

# Digitalització, emmagatzematge i transmissió d'àudio i vídeo

Alex Ribelles García

PID\_00176922



Universitat Oberta  
de Catalunya

[www.uoc.edu](http://www.uoc.edu)



*Els textos i imatges publicats en aquesta obra estan subjectes –llevat que s'indiqui el contrari– a una llicència de Reconeixement-NoComercial-SenseObraDerivada (BY-NC-ND) v.3.0 Espanya de Creative Commons. Podeu copiar-los, distribuir-los i transmetre'ls públicament sempre que en citeu l'autor i la font (FUOC. Fundació per a la Universitat Oberta de Catalunya), no en feu un ús comercial i no en feu obra derivada. La llicència completa es pot consultar a <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.ca>*

# Índex

|   |    |
|---|----|
| <b>Introducció</b> .....  | 5  |
| <b>Objectius</b> .....  | 6  |
| <b>1. Introducció</b> .....   | 7  |
| 1.1. Contingut del mòdul .....  | 7  |
| <b>2. Codificació i compressió de vídeo i àudio digital</b> .....                 | 9  |
| 2.1. Captació d'àudio i vídeo .....   | 10 |
| 2.2. El format intermedi comú d'imatge (CIF) .....                                | 12 |
| 2.2.1. Espai de color .....   | 12 |
| 2.2.2. Mostreig del vídeo i submostreig de color .....                            | 14 |
| 2.2.3. Respecte a mides estandarditzades en CIF .....                             | 15 |
| 2.3. Compressió de vídeo digital .....  | 16 |
| 2.4. Còdecs de vídeo digital amb pèrdues més habituals .....                      | 22 |
| 2.4.1. Emmagatzematge .....   | 22 |
| 2.4.2. Transmissió en temps real .....  | 22 |
| 2.4.3. El cas intermedi: vídeo sota demanda .....                                 | 23 |
| 2.4.4. El receptor .....  | 23 |
| 2.4.5. Exemples d'estàndards .....  | 23 |
| 2.5. Compressió d'àudio digital .....   | 25 |
| 2.6. Diferenciant contenidor i còdec .....  | 27 |
| <b>3. L'estàndard MPEG-2</b> .....  | 31 |
| 3.1. Altres estàndards relacionats: JPEG/JPEG2000, MPEG-1,<br>H.261 i H.263 ..... | 31 |
| 3.2. Què és l'MPEG-2 .....  | 32 |
| 3.3. Codificació i compressió en MPEG-2 .....                                     | 33 |
| 3.3.1. Els fluxos elementals de dades .....                                       | 36 |
| 3.3.2. Emmagatzematge i transmissió .....   | 39 |
| 3.3.3. Perfils i nivells de MPEG-2 .....  | 39 |
| <b>4. L'estàndard MPEG-4 i H.264/AVC</b> .....                                    | 42 |
| 4.1. L'estàndard i les seves patents .....  | 43 |
| 4.2. H.264 o MPEG-4 Part 10 .....   | 43 |
| 4.3. Emmagatzematge .....   | 44 |
| 4.4. Transmissió .....  | 44 |
| 4.5. Perfils i nivells .....  | 45 |
| 4.6. Utilització de l'H.264 en la indústria .....                                 | 46 |
| <b>5. Transmissió de vídeo en una xarxa IP (streaming)</b> .....                  | 48 |

---

|              |  |           |
|--------------|--|-----------|
| 5.1.         | Les xarxes IP .....                                      | 48        |
| 5.1.1.       | Protocol de xarxa ( <i>Internet protocol</i> , IP) ..... | 49        |
| 5.1.2.       | Protocols de transport (UDP, TCP i SCTP) .....           | 50        |
| 5.2.         | Tècniques de multidifusió .....                          | 52        |
| 5.3.         | Protocols de transmissió en temps real .....             | 54        |
| 5.3.1.       | Protocols de temps real (RTP, RTCP i RTSP) .....         | 55        |
| 5.3.2.       | Altres protocols de transmissió en temps real .....      | 56        |
| <b>6.</b>    | <b>MPEG-2 TS (<i>transport stream</i>)</b> .....         | <b>58</b> |
| 6.1.         | Paquets i taules del <i>transport stream</i> .....       | 58        |
| <b>Resum</b> | .....  | <b>60</b> |

## Introducció

En el mòdul d'introducció s'ha caracteritzat la imatge estàtica, i s'ha posat l'accent en el procés de digitalització per a capturar-la. Tot i que s'ha anat més enllà explicant el procés posterior de codificació JPEG, ara és el moment de centrar-se en la captació, la codificació i la transmissió de contingut d'àudio i vídeo, informacions més complexes que obligaran a introduir mètodes especialitzats de compressió nous.

Els diferents formats audiovisuals que presentarem es poden dividir segons criteris diferents. Així, parlarem de formats de compressió amb pèrdues o sense si comporten cert nivell de degradació de la informació original o no, i veurem com es pot detectar aquesta degradació. Una altra classificació que veurem depèn de l'objectiu: ser transmès en temps real o no, o senzillament ser emmagatzemat.

Explicarem la diferència entre *còdec* i *contenidor*, i la il·lustrarem amb exemples actuals, recalçant especialment els més estesos avui dia.

Finalment, veurem dos exemples de plataformes de distribució actual de contingut multimèdia: Internet i la televisió digital terrestre.

Per a il·lustrar aquests estàndards, tant en aquest mòdul com en les pràctiques associades, utilitzarem programari propietari (Adobe Photoshop) i eines de codi lliure (MPEG Streamclip i VLC).



## Objectius

Els objectius que es pretenen aconseguir amb l'estudi d'aquest mòdul són els següents:

- 1.** Identificar les característiques dels còdecs més habituals de vídeo.
- 2.** Capacitar per a decidir la manera més adequada de publicar un contingut digital sota demanda.
- 3.** Identificar i diferenciar els contenidors multimèdia més freqüents.
- 4.** Capacitar per a establir les característiques oportunes del contingut digital per a emmagatzemar-lo i accedir-hi.
- 5.** Identificar les dificultats de transmissió multimèdia en xarxes IP.
- 6.** Capacitar per a decidir la via de transmissió més adequada per a un contingut digital.

## 1. Introducció

La digitalització dels canals de comunicació i de la informació mateix ha suposat la denominada **convergència digital** i, de retruc, un intercanvi fàcil de tot tipus d'informació digital. El vídeo, sens dubte, a causa de les seves necessitats específiques de volum de dades i cadència va ser dels últims que s'hi va integrar, però un cop fet el pas és a tot arreu, en el nostre dia a dia, com un servei més. El procés d'integració ha tingut, doncs, dos elements paral·lels: els estàndards de format digital de vídeo i la implementació de sistemes digitals de transmissió.

La transferència de vídeo i àudio digital en temps real (*streaming*) ja existia en comunicacions mòbils des dels anys setanta, si entenem com a tals els sistemes dedicats experimentals sense objectiu comercial. Avui, són bàsicament serveis de valor afegit la qualitat dels quals està adaptada a la capacitat i tipus de receptor-client (telèfon mòbil, tauletes, punts d'informació, etc.) gràcies a estàndards més o menys oberts que possibiliten donar aquest servei al vol. Els tractarem en aquest mòdul.

### 1.1. Contingut del mòdul

En els apartats següents descriurem els components d'un sistema de vídeo digital, i remarcarem especialment els estàndards, presentarem breument la primera fase d'un sistema digital, la captura de vídeo i la compressió.

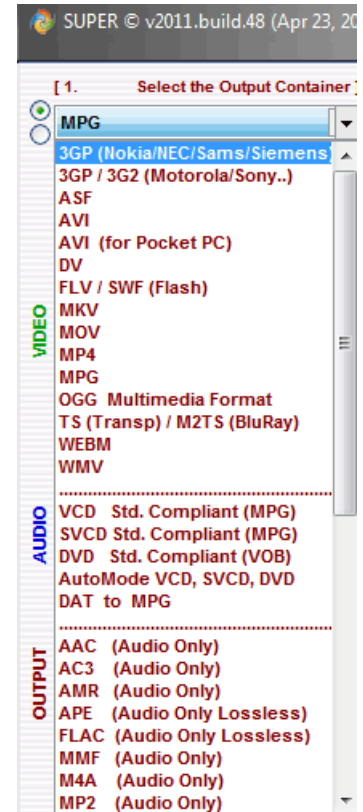
Dels estàndards actuals, veurem els més acceptats per la indústria, entre els quals destaquen clarament els especificats pel grup MPEG (Moving Pictures Experts Group) de la ISO (International Organization for Standardization), sense deixar de banda alguns altres de gran abast com FLV, OGG i 3GP. Una vista a la llista d'opcions de format d'exportació d'una aplicació de vídeo qualsevol és un reflex fidel dels formats de vídeo més utilitzats actualment.

Avui dia els sistemes de vídeo digital, com la televisió digital per cable, terrestre i de satèl·lit o el format DVD utilitzen l'estàndard MPEG-2, però la indústria ha adoptat clarament l'estàndard MPEG-4 per a desplegar sistemes de vídeo digital nous. En particular els sistemes de transmissió de dades basats en TCP/IP, com IPTV o 3G, sembla que adopten majoritàriament l'estàndard H.264/AVC, que és una part de l'estàndard MPEG-4 estandarditzat en conjunt amb la ITU-T. No obstant això, i per herència d'Internet, alguns estàndards de sobretaula altament eficients com el Flash tenen una vida per davant, i basen el seu funcionament en H.264.

L'estàndard MPEG-2 és presentat en l'apartat 3, mentre que el MPEG-4 i derivats es presenten en l'apartat 4. Com veurem, la força dels estàndards MPEG és la clara especificació de la manera de codificar i descodificar el contingut multimèdia.

Respecte a la transmissió de les dades generades, hi ha una solució per a cada tipus de mitjà possible que n'asseguri la **distribució**, i igualment hi ha una altra solució per a cada mitjà en què es vulguin **emmagatzemar**. Distribuir i emmagatzemar són les dues destinacions habituals d'aquesta informació audiovisual. Curiosament, per a xarxes IPTV les dues solucions són necessàries.

Com que no sempre s'utilitzen els estàndards MPEG-2 i MPEG-4 per a emmagatzemar i transmetre la informació, aquests mecanismes es presenten en apartats separats dels dels estàndards: l'apartat 5 mostra els mecanismes actuals d'emmagatzematge de vídeo, mentre que l'apartat 6 mostra els mecanismes de transmissió en una xarxa IP. Hi ha un cas particular de transmissió de MPEG: és el *transport stream*, comunament referenciat per MPEG-2 TS o MPEG-TS, molt utilitzat a les xarxes actuals de difusió digital i que presentem en l'últim apartat.



Els diferents formats de vídeo i àudio més habituals per a emmagatzematge o transmissió seleccionables en una aplicació com Super©



## 2. Codificació i compressió de vídeo i àudio digital

Codificació i compressió són conceptes tan lligats que és freqüent confondre'ls. La **codificació** és el procés en el qual es representa una informació de vídeo o àudio digital de manera diferent de l'original, sense perdre cap informació en el procés.

La majoria de codificacions que hi ha tenen com a objectiu comprimir la informació o ofuscar-la per a fer-la més segura (o tots dos). En aquesta assignatura només ens centrarem en el primer.

Així doncs, una codificació implica usualment una **compressió**, una reducció de la mida final del fitxer, sense perdre cap informació. Aquest comportament es coneix també com a **compressió sense pèrdues**. Lamentablement no té ni de lluny bons resultats amb vídeo o àudio, per això cal una tècnica complementària: la **compressió amb pèrdues**.

En general, la **compressió** fa referència a la reducció de la informació amb pèrdua d'aquesta o sense, com a mínim sense reduir-ne excessivament la qualitat.

Parlem llavors de compressió sense pèrdues o amb pèrdues. Aquesta última és la més utilitzada, ja que redueix molt més la mida de la informació que la compressió sense pèrdues. Si bé un canvi de la codificació sol pot aconseguir certa reducció del volum de dades, és sens dubte la compressió la clau de tot sistema d'emmagatzematge o distribució audiovisual. La combinació dels dos és la tècnica habitual de treball de tots els estàndards audiovisuals actuals.

Tot això és possible si es treballa amb informació digital: els sistemes d'àudio i vídeo actuals tenen una primera etapa de conversió d'analògic a digital, ja que la realitat que ens envolta és analògica per naturalesa.

Si aquesta primera etapa es fa amb equips professionals, el senyal involucrat no es comprimeix, sinó que només es codifica digitalment per a mantenir-ne la puresa en un format digital cru (*raw*) i ser processat més endavant per a comprimir-lo segons el que calgui.

### El format ZIP

Un exemple de codificació és el format ZIP, que sense perdre cap dada és capaç de representar la mateixa informació original, amb el benefici afegit que redueix la necessitat d'emmagatzematge d'aquesta informació.

### Vegeu també

En el subapartat 2.1 veurem com es fa aquesta conversió d'analògic a digital.

### Vegeu també

Alguns d'aquests formats per a mantenir la puresa del senyal es tracten en el subapartat 2.2.

En canvi, en entorns industrials i semiprofessionals la captura i la compressió van juntes, com és el cas del maquinari i programari que fa servir l'estàndard MPEG.

**Vegeu també**

L'estàndard MPEG es tracta en el subapartat 2.3 com una aproximació general a la compressió de vídeo.

## 2.1. Captació d'àudio i vídeo

La captació i codificació d'àudio és relativament senzilla comparada amb la de vídeo. Bàsicament, es mostreja el senyal d'àudio a una freqüència sempre dues vegades superior (com a mínim) a la freqüència màxima que pugui haver-hi en el senyal d'àudio.

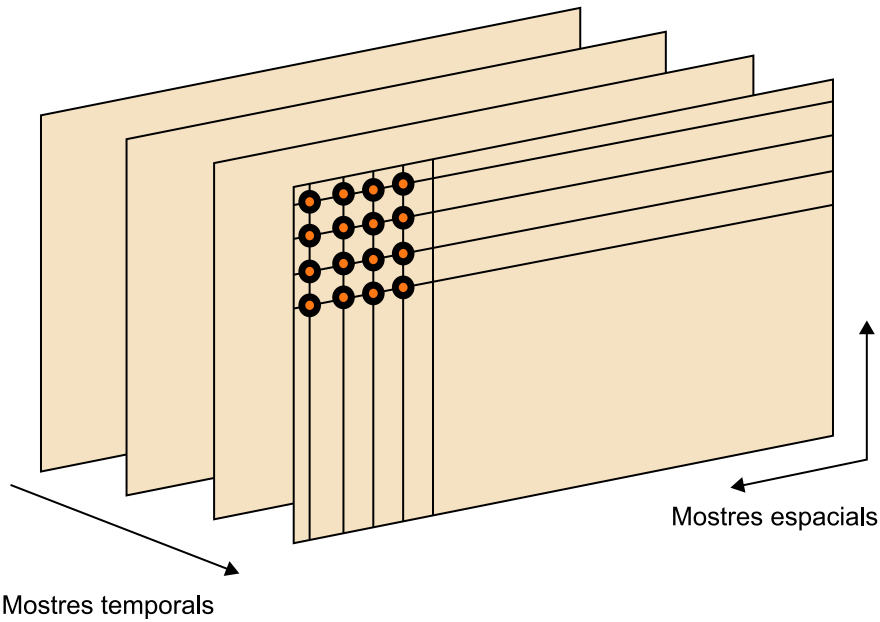
Així, si l'orella humana és capaç de captar freqüències d'àudio de fins a 16 kilohertz (kHz), és molt habitual que es mostregi a 44 o 48 kHz, és a dir, 44.000 o 48.000 mostres per segon.

Cada mostra representa un valor real de l'amplitud del senyal d'àudio, és a dir, té decimals, cosa que en dificulta la codificació en bits. Per a això, es converteixen aquests valors amb decimals a nombres enters, usualment en el rang de 0 a 255 (8 bits) o millor encara de 0 a 65.535 (16 bits). El format PCM d'àudio digital és el format cru per excel·lència en entorns professionals: no hi ha cap pèrdua en aquesta codificació, però això suposa una gran quantitat de bits per a emmagatzemar (CD).



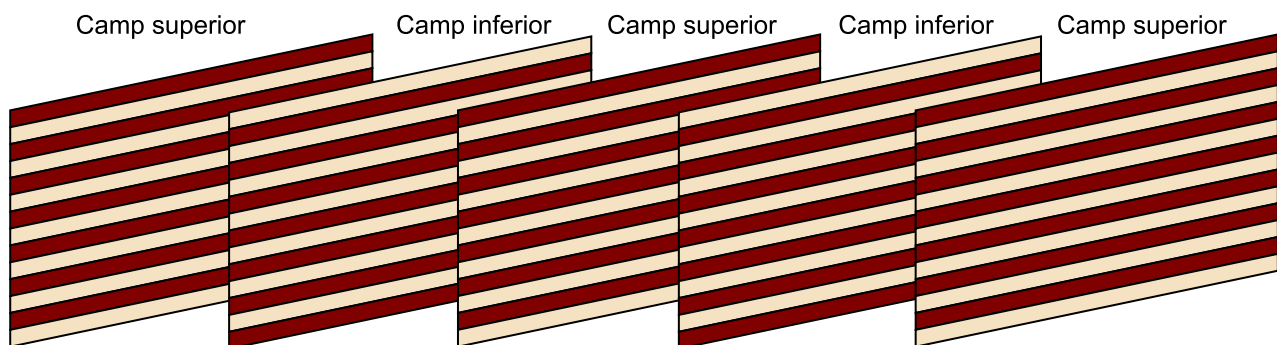
Mostreig d'un senyal analògic per a digitalitzar-lo

Respecte al vídeo, la representació natural d'un vídeo digital és una matriu tridimensional de píxels. Dues dimensions (horitzontal i vertical) representen un quadre o imatge i pertanyen al domini espacial (ample i alt), i la tercera dimensió és el temps.

