Les interfaces vidéo numériques : SDI & HD-SDI

1) Introduction 3
2) Le timecode 3
3) La définition Standard (SD) 5
A) Qu'entend on par définition standard ? 5
B) Norme SDI 5
a) Pal et NTSC 5
b) 4/3 et 16/9 5
c) Stucture d'échantillonnage 6
d) Dimensions utiles et incrustation de l'audio 7
e) Quantification 8
f) Débits 8
g) Type de liaison 8
C) SDTI 9
4) La haute définition (HD) 9
A) HDReady ou Full HD ? 10
B) Nomenclature 10
a) Résolution 10
b) Cadence 11
c) Balayage 13
d) Combinaisons de paramètres 14
C) Norme HD-SDI 14
5) La très haute définition (2K, 4K...) 15

1) Introduction

Les techniques de traitement multimédia et plus précisément celles de la vidéo se tournent de plus en plus vers la voie du numérique.

En effet, les outils numériques offrent des possibilités difficilement réalisables dans un fonctionnement 100% analogique. Toutefois, il ne faut pas perdre de vue que la captation et la restitution se font de manière analogique dans la majeure partie des cas.

Les oreilles et les yeux ne sont pas en mesure d'apprécier directement toutes les subtilités du langage binaire !

De la même manière que pour les signaux analogiques, il a fallu définir des normes pour standardiser la manipulation des signaux numériques.

Avant d'aborder ces notions, commençons par le timecode…

2) Le timecode

Le timecode (TC)est une référence temporelle liée à un contenu vidéo ou audio. C'est une suite de codes numériques générés à intervalles réguliers par un système de marquage de temps, pour la synchronisation et le marquage de matériaux enregistrés. Deux organismes, le SMPTE et l'EBU ont définit des normes respectivement pour le timecode en NTSC et en PAL.

Il s'exprime en heures, minutes, secondes et images ; il est enregistré parallèlement à chaque son et à chaque image. **HH :MM :SS :FF**

À chaque image, le timecode est composé de 4 informations :

- Le timecode s'exprime en heures, minutes, secondes et images. Dans l'absolu, le timecode commence à 00:00:00:00 (heures:minutes:secondes:images) et ne peut dépasser 23:59:59:24 (24 images pour l'EBU ou 29 pour le SMPTE).
- Le *userbits* : enregistré dans le code temporel à chaque image. Il est composé d'un « mot » de 32 bits, qui permet de mieux identifier l'enregistrement par un mot de 8 caractères (par exemple : la date du jour).
- Le contrôle d'erreurs (*controlbits*)
- Le contrôle de synchronisation (*synchronisation word*)

L'usage traditionnel en postproduction, est de débuter par 01:00:00:00 pour la première image d'un programme audiovisuel, par 02:00:00:00 pour la seconde, et ainsi de suite.

Afin de permettre la synchronisation, il faut y ajouter un *preroll* ou une amorce définie par un *start* et un décompte avant cette première image. Conventionnellement, l'amorce de mixage en France a une durée de 10 secondes. Le timecode de début (*start*) est donc à 00:59:50:00 (heure:minute:seconde:image) pour aboutir à l'adresse temporelle de la première image (01:00:00:00).

Les durées et timecodes spécifiques sont définis par les normes de diffusion (PAD) qui peuvent varier d'un diffuseur à l'autre.

On distingue plusieurs types de timecode en fonction de leur contexte d'utilisation :

- Audio : LTC, pour *Longitudinal Timecode* (« enregistré sur une piste audio ») Il est utilisé en vitesse de lecture normale et accélérée.
- Vidéo : VITC, pour *Vertical Interval Timecode* (incrusté dans l'image vidéo) une amélioration du LTC qui permet l'utilisation en lecture ralentie ainsi qu'à l'arrêt.
- MIDI : MTC, c'est un TC simplifié à l'image près (les *subframes* sont remplacés par des quarts d'image) utilisé depuis 1987 en musique, pour synchroniser des séquenceurs par exemple. C'est une passerelle entre le timecode et le MIDI
- Optique : Piste TC optique sur le film au cinéma (utilisée par le système DTS pour synchroniser un lecteur CD-ROM audio)
- Code Barre : C'est une façon d'inscrire visuellement des repères temporels sur une bande magnétique ou une image 35mm photochimique.

La problématique du timecode est que ce procédé ne peut assumer une synchronisation qu'à « l'image près ». C'était bien suffisant à l'origine mais de nos jours, ce système a bien des limites.

