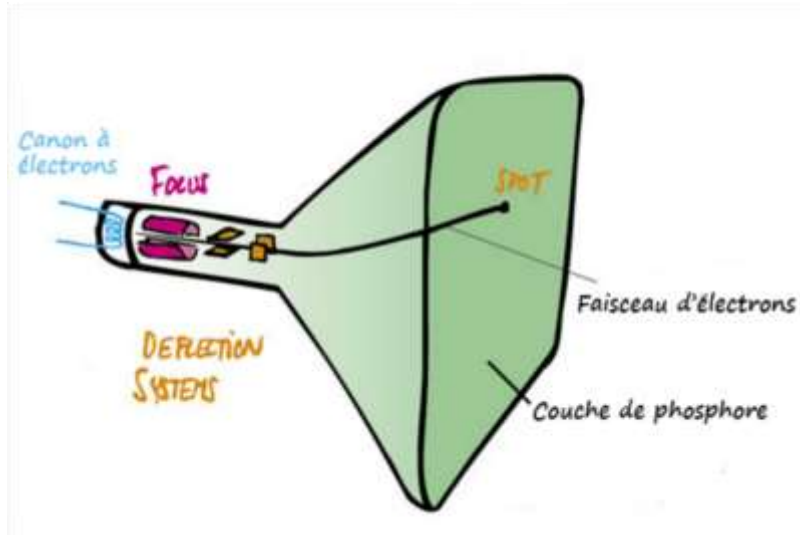


Le signal vidéo analogique en N/B

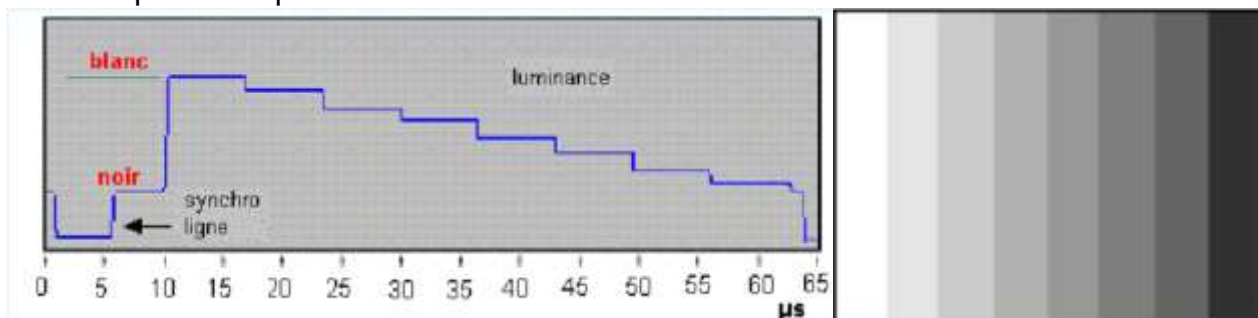
Le but de la télévision est de transmettre des images animées à distance. La première voie qui fut exploitée a été de copier le cinéma, donc de transmettre simultanément tout les points composant l'image. Chaque point lumineux était convertit en signal électrique et transmis. Ce système nécessitait donc autant d'émetteurs que de points dans l'image, l'idée fut rapidement abandonnée.

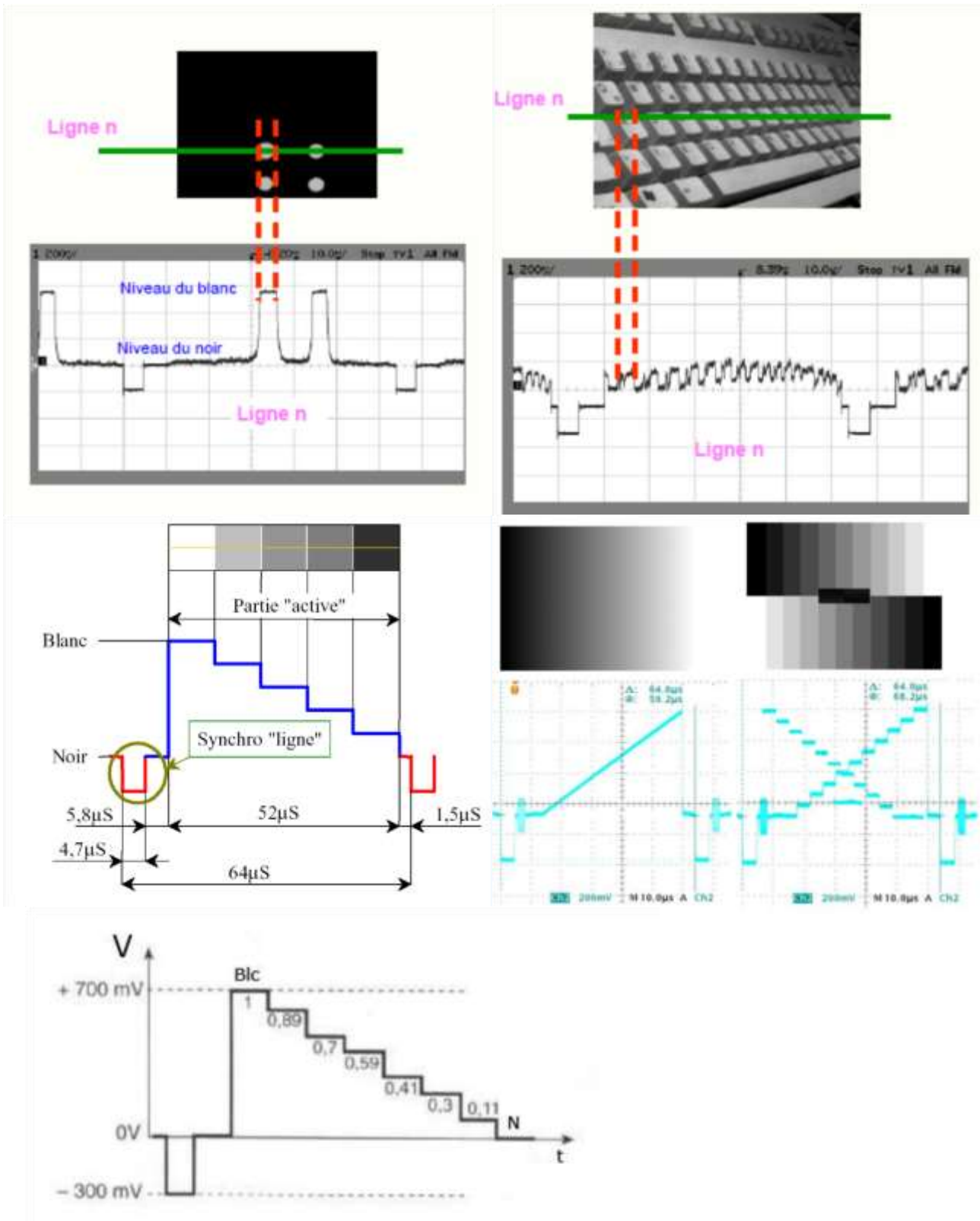
La télévision devait donc être un système séquentiel, l'image est décomposée à l'émission et recomposée à la réception.



La transmission des images s'effectue en convertissant à l'émission une image optique en signaux électriques proportionnels à la brillance de chacun des points constituant l'image et en les transmettant les uns après les autres sur un canal unique.

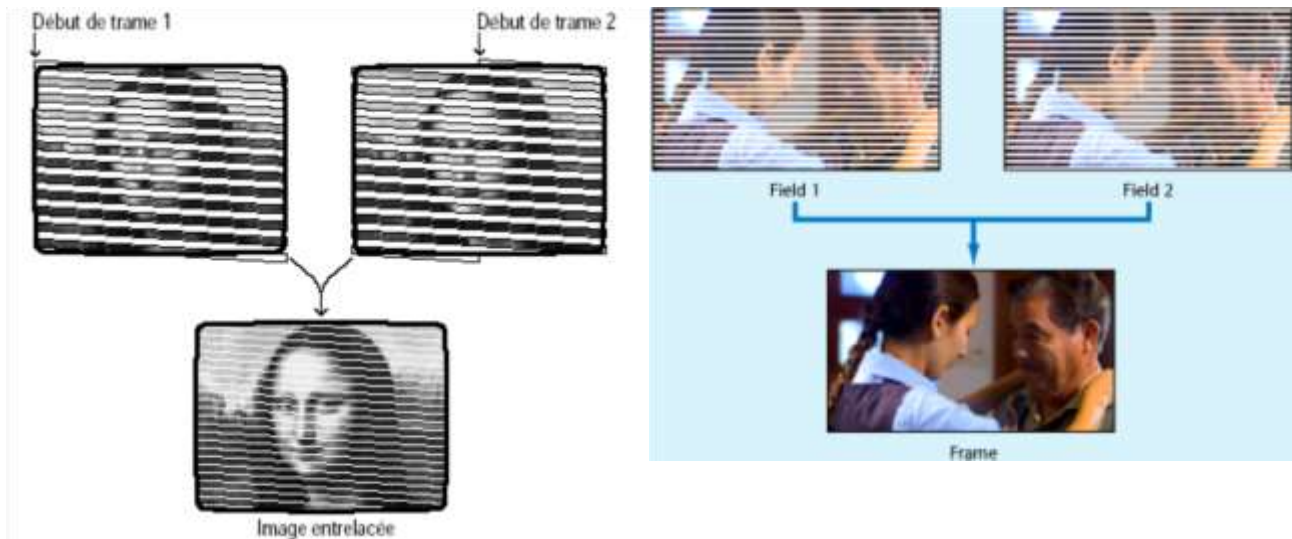
A la réception on opère la conversion inverse





La durée d'une image est fonction de la persistance rétinienne de l'œil humain (1/15 s) et de la fréquence du secteur (50Hz ou 60Hz), l'analyse complète d'une image dure 1/25 s en Europe et 1/30 s aux USA, valeur moitié de la fréquence du secteur, supérieure à la persistance rétinienne et proche de la cadence des images cinéma (24 i/s).

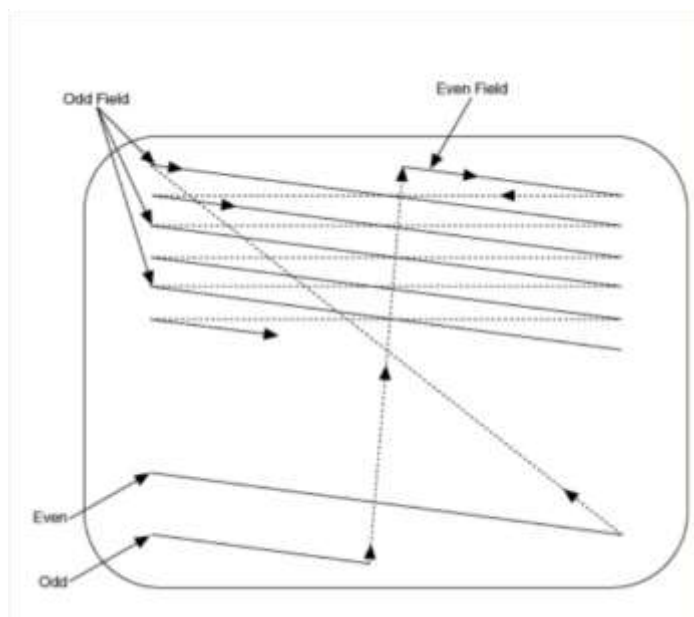
Pour éviter le scintillement de l'image, on analyse d'abord les lignes impaires et ensuite les lignes paires, c'est le balayage entrelacé.



L'intérêt de cette méthode est d'éviter un effet de papillonnement de l'image. En effet, lorsque le faisceau électronique balaie l'écran, les phosphores s'éclairent, puis l'intensité de la lumière décroît. Donc, au moment où le bas de l'image est formé (pleine luminosité), l'intensité du haut de l'image a légèrement baissé. Or, la fréquence de rafraîchissement de la télévision analogique est de 50 Hz (Pal et Secam) ou de 60 Hz (NTSC), en raison de la fréquence du courant domestique (50 Hz en Europe, 60 Hz aux États-Unis)[]; à cette fréquence, on peut voir une variation de contraste malgré la persistance rétinienne.

Si maintenant on effectue un balayage entrelacé, alors on balaie l'écran une ligne sur deux, et la variation de contraste a lieu entre deux lignes voisines et non plus entre le haut et le bas de l'écran. Les lignes étant très petites et resserrées, cela ne se voit pas, et l'on a un meilleur confort visuel.

Une image est décomposée en deux trames consécutives, une trame paire et une trame impaire. Une trame est une image de demi résolution verticale composée de 287,5 lignes[]. (sur les 625 lignes théoriques d'une image, il faut compter l'équivalent de 50 lignes pour le retour du spot en haut de l'écran, soit 575 lignes, donc 287,5 lignes par trame ; une trame commence ou se termine par une demie ligne). Une trame contient une ligne sur deux de l'image.



Le nombre de points par ligne détermine la définition horizontale, le nombre de lignes d'une image donne la définition verticale.

Pour effectuer un balayage entrelacé, le nombre de lignes par image doit être impair. En Europe on a adopté le système à 625 lignes et aux USA le 525.

En 625, l'oscillateur pilote fonctionne sur 31250 Hz, une division par deux donne la fréquence ligne 15625 Hz (625×25) et quatre divisions par 5 donne la fréquence de trame, 50 Hz.

En 625 lignes, définition verticale est donc de 625 points, la définition horizontale est de : $625 \times 4 / 3 = 830$ points

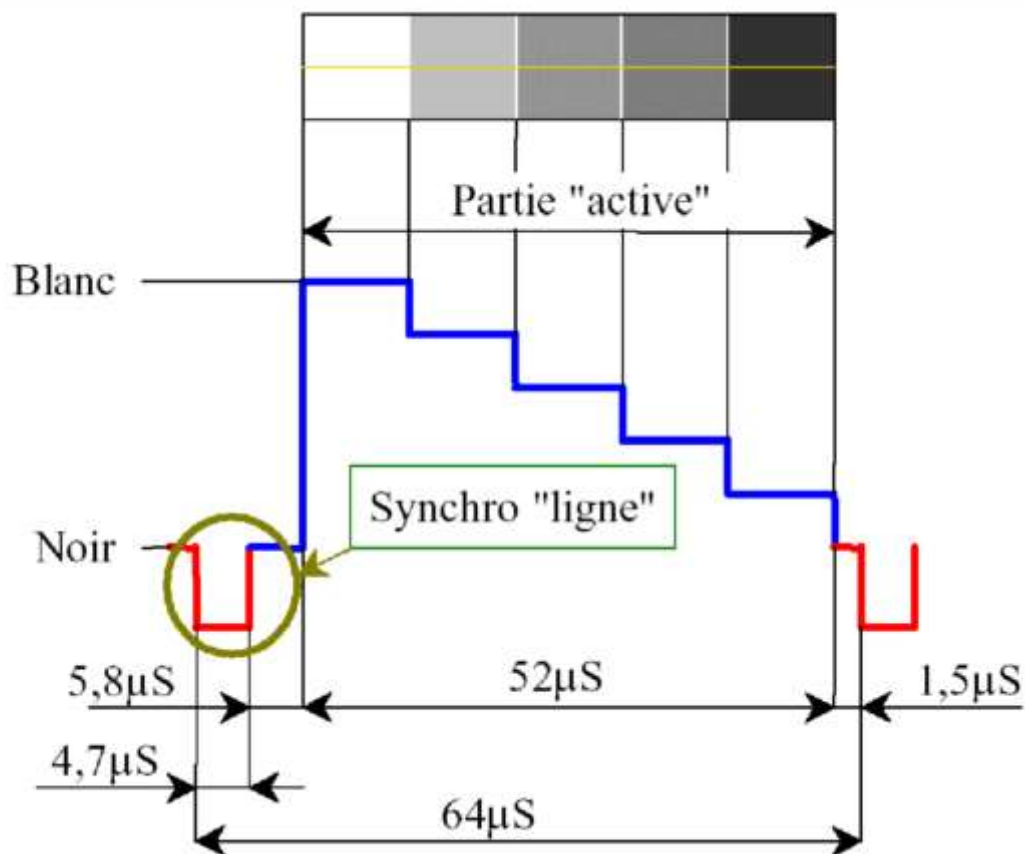
Si on analyse une image ayant successivement des points noirs et blancs, ils sont traduits par un signal électrique (vidéo) carré de fréquence : $625 \times 25 \times 830 / 2 = 6,5$ MHz

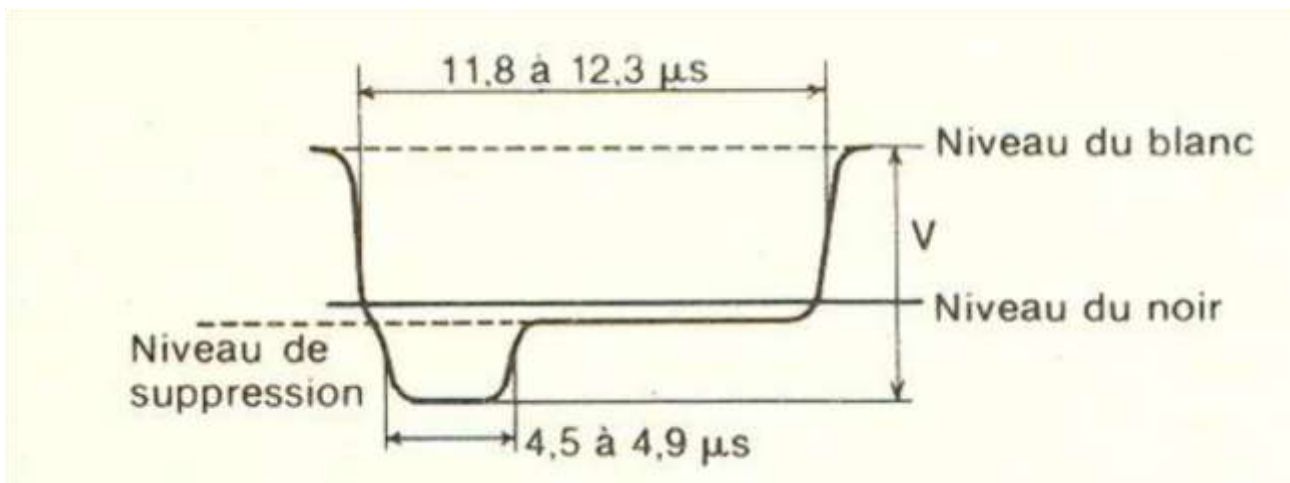
Ce cas étant limite, on adopte en pratique une bande passante légèrement inférieure de 5,5MHz.

En réduisant la bande passante vidéo, on diminue la définition horizontale sans agir sur la définition verticale imposée par le balayage, le point devient rectangulaire, et on perd en finesse.

A la fin de chaque ligne l'émetteur envoie un signal bref destiné déclencher l'oscillateur de relaxation du récepteur, le spot d'analyse du tube cathodique doit revenir très vite au bord gauche de l'écran pour balayer la ligne suivante.

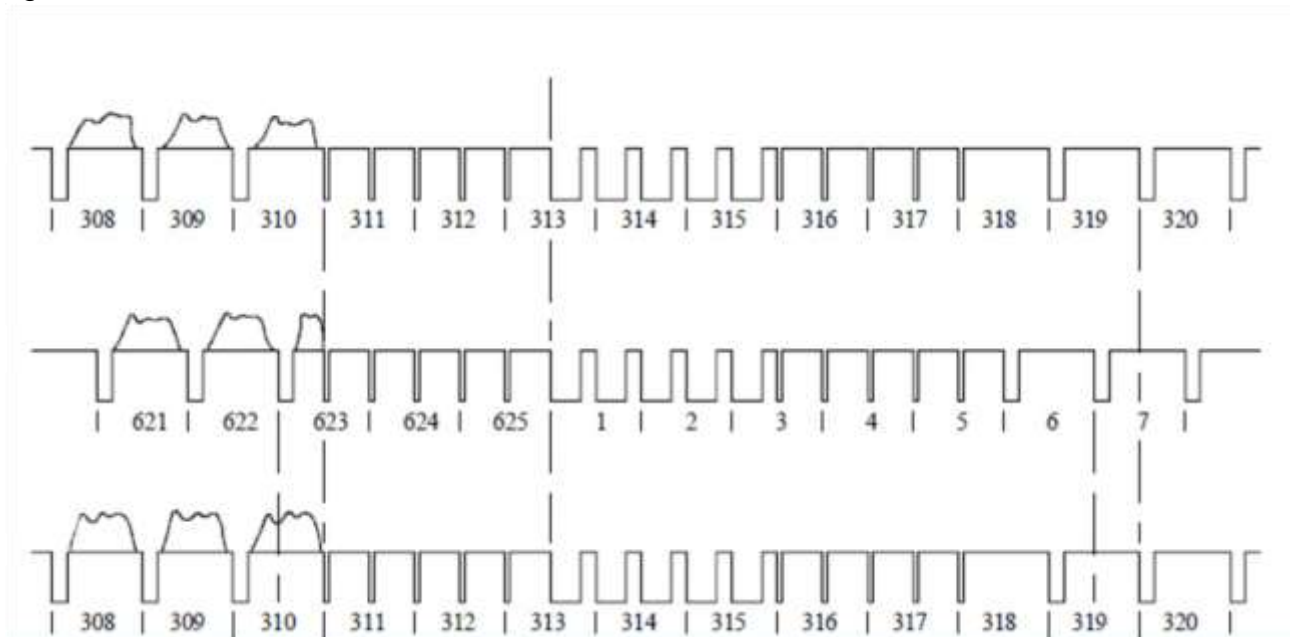
Ce signal bref s'appelle la synchro ligne.





A la fin de chaque trame, l'émetteur envoie un signal plus long destiné à faire revenir très vite le spot du bas droit de l'écran vers le haut gauche pour balayer la trame suivante. C'est la synchro trame.

La synchro ligne ne pouvant pas disparaître, la synchro trame doit contenir la synchro ligne.

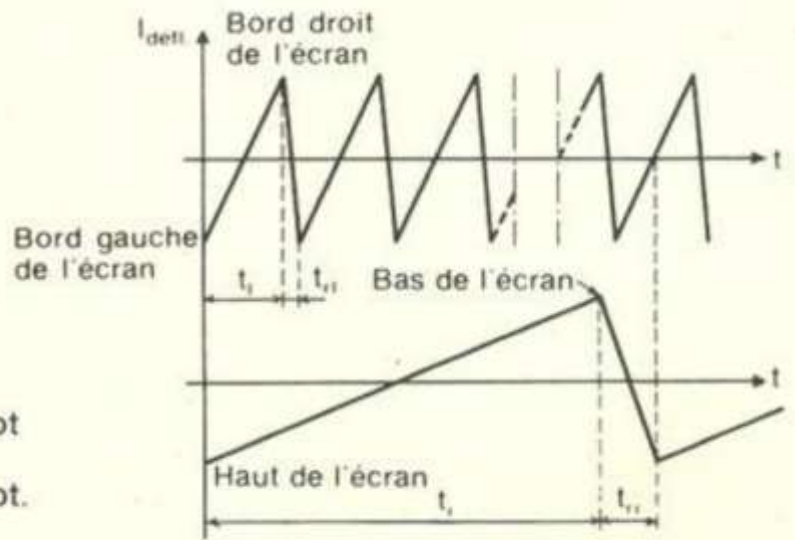


Les deux oscillateurs du récepteur, synchronisés par les « tops » de l'émetteur, fournissent les signaux en dents de scie destinés à assurer les déflexions verticales et horizontales. La synchro trame est composée de 5 tops synchro ligne inversés à double fréquence, précédés de 5 tops de pré égalisation et suivis de cinq tops de post égalisation, viennent ensuite 23 lignes d'effacement.

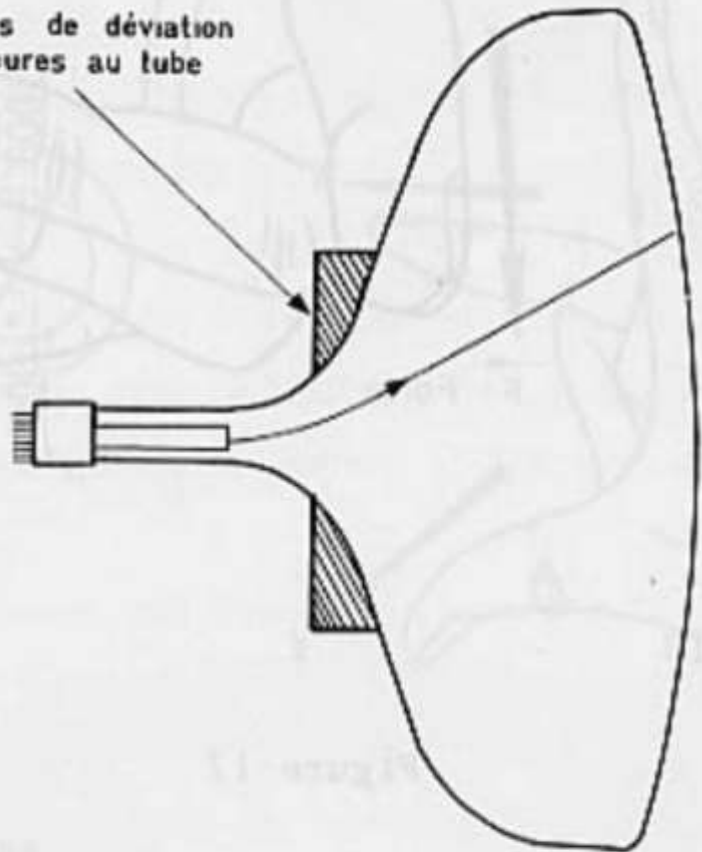
Ces lignes d'effacement, nécessaires à la remontée du spot, sont utilisées pour transporter les signaux télétexte, les anti-copie VHS, les signaux destinés au décodages de canal + (Be TV) analogique, etc...