Mostreig en tres dimensions del vídeo:  $x$ ,  $y$ ,  $t$ 

Aquesta representació facilita entendre que s'hauran de prendre mostres de cada quadre (moure's en el plànol espacial, **mostreig espacial**) i cada cert temps (moure's en el plànol temporal, **mostreig temporal**). Tal com es planteja, seria una codificació crua, però veurem a continuació que per certes raons reduïrem sempre el nombre de mostres que farem. Per això aquest format de codificació es denomina CIF (*common intermediate format*), perquè es troba en un punt mitjà entre la captura crua i la comprimida.

La forma natural de fer les mostres del senyal de vídeo és coneguda com a **mostreig progressiu**, és a dir, tal com planteja la imatge anterior interpretem el vídeo com una sèrie de quadres consecutius independents i mostregem espacialment l'un darrere l'altre. Però a causa dels requeriments de velocitat de mostreig necessaris, com que la informació de cada quadre és molt similar a la de l'anterior, podem mostrejar primer les línies parelles del primer quadre i després les imparelles d'aquest, les parelles del segon quadre, les imparelles d'aquest, etc.



Estructura de camps d'un vídeo entrellaçat

Aquest tipus de mostreig es denomina **mostreig entrellaçat** i permet captar la mateixa informació que el progressiu però amb un equip tècnicament més senzill. Pot semblar dràstic, però té alguns beneficis, motiu pel qual es fa servir habitualment i té uns efectes poc perceptibles en el moviment dels objectes. Aquesta opció està implementada en tot el programari o maquinari de captura de vídeo que hi ha avui dia. L'ull entrenat pot notar els efectes, especialment en zones àmplies d'imatge molt lluminoses, però en un monitor progressiu com qualsevol pantalla plana queden camuflats completament.

En resum, de cada quadre original passem a **dos camps** amb la meitat de resolució vertical (però la mateixa resolució horitzontal que l'original), entenent que cada camp és una imatge amb la meitat d'informació que el quadre original. La velocitat en camps/segon és el doble que quadres/segon, evidentment, però el flux final de dades generades és el mateix.

Aquest procediment no només redueix els requeriments tècnics de l'equip de captura, sinó que a més millora l'efecte de parpelleig que podria haver-hi reproduint un senyal de 25 quadres/segon respecte a un senyal de la meitat de definició però a 50 camps/segon. Encara que aquest avantatge va desaparèixer quan el receptor té un monitor progressiu: hi ha una memòria intermèdia que, després de rebre tots dos camps d'un quadre, els pot recompondre amb tota la definició.

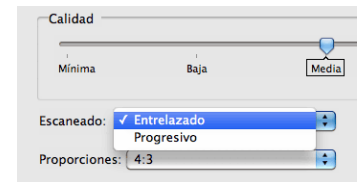
La norma PAL utilitzava aquest mecanisme en les seves càmeres per a capturar i transmetre per un canal de freqüències estret. Actualment, en plena era digital, segueix aplicant-se als senyals estàndards de la TDT i fins i tot als senyals HDTV d'alta definició 1080i.

## 2.2. El format intermedi comú d'imatge (CIF)

Com ja s'ha dit, una **codificació crua** és, en essència, una seqüència temporal de quadres, o de camps, és a dir, de la meitat de les línies d'un quadre quan hi ha entrellaçat. Cal determinar el format de cada quadre i el mecanisme de mostreig per a tenir una especificació completa del format CIF que acordin emissor i receptor.

### 2.2.1. Espai de color

Un quadre és una matriu rectangular de píxels que representa una imatge. El píxel està representat i localitzat per les seves dues coordenades X i Y, però a més té tres valors sencers que representen el color d'aquest en un espai de colors. Els espais de color més coneguts són RGB, YCbCr i, especialment, YUV.

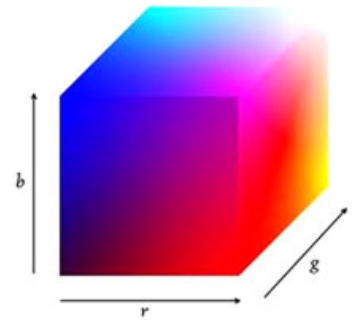


Opcions de captura de vídeo habituals

### Videoconferència

En una videoconferència (Skype, iChat, etc.), emissor i receptor han de tenir definit clarament el CIF: la mida del quadre d'imatge que es presenta, el nombre de colors que es reproduïxen, la cadència de quadres per segon, etc.

RGB representa un nivell de color per als colors bàsics vermell (R), verd (G) i blau (B). És un espai de color intuïtiu en el qual qualsevol color pot representar-se com una combinació d'aquests tres colors. Atès que la gamma és infinita en qualsevol, usualment es redueix a 256 valors per color, del 0 al 255. Així,  $R = 0$  és el vermell negre; i  $R = 255$ , el vermell més vívid i pur. Cal especificar, no obstant això, que aquesta codificació és usada per pura conveniència.



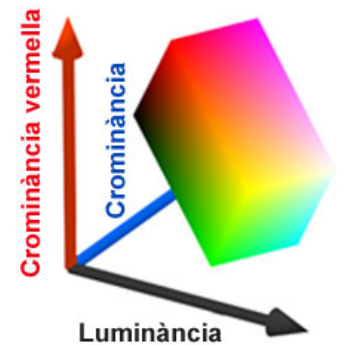
L'espai de color RGB

### Activitat 1

Amb l'Adobe Photoshop, obriu una imatge en color. A *Ventana/Canales* es mostren els components R, G i B. Confirmeu que allà on predomina un dels tres colors, el component corresponent té un valor més clar.



Els altres dos espais de color, en els quals es pot definir també qualsevol color amb tres valors, són equivalents a RGB, només que plantegen cada color com una combinació de blanc i negre i dos components de color. Així, a YCbCr el valor Y és el valor de luminància, Cb és el valor de crominància blau i Cr és el de crominància vermella. El valor de crominància verd pot deduir-se dels tres anteriors. L'espai YUV també segueix la mateixa filosofia.



L'espai de color YCbCr

La raó per la qual és més habitual utilitzar aquests dos últims espais és deguda a les característiques de l'ull: és menys sensible al color que a la lluentor o luminància. La comprovació és senzilla: posem un fil de color vermell a 5 metres d'un observador. Usualment detectarà el fil, però no el color. Sembla que el color era poc rellevant per a la supervivència dels nostres ancestres, com a mínim menys rellevant que la intensitat de llum.

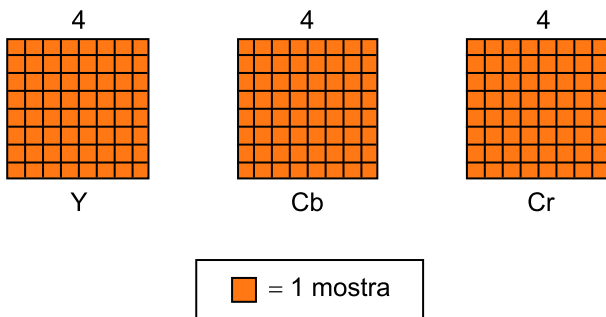
Sabent, llavors, que es pot representar cada quadre d'imatge pels seus tres quadres equivalents Y, Cb i Cr o Y, U i V, i sabent que la informació visual sobretot és al quadre Y, es podrà comprimir amb poques pèrdues (és a dir, una compressió més aviat suau) el quadre Y i comprimir amb més pèrdues (més compressió) els quadres Cb i Cr o U i V.

Així passa en l'estàndard JPEG de compressió d'imatge fixa. La primera etapa del procés és convertir la imatge des de l'espai de color RGB (tal com es genera en el sensor de càmera) a l'espai YCbCr.

### 2.2.2. Mostreig del vídeo i submostreig de color

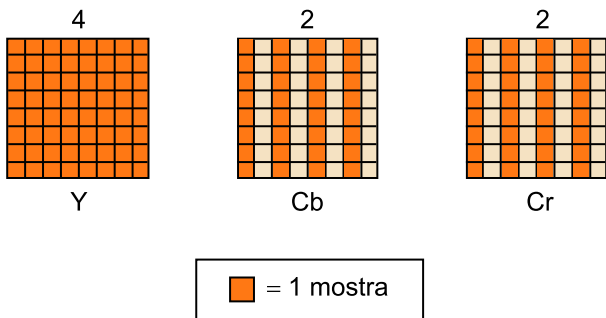
Aquesta compressió, que s'aplica més intensament al color, comença en el mer procés de mostreig de cada quadre. Es fa un mostreig més dens en el seu equivalent Y, i menys dens en els seus equivalents de color sense afectar seriosament la qualitat final. La sistemàtica de mostreig està estandarditzada en la família MPEG, basada en YCbCr o YUV segons el cas:

a) **4:4:4**: els tres components (Y, Cb i Cr) tenen la mateixa resolució i hi ha una mostra de cada component en cada píxel. Només és per a entorns professionals en els quals no es vol compressió (HDCAM, MPEG-2, H.264).



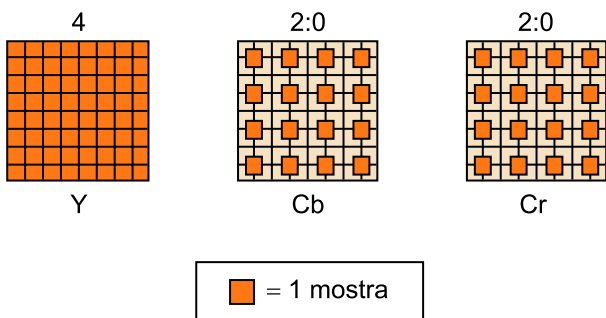
Mostres de cada component en el cas 4:4:4

b) **4:2:2**: conegut també com a *YUY2*, té 2 components de crominància per cada 4 de luminància (Digital Betacam, DVCPRO 50, MPEG-2, H.264)

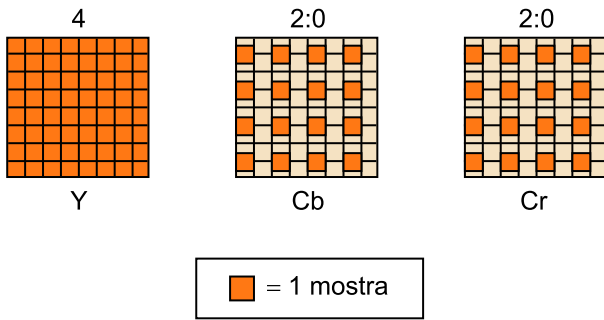


Mostres de cada component en el cas 4:2:2

c) **4:2:0**: conegut també com a *YV12*, té 1 component de crominància per cada 4 de luminància (DVD, MPEG-1, MPEG-2, MPEG-4).



Mostreig de cada component en el cas MPEG-1 (4:2:0)



Mostreig de cada component en el cas MPEG-2 (4:2:0)

En resum, els tres patrons tenen una distribució dels píxels que es mostregen en cada cas ben especificada. En el cas 4:4:4, es mostregen tots i cadascun dels píxels dels tres components de cada quadre. En el cas 4:2:2, es mostregen tots els píxels del component de luminància Y i la meitat de cada component de color. Finalment, en el cas 4:2:0, es mostregen tots els píxels del component de luminància Y, i només 1 de cada 4 en els components de color, variant segons la norma.

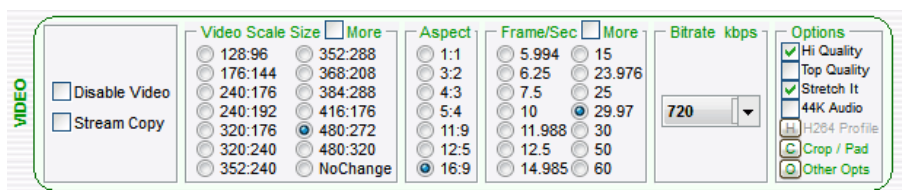
En tots, cada mostra aconseguida es representa posteriorment amb 8 o 10 bits, segons la precisió de la mostra que es vulgui.

### 2.2.3. Respecte a mides estandarditzades en CIF

Definida la sistemàtica de mostreig, queda per a definir la mida final que tindrà el vídeo digitalitzat resultant. També el format CIF inclou aquesta especificació:

Taula. Mides recomanades per al format CIF

| Format CIF         | Resolució de vídeo (ample × alt) |
|--------------------|----------------------------------|
| Sub QCIF           | 128 × 96                         |
| Quarter CIF (QCIF) | 176 × 144                        |
| CIF                | 352 × 288                        |
| DCIF               | 528 × 384                        |
| 4CIF               | 704 × 480                        |
| 16CIF              | 1.408 × 1.152                    |



Algunes aplicacions de codificació de vídeo expressen clarament els formats d'imatge en els quals es pot exportar, entre els quals hi ha els CIF (a la imatge, Super©)

Un format molt utilitzat en l'ambient de televisió es troba especificat en la recomanació ITU-R BT.601-5, que podeu veure en la imatge anterior, i té dues variants: NTSC i PAL.

Taula. Mides recomanades per a les variants NTSC i PAL del format ITU-R BT.601-5

| Paràmetre                         | NTSC<br>30 Hz | PAL/SE-<br>CAM<br>25 Hz |
|-----------------------------------|---------------|-------------------------|
| Camps per segon                   | 60            | 50                      |
| Línies per quadre complet         | 525           | 625                     |
| Mostres de Y per línia            | 858           | 864                     |
| Mostres de Cr i Cn per línia      | 429           | 432                     |
| Bits per mostra                   | 8             | 8                       |
| Flux de bits final                | 216 Mbps      | 216 Mbps                |
| Línies actives per quadre         | 480           | 576                     |
| Mostres actives per línia (Y)     | 720           | 720                     |
| Mostres actives per línia (Cr,Cb) | 360           | 360                     |

Fixeu-vos que calen 216 Mbits per segon per a emmagatzemar un vídeo codificat en aquest estàndard, o, de manera equivalent, un canal de 216 Mbps per a transmetre'l. Això supera àmpliament la disponibilitat actual de les xarxes de dades (Internet, 3G, etc.), per tant és imprescindible aplicar algun mecanisme de compressió de vídeo.

### 2.3. Compressió de vídeo digital

La quantitat d'informació utilitzada en una codificació crua de vídeo pot superar la capacitat d'emmagatzematge i transmissió del sistema de vídeo digital. Per això s'han desenvolupat i perfeccionat en les últimes dècades gran varietat de tècniques de compressió.

La compressió ha d'estar completament especificada tant en el codificador en origen com en el descodificador del destinatari, així s'assegura que la recepció és correcta. En l'auge d'Internet a la fi dels noranta van aparèixer innombrables **formats propietaris** (és a dir, exclusius de cada empresa) que obligaven a fer servir programes propis per a visualitzar-los, incompatibles amb els de la competència: per a RealVideo i RealAudio de RealNetworks calia l'aplicació RealPlayer, o el reproductor del Windows si el fitxer era format Windows Media Video de Microsoft. Fins ben entrat el segle XXI no van aparèixer aplicacions capaces de suportar diferents formats, cosa que va minimitzar el caos que hi havia.

El codificador converteix un senyal de vídeo prèviament capturat en format intermedi CIF a un format comprimit, que ha de ser conegut pel descodificador i que reconstrueix el senyal de vídeo intermedi per tal de ser presentat en un televisor o monitor. Generalment el parell codificador/descodificador s'anomena **còdec** (codificador/ descodificador).



En l'especificació d'una compressió, o còdec, es representa el vídeo original per una codificació. Idealment aquesta codificació ha de ser **eficient** (ha de fer servir la mínima quantitat d'informació possible, és a dir, bits) i **eficaç** (ha de representar de la manera més fidel possible el vídeo original). Lògicament aquests dos objectius es troben en conflicte i cal un compromís entre tots dos. Habitualment un estàndard de còdec presenta diversos **perfils** o configuracions amb diferents valors de compromís entre eficàcia i eficiència amb la finalitat d'adaptar-se a una quantitat més gran de sistemes de vídeo digital.

La classificació principal de les tècniques de compressió divideix els còdecs en dues grans famílies: sense pèrdues o amb pèrdues.

### 1) Compressió sense pèrdues (*lossless*)

Després del submostreig fet amb els components de color que ja hem vist, no es fa cap compressió addicional, per la qual cosa lògicament proporcionen la màxima qualitat visual. Usualment el submostreig que es fa és 4:4:4 a 8 o 10 bits per component, i s'utilitza en entorns de postproducció en els quals la imatge s'ha de mantenir a la màxima qualitat d'origen.

Els còdecs Matrox, Aja o Blackmagic es van popularitzar en aquest entorn per a vídeo de definició estàndard, i amb l'arribada de l'alta definició encara es fan servir en la versió HD. El format DPX de Kodak, basat en el seu clàssic sistema de digitalització Cineon, permet treballar en HD amb vídeo en pur RGB i mostreig 4:4:4 a 10 bits/mostra.

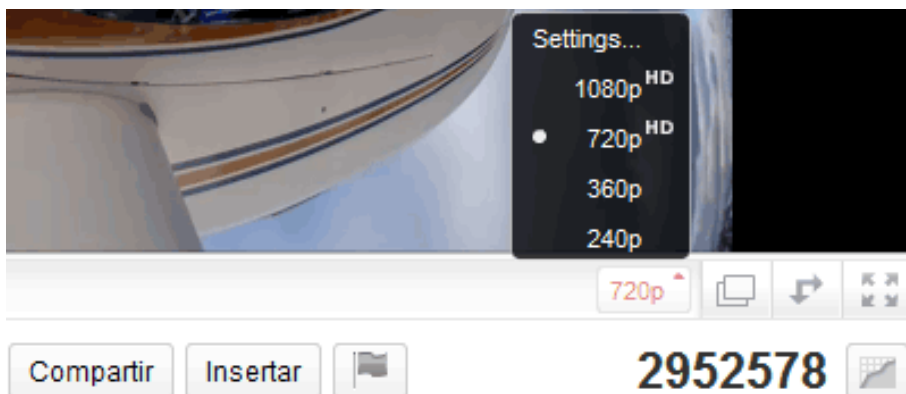
No obstant això, per a reduir les necessitats d'espai i d'amplada de banda de treball en HD (entre 1 i 1,5 Gbps) s'han començat a aplicar discretament còdecs de compressió amb pèrdues i a reduir així els requeriments d'emmagatzematge en agrupacions de discos.