En numérique et entre plusieurs machines audio (machines linéaires, ordinateur, etc.), il est nécessaire d'assumer une synchronisation précise à l'échantillon près. Il faudra alors utiliser d'autres méthodes, bien plus complexes…

Exemple de mire de barre normalisée avec affichage du timecode.

3) La définition Standard (SD)

A) Qu'entend on par définition standard ?

L'appellation définition standard fait référence aux signaux numériques correspondant aux standards analogiques à 625lignes/50Hz ou 525 lignes/60Hz. Les variantes PAL et NTSC sont ainsi transcrites sous formes numériques.

Par abus de langage on parle encore de PAL ou de NTSC alors que les signaux numériques sont fabriqués à partir du signal composante (Y,Dr,Db)!

B) Norme SDI

a) Pal et NTSC

La norme SDI (Serial Digital Interface) définit la manière dont les informations numériques doivent être mises en forme pour être transmise entre différents équipements audiovisuels sur une seule liaison série. Elle précise de nombreux paramètres :

Deux modes sont admis : PAL et NTSC, leurs nombres de lignes et cadences correspondant aux anciens systèmes analogiques :

- 625 lignes / 50 trames par secondes
- 525 lignes / 59,94 ou 60 trames par secondes

Ces deux modes ne sont pas compatibles. Impossible de mélanger des images PAL et des images NTSC directement. Il faudra convertir l'une ou l'autre, c'est la fastidieuse étape de transcodage qui permet d'adapter les résolutions et les cadences de défilement. En fonction des matériels et techniques utilisés, le résultat qualitatif pourra être très différent !

Ces deux formats ayant des caractéristiques relativement différentes, certains équipement ne prendront en charge que l'un ou l'autre des formats. Il sera impossible de visionner le contenu d'une cassette NTSC sur un magnétoscope ne prenant en charge que le PAL.

b) 4/3 et 16/9

Ces rapports définissent le rapport des dimensions d'images largeur/hauteur. On peut exprimer ce ratio sous la forme du résultat de l'opération : 1,33 ou 1,77 pour les cas les plus courants.

Les images SDI sont traitées en 4/3, mais il est possible d'obtenir un format 16/9 artificiellement par anamorphose (déformation des pixels) de la même manière qu'au cinéma.

image au proportions correctes (2,35).

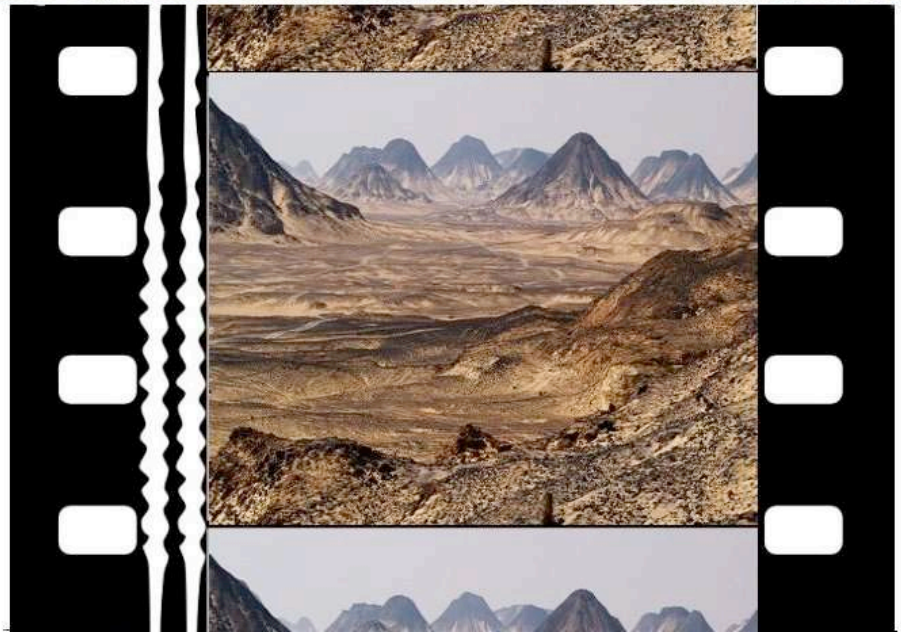


image enregistrée de manière anamorphosée(1,33).

