

REPUBLIQUE DU CAMEROUN
Paix - Travail – Patrie

UNIVERSITE DE YAOUNDE I

ECOLE NATIONALE SUPERIEURE
POLYTECHNIQUE



REPUBLIC OF CAMEROUN
Peace - Work – Fatherland

UNIVERSITY OF YAOUNDE I

NATIONAL ADVANCED SCHOOL
OF ENGINEERING

MASTER PRO 2 EN TELECOMMUNICATIONS

TELEVISION NUMERIQUE

Séquence 3 : COMPRESSION VIDEO-AUDIO, MPEG2 FLUX DE TRANSPORT

Equipe des concepteurs :

- Janvier FOTSING
- Pierre TSAFACK

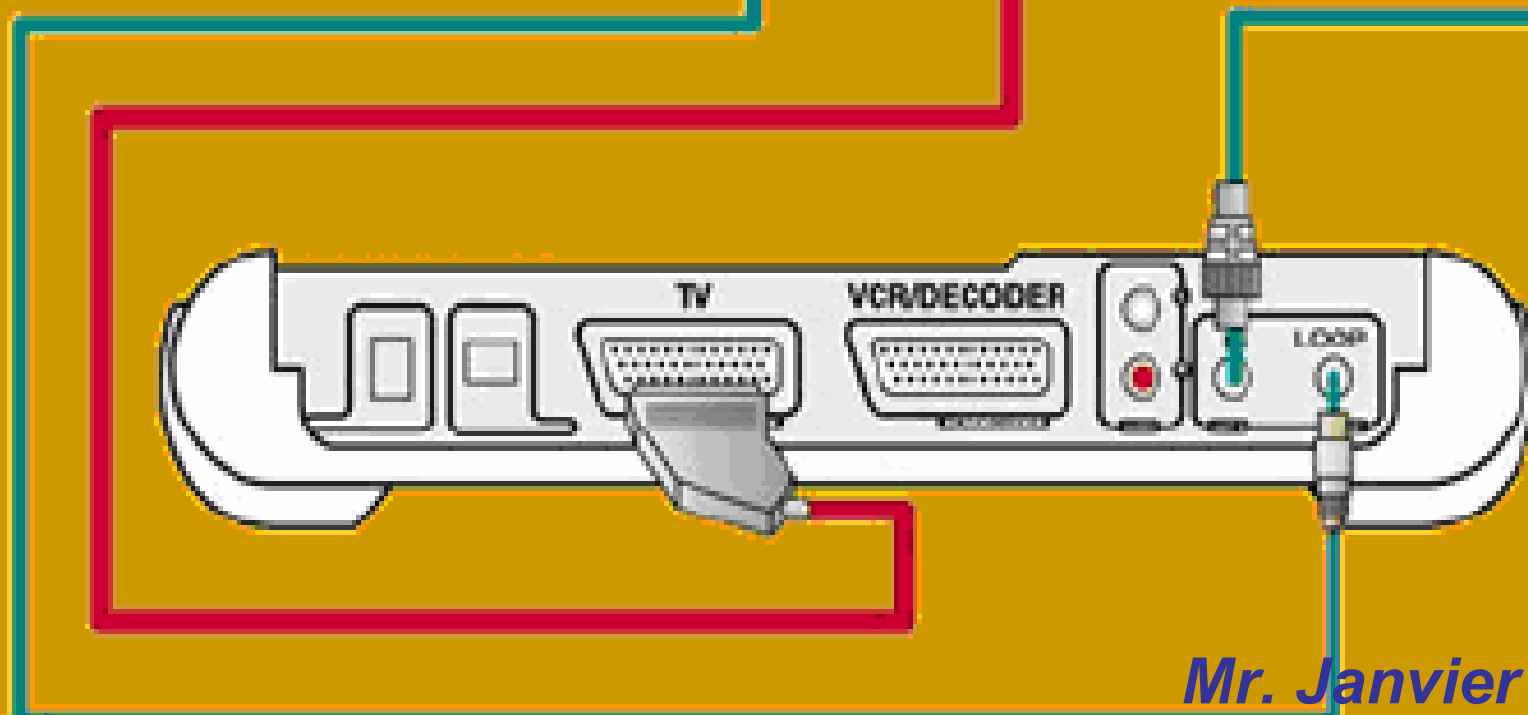
Le contenu est placé sous licence /creative commons/ de niveau 5 (Paternité, Pas d'utilisation commerciale, Partage des conditions initiales à l'identique)..



**Master
Professionalisant en
Télécommunications
(ENSP)**

Cours:

Télévision Numérique



Mr. Janvier FOTSING

Télévision Numérique

Cours: TVN 224

Compression vidéo-audio,
MPEG2 Flux de Transport

Master Professionnalisant en
Télécommunication

Plan de présentation

- Introduction
- Codage & multiplexage
- Codage vidéo
- Codage audio
- Le flux élémentaire
- Le packet Elementary Stream
- Le flux de programme et le flux de transport
- Télétex en DVB
- Services de télé et de données

Introduction

Le MPEG c'est un peu comme les poupées russes. Après avoir compressé le signal, on obtient un ES (Elementary Stream) qui va être découpé pour être encapsulé dans un PES (Packet Elementary Stream), lui même découpé et encapsulé dans des paquets. Chaque couche possède sa propre syntaxe.

De plus, groupe DVB a créé des extensions au MPEG pour transporter pratiquement toutes sortes de format de données (IP ou Open Tv par exemple) en plus de l'audio et la vidéo. les flux peuvent être cryptés et sont prévus pour être distribués par satellite, par câble ou par voie terrestre.

Autant dire que lorsque l'on parle MPEG et DVB, le champs d'action est vaste.

Codage & multiplexage

- *Codage*

Passage du signal audio ou vidéo en bande de base vers un format propre à être diffusé sur un DVB ou par sur un réseau type câble, satellite.

- Codage Vidéo
- Codage Audio
- Elementary Stream

- *Multiplexage*

Chaque composante est codée d'abord séparément puis regroupées lors du multiplexage pour faire un flux comportant toutes les informations du ou des programmes (vidéo, audio, données)

- ✓ Packet Elementary Stream
- ✓ Program Stream et Transport Stream

- *Les extensions*

Outre la vidéo et l'audio, les flux DVB peuvent transporter bien d'autres types de donnée :

- ☐ Le télétexte

Codage Vidéo

Le format vidéo de post production numérique utilise 270 Mbit/s de débit pour coder les images. Sachant qu'un canal satellite accepte autour de 38 Mbit/s (valeur la plus courante), on voit vite qu'il va falloir faire faire un sérieux régime à nos image.

En fait et après codage MPEG2, on arrivera à un débit utile de 3 à 8 Mbit/s pour un programme (audio + vidéo + données auxiliaires).

- ***Le grand principe du MPEG vidéo:***

"Ne jamais transmettre un élément d'image déjà transmis"

D'ou l'exploitation de la redondance temporelle :

Un élément commun à plusieurs images consécutives, n'est transmis qu'avec la première image. Pour les autres images, on ne transporte que sa position dans l'image. C'est le cas d'un plan fixe, où l'on voit une voiture qui va de gauche à droite. On transmet une fois le décor et la voiture et pour les images suivante on indique juste la position de la voiture.

Et de la redondance spatiale :

Si dans la même images, il y a trois voitures identique, elle n'est codée qu'une seule fois. En y ajoutant la position de chacune dans l'image, on peut reconstituer la scène.

Pour en savoir bien plus, lisez la suite.

Codage Vidéo

- *Codage spatial ou temporel*

Comme nous l'avons dit, la compression vidéo utilise les avantages des deux redondances (spatiale et temporelle). En MPEG, la redondance temporelle est d'abord réduite en utilisant les similitudes entre deux images successives. La plus grande partie possible de l'image courante est créée (ou prédictee) en utilisant l'information de l'image déjà émise. Quand on utilise cette technique, il suffit de transmettre une image de différence qui élimine les différences entre l'image actuelle et l'image de prédiction. L'image de différence est ensuite soumise à une compression spatiale. Pour des raisons pratiques, il est plus facile d'expliquer la compression spatiale avant d'aborder la compression temporelle.

La compression spatiale utilise la similarité entre des pixels adjacents sur une surface unie et tient compte des fréquences spatiales dominantes existant dans les zones amont. Le JPEG utilise uniquement la compression spatiale dans la mesure où ce système est conçu pour la transmission des images fixes. Le JPEG peut cependant être employé pour la transmission de séquences d'images fixes. Dans cette application appelée Motion JPEG, le facteur de compression n'est pas aussi bon que si l'on utilisait le codage temporel, mais il sera cependant possible d'effectuer un montage du flux de bits pour effectuer un montage image par image.

Codage Vidéo

- ***Codage spatial***

En codage spatial, la première étape consiste à effectuer une analyse de fréquence spatiale à l'aide d'une transformée. Une transformée est un outil mathématique permettant de traduire une forme d'onde en différents domaines, et, dans notre cas, dans le domaine fréquentiel. Le résultat d'une transformée est une suite de coefficients décrivant l'amplitude de chaque composante fréquentielle présente dans le signal. Une transformée inverse reproduit le signal initial. Si les coefficients sont gérés avec une précision suffisante, la sortie de la transformée inverse doit être identique à la forme d'onde originale.

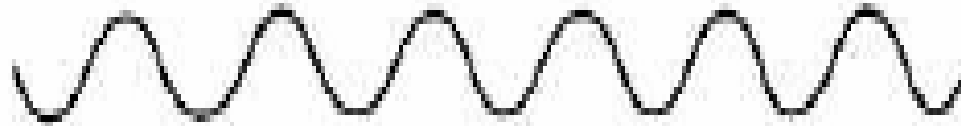
La transformée la plus répandue est la transformée de Fourier. Cette transformée cherche chaque fréquence comprise dans le signal d'entrée. Elle caractérise chaque fréquence en multipliant le signal d'entrée par un exemple de la fréquence cible appelée fonction de base, et en intégrant le produit obtenu. La figure ci-dessous montre que, lorsque la forme de signal d'entrée ne contient pas de composante à la fréquence cible, l'intégrale sera nulle mais, s'il en comporte une, l'intégrale constituera un coefficient caractérisant l'amplitude de cette composante.

