La compression vidéo

M3201 : Culture scientifique et traitement de l'information

E. CHOUZENOUX

emilie.chouzenoux@univ-mlv.fr

Plan

1) Rappels sur la vidéo numérique
2) Enjeux et principes de la compression
3) La compression des images
4) La compression des vidéos
5) La compression audio
6) La norme MPEG4

I - Rappels sur la vidéo numérique

Qu'est-ce qu'une vidéo numérique ?

- Vidéo = Son + Images

- Caractéristiques importantes :
  - Nombre d'images par secondes
  - Nombre de pixels par image
  - Codage de la couleur
  - Fréquence d'échantillonnage
  - Niveaux de quantification

➡ Ces caractéristiques influent sur le poids de la vidéo.

Calcul du poids d'une vidéo

- Le poids d'une vidéo numérique non compressée est donné par :

Poids des images = 3 couleurs × Nombre de pixels par image × Nombre d'images par seconde × Nombre de bits × Durée

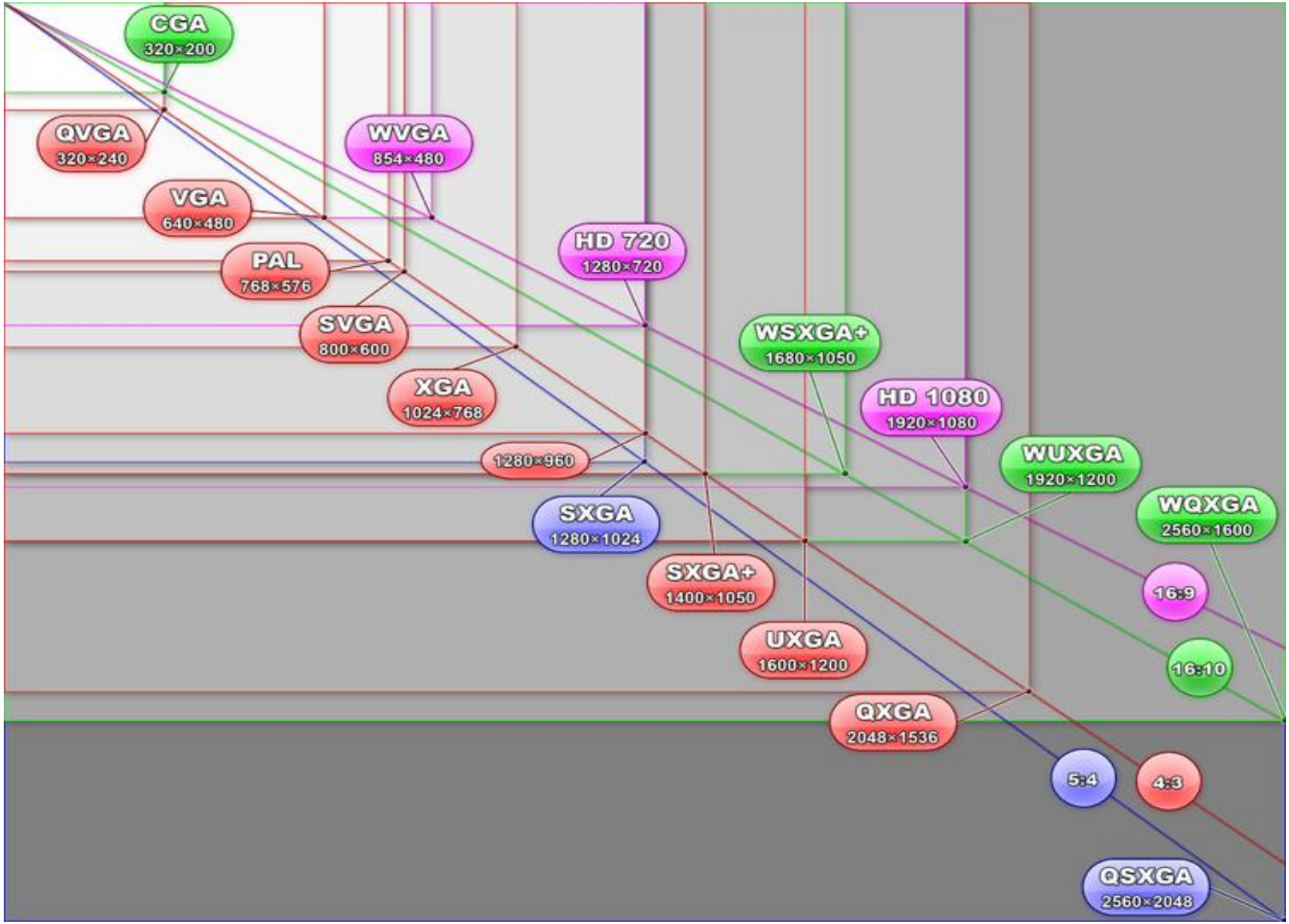
Poids du son = Nombre de pistes × Fréquence d'échantillonnage × Nombre de bits × Durée

Poids total = Poids des images + Poids du son

Quelques données numériques

- Nombre d'images par secondes :
  - Cinéma : 24 im/s
  - Télévision : 25 im/s ou 30 im/s (dépend du format)
- Résolutions d'écrans

➡ Voir schéma



Source :Wikipédia

Quelques données numériques

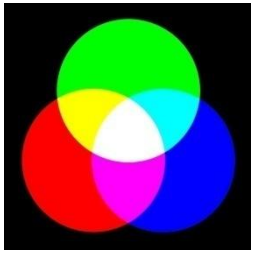
- Nombre d'images par secondes :
  - Cinéma : 24 im/s
  - Télévision : 25 im/s ou 30 im/s (dépend du format)
- Résolutions d'écrans

  ➡ Voir schéma
- Quantification
  - Image : 8 / 10 bits
  - Son : 16 / 24 bits
- Fréquence d'échantillonnage
  - Son : 44.1 KHz

La transmission de la couleur

Solution simple: Envoyer les composantes de lumière rouge, verte et bleue séparément. Puis additionner les signaux reçus.

Propriétés : Rouge + vert + bleu en quantités égales = blanc
Absence de signal = noir



Problème : Trop coûteux en matériel (3 fils !), et en quantité d'information (redondance spatiale).

Solution : Représentation luminance/chrominance, et transmission sur un seul fil, grâce à la modulation (voir cours 1ère année).

Représentation luminance/chrominance

Les formats vidéos utilisent des bases de couleurs différentes, toutes basées sur le principe de la décomposition YCrCb:

Y est la **luminance** = le niveau de gris de l'image.

Les **chrominances rouge (Cr) et bleu (Cb)**, portent l'information de la couleur.

La base YDrDb

$Y = 0{,}30R + 0{,}59V + 0{,}11B$

$D_R = -1{,}902\ (R - Y)$ « chrominance rouge »

$D_B = 1{,}505\ (B - Y)$ « chrominance bleue »

Utilisée dans le format SECAM

Représentation luminance/chrominance

Les formats vidéos utilisent des bases de couleurs différentes, toutes basées sur le principe de la décomposition YCrCb:

Y est la **luminance** = le niveau de gris de l'image.

Les **chrominances rouge (Cr) et bleu (Cb)**, portent l'information de la couleur.

La base YIQ

$$Y = 0{,}30R + 0{,}59V + 0{,}11B$$

$$I = 0{,}27(B - Y) + 0{,}74(R - Y)$$

$$Q = 0{,}41(B - Y) + 0{,}48(R - Y)$$

Utilisée dans le format NTSC

Représentation luminance/chrominance

Les formats vidéos utilisent des bases de couleurs différentes, toutes basées sur le principe de la décomposition YCrCb:

Y est la **luminance** = le niveau de gris de l'image.

Les **chrominances rouge (Cr) et bleu (Cb)**, portent l'information de la couleur.

La base YUV

$$Y = 0{,}30R + 0{,}59V + 0{,}11B$$
$$U = 0{,}493\ (B - Y)$$
$$V = 0{,}877\ (R - Y)$$

Utilisée dans le format PAL

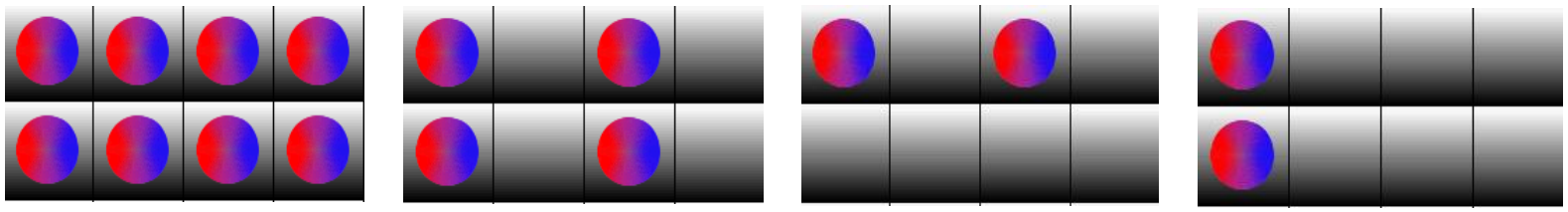
Sous-échantillonnage de la chrominance

L'œil est moins sensible aux variations de chrominance.

➡ On envoie moins d'échantillons de chrominance que de luminance.

- ✓ On raisonne par groupes de 4 pixels consécutifs.
- ✓ Traitement différent des lignes paires et impaires.

Mesure de Y, Cr et Cb ↙ Mesure de Y uniquement ↙



II - Enjeux et principes de la compression

Enjeux de la compression

Compresser = une nécessité

216Mbit/s

Débit utile vidéo

4:2:2 SD

2 à 8 Mbit/s

chaîne de TV numérique terrestre SD 576i

9,8 Mbit/s max

DVD vidéo

quelques Mbit/s

ADSL

25 Mbit/s

Caméscope DV

Stratégies de compression

En studio : pas nécessaire de compresser beaucoup pour les liaisons entre équipements
Enregistrement, stockage, transmission : nécessaire de compresser
En particulier : **diffusion télévisuelle grand public** où la compression est parfois poussée jusqu'à la limite d'apparition de défauts (TV avec l'ADSL…)

Deux stratégies de réduction de débit :

- **Sans pertes** = L'information est préservée. L'opération est réversible.

Ex : Algorithme de codage entropique de Huffman (ex: JPEG)

- **Avec pertes** = On ne peut pas reconstituer intégralement l'information originale
  - Suffisant la plupart du temps
  - Taux de compression plus importants
  - Résultats variables en fonction de la nature des images

- **Taux de compression** :

$$\eta = \frac{\text{débit sans compression}}{\text{débit avec compression}} = \frac{\text{volume sans compression}}{\text{volume avec compression}}$$

Stratégies de compression

Beaucoup de redondance dans les phénomènes naturels !

Dans une vidéo, où se trouve la redondance ?

Vidéo = son + image

- Dans le son :
  - **Redondance temporelle**
  - **Redondance statistique**
  - **Propriétés physiologiques**
- Dans l'image :
  - **Redondance spatiale**
  - **Redondance temporelle**
  - **Redondance statistique**
  - **Propriétés physiologiques**

Principes généraux

Redondance spatiale : Dans une image, les pixels voisins ont des intensités proches.

➡ Transformée en Cosinus Discrète.

Redondance temporelle : Entre deux images ou deux échantillons de signaux sonores consécutifs, il y a peu de changements.

➡ Codage différentiel / Compensation de mouvement.

Propriétés physiologiques: On tient compte des propriétés de l'acuité auditive et visuelle.

Ex: les détails fins d'une image n'ont pas besoin d'être codés avec autant de précision que les parties essentielles.

Ex: le masquage fréquentiel pour les signaux sonores.

➡ Quantification non uniforme.

Redondance statistique : Certains symboles reviennent plus fréquemment que d'autres.

➡ Codage entropique.

III – La compression des images

La compression JPEG

- JPEG (*Joint Photographic Expert Group*) :
norme ISO de compression d'images fixes avec pertes, standardisée en 1992.

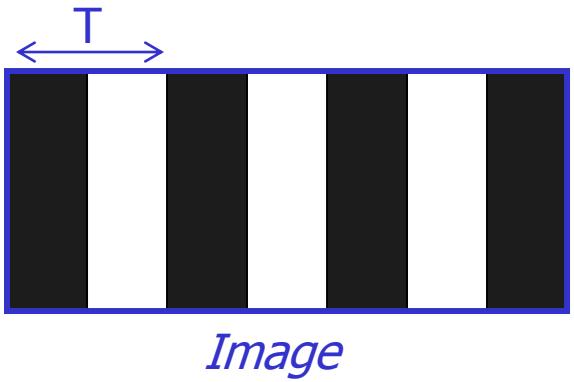
- JPEG prend en compte les failles du système de perception humaine, de manière à réduire la quantité d'information de l'image qui sera codée.

- L'algorithme de compression repose principalement sur :
  - ✓ Une transformation mathématique de l'image
    la DCT (*Discrete Cosine Transform*)
  - ✓ Un codage à longueur variable
    le codage de Huffman

- Formats source
  R, V, B en 4:4:4
  Y, Cr et Cb en 4:2:2

La notion de fréquences dans une image

Une image peut présenter **une fréquence spatiale horizontale** et une **fréquence spatiale verticale**.



Si l'œil parcourt de gauche à droite cette image,

⟶ sensation de variation cyclique de la luminance.

Il existe une période spatiale horizontale.

La fréquence spatiale représente le nombre de cycles par unité de longueur.

La période spatiale verticale, par contre, est nulle.

Pour les signaux temporels, on utilise la **transformée de Fourier** pour passer dans le domaine des fréquences.

Pour les images, on peut, entre autres, passer dans le domaine des fréquences par la **Transformée en Cosinus Discrète** (TCD, ou DCT – Discrete Cosinus Transform)

La Transformée en Cosinus Discrète

La DCT porte sur des blocs de 8×8 pixels.

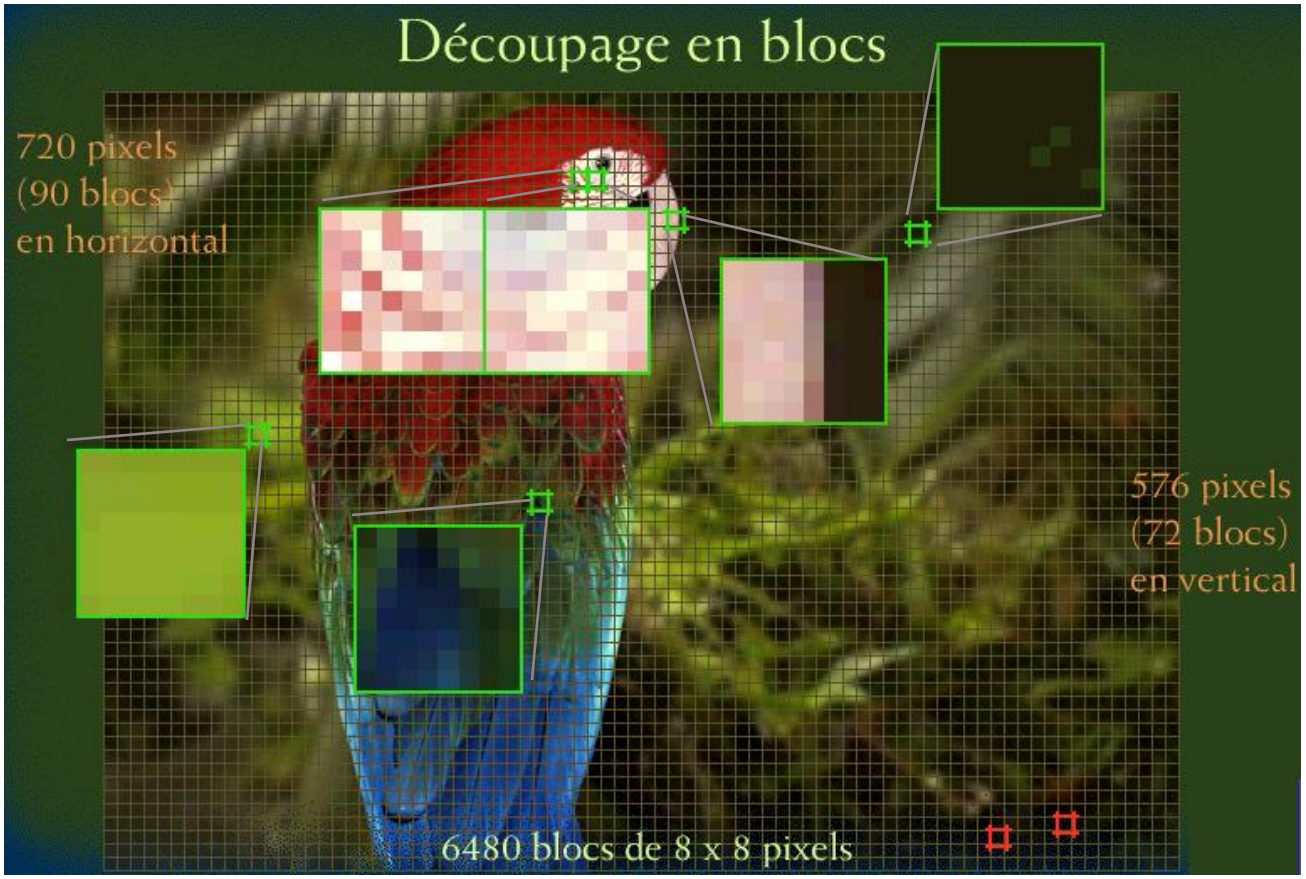
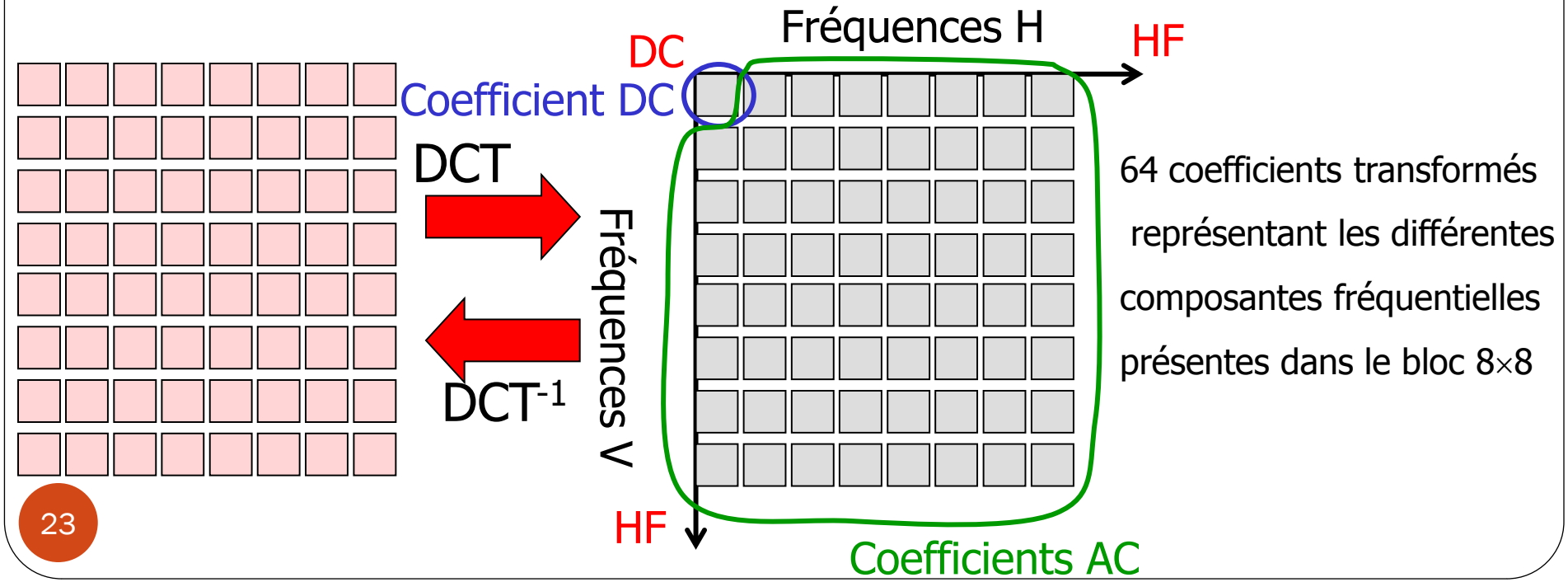


Image INA Formation

La Transformée en Cosinus Discrète

La DCT fournit une **représentation spectrale bidimensionnelle** (H et V) du bloc.

