

Réseaux Multimédia

© 2002 Damien Magoni

Toutes les illustrations

© 2001 Pearson Education Limited – Fred Halsall

Contenu

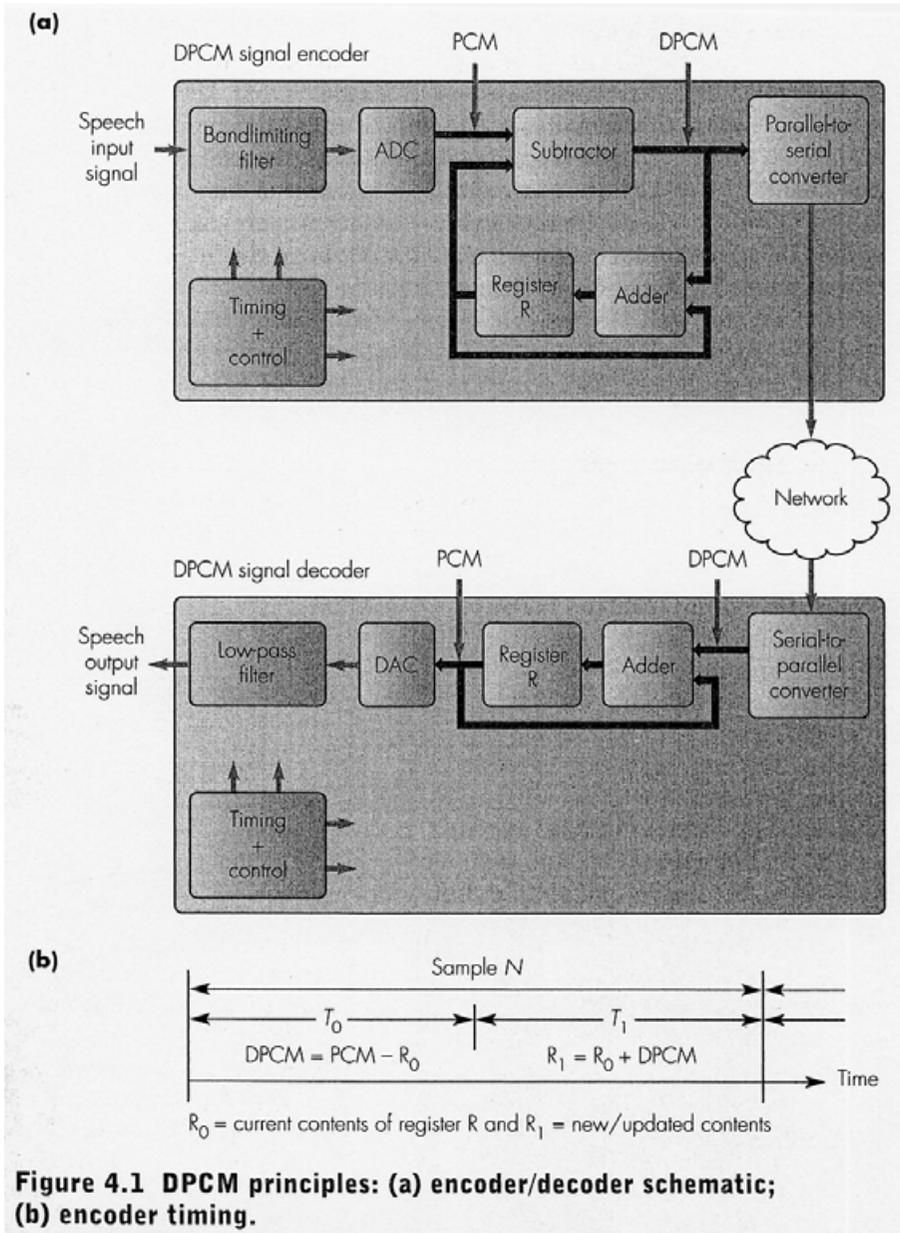
- Représentation des informations multimédia
 - Numérisation
 - Structure d'un encodeur
 - Structure d'un décodeur
 - Représentation du texte, des images et de l'audio/vidéo
- Compression du texte et des images
- Compression audio et vidéo
- Standards pour les communications multimédia

Partie 3

Compression audio et vidéo

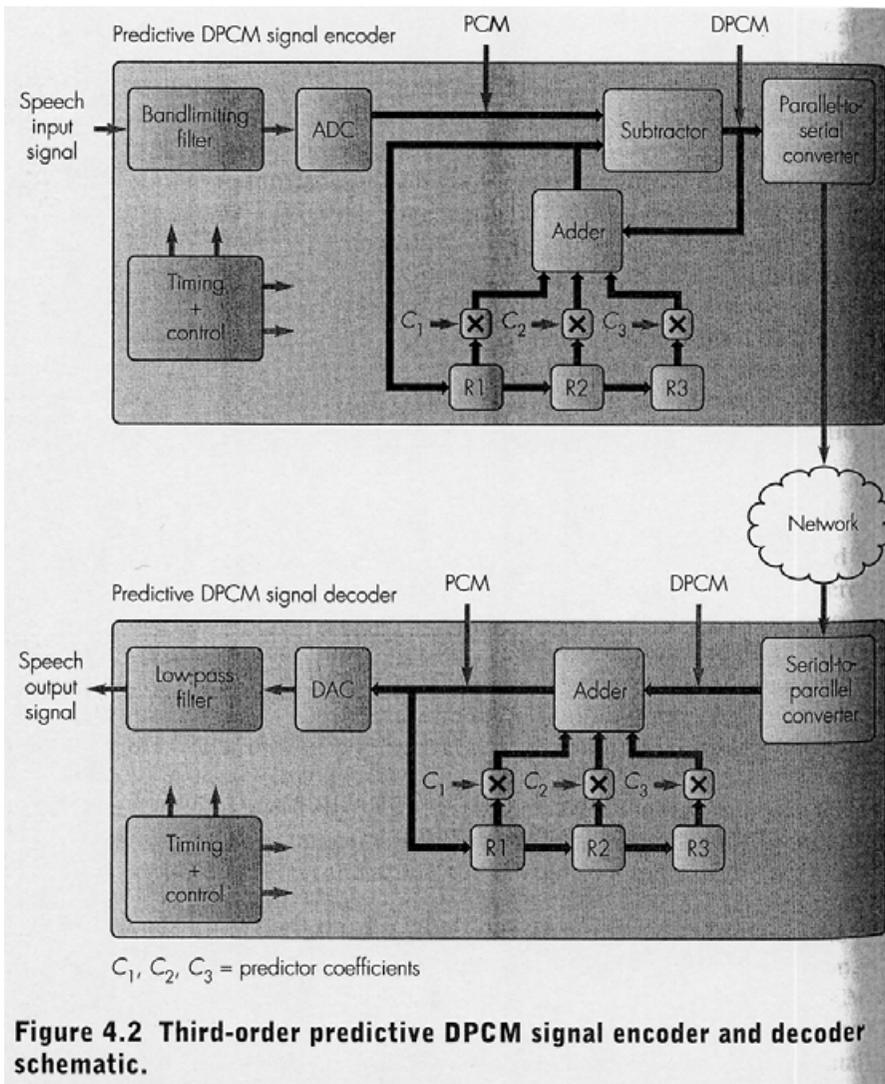
Codage DPCM

- Differential Pulse Code Modulation (DPCM)
- Utilisé pour coder des signaux audio car les variations en amplitude entre 2 échantillons sont plus faibles que les valeurs absolues
- Le gain de codage est typiquement de 1 bit
- L'échantillon précédent est stocké dans un registre
- La précision de chaque valeur différentielle (résidu) est déterminée par la précision de la valeur stockée dans le registre (les erreurs de quantification peuvent se cumuler)



Codage DPCM prédictif

- Pour avoir une meilleure précision sur la valeur du signal précédent, on utilise plusieurs valeurs précédentes pondérées par des coefficients
- Les pondérations sont déterminées par les coefficients prédictifs
- Cette technique permet de coder un signal sur 6 bits avec la même qualité qu'en PCM (au lieu de 12 ou 16 bits)



Codage ADPCM

- Adaptative Differential Pulse Code Modulation (ADPCM)
- Principe : faire varier le nombre de bits utilisés pour coder le résidu (plus la variation est faible, moins on utilise de bits)
- Norme ITU-T G.721
 - DPCM avec prédicteur du 8ème ordre
 - Codage sur 6 ou 5 bits selon le résidu
- Norme ITU-T G.722
 - Codage en sous-bande
 - La BP est étendue de 50Hz à 7kHz
 - Signal filtré en deux sous-bandes codées indépendamment avec des débits différents

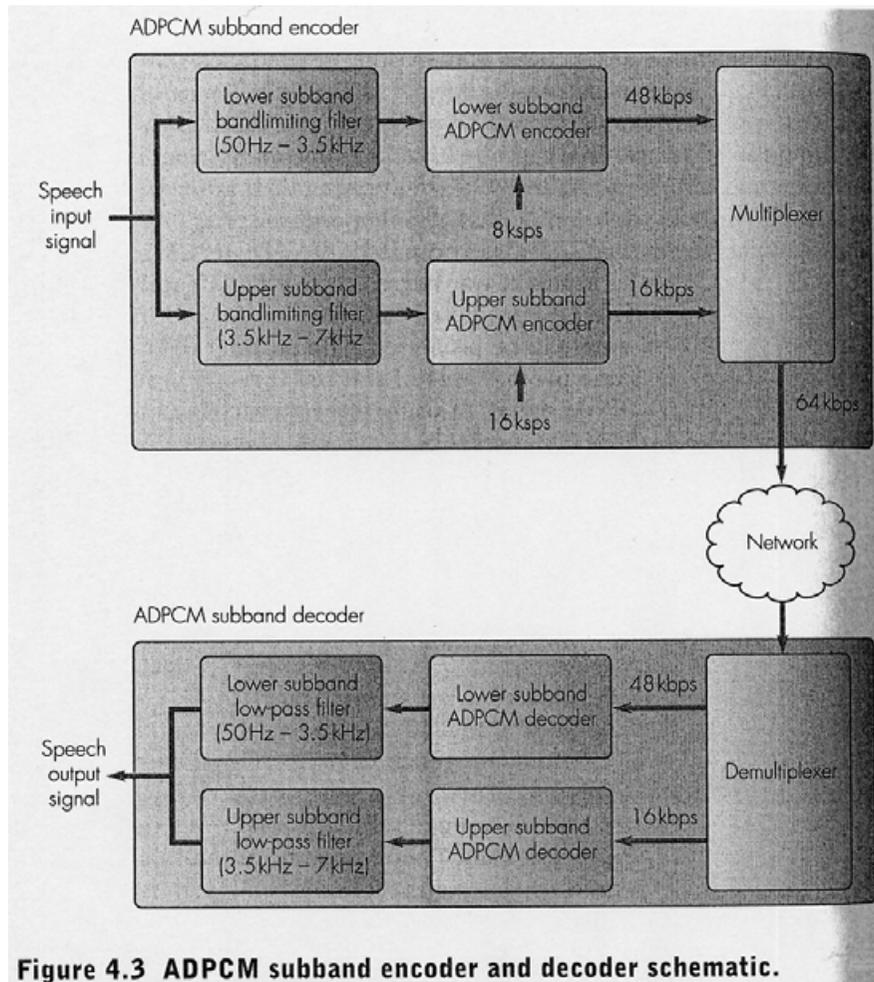


Figure 4.3 ADPCM subband encoder and decoder schematic.

