

## CHAPITRE 3

### COMPRESSION DU SIGNAL VIDEO ET AUDIO

#### 3.1. Définitions

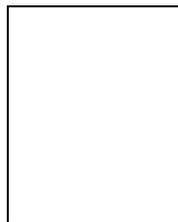
- Dans la compression, on veut réduire de façon significative la taille des fichiers graphiques puis photographiques et vidéo en vue de leur stockage ou transmission (diffusion).
- La compression peut être avec ou sans pertes (réversible) selon le but et le taux de compression.
- L'objectif de la compression est aussi de réduire le débit (la quantité d'information traitée) des formats suivants :

4 :2 :0	124 Mbits/s
4 :1 :1	124 Mbits/s
4 :2 :2	166 Mbits/s
4 :4 :4	248 Mbits/s

#### 3.2. Techniques élémentaires

##### 3.2.1. Réduction spatiale

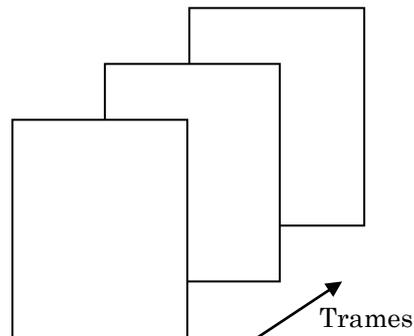
- Exploiter la redondance spatiale présente dans toute image naturelle : pixels voisins souvent identiques
- Cette redondance existe dans chaque image de la séquence
- Exploiter le fait que l'œil est moins sensible aux détails



**Figure 3.1.** La redondance spatiale.

### 3.2.2. Réduction temporelle

- Exploiter la redondance temporelle présente dans toute séquence vidéo
- Cette redondance est très importante entre 2 images consécutives (sauf “*shot change*”)
- Exploiter le fait que l’œil est moins sensible après un changement de scène (“*shot change*”)



**Figure 3.2.** La redondance temporelle.

### 3.3. La compression d’images fixes JPEG

- L’ISO (International Standard Organization) a crée en 1990 un groupe de travail international baptisé JPEG (Joint Photographic Experts Group) pour élaborer une norme de compression d’images photographiques de résolutions variables, sous formes  $YD_rD_b$  ou  $RVB$ .
- La compression JPEG peut être avec ou sans pertes (réversible), selon le but et le taux de compression recherché, mais les utilisations les plus courantes sont avec pertes, ce qui permet d’atteindre, selon la nature d’image, des taux de compression supérieur à 10 sans dégradation notable de qualité.

La compression JPEG se décompose en 6 étapes :

#### 3.3.1. Décomposition en blocs

- L’image d’origine sous forme  $YC_rC_b$  est découpée en blocs élémentaires de 8x8 pixels.
- Pour une image au format CCIR 601 de 720x576, un total de 6480 blocs de luminance  $Y$  et 3240 blocs pour chacune des composantes  $C_r$  et  $C_b$ .

- Chacun de ces blocs forme une matrice de 64 nombres de 0 à 255 (résultats de la numérisation sur 8 bits) pour la luminance et de -128 à +127 pour les composantes  $C_r$  et  $C_b$ .
- Ces valeurs représentent l'intensité lumineuse d'un pixel.

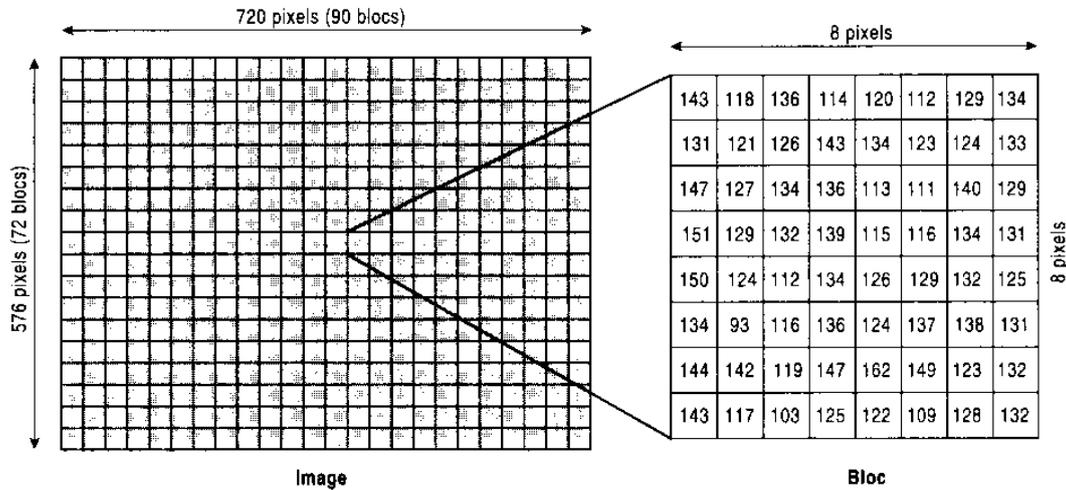


Figure 3.3. Découpage en blocs de 8x8 pixels.

### 3.3.2. Transformation en DCT

- La DCT décompose le signal en une série de fonction en cosinus harmoniques en phase avec le signal d'origine, ce qui réduit de moitié le nombre de coefficients nécessaires par rapport la DFT.
- Cette transformation appliquée à chacun des blocs ( $YC_rC_b$ ) génère pour chacun d'eux une nouvelle matrice 8x8 composée des coefficients des composantes de fréquence spatiales.
- La valeur de ces coefficients diminue rapidement en s'éloignant de l'origine de la matrice, qui se termine généralement par une série de 0.

$$G(f_x, f_y) = \frac{1}{4} C(f_x) C(f_y) \sum_{x=0}^7 \sum_{y=0}^7 g(x, y) \cos\left((2x+1)f_x \cdot \frac{\pi}{16}\right) \cos\left((2y+1)f_y \cdot \frac{\pi}{16}\right)$$

$$C(f) = \begin{cases} 1/\sqrt{2}, & f = 0 \\ 1, & f > 0 \end{cases} \quad (3.1)$$

$f_x, f_y$  : Les fréquences spatiales

$G(f_x, f_y)$  : Les coefficients DCT



- Les coefficients seront enfin quantifiés avec une précision de plus en plus faible lorsque la fréquence augmente, permettant ainsi une réduction supplémentaire de la quantité d'information nécessaire pour coder un bloc.

$$G_Q(f_x, f_y) = \text{round} \left( \frac{G(f_x, f_y)}{Q(f_x, f_y)} \right) \quad (3.2)$$

$f_x, f_y$  : Fréquences spatiales

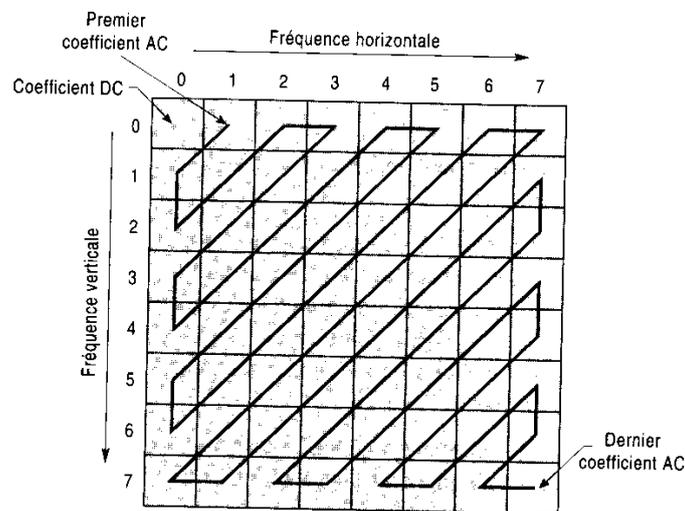
$G_Q(f_x, f_y)$  : Coefficients DCT quantifiés

$G(f_x, f_y)$  : Coefficients DCT non quantifiés

$Q(f_x, f_y)$  : Pas de quantification

### 3.3.4. Lecture en zigzag

A l'exception du coefficient DC traité séparément, les 63 coefficients AC sont lus en zigzag pour transformer la matrice en un flot de données série adapté aux prochaines étapes du processus.