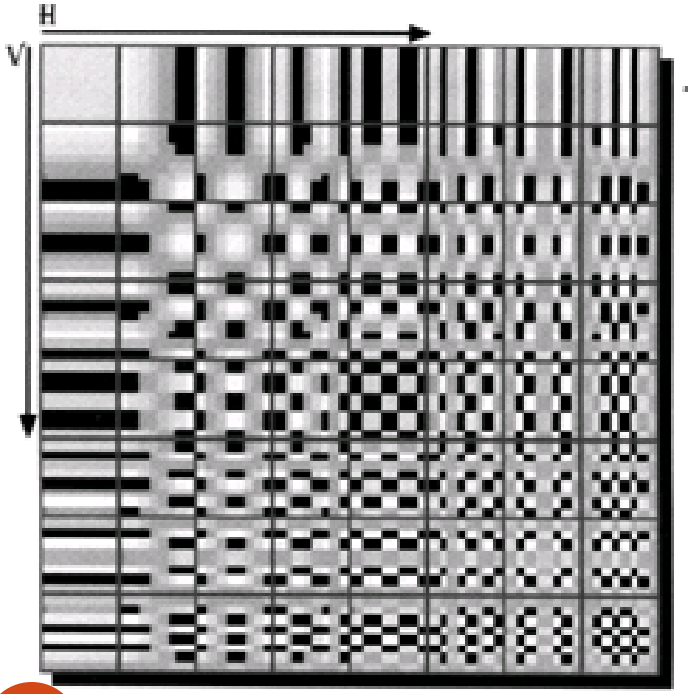
Le bloc est décomposé en une combinaison linéaire de « fonctions » images de base, de la même manière que la transformée de Fourier décompose un signal temporel en une somme de fonctions sinus et cosinus.



La Transformée en Cosinus Discrète

Les « fonctions » de base de la DCT sont représentées dans le tableau T ci-dessous.

Chaque coefficient de la DCT représente la **contribution** de l'image située à la même place dans la matrice dans la construction de l'image.

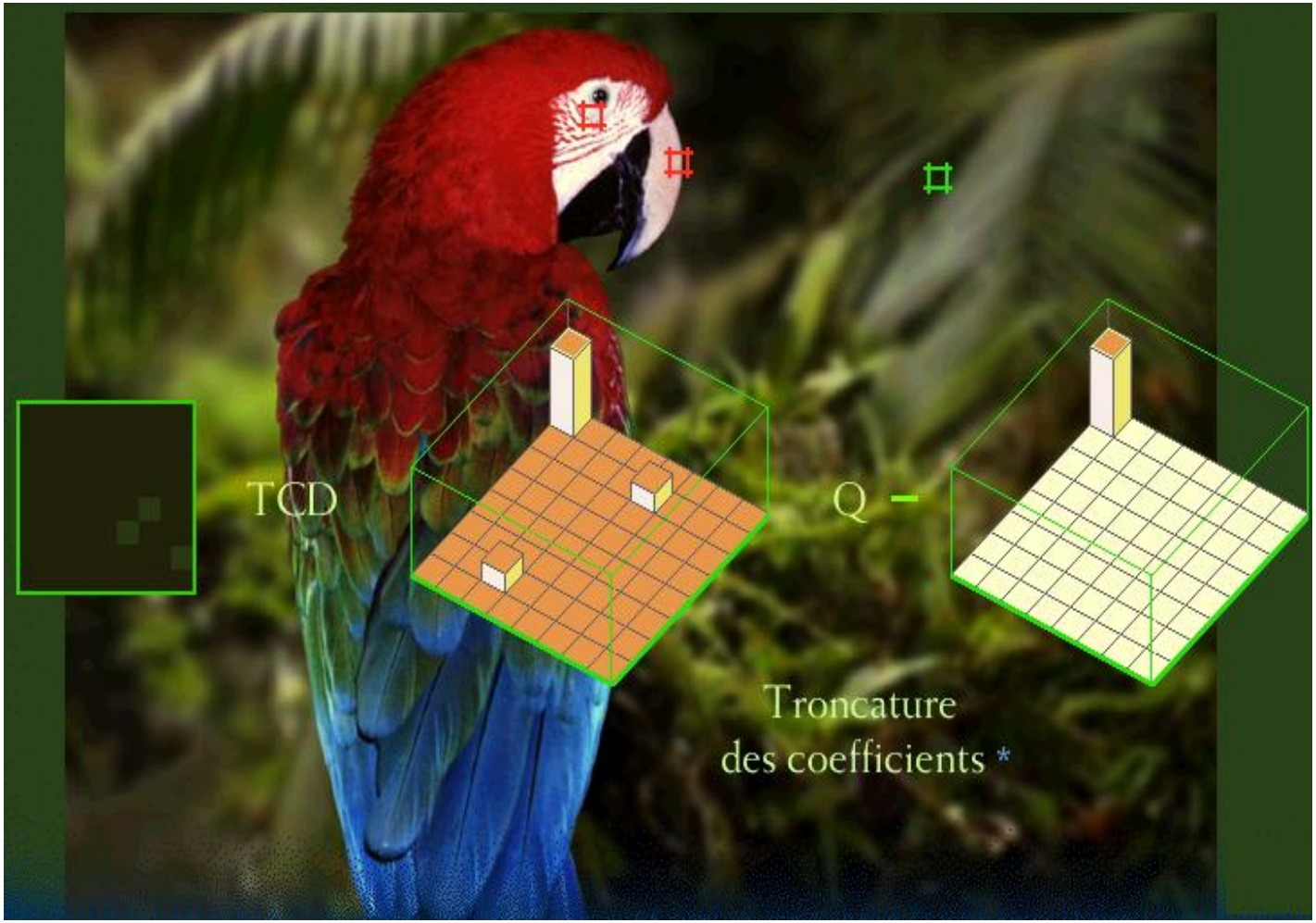


T

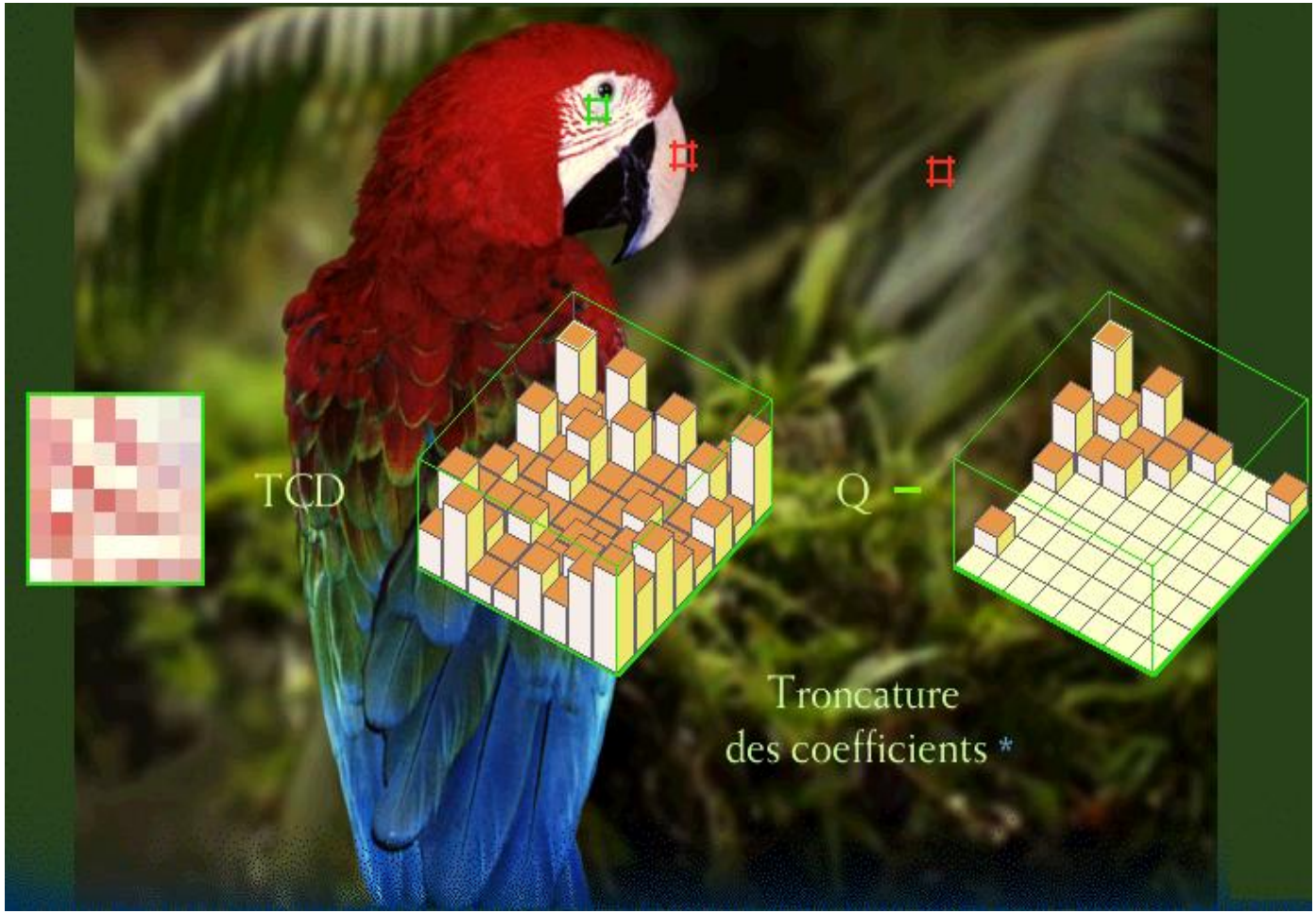
Exemple : La matrice des coefficients DCT d'un bloc de pixel est le suivant :

2	3	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Exemple 1 : Bloc homogène



Exemple 2 : Bloc hétérogène



Propriétés de la DCT

- On constate que les coefficients **hautes fréquences** portent les informations **de détails** de l'image.

  Les blocs des zones homogènes d'une image ont des coefficients non négligeables en BF, et négligeables en HF (exemple 1).

  Les blocs des zones hétérogènes ont des coefficients importants en HF (exemple 2).

- Puisque les hautes fréquences portent l'information de détail de l'image, on va diminuer leur valeur, voire l'annuler. C'est l'opération de **quantification**. L'image va être **dégradée**, mais compte-tenu de l'imperfection de l'œil humain, cette dégradation sera **peu** ou **pas perceptible** (tout dépendra de la sévérité de la quantification).

La quantification

La norme JPEG fournit des **tables**, qui permettent de retoucher la valeur des coefficients DCT de manière à **diminuer la valeur de ceux qui contribuent le moins au sens de l'image**. Les coefficients de **haute fréquence** sont les **plus réduits**.

- Le rapport de compression dépend directement de la table utilisée.
- Des tables de quantification sont définies pour **chaque composante** (Y, Cr et Cb).
- La quantification est **moins sévère pour la luminance** car l'œil y est plus sensible.

240	4	-10	7	11	19	24	29
-25	-13	6	8	-12	28	29	26
-16	-13	7	-11	19	27	33	27
6	7	10	13	24	-42	39	30
8	10	17	-27	43	-53	50	-37
11	16	-26	31	39	51	55	45
23	31	38	42	50	59	-59	49
35	45	46	48	55	49	50	-48

Y

/

8	6	5	8	12	20	25	30
6	6	7	10	18	28	30	27
7	7	8	12	20	28	35	28
7	8	11	15	25	44	40	31
9	11	19	28	44	55	52	38
12	25	28	32	41	55	56	46
25	32	39	44	52	60	60	51
36	46	48	48	56	50	54	50

Q_Y

$$Y' = \mathrm{Fix}(Y/Q_Y)$$

La quantification

La norme JPEG fournit des **tables**, qui permettent de retoucher la valeur des coefficients DCT de manière à **diminuer la valeur de ceux qui contribuent le moins au sens de l'image**. Les coefficients de **haute fréquence** sont les **plus réduits**.

- Le rapport de compression dépend directement de la table utilisée.
- Des tables de quantification sont définies pour **chaque composante** (Y, Cr et Cb).
- La quantification est **moins sévère pour la luminance** car l'œil y est plus sensible.

240	4	-10	7	11	19	24	29
-25	-13	6	8	-12	28	29	26
-16	-13	7	-11	19	27	33	27
6	7	10	13	24	-42	39	30
8	10	17	-27	43	-53	50	-37
11	16	-26	31	39	51	55	45
23	31	38	42	50	59	-59	49
35	45	46	48	55	49	50	-48

DCT(Y)

/

24	18	15	24	36	60	75	90
18	18	21	30	54	84	90	81
21	21	24	36	60	84	105	84
21	24	33	45	75	132	120	93
27	33	57	84	132	165	156	114
36	75	84	96	123	165	168	138
75	96	117	132	156	180	180	153
108	138	144	144	168	150	162	150

Q_Y

$$Y' = \mathrm{Fix}(Y/Q_Y)$$

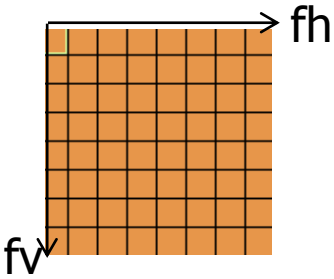
L'arrondi

a	Fix(a)
-1.2	
-0.3	
0.5	
4.3	
1.1	
-1.5	
6.7	
4	
-2.7	

Cette définition de l'arrondi permet d'avoir un maximum de coefficients proches de 0.

Illustration de l'effet de la quantification

Conservation des 64 coefficients ⟶ pas de perte



Conservation de 32 coefficients

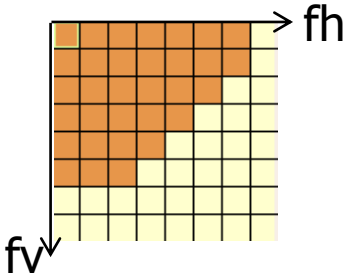
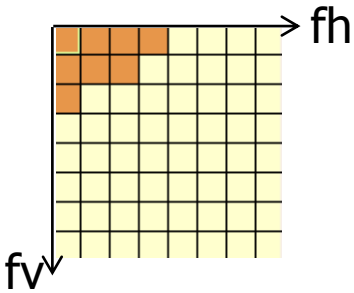


Illustration de l'effet de la quantification

Conservation de 8 coefficients



Conservation de 4 coefficients

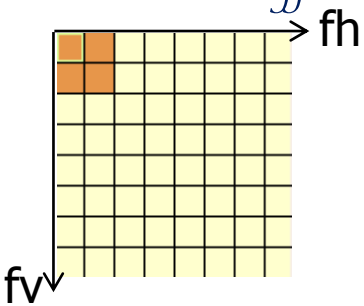
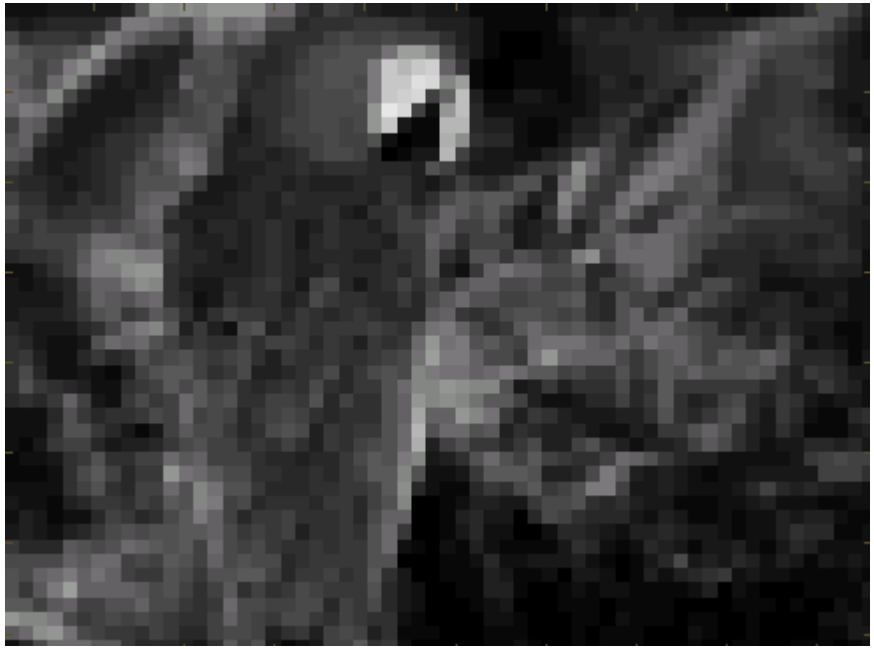
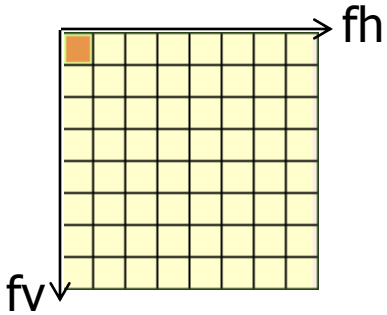


Illustration de l'effet de la quantification



Conservation d'un coefficient unique

⟶ la composante continue (DC)



Un codeur peut travailler

- À taux de compression constant (au détriment de la qualité).
- A qualité constante : la performance du codeur en matière de taux de compression risque d'être médiocre.