Codage Vidéo

Signal
d'entrée

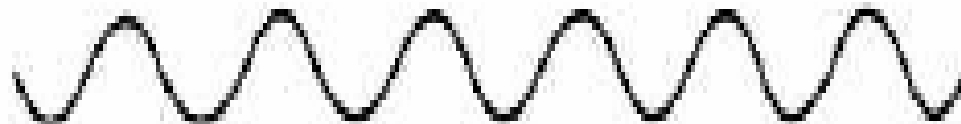


Fonction
de base

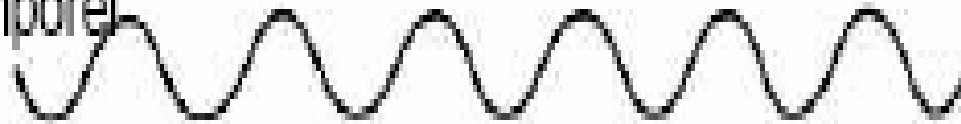


Pas de
corrélation
si la
fréquence
est
différente

Signal
d'entrée



Fonction
de base



Corrélation
élevée
si la
fréquence
est
identique

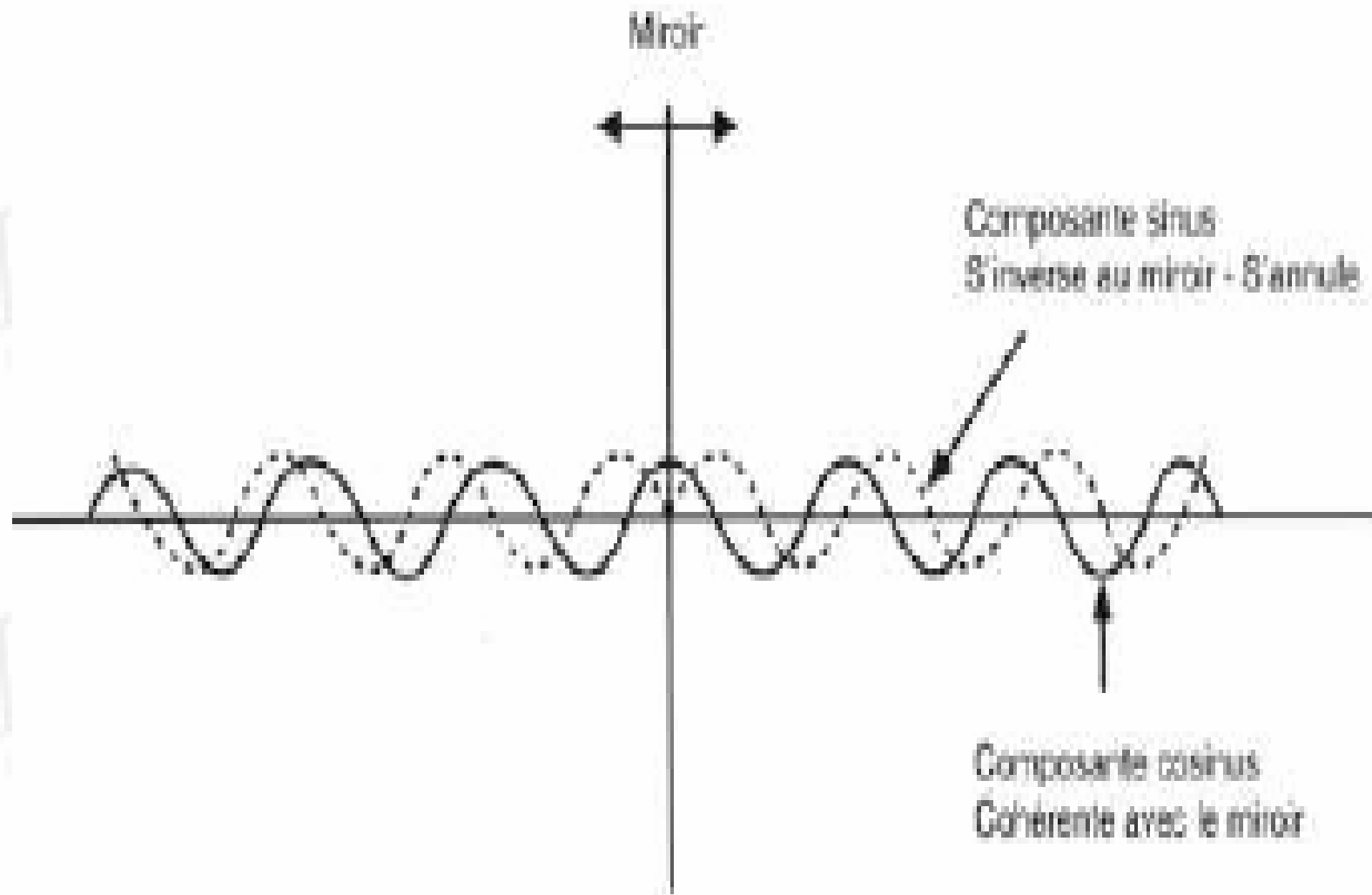
2.2 Codage spatial ou temporel

Codage Vidéo

- ***Codage spatial (suite)***

Le résultat sera tel qu'on vient de le décrire si la composante de fréquence est en phase avec la fonction de base. Cependant, si la composante de fréquence est en quadrature avec la fonction de base, l'intégrale sera quand même nulle. Il conviendra donc d'effectuer deux recherches avec deux composantes en quadrature, de la fonction de base de façon à bien détecter chaque composante du signal d'entrée. La transformée de Fourier présente l'inconvénient de nécessiter des coefficients pour les composantes sinus et cosinus de chaque fréquence. Dans la transformée cosinus, la forme de signal d'entrée est complétée avec son image temporelle avant multiplication par la fonction de base.

Codage Vidéo



Codage Vidéo

- ***Codage spatial (suite)***

L'illustration ci dessus montre que cette opération de « miroir » élimine toutes les composantes sinus et double les composantes cosinus. La fonction de base sinus ne se révèle donc plus utile et un seul coefficient reste alors nécessaire pour chaque fréquence. La Transformée Cosinus Discrète (DCT) est la version échantillonnée de la transformée cosinus; elle est énormément utilisée sous forme bi-dimensionnelle en MPEG. Le bloc de 8x8 pixels est changé en bloc de 8x8 coefficients. Comme la transformation réside en une multiplication par une fraction, il se produit un allongement du mot provenant du fait que les coefficients ont une longueur supérieure à celle des valeurs de pixels. Un bloc de pixels 8 bits devient alors un bloc de coefficients 11bits. Une DCT ne constitue pas alors une compression, mais elle obtient, en fait, le résultat inverse.

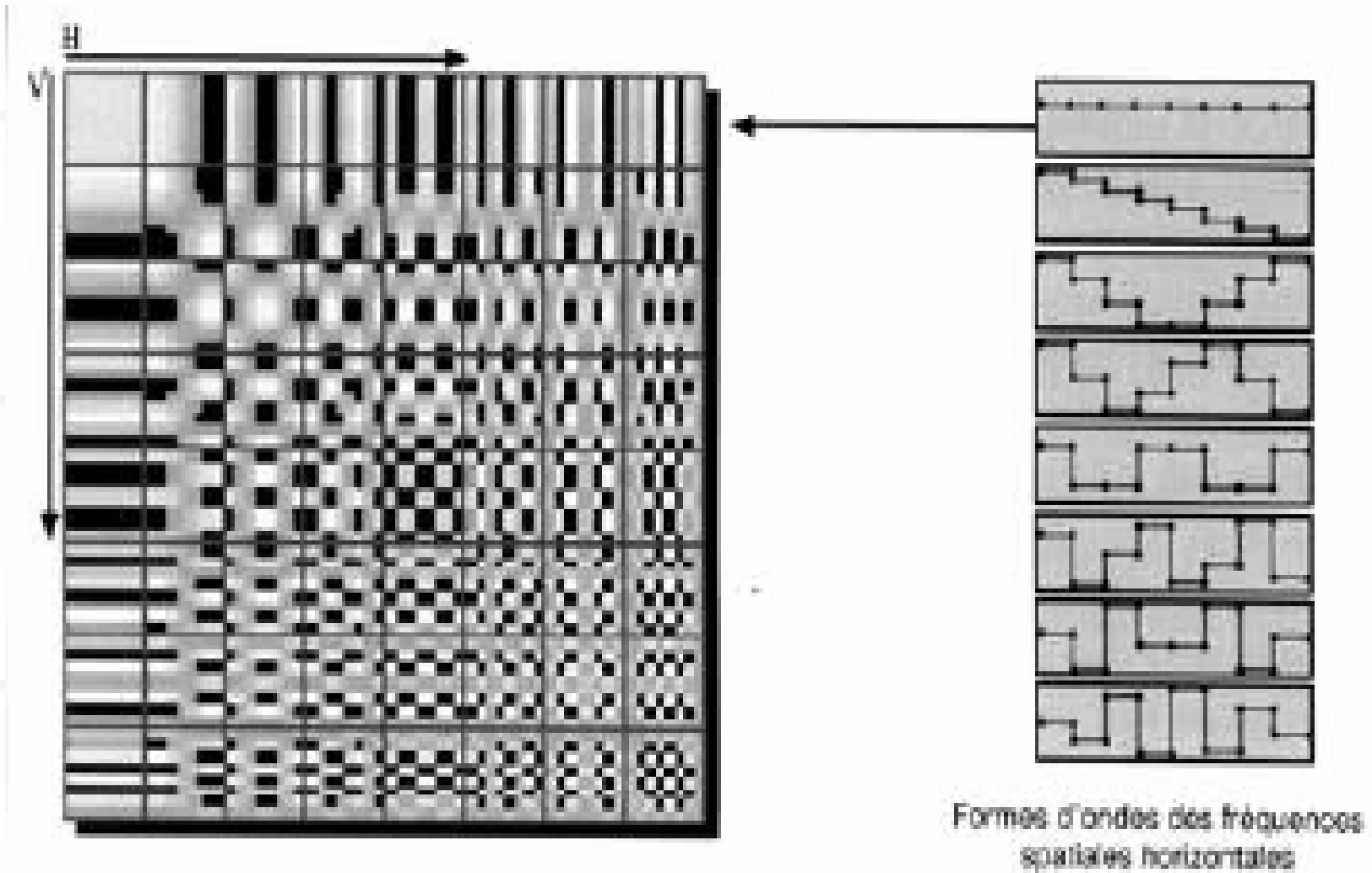
Codage Vidéo

Codage spatial (suite)

Cependant, la DCT convertit la source de pixels en une forme facilitant la compression. La figure suivante montre le résultat d'une transformée inverse des coefficients individuels d'une DCT pour un bloc 8x8. La case supérieure gauche représente la composante continue ou luminance moyenne pour l'ensemble du bloc. Aller vers la droite dans la rangée supérieure correspond à une augmentation de la fréquence spatiale horizontale. Descendre dans la colonne de gauche correspond à augmenter la fréquence spatiale verticale.

Dans les images réelles, diverses fréquences spatiales verticales et horizontales peuvent se produire simultanément et un coefficient en un certain point en représente toutes les combinaisons possibles.

Codage Vidéo



Codage Vidéo

Codage spatial (suite)

La figure montre également les coefficients sous une forme d'onde horizontale unidimensionnelle. La combinaison de ces formes d'onde avec diverses amplitudes et une polarité quelconque doit permettre de reproduire toute combinaison des 8 pixels. La combinaison des 64 coefficients de la DCT-2D permettra de reconstituer le bloc initial de 8x8 pixels. En ce qui concerne les images couleur, il est clair que les signaux de différence de couleur devront également être traités. Les signaux Y, Cr et Cb seront donc assemblés en réseaux séparés de 8x8 pixels et traités séparément.

Dans la plupart des signaux correspondant aux images de programme, la majorité des coefficients a une valeur nulle ou proche de zéro. Il ne sera donc point nécessaire de les transmettre. Il en résulte une compression non négligeable sans perte véritablement conséquente. Si un facteur de compression supérieur est nécessaire, la longueur des mots des coefficients non-nuls devra être réduite.

Cette réduction diminuera la précision de ces coefficients et introduira, par conséquent, des pertes au cours de ce traitement. Avec un peu de soin, on pourra introduire ces pertes de la façon la moins apparente possible pour le spectateur.

Codage Vidéo

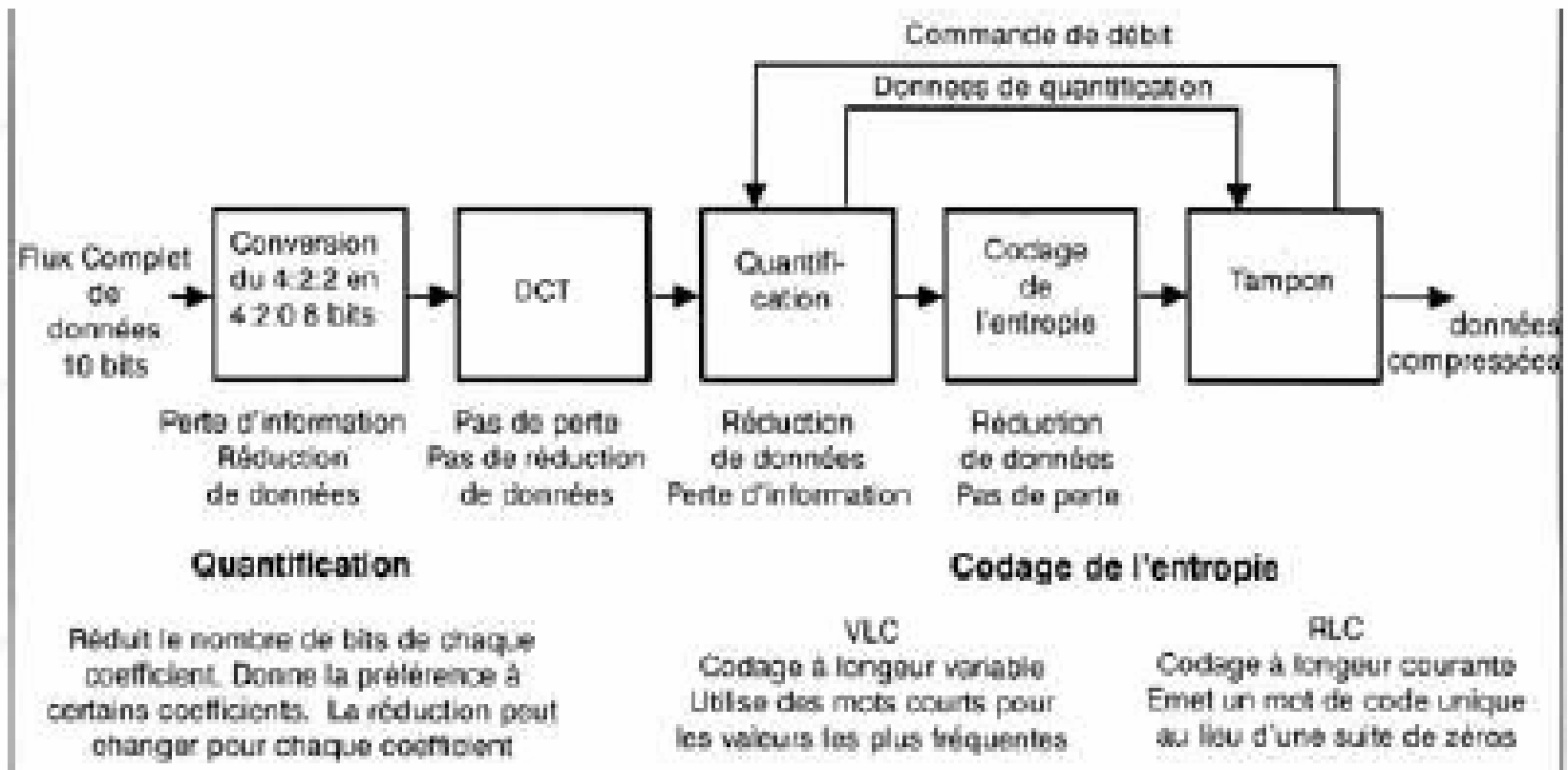
- ***Codage de l'entropie***

Dans une image vidéo animée, toutes les fréquences spatiales ne sont pas présentes simultanément; la matrice de coefficients de la DCT comportera donc des termes nuls. Malgré la scrutation, des termes nuls apparaîtront encore au milieu des coefficients non nuls. Le codage RLC (Run Length Coding) permet de gérer plus efficacement ces coefficients. Quand une suite de valeurs identique, comme des zéros, existe, le codage RLC émet simplement le nombre de zéros plutôt que toute la suite de bits nuls.