Les deux oscillateurs du récepteur, synchronisés par les "tops" de l'émetteur, fournissent les signaux en dents de scies destinés à assurer les défections horizontales et verticales.

- t_l durée d'une ligne
- t_{rl} temps de retour du spot
- t_i durée d'une trame
- t_{ri} temps de retour du spot.



**Bobines de déviation
extérieures au tube**



DEFLEXION ELECTROMAGNETIQUE

Figure 16

Le signal vidéo analogique couleur

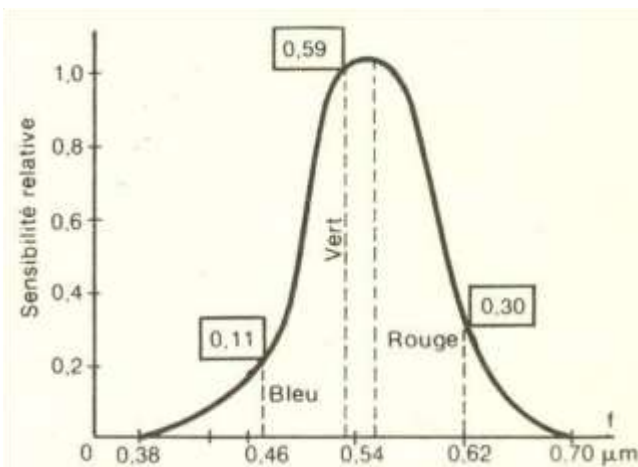
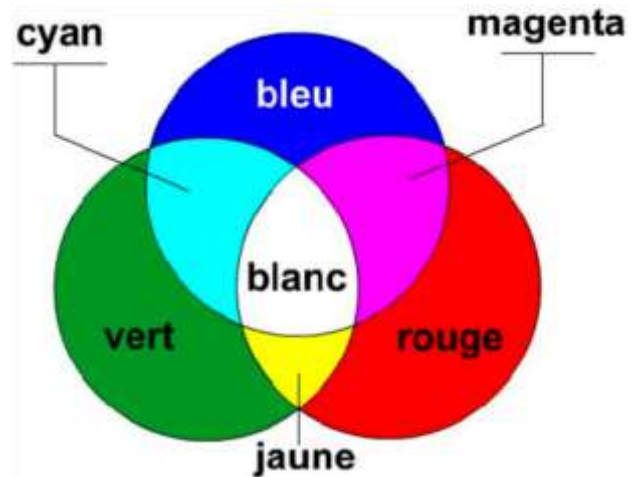
En télévision, on utilise la trichromie additive. Si on illumine un écran blanc au moyen de trois projecteurs respectivement rouge, vert et bleu on obtient :

- au centre une tache blanche formée par l'addition des trois couleurs primaires - à la périphérie, trois taches celles des primaires
- au stade intermédiaire, trois taches obtenues par les mélanges suivants :

Rouge + vert = jaune

Bleu + rouge = magenta

Vert + bleu = cyan



La Commission Internationale de l'éclairage a fixé les trois primaires :

rouge : 700 nm

vert : 546,1 nm

bleu : 435,8 nm

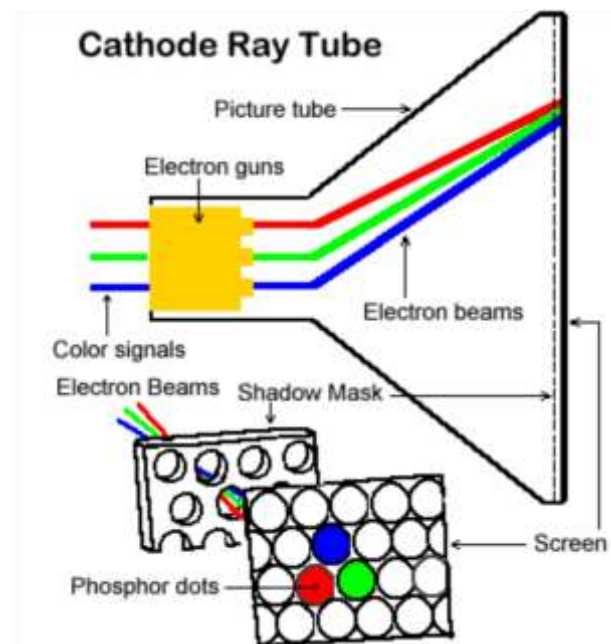
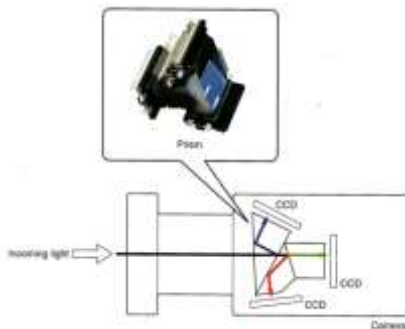
Malheureusement, l'oeil humain n'est pas sensible de la même façon à ces trois couleurs.

Cette courbe a été établie par la CIE en 1924 pour l'observateur moyen elle a permis d'établir la relation fondamentale en télévision :

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

Y est appelé la LUMINANCE, c'est en fait le signal noir et blanc

Le système doit être compatible, les émissions en couleur doivent être reçues en noir et blanc sur les téléviseurs noir et blanc. Le système doit être rétro compatible, les émissions en noir et blanc doivent être reçues en noir et blanc sur les téléviseurs couleur.



On constate qu'une image très fine en N/B peut être coloriée avec des touches colorées assez grossières, donc, les informations de chrominance peuvent être transmises avec une bande passante relativement étroite.

Du fait des compatibilités, les informations de chrominance doivent s'annuler lors de l'analyse d'une image N/B.

Les combinaisons R-Y, B-Y, V-Y satisfont à cette condition, en effet pour une image N/B, on a : $R=V=B$, compris entre 0 (noir) et 1 (Blanc).

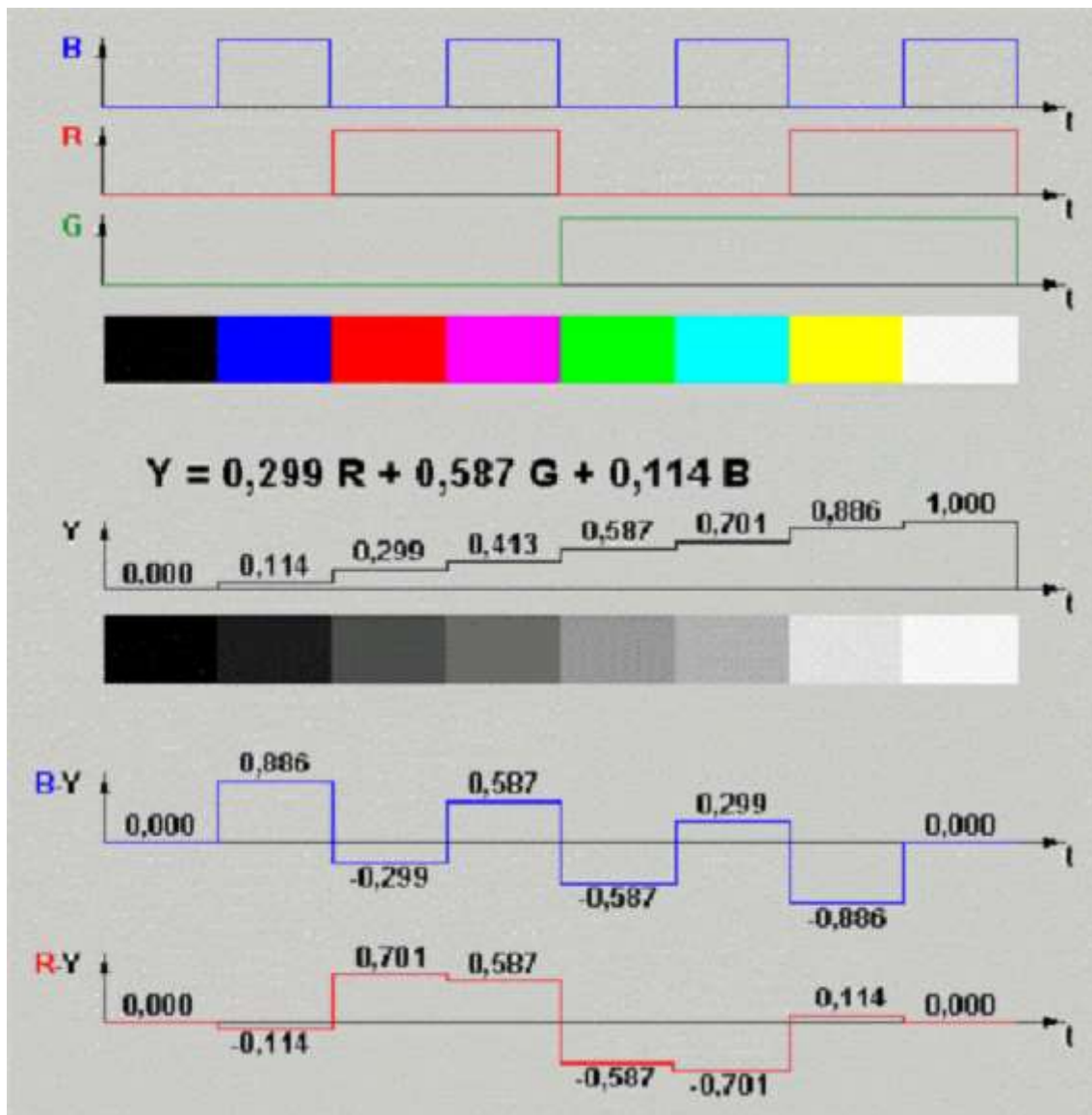
Pour une image blanche, on a : $R=V=B=1$ et $Y=1$, donc : $R-Y=0$, $B-Y=0$, $V-Y=0$

Même raisonnement pour les noirs et gris.

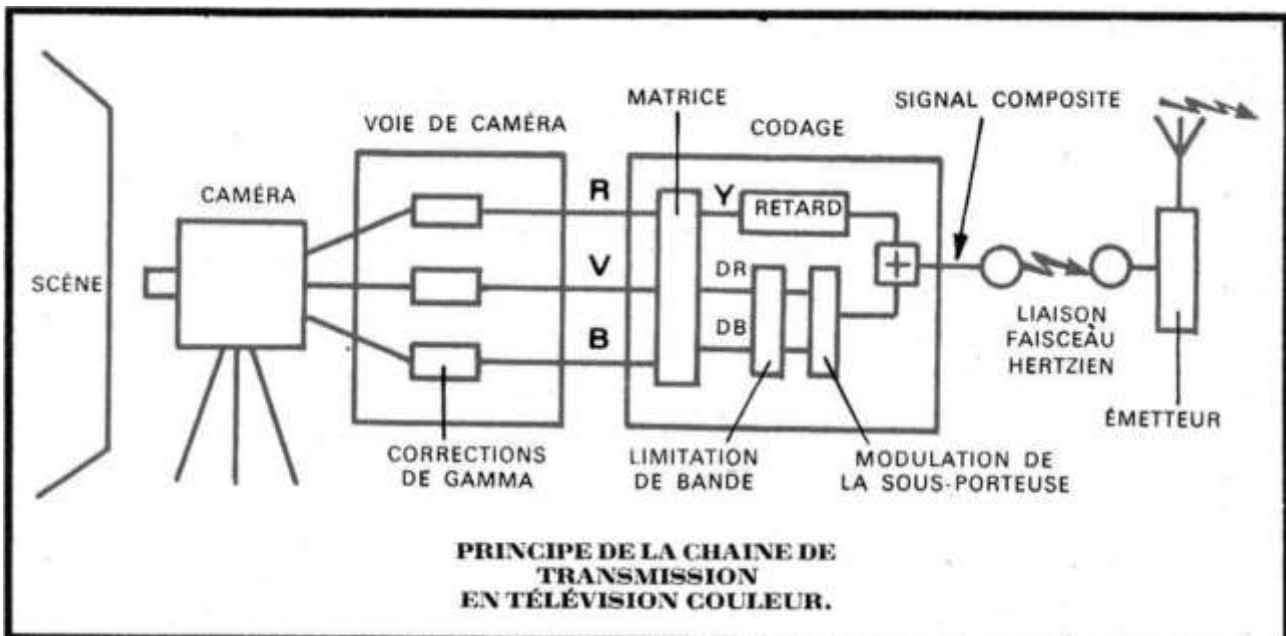
La difficulté consiste à combiner la luminance (Y) et la chrominance (R-Y , B-Y, V-Y) dans le canal vidéo large (en principe) de 6,5 MHz.

Il n'est pas nécessaire de transmettre les trois signaux de chrominance si on connaît leur somme et deux différences.

On choisit R-Y et B-Y car leurs différences sont les plus importantes et donnent donc les informations maximales.

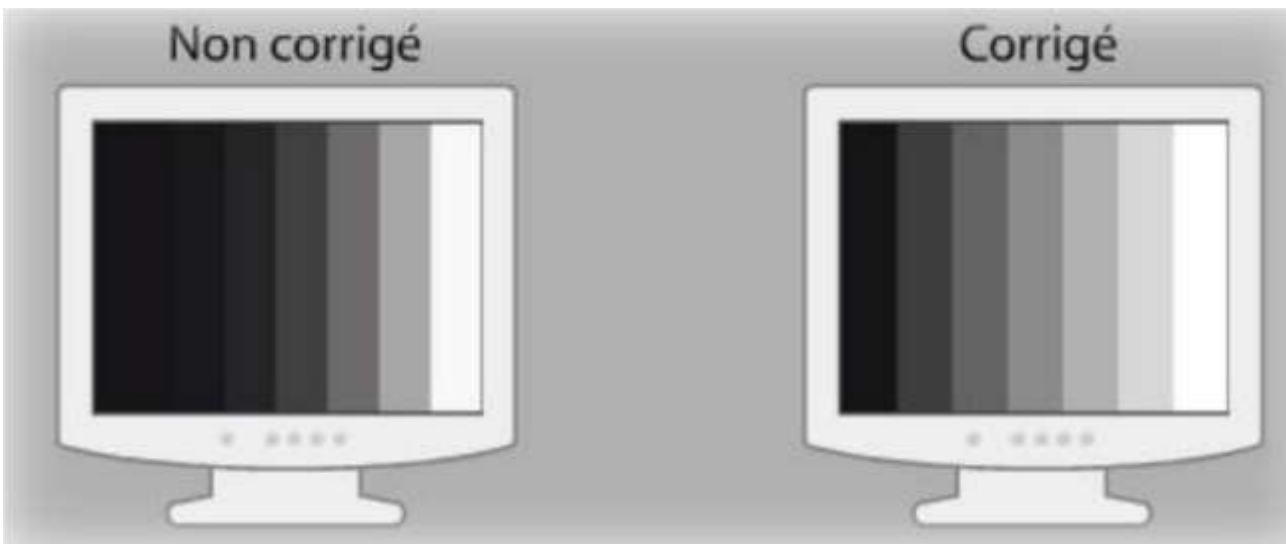


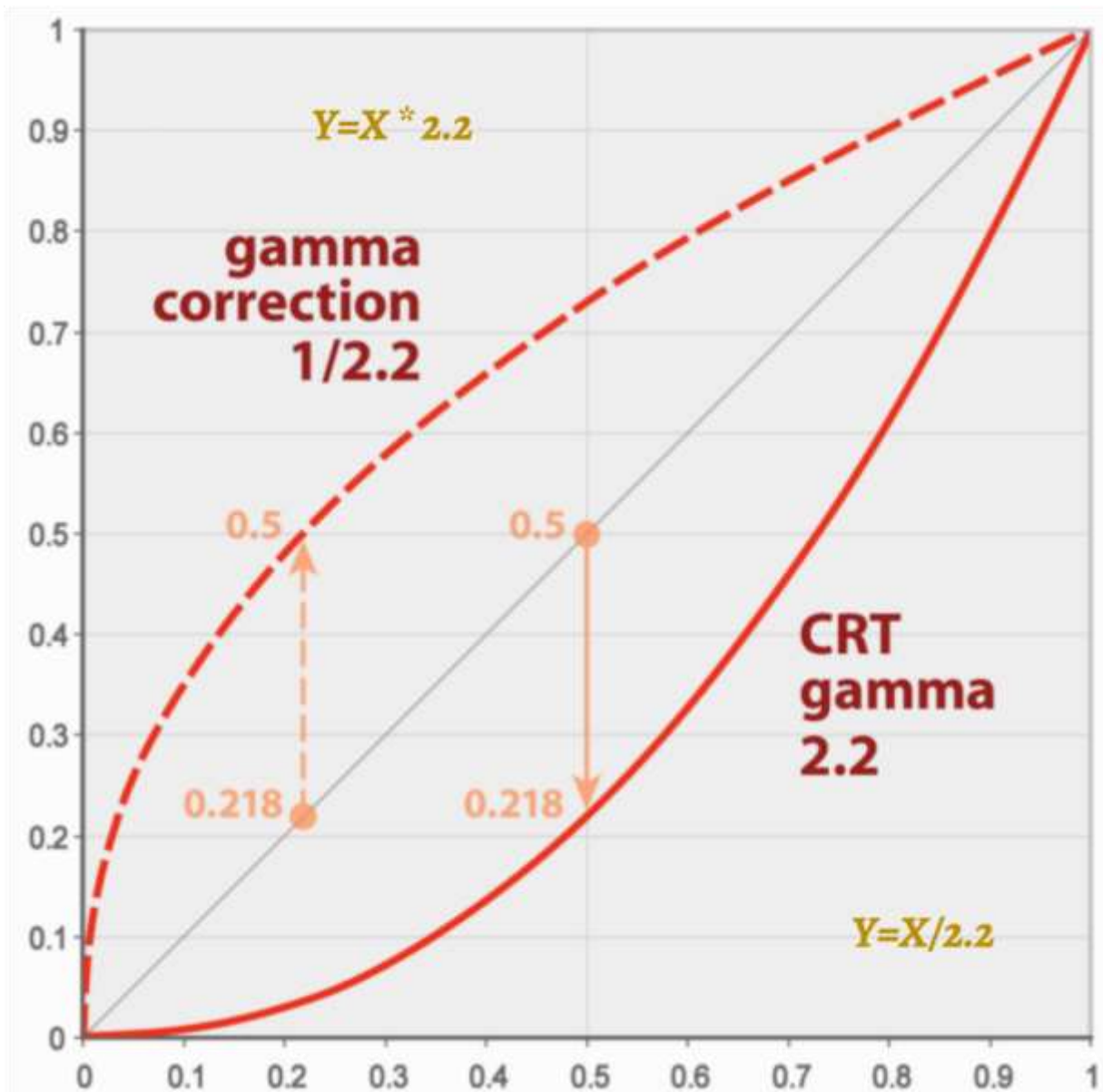
La difficulté réside dans le fait de placer deux signaux de chrominance sur une seule sousporteuse.



Le problème du gamma des moniteurs s'est posé dès l'apparition de la télévision dans la première moitié du XXe siècle. Les ingénieurs vidéo de l'époque furent confrontés à une restitution très sombre et contrastées des images sur les postes de télévision. La cause de ce phénomène a été très vite cernée. La luminosité émise par les luminophores d'un tube cathodique n'est pas proportionnelle à la tension électrique appliquée dans le tube.

Si une tension de 1 Volt donne la luminance maximum de 100, une tension moyenne de 0,5 V ne donnera pas une luminance de 50 % mais plutôt une luminance plus faible de seulement 20 % et le phénomène ira en s'amplifiant à mesure qu'on se rapproche des valeurs proches du zéro.





La correction de gamma se fait à l'émission.

R,V,B deviennent R',B',V'

Les trois couleurs R', B' ,V' passent ensuite dans un matricage, on obtient :

- R'-Y'
- B'-Y'
- Y' (0.3R'+0.59V'+0.11B')

Les deux composantes de chrominance passent par un filtre passe bas limitant leur bande passante à 1.5MHz et ensuite on leur affecte un coefficient de pondération de 0.9 pour R'-Y' et de 0.5 pour B'-Y'.

Nous obtenons enfin un signal COMPOSANT

Bande passante :Y=5MHz

R-Y=1.5MHz

B-Y=1.5MHz

La dernière étape consiste à ajouter ces composantes de chrominance (R-Y et B-Y) à la luminance (Y le noir et blanc) pour ne plus obtenir qu'un seul signal à transmettre.

Les deux informations de chrominance sont transmises simultanément par la modulation d'amplitude en quadrature de phase d'une sous-porteuse par les signaux R-Y et B-Y.

Modulation

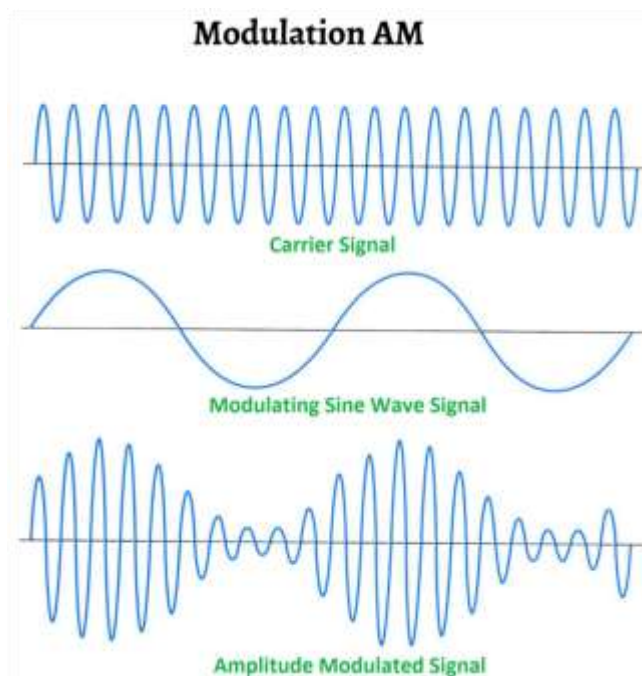
Pourquoi module-t-on un signal ?

La réception d'un signal nécessite des antennes dont les dimensions dépendent de la longueur d'onde du signal (en général de l'ordre de $l / 2$).

Un signal haute fréquence HF sera facilement transmissible [H.F correspond à des fréquences $F > 100$ MHz soit des longueurs d'onde $l = c / F$ ($c =$ vitesse de déplacement de l'onde, assimilée à la vitesse de la lumière) donc $l < 300 / 108 = 3$ m ; soit une antenne de longueur inférieure à 1,5m . Par contre , pour les signaux B.F ($f < 20$ Hz) la longueur d'onde sera beaucoup plus grande et cela nécessiterait des antennes démesurées et le signal serait rapidement atténué. Le but de la modulation est de translater le spectre d'un signal B.F [sons, musique , parole] vers les H.F pour pouvoir le transmettre facilement par voie hertzienne. La radio , la Télévision , les lignes téléphoniques utilisent le procédé de modulation . Le signal H.F est appelé PORTEUSE . Le signal B.F est appelé SIGNAL MODULATEUR .