La nature de l'image intervient sur l'efficacité de la compression.

- Une image simple peut être de bonne qualité avec un fort taux de réduction. Pour ce même taux, une image complexe sera certainement de piètre qualité…

Le balayage

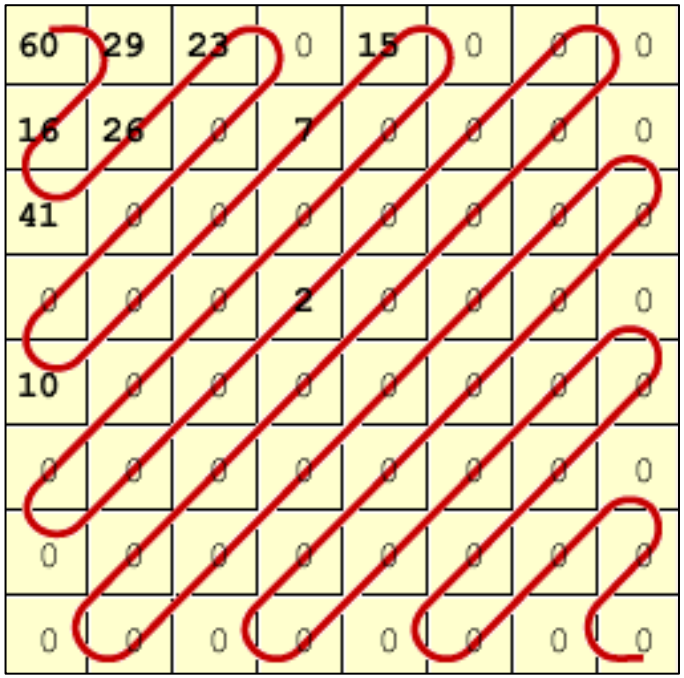
1) Les coefficients de la DCT sont **balayés en zig-zag**, des BF vers les HF.

2) Les suites de 0 sont transmises sous forme condensée (*Run Length Encoding*, RLE) :
couple (nb de « 0 » précédant le coeff non nul, coeff).

Exemple :

1)

2)



Le codage

Le résultat du RLE est ensuite codé, à l'aide d'un codeur **à longueur variable (=codeur entropique)**

Le principe des algorithmes de codage statistique est d'utiliser les probabilités d'occurrence de chaque symbole dans une séquence de symboles émanant de la source (même principe que le MORSE).

Codeur entropique de Shannon-Fano

L'algorithme comporte les étapes suivantes :

1) Les probabilités d'apparition de chaque symbole sont placées dans un tableau trié par ordre décroissant de probabilités.

2) Le tableau est coupé en deux groupes de symboles S_0 et S_1 dont la somme des probabilités de chaque groupe avoisine 0.5.

3) Le groupe S_0 est codé par un "0" et S_1 par un "1".

4) Si un groupe S_i n'a qu'un seul élément, c'est une feuille terminale, sinon la procédure reprend récursivement à l'étape 2 sur le groupe S_i.

Exemple

Pour illustrer cet algorithme, nous allons coder la phrase suivante :

"Le codage est indispensable"

Pour simplifier, nous n'allons pas prendre en compte le symbole espace (blanc). Tous ces symboles émanent de l'alphabet suivant :

$$\text{Alphabet} = \{L, E, C, O, D, A, G, S, T, I, N, P, B\}$$

Cet Alphabet a N=13 symboles.

La phrase se compose de 24 'mots' issus de l'alphabet.

1) Calcul des probabilités d'apparition des symboles

Symbole	Nombre de fois	Probabilité
E	5	0,208333333
S	3	0,125
A	2	0,083333333
D	2	0,083333333
I	2	0,083333333
N	2	0,083333333
L	2	0,083333333
B	1	0,041666667
G	1	0,041666667
P	1	0,041666667
T	1	0,041666667
O	1	0,041666667
C	1	0,041666667

2) Division en deux blocs

E	5
S	3
A	2
D	2
I	2
N	2
L	2
B	1
G	1
P	1
T	1
O	1
C	1

S

S_0

E	5
S	3
A	2
D	2

I	2
N	2
L	2
B	1
G	1
P	1
T	1
O	1
C	1

S_1

3) Affectation des codes

E	5
S	3
A	2
D	2
I	2
N	2
L	2
B	1
G	1
P	1
T	1
O	1
C	1

S

0

1

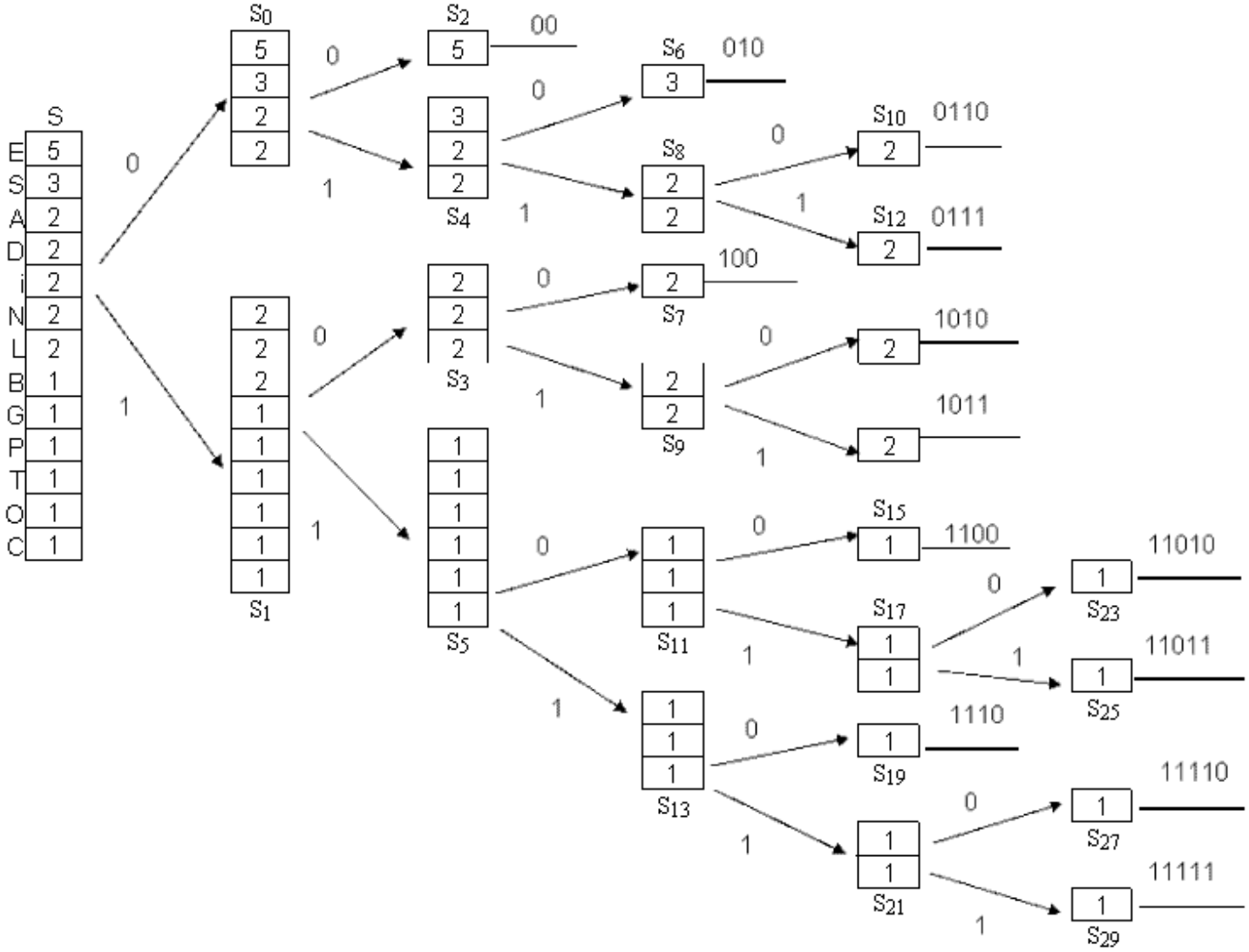
S_0

E	5
S	3
A	2
D	2

I	2
N	2
L	2
B	1
G	1
P	1
T	1
O	1
C	1

S_1

4) Ré-itération de la procédure



Codage résultant

Symbole	Probabilité	Code				
E	5/24				0	0
S	3/24			0	1	0
A	2/24		0	1	1	0
D	2/24		0	1	1	1
I	2/24			1	0	0
N	2/24		1	0	1	0
L	2/24		1	0	1	1
B	1/24		1	1	0	0
G	1/24	1	1	0	1	0
P	1/24	1	1	0	1	1
T	1/24		1	1	1	0
O	1/24	1	1	1	1	0
C	1/24	1	1	1	1	1

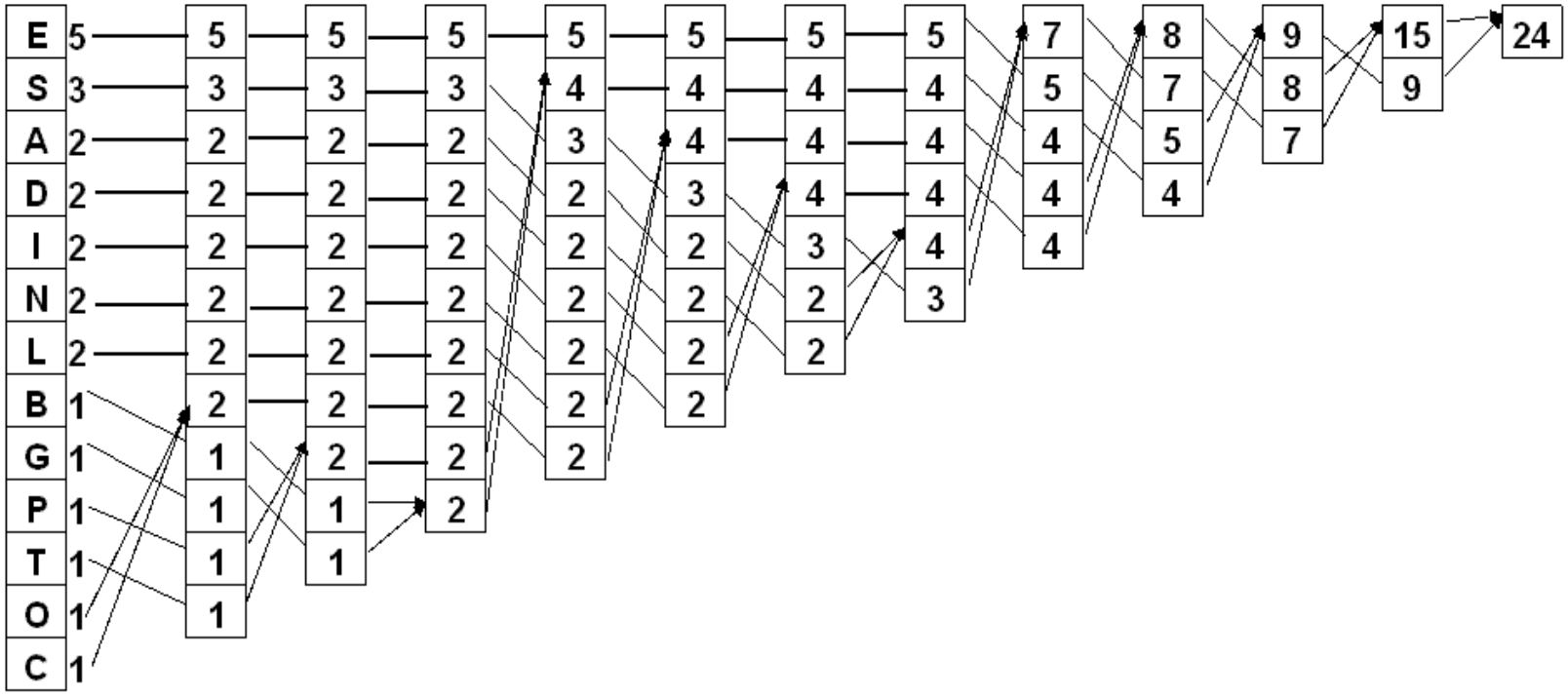
Le codeur entropique de Huffman

- Comme pour le codage de Shannon-Fano, les probabilités d'apparition des symboles sont placées dans un tableau trié par ordre décroissant de probabilités.
- L'algorithme de Huffman est implémenté suivant une structure d'arbre.
- Le principe de cet algorithme consiste à regrouper les deux symboles de probabilités la plus faible pour en faire un nouveau symbole dont la probabilité est la somme des probabilités de ces deux symboles.
- On itère cette opération et à chaque étape le nombre de symboles diminue. On construit de cette manière un arbre dont les feuilles sont les symboles à coder et les embranchements les codages intermédiaires.
- C'est ce codeur qui est souvent utilisé pour le format JPEG.

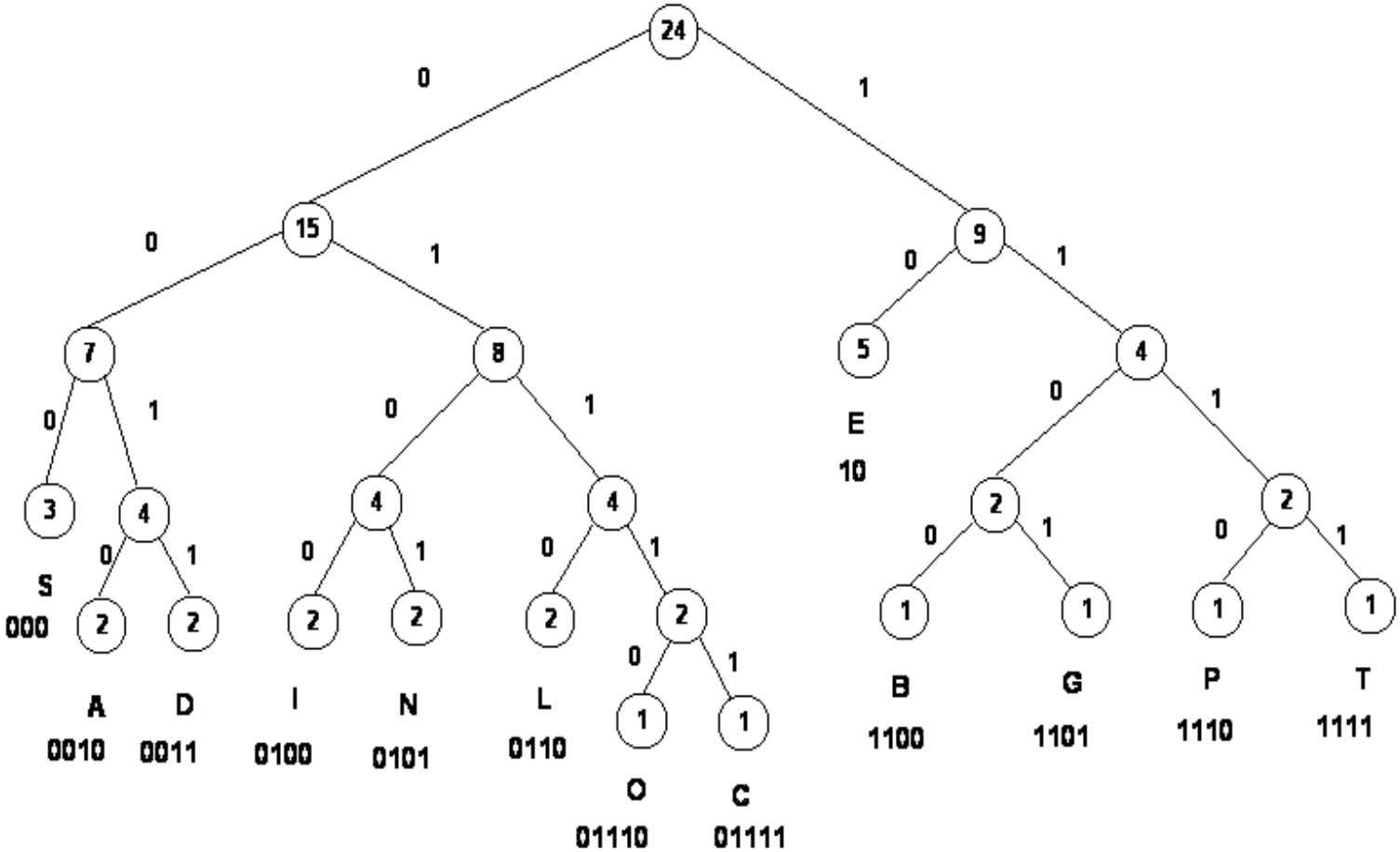
Exemple

Pour illustrer cet algorithme, nous allons coder :

"Le codage est indispensable"



Arbre de Huffman



Code résultant

Symbole	Probabilité	Code				
E	5/24				1	0
S	3/24			0	0	0
A	2/24		0	0	1	0
D	2/24		0	0	1	1
I	2/24		0	1	0	0
N	2/24		0	1	0	1
L	2/24		0	1	1	0
B	1/24		1	1	0	0
G	1/24		1	1	0	1
P	1/24		1	1	1	0
T	1/24		1	1	1	1
O	1/24	0	1	1	1	0
C	1/24	0	1	1	1	1