On peut étudier la probabilité de répétition de certaines valeurs particulières de coefficients dans la vidéo réelle. En pratique, certaines valeurs se rencontrent fréquemment et d'autres moins souvent. Cette information statistique peut être utilisée pour effectuer ultérieurement une compression à longueur variable (VLC). Les valeurs les plus fréquentes sont codées en mots de code courts et les autres en mots plus long. Pour faciliter la désérialisation on peut utiliser un mot de code comme préfixe pour les autres.

Codage Vidéo

- **Codeur spatial**



Codage Vidéo

- ***Codeur spatial***

Le signal d'entrée est supposé être à la norme 4:2:2 série en 8 ou 10 bits (SDI). Le MPEG n'utilise cependant qu'une résolution de 8 bits; aussi, un étage de traitement sera nécessaire pour arrondir les valeurs si le signal entrant est à 10 bits. Les profils MPEG effectuent un échantillonnage du type 4:2:0; un étage d'interpolation verticale/filtre passe-bas sera alors nécessaire. L'arrondi et le sous-échantillonnage couleur introduisent une légère mais irréversible perte d'information, mais aussi une réduction du débit. Le format d'entrée de la scrutation d'écran doit être tel qu'il puisse être converti en blocs de 8x8 pixels. L'étage DCT transforme l'information d'image dans le domaine fréquentiel. La DCT n'effectue pas de compression par elle même.

Codage Vidéo

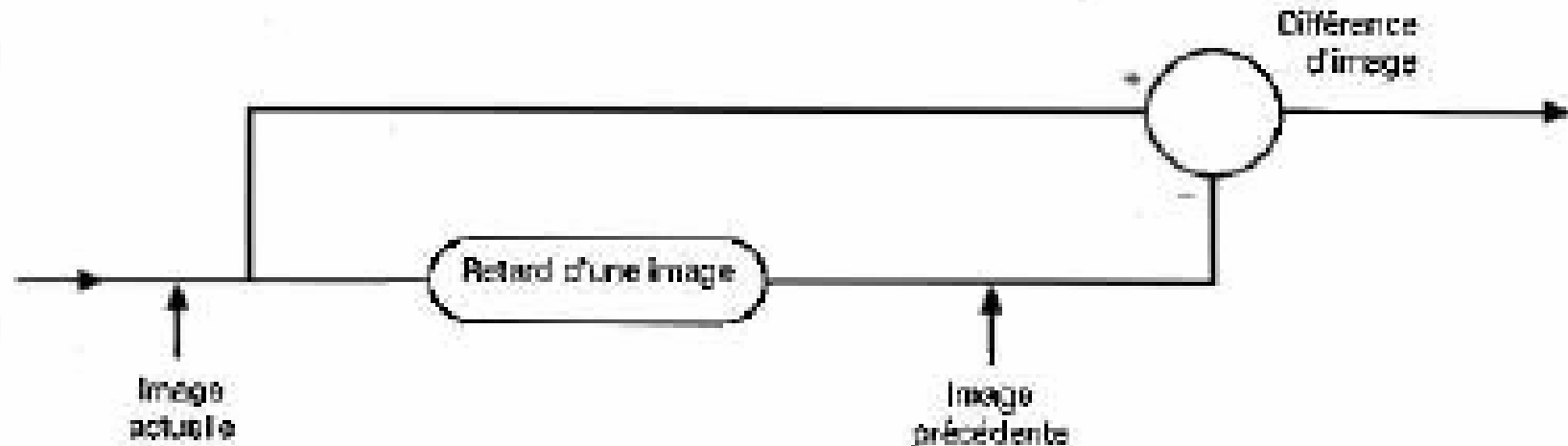
- ***Codeur spatial***

Dans le décodeur, le flux de bits est désérialisé et le codage de l'entropie est inversé pour reproduire les coefficients pondérés. La pondération inverse est ensuite appliquée et les coefficients sont replacés dans les matrices en fonction de la scrutation en zig-zag afin de recréer les matrices DCT. A la suite de la transformée inverse, le bloc de 8x8 pixels est recréé. Pour obtenir un signal de sortie d'écran balayé, les blocs sont stockés dans une RAM qui est lue ligne par ligne.

Codage Vidéo

- ***Codage temporel***

La redondance temporelle peut être exploitée par inter-codage, en transmettant uniquement les différences entre images. La figure suivante montre qu'un retard d'une image suivi d'un étage de soustraction permet de calculer la



Codage Vidéo

- ***Codage temporel***

La différence est une image en soi et elle peut être compressée en tant que telle par le codeur spatial comme cela a été précédemment décrit. Le décodeur inverse le codage spatial et ajoute l'image de différence à l'image précédente pour obtenir l'image suivante. Ce système simple présente quelques inconvénients.

D'abord, et comme seules les différences sont transmises, le décodage ne peut commencer dès le début de la transmission. Cette limitation rend difficile, pour le décodeur, la faculté de fournir une image après une commutation d'un flux de données vers un autre (ce qui se produit si l'observateur change de canal). De plus, et si une partie des données de différence est incorrecte, l'erreur persiste indéfiniment dans l'image. La solution à ces problèmes consiste à utiliser un système qui n'est pas totalement différentiel.

Codage Vidéo

- ***Codage bi-directionnel***

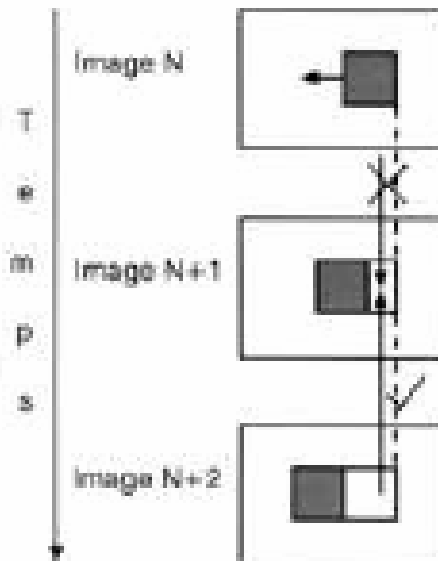
Quand un objet se déplace sur un fond, il cache le fond du côté qui avance et le révèle à l'opposé. Le fond qui est révélé nécessite la transmission de nouvelles données car cette partie de fond était auparavant masquée et il n'existe aucune information correspondante dans l'image précédente.

Un problème similaire se produit quand la caméra effectue un panoramique: de nouvelles zones apparaissent dans l'image et on ignore tout d'elles. Le MPEG aide à minimiser ce problème en utilisant un codage bidirectionnel qui permet de prendre des informations avant et après l'image courante. Si un fond est en cours de révélation, il sera présent dans une image future mais l'information peut être utilisée à posteriori pour recréer une image antérieure. La figure suivante montre le concept d'un codage bidirectionnel.

Codage Vidéo

- ***Codage bi-directionnel***

Sur la base d'un macrobloc, une image codée bidirectionnellement peut obtenir les données compensées en mouvement de la part d'une image antérieure ou d'une image ultérieure ou même utiliser un moyennage des deux ensembles de données. Le codage bidirectionnel réduit considérablement la quantité de données de différence nécessaire à l'amélioration du débit.



Codage Vidéo

- ***Images I, P et B***

En MPEG, trois différents types d'image sont nécessaires pour effectuer le codage différentiel et le codage bidirectionnel avec un minimum d'erreurs de propagation: Les images I sont intra-codées et ne nécessitent pas d'information supplémentaires pour être décodées.

Elles nécessitent beaucoup de données comparativement aux autres types d'image et c'est pourquoi elles ne sont transmises que quand cela est nécessaire. Elles consistent essentiellement en coefficients de transformée et n'ont pas de vecteur de mouvement. Elles autorisent la commutation de voies et bloquent la propagation des erreurs.

Codage Vidéo

- **Images I, P et B**

Les images P sont celles qui sont déduites d'une image antérieure qui peut être de type I ou P. Les données d'une image P sont constitués de vecteurs décrivant où chaque macrobloc doit être pris dans l'image précédente et des coefficients non transformés décrivant la correction ou les données de différence à ajouter à ce macrobloc. Les images P comportent pratiquement la moitié des données d'une image I.

Les images B sont prédites bidirectionnellement à partir d'images antérieures ou postérieures et de type I ou P.

Les données des images de type B consistent en vecteurs décrivant l'endroit où les données doivent être prises dans les images antérieures ou postérieures. Elles contiennent également les coefficients de transformée fournissant la correction

Codage Vidéo

- ***Un compresseur MPEG***

Les figures a, b et c montrent une structure bidirectionnelle typique de compensateur de mouvement. Après avoir subi un pré-traitement, la vidéo entre dans une série de mémoires à image susceptibles d'être contournées pour changer l'ordre des images. Les données entrent dans l'étage de soustraction et le calculateur de mouvement. Pour créer une image I (reportez-vous à la figure a), la sortie du retardateur d'entrée est sélectionnée et le soustracteur inhibité de façon que les données « by-passent » directement vers le codeur spatial.

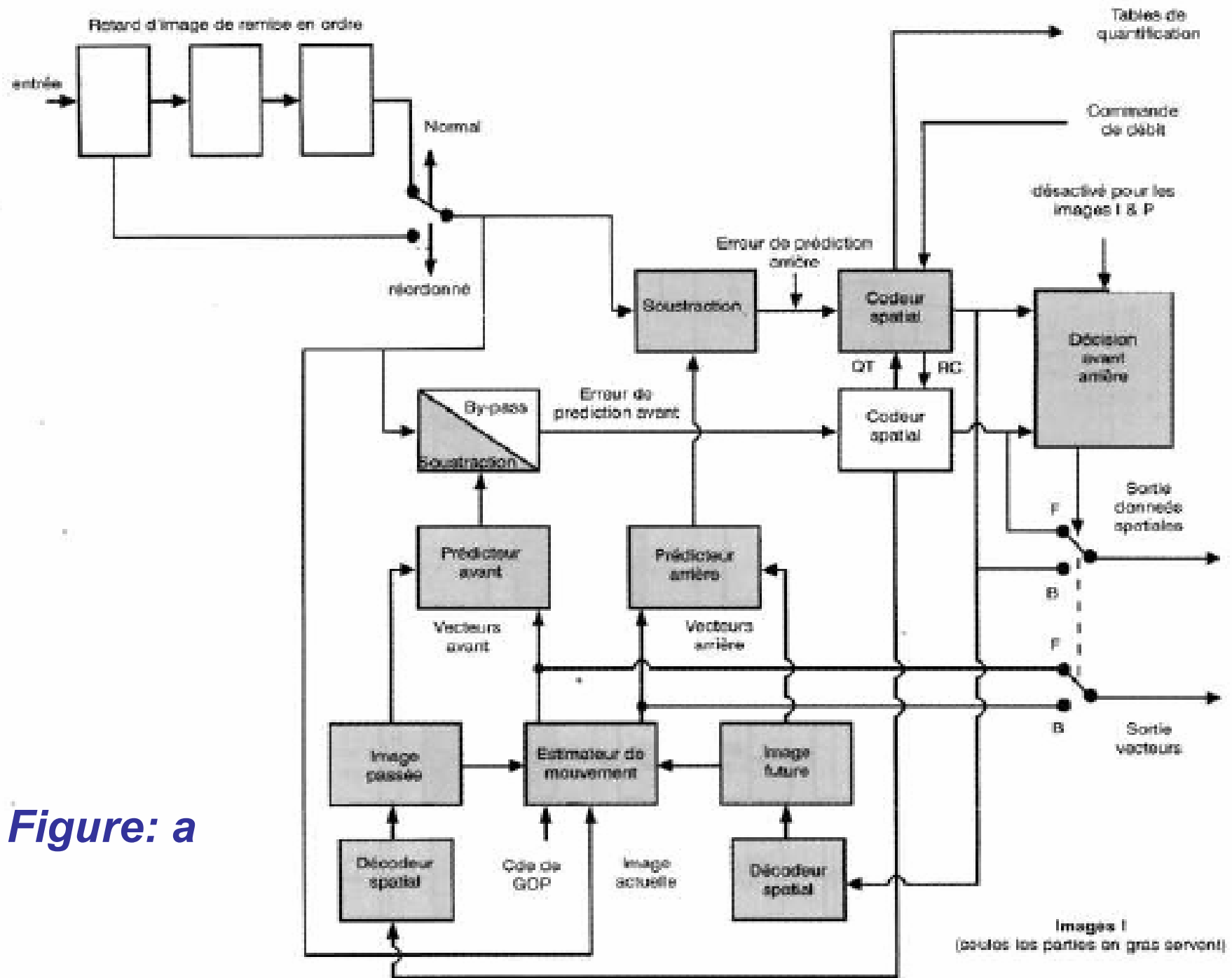


Figure: a

Codage Vidéo

- ***Un compresseur MPEG***

La sortie du soustracteur alimente également une mémoire d'image capable de stocker quelques images. L'image I est conservée en mémoire.