En resum, les tècniques sense pèrdues queden limitades a entorns específics i dedicats, i sempre tenen taxes de compressió molt baixes (per no dir gairebé inexistents si s'apliquen sobre material de vídeo visualment ric i molt canviant).

### 2) Compressió amb pèrdues (*lossy*)

Comprimir amb pèrdues no és renunciar a una qualitat final de treball, sinó que es defineix la qualitat necessària per al destinatari final i es comprimeix fins a arribar-hi. Un clip informatiu penjat al YouTube no ha de ser de màxima qualitat si té l'objectiu d'informar, i un curtmetratge pot necessitar ser HD per tal de presentar la riquesa cromàtica dels seus escenaris mitjançant la mateixa plataforma. En tots dos casos, en el servidor del YouTube hi ha el fitxer de

vídeo en certa qualitat inicial, i el YouTube el redueix visualment segons la definició requerida pels possibles destinataris, sempre en sentit descendent de la qualitat original a inferiors:



Opcions de visualització d'un vídeo 1080p al YouTube

En el cas de vídeo generat en temps real (*live streaming*), la compressió amb pèrdues és l'aposta necessària i imprescindible per a facilitar que la imatge i el so arribin amb la cadència necessària per a ser entesos. No hi ha temps material per a fer una codificació òptima com es faria en un vídeo emmagatzemat, obligacions del directe. I així i tot no assegura l'èxit en aquest objectiu perquè hi ha altres elements que ho poden impedir, com veurem més endavant.



Molts canals de televisió també emeten en temps real per Internet (a la imatge, canal de notícies 3/24)

Les tècniques més comunes de compressió es basen a eliminar la **redundància espacial i temporal** presents en el vídeo cru, característiques habituals del vídeo:

a) **Redundància espacial:** es basa en la suposició que els píxels d'una zona d'una imatge són molt semblants als de les zones que l'envolten, excepte on hi hagi un perfil (la transició d'un objecte a un altre representat en la imatge). Com vam veure en el mòdul anterior, l'estàndard JPEG d'imatge es basa en aquestes similituds per a comprimir la imatge fixa, i emmagatzema només les lleus diferències entre grups de píxels i els seus veïns. Veurem que l'estàndard MPEG també ho aplica, en aquest cas a vídeo, en alguns dels seus quadres d'imatge.

Zona d'imatge amb redundància espacial **baixa**



Zona d'imatge amb redundància espacial **alta**

Una imatge amb zones molt diferenciades de detall

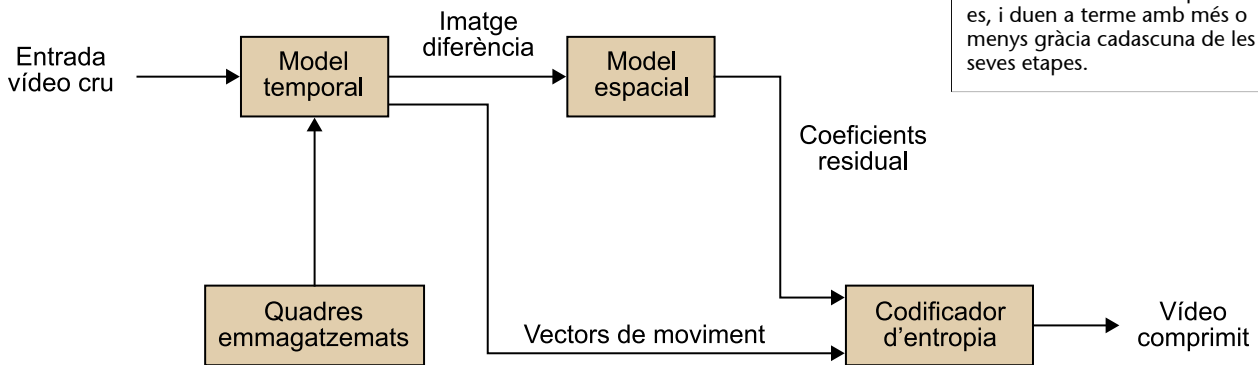
En altres paraules, les zones homogènies de la imatge són candidates a una bona compressió espacial amb poc esforç. No obstant això, el detall, és a dir, la riquesa de la imatge, és la que es posa en joc. El **compromís entre compressió i qualitat d'imatge** és un bon plantejament de compressió.

b) **Redundància temporal:** es basa en la suposició que hi ha poques diferències entre un fotograma i els immediatament anteriors o posteriors.



Un escombratge suau de la càmera d'esquerra a dreta sobre un aparador presenta molt poques diferències entre dos fotogrames consecutius del vídeo generat. Sense necessitat de calcular la imatge diferència es pot assegurar que hi ha una redundància temporal alta

La manera de detectar aquestes redundàncies i d'aplicar una compressió específica per a cadascuna es pot representar en l'esquema següent, que representa un **codificador genèric de vídeo amb pèrdues (còdec)**:



Model bàsic d'un codificador de vídeo amb pèrdues

#### Exemples

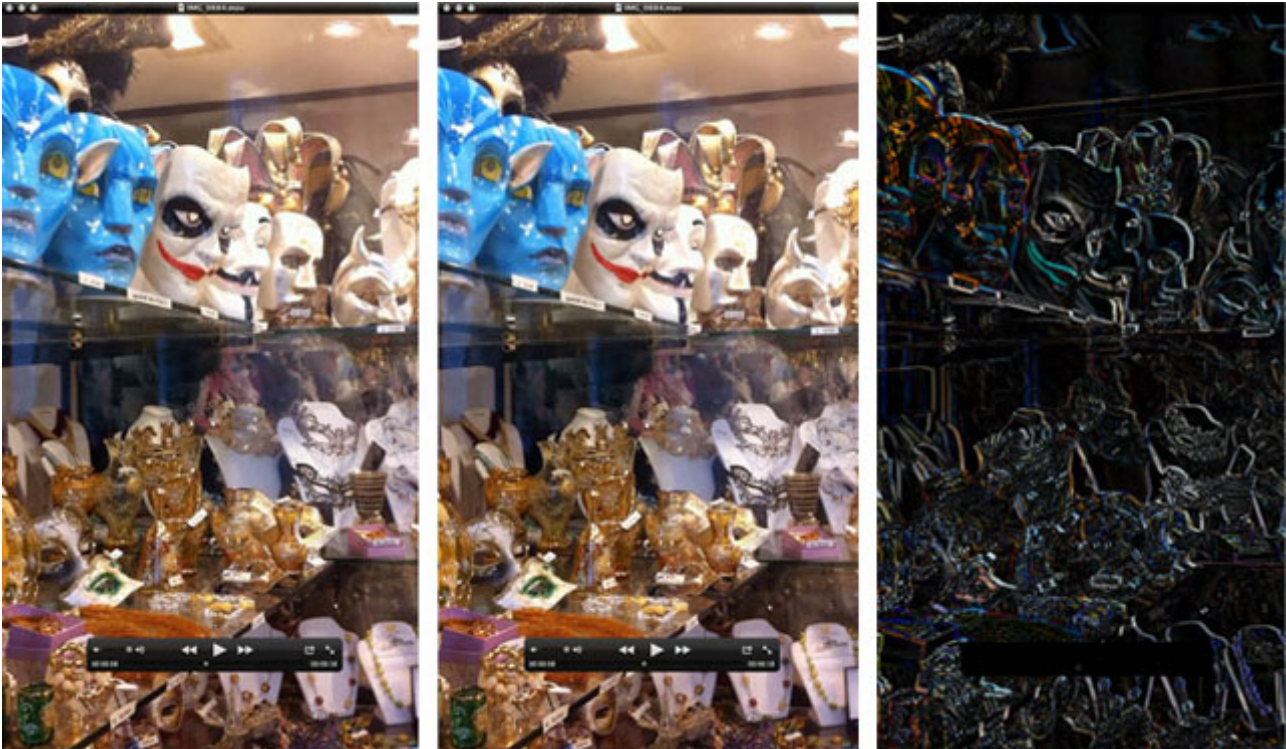
La família MPEG, H.264, Theora, VC-1, etc. són exemples de còdec que utilitzen un model de codificació amb pèrdues, i duen a terme amb més o menys gràcia cadascuna de les seves etapes.

La primera fase és el **model temporal**, que a partir dels quadres emmagatzemats del vídeo llegits fins al moment, **preveu el quadre següent** que arribarà per l'entrada de vídeo cru. Aquesta previsió usualment no està especificada i depèn de cada fabricant, i condiciona bona part de la qualitat final de tot el procés. Per exemple, un codificador MPEG-2 de l'empresa Tandberg utilitza un plantejament diferent de previsió del quadre següent que el codificador MPEG-2 de l'empresa Thomson. Sens dubte el funcionament del model temporal de cadascuna d'aquestes empreses és un secret ben guardat.

Una vegada que el model temporal ha calculat com hauria de ser el quadre següent, rep el quadre real de vídeo cru i fa dues operacions:

**a)** Primer, calcula la **imatge diferència**. En el millor dels casos, la previsió serà tan bona que la imatge diferència serà una sèrie de zeros, però la majoria de vegades conté alguna informació visual que plasma l'error d'aquesta previsió. Aquesta imatge error, i no el vídeo real, és el que seguirem comprimint en la fase següent.





Imatge real (esquerra), imatge prevista (centre) i imatge diferència (dreta)

b) Segon, com que estem treballant amb les imatges d'un vídeo en què usualment hi ha moviment, es calcula un conjunt de dades, anomenades **vectors de moviment**, que especifiquen si s'ha detectat moviment de parts de la imatge i on. Ajudarà molt en el moment de descodificar el vídeo.

La fase següent la centra el **model espacial**. La imatge diferència és comprimida utilitzant la similitud dels seus píxels: **comprimir una diferència** seria bàsicament un resum en tres paraules de tot el procés. Cada sistema de compressió de vídeo utilitza aquí les seves pròpies tècniques de compressió d'imatge, encara que les utilitzades per, per exemple, H.264, no disten molt de les del JPEG. A més, la majoria tenen pèrdues.

I amb aquesta idea al cap, com el JPEG, el resultat és una sèrie de valors que també són denominats *coeficients espacials*, i que fan el mateix paper: representen la imatge resultant amb un altre vocabulari matemàtic.

Finalment, tant els coeficients de la imatge com els vectors del moviment que s'hi han detectat són empaquetats de manera eficient en el **codificador d'entropia**, anomenat així perquè analitza la freqüència d'aparició d'aquests coeficients i vectors, i els assigna menys bits com més freqüents són. Així redueix la mida final del fitxer o del flux de dades de transmissió al mínim possible (o al valor més proper a aquest mínim).

El resultat és una sèrie binària compacta que conté tota la informació necessària per a poder ser descomprimida. Els sistemes de descompressió estan clarament definits, de manera que en aquest cas no hi ha cap diferència entre decodificadors d'una empresa o d'una altra.

## Activitat 2

Feu l'exercici 2 indicat en l'annex del mòdul 2, que presenta els formats de vídeo en què es pot exportar amb les aplicacions que es fan servir en l'assignatura.

### 2.4. Còdecs de vídeo digital amb pèrdues més habituals

Tenint en compte les particularitats de cada sistema d'àudio i vídeo digital s'han desenvolupat diferents estàndards de còdec basats en el model anterior. Com hem dit, és habitual que s'especifiqui completament el decodificador i que es deixi a l'habilitat de la indústria el fet de construir el codificador eficaç, fomentant la competència i la investigació en nous algorismes de previsió de quadre. Així, és freqüent que fent servir el mateix estàndard s'aconsegueixi una qualitat més bona i menys amplada de banda amb el pas del temps.

Els dos entorns habituals d'aplicació dels còdecs són l'emmagatzematge i la transmissió d'àudio/vídeo en temps real (*streaming*).

#### 2.4.1. Emmagatzematge

En els sistemes de codificació adreçats a emmagatzematge, el codificador té accés a tot el material que s'ha de codificar, i per això pot fer previsions més fiables i aconseguir taxes altes de compressió mantenint la qualitat audiovisual. El temps que tard a generar el fitxer final no és un factor crític, per la qual cosa els requeriments de capacitat de càlcul i necessitat de memòria no són exageradament elevats. Això permet que hi puguin haver codificadors de programari dignes per a complir aquesta missió i fins i tot funcionals en equips informàtics domèstics.

#### 2.4.2. Transmissió en temps real

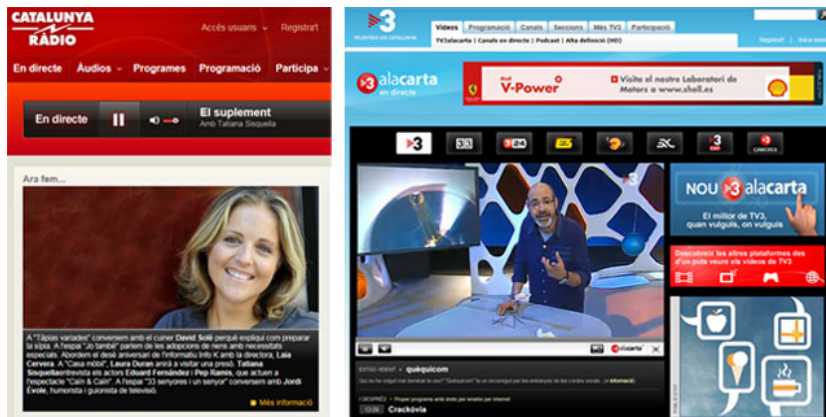
En els sistemes de transmissió en temps real com ràdio per Internet o la televisió basada en el protocol IP (IPTV), el material que hi ha per a fer la codificació arriba a la mateixa velocitat que es genera i s'ha de codificar a aquesta velocitat, per la qual cosa cal un equip d'alta disponibilitat, fiabilitat, una gran capacitat de memòria i amb una capacitat de càlcul excel·lent. La majoria són equips dedicats de maquinari d'alt valor situats en l'empresa que proveeix el servei: la majoria de webs de companyies de difusió actuals de ràdio i televisió en són un exemple.

#### MPEG-2

Si recordeu les primeres codificacions en MPEG-2 a mitjan noranta i us fixeu en les que podreu generar en aquest curs us adonareu del gran canvi de qualitat d'aquest format.



En altres serveis de transmissió en temps real menys ambiciosos (com per exemple la videoconferència), en els quals la qualitat visual no és rellevant i la cadència de les imatges no s'assegura, els requisits del codificador es relaxen i poden existir fins i tot en programari (FaceTime, Messenger, etc.).



La majoria de les empreses de ràdio i televisió emeten també part de la programació per Internet en paral·lel a l'emissió pel mitjà habitual

### 2.4.3. El cas intermedi: vídeo sota demanda

Accedir a un vídeo del YouTube (*video on demand*, 'vídeo sota demanda') sembla que es troba entre tots dos casos, ja que és visualitzar en temps real un material emmagatzemat prèviament en un servidor remot. No obstant això, el que és crucial en aquest cas és poder assegurar la cadència de recepció, i això s'ha previst prèviament en la generació del material, que s'ha creat en un format audiovisual fàcilment transportable i que pugui començar a visualitzar-se des que les primeres dades arriben al client.

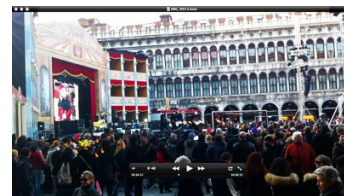
Aquesta previsió del format en el seu emmagatzematge fa que clàssicament aquest cas s'inclouï en la transmissió en temps real, i així ho farem a partir d'aquest moment encara que amb puntualitzacions. Per això, serveis de vídeo com Netflix, Hulu, Google TV o la majoria de serveis a la carta són de transmissió en temps real.

### 2.4.4. El receptor

El receptor (tant de transmissió en temps real com de lectura de fitxers emmagatzemats) té un descodificador que no necessita grans prestacions, ja que els sistemes de codificació són els que fan el treball d'anàlisi i generen la informació mínima necessària per a reconstruir el material audiovisual. Així, hi ha programari domèstic capaç tant de reproduir un Blu-ray en alta definició com de visualitzar una conferència emesa per Internet a l'altra banda del món.

### 2.4.5. Exemples d'estàndards

Actualment els dos estàndards de còdecs més habituals en els sistemes d'IPTV són H.264 (també anomenat *MPEG-4 Part 10* o *AVC*), desenvolupat en conjunt entre ITU-T i MPEG, i VC-1 (nucli de Windows Media Audio i Windows Media



VLC, QuickTime Player, Windows Media Player, etc. són exemples de reproductors de programari de propòsit general executables en entorns domèstics de poques prestacions

Video 10) desenvolupat per Microsoft. No obstant això, després de la compra d'On2 per part de Google i de les primeres proves de Google TV amb el còdec VP8 el 2010 el mercat pot alterar-se completament en pocs anys.

Per als sistemes que utilitzen xarxes basades en IP, com Internet i les xarxes locals, o serveis IP sobre 3G com a mitjà de transport, les compressions més utilitzades són:

1) **MPEG, H.264/MPEG-4 AVC i 3GPP**: formats estàndards, principalment utilitzats per QuickTime, en la seva línia de servidors Darwin/QuickTime Stream Server i en el reproductor QuickTime. Igualment Adobe en el Flash Video Server utilitza H.264 com a còdec des de la versió 9. Cal destacar que la utilització dels estàndards MPEG requereix pagar llicència tant en servidors com en reproductor. Per la seva qualitat, H.264/MPEG-4 també s'utilitza en emmagatzematge de vídeo en Blu-ray.

2) **Windows Media 10**: codificació propietària desenvolupada per Microsoft, té la seva fortalesa en la transmissió en temps real de baixa amplada de banda (com els d'Internet). Pot ser servit per la línia de servidors Microsoft utilitzant el servei Windows Media Services i reproduït pel Windows Media Player. Aquesta codificació requereix un permís especial i pagar llicència a Microsoft per a poder ser utilitzat en una altra línia de servidor/reproductors.

3) **Theora**: format obert, amb l'atractiu principal que no cal pagar llicència per a utilitzar-lo. Té una gamma d'estàndards en desenvolupament continu que pretén competir amb la línia d'estàndards MPEG-4. La codificació de vídeo Theora es basa en la codificació On2's VP3 i el seu contenidor més conegut per a emmagatzematge és el format Ogg.