On peut procéder de deux manières :

1. Modulation d'amplitude de la porteuse
2. Modulation de la fréquence du signal H.F



•L'onde du haut est la porteuse (carrier) celle qui va être modulée

•la seconde est le signal à transmettre (celui qui va moduler)

•la troisième est la porteuse modulée en amplitude par l'onde du haut l'intérêt ici est de pouvoir transmettre un signal qui a la même bande de fréquence que Y (donc non "mélangeable" avec lui) sur une fréquence unique (la fréquence de

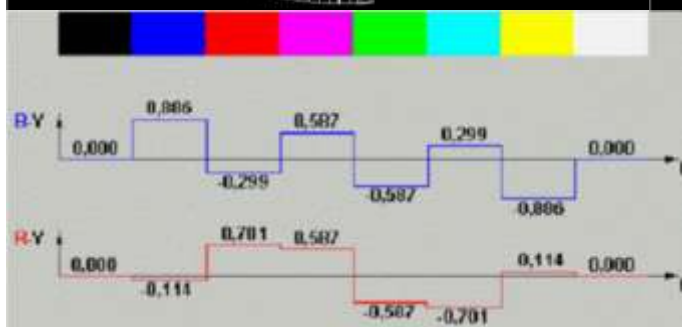
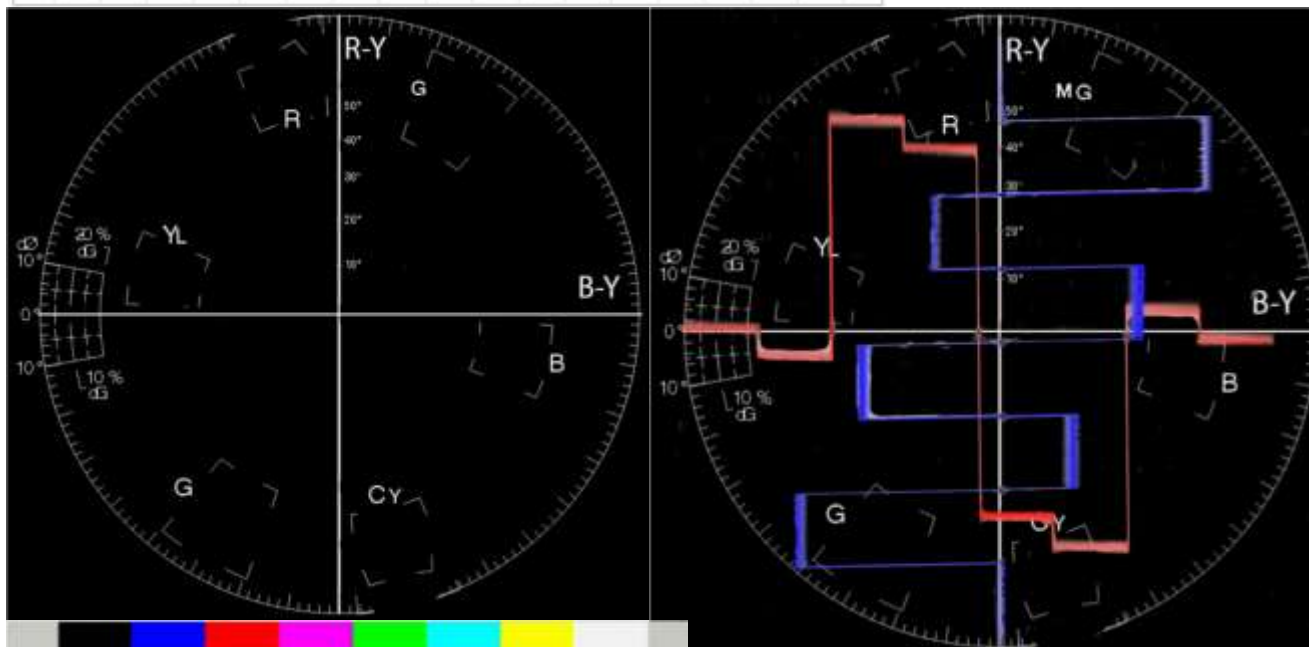
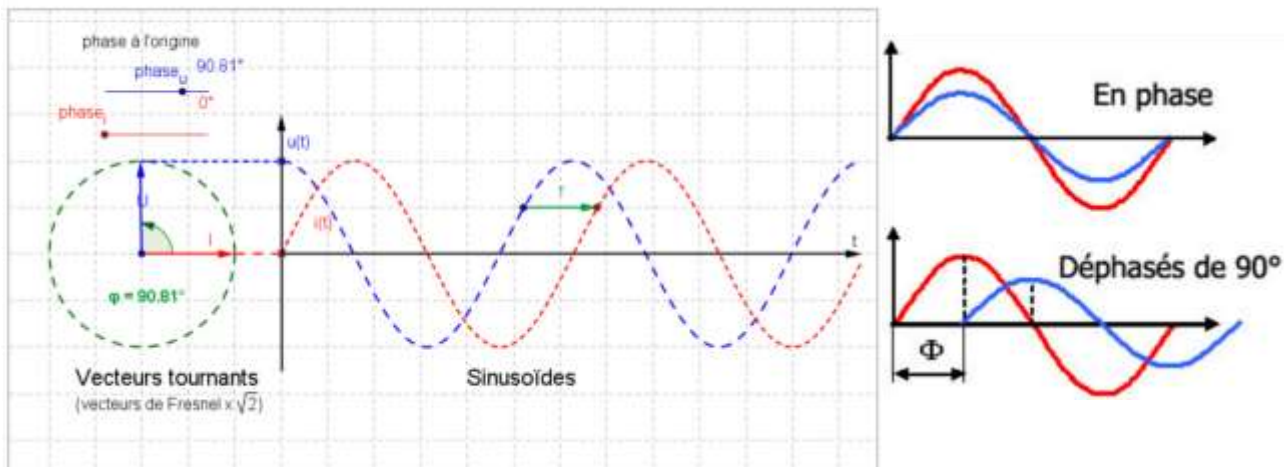
la porteuse)

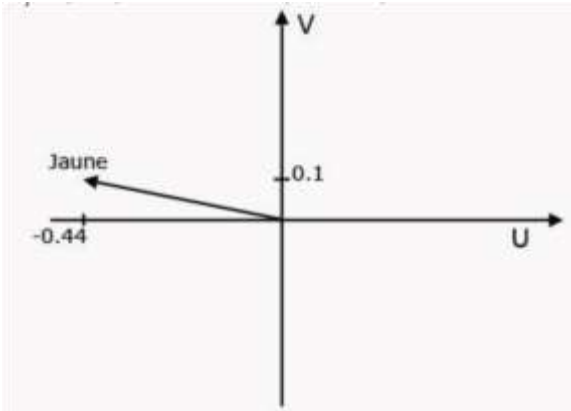
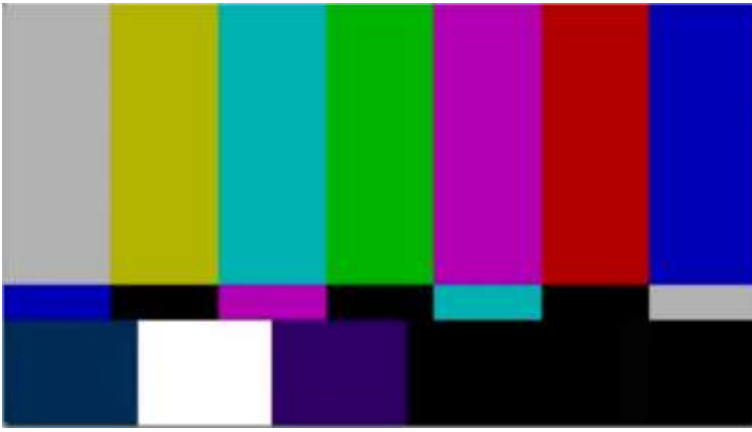
Pour transmettre les deux signaux de chrominance, on utilise deux ondes secondaires (dites sous-porteuse par opposition à l'onde porteuse qui véhicule le signal TV) de fréquence 4.43 MHz en PAL mais décalées (déphasées de 90°) entre elles (quadrature)

Ces sous-porteuses sont ensuite modulées en amplitude (une par R-Y et l'autre par B-Y), sans porteuse, puis additionnées l'une avec l'autre.

La modulation totale est représentée par un vecteur modulé à la fois en phase (la teinte) et en amplitude (la saturation) et variant en permanence selon les couleurs. Le signal ainsi obtenu est ensuite tout simplement ajouté au signal de luminance.

Déphasage





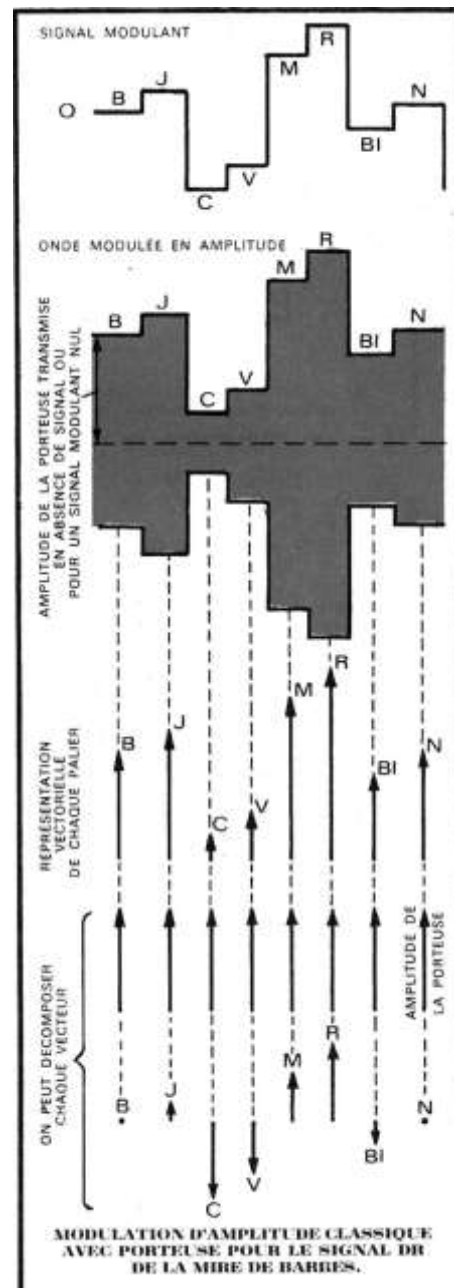
Prenons le jaune comme exemple. (le jaune de la mire « barres couleurs »)

$$Y_j = 0.89$$

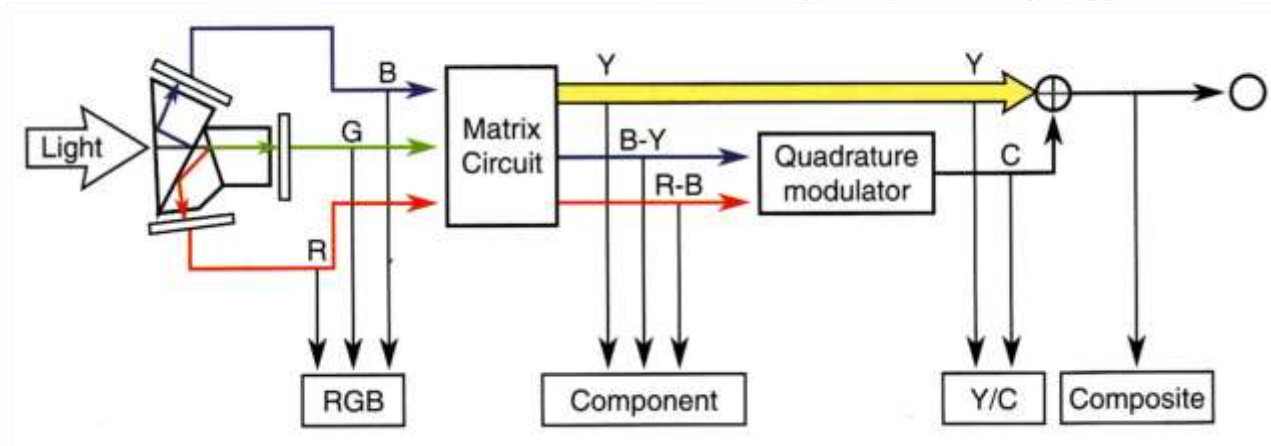
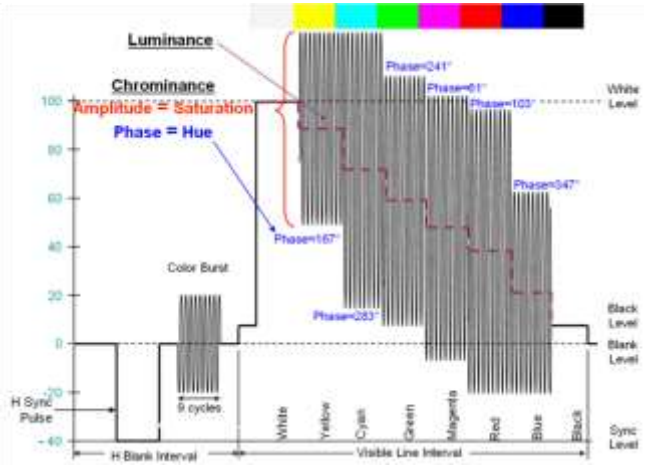
$$V_j = (R-Y) * 0.9 \text{ donc : } (1 - 0.89) * 0.9 = 0.1$$

$$U_j = (B-Y) * 0.5 \text{ donc : } (0 - 0.89) * 0.5 = - 0.44$$

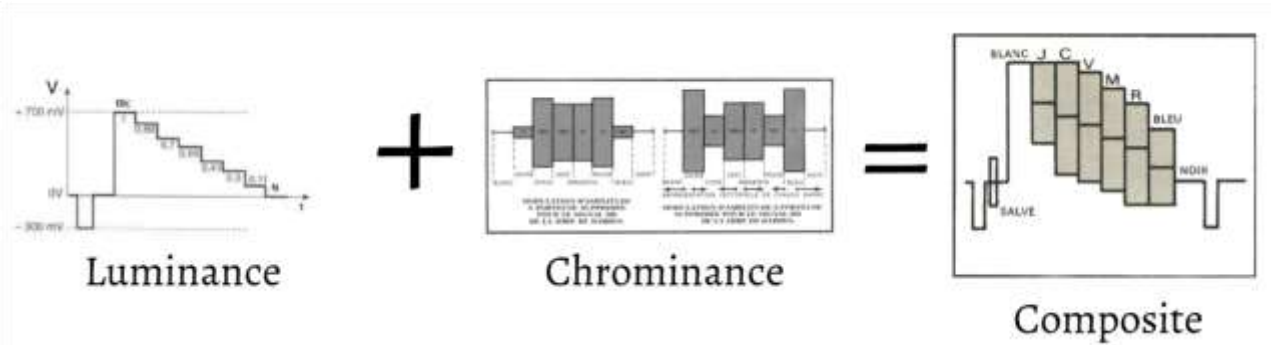
La modulation d'amplitude à porteuse supprimée. Le schéma représente une modulation d'amplitude classique avec porteuse ou on a choisi comme signal modulant le signal D_r (R-Y différence rouge) de la mire de barre. On voit d'après la forme de l'onde modulée, que lorsque le signal modulant est nul (noir ou blanc) on transmet la porteuse, pour une valeur positive du signal, elle augmente, pour une valeur négative, elle diminue.



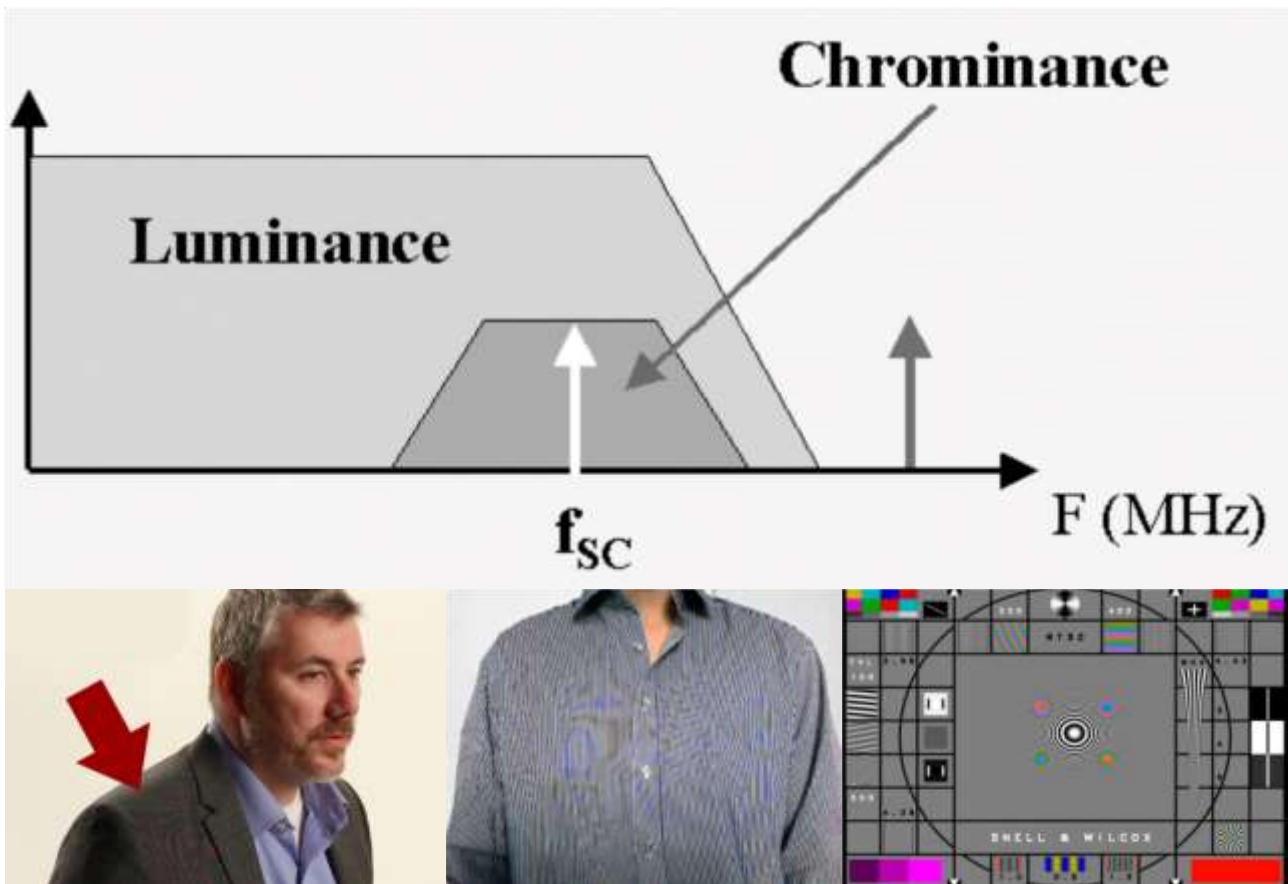
A la réception, pour démoduler une onde modulée en amplitude avec porteuse supprimée, il est nécessaire de régénérer la sous-porteuse, il faut lui donner la phase convenable et c'est dans ce but que des salves de référence (burst) sont transmises avec la phase de 180° sur les paliers arrières.



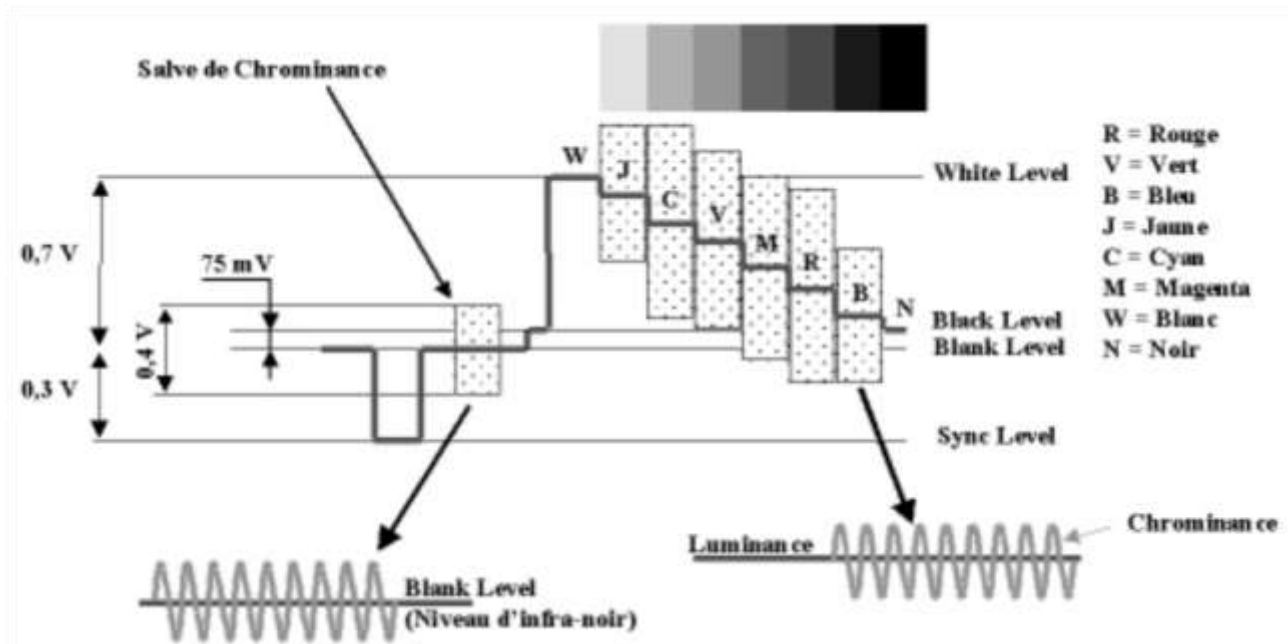
Les signaux de chrominance (les 4.43Mhz modulés par r-y et b-y) sont ajoutés au signal de luminance .

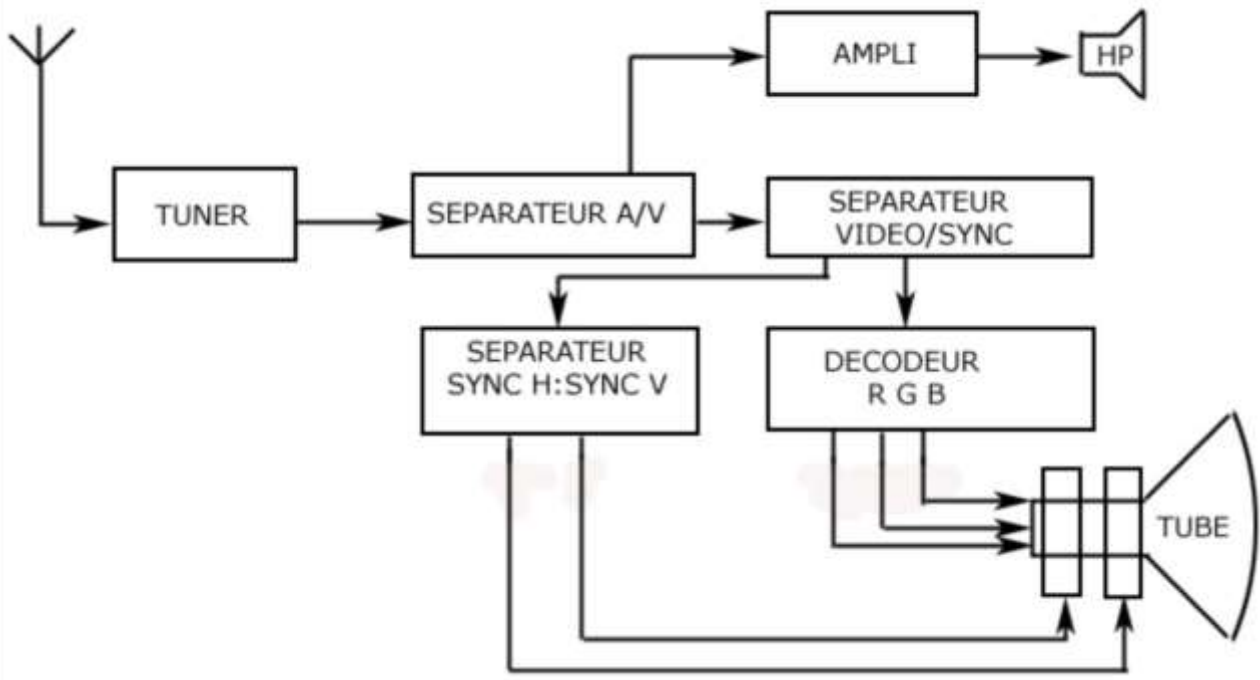


Du point de vue fréquentiel, le 4.43Mhz de la chrominance remplace le 4.43 Mhz de la luminance.



Dans un signal composite on décode la chroma en comparant le 4.43MHz contenu dans la vidéo avec l'oscillateur interne (resynchronisé à chaque ligne par le burst). La différence de phase donne la teinte l'amplitude donne la saturation.