Pour créer une image P (reportez-vous à la figure b), les tampons d'entrée sont by-passés de façon à prendre en compte l'image future. Le calculateur de mouvement compare l'image I présente dans la mémoire de sortie et l'image P de la mémoire d'entrée pour générer les vecteurs de mouvement avant. L'image I est déplacée par ces vecteurs pour réaliser une image prédictee P. Cette image prédictee P est soustraite de l'image P courante pour générer l'erreur de prédiction qui est codée spatialement et émise conjointement avec les vecteurs. L'erreur de prédiction est également ajoutée à l'image prédictee P pour créer une image P décodée localement et sauvegardée également dans la mémoire de sortie.

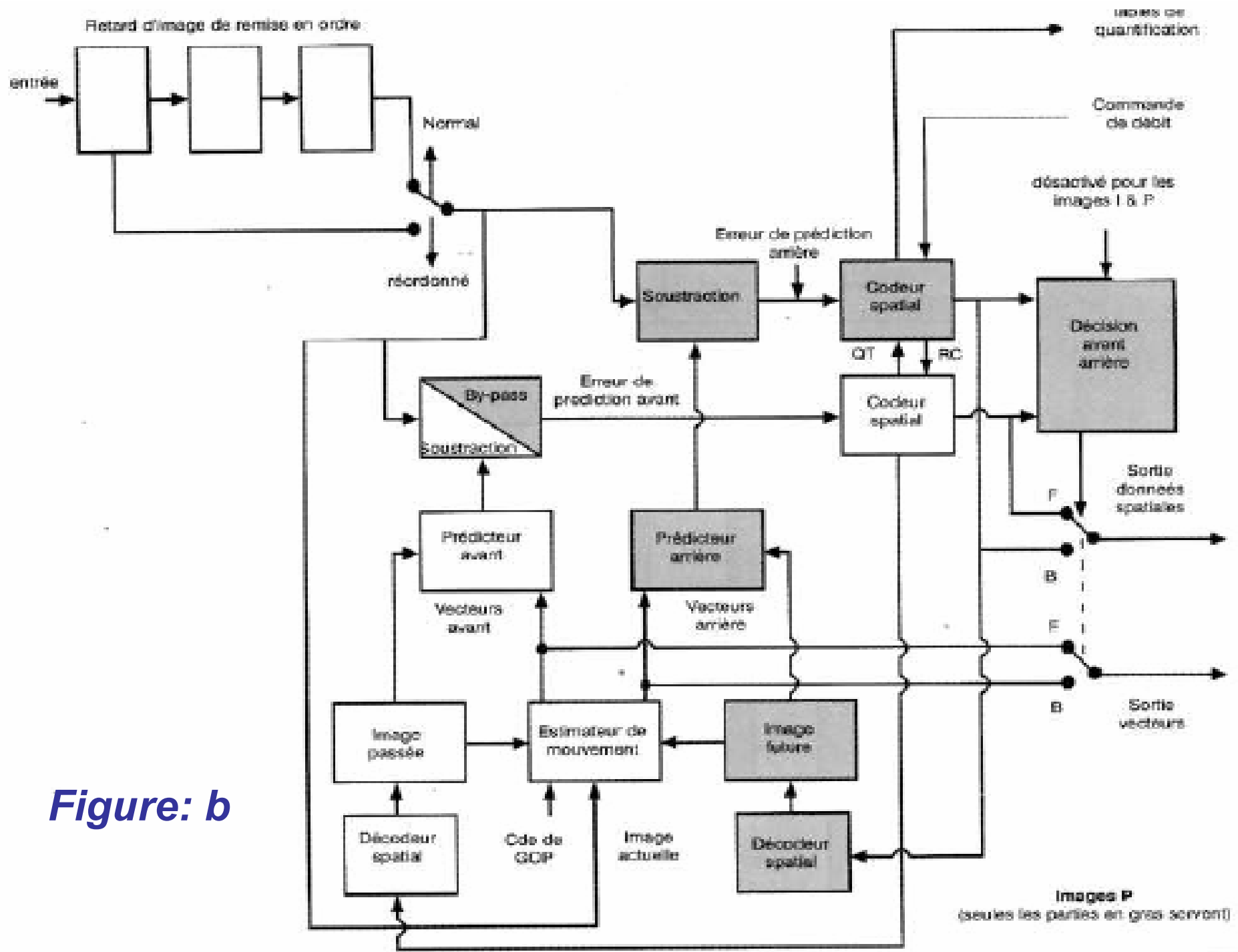


Figure: b

Images P
(seules les parties en gras seront)

Codage Vidéo

- ***Un compresseur MPEG***

Pour créer une image B (reportez- vous à la figure c), rappelez- vous que la mémoire de sortie contient une image I et une image P. On peut maintenant sélectionner une image B sur le tampon d'entrée. Le compensateur de mouvement compare l'image B à l'image I qui précède et l'image P qui suit pour déterminer les vecteurs bi-directionnels.

Les compensation de mouvement avant arrière sont réaliser de façon à créer deux images prédites B. Elles sont soustraites de l'image B courante. Macrobloc par macrobloc, les données avant et arrière sont choisies en fonction de leur moindre différence. les différences d'image sont alors codées spatialement et émises conjointement avec les vecteurs.

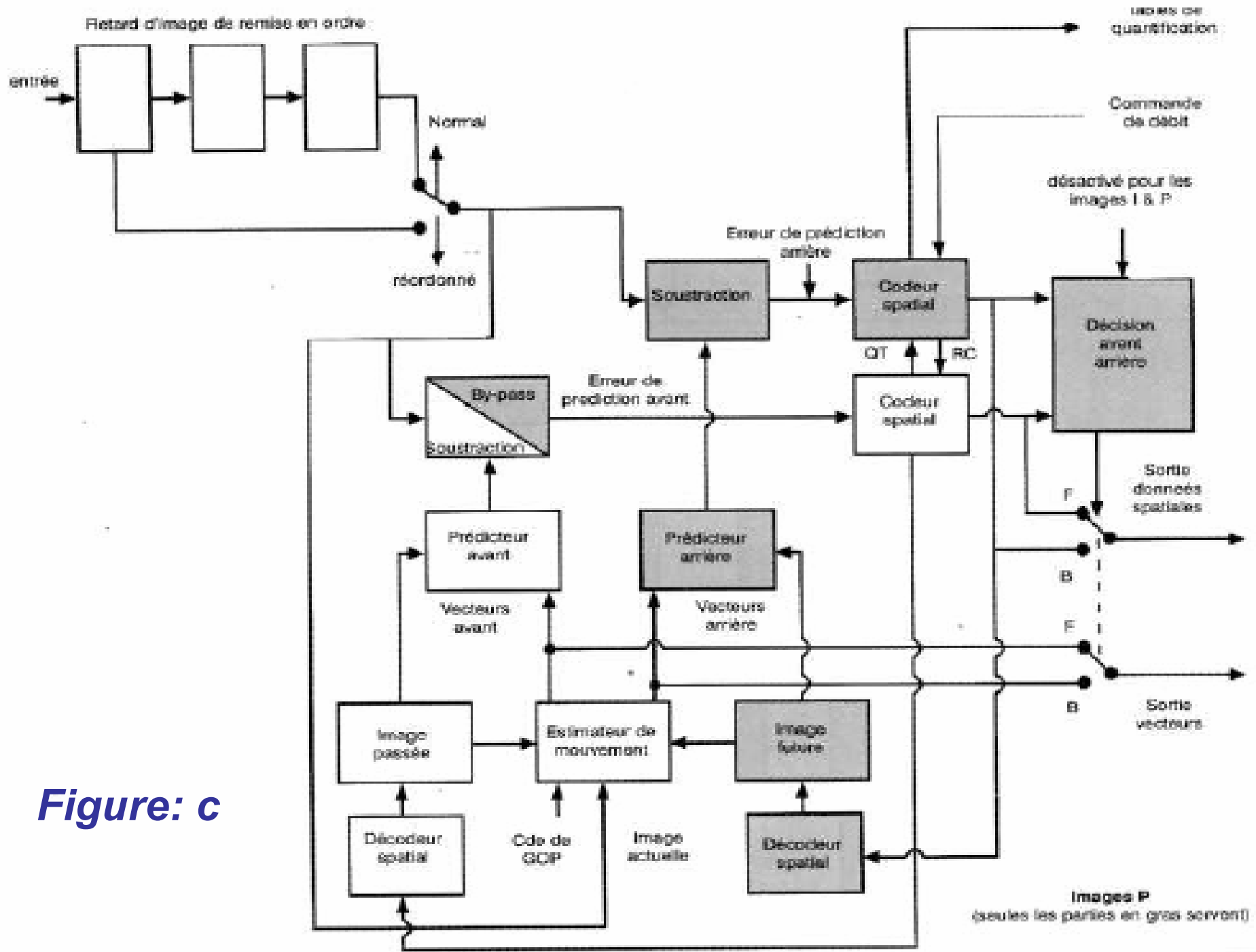
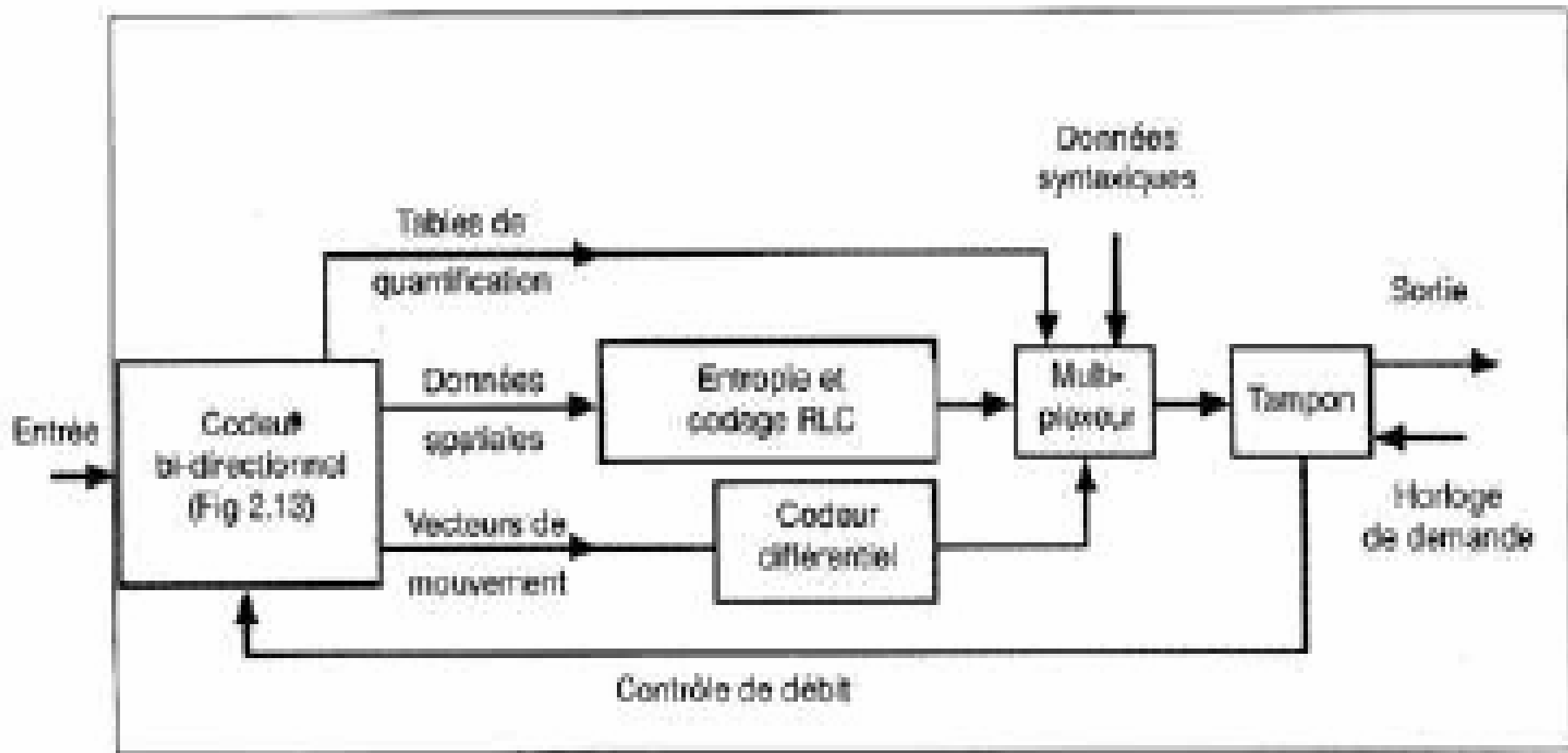


Figure: c

Codage Vidéo

- *Synoptique d'un codeur MPEG*



Codage Vidéo

- ***Profils et niveaux***

Le MPEG peut s'utiliser dans diverses applications nécessitant des performances et des complexités diverses. A l'aide des outils de codage définis dans le MPEG, il existe des milliers de combinaisons possibles. Dans un but de simplification, le MPEG est divisé en Profils, chaque profil étant lui-même subdivisé en Niveaux . Un profil constitue à la base la palette des caractéristiques d'un codage d'une certaine complexité. Un niveau est en fait un paramètre définissant par exemple la taille de l'image ou le débit du flux de bits.