4) **RealVideo**: codificació propietària, desenvolupada per RealNetwork, empresa pionera en transmissió en temps real d'àudio i vídeo a Internet actualment en desús. La seva fortalesa era la transmissió en temps real de baixa amplada de banda. Podia ser servit per la línia de servidors Helix i reproduït pel RealPlayer (o els antecessors HelixPlayer i PlayerOne) tots de l'empresa RealNetwork. Aquesta codificació requeria un permís especial i pagar llicència a RealNetwork per a poder ser utilitzat en una altra línia de servidor/reproductors, per la qual cosa el cost d'aquest còdec i la falta d'inversió en innovació ha suposat que hagi desaparegut gairebé del tot del mercat actual.

5) **VP8**: és el còdec obert de vídeo en transmissió en temps real de WebM, una nova proposta de format de Google per a ser utilitzat en HTML 5. De qualitat semblant a H.264, es basa en una llibreria de programari lliure denominada *x264* sota llicència GNU. El seu contenidor es basa en el contenidor obert Matroska.



## 2.5. Compressió d'àudio digital

Les tècniques de compressió d'àudio són cronològicament anteriors a les de vídeo. Com en aquest, hi ha tècniques amb pèrdues i sense pèrdues, algunes d'específiques per a transmetre en temps real i unes altres per a emmagatzemar, i totes es basen en la **reducció de la redundància** que hi ha en el senyal per a minimitzar el flux de dades generat per segon sense afectar la qualitat segons la seva necessitat.

Del gran ventall d'estàndards que hi ha, els uns se centren en la codificació de la veu humana per a transmetre en xarxes digitals. Serveis de transmissió de veu en temps real com la telefonia fixa, la telefonia mòbil o Internet aconseguen grans taxes de compressió en centrar-se en les característiques específiques d'aquest senyal. És habitual que es basin en tècniques de predicció, codificant, com en imatge i vídeo, les diferències entre la informació predita d'àudio i la real.

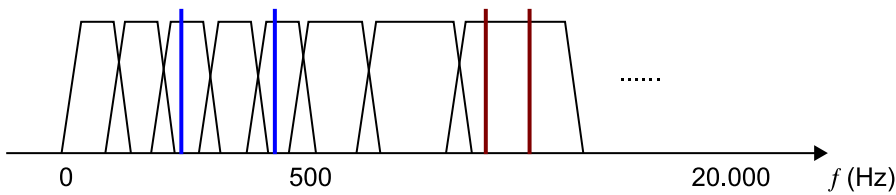
Com a exemples il·lustratius:

- Les xarxes **GSM** i **UMTS** codifiquen l'àudio en format AMR, tant en l'emissió com en l'emmagatzematge de missatges a la bústia de veu (extensió .amr), i és el format base del grup 3GPP per a futures xarxes mòbils de quarta generació.
- **Skype** es basa en el seu propi estàndard Silk (*super wideband audio codec*).

Un altre gran grup, més interessant en el nostre cas, té com a objectiu aconseguir la màxima qualitat de so amb el mínim nombre de bits per segon, i fins i tot possibilitar so multicanal. Es basa majoritàriament en la **codificació perceptual**, un avenç en el món de la codificació d'àudio implementat per primera vegada per Philips en el seu ja desaparegut sistema d'àudio compacte digital (DCC), que pretenia substituir el casset de cinta.

La codificació perceptual neix dels estudis sobre com l'ésser humà interpreta el so, i apareixen tres patrons que obren la porta a comprimir l'àudio sense una pèrdua audible de qualitat:

1) L'orella no detecta tots els sons que rep, es comporta com **un banc de filtres** que més o menys se solapen en freqüència. Dos sons situats en bancs diferents són diferenciats clarament pel cervell, però si són al mateix banc de freqüències, no.

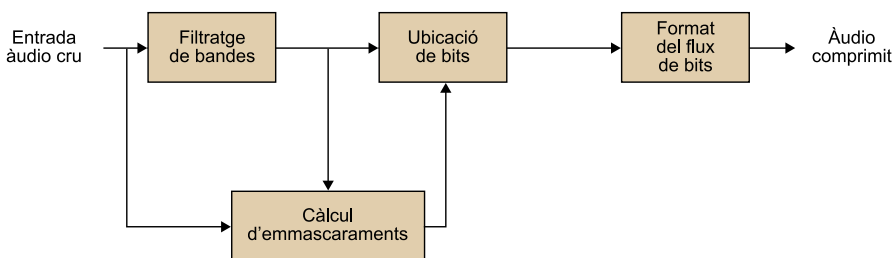


Dos tons (les dues línies verticals de l'esquerra) situats en freqüències de bandes diferents són diferenciats per l'orella, mentre que dos situats a la mateixa banda (les dues línies verticals de la dreta) són interpretats com un de sol

2) La sensibilitat de l'orella davant un so canvia si hi ha més sons. D'alguna manera, uns sons n'emascaren uns altres, i els fan indetectables per l'orella. Aquest fenomen es denomina **emascarament freqüencial**.

3) Més i tot, si un so feble que estem sentint es deté, es tardarà cert temps fins a poder sentir un so feble una altra vegada. Aquest segon fenomen es denomina **emascarament temporal**.

El model general de codificació perceptual pot expressar-se gràficament de la manera següent:



Model de codificació perceptual

Primer es fa un filtratge de bandes, i es rebutja en cada una els tons que comparteixen banda però que no seran detectats. Un calculador d'emascarament temporal i de freqüència rep aquestes dades i, agafant com a referència el senyal original, envia o no dades a ser representades per bits. Finalment, el flux de bits rep un format depenent de l'estàndard específic.

El sistema està fonamentat en les dades experimentals recollides per milers de proves en voluntaris, de manera que el model psicoacústic es pot interpretar com un comportament mitjà de l'orella humana. Hi ha subjectes, per tant, capacitats acústicament millor que uns altres que poden detectar la pèrdua de qualitat inherent a aquesta codificació.

Dels còdecs d'àudio habituals actualment, la majoria segueixen aquest model o variacions d'aquest, i uns altres han tendit cap a sistemes de compressió més conservadors:

- El format de so **MP2** (anomenat també *Musicam*) és l'habitual en fitxers d'àudio professional.

- L'estàndard MPEG-1 Audio Layer 3 nascut a partir de MP2, és un clàssic de la família de còdecs amb pèrdues, es va definir per a emmagatzematge d'àudio digital domèstic. Es coneix com a **MP3**.
- **Vorbis** també és un estàndard d'àudio amb pèrdues resultat d'un projecte de programari lliure ofert com a alternativa a l'estàndard propietari MP3, que ofereix una qualitat d'àudio similar i fins i tot millor en velocitats baixes de transmissió o lectura (< 64 kbps).
- **AAC** (Advances Audio Coding) posterior i de més qualitat que l'MP3, es troba en equips domèstics portàtils d'àudio d'Apple (iPod, iPhone, iPad), la família de mòbils Android, Sony (PlayStation 3), Nintendo (Dsi, Wii), etc. Suporta àudio multicanal (fins a 48 canals) i és una codificació amb pèrdues.
- **Dolby AC-3** (conegut com a *Dolby Digital*) és un estàndard propietari amb pèrdues nascut per al cinema, però la seva alta eficiència (320 kbps per 5 canals d'àudio d'alta fidelitat més un sisè canal de so envoltant) i una bona relació amb els fabricants d'equips el va alçar com a estàndard d'àudio per a DVD i Blu-ray, i fins i tot per a la televisió digital als Estats Units (i, per empatia, també accessible en algunes televisions europees).
- **Dolby Digital Life**, no obstant això, és un estàndard de codificació en temps real, adreçat al mercat multimèdia, especialment el sector de videojocs.
- **Windows Media Audio** (WMA) és l'aposta de Microsoft per un estàndard de codificació. Existeix tant amb pèrdues com sense pèrdues (WMA Lossless).
- I, només com a comentari, **RealAudio** va ser un estàndard de compressió amb pèrdues específicament dissenyat per a transmissió en temps real, molt utilitzat en els primers anys d'Internet, però la seva naturalesa propietària i la falta de millores per part de l'empresa RealNetworks el va portar a desaparèixer fa uns anys.

## 2.6. Diferenciant contenidor i còdec

Un **format de contenidor** és un format d'arxiu digital que emmagatzema determinat tipus d'informació codificada amb un còdec estàndard o diversos.

És habitual confondre aquests dos conceptes, i barrejar el nom del còdec amb el del contenidor. Així, Flash Video és un format contenidor l'extensió de fitxer del qual és .flv o .f4v i en què el vídeo pot estar codificat en H.264 o en un altre còdec com Sorenson Spark o VP-6, mentre que l'àudio pot estar codificat en MPEG-1 Audio Layer 3.

En alguns casos el nom del contenidor acaba substituint el del còdec per costum o simplificació.

Els contenidors més complexos són capaços de suportar diferents còdecs d'àudio i vídeo, gestionar subtítols, capítols i metadades (o *tags*). Un dels rols més importants del contenidor és propiciar informació temporal per a poder sincronitzar la reproducció de més d'un flux de dades o *stream* simultàniament.



Un exemple d'associació pura entre còdec de vídeo i contenidor

#### MP3

El contenidor pel còdec MPEG-1 Audio Layer 3 és comunament conegut com a MP3 i l'extensió dels arxius és .mp3, però és habitual parlar de *codificació MP3*.



Bona part de les aplicacions de codificació d'àudio i vídeo indiquen clarament els còdecs de vídeo i àudio possibles d'un contenidor seleccionat (a la imatge, Super©)

Algunes característiques que diferencien els contenidors entre si són les següents:

- La popularitat (i suport per part de les aplicacions).
- La sobrecàrrega o l'*overhead* (diferents contenidors presenten diferents mides d'arxiu per al mateix contingut).
- El suport de còdecs (per exemple, alguns contenidors com l'AVI no suporten codificacions amb quadres B).
- El suport de subtítols o altres característiques avançades.
- Que sigui apte per als servidors de transmissió en temps real com a format d'entrada.

La majoria dels contenidors tenen la possibilitat de fluxos de bits variables tant per a àudio com a vídeo. Alguna de les excepcions més notòries és per exemple el contenidor AVI de Microsoft, que no ho permet. Actualment quan un contenidor no té alguna característica desitjada, els diferents programes fan extensions per a poder suportar-les, encara que moltes vegades aquestes extensions són incompatibles entre si, i compliquen l'escenari.

La taula següent resumeix les característiques dels contenidors més populars:

Característiques dels contenidors més populars

| Contenidor | Còdec de vídeo suportat               | Còdec d'àudio suportat | Capacitat per a transmissió en temps real | Capacitat per a 3D |
|------------|---------------------------------------|------------------------|---|--------------------|
| .3gp       | MPEG-4 Part II<br>H.264/MPEG-4<br>AVC | AAC<br>AAC v2          | –   | –                  |

| Contenedor                                  | Còdec de vídeo suportat  | Còdec d'àudio suportat                            | Capacitat per a transmissió en temps real | Capacitat per a 3D |
|---|--|---|---|--------------------|
| .avi  | la majoria excepte H.264/AVC   | la majoria  | Sí  | –                  |
| .divx                                       | MPEG-4 Part II   | MPEG-1 Audio<br>AC-3<br>PCM                       | Sí  | –                  |
| (Flash Video)<br>.f4v                       | H.264/MPEG-4<br>AVC  | AAC<br>MPEG-1 Audio                               | Sí  | –                  |
| (Flash Video)<br>.flv                       | H.264/MPEG-4<br>AVC<br>Sorenson Spark<br>VP6                                 | AAC<br>MPEG-1 Audio<br>PCM                        | Sí  | –                  |
| (Matroska)<br>.mkv<br>.mka<br>.mks<br>.mk3d | la majoria   | la majoria  | Sí  | sí                 |
| (MPEG)<br>.mp4                              | MPEG-2 Part II<br>H.264/MPEG-4<br>AVC<br>H.263<br>VC-1                       | AC-3<br>MPEG-2<br>MPEG-4<br>Vorbis                | Sí  | sí                 |
| (MPEG)<br>.mpg<br>.mpeg                     | MPEG-1<br>MPEG-2   | MPEG-1 Layer I,<br>II, III                        | Sí  | –                  |
| (MPEG Program<br>Stream)<br>.pg             | MPEG-1<br>MPEG-2<br>MPEG-4 Part II<br>H.264<br>VC-1                          | MPEG-1 Layer I,<br>II, III                        | Sí  | –                  |
| (MPEG Transport<br>Stream)<br>.ts           | MPEG-1<br>MPEG-2<br>MPEG-4 Part II<br>H.264<br>VC-1                          | MPEG-1 Layer I,<br>II, III                        | Sí  | –                  |
| (Blu-ray Trans-<br>port Stream)<br>.m2ts    | MPEG-2 Part II<br>H.264<br>VC-1  | Dolby AC-3<br>Dolby Digital Plus<br>DTS<br>DTS HD | No aplicable                              | sí                 |
| (AVID)<br>.mxf                              | la majoria   | la majoria  | No aplicable                              | –                  |
| (XIPH.ORG)<br>.ogg                          | Theora<br>i molts altres   | Vorbis<br>i molts altres                          | Sí  | –                  |
| (Apple)<br>.qt<br>.mov                      | MPEG-1<br>MPEG-2 part II<br>MPEG-4 Part II<br>H.264/MPEG-4<br>AVC<br>Cinepak | AAC<br>MPEG-1 Audio<br>PCM                        | Sí  | –                  |
| (Real)<br>.rmvb                             | RealVideo  | RealAudio<br>AAC<br>Vorbis                        | Sí  | –                  |

| Contenedor                  | Còdec de vídeo suportat      | Còdec d'àudio suportat | Capacitat per a transmissió en temps real | Capacitat per a 3D |
|-----------------------------|------------------------------|------------------------|---|--------------------|
| (Microsoft)<br>.wma<br>.wmv | la majoria excepte H.264/AVC | la majoria             | Sí  | –                  |
| (Google)<br>WebM            | VP8                          | Vorbis                 | Sí  | sí                 |

### 3. L'estàndard MPEG-2

El grup MPEG (Moving Pictures Experts Group) de la ISO (International Organization for Standardization) és el grup més destacat i acceptat per la indústria pel que fa a estandardització de vídeo digital.

Després de l'èxit obtingut amb l'estàndard MPEG-1 el 1993, el grup va decidir fer un estàndard més complet i millorat, adreçat a un mercat potencial com la radiodifusió de televisió digital, atractiu i de gran escala. Finalitzat el 1995, l'estàndard té millores en la compressió respecte al seu antecessor, codificació entrelaçada (a més de progressiva) i gran flexibilitat a causa de la utilització de **perfils i nivells**. L'MPEG-2 estandarditza tant el vídeo com l'àudio que l'acompanya.

Els perfils i nivells són variacions de qualitat de compressió que permeten adaptar l'estàndard a les exigències dels sistemes particulars, a més d'oferir als fabricants regles més clares respecte a la conformitat dels seus productes amb l'estàndard. Els perfils i nivells de l'MPEG-2 els veurem més endavant.

L'estàndard MPEG-2 és molt utilitzat en l'actualitat, encara que lentament és substituït per l'MPEG-4 Part 10/H.264/AVC. És utilitzat per la majoria dels sistemes de difusió de televisió digital (DVB), tant per cable (DVB-C), satèl·lit (DVB-S) o terrestre (DVB-T), i és present al mercat domèstic amb el format d'emmagatzematge DVD.

Sense ànim d'entrar en els detalls profundament tècnics, farem una revisió general de l'estàndard a partir de les parts que el conformen: primer, veurem l'apartat més important, la definició del format multimèdia. Els fluxos de dades i la manera d'empaquetar-los els veurem molt per sobre.

L'emmagatzematge i la transmissió de fluxos MPEG-2 s'estudia amb més detall més endavant, aquí només ens hi aproximarem. Tancarem l'apartat amb una breu ressenya del còdec estandarditzat MPEG-2 i una descripció dels perfils i nivells de qualitat permesos.

#### 3.1. Altres estàndards relacionats: JPEG/JPEG2000, MPEG-1, H.261 i H.263

Alguns estàndards anteriors a l'MPEG-2 van ser font d'inspiració per a aquest i en algun cas són encara vigents:

1) **JPEG/JPEG2000**. Similar a l'MPEG, el JPEG (Joint Photographic Experts Group) és un grup de treball d'ISO, en aquest cas especialitzat en la compressió d'imatges. El 1992 va acabar l'estàndard JPEG, que encara és utilitzat. La compressió espacial de la codificació de vídeo digital està fortament lligada a les tècniques presentades pel grup JPEG.

**JPEG**

ISO/IEC 10918-1 / ITU-T Recommendation T.81.

El 2000 es va presentar una versió millorada de l'estàndard conegut com a *JPEG2000*, que aposta per un sistema de codificació nou i més eficient que a més permet codificar de manera escalada i fins i tot amb compressió sense pèrdues. No va tenir tanta popularitat perquè era de llicència.

**JPEG2000**

ISO/IEC 15444.

2) **MPEG-1**. El primer estàndard del grup MPEG, és l'MPEG-1. Culminat el 1993 actualment encara s'utilitza. El cas més notori d'èxit és el còdec d'àudio MP3.

**MPEG-1**

ISO/IEC 11172.

Per a vídeo, el còdec MPEG-1 va ser utilitzat en el format vídeo CD (o VCD), actualment reproduïble en la majoria dels reproductors DVD i de qualitat similar a la d'un vídeo VHS domèstic.

**MP3**

Especificat en la norma MPEG-1 Part 3 Audio Layer 3.

3) **H.261 i H.263**. El 1993 la ITU-T estandarditza l'H.261, un còdec per a serveis de videoconferència de baixa velocitat (es transmetia per RDSI a 64 kbps o múltiples d'aquest). Pensat per a xarxes commutades com la de telefonia analògica però amb serveis digitals com RDSI, actualment és l'únic context en què encara s'utilitza en general per raons de compatibilitat amb programes antics.