**Figure 3.6.** Lecture de la matrice en zigzag.

DC	AC01	AC10	AC20	AC11	...								
5	0	0	-2	0	-1	0	0	0	0	-1	0	-1	...

**Figure 3.7.** Ordonnancement des coefficients en zigzag.

### 3.3.5. Réduction de la redondance

#### 3.3.5.1. Codage RLC (Run Length Coding)

Lorsqu'une source d'information émet des éléments de message pouvant comporter des suites relativement longues d'éléments identiques, ce qui est le cas de la DCT après seuillage et quantification, il est intéressant de ne pas coder individuellement chaque élément, mais plutôt de coder le couple Valeur/Nombre de répétition, ce qui donne un facteur de réduction de débit d'autant plus important que la suite est plus longue.

Ce codage ne perd aucune information (réversible)

Ce type de codage est couramment utilisé pour la compression en vue de stockage en informatique (Compression ZIP).

Il peut être combiné à un codage à longueur variable (VLC) pour en augmenter l'efficacité.

#### 3.3.5.2. Codage VLC (Variable Length Coding)

Ces systèmes de codage sont basés sur le fait que la probabilité d'apparition d'un élément (ou message) codé sur  $n$  bits généré par une source parmi les  $2^n$  puissances possibles n'est pas équivalente pour tous les éléments.

On aura donc intérêt à coder sur moins de bits les éléments d'apparition fréquente et sur plus de bits les éléments plus rares, pour obtenir une longueur moyenne inférieure à la longueur fixe, ce qui permet une réduction de débit.

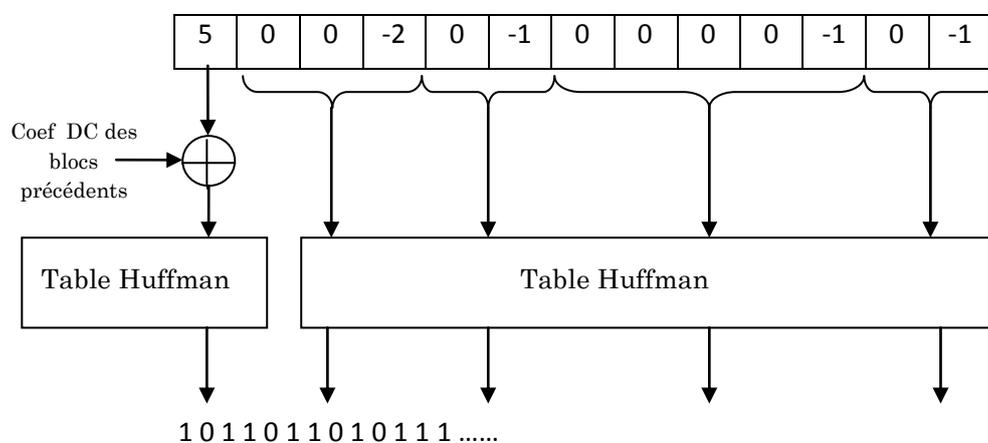
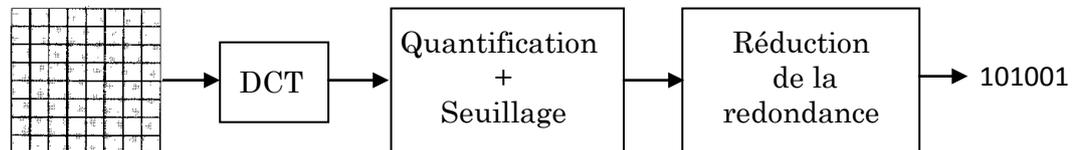


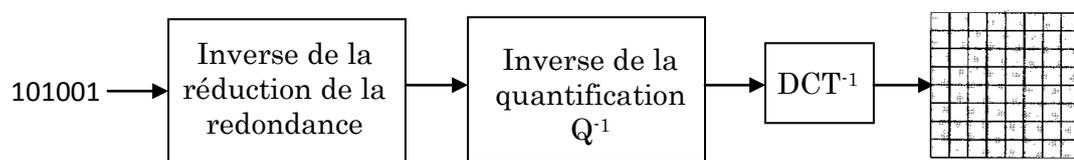
Figure 3.8. Réduction de la redondance.

Le but de codage à longueur variable, dont le plus connu est celui de Huffman, est de s'approcher le plus possible au débit entropique (nombre de bits par élément aussi proche que possible de l'entropie de la source).

### 3.4. Encodeur/Décodeur JPEG



*Figure 3.9. Schéma bloc de l'encodeur JPEG.*



*Figure 3.10. Schéma bloc de décodeur JPEG.*

### 3.5. La compression d'images animées MPEG

- Le groupe MPEG (Moving Pictures Experts Group) est créé par ISO en 1990.
- Le travail de ce groupe s'est concrétisé tout d'abord en 1992 par la norme ISO/IEC 11172, connue sous le nom MPEG-1.
- Le but principal de MPEG-1 est de permettre le stockage et la reproduction sur support CD-ROM ou CD-I avec un débit de l'ordre de 1.5 Mb/s (image + son).
- Outre la redondance spatiale intrinsèque d'une image fixe utilisée dans le codage JPEG, le codage d'images animées permet en effet d'exploiter la forte redondance temporelle des images successives qui constituent une séquence vidéo.

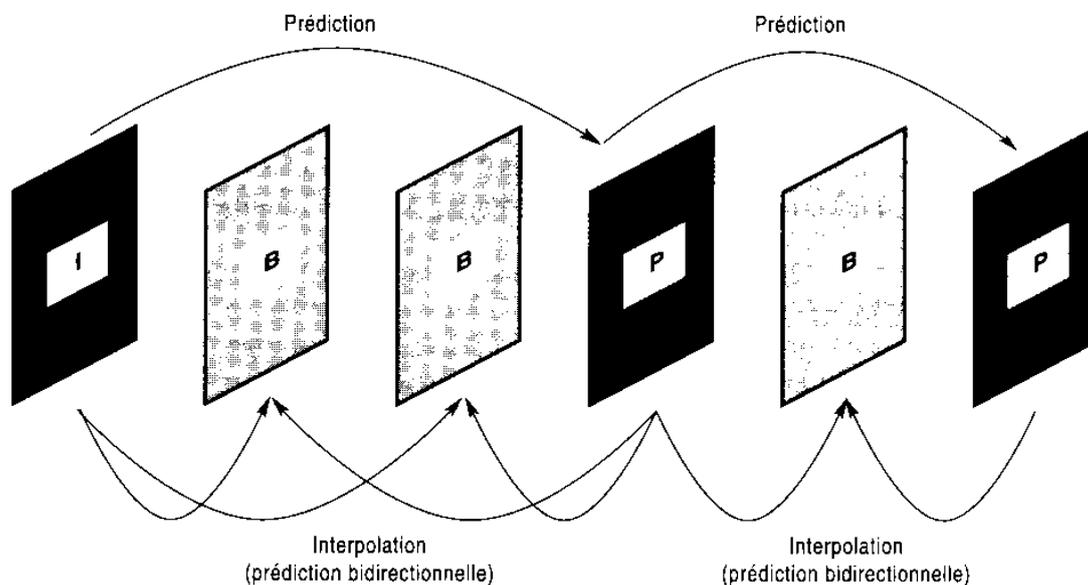
#### 3.5.1. Le codage vidéo MPEG-1

- L'objectif principal est d'atteindre un débit constant total de 1.5 Mb/s (débit d'un CD-ROM simple vitesse) dont 1.15 Mb/s pour la partie purement vidéo de MPEG-1, les 350 kb/s restant étant utilisés par le son (stéréo) et données auxiliaires.

- Le MPEG-1 utilise les mêmes étapes de JPEG et une autre technique de prédiction avec compensation de mouvement.