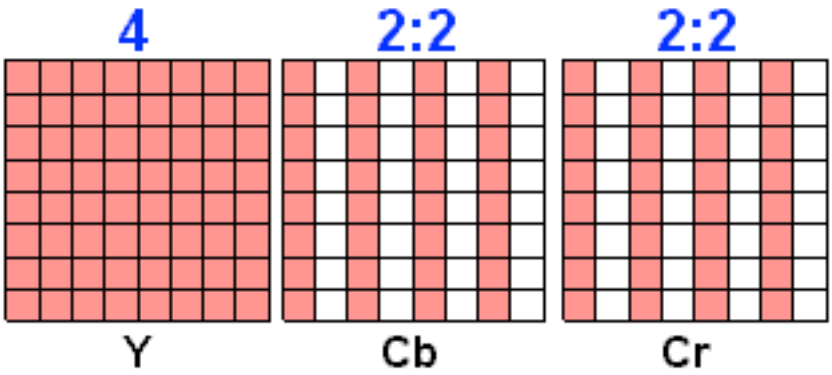
Les images sont donc enregistrées déformée, elles retrouveront les bonnes proportions lors de l'affichage. On prendra soin de préciser si l'enregistrement est en 4/3 ou en 16/9 pour permettre un affichage avec les bonnes proportions.

Il existe toutefois une variante en 16/9 natif dont la résolution est de 1024x576 pixels.

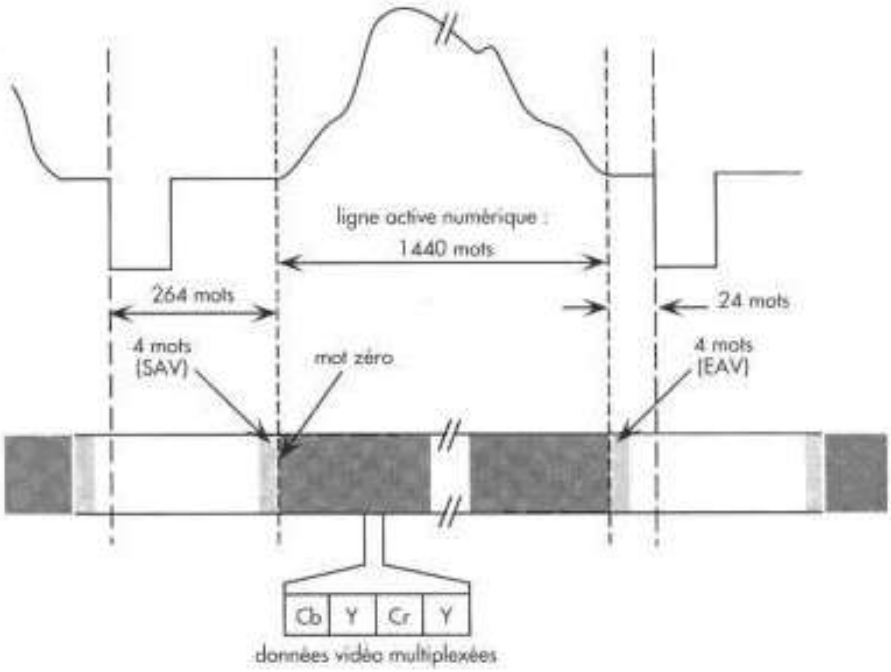
c) Stucture d'échantillonnage

La numérisation s'effectue avec une fréquence d'échantillonnage de 13,5 Mhz, avec une structure d'échantillonnage en 4 :2 :2 orthogonale.

Il y a donc autant de points d'information de luminance que de points d'information de chrominance au total (Cb+Cr).



Sur chaque ligne numérique on trouvera dans l'ordre une valeur Cb, une valeur Y, une valeur Cr et une autre valeur Y, et ainsi de suite.



En tout il y a 720 valeurs de luminances utiles par ligne de vidéo et 360 valeurs de Cb et 360 de Cr. Mais l'image est échantillonnée à 13,5 MHz pour la luminance et 6,75 MHz pour la chrominance ce qui nous donne 864 échantillons de luminance et 432 échantillons de chrominance.

Les valeurs « en trop » font donc partie de la suppression de ligne et peuvent contenir des informations de synchronisation, de time code ou de télétexte.

d) Dimensions utiles et incrustation de l'audio

Pour chaque image on a 625 (525) lignes, mais seulement 576 (480) lignes utiles pour la vidéo ce qui donne 25 lignes de suppression pour une trame et 24 lignes pour l'autre. Durant les suppressions lignes, le signal SDI peut contenir jusqu'à 16 canaux audio en 48 kHz encodé en 24 bits (audio embedded).