Codage prédictif linéaire

- Linear Predictive Coding (LPC)
- Le signal est analysé et traité avant d'être codé
- Codage lié à la perception d'un signal vocal par les oreilles :
 - Pitch : l'oreille est plus sensible aux fréquences 2-5kHz que les plages supérieures ou inférieures
 - Period : durée du signal
 - Loudness : quantité d'énergie présente dans le signal
- Les sons de la voix sont synthétisés suivant deux types : *voiced* (e.g. m,v,l), *unvoiced* (e.g. f, s)
- La voix est échantillonnée puis quantifiée. Un bloc d'échantillons (segment) est analysé pour définir les paramètres de la parole

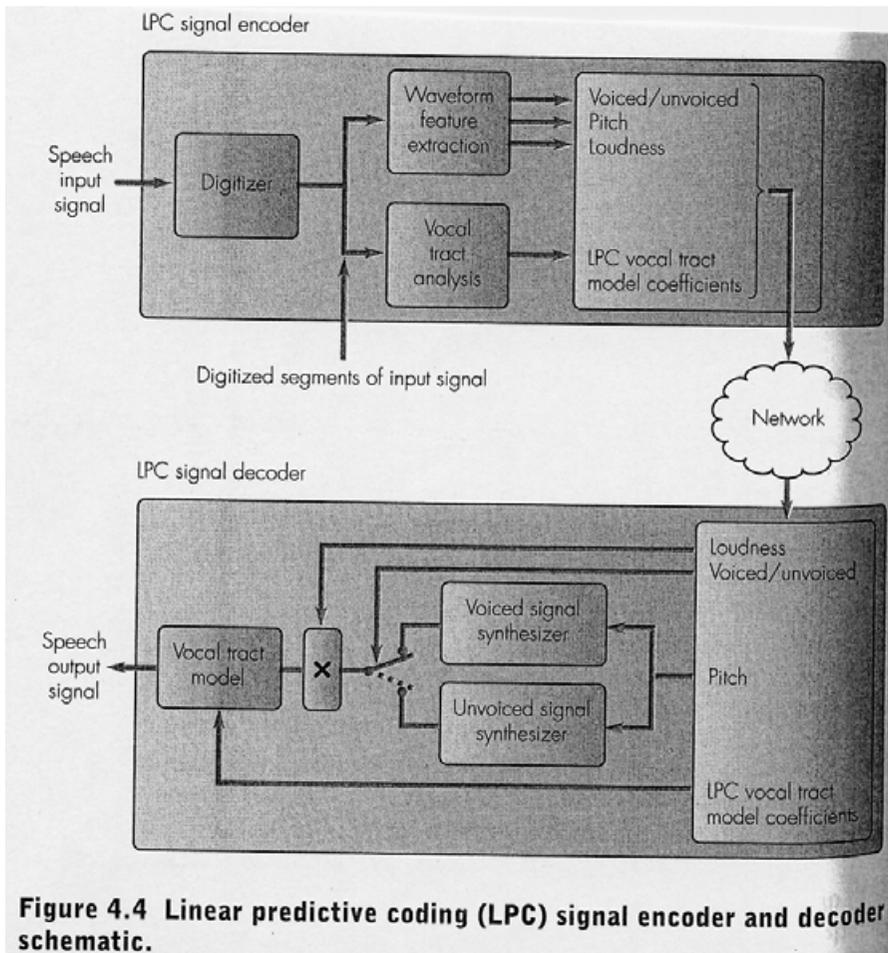


Figure 4.4 Linear predictive coding (LPC) signal encoder and decoder schematic.

Codage CELP

- Code-Excited Linear Prediction (CELP) model
- Au lieu de traiter chaque segment indépendamment, seul un ensemble limité de segments est utilisé
- Ces segments sont nommés modèles de forme d'onde et l'ensemble forme un dictionnaire des modèles
- Normes ITU-T G.728, G.729, G.729(A) et G.723.1
- Tous ces codeurs ont une latence (*delay*) de traitement nécessaire à la reconstruction du signal ainsi qu'une latence algorithmique liée au *buffering* des échantillons

Caractéristiques des codeurs CELP

Standard	Bit rate	Total delay	Application
G.728	16 kbps	0.625 ms	Low bit rate telephony
G.729	8 kbps	25 ms	Cellular telephony
G.729(A)	8 kbps	25 ms	Digital simultaneous voice and data
G.723.1	5.3/6.3 kbps	67.5 ms	Video and IP telephony

Codage perceptif

- LPC et CELP utilisés principalement pour la téléphonie donc pour la compression de la parole
- Il existe des codeurs perceptifs plus généraux pour la compression de l'audio
- Ils utilisent un modèle psycho-acoustique qui utilise les limitations de l'oreille humaine :
 - La sensibilité de l'oreille dans sa plage de 15Hz à 20kHz est non linéaire, sa dynamique est de l'ordre de 96dB
 - Le masquage fréquentiel : une fréquence de forte amplitude masque les fréquences proches d'amplitude moindre (fig. 4.6 pour 1, 4 et 8kHz)
 - Le masquage temporel : l'oreille requiert un temps minimum pour entendre un son faible après un son fort

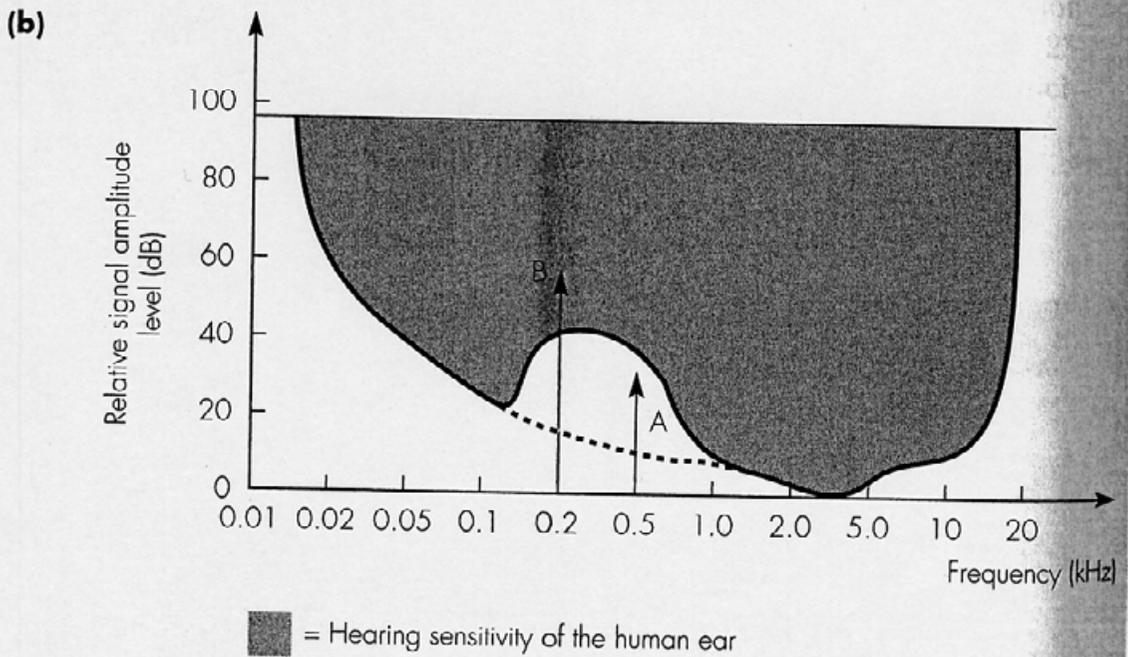
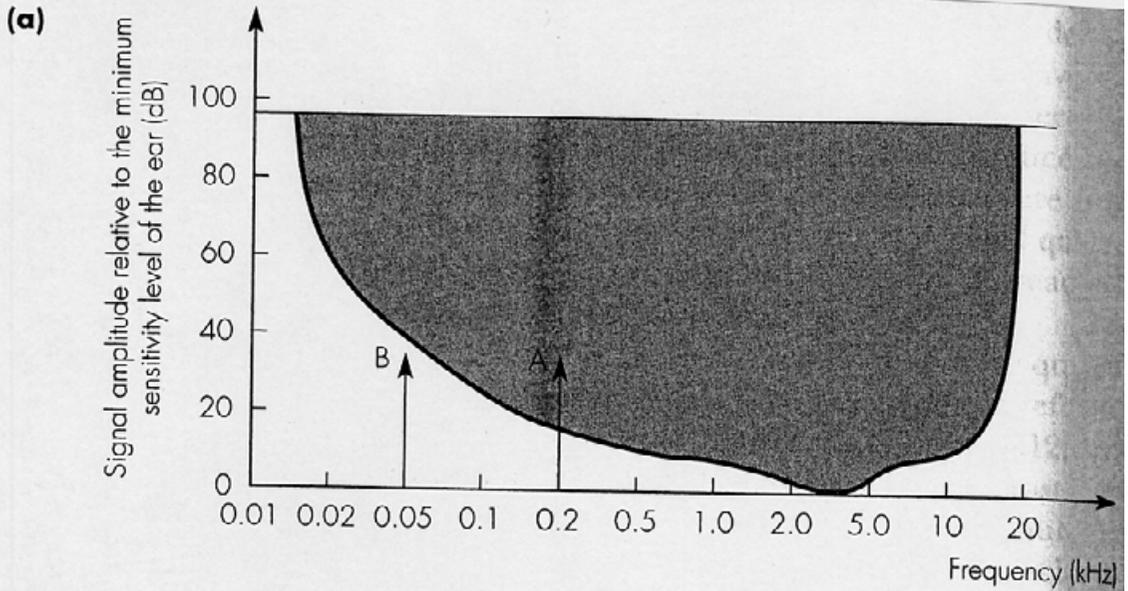


Figure 4.5 Perceptual properties of the human ear: (a) sensitivity as a function of frequency; (b) frequency masking.

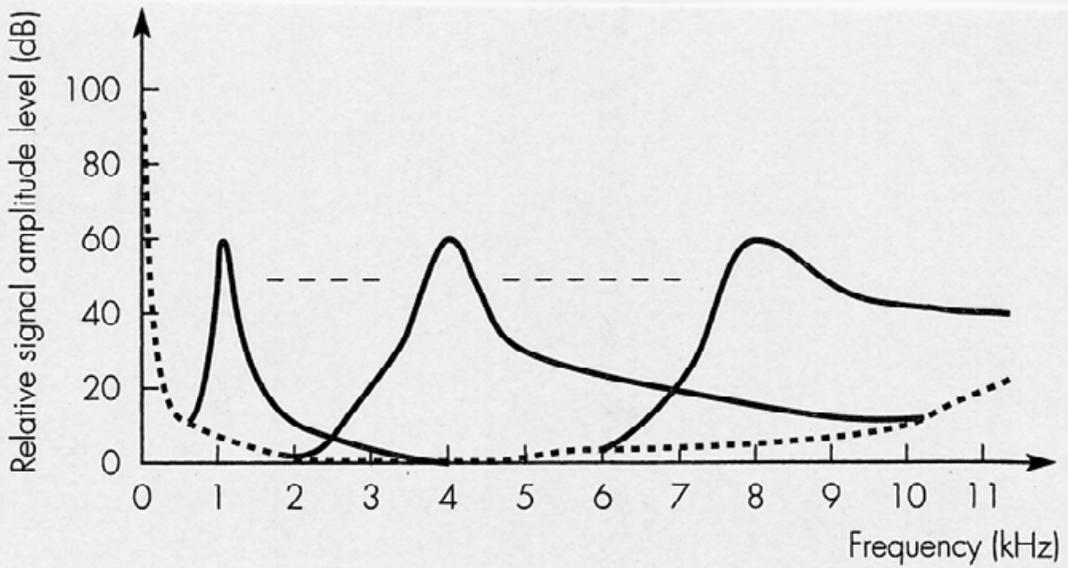


Figure 4.6 Variation with frequency of effect of frequency masking.