### 3.5.1.1. Les types d'image de MPEG

MPEG définit trois types d'images :



**Figure 3.10.** Enchaînement des trois types d'images MPEG.

#### Les images I (Intra) :

- Sont le point d'entrée obligatoire lors de l'accès à une séquence.
- Codées sans aucune référence à d'autres images, comme en JPEG.
- Elles contiennent tous les éléments nécessaires pour leur reconstruction par le décodeur.
- Le taux de compression des images I est relativement faible.

#### Les images P (Prédites) :

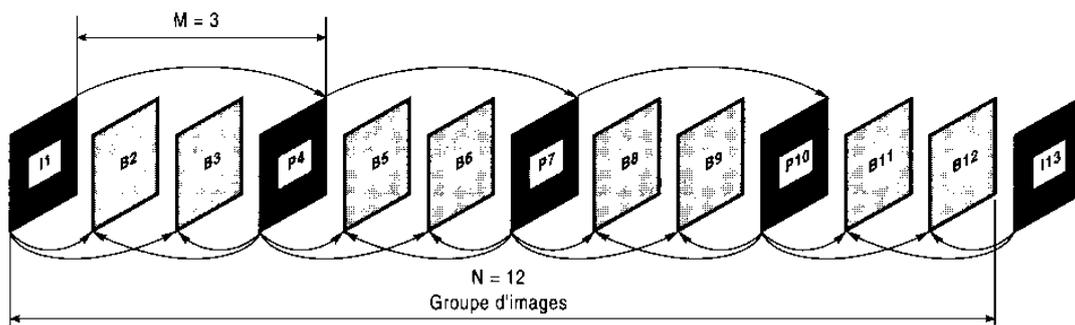
- Sont codées par rapport à l'image de type I ou P précédentes, grâce aux techniques de prédiction avec compensation de mouvement.
- Les images P propagent les erreurs de codage.
- Le taux de compression est plus important que celui des images I.

#### Les images B (Bidirectionnelles) :

- Sont codées par interpolation entre les deux images de types I ou P précédente et suivante qui les encadrent.
- Comme elles ne sont pas utilisées pour définir d'autres images, elles ne propagent pas les erreurs de codage.
- Ces images offrent le taux de compression le plus fort.

Deux paramètres  $M$  et  $N$  définissent la manière dont les images I, P et B s'enchaînent.

- $M$  est la distance (en nombre d'images) entre deux images P successives.
- $I$  est la distance entre deux images I successives.
- Les paramètres les plus couramment utilisés sont  $M = 3$  et  $N = 12$ .



**Figure 3.11.** Exemple de groupe d'image  $M = 3$  et  $N = 12$ .

Avec les paramètres définis ci-avant ( $M = 3$  et  $N = 12$ ), le mode de codage des images successives se traduit par la correspondance numéro ? (type d'image) ci-dessous :

1(I)2(B) 3(B)4(P)5(B)6(B)7(P)8(B)9(B)10(P)11(B)12(B)13(I)14(B)15(B)16(P)...

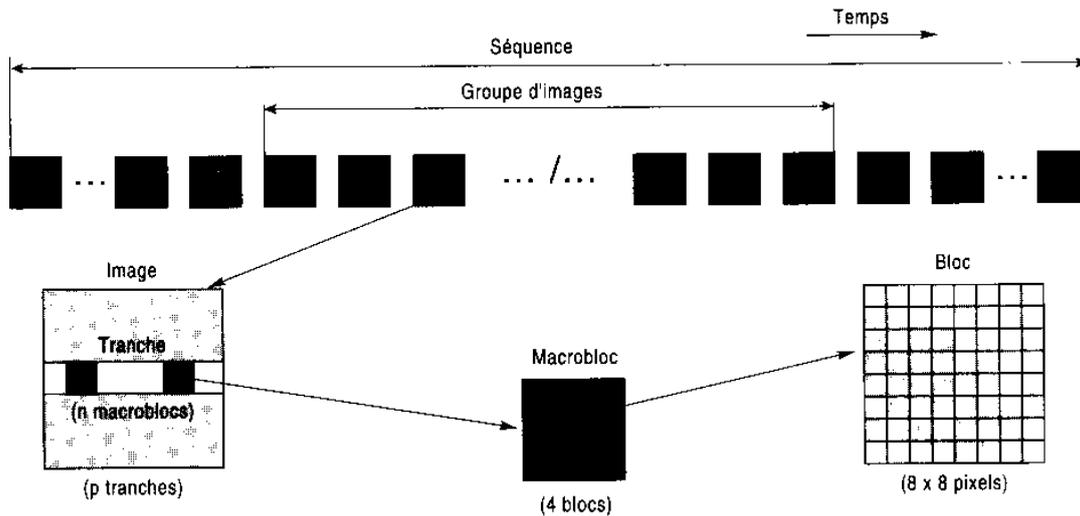
Cependant pour coder ou décoder une image B (Bidirectionnelle), le codeur et le décodeur auront besoin de l'image I ou P qui la précède et de l'image P ou I qui la suit.

L'ordre de l'image sera donc modifié avant le codage de sorte que codeur et décodeur disposent avant les images B des images I et/ou P nécessaire à leur traitement, soit :

1(I)4(P)2(B)3(B)7(P)5(B)6(B)10(P)8(B)9(B)13(I)11(B)12(B) 16(P)14(B) 15(B) ...

### 3.5.1.2. Décomposition en couches d'une séquence vidéo MPEG

MPEG définit également une hiérarchie de couches (Layers) à l'intérieur d'une séquence d'images.



**Figure 3.12.** Hiérarchie de couches de la séquence en bloc.

On trouve en partant du niveau le plus haut :

**Séquence :**

Couche la plus élevée, elle détermine le contexte dans lequel cette séquence est définie (paramètres vidéo de base tels que norme de balayage, etc.).

**Groupe d'images (Group of Pictures GoP):**

C'est la couche permettant l'accès aléatoire à la séquence, avec une première image de type I. Dans le cas décrit ci-avant, le groupe d'image en comporte 12 (1xI, 3xP, 8xB).

**Image (Picture) de type I, B et P:**

C'est la couche d'affichage élémentaire

**Tranche (Slice):**

Est défini comme une suite de macro blocs contigus, l'ensemble des slices doit couvrir toute l'image, sans chevauchement entre elles. Une slice peut théoriquement aller d'un seul macrobloc à l'image entière, mais en pratique, il s'agira le plus souvent d'une rangée horizontale complète de macro blocs.

**Macrobloc (Macroblock) :**

De taille 16x16 pixels (4 blocs de luminance et 2 blocs de chrominance 1xCr et 1xCb chacun de 8x8 pixels) : c'est la couche où s'effectue la prédiction avec compensation de mouvement.

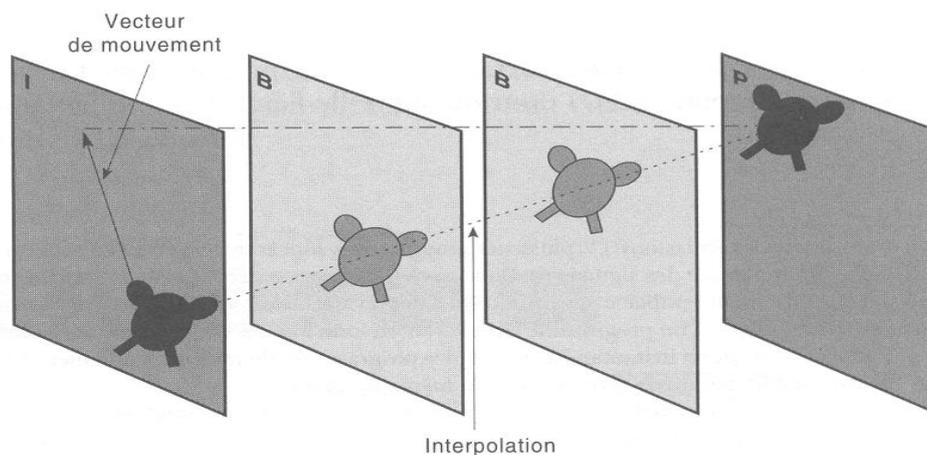
### Bloc (Block) :

De 8x8 pixels, c'est comme en JPEG, la couche où s'effectue la DCT.