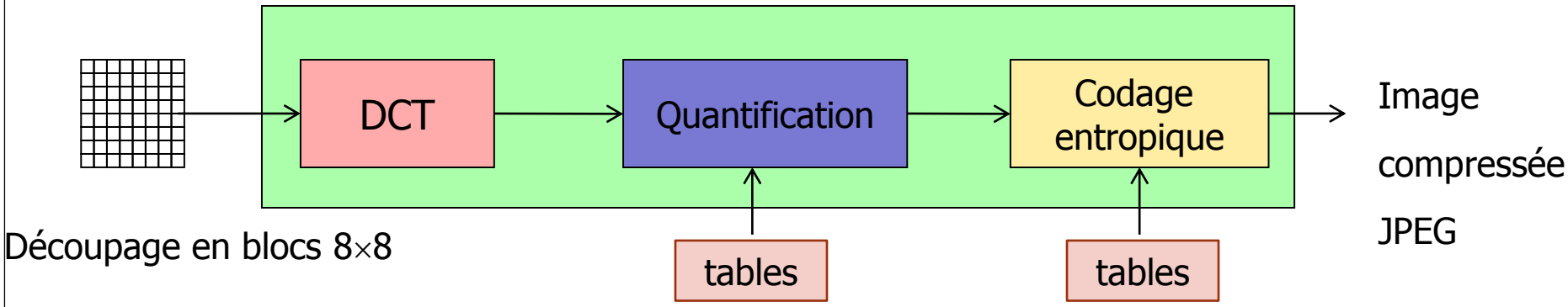
Pour résumer : La compression JPEG

Pour chacune des 3 matrices de composantes Y, Cr et Cb :

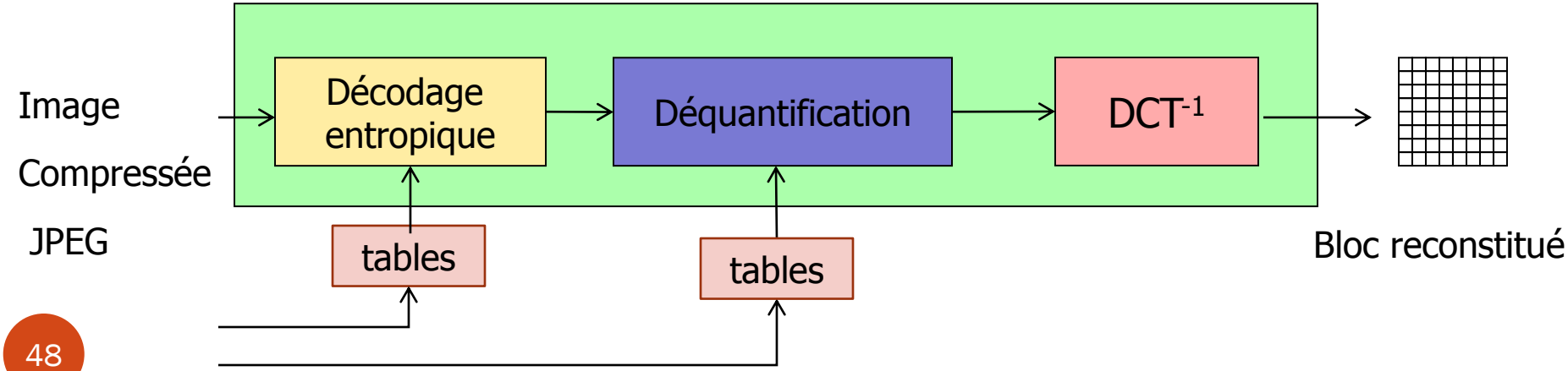
1. Décomposition en blocs de 8×8 pixels par composante
2. Passage du domaine spatial au domaine fréquentiel par DCT
3. Quantification et seuillage des coefficients DCT
4. Balayage en zig-zag + RLE
5. Codage entropique

Synoptiques fonctionnels d'un codeur et d'un décodeur JPEG

Codeur



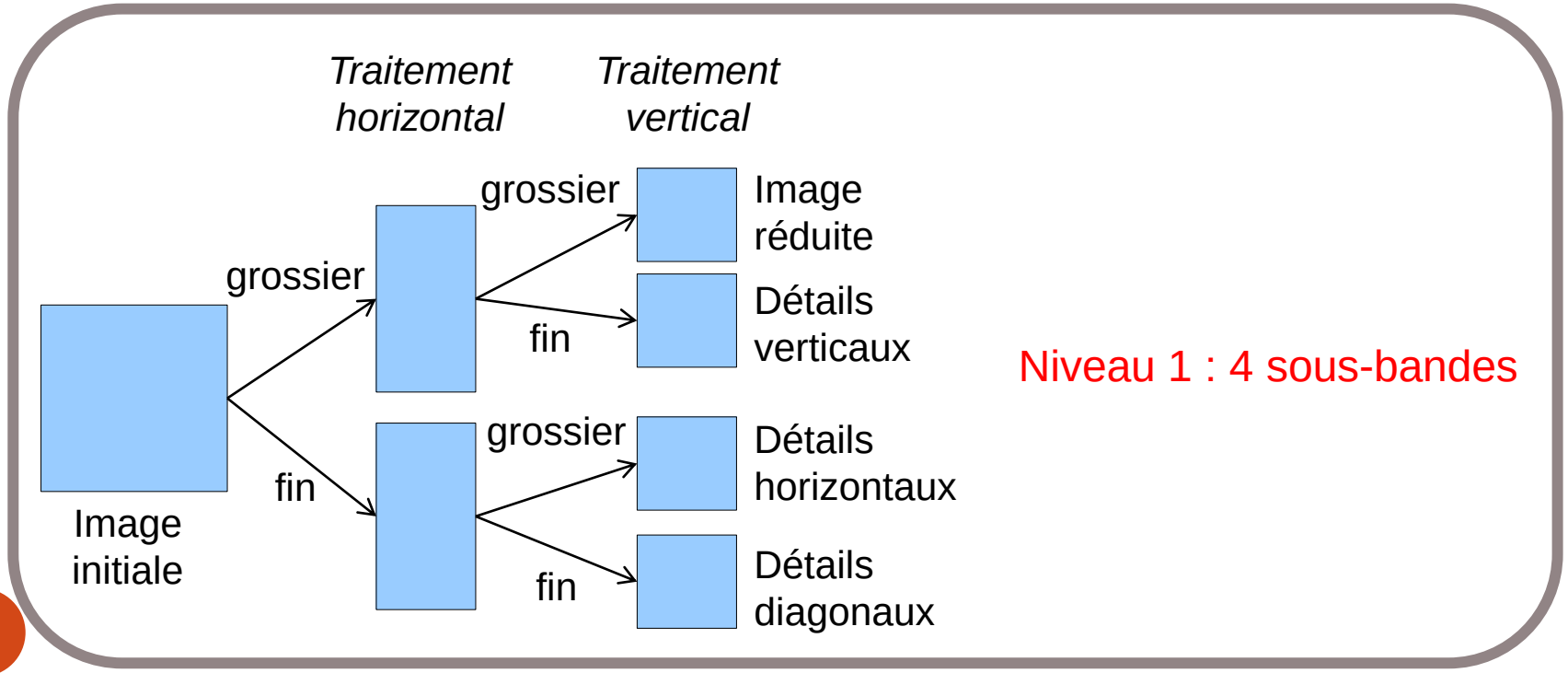
Décodeur



Principes de la compression JPEG2000

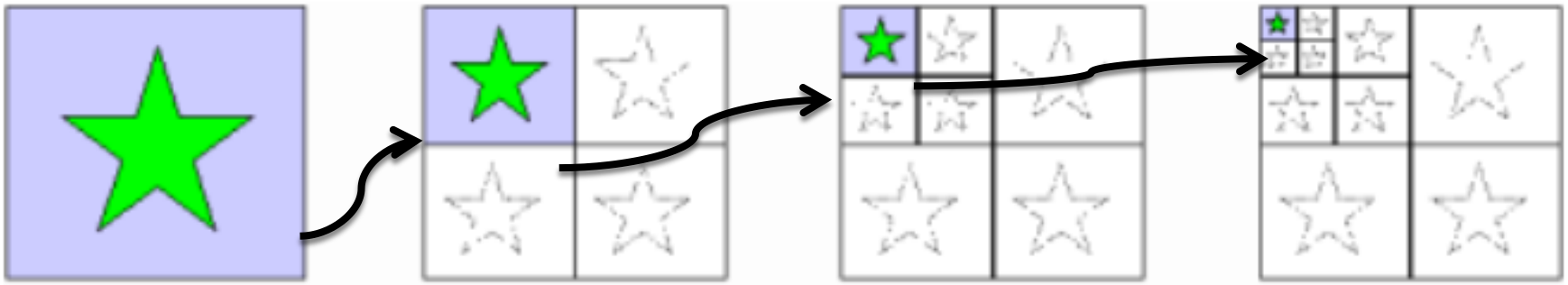
Repose sur une analyse multi-résolution de l'image, appelée **la transformée en ondelettes.**

A chaque niveau de résolution, création de nouvelles images, de résolution inférieure, appelées sous-bandes. L'une contient les détails grossiers (sous-bande d'approximation), les autres contiennent les détails fins (sous-bandes de détails).



Aspect multi-résolution

- La transformée peut être réitérée sur la sous-bande d'approximation.
- Il en résulte une représentation **hiérarchisée** de l'information contenue dans l'image.
- Pour une image naturelle, l'intensité est concentrée dans la sous-bande d'approximation, ce qui est utile pour la compression.
- Cette transformée est réversible.

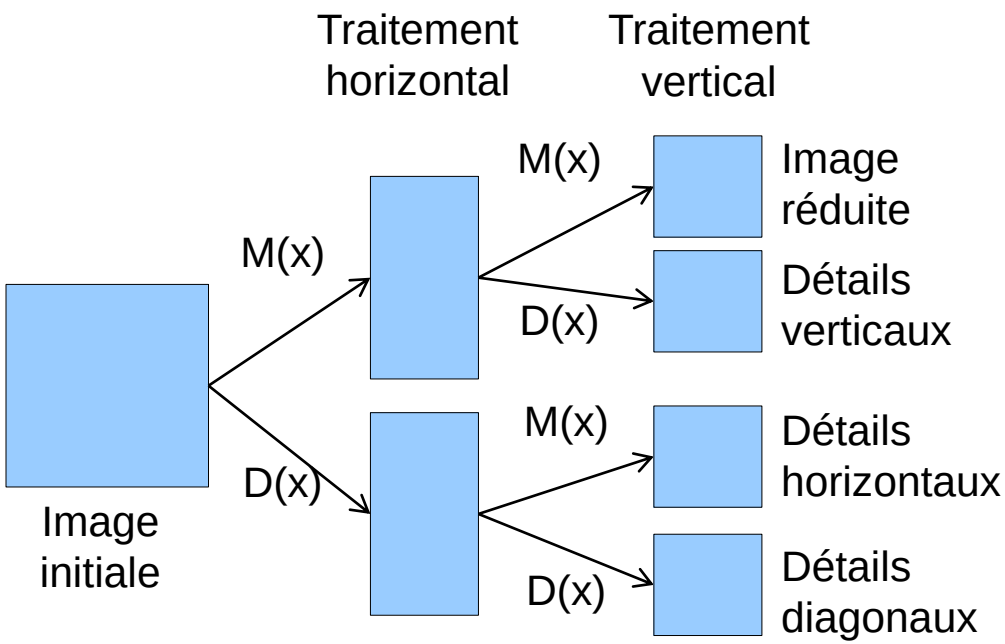


Niveau 0 *Niveau 1* *Niveau 2* *Niveau 3*

Exemple : base de Haar

Utilise deux fonctions :

- Approximation obtenue par **moyennage** de 2 pixels voisins.
- Détails obtenus par **différences** entre 2 pixels voisins.



Exemple : base de Haar

Image originale

Transformée dans la base de Haar sur un niveau

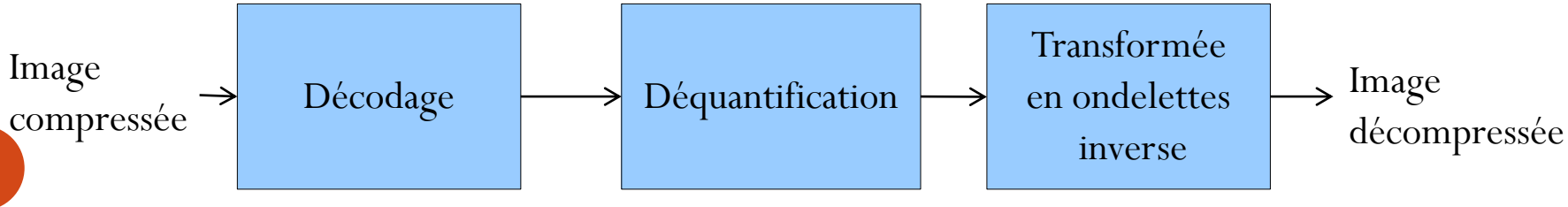
Synoptiques fonctionnels d'un codeur et d'un décodeur JPEG2000

- Comme la DCT, la transformée par ondelettes n'induit **pas de pertes.**

- Les pertes sont introduites par une opération de **quantification sur les sous-bandes de détails.** Suppression des détails en dessous d'un seuil donné.

Codeur

Décodeur



Avantages de JPEG2000

- Pas d'effet de blocs.

En JPEG, il provient du découpage en blocs de 8×8 pixels.

- Meilleurs taux de compression.

En JPEG, la redondance est exploitée à l'intérieur d'un bloc, pas sur l'ensemble de l'image, contrairement à la transformée par ondelettes.

- Taille de fichier prévisible. Taux de compression directement programmable (formule mathématique liant le taux et le seuil).

Utilisation

- Compression JPEG2000 pour les images fixes

- MPEG4 : Pour habillage des structures 3D par des images naturelles (fixes)

  Intérêt : le multicouche permet d'adapter le rendu des détails à la distance de visualisation (position de l'utilisateur dans l'environnement 3D).

IV - La compression des vidéos

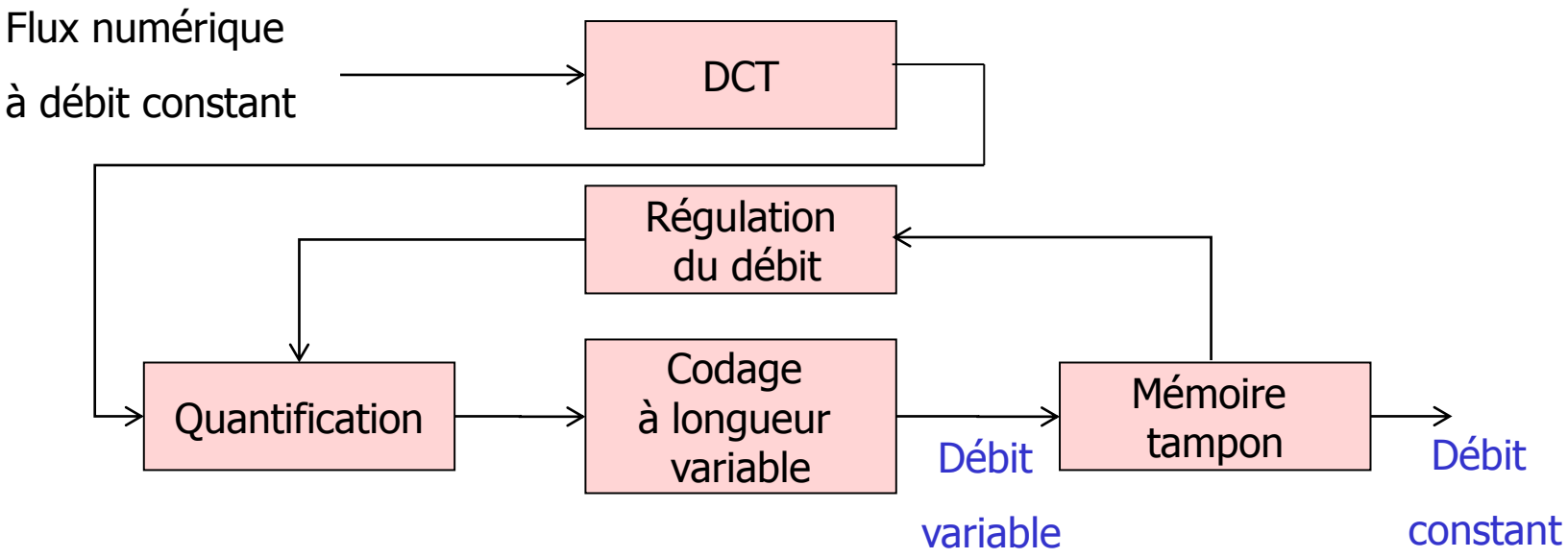
La compression MJPEG

- *Moving* JPEG (MJPEG) **n'est pas une norme** !! C'est un algorithme propriétaire, donc inadaptable d'un appareil à l'autre.

- MJPEG consiste en la **compression des images successives** d'une vidéo de façon indépendante.

- **Origine** : Le montage d'images compressées dépendantes les unes des autres est difficile car il y a rupture de la séquence. Les monteurs ont appliqué la compression JPEG individuellement aux photogrammes constituant une vidéo ⇨ « MJPEG ».

Fonctionnalité d'un codeur MJPEG



- Afin de garder un débit constant, le signal est stocké dans une mémoire tampon.
- Lorsque la mémoire déborde (par exemple, car images complexes donc DCT large), on sous-quantifie les prochaines images en changeant de table de quantification, voire on supprime certaines images.
- Les tables de quantification utilisées sont celles normalisées dans JPEG.

La compression MPEG

La compression MPEG repose sur la transformation DCT de l'image, comme dans JPEG et MJPEG.

L'algorithme part du principe que deux images successives dans une vidéo se ressemblent : on peut alors **transmettre uniquement les informations qui ont changé** entre les deux images.

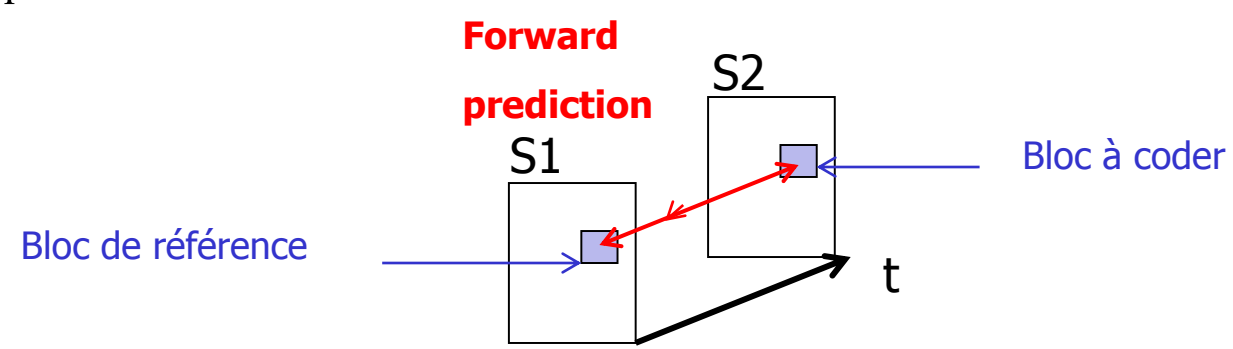
Ainsi, les images d'une vidéo peuvent être codées de deux manières :

- **En intra** : c'est-à-dire, comme une image indépendante des autres images de la vidéo, selon un algorithme identique à celui de JPEG.
- **En prédiction** : on ne transmet que les différences (compressées) entre l'image courante et l'image intra la précédant (ou la suivant).