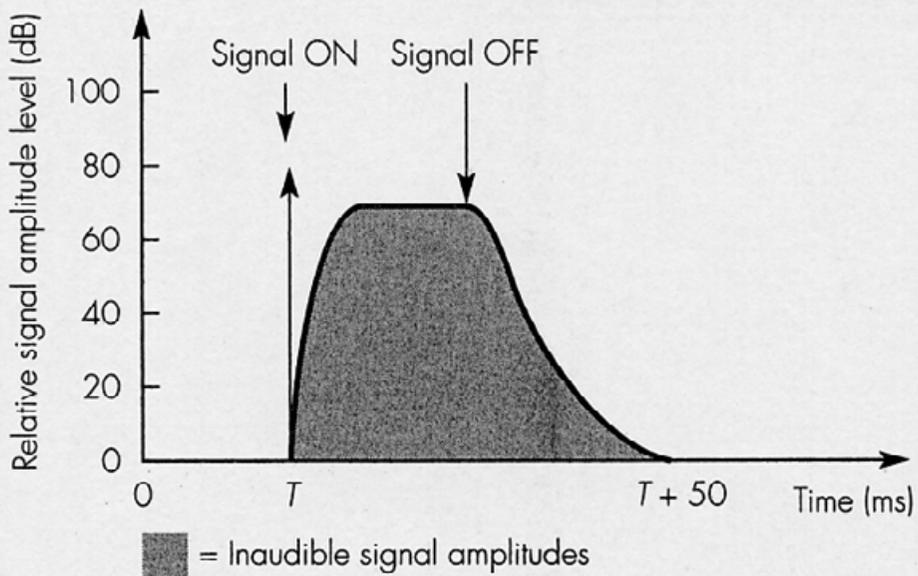
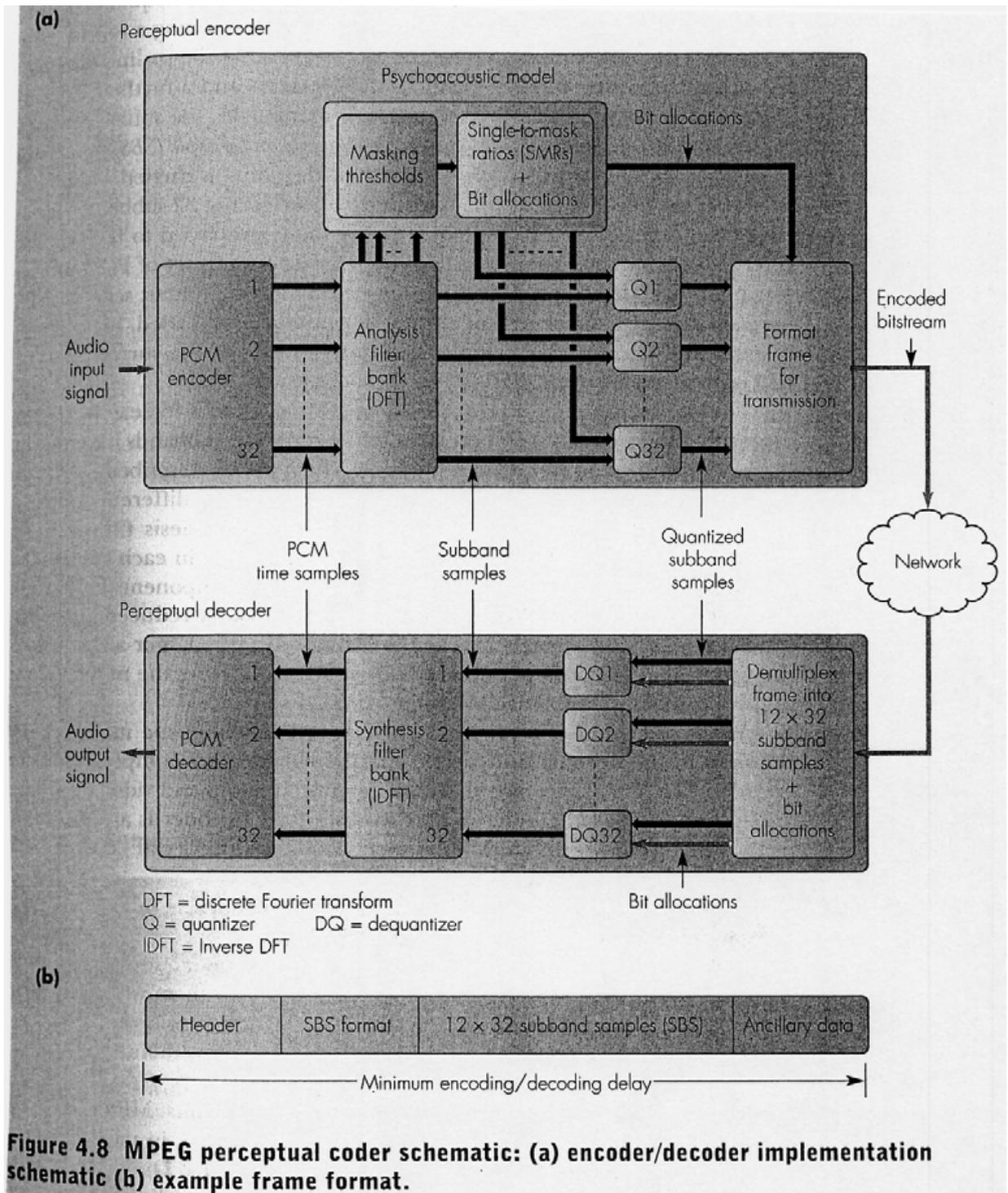


Figure 4.7 Temporal masking caused by a loud signal.

Codeurs audio MPEG

- Ils utilisent le codage perceptif
- Le signal est numérisé en PCM avec une fréquence d'échantillonnage et un nombre de bits par échantillon définis par l'application
- La BP disponible est divisée en un nombre de sous-bandes de même largeur
- Une banque de filtres fait correspondre chaque ensemble de 32 échantillons PCM successifs en un équivalent de 32 composantes fréquentielles (une par sous-bande) par Transformée de Fourier Discrète
- Chaque échantillon sous-bande mesure donc l'amplitude d'une de ces 32 fréquences durant un intervalle temporel de 32 échantillons PCM
- Dans un codeur basique, la durée de chaque segment est égal au temps de stockage de 12 ensembles successifs de 32 échantillons de sous-bandes soit 384 échantillons PCM
- Un ensemble de facteurs d'échelle est utilisé pour déterminer la précision de quantification (et donc l'allocation du nombre de bits) afin de coder avec plus de précision les régions où l'oreille est plus sensible



Codeurs audio MPEG (suite)

- Le modèle psycho-acoustique n'est pas requis dans le décodeur et celui-ci est donc plus simple que le codeur
- Norme ISO 11172-3 comprenant 3 niveaux (*layers*)

Layer	Application	Compressed bit rate	Quality	Input-to-output delay
1	Digital audio tape	32-448 kbps	Hi-fi quality at 192 kbps per channel	20 ms
2	Digital audio/video broadcasting	32-192 kbps	Near CD quality at 128 kbps per channel	40 ms
3	CD quality audio over low bit rate channels	64 kbps	CD quality at 64 kbps per channel	60 ms

Codeurs audio Dolby

- Les codeurs MPEG incluent l'information sur l'allocation des bits dans les échantillons (car elle est définie dynamiquement) ce qui consomme de la bande passante
- L'idée est de fixer le schéma d'allocation des bits pour chaque sous-bande, celui-ci doit alors être connu du décodeur
- Dolby Acoustic Coder (AC-1) :
 - Définit 40 sous-bandes à 32 ksps, un peu plus à 44.1 et 48 ksps (typiquement 512 kbps pour de la stéréo)
 - Mode d'allocation des bits adaptatif vers l'avant (*forward adaptive bit allocation mode*)

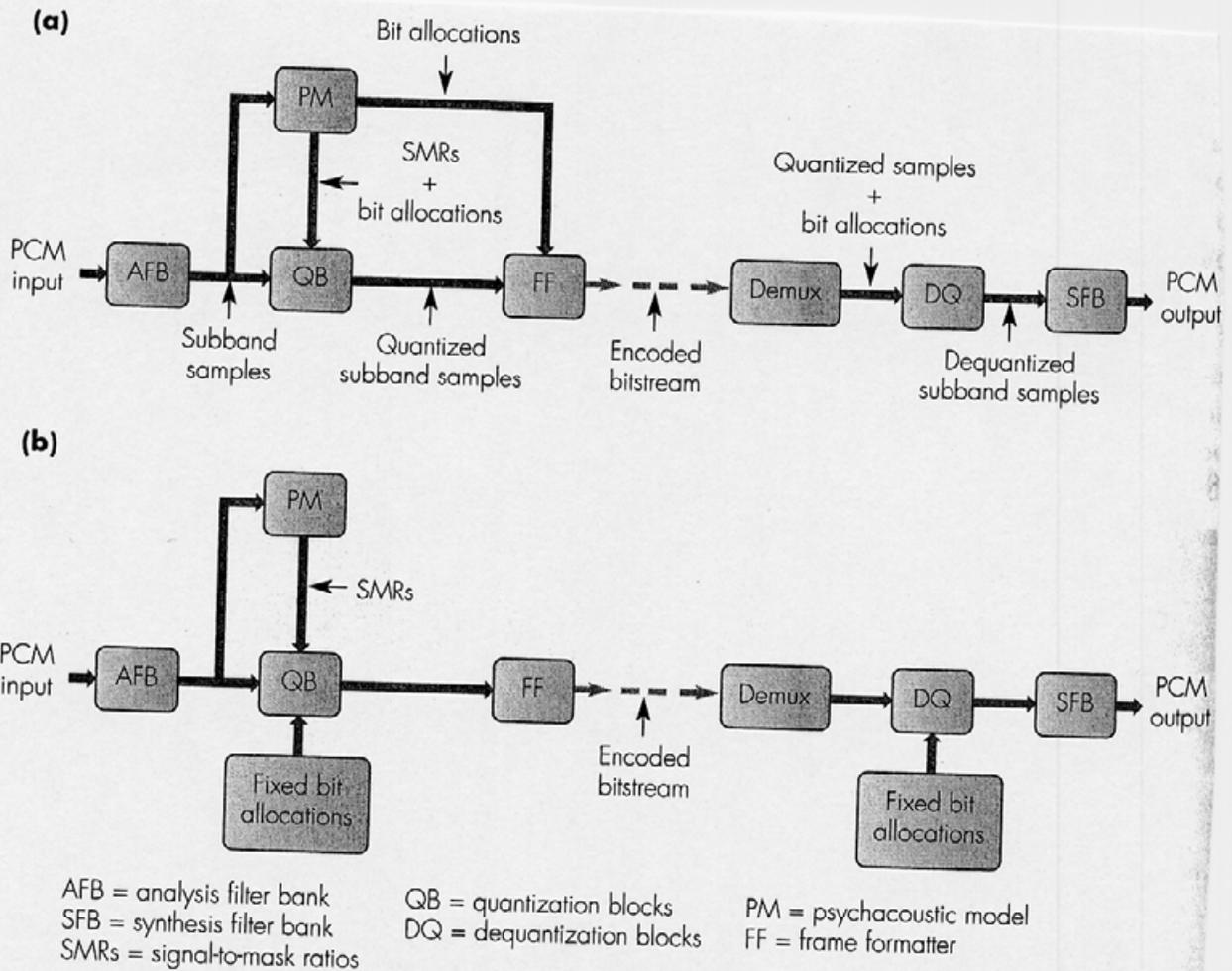


Figure 4.9 Perceptual coder schematics: (a) forward adaptive bit allocation (MPEG); (b) fixed bit allocation (Dolby AC-1).

Codeurs audio Dolby (suite)

- Dolby AC-2 :
 - Permet une allocation de bits adaptative tout en minimisant l'*overhead* dans le flux codé
 - Le décodeur doit posséder une copie du modèle psycho-acoustique
 - Chaque *frame* contient, en plus des échantillons, les coefficients fréquentiels présents dans le segment de forme d'onde : enveloppe spectrale codée
 - Mode d'allocation des bits adaptatif vers l'arrière (*backward adaptive bit allocation*)
 - Audio compressé de qualité Hi-fi à 256 kbps
 - Utilisé dans les cartes PC
- Dolby AC-3 :
 - Utilise un mode d'allocation des bits hybride (avant/arrière)
 - Comme AC-2 + un modèle psycho-acoustique supplémentaire utilisé pour calculer la différence entre l'allocation de bits de PM_B et de PM_F (par *forward adaptive bit allocation*)
 - Utilisé pour l'HDTV (débit stéréo typique de 192 kbps)

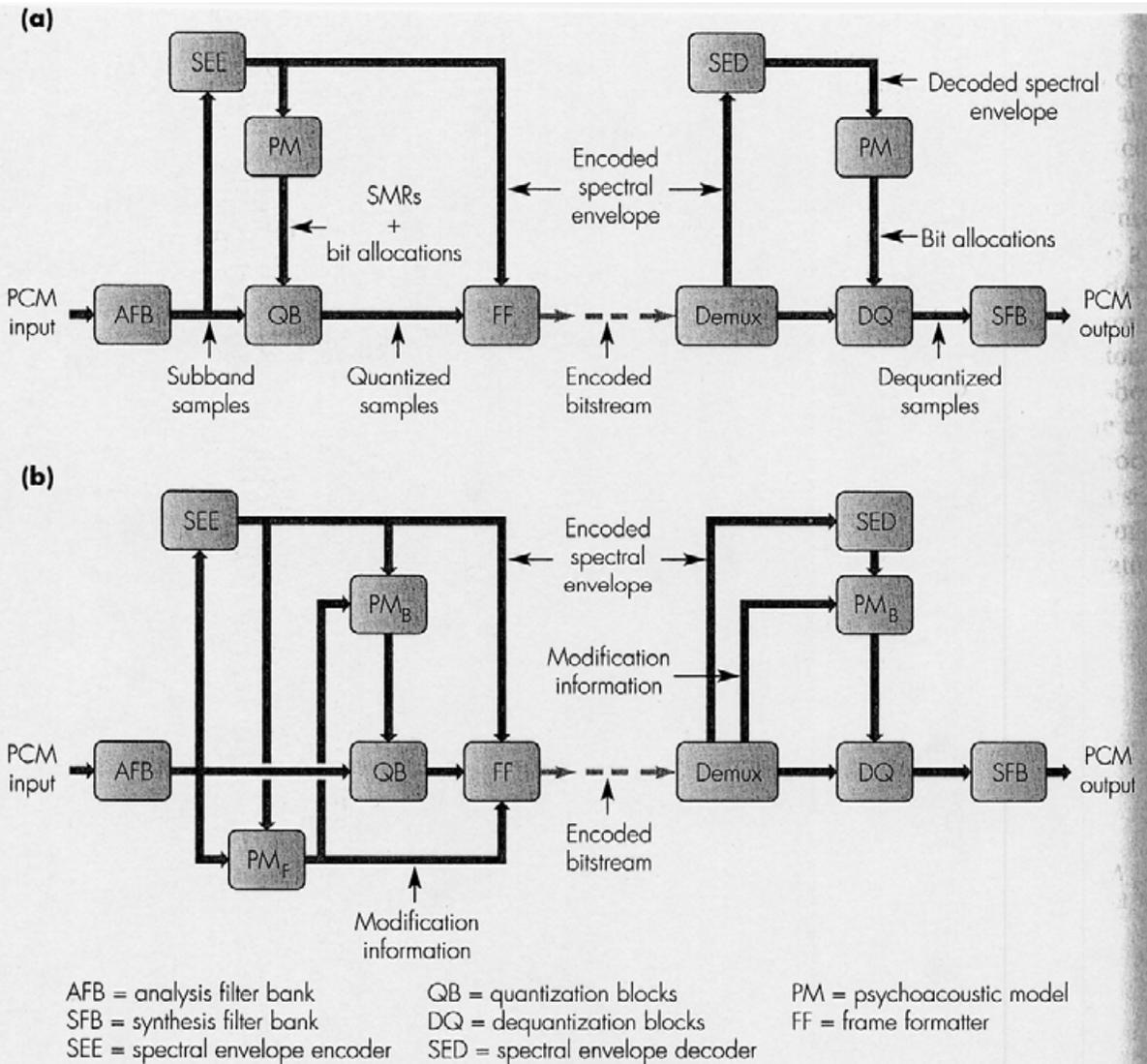
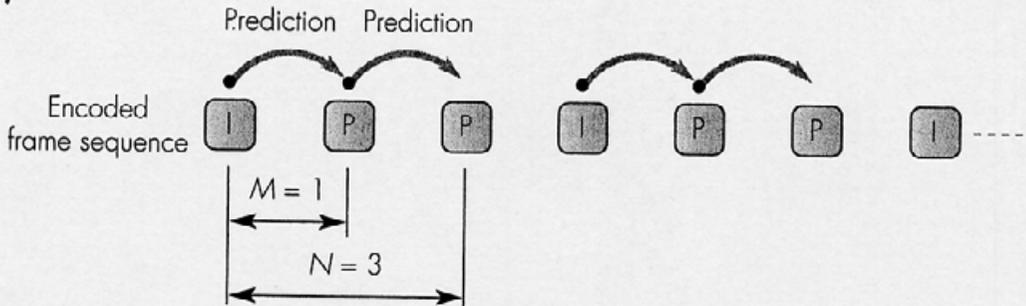