- En raison de découpage de l'image en nombre entier de macroblochs de 16x16, la définition horizontale de MPEG-1 est réduite à 352 pixels en luminance (22 macroblochs) et non pas 360 comme pour l'image SIF d'origine (En effet, 360 n'est pas multiple de 16).
- La résolution MPEG-1 est donc 352x288 @ 25 Hz (soit 22x18=396 macro-blocs) pour les images dont l'origine est à 625 lignes et 352x240 @ 30 Hz (soit 22x15=330 macro-blocs) pour les images d'origine 525 lignes.

#### 3.5.1.3. L'estimation et la compensation de mouvement

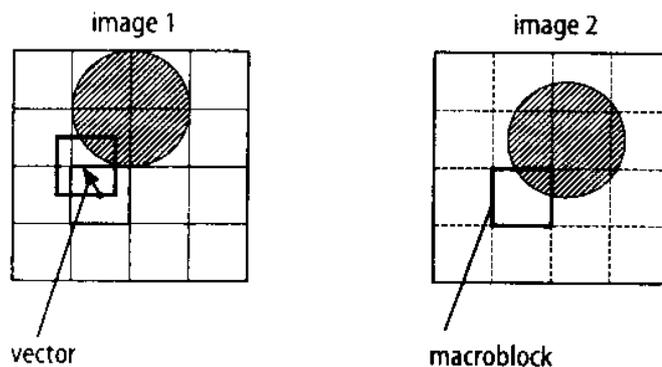
- Entre deux images successives, les objets en mouvement entraînent une différence entre les zones correspondantes au départ sur la première image et à l'arrivée sur la deuxième.
- Il n'y a pas de corrélation temporelle évidente entre ces zones.



**Figure 3.12.** Représentation schématique de la compensation de mouvement.

- L'estimation de mouvement consiste donc à trouver un vecteur assurant la correspondance entre une zone d'arrivée sur la deuxième image et une zone de départ sur la première.
- Cette recherche se fait au niveau de macro bloc (16x16) pixels, en déplaçant, à l'intérieur d'une fenêtre de recherche, un macro bloc de l'image en cours au voisinage de cette position dans l'image précédente de manière à trouver le plus ressemblant (Block matching), ce qui donne un vecteur de mouvement, qui s'appliquera à toute les composantes du macro bloc ( $YC_r C_b$ ).

- Seuls les macros blocs différent d'une image à l'autre auront ainsi besoin d'être codés, d'où une réduction significative de la qualité d'information à transmettre.
- Cette comparaison se faisant entre une image P et une image I ou deux images P, les vecteurs de mouvements peuvent être d'amplitude relativement importante (en raison de la distance entre ces deux images, 3 dans le cas  $M=3$ ,  $N=12$ ).
- Pour les images B, les vecteurs intermédiaires sont obtenus par une interpolation qui tient compte de leur position temporelle entre les images I et/ou P.
- On calcule ensuite la différence entre le bloc à coder et le bloc que l'estimation de mouvement à trouver le plus proche, et cette différence est transmise après un codage similaire à celui des blocs des images (DCT, Quantification, VLC).
- La taille moyenne des objets étant supérieure au macro bloc, il y a généralement corrélation entre les vecteurs de mouvement associés à deux macro blocs contigus, d'où l'intérêt d'utiliser une méthode de codage différentielle (DPCM) avec codage entropique pour transmettre ces vecteurs.
- Si pour un bloc donné la compensation de mouvement n'a pas donné de résultat exploitable, ce qui est relativement peu fréquent, le bloc est codé en Intra.



**Figure 3.13.** Exemple d'estimation de mouvement.

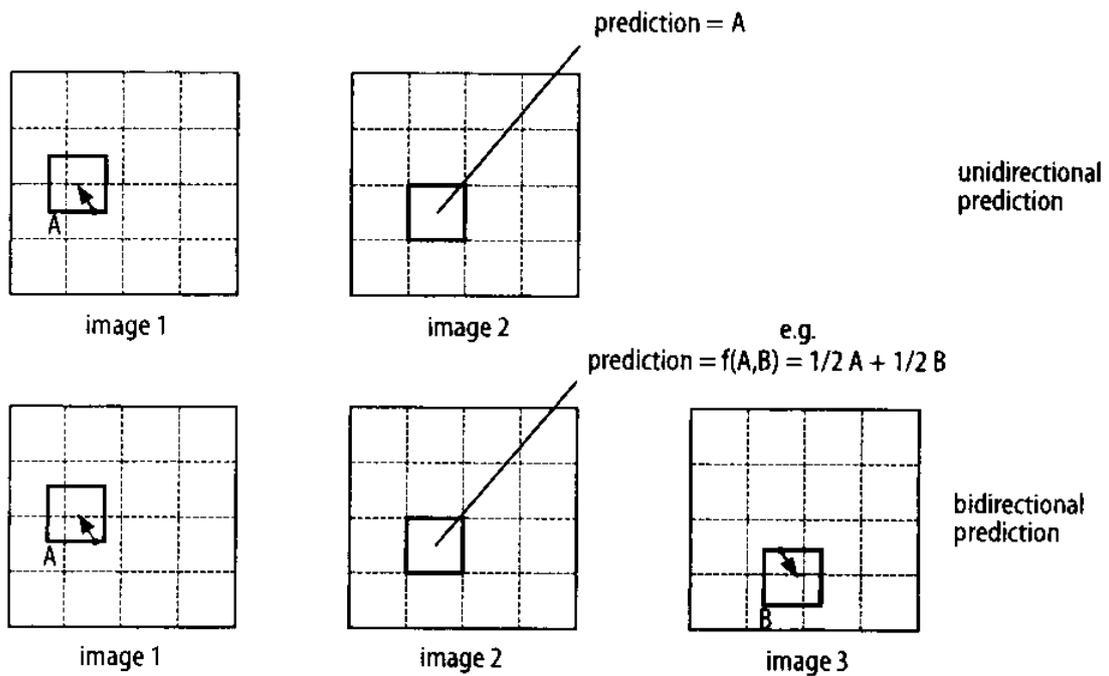


Figure 3.14. La prédiction unidirectionnelle et bidirectionnelle.

### 3.5.1.4. Régulation de débit

- Pour s'adapter aux contraintes du canal (de transmission ou d'enregistrement) et du buffer d'entrée spécifié pour le décodeur MPEG de référence, le débit doit généralement être constant en sortie du codeur.

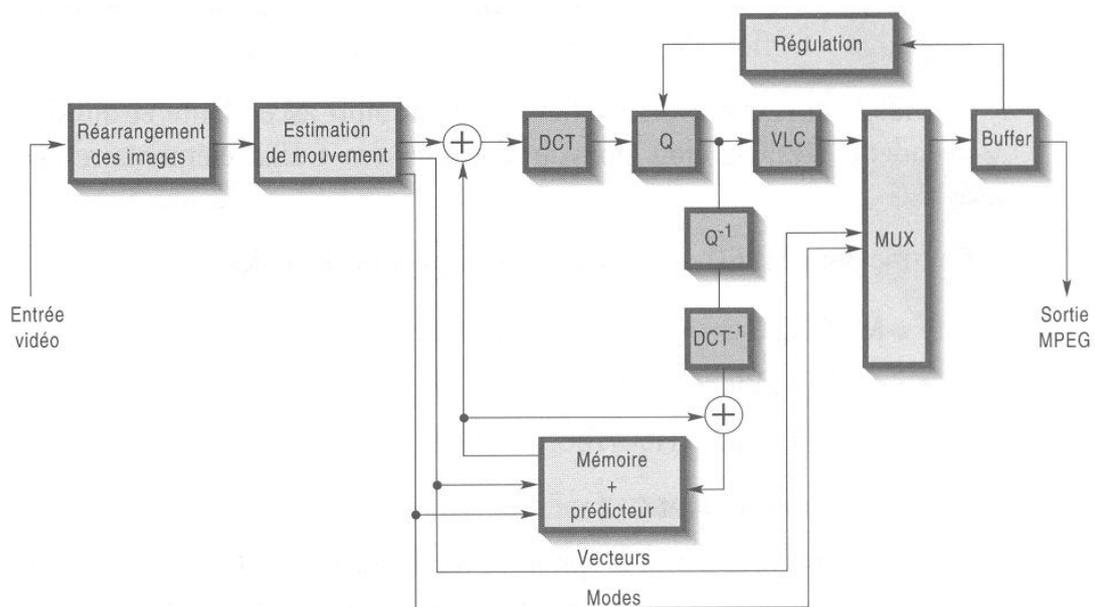
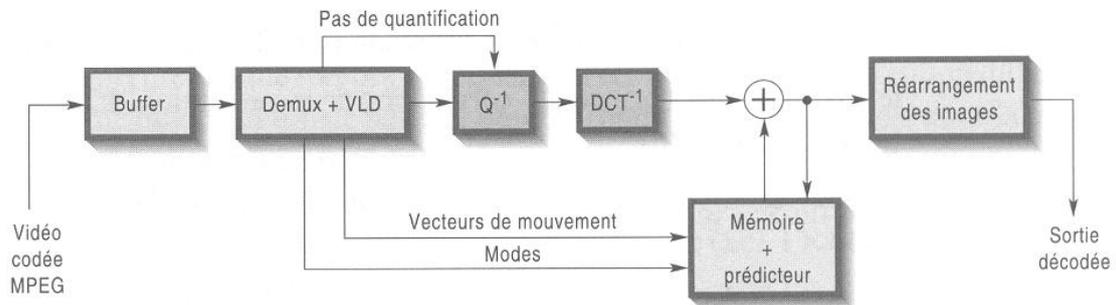