Différents types de connecteurs de la vidéo analogique.



connecteur	utilisation	brochage
DB25	liaison PC – écran externe	5 signaux séparés toutes les composantes R,V,B, synchro ligne et synchro trame séparées
PERITEL ou SCART	équipements vidéo (TV, DVD, camera, décodeur...)	<ul style="list-style-type: none"> • signal video composite et • 2 signaux séparés (SVHS) ou 3 composantes RVB
SVHS	équipements vidéo (TV, DVD, camera, décodeur...)	2 signaux séparés : <ul style="list-style-type: none"> • luminance Y • chrominance PAL/SECAM/NTSC
3 Cinch	matériel professionnel	3 signaux séparés : <ul style="list-style-type: none"> • luminance Y • couleur R-Y • couleur B-Y
1 Cinch ou 1 Jack	équipements vidéo (TV, DVD, camera, décodeur...)	un seul signal vidéo composite <ul style="list-style-type: none"> • luminance Y + chrominance PAL/SECAM/NTSC

PAL, SECAM, NTSC

Le système PAL

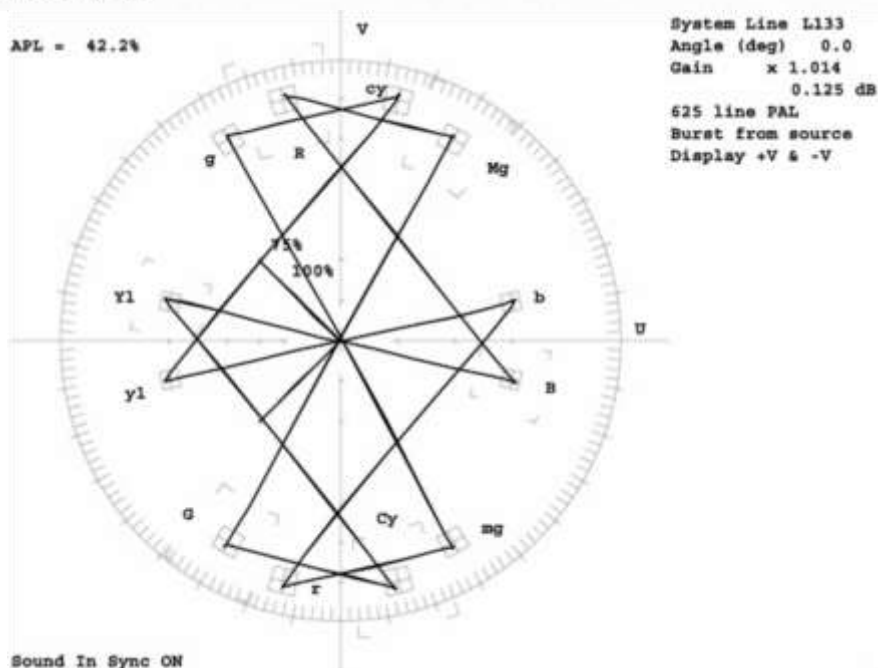
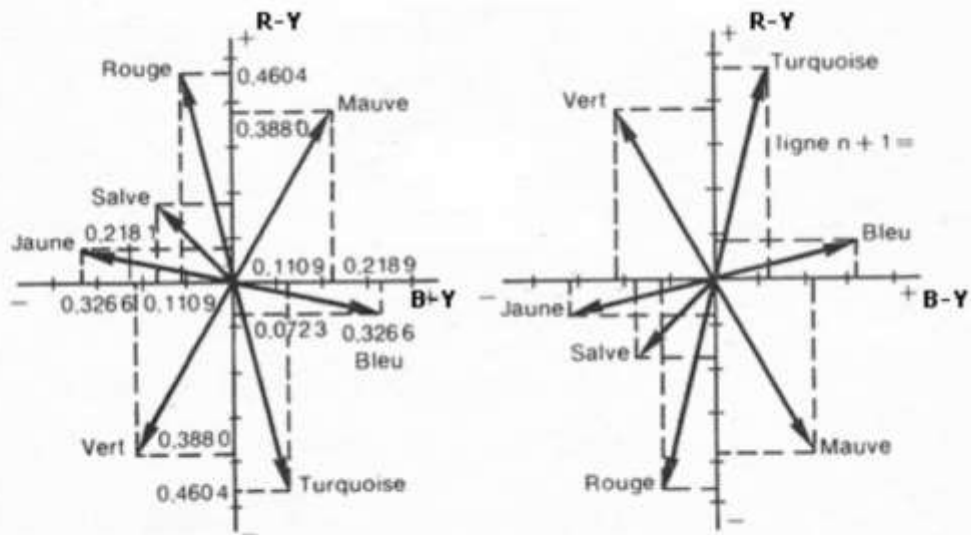
Tout ce qui a été vu précédemment est valable pour le PAL et le NTSC, seules les fréquences changent : NTSC : 30i/s , sc : 3.58MHz, PAL : 25i/s, sc : 4.43MHz.

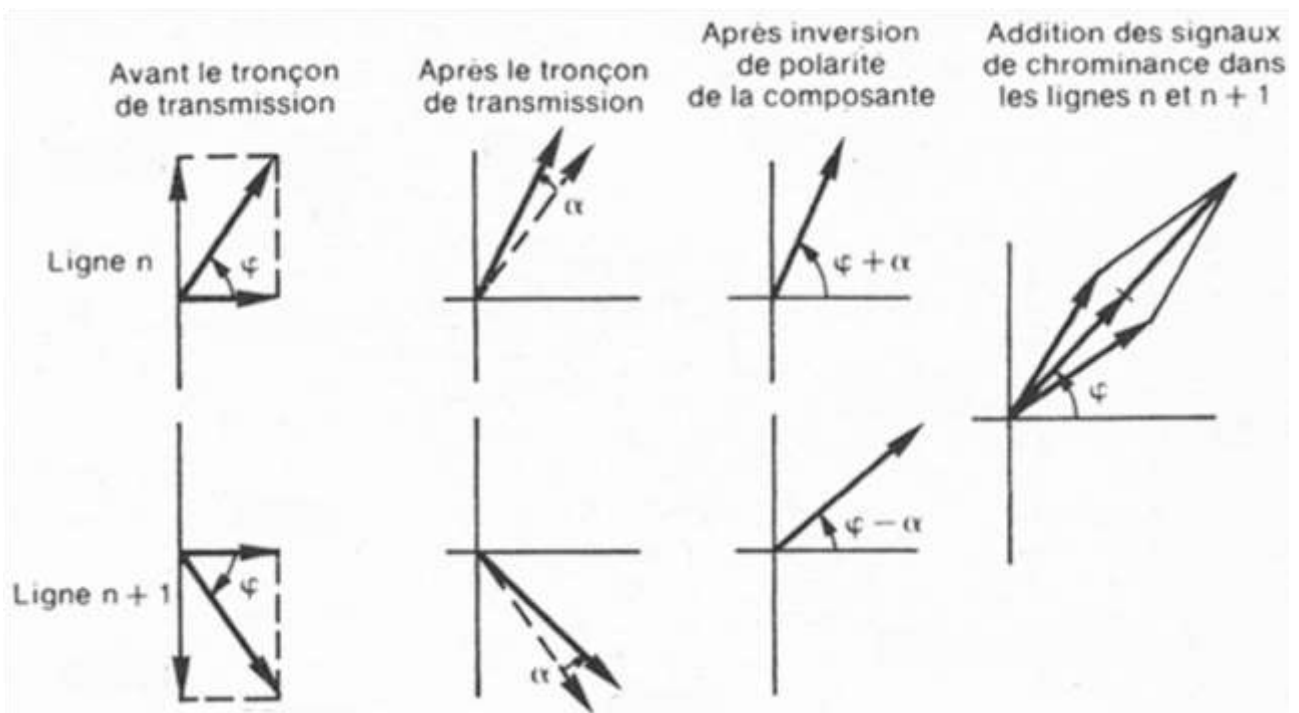
Le problème du NTSC est qu'il est très sensible à toute variation accidentelle de phase en cours de transmission, ce qui produit des modifications de teintes à la réception (d'où son surnom : Never Twice the Same Color).

Le PAL (Phase Alternated Line) remédie à cette sensibilité, en effet R-Y voit sa phase inversée de 180° à la fréquence ligne.

Le signal de chrominance reçu sur une ligne est mis en mémoire durant une ligne de façon à pouvoir être combiné avec le signal de la ligne suivante.

Si la transmission de ces deux composantes est affectée d'un déphasage, leur combinaison électronique permet d'éliminer le défaut.



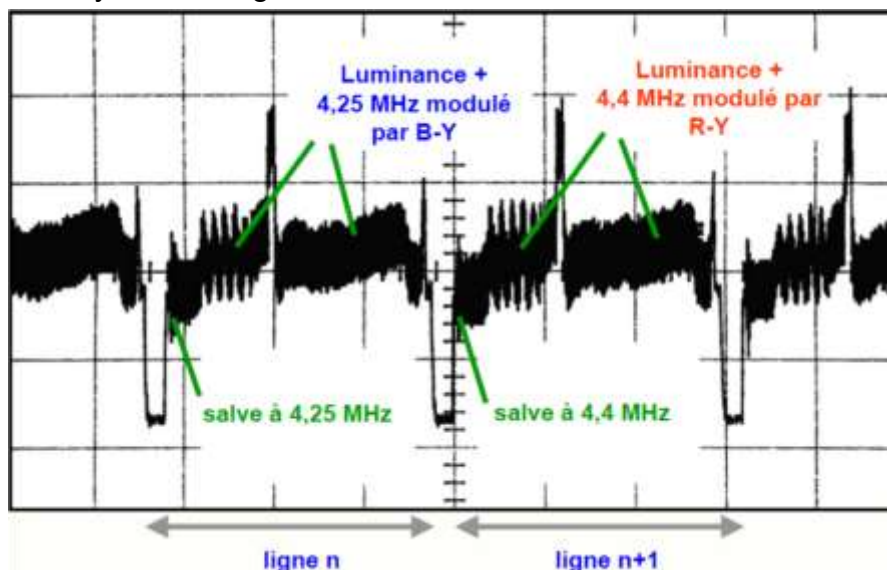


Le SECAM

Le SECAM (pour SEquence de Couleurs Avec Mémoire) est basé sur le même principe que le NTSC. Il en diffère cependant par le fait qu'il utilise une ligne à retard de $64 \mu\text{s}$ pour ne transmettre, à chaque instant, qu'un seul des deux signaux de chrominance, en alternant séquentiellement l'un et l'autre signal de ligne à ligne. Cela est obtenu en modulant, en fréquence, une onde sous-porteuse. Afin d'éviter tout risque d'erreur, une fréquence de sous-porteuse différente est attribuée à chaque couleur (4,40 MHz et 4,25 MHz).

Le SECAM a été mis au point par l'ingénieur français Henri de France:

- il procure une image en couleur d'aussi bonne qualité que le PAL
- il a été adopté pour éviter la pénétration des téléviseurs étrangers (PAL, NTSC) sur le marché français
- d'où l'appellation Système Élégant Contre les Américains !

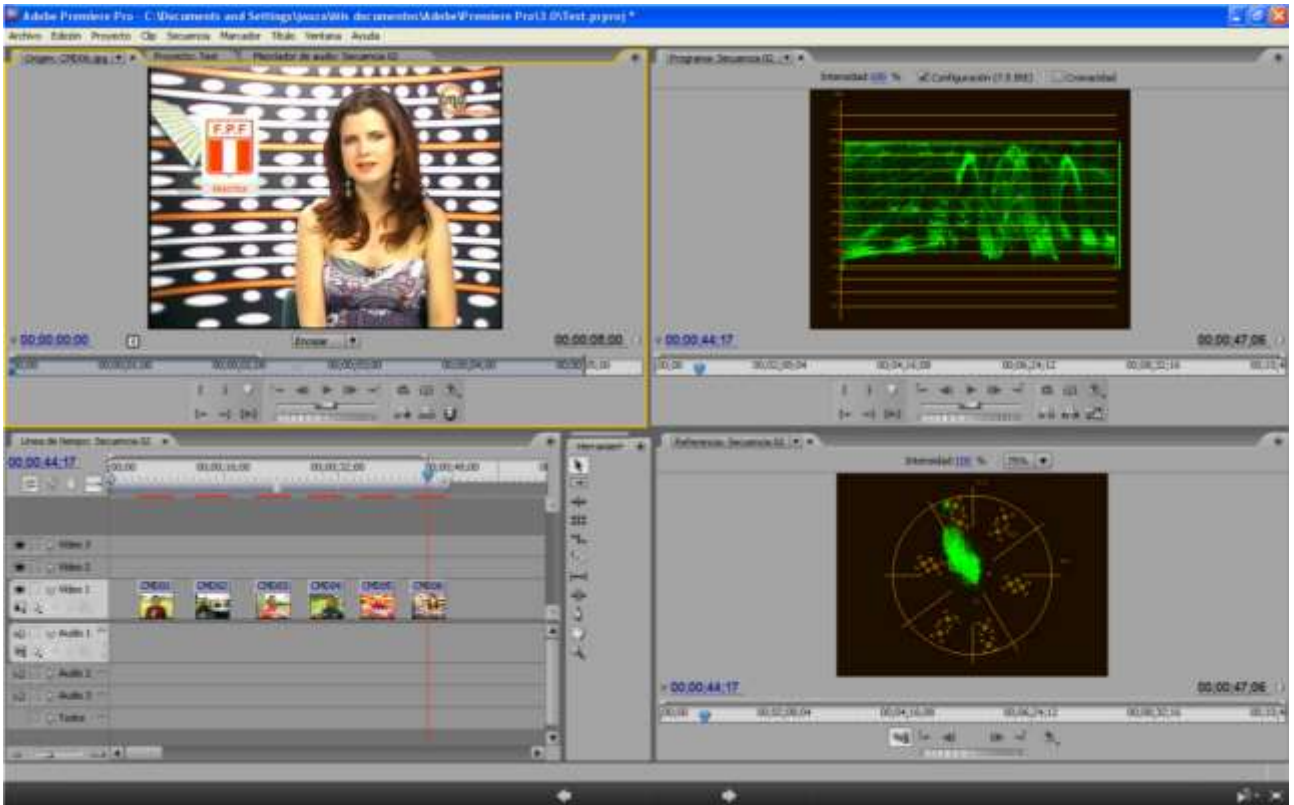


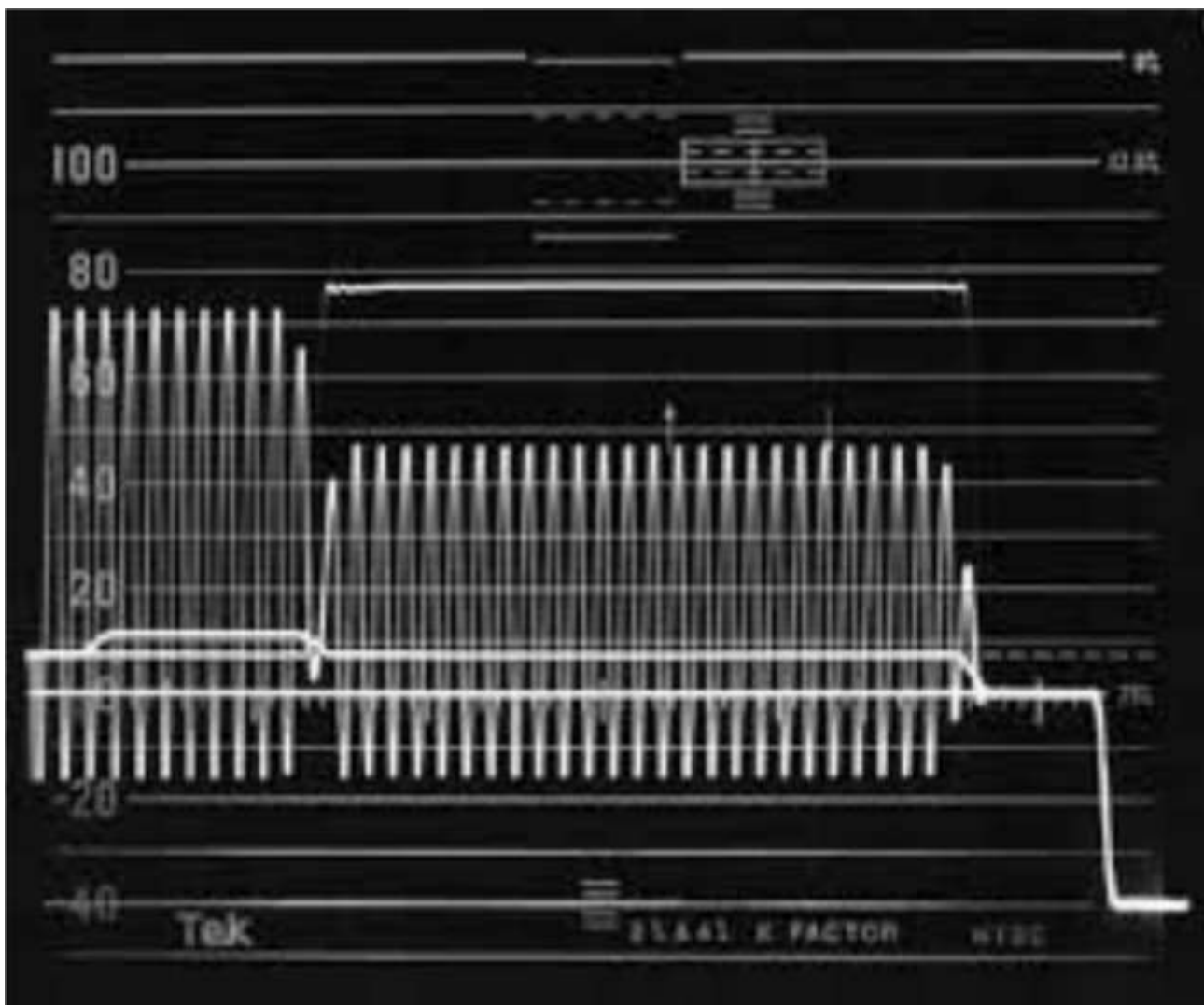
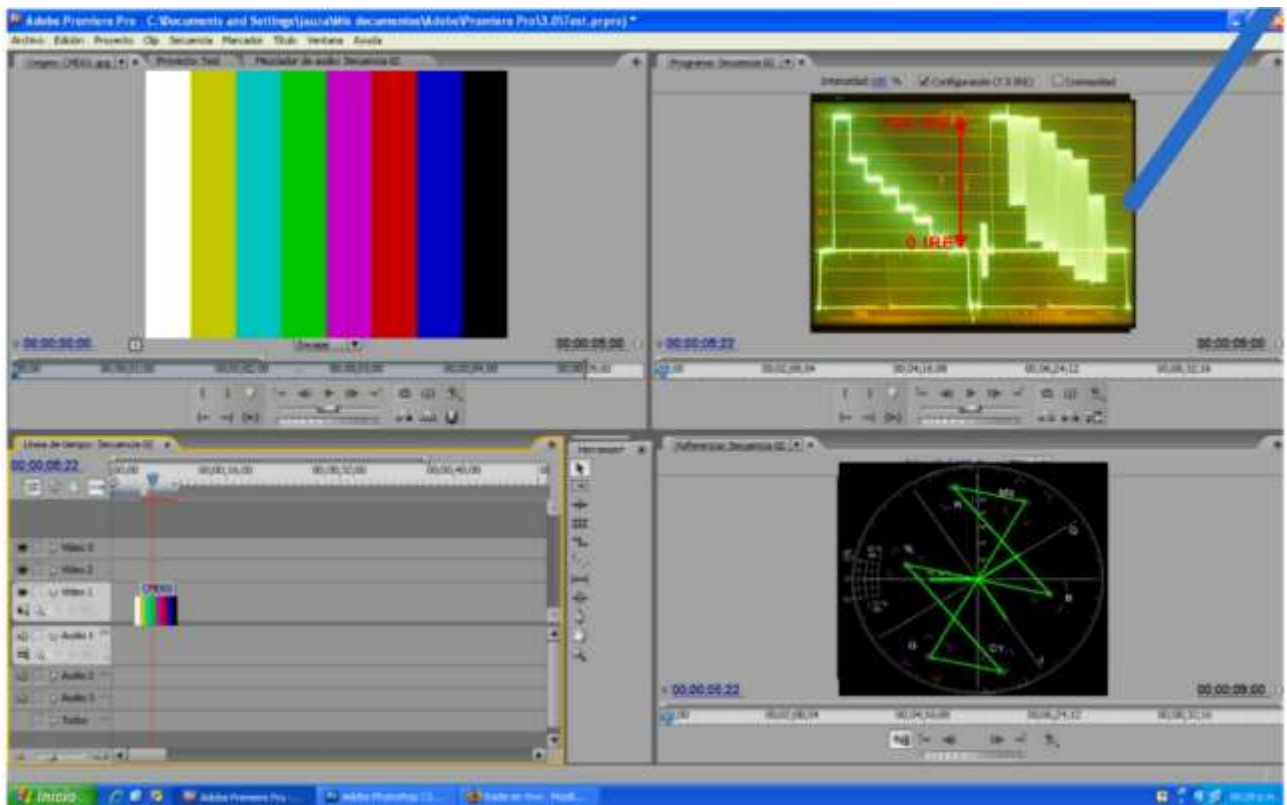
standard	fréquence des sous-porteuses couleur	modulation	remarques
NTSC	<p>B-Y : $f_c = 3,579545$ MHz</p> <p>R-Y : $f_c = 3,579545$ MHz (déphasée de $+90^\circ$)</p>	AM sans porteuse	<ul style="list-style-type: none"> - on transmet simultanément R-Y et B-Y - les deux sous-porteuses sont de même fréquence mais déphasées - des déphasages parasites durant la transmission faussent souvent les couleurs à l'arrivée - d'où l'appellation humoristique Never The Same Color !
PAL	<p>B-Y : $f_c = 4,433619$ MHz</p> <p>R-Y : $f_c = 4,433619$ MHz (déphasée de $\pm 90^\circ$)</p>	AM sans porteuse	<ul style="list-style-type: none"> - le PAL est une amélioration du NTSC (sous-porteuse de R-Y déphasée de $+ ou - 90^\circ$ une ligne sur deux) - on transmet simultanément R-Y et B-Y - meilleur rendu des couleurs
SECAM	<p>B-Y : $f_{B-Y} = 4,2500$ MHz</p> <p>R-Y : $f_{R-Y} = 4,40625$ MHz</p>	FM	<ul style="list-style-type: none"> - SECAM = Séquentiel à mémoire - on ne transmet qu'une couleur par ligne (l'autre couleur est prise de la ligne précédente) \Rightarrow pas de mélange des couleurs - bon rendu des couleurs

Waveform et vecteurscope

WF impossible de voir les couleurs!!! donc apparition du vecteurscope il affiche le burst à 0° vers la gauche et peut afficher toutes les couleurs. teintes égales angle avec le bruit et la saturation est donnée par la distance à l'origine. impossible de savoir où se trouvent ces couleurs dans l'image donc besoin des deux systèmes.

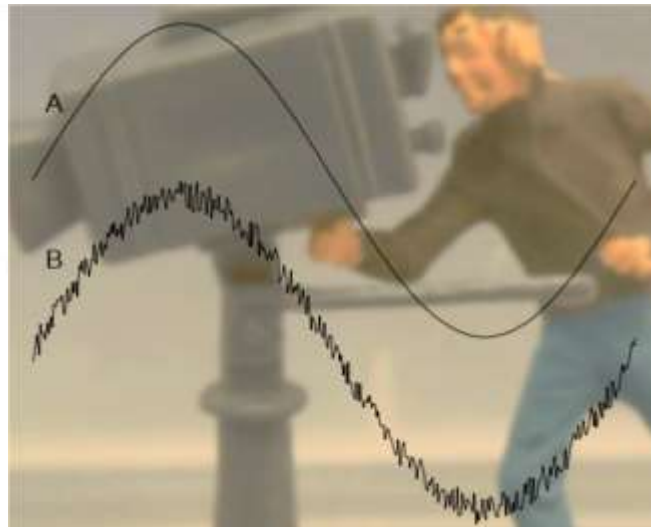
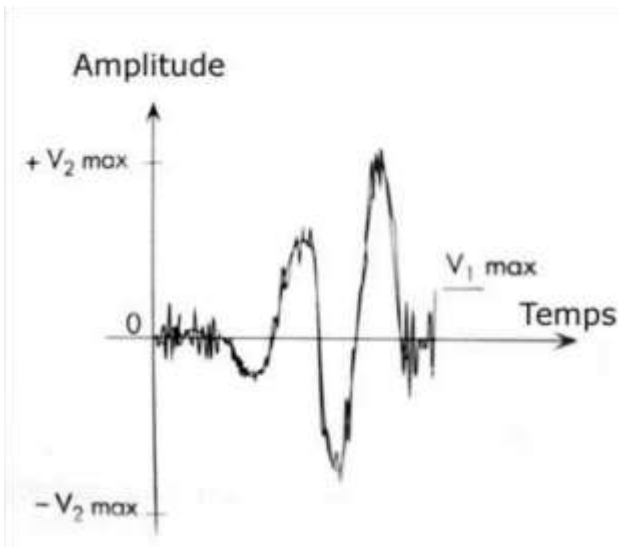






Conversion Analogique vers Numérique

- Le signal analogique est celui qui représente le plus fidèlement les variations d'un phénomène physiologique.
- Problème : le bruit (souffle, parasite) qui se superpose au signal lors de l'enregistrement ou diffusion...
- le bruit est quantifié par le rapport entre l'amplitude maximale du signal d'origine et celle du bruit qui lui est ajouté. (rapport signal/bruit en dB)
- 40 dB en VHS
- 60 dB en bétacam SP
- S/B (rapport signal/bruit) = $20 \log V_2 \max(\text{signal d'origine}) / V_1 \max(\text{bruit})$



En première génération, le bruit n'est pas très gênant, le problème est qu'il s'accumule au fil des générations (copies).

En effet, lors de la copie d'un signal « soufflé » il est impossible de distinguer ce qui est signal original et ce qui est souffle puisque le signal analogique est imprévisible.

Le langage binaire (base du numérique) se compose d'une suite de niveaux 1 ou 0, il est donc prévisible, après un 1 vient soit un 1 ou un 0, même chose après un 0. Le signal binaire peut donc être régénéré après chaque copie (élimination du souffle). Ce qui autorise un grand nombre de traitements sans pertes.

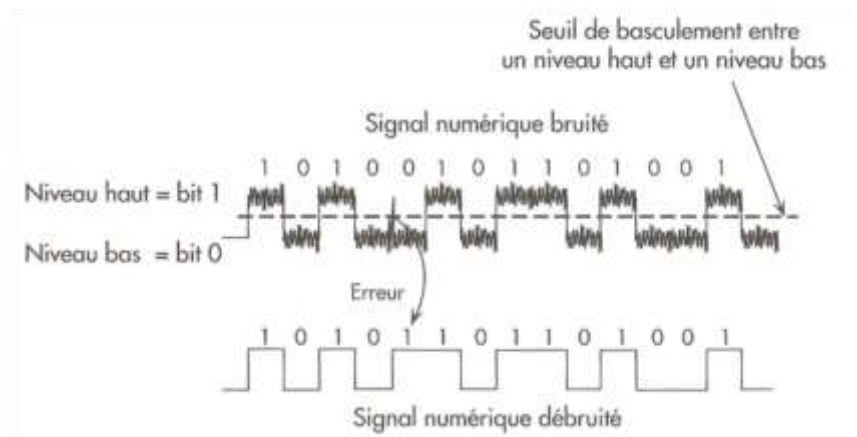
- Précisons que les capteurs ccd d'une caméra sont des dispositifs numériques traitant des signaux analogiques.