Figure 4.10 Perceptual coder schematic: (a) backward adaptive bit allocation (Dolby AC-2); (b) hybrid backward/forward adaptive bit allocation (Dolby AC-s).

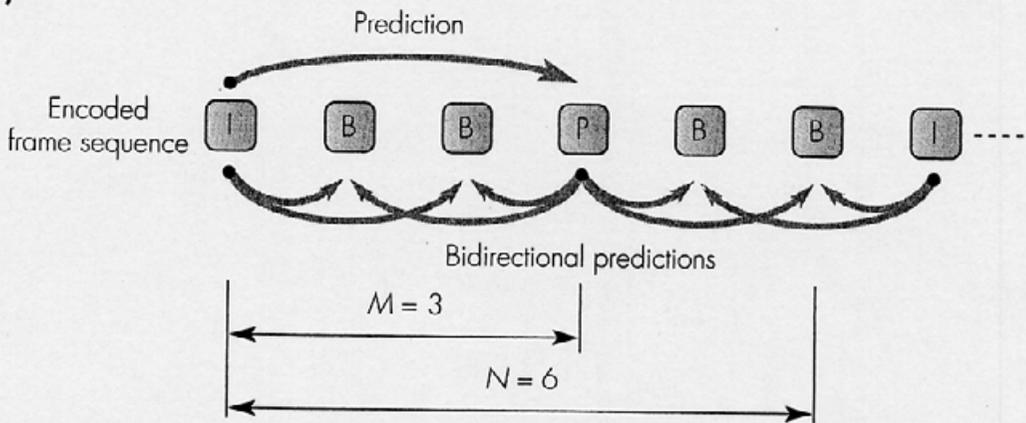
Compression vidéo

- MJPEG : JPEG appliqué à chaque image (ratio faible de 10:1 à 20:1)
- Exploiter la redondance temporelle : codage prédictif
- 3 types :
 - Images I indépendantes
 - Images P prédictives (a)
 - Images B prédictives bidirectionnelles (b)
- Nb d'images N entre deux images I est un Group Of Pictures et varie entre 3 et 12

(a)



(b)



M = prediction span N = group of pictures (GOP) span

(c)

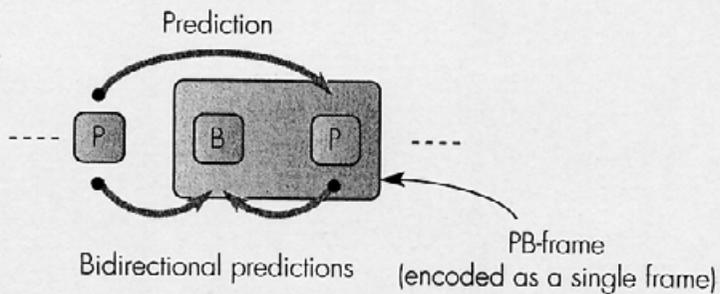
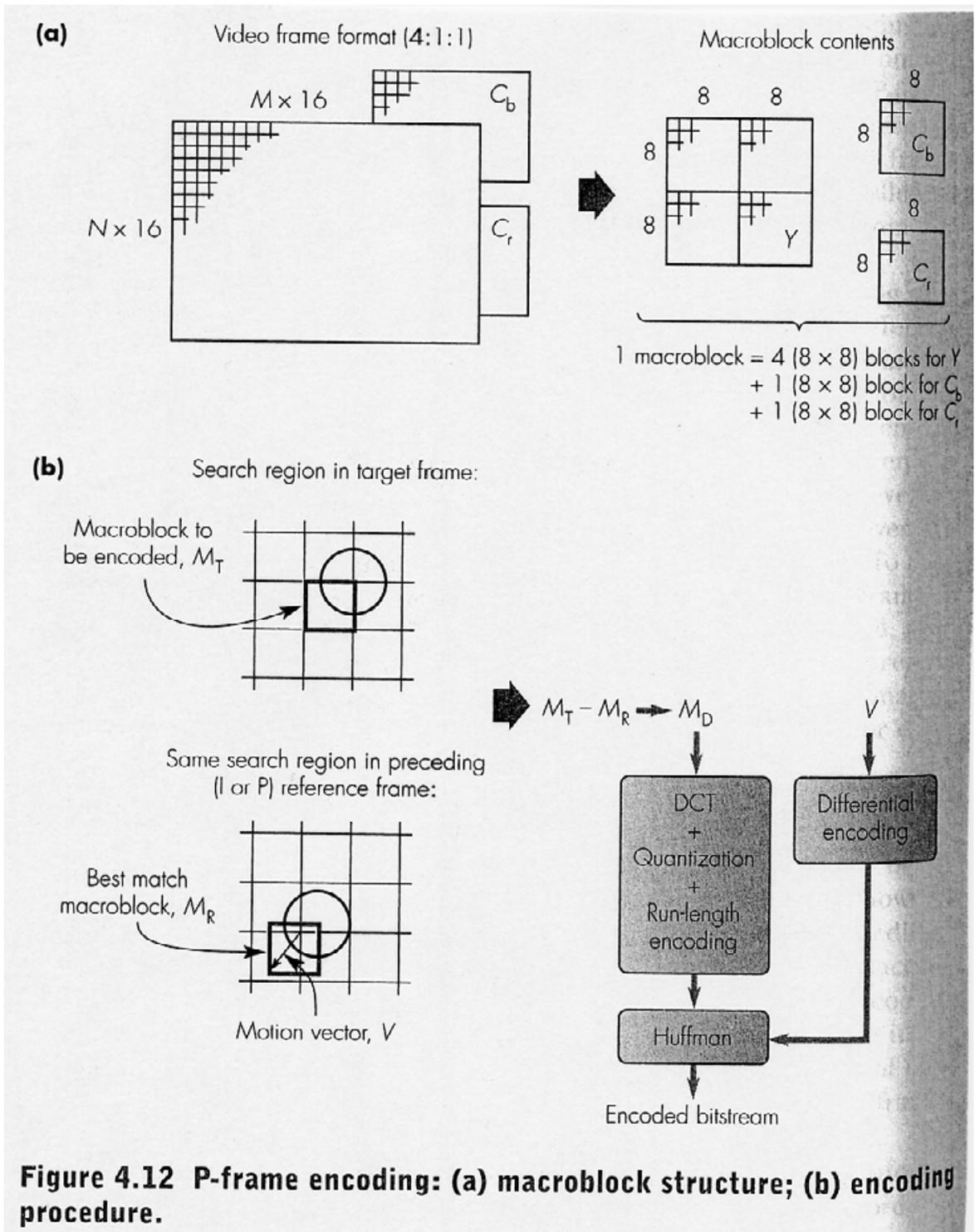


Figure 4.11 Example frame sequences with: (a) I- and P-frames only; (b) I-, P- and B-frames; (c) PB-frames.

Encodage P

- Chaque image divisée en macroblocs de 16x16 pixels
- Chaque macrobloc a une adresse
- Si un macrobloc X dans P est le même que le macrobloc X dans I alors on stocke juste son adresse
- Sinon on le cherche dans I et on stocke :
 - le *motion vector* pour l'atteindre (le décalage peut être au macrobloc près ou au pixel près)
 - l'erreur de prédiction : trois matrices Y , C_b et C_r contenant les écarts entre le bloc cible et le bloc de référence
- Sinon on l'encode en entier



Encodage B

- Un mouvement est estimé par rapport à l'image I ou P précédente et l'image I ou P suivante
- Le vecteur de mouvement et les matrices de différences sont calculés dans le cas précédant, le cas suivant et le cas moyenne précédant/suivant : le jeu ayant le moins d'écart avec la cible est choisi puis encodé comme dans une image P

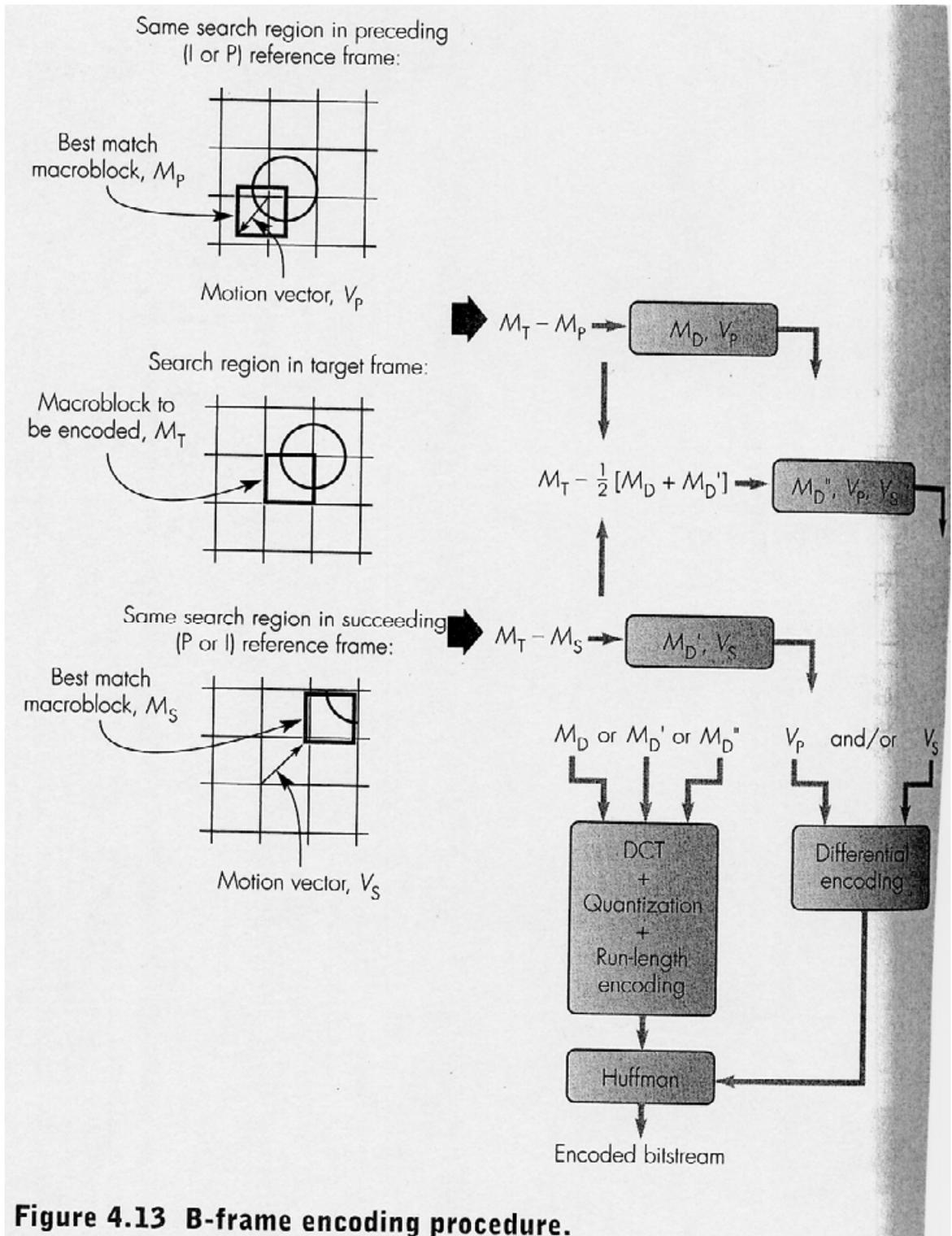


Figure 4.13 B-frame encoding procedure.