El 1998 la ITU-T estandarditza l'H.263, un còdec de vídeo de més qualitat i menys flux de bits resultant (30 kbps). En aquest còdec apareix el concepte de diferents perfils de codificació (variacions de qualitat en la compressió que van generar inicialment alguns problemes de compatibilitat entre els fabricants, resolts quan en l'H.264 s'aclareixen les especificacions).

### 3.2. Què és l'MPEG-2

L'estàndard MPEG-2 està completament definit en la ISO/IEC 13818. Aquesta norma està dividida en parts segons el tema. Cada part es considera un estàndard en si mateix; per tant, de manera més correcta: l'MPEG-2 és un conjunt d'estàndards.

Les parts que conformen l'estàndard MPEG-2 són, entre d'altres, les següents:

- El sistema: com sincronitzar i conjuntar els fluxos de vídeo i àudio en un sol flux de dades.
- El còdec de vídeo per a senyals entrelaçats i progressius.
- El còdec d'àudio, que és una extensió de l'MPEG-1 Audio (també conegut com a *MP3*) capacitat per a so multicanal.



- El control del flux per part de l'usuari final (per a reproduir, posar la pausa, situar-se en un punt en particular, estructurar en escenes, etc.).

La propietat intel·lectual de l'estàndard MPEG-2 és compartida per diverses corporacions. Més de 500 patents formen l'estàndard. Els venedors de productes i serveis basats en l'estàndard MPEG-2 han de pagar per a explotar la llicència. La institució beneficiària és MPEG-LA (<http://www.mpegla.com/avc/>), que administra el conjunt de patents MPEG.

### 3.3. Codificació i compressió en MPEG-2

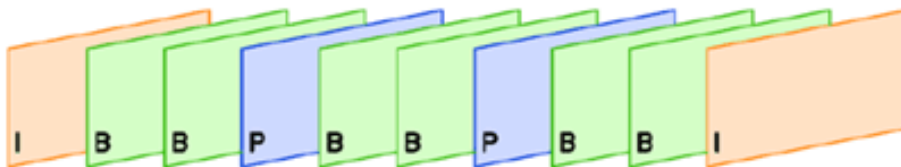
Quan seiem còmodament per a gaudir d'un DVD al reproductor o d'un fitxer en format MPEG-2 a l'equip multimèdia o a l'ordinador i volem anar a un punt en particular, l'avanç ràpid o el retrocés del nostre comandament a distància tenen un comportament diferent del que tenien aquestes funcions en el vell vídeo VHS.

Abans, quan avançàvem o rebobinàvem una cinta de vídeo, podíem veure amb més o menys claredat la pel·lícula accelerada, però ara la imatge va a trompades, i fa salts de segons a segons, presentant instantànies de la pel·lícula.

On són la resta de fotogrames? Per què no veiem els actors movent-se de manera accelerada? La culpa és del GOP (el grup d'imatges).

#### 1) MPEG i GOP

La imatge en moviment pot comprimir-se amb més pèrdues que la fixa, ja que entre una imatge i la següent usualment hi ha poques diferències (excepte un canvi de pla o una transició ràpida). La intenció és, de totes, mantenir-ne algunes d'intactes (és a dir, amb una compressió similar al JPEG, denominada *intraframe*) que serviran de referència a la resta, i es calcularà com a diferència o desplaçament d'aquestes. Les imatges de referència són denominades *I*.



#### Exemple de GOP de $M = 9$ i $N = 3$ (ordre de visualització) Imatges I

La imatge I (*intraframe*) no té cap referència amb les altres, es comprimeix espacialment ("a la JPEG", és a dir, es divideix en blocs de  $8 \times 8$  píxels i es transforma, es quantifica, es codifica, etc.). Es codifiquen soles, per això es denominen *intraframe*.

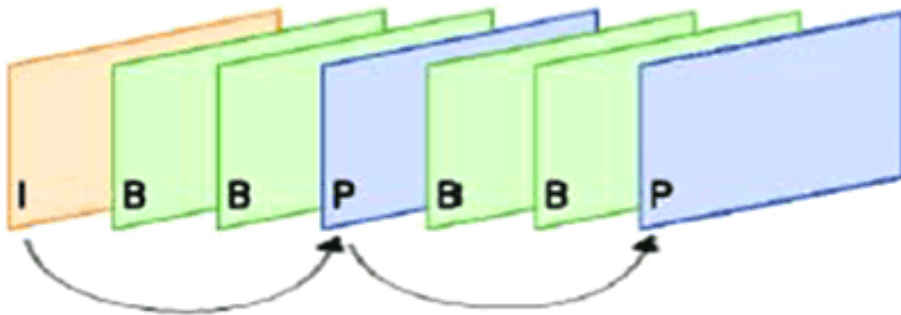
Aquestes són les que ens apareixen en pantalla quan fem una recerca ràpida cap endavant o enrere al reproductor DVD o programari: un reproductor de 39,95 € no necessita altres imatges (això implica, memòria, molt cara avui dia), només uns quants càlculs per a poder descomprimir la imatge I. És una manera econòmica de presentar-nos el punt aproximat de la pel·lícula on som.

Si codifiquéssim totes les imatges d'una seqüència com a imatges I, no tindríem una gran taxa de compressió i la pel·lícula no cabria en un DVD o el fitxer seria de desenes de gigabytes. Cal buscar un mètode més agressiu de compressió per a la resta d'imatges, i tenim l'oportunitat, ja que l'ull humà davant objectes en moviment no destria amb claredat; aquest mètode es denomina **estimació de moviment**.

## 2) Imatges P i l'estimació de moviment

Tenint els "pilars" I d'una seqüència d'imatges, les imatges, que hi ha entre dues imatges I es poden codificar amb estimació de moviment de dues maneres: imatges P i B.

Les imatges P (predictives) es calculen a partir de la imatge I o la imatge P immediatament anterior (no vol dir que sigui l'anterior, sinó l'última I o P que s'hagi codificat). Són també de gran qualitat encara que de menys que la I, ja que es calculen amb una estimació del moviment que hi ha en la imatge (es divideix la imatge en macroblocs de  $16 \times 16$  píxels i s'intenta veure quant s'han mogut respecte a la imatge que prenen de referència, sigui I o P).

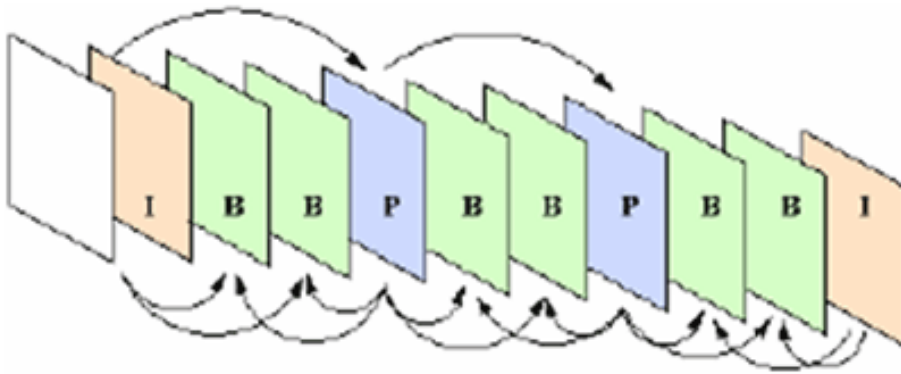


La intenció és que d'aquesta imatge només en guardem uns quants píxels i molts vectors de moviment que indiquin quant s'han mogut els blocs de píxels de la imatge I o P de referència. Aquest procés estalvia moltíssims bits. El problema és, doncs, calcular els vectors de moviment, és a dir, analitzar les diferències entre la imatge P i la de referència (I o P), comparar-les, veure si hi ha blocs de píxels similars i detectar quant s'han desplaçat en horitzontal i vertical.

En seqüències estàtiques (la càmera quieta i els objectes de la imatge en repòs), les imatges P són senzillament nul·les. En seqüències en moviment (i sobretot en una persecució policíaca), les imatges P "s'engreixen" perquè han d'emmagatzemar el moviment dels píxels de la imatge de referència, i si a més hi ha objectes nous, els han d'afegir.

## 3) Imatges B

Les imatges B (bidireccionals predictives) són el cas més extrem de compressió. A partir de dues imatges de tipus I o P, es dedueix la imatge B intermèdia, que conté més informació d'estimació de moviment augmentant la compressió, però d'una qualitat relativament baixa. Vegeu que, si bé les P usaven estimació de moviment en un sol sentit (des de la I o P anterior fins a aquestes), les B fan servir els dos sentits (l'anterior i la posterior I o P).



#### 4) Propagació d'errors (qualitat visual) i compressió: dilema etern

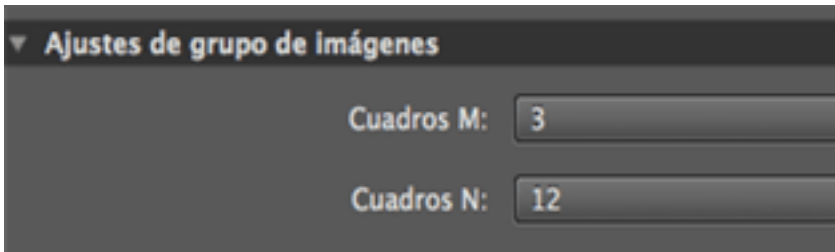
Les imatges I contenen errors, ja siguin propis de la codificació estil JPEG (una codificació amb pèrdues), ja siguin externs provocats per la transmissió (per exemple una antena mal instal·lada per a la TDT) o per processos d'enregistrament (un DVD en males condicions). Ara bé, si la resta d'imatges P i B depèn d'aquestes, és imaginable pensar que aquests errors de base es propagaran i les afectaran.

També les P poden contenir errors, ja sigui heretats de la I o P que depenguin, ja sigui per un error en la transmissió o enregistrament. Les B, senzillament, són les últimes víctimes de tots els errors que s'hagin produït de tota classe.

El valor N és la distància entre dues imatges I. Repassant el gràfic que representa un GOP un parell de pàgines abans, es dedueix que el GOP ha de tenir una mida de N imatges, ja que comença amb una imatge I i acaba just abans de la imatge I següent. Un valor de 12 és un equilibri entre una gran distància entre dues imatges de referència I i una compressió interessant amb una propagació d'errors acceptable.

Augmentar aquesta distància suposaria deixar en mans d'imatges P i B la qualitat de la seqüència i arriscar que petits errors de les imatges I s'amplifiquessin en tot el GOP. Reduir-la seria millorar la qualitat però també augmentar el pes de la seqüència i reduir la compressió.

El valor  $M$  és la distància entre una imatge  $I$  i la imatge de tipus  $I$  o  $P$  que hi hagi després. En l'exemple d'abans, és 3. Com més gran és el valor de  $M$ , més errors contindrà l'estimació de moviment, però més alta serà la compressió, dilema etern. En una exportació de projecte a MPEG-2, per exemple, l'Adobe Premiere deixa l'elecció a l'usuari:



Exemple de selecció dels valors  $N$  i  $M$  en una codificació MPEG-2 (cas Adobe Premiere)

### Exemples

- DVD: en fer una recerca d'imatge amb el comandament a distància en un DVD, el vídeo no flueix dinàmicament en pantalla, sinó a salts. Això és perquè va saltant d'imatge  $I$  a imatge  $I$ .
- En codificar una pel·lícula en MPEG-2, cada canvi de pla força l'inici d'un GOP. Així, la primera  $I$  conté els objectes nous.
- Reproducció en temps real per Internet: un GOP de  $N = 12$  i  $M = 3$  és habitual.
- Edició de vídeo no lineal: com més imatges  $I$  hi hagi, millor. Fins i tot hi ha l'estàndard M-JPEG, que està format exclusivament per imatges  $I$ .

### 3.3.1. Els fluxos elementals de dades

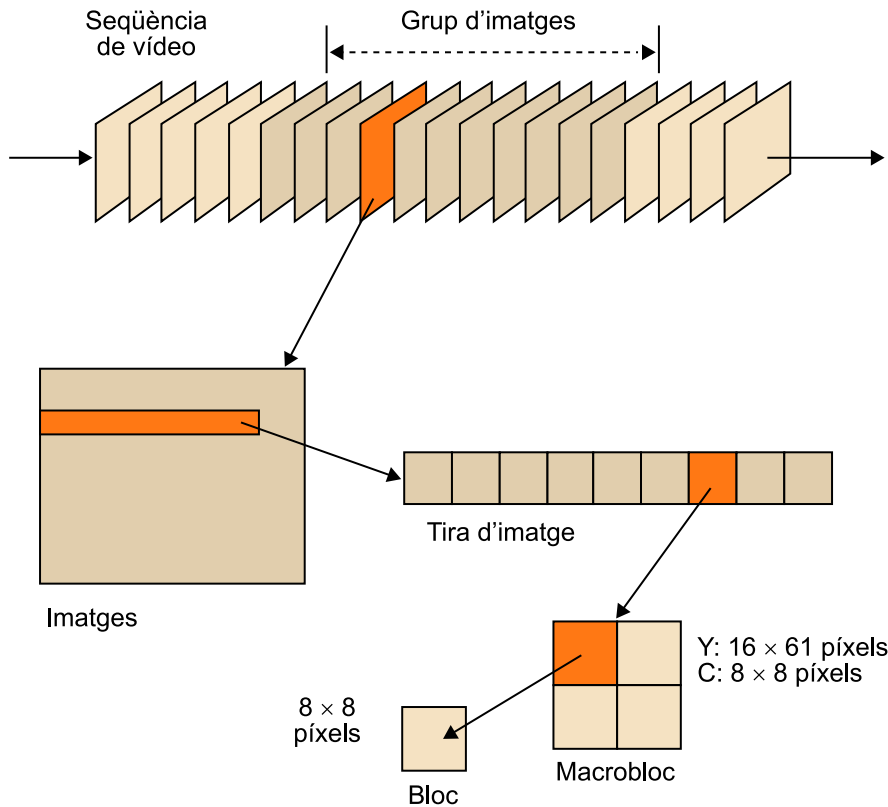
Una vegada generades les imatges  $I$ ,  $P$  i  $B$ , a partir d'aquestes cal construir un flux final de bits per emmagatzemar o transmetre, i aquest procés té diverses etapes.

Comencem definint el **flux elemental** (*ES*, *elementary stream*), que és senzillament la sortida del codificador i conté tota la informació necessària perquè un descodificador creï una aproximació tan precisa com sigui possible al vídeo original.

Hi ha dos tipus diferents d'ES, els d'àudio (*audio elementary streams*, AES) i els de vídeo (*video elementary streams*, VES). A continuació es descriuen tots:

#### 1) Flux elemental de vídeo (VES)

Un VES és una seqüència de vídeo; habitualment la sortida del codificador ja es troba en format VES. Com es veu en la figura següent, el flux no conté cada imatge per separat, sinó que s'estructura en grups d'imatges (GOP, *group of pictures*), la unitat de treball de qualsevol reproductor que vulgui presentar aquest flux (un reproductor DVD, un programa de reproducció, etc.).



Esquema del funcionament per grup d'imatges

## 2) Construcció del VES

Cada imatge es converteix en tres matrius rectangulars de valors, cadascuna amb els valors de Y, Cb i Cr, respectivament. En general, la matriu Y és quatre vegades més gran que les altres dues, ja que, com hem vist, és més rellevant la luminància (Y) que els components de color (Cb, Cr).

Segons el contingut i la comparativa que fa entre imatges consecutives, el codificador pren la decisió de definir una de les imatges com de tipus I, és a dir, codificable de manera independent i pilar bàsic per a la codificació de les immediatament següents. És habitual que, per exemple, la primera imatge després d'un canvi de pla es defineixi com a I, ja que no té relació amb les anteriors, però el criteri és molt més complex i cada fabricant aplica el que creu més convenient (l'algorisme no està definit en l'estàndard).

A partir de la imatge I, es calculen les imatges P posteriors i les B tal com s'ha dit. És important destacar que dins d'un GOP sempre es manté la mateixa seqüència d'imatges, una vegada definit el M i N de la codificació.

És important recalcar que l'MPEG no especifica en els seus estàndards com ha de ser un codificador, sinó que especifica la sortida del codificador, per a assegurar-se compatibilitat amb el descodificador. Així, codificadors i descodificadors de diferents fabricadors poden interoperar, i això també permet a la indústria investigar i millorar les tècniques de codificació sense sortir de l'estàndard.

L'estàndard en canvi sí que especifica un mecanisme de descodificació (un descodificador), però ha de ser pres com a referència, i cal deixar lliure la indústria per a canviar-lo sempre que aconsegueixi el mateix resultat que l'estàndard (en aquest aspecte es pensa en millores de *performance* i adaptació del mètode a maquinari específic).

En l'MPEG-2 hi ha múltiples implementacions de codificadors i descodificadors, algunes per a maquinari i moltes per a programari. La qualitat de codificació MPEG-2 ha millorat sensiblement des que va sortir i la seva qualitat és alta fins i tot amb fluxos de menys de 2 Mbps. És un estàndard madur, però amenaçat per l'MPEG-4, més actualitzat i que el supera en tots els aspectes, excepte l'econòmic.

Habitualment les solucions d'IPTV utilitzen equipament dedicat i costós per a la codificació. La descodificació es fa a l'equip del client, majoritàriament en programari (navegadors d'Internet) però hi ha algunes excepcions en maquinari (televisió per cable).

### 3) Flux elemental d'àudio (AES)

Un AES és la sortida del codificador d'àudio. Aquest cas és molt diferent del de compressió vídeo, aquí no hi ha diferents tipus de quadres, els quadres d'àudio són tots codificats de la mateixa manera, per la qual cosa tots tenen la mateixa mida.

### 4) Flux elemental empaquetat (PES)

A l'hora d'agrupar el *flux de dades* de vídeo (VES) i el d'àudio (AES) en un sol *flux* o *stream*, cal anar amb compte de no generar retards entre tots dos (enviar massa vídeo i poc àudio suposaria no tenir prou àudio o tenir-lo retardat respecte al vídeo). Cal dividir els dos *fluxos* en paquets d'una mida adequada abans de ser posats seqüencialment l'un darrere l'altre. Aquest procés es denomina *empaquetat*, i suposa indicar en cada paquet el tipus i uns identificadors perquè el receptor pugui reconstruir-los correctament, com també dades addicionals per a assegurar la sincronització (quin fragment d'àudio va amb quin fragment de vídeo). El flux total és el PES (*packetized elementary stream*).