Figure 3.15. Schémas du codeur MPEG-1.

- Pour régler le débit en sortie du codeur, on utilise en fin de processus une mémoire tampon (FIFO) dont on surveille l'état de remplissage de manière à le maintenir entre certaines limites en jouant sur la résolution des coefficients de quantification, dont l'influence sur le débit est majeure.



**Figure 3.16.** Schémas du décodeur MPEG-1.

### 3.6. Le codage MPEG-2

- MPEG-2 constitue la norme pour le codage de source du système défini par le DVB.
- MPEG-2 peut être décrite comme une boîte à outil de compression plus complexe que MPEG-1.
- MPEG-2 se compose de 3 parties :
  - ✓ MPEG-2 système (ISO/IEC 13818-1)
  - ✓ MPEG-2 vidéo (ISO/IEC 13818-2)
  - ✓ MPEG-2 audio (ISO/IEC 13818-3)

#### 3.6.1. Profils et niveaux MPEG-2

- MPEG-2 comporte 5 profils (profiles) qui déterminent le jeu d'outils de compression utilisé, donc le compromis entre taux de compression et coût de décodeur.
- MPEG-2 comporte 4 niveaux (levels) définissent la résolution de l'image.

##### 3.6.1.1. Les niveaux

- Le niveau *low* (bas) correspond à la résolution SIF utilisée en MPEG-1.
- Le niveau *main* (principal) correspond à la résolution 4:2:0 (normale) (jusqu'à 720x576).
- Le niveau *high-1440* (haut-1440) est destiné à la TVHD (jusqu'à 1440-1152).
- Le niveau *high* (haut) est optimisé pour la TVHD (jusqu'à 1920-1152).

### 3.6.1.2. Les profils

- Le profil *simple* est destiné à simplifier le codeur et le décodeur, au détriment du taux de compression, en n'utilisant pas de prédiction de type B (Bidirectionnelle).
- Le profil *main* correspond actuellement au meilleur compromis qualité/taux de compression, en utilisant les 3 types d'images (I, P, B), au prix d'un codeur et d'un décodeur plus complexes.
- Les profils *scalables* (codage hiérarchique) sont prévus pour des utilisations ultérieures et permettront de transmettre une image de base (base layer) en terme de résolution spatiale (spatially scalable profile) ou de quantification (SNR scalable profile) ainsi que des informations supplémentaires séparées (enhanced layer), permettant d'améliorer ses caractéristiques, par exemple pour transmettre la même émission en définition standard et HD ou permettre une réception de qualité acceptable en cas de réception difficile et de qualité optimale dans de bonnes conditions (par exemple pour la télévision numérique terrestre).
- Le profil *high* est prévu pour des applications de télédiffusion HDTV (format 4 :2 :0 ou 4 :2 :2).

La combinaison la plus importante à court terme, car c'est celle retenue pour les applications de télédiffusion numérique grand public en Europe, est dite *main profile at main level* (MP@ML).

Elle correspond au codage d'images entrelacées au format 4 :2 :0 de résolution 720x480 (30 Hz) ou 720x576 (25 Hz) avec une boîte à outils permettant le codage d'image de type I, P et B.

Selon le compromis qualité/débit recherché et la nature des images, le débit sera compris entre 4 Mb/s (qualité d'image équivalente à celle d'une image codée en PAL ou SECAM) et 9 Mb/s (qualité voisine d'une image de studio CCIR 601).

Tout le processus de codage des images animées décrit pour MPEG-1 s'applique à MPEG-2 (MP@ML), notamment la hiérarchie de couches, mis à part la résolution d'origine et le traitement des images entrelacées.

Une différence est à noter pour les slices en MPEG-2, elles ne couvrent pas nécessairement toute l'image, et elles doivent de plus se composer uniquement de macroblocs contigus situés sur la même ligne horizontale.

### 3.6.2. Modes de prédiction spécifique à MPEG-2 (images entrelacées)

La figure 3.17 représente la séquence temporelle de la position verticale des lignes des trames successives dans un système entrelacé.

Pour le codage intra des images entrelacées, MPEG-2 permet de choisir entre deux structures d'images dites *frames* (structure image) ou *field* (structure trame).

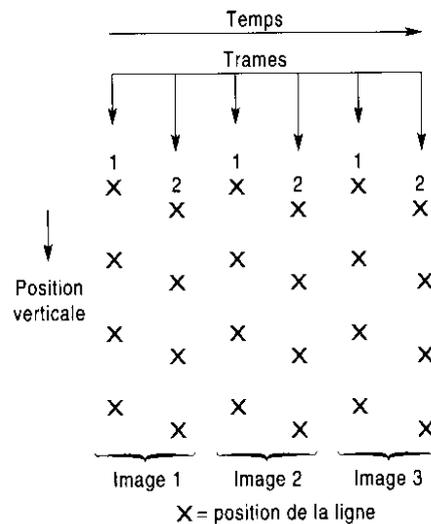


Figure 3.17. Position des lignes de trames successives dans un système entrelacé.

#### 3.6.2.1. Structure image (progressive)

Elle est appropriée en cas où il y a peu de mouvement entre deux trames successives. Les blocs et macroblocs sont découpés dans l'image complète et la DCT s'effectue donc sur des points verticaux distants de 20 ms dans le temps, ce qui ne pose pas de problème si les deux trames diffèrent peu.

Dans ce cas, il reste toutefois possible de coder les blocs les plus animés en mode inter-trame, c. a. d. en les découpant dans une trame.