La conversion

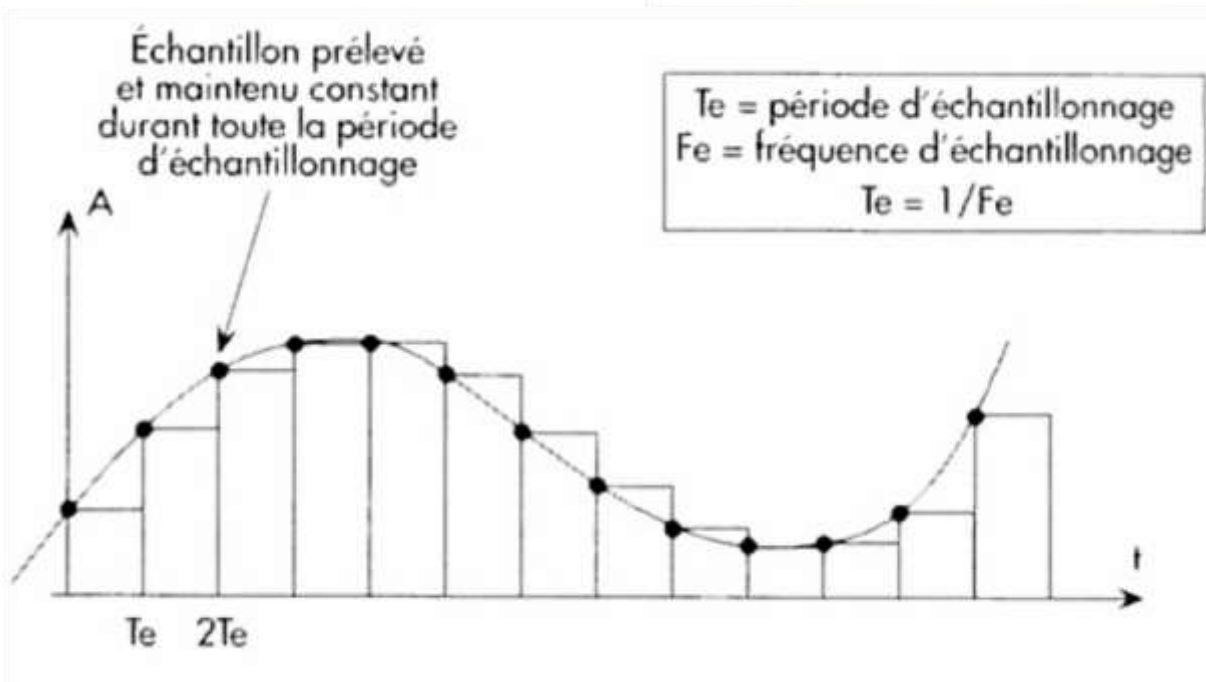
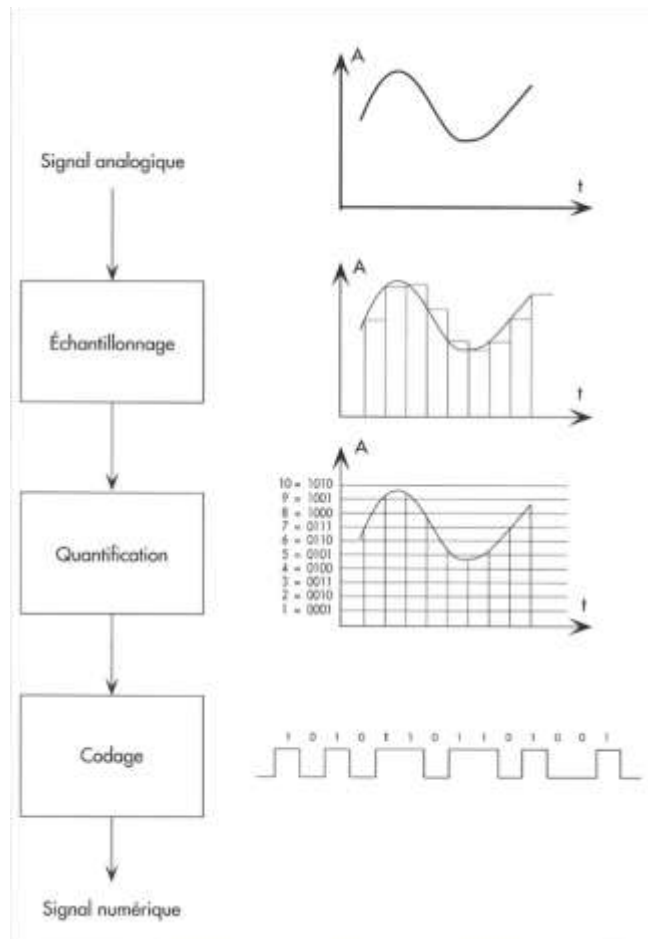
analogique /

numérique (digital

en version anglaise) 1 l'échantillonnage

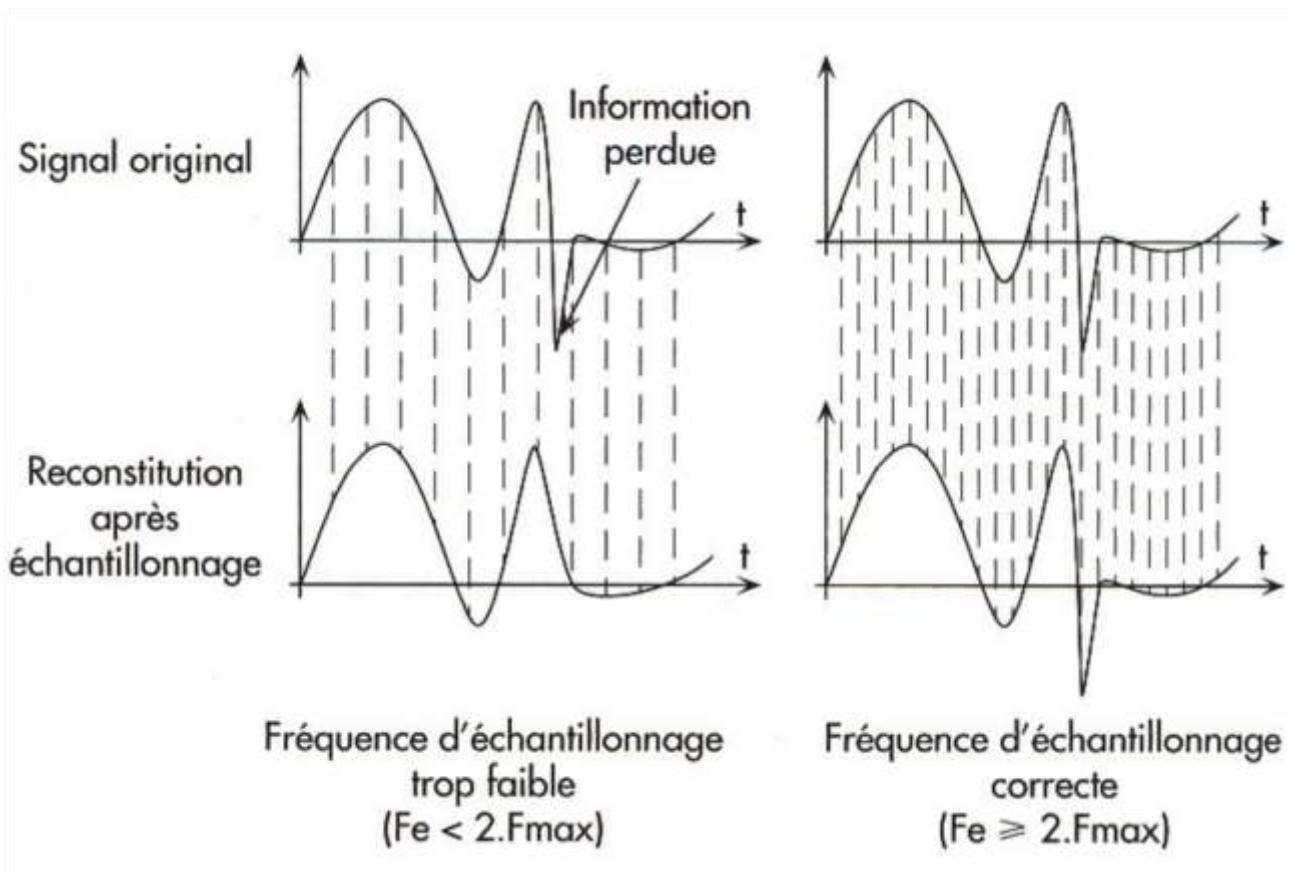


on prélève ponctuellement l'amplitude du signal analogique, et on garde cette amplitude jusqu'à l'échantillon suivant.



La loi de SN permet de déterminer la fréquence d'échantillonnage minimale à choisir.

- la fréquence d'échantillonnage doit être supérieure ou égale à deux fois la fréquence maximale à échantillonner
- Si l'on respecte la loi de Shannon et Nyquist, l'information après échantillonnage sera une représentation intégrale du signal original.



- La fréquence d'échantillonnage doit donc tenir compte de la bande passante du signal vidéo qui est théoriquement de 6,5Mhz.

Ce qui donne une fréquence d'échantillonnage de 13Mhz minimum

- En outre il était souhaitable que cette norme tienne compte des différents systèmes analogiques existants (PAL, NTSC ect)
- la fréquence ligne en 625 L est de 15625 Hz
- La fréquence ligne en 525 L est de 15734 HZ

Après divers tests une fréquence d'échantillonnage de luminance égale à 13,5 MHz a été choisie.

$$13,5 \text{ MHz} = 864 \times \text{syst } 625$$

$$= 858 \times \text{syst } 525 \text{ et}$$

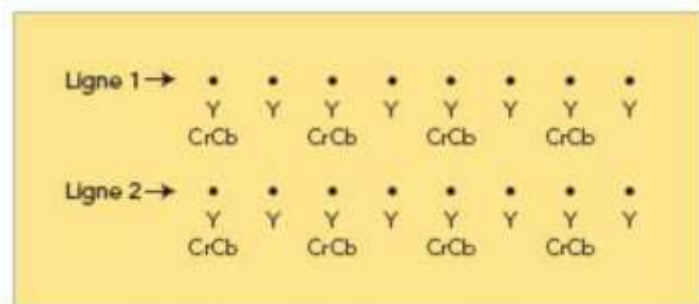
répond à la lois de SN.

- Les signaux de chrominances sont échantillonnés à une fréquence plus faible égale à 6,75 MHz.
- Cette norme a été nommée 4 :2 :2
- ces nombres représentent le rapport entre la fréquence d'échantillonnage de la luminance et des composantes de chrominance.
- le 4 remonte au début des travaux de numérisation, lorsque l'on envisageait de numériser du composite, la fréquence d'échantillonnage était de 17,7 MHz , soit 4 x la fréquence de la sous porteuse couleur. le système a été abandonné, mais le nom est resté.

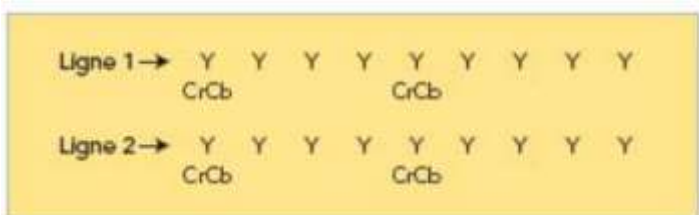
- le 4 :2 :2 est la norme professionnelle, il en existe d'autres, plus légères ou de haute qualité.

Déclinaisons du 4 :2 :2

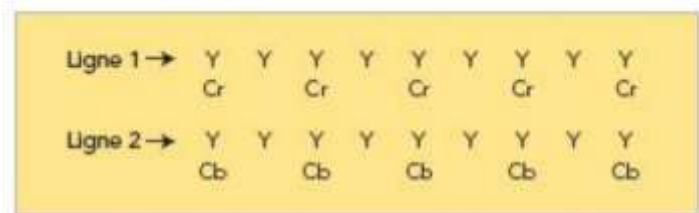
- 4 :2 :2 la luminance est échantillonnée à 13,5 MHz, les composantes de chrominance à 6,75 MHz (bétacam digital)
 - 4 :2 :2 :4 Codage 4 :2 :2 incluant un signal de découpe échantillonné à 13,5 MHz
 - 4 :1 :1 La luminance est toujours échantillonnée à 13,5 MHz, mais les composantes de chrominance sont échantillonnées à 3,375 MHz (soit $\frac{1}{4}$ de la luminance) ex : DVCPRO25
 - 4 :2 :0 Luminance échantillonnée à 13,5 MHz, chrominance à 6,75 Mhz mais une ligne sur deux.
- Ex : DVD, DV, DVCAM
- 4 :4 :4 Toutes les composantes sont échantillonnées à 13,5 MHz, ce type de codage est employé par les stations graphiques .
 - 4 :4 :4 :4 le précédent, incluant un signal de découpe à 13,5 MHz.



Echantillonnage à 4:2:2 des signaux de luminance et de différence des couleurs



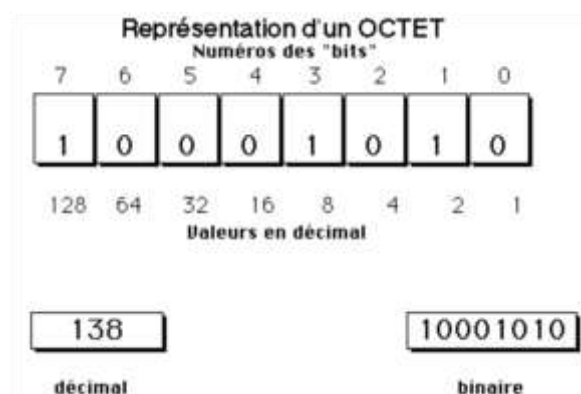
Echantillonnage 4:1:1



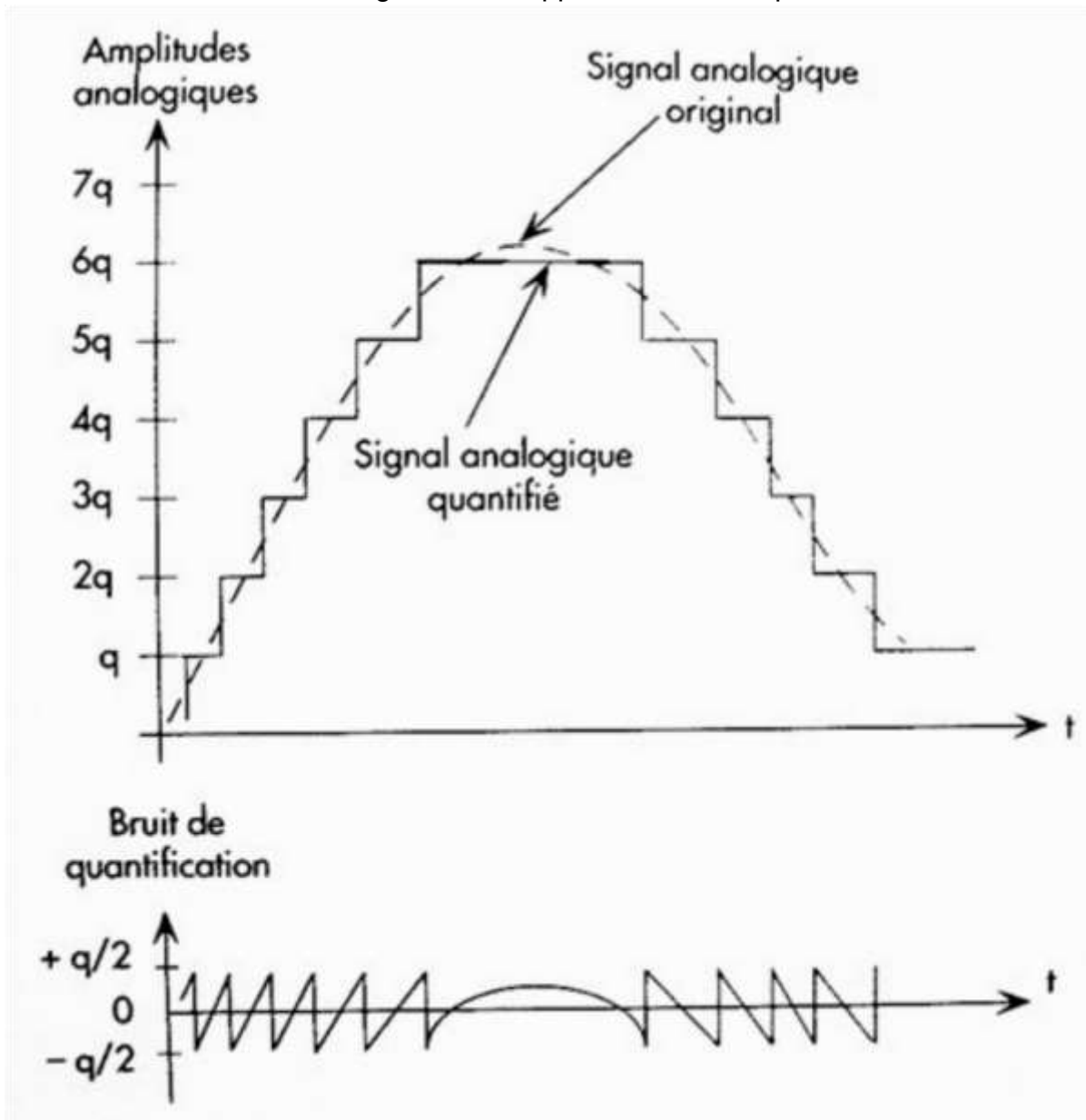
Le 4:2:0 uniformise la résolution verticale et horizontale des couleurs lors de l'utilisation de pixels carrés.

2 La Quantification

- La quantification est la deuxième étape du processus de numérisation.
- elle a pour but de faire correspondre chaque amplitude échantillonnée à un nombre entier exprimé en base 2.
- n chiffres permettent de coder $N = 2^n$
- Dans cette étape le signal analogique qui peut prendre une infinité de valeurs, est converti en un signal constitué d'un nombre fini de valeurs numériques.
- Des erreurs d'arrondis par défaut ou par excès sont donc inévitables, car à plusieurs valeurs proches, mais différentes, correspondra une seule valeur binaire .
- la précision du signal converti sera donc directement liée au nombre de valeurs disponibles.



- en fonction de la plage à quantifier, on définit une échelle constituée d'un nombre fini d'intervalles « q » appelés « pas de quantification » ou « échelons de quantification » ou « quantum » .
- l'erreur de quantification ne dépasse jamais « $q/2$ » .
- La succession des erreurs engendrées s'appelle « bruit de quantification »



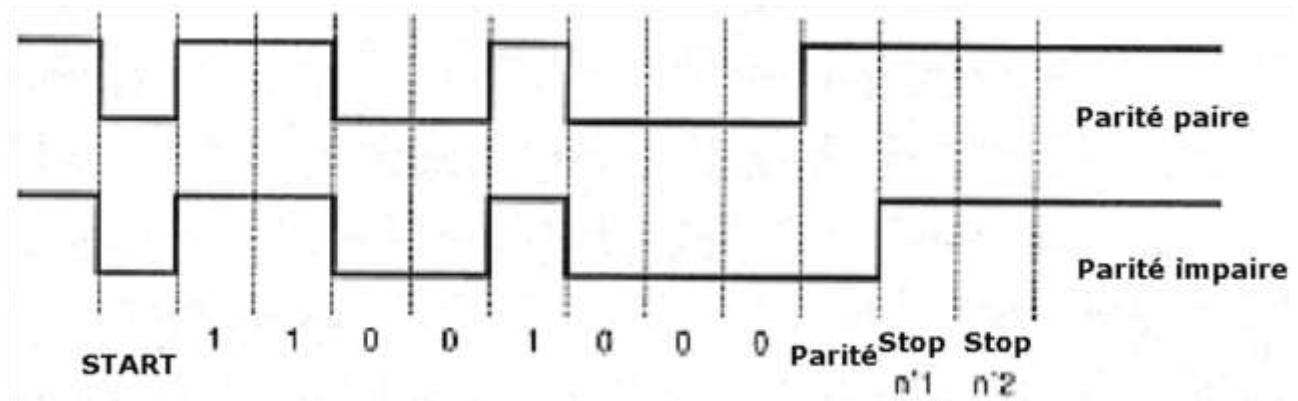
- A l'origine, la norme 4 :2 :2 spécifiait une quantification sur 8 bits des composantes du signal vidéo, ce qui permet de disposer de 256 niveaux possibles.
- Par la suite, pour les productions haut de gamme, la norme a évolué vers le 10 bits, ce qui permet 1024 niveaux.

3 Les corrections d'erreurs

Si toute la chaîne de traitement vidéo fonctionnait idéalement, les signaux numériques pourraient être enregistrés tels quels, ce n'est malheureusement pas le cas.

- les erreurs peuvent provenir :
- des supports d'enregistrement

- des mécanismes d'entraînement des magnétoscopes
- voies de transmission
- etc...
- en numérique, une erreur est soit une inversion de données (1 devient 0), soit une absence de donnée.



- L'entrelacement des données

La plupart des erreurs qui perturbent un signal numérique affectent en général plusieurs dizaines de bits consécutifs, c'est pourquoi a été inventé le principe du brassage de données.

- Le but est d'éloigner des données à l'origine consécutives, dans le but de disperser tout paquet d'erreur.
- Les erreurs deviennent donc plus petites et dispersées.
- Les données sont dispersées suivant un code connu du codeur et du décodeur.

compression intro

Pourquoi compresser ?

PRINCIPES DE COMPRESSION NUMERIQUE

Le signal numérique présente d'énormes avantages, mais consomme énormément d'espace.

une image codée en 4 :2 :2 sur 8 bits occupe un espace de 830 Kilo-octet (partie active) ce qui fait 21 Mbytes la seconde.

- autrement dit sur un CDROM de 640 MB, on pourrait stocker 30 sec de vidéo non compressée, sur un DVD de 4,7GB, 4mn.

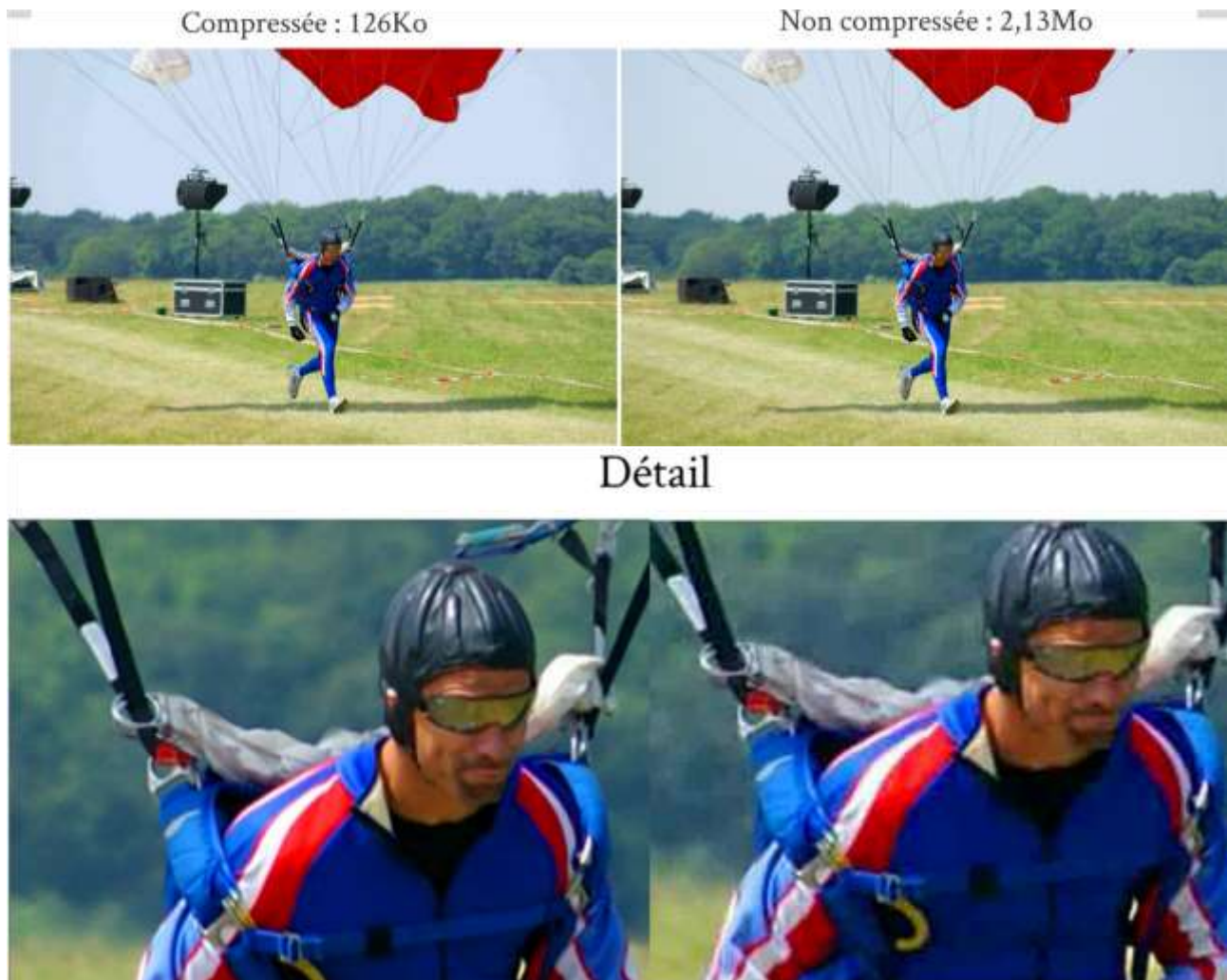
Si l'on veut diffuser et stocker de grandes quantités de données vidéo, il est nécessaire de compresser.