### 3.3.2. Emmagatzematge i transmissió

Generalment cal combinar diversos PES (almenys un àudio i un vídeo) per a crear un contingut multimèdia que serà reproduït més endavant. Dues possibilitats de tractament sorgeixen llavors: emmagatzemar el contingut per a la seva reproducció posterior o transmetre'l (en el nostre cas, per una xarxa IP).

#### 1) Emmagatzematge

L'estàndard de l'MPEG-2 no especifica cap format de contenidor d'arxiu. Això també passa amb l'MPEG-1, H.263 i d'altres, per la qual cosa poden ser utilitzats molts contenidors: MP3 per a àudio, MOV per a vídeo, etc.

#### 2) Transmissió

L'estàndard del sistema de l'MPEG-2 (ISO/IEC 13818-1) defineix dos mètodes de combinació de les dades d'àudio, vídeo i contingut associat per a la transmissió en un únic flux de dades:

- **Program streams** (conegut com a *MPEG-2 PS* o simplement *MPEG-PS*)
- **Transport streams** (conegut com a *MPEG-2 TS* o simplement *MPEG-TS*)

Un canal (com un canal de TDT) està format almenys per un PES de vídeo i un PES d'àudio. El procés d'ajuntar diversos PES que seran reproduïts conjuntament es diu **multiplexació**. Per a aconseguir una reproducció conjunta sincronitzada (sense desfasament entre àudio i vídeo) cal enviar informació de rellotge en la multiplexació.

Mentre que l'MPEG-PS permet transportar un únic canal, l'MPEG-TS permet enviar diversos simultàniament (cadascun amb una sincronització de rellotge independent). L'MPEG-TS incorpora a més mecanismes de detecció i correcció d'errors en la transmissió, tan habituals en xarxes públiques com en Internet o 3G.

Respecte a la transmissió per xarxa, es defineix un conjunt de protocols per a la transmissió de contingut multimèdia per una xarxa IP que veurem més endavant. Cal destacar, entre tots, per a sistemes de vídeo digital de temps real, el protocol RTP (*real time protocol*).

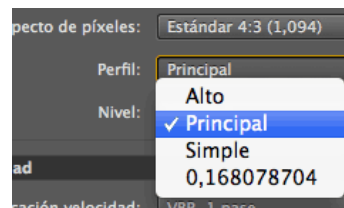
### 3.3.3. Perfils i nivells de MPEG-2

L'MPEG-2 va ser dissenyat per a abastar la màxima quantitat de sistemes de vídeo digital presents en el moment en què es va fer l'estàndard. Això implica disposar de diferents graus de qualitat per a cada aplicació: generalment determinats per la taxa de bits i la resolució de la codificació. Els graus de qualitat estan definits en l'estàndard en perfils (*profiles*) i nivells (*levels*).

### 1) Perfils

El perfil defineix la resolució de l'espai de colors i l'escalabilitat del flux de bits. Les opcions més habituals són les següents:

- **Perfil alt (*high profile*, HP):** destinat a codificar senyals d'alta definició. En realitat, mai s'utilitza, ja que l'MPEG-4 és molt més efectiu en aquest cas.
- **Perfil principal (*main profile*, MP):** el més habitual, s'adreça a aconseguir una qualitat estàndard de vídeo. Activa l'ús d'imatges I, P i B en la codificació, i permet codificar senyals de vídeo entrellaçats.
- **Perfil simple (*simple profile*, SP):** necessari quan l'*stream* generat va adreçat a equips de baixes prestacions tècniques (telefonía mòbil, PDA, tauletes, etc.). No utilitza imatges de tipus B, la qual cosa simplifica el càlcul de la descodificació i la necessitat de memòries intermèdies.



Perfils d'exportació (Adobe Premiere)

### 2) Nivells

El nivell defineix la resolució d'imatge, la quantitat de mostres de luminància (Y) per segon, la quantitat de capes d'àudio i vídeo i la màxima taxa de bits per perfil.

- **Nivell alt (*high level*, HL):** 1920x152, 80 Mbps. Per a alta definició, mai utilitzat.
- **Nivell alt 1440 (*high-1440*, H-14):** 1440x1152, 60 Mbps. També per a alta definició, mai utilitzat.
- **Nivell principal (*main level*, ML):** 720x576, 15 Mbps. L'habitual per a imatges de qualitat estàndard.
- **Nivell baix (*low level*, LL):** 288x352, 4 Mbps.



Nivells d'exportació (Adobe Premiere)

### 3) Combinacions de perfils i nivells

No s'utilitzen totes les combinacions de perfils i nivells. La notació utilitzada és perfil@nivell. Les més habituals són les següents:

Taula. Combinacions de perfils i nivells

| Perfil@nivell de resolució | Hz                     | Mostreig       | Mbps          | Exemple d'aplicació                       |
|----------------------------|------------------------|----------------|---------------|---|
| Sp@ll                      | 176 x 144              | 15 4:2:0       | 0.096         | Tauletes, mòbils                          |
| Sp@ml                      | 352 x 288<br>320 x 240 | 15 4:2:0<br>24 | 0.384         | PDA                                       |
| Mp@ll                      | 352 x 288              | 30 4:2:0       | 4             | Descodificadors                           |
| Mp@ml                      | 720 x 480<br>720 x 576 | 30 4:2:0<br>25 | 15 (DVD: 9.8) | DVD, TDT, televisió per cable i satèl·lit |



| Perfil@nivell de resolució | Hz                           | Mostreig       | Mbps | Exemple d'aplicació |
|----------------------------|------------------------------|----------------|------|---------------------|
| 422P@ml                    | 720 × 480<br>720 × 576       | 30 4:2:2<br>25 | 50   | Sony IMX            |
| 422P@h-14                  | 1.440 × 1.080<br>1.280 × 720 | 30 4:2:2<br>60 | 80   | Reservat            |
| 422P@hl                    | 1.920 × 1.080<br>1.280 × 720 | 30 4:2:2<br>60 | 300  | Reservat            |

#### 4) L'MPEG-2 en la indústria

L'MPEG-2 s'utilitza avui tant per a emmagatzemar com per a transmetre àudio i vídeo. Alguns dels sistemes que fan servir l'MPEG-2 són DVD i els estàndards de TDT, satèl·lit i cable a Europa (DVB) i Amèrica (ATSC). En cada cas s'implementa només la porció necessària de l'estàndard.

Uns exemples reals merament il·lustratius de l'MPEG-2 per a emmagatzematge són els següents:

##### a) DVD (*digital video disc*)

- Contenedor de vídeo: MPEG-2 Program Stream.
- Codificació de vídeo: MPEG-2.
- Codificació d'àudio: PCM, MP2 (Musicam), Dolby Digital o DTS a Europa, i Dolby Digital o DTS als Estats Units.

Resolució (px):

- NTSC: 720×480, 704×480, 352×480, 352×240.
- PAL: 720×576, 704×576, 352×576, 352×288.
- Taxa de bits d'àudio + vídeo: pic, 15 Mbps; mínim, 300 Kbit/s.
- Relació d'aspecte: 4:3, 16:9, 2.21:1.
- Taxa de quadres: 29.97 quadres/segons (NTSC), 25 quadres/segons (PAL).
- YUV 4:2:0.

##### b) DVB (*digital video broadcast*)

Inclou la televisió digital terrestre, satèl·lit i per cable.

Resolucions possibles:

- SDTV: 720, 640, 544, 480 o 352×480 píxels, 24/1.001, 24, 30/1.001 o 30 quadres/segons.
- 720, 704, 544, 480 o 352 × 576 píxels, 25 quadres/segon.
- 352 × 240 píxels, 24/1.001, 24, 30/1.001 o 30 quadres/segon.
- 352 × 288 píxels, 25 quadres/segon.

## 4. L'estàndard MPEG-4 i H.264/AVC

Després de l'MPEG-2, finalitzat el 1995, el Moving Picture Experts Group (MPEG) continua el seu treball d'estandardització. El 1993 (abans d'acabar l'MPEG-2) començar a treballar en l'estàndard MPEG-4, que abasta molts més aspectes dels sistemes de vídeo digital que els seus predecessors MPEG-1 i MPEG-2.

La part 2 de l'estàndard MPEG-4 (ISO/IEC 14496-2), coneguda com a *MPEG-4 Visual*, especifica la codificació i descodificació de vídeo digital, i es va acabar d'estandarditzar el 1999.

L'estàndard H.264 es va iniciar per part del grup de treball Video Coding Experts Group (VCEG) de la International Telecommunication Union (ITU-T). Les últimes etapes del treball les va dur a terme la Joint Video Team (JVT), un grup conformat pels grups VCEG i MPEG. L'estàndard final va ser publicat en conjunt el 2003. Per tant s'ha d'entendre que MPEG-4 i H.264 no són sinònims; H.264 és una part de l'estàndard MPEG-4.

### Nota

La publicació conjunta de l'estàndard el 2003 va ser per a MPEG com la part 10 de l'estàndard MPEG-4 (ISO/IEC 14496-10) i per a ITU-T com l'H.264.

L'H.264 també s'anomena *H.264 Advanced Video Coding (AVC)*, per tant, les formes correctes d'anomenar aquest estàndard són *MPEG-4 Part 10* o *H.264/AVC*.

L'MPEG-4 Visual i l'H.264, tot i que són contemporanis, són molt diferents en concepció: el primer apunta a la flexibilitat abastant la màxima quantitat de sistemes de vídeo digitals possibles, mentre que l'H.264 apunta a l'eficàcia en la compressió i a la confiabilitat de la transmissió.

La codificació utilitzada en la majoria dels sistemes d'IPTV és H.264, per això dedicarem aquest apartat exclusivament a aquesta part de l'estàndard MPEG-4. Per a entendre-ho millor, l'estructura d'aquesta secció es presenta de manera molt semblant a la de l'MPEG-2, començant pel detall dels estàndards que la componen, seguint per alguns detalls del format, els mecanismes de transmissió i emmagatzematge, els perfils i nivells definits i acabant amb l'aplicació actual en la indústria.

El contenidor d'aquest còdec està ben especificat, i es basa en un format estàndard que curiosament es va inspirar en el de QuickTime d'Apple: l'ISO Media File Format. L'extensió dels arxius MPEG-4 és .mp4.

#### 4.1. L'estàndard i les seves patents

L'MPEG-4 està estandarditzat en la norma ISO/IEC 14496. Com l'MPEG-2, es compon de diverses parts, entre les quals hi ha les següents:

- El sistema: descriu la sincronització i la transmissió simultània d'àudio i vídeo.
- El vídeo: el còdec de compressió per a elements visuals (vídeo, textures, imatges sintètiques, etc.). Un dels molts perfils definits en la part 2 és l'ASP (*advanced simple profile*).
- L'àudio: el conjunt de còdecs de compressió per a la codificació de fluxos d'àudio; inclouen variants d'AAC (*advanced audio coding*) i eines de codificació d'àudio i parla.
- Transport sobre xarxes IP: especifica un mètode per a transportar contingut MPEG-4 sobre xarxes IP.
- H.264-AVC (*advanced video coding*): un còdec estandarditzat de senyals de vídeo.
- Enginyeria d'aplicació i descripció d'escenes (BIFS): per a contingut interactiu 2D i 3D.
- Format per a mitjans audiovisuals basat en ISO: un format d'arxius per a emmagatzemar contingut multimèdia.
- Extensions per a la gestió i protecció de propietat intel·lectual (IPMP).
- El format d'arxiu MPEG-4: el format d'arxiu de contenidor designat per a continguts MPEG-4.
- El format d'arxiu AVC: per a emmagatzemar vídeo.

És important ressaltar la part 10 de l'estàndard, que especifica l'H.264; i la part 8, que especifica el transport de MPEG-4 sobre una xarxa IP.

Com les parts 1 i 2 de l'MPEG-2, per a utilitzar i explotar aquest estàndard cal pagar llicència a la institució MPEG-LA.

#### 4.2. H.264 o MPEG-4 Part 10

En filosofia, la codificació H.264 no difereix gaire de l'MPEG-2. S'especifica el format de codificació i com descodificar-lo (i es deixa un cop més lliure la indústria per a fer codificadors eficients).

També utilitza la nomenclatura de *perfils* i *nivells* per a definir variants de resolució i qualitat de vídeo. Actualment relativament pocs perfils es troben completament definits en l'estàndard, al començament només van ser definits tres perfils: *baseline*, *main* i *extended*; no obstant això avui sembla que es perfila amb més acceptació al mercat el perfil *high*, com ja veurem.

Una millora respecte a l'estàndard MPEG-2 és que se separa la codificació de la transferència, i això en facilita la implementació.

El format de sortida del codificador es diu VCL (*video coding layer*), que bàsicament és una seqüència de bits representant vídeo codificat. La jerarquia de codificació de vídeo MPEG-2 (GOP, quadres, blocs) queda pràcticament inalterada.

Un problema habitual de l'MPEG-2 és que si hi ha diferents subfluxos o serveis continguts en el mateix flux de bits, en commutar d'un a un altre el descodificador perdia qualitat uns instants. Contra això, en els perfils més alts de l'H.264, hi ha la possibilitat d'utilitzar dos tipus extra de quadres (a més dels I, P, B):

- **SP (*switching P*)** per a facilitar canviar de flux de codificació (conté macroblocs I o P)
- **SI (*switching I*)** per a facilitar canviar de flux de codificació (conté macroblocs I).

### 4.3. Emmagatzematge

L'estàndard MPEG-4 fa una especificació d'emmagatzematge en disc, com ja hem dit. Es basa en l'estàndard ISO Media File Format, que s'inspira en el format QuickTime d'Apple. L'extensió dels arxius MPEG-4 és .mp4.

Molts altres formats contenidors tenen aquesta codificació, com per exemple 3GPP, MOV o Flash Video.

### 4.4. Transmissió

Tant l'MPEG-4 com l'H.264 no defineixen un sistema específic de transport per a la transmissió.

Amb l'esmena 3 del sistema MPEG-2 (ISO/IEC 13818-1) és possible transportar MPEG-4 i en particular H.264 de la manera següent:

- *Program streams* (conegut com a *MPEG-2 PS* o simplement *MPEG-PS*).

- *Transport streams* (conegut com a *MPEG-2 TS* o simplement *MPEG-TS*), que permet fer conviure canals TDT en H.264/AVC en HD amb els MPEG-2 en SD.

Respecte a la transmissió per xarxa IP, s'ha actualitzat el protocol de temps real RTP (*real time protocol*) per a la transmissió de MPEG-4 i H.264. És relativament senzill transmetre contingut multimèdia en una xarxa de paquets IP, fins i tot si ens limitem al context d'IPTV no només s'utilitza RTP. A les xarxes actuals per a transmissió de vídeo digital per Internet, hi ha una clara preferència per protocols basats en TCP i en particular els basats en HTTP perquè són compatibles amb les configuracions freqüents de tallafocs i servidors intermediaris.

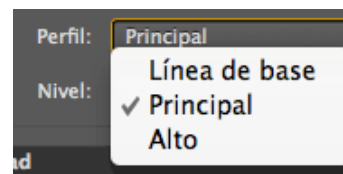
#### 4.5. Perfils i nivells

Com l'MPEG-2, l'estàndard H.264/AVC té diferents graus de qualitat. Els graus de qualitat estan definits en l'estàndard en perfils (*profiles*) i nivells (*levels*).

##### 1) Perfils

Hi ha set perfils en l'estàndard, però ens centrarem en els tres habituals, indicats en l'Adobe Premiere:

- **Línia de base** (*baseline profile, BP*): usat per a dispositius senzills com reproductors portàtils i telefonia mòbil, i per a transmissió en temps real de vídeo per xarxa IP.
- **Principal** (*main profile, MP*): dirigit a difusió (TDT HD) i emmagatzematge (Blu-ray), proveeix alta definició.
- **Alt** (*high profile, HP*): per a aplicacions professionals que requereixen alta definició de crominància i luminància (4:4:4).



Perfils d'exportació (Adobe Premiere)

##### 2) Nivells

Representats amb un nombre de l'1,0 al 5,1, són configuracions per a velocitats de transmissió creixents (des de 64 kbps en 1.0 fins a 240 Mbps en 5,1).

#### 4.6. Utilització de l'H.264 en la indústria

L'H.264 sempre és triat en sistema de vídeo digital d'alta resolució, com les xarxes d'IPTV i la televisió en alta definició terrestre, per cable o per satèl·lit. S'espera que amb el pas del temps els diferents sistemes de transmissió de vídeo migrin a aquest format i reemplaçin els tradicionals H.262 i MPEG-2, tal com ha passat a França, que des del començament ha apostat per aquest format en les emissions TDT.

Alguns dels sistemes que utilitzen o plantegen utilitzar H.264 són els que s'esmenten a continuació.

##### 1) Per a emmagatzematge

Blu-ray Disc del Blu-ray Disc Association (BDA). Utilitza un perfil alt, HP. Com a característiques més conegudes de Blu-ray tenim les següents:

- Contenedor de vídeo: BDAV MPEG-2 Transport Stream (fitxers d'extensió .bdav), amb el qual l'enregistrament d'una emissió de TDT en HD pot emmagatzemar-se directament sense alterar el format en absolut.
- Codificació de vídeo: H.264/MPEG-4 AVC (o VC-1 de Microsoft).
- Codificació d'àudio: Dolby Digital, DTS, Dolby Digital Plus, DTS-HD, PCM.