e) Quantification

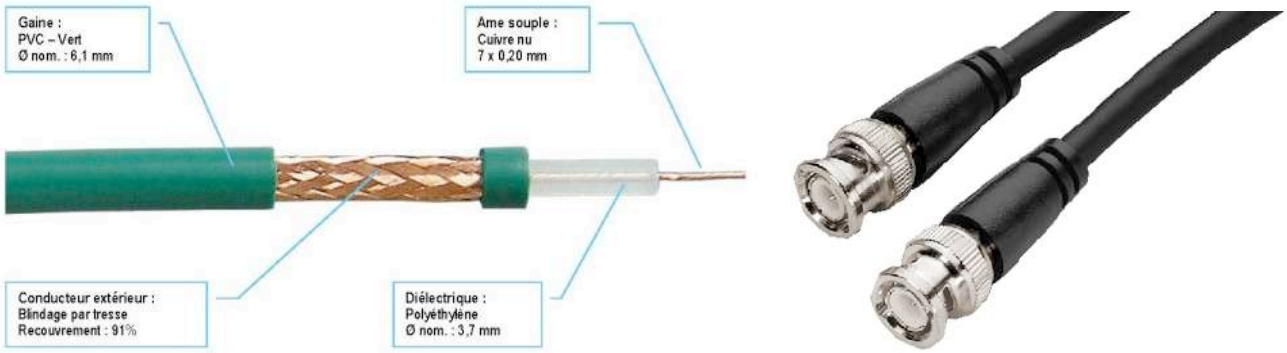
La quantification s'effectue généralement sur 10 bits (1024 valeurs). Cependant on interdit les valeurs inférieures à 64 et supérieurs à 940. Cela laisse une marge dans le cas de signaux analogiques incorrectement calibrés. Il est possible de fonctionner sur 8 bits, les valeurs évoluent alors entre 16 et 235.

f) Débits

Les vidéos dont la résolution utile est de 720x576 pixels ont donc un débit brut de 270 Mbits/s et un débit net de 207 Mbits/s (légèrement moins pour la version NTSC)

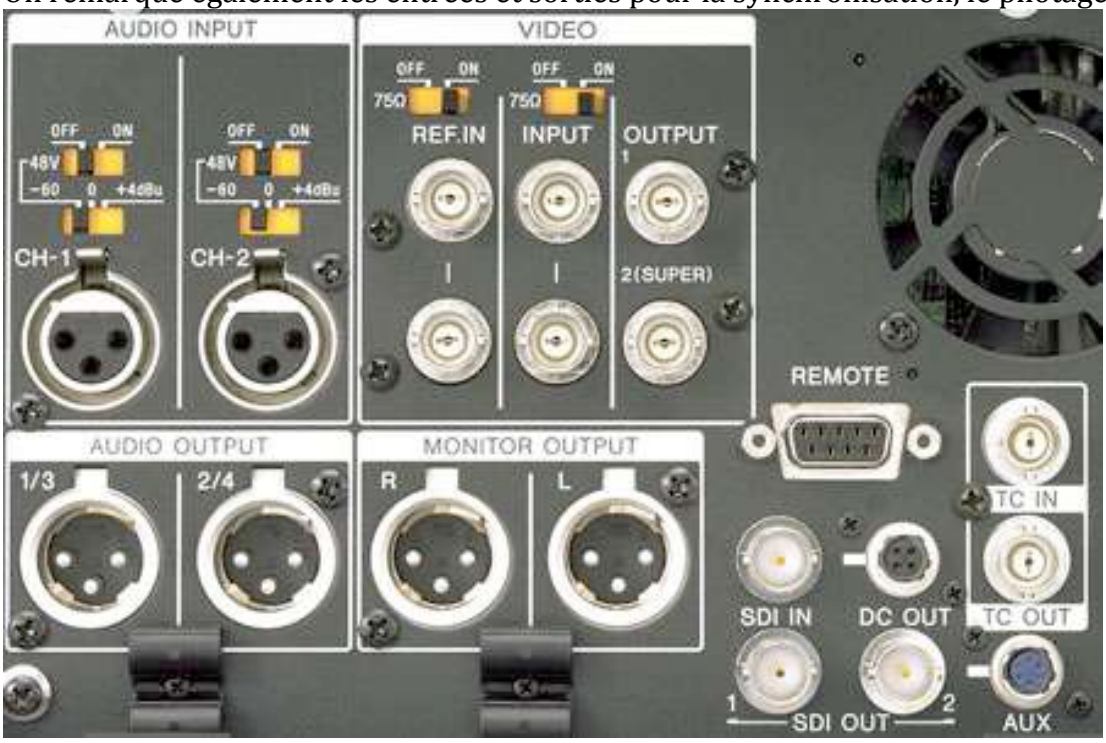
g) Type de liaison

Le câble utilisé pour relier des équipements en SDI est de type coaxial, terminé par des connecteurs BNC. Un dispositif de correction d'erreur garanti un taux d'erreur indétectable en dessous d'une distance de 300m.