Implémentation

- Schéma d'implémentation du codage des images I (a), des images P (b) et des images B (c)
- Le format d'un flot de bits d'un macrobloc est décrit en (d)
- Le taux de compression :
 - Pour images I varie de 10:1 à 20:1
 - Pour images P varie de 20:1 à 30:1
 - Pour images B varie de 30:1 à 50:1

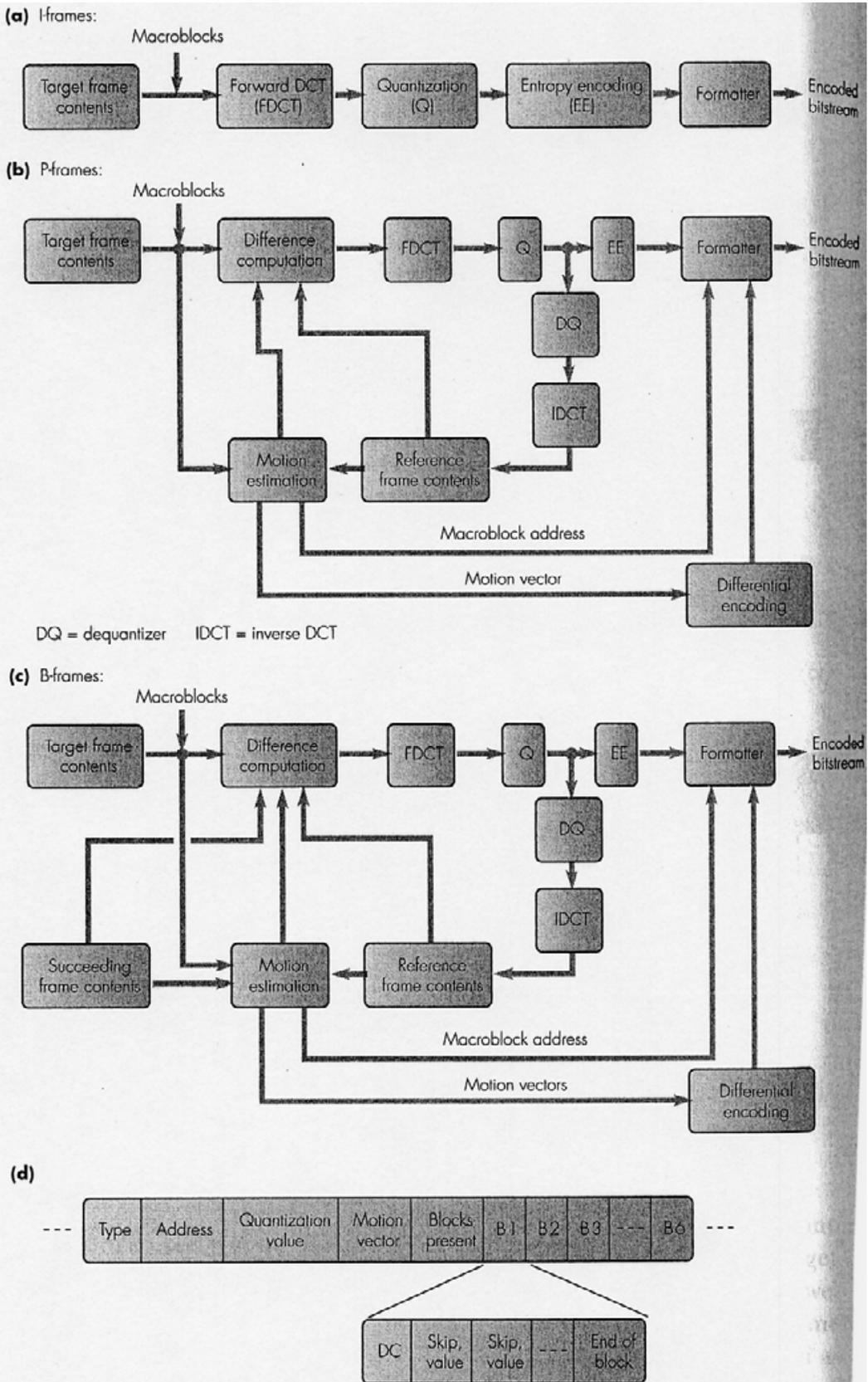


Figure 4.14 Implementation schematics: (a) I-frames; (b) P-frames; (c) B-frames; (d) example macroblock encoded bitstream format.

H.261

- Défini par ITU-T pour video stream sur ISDN
- Débit = $p \times 64\text{kbps}$ (p de 1 à 30)
- Format de numérisation CIF (videoconférence) ou QCIF (videophone)
- Résolution :
 - CIF $Y=352 \times 288$ ($C_b = C_r = 176 \times 144$)
 - QCIF $Y=176 \times 144$ ($C_b = C_r = 88 \times 72$)
- Balayage progressif non entrelacé de 30 fps pour CIF et 15 ou 7,5 fps pour QCIF
- Utilise des Groups Of (macro) Blocks (GOB) de 11x3 blocs
- FIFO pour produire un débit constant (CBR)

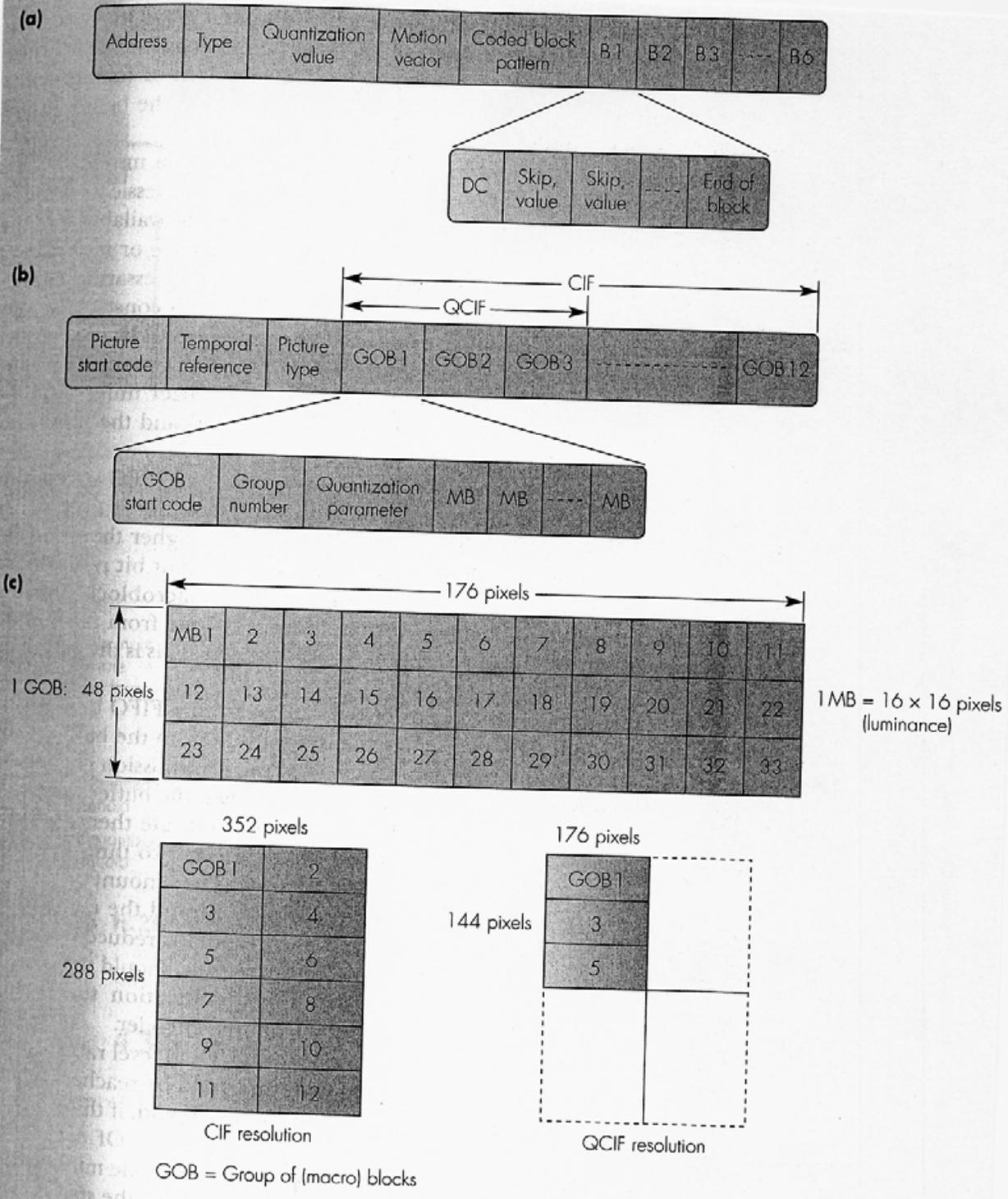
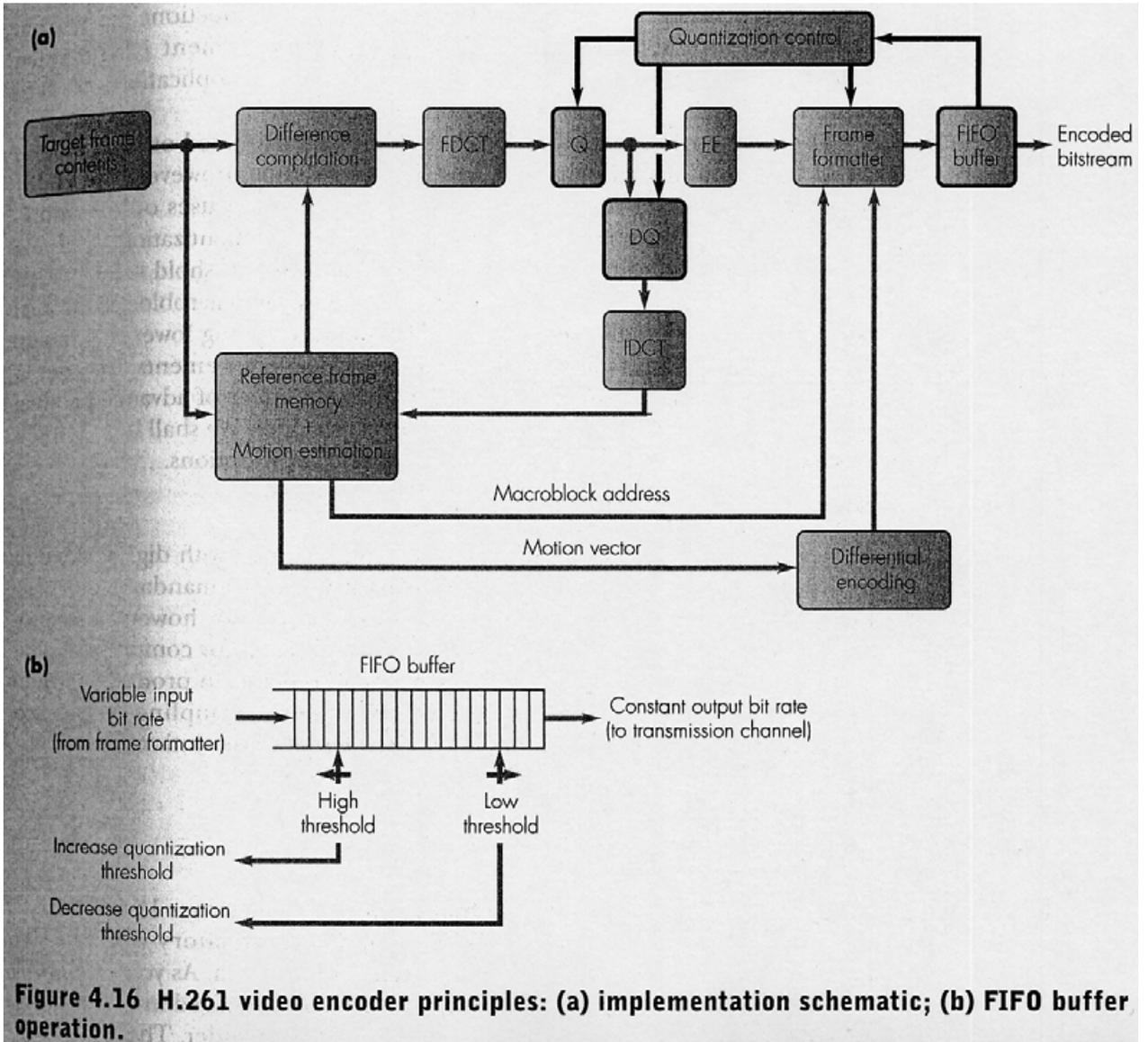


Figure 4.15 H.261 encoding formats: (a) macroblock format; (b) frame/picture format; (c) GOB structure.