L'art de la compression numérique en vidéo est de supprimer certaines informations de l'image et d'en simplifier d'autres.

2 types de compression.

1 La compression sans pertes ou « lossless » effectue un traitement transparent, permettant de retrouver intégralement les données d'origine après décompression. Malheureusement, elle ne conduit qu'à un taux de compression très faible.

2 La compression avec pertes ou « lossy » aboutit à un taux de compression beaucoup plus élevé, mais impose de négliger certaines informations de l'image. Si elle se fait dans des proportions raisonnables, la compression peut passer inaperçue, on parle alors de compression virtuellement transparente. Si la compression est élevée, elle entraîne des distorsions et artéfacts plus ou moins visibles.



Défauts typiques de la compression.

- effet de bloc : des blocs de pixels deviennent visibles (aspect de mosaïque) taux de compression excessif.
- effet de halo : une frange apparaît sur les contours. Débit trop faible - effet de blurring : détails moins nets. Compression trop élevée.
- bruit de quantification « effet vitre sale » , le bruit de quantification du à la digitalisation s'est accumulé lors de plusieurs cycles de conversion A/D et D/A.
- Mosquito noise : du bruit apparaît sur des transitions d'éléments en mouvement, sous la forme de points noirs. Quantification sur un nombre de bits trop faible.

Les normes de compression vidéo

Il existe deux formes de compression en vidéo : la compression spatiale ou MJPEG et la compression temporelle ou MPEG. Toutes deux avec pertes

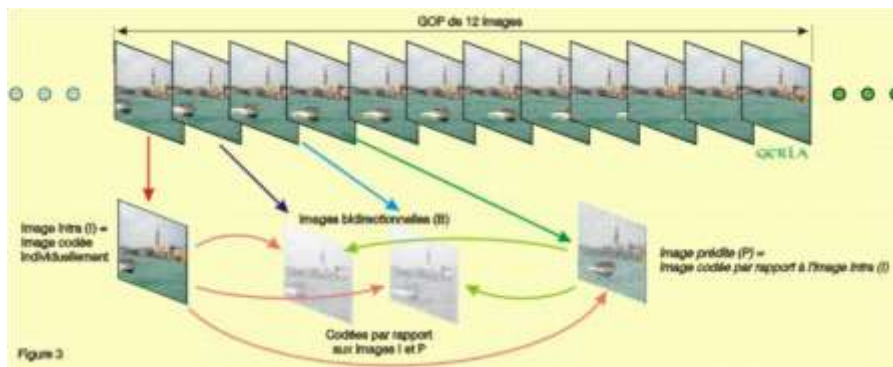
Compression spatiale

Une compression spatiale compresse chaque image indépendamment les unes des autres, c'est une compression Intra-Image

Compression temporelle

MPEG se base sur les redondances entre les différentes images d'une séquence à 25 images/seconde .

C'est une compression Inter-Image



Compression JPEG

Joint Picture Experts Group, JPEG, est un algorithme de compression : Il transforme une image de manière à lui faire prendre moins de place, quitte à perdre en qualité. exemple :



Norme de compression des images fixes instaurée en 1992 créée par un groupe d'experts de la photographie.

Compression avec pertes mais qui ne sont pas forcément visible pour l'oeil humain. On quantifie l'image sur 8 bits soit 256 valeurs différentes. Et le taux de compression est de 1 à 50

Le MJPEG est une adaptation pour la vidéo, celle-ci n'étant qu'une succession rapide de photos. (M = motion).

Ce système n'ayant jamais été normalisé, les fabricants ont développé des systèmes incompatibles entre eux.

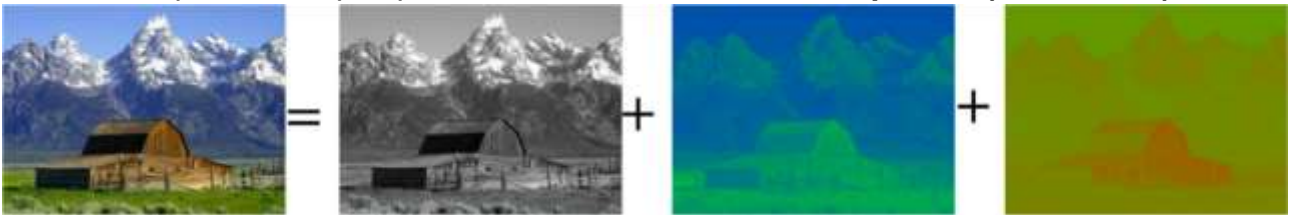
Depuis est apparu le système DV qui utilise une compression de type MJPEG, mais dont l'algorithme a été normalisé.

Une image est composée de milliers de petits carrés, les pixels. On peut les observer en zoomant sur l'oeil du chat :



Ce qu'il est intéressant de voir sur cette photo, c'est qu'il y a en général peu de différence entre deux pixels voisins.

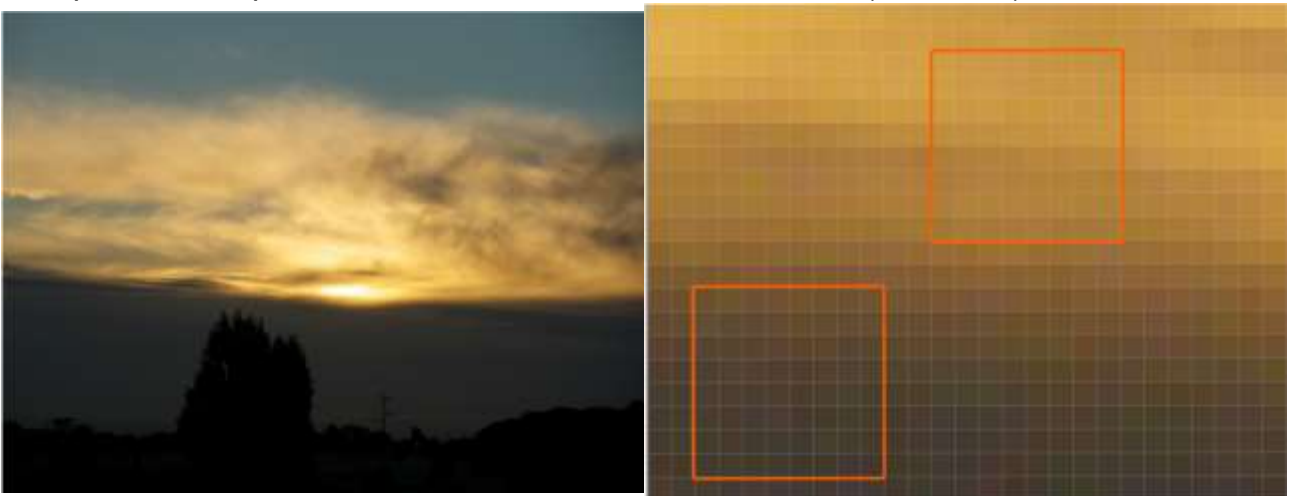
Une chose qu'il ne faut pas perdre de vue : on travail en **composant (Y, R-Y, B-Y)**



Phase 1

la décomposition en blocs et sous échantillonnage de la chrominance :

La première étape consiste donc à découper l'image en macroblocs de 8 pixels sur 8. Chaque bloc comprend donc 64 nombres variant de 0 à 255 (en 8 bits).



On va donc représenter une image en utilisant 3 tableaux, un par composante, par exemple :

Y				Cb				Cr			
141	150	143	147	170	172	174	179	85	84	79	83
142	150	147	144	169	170	168	179	84	80	82	86
140	139	143	145	169	171	173	175	82	89	85	84
138	145	150	148	171	172	170	170	82	87	86	83

Mais comme l'oeil est moins sensible aux variations de chrominances, il ne sert à rien de garder autant de valeurs sur les deux derniers tableaux!

En faisant la moyenne de chaque bloc de deux pixels de côté, on obtient les nouveaux tableaux. (Le Y n'est pas modifié)

Y				Cb				Cr			
141	150	143	147	170	170	175	175	83	83	83	83
142	150	147	144	170	170	175	175	83	83	83	83
140	139	143	145	171	171	172	172	85	85	85	85
138	145	150	148	171	171	172	172	85	85	85	85

Et on ne garde qu'une seule valeur sur les 4 identiques :

Y				Cb		Cr	
141	150	143	147	170	175	83	83
142	150	147	144	171	172	85	85
140	139	143	145				
138	145	150	148				

En effectuant cette opération relativement simple, on a déjà divisé par 4 la taille des données de chrominance à compresser.

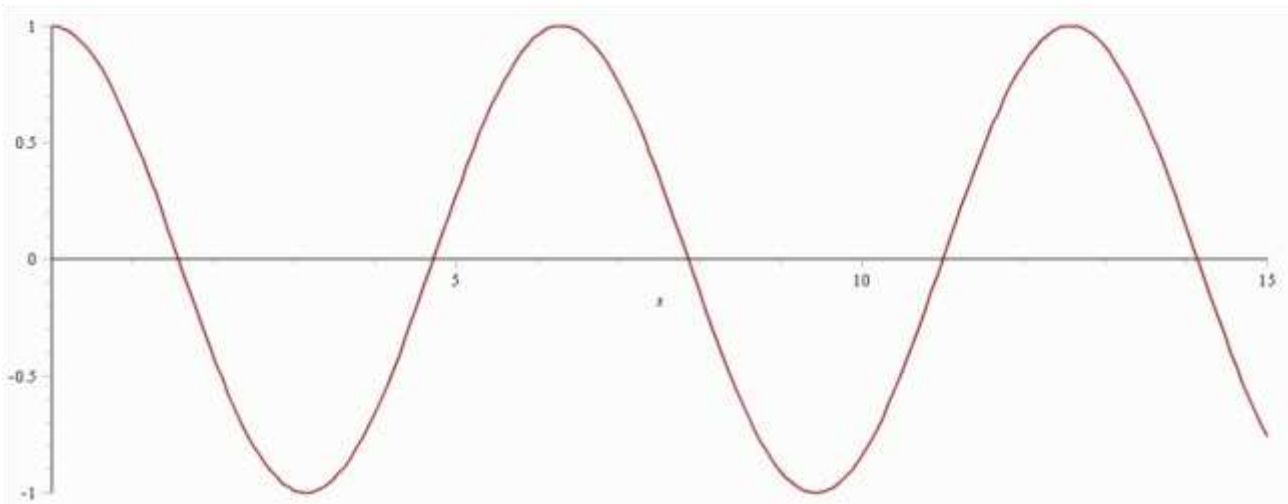
Phase 2

transformée en cosinus discret (DCT)

Appliquer une DCT à ces macroblochs revient à les faire passer du domaine temporel au domaine fréquentiel.

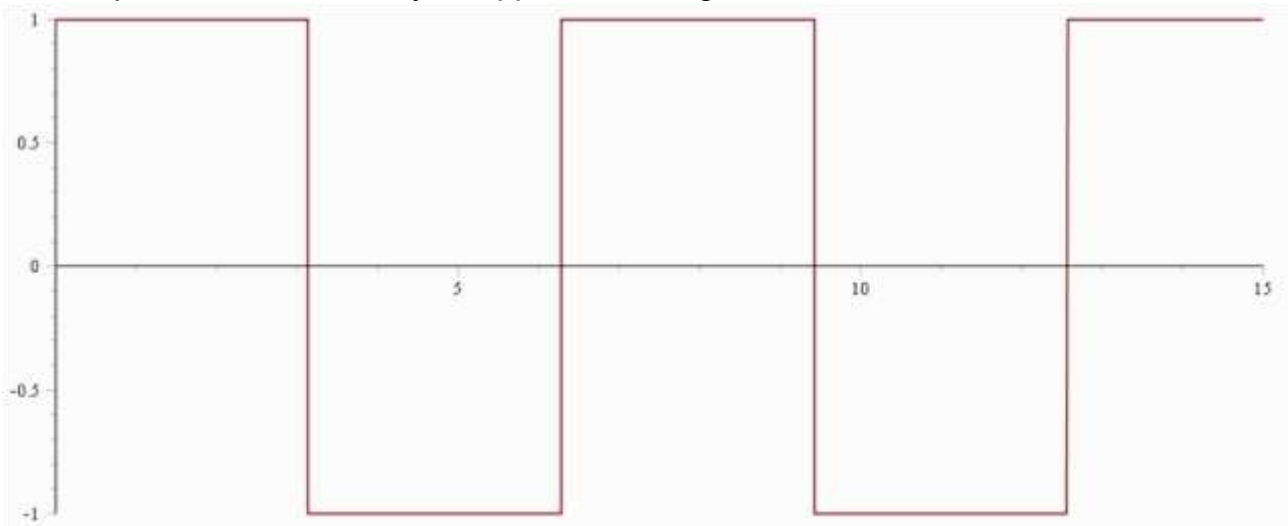
- dans le domaine temporel les valeurs sont décrites les unes à la suite des autres, c'est l'état initial du signal.
- dans le domaine fréquentiel, le signal est décrit par son spectre, par ses composantes spatiales, représentant tout les motifs géométriques qu'il est possible de représenter avec un bloc de 64 valeurs.

La transformée en cosinus discrète le principe, première chose, un cosinus c'est ça :

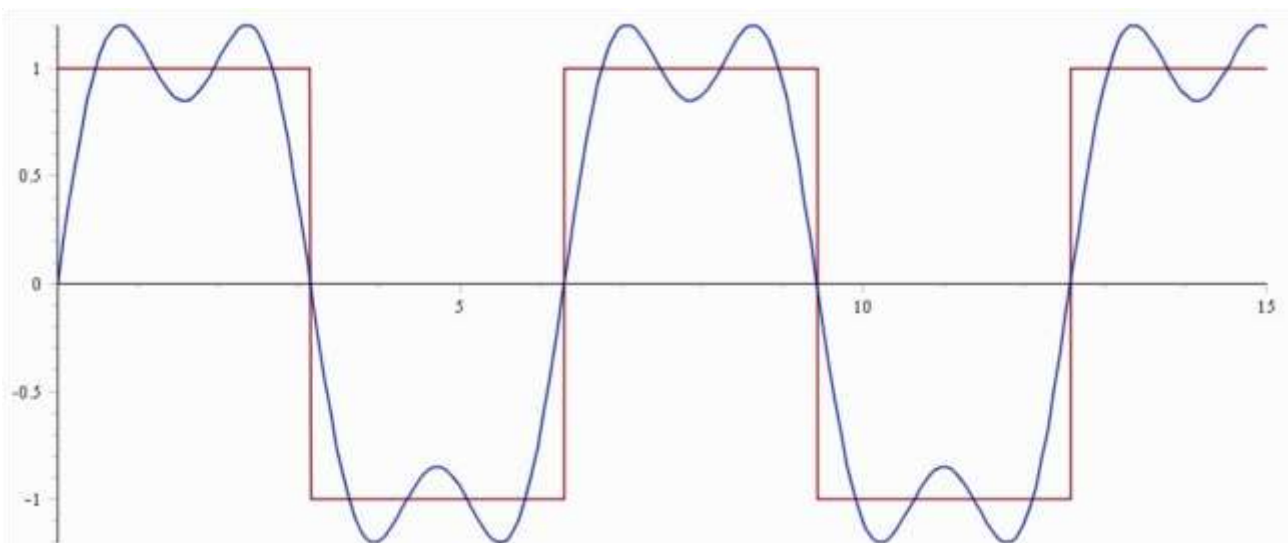


discrèt, fait référence à la discontinuité du signal.

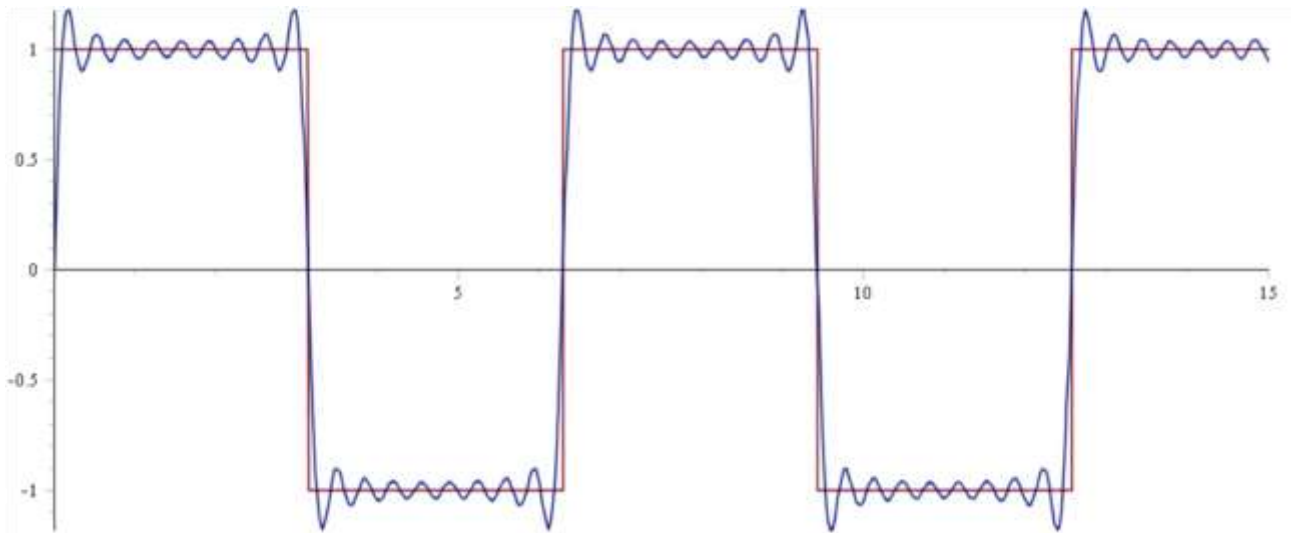
Que se passe-t-il si l'on essaye d'approcher ce signal en utilisant des cosinus?



Ça donne ça :



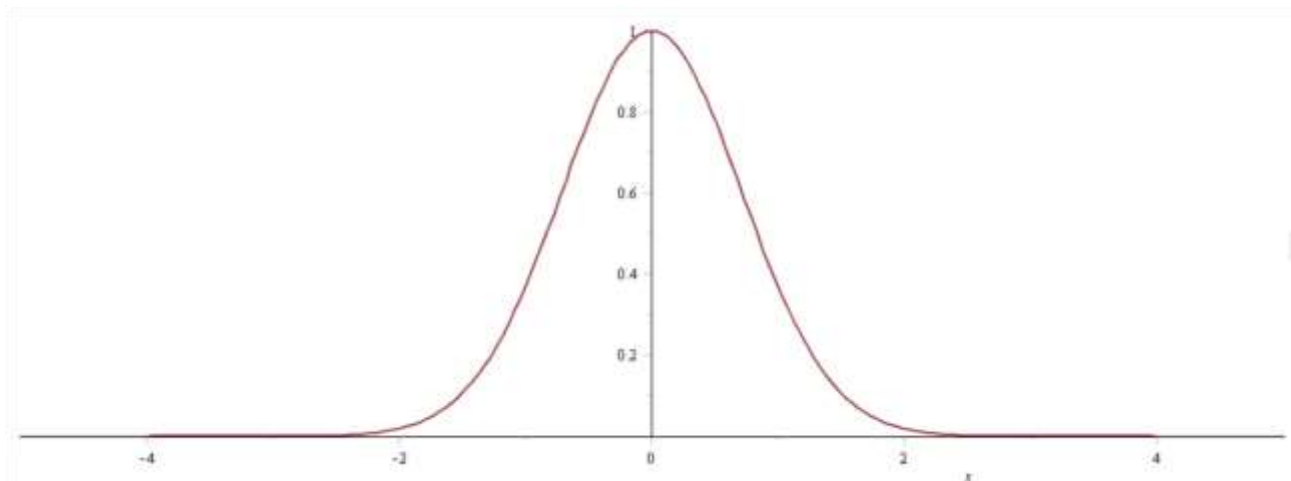
Si on rajoute d'autres cosinus, disons 8, qui oscillent de plus en plus rapidement, le signal carré est bien mieux approché :



Il faudrait en fait sommer une infinité de cosinus pour avoir exactement le signal carré.

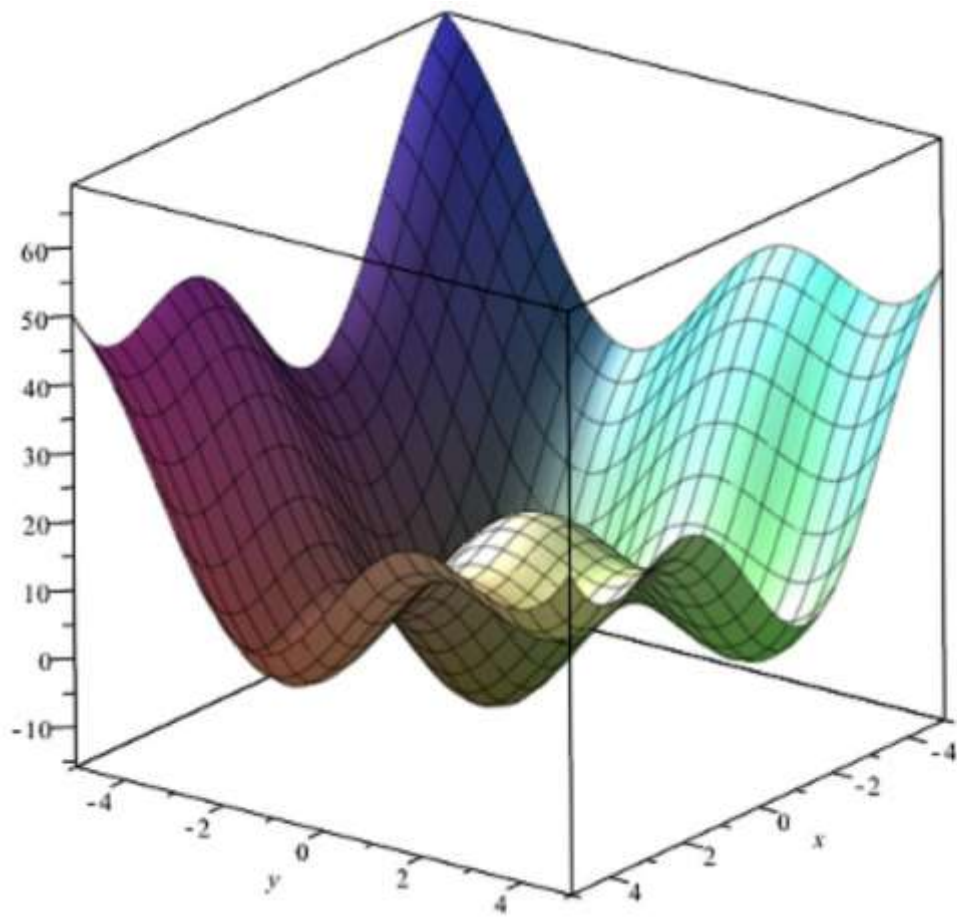
Dans le cas des signaux périodiques comme celui-ci (avec un motif qui se répète), on dit que l'on fait la décomposition en série de Fourier du signal.