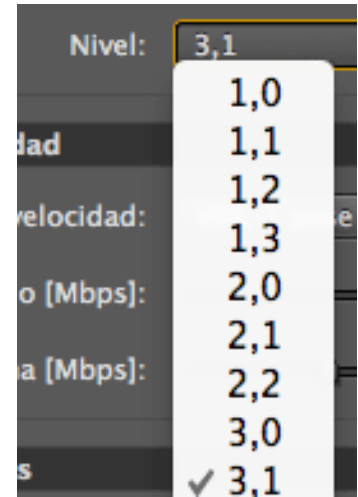
Combinacions de resolució i taxa de quadres en modes progressius:

- **1080p**: 1.920 × 1.080 píxels a 24/23,97 quadres/segons (Estats Units/Europa).
- **720p**: 1.280 × 720 píxels a 50/59,94 quadres/segons (Estats Units/Europa).
- Altres combinacions progressives (1.440 × 1.080, 1.280 × 720) a 24 quadres/segons (mode cinema).

Combinacions de resolució i taxa de quadres en modes entrelaçats:

- **1080i**: 1.920 × 1.080 píxels a 25/29,97 quadres/segons (Estats Units/Europa), o 1.440 × 1.080 píxels a 25/29,97 quadres/segons (Estats Units/Europa).
- **720i**: 1.280 × 720 píxels a 25/29,97 quadres/segons (Estats Units/Europa).

Relació d'aspecte 16:9 (encara que hi ha dos modes 4:3 poc utilitzats). El pic de la taxa de bits d'àudio + vídeo és de 36 Mbps; i el YUV, 4:2:0.



Nivells d'exportació

El format d'emmagatzematge HD-DVD de Toshiba, ja desaparegut, tenia característiques similars.

## **2) Per a difusió**

El consorci europeu DVB des de l'any 2004 va apostar per aportar (si no reemplaçar) a la codificació H.262/MPEG-2 actual aquest estàndard nou, tant en terrestre (DVB-T) com satèl·lit (DVB-S) i cable (DVB-C). El nou estàndard DVB-T2, que s'implantarà a partir de l'any 2015 a Europa, optimitzarà l'aprofitament de les freqüències de canal per a donar cabuda a més canals, cosa que podria facilitar la migració a H.264.

El comitè nord-americà ATSC especifica H.264 i VC-1 com a estàndards per a la transmissió terrestre de televisió. Igualment l'associació ARIB japonesa incorpora H.264/AVC en el seu sistema de TDT (ISDB-T).

En televisió per satèl·lit és present en la majoria de plataformes (Digital+, BBC HD, Euro1080, etc.).

## **3) Per a xarxa de telefonia cel·lular mòbil**

The 3rd Generation Partnership Project (3GPP) inclou l'H.264/AVC com a opcional. Actualment (l'any 2011) és infreqüent poder rebre aquest servei a causa de l'amplada de banda necessària.

## 5. Transmissió de vídeo en una xarxa IP (*streaming*)

Les tècniques per a la transmissió de vídeo a les xarxes IP habitualment són conegudes com a *flux de dades* o *streaming*, que es podria definir com la transmissió en temps real (en directe) d'àudio i vídeo sobre una xarxa.

Abans que apareguessin les tècniques de transmissió en temps real, les aplicacions multimèdia usaven Internet només per a transferir arxius. Així, una vegada que els continguts eren descarregats completament podien ser reproduïts. Actualment, amb la transmissió en temps real, és possible anar visualitzant un contingut multimèdia a mesura que és transferit.

Com hem vist en els estàndards digitals anteriors, els fluxos de dades d'àudio i vídeo s'empaqueten formant cadascun un flux elemental (ES) que a més guarda informació de sincronia entre si. A continuació es combinen (almenys un de vídeo i un altre d'àudio) en un flux elemental empaquetat (PES), i es possibilita que se'n combinin més.

A partir d'aquest moment podem fer dues coses:

- Emmagatzemar aquest contingut en un contenidor per a reproduir-lo més endavant.
- Transmetre'l (en el nostre cas per una xarxa IP).

En apartats anteriors hem vist per sobre com es condicionen els diferents *streams* per a ser emmagatzemats, ara veurem com enviar-los per la Xarxa.

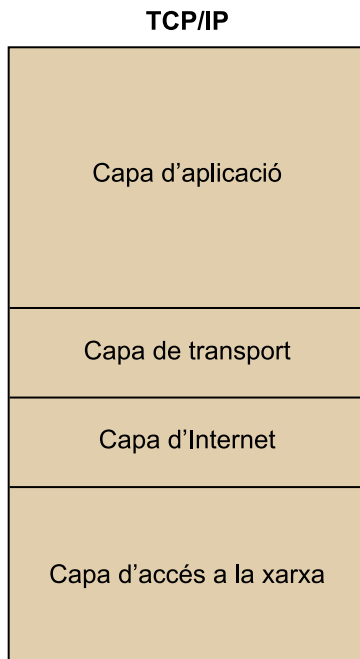
Hi ha múltiples tècniques de transmissió en temps real segons la realitat particular del sistema de vídeo digital. En aquest apartat veurem les més importants en el context de les xarxes basades en IP, com 3G, IPTV i Internet. Farem una introducció als protocols de les xarxes IP (sense aprofundir-hi, ja que no és objectiu d'aquesta assignatura) i als protocols de transmissió en temps real després.

### 5.1. Les xarxes IP

Una xarxa pot ser analitzada des de diferents punts de vista. Per exemple, la podríem estudiar elèctricament, mesurant les tensions dels senyals que hi circulen i els tipus de connectors necessaris. Aquest punt de vista s'anomena *físic* o *d'accés a la Xarxa*, i és més propi d'investigadors i instal·ladors.



## Esquema



Arquitectura de xarxa TCP/IP

Un segon punt de vista seria el d'enllaç, que estudia la identificació de les targetes de xarxa i el protocol d'intercanvi de bits entre aquestes. També és més propi de desenvolupadors i programadors, situats en un ús més abstracte.

Els punts de vista que ens interessin són els dos següents: el de Xarxa (en el qual parlem de dades que hi circulen des d'un equip amb una adreça a un altre o uns altres), i el de transport (mecanismes per a assegurar que les dades arriben correctament i, si no ho fan així, es demana reexpedició de les dades o no). Sobre aquesta quarta i última capa ja tindríem les aplicacions informàtiques que usen la Xarxa. En aquesta secció fem esment exclusivament a aquestes dues últimes capes (xarxa i transport) de les xarxes IP, amb la finalitat d'entendre els mecanismes de transmissió en temps real presentats seguidament.

### 5.1.1. Protocol de xarxa (*Internet protocol, IP*)

Les xarxes IP són un territori inestable i sempre canviant de ruta des d'un punt a un altre. Perquè un paquet de dades arribi a la destinació, ha de tenir-la indicada en cadascun dels paquets que conformen les dades que es volen enviar. El protocol IP és un protocol robust, que no espera que hi hagi un camí assegurat i estable entre origen i destinació, sinó que les dades es divideixen en paquets que s'envien de manera consecutiva i cadascun "navega" en una xarxa plena de paquets fins que arriba a la destinació que tenen marcada.

Dit tècnicament, el protocol IP és **no orientat a connexió per a la comunicació mitjançant una xarxa de paquets commutats**. És el protocol d'Internet i dels sistemes d'IPTV.

Aquests paquets són enviats amb informació extra per a poder fer accions relatives a garantir la seva integritat, però no assegura que les dades realment arribin a destinació (d'això se n'encarrega el protocol superior, el protocol de transport).

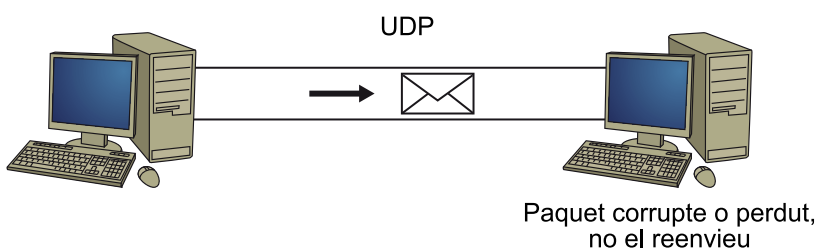
### 5.1.2. Protocols de transport (UDP, TCP i SCTP)

Per sobre d'aquest "carter" que és el protocol IP, sempre hi ha un protocol de transport que té la responsabilitat d'assegurar la transferència sense errors de dades entre l'emissor i el receptor. No obstant això, ha de ser àgil o la seva acció pot alentir la celeritat amb què es reben les dades. Els bits d'aquest protocol van dins del paquet de xarxa.

Depenent del tipus de servei que vulguem, podem triar un dels tres protocols de transport següents:

#### 1) UDP (*user datagram protocol*)

Protocol lleuger que no dóna garantia sobre la pèrdua de paquets ni la recepció per duplicat de paquets. Només té mecanismes de revisió mèdica de dades, si cal gestionar els errors en la transmissió s'hauran de fer a la banda del destinatari (en l'aplicació que els rebí). És el protocol de transport natural per a transmissió en temps real de vídeo i àudio.

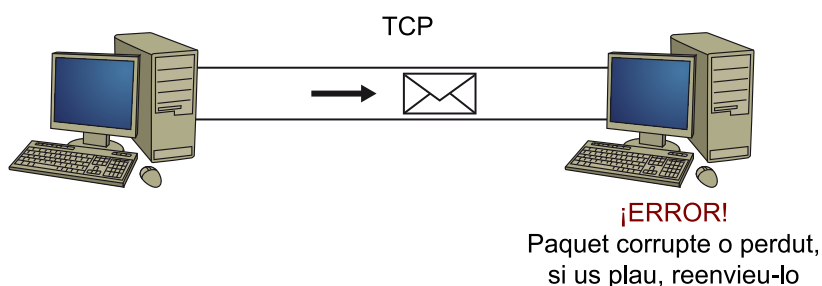


En el protocol UDP una pèrdua, error o duplicació de paquet no genera cap petició de reexpedició per part del destinatari

La raó per la qual moltes videoconferències, webs pirates de sèries de televisió no estrenades a Europa o les sessions de Spotify fan servir aquest protocol és perquè consumeix pocs bits del total de bits de dades i és ràpid a inicialitzar-se (ja que no fa gaire gestió). Certament és menys segur que uns altres, però en transmissió en temps real es considera inútil una retransmissió de la informació demanada, ja que aquesta arribaria a destemps (la transmissió en temps real presenta restriccions de temps real).

## 2) TCP (*transmission control protocol*)

Aquest protocol ens assegura que els paquets arriben i que ho fan en ordre. Això s'aconsegueix amb un mecanisme en el qual quan no es rep la confirmació del paquet per part del destinatari durant cert temps, aquest paquet es reenvia. Per això és possible detectar paquets perduts i demanar que es reenviïn. En el cas de transmissió de *stream*, quan es perden paquets, la retransmissió augmenta el retard i el consum de l'amplada de banda, la qual cosa pot provocar que es buidi la memòria intermèdia del reproductor (i que per tant s'interrompi la reproducció).



En canvi, en el protocol TCP s'assegura la recepció correcta de tots i cadascun dels paquets

En xarxes sense mecanismes per a assegurar la qualitat de servei, com Internet, avui s'opta per la transmissió en temps real sobre TCP perquè pèrdues esporàdiques de paquets repercuteixen en la qualitat percebuda de manera molt menys sensible en TCP que en UDP. En canvi, en xarxes d'IPTV, amb qualitat de servei i sense congestió, el mecanisme de transmissió en temps real encara és el tradicional basat en UDP.

## 3) SCTP (*stream control transmission protocol*)

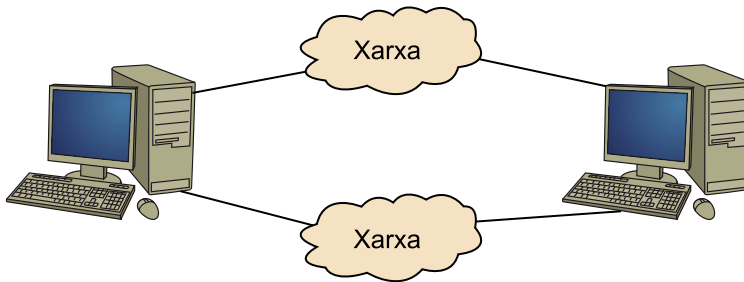
Aquest protocol de transport és una solució de compromís entre els tradicionals UDP i TCP. Tot i que està completament estandarditzat ha tingut una difusió lenta. Té una utilitat potencial en la transmissió en temps real de vídeo en xarxes sense fil i altres contextos amb grans pèrdues puntuals de dades.

### Mecanisme NAT

Aquesta difusió lenta és deguda especialment al fet que és incompatible amb el mecanisme NAT (*network adress translation*, 'traducció de l'adreça de xarxa'). Aquest mecanisme és el que, per exemple, permet que una xarxa local d'una empresa amb els equips amb adreces IP locals puguin comunicar-se amb l'exterior amb equips remots amb adreces IP completament diferents. L'encaminador de l'empresa s'encarrega del NAT, i fa possible que tots accedeixin a Internet.

Aquest protocol es basa en datagrames, però a diferència de l'UDP té un registre dels nombres de seqüència amb els quals fa controls de paquets fora d'ordre i pèrdues que seran retransmeses. A diferència del TCP, aquest protocol permet, en una mateixa connexió (origen SCTP, destinació SCTP), diferents

adreces IP. O sigui, que cadascun dels extrems de connexió pot tenir diferents IP (mantenint el port de connexió) i els paquets seran enviats indistintament per qualsevol IP i a qualsevol IP corresponent a cada extrem SCTP.



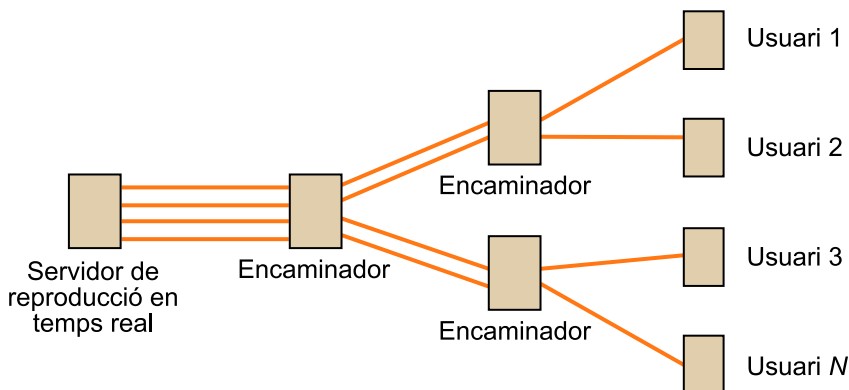
El protocol SCTP possibilita diferents adreces IP per un mateix port

També suporta el trànsit sobre la mateixa connexió de diferents tipus de paquets mantenint el nombre de seqüència per a cadascun.

## 5.2. Tècniques de multidifusió

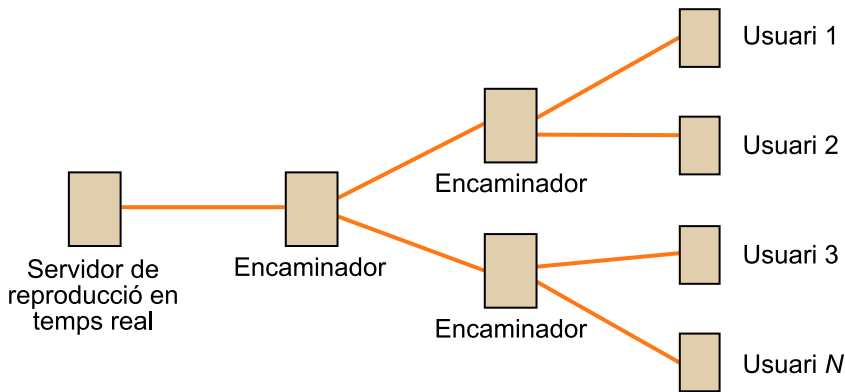
A les xarxes IP hi ha diferents tècniques per a difondre la informació, que enumerarem tot seguit:

1) **Unidestinió:** l'enviament d'un punt a un altre en una xarxa s'anomena *unidifusió*. El mecanisme habitualment utilitzat per totes les aplicacions té un nombre N màxim d'usuaris limitat.



N usuaris visualitzant simultàniament un programa de televisió per Internet. Amb la tècnica d'unidestinió el servidor ha d'enviar l'*stream* per a cada usuari

2) **Multidestinió:** enviament de la informació a múltiples destinacions simultàniament amb l'ajuda d'encaminadors preparats que, si estan enviant l'*stream* a un usuari, en rebre una altra petició, en dupliquen les dades. El servidor queda molt més descarregat, però obliga a tenir encaminadors capacitats per a aquest protocol.



Amb la tècnica de multidestinatció, el servidor envia inicialment un *stream* al primer usuari. Els altres reben el seu per reexpedicions dels encaminadors, que s'adonen que hi ha més d'un usuari amb la mateixa petició

**3) Difusió:** l'enviament a tots els nodes en una xarxa es denomina *difusió àmplia*. Lògicament només s'aplica en contextos reduïts, per exemple per al descobriment de recursos en una xarxa local (LAN).

La tècnica de multidestinatció és fàcil de descriure i difícil d'aplicar. La multidestinatció ha d'usar l'estratègia més eficient per a enviar els missatges sobre cada usuari de la xarxa (amb la finalitat de disminuir el consum d'amplada de banda), utilitzant cada enllaç com a màxim una vegada perquè cada paquet sigui difós i creant còpies en els encaminadors quan els enllaços en les destinacions es divideixen. Habitualment això s'aconsegueix amb **arbres d'expansió**, i sorgeix un protocol per a mantenir aquest arbre a través dels encaminadors.

Des del punt de vista de l'emissor i els receptors el procés és senzill: l'emissor envia la informació una sola vegada a una adreça IP especial (de manera idèntica a una difusió unidestinatció); els receptors se subscriuen al grup de multidestinatció associat a aquesta IP especial i reben una còpia de tot el que envia l'emissor. Per tant, abans de l'enviament de la informació en multidifusió, s'han d'establir una sèrie de paràmetres.

Per a poder rebre-la, cal establir el que es denomina *grup multidestinatció*. En IPv4 les IP especials pertanyen al rang 224.0.0.0 a 239.255.255.255 i el protocol utilitzat per a gestionar i associar-se als grups de multidestinatció es diu *Internet group management protocol* (IGMP).