Le profil simple n'a été seulement défini qu'au niveau Principal (Simple Profile at Main Level = SP@ML).

Codage Vidéo

- ***Profils et niveaux***

La plupart des applications de diffusion nécessite le Profil principal au Niveau Principal (Main Profile at Main Level = MP@ML), appellation du MPEG utilisé en télévision standard.

Un opérateur de télévision numérique a le soucis économique de diffuser un maximum de programme en un minimum de débits (comptez 20 millions par an pour louer un transpondeur) tout en gardant une qualité optimale. Une simple réduction de débit ferait apparaître, sans aucun doute, toutes la panoplie des artefact du MPEG.

Codage Vidéo

- ***Multiplexage statique***

En Multiplexage Statistique, le débit d'un codeur n'est plus fixe mais alloué dynamiquement en fonction de la complexité de l'image à coder.

En pratique on alloue un débit total à un ensemble de codeur qui attaque le même multiplexeur. A chaque image, chaque codeur calcule la complexité de l'image courante et fait une demande d'allocation de débit au gestionnaire du multiplexage statistique. Celui ci examine toutes les demandes et renvoie à chaque codeur, la valeur du débit qu'il peut utiliser. Ce système optimise l'utilisation du débit dans un multiplex et permet une amélioration globale des programmes. Son défaut réside en un léger retard entre une demande de débit et l'allocation de ce débit. Ce défaut est compensé par les codeurs double passe.

Encodage: Choix du codec

- ***Principaux Codecs vidéos***

1. MPEG-1, qui n'est plus utilisé, à moins de posséder quelques VCD.
2. MPEG-2, qui est le codec utilisé notamment par les DVD vidéo standard et la télévision numérique.
3. Les implémentations du MPEG-4 ASP, comme DivX et XviD
4. MJPEG, pour le montage vidéo image par image
5. MJPEG2000, évolution de MJPEG pour le cinéma numérique, l'imagerie de précision (médicale, spatiale...)
6. Les implémentations de MPEG-4 AVC / H.264 : X264, ATEME.

Codage Audio

La compression audio qui est entièrement basée sur les caractéristiques de l'audition humaine, s'accompagne de pertes et ne peut être abordée sans une étude préalable de ce sens. Il est surprenant de constater que l'audition humaine, et particulièrement en stéréo, a un pouvoir discriminatoire bien supérieur à celui de la vision et c'est pourquoi la compression audio doit être envisagée avec encore plus de précautions.

Comme la compression vidéo, la compression audio nécessite plusieurs niveaux de complexité en fonction du facteur décompression souhaité.

Codage Audio

- **Codage en sous-bandes**

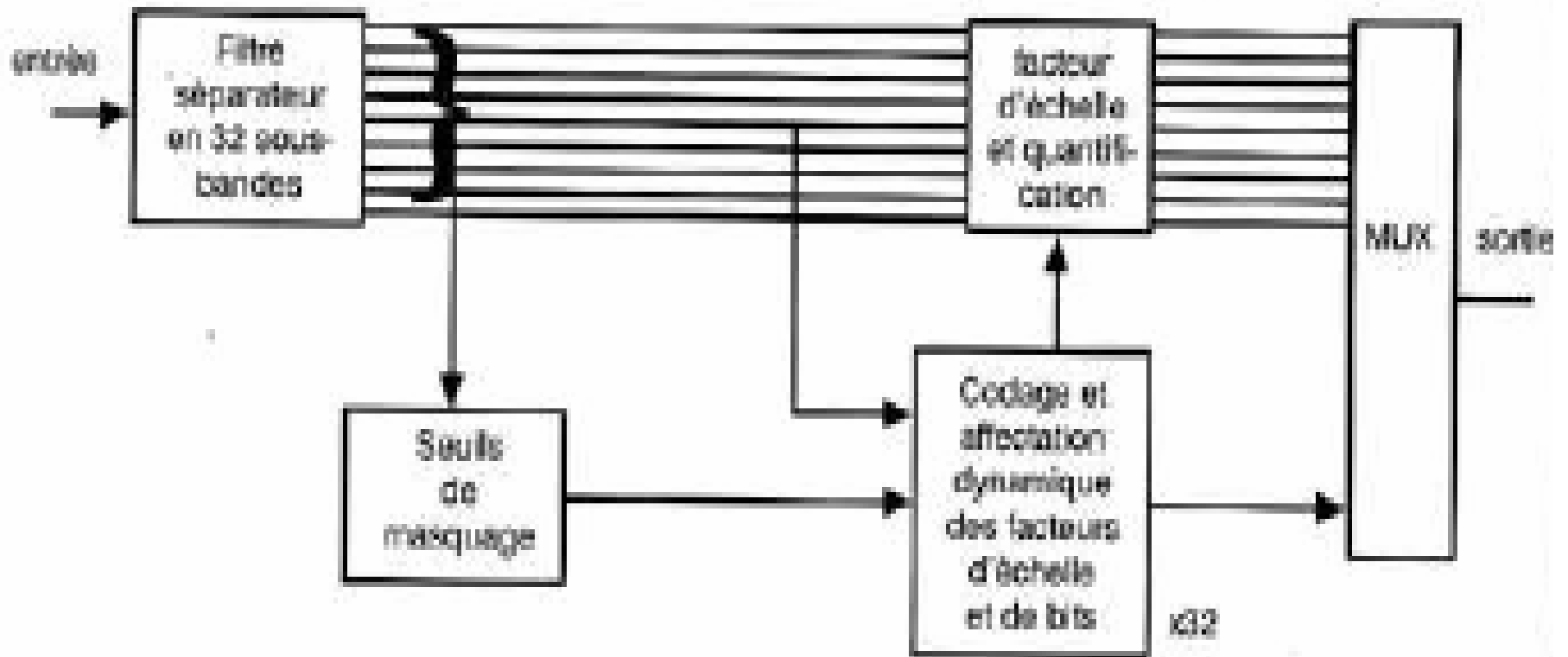
La figure suivante montre un compresseur à bandes séparées. Le filtre séparateur de bandes est un jeu de filtres à phase linéaire, ayant tous la même largeur de bande et qui se recouvrent. La sortie de chaque bande consiste en des échantillons représentatifs de la forme d'onde. Dans chaque bande de fréquence, l'entrée audio est amplifiée au maximum avant la transmission. Chaque niveau est ensuite ramené à sa valeur initiale.

Le bruit introduit par la transmission est ainsi réduit dans chaque bande. Si l'on compare la réduction de bruit au seuil d'audition, on s'aperçoit qu'un bruit plus important peut être toléré dans certaines bandes du fait de l'action du masquage. Par conséquent, il est possible, dans chaque bande, de réduire la longueur des mots d'échantillons après la compression.

Cette technique réalise une compression parce que le bruit introduit par la perte de résolution est masqué. La figure ci dessous montre un codeur simple à bandes séparées, comme ceux utilisés dans la Couche 1 du MPEG. L'entrée audio numérique alimente un filtre de séparations de bandes qui divise le spectre du signal en un certain nombre de bandes.

Codage Audio

- Codage en sous-bandes***



Codage Audio

- ***Codage en sous-bandes***

En MPEG, ce nombre est de 32. L'axe des temps est divisé en blocs d'égale longueur. Dans la couche 1 de MPEG, il y a donc 384 échantillons du signal d'entrée, ce qui se traduira, en sortie du filtre, par 12 échantillons dans chacune des 32 bandes. A l'intérieur de chaque bande, le niveau est amplifié par multiplication jusqu'à sa valeur maximale. Le gain nécessaire est constant pour la durée du bloc et un seul facteur d'échelle est transmis avec chaque bloc pour chaque bande de façon à pouvoir renverser le processus au décodage.

La sortie du groupe de filtres est également analysée afin de déterminer le spectre du signal d'entrée. Cette analyse permet de réaliser un modèle de masquage permettant de déterminer le degré de masquage que l'on peut attendre dans chaque bande.

Codage Audio

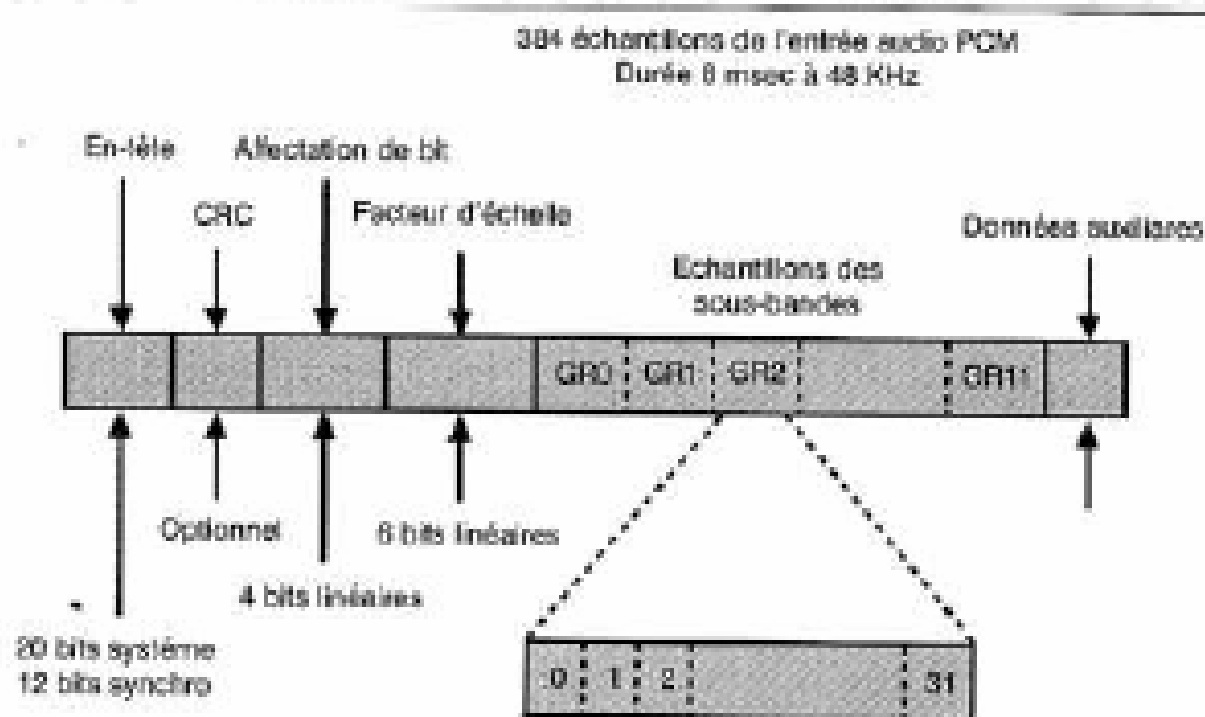
- ***Codage en sous-bandes***

Dans chaque bande, plus le masquage est agissant, moins l'échantillon doit être précis. La précision d'échantillon est alors réduite par requantification en vue de diminuer la longueur des mots. Cette réduction est aussi constante pour chaque mot dans la bande, mais les différentes bandes peuvent utiliser des longueurs de mot différentes. La longueur de mots doit être transmise comme un code d'affectation de bits afin de permettre au décodeur de désérialiser convenablement le flux de

Codage Audio

- **Couche 1 du MPEG**

La figure suivante montre un flux de bits audio MPEG Niveau 1.



