormis la France). La norme PAL définit le signal vidéo par balayage de 625 lignes télévisuelles par trame, à la fréquence d'image de 25 images par seconde.	14, 44
PCM	Pulse Code Modulation. Nom américain pour désigner la modulation par impulsions codées (MIC). Cette technique, utilisée principalement en téléphonie, convertit un signal analogique en un signal de téléphonie numérique à 64 [kb/s]. En toute rigueur, on ne devrait pas parler de modulation.	109, 223
PIN	Personal Identification Number. Code (mot de passe) nécessaire à chaque connexion d'un GSM au réseau.	213
PIRE	Puissance Isotrope Rayonnée Équivalente Il s'agit du produit de la puissance d'émission d'une antenne par son gain dans une direction spécifique.	196, 204
PLMN	Public Land Mobile Network. Il s'agit du réseau GSM, DCS ou PCS d'un opérateur dans un pays. Le "Network Color Code" identifie un PLMN dans un pays.	213
PM	Phase Modulation. Modulation de phase. Technique par laquelle on module la phase instantanée d'une porteuse au moyen du signal modulant à transmettre.	48
polarisation	Terme qui qualifie l'orientation du phaseur du champ électrique généré par une antenne. La polarisation est généralement linéaire (horizontale ou verticale) mais elle peut aussi être circulaire.	201
protocole	En télécommunications : ensemble de règles permettant le dialogue entre deux couches équivalentes d'entités différentes.	9
PUK	PIN Unblocking Key. Code nécessaire au déverrouillage d'une carte SIM.	213
QAM	Quadrature Amplitude Modulation. Modulation d'amplitude en quadrature. Cette technique de modulation (analogique ou numérique) consiste à moduler un signal pour deux signaux haute fréquence en quadrature de phase.	40, 161
RNIS	Réseau Numérique à Intégration de Services. Désigne le réseau téléphonique numérique. Au niveau du réseau, les signaux numériques utiles sont transmis à des multiples de 64[kb/s].	11, 215
roaming	Nom anglais pour désigner le fait qu'un utilisateur de GSM peut se déplacer d'une cellule à l'autre ou d'un réseau à un autre sans rupture de connexion. L'abonné qui utilise sa carte SIM est facturé par son opérateur. Cette opération est rendue possible grâce aux accords de roaming conclus entre les différents opérateurs.	213, 215
RTC	Réseau Téléphonique Commuté. Terme technique désignant le réseau téléphonique fixe.	208
SDH	Synchronous Digital Hierarchy. Hiérarchie normalisée de multiplexage pour signaux optiques. L'équivalent américain est SONET.	11
SIM	Subscriber Identity Module. Micro-processeur implanté dans une carte. Par extension, on parle de la carte SIM. Elle est insérée dans un GSM pour réaliser une série de fonctions et contenir une mini-base de données.	213
SMS	Short Message Service. Système permettant l'envoi de messages comprenant au plus 160 caractères (de 7 bits), soit 140 bytes, à un téléphone GSM.	223
SNR	Signal to Noise Ratio. Rapport signal à bruit. Il s'agit du quotient de la puissance du signal utile à celle du bruit considéré. Le SNR est un des moyens utilisé pour caractériser la qualité d'un canal de transmission.	19, 114
SSB	Single Sideband Modulation. Technique de modulation d'amplitude constituée seulement de la bande latérale supérieure ou inférieure du signal modulant. Ce type de modulation est couramment utilisé en radiodiffusion de signaux de télévision.	40, 44