Depenent de la informació es tria el mecanisme ideal per a difondre-la. En **IPTV**, la transmissió de canals en directe habitualment es fa amb multidestinatció, ja que s'espera que hi hagi molta audiència (l'enviament individual a cada espectador –unidestinatció– suposaria un dimensionament inadmissible de la xarxa).

En canvi, el **vídeo sota demanda (VoD)**, per naturalesa, és punt a punt i per tant se sol utilitzar la tècnica d'unidestinatció. Però això no és una regla, inclusivament en el context d'IPTV, per exemple, en els casos següents:

- Hi ha tècniques per a reduir el temps de zàping entre els canals en directe (enviats per multidestinació) que impliquen l'enviament unidestinació d'una ràfega inicial de gran amplada de banda per a omplir ràpidament la memòria intermèdia del receptor i reproduir ràpidament després del canvi de canal.
- D'altra banda, en els sistemes de VoD, quan un contingut és molt sol·licitat sota demanda, pot ser reproduït en forma de nVoD (*near VoD*), que implica l'enviament multidestinació i una espera inicial d'uns minuts per a rebre la propera transmissió multidestinació. Una tècnica una mica més sofisticada implica enviar una bona porció del VoD per unidestinació i emmagatzemar a la memòria intermèdia la transmissió multidestinació antiga més propera en el temps. Això evita l'espera inicial del sistema nVoD però exigeix tenir un espai de memòria molt més gran per a emmagatzemar. Aquest tipus de mecanisme mixt encara s'ha d'investigar més.

Com a conclusió, en els sistemes de vídeo digital especialitzats com IPTV, la utilització eficient de les tècniques de difusió suposen en una qualitat d'experiència més bona i més aprofitament dels recursos de la xarxa.

La situació en el context d'Internet és molt diferent. Al començament d'Internet, la multidestinació no estava estandarditzada i era difícilment suportada pels sistemes operatius i maquinari dedicat de l'època. Avui, per raons comercials, es manté aquesta restricció sobre Internet, per tant, no se suporta multidestinació (i per raons lògiques d'escala tampoc la difusió). Una esperança per al suport de multidestinació a Internet és amb el protocol IPv6, la nova versió del protocol IP, que nativament el suporta.

Si Internet suportés multidestinació, els sistemes dedicats d'IPTV presentarien una forta competència, que seria una oferta de contingut global per a un públic global (per aquest motiu no és atractiu per als proveïdors de connectivitat oferir aquests beneficis als proveïdors de serveis d'Internet). Els sistemes de vídeo a Internet actuals fan, segons les seves possibilitats, una multidestinació en capa d'aplicació, amb el desplegament global de centres de dades i la rèplica intel·ligent de la informació.

Tot i que són tècniques de multidestinació més eficients, el problema subsisteix en el tram més compromès: la xarxa d'accés, en què aquestes xarxes només poden fer distribució unidestinació (YouTube, Google Video, Netflix, etc.).

### 5.3. Protocols de transmissió en temps real

Hi ha una tercera categoria de protocols dissenyats específicament per a transmissió en temps real, coneguts com a *protocols de temps real*, que intenten assegurar l'enviament de fluxos de dades amb una cadència assegurada, tal com demana el material audiovisual.



Tant la multidestinació com la difusió estan restringides actualment a Internet

### 5.3.1. Protocols de temps real (RTP, RTCP i RTSP)

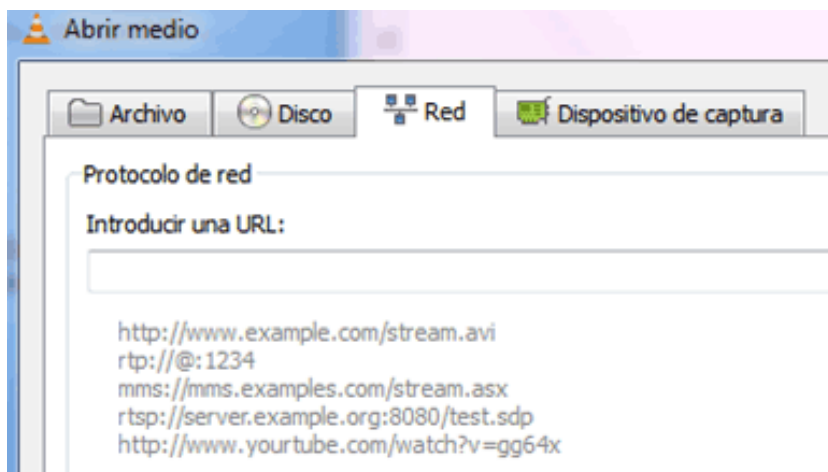
Els protocols de temps real són els següents:

1) **RTP (*real time transport protocol*)**. És un protocol de transport desenvolupat per transmissió en temps real, usualment muntat sobre UDP. El transport de RTP sobre TCP també està definit, però de totes maneres s'utilitza menys perquè té més latència i sobrecàrrega que l'UDP, cosa que suposa un desavantatge per a aplicacions de temps real. Inclou dades extra no presents en el TCP, com marcatge de temps i nombre de seqüència per si arriba desordenadament o si es perd alguna cosa, fet que contribueix a facilitar el transport de manera contínua. També hi ha dades de control que permeten al servidor fer la transmissió en temps real a una taxa de flux definida i correcta.

Lamentablement, el protocol RTP no té informació de sincronització, ni mecanismes per a recuperar errors nadius. Una sèrie de còdecs es troben estandaritzats per a poder ser enviats sobre RTP, diversos dins de la línia MPEG-2.

#### Marcació de temps

Gràcies a la marcatge de temps, es guarda en cada paquet l'hora en què és transmès a partir d'un rellotge de referència, cosa que facilita la sincronització.



En indicar en l'aplicació l'adreça de la font de transmissió en temps real, cal indicar el protocol de temps real en el qual s'està transmetent (a la il·lustració, el programa VideoLAN)

2) **RTCP (*real time transport control protocol*)**. És usat en conjunció amb el protocol RTP per a rebre informes estadístics. Permet, per exemple, detectar errors en l'arbre de distribució de clients multidestinació, el nombre de paquets perduts i estadístiques de fluctuacions (*jitters*).

Hi ha informes enviats per l'emissor i informes enviats pel receptor. Per exemple, els de l'emissor són habitualment el volum de quadres enviats o quantitat de bytes, i els informes del receptor presenten els quadres perduts o la taxa de quadres lliurats.

Una característica particular que té aquest protocol és que per a amortitzar la sobrecàrrega que suposa consumir bits per al protocol i no per a dades (*overhead*), es poden reunir diversos missatges RTCP i enviar-los en un missatge RTCP compost. Els paquets han d'estar compostos almenys per un missatge de

receptor o emissor i el nom del participant, i han de ser enviats periòdicament sense arribar a consumir el 5% de l'amplada de banda de la sessió. Els paquets RTCP es transporten sobre UDP.

El protocol RTCP no permet encriptació, autenticació ni autorització. Una actualització amb aquestes funcionalitats és l'estàndard SRTCP (*secure real time transport protocol*).

**3) RTSP (*real time streaming protocol*).** Controla dades multimèdia de temps real, i ofereix la possibilitat d'interactuar amb el reproductor, més o menys com un vídeo reproductor domèstic. El protocol RTSP permet reproduir, posar la pausa, avançar i més. També pot reaccionar a congestions a la xarxa i reduir l'amplada de banda.

Els comandaments estandarditzats en RTSP són els següents:

- **Describe:** amb aquest comandament es rep una descripció del recurs, una llista dels *streams* de dades de vídeo necessaris.
- **Setup:** s'indiquen els paràmetres de configuració de l'*stream*.
- **Play:** reproduïx l'*stream* especificat.
- **Pause:** atura la reproducció de l'*stream* indicat, que es pot reiniciar amb un *play*.
- **Teardown:** atura i allibera els recursos utilitzats, i finalitza l'*stream*.

El protocol RTSP s'inspira en el protocol web (HTTP 1.1) però el millora perquè pot mantenir l'estat de la connexió (l'HTTP no manté estat) i perquè tots dos (client i servidor) poden fer comandes. L'RTSP suporta l' RTP com a protocol de transport.

Una de les utilitats que té és oferir una forma inicial de triar el canal de distribució òptim cap al client.

#### **Triar el canal**

Alguns clients poden tenir filtrats en el tallafoc els paquets UDP, i per això el servidor de reproducció en temps real hauria de proveir la possibilitat d'escollir entre diferents protocols de transport com UDP o TCP o UDP multidesinació.

### **5.3.2. Altres protocols de transmissió en temps real**

Encara ens queda enumerar uns quants protocols de temps real més, que són els següents:

**1) RDT (*real data transport*).** És un protocol propietari per a transmetre àudio/vídeo desenvolupat per RealNetworks el 1995. Com l'RTP, treballava en conjunt amb l'RTSP per a controlar la transmissió en temps real, i com que no tenia més avantatges que el seu equivalent estàndard RTP, la seva continuïtat està en dubte.



**2) HTTP (*hypertext transfer protocol*).** Tot i que va ser dissenyat amb finalitats completament diferents (és el protocol pel qual podem fer navegació web), l'HTTP s'utilitza per la transmissió en temps real d'àudio i vídeo per diferents raons. El motiu principal és que la majoria de les connexions a Internet està protegides per un tallafoc o un servidor intermediari, per la qual cosa la manera més senzilla de transmetre en temps real un contingut en aquest context és amb l'HTTP.

Lògicament té més sobrecàrrega (*overhead*) que la família de protocols de temps real (RTP, RTCP i RTSP) i per això no és ben considerat en molts contextos, per exemple pel grup MPEG. No obstant això, com ja s'ha dit, és una opció vàlida per a Internet, perquè hi ha tallafocs i servidors intermediaris, i té una manera simple d'afrontar els errors per pèrdues a la xarxa (mitjançant retransmissió).

La transmissió en temps real (*streaming*) sobre HTTP no està gaire estandaritzada, tot i que és la més utilitzada per les ràdios a Internet (<http://www.shoutcast.com/>).

Actualment amb HTTP es poden transmetre diversos tipus de formats de vídeo. A més de fluxos de dades de vídeo crus (com els acceptats per RTP), amb HTTP es poden enviar MPEG-TS, MPEG-PS, OGG, ASF i altres.

**3) MMS (*Microsoft media services protocol*).** És un protocol propietari per a transmetre àudio i vídeo. Desenvolupat per Microsoft, va ser abandonat oficialment el 2003 i va deixar de rebre suport el 2008.

Operava sobre TCP, UDP o HTTP (client i servidor negociaven quin triaven), o el servidor triava unívocament segons l'estat de la xarxa. Si el client no podia negociar una bona connexió utilitzant MSS sobre UDP (abreujada *MMSU*) llavors ho intentava amb MMS sobre TCP (abreujada *MMST*). Si això fallava, utilitzava una versió modificada de HTTP per a establir la connexió (abreujada *MMSH*). En cada pas anava creixent la sobrecàrrega i es feia més ineficient.

## 6. MPEG-2 TS (*transport stream*)

L'MPEG TS és un protocol que ofereix un mecanisme per a **multiplexar**, és a dir, combinar en un sol *stream* els *streams* d'àudio, vídeo i dades, i poder transmetre'ls per una xarxa. És molt utilitzat en els sistemes de vídeo digital de televisió per cable i TDT.

Com ja hem dit, un *flux elemental* o *elementary stream* (ES) és bàsicament la sortida del codificador. Hi ha ES de dos tipus: els de vídeo (VES) i els d'àudio (AES). Per a gestionar els diferents ES, són dividits en paquets de mides diferents, segons les característiques de l'aplicació i del descodificador. El procés de partició en paquets de l'ES s'anomena *empaquetat* i un ES empaquetat és un *packetized elementary streams* (PES).

Quan s'utilitza l'MPEG-TS, diversos PES poden ser transmesos de manera conjunta en un mateix *stream* TS, procés que s'anomena *multiplexació*. En la majoria de les aplicacions cal àudio i vídeo, i per tant es requereix la multiplexació d'almenys un PES d'àudio i d'un PES de vídeo. Els PES transmesos dins d'un TS són agrupats en programes. Habitualment un *stream* TS transmet diversos canals de televisió simultàniament, cadascun és un programa format per almenys un PES de vídeo i un PES d'àudio. A més dels PES, un *stream* TS transmet informació de control en format de taules.

L'MPEG-TS no ofereix simplement una manera adequada de fer la multiplexació dels diferents ES, sinó que també ataca el problema de recrear el rellotge de la font en cada receptor, per aconseguir així una descodificació i un sincronisme correctes de l'àudio i del vídeo.

### 6.1. Paquets i taules del *transport stream*

Els paquets TS tenen una mida fixa de 188 bytes. Un flux TS està compost per un programa o més; per a identificar a quin PES pertany a quin programa, el TS envia periòdicament taules amb aquesta informació.

Dins de la capçalera dels paquets TS hi ha un camp anomenat *packet identifier* (PID) que identifica el contingut específic d'aquest paquet TS. El contingut d'un paquet TS pot ser:

- part d'un PES d'àudio/vídeo
- informació de control, conegut com a *program specific information* (PSI)

#### Taules del flux transport stream

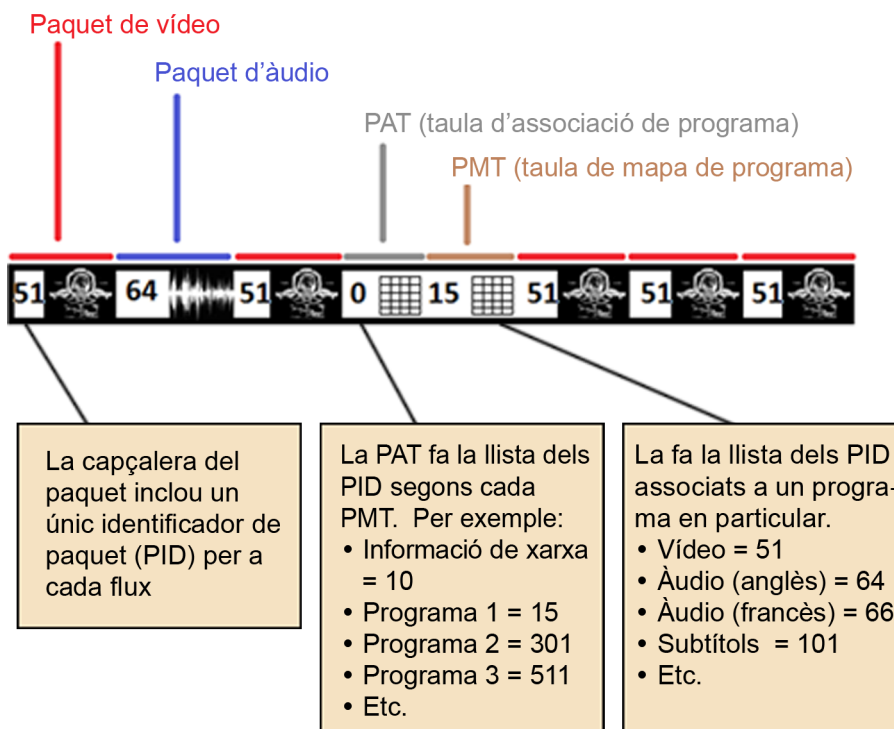
En aquestes taules es pot, per exemple, enviar informació de la programació present en el flux.

La PSI està composta per quatre taules diferents:

- la *program association table* (PAT)
- la *program map table* (PMT)
- la *conditional access table* (CAT)
- la *network information table* (NIT)

En aquestes taules hi ha tota la informació necessària per a extreure correctament cada *stream* de vídeo, àudio i dades i seleccionar així els diferents programes en el receptor. Les taules PAT i PMT sempre són presents dins de l'*stream* TS, mentre que les altres dues poden dependre de l'*stream* TS que s'està transmetent.

A la taula PAT (identificada amb PID = 0 × 0000), s'associa el nombre de programa amb el PID de la PMT per a aquest programa. Per exemple, la figura següent mostra que en els paquets amb PID número 15 seran transmesos els paquets corresponents a la PMT del programa 1. Al seu torn a la taula PMT, i entre altres coses, hi ha quins PID i, per tant, quins ES estan associats a quins programes; en aquest exemple es defineix que el PID número 64 és un ES d'àudio corresponent al programa número 1.



## Resum

La captació d'àudio i vídeo és més difícil que la captura d'imatge: la naturalesa entrelaçada del mostreig i els tipus de mostres de luminància i crominància segons la densitat d'informació desitjada per píxel. Però encara més difícil, si és possible, és la gestió del volum de dades digitals generades en la captura, que obliga a aplicar tècniques de compressió.

Hi ha definides dues grans famílies de compressió d'àudio i vídeo: les que generen pèrdues d'informació en la compressió i les que no en generen. L'estudi de la resposta humana davant la redundància espacial i temporal fa possible trobar tècniques de compressió amb pèrdues que minimitzen la informació visual perduda, per la qual cosa actualment són les més utilitzades. En el cas de l'àudio, el model psicoacústic fa possible el mateix criteri.

Els estàndards de compressió d'àudio i vídeo es conjunten en un contenidor, un format d'emmagatzematge de fitxer o de *flux* de dades que en facilita l'intercanvi i interoperabilitat entre diferents usuaris. Entre aquests estudiem amb detall l'MPEG-2, que es caracteritza per una estructura de grups d'imatges, combinació de tres tipus de quadres (P, B i I) i un sistema definit de perfils i nivells. Les tècniques de compressió espacial i detecció de moviment utilitzades són un compromís entre grau de compressió i la propagació d'errors visuals. Segons tot això, l'MPEG-4 AVC/H.264 té millores substancials que li confereixen més qualitat a velocitats de transmissió més baixes.

Finalment, s'enumeren els problemes relatius en la transmissió en directe de dades audiovisuals per xarxes de dades, es destaquen els protocols de transport i de temps real seleccionables per a assegurar la cadència de la recepció en situacions de pèrdua, error o duplicació de dades, i s'analitzen els mecanismes de transmissió de dades entre els encaminadors de la xarxa per a reduir la càrrega inherent sobre la infraestructura de comunicacions i, per tant, el cost.