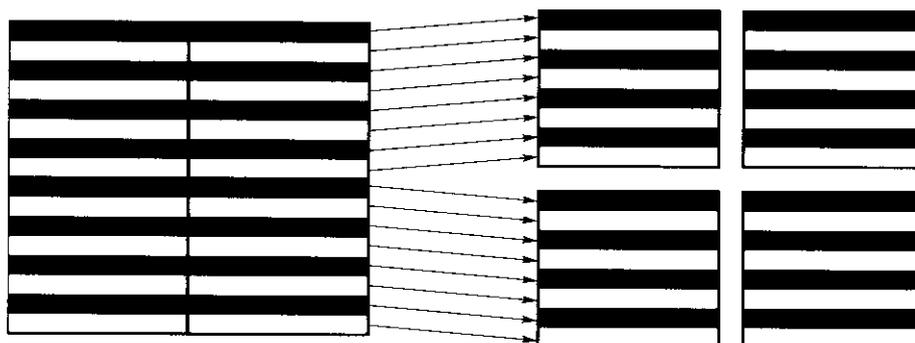
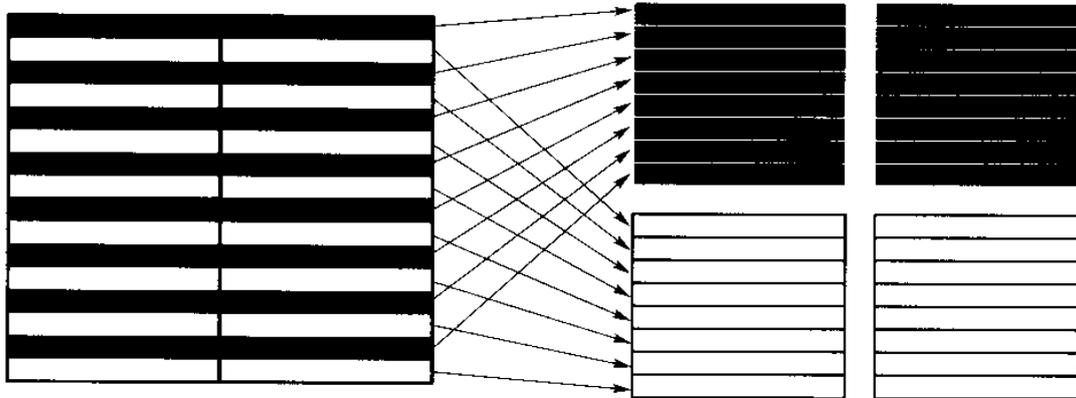


Figure 3.18. Découpage des blocs dans les macroblocs en mode image.

### 3.6.2.2. Structure trame (entrelacée)

Elle est préférable lorsque le mouvement est important d'une trame à l'autre : dans ce cas, afin d'éviter un contenu en fréquences verticales élevées qui réduirait l'efficacité de la compression après DCT, le découpage des macroblocs est fait en considérant chacune des trames comme une image indépendante à l'intérieur de laquelle sont pris les blocs.



*Figure 3.19. Découpage des blocs dans les macroblocs en mode trame.*

En ce qui concerne l'estimation de mouvement, plusieurs modes sont également prévus. Un macrobloc peut être prédite en mode image, trame ou mixte

- Dans le mode image, un macrobloc formé dans la trame impaire sert à prédire un bloc correspondant dans la prochaine trame impaire, et de même pour les blocs de la trame paire, et la prédiction se fait donc sur une durée de 40 ms (2 trames).
- Dans le mode trame, la prédiction d'un bloc est effectuée à partir d'un bloc de la trame précédente, et les vecteurs de mouvement correspondent alors à une durée de 20 ms.
- Dans le mode mixte, les blocs sont prédits à partir de deux blocs appartenant à deux trames.

### 3.6.3. Description du processus de codage MPEG-2

Tout comme en MPEG-1, la norme ne définit pas explicitement de méthode de codage, mais seulement la syntaxe régissant le train binaire en sortie du codeur, ce qui laisse une grande liberté à son concepteur.

A partir d'une image numérisée au format 4 :2 :0 (main profile), le codeur choisit pour chaque image son type (I, P ou B) et si elle doit être codée en mode *frame* ou *field*. Le codeur doit ensuite estimer les vecteurs de mouvement pour chaque macrobloc de 16x16 pixels.

Dans le cas général où le codeur est capable de générer des images B (bidirectionnelles), il devra réordonner les images avant codage et transmission.

L'unité de codage de base est le macrobloc, composé de 4 blocs de luminance de 8x8 pixels et (dans le cas du format 4 :2 :0) de 2 blocs de chrominance (un  $C_b$  et un  $C_r$ ) de 8x8 pixels qui couvrent la même zone de l'image.

Tous les macroblocs de l'image sont codés en séquence de la gauche vers la droite et du haut vers le bas, et un mode de codage est choisi indépendamment pour chacun d'eux.

Une fois le mode de codage choisi, la prédiction avec compensation de mouvement du contenu du bloc est faite à partir de l'image de référence (I ou P) passée (cas des images P) et éventuellement future (cas des images B). La prédiction est soustraite des données réelles du macrobloc, ce qui donne le signal d'erreur de prédiction.

Dans l'image à structure frame, le codeur devra choisir d'effectuer la DCT en mode frame ou field, ceci dépend principalement de l'amplitude du mouvement entre les deux trames de l'image.

Le signal d'erreur est ensuite séparé en bloc de 8x8 auquel la DCT est appliquée. Chaque bloc de coefficients résultant est quantifié et balayé en zigzag pour former une suite de coefficients.

L'information auxiliaire nécessitée par le décodeur pour reconstruire le bloc (mode de codage, vecteurs de mouvement...) est codée et les coefficients quantifiés sont codés en utilisant une table VLC (codage de Huffman).

Une unité de contrôle de débit surveille l'état de remplissage du FIFO de sortie, et utilise cette information en retour pour contrôler le nombre de bits que le codeur générera pour les blocs suivants, en jouant principalement sur les coefficients de quantifications.

On obtient alors en sortie du codeur un train binaire complet utilisable par un décodeur.

Pour augmenter la qualité d'image décodée, le codeur lui-même stock et décode (déquantification des coefficients puis DCT inverse) les images I et P, référence pour reconstruire d'autres images obtenues par prédiction avec compensation de mouvement dans le décodeur, et calcule un signal d'erreur qui est ajouté au signal de prédiction.

### **3.6.4. Description du processus de décodage MPEG-2**

Comme nous l'avons déjà dit, le décodage est plus simple que le codage, car il n'a pas à effectuer d'estimation de mouvement qui est l'une des parties les plus complexes du codeur.

Le tampon (Buffer) d'entrée reçoit les données du canal de transmission, et le décodeur lit le train binaire jusqu'à trouver le début d'une image, son type (I, P ou B) et sa structure (frame ou field).

Il débute le décodage à la première image I qu'il doit stocker en mémoire, ainsi que l'image P suivante pour servir de référence aux images P ou B qui en dépendent.

Pour les images P ou B, il consiste à construire la prédiction de chaque macrobloc à partir de son type, des vecteurs de mouvement et des images de référence en mémoire. Le décodeur lit, décode et quantifie les coefficients DCT de l'erreur de prédiction transmise pour chaque bloc de 8x8, et après transformation DCT inverse, ajoute le résultat à la prédiction.

L'image est reconstruite lorsque tous les macroblochs ont été traités.

La dernière étape du décodage est la remise des images dans l'ordre initial d'affichage.

Comme on l'a vu, le besoin de mémoire pour le décodeur est d'environ 3 images (deux images de référence et l'image en cours de reconstruction), soit pour une image 4 :2 :0 environ 16 Mbits.

### **3.7. La norme de compression MPEG-4.10 ou H264/AVC**

Cette norme de compression vidéo est issue des travaux d'une équipe commune (JVT, Joint Video Team) composée de membres du Video Coding Expert Group (VCEG) de l'ITU-T et du Motion Pictures Expert Group (MPEG) de l'ISO-IEC, d'où sa double appellation de H 264 et MPEG-4.10.

L'appellation H264/AVC (Advanced Video Coding) est également souvent utilisée.

Cette norme présente l'avantage d'augmenter considérablement l'efficacité de compression (gain d'au moins 50% par rapport à MPEG-2), ce qui est particulièrement intéressant en vue de la télévision à haute définition (TVHD) qui en MPEG-2 demande une bande passante de 15 à 18 Mb/s.

C'est pourquoi, en 2004, le DVB a autorisé et défini son utilisation en alternative au MPEG-2 pour la TV par satellite, câble, terrestre ou tout autre moyen de transmission à large bande (ADSL par ex)

Le standard H264/AVC se compose de deux couches : une couche de codage vidéo (Video Coding Layer ou VCL) dont le rôle est de représenter de façon la plus compacte possible le contenu vidéo et une couche d'abstraction du réseau (Network Abstraction Layer ou NAL) qui formate la représentation VCL de manière la plus appropriée à sa destination (transport ou stockage)

- La couche (NAL) formate les données issues du codage VCL en (unités NAL) qui contiennent un nombre entier d'octets. Selon la destination du stream généré (transport stream ou bit stream pur), les unités NAL sont ou non précédées d'un préfix de démarrage.
- La partie (VCL) représente le codage proprement dit de H264. Tout comme MPEG-1 et 2, c'est un hybride de prédiction inter-images, qui exploite les dépendances statistiques spatiales (intra image) au moyen d'un codage par transformée du résidu de prédiction.