Codage Audio

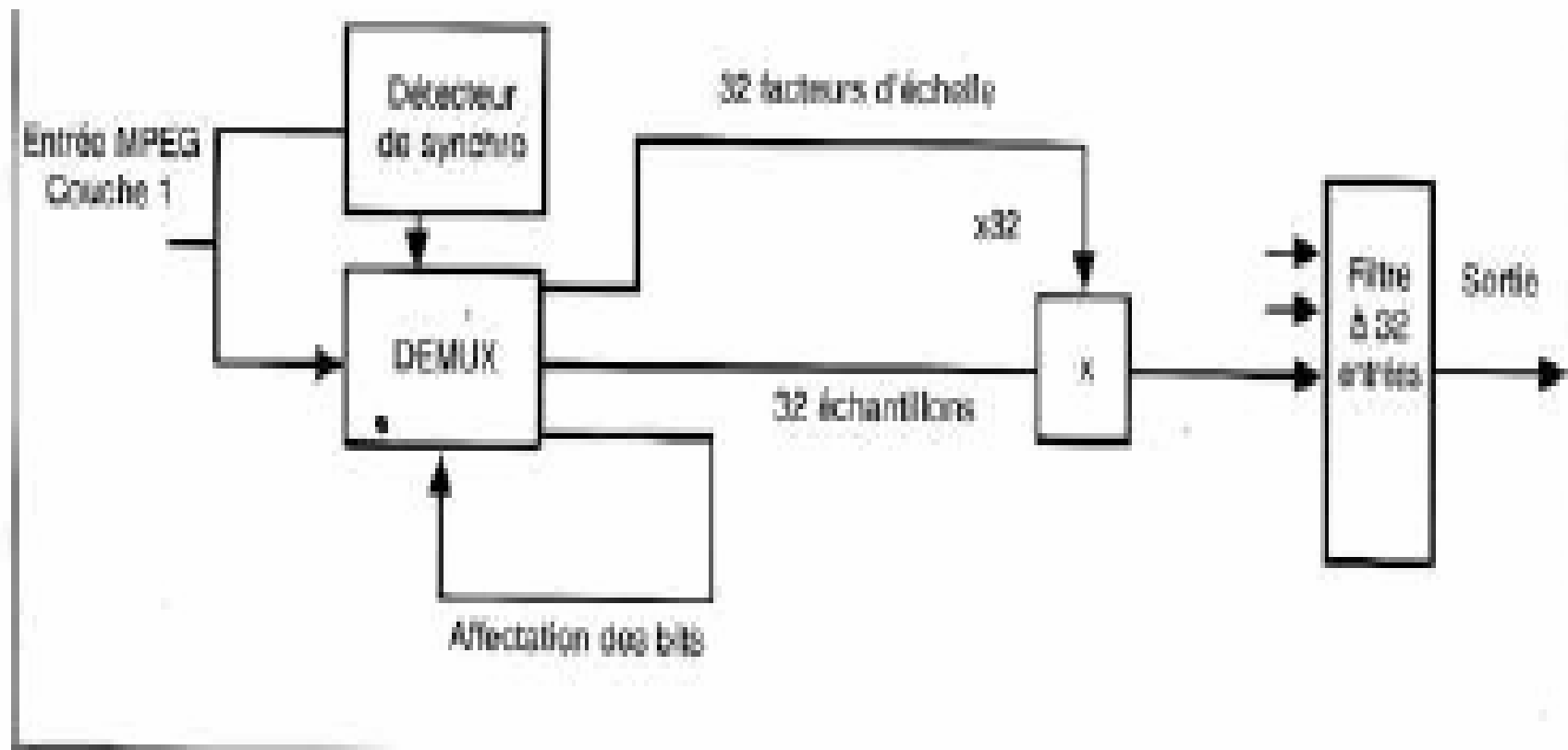
- ***Couche 1 du MPEG***

Après le mot de synchronisation et l'en-tête, il y a 32 codes d'affectation de bits de 4 bits chacun. Ces codes décrivent la longueur des mots des échantillons dans chaque sous-bande. Viennent ensuite les 32 facteurs d'échelle utilisés par la compression dans chaque bande. Ces facteurs d'échelle sont indispensables pour rétablir le bon niveau au décodage. Les facteurs d'échelle sont suivis des données audio de chaque bande.

Voici un décodeur du type Couche 1.

Codage Audio

- **Couche 1 du MPEG**



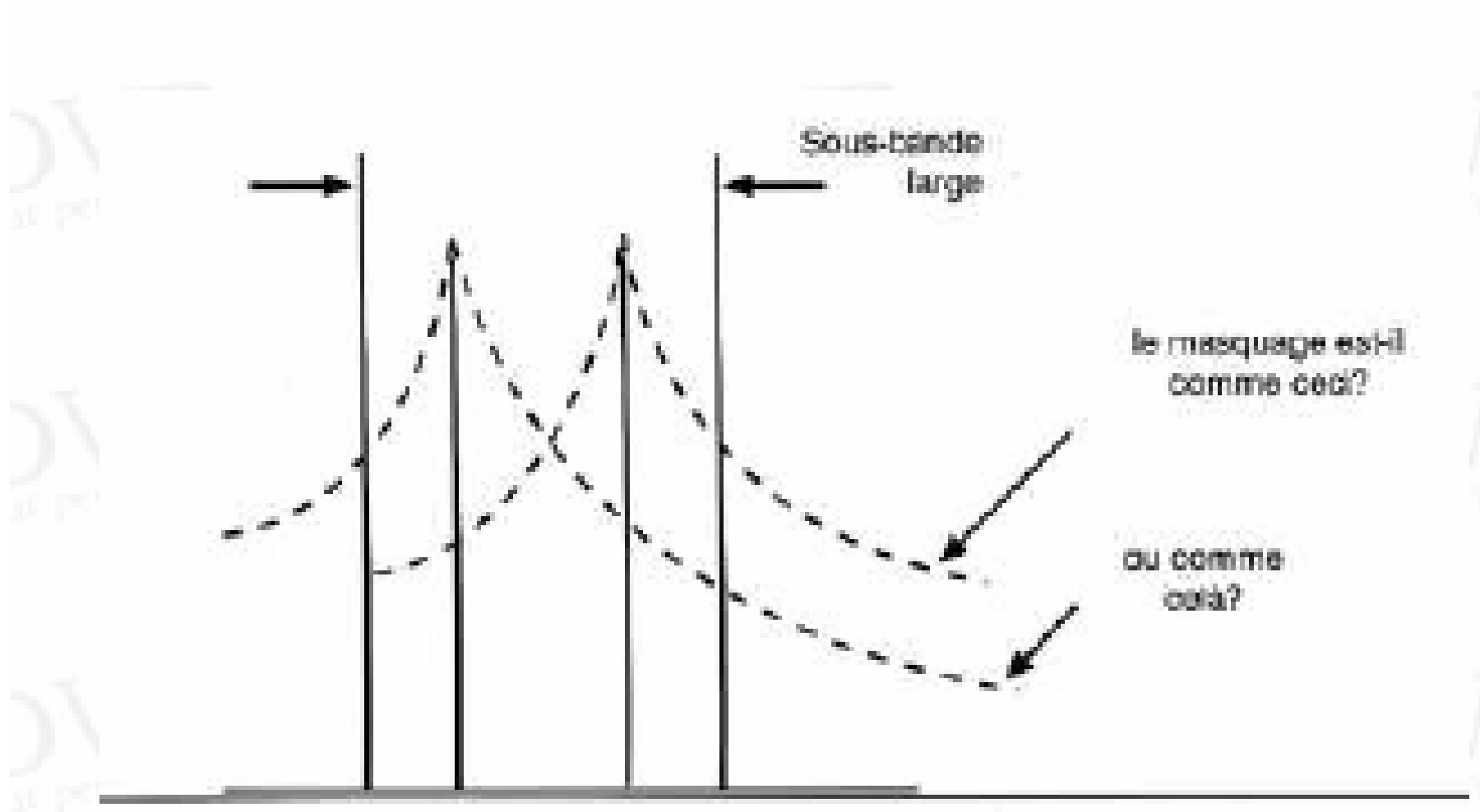
Codage Audio

- ***Couche 1 du MPEG***

Le mot de synchronisation est détecté par le générateur de temps qui déséréalise les bits d'affectation et les données de facteur d'échelle. L'affectation de bits permet ensuite la désérialisation des échantillons à longueur variable. La requantification inverse et la multiplication par l'inverse du facteur de compression sont appliquées de façon à ramener le niveau de chaque bande à sa bonne valeur. Les 32 bandes sont ensuite rassemblées dans un filtre de recombinaison pour rétablir la sortie audio.

Codage Audio

- **Couche 2 du MPEG**



Codage Audio

- Couche 2 du MPEG

La figure ci-dessus montre que, lorsque le filtre de séparation de bandes est utilisé pour créer le modèle de masquage, l'analyse de spectre n'est pas très précise dans la mesure où il n'y a que 32 sous-bandes et que l'énergie est répartie dans la totalité de la bande. On ne peut pas trop augmenter le plancher de bruit car, dans le pire des cas, le masquage n'agirait pas. Une analyse spectrale plus précise autoriserait un facteur de compression plus élevé. Dans la couche 2 du MPEG, l'analyse spectrale est effectuée à l'aide d'un processus séparé.

Codage Audio

- **Couche 2 du MPEG**

Une FFT à 512 points est effectuée directement à partir du signal d'entrée pour le modèle de masquage. Pour améliorer la précision de la résolution de fréquence, il faut augmenter l'excursion temporelle de la transformée, ce qui est effectué en portant à 1152 échantillons la taille du bloc.

Bien que le synoptique de la compression de bloc soit identique à celui de la couche 1 du MPEG, tous les facteurs d'échelle ne sont pas transmis dans la mesure où, dans les images de programme, ils présentent un degré de redondance non négligeable. Le facteur d'échelle de blocs successifs excède 2 dB dans moins de 10 % des cas et on a avantage à tirer parti de cette caractéristique en analysant les groupes de 3 facteurs d'échelle successifs.

Sur les programmes fixes, seul un facteur d'échelle sur trois est transmis. A mesure de l'augmentation de la variation dans une bande donnée, deux ou trois facteurs d'échelle sont transmis. Un code de sélection est également transmis pour permettre au décodeur de déterminer ce qui a été émis dans chaque bande. Cette technique permet de diviser par deux le débit du facteur d'échelle.

Codage Audio

- ***Couche 2 du MPEG: Codage à transformée***

Les couches 1 et 2 du MPEG sont basées sur les filtres séparateurs de bande dans lesquels le signal est toujours représenté comme une forme d'onde. La couche 3 utilise de son côté un codage à transformée comme celui utilisé en vidéo. Comme indiqué plus haut, l'oreille effectue une espèce de transformée sur le son incident et, du fait du facteur de résonance Q de la membrane basilaire, la réponse ne peut augmenter ou diminuer rapidement.

Par conséquent, si un signal audio est transformé dans le domaine fréquentiel, il n'est plus nécessaire de transmettre les coefficients trop souvent. Ce principe constitue la base du codage à transformée. Pour des facteurs de compression plus élevés, les coefficients peuvent être requantifiés, ce qui les rend moins précis. Ce processus génère du bruit qui pourra être placé à des fréquences où le masquage est le plus fort. Une caractéristique secondaire d'un codeur à transformée est donc que le spectre d'entrée est connu très précisément, ce qui permet de créer un modèle de masquage très fidèle.

Codage Audio

- **Couche 3 du MPEG**

Ce niveau complexe de codage n'est en réalité utilisé que lorsque les facteurs de compression les plus élevés sont nécessaires. Il comporte quelques points communs avec la couche 2.

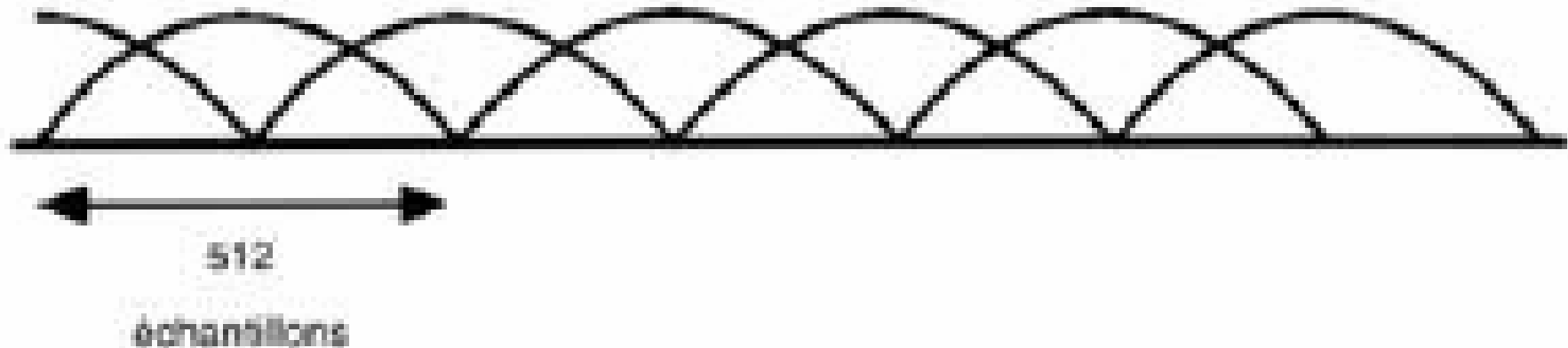
Une transformée cosinus discrète à 384 coefficients de sortie par bloc est utilisée. On peut obtenir ce résultat par un traitement direct des échantillons d'entrée mais, dans un codeur multi-niveau, il est possible d'utiliser une transformée hybride incorporant le filtrage 32 bandes des couches 1 et 2. Dans ce cas, les 32 sous-bandes du filtre QMF (Quadrature Mirror Filter) sont ensuite traitées par une Transformée Cosinus Discrète Modifiée (MDCT) à 32 bandes pour obtenir les 384 coefficients.