H.263

- Défini par ITU-T pour video sur réseau téléphonique commuté public (RTC) et sans fil (DECT)
- Débit < 64kbps
- Format de numérisation QCIF ou sub-QCIF
- Résolution :
 - QCIF Y=176x144 ($C_b = C_r = 88x72$)
 - S-QCIF Y=128x96 ($C_b = C_r = 68x64$)
- Peu d'images I et utilisation d'images PB
- Balayage progressif non entrelacé de 15 ou 7,5 fps pour QCIF
- Si un bloc est erroné, le décodeur cherche un marqueur de synchro dans le GOB suivant

Minimiser les erreurs

- Trois techniques :
 - *Error tracking* (4.17) : un message NAK avec n° d'image et de GOB pour renvoi de l'image I
 - *Independent segment decoding* (4.18) : chaque GOB dans une image est traité comme une sous video indépendante des autres GOB de l'image (l'estimation de mouvement est donc faite sur un GOB et non une image)
 - *Reference picture selection* (4.19) : comme l'error tracking avec 2 modes :
 - NACK : utilise le GOB de l'image I pour renvoi
 - ACK : seules les images ACK sont utilisées en référence (marche bien quand RTT est petit)

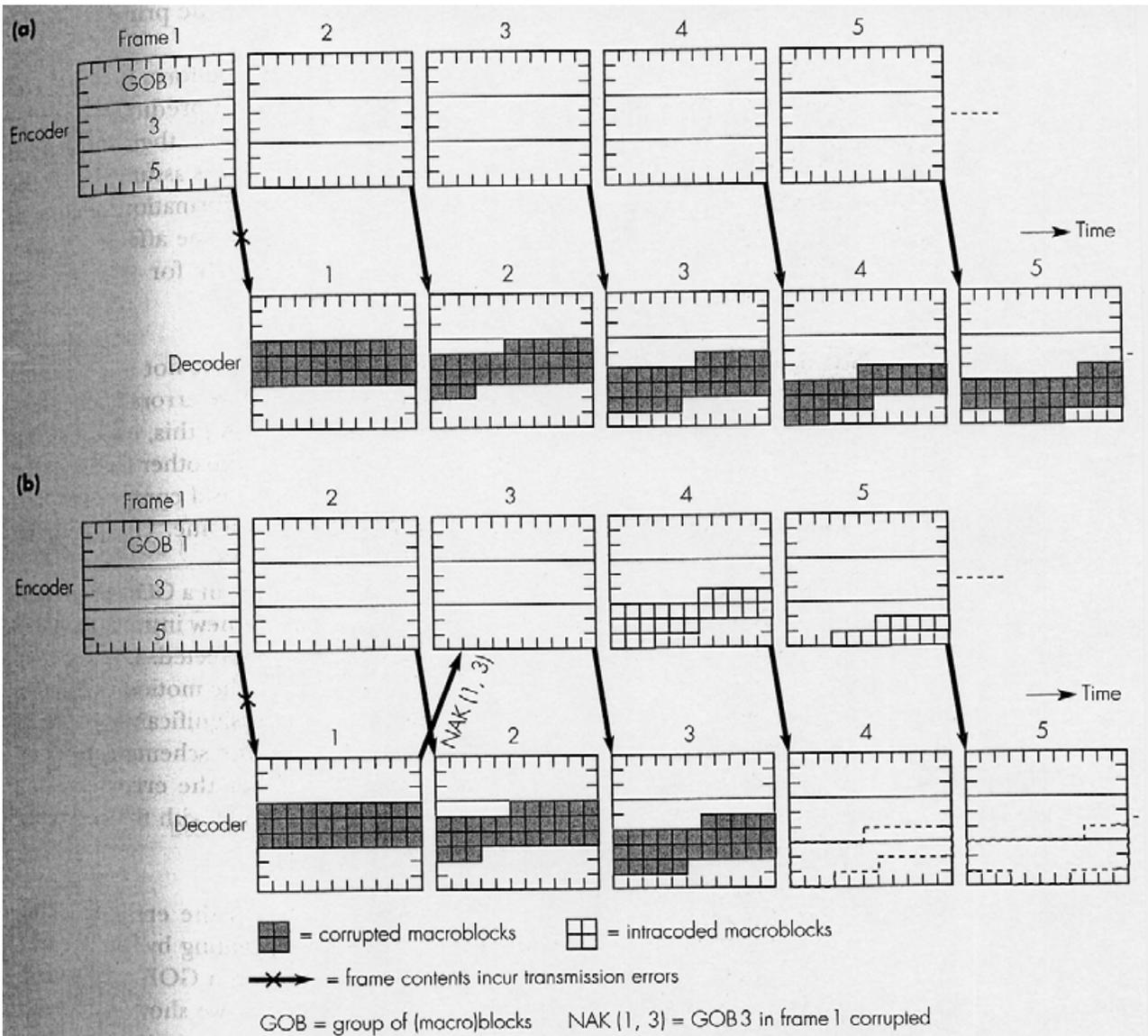
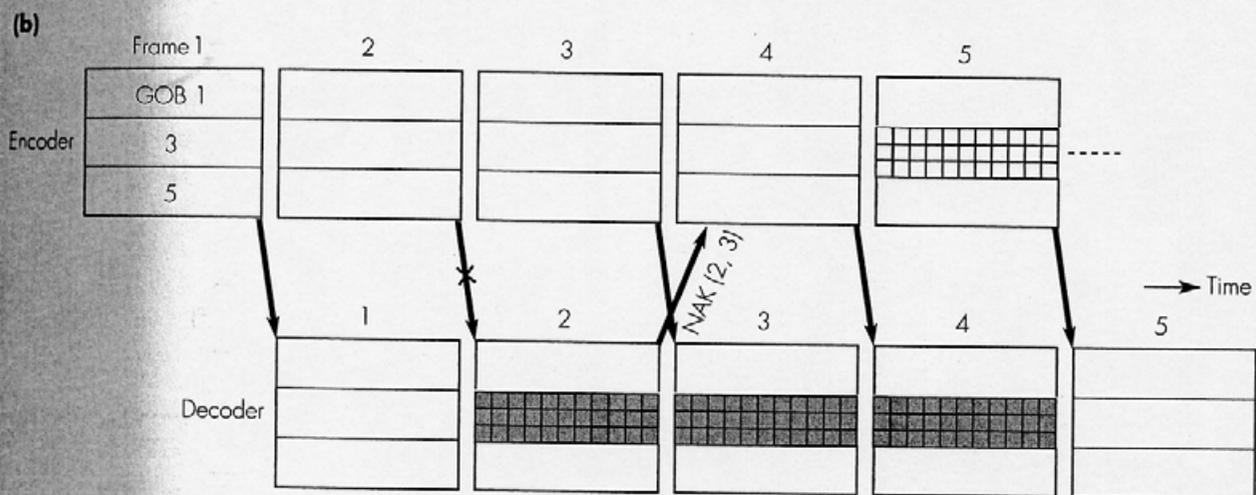
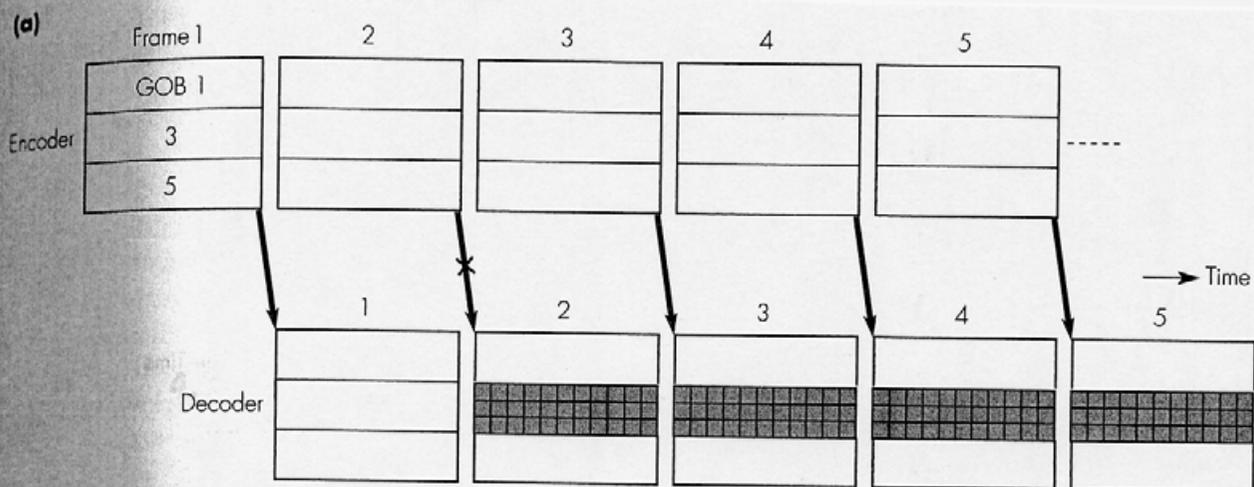


Figure 4.17 H.263 error tracking scheme: (a) example error propagation; (b) same example with error tracking applied.



■ = corrupted GOB

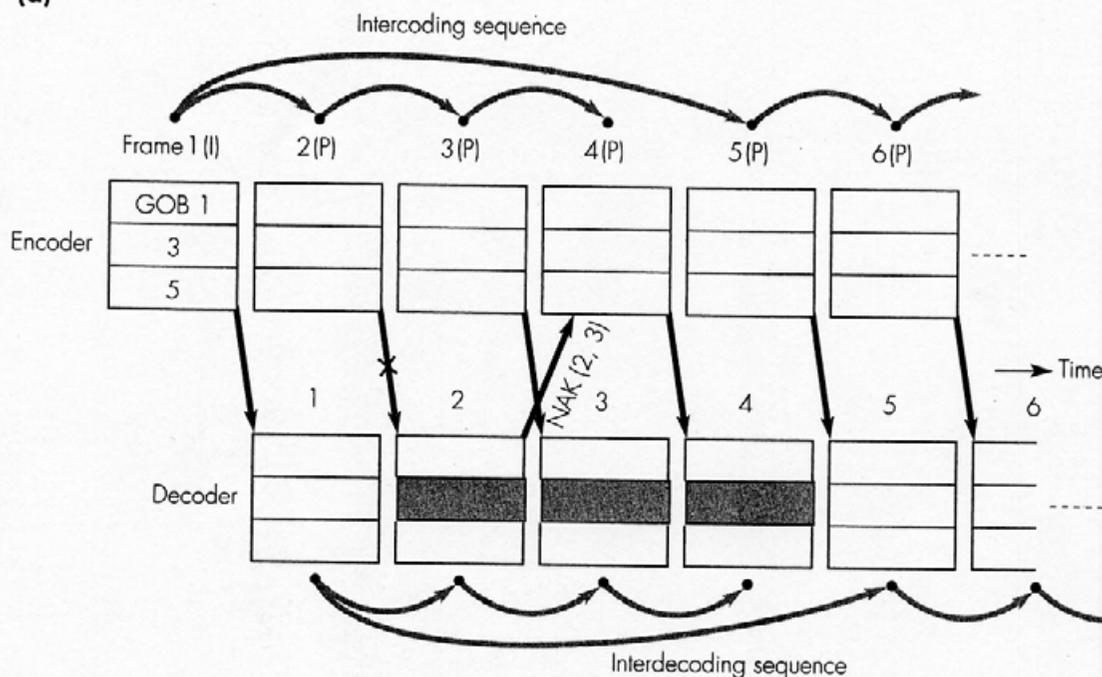
□ = intracoded GOB

✕ = frame contents incur transmission errors

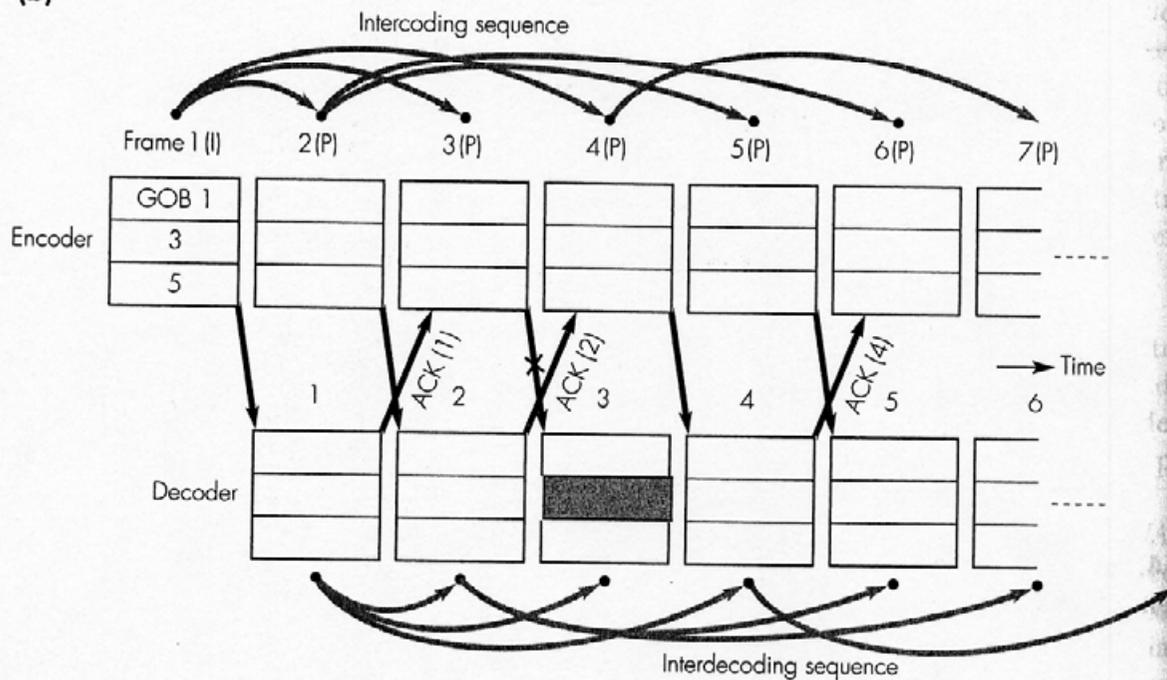
GOB = group of (macro)blocks NAK (2, 3) = GOB 3 in frame 2 corrupted

Figure 4.18 Independent segment decoding: (a) effect of a GOB being corrupted; (b) when used with error tracking.