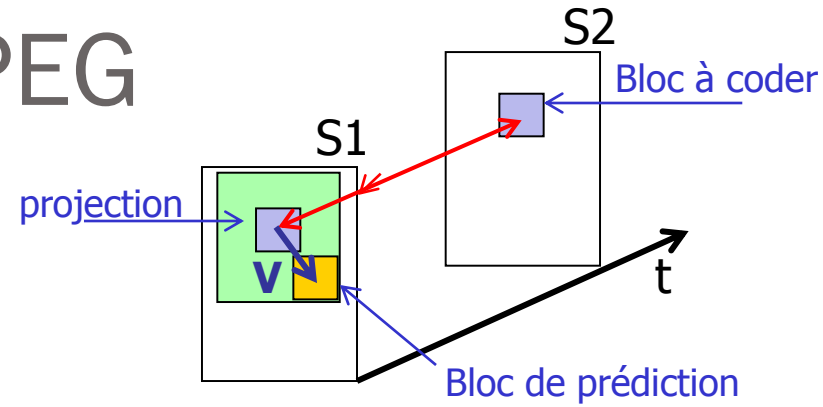
La compression MPEG

① Mesure de la différence entre deux images successives

- Image S1 codée en intra.
- Calcul pour chaque bloc de 8×8 pixels de « l'erreur » D= S2-S1. D est une matrice de 8×8 valeurs.
  - ☞Si les deux images se ressemblent, les coefficients de la matrice sont proches.
- Calcul de la DCT de la matrice D.
  - ☞Si les deux images se ressemblent, les coefficients **BF** de la matrice D sont **plus importants** que les coefficients HF. Donc on peut les quantifier **efficacement**
- Au lieu de transmettre l'image S2, on transmet la matrice des coefficients DCT de D quantifiés, qui est bien moins volumineuse.



La compression MPEG



② Estimation du mouvement subi par les blocs

Si beaucoup de mouvement,

- S2 est très différente de S1
- Les coefficients de la matrice D sont très différents
- Les coefficients DCT haute fréquence de D sont non négligeables : on ne peut pas les quantifier sans dégradation importante.

Le codeur va **estimer le déplacement du bloc dans l'image**.

Le codeur ouvre une fenêtre de recherche dans S1 autour du bloc de référence.

Il balaie la fenêtre et calcule pour chaque bloc rencontré dans la fenêtre l'erreur (opération de *block matching*).

Le bloc de S1 qui minimise l'erreur est retenu.

Le codeur transmet

- Les **coefficients DCT quantifiés** de la matrice d'erreur
- Le **vecteur de mouvement du bloc.**

Illustration de la Forward-Prediction

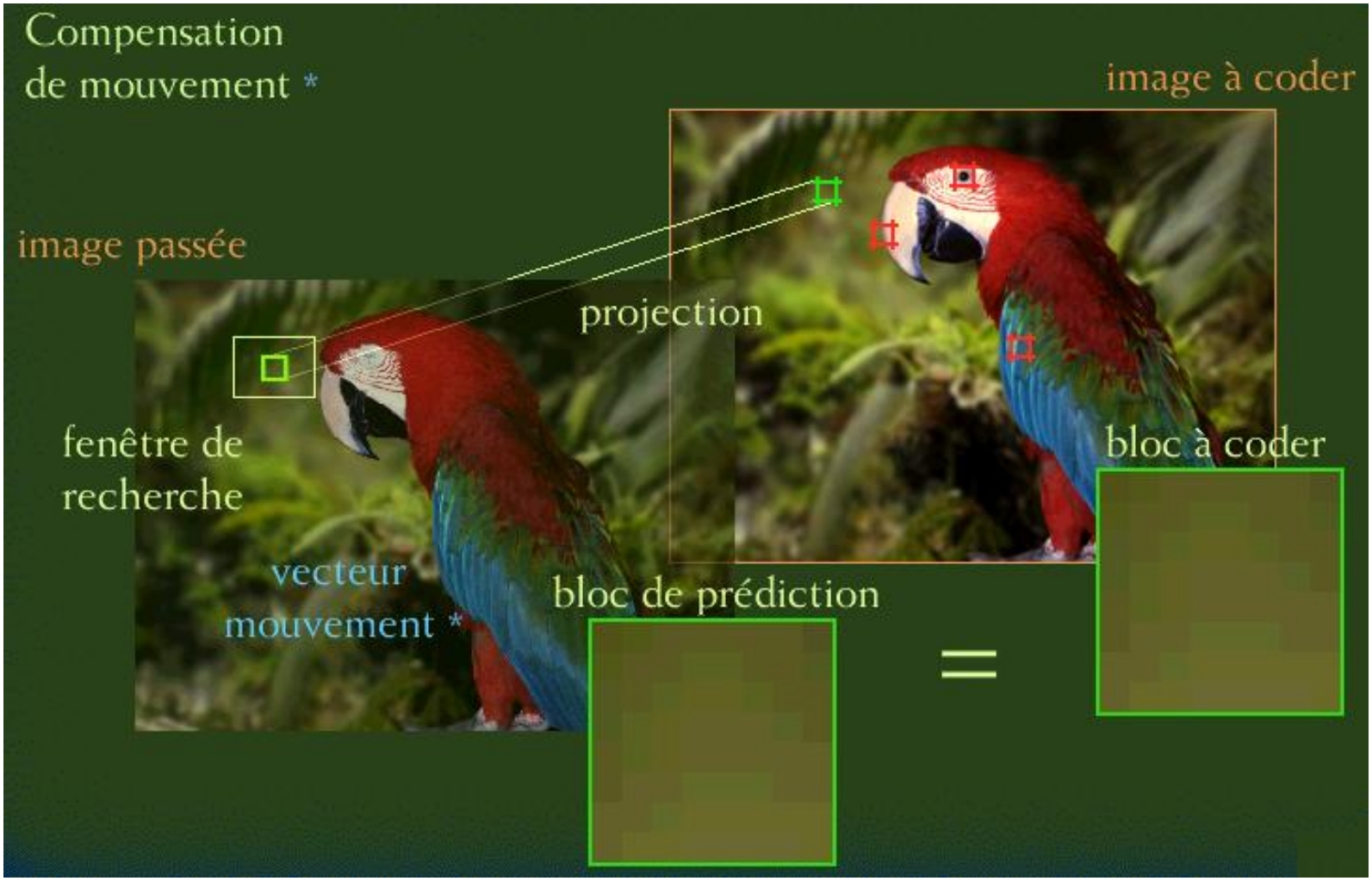
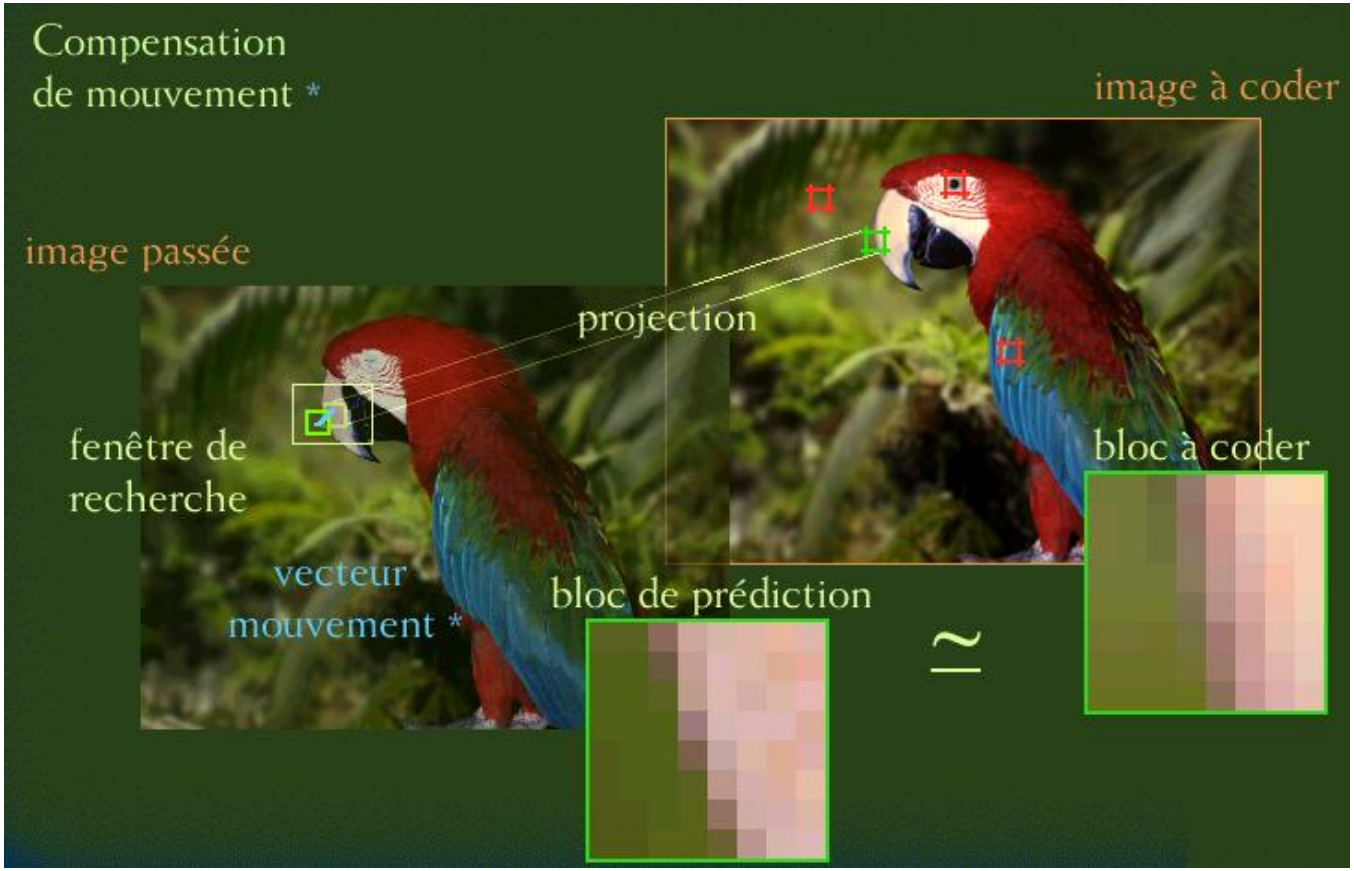


Illustration de la compensation de mouvement



☞ En réalité, on ne transmet pas un vecteur de mouvement pour chaque bloc, mais pour chaque **macro-bloc = 4 blocs adjacents.**

Les types d'images d'une séquence codée en MPEG

Dans une même séquence vidéo, trois types de codage des images :

Image I : en Intra

Image P (prédite, *forward prediction*) : par rapport à une image passée

Image B (bidirectionnelle, *forward prediction* et *backward prediction*) : par rapport à une image passée et une image future. Le codeur transmet la moyenne des erreurs et transmet les 2 vecteurs.

Le codeur choisit le mode de prédiction le plus rentable.

A l'intérieur d'une image P : macroblocs I ou P.

A l'intérieur d'une image B : macroblocs I, P ou B.

En moyenne, pour un même contenu,

volume(P) = volume(I) / 2

volume(B) = volume(I) / 4

(dépend de l'algorithme utilisé par le constructeur)

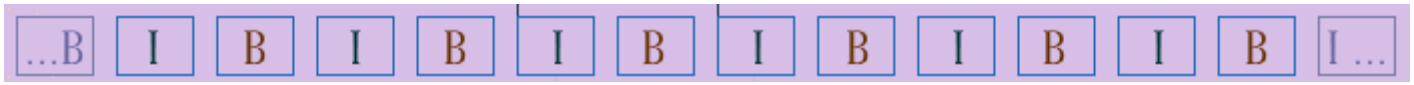
Le GOP

Le GOP (*Group Of Picture*) définit l'enchaînement des images I, P et B dans le flux vidéo.

Le GOP commence toujours par une image I.

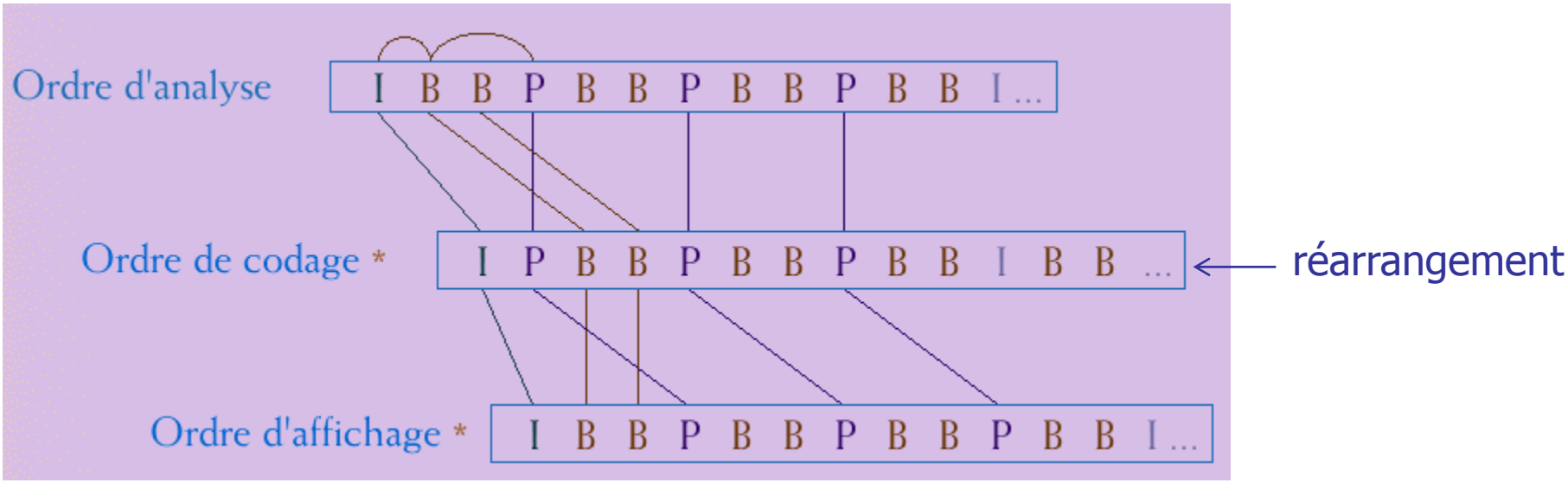
On note GOP(M,N), où M est la distance entre deux images P et N la distance entre deux images I.

Question : Un monteur peut-il travailler sur des vidéos dont les GOP sont représentés ci-dessous ?

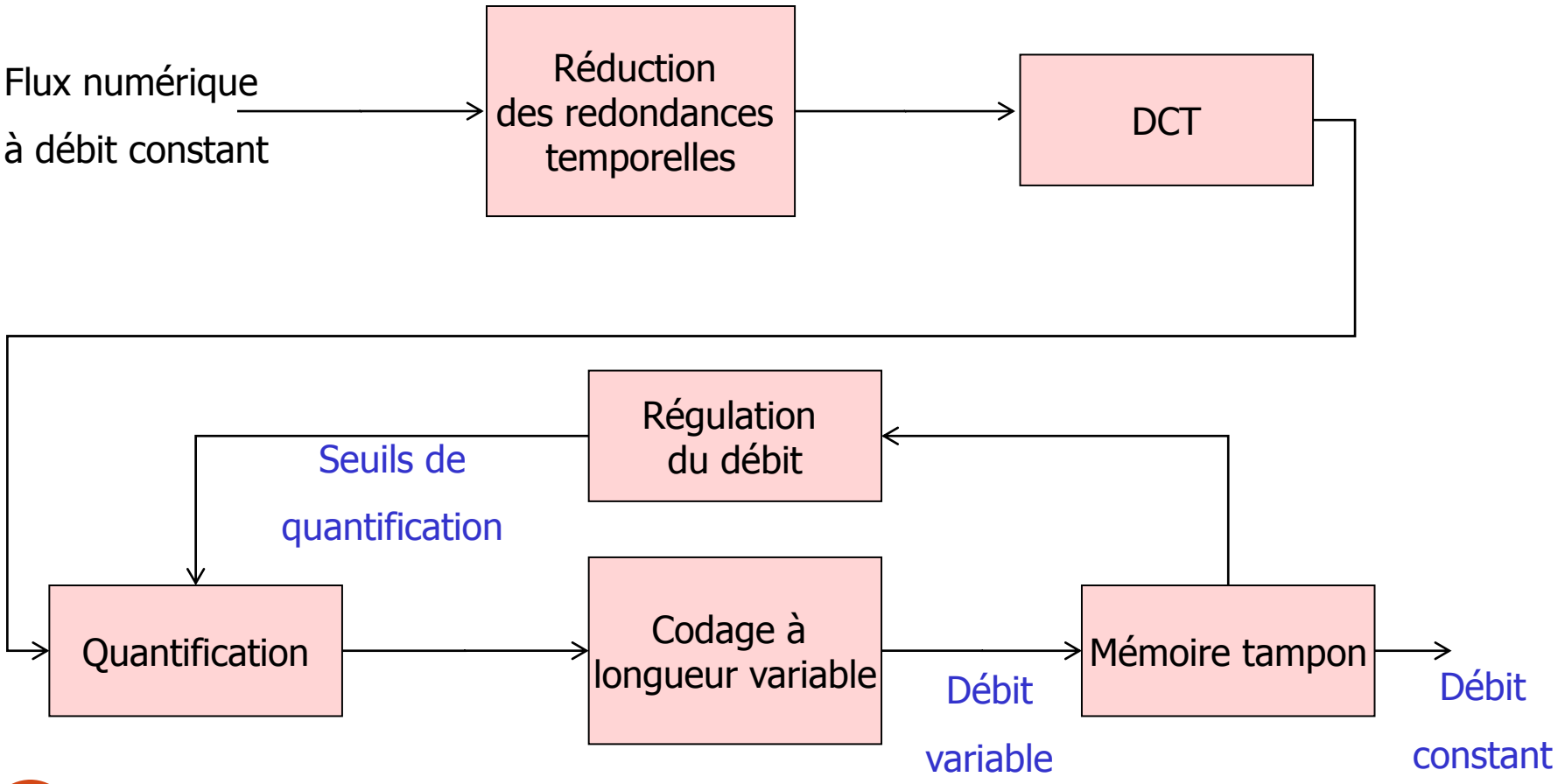


Réarrangement du GOP

Si l'enchaînement des images contient des images B, le codage et décodage ne peuvent être effectués dans l'ordre d'arrivée de la séquence : **réarrangement** préalable des images.