On retrouve deux sorties SDI sur le panneau arrière de ces magnétoscopes.

On remarque également les entrées et sorties pour la synchronisation, le pilotage...



La norme SDI définit ainsi le fonctionnement de l'interface vidéo et audio numérique sans compression. C'est la liaison par excellence dans le monde professionnel pour la vidéo numérique en définition standard.

C) SDTI

Une variante de la norme SDI appelé SDTI (Serial Data Transport Interface) utilise le même principe et les mêmes caractéristiques à ceci près que les données peuvent être compressées et véhiculées par l'interface avec une mise en forme différente pour les contenus.

Il est alors nécessaire qu'émetteur et récepteur utilisent le même protocole de communication pour garantir le fonctionnement des échanges.

Ce système est notamment utilisé avec la famille DV lors de la transmission sur une liaison firewire utilisée en mode isochrone.

4) La haute définition (HD)

L'évolution constante des technologies multimédias à la fin des années 90 s'est trouvée confrontée à un problème crucial. En effet la norme SDI imposait un cadre trop stricte pour profiter des nouveaux équipements (écrans de grande dimension etc…).

Une image SD n'est que la transcription numérique des anciens standards hérités de la télévision. A l'époque les techniques étaient bien moins abouties qu'aujourd'hui.

On a donc décidé de mettre en place une nouvelle norme pour répondre à ce besoin. Le concept de départ consiste à multiplier les dimensions de l'image SD par 2 avec des proportions 16/9, plus proches des proportions cinéma.

A) HDReady ou Full HD ?

Tout équipement audiovisuel qui est apte à afficher un signal en Haute Définition est reconnaissable par la présence du **label HDReady**. Pour obtenir l'autorisation d'utiliser le label, le fabricant s'engage à respecter les caractéristiques suivantes :

- Disposer de 720 lignes d'affichage minimum.
- Disposer de connecteurs analogiques YPbPr ou numériques DVI ou HDMI.
- Accepter les formats vidéo en 1280 x 720 à 50 et 60 Hz en progressif (720p), ou en 1920 x 1080 à 50 et 60 Hz en entrelacé (1080i).
- Disposer d'entrées DVI ou HDMI compatibles avec la protection HDCP.

Full HD est une appellation commerciale qui signifie qu'un appareil (en général un téléviseur ou un moniteur vidéo) est capable d'afficher en définition native, une image constituée de 1 080 lignes, et de 1 920 pixels par ligne, et ce, quel que soit son mode de transmission (entrelacé ou progressif).

Il ne s'agit là que d'appellations commerciale. De manière générale la haute définition concerne toutes les vidéos dont la résolution est supérieure à celle de la définition standard.

On pourra évidemment distinguer différents cas de figures proposant des résolutions, cadences et modes de balayage différents.

B) Nomenclature

a) Résolution

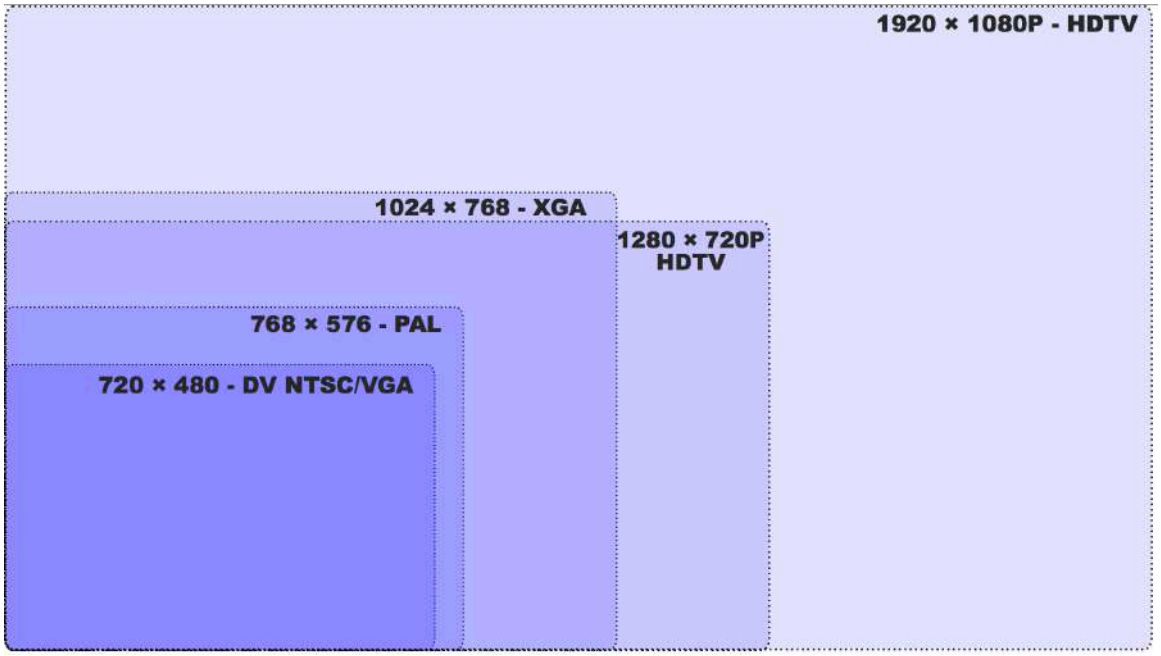
La résolution d'une image ou d'une vidéo décrit le nombre de pixels de sa partie utile. On a limité ces résolutions aux cas suivants :