(a)



(b)



I = intracoded frame

P = predicted/intercoded frame

Figure 4.19 Reference picture selection with independent segment decoding: (a) NAK mode; (b) ACK mode.

MPEG-1

- Format SIF
- Utilise des macroblocs 16x16 :
 - NTSC $Y=352 \times 240$ ($C_b=C_r=176 \times 120$)
 - PAL $Y=352 \times 288$ ($C_b = C_r = 176 \times 144$)
- Balayage progressif interlacé de 30 fps pour NTSC et 25 fps pour PAL
- *Timestamp* dans les images
- Nb blocs entre 2 *stamps* = *slice*
- Une video = séquence divisée en Group Of Pictures (GOP)
- Chaque image comprend jusqu'à N *slices*

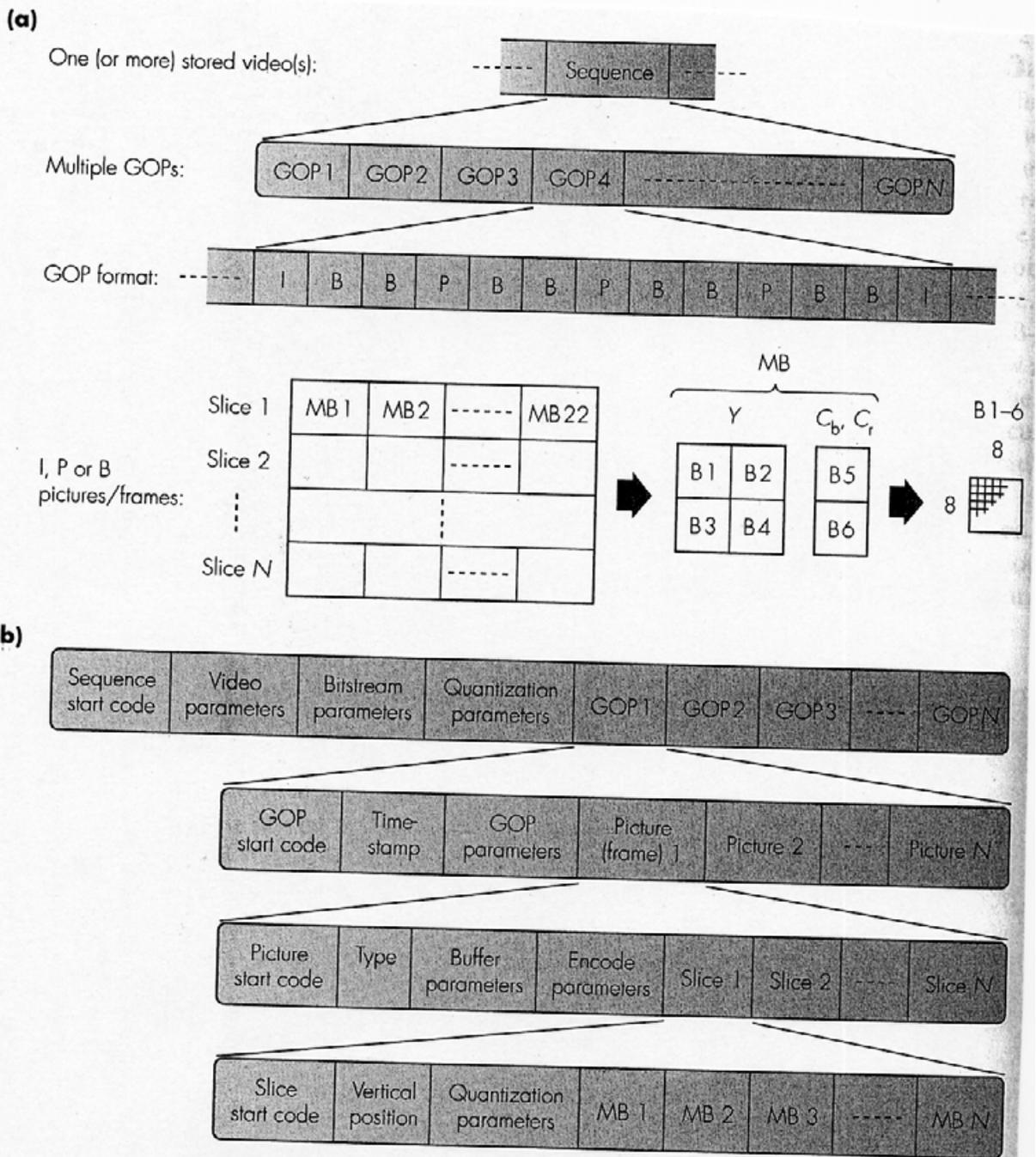


Figure 4.21 MPEG-1 video bitstream structure: (a) composition; (b) format.

MPEG-2

- 4 niveaux de qualité (low, main, high 1440, high) et 5 profils
- MP@ML :
 - 720x480/576 à 30/25 Hz
 - 4 à 15Mbps
- HDTV :
 - US : ATV = 16/9, 1920x1152, MP@HL, AC-3
 - Europe : DVB = 4/3, 1440x1152, SSP@H1440, MP-2
 - Asia : MUSE = 16/9, 1920x1035, MP@HL, AC-3
 - Prod : ITU-R = 16/9, 1920x1152, interlacé 4:2:0

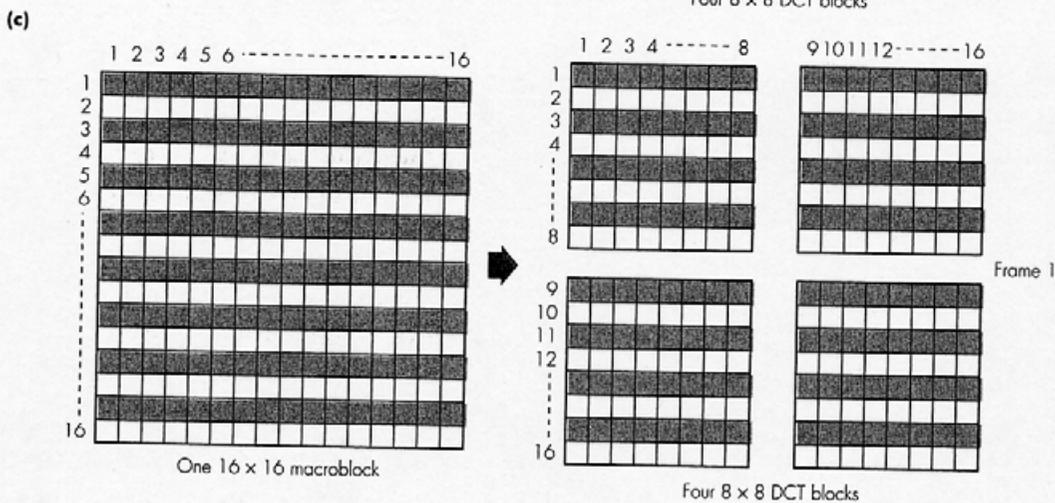
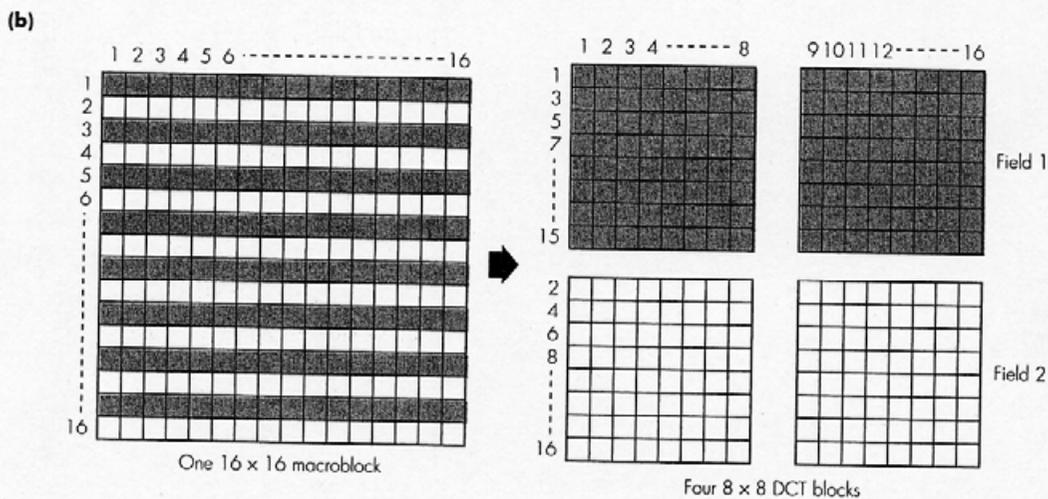
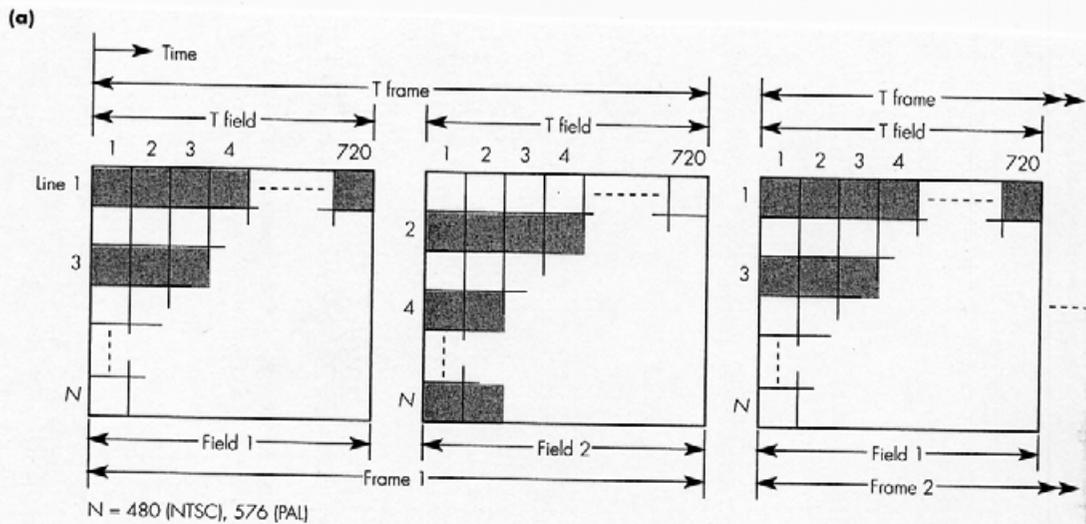
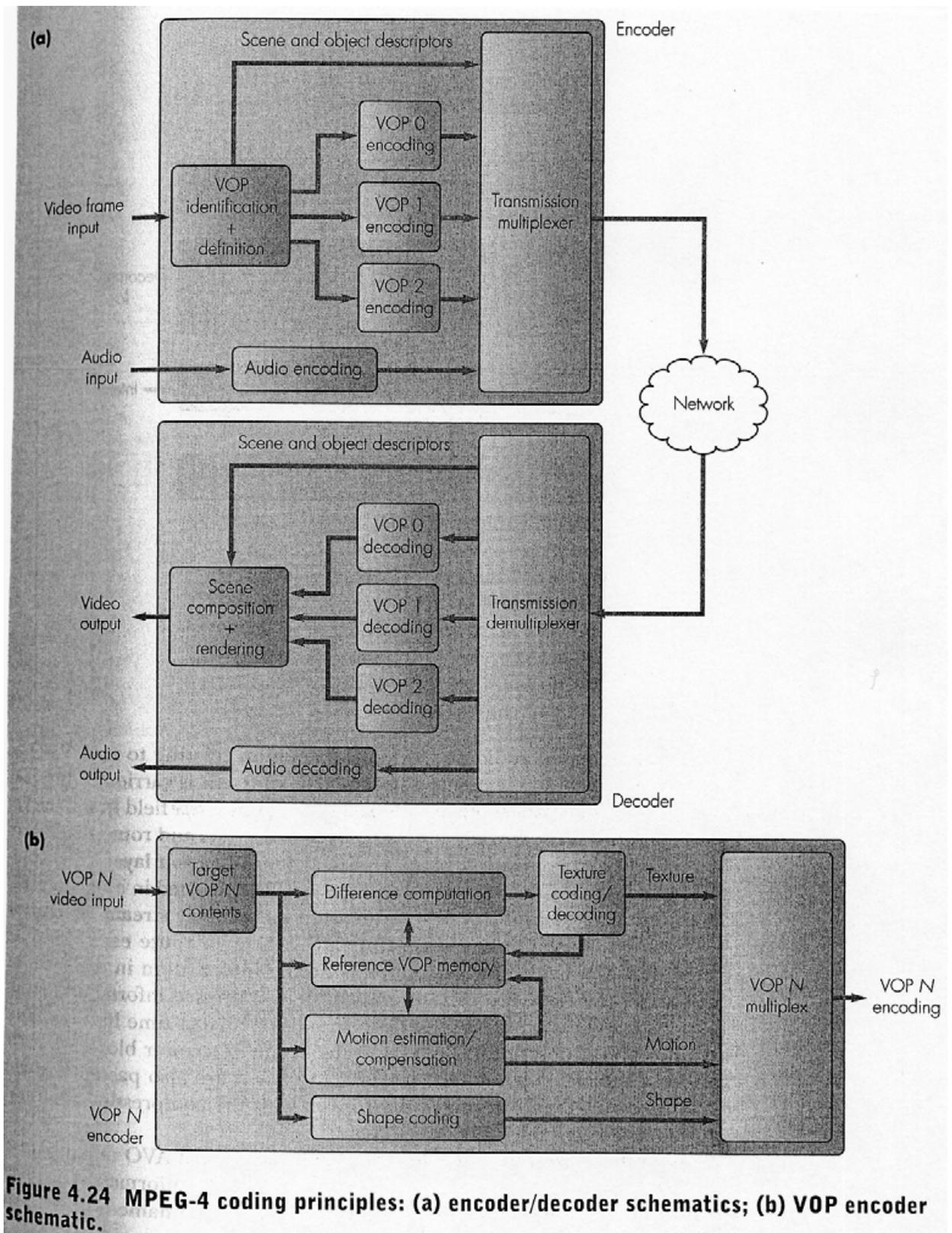