Synoptique du codeur MPEG



Paramètres des normes MPEG

Structure d'échantillonnage

- En théorie, un algorithme de compression MPEG accepte n'importe quel format en entrée (4:2:2, 4:2:0, SIF, SDTV, HDTV, …).

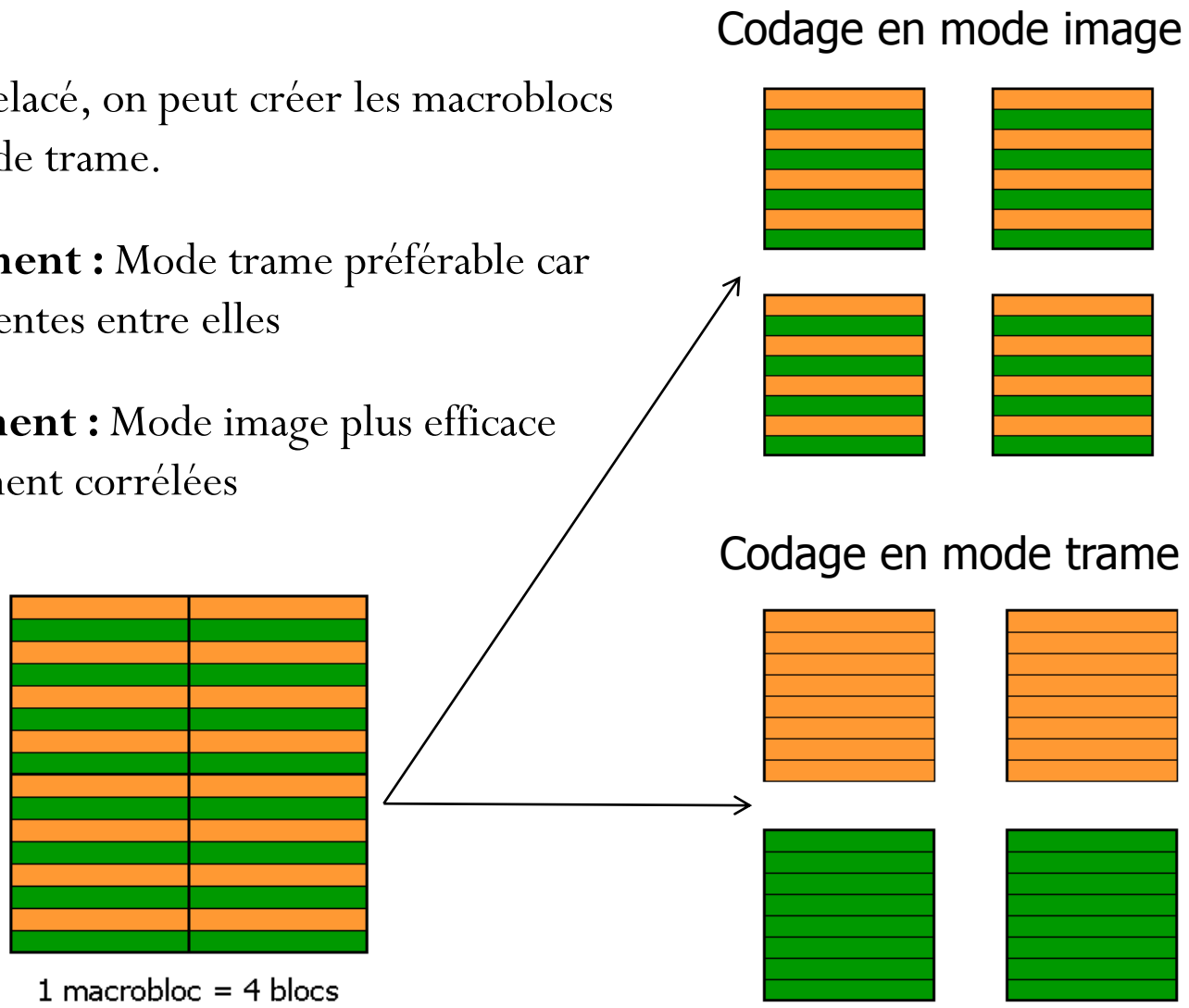
- La restriction se fait au niveau des profils (recommandations) décrits dans les normes.

Paramètres des normes MPEG

Dans un signal vidéo entrelacé, on peut créer les macroblocs en mode image ou en mode trame.

- **Image avec mouvement :** Mode trame préférable car les trames sont peu cohérentes entre elles

- **Image sans mouvement :** Mode image plus efficace car les trames sont fortement corrélées



La norme MPEG 1

Application : l'enregistrement

- CD-ROM, disques optiques
- VCD (magnétoscope numérique en Chine)
- Utilisée par la TV numérique aux USA en attendant la mise au point du MPEG2

Débit

- Comprime des images animées + son stéréo avec un débit de 1.5Mbit/s

Format d'entrée

- En général, format SIF
- Performances réduites pour un signal entrelacé (ex. : signal TV…)

MPEG1 définit **3 types d'images** : I, P, B et **2 paramètres** : GOP(M, N)

La norme MPEG 2

La norme **MPEG2** a été mise au point pour **le stockage et la diffusion de la vidéo numérique.** C'est la norme de compression utilisée dans la **télévision numérique terrestre** (TNT ou DVB-T).

Amélioration de MPEG1

- En tenant compte de l'entrelacement
- En rajoutant des outils de gestion des erreurs de transmission

Formats d'entrée acceptés

- 4:2:2, 4:2:0, 4:4:4
- EDTV, HDTV, futurs formats super HDTV
- RVB et composantes
- Formats informatiques
- Balayage entrelacé ou progressif : à l'intérieur d'une même image, le codeur peut choisir de traiter des blocs trame à trame ou en désentrelacé.

La norme MPEG 2

S'adapte à diverses qualités d'images en télévision

- 2 Mbit/s : qualité comparable au VHS
- 3 à 5 Mbit/s : qualité comparable au PAL
- 8 à 10 Mbit/s : qualité comparable au 4:2:2

Définit

- 4 niveaux (définition des paramètres utilisés)
- 6 profils (applications dédiées)
- 3 types d'images : I, P, B
- 2 paramètres : M et N

La donnée d'un couple profil/niveau garantit l'interopérabilité des équipements.

Profils et niveaux de la norme MPEG 2

CODEC bas coût TV numérique — Studio

4 niveaux *	6 profils *					
	Simple pas d'images B	Main 4:2:0	SNR TV améliorée	Spatial TVHD	High 4:2:2	4:2:2 post-prod.
		structures »				
High TVHD		80			100	
High 1440 TVHD		60	60	60	80	
Main TV	15	15	15	15	20	50
Low 1/4 TV		4				

TV numérique actuelle

(débits en Mb/s)

Difficulté technique croissante

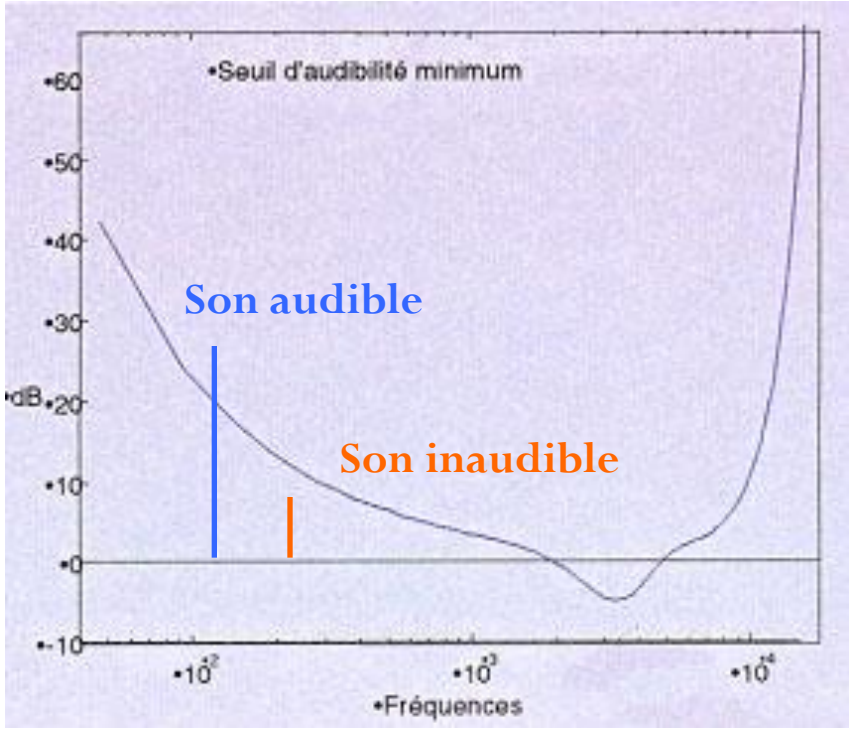
Qualité croissante

V – La compression audio

Seuil d'audibilité minimum

La compression **audio MPEG** exploite les limites de l'audition humaine.

A une fréquence donnée, notre oreille ne peut percevoir un son qu'à condition que son niveau de pression acoustique soit supérieur à un seuil. Le seuil est différent pour chaque fréquence de la bande audible.



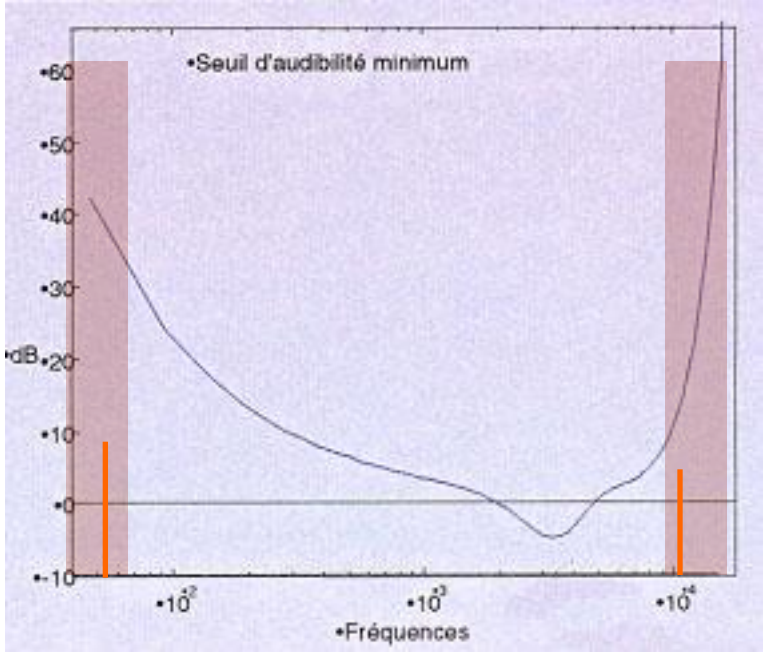
Courbe de sensibilité de l'oreille humaine dans un environnement calme

Seuil d'audibilité minimum

Conséquence :

Il est possible de supprimer « plus ou moins » (en fréquences) les sons hautes et basse fréquences.

Par exemple, l'enregistrement au format mp3 à 64kbit/s supprime **tous les sons de fréquences supérieures à 11kHz**. Dans ce cas une forte compression est obtenue au prix d'une baisse de la qualité d'écoute.

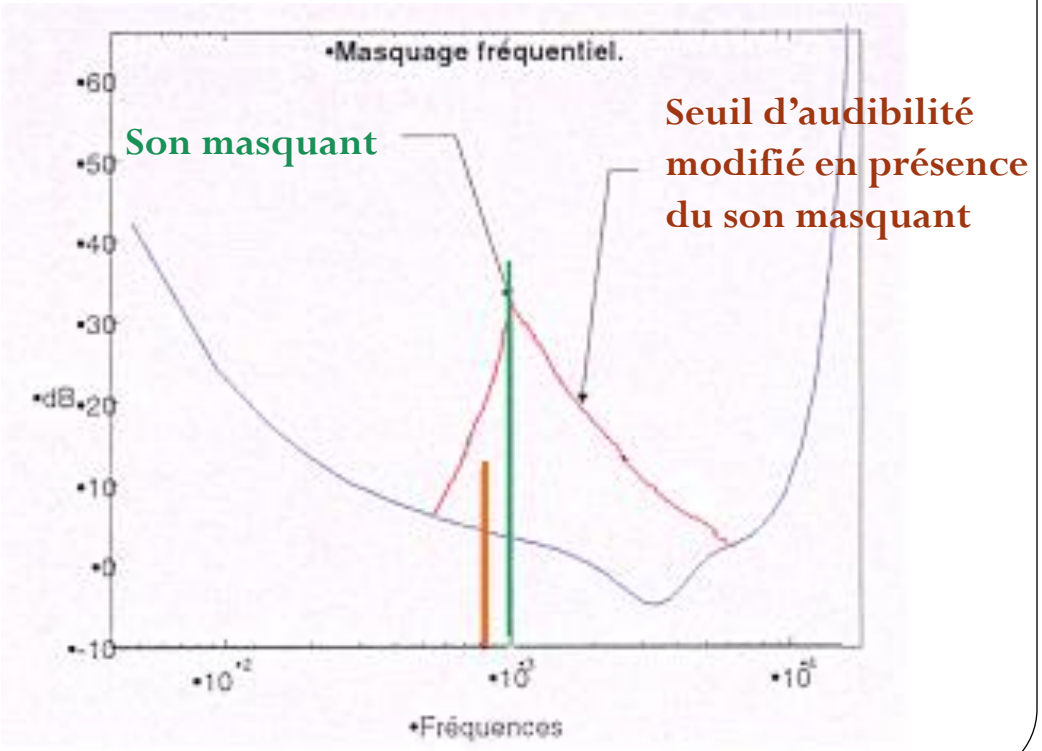


Courbe de sensibilité de l'oreille humaine dans un environnement calme

Le masquage fréquentiel

Un son d'amplitude supérieure au seuil d'audibilité peut devenir inaudible en présence d'un autre son pur de fréquence proche et d'amplitude plus importante. C'est l'effet de **masquage fréquentiel**.

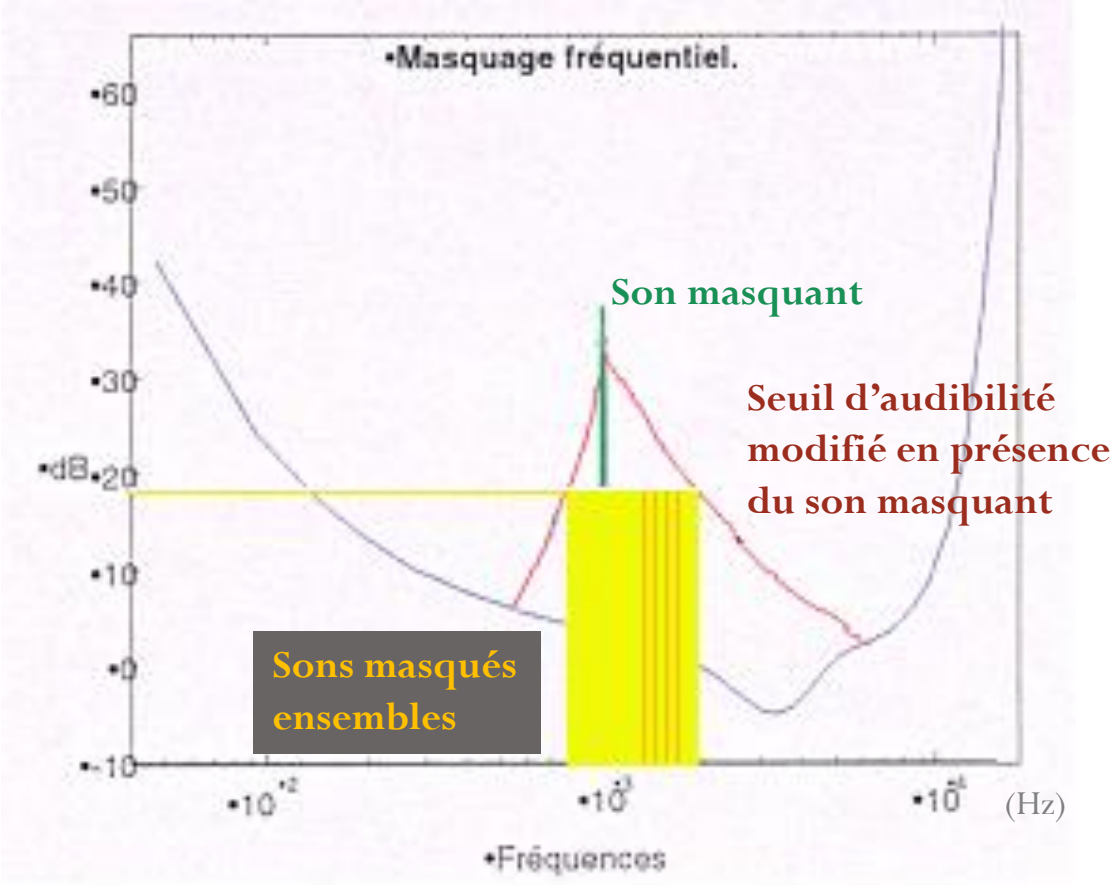
- Le seuil d'audibilité est relevé en présence du son dit « son masquant ».