Mais le même type de décomposition existe pour des signaux non périodiques comme celui-ci :

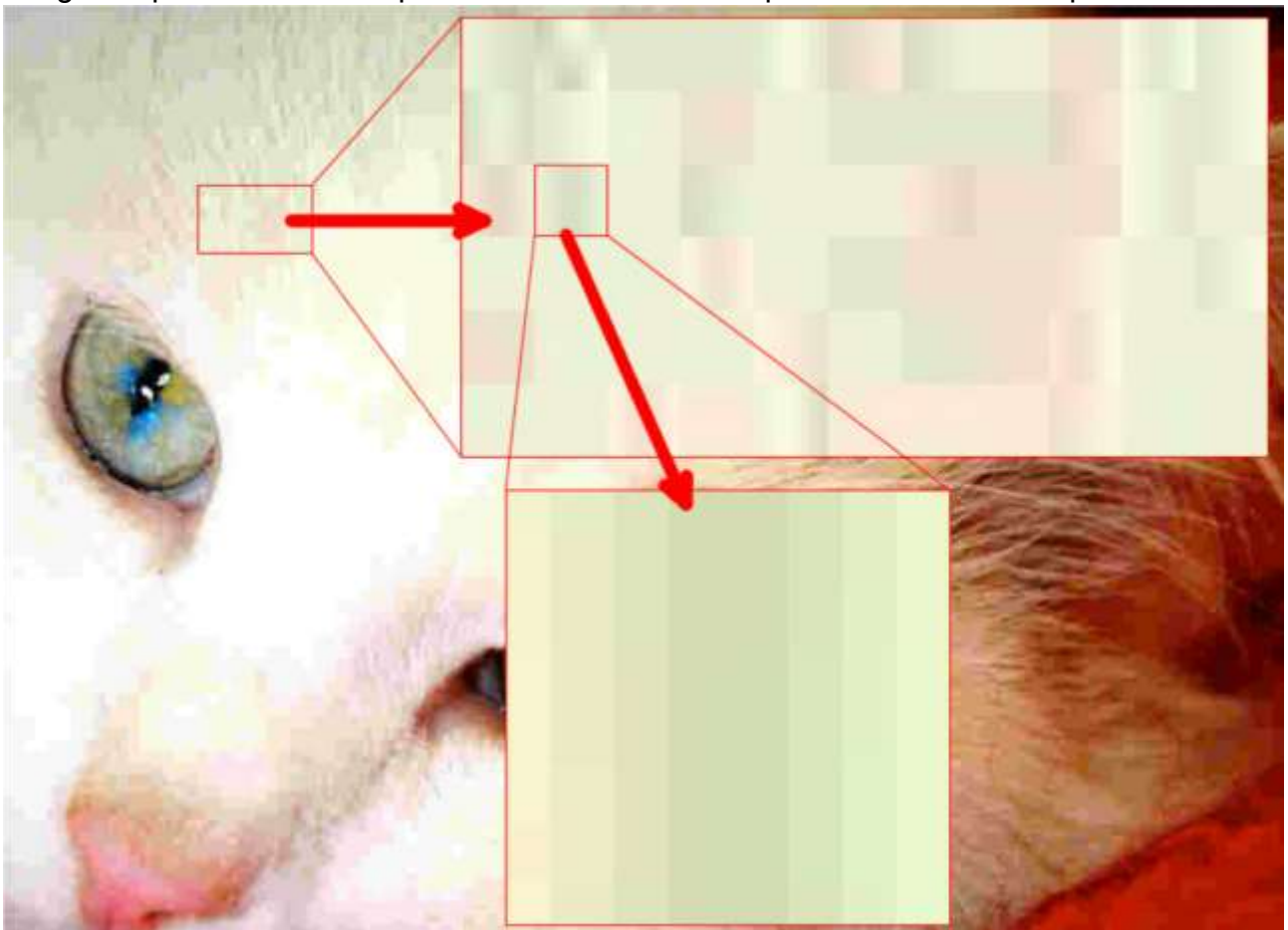


On parle de transformée de Fourier.

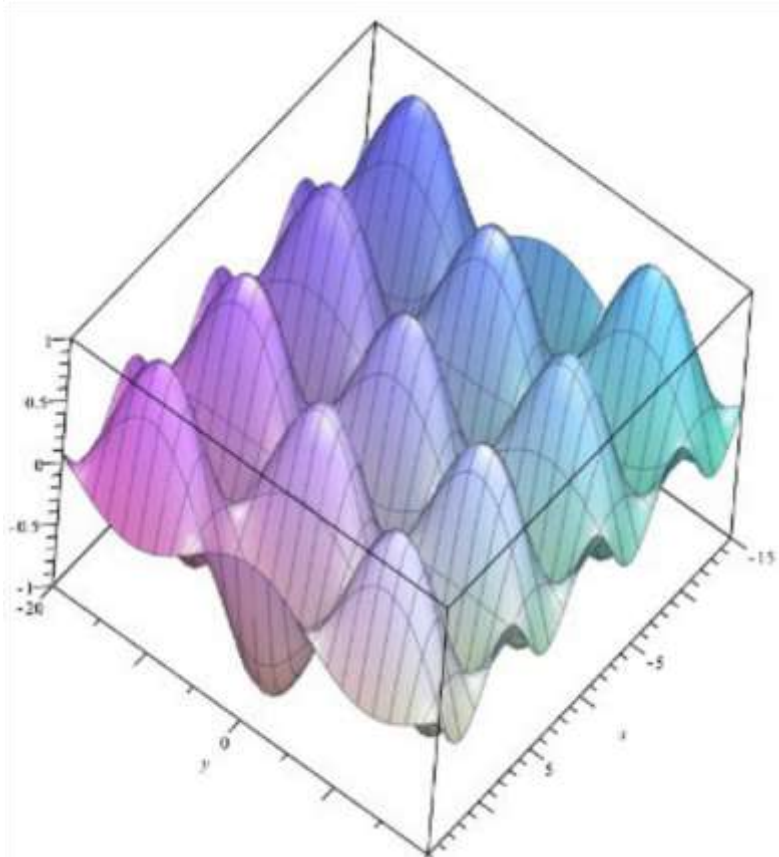
Imaginons maintenant le cas d'un signal en deux dimensions, comme une image :



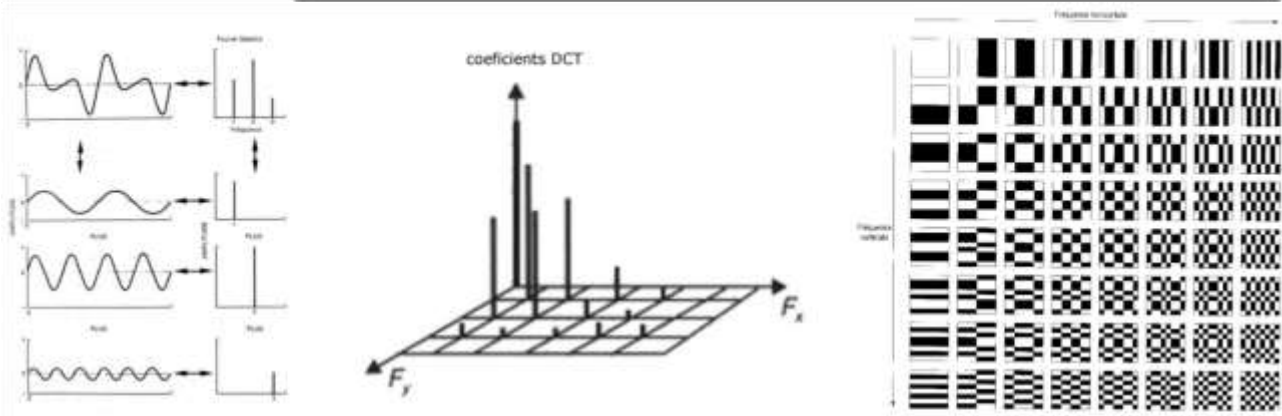
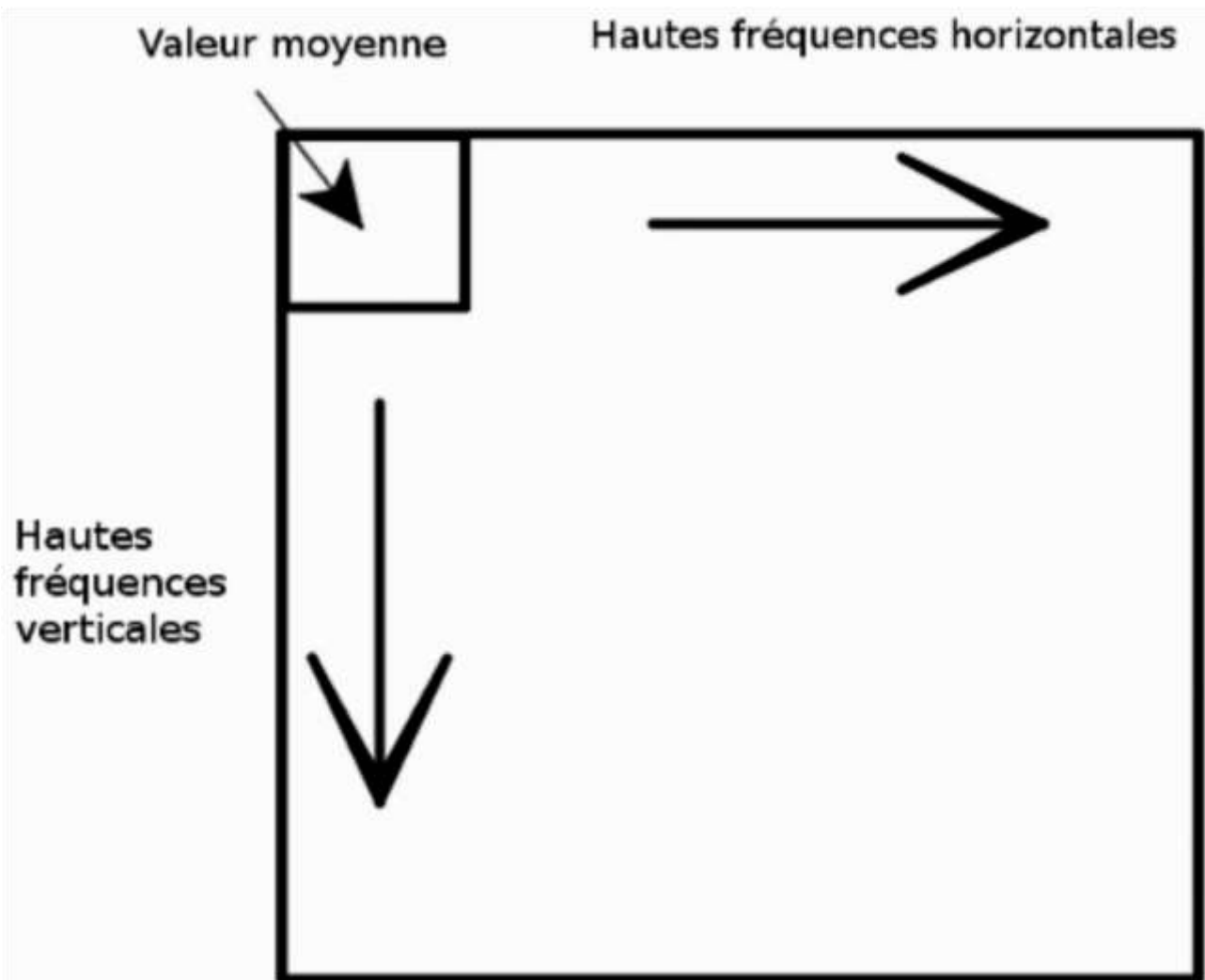
Imaginez que l'altitude d'un point sur la surface correspond à la valeur d'un pixel.



Ce type de surfaces peut également se décomposer en une somme de cosinus, qui vont devoir osciller dans deux directions :



On écrit ensuite dans un nouveau tableau, lui aussi de taille 8, les amplitudes de chacun des cosinus composant le signal. Ils sont classés en faisant apparaître les hautes fréquences vers le coin inférieur droit du tableau.



Chaque bloc sera converti en fréquence par la DCT. L'encodeur ne sait pas à l'avance quelles fréquences sont présentes dans le bloc, il doit donc couvrir toutes les fréquences possibles avec un bloc de 64 pixels.

Ce qui veut dire que chaque bloc est encodé comme une combinaison linéaire de 64 motifs dont les fréquences croissent du haut vers le bas et de la gauche vers la droite .

Comme nous encodons un bloc de 64 pixels et que nous avons 64 motifs possibles, chaque motif se verra assigné d'un coefficient qui représentera l'importance de ce motif dans le bloc analysé.

Notons que la DTC ne compress pas, elle ordonne fréquemment, elle facilite la compression, elle est de plus totalement réversible.

La compression JPEG se base sur le fait que dans une image courante, il y a en général beaucoup plus de basses fréquences, les coefficients les plus élevés se trouveront n général en haut et à gauche. La DTC a donc la propriété remarquable de concentrer l'énergie du bloc sur peu de coefficients situé sur l'angle supérieur gauche.

DCT

$$Img(x,y) \rightarrow F(u,v) \quad F(u,v) = \frac{2}{N} c(u).c(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} Img(x,y) \cdot \cos \left[\frac{\pi}{N} u \left(x + \frac{1}{2} \right) \right] \cdot \cos \left[\frac{\pi}{N} v \left(y + \frac{1}{2} \right) \right]$$

DCT inverse

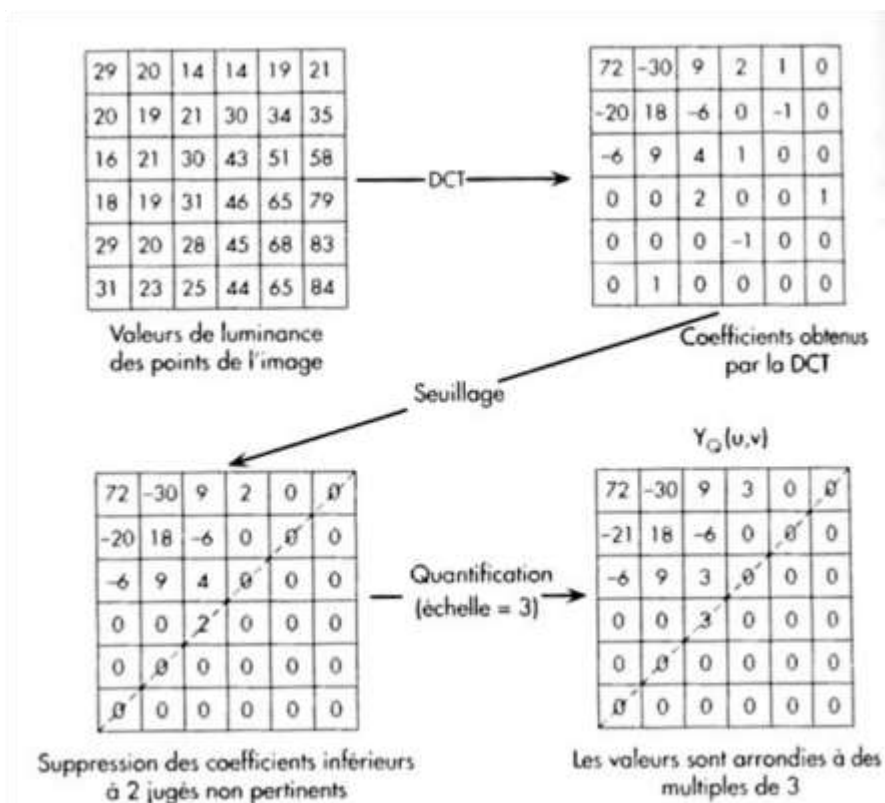
$$F(u,v) \rightarrow Img(x,y) \quad Img(x,y) = \frac{2}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} c(u).c(v).F(u,v) \cdot \cos \left[\frac{\pi}{N} u \left(x + \frac{1}{2} \right) \right] \cdot \cos \left[\frac{\pi}{N} v \left(y + \frac{1}{2} \right) \right]$$

Phase 3

Quantification des coefficients DCT.

La vision humaine étant moins sensible aux détails fin qu'aux plages uniformes, on va « sous coder » les fréquences spatiales élevées.

Les valeurs situées dans les hautes fréquences seront arrondies, codées sur peu de bits, parfois même éliminées.



Dans l'exemple ci-contre, après la DCT, tous les coefficients inférieurs à deux ont été supprimés, ensuite, les coefficients ont été arrondis à des multiples de trois (les valeurs 2 et 3 sont des exemples plus ils sont élevés, plus l'image sera détériorée).

Cet exemple illustre une façon de compresser en JPEG, il existe de nombreux codecs utilisant des algorithmes différents.

Autre solution, la matrice de quantification :

Exemple de matrice de quantification

5	9	13	17	21	25	29	33
9	13	17	21	25	29	33	37
13	17	21	25	29	33	37	41
17	21	25	29	33	37	41	45
21	25	29	33	37	41	45	49
25	29	33	37	41	45	49	53
29	33	37	41	45	49	53	57
33	37	41	45	49	53	57	61

Chaque amplitude est divisée par le nombre correspondant dans le tableau. Par exemple, la valeur moyenne est divisée par 5, et l'amplitude associée à la plus haute fréquence est divisée par 61.

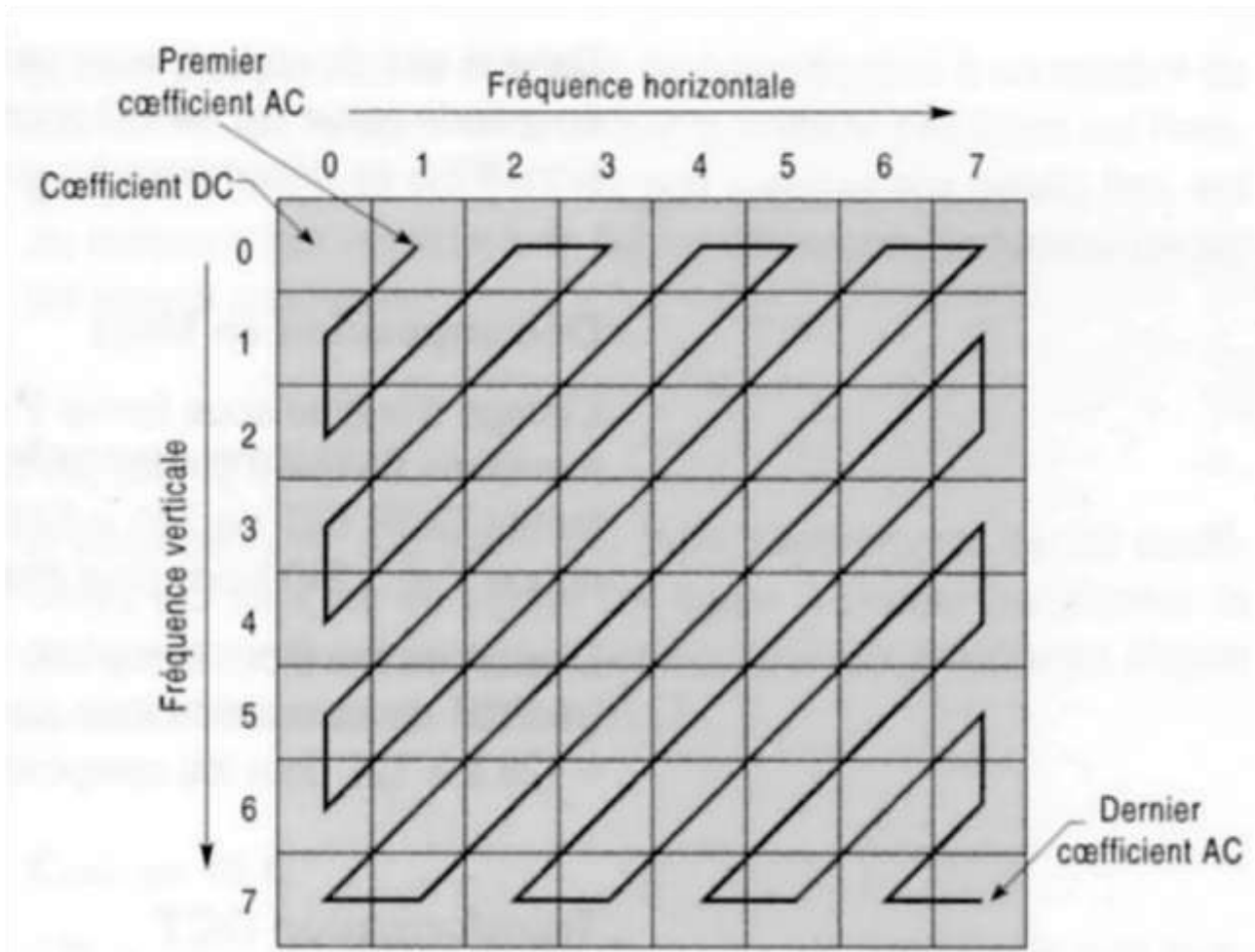
On voit que les amplitudes en bas à droite vont être divisées par des coefficients beaucoup plus grands. En arrondissant à l'entier le plus proche, elles seront probablement ramenées à 0. C'est à ce moment que l'on perd en qualité. En effet, pour retrouver l'amplitude de départ, il suffit de multiplier par le coefficient dans la matrice de quantification (aux arrondis près). Sauf pour 0 ! On aura beau multiplier par ce qu'on veut, on aura toujours 0, donc ces valeurs sont perdues, et l'image affichée lors de la décompression ne sera pas exactement la même.

Phase 4

Balayage en zigzag de la matrice.

La matrice obtenue après quantification est balayée en zigzag, de sorte que les coefficients soient réarrangés sous forme d'un vecteur.

On part des basses fréquences vers les hautes fréquences.



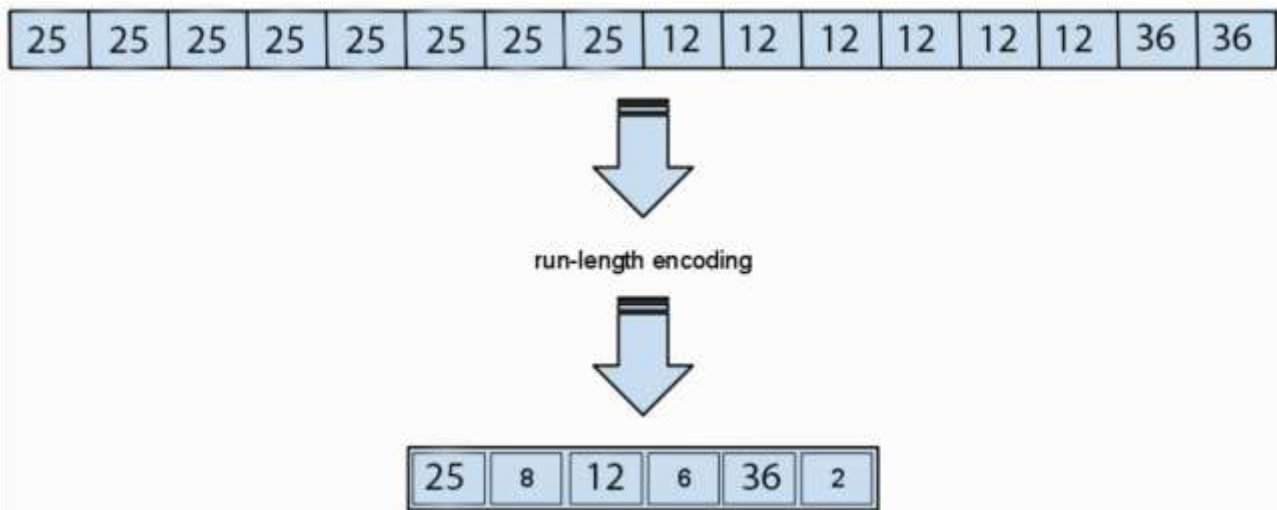
Phase 5

Codage à longueur variable .

Le vecteur formé contient des suites plus ou moins longues de coefficients de valeurs identiques, parfois nulles.

Plutôt que de les coder systématiquement, le codage à longueur variable (RLE) formate le vecteur sous la forme de paires de données.

La première indique la valeur du coefficient, la seconde le nombre de fois ou il est répété.



Phase 6

Codage entropique.

Le principe du codage entropique est d'attribuer le plus court au coefficients les plus fréquents, et le code le plus long aux coefficients statistiquement les moins fréquents. Ce codage est réalisé après une analyse statistique de l'image compressée.

NB : En image fixe, on code ensuite sur le nombre de bits nécessaires par images, en MJPEG, on code sur le nombre de bits nécessaires à la séquence.

Algorithme de Huffman

Il va permettre de donner précisément les bits à écrire pour constituer le fichier final (les 0 et 1). Par exemple, 5 peut devenir '010'. Il n'existe cependant pas de dictionnaire universel : la traduction dépend de la suite de caractères que l'on veut écrire. Dans un autre fichier, 5 pourrait devenir « 1001 ».

Alors comment sont définies les traductions ? En se basant sur la fréquence d'apparition de chacun des caractères.

Prenons un exemple : comment écrire la phrase « LE HOLLANDAIS VOLANT » en utilisant le moins de place possible ?

Certains caractères, comme le L et le A, sont répétés plusieurs fois. Il faut trouver une suite de 0 et de 1 suffisamment courte pour ne pas perdre de place inutilement (imaginez que L devienne '01000111100000', on aurait 4 fois ça dans le texte final !)

On commence par classer chacun des caractères selon leur nombre d'apparitions dans la phrase.