Figure 4.22 MPEG-2 DCT block derivation with I-frames: (a) effect of interlaced scanning; (b) field mode; (c) frame mode.

MPEG-4

- Audio/Video et multimedia interactif, fonctionnalités basées sur le contenu (objets A/V)
- Les objets AVO sont définis en langage BIFS, possèdent un descripteur, puis sont compressés
- Chaque image est segmentée en un nb de plans d'objets video (VOP) correspondant chacun à un AVO
- Les données sont envoyées dans un *transport stream* TS contenant un multiplexage de *packetized elementary stream* PES



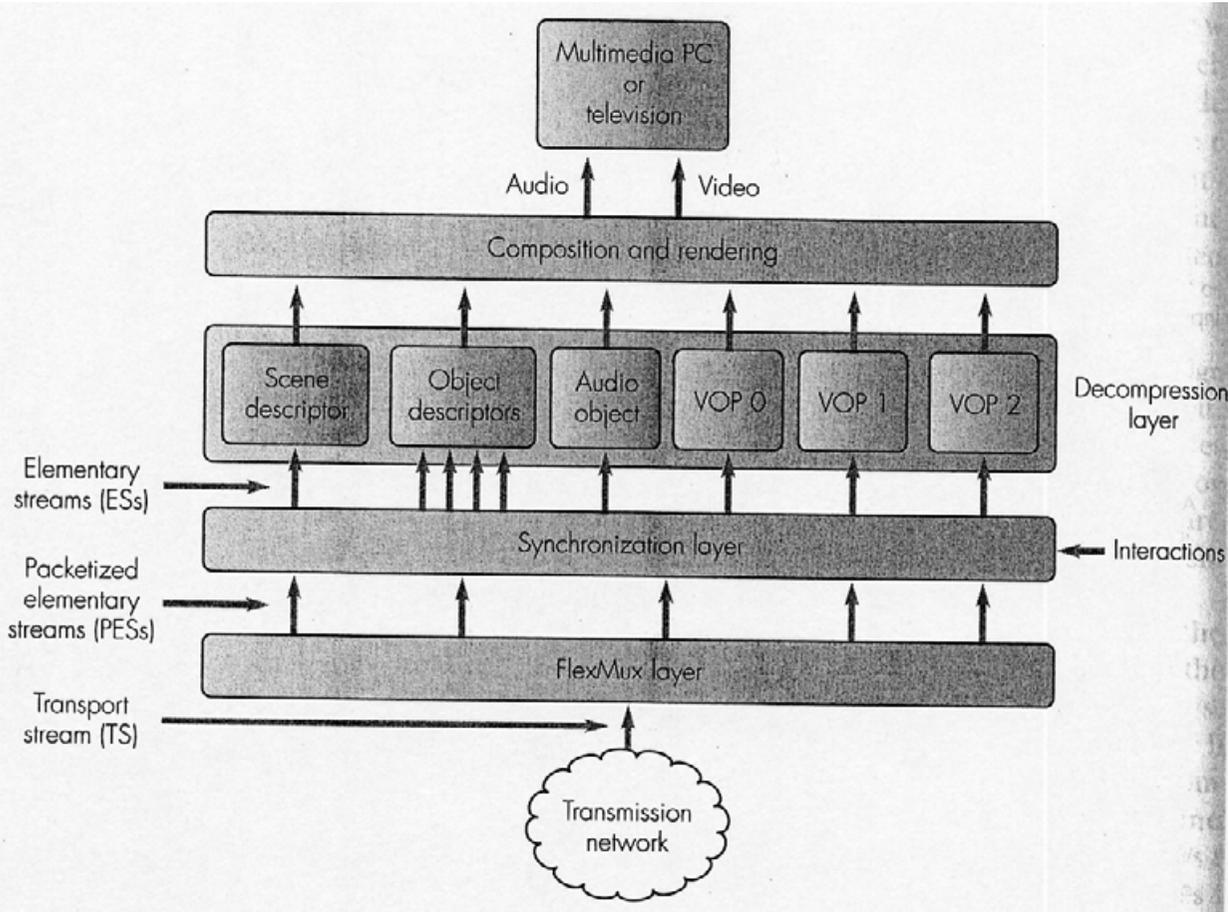


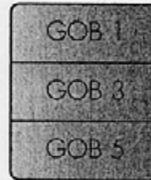
Figure 4.25 MPEG-4 decoder schematic.

Résistance aux erreurs

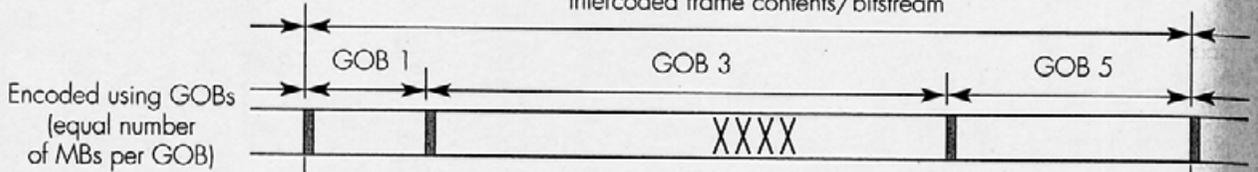
- Paquets video de taille fixe au lieu de GOBs :
 - Séparés par des marqueurs de resynchronisation
 - Longueur dépend du débit du lien (512 bits pour 24kbps)
 - La quantité d'info compressée dans un GOB varie : plus il est long plus il risque d'être transmis avec erreur
- Codage *reversible Variable Length Coding* :
 - Codage entropique pour les *motion vectors* et DCT basé sur Huffman mais pouvant être décodé dans les deux sens (préfixe ou suffixe)

(a)

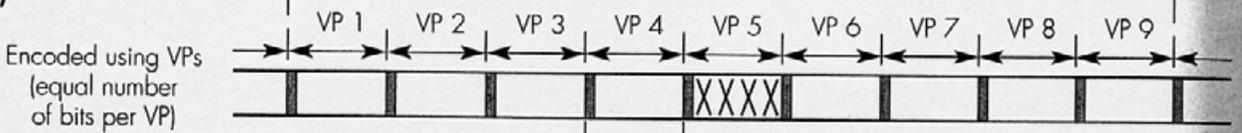
Uncoded frame contents (QCIF)



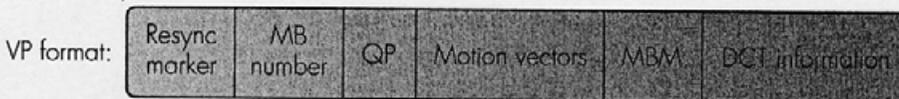
Intercoded frame contents/bitstream



(b)



(c)



GOB = group of (macro) blocks MB = macroblock VP = video packet
QP = quantization parameters MBM = motion boundary marker
X = bit errors

Figure 4.26 MPEG-4 encoding: (a) conventional GOB approach; (b) using fixed-length video packets; (c) video packet format.

Standards audio

Table 4.3 Summary of Speech compression standards and their applications.

<i>Standard</i>	<i>Compression Technique</i>	<i>Compressed bit rate (kbps)</i>	<i>Speech Quality</i>	<i>Example applications</i>
G.711	PCM + companding	64	Good	PSTN/ISDN telephony
G.721	Adaptive differential PCM (ADPCM)	32 16	Good Fair	Telephony at reduced bit rates
G.722	ADPCM with subband coding	64 56/48	Excellent Good	Audio conferencing
G.726	ADPCM with subband coding	40/32 24/16	Good Fair	General telephony at reduced bit rates
LPC-10	Linear predictive coding (LPC)	2.4/1.2	Poor	Telephony in military Networks
G.728	Code-excited LPC (CELP)	16	Good	Low delay/low bit rate Telephony
G.729	CELP	8	Good	Telephony in cellular (radio) networks
G.729(A)	CELP	8	Good	Simultaneous telephony and data (fax)
G.723.1	CELP	6.3 5.3	Good Fair	Video and Internet Telephony

Table 4.4 Summary of compression standards for general audio.

<i>Standard</i>		<i>Compressed bit rate (kbps)</i>	<i>Quality</i>	<i>Example applications</i>
MPEG Audio	Layer 1	32 – 448	Hi-fi quality at 192 kbps	Digital audio cassettes
	Layer 2	32 – 192	Near CD at 128 kbps	Digital audio and digital video broadcasting
	Layer 3	64	CD quality	CD-quality over low bit rate channels
Dolby audio coders	AC-1	512 kbps	Hi-fi quality	Radio and television satellite relays
	AC-2	256 kbps	Hi-fi quality	PC sound cards
	AC-3	192 kbps	Near CD quality	Digital video broadcasting

Standards video

Table 4.5 Summary of video compression standards.

<i>Standard</i>	<i>Digitization format</i>	<i>Compressed bit rate</i>	<i>Example applications</i>
H.261	CIF/QCIF	× 64 kbps	Video telephony/conferencing over × 64 kbps channels (p = 1-30) and LANs
H.263	S-QCIF/QCIF	<64 kbps	Video telephony/conferencing and security surveillance over low bit rate channels
MPEG-1/ISO11172	SIF	<1.5Mbps	Storage of VHS-quality video on CD-ROMs
MPEG-2/ISO 13818			
Low	SIF	<4Mbps	Recording of VHS-quality video
Main	4:2:0	<15Mbps	Digital video broadcasting
	4:2:2	<20Mbps	
High 1440	4:2:0	<60Mbps	High definition television (4/3 aspect ratio)
	4:2:2	<80Mbps	
High	4:2:0	<80Mbps	High definition television (16/9 aspect ratio)
	4:2:2	<100Mbps	
MPEG-4	Various	5kbps-tens Mbps	Versatile multimedia coding standard
MPEG-7	—	—	Structure and content descriptions of compressed multimedia information -search engines