Comme MPEG-4, H264 supporte le codage d'images au format 4:2:0 entrelacées ou progressives, mais dans le cas de H264, une séquence peut comporter les deux types d'images.

Les dernières extensions de la norme (FRExt) permettent aussi de coder des images au format 4:2:2 et 4:4:4 avec différents espaces de couleurs (YCbCr, RGB) et autorisent une précision supérieurs à 8 bits (10 ou 12) pour les échantillons vidéos.

L'amélioration très importante de l'efficacité du codage H264 par rapport à MPEG-2 ne provient pas de l'adoption d'un algorithme « révolutionnaire » mais d'une somme d'améliorations relativement peu importantes prises séparément.

C'est l'utilisation combine :

- De modes de prédiction plus sophistiqués, qui peuvent varier à l'intérieur même d'une image et se référer à un nombre plus important d'images successives.
- D'une transformée entière au lieu de la DCT pouvant porter sur des blocs 4\*4 ou 8\*8.

- D'un codage entropique adaptatif plus efficace (CAVLC et CABAC) qui permet d'atteindre ce résultat.

Le codage entropique est « adapté au contexte » et peut utiliser deux modes dont l'un appelé CABAC (Context Adaptive Binary Arithmetic Coding) permet à lui seul de réduire d'environ de 10 à 15% le débit nécessaire au codage d'une image à qualité égale par rapport à l'autre dit CAVLC (Context Adaptive Variable Length Coding), qui est déjà plus performant que le simple VLC de MPEG-2.

L'utilisation de nouveaux outils (PicAFF, Picture Adaptive Frame Field et MBAFF, Macro Block Adaptive Frame Field) permet un meilleur traitement des images entrelacées.

Enfin, un filtre de «déblocage » permet de diminuer sensiblement la visibilité des blocs – un défaut commun à la plupart des systèmes de compression vidéo- et en conséquence de réduire d'environ 5 à 10% le débit à qualité d'image subjective égale.

Tout comme MPEG-2, H264/AVC comporte plusieurs profils et niveaux.

### **3.8. La compression des signaux audio**

L'utilisation de l'audio numérique est devenue courante pour le grand public depuis l'apparition de disque compact (CD) au début des années 1980.

Le but étant d'obtenir une qualité de haute fidélité, une bande passante d'au moins 20 KHz était nécessaire, ce qui impliquait donc une fréquence d'échantillonnage supérieur à 40 KHz, c'est une valeur de 44.1 KHz qui a finalement été retenue.

Il fallait également assurer un rapport signal/bruit et une dynamique élevés (supérieurs à 80 dB).

La numérisation d'un signal analogique (Pulse Code Modulation, PCM), introduit un bruit de quantification correspondant à l'incertitude sur le bit de poids le plus faible, ce qui se traduit par un rapport signal/bruit de 6dB par bit de quantification, soit 96 dB avec la numérisation sur 16 bits retenue.

Il en résulte un débit de  $44.1 \times 16 \times 2 = 1411.2$  Kb/s pour un signal stéréophonique.

Un tel débit est acceptable sans compression pour le Compact Disc audio, car sa capacité de 640 Mo permet d'assurer dans ces conditions environ 74 minutes de son stéréophonique haute-fidélité.

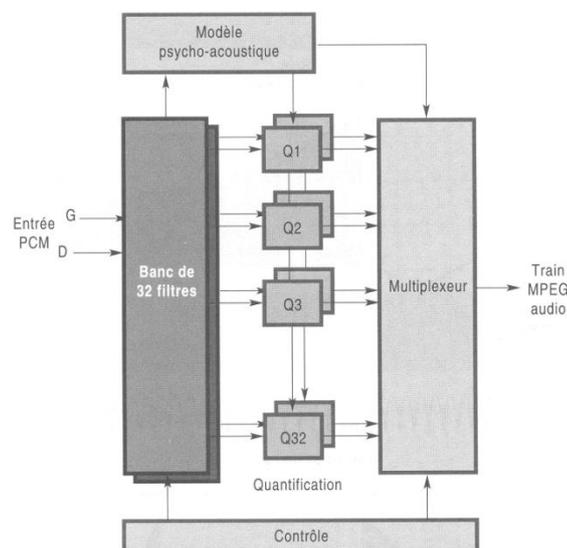
Deux autres fréquences d'échantillonnage sont couramment utilisées : 32 KHz (D2-MAC, NICAM, télécommunications...) et 48 KHz (enregistrement de studio, cassette audio numérique DAT...).

Tout comme cassette audio numérique DCC, les normes MPEG-1 et 2 prévoient la possibilité d'utiliser comme source des signaux audio échantillonnés à l'une des 3 fréquences ci-dessus.

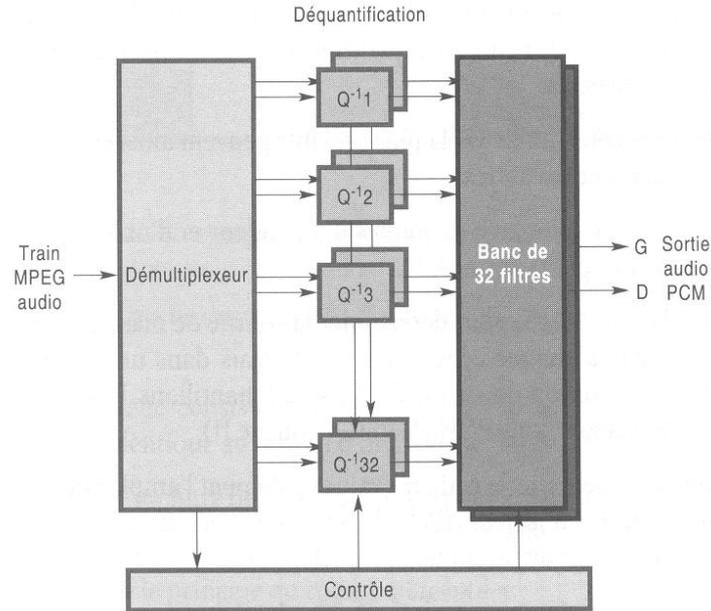
Les débits qu'elles procurent sont évidemment quelque peu différent, mais reste du même ordre de grandeur (de 1 à 1.5 Mb/s), bien trop élevés pour l'utilisation telle quelle en son d'accompagnement d'une image codée à la norme MPEG-1 ou MPEG-2.

Comme pour la vidéo, on est donc confronté au besoin de compresser les données issues de la numérisation du signal audio.

### 3.8.1. Principes de la compression audio MPEG



**Figure 3.20.** Schéma théorique du codeur MPEG audio.



**Figure 3.21.** Schéma théorique du décodeur MPEG audio.

### 3.8.2. Les couches du codage audio MPEG-1

Les normes MPEG audio définissent trois couches (layers) de codage, qui diffèrent par leur taux de compression pour une qualité audio perçue donnée.

La norme de télévision numérique DVB prescrit pour le son l'utilisation des couches I et II de la spécification MPEG-1 audio qui prévoit quatre modes de transmission principaux :

- Stereo : les voies G et D sont codées de manière complètement indépendante.
- Joint\_stereo : exploitation de la redondance entre les voies gauche et droite afin de réduire le débit.
- Dual\_channel : deux voies sont indépendantes (par exemple son bilingue).
- Mono : une seule voie son.

#### 3.8.2.1. La couche I ou pré-MUSICAM

Elle utilise l'algorithme PASC (Precision Adaptive Subband Coding) développé par PHILIPS pour sa cassette audio numérique (DCC).

Elle utilise un débit fixe parmi 14 possibles (de 32 à 448 Kb/s) ; la qualité haute fidélité nécessite 192 Kb/s par voie audio (384 Kb/s en stéréo).