TA	Timing Advance. Le décalage temporel utilisé pour prévenir les collisions entre messages envoyés par différents mobiles vers une station de base dans un réseau GSM.	222
TCP	Transmission Control Protocol. Protocole de transport utilisé pour communiquer sur Internet. TCP se charge de numérotter les paquets et gère les acquits (ou accusés de réception).	9
TDMA	Time Division Multiple Access. Technique de répartition de ressources par multiplexage temporel. Cette technique prévoit un mécanisme d'accès.	211, 218
TMSI	Temporary Mobile Subscriber Identity. Numéro attribué temporairement à un utilisateur GSM en fonction de sa localisation.	213
trame	En traitement d'images, la trame est la grille d'échantillonnage. On considère généralement la trame carrée mais la trame peut aussi être rectangulaire ou hexagonale. Dans le cas du format entrelacé, la trame désigne une image ne contenant que les lignes paires ou impaires de l'image. En télécommunications, trame désigne un ensemble d'informations numériques temporelles constituant un tout.	219
transcodage	Aussi appelé transrating. Il s'agit d'un procédé de changement du débit d'un signal comprimé.	224
TRAU	Transcoding Rate and Adaptation Unit. Unité de transcodage utilisée dans les réseaux GSM pour convertir un signal de 13 [kb/s] en un signal de 64 [kb/s] et vice-versa.	224
UDP	User Datagram Protocol. Protocole de transport utilisé pour la transmission d'informations sur Internet lorsqu'il s'agit de minimiser le délai de transmission. UDP numérote les paquets mais, contrairement à TCP, il ne vérifie pas si les paquets sont bien arrivés à destination.	9
UMTS	Universal Mobile Telecommunications System. Nom du standard de téléphonie mobile de troisième génération pour l'Europe.	207, 211
V90	Norme pour la transmission par modem dans le réseau téléphonique. Le débit peut atteindre 56 [kb/s].	162
VLR	Visitor Location Register. Registre local d'une zone comprenant plusieurs cellules d'un réseau GSM. Ce registre contient l'identité des utilisateurs présents dans la zone.	215, 216
VSB	Vestigial Sideband. Terme désignant la modulation d'amplitude à bande latérale résiduelle. Il s'agit d'une technique de modulation d'amplitude qui conserve la totalité d'une des bandes latérales du signal modulant ; de l'autre bande, il ne reste qu'un résidu. Si ce résidu est nul, on parle de modulation à bande latérale unique (SSB).	40, 46
X25	Série de protocoles, définis par l'ITU, destinés à la transmission de données. Leur utilisation est aujourd'hui largement supplantée par l'utilisation des protocoles à technologie Internet.	215
XOR	eXclusive OR. Fonction logique du OU exclusif. Le résultat de la fonction vaut 0 si les deux états sont à 0 ou à 1. Il vaut 1 dans les autres cas.	219

Bibliographie

- [1] A. Ambardar. Digital signal processing. Brooks/Cole Publishing Company, second edition, 1999. 56
- [2] J. Bic, D. Duponteuil, and J. Imbeaux. Elements of digital communications. John Wiley & Sons, 1991. 7
- [3] M. Capinski and E. Kopp. Measure, integral and probability. Springer, 1999. 70
- [4] M. Charbit. Éléments de théorie du signal : les signaux aléatoires. Ellipses, 1990. 91
- [5] L. Couch. Digital and analog communication systems. Prentice Hall, sixth edition, 2001.
- [6] P. Delogne. Introduction aux télécommunications, Notes de cours, 1988.
- [7] P. Delogne. Signaux de télécommunications, Notes de cours, 1990.
- [8] I. Gradshteyn, I. Ryzhik, and A. Jeffrey. Table of integrals, series and products. Academic Press, fifth edition, 1994. 51
- [9] M. Hall, L. Barclay, and M. Hewit. Propagation of radiowaves. The Institute of Electrical Engineers, 1996.
- [10] S. Haykin. Communication system. John Wiley & Sons, 1994.
- [11] G. Heine. GSM networks : protocols, terminology, and implementation. Artech House, 1999. 213
- [12] J. Kraus and D. Fleisch. Electromagnetics with applications. McGraw-Hill, fifth edition, 1999.
- [13] X. Lagrange, P. Godlewski, and S. Tabbane. Réseaux GSM-DCS. Hermès, troisième édition, 1997. 221
- [14] C. Paul and S. Nasar. Introduction to electromagnetic fields. McGraw-Hill, 1987. 179, 187
- [15] E. Rivier. Transmission numérique multimédia. Eyrolles, 1998.
- [16] E. Roubine. Distributions–Signal. Éditions Eyrolles, 1990. 99
- [17] B. Sklar. Digital communications : fundamental and applications. Prentice Hall, 1988. 98, 176
- [18] J. Tisal. Le réseau GSM. L'évolution GPRS : une étape vers UMTS. Dunod, troisième édition, 1999. 223
- [19] F. Vanden Berghen. Télégraphie : une histoire branchée. Crédit Communal, 1998. 2
- [20] B. Walke. Mobile Radio Networks : networking, protocols and traffic performance. John Wiley & Sons, 2002. 225
- [21] J. Walrand and P. Varaiya. Communication networks : a first course. McGraw-Hill, 1998.
- [22] R. Yates and D. Goodman. Probability and stochastic processes : a friendly introduction for electrical and computer engineers. John Wiley & Sons, 1999. 77, 78