- **1920x1080**
- **1280x720**
  De plus on a conservé la compatibilité avec les formats SD :
- **720x576 (PAL)**
- **720x480 (NTSC)**

On ne fait référence la plupart du temps qu'au nombre utile en sous-entendant le nombre de point par ligne.

De même, le nombre de ligne total n'a plus la même importance que précédemment. En effet, la synchronisation ne se fait plus de la même manière.

Il fallait éteindre le canon électron et lui donner le temps de se repositionner entre les lignes pour former une image sur le téléviseur cathodique, ce fonctionnement n'est plus tout à fait adapté pour un écran plat LCD ou plasma…



Représentation des différentes résolution SD et HD.

b) Cadence

En SD les images défilent à raison de 25, 30 et 29,97 images par secondes. En HD on distingue bien d'autres cadences:

- Les cadences de 25 et 30 images par secondes sont évidemment supportées.

- On trouvera également la cadence de 29,97 images par secondes. Cette valeur particulière appelée drop frame est un héritage du système NTSC. Lors du passage à la couleur, le décodage de la chrominance posait des problèmes. Les ingénieurs avaient trouvé comme solution de ralentir très légèrement la cadence des images dans un rapport de 1 pour 1000.

  En pratique, le time code incrémente directement de 3 images au lieu d'une à chaque changement de minutes :

  01:23:59:28
  01:23:59:29
  01:24:00:02 << on a sauté (droppé) des images (frame)
  01:24:00:03

  On enlève 2 images toutes minutes sauf pour les minutes multiples de dizaines où le time code défile correctement :

  01:29:59:28
  01:29:59:29
  01:30:00:00 << on vient de passer à la dizaine "30" minutes, pas de saut.
  01:30:00:01

  Il faut bien comprendre que les images ne sont pas supprimées littéralement, seul le time code est modifié.

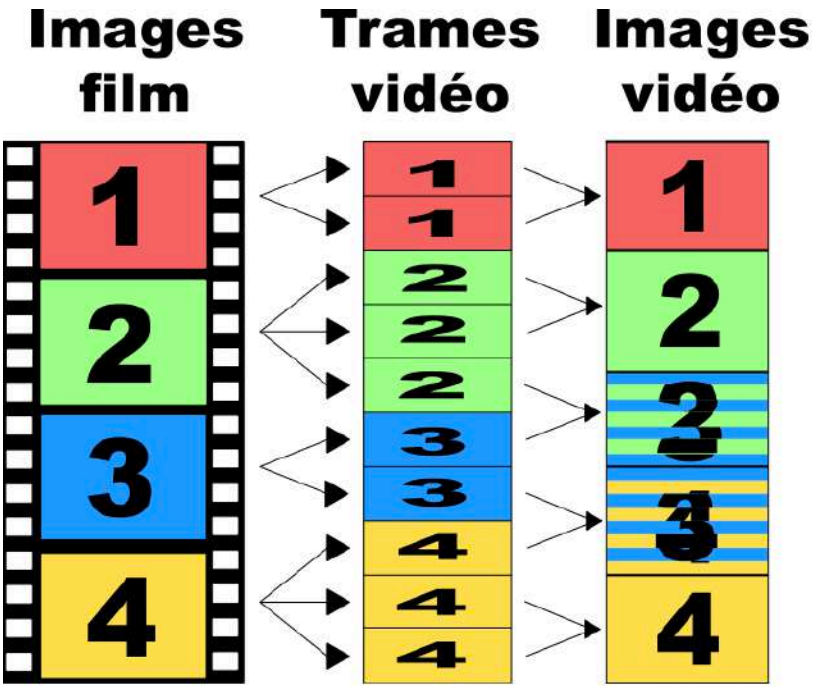
- Le cinéma utilise historiquement un défilement à raison de 24 images par seconde. La HD permet d'utiliser cette cadence de même que 23,976, correspondant au ralentissement de 1 pour 1000 (comme pour le NTSC couleur).