Variation du seuil d'audibilité masquage fréquentiel

Le masquage fréquentiel

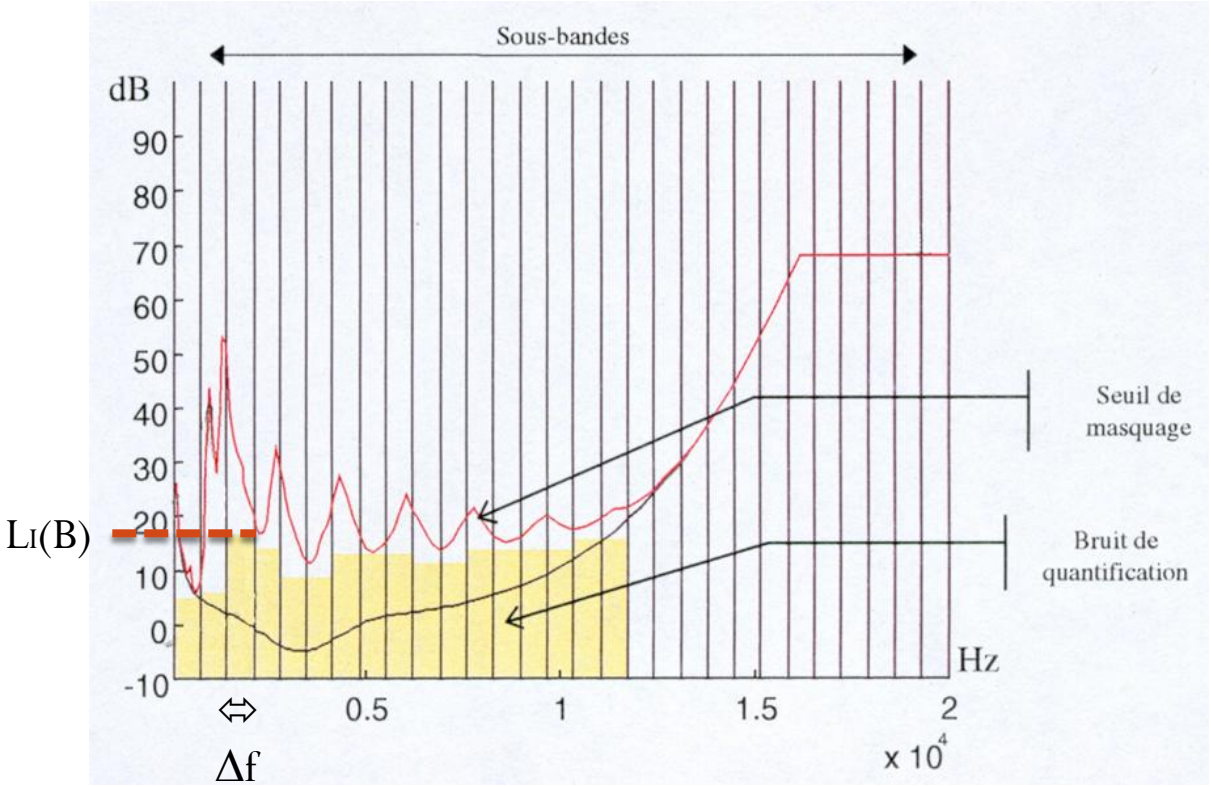
Plusieurs sons de fréquences proches les unes des autres peuvent être masqués ensemble par relèvement du seuil d'audibilité…



Variation du seuil d'audibilité masquage fréquentiel

Le masquage fréquentiel

Quantification du signal sonore ⟶ bruit de quantification



Dans la bande de fréquence Δf, le bruit de quantification ne sera **pas audible si le coefficient le plus élevé de ce bruit correspond à LI**. (Idem dans les autres bandes)

Plus le bruit est grand, plus la profondeur de quantification est faible.

$$L_I(S)-L_I(B) = 6n+2$$

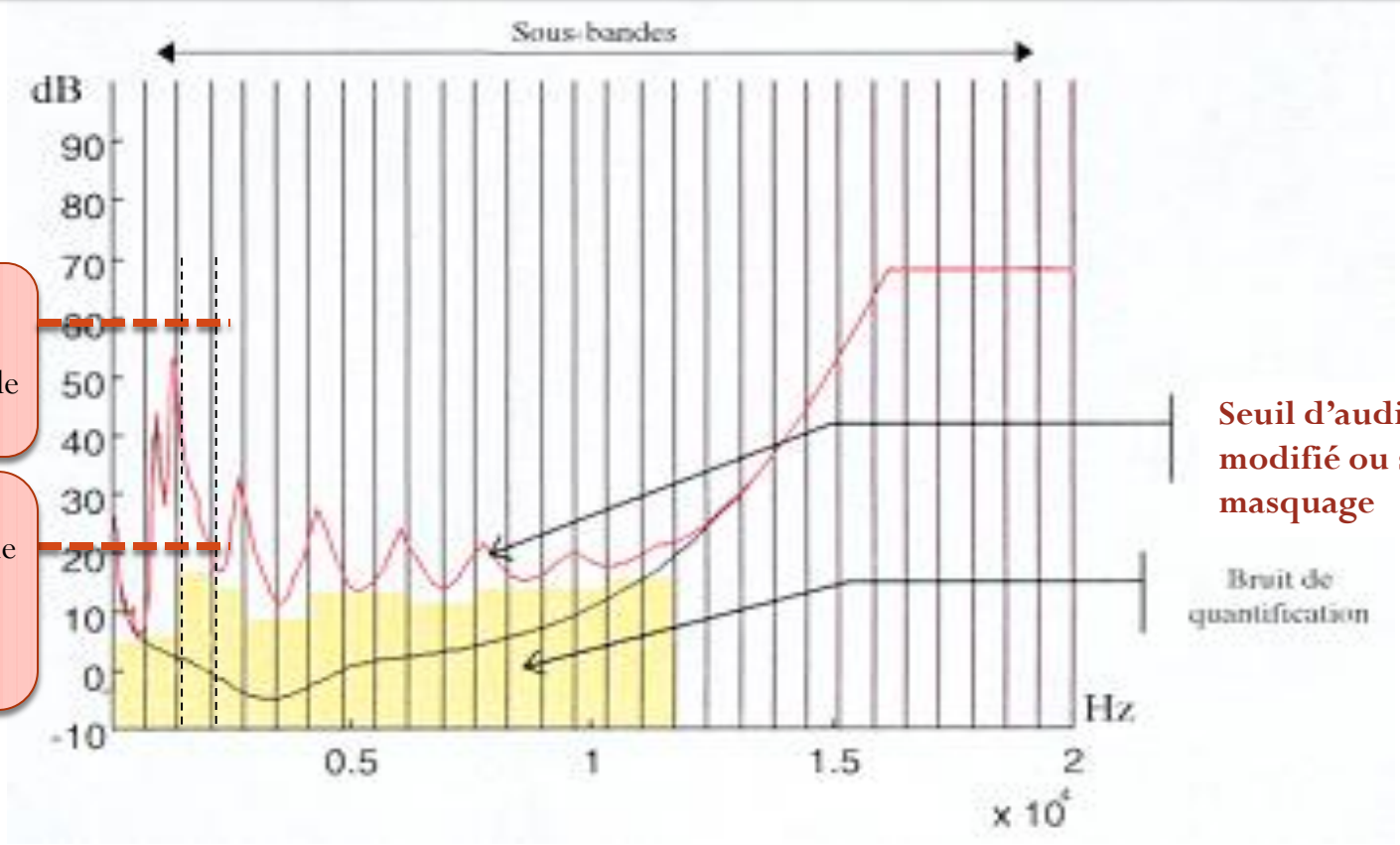
➢ On a donc intérêt à exploiter le seuil de masquage pour optimiser au mieux la quantification (augmenter le bruit donc diminuer la profondeur de quantification) et ce, en découpant le spectre en bandes de largeurs Δ f.

Le masquage fréquentiel

Exemple : Soit le signal sonore dans la bande de fréquence Δf de niveau maximal 60dB.

$L_I(S)$ = Niveau maximal du signal sonore dans la bande de fréquences Δf

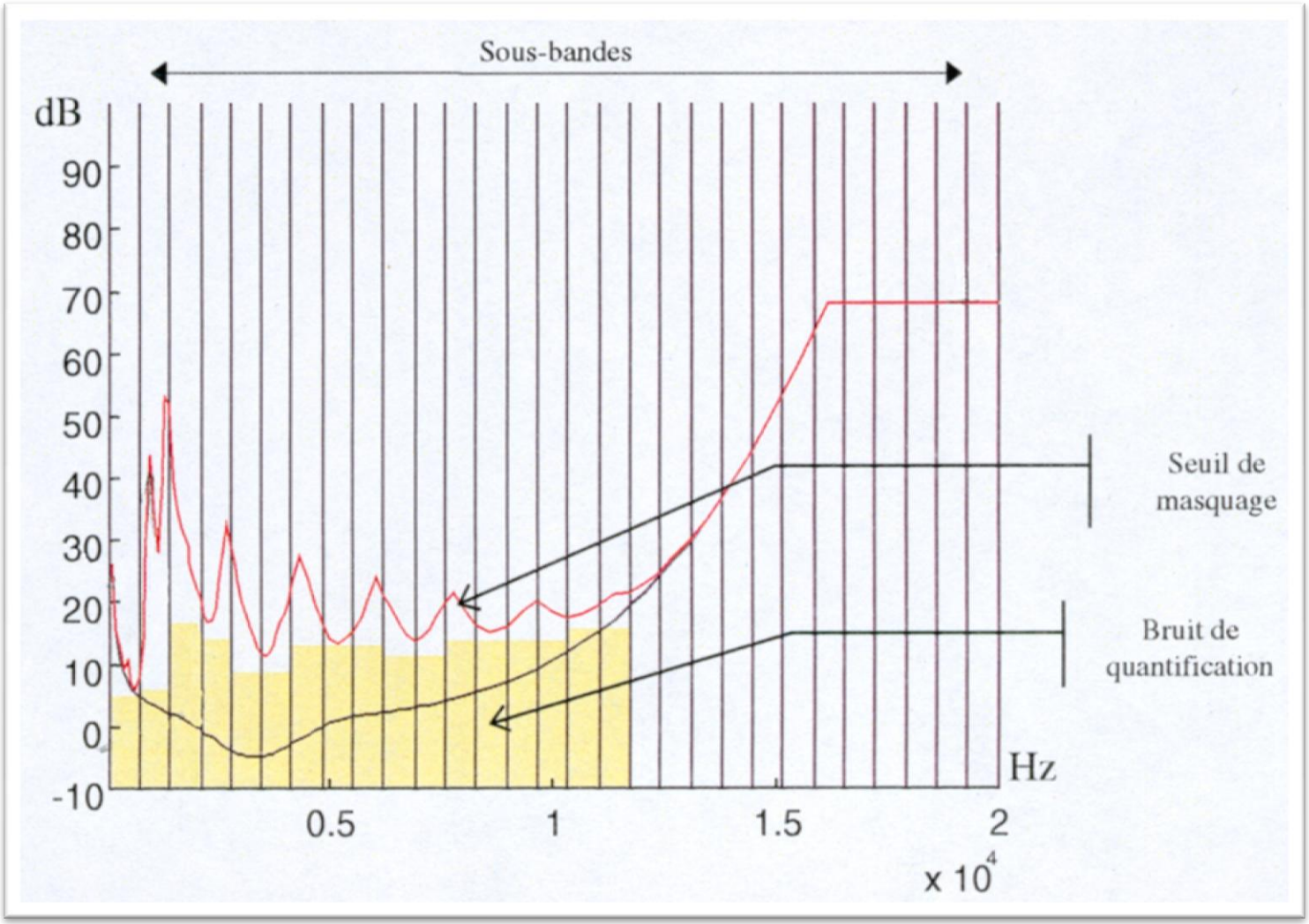
$L_I(B)$ = Niveau maximal du bruit de quantification, inaudible, dans la bande Δf .



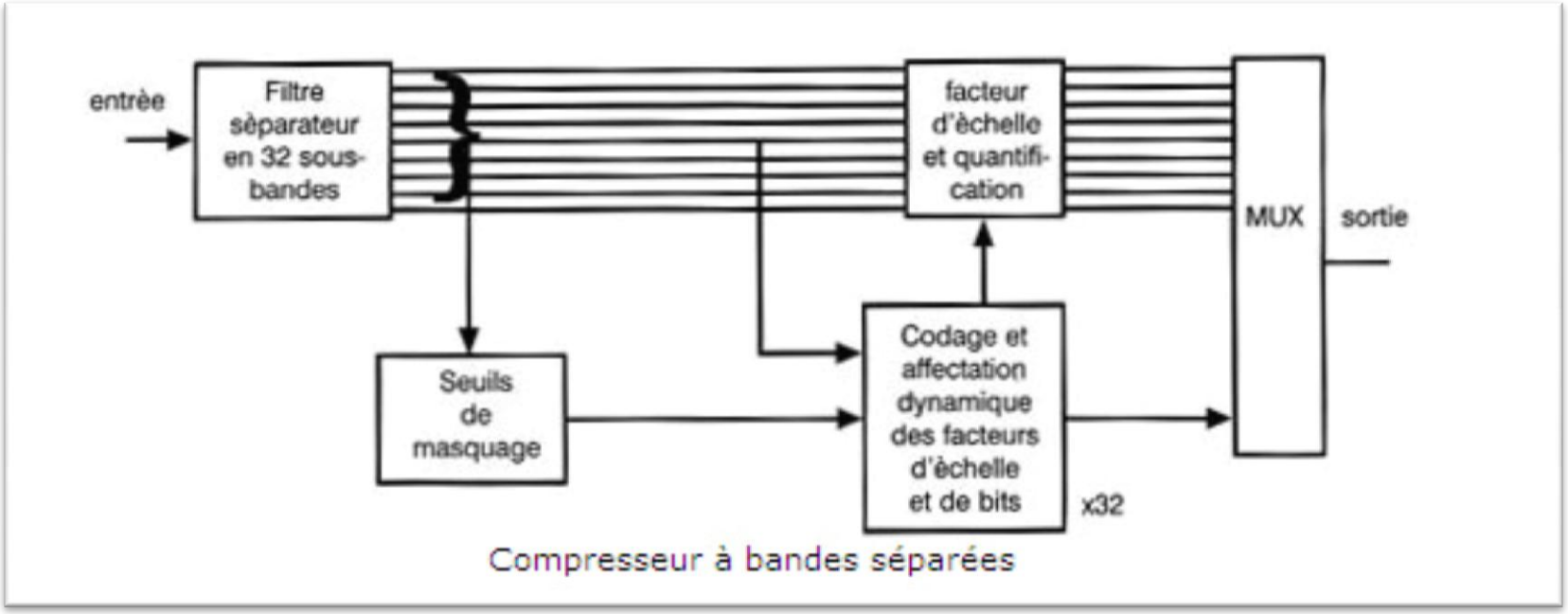
Le masquage fréquentiel

- Traitement sur une trame audio de 8ms
- Le spectre audible est divisé en 32 sous-bandes.
- Pour chaque sous-bande, on détermine la fréquence dominante et on en déduit la nouvelle courbe de seuil d'audibilité (à l'aide de modèles psycho-acoustiques).
- Ces modèles sont sauvegardés dans la ROM des codeurs.
- Dans chaque sous-bande, le codeur choisit le pas de quantification adapté au niveau du seuil d'audibilité.
- Le flux numérique contient les valeurs du pas de quantification de chaque sous-bande pour que le décodage soit possible.