Son avantage principal est la relative simplicité d'implémentation du codeur et du décodeur.

### **3.8.2.2. La couche II**

Son algorithme est connu sous le nom de MUSICAM, est le standard retenu pour la radio (DAB) et la télévision (DVB) numériques Européennes.

Elle permet d'obtenir une qualité équivalente avec un débit moindre (réduction de 30 à 50%) que pour la couche I, au prix d'un accroissement modéré de la complexité du codeur et du décodeur.

Le débit, fixe, peut être choisi de 32 à 192 Kb/s par voie, la qualité subjective HiFi étant obtenu à partir de 128 Kb/s par voie, soit 256 Kb/s en stéréo.

### **3.8.2.3. La couche III**

C'est le fameux format MP3, qui utilise un modèle psycho-acoustique différent dit modèle 2, un codage de Huffman et une analyse du signal basée sur la DCT au lieu de codage en sous-bandes des couches I et II. Les deux types de codages joint\_stereo sont autorisés.

Elle permet le débit variable et un taux de compression environ 2 fois plus élevé que la couche 2, au prix d'une complexité nettement supérieure du codeur et du décodeur, ainsi que d'un temps de codage/décodage plus long. La qualité HiFi est obtenue dès 64 Kb/s par voie (128 Kb/s en stéréo).

La norme MPEG-2 audio reprend l'essentiel de MPEG-1 et prévoit en outre une possibilité d'extension multicanaux permettant la transmission de son à voie multiples tout en restant compatible avec le mode stéréo MPEG-1 de base, ce qui permet une évolution ultérieure de la spécification DVB.

La norme MPEG-2 audio prévoit de plus la possibilité d'utiliser des fréquences d'échantillonnage égales à la moitié des valeurs standards (16 KHz/22.04KHz/24KHz), doublant ainsi la durée des trames et divisant le débit par deux, au prix bien sûr d'une bande passante réduite de moitié.

### **3.8.3. Autre codage audio : Dolby Digital (AC3), DTS**

Le son multicanal devenant plus important avec la généralisation des installations audio de type « home cinéma » (Home Theater), le DVB a ajouté à la norme la possibilité de transmettre un son de type 5.1 utilisant le système Dolby Digital (AC3) ou DTS, le MPEG-2 audio avec extension multicanal n'ayant pas rencontré un grand succès, même en Europe.

Le son AC3 ou DTS est transmis sous forme de PES en tant que données privées et doit être signalisé s'il est présent.

Des descripteurs DVB publics AC3 descriptor et DTS descriptor ont été définis à cet effet.

Le son de base MPEG-1 couche II (ISO/IEC11172-3) ou MPEG-2 multicanal (ISO/IEC13818-3) doit néanmoins toujours être transmis d'une part pour que le parc des récepteurs existants puisse recevoir en mono ou stéréo le son des programmes utilisant le son multicanal, d'autre part parce que le décodeur AC3 et DTS n'est en général pas inclus dans les récepteurs pour en réduire le coût et éviter d'avoir à payer deux fois les royalties du système (ces décodeurs étant généralement inclus dans les amplificateurs « home cinéma » récent)

De ce fait, le flux est généralement transmis tel quel à l'amplificateur multicanal au moyen d'une interface SP-DIF (Sony Philips Digital Inter Face) sous forme électrique ou optique.

Le Dolby Digital est un codage perceptuel basé sur les mêmes principes que le codage MPEG, qui offre l'avantage par rapport au MPEG-2 multicanal d'être utilisé depuis de nombreuses années dans les cinémas et actuellement par les DVD, pour lesquels il est devenu le standard de fait.

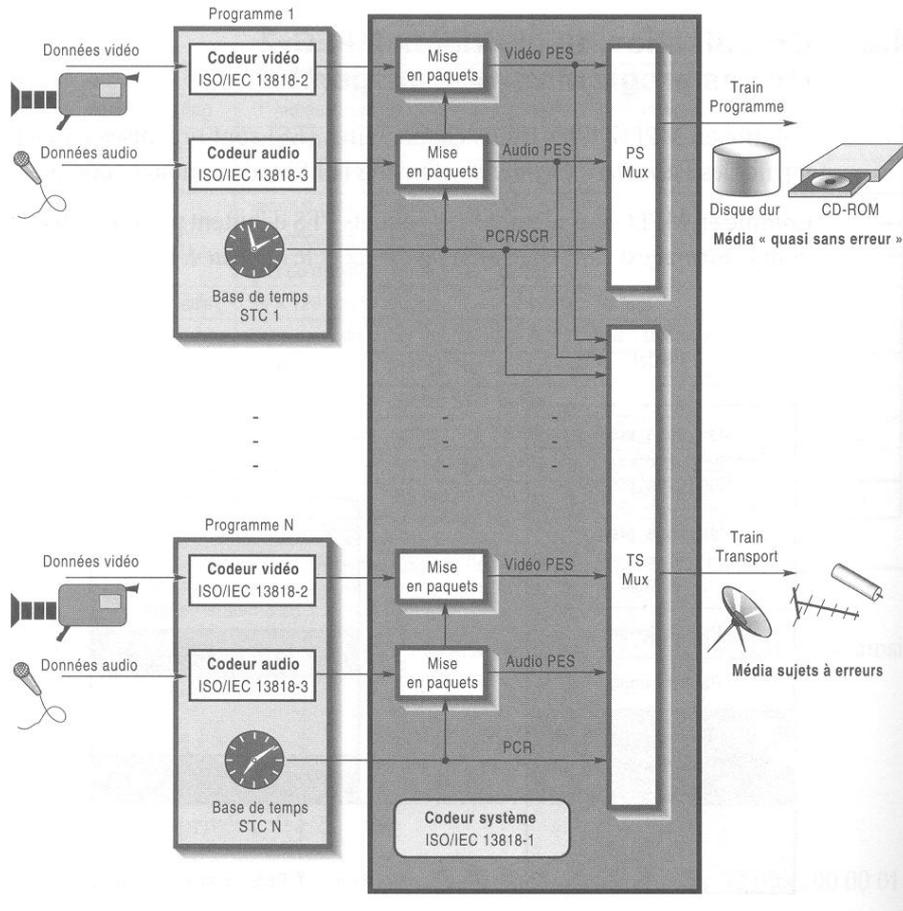
Il supporte des fréquences d'échantillonnage de 32, 44.1 et 48 KHz et des débits audio de 32 à 640 Kb/s, et il peut coder l'audio sous forme mono, stéréo ou multicanal 5.1.

### **3.9. Organisation du multiplex MPEG-2**

Comme en MPEG-1, les trains élémentaires (ES) sont organisés en paquets pour former les Packetized Elementary Streams (PES) vidéo, audio et données privées.

Le train programme de MPEG-2 est créé d'un ou plusieurs PES qui doivent obligatoirement partager la même horloge de référence.

- Ce type de train est destiné à des applications où le dispositif de stockage ou de transmission est susceptible de n'introduire que peu d'erreurs (les applications multimédia sur CD-ROM ou disque dur).
- Ses paquets peuvent donc être relativement longs (par exemple 2048 octets).
- Ce type de multiplex MPEG-2 est utilisé sur le disque vidéo numérique (DVD).



**Figure 3.22.** Schéma conceptuel de la génération des trains programme et transport MPEG-2.

Le train transport de MPEG-2 est principalement destiné au transport de programmes TV à longue distance sur des supports ou des milieux susceptible d'introduire un taux d'erreurs assez élevé.

- La longueur des paquets doit être relativement courte pour permettre l'introduction de dispositifs de correction d'erreurs efficaces.
- La longueur des paquets transport de MPEG-2 a donc été fixée à 188 octets, valeur retenue notamment pour les émissions par satellite, câble ou terrestres à la norme européenne DVB.
- Ce type de train est destiné à combiner plusieurs programmes ne partageant pas forcément la même horloge système (STC) à l'intérieur d'un même multiplex.
- Les différents PES (vidéo, audio...) formant un programme donné doivent néanmoins partager la même horloge afin de pouvoir être synchronisés par le décodeur.