Un astucieux système appelé pulldown permet de passer facilement de 24 à 30 images par seconde en recombinant les trames entre elle pour recréer de nouvelles images.

On pourra ainsi lire une vidéo tournée initialement à 24 images par secondes et la recadencer à raison de 30 images par secondes.

Le Timecode sera impacté en conséquence, mais la durée du programme est identique. Il n'a pas été ralenti ni accéléré.

En revanche de nouvelles images seront crées à partir des images de départ



Fonctionnement du pulldown

Dans le cas du 23,976, on a tout simplement combiné ce procédé avec un ralentissement de 1 pour 1000 pour obtenir la meilleure compatibilité possible au niveau de la couleur. La durée du programme est conservée, mais le time code ne défile pas à la même vitesse.

Pour passer de 24 à 25 images par secondes, on accélère la cadence du programme de 4%. La durée du programme sera donc modifiée. Cela pose des problèmes évidents de synchronisation image et son. Si cette modification de vitesse est acceptable visuellement, on observe un changement de tonalité pour le son si on l'accélère(son plus aigu). Il faudra corriger ce défaut en faisant une harmonisation.

On trouve également en HD les cadences de 50, 59.94 et 60 images par secondes correspondant au doublement des cadences SD, c'est à dire à la fréquence des trames entrelacées.

Soit au total 8 possibilités: **23.976, 24, 25, 29.97, 30, 50, 59.94, 60**

c) Balayage

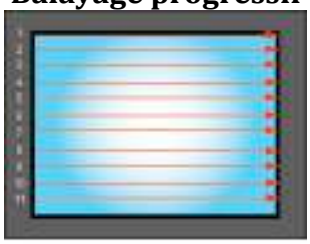
Deux modes de balayages sont disponibles le mode entrelacé et le mode progressif. Le mode entrelacé est un héritage de la télévision, le mode progressif correspond plutôt au fonctionnement des outils informatique ou du cinéma.

On indique par un **« i »** ou un **« p »** le mode de balayage choisi.

Balayage entrelacé

1er champ: Impair — 2ème champ: Pair — Une image complète en balayage entrelacé

Balayage progressif



Une image complète en balayage progressif

La mention **« psf »** (Progressive Segmented Frame) est utilisée lorsque l'on souhaite manipuler des contenus progressifs en utilisant du matériel et des formats entrelacés.

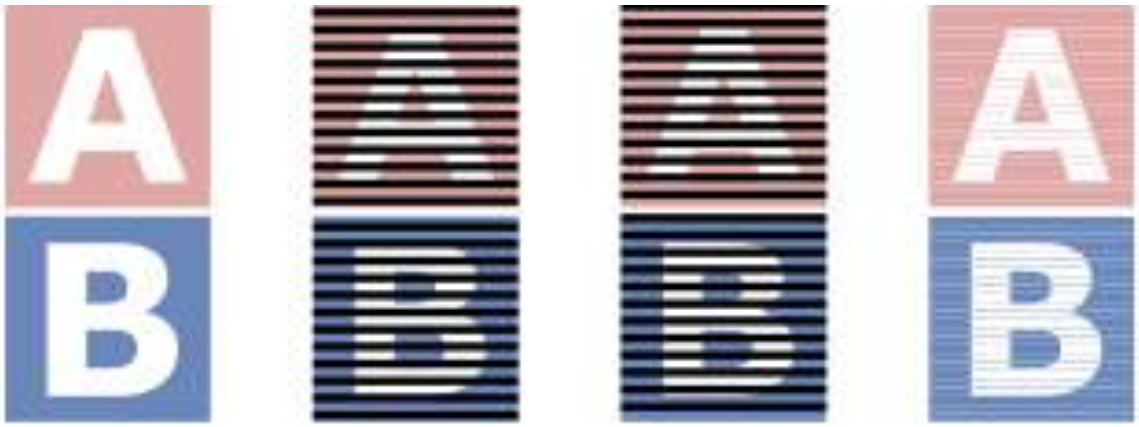


Image pleine -> première trame + deuxième trame -> image reconstituée

La cadence utilisée fera alors référence au nombre de trame ou bien au nombre d'images par secondes.

d) Combinaisons de paramètres

Une vidéo sera donc définie par ces 3 paramètres :

Résolutions	Cadences	Balayage
1920x1080	23.976	i
1280x720	24	p
720x576	25	psf
720x480	29.97	
	30	
	50	
	59.94	
	60	

- Ainsi, **720p25** correspond à une vidéo de résolution utile 1280x720 pixels, dont le balayage est progressif et affiché à raison de 25 images pleines par secondes.