Deux tailles de fenêtres sont utilisées pour éviter les pré-oscillations aux transitions. La commutation de fenêtre est commandée par le modèle psycho-acoustique. Il a été trouvé que le pré-écho n'apparaissait dans l'entropie que quand elle était supérieure au niveau moyen. Pour obtenir le facteur de compression le plus élevé, une quantification non-uniforme des coefficients est effectuée selon le codage de Huffman. Cette technique attribue les mots les plus courts aux valeurs de code les plus fréquentes.

Codage Audio

- **Codage AC-3**

La technique de codage audio AC-3 est utilisée avec le système ATSC à la place d'un des systèmes de codage audio MPEG. Le système AC-3 est basé sur une transformée et obtient le gain de codage en requantifiant les coefficients de fréquence. L'entrée PCM d'un codeur AC-3 est divisée en blocs par des fenêtres qui se chevauchent comme indiqué ci dessous.



Le flux élémentaire

Première couche syntaxique du MPEG, l'ES (Elementary Stream) est le signal brut en sortie de compression. Dans la réalité, on ne travaille jamais directement avec ce type de signal. Il reste caché au cœur des codeurs.

Un flux élémentaire ne contient rien de plus que ce qui est nécessaire au décodeur pour s'approcher au mieux de l'image. En MPEG, la syntaxe du signal compressé est strictement définie de façon à être sûr que les décodeurs puissent l'interpréter. Le codeur n'a pas d'autre définition que l'obligation d'émettre selon la syntaxe convenable. L'avantage de ce concept est qu'il s'approche du monde réel dans lequel il y a vraisemblablement plus de décodeurs que de codeurs.

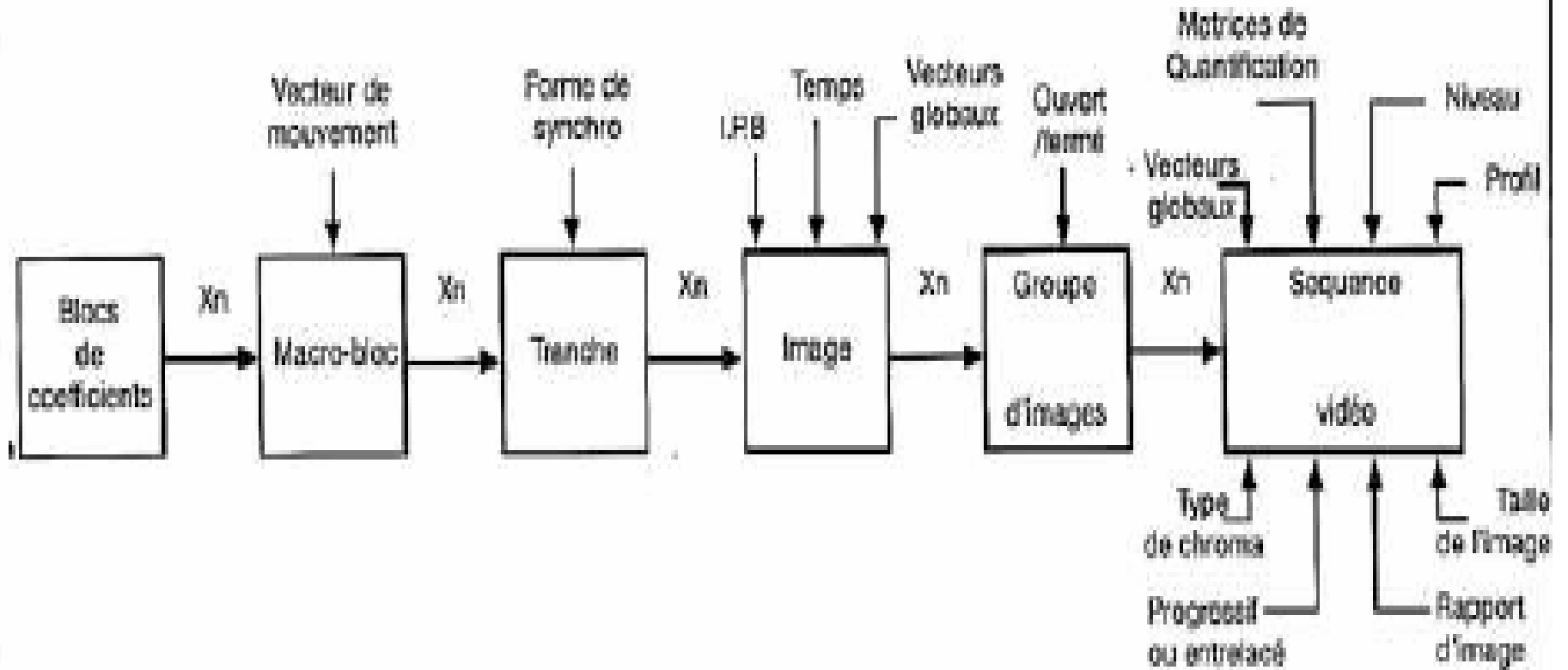
En standardisant le décodeur, on peut diminuer sensiblement le coût de sa fabrication. Inversement, le codeur peut être bien plus compliqué et plus onéreux et il peut procurer une forte amélioration de l'image sans être pénalisé par un coût excessif. Quand le codeur et le décodeur présentent des degrés de complexité différents, le système est dit « asymétrique ».

Le flux élémentaire

- ***Flux élémentaire vidéo***

La figure ci dessous montre la structure du flux élémentaire vidéo. L'unité fondamentale de l'information d'image est le bloc DCT qui représente un réseau de 8 x 8 pixels pouvant être Y, Cr ou Cb. La composante continue est d'abord émise avec une précision supérieure à celle des autres coefficients. Viennent ensuite les autres coefficients et le code de fin de bloc (EOB). Les blocs sont groupés en macroblocs qui constituent l'unité fondamentale d'une image et qui peuvent faire l'objet de la compensation de mouvement. Chaque macrobloc comporte, dans son en-tête, un vecteur de mouvement bi-dimensionnel.

Le flux élémentaire



Le flux élémentaire

- ***Flux élémentaires audio***

Plusieurs sortes d'audio peuvent être insérées dans un multiplex MPEG-2. Ces catégories comprennent l'audio codée selon les normes des couches 1, 2 ou 3 du MPEG ou l'ATSC. Le type de codage doit être inclus dans un indicateur que le décodeur lira de façon à faire appel au type de décodage approprié.

Le processus de compression audio est sensiblement différent de celui de la vidéo. Il ne comporte rien d'équivalent aux différents types d'image I, P ou B et les trames audio contiennent toujours la même quantité de données audio. Il n'y a également pas d'équivalence au codage bi-directionnel et les trames audio ne sont pas transmises autrement qu'en séquence naturelle.

Le flux élémentaire

En MPEG-2 audio, l'indicateur de l'en-tête de séquence contient le numéro de la couche utilisée pour la compression audio et le type de compression employé (stéréo par exemple) ainsi que la fréquence d'échantillonnage initiale.

La séquence audio est organisée en unités d'accès (AU = Access Units) qui constitueront les trames audio codées. Si le codage AC-3 est utilisé, cette disposition est répétée dans l'en-tête de séquence. L'unité d'accès audio est une trame synchronisée en AC-3.

La trame AC-3 représente une excursion de temps équivalente à 1536 échantillons audio, ce qui fait 32 ms pour un échantillonnage à 48 KHz et 48 ms en 32 KHz.

Le Packet Elementary Stream

Pour des raisons de commodité, les flux élémentaires continus transportant de l'audio ou de la vidéo en provenance de compresseurs doivent être divisés en paquets. Ces paquets sont identifiés par des en-têtes contenant des marqueurs temporels pour la synchronisation. A partir des PES, on peut créer des Flux de Programme et des Flux de Transport.

Le Packet Elementary Stream

- **Paquets PES**

Dans le flux élémentaire en paquets (PES), un flux élémentaire continu est divisé en paquets d'une taille adaptée à l'application. La taille peut atteindre plusieurs centaines de kilo-octets et varie en fonction de l'application. Chaque paquet est précédé par un en-tête PES. Le paquet commence par un préfixe de 24 bits (préfixe de départ) et un identificateur de flux caractérisant le contenu du paquet (audio ou vidéo), et si nécessaire, le type de codage audio. Ces deux paramètres (préfixe de départ et identificateur de flux) constituent le code de départ identifiant le code de début du paquet.

Il est important de ne pas confondre le paquet d'un PES et le paquet plus petit utilisé dans les flux de transport, ces deux éléments portant malheureusement le même nom.

Comme le MPEG ne définit que le flux de transport, et non le codeur, un concepteur peut décider de fabriquer un multiplexeur convertissant les flux élémentaires en flux de transport en une seule étape. Dans ce cas, les paquets PES ne peuvent exister sous une forme identifiable, mais ils sont logiquement partie intégrante de la charge du flux de transport. C'est d'ailleurs majoritairement le cas.

Le Packet Elementary Stream

- ***Marqueurs temporels***

Après la compression, les images sont extraites de la séquence à cause du codage bi-directionnel. Elles nécessitent une quantité variable de données et sont sujettes à un retard variable en raison du multiplexage et de la transmission. Des marqueurs temporels sont périodiquement incorporés dans chaque image afin de verrouiller la synchronisation entre l'audio et la vidéo. Un marqueur temporel est un nombre de 33 bits constitué par la valeur d'un compteur piloté par une horloge à 90 KHz. Cette horloge est le résultat de la division par 300 de l'horloge de programme à 27 MHz .

Le flux élémentaire

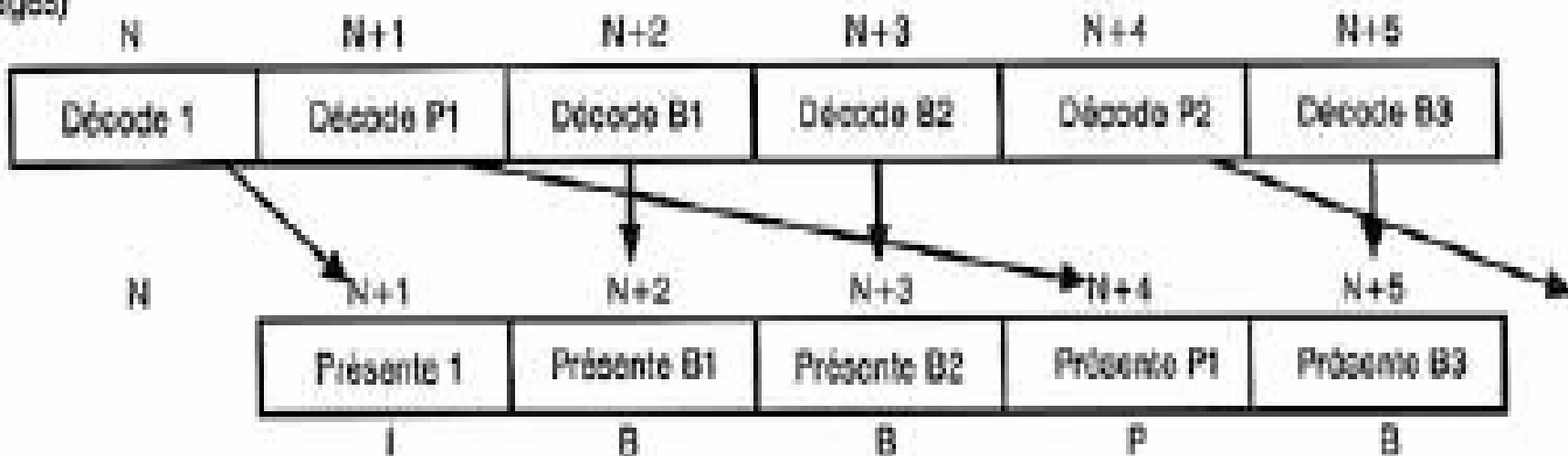
- ***PTS/DTS***

Quand on utilise un code bidirectionnel, une image doit avoir été décodée quelque temps avant d'être présentée de façon à pouvoir être utilisée comme source de données pour une image B. Bien que, par exemple, l'on puisse présenter les images dans l'ordre IBBP, elles seront transmises dans l'ordre IPBB. Il existera donc deux types de marqueurs temporels.

Le Packet Elementary Stream

	I	P1	B1	B2	P2	B3
PTS	N+1	N+4	N+2	N+3	N+7	N+5
DTS	N	N+1	—	—	N+4	—

Temps (en images)



Le Packet Elementary Stream

- ***PTS/DTS***

Le marqueur temporel de décodage (DTS = Decode Time Stamp) indique quand une image doit être décodée, tandis que le marqueur temporel de présentation (PTS = Presentation Time Stamp) indique quand l'image doit être présentée à la sortie du décodeur.