Sous-bandes audio MPEG

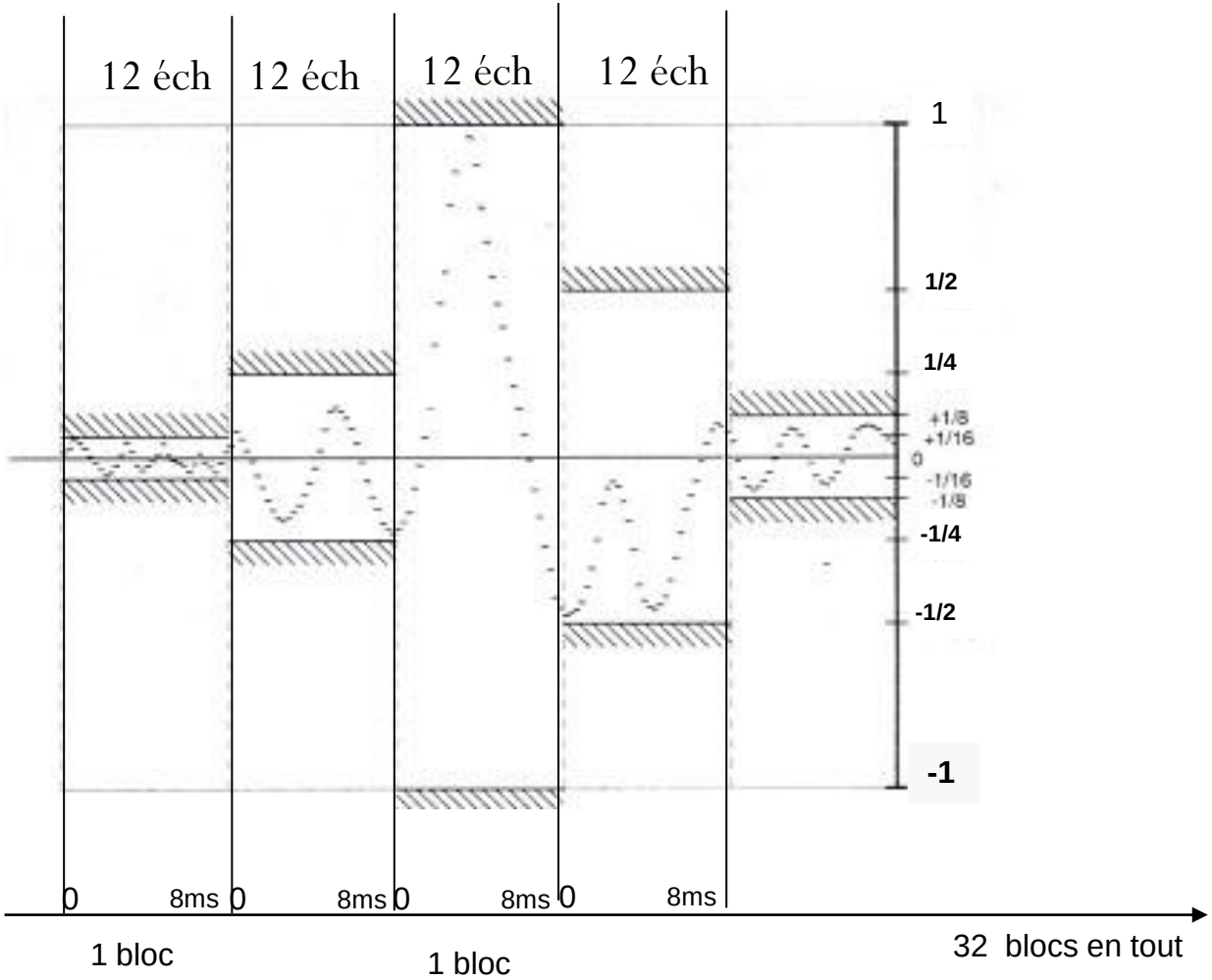


Sous-bandes audio MPEG



Compresseur à bandes séparées

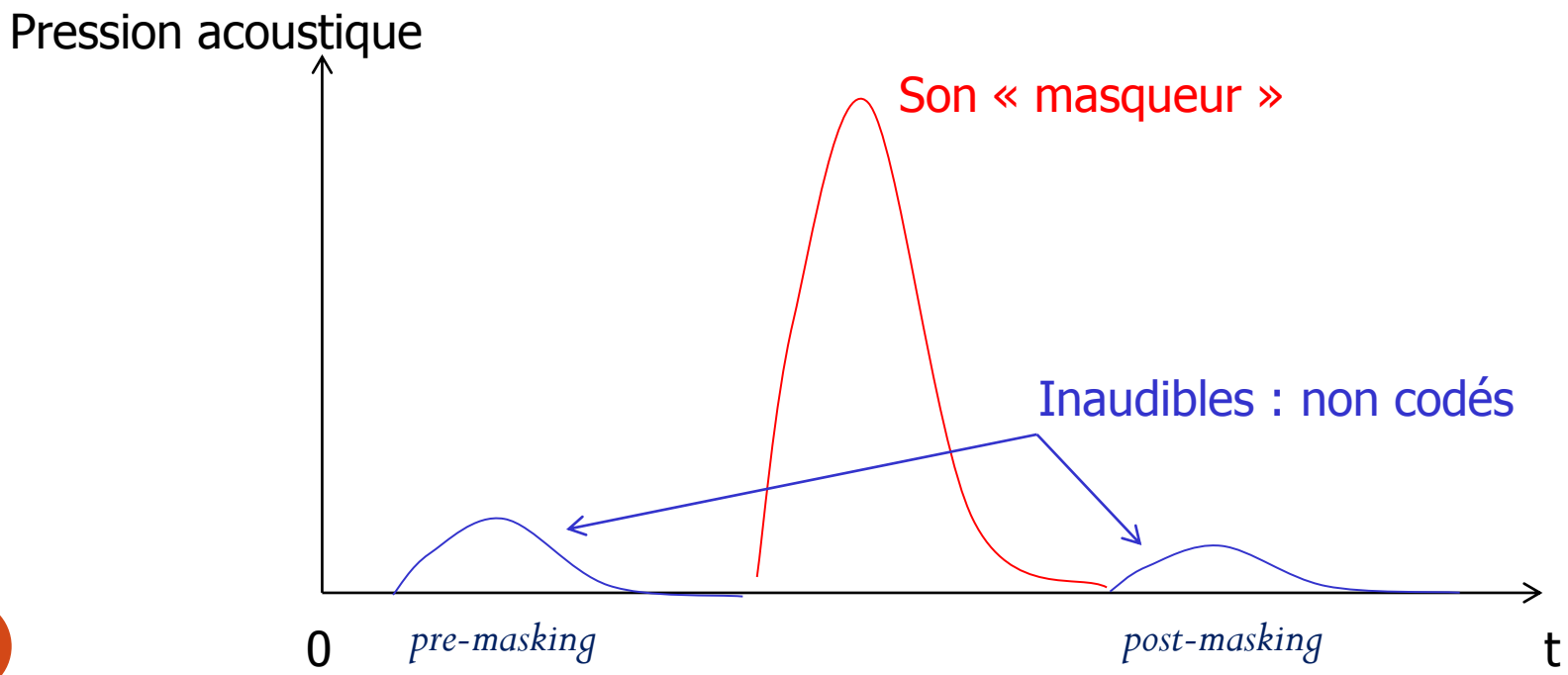
Sous-bandes audio MPEG



Le masquage temporel

L'oreille n'entend pas un son s'il est produit dans un court délai avant ou après un son **de même fréquence** de plus grande amplitude.

↳ On ne code pas les sons masqués temporellement.



Le masquage de canal

L'oreille humaine est incapable de détecter la provenance d'un son s'il a une fréquence trop basse.

↳ Coder en mono les sons basse fréquence (= **Joint Stéréo**).

Remarques :

✦Pour obtenir la meilleure qualité possible en **mode Stéréo**, le débit doit être assez élevé.

✦ Aux faibles débits, le son **mono** est préférable au **Joint Stéréo** et au **Stéréo**.

La compression audio MPEG 1

- Fréquences d'échantillonnage : 32 à 48 kHz
- Débits : 32 à 384 kbit/s
- Trois niveaux (*layers*) sont définis :

Niveau	**Technique**	**Utilisation Performances**
Niveau 1	Masquage fréquentiel	Qualité CD
Niveau 2	Masquage fréquentiel Masquage temporel	Qualité CD Utilisé dans le DVB-S
Niveau 3	Masquage fréquentiel avec largeur des sous-bandes adaptée Masquage temporel Codage entropique Exploitation de la cohérence entre les voies stéréo	« MP3 »

La compression audio MPEG 2

Elle reprend l'essentiel de la norme MPEG1 audio.

Fréquences d'échantillonnage : 16 à 24 kHz
Débits qui peuvent descendre jusqu'à 8 kbits/s

Gère les signaux multicanaux jusqu'à 7 voies pleine bande + 1 canal audio seulement pour les basses fréquences (audio « surround »7.1 et 5.1)

Un décodeur MPEG1 AUDIO saura décoder les deux canaux stéréo d'un signal MPEG2 AUDIO

VI - La compression MPEG 4

La norme MPEG 4

- Développée par l'ISO (1999)

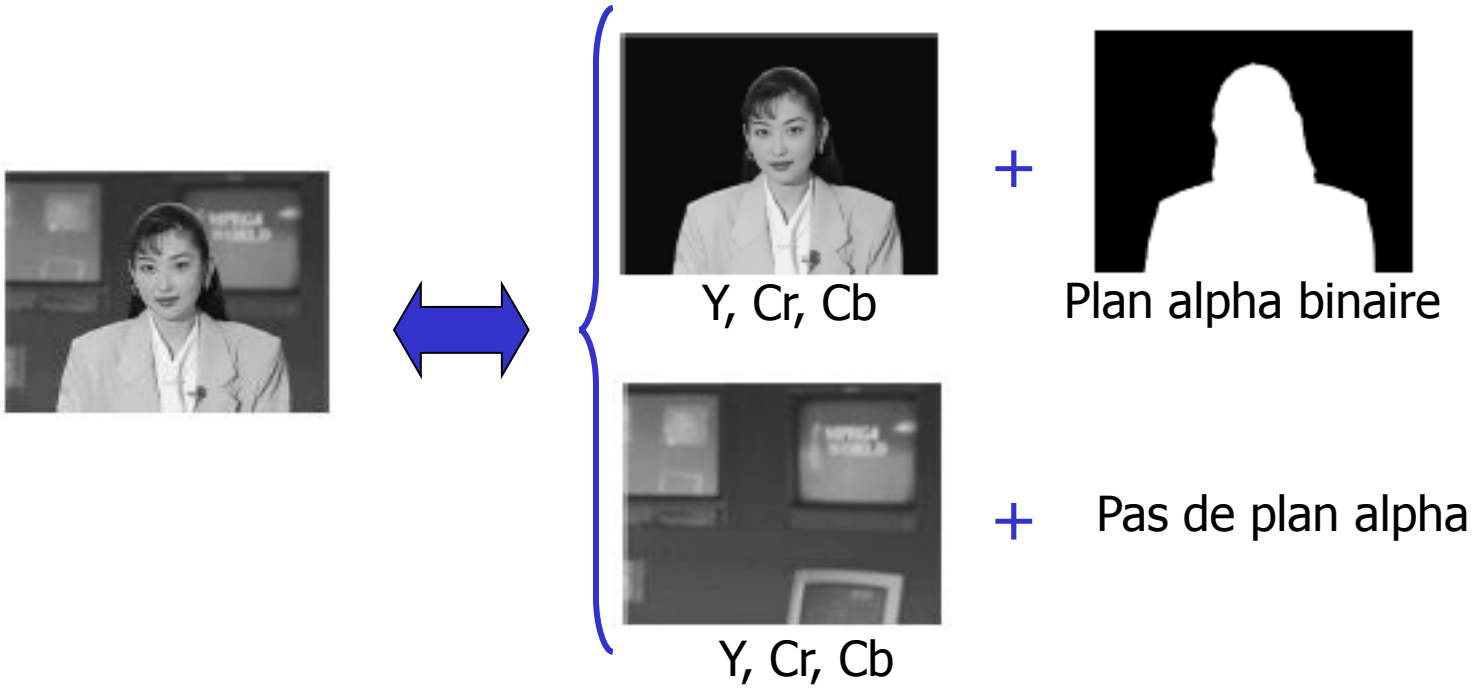
- Concerne
  La TV numérique (compression audio/vidéo naturelle)
  Les applications graphiques 2D/3D (synthèse d'images)
  Le Web (interactivité, accès, transmission)

- Nouveauté majeure : découpage des scènes en « objets »

Codage de la vidéo : le plan alpha

Un objet est décrit par

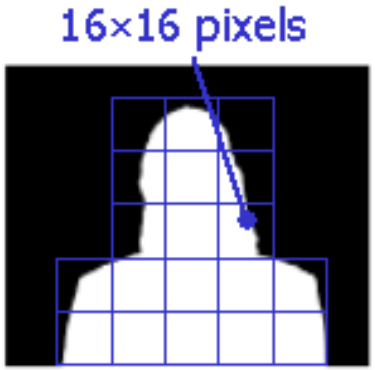
- Ses composantes Y, Cr, Cb : codées comme en MPEG 1 et 2 (sauf images fixes, cf. supra)
- Un plan alpha
  - Codé en binaire : 1 à l'intérieur de l'objet, 0 ailleurs
  - Ou codé sur 8 bits : définit en plus des niveaux de transparence



Codage de la vidéo : le plan alpha

Codage d'un plan alpha binaire : technique du codage de contour

- L'objet est découpé en blocs 16×16 pixels
- Plusieurs (7) types de BAB (Bloc Alpha Binaire) sont définis par la norme, en fonction du mouvement et de la « déformation » subis dans la scène
- Le codage de chaque type de BAB est défini par la norme : Intra, ou Inter avec compensation de mouvement
- Chaque bloc de l'objet est donc codé suivant son BAB



Codage d'un plan alpha sur 8 bits

- Composante binaire : par technique de codage de contour
- Les niveau de gris des blocs intérieurs (= niveaux de transparence) sont ensuite codés par DCT.

Codage de la vidéo : les sprites

Sprite : *« une grande image composée de tous les pixels d'un objet donné visibles à un moment ou un autre de la séquence »*

Ex. : un sprite d'objet « fond » = une vue panoramique d'un paysage sans les personnages passant au premier plan

C'est une image fixe : elle est codée par DCT en Intra et émise au début du flux MPEG.

Intérêt

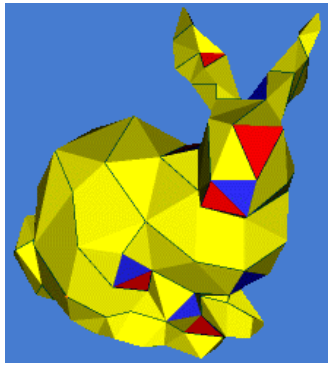
- Economie de bits
- Manipulations intéressantes sur la vidéo

Intégration de la 3D

- Pour habillage des structures 3D par des images naturelles (fixes), MPEG4 utilise la technique de compression par ondelettes, plus efficace que la DCT.

- La norme inclut également des algorithmes permettant la description des visages et leur animation.

- Elle permet aussi la construction d'objets par treillis actifs
  - Les objets sont décrits par un treillis triangulaire
  - Le décodeur calcule l'évolution de la texture d'une scène à l'autre pour chaque triangle par interpolation



Un treillis

La norme MPEG 4 audio

Codage des sons naturels

- Débits de 2 à 64 kbit/s
- La norme n'a pas pu proposer 1 seul algorithme pour tous les types de sons (musique, parole, etc. …)
- Elle inclut plusieurs standards de codage, adapté à chaque type de signal sonore
- Les algorithmes proposés dans la norme MPEG2 audio sont notamment repris.

Synthèse de sons : la norme inclut plusieurs algorithmes de génération de sons (pluie, bruits de pas, instruments de musique…).

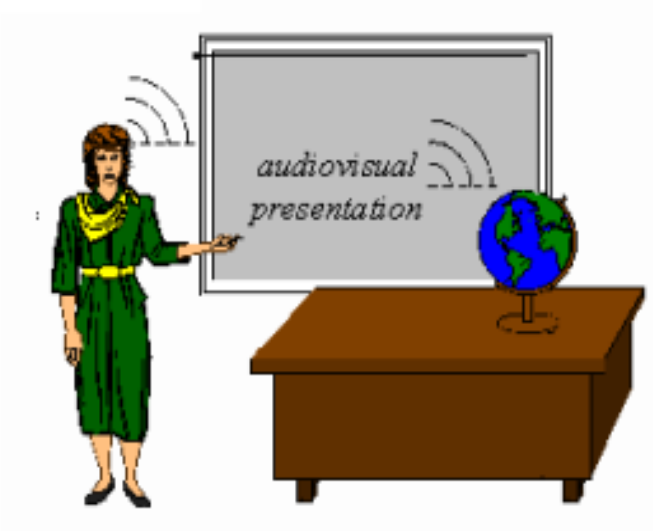
Description de scènes

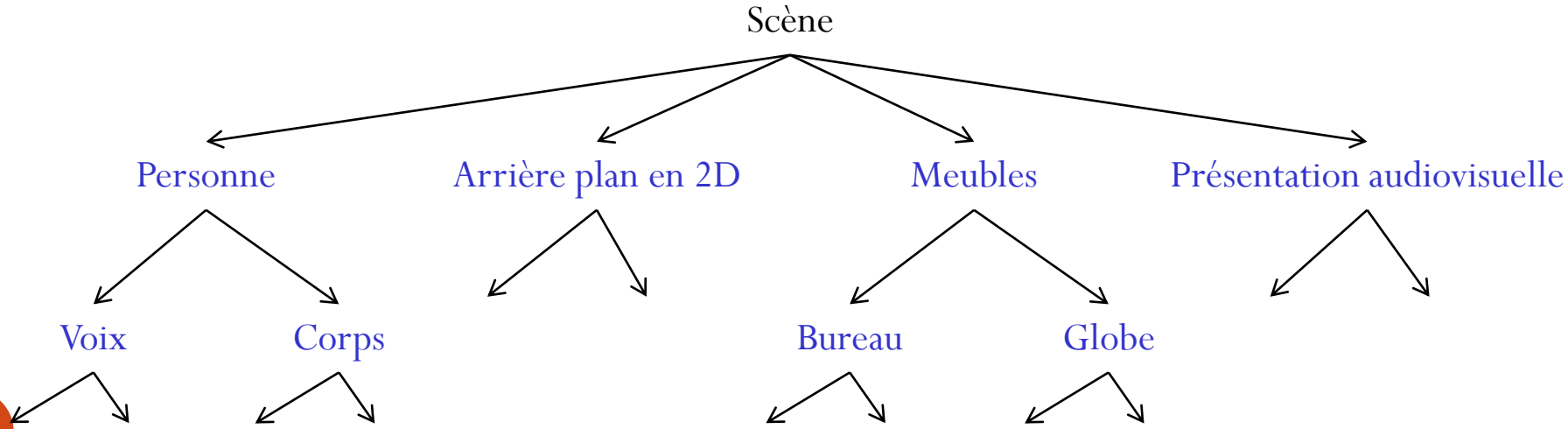
- Les scènes sont organisées sous forme d'une structure hiérarchique d'objets.

- La norme inclut un standard de description de la scène en flux binaire : BIFS (*Binary Format for Scene Description*).

- Elle permet l'utilisation du VRML (langage de modélisation de la réalité virtuelle) pour permettre l'interface de la 3D et de la vidéo.

Description de scènes





Profils et niveaux MPEG 4

1. **Profils visuels** : codage visuel des données naturelles, synthétiques ou hybrides
2. **Profils audio** : nombre d'outils implémentés et le débit visé
3. **Profils graphiques** : éléments graphiques et textuels utilisables dans une scène
4. **Profils de descripteurs de scène** : types d'informations pouvant constituer une scène MPEG4 (audio, 2D, 3D…)
5. **Profils de descripteurs d'objets** : outils disponibles pour décrire un objet (descripteur d'objet, synchronisation, information sur le contenu des objets)

Profils et niveaux MPEG 4

	NIVEAU 1	NIVEAU 2	NIVEAU 3	NIVEAU 4	NIVEAU 5
VISUEL (naturels)	SIMPLE VISUAL	SIMPLE SCALABLE	CORE VISUAL	MAIN VISUAL	N-BIT VISUAL
VISUEL (synthétiques Hybrides)	SIMPLE FACIAL ANIMATION VISUAL	SCALABLE TEXTURE VISUAL	BASIC ANIMATED 2D TEXTURE VISUAL	HYBRID VISUAL	
AUDIO	SPEECH	SYNTHESIS	SCALABLE	MAIN	
GRAPHIQUE	SIMPLE 2D GRAPHICS	COMPLETE 2D GRAPHICS	COMPLETE GRAPHICS		
DESCRIPTEUR DE SCENE	AUDIO	SIMPLE 2D	COMPLETE 2D	COMPLETE	
DESCRIPTEUR D'OBJET	MAIN				

Conclusion

- La compression vise à réduire le poids d'un fichier numérique et permet ainsi une occupation mémoire limitée et une vitesse de transmission plus rapide.

- Les schémas de compression reposent sur des outils mathématiques avancés. Parmi eux : La transformée en cosinus discrète, la transformée en ondelettes, le codage entropique, l'estimation de mouvements, la transformée de Fourier,…

- La compression se déroule souvent en plusieurs étapes, avec ou sans pertes, qui exploitent les redondances présentes dans les fichiers.

- Pour chaque format de fichier, une norme décrit précisément ces étapes et leurs paramètres de réglage, en fonction des applications visées (**profil**) et le niveau de qualité voulu (**niveau**).