- De même **1080i50** correspond à une vidéo de résolution utile 1920x1080 pixels, dont le balayage est entrelacé et affiché à raison de 50 trames par secondes.

- **576i50** permet de qualifier une image SD avec la nomenclature spécifique de la HD.

- **50i** et **25p** sont donc des modes comparables du point de vue de la quantité d'information, seul le mode de balayage diffère.

Toutes les combinaisons ne sont pas possibles :

- **25i** correspondrait par exemple à un programme cadencé à 25 trames par secondes avec un balayage entrelacé !

- **1080p50** et **1080p60** sont les formats les plus gourmands. La quantité d'information est gigantesque et rend la diffusion de ces formats très difficile. A l'avenir il y a fort à parier que cela deviendra la norme, mais aujourd'hui, ils pèsent encore un peu trop lourd.

C) Norme HD-SDI

La norme HD-SDI est donc l'évolution de la version SD permettant de manipuler ces vidéos en haute définition en tirant parti des améliorations technologiques.

- Il existe de nombreuses variantes possibles pour les vidéos en HD. On se limitera aux signaux **1080i et 720p pour le HD-SDI**. Ces deux formats sont comparables du point de vue qualité générale, ainsi qu'en terme de débit. Les pays fonctionnant historiquement en NTSC préfèrent aujourd'hui le 720p alors que les pays qui utilisaient le PAL ont plutôt tendance à travailler en 1080i.

- Les vidéos en 1080p pèsent deux fois plus lourds et sont aujourd'hui peu répandues.

- Les images seront toujours en **16/9** contrairement à la SD où le 4/3 était la norme.

- L'échantillonnage se fait à **74,25 Mhz** pour la luminance et à **37,125 Mhz** pour la chrominance (5,5 x plus que pour la SD). On trouvera soit la nomenclature **4 :2 :2 HD** ou bien **22 :11 :11** pour décrire la structure d'échantillonnage.

- La quantification se fait toujours sur 8 ou 10 bits comme dans le cas précédent. Ainsi le débit brut maximal total d'une liaison HD-SDI sera de **(74,25 + 37,125 + 37,125) = 3,375 x (22 + 11 + 11) x 10 = 1,485 Gbits/s.**

- On pourra au besoin utiliser des liaison double HD-SDI pour transporter des signaux en **22 :22 :22** avec un débit proche de **3Gbits/s** ou pour des images **3D relief.**

- Les mêmes liaisons coaxiales à embout BNC peuvent être utilisées que pour le signal SD. Seule la qualité du câble change. Cela facilite d'autant la mise en œuvre technique de cette évolution de la norme.

5) La très haute définition (2K, 4K…)

Aujourd'hui la Haute définition est la norme par excellence pour ce qui est de la vidéo numérique au niveau du grand public. Mais pour les mêmes raisons qui ont rendu la définition standard obsolète on travaille sur de nouveaux formats dont les résolutions sont plus grandes, notamment pour le cinéma numérique.

- 2K : 2048 x 1080 pixels (Comparable à la HD)
- 4K : 4096 x 2160 pixels
- 8K : 8192 x 4320 pixels
- …

Ces formats d'images amenés à se développer à l'avenir sont encore peu utilisés en dehors du cinéma numérique.

Ils demandent des débits et des espaces de stockages considérables, sans même parler des problèmes liés au traitement de cette énorme quantité d'information.

Pour du 4K non compressé: 4096x2160 pixels par image quantifiés sur 12 bits donnent une image d'environ 40 Mo. A raison de 50 images par secondes, on trouve un débit de l'ordre de 16Gb/s. Un film de 90 minutes pèserait alors 10To…pour la vidéo utile uniquement ! (multiplier par 4 pour du 8K, par 16 pour du 16K…)

De plus, il faudra disposer d'une image d'au moins 100 pouces de diagonale (2,5m) pour profiter de ces nouvelles technologies dans de bonnes conditions… En dessous de cette taille, l'œil ne ferait pas vraiment la différence avec de la HD standard.

On peut néanmoins réfléchir à l'utilisation de ces formats pour une utilisation détournée. La stéréoscopie fonctionne en projetant une image pour chaque œil. Des systèmes plus évolués permettent l'utilisation simultanée de 4 ou 16 images pour tenir compte des déplacements éventuels du spectateur (mouvement de tête, etc). On peut alors utiliser une image 8K qui est en fait composée de plusieurs sources de résolution inférieure, à la manière de la 3D en side by side.