Ici, cela donne :

L 4

A 3

(espace) 2

O 2

N 2

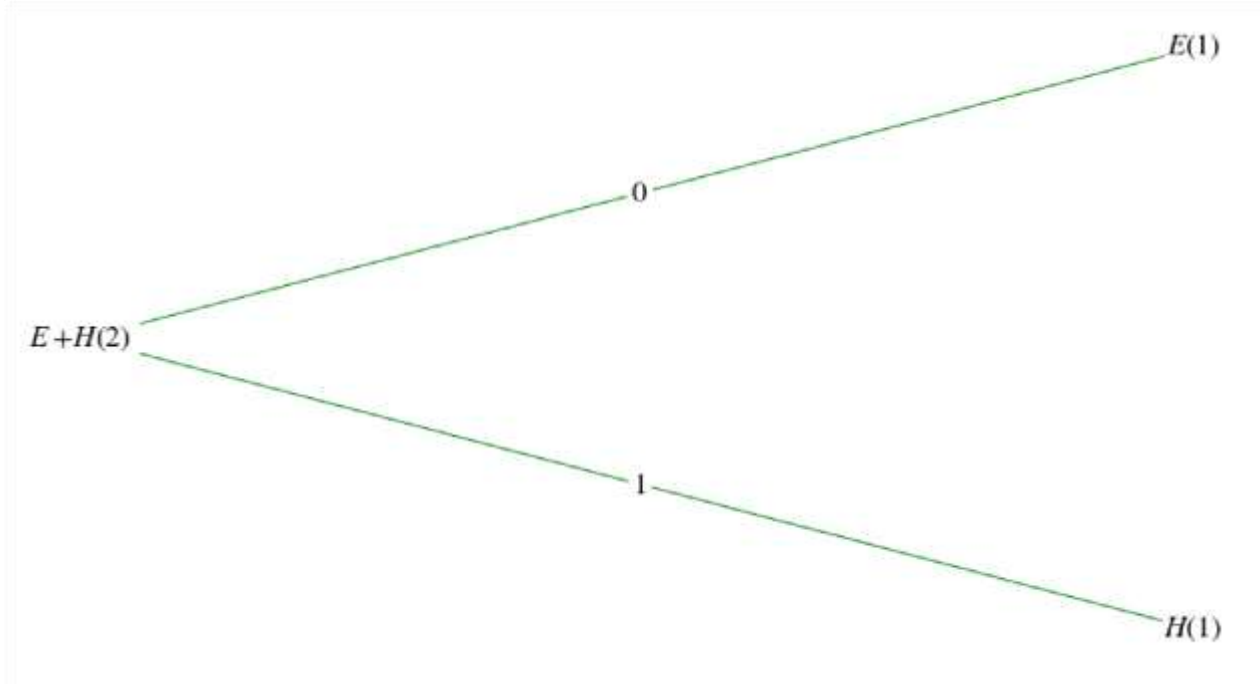
E 1

H 1

D 1

I 1
S 1
V 1
T 1

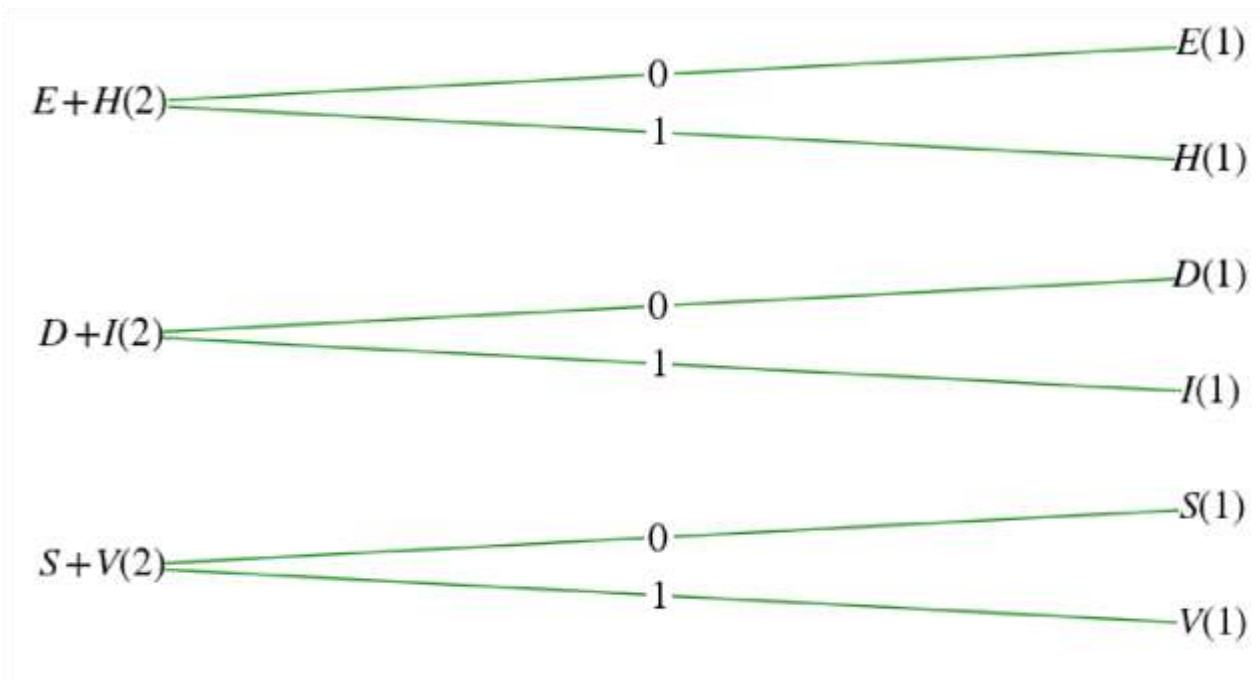
Puis, on forme un arbre à partir de ces lettres, en commençant par les deux lettres apparaissant le moins souvent, par exemple E et H. On dit alors que E et H sont les feuilles de l'arbre, et que E+H est un nœud.



Comme les deux lettres apparaissent une fois, on considère ce « couple » comme une lettre apparaissant deux fois. On se retrouve avec les nouvelles lettres : L 4

A 3
(espace) 2
O 2
N 2
E+H 2
D 1
I 1
S 1
V 1
T 1

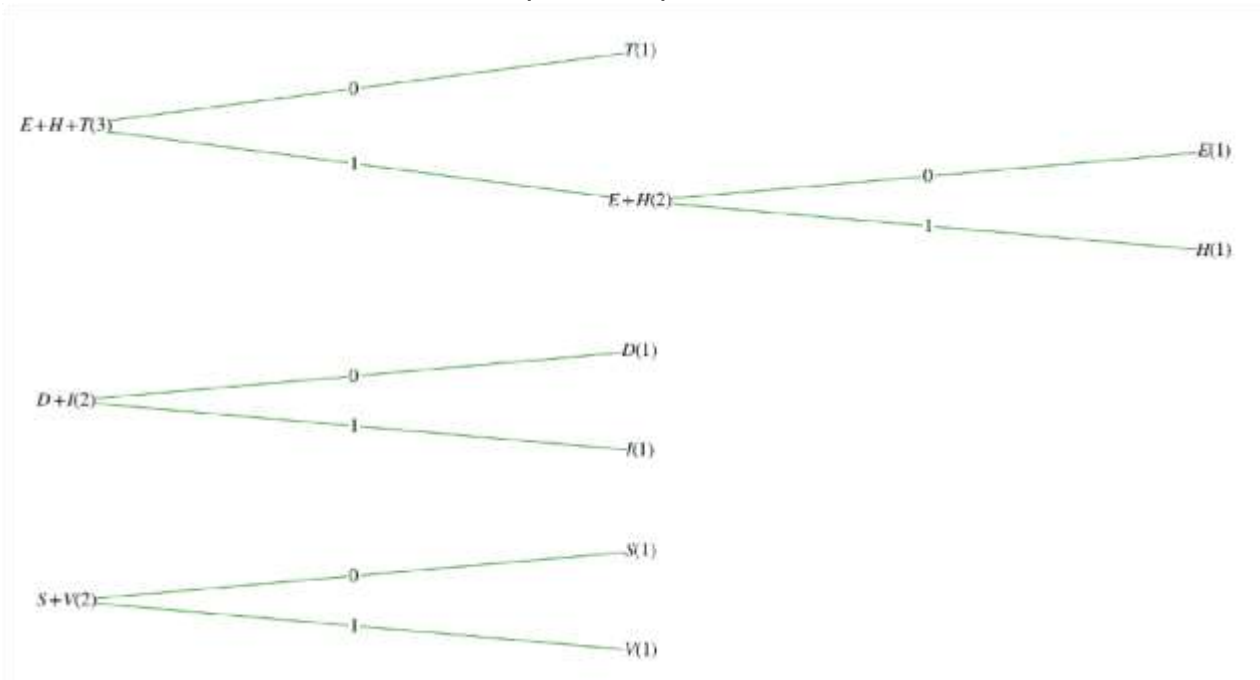
On ajoute un nouveau nœud à l'arbre, en prenant deux lettres parmi celles qui apparaissent le moins souvent :



Il nous reste donc ceci :

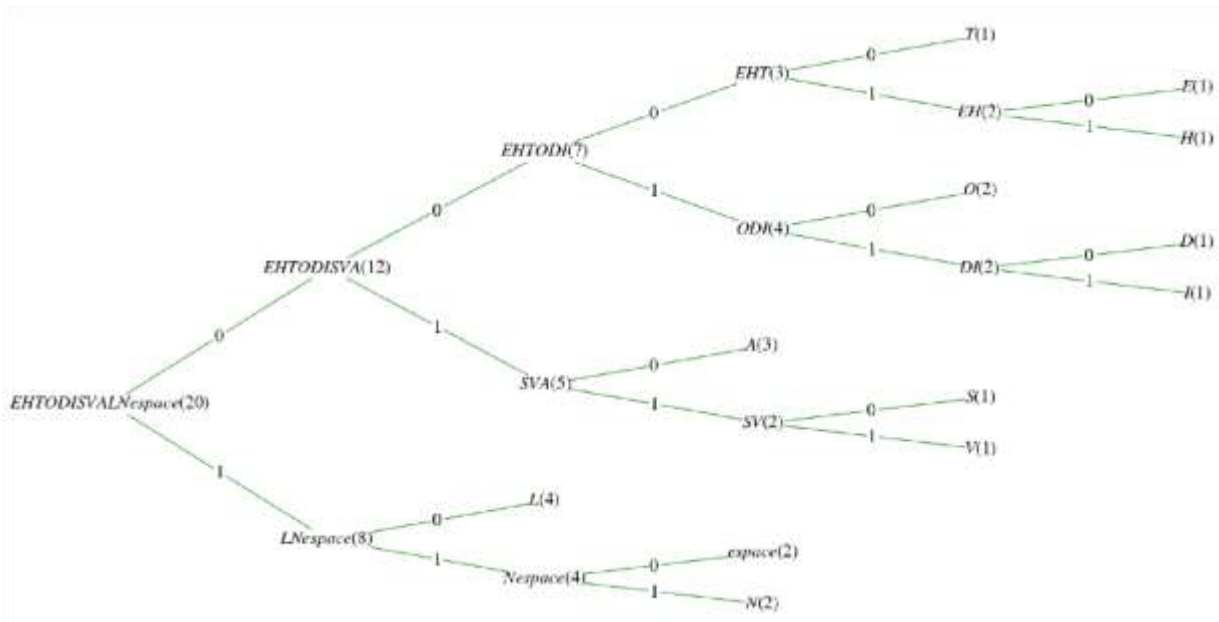
- L 4
- A 3
- (espace) 2
- O 2
- N 2
- E+H 2
- D+I 2
- S+V 2
- T 1

On va maintenant associer le T avec, par exemple, E+H :



On a donc une nouvelle « lettre », de poids 3, E+H+T.

En continuant de la même manière jusqu'à insérer toutes les lettres, on trouve par exemple cet arbre. Il n'est pas unique, cela dépend des lettres que l'on choisit d'associer.



Il suffit alors de parcourir l'arbre jusqu'à ses feuilles pour en déduire le code correspondant à chaque lettre !

Pour atteindre le T, on ne passe que sur des branches étiquetées avec un 0, donc T devient '0000'. Pour aller sur le L, on va d'abord vers le bas puis vers le haut : L devient '10', et c'est le code le plus court.

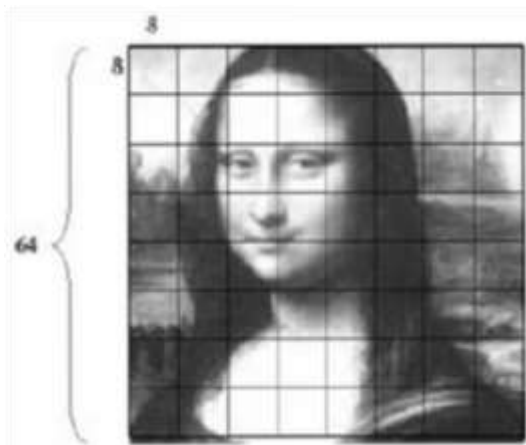
Enfin, la traduction d'une lettre ne peut être le début d'une autre : cela voudrait dire qu'une feuille est en fait un nœud qui permet d'atteindre une autre lettre. En utilisant ce dictionnaire, la phrase 'LE HOLLANDAIS VOLANT' s'écrit :

10 00010 110 00011 0010 10 10 010 111 00110 010 00111 0110 110 0111 0010 10 010 111 0000 (normalement il n'y a pas d'espaces, ici ils séparent les caractères).

Dans la compression JPEG, cet algorithme est appliqué à des nombres entiers et pas à des lettres, mais le principe est exactement le même. En mettant bout à bout toutes les suites de 0 et de 1 correspondant à chaque bloc, on écrit petit à petit le fichier JPEG.

Résumé

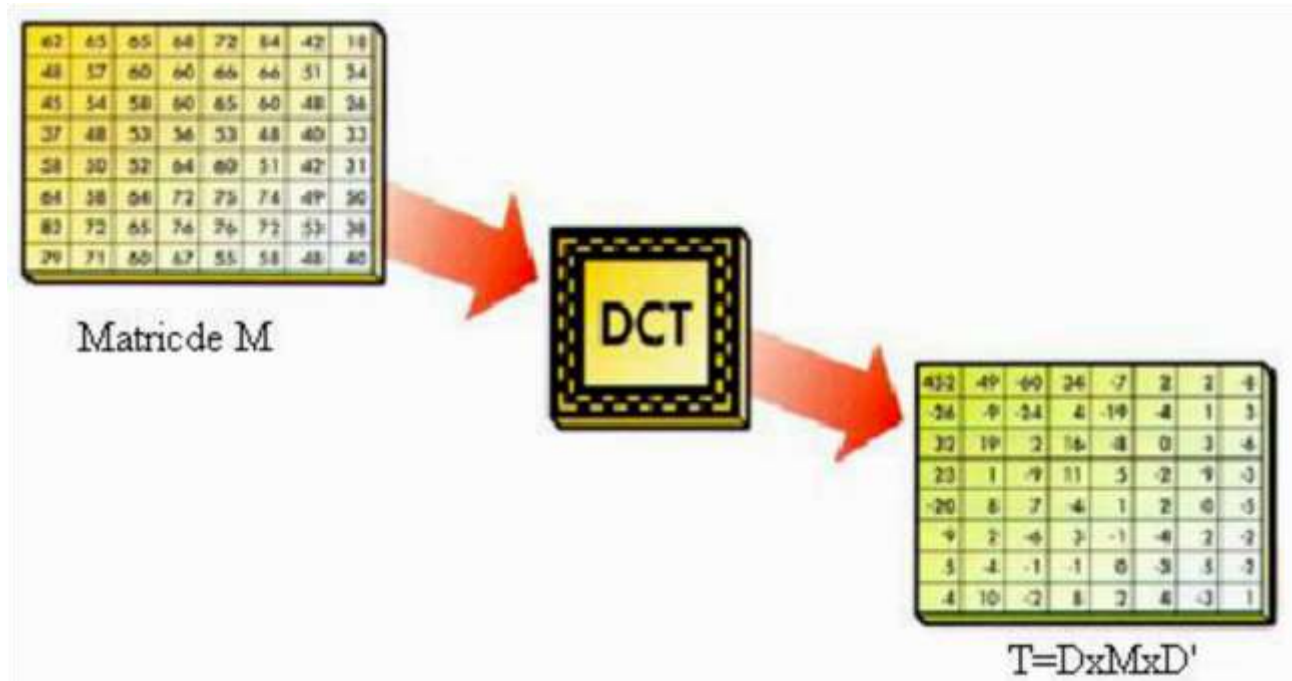
1 Découpage:



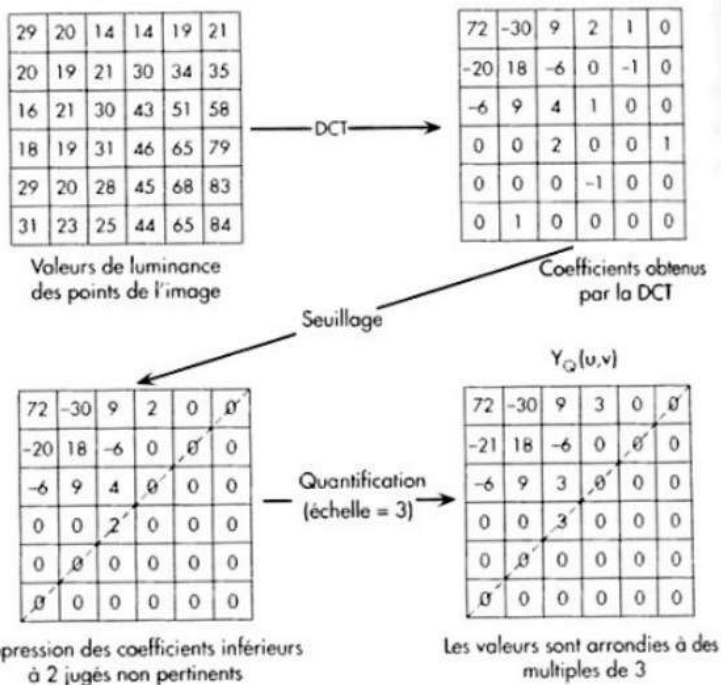
On divise l'image en blocs de 8X8 pixels.

2 Transformée en Cosinus Discret (DCT)

On applique une Transformée en Cosinus Discret à chacun des blocs. On obtient alors des coefficients représentant les différentes fréquences de l'image pour un canal donné. En haut à gauche se trouve les basses fréquences, en bas à droite les hautes fréquences.



3 Seuillage et Quantification



Le seuillage va éliminer les coefficients liés aux hautes fréquences si ils ne sont pas pertinents (peu représentés), la quantification va attribuer un commun diviseur à la matrice dans le but d'obtenir des suites de coefficients identiques.

4 Zig-Zag

Scanner en Zig-Zag les coefficients suivant un algorithme (ZigZag) pour obtenir une longue série de zéro.



la même image compressée à 1/100



La compression temporelle ou MPEG

Dans la compression spatiale en vidéo (MJPEG), chaque image était compressée, indépendamment des autres images. Chaque image contient donc l'entièreté des informations qui lui sont nécessaires.

La compression temporelle (MPEG) tient compte des redondances entre images.

Une séquence de type MPEG se compose de trois types d'images :

Les images I (intra)

Une image I est codée en mode intra-image en JPEG, elle contient toutes les informations nécessaire à son décodage.

Les images P (prédites)

Comme leur nom l'indique, les image P sont prédites à partir des images I, à l'aide des vecteurs de mouvements, c'est-à-dire le calcul du déplacement des éléments de l'image.

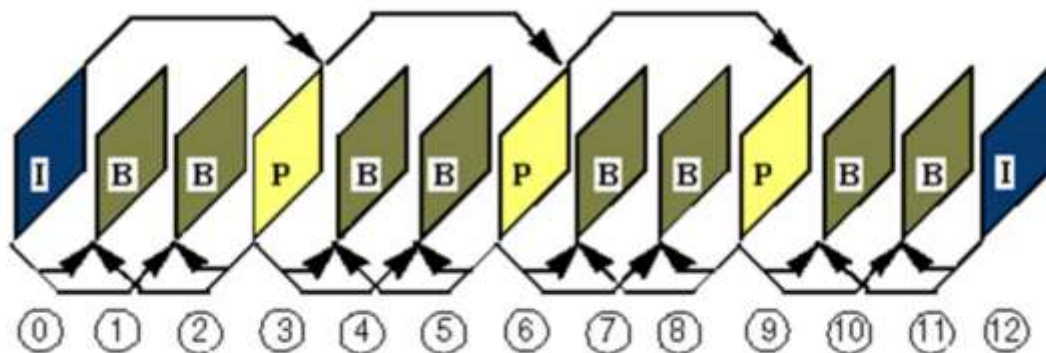
Une image P est typiquement trois fois moins lourde qu'une image I.

Les images B (bidirectionnelles)

Elles sont calculées bi-directionnellement à partir des images I et des images P à l'aide des vecteurs de mouvement.

Une image B est typiquement trois fois moins lourde qu'une image P

Une séquence MPEG composée des trois types d'images précitées s'appelle un GOP (group of pictures).



Un gop commence par une image I et se termine par l'image précédant l'image I suivante. Un GOP peut se résumer à une seule image I, d'une séquence IP, IB, ou IPB.

Dans un GOP, nous avons deux paramètres : M étant l'intervalle entre deux images P (ou entre I et P) et N, qui est l'intervalle entre deux images I (égale à la longueur du GOP). Dans cet exemple, nous avons un M=3 et un N=12.

L'estimation du mouvement en MPEG

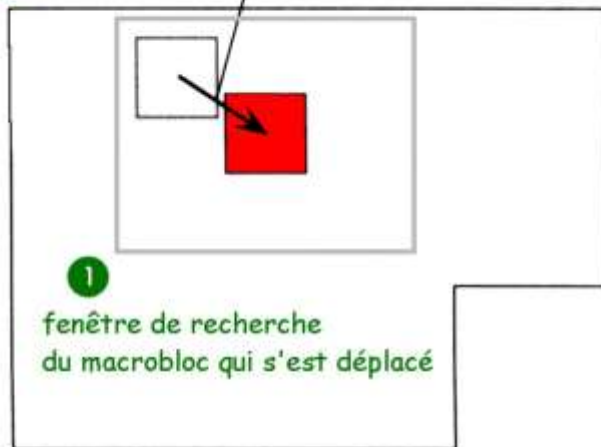
L'estimation de mouvement consiste à construire une image à partir d'une image précédente et d'informations relatives aux déplacements de ses composantes.

- l'estimation de mouvement est effectuée sur des macroblocs, typiquement constitué de l'association de 6 blocs : 4 pour la luminance et 2 pour la chrominance
- si on trouve sur 2 images successives deux macroblocs semblables, mais à des emplacements légèrement différents, on transmet une seule fois ce macrobloc
- pour reconstruire l'image actuelle il faudra indiquer sa nouvelle position sur l' image actuelle grâce à un vecteur mouvement

ancienne position
du macrobloc trouvée

3 calcul du vecteur de déplacement

2



1
fenêtre de recherche
du macrobloc qui s'est déplacé

Image i-1

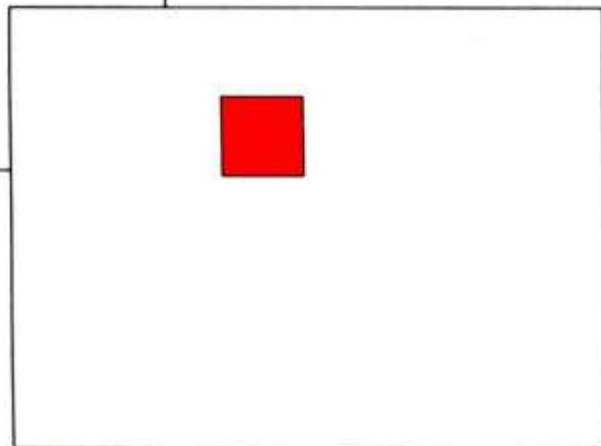
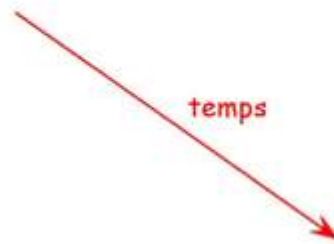
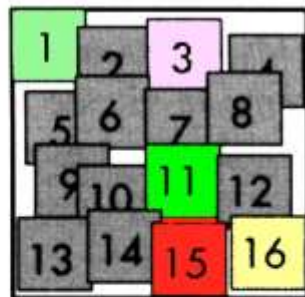


Image i



Macroblocks de l'image i-1
utilisés pour calculer l'image i

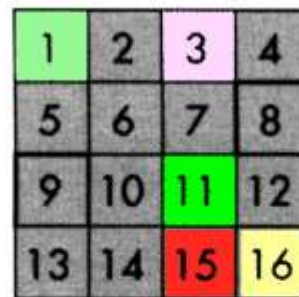


Image i prédite grâce aux
macroblocks de l'image i-1 et
aux vecteurs de mouvement

Correction des erreurs de prédiction

comparaison de l' image prédite avec la vraie nouvelle image pour générer des données d'erreur de prédiction



Résumé

- recherche des macroblocs semblables entre une image i et la précédente $i-1$
- calcul des vecteurs de mouvement caractérisant les déplacements des macroblocs
- construction d'une image prédite en utilisant ces vecteurs de mouvement
- comparaison de cette image prédite avec la vraie nouvelle image pour générer des données d'erreur de prédiction
- calcul des images B

Le principal inconvénient du système MPEG est que l'on n'a pas accès à chaque image de la séquence, en effet une image B dépendant en partie de données situées après elle, on ne pourra pas couper à cet endroit en montage.

Du fait de cet inconvénient, le MPEG est surtout un format de diffusion (DVD, VCD, Bouquets numériques etc....).

En diffusion, le GOP est de 12, en DVD de 15, en Betacam SX (acquisition broadcast) de 2.

Il est évident que plus le GOP est long, plus la compression est importante et plus les vecteurs de mouvement devront être précis.