Les images B sont décodées et présentées simultanément, c'est pourquoi ils ne comportent qu'un PTS. Quand une séquence IPBB est reçue, les images I et P doivent être décodées avant la première image B. Pendant que l'image P est en cours de décodage, l'image I est présentée en sortie de façon à pouvoir être suivie des images B.

Le Packet Elementary Stream

- ***PTS/DTS***

La figure ci-dessus montre que, quand une unité d'accès contenant une image I est reçue, les deux marqueurs DTS et PTS doivent exister dans l'en-tête et qu'ils doivent être séparés d'au moins une période d'image.

Si le codage bi-directionnel est utilisé, une image P doit venir ensuite et cette image est également dotée d'un PTS et d'un DTS, mais la séparation entre les deux marqueurs doit être de trois périodes d'image pour permettre l'intervention des images B. Ainsi, si une séquence IPBB est reçue, l'image I est retardée d'une image, l'image P de trois images et les deux images B n'ont aucun retard. La présentation devient alors IBBP.

En clair, si la structure de GOP est changée de façon à ce qu'il y ait plus d'images B insérées entre les images I et P, la différence entre le DTS et le PTS des images P sera augmentée.

Le Flux de Programme et le Flux de Transport

Les flux de programme et transport constituent deux façons d'assembler plusieurs flux de paquets PES.

- ***Enregistrement contre transmission***

Pour une qualité d'image donnée, le débit de données variera selon le contenu de l'image. Un canal à débit variable donnera donc le meilleur résultat. Dans le domaine des transmissions, la plupart des canaux sont fixes et le débit global de bits est maintenu constant par l'addition d'un bourrage (données sans signification).

Dans le cas du DVD, l'utilisation de ce bourrage constitue un gaspillage d'espace de stockage. Cependant, on peut accélérer ou ralentir la vitesse d'un moyen de stockage soit physiquement, soit dans le cas d'une unité de disque, en changeant les demandes de transfert de données. Cette approche permet d'avoir un canal à débit variable sans perte de capacité. Quand un support est relu, on peut ajuster la vitesse de façon à conserver le tampon à moitié plein quel que soit le débit dynamique qui peut être dynamiquement ajusté.

Le Flux de Programme et le Flux de Transport

- ***Introduction aux flux de programme***

Un flux de programme est un multiplex de paquets PES qui véhicule plusieurs flux élémentaires préalablement codés à partir de la même horloge mère ou de la même référence. Le flux peut être constitué d'un flux vidéo et de ses flux audio associés ou seulement un canal de multi programmes audio.

Le flux élémentaire vidéo est divisé en unités d'accès contenant chacune des données compressées décrivant une image. Ces images sont du type I, P ou B et chacune comporte un numéro d'unité d'accès qui indique la séquence d'affichage convenable. Une unité d'accès vidéo devient un paquet de flux de programme.

En vidéo, la taille de ces paquets varie. Un paquet d'image I sera, par exemple, plus grand que celui d'une image B. Les unités d'accès de l'audio numérique sont généralement de la même taille et plusieurs d'entre elles constituent un paquet de flux de programme.

Le Flux de Programme et le Flux de Transport

- ***Flux de transport***

Un flux de transport est bien plus qu'un simple multiplexage de plusieurs paquets PES. Dans les flux de programme, les marqueurs temporels sont suffisants pour recréer l'axe des temps parce que l'audio et la vidéo sont verrouillées sur une référence commune.

Pour télécharger des données dans un réseau éloigné, un impératif supplémentaire est nécessaire dans le décodeur pour recréer l'horloge de chaque programme. Cela impose un élément supplémentaire dans la syntaxe afin de fournir la référence d'horloge de programme (PCR = Program Clock Reference).

Le flux de transport véhicule plusieurs programmes différents utilisant chacun un débit et un facteur de compression susceptibles de varier dynamiquement, même si le débit global reste constant. Cette propriété est appelée multiplexage statistique et permet à un programme supportant des images difficiles de prendre de la bande à un programme supportant des images plus statiques.

Téletexte en DVB

- *Rappel*

En analogique sur 625 lignes, 575 transportent la vidéo "utile", cad l'image que l'on voit. 50 lignes sont utilisées pour la suppression trame (synchronisation verticale de la vidéo). C'est dans ces lignes "non visibles" que l'on insère le téletexte.

En vidéo numérique dite 422, on ne numérise que les 575 lignes de vidéo utiles. Donc à la trappe les lignes de téletexte. Cependant on arrive à les transporter sous forme de données auxiliaires comme l'audio. Voir norme 422 (itu 601)

Téletexte en DVB

- **DVB**

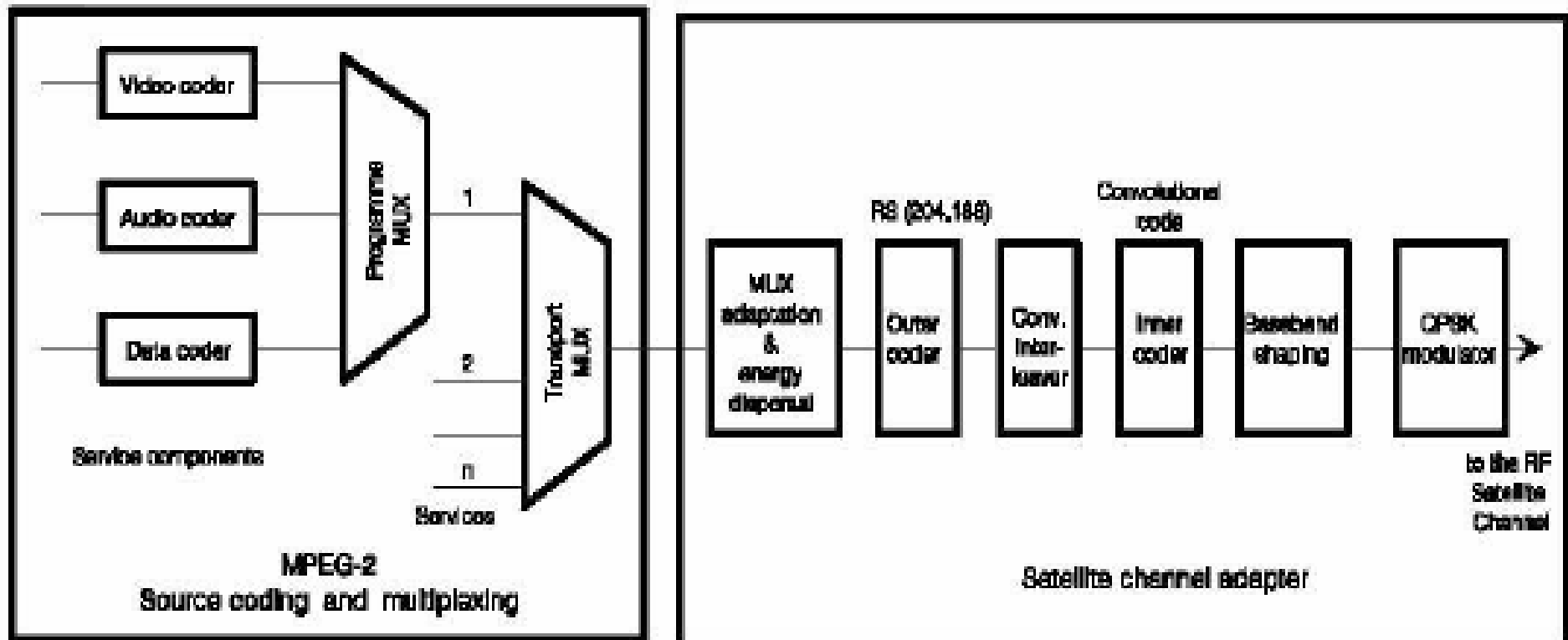
Le transport du téletexte est prévu. Au codage, on extrait les données du téletexte de la vidéo et on les encode dans un PID propre. Ainsi un service TV est composé de 3 composantes ayant chacun un PID (1 vidéo, 1 audio, 1 téletexte) Lorsque ton terminal décode le service, il réinsère les données de téletexte dans la vidéo. Le décodeur téletexte de ta télé peut alors fonctionner. Le débit d'une composante téletexte, dépend de la richesse du contenu éditorial. Comptez 120 kbit/s pour un téletexte à la française et plus de 300 kbit/s pour un téletexte à l'italienne.

Services de télé et de données

- ***Format de codage des services de télévision : MPEG-2.***
 - Débits vidéo :
 - SD (Standard Definition) (720 x 576 lignes) : de 3 à 6 Mbps.
 - HD (High Definition) (1280 x 720 lignes) : de 18 à 20 Mbps.
 - Débits audio : quelques 100 kbps.
 - Trames de 188 octets (dont un octet de synchronisation).
 - Services multiplexés en TDM.

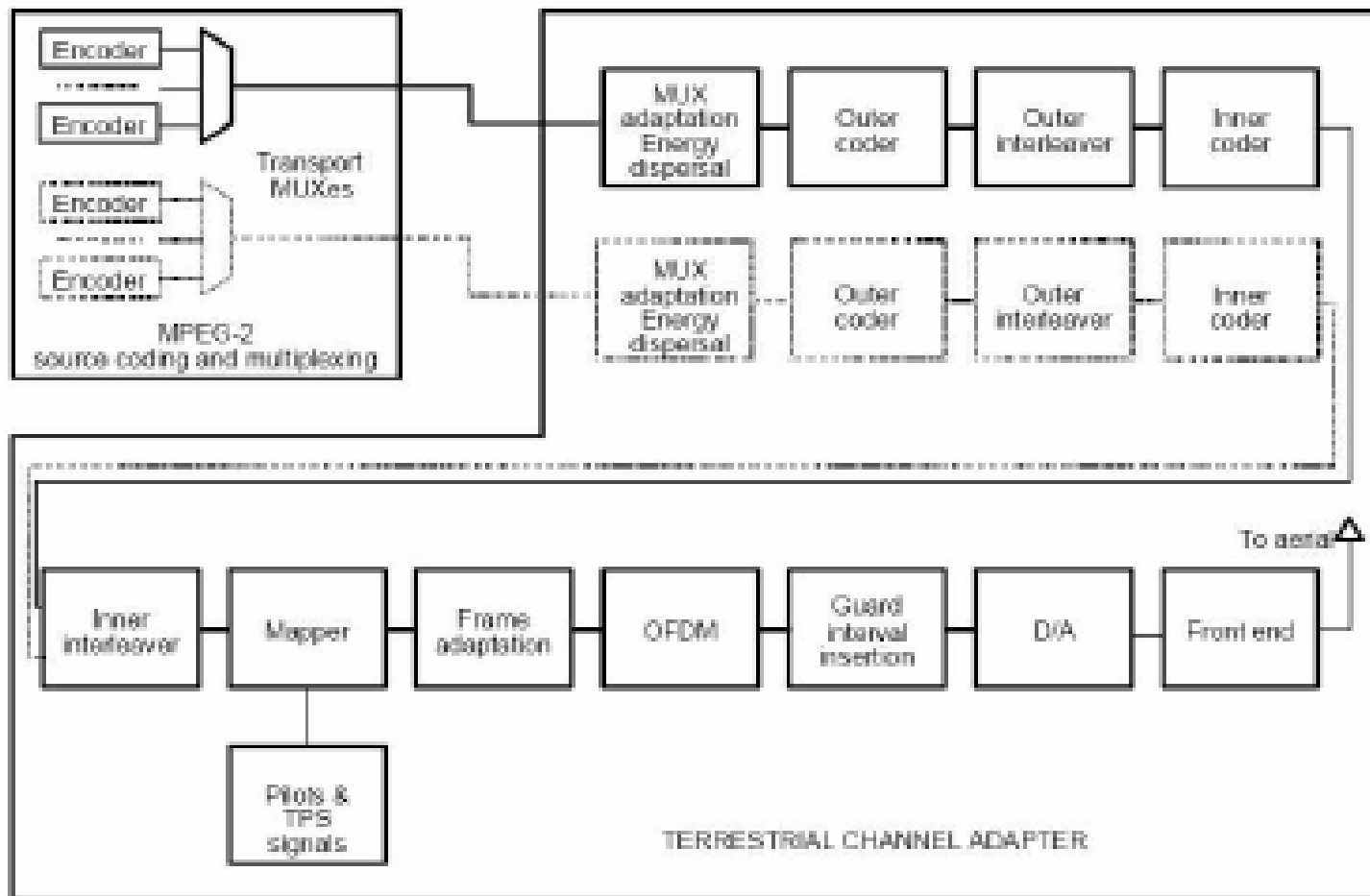
Services de télé et de données

Adaptation de signaux télé bande de base à la sortie d'un multiplex MPEG-2 aux caractéristiques d'un canal satellite.



Services de télé et de données

- Adaptation d'un flux MPEG-2 à un canal de transmission terrestre.



Référence Bibliographique

- DVB Digital Video Broadcasting: *Benoît ESCRIG*
- Diffusion vidéo numérique : le DVB-T :
Philippe Ciblat
- Les bases du MPEG V1.0 sur
<http://perso.libertysurf.fr/IPhilGood>